Molecular design opportunities presented by solvent-exposed regions of target proteins

Xiangyi Jia,a,#, Ji Yu,a,#, Zhongxia Zhoua, Jacob Kongstedb, Yuning Songc, Christophe Pannecouque,d, Erik De Clercq,d, Dongwei Kanga,*, Vasanthanathan Poongavanamb, *,†, Xinyong Liu,a,* and Peng Zhan,a,*

aDepartment of Medicinal Chemistry, Key Laboratory of Chemical Biology (Ministry of Education), School of Pharmaceutical Sciences, Shandong University, 44 West Culture Road, 250012 Ji’nan, Shandong, PR China.
bDepartment of Physics, Chemistry and Pharmacy, University of Southern Denmark, DK-5230 Odense M, Denmark.
cDepartment of Clinical Pharmacy, Qilu Hospital of Shandong University, Jinan, China.
dRega Institute for Medical Research, Laboratory of Virology and Chemotherapy, K.U. Leuven, Herestraat 49 Postbus 1043 (09.A097), B-3000 Leuven, Belgium.

This is the author manuscript accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/med.21581.

This article is protected by copyright. All rights reserved.
# These authors contributed equally.

* To whom correspondence should be addressed.

zhanpeng1982@sdu.edu.cn

†Present address: Biomedicinskrt Centrum (BMC), Department of Chemistry, Uppsala University, Husargatan 3, 75237 Uppsala, Sweden.

Abstract: Solvent-exposed regions, or solvent-filled pockets, within or adjacent to the ligand-binding sites of drug-target proteins provide opportunities for substantial modifications of existing small-molecular drug molecules without serious loss of activity. In this review, we present recent selected examples of exploitation of solvent-exposed regions of proteins in drug design and development from the recent medicinal-chemistry literature.

Graphical Abstract
Keywords: Solvent-accessible region, Protein/solvent interface, Solvent-exposed region, Solvent-filled pocket, Drug design, Medicinal chemistry

1. Introduction

Drug development is a time-consuming, multi-disciplinary campaign that ranges from lead discovery and optimization, target validation, preclinical evaluation, and clinical trials to regulatory approval [1]. Many apparently promising candidates are lost during the process for a variety of reasons, such as inappropriate pharmacokinetic (PK) properties or drug toxicity, while other agents fail to prove clinically useful, for example due to rapid development of drug resistance. Therefore, design of novel molecules with improved potency and drug-like properties remains an extremely important topic [2]. Fortunately, the biologically relevant chemical space is immense [3,4]. In particular, the structures of protein-based drug targets exhibit a high degree of plasticity, providing, for example, opportunities to make substantial modifications to known compounds in order to explore novel protein-inhibitor interactions in the drug-binding cavity or to exploit solvent-exposed regions, or solvent-filled pockets, within or adjacent to the ligand-binding sites [4].

In the structure-based lead optimization phase of drug development, open, solvent-exposed protein regions are often regarded as prospective binding sites for
the optimization of a chemical series in order to gain additional and specific protein-ligand interactions and to increase affinity for a target protein [5]. Charged and polar functional groups or water-solubility-enhancing groups (also referred to as solvent-exposed fragments or solvent-friendly fragments) are commonly added to a molecular scaffold as a means to improve the solubility, PK characteristics, and drug-like properties of a prospective drug candidate. Another approach is to fix the bio-active conformation of the ligand (rigidify the molecule).

Solvent-exposed regions can also be exploited to design dual-targeting/multi-targeting ligands, multivalent ligands, antibody-recruiting small molecules, proteolysis-targeting chimeras (PROTACs), DNA-encoded chemical libraries and affinity-based small molecular probes, and so on. A schematic representation of solvent exposed region-based ligand optimization is shown in Figure 1.

**Figure 1.** Illustration of ligand optimization through solvent exposed region on target.

![Diagram showing ligand optimization through solvent exposed region on target.](image)
From a thermodynamical point of view the process of ligand binding optimization can be seen as an optimization of the underlying binding energy, i.e. the change in the binding Gibbs energy. Such an energy change can be discussed in terms of its two components, the enthalpic and entropic contributions. Usually the desolvation entropy related to release of water molecules at the binding site together with the conformational entropy dominate the entropic contribution, whereas the enthalpic contribution is dominated by the desolvation enthalpy together with the so-called interaction enthalpy, i.e. the enthalpy change due to formation of drug-target interactions. Since the enthalpic contribution mainly originates from specific intermolecular interactions while the entropic contribution reflects unspecific interactions different optimization strategies should be followed in order to optimize either of these contributions. Moreover, since the enthalpic optimization is usually more difficult, starting the optimization procedure from enthalpic leads is generally advisable [6]. A detailed discussion of entropy-enthalpy compensation in ligand optimization is beyond the scope of this review.

Therefore, this review mainly focusses from medicinal chemistry point of view.

In general, the strategy of exploiting solvent-exposed regions has enormous potential for enhancing binding affinity, improving selectivity, overcoming drug resistance, and improving solubility or PK properties. In this review, we present some noteworthy examples of exploitation of solvent-exposed regions in drug
design and development from the recent medicinal-chemistry literature. Our aim is to suggest the potential generality of these approaches, and inspire researchers to apply them and related ideas to other drug discovery projects. It should be pointed out that the criteria for selecting examples in this review are not comprehensive, but typical.

2. Exploitation of solvent-exposed regions for structure-based drug design and development

2.1 To increase affinity or selectivity for a target protein by targeting solvent-exposed region

Based on the traditional concept of drug design, a potent ligand should smoothly enter and maximally occupy the binding pocket, interacting efficiently with the residues around the binding site. Therefore, the larger the binder, the stronger the interaction with residues around the pocket, provided that the ligand can smoothly enter the binding pocket. Consequently, structure-based design is typically focused on modification of a chemical series to gain additional binding contacts and to increase the binding affinity for the target protein [7].

Usually, detailed structural study of protein/binder complexes is a prerequisite for identifying suitable locations for rational optimization. For example, crystallographic studies of complexes of larger HIV-1 nonnucleoside
reverse transcriptase (RT) inhibitors (NNRTIs) with RT [8,9] showed that the Pro236 “hairpin loop” in RT has a structure close to the apo conformation, forming a relatively open pocket; thus, additional substituents introduced into the binding NNRTIs at positions that will be located in this solvent-exposed, open region should be well tolerated [8].

Nevirapine (1) was the first NNRTI approved for medical use in the United States, in 1996. Several 8-substituted nevirapine analogues were found to show excellent antiviral potency against a broad range of prevalent RT mutants and they also had favorable PK properties [10-14]. Especially, 2 and 3, phosphonate analogues of nevirapine, demonstrated excellent anti-viral activities against the WT and mutant Y181C HIV-1 (Figure 2). X-Ray crystallographic investigation revealed that while the dipyridodiazepinone platform of 3 (PDB ID: 4I7F) binds in a similar overall conformation to that of 1, it can be better accommodated in the mutant binding pocket. In addition, the extended C-8 substituent of the dipyridodiazepinone scaffold likely forms additional favorable interactions with RT (including with the main chains of Val106 and Ser105) that may stabilize binding of these 8-substituted nevirapine derivatives even with NNRTI-resistant mutations (Figure 3) [15].
Figure 2. Discovery of (A) 8-substituted nevirapine-based and (B) pyridinone-based NNRTIs targeting the Pro236 “hairpin loop” at the protein/solvent interface.

Figure 3. Crystal structures of HIV-1 RT in complex with: (A) 3 (PDB ID: 4I7F) and (B) 5 (PDB ID: 2BE2).

Compared with pyridinone derivative 4, R221239 (5) contains a flexible linker and a linked furan moiety that permits close interactions with Val106, Phe227, and Pro236 (Figure 3). These additional binding contacts improved the potency of...
R221239 against a panel of HIV-1-resistant strains bearing the V106A, Y188L, and F227C mutations (Figure 2) [16].

**Figure 4.** Structure-based optimization targeting the P236 “hairpin loop” at the protein/solvent interface to obtain structurally elaborated DAPY-based NNRTIs.

In an effort to more fully explore the Structure-activity relationships (SARs) of the diarylpyrimidine (DAPY)-based NNRTIs (exemplified by 6, etravirine, ETV) and potentially attenuate the resistance of the existing mutants, a further investigation of the interactions of the DAPYs with the solvent-exposed region of RT resulted in the identification of piperidine-linked aminopyrimidine and thiophene[3,2-d]pyrimidine derivatives, which exhibited broad-spectrum activity with low (single-digit) nanomolar EC_{50} values toward a panel of WT, single-mutant, and double-mutant HIV-1 strains [17,18]. Compared with 6, piperidine-linked aminopyrimidine derivatives 7 and 8 possess broad potency against resistant mutant viruses (including the K103N/Y181C and Y188L mutants). Compared with ETV, the piperidine-linked thiophene[3,2-d]pyrimidines 9 and 10 were exceptionally active against the whole viral panel, including wild-type (WT), L100I, K103N, E138K, Y181C, Y188L, F227L/V106A and
K103N/Y181C (Figure 4) [18]. Importantly, 9 has lower cytotoxicity (CC_{50} > 227 μM) and a huge selectivity index (SI) value of >159101. Moreover, 9 also displayed favorable PK and safety profiles in rats *in vivo* [18]. Crystal structure analyses of these molecules indicated that, besides a hydrogen bond from the aminopyrimidine imino group to the backbone carbonyl of Lys101, the piperidine nitrogen probably forms an additional hydrogen bond with the main chain of Lys103 *via* a bridging water molecule [17].

HIV-1 RT contains both DNA polymerase and ribonuclease H (RNase H) domains, converting the viral genomic RNA to dsDNA in infected host cells. Because of its key role in HIV genome replication, the RT-associated RNase H has recently been proven to be an appealing target [19]. The 3-hydroxypyrimidine-2,4-dione (HPD) derivative 11 was previously reported as a potent and selective RNase H inhibitor, but it lacked antiviral activity in cell culture. Recently, taking account of the solvent-exposed region of RNase H, a critical redesign of the HPD chemotype was conducted, featuring an additional group at the C5 position; this afforded double-winged HPD 12 with significantly enhanced RNase H inhibitory activity and robust antiviral potency (Figure 5)[20].

**Figure 5.** Optimization of peripheral substituents for the protein/solvent interface, affording a more potent HIV RNase H inhibitor.
Figure 6. Optimization of peripheral substituents for the solvent-exposed region of gp120, affording more potent HIV inhibitors.

Computational, thermodynamic, and crystallographic data have been used to optimize a series of small-molecular ligands targeting the Phe43 cavity of HIV gp120 [21,22]. Interestingly, biological evaluation showed that compounds 16 and 17, derived from CD4 mimics 13-15, could form additional binding contacts with the solvent-exposed region of gp120 (Figure 6), and inhibited HIV-1 entry into target cells with substantially higher activity and neutralization breadth than previous analogues, while retaining high selectivity for HIV-1 [23,24].

G protein-coupled receptors (GPCRs) are central to many physiological processes. The phenylethylamine chain in the adenosine A2A receptor antagonist ZM241385 (18) and long and flexible piperazine-bearing substituent of UK-432097 (19) extended above the principal adenosine binding cavity, are mostly directed toward the more solvent-exposed extracellular region. Different adenosine receptors have different residues located in this area, and these interactions appear to be important in designing synthetic A2A selective antagonists (Figure 7) [25].

Based on this information, further optimization of ZM241385 and UK-432097 resulted in the identification of 20 and 21 as potent and selective adenosine A2a
receptor antagonists, respectively, with low nanomolar affinity toward the A2a receptor and good selectivity with respect to the A1 receptor (400-fold in some cases) [26,27].

![Chemical structures](image)

**Figure 7.** Novel derivatives of ZM241385 (18) and UK-432097 (19) as potent and selective adenosine A2a receptor antagonists by targeting solvent-exposed region.

![Chemical structures](image)

**Figure 8.** Structure-based design of highly selective and potent GRK2 inhibitors. 
(a) Structural optimization of Paroxetine via solvent-exposed region based on the binding modes of 23 (PDB 5UKM) in the GRK2 active site. The pyrazole of 23 extends out of the active site toward solvent. (b) Rational lead hybridization based on crystallographic superposition of GRK2 in complex with 24 (salmon) and 25 (purple) (PDB entries 4PNK and 3PVW, respectively), by taking full advantage of solvent-exposed region.

Regulation of GPCRs is controlled by GPCR kinases (GRKs), some of which have been implicated in heart failure. In heart failure, the β-adrenergic receptors...
(βARs) become desensitized and uncoupled from heterotrimeric G proteins. This process is initiated by G protein-coupled receptor kinases (GRKs), some of which are upregulated in the failing heart, making them promising drug targets. The selective serotonin reuptake inhibitor, paroxetine, was previously identified as a GRK2 inhibitor. Utilizing a structure-based drug design approach, paroxetine (22) was modified toward the solvent to generate a highly potent and selective GRK2 inhibitor 23, with an IC₅₀ of 30 nM against GRK2 and greater than 230-fold selectivity over other GRKs and kinases. Furthermore, 23 displayed a 100-fold improvement in cardiomyocyte contractility assays over paroxetine and a plasma concentration higher than its IC₅₀ for over 7 h. In addition, 23 was crystallized in complex with GRK2 to give insights into the structural determinants of potency and selectivity of these inhibitors (Figure 8A) [28].

Rational lead hybridization based on crystallographic overlays is an efficient drug discovery approach [29]. GSK180736A (25), developed as a Rho-associated coiled-coil kinase 1 (ROCK1) inhibitor, was identified as an inhibitor of GRK2 and co-crystallized in the active site (PDB code: 4PNK). Guided by its binding pose overlaid with the binding pose of a known potent GRK2 inhibitor 24 (PDB code: 3PVW), a library of hybrid inhibitors was developed. This campaign produced several compounds possessing high potency and selectivity for GRK2 over other GRK subfamilies, PKA, and ROCK1. The most selective compound, 26 (CCG-224406), had an IC₅₀ for GRK2 of 130 nM, >700-fold selectivity over other GRK subfamilies, and no detectable inhibition of ROCK1. Four of the new inhibitors were crystallized with GRK2 to give molecular insights into the binding and kinase selectivity of this class of inhibitors (Figure 8B) [30].

This article is protected by copyright. All rights reserved.
Kinase inhibitors have proved to be drugs of major therapeutic and economic importance. A useful strategy for the design of highly potent or subtype-selective kinase inhibitors is to target specific residues near a solvent-exposed region of the protein (Figure 9) [31-34].

Since the G1202R mutation is located at the solvent front of the anaplastic lymphoma kinase (ALK) domain close to the inhibitor-binding pocket, a series of new diarylaminopyrimidine analogues was designed by incorporating a resorcinol moiety or 1,2,3-triazole moiety into the scaffolds 27 and 29 to interact with the ALK kinase domain where G1202R is located [35,36]. The resulting compounds (exemplified by 28 and 30) showed high activity against most ALK-resistant mutants, especially the challenging G1202R. In particular, compound 30 showed potent activity against a variety of frequently occurring crizotinib-resistant mutants, notably the L1196M mutant (IC$_{50} = 3.1$ nM) identified as the "gatekeeper" mutation and the G1202R mutant (IC$_{50} = 8.7$ nM), which confers resistance to all ALK clinical candidates [30].

Further, the overall activity of 4-aminoquinoline 31 was significantly improved by introducing a (piperidin-4-yl)-1H-pyrazol-4-yl moiety (affording compound 32) into the molecule to establish additional interactions with polar amino acid residues at the solvent front [37].

Crizotinib (33), an ALK receptor tyrosine kinase inhibitor approved by the U.S. FDA in 2011, is efficacious in ALK- and ROS-positive patients. The cocrystal structure of WT ALK and ceritinib suggested that the 4-(2-(isopropylsulfonyl)aryl)aminopyrimidine core occupies the kinase domain.
and forms a hydrogen bond with Met1199. The water-soluble "tail" piperidine motif exposed to the solvent has been investigated extensively in recent years, and modification of the "tail" is well tolerated, in terms of maintaining potency. The successful implement of structure-based and lipophilic-efficiency-guided drug design afforded aminopyridine 34, which displayed high potency against a variety of engineered ALK mutant cell lines and also potently inhibited proliferation of H3122-L1196M (a crizotinib-resistant cell line); it also has a favorable PK profile (86% bioavailability in the rat, about a 1 h half-life in rat) [38].

Figure 9. Optimization of peripheral substituents for the solvent-exposed regions of kinases, affording advanced inhibitors. The regions colored blue can considered as flexible sites that can tolerate a range of substituents, allowing us to generate new ligands having druggable properties, without loss of potency.

This article is protected by copyright. All rights reserved.
Further optimization of c-Met kinase inhibitors 35 and 37 was performed to improve protein-binder interactions at regions of the binding site near the solvent front, resulting in the identification of compounds 36 and 38, respectively [39,40].

Similarly, further optimization at position(s) within existing scaffolds (39, 41) that orient towards the solvent without specific hydrogen bond formation resulted in the discovery of selective PI3K inhibitors (40, 42) [41,42].

High-throughput screening of an in-house compound library for activity against TYK2 and JAK1-3 produced aminoindazole 43 as a hit molecule with a moderate inhibitory activity towards TYK2 (IC₅₀ = 650 nM), as well as a 10-fold JAK2 selectivity (Figure 9). Scaffold hopping of the aminoindazole core resulted in the identification of 3-amino-1,5-dihydro-4H-pyrazolopyridin-4-one 44 as a novel chemical scaffold for TYK2 inhibitors. Interestingly, initial SAR investigation showed that this platform could have a vertically flipped binding mode.

Introduction of a 1-methyl-3-pyrazolyl group at the 7-position of 45 directed toward the solvent-exposed region led to a significant improvement in TYK2-inhibitory potency, and further modification resulted in the identification of 46 [43].

Computational chemistry studies on the crystal structure of Aurora kinase A uncovered potential to improve the potency of ligands by reducing the steric clash.
with Leu139 and Val147 and by enhancing hydrophobic interaction with Gly216 and Leu139 residues in the solvent interface position of the protein. Based on these insights, rational redesign of furanopyrimidine 47 resulted in the discovery of quinazoline 48 [44,45].

SARs analysis suggested that a heterocyclic alkane at the solvent-exposed region in sunitinib (49) could substantially improve the inhibitory potency towards both FMS-like tyrosine kinase 3 (FLT3) and proliferation of FLT3-dependent human acute myelocytic leukemia (AML) cell line MV4-11. Indeed, 50 (FLT3, IC50 = 5.3 nM) is an efficacious and selective FLT3 inhibitor with favorable drug-like profiles (oral bioavailability, Cmax, Tmax, T1/2 and AUC in mice) [46].

The incorporation of various binary aromatic heterocycles at position 4 of pyrazole in the 1-H-pyrazole-3-carboxamide lead compound 51, resulted in the identification of compound 52. Then, different hydrophilic fragments were introduced in the solvent-accessible region of 52, leading to the discovery of compound 53 (FN-1501), which has potent activity against FLT3, CDK2, CDK4 and CDK6 and high anti-proliferative activity against MV4-11 cells (IC50 = 0.008 μM). 54 is safer acute-toxicity studies in mice (LD50 = 186 mg/kg). In addition, injections of 54 at 30 (mg/kg)/d significantly induced tumor regressions, and the percent-tumor-regression (PTR) rate on the 21st day was 78.95% [47].

This article is protected by copyright. All rights reserved.
Based on the binding model of 54 in the ATP-binding site of FGFR3, the indazole scaffold is located in the binding pocket and makes hydrogen bonds with Ala232 and Glu230; the vinyl pyrazole moiety extends toward the solvent, forming key hydrophobic interactions. Based on a binding model of compound 55 in the ATP-binding site of FGFR1 with a disclosed crystal structure of the FGFR1 kinase domain (PDB ID: 3WJ6), the benzimidazole fragment is considered to extend from the indazole 3-position towards the solvent-exposed region.

According to the fragment hybridization approach, the benzimidazole fragment of 55 was introduced into the indazole scaffold of 54, leading to the discovery of a potent pan-FGFR inhibitor 56 (FGFR1-4, IC\textsubscript{50} values ranging from 0.9 to 6.1 nM). Moreover, 56 indeed nearly complete inhibition of tumor growth in a xenografted mouse model [48].

Compound 57 potently inhibited RET kinase at both 100 and 20 μM. Molecular simulation suggested that heterocyclic substitution at the bromine would be solvent-exposed, and would be tolerated. From 57, the solvent region and the allosteric pocket were modified to furnish molecules 58 and 59, respectively. Thus, 58 and 59 were combined to afford compound 60. Further structural optimization (exemplified by 61 and 62) resulted in the advanced RET/VEGFR2 dual inhibitor Pz-1 (63), with isoxazole in the allosteric pocket and pyrazole in the solvent-accessible region [49].

This article is protected by copyright. All rights reserved.
Compound 64 was a previously reported tetrahydropyrido[4,3-d]pyrimidine-based Hsp90 inhibitor (in the fluorescence polarization assay, IC$_{50}$ = 0.101 μM). The molecular simulation demonstrated that 64 lay deep in the ATP-binding pocket of Hsp90 (Figure 10). The methyl group at the 2-position of the pyrimidine ring was exposed to the solvent, and substitution at this region should be tolerated. Thus, focusing on 64, the structure-activity (property) relationships of novel Hsp90 inhibitors were extensively investigated, resulting in the identification of compound 65 with improved Hsp90 inhibition (IC$_{50}$ = 0.028 μM) and antiproliferative activities (IC$_{50}$ = 0.14 μM (HCT116); IC$_{50}$ = 0.10 μM (MCF-7)) [50].

Starting from a previously reported series of benzoic esters 66, the chemical space in the solvent interface position of the PDE4-catalytic pocket was explored, focusing on new series of ortho-substituted derivatives (Figure 11). These SAR
efforts resulted in the identification of pyrrolidinyl and thiazolidinyl esters ((S*,S**)\textsuperscript{-67} and (S*,S**)\textsuperscript{-68}) as efficacious \textit{in vitro} PDE4 inhibitors \cite{51}. 

![Chemical structure](image)

**Figure 11.** Design concept of PDE4 inhibitors: optimization of peripheral substituents for the solvent-exposed region.

A structure-guided molecular hybridization approach to the modification of the structure of oral PDE4 inhibitor \textsuperscript{69} by introducing fragments from \textsuperscript{70} and \textsuperscript{71} led to the identification of dichloropyridyl-urea-pyridazinone \textsuperscript{72} and naphthyridin-5(6H)-one \textsuperscript{73}, respectively, with picomolar (pM) enzymatic potencies. The \textit{meta}-position of the pyridazinone 6-phenyl ring was oriented toward the solvent-filled pocket. Consequently, various functional groups with a range of physicochemical properties could be incorporated into the new series in the region occupying the solvent-filled pocket of PDE4. Compounds \textsuperscript{74} and \textsuperscript{75} were identified as potent, long-acting, and potentially safe inhaled PDE4 inhibitors. Compound \textsuperscript{76} bearing a 1,6-naphthyridin-5(6H)-one N-oxide moiety

This article is protected by copyright. All rights reserved.
was the most potent derivative of 73, with IC$_{50}$ of 45 pM in enzymatic assay and 75 pM in cellular assay [52].

### 2.2 To improve solubility by introducing solubilizing groups into a solvent-exposed region

Drug solubility is fundamentally dependent on the inherent physical properties of a compound. During lead optimization, polar and charged substituent groups (through salt formation or introduction of hydrophilic functional groups) are usually added to the open, solvent-exposed protein pocket of an inhibitor platform with aim of improving the solubility or PK properties of a drug candidate, without sacrificing the required potency [53]. For example, members of many NNRTI families contain a group that sits between Val106 and Pro236, and points toward the solvent-exposed region. This feature is very useful for modulation of the physicochemical properties of the NNRTIs.

#### Figure 12. Examples of NNRTIs modification by introducing solubilizing substituents in the tolerant region to improve solubility and PK profiles. These solubilizing groups are typically attached to a substitution point on the parent molecule that possesses a high degree of SAR tolerance and where modification usually neither enhances nor interferes with the target potency.

This article is protected by copyright. All rights reserved.
Recently, structure-based rational design has guided the installation of solubilizing substituents into the “Het-NH-Ph” and diaryltriazine (DATA) scaffolds (Figure 12). “Het-NH-Ph” NNRTIs 72, 73 showed substantially improved solubility compared with lead molecule 71 and the structurally related drugs ETR and RPV, while retaining low-nanomolar anti-HIV activity [54]. DATA derivatives 74 and 75 also showed low-nanomolar activity similar to that of ETR against WT HIV-1 and key resistant strains [55]. Subsequently, the determined co-crystal structures of 74 and 76 bound to RT revealed several salient features considered to be related to the substantial improvement of solubility (Figure 13) [55]. Specifically, the solubility enhancements appeared to result from strategic placement of the novel morpholinylalkoxy moiety at another RT/solvent interface (the entrance channel near Glu138) of the NNRTIs binding pocket [55].

**Figure 13.** Illustration of the X-ray crystal structures of 72 (pink, PDB ID: 4KKO) and 76 (yellow, PDB ID: 4O44) with HIV-1 RT; the morpholinoethoxy side chain in 72 projects towards Glu28 [54,55].
Other investigations have confirmed the utility of this methodology, which might be quite generally applicable [56]. For example, solubilizing moieties including morpholine were incorporated into the solvent-exposed region of the poorly soluble bromodomain inhibitor 77 to improve the polarity and charge. Notably, 78 and 79 displayed 200-500 fold improvement of solubility, though with about 3-5 fold reduced potency (Figure 14) [56]. Solubilizing groups have frequently been appended to kinase inhibitor drugs whose solubility is insufficient for further development. In general, such groups are located at substitution sites that have minimal effects on target binding affinity.

**Figure 14.** Discovery of bromodomain inhibitors 78 and 79 with improved aqueous solubility by targeting the solvent-exposed region.
**Figure 15.** Selected Lck kinase inhibitors containing solubilizing features (highlighted in blue).

Incorporation of a solubilizing tether that bears a weakly basic and polar tertiary amine group to the solvent-exposed region (the 6- and 7-positions of the fused phenyl ring) of the phenyl ring of Lck inhibitor 80 (<8 μM) afforded derivatives with improved solubility (>100 μM) and superior *in vitro* and cellular potencies, as demonstrated by compound 81 (Figure 15) [57].

**Figure 16.** Discovery of LpxC inhibitor 83 with improved aqueous solubility and drug-like profile by targeting the solvent-exposed region.

Compound 82 shows robust LpxC inhibition and antibacterial potency, but has poor solubility and undesirable *in vitro* cytotoxicity. Molecular simulation studies of the binding of hydroxamate 82 suggested that the 4-position of the benzene
moiety is oriented toward the large, mostly unoccupied solvent-exposed region. Therefore, this region was targeted for the introduction of water-solubilizing (or with lower hydrophobicity) moieties; this led to the identification of compound 83, which has improved water-solubility, reduced cytotoxicity to mammalian cell lines and plasma-protein binding (may cause changes of pharmacokinetic parameters), while retaining high \textit{in vitro} and \textit{in vivo} anti-pseudomonal potency (Figure 16) [58].

To sum up, we envision that further modifications of existing scaffolds with additional solubilizing moieties (such as oxetanyl sulfoxide), which extend appropriately into the protein-ligand interface or the solvent-facing region, will afford more potent molecules and may well be an effective way to develop back-up compounds with improved water-solubility, ADMET properties or favorable pharmacological profiles.

It should be noted that an underlying assumption of this methodology is that the required partial desolvation of the exposed polar moiety does not come with a substantial penalty in terms of binding affinity. Even though this design strategy seems to have been well accepted, a systematic study of the thermodynamics of solvation/desolvation of such exposed polar substituents at solvent-accessible regions is urgently needed [59]. It is rarely considered that solvent effects from water reorganization in the first hydration shell of protein–ligand complexes can have a significant impact on binding.

This article is protected by copyright. All rights reserved.
2.3 To improve drug-like properties (solubility irrelevant) by solvent-exposed region

**Figure 17.** Discovery of bioactive molecules with improved drug-like properties (solubility irrelevant) by solvent-exposed region.

This article is protected by copyright. All rights reserved.
Recently, by targeting solvent-exposed region, further optimization of PI3Kδ inhibitor 84 afforded compounds 85 and 86, with good lung retention and cell potency, which could be taken forward to in vivo studies where significant target engagement could be demonstrated (Figure 17A). This study suggested that addition of basic substituents to a region of 41 pointing to solvent was tolerated (enzyme inhibition pIC$_{50}$ > 9), with aim to carefully manipulate of the potency, pKa and lipophilicity [60].

Among the tetrazolyl HIV-1 NNRTIs, it was found that the addition of a substituent (i.e. an alkynyl fragment) at the para-position of the anilide in 87 (exemplified by 88a-d) greatly enhanced the metabolic stability (the half-lives) in rat liver microsomes (RLM) with retention of robust activity against the K103N/Y181C double mutant RT (Figure 17B) [61].

Using the previously reported aminopyrazole 89 as a lead compound inhibiting brain-penetrant aminopyrazole leucine-rich repeat kinase 2 (LRRK2), further structural optimization targeting the solvent-exposed region of the ATP-binding site resulted in the identification of highly potent, selective, and brain-penetrant inhibitors 90 and 91, with substantially improved stability, CYP inhibition, and induction liabilities, without compromising potency (Figure 17C) [62].
To improve the PK properties of the existing potent and selective Btk inhibitor CGI-1746 (92), structure-based design efforts were focused on both the solvent-exposed region (the morpholine–amide extends into the solvent) and the H3 binding pocket; this work led to the clinical candidate GDC-0834 (93), which kept the potency and selectivity of CGI-1746, but showed significantly improved preclinical PK profiles (Oral bioavailability (F) = 35%) (Figure 17D) [63].

The 7- and 8-positions on the benzoxazepinone core of receptor interacting protein 1 (RIP1) kinase inhibitor 94 are exposed to the solvent-accessible region. Consequently, optimization of the center platform and careful modification of neighboring substituents by utilizing structure-based design resulted in the identification of the cyano-substituted derivative 95 as a highly active and brain-penetrating RIP1 kinase inhibitor with favorable oral bioavailability and excellent ADMET properties ($t_{1/2} = 210$ min, in human and mouse liver microsomes; high plasma exposure ($AUC = 658 \text{ ng} \cdot \text{h/mL}$) and a moderate plasma duration ($MRT = 3.1$ h) in mice) (Figure 17E) [64].

This article is protected by copyright. All rights reserved.
2.4 To design prodrugs by targeting solvent-exposed region

Figure 18. The representative prodrugs (with significantly increased aqueous solubility) bearing caged moieties in solvent-exposed region.

Prodrugs have been explored to address the issues of solubility-limiting absorption and high excipient load. The surface-facing prodrugs will be metabolized prior to binding the intended target, even so, many prodrug groups have been found to be located at the solvent-accessible region, either intentionally or accidentally. This can not but be said to be an interesting phenomenon. For instance, an \(N\)-propionyl sulfonamide derivative GW695634 (97), designed as an amide prodrug of GW678248 (96), showed increased aqueous solubility and bioavailability in clinical trials (Figure 18) [65].

This article is protected by copyright. All rights reserved.
The N-2 position of pyridazinone 98, a potent NNRTI with limited solubility, was derivatized into a series of prodrugs, including hydroxymethyl esters and carbonates, as well as a phosphate. These prodrugs could effectively deliver 98 to rat plasma upon oral administration at 50 mg/kg. Improvements of 4.3- to 8.6-fold in the 24-hour exposure of 98 (over that of the prototype) were achieved, while the hydroxymethyl derivative 99a and the prodrugs disappeared completely after administration [66].

Adverse effects and poor water-solubility (0.08 mg/mL) limit the usage of the benzamide-based histone deacetylases (HDACs) inhibitor CI-994 (100). To overcome these limitations, two glucuronide prodrugs, compound 101 and carbamate prodrug 102, were designed, focusing on the solvent-facing region (Figure 18). The solubility of these prodrugs exceeded 1 mg/mL, and they showed improved stability at pH 2.1 and 7.0. When cleaved by glucuronidase, they showed the equivalent antiproliferative potency to CI-994 against non-small cell lung cancer cell line NCI-H661. In the absence of glucuronidase, the prodrugs showed decreased cytotoxicity, compared with the parent compound [67].

2.5 To favor a bio-active conformation by rigidification of small molecules

Complementarity between the target protein and bound ligand conformation is a prerequisite for high potency. In structure-based drug design, the initial goal is to design compounds that complement the binding pocket of a given target. However, proteins are quite flexible and undergo conformational changes during induced fitting with a flexible binder. Besides, a single compound can show more than one conformation, so, based on the available three-dimensional structures, a
strategy of constraining the conformation of flexible molecules (e.g., forming a
macrocycle between two solvent-exposed substituent positions of a given binder)
is well-established to improve potency, selectivity, and pharmacokinetics or
ADME-related properties (e.g. metabolic stability), while having a minimal effect
on the therapeutically relevant binding interaction [68,69].

Indeed, this approach has yielded new inhibitors with improved PK properties as
well as increased enzyme affinity, as exemplified by a wide range of macrocyclic
antiviral agents, such as hepatitis C virus (HCV) NS5B inhibitors (i.e. 103,
TMC647055) [70], HCV NS3/4A protease inhibitors (104, danoprevir; 105,
vaniprevir; 106, grazoprevir; 107, simeprevir; 108, paritaprevir; 109, voxilaprevir;
110, glecaprevir) [71-73], HIV-1 integrase inhibitors 111-114 [74], HIV-1
integrase-LEDGF/p75 interaction inhibitors 115-120 [75] and HIV-1 gp120
antagonist 122 (Figure 18) [76].

For example, taking peptidomimetic HIV-1 gp120 antagonist 121 as a lead
compound, cyclic peptide triazole (cPT) derivatives were obtained via cyclization
of two solvent-exposed substituent positions, thereby locking the molecule in an
active conformation. Especially, the six-residue cPT 122 (AAR029b) exhibited
submicromolar antiviral activities in inhibiting cell infection and triggering gp120
shedding, which causes irreversible virion inactivation. Importantly, cPTs
exhibited high stability to trypsin and chymotrypsin, in contrast with the
significant susceptibility of the corresponding linear PTs [76].

This structure-guided conformational restriction approach is also nicely illustrated
by the identification of cyclic peptidomimetic inhibitors of inducible nitric oxide

This article is protected by copyright. All rights reserved.
synthase [77], inhibitors of tricyclic indole diazepinone myeloid cell leukemia-1 [78] and the WD repeat domain 5 protein inhibitors (as exemplified by 124 and 125) (Figure 19E) [79]. Based on the linear peptidomimetic 123, structure-based design resulted in the identification of macrocyclic peptidomimetics 124 and 125 (MM-589), which could bind to WD repeat domain 5 (WDR5) and inhibit the WDR5-mixed lineage leukemia (MLL) protein-protein interaction. Especially, compound 125 potently binds to WDR5 (IC$_{50}$ = 0.90 nM) and inhibits the MLL H3K4 methyltransferase (HMT) activity (IC$_{50}$ = 2.7 nM, >45 times better than the linear peptidomimetic 1). Compound 125 could strongly and selectively inhibit the growth of human leukemia cells bearing MLL translocations (Figure 19) [79].

The significance of conformation as a determinant of passive permeability has also been illustrated in macrocycles (cyclic peptides), in which shielding of polar amides of backbone from the solvent can be achieved either by steric occlusion or through intramolecular hydrogen bonding [80]. However, there is little information to guide the design of synthetic macrocycles for favorable target-binding activity or bioavailability. To address this knowledge gap, Villar et al. proposed rules, “Villar's rule of macrocycles”, for designing macrocyclic agents with favorable oral bioavailability by controlling the physiochemical properties [81,82].

According to Villar's rules, a macrocycle should be orally bioavailable if it has 1) molecular weight of 600-1200 daltons, 2) clogP value of −2 to 6, 3) polar surface area of 180-320 Å$^2$, 4) number of H-bond acceptors in the range of 12-16, 5) number of H-bond donors < 12 and 6) number of rotatable bonds < 15 [81].

This article is protected by copyright. All rights reserved.
All in all, conformation restriction strategy via solvent-exposed region generally has several advantages: (1) Entropy loss of target protein-molecule interaction decreases and potency increases; (2) Additional interaction with target protein may occur due to the introduction of linker to enhance its potency; (3) After cyclization, polar groups may be concealed and the number of free rotating bonds decreases, resulting in increased permeability. Even so, ring closure presents a unique synthetic challenge in devising routes to access macrocycles [83].

**Figure 19.** Identification of macrocyclic compounds by rigidification of the bioactive conformation in the solvent-exposed region of the drug target. The section in “solvent-exposed region” was colored in blue.
2.5 To explore water-binding pockets (structural water molecules)

Water molecules often appear around binders in protein crystal structures. Careful investigation of water molecules and their energetics can significantly contribute to a successful drug design campaign [84]; there are many well-known examples, such as nonpeptidic urea inhibitors of HIV protease [85]. Structural biology and computational chemistry are useful to exploit water-binding pockets in structure-based design [86,87]. Displacing a water molecule that mediates protein-ligand binding by adding an appropriate moiety into the ligand to generate an isostere of the ligand-water complex can result in remarkable improvements in activity by establishing new hydrogen bond interactions directly between the protein and binder [88]. For example, the X-ray structure of influenza hemagglutinin (H3 - HK68: A/Hong Kong/1/1968, PDB ID: 5T6N) bound with the antiviral drug Arbidol (126) revealed the presence of a highly ordered water molecule adjacent to Arbidol, which was exploited for structure-based discovery of Arbidol derivatives (Figure 20). Introducing a meta-hydroxyl group into the thiophenol moiety of Arbidol to replace the structured water molecule in the binding pocket gave 127, which exhibited a remarkable improvement in potency against both H1 (98-fold) and H3 (1150-fold) hemagglutinin subtypes [89].

Recently, Horbert et al. employed WaterMap to find weakly bound water, and used the results to guide structural optimization (by using a phenolic hydroxy to...
replace a structural water molecule in the ATP-binding site) of
3,5-diaryl-pyrazin-2(1H)-ones (128, IC$_{50} = 0.5$ μM) as platelet-derived growth factor receptor (PDGF-R) β inhibitors, obtaining a 25-fold increase in inhibitory activity (129, IC$_{50} = 0.02$ μM) [90].

The treatment of structured water, especially prediction of the influence of water molecules on ligand binding, is still very challenging, partly because of difficulties in resolving water molecules by crystallography. In addition, it is difficult to predict which water molecules are obligate in the binding pocket and which can be displaced by a new binder [91,92]. Consequently, new computational methodologies are needed to predict the position and thermodynamic properties of water molecules (i.e., hydration sites) in the binding pocket [93]. At present, it is possible to identify areas of unfavorable bound waters based on Molecular Dynamics simulations and Grid Inhomogeneous Solvation Theory, which can be used improve the ligand binding by replacing so-called unhappy waters [94].

**Figure 20.** Structure-based optimization of (A) Arbidol (126) and (B) PDGF-R β inhibitor 128 based on utilization of the water-binding pocket. Details of WaterMap calculation is available free of charge via the Internet at https://pubs.acs.org/doi/10.1021/jm500373x.
2.6 To design dual-/multi-targeting ligands by exploiting solvent-exposed region

Multitarget-directed ligands (MTDLs), which are single biological agents designed to target multiple proteins relevant to a disease, are considered a promising approach to solve issues such as limited efficiency, poor safety and resistance profiles of individual targets [95]. Thus, MTDLs represent an efficient alternative strategy to the use of drug combinations. However, designing MTDLs with predefined pharmaceutical profiles is still a substantial challenge [96]. Nevertheless, the availability of detailed structural information of target proteins and the latest findings in bioinformatics have yielded a way to discover multi-targeting agents.

Key to the rational design of pharmacophore-fused MTDLs is to identify a tolerant region in the drug target. For example, structural biology studies suggested that the phenyl moiety in the N-1 substituent of the 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine (HEPT)-type HIV-1 NNRTIs (i.e., 130) is situated in the solvent-exposed region controlled by the Pro236 loop, where structural modifications are tolerated [97]. Based on these analyses, and in order to address the urgent need for new agents to reduce the global occurrence and spread of AIDS, several series of RT/IN (integrase) dual inhibitors were designed by incorporation of an IN pharmacophore element (i.e. 131) at this tolerant region of a previously reported potent NNRTI (Figure 21). Many of these inhibitors show potency against RT and IN at micromolar level, also inhibit HIV at nanomolar level (compounds 132 and 133). These results also

This article is protected by copyright. All rights reserved.
confirm that the installation of a second pharmacophore in these NNRTIs does not seriously influence their binding conformation with RT [98].

**Figure 21.** Discovery of RT/IN dual inhibitors, 132 and 133, by combining RT inhibitor 130 with IN inhibitor 131.

Furthermore, the MTDLs strategy has been nicely exemplified by the identification of a wide range of bifunctional HDAC inhibitors through combining the pharmacophoric features of HDAC inhibitors (i.e. 134, SAHA) and other anticancer agents (phosphodiesterases, poly(ADP-ribose)polymerase, protein kinase, CDK2, etc.) into one molecule, to take advantage of the presence of large hydrophobic patches at the HDAC surface rim [99]. Some encouraging results have been reported (Figure 22), such as dual-acting HDAC and topoisomerase II inhibitor 135 [100], triple HDAC and topoisomerase I/II inhibitor 136 [101], photoactivatable platinum(IV) complex cis,trans-[Pt(N3)2(Sub)2(tBu2bpy)] (137) targeting genomic DNA and HDAC [102], dual-acting estrogen receptor and HDAC inhibitor 138 [103], tamoxifen-HDACi conjugate 139 and ethynyl-estradiol-HDACi conjugate 140 [104], chimeric
3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR)-HDAC inhibitor [105], dual-acting androgen receptor (AR) and HDAC inhibitor [106], stilbene-SAHA chimeric molecule [107], nitric oxide (NO)-donor HDAC inhibitor [108], chimeric c-Src kinase and HDAC inhibitor [109], and dual-acting phosphodiesterase 5 (PDE5) and HDACs inhibitors [110], JAK2-HDAC dual inhibitor [111], dual-acting phosphodiesterase 5 (PDE5) and HDACs inhibitors [112], JAK1-HDAC dual inhibitor [113], LSD1-HDAC dual inhibitor [114], IDO1-HDAC dual inhibitor [115] and NAMPT-HDAC dual inhibitor [116], mTOR-HDAC dual inhibitor [117]. The design rationale for MTDLs is underpinned by deep insights derived from structural bioinformatics and structural biology.

**Figure 22.** Multitarget-directed HDACIs. The incorporated HDAC inhibitor features are colored red.

This article is protected by copyright. All rights reserved.
Based on structural data from the analysis of the co-crystal structure of ligands with Bcl-xL or Mcl-1, solvent regions were identified. Compounds 157 and 158 were found to be potent inhibitors of Mcl-1 and Bcl-xL with IC$_{50}$ values of 0.54 and 0.15 μM, respectively. The carboxyl moiety of 157 and the benzoyl group of 158 extend into the solvent region, where sufficient space exists for further structural decoration. The large space in the solvent-interaction region of either Mcl-1 or Bcl-xL might be used to introduce another target-binding moiety. Based on these analyses, a structure-based hybridization strategy was employed to discover novel Mcl-1/Bcl-xL dual inhibitors (159, 160) with a suitable linker based on 157 and 158 (Figure 22). Notably, compound 160 displayed the potent dual-inhibitory activities (Mcl-1, IC$_{50}$ = 0.088 μM; and Bcl-xL, IC$_{50}$ = 0.0037 μM) (Figure 23) [118].

This hybridization approach is also nicely illustrated by the identification of bifunctional ALK-Hsp90 inhibitors 163 and 164 (derived from 161 and 162)[119] and MEK-PI3K inhibitor 167 (ST-168, derived from 165 and 166) [120].
Undoubtedly, the increasing efforts in recent years to discover MTDLs through rational combination of pharmacophoric motifs of different parent molecules will lead to new advances in drug discovery.

2.7 To design bisubstrate (heterobivalent) ligands by exploiting solvent-exposed region

A relatively new strategy in drug discovery is the development of bisubstrate ligands. Bisubstrate (bivalent) binders comprise two conjugated components, each targeted to a binding pocket in a multi-substrate protein or heterodimer. The significant advantage of bisubstrate inhibitors is their ability to simultaneously bind more than one orthosteric binding sites and to form more interactions with the target, thus providing markedly enhanced binding affinity and (subtype) selectivity, when compared with single-site binders [121].

This article is protected by copyright. All rights reserved.
In 2015, a class of classical heterobivalent binders and a class of more integrated dual acting ligands were reported, incorporating ZM 241385 (168) as an adenosine A2A receptor antagonist and ropinirole (169) as a dopamine D2 receptor agonist, via the solvent-exposed region. The best molecules (exemplified by 170 and 171) maintained the potency of the parent pharmacophores at both targets (adenosine A2A and dopamine D2). Additionally, the integrated dual acting molecules (exemplified by 172) also demonstrated promising efficacies in blood-brain barrier permeability assay, whereas the heterobivalent binders 170 and 171 are potentially more suited as molecular probes (Figure 24) [122].

Figure 24. Chemical structures of the A2AR antagonist ZM 241385 (168) and D2R agonist ropinirole (169) and the designed heterobivalent binders and integrated dual acting ligands.

Bisubstrate (heterobivalent) ligands are an attractive option for the discovery of novel bifunctional anti-HIV agents. For example, the proximity (10-15 Å) of the nucleoside HIV-1 reverse transcriptase inhibitor (NRTI) and NNRTI binding sites

This article is protected by copyright. All rights reserved.
in RT and the crosstalk between these sites provide the opportunity to create a single “bifunctional” (chimeric) molecule that targets both sites simultaneously (Figure 25). This study illustrates the identification of a novel dual RT inhibitor utilizing d4T (NRTI) and a TMC derivative (173, a diarylpyrimidine-based NNRTI) connected via a poly(ethylene glycol) (PEG) linker in the solvent-exposed region. The triphosphate moiety of the d4T-4PEG-TMC bifunctional binder was successfully incorporated into RT in a base-specific manner. Furthermore, 174 shows low nanomolar activity, affording 4.3-fold and 4300-fold improvement of polymerization inhibition in vitro compared with the parent TMC derivative and d4T, respectively [123].

The bifunctional chimeric HIV inhibitor 177 was obtained by covalently combining a small-molecular CCR5 antagonist, LGC240 (176), and a gp120 peptide triazole antagonist, UM15 (175) (Figure 25). Interestingly, 177 displayed the specific targeting profiles of the two separate fragments, and meanwhile showed low nanomolar to sub-nanomolar activities in inhibiting replication of different pseudoviruses in cell culture, being significantly more potent than a non-covalent mixture of the individual fragments. These results suggest that creation of bifunctional chimeras via the solvent-exposed region that coordinately target HIV-1 gp120 and host-cell co-receptor proteins at the virus-cell interface can result in increased potency [124]. These studies reconfirm the power of the
rational design of heterobivalent binders as a general approach to achieve high affinity and potency.

**Figure 25.** Schematic representation of heterobivalent HIV-1 RT ligand 174 consisting of NRTI and NNRTI and bifunctional chimera 177 that coordinately targets HIV-1 envelope gp120 and host-cell CCR5 co-receptor.

### 2.8 To design multivalent ligands by exploiting solvent-exposed region

The structure-based design of multivalent ligands offers an attractive strategy for inhibiting of protein recognition and other biological functions mediated by proteins with multiple binding sites [125]. Representative examples include the design of multivalent binders targeting subtypes of the AB(5) bacterial toxin family [126] and influenza neuraminidases (NAs) [127-129].

In this section, we will focus on recent structure-based design of multivalent influenza neuraminidase inhibitors. Zanamivir (178) and oseltamivir are key antiviral drugs targeting influenza NA, but drug-resistant viruses bearing mutant NAs with impaired binding affinity are increasingly emerging. Structural analysis indicates that the C-7 hydroxyl group is exposed to the solvent, suggesting it would be an appropriate location to link the monomers, since derivatization of zanamivir at this position should not substantially reduce its affinity for the active site of NA. The spatial arrangement of the ligand-binding sites of NA provides the possibility to design branched multivalent binders with 2- or 4-fold symmetry.

This article is protected by copyright. All rights reserved.
Indeed, structural knowledge of both drug-binding sites and their spatial arrangement on the homotetrameric NA has been utilized to design several dimeric (179-181) or tetravalent (182,183) zanamivir molecules (Figure 26), which displayed significantly increased NA binding affinity, NA-inhibitory potency, and antiviral efficacy, compared with zanamivir [127-129]. The design methodology in those studies may also be useful to develop other compounds that target oligomeric proteins with multiple binding sites.

Figure 26. Rational design of dimeric or tetravalent zanamivir conjugates.

3. Exploitation of solvent-exposed regions for rational design of protein degradation and drug delivery systems

3.1 To design proteolysis targeting chimeras (PROTACs) by exploiting solvent-exposed region

This article is protected by copyright. All rights reserved.
Targeted protein degradation, using heterobifunctional small molecules (PROTACs) to remove protein targets from within cells, has emerged as a novel strategy for drug development, with the opportunity of providing therapeutic interventions not achievable with existing occupancy-based enzyme inhibition approaches [130-132]. Small-molecular-weight synthetic PROTACs (185-189) have been used to selectively degrade various specific proteins (Figure 26), including pirin [133], sirt2 [134], BET protein [135], androgen receptor [136], and BRD4 protein [137].

**Figure 27.** Representative examples of proteolysis targeting chimeras.

PROTACs consist of a targeting ligand (warhead) for the specific protein to be degraded, an E3 ubiquitin ligase recruitment binder, and a suitable linker connecting the two binders (Figure 27). A key step is to find a solvent-facing solubilizing group vector in the warhead (as exemplified by the pyrrolidine moiety.

This article is protected by copyright. All rights reserved.
in 184), at which to attach the linker, by analysis of the X-ray crystal structure of the warhead bound to the target protein. The incorporated linker should not significantly decrease the inhibitory potency. Recent studies exemplify how more potent warheads do not necessarily result in more potent PROTACs, and underscore how the ability to form a strongly bound ternary complex is of the utmost importance to the mechanism of action of PROTACs [138]. Therefore, it will be critical in future design of bivalent degraders to take full account of the kinetic and thermodynamic properties of the whole ternary system [139].

3.2 To design antibody-recruiting small molecules by exploiting solvent-exposed region

Antibody-recruiting molecules (ARMs), which are synthetic systems capable of modulating immunological functions, are an emerging field of research in synthetic immunology. Small-molecular ARMs can enhance antibody binding to disease-relevant viruses or cells, resulting in their immune-mediated clearance [140].
In 2009, compound 190 (derived from 191) was disclosed as an ARM that targets HIV gp120. This compound inhibits virus replication via two mutually reinforcing mechanisms, namely, inhibition of virus entry and antibody-mediated killing [141]. In 2014, the same lab reported the structure-guided optimization of ARMs targeting HIV gp120, which generated the advanced analogue 192 (derived from 193). This is ~1000-fold more active in gp120-binding and MT-2-based antiviral assays than 190 (Figure 28) [142].

Prostate-specific membrane antigen (PSMA) is a membrane-bound glutamate carboxypeptidase expressed at high levels in many forms of prostate cancer and is thought to be an important target for the drug design of prostate carcinoma. A series of ARMs that enhance antibody-mediated immune recognition of prostate cancer cells was recently disclosed [143,144]. Notably, compounds 194 and 195

**Figure 28.** Antibody-recruiting small molecules 190 and 192 that target HIV gp120, and 191 and 193 that target prostate-specific membrane antigen.
exhibited high affinity for PSMA and can bind to anti-DNP antibodies. Again, the identification of the solvent-facing region in the warheads (colored in red) bound to the target protein proved the key for structure-guided design of ARMs.

3.3 To design other new drug delivery systems by exploiting solvent-exposed region

Most small-molecular drugs are short-lived species in the circulatory system, being rapidly eliminated by glomerular filtration in the kidney. Human serum albumin (HSA) is an excellent carrier for drug delivery and for improving the PK properties (e.g., extending the half-life) of drugs. Noncovalent binding to HSA can protect small-molecular drugs against enzymatic degradation and renal clearance \textit{in vivo} [145].

Several HSA-derived drug delivery technologies have been developed. Especially \textit{in vivo} non-covalent endogenous HSA targeting has received widespread attention. For example, the clinically relevant glucagon-like peptide 1 (GLP-1) was linked with diflunisal (196, albumin binder) and indomethacin (197, albumin binder) to generate a divalent GLP-1 derivative 198 with a big boost in circulatory half-life and absorption time relative to its monovalent equivalent in PK studies (Figure 29) [146]. Again, the key was finding a suitable tethering site for the divalent HAS-conjugate. Similar insights have proved useful in the
structure-based exploitation of other drug delivery systems, including human serum amyloid P component, polyamines, and antibody-drug conjugates (including ARMs) [147-149].

![Diagram of indomethacin and diflunisal binding to HSA](image)

**Figure 29.** Typical examples of in vivo non-covalent endogenous HSA targeting approach.

4. Exploitation of solvent-exposed regions in other fields

4.1 To design affinity-based small-molecular probes by exploiting solvent-exposed region

Activity- and affinity-based probes have tremendous potential in chemical biology and drug discovery, as they provide useful chemoproteomic tools in target validation of low-molecular-weight drug candidates [150,151]. Usually, a functional affinity-based small-molecular probe (exemplified by WDR5 probe 200 or Keap1 probe 202) is composed of a potent low-molecular-weight binder as the recognition group (warhead), a linker and a reporter (fluorophores group or...
biotin). In general, the solvent-facing moiety of a small-molecule inhibitor
(4-position substituent group of aromatic ring A in 199, the methoxy groups in
201) is available as an attachment point for the fluorophore or other reporter
without compromising the desired binding affinity (Figure 30) [152,153]. The
disclosed X-ray crystal structure of protein-ligand complexes provides a
reasonable basis for the discovery of affinity-based small-molecular probes.

Figure 30. Structure-based design of affinity-based small-molecular probes.

4.2 To construct DNA-encoded chemical libraries (DELs) by exploiting
solvent-exposed region

DNA-encoded chemical libraries (DELs) are collections of molecules,
individually linked to conjugated DNA sequence using as amplifiable
identification barcodes. The huge combinatorial libraries were usually prepared

This article is protected by copyright. All rights reserved.
from split-and-pool synthetic procedures from individual molecules with the related DNA tag. Then the combinatorial libraries can be stored as a mixture, allowing bioactivity screening. In recent years, DELs have attracted considerable attention as an effective approach for the discovery of new chemical equity for drug targets [154]. With the increased application of DEL in drug discovery, many new compounds are emerging in the medicinal chemistry literature, as exemplified by receptor interacting protein 1 kinase inhibitor 203 and tankyrase 1 inhibitor 204 with drug-like properties (Figure 31) [155,156]. A key factor for the success of the DNA-encoded chemistry is the chemical diversity of the libraries, which enables the deep exploration of chemical spaces, being significantly superior to traditional high-throughput screening (HTS) approaches [157]. Another key factor for the success of the technology is that the conjugated DNA tags should be attached to a solvent-exposed region of the core scaffold (The positions that the DNA tags were attached was shown in Figure 30).

Figure 31. Discovery of bioactive compounds from DNA-encoded chemical library.
5. Discussion and prospects

**Figure 32.** Schematic diagram: opportunities for utilizing solvent-exposed regions in drug design.

Drug discovery is a complex, costly and time-consuming process that involves multiple steps, ranging from target validation to clinical trials and regulatory approval. In particular, the search for drugs with novel action mechanisms has become extremely expensive and risky. On the other hand, follow-on programs to optimize or extend the scope of lead compounds with proven mechanisms are less risky and costly, but can nevertheless produce drugs with meaningful differentiations from the lead.

The increasing availability of high-resolution X-ray co-crystallography of proteins with ligands, molecular simulation technology and abundant SAR conclusions have greatly facilitated the identification of tolerated solvent-exposed regions in the specific drug target. Consequently, in follow-on-based drug discovery...
campaign, such solvent-exposed regions provide a broad space for substantial modifications of existing compounds, as well as providing a basis for structure-based rational design of protein degradation and drug delivery systems, with the aim of enhancing binding affinity, improving selectivity, overcoming drug resistance, and improving solubility, tissues retention or pharmacokinetic properties (Figure 32). The possibility of discovering novel chemical entities that lie in unexplored regions of the crowded intellectual property landscape makes such follow-on programs attractive.

In structure optimization, even very minor structural changes of a ligand’s kinetics, thermodynamic binding, efficacy and drug-like profile can have a remarkable impact. Therefore, a key factor for the success of exploitation of solvent-exposed regions is the chemical diversity of the compound libraries. Deeper computational simulation of molecular systems to establish the structure-energy-function relationships of ligand-protein binding processes will also be important.

One of the main issues in computational drug design targeting solvent-exposed regions is the inherent flexibility of the tolerated region. Conformational flexibility is of the utmost importance in several important molecular recognition processes, including ligand-binding events, in which induced fit can result in significant structural rearrangement of the protein. Moreover, water has a
profound effect on the dynamics of biomolecules and influences many biological processes. The mechanism (or mechanisms) of enthalpy-entropy compensation in water-involved protein-ligand binding remains controversial [158]. Therefore, it is very challenging to manage receptor flexibility and solvent effects, especially to accurately predict binding free energies of protein-ligand complexes in various solvation environments with current computational methods.

Solvent-exposed region-guided modification, namely, combining the frameworks from two ligands, can results in large, complex ligands, including dual-/multi-targeting ligands, bisubstrate (heterobivalent) ligands, multivalent ligands, PROTACs, and antibody-recruiting molecules. The physicochemical properties of these larger molecules are generally less drug-like than those of smaller compounds. However, if the degree of framework overlap is maximized and the size of the combined ligands is minimized, drug-like ligands can be obtained. Focused or diversity-based screening may afford a route to smaller and more compact molecules [159].

Privileged structure is a very important concept in medicinal chemistry [160]. Interestingly, some privileged (solvent-friendly) moieties, including sulfonyl, sulfonamide, morpholine, piperidine and piperazine, have been introduced into solvent-exposed regions to improve aqueous solubility, cellular permeability, and metabolic stability [161-165]. In addition, clusters are novel and potentially useful
privileged moieties for the inhibition of certain enzymes, since they can fill particularly large, flexible, or open active sites in unique ways [166,167].

As described above, structurally conserved bound water molecules can also play a critical role. In computer-assisted rational drug design, particular emphasis should be placed on so-called unconventional interactions between water and heterocyclic nitrogens, as exemplified by lone-pair-π (lp-π) interactions [168,169].

It should be pointed out that the water is considered poorly understood [170], consequently, there are still some challenges in the rational treatment of structured water molecules in drug design.

Finally, it is our hope that the strategic advances outlined in this article will inspire further exploitation of solvent-exposed regions via an integrative multi-disciplinary approach (machine-learning approach) in the future development of drug candidates [171].

AUTHOR INFORMATION

Corresponding Authors

*D.K.: e-mail, kangdongwei@126.com; phone, 086-531-88382005;

*V.P.: e-mail, nathan@sdu.dk; 0046722598320;

*X.L.: e-mail, xinyongl@sdu.edu.cn; phone, 086-531-88380270;

This article is protected by copyright. All rights reserved.
Author Manuscript

*P.Z.: e-mail, zhanpeng1982@sdu.edu.cn; phone, 086-531-88382005

ORCID

Vasanathanathan Poongavanam: 0000-0002-8880-9247

Peng Zhan: 0000-0002-9675-6026

Author Contributions

All authors contributed to writing the manuscript. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Biographies

Xiangyi Jiang completed his B.S. degree at Shandong University in 2018 and then started work with Professor Xinyong Liu as a graduate student in the School of Pharmaceutical Science, Shandong University. His current research focuses on the discovery of novel HIV inhibitors.

Ji Yu completed his B.S. degree at Shandong University in 2016 and then started work with Associate Professor Peng Zhan and Professor Xinyong Liu as a graduate student in the School of Pharmaceutical Science, Shandong University. His current research focuses on the discovery of novel non-nucleoside HBV inhibitors.

This article is protected by copyright. All rights reserved.
Zhongxia Zhou was born in 1991 in Linyi, Shandong province, China. In 2014, she graduated from Shandong University and obtained her BS degree. She is currently studying for her doctor’s degree in the Department of Medicinal Chemistry of the School of Pharmaceutical Sciences in Shandong University. So far, she has published eight SCI academic peer-reviewed papers.

Jacob Kongsted completed his Ph.D. in Theoretical and Computational Chemistry at the University of Copenhagen in 2005, followed by post-doctoral training at Lunds University (Sweden) and University of Aarhus (Denmark). In 2009 he moved to University of Southern Denmark where he is currently full Professor of computational quantum chemistry. Kongsted has especially contributed to this field by development of novel quantum chemistry embedding methods as well as application of these computational models to systems of biophysical and biochemical interest. Recent research use quantum chemistry methods for rational design of light-induced biological functional materials and advanced drug design.

Yu’ning Song was born in 1984 in Jinan, Shandong province, China. In 2007, she graduated from China Pharmaceutical University. Then she earned her M.S. and Ph.D. in pharmacology from Shandong University in 2008 and 2010, respectively. She is currently working as a licensed pharmacist in the Department of Clinical Pharmacy, Qilu Hospital of Shandong University.

Christophe Pannecouque graduated in pharmaceutical sciences at the Rega Institute for Medical Research of the Katholieke Universiteit Leuven in 1990. He obtained his PhD in medicinal chemistry at the same university in 1995, and

This article is protected by copyright. All rights reserved.
joined the group of Professor Erik De Clercq as a postdoctoral fellow. In 2003, he became an associate professor. In the field of virological research, he has unravelled the modes of action of several classes of new products with anti-HIV activity. Recently, his research has been focused on cellular targets interfering with HIV replication and on regulation of the (cyto)pathogenicity of the virus.

**Erik De Clercq** has M.D. and Ph.D. degrees and has taught at the Katholieke Universiteit Leuven (and Kortrijk) Medical School, Belgium, where he was Chairman of the Department of Microbiology and Immunology until September 2006. He is currently Emeritus Professor of K.U. Leuven University, Member of the Belgian (Flemish) Royal Academy of Medicine and the Academia Europaea, and Fellow of the American Association for the Advancement of Science. In 2008, he was elected European Inventor of the Year (Lifetime Achievement Award), and in 2010 he, together with Dr. A. S. Fauci, was Laureate of the Dr. Paul Janssen Award for Biomedical Research. He is the (co)inventor of a number of antiviral drugs (valaciclovir, brivudin, cidofovir, adefovir, and tenofovir).

**Dongwei Kang** was born in 1990 in Liaocheng, Shandong province, China. In 2012, he graduated from the School of Hebei University of Technology with his BS degree. He earned his M.S. degree and Ph.D. in medicinal chemistry from Shandong University in 2015 and 2018, respectively. He is currently working as a postdoctoral researcher in the Department of Medicinal Chemistry of the School of Pharmaceutical Sciences at Shandong University.

**Vasanthanathan Poongavanam** obtained his PhD in Medicinal Chemistry from the University of Copenhagen in 2010, and is currently employed as a
postdoctoral researcher at Uppsala University, Sweden. Prior to that, he was a postdoctoral researcher at Southern Denmark University, Denmark (2013-2015), Denmark and the University of Vienna, Austria (2011-2012). His research activities involve cheminformatics, drug design and development, biomolecular simulation and computational medicinal chemistry.

**Xinyong Liu** received his B.S. and M.S. degrees from the School of Pharmaceutical Sciences, Shandong University, in 1984 and in 1991, respectively. From 1997 to 1999 he worked at Instituto de Quimica Medica (CSIC) in Spain as a senior visiting scholar. He obtained his Ph.D. from Shandong University in 2004. He is currently a full Professor of the Institute of Medicinal Chemistry, Shandong University. His research interests include rational drug design, synthesis and antiviral evaluation of a variety of molecules that interact with specific enzymes and receptors in the viral life cycle (HIV, HBV, HCV, and FluV). Other ongoing programs include studies of the molecular modification and structure-activity relationships of some natural products used to treat cardio/vascular diseases, and drug delivery research using PEGylated small-molecular agents.

**Peng Zhan** obtained his B.S. degree from Shandong University, China, in 2005. Then he earned his M.S. degree and Ph.D in medicinal chemistry from Shandong University in 2008 and 2010, respectively. He subsequently joined the research group of Professor Xinyong Liu as a Lecturer (2010-2012). From 2012 to 2014, he worked as a Postdoctoral Fellow funded by JSPS (Japan Society for the Promotion of Science) in the Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Japan. He is currently an associate professor in the
Institute of Medicinal Chemistry, Shandong University. His research interests involve the discovery of novel antiviral, anticancer and neurodegenerative diseases-related agents based on rational drug design and combinatorial chemistry approaches.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from the National Natural Science Foundation of China (NSFC Nos. 81573347, 81603028), Young Scholars Program of Shandong University (YSPSDU No. 2016WLJH32), the Fundamental Research Funds of Shandong University (No. 2017JC006), Key Project of NSFC for International Cooperation (No. 81420108027), the Natural Science Foundation of Shandong Province (No. ZR2016HB26), and Key research and development project of Shandong Province (No. 2017CXGC1401). All figures showing binding modes were generated using PyMol (www.pymol.org).

ABBREVIATIONS USED

ALK, anaplastic lymphoma kinase; AR, androgen receptor; ARMs, antibody-recruiting molecules; cPTs, cyclic peptide triazoles; DEL, DNA-encoded chemical library; FLT3, FMS-like tyrosine kinase 3; GLP-1, glucagon-like peptide 1; HDACs, histone deacetylases; HEPT, 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine; HPD, 3-hydroxypyrimidine-2,4-dione; HMGR, 3-hydroxy-3-methylglutaryl coenzyme A reductase; HAS, human serum albumin; IN, integrase; LRRK2, leucine-rich repeat kinase 2; MTDLs, multitarget-directed ligands; NA, neuraminidase; NRTI, nucleoside reverse transcriptase inhibitor; NNRTIs, nonnucleoside reverse transcriptase inhibitors; NO, nitric oxide; PDE5, phosphodiesterase 5; PDGF-R, platelet-derived growth factor receptor; PK, pharmacokinetic; PROTACs,
proteolysis targeting chimeras; PSMA, prostate-specific membrane antigen; RT, reverse transcriptase; RNase H, ribonuclease H; SAR, structure-activity relationship

REFERENCES


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Transcriptase Inhibitors with Improved Potency against Resistance-Associates


19. Wang, X.; Gao, P.; Menéndez-Arias, L.; Liu, X.; Zhan, P.


Double-Winged 3-Hydroxypyrimidine-2,4-diones: Potent and Selective


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


40. Ma, Y.; Sun, G.; Chen, D.; Peng, X.; Chen, Y. L.; Su, Y.; Ji, Y.; Liang, J.; Wang, X.; Chen, L.; Ding, J.; Xiong, B.; Ai, J.; Geng, M.; Shen, J. Design and This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


67. Thomas, M.; Clarhaut, J.; Tranoy-Opalinski, I.; Gesson, J. P.; Roche, J.; Papot, S. Synthesis and biological evaluation of glucuronide prodrugs of the

This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


86. Seo, J.; Igarashi, J.; Li, H.; Martasek, P.; Roman, L. J.; Poulos, T. L.; Silverman, R. B. Structure-based design and synthesis of


This article is protected by copyright. All rights reserved.
93. Cuzzolin, A.; Deganutti, G.; Salmaso, V.; Sturlese, M.; Moro, S. 
AquaMMapS: An Alternative Tool to Monitor the Role of Water Molecules 

94. Schauperl, M.; Podewitz, M.; Waldner, B.J.; Liedl, K.R. Enthalpic and 
4600-4610.

95. Morphy, R.; Rankovic, Z. Designed multiple ligands. An emerging drug 

96. (a) Zhan, P.; Liu, X. Designed multiple ligands: an emerging anti-HIV drug 
discovery paradigm. *Curr. Pharm. Des.* **2009**, *15*, 1893-1917. (b) Zhan, P.; Liu, 
1743-1758.

97. Chen, W.; Zhan, P.; Wu, J.; Li, Z.; Liu, X. The development of HEPT-type 
HIV non-nucleoside reverse transcriptase inhibitors and its implications for 

98. (a) Wang, Z.; Bennett, E. M.; Wilson, D. J.; Salomon, C.; Vince, R. Rationally 
**2007**, *50*, 3416-3419. (b) Wang, Z.; Vince, R. Synthesis of pyrimidine and 
quinolone conjugates as a scaffold for dual inhibitors of HIV reverse transcriptase 
Vince, R. Design and synthesis of dual inhibitors of HIV reverse transcriptase and 

This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


112. Huang, Y.; Dong, G.; Li, H.; Liu, N.; Zhang, W.; Sheng, C. Discovery of Janus Kinase 2 (JAK2) and Histone Deacetylase (HDAC) Dual Inhibitors as a


117. Chen, Y.; Yuan, X.; Zhang, W.; Tang, M.; Zheng, L.; Wang, F.; Yan, W.; Yang, S.; Wei, Y.; He, J.; Chen, L. Discovery of a Novel Dual Histone Deacetylases (HDACs) and Mammalian Target of Rapamycin (mTOR) Target

This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


127. Macdonald, S. J.; Cameron, R.; Demaine, D. A.; Fenton, R. J.; Foster, G.;
B.; Jones, H. T.; McConnell, D. B.; McKimm-Breschkin, J.; Mills, G.; Nguyen,
V.; Owens, I. J.; Parry, N.; Shanahan, S. E.; Smith, D.; Watson, K. G.; Wu, W. Y.;
Tucker, S. P. Dimeric zanamivir conjugates with various linking groups are
potent, long-lasting inhibitors of influenza neuraminidase including H5N1 avian

Wong, C. H. Synergistic effect of zanamivir-porphyrin conjugates on inhibition of
neuraminidase and inactivation of influenza virus. J. Med. Chem. 2009, 52,
4903-4910.

Yan, J.; Gao, GF.; Li, X. Structure-Based Tetravalent Zanamivir with Potent
Inhibitory Activity against Drug-Resistant Influenza Viruses. J. Med. Chem. 2016,
59, 6303-6312.

130. (a) Crews, C. M. Inducing Protein Degradation as a Therapeutic Strategy. J.
Med. Chem. 2018, 61, 403-404. (b) Cromm, P. M.; Crews, C. M. Targeted Protein
24, 1181-1190. (c) Bondeson, D. P.; Smith, B. E.; Burslem, G. M.; Buhimschi, A.
Crews, C. M. Lessons in PROTAC Design from Selective Degradation with a
Smith, B. E.; Lai, A. C.; Jaime-Figueroa, S.; McQuaid, D. C.; Bondeson, D. P.;

This article is protected by copyright. All rights reserved.
Toure, M.; Dong, H.; Qian, Y.; Wang, J.; Crew, A. P.; Hines, J.; Crews, C. M.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.
Guo, X. K.; You, Q. D. Design, synthesis, and initial evaluation of affinity-based

Evaluation of Affinity-Based Small-Molecule Probes for Fluorescent
Visualization and Specific Detection of Keap1. *J. Med. Chem.* 2016, 59,
7305-7310.


155. Harris, P. A.; King, B. W.; Bandyopadhyay, D.; Berger, S. B.; Campobasso,
N.; Capriotti, C. A.; Cox, J. A.; Dare, L.; Dong, X.; Finger, J. N.; Grady, L. C.;
Hoffman, S. J.; Jeong, J. U.; Kang, J.; Kasparcova, V.; Lakdawala, A. S.; Lehr,
Schaeffer, M. C.; Summerfield, J. D.; Swift, B. A.; Totoritis, R. D.; Ward, P.;
Library Screening Identifies Benzo[b][1,4]oxazepin-4-ones as Highly Potent and
Monoselective Receptor Interacting Protein 1 Kinase Inhibitors. *J. Med. Chem.*
2016, 59, 2163-2178.

Schüler, H.; Neri, D.; Franzini, R. M. Tankyrase 1 Inhibitors with Drug-like
Properties Identified by Screening a DNA-Encoded Chemical Library. *J. Med.
Chem.* 2015, 58, 5143-5149.

This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Simulations of β-Galactosidase Suggest that Both Modes Can Co-exist.

*Chemistry.* 2018, 24, 5849-5859.
