

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

Ruchi Lohia,[‡] Reza Salari,[‡] and Grace Brannigan^{*,‡,¶}

[‡]*Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA*

[¶]*Department of Physics, Rutgers University, Camden, NJ, USA*

E-mail: 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 mutation encodes a hydrophobic-to-hydrophobic mutation at the midpoint of the prodomain of precursor brain-derived neurotrophic factor (BDNF), one of the earliest mutations to be associated with neuropsychiatric disorders, for which the underlying molecular mechanism is unknown. Here we report on ~ 0.5 ms fully-atomistic explicit solvent temperature replica exchange molecular dynamics simulations of the 91 residue prodomain, for both the V66 and M66 sequence with neutral and protonated adjacent histidine. We show that the highly disordered prodomain can be meaningfully divided into domains based on sequence alone. We find that the tertiary contact preference is determined by individual domains' mean hydrophobicity, net charge and its distribution. Furthermore,

[†]A footnote for the title

we find the prevalence of Met-Met interactions in M66 sequence, these interactions have been under-appreciated in IDPs. The simulations were able to correctly reproduce the location of both local and non-local secondary changes due to the mutation when compared with NMR spectroscopy.¹ We find that the local structure change is mediated via entropic and sequence specific effects. Interestingly, we find that the non-local secondary structure change is not mediated via change in tertiary contact preference but rather it is mediated via change in local secondary structure preference.

Introduction

The physiological significance of intrinsically disordered proteins (IDPs), which can explore a wide range of conformational ensembles in their functional form,²⁻⁵ is now well-established. More than 33% of eukaryotic proteins contain disordered regions longer than 30 residues,⁴ 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.² 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 DISPROT⁹ 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.

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 nucleotide 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 translational 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} Hydrophobicity dependent compaction has been observed in disordered domain of poly(A)-binding protein.²³

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;^{14,24} 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.^{7,25} and references therein). Molecular dynamics (MD) simulations have played a significant role in understanding IDP structure and dynamics,^{26–32} 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^{33,34} 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.³³

Many mutations 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 mutation (rs6265) in the pro-domain region of Brain-derived Neurotrophic Factor (BDNF),³⁵ a signaling protein that retains a critical role in neurogenesis and synaptogenesis throughout adulthood (Fig 1b).³⁶ It has been implicated in maintenance of the hippocampus and the mechanism underlying action of numerous antidepressants,^{37,38} including rapidly acting low-dose ketamine.³⁹ An extensive library of genome-wide association (and even earlier) studies have repeatedly identified the Val66Met mutation 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.^{35,39–42} BDNF plays role in orientation selectivity in the visual system^{43–45} and recently, Val66Met has also been associated with reduced visuomotor associative learning and the sensitivity to action observation⁴⁶

Difficulties in obtaining unambiguous disease associations at the proBDNF Val66Met mutation using GWAS are paralleled by challenges in characterizing its effects on the properties

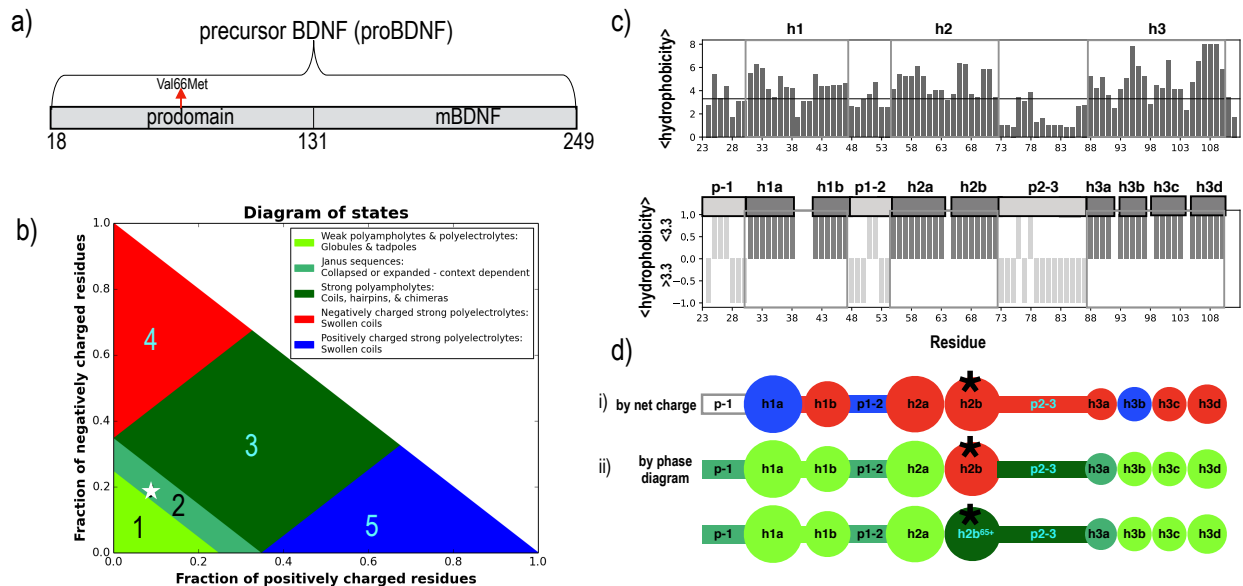


Figure 1: Prodomain and its domain decomposition. a) proBDNF consists of two domains: the prodomain and mature BDNF (mBDNF). b) Diagram of states reported by Ref.^{10,11} based on fraction of positively and negatively-charged residues. As indicated with star, the neutral and protonated BDNF prodomain lies in the Janus region. c) Mean hydrophobicity for the V66 prodomain based on a sliding window of 3 residues (top) and hydrophobicity based domain identification of prodomain sequence (bottom) further described in methods. Eight hydrophobic domains (darkgrey) are identified along with 3 domain of low hydrophobicity (dimgrey). d) The cartoon representation of the prodomain where i) domains colored according to net charge; negative charge (red), positive charge (blue) and neutral (white) ii) domains are colored according to the region found in diagram of states for neutral (top) and protonated prodomain (bottom). The domain h2b contains the mutation and is marked with star. Parts b) is generated with CIDER.⁴⁷

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.⁴⁸

It was subsequently revealed that the cleaved proddomains (~ 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.¹ NMR measurements on the prodomain confirmed significant intrinsic disorder for both forms, with differential secondary structure preference around residue 66.¹ 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 mutation, but additional NMR experiments implicated residue 66 in binding of M66 prodomain to SorCS2.¹

The Val66Met is present in a region with high density of negative charged residues (D61,E64,E68,E69) (Fig 1c). In this scenario, residue H65 can exist in protonated or neutral charge state in vivo due to its low pKa. In order to capture the effects of histidine protonation states on Val66Met, we study the Val66 and M66 prodomain in presence of both neutral H65 and protonated H65. For the ease of understanding our observations, the prodomain sequence is divided into hydrophobic domains (Fig 1c).

In this work, we report on unbiased fully-atomistic replica-exchange MD simulations of the 90 residue BDNF prodomain in explicit solvent, for V66, V66⁶⁵⁺, M66 and M66⁶⁵⁺ forms. This sequence falls at the boundary of the Janus and globular domains in the diagram proposed by Das and Pappu.^{10,11}

Results and discussion

Prodomain Sequence Decomposition

The BDNF prodomain is 91 residues long and residue-level analysis for the entire protein is challenging to interpret. To simplify analysis, the prodomain sequence was divided into subdomains based on sequence hydrophobicity. Eight hydrophobic domains (h) and 3 linkers with lower hydrophobicity (p) and higher disorder were identified (Fig 1c). The properties of each domain are summarized (Table 1).

We characterized each domain on the phase diagram and calculated its κ as proposed by Pappu et al.^{11,49} κ measures charge segregation in a linear sequence; low κ and high κ is realized for well mixed sequences leading to random coil conformations and sequences with segregation of oppositely charged residues leading to hairpin like conformations, respectively. Two unique hydrophobic domains with high NCPR (≥ 0.25) are identified: h2b/h2b⁶⁵⁺ and h3a.

Table 1: Hydrophobic and linker domains identified in Pro region of BDNF^a.

Domain	N	NCPR	<H>	FCR	f-	f+	κ	Sequence	R	P
p-1	8	0.00	0.37	0.25	0.13	0.13	0.8	EANIRGQG	2	0.00
h1a	8	0.13	0.52	0.13	0.00	0.13	1.0	GLAYPGVR	1	0.12
h1b	6	-0.17	0.49	0.17	0.17	0.00	0.1	TLESVN	1	0.00
p1-2	7	0.29	0.33	0.29	0.00	0.29	1.0	PKAGSR	2	0.16
h2a	9	-0.11	0.58	0.11	0.11	0.00	0.7	GLTSLADTF	1	0.00
h2b(V66)	8	-0.38	0.54	0.38	0.38	0.00	0.3	HVIEELLD	4	0.00
h2b(M66)	8	-0.38	0.50	0.38	0.38	0.00	0.3	HMIEELLD	4	0.00
h2b ⁶⁵⁺ (V66)	8	-0.25	0.54	0.50	0.38	0.13	0.6	H ⁺ VIEELLD	3	0.00
h2b ⁶⁵⁺ (M66)	8	-0.25	0.50	0.50	0.38	0.13	0.6	H ⁺ MIEELLD	3	0.00
p3	15	-0.13	0.21	0.53	0.33	0.20	0.1	EDQKVRPI NEENNKDA	3	0.06
h3a	4	-0.25	0.45	0.25	0.25	0.00	N/A	DLYT	2	0.00
h3b	5	0.20	0.60	0.20	0.00	0.20	N/A	RVMLS	1	0.00
h3c	5	-0.20	0.49	0.20	0.20	0.00	N/A	QVPLE	1	0.20
h3d	7	-0.14	0.70	0.14	0.14	0.00	1.0	PLLFLLE	1	0.14
V66 seq	91	-0.09	0.44	0.26	0.2	0.09	0.2		2	0.07
M66 seq	91	-0.09	0.44	0.26	0.2	0.09	0.2		2	0.07
V66 ⁶⁵⁺ seq	91	-0.08	0.44	0.28	0.18	0.10	0.2		2	0.07
M66 ⁶⁵⁺ seq	91	-0.08	0.44	0.28	0.18	0.10	0.2		2	0.07

^aNumber of residues (N), Net charge per residue (NCPR), Mean hydrophobicity <H>, Fraction of charged residues (FCR), fraction of positive (f+) and negative charge residues (f-), κ and region(R) in phase diagram as defined by Das et al,¹¹ fraction of Proline residues (P). κ was calculated using CIDER.⁴⁷

The domain h2b has several unique properties: i) it contains the Val66Met mutation ii) is located at the sequence midpoint iii) its the only strong polyelectrolyte domain and has the highest NCPR among all the domains. It contains a mix of acidic and hydrophobic residues. Strong polyelectrolytes are predicted to behave as swollen coils,¹¹ thus the hydrophobic residues in this solvated expanded coil domain would be more accessible to the solvent than other hydrophobic residues in compact globule forming domains, if present as an isolated peptide. The Val66Met mutation slightly reduces the hydrophobicity of domain h2b (Table 1). Protonation at residue His65 moves the entire h2b domain from a strong polyelectrolyte region with low κ to a strong polyampholyte region with high κ .

The domain h3a is a unique hydrophobic Janus domain with high NCPR. Janus sequences have intermediate compositional biases and their conformations are context dependent .⁵⁰ The SNP domain h2b and the Janus domain h3a are separated with the long (15 residue) strong polyampholyte linker p3 which has well mixed charge ($\kappa = 0.1$). All remaining hydrophobic domains are classified as weak polyampholyte and are predicted to have compact globule conformations to shield hydrophobic residues if present as an isolated peptide .⁵⁰ The domains h1a and h3b are positively charged and all the remaining hydrophobic domains are negatively charged (Fig 1c).

The BDNF prodomain by itself is a boundary region polyampholyte with very low negative NCPR (Table 1). However, by coarse-graining the BDNF sequence, we were able to identify it as an arrangement of eight hydrophobic domains interconnected by two less hydrophobic disorder promoting domains.

Comparison of experimental observables and their computational analogues

Chemical shifts for BDNF prodomain indicates very low propensity of residual secondary structure. In order to get accurate structural characterization of prodomain with MD, we ran T-REMD simulations of 30 residue fragment of V66 prodomain with several commonly used forcefield and water model combinations. Amber99sb*-ildn-q^{51,52} with Tip4p-D⁵³ yielded the best agreement with experimental data obtained by NMR¹ (FigS1) and was used for the full prodomain simulations. This is consistent with the force field comparison study by Robustelli et al which observed that for IDP's with little or no secondary structure, Amber99sb*-ildn-q with Tip4p-D gives good agreement with experimental NMR measurements .⁵⁴

In order to assess the validity of MD-generated ensembles, we compared the MD ensembles with experimental data obtained by NMR spectroscopy and NMR diffusion measurements (Fig 2a, FigS2, FigS3a) . Fig 2a shows the C_α chemical shifts calculated from the

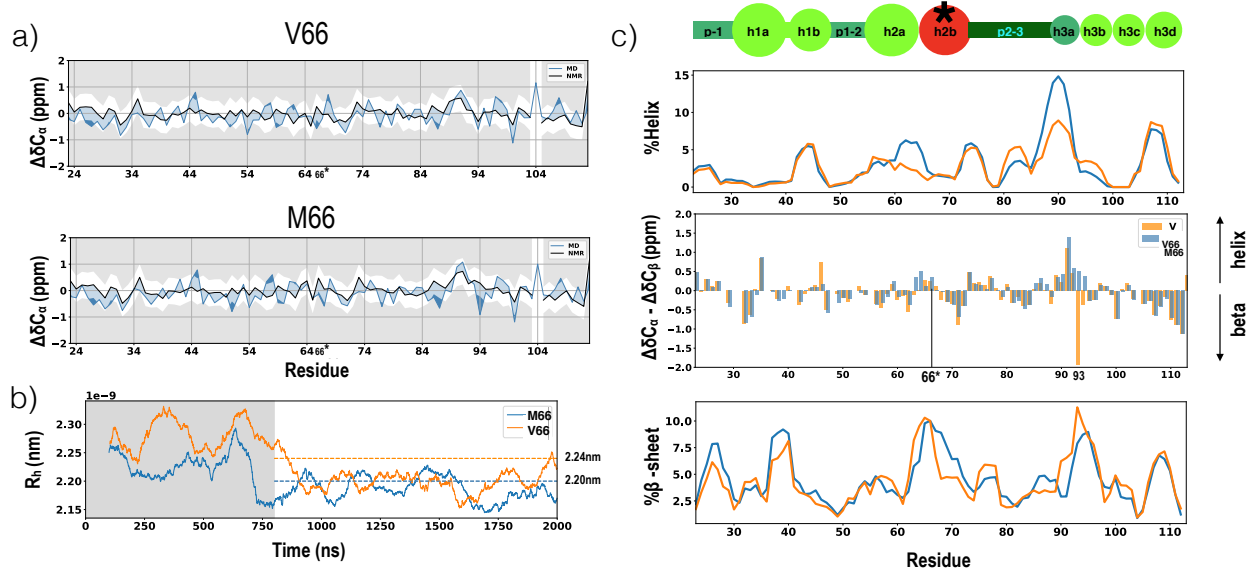


Figure 2: **Comparison of MD and NMR observables.** a) Comparison of chemical shifts from MD ensembles (blue lines, at 300K) and NMR chemical shifts (black lines, at 280K) from Anastasia et al.¹ b) Hydrodynamic radius (R_h) at 100 ns moving window for V66 and M66 sequence vs the simulation time. The R_h for both the sequences converge after 800ns of simulation time. The NMR diffusion experiment values from Anastasia et al.¹ are represented with dashed lines. c) Helix (top) or β (bottom) propensity at each simulated residue defined as the probability of a given residue being part of a sequence of four or more consecutive residues whose dihedral angles place them in the helical region or beta region of the Ramachandran map (further described in methods). C_{α} - C_{β} secondary chemical shifts for each sequence from Anastasia et al.¹ (middle). Positive differences indicate helical structure and negative differences indicate β structure.

simulations using SPARTA+⁵⁵ and compares them with the NMR chemical shifts. We obtain good agreement with NMR chemical shifts: the discrepancy at each residue is <0.7 ppm, which is less than the individual SPARTA+ prediction uncertainties of ~ 1 ppm.⁵⁵

The simulated hydrodynamic radii (R_h) calculated using Hydropro⁵⁶ of V66 (2.23 ± 0.02 nm) and M66 (2.20 ± 0.02 nm) sequence are in excellent agreement with the experimental values from NMR diffusion measurements by Anastasia et al.¹ (2.24 ± 0.1 nm and 2.20 ± 0.1 nm respectively) (Fig 2b). The R_h for protonated His65 (FigS3b) are also in excellent agreement with the experimental values (V66⁶⁵⁺: 2.24 ± 0.02 nm, M66⁶⁵⁺: 2.21 ± 0.02 nm). This approach has been used earlier to validate simulations of disordered proteins^{57,58} and our results support the use for Tip4p-D for such simulations^{53,54}.

Anastasia et al ¹ observed an increase in helical tendency for the M66 sequence within domain h2 and h3ab and an increase in β tendency within domain h3b in the V66 sequence (Fig 2c). To investigate the effect of the mutation on residual secondary structure for the simulated prodomain, we compared the helix tendency at every residue for V66 and M66 simulations (Fig 2c).

Consistent with NMR experiments, the M66 sequence demonstrates an increased tendency of forming helices within domain h2 and h3ab relative to the same domains in the V66 sequence (Fig 2d). We also find that the residue R93 in domain h3b of V66 sequence has a higher tendency to form β -sheets than in the M66 sequence, in agreement with NMR experiments (Fig 2c). We get consistent observations with NMR data ¹ when V66⁶⁵⁺ and M66⁶⁵⁺ simulations are compared (FigS3c).

Interdomain interactions (prototertiary structure)

To identify favorable long-range interactions we calculated the normalized probability of inter-domain C_α interactions (Fig 3a, FigS5a) as described in the Methods. We find that hydrophobic domains have a higher probability of forming long-range contacts, even for domains separated by sequence. The inter-domain contacts for the least hydrophobic, strong polyampholyte p3 linker decay with the distance along the sequence as we would expect for a random coil.

Two out of 8 hydrophobic domains in the prodomain are positively charged, and all the remaining domains are negatively charged. Consistently, most of the observed long-range contacts either include positively charged domains h1a or h3b and/or negatively-charged, high NCP domains, h2b and/or h3a (Fig 3a, FigS5a). Furthermore, the oppositely charged terminal domains have frequent contacts. Overall, we find that even for hydrophobic domains, electrostatic interactions do contribute to inter-domain contacts.

We find that frequent contacts are always formed between neighboring domains, except at h2b/h2b⁶⁵⁺-p3 and h3c-h3d. Three sequential same charged residues (Asp72,Glu73,Asp74)

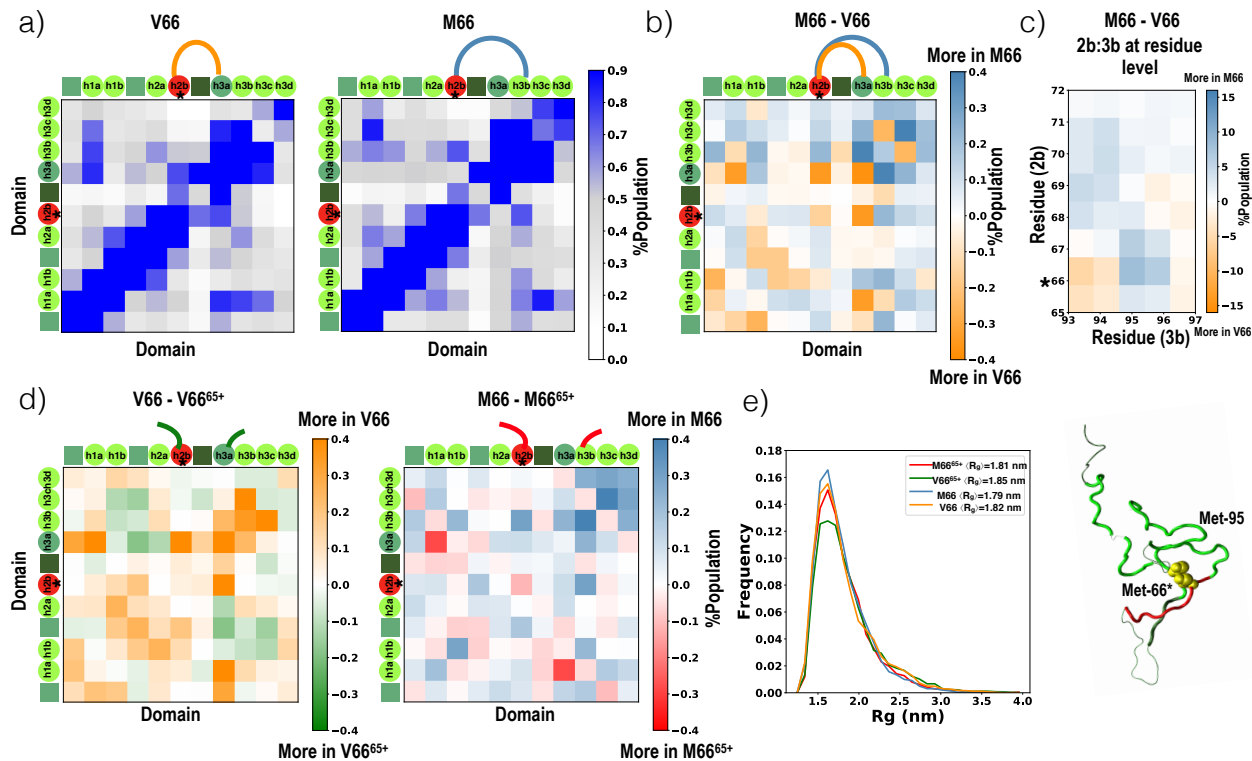


Figure 3: Effect of Val66Met on inter-domain contacts. a) Inter-domain contact probability for V66 (left) and M66 (right). The maps are annotated with cartoon representation of the prodomain; circles and squares represent hydrophobic domains and linker regions respectively and are colored as in Fig1. Two domains are in contact if the distance between any inter-domain pair of C_{α} atoms is within 8\AA . The contact probability for any domain pair is also normalized with the product of length of each domain. The most frequent long-range contact formed by the SNP domain is represented with an orange (V66) or blue (M66) edge connecting the two domains. b) Difference between the contact probabilities shown in a); interactions more frequently found in M66 or V66 are in blue and orange respectively. c) Contact probability at each residue shown in b) within domains h2b and h3b (top). Representative conformation of M66 sequence showing Met66-Met95 contact, colored as in Fig1 with Methionine in yellow. Thick lines represents hydrophobic domains whereas thin line represents linker regions (bottom). d) Difference between the contact probabilities of neutral H65 and protonated H65 for V66 (left) and M66 (right) sequence respectively. e) R_g distributions across all four simulations.

at the interface between h2b/h2b⁶⁵⁺ and p3 and two sequential Proline residues (Pro104,Pro105) at the interface between h3b and h3d probably explains the less frequent contacts between these two neighboring domains (Table 1). We would expect the strong hydrophobic polyelectrolyte h2b domain to form frequent contacts with all other domains, due its exposed hydrophobic residues and low κ (Table 1). Indeed, the SNP domain h2b forms strong long

range contacts with positively charged domain h3b and/or the negatively charged h3a.

In order to understand the effect of Val66Met on inter-domain contacts, we compared the contact maps for each sequence (Fig 3b). Val66Met shifts the inter-domain h2b-h3a contact to h2b-h3b. To understand the origin of the more favorable interaction between the SNP domain and h3b, we zoom in on the residue level interactions between these two domains (Fig 3c). Both V66 and M66 sequence form their strongest interactions between hydrophobic residues, Val66-Val94 (9%) and Met66-Met95 (10%) respectively (FigS4). Interestingly, we find that Met95 forms frequent contact with Met66 and not Val66 (Met66-Met95 (10%) vs Val66-Val95 (2%)). This is consistent with the results of ab-initio calculations: Gomez-Tamayo et al has reported that Met-Met interactions are stronger than Met-Leu, Met-Phe or aromatic-aromatic interactions due to polarizability of sulphur .⁵⁹

Apart from direct effects on contacts involving SNP, we also observe indirect effects on contacts not involving the SNP. h3a-h3b and h3a-h1a are observed frequently in M66 and V66 sequence respectively (Fig 3b). Even when H65 is protonated, Val66Met changes the maximum contact preference for the Janus domain h3a; gain of contacts at h3a-h3b, h3a-h1a and loss of h3a-h2a in M66⁶⁵⁺ when compared with V66⁶⁵⁺ (FigS5b).

In order to understand the effect of protonation on the prodomain ensemble, we compared the inter-domain contact maps for each protonation state (Fig 3d). h2b⁶⁵⁺ forms fewer contacts with the remaining domains for both V66 and M66 when compared with h2b. Infact, we observe more self-interactions of h2b⁶⁵⁺, as expected for a strong polyampholyte domain with high κ when compared with strong polyelectrolyte domain with low κ (Table 1). Apart from direct effects on contacts involving h2b⁶⁵⁺, protonation causes loss and gain of h3a-h1a contacts in V66 and M66 sequences respectively. This sensitivity to both mutation and protonation of other domains is consistent with context dependence of the Janus sequence h3a.

There are slight differences in the radius of gyration (R_g) distributions across all four simulations (Fig 3e): 1) The sequence with Met66 has slightly lower R_g when compared with

its Val66 form. For M66 sequence, the SNP domain h2b/h2b⁶⁵⁺ forms frequent contacts with the rest of the domains when compared with the Val66 sequence (Fig 3b). This could explain the slightly more compact states of M66 and M66⁶⁵⁺ when compared with V66 and V66⁶⁵⁺. 2) The His65 protonated form is more expanded when compared with its respective non-protonated states. For both V66 and M66 sequence, protonation at 65 reduces inter-domain contacts at h2b (Fig 3d), which is consistent with the slight increase in R_g upon protonation.

Intradomain interactions and residual secondary structure

We find frequent intra-domain contact probability at every domain except at p3 linker and domain h3c (Fig 3a, FigS5a). Consistent with p3 linker being a strong polyampholyte with low κ , we find less self interaction in this domain (Table 1). Domain h3c show less self-interaction due to its highest fraction of Proline (20%) and its placement at the center of the domain (Table 1).

The NMR chemical shifts indicate that mutation increases the helicity at both the SNP domain h2b and Janus domain h3a. In agreement with NMR, we find an increase in helical structure at both domains in the M66 sequence (Fig 1c). Comparing the length of secondary structure formed at each residue (Fig 4a, FigS8b) reveals an even stronger effect of the mutation: in both protonated and neutral His65 states, Val66Met consistently increases the frequency of long helix formed within domain h2.

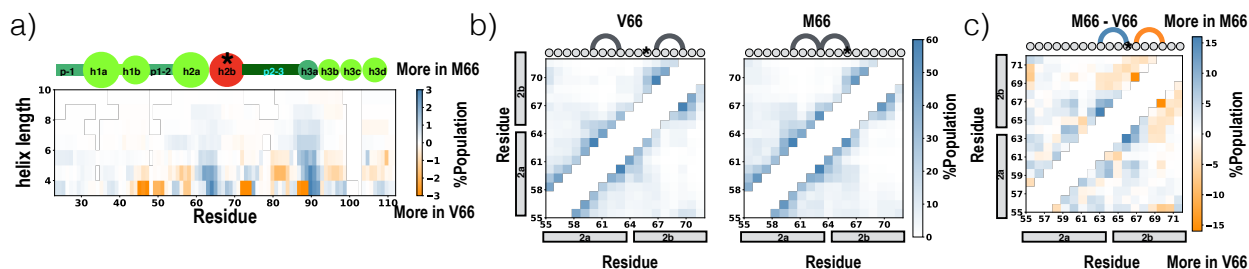


Figure 4: a) Difference in helix length probability at each residue in the prodomain. b) Intra-domain contact probability for V66 (left) and M66 (right). Each residue in domain 2 is annotated with a circle representation and the most preferred contact is represented with an edge. c) Difference between the contact probabilities shown in b); interactions more frequently found in M66 or V66 are in blue and orange respectively.

Reduced entropic cost of helix formation in M66 when compared with V66 possibly contributes to the increased helix formation in the h2 domain. Creamer et. al. ranked the entropic cost of helix formation for apolar side chains using simulations of a (Ala)₈ sequence with the guest amino acid at the center, and reported higher entropic cost of helix formation for Valine when compared with Methionine.⁶⁰ In general, C^β-branched amino acids, such as Valine, have more restricted side-chain rotamers in helical conformation when compared with non-C^β-branched amino acids.

The helical structure in M66 is also stabilized by local sequence. We considered the intra-domain contact map of domain h2 and find that for both M66 and M66⁶⁵⁺, Met66(i) is frequently contacting Phe63(i-3) when compared with any other intra-domain contact at residue 66: M66-F63 is 50% stronger when compared with M66-E69 (Fig 4b, FigS8b). In the V66 sequence, it is far more likely that the Ile67(i+1)-Leu70(i+4) contact is formed (Fig 4c). A previous study from Faure et al.⁶¹ analyzed the frequency of two amino-acids contact within 1230 protein chains from the Protein DataBank (PDB) and found that Met has also a strong affinity for Phe and other Methionines. In fact, we find that the largest change in intra-domain contacts from V66 to M66 is the gain of contact at M66-F63 (20% stronger in M66 when compared with V66) followed by loss of contact at Ile67(i+1)-Leu70(i+4) (Fig 4b).

When the M66 sequence is protonated at His65, the helix propensity shifts from helix at residues 59-66 to even longer helix at residues 59 - 69 (Fig S8a). When protonated, both V66 and M66 sequences gain contacts with His65, Met/Val66 and Ile67 (FigS8b). These increases in intra-domain contact frequency at h2b⁶⁵⁺ are consistent with expectations, since protonation at H65 reduces the net charge and increases κ of the otherwise negatively charged polyelectrolyte domain with low κ (Table 1).

Secondary structure (intra-domain) and tertiary structure (inter-domain) coupling

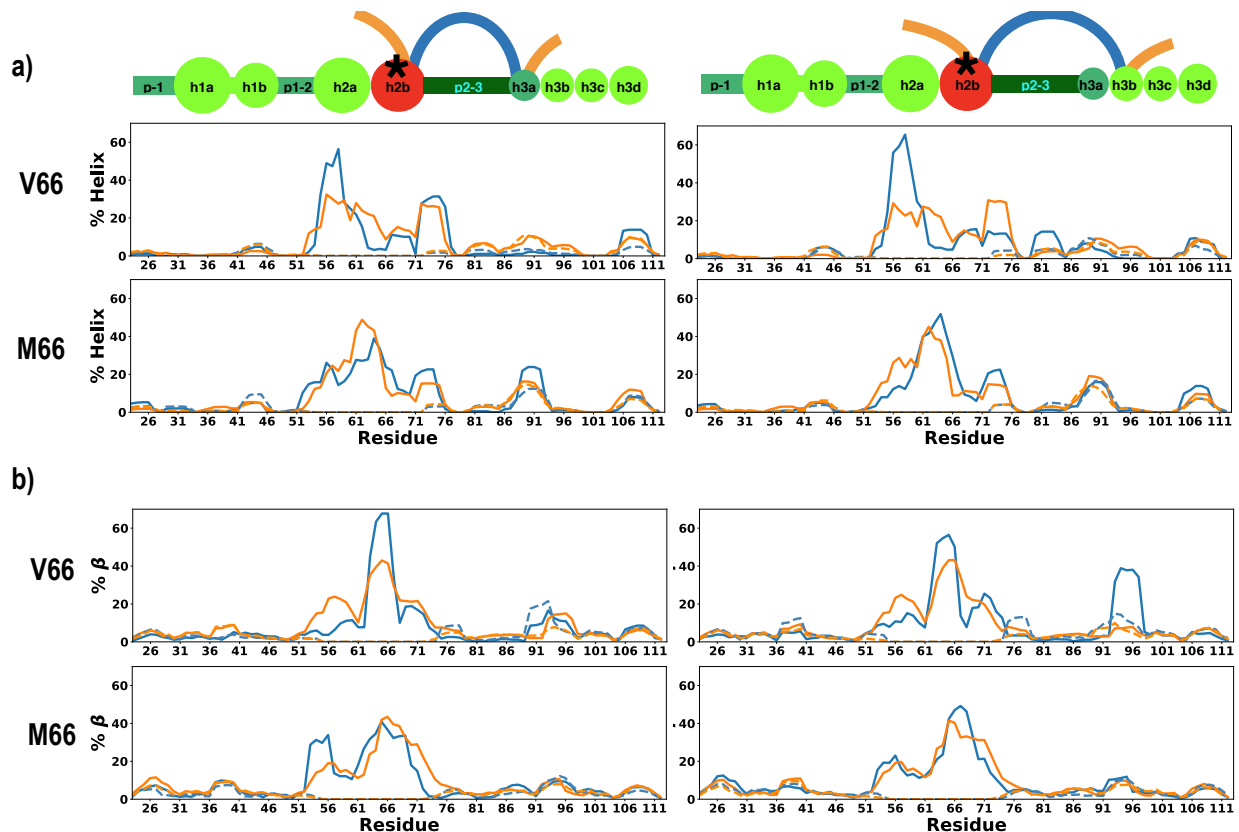


Figure 5: **h2b-h3ab contact in prodomain and domain h2 secondary structure coupling.** Helix (a) and β (b) propensities at each residue in V66(top) and M66(bottom). Frames were first clustered by whether the inter-domain h2b-h3a (left) or h2b-h3b (right) contact was formed (blue) or broken (orange), and then by whether the respective secondary structure was present in h2b (solid) or absent (dashed).

Both our own MD results and previous results from NMR indicate the mutation causes local secondary structure change in domain h2 and non-local secondary structure changes in domain h3a and h3b (Fig 4c). We hypothesized that since the Val66Met mutation affects the contact frequency of domain h2b with h3ab, the secondary structure change in h3ab could be due entirely to the change in inter-domain contacts. To test this, we divided the frames into 4 clusters, representing two independent collective variables with two possible values each: either the inter-domain contact is formed or broken, and either domain h2 does

or does not have significant secondary structure.

First, we clustered frames by whether the SNP domain h2b is in contact with h3ab and then by whether helical structure was present in h2. Fig 5 compares the helix propensities at each residue for each cluster in V66 and M66 sequence. In V66 sequence, formation of h2b-h3a and/or h2b-h3b disrupts helicity at residue 66 and within Janus domain. We observe likewise for V66 protonated sequence (FigS11). In M66 sequence, we find that both the contacts h2b-h3a and h2b-h3b stabilizes helicity at residue 66. Additionally, when h2b-h3a is formed, the domain h3a helix has twice as much probability to be formed simultaneously with h2 helix when compared with the remaining 3 clusters Fig 5. This suggests that the h2b-h3a inter-domain contact is necessary but not sufficient for increased helix formation in h3a. We observe likewise for M66 protonated sequence (FigS11b).

Next, we looked at the 4 clusters made with β structure at h2 as dependent variable for each contact (Fig 5b). In V66 sequence, we find that both the contacts h2b-h3a and h2b-h3b stabilizes beta at either domains. Additionally, when h2b-h3b is formed, the domain h3b β has much stronger correlation with domain h2 β when compared with the remaining 3 clusters Fig 5. This suggests that the h2b-h3b inter-domain contact is necessary but not sufficient for increased β formation in h3b. In M66 sequence, we do not observe any such correlation.

To summarize, we find that the direct contact between the SNP domain and h3ab predicts β and helicity locally in V66 and M66 sequence respectively. This was further confirmed by looking at the intra-domain contact probability within h2 when either contact is formed for V66 vs M66 sequence (Fig 6a). M66 sequence forms helix supporting contacts at residue 66(i) whereas V66 doesn't form any strong intra-domain contact at 66 when h2b-h3ab is formed. For non-local secondary structure change at h3ab, the h2b-h3ab inter-domain contact is necessary and but is not sufficient. Intra-domain contact at h2 influences intra-domain contact at h3ab when h2b-h3ab inter-domain contact is formed.

Few studies have reported helix-helix stabilization in IDP's. Feuerstein et al ⁶² reported

partly stabilized helical structure by long-range tertiary in disordered fragment of the hepatitis C virus protein NS5A with PRE measurement from NMR. Conicella et al.²² showed that TDP-43 undergoes liquid-liquid phase separation in part via helical contacts that are enhanced by self-interaction and can be disrupted by disease mutation. However, identifying the type of interaction at residue level has been difficult due to disorder.

To determine the type of residue contacts between h2 and h3ab stabilizing β and helicity in V66 and M66 sequence, we plotted the inter-domain contact probability when respective secondary structure is formed at h2(Fig 6a). In V66 the beta pairing at h2b-h3b is supported via combination of backbone hydrogen bonds at V66:S92, electrostatic and hydrophobic contacts at residues E64:R93, V66:V94 respectively (Fig 6b). We do not find these beta sheet structures with h2b:h3a contact in V66 protonated states (Fig S10,S12). This is likely due to electrostatic repulsions between residue H65+ and R93. In M66 the helix coupling of h2b-h3ab is supported via hydrophobic contacts at residues M66 and L70 (Fig 6c). In protonated states, helix within h3ab is supported via hydrophobic contact with residue M66 as well (FigS11c).

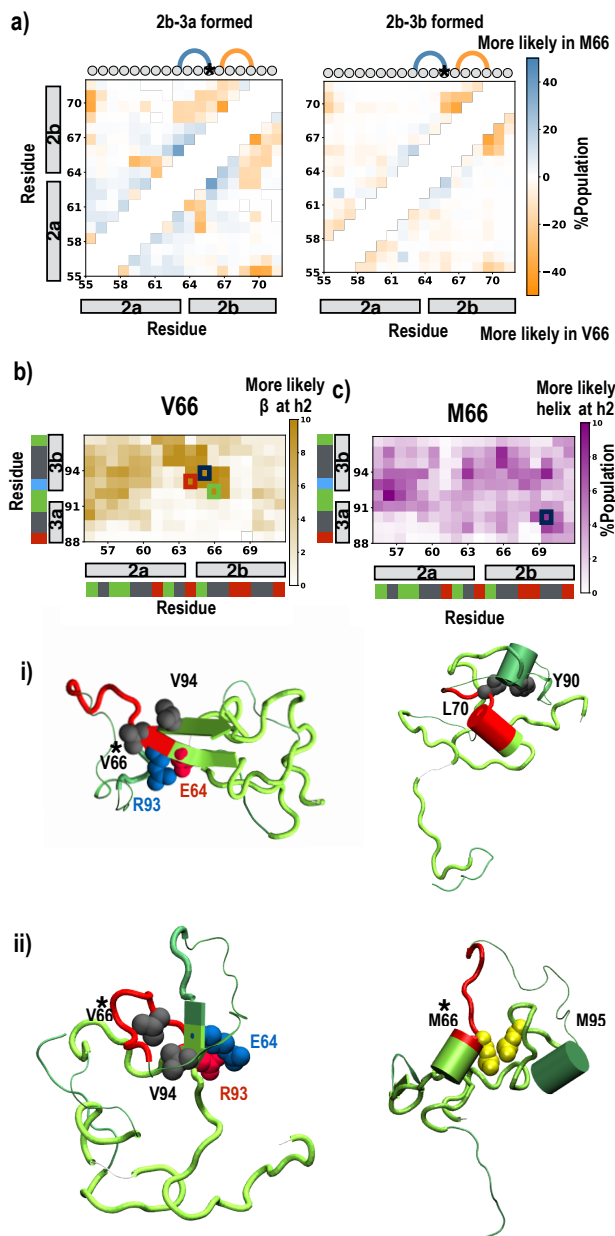


Figure 6: **Intra-domain and inter-domain contacts at residue level.** Difference between contact probability for M66 and V66 sequence at each residue in domain h2 when inter-domain contact is present between h2b-h3a (left) and h2b-h3b (right). The strongest contact difference is represented with an orange (V66) or blue (M66) edge connecting the two residues. b) Contact probability at each residue between domains h2 and h3ab when helix (b) or β (c) within domain h2 is formed in V66 and M66 sequence respectively. Preferred hydrophobic interactions are annotated with black squares, backbone contacts with green squares and opposite charge contacts with red squares. The label of the plots are colored according to residue type: blue-basic, red-acidic, green-polar and Gly, grey-hydrophobic. i) and ii) are representative conformations from 2 different replicas for V66 (left) and M66(right) with chain colored as in Fig1 and preferred residue-level contacts in VDW representations, with residues colored with residue type.

Summary and Conclusion

We have carried out ~ 0.5 ms of fully-atomistic MD simulation of the 91 residue prodomain of brain-derived neurotrophic factor with protonated and neutral His65 state, both with and without the disease-associated Val66Met mutation. These long simulations successfully reproduced the experimentally observed chemical shift and R_h . In respect to effect of the Val66Met mutation on residual secondary structure observed experimentally, the simulations were able to correctly reproduce the location of both local and non-local secondary changes due to the Val66Met mutation in the prodomain sequence.

Due to highly dynamic nature of disordered proteins, identifying their tertiary contacts preference is very challenging. In this study, we showed that coarse-graining a highly disordered long prodomain into domains of 4 or more residues based on the hydrophobic residues distribution along the sequence can give us meaningful insights into its "tertiary" contact network.

We show that the intra-domain and inter-domain contact frequency at each domain is determined by its mean hydrophobicity in addition to NCPR, charge residue distribution as proposed by Das et al. and its proline content.

Each identified hydrophobic domains lie in the folded protein region in the mean charge vs mean hydrophobicity plot except the linker domain p3. We identified two unique hydrophobic domains, the Val66Met mutation containing strong polyelectrolyte domain h2a and the Janus domain h3a connected via a unique highly disordered long linker region p3. These unique domains have biological significance as well: The sequence h2-p3-h3abc is essential for intracellular trafficking of proBDNF.⁶³

The functional importance of flexible linker regions in ordered has been intensively analyzed. Flexible linkers allow the connecting domains to freely twist and rotate to recruit their binding partners in ordered proteins. The presence of linker region at the separation of SNP domain and h3ab is probably essential in this protein design.

In alignment with the sequence characteristics of Janus sequences, we find that Janus

domains has the unique ability to forms both helix or β structures with high propensity. We observe that Janus domain secondary structure and inter-domain contacts is sensitive to SNP domain secondary structure and mutation/protonation within SNP domain respectively.

At residue level tertiary contact preference, we observe that Met-Met interactions drives tertiary contact changes in the prodomain due to Val66Met. Met-Met interactions have been shown to stabilize tertiary contacts in folded proteins and membrane proteins, but their role has not been investigated in disordered proteins. When the adjacent histidine is protonated, the frequency of tertiary contacts are reduced, including Met-Met interactions. This is likely due to less exposed hydrophobic residues in now strong polyampholyte SNP domain when compared with polyelectrolyte SNP. This suggests that Met-Met interactions could play a role in IDP tertiary contacts if present in a polyelectrolyte/expanded coil conformation region.

Interestingly, we find that for non-local secondary structure change, the inter-domain contact between local and non-local domain is necessary and but is not sufficient. Intra-domain contact at local domain influences intra-domain contact at non-local domain when inter-domain contact is formed between the two domains. Staller et al ⁶⁴ have earlier reported that in disordered acidic activation domain (Gcn4), the acidic residues keep key hydrophobic residues exposed to solvent and binding partners. In M66, mutation-Janus stabilizes helix structures in either domain mediated via hydrophobic interactions, whereas in V66, mutation-Janus stabilizes β structures within Janus domain mediated via combination of backbone hydrogen bonding, hydrophobic and electrostatics.

We were able to get insights into the mechanism via which hydrophobic mutations can have effects on proteins secondary structure, non-local secondary structure and tertiary contacts. We compare the effects of a hydrophobic mutation and adjacent histidine protonation. i) Mutation changes tertiary contact preference via Met-Met interactions. Whereas, protonation weakens inter-domain contact ii) Mutation changes intra-domain contact preference due to local sequence effects and residue entropic cost of helix formation. Protonation increases

intra-domain contact frequency due to reduced electrostatic repulsions in the otherwise negatively charged SNP domain.

Anastasia et al¹ observed differential kinetics for interactions between the BDNF prodomain and SorCS2; M66 binds more preferably at SNP domain (H65 to L71) with SorCS2, whereas V66 binds weakly overall but more strongly with domain h3a and h3b (residues Y90 to V94). The stronger binding at residue M66 could be attributed to either a) ability of M66 to form alpha-helix at residue 66 when compared with V66 or b) stronger Met-Met contact at residue M66 with the SorCS2 Met residues, or a combination of both a) and b).

In our simulations, we observe that helicity within the SNP domain is coupled with contact formation between the SNP domain and Janus domain hydrophobic residues. The SorCS2 sequence also has a Janus domain patch on its surface. Thus, M66 forms could form simultaneous helix and contact formation within the SNP domain and Janus domain on SorCS2 surface. In V66, instead of the SNP, the Janus domain may interact strongly with SorCS2 surface.

Protonation increases self interaction but reduces inter-domain contacts, thus, very slightly low pH can mediate the interaction between the SNP domain and SorCS2 surface.

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-Tasser⁶⁵⁻⁶⁷, Robetta and Modeller⁶⁸, and all were simulated in a water box at 600K for 50 ns at a constant volume. From the six resulting trajectories, 64 structures with correct proline isomers were selected (based on at least 2ps time interval); in total, our study included 64 unique prodomain structures. All structures were cooled to 300K for 1ns, while prolines were restrained in trans-conformation. Each V66-Hip65 replica was placed in a dodecahedron

water box with 30,500 TIP4P-D⁶⁹ water molecules and a 0.15M salt concentration (NaCl) for a total system size of approximately 124,000 atoms. The same volume for each replica was ensured by fixing the simulation box of each replica to the average box size (11 nm).

Molecular Dynamics Simulation

For the simulations we use the amber99SB*-ILDN-q force field⁷ and the GROMACS 5.1.2 simulation package^{70,71} with a time step of 2 fs. 64 replicas are used with temperatures ranging from 300-385K, with exponential spacing. System was simulated using T-REMD⁷² with an exchange frequency of 1ps for 2 μ s, giving a total simulation time of 128 μ s with NVT ensemble for each system. A different random seed was used for the Langevin dynamics of each replica. Long-range electrostatics are calculated using the particle mesh Ewald (PME) method⁷³ with a 1 nm cutoff and a 0.12 nm grid spacing. Periodic boundary conditions are also used to reduce system size effects. The average exchange acceptance probability ranged between 0.19-0.23 which is quite good. For both V66 and M66 groups, about 800ns for each replica were discarded for equilibration purposes.

M - .36, V- .77, V-hip-11, 4.47 , replica 38 becomes expanded, all the frames from this replica is removed giving 1.66 remaining replicas, M-hip-11, 1.16.

Domain identification

To be written

Chemical Shifts calculation

Prior to the present study, Anastasia et al¹ 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+⁷⁴ and SSP.⁷⁵ For comparison with simulation data, we reinterpreted the chemical shifts directly from,¹ deposited at Biological Magnetic Resonance Bank. Chemical shifts (CA,CB,CO,N,HN) were generated from MD trajectories using

SPARTA+⁵⁵ and the NMR data (This calculation used did not use HA chemical shifts.)

Secondary structure calculation

The length of helix formed at each residue was calculated by determining the number of consecutive residues in which the dihedral angles satisfied $\phi < 0^\circ$ and $-120^\circ < \psi < 50^\circ$, as in.^{76,77}

Contact maps

If distance between C_α - C_α atoms of two residues is 8\AA or less, noted C_α . Depending on the distance criterion used, two domains are in contact, if the distance between at-least one pair of residue from the domain satisfy the cutoff. To closely look at the prodomain conformations, including backbone-backbone contacts, backbone-sidechain contacts and sidechain-sidechain contacts, we looked at the C_α contact preference between domains with larger cutoff (8\AA).

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. This research used resources from the Rutgers Discovery Informatics Institute, which are supported by Rutgers and the State of New Jersey.⁷⁸

References

- (1) Anastasia, A.; Deinhardt, K.; Chao, M. V.; Will, N. E.; Irmady, K.; Lee, F. S.; Hempstead, B. L.; Bracken, C. *Nat. Commun.* **2013**, *4*, 2490.
- (2) Uversky, V. N. *Biochim. Biophys. Acta* **2013**, *1834*, 932–51.
- (3) Panchenko, A. R.; Babu, M. M. *Curr. Opin. Struct. Biol.* **2015**, *32*, viii–x.
- (4) Ward, J.; Sodhi, J.; McGuffin, L.; Buxton, B.; Jones, D. *J. Mol. Biol.* **2004**, *337*, 635–645.
- (5) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
- (6) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. *FEBS J.* **2005**, *272*, 5129–5148.
- (7) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. *Chem. Rev.* **2014**, *114*, 6561–6588.
- (8) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–40.
- (9) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2007**, *35*, D786–D793.
- (10) Das, R. K.; Ruff, K. M.; Pappu, R. V. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112.
- (11) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392–7.
- (12) Vacic, V.; Markwick, P. R. L.; Oldfield, C. J.; Zhao, X.; Haynes, C.; Uversky, V. N.; Iakoucheva, L. M. *PLoS Comput. Biol.* **2012**, *8*, e1002709.
- (13) Larini, L.; Gessel, M. M.; LaPointe, N. E.; Do, T. D.; Bowers, M. T.; Feinstein, S. C.; Shea, J.-E. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8916.

- (14) Ganguly, D. et al. *PLOS Comput. Biol.* **2015**, *11*, e1004247.
- (15) Viet, M. H.; Nguyen, P. H.; Derreumaux, P.; Li, M. S. *ACS Chem. Neurosci.* **2014**, *5*, 646–657.
- (16) Viet, M. H.; Nguyen, P. H.; Ngo, S. T.; Li, M. S.; Derreumaux, P. *ACS Chem. Neurosci.* **2013**, *4*, 1446–1457.
- (17) Truong, P. M.; Viet, M. H.; Nguyen, P. H.; Hu, C.-K.; Li, M. S. *J. Phys. Chem. B* **2014**, *118*, 8972–8981.
- (18) Zhan, Y. A.; Wu, H.; Powell, A. T.; Daughdrill, G. W.; Ytreberg, F. M. *Proteins Struct. Funct. Bioinforma.* **2013**, *81*, 1738–1747.
- (19) Xu, L.; Shan, S.; Wang, X. *J. Phys. Chem. B* **2013**, *117*, 6206–6216.
- (20) Bah, A.; Forman-Kay, J. D. *J. Biol. Chem.* **2016**, *291*, 6696–6705.
- (21) He, Y.; Chen, Y.; Mooney, S. M.; Rajagopalan, K.; Bhargava, A.; Sacho, E.; Weninger, K.; Bryan, P. N.; Kulkarni, P.; Orban, J. *J. Biol. Chem.* **2015**, *290*, 25090–25102.
- (22) Conicella, A. E.; Zerze, G. H.; Mittal, J.; Fawzi, N. L.; Alexander Conicella, A. E.; Zerze, G. H.; Mittal, J.; Fawzi Correspondence, N. L.; Conicella, A. E.; Fawzi, N. L.; Zerze, G. H.; Mittal, J.; Fawzi, N. L. *Struct. Des.* **2016**, *24*, 1537–1549.
- (23) Riback, J. A.; Katanski, C. D.; Kear-Scott, J. L.; Pilipenko, E. V.; Rojek, A. E.; Sosnick, T. R.; Drummond, D. A. *Cell* **2017**, *168*, 1028–1040.e19.
- (24) Ieřmantavičius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. *J. Am. Chem. Soc.* **2013**, *135*, 10155–10163.
- (25) Mittag, T.; Forman-Kay, J. D. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.

- (26) Stanley, N.; Esteban-Martín, S.; De Fabritiis, G. *Prog. Biophys. Mol. Biol.* **2015**, *119*, 47–52.
- (27) Ithuralde, R. E.; Roitberg, A. E.; Turjanski, A. G. *J. Am. Chem. Soc.* **2016**, *138*, 8742–8751.
- (28) Knott, M.; Best, R. B.; Hummer, G.; de Bakker, P.; Word, J. *PLoS Comput. Biol.* **2012**, *8*, e1002605.
- (29) Invernizzi, G.; Lambrughi, M.; Regonesi, M. E.; Tortora, P.; Papaleo, E. *Biochim. Biophys. Acta* **2013**, *1830*, 5236–47.
- (30) Abeln, S.; Frenkel, D. *PLoS Comput. Biol.* **2008**, *4*, e1000241.
- (31) Yedvabny, E.; Nerenberg, P. S.; So, C.; Head-Gordon, T. *J. Phys. Chem. B* **2015**, *119*, 896–905.
- (32) Levine, Z. A.; Shea, J.-E. *Curr. Opin. Struct. Biol.* **2017**, *43*, 95–103.
- (33) Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J.-E. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2758–63.
- (34) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. *Arch. Biochem. Biophys.* **2008**, *469*, 132–41.
- (35) Notaras, M.; Hill, R.; van den Buuse, M. *Mol. Psychiatry* **2015**, *20*, 916–930.
- (36) Korte, M.; Carroll, P.; Wolf, E.; Brem, G.; Thoenen, H.; Bonhoeffer, T. *Proc. Natl. Acad. Sci.* **1995**, *92*, 8856–8860.
- (37) Autry, A. E.; Monteggia, L. M. *Pharmacol. Rev.* **2012**, *64*, 238–258.
- (38) Björkholm, C.; Monteggia, L. M. *Neuropharmacology* **2016**, *102*, 72–79.

- (39) Autry, A. E.; Adachi, M.; Nosyreva, E.; Na, E. S.; Los, M. F.; Cheng, P.-f.; Kavalali, E. T.; Monteggia, L. M. *Nature* **2011**, *475*, 91–95.
- (40) Soliman, F. et al. *Science* **2010**, *327*, 863–6.
- (41) Chen, Z.-Y.; Bath, K.; McEwen, B.; Hempstead, B.; Lee, F. *Novartis Found. Symp.* **2008**, *289*, 180–8; discussion 188–95.
- (42) Verhagen, M.; van der Meij, A.; van Deurzen, P. A. M.; Janzing, J. G. E.; Arias-Vásquez, A.; Buitelaar, J. K.; Franke, B. *Mol. Psychiatry* **2010**, *15*, 260–71.
- (43) Huang, Z.; Kirkwood, A.; Pizzorusso, T.; Porciatti, V.; Morales, B.; Bear, M. F.; Maffei, L.; Tonegawa, S. *Cell* **1999**, *98*, 739–755.
- (44) Liu, B.-h.; Li, Y.-t.; Ma, W.-p.; Pan, C.-j.; Zhang, L.; Tao, H. *Neuron* **2011**, *71*, 542–554.
- (45) Gao, M.; Maynard, K. R.; Chokshi, V.; Song, L.; Jacobs, C.; Wang, H.; Tran, T.; Martinowich, K.; Lee, H.-K. *J. Neurosci.* **2014**, *34*, 10770–9.
- (46) Taschereau-Dumouchel, V.; Héту, S.; Michon, P.-E.; Vachon-Pressseau, E.; Massicotte, E.; De Beaumont, L.; Fecteau, S.; Poirier, J.; Mercier, C.; Chagnon, Y. C.; Jackson, P. L. *Sci. Rep.* **2016**, *6*, 34907.
- (47) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O.; Pappu, R. V. *Biophys. J.* **2017**, *112*, 16–21.
- (48) Feng, D.; Kim, T.; Özkan, E.; Light, M.; Torkin, R.; Teng, K. K.; Hempstead, B. L.; Garcia, K. C. *J. Mol. Biol.* **2010**, *396*, 967–984.
- (49) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8183–8.
- (50) Das, R. K.; Ruff, K. M.; Pappu, R. V. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112.

- (51) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950–8.
- (52) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.
- (53) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (54) Robustelli, P.; Piana, S.; Shaw, D. E. *Proc. Natl. Acad. Sci.* **2018**, *115*, E4758–E4766.
- (55) Shen, Y.; Bax, A. *J. Biomol. NMR* **2010**, *48*, 13–22.
- (56) Ortega, A.; Amorós, D.; García de la Torre, J. *Biophys. J.* **2011**, *101*, 892–8.
- (57) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmü, H.; Grubmüller, H. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.
- (58) Meng, F.; Bellaiche, M. M. J.; Kim, J.-Y.; Ul, G.; Zerze, H.; Best, R. B.; Chung, H. S. **2018**,
- (59) Gómez-Tamayo, J. C.; Cordoní, A.; Olivella, M.; Mayol, E.; Fourmy, D.; Pardo, L. *Protein Sci.* **2016**, *25*, 1517–24.
- (60) Creamer, T. P.; Rose, G. D. *Nat. Sci.* **1992**, *89*, 5937–5941.
- (61) Faure, G.; Bornot, A.; de Brevern, A. G. *Biochimie* **2008**, *90*, 626–639.
- (62) Feuerstein, S.; Solyom, Z.; Aladag, A.; Favier, A.; Schwarten, M.; Hoffmann, S.; Willbold, D.; Brutscher, B. *J. Mol. Biol.* **2012**, *420*, 310–323.
- (63) Chen, Z.-Y.; Ieraci, A.; Teng, H.; Dall, H.; Meng, C.-X.; Herrera, D. G.; Nykjaer, A.; Hempstead, B. L.; Lee, F. S. *J. Neurosci.* **2005**, *25*, 6156–66.

- (64) Staller, M. V.; Holehouse, A. S.; Swain-Lenz, D.; Das, R. K.; Pappu, R. V.; Cohen, B. A. *Cell Syst.* **2018**, *6*, 444–455.e6.
- (65) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. *Nat. Methods* **2014**, *12*, 7–8.
- (66) Roy, A.; Kucukural, A.; Zhang, Y. *Nat. Protoc.* **2010**, *5*, 725–738.
- (67) Zhang, Y. *BMC Bioinformatics* **2008**, *9*, 40.
- (68) Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (70) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (71) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. *SoftwareX* **2015**, *1-2*, 19–25.
- (72) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (73) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (74) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213–23.
- (75) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795–2804.
- (76) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.
- (77) Iglesias, J.; Sanchez-Martínez, M.; Crehuet, R. *Intrinsically Disord. Proteins* **2013**, *1*, e25323.
- (78) Parashar, M.; Brennan-Tonetta, M.; Rodero, I.; Villalobos, J. J. **2018**,