The free energy folding penalty accompanying binding of intrinsically disordered α-helical motifs

San Hadži | Jurij Lah

Abstract
Intrinsically disordered proteins (IDPs) are abundant in eukaryotic proteomes and preform critical roles in many cellular processes, most often through the association with globular proteins. Despite lacking a stable three-dimensional structure by themselves, they may acquire a defined conformation upon binding globular targets. The most common type of secondary structure acquired by these binding motifs entails formation of an α-helix. It has been hypothesized that such disorder-to-order transitions are associated with a significant free energy penalty due to IDP folding, which reduces the overall IDP-target affinity. However, the exact magnitude of IDP folding penalty in α-helical binding motifs has not been systematically estimated. Here, we report the folding penalty contributions for 30 IDPs undergoing folding-upon-binding and find that the average IDP folding penalty is +2.0 kcal/mol and ranges from 0.7 to 3.5 kcal/mol. We observe that the folding penalty scales approximately linearly with the change in IDP helicity upon binding, which provides a simple empirical way to estimate folding penalty. We analyze to what extent do pre-folding and bound-state dynamics (fuzziness) reduce the folding penalty and find that these effects combined, on average, reduce the folding cost by around half. Taken together, the presented analysis provides a quantitative basis for understanding the role of folding penalty in IDP-target interactions and introduces a method estimate this quantity. Estimation and reduction of IDP folding penalty may prove useful in the rational design of helix-stabilized inhibitors of IDP-target interactions.

Statement: The α-helical binding motifs are ubiquitous among the intrinsically disordered proteins (IDPs). Upon binding their targets, they undergo a disorder-to-order transition, which is accompanied by a significant folding penalty whose magnitude is generally not known. Here, we use recently developed statistical-thermodynamic model to estimate the folding penalties for 30 IDPs and clarify the roles of IDP pre-folding and bound-state dynamics in reducing the folding penalty.
1 | INTRODUCTION

Intrinsically disordered proteins (IDPs) are abundant class of proteins, particularly in the eukaryotic proteomes where intrinsic disorder is present in around 40% of proteins. Due to their specific amino acid composition, which is depleted in hydrophobic but enriched in charged residues, these proteins do not fold cooperatively into a stable three-dimension structure but populate an ensemble of disordered, coil-like conformations. Nevertheless, they preform critical roles in transcription regulation and signal transduction and play a key role in the liquid–liquid phase separation. In most cases, the function of IDPs is intricately linked to the interactions with globular protein targets. Upon binding their targets some IDPs become structured, most commonly this entails a formation of α-helix, however, they may also remain dynamic and partially disordered, forming the so-called fuzzy complexes. It has been hypothesized that the interactions involving a disorder-to-order transition have reduced IDP-target affinity because of the IDP folding penalty—the free energy for transition from the disordered IDP ensemble to the ordered, target-bound ensemble. Given that IDP folding is entropically unfavorable, the overall association is, therefore, driven by the energetically favorable IDP-target interactions. Adding to the complexity of the process, the IDPs are not completely disordered in the unbound state and also do not fold into a complete α-helix in the target-bound state. Both of these phenomena, referred to as IDP pre-folding and fuzziness, have been identified as possible mechanisms that reduce the folding penalty and thereby increase IDP-target affinity.

For many IDPs, it has been observed that in their unbound, free state they may transiently populate the target-bound conformation. These nascent helical structures have been referred to as pre-structured motifs or pre-folded elements, and it has been assumed that this partial folding can effectively pre-pay a part of the folding penalty. We have recently compiled a database of experimental helicities of disordered peptides that undergo folding-upon-binding and suggested that pre-folded structures may be a general feature of α-helical binding motifs. This conclusion was based on the observation that the same IDP residues, which are responsible for target binding tend to have a significantly higher propensity to form helical structure. This particularly holds for leucine, which mediates the majority of interactions and has high preference for helical structure. Therefore, in many binding motifs, particularly those rich in leucine, nascent helical structure and binding are to some extent coupled. In addition to pre-folding, the IDP folding penalty can be also reduced by the increased dynamics in the target-bound IDP ensemble (fuzziness). However, the discussions on the role of folding penalty in the IDP-target interactions and its regulation by pre-folding or fuzziness critically lack some quantitative estimates. In other words, how large exactly is the IDP folding penalty?

There have been only a few attempts to quantitatively estimate the magnitude of folding penalty. Perhaps most systematic attempt has been the study from Thelium et al. who estimated this quantity indirectly by comparing the free energy of association of IDP-target and globular-globular complex. They observed a difference 2–2.5 kcal/mol and suggested that the lower affinity for IDP-target complexes is due to folding penalty. Another approach is based on the surface-area parametrizations and was used to estimate ΔGfold for the CcdA peptide, and has recently been improved for the estimation of the IDP folding entropy. Statistical-thermodynamic models have been used to estimate thermodynamic folding contributions for two IDP systems, but have not been applied to a larger IDP sample. The estimation of ΔGfold for IDPs that fold upon binding is more complicated as it seems initially. First, there is often a considerable confusion between the thermodynamic view of coupled folding and binding and the kinetic picture of this process (conformational selection vs. induced fit). Given that the thermodynamic contributions are evaluated using functions of state, the estimation of folding contribution is independent of the actual path connecting the initial and final state (conformational selection or induced fit). Thus, the kinetic mechanism of folding-upon-binding transition is irrelevant for the discussion of the thermodynamic (equilibrium) contributions of IDP folding penalty. Second, the experimental characterization of the overall IDP-target association in terms of affinity will provide only the sum of the IDP folding and the binding contributions, and one cannot, by simple means, determine these two contributions separately. Third, as mentioned above, folding upon binding is not a simple two-state process. Since helix-coil equilibria in peptides are characterized by low degree of cooperativity the unbound IDPs exist in
an ensemble of conformations with varying degrees of helicity. For this reason, description of folding-upon-binding transition as a two-state process can be considered only as an approximation. It, therefore, appears that a better description of the folding-upon-binding transitions should consider the transition between the two IDP ensembles: the unbound and the target-bound one.

We have previously shown that the structure and energetics of the α-helical IDP ensembles can be conveniently described using the helix-coil theory. In principle for the IDPs in their free, unbound state the sequence alone should provide sufficient information to describe this ensemble, since helix-coil parameters are known for each amino acid and the sequence-ensemble relation is rather well understood. On the other hand, for the IDP ensemble in the target-bound state a different model is needed that describes how formation of target interactions affects the IDP conformation. Previously, we developed a modified helix-coil model where the target-bound IDP ensemble is assumed to be constrained by several strong IDP-target interactions (hotspots). For IDPs folding to α-helix, these hotspots are required to adopt a helical confirmation; therefore, the formation of hotspot interactions also drives folding of IDP into helical ensemble. We explained how the distribution of hotspot residues along the IDP sequence limits the number of possible microstates in the IDP-target ensemble, while the IDP helix propensity (tendency to form helix) determines the probabilities of these microstates. Thus, depending on the arrangement of hotspots on the IDP and IDP folding propensity different target-bound ensembles can emerge, which range from dynamic ones with high degree of conformational heterogeneity to the more rigid and ordered ensembles. We have shown that hotspot-constrained ensemble model can accurately reproduce the target-bound IDP dynamics as observed by NMR and the effect of IDP mutations on the IDP-target affinity.

Using these tools, we now address the fundamental question of folding-upon-binding interactions and estimate the value of IDP folding penalty for 30 IDP systems harboring the α-helix binding motif. Furthermore, we discuss the role of pre-folding and target-bound fuzziness in quantitative terms and show that these two effects combined can almost halve the value of IDP folding penalty. An empirical relation is introduced that can be used to approximately estimate the IDP folding penalty from the gain in IDP helicity upon binding. A comparison of the folding penalty contribution with the overall IDP-target affinities shows no correlation between these two quantities, suggesting that folding penalty is not a major determinant of IDP-target affinity in general; however, its role depends on the system. For weakly interacting IDP-target systems the relative contribution of folding penalty can be quite substantial (almost 50%) and thus importantly determines the IDP-target association. Importantly, reducing the folding penalty can be used to optimize IDP-target affinity, and the introduced empirical relation provides a quantitative basis for understanding the relationship between helix-stabilization and IDP folding penalty.

2 RESULTS AND DISCUSSION

2.1 The folding penalty of 30 α-helical binding motifs

We have compiled a dataset of 30 IDPs that fold into α-helix upon binding their targets (Table 1, Table S1). This dataset is based on the previously published collection of 65 experimentally characterized IDPs that fold-upon-binding. For these IDPs, both the structure of the IDP-target complex as well as the average helicity in the unbound state are known from the literature (Table S1). The unbound helicity, αUNETBOUND, has been calculated from the circular dichroism data (see Section 3, Equation 2). For these IDPs, we find that on average they possess 21% of helical structure in their unbound state. For each IDP in the dataset, we use the Liefson–Roig helix-coil theory to describe the unbound state ensemble, that is to obtain the probability of each microstate with a given helicity. The average ensemble helicity is then calculated as the weighted sum of all microstates multiplied by their helicities (Equation 5), and should in principle match the experimentally determined αUNETBOUND. We use a heteropolymer version of Liefson–Roig model (Equations 3–5) that, based on the IDP sequence, assigns specific helix propagation constants \( w_i \) to each amino acid. However, helix propagation constants alone are not sufficient to reproduce the experimental value of the αUNETBOUND, most likely because there are additional interactions affecting helix stability which are not included in our treatment (e.g., interactions between sidechains, N- and C-capping interactions, helix macrodipole). Even the most advanced versions of helix-coil algorithms such as AGADIR, that consider many of these interactions, generally underestimate the absolute helix content for IDPs. We, therefore, introduce a calibration factor that multiplies all \( w_i \) constants by a given value and is determined for each IDP separately such that the model-calculated helicity matches the experimental value (see Section 3). These calibration factors range between 1 and 2 and are listed in Table S1. In short, we consider helix propensity of each amino acid separately using \( w_i \) constants, while all other interactions are treated empirically.
In con-
Depending
In short, the
The magnitude of IDP
3
and shown in Figure
4o f1 1
HAD
21
Δ
Gfold
IDP to target
ΔG-fold
IDP to target
ΔG-fold
ACTR1046−1058 to CBP
1.2
NDR62−69 to S100B
2.6
BAK76−96 to BCL2
2.7
NOXA_A13−47 to MCL-1
2.3
BAX49−83 to BCL2
2.6
NOXA_B64−98 to MCL-1
0.9
BECN1105−130 to BLC2
3.0
NTAIL486−505 to protein P
2.2
BID76−112 to BLC2
2.3
p53_92−111 to S100B
2.4
BMF24−158 to BCL2
0.7
PaaA16−34 to ParE2
1.8
Calcineurin396−414 to Calmodulin
2.2
PaaA235−63 to ParE2x
1.3
CeRED163−223 to SMU-1
2.4
Phd56−73 to Doc
2.2
cMyb291−315 to CBP
1.5
Phd64−73 Doc
1.7
E2A9−27 to CBP
1.4
pKID102−132 to CBP
2.3
FCP1324−358 to RAP74
1.3
PUMA102−132 to MCL-1
2.0
FOXO3a-CR2C462−483 to CBP
0.7
p65(ReLA)923−316 to IkBβ
3.0
HBZ33−56 to CBP
3.2
SID(MAD1)7−14 to Sin3A
1.1
IA3−68 to Proteinase A
3.5
TPX2310−41 to Aurora-A
1.5
MLL2846−2859 to CBP
0.7
TRTK12265−276 to protein16A
1.0

Note: Further details (IDP sequences, UNIPROT and PDB codes of IDP-target complexes) are provided in Table S1. The estimated the average uncertainty of the ΔGfold arising from different sources is ±1.2 kcal/mol (see Section 3).
Abbreviation: IDP, intrinsically disordered protein.

using the calibration factor that is applied evenly to all residues. Using these calibrated propagation constants, we calculate the partition function and the properties of the unbound IDP ensemble. As an example, we show the distribution of microstates in the unbound IDP ensemble and its average helicity for PaaR and E2A peptides (Figure S1).

To describe the target-bound IDP ensemble, where IDP acquires more helical conformations, we use the hotspot-constrained helix-coil model as described before. In contrast to the unbound ensemble, in the target-bound ensemble the IDP conformations are constrained by different interactions with the target. The model assumes that the IDP residues which mediate strongest interactions with the target (hotspots) must adopt a helical confirmation, as it is observed in the structure of the IDP-target complex. Due to the cooperativity of helix-coil transition, the neighboring IDP residues will also tend to form a helix, but are not required per se, which leaves the room for target-bound dynamics and conformational heterogeneity. Depending on the distribution of hotspots, the bound-state IDP ensemble can range from an ordered one with mostly helical conformations, to the more dynamic ensemble where several microstates with different degrees of helicity can satisfy the target-interaction pattern. To identify the IDP hotspots, we analyze the high-resolution structures of IDP-target complexes using a computation tool PPCheck. In short, the obtained distribution of hotspots on the IDP sequence, together with the calibrated helix propagation constants are used to define the target-bound IDP ensemble and its partition function (Equation 6). An example of target-bound IDP ensembles for PaaR and E2A peptides (probability distribution for different microstates) is shown in Figure S1, and shows how hotspots increase the probability of helical microstates, resulting in the folding-upon-binding transition.

The folding free energy ΔGfold can be estimated from the obtained partition functions of target-bound and unbound IDP ensembles (Equation 7). The ΔGfold thus represents the free energy associated with the transition from the native, unbound IDP to the target-bound IDP ensemble and does not include any energetic contributions due to IDP-target interactions (Figure 1, middle). The target-bound ensemble includes only intramolecular interactions and is fully solvated, since wi constants correspond to the helix-coil equilibria in water. As expected the ΔGfold values are positive, since IDPs do not spontaneously adopt target-bound confirmations with high helicity; these are stabilized because of favorable IDP-target interactions. The obtained values of ΔGfold for all IDPs are listed in Table 1 and shown in Figure 2. They span from 0.7 kcal/mol (FOXO3A, BMF) to 3.5 kcal/mol (IA3) with the average value of 2.0 kcal/mol. These values thus provide the first systematic estimate of folding penalty for a larger IDP dataset. By considering different sources of errors, we estimate that the average uncertainty in reported values is 1.2 kcal/mol (see Section 3).
There have been only few attempts to estimate $\Delta G_{\text{fold}}$ using different approaches. Teilum and coworkers compared the association affinities between IDP-globular and globular-globular complexes.\textsuperscript{16} They observed that the average association free energy of IDP-globular complexes is $2-2.5$ kcal/mol lower compared to the globular-globular complexes. This difference has been attributed to the IDP folding penalty and the reported value is in a similar range as our average $\Delta G_{\text{fold}}$ contribution. The study, however, included also some structurally different IDP motifs (not only $\alpha$-helical ones) and the assumption that the difference in the average affinities equals the average folding penalty is somewhat questionable. Another indirect approach relies on the use of empirical thermodynamic relationships, which are used to dissect the overall parameters of IDP-target association in to folding and binding contributions. The approach, which has been pioneered by Spolar and Record,\textsuperscript{22} has recently been re-calibrated to analyze IDP-target interactions.\textsuperscript{17} However, the mentioned studies estimate only the conformational entropy difference, not the folding free energy. Another study also estimated the enthalpic contribution using empirical surface-area relations to obtain the folding free energy for CcdA peptide, however the resulting value has a very high error margin.\textsuperscript{16} The problem with the approaches based on the empirical parameterizations is that they are indirect, and that the empirical coefficients are associated with a significant uncertainty. Thus, one is confronted with the problem of deriving the value of folding penalty from a difference of several large opposing contributions (entropy, enthalpy) resulting in high uncertainty. A more direct experimental method is the helicity perturbation approach, which has been developed to estimate the folding penalty of CcdA protein.\textsuperscript{23} According to the method IDP mutations are introduced to the solvent exposed IDP positions in order to change the helix folding energetics by a known amount (based on the values from helix propensity scales). Next, the effects of mutations on the IDP-target affinity and IDP helicity are measured and these differences are analyzed using the appropriate thermodynamic model to extract the folding penalty. In the case of CcdA this yielded $\Delta G_{\text{fold}} = 1.9$ kcal/mol.\textsuperscript{23} Finally, a statistical-mechanical approach to this problem has been pioneered by Rajasekaran et al. who used the Wako–Saito–Muñoz–Eaton model with two values of entropic penalty for the ordered and disordered
residues. Unfortunately, the study estimated only the entropic contribution, free energy has not been estimated. We recently analyzed the folding-upon-binding interaction of the HigA2 peptide, which is characterized by the unusually tight, picomolar affinity. Using a similar approach as described here, we estimated the folding penalty of HigA2 as 2.8 kcal/mol. In that study, the HigA2 unbound ensemble was treated in the same way as here, while the target-bound ensemble was defined a bit differently, not using the hotspot-constrained model. Rather all residues in HigA2 which are helical in the IDP-target complex were constrained to helix conformation, not just hotspots. A recalibration of $\Delta G_{\text{fold}}$ using the hotspot-constrained model for the target-bound ensemble gives a similar value, 2.6 kcal/mol. This is expected since HigA2 forms well ordered, enthalpy-optimized complex and the contribution of bound-state disorder in expected to be small. Collectively, our estimates place the folding penalty of $\alpha$-helical binding motifs in the 0.7–3.5 kcal/mol range, depending on the IDP. The average IDP folding penalty is 2.0 kcal/mol, which is in line with some previous indirect and more direct approaches.

2.2 Pre-folding and bound-state fuzziness reduce the folding penalty

It has been hypothesized that IDPs can employ different mechanisms to reduce the folding penalty such as pre-folding (IDP is already partially helical in the unbound state) and fuzziness (IDP remain partially disordered in the target-bound state). While the existence of these two phenomena is now clearly established, their role on the IDP folding penalty has, however, not been described quantitatively. Given that $\Delta G$ is state function it is possible to calculate the free energy for a transition between any states, even hypothetical states. To this aim, we define two hypothetical states where IDP pre-folding or fuzziness are absent (coil and rigid ensembles on Figure 1 shown to the left and right). The pre-folding contribution $\Delta G_{\text{prefold}}$ can be estimated by considering the transition from a hypothetical coil ensemble (no pre-folding, helicity = 0) to the native partially helical ensemble (helicity = $\alpha_{\text{UNBOUND}}$) (Equation 7). The $\Delta G_{\text{prefold}}$ contributions for all IDPs are listed in Table S1 and range from −0.5 to −2.5 kcal/mol, with the average value of −1.0 kcal/mol (Figure 2). These contributions thus show that without pre-folding (if IDPs would start from a coil conformation) the folding penalty would be even more unfavorable; on average by 1 kcal/mol. As expected, the values of $\Delta G_{\text{prefold}}$ correlate with the unbound IDP helicity, that is more helical IDPs have higher values of $\Delta G_{\text{prefold}}$ (Figure S2). In general, IDP pre-folding can reduce the folding penalty by about 30% ($\Delta G_{\text{fold}} - \Delta G_{\text{prefold}}$), compared to transitions where IDPs starts from a coil ensemble with zero helical structure.

To estimate the extent to which IDP bound-state heterogeneity and dynamics (fuzziness) reduce the folding penalty, we define a hypothetical rigid target-bound IDP ensemble (Figure 1, right). In this hypothetical ensemble all IDP residues (not only hotspots), which are helical in the structure of IDP-target complex are required to populate helical conformation. Consequently, the conformational heterogeneity and dynamics are reduced. The contribution due to fuzziness $\Delta G_{\text{fuzzy}}$ can be then estimated by considering the transition from a hypothetical rigid ensemble (less dynamics) to the native target-bound ensemble as defined by hotspot-constrained model (Equation 7). Note that in some cases, IDPs can establish many hotspots with their targets, therefore the native target-bound ensemble is already highly ordered and there may be no difference between the hypothetical rigid and the native, hotspot-constrained ensemble, leading to $\Delta G_{\text{fuzzy}} = 0$. In other words, the rigid ensemble can be considered as a limiting case of the hotspot-constrained ensemble when IDPs is constrained by many hotspots. The values of $\Delta G_{\text{fuzzy}}$ are listed in Table S1 and range from 0 (no fuzziness, the native ensemble is essentially rigid and ordered) to −1.7 kcal/mol (native ensemble is highly dynamic), with the average value of −0.8 kcal/mol (Figure 2). These results show that the energetic contribution of binding motif fuzziness can reduce the IDP folding penalty by roughly 30% ($\Delta G_{\text{fuzzy}}$/$\Delta G_{\text{fold}} - \Delta G_{\text{fuzzy}}$) relative to the process where IDP folds into a rigid ensemble. Collectively, the above analysis shows that pre-folding and target-bound dynamics can considerably reduce the IDP folding penalty. For a hypothetical process where IDPs would fold-upon-binding starting from a completely disordered conformation and end in a rigid ensemble the average folding penalty would be almost twice of what is observed (3.8 kcal/mol). Thus, the combined effect of pre-folding and fuzziness roughly halves the magnitude of IDP folding penalty.

2.3 Empirical correlation relates folding penalty to the difference in IDP helicity upon binding

Thermodynamic considerations suggest that the IDP folding penalty should correlate with the degree of IDP folding upon binding, that is the difference in IDP helicity: $\Delta \alpha = \alpha_{\text{BOUND}} - \alpha_{\text{UNBOUND}}$. Indeed, a common strategy of optimizing IDP-target affinity is to increase the unbound IDP helicity, thus reducing $\Delta \alpha$ and...
For several systems high affinities were observed. The correlation between folding penalty and gain in \( \Delta \alpha \) is very similar (we observed that the average per-residue folding propensity confirms the existence of a correlation between these scatter around \( \Delta \alpha \) to their similar amino acid composition. Using the values \( \Delta \alpha \) to be approximated as linear (0.4 < \( \Delta \alpha \) < 0.8, see Figure S3). The coefficients of linear equation are given in Equation (1).

This relation provides a useful way to estimate the \( \Delta G_{\text{fold}} \) contribution from the IDP helicity change (\( \Delta \alpha \)). In our case we used structural data from IDP-target complexes to determine \( \alpha_{\text{BOUND}} \) and CD data to determine \( \alpha_{\text{UNBOUND}} \). For the cases where high-resolution structure of the target-bound complex is not known, \( \alpha_{\text{BOUND}} \) can be also estimated from the difference CD spectra (complex-target, see Section 3). In such cases, however, it should be verified that the target does not contribute to the difference CD spectrum, that is the structural change must be only due to IDP folding. We also verified that the observed correlation between \( \Delta G_{\text{fold}} \) and \( \Delta \alpha \) is not due underlying correlation with either \( \alpha_{\text{UNBOUND}} \) or \( \alpha_{\text{BOUND}} \).

We find that \( \Delta G_{\text{fold}} \) does not correlate with either \( \alpha_{\text{UNBOUND}} \) or \( \alpha_{\text{BOUND}} \) alone (Figure S4).

**2.4 How folding penalty affects the IDP-target affinity**

The obtained values of IDP folding penalty, on average 2.0 kcal/mol, are comparable in magnitude to a strong IDP-target interaction, for example, formation of a hotspot (>1 kcal/mol). This suggests that, in order to reach high IDP-target affinities, the folding penalty could be simply compensated by additional hotspot interactions. To investigate how the values of folding penalty compare with the association free energy (\( \Delta G_{\text{association}} \)), we searched the literature and found data for 21 out of 30 IDP systems (Table S1). Note that both \( \Delta G_{\text{fold}} \) and \( \Delta G_{\text{association}} \) can be compared directly since these are standard free energies and as such correspond to the difference in free energy of reactants and products when all reaction components are at standard-state conditions (1 M concentration). However, the fraction of IDP associated to its target also depends on the protein concentration; therefore, the low affinity can be compensated by higher concentrations of target and IDP. Unfortunately, the protein concentrations in vivo are not known for most systems, therefore a comparison that would consider both the \( \Delta G_{\text{association}} \) and protein concentration together with the \( \Delta G_{\text{fold}} \) cannot be made at present. The average IDP-target association free energy for IDPs included in this study is ~8.7 kcal/mol and a similar value also appears in larger datasets of IDP-target affinities.\(^{15,24}\) For several systems high affinities were observed (\( K_D = 1 \text{nM} \)) and even picomolar affinities have been described for some IDPs that fold upon binding.\(^{19}\) The relative contribution of folding penalty, estimated as \( \Delta G_{\text{fold}}/\Delta G_{\text{association}} \) is on average 20%. However, the importance of the folding penalty depends strongly on the IDP-target system and its association affinity. For

\[
\Delta G_{\text{fold}} = (5.5 \pm 0.7 \Delta \alpha - 1.4 \pm 0.5) \text{ kcal/mol.} \quad (1)
\]
systems with low affinity ($K_D > 10 \ \mu M$), the folding penalty represents a major contribution to the association free energy ($\Delta G_{\text{fold}}/\Delta G_{\text{association}} > 50\%$), while for strong binders it is almost negligible ($\Delta G_{\text{fold}}/\Delta G_{\text{association}} < 10\%$) as it is readily compensated by the significant binding contribution.

The IDP-target affinity can be modulated by changing the IDP folding penalty. For example, mutating IDP residues by introduction of helix breakers (e.g., glycine) has been shown to significantly reduce the IDP-target affinity. On the other hand, substitutions with the helix promoting residues, such as alanine, leucine or arginine can increase IDP helicity, reduce the folding penalty and consequently increase the IDP-target affinity. Stabilization of helical binding motifs, particularly using the hydrocarbon stapling, has proven to be an efficient strategy for the rational design of peptide inhibitors of IDP-target interactions (reviewed in Walensky and Bird). In this strategy, a covalent bond (hydrocarbon staple) is introduced between two modified residues on the same side of helix (usually between residues $i$ and $i + 4$), which effectively stabilizes helical structure. As shown above in a certain range, the relation between $\Delta G_{\text{fold}}$ and $\Delta \alpha$ can be considered as linear, and based on Equation (1), the stabilization of helicity by 10% can reduce the average $\Delta G_{\text{fold}}$ by 0.55 kcal/mol. This translates into a roughly 2–3 times stronger IDP-target affinity. Interestingly, the same magnitude of stabilization has been observed recently for the c-Myc-KIX peptide inhibitor, which has been stabilized using several conservative K to R substitutions. In this modified variant IDP helicity increased by around 10% leading to the affinity increase of $-0.5$ kcal/mol. A similar, around $-0.2$ to $-0.3$ kcal/mol stabilization per 10% helicity increase have been also reported for other systems. Removing the folding penalty altogether (assuming the average value of $\Delta G_{\text{fold}} = 2$ kcal/mol) could theoretically lead to more than an order of magnitude increase of the IDP-target affinity. About an order of magnitude increase in affinity has been reported for different peptide inhibitors of BCL-2 family proteins, which were significantly stabilized using hydrocarbon staples. Thus, increase in IDP helicity gives a certain, but not unlimited, room for the affinity modulation in the design of competitive inhibitors of IDP-target interactions.

Taken together, we provide the first direct estimates of IDP folding penalty, a quantity that has eluded for many years. The presented analysis clarifies the roles of IDP pre-folding and bound-state dynamics (fuzziness) in reducing the folding penalty. The empirical relation in Equation (1) provides a simple way to approximately estimate the folding penalty form the change in IDP helicity upon binding, but also gives a quantitative tool that can be used optimization of IDP-target binding affinity via reduction of IDP folding penalty.

3 | MATERIALS AND METHODS

3.1 | Estimation of $\alpha_{\text{UNBOUND}}$ from the circular dichroism spectra

Published circular dichroism spectra of studied IDPs were analyzed as described previously. Briefly, the published spectra were digitalized to obtain the value of CD intensity at 222 nm. The intensity was then converted to the per-residue molar ellipticity [$\theta$] and the unbound IDP helicity was calculated as:

$$\alpha_{\text{UNBOUND}} = \frac{[\theta] - [\theta]_{\text{COIL}}}{[\theta]_{\text{HELIX}} - [\theta]_{\text{COIL}}}. \quad (2)$$

where $[\theta]_{\text{COIL}} = 2,220 - 53T$ and $[\theta]_{\text{HELIX}} = (-44,000 + 250T) (1 - 3/N)$ represent molar ellipticities at 222 nm of peptide bonds in coil or helix conformation respectively, $T$ is temperature in °C and $N$ is the number of residues in the peptide. The obtained $\alpha_{\text{UNBOUND}}$ value represents the fraction of helical residues in the IDP, that is, fractional helicity, which is listed in Table S1 for all IDPs.

3.2 | Analysis of IDP hotspot residues and estimation of $\alpha_{\text{BOUND}}$

Structures of IDP-target complexes were analyzed with PPCheck to identify hotspot residues, while KFC2 was used to assess the variability in hotspot determination and calculation of the corresponding error in $\Delta G_{\text{fold}}$ estimation. In all cases hotspot residues were also the one that adopted helical conformation according to STRIDE definitions (see Equation 6).

The fractional helicity in the target-bound state ($\alpha_{\text{BOUND}}$) was estimated from the structures of IDP-target complexes (PDB codes are listed in Table S1). Structure of IDP was analyzed with program STRIDE to identify residues in helical conformation and $\alpha_{\text{BOUND}}$ was calculated by dividing the number of residues in helical conformation with the total number of residues. In absence of high-resolution structures of IDP-target complex to estimate $\alpha_{\text{BOUND}}$, one may also estimate $\alpha_{\text{BOUND}}$ directly from CD spectra, as described before. Briefly, the CD spectra of target and IDP-target complex are recorded. The CD spectrum of the target-bound IDP is obtained as a difference spectrum (IDP-target) – (target). Importantly, the sample of IDP-target complex must be prepared at conditions where all IDP is target-bound and the
fraction of free IDP is negligible (sufficiently high concentrations of complex). The resulting spectra can then be analyzed using Equations (1) and (2).

3.3 Statistical-thermodynamic model

We used Ising-like model of helix-coil transition developed by Liefson and Roig.33 Briefly, the model assumes that the formation of helix can be separated into nucleation and elongation events. Helix nucleation is quantified by the nucleation constant \( \nu \) and describes the restriction of \( \Phi \) and \( \Psi \) dihedral angles to those corresponding to the helix conformation, but does not include formation of \( i-i+4 \) hydrogen bonds. Nucleation is an entropically unfavorable process and a commonly adopted value of \( \nu \) value is 0.048 for all residues.34 Helix propagation is quantified by the propagation constant \( w_i \), which describes transition of a residue from coil to the hydrogen-bonded helical conformation. This process is more favorable due to formation of \( i-i+4 \) hydrogen bonds, therefore the residues proceeding \( i \)-th residue are required to be in the helical conformation. The values of propagation constants depend on the amino acid type and several amino acid propensity scales have been published.14,34,35 Here, we use the scale from Pace and Sholtz.14 Since the scale is given as relative to alanine, we use \( w_{\text{Ala}} = 1.61 \) to convert it to absolute scale.34 Liefson and Roig defined eight possible residue triplet states with the corresponding statistical weights and developed a matrix formalism to assign these weights to all possible microstates in a given peptide. The statistical weight of each microstate is the product of different \( v \) and \( w_i \) weights. Commonly a \( 3 \times 3 \) from of matrix is used to assign these weights to each \( i \)-th residue:

\[
M_i = \begin{pmatrix}
  w_i & v & 0 \\
  0 & 0 & 1 \\
  v & v & 1 
\end{pmatrix}.
\] (3)

The partition function \( Z_U \) for a given for a polypeptide with \( N \) residues in its free, unbound state is obtained via multiplication of matrices in the order of the amino acid sequence till the last (\( N \)-th) residue:

\[
Z_U = (0 \ 0 \ 1) \prod_{i=1}^{i=N} M_i \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \sum_{i=N}^{i=1} M_i e^+.
\] (4)

End vectors (\( e \) and \( e^+ \)) ensure that the terminal residues are weighted either by 1 or \( v \), but never with the \( w \) weight. One of the central quantities calculated from the partition function that can be compared to the experimental data is the average peptide helicity:

\[
\alpha = \frac{d\ln Z_U}{d\ln w} \frac{1}{N-2}.
\] (5)

Here, the \( N-2 \) term refers to the maximal number of helical residues in the peptide with \( N \) residues. The model-calculated helicity obtained using Equation (5) and the set of \( w_i \) weights from the propensity scale can be directly compared to the experimental value of \( \alpha_{\text{UNBOUND}} \). However, the set of helix propagation constants \( w_i \) alone appear to be insufficient to accurately reproduce the absolute values of IDP helicity in the unbound state. Even much sophisticated algorithms based on the Liefson-Roig formalism such as AGADIR with many additional parameters other than \( w_i \) (e.g., capping and sidechain interaction), systematically underestimate the absolute helicity of IDPs, as we have shown recently.13 To amend for these missing interactions, we use an empirical calibration factor that multiplies all \( w_i \) constants by a constant (determined for each IDP separately) such that the model calculated helicity (Equation (5)) corresponds to the experimental one. This factor is determined using Brent minimization method that minimizes the difference between model calculated and experimental helicity. These empirical factors range between 0.98 to 1.97, with the average value of 1.37 and are listed in Table S1.

We recently developed a modified version of helix-coil model to describe target-bound IDP ensembles.20 This hotspot-constrained helix-coil model assumes that in the target-bound ensemble a subset of residues that from strong>IDP contacts (hotspots) must adopt a fix conformation, which in the case of \( \alpha \)-binding motifs is the helical conformation. The partition function \( Z_B \) is defined as set of microstates \( m \):

\[
Z_B = \{m \in Z_U | i \in (H \cap X) = w_i \}.
\] (6)

The \( Z_B \) is a subset of \( Z_U \) containing call \( i \) residues that are both helical and hotspots (\( i \) is element of \( H \) and \( X \) sets) must have the weight equal to \( w_i \). The technical procedure how to obtain \( Z_B \) as a subset of \( Z_U \) is outlined in Hadži et al.20 A modification of Equation (6) is used to define the hypothetical rigid target-bound ensemble (see Figure 1) and obtain its partition function \( Z_{B\text{rigid}} \). In this case, \( i \) residues are required to be elements of helical set (\( i \in (H) \)), but not also hotspot set. In other words, all residues which adopt helix conformation in the IDP-target complex have the weight equal to \( w_i \). This means that there are fewer possible microstates in the ensemble.
The differences in free energy for the transitions shown in Figure 1 were calculated as:

\[
\Delta G_{\text{fold}} = -RT \ln \frac{Z_B}{Z_U}; \Delta G_{\text{prefold}} = -RT \ln \frac{Z_U}{(1 + v)^{N-2}}; \\
\Delta G_{\text{fuzzy}} = -RT \ln \frac{Z_B}{Z_{B,\text{rigid}}}. (7)
\]

3.4 Uncertainties related to the estimation of the folding penalty

There are two main sources of errors that affect the estimation of folding penalty. The first is related to estimation of IDP helix propensity and stems from (a) the uncertainty in the helix propagation constants \( w_i \) and (b) from the measurement errors of CD intensity. Here, we used the \( w_i \) constants from the consensus helix propensity scale of Pace and Scholtz. In order to estimate the uncertainty in \( \Delta G_{\text{fold}} \) originating from the use of \( w_i \) constants, we recalculated the \( \Delta G_{\text{fold}} \) using two alternative helix propensity scales. The average difference in the \( \Delta G_{\text{fold}} \) calculated using these. Two scales (with respect to the one calculated using Pace\&Scholtz scale) is 0.1 and 0.4 kcal/mol, respectively. The error in the CD signal intensity is mainly the result of inaccurate concentration of sample, since the CD signals are converted to molar units and compared to the molar per residue ellipticities of pure helix and coil conformations. This leads to the uncertainty in the calculated helicity and subsequently in the IDP helix propensity. Other sources of error related to CD measurements arise from the inaccuracy of the signal intensity, contributions arising from aromatic residues and a possible presence of other, non-helical conformations that would contribute to the CD intensity signal. We tentatively estimate that all these factors together likely add around ±5% error to the \( \alpha_{\text{UNBOUND}} \) value. A recalculations of folding penalties values using IDP helicities shifted by the value of CD error, shows that this contributes ±0.4 kcal/mol of average error to \( \Delta G_{\text{fold}} \) estimates.

The second source of error in the folding penalty estimation is related to the representation of the target-bound IDP. Here, the problem concerns the uncertainty associated with the determination of IDP-target hotspots using computational methods. To address this, we use a different computational method KFC2 to identify IDP hotspots, and recalculate the folding penalty using hotspot distributions predicted by KFC2 tool. The folding penalty is not very different from the one calculated using PPCheck hotspots, the average difference is 0.3 kcal/mol, while for three IDPs the folding penalty differs by around 1 kcal/mol. All together these factors contribute to around 1.2 kcal/mol of error to the folding penalty values listed in Table 1.

AUTHOR CONTRIBUTIONS
San Hadži: Conceptualization (lead); investigation (lead); writing – original draft (lead). Jurij Lah: Validation (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS
This work was supported by the grants P1-0201 and J1-1706 from the Slovenian Research Agency to San Hadži and Jurij Lah. Open access funding enabled and organized by Projekt DEAL.

ORCID
San Hadži https://orcid.org/0000-0002-5615-7918

REFERENCES
1. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ. Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform. 2000;11:161–171.
2. Uversky VN. What does it mean to be natively unfolded? Eur J Biochem. 2002;269:2–12.
3. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans. 2016;44:1185–1200.
4. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009;19(1):31–38.
5. Yan J, Dunker AK, Uversky VN, Kurgan L. Molecular recognition features (MoRFs) in three domains of life. Mol Biosyst. 2016;12(3):697–710.
6. Tompa P, Fuxreiter M. Fuzzy complexes: Polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 2008;33(1):2–8.
7. Fuxreiter M. Fold or not to fold upon binding - does it really matter? Curr Opin Struct Biol. 2019;54:19–25.
8. Sharma R, Raduly Z, Miskei M, Fuxreiter M. Fuzzy complexes: Specific binding without complete folding. FEBS Lett. 2015;589:2533–2542.
9. Dunker AK, Garner E, Guilliot S, et al. Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. Pac Symp Biocomput. 1998;473–484.
10. Flock T, Weatheritt RJ, Latysheva NS, Babu MM. Controlling entropy to tune the functions of intrinsically disordered regions. Curr Opin Struct Biol. 2014;26:62–72.
11. Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J Mol Biol. 2004;338:1015–1026.
12. Kim DH, Han KH. Transient secondary structures as general target-binding motifs in intrinsically disordered proteins. Int J Mol Sci. 2018;19:3614.
13. Hadži S, Purić S, Zavrtanik U, Vranken W. What stabilizes pre-folded structures in the intrinsically disordered α-helical binding motifs? bioRxiv. 2022;478151. https://doi.org/10.1101/2022.01.28.478151.
14. Pace CN, Scholtz JM. A helix propensity scale based on experimental studies of peptides and proteins. Biophys J. 1998;75(1):422–427.
15. Teilum K, Olsen JG, Kragelund BB. Globular and disordered-the non-identical twins in protein-protein interactions. Front Mol Biosci. 2015;9(2):40.
16. Drobnak I, De Jonge N, Haesaerts S, Vesnaver G, Loris R, Lah J. Energetic basis of uncoupling folding from binding for an intrinsically disordered protein. J Am Chem Soc. 2013;135(4):1288–1294.
17. Theisen FF, Staby L, Tidemand FG, et al. Quantification of conformational entropy unravels effect of disordered flanking region in coupled folding and binding. J Am Chem Soc. 2021;143(36):14540–14550.
18. Rajasekaran N, Gopi S, Narayan A, Naganathan AN. Quantifying protein disorder through measures of excess conformational entropy. J Phys Chem B. 2016;120:4341–4350.
19. Zavrtanik U, Hadži S, Lah J. Unraveling the thermodynamics of ultra-tight binding of intrinsically disordered proteins. Front Mol Biosci. 2021;31(8):726824.
20. Hadži S, Loris R, Lah J. The sequence-ensemble relationship in fuzzy protein complexes. Proc Natl Acad Sci U S A. 2021;118(37):e2020562118.
21. Sukhwal A, Sowdhamini R. PPcheck: A webserver for the quantitative analysis of protein protein interfaces and prediction of residue hotspots. Bioinform Biol Insights. 2015;9:141–151.
22. Spolar RS, Record MT Jr. Coupling of local folding to site-specific binding of proteins to DNA. Science. 1994;263(5148):777–784.
23. Hadži S, Mernik A, Podlipnik Č, Loris R, Lah J. The thermodynamic basis of the fuzzy interaction of an intrinsically disordered protein. Angew Chem Int Ed Engl. 2017;56(46):14494–14497.
24. Huang Y, Liu Z. Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chemistry. 2013;19(14):4462–4467.
25. Poosapati A, Gregory E, Borcherds WM, Chemes LB, Daughdrill GW. Uncoupling the folding and binding of an intrinsically disordered protein. J Mol Biol. 2018;430:2389–2402.
26. Suetaka S, Oka Y, Kunihara T, Hayashi Y, Arai M. Rational design of a helical peptide inhibitor targeting c-Myb–KIX interaction. Sci Rep. 2022;12:816.
27. Iešmantavičius V, Dogan J, Jemth P, Teilum K, Kjaergaard M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. Angew Chemie – Int Ed. 2014;53:1548–1551.
28. Walensky LD, Bird GH. Hydrocarbon-stapled peptides: Principles, practice, and progress. J Med Chem. 2014;57(15):6275–6288.
29. Walensky LD, Kung AL, Escher I, et al. Activation of apoptosis in vivo by a hydrocarbon-stapled BH3 helix. Science. 2004;305(5689):1466–1470.
30. Errington N, Iqbalsyah T, Doig AJ. Structure and stability of the alpha-helix: Lessons for design. Methods Mol Biol. 2006;340:3–26.
31. Zhu X, Mitchell JC. KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density and plasticity features. Proteins. 2011;79(9):2671–2683.
32. Heing M, Frishman D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. 2004;32:W500–W502.
33. Lifson S, Roig A. On the theory of helix-coil transition in polypeptides. J Chem Phys. 1961;34:1963–1974.
34. Chakrabartty A, Kortemme T, Baldwin RL. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. Protein Sci. 1994;3(5):843–852.
35. Andersen NH, Tong H. Empirical parameterization of a model for predicting peptide helix/coil equilibrium populations. Protein Sci. 1997;6:1920–1936.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hadži S, Lah J. The free energy folding penalty accompanying binding of intrinsically disordered α-helical motifs. Protein Science. 2022;31(7):e4370. [https://doi.org/10.1002/pro.4370](https://doi.org/10.1002/pro.4370)