Expert Opinion

informa
healthcare

Ligand efficiency indices for effective drug discovery

Cele Abad-Zapatero
Abbott Laboratories, Department of Structural Biology, D-R46Y, AP-10, LL-07, Abbott Park, IL 60064-6098, USA

Successful drug discovery requires the optimization of a large number of variables ranging from strictly physicochemical parameters such as molecular weight to more complex parameters related to toxicity and bioavailability. Presently, structure-based methodologies influence many aspects of the drug discovery process from lead discovery to the final preclinical characterization. However, critical biological issues along the path to the market have diminished the impact and power of this methodology. The physicochemical properties of the novel chemical entities designed and guided by structural methods have become the subject of intense scrutiny from lead discovery to drug candidate. The idea of ligand efficiency (binding energy/non-hydrogen atoms) has recently emerged as a useful guide to optimize fragment and lead selection in the discovery process. More generalized concepts of ligand efficiency, related to efficiency per dalton and per unit of polar surface area, have also been introduced and will be discussed in the broader context. Preliminary results and trends obtained using ligand efficiencies as guides are reviewed and their future application to guide drug discovery will be discussed, as well as their integration into the structure-based drug design methods to make them more effective and numerically robust.

Keywords: computer-assisted drug discovery, fragment-based drug design, lead optimization, ligand efficiency indices, structure-based drug discovery

Expert Opin. Drug Discov. (2007) 2(4):469-488

1. Introduction

Successful delivery of novel pharmaceutical entities to the clinic is a multidimensional optimization process that involves elaborate medicinal chemistry, detailed *in vitro* enzymology, complex cell and organismic biology to assess effectiveness as well as toxicology and lengthy clinical studies. The impact of the advances in the related areas of molecular and structural biology, assay development, human genomics, proteomics and other related fields, although noticeable, is still far from dramatic. The number of variables to be taken into consideration seems to be expanding rather than diminishing. For a number of years now, it was hoped that powerful technologies derived from the physical sciences such as nuclear magnetic resonance (NMR) spectroscopy and X-ray diffraction applied to targets of interest would dramatically expedite the drug discovery process. Although this expectation has been fulfilled in certain cases, the effectiveness of the entire drug discovery process is under scrutiny [1]. A more streamlined and effective framework is needed to facilitate the drug discovery process, making it more efficient.

The concept of using protein structure information to accelerate the drug discovery process began to materialize in the pharmaceutical industry in the early 1980s. Although initially conceived only to expedite the lead optimization process, structure-based drug design (SBDD) now includes and supports virtually all aspects in the drug-discovery process [2,3]. The technologies of molecular biology, protein expression and purification have matured to the point of not being rate limiting in

providing target samples. Similarly, the latest technologies of macromolecular crystallography [4,5] combined with the routine access to third-generation synchrotron sources by the pharmaceutical industry [101] has expedited the data acquisition, structure determination and refinement of a large number of target–inhibitor complexes. There is still a significant gap between the purified samples of protein targets and the routine production of crystals of suitable quality for SBDD. The challenges are still mainly in terms of the unique properties of each particular human target, the possibilities of post-translation modifications and the subtleties of growing high-quality crystals in the presence or absence of ligands. Nonetheless, in the last few years, the extended efforts of several structural genomics programs have resulted in effective strategies to address these problems [5-7] by screening methods that optimize protein stability and protein crystallization. Thus, experimental structures of specific drug targets or of some structural homologs are becoming more accessible [5]. In the near future, as the different structural genomics initiatives reach maturity, more experimental structures related to therapeutic targets will continue to fill the existing gaps with synchrotron sources still playing a dominant role [8]. Except for the unique challenges of membrane protein targets, some structural information is available or will soon be available, either for the target themselves or for closely related homologs in public domain databases [9] or in the laboratories of the pharmaceutical industry. This three-dimensional information is being used and will continue to be used in many different ways to accelerate the drug discovery process.

However, the complexity of the biological issues directly related to drug discovery (i.e., target validation, toxicity, pharmacokinetics and many others) has limited the impact of SBDD and other technologies in the drug discovery process. As a consequence, there is surge of interest in making the drug discovery methodology more effective by incorporating, early in the design and evaluation of the molecular entities, the most favorable physicochemical properties (i.e., molecular weight, solubility, polar surface area, ClogP and others), to increase the probability of finding clinical candidates. The importance of those properties in the design of therapeutically effective compounds has always been recognized in an implicit manner, connected to the experience or insight of the individual medicinal chemists or savvy project leaders. However, there is a need to put those intangible qualities and knowledge into a numerical and quantitative framework that will make the process more robust and the favorable outcome more probable.

In this review, the generalized concepts of ligand efficiencies will be introduced and put into the context of other ways of assessing ligand efficiency. The application of these concepts to fragment-based drug design, library design and virtual screening is reviewed or outlined and their use is illustrated with initial examples from the literature as well as using retrospective analysis in some concrete targeted projects. A general outline of how to incorporate those concepts into the drug discovery process is suggested. Although the focus of the review is on structure-based approaches, the concepts related to ligand efficiency can be applied to any drug discovery methodology. It is suggested that formalisms based on the concepts of ligand efficiencies will facilitate the drug discovery process in the near future by providing a more adequate numerical and statistical treatment of the drug discovery process.

2. Definitions

An initial attempt to quantify the binding affinity of a ligand in relation to the number of non-hydrogen atoms was presented by Kuntz *et al.* [10]. An extension of their original concept was given an explicit formulation by Hopkins *et al.* [11]. They defined 'ligand efficiency' (LE) numerically as the quotient of ΔG and the number of non-hydrogen atoms of the compound:

$$LE = \Delta g = (\Delta G)/N \quad (1)$$

where $\Delta G = -RT\ln K_i$ and N is the number of non-hydrogen atoms. The units of LE (Δg) are kcal/mol per non-hydrogen atom. For reference purposes it is useful to note that for K_i = 1 nM at 300° K, a compound has a binding energy of -12.4 kcal/mol. Thus, a 1 nM compound consisting of 25 non-hydrogen atoms will have a ligand efficiency (LE) of ~ 0.5 kcal/mol/non-hydrogen atom.

As non-hydrogen atoms can be of many different types and a key property of a compound is its molecular weight (MW), a natural extension of the concept of ligand efficiency has been introduced recently that is based on expressing the binding affinity as $pK_i = -\log K_i$ and using MW as reference expressed in kDa (Box 1 of [12]). Three related efficiency indices: per cent efficiency index (PEI); binding efficiency index (BEI) and surface efficiency index (SEI) have been introduced and their mathematical definitions and reference values are summarized in Table 1 for convenience.

PEI is the fractional (0 – 1 scale) inhibition of a compound divided by the MW in kDa. BEI is the binding efficiency index relating potency to molecular weight on a per kDa scale and SEI is the surface efficiency index monitoring the potency gains as related to the increase in polar surface area (PSA) referred to 100 Å^2. It should be noted that K_i, K_d, and IC_{50} values are not strictly interchangeable, but for the purposes of the indices described here, it is not necessary to make a distinction. It is assumed that comparisons will always be made using similar measurements. The units of kDa for MW used to calculate BEI simplify the arithmetic calculations. The choice of the 100-Å^2 polar surface area as a normalizing factor in SEI is convenient and also relates to the results of

Table 1. Names, definitions and idealized reference values for ligand efficiency indices.

Name	Definition	Reference value*
PEI	Per cent inhibition (as a fraction: 0 – 1.0)	2.7
BEI	$p(K_i)$, $p(K_d)$ or $p(IC_{50})^{\ddagger}$/MW(kDa)	27
SEI	$p(K_i)$, $p(K_d)$ or $p(IC_{50})^{\ddagger}$/(PSA/100 Å^2)	18‡
LE	ΔG/N(non-hydrogen atoms)	-0.50

*Reference values are calculated for each index using the following idealized values (units have been omitted in the table):

Per cent inhibition of 90% in a fractional scale (0.9).

Molecular weight = 333 Daltons (0.333 kDa). This value of MW was found to be a suitable reference point as it makes the arithmetic easy and is also near the mean value of MW for a large sample of marketed oral drugs [60,61].

K_i or IC_{50} = 1.0 nM; pK_i = 9.00.

Van der Waals PSA = 50 Å^2.

ΔG = -12.4 kcal/mol; N (non-hydrogen atoms) = 25.

$^{\ddagger}p(K_i)$, $p(K_d)$, or $p(IC_{50})$ is defined as -log(Ki), -log(Kd), or -log(IC_{50}), respectively. These related quantities have been referred to as affinity [32] and the affinity per Dalton used by this author as a measure of 'efficiency of ligand binding' differs from BEI only by a factor of a thousand.

ΔG= -RT lnKi ; <MW> for a 'drug-like' atom is assumed to be 13.3 (11). Ideally, it would require 25 of those non-hydrogen atoms to reach 333 Da.

Note: By the definition, for any given compound the ratio of BEI/SEI is equal to 10*(PSA/MW).

BEI: Binding efficiency index; LE: Ligand efficiency; MW: Molecular weight; PEI: Per cent efficiency index; PSA: Polar surface area; SEI: Surface efficiency index.

Palm *et al.* who found a sharp change in oral bioavailability for compounds with polar surface areas near 100 Å^2 [13].

Although in their definitions, BEI and SEI for each ligand share the potency (as pK_i or equivalent), these two variables are not correlated. Correlation coefficients for different test cases typically are in the range of 0.17 – 0.20 (Metz pers. commun.) and, thus, should be considered as two independent parameters. Excluding the sign, an approximate conversion factor of 54 between LE and BEI can be estimated for a 1 nM compound (MW = 333), containing 25 non-hydrogen atoms with an mean MW per 'drug-like' atom of 13.3 daltons [11]. Given the related definitions of LE and BEI, there is an excellent correlation (~ 0.9) between the corresponding values for each specific target, from which the conversion factor between the two can be obtained.

One should compare the definitions of these new variables with the standard wisdom of the field as summarized in the Lipinski's 'rule of five' and others (i.e., rule of three [14]) that are being used as criteria in the drug discovery process more as 'rules-of-thumb' or filters, than as a rigorous mathematical framework to optimize the drug discovery process (Table 2). BEI provides a continuous numerical scale for one of Lipinski's variables (MW) and implicitly (via PSA), SEI provides an analogous scale for all the others related to solubility and favorable PK. They can be used to rank compounds on a logarithmic scale similar to the pH scale, but their use does not suggest or imply broad cut-offs based on threshold values.

As introduced here, efficiency indices only provide a numerical framework for three critical variables (potency, MW and PSA) by combining them into two (BEI, SEI) and providing similar and continuous numerical scales for ranking, comparing and optimizing their values in a simple, two-dimensional 'optimization' plane [12]. Naturally, other variables of importance could be added in a similar format and scale to provide a multidimensional framework for future analytical or numerical optimization.

Similarly to the Richter scale used for expressing the magnitude of seismic phenomena, relatively small changes (~ 1) in SEI or BEI reflect significant changes in the fingerprint of the compounds they represent in the optimization plane. As defined, the ratio BEI:SEI is independent of the pK_i values and is directly proportional to PSA/MW (see note in Table 1). The latter ratio can be considered a crude descriptor of solubility and for a specific compound series is well correlated to their predicted solubility (Abad-Zapatero, unpublished data). In the optimization plane, polarity decreases as the compounds move to the right, improving their SEI and shifting the compound towards more favorable drug-like properties (see below).

3. Efficiency indices in the hit-to-lead process

After target selection, weaving through the list of high-throughput screening (HTS) hits is one of the most critical steps of successful drug discovery. Insightful choices of the correct chemical groups would likely reduce the attrition rate and maximize the probability of a successful clinical candidate. The significance of very high values in PEI or BEI to indicate the presence of 'nuisance' compounds or irreversible inhibitors in protein tyrosine phosphatase 1B (PTP1B) has been discussed [12,15]. In PTP1B, the isoquinoline diols (BEI ~ 33) illustrated as class F (Figure 1) were potent oxidants of the active site Cys215.

Several computational approaches have been implemented at different laboratories to make the selection process more robust (16 – 18). The specifics of the different approaches can be found in the original publications, but a common theme is to downplay the value of potency alone or even LE only for single hits. The computational approaches implemented are directed towards trying to prioritize the possible series based on a combined profile that would include not only LE, but also selectivity and absorption, distribution, metabolism,

Table 2. Binding efficiency indices compared with threshold criteria in preclinical drug discovery.

Variable	Efficiency indices: BEI, SEI	Rule-of-3*	Rule-of-5‡
ClogP	Implicit scale via SEI§	≤ 3	≤ 5
Number of N	Implicit scale via SEI§	Yes	Yes
Number of O	Implicit scale via SEI§	< 9	< 10
H-bond donors	Implicit scale via SEI§	≤ 3	≤ 5
MW	Cont. scale in kDa in BEI§	≤ 300	≤ 500
Number of rotatable bonds¶	Not included explicitly	≤ 3	≤ 10
PSA¶	Cont. scale, explicitly included in SEI	≤ 60 Å^2	≤ 140 Å^2

*Congreve *et al*. [14].

‡[62].

§See Table 1 for definitions.

¶Later suggested by Veber *et al*. [63].

Number of rotatable bonds is an entropy-related term and if the values of K_i (or related) used to compute SEI/BEI are experimental, those values would contain an entropy component in them. However, if the K_i values are obtained by computational methods, the issue is much more complex, but the relative ranking of compounds is still possible in terms of the 'predicted' SEI/BEI.

'Astex Rule of Three' is a trademark of Astex Technology.

BEI: Binding efficiency index; MW: Molecular weight; PSA: Polar surface area; SEI: Surface efficiency index.

elimination and toxicity (ADMET) data. Only by looking at this combined fingerprint would it be possible to judge fully the potential of the different series. The examples presented [17] show how the distinctions between 'lead-like' and 'drug-like' might be arbitrary; both classes of hits are good starting points for optimization, even if their affinity for the targets is only moderate-to-low. They conclude that the most promising compounds have a favorable LE value and few liabilities that can be addressed by medicinal chemistry. Keseru and Makara have examined the changes from hits to leads in a sizeable (~ 40 targets) sample and find that for 87% of targets the LE efficiency improved with a mean increase of 0.07, with certain classes (mGluR5) gaining almost 50% from hit to lead [18]. Adding SEI to the ligand efficiencies related to potency/per-dalton (LE, BEI) during the hit-to-lead selection process could facilitate the decision-making.

The data are still scarce, but as more laboratories develop integrated, computationally driven, decision-making protocols for analyzing and selecting their leads from the pools of HTS hits, it is quite likely than these initial trends of increased efficiency (potency/MW) will be confirmed. It is still uncertain how best to include ADMET properties when scoring screening leads before medicinal chemistry optimization begins and whether any efficiency indices can be suitably defined.

4. Efficiency indices in fragment-based lead discovery

Fragment-based approaches have now become mainstream in drug discovery from screening [19], to lead discovery [20,21] and lead optimization [22]. An extensive review of the field has been published recently [23]. It is precisely through the goal of optimizing the stages of 'fragment-like', 'lead-like' and 'drug-like' that the concept of 'ligand efficiency' has come to a more clear formulation.

After the initial paper by Kuntz and co-workers, the concept of LE developed, together with the ideas and strategies of fragment-based drug design. It was implicit in the issues raised by Rees *et al.* [20] and formulated explicitly by Hopkins and co-workers [11] as in Equation 1. From the initial analysis of 160 high-affinity ligands, Kuntz and colleagues inferred that the binding of ligands to their target was dominated by a steep gain of ~ 1.5 kcal/mol/heavy atom for compounds of ~ 15 non-hydrogen atoms and reached a plateau of 12 kcal/mol for larger ligands (> 25 atoms). These initial conclusions have been revised [11,24,25] to suggest a more realistic value of ~ 0.3 kcal/mol for the contribution of a standard 'drug-like' atom with a mean molecular mass of ~ 13.3 Da. This gentler slope of binding energy/atom has been documented in several fragment-based lead discovery examples [21] and has also been uncovered by using a more restrictive set of ligands in Kuntz's original data set by two recent analysis: Leach and co-workers [24] and Hajduk [25]; the latter by restricting the analysis to ligands with 5 – 25 atoms. Thus, from several viewpoints it appears that a consensus LE of 0.30 kcal/mol/non-hydrogen atom added would be a realistic minimum gain as the compounds evolve from 'lead-like' to 'drug-like'.

This optimized value should be compared with the trends observed when relating drugs to the screening leads from which they were derived. In their original analysis [26], Oprea and co-workers estimated a median molecular weight increase of 69 Da. Further analysis of a larger data set resulted in a much smaller average increase of only 42 Da (from 272 to 314 Da) [19,24]. If the average 'drug-like' heavy atom has a mean molecular mass of 13.286 Da [11], the implication is that, ideally, the structural changes from lead to drug should involve only the addition of 3 to 5 additional heteroatoms to the original scaffold. Proudfoot [27] has documented how closely related drugs are to their parent leads, showing that the

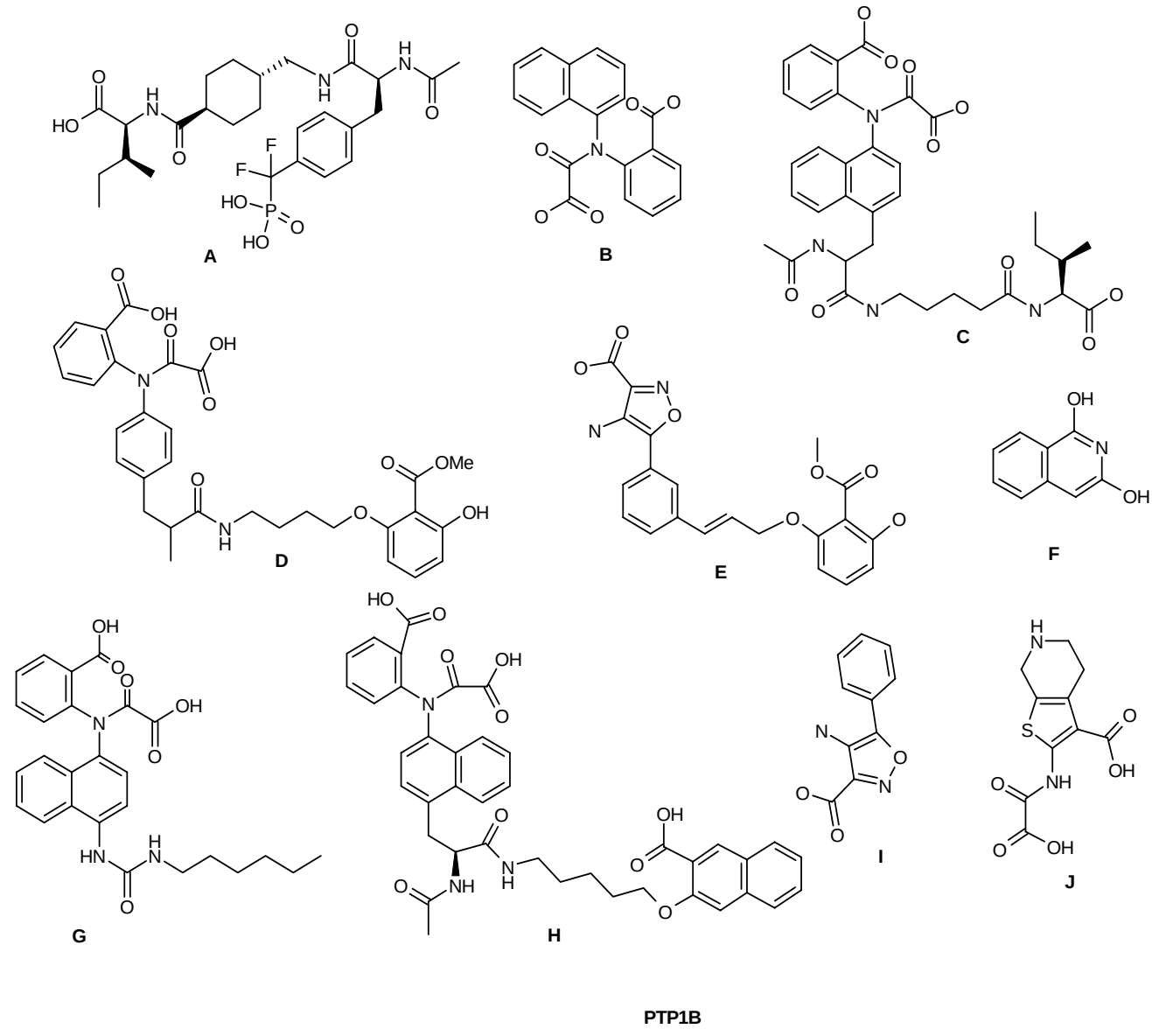


Figure 1. Schematic diagrams of the structures and classes of inhibitors for PTP1B. The letters refer to the chemical classes illustrated in **Figures 2 – 5** and discussed in the text. A: difluorophosphonates; B: single-site naphthyl oxamates; C: 2-site oxamates with amino acid; D: 2-site oxamate salicylate; E: 2-site isoxazole salicilylate; F: isoquinoline diols; G: 1-site oxamate with linkers; H: 2-site oxamate with naphthyl; I: single-site isoxazoles; J: represents pre-existing 1-site PTP1B inhibitors (2-(oxalyl-amino) carboxylates) from other competing groups [52,53]. Class J is not included in **Figures 2** and **3** for clarity. The relative position of this class in the SEI/BEI plane is illustrated with one compound in **Figure 4** (SEI = 3.7, BEI = 16.0) and two in **Figure 5** (purple squares) with efficiency indices calculated using the K_i values obtained by internal assays as opposed to the ones available in the indicated publications. Class F is not represented in the graphs because it would be out of range (BEI ~ 33).

majority of the drugs analyzed (up to 2003) were within 25% of the lead MW and within a very limited range of CLogP values. A significant example that even surpasses those criteria is the discovery of the Factor Xa inhibitor BAY-59-7939 [17,28]. Although this compound is not yet a marketed drug (presently in Phase II), starting with a 'drug-like' compound with an IC_{50} = 20 µM (MW = 421 Da, BEI = 11.2, SEI = 7.6) four minor changes in the scaffold yielded a compound with an IC_{50} = 0.7 nM with a MW increase of only 14 Da (BEI = 21.0, SEI = 10.4). These small changes shifted both efficiency indices towards more optimum values rather dramatically. Very large shifts in potency (EC_{50} changed from > 10 µM to 4 nM) have also been achieved as 'latent hits' in screening collections [29] and are uncovered within privileged scaffolds [30] by just changing one methoxy group into the corresponding hydroxyl analog (BEI changed from < 17.7 to ~ 31).

These singular examples aside, it should be noted that the overall figures are only reflecting mean trends covering many different targets. As has been recently noted [31], the molecular properties of specific drugs differ significantly for different targets or target families [32] and broad generalizations of 'drug likeness' should be made with caution. To assess whether or not any multitarget generalizations related to LE are possible, existing data from previous SBDD efforts must be analyzed. Data for individual targets and also across target families should be re-examined critically and there are already some initial studies available [32]. The application of any insights gained from these analyzes should enhance and enrich the future development of SBDD.

5. Efficiency indices in fragment-based ligand optimization

As indicated in the concise communication by Hopkins and colleagues, the efficiency of the initial hits or leads is the critical parameter for further optimization, not the potency alone [11]. It is becoming well accepted that differences in the efficiencies of the initial fragments can have a large impact in what might be possible along the optimization path of a chemical series for a particular target [11,21]. However, it is less clear about the path that fragments follow on optimization and the available data suggests that this is most likely to be target dependent [21,33]. Using a retrospective lead deconstruction analysis, Hajduk has suggested recently [25] that following an idealized path in the optimization of a fragment, the successive gains in pK_D can be related to the corresponding compound molecular weight by the linear least-squares relationship:

$$MW \approx m \cdot pK_D + b \quad (2)$$

Clearly, if Equation 2 is true, a similar equation can be fitted, where pK_D can also be expressed as a function of MW with the corresponding change in the values of the slope (m) and intercept (b):

$$pK_D \approx m \cdot MW + b \quad (3)$$

This is an intriguing concept that could have significant implications for future fragment-based SBDD and was suggested by Hopkins *et al.* [11] as a possible trend. However, a few qualifications are appropriate. The least squares analysis is limited to a specific range of values ($3.3 < pK_D < 9.3$; $132 < MW < 813$ Da) and it is assumed that the pK_D values do not have any errors associated with them. The different values for the slopes (m) and intercepts (b) for the 18 highly optimized drug leads (derived from 15 targets) can be derived from the supplementary material [25]. Hajduk suggested that the resulting slopes are approximately constant and that the y-intercepts are different from zero and vary for the various targets (Figure 2 of [25]). Although the correlations coefficients are high for the small sets of three to six compounds that are used in the separate least-squares fits, one should keep in mind that by extrapolation a compound with a MW of 0 should have no binding energy ($K_D > 1$ M). Therefore, values of b significantly different from 0 should be considered with caution, as they are outside of the fitted pK_D range of values. The significance, if any, of the different intercept values for the different targets is uncertain at this time. Requiring that the least squares lines pass near the origin would result in a wider variation of the linear slopes for the different optimized series in the different targets.

General expressions for the relationship between BEI and the slope and intercept of Equations 2 and 3 can be easily obtained. By simple algebraic manipulations on Equation 2 (i.e., inversion, followed by multiplication successively by pK_D and 1000), it can be shown that:

$$BEI \approx 1000/(m + b/pK_D) \quad (4)$$

Analogous manipulations on Equation 3 can be used to obtain the corresponding equation:

$$BEI \approx 1000 \cdot \langle m + b/(MW) \rangle \quad (5)$$

(Note: BEI = 1000·(pK_D/MW). The generalized Equations 4 and 5 show that if the intercepts of the linear regressions (Equations 2 and 3) are near 0, the BEI is essentially constant and proportional to the slope of the regression lines; the factor of a thousand derives from the different scales for MW. This is the original trend suggested by Hopkins *et al.* [11] and also observed in later work in certain targets [21]. Otherwise, a minor correction (~ 20%) needs to be applied that is more significant in the low pK_D (< 3 or K_D > mM) ranges. This simple analysis suggests that if MW is linearly related to pK_D (Equations 2 and 3), then the slope of the line is directly related to BEI and should stay approximately constant through the pK_D (or MW) range and specially at the most significant pK_D values (pK_D > 6 or K_D < µM), where the correction is rather small and the pK_D values more accurate. This is probably the trend observed by Hajduk in his deconstruction analysis. The average value of the slope of 64 (± 18) (for Equation 2) for the various targets discussed in [25] would correspond to BEI ~ 15.6, assuming a value of b near 0.

An interesting implication of Hajduk's analysis is the use of his findings to develop prediction maps to enable quantitative assessments of lead identification and optimization. This line of reasoning permits one to estimate the size of a ligand for a suggested target potency, given the pK_D of the initial fragment. The analysis is attractive, but it derives from the constancy of BEI through the optimization process and can be simply stated: if the BEI stays constant, an n-fold improvement in pK_D would require a corresponding increase in the MW (see Table 1). Hopkins and co-workers [11] presented a similar reasoning. One could ask: if the BEI is approximately constant during the optimization process, which variable is being optimized? A simple extension of the data presented for the BCL-x_L target (Figure 1 of [25]) shows that although BEI is approximately constant the corresponding values for SEI increase from 5.5 (compound 6 of [25]) to 7.3 for the clinical candidate (compound 1, AB-737). Other issues such as target specificity might have been addressed during the optimization process for which the values of SEI probably has no relevance.

Those qualifications not withstanding, the analysis shows how the BEI can be used to separate highly druggable targets [34] that maintain a high efficiency (i.e., protein kinases BEI ~ 23) throughout the SBDD process, from the more challenging ones where achievable efficiencies are much lower (i.e., BCL-x_L, BEI ~ 12). A more extended analysis of fragment optimization pathways in different targets is needed to confirm whether the trend suggested by the deconvolution analysis illustrated by Hajduk is sustained prospectively without the knowledge of the 'ideal' back-fragmentation of the final drug candidate.

6. Efficiency indices in library design

The simplicity of threshold rules (Lipinski's and others) made them easily accepted and implemented as *in silico* filters to assess the 'drug-like' qualities of libraries and compound collections. Other filters, some of them based on various mathematical models ([12,35,36] and references therein), developed later and are now part of the standard filtering process of libraries at design or synthesis stage. It would seem only a natural extension to attempt to design prospective libraries with the objective of maximizing BEI and SEI, separate on in combination. However, the effectiveness of this plan is limited by the availability of accurate estimates of the binding affinities (i.e., K_i or equivalents) (see below). Although progress is being made in the field of calculating ligand-binding affinities by various methods, the results are still within ~ 1 kcal for congeneric series and larger for more distant compounds [24,37,38]. This implies an uncertainty in the K_i of ~ 10-fold. For library design, the problem is compounded by the uncertainties in the mode of binding (variations in the points of interaction between ligand and protein, i.e., poses), when the resulting compound departs significantly from the original fragment. However, it should be pointed out that once the structure of the target with the core fragment is known, all that is needed is a relative ranking of the compounds in terms of K_i to evaluate the quality of the resulting compounds. Given an estimate of K_i, it is very simple to obtain BEI and SEI. Often, the K_i values for a subset of related compounds are available from the experimental results of the project and it should be possible to scale the theoretical results obtained for the measured compounds and provide a reasonable reference for the unknown compounds, enough to permit the relative ranking for various library ideas. Although conceptually feasible, the cooperation and synergy between the groups performing the activity assays and the computational groups will be necessary to address technical issues such as a wide range of compound activities and sizes in order to provide a satisfactory and meaningful scaling.

Examining the values of BEI and SEI individually or in combination in the SEI-BEI optimization plane would allow a relative ranking of different libraries that would facilitate the selection process for further development. A recent example should be noted in the scaffold oriented synthesis of thienopyrazoles as kinase inhibitors [39] where it is suggested that kinase inhibitors with high binding affininity (BEI ~ 24) represent excellent starting compounds for subsequent libraries focussed on additional potency and selectivity. Although no specific efficiency gains are given, trends of increasing efficiency have been also shown with linear library synthesis and elaborations for different target classes [33]. The advantage that these efficiency indices offer is the combination of three variables into two, allowing for easier comparison of these three critical properties at the same time in any library design or improvement effort. As an example, the values of the density of various substances are a better way to compare them than the separate measurements of their respective masses and volumes. Some more examples are needed to provide validity to this approach and to explore the value of efficiency indices in library design (see below).

7. Efficiency indices in virtual screening

In contrast to library design, the process of virtual screening of large existing compound collections or virtual libraries depends on the availability of a reasonable estimate of the binding affinity between the ligand and the structure of the target. This is a rather complex problem in SBDD and recent reviews address the advances and the technical issues that are still challenging the field [24,37]. Critical issues are: i) identifying the correct conformation of the ligand; ii) select the correct binding mode or 'pose' for the given ligand conformation; iii) describing conceptually and numerically the best scoring function; and iv) obtaining a reasonable speed so as to make the process computationally practical. The incorporation of algorithms that include ligand- and receptor-induced fit effects is a significant development that will improve the estimates of the successive binding affinities between ligands and receptors [40,41]. Critical assessments of the docking programs, scoring functions and other related issues in the field have been recently reviewed [24].

As indicated above, even with only approximate values (within ~ 10-fold) of the binding affinities between the ligand and the target, it is possible to relatively rank chemotypes, scaffolds, fragments or even individual compounds in terms of their BEI/SEI values. This can be done in a manner similar to what is presently done with the binding affinities or free energies calculated by the programs. Ranking the possible hits in terms of their respective BEI values would provide a relative measure of their efficiency per unit mass added. An initial example of the use of ligand efficiencies to asses the quality of the resulting inhibitors in the virtual screening against metalloenzymes has been reported [42]. In addition, the relative values of SEI would serve as a measure of their potency per PSA exposed. Combining the relative values of the different compounds in the optimization plane (SEI-BEI) would serve to identify where the compounds stand, compare them with available existing chemical entities from the competition or from earlier efforts from the project. It could also help to project the future development of the different chemotypes, scaffolds or fragments and to assess the druggability of the target [34]. Undoubtedly, as the tools and algorithms of the computational chemists improve, so will the predictive value of these models.

8. Efficiency indices in lead optimization

At present, there are no published examples of the prospective use of efficiency indices (both BEI and SEI) in SBDD to guide or at least facilitate the discovery process. However, some insights can be gained as to their potential to guide SBDD efforts by analyzing retrospectively the data available for other targets. The analysis can benefit from the extensive data available in proprietary internal databases as will be shown with the first example, PTP1B. However, even the limited published data alone (FBPase) can be sufficient to illustrate a different approach to drug discovery, using efficiency indices as guides.

8.1 PTP1B

In the past decade, human PTP1B has been the target of extensive SBDD efforts in many laboratories because of its relevance in biological processes related to insulin signaling and Type 2 diabetes [43]. The medicinal chemistry efforts in the author's laboratory have been published elsewhere [44-51] and a brief summary of this fragment-based strategy has been recently presented [15].

The set of 122 compounds for which a refined three-dimensional structure of the PTP1B–ligand complex was known were selected for analysis. The compounds were grouped into nine chemical classes (A – I) based on the structure and mode of binding as was outlined by Abad-Zapatero *et al.* [15] and illustrated in Figure 1. An additional group, class J (2-(oxalyl-amino)-carboxylates), has been added to represent existing PTP1B inhibitors by other groups [52,53] (Figure 1). SEI, BEI values were calculated for the different compounds and subject to the statistical analysis summarized in Figures 2 and 3. The use of the two indices individually (Figures 2 and 3) or in the combined SEI-BEI optimization plane (Figure 4) allows a different perspective on the merit of the different chemical classes.

Looking at the most relevant classes for the discussion (B – E, G, I in Figure 2), the BEI values of the individual compounds can be clustered within each class and the mean value of each class can be compared statistically with the one from the other classes. Class B, 1-site naphthyl oxamates, has the highest mean BEI and is statistically superior (95% confidence level) to classes C, D, E and G. This result confirms numerically the notion that the 2-site compounds, although more potent (typically low nM), are less efficient than the 1-site napththyl oxamates (at best only 1 µM) given their larger MW [54]. This class of compounds (class B) is also more efficient than class G, which includes the 1-site oxamates designed to bind like the Tyr-P residues that they are trying to mimic by interacting with Asp48 in PTP1B. The analysis also shows that the mean of class I (1-site isoxazoles) is not very different from class B, achieving similar efficiency with a different core. In fact, some of the high ranked compounds in class E can compare very favorably with the best compounds in class B in terms of BEI, but not in terms of potency alone.

Figure 3 makes a similar analysis of the compounds in the different classes based on their efficiency per unit of PSA exposed (SEI). Overall, class B is still the one with the highest value, but two insights are apparent. First, classes D and E are now more distinct from C showing the effect of having fewer polar groups and more compounds within class E are comparable with the better ones in class B. More importantly, the analysis shows clearly how poorly all of the compounds of the different series rank in terms of potency per PSA (BEI < 6). All of the compounds have several polar groups, especially in

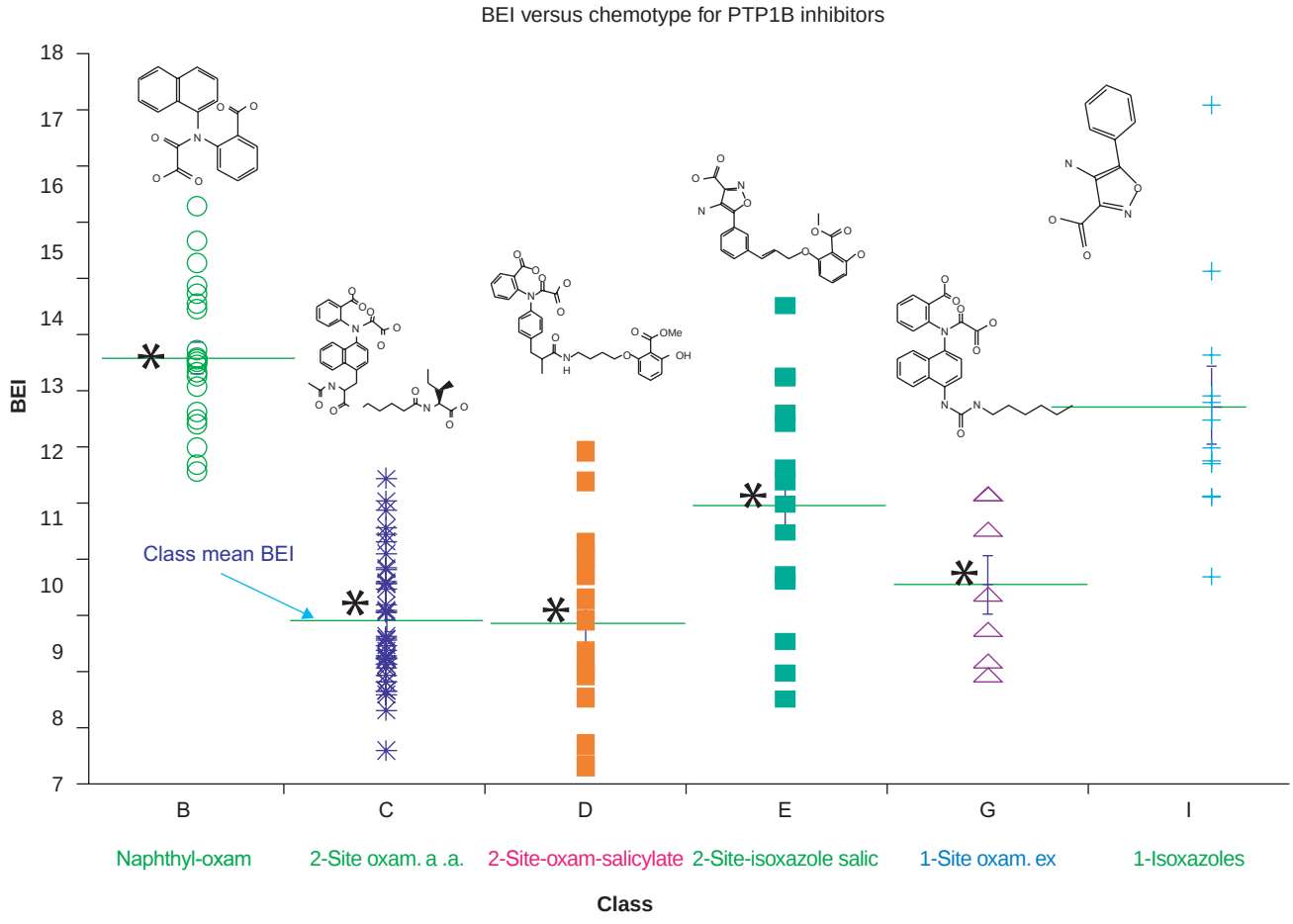


Figure 2. BEI analysis of the PTP1B inhibitors. BEI versus chemical class. The bar represents the class mean and * indicates that the mean of class B is significantly different (95% level) from all the others. Class B has a better BEI than any of the others, although one individual compound of I is superior.

The statistical analysis and the plots were carried out using JMP version 5.1.1 (SAS Institute 1989 – 2004).

BEI: Binding efficiency index; PTP1B: Protein tyrosine phosphatase 1B.

classes C – E where the charges and separated by long linkers and fully exposed.

The combined characteristics of all the compounds are quite apparent when their SEI/BEI values are plotted in the optimization plane (SEI-BEI). It is apparent that the class of compounds placed more favorably in the plane is class B (green open circles), reflecting the superior values of SEI and BEI. However, these compounds lacked a critical ingredient of a good therapeutic agent against PTP1B: they lacked phosphatase selectivity [15]. Overlapping with the disperse cloud of points representing class B, there are several favorable compounds of class E (double site isoxozales, green filled squares), which as indicated earlier, also had favorable physicochemical characteristics. It is worth noting that the only cell active compound for this target was among the class E and its location in the plane and structure are shown on Figure 4 (blue rectangle). This analysis notwithstanding, the centroid of the 122 compounds analyzed had SEI, BEI values (~ 5.2, 12.5; respectively) indicating how far they were from a 'reasonable' target value of 18 and 27, respectively (Figure 5, Table 1). The vast majority of the compounds (n = 114, root-mean-square error [RMSE] = 1.39) extend closely along a line with a slope (BEI/SEI) equal to ~ 3.2 or PSA/MW ~ 0.32. This high ratio reflects the polar nature of the inhibitors fitting into a very polar binding site. It should be noted that as BEI and SEI have the same numerator (Table 1), it is justified to assume that the linear fit of the two variables should pass through the origin (Figure 5). The conventional linear least squares achieved a similar RMSE with an R^2 = 0.68 (n = 114) and intercept near 0.

Although no rigorous comparison has been made between the low SEI values and poor permeability/bioavailability of all the PTP1B inhibitors discussed in this work, the trend is certainly clear. Essentially, all the compounds (Figure 1) are negatively charged and, therefore, the dominant factor in

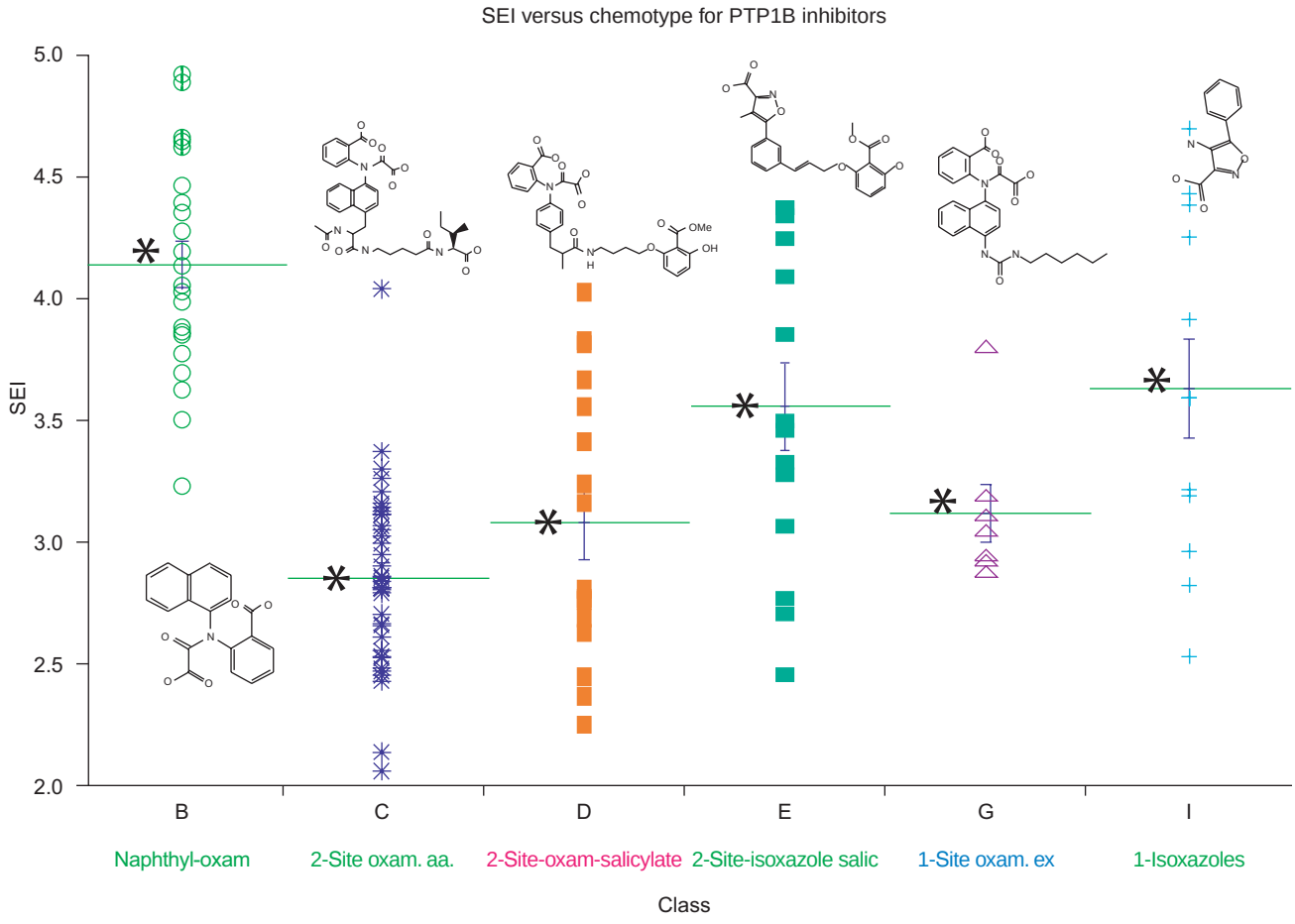


Figure 3. SEI analysis of the PTP1B inhibitors. SEI versus chemical class. Overall the values of SEI are rather low for all the classes represented. Even so, classes B, E and I rank the best, with class B being statistically different (95% level) than all the others.
PTP1B: Protein tyrosine phosphatase 1B; SEI: Surface efficiency index.

determining their permeability is their PSA, which is approximately inversely proportional to their bioavailability score (ABS, see Figure 10 of [55] and FBPase section below). The smaller compounds moved towards relatively high BEI values, but the polar nature of the PTP1B active site made it very difficult to optimize the SEI (SEI < 5.5) values to anywhere near the diagonal, where both BEI and SEI would be comparable suggesting more 'drug-like' properties. Figures 4 and 5 also show the position in the optimization plane of the other significant PTP1B inhibitors (class J, purple squares) that would also be negatively charged at physiological pH ranges. For comparison, two known drugs directed towards rather different biological targets are also shown: Iressa®, an anticancer agent targeting epidermal growth factor (SEI = 11.2, BEI = 17.0); and Haloperidol®, an antipsychotic agent drug targeting dopamine 2 receptors (23.3, 25.0, respectively) [12,56].

8.2 FBPase

Fructose-1,6-bisphosphatase (FBPase) is a homotetrameric enzyme that plays a critical role in the regulation of hepatic glucose output. The expression and activity of FBPase is upregulated in patients and in animal models of Type 2 diabetes (T2DM). There have been several reports of different small molecule inhibitors of FBPase as therapeutic agents for the treatment of T2DM (Figure 1 of [57]). Abbott's SBDD efforts concentrated on a novel benzoxazole-2-benzenesulfonamide core, which was characterized by X-ray crystallographic studies to bind at the tetramer interface in a unique mode [57].

The medicinal chemistry effort explored the structure–activity relationship (SAR) of the initial core by substitutions around the benzoxazole ring (4-7 substitutions, Figure 6A), replacements of the phenyl ring itself (Phen-rep, Figure 6B) and substitutions around the phenyl ring (Phen-sub, Figure 6C) and (Figure 7) [58]. An analysis similar to the one presented earlier for PTP1B can be performed based on a more limited set of compounds, suggesting that this *modus operandi* should also be useful during active projects. Figure 7 shows the plot of BEI versus chemotype, where the cofactor AMP and a reference compound (Metabasis both active and prodrug 9, 10; Figure 6) have also been included for comparison.

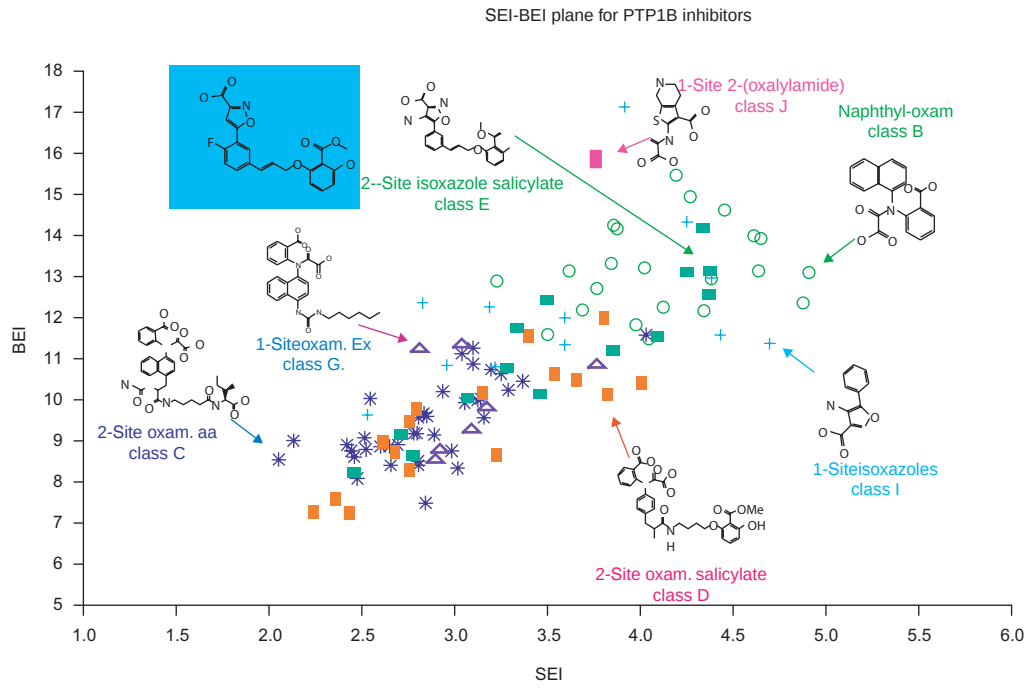


Figure 4. Bi-variate analysis of the different inhibitor classes in the optimization plane. The class symbols are indicated: class B: green open circles; C: blue stars; D: fill red squares; E: fill green squares; G: open blue triangles; I: green crosses; J: purple square. The compound in the filled blue box (upper left) was the only compound found to be active in cell assays. Chemical classes as in **Figures 2** and **3**.
BEI: Binding efficiency index; PTP1B: Protein tyrosine phosphatase 1B; SEI: Surface efficiency index.

Even with a limited number of compounds, the statistical analysis of the mean of classes clearly showed that classes represented by compounds 6 and 7 (5, 7-substituted benzoxazoles, Figure 6) were superior with mean BEIs around 12 and the 7-substitution being better than 5. These data were consistent with the SAR in that position 4 of the benzoxazole was the most restrictive to any kind of substitution, whereas at position 5 small substituents were tolerated and accepted at position 6. The most promising SAR followed from position 7 were larger substituents were possible yielding the most potent (0.57 µM) compound. Structural data showed that position 7 pointed towards the outside of the tetramer, whereas the other only allowed limited substitutions because they interfered in various ways at the tetramer interface (Table 3) [58].

More dramatic was the effect of the phenyl replacements where a limited SAR effort (Table 1 of [58]) showed that the naphthyl moiety was clearly superior (Figure 7) with a BEI > 15 (Table 3, with SEI, BEI values). The limited SAR on the phenyl substitutions did not yield a similarly robust response. The analysis of SEI versus chemotype also favored the results of the phenyl replacement as being more effective to produce compounds with superior BEI and SEI. It should be noted that, although the 7-benzoxazole substitutions gave the most potent compound of the medicinal chemistry effort (0.57 µM), the analysis based on the efficiency indices suggested that the phenyl replacement chemistry was more effective in yielding favorable compounds and pointed to the naphthyl substitution (compound 4, Table 3) as the highest ranking compound (Figure 7). The combined plot in the SEI-BEI optimization plane also confirmed this compound as the one having the best balance of potency/Da and potency/PSA (Figure 8). In fact, when this compound was tested it showed excellent pharmacokinetic (PK) properties (Figure 8) [58]. Unfortunately, compound 4 did not give the expected target-related therapeutic effects in animal models and no other compounds were tested for PK.

The continuous SEI values of the compounds from the different FBPase series relate well to the discrete bioavailability score (ABS) developed by Martin [55]. This is consistent with the observation that an approximate value for the ABS scores of three hypothetical anionic compounds with PSA in the mid-ranges of the intervals selected by Martin (PSA values of 50, 100 and 175 $Å^2$ with scores of 0.85, 0.56 and 0.17, respectively) can be estimated by a simple, inversely proportional relationship. The bioavailability scores can be obtained by:

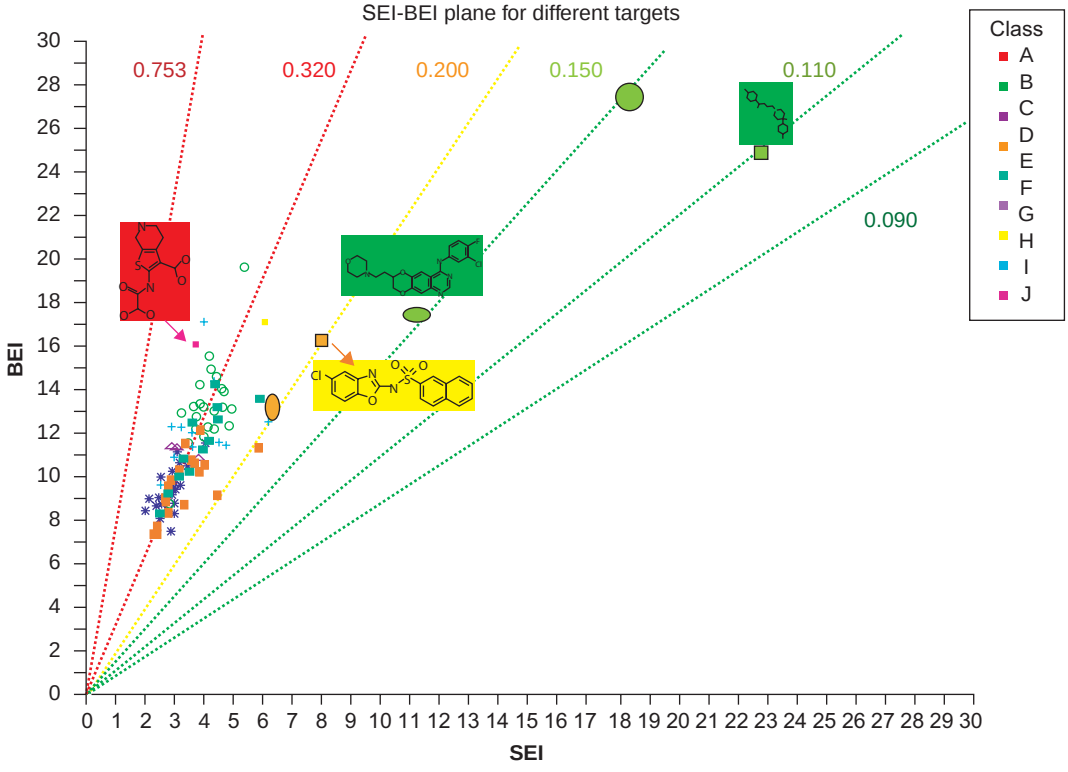


Figure 5. Bird's-eye view of the optimization plane for different targets. The 122 compounds for PTP1B extend along a region of the plane that is far from the diagonal line where approximately both BEI and SEI are optimized, favoring 'drug-like' properties. This is a reflection of the highly polar nature of the PTP1B inhibitors. Class J is defined in **Figure 1** and illustrated with one structure. The relative position of compounds for other targets is illustrated as follows. The vertical ochre oval marks the approximate position of the centroid of the FBPase compounds discussed in the text and illustrated in **Figures 7** and **8** and the ochre square is compound 4 (**Table 3**). The horizontal green oval represents the position of Iressa® (anticancer agent targeting epidermal growth factor receptor kinase) and the green square corresponds to Haloperidol® (a potent antipsicotic agent directed towards dopamine receptors in the CNS). The green circle is the position of the reference values given in **Table 1** for a 1-nM compound (PSA of 50 Å², MW of 333 Da). The different lines colored in a 'traffic-light spectrum' from red to green (left to right) correspond to the PSA/MW ratios indicated (0.750 – 0.090) showing the gradation from one PTP1B inhibitor, class J compound (potent, but very polar) to Haloperidol, a small and very potent antipsicotic agent. Compounds with similar PSA/MW ratios probably correspond to inhibitors with similar binding pockets binding in the same pose as illustrated for PTP1B. Class symbols as in **Figures 2 – 4**.

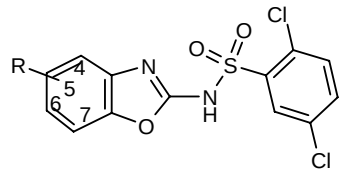
BEI: Binding efficiency index; PSA: Polar surface area; PTP1B: Protein tyrosine phosphatase 1B; SEI: Surface efficiency index.

$$ABS \approx 40/\langle PSA \rangle \quad (6)$$

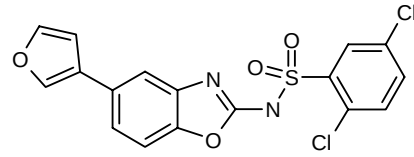
or:

$$ABS \approx 0.4/\langle PSA/100 \rangle \quad (7)$$

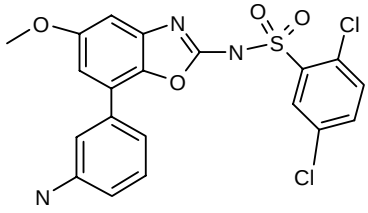
Where $R^2 = 0.88$, n = 3. The definition of SEI as inversely proportional to PSA, using a reference value of 100 Å² (Table 1) is, therefore, analogous to the discrete ABS scale and could provide a sliding scale of permeability/bioavailability related to potency at least for negatively charged compounds. As indicated earlier, the large PSA of the PTP1B inhibitors results in low SEI values that are correlated with poor ABS scores. The generality and consequences of these observations are still unclear. A more rigorous analysis of the relationship between ligand efficiencies and ADME properties for different targets and across a variety of chemical series is needed. The approximate position of the centroid of the best compounds discussed for this target is represented in Figure 5 (ochre oval), in relation to the more polar PTP1B inhibitors and the reference values for two known drugs (see above) [56].



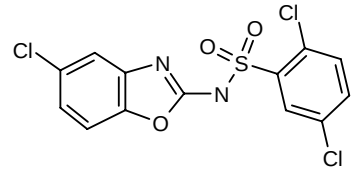
A



6

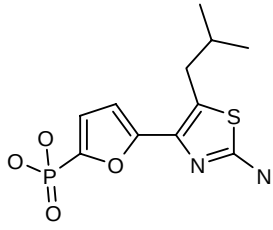


7



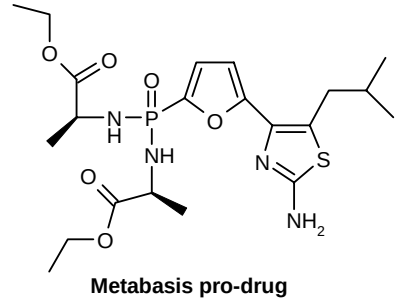
Lead benzoxazole

8



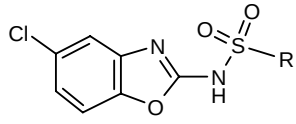
Metabasis

9

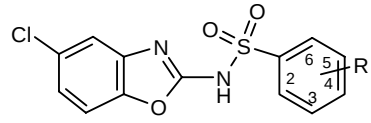


Metabasis pro-drug

10



B: Phen-rep

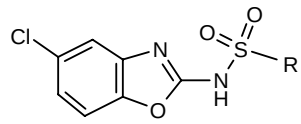


C: Phen-Sub

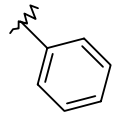
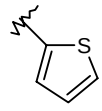
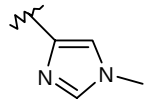
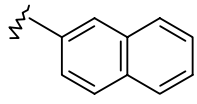
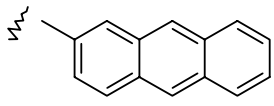
FBPase

Figure 6. Schematic diagrams of the structures and classes of inhibitors for FBPase. The letters refer to the chemical classes illustrated in **Figures 7** and **8** and represent a chemical strategy (A – C). **A**: four different classes derived from substitutions around the benzoxazole ring at positions: 4, 5, 6 and 7. **B**: phenyl-replacement detailed in **Table 3**. **C**: phenyl substitutions. Compound **6** is an important example of 5-benzoxazole substitutions. Compound **7** was the most potent of the double substituted benzoxazole 5-CH_3O, 7-(3-NH_2-Ph). **9** and **10** illustrate compounds from Metabasis ([57,58] and references herein).

Table 3. Efficiency SAR for the phenyl replacements of benzoxazole cores in FBPase.



Phenyl-rep SAR table

Compound	R	FBPase (IC_{50} μM)	BEI	SEI
1		7.7	16.6	7.1
2		13	15.5	6.8
3		9.4	16.1	5.6
4		2.5	15.8	7.9
5		3.4	13.4	7.6
1a	Me	> 50	< 17.5	< 5.9
1b	n-Bu	> 50	< 14.9	< 5.9

Adapted from [58].

BEI and SEI were calculated form the IC_{50} column using the MW and the standard 2D-polar surface area as indicated in **Table 1**.

For 1a and b an IC_{50} of 50 μM was used.

BEI: Binding efficiency index; FBPase: Fructose-1,6-bisphosphatase; SAR: Structure–activity relationship; SEI: Surface efficiency index.

9. Efficient structure-based drug discovery

How can we incorporate LE indices to provide effective numerical markers during the drug discovery process? An overall system is suggested in Figure 9. Ideas for novel compounds could come from individual suggestions (i.e., chemists, crystallographers, modelers) or from the present or future software tools [59]. Existing software can generate the three-dimensional structure of the suggested compound in the context of a closely related protein–ligand complex (boxes 2 and 3 Figure 9) (see [59] and references therein). Scaling of the theoretically predicted K_i values (or IC_{50} values) to the

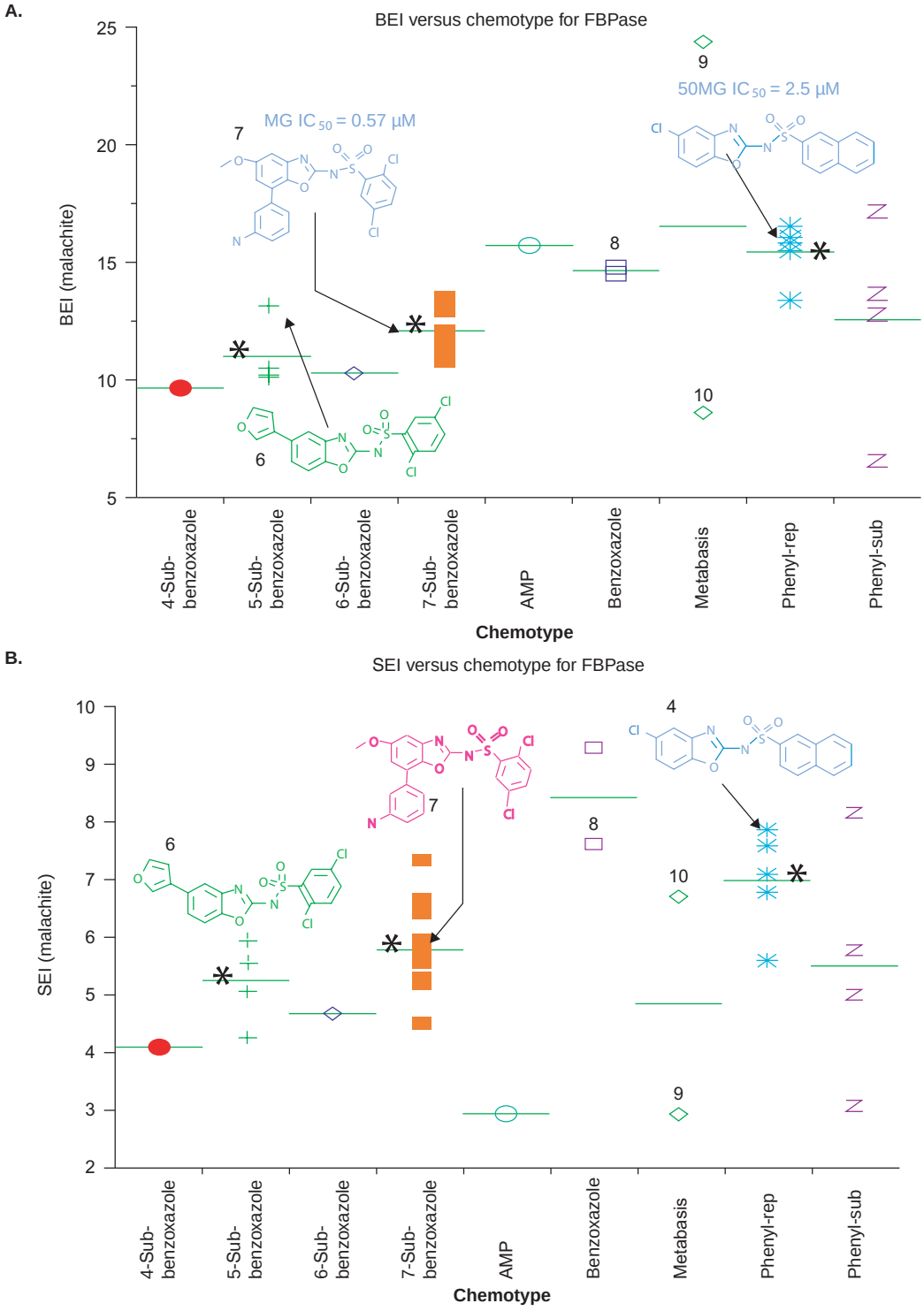


Figure 7. BEI-SEI analysis of the of the FBPase inhibitors. Analysis of the FBPase compounds in terms of BEI (**A**) and SEI (**B**) versus chemical class as in **Figures 2** and **3.** The different classes have been described in **Figure 6.** The compounds resulting from the phenyl replacement (black asterisks) are statistically better (90% level) than the other classes. The lower confidence level compared with PTP1B is related to the smaller number of members per class. Numbers next to compounds refer to **Figure 6** and **Table 3**.

The statistical analysis and the plots were carried out using JMP version 5.1.1 (SAS Institute 1989 – 2004).

BEI: Binding efficiency index; FBPase: Fructose-1,6-bisphosphatase; PTP1B: Protein tyrosine phosphatase 1B; SEI: Surface efficiency index.

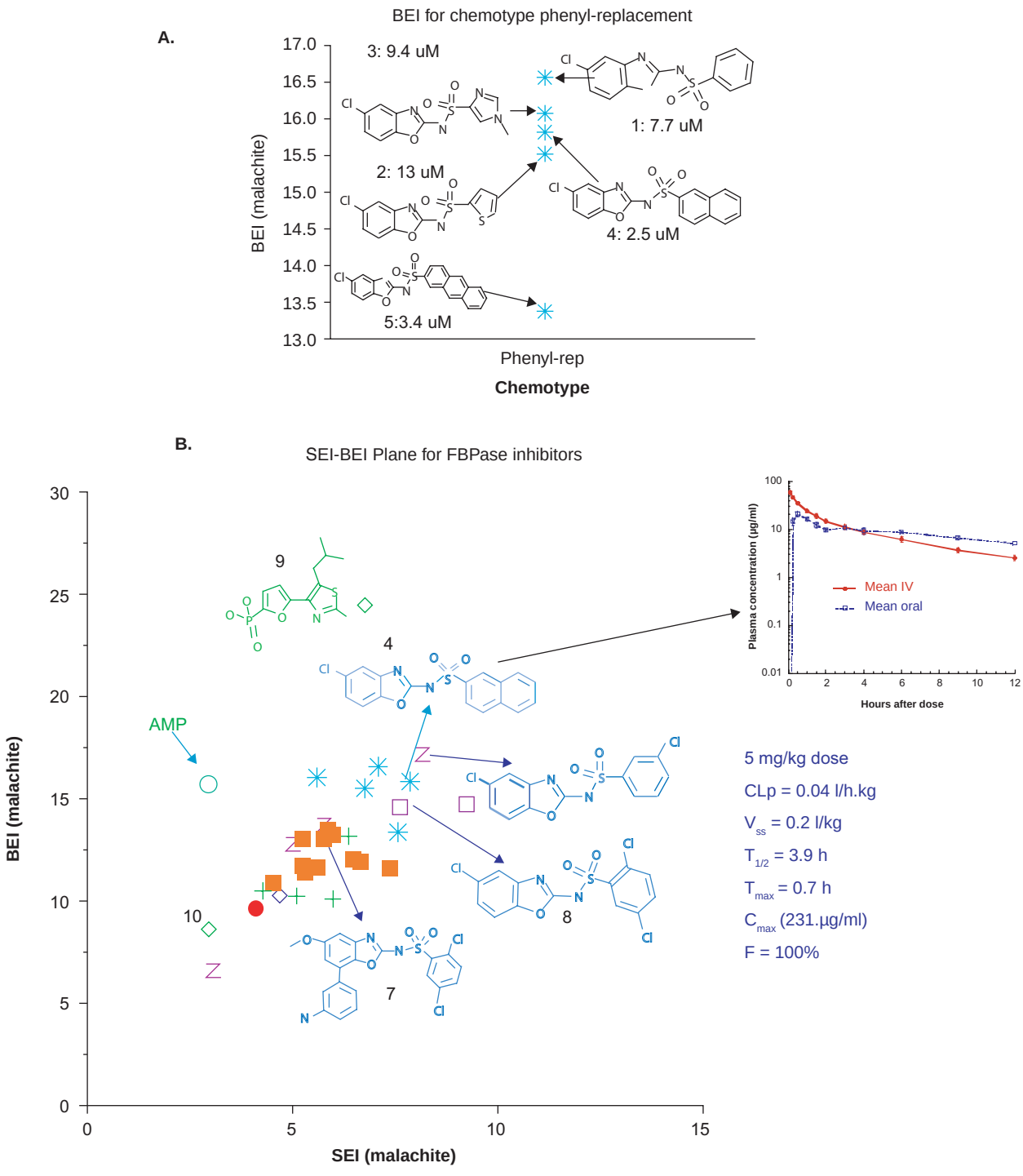


Figure 8. BEI-SEI analysis of the of the FBPase inhibitors. A. Detail of the compounds included in the phenyl-replacement class (blue asterisks, **Table 3** except compounds 1a – b) separated in terms of the different values of BEI. All the FBPase inhibitors discussed are Lipinski compliant. **B.** Bi-variate analysis of the different classes in the optimization plane. The symbols are all the same throughout. 4, 5, 6, 7-substituted benzoxazoles: filled red circles, green crosses, blue open diamonds, filled red squares, respectively. Phenyl substitutions (z), as well as a few reference compounds (AMP, **Figure 6**, **Table 3**) are also represented. Numbers next to compounds refer to **Figure 6** and **Table 3**.

The statistical analysis and the plots were carried out using JMP version 5.1.1 (SAS Institute 1989 – 2004).
BEI: Binding efficiency index; FBPase: Fructose-1,6-bisphosphatase; SEI: Surface efficiency index.

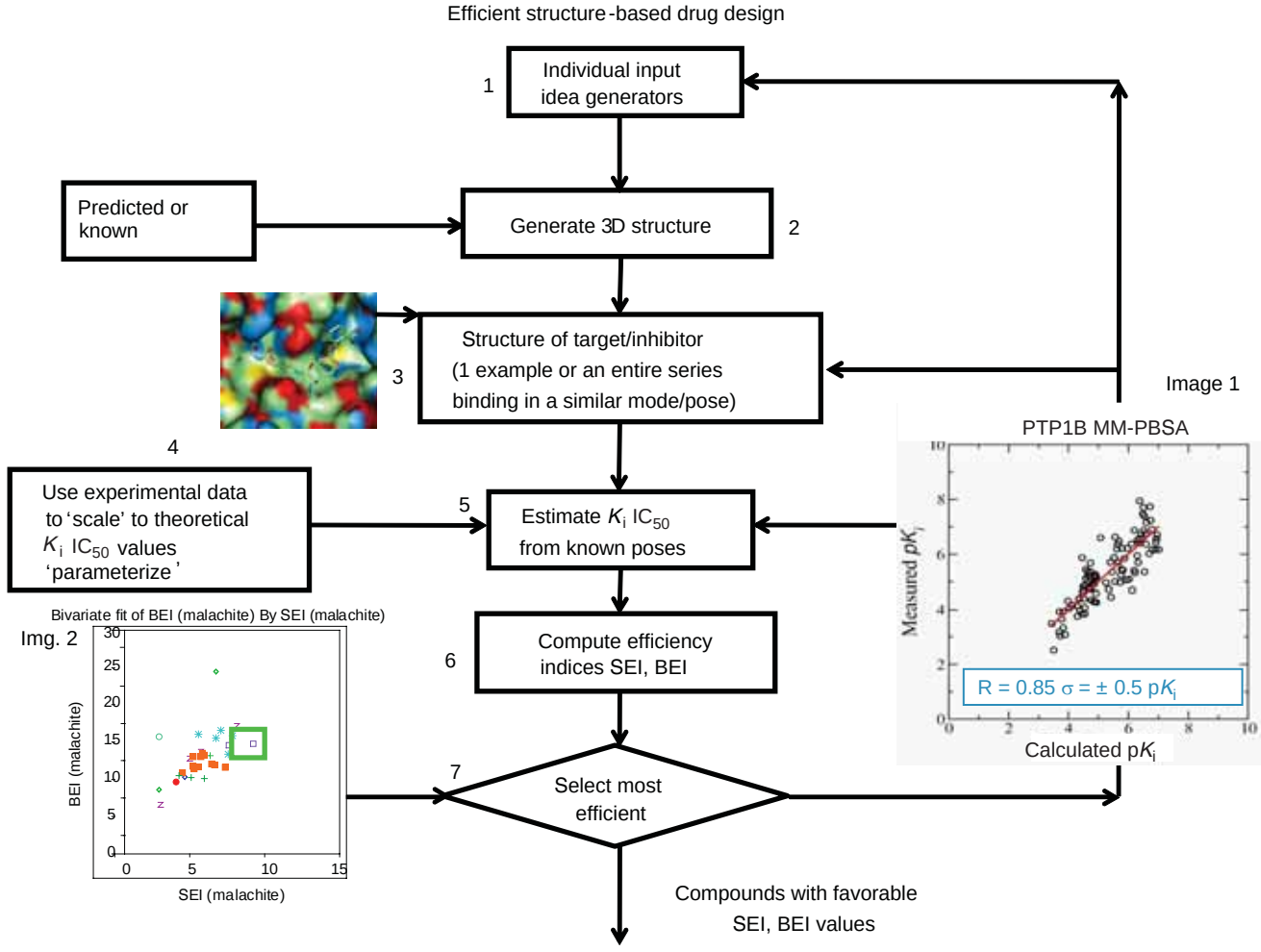


Figure 9. Efficient structure-based drug design flow chart. Incorporation of the concepts of BEI and SEI into the overall ligand- or structure-based drug discovery process. Details are discussed in the text. The plot on the right is meant to represent the scaling of the theoretical K_i values to the experimentally measured K_i values as a possible way to improve the relative values of the theoretically computed K_i values for any target. The example is for PTP1B (122 compounds, courtesy of S Muchmore and S Brown. Correlation coefficient, R = 0.85). The plot on the left is a reduced version of **Figure 8B** (above) and is only intended to illustrate the placement of the compound(s) of interest in the SEI, BEI plane.

BEI: Binding efficiency index; PTP1B: Protein tyrosine phosphatase 1B; SEI: Surface efficiency index.

experimentally obtained ones will provide a reasonable (semi-iempirical or parameterized) framework to compare the different potencies (boxes 4 and 5, image 1, Figure 9). Lastly, and this is the key step in the efficiency-driven drug discovery concept, the efficiency indices SEI and BEI are computed and mapped in the SEI-BEI optimization plane (box 6, image 2, Figure 9). Compounds, series and modifications are selected in terms of chemical feasibility or accessibility and simultaneous optimization of SEI and BEI (box 7, Figure 9). An iterative process using these figures of merit should produce highly efficient compounds (per Dalton, per unit of PSA), which will have a higher probability of being selected for further clinical development.

An independent validation of the use of SEI/BEI as guideposts for future drug discovery might be to document and analyze by rigorous statistical criteria the distribution of (SEI, BEI) values for all the current drugs in the market, related to their corresponding targets, and to follow the trends as the candidates moved through the different stages of drug discovery and development. Knowing the position of the centroids of the two-dimensional distributions for various targets and the associated dispersions (related to step 7 and image 2, Figure 9) would aid in navigating the turbulent waters of drug discovery with a more sound, statistically based, compass.

Irrespective of whether the initial leads are 'lead like' or 'drug like' the efficiency indices could provide a numerical guide of where the compounds initially are, as given by their SEI/BEI values, which direction they need to move and the distance that they need to 'travel' in the plane to reach

optimum values. As the software tools, computational algorithms combined with more available data and speed of the different steps in the outlined process improve, so will the predictive value and power of the overall approach suggested in Figure 9.

10. Conclusion

Ligand efficiency was originally formulated to provide a simple way of assessing the quality of hits in HTS efforts. Extensions of this concept beyond the efficiency per mass unit (LE, BEI) to efficiency per unit of polar surface area (SEI) have been suggested, in an attempt to provide more rigorous and encompassing parameters to guide and quantify the drug discovery process. These parameters have not been used in any prospective manner yet to provide any conclusion as to the value of the original concept. However, retrospective analysis on existing proprietary databases are beginning to appear that suggest the use of these indices to analyze the optimization process from fragment to lead, to clinical candidate. In addition, conventional SAR data can be reframed into a combined SEI/BEI optimization pathway that could be used prospectively in future SBDD projects and drug discovery strategies in general. The seeds are still germinating or possibly are beginning to sprout. More time and nurturing are needed to see what emerges from the soil.

11. Expert opinion

The notion that small and hydrophobic chemical entities make good drugs has been part of medicinal chemistry probably since Paul Ehrlich introduced the concept of chemotherapy at the beginning of the twentieth century. Nowadays, the relative facility associated with modern methods of combinatorial chemistry, improved methods of chemical synthesis and the wide availability of *in vitro* assays has resulted in an explosion of data relating chemical entities to enzymatic activity. This bewildering amount of data drives the vast majority of drug discovery projects from the initial HTS campaigns, to lead discovery, lead validation and optimization and onto preclinical candidates. It would appear that criteria better than the intuition of the experienced medicinal chemists are needed to guide the drug discovery process in an effective manner.

Initially developed to understand and quantify the binding affinity of compounds in relation to the their number of non-hydrogen bonds, the concepts related to ligand efficiency are beginning to take hold in some areas related to medicinal chemistry. In different formulations, the size-related ligand efficiencies indices (LE, BEI; potency/Dalton) are beginning to be used for selecting hits for further elaboration into leads, to identify optimum fragments for further fragment-based drug design and some possible trends have been suggested as important during the optimization process.

Less extended is the use of efficiency as related to PSA (SEI; potency/unit of PSA) as it represents a novel concept that might take a longer time to permeate the medicinal chemistry community and whose merit has still not been proved. A critical finding for its acceptability would be to unambiguously show that there is a strong correlation between high SEI values and favorable PK properties for different targets. Retrospective analysis of lead optimization pathways and SBDD efforts are beginning to uncover suggestive relationships in terms of pK_D, MW and ligand efficiencies. If confirmed, these observed trends could guide future drug discovery efforts in a more direct and effective manner by using continuous scales along the two critical directions of drug discovery: molecular size and solubility. Most likely, other efficiency indices related to cell or *in vivo* activity/toxicity will be developed in an analogous form, which could be incorporated as additional variables to be optimized in the future by analytical, statistical or numerical methods. A most pressing question that needs to be answered rigorously is whether any combination of these indices can be used as a sound predictor of the biological issues related to pharmacokinetics, metabolism or toxicity.

Disguised in different mathematical formulations, efficiency indices are here to stay especially as they relate to experimental pK_i values. They will provide simple, easy to follow indicators along the preclinical drug discovery path. The fundamental question that needs to be answered as rigorously as possible is whether they will turn out to be simple numerical rules-of-thumb or whether, by the virtue of an optimal formulation or definition, they will become the numerical foundation for a more quantitative and statistically sound treatment of drug discovery in the future. More questionable is their utility as guideposts when derived from theoretical estimates of K_i. A critical development for this dream to materialize in the next decade would be the possibility of estimating binding affinities (as K_i values or analogous quantities) between the ligand and the target with improved accuracy at least within congeneric species. This milestone would permit a more predictive, quantitative, effective and efficient drug discovery.

Acknowledgements

The author would like to thank S Muchmore and S Brown for the plot of the theoretical versus experimental K_i values of PTP1B included in Figure 9. The author also wishes to acknowledge discussions and critical reading of earlier versions of the manuscript by J Greer, P Hajduk, CT Lin, K Longenecker, Y Martin, J Metz, G Stamper, K Stewart, V Stoll, B Szczepankiewicz, TW Von Geldern, H Zhao and I Akritopoulou-Zanze. The suggestions of the referees to improve the manuscript are also appreciated.

Bibliography

1. CAMPBELL SF: Science, art and drug discovery: a personal perspective. *Clin. Sci. (Lond)* (2000) 99(4):255-260.
2. CONGREVE M, MURRAY CW, BLUNDELL TL: Structural biology and drug discovery. *Drug Discov. Today* (2005) 10(13):895-907.
3. TINTELNOT-BLOMLEY M, LEWIS RA: A critical appraisal of structure-based drug design. *IDrugs* (2006) 9(2):114-118.
4. WALSH MA, DEMENTIEVA I, EVANS G, SANISHVILI R, JOACHIMIAK A: Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr. D Biol. Crystallogr.* (1999) 55(Pt. 6):1168-1173.
5. DAUTER Z: Current state and prospects of macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* (2006) 62(Pt. 1):1-11.
6. VEDADI M, NIESEN FH, ALLALI-HASSANI A *et al.*: Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. USA* (2006) 103(43):15835-15840.
7. ADAMS MD, CELNIKER SE, HOLT RA *et al.*: The genome sequence of *Drosophila melanogaster*. *Science* (2000) 287(5461):2185-2195.
8. JIANG J, SWEET RM: Protein Data Bank depositions from synchrotron sources. *J. Synchrotron Radiat.* (2004) 11(Pt. 4):319-327.
9. BERMAN HM, WESTBROOK J, FENG Z *et al.*: The Protein Data Bank. *Nucleic Acids Res.* (2000) 28(1):235-242.
10. KUNTZ ID, CHEN K, SHARP KA, KOLLMAN PA: The maximal affinity of ligands. *Proc. Natl. Acad. Sci. USA* (1999) 96(18):9997-10002.
11. HOPKINS AL, GROOM CR, ALEX A: Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* (2004) 9(10):430-431.
12. ABAD-ZAPATERO C, METZ JM: Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* (2005) 10(7):464-469.
13. PALM K, STENBERG P, LUTHMAN K, ARTURSSON P: Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* (1997) 14(5):568-571.
14. CONGREVE M, CARR R, MURRAY C, JHOTI H: A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* (2003) 8:876-877.
15. ABAD-ZAPATERO C, STAMPER GF, STOLL VS: Synergistic use of protein-crystallography and solution-phase NMR spectroscopy in structure-based drug design: strategies and tactics. In: *Fragment-based Approaches in Drug Discovery*. Jahnke W, Erlanson DA (Eds), Wiley-VCH Verlag GmbH & Co., Weinheim, Germany (2006):249-266.
16. SCHNECKE V, BROSTROM J: Computational chemistry-driven decision making in lead generation. *Drug Discov. Today* (2006) 11:43-50.
17. WUNBERG T, HENDRIX M, HILLISCH A *et al.*: Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discov. Today* (2006) 11(3-4):175-180.
18. KESERU GM, MAKARA MG: Hit discovery and hit-to-lead approaches. *Drug Discovery Today* (2006) 11(15/16):741-748.
19. LEACH AR, HANN MM: Fragment screening: an Introduction. *Mol. Biosyst.* (2006) 2:429-446.
20. REES DC, CONGREVE M, MURRAY CW, CARR R: Fragment-based lead discovery. *Nat. Rev. Drug Discov.* (2004) 3:660-672.
21. CARR RA, CONGREVE M, MURRAY CW, REES DC: Fragment-based lead discovery: leads by design. *Drug Discov. Today* (2005) 10(14):987-992.
22. HAJDUK PJ, SHEPPARD G, NETTESHEIM DG *et al.*: Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* (1997) 119:5818-5827.
23. JAHNKE W, ERLANSON DA: *Fragment-based Approaches in Drug Discovery*. Jahnke W, Erlanson DA (Eds), Wiley-VCH Verlach GmbH, Weinheim, Germany (2006).
24. LEACH AR, SHOICHET BK, PEISHOFF CE: Prediction of protein–ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* (2006) 49(20):5851-5855.
25. HAJDUK PJ: Fragment-based drug design: how big is too big? *J. Med. Chem.* (2006) 49(24):6972-6976.
26. OPREA TI, DAVIS AM, TEAGUE JJ, LEESON PD: Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comput. Sci.* (2001) 41:1308-1315.
27. PROUDFOOT JR: Drugs, leads, and drug-likeness: an analysis of some recently launched drugs. *Bioorg. Med. Chem. Lett.* (2002) 12(12):1647-1650.
28. ROEHRIG S, STRAUB A, POHLMANN J *et al.*: Discovery of the novel antithrombotic agent 5-chloro-*N*-({(5S)-2-oxo-3-[4-(3-oxomorpholin-4-yl)phenyl]-1,3-oxazolidin-5-yl}methyl)thiophene-2-carboxamide (BAY 59-7939): an oral, direct Factor Xa inhibitor. *J. Med. Chem.* (2005) 48(19):5900-5908.
29. MESTRES J, VEENEMAN GH: Identification of 'latent hits' in compound screening collections. *J. Med. Chem.* (2003) 46(16):3441-3444.
30. BEMIS GW, MURCKO MA: The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* (1996) 39(15):2887-2893.
31. VIETH M, SUTHERLAND JJ: Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* (2006) 49(12):3451-3453.
32. MORPHY R: The influence of target family and functional activity on the physicochemical properties of pre-clinical compounds. *J. Med. Chem.* (2006) 49(10):2969-2978.
33. BLANEY J, NIENABER V, BURLEY SK: Fragment-based lead discovery and optimization using X-ray crystallography, computational chemistry, and high-throughput organic synthesis. In: *Fragment-based Approaches in Drug Discovery*, Jahnke W, Erlanson DA (Eds), Wiley-VCH, Weinheim, Germany (2006):215-248.
34. HAJDUK PJ, HUTH JR, FESIK SW: Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* (2005) 48(7):2518-2525.
35. SADOWSKI J, KUBINYI H: A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* (1998) 41(18):3325-3329.
36. FRIMURER TM, BYWATER R, NAERUM L, LAURITSEN LN, BRUNAK S: Improving the odds in discriminating 'drug-like' from 'non drug-like' compounds. *J. Chem. Inf. Comput. Sci.* (2000) 40(6):1315-1324.

37. WARREN GL, ANDREWS CW, CAPELLI AM *et al.*: A critical assessment of docking programs and scoring functions. *J. Med. Chem.* (2006) 49(20):5912-5931.
38. PEARLMAN DA: Evaluating the molecular mechanics poisson-Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* (2005) 48(24):7796-7807.
39. AKRITOPOULOU-ZANZE I, DARCZACK D, SARRIS K *et al.*: Scaffold oriented synthesis. Part 1: design, preparation, and biological evaluation of thienopyrazoles as kinase inhibitors. *Biorg. Med. Chem. Lett.* (2006) 16:96-99.
40. SHERMAN W, BEARD HS, FARID R: Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des.* (2006) 67(1):83-84.
41. SHERMAN W, DAY T, JACOBSON MP, FRIESNER RA, FARID R: Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* (2006) 49(2):534-553.
42. IRWIN JJ, RAUSHEL FM, SHOICHET BK: Virtual screening against metalloenzymes for inhibitors and substrates. *Biochemistry* (2005) 44:12316-12328.
43. PEI Z, LIU G, LUBBEN TH, SZCZEPANKIEWICZ BG: Inhibition of protein tyrosine phosphatase 1B as a potential treatment of diabetes and obesity. *Curr. Pharm. Des.* (2004) 10(28):3481-3504.
44. SZCZEPANKIEWICZ BG, LIU G, HAJDUK PJ *et al.*: Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* (2003) 125(14):4087-4096.
45. XIN Z, OOST TK, ABAD-ZAPATERO C *et al.*: Potent, selective inhibitors of protein tyrosine phosphatase 1B. *Bioorg. Med. Chem. Lett.* (2003) 13(11):1887-1890.
46. LIU G, SZCZEPANKIEWICZ BG, PEI Z *et al.*: Discovery and structure-activity relationship of oxalylarylaminobenzoic acids as inhibitors of protein tyrosine phosphatase 1B. *J. Med. Chem.* (2003) 46(11):2093-2103.
47. LIU G, XIN Z, LIANG H *et al.*: Selective protein tyrosine phosphatase 1B inhibitors: targeting the second phosphotyrosine binding site with non-carboxylic acid-containing ligands. *J. Med. Chem.* (2003) 46(16):3437-3440.
48. PEI Z, LI X, LIU G *et al.*: Discovery and SAR of novel, potent and selective protein tyrosine phosphatase 1B inhibitors. *Bioorg. Med. Chem. Lett.* (2003) 13(19):3129-3132.
49. LIU G, XIN Z, PEI Z *et al.*: Fragment screening and assembly: a highly efficient approach to a selective and cell active protein tyrosine phosphatase 1B inhibitor. *J. Med. Chem.* (2003) 46(20):4232-4235.
50. XIN Z, LIU G, ABAD-ZAPATERO C *et al.*: Identification of a monoacid-based, cell permeable, selective inhibitor of protein tyrosine phosphatase 1B. *Bioorg. Med. Chem. Lett.* (2003) 13(22):3947-3950.
51. ZHAO H, LIU G, XIN Z *et al.*: Isoxazole carboxylic acids as protein tyrosine phosphatase 1B (PTP1B) inhibitors. *Bioorg. Med. Chem. Lett.* (2004) 14(22):5543-5546.
52. IVERSEN LF, ANDERSEN HS, BRANNER S *et al.*: Structure-based design of a low molecular weight, nonphosphorus, nonpeptide, and highly selective inhibitor of protein-tyrosine phosphatase 1B. *J. Biol. Chem.* (2000) 275(14):10300-10307.
53. ANDERSEN HS, IVERSEN LF, JEPPESEN CB *et al.*: 2-(Oxalylamino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases. *J. Biol. Chem.* (2000) 275(10):7101-7108.
54. SZCZEPANKIEWICZ BG, LIU G, HAJDUK PJ *et al.*: Discovery of a potent, selective protein tyrosine phosphatase 1B Inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* (2003) 125:4087-4096.
55. MARTIN YC: A bioavailability score. *J. Med. Chem.* (2005) 48(9):3164-3170.
56. WISHART DS, KNOX C, GUO AC *et al.*: DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* (2006) 34(Database issue):D668-D672.
57. VON GELDERN TW, LAI CH, GUM RJ *et al.*: Benzoxazole benzesulfonamides are novel allosteric inhibitors of fructose-1,6-bisphosphatase with a distinct binding mode. *Biorg. Med. Chem. Lett.* (2006) 16:1811-18115.
58. LAI CH, GUM RJ, DALY M *et al.*: Benzoxazole benzenesulfonamides as allosteric inhibitors of fructose-1,6-bisphosphatase. *Biorg. Med. Chem. Lett.* (2006) 16:1807-1810.
59. STEWART KD, SHIRODA M, JAMES CA: Drug Guru: a computer software program for drug design using medicinal chemistry rules. *Bioorg. Med. Chem.* (2006) 14(20):7011-7022.
60. WENLOCK MC, AUSTIN RP, BARTON P, DAVIS AM, LEESON PD: A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* (2003) 46(7):1250-1256.
61. VIETH M, SIEGEL MG, HIGGS RE *et al.*: Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* (2004) 47(1):224-232.
62. LIPINSKI CA: Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* (2000) 44(1):235-249.
63. VEBER DF, JOHNSON SR, CHENG HY *et al.*: Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* (2002) 45(12):2615-2623.

Website

101. http://www.imca.aps.anl.gov Industrial macromolecular crystallography association web site (2006).

Affiliation

Cele Abad-Zapatero
Center for Pharmaceutical Biotechnology,
Molecular Biology Research Building,
Room 3152, University of Illinois at Chicago,
900 So. Ashland St. m/c 870, Chicago,
IL 60607-7173, USA
Tel: +1 312 355 4105; Fax: +1 312 413 9303;
E-mail: caz@uic.edu