Subject: Enumeration of Libraries
Posted by frytz on Fri, 04 Feb 2022 15:16:59 GMT
View Forum Message <> Reply to Message

I am trying to enumerate a library of compounds. A+B+C.

A+C are predefined while B should be more flexible in length and atomic constitution.

I am using U, Np as indicators to show where the molecules should be joined.

I am not able to define U-X-Y-Np (B) in a flexible manner to join A+C in many different ways. 8o

How can I draw fragment B in a way that it will accept many different structures and features?

Thank you for your help.

## File Attachments

1) Screenshot 2022-02-04 161317.png, downloaded 973 times




Subject: Re: Enumeration of Libraries
Posted by thomas on Thu, 10 Feb 2022 10:04:02 GMT
View Forum Message <> Reply to Message

DataWarrior's concept behind the combinatorial enumeration is the definition of substructure based generic reaction. The reactant substructures are supposed to be substructures of real building blocks. That means they should contains the functional groups which are needed for the reaction to work. Atom and bond query features as well as exclude groups can be used as constraints, such that a substructure search in real building blocks only matches those compounds, which are reactive enough for the reaction to work.

Different from this concept combinatorial libraries are sometimes defined using pre-processed building blocks, where a real chemical functionality is replaced by pseudo atoms (U,Np,...), which basically represent a connection point. Two or more lists of pre-processed (sometimes called 'clipped reactants') can be joined by removing matching pseudo atoms and connecting their neighbours by a new single bond. This concept has its limitations, but is fast to do once the clipped libraries exist. While DataWarrior's reactor was not built with this concept in mind, it can be

used (with limitations) for both, pre-processing building blocks to build a clipped library and to process multiple clipped libraries to generate product structures. The limitation I see is that a clipped library that contains multiple connection points (pseudo atoms) must not contain building blocks of type U-X-Np and U-X---Y-Np at the same time. You can define the generic reaction as follows.

If some of your B reactants have U and Np connected to the same atom, then these are missed by the reaction above. For those you may do a second run with this reaction and afterwards merge the compounds generated by both runs.

(Note: there is an invisible single bond in the second reactant between ? and Np)

Hope this answers the question...

Thomas

## File Attachments

1) t.png, downloaded 1552 times




2) t2.png, downloaded 1475 times

