Q&A
27 October 2021
Statistics Are a Girl's best Friend: Expanding the mechanistic Study Toolbox with Data Science
Dr Anat Milo (Ben-Gurion University of the Negev)

Q1: So how little data do you think we can get away with? But the little data side against them?

It depends on a few things. So, rule of thumb I like to go with 15 data points because one of the rules of thumb is five data points per parameter or else. But again, that's not great either because it's a lot of parameters per data set. But if you have a simple question where your mechanism is fairly simple and it's based on the components of the reaction, then you don't need many data points to look at it. I think that one of the things that is at the core of this is that we design our datasets in advance to answer some questions. So, if we take aldehydes, we'd put something electron donating and withdrawing at the two position, then the same at the three position, and then at the four position, and then we put something big at the two, three, four position, and so we're kind of probing things. If the mechanism is fairly complex, we'd need more data points to get a reliable correlation that validates, so it really depends on the system. You can go very small if you're not describing things that aren't very complex then the less data you have, the less parameters you're allowed to use statistically.

Q2: Looking or your lab it looks to me like you've got some flow control systems and stuff there to actually make sure that the experiments are probably more reproducible then doing them all individually by hand. What the interplay there between actually doing experiments and repeating them, and that sort of variability that clearly always crept in when I was trying to do experiments in the lab on that?

I think one of the key points is that when you're looking at data that you want to analyse statistically, you want to know your error. So, it's really important to be able to reproduce the experiment and then to look at the error because it could be that you're producing a model that the error is small compared to the experimental error or vice versa. We always want to know that your experimental error is smaller than which you can get with the with the model. Every reaction that we run is at least in duplicate, if not more. I have here a picture, this is our bigger lab room. We have a smaller one which is temperature and humidity controlled. I had some talks with the people who are in charge of the infrastructure at the university, so we are now in new labs and we designed it. But before it was pretty difficult, so I used to have in my last slide our temperature and humidity controller because that is critical. Again, that's why high throughput or automation are really important, because if you can, you know take out that variable, our human contribution becomes a more interesting when we're not just trying to get things to work the same way

Q3: When you're only doing a relatively small number of reactions it's at least conceivable to do them, even that it's a huge amount of effort. But being able to do things high throughput on small scale, so what with some of these reactions I'm never quite sure how smaller scale could one realistically do them and then be representative of doing them on a moderate scale. Or maybe that doesn't need to be, but you know the less you have to do, the more things you can make I guess is the way I look at it. Is that realistic or does one really need to still handle this on a normal lab scale?

It really depends that I know that people have done things on a nano-mole scales, you can do this manually. You can take a 96 well plate or an even bigger one and load it yourself and you can get results. I think that normally they at least correspond with what you'd get on a larger scale. They're telling you something that when you move to the largest scale, you probably have to tweak a little bit to optimise it, but it will definitely tell you something about the components and how they work together. But in in our lab I think that the smallest scales we do are in GC vials, we haven't got smaller than that, but that's pretty small.

It's really nice to see a group where the experiments and the modelling are being done by the same group of people, so that the models really have feel for the experiments and the experiment (maybe it's the same people doing them). It's really important to have a feel for both sides in my view so that you know what to expect and what to what you need to do, and I think it makes a huge difference to the reliability of these things. Thank you.

I truly believe that, and I keep telling the students that they have to go into the lab and do the models as well. The best thing is knowing all the aspects of what you're doing.