By Robert Gotwals

Computing CoVID-19: A Presentation of the North Carolina School of Science and Math, Summer 2020

COMPUTING CoVID-19: SUMMER 2020

COMPUTING A DRUG:
GENERATING A SCREENING SET

Developer:
Robert Gotwals
May 21. 2020

1

INTRODUCTION

The real power of a computer, and of computational science, is that it can do a lot of repetitive calculations or tasks very quickly, much more quickly than can be done by a human.

A great example of that is in what is known as *screening*. Using computers, we can generate hundreds to thousands of potential drug compounds computationally, and test, or screen, these compounds against some target receptor, such as the ACE2 protein receptor, the part of a host cell that attaches to the SARS-CoV-2 virus and facilitates its entry into the host cell, where it replicates.

Figure 1 shows an example workflow for a screening process. In this example, we are starting with 605 million (!) compounds. We then use two different docking processes, one using a docking software package known as Smina and the next using the Glide docking software. This eliminates all but 2783 compounds! Then, a machine learning algorithm known as clustering reduces that set to 659 potential compounds. We then screen those against pharmacokinetic (PK) parameters that you learned about previously. These PK parameters are mostly ADME (absorption-distribution-metabolism-excretion) considerations, and include measures such as logP, logS, and Lipinski scoring. We are now down to 144 potential compounds. Molecular dynamics (MD, a computational chemistry technique) brings our numbers down to 29. Toxicity studies eliminate all but 16 of those drugs that will undergo further testing, including synthesis in a synthetic organic chemistry lab for testing in animal models, such as mice.
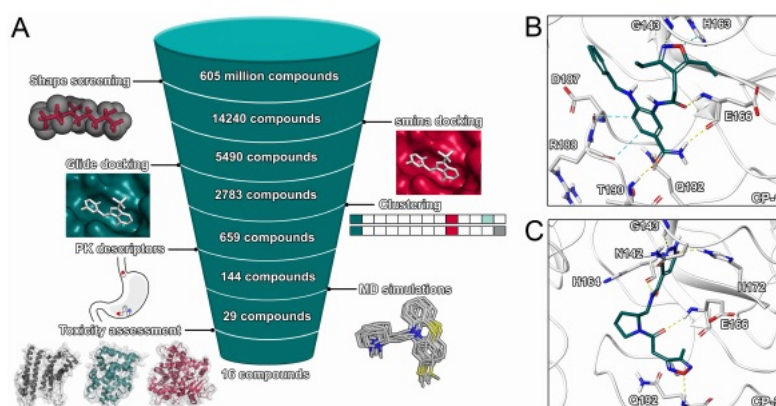
Figure 1: Workflow of a large drug discovery screening process [2]

So how do we get the 608 million compounds to start the process? In drug discovery, a great example of that is in a process known as *chemical transformations*. This is a relatively simple concept: we start with a lead drug (like you did in the morning lab) and modify the structure computationally hundreds if not thousands of times, not unlike the very small lead drug modification you did in the morning lab! The difference, of course, is that the computer does these transformations for you.

Medicinal chemists are very knowledgeable about the types of chemical transformations that are of the most value in designing a new drug. Segall *et. al.* ([1]) describes how a chemical transformation library was developed for the NOVA module, found in the StarDrop software. Figure 2 shows the seven broad categories of transformations, also described in the literature description below:

Table 1. Distribution of Transformations between Groups

group	number of transformations
functional group addition	20
linker modification	54
remove atom	5
ring addition	13
ring modification	26
ring removal	4
terminal group exchange	84
total	206

Figure 2: Categories of transformation types in StarDrop [1]

> Two hundred and six transformations were generated by study of medicinal chemistry literature and observation of the optimization steps between known drugs and the lead molecules from which they were derived.
>
> The transformations were divided into seven broad groups: functional group addition, linker modification, remove atom, ring addition, ring modification, ring removal, and terminal group exchange.

Table/Figure 3 shows a sample screenshot of some of the types of transformation rules that are available in StarDrop. Some of them, like the addition of a methyl group (CH_3-) to an amine (NH_2) is a very simple transformation, and very similar to the transformations you did with the first lab, adding four alkyl groups (methyl, ethyl, propyl, and butyl) to a carbon. Some transformations, however, are much more complicated, such as adding a sulfonamide group to a benzene ring or adding a carbon to a carbonyl (C=O). Generally, however, these are fundamental chemical transformations that you will study in significant detail when you take organic chemistry as a sophomore in college!

Table 2. Example of Transformation Rules

Group	Transformation Name	Illustration	
Functional Group Addition	Methyl addition to amine		
	Sulfonamide addition to benzene		
Linker Modification	Secondary carbon to carbonyl		
	Ester to amide linker		

Figure 3: Partial screenshot of transformation rules from StarDrop [1]

Generating hundreds to thousands of new compounds by applying one or more transformations to the lead drug is a computationally simple process, but one that requires a relatively significant amount of time. For example, for this lab, you will be applying *one* transformation to the lead drug that you modified in the morning lab, the experimental lead drug from Diamond Light Source's "XChem"

<div align="right">

2

</div>

STUDENT ACTIVITY

NOTE! The majority of the steps for the activity will be demonstrated in the webinar. These instructions are meant only as short reminders of the steps you need to take to effectively modify the lead drug and test the new compounds!

2.1 PART 1 – PREPARATION OF THE LEAD DRUG

For this part, we want to import the lead drug (TRY-UNI-714a760b-6) from the file "lead-drug.csv" (on Canvas) into StarDrop. You might want to consider double-clicking on the name, change TRY-UNI-714a760b-6 to "Lead Drug", something easier to read!

2.2 PART 2 –PERFORMING CHEMICAL TRANSFORMATIONS

For this calculation, we want to use the NOVA module in StarDrop. Make sure your lead drug is selected, and proceed to the NOVA tab.

You can start the transformation process by clicking on the arrow in the bottom-right corner of the window.

Select "Chemistry Transformations" from the list of menu items. We want to only transform the chlorobenzene portion of the molecule, which is located on the right side of the lead drug. Draw a circle around everything except the chlorobenzene ring.

Now you are at the Transformations menu page. Click on "Local" to DESELECT all of the checked boxes. Now, click on the dropdown arrow for "Functional Group Addition". We want to add these functional groups to benzene:

1. Carboxyl addition to benzene

2. Chloro addition to benzene

3. Fluoro addition to benzene

4. Hydroxyl addition to benzene

Click "Next", taking you to the Control Output window. We want to generate four (4) generations of new compounds. Typically, we would generate many generations to create a large number of transformed compounds!

Clicking on "Next" gets you to the "Apply Filters" window. We'll ask StarDrop not to include any chemically undesirable fragments. Leave the "undesirable" box checked.

Clicking on "Next" gets you to a "Create New Data Set" window. Call your dataset "Transformations". Clicking on the "Finish" button will run the transformations. We have not asked for many transformations or many generations, so this run should take place relatively quickly. You should see the words "Transformations: Started" in the panel at the bottom left, and a radial graphic will appear in the window on the right when things are done.

2.3   PART 3 – CLEANING UP THE TRANSFORMATIONS DATA SET

It is helpful if we clean things up a little, and we'll do that in this section:

1. On the bottom left, you will see two tabs, leaddrug and Transformations. From the View meno, select "leaddrug", then from the File menu, select "Close Data Set". You should now only have your Transformations tab on the bottom right. You don't have to save it.

2. At the top right, click on the icon that looks like a grid, or spreadsheet. You should now have a listing of compounds. QUESTION: how many compounds (including the lead, or parent, drug do you have?

3. There are columns we don't need. You can Control-click or right click on ID and Transformations (one at a time) and delete them, leaving you with the structure and its generation.

Your dataset is now ready for analysis!

2.4   PART 4 – ANALYSIS

1. Clicking on the "Show Details" button at the bottom will show you how many compounds you have, and the transformations that you performed. Typically, a scientist would transcribe notes to him/herself about this particular dataset.

2. Click just under the magnifying glass at the top right. This should select all of your compounds. We want to run some models, so select the "Modeling" tab. Run: logP, logS, 2C9pKi (binding affinity), MW (molecular weight), and rotatable bonds.

Now you are ready to conduct an analysis of your results. You will use the "Visualization" tab. Once in the Visualization panel, you might need to go to View and select the Transformations data set. Here are some analyses to conduct:

1. Histogram: choose the Histogram option

    (a) plot Generation on the x-axis and logP on the y. If I need compounds with a low logP value, which generation of compounds would I select?

    (b) Now plot Generation vs. logS. Same thing, I need a low logS. Which generation is best?

    (c) 2C9 pKi is a measure of how well a drug binds to a receptor, for example, how well a drug binds to the ACE2 protein receptor on a host cell. If we can design a drug to bind to the ACE2 receptor, then there is no room for the virus to bind and do its damage. The higher the pKi, the better it binds. Given that, which generation has the best pKi binding?

2. Cluster Analysis Clustering is a good way to find compounds that have some similarities and put them in families, or *clusters*. Cluster analysis is an important task in *machine learning*.

    (a) From Data Tools, go to Clustering. We want to cluster on 2C9 pKi, with an 80

    (b) Using Histogram, plot Generation (x-axis) and 2C9 pKi on y. Which generation is represented the most?

    (c) Using the 3D plot, plot logP on x, logS on y, and 2C9 on z. Take a look at the Cluster 3 compounds. They are clustered how: high logP, low logS? High logP, high logS? Low logP, low logS? Low logP, high logS?

2.5   PART 5 – PERFORMING A SCREENING STUDY

For this part, we want to screen our 575 compounds to reduce the number of targets. We'll use two criteria (one at a time):

1. We know that compounds with a logP less than or equal to 4 should be OK, but there is a possibility that compounds with a logP > 4 might also be candidates. Our first screening is going to be for that parameter.

2. From those compounds that have logP > 4, we want to screen them against a protein. We'll screen them against PDB ID:6LU7, the crystal structure of COVID-19 main protease.

As above, you'll need to rely on your attention during the webinar, but some instructions/guidelines are provided below:

1. To filter out compounds, go to Data Set –> Filter. You want logP with a value <= 4, and you want those compounds that meet that criteria to be deleted. It's actually a little safer to move them to a new dataset so you don't lose the original 575 compounds! How many compounds do you now have?

2.6   PART 5 – SCREENING AGAINS PDB ID: 6LU7

You have seen PDB ID:6LU7 in this course, it being the crystal structure of COVID-19 main protease in complex with an inhibitor N3. If you recall, the N3 inhibitor – a drug that inhibits the action of this protein, a protease, a protein enzyme that breaks down other proteins and amino acids – is identified as PJE in this structure. We want to use PJE as a reference, and compare, or *screen*, the compounds remaining in our dataset agaist this reference. What we will get is a value from 0 to 1 – the closer to 1, the better the comparison.

1. To run your screening, use the torch3D tool. You will import the 6LU7 protein, extract from protein, and use PJE as the reference. Run 100 conformations of the tool. Once the run is done (this takes a while!), filter out all 6LU7 runs that are less than 0.5. How many compounds do you now have?

REFERENCES

[1] Segall, Matthew, et al. "Applying medicinal chemistry transformations and multiparameter optimization to guide the search for high-quality leads and candidates." Journal of chemical information and modeling 51.11 (2011): 2967-2976.

[2] Fischer, André, et al. "Inhibitors for novel coronavirus protease identified by virtual screening of 687 million compounds." ChemRxiv Prepr (2020).