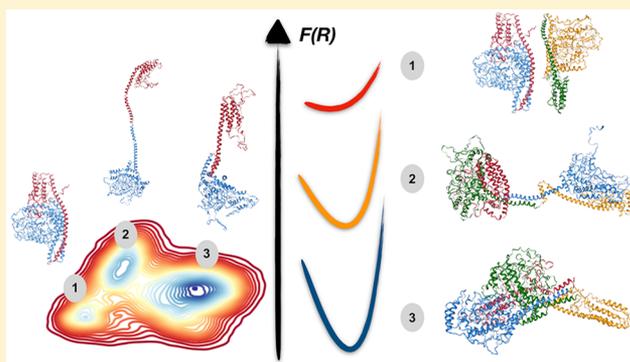


# Disorder Mediated Oligomerization of *DISC1* Proteins Revealed by Coarse-Grained Molecular Dynamics Simulations

Julien Roche<sup>†</sup> and Davit A. Potoyan<sup>\*,‡,¶,†</sup><sup>†</sup>Roy J. Carver Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011, United States<sup>‡</sup>Department of Chemistry, Iowa State University, Ames, Iowa 50011, United States<sup>¶</sup>Bioinformatics and Computational Biology Program, Iowa State University, Ames, Iowa 50011, United States

## S Supporting Information

**ABSTRACT:** Disrupted-in-schizophrenia-1 (*DISC1*) is a scaffold protein of significant importance for neuro-development and a prominent candidate protein in the etiology of mental disorders. In this work, we investigate the role of conformational heterogeneity and local structural disorder in the oligomerization pathway of the full-length *DISC1* and of two truncation variants. Through extensive coarse-grained molecular dynamics simulations with a predictive energy landscape-based model, we shed light on the interplay of local and global disorder which lead to different oligomerization pathways. We found that both global conformational heterogeneity and local structural disorder play an important role in shaping distinct oligomerization trends of *DISC1*. This study also sheds light on the differences in oligomerization pathways of the full-length protein compared to the truncated variants produced by a chromosomal translocation associated with schizophrenia. We report that oligomerization of full-length *DISC1* sequence works in a nonadditive manner with respect to truncated fragments that do not mirror the conformational landscape or binding affinities of the full-length unit.



## INTRODUCTION

Regulation, recognition, and cell signaling involves the coordinated actions of many players, including scaffold proteins that provide selective spatial and temporal coordination among interacting proteins. Scaffold proteins have typically no catalytic activity but instead, they bring together interacting proteins within a signaling pathway to facilitate or modify the specificity of their interactions. A common characteristic of scaffold proteins is the extent of structural disorder.<sup>1</sup> The structural disorder is associated with several key attributes of scaffold proteins, such as the ability to bind multiple partners, mediating transient interactions, and undergoing a complex array of regulatory post-translational modifications.<sup>2</sup>

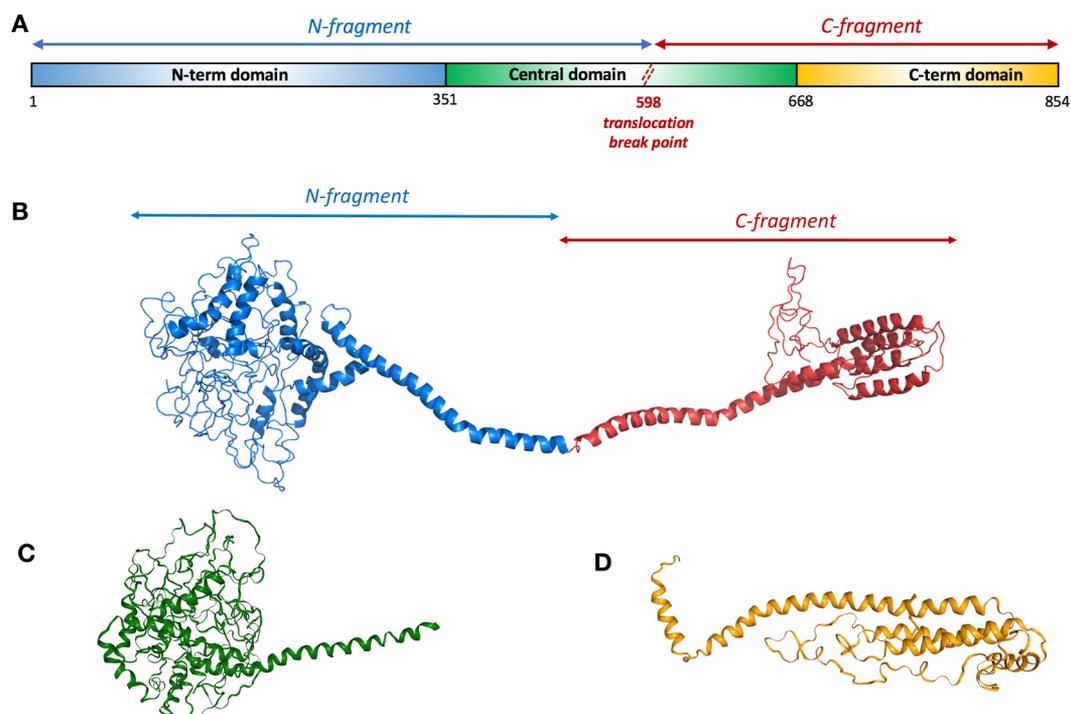
To understand the role of structural disorder in mediating the function of scaffold proteins, we carried out extensive molecular simulations on *DISC1*, a canonical scaffold protein involved in neuron differentiation and migration. *DISC1* was originally identified at the break-point of a (1:11) chromosomal translocation in a Scottish family with a high loading of major mental illness,<sup>3,4</sup> and it is now established by considerable genetic evidence as a risk factor for a wide range of psychiatric disorders.<sup>5–8</sup> Beyond the balanced chromosomal translocation, multiple other sequence variations of *DISC1* are associated with schizophrenia, depression, and

autism.<sup>9,10</sup> The product of *DISC1* is a scaffold protein that regulates the activity, mostly through inhibition, of a considerable number of enzymes of interrelated signaling cascades in the central nervous system.<sup>11–13</sup> Despite growing appreciation of its role in the etiology of mental disorders, almost nothing is known about the structural properties of *DISC1* by itself and in complex with its protein partners. *DISC1* is an 854-residue long protein that shares many features with other long scaffold proteins containing fully or partially disordered regions. Bioinformatics predictions suggest that the first 350 amino acids of *DISC1* are predominantly disordered (Figure S1) whereas the remainder of the protein is rich in  $\alpha$ -helical and coiled-coil motifs. The central region, which is essential for *DISC1* self-association, is predicted to form a discontinuous coiled-coil domain with no continuous stretches greater than 50 amino acids.<sup>14</sup> Specifically the N-terminus is rich in glycine (12.6%), serine (15.2%) and charged amino acids (35 negatively charged and 38 positively charged amino acids). Both bioinformatics predictions (as shown in Figure S1) and high-throughput expression of truncated fragments suggest

Received: August 5, 2019

Revised: October 15, 2019

Published: October 15, 2019



**Figure 1.** Domain organization of (A) full length, (B) N-terminal and (C) C-terminal, and (D) N-terminal fragments of *DISC1* sequence together with representative structures obtained through molecular dynamics simulations. Consistent color coding of structural domains is used throughout the manuscript with blue/green corresponding to N-terminal domain(s) and red/orange to C-terminal domain(s).

that the N-terminus is indeed significantly disordered in solution.<sup>15</sup>

An early study has predicted the presence of two UVR domains in this region, consisting of two  $\alpha$ -helices, packed against each other in an antiparallel hairpin fold.<sup>16</sup> The C-terminus contains two regions with coiled-coil propensity and two leucine zippers motifs.<sup>14</sup> Zhang and co-workers have recently solved the atomic structure of *DISC1* C-terminal tail (aa. 765–835) in complex with Nde1, revealing a helix-turn-helix structure,<sup>17</sup> which was later confirmed by SAXS.<sup>18</sup> Overall, the C-terminal region appears to possess the characteristics of a series of helical bundles that mediate protein–protein interactions.

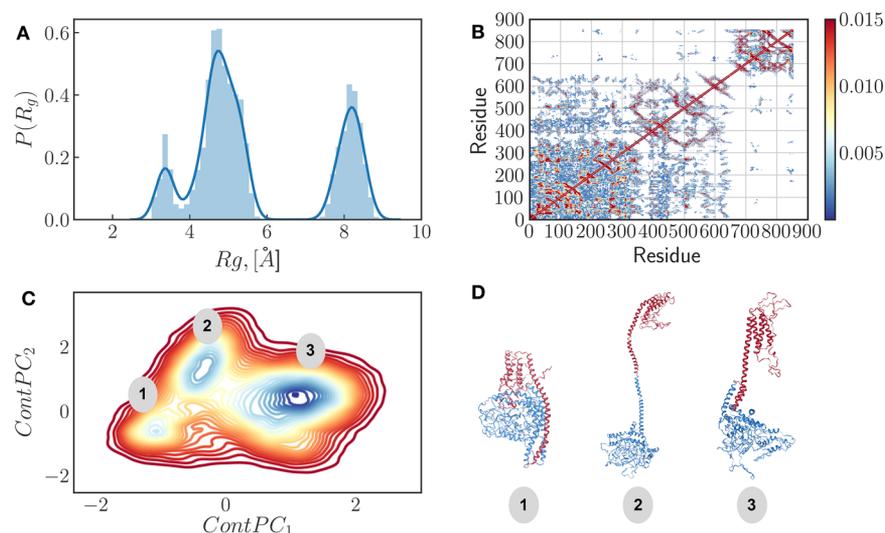
Wagner and co-workers have shown that the full-length protein exists in equilibrium between octamers and dimers, but this equilibrium is delicate and examples have been reported where single point mutations or truncations affect the oligomerization state of *DISC1*.<sup>19–21</sup> The oligomerization state of *DISC1* appears to be a delicate equilibrium between dimeric and octameric species, potentially modulated by PKA-induced phosphorylation.<sup>22</sup> Disease-associated mutations such as S704C have also been reported to affect the oligomerization state of *DISC1*.<sup>19</sup> Yet, the structure of *DISC1* oligomers and pathways to oligomerization remain largely unknown.

To unravel the specific contributions of conformational dynamics and structural disorder to the formation of oligomers, we have conducted extensive long-time scale simulations using a coarse-grained predictive model of the full-length *DISC1* and the truncated fragments resulting from the (1:11) chromosomal translocation associated with schizophrenia.<sup>3,4</sup> Through a set of constant temperature simulations, we first explored equilibrium conformations ensemble of all monomeric units revealing large differences in the conformational ensembles sampled by the full-length

protein and the N- and C-fragments. Subsequently, we have run a different set of constant temperature simulations and umbrella sampling simulations in order to study candidate oligomerization structures and their oligomerization free energy landscapes. By employing various measures for protein contact analysis in the form of contact frequency maps, intradomain contact distributions, and principal component analysis in the space of residue–residue contacts we have revealed the interplay of contacts that shapes distinct conformations preferences and binding modes of subunits.

## METHODS

We have used Phyre2<sup>23</sup> to create an initial 3D model of the full-length *DISC1* based on homology modeling. This initial structural model (691 amino acids over a total of 854 residues were modeled at  $\geq 90\%$  confidence) was further refined with ModRefiner<sup>24</sup> and used for extended MD simulations with the coarse-grained protein force field AWSEM (associative memory, water mediated, structure and energy model).<sup>25</sup> As a coarse-grained molecular model for proteins, AWSEM resolves each amino acid with three interacting sites corresponding to  $C_\alpha$ ,  $C_\beta$ , and O atoms. The remaining atoms can be reconstructed after employing appropriate geometric constraints of the polypeptide chain. The Hamiltonian of AWSEM defining the interaction between sites includes both knowledge-based terms acting at local secondary structure level and physics-based terms such as hydrogen bonding and water-mediated interactions for capturing interactions at the level of tertiary structures.<sup>26</sup> The judicious combination of physical and knowledge-based terms has been shown to perform well for *de novo* predictions of the majority of proteins with fold-able landscapes including native states and allosteric conformational states of single and multidomain proteins.<sup>26–28</sup> Thanks to the predictive nature of the AWSEM



**Figure 2.** (A) Distribution of radii of gyration for full monomer unit of *DISC1*. (B) Contact frequency map showing the  $C_{\alpha}$ – $C_{\alpha}$  contacts that are observed more frequently in simulations. (C) Probability distribution of simulated trajectories onto principal components in the space of physical  $C_{\alpha}$ – $C_{\alpha}$  contacts that correspond to high contact frequencies. (D) Representative structures from simulations corresponding to distinct regions of the first two principal components.

model, it has found wide applicability for solving various biophysical problems including predictions of monomeric protein units, protein aggregation,<sup>29,30</sup> protein–DNA assembly,<sup>31,32</sup> and proteins with unstructured regions<sup>33</sup> and recently also for natively disordered proteins.<sup>34</sup> In this work, we have used the AWSEM Hamiltonian supplemented with electrostatic potential<sup>35</sup> and multiple fragment memories<sup>33,36</sup> in order to account for disordered nature of the *DISC1* sequence and model conformational heterogeneity. The conformational heterogeneity stems from the interplay of local disorder and tertiary interactions which give *DISC1* its signature elongated shape. Both secondary structural disorder and tertiary interactions are accounted for by the AWSEM coarse-grained Hamiltonian through knowledge-based and physics-based terms, respectively.

All of the molecular dynamics simulations were run using the LAMMPS molecular dynamics simulator. The Langevin integrator in LAMMPS was employed with a 5 fs integration time step and with a damping time of  $10^3$  fs [CG – time]. Simulations were run for 20–50 million steps until convergence was reached, which was checked by comparing different sets of independent runs for agreements between first few eigenvectors of contact-based principal components and root-mean-square inner product (Supporting Information, Figures S2 and S3). Conformations of monomers and dimers were saved at 5000 intervals.

We note that due to coarse-grained nature of the AWSEM force-field mapping between coarse-grained time [CG-time] and wall-time needs to account for acceleration in sampling resulting from the absence of solvent and protein degrees of freedom which smoothens the free energy landscapes of proteins.<sup>26</sup> Previous works employing AWSEM for modeling protein dynamics<sup>32,33</sup> have shown that approximately ~10–100 fold acceleration is achieved with proteins of a size comparable to *DISC1*.

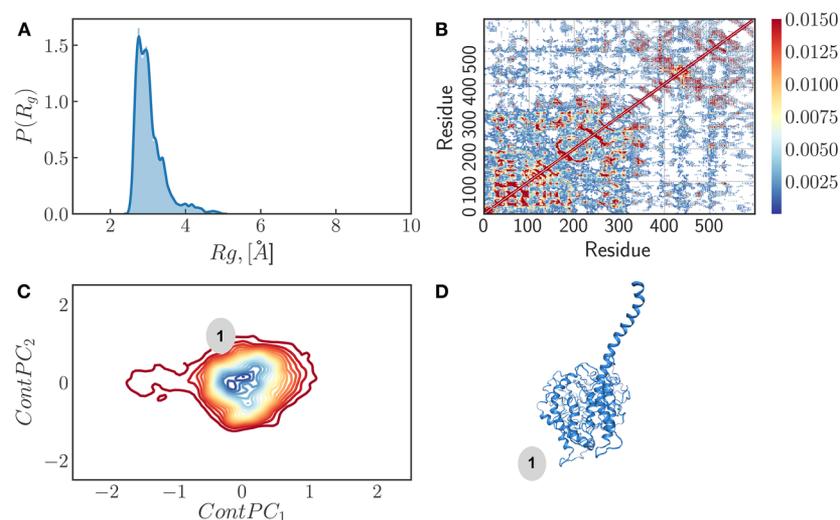
For *de novo* prediction of dimers, we have initiated monomers in parallel and antiparallel orientation and carried out short simulated annealing runs followed by constant temperature simulations at  $T = 300$  K. Umbrella sampling

simulations were initiated using conformations generated from monomer and C- and N-terminal fragment simulations. We have considered both parallel and antiparallel orientations. Center of mass harmonic restraint  $U_i(R_i) = k_{com}(R_{com} - \langle R_i \rangle)^2$  for umbrella windows  $i = 1–50$  have been applied to the groups of residues spanning the central regions of the *DISC1* (Figure 1) in the case of full-length monomer and all residues in the case of N and C-terminal fragments. We used  $k_{com} \sim 0.12$  kcal/Å<sup>2</sup> for spring constants and 50,  $\Delta\langle R_i \rangle = 2$  Å separated windows which were generated from the bound states of fragments. The weighted histogram analysis method (WHAM) has been used for reconstructing the unbiased estimates of free energy  $F(R_{com})$  as a function of the center of mass coordinate  $R_{com}$ .

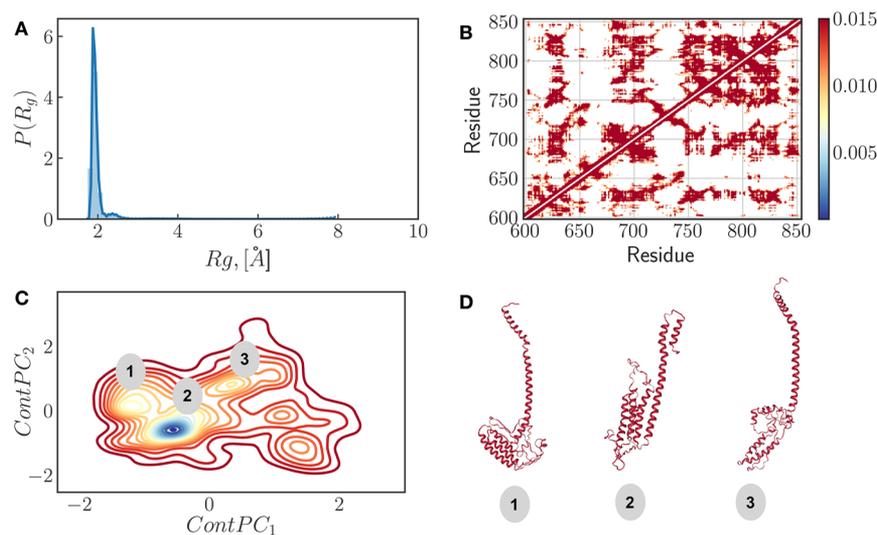
For the contact analysis, we have considered  $C_{\alpha}$  atoms by excluding contacts between  $(i, j)$  pairs of residues that are separated by  $|i - j| > 3$  in 1D sequence space. We have used distances satisfying  $r_{ij} \leq 8$  Å condition in 3D space as the definition of contacts. Employing a properly filtered definition of contacts, we have carried out a principal component analysis in the space of physical contacts which reveals nonlocal contact making and breaking in conformational ensembles of monomers and oligomers. Examining the first few principal components reveals the interfacial hot spots which trigger most variable contact changes and are likely driving forces for slowest conformational transitions.<sup>31,37,38</sup>

## RESULTS AND DISCUSSION

**In Its Monomeric Form *DISC1* Is in Equilibrium between Extended and Compact States.** To fully describe the mechanisms of oligomerization of *DISC1*, we first examined the conformational ensemble populated by the full-length monomeric protein. We achieved an extensive sampling of conformations through 20 independent simulations that were initiated by performing simulated annealing runs followed by more than 200  $\mu$ s simulations at constant temperature  $T = 300$  K (Figure S2). The conformational ensemble generated shows a heterogeneous distribution of mostly helical structures, with significant structural disorder localized within N-terminal



**Figure 3.** (A) Distribution of radii of gyration for N-terminal fragment unit of *DISC1*. (B) Contact frequency map showing the  $C_{\alpha}$ - $C_{\alpha}$  contacts that are observed more frequently in simulations. (C) Probability distribution of simulated trajectories onto principal components in the space of physical  $C_{\alpha}$ - $C_{\alpha}$  contacts that correspond to high contact frequencies. (D) Representative structures from simulations corresponding to distinct regions of the first two principal components.



**Figure 4.** (A) Distribution of radii of gyration for C-terminal fragment unit of *DISC1*. (B) Contact frequency map showing the  $C_{\alpha}$ - $C_{\alpha}$  contacts that are observed more frequently in simulations. (C) Probability distribution of simulated trajectories onto principal components in the space of physical  $C_{\alpha}$ - $C_{\alpha}$  contacts that correspond to high contact frequencies. (D) Representative structures from simulations corresponding to distinct regions of the first two principal components.

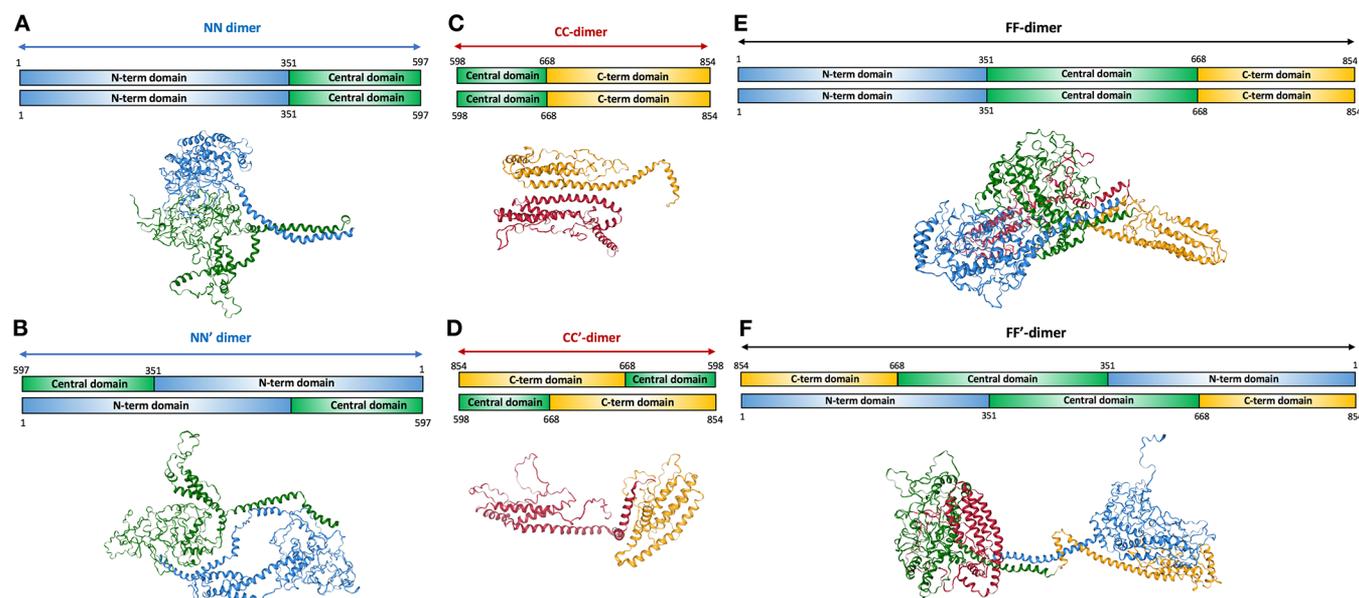
domain (Figure 1,2), in good agreement with bioinformatic secondary structure predictions (Figure S1). In addition, the simulations show that the C-termini of the generated conformations form bundles of short helices, in good agreement with the recently reported solution structure of the C-terminal fragment<sup>39,40</sup> (aa. 765–835)

To characterize and disentangle global conformational heterogeneity and local structural disorder in *DISC1* ensemble, we employed two separate measures; the radius of gyration and various contact-metrics in the form of contact frequency map, contact-based PCA (Figures 2 and 3) in addition to standard quantification of local per residue-based secondary structure elements (Figure S2).

In the absence of well-defined 3D structures, the radius of gyration has been a central quantity for characterizing the extent of structural disorder in proteins. However, in situations where there is significant secondary structural content present,

global measures of disorder such as radius of gyration, end-to-end distance, and structure factors can be less informative.<sup>41,42</sup> Contact-based measures, on the other hand, have the advantage of revealing a more detailed picture of disorder throughout the sequence as well as revealing dynamic and thermodynamic driving forces that shape energy landscapes of disordered proteins.<sup>31,37,38</sup>

The distribution of radius of gyration clearly shows the presence of at least three major populations, with an average radius of gyrations of 3.5, 4.6, and 8.2 Å (Figure 2A). Analyzing frequency of contact formation in *DISC1* ensemble we found a specific pattern of contacts within the N- and C-terminal domains while the central domain shows largely smeared contacts with the N-terminal domain (Figure 2B). We then performed a principal component analysis (PCA) of the distances of statistically significant contact between residue pairs, filtered to exclude noisy transient contacts via condition



**Figure 5.** Various combinations of dimeric constructs studied here, with (A) the parallel N-terminal fragment dimer (NN), (B) antiparallel fragment dimer (NN'), (C) parallel C-terminal fragment dimer (CC), antiparallel C-terminal fragment (CC'), (E) parallel full-length dimer (FF), and (F) antiparallel full-length dimer (FF').

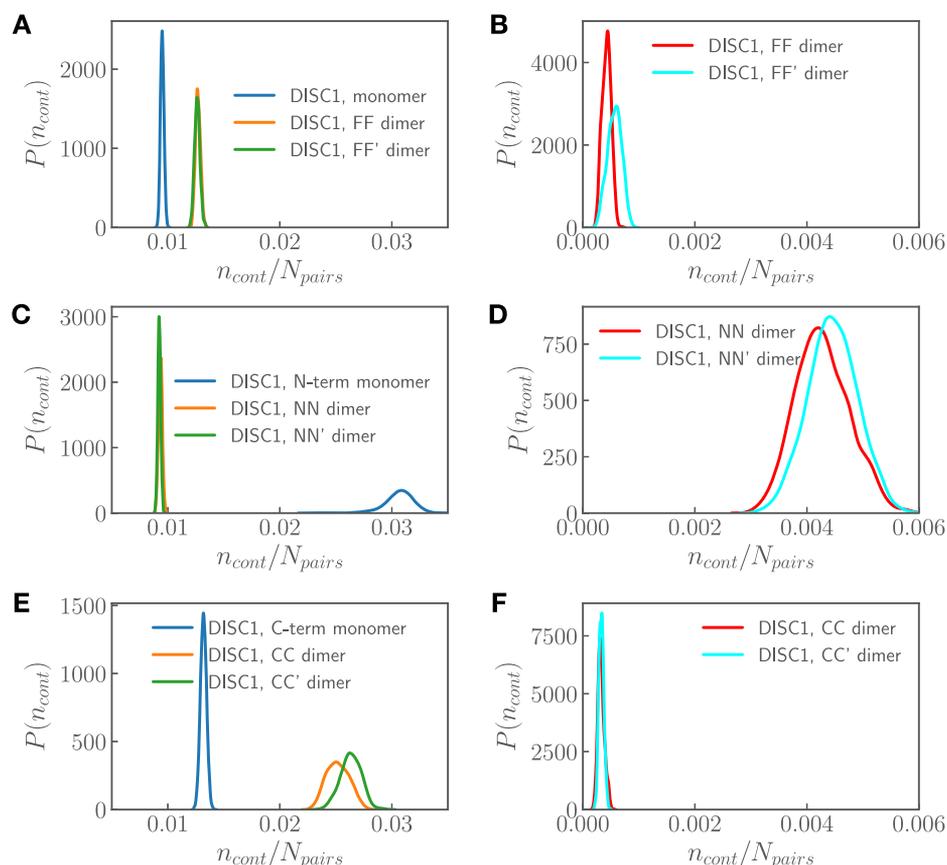
imposed on probability density  $p > 0.001$ , identified in the frequency map. The first principal components identify sets of contacts which contribute the most to the variance, and these pairs typically correlate with more global structural changes. The probability distribution of conformational ensemble onto the first two contact principal modes shows indeed that the high variance contacts between the central and N-terminal domains are significant drivers of observed conformational heterogeneity (Figure 2C). In good agreement with the distribution of radius of gyration, we found with the principal component analysis three major populations within the conformational ensemble of *DISC1*, corresponding to compact, semicompact, and extended states (labeled 1, 2, and 3 respectively in Figure 2, parts C and D).

**The N- and C-Terminal Fragments Are Conformationally Less Heterogeneous than the Full-Length *DISC1* but Not Less Disordered.** Examining the N-terminal fragment of *DISC1* (residue 1–597, see Figure 1C) we found drastically different conformational behavior. The radius of gyration distribution shows unimodal distribution largely in this case as opposed to well-defined multimodal features of full-length *DISC1* (Figure 3A). The decrease in conformational heterogeneity, however, does not imply a decrease in structural disorder. We observed indeed an increase in local disorder, which is likely driven by the absence of contacts formed between the N-terminal and central domains (Figure 3B). The principal component analysis of statistically significant contact-making residue pairs, filtered to exclude noisy transient contacts via condition  $p > 0.001$ , revealed that contact making and breaking do not drive any large scale coherent structural changes (Figure 3C). Instead, contact-making and breaking are highly delocalized and irregular as one would expect from protein with a significant amount of disorder. Overall, the conformational ensemble of the N-terminal fragment is relatively homogeneous, with a single major population corresponding to a compact and largely disordered state (Figure 3D).

We then examined the C-terminal fragment (residue 598–854, see Figure 1C). We found that the conformational features of this fragment are significantly different from that of the full-length *DISC1* and N-terminal fragment. In particular, we observed much less structural disorder relative to the N-terminal fragment, although the radius of gyration distribution shows a non-negligible degree of conformational heterogeneity (Figure 4A). This heterogeneity appears to be driven by frustration to form bundled  $\alpha$ -helices versus the conformational entropy favored elongated forms. Analysis of contact frequency maps shows that the conformational ensemble is populated with relatively rigid local  $\alpha$ -helical segments that can adopt multiple different global folds due to the elongated nature of the chain (Figure 4B). Principal component analysis of statistically significant contacts  $p > 0.001$  reveals that indeed the most variant set of contacts drive large scale structural changes through various sets of contact making and breaking (Figure 4C). On the basis of this analysis, the C-terminal fragment seems to populate a heterogeneous ensemble of helical bundles, within its most compact form, up to four short helices (Figure 4D).

Overall, the conformational ensemble of the N-terminal and C-terminal fragments are significantly different from that of the full-length *DISC1*. The significant difference between N and C terminal contact patterns is suggesting that interdomain contacts participate in the structural heterogeneity of the full-length protein. This interplay between structural heterogeneity on a local and global scale also hints at the large conformational space accessible to dimeric structures.

**Both Conformational Heterogeneity and Structural Disorder Mediate the Oligomerization of *DISC1*.** Next, we analyze various dimeric combinations of *DISC1* and investigate how conformational heterogeneity and local structural disorder mediate the oligomerization properties of *DISC1*. Disordered regions in proteins are usually viewed as transient states that can facilitate protein–protein interfaces upon binding.<sup>43</sup> The discovery of the widespread presence of fuzzy complexes,<sup>44–46</sup> however, has shown that the spectrum



**Figure 6.** Distributions of  $C_{\alpha}$ – $C_{\alpha}$  contacts evaluated for (A) intrachain contacts of full length *DISC1* monomer and dimer units, (B) interchain contacts for full length *DISC1* dimer units, (C) intrachain contacts of N-terminal *DISC1* monomer and dimer units. (D) interchain contacts for *DISC1* N-terminal fragment dimer units, (E) intrachain contacts of C-terminal *DISC1* monomer and dimer units, and (F) interchain contacts for *DISC1* C-terminal fragment dimer units.

of action for disordered proteins is much broader.<sup>47,48</sup> Many cases have been uncovered where structural disorder is not only maintained but even generated through protein–protein and protein–nucleic acid interactions.<sup>49–51</sup> With *DISC1*, we found a clear case supporting the role of both conformational heterogeneity and local structural disorder in facilitating the formation of oligomerization interfaces.

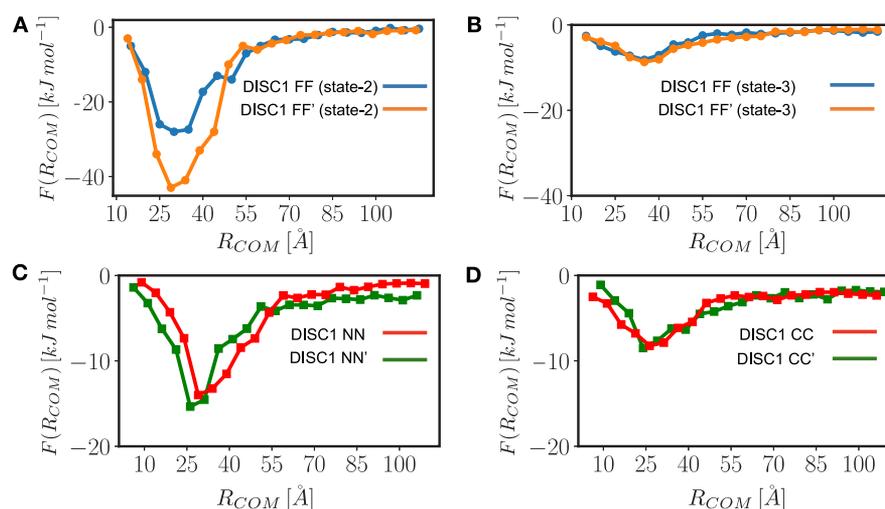
To shed light on the complexities of *DISC1* oligomerization pathways reported by experiments, we carried out two kinds of simulations aimed at (i) predicting a structural model of the full-length and truncated dimers of *DISC1* and (ii) sampling the free energy landscape associated with these oligomerization pathways. For the structural prediction part, we generated two sets of initial 3d models for each construct, representing the parallel and antiparallel orientation of the subunits (Figure 5).

To generate an ensemble of dimer states, we first ran short time simulated annealing steps followed by long time scale constant temperature molecular dynamics simulations. We have found that the dimeric constructs initiated from extended conformations of the full-length *DISC1* lead to both extended and collapsed dimers corresponding to parallel and antiparallel orientations, respectively. Furthermore, dimerization is accompanied by large scale rearrangement of intra- and interdomain contacts (Figure S4). Dimerization of the N- and C-terminal fragments are also accompanied by large-scale structural rearrangements. We quantified the interplay between intra- and interdomain contacts by computing the fraction of contacts formed within and between monomeric versus

dimeric units. The fraction of contacts is evaluated by computing the ratio  $n_{cont}(t)/N_{pairs}$ , where  $n_{cont}(t)$  is the number of  $C_{\alpha}$ – $C_{\alpha}$  distances which satisfy the  $r_{ij} \leq 8 \text{ \AA}$  condition for every state in the sampled conformational ensemble indexed by time coordinate  $t$  (multiple equilibrium trajectories have been combined together). The  $N_{pairs}$  is the total number of  $C_{\alpha}$ – $C_{\alpha}$  pairs, which provides a normalization measure for comparing fragments vs monomers and dimers (Figure 6).

Examining the intradomain contact distributions for the full-length *DISC1*, we found that dimerization results in an enhancement of contact formation within each subunit relative to the monomeric form (Figure 6A). This can be explained by the fact that conformational ensemble of *DISC1* monomer includes various collapsed, exposed and intermediate forms which interconvert via extensive contact-making/breaking transitions whereas, in the dimeric form, subunits are locked into one major conformational state. We observed no significant difference in terms of intradomain contact distribution between parallel (FF) and antiparallel dimeric forms (FF'). The distribution of interdomain contacts (Figure 6B), however, reveals that dimers with antiparallel orientation (FF') form more contacts, due to their collapsed nature, than dimer with parallel orientation (FF'). These results suggest that the flexible nature of the *DISC1* manifests not only in conformational heterogeneity at the monomeric level but also in the dimeric form.

In the case of the dimers formed by the N-terminal fragment of *DISC1* (NN and NN'), we found that dimerization results



**Figure 7.** Free energy profiles  $F_{com}$  of binding/unbinding of the different *DISC1* dimers with parallel/antiparallel orientations as a function of center of mass distance between monomeric units  $R_{com}$ . (A) Free energy of dimerization with parallel and antiparallel orientations of full length sequence. (B) Free energy profile of binding between collapsed states of the full length sequence. (C) Free energy profile for N-terminal fragment dimerization with parallel and antiparallel orientation. (D) Free energy profile for C-terminal fragment dimerization with parallel and antiparallel orientation.

in a loss of intradomain contacts relative to the free monomeric unit (Figure 6C and Figure S5). This suggests that the formation of the dimer interface in the N-terminal region is the result of a trade-off between inter- and intradomain contacts. We observed no significant difference in terms of intradomain contact distribution between the parallel (NN) and antiparallel dimeric forms (NN'). The distribution of interdomain contacts (Figure 6D), however, reveals that dimers with an antiparallel orientation of N-terminal domain (NN') are forming a more extensive network of contacts relative to its parallel counterpart (NN). These observations suggest that structural disorder in the N-terminal region is a significant driver of oligomerization by promoting the formation of an extensive network of interdomain contacts. To ensure that our findings regarding these subtle differences of contact distributions between NN and FF dimers shown in panels B, D, and F are statistically significant we have subjected the distributions depicted in (Figure 6BD) to the *t* test, finding  $p \ll 0.001$ , which allows us to reject the null hypothesis with high confidence.

For dimers formed by the C-terminal fragment of *DISC1* (CC and CC'), we found that the distribution of intradomain contacts is similar to that of the full-length *DISC1* (Figure 6E and Figure S6), suggesting that dimerization of the C-terminal region results in a large gain of intradomain contacts, which is enough to offset the loss of contacts by the N-terminal fragment upon the formation of full-length dimers. The distinction in terms of intradomain contact distribution between parallel (CC) and antiparallel dimeric forms (CC') appears again negligible. The extent of interdomain contacts (Figure 6F) is notably smaller than that of the N-terminal fragment and full-length *DISC1* (keeping in mind that  $N_{pairs} \sim L^2$  with  $L$  being sequence length, see also Figure S7) which can be rationalized by the absence of structural disorder in the C-terminal region.

Finally, to characterize the oligomerization propensity of the various dimeric conformational states described above, we have computed the free energy profiles of dimerization (Figure 7) for each putative dimer by carrying out umbrella sampling

simulations with a harmonic bias applied to both subunits. The free energy profiles obtained for the full dimer initiated from the extended monomeric state (structure 2 in Figure 2D) show that the parallel (FF) and antiparallel conformations (FF') have drastically different propensities for dimerization (Figure 7A). For reference, we have also calculated the free energy profiles of association dimers initiated from the collapsed monomeric state (structure 3 in Figure 2D). We observed that such dimers maintain their collapsed states, suggesting that in the absence of exposed dimerization surfaces, the free energy gain is significantly diminished (Figure 7B). Such a large free energy gap between collapsed and extended dimeric forms also suggests that both conformational heterogeneity and structural disorder, which modulated the extent of exposed surface area, are important for the formation of stable dimers.

Examining the free energy profiles of dimerization obtained for the N-terminal fragments of *DISC1* with parallel (NN) and antiparallel orientations (NN'), we found no significant difference free energy gap between both dimeric forms (Figure 7C). This could be rationalized by the absence of the central and C-terminal domains which provide most of the conformational heterogeneity observed for the full-length protein (Figure 2). These results suggest that structural disorder by itself (i.e., without conformational heterogeneity) is enough to distinguish the parallel and antiparallel dimer conformations.

Similarly, the free energy profiles of dimerization computed for the C-terminal fragments of *DISC1* shows no significant difference between dimers with parallel (CC) and antiparallel orientations (CC') (Figure 7D), which again suggest that both structural disorder and conformational heterogeneity are required to form stable *DISC1* dimers.

## CONCLUSIONS

In this paper, we present a detailed molecular level study of the behavior of *DISC1* protein, which is believed to rely on a disorder to act as a major hub for regulatory pathways connected with neural development. Specifically, we have shed light on the role of conformational heterogeneity and local structural disorder in the oligomerization pathway of the full-

length *DISC1* and its N- and C-terminal fragments. In this paper, we make a point to emphasize this distinction between local and global disorder because often in literature disorder is meant to imply only one type: the absence of secondary structural elements. For the example of *DISC1*, however, experiments<sup>39,40</sup> and our simulations have found that the protein chain can also be conformationally flexible in solution due to its elongated shape. Therefore, we quantified the degree of local disorder arising from the lack of secondary structure but also the degree of global disorder arising from the elongated nature of the protein, which is entropically driven to sample a myriad of conformations.

By carrying out extensive sampling with coarse-grained molecular dynamics simulations using the predictive energy landscape based protein force field AWSEM, we have uncovered a complex interplay between intra- and interdomain contact frustration which shapes the conformational energy landscapes of *DISC1* dimers.

We found that the interplay between intra- and interdomain contacts drive the monomeric subunits of *DISC1* into adopting a heterogeneous conformational ensemble with collapsed and exposed surfaces. We then go on to show that conformational heterogeneity and disorder translate into disparate binding pathways leading to the formation of collapsed, transiently bound, and exposed dimeric structures. We found that collapsed dimeric states have marginal oligomerization affinity, whereas exposed states lead to the formation of stable dimers. By analyzing the N and C-terminal fragments, we uncover that the structural disorder localized in the N-terminal region is a significant driver of dimerization. The C-terminal region, on the other hand, is responsible for generating conformational heterogeneity and entropically favored exposed states, which have a higher propensity for forming dimers and most likely serve as stepping stones for the formation of higher-level oligomers. Our simulation also predicts the existence of collapsed dimeric states, which may act deep kinetic traps and serve as a powerful regulatory mechanism for controlling the oligomerization pathways and ultimately the function of *DISC1*.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jpcc.9b07467](https://doi.org/10.1021/acs.jpcc.9b07467).

Methods used and additional analysis employed (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [potoyan@iastate.edu](mailto:potoyan@iastate.edu).

### ORCID

Julien Roche: [0000-0003-1254-1173](https://orcid.org/0000-0003-1254-1173)

Davit A. Potoyan: [0000-0002-5860-1699](https://orcid.org/0000-0002-5860-1699)

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

J.R. acknowledges the Iowa State University College of Liberal Arts and Sciences and the Roy J. Carver Charitable Trust for their support. D.A.P. acknowledges financial support from the Caldwell Foundation of Iowa State.

## ■ REFERENCES

- (1) Balazs, A.; Csizmok, V.; Buday, L.; Rakács, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. High levels of structural disorder in scaffold proteins asexemplified by a novel neuronal protein, CASK-interactive protein1. *FEBS J.* **2009**, *276*, 3744–3756.
- (2) Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Intrinsic disorder in scaffold proteins: getting more from less. *Prog. Biophys. Mol. Biol.* **2008**, *98*, 85–106.
- (3) Millar, J. K.; Wilason-Annan, J. C.; Anderson, S.; Christie, S.; Taylor, M. S.; Semple, C. A.; Devon, R. S.; Clar, D. M. S.; Muir, W. J.; Blackwood, D. H.; et al. Disruption of two novel gene by a translocation co-segregating with schizophrenia. *Hum. Mol. Genet.* **2000**, *9*, 1415–1423.
- (4) Millar, J. K.; Christie, S.; Anderson, S.; Lawson, D.; Loh, D. H.-W.; Devon, R. S.; Arveiler, B.; Muir, W. J.; Blackwood, D. H.; Porteous, D. J. Genomic structure and localisation within a linkage hotspot of Disrupted in Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia. *Mol. Psychiatry* **2001**, *6*, 173–178.
- (5) Tomppo, L.; Hennah, W.; Miettunen, J.; Jarvelin, M. R.; Veijola, J.; Ripatti, S.; Lahermo, P.; Lichtermann, D.; Peltonen, L.; Ekelund, J. Association of variants in *DISC1* with psychosis-related traits in a large population cohort. *Arch. Gen. Psychiatry* **2009**, *66*, 134–141.
- (6) Hotta, Y.; Ohnuma, T.; Hanzawa, R.; Shibata, N.; Maeshima, H.; Baba, H.; Hatano, T.; Takebayashi, Y.; Kitazawa, M.; Higa, M.; et al. Association study between Disrupted in-Schizophrenia-1 (*DISC1*) and Japanese patients with treatment-resistant schizophrenia (TRS). *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **2011**, *35*, 636–639.
- (7) Tomppo, L.; Ekelund, J.; Lichtermann, D.; Veijola, J.; Jarvelin, M. R.; Hennah, W. *DISC1* conditioned GWAS for psychosis proneness in a large Finnish birth cohort. *PLoS One* **2012**, *7*, No. e30643.
- (8) Porteous, D. J.; Thomson, P. A.; Millar, J. K.; Evans, K. L.; Hennah, W.; Soares, D. C.; McCarthy, S.; McCombie, W. R.; Clapcote, S. J.; Korth, C.; et al. *DISC1* as a genetic risk factor for schizophrenia and related major mental illness: response to Sullivan. *Mol. Psychiatry* **2014**, *19*, 141–143.
- (9) Mouaffak, F.; Kebir, O.; Chayet, M.; Tordjman, S.; Vacheron, M. N.; Millet, B.; Jaafari, N.; Bellon, A.; Olie, J. P.; Krebs, M. O. Association of Disrupted in Schizophrenia (*DISC1*) missense variants with ultra-resistant schizophrenia. *Pharmacogenomics J.* **2011**, *11*, 267–273.
- (10) Teng, S.; Thomson, P. A.; McCarthy, S.; Kramer, M.; Muller, S.; Lihm, J.; Morris, S.; Soares, D. C.; Hennah, W.; Harris, S.; et al. Rare disruptive variants in the *DISC1* interactome and regulome: association with cognitive ability and schizophrenia. *Mol. Psychiatry* **2018**, *23*, 1270–1277.
- (11) Camargo, L. M.; Collura, V.; Rain, J.-C.; Mizuguchi, K.; Hermjakob, H.; Kerrien, S.; Bonnert, T. P.; Whiting, P. J.; Brandon, N. J. Disrupted in Schizophrenia 1 interactome: evidence for the close connectivity of risk gene and potential synaptic basis for schizophrenia. *Mol. Psychiatry* **2007**, *12*, 74–86.
- (12) Bradshaw, N. J.; Porteous, D. J. *DISC1*-binding proteins in neural development, signaling and schizophrenia. *Neuropharmacology* **2012**, *62*, 1230–1241.
- (13) Lipina, T. V.; Roder, J. C. Disrupted-In-Schizophrenia-1 (*DISC1*) interactome and mental disorders: impacts of mouse models. *Neurosci. Biobehav. Rev.* **2014**, *45*, 271–294.
- (14) Soares, D. C.; Carlyle; Bradshaw, N. J.; Porteous, D. J. *DISC1*: structure, function and therapeutic potential for major mental illness. *ACS Chem. Neurosci.* **2011**, *2*, 609–632.
- (15) Yerabham, A. S.; Mas, P. J.; Decker, C.; Soares, D. C.; Weiergräber, O. H.; Nagel Steger, L.; Willbold, D.; Hart, D. J.; Bradshaw, N. J.; Korth, C. A structural organization for the Disrupted in Schizophrenia 1 protein, identified by high-throughput screening, reveals distinctly folded regions, which are bisected by mental illness-related mutations. *J. Biol. Chem.* **2017**, *292*, 6468–6477.
- (16) Sanchez-Pulido, L.; Ponting, C. P. Structure and evolutionary history of *DISC1*. *Hum. Mol. Genet.* **2011**, *20*, R175–R181.

- (17) Ye, F.; Kang, E.; Yu, C.; Jacob, F.; Yu, C.; Mao, M.; Poon, R. Y. C.; Kim, J.; Song, H.; Ming, G.-L.; et al. DISC1 regulates neurogenesis via modulating kinetochore attachment of Ndel1/Nde1 during mitosis. *Neuron* **2017**, *96*, 1204.
- (18) Yerabham, A. S. K.; Mas, P. J.; Decker, C.; Soares, D. C.; Weiergraber, O. H.; Nagel Steger, L.; Willbold, D.; Hart, D. J.; Bradshaw, N. J.; Korth, C. A structural organization for the Disrupted in Schizophrenia 1 protein, identified by high-throughput screening, reveals distinctly folded regions, which are bisected by mental illness-related mutations. *J. Biol. Chem.* **2017**, *292*, 6468–6477.
- (19) Narayanan, S.; Arthanari, H.; Wolfe, M. S.; Wagner, G. Molecular characterization of disrupted in schizophrenia-1 risk variant S704C reveals the formation of altered oligomeric assembly. *J. Biol. Chem.* **2011**, *286*, 44266–44276.
- (20) Leliveld, R. S.; Bader, V.; Hendriks, P.; Prikulis, I.; Sajjani, G.; Requena, J. R. R.; Korth, C. Insolubility of Disrupted-In-Schizophrenia 1 disrupts oligomer-dependent interactions with nuclear distribution element 1 and is associated with sporadic mental disease. *J. Neurosci.* **2008**, *28*, 3839–3845.
- (21) Leliveld, R. S.; Hendriks, P.; Michel, M.; Sajjani, G.; Bader, V.; Trossbach, S.; Prikulis, I.; Hartmann, R.; Jonas, E.; Willbold, D.; et al. Oligomer assembly of the C-terminal DISC1 domain (640–854) is controlled by self-association motifs and disease associated polymorphism S704C. *Biochemistry* **2009**, *48*, 7746–7755.
- (22) Ishizuka, K.; Kamiya, A.; Oh, E. C.; Kanki, H.; Seshadri, S.; Robinson, J. F.; Murdoch, H.; Dunlop, A. J.; Kubo, K.; Furukori, K.; et al. DISC1-dependent switch from progenitor proliferation to migration in the developing cortex. *Nature* **2011**, *473*, 92–96.
- (23) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858.
- (24) Xu, D.; Zhang, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **2011**, *101*, 2525–2534.
- (25) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503.
- (26) Papoian, G. A.; Wolynes, P. G. AWSEM-MD. *Coarse-Grained Modeling of Biomolecules* **2017**, 121–190.
- (27) Schafer, N. P.; Kim, B. L.; Zheng, W.; Wolynes, P. G. Learning to fold proteins using energy landscape theory. *Isr. J. Chem.* **2014**, *54*, 1311–1337.
- (28) Lin, X.; Schafer, N. P.; Lu, W.; Jin, S.; Chen, X.; Chen, M.; Onuchic, J. N.; Wolynes, P. G. Forging tools for refining predicted protein structures. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 9400–9409.
- (29) Zheng, W.; Tsai, M.-Y.; Wolynes, P. G. Comparing the aggregation free energy landscapes of amyloid Beta (1–42) and amyloid Beta (1–40). *J. Am. Chem. Soc.* **2017**, *139*, 16666–16676.
- (30) Chen, M.; Tsai, M.; Zheng, W.; Wolynes, P. G. The aggregation free energy landscapes of polyglutamine repeats. *J. Am. Chem. Soc.* **2016**, *138*, 15197–15203.
- (31) Potoyan, D. A.; Bueno, C.; Zheng, W.; Komives, E. A.; Wolynes, P. G. Resolving the NF $\kappa$ B heterodimer binding paradox: strain and frustration guide the binding of dimeric transcription factors. *J. Am. Chem. Soc.* **2017**, *139*, 18558–18566.
- (32) Potoyan, D. A.; Zheng, W.; Komives, E. A.; Wolynes, P. G. Molecular stripping in the NF- $\kappa$ B/I $\kappa$ B/DNA genetic regulatory network. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 110–115.
- (33) Potoyan, D. A.; Zheng, W.; Ferreira, D. U.; Wolynes, P. G.; Komives, E. A. PEST control of molecular stripping of NF $\kappa$ B from DNA transcription sites. *J. Phys. Chem. B* **2016**, *120*, 8532–8538.
- (34) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: a coarse-grained force field for intrinsically disordered proteins. *J. Phys. Chem. B* **2018**, *122*, 11115–11125.
- (35) Tsai, M.-Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G. Electrostatics, structure prediction, and the energy landscapes for protein folding and binding. *Protein Sci.* **2016**, *25*, 255–269.
- (36) Chen, M.; Lin, X.; Zheng, W.; Onuchic, J. N.; Wolynes, P. G. Protein folding and structure prediction from the ground up: The atomistic associative memory, water mediated, structure and energy model. *J. Phys. Chem. B* **2016**, *120*, 8557–8565.
- (37) Lindsay, R. J.; Pham, B.; Shen, T.; McCord, R. P. Characterizing the 3D structure and dynamics of chromosomes and proteins in a common contact matrix framework. *Nucleic Acids Res.* **2018**, *46*, 8143–8152.
- (38) Lindsay, R. J.; Siess, J.; Lohry, D. P.; McGee, T. S.; Ritchie, J. S.; Johnson, Q. R.; Shen, T. Characterizing protein conformations by correlation analysis of coarse-grained contact matrices. *J. Chem. Phys.* **2018**, *148*, No. 025101.
- (39) Ye, F.; Kang, E.; Yu, C.; Qian, X.; Jacob, F.; Yu, C.; Mao, M.; Poon, R. Y.; Kim, J.; Song, H.; et al. DISC1 regulates neurogenesis via modulating kinetochore attachment of Ndel1/Nde1 during mitosis. *Neuron* **2017**, *96*, 1041–1054.
- (40) Wang, X.; Ye, F.; Wen, Z.; Guo, Z.; Yu, C.; Huang, W.-K.; Ringeling, F. R.; Su, Y.; Zheng, W.; Zhou, G.; et al. Structural interaction between DISC1 and ATF4 underlying transcriptional and synaptic dysregulation in an iPSC model of mental disorders. *Mol. Psychiatry* **2019**, 1–15.
- (41) Potoyan, D. A.; Papoian, G. A. Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J. Am. Chem. Soc.* **2011**, *133*, 7405–7415.
- (42) Winogradoff, D.; Echeverria, I.; Potoyan, D. A.; Papoian, G. A. The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects. *J. Am. Chem. Soc.* **2015**, *137*, 6245–6253.
- (43) Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **2014**, *114*, 6561–6588.
- (44) Fuxreiter, M. Fuzziness in protein interactions: A historical perspective. *J. Mol. Biol.* **2018**, *430*, 2278–2287.
- (45) Miskei, M.; Antal, C.; Fuxreiter, M. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher order assemblies. *Nucleic Acids Res.* **2017**, *45*, D228–D235.
- (46) Fuxreiter, M. Fuzziness: linking regulation to protein dynamics. *Mol. Biosyst.* **2012**, *8*, 168–177.
- (47) Uversky, V. N. Intrinsic Disorder, Protein–Protein Interactions, and Disease. *Adv. Protein Chem. Struct. Biol.* **2018**, *110*, 85–121.
- (48) Fonin, A. V.; Darling, A. L.; Kuznetsova, I. M.; Turoverov, K. K.; Uversky, V. N. Intrinsically disordered proteins in crowded milieu: When chaos prevails within the cellular gumbo. *Cell. Mol. Life Sci.* **2018**, *75*, 3907–3929.
- (49) Uversky, V. N. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* **2011**, *40*, 1623–1634.
- (50) Gurevich, V. V.; Gurevich, E. V.; Uversky, V. N. Arrestins: structural disorder creates rich functionality. *Protein Cell* **2018**, *9*, 986–1003.
- (51) Uversky, V. N. Intrinsic disorder-based protein interactions and their modulators. *Curr. Pharm. Des.* **2013**, *19*, 4191–4213.