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

Ruchi Lohia, Reza Salari, Grace Brannigan

Submitted date: 29/08/2019 • Posted date: 29/08/2019
Licence: CC BY-NC-ND 4.0

Citation information: Lohia, Ruchi; Salari, Reza; Brannigan, Grace (2019): Sequence specificity despite intrinsic disorder: how a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein. ChemRxiv. Preprint.

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) in precursor brain-derived neurotrophic factor (BDNF) is one of the earliest SNPs to be associated with neuropsychiatric disorders, and the underlying molecular mechanism is unknown. Here we report on over 250 μs of fully-atomistic, explicit solvent, temperature replica exchange molecular dynamics (MD) 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 change in local structure is mediated via entropic and sequence specific effects. We developed a hierarchical sequence-based framework for analysis and conceptualization, which first identifies “blobs” of 5-15 residues representing local globular regions or linkers. We use this framework within a novel test for enrichment of higher-order (tertiary) structure in disordered proteins; the size and shape of each blob is extracted from MD simulation of the real protein (RP), and used to parameterize a self-avoiding heterogenous polymer (SAHP). The SAHP version of the BDNF prodomain suggested a protein segmented into three regions, with a central long, highly disordered polyampholyte linker separating two globular regions. This effective segmentation was also observed in full simulations of the RP, but the Val66Met substitution significantly increased interactions across the linker, as well as the number of participating residues. The Val66Met substitution replaces β-bridging between Val66 and Val94 (on either side of the linker) with specific side-chain interactions between Met66 and Met95. The protein backbone in the vicinity of Met95 is then free to form β-bridges with residues 31-41 near the N-terminus, which condenses the protein. A significant role for Met/Met interactions is consistent with previously-observed non-local effects of the Val66Met SNP, as well as established interactions between the Met66 sequence and a Met-rich receptor that initiates neuronal growth cone retraction.
| File Name                        | Size   | View Options     |
|---------------------------------|--------|------------------|
| SI.pdf                          | 24.16 MiB | view on ChemRxiv, download file |
| IDP_Val66Met_chem.pdf           | 8.06 MiB | view on ChemRxiv, download file |
Sequence specificity despite intrinsic disorder: how a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein

Ruchi Lohia, Reza Salari, Grace Brannigan

1 Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA
2 Department of Physics, Rutgers University, Camden, NJ, USA

* grace.brannigan@rutgers.edu(GB)
S1 Fig. Force field comparison. We ran T-REMD simulations of a 30 residue fragment of the V66 prodomain with several commonly used force field and water model combinations. (a) Comparison of ΔδC_α secondary chemical shifts at 280K from MD ensembles for a99sb*-ildn-q [1,2] with Tip4p-D [3], c36m [4], a99sbws [1,5], a03sbws [5,6], a99sb-ildn with Tip3p [7], calculated using SPARTA+ [8] and NMR from Ref. 9. (b) R_g vs the simulation time, using a 100 ns moving window on left and R_g distribution for each force field on right. Tip3p and a03sbws generates most collapsed and expanded R_g distribution respectively. The equilibration time and (R_g) is shown with vertical and horizontal dashed lines for each force field.

Table S1. Summary of force field comparison simulations.

| Force field                  | ΔδC_α  | 〈R_g〉         | equilibration time | no of replica |
|------------------------------|--------|-----------------|--------------------|---------------|
| a99sb*-ildn-q (Tip4p-D)      | 0.355  | 1.270 ± 0.007   | 200 ns             | 36            |
| a99sbws (Tip4p/2005)         | 0.425  | 1.277 ± 0.007   | 200 ns             | 36            |
| c36m (Tip3p)                 | 0.350  | 1.306 ± 0.007   | 200 ns             | 30            |
| a99sb-ildn (Tip3p)           | 0.617  | 0.922 ± 0.003   | 200 ns             | 32            |
S2 Fig. Effects of temperature and Val66Met mutation on helix propensity around residue 66. Frequency of helix of a given length at residue 66 in V66 (top) and M66 (bottom) in the temperature range of 300K to 385 K. With the increase in temperature the color transitions from cooler (blue) to hotter (red). It is entropically unfavorable for V66 and its neighboring residue to be simultaneously in the helical region of the Ramachandran map, as indicated by the decreasing helical propensity with increasing temperature. For longer helices, the trend will depend more on the additional side-chains in the helix, and the trend with temperature is reversed, but it remains weaker than the analogous trend for the M66 sequence. Errors represent standard error of a Bernoulli trial with n number of samples, where n is the product of total number unique replicas forming the helix of given length at residue 66 at a given temperature and average number of roundtrips per replica, 17.

S0.1 Heterogeneous behavior of individual domains

Disordered proteins can be well-described by Flory scaling theory $\langle R_{\mid i-j\mid} \rangle = A |i-j|^\nu$, where $\langle R_{\mid i-j\mid} \rangle$ is the ensemble-averaged internal distance, $|i-j|$ is residue separation along the chain, and $\nu$ is the Flory scaling coefficient [10]. Larger values of $\nu$ correspond to swollen coils, while smaller values correspond to compact globules [11]. In particular, when $\nu=0.6$ (“good solvent”) the protein maximizes its interaction with solvent, and for $\nu=0.33$ (“poor solvent”), the protein maximizes self-interactions. The special intermediate case of $\nu=0.5$ is called a “theta solvent” [10]. Most IDPs that obey this scaling behavior have $\nu>0.5$ [11,14].

As shown in Fig S3 Fig 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. Each identified individual domain does obey a power law, and we calculated $A$ and $\nu$ for each domain as if it was isolated from rest of the protein (Fig S3 Fig). The highest observed value of $\nu$ was in h2b and h3c domain. This is in agreement with strong polyelectrolyte nature of h2b and high content of Proline residue (20%) in h3c.

**Method** We calculated the average distance between the first atom (N) and last atom (O) for all residue pairs of a given sequence as a function of sequence separation $|i-j|$ using g_traj. Errors before fitting were calculated as the standard error in the mean, where $n = 1088$ is the product of total number of replicas simulated (64) and average number of roundtrips per replica (17). $\nu$ was calculated by linear fit of...
Each of the prodomain were fitted prefactor of 0.59 and residue separation of 3 or more residues. V66 M66 h1a h1b h2a h2b h3a h3b h3c h3d.

S3 Fig. Scaling behavior of each identified domain. Ensemble averaged interchain distance profiles for the entire V66 and M66 prodomain and each blob in the sequence. Theoretical polymer scaling limits are shown with grey lines (prefactor $A = 0.59$ nm) (top). Flory exponents for each blob (bottom).

$\ln(\langle R_{ij} \rangle)$ vs $\ln(|i - j|)$ weighted by each point’s pre fit error with fixed $A$ of 0.59 nm. To exclude the short-range backbone rigidity, distances with $|i - j| < 3$ were not fit.
S4 Fig. Effect of perturbing monomer properties on freely-jointed, self-avoiding heteropolymer Contact probability maps from MC simulations, analogous to those in Figure 5a of the main text, in which the blob $p_3$ is swapped with every other blob in the chain, with the new location represented by the purple square in the graph annotation. As the $p_3$ blob is shifted along the chain, $p_3$ and $p_1$ consistently bound a white “forbidden” region that has little interaction with the rest of the protein.
**S5 Fig.** β-pairing of each blob pair β propensities at each residue in V66 sequence (top) and M66 sequence (bottom) for four clusters. Frames were first clustered by whether the X-Y contact was formed (purple) or broken (green), and then by whether β structure was present in X (solid) or absent (dashed). X represents p1 and is annotated at the top panel and Y represents other blobs identified in the sequence and is annotated on the left for each panel. 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).
S6 Fig. β-pairing of each blob pair. Same as Fig S5, where X represents h1a (left) or h1b (right).
**S7 Fig.** \(\beta\)-pairing of each blob pair. Same as Fig S5, where X represents h2a (left) or h2b(right).
**S8 Fig.** β-pairing of each blob pair. Same as Fig S5, where X represents h3a (left) or h3b (right).
**S9 Fig.** β-pairing of each blob pair. Same as Fig S5, where X represents h3c (left) or h3d (right).
S10 Fig. β-pairing of each blob pair. Same as Fig S5, where X represents p2 (top) or p3 (bottom).
S11 Fig. Residue level contacts for the entire prodomain. Contact probability between every residue pair for V66 (left) and M66 (middle) and M66-V66(right). Two residue pairs are in contact if the distance between Cα-Cα atoms between the two residues are 0.8nm or less. 

b) A linear network of transient tertiary contacts shown in a). The contact networks were build using Cytoscape [15] with a linear representation of residues. Each protein residue comprises a node in the network, with interactions between residues represented as edges. The strength of individual interactions can be interpreted by the thickness of the edge line on the network diagram. If the separation between residues forming the contact is more than 3, its edge is drawn above the node; otherwise, the edge is drawn at the bottom of the node. To focus on significant interactions, interactions showing more than 4% persistence were considered in network visualization.
S12 Fig. Residue level contacts for the entire prodomain. Contact probability between every residue pair for V66 (left) and M66 (middle) and M66-V66(right). Two residue pairs are in contact if the distance between backbone-backbone atoms between the two residues are 0.4nm or less (1st row), if the distance between non hydrogen sidechain-siechain atoms between the two residues are 0.4nm or less (2nd row), if the distance between non hydrogen sidechain-siechain atoms between the two hydrophobic residues are 0.4nm or less (3rd row), if the two residue pairs are forming a salt bridge with the distance between the donor and acceptor atoms < 0.32nm (4th row).
References

1. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950–8. doi:10.1002/prot.22711.

2. 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(3):712–725. doi:10.1002/prot.21123.

3. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015;119(16):5113–5123. doi:10.1021/jp508971m.

4. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71–73. doi:10.1038/nmeth.4067.

5. Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014;10(11):5113–5124. doi:10.1021/ct500569b.

6. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. J Phys Chem B. 2009;113(26):9004–9015. doi:10.1021/jp901540t.

7. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc. 1981;103(2):335–340. doi:10.1021/ja00392a016.

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

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

10. Flory PJ. The Configuration of Real Polymer Chains. J Chem Phys. 1949;17(3):303–310. doi:10.1063/1.1747243.

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

12. 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(40):16155–16160. doi:10.1073/pnas.1207719109.

13. Zerze GH, Best RB, Mittal J. Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations. J Phys Chem B. 2015;119(46):14622–14630. doi:10.1021/acs.jpcb.5b08619.
14. Meng F, Bellaiche MMJ, Kim JY, Zerze GH, Best RB, Chung HS. Highly Disordered Amyloid-β Monomer Probed by Single-Molecule FRET and MD Simulation. Biophys J. 2018;114(4):870–884. doi:10.1016/j.bpj.2017.12.025.

15. Ahlstrom LS, Baker JL, Ehrlich K, Campbell ZT, Patel S, Vorontsov II, et al. Network visualization of conformational sampling during molecular dynamics simulation. J Mol Graph Model. 2013;46:140–9. doi:10.1016/j.jmgm.2013.10.003.
Sequence specificity despite intrinsic disorder: how a disease-associated Val/Met polymorphism rearranges tertiary interactions in a long disordered protein

Ruchi Lohia¹, Reza Salari¹, Grace Brannigan¹,²

1 Center for Computational and Integrative Biology, Rutgers University, Camden, NJ, USA
2 Department of Physics, Rutgers University, Camden, NJ, USA

* grace.brannigan@rutgers.edu(GB)

Abstract

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) in precursor brain-derived neurotrophic factor (BDNF) is one of the earliest SNPs to be associated with neuropsychiatric disorders, and the underlying molecular mechanism is unknown. Here we report on over 250 µs of fully-atomistic, explicit solvent, temperature replica exchange molecular dynamics (MD) 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 change in local structure is mediated via entropic and sequence specific effects. We developed a hierarchical sequence-based framework for analysis and conceptualization, which first identifies “blobs” of 5-15 residues representing local globular regions or linkers. We use this framework within a novel test for enrichment of higher-order (tertiary) structure in disordered proteins: the size and shape of each blob is extracted from MD simulation of the real protein (RP), and used to parameterize a self-avoiding heterogenous polymer (SAHP). The SAHP version of the BDNF prodomain suggested a protein segmented into three regions, with a central long, highly disordered polyampholyte linker separating two globular regions. This effective segmentation was also observed in full simulations of the RP, but the Val66Met substitution significantly increased interactions across the linker, as well as the number of participating residues. The Val66Met substitution replaces β-bridging between Val66 and Val94 (on either side of the linker) with specific side-chain interactions between Met66 and Met95. The protein backbone in the vicinity of Met95 is then free to form β-bridges with residues 31-41 near the N-terminus, which condenses the protein. A significant role for Met/Met interactions is consistent with previously-observed non-local effects of the Val66Met SNP, as well as established interactions between the Met66 sequence and a Met-rich receptor that initiates neuronal growth cone retraction.

Author summary

Intrinsically disordered proteins are proteins that have no well-defined structure in at least one functional form. Mutations in one amino acid may still affect their function.
significantly, especially in subtle ways with cumulative adverse effects on health. Here we report on molecular dynamics simulations of a protein that is critical for neuronal health throughout adulthood (Brain-derived Neurotrophic Factor). We investigate the effects of a mutation carried by 30% of human population, which has been widely studied for its association with aging-related and stress-related disorders, reduced volume of the hippocampus, and variations in episodic memory. We identify a molecular mechanism in which the mutation may change the global conformations of the protein and its ability to bind to receptors.

Introduction

The physiological significance of intrinsically disordered proteins (IDPs), which can explore a wide range of conformational ensembles in their functional form, is now well-established [1–5]. 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 [6] and cell signaling [7–9]. Long intrinsically disordered regions are particularly abundant among cancer [10] and neurodegenerative-associated proteins [11,12].

IDP amino acid sequences tend to be low-complexity [13,14] and include numerous charged residues, often in long repeats [1,15]. In contrast to ordered proteins, in which a complex sequence encodes a well-defined tertiary structure, an IDP sequence determines a heterogeneous conformational ensemble [16–18]. More than 35% of IDPs reported in DISPROT [19] 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 1c) [20–23]. 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 [24]. More than 25% of disease-associated missense single nucleotide polymorphisms (SNPs) are found in IDPs [25]. Although detectable, the relatively subtle functional effects of these SNPs 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 [26–32] have demonstrated clear effects of single charged-residue insertion, deletion, or substitutions on conformational ensemble and aggregation of IDPs monomers. Simple electrostatic models predict that modifications of residue charge will directly affect ensemble properties [20,26,33,34]. Locally, such mutations can modulate residual secondary structure preferences via forming or breaking local salt-bridges or by introducing helix breaking residues [27,31,35].

For IDPs with a relatively low fraction of charged residues, typical of the Janus region of the state diagram proposed by Das and Pappu [20,21] (Fig 1), more subtle differences among neutral amino acids play an increasingly important role in determining the ensemble. More than 40% of disease-associated IDP polymorphisms annotated in the human UniProtKB/Swiss-Prot database [30] are substitutions between two charge-neutral residues. 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 [27,37,38]; 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 [27].

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. [11,39] and references therein). Molecular dynamics (MD) simulations have played a significant role in understanding IDP structure and dynamics [40–45], 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 [46,47] 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 [40].

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 precursor Brain-derived Neurotrophic Factor (proBDNF), a signaling protein that retains a critical role in neurogenesis and synaptogenesis throughout adulthood [48,49] (Fig 1a). It has been implicated in maintenance of the hippocampus [50,51], orientation selectivity in the visual system [52–54] and the mechanism underlying action of numerous antidepressants [55,56], including rapidly acting low-dose ketamine [57]. 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 [57–61].

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 [62]. The prodomain sequence falls in the Janus sequence region in the phase diagram proposed by Das and Pappu [20,21].

It was subsequently revealed that the cleaved prodomains (∼90 residues) are found in monomeric states in vivo, and the M66 (but not V66) form binds to SorCS2 (sortilin-related VPS10p domain containing receptor 2), leading to axonal growth cone retraction [63] and eliminated synapses in hippocampal neurons [64]. NMR measurements on the prodomain confirmed significant intrinsic disorder for both forms, with differential secondary structure preference around residue 66 [63]. 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 [63].

In this work, we aimed to provide insight into the following questions: (1) What interactions drive the secondary structure change local to residue 66 observed through NMR? (2) How can we meaningfully detect tertiary interactions in a long disordered
protein? (3) Do effects on tertiary interactions explain the non-local secondary structure changes previously observed through NMR? (4) How and why does the Val66Met mutation change tertiary interactions, especially as a charge-neutral mutation? To achieve these aims, we conducted unbiased fully-atomistic replica-exchange MD simulations of the 90 residue BDNF prodomain in explicit solvent, for V66 and M66 forms.

We begin by identifying globular regions, or blobs within the protein using a sequence-based approach based on residue hydrophobicity; this is useful for both conceptualizing the long disordered protein in the absence of a well-defined topology, as well as focusing the analysis. We then compare our simulation results with previous NMR results of Anastasia et al [63] and discuss the effects of the Val66Met SNP on residual secondary structure. We propose and apply an approach for decoupling short-range structural correlations from long-range structural correlations, by comparison with a simplified polymer model parameterized from the MD trajectories. We then discuss the effect of the Val66Met SNP on the network of correlated β strands between distant residues, illustrating how effects of the mutation propagate to tertiary contacts in which the mutation is not involved. Finally, we identify individual residue sidechains that drive the observed effects on this network. Our results suggest an important and previously-unconsidered role for specific Met-Met interactions in transducing the effects of the BDNF Val66Met SNP, and confirm the presence of weak but long-range structural correlations in a disordered protein.

Results and discussion

Prodomain Sequence Decomposition

The region of the BDNF prodomain studied using NMR [63], and simulated here, is 91 residues long. Conceptualization of long structured proteins relies heavily on the consecutive secondary structure elements that form the protein’s topology, allowing for a coarse cartoon-style representation. No such approach for constructing an IDP topology has been available. Our original motivation for identifying globular segments in the sequence was to improve statistical power in analyzing contacts, but we found the resulting topological description to be broadly useful for interpretation of results. We thus present this conceptual tool upfront for clarity.

To avoid ambiguity, we restrict use of the term “domain” to refer to the two major BDNF domains (mature domain and prodomain), and instead specify three levels of hierarchy below the domain level: The prodomain contains multiple “regions”, regions contain “groups”, and groups contain “blobs”. Blobs and groups were identified by sequence alone, as described in Methods, while regions were identified by Monte Carlo simulation of a simplified polymer representing the blobs.

The sequence-analysis approach outlined in Methods divides the sequence into alternating groups, classified as either hydrophobic (h groups) or non-hydrophobic (p groups). The prodomain is composed of six such groups, notated as p1-h1-p2-h2-p3-h3 from N-terminus to C-terminus. The h groups are further divided into blobs (Fig 1b), indexed with a letter. Each hydrophobic group contains two to four blobs: h1 contains h1a and h1b, h2 contains h2a and h2b, and h3 contains h3a, h3b, h3c, and h3d. We denote multiple consecutive blobs within a group by multiple letters: h3ab indicates the stretch of residues between the beginning of blob h3a and the end of blob of h3b. Each p group consists of just one blob. The results in Section 2 led us to further designate Region I (containing p1 through h2), Region II (comprised of p3) and Region III (comprised of h3).

Since each blob sequence has its own properties (Table I), this process also
suggested a new, more tractable conceptualization of the long, disordered BDNF prodomain. Each blob can be analyzed individually according to Das and Pappu metrics \[21\] (Fig 1) or Uversky metrics \[66\] (Fig 1), while several other sequence properties of each blob are shown in Table 1. The Das and Pappu phase diagram predicts the compactness of IDPs based on their fraction of positively (f+) and

---

**Fig 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 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 blob. Eight hydrophobic “h” blobs (dark grey) are identified along with 3 “p” blobs of low hydrophobicity (light grey). c) The diagram of IDP states proposed by Das and Pappu \[21\], based on fraction of positive \( (f^+) \) and negative \( (f^-) \) charged residues, and annotated by the location of the whole BDNF prodomain sequence and each blob identified in b. d) Location of blobs on an Uversky diagram \[66\] of IDPs and globular proteins, as a function of absolute net charge per residue \( (|\text{NCPR}|) \) and \( \langle H \rangle \), with the boundary line between folded and disordered proteins given by the equation in the legend. e) Blobs 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 \[21\] diagram in c) (ii). The blob h2b contains the Val66Met mutation and is marked with star. Additional properties of the blob sequences can be found in Table 1.
Table 1. Sequence based properties of hydrophobic and linker blobs identified in the BDNF prodomain, as shown in Fig 1.

| Region | Group | Blob | N\(^a\) | NCPR\(^b\) | (H)\(^c\) | FCR\(^d\) | f^e- | f^+\(^f\) | κ\(^g\) | Sequence | R\(^h\) | P\(^i\) |
|--------|-------|------|--------|-----------|---------|---------|-------|--------|--------|----------|--------|--------|
| I      | p1    | p1   | 8      | 0.00     | 0.37    | 0.25    | 0.13  | 0.13   | 0.8    | EANIRGGQG | 2      | 0.00   |
|        | h1    | 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    | p2   | 7      | 0.29    | 0.34    | 0.29    | 0.00  | 0.29   | 1.0    | GPKAGSR   | 2      | 0.14   |
| II     | p3    | p3   | 15     | -0.13   | 0.21    | 0.53    | 0.33  | 0.20   | 0.1    | EDQKVRP   | 3      | 0.06   |
| III    | h3    | 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    | 0.20  | 0.20   | 0.20   | RVMLS     | 1      | 0.00   |
|        |      | h3c  | 5      | -0.20   | 0.49    | 0.20    | 0.20  | 0.20   | 0.20   | QVPLE     | 1      | 0.20   |
|        |      | h3d  | 7      | -0.14   | 0.70    | 0.14    | 0.14  | 0.14   | 1.0    | PLLFLE    | 1      | 0.14   |

\(^a\) Number of residues in the blob  
\(^b\) Net charge per residue  
\(^c\) Mean hydrophobicity, average of Kyte-Dolittle scores for each residue in the blob 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 \(\kappa\) as defined by Das and Pappu, calculated using CIDER  
\(^h\) Region in phase diagram proposed by Das and Pappu, (Fig 1c)  
\(^i\) Fraction of Proline residues

negatively (f-) charged residues (Fig 1c). Hydrophobic blobs h2b and blob h3a lie in the strong polyelectrolyte and Janus sequence region respectively. All the remaining hydrophobic blobs are classified as weak polyampholytes and, as isolated peptides, would be predicted to have compact globule conformations to shield hydrophobic residues [21]. Linker blobs p1 and p2 also lie in the Janus sequence regions, while blob p3 lies in the strong polyanalyte region with low value of the charge distribution parameter \(\kappa\) [21], and is predicted to have random coil conformations if present as an isolated peptide.

The Uversky diagram characterizes proteins as globular or intrinsically disordered based on their normalized mean hydrophobicity and absolute net charge per residue (|NCPR|) (Fig 1d). 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 blobs h2b and h3a, all hydrophobic blobs 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 blob h2b containing Val66Met has several unique properties among the identified blobs: 1) it is located at the sequence midpoint 2) it is the only strong polyelectrolyte blob 3) it has the strongest NCPR (-0.38) among all the blobs 4) The sequence is composed almost entirely of two competing residue types, yielding the uncommon mix of a highly-charged, hydrophobic blob. Considering mean hydrophobicity alone, Uversky et al. 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
More specifically, this blob 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 [68–72]. Staller et al. [72] 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 blob h3a is a unique hydrophobic Janus blob with high NCPR. Janus sequences have intermediate compositional biases and their conformations are context dependent [21,73]. The SNP blob h2b and the Janus blob h3a are separated by the long (15 residue) strong polyampholyte linker p3, which has well mixed charge ($\kappa = 0.1$). The blobs h1a and h3b are positively charged and all the remaining hydrophobic blobs are negatively charged (Fig 1e).

Comparison of experimental observables and their computational analogues

NMR spectroscopy [63] has previously confirmed the intrinsic disorder of the prodomain. 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 [74,75]. Piana et al. [75] 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 a03sbws [76,77] with Tip4p/2005 [78], a99sbws [77,79] with Tip4p/2005 [78], a99sb*-ildn-q [79,80] with Tip4p-D [75] and c36m [81] with Tip3p [82] on 30 residue fragments of the V66 prodomain using temperature replica exchange molecular dynamics (T-REMD), further described in SI. To minimize the effects of loss of long-range contacts in the 30 residue fragment, only $\Delta \delta C_\alpha$ secondary chemical shifts were compared; $\Delta \delta C_\beta$ is more dependent on $\beta$-pairing within the sequence. Among all the force fields tested only a03sbws with Tip4p/2005 and a99sb-ildn with Tip3p gave significant deviations with NMR. The three remaining force fields gave good comparison ($\Delta \delta C_\alpha$ RMSD <0.5 ppm) [S1 Fig, S1 Table]. This is also consistent with the force field comparison study by Robustelli et al. [83], which observed that for IDPs with little or no secondary structure, both c36m and a99sb*-ildn-q with Tip4p-D yielded the best agreement with experimental NMR measurements.

The a99sb*-ildn-q/Tip4p-D forcefield was used for the full prodomain MD simulations further described in Methods. Fig 2 shows the $\Delta \delta C_\alpha$ and $\Delta \delta C_\beta$ secondary chemical shifts calculated from the full-length simulations using SPARTA+ [84] (further described in methods) and compares them with the NMR secondary chemical shifts obtained from Anastasia et al. [63] for V66 sequence and M66 sequence. We obtain good agreement with NMR secondary chemical shifts: the discrepancy at each residue is <0.7 ppm, which is less than the individual SPARTA+ prediction uncertainties of ~1 ppm [84].

Comparison of the simulated hydrodynamic radius ($R_h$) generated from MD and NMR/SAXS is an additional useful validation measure. $R_h$ was calculated from the trajectory using Hydropro [87] (further described in methods). Hydrodynamic
Fig 2. Comparison of MD and NMR observables. a) $\Delta \delta C_{\alpha}$ (top), $\Delta \delta C_{\beta}$ (middle), $\Delta \delta C_{\alpha}-\Delta \delta C_{\beta}$ (bottom) secondary chemical shifts from NMR at 280K (black lines) [63] and MD at 300K for V66 (a) and M66 (b). The gray region represents a discrepancy of more than 1 ppm from NMR secondary chemical shifts. Root-mean-squared deviation (RMSD) represents the deviation between the NMR and MD values. Error at each residue is calculated as the standard error in the mean, where $n = 1088$ is the product of total number of replicas simulated and average number of roundtrips per replica.

Effects of Val66Met on local and non-local secondary structure

Anastasia et al. [63] reported an increase in helical tendency for the M66 sequence within blob h2 and h3ab and an increase in $\beta$ tendency within blob h3b in the V66 sequence (Fig 3a). Consistent with these NMR experiments [63], the M66 sequence demonstrates an increased tendency of forming helices within blob h2 and h3a relative to the same blobs in the V66 sequence (Fig 2b).

Comparing the length of secondary structure formed at each residue (Fig 3c) 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 group h2.

In general, $C^3$-branched amino acids, such as Valine, have more restricted side-chain rotamers in helical conformation when compared with non-$C^3$-branched amino acids. Creamer et. al. ranked the entropic cost of helix formation for apolar side chains using simulations of an (Ala)$_8$ sequence with the guest amino acid at the center, and reported a higher entropic cost of helix formation for Valine when compared with Methionine [88]. In our simulations, the likelihood that V66 will be in a short helix decreases with...
Fig 3. Effects of Val66Met on contacts at residue level. a) Differential secondary chemical shifts for the V66 sequence and M66 sequence from NMR [63]. Values on top are equivalent to the two NMR curves shown in Figure 2c, while the difference between the two curves is shown at the bottom. b) Helix (top) or \( \beta \) (bottom) propensity for each simulated residue of the 300K replica, 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 (left) region or \( \beta \) (right) region of the Ramachandran map (further described in methods). Errors represent standard error of a Bernoulli trial with \( n \) samples, where \( n = 1088 \) is the product of the total number replicas (64) and average number of roundtrips per replica (17). c) Difference (M66-V66) between probabilities of secondary structure formation of a given length, for helix (top) and \( \beta \) (bottom). d) Contact probability for each residue pair within the h2 group for each sequence. Each residue in group h2 is annotated with a circle representation and the contacts with at least 50% population are represented with an edge. e) Difference between the contact probabilities shown in d)

temperature, while the opposite effect is observed for M66 [S2 Fig]. These trends are consistent with an increased entropic cost for helix formation at V66 relative to M66.

The helical structure in M66 is also stabilized by local sequence, including the favorable interaction between Met66(i) and Phe63(i-3). MD simulations have previously shown the stability of a sulfur-aromatic contact in a model helix [89]. Fig 3d shows the residue level contact map within group h2. For the M66 sequence, Met66(i) more frequently contacts Phe63 (i-3) than any other residue within the blob: M66-F63 is formed 60% more often than M66-E69 (Fig 3d). We find that the largest change in intrablob contacts 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) (Fig 3e). This is also consistent with previously identified Met-Phe interactions [89–92].

While the effects of the Val66Met mutation on secondary structure in the blob which contains residue 66 (h2b) are not unexpected, we also observed an effect on secondary structure in group h1 and blobs h3a and h3b within group h3. As shown in Fig 3c, the increased frequency of long helices for blob h3a in the Met66 sequence is comparable to the increase in blob h2b. We consider the possible tertiary origins of the non-local effects on \( \beta \) propensity in .
Regions of tertiary enrichment

The potential number of residue-residue contacts in the prodomain is $91 \times 90/2 \sim 4000$, and each contact is formed infrequently. Detecting significant differences for numerous weak signals is statistically prohibitive, even given the long simulations presented here. Dividing the sequence into blobs based on sequence hydrophobicity, as described in Methods, helps address this analysis challenge (Fig 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.

Fig 4. Detection of Tertiary Enrichment. To decouple short-range from long-range structural correlations, this work grouped segments of the protein into blobs using sequence, and then compared contacts between the blobs to those expected for an analogous self-avoiding heteropolymer (SAHP). The SAHP was parameterized by extracting local properties (size and shape) of blobs from the real protein (RP) trajectory.
We expect that even for a freely-jointed, self-avoiding heteropolymer (SAHP), contact probability between monomers would depend on monomer shape and separation, although a SAHP does not have tertiary structure. Inspired by the Kuhn treatment of real polymers, we propose that the expected intermonomer contact frequency in a SAHP can be a useful reference for detecting specific tertiary interactions (Fig 4), as long as the monomers mimic the blobs of the real protein (RP). In support of this approach, we find that within a given blob, the protein obeys Flory polymer scaling laws. The exponent varies across blobs (Section S1, S3 Fig), capturing the intrinsic heterogeneity of the long polymer.

Fig 5. Effect of Val66Met on contacts between blobs. a) Blob-blob contact probability for the V66 self-avoiding heteropolymer (SAHP). The black boxes mark the regions identified. b) Blob-blob contact probability shown in a) for the real protein (RP); V66 (left) and M66 (right) sequences. The x and y axes are annotated with cartoon representations of the prodomain; circles are drawn to the scale of each blob’s size. c) %Population of contacts (top) and enrichment in RP contacts with respect to SAHP (bottom) for each region. The errors represent standard errors (n =1088 as described in Methods). d) Difference between the contact probabilities shown in b). e) Differences shown in d) with respect to SAHP; interactions more frequently found in M66 or V66 are in blue and orange respectively.

The predicted contact probabilities for this freely-jointed, self-avoiding heteropolymer are shown in Fig 5. In the SAHP version of the prodomain, the chain is visibly segmented at the boundary between the h2b and p3 blobs. As shown in S4 Fig, shifting the p3 bead within the SAHP chain shifts the visible segmentation boundary, confirming that the p3 blob defines the segmentation. Based on this expectation, we define three regions: the pre-p3 blobs are “Region I”, p3 is “Region II”, and the post-p3 blobs are “Region III”. SAHP Blobs within Region I are in contact for 61% of the frames, while blobs within Region III are in contact in 81% of the frames. In comparison, the average contact probability between Regions I and III is only 10%(Fig 5).

Fig 5b shows the probability of blob-blob contacts for both the V66 and M66
sequences of the RP, calculated analogously to those in the SAHP. The frequencies of contacts within Region I and within Region III were quantitatively consistent with the SAHP predictions. The total number of blob-blob contacts within Region I was enriched by 1.2 times that expected for the SAHP. Within Region III, the total number was depleted by 0.9 times the expected value (Fig 5c).

In contrast, contacts between blobs on either side of the long p3 linker are more common in the RP than in the SAHP, and are also affected by the substitution at residue 66 (Fig 5c,d,e). Contacts between pre-linker Region I and post-linker Region III are about three times as common in the RP as in the SAHP, indicating specific tertiary interactions beyond those expected for a polymer undergoing a random-walk. Quantitatively, enrichment in the V66 sequence is 2.9±0.1 while enrichment in the M66 sequence is 3.4±0.1. The increased number of cross linker contacts are also consistent with the lower mean hydrodynamic radius (Fig 8a) and radius of gyration (Fig 8a) for the M66 sequence.

Effects of Val66Met on the β-pairing network

Fig 6. Secondary structure coupling between blobs on each side of the p3 linker. β propensities at each residue in V66 sequence (top) and M66 sequence (bottom) for four clusters. Frames were first clustered by whether the h3b-h2b (a) or h3b-h1a (b) contact was formed (purple) or broken (green), and then by whether β structure was present in h2b (solid) (a) or h1a (b) or absent (dashed). The dark-gray window indicates the contacting blob that is constrained to have high or vanishing values by construction of the cluster, while the white window indicates the contacting blob without constrained secondary structure. If the contact is coupled to simultaneous β-strand formation, the peak within the white window for the solid purple curve should be significantly higher than other curves. 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).

To test whether the changes we observed in tertiary contacts at the blob, group, or region level could be due to a change in partnering β-strands, we applied a clustering approach. All frames were divided into 4 clusters, representing two independent collective variables with two possible values each: either a certain contact between blobs X and Y is formed or broken, and any residue in blob X is found within a stretch of 4 sequential residues in β conformation. The four clusters are thus represented as (contacting,absent), (contacting,present), (distant,absent), and (distant,present).

For each cluster, we calculated β propensity across all residues. (Figure 6) If the X-Y contact reflects correlated β-strands, we expect a peak at residues in blob Y in the
(contacting,present) cluster that is significantly higher than the signal for all other clusters. If the secondary structure in Y is used for clustering instead, the reciprocal peak (at blob X) should be reproduced. Furthermore, unless there are higher-order correlations between multiple sets of \( \beta \)-strands, \( \beta \) propensity should not depend on cluster for all residues not in blob X or Y.

This clustering process on all frames was carried out for all possible X and Y blobs, provided X and Y were not in the same group (S5-S10 Fig). For most pairs, there was no correlating peak in \( \beta \) structure. For some pairs, a peak was present in one direction but the reciprocal peak was not present in the opposite direction. This result reflected longer \( \beta \)-strands that extended to a neighboring blob, which had the true peak. Only one symmetrically significant peak (indicating correlated \( \beta \) structure) was observed per sequence. Both involved the h3b blob, but the partner blob shifted from h2b in the V66 sequence to h1a in the M66 sequence (Figure 6).

Despite loss of correlated \( \beta \)-pairing, the contact between h2b and h3b is actually more probable in the M66 sequence than in the V66 sequence (Figure 5I). As discussed in , this result reflects a significant change at the residue level. In the M66 sequence, specific interactions between M66 and side-chains of residues within h3b form the contact, rather than backbone-backbone interactions. As the h3b side-chains stabilize the contact with h2b, the backbone of h3b is then free to pair with h1a, increasing the number of favorable long-range contacts and condensing the M66 sequence overall.

**Noteworthy residue-residue interactions stabilizing tertiary contacts**

As shown in the previous section, the Val66Met substitution causes loss of correlated \( \beta \)-strands between blobs h2b and h3b, while introducing correlated \( \beta \)-strands between blobs h3b and h1a. We consider here the effects of the substitution on these contacts at residue level. As shown in the absolute residue-residue contact probability maps (Fig 7a), both sequences frequently form contacts between hydrophobic residues in blobs h2b and h3b. The residue pairs most frequently forming the contact shift from Val66-Val94 in the V66 sequence to Met66-Met95 in the M66 sequence (Fig 7b). The residue-level contact maps also show a high probability of contacts between D72 and T91 in the Val66 but not Met66 sequence. As illustrated in Figure 7c, these contacts (between \( \alpha \) carbons) are stabilized by salt-bridges between R93 and D73, in a conformation that is incompatible with a side-chain contact between Val/Met66 and Met95.

Met95 is the only other Methionine in the simulated sequence. The role of specific Met-Met interactions due to polarizable sulfur atoms is often under-appreciated, but are common in structures of folded proteins [90]. Using ab initio calculations, Gómez-Tamayo et al [92], predicted that Met-Met interactions are stronger than Met-aromatic or aromatic-aromatic interactions, due to the polarizability of Sulfur. Although the fixed-charge forcefield we are using (a99sb*-ildn-q) cannot explicitly capture polarizability, Gómez-Tamayo et al. demonstrate that this forcefield preserves rankings of strong side-chain interactions involving Methionine. 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) (Fig 7a and b). Methionine-aromatic interactions also contribute to the increased number of Region I-III contacts: M66, but not V66, forms a frequent contact with F108 in blob h3d, which is also consistent with the favorable interactions between Met-Phe residues [89–91] (Fig 7a and b).

To determine which residue contacts between h2b and h3ab couple the secondary structure within the two blobs, we decomposed the residue-level contact maps into nine
Fig 7. Effect of secondary structure in group h2 on which residues form the cross-boundary h2-h3 contact. a) Contact probability at each residue in h2b with each residue in h3 for V66 (left) and M66 (right). b) Difference between the contact probabilities shown in a). c) Representative conformation of M66 sequence and V66 sequence showing preferred residue-level contacts in VDW representations, with residues colored by residue type: blue:basic, red:acidic, cyan:polar, grey:hydrophobic, Met: yellow and chain colored as in Fig 1. Tubes represent hydrophobic “h” blobs whereas lines represent non-hydrophobic linker “p” blobs. d) Contact probability between residues 63-69 and each residue in h3ab, when respective secondary structure is formed at each residues, for both the V66 and M66 sequences. Residue labels are colored according to residue type: blue:basic, red:acidic, grey:hydrophobic/polar and Met: yellow.

clusters. Each cluster was specified by two collective variables with three possible values each: secondary structure (helix, β, or coil) around residue 66 and secondary structure (helix, β, or coil) in h3ab (Figure 7d). The β-pairing at h2b-h3ab is stabilized via a combination of backbone hydrogen bonds between V66 and S92, salt bridge between E64 and R93, and hydrophobic interactions between V66 and V94. The V66-M95 contact was only formed frequently within the (h2b - coil, h3ab - helix) cluster, and since this cluster was a very small part of the overall population, the contact overall was rare as well. (Figure 7d) This cluster was more common in the M66 sequence, and contributes to the non-local increase in helicity around residue 95 (Figure 3b).
Summary and Conclusion

We have carried out over 250 \( \mu s \) 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 secondary 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 \([21]\), can be meaningfully divided into 11 blobs based on sequence hydrophobicity alone. Among 8 hydrophobic blobs, we identified 2 blobs in the disordered region: the strong polyelectrolyte blob h2b (which contains Val66Met), and the Janus blob h3a. These are connected via the highly disordered long linker p3. The groups containing these unique blobs have biological significance as well: The sequence h2-p3-h3 is essential for intracellular trafficking of proBDNF \([94]\).

The sequence decomposition approach suggested a tractable approach for coarse-graining analysis, by reducing the initial number of potential contacts from over 4000 (S11-S12 Fig) to 55 (Fig 5), while increasing the number of observations for each contact. Furthermore, it allowed us to isolate the most sensitive regions of the protein for examination at the residue level. This 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. Our conclusions further 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 its local blob h2b, the Val66Met mutation affects local contact preference due to local sequence effects (preferred Met-Phe contacts) and the reduced entropic cost of helix formation for the Metionine sidechain.

The long, disordered, exposed region II linker segregates the blob-level contact probability map: blobs within Region I or Region III have a high probability of contact, while Region I-III contacts are far less probable. We consistently observed this segregation in both simple self-avoiding hetropolymer simulations with beads mimicking identified blobs, and actual prodomain simulations. Val66Met increases the frequency of Region I-III contacts. We find here that the dominant mechanism involves replacing \( \beta \)-strand coupling between group h2 of Region I and group h3 of Region III with favorable Met/Met side-chain interactions between the same groups. The group h3 backbone is then exposed for interactions with the backbone of group h1, also of Region I. The non-local increase in helicity in group h3 may reflect stabilization of non-\( \beta \) structure by the Met-Met interactions.

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. In general, our study supports previous observations \([92,95]\) that Methionine plays a distinct role from true aliphatic residues in determining protein structure, and highlights the importance of mimicking its unique properties within fixed-charge forcefields.

Anastasia et al. \([63]\) 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 blob h2a and h3b (residues Y90 to V94). The stronger binding at residue M66 could be attributed to either A) Helix propensity at the SNP blob at M66. This SNP blob 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 [96]. 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 [97][99], Rosetta [100] and Modeller [101], 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 2 ps time interval); in total, our study included 64 unique prodomain structures. All structures were cooled to 300K for 1 ns, while prolines were restrained in trans-conformation. Each V66 replica was placed in a dodecahedron water box with 30,500 Tip4p-D [75] 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).

MD Simulations
For the simulations we use the a99sb*-ildn-q force field [79][80] and the GROMACS 5.1.2 simulation package, [102][103] with a time step of 2 fs. Long-range electrostatics are calculated using the particle mesh Ewald (PME) method [104], 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 [105] with an exchange frequency of 1 ps 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 300-385 K, 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 51.2 μs (800 ns × 64) trajectories were discarded for equilibration purposes, determined by plateauing of $R_g$ (Fig 5h). Simulation convergence was monitored using several metrics (Fig 5). Over the course of remaining 76.8 μs (1.2 μs × 64) simulations, for both the V66 and M66 sequence each replica completes a minimum of 5 roundtrips and an average of 17 roundtrips (Fig 5b).

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 the standard error in the mean, where $n = 1088$ is the product of total number of replicas simulated (64) and average number of roundtrips per replica (17).
Blob identification

Mean hydrophobicity ($\langle H \rangle$) at each residue is defined as the average Kyte-Dolittle score with a window size of 3 residues, scaled to fit between 0 and 1. Any stretch of four or more residues with $\langle H \rangle > 0.37$ is classified as a hydrophobic or h blob and any stretch of four or more residues with $\langle H \rangle \leq 0.37$ is classified as a non-hydrophobic linker or “p” blob. Multiple consecutive hydrophobic blobs without a “p” blob separating them are classified as a single group.

Secondary Chemical Shifts

Prior to the present study, Anastasia et al. measured chemical shifts for the BDNF prodomain (residues 21-113) using NMR, and then used backbone NMR secondary chemical shifts to predict secondary structure via TALOS+ and SSP. For comparison with simulation data, we reinterpreted the chemical shifts directly from.
deposited at Biological Magnetic Resonance Bank (BMRB). Secondary chemical shifts are calculated as follows: \( \Delta \delta_{C_\alpha,MD} = (\delta_{C_\alpha,MD} - \delta_{C_\alpha,RC(300K)}) \) for MD and \( \Delta \delta_{C_\alpha,NMR} = (\delta_{C_\alpha,NMR} - \delta_{C_\alpha,RC(280K)}) \), where \( \delta_{C_\alpha,MD} \) are predicted \( \delta_{C_\alpha} \) chemical shifts from MD simulation ensembles using SPARTA+ [84] and \( \delta_{C_\alpha,NMR} \) were obtained from BMRB [63]. Random coil chemical shifts (\( \delta_{C_\alpha,RC} \)) for the 91 residue BDNF prodomain were obtained using POTENCI [108] at pH 7, 0.15 M ion concentration and 280K and 300K for NMR and MD respectively. Error at each residue are calculated as the standard error in the mean, where \( n = 1088 \) is the product of total number of replicas simulated (64) and average number of roundtrips per replica (17).

Hydrodynamic radius calculation

The values for the Hydropro [87] parameters were: atomic level model with shell-method calculation, \( a = 0.29 \) nm, 6 minibead iterations, and \( \sigma = 0.1 \) 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 molecular weight of the peptide was equal to 10044 Da (V66 sequence) or 10076 Da (M66 sequence). The resultant translational diffusion constants were then used for calculating \( R_h \) using the Stokes-Einstein equation. Error is calculated as the standard error in the mean, where \( n = 1088 \) is the product of total number of replicas simulated (64) and average number of roundtrips per replica (17).

Secondary structure calculation

Helix propensity or \( \beta \) 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 \( \beta \) region of the Ramachandran space. The helical region is defined as \(-100^\circ < \phi < -30^\circ \) and \(-120^\circ \leq \psi \leq 50^\circ \) [42,109,110]. The \( \beta \) region is defined as \( \phi < -80^\circ \) and \( 50^\circ < \psi < -120^\circ \). The length of SS (SS-map) [111] were calculated with the above defined helical and \( \beta \) 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).

Blob-level contact maps

The excess distance between any two blobs \( i \) and \( j \) is defined as

\[
d_{e,ij} = |\vec{r}_i - \vec{r}_j| - (R_{g,i} + R_{g,j})
\]

where \( \vec{r}_i \) is the position vector of a blob \( i \) defined as the mean of its N-termini N atom and C-termini O atom coordinates, calculated using g traj utility of Gromacs. Two blobs \( i \) and \( j \) are in contact if the excess distance \( (d_{e,ij}) \) between the two is less than 0.55 nm. At residue level, two residues are in contact if the distance between C\( \alpha \) atoms of the two residues is 0.8 nm or less. Presented statistical uncertainties are the standard error in the mean, with \( n \) is the product of total number of replicas forming the given contact and the 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 blobs identified in Fig 1. The separation between monomers \( i \) and \( i + 1 \) (analogous to the Kuhn length for a
homopolymer [93]) was constrained to be half the end to end distance for each of the analogous blobs:

$$|\vec{r}_{i-1} - \vec{r}_i| = \frac{\langle R_{\text{etoe},i-1} \rangle + \langle R_{\text{etoe},i} \rangle}{2}$$  \hspace{1cm} (2)$$

where $\langle R_{\text{etoe},i} \rangle$ was determined from the coordinates of blob $i$ residues in the MD simulations, shown in Figure 9a.

Two monomers $i$ and $j$ are considered to be overlapping if

$$|\vec{r}_i - \vec{r}_j| = \frac{d_{e,ij}}{(R_g,i) + (R_g,j)} + 1 < a$$  \hspace{1cm} (3)$$

where $\langle R_{g,i} \rangle$ was determined from the coordinates of residues in blob $i$ in the MD simulations (Figure 9a), and $a$ is a constant. In the MD simulations of the real protein, we observed that $\frac{d_{e,ij}}{(R_g,i) + (R_g,j)} \geq -0.7$ for almost all frames (Figure 9b), 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 $|\vec{r}_{i-1} - (\vec{r}_i + \vec{\delta r})| = \frac{(R_{\text{etoe},i-1} + (R_{\text{etoe},i}))/2}$, satisfying Eq 4) the translation $\vec{r}_j \rightarrow \vec{r}_j + \delta_r$ was applied for all $j \geq i$.

Any trial move that caused an overlap according to Eq. 3 was rejected, while all others were accepted. The MC simulation was run for 500,000 steps (50,000 steps per moveable bead); additional steps did not change the outcome in Fig 5a.

Fig 9. Parameterization of self-avoiding heteropolymer. Average end to end distance ($R_{\text{etoe}}$) vs Average radius of gyration ($R_g$) for each blob of V66 sequence. Error at each residue is calculated as the standard error in the mean, where n = 1088 is the product of total number of replicas simulated and average number of roundtrips per replica. The errors were smaller than the circles used for the representation of each blob. b) The distribution of normalized excess distances across all blob-pairs in the V66 RP, where $|i - j| > 1$. c) Relationship between the radius of gyration $R_g,i$, end to end distance $R_{\text{etoe},j}$, and excess distance $d_{ij}$, calculated for each blob or blob pair using a RP trajectory. d) The SAHP is a chain with each monomer representing a blob of the real protein and modeled as a hard sphere. Each monomer $i$ has radius $aR_g,i$ and is separated from monomer $i + 1$ by bond length $(R_{\text{etoe},i} + R_{\text{etoe},i+1})/2$. Bond lengths are constrained and bond angles can rotate freely.
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 and the Rutgers Discovery Informatics Institute, which are supported by Rutgers and the State of New Jersey.

Supporting Information

S1 Fig. Force field comparison.

S1 Table Force field comparison summary.

S2 Fig. Effects of temperature and Val66Met mutation on helix propensity around residue 66.

Section S1 Heterogeneous behavior of individual domains.

S3 Fig. Scaling behavior of each identified blob and entire prodomain.

S4 Fig. Effect of perturbing monomer properties on freely-jointed, self-avoiding heteropolymer

S5-S10 Fig. $\beta$-pairing for each domain pair 4 cluster.

S11-S12 Fig. Residue level contacts for the entire prodomain including backbone-backbone, sidechain-sidechain, salt bridge and hydrophobic contacts.

References

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

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

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

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

5. Uversky VN. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. Front Phys. 2019;7:10. doi:10.3389/fphy.2019.00010.

6. Minezaki Y, Homma K, Kinjo AR, Nishikawa K. Human Transcription Factors Contain a High Fraction of Intrinsically Disordered Regions Essential for Transcriptional Regulation. J Mol Biol. 2006;359(4):1137–1149. doi:10.1016/j.jmb.2006.04.016.
7. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets: The roles of intrinsic disorder in protein interaction networks. FEBS J. 2005;272(20):5129–5148. doi:10.1111/j.1742-4658.2005.04948.x.

8. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16(1):18–29. doi:10.1038/nrm3920.

9. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, et al. Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. J Proteome Res. 2007;6(5):1899–916. doi:10.1021/pr060393m.

10. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. J Mol Biol. 2002;323(3):573–584. doi:10.1016/S0022-2836(02)00969-5.

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

12. Buée L, Bussière T, Buée-Scherrer V, Delacourte A, Hof PR. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. These authors contributed equally to this work. Brain Res Rev. 2000;33(1):95–130. doi:10.1016/S0165-0173(00)00019-9.

13. Weathers EA, Paulaitis ME, Woolf TB, Hoh JH. Insights into protein structure and function from disorder-complexity space. Proteins Struct Funct Bioinforma. 2006;66(1):16–28. doi:10.1002/prot.21055.

14. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins Struct Funct Genet. 2001;42(1):38–48. doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3.

15. Jorda J, Xue B, Uversky VN, Kajava AV. Protein tandem repeats - the more perfect, the less structured. FEBS J. 2010;277(12):2673–2682. doi:10.1111/j.1742-4658.2010.07684.x.

16. Dyson HJ, Wright PE. Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol. 1998;5(7):499–503. doi:10.1038/739.

17. Mukhopadhyay S, Krishnan R, Lenke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proc Natl Acad Sci. 2007;104(8):2649–2654. doi:10.1073/pnas.0611503104.

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

19. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. Nucleic Acids Res. 2007;35(Database issue):D786–93. doi:10.1093/nar/gkl893.

20. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102–112. doi:10.1016/j.sbi.2015.03.008.
21. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc Natl Acad Sci U S A. 2013;110(33):13392–7. doi:10.1073/pnas.1304749110.

22. Sawle L, Ghosh K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. J Chem Phys. 2015;143(8):085101. doi:10.1063/1.4929391.

23. Firman T, Ghosh K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. J Chem Phys. 2018;148(12):123305. doi:10.1063/1.5005821.

24. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept. Annu Rev Biophys. 2008;37(1):215–246. doi:10.1146/annurev.biophys.37.032807.125924.

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

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

27. Ganguly D, Chen J. Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. PLOS Comput Biol. 2015;11(4):e1004247. doi:10.1371/journal.pcbi.1004247.

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

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

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

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

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

33. Bah A, Forman-Kay JD. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. J Biol Chem. 2016;291(13):6696–6705. doi:10.1074/jbc.R115.695056.
34. He Y, Chen Y, Mooney SM, Rajagopalan K, Bhargava A, Sacho E, et al. Phosphorylation-induced Conformational Ensemble Switching in an Intrinsically Disordered Cancer/Testis Antigen. J Biol Chem. 2015;290(41):25090–25102. doi:10.1074/jbc.M115.658583.

35. Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure. 2016;24(9):1537–49. doi:10.1016/j.str.2016.07.007.

36. Yip YL, Famiglietti M, Gos A, Duek PD, David FPA, Gateau A, et al. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. Hum Mutat. 2008;29(3):361–366. doi:10.1002/humu.20671.

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

38. Feuerstein S, Solyom Z, Aladag A, Favier A, Schwarten M, Hoffmann S, et al. Transient Structure and SH3 Interaction Sites in an Intrinsically Disordered Fragment of the Hepatitis C Virus Protein NS5A. J Mol Biol. 2012;420(4-5):310–323. doi:10.1016/j.jmb.2012.04.023.

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

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

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

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

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

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

45. Levine ZA, Shea JE. Simulations of disordered proteins and systems with conformational heterogeneity. Curr Opin Struct Biol. 2017;43:95–103. doi:10.1016/j.sbi.2016.11.006.

46. Levine ZA, Larini L, LaPointe NE, Feinstein SC, Shea JE. Regulation and aggregation of intrinsically disordered peptides. Proc Natl Acad Sci U S A. 2015;112(9):2758–63. doi:10.1073/pnas.1418155112.
47. Pappu RV, Wang X, Vitalis A, Crick SL. A polymer physics perspective on driving forces and mechanisms for protein aggregation. Arch Biochem Biophys. 2008;469(1):132–41. doi:10.1016/j.abb.2007.08.033.

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

49. Davies AM. Regulation of neuronal survival and death by extracellular signals during development. EMBO J. 2003;22(11):2537–45. doi:10.1093/emboj/cdg254.

50. Pezawas L, Verchinski BA, Mattay VS, Callicott JH, Kolachana BS, Straub RE, et al. The brain-derived neurotrophic factor val66met polymorphism and variation in human cortical morphology. J Neurosci. 2004;24(45):10099–102. doi:10.1523/JNEUROSCI.2680-04.2004.

51. Benjamin S, McQuoid DR, Potter GG, Payne ME, MacFall JR, Steffens DC, et al. The Brain-Derived Neurotrophic Factor Val66Met Polymorphism, Hippocampal Volume, and Cognitive Function in Geriatric Depression. Am J Geriatr Psychiatry. 2010;18(4):323–331. doi:10.1097/JGP.0b013e3181cabb2b.

52. Huang ZJ, Kirkwood A, Pizzorusso T, Porciatti V, Morales B, Bear MF, et al. BDNF Regulates the Maturation of Inhibition and the Critical Period of Plasticity in Mouse Visual Cortex. Cell. 1999;98(6):739–755. doi:10.1016/S0092-8674(00)81509-3.

53. Liu Bh, Li Yt, Ma Wp, Pan Cj, Zhang L, Tao H. Broad Inhibition Sharpens Orientation Selectivity by Expanding Input Dynamic Range in Mouse Simple Cells. Neuron. 2011;71(3):542–554. doi:10.1016/j.neuron.2011.06.017.

54. Gao M, Maynard KR, Chokshi V, Song L, Jacobs C, Wang H, et al. Rebound Potentiation of Inhibition in Juvenile Visual Cortex Requires Vision-Induced BDNF Expression. J Neurosci. 2014;34(32):10770–10779. doi:10.1523/JNEUROSCI.5454-13.2014.

55. Autry AE, Monteggia LM. Brain-Derived Neurotrophic Factor and Neuropsychiatric Disorders. Pharmacol Rev. 2012;64(2):238–258. doi:10.1124/pr.111.005108.

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

57. Autry AE, Adachi M, Nosyreva E, Na ES, Los MF, Cheng Pf, et al. NMDA receptor blockade at rest triggers rapid behavioural antidepressant responses. Nature. 2011;475(7354):91–5. doi:10.1038/nature10130.

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

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

60. Verhagen M, van der Meij A, van Deurzen PAM, Janzing JGE, Arias-Vásquez A, Buitelaar JK, et al. Meta-analysis of the BDNF Val66Met polymorphism in major depressive disorder: effects of gender and ethnicity. Mol Psychiatry. 2010;15(3):260–71. doi:10.1038/mp.2008.109.
61. Notaras M, Hill R, van den Buuse M. The BDNF gene Val66Met polymorphism as a modifier of psychiatric disorder susceptibility: progress and controversy. Mol Psychiatry. 2015;20(8):916–30. doi:10.1038/mp.2015.27.

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

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

64. Giza JI, Kim J, Meyer HC, Anastasia A, Dincheva I, Zheng CI, et al. The BDNF Val66Met Prodomain Disassembles Dendritic Spines Altering Fear Extinction Circuity and Behavior. Neuron. 2018;99(1):163–178.e6. doi:10.1016/j.neuron.2018.05.024.

65. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.

66. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000;41(3):415–27.

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

68. Brzovic PS, Heikaus CC, Kisselev L, Vernon R, Herbig E, Pacheco D, et al. The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Mol Cell. 2011;44(6):942–953. doi:10.1016/j.molcel.2011.11.008.

69. Uesugi M, Nyanguile O, Lu H, Levine AJ, Verdine GL. Induced alpha helix in the VP16 activation domain upon binding to a human TAF. Science. 1997;277(5330):1310–3. doi:10.1126/science.277.5330.1310.

70. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. Solution Structure of the KIX Domain of CBP Bound to the Transactivation Domain of CREB: A Model for Activator:Coactivator Interactions. Cell. 1997;91(6):741–752. doi:10.1016/S0092-8674(00)80463-8.

71. Canales Á, Rössinger M, Sastre J, Felli IC, Jiménez-Barbero J, Giménez-Gallego G, et al. Hidden α-helical propensity segments within disordered regions of the transcriptional activator CHOP. PLoS One. 2017;12(12):e0189171. doi:10.1371/journal.pone.0189171.

72. Staller MV, Holehouse AS, Swain-Lenz D, Das RK, Pappu RV, Cohen BA. A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain. Cell Syst. 2018;6(4):444–455.e6. doi:10.1016/j.cels.2018.01.015.

73. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr Opin Struct Biol. 2015;32:102–112. doi:10.1016/J.SBFI.2015.03.008.
74. Mercadante D, Milles S, Fuertes G, Svergun DI, Lemke EA, Grätzer F. Kirkwood–Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. J Phys Chem B. 2015;119(25):7975–7984. doi:10.1021/acs.jpcb.5b03440.

75. Piana S, Donchev AG, Robustelli P, Shaw DE. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. J Phys Chem B. 2015;119(16):5113–5123. doi:10.1021/jp508971m.

76. Best RB, Hummer G. Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. J Phys Chem B. 2009;113(26):9004–9015. doi:10.1021/jp901540t.

77. Best RB, Zheng W, Mittal J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. J Chem Theory Comput. 2014;10(11):5113–5124. doi:10.1021/ct500569b.

78. Abascal JLF, Vega C. A general purpose model for the condensed phases of water: TIP4P/2005. J Chem Phys. 2005;123(23):234505. doi:10.1063/1.2121687.

79. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950–8. doi:10.1002/prot.22711.

80. 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(3):712–725. doi:10.1002/prot.21123.

81. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71–73. doi:10.1038/nmeth.4067.

82. Jorgensen WL. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. J Am Chem Soc. 1981;103(2):335–340. doi:10.1021/ja00392a016.

83. Robustelli P, Piana S, Shaw DE. Developing a molecular dynamics force field for both folded and disordered protein states. Proc Natl Acad Sci U S A. 2018;115(21):E4758–E4766. doi:10.1073/pnas.1800690115.

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

85. Rauscher S, Gapsys V, Gazda MJ, Zweckstetter M, de Groot BL, Grubmüller H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. J Chem Theory Comput. 2015;11(11):5513–5524. doi:10.1021/acs.jctc.5b00736.

86. Meng F, Bellaiche MMJ, Kim JY, Zerze GH, Best RB, Chung HS. Highly Disordered Amyloid-β Monomer Probed by Single-Molecule FRET and MD Simulation. Biophys J. 2018;114(4):870–884. doi:10.1016/j.bpj.2017.12.025.

87. 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(4):892–8. doi:10.1016/j.bpj.2011.06.046.
88. Creamer TP, Rose GD. Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. Proc Natl Acad Sci U S A. 1992;89(13):5937–41. doi:10.1073/pnas.89.13.5937.

89. Viguera AR, Serrano L. Side-chain interactions between sulfur-containing amino acids and phenylalanine in alpha-helices. Biochemistry. 1995;34(27):8771–9.

90. Faure G, Bornot A, de Brevyn AG. Protein contacts, inter-residue interactions and side-chain modelling. Biochimie. 2008;90(4):626–639. doi:10.1016/j.biochi.2007.11.007.

91. Valley CC, Cembran A, Perlmutter JD, Lewis AK, Labello NP, Gao J, et al. The methionine-aromatic motif plays a unique role in stabilizing protein structure. J Biol Chem. 2012;287(42):34979–91. doi:10.1074/jbc.M112.374504.

92. Gómez-Tamayo JC, 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(8):1517–24. doi:10.1002/pro.2955.

93. Rubinstein M, Colby RH. Polymer physics. Oxford University Press; 2003.

94. Chen ZY, Ieraci A, Teng H, Dall H, Meng CX, Herrera DG, et al. Sortilin controls intracellular sorting of brain-derived neurotrophic factor to the regulated secretory pathway. J Neurosci. 2005;25(26):6156–66. doi:10.1523/JNEUROSCI.1017-05.2005.

95. Lim JM, Kim G, Levine RL. Methionine in Proteins: It’s Not Just for Protein Initiation Anymore. Neurochem Res. 2019;44(1):247–257. doi:10.1007/s11064-017-2460-0.

96. Leopur N, Chataigner LMP, Janssen BJC. Structural insights into SorCS2–Nerve Growth Factor complex formation. Nat Commun. 2018;9(1):2979. doi:10.1038/s41467-018-05405-z.

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

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

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

100. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32(Web Server issue):W526–31. doi:10.1093/nar/gkh468.

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

102. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1-3):43–56. doi:10.1016/0010-4655(95)00042-E.
103. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1-2:19–25. doi:10.1016/j.softx.2015.06.001.

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

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

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

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

108. Nielsen JT, Mulder FAA. POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. J Biomol NMR. 2018;70(3):141–165. doi:10.1007/s10858-018-0166-5.

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

110. García AE, Sanbonmatsu KY. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. Proc Natl Acad Sci U S A. 2002;99(5):2782–7. doi:10.1073/pnas.042496899.

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

112. Parashar M, Brennan-Tonetta M, Rodero I, Villalobos JJ. High Performance Computing at the Rutgers Discovery Informatics Institute. 2018;(November). doi:10.13140/RG.2.2.11579.87846.
