Report from Dagstuhl Seminar 24451

Machine Learning for Protein-Protein and Protein-Ligand Interactions

Anne-Florence Bitbol^{*1}**, Jennifer Listgarten**^{*2}**, Tomas Pluskal**^{*3}**,**
Anton Bushuiev$^{\dagger 4}$**, and Roman Bushuiev**$^{\dagger 5}$

1 **EPFL – Lausanne, CH.** anne-florence.bitbol@epfl.ch
2 **University of California – Berkeley, US.** jennl@berkeley.edu
3 **IOCB – Prague, CZ.** tomas.pluskal@uochb.cas.cz
4 **Czech Technical University – Prague, CZ.** anton.bushuiev@cvut.cz
5 **The Czech Academy of Sciences – Prague, CZ.** roman.bushuiev@uochb.cas.cz

Abstract

Dagstuhl Seminar 24451 focused on how machine learning (ML) is revolutionizing computational biology and chemistry by enhancing the prediction and design of protein-protein and protein-ligand interactions. Key topics included integrating biological and chemical knowledge into ML models, addressing data quality and availability issues, and fostering interdisciplinary collaborations. Theoretical discussions explored representation learning, generative models, and protein language models as efficient alternatives to traditional methods. Practical sessions emphasized the importance of experimental constraints in ML workflows and proposed standards for balanced datasets. The seminar concluded by encouraging collaboration between computational and wet-lab researchers, setting the groundwork for future innovations in protein science and drug discovery.

Seminar November 3–8, 2024 – https://www.dagstuhl.de/24451
2012 ACM Subject Classification Applied computing → Life and medical sciences
Keywords and phrases biological machine learning, ligand, molecular interactions, protein
Digital Object Identifier 10.4230/DagRep.14.11.1

1 Executive Summary

Tomas Pluskal (IOCB – Prague, CZ)
Anne-Florence Bitbol (EPFL – Lausanne, CH)
Jennifer Listgarten (University of California – Berkeley, US)

The Dagstuhl Seminar 24451, titled "Machine Learning for Protein-Protein and Protein-Ligand Interactions", convened leading experts from computational biology, chemistry, and machine learning (ML) to explore advancements and challenges in understanding biomolecular interactions. This was the first seminar of its kind, and it aimed to address pressing issues such as integrating domain knowledge into ML models, ensuring data availability and quality, and fostering effective interdisciplinary collaboration. The event facilitated discussions on theoretical advancements, practical challenges, and future research directions in leveraging ML for protein science and drug discovery.

[*] Editor / Organizer
[†] Editorial Assistant / Collector

Machine Learning for Protein-Protein and Protein-Ligand Interactions
Dagstuhl Reports, Vol. 14, Issue 11, pp. 1–15
Editors: Anne-Florence Bitbol, Jennifer Listgarten, and Tomas Pluskal
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

A central theme of the seminar was the exploration of ML techniques for predicting protein-protein and protein-ligand interactions with improved accuracy and interpretability. Topics included representation learning, generative modeling, and the role of inductive biases in structuring ML models. The emergence of protein language models (pLMs) was a particular highlight, as they offer sustainability and efficiency advantages over traditional sequence alignment-based methods. These models were recognized for their ability to enhance structural and functional predictions, particularly in cases where evolutionary data is scarce or unreliable.

Another critical discussion focused on benchmarking ML models for biomolecular interactions. Participants debated the limitations of current datasets, particularly in terms of negative data and class imbalances, and emphasized the need for standardized benchmarks. Reliable benchmarking was identified as essential for validating new approaches, particularly in enzyme design and protein interaction network analysis.

The seminar also tackled real-world challenges in applying ML to protein science. One of the most pressing issues was the scarcity of high-quality datasets, particularly for protein-ligand interactions. Participants proposed new standards for curating balanced datasets and ensuring the inclusion of negative data to improve ML model training. Additionally, discussions highlighted the importance of integrating experimental constraints into ML workflows, which would enhance cost-effectiveness in protein engineering and directed evolution.

The integration of computational and wet-lab research was another key topic. Effective collaboration between these fields remains a major challenge due to differing methodologies and objectives. Strategies were proposed to improve synergy between computational modeling and experimental validation, including the development of interdisciplinary training programs and shared data repositories.

Inspired by previous Dagstuhl meetings, the seminar adopted a discussion-driven format, encouraging open exchanges and collaboration. Several actionable outcomes emerged, including initiatives to enhance data accessibility, develop standardized benchmarking frameworks, and refine ML model architectures for improved performance in biomolecular applications.

The event successfully bridged computational and experimental research, paving the way for future innovations in protein science and drug discovery. Moving forward, participants emphasized the need for continued interdisciplinary collaboration, the development of more reliable datasets, and the refinement of ML techniques to better capture the complexity of protein interactions. The outcomes of this seminar are expected to significantly influence the future of machine learning applications in biology and chemistry, setting the stage for groundbreaking advancements in the field.

2 Table of Contents

3 Overview of Talks

3.1 Reliable protein-protein interaction prediction and benchmarking (talk)

Arne Elofsson (Stockholm University – Solna, SE)

This session addressed the challenges in predicting protein-protein interactions, focusing on whether high-accuracy predictions require detailed structural data like AlphaFold models. It suggests that in real-world scenarios, experimental data are often context-specific, raising questions about the optimal benchmark for these predictions and whether general models are sufficient. Key challenges include predicting interactions within similar (paralogs vs orthologs) proteins versus novel or unrelated ones, and distinguishing merely whether two proteins interact from understanding how they interact. Coevolution is a valuable signal if it can be found, which is not always easy. The potential for a progressive approach – starting with simple protein pairs and building up to complex multimers – was discussed, allowing to mapping broader protein networks. This is however still far from automated. Specific issues with data quality, especially binding affinity data, were raised. Data availability and quality are here major limiting factors. There is a lack of high-quality, standardized test datasets. This data limitation impacts both protein-protein and protein-small molecule interaction and binding affinity predictions. The session also discussed potential benchmarking improvements, including creating "clean" datasets that address issues of human protein bias, data leakage, and proper deduplication. There is a need for balanced, fair train/test sets that account for data imbalances, and robust negative datasets.

3.2 Protein language models, foundation models (talk)

Burkhard Rost (TU München, DE)

Large Language Models for proteins, namely protein Language Models (pLMs), have begun to provide an important alternative to capturing the information encoded in a protein sequence in computers. Arguably, pLMs have advanced importantly to understanding aspects of the language of life as written in proteins, and through this understanding, they are becoming an increasingly powerful means of advancing protein prediction, e.g., in the prediction of molecular function as expressed by identifying binding residues or of variant effects upon molecular function. The so-called pLM embeddings implicitly capture the information. Therefore, they suffice as exclusive input into downstream supervised methods for protein prediction. Over the last 33 years, evolutionary information from multiple sequence alignments (MSAs) has been the most successful universal key to the success of protein prediction. For many applications, MSA-free pLM-based predictions now have become significantly more accurate. The reason for this often is a combination of two aspects: embeddings condense the grammar so efficiently that downstream prediction methods succeed with few free parameters (i.e. unusually small models for the era of deep neural networks) and pLM-based methods provide protein-specific solutions. As additional benefit, once the pLM training has been completed, pLM-based solutions often consume much fewer resources than MSA-based solutions. In fact, here we appeal to the community to rather optimize foundation models than retraining new ones and to evolve incentives to create new solutions that require fewer resources even at some loss in accuracy. Although pLMs have not, yet, succeeded to entirely replace the body of solutions developed over three decades, they clearly are rapidly advancing as THE universal key for protein prediction. Fine-tuning foundation pLMs enhances efficiency and accuracy of solutions, in particular in cases with few experimental annotations. pLMs facilitate the integration of computational and experimental biology, of AI and wet-lab, in particular toward a new era of protein design.

4 Working groups

4.1 Importance of MSAs and homology

Anne-Florence Bitbol (EPFL – Lausanne, CH)

Models based on MSAs are powerful at capturing coevolution, and are used in AlphaFold2 and AlphaFold3. Single-sequence based models capture coevolution at least to some extent, and appear to "figure out" MSAs. Single-sequence models can infer homology without MSAs, as shown by methods like PoET and ProtMamba, raising questions about training objectives, sequence order, and optimization. Pairwise alignments offer simpler alternatives, especially for rare proteins and data-limited cases. Strategies like data augmentation (e.g., VAEs), sequence-to-MSA models, and leveraging ancestral sequences show promise, though biases in inferred data remain a challenge. These approaches aim to balance evolutionary information use while addressing data limitations and enhancing downstream task performance.

4.2 Protein language models, foundation models

Anne-Florence Bitbol (EPFL – Lausanne, CH), Ilia Igashov (EPFL – Lausanne, CH)

With the rapid growth of the available computational capabilities, the computational biology community has extensively explored the capabilities of protein language models (PLMs), self-supervised methods designed to model the distribution of the amino acids in protein sequences. While these models have demonstrated high performance in addressing various tasks (protein structure prediction, protein design, variation effect prediction, etc.), many questions and challenges in their training and applications remain open. From the algorithmic perspective, it is still unclear if the highest efficiency of PLMs is achieved in the autoregressive setting, or other masked modeling strategies (including recently introduced discrete diffusion) can further improve the performance of these models. From the data perspective, there is no consensus if the available sequence data is sufficient to train a fully generalizable model. Besides, can PLMs help go outside of the sequence distribution induced by a small fraction of sequences available for learning? Besides, is it beneficial to include additional protein representations such as protein structure and train bilingual PLMs? It was additionally discussed if fine-tuning of generally-trained PLMs to specific tasks using smaller datasets can pose the optimal solution to the community in terms of both performance and sustainability.

4.3 Defining similarity between biomolecular interactions, fast search

Anton Bushuiev (Czech Technical University – Prague, CZ), Roman Bushuiev (The Czech Academy of Sciences – Prague, CZ), Josef Sivic (Czech Technical University – Prague, CZ), Martin Steinegger (Seoul National University, KR)

We discussed the ongoing challenges in defining and efficiently comparing protein-protein and protein-ligand interactions at a large scale, particularly in integrating both geometric and biochemical aspects of their similarity. Despite various approaches, reaching a consensus on interface definitions was challenging. We explored existing methods, including alignment-based approaches (more precise but slower) and alignment-free methods (fast approximations). We concluded that a detailed empirical comparison is necessary to establish biologically relevant definitions of similarity and to develop fast algorithms applicable to large-scale interaction data, providing meaningful biological insights. Looking ahead, we recommended creating a benchmark to evaluate current tools and to use this foundation for developing more advanced and reliable machine-learning models.

4.4 Evolution and evolutionary paths

Simona Cocco (ENS – Paris, FR), Alexander Schug (Jülich Supercomputing Centre, DE), Martin Weigt (Sorbonne University – Paris, FR)

This session aimed to explore methods for characterizing and predicting evolution and evolutionary pathways of biomolecules, focusing on tracing accumulated mutations and navigating complex sequence fitness landscapes. Key questions included how to reconstruct pathways using current evolutionary endpoints, especially when dealing with less structured proteins or promiscuous intermediate states. Discussions addressed the best models for fitness landscapes, including structural and ligand-binding considerations, and the dynamics of protein evolution. Ancestral protein reconstruction, phylogenetic tree construction, and defining optimal transition pathways or even ensembles between sequences were also considered. Emphasis was on considering data from multiple sources, including multiple sequence alignments, short-term lab evolution experiments, and viral evolution, to refine models and improve analysis or even predictions of evolvability, evolutionary trajectories or, more generally, whole fitness landscapes.

4.5 Reliable protein-protein interaction prediction and benchmarking

Simona Cocco (ENS – Paris, FR), Alexandre Bonvin (Utrecht University, NL)

This session discussed the state-of-the-art in predicting protein-protein interactions, using both structural methods (e.g. AlphaFold-based) and sequence based, highlighting the current challenges and limitations. The question of reliably benchmarking such predictions were also discussed, especially how to define such benchmarks (including negative data) and define independent test sets.

4.6 Fine-tuning foundation models

Arne Elofsson (Stockholm University – Solna, SE)

This session explored the practice of fine-tuning foundation models, particularly focusing on its benefits and limitations within various life science applications. Topics included definitions and distinctions in model customization, such as the advantages of fine-tuning versus training smaller models from scratch and strategies to mitigate common challenges like overfitting. Participants reviewed case studies where fine-tuning substantially improved model performance for specialized tasks, such as analyzing natural compounds or RNA. Techniques such as LoRA and SetFit were highlighted for efficient model adaptation with limited data. The session concluded with future directions, emphasizing scalable and targeted fine-tuning strategies for complex biological data.

4.7 Protein dynamics as input (and/or output?) for machine learning

Arne Elofsson (Stockholm University – Solna, SE), Sergei Grudinin (CNRS – St. Martin-d'Hères, FR)

This session delved into the intersection of protein dynamics and machine learning, examining how dynamic data can be effectively utilized as both input and output in predictive models. Key discussions focused on the challenges and potential of integrating diverse data sources – such as molecular dynamics (MD) simulations, NMR, Cryo-EM, and HD-XMS – into machine learning frameworks to explore energy landscapes, conformational pathways, and reaction rates. The session further explored the limitations of equilibrium assumptions, the need for efficient data mapping, and strategies for improving MD-based predictions, including using AI to approximate MD outputs and accelerate simulations. These insights aim to advance the predictive power of machine learning in understanding and designing dynamic protein systems.

4.8 Protein symmetries, 3D alignment

Sergei Grudinin (CNRS – St. Martin-d'Hères, FR)

With the rapid growth of the available computational capabilities, structural and sequence databases, we see a rapid growth of architectures with different loss functions and kernels. We discussed and explored methods for 3D point cloud alignment, focusing on the constrained Procrustes problem and registration, concluding that these are relatively straightforward with stable solutions. Additionally, we discussed discrete alignment, specifically pairwise sequence alignment using a differentiable Needleman–Wunsch approach, and its integration into end-to-end architectures. These techniques hold promise for efficient database searches and as potential kernels within end-to-end networks.

4.9 Protein tokenization in foundation models

Sergei Grudinin (CNRS – St. Martin-d'Hères, FR), Anne-Florence Bitbol (EPFL – Lausanne, CH), Anton Bushuiev (Czech Technical University – Prague, CZ), Julius Wenckstern (EPFL – Lausanne, CH)

Models such as transformers typically require discrete data representations represented with "tokens". In the case of proteins, sequences are naturally discrete, with 20 natural amino acids. However, protein structure is continuous, and deformations and dynamics operate in continuous space too. Discretization schemes have been developed for protein structure, with success. Multi-modal models like ESM3 and SaProt highlight the potential of tokenization to enhance protein representation, opening doors for prompting strategies in protein design. Is tokenization necessary? How to do it best? How to evaluate its performance?

4.10 Uncertainty estimation in ML for small molecules

Jessica Lanini (Novartis AG – Basel, CH), Ilia Igashov (EPFL – Lausanne, CH)

Uncertainty estimation is essential in machine learning for small molecules, influencing tasks like ranking, regression, and optimization. Frequentist methods, such as adapting ranking models to output probabilities, have shown promise in protein fitness prediction, but their consistency and applicability to multidimensional data remain debated. Support vector index machines and deep ensemble methods are proposed alternatives, emphasizing data variability and distance. Overconfidence, driven by over-parametrization and prolonged training, poses a major challenge, mitigated by techniques like temperature scaling. In multi-parameter optimization, uncertainty helps navigate problem landscapes, prioritizing less uncertain solutions. Discussions also highlighted the need for calibration after model fine-tuning and the potential of uncertainty estimation to replace traditional test error metrics by providing deeper insights into data distributions, especially in generative tasks like de novo drug design.

4.11 Protein binder design: are we there yet?

Andrew Leach (University of Manchester, GB), Ilia Igashov (EPFL – Lausanne, CH), Petr Kouba (Czech Technical University – Prague, CZ), Armita Nourmohammad (University of Washington – Seattle, US)

The session was divided into two discussions focusing on protein–small molecule and protein–protein interactions, respectively. The discussion on protein–small molecule binder design was initiated by a short presentation and covered several topics: the objectives of molecular generation, the desired outputs from such approaches, methods for sanity-checking the outputs, actions to take after validation, and strategies for evaluating the models. In the discussion on protein–protein binder design, the focus was on current capabilities with existing tools, the benchmarks and evaluation standards in the field, and the most promising methodological approaches. These points were then discucces in depth.

4.12 Coevolution at the scale of PPI networks

Cyril Malbranke (EPFL – Lausanne, CH), Arne Elofsson (Stockholm University – Solna, SE)

This session on coevolution in protein-protein interactions (PPIs) explored recent advancements and challenges in understanding coevolutionary signals within PPIs, with a focus on how these signals can improve the specificity and reliability of computational models. Discussions addressed the efficacy of methods such as Direct Coupling Analysis (DCA) for filtering AlphaFold (AF2) models, the integration of structural contact information, and the reliability of multi-sequence alignments (MSA) in binder design. Additionally, the session evaluated tools like PPI-3D and Foldseek multimer for scanning protein interfaces, examining whether these resources can contribute to a more nuanced understanding of transient and stable PPIs. This collaborative session aims to refine existing computational approaches and identify future directions, particularly in enhancing model precision and exploring new applications of coevolution in PPI research.

4.13 Injecting inductive biases

Hunter Nisonoff (University of California – Berkeley, US), Bruce Wittmann (Microsoft – Redmond, US)

This session focused on the integration of inductive biases into machine learning models to enhance the prediction of protein-protein interactions (PPI), protein inverse folding, and protein function. Key discussions revolved around the necessity and methods of incorporating inductive biases, such as data representation, model architecture, and loss functions. Participants highlighted the importance of understanding and leveraging physical theories, such as molecular orbital theory, to improve model accuracy. The session also addressed the challenges of limited and noisy data, proposing solutions like data augmentation and conditioning information. Specific problems, such as inverse folding and protein-ligand interactions, were examined, with suggestions to model side chains and use energy functions to guide generative models. The session concluded with a consensus on the need for tailored inductive biases depending on the problem at hand and the potential benefits of integrating dynamics and multimodal data into the models.

4.14 Enzymatic de novo reaction prediction

Tomas Pluskal (IOCB – Prague, CZ), Aalt-Jan van Dijk (University of Amsterdam, NL)

We discussed the task of designing new enzymes and considered in particular that it would make sense to start from a specific family of interest to keep the problem somewhat small. Here, it is important to consider that the amount of information available for different enzyme families varies enormously (ranging from "virtually nothing except sequences" to "hundreds/thousands of experimentally characterized family members"). Various approaches were explored, and various details in particular of using Alphafold3 were discussed. We also discussed that different levels of "new" are used in different contexts – e.g. new product, new substrate, new product based on new substrate. One concrete action point is to come up with a roadmap which would help to clarify and formalize the different levels and interpretations of "new" that are relevant in this field.

4.15 Enzymes (Enzymatic reaction prediction)

Tomas Pluskal (IOCB – Prague, CZ), Simon Mathis (University of Cambridge, GB)

This discussion examined key challenges and opportunities in enzyme engineering and de novo enzyme design, focusing on data quality, predictive modeling, and experimental validation. Enzyme databases like BRENDA were identified as lacking the rigor of structural repositories such as PDB, with uncertainties around catalytic sites limiting progress. A CASP-like competition (CAEP) was proposed to benchmark enzyme property predictions. Central questions included understanding enzyme promiscuity, predicting substrate scope, and designing enzymes for novel substrates. Advances in similarity metrics, experimental mapping, and large-scale datasets were highlighted as critical for substrate re-purposing and specificity prediction. Emerging tools, such as density functional theory (DFT) for metalloenzymes, and studies on enzyme-substrate specificity screens underscored the need for standardized benchmarks and interdisciplinary methods to drive innovation in enzyme research.

4.16 Enzyme engineering

Juho Rousu (Aalto University, FI), Simon Mathis (University of Cambridge, GB)

This session focused on enzyme engineering, in particular optimizing properties such as solubility and thermostability, in contrast to de novo enzyme design which focuses in enzyme function and specific activity. The participants noted that enzyme engineering differs from general protein engineering only in small details. It was noted that there are three main data sources available for the task, namely deep mutational scanning data, small enzyme specific datasets generated in-house as well as data in literature. These can be supplemented with physics-based simulated data. In terms of engineering enzymes in industrial scale, stability takes priority over activity, as overall yield over time stays better with a stable, not optimally active enzyme than vice versa. Pharma industry may also apply for stereoselectivity, as stereochemistry is often important for drugs. In therapeutic applications non-immunogenic, extremely selective enzymes are preferred. To move forward, communication with the wet lab community was deemed important in order to extract important quantities to consider as well as sources for noise and biases in the data. The problem of building databases to support machine learning initiatives was also discussed. It was noted that incentivising wet-lab researchers to submit high-quality data needs to be considered, otherwise the buy-in can remain low.

4.17 Predicting graph-structured output

Juho Rousu (Aalto University, FI), Roman Bushuiev (The Czech Academy of Sciences – Prague, CZ)

The discussion examined recent advances and limitations in molecular generation methods, particularly diffusion and flow-matching models, and the challenges in encoding geometric and chemical information efficiently. Key points included the trade-off between generating novel molecules and maintaining chemical validity, the impact of initial noise distribution, and the need for effective graph encoding. Classical SMIRKS-encoded reactions were also considered, with a caution to avoid generating chemically invalid outputs that could discourage chemists. The group also explored the balance between SMILES and graph-based approaches, recognizing no consensus on which is superior. Benchmarking challenges and the utility of foundation models for molecular tasks were also highlighted, alongside an emphasis on rigorous evaluation and the potential value of integrating machine learning into chemists' workflows.

4.18 Small molecules: Deep Learning for Molecular Property and Activity Prediction

Andrea Volkamer (Universität des Saarlandes – Saarbrücken, DE), Andrew Leach (University of Manchester, GB)

The group introduced their interests and questions and then split to discuss chemical space and multi-parameter optimisation. The chemical space group concentrated on the roughness of chemical space (e.g., activity cliffs) and examined the extent of the chemical space exploration based on mass spectrometry data. The multi-parameter optimization group focused on machine learning approaches for optimization, considering issues of generalization and the availability of training data.

4.19 Efficient training, small language models, linear transformers, linear-scaling models

Bruce Wittmann (Microsoft – Redmond, US), Christian Dallago (Nvidia – München, DE), Sergei Grudinin (CNRS – St. Martin-d'Hères, FR), Cyril Malbranke (EPFL – Lausanne, CH)

This session focused on strategies to enhance the efficiency of training large language models for protein sequences. Key discussions included the exploration of linear models, such as state space models, as alternatives to traditional transformers to trade complexity for speed. The session also covered tools for efficient training, such as flash attention, which optimizes memory access and data loading processes. The importance of curated and high-quality training data was emphasized, with datasets like SwissProt and UniProt highlighted for their relevance. Participants discussed the necessity of data pre-processing, including the identification of co-evolving amino acids and the use of classifiers to improve data quality. The session concluded with a consensus on the need for interdisciplinary collaboration and the development of standardized protocols to maximize the impact of machine learning in protein modeling.

4.20 Injection of laboratory constraints into ML models and workflows

Bruce Wittmann (Microsoft – Redmond, US), Hunter Nisonoff (University of California – Berkeley, US)

In protein engineering, there has been a large interest in leveraging predictive and generative models to design libraries of sequences. However, in practice it can be very expensive to synthesize full-length genes that are proposed by the machine learning methods. In contrast, experimental directed evolution techniques typically use cheap laboratory techniques to build variants of existing genes very cheaply. This conversation was proposed to discuss how we can incorporate experimental constraints with machine learning models to design large libraries with reasonable cost.

4.21 Machine learning-guided directed evolution

Bruce Wittmann (Microsoft – Redmond, US)

This session focused on the efficient integration of machine learning (ML) into the directed evolution pipeline to enhance protein engineering. Key discussions included the application of ML algorithms to predict beneficial mutations, optimize experimental design, and interpret complex datasets. Emphasis was placed on incorporating domain-specific knowledge and leveraging high-throughput screening data to train robust models. The session highlighted the potential of ML to accelerate iterative cycles of mutation and selection, thereby reducing the time and cost associated with traditional directed evolution methods. Challenges such as data quality, model interpretability, and the integration of ML predictions with experimental workflows were addressed. The session concluded with a consensus on the need for interdisciplinary collaboration and the development of standardized protocols to maximize the impact of ML in directed evolution.

Participants

- Anne-Florence Bitbol
EPFL – Lausanne, CH
- Sebastian Böcker
Friedrich-Schiller-Universität Jena, DE
- Alexandre Bonvin
Utrecht University, NL
- Anton Bushuiev
Czech Technical University – Prague, CZ
- Roman Bushuiev
The Czech Academy of Sciences – Prague, CZ
- Alessandra Carbone
Sorbonne University – Paris, FR
- Alberto Cazzaniga
AREA Science Park – Trieste, IT
- Simona Cocco
ENS – Paris, FR
- Francesca Cuturello
AREA Science Park – Trieste, IT
- Christian Dallago
Nvidia – München, DE
- Arne Elofsson
Stockholm University – Solna, SE
- Sergei Grudinin
CNRS – St. Martin-d'Hères, FR
- Ilia Igashov
EPFL – Lausanne, CH
- Petr Kouba
Czech Technical University – Prague, CZ
- Jessica Lanini
Novartis AG – Basel, CH
- Andrew Leach
University of Manchester, GB
- Jennifer Listgarten
University of California – Berkeley, US
- Cyril Malbranke
EPFL – Lausanne, CH
- Hiroshi Mamitsuka
Kyoto University, JP
- Céline Marquet
TU München – Garching, DE
- Simon Mathis
University of Cambridge, GB
- Stanislav Mazurenko
Masaryk University – Brno, CZ
- Remi Monasson
ENS – Paris, FR
- Hunter Nisonoff
University of California – Berkeley, US
- Armita Nourmohammad
University of Washington – Seattle, US
- Tomas Pluskal
IOCB – Prague, CZ
- Burkhard Rost
TU München, DE
- Juho Rousu
Aalto University, FI
- Alexander Schug
Jülich Supercomputing Centre, DE
- Josef Sivic
Czech Technical University – Prague, CZ
- Martin Steinegger
Seoul National University, KR
- Aalt-Jan van Dijk
University of Amsterdam, NL
- Pablo Varas Pardo
Institute of Mathematical Sciences – Madrid, ES
- Andrea Volkamer
Universität des Saarlandes – Saarbrücken, DE
- Martin Weigt
Sorbonne University – Paris, FR
- Julius Wenckstern
EPFL – Lausanne, CH
- Bruce Wittmann
Microsoft – Redmond, US
- Xiaotong Xu
Utrecht University, NL
- Omri Yakir
Tel Aviv University, IL
- Lenka Zdeborova
EPFL – Lausanne, CH