

# Targeting Disordered Proteins with Small Molecules Using Entropy

Gabriella T. Heller, Pietro Sormanni, and Michele Vendruscolo

*Department of Chemistry, University of Cambridge, UK*

## Abstract

The human proteome includes many proteins that are fully disordered or that contain large disordered regions. Although these proteins are closely linked with a wide range of human diseases, no clinically-approved drug targets them in their monomeric forms. This situation arises, at least in part, from the current lack of understanding of the mechanisms by which small molecules can bind disordered proteins. To explore possible solutions to this problem, we discuss how an overall decrease in the free energy associated with intermolecular binding can originate from different combinations of enthalpic and entropic contributions. We then propose a mechanism of binding by which small molecules can affect the conformational space of a disordered protein by creating an entropic expansion of the system in which additional conformations of the protein are increasingly populated. This mechanism can be exploited in strategies to modulate the behaviour of disordered proteins, as even small changes in their conformational spaces can drastically affect their properties.

**Keywords:** disordered proteins, binding, entropy, small molecule, entropic expansion

## Disordered proteins and disease

Disease-modifying proteins involved in cancer, neurodegeneration, cardiovascular disease, and diabetes make up about one-third of those encoded by the human genome (**Figure 1a**). Of these proteins, only about 15% are currently considered ‘druggable’, as they are known or predicted to interact with drugs [1]. Essentially all clinically-approved small molecule therapeutics target structured domains [2, 3], despite the fact that intrinsically disordered proteins or intrinsically disordered regions (see Glossary) of otherwise ordered proteins are also commonly involved in disease [4–7] (**Figure 1a**). These disordered proteins, which lack a well-defined stable structure, exist in a dynamic equilibrium of conformationally distinct states.

Proteins with more than 40 consecutive disordered residues are reported to make up one-third to half of the human proteome [8, 9]. These proteins exhibit widely varying degrees of disorder, and this disorder is rather evenly distributed. Our analysis using the *s2D* method (see Glossary) [10] indicates that disordered proteins are approximately 40% of the protein-coding human genome (**Figure 1**). This result was obtained by defining disordered proteins as those that contain more than 40% of their residues in regions of at least 40 consecutive disordered amino acids following similar previous conventions [8, 9]. It is increasingly recognized that disorder serves a biological role, as conformational heterogeneity granted by disordered regions enables proteins to exert diverse functions in response to stimuli. Unlike structured proteins, which are essential for catalysis and transport, disordered proteins appear crucial for regulation and signaling, acting as network hubs interacting with a wide range of biomolecules, rather than just one [4, 11–17].

Given the variety of their functions, dysregulation of disordered proteins can give rise to disease, including cardiovascular disorders, diabetes, cancer, and neurodegeneration [4, 6, 7]. There is, however, an underrepresentation of disordered proteins encoded by the so-called ‘druggable genome’ (**Figure 1c**), which is the set of all proteins reported or predicted to bind small molecules [1–3]. Even in cases in which proteins with disordered regions are targeted, most drugs directly target structured domains of these proteins. Overall, despite their high prevalence in disease, disordered proteins are not targeted by clinically-available drugs. Here we discuss possible strategies to modify this situation to identify opportunities to exploit this untapped potential.

## Small molecules binding to disordered proteins

Major advances have been recently made in understanding the molecular roles disordered proteins play in disease [6, 7, 18]. However, the development of therapeutics that target disordered proteins is still in its infancy, in part because the highly dynamic nature of these proteins renders them difficult to study experimentally. For example, in the case of Alzheimer's disease, despite the enormous efforts over the past two decades to develop drugs capable of inhibiting the aggregation process of the disordered amyloid  $\beta$  peptide, currently no compound that effectively does so has entered clinical use [18–20]. A recently proposed approach to obtain drugs targeting disordered regions relies on the computational docking of small-molecule fragments against an ensemble of representative conformations of the protein of interest [21]. Its application to  $\alpha$ -synuclein, a disordered protein involved in Parkinson's disease, identified a compound that inhibits the aggregation of  $\alpha$ -synuclein [21]. However, it is still poorly understood whether this compound binds directly to the monomeric protein rather than to its aggregated species. A clearer example of direct targeting of monomeric disordered proteins is the oncoprotein c-Myc [22–24]. High-throughput screening yielded a series of compounds interacting with its disordered regions, which prevent binding to its partner, Max. The mechanism of these drug-binding interactions, however, remains unclear and these compounds have not yet entered clinical use [22–24, 25].

Disordered proteins populate ensembles of many conformations, each with its own occupation probability. The behaviour of disordered proteins is governed by these ensembles and can thus be drastically different from that of any individual conformation. Upon interacting with other molecules such as protein-binding partners, disordered proteins may pay an entropic cost because they restrict their conformation space in the bound form, which can be compensated by an enthalpic gain [11, 26]. Conversely, in an alternative scenario a change in the behaviour of disordered proteins may be achieved through the use of small molecules, such that the conformational space of a disordered protein is not restricted, but rather, entropically expanded (see Glossary) by new, transiently bound states with varying probabilities. In the following we discuss potential mechanisms through which small molecules could be effective at targeting monomeric disordered proteins.

## Thermodynamics of protein-ligand binding

The binding of two molecules occurs spontaneously when it is associated with an overall decrease in free energy ( $\Delta G < 0$ ), where  $\Delta G$  indicates the difference between the free energy  $G$  of the final state and that of the initial state. This difference can be expressed as the sum of enthalpic and entropic contributions

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

where the change in enthalpy ( $\Delta H$ ) is determined by a variety of interatomic forces, including electrostatic, van der Waals and hydrogen bonding interactions, and the entropic contribution ( $\Delta S$ ) represents the change in the size of the conformational space available to the overall system, including the protein, ligand, and solvent molecules.

Enthalpic and entropic factors can either contribute favorably or unfavorably to  $\Delta G$ , resulting in the four possible modes: 1)  $\Delta H > 0$ ,  $\Delta S < 0$ , 2)  $\Delta H < 0$ ,  $\Delta S < 0$ , 3)  $\Delta H < 0$ ,  $\Delta S > 0$ , and 4)  $\Delta H > 0$ ,  $\Delta S > 0$ . Only modes 2-4 yield negative  $\Delta G$  values, thus leading to binding. Protein-ligand binding systems can be characterized experimentally into one of these four modes using, for example, isothermal titration calorimetry (ITC) [27]. ITC experiments allow direct, in-solution, label-free determination of both  $\Delta G$  and  $\Delta H$  for a protein-ligand binding system, including contributions from the solvent. The difference of these observed values can be used to calculate  $-T\Delta S$  using Eq. 1 [27-29]. While many protein-ligand binding events are driven by enthalpic factors, in some cases entropy can contribute favorably towards a negative change in free energy, and thus result in binding.

To better understand the role of entropy in protein-ligand binding interactions, we reviewed all entries in the Binding Database (BindingDB) for which there are ITC data (139 unique, non-mutant entries). We categorized these entries according to the magnitude of the entropic contributions [28, 30-33] (**Figure 2a**). Some enthalpically-favorable interactions come at an entropic cost (black points). This compromise is referred to as enthalpy-entropy compensation.

In rational drug design, it is possible to optimize the enthalpy to promote binding to a target,

and occasionally the entropy is also optimized. This emphasis is reflected by the distribution of the entropic contributions to binding across the BindingDB (**Figure 2b**). We note that strong entropic contributions to protein-ligand binding are highly underrepresented in the BindingDB. Furthermore, the vast majority of the entropically-favourable interactions in **Figure 2** are of hydrophobic nature, as binding originates from a large positive solvation entropy [27]. Before interacting, protein and ligand are solvated separately. Upon binding, which in this context usually involves the burial of hydrophobic surfaces, many water molecules from the separate hydration shells surrounding the protein and ligand are freed into bulk solvent. This increases the number of conformational states of the water molecules, and thus, of the system as a whole, overcoming the loss of entropy due to restraining the molecule and binding site [27, 28].

Interestingly, however, cases have been reported (highlighted as squares in **Figure 2a**) of entropy-driven binding where solvation effects may not fully explain the increase in entropy upon binding. The first of these cases is the binding of a  $\text{Ca}^{2+}$  ion to phospholipase D  $\beta\text{C}2$ , or to phospholipase D  $\alpha\text{C}2$  [34]. The second is the binding of a  $\text{Zn}^{2+}$  ion to conantokin-G or to conantokin-T [35]. These changes in entropy seem to be arising from conformational changes in the proteins themselves, rather than from released solvent molecules. More specifically, the binding of metal ions may expose hydrophobic groups to water, causing the backbones to behave in a disordered manner [34, 35]. The increased conformational flexibility enables the protein target to occupy more states, thereby increasing the entropy of the system (**Figure 2c**).

Such induced disorder-upon-binding mechanisms have been proposed and validated for other classes of proteins. For example, one ligand, PD173955, has been optimised to promote disorder of the Bcr-Abl kinase activation loop and therefore increases overall binding affinity as compared to its parent compound, imatinib [36]. Along similar lines, variable differences in the conformational entropy between free and bound states of galectin-3 have been observed upon binding to different carbohydrate differences. The extent of this induced conformational entropy has been shown to control the affinity [37]. These studies, among others, [38-41] support the validity of ligand binding through entropic expansion for disordered proteins.

In contrast to traditional binding scenarios in which a drug locks a protein in an inactive state, the entropy-driven binding scenario may appear unconventional due to the multiplicity of low populated conformations in which the protein is weakly interacting with the ligand (**Figure**

**2c).** This mechanism of interaction suggests a more general definition of ‘binding’ than one in which two molecules form a stable association through strong intermolecular interactions. Within this framework any interaction in which a ligand significantly affects the conformational space of a protein may be considered binding. This view raises two issues (see Outstanding Questions Box): firstly, by this definition, solutes affecting protein hydration, high salt concentrations, and crowding agents that affect the conformational space of a disordered protein are be considered as ‘ligands’ even though they are not traditionally considered as such. Although discussing binding in this way may be unconventional, we are forced here to consider it because it is the natural consequence of entropic expansion.

A second concern is how these interactions can be specific, as binding interactions governed by a network of weaker interactions to induce favourable entropic changes may not be localized to a single binding region. We propose, however, that a combination of entropically and enthalpically favourable interactions can be used to optimize, specific interactions that can alter the behaviour of disordered regions. In other words, one way in which ligands may bind disordered regions consists of using relatively weak enthalpic interactions to ensure specificity, but relying on entropic factors for increasing the binding free energy. Mutagenesis studies have confirmed that drugs found to bind to the disordered region of c-Myc have specific binding regions 11-29 residues long within the 84-residue long disordered bHLHZip domain [42], thus suggesting that specificity may be achievable when binding to disordered regions. However, it is possible that these small molecules may bind to other proteins as some of their constituent scaffolds have been shown to bind a large number of targets with weak or moderate affinity [43].

### **Using entropy to target disordered proteins**

Enthalpy-focused rational drug design, which is widely used for targeting globular proteins [27], has not yet led to major advances for disordered proteins. While focusing on favourable enthalpic contributions may be a good strategy for targeting structured proteins, we suggest that exploiting entropically-favourable binding may provide novel opportunities for obtaining drugs to target monomeric disordered proteins. We anticipate that many drugs targeting disordered proteins will populate the bottom quadrants of **Figure 2c** by similar mechanisms to those described above for metal-binding proteins. Much like phospholipase D and

conantokins, the conformational space available to a polypeptide chain may be susceptible to entropic expansion upon the introduction of a ligand, thus resulting in an increase the number of states accessible to the disordered polypeptide chain itself (**Figure 3**). Such an increase of entropy may result in favorable protein-ligand interactions, and be modulated by hydrophobic effects, electrostatics,  $\pi$ -effects, and van der Waals forces.

This type of entropically-favourable binding may be more effective than targeting disordered proteins with strong enthalpic or solvent-mediated entropic interactions that stabilize a single conformation. Such types of interactions indeed occur for disordered proteins, such as the many disorder-to-order transitions reported in the literature [9, 12, 44]. These interactions, however, tend to involve large interfaces, which currently can be disrupted by binding the structured partner to block the binding region of the disordered protein [11]. We propose that small molecules can also be developed to directly target monomeric disordered proteins in order to alter their behaviour via entropic expansion. We illustrate this concept in **Figure 3**, which represents the limited conformational space of an unbound disordered protein being expanded upon the introduction of a drug. The mechanisms by which this type of binding can happen include the disruption or creation of contacts yielding more extended conformations or more structured ones, expanding the conformational space of the bound protein with a corresponding overall increase in entropy. As new states of the protein are introduced upon the addition of a drug, the statistical weights of existing states in the unbound ensemble may decrease, thus providing a potential avenue to significantly alter the behaviour of the disordered target.

This concept may be used to rationalize simulations performed recently on a peptide fragment of c-Myc and its binding partner, the small molecule 10058-F4. The c-Myc peptide was observed to occupy an increasingly large number of low populated states upon binding 10058-F4 as compared to the unbound form [45]. We interpret this observation as an example of the extended conformational space of a bound disordered protein and thereby a potential increase in the overall entropy of the system. Similar analyses on another c-Myc peptide and its binding partner 10074-A4 further suggest that drug binding to c-Myc occurs weakly at many binding sites along the binding region [46]. These reports are consistent with the hypothesis of an entropy-increasing mechanism, as described above. Further studies will be necessary to confirm these conclusions.

Additionally, a second relevant example is the MDM2 protein, which contains a large disordered lid region. Upon binding some classes of p53/MDM2 antagonists, the base of the lid region becomes ordered via contacts with the ligand with an accompanying increase in flexibility of the remainder of the disordered region as compared to the disordered state [47,48]. In the current attempts to curtail the toxic behavior of disordered proteins, likely milestones may include the fine-tuning of flexible disordered regions such as the flexible lid of MDM2.

### **Therapeutic potentials of the entropic expansion mechanism of binding**

Conformational ensembles of disordered proteins are extremely sensitive to external factors. Post-translational modifications [49], point mutations [50], and pH changes [51] can significantly reweight the states within an ensemble, including introducing nontrivial states and significantly decreasing the weights of existing ones. Small changes within ensembles can have profound physiological effects. In the case of mutational variants of  $\alpha$ -synuclein, it was reported that relatively small differences in the ensembles, and more specifically in the populations of  $\beta$ -strands [50] and polyproline II [52] secondary structure elements, correspond to large differences in the aggregation rates, thus affecting the formation of neurotoxic species.

Small molecules can have a similar potential to change the ensembles of disordered proteins from one that is disease-promoting to one that may be disease-preventing, thereby enabling them to have a therapeutic effect. Earlier in this article, we described drug development opportunities aimed at entropically expanding the ensemble, although reweighting, restricting, or shifting the ensemble are also candidate strategies to change the behaviour of disordered proteins and eliminate disease-promoting conformations.

As is the case of c-Myc described earlier, drugs that target disordered regions have shown promising initial results both *in vitro* and *in vivo*, and currently more metabolically-stable analogs are in development [6, 23, 53]. An ability to modify the behaviour of disordered proteins via the use of small molecules may have profound implications for disease. For instance, the major neurotoxic agents responsible in Alzheimer's disease are soluble oligomeric species formed by the aggregation of the amyloid  $\beta$  peptide [20, 54, 55]. We thus believe that particularly effective therapeutic agents against aggregation may inhibit the formation of oligomeric assemblies via entropic expansion of monomeric disordered peptides.

Because many of the toxic oligomeric species involved in neurodegeneration are largely comprised of disordered proteins, reweighting the monomeric ensembles of these disordered proteins may slow down the formation of toxic oligomeric species. As new states are introduced, the populations of existing, potentially disease-promoting states will decrease, and such an expansion of the ensemble can be achieved by the introduction of certain small molecules that increase the entropy of the overall system.

### **Concluding remarks**

In most cases the binding of ligands to structured proteins is either driven by enthalpic factors or by solvent-mediated entropic interactions such as hydrophobic attraction. In the case of disordered proteins, however, we expect that favourable entropic contributions arising from the proteins themselves rather than from the solvent may represent an alternative drug-discovery strategy, which can be then further tunable using enthalpic contributions. Disordered proteins have been considered as ‘untargetable’ so far because they do not readily lend themselves to enthalpy-driven binding, as no tight binding small-molecules have been identified to interact with them (see Outstanding Questions Box). They may be, however, more amenable to binding via entropic expansion, which may only require weak enthalpic interactions between proteins and ligands. We anticipate that by exploiting recent developments in experimental measurements and molecular dynamics simulations it will become possible to obtain increasingly accurate estimates of entropic contributions of disordered protein-ligand systems, which will hopefully facilitate the rational development of small molecules targeting disordered proteins.

### **Acknowledgements**

The authors thank Dr. Francesco Aprile and Dr. Carlo Camilloni for helpful discussions. GTH acknowledges the Churchill Scholarship for funding.

## Glossary

**BindingDB (Binding Database).** An online database of measured binding affinities (<http://www.bindingdb.org>). Entries are mainly protein considered to be drug-targets with small drug-like molecules.

**Disordered proteins or disordered regions.** Proteins or protein regions that under native conditions do not populate a well-defined conformation, but rather a heterogeneous ensemble of states.

**Entropic expansion.** An increase of entropy of a disordered protein upon the introduction of a ligand, whereby the increase arises because the disordered protein populates even more states than it did in the unbound form.

**Isothermal titration calorimetry (ITC).** An experimental technique that can be used to determine thermodynamic parameters from a binding interaction.

**s2D method.** A computational method to simultaneously predict disordered regions and secondary-structure populations of proteins from their amino acid sequences.

## References

1. Griffith M, et al. (2013) DGIdb: mining the druggable genome. *Nat Methods* 10(12):1209–10.
2. Russ AP, Lampel S (2005) The druggable genome: An update. *Drug Discov Today* 10:1607–1610.
3. Hopkins AL, Groom CR (2002) The druggable genome. *Nat Rev Drug Discov* 1:727–730.
4. Babu MM, van der Lee R, de Groot NS, Gsponer J (2011) Intrinsically disordered proteins: Regulation and disease. *Curr Opin Struct Biol* 21:432–440.
5. Tompa P (2012) Intrinsically disordered proteins: A 10-year recap. *Trends Biochem Sci* 37:509–516.
6. Metallo SJ (2010) Intrinsically disordered proteins are potential drug targets. *Curr Opin Chem Biol* 14:481–488.
7. Uversky VN, Oldfield CJ, Dunker AK (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215–246.
8. 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.
9. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z (2002) Intrinsic disorder and protein function. *Biochemistry* 41:6573–6582.
10. Sormanni P, Camilloni C, Fariselli P, Vendruscolo M (2015) The s2D Method: Simultaneous Sequence-Based Prediction of the Statistical Populations of Ordered and Disordered Regions in Proteins. *J Mol Biol* 427(4):982–996.
11. Flock T, Weatheritt RJ, Latysheva NS, Babu MM (2014) Controlling entropy to tune the functions of intrinsically disordered regions. *Curr Opin Struct Biol* 26:62–72.

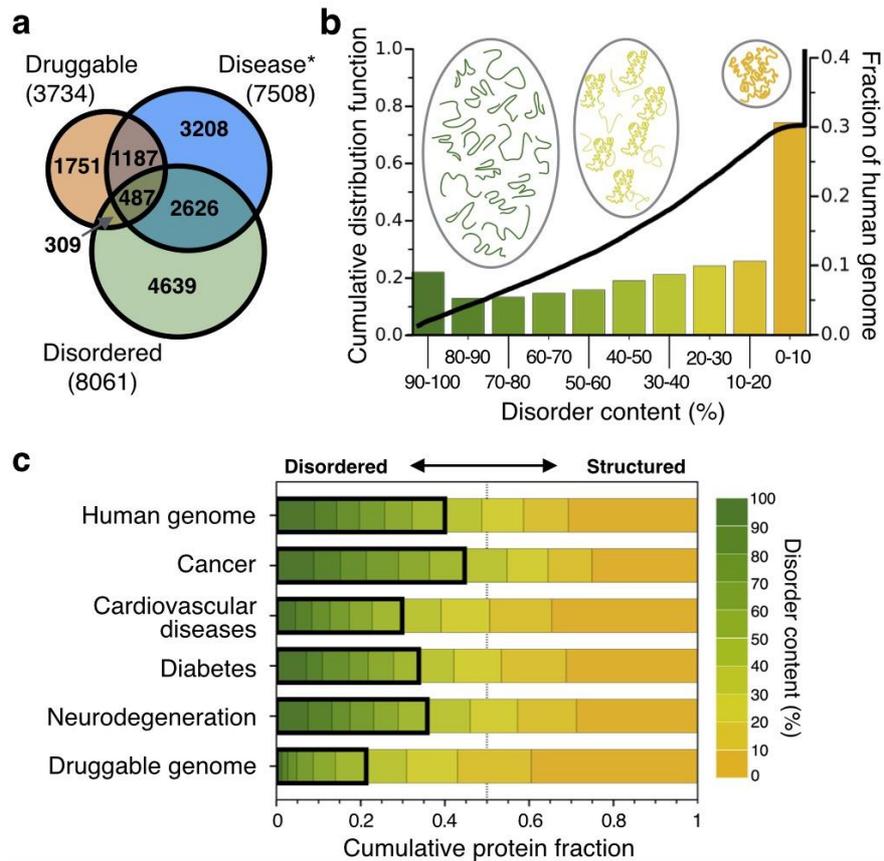
12. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 6:197–208.
13. Tompa P, Han K (2012) Intrinsically disordered proteins. *Phys Today* 65:64.
14. Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 293:321–331.
15. Uversky VN (2011) Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol* 43:1090–1103.
16. Habchi J, Tompa P, Longhi S, Uversky VN (2014) Introducing protein intrinsic disorder. *Chem Rev* 114(13):6561–6588.
17. Tompa P (2012) Intrinsically disordered proteins: A 10-year recap. *Trends Biochem Sci* 37:509–516.
18. Hardy J, Selkoe DJ (2002) The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297:353–356.
19. Rafii MS, Aisen PS (2009) Recent developments in Alzheimer's disease therapeutics. *BMC Med* 7:7.
20. Cohen SI a, et al. (2013) Proliferation of amyloid- $\beta$ 42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci U S A* 110:9758–63.
21. Tóth G, et al. (2014) Targeting the intrinsically disordered structural ensemble of  $\alpha$ -synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease. *PLoS One* 9(2):e87133.
22. Hammoudeh DI, Follis AV, Prochownik E V., Metallo SJ (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J Am Chem Soc* 131:7390–7401.
23. Follis AV, Hammoudeh DI, Daab AT, Metallo SJ (2009) Small-molecule perturbation of competing interactions between c-Myc and Max. *Bioorganic Med Chem Lett* 19:807–810.
24. Berg T (2010) Small-molecule modulators of c-Myc/Max and Max/Max interactions. *Curr Top Microbiol Immunol* 348:139–149.

25. Cuchillo R, Michel J (2012) Mechanisms of small-molecule binding to intrinsically disordered proteins. *Biochem Soc Trans* 40(5):1004–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22988855>.
26. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6:352–358.
27. Bronowska A (2011) Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. *Thermodyn - Interact Stud - Solids, Liq Gases*:1–49.
28. Klebe G (2015) Applying thermodynamic profiling in lead finding and optimization. *Nat Rev Drug Discov* 14(2):95–110.
29. Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys* 42:121–42.
30. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK (2007) BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res* 35.
31. Chen X, Lin Y, Gilson MK (2001) The binding database: Overview and user's guide. *Biopolymers* 61:127–141.
32. Chen X, Lin Y, Liu M, Gilson MK (2002) The Binding Database: data management and interface design. *Bioinformatics* 18:130–139.
33. Chen X, Liu M, Gilson MK (2001) BindingDB: a web-accessible molecular recognition database. *Comb Chem High Throughput Screen* 4:719–725.
34. Zheng L, Krishnamoorthi R, Zolkiewski M, Wang X (2000) Distinct Ca<sup>2+</sup> binding properties of novel C2 domains of plant phospholipase D $\alpha$  and  $\beta$ . *J Biol Chem* 275:19700–19706.
35. Prorok M, Castellino FJ (1998) Thermodynamics of binding of calcium, magnesium, and zinc to the N- methyl-D-aspartate receptor ion channel peptidic inhibitors, conantokin-G and conantokin-T. *J Biol Chem* 273:19573–19578.

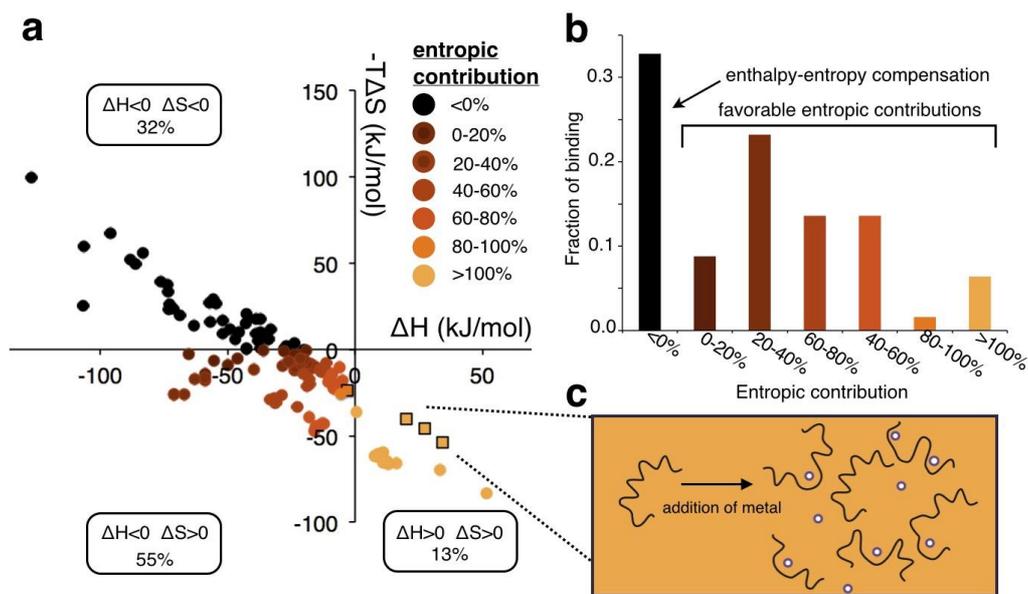
36. Crespo A, Fernández A (2008) Induced disorder in protein-ligand complexes as a drug-design strategy. *Mol Pharm* 5(3):430–437.
37. Diehl C, et al. (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc* 132(41):14577–14589.
38. Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. *Nature* 448(7151):325–329.
39. Chang C-EA, McLaughlin WA, Baron R, Wang W, McCammon JA (2008) Entropic contributions and the influence of the hydrophobic environment in promiscuous protein-protein association. *Proc Natl Acad Sci U S A* 105(21):7456–7461.
40. Baron R, McCammon JA (2008) (Thermo)dynamic role of receptor flexibility, entropy, and motional correlation in protein-ligand binding. *ChemPhysChem* 9(7):983–988.
41. Tzeng S-R, Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. *Nature* 462(7271):368–372.
42. Follis AV, Hammoudeh DI, Wang H, Prochownik E V., Metallo SJ (2008) Structural Rationale for the Coupled Binding and Unfolding of the c-Myc Oncoprotein by Small Molecules. *Chem Biol* 15(11):1149–1155.
43. Mendgen T, Steuer C, Klein CD (2012) Privileged scaffolds or promiscuous binders: A comparative study on rhodanines and related heterocycles in medicinal chemistry. *J Med Chem* 55(2):743–753.
44. Wright PE, Dyson HJ (2009) Linking folding and binding. *Curr Opin Struct Biol* 19:31–38.
45. Michel J, Cuchillo R (2012) The impact of small molecule binding on the energy landscape of the intrinsically disordered protein C-Myc. *PLoS One* 7.
46. Jin F, Yu C, Lai L, Liu Z (2013) Ligand Clouds around Protein Clouds: A Scenario of Ligand Binding with Intrinsically Disordered Proteins. *PLoS Comput Biol* 9.
47. Michelsen K, et al. (2012) Ordering of the N-terminus of human MDM2 by small molecule inhibitors. *J Am Chem Soc* 134(41):17059–17067.

48. Bueren-Calabuig J a., Michel J (2015) Elucidation of Ligand-Dependent Modulation of Disorder-Order Transitions in the Oncoprotein MDM2. *PLOS Comput Biol* 11(6):e1004282. Available at: <http://dx.plos.org/10.1371/journal.pcbi.1004282>.
49. Xiang S, et al. (2013) Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure* 21:2162–2174.
50. Camilloni C, Vendruscolo M (2013) A relationship between the aggregation rates of  $\alpha$ -synuclein variants and the  $\beta$ -sheet populations in their monomeric forms. *J Phys Chem B* 117:10737–10741.
51. Hong W, et al. (2005) Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. *J Biol Chem* 280:27029–27034.
52. Porcari R, et al. (2015) The H50Q Mutation Induces a 10-fold Decrease in the Solubility of  $\alpha$ -Synuclein. *J Biol Chem* 290(4):2395–2404.
53. Jeong K-C, Ahn K-O, Yang C-H (2010) Small-molecule inhibitors of c-Myc transcriptional factor suppress proliferation and induce apoptosis of promyelocytic leukemia cell via cell cycle arrest. *Mol Biosyst* 6:1503–1509.
54. Jarrett JT, Lansbury PT (1993) Seeding “one-dimensional crystallization” of amyloid: A pathogenic mechanism in Alzheimer’s disease and scrapie? *Cell* 73:1055–1058.
55. Knowles TPJ (2009) An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science (80)* 1533:1533–1537.
56. Uversky VN, et al. (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* 10 Suppl 1:S7.
57. Andrea Giandanti AD (2014) On the abundance of Intrinsically Disordered Proteins in the Human Proteome and its Relation to Diseases: There is no Enrichment. *Biophys Bioeng Lett* 7(2):27–32.

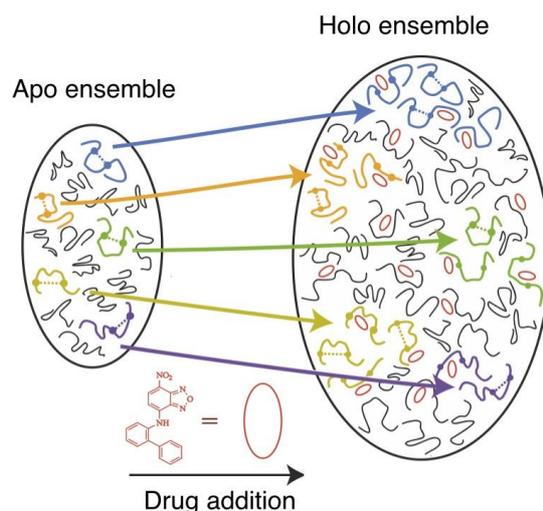




**Figure 1. Prevalence of disorder in some common human diseases.** (a) Venn diagram of three subsets of the human proteome. Proteins are defined as ‘disordered’ if they contain more than 40% of their residues in regions of at least 40 consecutive disordered amino acids, as ‘druggable’ if they are known or predicted to interact with drugs [1], and as ‘disease-related’ or ‘disease-modifying’ (disease\*) if they are involved in cancer, diabetes, neurodegeneration or cardiovascular diseases (proteins in these groups were determined with a keyword method adapted from Refs. [56, 57]). (b) Fraction of proteins encoded by the human genome (right axis) binned according to their content of structural disorder (x-axis). Green bins represent highly disordered proteins, and orange bins structured ones. The black line is the cumulative distribution function (left axis). Cartoons illustrate ensembles of three proteins with varying disorder content. (c) Comparison of the amount of protein disorder encoded by the human genome, by the druggable genome, and in disease-related proteins. Proteins are binned horizontally by disorder content (color-bar). Black boxes represent the fraction of disordered proteins as defined in (a). The analysis of disorder was performed using the s2D method [10]; an individual residue was considered disordered if its populations of  $\alpha$ -helix and  $\beta$ -strand were smaller than 0.5.



**Figure 2. Entropic contributions to protein-ligand binding in the BindingDB.** (a) Entries from the BindingDB are plotted according to their enthalpic ( $\Delta H$ , x-axis) and entropic ( $-T\Delta S$ , y-axis) contributions. Data points are representative of all unique, non-mutant entries in the BindingDB and are coloured according to their entropic contributions. Squares represent entries in which the release of solvent molecules is not considered to be the cause of the change in entropy, but rather the many conformations of the backbone itself. These cases include:  $\text{Ca}^{2+}$  binding to phospholipase D  $\beta\text{C}2$ ,  $\text{Ca}^{2+}$  binding to phospholipase D  $\alpha\text{C}2$  [34]  $\text{Zn}^{2+}$  binding to conantokin-G, and  $\text{Zn}^{2+}$  binding to conantokin-T [35]. (b) Distribution of entropic contributions to  $\Delta G$ . (c) Illustration of a possible entropy-driven binding of  $\text{Zn}^{2+}$  to conantokins, where the free energy decreases in the presence of  $\text{Zn}^{2+}$ . This entropy may arise from the exposure of the backbone to the solvent, thereby introducing new states in the bound form.



**Figure 3. Entropic expansion of the conformational space of a disordered protein upon binding a small molecule.** A small molecule can interfere with the intramolecular interactions within a disordered protein, thereby increasing its entropy by changing the statistical weights of the conformations that it populates, in some cases by increasing their number. The ensemble on the left represents the conformational space of an unbound disordered protein, and the one on the right represents the remodelling of this conformational space upon binding a drug (shown in red). Different colors represent how various states may be affected upon drug binding.