HHS Public Access

Author manuscript

ChemMedChem. Author manuscript; available in PMC 2020 April 22.

Published in final edited form as:

ChemMedChem. 2019 January 08; 14(1): 119–131. doi:10.1002/cmdc.201800537.

Chemical Space Overlap with Critical Protein–Protein Interface Residues in Commercial and Specialized Small-Molecule Libraries

Yubing Si[a], **David Xu**[b],[c], **Khuchtumur Bum-Erdene**[a], **Mona K. Ghozayel**[a], **Baocheng Yang**[d], **Paul A. Clemons**[e], **Samy O. Meroueh**[a],[b]

[a]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202 (USA)

[b]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, 46202 (USA)

[c]Department of BioHealth Informatics, Indiana University School of Informatics and Computing, Indianapolis, IN, 46202 (USA)

[d]Henan Provincial Key Laboratory of Nanocomposites and Applications, Institute of Nanostructured Functional Materials, Huanghe Science and Technology College, Zhengzhou, Henan 450006 (China)

[e]Chemical Biology and Therapeutics Science Program, Broad Institute of Harvard and MIT, Cambridge, MA, 02142 (USA)

Abstract

There is growing interest in the use of structure-based virtual screening to identify small molecules that inhibit challenging protein–protein interactions (PPIs). In this study, we investigated how effectively chemical library members docked at the PPI interface mimic the position of critical side-chain residues known as “hot spots”. Three compound collections were considered, a commercially available screening collection (ChemDiv), a collection of diversity-oriented synthesis (DOS) compounds that contains natural-product-like small molecules, and a library constructed using established reactions (the “screenable chemical universe based on intuitive data organization”, SCUBIDOO). Three different tight PPIs for which hot-spot residues have been identified were selected for analysis: uPAR·uPA, TEAD4·Yap1, and CaVα·CaVβ. Analysis of library physicochemical properties was followed by docking to the PPI receptors. A pharmacophore method was used to measure overlap between small-molecule substituents and hot-spot side chains. Fragment-like conformationally restricted small molecules showed better hot-spot overlap for interfaces with well-defined pockets such as uPAR·uPA, whereas better overlap was observed for more complex DOS compounds in interfaces lacking a well-defined binding site

smeroueh@iu.edu.

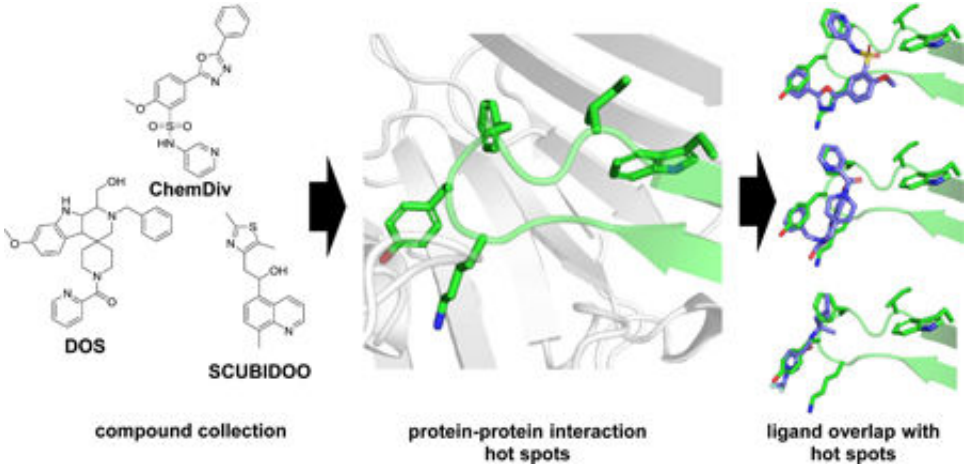
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under : https://doi.org/10.1002/cmdc.201800537.

Conflict of interest

The authors declare no conflict of interest.

such as TEAD4·Yap1. Virtual screening of conformationally restricted compounds targeting uPAR·uPA and TEAD4·Yap1 followed by experimental validation reinforce these findings, as the best hits were fragment-like and had few rotatable bonds for the former, while no hits were identified for the latter. Overall, such studies provide a framework for understanding PPIs in the context of additional chemical matter and new PPI definitions.

Graphical Abstract



Keywords

compound libraries; computational chemistry; protein–protein interactions; virtual screening

Introduction

Protein–protein interactions (PPIs) range in binding affinities from weak to tight.[1–3] PPIs have been classified into three categories: primary, secondary, and tertiary.[4] Primary interactions consist of a linear peptide segment bound to a protein receptor. A secondary interaction typically involves an α-helix or β-turn that is ensconced into a well-defined binding cavity. Tertiary interactions are more complex, usually containing multiple α -helices and β-strands over large interfaces.[5] For primary interactions with druggable and relatively weak affinities (K_d > 0.1 μm), such as bromodomains, small-molecule inhibitors have been successfully developed; several compounds have advanced to clinical trials and one compound was FDA-approved.[6] Disruption of secondary and tertiary interactions with small molecules is more challenging. Small-molecule inhibitors of tight secondary and tertiary PPIs are relatively rare. Successful examples include inhibitors of MDM2·p53, but this interaction is weak (K_d ≈1 μm).[7] For tight interactions (K_d <100 nm), there are only a handful of small-molecule inhibitors for interactions such as Bcl-xL·Bak,[8] IL-2·IL-2Rα,[9] and more recently KEAP1·NRF2.[10]

There is growing interest in applying computational methods for the discovery of small-molecule PPI inhibitors, particularly for tight secondary and tertiary interactions. We have successfully used virtual screening to identify inhibitors of the tight tertiary interaction

between the urokinase receptor uPAR and its protein ligand uPA with compounds ranging from sub-micromolar to micromolar affinity.[11–13] In one case, we screened multiple structures that were sampled from explicit-solvent molecular-dynamics simulations and identified a sub-micromolar affinity compound.[11] The predicted structure of the compound was independently confirmed by X-ray crystallography.[14] More recently, we introduced a fingerprint method that uses the native protein ligand as a guide to identify small molecules that mimic the interaction of the protein ligand.[13] Using the fingerprint method, we identified several hit compounds with single-digit micromolar affinities. Finding highly potent inhibitors of tight PPIs by virtual screening of commercial libraries remains extremely challenging, but we have shown that quality hit compounds with single-digit micromolar binding affinities can be identified.

Although progress has been made in the development of scoring and docking methods to enrich compound collections for the discovery of PPI inhibitor hit compounds, relatively little work has been done on the suitability of existing commercial and specialized collections for PPI drug discovery. Data-bases such as iPPI,[15] 2P2I,[16] and TIMBAL[17] have been created to explore additional chemotypes for PPIs. A few studies have explored whether commercial and specialized libraries contain small molecules that are suitable for disrupting PPIs.[18–20] These studies suggest that small-molecule PPI inhibitors tend to cover different chemical space than enzyme inhibitors. Another approach to determine whether a compound collection is enriched with compounds that are suitable for disrupting PPIs is to explore how effectively compounds mimic the side chains of critical amino acids of the PPI ligand protein, or whether compounds engage critical residues on the PPI receptor protein.[21] Small molecules that better mimic or engage critical residues are expected to be more effective inhibitors of PPIs.[22–24]

In this study, we explored how effectively small molecules in a commercial library (ChemDiv), a collection of diversity-oriented synthesis (DOS) libraries, and the “screenable chemical universe based on intuitive data organization” (SCUBIDOO) compound collections mimic the positions of critical residues at tight protein–protein interfaces. We employed a combined docking and pharmacophore approach to measure overlap between chemical structures and the side chains of interface residues. We selected three tight PPIs for which extensive alanine-scanning studies have identified critical hot-spot residues; these include the interaction between the urokinase receptor (uPAR) and its serine proteinase ligand urokinase (uPA), the transcription factor TEAD and its co-activator Yap (TEAD·Yap), and the α- and β-subunits of the voltage-gated calcium channels ($Ca_V\alpha \cdot Ca_V\beta$). Our findings suggest that smaller conformationally restricted compounds show excellent overlap with hot spots. In a proof-of-concept study, we experimentally validated the compounds identified from virtual screening of a library of conformationally restricted commercially available compounds against the PPIs of uPAR·uPA and TEAD·Yap.

Results

Analysis of compound collection physicochemical properties

Several types of chemical libraries have been created over the years that may be suitable in identifying PPI inhibitors. Among them, diversity-oriented synthesis (DOS) represents a

strategy to efficiently generate compound collections with a high degree of structural diversity[25,26] and to produce new biological probes.[27–33] The DOS approach aims to achieve buildingblock, functional group, stereochemical, and skeletal diversity, occupying a middle ground between the structural complexity of natural products and the efficiency of commercially available synthetic libraries.[34] In addition to commercially available and DOS compounds, there is an increasing number of specialized libraries that have been constructed for use in virtual screening. These libraries were designed to overcome some of the shortcomings of commercial libraries, namely the limited chemical space commercial libraries cover and the large number of nuisance compounds.[35–37] The SCUBIDOO library[38] is one example. It consists of compounds constructed by combining building blocks from 58 organic reactions known to the pharmaceutical field.

To explore the structural overlap between small molecules and amino acid side chains at protein–protein interfaces, three compound collections from both commercial and non-commercial sources were selected. The first was the commercially available library from ChemDiv, Inc. (ChemDiv), which is frequently used in compound screening. The second was a collection of DOS libraries[33,39] from the Broad Institute (DOS), and the third was SCUBIDOO.[38] The physicochemical properties of the compounds in each of the three collections were determined (Figure 1). Compounds in DOS had a mean molecular weight (MW) of 512±94 Da, which is larger than the mean molecular weight of compounds in ChemDiv (MW=410±74 Da) and SCUBIDOO (MW=327±49 Da) (Figure 1A). Lipophilicity, characterized here by computed Alog*P* values, plays a crucial role in determining solubility. Compounds in ChemDiv had a mean Alog*P* of 3.6±1.5 (Figure 1B). ChemDiv compounds are predicted to be generally less soluble than compounds from DOS or SCUBIDOO considering their substantially lower mean Alog*P* values of 2.3±1.6, and 1.3±1.8, respectively.

To gain insight into the three-dimensional characteristics and complexity of compounds in these collections, the number of chiral centers was counted (Figure 1C). DOS compounds had generally more chiral centers (3.6±0.8), compared with SCUBIDOO (2.4±0.7) and ChemDiv compounds (0.7±1.2). These data confirmed that DOS compounds are generally more stereochemically complex than SCUBIDOO compounds. ChemDiv compounds showed a remarkably low number of chiral centers per compound compared with both DOS and SCUBIDOO.

A measure of flexibility is the number of rotatable bonds in a compound (Figure 1D). There were more rotatable bonds on average per compound in DOS (RB=6.8±2.1) compared with SCUBIDOO (RB=3.8±1.1) and ChemDiv (RB=5.5±2.1). The molecular weight of a compound was generally correlated with the number of rotatable bonds in the DOS (Pearson's r= 0.68) and ChemDiv collections (r=0.55), but not in the SCUBIDOO library (r=0.13). There was a similar trend in the correlations between molecular weight and number of chiral centers in DOS (r=0.31), ChemDiv (r=0.20), and SCUBIDOO (r=0.08). Similarly, the number of chiral centers and rotatable bonds was only correlated in the DOS collection (r=0.27).

Polar surface area (PSA) is a commonly used descriptor to provide insight into compound permeability and oral bioavailability.[40] In addition, PSA is often used in combination with the number of rotatable bonds to reflect molecular flexibility.[41] Compounds in DOS had the highest PSA values with a mean of 104±31 $Å^2$, followed by ChemDiv (PSA=84±27 $Å^2$) and SCUBIDOO (PSA=66±28 $Å^2$) (Figure 1E). We found that PSA and the number of rotatable bonds were most strongly correlated in DOS (r=0.60) and ChemDiv (r=0.40), but not in SCUBIDOO (r=−0.02).

The number of hydrogen-bond donors (HBDs) and hydrogen-bond acceptors (HBAs) are important parameters related to compound polarity and membrane permeability. It has been suggested that the number of HBDs may be more important than the number of HBAs to enhance bioavailability and membrane permeability of lead compounds.[40,42,43] We found that DOS compounds had the highest number of HBDs and HBAs (HBD=2.0±0.9; HBA=5.2±1.6), followed by ChemDiv (HBD= 1.0±0.8; HBA=3.8±1.4) and SCUBIDOO (HBD=1.2±0.9; HBA=2.6±1.4) (Figure 1F,G). Among all the physicochemical properties that we have considered, only the ring count was similar among the three collections (Figure 1H). Because SCUBIDOO was built by limiting the number of building blocks, compounds in SCUBIDOO have lower molecular weight and fewer rotatable bonds and chiral centers than compounds in DOS and ChemDiv.

We next compared the distributions of these eight physicochemical properties of the three collections using principal component analysis (PCA). Compounds from each of the three collections were projected onto the first two principal components (Figure 2A). In the first principal component, the distributions were separated such that there are distinct peaks for each of the three sources ($PC1_{DOS}$ =−1.6±1.7; $PC1_{ChemDiv}$ = 0.1±1.3; $PC1_{SCUBIDOO}$ =1.5±0.9). This effect is observed qualitatively in the marginal distributions for PC1 for each compound collection, where each set was projected onto different ranges of the first principal component. In the second principal component, however, the marginal distributions of DOS ($PC2_{DOS}$ = −0.5±0.9) and SCUBIDOO ($PC2_{SCUBIDOO}$ =−0.5±1.1) overlapped and each partially overlapped with the marginal distribution of ChemDiv ($PC2_{ChemDiv}$ =0.9±1.0). The loadings of the PCA are provided to illustrate the relative contributions of each of the eight input descriptors (Figure 2B). The total variance explained by the first two principal components were 44% and 18%, respectively. When the third principal component is included, the cumulative variance explained reached 76%.

To gain further insight into the structure of compounds in these three collections, the molecular shape diversity of each was evaluated using principal moments of inertia (PMI) (Figure 3). PMI plots represent the shape distribution of a collection of molecules. The three vertices of the triangular plot represent the extremes of molecular geometry. The top-left, top-right, and bottom corners correspond to small molecules with linear (e.g., diacetylene), spherical (e.g., adamantane), and disc-like (e.g., benzene) shapes, respectively. Compounds in SCUBIDOO (I_1/I_3 =0.26±0.12; I_2/I_3 =0.88±0.08) were predominantly linear relative to compounds in ChemDiv and DOS. ChemDiv (I_1/I_3 =0.31±0.13; I_2/I_3 =0.84±0.09) and DOS (I_1/I_3 = 0.33±0.12; I_2/I_3 =0.82±0.09) compounds were primarily along the diagonal between linear and disc-like structures. Despite the substantial difference in the number of chiral centers and rotatable between DOS (chiral center=3.6±0.8; RB=6.8±2.1) and ChemDiv

(chiral center=0.7±1.2; RB=5.5±2.1), there was little difference in the shape diversity between the two collections.

Compound overlap with protein–ligand hot spots at protein–protein interaction interfaces

We wondered if the chemical scaffolds found in each compound collection would yield compounds that could overlap and mimic the physicochemical properties of amino acid side chains at protein–protein interfaces. To explore this question, we selected three high-affinity (K_d <100 nm) PPIs with three distinct binding motifs: 1) a β-turn motif at the interface between the urokinase receptor and its serine proteinase ligand urokinase (uPAR·uPA); 2) a twisted-coil motif at the Ω-loop between the transcription factor TEAD and its co-activator Yap (TEAD·Yap); and 3) an a-helix motif between the α-and β-subunits of the calcium channel ($Ca_V\alpha \cdot Ca_V\beta$).

We first docked compounds from the ChemDiv, DOS, and SCUBIDOO to the protein receptor at each of the three interaction interfaces. A ligand-based pharmacophore approach was used to identify compounds that overlap with interface residues on the protein ligand. Six possible pharmacophore features can be assigned to the side chains of protein residues: hydrogen-bond acceptor (A) and donor (D), hydrophobic group (H), negatively (N) and positively (P) charged group, and aromatic ring (R). These pharmacophore sites are characterized by type, location, and, if applicable, directionality.

uPAR·uPA—The uPAR·uPA interface consists primarily of a β-turn on the protein ligand uPA ensconced in a large pocket on the protein receptor uPAR, leading to an interaction that is both tight (K_d = 1 nm) and stable ($k_{off} = 10^{-4}\ s^{-1}$) (Figure 4A,B). At the uPAR·uPA interface, the side chain of five hot-spot residues from uPA extend into the hydrophobic pocket of uPAR: Lys-23, Tyr-24, Phe-25, Ile-28, and Trp-30 (Figure 4C).[44] We used pharmacophores to represent the position and physicochemical properties of the side chains of these five residues. For a given pharmacophore, we compared each docked compound and determined whether it had a chemical moiety that overlaps with the pharmacophore with the appropriate physicochemical property. For example, a compound that possesses a benzene group that occupies the same position as an aromatic ring (pharmacophore) of a tyrosine residue on uPA is expected to mimic the properties of the tyrosine residue in the interaction between uPAR·uPA and disrupt binding. In the pharmacophore model, the ε-amine on Lys-23 was modeled using a positive charge feature, while the benzene rings of Tyr-24 and Phe-25 were modeled using aromatic ring features. We assigned separate aromatic ring features to the benzene and pyrrole rings on the indole of Trp-30. The aliphatic side chain of Ile-28 was represented by a hydrophobic feature centered at the Cα−Cγ1 bond. We searched for compounds containing functional moieties that matched a corresponding pharmacophore feature.

There were 31387, 3157, and 11 compounds in ChemDiv that overlap with one, two, and three distinct hot spots on uPA, respectively (Figure 4D). For DOS, fewer compounds were found to overlap with hot spots. In total, 19210, 952, and 10 compounds overlapped with one, two, and three hot spots, respectively (Figure 4E). The number of overlaps for

SCUBIDOO was similar to DOS, with 17992, 1033, and 8 compounds that overlapped with one, two, and three hot spots, respectively (Figure 4F).

There are differences in the degree of overlap with individual hot spots among the three collections. At Lys-23, DOS and SCUBIDOO compounds showed dramatically greater overlap compared with ChemDiv. A total of 1094 and 773 compounds from DOS and SCUBIDOO, respectively, overlapped with Lys-23 compared with only 72 for ChemDiv. Similarly, for compounds that overlapped with Lys-23 and one other hot spot, we found 161 DOS and 184 SCUBIDOO compounds, compared with only 11 in ChemDiv. This finding might be attributed to the larger average Alog*P* value of ChemDiv (Alog*P*=3.6±1.6), which contains fewer polarizable moieties than DOS (Alog*P*=2.3±1.6) and SCUBIDOO (Alog*P*=1.3±1.8). Analogous trends were observed for both Ile-28 and Trp-30. At Ile-28, 784 ChemDiv and 1674 DOS compounds overlapped with the residue, while only 70 SCUBIDOO compounds showed overlap with this residue. At Trp-30, none of the compounds in ChemDiv or SCUBIDOO overlapped with either the indole or benzene rings of the side chain, while a total of 1772 compounds from DOS overlapped with Trp-30, including 213 compounds that overlap Trp-30 and at least one other residue. The latter finding may be attributed to the fact that Trp-30 is positioned outside of the deep binding pocket on uPAR, and may be more difficult to reach by the smaller compounds in ChemDiv and SCUBIDOO. When comparing the DOS compounds that overlap with Trp-30 to all DOS compounds, there was a slight shift in the distribution of aromatic rings from 4.0±0.9 to 4.3±0.9, the distribution of rotatable bonds from 6.8±2.1 to 7.0±2.1, and the distribution of chiral centers from 3.6±0.8 to 3.7±0.7. Relative to DOS and SCUBIDOO, the number of compounds in ChemDiv that overlapped with Tyr-24 and Phe-25 was nearly double when considering only compounds that overlap with a single hotspot. Similarly, there was more than a threefold increase in the number of compounds that overlapped with Tyr-24 or Phe-25 and one other hot spot. Although DOS compounds have proven successful in producing probe molecules after high-throughput screening, many of these compounds have high molecular weight.[39] Indeed, in our study, compounds from DOS that overlap with hot spots had higher molecular weight (MW=536:97 Da) than those from ChemDiv (MW=408±71 Da) and SCUBIDOO (MW=332±45 Da).

A question of interest is whether compounds that are predicted by docking studies to have higher binding affinities will show different overlap with hot-spot residues at the protein–protein interface. To address this question, we repeated the above analysis except that only compounds 1) that are 500 Da or less, and 2) that are among the top 1000 highest-scoring from a docking simulation were considered. There were 603, 433 and 605 compounds from ChemDiv, DOS, and SCUBIDOO that matched at least one hot spot (Figure 4D,E,F). ChemDiv and SCUBIDOO compounds all possessed substituents that occupy the same position as Lys-23, Tyr-24 or Phe-25 side chains on uPA, while none overlapped with hot spots on uPA located at Ile-28 and Trp-30. The DOS collection had only three compounds that overlapped these residues (Figure 4E). Furthermore, except for SCUBIDOO, none of the compounds in ChemDiv or DOS simultaneously showed an overlap with three hot spots. The four compounds from SCUBIDOO overlapped with Lys-23, Tyr-24, and Phe-25. It is interesting that compounds from SCUBIDOO, which are generally smaller in size than DOS and ChemDiv compounds, are able to mimic more hot spots at the protein–protein interface.

TEAD·Yap—The PPI between TEAD and Yap occurs over a large interface of 1300 Å^2.[45] The TEAD binding domain of Yap wraps around the globular structure of TEAD via three interfaces (Figure 5A). The twisted-coil region of Yap at the W-loop is most critical for complex formation (Figure 5B). There are six hot spots at this region, with four hydrophobic residues (Met-86, Leu-91, Phe-95, and Phe-96), one charged residue (Arg-89), and one polar residue (Ser-94). The aliphatic side chains of Met-86 and Leu-91 were identified as hydrophobic features in the pharmacophore model, the side chain of Arg-89 was modeled using a positive charge feature, and the benzene rings of Phe-95 and Phe-96 were modeled using aromatic ring features. The hydrogen and oxygen atom on the side chain of Ser-94 were identified as a hydrogen-bond donor and acceptor, respectively (Figure 5C). In the native complex, two hydrogen bonds are formed between Ser-94 on Yap and Glu-240 and Tyr-406 on TEAD.[46]

There were 47061, 8862, 403, and 5 compounds in ChemDiv that overlapped with one, two, three, and four different hot spots on Yap, respectively (Figure 5D). From the DOS collection, 50976, 9570, 425, and 3 compounds overlapped with one, two, three, and four different hot spots on Yap, respectively (Figure 5E). From the SCUBIDOO library, only 35369, 4341, and 59 compounds overlapped with one, two, and three hot spots simultaneously, while no compounds matched four hot spots (Figure 5F). Among compounds that overlapped with multiple hotspots, ChemDiv and DOS had similar distributions while SCUBIDOO had approximately half the number of compounds that overlapped with two hotspots and approximately 15% of the number of compounds that overlapped with three hotspots when compared with the other two collections.

In all three sets, approximately a quarter of all compounds overlapped with the hydrophobic pharmacophore on Met-86. Similarly, about 10% in each collection overlapped with the hydrophobic Leu-91 and aromatic Phe-95 residues on Yap. At Phe-96, DOS had substantially more compounds that overlapped with the residue either alone or with one other residue compared to the other two collections. Overlap with Arg-89, Ser-94, and Phe-95 were less prevalent in ChemDiv and SCUBIDOO compared to DOS. Moreover, no compounds in SCUBIDOO matched the pharmacophore of Ser-94. This observation may be attributed to the lower molecular weight and more limited structural diversity in this library, which limits the ability of compounds to overlap with residues that are outside the immediate binding pocket.

As with uPAR·uPA, we explored hot-spot overlap of compounds that were predicted by docking to have the highest binding affinities to the protein receptor, in this case TEAD4. A collection of compounds was created by selecting 1000 compounds with the highest binding scores and molecular weights less than 500 Da. In ChemDiv, there were 231 and 22 compounds that overlapped with one or two hot spots, respectively (Figure 5D). In DOS, there were 390, 126, and 6 compounds that match one, two, and three hot spots, respectively (Figure 5E). From SCUBIDOO, we found a total of 354 and 39 compounds that matched the side chain of one or two hot spots, respectively (Figure 5F). The compounds in ChemDiv and SCUBIDOO mainly mapped onto the hydrophobic residues of Met-86, Leu-91, and Phe-95, while DOS compounds generally had overlap with Ser-94 and Met-86. Interestingly, no compound overlapped with Phe-96 from any of the three collections. Moreover, in DOS,

there were six compounds that with three hot spots simultaneously. All six compounds overlapped with Met-86 and have 6.8±1.9 rotatable bonds and 3.8±1.7 chiral centers. One compound overlapped with the nearby Arg-89, which is located at the periphery of the binding pocket. This compound is structurally complex (RB=7, chiral centers=3), but its molecular weight is only 345.3 Da, which suggests that compound flexibility may play a larger role than molecular weight in targeting the flat TEAD·Yap pocket.

$Ca_V\alpha \cdot Ca_V\beta$—The structural basis of the $Ca_V\alpha \cdot Ca_V\beta$ interaction was revealed in a co-crystal structure between $Ca_V\beta_3$ and the a-interacting domain (AID) of $Ca_V\alpha$ ($Ca_V\alpha_{AID}$) (Figure 6A). $Ca_V\alpha_{AID}$ is a 25-residue α-helix that binds to a well-defined groove on the GK domain of $Ca_V\beta_3$ (Figure 6B). Previous biophysical studies combined with site-directed mutagenesis on $Ca_V\alpha_{AID}$ revealed the presence of three hot-spot residues on the α-helix at Tyr-437, Trp-440, and Ile-441.[47] The side chains of these residues are ensconced into three sub-sites. The aliphatic side chain of Ile-441 was identified as a hydrophobic feature in the pharmacophore model, while the benzene ring of Tyr-437 was modeled using an aromatic ring feature. We assigned separate aromatic ring features to the benzene and pyrrole rings on the indole of Trp-440 (Figure 6C).

There were 59171, 38940, and 2099 compounds in ChemDiv that overlapped with one, two, and three hot spots, respectively (Figure 6D). From DOS and SCUBIDOO, substantially fewer compounds were found to overlap with multiple hot spots. There were 58219, 13350, and 476 compounds from DOS that overlapped with one, two, and three separate hot spots, respectively (Figure 6E), and correspondingly 45097, 15757, and 40 from SCUBIDOO (Figure 6F). While the number of compounds that overlap with a single hot spot are similar, the number of compounds that overlap with two or three hot spots is substantially different for ChemDiv compared to DOS and SCUBIDOO. The largest difference was seen in compounds that overlap with either Tyr-437 or Trp-440 as well as one other hot spot. In ChemDiv, approximately 35000 compounds overlapped with one of the two aromatic hot spots as well as another residue. If a compound in ChemDiv was overlapping with multiple residues, it was often at both the Tyr-437 and Trp-440 sites, while fewer compounds overlapped with one of these two aromatic residues and the hydrophobic Ile-441. This trend was also observed in SCUBIDOO, where there was much higher co-occurrence of a combination of Tyr-437 and Trp-440 than between one of the two aromatic residues and Ile-441. Among DOS compounds, in contrast, the co-occurrence rate between the three residues was more uniform, with 7980, 9767, and 8953 compounds that overlapped with Tyr-437, Trp-40, and Ile-441 and one other residue, respectively.

We again considered compounds with highest predicted binding affinities and low molecular weight. The top highest-scoring 1000 compounds with molecular weights less than 500 Da were selected based on docking score (Figure 6D–F). There were 936, 812, and 951 compounds for ChemDiv, DOS, and SCUBIDOO matching at least one hot spot. It is interesting that despite the higher complexity of DOS compounds (chiral centers=3.5±1.0) fewer matched hot spots than ChemDiv and SCUBIDOO compounds, which are less rich in chiral centers (ChemDiv chiral centers=0.4±0.7; SCUBIDOO chiral centers=1.3±1.0). Interestingly, there were twice as many compounds in ChemDiv and SCUBIDOO that overlapped with at least two hot spots than DOS. Remarkably, there were 50 compounds in

ChemDiv that overlapped with three hot spots relative to only two each for DOS and SCUBIDOO.

Virtual screening of commercial library against two protein–protein interactions

The discovery that conformationally restricted fragment-like libraries such as SCUBIDOO can be effective in mimicking protein–ligand hot spots for interfaces with well-defined pockets prompted us to test this observation with virtual screening and experimental validation. The ChemDiv commercial library was filtered for conformationally restricted compounds and tested against the uPAR·uPA interaction. The ChemDiv library was filtered for compounds between 350 and 500 Da having six or fewer rotatable bonds as well as high predicted solubility (Alog$P \leq 4$). The resulting collection of 50,893 compounds was docked to uPAR at the uPAR·uPA interface and to TEAD at the TEAD·Yap interface. Compounds were ranked based on their overlap with hot spots located at the protein–protein interface. The top 85 compounds were selected and tested for inhibition of uPAR·uPA and TEAD·Yap using a previously developed fluorescence polarization assay. The compounds were first tested in duplicate at a single concentration of 50 μm (Figure 7A). For the uPAR·uPA interface, five compounds, namely 1 (IPR-3247), 2 (IPR-3260), 3 (IPR-3271), 4 (IPR-3288) and 5 (IPR-3305) inhibited more than 25% (Supporting Information Table S1). These compounds were also tested in a concentration-dependent manner (Figure 7B). All five compounds inhibited the FP assay, with K_i values in the double-digit micromolar range. All five were predicted to overlap with two aromatic hot spots on uPA, Tyr-24 and Phe-25 (Figure 7C). Compound 2, however, ap-pears to have poor solubility at concentrations that are higher than 25 μm. Compounds **3** and **5** also exhibited poor solubility at higher concentrations. The two compounds with reasonably good solubility were **1** and **4** with K_i values of 25 and 62 μm, respectively.

Conformationally restricted compounds from ChemDiv were docked against TEAD at the TEAD4·Yap interface. The top 81 compounds from virtual screening against the Ω-loop pocket on TEAD4 were experimentally validated using a fluorescence polarization assay. Compounds were tested in duplicate at an initial concentration of 50 μm (Figure 7D). No compounds were found to inhibit more than 15%. TED-97, a previously discovered small-molecule inhibitor of TEAD4·Yap1 was used as positive control. A concentration-dependent study was performed for compounds that inhibited by 15%, but none exhibited concentration-dependent inhibition. The lack of hits confirms the highly challenging nature of this interaction and the fact that to successfully inhibit this target, larger more complex compounds such as DOS may be required. It is interesting to note that compounds in ChemDiv are generally different from the more complex DOS library compounds (see Figure 1). Furthermore, the overlap of ChemDiv compounds and DOS compounds with hot spots at the TEAD·Yap interface is different for both the whole library than for the top 1000 compounds (see Figure 5).

The ranking of the filtered and unfiltered compound sets from uPAR was compared to determine whether pre-filtering by rotatable bonds, molecular weight, and Alog$P$ led to further enrichment of the original library for hit compounds. We identified the ranking of the five active compounds using the distribution of docking scores of the compounds in the

unfiltered ChemDiv library (Supporting Information Figure S1). For the unfiltered library, we found that four of the five hit compounds have docking scores that puts them between the 25th and 75th percentiles. In other words, had we strictly ranked the unfiltered library, these hit compounds would not have been selected as they would not have been among the top candidates chosen for experimental validation. Only compound **3** (IPR3271) would have been among the top 100 compounds in the unfiltered library (i.e., without physicochemical filters).

Discussion

Three chemical libraries were compared for their ability to mimic the positions of amino acid side chains of critical residues at protein–protein interfaces: 1) a commercial library from ChemDiv; 2) a collection of diversity-oriented synthesis (DOS) libraries from the Broad Institute; and 3) SCUBIDOO, a non-commercial library built by combining building blocks of common reactions in medicinal chemistry. The physicochemical properties of each of these three collections was explored. DOS compounds were larger, had more chiral centers, and were more flexible as evidenced by a larger number of rotatable bonds compared with ChemDiv and SCUBIDOO. Similarly, the molecular weight distributions of the compound collections correlates well with the number of rotatable bonds and chiral centers in DOS and ChemDiv, but not SCUBIDOO. Principal component analysis revealed a primary separation of all three collections in the first component, while the second principal component did not distinguish DOS from SCUBIDOO libraries. Finally, principal moment of inertia analysis suggests that compounds in the SCUBIDOO library are enriched for linear compounds, while ChemDiv and DOS contain compounds with more disc-like and some globular characteristics.

We explored the ability of these three compound collections to mimic the physicochemical properties of amino acid side chains of hot spots at the interface of three PPIs. Three tight PPIs were selected: 1) a β-turn motif at the interface between the urokinase receptor and its serine proteinase ligand urokinase (uPAR·uPA), 2) a twisted-coil motif at the Ω-loop between the transcription factor TEAD and its co-activator Yap (TEAD·Yap), and 3) an a-helix motif between the α-and β-subunits of the calcium channel ($Ca_V\alpha \cdot Ca_V\beta$). In each system, hot-spot residues were identified on the protein ligand at each PPI interface. Each library showed different overlap with individual hot spots in each PPI. For example, for uPAR·uPA, ChemDiv contains more compounds that overlap with the two aromatic residues Tyr-24 and Phe-25, while DOS and SCUBIDOO possess many more compounds that overlap with the positively charged Lys-23 residue. Similarly, DOS was twice as likely to overlap with Ile-28 than ChemDiv and 20 times as likely than SCUBIDOO. DOS was also the only collection to have compounds that overlapped with the more distant Trp-30 hot spot. Similar trends were observed in TEAD·Yap, where DOS compounds were better able to mimic Arg-89, Ser-94, and Phe-96 than those of ChemDiv and SCUBIDOO, but ChemDiv compounds were better able to mimic Phe-95 than the other two collections. In $Ca_V\alpha \cdot Ca_V\beta$, we found similar distributions in the number of compounds that overlapped with a single hot spot, but more than two and four times the number of compounds that overlapped with two and three distinct hot spots in ChemDiv compared to DOS. The gap is

further exacerbated when comparing ChemDiv and SCUBIDOO, where there were two and 50 times as many compounds that overlapped with two and three hot spots, respectively.

We refined our analysis, except that the top 1000 best scoring compounds by docking simulation were selected, subject to a molecular weight filter. This new set of compounds was enriched for small molecules predicted to bind to the target with high affinity. Interestingly, there were many more small molecules from ChemDiv and SCUBIDOO that overlapped with at least two hot spots for uPAR·uPA and $Ca_V\alpha \cdot Ca_V\beta$. In fact, for the $Ca_V\alpha \cdot Ca_V\beta$ interaction, there were more than 50 compounds in ChemDiv that overlapped with three hot spots, compared with only two for both DOS and SCUBIDOO. For TEAD4·Yap1, the DOS collection showed substantially greater overlap. These results can be explained by the fact that when compounds are ranked by their binding affinity, the resulting set is likely enriched for small molecules that make the best shape complementarity with existing pockets and therefore result in the largest solvent-accessible surface area change. In the case of uPAR and $Ca_V0\beta$, both receptors possess well-defined binding pockets. Considering that ChemDiv and SCUBIDOO possess a large number of small fragment-like compounds, it is more likely that these compounds will fit best into pockets and sub-pockets of these receptors. DOS compounds, on the other hand, are larger and more flexible, and may not score as well as SCUBIDOO and ChemDiv library small molecules, which may explain why there are fewer of these compounds that overlap with two or more hot spots in the top 1000 set. Because TEAD4 has a shallow pocket, DOS compounds will probably score higher, as for a flat surface, a larger compound is much more likely to lead to greater surface burial and therefore a better score. These results were tested by carrying out a virtual screen of more than 50000 compounds from the ChemDiv library against uPAR and TEAD. We selected conformationally restricted compounds between 350 and 500 Da in MW. Interestingly, several hits emerged in uPAR, and the most promising among them are fragment-like hits that have 2–3 rotatable bonds. No hits were identified for TEAD4.

Our work suggests that small fragment-like compounds that are conformationally restricted can be good candidates for PPI with well-defined pockets, while larger more complex compounds may be a better option for PPI with large but flat and featureless surfaces. The affinity of large hit compounds is generally driven by entropy, so these compounds may not show optimal fit to smaller, well-defined binding pockets. Hence, improving the enthalpy of binding becomes a significant challenge that is only accomplished by substantial modifications to the compound structure. Small fragment-like compounds are generally driven by enthalpy. These compounds tend to show optimal fit into pockets. It is easier to modify these compounds by extending them into neighboring pockets while at the same time improving entropy of binding.

Experimental Section

Ligand preparation:

Three libraries were chosen to compare how compounds mimicked the hot-spot residues on the different PPI interfaces. The first is the ChemDiv library, which consists of over 1.7 million compounds that were retrieved from the ZINC15 website.[48] The second collection is a set of diversity-oriented synthesis (DOS) libraries prepared at the Broad Institute

(Supporting Information Table S2). This collection consists of 100903 compounds from 32 libraries that were synthesized using DOS principles applied to 11 different reaction pathways.[33,39] The third collection is from SCUBIDOO, which was developed from 58 organic reactions known in the pharmaceutical field totaling over 21 million compounds.[38] The SCUBIDOO website provides three different representative samples created using a stratified sampling algorithm. We selected the M library from SCUBIDOO consisting of 99977 compounds. Compounds predicted to be pan-assay interference compounds (PAINS) were identified and filtered using the PAINS1, PAINS2, and PAINS3 filters in Canvas,[49,50] resulting in 1428800 compounds for ChemDiv, 99663 compounds for DOS, and 93074 compounds for SCUBIDOO. ChemDiv compounds retrieved from ZINC were previously prepared through their internal workflow.[51] Compounds from DOS and SCUBIDOO were prepared using LigPrep.[52] Epik was used for protonation-state assignment and tautomer generation.[53] The OPLS_2005 force field was used for minimization and the ionization sates were generated at pH 7.[54] Compounds were desalted to exclude additional molecules such as counter ions in salt and water molecules and tautomers were generated. For DOS and SCUBIDOO, stereoisomers were generated by retaining specified chiralities and varying those where the stereochemistry of the chiral center were undefined. Up to 32 different stereoisomers per ligand were generated in this manner, resulting in 127483 compounds for DOS and 125917 compounds for SCUBIDOO. The size of each compound library was normalized by random sampling to 125917 compounds.

Principal component analysis:

Principal component analysis (PCA) aims to simplify high-dimensional data by projecting the data onto a new set of dimensions that most effectively captures the variance in the data. We used it to visualize similarities and differences between the physicochemical properties of different collections of compounds. Eight physicochemical features were used for PCA: molecular weight (MW), Alog*P*, number of hydrogen-bond acceptors (HBA), number of hydrogen-bond donors (HBD), number of rotatable bonds (RB), polar surface area (PSA), number of chiral centers, and number of rings. The mean of each feature is shifted to zero and each feature is scaled to have unit variance prior to the analysis. PCA is calculated using singular value decomposition in R (version3.2.3).

Principal moment of inertia:

Principal moment of inertia (PMI) descriptors provide an intuitive notion of the three-dimensional shape diversity of the various compound data sets. Lowenergy conformations are identified for each molecule in the data set, and three PMI values (I_1, I_2 and I_3; where $I_3 \geq I_2 \geq I_1$) are calculated for each conformer. Normalized ratios of PMI (I_1/I_3 and I_2/I_3) are then calculated and plotted on a triangular graph, with the vertices (0,1), (0.5,0.5), and (1,1) representing a perfect linear (diacetylene), disc-like (benzene), and spherical (adamantane) compound, respectively. The moments of inertia in the three spatial dimensions were calculated by the *calculate_pmi*.py script in Schrödinger.

Protein preparation:

The structures of the uPAR·uPA (PDB ID: 3BT1), TEAD·Yap (PDB ID: 3KYS), and $Ca_V\alpha \cdot Ca_V\beta$ (PDB ID: 1VYT) interactions were retrieved and prepared using the Protein

Preparation Wizard using the Schrçdinger Suite.[52,55] Bond orders were assigned, hydrogen atoms were added, and disulfide bonds were created. Water residues and additional ions and heteroatom groups were discarded. Missing side chains and loops were introduced using the Prime module.[56] The resulting protein structures were protonated at pH 7.0 using PROPKA.[57]

Virtual screening:

The compound library was docked to the prepared protein structures using AutoDock Vina (Version 1.1.2).[58] The binding pocket was centered at each of the interfaces with a box with dimensions of 21 Å×21 Å×21 Å. All other parameters were set to default values. The docked conformations were converted back to MOL2 format using inhouse Python scripts for additional analysis. Glide was used to score the Vina-docked binding modes in place using the GlideHTVS scoring function.[59,60]

Ligand pharmacophore:

A previously described[13] pharmacophore-based approach was adapted and used to identify how docking compounds overlapped with and mimicked known hot spots of the protein ligand in each of the protein–protein complexes. A set of pharmacophore hypotheses was constructed corresponding to the physicochemical properties of the protein ligand side chain using the Phase package in Schrödinger.[61,62] Phase has six built-in types of pharmacophore features: 1) hydrogen-bond acceptor (A); 2) hydrogen-bond donor (D); 3) hydrophobe (H); 4) negative ionizable (N); 5) positive ionizable (P); and 6) aromatic ring (R). The docked ligand conformation was used for pharmacophore calculations with an inter-site distance matching tolerance of 2.0 Å. Hydrogen-bond acceptor sites were positioned on atoms that carry one or more donatable lone pairs, while hydrogen-bond donor sites were centered on each electrophilic site. Negative and positive ionizable sites were modeled as single points located on a formally charged atom, or at the centroid of a group of atoms over which the ionic charge is shared. Rings, isopropyl groups, *tert*-butyl groups, various halogenated moieties, and aliphatic chains are treated as hydrophobic sites.[61] Aromatic rings were distinguished from other hydrophobic groups and designated as a separate type of pharmacophore feature (i.e., “R” rather than “H”). In these cases, a single site was placed at the centroid of each aromatic ring. In aromatic ring pharmacophores, a normal vector was projected orthogonal to the plane of each ring. Similarly, in positive and negative ionizable pharmacophores (i.e., “P” and “N”, respectively), a vector parallel to the plane of the respective atom was formed.

Compound conformers from virtual screening were used to identify matches without refinement using Phase’s default fitness function. In this scoring function, three factors are used to describe the degree to which a compound matches a pharmacophore feature: 1) the site score, which describes how well the compound superimposes the pharmacophore feature, 2) the vector score, which describes the cosine angle between the normal vector of an aromatic ring on the compound with an aromatic feature, and 3) the volume score, which describes the proportion of the total volume of the pharmacophore feature overlapped by the compound. The vector score ranges from −1.0 (antiparallel) to 1.0 (parallel) in positive and negative ionizable pharmacophores, and 0.0 (perpendicular) to 1.0 (parallel) in aromatic

pharmacophores. Vector scores less than 0.8, corresponding to angles more than ±37°, were rejected. Compounds with either root-mean-square deviation (RMSD) overlap greater than 1.2 Å or that did not overlap with the pharmacophore feature were discarded. All other parameters were set at default values. The remaining compounds that matched a given pharmacophore were retained without sorting compounds by Phase's internal scoring function.

Fluorescence polarization (FP) assay:

Polarized fluorescence intensities were measured using EnVision Multilabel plate readers (PerkinElmer, Waltham, MA, USA) with excitation and emission wavelengths of 485 and 535 nm, respectively.[11] Samples were prepared in Thermo Scientific Nunc 384-well black microplate in duplicates. First, the compounds were diluted in DMSO and further diluted in 1× PBS buffer with 0.01% Triton X-100 for a final concentration of 200–0.2 3μm. Triton X-100 was added to the buffer to avoid compound aggregation. 5 mL of the compound solution and 40 μL of PBS with 0.01% Triton X-100 containing uPAR was added to the wells and incubated for at least 15 min to allow the compound to bind to the protein. Finally, 5 μL of fluorescent AE147-FAM peptide solution was added for a total volume of 50 μL in each well resulting in final uPAR and peptide concentrations of 250 nm and 100 nm respectively. For the TEAD4·Yap1 FP assay, final concentrations of 64 nm GST-TEAD4 (217–434) and 16 nm FAM-labeled Yap1 peptide (residues 60–99) were used. The final DMSO concentration was 2% *v/v*, which had no effect on the binding of the peptide. Controls included wells containing only the peptide, and wells containing both protein and peptide, each in duplicate to ensure the reproducibility of the assay. When compounds were insoluble and visible precipitation was observed, the data points at high concentrations were not included in the calculation of IC_{50} values. Inhibition constants were calculated from the IC_{50} values using the K_i calculator available at http://sw16.im.med.umich.edu/software/calc_ki/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported by the US National Institutes of Health (R01CA197928 to S.O.M.) and the American Cancer Society Research Scholar Grant RSG-12-092-01-CDD (S.O.M.). This work was done using resources provided by the Open Science Grid, which is supported by the US National Science Foundation award 1148698, and the US Department of Energy's Office of Science. We are grateful to Joshua Bittker, Karen Emmith, and Carol Mulrooney from the Broad Institute Compound Management team, and to Vlado Dancik, for assistance with DOS structure cleanup and compound registration.

References

[1]. Bahadur RP, Chakrabarti P, Rodier F, Janin J, J. Mol. Biol. 2004, 336, 943–955. [PubMed: 15095871]

[2]. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C, Structure 2010, 18, 1233–1243. [PubMed: 20947012]

[3]. Smith MC, Gestwicki JE, Expert Rev. Mol. Med. 2012, 14, e16.

[4]. Arkin MR, Tang Y, Wells JA, Chem. Biol. 2014, 21, 1102–1114. [PubMed: 25237857]

[5]. Bogan AA, Thorn KS, J. Mol. Biol. 1998, 280, 1–9. [PubMed: 9653027]

[6]. Souers AJ, Leverson JD, Boghaert ER, Ackler SL, Catron ND, Chen J, Dayton BD, Ding H, Enschede SH, Fairbrother WJ, Huang DC, Hymowitz SG, Jin S, Khaw SL, Kovar PJ, Lam LT, Lee J, Maecker HL, Marsh KC, Mason KD, Mitten MJ, Nimmer PM, Oleksijew A, Park CH, Park CM, Phillips DC, Roberts AW, Sampath D, Seymour JF, Smith ML, Sullivan GM, Tahir SK, Tse C, Wendt MD, Xiao Y, Xue JC, Zhang H, Humerickhouse RA, Rosenberg SH, Elmore SW, Nat. Med. 2013, 19, 202–208. [PubMed: 23291630]

[7]. Pazgier M, Liu M, Zou G, Yuan W, Li C, Li C, Li J, Monbo J, Zella D, Tarasov SG, Lu W, Proc. Natl. Acad. Sci. USA 2009, 106, 4665–4670. [PubMed: 19255450]

[8]. Petros AM, Dinges J, Augeri DJ, Baumeister SA, Betebenner DA, Bures MG, Elmore SW, Hajduk PJ, Joseph MK, Landis SK, Nettesheim DG, Rosenberg SH, Shen W, Thomas S, Wang X, Zanze I, Zhang H, Fesik SW, J. Med. Chem. 2006, 49, 656–663. [PubMed: 16420051]

[9]. Wilson CG, Arkin MR, Curr. Top. Microbiol. Immunol. 2011, 348, 25–59. [PubMed: 20703966]

[10]. Davies TG, Wixted WE, Coyle JE, Griffiths-Jones C, Hearn K, McMenamin R, Norton D, Rich SJ, Richardson C, Saxty G, Willems HM, Woolford AJ, Cottom JE, Kou JP, Yonchuk JG, Feldser HG, Sanchez Y, Foley JP, Bolognese BJ, Logan G, Podolin PL, Yan H, Callahan JF, Heightman TD, Kerns JK, J. Med. Chem. 2016, 59, 3991–4006. [PubMed: 27031670]

[11]. Khanna M, Wang F, Jo I, Knabe WE, Wilson SM, Li L, Bum-Erdene K, Li J, Khanna WSG,R, Meroueh SO, ACS Chem. Biol. 2011, 6, 1232–1243. [PubMed: 21875078]

[12]. Liu D, Xu D, Liu M, Knabe WE, Yuan C, Zhou D, Huang M, Meroueh SO, Biochemistry 2017, 56, 1768–1784. [PubMed: 28186725]

[13]. Xu D, Bum-Erdene K, Si Y, Zhou D, Ghozayel MK, Meroueh SO, ChemMedChem 2017, 12, 1794–1809. [PubMed: 28960868]

[14]. Rullo AF, Fitzgerald KJ, Muthusamy V, Liu M, Yuan C, Huang M, Kim M, Cho AE, Spiegel DA, Angew. Chem. Int. Ed. 2016, 55, 3642–3646; Angew. Chem. 2016, 128, 3706–3710.

[15]. Labbé CM, Kuenemann MA, Zarzycka B, Vriend G, Nicolaes GA, Lagorce D, Miteva MA, Villoutreix BO, Sperandio O, Nucleic Acids Res. 2016, 44, D542–547. [PubMed: 26432833]

[16]. Bourgeas R, Basse MJ, Morelli X, Roche P, PLoS One 2010, 5, e9598.

[17]. Higueruelo AP, Jubb H, Blundell TL, Database 2013, 2013, bat039.

[18]. Villoutreix BO, Kuenemann MA, Poyet JL, Bruzzoni-Giovanelli H, Labbe C, Lagorce D, Sperandio O, Miteva MA, Mol. Inform. 2014, 33, 414–437. [PubMed: 25254076]

[19]. Jin X, Lee K, Kim NH, Kim HS, Yook JI, Choi J, No KT, J. Mol. Graph. Model. 2018, 79, 46–58. [PubMed: 29136547]

[20]. ReynHs C, Host H, Camproux AC, Laconde G, Leroux F, Mazars A, Deprez B, Fahraeus R, Villoutreix BO, Sperandio O, PLoS Comput. Biol. 2010, 6, e1000695.

[21]. Xu D, Si Y, Meroueh SO, J. Chem. Inf. Model. 2017, 57, 2250–2272. [PubMed: 28766941]

[22]. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS, Proc. Natl. Acad. Sci. USA 2000, 97, 8950–8954. [PubMed: 10908667]

[23]. Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S, Bioinformatics 2009, 25, 621–627. [PubMed: 19176554]

[24]. Grosdidier S, Fernandez-Recio J, BMC Bioinf. 2008, 9, 447.

[25]. Spring DR, Org. Biomol. Chem. 2003, 1, 3867–3870. [PubMed: 14664374]

[26]. Spandl RJ, Bender A, Spring DR, Org. Biomol. Chem. 2008, 6, 1149–1158. [PubMed: 18362950]

[27]. Cordier C, Morton D, Murrison S, Nelson A, O'Leary-Steele C, Nat. Prod. Rep. 2008, 25, 719–737. [PubMed: 18663392]

[28]. Tan DS, Nat. Chem. Biol. 2005, 1, 74–84. [PubMed: 16408003]

[29]. Lenci E, Menchi G, Trabocchi A, Org. Biomol. Chem. 2016, 14, 808–825. [PubMed: 26632306]

[30]. O'Connor CJ, Beckmann HS, Spring DR, Chem. Soc. Rev. 2012, 41, 4444–4456. [PubMed: 22491328]

[31]. Comer E, Duvall JR, duPont Lee MT, Future Med. Chem. 2014, 6, 1927–1942. [PubMed: 25495985]

[32]. Scott DE, Bayly AR, Abell C, Skidmore J, Nat. Rev. Drug Discovery 2016, 15, 533–550. [PubMed: 27050677]

[33]. Gerry CJ, Schreiber SL, Nat. Rev. Drug Discovery 2018, 17, 333–352. [PubMed: 29651105]

[34]. Dandapani S, Marcaurelle LA, Nat. Chem. Biol. 2010, 6, 861–863. [PubMed: 21079589]

[35]. Galloway WR, Bender A, Welch M, Spring DR, Chem. Commun. 2009, 2446–2462.

[36]. Galloway WR, Spring DR, Expert Opin. Drug Discovery 2009, 4, 467–472.

[37]. Haggarty SJ, Curr. Opin. Chem. Biol. 2005, 9, 296–303. [PubMed: 15939332]

[38]. Chevillard F, Kolb P, J. Chem. Inf. Model. 2015, 55, 1824–1835. [PubMed: 26282054]

[39]. Marcaurelle LA, Comer E, Dandapani S, Duvall JR, Gerard B, Kesavan S, Lee MDT, Liu H, Lowe JT, Marie JC, Mulrooney CA, Pandya BA, Rowley A, Ryba TD, Suh BC, Wei J, Young DW, Akella LB, Ross NT, Zhang YL, Fass DM, Reis SA, Zhao WN, Haggarty SJ, Palmer M, Foley MA, J. Am. Chem. Soc. 2010, 132, 16962–16976. [PubMed: 21067169]

[40]. Leeson PD, Adv. Drug Deliv. Rev. 2016, 101, 22–33. [PubMed: 26836397]

[41]. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD, J. Med. Chem. 2002, 45, 2615–2623. [PubMed: 12036371]

[42]. Leeson PD, Davis AM, J. Med. Chem. 2004, 47, 6338–6348. [PubMed: 15566303]

[43]. Baell J, Congreve M, Leeson P, Abad-Zapatero C, Future Med. Chem. 2013, 5, 745–752. [PubMed: 23651089]

[44]. Magdolen V, Rettenberger P, Koppitz M, Goretzki L, Kessler H, Weidle UH, Konig B, Graeff H, Schmitt M, Wilhelm O, Eur. J. Biochem. 1996, 237, 743–751. [PubMed: 8647121]

[45]. Chen L, Chan SW, Zhang X, Walsh M, Lim CJ, Hong W, Song H, Genes Dev. 2010, 24, 290–300. [PubMed: 20123908]

[46]. Li Z, Zhao B, Wang P, Chen F, Dong Z, Yang H, Guan KL, Xu Y, Genes Dev. 2010, 24, 235–240.

[47]. Van Petegem F, Duderstadt KE, Clark KA, Wang M, Minor DL Jr., Structure 2008, 16, 280–294. [PubMed: 18275819]

[48]. Sterling T, Irwin JJ, J. Chem. Inf. Model. 2015, 55, 2324–2337. [PubMed: 26479676]

[49]. Duan JX, Dixon SL, Lowrie JF, Sherman W, Mol J. Graphics Modell. 2010, 29, 157–170.

[50]. Sastry M, Lowrie JF, Dixon SL, Sherman W, J. Chem. Inf. Model. 2010, 50, 771–784. [PubMed: 20450209]

[51]. Irwin JJ, Shoichet BK, J. Chem. Inf. Model. 2005, 45, 177–182. [PubMed: 15667143]

[52]. Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W, J. Comput.-Aided Mol. Des. 2013, 27, 221–234. [PubMed: 23579614]

[53]. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M, J. Comput. Aided Mol. Des. 2007, 21, 681–691. [PubMed: 17899391]

[54]. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM, J. Comput. Chem. 2005, 26, 1752–1780. [PubMed: 16211539]

[55]. Greenwood JR, Calkins D, Sullivan AP, Shelley JC, Comput J. Aided Mol. Des. 2010, 24, 591–604.

[56]. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA, Proteins Struct. Funct. Bioinf. 2004, 55, 351–367.

[57]. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH, Chem J. Theory Comput. 2011, 7, 525–537.

[58]. Trott O, Olson AJ, J. Comput. Chem. 2010, 31, 455–461. [PubMed: 19499576]

[59]. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS, J. Med. Chem. 2004, 47, 1739–1749. [PubMed: 15027865]

[60]. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL, J. Med. Chem. 2004, 47, 1750–1759. [PubMed: 15027866]

[61]. Dixon SL, Smondyrev AM, Rao SN, Chem. Biol. Drug Des. 2006, 67, 370–372. [PubMed: 16784462]

[62]. Dixon SL, Smondyrev AM, Knoll EH, Rao SN, Shaw DE, Friesner RA, J. Comput.-Aided Mol. Des. 2006, 20, 647–671. [PubMed: 17124629]

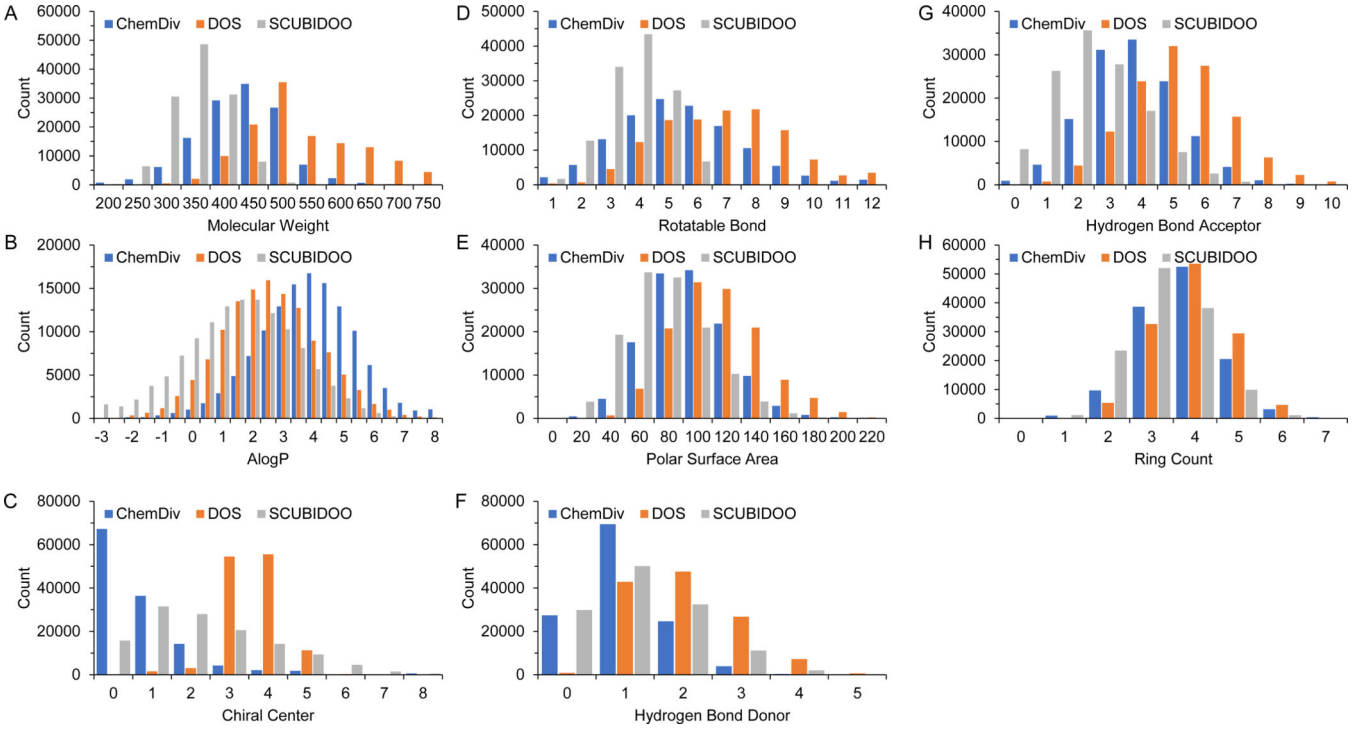
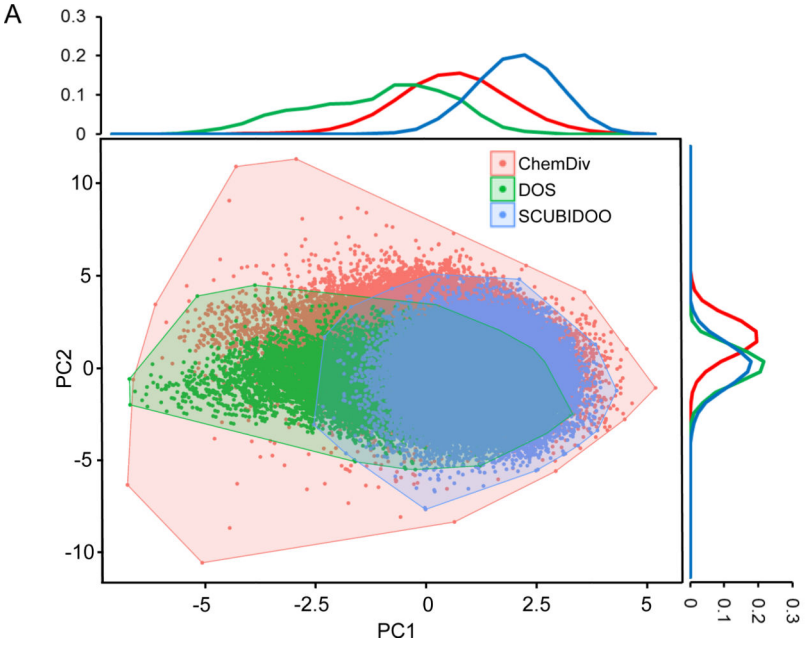


Figure 1.

Distributions of individual physicochemical properties of the ChemDiv, DOS, and SCUBIDOO compound collections highlight differences in size, flexibility, and complexity.

A



B

	Scale	PC1	PC2	PC3
ALogP	1.907	-0.157	0.692	-0.056
Hydrogen-bond acceptor	1.860	-0.439	-0.082	-0.241
hydrogen-bond donor	0.965	-0.300	-0.338	-0.061
molecular weight	106.326	-0.493	0.135	0.189
polar surface area	32.87	-0.448	-0.152	-0.330
rotatable bond	2.220	-0.421	-0.044	-0.061
chiral center	1.719	-0.132	-0.407	0.733
ring count	0.975	-0.233	0.438	0.499
standard deviation		**1.865**	**1.210**	**1.064**
component variance		**0.435**	**0.183**	**0.142**
cumulative variance		**0.435**	**0.618**	**0.760**

Figure 2.

Principal component analysis (PCA) of physicochemical property space for ChemDiv (red), DOS (green), and SCUBIDOO (blue) illustrates similarities and differences between these collections. A) Compounds from each of the three collections were projected onto the first two principal components, each of which is also represented by its marginal distribution on a per-collection basis. B) Loading levels for the first three principal components (>75% cumulative variance) on the input features.

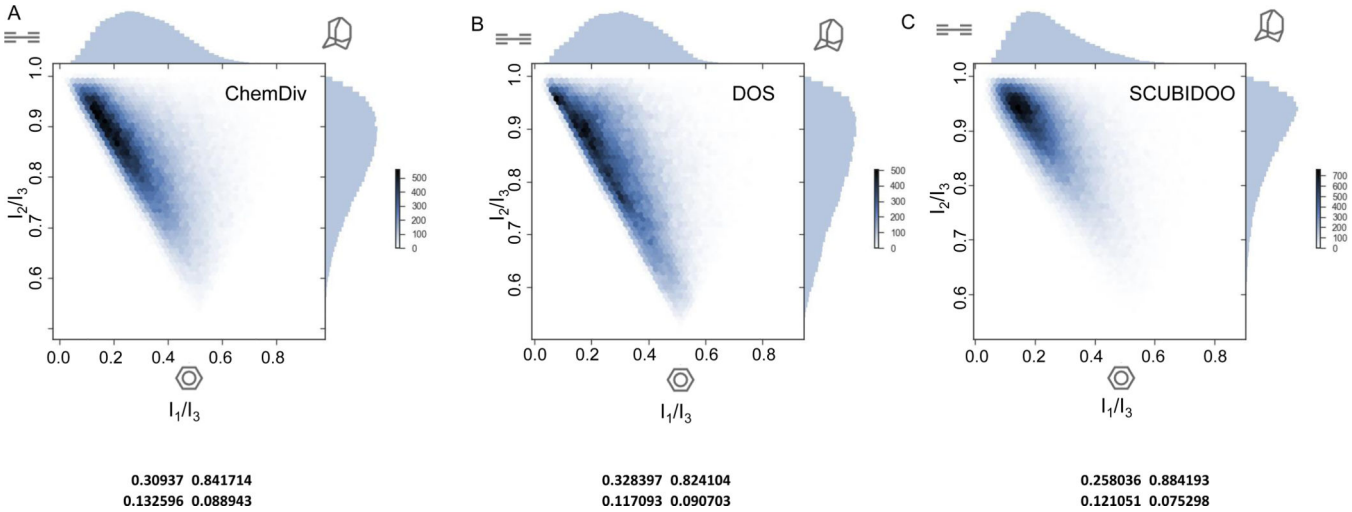
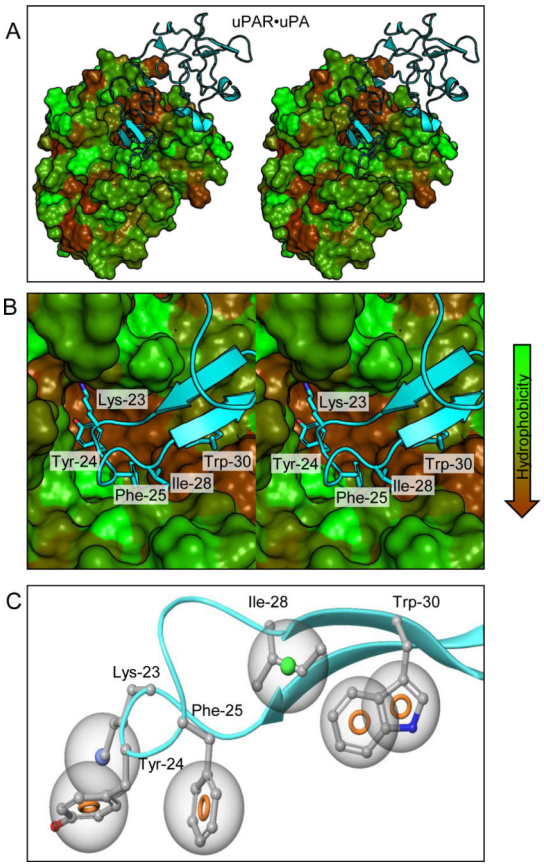


Figure 3.

Principal moments of inertia (PMI) illustrate the shape diversity of three compound collections. The top left-hand corner represents a linear molecule (e.g., diacetylene), the top right-hand corner represents a spherical molecule (e.g., adamantane) and the bottom corner represents a disc-like molecule (e.g., benzene).



D

ChemDiv	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Lys-23	72	11	0	12	1	0
Tyr-24	13370	3118	11	64	100	0
Phe-25	17161	3040	11	427	99	0
Ile-28	784	145	11	0	0	0
Trp-30	0	0	0	0	0	0
total	31387	3157	11	503	100	0

large small

E

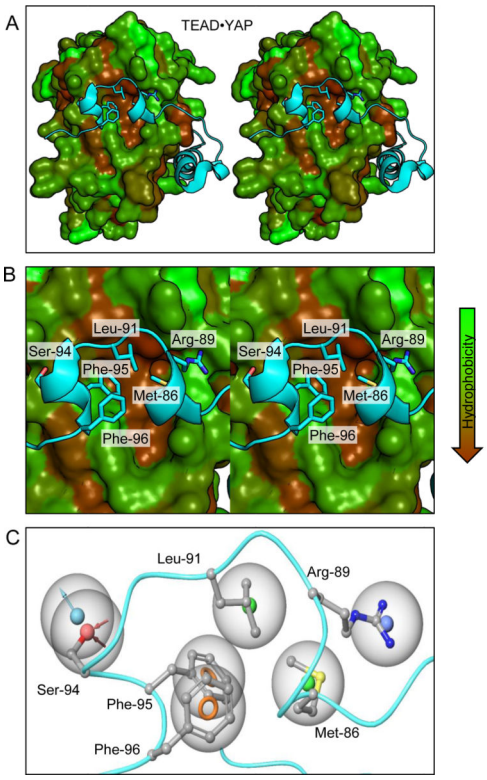
DOS	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Lys-23	1094	161	2	13	6	0
Tyr-24	8579	750	7	104	31	0
Phe-25	6304	541	6	277	35	0
Ile-28	1674	246	8	2	0	0
Trp-30	1559	206	7	1	0	0
total	19210	952	10	397	36	0

F

SCUBIDOO	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Lys-23	773	184	8	97	64	4
Tyr-24	7618	954	8	38	85	4
Phe-25	9531	918	8	361	61	4
Ile-28	70	10	0	0	0	0
Trp-30	0	0	0	0	0	0
total	17992	1033	8	496	105	4

Figure 4.

The protein complex of uPAR and uPA peptide. A) uPAR is shown as a surface colored by hydrophobicity, with more hydrophobic residues in brown and more hydrophilic residues in green. The uPA peptide is shown as a cyan ribbon. B) The uPAR·uPA complex and the five hot spots (side chains in stick representation) on uPA (cyan ribbon). C) Pharmacophore features for uPA are shown as small colored spheres: positive charge (Lys-23 in dark blue), hydrophobic (Ile-28 in green), and aromatic rings (Tyr-24, Phe-25, Trp-30 in tan). D) The matching count for compounds in ChemDiv that overlap with one, two, or three distinct hot spots on uPA using the whole collection and the top 1000 compounds (highest docking scores with MW ≤ 500 Da), respectively. E) The matching count for compounds in DOS that overlap with one, two, or three distinct hot spots on uPA using the whole collection and the top 1000 compounds, respectively. F) The matching count for compounds in SCUBIDOO that overlap with one, two, or three distinct hot spots on uPA using the whole library and the top 1000 compounds, respectively.



D

ChemDiv	whole collection				top 1000 (<500 Da)			
	1	2	3	4	1	2	3	4
Met-86	25855	7869	395	5	108	16	0	0
Arg-89	28	6	0	0	6	2	0	0
Leu-91	9890	4336	381	5	43	9	0	0
Ser-94	84	81	23	4	1	0	0	0
Phe-95	11108	5372	393	5	73	17	0	0
Phe-96	96	60	17	1	0	0	0	0
total	47061	8862	403	5	231	22	0	0

large small

E

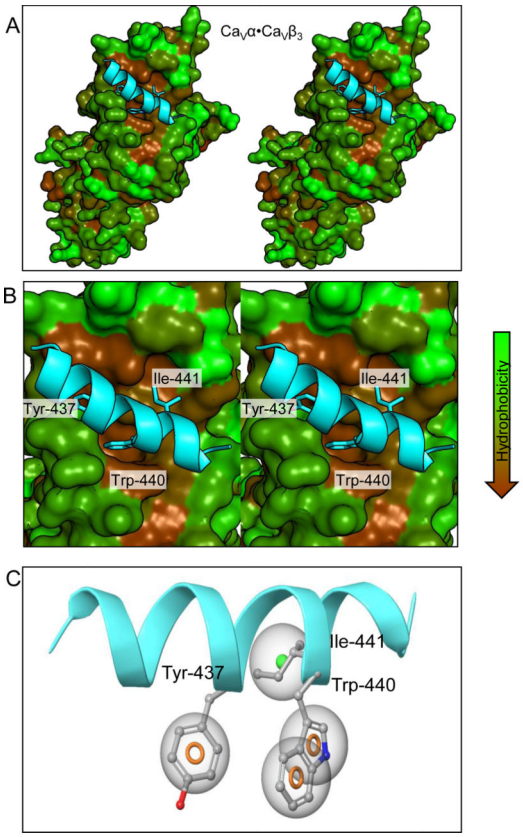
DOS	whole collection				top 1000 (<500 Da)			
	1	2	3	4	1	2	3	4
Met-86	29112	7746	399	3	117	83	6	0
Arg-89	2099	881	97	1	26	9	1	0
Leu-91	9501	3459	305	3	46	14	3	0
Ser-94	5481	3112	229	3	160	97	5	0
Phe-95	4580	3704	224	2	41	49	3	0
Phe-96	203	238	21	0	0	0	0	0
total	50976	9570	425	3	390	126	6	0

F

SCUBIDOO	whole collection				top 1000 (<500 Da)			
	1	2	3	4	1	2	3	4
Met-86	23848	4180	59	0	152	38	0	0
Arg-89	89	17	3	0	6	1	0	0
Leu-91	5674	2190	55	0	89	20	0	0
Ser-94	0	0	0	0	0	0	0	0
Phe-95	5735	2283	58	0	107	19	0	0
Phe-96	23	12	2	0	0	0	0	0
total	35369	4341	59	0	354	39	0	0

Figure 5.

The protein complex of TEAD and YAP peptide. A) TEAD is shown as a surface colored by hydrophobicity, with more hydrophobic residues in brown and more hydrophilic residues in green. The YAP peptide is shown as a cyan ribbon. B) The TEAD·YAP complex and the six hot spots (side chains in stick representation) on YAP (cyan ribbon). C) Pharmacophore features for YAP are shown as small colored spheres: hydrophobic (Met-86 and Leu-91 in green), positive charge (Arg-89 in dark blue), hydrogen-bond acceptor (Ser-94 in red), hydrogen-bond donor (Ser-94 in light blue), and aromatic rings (Phe-95, Phe-96 in tan). D) The matching count for compounds in ChemDiv that overlap with one, two, or three distinct hot spots on YAP using the whole collection and the top 1000 compounds (highest docking scores with MW ≤ 500 Da), respectively. E) The matching count for compounds in DOS that overlap with one, two, or three distinct hot spots on YAP using the whole collection and the top 1000 compounds, respectively. F) The matching count for compounds in SCUBIDOO that overlap with one, two, or three distinct hot spots on YAP using the whole library and the top 1000 compounds, respectively.



D

ChemDiv	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Tyr-437	26800	33684	2099	97	509	50
Trp-440	28235	35170	2099	226	529	50
Ile-441	4136	9026	2099	10	68	50
total	59171	38940	2099	333	553	50

large

small

E

DOS	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Tyr-437	20509	7980	476	270	104	2
Trp-440	25864	9767	476	327	90	2
Ile-441	11846	8953	476	62	108	2
total	58219	13350	476	659	151	2

F

SCUBIDOO	whole collection			top 1000 (<500 Da)		
	1	2	3	1	2	3
Tyr-437	19108	12144	40	99	478	2
Trp-440	22112	15168	40	301	544	2
Ile-441	3877	4202	40	2	72	2
total	45097	15757	40	402	547	2

Figure 6.

The protein complex of $Ca_V\beta$ and $Ca_V\alpha$ peptide. A) $Ca_V\beta_3$ is shown as a surface colored by hydrophobicity, with more hydrophobic residues in brown and more hydrophilic residues in green. The $Ca_V\alpha$ peptide is shown as a cyan ribbon. B) The $Ca_V\alpha$·$Ca_V\beta_3$ complex and the three hot spots (side chains in stick representation) on $Ca_V\alpha$ (cyan ribbon). C) Pharmacophore features for $Ca_V\alpha$ are shown as small colored spheres: hydrophobic (Ile-441 in green) and aromatic rings (Tyr-437, Trp-440 in tan). D) The matching count for compounds in ChemDiv that overlap with one, two, or three distinct hot spots on $Ca_V\alpha$ using the whole collection and the top 1000 compounds (highest docking scores with MW ≤ 500 Da), respectively. E) The matching count for compounds in DOS that overlap with one, two, or three distinct hot spots on $Ca_V\alpha$ using the whole collection and the top 1000 compounds, respectively. F) The matching count for compounds in SCUBIDOO that overlap with one, two, or three distinct hot spots on $Ca_V\alpha$ using the whole library and the top 1000 compounds, respectively.

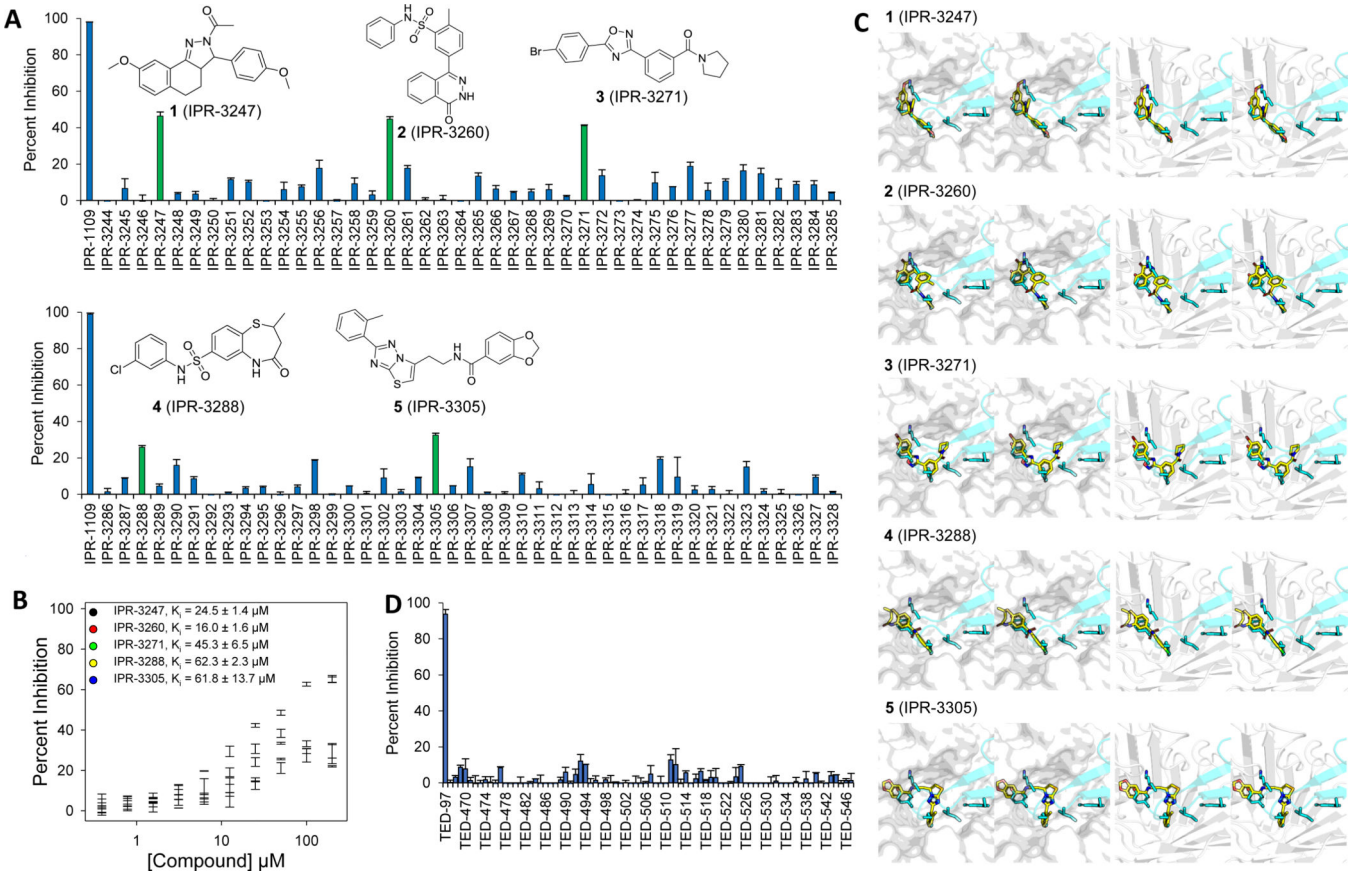


Figure 7.

A) A set of 85 compounds from virtual screening were tested at a single concentration of 50 μM in a fluorescence polarization (FP) assay for inhibition of uPAR·AE147 interaction. Structures are provided for compounds **1–5** (green bars) that were advanced to concentration-response experiments. B) Concentration-dependent FP assay measuring the inhibition of uPAR·AE147-FAM peptide interaction by compounds. At high concentrations, some compounds were insoluble and high-concentration data points were omitted from curve-fitting. C) Virtual screening binding modes of **1–5** (yellow sticks) in the uPAR·uPA binding pocket. uPAR is shown as a white surface and ribbons on the left and right, respectively. uPA is shown as a partially transparent cyan ribbon. The side chains of four interface residues on uPA are shown as a stick representation and colored cyan. D) A set of 81 compounds from virtual screening were tested at a single concentration of 50 μm for inhibition of a FP-based GST-TEAD4·Yap interaction assay. Compounds were selected to target the W-loop pocket on TEAD4 and were limited to be conformationally restrictive (i.e., 350 ≤ MW ≤ 500, RB ≤ 6, Alog*P* ≤ 4).