

Alexey Zakharov

Behind the Mask

November 4, 2022

Barr: Good morning. Today is November 4, 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. Alexey Zakharov. Dr. Zakharov is an informatics leader for an early therapeutic discovery project team in the Division of Preclinical Innovation at NCATS [National Center for Advancing Translational Sciences]. Today he's going to be speaking about his COVID-19 research and experiences. Thank you very much for being with me.

Zakharov: Good morning, Gabrielle. Thank you for having me today.

Barr: Absolutely. To begin, please describe the research you were a part of in the early part of the pandemic that identified via the computer 73 combinations of 32 drugs that could possibly combat SARS-CoV-2 and then tested them in vitro.

Zakharov: Sure. That's a great question. When the pandemic started, NCATS initiated an award to support and promote development of anti-SARS-CoV-2 treatments. Many of my coworkers were working on developing relevant assays for screening compounds. They developed both phenotypic as well as covering different mechanisms of action, like biochemicals. NCATS is utilizing a quantity of high throughput screening platforms. So whatever we screen, we are usually doing it in dose response manner so we can see good curves for each compound. Once assays are developed, the first aim is of course to repurpose the existing drug available on the market. NCATS has a long history of development, and screening of approved drugs. We have one collection called the "NCATS Pharmaceutical Collection" or NPC, which is a publicly accessible collection of drugs that are approved by different countries and amenable to high-throughput screening. NPC has more than 2,800 compounds ready for screening.

Barr: That's a lot!

Zakharov: Yeah, it is. It's really a lot, and it's growing. We very recently extended that—in one of our publications we put it up to 10,000. Back then we were screening to prioritize the 2,800 compounds using the assay we developed. Of course, NCATS is actively supporting open science concepts, so whatever the screen, we decided to put it in an open data portal to share with the community. This open data portal presents data from assays and information about the assay execution protocol, information about targets, mechanisms of action, and so on—all of the information we have for this particular library's screening result. Of course, this repurposing effort helped us reveal and confirm some active compounds—including Remdesivir, which was authorized back then by FDA for clinical use. Despite some progress in drug repurposing, it was very clear that the found drugs cause side effects and also basic resistance due to rapid viral mutation. This was the point where our combination drug repurposing research starts. Drug combination is useful in treating viral infection due to the fact that they can substantially lower the risk of developing resistance. Also, the antiviral action of combination drugs may be stronger than either of the drugs alone. This effect is called synergy.

Primarily we focus on screening compounds in live viruses, which requires a BSL-3 [biosafety level 3] facility. Such assay translates better to the clinic, but it has limited throughput. Basically, it's hard to screen millions of compounds under that condition. Because of that we decided to utilize in silico [computer] methods, such as AI [artificial intelligence], machine learning and knowledge graph approaches, because they can really accelerate the early drug discovery process. We first identify 76 individual drug candidates for repurposing in combination therapy against SARS-CoV-2. If you compare all the possible pairwise combinations across these drugs, you end up with 2,850 unique binary combinations. From this amount of all possible (combinations) we decided we need to prioritize them because it was just too many. We applied the different approaches and tried to use some different mechanisms of action—targeting viruses at different stages of life—to find the best combination of compound. It will probably increase the chance of probability of synergy which we can use for both of them. On top of that, we also try to apply different silicon tools such as the Chemotext. Or, more recently, developed by our collaborator from University of North Carolina's COVID -KOP approach, together with QSAR models, which model major drug interactions. We're all trying to apply that technology to eliminate compounds that might have negative drug interactions—like having exhibited side effects or having been previously tested and the literature suggests the outcome was not great. From this list we prioritized and came up with 73 combinations from the 2,800, which covered 32 drugs combined together. We screened those dose response titrations with a 6x6 matrix—so basically each compound has six dilutions and they kind of create this matrix cube measuring 6x6. From the 73 combos, we identified 17 synergistic and eight antagonistic combinations. Four of this eight actually show both synergistic and antagonistic interaction at the different concentrations.

Barr: Have any of them moved on to being tested further?

Zakharov: Yeah. We tried to organize possible clinical trials for some of them, but at the time, it was hard to find animal models available. That was one of the obstacles we were facing.

Barr: You do a lot on the technological side. Could you talk us through developing the computer algorithms that allow you to look at all these different combinations and situations?

Zakharov: That's always what you want to do—develop some new methods and approaches and apply them. Especially in the pandemic, it's not just about the methods and validations. It's about getting results. We need to incorporate whatever knowledge was existing at that time into the modeling effort and apply as many different models and as much computational technology as possible. If you look at the ligand-based and structure-based, each approach has their own pros and cons. Basically, when they are applied altogether, you look at the problem from the different angles, so the success rate will increase.

Barr: That brings us to your next study. Can you talk about your participation in an effort that proposed an approach for looking at synergistic drug combinations against COVID-19 through machine learning, and what some of the initial complications included?

Zakharov: I want to emphasize the outcomes of the 16 synergistic and antagonistic results. Most notably, you usually look at the synergy when you combine it to show that it's good, but even when you notice two drugs show antagonistic effects, it's also alarming. Of course, it's in vitro, but maybe we don't want it in the clinic. Most notably, what we found among this antagonistic combination is that there is a strong antagonistic effect between remdesivir and anti-malarial drugs—such as hydroxychloroquine, mefloquine, and amodiaquine. That was kind of an interesting finding, especially taking into account

that later on the FDA withdrew this emergency authorization for hydroxychloroquine. The good part of applying our platform is that we actually see this antagonism in vitro. For the synergistic compound, we found that nitazoxanide actually synergizes very good with Remdesivir, Umifenovir and amodiaquine very well. One of the combinations that was really strong was between nitazoxanide and Remdesivir, and we also found what concentration will absorb this synergy in the plasma and lung. We were thinking this combo has interesting potential for the clinical trials. But as I said before, unfortunately we didn't reach the efficacy study with the animal models back then.

Barr: Why did certain combinations work better together than others? What properties made them work better?

Zakharov: That's what we're looking at, to understand what's going on over there. They found nitazoxanide, for example, shows some complex antiviral mechanisms. They go over the different virus life cycles and what boosts up synergy. But again, in reality, we really need to deep dive into the mechanism of action and try to understand what is behind this synergy.

Barr: Can you talk a little bit about looking at synergistic drug combinations through machine learning and what some of the complications were? Discuss the ComboNet architecture that you and your team use.

Zakharov: That's a good point. After successfully discovering synergistic combination against SARS-CoV-2 using these knowledge graph approach, we switched to machine learning or AI methods. Since the nature of these methods are different, the technology is different. We were hoping that these methods could bring new and interesting results which are not achieved by other methods. That's kind of like what I referenced before—when you use the different methods, basically each of them can bring something that the other one cannot. Machine learning and AI methods are really great for navigation of a large chemical space, and they are actively utilized by the scientific community for the virtual screen and compound prioritization. As I mentioned, when we screened the NPC collection, we ended up with about 153 relatively potent single agents, but in drugs, with AC50 less than 30 micromole. If we look at all the possible combinations among them, we end up with more than 11,000 combinations. It's a lot of time and resources to screen 11,000 combinations, I would say, so we definitely needed to apply AI and machine learning to prioritize screening. That's what we actually decided to do back then. Of course, using the AI for the combos is not something new. There are studies in the literature. We had good experience in the past with that approach, so that's why we decided to utilize it for this.

We collaborate with MIT [Massachusetts Institute of Technology] because they actually developed a new AI method called ComboNet, which consists of two parts. One part is a graph convolution network. The idea is that the approach learns the continuous representation of the unique pieces of the molecules. This representation contains both structural features of molecules and the predicted drugs. The second part of this architecture models drug-disease association. This is a lineal function that learns how the biological targets and structural features of molecules relate to antiviral activity and synergy. This unique deep learning architecture was really interesting to apply for the modeling and selection of prospective combos among more than 11,000 combinations. They trained on ComboNet and applied the prioritization of compounds. As a result, we picked up the top 30 candidates from that and experimentally tested them in the same live virus assay, which they have access to. From assay validation, we actually found two combinations: remdesivir with reserpine and remdesivir with IQ-1S. They show strong synergy invitro. Importantly, they also found that this IQ (IQ-1S) drug combination has low cytotoxicity. Reserpine is an FDA approved drug primarily used for peripheral anti-hypertension. It

has a moderate potency, but when mixed with remdesivir, it really boosts up the activity and shows synergy. IQ-1S demonstrated single digit micromole activity itself, but in the mixture with remdesivir, it really boosts up the activity. What I want to emphasize is that when we're applying this technology, we bring something here that we didn't find using the knowledge graph. But when we use the knowledge graph, we found things we could not get with deep learning. I would say all the approaches are important, especially when you put them together.

Barr: Will you please explain why compounds that bind to the human ACE2 [angiotensin-converting enzyme 2] proteins can interrupt SARS-CoV-2 replication without damaging ACE2's natural enzymatic function, and why those are promising candidates?

Zakharov: We actually published several works related to that. ACE [angiotensin-converting enzyme] and ACE2 are key enzymes in the renin-angiotensin-aldosterone system, which is implicated in renal, pulmonary, immune, and cardiovascular functions. Angiotensin-1 is cleaved into angiotensin-2 by ACE. ACE2 is responsible for conversion of angiotensin-2 into angiotensin. Due to the complex role of that system and the inhibition of ACE2, enzymatic activity will lead to increase of inflammation, fibrosis, oxidative stress, and vasoconstriction. At the same time, ACE2 is a key entry element of viral entry and inhibition of the SARS-CoV-2 spike protein interaction with ACE2 without of the implication is ACE2 enzymatic activity is the good goal to find the treatment for COVID-19. The aim of the study which we conducted was to develop allostatic binders as an anti-SARS-CoV-2 agents. So, allostatic it means it's not binding on the active side so it should not interfere with the enzymatic activity. To achieve this goal, we developed a cascade of assays, basically, and set up a triaging system which can detect the putative allostatic binding to the human ACE2 enzyme, which decreases SARS-CoV-2 replication without impacting ACE2 natural enzymatic function. That was a complex system that our biology group set up and developed. We screened a small diverse library collection initially and got a success rate of four percent. We're using this small library to build AI and machine learning models. A really nice idea of that approach is that you can use the small library of 3000 compounds, to build a machine learning model, and then make a prediction on hundreds of thousands of compounds and select some new chemotypes. That's actually what we did. They would run our model on huge, big libraries and we came up with 73 novel conformant ACE-2 binders with Kd value as low as six nanomoles. We tested these compounds in the follow-up with SARS-CoV2 live virus assays, and we confirmed an inhibitory activity. Basically, they were active there. These compounds are a good starting point for further medical chemist studies.

Barr: Can you talk about any further research that's in progress or being planned to look at the antiviral mechanisms of the five compounds your team looked at to inhibit SARS-CoV-2 replication?

Zakharov: That's a good question. There's a lot of work which can be done to dive deeper into understanding the mechanism of action. Of course, ideally, once you find the binder, you want to see how it binds. For example, co-crystallization of the compound with ACE2 enzymes would be great to get. The other thing that would be interesting to see is at what stage of viral entry it actually works. For example, we should try to detect the compound destruct interaction between spike and ACE2. Basically, if it destructs a protein-protein interaction. If it's not doing that, maybe it's perform conformational change of a spike with ACE2 complex itself and inhibit viral membrane fusion with this cell membrane. All of that requires further additional experiments to elucidate mechanisms and at what stage it's actually blocked. That's a really exciting opportunity.

Barr: Definitely. Are there any other COVID-19 studies and initiatives you have been involved in, or any projects that you hope to take on in the future?

Zakharov: Yeah, absolutely. One study we did is related a little bit, and that's an interesting study because in that study, we developed a hybrid in-silico approach, which we used to find novel inhibitors of SARS-CoV-2 variants. As you know, viruses tend to mutate and adapt to the treatments, so to develop new inhibitors is an important goal for the majority of viruses. In that study, we decided to develop the hybrid approach, which utilizes data generated right from the assay and from in-silico part we used both the machine learning and pharmacophore-based modeling. Again, you have different techniques, if you will—one of them ligand-based and the other one is pharmacophore-based. So, we combined these two with the screening results to build the predictive models and run them in the active learning loop strategy. What that means, basically we screen the database, and we use the database to build a model. When we apply this model for the big, larger space with compounds, we select compound test them. When we rebuild the model, we update the model and we do it in iterative loops. What it gives to us? It gives us, like making models and smarter in each iteration, plus it's helped us to navigate space better, if you will. We did the two rounds of the virtual screening, and we actually found 100 compounds which show activity against live virus and show activity in live virus assays. That was a really good hit rate. It was about 60%, way higher than what you usually get from the high throughput screening. They analyze the compounds, which we get and we clustered them. We found three chemotypes that haven't been published previously. This is completely new chemical space against SARS-CoV-2 infection.

Once you get this, of course, because that data came from the phenotypic screen as I mentioned, you want to elucidate possible mechanisms of action, so we decided to screen this compound against the different targets—whatever assays were available at that time. We ran them against main protease to see if any compounds are active on that target, against the receptor binding domain of the spike protein as well as ACE2 binders (the same cascade, which we developed). We again found 6 compounds bind to ACE2 without inhibition of enzymatic assays. But in this study, we move a little bit further, we tested these 6 compounds in the unique assay (pseudo particle entry assay) to check if this compound interferes with viral entries. What we found is that this compound indeed inhibits viral entry in the pseudo particle assay where we've over-expressed ACE2 proteins. We also checked whether this compound binds the ACE2 and to the spikes, and we found they don't. Since they don't bind ACE2, we propose that the activity should be independent from the ACE protein sequence, in this case. They should be active against different strains of SARS-CoV-2. The test was against two different mutant strains—there was a South African and U.K.—and compared the wild type with this pseudo particle entry assay. We found that they all have activity and some of them were even more active for the mutant rather than the wild type. That was one of the interesting studies.

Besides that, if we're talking about general projects, I want to mention that NCATS is actively participating in the Anti-Viral Program for Pandemics (APP). This is a multi-agency initiative to develop safe and effective oral antivirals. In addition to SARS-CoV-2, NCATS is working on the other viruses within the scope of the APP program that represent known threats because of their pandemic potential. The whole list can be found on the NCATS website. Since I am currently leading the AI research group at NCATS within the APP program, I'm participating in multiple antiviral projects, particularly trying to develop inhibitors for the Bunyaviruses and Flaviviruses right now.

Barr: That's very interesting. Can you talk a little bit more about work being done on the variants and how that's impacted SARS-Cov-2 research? It seems like as soon as progress is made, another variant throws a curveball.

Zakharov: All these viruses definitely have an impact on research. NCATS conducts enormous support to trace activity of different SARS-CoV-2 variants. As I mentioned before, they established an open data portal in collaboration with ACTIV [Accelerating COVID-19 Therapeutic Interventions and Vaccines] and industry partners and have compiled a database of invitro therapeutic activity against different SARS-CoV-2 variants from prioritized sets of publications. Currently everyone can see on the website more than 300 data resources with more than 40,000 data points—such as how all of these different variants affected certain activities, and stuff like that. That’s really a great and unique portal. Besides data collection, our biologist group is actively working on development of corresponding assays, which can cover different variants and sequential screenings of compound libraries. It was sort of pacing between assay development and frequency of that mutant appearing. I also think that the important part of computational studies, in this case, is to predict what new variants would become dominate next, so we can perform the assay development and screening ahead of time. We will be able to find the allostatic modulators of the target, which show some promising activity across multiple variants.

Barr: Hopefully so. There have been lots of possibilities for drugs, but many of the repurposing studies and trials for COVID-19, in all of its stages and levels of severity, have been relatively unsuccessful when it comes to actually treating people with the disease. Can you comment a little bit about why that is and how you hope computing will expedite the process and ensure that safer options are available for people?

Zakharov: Unsuccessful outcomes from the clinical trials is very tough question since each study might have their own unique reason for that. Overall, with a new pandemic, we are stepping on uncharted territory where we learn on the go and outcomes from the trials just reflect that. Regarding computational studies, it’s already proven that utilization of silicon methods can accelerate translational science and can help make studies more efficient. Indeed, if you look on low throughput screening assays using BSL-3 facilities, it’s hard to experimentally test large numbers of single agents or even combinations of them, especially if you do thousands of single agents and combinations would just be more and more, like a geometrical progression. The other thing is that complexity is increasing with the fact that, usually, we apply a cascade of assays to eliminate assay artifacts or cytotoxic compounds, to select more potent and selective compound with desire ADME profile. It’s not just one assay across hundreds of thousands, it’s multiple assays which need to be conducted. Applying in silico modeling at each stage can reduce the number of experiments that need to be conducted and make the drug discovery pipeline more efficient—and speed up the overall process. In this case we really can triage the compounds with the cascade of assays and then use this refined data to build a model, project this model to the new chemical space, and prioritize compounds for screens from that. You don’t need to screen all of these millions of compounds, but you prioritize and select the best from them.

Barr: That’s interesting. In addition to being a scientist, you’re also a person who’s been living through this pandemic for the past two and a half years. What have been some personal challenges and opportunities for you that the pandemic has presented?

Zakharov: That’s a good question. Once the pandemic happened, personally challenging and maybe for many others as well was to reorganize the workstyle from the office work to the remote work, like work from home. We need to shift and have all possible communication between team members done over the virtual meetings and not in person like we used to have it before. That was a challenging thing—to reorganize all of that and communication with the group, with the collaborators and conduct the experiments. At the same time, I would say this virtual environment it’s also opened up a huge opportunity to set up collaboration across different institutions very quickly. Before, you needed to

travel. Now, participation in the conference also was done virtually, so you don't need to spend time to go somewhere. You just hop-in a virtual meeting and can present your findings. That was a really great opportunity, and also like the fact that we can share and present important studies and the results and brainstorming with collaborators who are miles away—and do it like its fast fashion! I would say that kind of change all of this style and that were the challenges. The other opportunity it was really great because all the people around were focusing on this pandemic problem. The amount of publication really increased with that for these kind of topics.

Barr: Is there anything else you'd like to share about your COVID-19 research or experiences?

Zakharov: That's pretty much it. It's really a great journey and development and the efforts of multiple people—chemists, biologists and informatics—all working together in NCATS. It was really an interesting journey. A lot of findings—a lot of good basic and applied science was conducted and published.

Barr: Definitely. Thank you so much for all your work with the pandemic, and I wish you only the best—and continued health.

Zakharov: Thank you so much for having me today. It was really a great opportunity to present our science at NCATS. Thank you, Gabrielle.