

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

Ruchi Lohia<sup>1</sup>, Reza Salari<sup>1</sup>, Grace Brannigan<sup>1,2\*</sup>,

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

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

\* grace.brannigan@rutgers.edu(GB)

## Abstract

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

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

## Author Summary

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

## Introduction

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

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

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

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

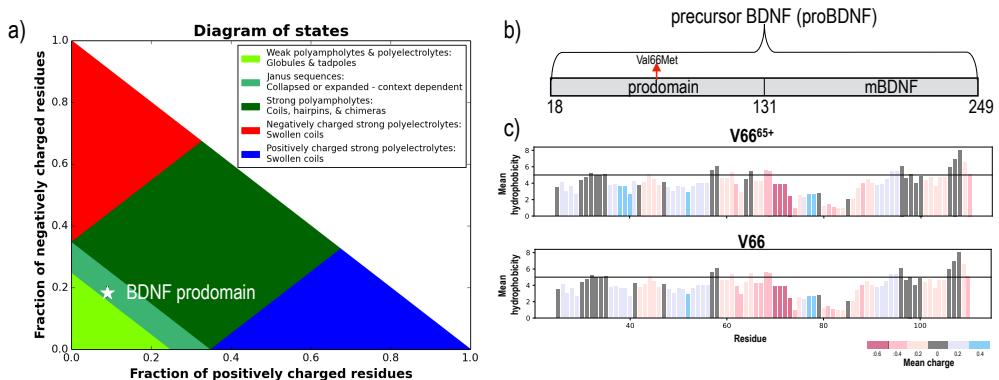
For IDPs with a relatively low fraction of charged residues, typical of the Janus region of the state diagram proposed by Das and Pappu [10,11] (Fig 1a), more subtle differences among neutral amino-acids play an increasingly important role in determining the ensemble. More than 15% of disease-associated IDP polymorphisms are substitutions between two charge-neutral residues. [12] The extent to which such substitutions in IDPs can affect non-local aspects of the conformational ensemble is uncertain; such substitution directly affects short-range interactions, and structure-based coupling between distant residues in IDPs is expected to be weak. Nonetheless, correlations between secondary structure of distant residues has been frequently observed in IDPs [14,23]; for example, several cancer mutations in transactivation domain of tumor suppressor p53 can lead to helicity changes in residues sequentially far away from the mutation sites [14].

In structured proteins, contacts between residues distant along the sequence are

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

Many SNPs in IDPs are associated with neurological, aging-associated neurodegenerative, or psychiatric disorders; despite an exponential increase in the amount of available genetic data, identifying the genetic origins of such disorders has proven remarkably challenging, with few variants identified as replicable predictors of disease. One of the earliest identified variants is the Val66Met SNP (rs6265) in the pro-domain region of Brain-derived Neurotrophic Factor (BDNF), [33] a signaling protein that retains a critical role in neurogenesis and synaptogenesis throughout adulthood (Fig 1b). [34] It has been implicated in maintenance of the hippocampus and the mechanism underlying action of numerous antidepressants, [35, 36] including rapidly acting low-dose ketamine. [37] An extensive library of genome-wide association (and even earlier) studies have repeatedly identified the Val66Met SNP as reducing hippocampal volume and episodic memory, as well as predicting increased susceptibility to neuropsychiatric disorders including schizophrenia, bipolar, and unipolar depression, but associations have been inconsistent and population dependent. [33, 37–40]

Difficulties in obtaining unambiguous disease associations at the proBDNF Val66Met SNP using GWAS are paralleled by challenges in characterizing its effects on the properties of the BDNF prodomain using structural techniques. A crystal structure of a



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

homologous neurotrophic factor in complex with a shared receptor, revealed a well-defined volume corresponding to the prodomain, but which lacked resolvable density. [41]

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

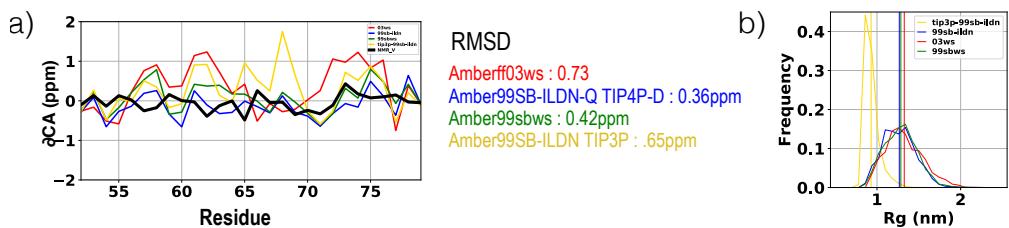
The Val66Met is present in a region with high density of negative charged residues (D61,E64,E68,E69). In this scenario, residue his65 can exist in protonated or neutral charge state *in vivo* due to it’s low pKa. In order to capture the effects of histidine protonation states on Val66Met, we study the Val66 and M66 prodomain in presence of both neutral His65 and protonated His65.

In this work, we report on unbiased fully-atomistic replica-exchange MD simulations of the 90 residue BDNF prodomain in explicit solvent, for V66, V66<sup>65+</sup>, M66 and M66<sup>65+</sup> forms. This sequence falls at the boundary of the Janus and globular domains in the diagram proposed by Das and Pappu. [10,11]

## Results and Discussion

### Selection of force field

In order to get accurate structural characterization of proBDNF with MD, we did 500ns of T-REMD simulations of 30 residue fragment of V66 proBDNF with few popular ff and water model combinations and proceed with the one which gave best agreement with experimental data obtained by NMR. Fig 2a compares the CA shifts for Amber99sb\*-ildn-q with Tip4p-D, Amber99sbws, Amberff03sbws, Amber99sb-ildn with Tip3P and NMR. Amber99sb-ildn with Tip4p-D and Amber99sbws gives good agreement with NMR CA chemical shift. We also compared Rg distribution for the tested ff (Fig 2b). Tip3P generates very collapsed ensembles and the remaining three ff generates similar Rg distribution. Tip3P is known to produce structures which are too collapsed relative to experiment. Since, the remaining three ff gives comparable Rg, we proceed with Amber99sb\*-ildn-q with Tip4p-D because it gives the best match with experiments.



**Fig 2. Ff comparison.** (a) Comparison of calculated chemical shifts from MD ensembles at 280K and NMR chemical shifts from [42] at 280K, as described in methods. (b) Rg distribution for each ff.

### Predictions from MD trajectories vs NMR data

In order to access the validity of MD generated ensembles, we compare the MD ensembles with experimental data obtained by NMR spectroscopy and NMR diffusion

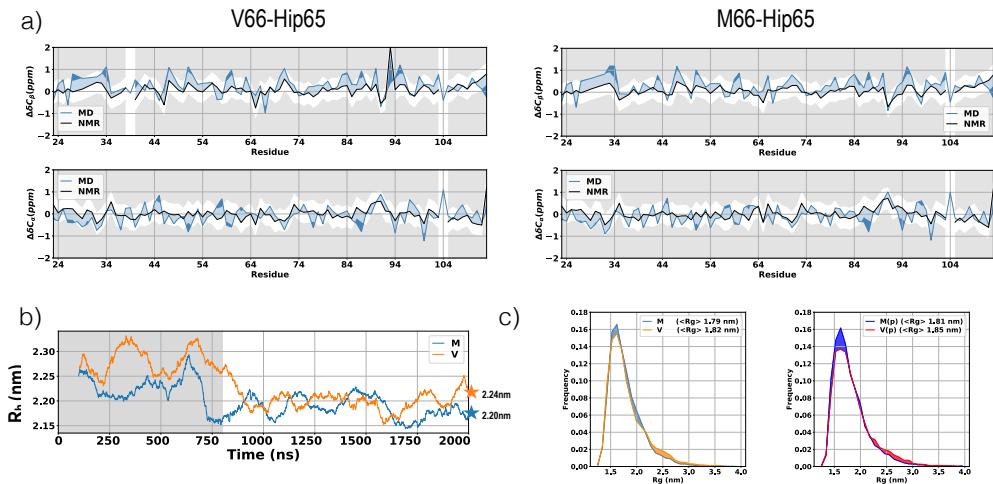
measurements. Fig 3a compares MD chemical shift and NMR chemical shifts. We get good agreement with NMR chemical shifts; deviations at each residue is <0.5 ppm. Although there are some localized discrepancies for certain residue types, we get discrepancy at same residues for all four simulations. Thus, specific discrepancy is probably reflecting residual force-field inaccuracies. Consistent with intrinsic disorder, helix and  $\beta$  propensity for each residue was low for both MD at 300K and NMR data at 280K. The simulated hydrodynamic radii (calculated using hydropro) of V66 (2.21 nm) and M66 (2.18 nm) are in excellent agreement with the experimental values (2.24 nm and 2.20 nm respectively) (Fig 3b). This, confirms that the overall dimensions of the proteins is reasonable.

### Our simulations are well converged

Most of the IDP simulations studies have been performed on smaller IDP fragments (residues 3-40). We performed the explicit solvent replica simulations of 91 residues, which was computationally challenging and thus we carefully accessed the convergence of our simulations. All replicas were able to diffuse in the temperature range 300K to 385K (replica round trip number >7) (Fig S1). The Rg for V66 and M66 converge after 800ns of simulation at 300K (Fig 3b). We discarded the first 800 ns of the trajectories as conformational equilibration. The Rg distribution of all four simulations are symmetric/unimodal (Fig 3c). Mostly, asymmetric form of Rg distribution has found to be indicative of lack of convergence.

### Val66Met changes local and non-local secondary structure

**Both mutation at residue 66 and protonation at 65 increases helix formation at residue 66.** NMR studies found differential chemical shifts at residues 66 and 93 ( marked with stars ) for Val66Met (Fig 4a). Consistent with NMR data, both Val66Met and Val<sup>65+</sup>66Met<sup>65+</sup> changes the frequency of long helix ( >6 residues ) formed at residue 63 and 93 and beta sheets ( >4 residues ) at residue 93 (Fig 4b). However, how the mutation causes this differential chemical shifts is not known. To look into the effect of mutation on residual secondary structure, we compare the helix tendency at every residue for each of the four simulation (Fig 4c). M66 and M66<sup>65+</sup> has



**Fig 3. Comparison of MD and NMR observables.** a) Comparison of calculated chemical shifts from MD ensembles at 300K and NMR chemical shifts from [42] at 280K, as described in methods. b) Rh at 100 ns moving window for V66<sup>65+</sup> and M66<sup>65+</sup> vs simulation time. The Rh for V66<sup>65+</sup> and M66<sup>65+</sup> converge after 800ns of simulation at 300K. c) Rg distribution for each simulation

increased tendency of forming  $\alpha$ -helix at residue 66 when compared with V66 and V66<sup>65+</sup> (Fig 3b). To get more insight at the helix formed at 66, we look at the population of each helix length formed at 66 (Fig 4d). Only M66 and not V66 has long helix formation at residues 59-67 (9 residues). Additionally, M66<sup>65+</sup> forms even longer helix 62- 71 (10 residues). Increased helix formations in M form is consistent with earlier observation, where an increased entropic cost of helical formation for the valine side-chain, which can access only one of three possible side-chain conformations [63]. Protonated histidine favors longer helical structure at residue 66 due to helix favored salt-bridge formation Glu61<sup>65+</sup> and Hip65-GLU69.

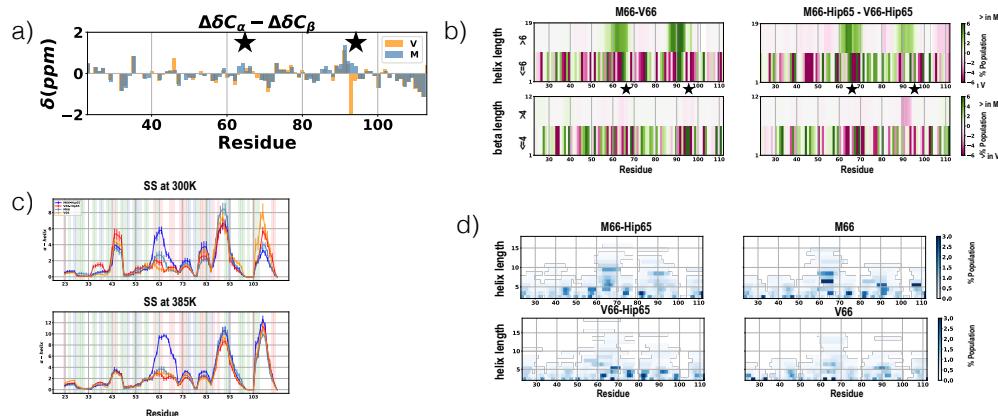
We also look at the temperature dependence of helical structure in prodomain. At 385K, the helix at residue 66 increases (>2%) for all V66, M66, V66<sup>65+</sup>, but M66<sup>65+</sup> has strongest preference (>5%) (Fig 4c).

**Only Val66 and not Met66 forms long beta sheet structures at residue 66 and 93.** We next examined the backbone contacts formed between residues. V66 and M66 forms differential backbone contact at residue 66 (Fig 5a). V66 forms beta sheet structures with residue 92. This is also consistent with NMR observation, where Val66Met changes CS at residue 93 (Fig 4a). We further test if the beta sheet structures at residue 66-95 in V66 is more favored in V66 than in M66 and is not a

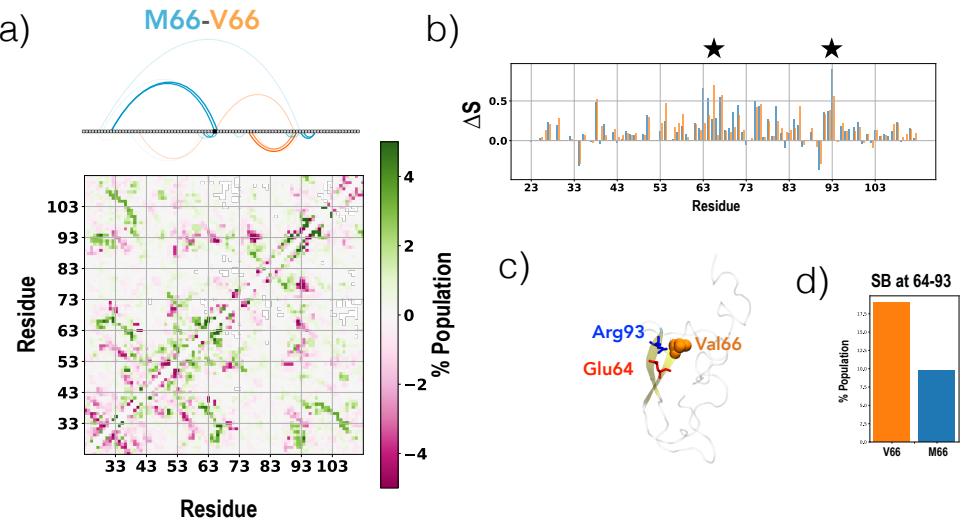
limitation of simulation convergence. To verify this observation, we performed simulations with backbone restraints of 1Kj/mole/nm<sup>2</sup> of a V66 frame having beta sheet structures at residue 66-95 with Val66 mutated to Met. We find that M66 has higher loss in entropy at 93 (Fig 5b). Additionally, 66-92 beta structures forms 64-93 salt bridge simultaneously in V66(Fig 5c). In our restrained simulations M66 has 50% less probability of forming salt-bridge at residue 64-93 (Fig 5d). This result suggests that beta at 66-95 is less favored in M66 due to higher entropic and energetic cost.

### M66 supports helix formation at residue 93

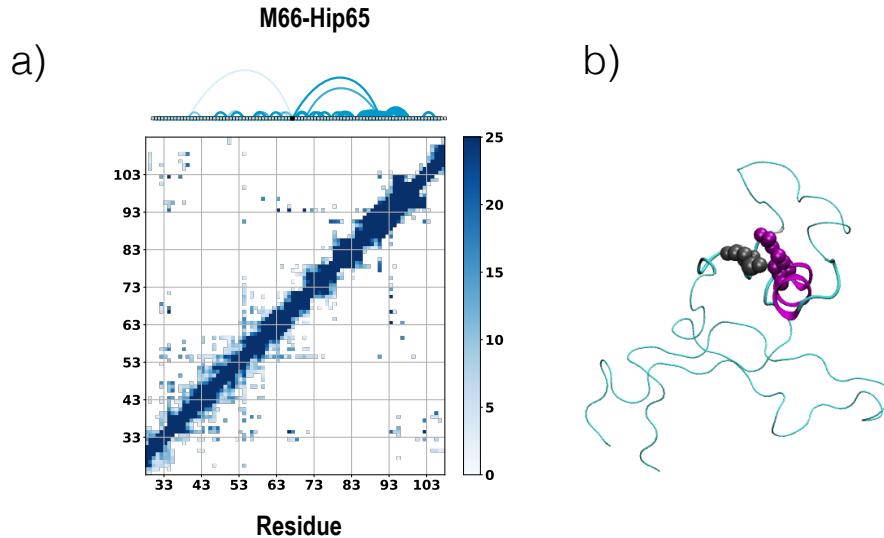
We next examined the increased helix tendency at residue 93 from Val66Met substitution. We find structures forming helix at 93 are in contact with residue 66 in at-least 25% of it's population in M66<sup>65+</sup> (Fig 6). The gain of this residue specific interaction in M66, probably increases helix formation at 93 when compared with V66.



**Fig 4. Simulation predicted helical structure and it's comparison with experiments.** a) CA-CB secondary chemical shifts for V66 and M66 from [42] at 280K. Positive difference indicate helical structure and negative differences indicate beta structure. Residues 63-67 and residues 93-95 have slightly higher tendency of being helical in M66 ( marked with stars ). b) Difference in helix length (top) and beta length (bottom) for each residue. In agreement with the experiment, we find higher tendency of forming longer helix at the regions marked with stars. V66 has higher tendency of forming beta at residue 93. c) STRIDE predicted secondary structure at each residue at 300K (top) and 385K (bottom). Protonated his65 has increased tendency of forming long helix at residue 66 only for M66. The background of the plots are colored according to residue type: blue-basic, red-acidic, green-polar, white-hydrophobic. d) Helix length distribution at each residue when 66 is in the helix region of ramachandran map (methods)



**Fig 5. Backbone contacts at residue 66.** A) Difference in V66 and M66 backbone contacts. Contact difference >5% are also shown with network representation. b) Change in entropy at every residue from random coil when 66-95 are in beta sheet structures. c) VMD representation of a V66 frame when 66 and 95 forms backbone h-bond. d) Population of salt bridge formed at 64-93 when 66 is in beta sheet structure with 92.



**Fig 6. M66 supports helix formation at residue 93** a) Backbone tertiary-contact network for  $M66^{65+}$  when 93 is in helix. Met66 stabilizes the helix formed at residue 93. b) VMD representation of a frame forming helix at 93 and contact at 66-93. 66 and 93 are shown in vanderwaal representation with grey and magenta color respectively.

#### Long range contacts involve residue 66 or the termini

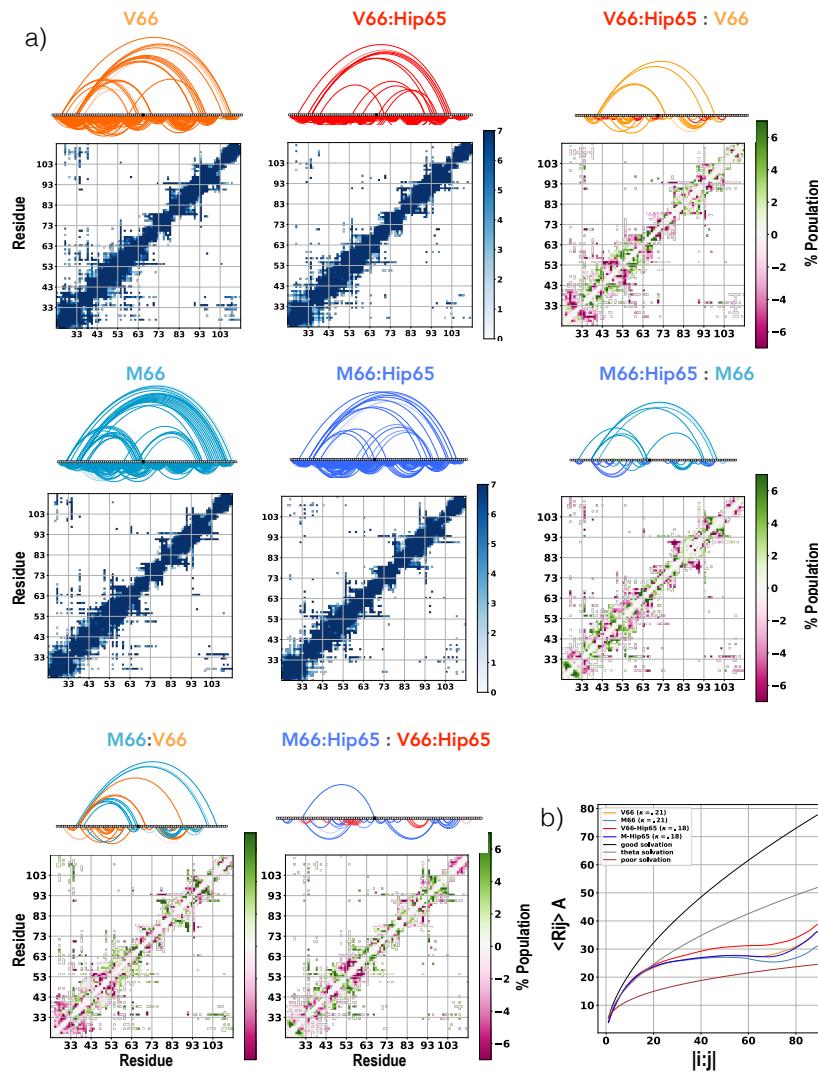
Conventional contact maps provided little useful insight into tertiary structure for either forms of the protein, which was consistent with the intrinsic disorder and frequent

transient interactions among neutral side-chains. Identification of contact residues  
yielded several persistent, weak long-range interactions at 300K for both sequences,  
which can be represented along a single axis (Fig 6).  
178  
179  
180

Prodomain has well defined regions forming long-range tertiary contacts. From the  
tertiary contact maps from all four simulations, three regions are identified which forms  
high density of long range contacts; -1': residues 31 to 35, 0': residues 57 to 69, +1':  
94-98 and +2': 106-109 (Fig 6). All the three regions identified has high density of  
hydrophobic residues (Fig ??a). It has been frequently observed that unfolded proteins  
form strong hydrophobic contacts. Thus, it's not very surprising that the disordered  
prodomain has the presence of strong hydrophobic contacts. Charges in the  
hydrophobic regions seem to determine the strength of contact formed between two  
hydrophobic regions. 0' region, which is negatively charged forms contact with  
positively charged +1' region or neutral -1' region.  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190

**Effects of H65 protonation states on long range tertiary contacts.** These  
contact maps for the entire ensemble indicate far more tertiary contacts between the  
sequence midpoint and the terminal domain of the sequence for V66 and M66 when  
compared with the protonated H65. Hip65 loses long range contacts for both V66 and  
M66. It has been earlier observed that fraction and distribution of charged residues  
along the disordered protein sequence determines the long range contacts. Several  
residues near the mutation (residues E64,E68, and E69) are negatively charged. Both  
V66 and M66 loses salt-bridge with the gain of positive charged residue in the  
otherwise negatively charged region of the sequence.  
191  
192  
193  
194  
195  
196  
197  
198  
199

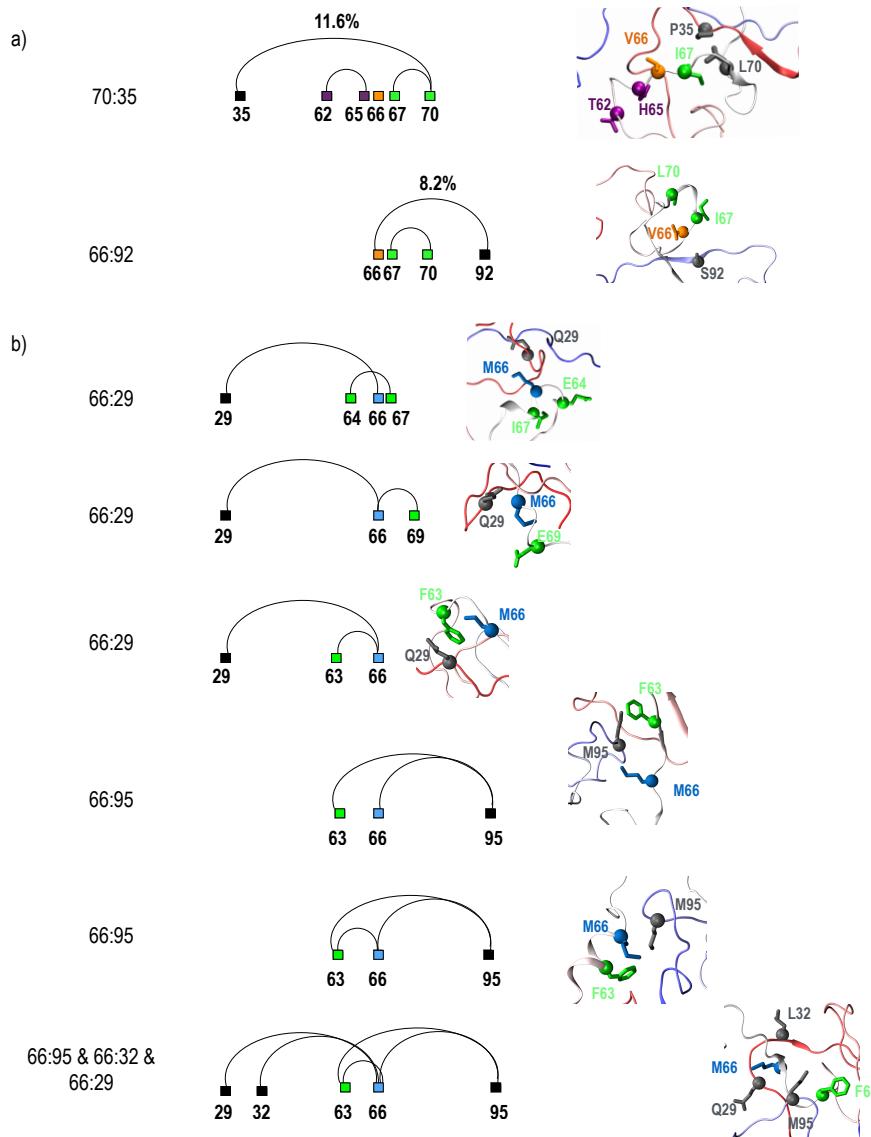
**V66 collapse is driven by electrostatic contacts at center whereas M66  
collapse is driven by hydrophobic contacts.** We look at the effect of Val66Met on  
long-range contacts. Even though the prodomain is disordered we identify loss of  
specific residue contact at 66. M66 and  $M66^{65+}$  forms strong hydrophobic contacts at  
residue M66-Y34 when compared with V66 and  $V66^{65+}$  respectively. In V66, Y34 forms  
strong contact with residue L70. However, this hydrophobic contact is formed with  
simultaneous salt bridge formation between residue 27 and 72. The strong hydrophobic  
contacts at residue 66 forms comparatively collapsed structures in M66, M66 (p) when  
compared with V66 and V66 (p).  
200  
201  
202  
203  
204  
205  
206  
207  
208



**Fig 7. Linear networks of transient tertiary contacts.** a) The backbone tertiary-contact network is made for V66 and M66, with each residue serving as a node in the network, as described in Methods. A contact is formed if two residues are within .45nm of each other. If the two residues forming contact are more than 24 residues apart the edge is drawn on the top of the node, otherwise the edge is drawn at the bottom of the node. Backbone interactions serve as edges between individual network nodes; the thickness and the transparency of the edge corresponds to the strength of the contact. Contacts observed in 37 or more replicas are only visible. Gain of protonation states loses contacts formed from 0' region for both V66 and M66. V66 to M66 gains contact at residue 66-34 for both neutral and protonated histidine at residue 65. b) Internal residue level scaling for each peptide.

## Conclusion

We have carried out .5ms of fully-atomistic MD simulation of the 90 residue prodomain of brain-derived neurotrophic factor with protonated and neutral His65 state, for both



**Fig 8. Long range contacts are correlated with short range contact at residues 63, 65, 66 or 67, 69.** Two strongest non-correlated long range contact formed in V66 a) and M66 b).

with and without the disease-associated Val66Met mutation.

Anastasia et al [42] observed differential kinetics for interactions between BDNF prodomain and SorCS2; M66 binds more strongly at residues H65 to L71 with SorCS2, whereas V66 binds more strongly at residue Y90 to V94. The stronger binding at residue M66 could be attributed to either a) ability of M66 to form alpha-helix at residue 66 when compared with V66 or b) stronger hydrophobic contact at residue M66 with the SorCS2 hydrophobic residues or combination of both a) and b).

Residue 66 possess several meaningful properties, beyond including the  
disease-associated mutation of our original interest. It is a) neutral residues inserted in a  
stretch of acidic residues constituting the most highly charged region of the protein and  
b) directly adjacent to the sequence midpoint at E68. We have not yet isolated which of  
these contribute to the critical role of residue 66 in forming strong tertiary interactions,  
and it is possible that overlap of (a) with (b) is intrinsic to the protein design.

## Materials and Methods

### System setup

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

### Molecular Dynamics Simulation

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

Each replica was then simulated using T-REMD [52] with an exchange frequency of  
1ps for 650 ns, giving a total simulation time of 78 μs with NVT ensemble. A different

random seed was used for the Langevin dynamics of each replica. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method [53], with a 1 nm cutoff and a 0.12 nm grid spacing. Periodic boundary conditions were also used to reduce system size effects. Bonds with H-atoms were constrained using the LINCS (linear constraint solver) algorithm [54]. The time constant for temperature coupling is 1.5ps. The average exchange acceptance probability ranged between 0.13-0.23 for both V66 and M66. For both V66 and M66 groups, about 500 ns for each replica were discarded for equilibration purposes.

## Analysis of MD Trajectories

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

### Helix length calculation

The length of helix formed at each residue was calculated by determining the number of consecutive residues in which the dihedral angles satisfied  $\phi < 0^\circ$  and  $-120^\circ < \psi < 50^\circ$ , as in [56, 57].

### Tertiary contacts network

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

interactions, interactions showing more than 3% persistence were considered in network  
275 visualization.

## Acknowledgments

The authors are grateful to Dr. Clay Bracken and Dr. Barbara Hempstead of Weill  
277 Cornell Medical Center for helpful discussions. Computational time was provided  
279 through XSEDE resources via NSF MCB110149.  
280

## 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. 282
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. 284
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. 287
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. 290
5. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J Mol Biol*. 2004;337(3):635–645. doi:10.1016/j.jmb.2004.02.002. 292
6. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J*. 2005;272(20):5129–5148. doi:10.1111/j.1742-4658.2005.04948.x. 295
7. Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem Rev*. 2014;114(13):6561–6588. doi:10.1021/cr400514h. 298

8. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol.* 2011;21(3):432–40. doi:10.1016/j.sbi.2011.03.011. 300  
301  
302
9. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 2007;35(Database):D786–D793. doi:10.1093/nar/gkl893. 303  
304  
305
10. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol.* 2015;32:102–112. doi:10.1016/j.sbi.2015.03.008. 306  
307  
308
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. 309  
310  
311
12. Vacic V, Markwick PRL, Oldfield CJ, Zhao X, Haynes C, Uversky VN, et al. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Comput Biol.* 2012;8(10):e1002709. 312  
313  
314  
315
13. Larini L, Gessel MM, LaPointe NE, Do TD, Bowers MT, Feinstein SC, et al. Initiation of assembly of tau(273–284) and its ΔK280 mutant: an experimental and computational study. *Phys Chem Chem Phys.* 2013;15(23):8916. 316  
317  
318  
319
14. Ganguly D, Chen J, Dyson H, Wright P, Uversky V, Oldfield C, et al. Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. *PLOS Comput Biol.* 2015;11(4):e1004247. doi:10.1371/journal.pcbi.1004247. 320  
321  
322  
323
15. Viet MH, Nguyen PH, Derreumaux P, Li MS. Effect of the English Familial Disease Mutation (H6R) on the Monomers and Dimers of Aβ40 and Aβ42. *ACS Chem Neurosci.* 2014;5(8):646–657. doi:10.1021/cn500007j. 324  
325  
326
16. Viet MH, Nguyen PH, Ngo ST, Li MS, Derreumaux P. Effect of the Tottori Familial Disease Mutation (D7N) on the Monomers and Dimers of Aβ 327  
328

- <sub>40</sub> and A $\beta$  <sub>42</sub>. ACS Chem Neurosci. 329  
2013;4(11):1446–1457. doi:10.1021/cn400110d. 330
17. Truong PM, Viet MH, Nguyen PH, Hu CK, Li MS. Effect of Taiwan Mutation 331  
(D7H) on Structures of Amyloid- $\beta$  Peptides: Replica Exchange Molecular 332  
Dynamics Study. J Phys Chem B. 2014;118(30):8972–8981. 333  
doi:10.1021/jp503652s. 334
18. Zhan YA, Wu H, Powell AT, Daughdrill GW, Ytreberg FM. Impact of the K24N 335  
mutation on the transactivation domain of p53 and its binding to murine 336  
double-minute clone 2. Proteins Struct Funct Bioinforma. 2013;81(10):1738–1747. 337  
doi:10.1002/prot.24310. 338
19. Xu L, Shan S, Wang X. Single Point Mutation Alters the Microstate Dynamics of 339  
Amyloid  $\beta$ -Protein A $\beta$ 42 as Revealed by Dihedral Dynamics Analyses. J Phys 340  
Chem B. 2013;117(20):6206–6216. doi:10.1021/jp403288b. 341
20. Bah A, Forman-Kay JD. Modulation of Intrinsically Disordered Protein Function 342  
by Post-translational Modifications. J Biol Chem. 2016;291(13):6696–6705. 343  
doi:10.1074/jbc.R115.695056. 344
21. He Y, Chen Y, Mooney SM, Rajagopalan K, Bhargava A, Sacho E, et al. 345  
Phosphorylation-induced Conformational Ensemble Switching in an Intrinsically 346  
Disordered Cancer/Testis Antigen. J Biol Chem. 2015;290(41):25090–25102. 347  
doi:10.1074/jbc.M115.658583. 348
22. Alexander Conicella AE, Zerze GH, Mittal J, Fawzi Correspondence NL, 349  
Conicella AE, Fawzi NL. ALS Mutations Disrupt Phase Separation Mediated by 350  
 $\alpha$ -Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Struct 351  
Des. 2016;24:1537–1549. doi:10.1016/j.str.2016.07.007. 352
23. Iešmantavičius V, Jensen MR, Ozenne V, Blackledge M, Poulsen FM, Kjaergaard 353  
M. Modulation of the Intrinsic Helix Propensity of an Intrinsically Disordered 354  
Protein Reveals Long-Range Helix–Helix Interactions. J Am Chem Soc. 355  
2013;135(27):10155–10163. doi:10.1021/ja4045532. 356

24. Mittag T, Forman-Kay JD. Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol.* 2007;17(1):3–14. doi:10.1016/j.sbi.2007.01.009. 357
25. Stanley N, Esteban-Martín S, De Fabritiis G. Progress in studying intrinsically disordered proteins with atomistic simulations. *Prog Biophys Mol Biol.* 2015;119:47–52. doi:10.1016/j.pbiomolbio.2015.03.003. 359  
360  
361
26. Ithurralde 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. 362  
363  
364
27. Knott M, Best RB, Hummer G, de Bakker P, Word J. A Preformed Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. *PLoS Comput Biol.* 2012;8(7):e1002605. 365  
366  
367  
368
28. Invernizzi G, Lambrughi M, Regonesi ME, Tortora P, Papaleo E. The conformational ensemble of the disordered and aggregation-protective 182–291 region of ataxin-3. *Biochim Biophys Acta.* 2013;1830(11):5236–47. 369  
370  
371  
372
29. Abeln S, Frenkel D. Disordered flanks prevent peptide aggregation. *PLoS Comput Biol.* 2008;4(12):e1000241. doi:10.1371/journal.pcbi.1000241. 373  
374
30. Yedvabny E, Nerenberg PS, So C, Head-Gordon T. Disordered Structural Ensembles of Vasopressin and Oxytocin and Their Mutants. *J Phys Chem B.* 2015;119(3):896–905. doi:10.1021/jp505902m. 375  
376  
377
31. Levine ZA, Larini L, LaPointe NE, Feinstein SC, Shea JE. Regulation and aggregation of intrinsically disordered peptides. *Proc Natl Acad Sci U S A.* 2015;112(9):2758–63. doi:10.1073/pnas.1418155112. 378  
379  
380
32. Pappu RV, Wang X, Vitalis A, Crick SL. A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch Biochem Biophys.* 2008;469(1):132–41. doi:10.1016/j.abb.2007.08.033. 381  
382  
383

33. Notaras M, Hill R, van den Buuse M. The BDNF gene Val66Met polymorphism 384  
as a modifier of psychiatric disorder susceptibility: progress and controversy. Mol 385  
Psychiatry. 2015;20(8):916–930. doi:10.1038/mp.2015.27. 386
34. Korte M, Carroll P, Wolf E, Brem G, Thoenen H, Bonhoeffer T. Hippocampal 387  
long-term potentiation is impaired in mice lacking brain-derived neurotrophic 388  
factor. Proc Natl Acad Sci. 1995;92(19):8856–8860. doi:10.1073/pnas.92.19.8856. 389
35. Autry AE, Monteggia LM. Brain-derived neurotrophic factor and 390  
neuropsychiatric disorders. Pharmacol Rev. 2012;64(2):238–258. 391  
doi:10.1124/pr.111.005108. 392
36. Björkholm C, Monteggia LM. BDNF - a key transducer of antidepressant effects. 393  
Neuropharmacology. 2016;102:72–79. doi:10.1016/j.neuropharm.2015.10.034. 394
37. Autry AE, Adachi M, Nosyreva E, Na ES, Los MF, Cheng Pf, et al. NMDA 395  
receptor blockade at rest triggers rapid behavioural antidepressant responses. 396  
Nature. 2011;475(7354):91–95. doi:10.1038/nature10130. 397
38. Soliman F, Glatt CE, Bath KG, Levita L, Jones RM, Pattwell SS, et al. A 398  
genetic variant BDNF polymorphism alters extinction learning in both mouse 399  
and human. Science. 2010;327(5967):863–6. doi:10.1126/science.1181886. 400
39. Chen ZY, Bath K, McEwen B, Hempstead B, Lee F. Impact of genetic variant 401  
BDNF (Val66Met) on brain structure and function. Novartis Found Symp. 402  
2008;289:180–8; discussion 188–95. 403
40. Verhagen M, van der Meij A, van Deurzen PAM, Janzing JGE, Arias-Vásquez A, 404  
Buitelaar JK, et al. Meta-analysis of the BDNF Val66Met polymorphism in 405  
major depressive disorder: effects of gender and ethnicity. Mol Psychiatry. 406  
2010;15(3):260–71. doi:10.1038/mp.2008.109. 407
41. Feng D, Kim T, Özkan E, Light M, Torkin R, Teng KK, et al. Molecular and 408  
Structural Insight into proNGF Engagement of p75NTR and Sortilin. J Mol Biol. 409  
2010;396(4):967–984. doi:10.1016/j.jmb.2009.12.030. 410

42. Anastasia A, Deinhardt K, Chao MV, Will NE, Irmady K, Lee FS, et al. Val66Met polymorphism of BDNF alters prodomain structure to induce neuronal growth cone retraction. *Nat Commun.* 2013;4:2490. doi:10.1038/ncomms3490. 411  
412  
413
43. Holehouse AS, Das RK, Ahad JN, Richardson MOG, Pappu RV. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. *Biophys J.* 2017;112(1):16–21. doi:10.1016/j.bpj.2016.11.3200. 414  
415  
416
44. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods.* 2014;12(1):7–8. 417  
418  
419
- doi:10.1038/nmeth.3213.
45. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc.* 2010;5(4):725–738. 420  
421  
422
- doi:10.1038/nprot.2010.5.
46. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008;9(1):40. doi:10.1186/1471-2105-9-40. 423  
424
47. Šali A, Blundell TL. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J Mol Biol.* 1993;234(3):779–815. doi:10.1006/jmbi.1993.1626. 425  
426
48. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 1983;79(2):926–935. doi:10.1063/1.445869. 427  
428  
429
49. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct Funct Bioinforma.* 2010; p. NA–NA. doi:10.1002/prot.22711. 430  
431  
432
50. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun.* 1995;91(1-3):43–56. doi:10.1016/0010-4655(95)00042-E. 433  
434  
435
51. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1-2(June 2016):19–25. 436  
437  
438  
439
- doi:10.1016/j.softx.2015.06.001.

52. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett.* 1999;314(1-2):141–151. doi:10.1016/S0009-2614(99)01123-9. 440  
441  
442
53. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *J Chem Phys.* 1995;103(19):8577–8593. doi:10.1063/1.470117. 443  
444  
445
54. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem.* 1997;18(12):1463–1472. doi:10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H. 446  
447  
448
55. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph.* 1996;14(1):33–8, 27–8. 449  
450
56. Nodet G, Salmon L, Ozenne V, Meier S, Jensen MR, Blackledge M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J Am Chem Soc.* 2009;131(49):17908–17918. doi:10.1021/ja9069024. 451  
452  
453  
454
57. Iglesias J, Sanchez-Martínez M, Crehuet R. SS-map. Intrinsically Disord Proteins. 2013;1(1):e25323. doi:10.4161/idp.25323. 455  
456
58. Ahlstrom LS, Baker JL, Ehrlich K, Campbell ZT, Patel S, Vorontsov II, et al. Network visualization of conformational sampling during molecular dynamics simulation. *J Mol Graph Model.* 2013;46:140–9. doi:10.1016/j.jmgm.2013.10.003. 457  
458  
459
59. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR.* 2009;44(4):213–23. doi:10.1007/s10858-009-9333-z. 460  
461  
462
60. Marsh JA, Singh VK, Jia Z, Forman-Kay JD. Sensitivity of secondary structure propensities to sequence differences between  $\alpha$ - and  $\gamma$ -synuclein: Implications for fibrillation. *Protein Sci.* 2006;15(12):2795–2804. doi:10.1110/ps.062465306. 463  
464  
465
61. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M. The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of 466  
467

Ordered and Disordered Regions in Proteins. *J Mol Biol.* 2015;427(4):982–996. 468  
doi:10.1016/j.jmb.2014.12.007. 469

62. Shen Y, Bax A. SPARTA+: a modest improvement in empirical NMR chemical 470  
shift prediction by means of an artificial neural network. *J Biomol NMR.* 471  
2010;48(1):13–22. doi:10.1007/s10858-010-9433-9. 472
63. Creamer TP, Rose GD. Side-chain entropy opposes  $\alpha$ -helix formation but 473  
rationalizes experimentally determined helix-forming propensities ( $\alpha$ -helix/protein 474  
folding/protein engineering). *Nat Sci.* 1992;89(250):5937–5941. 475

## Supporting Information

**S1 Fig. Mixing of replicas during the simulation.** a) Mean square displacement 477  
(MSD) of each replica in the temperature range (300K-420K) for both V66 and M66. 478  
Replicas visiting 300K are colored green and remaining replicas are colored red. b) 479  
Population of each replica at 300K. c) Number of round trips completed by each replica. 480

**S2 Fig. Temperature dependence of helical length around residue 66.** a) 481  
Helical propensity at residues 50–77, determined from STRIDE, for helix of length 10,11 482  
and 12 containing residue 66, with curves colored according to temperature. M66 folds 483  
into a 12 residue helix at residues 62–74 at high temperature. Representative 484  
conformations are shown, colored by secondary structure, with residue 66 in stick 485  
representation. b) Total number of replicas at which lengths for helices containing 486  
residue 66 is observed, for a range of temperatures. The helices of longer lengths are 487  
formed in 4 or more replicas at high temperatures. 488

**S3 Fig. Backbone hydrogen bonding partners of residue 66 and length of 489  
beta sheet formed at every residue at 300K .** a) Population of backbone 490  
hydrogen bonding of residue 66 with all other residues in the sequence. Residue 66 491  
forms a weak contact with residue at +4' region b) Length of beta strands formed at 492  
each residue. Pairing regions show higher density of beta of length 3 or more for both 493  
V66 and M66. The strand length formed at each residue was calculated by determining 494

the number of consecutive residues in which the dihedral angles satisfied  $\psi > 50^\circ$  and  $\phi < -90^\circ$  or  $\psi < -120^\circ$  and  $\phi < -90^\circ$ .  
495

**S4 Fig. Distribution of dihedral angles at residue 65,66 and 67 for each cluster at 300K.**  
497  
498

**S5 Fig. Distribution of dihedral angles with temperature at residue 65,66 and 67 at each cluster.** Dotted lines correspond to  $-120^\circ < \psi < 50^\circ$ .  
499  
500

**S6 Fig. Linear networks of transient tertiary contacts** a) Hydrogen-bonding and b) salt-bridging pairs are shown for the entire ensemble and then decomposed by cluster at high temperatures (398K-420K). The backbone tertiary-contact network is made for V66 and M66, with each residue serving as a node in the network, as described in Methods. Residues at the pairing regions are colored black. Backbone interactions serve as edges between individual network nodes; the thickness of the edge corresponds to the strength of the hydrogen bond, and the transparency of the edge increases as its frequency increases. If residue 66 or its nearby residues (A51-P79) are involved in hydrogen-bond formation then the edge is drawn above the node; otherwise, it is drawn at the bottom of the node. Fewer contact pairs are observed at high temperatures. V66 forms stronger backbone hydrogen-bonding for beta bridge formations and M66 forms stronger backbone hydrogen-bonding for helix formation at residue 66. Cluster HHH in M66 simultaneously forms salt-bridge and hydrogen bonds from residues near the SNP.  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513

**S7 Fig. Secondary structure propensity, determined using STRIDE from MD trajectories, at each residue, with curves colored according to temperature.**  
514  
515  
516

**S8 Fig. Length of helix formed at every cluster at 300K.** Cluster HHH and Cluster XHB can form long cooperative helices at residue 66. A Smaller helix with 8 or fewer residues forms near residue 66 for clusters BXH,BBB and HBB.  
517  
518  
519

**S9 Fig. Distribution of  $R_g$  for each cluster.**  $\langle R_g \rangle$  with temperature for a) entire ensemble and b) each cluster. The  $\langle R_g \rangle$  for V66 and M66 reverses with increase in temperature. We observe the same trend reversal for cluster HHH. c) Distribution of  
520  
521  
522

$R_g$  for each cluster at every temperature. The line color transitions from blue (cold) to red (hot) with increase in temperature from 300K to 420K. 523

**S10 Fig. Distribution of amplitude of the fitted gaussian curves with temperature.** Cluster BXH and XHB shows high amplitude of collapsed and expanded states, respectively at 300K. Cluster HHH shows increase in amplitude for collapsed structure at high temperature. 525  
526  
527  
528

**S11 Fig. Population densities of salt bridges and hydrogen bonds for V66 and M66 for each cluster at low temperatures (300K-317K).** Simultaneous salt bridge and hydrogen bonds formation stabilize cluster BXH and only salt bridge formation destabilize cluster XHB. a) Population densities of total number of salt bridges per frame vs  $R_g$  (normalized with respect to total number of frames). b) Population densities of total number of hydrogen bonds per frame vs  $R_g$  (normalized with respect to total number of frames). c) Population densities of total number of salt bridge per frame vs total number of hydrogen bonds per frame ( normalized with respect to total number of frames). 529  
530  
531  
532  
533  
534  
535  
536  
537