Sequence specificity despite intrinsic disorder: how a disease-associated Val/Met polymorphism shifts tertiary interactions in a long disordered protein

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

The role of electrostatic interactions and mutations that change charge states in intrinsically disordered proteins (IDPs) is well-established, but 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 over 250 μs of fully-atomistic, explicit solvent, temperature replica exchange molecular dynamics simulations of the 91 residue BDNF prodomain, for both the V66 and M66 sequence. The simulations were able to correctly reproduce the location of both local and non-local secondary changes due to the Val66Met mutation when compared with NMR spectroscopy. We find that the local structure change is mediated via entropic and sequence specific effects. We

1

show that the highly disordered prodomain can be meaningfully divided into domains based on sequence alone. Monte Carlo simulations of a self-excluding heterogeneous polymer, with monomers representing each domain, suggest the sequence would be effectively segmented by the long, highly disordered polyampholyte near the sequence midpoint. This is qualitatively consistent with observed interdomain contacts within the BDNF prodomain, although contacts between the two segments are enriched relative to the self-excluding polymer. The Val66Met mutation increases interactions across the boundary between the two segments, due in part to a specific Met-Met interaction with a Methionine in the other segment. This effect propagates to cause the non-local change in secondary structure around the second methionine, previously observed in NMR. The effect is not mediated simply via changes in inter-domain contacts but is also dependent on secondary structure formation around residue 66, indicating a mechanism for secondary structure coupling in disordered proteins.

Introduction

The physiological significance of intrinsically disordered proteins (IDPs), which can explore a wide range of conformational ensembles in their functional form, is now well-established.[1–4] More than 33% of eukaryotic proteins contain disordered regions longer than 30 residues,[3] many of which are involved in critical biological functions, including transcriptional regulation and cell signaling.[5] Long intrinsically disordered regions are particularly abundant among cancer and neurodegenerative-associated proteins.[6,7]

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[8] 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 (Figure 1c).[9,10] 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 nucelotide polymorphisms (SNPs) are found in IDPs;[11] 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[12–18] 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.[9,12,19,20] Locally, such mutations can modulate residual secondary structure preferences via forming or breaking local salt-bridges or by introducing helix breaking residues.[13,17,21]

For IDPs with a relatively low fraction of charged residues, typical of the Janus region of the state diagram proposed by Das and Pappu[9,10] (Figure 1c), 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.[11] The extent to which such substitutions in IDPs can affect non-local aspects of the conformational ensemble is uncertain; these substitutions directly affect 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;[13,22] for example, several cancer mutations in transactivation domain of tumor suppressor p53 can lead to helicity changes in residues sequentially

3

far away from the mutation sites.[13]

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.[6,23] and references therein). Molecular dynamics (MD) simulations have played a significant role in understanding IDP structure and dynamics ,[24–30] 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[34,35] (Figure 1a). It has been implicated in maintenance of the hippocampus, orientation selectivity in the visual system[36–38] and the mechanism underlying action of numerous antidepressants,[39,40] including rapidly acting low-dose ketamine.[41] 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.[33,41–44]
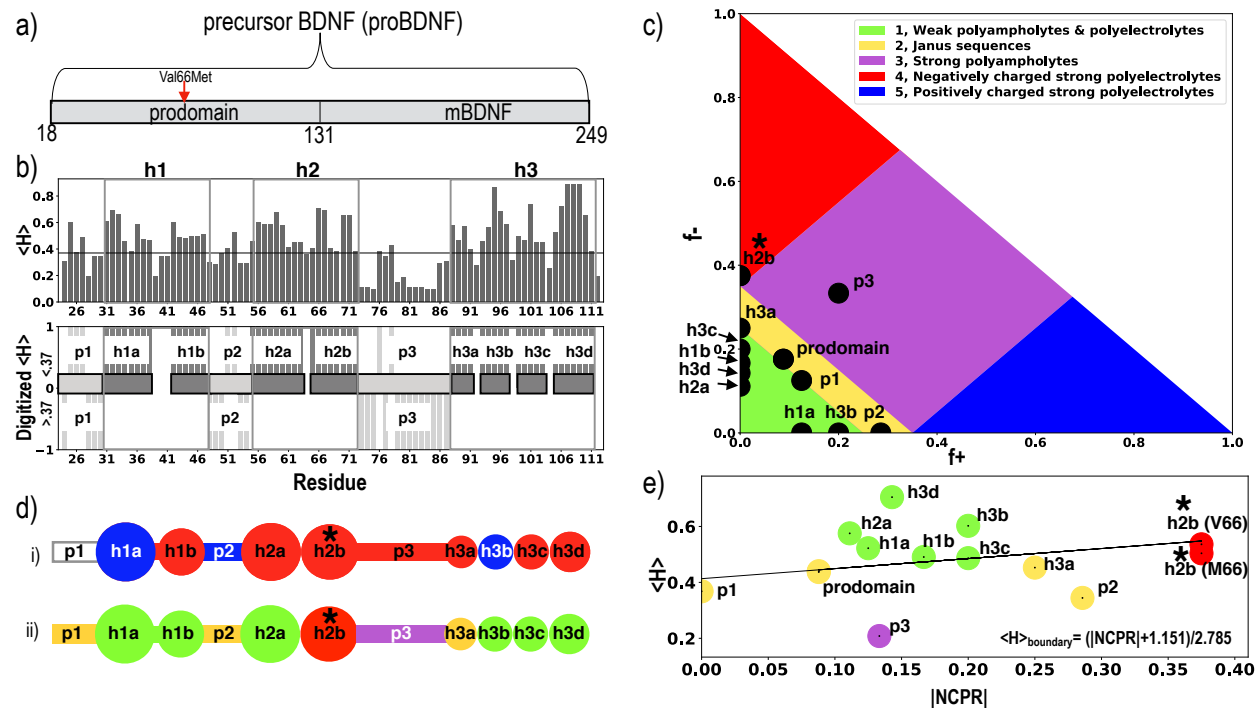

Figure 1: **Sequence-based decomposition of the BDNF prodomain.** a) The two functional domains of proBDNF: the disordered precursor prodomain considered in this manuscript and the structured mature domain BDNF (mBDNF). b) The mean hydrophobicity $\langle H \rangle$ per residue (top), given by the Kyte-Dolittle[45] score averaged over a three residue window, and scaled to fit between 0 and 1 was digitized (bottom) according to a cutoff at $\langle H \rangle = 0.37$. Three or more contiguous residues above the cutoff were identified as forming a hydrophobic domain. Eight hydrophobic domains (darkgrey) are identified along with 3 domains of low hydrophobicity (light grey). c) The diagram of IDP states proposed by Das and Pappu,[10] based on fraction of positive (f^+) and negative (f^-) charged residues, and annotated by the location of the whole BDNF prodomain sequence and each subdomain identified in b. d) Domains identified in b), colored according to net charge (i), with negative charge (red), positive charge (blue) and neutral (white) or region of the Das and Pappu diagram in c) (ii). e) Location of domains on an Uversky diagram[46] of IDPs and globular proteins, as a function of absolute net charge per residue ($|NCPR|$) and $\langle H \rangle$, with the boundary line between folded and disordered proteins given by the equation in the legend. The domain h2b contains the Val66Met mutation and is marked with star. Additional properties of the domain sequences can be found in Table 1.

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 lacked resolvable density.[47]

It was subsequently revealed that the cleaved prodomains ($\sim$ 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[48] and eliminated synapses in hippocampal neurons.[49] NMR measurements on the prodomain confirmed significant intrinsic disorder for both forms, with differential secondary structure preference around residue 66.[48] Tertiary contact distances from NOEs were not accessible, however, and uncertainty in interpretation of the NMR signal obscured non-local effects on secondary structure. Additional NMR experiments implicated residue 66 in binding of M66 prodomain to SorCS2.[48]

In this work, we report on unbiased fully-atomistic replica-exchange MD simulations of the 90 residue BDNF prodomain in explicit solvent, for V66 and M66 forms. This sequence falls in the Janus sequence region in the phase diagram proposed by Das and Pappu.[9,10] We identify domains using a sequence-based approach based on residue hydrophobicity, which suggests the heterogeneous sequence is composed of multiple globular and intrinsically disordered regions. We then use a hierarchical analysis approach to isolate which interdomain contacts reflect significant effects of the Val66Met substitution, identify the residue level interactions that contribute to these effects, and finally, discuss mechanisms underlying both local and non-local secondary structure effects previously observed through NMR.

Results and discussion

Prodomain Sequence Decomposition

The region of the BDNF prodomain studied using NMR,[48] and simulated here, is 91 residues long. The potential number of residue-residue contacts is $91 \times 90/2 \sim 4000$, and each contact is formed infrequently. Detecting significant differences for numerous weak signals is statisti-

Table 1: Sequence based properties of hydrophobic and linker domains identified in the Pro region of BDNF, as shown in Figure 1.

Domain	N[a]	NCPR[b]	$\langle H \rangle^c$	FCR[d]	f-[e]	f+[f]	κ^g	Sequence	R[h]	P[i]
p1	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	0.13
h1b	6	-0.17	0.49	0.17	0.17	0.00	0.1	TLESVN	1	0.00
p2	7	0.29	0.34	0.29	0.00	0.29	1.0	GPKAGSR	2	0.14
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
p3	15	-0.13	0.21	0.53	0.33	0.20	0.1	EDQKVRP 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	0.18	0.09	0.2		2	0.07
M66 Seq	91	-0.09	0.43	0.26	0.18	0.09	0.2		2	0.07

[a] Number of residues in the domain
[b] Net charge per residue
[c] Mean hydrophobicity, average of Kyte-Dolittle[45] scores for each residue in the domain scaled to fit between 0 and 1
[d] Fraction of charged residues
[e] Fraction of positively charged residues
[f] Fraction of negatively charged residues
[g] Charge distribution parameter κ as defined by Das and Pappu,[10] calculated using CIDER[50]
[h] Region in phase diagram proposed by Das and Pappu,[10] (Figure 1c)
[i] Fraction of Proline residues

cally prohibitive, even given the long simulations presented here. We were initially motivated to divide the sequence into domains based on sequence hydrophobicity, as described in Methods, to address this analysis challenge (Figure 1b). Such coarse-graining reduces the number of potential contacts to $11 \times 10/2 = 55$, while increasing the likelihood that any given contact will be formed. Three hydrophobic macrodomains were identified (Figure 1b): h1, h2, and h3. The h domains are separated by three charged linker ("p") domains : p1, p2, and p3. Each h macrodomain contains two to four microdomains, separated by one or two residues

each: h1 contains h1a and h1b, h2 contains h2a and h2b, and h3 contains h3a, h3b, h3c, and h3d.

Since each domain sequence has its own properties (Table 1), this process also suggested a new, more tractable conceptualization of the long, disordered BDNF prodomain. Each domain can be analyzed individually according to Das and Pappu metrics[10] (Figure 1c) or Uversky metrics[46] (Figure 1e), while several other sequence properties of each domain are shown in Table 1. The Das and Pappu phase diagram[10] predicts the compactness of IDPs based on their fraction of positively (f+) and negatively (f-) charged residues (Figure 1c). Hydrophobic domains h2b and domain h3a lie in the strong polyelectrolyte and Janus sequence region respectively. All the remaining hydrophobic domains are classified as weak polyampholyte and, as isolated peptides, would be predicted to have compact globule conformations to shield hydrophobic residues.[10] Linkers p1 and p2 also lie in the Janus sequence regions, while linker p3 lies in the strong polyampholyte region with low value of the charge distribution parameter κ,[10] and is predicted to have random coil conformations if present as an isolated peptide.

The Uversky diagram[46] characterizes proteins as globular or intrinsically disordered based on their normalized mean hydrophobicity and absolute net charge per residue ($|NCPR|$) (Figure 1e). The proteins falling above the boundary line are classified as globular proteins, while the ones below that line are generally classified as IDPs. With the exception of hydrophobic domains h2b and h3a, all hydrophobic domains identified here fall in the globular protein regions. Domains h2b, h3a and p1 fall on the disordered side of the boundary, while p2 and p3 are both deep in the disordered region.

The domain h2b containing Val66Met has several unique properties among the identified domains: 1) it is located at the sequence midpoint 2) it is the only strong polyelectrolyte domain 3) it has the strongest NCPR (-0.38) among all the domains 4) The sequence is composed almost entirely of two competing residue types, yielding the uncommon mix of a highly-charged, hydrophobic domain. Considering mean hydrophobicity alone, Uversky

8

et al.[46] found $\langle H \rangle \sim 0.48 \pm 0.03$ for a set of 275 folded proteins and $\langle H \rangle \sim 0.39 \pm 0.05$ for a set of 91 unfolded proteins. By this criteria, we would expect the h2b sequence to be folded: for V66-h2b, $\langle H \rangle \sim 0.54$, while for M66-h2b, $\langle H \rangle \sim 0.50$. The full Uversky diagram also considers NCPR, and the strong NCPR pushes h2b into the IDP region of the Uversky diagram.[46]

More specifically, this domain sequence (HV/MIEELLD) has hydrophobic residues at i, i+3, i+4 separated by acidic residues at i+1, i+2. Helix formation would thus segregate hydrophobic residues from acidic residues but would also increase the density of like-charge residues. Similar sequences are observed in the activation domains of transcription factors: a motif of alternating hydrophobic and acidic residues folds into an amphipathic helix upon binding, and the interactions between the amphipathic helix and the binding partner are mediated by hydrophobic residues, not charged residues.[51–55] Staller et al.[55] have earlier reported that in the disordered acidic activation domain of Gcn4, the acidic residues keep key hydrophobic residues exposed to solvent and binding partners.

The domain h3a is a unique hydrophobic Janus domain with high NCPR. Janus sequences have intermediate compositional biases and their conformations are context dependent.[10,56] The SNP domain h2b and the Janus domain h3a are separated by the long (15 residue) strong polyampholyte linker p3, which has well mixed charge ($\kappa = 0.1$). The domains h1a and h3b are positively charged and all the remaining hydrophobic domains are negatively charged (Figure 1d).

Comparison of experimental observables and their computational analogues

NMR chemical shifts (Figure 2a) for the Val66 and M66 prodomain indicated that they are disordered, lacking any stable secondary structure.[48] Many of the common force-field and water model combinations used for MD simulations are optimized for folded proteins, and are not recommended for IDPs.[57,58] Piana el al[58] showed that several such force-field and water

model combinations produced substantially more compact disordered states when compared with experiments. In order to predict accurate ensembles of the prodomain, we tested several force-field and water model combinations, optimized for IDPs, including Amberff03sbws[59,60] with TIP4P/2005,[61] Amber99sbws[60,62] with TIP4P/2005[61] and Amber99sb*-ildn-q[62,63] with Tip4p-D,[58] on 30 residue fragments of the V66 prodomain using temperature replica exchange molecular dynamics (T-REMD), further described in SI. Amber99sb*-ildn-q[62,63] with Tip4p-D[58] yielded the best agreement with experimental data obtained by NMR[48] (FigureS1) and was used for the full prodomain MD simulations further described in Methods. This is consistent with the force field comparison study by Robustelli et al., which observed that for IDPs with little or no secondary structure, Amber99sb*-ildn-q with Tip4p-D gives best agreement with experimental NMR measurements.[64]

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. Figure 2a shows the C_α and C_β chemical shifts calculated from the simulations using SPARTA+[65] and compares them with the NMR chemical shifts obtained from Anastasia et al.[48] for V66 sequence and M66 sequence (Figure S2). 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.[65]

The simulated hydrodynamic radii (R_h) calculated using Hydropro[66] for both the V66 ($R_h = 2.23 \pm .01$ nm) and M66 ($R_h = 2.20 \pm .01$ nm) sequences are in excellent agreement with the experimental values from NMR diffusion measurements[48] ($R_h = 2.24 \pm 0.1$ nm and $R_h = 2.20 \pm 0.1$ nm for the V66 and M66 sequence respectively) (Figure 2b). Comparison between R_h generated from MD and NMR/SAXS experiments has been used earlier to validate simulations of disordered proteins[57,67,68] and our results support the use of Tip4p-D for such simulations.[58,64] Use of Tip3P resulted in more compact ensembles (data not shown). The simulated radius of gyration (R_g) distribution for both the V66 and M66 sequences demonstrate that both prodomains populate closely overlapping ensembles (Figure 3a).
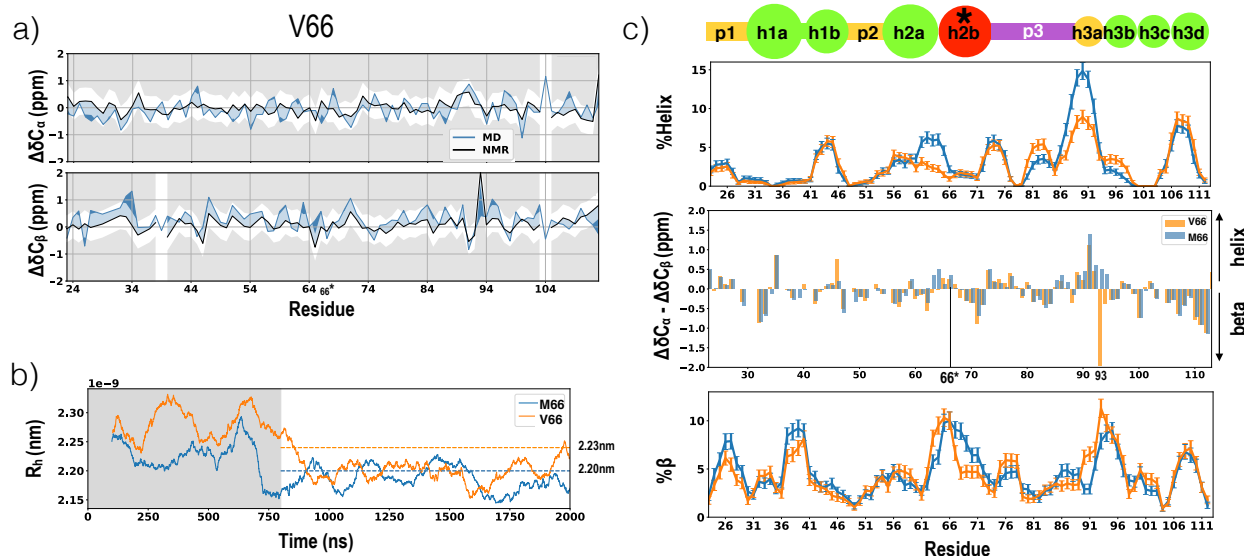
Figure 2: **Comparison of MD and NMR observables.** a) Chemical shifts from NMR at 280K (black lines)[48] and MD at 300K (blue lines), calculated using SPARTA+ .[65] The gray region represents a discrepancy of more than 0.5 ppm from NMR chemical shifts. b) Hydrodynamic radius (R_h) vs the simulation time, using a 100 ns moving window. The NMR diffusion experiment values[48] are represented with dashed lines, and the shaded region represents the discarded equilibration period. 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 (top) region or beta (bottom) region of the Ramachandran map (further described in methods). C_α-C_β secondary chemical shifts for each sequence from Anastasia et al.[48] (middle) are also shown. Positive differences indicate helical structure and negative differences indicate β structure. Errors represent standard error of a Bernoulli trial with n number of samples, where n is the product of total number replicas, 64 and average number of roundtrips per replica, 17.

Anastasia et al.[48] 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 (Figure 2c). Consistent with NMR experiments,[48] the M66 sequence demonstrates an increased tendency of forming helices within domain h2 and h3a relative to the same domains in the V66 sequence (Figure 2c). We also find that the residue R93 in domain h3b of V66 sequence has a $\sim 30\%$ higher tendency to form β-sheets than in the M66 sequence, in agreement with NMR experiments.

11

Heterogeneous behavior of individual domains

Disordered proteins can be well-described by Flory scaling theory $\langle R_{|i-j|} \rangle = An^\nu$, where $\langle R_{|i-j|} \rangle$ is the ensemble-averaged internal distance, |i-j| is residue separation along the chain, and ν is the Flory scaling coefficient.[69] Larger values of ν correspond to swollen coils, while smaller values correspond to compact globules.[10] In particular, when ν=0.6 ("good solvent") the protein maximizes its interaction with solvent, and for ν=0.33 ("poor solvent"), the protein maximizes self-interactions. The special intermediate case of ν=0.5 is called a "theta solvent".[69] Most IDPs that obey this scaling behavior have ν>0.5.[10,68,70,71] As shown in Figure 3b, the prodomain as a whole is not well fit by a single power law: for separations of 15 or fewer residues the prodomain falls in the "theta solvent" regime, while for separations of 20 or more residues it falls in the "poor solvent" regime.

Table 2: Average radius of gyration (R_g), Average end to end distance (R_{etoe}) and Flory exponent (ν) for each proregion domain, calculated from MD trajectories. Errors represent fit error of $\langle R_{|i-j|} \rangle$ vs An^ν, weighted by each point's standard deviation. Values for V66 and M66 sequences were within error.

Domain	$\langle R_g \rangle$ (nm)	$\langle R_{\text{etoe}} \rangle$ (nm)	ν (V66)	ν (M66)
p1	0.67 ±0.01	1.60 ±0.05		
h1a	0.67 ±0.01	1.60 ±0.06	0.49 ±0.11	0.53 ±0.11
h1b	0.53 ±0.01	1.21 ±0.04		
p2	0.62 ±0.01	1.38 ±0.05		
h2a	0.66 ±0.01	1.47 ±0.06	0.42 ±0.11	0.43 ±0.11
h2b	0.68 ±0.01	1.80 ±0.05		
p3	0.97 ±0.01	2.16 ±0.08	0.63 ±0.10	0.62 ±0.10
h3a	0.46 ±0.00	0.93 ±0.02		
h3b	0.55 ±0.01	1.23 ±0.03	0.58 ±0.08	0.54 ±0.08
h3c	0.56 ±0.00	1.27 ±0.03		
h3d	0.62 ±0.01	1.53 ±0.04		

Each identified individual domain does obey a power law, and we calculated A and ν for each domain as if it was isolated from rest of the protein (Table S1, Figure S3). For every domain except the p3 linker, the error across ν values was substantial, since N was short. However, each hydrophobic macro-domain h1,h2 and h3 can be fit by a power law, with the Flory exponents shown in Table 2. The lowest observed value of ν was observed in
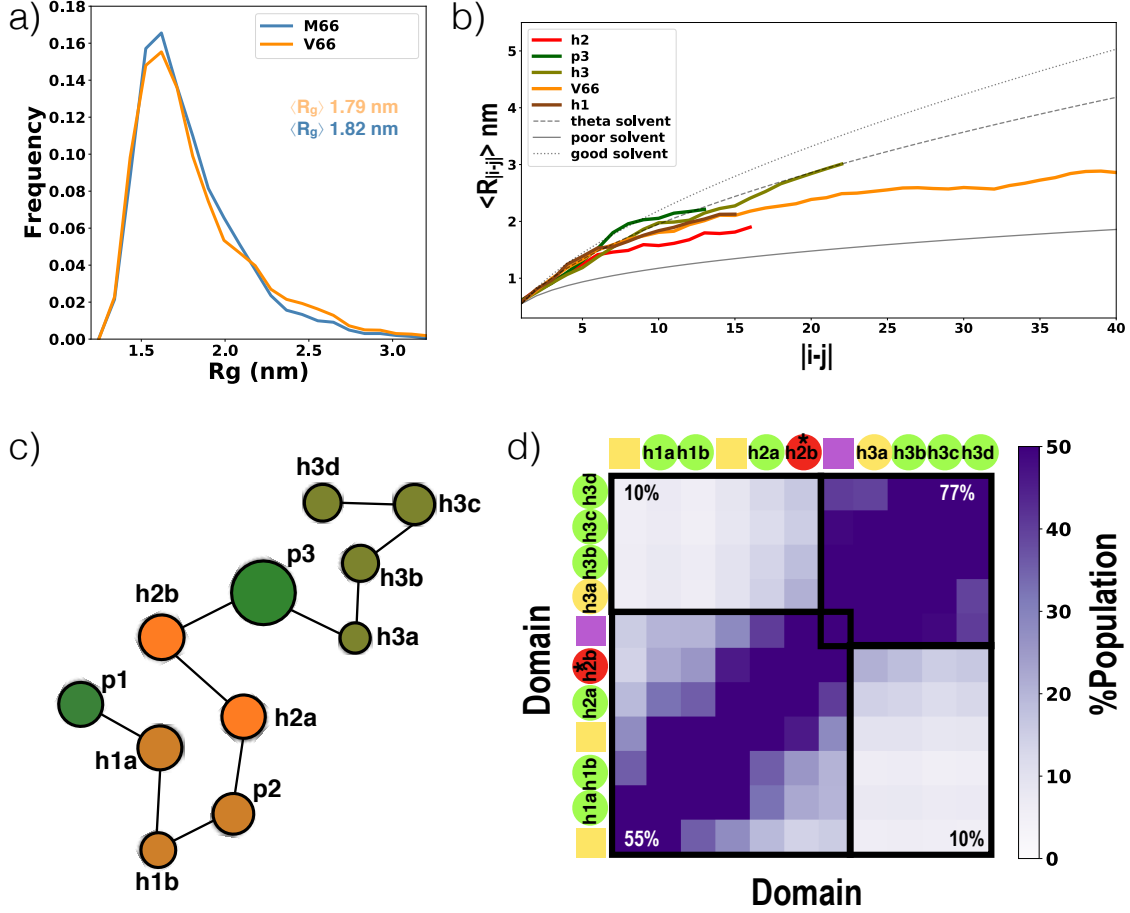
Figure 3: **Heterogeneity in the BDNF prodomain.** a) R_g distributions for the V66 and M66 sequence, with $\langle R_{g,M66} \rangle = 1.79 \pm 0.01$ nm and $\langle R_{g,V66} \rangle = 1.82 \pm 0.01$ nm. b) Ensemble averaged intrachain distance profiles for the V66 sequence, as well as each hydrophobic macro-domain and the disordered p3 linker. Theoretical polymer scaling limits are shown with dashed lines (prefactor A = 0.55 nm) while Flory exponents for each curve are given in Table 2. c) Model for the freely-jointed self-avoiding heteropolymer mimicking the BDNF prodomain, drawn to scale, with the size and separation between beads reflects R_g and R_{etoe} for each domain (Table 2). The domains are colored according to their Flory exponents: $\nu >= 0.6$ is green, $\nu <= 0.45$ is orange and the remaining ν in the "theta solvent" range is colored brown. d) Inter-domain contact probability for the model in (c), calculated as described in Methods. The black boxes mark the regions on either side of the h2b-p3 boundary, and each is annotated with the average contact probability for any domain-pair falling within it.

h2 ($\nu = 0.42 \pm 0.11$) and the highest value was in the adjacent p3 domain ($\nu = 0.63 \pm 0.1$), suggesting a sharp boundary at the h2-p3 interface, where the sequence shifts from highly collapsed ("poor solvent") to highly expanded ("good solvent") tendencies.

13

We expect that even for a heterogeneous freely-jointed, self-avoiding polymer, contact probability between monomers would depend on monomer shape and separation. To predict the inter-domain contact probability for a self-avoiding polymer composed of 11 monomers with properties similar to our domains (analogous to Kuhn segments[72]), we performed simple Monte Carlo calculations as described in methods. The monomer-monomer bond length (analogous to the Kuhn length) and excluded volume radius reflected the average end to end distance (R_{etoe}) and radius of gyration (R_g) for each domain from MD simulations (Table 2), and a schematic drawn to scale is shown in Figure 3c. The predicted interdomain contact probabilities for this freely-jointed, self-avoiding heteropolymer are shown in Figure 3d, suggesting that such a chain would effectively be segmented at the boundary between the domains mimicking h2b and p3. More specifically, domains on the N terminal side of p3 are in contact in 55% of the frames, while domains on the C terminal side are in contact in 77% of the frames, but the average contact probability between these two self-interacting regions is 10%. Shifting the p3 bead within the chain shifts the boundary (Figure S4). These calculations provide a useful reference expectation for interpreting interdomain contacts in the real sequence.

Interdomain and tertiary interactions

Figure 4a shows the probability of interdomain contacts for both the V66 and M66 sequences, calculated analogously to those in the self-avoiding heteropolymer in Figure 3d. Consistent with the MC predictions, the contact probability map for each sequence clusters into 4 visible regions, separated along both axes by domain p3. The total number of interdomain contacts that did not cross the boundary were quantitatively consistent with the self-avoiding heteropolymer, exhibiting frequencies that were only 20% more and 10% less than the self-avoiding heteropolymer frequencies on the N-terminal and C-terminal side respectively. These frequencies were the same for both the V66 and M66 sequences.

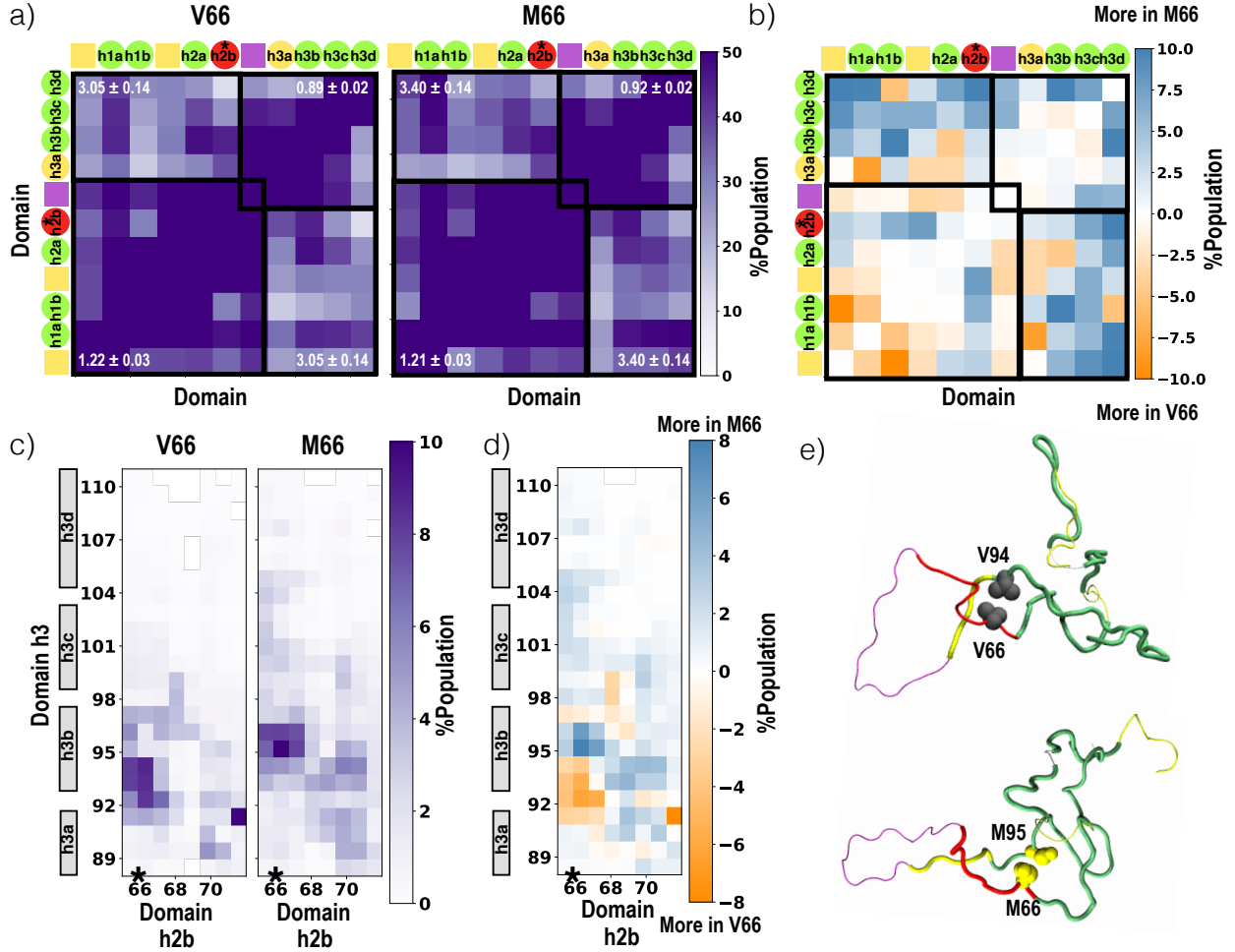In the real protein, contacts crossing the h2-p3 boundary are much less common than

Figure 4: **Effect of Val66Met on inter-domain contacts.** a) Inter-domain contact probability for the V66 (left) and M66 (right) sequences, calculated as described in Methods. The black boxes mark the regions identified in Figure 3d for the self-avoiding heteropolymer, and the enrichment for the total number of contacts, relative to the totals in Figure 3d, are annotated for each box. The errors represent standard errors (n =1088 as described in Methods). The x and y axes are annotated with cartoon representations of the prodomain; circles and squares represent hydrophobic domains and linker regions respectively and are colored as in Figure 1. 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 in h2b with each residue in h3 for V66 (left) and M66 (right). d) Difference between the contact probabilities shown in c); interactions colored as in b). e) Representative conformation of V66 sequence (top) and M66 sequence (bottom) showing Val66-Val94 and Met66-Met95 contact respectively, colored as in Figure1 with Methionine in yellow. Tubes represent hydrophobic domains whereas lines represents linker regions.

those on either side of the boundary, which is qualitatively consistent with the predictions for the self-avoiding heteropolymer (Figure 4a). Quantitatively, they are still about three

times more common than predicted from the self-avoiding heteropolymer, consistent with favorable interdomain interactions across the boundary. Furthermore, the extent of enrichment changes with the Val66Met mutation: the V66 sequence has $3.1\times$ enrichment while the M66 sequence has $3.4\times$ enrichment. The increased number of cross-boundary contacts are also consistent with the lower mean hydrodynamic radius (Figure 2b) and radius of gyration (Figure 3a) for the M66 sequence.

This difference suggested that the presence of residue 66 near the collapsed/expanded boundary could contribute to its conformational effects, and consistent with this hypothesis, the M66-h2b domain contacts every domain on the other side of the boundary more frequently than does the V66-h2b domain (Figure 4b).

To understand the origin of the favorable interaction between the SNP domain h2b and the C-terminal side of p3, we consider the contacts between the residues of these domains (Figure 4c, Figure S5). As shown in the absolute contact probability map (Figure 4c), the SNP domain h2b forms most frequent contacts with domain h3b in the C-terminal side of p3. At residue level, both sequences frequently form contacts between hydrophobic residues at domains h2b-h3b, but the residue pairs most frequently forming the contact shift from Val66-Val94 in the V66 sequence to Met66-Met95 in the M66 sequence (Figure 4c and d). This result was remarkably consistent with a role for residue 66 in driving these interdomain interactions, as well as sensitivity to the Val66Met substitution.

Specific Met-Met interactions between the two polarizable sulfur atoms are often under appreciated, but are common in structures of folded proteins.[73] Using ab initio calculations, Gómez-Tamayo et al. predicted that Met-Met interactions are stronger than Met-aromatic or aromatic-aromatic interactions.[74] These interactions are also captured reasonably well in the non-polarizable AMBER force field.[74] In these simulations, the Met66-Met95 contact was about five times as common (10% of frames) as the analogous Val66-Met95 contact (2% of frames). Methionine-aromatic interactions also contribute to the increased number of cross boundary contacts: M66, but not V66, forms a frequent contact with F108 in domain

16

h3d, which is also consistent with the favorable interactions between Met-Phe residues[73,75,76] (Figure 4d). While determining intermolecular interactions that cause compaction in the disordered domain of poly(A)-binding protein using small-angle X-ray scattering (SAXS), Riback2017 et al.[77] reported that increasing the net hydrophobicity produces increased domain compaction across multiple sets of mutations, with one exception: substituting the less hydrophobic Methionine with more hydrophobic Valine.

Intradomain interactions and residual secondary structure

NMR chemical shifts[48] indicate that the Val66Met mutation increases the helicity within 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 (Figure 2c). Comparing the length of secondary structure formed at each residue (Figure 5a and c) reveals an even stronger effect of the mutation that would not have been detectable via NMR: Val66Met consistently increases the frequency of long helices formed within domain h2.

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. 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.[78] In our simulations, the likelihood that V66 will be in a short helix decreases with temperature, while the opposite effect is observed for M66 (Figure S6).

The helical structure in M66 is also stabilized by local sequence: the preferred interaction of Met66(i)-Phe63(i-3). MD simulations have previously shown the stability of a sulfur-aromatic contact in a model helix.[75] Figure 5b shows the residue level contact map within domain h2. For the M66 sequence, Met66(i) is frequently contacting Phe63 (i-3) when compared with any other intra-domain contact at residue 66: M66-F63 is formed 60% more often than M66-E69 (Figure 5b). We find that the largest change in intra-domain contacts
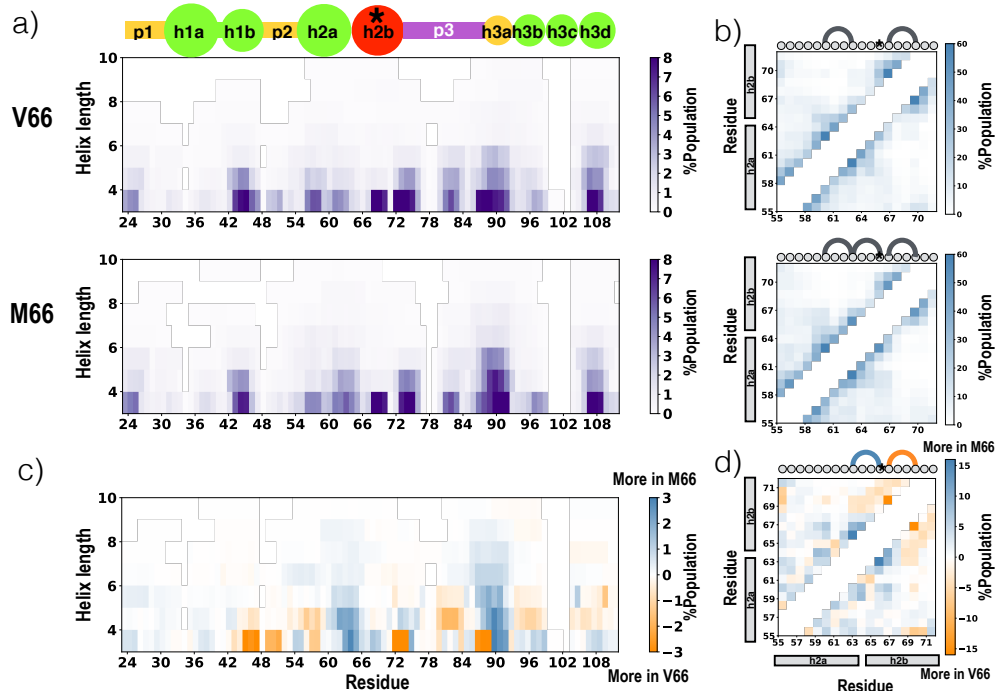
Figure 5: **Effects of Val66Met on intra-domain contacts at residue level.** a) Helix length probability at each residue and b) Contact probability for each residue pair within the h2 domain for V66 sequence (top) and M66 sequence (bottom). Each residue in domain h2 is annotated with a circle representation and the contacts with at-least 50% population are represented with an edge. c) Difference in helix length probability shown in a), d) Difference between the contact probabilities shown in b); interactions more frequently found in M66 or V66 are in blue and orange respectively.

from V66 to M66 is the gain of contact at M66-F63 (40% stronger in M66 when compared with V66) followed by loss of contact at Ile67 (i+1)-Leu70(i+4) (30% weaker in M66 when compared with V66) (Figure 5d). This is also consistent with previously identified Met-Phe interactions[73–76]

Secondary structure coupling via tertiary interactions

While the effects of the Val66Met mutation on secondary structure in the domain (h2b) which contains residue 66 are not unexpected, we also observed a correlated effect on secondary structure in domain h3. As shown in Figure 2c, the increased frequency of long helices for domain h3a in the Met66 sequence is comparable to the increase in domain h2b. This is

18

consistent with NMR results [48] showing a change in chemical shifts at residue 93 (Figure 2c), although this effect was not originally discussed in Reference.[48] We hypothesized that since the Val66Met mutation affects the contact frequency between domain h2b and h3ab, the secondary structure change in h3ab could be due entirely to the change in these inter-domain contacts. To test this, we divided all frames into 4 clusters, representing two independent collective variables with two possible values each: either the h2b-h3a inter-domain contact is formed or broken, and either domain h2 does or does not contain a stretch of secondary structure of a given length. This clustering process on all frames was carried out twice, for each of two possible secondary structure elements (4 continuous residues in helix or 4 continuous residues in β). Populations of the individual clusters are shown in Table 3. The percent population of clusters with both respective secondary structure and h2b-h3a contact formed does not differ significantly from the uncorrelated expectation, also shown in Table 3.

Table 3: Probabilities of simultaneous h2b-h3a interdomain contact and secondary structure formation in h2. Each clustering process divided the frames based on two categorical variables with two possible values each: either the h2b-h3a inter-domain contact is formed or broken, and either domain h2 does or does not contain four contiguous residues in the given secondary structure. Each sequence was clustered twice : once for helical secondary structure and once for β-strand. The expected cell frequencies, assuming no correlation, are shown in parenthesis. According to a Fisher exact test, $p > 0.01$ for all four contingency tables, even with generous estimates for the number of uncorrelated samples.

	V66		M66	
	h2 secondary structure		h2 secondary structure	
h2b-h3a contact	helix	not helix	helix	not helix
formed	3.6 (4.4)	34.3 (34.2)	6.6 (5.6)	32.9 (33.1)
broken	8.1 (7.3)	54.0 (54.8)	7.6 (8.6)	52.8 (52.7)
	β	not β	β	not β
formed	7.2 (8.5)	30.7 (30.4)	10.7 (9.2)	28.9 (29.2)
broken	15.2 (13.9)	46.9 (47.2)	12.7 (14.2)	47.7 (47.4)

For any cluster, the secondary structure in the SNP-domain h2 is constrained by the clustering process. Helical or β strand secondary structure propensities for all other residues, averaged across the frames in the cluster, are shown in Figure 6. It would be very surprising to observe correlations in secondary structure between non-contacting domains in converged
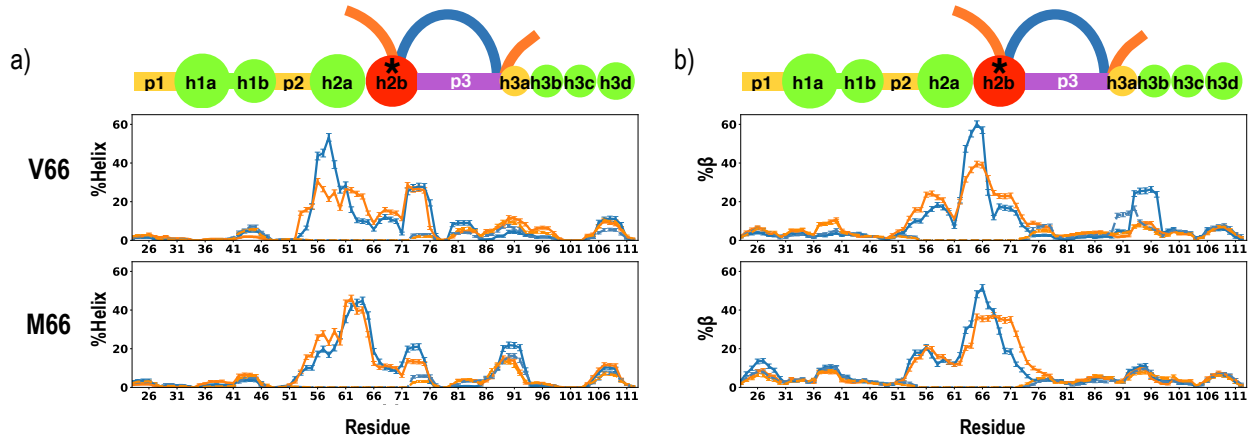
Figure 6: **Secondary structure coupling between domains on each side of the p3 linker.** Helix (a) and β (b) propensities at each residue in V66 sequence (top) and M66 sequence (bottom) for four clusters. Frames were first clustered by whether the inter-domain h2b-h3a contact was formed (blue) or broken (orange), and then by whether the respective secondary structure was present in h2 (solid) or absent (dashed). Errors represent standard error of a Bernoulli trial with n number of samples, where n is the product of total number of unique replicas in a given cluster and average number of roundtrips per replica, 17.

data; most likely such an observation would indicate a single dominant replica. Reassuringly, we detect no significant correlation of secondary structure in domains h2 and h3a unless the two domains are actually in contact: the orange solid and orange dashed lines overlap in Figure 6. If our hypothesis was supported, h3a secondary structure would vary between clusters with and without the h2b-h3a contact (blue vs. orange lines in Figure 6 would not overlap) but would not depend on secondary structure in h2 (blue solid vs blue dashed lines in Figure 6 would overlap). Overall, we find that whether this hypothesis is supported for either helix-helix or β-β coupling depends on whether residue 66 is Valine or Methionine.

Figure 6a compares the helix propensities at each residue for each cluster in the V66 and M66 sequences. In the V66 sequence, the clusters forming the h2b-h3a contact are less likely to be helical within h3a, regardless of whether h2 is in helix, supporting our original hypothesis that the presence of the h2 domain was sufficient to affect the secondary structure of the "context-sensitive" Janus domain h3a. In the M66 sequence, when a helix is formed in h2 and h2b is in contact with domain h3a, h3a is more likely to form a helix than if domain h2 is not helical (Figure 6a). Furthermore, if domain h2 is not helical, helical propensity

20

in domain h3a is indistinguishable from clusters in which the two domains are not even in contact. Thus, the interdomain contact is not sufficient to increase helical propensity: helix formation in domain h2 is also necessary for this effect in domain h3a. Consequently, our hypothesis is not supported for the M66 sequence. Overall, these results suggest that contact with the highly negatively charged h2b domain containing a Valine at residue 66 (regardless of its structure) is sufficient to slightly disrupt helical formation in the Janus domain h3a. When Val66 is replaced with Met, the interaction between the two domains becomes cooperative, consistent with helix-helix coupling.

Few studies have reported helix-helix stabilization in IDPs. Feuerstein et al.[79] reported partly stabilized helical structure by long-range tertiary contacts in a disordered fragment of the hepatitis C virus protein NS5A with PRE measurement from NMR. Conicella et al.[21] 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.

Interdomain coupling between β-strands is expected, even for disordered proteins: non-hairpin β-strand pairs cannot be local contacts. Any mutation in an IDP that disrupts local β-strand propensity should simultaneously reduce propensity of the partner strands. In the V66 sequence, we find that the h2b-h3a contact is associated with β-strand within h3a (Figure 6b). In the M66 sequence, we do not observe any such correlation.

To determine which residue contacts between h2b and h3ab couple secondary structure within the two domains, we calculated the contact probability at residue level for the cluster with h2b-h3a contact and β or helix within h2 for both sequences (Figure 7). In the V66 sequence, 25% of the frames simultaneously forming the h2b-h3a contact and helix within h2 also form β within h2. The β pairing at h2b-h3ab is stabilized via combination of backbone hydrogen bonds between V66 and S92, salt bridge between E64 and R93, and hydrophobic interactions between V66 and V94 (Figure 7a). In the M66 sequence, contact between h2 helices and h3ab is relatively unspecific and distributed among a number of possible
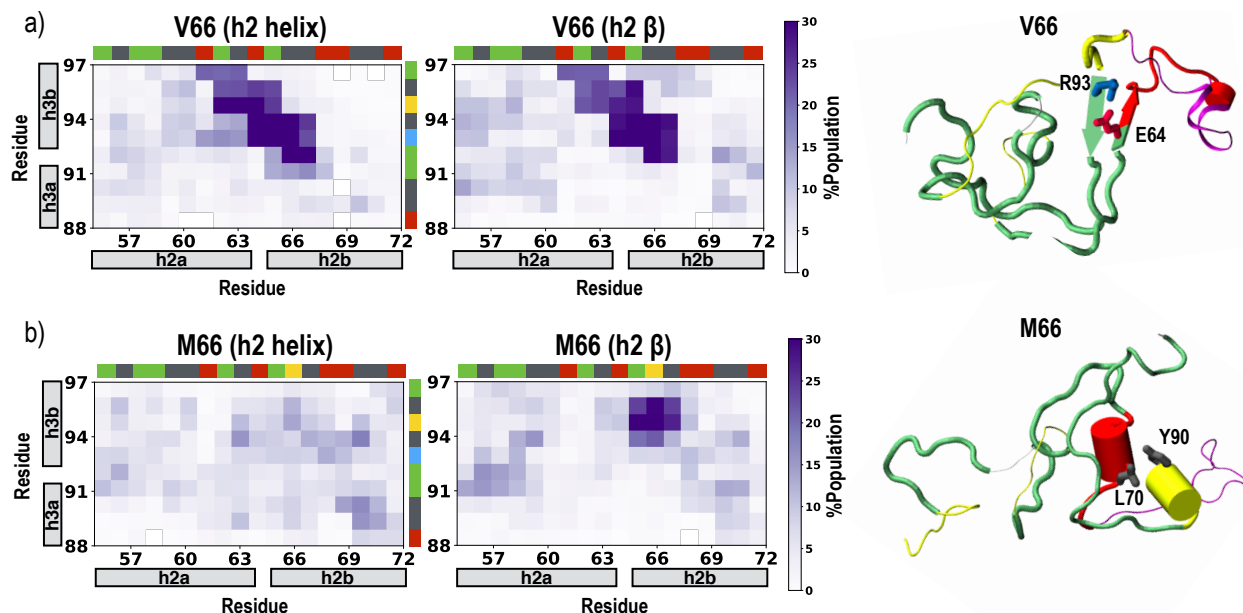
Figure 7: **Effect of secondary structure in domain h2 on which residues form the cross-boundary h2-h3 contact.** Contact probability between each residue in h2 and h3ab, when h2 contains either helix (left) or β (right) simultaneously with h2b-h3a contact, corresponding to the cluster represented by solid blue line in Figure 6, for both the V66 (a) and M66 (b) sequences. Axes are annotated with boxes colored according to residue type: blue:basic, red:acidic, green:polar or Gly, grey:hydrophobic and Met: yellow. Representative conformations in V66 sequence (top) and M66 sequence (bottom) for each contact probability with chain colored as in Figure 1 and preferred residue-level contacts in Licorice representations, with residues colored by residue type.

hydrophobic contacts, particularly L70-Y90, as well as the M66-M95 interaction discussed in the previous section (Figure 7b).

Summary and Conclusion

We have carried out over 250 μs of fully-atomistic explicit solvent MD simulation of the 91 residue prodomain of brain-derived neurotrophic factor, with and without the disease-associated Val66Met mutation. These long simulations successfully reproduced the experimentally observed chemical shifts and R_h. The simulations also correctly reproduced the location of both local and non-local secondary changes due to the Val66Met mutation in the prodomain sequence.

We find that the highly disordered 91 residue prodomain, which as a whole falls in the Janus sequence region of the Das and Pappu phase diagram,[10] can be meaningfully divided into 11 domains based on sequence hydrophobicity alone. Among 8 hydrophobic domains, we identified 2 domains in the disordered region: the strong polyelectrolyte domain h2b (which contains Val66Met), and the Janus domain h3a. These are connected via a highly disordered long linker region p3. These unique domains have biological significance as well: The sequence h2-p3-h3 is essential for intracellular trafficking of proBDN.[80]

The domain decomposition suggested a tractable approach for coarse-graining analysis, by reducing the initial number of potential contacts from over 4000 to 55, while increasing the number of observations for each contact. Furthermore, it allowed us to isolate the most useful regions of the protein for examination at the residue level. This domain decomposition method, simply based on sequence hydrophobicity, may be a generally useful strategy to identify functionally significant regions in proteomics investigations for any long disordered proteins, and our conclusions suggest an important role for disorder heterogeneity within disordered proteins.

We were able to identify mechanisms through which a charge-neutral mutation can affect a disordered proteins residual secondary structure and tertiary contacts, as well as how these effects can be propagated to non-local residual secondary structure. Within domain h2b the Val66Met mutation affects intra-domain contact preference due to local sequence effects (preferred Met-Phe contacts) and the reduced entropic cost of helix formation for the Methionine sidechain. Consistent with the sequence characteristics of Janus sequences, we find that the secondary structure of the Janus domain h3a is sensitive to contacts with the highly charged h2b domain. For the Met66 sequence, this sensitivity is only observed when the contacting h2b domain is also helical. We did not detect such extreme sensitivity in other domains, although the entire prodomain does fall in the Janus region on the Das and Pappu diagram in Figure 1b. Thus, this designation may only be predictive for shorter segments of the sequence that would also fall in the Janus region on their own.

The long, disordered, exposed p3 linker segregates the inter-domain contact probability map: domains on the same side of p3 have a high probability of contact whereas probability of contact between either side of p3 is less probable. We consistently observed this segregation in both simple self-avoiding hetropolymer simulations with beads mimicking identified domains, and actual prodomain simulations. Val66Met increases the frequency of contacts between domains on either side of p3 due to favorable Met-Met or Met-Phe contacts. These interactions have been shown to stabilize tertiary contacts in folded proteins and membrane proteins, but their role has not been investigated in disordered proteins. In general, our study supports the observation that Methionine is distinct from generic hydrophobic residues due to the presence of polarizable Sulphur.[74,81]

Anastasia et al.[48] observed differential kinetics for interactions between the BDNF prodomain and SorCS2; M66 binds more preferably at the SNP domain h2b (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) Helix propensity at the SNP domain in M66. This SNP domain helix segregates all the acidic and hydrophobic residues on either side of helix. Therefore, it is likely that this preformed structure will be helpful for binding B) preferred Met-Met contacts when compared with Val-Met contacts. We find frequent interactions of Met66 with the only other Methionine in the prodomain, and the SorCS2 surface is rich in exposed methionines.[82] Therefore, it is possible that the M66 prodomain binds strongly to SorCS2 due to preferred Met-Met contacts between residue Met66 and the SorCS2 Met residues or a combination of both A) and B). This hypothesis could be tested experimentally by mutation of the methionines in SorCS2.

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,[83–85] Rosetta[86] and Modeller,[87] 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 replica was placed in a dodecahedron water box with 30,500 TIP4P-D [58] 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 field[62,63] and the GROMACS 5.1.2 simulation package,[88,89] with a time step of 2 fs. Long-range electrostatics are calculated using the particle mesh Ewald (PME) method,[90] with a 1 nm cutoff and a 0.12 nm grid spacing. Periodic boundary conditions are also used to reduce system size effects. System was simulated using T-REMD[91] with an exchange frequency of 1ps for 2 μs, giving a total simulation time of 128 μs with NVT ensemble for each system. 64 replicas are used with temperatures ranging from 300-385K, with exponential spacing. A different random seed was used for the Langevin dynamics of each replica. The average exchange acceptance probability ranged between 0.19-0.23.

The minimum separation between the molecule and its image was less than 2 nm only <1% of the time for both sequences and these frames were discarded from all the analysis. Time-series of the relative measurements were generated every 100 ps. For both V66 and M66

groups, initial 0.5 μs (800 ns $\times$ 64) were discarded for equilibration purposes, determined by plateauing of R_h (Figure 2b). Over the course of remaining 76.8 μs (1.2 μs $\times$ 64) simulations, for both the V66 and M66 sequence each replica completes a minimum of 5 roundtrips and an average of 17 roundtrips (Figure S7).

Time-series of the radius of gyration R_g and end-to-end distance R_{etoe} were calculated using respectively the g_gyrate and g_polystat utilities of Gromacs. We took R_{etoe} as the distance between N-termini and C-termini N and O atoms respectively. Statistical uncertainties are provided for R_g, R_h as total standard deviation divided by the square root of the product of total number of replicas (64) and average number of roundtrips per replica, 17.

Domain identification

Mean hydrophobicity ($\langle H \rangle$) at each residue is defined as the average Kyte-Dolittle[45] score with a window size of 3 residues, scaled to fit between 0 and 1. Any stretch of four or more residues with $\langle H \rangle$ of 0.37 or more is classified as hydrophobic domain and from the remaining residues, stretch of four or more residues is classified as linker region.

Chemical Shifts calculation

Prior to the present study, Anastasia et al.[48] 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+ [92] and SSP.[93] For comparison with simulation data, we reinterpreted the chemical shifts directly from,[48] deposited at Biological Magnetic Resonance Bank. Random coil reference chemical shifts were obtained from the Poulsen IDP/IUP random coil chemical shifts[94–96] at pH 7, 280 K.

Hydrodynamic radius calculation

The values for the Hydropro [66] parameters were: atomic level model with shell-method calculation, a = 0.29 nm, 6 minibead iterations, σ = 0.1 to 0.2 nm. The temperature was taken to be 300 K, the solvent viscosity 0.01 Poise, the solvent density 1.0 gcm^{-3}, the partial specific volume of the peptide 0.7313 $cm^3 g^{-1}$ (V66 sequence) or 0.7304 $cm^3 g^{-1}$ (M66 sequence), and a molecular weight of the peptide equal to 10044 Da (V66 sequence) or 10076 Da (M66 sequence).

Secondary structure calculation

Helix propensity or β propensity is expressed as the probability of a given residue being part of a sequence of four consecutive residues whose dihedral angles place them in the helical region or β region of the Ramachandran space. The helical region is defined as -100° $<\phi$ $<$-30° and -120° $<\psi$ $<$50°.[26,97,98] The β region is defined as ϕ $<$-80° and 50° $<\psi$ $<$-120°. The length of SS (SS-map)[99] were calculated with the above defined helical and β region. The error bars are calculated with standard error of a Bernoulli trial with n number of samples, where n is the product of total number of unique replicas in a cluster and average number of roundtrips per replica, 17.

Self-avoiding heteropolymer simulation

The BDNF prodomain was approximated as a freely-jointed self-excluding heteropolymer with 11 monomers, each mimicking one of the domains identified in Figure 1. The separation between monomers i and $i+1$ (analogous to the Kuhn length for a homopolymer[72]) was constrained to be half the end to end distance for each of the analogous domains:

$$|\vec{r}_{i-1} - \vec{r}_i| = \frac{\langle R_{etoe,i-1} \rangle + \langle R_{etoe,i} \rangle}{2} \tag{1}$$

where $\langle R_{etoe,i} \rangle$ was determined from the coordinates of domain i residues in the MD

simulations, given in Table 2.

Two monomers i and j are considered to be overlapping if

$$|\vec{r}_i - \vec{r}_j| < a(\langle R_{g,i} \rangle + \langle R_{g,j} \rangle) \tag{2}$$

where $\langle R_{g,i} \rangle$ was determined from the coordinates of residues in domain i in the MD simulations (Table 2), and a is a constant. In the MD simulations of the real protein, we observed that the center of mass distances between domains were rarely less than 0.3 $(\langle R_{g,i} \rangle + \langle R_{g,j} \rangle)$, and thus we set $a = 0.3$.

The random walk was carried out using a simple Metropolis Monte Carlo, with the following move set: 1) a random bead $i > 0$ was selected, 2) a random displacement vector $\vec{\delta r}$ was generated in three cartesian dimensions, 3) $\vec{\delta r}$ was scaled so that $|\vec{r}_{i-1} - (\vec{r}_i + \vec{\delta r})| = (\langle R_{etoe,i-1} \rangle + \langle R_{etoe,i} \rangle)/2$, satisfying Eq 1, 4) the translation $\vec{r}_j \to \vec{r}_j + \delta_r$ was applied for all $j \geq i$.

Any trial move that caused an overlap according to Eq. 2 was rejected, while all others were accepted. The MC simulation was run for 50,000 steps (50,000 for each moveable bead); additional steps did not change the outcome in Figure 3.

Contact maps

Two domains i and j are in contact if the excess distance $(d_{e,ij})$ between the two is less than 0.55 nm, where excess distance is defined as $d_{e,ij} = \sqrt{(\vec{r}_i - \vec{r}_j)^2} - (R_{g,i} + R_{g,j})$, $\vec{r}_i$ is the position vector of a domain i defined as the mean of its N-termini N atom and C-termini O atom coordinates, calculated using g_traj utility of Gromacs. Two residues are in contact if the distance between C_α-C_α atoms between the two residues is 0.8 nm or less. Statistical uncertainties are calculated as total standard deviation of a given contact formation divided by the square root of the product of total number of replicas forming the given contact and average number of roundtrips per replica, 17.

Acknowledgments

Supporting Information Available

Figures showing comparison of various force field and NMR chemical shifts; comparison of M66 sequence chemical shifts with NMR chemical shifts; Ensemble averaged interchain distance profiles for V66 sequence for each domain; Effect of perturbing monomer properties on freely-jointed, self-avoiding heteropolymer; Residue level contacts for the entire prodomain; Effects of temperature and Val66Met mutation on helix propensity around residue 66; Number of round trip completed by each replica; a table of fitted Flory exponents for each domain.

References

(1) Uversky, V. N. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 932–51.

(2) Panchenko, A. R.; Babu, M. M. 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.

(3) Ward, J.; Sodhi, J.; McGuffin, L.; Buxton, B.; Jones, D. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **2004**, *337*, 635–645.

(4) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.

(5) Dunker, A. K.; Cortese, M. S.; Romero, P.; Iakoucheva, L. M.; Uversky, V. N. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **2005**, *272*, 5129–5148.

(6) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing Protein Intrinsic Disorder. *Chem. Rev.* **2014**, *114*, 6561–6588.

(7) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–40.

(8) 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. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **2007**, *35*, D786–D793.

(9) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112.

(10) Das, R. K.; Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392–7.

(11) Vacic, V.; Markwick, P. R. L.; Oldfield, C. J.; Zhao, X.; Haynes, C.; Uversky, V. N.; Iakoucheva, L. M. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput. Biol.* **2012**, *8*, e1002709.

(12) Larini, L.; Gessel, M. M.; LaPointe, N. E.; Do, T. D.; Bowers, M. T.; Feinstein, S. C.; Shea, J.-E. Initiation of assembly of tau(273-284) and its ΔK280 mutant: an experimental and computational study. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8916.

(13) Ganguly, D. et al. Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. *PLOS Comput. Biol.* **2015**, *11*, e1004247.

(14) Viet, M. H.; Nguyen, P. H.; Derreumaux, P.; Li, M. S. Effect of the English Familial Disease Mutation (H6R) on the Monomers and Dimers of Aβ40 and Aβ42. *ACS Chem. Neurosci.* **2014**, *5*, 646–657.

(15) Viet, M. H.; Nguyen, P. H.; Ngo, S. T.; Li, M. S.; 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*, 1446–1457.

(16) Truong, P. M.; Viet, M. H.; Nguyen, P. H.; Hu, C.-K.; Li, M. S. Effect of Taiwan Mutation (D7H) on Structures of Amyloid-β Peptides: Replica Exchange Molecular Dynamics Study. *J. Phys. Chem. B* **2014**, *118*, 8972–8981.

(17) Zhan, Y. A.; Wu, H.; Powell, A. T.; Daughdrill, G. W.; Ytreberg, F. M. 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*, 1738–1747.

(18) 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*, 6206–6216.

(19) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *J. Biol. Chem.* **2016**, *291*, 6696–6705.

(20) He, Y.; Chen, Y.; Mooney, S. M.; Rajagopalan, K.; Bhargava, A.; Sacho, E.; Weninger, K.; Bryan, P. N.; Kulkarni, P.; Orban, J. Phosphorylation-induced Conformational Ensemble Switching in an Intrinsically Disordered Cancer/Testis Antigen. *J. Biol. Chem.* **2015**, *290*, 25090–25102.

(21) 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. 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.

(22) Iešmantavičius, V.; Jensen, M. R.; Ozenne, V.; Blackledge, M.; Poulsen, F. M.; Kjaergaard, M. Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered Protein Reveals Long-Range HelixHelix Interactions. *J. Am. Chem. Soc.* **2013**, *135*, 10155–10163.

(23) Mittag, T.; Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **2007**, *17*, 3–14.

(24) 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.

(25) Ithuralde, R. E.; Roitberg, A. E.; Turjanski, A. G. Structured and Unstructured Binding of an Intrinsically Disordered Protein as Revealed by Atomistic Simulations. *J. Am. Chem. Soc.* **2016**, *138*, 8742–8751.

(26) Knott, M.; Best, R. B.; 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*, e1002605.

(27) Invernizzi, G.; Lambrughi, M.; Regonesi, M. E.; Tortora, P.; Papaleo, E. The conformational ensemble of the disordered and aggregation-protective 182-291 region of ataxin-3. *Biochim. Biophys. Acta* **2013**, *1830*, 5236–47.

(28) Abeln, S.; Frenkel, D. Disordered flanks prevent peptide aggregation. *PLoS Comput. Biol.* **2008**, *4*, e1000241.

(29) Yedvabny, E.; Nerenberg, P. S.; So, C.; Head-Gordon, T. Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants. *J. Phys. Chem. B* **2015**, *119*, 896–905.

(30) Levine, Z. A.; Shea, J.-E. Simulations of disordered proteins and systems with conformational heterogeneity. *Curr. Opin. Struct. Biol.* **2017**, *43*, 95–103.

(31) Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J.-E. Regulation and aggregation of intrinsically disordered peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 2758–63.

(32) Pappu, R. V.; Wang, X.; Vitalis, A.; Crick, S. L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys.* **2008**, *469*, 132–41.

(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*, 916–930.

(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*, 8856–8860.

(35) Davies, A. M. Regulation of neuronal survival and death by extracellular signals during development. *EMBO J.* **2003**, *22*, 2537–45.

(36) Huang, Z.; Kirkwood, A.; Pizzorusso, T.; Porciatti, V.; Morales, B.; Bear, M. F.; Maffei, L.; Tonegawa, S. BDNF Regulates the Maturation of Inhibition and the Critical Period of Plasticity in Mouse Visual Cortex. *Cell* **1999**, *98*, 739–755.

(37) Liu, B.-h.; Li, Y.-t.; Ma, W.-p.; Pan, C.-j.; Zhang, L.; Tao, H. Broad Inhibition

Sharpens Orientation Selectivity by Expanding Input Dynamic Range in Mouse Simple Cells. *Neuron* **2011**, *71*, 542–554.

(38) Gao, M.; Maynard, K. R.; Chokshi, V.; Song, L.; Jacobs, C.; Wang, H.; Tran, T.; Martinowich, K.; Lee, H.-K. Rebound potentiation of inhibition in juvenile visual cortex requires vision-induced BDNF expression. *J. Neurosci.* **2014**, *34*, 10770–9.

(39) Autry, A. E.; Monteggia, L. M. Brain-Derived Neurotrophic Factor and Neuropsychiatric Disorders. *Pharmacol. Rev.* **2012**, *64*, 238–258.

(40) Björkholm, C.; Monteggia, L. M. BDNF a key transducer of antidepressant effects. *Neuropharmacology* **2016**, *102*, 72–79.

(41) Autry, A. E.; Adachi, M.; Nosyreva, E.; Na, E. S.; Los, M. F.; Cheng, P.-f.; Kavalali, E. T.; Monteggia, L. M. NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses. *Nature* **2011**, *475*, 91–95.

(42) Soliman, F. et al. A genetic variant BDNF polymorphism alters extinction learning in both mouse and human. *Science* **2010**, *327*, 863–6.

(43) Chen, Z.-Y.; 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.

(44) Verhagen, M.; van der Meij, A.; van Deurzen, P. A. M.; Janzing, J. G. E.; Arias-Vásquez, A.; Buitelaar, J. K.; Franke, B. Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity. *Mol. Psychiatry* **2010**, *15*, 260–71.

(45) Kyte, J.; Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–32.

(46) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–27.

(47) Feng, D.; Kim, T.; Özkan, E.; Light, M.; Torkin, R.; Teng, K. K.; Hempstead, B. L.; Garcia, K. C. Molecular and Structural Insight into proNGF Engagement of p75NTR and Sortilin. *J. Mol. Biol.* **2010**, *396*, 967–984.

(48) Anastasia, A.; Deinhardt, K.; Chao, M. V.; Will, N. E.; Irmady, K.; Lee, F. S.; Hempstead, B. L.; Bracken, C. Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. *Nat. Commun.* **2013**, *4*, 2490.

(49) Giza, J. I.; Kim, J.; Meyer, H. C.; Anastasia, A.; Dincheva, I.; Zheng, C. I.; Lopez, K.; Bains, H.; Yang, J.; Bracken, C.; Liston, C.; Jing, D.; Hempstead, B. L.; Lee, F. S. The BDNF Val66Met Prodomain Disassembles Dendritic Spines Altering Fear Extinction Circuitry and Behavior. *Neuron* **2018**, *99*, 163–178.e6.

(50) Holehouse, A. S.; Das, R. K.; Ahad, J. N.; Richardson, M. O.; Pappu, R. V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys. J.* **2017**, *112*, 16–21.

(51) Brzovic, P.; Heikaus, C.; Kisselev, L.; Vernon, R.; Herbig, E.; Pacheco, D.; Warfield, L.; Littlefield, P.; Baker, D.; Klevit, R.; Hahn, S. The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15Using a Simple Protein Interface Forming a Fuzzy Complex. *Mol. Cell* **2011**, *44*, 942–953.

(52) Uesugi, M.; Nyanguile, O.; Lu, H.; Levine, A. J.; Verdine, G. L. Induced α Helix in the VP16 Activation Domain upon Binding to a Human TAF. *Science (80-.).* **1997**, *277*, 1310–1313.

(53) Radhakrishnan, I.; Pérez-Alvarado, G. C.; Parker, D.; Dyson, H.; Montminy, M. R.; Wright, P. E. Solution Structure of the KIX Domain of CBP Bound to the Transacti-

vation Domain of CREB: A Model for Activator:Coactivator Interactions. *Cell* **1997**, *91*, 741–752.

(54) Canales, Á.; Rösinger, M.; Sastre, J.; Felli, I. C.; Jiménez-Barbero, J.; Giménez-Gallego, G.; Fernández-Tornero, C. Hidden α-helical propensity segments within disordered regions of the transcriptional activator CHOP. *PLoS One* **2017**, *12*, e0189171.

(55) Staller, M. V.; Holehouse, A. S.; Swain-Lenz, D.; Das, R. K.; Pappu, R. V.; Cohen, B. A. A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. *Cell Syst.* **2018**, *6*, 444–455.e6.

(56) Das, R. K.; Ruff, K. M.; Pappu, R. V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112.

(57) Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Gräter, F. KirkwoodBuff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *J. Phys. Chem. B* **2015**, *119*, 7975–7984.

(58) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

(59) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the HelixCoil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

(60) Best, R. B.; Zheng, W.; Mittal, J. Balanced ProteinWater Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

(61) Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

(62) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–8.

(63) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.

(64) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci.* **2018**, *115*, E4758–E4766.

(65) 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*, 13–22.

(66) Ortega, A.; Amorós, D.; García de la Torre, J. Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys. J.* **2011**, *101*, 892–8.

(67) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmü, H.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.

(68) Meng, F.; Bellaiche, M. M. J.; Kim, J.-Y.; Ul, G.; Zerze, H.; Best, R. B.; Chung, H. S. Highly Disordered Amyloid-β Monomer Probed by Single-Molecule FRET and MD Simulation. **2018**,

(69) Flory, P. J. The Configuration of Real Polymer Chains. *J. Chem. Phys.* **1949**, *17*, 303–310.

(70) Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci.* **2012**, *109*, 16155–16160.

(71) Zerze, G. H.; Best, R. B.; Mittal, J. Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. *J. Phys. Chem. B* **2015**, *119*, 14622–14630.

(72) Rubinstein, M.; Colby, R. H. *Polymer physics*; Oxford University Press, 2003; p 440.

(73) Faure, G.; Bornot, A.; de Brevern, A. G. Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie* **2008**, *90*, 626–639.

(74) Gómez-Tamayo, J. C.; Cordomí, A.; Olivella, M.; Mayol, E.; Fourmy, D.; Pardo, L. Analysis of the interactions of sulfur-containing amino acids in membrane proteins. *Protein Sci.* **2016**, *25*, 1517–24.

(75) Viguera, A. R.; Serrano, L. Side-chain interactions between sulfur-containing amino acids and phenylalanine in alpha-helices. *Biochemistry* **1995**, *34*, 8771–9.

(76) Valley, C. C.; Cembran, A.; Perlmutter, J. D.; Lewis, A. K.; Labello, N. P.; Gao, J.; Sachs, J. N. The methionine-aromatic motif plays a unique role in stabilizing protein structure. *J. Biol. Chem.* **2012**, *287*, 34979–91.

(77) Riback, J. A.; Katanski, C. D.; Kear-Scott, J. L.; Pilipenko, E. V.; Rojek, A. E.; Sosnick, T. R.; Drummond, D. A. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* **2017**, *168*, 1028–1040.e19.

(78) Creamer, T. P.; Rose, G. D. Side-chain entropy opposes a-helix formation but rationalizes experimentally determined helix-forming propensities (a-helix/protein folding/protein engineering). *Nat. Sci.* **1992**, *89*, 5937–5941.

(79) Feuerstein, S.; Solyom, Z.; Aladag, A.; Favier, A.; Schwarten, M.; Hoffmann, S.; Willbold, D.; Brutscher, B. Transient Structure and SH3 Interaction Sites in an Intrinsically Disordered Fragment of the Hepatitis C Virus Protein NS5A. *J. Mol. Biol.* **2012**, *420*, 310–323.

(80) Chen, Z.-Y.; Ieraci, A.; Teng, H.; Dall, H.; Meng, C.-X.; Herrera, D. G.; Nykjaer, A.; Hempstead, B. L.; Lee, F. S. Sortilin controls intracellular sorting of brain-derived neurotrophic factor to the regulated secretory pathway. *J. Neurosci.* **2005**, *25*, 6156–66.

(81) Lim, J. M.; Kim, G.; Levine, R. L. Methionine in Proteins: It's Not Just for Protein Initiation Anymore. *Neurochem. Res.* **2019**, *44*, 247–257.

(82) Leloup, N.; Chataigner, L. M. P.; Janssen, B. J. C. Structural insights into SorCS2-Nerve Growth Factor complex formation. *Nat. Commun.* **2018**, *9*, 2979.

(83) 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*, 7–8.

(84) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738.

(85) Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **2008**, *9*, 40.

(86) Kim, D. E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526–31.

(87) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.

(88) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.

(89) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(90) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(91) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(92) 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*, 213–23.

(93) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **2006**, *15*, 2795–2804.

(94) Schwarzinger, S.; Kroon, G. J.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **2001**, *123*, 2970–8.

(95) Kjaergaard, M.; Poulsen, F. M. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **2011**, *50*, 157–165.

(96) Kjaergaard, M.; Brander, S.; Poulsen, F. M. Random coil chemical shift for intrinsi-

cally disordered proteins: effects of temperature and pH. *J. Biomol. NMR* **2011**, *49*, 139–149.

(97) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; 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*, 17908–17918.

(98) García, A. E.; Sanbonmatsu, K. Y. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2782–7.

(99) Iglesias, J.; Sanchez-Martínez, M.; Crehuet, R. SS-map. *Intrinsically Disord. Proteins* **2013**, *1*, e25323.

(100) Parashar, M.; Brennan-Tonetta, M.; Rodero, I.; Villalobos, J. J. High Performance Computing at the Rutgers Discovery Informatics Institute. **2018**,