Menghang Xia, Ph.D. and Ruili Huang, Ph.D.

Behind the Mask

November 14, 2022

Barr: Good morning. Today is November 14, 2022. My name is Gabrielle Barr, and I'm the Archivist with the Office of NIH History and Stetten Museum. Today, I have the pleasure of speaking with Dr. Menghang Xia, the Leader of Systems Toxicology Activities for the Toxicology in the 21st Century program, which is part of the Division of Preclinical Innovation in the Chemical Genomic Branch of NCATS [National Center for Advancing Translational Sciences], and her colleague, Dr. Ruili Huang, who is a Senior Scientist and Informatics Leader in the Toxicology in the 21st Century program, and they are going to be speaking about some of their COVID-19 research and experiences. Thank you both for being with me.

Huang: Well, thank you for inviting us.

Barr: Absolutely. To begin, will you please define what biological activity-based modeling is and what the advantages are of doing this type of modeling?

Huang: I don't know if people are familiar with predictive modeling. Traditionally people just use chemical structures to try to predict the biological activity for small molecule compounds against a biological target or disease, such as COVID-19 or another viral disease, or some other diseases such as cancer—or anything. Traditionally people use chemical structures to make a prediction. They know the activities of some compounds with known structure, and then when you come up with a new chemical with the new structure, you look at the structure of the new chemical to see if it looks similar to the existing chemical. Then you make a prediction of whether this new compound will be active on this new target or not. This is a traditional method called "structure-based modeling." People are probably more familiar with the term "QSAR"—quantitative structure-activity relationship. That's more the traditional part of predictive modeling.

Then there's this biological activity-based modeling. We don't rely on the chemical structures to make a prediction, but we will use the biological activity profiles of those compounds. At NCATS, we do a lot of high-throughput screening, so we generate a lot of biological activity data and activity profiles on those compounds. Then we can use this data to make predictions about the activities of those compounds against new targets or new diseases. We apply this new approach to try to find potential new drug candidates to treat SARS-CoV-2 or COVID-19. That's basically the concept behind the biology activity-based modeling. Basically, you apply a computational model based on the biological activity profiles of compounds and then make a prediction on their activity against, for example, COVID-19. The advantage of this approach compared to the traditional structure-based approach is that using this approach, you basically don't have to know the chemical structure. You can make a prediction on anything, so you can profile the activity of a compound or a substance—or anything—against a panel of assays and then just use this biological activity profile signature to predict this new activity against a new target.

Barr: Is that faster? Or is it difficult to learn the structure?

Huang: Sometimes. Not all substances have a defined chemical structure. That's the difficulty. For example, herbal medications are like a mixture. For a lot of natural products, we don't know the exact structure of the

chemicals, or we don't even know which chemicals are in there. Then we cannot use the traditional structure-based modeling, but we can apply this biological activity-based modeling, and then we can still make a prediction on those chemicals or substances. That's a major advantage. You can use this approach to actually discover new chemical structures or new scaffolds. That's the key structure within the chemical; we call it a "scaffold." That part is important for its biological activity. If you use the traditional structure similarity-based approach, you cannot discover any new structures. You can only find the chemicals with the structures similar to existing chemicals. But if we use this biological activity-based approach, then you can discover completely new structures, or new scaffolds, and then you can build on this new scaffold to discover completely new drugs. Maybe you can avoid some kind of toxicity with this new structure and improve activity. That's also another advantage of this new biological activity-based approach.

Barr: That's really interesting. Can you speak about how you and others have used the biological activity-based modeling to predict 311 potential compounds to be used against severe COVID-19?

Huang: We have this collection called the "NCATS Pharmaceutical Collection." It's a collection of known approved drugs or investigational drugs. We have profiled those compounds against a panel of assays at NCATS. We generated biological activity profiles on those compounds. Based on those biological activity profiles, we then build a model just to find out which profiles are important to treat SARS-CoV-2 or COVID-19. We'll apply those models by basically screening the entire NCATS library of about half a million compounds. Then we make a prediction of which compounds are going to be active against COVID-19. We actually selected 311 compounds that are likely to be active against COVID-19. Then we ran the experimental validation of those compounds—so we basically tested those compounds in the real SARS-CoV-2 assay to see if those compounds are active. That's how we discovered one third of those compounds were actually found to be active in this assay. That's pretty good accuracy compared to other traditional modeling. From this project, we discovered about 100 potential new drug candidates for COVID-19, so we can prioritize those compounds for further testing or investigation.

Barr: Are there certain characteristics of the ones that you found that could be potential candidates? When you did the prediction in the biological activity modeling, what were your findings in terms of the types of compounds that worked against SARS-CoV-2 and why those kinds of compounds work well?

Huang: Because we use this biological activity profile-based approach, we can compare the biological activity profile similarity between the new drugs and the existing drugs. Basically, if the hypothesis is that two compounds share similar activity profiles, then they share similar targets or modes of action—so that's how it works. They can target certain targets in the virus, so that's why they can work, or potentially work, to inhibit the virus or COVID-19. That's what we found—they have certain characteristics that can target or inhibit certain targets in this virus.

Barr: Can you please speak a little bit about why therapeutics that target the AP-1 [activator protein 1] signaling pathway have a positive effect against COVID-19?

Huang: This has also been actually from a data mining approach. I talked about the biological activity profiles. We have tested a lot of compounds in the SARS-CoV-2 assay, so we have this kind of signature for just SARS-CoV-2. We basically compared the signatures from other assays to see which ones are similar to the SARS-CoV-2 assay. That's how we found out that the AP-1 signaling pathway is highly correlated with the signature of the SARS-CoV-2 assay. That's why we thought maybe AP-1 is also an important target that we can use to treat SARS-CoV-2 or COVID-19. This is more like a hypothesis. We haven't really applied this or designed the experiments to test this. We found some new compounds that can interfere with this AP-1 signaling. We also found out those

compounds can inhibit SARS-CoV-2. We're not sure yet if those compounds actually go through the AP-1 signaling or if it's kind of a side pathway.

Xia: It has been reported that, during the viral infection, the AP-1 signal pathway was activated to upregulate the gene expression of the cytokine, which contributes to the inflammation. The AP-1 may play some role against the viral infection. That is the biology behind it.

Huang: That kind of serves as a validation of our approach. That still needs further validation experiments to further confirm, but right now it's more like a hypothesis.

Barr: Will you speak about your role in analyzing NCATS' Pharmaceutical Collection of approved drug screens and comparing their activity profiles with the results of a severe acute SARS-CoV-2 cytopathic effect assay?

Huang: Yeah, so that's actually what I just talked about. We compare the drug activity profiles generated on this NCATS Pharmaceutical Collection, and we screened that collection against a lot of different assay targets.

Barr: You also looked at the toxicity, right?

Huang: Yeah, we also looked at the toxicity. We noticed the hERG [human ether-a-go-go related gene] assay, and a lot of those compounds active in the SARS-CoV-2 assays, were also active in the human hERG assay. That can potentially result in cardiotoxicity. In addition to AP-1, we also found that hERG is another potentially toxic target for those compounds that can inhibit SARS-CoV-2. That's another finding from this study.

Barr: Can you talk a little bit more about your findings around the hERG gene?

Huang: We basically found a lot of the anti-SARS-CoV-2 compounds active in the human hERG, and that may be resulting in cardiotoxicity. In that sense, people can use this information to prioritize the chemicals that they want to advance for further testing of SARS-CoV-2—so they probably want to select compounds that don't inhibit hERG. We do see some compounds that inhibit hERG, and then they actually show certain structural characteristics. When people select those compounds, they should try to avoid those structural characteristics to avoid the hERG-inhibition part of their activity, and then select compounds that can inhibit SARS-CoV-2 without inhibiting hERG to try to minimize or avoid cardiotoxicity.

Barr: Will you please introduce the Tox21 program and its mission?

Xia: The Tox21 program is a multi-government agency collaborative effort among the EPA [U.S. Environmental Protection Agency], NTP [National Toxicology Program], FDA [U.S. Food and Drug Administration], and NCATS. Why do we run this program? Because currently there are over 80,000 chemical compounds in use in the United States. Every year about 2,000 manmade chemicals are introduced into our environment. The problem is that most of these chemicals have not been tested for their toxicity on human health. The traditional methods to test compound toxicity are using animal models, which are not only expensive and low throughput, but also sometimes fail to predict toxicity in humans. That is the reason the U.S. government formed the Tox21 program—in order to reduce animal use in toxicological studies and to advance in vitro toxicological testing in the 21st century. The mission of this program is to shift future toxicity testing from being focused on the traditional animal studies to less expensive and high-throughput in vitro methods. In order to achieve this mission, we use a battery of in vitro assays to screen hundreds of thousands of environmental chemical compounds. The data generated from this primary screening can be used to identify the mechanisms of

compound action and to prioritize the compounds for further in-depth study. We can also use this screening data to develop computational models to predict the effects of the unknown compounds. As Ruili mentioned that they use the computational models by using our data to predict the effects of the unknown compounds. Since we started the Tox21 program in 2011, we have developed and validated over 100 in vitro assays, which cover over 60 toxicologically relevant targets or signaling pathways in a high-throughput screening platform. We use this in vitro assay to screen the Tox21 10,000 compounds. So far, we have generated and uploaded more than 200 screening data sets containing over 100 million data points to public databases such as PubChem.

Barr: That's a lot. Do you know in what ways the data that you've made public has been used by different audiences?

Xia: Since we publish all these data in PubChem, a lot of people actually use our data. We also launched the data competition to use our data to build up the prediction model to predict the (effects of) compounds. A lot of computational scientists used our data sets.

Huang: People can use that data as training datasets to build prediction models for new compounds, to predict the toxic potential toxicity or toxicity targets for new compounds against other targets. The Tox21 Challenge was in 2014. We asked people to use the Tox21 data to build models for twelve different toxicity targets. That turned out pretty well. Most of the models achieved accuracies above 80% or 90%. We have this public website that posts the Tox21 data, and we see a lot of traffic on this website. Currently, we have thousands of users—people who downloaded data from our website to build models. Some people can also use this data to select potentially toxic compounds and follow-up on those compounds and study their toxicity mechanisms.

Barr: Will you discuss in depth your efforts to examine the toxicity of small molecule drugs specifically for COVID-19 using the Tox21 database, and what criteria you based your assessment on?

Xia: During this pandemic, we know many research groups tried to develop new drugs or repurpose older drugs for COVID-19 treatment. For example, remdesivir has been approved to treat COVID-19, and several other drugs have been approved for emergency use in the clinic, such as Paxlovid. Many others are still in the development stage. Many drug repurposing candidates with antiviral properties sometimes may have toxicological effects via some unknown mechanisms, which motivated us to check our previously generated Tox21 screening data sets to understand more about these compounds and their toxicological effects. The Tox21 10,000 "10K" Compound Collection actually includes the approved drugs and environmental chemical compounds. We also found that this library actually contains more than 20 of the COVID-19 drug candidates. As I mentioned before, this big data set was generated by screening the Tox21 10K Compound Collection against over 100 in intro assays that cover many cellular targets and pathways. For example, we found some of the drugs that were active in our nuclear receptor assays, such as androgen, estrogen, and pregnane X receptors. We also found several of the drugs modulate several signaling transduction pathways—for example, Nrf2/ARE and TGF beta signaling pathways, as well as the endoplasmic reticulum stress pathway. Our screening data sets also include cell viability, cytotoxic data in more than 10 cell types. We found that some drugs have a cytotoxic effect against a wide range of cell types. I feel that the Tox21 screening data is powerful because we can use the screening data to evaluate the COVID-19 drug candidates for their potential toxic effects and related pathways which provide valuable information for the drug development. Another benefit is that all these Tox21 screening data are publicly available, so anybody can access these data sets and perform their own data analysis based on their needs.

Barr: Are you aware of how researchers have used this data to further their own COVID-19 research, or how its advanced drugs being put on the market?

Xia: How to use our data sets? Ruili leads a couple of studies which actually used those type of data.

Huang: [inaudible.] We found a lot of those drugs showed activity in the cell viability assays. That means they're just basically cytotoxic. They can kill cells. That's another toxic effect. I don't know if any people outside of our group have [used it].

Xia: I'm not aware of that. Sometimes we need to do a Google search to find out. If people use our data, they will not tell us. It's publicly accessible (data).

Huang: COVID-19 is still relatively new—so they probably use the data, but maybe the paper hasn't come out, so we don't know.

Xia: Probably in the coming years, we will see those kinds of papers using our data.

Barr: That will be really interesting to note, I'm sure, for you guys. Have you all been involved in other COVID-19 research or do you have plans to further any of the studies or efforts that you've been part of?

Xia: Our main focus is on the Tox21 program, so the COVID-19 project actually is our side product. Sometimes the COVID-19 research group contacts us to get our help. But for us, we don't have a main project that is COVID-19 related.

Huang: I have the informatics group, so we have computational people. We're still continuing our effort of trying to optimize our models or build new models that can discover new potential anti-SARS-CoV-2 compounds. That's ongoing, so we're trying to develop new computational algorithms to improve the model efficiency. During the pandemic some of the other work had to slow down. That actually gave us more time to focus on data analysis and model building—even writing more manuscripts and publications. For me, part of my time is still spent on COVID-19 projects.

Barr: Can you speak a little bit about how you've applied your training and your expertise to COVID-19 and how some of your other work has informed how you approach the pandemic?

Xia: I have a Ph.D. degree in pharmacology from the State University of New York at Buffalo, and then I went to the University of California San Francisco for my postdoc trainee in molecular and cell biology. After that, I worked in two pharmaceutical companies, Amgen and Merck Research Lab, to identify and validate drug targets for drug development. In 2005, I joined the NIH Chemical Genomics Center, which is now NCATS. For Tox21 program, I am leading the major effort to develop and validate a group of the in vitro assays in a quantitative high-throughput screening platform. I have led my team to screening more than 100 in vitro assays that involve many targets and the pathways. I have the expertise in assay development, assay validation and compound screening, so I can use my knowledge and skill sets to evaluate some of the COVID-19 drug candidates.

Huang: For me, I actually have my original training in cancer. I have my Ph.D. in chemistry so I have knowledge in chemistry, and then I moved into computational biology, chemoinformatics, bioinformatics, data mining, and data analysis. I have a combined background in both chemistry and some biology. I learned a lot from my colleagues, like Menghang, who is a real good biologist. I have gained a lot of computational modeling skills. I have been building models for toxicity or trying to identify potential new candidates for different disease targets. I previously worked on anti-malarial drugs, building models that identify new anti-malarial drugs—and

also for anti-Zika and anti-Ebola compounds. Actually, we would use those models for those diseases to validate this new biological activity-based model. Then, when COVID-19 came, we just applied this new modeling approach as a new concept to identify new drugs for COVID-19. My background knowledge has well prepared me to help with combating this new pandemic.

Barr: In addition to being scientists, you are also people who've been living through the pandemic the past two and a half years. Can you discuss some of the opportunities and challenges that the pandemic has presented to you as individuals?

Xia: We had a big challenge during the pandemic because I'm the biologist, pharmacologist, and toxicologist. We always work in the lab, right? In March 2020, we started to telework at home full time. Then it changed to mostly teleworking and a little time working in the lab, with 30% capacity. Therefore, the major negative impact has been on our lab work. Also, the Tox21 screening is a team effort and needs multidisciplinary teams working together. During the pandemic, we just could not perform the online screening due to the limited lab capacity. Also, when we perform the online screening, we need 100% lab capacity. When we run the screening, we are nonstop (working) for four to six months, including some weekends. We just could not perform the big screening. As for some opportunities, during this pandemic, actually, we have had more time than before where we can summarize our previous research work and write scientific manuscripts. In the past two and a half years, I can't believe that we have published more than 40 scientific manuscripts and eight to ten book chapters, which is the most productive time we have ever seen. The flexibility we had during the pandemic actually made our scientists work more efficiently. Also, some of the young scientists in my group have young kids, so they can do the work, write the manuscript, and also be taking care of their kids. That is the good part.

Huang: For me, maybe there's more opportunities than challenges. The challenge for Menghang is an opportunity for me because I used to have to analyze all the data generated from Menghang's screens and assays. When her lab work slowed down, we had more time to focus on other areas like COVID-19. In the last couple of years, we spent most our time on COVID-19 projects, instead of traditionally focusing on just one project. But now we have more time to work on COVID-19 projects and develop those new models and new algorithms. Actually, like Menghang, we also had more publications in the past couple of years—like almost 20 papers per year. Also, we have a lot of collaborative papers together with Menghang. One challenge is, because we are computational people, we work on computers all the time. We need big powerful computers, but during the last few years, the informatics group were not allowed onsite at all. We were not allowed to go back to our office. We just had to work from home all the time—on my little laptop. I'd have to remote into my workstation in my office, and that depends on the internet speed. I have a couple of big monitors in my office, but only have access to the small laptop at home. That's a limiting factor. It's just not as efficient as working from my office. But other than that, for me, research-wise, our work has turned out pretty well. We got to contribute to this COVID-19 effort.

Barr: Is there anything else that you want to share about your research or about your experiences?

Xia: That's alright. I am very glad Ruili and I could do the interview together. I'm very pleased.

Huang: Yeah, we are very close collaborators. We work together all the time. I'm happy that, finally, starting in April, we were allowed back in our office. Now I go to work, and I actually go to my office, and I get to work on my big computer.

Xia: Yeah. This [Zoom] background is my office!

Huang: The [heating system] in our building broke. We have no heat in the office, so I'm working from home today.

Barr: Oh, no! Well, thank you both for all your work that you've been doing. I wish you both the best.

Huang: Thank you.

Xia: Thank you so much. Our pleasure.

https://pubmed.ncbi.nlm.nih.gov/33762619/
https://pubmed.ncbi.nlm.nih.gov/33623157/
https://pubmed.ncbi.nlm.nih.gov/35395401/
https://pubmed.ncbi.nlm.nih.gov/36449624/
https://pubmed.ncbi.nlm.nih.gov/33708112/
https://pubmed.ncbi.nlm.nih.gov/33689873/
https://pubmed.ncbi.nlm.nih.gov/34048893/
https://pubmed.ncbi.nlm.nih.gov/35275639/
https://pubmed.ncbi.nlm.nih.gov/36719788/