Exploring Key Orientations at Protein–Protein Interfaces with Small Molecule Probes

Eunhwa Ko,† Arjun Raghuraman,† Lisa M. Perez,‡ Thomas R. Ioerger,*§ and Kevin Burgess*†

†Department of Chemistry and ‡Laboratory for Molecular Simulation, Texas A & M University, Box 30012, College Station, Texas 77842, United States
§Department of Computer Science, Texas A & M University, College Station, Texas 77843-3112, United States

Supporting Information

ABSTRACT: Small molecule probes that selectively perturb protein–protein interactions (PPIs) are pivotal to biomedical science, but their discovery is challenging. We hypothesized that conformational resemblance of semirigid scaffolds expressing amino acid side-chains to PPI-interface regions could guide this process. Consequently, a data mining algorithm was developed to sample huge numbers of PPIs to find ones that match preferred conformers of a selected semirigid scaffold. Conformations of one such chemotype (1aaa; all methyl side-chains) matched several biomedically significant PPIs, including the dimerization interface of HIV-1 protease. On the basis of these observations, four molecules 1 with side-chains corresponding to the matching HIV-1 dimerization interface regions were prepared; all four inhibited HIV-1 protease via perturbation of dimerization. These data indicate this approach may inspire design of small molecule interface probes to perturb PPIs.

INTRODUCTION

Discovery of small molecules that perturb protein–protein interactions (PPIs) is often achieved by high-throughput screening (HTS), fragment- and structure-based strategies, molecular evolution of macrocycles, tethering target template click chemistry, and design of secondary structure mimics. However, the most prevalent method, HTS, gives disappointing hit rates relative to the cost and time expenditures involved, even if it is augmented by computational simulations based on matching virtual libraries with structural and physiochemical descriptors.

Compound libraries for HTS assembled to find small molecules that bind enzyme active sites, ion channels, and G protein-coupled receptors, and filtered for predicted oral bioavailabilities, may not be suitable for PPI targets; it has been suggested this is one reason for the poor hit rates. Despite this, there is no widely accepted notion of preferred small molecule chemotypes for these targets, except for small molecule mimics of ideal interface secondary structures.

These minimalistic mimics are comprised of nonpeptidic, semirigid skeletons that express amino acid side-chains, e.g., compound 1 (Figure 1). This led us to hypothesize a set of privileged small molecule chemotypes for perturbation of PPIs and how they could be applied. Specifically, favorable conformations of semirigid small molecules expressing amino acid side-chains could be compared with PPI interfaces to find the PPI that best matches the molecule; we call this concept Exploring Key Orientations (EKO).

GUIDELINES FOR A SET OF SMALL MOLECULE CHEMOTYPES TO PERTURB PROTEIN–PROTEIN INTERACTIONS

Expression of amino acid side-chains on semirigid small molecules is a valuable concept because interactions between interface side-chains dominate PPIs; however, in our view, secondary structure mimicry is not the overall end point because key interface regions are often formed from more than one, and/or from nonideal, secondary structure motifs. There are many PPI interface regions that cannot be mimicked by molecules that resemble one ideal secondary structure. Chemotypes for perturbing PPIs therefore should be based on comparing the orientations of the amino acid side-chains they project with those at protein–protein interfaces and not necessarily on secondary structure mimicry. Thus, the following structural design criteria were conceived for the chemotypes:

- synthetically accessible with any combination of at least three amino acid side-chains (e.g., Arg, Trp, His, etc.) to be incorporated onto a semirigid scaffold that has ...
- kinetically and thermodynamically accessible conformations (i.e., not too rigid) for induced fit to the protein binding partner but with ...
- only moderate loss of entropy on docking (i.e., the scaffold has only a few significant degrees of freedom that influence the side-chain orientations).
Properties like water solubilities, toxicities, cell permeabilities, shelf lives, etc., are important, but different, issues. Justification for selection of three side-chains as a starting point for these guidelines is as follows. In our estimation, two side-chain analogues would tend to have inadequate affinities and selectivities. Combinations of three side-chains were chosen because tripeptides frequently display high affinities and selectivities in cell biology (reviewed, e.g., RGD motifs). Four side-chain systems tend to be harder to prepare, and their allowed conformations probably would be too exclusive because this would involve matching eight coordinates (see below). The key parameters above are intended to be nonexclusive structural chemotype guidelines for identifying small molecules to perturb PPIs, and the EKO approach is specifically designed to work with these types of molecules.

**DEVELOPMENT OF THE EKO APPROACH**

Semirigid scaffolds presenting side-chains have multiple favorable core conformations. Implementation of EKO requires that side-chain orientations in these conformations be compared with projections of side-chains at protein–protein interfaces. If there is a good match in this comparison, then the small molecule might wholly or partially displace that protein from the PPI, thereby perturbing the interface. Complete dissociation at PPI interfaces is not essential because binding small molecules at PPIs may have biochemical ramifications even if they do not displace the protein binding partners (cf. allosteric binding).

Computational methods are required to achieve the levels of insight required to compare side-chain orientations of a small molecule with those at PPI interfaces. Fortunately, once these methods have been worked out, this facilitates data mining on a massive scale, i.e., systematic and sequential sampling of many structurally characterized PPIs.

The first step in developing EKO was to establish bases for comparisons (Figure 1). Amino acid Cα and Cβ coordinates are the best simple defining characteristic of side-chain orientations since Cβ–Cγ vectors and atoms downstream of these are relatively mobile and do not define overall direction. Preferred conformations of scaffolds that express only Cα and Cβ atoms, i.e., methyl-substituted ones, show how the molecular core tends to project any set of amino acid side-
chains. For example, favorable conformations of the Ala-Ala-Ala derivative 1aaa in medium with a continuous dielectric of 80 can be reduced to a set of six coordinates (3(Cα + Cβ)) that represent the scaffold bias when unperturbed by side-chain functionalities or explicit water molecule effects.

Orientations of side-chains in PPIs can be represented by Cα and Cβ coordinates from crystallographic data. Only side-chains that are physically able to interact with the protein binding partner need be considered, hence filters were devised. Thus only side-chains within X Å of a chain on the other protein are selected, where "X" is a user-defined parameter (set at 4 Å in this work). An interface residue is defined as a residue for which "any side-chain atom" of a protein comes within X Å of "any" non-H atom in the partner protein. For instance, the atoms Cβ, Cγ1, and Cγ2 of Val would be considered, as would the terminal –QH and Cβ of Tyr.

To implement EKO (Figure 2), preferred conformations of a semirigid small molecule with methyl side-chains must first be simulated; here quenched molecular dynamics (QMD)\textsuperscript{13,18} was used to do this, and only conformers within 3 kcal/mol of the most stable one identified were considered. This "3 kcal/mol cut-off" gave the following number of conformers for each stereoisomer of 1: LLL- (490), DLL- (490), LLD- (453), LLD- (512), LDD- (489), LLD- (511), DLD- (487), and DDD- (466). To save CPU, the analysis does not use all those conformations; instead, they are clustered into families with similar RMSDs based on Cα and Cβ coordinates, and poorly represented conformers from each family are removed (ca. 20–30% of the QMD conformers; see Table S1 and the surrounding discussion in the Supporting Information for the exact numbers and procedure).

The data mining algorithm developed here takes each preferred conformation as an input, expresses it as six coordinates \{3(Cα + Cβ)\}, and quantifies the "goodness of fit" of these on all combinations of three amino acid side-chains in all the structurally characterized PPI interfaces that are entered. Using the Texas A & M supercomputing facility, over 53,000 PPI interfaces corresponding to 15,736 structures were sampled in less than 6 h per 1aaa stereoisomer. For eight stereomers of 1aaa, EKO exposed a total of 391 unique PPI-interface regions where orientations of side-chains in preferred conformations matched those at interfaces with RMSDs ≤ 0.30 Å (Table S2, Supporting Information).

The output of this algorithm (Tables S3–S10, Supporting Information) is a relatively long list of interface regions that matched with preferred conformers of the featured compound. Data from mining a single isomer of 1 take too much space to show here, but Table 1 illustrates an EKO output for the 11 best "hit" interface overlays, and three others, from L,L,L-1aaa. Entries 1S, 16, and 23 are, in our view, biomedically significant PPI targets that would interest researchers considering synthesis of molecules with type 1 chemotypes.

In the procedure above, preferred conformations of the featured scaffold are calculated using truncated (Me-) side-chains, but they are overlaid on Cα and Cβ coordinates corresponding to combinations of particular interface amino acids. By making this comparison, EKO searches for intrinsic conformational biases of the scaffold with methyl side-chains that will be reinforced when the molecule binds a protein-binding partner in a hit PPI. Synergy occurs in these situations because the favored scaffold Cα–Cβ orientations coincide with the ways the rest of these side-chains are bound by the protein binding partner at the PPI interface.

EKO side-steps the most problematic issues encountered in simulations of small molecules interacting with protein surfaces by focusing on static interface regions in structurally characterized PPIs. Structural data clearly show the interface regions and the side-chain orientations circumventing the issue

---

**Table 1. Summary for Stereomer L,L,L-1aaa from EKO**

<table>
<thead>
<tr>
<th>Entry</th>
<th>PDB</th>
<th>Protein Homo- or Hetero-oligomer</th>
<th>RMSD (Å)</th>
<th>Residues (R1–R–R2–R3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1kn0</td>
<td>Rad52</td>
<td>0.14</td>
<td>H121–S119–D117</td>
</tr>
<tr>
<td>2</td>
<td>1nzc</td>
<td>nitrogenase</td>
<td>0.19</td>
<td>K145–D76–S257</td>
</tr>
<tr>
<td>3</td>
<td>1go0</td>
<td>trihydroxyanaphthalene reductase</td>
<td>0.23</td>
<td>P173–H122–V126</td>
</tr>
<tr>
<td>4</td>
<td>1j3u</td>
<td>aspartase</td>
<td>0.23</td>
<td>V236–T234–V232</td>
</tr>
<tr>
<td>5</td>
<td>1g7</td>
<td>TrwB</td>
<td>0.23</td>
<td>T352–D349–S346</td>
</tr>
<tr>
<td>6</td>
<td>1six</td>
<td>trypsin-ecotin</td>
<td>0.24</td>
<td>Me5–T83–L52</td>
</tr>
<tr>
<td>7</td>
<td>3pcb</td>
<td>3,4-PCD*</td>
<td>0.24</td>
<td>Q177–I75–K173</td>
</tr>
<tr>
<td>8</td>
<td>1cj</td>
<td>O-acetylserin sulfhydrolase</td>
<td>0.24</td>
<td>L268–S301–E303</td>
</tr>
<tr>
<td>9</td>
<td>26f</td>
<td>IS200 transposase</td>
<td>0.25</td>
<td>H60–V18–V107</td>
</tr>
<tr>
<td>10</td>
<td>1mtp</td>
<td>serpin (thermolipin)</td>
<td>0.26</td>
<td>Y200–T210–A218</td>
</tr>
<tr>
<td>11</td>
<td>1eef</td>
<td>heat-labile enterotoxin</td>
<td>0.26</td>
<td>T47–I39–E29</td>
</tr>
<tr>
<td>15</td>
<td>1hzh</td>
<td>AICAR Tfase\textsuperscript{a}</td>
<td>0.28</td>
<td>A218–L220–T222</td>
</tr>
<tr>
<td>16</td>
<td>3gpd</td>
<td>GAPDH\textsuperscript{c}</td>
<td>0.28</td>
<td>T228–M230–F232</td>
</tr>
<tr>
<td>23</td>
<td>1hpv</td>
<td>HIV-1 protease</td>
<td>0.29</td>
<td>L97–C95–I93</td>
</tr>
</tbody>
</table>

\textsuperscript{a}3,4-PCD: protocatechuate 3,4-dioxygenase.

\textsuperscript{b}AICAR Tfase: avian aminomimidazole-4-carboxamide ribonucleotide transformylase.

\textsuperscript{c}GAPDH: d-glyceraldehyde 3-phosphate dehydrogenase.
of how the small molecule and protein might flex to adapt to each other. EKO determines situations where the structurally characterized PPI and favored conformations of the small molecule have similar side-chain orientations: if there are no anomalies in the structural data, then those side-chains are sterically and physiochemically matched.

In unpublished work (in preparation), we have simulated all accessible conformers of eight well-known putative α-helical minimalist mimics and compared their side-chain orientations with those for ideal α-helical structures. None of the conformers of these compounds matched ideal α-helical structures with an RMSD of less than 0.30 Å. On the basis of this study, and other unpublished work, we suggest RMSD <0.30 Å is a stringent test for matching three side-chains in interface mimics.

Mining the eight stereoisomers of 1aaa on over 53,000 structurally characterized PPIs gave only two instances with RMSDs ≤0.30 Å in which the same three side-chains were implicated for different PPI targets. This observation suggests good selectivity is possible, probably via two origins. First, if L- and D-isomers of the protein encoded amino acids can be used at every position, then one particular combination is a 1 in 59,319 occurrence. Exactly the same reasoning would apply for tripeptides, but the second origin of selectivity, an entropy parameter, is more applicable to semirigid small molecules. Semirigid chemotypes access far fewer conformations than similar peptides, so they are statistically less likely to match at a PPI interface, hence the exclusivity of any hit found is higher.

Figure 3 shows six PPIs selected from the 391 that EKO overlaid on two “side-chain triplets” on the HIV-1 protease dimerization region: Leu97-Cys95-Ile93 (RMSD 0.29 Å; Figure 4) and Phe99-Leu97-Cys95 (0.33 Å). This dimer interface has hot spots at Cys95-Thr96-Leu97 at CH2SH, making the synthesis significantly more convenient. Mutation of Cys95 to Ala has little impact on the protease activity and has differential effects on HIV-1 protease activity.

Inhibition of HIV-1 protease was measured via a fluorescence-based assay, and then Zhang-Poorman kinetic analyses performed to determine if the inhibitors disrupt the dimer interface. Several intermediates in the syntheses of compounds 1 that have only two amino acid side-chains on the scaffold were tested, i.e., compounds 2. One of these (2a-H)
gave no measurable inhibition of the protease, while another four gave IC₅₀ values in the high micromolar range (2lai-tBu, 176.4 μM; 2li-H, 623.2 μM; 2lfl- Bu, S16.3 μM; 2fl-H, 418.7 μM; Table S11, Supporting Information). These data support the assertion made above that semirigid scaffolds bearing only two amino acid side-chains tend to give relatively poor binding affinities.

All four of the featured compounds 1 inhibited HIV-1 protease more effectively than the two side-chain intermediates 2 (Figure 4; 1lai-H: IC₅₀ = 3.7 ± 0.3 μM, Kᵢ = 0.38 ± 0.07 μM; 1fla-H: IC₅₀ = 46.5 ± 8.0 μM, Kᵢ = 0.93 ± 0.2 μM; 1lai-tBu: IC₅₀ = 111.1 ± 18 μM, Kᵢ = 19.4 ± 4.1 μM; 1fla-tBu, IC₅₀ = 54.9 ± 5.7 μM, Kᵢ = 21.0 ± 2.1 μM). Zhang–Poorman analyses indicated that all four compounds acted via dimerization disruption. The Kᵢ of the best hit, 1lai-H, is comparable with optimized HIV-1 dimerization inhibitors reported in the literature (Figure S13, Supporting Information).34–42 Known HIV-1 protease dimerization inhibitors tend to be long peptide sequences corresponding to two regions in the dimeric interface (Figure S13, Supporting Information). Conversely, the assayed compounds 1 are nonpeptidic small molecules with sequences corresponding to one set of three amino acids.

Inhibition by compounds that perturb HIV-1 protease dimerization tends to be inversely dependent on enzyme concentration.34,39 One compound, 1fla-tBu, was tested, and the inhibition of HIV-1 protease indeed decreased by a factor of 8 when a 5× higher HIV-1 protease dimer concentration was used (Supporting Information, Table S12). However, a similar inverse dependence on enzyme concentration also would be expected if the compounds formed aggregates that inhibited nonselectivity.43–45 This was a valid concern because the featured compounds contain hydrophobic side-chains (corresponding to alanine, isoleucine, leucine, and phenylalanine). Consequently, the possible aggregation issue was addressed by comparing inhibition in the presence and absence of 0.1% Triton X-100 (Table 2).43–45 Within experimental error, the initial velocities for a given compound do not change with and without Triton X-100, indicating the molecules are not promiscuous inhibitors acting through aggregation.

**Table 2. Comparison of Initial Velocities with and Without 0.1% Triton X-100**

<table>
<thead>
<tr>
<th>entry</th>
<th>compounds</th>
<th>normalized initial velocities with 0.1% Triton X-100 (fluorescence units/s)</th>
<th>initial velocities Without 0.1% Triton X-100 (fluorescence units/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1lai-tBu</td>
<td>5.84 ± 0.78</td>
<td>5.48 ± 0.46</td>
</tr>
<tr>
<td>2</td>
<td>1lai-H</td>
<td>5.28 ± 0.90</td>
<td>5.29 ± 0.49</td>
</tr>
<tr>
<td>3</td>
<td>1fla-tBu</td>
<td>3.42 ± 1.22</td>
<td>3.74 ± 0.69</td>
</tr>
<tr>
<td>4</td>
<td>1fla-H</td>
<td>2.21 ± 0.40</td>
<td>2.06 ± 0.31</td>
</tr>
</tbody>
</table>

*In buffer of 0.1 M sodium acetate, 1.0 M sodium chloride, 1.0 mM EDTA, 1.0 mM DTT, 10% DMSO, and 1.0 mg/mL of BSA, pH 4.7.

Control experiments showed that Triton X-100 at 0.1% (by volume) altered the initial velocities for cleavage of the fluorogenic substrate in the absence of compounds 1; consequently, a factor reflecting this perturbation was applied to give the normalized velocities indicated. Consequently, data in columns 3 and 4 can be compared, but relative rates for entries 1–4 cannot.
Verification that the featured compounds 1 acted as dimerization inhibitors was not possible via analytical gel ultracentrifugation since molecules 1 and HIV-1 protease have similar UV absorption profiles (e.g., at 280 nm; see Supporting Information) complicating detection. Consequently, qualitative cross-linking experiments analyzed by gel electrophoresis were undertaken instead.46 Figure 5 is an imaged gel showing the result of incubating the test compounds duplicated (see Supporting Information). A major limitation to the design of minimalist secondary structure mimics has been to generate structures that are selective for specific PPIs. A key innovation described here is to address that issue by using data mining for the reverse process: to find PPIs that match preferred small molecule conformations of the featured interface mimic. Moreover, the EKO approach can be applied irrespective of whether the small molecule resembles a secondary structure or not. It is the inverse of HTS where an assay is selected for a particular PPI and huge libraries are screened against it; consequently, EKO is chemistry-driven, whereas HTS and the other approaches currently used focus on the protein target selected. As far as we are aware, EKO is the first data mining approach to match PPIs with probes via virtual affinity selection from a huge PPI library using specific small molecule baits.

Researchers who are currently focused on one particular PPI may apply the EKO approach to single specific structurally characterized PPI targets; the chances of finding a good match are less than if it is coupled with data mining, but this approach can be rewarding (unpublished data). We believe the EKO method may lead chemists to identify PPIs that are most likely to be perturbed by molecules they have designed, and managers who oversee early stage drug discovery may decide to devote a fraction of their HTS budget to explore the cost effectiveness of EKO for PPI targets. We are currently exploring the feasibility of performing the EKO process using a powerful desktop personal computer to avoid the need for expensive computing equipment, and the initial indications are that this is viable. Further developments of this kind should significantly expand the user-base for this compound. At a minimum, it is an idea-generation method to inspire the design of small molecule interface probes to perturb PPIs.

**CONCLUSION**

A major limitation to the design of minimalist secondary structure mimics has been to generate structures that are selective for specific PPIs. A key innovation described here is to address that issue by using data mining for the reverse process: to find PPIs that match preferred small molecule conformations of the featured interface mimic. Moreover, the EKO approach can be applied irrespective of whether the small molecule resembles a secondary structure or not. It is the inverse of HTS where an assay is selected for a particular PPI and huge libraries are screened against it; consequently, EKO is chemistry-driven, whereas HTS and the other approaches currently used focus on the protein target selected. As far as we are aware, EKO is the first data mining approach to match PPIs with probes via virtual affinity selection from a huge PPI library using specific small molecule baits.

Researchers who are currently focused on one particular PPI may apply the EKO approach to single specific structurally characterized PPI targets; the chances of finding a good match are less than if it is coupled with data mining, but this approach can be rewarding (unpublished data). We believe the EKO method may lead chemists to identify PPIs that are most likely to be perturbed by molecules they have designed, and managers who oversee early stage drug discovery may decide to devote a fraction of their HTS budget to explore the cost effectiveness of EKO for PPI targets. We are currently exploring the feasibility of performing the EKO process using a powerful desktop personal computer to avoid the need for expensive computing equipment, and the initial indications are that this is viable. Further developments of this kind should significantly expand the user-base for this compound. At a minimum, it is an idea-generation method to inspire the design of small molecule interface probes to perturb PPIs.

**CONCLUSION**

A major limitation to the design of minimalist secondary structure mimics has been to generate structures that are selective for specific PPIs. A key innovation described here is to address that issue by using data mining for the reverse process: to find PPIs that match preferred small molecule conformations of the featured interface mimic. Moreover, the EKO approach can be applied irrespective of whether the small molecule resembles a secondary structure or not. It is the inverse of HTS where an assay is selected for a particular PPI and huge libraries are screened against it; consequently, EKO is chemistry-driven, whereas HTS and the other approaches currently used focus on the protein target selected. As far as we are aware, EKO is the first data mining approach to match PPIs with probes via virtual affinity selection from a huge PPI library using specific small molecule baits.

Researchers who are currently focused on one particular PPI may apply the EKO approach to single specific structurally characterized PPI targets; the chances of finding a good match are less than if it is coupled with data mining, but this approach can be rewarding (unpublished data). We believe the EKO method may lead chemists to identify PPIs that are most likely to be perturbed by molecules they have designed, and managers who oversee early stage drug discovery may decide to devote a fraction of their HTS budget to explore the cost effectiveness of EKO for PPI targets. We are currently exploring the feasibility of performing the EKO process using a powerful desktop personal computer to avoid the need for expensive computing equipment, and the initial indications are that this is viable. Further developments of this kind should significantly expand the user-base for this compound. At a minimum, it is an idea-generation method to inspire the design of small molecule interface probes to perturb PPIs.

**ASSOCIATED CONTENT**

Supporting Information

Development of the algorithm; figures, tables, and discussion to accompany the data mining experiments; procedures for HIV-1 protease inhibition and the Zhang–Poorman assays; literature on HIV-1 protease dimerization inhibitors; structures of all compounds tested; comparison of initial velocities in HIV-1 protease inhibition with and without Triton X-100; UV spectrum of 11fa; procedures for enzyme cross-linking experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

**AUTHOR INFORMATION**

**Corresponding Author**

burgess@tamu.edu

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank Mr. Dongyue Xin for performing the tests for aggregation via enzyme kinetics experiments and gratefully acknowledge Dr. Larry Dangott and the Texas A&M University Protein Chemistry Laboratory for help with the cross-linking experiments. HIV-1 protease (Q7K) was kindly provided by Dr. Celia Schiffer at University of Massachusetts Medical School. The National Institutes of Health (GM087981) and The Robert A. Welch Foundation (A-1121) provided financial support, and the Texas A&M Supercomputing Facility (http://sct.tamu.edu/) provided computing resources for the research reported in this paper.

**REFERENCES**
