

## ChEMBL and AtlasCBS: A Special Synergy within StarDrop

By Cele Abad-Zapatero. University of Illinois at Chicago. Center for Biomolecular Sciences and Institute of Tuberculosis Research. E-mail: [caz@uic.edu](mailto:caz@uic.edu).

Website: <https://www.uic.edu/labs/caz/index.html>.

ChEMBL is well established and highly reliable database and does not need any introduction. AtlasCBS is currently a server within the European Bioinformatics Institute <https://www.ebi.ac.uk/chembl/atlasCBS>. The concept of the Atlas of Chemico-Biological Space (AtlasCBS) was published in 2010 and the web tool was published and implemented in 2012<sup>1</sup>. It allows the representation, mapping and analysis of the content of Structure-Based Data Bases (SAR-Data Bases) in 'efficiency planes' defined by pairs of variables related to Ligand Efficiency Indices. A script to incorporate these ideas into the workflows of drug discovery within **StarDrop** was deposited and made available in 2013. The video accompanying the deposition of the AtlasCBS script in StarDrop gives further details. The purpose of this brief note is to document the combined use of ChEMBL and AtlasCBS within the StarDrop environment.

The exponential growth of the content of SAR-Data Bases makes it difficult to update on a regular basis the content of the three databases that were part of the AtlasCBS server in 2011: PDBBind, BindinDB, ChEMBL. It was suggested in a later publication<sup>2</sup> that an alternative -and possibly more effective- way to overcome this problem would be to use the concepts of the AtlasCBS in a more local and custom-driven implementation, rather than a central, web base application. This is what the combination of the ChEMBL and AtlasCBS scripts within StarDrop allows the user to do.

Naturally, one can do searches and download the datasets of interest using the ChEMBL website directly. Typically, one goes to the server enters the query, i.e. ChEMBL target ID and selecting the appropriate columns one can download the data as a (CSV) formatted file and continue on to import these data into StarDrop. Note that ChEMBL has already data columns available for LE, BEI and SEI, some of the earliest Ligand Efficiency Indices introduced early in 2000s as well as the LLE introduced later. From the CSV file, one needs to edit out the semicolons and replaced them by commas in order to successfully upload the data. Using the ChEMBL script permits to simplify all these steps by just using the ChEMBL targetID, after activating the appropriate Custom Script in your StarDrop application.

After this step, make sure that any rows containing 'invalid' entries (i.e.?) in the PChemvalue columns are deleted, otherwise the AtlasCBS script will fail. AtlasCBS only needs for the user to specify the column from which the 'pKi' values should be extracted to calculate the various Ligand Efficiency Indices. These variables are added as additional columns after the specified 'pKi' column. In addition to the previously discussed: SEI, BEI, NSEI, NBEI, nBEI, mBEI; two more are calculated (LEH, LEP) directly related to the initial LE defined by Hopkins 2004 as has been discussed in a more recent publication<sup>3</sup>.

Using the visualization tools, the user can now display the content of the selected data in the most suitable framework. For this discussion, the selection of a pair of efficiency indices:

polarity and size (x,y), i.e SEI, BEI or NSEI, nBEI, would be the suitable to make it easier to interpret the results. I am providing images of these planes for the Neuraminidase target (ChEMBL target ID: 2051) to highlight the main features (Fig. 1a,b). The compounds and figures illustrated below relate to the illustrations and discussion that can be found in Chapter 8, Fig. 8.2., pg. 113)<sup>4</sup>.

Further explorations of the most suitable Ligand Efficiency Indices to drive each drug-discovery project can be easily implemented within this framework, including the incorporation of proprietary compounds to follow the optimization trajectory in 'efficiency planes'. This is an excellent a graphical and analytical way to effectively drive drug discovery in combination with Multiple Parameter Optimization (MPO) within StarDrop<sup>4</sup>.

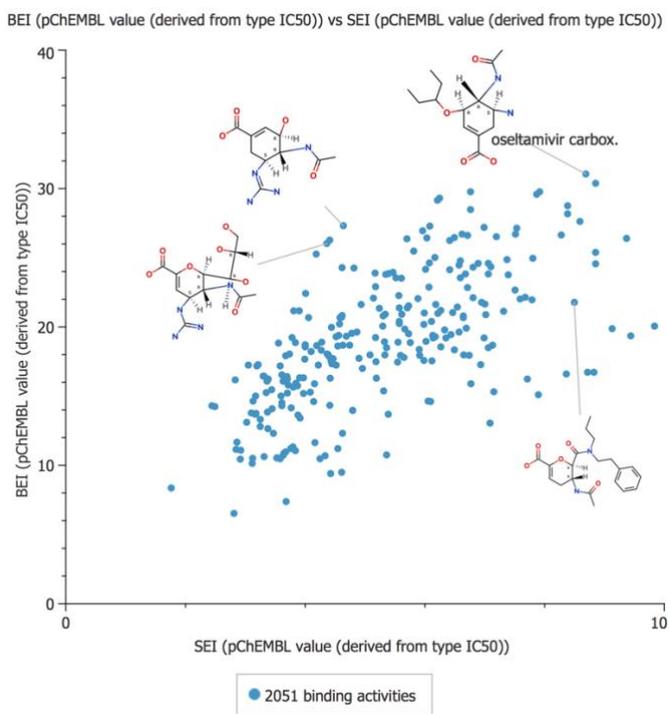
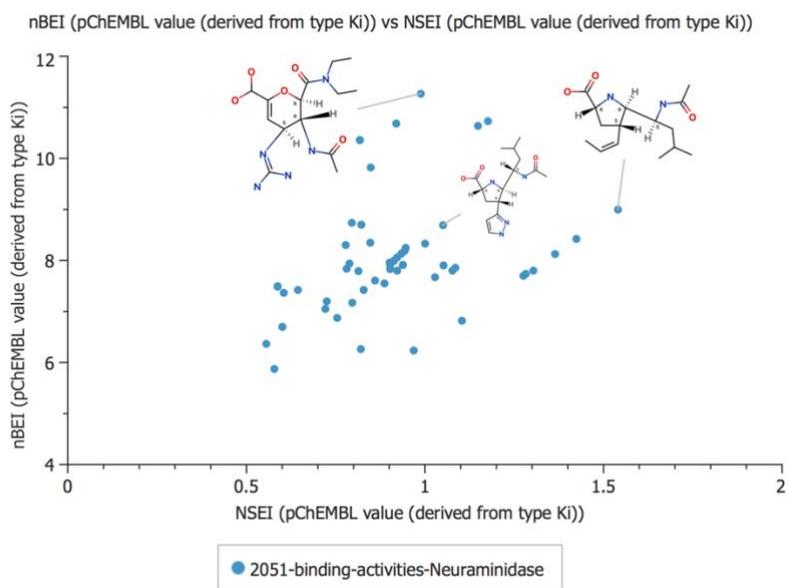


Figure 1a. BEI vs. SEI using pIC50 values. The compounds align along lines of slope  $10 \cdot \text{PSA}/\text{MW}$  (a measure of polarity, related to LogP). Optimum drug like molecules map on the upper right quadrant of both figures where the values for the two variables (x,y) are maximized. For reference, Oseltamivir carboxylate, the activated form of OSELTAMIVIR is at the top, right hand corner. Precursors, previous generation compounds (ZANAMIVIR and analogs) are also annotated further to the left. Polarity increases counter clockwise as PSA/MW increases. The slope is given by  $\text{BEI}/\text{SEI}$  as a continuous variable.  $\text{BEI} = \text{pKi}/\text{MW}(\text{KDa})$ .  $\text{SEI} = \text{pKi}/(\text{PSA}/100)$ .



**Figure 1b. nBEI vs. NSEI using pKi values. Note disposition of along lines with slopes equal to sum of N+O count (polarity). Polarity increases counterclockwise as the N+O count increases. Most polar compounds are in steepest lines. More drug like compounds and more potent are in the upper right-hand quadrant.**

#### Notes:

1. Cortés-Cabrera, A., Morreale, A., Gago, F. and **C. Abad-Zapatero**. [AtlasCBS: a web server to map and explore chemico-biological space](#). J. Comp. Aided Mol. Design. 2012. DOI 10.1007/s10822-012-9587-5.
2. **'Software and Techniques for Bio-Molecular Modeling'** edited by Dr. Azat Mukhametov and published by the Austin Publishing Group (2016). Chapter 2 describes the latest developments on the AtlasCBS concept, in particular the use of KNIME workflows to produce atlas-like plots and efficiency planes using simple KNIME modules.
3. Abad-Zapatero, C. Champness, E. J, M. D. Segall. Alternative variables in Drug Discovery: Promises and Challenges. *Future Medicinal Chemistry* (2014) 6(5) 577-593.
4. Abad-Zapatero, C. Ligand Efficiencies for Drug Discovery: Towards an Atlas-Guided Paradigm. (2013). Academic Press/ Elsevier.