

Published in final edited form as:

Curr Opin Struct Biol. 2011 June ; 21(3): 441–446. doi:10.1016/j.sbi.2011.02.005.

Evolution and Disorder

Celeste J. Brown^a, Audra K. Johnson^{a,1}, A. Keith Dunker^b, and Gary W. Daughdrill^c

^aDepartment of Biological Sciences, IBEST, University of Idaho, Moscow, ID 83844-3051; celesteb@uidaho.edu

^bCenter for Computational Biology & Bioinformatics, Biochemistry and Molecular Biology, Indiana University Schools of Medicine & Informatics, 410 W. 10th Street, Suite 5000, Indianapolis, IN 46202; kedunker@iupui.edu

^cDepartment of Cell Biology, Microbiology and Molecular Biology, Center for Drug Discovery and Innovation, 3720 Spectrum Blvd. Suite 321, University of South Florida, Tampa, FL 33612; gdaughdrill@usf.edu

Abstract

The evolution of disordered proteins or regions of proteins differs from ordered proteins because of the differences in their sequence composition, intramolecular contacts and function. Recent assessments of disordered protein evolution at the sequence, structural and functional levels support this hypothesis. Disordered proteins have a different pattern of accepted point mutations, exhibit higher rates of insertions and deletions, and generally, but not always, evolve more rapidly than ordered proteins. Even with these high rates of sequence evolution, a few examples have shown that disordered proteins maintain their flexibility under physiological conditions, and it is hypothesized that they maintain specific structural ensembles.

Introduction

Intrinsically disordered proteins (IDPs) perform essential functions in organisms from viruses to vertebrates, and their improper functioning is responsible for numerous disease states in humans, including some cancers and neurodegenerative diseases [1-6]. These proteins or regions of proteins do not form a compact globular structure and populate broad conformational ensembles that are characterized by molecular motions on multiple timescales [7]. IDPs have a different amino acid composition than ordered proteins, and they often are composed of low complexity sequences [8,9]. The various functions of IDPs range from flexible linkers, whose sole functional constraint appears to be the maintenance of flexibility, to displaying protein modification sites, in which the sites are functionally important, to molecular recognition, in which large regions of the IDP interacts with other proteins [2,4,5].

© 2011 Elsevier Ltd. All rights reserved.

corresponding author: Celeste J. Brown Department of Biological Sciences PO Box 443051 University of Idaho Moscow, ID 83844-3051 celesteb@uidaho.edu phone: 01-208-885-4012 fax: 01-208-885-7905.

¹current address: Department of Genome Sciences, Foege Building S-250, Box 355065, 3720 15th Ave NE, Seattle WA 98195-5065; audrakj@uw.edu

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The sequences of IDPs are expected to be less constrained because they do not form intramolecular, long-range contacts [10]. However, the sequence evolution of some regions within IDPs may be under selection for specific functions. For instance, protein modification sites may be constrained because changing these sites will have a deleterious effect on signaling events [11]. In addition, regions involved in protein-protein interactions are possibly constrained by intermolecular contacts [12]. Additionally, recent evolutionary studies suggest that function is not the sole factor constraining protein evolution [13]. Therefore, looking at the evolution of IDPs not only provides information about their evolution, but about genome evolution in general.

Accepted point mutations differ between disordered and ordered proteins

One way to describe how proteins evolve over time is to compare the sequences of homologous proteins, where homology depends upon the sequences having shared a common ancestor. From these comparisons, models of protein evolution can be inferred based upon the frequency with which different amino acids occur at the same position among the homologues. These models reflect the point mutations that are accepted by evolution because they have a neutral or positive effect upon protein function [14]. Evolutionary models are commonly used to identify or align homologous proteins, however, they also provide information about how proteins evolve.

Two recent studies show that there are differences in the accepted point mutations between disordered and ordered proteins. One study compared evolutionary models inferred from disordered and ordered protein sequences [15]. From these models, it was shown that disordered sequences have a greater likelihood of changing and that the changes are non-conservative compared with ordered sequences. The second study compared the conservation of secondary structure and intrinsic disorder when sequences were evolved under three evolutionary models [16]. This study showed that when disordered protein is evolved by order-biased evolutionary models, the disordered regions are not conserved. Two of the models used in this study were well-known substitutions matrices, PAM120 and BLOSUM62, the third model chose substitutions based on the frequency of the amino acids within the dataset being evolved. Each of these models reflect the evolution of ordered proteins. PAM120 was explicitly based upon protein sequences for which structures had been determined [14]. BLOSUM62 is implicitly based upon structured protein due to the removal of ambiguous alignments that contained gaps [17], when insertions and deletions occur more often in IDPs (Figure 1). In the third model based on amino acid frequencies, the frequencies were taken from the entire protein, not just the disordered region of the protein, which again biases the model away from disordered sequence [18]. Given that these models are biased toward evolutionary changes found in ordered protein, it is not surprising that disordered regions were lost. Additionally, studies of conserved disordered domains indicate that disordered regions, if not their sequences, are maintained over various evolutionary time scales [12,19-24]. Hence, disordered proteins evolve differently from ordered proteins to maintain their disordered structure.

Rates of evolution are generally but not always high in disordered proteins

Another way to describe how proteins evolve over time is to determine the rate at which their amino acid sequences change. An early study of a limited number of proteins containing both ordered and disordered regions showed that the disordered regions of most of these proteins evolve more rapidly than their ordered regions [25]. Because this study compared evolutionary rates within a protein, differences due to transcription and translation were held constant. Other early studies indicated that disordered proteins evolve rapidly by repeat expansions, especially in single amino acid repeats [26,27].

These early studies are supported by recent comparisons among the genomes of several organisms. Genomic studies necessarily rely upon prediction of disorder and order, since relatively few disordered proteins have been characterized structurally. These genomic studies indicate that disordered proteins, in general, evolve rapidly when compared with ordered proteins, although specific instances of slowly-evolving disordered proteins are found [24,28-30]. Indeed, a significant fraction of Pfam domains are predicted to be disordered, and the sequence conservation of these domains is similar to that of Pfam domains predicted to be ordered [21,31]. Identifying structural or biological factors to explain why various disordered regions exhibit high or low evolutionary rates remains a challenging, unsolved problem.

Several recent genomic studies describing biophysical factors that correlate with rates of protein evolution may help in solving this challenging problem. One factor that is found consistently in these studies is the generally slow rate of evolution for genes that are highly expressed, and this slow rate is at both the DNA and protein sequence levels [32]. Recent studies have shown that disordered proteins are very tightly regulated and their levels of gene expression are quite low, which correlates well with high rates of evolution for genes with low expression levels [33,34]. One question that can be asked is what the causative factors are in these correlations. In the case of genes with high expression levels, the slow rates of evolution are hypothesized to be due to the high cost of overcoming protein misfolding [13]. Since disordered proteins do not fold, misfolding is not a pressure that would act on these proteins no matter what their level of expression. Additionally, there appears to be strong selection to reduce the negative consequences of large concentrations of disordered protein by tightly controlling their expression at all levels of transcription, translation and protein degradation [33]. These results suggest that the correlation between low gene expression levels and high rates of evolution may be due to weakly-expressed, rapidly-evolving disordered proteins. Conversely, those disordered proteins that do not evolve rapidly may be less common, highly-expressed proteins.

Evolution of dynamic behavior

Do the large differences in the sequences of homologous IDPs lead to differences in dynamic behavior? To date only two studies have experimentally characterized the dynamic behavior of an IDP family [19,23]. In a study from our groups, NMR relaxation experiments were used to investigate the dynamic behavior of a conserved linker domain from the 70 kDa subunit of replication protein A (RPA70). We showed that dynamic behavior is conserved in the face of negligible sequence conservation. We also showed that dynamic behavior is under selection, even for very flexible IDPs that do not appear to interact with other proteins. In the study by Gage and co-workers, mesophilic and thermophilic variants of the anti-sigma factor, FlgM, were analyzed using circular dichroism. In this study, the thermophilic variant exhibited significant structure (resembling a molten globule) at mesophilic temperatures, and this structure was successively eliminated as temperature was increased. By contrast the mesophilic variant shows no evidence for structure at mesophilic temperatures. It is unclear what molecular features are responsible for this behavior.

The dynamic behavior of IDPs is a poorly understood molecular property that appears to be necessary for function and is probably under multiple levels of evolutionary selection. To understand the evolution of dynamic behavior it is necessary to determine realistic structural ensembles of IDPs from the same protein families. Intrinsically disordered proteins form ensembles of structures that can have a broad range of compactness, secondary structure content, and conformational dynamics. A general model based on local interactions has been proposed to describe the structural ensembles of intrinsically disordered proteins [35-37]. In this model, ensembles are generated using a database of dihedral angles that are observed in

the loop regions of folded proteins. Ensembles generated using this method can predict residual dipolar couplings and small-angle X-ray scattering data for intrinsically disordered proteins with high accuracy. The results from both of these studies suggest that any conformational biasing observed in the unfolded state is defined locally. Interestingly, one of the same groups had to modify the method described above to accurately model intrinsically-disordered β -synuclein [38]. The approach was modified to account for the presence of long-range structure by selecting models that satisfied long-range distance restraints. Transient long-range structure was also detected for the intrinsically disordered transactivation domain of the tumor-suppressor protein, p53 and the microtubule-associated protein, tau [39-41]. The functional consequences of the transient long-range structure observed for some intrinsically disordered proteins have not been determined. However, the presence of transient long-range structure is consistent with the notion that intrinsically disordered proteins have been selected to perform functions that require them to maintain a specific structural ensemble. It would be very surprising if all of the different selective pressures acting on intrinsically disordered proteins resulted in a single category of structural ensemble.

Evolution of functional sites within IDPs

One important function of IDPs is displaying sites for protein modification, for example, protein phosphorylation sites [5,42]. The evolutionary characteristics of IDPs, especially their rapid rate of evolution and their propensity for insertions and deletions (Figure 1), may result in these modification sites evolving differently over time than if they were in ordered regions. On the most recent time scale measured by single nucleotide polymorphisms within a species, phosphorylated serines and threonines seem to evolve at the same rate as their non-phosphorylated counterparts [43]. In comparisons of more distantly-related species, such as among mammals, phosphorylated sites are more conserved than non-phosphorylated sites, but phosphorylated sites within IDPs evolve more rapidly than sites in ordered regions [11,43]. In the most distant comparisons, such as between fungi and flies, the positions of phosphorylated sites within homologous IDPs shift, conserving a functional site somewhere within the IDP [44,45]. Thus the often rapid rate of evolution and the greater proportion of insertions and deletions in IDPs influence the perceived rate of evolution of phosphorylation sites at greater divergence times.

Evolution of coupled folding and binding

The most commonly observed function of IDPs is the folding of disordered segments when they bind to other proteins and/or DNA. This coupling of folding and binding is thought to balance the specificity and affinity of protein-protein and protein-DNA interactions, which is essential for the fidelity of many biological processes. If binding is not specific, then the wrong interactions will occur but specific binding requires an extensive molecular interface, which results in a high affinity. If the affinity is too high, then the interaction will last too long and processing will be disrupted. One solution to this problem of balancing specificity and affinity was the evolution of coupled folding and binding [46-48]. When folding and binding are coupled, one (or both) of the interacting partners maintains a state of high conformational entropy when not bound. Some of this conformational entropy is lost during binding because a more ordered structure is formed. This entropy loss reduces the overall affinity of the interaction and this affinity loss can be offset by additional compensating interactions in the bound state, leading to an increase in specificity. Highly specific interactions with modest affinities can be accommodated with this mechanism. Recent theoretical studies lend support to this hypothesis but a rigorous experimental test has not been performed [49].

The term “coupled folding and binding” suggests that both folding and binding occur simultaneously. This is probably not the case and there are at least two step-wise mechanisms, referred to as conformational selection (CS) and induced folding (IF), used to describe the process [48,50]. Both mechanisms assume that an equilibrium distribution of structures exist and some of these structures resemble the bound state and some of these structures do not resemble the bound state (Figure 2). An IDP is said to bind by CS if the structures that resemble the bound state occur at a high enough probability to selectively interact with the binding partner. By contrast, the IDP will bind by IF if there is a very low population of structures that resemble the bound state and all structures are competent for binding in a low affinity state. In this case, just about any structure in the ensemble can bind, and this binding is followed by a conformational rearrangement to form the specific contacts necessary for the final bound state. These two mechanisms are not mutually exclusive and represent two extreme cases of what is possible. In reality most IDP mechanisms probably share some features of both pathways. We propose that the conformational flexibility of the free IDP will determine the probability for a particular pathway. If the flexibility is high, then IF is more probable. If the flexibility is low, then CS is more probable.

Only a few groups have rigorously investigated the mechanism for coupled folding and binding reactions and these studies support the hypothesis that the relative flexibility of the free state of an IDP will determine whether the binding mechanism is induced fit or conformational selection. In a study by Sugase and Wright, the interaction between the phosphorylated kinase inducible activation domain (pKID) of the transcription factor CREB and the KIX domain of the CREB binding protein was investigated using relaxation dispersion [51]. Relaxation dispersion is a type of NMR experiment that allows one to define the on and off rates for the binding of individual amino acids. This information can be combined with knowledge of the conformational flexibility in the free state to distinguish between binding mechanisms that are predominantly induced fit or conformational selection. Interestingly, the results of this study showed evidence for both binding mechanisms in the two different alpha helices that form when pKID binds to KIX. In another study by Levy and Woynes, a computational approach was used to simulate transition state ensembles for protein-protein interactions where folding and binding are coupled and where folding precedes binding [52]. While this study did not strictly distinguish between conformational selection and induced fit binding mechanisms, it did show a strong correlation between the flexibility of the free state and the kinetics of binding. Binding occurs more rapidly when the flexibility of the free state is high, which is consistent with the induced fit binding mechanism, and more slowly when the flexibility of the free state is low, which is consistent with the conformational selection binding mechanism. A relationship between the flexibility of the free state and binding mechanism was also observed for different mutants of staphylococcal nuclease [53]. In this study, a single point mutant was used to switch the binding mechanism from conformational selection to induced fit. More work is required to determine whether there is a connection between IDP flexibility and their binding mechanisms.

Conclusions

Although disordered proteins often evolve at a fast rate and have different accepted point mutations, the conservation of their flexibility and function indicate that these proteins have important physiological roles in all organisms from viruses to vertebrates. As more comparisons are made among homologous disordered regions and proteins, the evolutionary constraints on flexibility and disordered protein functions, such as coupled folding and binding, will be elucidated. Aside from understanding the evolution of disordered proteins, per se, it is important to study the role of disordered proteins in the context of evolutionary theories that were developed to explain the evolution of ordered proteins.

Acknowledgments

CJB is supported by the National Institutes of Health (P20RR16448). GWD is supported by the National Science Foundation (Award # 0939014) and the American Cancer Society (RSG-07-289-01-GMC). AKD has been supported by the National Institutes of Health (R01LM007688 and R02GM071714) and more recently by the National Science Foundation (MCB-0849803).

References

1. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 1999; 293:321–331. [PubMed: 10550212]
2. Uversky V, Gillespie J, Fink A. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins.* 2000; 41:415–427. [PubMed: 11025552]
3. 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.
4. Tompa P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 2002; 27:527–533. [PubMed: 12368089]
5. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry.* 2002; 41:6573–6582. [PubMed: 12022860]
6. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol.* 2002; 323:573–584. [PubMed: 12381310]
7. Daughdrill, GW.; Pielak, GJ.; Uversky, V.; Cortese, MS.; Dunker, AK. Natively Disordered Proteins. In: Buchner, J.; Kiefhaber, T., editors. *Handbook of Protein Folding.* Vol. vol 2. Wiley-VCH; 2005. p. 275-357.
8. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophys J.* 2007; 92:1439–1456. [PubMed: 17158572]
9. Romero PR, Obradovic Z, Dunker AK. The complexity of disorder. *Biophysics Journal.* 1999; 76(Part 2 of 2):A1–A525.
10. Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics.* 2008; 9(Suppl 2):S1.
11. Chen SC, Chen FC, Li WH. Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. *Mol Biol Evol.* 2010; 27:2548–2554. [PubMed: 20534707]
12. Toth-Petroczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M. Malleable machines in transcription regulation: the mediator complex. *PLoS Comput Biol.* 2008; 4:e1000243. [PubMed: 19096501]
13. Wilke CO, Drummond DA. Signatures of protein biophysics in coding sequence evolution. *Curr Opin Struct Biol.* 2010; 20:385–389. [PubMed: 20395125]
14. Dayhoff, MO.; Schwartz, RM.; Orcutt, BC. A model of evolutionary change in proteins. In: Dayhoff, MO., editor. *Atlas of Protein Sequence and Structure.* Vol. vol 5. Natl. Biomed. Res. Found.; 1978. p. 345-352.
- *15. Brown CJ, Johnson AK, Daughdrill GW. Comparing models of evolution for ordered and disordered proteins. *Mol Biol Evol.* 2010; 27:609–621. [PubMed: 19923193] Models of protein evolution are inferred for characterized ordered and disordered protein families. IDPs show unique patterns of amino acid substitutions when compared to ordered proteins.
16. Schaefer C, Schlessinger A, Rost B. Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics.* 2010; 26:625–631. [PubMed: 20081223]
17. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A.* 1992; 89:10915–10919. [PubMed: 1438297]
18. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, et al. DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* 2007; 35:D786–793. [PubMed: 17145717]

19. Daughdrill GW, Narayanaswami P, Gilmore SH, Belczyk A, Brown CJ. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. *J Mol Evol.* 2007; 65:277–288. [PubMed: 17721672]
20. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: II. Functions of conserved disorder. *J Proteome Res.* 2006; 5:888–898. [PubMed: 16602696]
21. Chen JW, Romero P, Uversky VN, Dunker AK. Conservation of intrinsic disorder in protein domains and families: I. A database of conserved predicted disordered regions. *J Proteome Res.* 2006; 5:879–887. [PubMed: 16602695]
22. Ayme-Southgate AJ, Southgate RJ, Philipp RA, Sotka EE, Kramp C. The myofibrillar protein, projectin, is highly conserved across insect evolution except for its PEVK domain. *J Mol Evol.* 2008; 67:653–669. [PubMed: 18982379]
23. Molloy RG, Ma WK, Allen AC, Greenwood K, Bryan L, Sacora R, Williams L, Gage MJ. Aquifex aeolicus FlgM protein exhibits a temperature-dependent disordered nature. *Biochim Biophys Acta.* 2010; 1804:1457–1466. [PubMed: 20298817]
24. Toth-Petroczy A, Meszaros B, Simon I, Dunker AK, Uversky VN, Fuxreiter M. Assessing conservation of disordered regions in proteins. *The Open Proteomics Journal.* 2008; 1:46–53.
25. Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, Williams CJ, Dunker AK. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol.* 2002; 55:104–110. [PubMed: 12165847]
26. Huntley M, Golding GB. Evolution of simple sequence in proteins. *J Mol Evol.* 2000; 51:131–140. [PubMed: 10948269]
27. Tompa P. Intrinsically unstructured proteins evolve by repeat expansion. *Bioessays.* 2003; 25:847–855. [PubMed: 12938174]
28. Simon M, Hancock JM. Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins. *Genome Biol.* 2009; 10:R59. [PubMed: 19486509]
29. Lin YS, Hsu WL, Hwang JK, Li WH. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol.* 2007; 24:1005–1011. [PubMed: 17264066]
- **30. Liu J, Zhang Y, Lei X, Zhang Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biol.* 2008; 9:R69. [PubMed: 18397526]
Fast frequency rates of IDPs is evident from the high frequency of single nucleotide polymorphisms that affect the amino acid sequence within species.
31. Tompa P, Fuxreiter M, Oldfield CJ, Simon I, Dunker AK, Uversky VN. Close encounters of the third kind: disordered domains and the interactions of proteins. *Bioessays.* 2009; 31:328–335. [PubMed: 19260013]
- *32. Xia Y, Franzosa EA, Gerstein MB. Integrated assessment of genomic correlates of protein evolutionary rate. *PLoS Comput Biol.* 2009; 5:e1000413. [PubMed: 19521505] Using a statistical analysis comparing protein evolutionary rate with a large number of protein properties, it was shown that IDPs are negatively correlated with slow evolutionary rates.
33. Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science.* 2008; 322:1365–1368. [PubMed: 19039133]
34. Vavouri T, Semple JI, Garcia-Verdugo R, Lehner B. Intrinsic protein disorder and interaction promiscuity are widely associated with dosage sensitivity. *Cell.* 2009; 138:198–208. [PubMed: 19596244]
35. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A.* 2005; 102:17002–17007. [PubMed: 16284250]
36. Jha AK, Colubri A, Freed KF, Sosnick TR. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc Natl Acad Sci U S A.* 2005; 102:13099–13104. [PubMed: 16131545]
- **37. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci U S A.* 2008; 105:5762–5767. [PubMed:

- 18391200] Applies a novel method for estimating the structural ensembles of IDPs to the disordered transactivation domain of the tumor suppressor, p53.
38. Bernado P, Bertocini CW, Griesinger C, Zweckstetter M, Blackledge M. Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings. *J Am Chem Soc.* 2005; 127:17968–17969. [PubMed: 16366524]
 39. Lowry DF, Hausrath AC, Daughdrill GW. A robust approach for analyzing a heterogeneous structural ensemble. *Proteins.* 2008; 73:918–928. [PubMed: 18536020]
 40. Vise P, Baral B, Stancik A, Lowry DF, Daughdrill GW. Identifying long-range structure in the intrinsically unstructured transactivation domain of p53. *Proteins.* 2007; 67:526–530. [PubMed: 17335006]
 41. Lowry DF, Stancik A, Shrestha RM, Daughdrill GW. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins.* 2008; 71:587–598. [PubMed: 17972286]
 42. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res.* 2004; 32:1037–1049. [PubMed: 14960716]
 43. Landry CR, Levy ED, Michnick SW. Weak functional constraints on phosphoproteomes. *Trends Genet.* 2009; 25:193–197. [PubMed: 19349092]
 44. Holt LJ, Tuch BB, Villen J, Johnson AD, Gygi SP, Morgan DO. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science.* 2009; 325:1682–1686. [PubMed: 19779198]
 45. Tan CS, Jorgensen C, Linding R. Roles of “junk phosphorylation” in modulating biomolecular association of phosphorylated proteins? *Cell Cycle.* 2010; 9:1232. [PubMed: 20404523]
 46. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol.* 2002; 12:54–60. [PubMed: 11839490]
 47. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry.* 2005; 44:12454–12470. [PubMed: 16156658]
 - *48. Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol.* 2009; 19:31–38. [PubMed: 19157855] Detailed insights into the mechanism of coupled folding and binding are provided by NMR relaxation dispersion measurements.
 - **49. Liu J, Faeder JR, Camacho CJ. Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc Natl Acad Sci U S A.* 2009; 106:19819–19823. [PubMed: 19903882] Development of a theory to predict the relationship between binding affinity and disorder tendency.
 50. Dunker AK, Garner E, Guillot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac. Symp. Biocomput.* 1998; 3:473–484. [PubMed: 9697205]
 51. Sugase K, Dyson HJ, Wright PE. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature.* 2007; 447:1021–1025. [PubMed: 17522630]
 52. Levy Y, Cho SS, Onuchic JN, Wolynes PG. A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J Mol Biol.* 2005; 346:1121–1145. [PubMed: 15701522]
 53. Onitsuka M, Kamikubo H, Yamazaki Y, Kataoka M. Mechanism of induced folding: Both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants. *Proteins.* 2008; 72:837–847. [PubMed: 18260114]
 54. Gely S, Lowry DF, Bernard C, Jensen MR, Blackledge M, Costanzo S, Bourhis JM, Darbon H, Daughdrill G, Longhi S. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J Mol Recognit.* 2010; 23:435–447. [PubMed: 20058326]

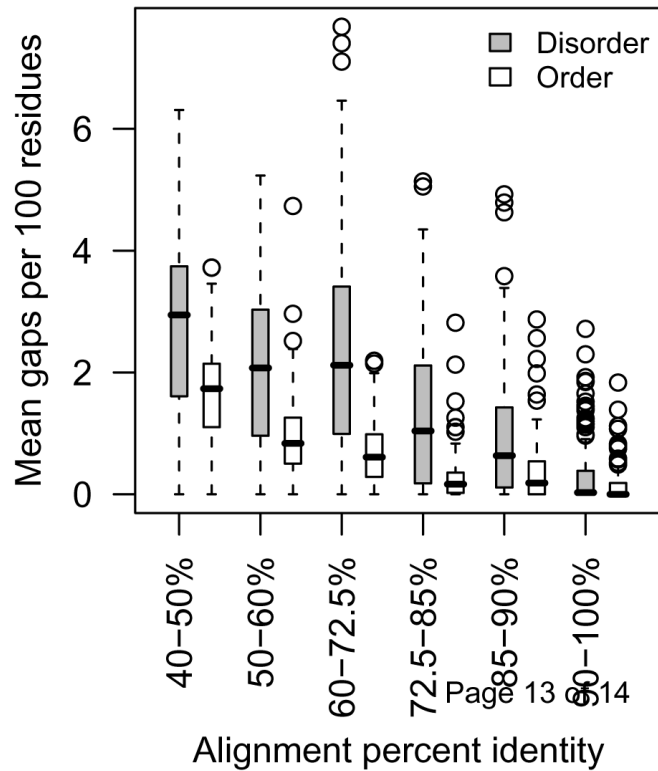


Figure 1.

Disordered proteins have more insertions and/or deletions than ordered proteins.

Homologues within families of structurally-characterized disordered regions or completely ordered proteins were aligned [15], and the alignments were compared at various levels of percent identity for the number of gaps of any length within the alignment. The box plots show median (dark bar), quartiles (grey boxes), 1.5x the interquartile range (dashed lines) and means that were outside of this range (circles). There are significantly more gaps within disordered regions of proteins than within the ordered proteins at all levels of percent identity (Wilcoxon rank sum tests, $p < 0.01$). (A. K. Johnson, MS Thesis: Comparing models of evolution between ordered and disordered proteins, University of Idaho, 2008.)

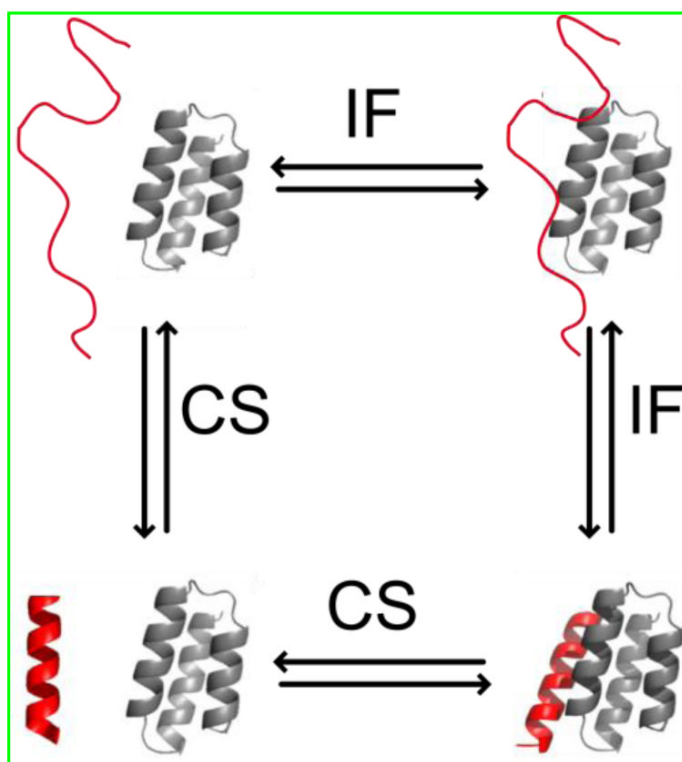


Figure 2. Thermodynamic cycle showing the relationship between induced folding (IF) and conformational selection (CS) binding mechanisms. A representative IDP structure is shown in red. This structure can either bind with low affinity and then fold into the native structure (IF) or can fold into the native structure and then bind (CS). Figure adapted from [54].