Methods 131 (2017) 22–32

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

In silico structure-based approaches to discover protein-protein interaction-targeting drugs

Woong-Hee Shin [a,1], Charles W. Christoffer [b,1], Daisuke Kihara [a,b,*]

[a] Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Computer Science, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 12 June 2017
Received in revised form 8 August 2017
Accepted 8 August 2017
Available online 9 August 2017

Keywords:
Protein-protein interaction
PPI drug
Virtual screening
Fragment-based drug discovery
Hotspot
Protein-ligand docking

ABSTRACT

A core concept behind modern drug discovery is finding a small molecule that modulates a function of a target protein. This concept has been successfully applied since the mid-1970s. However, the efficiency of drug discovery is decreasing because the druggable target space in the human proteome is limited. Recently, protein–protein interaction (PPI) has been identified as an emerging target space for drug discovery. PPI plays a pivotal role in biological pathways including diseases. Current human interactome research suggests that the number of PPIs is between 130,000 and 650,000, and only a small number of them have been targeted as drug targets. For traditional drug targets, *in silico* structure-based methods have been successful in many cases. However, their performance suffers on PPI interfaces because PPI interfaces are different in five major aspects: From a geometric standpoint, they have relatively large interface regions, flat geometry, and the interface surface shape tends to fluctuate upon binding. Also, their interactions are dominated by hydrophobic atoms, which is different from traditional binding-pocket-targeted drugs. Finally, PPI targets usually lack natural molecules that bind to the target PPI interface. Here, we first summarize characteristics of PPI interfaces and their known binders. Then, we will review existing *in silico* structure-based approaches for discovering small molecules that bind to PPI interfaces.

Contents

* Corresponding author at: Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.
E-mail address: dkihara@purdue.edu (D. Kihara).
[1] Co-first authors. These two authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ymeth.2017.08.006
1046-2023/

1. Introduction

The drug discovery process underwent a paradigm shift in the mid-1970s, when the human genome sequence was first cloned [1]. Before then, drug discovery focused on pharmacological and phenotypical characteristics of small molecules, and knowledge of a target molecule of a drug was not routinely considered. These days, however, the drug development process begins with identification of a drug target, which is usually a protein [1].

A core concept behind modern drug discovery is finding a small molecule that binds to a receptor protein in a binding pocket and modulates it. The target should have a well-formed binding pocket that interacts with a cognate small molecule, either a substrate or a cofactor. Proteins that have such binding pockets are referred to as druggable targets, and about 400 proteins have been identified as druggable [2] in the human proteome. Most of the targets fall into a small number of families: guanine nucleotide-binding protein (G protein)-coupled receptor, nuclear hormone receptor, ion channel, kinase, and protease [3,4]. Many drugs have been successfully developed on these targets. On the other hand, since most of the traditionally druggable target space has been extensively explored, drug discovery is now experiencing a drastic decrease in efficiency [5]. The pharmaceutical industry is facing a challenge to identify novel targets for treating diseases.

To improve the efficiency of drug development, protein-protein interactions (PPIs) have recently been spotlighted as a new type of drug targets for small molecules. PPIs play a key role in many biological processes and in many diseases [6], including cancer [7]. PPIs are expected to be successful drug targets for diseases that do not yet have effective treatments or drugs, which include diseases related to aging, e.g. Alzheimer's disease [8]. It is estimated that the human interactome contains 130,000 [9] to 650,000 PPIs [10], among which only a small fraction have been targeted for drugs [11]. In the last decade, more than 40 PPIs have been targeted, and several inhibitors have proceeded to clinical trials [12]. Table 1 lists current PPI drugs and targeted interfaces with the status of the drug development.

One successful class of small molecule protein-protein interaction inhibitors (SMPPIIs) is bromodomain inhibitors. Bromodomain recognizes acetylated histone tails to turn on genes [13]. There are already four SMPPIIs, I-BET762, CPI-0610, Ten-010, and OTX15, in clinical trials targeting cancer, which bind to the hydrophobic core of bromodomain and inhibit bromodomain-histone tail interaction [12]. They were discovered through various experimental methods. For example, I-BET762 was identified by cell-based high-throughput screening, and CPI-0610 was evolved from JQ1, a previously identified SMPPII for bromodomain [12].

For traditional drug targets, computer-aided drug design (CADD) has become a powerful and routinely-used approach that can effectively aid experimental discovery of new drugs [14–20]. Although there are some examples where CADD has been applied to SMPPII discovery [21–23], PPIs remain difficult targets since PPIs and their binding ligands have different characteristics from traditional targets. [24,25]. To understand the current status of CADD methods for PPI targets, here we review computational tools for discovering SMPPIIs, which try to overcome the hurdles that arise from the different nature of PPIs and compounds that interact with PPIs. In the following sections, we start by summarizing characteristics of PPIs and SMPPIIs. Next, we introduce databases that catalog SMPPIIs and PPI-SMPPII binding complexes. Subsequently, we review receptor-based CADD methods for SMPPII discovery. We also discuss successful examples of computational approaches to identify SMPPIIs and PPIs with disordered proteins.

2. Characteristics of PPIs and SMPPIIs

2.1. Structural features of PPIs

PPIs have physico-chemical characteristics distinct from those of protein-ligand binding pockets. Analysis of protein-protein complex 3 D structures shows that contact surfaces of PPIs range from 1,000 to 4,000 Å^2 [26], and the average area of the interfaces is 1,600 Å^2 [27], which is much larger than traditional ligand binding pockets, which vary from 300 to 1,000 Å^2 [26]. From a geometrical point of view, most PPI interfaces have planar shapes [28–30], except for intertwined interface structures [31], and they do not have a groove where a small molecule binds to. In contrast, traditional binding pockets have a concave protein surface [32].

A PPI interface can be further split into two regions: core and rim [33,34]. The core region is buried in the interface, while the rim surrounds the core and contains solvent accessible area [35]. The interface core residues are more hydrophobic, more conserved, and less mobile than rim residues [36,37]. Most PPIs are driven by the interaction between hydrophobic core regions. Thus, SMPPIIs are also hydrophobic.

Another characteristic of PPI target sites is that SMPPIIs bind to several sub-pockets in a PPI simultaneously. Fuller et al. examined crystal structures of SMPPIIs bound to PPI targets [38] using Q-SiteFinder, which predicts binding sites by calculating the interaction energy of a query protein surface and a methyl probe rolled across the surface [39]. Interestingly, it was found that SMPPIIs occupy six small pockets (volume ~55 Å^3) on average, whereas for traditional protein-ligand binding interfaces, the ligand occupies one or two large pockets (~260 Å^3) [38].

The last distinguishing feature of PPI target sites is its flexibility. A small molecule binding pocket is formed by transient surface fluctuation, which is not observed after protein-protein complex formation [40,41]. The different characteristics of PPI target sites and conventional drug binding sites are summarized in the upper part of Fig. 1.

2.2. Characteristics of SMPPIIs

Since PPIs have different characteristics from traditional drug binding pockets, SMPPIIs should also have features distinct from those of traditional drugs. After Morelli et al. examined chemical features of 39 ligands in the 2P2I database [42], he suggested the 'Rule-of-4 (RO4)' [43]: molecular weight (MW) > 400 Da, AlogP > 4, the number of rings (#Rings) > 4, and the number of hydrogen-bond acceptors (#HBA) > 4. Comparing RO4 with the Lipinski's rule-of-five, which defines drug-likeness of a molecule (MW < 500 Da, ClogP (computed logP) < 5, the number of hydrogen-bond donors (#HBD) < 10, and #HBA < 5) [44], SMPPIIs are relatively larger (i.e. MW is heavier), more hydrophobic, and more likely to form hydrogen bonds with proteins than traditional drugs. Due to their highly hydrophobic nature, designing SMPPIIs

Table 1
Example PPI targeting drugs.

PPI	Example Drug	Source[a]	Drug Status[b]	Drug PDB[c]	Example PDB Structures[d]
14-3-3	fusicoccin	TIMBAL	Experimental	5D3F	5D3F(C+)
Adenylyl Cyclase/GNAS	colforsin	TIMBAL	Experimental	–	–
Annexin A2/S100-A10	5-(4-Dimethylamino-phenyl)-3-hydroxy-1-(2-hydroxy-propyl)-4-(4-methyl-benzoyl)-1,5-dihydro-pyrrol-2-one	TIMBAL	Experimental	–	1BT6(C), 2HYW(M), 1A4P(M)
ARF1/Arno	brefeldin A	TIMBAL	Experimental	1R8Q	1R8S(C), 1R8 M(M)
Bcl2/Bax	navitoclax	2P2I, TIMBAL, & Ref [12]	Investigational	4LVT	2XA0(C), 1YSW(M+), 1F16(M)
BclXL/Bak	gossypol	2P2I and TIMBAL	Investigational	–	1BXL(C), 1YSG(M+), 2IMT(M)
Beta-catenin	2-phenoxy-, 2-[(5-methyl-2-furanyl)methylene]hydrazide	TIMBAL	Experimental	–	1G3J(C), 1T08(C)
BIII	gabapentin	TIMBAL	Approved	–	2LCM(M)
Bromodomain/histone	apabetalone	2P2I, TIMBAL & [12]	Investigational	4MR4	3UVW(C), 2YEL(M+), 2I8N(M)
c-Myc/Max	nadroparin	2P2I and TIMBAL	Approved	–	1NKP(C)
CCR5/gp120	maraviroc	[12]	Approved	4MBS	4MBS(M+)
CD154/TRAF	(2R)-{[(2′-[(biphenyl-3-ylmethyl)carbamoyl]-6′-{[(2R)-2-(pyrrolidin-1-ylmethyl)pyrrolidin-1-yl]carbonyl}-6-{[(2 R)-2-(1 H-pyrrol-1-ylmethyl)pyrrolidin-1-yl]carbonyl}-4,4′-bipyridin-2-yl)carbonyl] amino}(cyclohexyl)ethanoic acid	TIMBAL	Experimental	3LKJ	1FLL(C), 3LKJ(M+)
CD80 (B7-1)	2-(3-fluorophenyl)-6-methyl-4-(3-nitrophenyl)-1,6-dihydrodipyrazolo[3,4-b:3,4-d]pyridin-3(2 H)-one	TIMBAL	Experimental	–	1I8L(C)
CIAP1-BIR3/CASPASE-9	birinapant	2P2I	Investigational	–	3D9T(C), 3MUP(M+)
Clathrin	Pitstop 1	TIMBAL	Experimental	2XZG	4G55(M+), 2XZG(M+)
CRM1	verdinexor	TIMBAL	Investigational	–	3GB8(C)
Cyclophilins	cyclosporine	TIMBAL	Approved	1CWA	1AK4(C)
HDM2/p53	Nutlin-16	2P2I & TIMBAL	Experimental	1RV1	1YCR(C), 5TRF(M+)
HIF-1a	carvedilol	2P2I & TIMBAL	Approved	–	1H2K(C)
HPV-E2/HPV-E1	Bilh 434	2P2I & TIMBAL	Experimental	1R6N	1TUE(C), 1R6N(M+)
HRAS/SOS1	2-(2,4-dinitrophenyl)-N-(4-fluorophenyl)hydrazinecarbothioamide	2P2I & TIMBAL	Experimental	2LWI	1BKD(C)
IL-2/IL-2R	apremilast	2P2I & TIMBAL	Approved	–	1Z92(C), 1M48(M+), 1IRL(M)
Immunophilin FKBP1A	tacrolimus	TIMBAL	Approved	1FKJ	1B6C(C)
Integrase/LEDGF	raltegravir	2P2I	Approved	–	2B4J(C), 3LPT(M+)
Integrins	3-({1-[3-(4,5-dihydro-1H-imidazol-2-ylamino)-phenyl]-5-oxo-pyrrolidine-3-carbonyl}-amino)-3-pyridin-3-yl-propionic acid	TIMBAL & [12]	Experimental	–	1DZI(C)
KEAP1/NRF2	dimethyl fumarate	2P2I & TIMBAL	Approved	–	2FLU(C), 3VNG(M+), 1U6D(M)
MDM4/p53	WK298	2P2I & TIMBAL	Experimental	3LBJ	3DAB(C), 3LBJ(M+)
Menin/MLL	MI-2	2P2I & TIMBAL	Experimental	4GQ3	4GQ6(C), 4GQ3(M+)
Neuropilin-1	(S)-2-(3-(Benzo[c][1,2,5]thiadiazole-4-sulfonamido)thiophene-2-carboxamido)-5-guanidinopentanoic acid	TIMBAL	Experimental	3I97	4DEQ(C)
Plk1(PBD)	3-[3-chloro-5-(5-{[(1S)-1-phenylethyl]amino}isoxazolo[5,4-c]pyridin-3-yl)phenyl]propan-1-ol	TIMBAL	Experimental	–	1Q4K(C), 1Q4O(M)
PPAR-gamma/NCOA	rosiglitazone	TIMBAL	Approved	4EMA	1FM6(C+), 4EMA(M+)
Rac1	azathioprine	TIMBAL	Approved	–	1FOE(C)
Rad51/BRCA2	amuvatinib	TIMBAL	Investigational	–	1N0W(C)
RGS4/GNAI1	CCG-4986	TIMBAL	Experimental	–	–
S100B/p53	olopatadine	TIMBAL	Approved	–	1UWO(M)
SOD1	arimoclomol	TIMBAL	Investigational	–	1HL4(C), 4A7Q(M+)
STAT3	CHEMBL156046	TIMBAL	Experimental	–	–
STAT5	dasatinib	TIMBAL	Approved	–	–
Sur-2	adamanolol	TIMBAL	Experimental	–	–
Tak1/TAB	staurosporine	TIMBAL	Experimental	–	4L53(C+), 4L52(M)
TNFA/TNFA	ganoderic acid	2P2I & TIMBAL	Experimental	–	1TNF(C), 2AZ5(M+)
TNFR1A/TNFB	10-[3-(morpholin-4-yl)propyl]-4-(3-nitrophenyl)-11-sulfanylidene-5-thia-1,10-diazatricyclo[6.3.0.0^{2,6}]undeca-2(6),3,7-trien-9-one	2P2I	Experimental	1FT4	1TNR(C)
ToxT	virstatin	TIMBAL	Experimental	–	3GBG(M+)
Transthyretin/RBP	diflunisal	TIMBAL	Approved	3D2T	1QAB(C), 3D2T(M+), 1E3F(M)
Tubulin	griseofulvin	TIMBAL	Approved		4AQV(C)
UL42/UL30	BP1	TIMBAL	Experimental	–	1DML(C)

Table 1 (*continued*)

PPI	Example Drug	Source[a]	Drug Status[b]	Drug PDB[c]	Example PDB Structures[d]
VHL/HIF1A	(4R)-4-hydroxy-1-[(3-methylisoxazol-5-yl)acetyl]-N-[4-(1,3-oxazol-5-yl)benzyl]-L-prolinamide	2P2I & TIMBAL	Experimental	3ZRC	4AJY(C), 3ZRC(M+)
WDR5/MLL1	2-chloro-N-[2-(4-methylpiperazin-1-yl)-5- nitrophenyl]benzamide	2P2I & TIMBAL	Experimental	3SMR	4ESG(C), 3SMR(M+)
XIAP-BIR3/SMAC	birinapant	2P2I, TIMBAL, & [12]	Investigational	–	1G73(C), 1TFQ(M+), 1C9Q(M)
ZipA/FtsZ	N-methyl-N-[3-(6-phenyl[1,2,4]triazolo[4,3-B]pyridazin-3-yl)phenyl] acetamide	2P2I & TIMBAL	Experimental	–	1F47(C), 1S1 J(M+), 1F46(M)

[a] Sources of the information are from 2P2I, TIMBAL, or reference [12].
[b] Drug Status is according to the DrugBank classification criteria [115]: "Approved" means approved in some jurisdiction; "Investigational" means in some phase of the approval process in some jurisdiction; "Experimental" means shown by experiment to bind the target.
[c] In rows with a dash in the Drug PDB column, example structures marked with M+ or C+ contain a different ligand.
[d] In the example PDB entry column, (C) indicates a protein-protein complex structure; (M) indicates a single protein (not a complex); and + indicates the presence of a modulating ligand.

faces a challenge of bioavailability, which is defined as the proportion of a drug that reaches at the site of action [32]. The authors also mapped the SMPPIIs on the chemico-biological space. The space is described by two parameters: the binding efficiency index (BEI), which is defined as binding affinity divided by MW, and the surface efficiency index (SEI), defined as affinity divided by polar surface area [45]. The ligands in the 2P2I database have a mean BEI of 11.7 and a mean SEI of 7.2. These values are smaller than for traditional drugs, which have a mean BEI of 25.8 and a mean SEI of 14.5 [45]. The mean values of SMPPIIs correspond to the 'sub-optimal' region in terms of compound binding efficiency index (Fig. 4 in [45]), indicating that the ligands are hard to optimize (i.e. chemically modify) into appropriate drugs with proper oral-availability, pharmacokinetics, and ADME (Absorption, Distribution, Metabolism, and Excretion) behavior [44].

Villoutreix et al. analyzed physico-chemical features of molecules in the i-PPI database [46]. Similar to the 2P2I database analysis, SMPPIIs in the i-PPI database [47] are heavier (mean MW: 421 Da) and more hydrophobic (mean AlogP: 3.58) than regular drugs (mean MW: 341 Da, mean AlogP: 2.61) with statistical significance [46]. Interestingly, the mean topological polar surface area (TPSA) of SMPPIIs is 89 $Å^2$, whereas mean TPSA of regular drugs is 71 $Å^2$ [46]. This is a contradiction with the AlogP value result, which indicates SMPPIIs are more hydrophobic, but recalling that SMPPIIs are larger, they also have more polar atoms. Another interesting difference between SMPPIIs and traditional drugs is the number of multiple bonds and the number of aromatic rings in compounds. The median values of SMPPIIs are 17 multiple bonds (number of triple bonds, double bonds, and aromatic bonds), 2 benzene-like rings, and 13 aromatic carbons while those of the traditional drugs are 12 multiple bonds, 1 ring, and 9 aromatic carbons [46]. This rigidity of SMPPIIs might be beneficial to the binding affinity since a more flexible ligand incurs an entropic penalty. For example, increasing rigidity during lead optimization for MDM2/p53 interaction inhibitors improved binding affinities [48]. Fig. 2 shows two typical examples of PPI drugs, benzodiazepinedione and ABT-737. They have high MW (benzodiazepinedione: 581.2 Da, ABT-737: 813.4 Da), logP (benzodiazepinedione: 6.1, ABT-737: 9.2), polar surface area (PSA) (benzodiazepinedione: 87 $Å^2$, ABT-737: 164 $Å^2$), and multiple aromatic rings (benzodiazepinedione: 3, ABT-737: 5).

The authors of the i-PPI Database further analyzed the three-dimensional structure of the i-PPI database compounds [49]. They calculated 112 descriptors for SMPPIIs and traditional drug compounds and compared them. From an analysis of the moments of inertia of compound shapes, it was shown that existing drugs have rod- or disk-like structures, while SMPPIIs tend toward sphere-like conformations. Other descriptors, SHP2 and RDF070m, which account for atom distribution in the 3D space, showed that SMPPIIs have branched structures, which helps them bind to PPI interfaces by occupying multiple subpockets, as noted above [43]. The authors further selected four descriptors out of the 112 that were highly discriminative between SMPPIIs and regular drug molecules. The four descriptors were globularity, CW2, IW4, and EDmin3. Globularity measures the sphere-likeness of a compound. CW2 is the ratio between the hydrophilic surface area and the total surface area. IW4 measures imbalance between the barycenter of a molecule and the barycenter of its hydrophilic interacting regions. A high value of IW4 indicates that hydrophilic regions are localized in the molecule. EDmin3 is the third lowest local minimum of the interaction energy with a hydrophobic probe. Lower EDmin3 values indicate that a molecule can form more stable interactions with hydrophobic partners. SMPPIIs have lower EDmin3 and CW2 and higher IW4 and globularity than non-PPI drugs. From this result, an ideal SMPPII has a globular structure, a strong capacity to bind hydrophobic patches, a small proportion of exposed

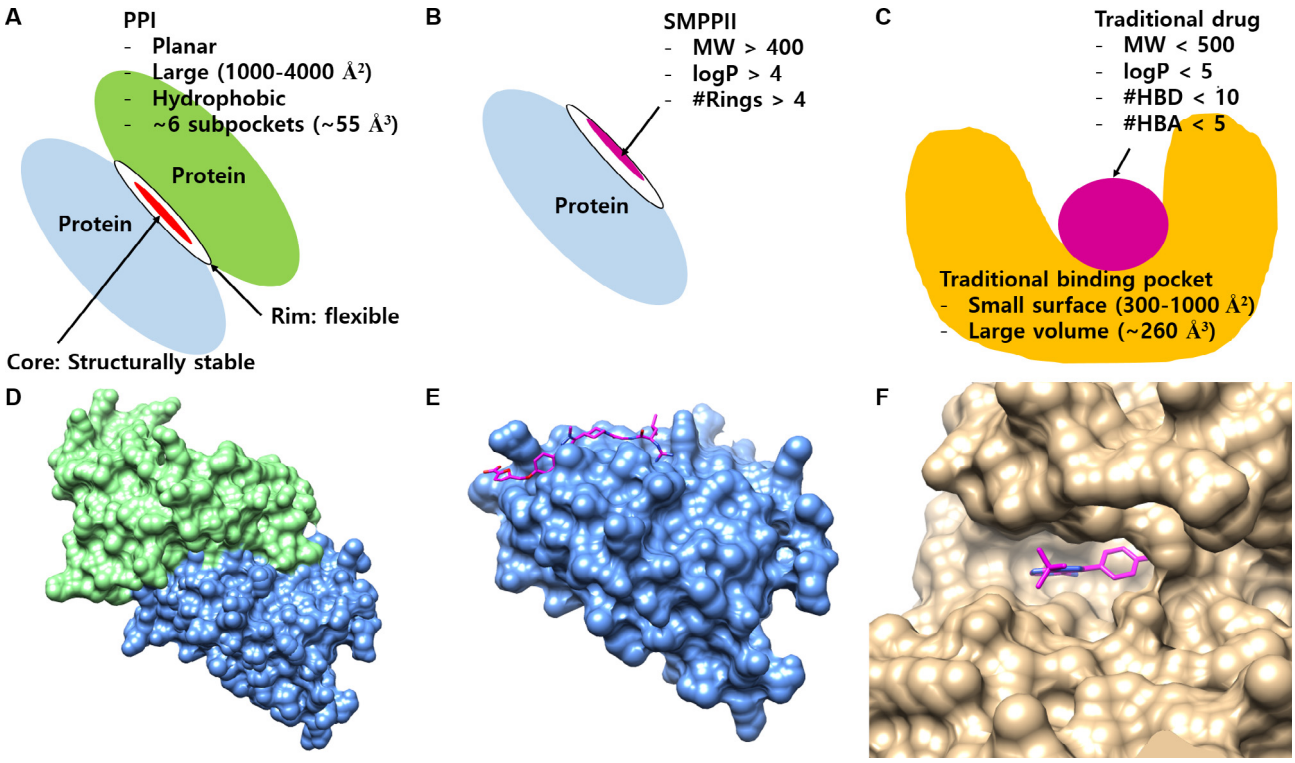

Fig. 1. A Illustration of PPI. Blue and green circles represent proteins. Red and white colored regions are core and surrounding rim regions of the PPI, respectively. B. SMPPII binds to PPI to inhibit binding. C. Illustration of a traditional binding pocket bound with a ligand. D. Example of PPI, a complex of IL-2 (blue) and IL-2α (green) (PDB ID: 1Z92). E. SP4206 (magenta), an inhibitor of IL-2 and IL-2α, binds with IL-2 (PDB ID: 1PY2). F. AC1L9LOG (magenta), an inhibitor of lymphocyte-specific kinase (gold), binds to its target protein (PDB ID: 1QPE). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

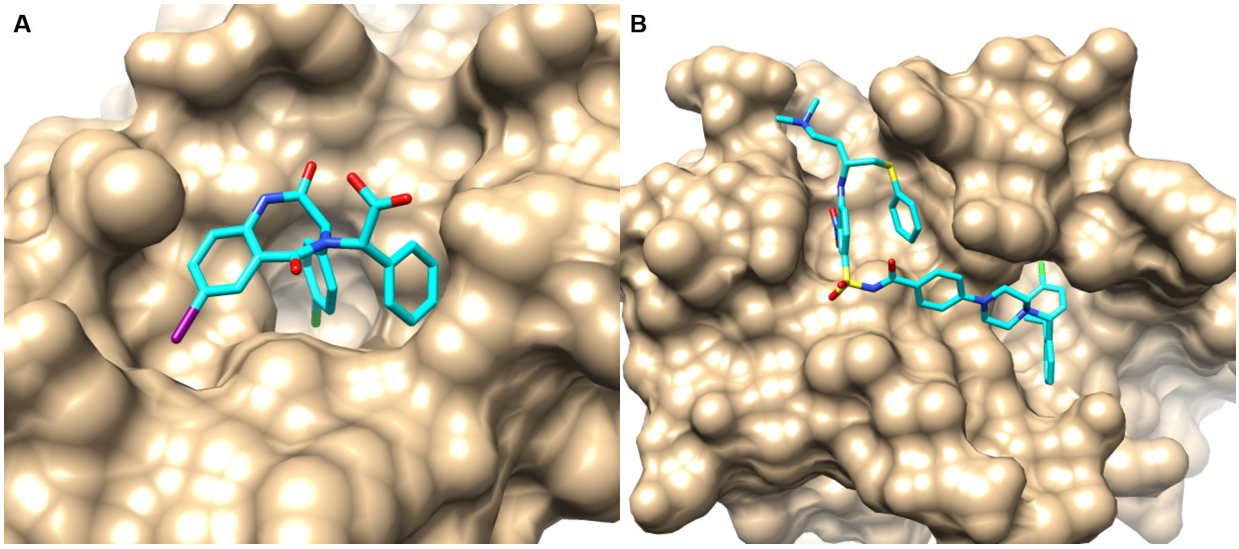

Fig. 2. Crystal structures of typical PPI drugs binding to their targets. A. Benzodiazepinedione binds to HDM2 (PDB ID: 1T4E). Benzodiazepinedione has three benzene-like rings out of four rings, 19 multiple bonds, 18 aromatic carbons. Its molecular weight, polar surface area, and logP are 581.2 Da, 87 Å^2, and 6.1, respectively. B. ABT-737 binds to Bcl-xL (PDB ID: 2YZJ) The drug has five benzene-like rings out of six rings, 34 multiple bonds, 30 aromatic carbons. Its molecular weight and polar surface area, and logP are 813.4 Da, 164 Å2, and 9.1, respectively. The ligands are colored in cyan and the protein surfaces are colored in gold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

hydrophilic surface with high concentrations of hydrophilic atoms [49]. The lower part of Fig. 1 summarizes the distinct features of SMPPIIs in comparison with traditional drug molecules.

3. Compound databases

Due to the importance of PPI targets, databases summarizing PPI drugs and their associated information have been developed. Table 2 lists PPI drug databases, as well as databases of PPI structures which have been experimentally solved or computationally modelled. PPI drug databases are briefly explained below.

3.1. 2P2I-DB

One such database is 2P2I-DB (http://2p2idb.cnrs-mrs.fr/) [42]. 2P2I-DB was constructed by data mining literature and exhaustive

Table 2
Databases of PPI drugs and protein complexes.

Database	URL	Description
2P2I-DB [42]	http://2p2idb.cnrs-mrs.fr/	known targets, along with inhibitors
TIMBAL [51]	http://mordred.bioc.cam.ac.uk/timbal/	known targets, along with inhibitors
iPPI-DB [47]	http://www.ippidb.cdithem.fr/	known targets, along with inhibitors
3D Complex [116]	http://www.3dcomplex.org/	experimentally solved protein complexes
3did [117]	http://3did.irbbarcelona.org/	protein domain interactions
Interactome3D [118]	http://interactome3d.irbbarcelona.org/	structural annotations of PPI networks
PDB [50]	http://www.wwpdb.org/	primary source of protein structure and complexes
EMDB [119]	http://www.emdatabank.org/	primary source of protein structures and complexes solved by electron microscopy

Listed in upper half of the table are databases of PPI drugs. Latter half lists databases of PPI structures.

search of the Protein Data Bank (PDB) [50]. It contains 17 PPI complexes and 56 small molecule inhibitors bound to their targets. It catalogs the planarity, eccentricity, secondary structure elements at the interface, gap volume, gap volume index, hydrogen bond density, salt bridges, buried surface area, root-mean-square deviation (RMSD) between bound and unbound structures, interface pocket volume, and number of interface residue segments.

3.2. TIMBAL

Another such database is TIMBAL (http://mordred.bioc.cam.ac.uk/timbal/) [51]. Initially, TIMBAL was curated manually, but it is currently maintained by searching ChEMBL with UniProt codes of the targets and their orthologs for relevant small molecule data in binding assays. TIMBAL currently catalogs 50 PPI targets, with 1137 UniProt entries, 8889 small molecules, and 1695 PDB entries.

3.3. iPPI-DB

Similarly to TIMBAL, iPPI-DB (http://www.ippidb.cdithem.fr/) [47] is not restricted to small molecules which have been co-crystalized with their target. iPPI-DB catalogs only inhibitors awarded world patents or published in medicinal chemistry journals. Targets with ambiguous information are excluded. Additionally, the literature must support the relation of the target to some disease. Also, the inhibitor must be a small non-peptidic compound, must not be metal based, must not be macrocyclic, and must not contain atoms different from C, N, O, S, P, and halogens. Also, results from certain experimental conditions are excluded. At time of writing, iPPI-DB contains 1756 inhibitors across 18 PPI families.

4. Computational methods 1: Screening full ligands

To aid the development of drugs for PPI targets, conventional structure-based screening methods, such as GOLD [17], Glide [18], Surflex-dock [19], and FlexX [20], have been successfully applied [22,52–54], often in combination with pharmacophore matching [55,56] (A pharmacophore method describes important molecular features of a ligand-receptor interaction, which can be useful to screen ligands that agree with the pharmacophore model of the interaction). They were successful particularly when the PPI interfaces analyzed have pockets where ligands can fit. On the other hand, there are efforts toward the development of new methods that address the distinct and difficult nature of PPI targets. In the next two sections we introduce computational methods developed or used for PPI targets. First, we introduce four methods that search full-ligands for PPIs. Then, in the subsequent section, we review fragment-based approaches applied to PPIs.

4.1. EleKit

EleKit is a method for predicting a ligand small molecule at a PPI interface of a receptor protein by evaluating the similarity of electrostatic potentials between the small molecule and the known protein ligand [57]. As input, it requires a structure of the receptor in complex with the ligand protein and another structure for the receptor in complex with a small molecule (Fig. 3). The structure of the receptor with a bound small molecule can be either experimentally determined or computationally modelled. First, the docking interface region between the receptor and ligand proteins is identified as the potential binding region (called a mask) of the small molecule ligand. Then, within the volume of this mask, electrostatic potentials are calculated at grid points from both the ligand protein and the small molecule. Finally, the small molecule will be predicted to interact with the receptor protein if the potentials from the small molecule and the ligand protein have a significant correlation (Spearman rank correlation coefficient) with each other.

The authors showed several cases where potentials of small molecule ligands at PPI interfaces exhibit high correlation to those of the ligand proteins, including the HIV-1 integrase: LEDGF/p57 complex and integrin: fibrinogen interactions, but also noted that some tested structures had insignificant correlations. EleKit was also tested in a virtual screening setting, where active compounds were sought among decoy compounds taken from the ZINC database [58]. Compounds were docked with MOE [59] to a receptor protein and their electrostatic potential correlation to that of its ligand protein was examined. Positive correlation was observed when active compounds were placed in their experimentally determined binding poses. Although further benchmarking and development may be needed to make the method practical, this method is interesting in that it explicitly considers the characteristics of a ligand protein of a receptor protein.

4.2. DARC and DARC2.0

DARC (Docking Approach using Ray-Casting) [25] is specifically designed to handle the shallow, flat pockets, which is typically observed in PPI sites. DARC works by comparing the topology of a protein surface pocket to the topology of a bound ligand at the binding site. First, an origin is selected at 30 Å from the center of the pocket in the direction of the protein's center of mass. From this origin, rays are cast through points on the pocket surface (Fig. 4). Based on how these rays intersect the pocket and the ligand, the following score is computed:

$$\text{DARC overlap score} = \sum_{R \in \text{rays}} \begin{cases} c_1 \cdot (\rho_{\text{ligand}} - \rho_{\text{pocket}}), \text{if } \rho_{\text{pocket}} < \rho_{\text{ligand}} \\ c_2 \cdot (\rho_{\text{pocket}} - \rho_{\text{ligand}}), \text{if } \rho_{\text{ligand}} < \rho_{\text{pocket}} \\ c_3, \text{if R does not intersect ligand} \\ c_4, \text{if R does not intersect pocket} \end{cases} \quad (1)$$

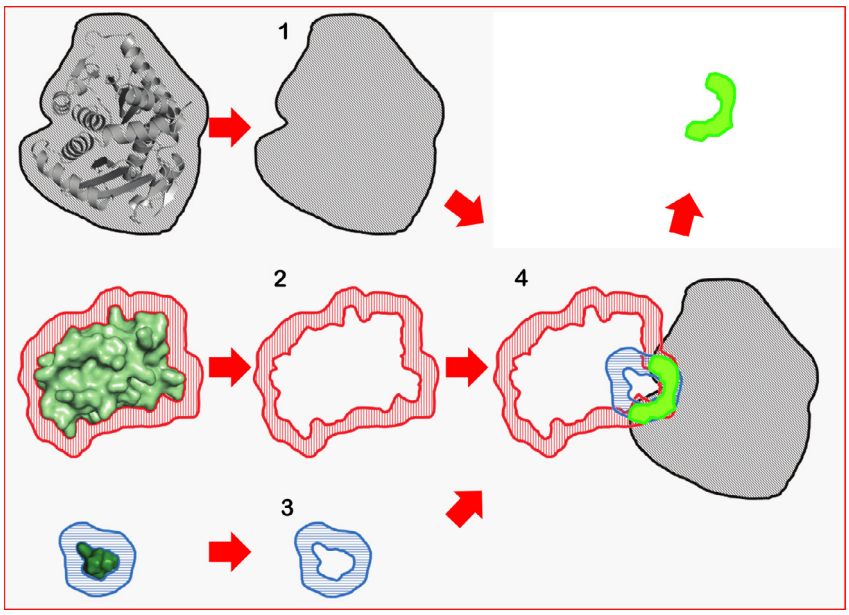

Fig. 3. Overview of the EleKit algorithm. A receptor protein (RP) is shown in gray, the ligand protein (LP) is in green, and the ligand small molecule (LSM) is shown at the bottom. (1) a mask of RP (gray) is created; (2) a near-but-not-inside mask of LP (red) is created; (3) a near-but-not-inside mask of LSM (blue) is created, (4) the logical conjunction of the three masks is used to select points to correlate from the electrostatic potentials of LP and LSM. The figure is Fig. 2 of the original paper of the Elekit method (Electrostatic similarities between protein and small molecule ligands facilitate the design of protein-protein interaction inhibitors, Voet A, Berenger F, Zhang KYJ, PLOS One, 8: e75762, https://doi.org/10.1371/journal.pone.0075762.g002). The caption is modified from the original. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

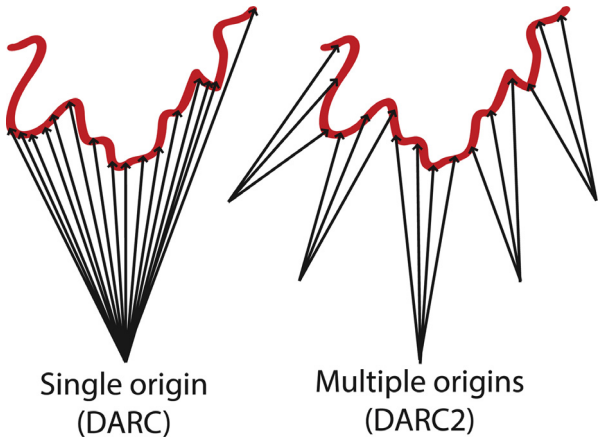

Fig. 4. DARC and DARC 2.0 ray casting. Left, In DARC, ray casting is performed from an origin behind the pocket. Right, In DARC 2.0, ray casting is performed from four additional origins. (The figure is modified from Fig. 2 of the original paper of DARC 2.0: DARC 2.0: Improved docking and virtual screening at protein interaction sites. Gowthaman R, Lyskov S, Karanicolas J, PLOS One, 10: e0131612, https://doi.org/10.1371/journal.pone.0131612.g002). The caption is modified from the original.

where ρ_{pocket} and ρ_{ligand} are the distance at which a given ray intersects the pocket and ligand, respectively. c_1 to c_4 are parameters that have positive values, optimized on a training dataset of known SMPPIIs and PPI targets they bind. Probable binding poses of a compound have lower DARC scores. The first term with c_1 applies to rays that reach the pocket before the ligand, indicating underpacking between protein and ligand. The second term applies to rays that reach the ligand before the pocket, indicating steric clash. The third term is for penalizing rays that do not intersect the ligand, i.e. when the ligand is small relative to the pocket. The last term is for rays that do not intersect the pocket, indicating that the ligand is larger than the pocket. Thus, a ligand binding pose on a pocket is evaluated by shape complementarity between the ligand and the pocket surface. Docking of a small molecule ligand is performed by optimizing binding poses of the ligand with a particle swarm optimization algorithm to minimize this score. For a ligand molecule, multiple conformations are generated, which are docked separately. DARC was applied to find inhibitors for human anti-apoptotic protein Mcl-1, where top scoring compounds were shown to have clear inhibition in a biochemical assay.

The authors further extended DARC in a subsequent version, DARC 2.0 [60]. First, additional ray casting is performed from four additional origins 45° from the original axis, capturing more topographic information (Fig. 4). A term incorporating electrostatic similarity was also added:

$$\text{Electrostatics score} = \sum_{\text{ligand atoms}} q_i \phi_i \tag{2}$$

where q_i is atom i 's partial charge as computed with molcharge [61], and ϕ_i is the electrostatic potential at its location. The electrostatic potential was set to zero in the protein's interior and a highly unfavorable value outside the defined binding site to penalize candidates leaving the binding site. To summarize, the interesting idea of the DARC algorithms is the use of ray casting, which is suitable for capturing geometric features of flat surface in PPI interfaces.

4.3. Exemplar

As illustrated earlier with Fig. 1, dealing with the plasticity of a protein surface is another challenge posed by PPI targets. It was reported that a small molecule-bound form of a PPI interface is usually different from either the isolated or the ligand protein bound form, highlighting the difficulty of the problem [40].

To consider alternative conformations of a PPI interface explicitly, this method first generates an ensemble of low-energy protein structures using Rosetta's Monte Carlo sampling implementation with an additional energy term which biases the sampling towards structures with deep pockets [62]. Then, from each pocket in the generated structure ensemble, a volumetric exemplar is con-

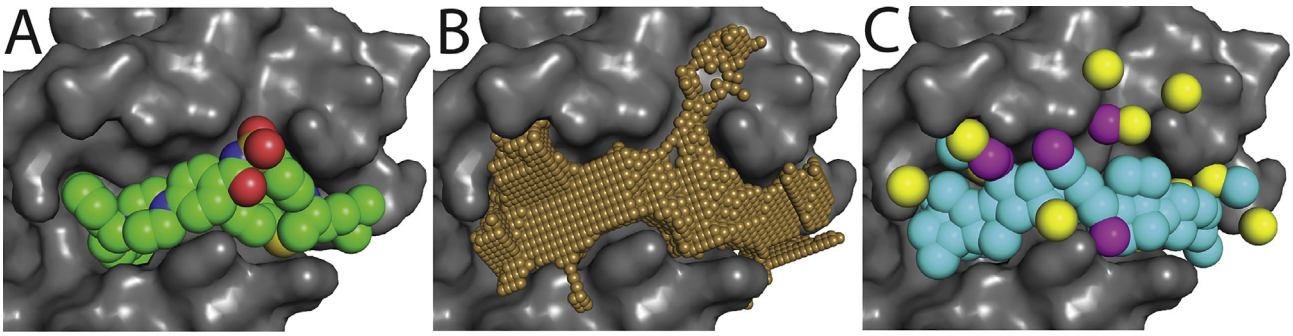

Fig. 5. Building "exemplars" from surface pockets. A, Bcl-xL (*grey surface*) is shown in complex with an inhibitor (*spheres*). B, The protein surface features a large pocket (*small spheres*) that is complementary in shape to the inhibitor. C, From this surface pocket, an exemplar is built: The exemplar is comprised of hydrogen bond donors (*yellow*) and acceptors (*magenta*) that complement surface polar groups on the protein, and hydrophobic atoms that fill the remainder of the surface pocket (*cyan*). (The figure is Fig. 1 in the original paper of this method: Selectivity by small-molecule inhibitors of protein interactions can be driven by protein surface fluctuations. Johnson Dk, Karanicolas J, PLoS Comp. Biol., 11: e1004081, https://doi.org/10.1371/journal.pcbi.1004081.g001). The caption is modified from the original. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

structed (Fig. 5). In the exemplar, locations for hydrogen bond partners for each donor and acceptor and polar groups are specified. The remaining volume is filled, representing hydrophobic regions. To screen a ligand library, a candidate ligand is superimposed to the exemplar using ROCS [63], which computes agreement of volumetric superimposition of two compounds.

5. Computational methods 2: Hot spots and fragment-Based methods

Since SMPPIIs often occupy multiple subpockets on their PPI target sites [43], fragment-based drug discovery (FBDD) has been highlighted as a successful alternative to full-ligand screening [64]. Successful examples of SMPPIIs using FBDD include an inhibitor targeting Bcl-X_L [65], which was evaluated as a combination treatment against solid tumors together with trametinib in a clinical trial in January 2017. In this case, FBDD was performed with experiments using nuclear magnetic resonance (NMR) for examining structure-activity relationships (SAR) of compounds. Computationally, FBDD on PPIs typically starts by finding a 'hotspot' and finding a fragment that binds to that hotspot. Hotspot residues are essential residues on the interface which contribute more than 2.0 kcal/mol to the binding free energy of the protein-protein interaction [66]. In this section, concepts of hotspot and recent *in silico* FBDD approaches are covered.

5.1. Hotspots on PPIs

Hotspot residues are energetically highly contributing residues in protein-protein interaction. In alanine scanning experiments, the hotspot residues alter binding free energy by more than 2.0 kcal/mol when mutated to alanine ($\Delta\Delta G_{bind} > 2.0$ kcal/mol). Since PPIs are much larger than traditional targets, FBDD approaches to design SMPPIIs start by finding hotspots on the targeted PPI and discovering fragments which bind to the hotspots with high affinities.

In typical PPIs, only 9.5% of interfacial residues are hotspot residues [67], and they are often clustered at the center of the interface [68]. The hotspots exist in both PPI partners, having complementary physico-chemical features to one another, by forming hydrogen bonds, salt bridges, or hydrophobic interaction [30]. In terms of amino acid composition, a systematic analysis revealed that most hotspot residues are tryptophan (21%), arginine (13%), or tyrosine (12%) [68]. Tryptophan is a unique amino acid, which has a bulky side-chain with hydrophobicity, can form π-π interaction, and hydrogen bonds [69]. Alanine mutation of tryptophan forms a large cavity, causing destabilization of the protein-protein complex [66]. In contrast, leucine, serine, threonine, and valine are rarely hotspots [67]. It is also shown that the hotspot residues are more conserved than other interfacial residues [70]. Another feature of PPI hotspot residues is their immobility. Kuttner and Engle performed steered molecular dynamics using crystal structures of IL2, MDM2, and PCNA [71], where a pulling force was applied on the surface residues. The results showed that highly immobile residues are located in the center of PPIs and form an interaction with SMPPIIs [71].

5.2. Characteristics of binding fragments on PPIs

In Section 2, we compared traditional drugs and SMPPIIs. Molecular characteristics of SMPPIIs, characterized by RO4, are bigger and more hydrophobic than regular drug compounds, characterized by RO5. Similarly, binding fragments for PPIs violate the Rule-of-three (RO3, MW < 300 Da, #HBD $\leq$ 3, #HBA $\leq$ 3, logP $\leq$ 3, #ROT $\leq$ 3, PSA $\leq$ 60 Å^2), which summarizes characteristics of regular fragments for FBDD that are used for traditional binding pockets [72]. Turnbull et al. statistically analyzed physico-chemical features of PPI binding fragments [30]: The average value of MW, logP, #ROT, and PSA are 278 Da, 2.48, 4.01, and 59.89 Å^2, respectively. PPI binding fragments have higher molecular weight and are more hydrophobic, as in the full-ligand comparison. They also measured ligand efficiency (LE, binding free energy divided by molecular weight). Average LE of PPI fragments is lower than that of regular fragments, as in full-ligand comparison [44].

5.3. Using conventional computational FBDD tools for identifying SMPPIIs

FBDD for PPI targets can be performed by applying existing conventional structure-based methods that dock whole ligands to target binding sites. Indeed, some successful cases were reported using protein-ligand docking (PLD) and pharmacophore search [73]. However, applying PLD for fragment binding can have two problems. First, since scoring functions work in an atom-additive way, small fragments have weak scores and may not be able to find a correct binding mode. Second, although fragments bind in a cooperative fashion [74], PLD generally samples one fragment per run, so it cannot consider cooperativity [75].

There are also studies that used a computational FBDD method for PPI targets. The FTMap program [76] performs docking of fragments computationally on a protein surface, somewhat mimicking the fragment docking process, which is called multiple solvent crystal structures (MSCS). It uses 16 organic molecules as probes: ethanol, isopropanol, isobutanol, acetone, acetaldehyde, dimethyl

ether, cyclohexane, ethane, acetonitrile, urea, methylamine, phenol, benzaldehyde, benzene, acetamide, and N,N-dimethylformamide. The probes are docked on the target protein surface using the fast Fourier transform algorithm, and the 2000 lowest binding poses for each probe are minimized with the CHARMM potential and clustered with a 4 Å radius. Then, from the top six clusters, ranked by the Boltzmann-averaged energy of the docked poses, hotspots are predicted [76]. It was shown that FTMap was able to identify hotspots on PPIs [77,78].

Molecular dynamics (MD) simulation, mimicking MSCS, is also used to predict hotspots [79]. The target protein structure is solvated with a mixture of explicit water and small molecule probes. In MD production stage, the probes frequently appear around hotspot regions while moving along the potential energy surface. By counting the frequency of the probes in trajectory, hotspot residues can be identified. Wang and Yang his used a MD-based MSCS method to predict hotspots on several PPIs successfully [80–82].

Chenglong Li and his colleague developed a program called MLSD (Multiple Ligand Simultaneous Docking) [83]. The program, as its name indicates, docks two compounds simultaneously using a scoring function that evaluates docking of each molecule on the target protein with the AutoDock4 scoring function [83] and interaction between the compounds. MLSD was applied to design a SMPPII for the IL-6/GP130 interface [84]. Before running MLSD, the authors identified hotspot residues by molecular dynamics simulation, and defined a docking box that covers the hotspots. A fragment library that binds to the hotspot residues was built from the known inhibitor and its analogues. MLSD identified potent fragment combinations, and the combinations were further grown to full ligands, which were found to be similar to two existing drugs (drug repositioning). The identified drugs from repositioning were validated as SMPPIIs by experiment [84].

5.4. PocketQuery and AnchorQuery

The Camacho group developed webservers called PocketQuery [85] and AnchorQuery [86], a set of computational methods aimed for drug discovery for PPI targets. For a query PPI complex, PocketQuery first predicts and ranks the interface residues in terms of druggability. The druggability of a residue is predicted by a machine learning method, support vector machine (SVM), which considers following features of a residue: the change of free energy of a residue upon complexation (ΔG^{FC}), estimation of the change of free energy of an alanine mutation ($\Delta\Delta G^{R}$), the change of solvent accessible surface area upon complexation (ΔSASA), ΔSASA in%, a sequence conservation score, and an evolutionary rate of the residue [85]. From the PocketQuery result, AnchorQuery suggests binding fragments for the highly druggable residues by matching the residues with pharmacophores of fragments in a pre-generated library [86]. The fragment library is biased towards anchor fragments, which mediates PPI complex formation. Thus, the library mainly composed of side-chains of amino acids and their analogues. In addition, to solve the synthetic availability issue of FBDD, the library is focused on multi-component reactions [86]. Distinct from general organic synthesis, which is sequential and a multi-step process, multi-component reactions yield the target molecule in one-pot reactions from three or more starting materials [87]. Using AnchorQuery, the authors were able to identify MDM2 inhibitors [86].

6. Successful cases of SMPPII discovery

In this section, successful examples of SMPPIIs from computational structure-based discovery are illustrated, including full-ligand screening and FBDD.

Full-ligand screening has been used to discover an inhibitor for the Kelch-like ECH-associated protein 1 (Keap1) and nuclear factor erythroid 2-related factor 2 (Nrf2) interaction. Activation of Nrf2 initiates an antioxidant and anti-inflammatory response. Keap1 represses the activation of Nrf2 by binding to it [88]. The Xing group screened the Specs database (153,611 molecules) with Glide [18] for the PPI interface on Keap1's side, and selected 90 compounds that showed higher binding affinities than a known inhibitor. Clustering the molecules with a threshold of Tanimoto coefficient 0.8 resulted 61 novel compounds. The authors further validated the compounds with experiment and obtained an inhibitor with lower micromolar K_D values [52].

In another example, full-ligand screening was successfully applied to the PPI of Small Ubiquitin-like Modifier (SUMO), a target protein, and proteins with the SUMO interaction motif (SIM). The covalent interaction between SUMO and its target, which is called sumoylation, regulates various processes, including DNA replication and repair, chromosome packing and dynamics, and cell proliferation and also known to be involved in important pathogenicities such as neurodegenerative diseases and cancer [89]. The reaction is mediated by the recognition of SUMO proteins by the SIM of a target protein. Voet et al. discovered an inhibitor of the SUMO-SIM interaction using Elekit [57]. First, they filtered compounds by considering a total negative charge and the number of hydrogen bond acceptor to be larger than donor atoms to satisfy complementarity to the positively-charged SUMO interface. The filtered compounds were docked to the interface using GOLD [17]. The poses were further validated with electrostatic similarity with DAXX peptide, a part of a SIM protein, using Elekit. The top four scored molecules showed micromolar K_D and IC_{50} values [90].

In an example of FBDD, Yu and colleagues recently discovered a dimerization inhibitor which binds to signal transducer and activator of transcription 3 (STAT3) dimer interface using MLSD [91]. When STAT3 is phosphorylated by a kinase, it dimerizes and translocates into the nucleus to bind to DNA, causing cancer [92]. Thus, targeting STAT3 can be an anti-cancer therapy. From a fragment database, 963 fragment combinations (FCs) were generated and docked to around known hotspots of the STAT3 monomer. The top 25 FCs were selected and linked with linear linker if the distance between the fragments was between 3 and 5 Å. Full-grown ligands were docked again on the interface. They selected 10 compounds if the binding affinity of the full-grown molecule was stronger than that of the FC. With experiment, they found an inhibitor that showed about 90% of inhibition percentage and micromolar IC_{50} [91].

In another example of FBDD, a MDM2-p53 inhibitor was identified using AnchorQuery [93]. p53 is a tumor suppressor that is involved in numerous human cancers. Around half of them are associated with mutation of the protein and the other half are linked to the interaction of p53 and MDM2. Since MDM2 is a negative regulator of p53, inactivating the protein [94], inhibiting the MDM2 and p53 interaction is a promising strategy for an anti-cancer drug. From a PDB structure of MDM2, AnchorQuery [86] detected four subpockets where fragments can bind; indole ring, aliphatic substituent, and benzyloxy group for Trp23, Phe19, and Leu26 subpockets, respectively. From the fragment information, they found a hit compound that is consistent with the identified fragments, which showed a K_i of 600 nM. They further optimized the compound successfully and improved K_i to 20 nM [93].

7. Intrinsically disordered proteins as drug targets

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) are proteins (or parts of proteins) that lack a stable native tertiary structure under physiological conditions

[95]. IDPs are flexible and form multiple PPI interfaces [96] involving many interactions in PPI networks. Among disease-associated proteins, more than 60% have IDRs longer than 30 amino acids [97]. There are examples of drugs targeting interactions with IDPs. NY2264 is an inhibitor targeting Myc, an IDP, and Myc-associated factor X (MAX) interaction [98]. Nutlin, an inhibitor of the p53-MDM2 interaction, binds to MDM2 and releases p53, which has a long IDR at its N-terminus [99,100].

It is not straightforward to target PPIs with IDPs because IDPs do not have solid structures and change conformations upon binding. One approach is to generate many conformations, "protein clouds", by dynamics simulation. The Lai group generated multiple conformations of c-Myc using all-atom replica exchange molecular dynamics simulation and found inhibitors for c-Myc/MAX interaction by virtual screening with Glide [101]. Identified compounds' inhibition was confirmed by a cell-based experiment.

Recently our group developed a computational method called IDP-LZerD [102], which predicts docking conformation of an IDP with its structured binding partner protein. This program would be helpful to find hotspots at the interface of PPIs of IDPs for drug design.

8. Conclusions

This review summarizes current computational structure-based approaches for discovering SMPPIIs. They are designed to overcome the hurdles of PPIs, which are different from conventional binding pockets in target proteins. Although these methods have been successfully applied in a couple of cases, the application has been limited to a small number of PPIs, which is partly due to the limited number of known structures of PPIs. To augment the limited structure data of PPIs, a potential future direction is to use computational protein-protein docking programs [103,104] to model PPIs, which are now available for various types of PPIs [102,105].

In the future, an alternative approach to modulating PPIs will become available: modifying proteins themselves by taking advantage of recent advances in gene editing techniques using CRISPR/CAS9 and TALEN [106–108]. An advantage of gene editing is that modification will be inherited by the next generation [109,110]. Techniques for modulating protein-protein interactions have been accumulated by artificial protein design experiments [111–114].

Currently, over 90% of PPIs in the human interactome are still untargeted. Therefore, there is a still a vast landscape of PPI targets to explore. For that, many tasks in computational method development remain, for example, prediction of PPI druggability.

Acknowledgements

This work was partly supported by the National Institutes of Health (R01GM123055, R01GM097528) and the National Science Foundation (IIS1319551, DBI1262189, IOS1127027, DMS1614777).

References

[1] G.P. Belfield, S.J. Delaney, Biochem. Soc. Trans. 34 (2006) 313–316.
[2] L. Jin, W. Wang, G. Fang, Annu. Rev. Pharmacol. Toxicol. 54 (2014) 435–456.
[3] A.L. Hopkins, C.R. Groom, Nat. Rev. Drug Discov. 1 (2002) 727–730.
[4] R. Santos, O. Ursu, A. Gaulton, A.P. Bento, R.S. Donadi, C.G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T.I. Opera, J.P. Overington, Nat. Rev. Drug Discov. 16 (2017) 19–34.
[5] J.W. Scannell, A. Blanckley, H. Boldon, B. Warrington, Nat. Rev. Drug Discov. 11 (2012) 191–200.
[6] M.W. Gonzalez, M.G. Kann, PLoS Comput. Biol. 8 (2012) e1002819.
[7] A.A. Ivanov, F.R. Khuri, H. Fu, Trends Pharmacol. Sci. 34 (2013) 393–400.
[8] L.-G. Milroy, T.N. Grossmann, S. Hennig, L. Brunsveld, C. Ottmann, Chem. Rev. 114 (2014) 4695–4748.
[9] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M.A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J.M. Sahalie, S. Cevik, C. Simon, A.-S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R.R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M.E. Cusick, F. P. Roth, D.E. Hill, J. Tavernier, E.E. Wanker, A.-L. Barabási, M. Vidal, Nat. Meth. 6 (2009) 83–90.
[10] M.P.H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H.J. An, M. Lappe, C. Wiuf, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 6959–6964.
[11] D. Rognan, Med. Chem. Commun. 6 (2015) 51–60.
[12] M.R. Arkin, Y. Tang, J.A. Wells, Chem. Biol. 21 (2014) 1102–1114.
[13] P. Filippakopoulous, S. Knapp, Nat. Rev. Drug Discov. 13 (2014) 337–356.
[14] W.-H. Shin, C.W. Christoffer, J. Wang, D. Kihara, J. Chem. Inf. Model. 56 (2016) 1676–1691.
[15] X. Zhu, W.-H. Shin, H. Kim, D. Kihara, J. Chem. Inf. Model. 56 (2016) 1088–1099.
[16] W.-H. Shin, X. Zhu, H.G. Bures, D. Kihara, Molecules 20 (2015) 12842–12861.
[17] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, J. Mol. Biol. 267 (1997) 727–748.
[18] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, J. Med. Chem. 47 (2004) 1739–1749.
[19] A.N. Jain, J. Med. Chem. 46 (2003) 499–511.
[20] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, J. Mol. Biol. 261 (1996) 470–489.
[21] L.R. Vidler, P. Filippkopoulos, O. Fedorov, S. Picaud, S. Martin, M. Tomsett, H. Woodward, N. Brown, S. Knapp, S. Hoelder, J. Med. Chem. 56 (2013) 8073–8088.
[22] S.M. Vogel, M.R. Bauer, A.C. Joerger, R. Wilcken, T. Brandt, D.B. Veprintsev, T.J. Rutherford, A.R. Fersht, F.M. Beockler, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 16906–16910.
[23] H.-P. Sun, Z.-Y. Jiang, M.-Y. Zhang, M.-C. Lu, T.-T. Yang, Y. Pan, H.-Z. Huang, X.-J. Zhang, Q.-D. You, MedChemComm 5 (2014) 93–98.
[24] D.M. Krüger, G. Jessen, H. Gohlke, J. Chem. Inf. Model. 52 (2012) 2807–2811.
[25] R. Gowthaman, R.A. Miller, S. Rogers, J. Khowsathit, L. Lan, N. Bai, D.K. Johnson, C. Liu, L. Xu, A. Anbanandam, J. Aubé, A. Roy, J. Karanicolas, J. Med. Chem. 59 (2016) 4152–4170.
[26] R. Sable, S. Jois, Molecules 20 (2015) 11569–11603.
[27] L. Lo Conte, C. Clothia, J. Janin, J. Mol. Biol. 285 (1999) 2177–2198.
[28] J.A. Wells, C.L. McClendon, Nature 450 (2007) 1001–1009.
[29] M. Brylinski, M. Gao, J. Skolnick, Phys. Chem. Chem. Phys. 13 (2011) 17044–17055.
[30] M. Gao, J. Skolnick, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 22517–22522.
[31] M. Gao, J. Skolnick, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 3784–3789.
[32] A.P. Turnbull, S.M. Boyd, B. Walse, Res. Rep. Biochem. 4 (2014) 13–26.
[33] N. Tuncbag, G. Kor, O. Keshin, A. Gursoy, R. Nussinov, Brief. Bioinform. 10 (2009) 217–232.
[34] P. Chakrabarti, J. Janin, Proteins 47 (2002) 334–343.
[35] R.P. Bahadur, P. Chakrabarti, F. Rodier, J. Janin, Proteins 53 (2003) 708–719.
[36] A. David, M.J.E. Sternberg, J. Mol. Biol. 427 (2015) 2886–2898.
[37] C. Yan, F. Wu, R.L. Jernigan, D. Dobbs, V. Honavar, Proteins J. 27 (2008) 59–70.
[38] J.C. Fuller, N.J. Burgoyne, R.M. Jackson, Drug Discov. Today 14 (2009) 155–161.
[39] A.T.R. Laurie, R.M. Jackson, Bioinformatics 21 (2005) 1908–1916.
[40] A.R. Arkin, M. Randal, W.L. DeLano, J. Hyde, T.N. Luong, J.D. Oslob, D.R. Raphael, L. Taylor, J. Wang, R.S. McDowell, J.A. Wells, A.C. Braisted, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 1603–1608.
[41] S. Eyrisch, V. Helms, J. Med. Chem. 50 (2007) 3457–3464.
[42] M.J. Basse, S. Berzi, R. Bourgeois, S. Bouzidi, B. Chetrit, V. Hamon, X. Morelli, P. Roche, Nucl. Acids Res. 41 (2013) D24–D27.
[43] X. Morelli, R. Bourgeois, P. Roche, Curr. Opin. Chem. Biol. 15 (2011) 475–481.
[44] C.A. Lipinski, Drug Discov. Today Technol. 1 (2004) 337–341.
[45] C. Abad-Zapatero, J.T. Metz, Drug Discov. Today 10 (2005) 464–469.
[46] O. Sperandio, C.H. Reynès, A.-C. Camproux, B.O. Villoutreix, Drug Discov. Today 15 (2010) 220–229.
[47] C.M. Labbé, M.A. Kuenemann, B. Zarzycka, G. Vriend, G.A.F. Nicolaes, D. Lagorce, M.A. Miteva, B.O. Villoutreix, O. Sperandio, Nucl. Acids Res. 44 (2016) D542–D547.
[48] M. Aeluri, S. Chamakuri, B. Dasari, S.K.R. Guduru, R. Jimidi, S. Jogula, P. Arya, Chem. Rev. 114 (2014) 4640–4694.
[49] M.A. Kuenemann, L.M.I. Bourbon, C.M. Labbé, B.O. Villoutreix, O. Sperandio, J. Chem. Inf. Model. 54 (2014) 3067–3079.
[50] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucl. Acids Res. 28 (2000) 235–242.
[51] A. P. Higueruelo, H. Jubb, T. L. Blundell, Database (Oxford) (2013) bat039.
[52] C. Zhuang, S. Narayanapillai, W. Zhang, Y.Y. Sham, C. Xing, J. Med. Chem. 57 (2014) 1121–1126.
[53] L. Borriello, M. Montès, Y. Lepelletier, B. Leforban, W.-Q. Liu, L. Demange, B. Delhomme, S. Pavoni, R. Jarray, J. Boucher, S. Dufour, O. Hermine, C. Garbay, R. Hadj-Slimane, F. Raynaud, Cancer Lett. 349 (2014) 120–127.
[54] S. Betzi, A. Restouin, S. Opi, S.T. Arold, I. Parrot, F. Guerlesquin, X. Morelli, Y. Collette, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 19256–19261.
[55] X. Xue, J.-L. Wei, L.-L. Xu, M.-Y. Xi, X.-L. Xu, F. Liu, X.-K. Guo, L. Wang, X.-J. Zhang, M.-Y. Zhang, M.-C. Lu, H.-P. Sun, Q.-D. You, J. Chem. Inf. Model. 53 (2013) 2715–2729.
[56] H.-P. Sun, Z.-Y. Jiang, M.-Y. Zhang, M.-C. Lu, T.-T. Yang, Y. Pan, H.-Z. Huang, X.-J. Zhang, Q.-D. You, Med. Chem. Commun. 5 (2014) 93–98.
[57] A. Voet, F. Berenger, K.Y.J. Zhang, PLoS One 8 (2013) e75762.
[58] J.J. Irwin, B.K. Shoichet, J. Chem. Inf. Model. 45 (2005) 177–182.

[59] Chemical Computing Group Inc (2011) Molecular Operating Environment (MOE) 10th edition. 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7: Chemical Computing Group Inc.
[60] R. Gowthaman, S. Lyskov, J. Karanicolas, PLoS One 10 (2015) e0131612.
[61] OpenEye Scientific Software, Santa Fe, NM. ZAP toolkit 2.2.0. Available: http://www.eyesopen.com.
[62] D.K. Johnson, J. Karanicolas, PLoS Comput. Biol. 11 (2015) e1004081.
[63] P.C. Hawkins, A.G. Skillman, J. Nicholls, J. Med. Chem. 50 (2007) 74–82.
[64] L.N. Makley, J.E. Gestwicki, Chem. Biol. Drug. Des. 81 (2013) 22–32.
[65] A.M. Petros, J. Dinges, D.J. Augeri, S.A. Baumeister, D.A. Betebenner, M.G. Bures, S.W. Elmore, P.J. Hajduk, M.K. Joseph, S.K. Landis, D.G. Nettesheim, S.H. Rosenberg, W. Shen, S. Thomas, X. Wang, I. Zanze, H. Zhang, S.W. Fesik, J. Med. Chem. 49 (2006) 656–663.
[66] A.A. Bogan, K.S. Thorn, J. Mol. Biol. 280 (1998) 1–9.
[67] I.S. Moreira, P.A. Fernandes, M.J. Ramos, Proteins 68 (2007) 803–812.
[68] E. Cukuroglu, A. Gursoy, O. Keskin, Nucl. Acids Res. 40 (2012) D829–833.
[69] U. Samanta, D. Pal, P. Chakrabarti, Proteins 38 (2000) 288–300.
[70] M. Guharoy, P. Chakrabarti, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15447–15452.
[71] Y.Y. Kuttner, S. Engle, J. Mol. Biol. 415 (2012) 419–428.
[72] M. Congreve, R. Carr, C. Murray, Drug Discov. Today 8 (2003) 876–877.
[73] B. Yu, Z. Huang, M. Zhang, D.R. Dillard, H. Ji, ACS Chem. Biol. 8 (2013) 524–529.
[74] P.C. Nair, A.K. Malde, N. Drinkwater, A.E. Mark, A.C.S. Med, Chem. Lett. 3 (2012) 322–326.
[75] H. Li, C. Li, J. Comput. Chem. 31 (2010) 2014–2222.
[76] D. Kozakov, L.E. Grove, D.R. Hall, T. Bohnuud, S.E. Mottarella, L. Luo, B. Xia, B. Beglov, S. Vajda, Nat. Protoc. 10 (2015) 733–755.
[77] B.S. Zerbe, D.R. Hall, S. Vajda, A. Whitty, D. Kozakov, J. Chem. Inf. Model. 52 (2012) 2236–2244.
[78] D. Kozakov, D.R. Hall, G.-Y. Chuang, R. Cencic, R. Brenke, L.E. Grove, D. Belgov, J. Pelletier, A. Whitty, S. Vajda, Proc. Natl. Acad. Sci. U.S.A. 108 (2011) 13528–13533.
[79] P. Ghanakota, H.A. Carlson, J. Med. Chem. 59 (2016) 10383–10399.
[80] C.Y. Yang, S. Wang, A.C.S. Med, Chem. Lett. 2 (2011) 280–284.
[81] C.Y. Yang, S. Wang, A.C.S. Med, Chem. Lett. 3 (2012) 308–312.
[82] C.Y. Yang, PLoS One 10 (2015) e0118671.
[83] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, J. Comput. Chem. 30 (2009) 2785–2791.
[84] H. Li, H. Xiao, L. Lin, D. Jou, V. Kumari, J. Lin, C. Li, J. Med. Chem. 57 (2014) 632–641.
[85] D.R. Koes, C.J. Camacho, Nucl. Acids Res. 40 (2012) W387–W392.
[86] D. Koes, K. Khoury, Y. Huang, W. Wang, M. Bista, G.M. Popowicz, S. Wolf, T.A. Holak, A. Dömling, C.J. Camacho, PLoS One 7 (2012) e32839.
[87] A. Dömling, I. Ugi, Angew. Chem. Int. Ed. 39 (2000) 3168–3210.
[88] K. Itoh, N. Wakabayashi, Y. Katoh, T. Ishii, K. Igarashi, J.D. Engel, M. Yamamoto, Genes Dev. 13 (1999) 76–86.
[89] K. Bettermann, M. Benesch, S. Weis, J. Haybaeck, Cancer Lett. 316 (2012) 113–125.
[90] A.R.D. Voet, A. Ito, M. Hirohama, S. Matsuoka, N. Tochio, T. Kigawa, M. Yoshida, K.Y.J. Zhang, Med. Chem. Commun. 5 (2014) 783–786.
[91] W. Yu, C. Li, W. Zhang, Y. Xiu, S. Li, J. Lin, K. Yu, M. Liu, L. Yang, J. Luo, Y. Chen, H. Sun, L. Kong, J. Med. Chem. 60 (2017) 2718–2731.
[92] C.Y. Liu, J.C. Su, M.H. Ni, L.M. Tseng, P.Y. Chu, D.S. Wang, W.T. Tai, Y.P. Kao, M. H. Hung, C.W. Shiau, K.F. Chen, Breast Cancer Res. Treat. 146 (2014) 71–84.
[93] E. Surmiak, C.G. Neochoritis, B. Musielak, A. Twarda-Clapa, K. Kurpiewska, G. Dubin, C. Camacho, T.A. Holak, A. Dömling, Eur. J. Med. Chem. 126 (2017) 384–407.
[94] C.J. Brown, S. Lain, C.S. Verma, A.R. Fersht, A.R. Lane, Nat. Rev. Cancer 9 (2009) 62–873.
[95] A.K. Dunker, C.J. Brown, J.D. Lawson, Z. Obradovic, Biochemistry 41 (2002) 6573–6582.
[96] D. Marasco, P.L. Scognamiglio, Int. J. Mol. Sci. 16 (2015) 7394–7412.
[97] S.J. Metallo, Curr. Opin. Chem. Biol. 14 (2010) 481–488.
[98] S. Fletcher, E.V. Prochownik, Biochim. Biophys. Acta 1849 (2015) 525–543.
[99] N. Bharathan, K. Bharatham, A.A. Shelat, D. Bashford, J. Chem. Info. Model. 54 (2014) 648–659.
[100] Q. Zhang, S.X. Zeng, H. Lu, Sub-Cell. Biochem. 85 (2014) 281–319.
[101] C. Yu, X. Niu, F. Jin, Z. Liu, C. Jin, L. Lai, Sci. Rep. 6 (2016) 22298.
[102] L.X. Peterson, A. Roy, C. Christoffer, G. Terashi, D. Kihara, PLoS Comput. Biol. 13 (2017) e1005485.
[103] V. Venkatraman, Y.D. Yang, L. Sael, D. Kihara, BMC Bioinform. 10 (2009) 407.
[104] M.F. Lensink, S. Velakan, S.J. Wodak, Proteins 85 (2017) 359–377.
[105] J. Esquivel-Rodriguez, Y.D. Yang, D. Kihara, Proteins 80 (2012) 1818–1833.
[106] J.A. Doudna, E. Charpentier, Science 346 (2014) 1258096.
[107] J. Boch, Nat. Biotechnol. 29 (2011) 135–136.
[108] Y.-K. Kim, G. Wee, J. Park, J. Kim, D. Baek, J.-S. Kim, V.N. Kim, Nat. Struct. Mol. Biol. 20 (2013) 1458–1464.
[109] W. Jiang, B. Yang, D.P. Weeks, PLoS One 9 (2014) e99225.
[110] A.E. Friedland, Y.B. Tzur, K.M. Esvelt, M.P. Colaiácovo, G.M. Church, J.A. Calarco, Nat. Meth. 10 (2013) 741–743.
[111] G. Schreiber, S.J. Fleishman, Curr. Opin. Str. Biol. 23 (2013) 903–910.
[112] L. Rosenfeld, M. Heyne, J.M. Shefman, N. Papo, Trends Biochem. Sci. 41 (2016) 421–433.
[113] S. Michielssens, J.H. Peters, D. Ban, S. Pratihar, D. Seeliger, M. Sharma, K. Giller, T.M. Sabo, S. Becker, D. Lee, C. Griesinger, B.L. de Groot, Angew. Chem. Int. Ed. Engl. 53 (2014) 10367–10371.
[114] J. Karanicolas, J.E. Corn, I. Chen, L.A. Joachimiak, O. Dym, S.H. Peck, S. Albeck, T. Unger, W. Hu, G. Liu, S. Delbecq, G.T. Montelione, C.P. Spiegel, D.R. Liu, D. Baker, Mol. Cell. 42 (2011) 250–260.
[115] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A.C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z.T. Dame, B. Han, Y. Zhou, D.S. Wishart, Nucl. Acids Res. 42 (2014) D1091–D1097.
[116] E.D. Levy, J.B. Pereira-Leal, C. Chothia, S.A. Teichmann, PLoS Comput. Biol. 2 (2006) e155.
[117] R. Mosca, A. Ceol, A. Stein, R. Olivella, P. Aloy, Nucl. Acids Res. 42 (2014) D374–D379.
[118] R. Mosca, A. Ceol, P. Aloy, Nat. Meth. 10 (2013) 47–53.
[119] C. Lawson, A. Patwardhan, M.L. Baker, C. Hryc, E.S. Garcia, B.P. Hudson, I. Lagerstedt, S.J. Ludke, G. Pintilie, R. Sala, J.D. Westbrook, H.M. Berman, G.J. Kleywegt, W. Chiu, Nucl. Acids Res. 44 (2016) D395–D403.