Bioactive molecules - Wet lab practice.

David Luna, Alejandro Sáinz & Raquel Ventura

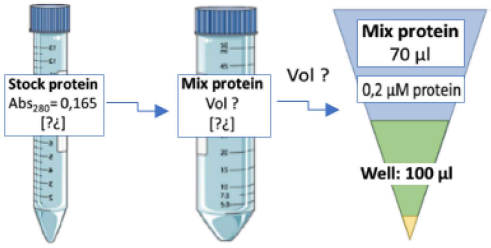
1 Previous calculations for the screening procedure.

In this experiment we are going to desin a small molecules screening. Our target molecule is a protein from SARS-COV 2 called 3CLPro. This protease is involved in viral polyprotein cleavage. Our aim is to find compounds which inhibit the process. The protein has been already purified, so we must check its crystalographic structure. After the structure is defined, is good to prove 3CLPro behaviour in solution. In order to see if it's folded, we characterize the protein, measuring flourescence and CD spectra.

The method we use to determine potential inhibitors is a small screening of a known compound library, proving directly the activity in a 96 wells plate. In the lateral 1 and 12 columns we put the control solution. The quantities needed of each starting solution are the following ones[1]:

1.1 Protein.

Based on the data shown on the following picture, we can easily calculate the missing quantities.



$$[\text{Stockprotein}] \rightarrow \epsilon \cdot l \cdot [\text{S.Prot.}] = \text{Abs} \rightarrow$$

$$0.165 = 32890\ \text{M}^{-1}\text{cm}^{-1} \cdot 0.05\ \text{cm} \cdot [\text{S.Prot.}]$$

$$[\textbf{Stock protein}] = 1.033 \cdot 10^{-4}\ \text{M} = 0.103\ \text{mM}$$

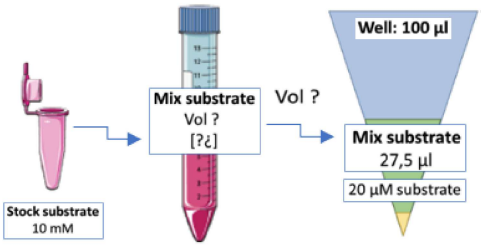
$$\textbf{Vol}_1 = 70\ \mu\text{L} \cdot 96\ \text{wells} \cdot 1.1 = 7392\ \mu\text{L} = 7.39\ \text{ml}$$

$$[\textbf{Mix protein}] = 0.2\ \mu\text{L} \cdot \left(\frac{100\ \mu\text{L}}{70\ \mu\text{L}}\right) = 0.286\ \mu\text{M}$$

[1] All quantities are going to be computed with an excess of 10%. Those factors are written in red in the following calculations.

$$\mathbf{Vol}_2 = \frac{0.286\ \mu\text{M}\ \cdot\ 7.4\ \text{ml}}{0.1033\ \text{mM}} = 20\ \mu\text{L of protein mix}\ ;\ 7.37\ \text{ml of buffer}$$

1.2 Substrate.



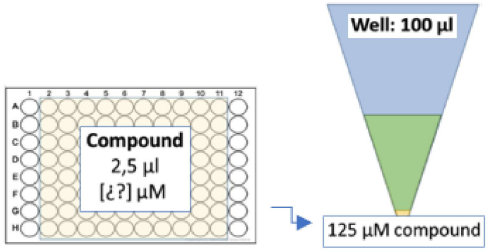
$$[\text{Stockprotein}] = 10\ \text{mM}$$

$$\mathbf{Vol}_1 = 27.5\ \mu\text{L} \cdot 96\ \text{wells} \cdot 1.1 = 2904\ \mu\text{L} = 2.94\ \text{ml}$$

$$[\mathbf{Mix\ substrate}] = 20\ \mu\text{M} \cdot \left(\frac{100\mu\text{L}}{27.5\ \mu\text{L}}\right)$$

$$\mathbf{Vol}_2 = \frac{72.7\ \mu\text{M}\ \cdot\ 2.94\ \text{ml}}{10\ \text{mM}} = 21\ \mu\text{L of substrate mix}\ ;\ 7.37\ \text{ml of buffer}$$

1.3 Compound.



$$[\text{Compound}] = \frac{125\ \text{mM}\ \cdot\ 100\ \text{ml}}{2.5\ \mu\text{L}} = 5000\ \mu\text{M}$$

2 Final data analysis.

We have now the data corresponding to the fluorescence screening done in the lab. Our set of data has two easily distinguishable types of data: the control group, that doesn't inhibit the protease, and our ligands.

Our aim is to select the best 10 ligands, in terms of inhibition potential. First we will explain our reasoning, and then we will present our results.

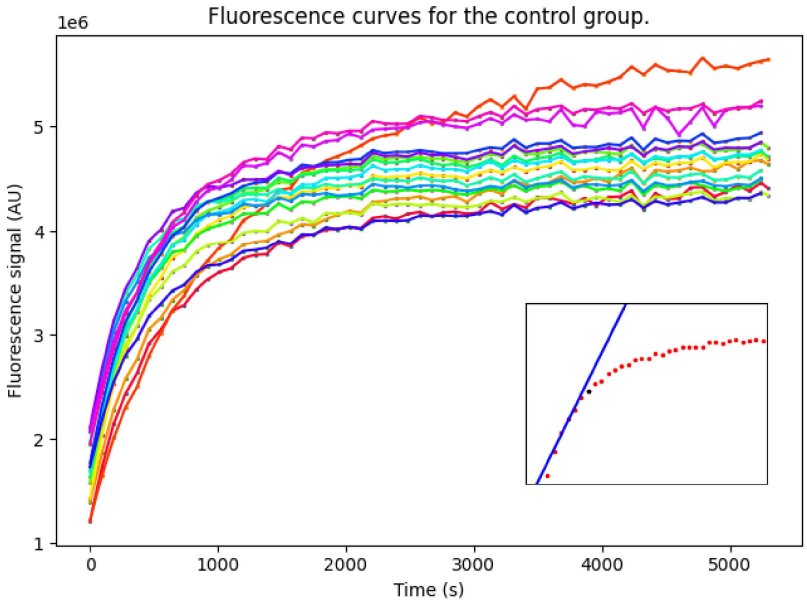
To find the inhibition potential of a compound in a fluorescence experiment, we have to compute the initial slope of those curves respect to time. In our case, we considered that the initial slope was determined by 5 points, from the second to the sixth:

- We did not consider the first one because it carries a lot of experimental error, so it's not as reliable as the others. One can clearly see that all the curves differ a lot around that point.

- The reason to stop at the sixth point is simple. Just from looking at the graph, it can be easily seen that the seventh point deviates from a straight line in a way much more noticeable than the previous ones, which shows the change in behaviour of the curve (not a line anymore). Then, we will exclude all points forward, because we think that they will underestimate our initial slope, and should not be part of that calculation.

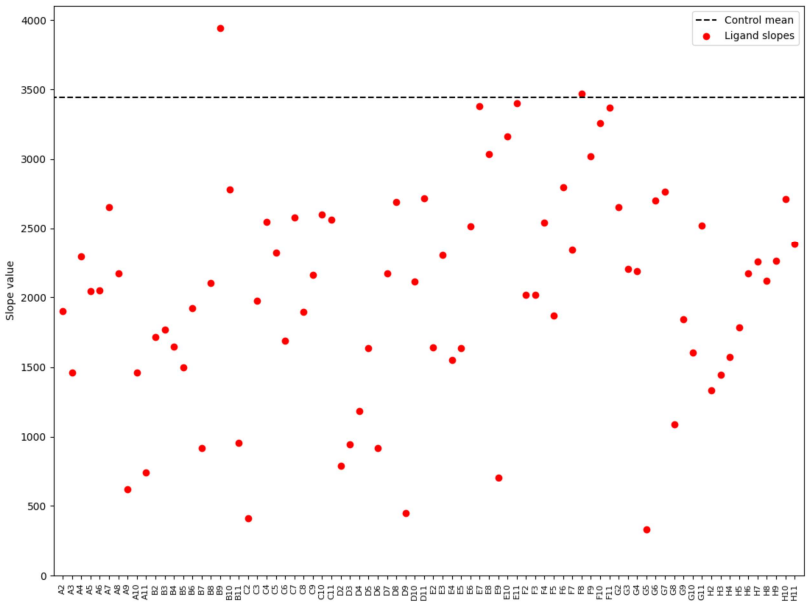
  To make proof of our claim, see the embedded box in the first graph. The seventh dot is marked in black, and we see that the tendency of the curve starts to change from a line around there.

In order to visually show what we are doing, we include the next graph, which is the fluorescence curves for the control group:



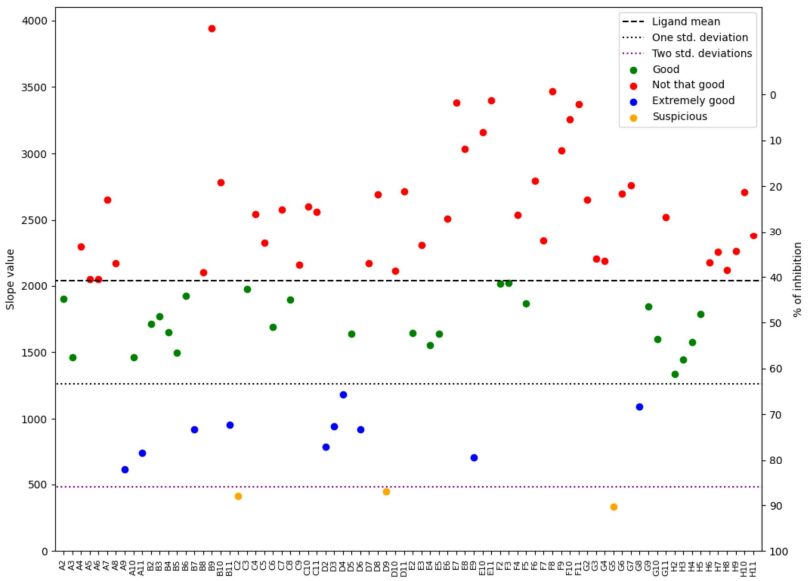
We see that the first point changes a lot between different controls; that is the reason for its lack of reliability.

Once the slopes have been calculated, we have to classify them in some way. Let's plot the mean slope for the control group and the slopes of the ligands, just to see if they inhibit the protein or do nothing at all.



They are (except one) all below the control mean slope, so we now know they all inhibit the protease in some form. However, how do we select the best 10 compounds from the bunch?

We designed a procedure to identify those ligands. First, we calculate the mean slope and its standard deviation. Then, we plot the slopes of the ligands along with the mean, the mean minus one standard deviation, and the mean minus two standard deviations. The plot in question:



Then, depending on where the points lie in the graph, we classify them as:

- *Not that good*: if they fall above the mean slope it means that, although they are able to inhibit the protease, their inhibition potential is not as good as the majority of the ligands. We will not be considering them as part of our possible choices.

- *Good*: this tier is where compounds below the mean and within 1 standard deviation fall They will work as inhibitors, but they are far from extraordinaire.

- *Extremely good*: this ligands exhibit an incredible inhibition capacity upon the protease, and will be the compound pool from where we will be choosing our top 10 selection. They are within 1 and 2 standard deviations of the mean slope.

- *Suspicious*: these three compound have a slope so low that it begs the question: Is this due to their raw inhibition capacity, or caused by some error in the experimental procedure? We know that sometimes, due to the coloring of the solution, fluorescence essays do not work well. Because we do not know the answer to this question, we will discard them as candidates.

Now, based on only this data we can not make a clear choice. The appropriate procedure would be to repeat the screening for the so called *suspicious* points with the correct coloration to avoid the same potential mistake.

Depending on the results, our top 10 choice will have some compounds from the *extremely good* bunch and, potentially (depending on whether their slope values are correct or not), from the *suspicious* group.