CellPress

Targeting disordered proteins with small molecules using entropy

Gabriella T. Heller, Pietro Sormanni, and Michele Vendruscolo

Department of Chemistry, University of Cambridge, Cambridge, UK

The human proteome includes many disordered proteins. Although these proteins are closely linked with a 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 bind proteins that do not fold into well-defined conformations. To explore possible solutions to this problem, we discuss quite generally how an overall decrease in the free energy associated with intermolecular binding can originate from different combinations of enthalpic and entropic contributions. We then consider more specifically a mechanism of binding by which small molecules can affect the conformational space of a disordered protein by creating an entropic expansion in which more conformations of the protein become populated.

Disordered proteins and disease

Disease-modifying proteins involved in cancer, neurodegeneration, cardiovascular diseases and diabetes comprise about one-third of those encoded by the human genome (Figure 1A). Of these proteins, only approximately 22% are currently considered 'druggable', as they are known or predicted to interact with drugs (Figure 1A) [1–3]. Moreover, 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 have been reported to comprise one-third to one-half of the human proteome [8,9]. These proteins exhibit widely varying degrees of disorder, and this disorder is rather evenly distributed. An analysis using the s2D method [10] indicated that disordered proteins correspond to 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, consistently with similar previous conventions [8,9].

Corresponding author: Vendruscolo, M. (mv245@cam.ac.uk).

Keywords: disordered proteins; binding; entropy; small molecule; entropic expansion.

0968-0004/

 http://dx.doi.org/10.1016/j.tibs.2015.07.004

Following an initial surprise after its discovery, it is now increasingly recognised that disorder serves a biological role, because 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 signalling, acting as network hubs interacting with a wide range of biomolecules [4,5,11–16].

Given the variety of their functions, dysregulation of disordered proteins can give rise to a variety of diseases including cardiovascular disorders, diabetes, cancer, and neurodegeneration [4,6,7]. However, there is an underrepresentation of disordered proteins among those encoded by the current 'druggable genome' (Figure 1C). Even in cases in which proteins with disordered regions are targeted, most drugs are directed towards the 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 of disordered proteins in disease [4–9,11–16]. 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 it difficult to study them 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 β peptide, currently no compound that effectively does so

Glossary

Binding Database (BindingDB): an online database of measured binding affinities (http://www.bindingdb.org). Entries are mainly proteins considered to be targets of 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 the size of the conformational space of a disordered protein upon the introduction of a ligand, whereby the bound protein populates even more states than the unbound protein.

Isothermal titration calorimetry (ITC): an experimental technique that can be used to determine thermodynamic parameters for a binding interaction.

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

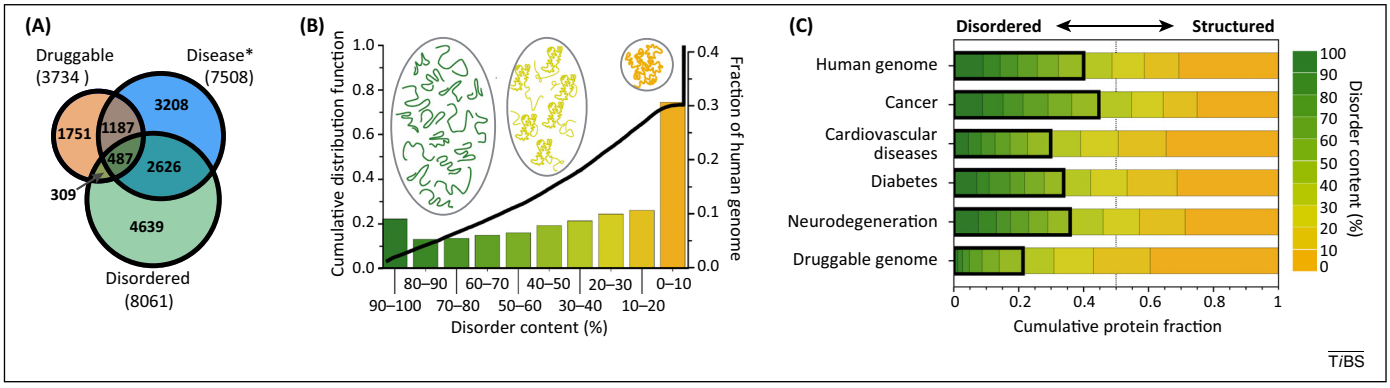

Figure 1. Prevalence of protein disorder in some common human diseases. **(A)** Venn diagram of three subsets of the human proteome. Proteins are defined as 'disordered' if they contain more than 40% of their residues in regions of at least 40 consecutive disordered amino acids, as 'druggable' if they 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 [55,56]). **(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 disordered content (colour 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 α-helical and β-strand populations are smaller than 0.5.

has entered clinical use [17–19]. 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 [20]. Its application to α-synuclein, a disordered protein involved in Parkinson's disease, identified a compound that inhibits the aggregation of α-synuclein [20]. However, it is still poorly understood whether this compound binds more preferentially the monomeric protein than its aggregated species. A clearer example of direct targeting of monomeric disordered proteins is the case of the oncoprotein c-Myc [21–23]. A recent high-throughput screening yielded a series of compounds, which interact with its disordered regions and prevent binding to its partner, Max. However, the mechanism of these drug-binding interactions remains unclear and these compounds have not yet entered clinical use [21–24].

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 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 their conformation space is restricted in the bound form, which can be compensated by an enthalpic gain [11,25]. 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 by new, transiently bound states. In the following, we discuss these and other 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 ΔG 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 (Equation 1):

$$\Delta G = \Delta H - T\Delta S \quad [1]$$

where the change in enthalpy (ΔH) is determined by a variety of interatomic forces, including electrostatic, van der Waals, and hydrogen-bonding interactions, and the entropic contribution ΔS 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 favourably or unfavourably to ΔG, resulting in the four possible modes: (i) $\Delta H>0$, $\Delta S<0$; (ii) $\Delta H<0$, $\Delta S<0$; (iii) $\Delta H<0$, $\Delta S>0$; and (iv) $\Delta H>0$, $\Delta S>0$. Only modes (ii–iv) yield negative ΔG values, thus lead to binding. Protein-ligand binding systems can be characterised experimentally into one of these four modes using, for example, isothermal titration calorimetry (ITC) [26]. ITC experiments allow direct, in-solution, label-free determination of both ΔG and ΔH 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 Equation 1 [26–28]. While many protein-ligand binding events are driven by enthalpic factors, in some cases entropy can contribute favourably 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 thermodynamic data (139 unique, non-mutant entries). We categorised these entries according to the magnitude of the entropic contributions [27,29–32] (Figure 2A). Some enthalpically favourable interactions come at an entropic cost (black points in Figure 2A). This compromise is commonly referred to as enthalpy–entropy compensation.

In rational drug design, it is possible to optimise enthalpic contributions to promote binding to a target, and occasionally the entropy is also optimised. This emphasis is reflected by the distribution of the entropic contributions to binding across the BindingDB (Figure 2B). We note that