Integrating Generative and Experimental Platforms for Biomolecular Design

Contents

1 Workshop Summary

Motivation. There is a dire need to reduce cost and time to find solutions for medical, industrial, and environmental challenges. To this end, biomolecular design aims to artificially engineer proteins, ligands, and nucleic acids (e.g. DNA, RNA) towards functions unobserved in nature. Recently, generative machine learning (ML) has demonstrated remarkable potential to design novel and functional biomolecules [8]. Such extraordinary progress has been driven by the confluence of decades-worth of scientific experimentation data and the exponential growth in capabilities of both discriminative and generative ML. However, there remains a critical gap between biomolecular engineering and ML: many ML studies today strive for state-of-the-art outcomes on static benchmarks, often in isolation from their experimental counterparts. This disconnect may lead to misaligned evaluation metrics and impactful biological problems being overlooked. With generative modeling as a core focus of the current ML landscape, collaboration between biologists and ML experts is pivotal in ensuring research is geared toward addressing the most pressing challenges with experimental validations.

Overview. Our proposed workshop aims to align real-world biological problems with generative machine learning and foster interdisciplinary collaborations by appealing to experimental and computational attendees. The central themes of this workshop are:

1. Generative ML for biomolecular design.
   - *Inverse design.* While generative models are capable of designing diverse and novel biomolecules, an unsolved challenge is designing biomolecules for specific properties and constraints.
   - *Modeling biomolecular data.* The complexities of biological processes and the inherent experimental limitations are challenging for dataset generation and modeling. There is a need for domain-specific algorithms, capable of learning from both sparse and abundant data, that can uncover the governing mechanisms in these complex systems.
2. Integrating generative ML into the workflow of experimentalists.
   - *Adaptive experimental design.* After testing designs and conducting experiments, a unsolved challenge is determining how to best use the data to inform subsequent designs and experiments. Active learning, reinforcement learning, and Bayesian ML provide foundational approaches to develop relevant techniques.
3. Biological problems ripe for ML and development of useful *in-silico* oracles and benchmarks.
   - *Problem settings.* The workshop will foster dialogue on important biological problems that have abundant data or high-throughput data generation methods but lack satisfactory ML solutions.
   - *In-silico benchmarks.* The increasing utilization of *in-silico* oracles warrants investigation of how closely they align with real-world experiments. The workshop will encourage the development of robust and informative oracles.

Publication. The workshop is partnering with *Cell Systems* for a special collection, where select papers will be invited for fast-track submission. This collaboration aims to elevate impact of the presented work and help attract top-tier submissions, especially from the biology sector. See Section 3 for details.

2 Logistics

2.1 Format

Our workshop will **be held in-person** with **virtual arrangements** for those unable to attend in-person. It will be a **large-attendance talk** format enriched with **two poster sessions**, a **panel discussion**, **contributed talks** for best papers at the workshop, and multiple **networking sessions**. We believe this format will strike a balance between delivering scientific advancements and building connections between the computationalists and experimentalists attending the workshop. We have planned multiple poster sessions and networking events to discourage attendees from breaking off into small, familiar groups and instead establish relationships with potential collaborators.

Virtual engagement. We will offer virtual engagement options for those unable to attend in person. Attendees will be able to access recorded sessions, posters, slides, and papers through our workshop website. We will require all accepted submissions to send us their posters and camera-ready papers as PDFs. In addition, we will provide links to papers published at our workshop and successfully fast-tracked in *Cell Systems*.

2.2 Accessibility

Website. We will create a website[1] for the workshop. Accepted ML-track papers and biology-track abstracts will be released on our website before the workshop and maintained afterwards. We will feature talk titles, abstracts as well as the final schedule for attendees. The sessions for each poster will also be made available as a public Google Sheets document. We will use OpenReview for the review process.

Resources. To promote mutual understanding between experimentalists and computationalists for productive and insightful discussions, we will provide tutorial videos covering key concepts in biomolecular design and generative ML. Furthermore, presenters will receive guidance on making their content more accessible and engaging for all attendees. We also will organize social mingles to promote networking between our diverse audience.

[1] Tentative URL https://www.mlbiodesign.io

Travel awards. To foster diversity, equity, and inclusion (DEI), we will provide free workshop registration and fund travel for selected applicants using sponsorship funding, where the priority will be given to *students and minority groups based on DEI*.

2.3 Audience

Our audience would be diverse given the interdisciplinary nature of our workshop. Based on discussions with related workshop organizers[2], we anticipate around **200 attendees**. Our primary goal is to attract researchers working at the intersection of machine learning and biology. We also seek to engage pure ML researchers who are looking for applications for their work, as well as biologists who are exploring new machine learning techniques to address their problems. Additionally, we aim to draw in industry researchers actively involved in this field.

2.4 Broad Outreach

Our organizing and program committee members will leverage their extensive academic networks to increase workshop awareness in both generative ML and biology. This includes sharing event details within their institutions and with peers in the field. We will also collaborate with our sponsors and our connections in industry research labs, such as Valence, Recursion, Amgen, Genentech, Intel, and Microsoft, to disseminate information about this workshop. To extend our reach, we will employ various social media platforms, including Twitter, Facebook, LinkedIn, WeChat, and blog posts to foster interactions and engage with the general audience of the workshop.

2.5 Timeline

Main workshop deadlines:

- Workshop submission deadline: February 3rd, 2024.
- Workshop accept/reject notification date: March 3rd, 2024.

We will then follow-up with selected papers for fast-tracking to *Cell Systems*, where decisions are tentatively scheduled by May 2024.

3 Submissions

Submission tracks. The workshop submission is designed to attract high-quality original papers at the intersection between biomolecular design and generative AI. We will provide two separate submission tracks for topics described in section 1:

- **Machine learning track**. This track will feature generative machine learning advancements for biomolecular design where results are entirely *in silico*. The topics of the papers will include inverse design, biomolecular data modelling, and adaptive experimental design, *in silico* benchmarks.
- **Biology track**. This track will consist of papers which have *wet lab* experimental results. We will welcome hybrid works employing ML for experimental biomolecular design problems, as well as biological problem settings (e.g. high throughput techniques, single cell analysis) relevant to generative machine learning.

By providing a biology track and partnering with *Cell Systems*, we hope to attract researchers from hybrid labs, as well as biologists who are pursuing state-of-the-art ML techniques for their research.

Submission guidelines. Both tracks will accept submissions up to 5 pages in length (excluding appendix). The biology track will also consider extended abstracts (up to 2 pages), similar to standard biological conferences. All submissions will be non-archival. We will be explicit in our Call for Papers that papers previously published at an archival venue will be rejected.

Review process. The review process will be double-blind, and they will be conducted through OpenReview. We anticipate around *100* submissions, based on discussions with organizers with previous related workshops. We aim to accept around *60* papers. We will recruit up to *100* reviewers from diverse backgrounds. (see section 10). Each submission will receive *3* reviews and each reviewer will review up to *3* submissions.

Conflict of interest. To prevent conflicts of interest, reviewers will not evaluate submissions from their department or from collaborators within the past 5 years.

Awards. Accepted papers with exceptional quality will be recognized through Best Paper and Distinguished Paper awards, as agreed upon by review scores and area chairs. Outstanding accepted papers will be selected for a series of contributed talks, offering a platform for further discussion among workshop attendees.

Partnership with *Cell Systems* In collaboration with *Cell Systems*, authors can opt for their paper to be considered for a fast-track review at *Cell Systems*. Based on the workshop reviews, our organizing committee and editors at *Cell Systems* will select high quality papers for an additional round of review at *Cell Systems*. Accepted papers at *Cell Systems* will form a special collection. This will be similar to the *Cell Systems*-RECOMB partnership.

[2]Specifically, organizers from Machine Learning for Structural Biology workshop at NeurIPS.

4 Tentative Schedule

We aim to create interdisciplinary discussions surrounding the formidable challenges in biology and how generative machine learning can tackle them. To create an engaging and inclusive workshop appealing to a diverse audience, our program features a variety of sessions. These sessions encompass distinguished keynote speeches (**invited talks**), selected contributed machine learning papers, experimental biology and *in-silico* modelling abstracts (**contributed talks**), engaging **poster sessions**, an insightful **panel discussion**, and **social mingles** in between and after sessions.

Invited talks. Each invited talk is structured with 25 minutes dedicated to the presentation and an additional 5 minutes reserved for questions. Each talk will focus on a theme of the workshop (see section 1). We have invited speakers based on their well-known expertise, significant yet diverse scientific achievements, and exceptional presentation skills (see section 5.1).

Contributed talks. For contributed talks, we will employ a rigorous peer-review selection process, guided by the diversity of topics and high reviewer scores, ensuring that we spotlight outstanding and impactful submissions. Our goal is that the talks strike a balance between biology and machine learning.

Poster sessions. Poster sessions follow contributed talks, offering a broader range of topics and a space for more personal and detailed conversations. Discussions around posters will foster connections and idea exchange amongst our participants. We purposely planned multiple poster sessions so presenters can also visit other interesting works.

Panel discussion. The panel discussion will spotlight young, rising leaders at the intersection of ML and biotechnology to discuss the future of generative AI as applied to biomolecular design. We will identify current misalignment and challenges, as well as unsolved biological problems where generative ML is set to disrupt the stage. The panel discussion will be moderated by a workshop organizer, and will include a Q&A session with the audience. (see section 5.2).

Social mingles. A key objective of the workshop is to bring experimentalists and computationalists together for collaborations. With support from our generous sponsors, we plan to provide lunch within the venue to foster continued interaction. Additional lunch activities, such as topic-specific, technique-based, or self-organized breakout sessions, can be arranged if the venue layout permits. We will also organize an after party to encourage further networking and idea exchange.

Tentative schedule	
8:50 - 9:00 AM	Open remarks
9:00 - 9:30 AM	Invited talk 1
9:30 - 10:00 AM	Invited talk 2
10:00 - 10:15 AM	Coffee break
10:15 - 10:45 AM	Contributed talks (3 talks)
10:45 - 11:45 AM	Poster session 1
11:45 - 1:00 PM	Lunch break
1:00 - 1:30 PM	Invited talk 3
1:30 - 2:00 PM	Invited talk 4
2:00 - 2:15 PM	Coffee break
2:15 - 3:00 PM	Panel discussion
3:00 - 3:30 PM	Contributed talks (3 talks)
3:30 - 4:30 PM	Poster session 2
4:30 - 4:45 PM	Awards, closing remarks

5 Invited Speakers and Panelists

5.1 Invited Speakers

- Lucy Colwell (lcolwell@google.com, confirmed) is an University Associate Professor at Cambridge University and Research Scientist at Google. Her work has been used to predict protein interactions and provide functional annotations of proteins used widely in computational biology [1]. Professor Colwell brings a unique perspective of working in protein design within Google and using compute at scale.

- Bruno Correia (bruno.correia@epfl.ch, confirmed) is an Associate Professor at École Polytechnique Fédérale de Lausanne. Professor Correia's work on *de novo* protein design nearly always involves experimental validation. His work involving the design of novel protein interactions has been recognized in journals such as *Nature* [2]. His vast experience in collaborations with geometric deep learning experts provides an example of successful integration between cutting edge ML and experimental biology.
- Debora Marks (debbie@hms.harvard.edu, confirmed) is a Professor of Systems Biology at Harvard Medical School. Her work has been seminal in understanding the role of evolutionary couplings (e.g. multiple sequence alignments) for structure prediction [5]. Recently, her works has focused on predicting function from RNA sequences/genetic mutations and forecasting viral evolution [6]. Professor Marks' vast knowledge and impact in computational biology, mathematics, and computing reflect the themes of our workshop.
- Alex Rives (arives@cs.nyu.edu, confirmed) is the founder and CEO of EvolutionaryScale. He led the development of ESM and ESMFold [4] at Meta which have become immensely useful tools in computational biology for protein structure prediction. His experience leading a multi-disciplinary team at Meta provides a perspective on how experimental biology problems can be tackled with scale.

5.2 Invited Panelists

- Amirali Aghazadeh (amiralia@gatech.edu, confirmed) is an Assistant Professor at Georgia Institute of Technology. His research focuses on developing machine learning methods for addressing challenges in AI for Science. He leverages generative ML and language modeling for various biological applications, spanning genomics, function prediction, and sequence-based design. He also pioneers algorithms suitable for large-scale datasets and combinatorial spaces.
- Mohammed AlQuraishi (m.alquraishi@columbia.edu, confirmed) is an Assistant Professor at Columbia University. His work is at the intersection of ML, biophysics, and systems biology, focusing on both molecular and systems-level perspectives of proteomes. His active collaborations with experimental scientists closely resonate with the themes of our workshop.
- Noelia Ferruz (nfccri@ibmb.csic.es, confirmed) is a Group Leader at the Barcelona Institute of Molecular Biology. Her team focuses on protein design for various applications, spanning health and sustainability. Her work involves harnessing deep unsupervised generative models and large-scale molecular dynamics for protein design, as well as validating and refining computational models through experimental characterizations.
- Gevorg Grigoryan (ggrigoryan@generatebiomedicines.com, confirmed) is the co-founder and CTO at Generate Biomedicines. He leads on-demand drug design platform development based on generative models to create novel protein structures with desired properties. Additionally, he is a research associate professor at Dartmouth College, and his research there focuses on understanding the protein sequence-structure relationship, protein folding, interaction, and function from data.
- Kotaro Tsuboyama (ktsubo@iis.u-tokyo.ac.jp, tentative) is a Lecturer at University of Tokyo. His research focuses on experimental high-throughput analytical methods to study protein properties and functions, with an emphasis on de novo designed proteins. He employs biological and machine learning methods to study protein interaction prediction, protein stability, and cytosolic delivery of proteins.
- Rohit Singh (rsingh@csail.mit.edu, confirmed) is a Research Scientist at MIT CSAIL and an incoming Assistant Professor at Duke University. His research focuses on harnessing machine learning to advance precision diagnostics and therapeutics. He employs methods such as causal inference, language models, and network alignment to identify potential drug targets and design therapeutic interventions.

6 Organizers and Biographies

Website page and email address in colored hyperlink. In bold we have highlighted previous organizing or related experience.

- Chenghao Liu (chenghao.liu@mail.mcgill.ca) is a PhD candidate at McGill University and Mila - Québec AI Institute, advised by Dima Perepichka and Yoshua Bengio. He is a co-founder of Dreamfold, a protein design start-up. He is a chemist by training, and his research is now focused on developing generative active learning methods for materials and drug discovery. **He is a co-organizer of the CQMF chemistry conference.**
- Jarrid Rector-Brooks (jarrid.rector-brooks@mila.quebec) is a PhD candidate at the Université de Montréal and Mila - Québec AI Institute advised by Yoshua Bengio and Sarath Chandar. He is a co-founder of Dreamfold, a protein design start-up. His research aims to develop improved generative models specifically for the design of therapeutics with an eye towards high-throughput adaptive experimental design. **This is his first time co-organizing a workshop.**
- Jason Yim (jyim@mit.edu) is a PhD candidate at the Massachusetts Institute of Technology (MIT) within Computer Science and Artificial Intelligence Laboratory advised by Tommi Jaakkola and Regina Barzilay. His research focuses on developing generative model for sequential and geometric scientific data as well as experimental design in biological experiments. He has previously worked as a research engineer at DeepMind and interned at Microsoft AI4science on scientific applications with ML. **He is a co-organizer of the MoML conference and previously lead organizer of the Johns Hopkins University hackathon HopHacks .**
- Soojung Yang (soojungy@mit.edu) is a PhD student at the Massachusetts Institute of Technology (MIT), advised by Rafael Gómez-Bombarelli. Her research focuses on exploring the protein conformational landscape by harnessing the power of machine learning and multiscale molecular simulations. Additionally, she collaborates with experimental scientists on therapeutic peptide discovery projects. **This is her first time co-organizing a workshop.**

- Sidney Lisanza (lisanza@uw.edu) is a graduate student at the University of Washington within the Institute for Protein Design. He develops ML tools to expedite the protein design process both at the lead discovery stage and the subsequent optimization of candidates. He enjoys time with friends/family, being outside, listening to music, and preferably doing all simultaneously. **This is his first time co-organizing a workshop.**
- Francesca-Zhoufan Li (fzl@caltech.edu) is a bioengineering graduate student at the California Institute of Technology (Caltech), advised by Frances Arnold and Yisong Yue. Leveraging her experimental background, she focuses on predicting engineered protein functions with a particular interest in expediting wet-lab enzyme engineering. Her collaboration with Microsoft Research involves investigating the efficacy of pretrained protein language models for various protein engineering tasks. **This is her first time co-organizing a workshop.**
- Pranam Chatterjee (pranam.chatterjee@duke.edu) is an Assistant Professor of Biomedical Engineering and Computer Science at Duke University. Research in his Programmable Biology Group exists at the interface of computational design and experimental engineering, specifically employing generative ML to design programmable proteins for genome, proteome, and cell engineering. He completed his SB, SM, and PhD from MIT and is the founder of two startups, Gameto, Inc. and UbiquiTx, Inc., that leverage AI to design the next generation of fertility solutions and therapeutic biologics, respectively.
- Tommi Jaakkola (tommi@csail.mit.edu) is the Thomas Siebel Professor of Electric Engineering and Computer Science at MIT. His research covers theory, algorithms, and applications of machine learning, from statistical inference and estimation to natural language processing, computational biology, as well as recently machine learning for chemistry. His awards include Sloan research fellowship, AAAI Fellow, and many publication awards across the research areas.
- Regina Barzilay (regina@csail.mit.edu) is a School of Engineering Distinguished Professor of AI & Health in the Department of Computer Science and the AI Faculty Lead at MIT Jameel Clinic. She develops machine learning methods for drug discovery and clinical AI. In the past, she worked on natural language processing. Her research has been recognized with the MacArthur Fellowship, an NSF Career Award, and the AAAI Squirrel AI Award for Artificial Intelligence for the Benefit of Humanity.
- David Baker (dabaker@uw.edu) is the Henrietta and Aubrey Davis Endowed Professor in Biochemistry at the University of Washington. He serves as the director of the Rosetta Commons, a consortium of labs and researchers that develop biomolecular structure prediction and design software. His work has pioneered computational and experimental methods for protein design including RosettaFold which won Science magazine's breakthrough of the year in 2021.
- Frances Arnold (frances@cheme.caltech.edu) is the Linus Pauling Professor of Chemical Engineering, Bioengineering, and Biochemistry and the Director of the Donna and Benjamin M. Rosen Bioengineering Center at Caltech. Renowned for pioneering directed evolution, she has been recognized by numerous awards, including the 2018 Nobel Prize in Chemistry. She has co-founded companies such as Gevo and Provivi and served on the boards for companies including Alphabet, Illumina, and Generate Biomedicines. Since January 2021, she has been an external co-chair of President Joe Biden's Council of Advisors on Science and Technology (PCAST).
- Yoshua Bengio (yoshua.bengio@mila.quebec) is a Full Professor in the Department of Computer Science and Operations Research at Université de Montréal, as well as the Founder and Scientific Director of Mila and the Scientific Director of IVADO. Considered one of the world's leaders in artificial intelligence and deep learning, he is the recipient of the 2018 A.M. Turing Award. He is a Fellow of both the Royal Society of London and Canada, an Officer of the Order of Canada, and a Canada CIFAR AI Chair.

7 Diversity, Equity, and Inclusion

We are dedicated to the cause of diversity, equity, and inclusion (DEI) in our proposed workshop. Our efforts span a wide spectrum of academic disciplines, cultural backgrounds, personal experiences, and identities, ensuring an inclusive environment for all. We are committed to creating a space where each individual feels valued and welcomed, regardless of ethnicity, gender, sexual orientation, affiliations, nationality, seniority, abilities, socioeconomic status, religion, backgrounds, experiences, viewpoints, perspectives, and beyond.

Our organizing committee members come from **7 academic institutions**[3] with expertise in more than **6 academic disciplines**[4]. Our organizers identify with more than **8 cultural backgrounds**[5] with **4 female organizers**[6] and include first-generation students[7].

We have **6 professors** and **6 PhD student** organizers who work on a wide range of problems related to biomolecular design. Some of our student organizers are first time workshop organizers, while others bring prior experience from organizing conferences or events in various organizations (see Section 6). We provide many industry viewpoints through interning or working at tech companies such as DeepMind and Microsoft Research on scientific applications with ML. Several of organizers are founders or scientific advisors to biotech start-ups and established pharmaceutical companies. The experience level of our student organizers span from 1st to 5th year PhD students while our professors span from assistant to full professors.

We have thoughtfully invited speakers and panelists with DEI in mind. Each speaker and panelist represents a different academic institution or industry lab. Two out of four of our speakers are female and are accomplished researchers in computational biology. Our panelists

[3] Mila, MIT, Caltech, University of Washington, Duke University, McGill University, and Université de Montréal.
[4] Including computer science, computational biology, biochemistry, physical chemistry, chemical engineering, and bioengineering
[5] Including Canada, China, Finland, France, India, Israel, Kenya, and South Korea
[6] Soojung, Francesca, Regina, and Frances
[7] Jarrid, Soojung, and Francesca

and speakers span three continents, Asia, Europe, and North America. We selected speakers who come from different backgrounds (e.g. computer science, experimental biology) and whose topics are orthogonal.

In our commitment to fostering DEI, we are also dedicated to increasing global participation in our workshop. To facilitate international travel to the extent within our control, we will collaborate with the ICLR main organizers and provide assistance in obtaining invitation letters for those who may require them for visa and travel-related purposes. Our aim is to ensure that individuals from across the globe have the opportunity to join us, share their insights, and contribute to our collective learning and growth.

Finally, our workshop will promote a venue for safe, non-judgemental, and respectful discourse. We will set-up anonymous communication channels (phone, email, and form) to report any misbehaviour or DEI related issues.

8 Previous Related ICLR Workshops

An undoubtedly impactful application is combining ML with scientific applications. We list the most related workshops at previous ICLR conferences in order of relevance (first is most relevant).

- Machine Learning for Drug Discovery Workshop. ICLR 2022-2023. This workshop focuses on optimizing and discovering therapeutic candidates. Our workshop focuses on general purpose design of biomolecules – proteins, RNA/DNA – and more so on the generative perspective. The broader scope allows us to explore diverse applications; for example, enzyme engineering can significantly contribute to plastic degradation and gene editing with CRIPSR. Both workshops emphasize bringing ML closer to real-world evaluation and using ML as a core part of biological experimentation. An distinctive feature of our workshop is our aim to encourage the involvement of experimentalists, who have been underrepresented in prior ML conferences and workshops, despite their crucial role in generative biology. To achieve this goal, our workshop offers dedicated tracks and collaboration with a high impact biology journal, and our speakers and panelists come from a wide variety of backgrounds, ranging from experimentation to computational modeling.
- Deep Generative Models for Highly Structured Data. ICLR 2019 & 2022. This workshop focuses on incorporating structure into generative models for downstream use in real-world data modalities. While proteins and biological data is one application, it is not the focus whereas it is in our workshop. Furthermore, we focus on how to integrate generative models into real-world biological experiments.
- Machine Learning for Materials. ICLR 2023. The goals of this workshop are similar to ours but focused on materials. Here they emphasize identifying the unique challenges with designing useful materials and uncovering the meaningful tasks for novel ML techniques to be developed. We share the common aim of emphasizing the complexity of working with biomolecules and the need for directing ML research towards the most pressing problems in biomolecular design.
- Physics for Machine Learning. ICLR 2023. This workshop focuses on developing ML applications for physics. They also emphasize bringing ML closer to real-world applications in the physical sciences.

Why this workshop (and why now). It is an exciting time to work at the intersection of biology and ML. AlphaFold2 [3] undoubtedly shifted the direction of many ML researchers to work on structural biology. Experimental biology is reaching a inflection point of data generation becoming cheaper and faster than ever. Works such as RFdiffusion [8] published in *Nature* is a example of novel ML research combined with experimental validation to achieve unprecedented biomolecular design success. The advent of ChatGPT and foundation models provides a possible road map to developing interactive systems between scientists and AI for scientific discovery [7]. However, the ChatGPT paradigm cannot be transferred to biology where data cannot be readily annotated or generated at scale. There will need to be important breakthroughs in generative AI tailored to biology and adaptive experimental design to reduce cost and time. We believe the time is ripe to bring together the different disciplines to chart a path towards rapid scientific discovery in the age of AI.

9 Sponsors

Google Deepmind and Dreamfold have confirmed to be sponsors for the workshops, and Prescient Design has tentatively confirmed to be a sponsor. Their monetary contributions have yet to be confirmed but we expect at least $15,000. We are in active discussions with other industrial research labs for sponsorship. The money from sponsors will be used in the following ways ordered from highest to lowest priority:

1. Registration and travel grants for minority groups and students.
2. Lunch catering.
3. Best paper awards.
4. Post conference gathering at a nearby restaurant.

10 Program Committee

The role of program committee members will be to help review the workshop submissions. Each organizer will recruit enough reviewers to provide 3 reviews for each submission. Since each organizer comes from a different lab with different personal connections, we expect to have a rich pool or reviewers. We have prepared a list of reviewers who have tentatively confirmed to be reviewers. We have additional ways throughout our diverse connections to seek more reviewers if needed.

Patrick J. Almhjell, Ph.D., Simon Axelrod, Ph.D., Ava P. Amini, Ph.D., Emmanuel Bengio, Ph.D., Joey Bose, Ph.D., Tim-Henrik Buelles, Ph.D., Bianca Dumitrascu, Ph.D., Michael Galkin, Ph.D., Alex Hernandez-Garcia, Ph.D., Riashat Islam, Ph.D., Kadina Johnston, Ph.D., Michal Koziarski, Ph.D., Alex X. Lu, Ph.D., Pablos Lemos, Ph.D., Ge Liu, Ph.D., Sulin Liu, Ph.D., Santiago Miret, Ph.D., Ariane Mora, Ph.D., Nathan Frey, Ph.D., Andrei Nica, Ph.D., Ladislav Rampasek, Ph.D., Seongok Ryu, Ph.D., Lena Simine, Ph.D., Almer van der Sloot, Ph.D., Alexander Tong, Ph.D., Kevin K. Yang, Ph.D., Zichao Yan, Ph.D., Bruce Wittmann, Ph.D., Zach Wu, Ph.D., Tara Akhound-Sadegh, Lucas Arnoldt, Ron Boger, Shahar Bracha, Paul Bertin, Maria Carreira, Tianlai Chen, Itamar Chinn, MinGyu Choi, Felix Faltings, Jacob Gershon, Prashant Govindrajan, Sarah Gurev, Guillaume Huguet, Ian Humphreys, Moksh Jain, Andrew Kirjner, Maksym Korablyov, Daniel Levy, Seokhyun Moon, Sean Murphy, Peter Mikhael, Juno Nam, Lena Nehale Ezzine, Umesh Padia, Raman Samusevich, Wenxian Shi, Luca Thiede, Brian Trippe, Veronica Tarka, Tony Tu, Pascal Sturmfels, Akshay Subramanian, Allen Tao, Hannes Stark, Sophia Vincoff, Sasha Volokhova, Rachel Wu, Jason Yang, Dinghuai Zhang, Yinnuo Zhang, Wonho Zhung.

References

[1] Maxwell L Bileschi, David Belanger, Drew H Bryant, Theo Sanderson, Brandon Carter, D Sculley, Alex Bateman, Mark A DePristo, and Lucy J Colwell. Using deep learning to annotate the protein universe. *Nature Biotechnology*, 40(6):932–937, 2022.

[2] Pablo Gainza, Sarah Wehrle, Alexandra Van Hall-Beauvais, Anthony Marchand, Andreas Scheck, Zander Harteveld, Stephen Buckley, Dongchun Ni, Shuguang Tan, Freyr Sverrisson, et al. De novo design of protein interactions with learned surface fingerprints. *Nature*, pages 1–9, 2023.

[3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873): 583–589, 2021.

[4] Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

[5] Debora S Marks, Lucy J Colwell, Robert Sheridan, Thomas A Hopf, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Protein 3d structure computed from evolutionary sequence variation. *PloS one*, 6(12):e28766, 2011.

[6] Nicole N Thadani, Sarah Gurev, Pascal Notin, Noor Youssef, Nathan J Rollins, Daniel Ritter, Chris Sander, Yarin Gal, and Debora S Marks. Learning from prepandemic data to forecast viral escape. *Nature*, pages 1–8, 2023.

[7] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[8] Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976): 1089–1100, 2023.