Compound Enumeration using Reaction Workflows

Jameed Hussain, Gianpaolo Bravi and Mike Hartshorn

Introduction

The ability to enumerate virtual chemical structures is vital in the design and synthesis of chemical arrays. It is now straight-forward to enumerate many compounds for a given chemical reaction. However, the synthesis of compounds typically involves several steps. In addition, the drive to improve the efficiency of multistep synthesis, convergent synthesis is increasingly common. These more complex reaction schemes can present a challenge for compound enumeration systems.

Reaction Workflows

Reaction Workflows (Rw) is a new web-based application developed by Dotmatics to perform compound enumeration. The application uses a graph to represent a reaction workflow with nodes to represent the reagents and reactions. The products of a reaction can also be passed to another reaction. These reaction workflows are akin to the reaction schemes we are familiar within compound synthesis.
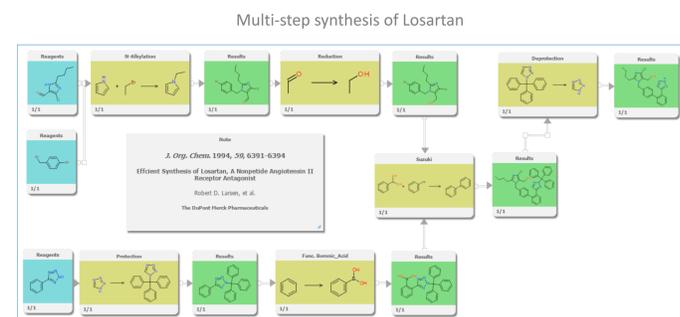
Simple to use
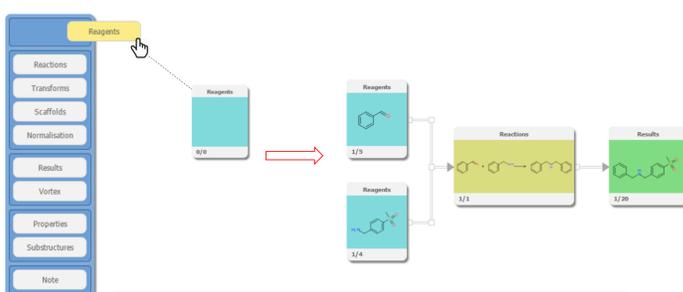
A simple drag and drop user interface to build chemical reaction schemes

The application uses an intuitive drag and drop interface to allow users to build up a reaction workflow. These can be run and the resulting products formed can easily visualized and downloaded. The intuitive interface means it is straight-forward for bench chemists to build simple or more complex reaction schemes. Multi-component reactions schemes can be built easily - an example of a four component Ugi reaction built with Rw is shown below.

A four-component Ugi reaction

Example of a multi-component Ugi reaction.
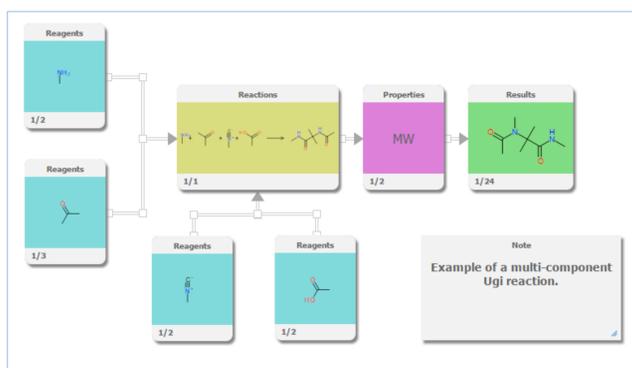
Rw allows the products of a reaction to be passed to another reaction. This makes it possible to put together a complex multi-step synthesis. This is illustrated below with a workflow for the convergent synthesis of Losartan[1].

Multi-step synthesis of Losartan

J. Org. Chem. 1994, 59, 6391-6394
Efficient Synthesis of Losartan, A Nonpeptide Angiotensin II Receptor Antagonist
Robert D. Larsen, et al.
The DuPont Merck Pharmaceuticals

Extensive Reaction List

The generation of SMIRKS (or RXN files) for a reaction can be difficult for non-expert users. Therefore, the application has an extensive reaction list available[2] which means in most cases the user simply needs to select a reaction. In the situations where a reaction is not available, a chemical sketcher can be used to define a reaction.
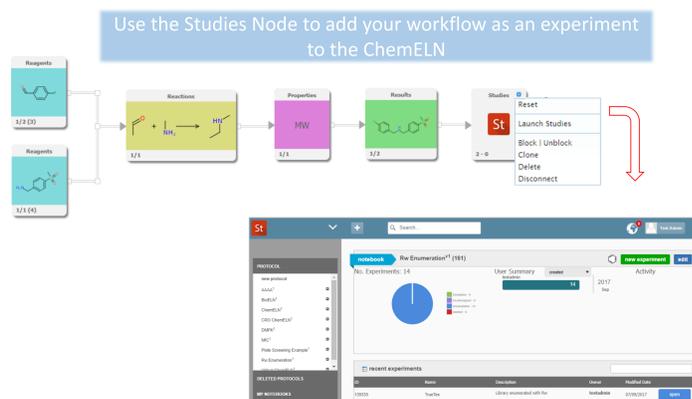

Extensive list of pre-built chemical reactions - straightforward to add more

ChemELN Integration

A reaction workflow can be added as a chemical experiment to the ChemELN using the Studies node. This will add the reagents, reaction and the products details to a new experiment in the ChemELN.
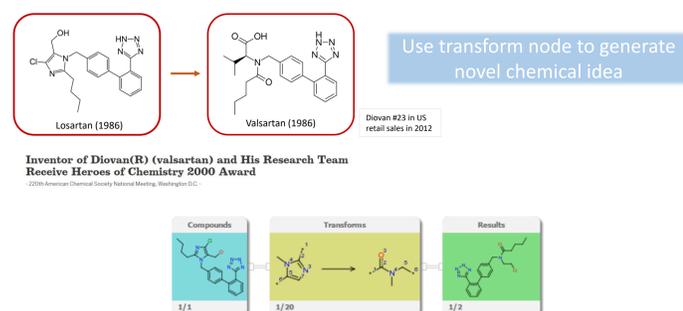
Use the Studies Node to add your workflow as an experiment to the ChemELN

Scaffold based enumeration

Markush type scaffold or R-group enumeration is also available in Rw. This type of representation appears in chemical patents and is more common in the lead optimization phase of drug discovery. Additionally, it is possible to mix reaction and scaffold enumeration in a single workflow which makes it easy to use real structures in an R-group enumeration.


Scaffold based enumeration

Mix reaction and scaffold enumeration in single workflow
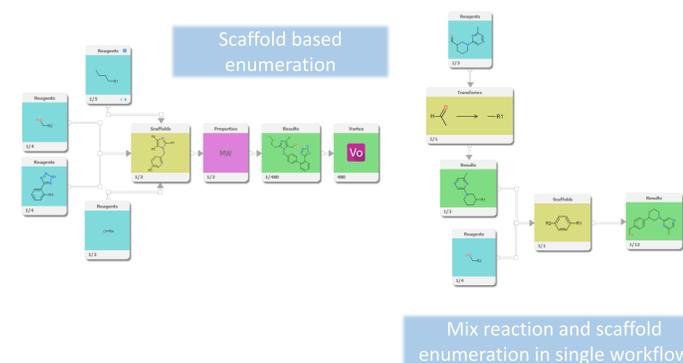
Idea Generation

The application has list of small bioisoteric changes or transforms that can be applied to compounds. These can be used to generate novel chemical ideas. Additionally, any external source of transforms (such as matched molecular pair analysis) can also be used in Rw.


Losartan (1986) → Valsartan (1986)

Diovan #23 in US retail sales in 2012

Use transform node to generate novel chemical idea

Inventor of Diovan(R) (valsartan) and His Research Team Receive Heroes of Chemistry 2000 Award
220th American Chemical Society National Meeting, Washington D.C.

Structure Normalization

A common task that needs to be performed when working with structures from external sources (suppliers etc.) is normalization. Here structures are standardized to a preferred representation (e.g. change nitro groups to pentavalent or change separated). This can be carried out in Rw with the Normalization node where the preferred representation can be selected and applied to a set of structures.

Select the representation rules to use. Then apply to clean and standardise your structures

---
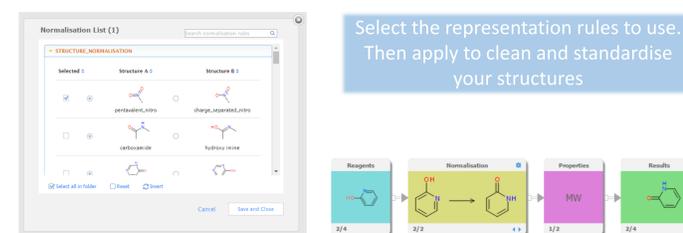
Compound Annotation and Filtering

The property node can be used to calculate properties for compounds. It can also be used to filter compounds based on values that are set by the user. These filters can be saved as a profile and shared with others. The substructure node can also be used to annotate and filter based on substructures. This is particularly important when one needs to remove building blocks from an array which contain functional groups incompatible with the reaction specified.
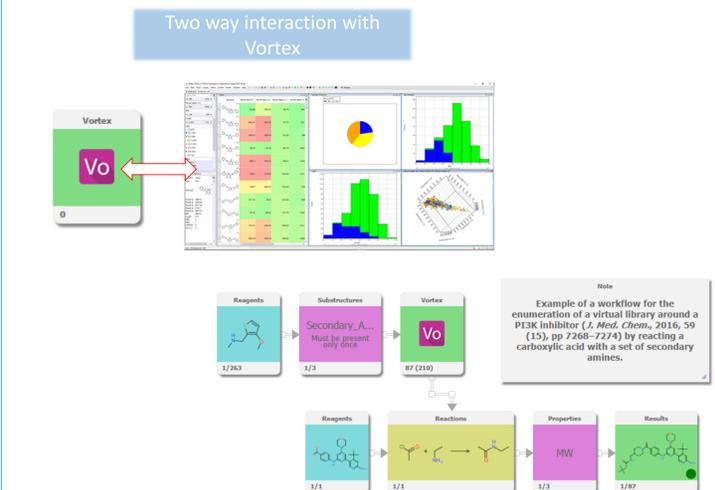

Filter using substructures or properties
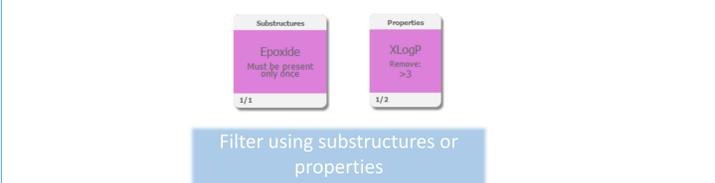
Desalting and Neutralization

Additional compound "cleaning" can be performed within Rw. This includes desalting (removing the salt for a compound) and neutralization (removing the charge from a compound). The functionalities are performed by the Normalisation node.

Clean (desalt/neutralise) structures

Two way interaction with Vortex

Vortex is an interactive data visualization and analysis application which can be used to explore large chemical datasets. It can be used in conjunction with Rw for visualisation or advanced profiling and selection. This is extremely useful when a selection needs to be made from a large list of building blocks or products.

Two way interaction with Vortex

Example of a workflow for the enumeration of a virtual library around a PI3K inhibitor (*J. Med. Chem.*, 2016, 59 (15), pp 7268–7274) by reacting a carboxylic acid with a set of secondary amines.

References

1. Larsen, Robert D., et al. "Efficient synthesis of losartan, a nonpeptide angiotensin II receptor antagonist." *The Journal of Organic Chemistry* 59.21 (**1994**): 6391-6394.
2. Hartenfeller, Markus, et al. "A collection of robust organic synthesis reactions for in silico molecule design." *Journal of chemical information and modeling* 51.12 (**2011**): 3093-3098.