An Intuitive Workflow to Enumerate and Explore Large Virtual Libraries

Aishling Cooke, James Chisholm, Ed Champness, Peter Hunt, Matthew Segall, Tamsin Mansley
Optibrium Ltd., Cambridge, UK., info@optibrium.com

Introduction

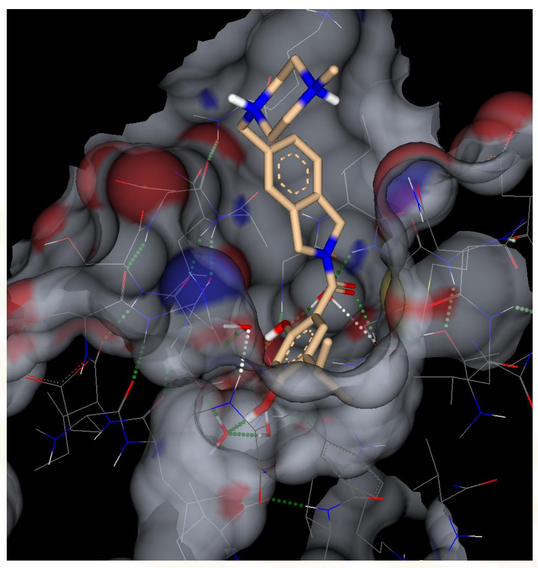
Enumeration of a virtual library based on cores or scaffolds of interest helps to quickly explore potential substituents around hit or lead series and prioritise strategies that are most likely to yield high quality compounds. To be most effective, a library enumeration workflow should seamlessly combine a number of important elements:

- Search of commercial compound providers or internal collections to find the most relevant building blocks that are readily available
- Flexible 'clipping' of the building blocks to define the corresponding R-groups
- Easy enumeration of a virtual library using the resulting R-groups
- A seamless link to predictive models, docking and multi-parameter optimisation to quickly prioritise the highest quality products for synthesis
- A link from the products to the corresponding building blocks for reagent ordering

In this poster, we will illustrate a visual and intuitive workflow linking all of these capabilities. All of the steps were performed in the StarDrop™ software [1], linked with eMolecules™ web services [2] for building block searches and LeadIT™ [3] for docking.

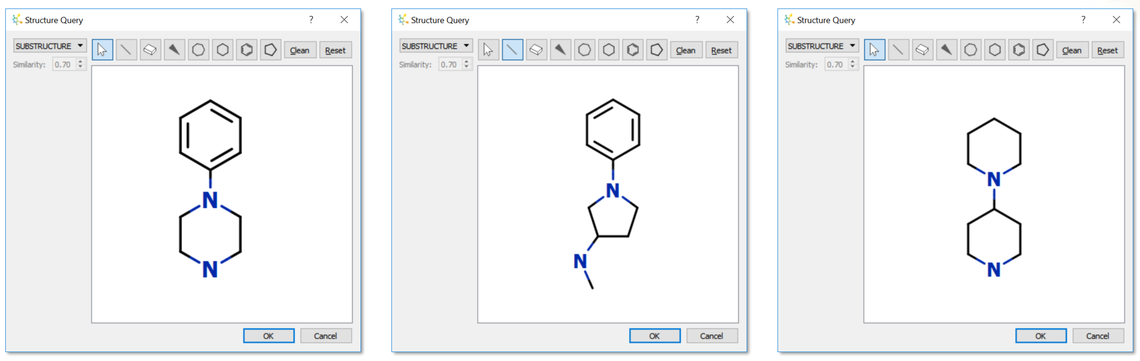
Example

The crystal structure on the right (PDB 2XJX) shows the binding site of heat shock protein 90 (HSP90) with a co-crystalised ligand. The beta resorcinol group forms a tight hydrogen bond network in the binding site, but the 5-(piperazin-1-ylmethyl)-isoindoline does not form any strong interactions with the protein. We will explore the hypothesis that high quality compounds with better affinity can be proposed through enumeration and prioritisation of a virtual library, based on an amide coupling reaction with a beta resorcylic acid core and commercially available amines.



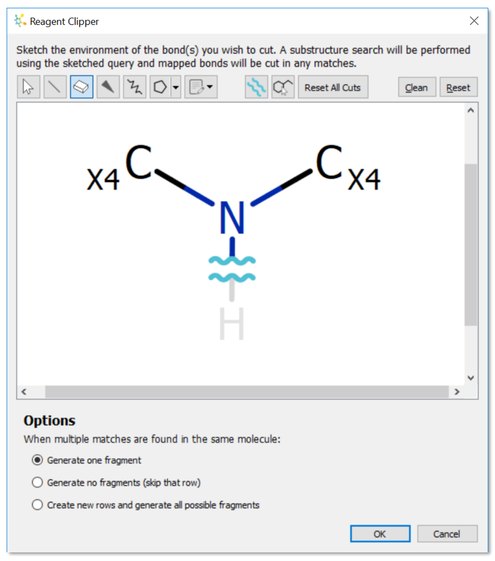
Building Block Search

Searches for building blocks may be performed against an internal collection or commercial vendor catalogues. Here, we show some example substructure searches of the eMolecules catalogue for suitable secondary amines.



R-group Clipping

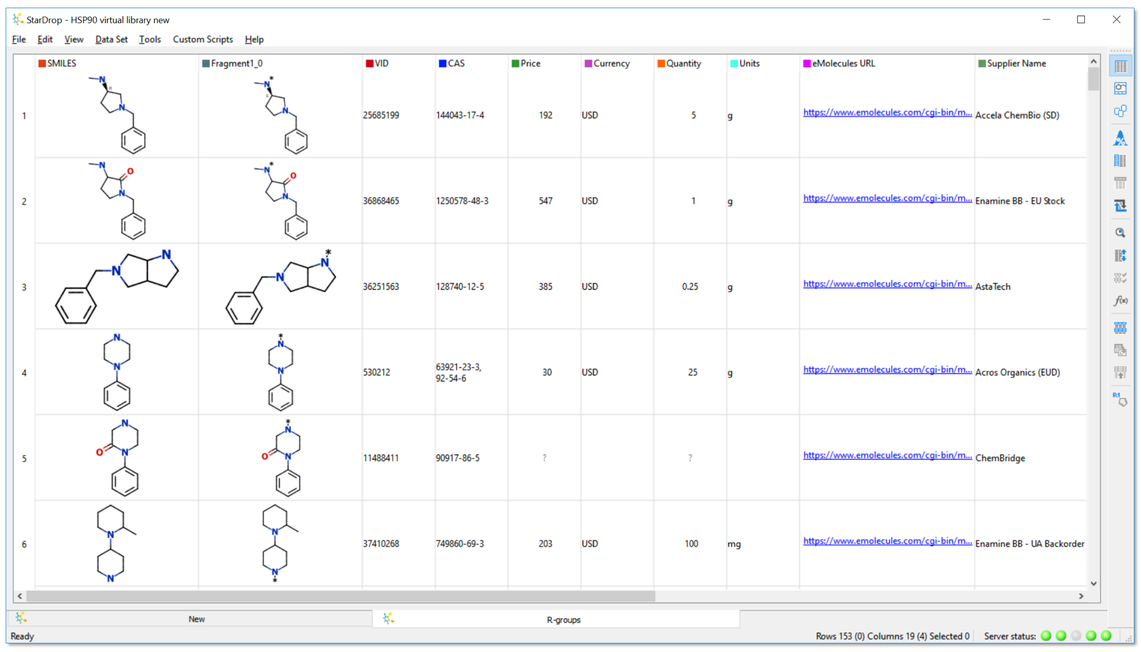
The building blocks can be clipped using a flexible substructure search tool that enables the reaction centre and the attachment point of the resulting R-group to be precisely defined. Here we define a secondary amine, ensuring that the adjacent carbons are sp^3 hybridised:



It is also important to provide options to handle building blocks in which the reaction centre occurs multiple times. The safest option, selected above, is to choose not to generate an R-group because such a building block may result in multiple products from the same reaction.

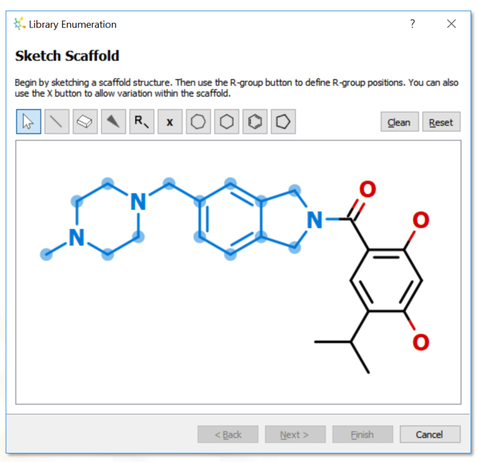
R-group and Building Block Information

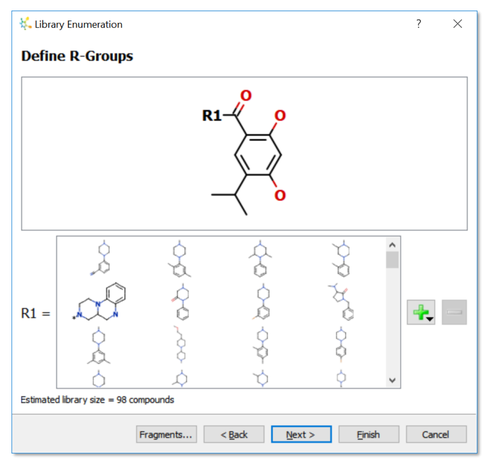
The result of the searches and clipping is a data set containing R-groups and the corresponding building blocks with ordering and availability information.



Library enumeration

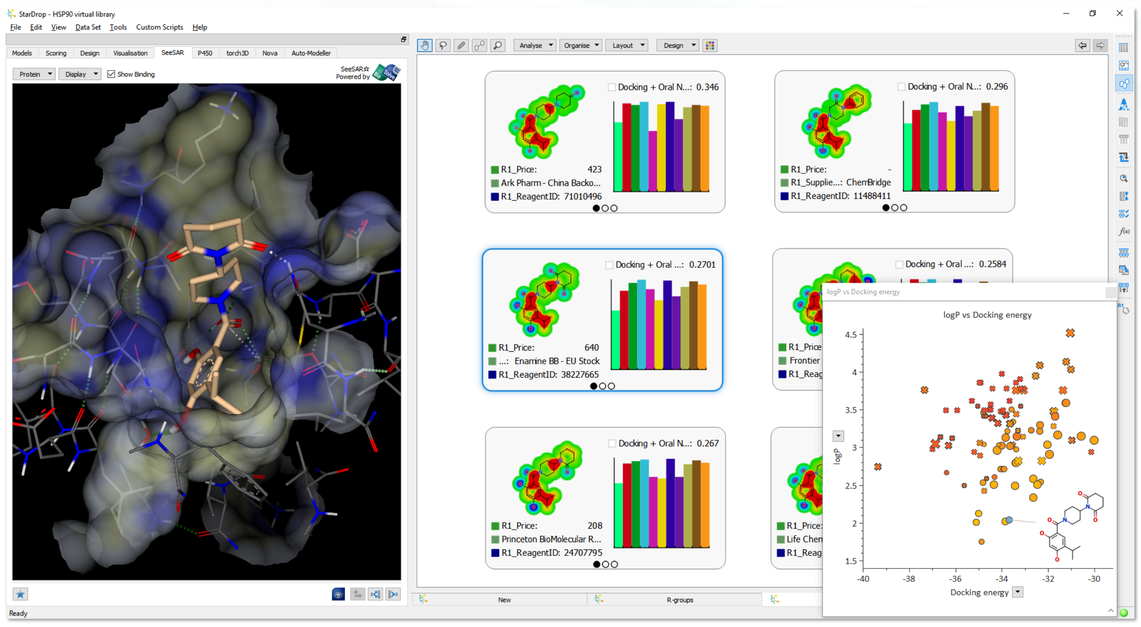
These R-groups can then be used in a library enumeration. This example illustrates a library based on the co-crystalised ligand, where the highlighted atoms are replaced by the secondary amine building blocks on a resorcinol scaffold.





Prioritisation of Products

Predictive models can be applied to assess the resulting products. In the example below, the products have been docked with the HSP90 protein structure using FlexX™ [4] in LeadIT [3]. A broad range of physicochemical, absorption, distribution, metabolism and excretion (ADME) properties have also been predicted using StarDrop's ADME QSAR models. The resulting docking scores and property predictions have been assessed against a multi-parameter scoring profile using Probabilistic Scoring [5] to identify those products with the best chance of achieving the required property criteria.



The building block information is retained for each product making it easy to select the best compounds and order the corresponding reagents.

Conclusion

When integrated within a comprehensive environment for compound data analysis, design and predictive modelling, library enumeration can provide an efficient approach to quickly explore and prioritise many synthetically accessible optimisation strategies.

References

[1] StarDrop: www.optibrium.com/stardrop
[2] eMolecules: www.emolecules.com
[3] LeadIt: www.biosolveit.de/LeadIT/
[4] FlexX: www.biosolveit.de/FlexX/
[5] M.D. Segall (2012) *Curr. Pharm. Des.* 18(9) pp. 1292-1310