Drug Discovery Today • Volume 00, Number 00 • August 2010 PERSPECTIVE

ELSEVIER

feature

Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation

Cele Abad-Zapatero[1,*], Ognjen Perišić[1,2], John Wass[3], A. Patrícia Bento[4], John Overington[4], Bissan Al-Lazikani[4,5] and Michael E. Johnson[1]

We propose a numerical framework that permits an effective atlas-like representation of chemico-biological space based on a series of Cartesian planes mapping the ligands with the corresponding targets connected by an affinity parameter (K_i or related). The numerical framework is derived from the concept of ligand efficiency indices, which provide a natural coordinate system combining the potency toward the target (biological space) with the physicochemical properties of the ligand (chemical space). This framework facilitates navigation in the multidimensional drug discovery space using map-like representations based on pairs of combined variables related to the efficiency of the ligands per Dalton (molecular weight or number of non-hydrogen atoms) and per unit of polar surface area (or number of polar atoms).

Introduction

Successful drug discovery requires the optimization of a large number of variables in two different domains: chemical and biological. The range of variables expands from the strictly physicochemical properties of the ligand, such as molecular weight (MW), to the more complex variables related to its affinity toward the target and to bioavailability and toxicity in the clinic. Navigating through this N-dimensional space using the potency of the ligand toward the target (i.e. K_i, IC_{50} or related metrics) might result in good inhibitors but does not necessarily result in successful drugs in the clinic. The impact of the advances in the related fields of molecular and structural biology, assay development, human genomics, proteomics, fragment-based drug discovery, and other technologies on this process, although noticeable, is still far from dramatic [1,2].

Conceptually, the introduction of filtering criteria – such as the rule of five [3] a decade ago and the rule of three in fragment-based strategies [4] – was expected to expedite the discovery of chemical entities with effective potency in the clinic. Although this expectation has been fulfilled in certain cases, the effectiveness of the entire drug discovery process has been, and remains, under a degree of scrutiny [2,5].

A more streamlined and effective framework is needed to facilitate and improve the drug discovery process. Toward that ultimate goal and for several years now, the number of databases dedicated to storing and annotating target–ligand complexes (SAR databases) has grown dramatically as a natural extension of the information contained at the Protein Data Bank [6]. Currently, PDBBind [7,8], BindingDB [9,10], DrugBank [11], MOAD [12,13], and WOMBAT [14] and ChEMBL [15] (http://www.ebi.ac.uk/chembldb), among others, have extended dramatically the number of observations in chemico-biological space (CBS). These resources have amassed vast amounts of high-quality data connecting ligands to their targets (PDBBind, BindingDB, MOAD and ChEMBL) complemented, when available, with structural information and relating them to relevant therapeutic entities (DrugBank; http://www.redpoll.pharmacy.ualberta.ca/drugbank).

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2010.08.004

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

The vastness of chemical space [16] and certain tools to navigate it [17] have already been presented and discussed. Briefly, chemical space has previously been considered as being analogous to the cosmological universe with chemical compounds populating the space instead of stars [16]; however, the existence of this vast and ever-increasing amount of data combining compounds and targets mentioned previously, without sound underlying principles, has been compared to the astronomy before the formulation of Kepler's Laws of planetary motion [18]. Is it possible to find certain generalizations that can effectively describe and represent CBS? Can those insights be translated into a more efficient drug discovery process?

Two conceptual trends in drug discovery that have emerged in the past decade could be combined to provide an effective mapping of CBS. One is the concept of 'chemogeography' and the corresponding chemical global positioning system (ChemGPS) [17] as an aid to navigate chemical space. The other is the introduction of combined variables referred to as ligand efficiency (LE) or ligand efficiency indices (LEIs) to relate the potency of compounds to their size and polarity [19,20].

Any integration of chemical space with biological (or affinity) space raises many important issues related specifically to biological data that should be mentioned upfront. A compound's physiochemical properties can be easily measured or calculated and, in addition, are very well defined. By contrast, the biological data associated with the target, in particular those data related to ligand–target binding affinity, should be always be considered with caution. Variation in the apparent affinity of a ligand toward its target can be strongly dependent upon specific assay details, reproducibility within a specific assay, variability between different assays, sample size and species variation, among other things.

In spite of the above caveats, we propose that a combined representation of CBS has merit. In science and technology, the introduction of new variables is often a first step in helping to provide a new framework that can be effective in solving long-standing problems. We suggest that LEIs can play such a part in drug discovery. They naturally connect the physicochemical properties of the ligand (e.g. MW and polar surface area [PSA], among others) with the biological target via the affinity parameter K_i (or other related measures of affinity).

In this article, we briefly summarize the earlier concepts used to navigate chemical space and present an algebraic framework based on LEIs that permits an intuitive, graphical representation of the combined CBS. We explain the appearance and overall properties of this representation and provide some examples of the use of this framework in drug discovery. It is presented as an initial concept for future application and development. How this framework and graphical representation could facilitate drug discovery in the future will require its application to specific examples in a rigorous, prospective and predictive manner.

Background concepts

Chemical space

The concept of chemogeography and the software navigating system ChemGPS were introduced to relate chemical space to drug space (i.e. subset of compounds that correspond to marketed drugs). The coordinates of ChemGPS maps were based on the *t*-scores derived from principal component analysis of 72 descriptors of various physicochemical and topological properties of the chemical entities [17]. From the drug discovery perspective, the challenge is to identify regions of chemical space that contain biologically active compounds for particular biological targets [16]. Lipinski and Hopkins hypothesized that within the continuum and vastness of chemical space, there would be enclosed, discreet regions occupied by compounds with specific affinities towards particular biological targets or gene families (i.e. proteases, kinases, GPCRs and others). The question of which variables (or coordinate systems) would enable such segregation, however, was left open.

Efficiency indices

A first attempt to quantify the binding affinity of a ligand in relation to the number of non-hydrogen atoms was presented by Kuntz *et al.* [21]. In an extension of their original concept, Hopkins *et al.* [19] defined LE numerically as the quotient between ΔG and the number of non-hydrogen atoms of the compound:

$$\mathrm{LE} = \Delta g = \frac{\Delta G}{\mathrm{NHEA}} \qquad (1)$$

where $\Delta G = -RT \ln K_i$ and NHEA is the number of non-hydrogen atoms (or heavy atom count). Dimensional analysis showed the units of LE to be kcal/mol per non-hydrogen atom (Table 1). Variations and extensions of this concept continue to appear in the literature [22] and are gaining a much wider acceptance among medicinal chemists [23]. In particular, the concept of 'group efficiency' has been proposed as an extension of the original idea to estimate the binding efficiency of parts of a molecule or of groups added to an existing fragment or lead molecule [24].

Because non-hydrogen atoms can be of many different types and a key property of a compound is its MW, a natural extension of the concept of LE (Eq. (1)) was introduced soon thereafter, which excluded the energy units. Binding efficiency index (BEI) is a simpler index

TABLE 1

Names, definitions and idealized reference values for various definitions of ligand efficiencies.

Name	Definition	Example value[a]	Eq. no.
LE[b,c]	ΔG/NHEA (non-hydrogen atoms)	−0.50[c]	(1)
BEI	p(K_i), p(K_d), or p(IC_{50})/MW (kDa)	27	(2)
SEI	p(K_i), p(K_d), or p(IC_{50})/(PSA/100 Å^2)	18	(3)
NSEI	NSEI = $-\log_{10} K_i$/(NPOL) = pK_i/NPOL(N,O)	1.5	(4)
NBEI	NBEI = $-\log_{10} K_i$/(NHEA) = pK_i/(NHEA)	0.36	(5)
***n*BEI**[d]	*n*BEI = $-\log_{10}[(K_i/\mathrm{NHEA})]$	10.25	(6)
***m*BEI**	*m*BEI = $-\log_{10}[(K_i/\mathrm{MW})]$	11.5	(7)

By definition, for any given compound the ratio of BEI/SEI is equal to 10(PSA/MW) or NBEI/NSEI = NPOL(N,O)/NHEA(non-H) = 0.36/1.5 = 6/25 = 0.24.

[a] Reference values and definitions of BEI,SEI adapted from Ref. [29]. K_i or IC_{50} = 1.0 nM; molecular weight = 333 Da (or 0.333 kDa). This value of MW is near the mean value of MW for a large sample of marketed oral drugs [31,32]. Van der Waals PSA = 50 Å^2. For the atom-related definitions, LEIs are calculated for each index using the following idealized values (units have been omitted in the table): K_i or IC_{50} = 1.0 nM; p(K_i) = –log K_i = 9.00; MW = 0.333 kDa; NHEA = number of heavy atoms (non-hydrogen in the compound) = 25; NPOL = number of polar atoms (N,O) = 6.

[b] $\Delta G = -12.4$ kcal/mol, $\Delta G = -RT \ln K_i$, assuming K_i = 1.0 nM, T = 300 K; NHEA (non-hydrogen atoms) = 25; corresponding to a mean MW/atom of 13.3 Da.

[c] Hopkins *et al.* [19].

[d] This small change in the definition of NBEI (taking the $\log_{10}$ after the ratio of (K_i/NHEA) was found to be crucial for the separation of the compounds in the *n*BEI–NSEI plane (see text and Supplementary material II). Similarly, for *m*BEI–NSEI pair (Eqs. (7) and (9)).

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

BOX 1

Simple is beautiful

The information necessary to map compounds, chemical series and efficiency data on efficiency space is readily available using conventional software tools. The overall concept is summarized in Fig. 3.

- Given the empirical formula or SMILES string of a compound and a measure of affinity (K_i) or activity (IC_{50}), extract the following information. NPOL (number of N, O atoms), NHEA (number of non-hydrogen atoms), MW and tPSA (or any other estimate of the PSA). ($pK_i = -\log_{10} K_i$).
- Calculate the different LEIs as defined in Table 1. Create a spreadsheet with all the compounds including 'reference' compounds and/or compounds from other groups.
- Plot: NSEI, *n*BEI (*x*, *y*). Other LEI planes (NSEI, NBEI; SEI, and BEI) can also be used for fine-tuning the analysis. In NSEI–*n*BEI (and NSEI–*m*BEI), the compounds will appear along a series of lines with slopes equal to NPOL.
- If the reference compound has N polar atoms, it will appear on the corresponding NPOL line. Highly polar, very size-efficient compounds will plot to the left (in the upper left part) of the plot; compounds that are very potent and very hydrophobic will map to the lower right.
- The line NPOL = *N* basically divides the plane in two halves. Compounds with NPOL = *N* + 1 will map to the left of the reference line; compounds with NPOL = *N* − 1 will map to the right of the corresponding line.
- The position of a 'hypothetical' compound with several polar atoms different from N will map in the corresponding part of the map [30], in relation to the starting compound with N polar atoms (N + O). If no affinity data are available, the compound cannot be placed uniquely along the new NPOL line; once a measurement (or theoretical estimate) has been obtained, the target–ligand pair can be located unambiguously in the plane (Fig. 3 legend: solid arrows *versus* dashed arrows).
- Strategy: try to optimize (maximize) both variables together in the plane insofar as possible. Compounds or series that move towards the diagonal of the plot optimize both variables simultaneously.
- Series with the same number of polar atoms (irrespective of where they occur in the structure of the ligand!) will map along the corresponding NPOL line, as determined by the corresponding potency: more potent compounds will move up; less potent compounds will move down, along the line (Fig. 3). The most effective way to explore alternative regions of the plane is probably to change the number of polar atoms of the compound and/or series by using 'replacement' rules [30]. This is equivalent to jumping from 'line-to-line' in the NSEI–*n*BEI (NSEI–*m*BEI) plane ('line hopping').
- For certain targets and chemical series, when an optimum compromise has been found for both NSEI and *n*BEI (NSEI–*m*BEI) and solubility issues are encountered, the best strategy might be to move to the left (increase solubility by adding polar atoms), but the efficiency per atom added (*n*BEI, *m*BEI) should be retained or increased as much as possible.

based on the ratio of the binding affinity given as $pK_i = -\log K_i$, using MW as reference expressed in kiloDaltons (Eq. (2), Table 1; in addition, see Ref. [2] for further details). This concept of ligand efficiency in various definitions is gaining wider acceptance in the medicinal chemistry literature (see Refs. [24–26] and references therein), especially in relation to guiding fragment-based strategies to expedite drug discovery [2,27]. It has also been used to dissect the most efficient fragment in the natural product argifin, a cyclopentapeptide inhibitor of chitinase activity [28].

Although it has always been considered to be of great importance, much less direct use has been made of the concept of ligand efficiency related to compound polarity. A specific definition of this concept (the surface efficiency index, or SEI) has been defined using the affinity of the ligand (pK_i) and the PSA of the ligand scaled to 100 $Å^2$ [20,29]. We propose that effective use of the concept of LEIs should involve the use of multiple, non-redundant variables, within a numerical formulation that uses similar numerical scales. It is likely that the combination of several complementary LEIs will have the highest predictive value. An initial illustration of the concept of combining SEI-BEI in a Cartesian plane has been published for inhibitors of human protein tyrosine phosphatase 1B (PTP1B) [29].

Selecting the optimum variables to map CBS

Table 1 lists several definitions of LEIs that have been explored and compared in proposing these concepts. Initial analyses favored formulations that directly related potency (K_i, IC_{50} or related measurements) to the MW and the PSA of the ligands, using BEI and SEI, respectively [20]. This indicated that the distribution of chemical compounds in the optimization plane could be understood in terms of a wedge of varying ratios of PSA-MW across the SEI-BEI plane (Fig. 5 of Ref. [29]). In the Supplementary Data associated with this article (section I, Supplementary Fig. S1a), we show confirmation of this assertion using PDBBind data.

To refine and improve on this concept, two formulations related to the heavy atom count [19] and the number of polar atoms (NPOL = -number of O and N) were examined. Such a framework has the advantages that medicinal chemists find it intuitive and it relates better to previous work in the field [19,22]. The working definitions of LEIs related to the number and type of atoms present in the ligand that have been explored in this work are shown in Table 1. An example of the distribution of target–ligand pairs in the 2007 version of PDBBind using variables NSEI–NBEI is shown in the Supplementary Data (section I, Supplementary Fig. S1b). This list cannot be considered to be exhaustive and, indeed, it is possible that additional LEIs will be defined in the future to optimize the drug discovery process. It will be desirable, however, that all can be related to one another in similar scales to provide additional dimensions for future numerical or statistical optimization.

Although the representation of CBS given by the planes SEI–BEI and NSEI–NBEI can be useful for certain applications, we would like to emphasize the representation in Cartesian planes defined by the variables NSEI–*n*BEI and NSEI–*m*BEI (*x*–*y*; Table 1 and Fig. 1a and b, respectively). These figures depict the distribution of the PDBBind database compounds in the planes defined by variables *n*BEI *versus* NSEI and *m*BEI *versus* NSEI. The plots correspond to the distribution of lines in the plane described by Eqs. (8) and (9), respectively, as shown in the Supplementary Data, Section II (Supplementary Box S1). The change in the

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

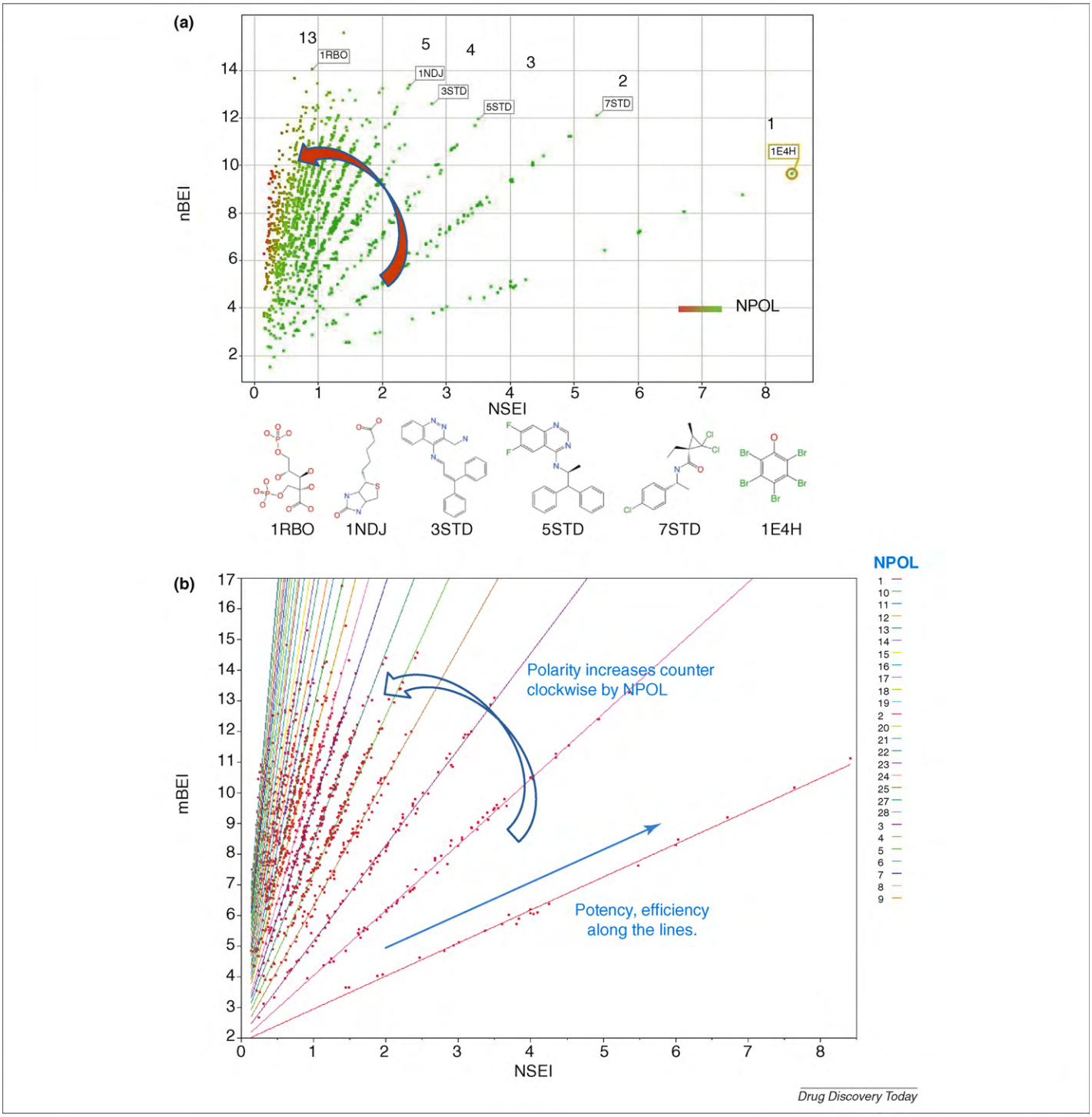

FIGURE 1

Representation of the content of PDBBind in a Cartesian plane defined by LEIs related to the number of atoms of the ligand. (a) Representation in a Cartesian (atlas-like) map of the entries in the PDBBind (2007) refined set in the *n*BEI–NSEI plane. Each point corresponds to a target–ligand complex (one PDB access code, 1283 entries), and the slope of each line corresponds to a certain number of polar atoms in the ligand, as indicated (upper right-hand panel). Intercept is related to the number of heavy (non-hydrogen) atoms in the ligand: $\log_{10}$(NHEA). NPOL range 1–28 (see Table 1 for definitions and Box 1 for the algebraic derivations). The chemical structures of the high efficiency ligands in the data set (PDB accession codes highlighted) have been shown below to show the progressive increase in the number of polar atoms and polarity of the ligands, as the slope increases counterclockwise. (b) Representation in a Cartesian (atlas-like) map of the entries in the PDBBind (2007) refined set in the *m*BEI–NSEI plane. The same set in the atom-related efficiency indices *m*BEI–NSEI. The intercept is related to the MW of the ligand as $\log_{10}$(MW). The compounds map along lines of slope defined by the number of polar (N and O: NPOL) atoms in the ligand ranging from 1 to 28 (see side panel). The statistical analysis was performed as indicated in the Supplementary Data. NPOL range 1–28 (see side panel). See Table 1 and Supplementary material II, Box 1 for the variable definitions and algebraic derivations. Images prepared with JMP7 (SAS Institute) and Spotfire™.

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

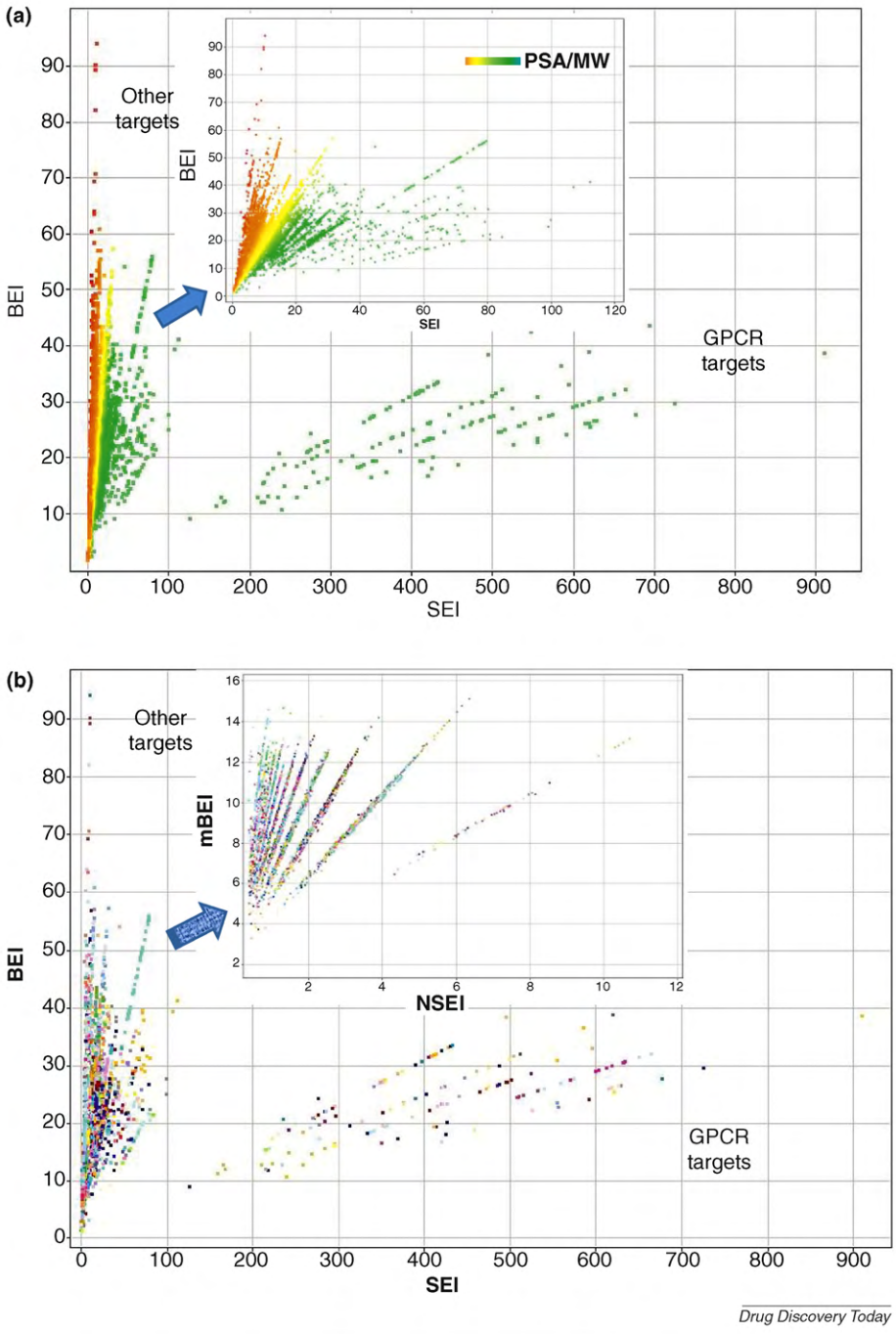

FIGURE 2

Representation of the content of the WOMBAT drug database illustrating the separation between GPCR and non-GPCR targets. (**a**) Separation of the conventional and GPCR targets in an optimization plane using BEI–SEI as variables with a wide range of values: SEI (0, 925); BEI (0, 100). The large values of the extreme target–ligand complexes correspond to very potent and very small (top left, high BEI and low SEI values: glycine bound to NMDA/gly) or very potent and extremely hydrophobic (lower right, high SEI and low BEI: amitriptyline bound to α1 receptor). Data access courtesy of T. Oprea [14] as included in WOMBAT 2007. 'Geographically', the separation between the GPCRs and the more 'conventional' targets in this representation could be equivalent to a major ocean separating two different continents. The inset shows a magnified view of the region corresponding to the conventional targets [SEI (0, 120), BEI (0, 100)] using the same variables BEI–SEI and an adjusted scale to fit into the available space (see Table 1 for definitions). The color gradient in both images is related to 10(PSA/MW) (slope of the lines) from minimum (0.04, dark green), first quartile (1.23), median (1.92), third quartile (2.95) and maximum 11.38, red). The non-GPCR targets account for 90.5% of the sample size. The blue arrow indicates simply a scale change. (**b**) The same data and variable range as before with the various target–drugs complexes colored by target. The inset represents the region corresponding to the conventional targets (as in Fig. 2a) but magnified in a different scale and represented with the variables *m*BEI-NSEI [NSEI: (0, 12), *m*BEI: (2, 16)] to emphasize the separation of the different lines by the number of polar atoms (NPOL) of the ligands. The colors of the different squares correspond to different targets. The textured blue arrow indicates a change in scale and a representation in different variables. Images prepared with Spotfire™.

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

definition of the atom-related related variables (NBEI = $-\log_{10} K_i$/NHEA, as opposed to *n*BEI = $-\log_{10}[(K_i/\text{NHEA})]$; Table 1) and their combination in the NSEI–*n*BEI plane (Eq. (8)) has enabled a clear, fan-like separation of all the compounds in the database in terms of the number of polar atoms (NPOL, slope) and number of heavy atoms (NHEA, intercept: $\log_{10}$(NHEA)). For clarity, the regression analysis of all the lines ranging from NPOL = 1 to NPOL = 28 ($R^2 > 0.99$) are shown only in Fig. 1b (Supplementary Data, Section II, Supplementary Box S2).

A similar result is obtained in the *m*BEI–NSEI representation, as depicted in Fig. 1b where the intercept is $\log_{10}$(MW) (Eq. (9)). This alternative representation enables a better separation of the compounds along the intercept ($\log_{10}$(NHEA) *versus* $\log_{10}$(MW)); Fig. 1a and b). Because of their fan-like nature, we refer to these plots as 'fan-plots' of the database, providing a rapid visual map of the content of any target–ligand database and their relative affinities (as given by K_i or related parameters).

It is important to emphasize that in either of these representations of the contents of PDBBind database, any small ligand can be placed along a specific line in the diagram given only its chemical composition: number of non-hydrogen atoms (NHEA or MW related to the intercept) and number of polar atoms (NPOL or the PSA/MW ratio, related to the slope of the line). Within that line, the unique position of the target–ligand pair on the map is determined only when K_i values (experimental or calculated) are available with respect to a specific target. The same ligand with different affinities towards *N* separate targets will be represented by *N* points located along the line; the higher the potency, the farther away from the origin. In this way, the uncertainty in the affinity parameter(s) indicated before as a weakness in representing biological data will be reflected by a 'sliding' position of the target–ligand pairs along the line, corresponding to the physicochemical properties of the ligand (PSA/MW, NPOL values defining the slope of the lines). Thus, the *x* (SEI-like) and *y* (BEI-like) coordinates of any target–ligand complex in the plane represent the polarity and size-related components, respectively, of the efficiency of that ligand towards the total LEI of the compound.

The concept of an atlas-like representation of CBS

The previous analysis shows that it is possible to represent visually the connection between chemical ligands and biological targets in a two-dimensional, Cartesian plane. Any target–ligand complex in any database (as illustrated for PDBBind and WOMBAT in Figs. 1 and 2) can be represented by a point in the plane, given the physicochemical characteristics of the compound (i.e. PSA, MW or NPOL, NHEA) and its affinity towards a specified target (K_i). Obviously, this can be done by computing the corresponding variable pairs (SEI, BEI; NSEI, NBEI; or *n*BEI, NSEI or other combinations) and plotting

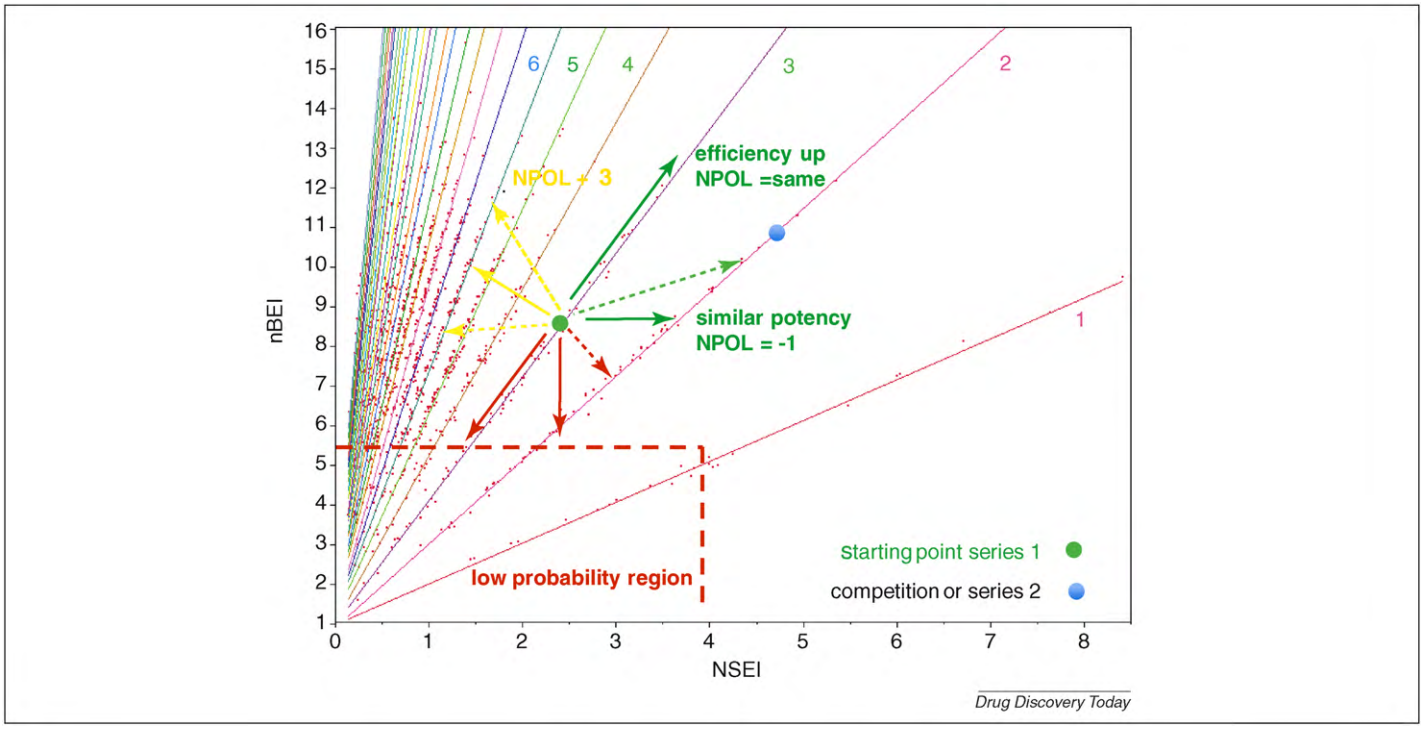

FIGURE 3

Schematic representation of the application of the AtlasCBS concepts in drug discovery (Box 1). The concepts presented and illustrated in this work suggesting a graphical representation to guide drug discovery have been summarized in a schematic representation in the *n*BEI–NSEI plane. The guiding strategies and concepts have been summarized in Box 1. Given the empirical formula of a compound, it can be placed anywhere along a specific line based on NPOL and $\log_{10}$(NHEA). An experimental affinity towards a target (K_i or related) allows placement of the target–ligand complex on a unique point in the plane (green dot). Strategies to modify this starting compound have been outlined in the diagram. Adding three polar atoms will move the compound to the left (yellow arrows, more polar). A solid arrow is meant to indicate an experimental affinity measurement; dashed arrows (three colors) mark the possible trajectory based on theoretical estimates of the binding affinity. Retaining the same number of polar atoms will move the trajectory along the corresponding NPOL line up or down depending on the resulting potency. A blue dot suggests the position of an alternative series or the position of a compound from the competition. The dashed red box outlines the region of the CBS space, where very few successful drugs have been documented. Details on how to use the optimization planes related to LEIs as variables to optimize drug discovery have been published before [29]. The target–ligand complexes included in the refined PDBBind set were statistically modeled by NPOL (Supplementary Data, section II, Box S2) and are included as background for the discussion to simulate other existing complexes.

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

them in a map-like fashion. The collection of maps of CBS in different coordinates (physical and atomic) for different regions of the plane and at different scales is what we would like to call 'AtlasCBS' (an atlas-like representation of CBS). This collection could be conceived as a series of pages (physical and/or electronic) that can be plotted, stored, searched, updated and used as a graphical representation of CBS and as a guide for drug discovery.

In addition to the angular coordinate related exclusively to the properties of the ligand, the proposed two-dimensional representations of CBS also contain information about the biological target, implicitly included in the definitions of the LEIs, via the affinity variable. The relationship to the target in the map is given by the radial distance of the point (representing the target–ligand complex) to the origin, analogous to its radial component (Fig. 2a). Thus, a compound or a series of compounds with nearly identical PSA/MW ratios can be placed anywhere along a certain line in the plane given its PSA/MW value. It is only when a pK_i value is available that it is possible to place unambiguously the compound(s) at a specific point(s) within the line given their PSA/MW ratio. The most polar ligands within the different targets lay on the left, and they become progressively less polar (more drug-like) as the PSA/MW ratio becomes smaller, as shown for an extensive drug discovery effort directed towards human PTP1B (Fig. 5 of Ref. [29]). How potent the compounds are for a specific target is proportional to the distance from the origin along the appropriate line. The power of this representation to illustrate a wide range of physicochemical properties of the ligands, as they relate to the targets, is shown in Fig. 2a. The content of the database WOMBAT has been represented in the BEI–SEI plane using a wide range of values for both variables to illustrate the separation in chemical space between conventional (non-GPCR, upper left) and GPCR targets.

This conceptual separation is still valid for the representations of the optimization plane in efficiency indices related to the atomic composition of the ligands, namely NSEI, NBEI (Supplementary Data, section I) and NSEI, *n*BEI (Fig. 2b, inset). In addition, in the latter plane, the separation of the different lines is more dramatic and more intuitive because the slopes are given by the NPOL of the ligand, and it varies dramatically for the low slope values (1, 2, 3, *etc.*); it approaches a narrowly separated family of lines at high NPOL values.

A word of caution is appropriate here because of the wide variability of the biological data in any biological project, as indicated above. LEIs can be calculated based on K_is, K_ds and IC_{50}s as the most notable in measuring the affinity of the ligands towards the target. For the most part and preferably, the values should be experimental. Theoretical values could be considered only as a guide and approximation (Box 1). In any serious analysis involving LEIs, the values should only be compared if the same variable is used to calculate the relative efficiencies. For comparison of internal compounds against compounds from different or competing groups, the affinity should be measured under the same assay conditions. In addition, this representation of CBS is focused on ligands that inhibit the activity of the biological target (i.e. antagonists of receptors). Possibly, alternative efficiency indices could be devised in the future for agonists and a similar representation could be found. In the future, it might also be possible to add a third dimension related to the efficacy of the ligands in *in vivo* assays (Fig. 3).

The atlas-like representation of CBS can also be related to geographical maps in a different way. The macroscopic variables (PSA, MW) would correspond to exploring the physicochemical properties of the ligands (analogous to mountains, rivers and physical features of geographical maps), and the atomic properties (i.e. NPOL, NHEA) would be analogous to more discrete features of the landscape such as cities, major highways, or even metro stops or blocks at a magnified scale. We present some applications of the use of this atlas-like representation (and related optimization planes) in two areas of drug discovery in the Supplementary Data (section III); probably, others will follow.

Applications

Two examples of the application of the AtlasCBS representation to drug discovery are presented in Supplementary Data (section III): trajectories in drug discovery space, as illustrated for human PTP1B (Supplementary Fig. S2a–d) (application 1) [33–40], and mapping of compounds with inhibitory and pharmacological activity for the HIV-I protease (Supplementary Fig. S3a–c) (application 2) [9,10].

Concluding remarks: a natural representation of CBS

The affinity (or potency) between the ligand and the target (K_i, IC_{50} or related parameters) has probably been the dominant parameter in drug discovery since reliable estimates could be measured *in vitro* or *in cell* assays. The knowledge and experience acquired within the past decade or so regarding the constraints imposed by the physicochemical properties of the ligands for the success of preclinical drug discovery programs has been incorporated into empirical rules of thumb that have guided the discovery effort [3,4]. These guidelines, however, do not consider potency towards the target and thus have only limited predictive value and effectiveness in the tortuous road of drug discovery. We wish to suggest that LEIs offer a more natural and effective combination of variables that can graphically aid in mapping CBS. These combined variables encompass the more established variables (i.e. LogP, MW, and PSA) and provide a graphical representation of the guiding criteria that are being used currently (e.g. the rule of five). We suggest that these novel variables could, in the near future, make the drug discovery process more effective. How these ideas can be incorporated into the overall drug discovery process has been presented before [29].

We suggest that LEIs relating potency to two crucial physicochemical properties of the ligand (i.e. size and polarity) can be defined in two different but related ways: (i) relating to physicochemical properties such as MW and PSA (i.e. BEI, SEI; *m*BEI) and (ii) referring to the atomic composition of the ligand as given by NHEA and NPOL (i.e. NBEI, *n*BEI, and NSEI), respectively. We propose that suitable combinations of pairs of these variables (combining size and polarity-related variables in a Cartesian plane) can effectively aid in mapping the chemical space of available chemical matter and relate it to the targets towards which they have a measurable affinity. The representation is akin to an atlas, representing how the different physicochemical properties of the ligands (angular coordinate) relate to their potency towards biological targets (radial coordinate).

It is anticipated that wider use of this conceptual, numerical and graphical representation of the extensive public and proprietary target–ligand databases of inhibitors and commercial drugs at different stages, along the approval pathway could point the way towards a more effective and efficient drug discovery process. Possibly, this framework could be instrumental in identifying regions of CBS with a higher probability of yielding clinical drugs for specific targets. This 'drug likelihood' estimate (the drug efficiency index) could be incorporated into the proposed framework as an elevation coordinate (or additional coordinate), thus providing an effective mapping of the drug discovery landscape.

Acknowledgements

Access to the data in WOMBAT granted by T. Oprea is fully appreciated. The authors

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004

appreciate the constructive discussions, criticism and support given to the project by H. Lu and his group in the Department of Bioinformatics and Bioengineering at UIC. Comments and discussions with Debbie Mulhearn and computer support by B. Santasiero are also greatly appreciated. Kent Stewart, Yvonne Martin and James Metz from Abbott Laboratories read previous versions of the manuscript and provided valuable criticism. Prof. Federico Gago from the University of Alcala de Henares, Madrid, Spain, provided insightful comments. The contribution of P. Bento at EMBL-EBI in preparing the figures is greatly appreciated, as is the hospitality of the entire ChEMBL group at the Wellcome Trust campus, Hinxton, UK. The hospitality of the Parc Cientific Barcelona supported by a grant from the AGAUR Foundation for the completion of this work is greatly appreciated. The medicinal chemistry and computational groups at Merck-Serono (Wolfgang Sauer, Serge Christman-Franck, Mireille Krier and colleagues) were exposed first to these ideas, and their positive response gave the impetus and motivation to continue. We acknowledge partial support for this work from the Center for Pharmaceutical Biotechnology (O.P) and the EMBL (A.P.B). The ChEMBL database is independently supported by the Wellcome Trust.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.drudis.2010.08.004.

References

1 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108
2 Hajduk, P.J. and Greer, J. (2007) A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discov.* 6, 211–219
3 Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* 44, 235–249
4 Congreve, M. *et al.* (2003) A 'rule of three' for fragment-based lead discovery? *Drug Discov. Today* 8, 876–877
5 Campbell, S.F. (2000) Science, art and drug discovery: a personal perspective. *Clin. Sci.* 99, 255–260
6 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242
7 Wang, R. *et al.* (2005) The PDBbind database: methodologies and updates. *J. Med. Chem.* 48, 4111–4119
8 Wang, R. *et al.* (2004) The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J. Med. Chem.* 47, 2977–2980
9 Liu, T. *et al.* (2007) BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities. *Nucleic Acids Res.* 35, D198–D201 Database issue
10 Chen, X. *et al.* (2001) The binding database: overview and user's guide. *Biopolymers* 61, 127–141
11 Wishart, D.S. *et al.* (2006) DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. *Nucleic Acids Res.* 34, D668–D672
12 Smith, R.D. *et al.* (2006) Exploring protein–ligand recognition with Binding MOAD. *J. Mol. Graph. Model.* 24, 414–425
13 Hu, L. *et al.* (2005) Binding MOAD (Mother Of All Databases). *Proteins* 60, 333–340
14 Oprea, T.I. *et al.* (2007) Lead-like, drug-like or "Pub-like": how different are they? *J. Comput. Aided Mol. Des.* 21, 113–119
15 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
16 Lipinski, C. and Hopkins, A. (2004) Navigating chemical space for biology and medicine. *Nature* 432, 855–861
17 Oprea, T.I. and Gottfries, J. (2001) Chemography: the art of navigating in chemical space. *J. Comb. Chem.* 3, 157–166
18 Abad-Zapatero, C. (2007) A sorcerer's apprentice and the Rule of Five: from rule of thumb to commandments and beyond. *Drug Discov. Today* 12, 995–997
19 Hopkins, A.L. *et al.* (2004) Ligand efficiency: a useful metric for lead selection. *Drug Discov. Today* 9, 430–431
20 Abad-Zapatero, C. and Metz, J.M. (2005) Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* 10, 464–469
21 Kuntz, I.D. *et al.* (1999) The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9997–10002
22 Reynolds, C.H. *et al.* (2007) The role of molecular size in ligand efficiency. *Bioorg. Med. Chem. Lett.* 17, 4258–4261
23 Perola, E. (2010) An analysis of the binding efficiencies of drugs and their leads in successful drug discovery programs. *J. Med. Chem.* 53, 2986–2997
24 Verdonk, M.L. and Rees, D.C. (2008) Group efficiency: a guideline for hits-to-leads chemistry. *ChemMedChem* 3, 1179–1180
25 Roche, D. *et al.* (2009) Discovery and structure–activity relationships of pentanedioic acid diamides as potent inhibitors of 11beta-hydroxysteroid dehydrogenase type I. *Bioorg. Med. Chem. Lett.* 19, 2674–2678
26 Lepifre, F. *et al.* (2009) Discovery and structure-guided drug design of inhibitors of 11beta-hydroxysteroid-dehydrogenase type I based on a spiro-carboxamide scaffold. *Bioorg. Med. Chem. Lett.* 19, 3682–3685
27 Congreve, M. *et al.* (2008) Recent developments in fragment-based drug discovery. *J. Med. Chem.* 51, 3661–3680
28 Andersen, O.A. *et al.* (2008) Structure-based dissection of the natural product cyclopentapeptide chitinase inhibitor argifin. *Chem. Biol.* 15, 295–301
29 Abad-Zapatero, C. (2007) Ligand efficiency indices for effective drug discovery. *Exp. Opin. Drug Discov.* 2, 469–488
30 Stewart, K.D. *et al.* (2006) Drug Guru: a computer software program for drug design using medicinal chemistry rules. *Bioorg. Med. Chem.* 14, 7011–7022
31 Wenlock, M.C. *et al.* (2003) A comparison of physicochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* 46, 1250–1256
32 Vieth, M. *et al.* (2004) Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* 47, 224–232
33 Scapin, G. *et al.* (2003) The structural basis for the selectivity of benzotriazole inhibitors of PTP1B. *Biochemistry* 42, 11451–11459
34 Iversen, L.F. *et al.* (2000) Structure-based design of a low molecular weight, nonphosphorus, nonpeptide, and highly selective inhibitor of protein-tyrosine phosphatase 1B. *J. Biol. Chem.* 275, 10300–10307
35 Andersen, H.S. *et al.* (2000) 2-(oxalylamino)-benzoic acid is a general, competitive inhibitor of protein-tyrosine phosphatases. *J. Biol. Chem.* 275, 7101–7108
36 Szczepankiewicz, B.G. *et al.* (2003) Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* 125, 4087–4096
37 Xin, Z. *et al.* (2003) Identification of a monoacid-based, cell permeable, selective inhibitor of protein tyrosine phosphatase 1B. *Bioorg. Med. Chem. Lett.* 13, 3947–3950
38 Ala, P.J. *et al.* (2006) Structural insights into the design of nonpeptidic isothiazolidinone-containing inhibitors of protein-tyrosine phosphatase 1B. *J. Biol. Chem.* 281, 38013–38021
39 Ala, P.J. *et al.* (2006) Structural basis for inhibition of protein-tyrosine phosphatase 1B by isothiazolidinone heterocyclic phosphonate mimetics. *J. Biol. Chem.* 281, 32784–32795
40 Klopfenstein, S.R. *et al.* (2006) 1,2,3,4-Tetrahydroisoquinolinyl sulfamic acids as phosphatase PTP1B inhibitors. *Bioorg. Med. Chem. Lett.* 16, 1574–1578

***Cele Abad-Zapatero*[1,*]**
***Ognjen Perišić*[1,2]**
***John Wass*[3]**
***A. Patrícia Bento*[4]**
***John Overington*[4]**
***Bissan Al-Lazikani*[4,5]**
***Michael E. Johnson*[1]**

[1]*Center for Pharmaceutical Biotechnology, MBRB Building, University of Illinois at Chicago (MC870), 900 S. Ashland Street, Room 3020, Chicago, IL 60607, USA*
[2]*Computational Science. Department of Chemistry and Applied Biosciences - ETH Zurich USI Campus, via Giuseppe Buffi 13. CH - 6900 Lugano*
[3]*Quantum Cat consultants, 602 Forest Hill, Rd. Lake Forest, IL 60045, USA*
[4]*European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK*
[5]*Current address: The Institute of Cancer Research, Haddow Laboratories, Belmont, Sutton, Surrey, UK.*

**Corresponding author*
caz@uic.edu, xtalp1@aol.com (C. Abad-Zapatero)

Please cite this article in press as: Abad-Zapatero, C. et al., Ligand efficiency indices for an effective mapping of chemico-biological space: the concept of an atlas-like representation, Drug Discov Today (2010), doi:10.1016/j.drudis.2010.08.004