ELSEVIER

Targeting intrinsically disordered proteins at the edge of chaos

Hao Ruan[1,‡], Qi Sun[1,‡], Weilin Zhang[1], Ying Liu[1,2] and Luhua Lai[1,2,3]

[1] BNLMS, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China
[2] Center for Quantitative Biology, Peking University, Beijing 100871, China
[3] Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China

Intrinsically disordered proteins or intrinsically disordered regions (IDPs or IDRs) are those that do not fold into defined tertiary structures under physiological conditions. Given their prevalence in various diseases, IDPs are attractive therapeutic targets. However, because of the dynamic nature of the IDP structure, conventional structure-based drug design methods cannot be directly applied. Thanks to recent progress in understanding the mechanisms underlying IDP and ligand interactions, computational strategies for IDP-targeted rational drug discovery are emerging. Here, we summarize recent developments in computational IDP drug design strategies and their successful applications, analyze the typical properties of reported IDP-binding compounds (iIDPs), and discuss the major challenges ahead as well as possible solutions.

Introduction

Proteins function via an array of conformations, from ordered to completely disordered, demonstrating a range of thermodynamic fluctuations. IDPs or IDRs lack a well-defined structure and have recently attracted increased attention [1–4]. They typically comprise a high proportion of polar and charged amino acid residues and a depletion of bulk hydrophobic side chains [5]. IDPs/IDRs have a variety of structural subtypes, from highly disordered, expanded conformational ensembles, to compact, disorder-restricted states [5,6]. IDPs/IDRs are abundant in many organisms, and their proportion increases sharply from bacteria to higher organisms [7,8]. IDPs commonly function as hubs in protein interaction networks, and often have central roles in regulating signaling pathways and participate in the assembly of molecular machines [9]. Recent studies have further shown that disordered proteins, which are characterized by low sequence complexity, can promote phase separation under physiological conditions [10,11].

IDPs are prevalent in various human diseases and are considered potential drug targets [12,13]. Although the human genome contains a large number of possible drug targets, only ~2% of human proteins are known to interact with approved drugs [14,15]. Many remaining targets, particularly IDPs such as c-Myc, p53, and Ras, have been classified as 'difficult to drug' or 'yet to be drugged' [16].

Several recent examples of IDP ligand discovery have been reported, and computational methods for IDP drug design are also emerging. However, our understanding of the thermodynamics and kinetics governing IDP–ligand interactions, and a rational and generally applicable strategy for IDP drug discovery, remain lacking. Here, we briefly introduce the necessity of targeting IDPs and their druggability, and then give recent examples of IDP ligand discovery. We also provide an analysis of known iIDPs, insights gained from binding mechanism studies between IDPs and their ligands, and a summary of rational drug design approaches for IDPs. We conclude by discussing remaining challenges and possible solutions for IDP-targeting drug design.

IDPs as potential drug targets

Given their central roles in numerous cellular processes, IDP availability in cells is controlled at multiple levels, such as by regulating mRNA transcription and clearance, protein translation,

Corresponding author: Lai, L. (lhlai@pku.edu.cn)
‡ These authors contributed equally to this work.

https://doi.org/10.1016/j.drudis.2018.09.017

and protein degradation [17,18]. Strategies for maintaining optimal IDP levels are evolutionarily conserved from *Schizosaccharomyces pombe* to humans [18]. Altering IDP availability can produce undesirable results, such as reduced signaling fidelity and acquisition of potentially harmful aberrant functions [17]. IDPs are enriched in cardiovascular disease, diabetes, cancer, and neurodegenerative disease-related proteins. This phenomenon inspired the disorder in disorders (D^2) concept, which emphasizes the prevalence of disorder in proteins associated with diseases (disorders) [12]. For example, disordered regions are prevalent in cancer-related proteins, including the oncogene, c-Myc, and the cancer suppressor gene, *p53*, and the altered abundance of these proteins contributes to cancer development [19]. Additionally, mutations in some disordered proteins, such as amyloid β-peptide (Aβ) and α-synuclein, can promote aggregation, leading to the development of neurodegenerative diseases [20].

Given that IDPs are common in many diseases, there is potential to utilize our knowledge of protein disorders to develop drugs targeting these proteins [12,13,21–26]. However, because of the ensemble nature of IDP conformation, conventional structure-based drug design methods cannot be directly applied. Most successful strategies that have facilitated the recovery of function for disease-associated disordered proteins have avoided directly targeting IDPs. For example, strategies have been developed to target enzymes that modify disordered proteins to recover their normal availability, the ordered domains of disordered proteins, and ordered IDP-binding partners, such as for p53–MDM2 interaction inhibitors that bind to MDM2 [27–29].

An underexplored approach involves the design of ligands that directly bind to IDPs, change their conformational ensemble, and compete with endogenous binding partners. This strategy can also target misfolded IDPs, which interact with undesired partners or aggregate. Although IDPs are suggested to have similar druggability to ordered proteins, IDPs often contain more binding cavities, because of the nature of coupled folding and binding processes, and IDPs often have weaker binding affinity compared with structured proteins [30,31]. These observations suggest that small-molecule ligands could be designed to specifically bind to IDPs and disrupt endogenous interactions. Advancement of these strategies might be possible to convert the vast number of IDPs into druggable targets and pave the way for exciting therapeutic discoveries.

Examples of ligand discovery targeting IDPs

Despite difficulties associated with ligand discovery for direct IDP targeting, there are several reported examples, most of which are involved in protein–protein interaction networks. Most IDP-binding ligands were discovered by high-throughput experimental screening of chemical compound libraries. These examples provide valuable clues for enhancing our understanding of IDP–ligand interactions and suggest strategies for the development of rational design methods to target IDPs (Table 1).

The first example is c-Myc, an oncogenic transcription factor with a helix-loop-helix leucine zipper structure, which is commonly overexpressed in human cancers [32]. The heterodimer formed between c-Myc and Myc-associated ligand Max is essential for its functions in DNA binding and transcriptional activation. Using a yeast two-hybrid approach, several low-molecular-weight compounds that bind to c-Myc and inhibit c-Myc–Max heterodimer formation were discovered [32]. Fluorescence polarization (FP) assay, circular dichroism (CD), and nuclear magnetic resonance (NMR) experiments further showed that three compounds bind at three independent disordered regions of c-Myc, providing evidence for small-molecule-binding sites in disordered proteins [33,34]. Follow-up studies discovered additional small molecules [35–41]. For example, using protein-fragment complementation assay-based high-throughput screening (HTS), the compound sAJM589 was found to inhibit the c-Myc–Max interaction with an IC_{50} of 1 μM [41]. Additionally, several molecules that inhibit the c-Myc–Max interaction were discovered by molecular docking screening [42].

The oncogenic fusion protein, EWS-FLI1, is common in Ewing's sarcoma family tumors, and binding of this molecule to RNA helicase (RHA) is essential for tumor maintenance [43]. Using surface plasmon resonance (SPR) to screen a library of 3000 small molecules, the compound NSC635437 was found to bind to the disordered EWS-FLI1 [43]. Through an aromatic optimization strategy, YK-4-279, based on NSC635437, was then discovered to significantly disrupt the interaction between EWS-FLI1 and recombinant GST-$RHA_{647–1075}$. In addition, YK-4-279 functionally inhibited EWS-FLI1, promoted toxicity in tumor cell lines containing EWS-FLI1, and reduced tumor volume in a Ewing's sarcoma xenograft model.

The disordered protein, p27, regulates cell proliferation by interacting with cyclin-dependent kinases (CDKs), and disrupting this interaction could prevent breast cancer cell migration [44]. Using NMR spectroscopy to screen a fragment library, nine molecules that blocked the interaction between p27 and CDKs were identified [44]. These compounds can be divided into two groups, each of which specifically binds to distinct, but partially overlapping, regions of the disordered kinase inhibitory domain of p27. Binding sites within this p27 kinase inhibitory domain contain most of the aromatic residues found in full-length p27. Binding of small-molecule inhibitors to this region blocks the ability of p27 to sequester CDKs. For example, the group 2 molecule, SJ403, can sequester p27 and inhibit interaction between p27 and Cdk2/cyclin A, thus partially restoring Cdk2/cyclin A activity.

Nuclear protein 1 (NUPR1) is a multifunctional disordered protein overexpressed in pancreatic tumors. By using fluorescence thermal denaturation to screen 1120 US Food and Drug Administration (FDA)-approved drugs, 15 compounds that significantly alter the temperature denaturation profiles of NUPR1 were found [45]. Their binding affinity and binding regions were further evaluated by isothermal titration calorimetry (ITC) and NMR, respectively. Among them, the compound trifluoperazine was found to disrupt the interaction between NUPR1 and its binding partner, MSLI1, in a cell-based assay. Tumor growth inhibition was also observed after daily treatment with trifluoperazine in a pancreatic ductal adenocarcinoma (PDAC) xenograft tumor mouse model. By combining NMR experiments and sequence hydrophobicity analysis, these compounds were shown to primarily bind to hydrophobic regions of NUPR1.

Both disordered AF4 and AF9 are fusion partners for the mixed-lineage leukemia (MLL) gene, and leukemic cell lines require the AF4–AF9 interaction to survive and self-renew [46]. They bind to each other and promote reciprocal translocations commonly oc-

TABLE 1

Reported IDP-binding compounds[a]

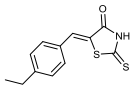
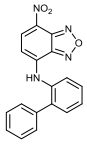
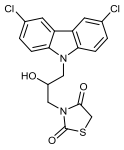
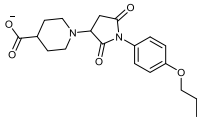
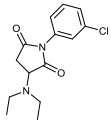
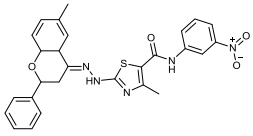
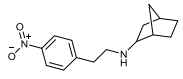
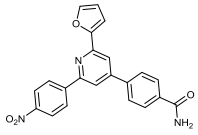
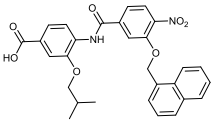
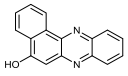
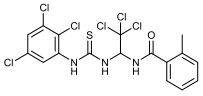
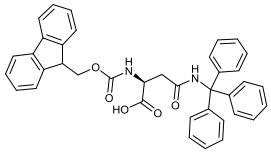
Target	Compound	Structure	Identification	*In vitro* activity	Cytoactivity	Refs
c-Myc	10058-F4		HTS	SPR: 39.7 ± 8.1 μM	HL-60: 49 μM	[32,35,39]
	10074-G5		HTS	SPR: 31.7 ± 24.9 μM	HL-60: 13.5 ± 2.1 μM	[32,37,39]
	10074-A4		HTS	SPR: 36.3 ± 9.0 μM	HL-60: 15.1 ± 2.3 μM	[32,34]
	10031-B8		HTS	FP: 16 ± 4 μM	Not available	[32,34]
	10075-G5		HTS	FP: 24 ± 4 μM	Not available	[32,34]
	10050-C10		HTS	FP: 0.9 ± 0.3 μM	Not available	[32,34]
	10009-G9		HTS	FP: 40 ± 10 μM	Not available	[32,34]
	KJ-Pyr-9		HTS	BSI: 6.5 ± 1.0 nM	BL: 1–2.5 μM	[38]
	JKY-2-169		RD	EMSA: 11.6 ± 2.3 μM	HL-60: 19.9 ± 1.6 μM	[40]
	sAJM589		HTS	PCA: 1.8 ± 0.03 μM	HL-60: 1.2 ± 0.09 μM	[41]
	PKUMDL-YC1204		VS	SPR: 0.55 ± 0.14 μM	HL-60: 8.80 ± 0.26 μM	[42]
	PKUMDL-YC1205		VS	SPR: 18 ± 12 μM	HL-60: 40.0 ± 1.9 μM	[42]

TABLE 1 (*Continued*)

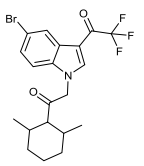
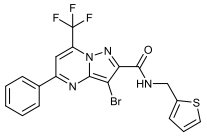
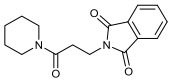
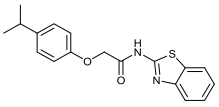
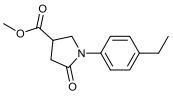
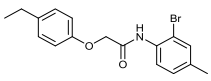
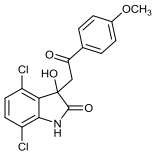
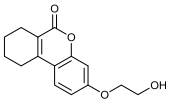
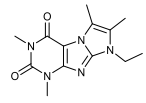
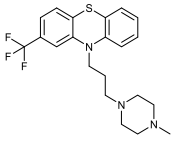
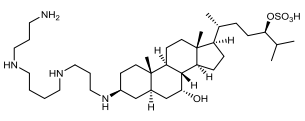
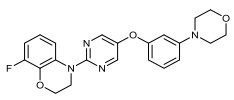
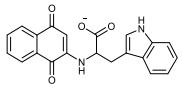
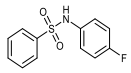
Target	Compound	Structure	Identification	*In vitro* activity	Cytoactivity	Refs
	PKUMDL-YC1301		VS	SPR: 77 ± 22 μM	No inhibition	[42]
	10019-D3		RD	FP: 11 ± 4 μM	Not available	[32,34]
	5140069		VS	EMSA: 6 ± 3 μM	No inhibition	[36]
	6525237		VS	EMSA: 19 ± 6 μM	No inhibition	[36]
	6569963		VS	EMSA: 30 ± 10 μM	No inhibition	[36]
	5360134		VS	EMSA: 60 ± 20 μM	HL-60: 23 μM	[36]
EWS-FLI1	YK-4-279		HTS	SPR: 9.48 μM	TC32: 0.9 μM	[43]
p27	SJ710		HTS	NMR: 4.8 ± 1.3 mM	Not available	[44]
	SJ403		HTS	NMR: 2.2 ± 0.3 mM	Not available	[44]
NUPR1	Trifluoperazine		HTS	ITC: 5.2 μM	Cell viability: 26 ± 7% (10 μM)	[45]
AF9	PFWT peptide	Pen-LWVKIDLDLLSRV	Designed by mimicking binding partner AF4	Enzyme- linked protein-binding assay: showed efficacy at 10 μg/ml	B1 cell: LC_{50} 20 μM	[46]
TipA	Promothiocin A, nosiheptide, thiostrepton	Thiopeptide, highly modified multicyclic peptide antibiotics	BC	NMR confirmed	Showed significant efficacy	[47]
PTP1B	MSI-1436		BC	ITC (two-site model): 0.3 <5 μM and 3 μM		[48]

TABLE 1 (*Continued*)

Target	Compound	Structure	Identification	*In vitro* activity	Cytoactivity	Refs
Aβ peptide	SEN1576		RD	SPR: 3.8 μM	SH-SY5Y: 23 μM	[51]
	NQTrp		RD	ThT: 50 nM	Increased viability	[49,50]
α-synuclein	ELN484228		VS	Not available	Showed efficacy at 1 μM	[52]

[a] Abbreviations: BC, by chance; BSI, back-scattering interferometry; EMSA, electrophoretic mobility shift assay; RD, rational design; ThT, thioflavin T.

cur in this disease. The AF4–AF9 protein complex is a promising target for leukemia therapy. Based on peptide-mapping experiments, the peptide PFWT was designed by mimicking the AF9-binding site of AF4 [46]. PFWT inhibited the AF9–AF4 interaction in both an enzyme-linked protein-binding assay and in cultured cells. PFWT also inhibited the growth of several leukemia cell lines with less toxicity to hematopoietic progenitor cells.

The bacterial thiostrepton induced protein A (TipA) is a transcriptional regulator that represents a new multidrug resistance system [47]. Thiopeptide antibiotics, including thiostrepton, nosiheptide, and promothiocin A, interact with TipA, forming a covalent bond with a cysteine residue in TipA. Thiopeptide antibiotic binding induces folding of the disordered N-terminal region of the TipA antibiotic-binding domain, promotes a conformational switch in TipA, and upregulates *tipA* transcription, resulting in increased resistance. Notably, although they display a range of different sizes and structures, thiopeptide antibiotics specifically bind to TipA by interacting with its conserved antibiotic-binding domain, inducing folding of the disordered N terminus [47].

Protein tyrosine phosphatase 1B (PTP1B) is a validated therapeutic target for diabetes, obesity, and breast cancer. It has a positive role in obesity and insulin resistance, as well as in breast tumorigenesis. The aminosterol natural product, MSI-1436, from dogfish sharks, can allosterically inhibit the catalytic activity of PTP1B [48]. Dynamic light scattering and NMR experiments showed that MSI-1436 binds to the disordered C terminus of PTP1B, induces a conformational change that brings the C and N termini closer together, and locks PTP1B in an inactive state.

The accumulation and aggregation of Aβ is crucial for the development and progression of Alzheimer's disease (AD). Thus, inhibition of Aβ aggregation and fibril formation has been proposed as a potential therapeutic strategy for AD. Many compounds that inhibit aggregation and/or fibrillation have been found, both via experimental and virtual screening (VS) [49–51]. They function by accelerating aggregation, reducing fibril fragmentation, or removing toxic oligomers. Similarly, α-synuclein is the main component of filamentous inclusions, which are implicated in several neurodegenerative disorders, including Parkinson's disease (PD). The compound, ELN484228, was found to target α-synuclein, by VS [52].

Typical features of IDP-binding compounds

In the above examples, compounds specifically bind to IDPs, mostly by hydrophobic interactions among many polar interactions. This binding can reshape conformation ensemble, induce folding, change the overall protein structure with regards to folded and disordered domains, or prevent aggregation. Given the difficulties in developing iIDPs, learning from the chemical properties and structural features of successful examples will provide valuable insights into criteria that are useful in identifying new iIDPs. We collected 30 groups of reported iIDPs and analyzed their chemical features (Table 1). Several guidelines were applied in selecting iIDPs for analysis, with a major focus on identifying conserved and unbiased physicochemical properties and structural features. Overall, we selected compounds according to the following parameters: (i) compounds must directly bind to an IDP; (ii) only one representative compound was selected from each chemical analog group to avoid bias towards groups with large numbers of reported compounds; and (iii) peptides and macrocyclic compounds were removed.

Among the 30 groups of reported iIDPs, 26 representative compounds were selected for analysis. Fifteen commonly used molecular descriptors were calculated using XLogP3 and Canvas software (Schrödinger, Inc., version 10.2) [53]. For comparison, the same descriptors were also calculated for FDA-approved drugs from the DrugBank Database. Key physicochemical property distributions for iIDPs and FDA-approved drugs are shown in Fig. 1.

The most widely used drug-likeness filter is Lipinski's Rule of Five (ROF), which estimates the solubility, absorption, and permeability of a drug [54]. We found that >77% of iIDPs have an overall good adherence to the ROF. Molecular weights ranged from 246 to 685, with a mean of 377 (Fig. 1a). Furthermore, the molecular weight cutoff of 480 proposed in the Ghose Filter to indicate high drug-likeness was met by 77% of iIDPs [55]. Given that most iIDPs have lower molecular weights than known PPI interface inhibitors, further optimization is possible [56]. Compared with the broad distribution of logP values observed for FDA compounds (average logP: 1.68), iIDPs display a narrower distribution and tend to be more hydrophobic (average logP: 3.72), showing a limited number of hydrogen bond acceptors and donors

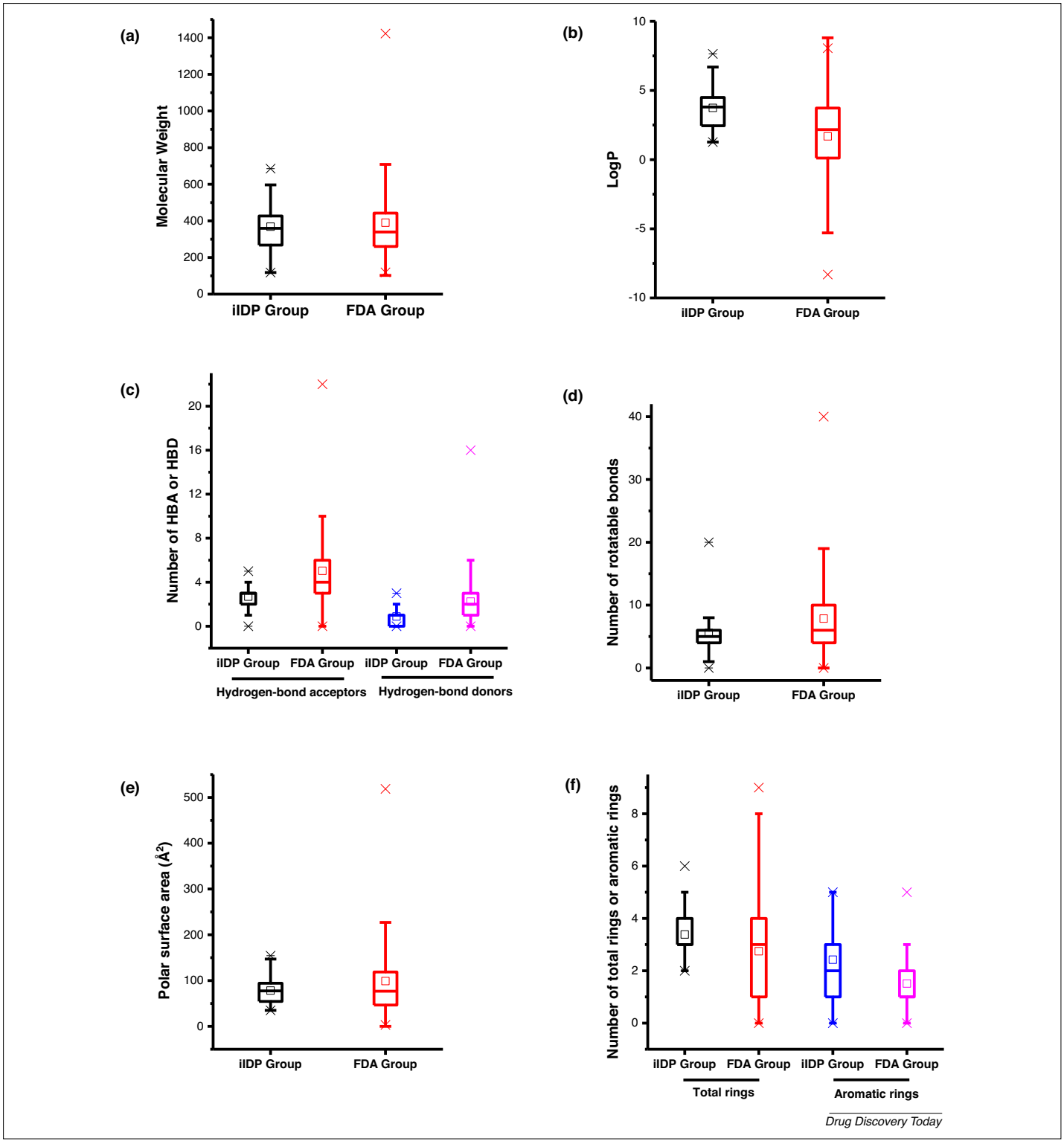


FIGURE 1

Key physicochemical property distributions of intrinsically disordered protein (IDP)-binding compounds (blue and black histogram) and the US Food and Drug Administration (FDA)-approved group (red and dark cyan histogram): (a) molecular weight, (b) LogP, (c) number of hydrogen-bond donors and acceptors, (d) number of rotatable bonds (NRB), (e) polar surface area (PSA), and (f) number of rings (NOR).

(Fig. 1b,c). With the exception of MSI-1436, these iIDPs have no more than three hydrogen-bond donors or five hydrogen-bond acceptors, which fits well within the ROF. Similar to FDA-approved drugs, a large portion of iIDPs abide by the Veber Rules (NRB $\leq$10 and PSA $\leq$140 Å^2), indicating a high probability of oral bioavailability [57] (Fig. 1d,e). All iIDPs also contain at least two rings, adhering to the MACCS-II Drug Data Report (MDDR) Rule (NOR $\geq$3) [58]. Furthermore, most of these rings are aromatic and contribute to hydrophobic interactions with mainly rigid scaffolds (Fig. 1f).

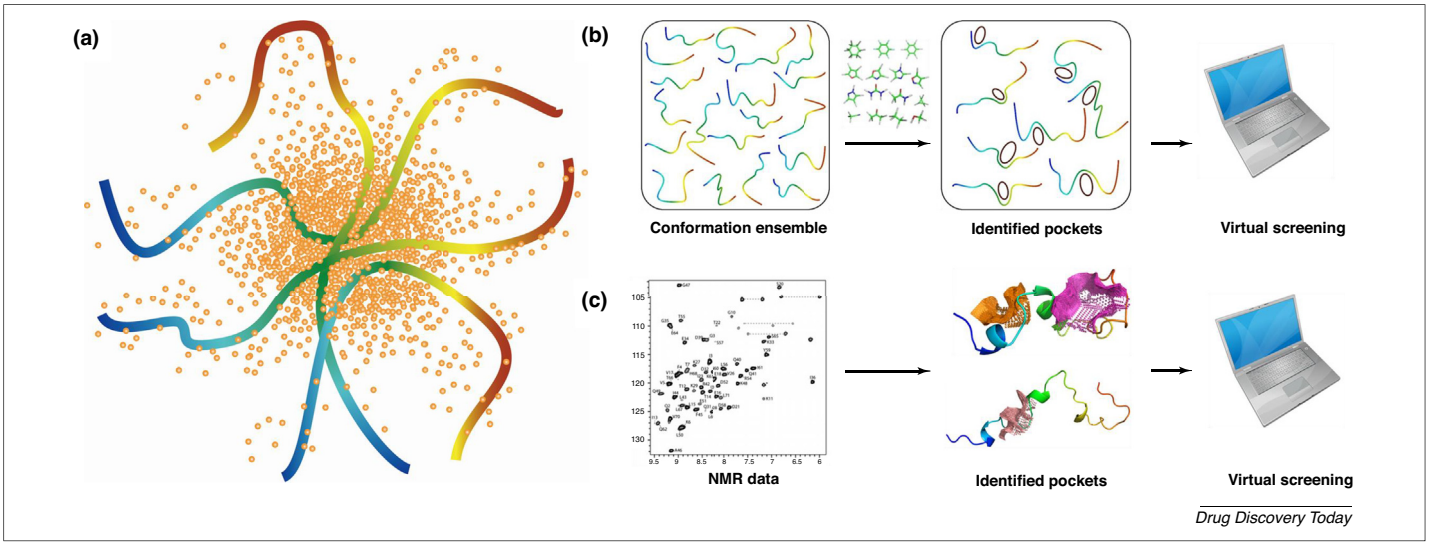


FIGURE 2

The fuzzy intrinsically disordered protein (IDP)–ligand interactions and methods for drug design targeting IDPs used in [49] and [39], respectively. (a) The IDP (color ribbons) remains disordered after ligand binding (yellow dots) and the binding site is also dynamic. The ligand shows variable binding probabilities in different IDP regions, reflecting binding specificity. In this case, the ligand prefers the center green region of the IDP. (b) The α-synuclein conformation ensemble was generated by molecular dynamics simulations restrained with nuclear magnetic resonance (NMR)-derived paramagnetic relaxation enhancement distance thresholds; 100 structures were randomly chosen, 22 conformations of which were selected for further analysis. The computational fragment probe mapping was then performed, and eight binding pockets were identified, located in eight different structures. A library of compounds was individually docked to each of the eight pockets. (c) The apo and holo structures for c-Myc$_{370–409}$ were built according to reported dihedral angles derived from NMR data; further refinement was performed. Three binding sites were predicted by the CAVITY program, and two libraries of compounds were individually docked into these pockets.

To summarize, iIDPs are generally both more hydrophobic and aromatic, and contain more rings than conventional drugs. Our analysis of known iIDPs revealed their key structural and physicochemical features, which could provide guidelines for future IDP drug discovery. Critically, this analysis should be further validated when more iIDPs become available.

Mechanisms of IDP–ligand interactions

Interactions between IDPs and their ligands are dynamic (Fig. 2a). Ligand binding causes a population shift in the IDP conformation ensemble, although their overall structure remains disordered [59–61]. Additionally, ligands often bind to IDPs at multiple sites, thus exhibiting 'ligand clouds' or 'specific-diffuse' binding mechanisms [60,62]. In the case of the c-Myc bHLH zip domain, the reported compounds bound at three independent binding sites, and both NMR experiments and molecular dynamics (MD) simulations confirmed that c-Myc remains disordered after ligand binding [33,34,59,60]. Additionally, ligands retain considerable mobility around c-Myc, accompanying its conformational changes, referred to as 'ligand clouds around protein clouds' [60]. Consequently, binding specificity between protein clouds and ligand clouds is reflected in binding probabilities, with the term 'specific diffuse' describing this dynamic and specific binding mechanism [62]. For Aβ, it has similarly been found that small-molecule ligands bind at different transient sites, and these interactions are dynamic [50,63,64]. Additionally, some protein–protein interactions involving IDPs were also found to be dynamic (termed 'dynamic fuzzy complexes'). For example, the disordered protein NuMA and the protein 4.1G C-terminal domain form a dynamic complex. Both proteins exhibit multiple binding sites, with a high frequency of binding sites indicating specificity [65]. Recently, structural ensemble modulation mechanisms were proposed to describe small-molecule binding to IDPs [66]. Upon binding to a small molecule, there are three possibilities: (i) one of the states predominates and the overall entropy reduces ('entropic collapse'); (ii) the structure ensemble redistributes without overall entropy change ('isentropic shift'); and (iii) the number of conformation states increases with an overall increase of entropy ('entropic expansion'). Given the difficulties associated with IDP–ligand binding measurements, we still lack a deep understanding of the mechanisms underlying these interactions [67].

Targeting IDPs via rational drug design

Over the past decade, computer-aided drug design (CADD) has significantly increased hit rates compared with HTS and has been widely used at various stages of drug discovery [68]. However, given the dynamic and ensemble nature of IDP structure, conventional structure-based design methods cannot be directly applied. Therefore, novel rational drug design methods for IDPs need be developed, as discussed below.

Zhu *et al.* developed a fragment-based mapping strategy to identify druggable pockets in the disordered monomeric Aβ peptide [69]. In this method, 45 conformation populations were identified by clustering the MD simulation trajectories. Small organic molecular fragments were then docked onto the surface of the Aβ peptide, with surface sites capable of binding multiple fragments defined as 'hot spots'. Neighboring hot spots were identified as potential binding pockets. Using this method, eight potential binding pockets were identified in α-synuclein, and VS was performed [52] (Fig. 2b). A library of 33 000 compounds was individually docked into the eight pockets, and cellular assays demonstrated that one identified compound, ELN484228, re-

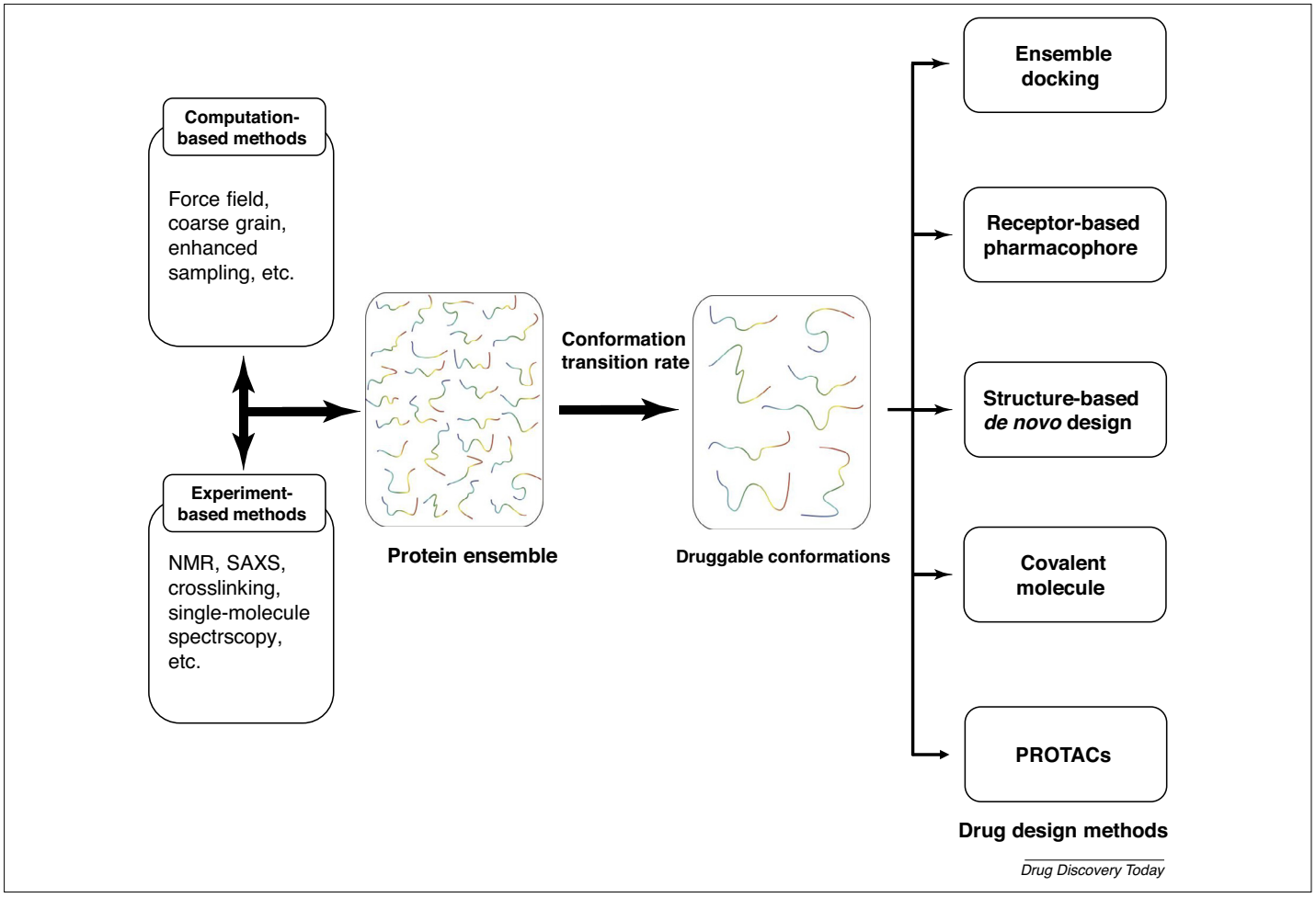


FIGURE 3

Proposed strategies for intrinsically disordered protein (IDP) drug design. In the first step, computation-based methods and experiment-based methods are either combined or used individually to generate an IDP ensemble. In the second step, druggable conformations are selected according to conformation clustering by considering the conformation transition rate. In the third step, five possible drug design approaches can be used to discover ligands for IDPs.

duced the cellular dysfunction mediated by elevated α-synuclein levels. However, whether ELN484228 can directly bind to α-synuclein remains to be verified both *in vitro* and *in vivo*.

A rational design approach was used to identify seven c-Myc-binding compounds [42] (Fig. 2c). The apo and holo c-Myc$_{370–409}$ conformations were derived from MD simulations refined with NMR data [60]. Three potential druggable pockets were identified using the binding site detection program, Cavity, and subjected to VS with a chemical library containing 210 000 compounds [70]. Among the top-ranking compounds, 273 were experimentally tested using CD, SPR, and NMR to assess *in vitro* binding. Five molecules capable of binding to c-Myc$_{370–409}$ *in vitro* showed significant cell growth inhibition, mRNA-level reduction of c-Myc target genes, and cell cycle arrest. This is the first successful example of c-Myc targeting via rational drug design, and ligands with increased potency are being developed.

Challenges and perspectives

Diverse functions and nonclassical regulation mechanisms make IDPs difficult to target. Specifically, IDPs can function as entropic chains, display sites, effectors, or assemblers, and IDP regulatory proteins commonly contain structured domains, disordered motifs, and linkers [6,71]. Many IDPs significantly impact the function and assembly of multiprotein molecular machines. Although these functions might necessitate the development of distinct drug design methods for various types of IDP, it will be necessary to characterize the dynamic conformation ensemble first. It is clear that improved drug design methods for targeting protein conformation ensembles remain a challenge (Fig. 3).

Combined with proper sampling methods, force field-based simulations can be used to obtain IDP ensembles. Given the existing bias of secondary structures and overestimation of solute–solute and intraprotein interactions, force fields originally developed for ordered proteins are not completely transferable to IDPs. However, significant progress in IDP force field development has been made, including: (i) refining backbone dihedral parameters to balance the propensity of secondary structures; (ii) balancing protein–water and protein–protein interactions to avoid overly collapsed structures; (iii) and developing residue-specific force fields, as well as implementing polarizable force fields [72–75]. A coarse-grained model for multiscale enhanced sampling was developed to provide the semi-

quantitative secondary structure of IDPs [76]. However, because of the limitations of theoretical model accuracy and sampling timescales, current computation-based methods alone might not be able to generate accurate IDP structure ensemble. In addition to computation-based methods, experiment-based methods, such as small-angle X-ray scattering (SAXS), NMR, single-molecule spectroscopy, and chemical crosslinking, provide valuable information for constructing conformation ensembles [77]. Restrictions of the experimental approaches include: (i) some experimental methods provide average conformation information rather than heterogeneous state information; (ii) one particular experimental method typically provides specific properties; and (iii) random and systematic errors are inevitable for experimental studies [78]. Thus, computational and experimental methods are often used in combination to more fully characterize IDP structures [78,79].

With current knowledge of IDP conformation ensemble, it remains unclear how to choose conformations for optimal drug design. Two recently proposed approaches were based on multiple representative conformations, without considering conformation transition rates and transition pathways [42,52]. We believe that it will be important to understand conformation transition pathways between apo and holo states, so as to more efficiently trap IDPs in nonfunctional conformations.

We propose several potential solutions for designing molecules to target selected conformations, as outlined below.

One possibility involves using ensemble docking to accommodate protein flexibility. In the case of the transactivation response element (TAR) RNA from HIV type 1, ~51 000 small molecules were docked independently into 20 conformers of the TAR^{NMR-MD} ensemble, yielding six active small molecules [80]. All six compounds bound to TAR using different modes. Although this ensemble docking strategy accommodates protein flexibility, the potential for multiple binding sites between small molecules and an IDP is neglected. Previous studies showed that six active compounds targeting c-Myc display 'multiconformational affinity'. That is, these compounds tend to bind to different conformations with similar affinity rather than bind to one preferred group of conformations. Therefore, the 'multiconformational-affinity' strategy for rational ligand design targeting IDPs was proposed [42].

Alternatively, receptor-based pharmacophore models can be used for VS. In this case, the same residues within conformation groups can be used to generate pharmacophore models, where the shape of pharmacophore features is flexible. The deformable pharmacophore models can then be applied for feature-based filtering.

Multitarget structure-based *de novo* drug design methods can also be extended to accommodate multiconformation drug design. For example, LigBuilder is a multiple purpose *de novo* design program that was recently developed to handle multiple targets [81–83]. In this system, initial fragments from a library are placed into possible binding sites, and new ligand structures are generated by the growth scheme. Compounds are then evaluated according to fitness calculation, and top-ranking candidates are retained for the next round of growth, until growth termination criteria are met.

Covalent ligand design is another potential strategy for IDP targeting. Covalent compounds bind to their targets for long periods of time and at low doses. Although IDPs contain many reactive residues, such as cysteine, lysine, and serine, selective modification can be achieved by considering both the environment and reactivity of modified residues. For example, it was shown that covalently modified cysteines normally have lower pK_a values, higher solvent exposure, and a nearby druggable ligand-binding pocket [84]. Additionally, crucial post-translational modification sites in IDPs represent an alternative choice for designing covalent ligands. When a suitable site is chosen, covalent fragment libraries can be used for virtual or experimental screening. Previously identified IDP-binding ligands can also be modified to include reactive groups. Following the virtual covalent reaction, MD simulations can be utilized to investigate ensemble conformational change. Furthermore, direct screening experiments can also be conducted. For example, reactive fragments can be divided into pools of several molecules for mass spectrometry screening after incubation with an IDP. Compound optimization can then be performed based on initial hits.

The other possibility is to use proteolysis-targeting chimaera (PROTAC) technology to degrade target IDPs. PROTACs exploit the cellular proteasome to selectively degrade target proteins [85]. IDP availability in cells is precisely controlled. Therefore, the PROTAC strategy has remarkable potency for IDPs, and can be used to directly target IDPs and reduce their cellular concentration, or to inhibit proteins that promote IDP degradation and thus increase IDP cellular concentration.

Concluding remarks

IDPs are closely associated with a range of diseases. Thus, developing drug design strategies to target IDPs opens new possibilities for disease intervention, and can provide improved arrays of chemical probes for understanding functions of IDPs. The field of IDP targeting via rational drug design is still in its infancy. As our understanding of mechanisms governing interaction between IDPs and their targets or ligands improves, new drug design strategies will emerge. With optimized experimental and theoretical approaches, more efficient drug design methods and successful applications are expected. In addition, identification of IDP-binding ligands will help uncover conserved features of iIDPs, which will further enhance the rate of IDP ligand discovery. Thus, the era of using IDPs as druggable targets is approaching, and we are expecting both challenges and exciting discoveries ahead.

Acknowledgments

We thank Chen Yu, Hanzhong Liu, and other members of the Lai group for their contributions and helpful discussions. We appreciate the contributions of, and discussions with, Zhirong Liu, Fan Jin, and the Liu group members. The authors' research work on IDP drug design was supported in part by the National Natural Science Foundation of China (21633001) and the Ministry of Science and Technology of China (2016YFA0502303, 2015CB910300).

References

1 Wright, P.E. *et al.* (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331
2 Dyson, H.J. *et al.* (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208
3 Tompa, P. (2011) Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21, 419–425
4 Uversky, V.N. (2011) Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* 43, 1090–1103
5 Uversky, V.N. *et al.* (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* 41, 415–427
6 van der Lee, R. *et al.* (2014) Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114, 6589–6631
7 Xue, B. *et al.* (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* 30, 137–149
8 Oates, M.E. *et al.* (2013) D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* 41, D508–D516
9 Wright, P.E. *et al.* (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16, 18–29
10 Brangwynne, C.P. *et al.* (2015) Polymer physics of intracellular phase transitions. *Nat. Phys.* 11, 899–904
11 Banani, S.F. *et al.* (2017) Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* 18, 285–298
12 Uversky, V.N. *et al.* (2008) Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annu. Rev. Biophys.* 37, 215–246
13 Metallo, S.J. (2010) Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* 14, 481–488
14 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
15 Lazo, J.S. and Sharlow, E.R. (2016) Drugging undruggable molecular cancer targets. *Annu. Rev. Pharmacol. Toxicol.* 56, 23–40
16 Dang, C.V. *et al.* (2017) Drugging the 'undruggable' cancer targets. *Nat. Rev. Cancer* 17, 502–508
17 Babu, M.M. *et al.* (2011) Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* 21, 432–440
18 Gsponer, J. *et al.* (2008) Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science* 322, 1365–1368
19 Iakoucheva, L.M. *et al.* (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* 323, 573–584
20 Babu, M.M. (2016) The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* 44, 1185–1200
21 Cheng, Y.G. *et al.* (2006) Rational drug design via intrinsically disordered protein. *Trends Biotechnol.* 24, 435–442
22 Anurag, M. and Dash, D. (2009) Unraveling the potential of intrinsically disordered proteins as drug targets: application to *Mycobacterium tuberculosis*. *Mol. Biosyst.* 5, 1752–1757
23 Dunker, A.K. and Uversky, V.N. (2010) Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Curr. Opin. Pharmacol.* 10, 782–788
24 Uversky, V.N. (2012) Intrinsically disordered proteins and novel strategies for drug discovery. *Expert Opin. Drug Discov.* 7, 475–488
25 Chen, C.Y. and Tou, W.L. (2013) How to design a drug for the disordered proteins? *Drug Discov. Today* 18, 910–915
26 Ambadipudi, S. and Zweckstetter, M. (2016) Targeting intrinsically disordered proteins in rational drug discovery. *Expert Opin. Drug Discov.* 11, 65–77
27 Vassilev, L.T. *et al.* (2004) In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848
28 Shangary, S. *et al.* (2008) Temporal activation of p53 by a specific MDM2 inhibitor is selectively toxic to tumors and leads to complete tumor growth inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3933–3938
29 Chang, Y.S. *et al.* (2013) Stapled α-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3445–E3454
30 Liu, Z.R. and Huang, Y.Q. (2014) Advantages of proteins being disordered. *Protein Sci.* 23, 539–550
31 Zhang, Y. *et al.* (2015) Binding cavities and druggability of intrinsically disordered proteins. *Protein Sci.* 24, 688–705
32 Yin, X.Y. *et al.* (2003) Low molecular weight inhibitors of Myc–Max interaction and function. *Oncogene* 22, 6151–6159
33 Follis, A.V. *et al.* (2008) Structural rationale for the coupled binding and unfolding of the c-Myc oncoprotein by small molecules. *Chem. Biol.* 15, 1149–1155
34 Hammoudeh, D.I. *et al.* (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.* 131, 7390–7401
35 Wang, H.B. *et al.* (2007) Improved low molecular weight Myc–Max inhibitors. *Mol. Cancer Ther.* 6, 2399–2408
36 Mustata, G. *et al.* (2009) Discovery of novel Myc–Max heterodimer disruptors with a three-dimensional pharmacophore model. *J. Med. Chem.* 52, 1247–1250
37 Clausen, D.M. *et al.* (2010) In vitro cytotoxicity and in vivo efficacy, pharmacokinetics, and metabolism of 10074-G5, a novel small-molecule inhibitor of c-Myc/Max dimerization. *J. Pharmacol. Exp. Ther.* 335, 715–727
38 Hart, J.R. *et al.* (2014) Inhibitor of MYC identified in a Kröhnke pyridine library. *Proc. Natl. Acad. Sci. U. S. A.* 111, 12556–12561
39 Müller, I. *et al.* (2014) Targeting of the MYCN protein with small molecule c-MYC inhibitors. *PLoS One* 9, e97285
40 Jung, K.Y. *et al.* (2015) Perturbation of the c-Myc–Max protein–protein interaction via synthetic alpha-helix mimetics. *J. Med. Chem.* 58, 3002–3024
41 Choi, S.H. *et al.* (2017) Targeted disruption of Myc–Max oncoprotein complex by a small molecule. *ACS Chem. Biol.* 12, 2715–2719
42 Yu, C. *et al.* (2016) Structure-based inhibitor design for the intrinsically disordered protein c-Myc. *Sci. Rep.* 6, 22298
43 Erkizan, H.V. *et al.* (2009) A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nat. Med.* 15, 750–756
44 Iconaru, L.I. *et al.* (2015) Discovery of small molecules that inhibit the disordered protein, p27(Kip1). *Sci. Rep.* 5, 15686
45 Neira, J.L. *et al.* (2017) Identification of a drug targeting an intrinsically disordered protein involved in pancreatic adenocarcinoma. *Sci. Rep.* 7, 39732
46 Srinivasan, R.S. *et al.* (2004) The synthetic peptide PFWT disrupts AF4–AF9 protein complexes and induces apoptosis in t(4;11) leukemia cells. *Leukemia* 18, 1364–1372
47 Habazettl, J. *et al.* (2014) Structural basis and dynamics of multidrug recognition in a minimal bacterial multidrug resistance system. *Proc. Natl. Acad. Sci. U. S. A.* 111, E5498–E5507
48 Krishnan, N. *et al.* (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.* 10, 558–566
49 Roni, S.A. (2010) Complete phenotypic recovery of an Alzheimer's disease model by a quinone–tryptophan hybrid aggregation inhibitor. *PLoS One* 5, e11101
50 Convertino, M. *et al.* (2011) Disordered binding of small molecules to Aβ(12–28). *J. Biol. Chem.* 286, 41578–41588
51 O'Hare, E. *et al.* (2014) Novel 5-aryloxypyrimidine SEN1576 as a candidate for the treatment of Alzheimer's disease. *Int. J. Neuropsychopharmacol.* 17, 117–126
52 Tóth, G. *et al.* (2014) Targeting the intrinsically disordered structural ensemble of α-synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease. *PLoS One* 9, e87133
53 Cheng, T.J. *et al.* (2007) Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* 47, 2140–2148
54 Lipinski, C.A. *et al.* (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26
55 Ghose, A.K. *et al.* (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. a qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1, 55–68
56 Kuenemann, M.A. *et al.* (2014) Which three-dimensional characteristics make efficient inhibitors of protein-protein interactions? *J. Chem. Inf. Model.* 54, 3067–3079
57 Veber, D.F. *et al.* (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623
58 Opera, T.I. (2000) Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.* 14, 251–264
59 Michel, J. and Cuchillo, R. (2012) The impact of small molecule binding on the energy landscape of the intrinsically disordered protein c-myc. *PLoS One* 7, e41070
60 Jin, F. *et al.* (2013) Ligand clouds around protein clouds: a scenario of ligand binding with intrinsically disordered proteins. *PLoS Comput. Biol.* 9, e1003249
61 Ban, D. *et al.* (2017) A small molecule causes a population shift in the conformational landscape of an intrinsically disordered protein. *J. Am. Chem. Soc.* 139, 13692–13700
62 Heller, G.T. *et al.* (2017) Sequence specificity in the entropy-driven binding of a small molecule and a disordered peptide. *J. Mol. Biol.* 429, 2772–2779
63 Chebaro, Y. *et al.* (2012) Structures of Aβ17–42 trimers in isolation and with five small-molecule drugs using a hierarchical computational procedure. *J. Phys. Chem. B* 116, 8412–8422

64 Zhang, T. *et al.* (2014) Atomic and dynamic insights into the beneficial effect of the 1,4-naphthoquinon-2-yl-L-tryptophan inhibitor on Alzheimer's Aβ1–42 dimer in terms of aggregation and toxicity. *ACS Chem. Neurosci.* 5, 148–159
65 Wu, S.W. *et al.* (2017) The dynamic multisite interactions between two intrinsically disordered proteins. *Angew. Chem. Int. Ed. Engl.* 56, 7515–7519
66 Heller, G.T. *et al.* (2018) Structural ensemble modulation upon small-molecule binding to disordered proteins. *J. Mol. Biol.* 430, 2288–2292
67 Heller, G.T. *et al.* (2017) Methods of probing the interactions between small molecules and disordered proteins. *Cell. Mol. Life Sci.* 74, 3225–3243
68 Ban, F.Q. *et al.* (2017) Best practices of computer-aided drug discovery: lessons learned from the development of a preclinical candidate for prostate cancer with a new mechanism of action. *J. Chem. Inf. Model.* 57, 1018–1028
69 Zhu, M. *et al.* (2013) Identification of small-molecule binding pockets in the soluble monomeric form of the Aβ42 peptide. *J. Chem. Phys.* 139, 035101
70 Yuan, Y.X. *et al.* (2013) Binding site detection and druggability prediction of protein targets for structure-based drug design. *Curr. Pharm. Des.* 19, 2326–2333
71 Tompa, P. (2014) Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. *Chem. Rev.* 114, 6715–6732
72 Piana, S. *et al.* (2015) Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* 119, 5113–5123
73 Huang, J. *et al.* (2017) Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 48, 40–48
74 Wu, H.N. *et al.* (2017) Significantly improved protein folding thermodynamics using a dispersion-corrected water model and a new residue-specific force field. *J. Phys. Chem. Lett.* 8, 3199–3205
75 Mercadante, D. *et al.* (2015) Kirkwood–Buff approach rescues overcollapse of a disordered protein in canonical protein force fields. *J. Phys. Chem. B* 119, 7975–7984
76 Liu, X.R. and Chen, J.H. (2017) HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations. *Phys. Chem. Chem. Phys.* 19, 32421–32432
77 Bhowmick, A. *et al.* (2016) Finding our way in the dark proteome. *J. Am. Chem. Soc.* 138, 9730–9742
78 Bonomi, M. *et al.* (2017) Principles of protein structural ensemble determination. *Curr. Opin. Struct. Biol.* 42, 106–116
79 Huang, W. *et al.* (2016) Theoretical modeling of multiprotein complexes by iSPOT: Integration of small-angle X-ray scattering, hydroxyl radical footprinting, and computational docking. *J. Struct. Biol.* 196, 340–349
80 Stelzer, A.C. *et al.* (2011) Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* 7, 553–559
81 Zhang, W.L. *et al.* (2017) Computational multitarget drug design. *J. Chem. Inf. Model.* 57, 403–412
82 Wang, R.X. *et al.* (2000) LigBuilder: a multi-purpose program for structure-based drug design. *J. Mol. Model.* 6, 498–516
83 Yuan, Y.X. *et al.* (2011) LigBuilder 2: a practical *de novo* drug design approach. *J. Chem. Inf. Model.* 51, 1083–1091
84 Zhang, W.L. *et al.* (2017) Statistical analysis and prediction of covalent ligand targeted cysteine residues. *J. Chem. Inf. Model.* 57, 1453–1460
85 Lai, A.C. and Crews, C.M. (2017) Induced protein degradation: an emerging drug discovery paradigm. *Nat. Rev. Drug Discov.* 16, 101–114