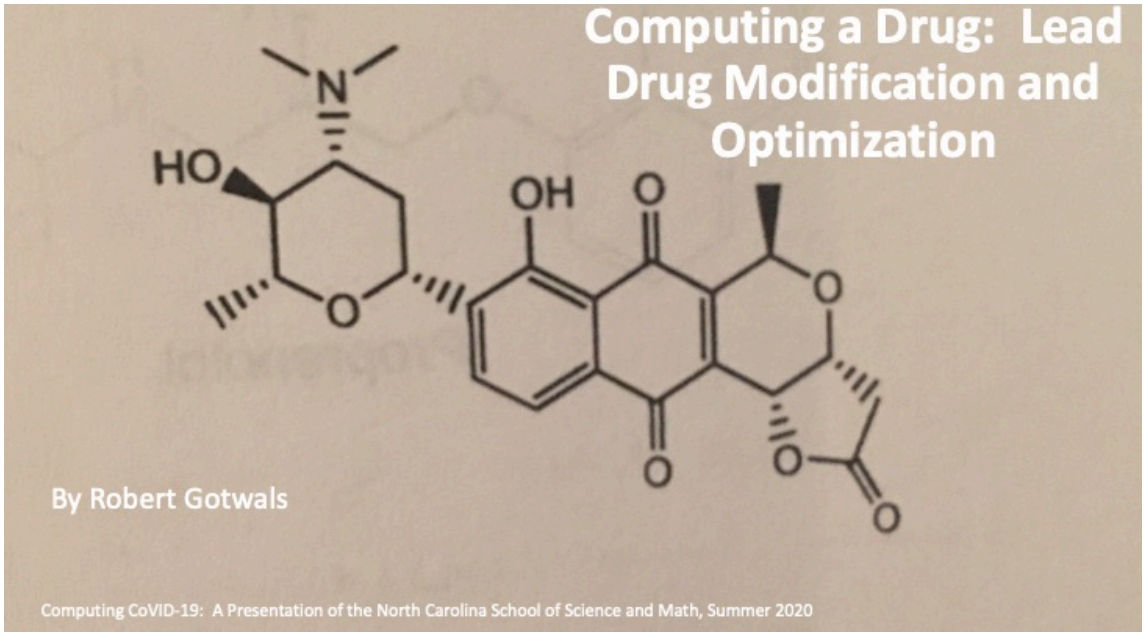


Computing CoVID-19: Summer 2020

Computing a Drug:

Lead Drug Modification and Optimization

Developer:
Robert Gotwals
May 11, 2020

1

Introduction

There are two primary efforts to reduce and/or prevent the coronavirus epidemic. The first is the development of a vaccine. A vaccine produces antibodies, proteins which attack the virus and reduce its effect. Successful antibodies not only reduce the number of people who can get sick, it also speeds up the development of *herd immunity*, a condition in which enough people are immune and the virus doesn't have enough hosts in which to thrive and reproduce.

The second effort is the development of a drug that will reduce or eliminate the symptomatic effects, including death, that a virus such as SARS-CoV-2 can generate. In this activity, we'll look at some of the ways in which new drugs are generated and evaluated for effectiveness.

Figure 1 shows a typical new drug development timeline, and you should be hearing a great deal about how this timeline has been significantly shortened in the search for a therapeutic drug. The typical timeline suggests 3-5 years for target identification and screening, 1-2 years for ADME(T) and pharmacokinetic/pharmacodynamics (PK/PD) studies,

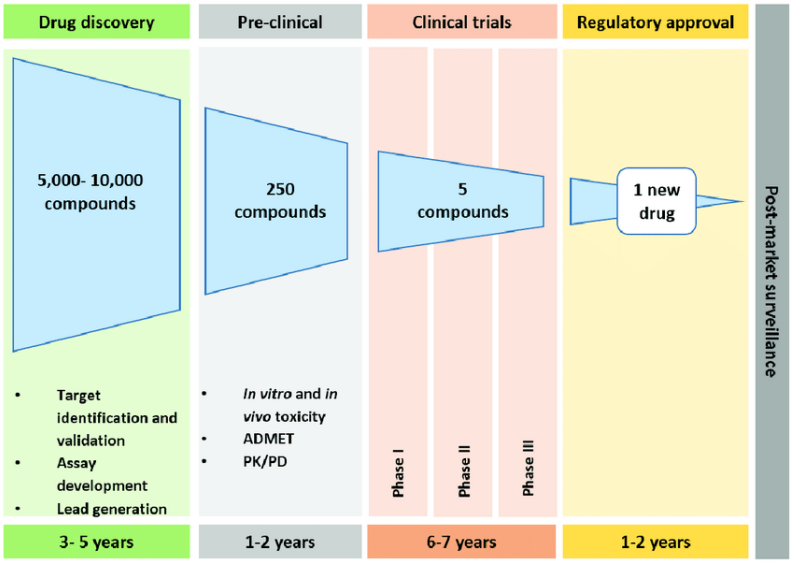


Figure 1: New drug development timeline [1]

then 6-7 years for clinical trials, and finally 1-2 years for regulatory approval, by the Food and Drug Administration (FDA) in the United States. One of the ways in which this timeline is being shortened for CoVID-19 is through "re-purposing" existing drugs, such as remdesivr (Gilead Sciences), a drug originally developed for diseases such as Hepatitis C. By re-purposing drugs, the drug discovery and pre-clinical steps can be ignored, and scientists can move directly into clinical trials. Human volunteers are currently participating in clinical trials.

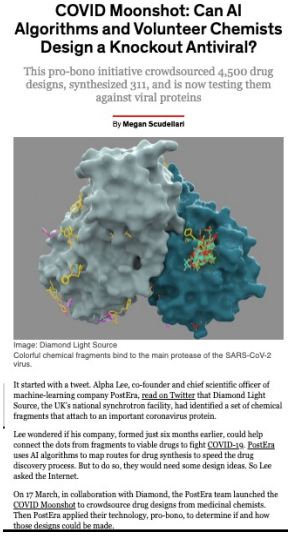
In recent developments (May/June 2020), the steroid dexamethasone was found to reduce deaths for patients on a ventilator or oxygen therapy by one-third. The paragraph below describes some findings from a drug trial in the United Kingdom:

> The UK RECOVERY trial is a Phase II/III randomised, controlled trial that began in March 2020 and is testing numerous potential treatments for Covid-19, including lopinavir-ritonavir, azithromycin, tocilizumab, convalescent plasma, and low-dose dexamethasone. The trial also had a hydroxychloroquine arm, but the UK Medicines and Healthcare Products Regulatory Agency (MHRA) suspended this due to no significant difference in mortality rate and hospital stay duration. More than 11,000 patients were enrolled in 175 NHS hospitals in the UK. A total of 2,104 patients were randomised to receive 6mg per day of dexamethasone for 10 days and then compared to 4,321 patients that received usual care alone (patients on ventilation, patients who required oxygen only, and those who did not need any respiratory assistance). Dexamethasone was found to reduce deaths in patients on ventilators by one-third and by one-fifth in patients who required oxygen only. No benefits were found in those who did not need respiratory assistance. With these results, the UK government approved the use of dexamethasone for Covid-19 patients needing oxygen and those on ventilators.([4])

For this lab, however, we wish to provide an example of *drug discovery*, which involves the screening of hundreds to thousands, even millions, of compounds. There are a myriad of ways to develop compounds for screening against some known target protein receptor. For example, the target protein receptor for SARS-CoV-2 is the ACE2 receptor, found on many host cells, especially cells in the lungs. In drug discovery, we want to develop a drug that will interact with the ACE2 receptor by connecting (binding) to it, thus preventing, or inhibiting,

the virus from doing the same thing. As such, we want to discover or develop a drug that has the right physical characteristics to bind tightly to that receptor.

Figure 3 shows how an organization, Diamond Light Source, helped to develop a crowd-sourcing program to identify new potential compounds. By using a variety of machine learning and other artificial intelligence techniques, the partners in the "Project Moonshot" consortium (Figure 2) are developing a library of compounds that can be screened against a variety of target proteins, including ACE2. The use of these advanced computational techniques are enabling scientists to shorten the timeline, especially the 3-5 year drug discovery portion of the timeline. The compound you will be studying in this lab, TRY-UNI-714a760b-6, is highlighted on the graphic of the dataset.



COVID Moonshot: Can AI Algorithms and Volunteer Chemists Design a Knockout Antiviral?

This pro-bono initiative crowdsourced 4,500 drug designs, synthesized 311, and is now testing them against viral proteins

By Megan Scudellari

Image: Diamond Light Source
Colorful chemical fragments bind to the main protease of the SARS-CoV-2 virus.

It started with a tweet. Alpha Lee, co-founder and chief scientific officer of machine-learning company PostEra, read on Twitter that Diamond Light Source, the UK's national synchrotron facility, had identified a set of chemical fragments that attach to an important coronavirus protein.

Lee wondered if his company, formed just six months earlier, could help connect the dots from fragments to viable drugs to fight COVID-19. PostEra uses AI algorithms to map routes for drug synthesis to speed the drug discovery process. But to do so, they would need some design ideas. So Lee asked the Internet.

On 17 March, in collaboration with Diamond, the PostEra team launched the COVID Moonshot to crowdsource drug designs from medicinal chemists. Then PostEra applied their technology, pro-bono, to determine if and how those designs could be made.

Figure 2: CoVID Project Moonshot

Dataset	Library Name	Compound SMILES	Modified Compound SMILES	Compound ID	Site	Confidence Annotation	Resolution (A)	Occupancy estimate	Z-Peak	PDB Status	PDB Entry
Mpro-x1077	DSIpoised_DMSO	N6CC=1C=CC[=CN1]N2CCCOCC2		Z1348371854	A - active	4 - High Confidence	1.45	0.23	5.61	Deposited	5RF6
Mpro-x0426	DSIpoised_DMSO	FC=1C=CC=CC1C(=O)NCCC=2C=CN=CC2		Z1310876699	A - active	4 - High confidence	1.43	0.62	5.79	Deposited	5RGK
Mpro-x1226	DSIpoised_DMSO	CCNCC1=CN(C)N=N1		Z1271660837	X - xtal contact (H80)	2 - Correct ligand, weak density	1.48	0.21	5.44	Deposited	5RFB
Mpro-x1237	DSIpoised_DMSO	CS(=O)(=O)CC1=NC=2C=CC=CC2N1		Z126932614	X - xtal contact (K12/D33)	2 - Correct ligand, weak density	1.41	0.17	4.35	Deposited	5RFD
Mpro-x1163	DSIpoised_DMSO	OCC1CN(CC=2C=CC=CC2)CCO1		Z1259086950	X - xtal contact (E240)	2 - Correct ligand, weak density	1.72	0.5	5.84	Deposited	5RGS
Mpro-x0104	DSIpoised_DMSO	CC(=O)NCCCC1=CNC=2C=CCCF=CC12		Z1220452176	A - active	2 - Correct ligand, weak density	1.59	0.78	6.47	Deposited	5R7Z
Mpro-x0107	DSIpoised_DMSO	CC(=O)NC=1C=NC=CC1C		Z1129283193	A - active	4 - High Confidence	1.88	0.36	5.47	Deposited	5RE4
Mpro-x0540	DSIpoised_DMSO	O=C1NCCC=1C=CN=CC1)NC2CCCCC2		Z111507846	A - active	4 - High Confidence	1.8	0.48	4.84	Deposited	5REH
Mpro-x2646	Moonshot	CC=1C=CN=CC1NC(=O)CC=2C=CC=C(C1)C2		TRY-UNI-714a760b-6	A - active - Moonshot	4 - High Confidence	1.83	0.74	5.41	Deposited	5RH2
Mpro-x2572	Moonshot	CC=1C=CN=CC1NC(=O)CC=2C=CC=C(C(N)C)C2		TRY-UNI-714a760b-20	A - active - Moonshot	4 - High Confidence	1.69	0.62	4.22	Deposited	5RGX
Mpro-x2649	Moonshot	CC(C)=O)NC=1C=NC=CC1C)C=2C=CC=C(C1)C2		TRY-UNI-714a760b-18	A - active - Moonshot	4 - High Confidence	1.69	0.8	4.51	Deposited	5RH3
Mpro-x2193	SpotFinder (Gyorgy Keseru)	CS(=O)(=O)c1ccc(cc1)N1CCNCC1		SF013	A - active	4 - High Confidence	1.57	0.6	7.8	Deposited	5RHD
Mpro-x0874	YORK3D	O=C([C@@H]1[C@H]2C=CC[C@@H]2C1)N		POB0129	A - active	4 - High Confidence	1.79	0.28	6.44	Deposited	5REZ
Mpro-x0887	YORK3D	OCC1(C2=NC=CC=C2)CCC1		POB0073	C - dimer (K137)	4 - High Confidence	1.65	0.26	5.65	Deposited	5RF0
Mpro-x2052	Cysteine Covalent Library (Weizmann Institute)	O=C(C1CCN(CCC1)=O)CC1)NC2=CC=C(OC)N=C2	OC1=CC=C]NC(=O)C2CCN(CC2)C(*)=O)C=N	PG-COV-42	B - active - covalent	2 - Correct ligand, weak density	1.56	0.38	4.82	Deposited	5RHE
Mpro-x2754	Cysteine Covalent Library (Weizmann Institute)	O=C(1CCN(CCC1)=O)CC1)NC)C2=CC=CC=C2	CN(C(=O)C1CCN(CCC1)C(*)=O)C1=CC=CC=C1	PG-COV-34	B - active - covalent	4 - High Confidence	1.76	0.72	5.51	Deposited	5RHF
Mpro-x0786	Cysteine Covalent Library (Weizmann Institute)	C1CC(=O)Nc1ccccc1)C(=O)NN2CCSCC2	CC(=O)Nc1ccccc1)C(=O)N1CCSCC1	PCM-0103072	B - active - covalent	2 - Correct ligand, weak density	1.6	0.24	5.77	Deposited	5REV
Mpro-x1348	Cysteine Covalent Library (Weizmann Institute)	COc1ccce2nc(NC(=O)C[N]nc12	COc1ccce2nc(NC(=O)C)nc12	PCM-0103067	B - active - covalent	4 - High Confidence	1.8	0.37	7.18	Deposited	5RFJ
Mpro-x0734	Cysteine Covalent Library (Weizmann Institute)	[O-][N+](=O)c1ccccc1N2CCN(CC2)C(=O)CCl	[O-][N+](=O)c1ccccc1N2CCN(CC2)C(=O)C	PCM-0103016	B - active - covalent	2 - Correct ligand, weak density	1.96	0.25	5.41	Deposited	5REM
Mpro-x1458	Cysteine Covalent Library (Weizmann Institute)	CC1C[N]CC(=O)C1CCN(CC1)C(=O)CCl	CC1C[N]CC(=O)C1CCN(CC1)C(=O)C	PCM-0102974	B - active - covalent	4 - High Confidence	1.9	0.37	6.43	Deposited	5RFY
Mpro-x1380	Cysteine Covalent Library (Weizmann Institute)	CICC1=O)N1CC1(CC1)C(=O)N2CCCCC2	CC(=O)N1CCC1(CC1)C(=O)N2CCCCC2	PCM-0102972	B - active - covalent	2 - Correct ligand, weak density	1.83	0.37	5.37	Deposited	5RFO
Mpro-x0705	Cysteine Covalent Library (Weizmann Institute)	Cc1ccc(cc1)C(=O)N2CCN(CC2)C(=O)CCl	CC(=O)N1CCN(CC1)C(=O)c1ccc(C)cc1	PCM-0102962	B - active - covalent	2 - Correct ligand, weak density	1.76	0.6	8.89	Deposited	5RGL
Mpro-x0831	Cysteine Covalent Library (Weizmann Institute)	Cc1ccccc1CN2CCCN(CC2)C(=O)CCl	Cc1ccccc1CN2CCCN(CC2)C(=O)C	PCM-0102911	B - active - covalent	4 - High Confidence	1.96	0.43	5.43	Deposited	5REY
Mpro-x1375	Cysteine Covalent Library (Weizmann Institute)	Fc1ccc(cc1)N(CC3(=O)c4ccccc4C3(C)=O)CCl	Fc1ccc(cc1)N(C3(*)=O)C=C2)C(=O)C	PCM-0102868	B - active - covalent	4 - High Confidence	1.8	0.36	6.33	Deposited	5RFN
Mpro-x0731	Cysteine Covalent Library (Weizmann Institute)	Cc1ccc(cc1)S(=O)(=O)N2CCN(CC2)C(=O)CCl	CC(=O)N1CCN(CC1)S(=O)(=O)c1ccc(C)cc1	PCM-0102759	B - active - covalent	2 - Correct ligand, weak density	1.86	0.56	7.99	Deposited	5RGN
Mpro-x1386	Cysteine Covalent Library (Weizmann Institute)	ClCC(=O)N1CCN(Cc2ccsc2)CC1	CC(=O)N1CCN(Cc2ccsc2)CC1	PCM-0102739	B - active - covalent	4 - High Confidence	1.7	0.36	5.92	Deposited	5RFS
Mpro-x1308	Cysteine Covalent Library (Weizmann Institute)	ClCC(=O)N1CCN(CC1)S(=O)(=O)c2ccc(Cl)cc2	CC(=O)N1CCN(CC1)S(=O)(=O)c2ccc(Cl)cc2	PCM-0102704	B - active - covalent	4 - High Confidence	1.78	0.38	6.28	Deposited	5RFF
Mpro-x0771	Cysteine Covalent Library (Weizmann Institute)	Cc1ccc(cc1Cn1)S(=O)(=O)N2CCN(CC2)C(=O)CCl	CC(=O)N1CCN(CC1)S(=O)(=O)c1ccc(C)cc1C	PCM-0102628	B - active - covalent	2 - Correct ligand, weak density	2.07	0.58	6.11	Deposited	5RGP
Mpro-x0759	Cysteine Covalent Library (Weizmann Institute)	Fc1ccc(cc1)C2CN(CCO2)C(=O)CCl	Fc1ccc(cc1)C2CN(CCO2)C(=O)C	PCM-0102615	B - active - covalent	4 - High Confidence	1.88	0.4	5.59	Deposited	5RER
Mpro-x0752	Cysteine Covalent Library (Weizmann Institute)	ClCC(=O)NCc1ccc2OCOc2c1	CC(=O)NCc1ccc2OCOc2c1	PCM-0102578	B - active - covalent	2 - Correct ligand, weak density	1.88	0.37	4.37	Deposited	5REO
Mpro-x1351	Cysteine Covalent Library (Weizmann Institute)	ClCC(=O)N1CCC(CC1)NC(=O)c2ccccc2	CC(=O)N1CCC(CC1)NC(=O)c2ccccc2	PCM-0102575	B - active - covalent	2 - Correct ligand, weak density	1.75	0.35	4.76	Deposited	5RFK
Mpro-x1374	Cysteine Covalent Library (Weizmann Institute)	Cc1ccc(cc1)N(CC3(=O)c4ccccc4C3(C)=O)CCl	Cc1ccc(cc1)N(C3(*)=O)c4ccccc4C3(C)=O	PCM-0102539	B - active - covalent	2 - Correct ligand, weak density	2.06	0.24	6.57	Deposited	5RFM

Figure 3: Screenshot of potential compounds) [3]

The process oftentimes begins with the identification of a *lead drug*. This is a drug that scientists believe will be effective, in our case as an ACE2 inhibitor. This is done using a variety of studies and experiments.

Typically, however, a lead drug will have a variety of problems that make it unsuitable.For example, it may not dissolve easily in the bloodstream, making it difficult to get to the host cells in the lungs. It may not have the characteristics, such as a lipophilicity (logP) value of less than 5, generally considered to be the value needed for the drug to be an oral drug, taken

by mouth.

One of the compounds identified by the Diamond Light "XChem" group is TRY-UNI-714a760b-6, shown in Figure 4. This is a new compound that has been shown to have potential effects as a CoVID-19 drug. The drug contains two benzene (phenyl) rings, two amine groups, a carbonyl (C=O), and a halogen (chlorine).

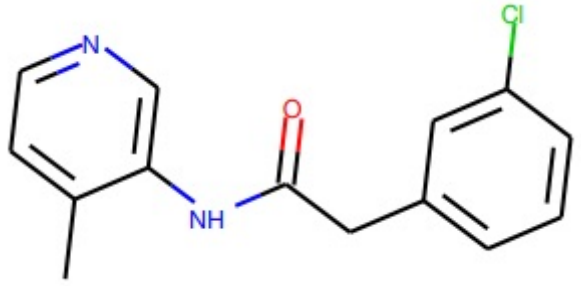


Figure 4: Graphic of a current lead drug (TRY-UNI-714a760b-6) [2]

The goal of modifying a lead drug is to find compounds that have most if not all of the structure of the lead drug, but have added chunks of organic "stuff" that will change one or more properties of the drug. For example, we might need the lead drug to be more soluble, so we will experiment with adding different organic "fragments" to specific locations on the lead drug. These "substituents are also called "R-groups", and are mostly carbon compounds. Almost all drug discovery software tools, including StarDrop, come with a built-in fragment library.

There are two fundamental ways to add substituents to a lead drug. The first is to use an existing "library" of fragments. Many drug companies, such as GlaxoSmithKline, develop these libraries, and they are typically closely guarded (proprietary) secrets. As a side note, one of the more interesting phenomenon of the search for a CoVID-19 vaccine or drug is the degree to which very competitive drug companies have been willing to share data. Imagine your soccer team, during the final championship game of the year, is willing to let the other team have your best player, and they let you have theirs. That's the level of sharing and cooperation that is being seen in the search for a drug or vaccine solution to this pandemic!

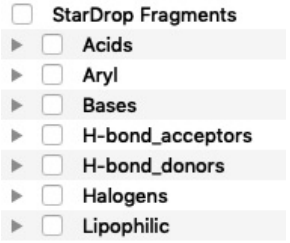


Figure 5: StarDrop Fragment Library

The second way, chemical transformations, uses fundamental principals from chemistry, mostly organic chemistry, to modify the lead drug. For example, we can say "convert one carbon in a benzene ring to an amine (NH_2)". The software then looks for places in the lead drug where it can apply that rule. In the next lab, you'll use chemical transformations to create hundreds of new compounds!

2

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!

For this lab, you will modify the lead drug shown in Figure 4. Specifically, we are going to add four substituents:

1. methyl group (CH_3-)
2. ethyl group (CH_3CH_2-)
3. propyl group ($CH_3CH_2CH_2-$)
4. butyl group ($CH_3CH_2CH_2CH_2-$)

These four substituents will be added to the carbon in the *para* position on the chlorobenzene group. This is the carbon directly opposite of the chlorine on the benzene ring on the right side of the lead drug. This is also known as the "4-position", with chlorine being on Carbon 1, then our substituents will go in Carbon 4.

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!

Now, following the example shown in the webinar, we want to establish the baseline ADME parameters for the lead drug. Using the Models tab, calculate these ADME parameters for the lead drug:

1. **logP**: this is a measure of lipophilicity, how well the drug will diffuse through lipid cell membranes, such as the phospholipid membrane found in most cells. A lower logP is generally better.

2. **logS**: this is a measure of how soluble (in water) the drug is. A lower logS is generally better.

3. **rotatable bonds:** a measure of how flexible the drug is. Bonds that are rotatable mean the drug can change shape, or *conformers*, when binding to a protein. Generally, fewer rotatable bonds is better, but not always!

4. **molecular weight (MW)**: this is in grams/mol, and a measure of how much the drug weighs. Drugs really need to be under 500 g/mol for them to be suitable for oral administration.

5. **2C9 pKi**: (this is a measure of how well the drug binds to the CYP2C9 protein enzyme, a good indicator of how well this drug will interact with a target protein).

2.2 Part 2 – R-group substitution

The next step is to use StarDrop and its "Nova" tool and its Library Enumeration to add the four substituents (methyl, ethyl, propyl, and butyl groups) to the 4-position on the chlorobenzene portion of the lead drug. You will need to follow the example done in the webinar for all of the steps, but when done, you should have a new library of lead drug-substituted compounds.

2.3 Part 3 – Merging of lead drug with substituted compounds

In this step, you need to do the following, based on the example from the webinar:

1. using Data Set –> Merge, you want to merge the new library dataset with the original lead drug data set.

2. rename the merged data set "Screening Compounds".

3. delete the original lead drug data set and the library data set, leaving one data set – Screening Compounds.

2.4 Part 4 – Analysis

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

1. Histogram: choose the Histogram option
   (a) plot logS (solubility) on the x-axis and logP on the y. If I need compounds with a low logP value, what should my logS value be – low, medium,or high?
   (b) Based on your answer to the previous question, how do I make that happen? Experiment with the molecular weight (MW) on the x-axis and logS on the y-axis. What size (low, medium, high) molecular weight compound do I need to get the logS (low, medium, high) value?
2. Plotting: select 2D Scatter from the Chart menu.
   (a) Plot logP vs. 2C9 pKi, then, from the More options icon (last one on the far right), choose "Customize" then "Lines" and then click on "Regression". The R^2 value tells you how well logP will be able predict the pKi value. The closer to 1 or -1 the value is, the better. What is the R^2 value for logP?
   (b) What about logS? Notice that the slope here is negative.
   (c) What about molecular weight? Is that a good predictor of pKi?
3. Function Generating: we want to calculate what is known as a *Hammett constant* for our library of R-substituted lead drugs. We add four R-groups – methyl, ethyl, propyl and butyl. By how much did the methyl group change the logP value? Did it go up or go down? The logP for the lead drug is approximately 3.028. The compound with a methyl group on it is 3.351, higher. How much higher? A simple subtraction provides the Hammett constant for methyl, represented by a sigma (σ) symbol:

$$\sigma = logP_{substituted} - logP_{leaddrug} \tag{1}$$

   (a) In StarDrop, go to Tools –> Function Editor
   (b) Double click on the logP value for the lead drug. You should see the value to four decimal places. Copy that value.
   (c) Make the new column name Hammett constant.

(d) Create the function: logP - 3.0277 (or the value you copied in Step 2).

(e) You should now have a new column called "Hammett constant". Control- or right-click onto the column heading and select Edit. Change the format to Decimal with three decimal places. The value of the constant for the lead drug should be zero.

(f) Using a plotting tool of your choice, what happens to the binding affinity (2C9 pKi) as the constant gets bigger?

References

[1] https://www.researchgate.net/profile/Niroshini_Nirmalan/publication/308045230/figure/fig1/AS:613907634872358@1523378394724/Drug-discovery-and-development-timeline-The-current-drug-approval-pipeline-can-png, accessed May 20,2020

[2] https://postera.ai/covid/submissions/714a760b-0e02-4b09-8736-f27f854f8c22, accessed May 20. 2020

[3] https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem/Downloads.html, accessed May 20,2020.

[4] https://www.clinicaltrialsarena.com/comment/dexamethasone-covid-19/, accessed July 3, 2020