Purposely engineered drug-target mismatches for entropy-based drug optimization

Ariel Fernández[1,2,*], Christopher Fraser[2] and L. Ridgway Scott[2]

[1] Instituto Argentino de Matemática "Alberto P. Calderón", CONICET, Saavedra 15, Buenos Aires 1083, Argentina

[2] Department of Computer Science, The University of Chicago, Chicago, IL 60637

* Corresponding author. Phone 608 609 4836; e-mail: ariel@uchicago.edu

Proteins are dynamic objects, often undergoing significant structural change and reducing their conformational possibilities upon binding to a ligand. Thus, unless dynamic information is incorporated, structure-based drug design becomes of limited applicability. Even within a dynamic approach, a rarely visited scenario arises as proteins increase their entropy content upon binding by locally enhancing conformational exploration in the complex. We argue that this binding mode is of primary importance in drug development, since it allows for drugs that are not optimized in the conventional way but feature mismatches with the target, suggesting a new class of molecular design based on entropy optimization. This possibility is illustrated in this opinion piece, which advocates the exploitation of dynamic information for drug design.

Revisiting the current paradigm in structure-based drug design

Structure-based drug design remains a challenging endeavor in molecular targeted therapy [1-4]. According to a current paradigm, the designer goal is to maximize the number and quality of favorable interactions across the drug-target interface [4]. This strategy amounts to an enthalpy-based optimization of the affinity [4], since favorable intermolecular interactions lower the enthalpy content of the drug-target complex. In thermodynamic terms, this strategy aims at minimizing the binding free-energy change, $\Delta G_{bind} = \Delta H_{bind} - T\Delta S_{bind}$ (G=Gibbs free energy, H=enthalpy, T=absolute temperature, S=entropy) by minimizing ΔH_{bind}, while assuming a relatively invariant entropic cost ($-T\Delta S_{bind}$) for the drug-ligand association. The latter assumption may be valid for relatively rigid targets but is unlikely to hold for flexible targets. In such cases, enthalpy reductions arising from favorable intermolecular contacts introduce conformational constraints in the target protein, with an entropic penalty ($-T\Delta S_{bind} > 0$) that compensates for the enthalpy decrease, thereby reducing the drug affinity. Thus, flexible targets pose a harder optimization problem, as the association free-energy minimization becomes mini-max problem, where two compensatory effects need to be optimized.

The limitations of the current strategy are apparent in the therapeutic interference with signaling pathways, where the molecular targets are typically kinases, the transducers of cell signals [5, 6]. Such targets present dynamic regions like the activation loop, the ATP-binding loop, etc., prone to undergo structural adaptation (induced fit) upon association [6, 7]. Thus, the affinity of a drug designed based solely on enthalpy optimization becomes compromised due to the compensatory entropy cost of the induced fit required

to make the favorable intermolecular interactions [8]. These observations suggest the need for a departure from enthalpy optimization when considering targets with flexible regions [9].

Indeed, many drugs of clinical importance in anticancer treatment that target kinases, such as *sunitinib* [10] or *sorafenib* [11], contain hydrophobic-polar mismatches at the drug-target interface within typical cutoff distances for electrostatic interactions [12], as shown in Fig. 1a-c. This fact is at odds with the prevailing emphasis on maximizing favorable intermolecular interactions. These mismatches indicate that these drugs were not enthalpically optimized. For that matter, their optimization was not the result of thermodynamic analysis. In fact, direct examination of the patent trails associated with their discovery (for example, US patents numbers 6573293, 7125905 and 7211600 for *sunitinib*) point to an affinity optimization obtained empirically by screening large compound libraries. For example, the earliest US-patent (#6,596,746) claiming the anticancer drug *dasatinib* covers an invention of 580 compounds. What transpires is that these compounds could not have been identified by rational design under the sole premise of enthalpy optimization. Rather, they were selected from an empirical high-throughput screening. Naturally, the therapeutic agent is ultimately selected beyond structure-related affinity considerations, as bioavailability, deliverability, safety, metabolic profile and PD/PK attributes are necessarily factored into the final optimization process.

This view is reinforced by the telling example of compound PD 173955, with a higher affinity than the well-known anticancer drug *imatinib* for the latter's primary target [13].

As shown below, compound PD173955 does not incorporate improvements in terms of enthalpy optimization relative to *imatinib*. In fact, with its hydrophobic-polar mismatches and less hydrogen bonding than *imatinib* across the drug-target interface, the PD173955 is not enthalpically optimized.

These observations introduce the problem of how to rationalize high (nanomolar) affinities of mismatched drugs. As shown and advocated in this opinion, these paradoxically high affinities require that we introduce novel design principles based on dynamic considerations [14]. Dynamic information is often encoded in the induced-fit state of a protein-ligand complex [6, 7] and often overlooked in the structural analysis. The history of this crucial oversight is long but poorly documented, with few references devoted to the subject [6, 15-17]. Our observations raise the possibility that many compounds with pharmacological potential could be optimized *vis-a-vis* the conformational entropy of the target. To our knowledge this fact remains largely unreported. This could be due to the fact that the pharmaceutical industry has a vested interest in not disclosing successful methodology, or simply because the novel approach remains largely unknown, although the entropy-driven binding of a fluorine-substituted form of the HTLV-1 Tax peptide is entirely compatible with our explanation [17]. Two molecular indicators of conformational entropy optimization - generally overlooked in structural analysis - will be reviewed and illustrated: a) mismatches across the protein/ligand interface and b) induced fit in drug-target complexes.

In general, conformational entropy optimization can be readily contrasted against more conventional cases of drug compounds with low solubility fitting snuggly into nearly rigid targets [1-4]. In such cases, the binding entropy changes are mostly associated with favorable solvent displacements and enthalpic optimization justifiably prevails as the dominant design strategy. However, the possibility of transcending enthalpic optimization by lowering the entropic cost of the drug-target association should be evaluated. This may well require considering the increasingly accessible dynamic information from NMR relaxation studies on induced conformational dynamics [15, 16]. We argue that this type of information may enable an entropy-optimization strategy, whereby the binding entropic cost is minimized by a judicious control of localized structural change. The molecular prototypes likely to emerge from this strategy share common features with the therapeutic agents analyzed in this opinion: they bind their targets loosely rather than snuggly, in order to enable hydration of polar groups occluded upon binding. This entropy-boosting strategy is reinforced by drug design constraints imposed to minimize the number of rotational degrees of freedom in the compound, so as to minimize the ligand entropy loss upon binding [18].

Drug-target mismatches as molecular indicators of entropy-optimized drugs

Fig. 1 illustrates a largely overlooked feature of drug-target complexes, namely, hydrophobic-polar mismatches across the drug-target interface. These mismatches arise from the hydration demands of the drug C-F and C-Cl polarized bonds or effective dipoles [19], unmet as the dipoles are brought within close proximity to apolar residues upon drug-target association (Fig. 1). These molecular indicators are not dealt with in the

standard structural rationale for drug affinity since they are at odds with prevailing enthalpy-based design optimization aimed at maximizing the number of favorable intermolecular contacts. Yet, basic thermodynamic considerations on drug-target binding (box 1) derived from Molecular Dynamics (MD) simulations [20-22] and rigorous statistical-mechanical analysis of the protein-water interface [23] enable us to resolve what appear to be "anomalously high affinities". It should be emphasized that such affinities were obtained experimentally [13] and later found to be in good quantitative agreement with our computed values [24]. To illustrate this phenomenon, we shall focus on the higher affinity of PD173955 for the Bcr-Abl kinase when compared with that of *imatinib*, purposely designed to be the primary Bcr-Abl kinase inhibitor [25]. Our analysis of the PD173955-(Bcr-Abl) complex (PDB.1M52) reveals mismatches between the drug chlorine atoms and protein side chains that frame their microenvironment upon association (Fig. 1d-e). This fact, coupled with the absence of favorable electrostatic interaction along the drug-target interface, leads us to infer that drug-target mismatches enhance the conformational entropy of the target protein as a means to enable hydration of the drug polar moieties and that this reduction in the binding entropy cost enhances the drug affinity. The expansion in conformational exploration upon ligand binding exposes the organic halogens to solvent, so their hydration requirements may be fulfilled. This induced dynamic state in the target protein favorably increases the entropy content, thereby increasing affinity and is actually corroborated by thermodynamic analysis. If in fact drug-target mismatches represent an entropically improved design, we would expect an unfavorable increase in the entropic term of the binding free energy, $-T\Delta\Delta S_{bind}>0$, when the two chlorines are replaced for hydrogens, thus removing the mismatches (Fig.

1d-e). This is indeed the case. A thermodynamic calculation at 200ns-equilibration time (box 1) for the initial and final state of the drug-target association process gives the following parameters:

PD173955+(Bcr-Abl) kinase association (PDB.1M52, T=303K):

ΔG_{bind} = -22.5 kcal/mol; ΔH_{bind} = -54.3 kcal/mol; $-T\Delta S_{bind}$ = 31.8 kcal/mol

Drug-target association resulting after chlorine→hydrogen substitution (x2) in PD173955: ΔG_{bind} = -14.5 kcal/mol; ΔH_{bind} = -54.0 kcal/mol; $-T\Delta S_{bind}$ = 39.6 kcal/mol

imatinib+(Bcr-Abl) kinase association (PDB.1IEP, T=303K):

ΔG_{bind} = -19.1 kcal/mol; ΔH_{bind} = -68.4 kcal/mol; $-T\Delta S_{bind}$ = 49.3 kcal/mol

Previous validation against experimental values of PD173955 anomalous affinity [13, 24] instills confidence in our computations that are adopted here as guidelines for the entropy-based design strategy. The computational results provide the key to solve the paradoxical high affinity of the PD173955 compound: The mismatches that occur at the drug-target interface upon binding promote an increase in the entropy content of the complex, thereby enhancing affinity. Thus, PD173955 binds to its target with higher affinity in spite of the fact that it makes only two intermolecular hydrogen bonds with Bcr-Abl kinase [13], while *imatinib* makes six [25]. The paradox is resolved by taking into account that PD173955 is an entropically optimized compound capable of boosting

the entropy content of its target to promote its own favorable hydration, while *imatinib* is not entropically optimized. The "PD173955 vs. *imatinib*" paradox and its resolution suggest a novel paradigm in drug design.

Sunitinib and *sorafenib*: two dynamically optimized anticancer agents

In this section we show that important anticancer drugs that interfere with signaling pathways controlling the fate of tumor cells are actually entropy-based optimal or suboptimal designs that do not result from prioritization of favorable intermolecular contacts. In accord with the previous arguments, we provide a design rationale for the hydrophobic-polar mismatches (Fig. 1a-c) between kinase inhibitors *sunitinib* and *sorafenib* and their respective targets, C-Kit and p38 Map kinase [10, 11].

Chemical intuition reinforced by thermodynamic computation (box 1) [6, 20-23] suggest that these mismatches promote exploration of conformation space by the target protein. Our rationale is that expansion of the accessible conformation space in the bound protein state is necessary to enable proper hydration of the otherwise occluded ligand polar moieties. This conjecture is corroborated by thermodynamic computations (200ns equilibration for all states), revealing the adverse impact on affinity associated with halogen→hydrogen substitution that results in mismatch removal:

Sunitinib+Kit kinase association (PDB.3G0E, T=303K):

ΔG_{bind} = -19.4 kcal/mol; ΔH_{bind} = -54.6 kcal/mol; $-T\Delta S_{bind}$ = 35.2 kcal/mol

Drug-target association resulting after fluorine→hydrogen substitution in *sunitinib*:

ΔG_{bind} = -12.6 kcal/mol; ΔH_{bind} = -52.1 kcal/mol; $-T\Delta S_{bind}$ = 39.5 kcal/mol

sorafenib+p38 Map kinase association (PDB.3GCS, T=303K):

ΔG_{bind} = -23.1 kcal/mol; ΔH_{bind} = -46.4 kcal/mol; $-T\Delta S_{bind}$ = 23.3 kcal/mol

Drug-target association resulting after fluorine→hydrogen substitutions (x4) in *sorafenib*:

ΔG_{bind} = -3.0 kcal/mol; ΔH_{bind} = -35.8 kcal/mol; $-T\Delta S_{bind}$ = 32.8 kcal/mol

The induced dynamic state associated with the drug-target mismatches is favorable and enhances the affinity of the drug. Thus, as the drug halogens are substituted for hydrogens, we obtain $\Delta\Delta G_{bind}>0$ and $\Delta\Delta S_{bind}<0$, reflecting an unfavorable transformation. These effects would not surface in standard structural underpinnings of the drug affinities. They suggest a saddle-point type optimization, where a boost in target conformational entropy is balanced against the competing tendency to make as many favorable intermolecular contacts as possible.

Controlling induced dynamic states through molecular design

This section is devoted to show how local unfolding of the target can be controllably induced by drug redesign and exploited to improve the drug affinity. Guided by the chemical intuition gained from the previous analysis, we propose and justify a strategy for the redesign of *imatinib* aimed at controlling its specificity towards one of its primary

cancer targets, the C-Kit kinase [26]. As previously shown [7], the motivation for discriminating C-Kit from the other primary *imatinib* target, Bcr-Abl kinase [25, 27] arises from the need to achieve higher safety in the treatment of the gastrointestinal stromal tumor (GIST). This safety is achieved by nanomolar inhibition of C-Kit, free from the potential cardiotoxicity associated with Bcr-Abl kinase inhibition [28]. This level of specificity has been achieved by redesigning *imatinib* into a better protector of the backbone hydrogen bond Cys673-Gly676 in C-Kit. The protective role is achieved by incorporating a methyl in the terminal ring of *imatinib* (Fig. 2), a re-design named WBZ_4 [7]. Since the under-protected (solvent-exposed) hydrogen bond is unique to C-Kit and is well protected in Bcr-Abl kinase, WBZ_4 binds selectively to C-Kit. Thus, the motivation behind the structure-based re-design of *imatinib* into WBZ_4 is the need to improve drug safety by increasing specificity. The approach took into account a singular molecular feature that discriminates protein targets with a common fold: the extent of protection of backbone hydrogen bonds. This suggested the re-design strategy of engineering a compound that could contribute intermolecularly to protect from water attack a uniquely unprotected hydrogen bond in the target. On the other hand, according to our thermodynamic calculations, the same level of affinity ($\Delta\Delta G_{bind} \approx 0$) would be obtained if we replace the added protective methyl in WBZ_4 for the de-protector fluorine.

This surprising prediction of replacing a polar for a nonpolar group to achieve the same selective affinity awaits experimental confirmation and can be rationalized when we consider the 5ns snapshot of the stable conformation in the molecular dynamics. WBZ_4

stabilizes the weak backbone bond Cys673-Gly676 by excluding surrounding water (Fig. 2, lower panel), hence decreasing the free energy of the drug-target complex. If the bond were not disrupted by the fluorinated *imatinib* (upper panel) it would seal the space and preclude hydration of the C-F drug dipole [19]. The disruption is entropically favorable, yielding $\Delta\Delta S_{bind}>0$ as *imatinib* is replaced for its fluorinated variant, since it removes a conformational constraint (the backbone bond) in the nucleotide-binding loop. In fact, this entropy boost relative to *imatinib* overcompensates for the net enthalpic loss associated with the disruption of the Cys673-Gly676 hydrogen bond. Thus, $\Delta\Delta H_{bind} - T\Delta\Delta S_{bind} < 0$ as we compare binding free energy changes in the fluorinated variant with those of *imatinib*. The net effect of this induced increase in entropy is an enhanced affinity of the fluorinated compound to levels comparable to those of WBZ_4 ($\Delta\Delta G_{bind}\approx 0$). However, the fluorinated *imatinib* would be entropically optimized, while WBZ_4 is an enthalpic improvement of *imatinib*.

The redesign of *imatinib* based on the dynamic behavior of its target upon association can be further justified by examining other drug-target complexes where water-sealing intramolecular interactions are disrupted to satisfy the ligand hydration requirements. An illustration of this entropy-based binding mode is found as we examine the inhibitory action of *gefitinib*, a therapeutic agent in the treatment of non-small-cell lung cancer [29]. In the *gefitinib*-EGFR kinase complex (PDB.2ITY), hydration of the drug chlorine and fluorine are precluded by sealing intramolecular interactions: the Leu788-Lys745 backbone hydrogen bond and the Lys745-Glu762 salt bridge, respectively (Fig. 1f-g). Thus, *gefitinib* proves to be an entropically improved ligand, capable of inducing the

conformational expansion of its target EGFR kinase by disrupting the preformed electrostatic interactions upon binding in order to satisfy its hydration requirements. This favorable induced dynamic state is corroborated by computing the detrimental effect of halogen→hydrogen substitution on the binding thermodynamic parameters: $\Delta\Delta G_{bind}$=7.0kcal/mol; $-T\Delta\Delta S_{bind}$=9.3kcal/mol.

This example again highlights the importance of incorporating dynamic information on the induced fit into approaches to drug design.

Conclusions and future perspectives

The fact that protein dynamics constitutes an essential component in targeted drug design has been recognized [8, 30-34], and its importance in targeting flexible proteins has been emphasized [6, 9]. Yet, the induced fit or structural change that the protein undergoes upon ligand binding is difficult to predict or control, except in localized regions of the target structure [6], introducing serious limitations in drug design. In his regard, the scenarios for drug action featured in this opinion inspire a revision of current strategies for structure-based design. We highlight the need for a paradigmatic shift to harness the dynamic nature of induced states in therapeutic targets. Accordingly, we advocate for the implementation of a design concept based on conformational entropy optimization through a control of the induced state. This concept is illustrated in Fig. 3 where enthalpy-based and entropy-based improvements of a lead compound are represented. We note that enthalpy optimization may be adequate (and in fact the only

possibility) for a near-rigid target, undergoing minimal structural adaptation upon binding. An enthalpy-based improvement of the lead compound (Fig. 3, upper panel) would not significantly affect the binding entropy ($\Delta\Delta S_{bind} \approx 0$). On the other hand, in target proteins with flexible regions, drug-target mismatches may actually enhance the affinity by expanding the explorations of conformation space in the target protein in order to fulfill the hydration requirements of the ligand (Fig. 3, middle panel). This scenario reflects the entropy-based improvement so far overlooked and advocated in this opinion piece. In consonance with the prevailing paradigm, a protein with a flexible region could be also targeted through an enthalpy-based improvement of the lead (Fig. 3, lower panel). This route would not be as advantageous as the dynamics-based design due to enthalpy-entropy compensation: Making a favorable contact with the flexible region imposes conformational constraints that result in entropy decrease.

As shown in this opinion piece, many therapeutic drugs actually fulfill the design premises of entropy-based optimization, departing considerably from the usual dictates that require a maximization of favorable contacts with the target. The dynamic information needed for entropy-optimized design can be encoded to serve as guidance to the drug designer and such information is becoming increasingly accessible via NMR-relaxation analysis of protein-ligand complexes [15, 16, 31]. This dynamic information has been often overlooked because it is signaled by mismatches across the target-drug interface that do not fit with current design premises. Nonetheless, when examined in the right light, these singularities may herald the advent of a new and powerful drug-design strategy.

Acknowledgments

The authors thank Dr. Jianping Chen for his help with molecular dynamics computations. The research of A. F. was partially supported by the National Research Council of Argentina (CONICET), the Institute of Biophysical Dynamics and the Computer Science Department at The University of Chicago. L. R. S. acknowledges partial support from NSF grant DMS-0920960.

References

1. Greer, J. *et al.* (1994) Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* 37, 1035–54

2. Merz, K. M. *et al.* Editors (2010) *Drug design: Structure and ligand-based approaches*. Cambridge Univ. Press, New York.

3. Whitesides, G. M. and Krishnamurthy, V. M. (2005) Designing ligands to bind proteins. *Quart. Rev. Biophys.* 38, 385-395

4. Freire, E. (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today* 13, 869-874

5. Huse, M. and Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell* 109, 275-282

6. Fernández, A., *et al.* (2008) Taming the induced folding of drug-targeted kinases. *Trends Pharm. Sci.* 30, 66-71

7. Fernández, A. *et al.* (2007) An anticancer C-Kit inhibitor is re-engineered to make it more active and less cardiotoxic. *J. Clin. Invest*. 117, 4044-4054

8. Frederick, K. K. *et al.* (2007) Conformational entropy in molecular recognition by proteins. *Nature* 448, 325-330

9. Teague, S. J. (2003) Implications of protein flexibility for drug discovery. *Nature Revs. Drug Disc.* 2, 527-541

10. Gajiwala, K. S. *et al.* (2009) KIT kinase mutants show unique mechanisms of drug resistance to *imatinib* and *sunitinib* in gastrointestinal stromal tumor patients. *Proc. Natl. Acad. Sci. USA* 106, 1542-1547

11. Simard. J. R. *et al.* (2009) Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. *J. Am. Chem. Soc.* 131, 13286-13296

12. Mueller, K. *et al.* (2007) Fluorine in Pharmaceuticals: Looking beyond intuition. *Science* 317, 1881-1886

13. Nagar, B. *et al.* (2002) Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and *Imatinib* (STI-571). *Cancer Research* 62, 4236-4243

14. Peng J. W. (2009) Communication Breakdown: Protein dynamics and drug design. *Structure* 17, 319-320

15. Diehl, C. *et al*. (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J. Am. Chem. Soc.* 132, 14577-14589

16. Stone, M. J. (2001) NMR Relaxation Studies of the Role of Conformational Entropy in Protein Stability and Ligand Binding. *Acc. Chem. Res.* 34, 379-388

17. Piepenbrink, K.H. *et al.* (2009) Fluorine substitutions in an antigenic peptide selectively modulate T-cell receptor binding in a minimally perturbing manner. *Biochem. J.* 423, 353-361

18. Knox, A. J. *et al*. (2005) Considerations in compound database preparation: "hidden" impact on virtual screening results. *J. Chem. Inf. Model.* 45, 1908-1919

19. Kwon, O. *et al.* (2010) Hydration dynamic at fluorinated protein surfaces. *Proc. Natl. Acad. Sci. USA* 107, 17101-17106

20. Case, D. A. *et al.* (2005) The Amber biomolecular simulation programs. *J. Comp. Chem.* 26, 1668-1688

21. Wang, J. *et al.* (2004) Development and testing of a general Amber force field. *J. Comput. Chem.* 25, 1157-1174

22. Fernández, A. (2001) Conformation-dependent environments in folding proteins. *J. Chem. Phys.* 114, 2489-2502

23. Fernández, A. and Lynch, M. (2011) Nonadaptive origins of interactome complexity. *Nature*, in press, doi:10.1038/nature09992, published online May 18, 2011.

24. Crespo, A. and Fernández, A. (2008) Induced disorder in protein-ligand complexes as a drug design strategy. *Mol. Pharmaceutics (ACS)* 5, 430-437

25. Schindler, T. *et al.* (2000) Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* 289, 1938-1942

26. Mol, C. D. *et al*. (2004) Structural basis for the autoinhibition and STI-571 inhibition of c-Kit tyrosine kinase. *J. Biol. Chem.* 279, 31655-31663

27. Donato, N. J. and Talpaz, M. (2000) Clinical use of tyrosine kinase inhibitors: Therapy for chronic myelogenous leukemia and other cancers. *Clin. Cancer Res.* 6, 2965-2966

28. Force, T., *et al.* (2007) Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nature Revs. Cancer* 7, 332-344

29. Yun, C. H. *et al.* (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* 11, 217-223

30. Carroll M. J. *et al*. (2011) Direct detection of structurally resolved dynamics in a multiconformation receptor-ligand complex. *J. Am. Chem. Soc.* 133, 6422-6428

31. Daly, N. L. *et al*. (2011) NMR and protein structure in drug design: application to cyclotides and conotoxins. *Eur. Biophys. J.* 40, 359-370

32. Bajaj, C. *et al*. (2011) A dynamic data structure for flexible molecular maintenance and informatics. *Bioinformatics* 27, 55-62

33. Lill, M. A. and Danielson, M. L. (2011) Computer-aided drug design platform using PyMOL. *J. Comput. Aided Mol. Des.* 25, 13-19

34. Durrant, J. D. *et al*. (2010) Including receptor flexibility and induced fit effects into the design of MMP-2 inhibitors. *J. Mol. Recognit.* 23, 173-182

Figure Captions

Figure 1. *Examples of inhibitor-target mismatches as signals of entropically optimized drug design.* Side chain atoms within 4Å of designated inhibitor halogen or halogen-substituted atoms and involved in nearby electrostatic interactions, all colored according to element (N: blue; O: red; C/H: white, F: light blue; Cl: green; S: yellow). The 4Å radius adopted is a typical cutoff distance for electrostatic interactions [12]. **a-c**. Microenvironments around organic halogen atoms or halogen-substituted groups of anticancer drugs *sunitinib* (**a**) and *sorafenib* (**b, c**) in complex with C-Kit and p38 Map kinase, respectively. The microenvironments were determined from structural coordinates deposited in PDB files with accession codes 3G0E (*sunitinib*+C-Kit) and 3GCS (*sorafenib*+p38 Map kinase). Mismatches between the polar halogen groups and the surrounding apolar atoms increase the likelihood of conformational rearrangement of the protein upon binding with partial disruption of the microenvironment local structure. Note that the two oxygen atoms from water molecules 458 and 482 shown in (**c**) are solvating Asp168 and the main chain carbonyl of Ile147 (not shown), and are not interacting with the chlorine. The backbone Ile166-Ile84 hydrogen bond (HB) is a target for disruption by the fluorine-substituted sorafenib. Similarly, the backbone Phe811-Ala814 hydrogen bond in the target protein C-Kit displayed in (**a**) is particularly prone to disruption by sunitinib. The main chain carbonyl oxygen of Asp810, which is 3.2Å away from the sunitinib fluorine also results in an O-F polar-polar mismatch. **d-e**. Hydrophobic-polar mismatches at the interface between kinase inhibitor PD173955 and its target Bcr-Abl kinase (PDB.1M52). Hydrophobic residues frame the microenvironments of the two chlorine atoms preventing their hydration, with the

backbone Ile113-Val270 hydrogen bond being disrupted by one of the PD173955 chlorine atoms. **f-g**. Drug-target mismatches between *gefitinib* and its target protein EGFR kinase in the PDB complex with accession 2ITY. The drug chlorine (**f**) and fluorine (**g**) substituents possess hydration requirements that can only be met through disruption of the water-sealing preformed electrostatic interactions in the protein: the Leu788-Lys745 backbone hydrogen bond (**f**) and the salt bridge (SB) Lys745-Glu762 (**g**), respectively.

Figure 2. *Induced dynamic state promoted by entropy-optimized drugs upon association with their target proteins*. Stable configurations of the nucleotide-binding loop prevailing in the range 3-200ns within a 200s-MD-equilibration of complexes with *imatinib* variants bound to the Kit kinase. The ligand and protein in the initial associated state are aligned with *imatinib*/C-Kit kinase in the PDB complex with accession PDB.1T46. Due to higher hydration requirements, the fluorinated *imatinib* (fluorine position in dashed green line, upper panel) disrupts the sealing intramolecular hydrogen bond C673-G676 of the C-Kit loop, promoting an increase in conformational entropy that overcompensates for the enthalpic loss associated with hydrogen-bond disruption (computations in main text). The stable loop conformation reveals an intercalated web of water molecules between Cys673 and Gly676 and mediated by the fluorine, all indicative of the bond disruption. On the other hand, WBZ_4 (a methylated variant of *imatinib*) is protective of the C673-G676 bond whose stability increases upon complexation [7]. Hence, WBZ_4 is an enthalpically optimized version of *imatinib* while the fluorinated variant is a re-design based on entropy optimization, and yet both have comparable affinities for C-Kit ($\Delta\Delta G_{bind} \approx 0$).

Figure 3. *Qualitative thermodynamic assessment of the improvement of lead compounds guided by enthalpic and entropic optimization on rigid (upper panel) and flexible targets (middle and lower panel).* The upper panel describes enthalpic optimization (only likely alternative) for a rigid target, the middle panel describes entropic optimization for a flexible target, and the lower panel describes enthalpic optimization for a flexible target. While the latter strategy is widely used, it generates mutually opposing enthalpy-entropy compensatory effects as favorable intermolecular contacts impose conformational constraints on the target. This leads us to uphold an entropic optimization strategy for flexible targets, as illustrated in the middle panel.

Box 1. Thermodynamic computations for drug-target association

When localized, the conformational adaptation of a target protein bound to a ligand may be assessed by performing classical molecular dynamic (MD) simulations following the protocol in ref. [6]. The initial state consists of the crystal structure of the kinase-ligand complex in contact with a pre-equilibrated statistical reservoir of surrounding water molecules. During solvent-pre-equilibration the protein structure is maintained rigid. To identify the induced folding or induced unfolding, we additionally perform simulations of the uncomplexed kinases in their free state by *in silico* removal of the drug/ligand followed by 200ns-equilibration with the solvent [6]. MD simulations generate conformational changes based on atomic-level integration of Newton's equations defined by an empirical potential, commonly known as force field, that

determines the forces between every pair of atoms in the system. The time integration generates a trajectory in conformation space for the many-body system subject to global constraints arising from the fact that the number of particles, pressure and temperature should remain constant and the distances between covalently bonded atoms should remain constant. These simulations were performed using an Amber9 package [20, 21] generating 200ns equilibrating trajectories.

This dynamic analysis enables computation of the enthalpy change (ΔH_{bind}) and entropy change (ΔS_{bind}) associated with drug binding. The enthalpy change is estimated from the difference in energy content for all pair-wise interactions between final state (drug-protein complex in equilibrium with water) and initial state (free protein and free drug individually equilibrated with water) [6, 20, 21]. The parameter ΔS_{bind} requires a computation of the regions in each dihedral torsional angle of the protein and drug that are accessible as the molecular conformation changes along the trajectory. Thus, for each drug-target association, three 200ns-simulations are carried out, one to equilibrate the complex with the solvent, one to equilibrate the free drug and one to equilibrate free protein with the solvent. The set of three simulations is repeated 10 times with different solvent initial configurations to ensure convergence in the estimation of the accessible region in conformation space. System size ($N \sim 10^5$) and computation N^2-scalability limited our microstate exploration. The timespan 200ns was found sufficient to trap the systems in a free energy minimum, as signaled by a root mean square atomic displacement RMSAD<1Å over an additional 100ns-run. The parameter ΔS_{bind} breaks down into two components $\Delta S_{bind} = \Delta S_{conf} + \Delta S_{solv}$, where ΔS_{conf}, ΔS_{solv} denote the changes in conformational and solvent entropy, respectively. The first term is computed

as $\Delta S_{conf} = R ln[W_f/W_i]$, where $R=1.98 cal K^{-1} mol^{-1}$ is the universal gas constant, the indices *f, i* denote final and initial state respectively, and W_f, W_i, represent the products of the lengths of available regions for each torsional angle of the protein chain and drug in the final and initial state, respectively [22]. The solvent contribution ΔS_{solv} is obtained following the protocol described in [23], by determining the net gain in hydrogen-bonding coordination with nearby water molecules undergone by interfacial water molecules as they become displaced to bulk solvent upon ligand-target association. The interfacial tension or free energy cost of spanning protein-water and ligand–water interfaces is determined by the decrease in the number, g, of hydrogen bonds with nearby water molecules relative to the bulk expectation value g=4 [23].

Abbreviations:

Bcr: Breakpoint cluster region

Abl: V-abl Abelson murine leukemia viral oncogene homolog 1 (also ABL1)

C-Kit kinase: Mast/stem cell growth factor receptor tyrosine-protein kinase

EGFR: Epidermal Growth Factor Receptor

HTLV: Human T-cell Lymphotropic Virus

Map: Mitogen-Activated Protein

PD: Pharmaco-dynamics

PK: Pharmaco-kinetics

Figure 1a

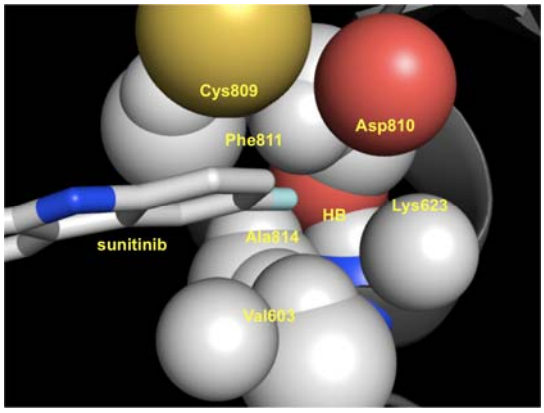


Figure 1b

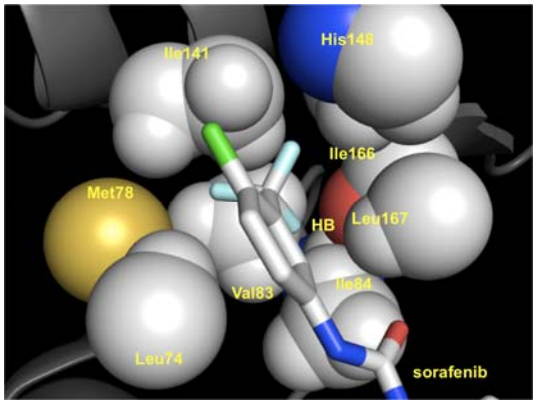


Figure 1c

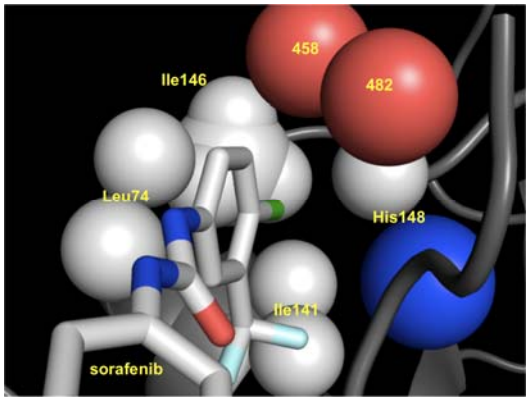


Figure 1d

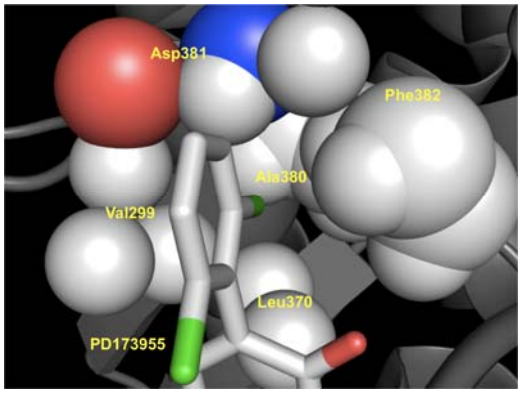


Figure 1e

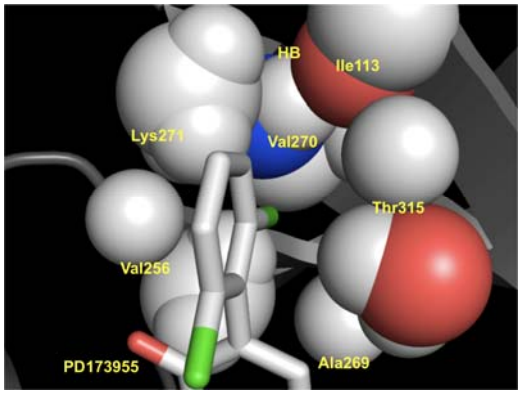


Figure 1f

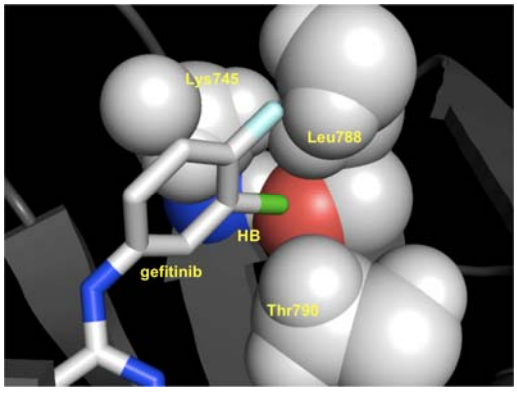


Figure 1g

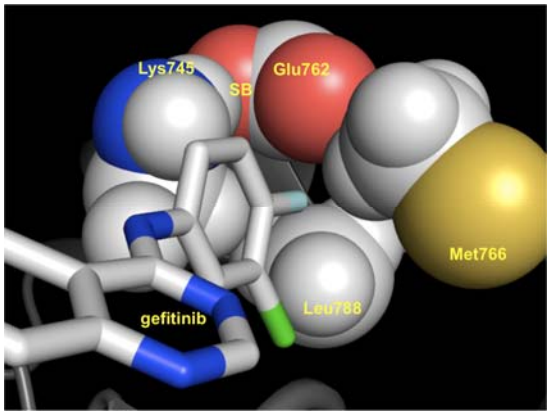


Figure 2

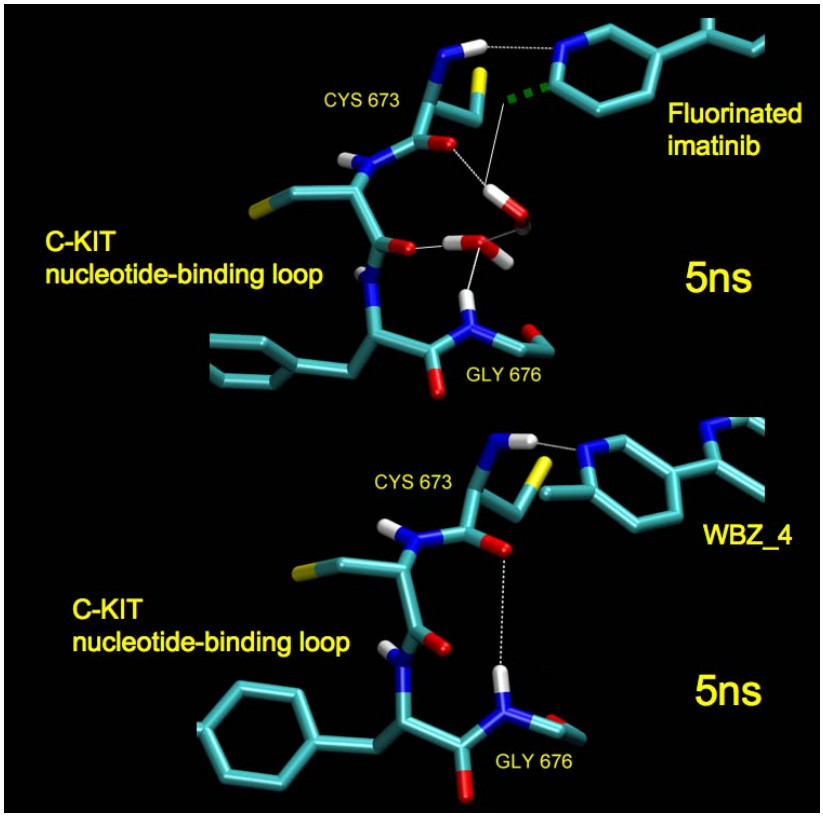
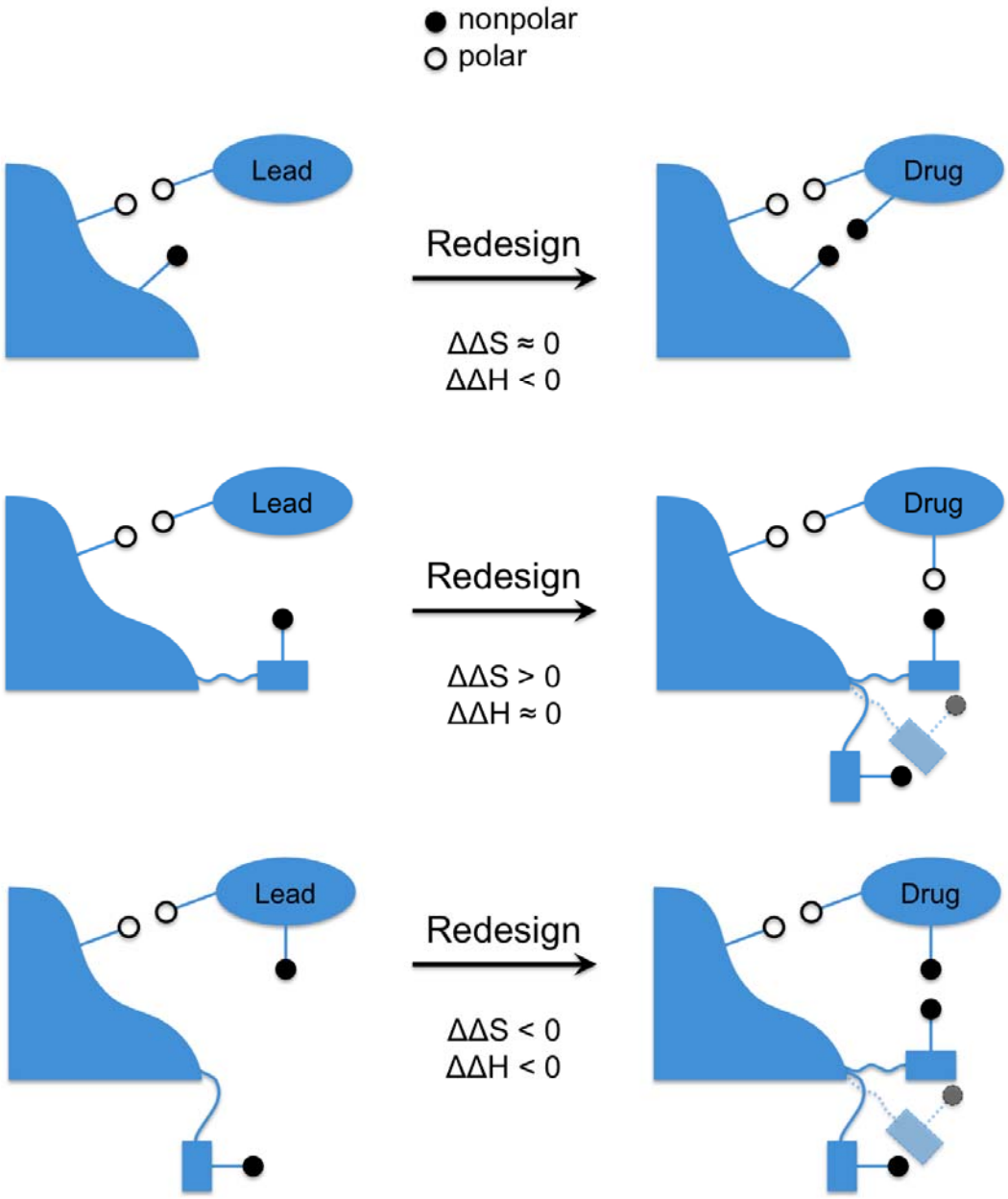


Figure 3


nonpolar
polar
Lead
Redesign
ΔΔS ≈ 0
ΔΔH < 0
Drug
Lead
Redesign
ΔΔS > 0
ΔΔH ≈ 0
Drug
Lead
Redesign
ΔΔS < 0
ΔΔH < 0
Drug