Beta Turn Propensity and Polymer Scaling Exponent Identify Intrinsically Disordered Proteins that Phase Separate

Elisia A. Paiz,[1] John J. Correia,[2] Nicholas C. Fitzkee,[3] Loren E. Hough,[4,5] and Steven T. Whitten[1]*

[1] Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666, United States
[2] Department of Cell and Molecular Biology, University of Mississippi Medical Center, Jackson, MS 39211, United States
[3] Department of Chemistry, Mississippi State University, Mississippi State, Mississippi 39762, United States
[4] Department of Physics, University of Colorado Boulder, Boulder CO 80309, United States
[5] BioFrontiers Institute, University of Colorado Boulder, Boulder CO 80309, United States

Abstract

The complex cellular milieu can spontaneously de-mix in a process driven in part by proteins that are intrinsically disordered (ID). We hypothesized that protein self-interactions that determine the polymer scaling exponent, ν, of monomeric ID proteins (IDPs), also facilitate de-mixing transitions into phase separated assemblies. We analyzed a protein database containing subsets that are folded, ID, or IDPs identified previously to spontaneously phase separate. We found that the subsets differentiated into distinct protein classes according to sequence-based calculations of ν and, surprisingly, the propensity in the sequence for adopting the β-turn. Structure-based simulations find that transient β-turn structures reduce the desolvation penalty of forming a protein-rich phase. By this mechanism, β-turns act as energetically favored nucleation points, which may explain the increased propensity for turns in IDPs that are utilized biologically for phase separation.

Introduction

Protein liquid-liquid phase separation (LLPS) is increasingly recognized as an important organizing phenomenon in cells. LLPS is a reversible process whereby complex protein mixtures spontaneously de-mix into liquid droplets that are enriched in a target protein; concomitantly, surrounding regions are depleted of that protein (Lee et al., 2013). The de-mixing transition is thought to provide temporal and spatial control over intracellular interactions by assembling collections of proteins into structures called membraneless organelles (Mitrea and Kriwacki, 2016), a key step in the regulatory function of P bodies, the nucleolus, and germ granules (Brady et al., 2017; Elbaum-Garfinkle et al., 2015; Mitrea et al., 2018). The physical mechanisms responsible for LLPS are not fully understood, but it is known to be facilitated primarily by proteins that are intrinsically disordered or that contain large intrinsically disordered regions (Uversky et al., 2015; Chong and Forman-Kay, 2016; Mitrea and Kriwacki, 2016).

Among intrinsically disordered regions (IDRs), the hydrodynamic size has been found to vary substantially with the primary sequence (Marsh and Forman-Kay, 2010; Tomasso et al., 2016). The hydrodynamic dimensions of the ensembles of disordered structures that exist in solution for biological proteins have been closely scrutinized (Uversky, 2002; Hofmann et al., 2012; Wuttke et al., 2014; English et al., 2019) and appears important for the biological function of IDRs. For example, some ID proteins (IDPs) regulate the remodeling of cellular membranes, and their size controls curvature at membrane surfaces (Snead et al., 2019; Zeno et al., 2019). With the increased interest in understanding the functional roles of IDPs and IDRs, the hydrodynamic size has also been used to extract information on the balance of intra- and intermolecular interactions for a protein and solvent combination (Wilkins et al., 1999; Uversky, 2002; Marsh and Forman-Kay, 2010). This is achieved by the polymer scaling exponent, ν, obtained from the dependence of size (e.g., hydrodynamic radius, R_h, or radius of gyration, R_g) on polymer length, N, in the power law relationship $R_h \propto N^{\nu}$. For $\nu \sim 0.3$, self-interactions prevail, meaning that protein chains have higher affinity for themselves than for the solvent (Wilkins et al., 1999). For $\nu \sim 0.6$, interactions between chain and solvent dominate (Kohn et al., 2004). For $\nu \sim 0.5$, the self-interactions of the polypeptide and its interactions with the solvent are net equal, resulting in a *theta* condition (Hofmann et al., 2012) whereby the chain dimensions depend on polymer properties alone, such as bond angles, lengths, and allowed rotations (Flory, 1949). Under theta conditions, the structural size of the protein matches expectations from a random walk by an ideal chain. Overall, however, as ν increases, the hydrodynamic dimensions of a protein similarly increase, and chain contact with the solvent becomes increasingly prominent. Because self-interactions required for de-mixing could perturb the net balance of intra- and intermolecular interactions for a protein and its solvent, the polymer scaling exponent likewise could be a predictor of LLPS potential among IDPs.

Here, we have tested the hypothesis that the ability of an IDP to phase separate can be detected from its hydrodynamic dimensions in the monomeric state. We conjecture that the self-interactions that facilitate the de-mixing transition will manifest in decreased R_h for IDPs that are competent to phase separate into protein-rich droplets when compared to IDPs that are not. Based on observations that the presence of ID in a protein (Meng et al., 2017) and the hydrodynamic size of an IDP (or IDR fragment) are both highly predictable from the primary sequence (Marsh and Forman-Kay, 2010; English et al., 2019), we show by analysis of a protein database containing subsets that are folded, ID, or identified previously (Vernon et al., 2018) to spontaneously phase separate that these three subsets differentiate into distinct protein classes according to sequence-based calculations of ν and, surprisingly, chain propensity for β-turn structures. Specifically, ν calculated from IDRs of proteins that phase separate were lower when compared to other ID sequences, consistent with our motivating hypothesis, while β-turn propensity was found to increase.

To investigate the structural characteristics that possibly underlie a role for the β-turn in facilitating LLPS, computer simulations of the protein conformational ensemble were used to evaluate the consequences of β-turn enrichment. Simulated ensembles indicate that solvent waters that are associated with disordered regions are fewer in number when the protein adopts a β-turn, since the tight turn removes backbone

hydrogen bonding sites. With fewer attached waters, this predicts a mechanism where transient sampling of the β-turn can reduce the cost of desolvation that occurs with LLPS (Reichheld et al., 2017), allowing β-turns to act as energetically favored nucleation points. Using the recently proposed framework of *stickers* and *spacers* as a molecular model for LLPS (Harmon et al., 2017; Yang et al., 2019; Martin et al., 2020), it is possible that transient β-turn structures could serve as stickers in the ensemble of otherwise randomly configured spacers.

Results

Hydrodynamic Size is Reduced in IDPs that exhibit LLPS (PS IDPs). We hypothesize that LLPS potential among IDPs can be predicted from the protein hydrodynamic size. To demonstrate the concept that protein self-interactions reduce R_h, and thus also reduce ν, **Figure 1** shows mean R_h that were measured for folded, chemically denatured, and IDPs (**Fig. 1B**). As expected, since IDPs are unfolded and known to have relatively few tertiary contacts, IDPs have larger experimental R_h when compared to folded proteins of similar N and, as a group, they have larger ν. Moreover, the hydrodynamic dimensions of IDPs are similar in magnitude to the non-globular and large sizes observed for chemically denatured proteins although the IDP measurements were performed under native-like conditions.

Also included in the figure are sequence predicted R_h for the ID regions of proteins that have been verified to exhibit phase separation behavior (Vernon et al., 2018), referred to hereafter as the PS IDP set. ID regions were predicted from the primary sequences using the GeneSilico MetaDisorder service (Kozlowski and Bujnicki, 2012). For the identified ID regions, R_h was then predicted using the net charge and intrinsic chain propensity for the polyproline II backbone conformation, both known to promote elongated hydrodynamic dimensions in disordered ensembles (Tomasso et al., 2016; English et al., 2017, 2019). Comparing the IDP set of experimental R_h to the set of predicted R_h from PS IDPs, two observations are immediately apparent. First, the PS IDP set does indeed exhibit a reduced slope (i.e., smaller ν) in its trend of size versus sequence length relative to the IDP set. Second, the R_h variance relative to the R_h trend line is larger for IDPs than it is for PS IDPs. This predicts constrained structural properties for the ID regions of proteins that phase separate, presumably to facilitate de-mixing behavior. Both sets yield trend lines that extrapolate to essentially identical y-axis intercepts, indicating similar chain flexibility. This intercept gives the pre-factor, R_o, in the power-law function, $R_h = R_o \cdot N^{\nu}$, and relates to the segmental length of the chain that can be considered as freely joined (Flory, 1969). Hence, the observed difference in ν is not from differences in average chain stiffness between the two IDP sets, but rather likely due to differences in intra-molecular interactions between sidechains.

A key difference between the two IDP sets is that the non-phase separating protein data uses experimental R_h, while the PS IDP set uses sequence predicted values. The methods developed to predict R_h from sequence for IDPs (Marsh and Forman-Kay, 2010; Tomasso et al., 2016; English et al., 2017) were found to be accurate when tested (Perez et al., 2014; English et al., 2019), and the IDPs used here consist primarily of the same IDPs that were used to train the predictions. R_h for these IDPs are thus accurately predicted from sequence (**Supplementary Information, Fig. S1**). Comparing sequence-predicted R_h from both sets, IDPs to PS IDPs, the same conclusions are drawn. Namely, PS IDPs have, as a group, smaller ν and reduced R_h variance. The identity, sequence, and R_h (measured or predicted) for each protein in **Figure 1** is provided in **Supplementary Information, Tables S1-S4**.

Branched Amino Acids are Rare in PS IDPs. Because the structural dimensions of IDPs are strongly correlated to sequence composition (Marsh and Forman-Kay, 2010; Tomasso et al., 2016; English et al., 2017, 2019), the differences in amino acid content of the two IDP sets were evaluated. **Figure 1C** shows the ratio obtained from the percent composition of each amino acid type in the two protein sets. Ratios above 1 represent enrichment in PS IDPs, whereas ratios lower than 1 are depletion. Amino acid types that

are enriched in PS IDPs were recognized by Forman-Kay and coworkers to have either pi-orbitals among the side chain groups (F, N, R, and Y) or highly exposed backbone peptide bonds (G and S), which also have pi-orbitals of bonded sp^2-hybridized atoms, predicting a role for pi-pi interactions in promoting de-mixing (Vernon et al., 2018). The same amino acid enrichment is found again in the analysis here. We find it notable, however, that the branched amino acids, I, L, and V, are the 3 most depleted amino acid types in PS IDPs when compared to the IDP set. The reduced I, L, and V content seems unrelated to abating the presence of bulky, hydrophobic side chains, since the aromatic amino acids, F and W, are not similarly depleted in PS IDPs (Vernon et al., 2018; Martin et al., 2020).

To investigate this issue, we compared the scaling exponent, ν, calculated individually for each protein to the combined percent composition of I, L, and V in the protein sequence. **Figure 2A** shows that IDPs and PS IDPs form separate protein classes according to their dataset averages. The average and standard deviation obtained from these calculations when applied to the sequences of folded proteins show that I, L, and V are not uniquely enriched in IDPs. Rather, in the database, I, L, and V are uniquely depleted in PS IDPs. Also, although the equations that predict R_h (and thus ν) from sequence were developed for IDPs, and not designed for use with folded proteins, it was shown recently that reversing the sequence of a foldable protein can yield an IDP with experimental R_h that is indeed accurately modeled by the IDP-trained predictors of hydrodynamic size (English et al., 2018).

β-turn Propensity is Increased in PS IDPs. Because I, L, and V have low intrinsic propensity for being found in a β-turn, as determined from surveys of stable protein structures (Levitt, 1978), and computer models have indicated a role for the β-turn in promoting liquid-liquid phase separation (Zhang et al., 2018), we compared ν calculated for the different proteins to the intrinsic chain propensity for forming the β-turn, again finding that IDPs and PS IDPs are separate protein classes (**Fig. 2B**). Specifically, the PS IDP set has increased intrinsic chain propensity for the β-turn, more so than both IDPs and folded proteins. Moreover, the PS IDP set is enriched in amino acid types that favor the β-turn and likewise depleted in amino acids types that disfavor the β-turn (**Supplementary Information, Fig. S2**).

Among the IDPs, Aβ(1-40) has the lowest ν and also the lowest β-turn propensity. Easily identified in **Fig. 2B**, this protein is the IDP that resides in the figure region that is mostly represented by folded proteins. Because α-synuclein is located in this figure where the IDP and folded averages meet (showing $\nu = 0.548$, β-turn propensity = 0.998), it could be that sequences with ν that closely match values from folded proteins have a tendency to form irreversible aggregates, like amyloid, if the sequence is not capable of folding and also if the β-turn propensity too low. Both Aβ(1-40) and α-synuclein are amyloid forming proteins (Garzon-Rodriguez et al., 1997; Li et al., 2018). If the β-turn propensity is high, a low value for ν instead indicates an IDP that is capable of LLPS behavior. Amino acid specific propensities for adopting β-turn structures from Levitt (Levitt, 1978) that were used to generate **Figures 2** and **S2** are provided in **Supplementary Information, Table S5**.

β-turn Structures Reduce Chain Associated Solvent Waters. We sought to determine the structural properties of turns that would enable them to form nucleation points in LLPS. An analysis of 1,000 turn and 1,000 non-turn structures was performed on the sequence GVPGVG (Whitten et al., 2008). For the structures containing a turn, a hydrogen bond was introduced between the H_N of V5 and O of V2. This sequence is derived from elastin-like polypeptide (ELP), a protein sequence known to undergo LLPS (Zhang et al., 2006; Lyons et al., 2013, 2014) and one where transient β-turns have been implicated in self-association (Reichheld et al., 2017; Zhang et al., 2018). A variety of structural measurements were taken on each ensemble, and statistical convergence was confirmed by comparing the average values of the first 500 structures to the average over the entire ensemble, using the standard error of the mean to estimate uncertainty. Using this approach, all measurements were statistically identical, suggesting that additional conformations did not alter the measurements beyond the first 500 structures.

Overall, the structures with β-turns are more compact and have lower accessible surface area and hydrophobic accessible surface area (**Table 1**). This makes sense because the hydrogen bond restricts the conformation of turn structures and limits surface exposure. However, because the structures are more compact, β-turn conformations will also tend to associate with fewer solvent waters. This is demonstrated by conditional hydrophobic accessible surface area (CHASA), a concept introduced by Fleming and coworkers (Fleming et al., 2005). In the calculation of CHASA, solvent waters are placed near hydrophilic groups in the protein based on the positions sampled in long molecular dynamics simulations of peptide solvation (Fleming et al., 2005). Sterically allowed waters are kept, and these waters are used to model solvation waters that make hydrogen bonds to the peptide. These additional solvation waters occlude the hydrophobic accessible surface. The rationale is that all hydrogen bonding groups in a disordered peptide will be satisfied, either by making hydrogen bonds to the peptide or to solvent waters, and these waters will alter the hydrophobic surface area.

CHASA is also found to be lower in β-turn versus non-turn ensembles (327 vs. 354 $Å^2$); however, adding the solvent waters occludes less hydrophobic surface area in the β-turn ensemble, meaning that turn conformations will on average be associated with fewer waters than non-turns (**Table 1**). This is supported by the number of sterically accessible waters placed by the CHASA algorithm: 37.1 ± 0.1 waters are associated with the turn structures, whereas 44.4 ± 0.1 waters are associated with the non-turns. If hydrophobic interactions are to be made between peptide segments, these solvent-associated waters must be removed. However, fewer waters must be removed when peptide segments are in β-turn conformations, which could make turn-turn association more favorable.

In addition, the fixed conformations of β-turns expose large contiguous regions of hydrophobic surface area relative to the random conformations (**Fig. 3**). The CHASA-placed solvation waters in GVPGVG are clustered when the peptide is in a β-turn conformation (**Fig. 3A**), exposing contiguous segments of hydrophobic accessible surface area for residues V2 and P3. Representative structures of non-turn conformations (**Fig. 3B**) do not exhibit these clusters. It is likely that unconstrained, non-turn conformations present a moving target for chain-chain interactions, making it more difficult for inter- and intra-chain association to occur. On the other hand, two fixed β-turns can associate and bury a large relative fraction of hydrophobic accessible surface area (110 $Å^2$ per turn; **Fig. 4**).

Discussion

For flexible chains, the size of the polymer scales with the number of subunits as $R_h \sim R_o \cdot N^{\nu}$. While R_o is determined by chain stiffness, ν is determined primarily by the relative strength of interactions of polymer subunits with either other subunits or the solvent. When subunit-solvent interactions are more favorable relative to subunit-subunit interactions, the chain is expanded. We have previously developed an experimentally based parameterization for predicting the polymer scaling exponent, ν, for disordered proteins based simply on sequence composition. Despite its simplicity, this prediction is remarkably accurate for a range of disordered proteins (Langridge et al., 2014; Perez et al., 2014; Tomasso et al., 2016; English et al., 2017; Yarawsky et al., 2017; English et al., 2018, 2019). This parameterization ignores such affects such as the distribution or clustering of charges (Das and Pappu, 2013), hydrophobic residues (Krishnan et al., 2008), or the relative spacing of more adhesive groups (i.e., stickers) and more inert groups (i.e., spacers) (Harmon et al., 2017; Martin et al., 2020). All of those effects will certainly contribute to the properties of an IDP, including its radius of hydration and propensity to form phase separated droplets.

The relative strengths of subunit-solvent and subunit-subunit interactions is a primary determinant of LLPS of polymers, and so we hypothesized that this simple, ensemble averaged, measure of polymer behavior would be related to the propensity of IDPs to spontaneously form protein-rich droplets. We queried a list of disordered proteins curated previously for their potential to phase separate, and shown to have a higher

propensity for potential cation-pi interactions (Vernon et al., 2018). We compared those proteins to a list of proteins whose radius of hydration has been measured in a monomeric state in solution. Note that we cannot exclude that these control proteins might have the propensity to phase separate under some buffer conditions, simply that they have not been observed to do so. Remarkably, the PS IDPs naturally separate from the non-PS IDPs on a plot of predicted (PS IDPs) or measured (non-PS IDPs) radius of hydration as a function of size (**Fig. 1**). This demonstrates that sequence composition in addition to the particular patterning of residues is important for determining the protein properties.

Next, we sought to determine if the composition of amino acids could reveal potential molecular mechanisms driving a preference for chain-chain relative to chain-solvent interactions. This mimics the idea that phase changes are stabilized by modular or multivalent interactions proposed by the Rosen lab (Li et al., 2012) and extends the idea from the Forman-Kay lab (Vernon et al., 2018) that pi-pi interactions stabilize coascervates. More recently Martin, et al incorporated this into the stickers and spacers model proposed by the Pappu lab (Harmon et al., 2017; Yang et al., 2019; Martin et al., 2020) where the interaction modules can be as small as single amino acids, in their case aromatic residues F and Y. We first analyzed the relative prevalence of amino acids in each of our groups and found significant depletion of I, L, and V residues in the PS IDPs as compared to the non-PS IDPs (**Fig. 2**). This is in addition to the depletion of hydrophobic residues in IDPs in general relative to their folded counterparts (Dunker et al., 2000). This is consistent with work from Forman-Kay and coworkers highlighting the importance of pi-orbitals in interactions between amino acids, and thus a preference among hydrophobic amino acids for those residues that can provide those interactions (Vernon et al., 2018).

Our parameterized prediction of ν relied on established relationships between the different amino acids and the common fold motifs. In the sequences we have studied so far, the sequence averaged propensity for folds (α-helix, β-turn, etc.) is sufficient to predict the polymer chain size. This dependence on local folding properties motivated us to consider whether particular fold types are more or less prevalent by prediction in the PS IDPs than in the non-PS IDPs. The relative depletion of I, L, and V is also consistent with an increase in β-turn propensity (**Fig. 2**). This finding is consistent with our previous work on the potential for β-turn structures, independent of specific sequence composition, to mediate chain-chain interactions and drive protein compaction and phase separation (Lyons et al., 2013, 2014). We previously focused on synthetic ELPs where increasing temperature drove an increase in both β-turn structure and propensity for phase separation. We hypothesized that that phase separation was favored by β-turn β-turn interactions. To test this hypothesis in that work, we modeled ELP interactions by generating a series of structural ensembles that incorporated differing amounts of β-turn bias throughout the chain (Zhang et al., 2018). Ensembles were then docked into dimer structures where sites of interaction were preferentially β-turns. Consistent with the hypothesis that β-turn β-turn interactions can be favorable relative to chain-solvent interactions driving chain compaction and phase separation, we found a strong preference for β-turn propensity predicted among the PS IDPs relative to the non-PS IDPs and folded proteins (**Fig. 2**).

Using the framework of stickers and spacers, it is possible that transient β-turns can serve as stickers in an ensemble of otherwise random conformations. Even a low population (2-5%) of β-turns (Zhang et al., 2018) may be sufficient to support self-association. As we have shown above, the partial collapse of a β-turn lowers the number of waters that must be removed during association, and it primes the backbone in a conformation that is favorable for the burial of hydrophobic surface area (**Fig. 3**). Energetically, this transient interaction can be driven by intrinsic turn propensity and a small, favorable hydrogen bond formation energy. On the other hand, a random, non-turn ensemble of conformations poses several challenges for self-association; these conformations must shed a larger number of waters when associating, and pairs of conformations must be selected that allow for significant hydrophobic burial. From the perspective of chain entropy, this represents a significant feat that is partially overcome when both associating chains are in β-turn conformations. Thus, our calculations are consistent with a model where

transient β-turn stickers are separated by long stretches of random, non-turn spacers, and this may explain the propensity for turns in IDPs that undergo LLPS.

In forming the β-turn β-turn complex, the assumption that macromolecular association is driven by the favorable energetics of hydrophobic burial has consequences that must be considered. The hydrophobic effect is known to be sensitive to temperature changes where increasingly cold conditions show progressively weaker potential to favor the nonspecific burial of hydrophobic groups from aqueous solution (Baldwin, 1986; Whitten et al., 2006). Because of this temperature dependence to the hydrophobic effect, the role of the β-turn in protein phase separation we have proposed is most likely applicable to heat induced phase separation, such as that described by the lower critical solution temperature (LCST) classification of phase separation (Quiroz and Chilkoti, 2015). While this doesn't exclude β-turns contribution, it is unlikely to apply or play a dominant role to the upper critical solution temperature (UCST) class where target proteins are miscible at higher temperatures and then de-mix when temperatures are lowered (Glatzel et al., 2011).

Lastly, we find it interesting that the intrinsic potential for self-interactions, as estimated from sequence-calculated ν, is on average lower (i.e., higher average ν) for sequences from proteins that adopt stable folded structures than it is for sequences from the PS IDP set; a result that is counterintuitive. This is shown in **Fig. 2B** and seems to emphasize the critical role of amino acid patterning for a folding reaction to occur and ultimately yield a globular structure stabilized by a network of intra-molecular interactions. Our compositional analysis of protein sequences specifically ignored patterning effects on structure. Nonetheless, two important implications about the character of IDRs and their role in LLPS emerge from these studies. First, IDRs that drive LLPS form a separate protein class from other IDRs and are identifiable from sequence by calculating the β-turn propensity and polymer scaling exponent, ν, which inversely relates to the ability of a protein to interact with itself. Second, a simple mechanism has been identified where LLPS can be primed in disordered regions by transitions to the β-turn and its concomitant desolvation. Molecular simulations additionally point to significant hydrophobic burial in the course of β-turn β-turn interactions that could act to stabilize the shift to a protein-rich phase.

Methods

***R_h* prediction.** The hydrodynamic dimensions of disordered protein ensembles depend strongly on sequence composition and R_h has been shown to be accurately predicted from the intrinsic chain bias for the polyproline II (PPII) conformation (Perez et al., 2014; Tomasso et al., 2016) and sequence estimates of the protein net charge (English et al., 2017, 2019). The equation to predict R_h is

$$R_h = 2.16\text{Å} \cdot N^{(0.503-0.11\cdot ln(1-f_{PPII}))} + 0.26 \cdot |Q_{net}| - 0.29 \cdot N^{0.5}, \qquad (1)$$

where N is the number of residues, f_{PPII} is the fractional number of residues in the PPII conformation, and Q_{net} is the net charge (English et al., 2019). f_{PPII} is estimated from $\sum P_{PPII,i}/N$, where $P_{PPII,i}$ is the experimental PPII propensity determined for amino acid type i in unfolded peptides (Elam et al., 2013) and the summation is over the protein sequence. Q_{net} is determined from the number of lysine and arginine residues minus the number of glutamic acid and aspartic acid.

Disorder prediction. The presence of intrinsic disorder in proteins and protein regions can be predicted from sequence with good confidence (Meng et al., 2017). The GeneSilico MetaDisorder service (Kozlowski and Bujnicki, 2012) was utilized to calculate the disorder tendency at each position in a sequence from the consensus prediction of 13 primary methods. Residues with a disorder tendency >0.5 are predicted to be disordered, while those with disorder tendency <0.5 are predicted to be ordered. To minimize

misidentification, we selected ID regions as those with at least 20 contiguous residue positions having disorder tendency ≥0.7.

Calculation of β-turn propensity. The propensity to form β-turn structures was calculated by $\sum scale_i/N$, where $scale_i$ is the value for amino acid type i in the β-turn frequencies from Levitt (Levitt, 1978) obtained from ProtScale at the ExPASy Bioinformatics Resource Portal (Gasteiger et al., 2003). The summation is over the protein sequence containing N number of amino acids. Normalized β-turn frequencies for the common amino acids are reproduced in **Table S5** of the **Supplementary Information**.

Computer generation of disordered ensembles. Structures of GVPGVG were generated by a random search of conformational space using a hard sphere collision model (Whitten et al., 2008). This model uses van der Waals atomic radii (Ramachandran et al., 1963; Iijima et al., 1987) as the only scoring function to eliminate grossly improbable conformations. The procedure to generate a random conformer starts with a unit peptide and all other atoms for a chain are determined by the rotational matrix (Jeffreys and Jeffreys, 1950). Backbone atoms are generated from the dihedral angles φ, ψ, and ω and the standard bond angles and bond lengths (Momany et al., 1975). Backbone dihedral angles are assigned randomly, using a random number generator based on Knuth's subtractive method (Knuth, 1981). (φ, ψ) is restricted to the allowed Ramachandran regions (Mandel et al., 1977) to sample conformational space efficiently. For peptide bonds, ω had a Gaussian fluctuation of ± 5% about the *trans* form (180°) for nonproline residues. Proline sampled the *cis* form (0°) at a rate of 10% (MacArthur and Thornton, 1991). Of the two possible positions of the Cβ atom in nonglycine residues, the one corresponding to L-amino acids was used. The positions of all other side chain atoms were determined from random sampling of rotamer libraries (Lovell et al., 2000). Structures adopting the type II β-turn were identified as those with (φ, ψ) angles of (-60°±15°, 120°±15°) and (80°±15°, 0°±15°) for P3 and G4, respectively, while also containing a hydrogen bond connecting the carbonyl oxygen of V2 to the amide proton of V5.

CHASA analysis and molecular docking. Computer generated structures, described above, were processed using the CHASA module (Fleming et al., 2005) of the LINUS software package (Srinivasan and Rose, 1995; Srinivasan et al., 2004). Two structures containing turns were docked using the GOLD/HERMES molecular docking software version 2020.1 (Jones et al., 1997). After hydrogen atoms were added, docking used the ChemPLP scoring function. The beta carbon on the third proline residue defined the binding site. Valine side chains were sampled using the built-in rotamer library, and all backbone torsions were held fixed in their original conformation. HERMES was used to calculate the buried hydrophobic accessible surface area upon formation of the complex.

Acknowledgements

This work was supported by the National Institutes of Health under grants R15GM115603 (S.T.W.), R25GM102783 (South Texas Doctoral Bridge Program; N. M. J. Blake, B. O. Oyajobi, and S.T.W.), R35GM119755 (L.E.H.), and R01AI139479 (N.C.F.), as well as the National Science Foundation under grant 1818090 (N.C.F.). No nongovernmental sources were used to fund this project. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF or NIH.

Competing interests: The authors declare that no competing interests exist.

References

Baldwin RL. 1986. Temperature dependence of the hydrophobic interaction in protein folding. *PNAS* **83**:8069–8072. doi:10.1073/pnas.83.21.8069

Brady JP, Farber PJ, Sekhar A, Lin Y-H, Huang R, Bah A, Nott TJ, Chan HS, Baldwin AJ, Forman-Kay JD, Kay LE. 2017. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc Natl Acad Sci USA* **114**:E8194–E8203. doi:10.1073/pnas.1706197114

Chong PA, Forman-Kay JD. 2016. Liquid-liquid phase separation in cellular signaling systems. *Curr Opin Struct Biol* **41**:180–186. doi:10.1016/j.sbi.2016.08.001

Das RK, Pappu RV. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *PNAS* **110**:13392–13397. doi:10.1073/pnas.1304749110

Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform Ser Workshop Genome Inform* **11**:161–171.

Elam WA, Schrank TP, Campagnolo AJ, Hilser VJ. 2013. Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites. *Protein Science* **22**:405–417. doi:10.1002/pro.2217

Elbaum-Garfinkle S, Kim Y, Szczepaniak K, Chen CC-H, Eckmann CR, Myong S, Brangwynne CP. 2015. The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *PNAS* **112**:7189–7194. doi:10.1073/pnas.1504822112

English LR, Tilton EC, Ricard BJ, Whitten ST. 2017. Intrinsic α helix propensities compact hydrodynamic radii in intrinsically disordered proteins. *Proteins* **85**:296–311. doi:10.1002/prot.25222

English LR, Tischer A, Demeler AK, Demeler B, Whitten ST. 2018. Sequence Reversal Prevents Chain Collapse and Yields Heat-Sensitive Intrinsic Disorder. *Biophys J* **115**:328–340. doi:10.1016/j.bpj.2018.06.006

English LR, Voss SM, Tilton EC, Paiz EA, So S, Parra GL, Whitten ST. 2019. Impact of Heat on Coil Hydrodynamic Size Yields the Energetics of Denatured State Conformational Bias. *J Phys Chem B* **123**:10014–10024. doi:10.1021/acs.jpcb.9b09088

Fleming PJ, Fitzkee NC, Mezei M, Srinivasan R, Rose GD. 2005. A novel method reveals that solvent water favors polyproline II over beta-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). *Protein Sci* **14**:111–118. doi:10.1110/ps.041047005

Flory PJ. 1969. Statistical mechanics of chain molecules. New York: Interscience Publishers.

Flory PJ. 1949. The Configuration of Real Polymer Chains. *J Chem Phys* **17**:303–310. doi:10.1063/1.1747243

Garzon-Rodriguez W, Sepulveda-Becerra M, Milton S, Glabe CG. 1997. Soluble Amyloid Aβ-(1–40) Exists as a Stable Dimer at Low Concentrations. *J Biol Chem* **272**:21037–21044. doi:10.1074/jbc.272.34.21037

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**:3784–3788. doi:10.1093/nar/gkg563

Glatzel S, Laschewsky A, Lutz J-F. 2011. Well-Defined Uncharged Polymers with a Sharp UCST in Water and in Physiological Milieu. *Macromolecules* **44**:413–415. doi:10.1021/ma102677k

Harmon TS, Holehouse AS, Rosen MK, Pappu RV. 2017. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**:e30294. doi:10.7554/eLife.30294

Hofmann H, Soranno A, Borgia A, Gast K, Nettels D, Schuler B. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *PNAS* **109**:16155–16160. doi:10.1073/pnas.1207719109

Iijima H, Dunbar JB, Marshall GR. 1987. Calibration of effective van der Waals atomic contact radii for proteins and peptides. *Proteins: Structure, Function, and Bioinformatics* **2**:330–339. doi:10.1002/prot.340020408

Jeffreys H, Jeffreys BS. 1950. Methods of mathematical physics. New York: Cambridge University Press.

Jones G, Willett P, Glen RC, Leach AR, Taylor R. 1997. Development and validation of a genetic algorithm for flexible docking11Edited by F. E. Cohen. *Journal of Molecular Biology* **267**:727–748. doi:10.1006/jmbi.1996.0897

Knuth DE. 1981. The Art of Computer Programming, Volume 2: Seminumerical Algorithms, 2nd ed. Reading, MA: Addison-Wesley.

Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *PNAS* **101**:12491–12496. doi:10.1073/pnas.0403643101

Kozlowski LP, Bujnicki JM. 2012. MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**:111. doi:10.1186/1471-2105-13-111

Krishnan VV, Lau EY, Yamada J, Denning DP, Patel SS, Colvin ME, Rexach MF. 2008. Intramolecular Cohesion of Coils Mediated by Phenylalanine–Glycine Motifs in the Natively Unfolded Domain of a Nucleoporin. *PLOS Computational Biology* **4**:e1000145. doi:10.1371/journal.pcbi.1000145

Langridge TD, Tarver MJ, Whitten ST. 2014. Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins* **82**:668–678. doi:10.1002/prot.24449

Lee CF, Brangwynne CP, Gharakhani J, Hyman AA, Jülicher F. 2013. Spatial Organization of the Cell Cytoplasm by Position-Dependent Phase Separation. *Phys Rev Lett* **111**:088101. doi:10.1103/PhysRevLett.111.088101

Levitt M. 1978. Conformational preferences of amino acids in globular proteins. *Biochemistry* **17**:4277–4285. doi:10.1021/bi00613a026

Li P, Banjade S, Cheng H-C, Kim S, Chen B, Guo L, Llaguno M, Hollingsworth JV, King DS, Banani SF, Russo PS, Jiang Q-X, Nixon BT, Rosen MK. 2012. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**:336–340. doi:10.1038/nature10879

Li Y, Zhao C, Luo F, Liu Z, Gui X, Luo Z, Zhang X, Li D, Liu C, Li X. 2018. Amyloid fibril structure of α-synuclein determined by cryo-electron microscopy. *Cell Res* **28**:897–903. doi:10.1038/s41422-018-0075-x

Lovell SC, Word JM, Richardson JS, Richardson DC. 2000. The penultimate rotamer library. *Proteins* **40**:389–408.

Lyons DF, Le V, Bidwell GL, Kramer WH, Lewis EA, Raucher D, Correia JJ. 2013. Structural and Hydrodynamic Analysis of a Novel Drug Delivery Vector: ELP[V5G3A2-150]. *Biophysical Journal* **104**:2009–2021. doi:10.1016/j.bpj.2013.03.040

Lyons DF, Le V, Kramer WH, Bidwell GL, Lewis EA, Raucher D, Correia JJ. 2014. Effect of Basic Cell-Penetrating Peptides on the Structural, Thermodynamic, and Hydrodynamic Properties of a Novel Drug Delivery Vector, ELP[V5G3A2-150]. *Biochemistry* **53**:1081–1091. doi:10.1021/bi400955w

MacArthur MW, Thornton JM. 1991. Influence of proline residues on protein conformation. *Journal of Molecular Biology* **218**:397–412. doi:10.1016/0022-2836(91)90721-H

Mandel N, Mandel G, Trus BL, Rosenberg J, Carlson G, Dickerson RE. 1977. Tuna cytochrome c at 2.0 A resolution. III. Coordinate optimization and comparison of structures. *J Biol Chem* **252**:4619–4636.

Marsh JA, Forman-Kay JD. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys J* **98**:2383–2390. doi:10.1016/j.bpj.2010.02.006

Martin EW, Holehouse AS, Peran I, Farag M, Incicco JJ, Bremer A, Grace CR, Soranno A, Pappu RV, Mittag T. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**:694–699. doi:10.1126/science.aaw8653

Meng F, Uversky VN, Kurgan L. 2017. Comprehensive review of methods for prediction of intrinsic disorder and its molecular functions. *Cell Mol Life Sci* **74**:3069–3090. doi:10.1007/s00018-017-2555-4

Mitrea DM, Cika JA, Stanley CB, Nourse A, Onuchic PL, Banerjee PR, Phillips AH, Park C-G, Deniz AA, Kriwacki RW. 2018. Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation. *Nat Commun* **9**:842. doi:10.1038/s41467-018-03255-3

Mitrea DM, Kriwacki RW. 2016. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling* **14**:1. doi:10.1186/s12964-015-0125-7

Momany FA, McGuire RF, Burgess AW, Scheraga HA. 1975. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J Phys Chem* **79**:2361–2381. doi:10.1021/j100589a006

Perez RB, Tischer A, Auton M, Whitten ST. 2014. Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. *Proteins* **82**:3373–3384. doi:10.1002/prot.24692

Quiroz FG, Chilkoti A. 2015. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nature Materials* **14**:1164–1171. doi:10.1038/nmat4418

Ramachandran GN, Ramakrishnan C, Sasisekharan V. 1963. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology* **7**:95–99. doi:10.1016/S0022-2836(63)80023-6

Reichheld SE, Muiznieks LD, Keeley FW, Sharpe S. 2017. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *PNAS* **114**:E4408–E4415. doi:10.1073/pnas.1701877114

Snead WT, Zeno WF, Kago G, Perkins RW, Richter JB, Zhao C, Lafer EM, Stachowiak JC. 2019. BAR scaffolds drive membrane fission by crowding disordered domains. *J Cell Biol* **218**:664–682. doi:10.1083/jcb.201807119

Srinivasan R, Fleming PJ, Rose GD. 2004. Ab initio protein folding using LINUS. *Meth Enzymol* **383**:48–66. doi:10.1016/S0076-6879(04)83003-9

Srinivasan R, Rose GD. 1995. LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins* **22**:81–99. doi:10.1002/prot.340220202

Tomasso ME, Tarver MJ, Devarajan D, Whitten ST. 2016. Hydrodynamic Radii of Intrinsically Disordered Proteins Determined from Experimental Polyproline II Propensities. *PLoS Comput Biol* **12**:e1004686. doi:10.1371/journal.pcbi.1004686

Uversky VN. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **11**:739–756. doi:10.1110/ps.4210102

Uversky VN, Kuznetsova IM, Turoverov KK, Zaslavsky B. 2015. Intrinsically disordered proteins as crucial constituents of cellular aqueous two phase systems and coacervates. *FEBS Letters* **589**:15–22. doi:10.1016/j.febslet.2014.11.028

Vernon RM, Chong PA, Tsang B, Kim TH, Bah A, Farber P, Lin H, Forman-Kay JD. 2018. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. *Elife* **7**. doi:10.7554/eLife.31486

Whitten ST, Kurtz AJ, Pometun MS, Wand AJ, Hilser VJ. 2006. Revealing the Nature of the Native State Ensemble through Cold Denaturation. *Biochemistry* **45**:10163–10174. doi:10.1021/bi060855+

Whitten ST, Yang H-W, Fox RO, Hilser VJ. 2008. Exploring the impact of polyproline II (PII) conformational bias on the binding of peptides to the SEM-5 SH3 domain. *Protein Sci* **17**:1200–1211. doi:10.1110/ps.033647.107

Wilkins DK, Grimshaw SB, Receveur V, Dobson CM, Jones JA, Smith LJ. 1999. Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **38**:16424–16431. doi:10.1021/bi991765q

Wuttke R, Hofmann H, Nettels D, Borgia MB, Mittal J, Best RB, Schuler B. 2014. Temperature-dependent solvation modulates the dimensions of disordered proteins. *Proc Natl Acad Sci U S A* **111**:5213–5218. doi:10.1073/pnas.1313006111

Yang Y, Jones HB, Dao TP, Castañeda CA. 2019. Single Amino Acid Substitutions in Stickers, but Not Spacers, Substantially Alter UBQLN2 Phase Transitions and Dense Phase Material Properties. *J Phys Chem B* **123**:3618–3629. doi:10.1021/acs.jpcb.9b01024

Yarawsky AE, English LR, Whitten ST, Herr AB. 2017. The Proline/Glycine-Rich Region of the Biofilm Adhesion Protein Aap Forms an Extended Stalk that Resists Compaction. *J Mol Biol* **429**:261–279. doi:10.1016/j.jmb.2016.11.017

Zeno WF, Thatte AS, Wang L, Snead WT, Lafer EM, Stachowiak JC. 2019. Molecular Mechanisms of Membrane Curvature Sensing by a Disordered Protein. *J Am Chem Soc* **141**:10361–10371. doi:10.1021/jacs.9b03927

Zhang Y, Trabbic-Carlson K, Albertorio F, Chilkoti A, Cremer PS. 2006. Aqueous Two-Phase System Formation Kinetics for Elastin-Like Polypeptides of Varying Chain Length. *Biomacromolecules* **7**:2192–2199. doi:10.1021/bm060254y

Zhang Y, Zai-Rose V, Price CJ, Ezzell NA, Bidwell GL, Correia JJ, Fitzkee NC. 2018. Modeling the Early Stages of Phase Separation in Disordered Elastin-like Proteins. *Biophys J* **114**:1563–1578. doi:10.1016/j.bpj.2018.01.045

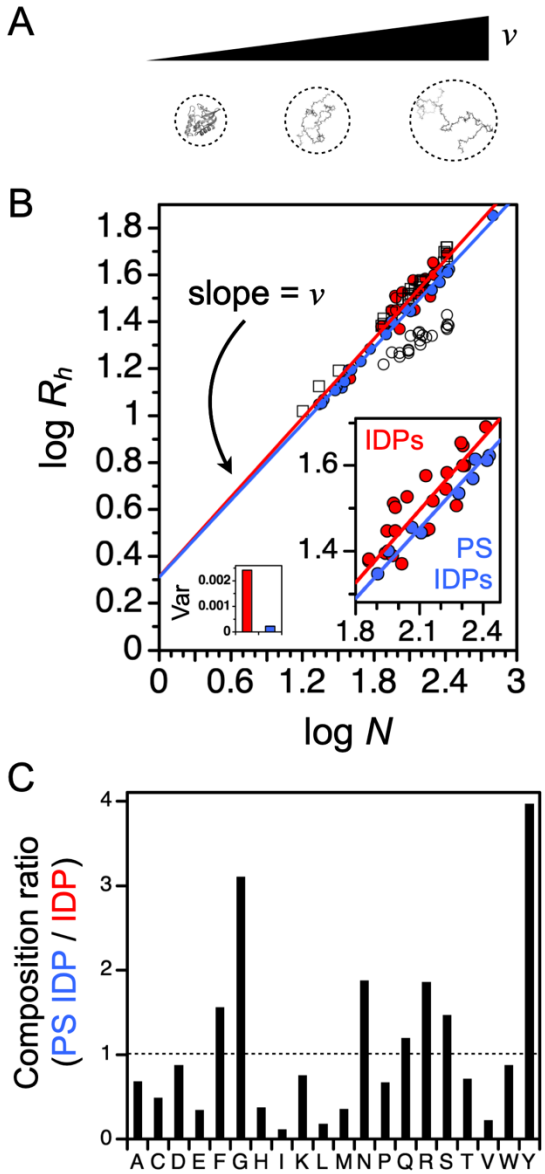

Figure 1. Hydrodynamic dimensions of proteins. A) Scaling exponent, ν, from the power law relationship, $R_h = R_o \cdot N^{\nu}$, is proportional to hydrodynamic size. **B)** Mean hydrodynamic radius, R_h (Å), measured for folded proteins (open circles), chemically denatured proteins (open squares), and IDPs (red circles) compared to the sequence length, N. Blue circles are R_h that were predicted for the ID regions of proteins known to phase separate (PS IDPs). Right inset: duplicated data from the main plot that highlights the reduced ν of the PS IDP set and the lower variance from its trend line, when compared to the IDP set. Left inset: variance, Var, of each set, IDP (red) and PS IDP (blue), calculated as $1/n \cdot \Sigma\ (x - \bar{x})^2$, where n is the number of proteins, x is log R_h, and $\bar{x}$ the trend line value determined from N. **C)** Composition ratio is the percent composition of each amino acid type, identified by its 1-letter code, for the PS IDP set divided by the IDP set.

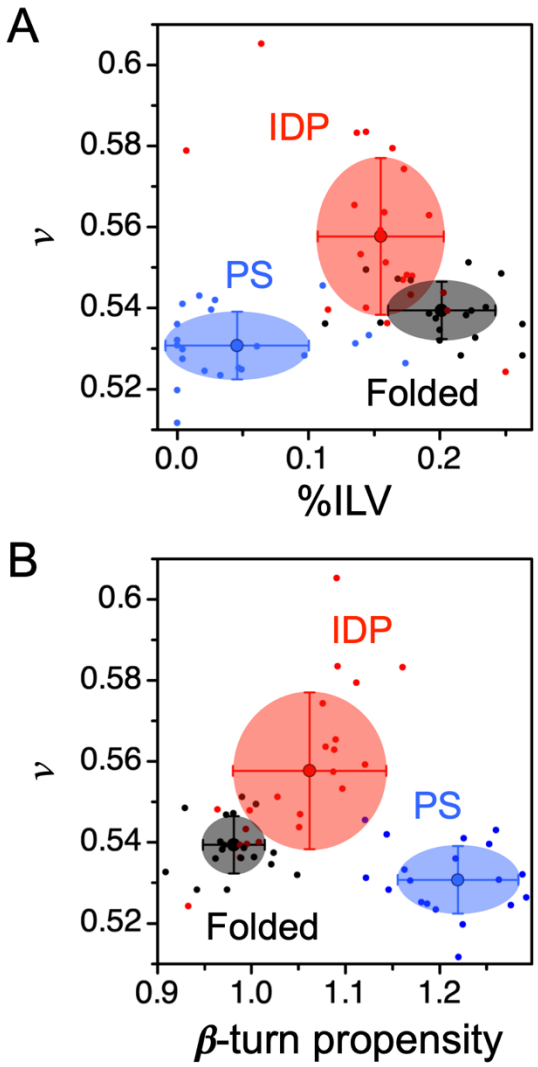

Figure 2. Scaling exponent and compositional differences of IDPs, PS IDPs, and folded proteins. Scaling exponent, ν, was calculated for each protein from R_h and N as $\nu = log(R_h/R_o)/logN$, using 2.16 Å for R_o obtained from simulated conformational ensembles (Langridge et al., 2014). R_o determined from y-axis intercepts of the trend lines for IDPs and PS IDPs in Figure 1 yields 2.1 Å, showing good agreement with computer simulated results for R_o. R_h used to calculate ν were experimental for IDPs, whereas PS IDPs and folded proteins used sequence predicted R_h by equation 1 in Methods. Large points and error bars show the averages and standard deviations for the IDP (red), PS IDP (blue), and folded protein (black) sets. The smaller points show individual proteins using the same coloring scheme. **A)** Comparison of ν to the combined percent composition of the branched amino acids, I + L + V. **B)** Comparison of ν to the β-turn propensity calculated from sequence.

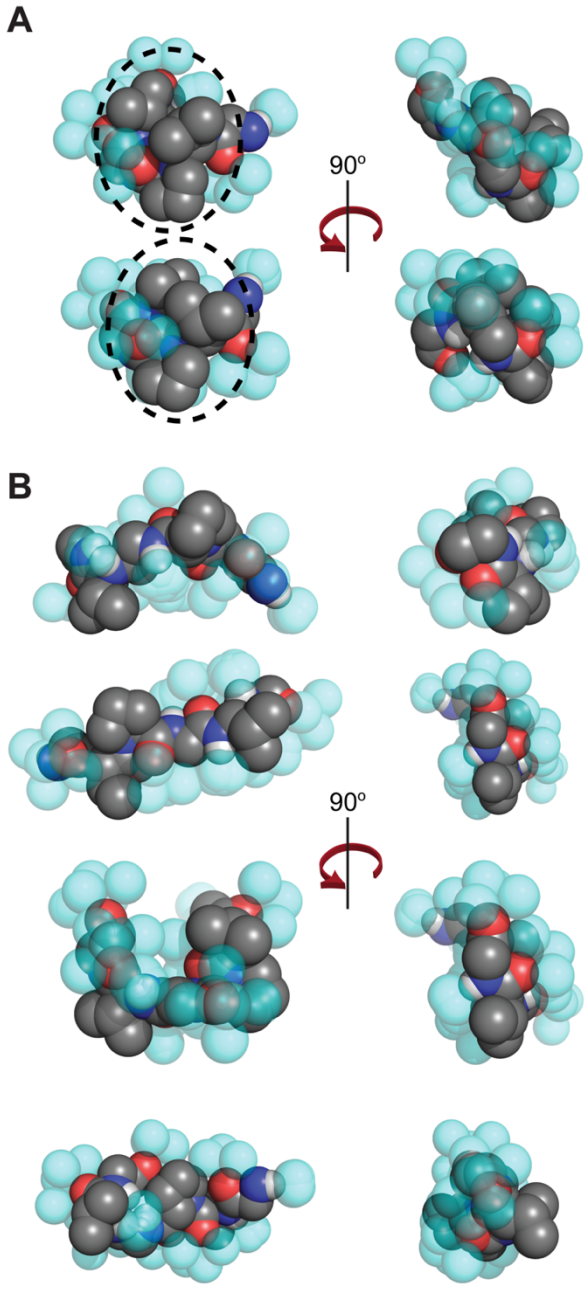

Figure 3. Representative structures of turn and non-turn surface area. Conformations are shown for the ELP repeat GVPGVG, including sterically allowed, CHASA-generated solvation waters (see text). **A)** Two representative turn conformation with a hydrogen bond between residues V2 and V5. Hydrophobic surface area from the V and P residues is clustered together and highlighted with a dashed black oval. The right-hand panels are generated by a 90° rotation about the indicated axis. **B)** ELP structures in four typical non-turn conformations. While the total hydrophobic surface area is greater than for the structures shown in panel A, it is less well-localized. CHASA solvation waters, associated with hydrogen bonding groups in the peptide backbone, are shown as semitransparent cyan spheres.

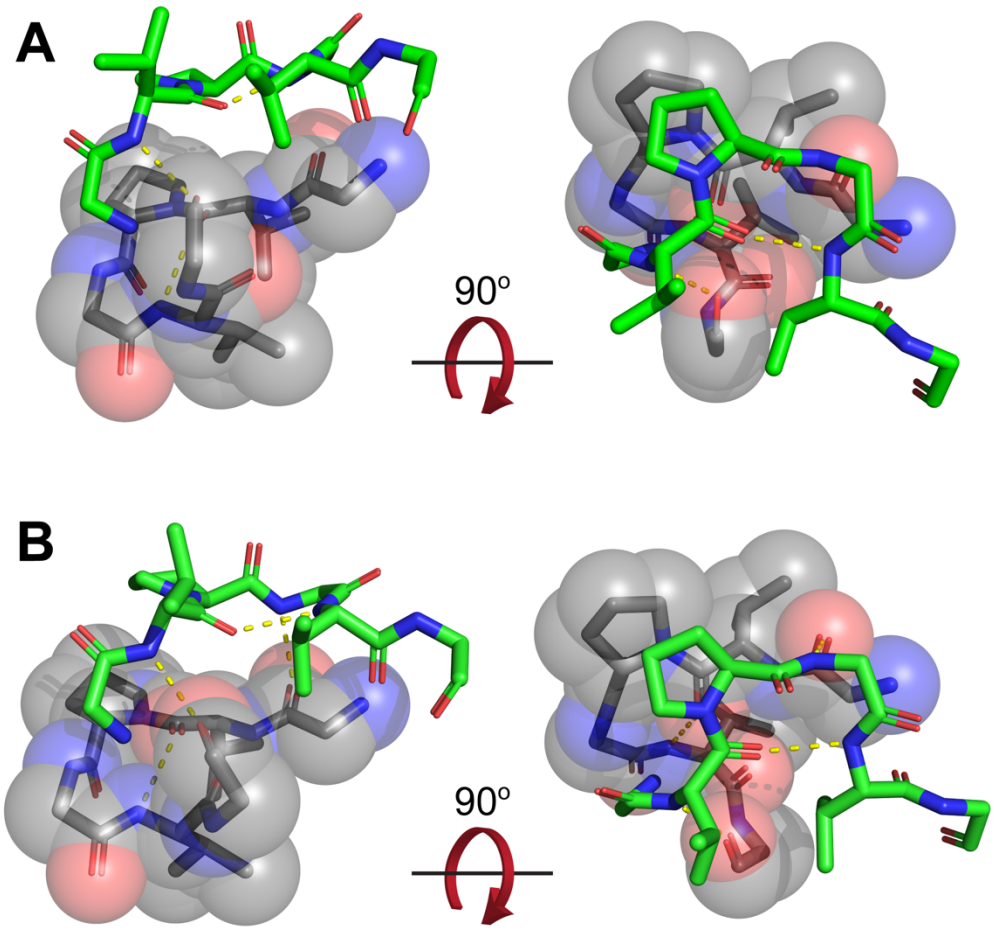

Figure 4. Representative complex from a β-turn β-turn interaction. GOLD-generated docking solution of two β-turns showing how turns might self-associate during LLPS, each turn burying approximately 110 $Å^2$ of hydrophobic surface area. The two turns are colored grey (semi-transparent space filling representation) and green (stick representation). Hydrogen bonds are rendered as dashed lines.

Tables

Table I. Structural properties of turn and non-turn ensembles.

	Non-Turn Ensemble [a]	**β-Turn Ensemble** [a]
Total ASA (Å^2)	738.3 ± 0.9	670.3 ± 0.5
Hydrophobic ASA (Å^2)	536.4 ± 0.8	488.9 ± 0.5
CHASA (Å^2)	353.6 ± 0.6	327.3 ± 0.6
Hydrophobic ASA Lost assuming backbone hydration (Hydrophobic ASA – CHASA, Å^2)	182.8 ± 1.0	161.6 ± 0.7
Number of Backbone Hydration Waters (CHASA maximum is 55)	44.4 ± 0.1	37.1 ± 0.1

[a] Uncertainties were calculated as the standard error of the mean.