Published in final edited form as:

J Am Chem Soc. 2013 January 9; 135(1): 167–173. doi:10.1021/ja3067258.

Exploring Key Orientations at Protein-protein Interfaces with Small Molecule Probes

Eunhwa Ko¹, Arjun Raghuraman¹, Lisa M. Perez², Thomas R. Ioerger^{3,*}, and Kevin Burgess^{1,*}

¹Department of Chemistry, Texas A & M University, Box 30012, College Station, TX 77842

²Laboratory for Molecular Simulation, Texas A & M University, Box 30012, College Station, TX 77842

³Department of Computer Science, Texas A & M University, College Station, TX 77843-3112

Abstract

Small molecule probes that selectively perturb protein-protein interactions (PPIs) are pivotal to biomedical science, but their discovery is challenging. We hypothesized conformational resemblance of semi-rigid 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 semi-rigid scaffold. Conformations of one such chemotype (**1aaa**; all methyl side-chains) matched several biomedically significant PPIs, including the dimerization interface of HIV-1 protease. Based on 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 PPIs is often achieved by high throughput screening (HTS), fragment- and structure-based strategies,^{1–3} molecular evolution of macrocycles,⁴ tethering,² target template click chemistry,⁵ and design of secondary structure

Corresponding Author: burgess@tamu.edu.

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 HIV1 protease inhibition with and without Triton X100; UV spectrum of **1fla**; procedures for enzyme crosslinking experiments. This material is available free of charge via the internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

The authors declare no competing financial interests.

mimics.⁶ 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.^{7,8}

Compound libraries for HTS assembled to find small molecules that bind enzyme active sites, ion channels, and G-coupled protein receptors, and filtered for predicted oral bioavailabilities,^{9,10} may not be suitable for PPI targets; it has been suggested this is one reason for the poor hit-rates.¹¹ 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.^{12,13} These *minimalist mimics* comprise of non-peptidic, semi-rigid skeletons that express amino acid side-chains, *eg* compound **1**¹⁴ (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 semi-rigid 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 PPIs

Expression of amino acid side-chains on semi-rigid small molecules is a valuable concept because interactions between interface side-chains dominate PPIs,¹⁵ but, 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 non-ideal, 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 (*eg* Arg, Trp, His, etc) to be incorporated onto a semi-rigid scaffold that has
- kinetically and thermodynamically accessible conformations (*ie* not too rigid) for induced fit to the protein binding partner but with
- only moderate loss of entropy on docking (*ie* 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 analogs 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,¹⁶ *eg* 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 8 coordinates (*see below*). The key parameters above intended to be non-exclusive *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

Semi-rigid 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 does not displace the protein binding partners¹⁷ (*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, *ie* systematic and sequential sampling of many structurally characterized PPIs.

The first step in developing EKO was to establish bases for comparisons. Amino acid $C\alpha$ and $C\beta$ coordinates are the best simple defining characteristic of side-chain orientations since $C\beta - C\gamma$ vectors and atoms downstream of these are relatively mobile and do not define overall direction.¹³ Preferred conformations of scaffolds that express only $C\alpha$ and $C\beta$ atoms, *ie* 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 ($3 \times C\alpha$ and $C\beta$) that represent the scaffold bias when unperturbed by side-chain functionalities or explicit water molecules effects.

Orientations of side-chains in PPIs can be represented by $C\alpha$ and $C\beta$ 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 $X \text{ \AA}$ of a chain on the other protein are selected, where “X” is a user-defined parameter (set at $4 \text{ \AA}$ in this work). An interface residue is defined as a residue for which “any side-chain atom” of a protein comes within $X \text{ \AA}$ of “any” non-*H* atom in the partner protein. For instance, the atoms $C\beta$, $C\gamma 1$, and $C\gamma 2$ of Val would be considered, as would the terminal -*OH* and $C\beta$ of Tyr.

To implement EKO (Figure 2), preferred conformations of a semi-rigid small molecule with methyl side-chains must first be simulated; here quenched molecular dynamics (QMD)^{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), LDL- (453), LLD- (512), LDD- (489), DLD- (511), DDL- (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 $C\alpha$ and $C\beta$ 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 material for the exact numbers and procedure).

The data mining algorithm developed here takes each preferred conformation as an input, expresses it as six coordinates $\{3 \times (C\alpha + C\beta)\}$, 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 was sampled in less than 6 h per **1aaa** stereoisomer. For eight stereoisomers 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).

The output of this algorithm (Tables S3 – S10) 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** takes too much space to show here, but Table 1 illustrates an EKO output for the eleven best “hit” interface overlays, and three others (red), from L,L,L-**1aaa**. Entries 15, 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 $C\alpha$ and $C\beta$ 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 $C\alpha$ - $C\beta$ 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 shows the interface regions and the side-chain orientations circumventing the issue 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. Based on 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 semi-rigid small molecules. Semi-rigid chemotypes access far fewer conformations than similar peptides, 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 (**a – f**) selected from the 391 that EKO found to match preferred conformers of **1aaa**. Brief descriptions of each given in the Figure caption to illustrate why they are high-interest targets for medicinal chemistry. This set of only six targets contains ones that impact cancer, neurodegenerative diseases, and type 2 diabetes.

As far as we are aware, there is no precedent for the EKO approach. For instance, CAVEAT^{27–30} is operationally distinct and has different objectives. CAVEAT searches for molecular frameworks that can position functional groups in specific relative orientations in enzyme active sites by screening *multiple ligands for a specified protein cavity*, almost invariably to find enzyme inhibitors, and it does *not* use a protein-binding partner to guide

the selection of the ligands. Conversely, EKO uses one protein in a PPI as a template to identify small molecules that may bind the other protein; there is no parallel for this in CAVEAT. There is also nothing in CAVEAT to sequentially address different regions in PPI interfaces. The main similarity connecting CAVEAT and EKO is that they both involve data mining: CAVEAT mines different small molecules for fit cavities in proteins selected by the user while EKO searches different PPIs given a user-defined starting molecule.

VALIDATION OF EKO USING HIV-1 PROTEASE

We looked for one test case to validate the EKO method experimentally. EKO revealed preferred conformations of **1aaa** 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-spots¹¹ at Cys95-Thr96-Leu97-Asn98-Phe99.³¹ Mutation of Cys95 to Ala has little impact on the protease activity,³² and presumably on the dimerization energy too, so **1lai**-H and **1fla**-H were prepared¹⁴ where methyl side-chains substitute for CH₂SH, making the synthesis significantly more convenient.

Inhibition of HIV-1 protease was measured via a fluorescence-based assay,³³ then Zhang-Poorman kinetic analyses³⁴ were 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, *ie* compounds **2**. One of these (**2la**-H) gave no measurable inhibition of the protease, while another four gave IC₅₀ values in the high micromolar range (**2la**-^tBu, 176.4 μM; **2li**-H, 623.2 μM; **2fi**-^tBu, 516.3 μM; **2fi**-H, 418.7 μM; Table S11). These data support the assertion made above that semi-rigid scaffolds bearing only two amino acid side-chains tend to give relatively poor binding affinities.

All of the 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, K_i = 0.38 ± 0.07 μM; **1fla**-H: IC₅₀ = 46.5 ± 8.0 μM, K_i = 0.93 ± 0.2 μM; **1lai**-^tBu: IC₅₀ = 111.1 ± 18 μM, K_i = 19.4 ± 4.1 μM; **1fla**-^tBu, IC₅₀ 54.9 ± 5.7 μM, K_i = 21.0 ± 2.1 μM). Zhang-Poorman analyses indicated that all four compounds acted via dimerization disruption. The K_i of the best hit, **1lai**-H, is comparable with optimized HIV-1 dimerization inhibitors reported in the literature (Figure S13).^{34–42} Known HIV-1 protease dimerization inhibitors tend to be long peptide sequences corresponding to *two* regions in the dimeric interface (Figure S13). Conversely, the assayed compounds **1** are *non-peptidic* 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, Table S12). However, a similar inverse dependence on enzyme concentration also would be expected if the compounds formed aggregates that inhibited non-selectively.^{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, indicating the molecules are *not* promiscuous inhibitors acting through aggregation.

Verification that the featured compounds **1** acted as dimerization inhibitors was not conveniently possible via analytical gel ultracentrifugation since molecules **1** and HIV1 protease have similar UV absorption profiles (*eg* at 280 nm; see supporting) complicating detection. Consequently, qualitative crosslinking experiments analyzed by gel electrophoresis were undertaken instead.⁴⁶ Figure 5 is an imaged gel showing the result of incubating the test compounds (200 μ M) with the protease (10 μ M) for 2 h at room temperature, then with a crosslinking agent (BS3) for 20 min at 37 °C. Quantitation was not attempted because of overlapping bands in the gels, possibly due to unfolding and autolysis of the enzyme under these conditions; nevertheless, the result is clear: all four of the test compounds suppressed the amount of crosslinked dimer formed under these conditions.

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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 crosslinking 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://sc.tamu.edu/>) provided computing resources for the research reported in this paper.

References

1. Rees DC, Congreve M, Murray CW, Carr R. *Nature*. 2004; 3:660–672.
2. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, Wells JA. *Proc. Natl. Acad. Sci. U.S.A.* 2000; 97:9367–9372. [PubMed: 10944209]
3. Zinzalla G, Thurston DE. *Future Med. Chem.* 2009; 1:65–93. [PubMed: 21426071]
4. Gartner ZJ, Tse BN, Grubina R, Doyon JB, Snyder TM, Liu DR. *Science*. 2004; 305:1601–1605. [PubMed: 15319493]
5. Kulkarni SS, Hu X, Doi K, Wang H-G, Manetsch R. *ACS Chem. Biol.* 2011; 6:724–732. [PubMed: 21506574]
6. Perez de Vega MJ, Martin-Martinez M, Genzalez-Muniz R. *Curr. Top. Med. Chem.* 2007; 7:33–62. [PubMed: 17266595]
7. Sperandio O, Miteva MA, Villoutreix BO. *Methods Princ. Med. Chem.* 2011; 48:435–465.
8. Meireles LMC, Mustata G. *Curr. Top. Med. Chem. (Sharjah United Arab Emirates)*. 2011; 11:248–257.
9. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. *Adv. Drug Delivery Rev.* 1997; 23:3–25.
10. Miller JL. *Curr. Top. Med. Chem.* 2006; 6:19–29. [PubMed: 16454755]
11. Wells JA, McClendon CL. *Nature*. 2007; 450:1001–1009. [PubMed: 18075579]
12. Ko E, Liu J, Burgess K. *Chem. Soc. Rev.* 2011; 40:4411–4421. [PubMed: 21483946]
13. Ko E, Ling J, Perez LM, Lu G, Schaefer A, Burgess K. *J. Am. Chem. Soc.* 2011; 133:462–477. [PubMed: 21182254]
14. Raghuraman A, Ko E, Perez LM, Ioerger TR, Burgess K. *J. Am. Chem. Soc.* 2011; 133:12350–12353. [PubMed: 21780756]
15. Conte LL, Chothia C, Janin J. *J. Mol. Biol.* 1999; 285:2177–2198. [PubMed: 9925793]
16. Ung P, Winkler DA. *J. Med. Chem.* 2011; 54:1111–1125. [PubMed: 21275407]
17. Thiel P, Kaiser M, Ottmann C. *Angew. Chem. Int. Ed.* 2012; 51:2012–2018.
18. O'Connor SD, Smith PE, Al-Obeidi F, Pettitt BM. *J. Med. Chem.* 1992; 35:2870–2881. [PubMed: 1495018]
19. Greasley SE, Horton P, Ramcharan J, Beardsley GP, Benkovic SJ, Wilson IA. *Nat. Struct. Biol.* 2001; 8:402–406. [PubMed: 11323713]
20. Floyd RA, Kotake Y, Towner RA, Guo W-X, Nakae D, Konishi Y. *J. Toxicol. Pathol.* 2007; 20:77–92.
21. Chuang D-M, Hough C, Senatorov VV. *Annu. Rev. Pharmacol. Toxicol.* 2005; 45:269–290. [PubMed: 15822178]
22. Hartmann A, Hunot S, Michel PP, Muriel M-P, Vyas S, Faucheux BA, Mouatt-Prigent A, Turmel H, Srinivasan A, Ruberg M, Evan GI, Agid Y, Hirsch EC. *Proc. Natl. Acad. Sci.* 2000; 97:2875–2880. [PubMed: 10688892]
23. Pasinelli P, Houseweart MK, Brown RH Jr, Cleveland DW. *Proc. Natl. Acad. Sci.* 2000; 97:13901–13906. [PubMed: 11095709]
24. D'Amelio M, Cavallucci V, Cecconi F. *Cell Death Differ.* 2010; 17:1104–1114. [PubMed: 19960023]
25. Herman GA, Stein PP, Thornberry NA, Wagner JA. *Clin. Pharmacol. Ther.* 2007; 81:761–767. [PubMed: 17392725]
26. Poelji PD, Dang Q, Erion MD. *Drug Discovery Today: Therapeutic Strategies*. 2007; 4:103–109.
27. Lauri G, Bartlett PA. *J. Comput.-Aided Mol. Des.* 1994; 8:51–66. [PubMed: 8035213]
28. Sefler AM, Lauri G, Bartlett PA. *Int. J. Peptide Protein Res.* 1996; 48:129–138. [PubMed: 8872530]
29. Etzkorn FA, Guo T, Lipton MA, Goldberg SD, Bartlett PA. *J. Am. Chem. Soc.* 1994; 116:10412–10425.

30. Yang Y, Nesterenko DV, Trump RP, Yamaguchi K, Bartlett PA, Drueckhammer DG. *J. Chem. Inf. Model.* 2005; 45:1820–1823. [PubMed: 16309289]
31. Todd MJ, Semo N, Freire E. *J. Mol. Biol.* 1998; 283:475–488. [PubMed: 9769219]
32. Davis DA, Dorsey K, Wingfield PT, Stahl SJ, Kaufman J, Fales HM, Levine RL. *Biochemistry.* 1996; 35:2482–2488. [PubMed: 8652592]
33. Matayoshi ED, Wang GT, Krafft GA, Erickson J. *Science.* 1990; 247:954. [PubMed: 2106161]
34. Zhang ZY, Poorman RA, Maggiora LL, Heinrikson RL, Kezdy FJ. *J. Biol. Chem.* 1991; 266:15591–15594. [PubMed: 1874717]
35. Lee S-G, Chmielewski J. *ChemBioChem.* 2010; 11:1513–1516. [PubMed: 20575134]
36. Shultz MD, Bowman MJ, Ham Y-W, Zhao X, Tora G, Chmielewski J. *Angew. Chem. Int. Ed.* 2000; 39:2710–2713.
37. Bowman MJ, Chmielewski J. *Bioorg. Med. Chem.* 2009; 17:967–976. [PubMed: 18337105]
38. Bannwarth L, Rose T, Dufau L, Vanderesse R, Dumond J, Jamart-Gregoire B, Pannecouque C, De Clercq E, Reboud-Ravaux M. *Biochemistry.* 2009; 48:379–387. [PubMed: 19105629]
39. Davis DA, Brown CA, Singer KE, Wang V, Kaufman J, Stahl SJ, Wingfield P, Maeda K, Harada S, Yoshimura K, Kosalaraksa P, Mitsuya H, Yarchoan R. *Antiviral Res.* 2006; 72:89–99. [PubMed: 16687179]
40. Vidu A, Dufau L, Bannwarth L, Soulier J-L, Sicsic S, Piarulli U, Reboud-Ravaux M, Onger S. *ChemMedChem.* 2010; 5:1899–1906. [PubMed: 20936621]
41. Quere L, Wenger T, Schramm HJ. *Biochem. Biophys. Res. Commun.* 1996; 227:484–488. [PubMed: 8967903]
42. Breccia P, Boggetto N, Perez-Fernandez R, Gool MV, Takahashi M, Rene L, Prados P, Badet P, Reboud-Ravaux M, Mendoza JD. *J. Med. Chem.* 2003; 46:5196–5107. [PubMed: 14613322]
43. McGovern SL, Caselli E, Grigorieff N, Shoichet BK. *J. Med. Chem.* 2002; 45:1712–1722. [PubMed: 11931626]
44. McGovern SL, Shoichet BK. *J. Med. Chem.* 2003; 46:1478–1483. [PubMed: 12672248]
45. Ryan AJ, Gray NM, Lowe PN, Chung C-w. *J. Med. Chem.* 2003; 46:3448–3451. [PubMed: 12877581]
46. Zutshi R, Franciskovich J, Shultz M, Schweitzer B, Bishop P, Wilson M, Chmielewski J. *J. Am. Chem. Soc.* 1997; 119:4841–4845.

Figure 1.

$C\alpha - C\beta$ Vectors depict how side-chains project at PPI interfaces (**a** and **b**) and in preferred conformations of semi-rigid small molecules that express amino acid side-chains (**c–e**). These side-chain orientations can be depicted as sets of 6 coordinates in three-dimensional space and overlaid; goodness of fit may be expressed using RMSD (root mean square values). The protein-protein interface shown is a region from heat-labile enterotoxin B₅ (pdb: 1eef).

Figure 2.

Preferred conformers of small molecules are determined and the side-chain C_{α} and C_{β} coordinates for each are recorded. The algorithm determines the interface residues of each PPI in a database and sequentially matches each combination of 3 interface side-chains with the 3 side-chains of the small molecule, and outputs goodness of fit for every overlay as RMSDs.

Figure 3.

Selected PPIs that match preferred conformations of **1aaa**. **(a)** 5-Aminoimidazole-4-carboxamide ribonucleotide transformylase (AICAR Tfase: an enzyme in purine biosynthesis. Cancer cells depend on *de novo* purine biosynthesis, hence they are vulnerable to AICAR Tfase inhibitors).¹⁹ **(b)** Inducible Nitric Oxide Synthase (iNOS; implicated in inflammatory, autoimmune diseases, and cancers).²⁰ **(c)** D-GlycerAldehyde-3-Phosphate DeHydrogenase (GAPDH; implicated in neurodegenerative apoptotic cell death).²¹ **(d)** Caspase 3 (pivotal in apoptotic neural cell death in Parkinson's,²² ALS,²³ Huntington's, and Alzheimer's diseases).²⁴ **(e)** DiPeptidyl Peptidase DDP-4 (a serine protease target for treatment of type 2 diabetes).²⁵ **(f)** Fructose-1,6-BisPhosphatase (FBP: modulates gluconeogenesis and is another target for controlling type 2 diabetes).²⁶

Figure 4.

Disruption of the HIV-1 protease dimer. (a) Structure of HIV-1 protease (1hvp) showing the relevant dimer interface region boxed that was exposed by EKO. (b) Original overlay identified on C-terminal ICL of the protease (RMSD 0.29 Å). Zhang-Poorman analyses of (c) **1lai**-tBu (100 μM), **1fla**-tBu (50 μM) and uninhibited HIV-1 (*ie* substrate only, no inhibitor) and (d) **1lai**-H (5 μM), **1fla**-H (50 μM) and uninhibited HIV-1. V = initial velocity, and Eo = total enzyme concentration. Based on at least three runs for each data point.

Figure 5.
Formation of crosslinked HIV1 protease dimer is suppressed by all four featured compounds
1. This experiment was duplicated (see supporting).

Table 1Summary for stereomer L,L,L-**1aaa** from EKO

entry	PDB	proteins	RMSD (Å)	residues (R ¹ -R ² -R ³)
1	1kn0	Rad52	0.14	H121-S119-D117
2	1n2c	nitrogenase	0.19	K145-D76-S257
3	1g0o	trihydroxynaphthalene reductase	0.23	P173-H122-V126
4	1j3u	aspartase	0.23	V236-T234-V232
5	1gl7	TrwB	0.23	T352-D349-S346
6	1six	trypsin-ecotin	0.24	Me5-T83-L52
7	3pcb	3,4-PCD ^a	0.24	Q177-175-K173
8	1fcj	<i>O</i> -acetylserine sulfinydrylase	0.24	L268-S301-E303
9	2f4f	IS200 transposase	0.25	H60-V18-V107
10	1mtp	serpin (thermopin)	0.26	Y200-T210-A218
11	1eef	heat-labile enterotoxin	0.26	T47-I39-E29
15	1thz	AICAR Tfase ^b	0.28	A218-L220-T222
16	3gpd	GAPDH ^c	0.28	T228-M230-F232
23	1hvp	HIV-1 protease	0.29	L97-C95-I93

^a3,4-PCD: protococatechuate 3,4-dioxygenase.^bAICAR Tfase: avian aminoimidazole-4-carboxamide ribonucleotide transformylase.^cGAPDH: D-glyceraldehyde 3-phosphate dehydrogenase.

Table 2Comparison of initial velocities with and without 0.1 % Triton X-100^a.

entry	compounds	normalized initial velocities ^b with 0.1 % Triton X-100 (fluorescence units/sec)	initial velocities without 0.1 % Triton X-100 (fluorescence units/sec)
1	1 ai- ^t Bu	5.84 ± 0.78	5.48 ± 0.46
2	1 ai-H	5.28 ± 0.90	5.29 ± 0.49
3	1 fa- ^t Bu	3.42 ± 1.22	3.74 ± 0.69
4	1 fa-H	2.21 ± 0.40	2.06 ± 0.31

^aIn 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 BSA, pH 4.7.

^bControl 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.