# Managing Chemical Libraries with Screening Assistant 2.0

**Vincent Le Guilloux[1*]; Stéphane Bourg[2]; Lionel Colliandre[1]; Alban Arrault[3]; Philippe Vayer[3]; Luc Morin-Allory[1]**

1) Institut de Chimie Organique et Analytique, ICOA, Laboratoire de Chémoinformatique, UMR CNRS 6005, Université d'Orléans, BP 6759, F-45067 Orléans Cedex2, France
2) Fédération de recherche "Physique et Chimie du Vivant", FR CNRS 2708, Rue de Charles Sadron, F-45071 Orléans Cedex2, France
3) Technologie Servier, 27 rue Eugène Vignat, 45000 Orléans, FRANCE
* Corresponding author: vincent.le-guilloux@univ-orleans.fr

## Introduction

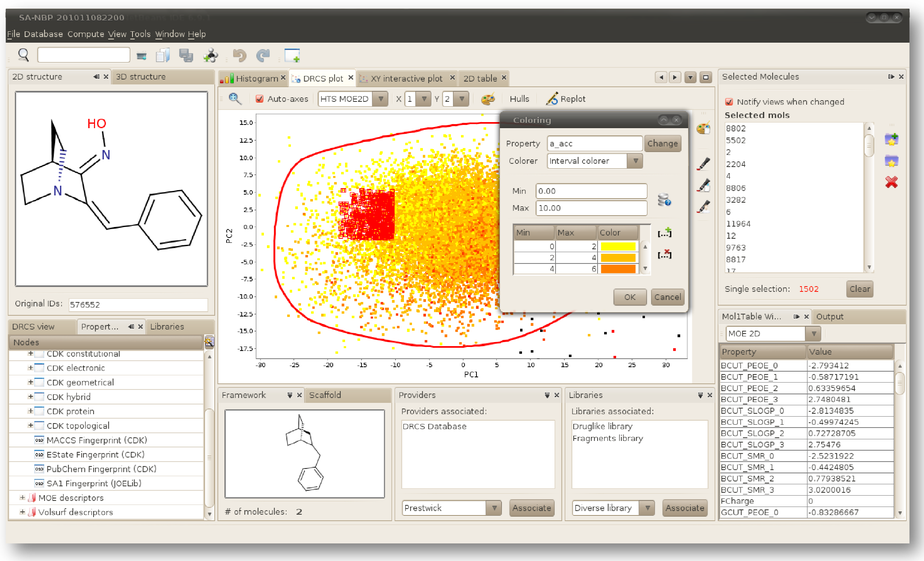
**Chemical libraries** are the daily bread of most drug discovery scientists. Up to millions of molecules can be screened in a single High Throughput campaign nowadays, while more advanced projects often contain hundreds to thousands of molecules. Faced to their increasing complexity and size, chemoinformatics and diversity analysis are appropriate tools to deal with chemical datasets[1]. Despite recent very promising efforts devoted to development of free and open-source chemoinformatics tools (e.g. Bioclipse), only few free software are specifically dedicated to assist the screening process and facilitate the selection, the analysis and the management of chemical libraries.

We present in this poster the second version of Screening Assistant[2] (**SA**), a free and open-source JAVA software dedicated to the storage and the analysis of small to very large chemical libraries. SA2 stores unique chemical structures using a MySQL database, and associates to the molecules various standard pre-computed descriptors as well as user-defined properties/descriptors that can be imported in a flexible way. Various standard and advanced chemoinformatics methods have been implemented, including **chemical space visualization/creation**, **substructure and similarity searches**, **diverse subset** extraction and **diversity indices** calculation. Its modular architecture, based on the **NetBeans Platform**, eases the addition of new functionalities to the software. The program and source code are **freely available** (GPL).

## Key features of SA 2.0

**Store, analyze and create chemical libraries using chemoinformatics methods**

**Modular and extendable software**




### Manage molecules and their properties

- Use IUPAC **InChI** identifier to store **unique** molecules (up to several millions)
- Associate molecules with **providers** and **libraries**
- **Import and organize new properties / fingerprints**

### Visualize and profile libraries

- Visualize libraries using **Delimited Reference Chemical Subspaces[3]** (DRCS) and **interactive 2D plots**
- Create your own DRCS model and contours
- Basic statistics: histogram on properties, provider repartition…
- Calculate various **diversity indices**

### Create new libraries

- Create custom **filtered libraries** using any available descriptor(s) and classical pre-computed HTS flag(s)
- Create **diverse libraries** (algorithm based on frameworks)
- Create scaffold-based, merged, complementary libraries…

### Chemoinformatics

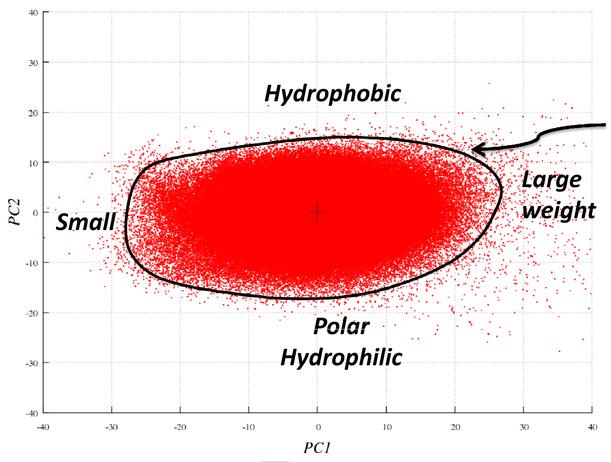
- Compute molecular descriptors and standard flags: **Lipinski RO5, Fragment RO3, Reactive, Warhead flags**…
- **Substructure search, similarity search**…
- **CDK** (which is used by default), JOELib and Marvin modules

## Delimited Reference Chemical Subspaces[3]

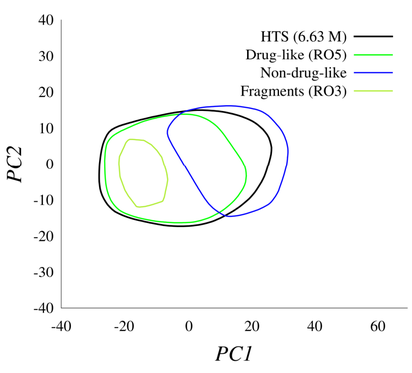
### DRCS-MOE 2D (HTS compounds)

Project molecules in viewable 2D space using a PCA model computed on a database of 6.63 million HTS molecules

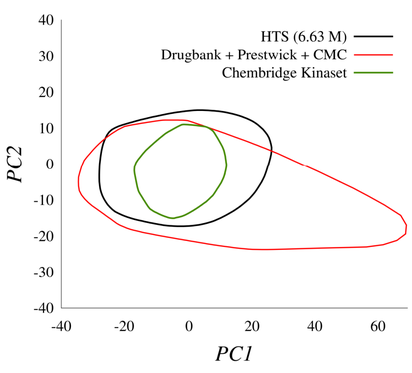
*See companion poster of Colliandre & al.*




**DRCS Contour based on convex hull calculation (encompasses 99.7 % of HTS molecules)**




**Creation of compounds specific subspace contours**




## Exploiting libraries: a simple workflow

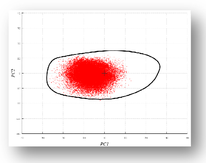
**In-house tools**

Based on the SDF file representing your library(ies), standardize your molecules (or create a standardization module in SA!) and compute descriptors that require proprietary software (e.g. MOE).

**SA 2.0**

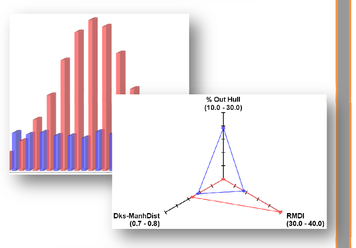
### *Prepare SA database*

- ✓ Import molecules and properties in a SA database
- ✓ Enable the computation of CDK descriptors, scaffolds…
- ✓ Eventually, create your own reference space



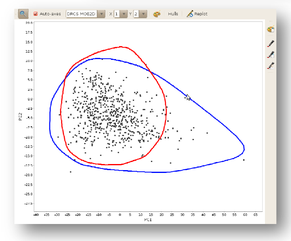
### *Analyze libraries*

- ✓ Assess chemical space coverage (DRCS), check outliers
- ✓ Compare distributions of properties / descriptors
- ✓ Compute diversity indices on various DRCS




### *Create filtered / diverse library*

- ✓ Create a new filtered library (removing reactive, non- drug-like molecules, using a single provider…)
- ✓ Create diverse subset based on this filtered library
- ✓ Export the new library for further processing



**In-house tools**

Perform experimental screening. You can subsequently import the results back into the database as new properties of molecules, and use the visualization facilities of SA to analyze the data and **link the chemical space to the biological space**.

## Conclusion & Perspectives

- ✓ Store and manage (large) chemical libraries
- ✓ Interactive visualization in various chemical spaces (DRCS)
- ✓ Easy creation of new libraries based on diversity and / or filtering procedures
- ✓ Comprehensive user / developer documentation
- ✓ Modular and documented architecture enabling the software to be extended / customized

### Under development

- ❖ Compute multiple **diversity indices** and associate intuitive visualization charts.
- ❖ Use of **Self Organized Map** to create non-linear chemical spaces representations

**http://sa2.sourceforge.net/**

## References

1. Sukuru SC. & al., **2009**, *J. Biomol. Screen.* 14(6):690-699
2. Monge, A & al., **2006**, *Mol. Divers.* 10(3):389:403
3. Le Guilloux, V & al., submitted to *J. Chem. Inf. Model.*