J Comput Aided Mol Des (2012) 26:995–1003
DOI 10.1007/s10822-012-9587-5

AtlasCBS: a web server to map and explore chemico-biological space

Álvaro Cortés-Cabrera · Antonio Morreale · Federico Gago · Celerino Abad-Zapatero

Received: 26 March 2012 / Accepted: 30 June 2012 / Published online: 14 July 2012

Abstract New approaches are needed that can help decrease the unsustainable failure in small-molecule drug discovery. Ligand Efficiency Indices (LEI) are making a great impact on early-stage compound selection and prioritization. Given a target-ligand database with chemical structures and associated biological affinities/activities for a target, the AtlasCBS server generates two-dimensional, dynamical representations of its contents in terms of LEI. These variables allow an effective decoupling of the chemical (angular) and biological (radial) components. BindingDB, PDBBind and ChEMBL databases are currently implemented. Proprietary datasets can also be uploaded and compared. The utility of this atlas-like representation in the future of drug design is highlighted with some examples. The web server can be accessed at http://ub.cbm.uam.es/atlascbs and https://www.ebi.ac.uk/chembl/atlascbs.

Á. Cortés-Cabrera · F. Gago · C. Abad-Zapatero
Departamento de Farmacología, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, Spain

Á. Cortés-Cabrera · F. Gago · C. Abad-Zapatero
Unidad Asociada Instituto de Química Médica (CSIC), Madrid, Spain

Á. Cortés-Cabrera · A. Morreale
Unidad de Bioinformática, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Campus UAM, c/Nicolás Cabrera 1, 28049 Madrid, Spain

Present Address:
C. Abad-Zapatero (✉)
Center for Pharmaceutical Biotechnology, University of Illinois at Chicago, MBRB building (MC870), Chicago, IL 60607, USA
e-mail: caz@uic.edu

Keywords Ligand efficiency indices · Chemico-biological space · Structure–activity databases · AtlasCBS server · Efficiency-based drug design · Efficiency planes

Introduction

The ever growing advances in the fields of structural biology, high-throughput screening and structure-based drug design have resulted in an exponential increase of the information related to targets, ligands, and their complexes that is stored in several databases (i.e., Structure–Activity-Relationship or SAR databases: BindingDB [1], ChEMBL [2], PDBBind [3], among others). The vastness of chemical space as it relates to medicinal applications has been recognized [4] and certain tools to aid in navigating it have been introduced [5, 6]. Nonetheless, an effective method to combine biological targets with the subset of ligands with which they interact and map them in chemico-biological-space (CBS) is lacking.

In the last few years the concept of ligand efficiency has taken hold in medicinal chemistry following the pioneering work of Hopkins et al. [7] who suggested an efficiency-based reference for lead selection taking into account the size of the ligand. The concept was extended to include efficiency in two complementary variables: size (molecular weight, MW, of the ligand) and polarity (Polar Surface Area, PSA) [8]. In this formulation, the size variable was defined as BEI = pK_i/(MW/1,000) and the polarity variable as SEI = pK_i/(PSA/100) where K_i is the binding affinity of the inhibitor (see Table 1). BEI is the binding efficiency index relating potency to MW on a per kDa scale and SEI is the surface efficiency index monitoring the potency gains as related to the increase in PSA referred to 100 $Å^2$. It was then proposed that ligand efficiency indices

Table 1 Definitions of ligand efficiency indices used in AtlasCBS

Acronym	Definition
BEI	pK_i/MW, pK_d/MW, pIC_{50}/MW[a]
SEI	pK_i/PSA, pK_d/PSA, pIC_{50}/PSA[b]
NSEI	$pK_i/NPOL$[c]
NBEI	pK_i/NHA[d]
nBEI	$-\log_{10}$ [K_i/NHA]
mBEI	$-\log_{10}$ [K_i/MW]

Examples of possible efficiency planes represented and available in the AtlasCBS server and a brief description of their characteristics and appearance

1. SEI, BEI (x, y). Slope of the lines 10*(PSA/MW). No intersect

2. NSEI, NBEI (x, y). Slope of the lines NPOL/NHA: a rational number. No intersect

3. NSEI, nBEI (x, y). Slope of the lines NPOL, intersect $\log_{10}$(NHA)

4. NSEI, mBEI (x, y). Slope of the lines NPOL, intersect $\log_{10}$(MW)

Any combination is possible but pairs of variables related to affinity/polarity (SEI-like, x) and affinity/size (BEI-like, y) are strongly recommended. With this choice of variables the polarity or physico-chemical characteristics (as given by PSA/MW or NPOL/NHA) of the chemical compounds increases counterclockwise as indicated above, mimicking a graphical representation of the variables considered in Lipinski's Rule of Five (Ro5) [13, 17] by changes in the angular coordinate (slope) of the lines

[a] MW: molecular weight (in kDa). $pK_i = -\log_{10} K_i$

[b] PSA: polar surface area scaled to 100 $Å^2$

[c] NPOL: number of polar atoms (N and O)

[d] NHA: number of heavy atoms (non-hydrogen)

(LEI) could be used to guide drug discovery. These initial ideas have been extended to include other indices such as those related to lipophilicity of the ligands: LLE, lipophilicity ligand efficiency; LLE_{Astex}, lipophilicity ligand efficiency for fragments; and LELP, lipophilicity in optimization and drug-likeness. In addition, the ligand efficiency decreases with the size of the ligand and thus size-corrected (or size-independent) ligand efficiencies (SILE) have also been introduced. A comprehensive review of the various LEI mentioned has been published recently [9], where the original concepts and the newly introduced variables have been fully discussed and documented in the context of the current practice of medicinal chemistry .

The concept of an atlas-like representation of CBS based on the use of two complementary LEI was introduced recently [10]. The proposed formulation allowed a representation in Cartesian planes (efficiency planes) that combined the affinity of the ligand towards the target and two important physico-chemical properties of the ligand: size and polarity. Briefly, a combination of efficiency indices defined in size (BEI, y) and polarity (SEI, x) can be represented on two-dimensional (2D) efficiency planes (see Table 1 for a summary of proposed variables). The rationale was that, as a first approximation, the drug discovery process attempts to optimize the combination of three variables: potency, size and polarity and therefore a graphical representation of the combination of these parameters could aid in the optimization process. Given a choice of axes and definition of variables, the appearance of the plots is characterized by a counterclockwise increase of the polarity of the compounds in the angular coordinate (chemistry) and a radial distance from the origin related to the affinity towards the target (biological) (see Table 1). Applications of this representation to several areas of drug discovery have also been described [10].

However, thus far, no friendly tool is available to connect 2D or 3D ligand structures plus their chemical properties (chemical space) with biological affinity/activity data pertaining to one or more target proteins (biological space). Moreover, a friendly tool is lacking that would allow the medicinal chemistry community to explore the value of the graphical representation of CBS and the efficiency (in size and polarity) of their chemical entities towards a certain target. The AtlasCBS server introduced and described here is such an application.

Materials and methods

LEI and molecular properties calculation

Several LEI are calculated, as described elsewhere, using the formulas contained in Table 1 and in Ref. [10]. Molecular properties such as atomic masses, number of polar (NPOL), non-hydrogen atoms (NHA), and PSA are polar surface area are calculated using the Chemistry Development Toolkit (CDK) [11, 12].

Web design and usage

The main web page for the server contains the five tabs shown in Fig. 1a, b: (a) **Main**: basic information about the server and its purpose, main references, contact information, and access to the main topics covered in the **Help** tab; (b) **Map viewer**: tools for uploading the data from existing databases and for visualizing and analyzing their content; (c) **Login**: required only if the user wants to have private database access; (d) **Help**: information on how to use the AtlasCBS server; and (e) **About**, details concerning the institutions and people involved in the project. The **Login** tab changes to **Manage data** upon a successful login allowing access to proprietary user data. Underneath the server there are three modules that provide all the functionalities for the tabs:

- **(A) Map viewer**. The graphical engine of the server represents data from different sources, allows visualization

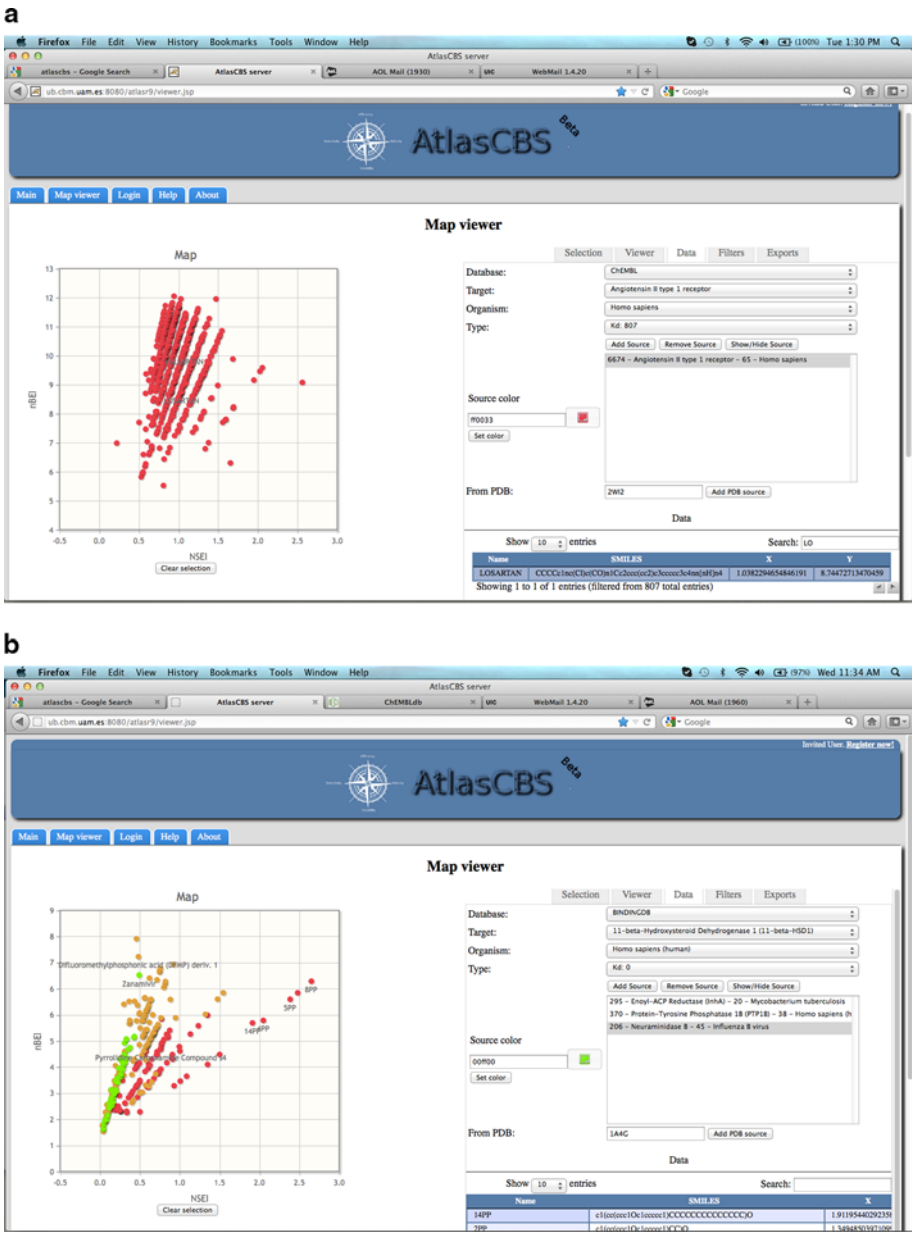

Fig. 1 Example screens presented by the AtlasCBS server. **a** The *upper left* tabs correspond to the main pages of the server: **Main**, **Map Viewer**, **Login** (for private access), **Help**, and **About** (see text). The left graphical panel within the page represents a typical efficiency plane (NSEI vs. nBEI) for the 807 entries found in ChEMBL for Angiotensin II receptor type I with K_d affinity values (see Type window). Each point in the plane represents a target-ligand pair. The angular coordinate (NPOL in this case) corresponds to the number of polar (N,O) atoms of the ligand increasing counterclockwise (NPOL = 3–12). The radial coordinate corresponds to the affinity of the ligands towards the target. The *top right* panel shows the different options for the management of the session within the **Map viewer** and choice of database, target and organism. Most importantly, the choice of LEI as Cartesian axes (x, y) (within **Viewer**) determines the appearance of the efficiency planes. The *lower right* panel shows the compounds selected in an alphabetical list. The values of the LEI variables are also shown in this window along with the SMILES strings of the corresponding compounds. The SMILES string of one of the selected compounds (LOSARTAN) is shown at the bottom, resulting from inputting 'LO' in the search box. It is also annotated on the left map in black fonts. Using the 'Duplicate window' tab, multiple windows can be displayed simultaneously to compare pages in the 'AtlasCBS' with different variables or scales, as in a real life atlas. **b** As in **a** but in this case the map in the NSEI-nBEI plane corresponds to the response of the server to three succesive PDB access codes: 2H7L, enoyl-ACP-reductase (InhA) from *Mycobacterium tuberculosis* (*purple*), 2CNE, human protein tyrosine phosphatase 1B (*amber*), and 1A4G, neuraminidase B (*green*). The colors were changed for contrast using the 'Source color' option. In addition to the compound markers corresponding to the three PDB entries, other compounds have been *highlighted* (black lettering 14PP, 7PP etc.). Of note, the marketed drug zanamivir occupies the top (most efficient) position of the neuraminidase B inhibitors (*green set*) (see Fig. 3 for a close-up)

of chemical structures, and provides efficient filtering and searching tools to compare and classify molecules and their efficiencies. The application opens in the **Data** tab within the **Map viewer** panel. First, the users should select any data source (target and organism) available within the AtlasCBS server, extracted from a previous release of BindingDB (www.binding.org), PDBBind (www.pdbbind.org) and ChEMBL (https://www.ebi.ac.uk/chembl), scrolling the corresponding server tabs, or load an external data set (see below). Any external set should include the compound name, structural description of the molecules (SMILES strings) and their affinity/activity (K_d, K_i, or IC_{50}) values.

1. **Data selection.** The user selects from the available data (K_i, IC_{50} or K_d) in the databases and uploads (**Add Source** tab) the data to the viewer. By default, the first map shown is nBEI versus NSEI, where the appearance is a set of lines of slope given by the number of polar (N + O) atoms, increasing counterclockwise. To generate other maps, the user selects x and y variables in the **Viewer** tab from the sets: SEI, BEI; NSEI, NBEI; NSEI with nBEI or mBEI, respectively. A brief definition of the different variables in tabular form is available by opening an adjacent pull-down tab, and other details are provided in a link. Any combination is possible but complementary pairs (LEI per size and polarity: y and x, respectively) are recommended. Given a map, molecules can then be selected by clicking on them on the map or from the list. The 2D structure of the selected molecule can be seen within the **Selection** tab together with its basic physico-chemical properties. A list of compound names, SMILES strings and LEI values (x, y) for the displayed compounds is shown on the lower right panel (Fig. 1). Molecules can be selected by a simple character search on the same panel or filtered (within the **Filter** tab) using either a "range of values" or the "Slope" option to choose those that share the same number of polar atoms (in the NSEI-nBEI plane). Selected compounds can be compared or used for similarity searches employing molecular fingerprints and Tanimoto coefficients. Other features include: (a) mixing and visualization of different data sources at the same time; (b) changing colors for the different data sources, with 'on' and 'off' capabilities; and (c) dynamic scaling of the axis to zoom in/out on particular areas of interest (Fig. 1). These features are intended to give the sense and the feel of a geographical atlas. Finally, it is possible to save and restore any working session as well as to export 'working datasets' and plots of specific efficiency planes of interest. The appearance of some of these options could be dependent on the browser's capabilities (Table 2; Figs. 1, 2).
2. **Relation to other databases.** Within **Data,** it is possible to input directly a PDB access code and the server will make available to the user all the affinity data present in BindingDB for that target. The activity or affinity parameter (IC_{50}, K_i, or K_d) with the larger number of entries is selected by default. As before, once a compound has been selected and appears in the selected list, the **Selection** tab, within the **Map Viewer** module, shows the compound's 2D structure together with its name, basic physico-chemical properties (MW, number of polar atoms, PSA, number of heavy [i.e. non-hydrogen] atoms) and a direct link to BindingDB (Fig. 2).

- **(B) Manage data (Private database manager).** This feature allows users to upload and process their own datasets in a secure way provided they register and accept the terms of usage of the site (access is granted using a valid e-mail address and a user-chosen password).

  1. Datasets containing affinity data are read in as semicolon-separated values ("CSV"), which are readily available from common spreadsheets such as those produced by Microsoft Excel or OpenOffice Calc. Compounds can also be added manually as long as the required data are correctly given in the specified format. Users can modify each field interactively and they can also use filtering and searching tools.
  2. Chemical library contents can be uploaded into the private data stream and the server will generate randomized affinity (K_i) values to explore/simulate a screening experiment (see below).

Table 2 Viewers characteristics

Characteristic	Java applet viewer	Javascript/HTML viewer
SMART filtering	✔	–
Similarity search	✔	–
Simple filters	✔	✔
Polarity highlighting	✔	–
Labels	✔	✔
Mixed sources	✔	✔
Save session	–	✔
Dynamic scale change	✔	✔

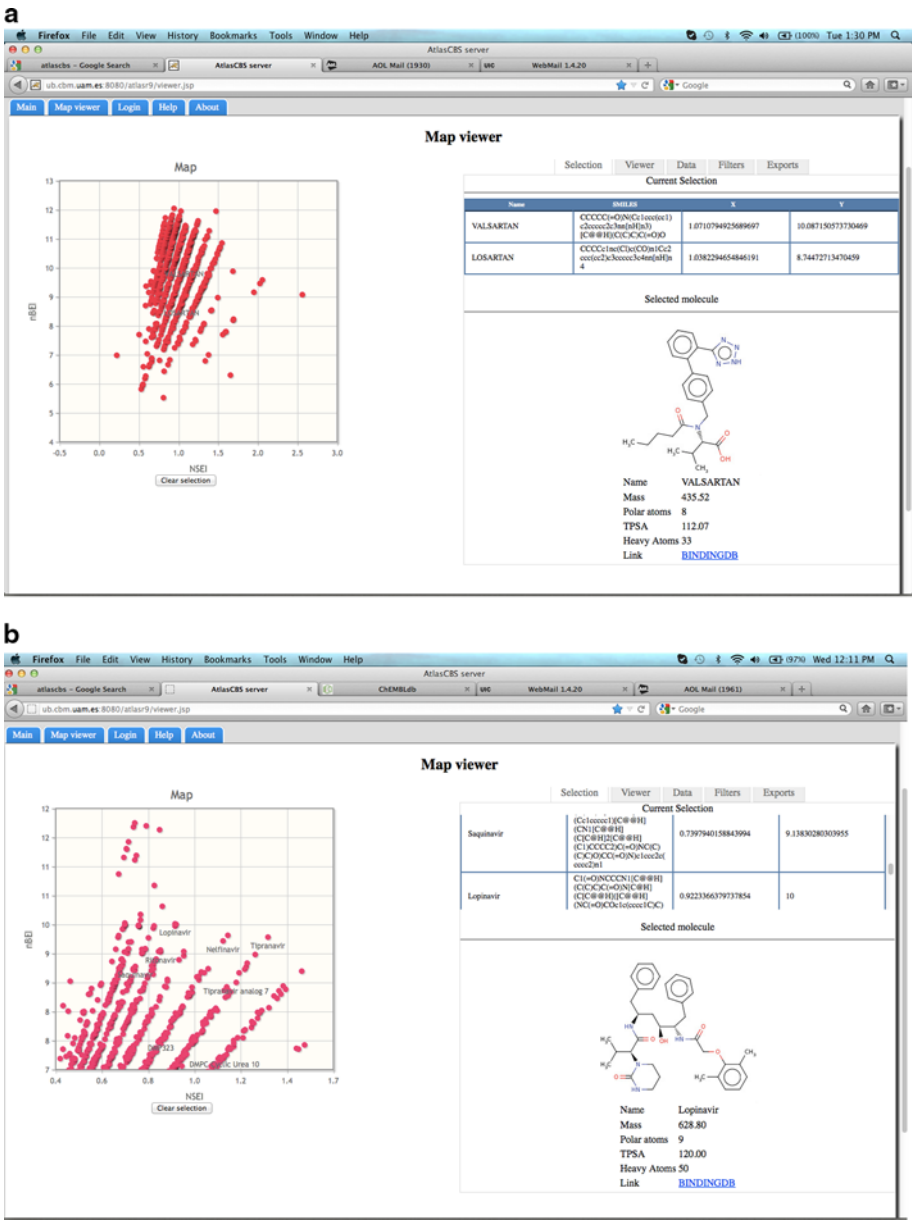

Fig. 2 Mapping efficiency and structure in chemico-biological space. **a** Image from the **Map Viewer** panel illustrating the position of a certain selected compound (VALSARTAN) targeting angiotensin II type 1 receptor (as in Fig. 1a) and the corresponding 2D structure within the **Selection** tab. The basic physico-chemical parameters of the compound are displayed as well as a direct link to BindingDB that can be used to look up some additional information. The relative efficiency of other compounds for the same target such as LOSARTAN (Fig. 1a) can be compared. **b** Close-up image from the **Map Viewer** panel showing a subset of the compounds extracted from BindingDB as a response to the PDB access code 1OHR (HIV-1 protease in complex with nelfinavir). The 2D structure of the selected compound is shown. The progressive migration towards higher efficiencies of the compounds targeting HIV-1 protease can be seen. The structure of Lopinavir, a later HIV-1 protease inhibitor to make it into the market, is shown for reference. The angular component (*slope of the lines*) relates to the number of polar atoms (N + O). The radial coordinate is given by the measured affinity towards the corresponding target. Different measurements for different targets (i.e. wild-type vs. mutant(s)) will all map along the same line. The thickness of the lines depends on the number of compounds with the same NPOL value but different number of non-hydrogen atoms ($\log_{10}$ (NHA)). See Table 1. The selection of this area (NSEI: 0.4–1.7; nBEI: 7–12) was made by highlighting a certain region of the full NSEI-nBEI plane with the mouse following representation on the plane of all the compounds with affinity for the HIV-1 protease target. The physico-chemical properties of the ligands are shown together with a direct link to BindingDB as above. Note how lopinavir, tipranavir and nelfinavir occupy the highest efficiency positions for the corresponding NPOL lines in relation to other analogs (e.g. tipranavir analog 7, *third line* from the *right*)

- **(C) Help.** It contains and explains the elements and functionalities of the server as well as references to background papers. There are also references to previous papers in the main portal of the server.

Implementation details

The server is organized in three layers: clients, application server and database. Each layer can communicate with the nearest neighbor but not beyond. The three layers have been implemented with the following elements for the different components, respectively: (1) Java, JavaScript, and HTML clients; (2) the Apache Tomcat servlet container; and (3) the MySQL database engine. The front-end is based on HTML, JavaScript and Java Server Pages (JSP) or a Java applet; Java servlets handle the data traffic between the interface and the database, using a Model-View-Controller (MVC) paradigm. The information contained in a current release of BindingDB (19/05/2012), PDBBind (v2011) and ChEMBLdb (v13) was imported into the MySQL server's database by standalone Java programs that also compute the molecular properties and the efficiency indices for the molecules. Data for three affinity parameters (K_i, IC_{50}, K_d) are extracted and retained within the server connected to the appropriate target and organism. In the case of BindingDB, the SDF (Structure Data File) was converted to SMILES format using RDKit (http://www.rdkit.org/) and then inserted into the database with the information provided for targets and organism in the property fields of the same SDF file, along with the calculated molecular properties by CDK [11, 12]. For PDBBind the data were transformed directly into SMILES strings from the PDB files and the affinity information in the annotated set was used. Finally, for ChEMBL, a subset of the original release for MySQL was imported, excluding extreme affinity/activity values and considering only data in nanomolar units for K_i, IC_{50} and K_d. Names or reference IDs for molecule in those databases were recorded in order to provide links from the platform to the original data sources.

Users-uploaded external data are processed on-the-fly, with a special servlet using the CDK and the AJAX technology [12]. The accepted format of the user's external database and examples are described in the help pages. For each entry, the CSV-formatted file should contain molecule name, SMILES code, type of affinity/activity variable (K_i, IC_{50}, K_d), and the affinity/activity value (in nanomolar concentration units).

A separate tab is available to upload the contents of proprietary chemical libraries (without any affinity data) within the **Manage data** tab. The files containing the chemical composition of the libraries should be in SMILES format separated by a tab character from the compound name. Only two items are required: compound number (e.g. AS0045) and the corresponding SMILES string. The server assigns internally random affinity (K_i) values to the compounds in the micro- to nanomolar range. The physicochemical properties of the compounds within the library are represented by the range of the angular coordinates in the efficiency planes. The regions of the efficiency plane accessible to different affinities are simulated by the random values of the affinity constants (radial direction) [13].

Visualization of the data is based on a Java applet that enables graphical representation and allows the display of multiple pages simultaneously, zooming in and out of user's predefined areas, and selection of compounds using SMARTS strings, similarity, or automatic detection. In the NSEI-nBEI (x, y) plane, compound series are easily followed by the slope of the lines that correspond to the number of polar atoms (N + O), which increases counterclockwise. We have also developed a javascript-based web application that avoids the use of Java but implements fewer features, although the result is browser-dependent. For a list of the features supported by each visualization module see Table 2.

Molecular fingerprints and Tanimoto coefficients

Molecular comparisons are based on the calculation of the CDK [11, 12] fingerprint, which encodes the topology of the compounds as bit strings. Bit strings are compared using the Tanimoto coefficient (Tc, Eq. 1), which evaluates, from 0 (not at all) to 1 (identity), the similarity between the compounds:

$$Tc = Nab/(Na + Nb - Nab) \quad (1)$$

where Nab are the bits in common for two compounds (a, b) and Na and Nb the number of bits activated in compounds a and b, respectively.

Results and discussion

LEI represent a relatively simple concept that naturally connects the chemistry and the biology of receptor-binding ligands via one or more affinity/activity variables (K_i, IC_{50}, or equivalent). Other expanded definitions of ligand efficiency proposed recently to expand the initial ideas [9] could further impact the future of drug discovery. In this work, we have developed an application based on a unified formulation of LEI (see Table 1 for definitions) that are calculated by weighting the affinity values with molecular properties such as the number of heavy (non-hydrogen, NHA) or polar atoms (NPOL = Number of N + O), the polar surface area (PSA), or the molecular weight (MW).

The combined use of two complementary LEI, namely (1) affinity/polarity (K_i combined with NPOL, PSA), x-axis, and (2) affinity/size (K_i combined with NHA, MW), y-axis, allows a very intuitive depiction of the database contents as a series of Cartesian diagrams that constitute an atlas-like representation of CBS. The characteristics and appearance of these plots ("efficiency planes") depend on the choice of variables (see Table 1) and can be examined using the AtlasCBS server.

A very interesting characteristic of these LEI variables is that pairs of complementary indices can be displayed on an efficiency plane that unveils some of the intricacies of CBS in an appealing graphical manner. Although any combination of affinity/polarity (x-coordinate) and affinity/size (y-coordinate) could be useful [10], of special interest are the NSEI-nBEI, NSEI-mBEI (x, y) efficiency planes (Table 1; Figs. 1, 2, 3). In this type of plot, the slope of the line occupied by each target-ligand pair (its angular coordinate) depends only on the chemical composition of the ligand (in this case given by the number of polar atoms). The unique position along the line (the radial coordinate) for each target-ligand pair will depend on the biological affinity/activity of the ligand for/on the given target [10]. The drug discovery efforts can then be represented on the suggested efficiency planes as 'trajectories' in CBS [10] and optimized trajectories could be devised or proposed in the future based on numerical or statistical criteria [14].

It has been shown in initial retrospective studies that the optimized ligand(s) within a series of analogues typically map on the upper, right-hand, quadrant of the efficiency planes (e.g. NSEI and nBEI, or similar), where both variables are maximized [10, 13, 14]. Examples are shown in Figs. 1b, 2b and 3. Therefore, in the near future, it is theoretically possible to envisage an automatic procedure that will optimize ligand efficiencies by replacing some molecular fragments and evaluating the LEI of the new candidates in an iterative fashion, until the best possible ligand(s) is(are) found [15]. It is precisely here that the future power of this methodology as a graphical and numerical guide in the search for better drug candidates should be apparent. In the future, other definitions [9], and possibly more variables, could be incorporated into this graphical framework to strengthen the prospective value of these concepts.

Other combinations of the variables defined in Table 1 can be selected and used in pairs to build up an "electronic atlas" composed of different "Cartesian maps" (i.e., pages or efficiency planes). These maps will depict complete views or selected regions of CBS at different scales, constituting what we refer to as an atlas-like representation of CBS (Fig. 2b). This AtlasCBS server is visual and dynamic

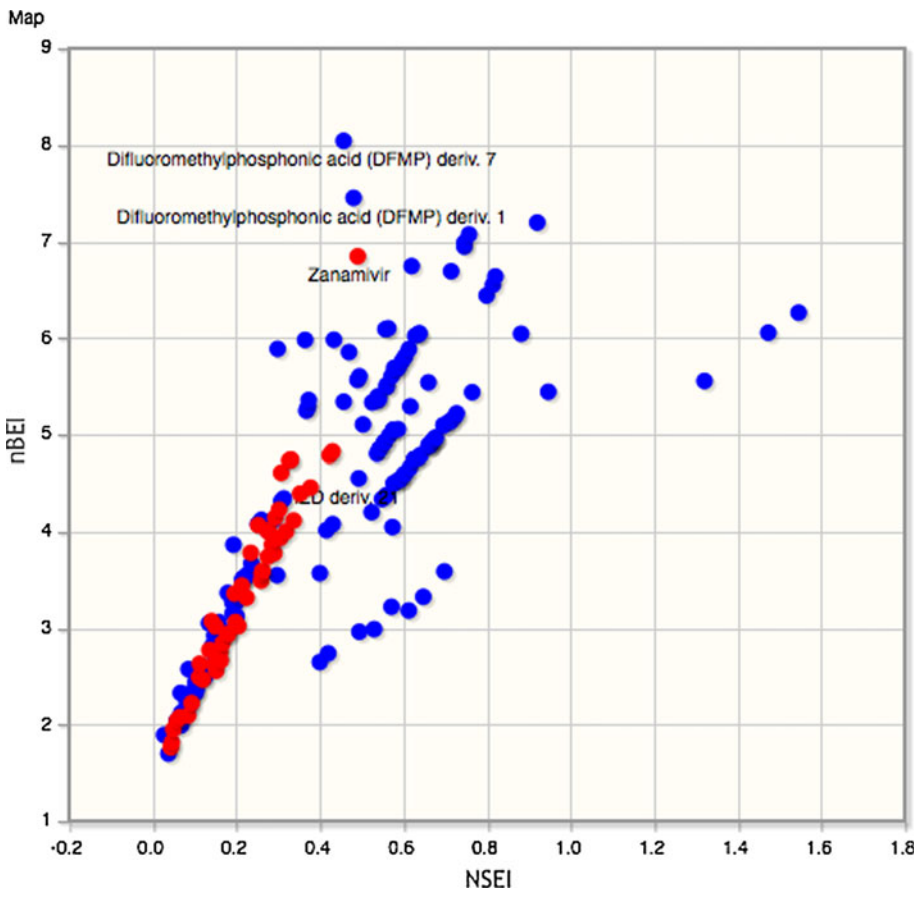

Fig. 3 Efficiency plane representation of the compounds with affinity for neuraminidase and PTP1B. The information was extracted from the BindingDB database upon entering the PDB access codes 1A4G (*red*) and 2CNE (*blue*) in the AtlasCBS server application. The plot illustrates how the marketed compound (zanamivir) optimizes the efficiency of the ligand in size (y-axis) and polarity (x-axis). This image was exported using the '**Exports**' option from the AtlasCBS server after removing the Enoyl ACP-reductase data shown in Fig. 1b

by nature and we would expect it to be extremely useful to help navigation through the 'vastness of chemical space' [4]. To the best of our knowledge, this is the first time a tool is presented that graphically displays, and naturally maps and classifies in a user-friendly way, the information stored in these ligand-receptor databases in terms of LEI. Apart from representing the chemico-biological content of public databases such as BindingDB, PDBBind and ChEMBL, our server also allows interested users to upload, map and compare their own proprietary datasets. Thus, the CBS of known targets can be easily explored and data from different sources can be visualized and compared. The URL for the server is http://ub.cbm.uam.es/atlascbs but a mirror is also available at the European Bioinformatics Institute (EBI) Hinxton campus server https://www.ebi.ac.uk/chembl/atlascbs. Both sites require registration only to grant secure and confidential access.

Some applications of the AtlasCBS concept have been reviewed lately by Abad-Zapatero and Blasi [10, 13] in several domains of the drug discovery process. Examples are the analysis and comparison of the contents of different databases (mapping of drugs vs. non-drugs), polypharmacology, fragment-based ligand design strategies, drug discovery trajectories and others. The AtlasCBS server presented here allows the exploration of these important areas of drug discovery dynamically on-line for the first time. We wish to encourage the drug discovery community to use this tool so that it can be further improved.

Besides the contents of the three SAR databases currently included in the server (BindingDB, PDBBind and ChEMBL), the internal structure of the application allows a direct link with the Protein Data Bank (PDB) and BindingDB in two effective ways. From the **Data** tab, the user can input a PDB accession code containing a target-ligand complex (for example 1A4G, for the target neuraminidase in complex with zanamivir). The application will use the information in the PDB entry to extract all the corresponding affinity data for that target from BindingDB and will represent the available ligands on the nBEI versus NSEI efficiency plane, with the entered compound highlighted for reference (see Figs. 1b, 2b, 3). A direct link with the information for any compound present in BindingDB is possible from the **Selection** tab and this facilitates the access to further information about the target-ligand pairs under study.

The examples presented above of the efficiency data for compounds targeted towards HIV-1 protease (Fig. 2b, extracted using PDB code 1OHR) and influenza virus neuraminidase (PDB code 1A4G, Fig. 3) illustrate a very important use of the efficiency planes presented by the AtlasCBS tool that could have an impact on future drug discovery efforts. Namely, that compounds with maximal efficiencies in size and polarity are often the best suited preclinical candidates for further development and often correspond to the successful marketed drug, as supported by other retrospective studies [13–15]. This notion can be extensively explored using the server as the available data can be represented in a variety of efficiency planes. To make it more effective and complete, the inclusion of additional LEI into the AtlasCBS framework is being considered and will be implemented in the near future. This will include size-related indices such as the original and commonly used ligand efficiency definition proposed by Hopkins et al. [7], and also size-independent (e.g. SILE) and polarity-related indices (e.g. LLE and others [9]).

Undoubtedly, the most interesting applications will be those for which LEI are used prospectively to guide the drug discovery process. Our suggestion would be to incorporate routinely the LEI framework into the drug discovery pipeline. In a recent study, Blasi et al. [16] devised a workflow to obtain better drug candidates targeting the transthyretin carrier protein (TTR) by combining LEI, pharmacophoric search and ligand docking. Briefly, a retrospective NSEI-nBEI map was first built with some known binders so as to select the most appropriate candidate for further improvement. Second, the core structure of the selected compound was used as a pharmacophore to search into a database of commercially available compounds. Those molecules fulfilling the pharmacophoric requirements were submitted to docking and the 80 top-scoring hits were selected. Third, these scores were transformed into estimated K_i values for calculating their 'theoretical' LEI. Finally, a prospective map was built and the 12 'most efficient' compounds (those having the highest values of NSEI-nBEI for the different NPOL lines) were selected for experimental tests of activity and pharmacokinetic behavior. The results with the compounds proposed for NPOL = 5 (four compounds, the most polar and easier to synthesize) so far confirm the prediction of being the most efficient in the experimental assay (Blasi and Quintana, personal communication). Using as a guide the suggestion that compounds with maximal efficiencies are likely to be good candidates for further development, we propose that the above strategy and the use of the AtlasCBS server would be advantageous to the drug-discovery community. This could set the basis for a more rigorous, numerically and efficiency-based drug discovery paradigm [15].

Conclusion

An effective web tool is presented that aims to facilitate the drug discovery process by providing an atlas-like representation of the CBS using LEI as descriptors. This web server allows the graphical visualization of database

contents as pages in a map-like environment, with different variables and scales. The CBS can be easily navigated to examine the efficiency of existing and prospective target-binding molecules differing in size and polarity. We propose that the atlas representation can be extremely useful as a guide in several areas of drug discovery, including mapping design efforts, exploring new design strategies and optimizing candidates in hit-to-lead campaigns.

Acknowledgments The assistance of the ChEMBL group in facilitating the data extraction for the AtlasCBS database is greatly appreciated. This work was supported by grants from Comunidad Autónoma de Madrid (S-BIO-0214-2006 and S2010-BMD-2457 to F.G. and A.M.), Fundación Severo Ochoa (AMAROUTO program to A.M.) and the Spanish Ministerio de Educación (FPU AP2009-0203 to A. C. and SAB2010-0037 to C. A-Z). The suggestions and comments from the reviewers of this work are greatly appreciated. The insightful comments, discussions and suggestions of Daniel Blasi of the Platform of Drug Discovery (Dr. J. Quintana, Director) within the Parc Científic Barcelona are also appreciated. We gratefully acknowledge the help of Mark Davies, John Overington and Peter Rose in our efforts to relate AtlasCBS to ChEMBL and PDB.

References

1. Liu T, Lin Y, Wen X, Jorissen RN, Gilson M (2007) Nucleic Acids Res 35(Database issue):D198
2. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2011) Nucleic Acids Res 40(Database issue):D1100
3. Wang R, Fang X, Lu Y, Yang CY, Wang S (2005) J Med Chem 48(12):4111
4. Lipinski C, Hopkins A (2004) Nature 432(7019):855
5. Oprea TI, Gottfries J (2001) J Comb Chem 3(2):157
6. Watson P, Verdonk M, Hartshorn MJ (2003) J Mol Graph Model 22(1):71
7. Hopkins AL, Groom CR, Alex A (2004) Drug Discov Today 9:430
8. Abad-Zapatero C, Metz JM (2005) Drug Discov Today 10(7):464
9. Hann MM, Keseru GM (2012) Nat Rev Drug Discov 11(5):355
10. Abad-Zapatero C, Perisic O, Wass J, Bento AP, Overington J, Al-Lazikani B, Johnson ME (2010) Drug Discov Today 15(19–20):804
11. Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E (2003) J Chem Inf Comput Sci 43(2):493
12. Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL (2006) Curr Pharm Des 12(17):2111
13. Abad-Zapatero C, Blasi D (2011) Mol Inform 30(2–3):122
14. Christmann-Franck S, Cravo D, Abad-Zapatero C (2011) Mol Inform 30(2–3):137
15. Abad-Zapatero C (2007) Expert Opin Drug Discov 2(4):469
16. Blasi D, Arsequel G, Valencia G, Nieto J, Planas A, Pinto M, Centeno NB, Abad-Zapatero C, Quintana J (2011) Mol Inf 30(2–3):161
17. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Adv Drug Deliv Rev 23:3