Meeting Report

Birte Höcker*, Ina Koch* and Janosch Hennig*

The 76th Mosbacher Colloquium: AI-driven (r) evolution in structural biology and protein design

https://doi.org/10.1515/hsz-2025-0184
Received July 24, 2025; accepted September 16, 2025;
published online September 30, 2025

Abstract: The 76th Mosbacher Kolloquium focused on recent advances in machine learning applications for structural biology and protein design. It covered topics spanning artificial intelligence-driven protein structure prediction, integrative modeling, generative protein design, and general applications in life sciences. With strong participation, high-caliber talks, and a clear focus on the integration of AI in biomolecular research, the meeting underscored the transformative role of machine learning in molecular biosciences and provided a vibrant platform for knowledge exchange across disciplines and generations.

Keywords: protein design; protein engineering; structural biology; machine learning; bioinformatics; artificial intelligence

Since its inception in 1949, the Mosbacher Kolloquium has been one of the most prestigious and long-standing scientific conferences in molecular biosciences in Germany. Organized annually by the German Society for Biochemistry and Molecular Biology (GBM), this meeting has become a key venue for presenting cutting-edge developments across biochemistry, molecular biology, and related disciplines. Each year, the Colloquium brings together established researchers, early-career scientists, and students to discuss recent breakthroughs and emerging challenges.

The 76th Mosbacher Kolloquium, held from March 19 to 22, 2025, in the town of Mosbach, Germany, focused on machine learning in structural biology and protein design. The meeting addressed how the field has been transformed by AI-based approaches, which now influence everything from protein structure prediction to rational biomolecular design and have provided ground-breaking contributions to structural biology data analysis.

Over the past decade, machine learning (ML) has become an essential tool in structural biology and protein design. Methods powered by ML have revolutionized biomolecular structure prediction, structural analysis, and *de novo* protein design. The advent of AlphaFold, alongside numerous other AI-based software tools, exemplifies both the enormous potential and the inherent challenges in this space. These challenges include not only data handling and storage, but also the meaningful integration of heterogeneous data from diverse experimental sources with ML-based structure prediction, as well as the careful examination and interpretation of biomolecular complex structure predictions.

Given the rapid pace of development in this multidisciplinary field, the Colloquium aimed to foster scientific exchange between experts and young researchers from disciplines including biology, biochemistry, bioinformatics, physics, and computer science.

The scientific program was curated by Birte Höcker, Janosch Hennig and Ina Koch. They led a traditional primer session before the official opening, designed for early-career scientists and participants new to the field. Ina Koch began by revisiting secondary structure prediction in the 1980s using graph-based knowledge acquisition – an early precursor to modern ML (Kaden et al. 1990). Her colleague Marcus Keβler introduced fundamental AI concepts. Janosch Hennig then discussed the positive impact of AlphaFold on experimental structure determination of larger biomolecular complexes, but also its limitations and the need for experimental validations of predictions going beyond single protein domains (Hennig 2025). Finally, Birte Höcker discussed the tremendous advances in protein design enabled by ML-based tools such as protein language and diffusion models (Dauparas et al. 2022; Ferruz and Höcker 2022; Jendrusch et al. 2025; Watson et al. 2023). This introductory session also framed the five major thematic areas of the

***Corresponding authors: Birte Höcker**, Department of Biochemistry III–Protein Design, University of Bayreuth, Bayreuth, Germany, E-mail: birte.hoecker@uni-bayreuth.de. https://orcid.org/0000-0002-8250-9462; **Ina Koch**, Faculty of Computer Science and Mathematics, Molecular Bioinformatics, Institute of Computer Science, Goethe University Frankfurt, Frankfurt am Main, Germany, E-mail: ina.koch@bioinformatik.uni-frankfurt.de; and **Janosch Hennig**, Department of Biochemistry IV, Biophysical Chemistry, University of Bayreuth, Bayreuth, Germany; and Molecular Systems Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany, E-mail: janosch.hennig@embl.de. https://orcid.org/0000-0001-5214-7002

conference: theory and development, machine learning in structural biology, AI-based structure prediction and integrative modelling, from protein engineering to AI-based protein design and AI applications in life sciences.

Before the welcome reception in the town hall of Mosbach and official start of the program, the Rainer Rudolph Foundation, which promotes and supports basic and applied research in protein biochemistry and biotechnology, presented their yearly young scientist awardees. The prizes went to Davide Amendola (Zürich) for investigating a bacterial predator arsenal using cryo-EM, to Alexander Braun (Bayreuth) for designing triosephosphate isomerases using generative language models, and to Britta Lipinski (Darmstadt) for engineering bispecific single-domain antibodies. The presentations of the three awardees fitted well to the topic of the colloquium highlighting the topicality of the meeting.

On the first day, Thursday, March 20, 2025, session I (theory and development), chaired by Clara Schoeder (University of Leipzig, Germany) and Ina Koch, opened with a keynote by David Jones (University College London, UK) on domain encyclopedias and rapid multi-domain protein structure search (Lau et al. 2024). Talks by Matthias Rarey (University of Hamburg, Germany) and Andrea Volkamer (Saarland University, Germany) explored intelligent computation in molecular design and data integration in drug discovery (Backenköhler et al. 2024; Harren et al. 2024), respectively, while Burkhard Rost (TU Munich, Germany) closed the session with insights into using Natural Language Processing (NLP) to decode the "language" of proteins (Heinzinger and Rost 2025).

Session II (machine learning in structural biology) was then chaired by Ora Furman (Hebrew University, Israel), and featured Sergey Ovchinnikov (MIT, USA) on inverting deep learning-based structure prediction for protein design (Zhang et al. 2025), Noelia Ferruz (CRG Barcelona) on generative protein language models and reinforcement learning in protein design optimization (Ferruz et al. 2022; Stocco et al. 2025) and Martin Steinegger (Seoul National University, South Korea) on Foldseek, a high-speed tool for large-scale protein structure comparison (Kim et al. 2025).

All sessions included early career investigator talks and these were given by Leif Seute (Heidelberg Institute for Theoretical Studies, Germany) on geometric learning of protein conformational ensembles, Lea Brönnimann (University of Bern, Switzerland) on applying deep learning to analyze antibody chain repertoires and David Medina-Ortiz (Leibniz Institute of Plant Biochemistry, Halle, Germany) presented an interpretable geometric deep learning framework for protein engineering.

The evening of the first full day was closed with exciting award lectures. The Bayer Pharmaceuticals PhD Award was given to Franziska Sendker (Harvard University, USA) for her work on protein fractal evolution. The Karl-Lohmann Prize was awarded to Davide Tamborrini (Biozentrum Basel, Switzerland) for his impressive work in which he determined cryo-EM structures of cardiac sarcomeres. The highly prestigious Feodor Lynen Medal was awarded to Tanja Kortemme (UCSF, USA). Her talk on her award-winning work on *de novo* protein design, including AI-based generation of conformationally switchable proteins (Guo et al. 2025), was awaited with great anticipation and received the deserved attention.

The second day, Friday, March 21, 2025, started with session III (AI-based structure prediction and integrative modelling). It was chaired by Saacnicteh Toledo Patino (OIST, Japan), and began with a talk by Yang Zhang (National University of Singapore), which introduced D-I-Tasser. Further developments in this long-standing structure prediction tool are now surpassing AlphaFold2 and 3, regarding structure prediction accuracy (Zheng et al. 2025).

Jan Kosinski (EMBL Hamburg, Germany) presented the impressive structural work on the nuclear pore complex for which AlphaFold was utilized to reach unprecedented completeness of a structural model of this ultra-large protein complex (Mosalaganti et al. 2022). Additionally, he introduced useful computational tools to enhance structure predictions by AlphaFold based on including experimental input (Kosinski 2024).

Session IV (from protein engineering to AI-based protein design), chaired by Alena Khmelinskaia (LMU Munich, Germany) was opened by Possu Huang (Stanford University, USA) who compared generative models for protein design and the importance of sufficient sampling in these methods (Lu et al. 2025). This was followed by Bruno Correia (EPFL, Switzerland) who examined the structural and functional protein landscape using computational design, which contextualized the latest methods in a forward-looking manner (Khakzad et al. 2023).

As a tradition of the Mosbacher Kolloquium, the Junior GBM is organizing and presenting a session. The Junior GBM consists of undergraduate and graduate students, and has become an integral and important part of this scientific society. As Junior GBM representatives, Frederike Knipp and Marcel Zimmeck introduced this session and Alex Rives (EvolutionaryScale, USA) as the first speaker, who talked about his large language models reflecting evolutionary protein diversity and which have been demonstrated to be extremely useful in protein design (Hayes et al. 2025). Stanislaw Dunin-Horkawicz (MPI, Tübingen) illustrated learning from AlphaFold2 prediction failures (Madaj et al. 2025).

The early career investigator talks by Sophie Binder (University of Bonn, Germany) on binder design for cancer

drug targets, validated via FACS, and Arman Simonyan (University of Copenhagen, Denmark) on ML approaches for peptide probes and drugs show-cased how easily the latest protein design frameworks are introduced into academic early drug discovery research with promising results. Anna Backberg (LMU Munich, Germany) and Kateryna Maksymenko (MPI Tübingen) presented their work on multi-geometry protein scaffold design and training-free clinically relevant binder prediction, respectively.

Friday evening was closed again with anticipated award lectures. The Eduard Buchner Prize was awarded to Dorothee Kern (Scripps, USA), who presented her work on predictive models of protein dynamics by classifying multiple sequence alignments as input for AlphaFold (Wayment-Steele et al. 2024) and a highly innovative ML tool that learns from the absence of experimental signals in NMR datasets to predict protein dynamics on the intermediate exchange time scale (Wayment-Steele et al. 2025).

The Otto Warburg Medal is topically rarely related to the theme of the Colloquium at which the medal is awarded, as was the case this year. Nevertheless, Matthias Hentze (EMBL Heidelberg, Germany) presented his impressive work, which laid the foundation for his pioneering the term and area of research called riboregulation, the post-transcriptional control of gene expression by RNAs (Horos et al. 2019; Huppertz et al. 2022).

The final day, Saturday, March 22, began with the GBM Young Investigator Session. The GBM Young Investigators are yet another sub-group of the GBM in which junior group leaders organize events, where science and topics, relevant to young independent group leaders are discussed. In this session, Andrea Thorn (University of Hamburg, Germany) spoke about the potential of AI applications for experiments in structural biology and the need for high-quality structural data. Following this, Maximilian Fürst (University of Groningen, Netherlands) gave a talk on big data in biochemistry and biophysics in the lab, and how it can be used for further development of AI-based protein design (Korbeld and Fürst 2025).

After the coffee break, session 5 (AI application in life sciences) began, chaired by Horst Lechner (TU Graz, Austria). Judith Zaugg (EMBL Heidelberg, Germany) opened with the presentation of an unsupervised deep learning tool for the analysis of cryo-electron tomography (cryo-ET) data (de Teresa-Trueba et al. 2023). Cryo-ET enables the determination of high-resolution biomolecular structures in a natural, cellular context. Data analysis remains a challenge that can be overcome using ML methods. Jens Meiler (University of Leipzig, Germany) provided insights into how artificial intelligence is revolutionizing protein structure prediction and therapy design, from small molecules to new modalities (Tang et al. 2025). Marharyta Makarova (University of Erlangen, Germany) and Salomé Guilbert (EPFL, Switzerland) presented early career talks about a novel rotamer library for nucleic acids as a basis for the protein design software MUMBO to enable the design of protein-nucleic acid complexes and *in silico* enzyme design for the degradation of short-chain per- and polyfluorinated alkyl substances, respectively. In his lecture on programmable generative protein design Gevorg Grigoryan (Somerville, USA) reflected on developments in protein design from both academic, industrial and medical perspectives, offering a fresh and broad view that rounded off the Colloquium perfectly (Ingraham et al. 2023).

Overall, approximately 470 participants from 18 countries took part in the conference, including 103 bachelor's or master's students, 142 doctoral candidates, 18 retired participants, and 207 with a doctorate. Eight industrial exhibitors supported the meeting. Three poster sessions were held over the course of the event, which were very well attended and in which over 100 posters were presented. A jury evaluated the posters and awarded three with prizes from the GBM and FEBS Open BIO and three further posters received book prizes.

The 76th Mosbacher Kolloquium successfully captured the spirit and urgency of the current wave of AI-driven research in structural biology and protein design. Common challenges were discussed such as standards according to guidelines of the field, the need for quality as well as quantity of data and how to share it, and careful testing of newly emerging computational tools. The development of AI-based tools will be a topic for many years that will focus more on explainable AI and applications for prediction of dynamics of protein structures and complexes. With the second-highest attendance in the history of the Mosbacher Kolloquium, the event resonated strongly with both the national and international scientific community. Attendees praised the conference as refreshing and intellectually stimulating.

Acknowledgments: We would like to gratefully acknowledge the GBM for support and the amazing and professional assistance by the GBM Head Office represented by Prof. Katharina Hieke-Kubatzky, Tino Apel and Ivana Jankovic-Bach.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: All authors wrote the manuscript.

Use of Large Language Models, AI and Machine Learning Tools: None declared.
Conflict of interest: The authors state no conflict of interest.
Research funding: None declared.
Data availability: Not applicable.

References

Backenköhler, M., Groß, J., Wolf, V., and Volkamer, A. (2024). Guided docking as a data generation approach facilitates structure-based machine learning on kinases. *J. Chem. Inf. Model.* 64: 4009–4020.

Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R.J., Milles, L.F., Wicky, B.I.M., Courbet, A., de Haas, R.J., Bethel, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378: 49–56.

de Teresa-Trueba, I., Goetz, S.K., Mattausch, A., Stojanovska, F., Zimmerli, C.E., Toro-Nahuelpan, M., Cheng, D.W.C., Tollervey, F., Pape, C., Beck, M., et al. (2023). Convolutional networks for supervised mining of molecular patterns within cellular context. *Nat. Methods* 20: 284–294.

Ferruz, N. and Höcker, B. (2022). Controllable protein design with language models. *Nature. Mach. Int.* 4: 521–532.

Ferruz, N., Schmidt, S., and Höcker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. *Nat. Commun.* 13: 4348.

Guo, A.B., Akpinaroglu, D., Stephens, C.A., Grabe, M., Smith, C.A., Kelly, M.J.S., and Kortemme, T. (2025). Deep learning-guided design of dynamic proteins. *Science* 388: eadr7094.

Harren, T., Gutermuth, T., Grebner, C., Hessler, G., and Rarey, M. (2024). Modern machine-learning for binding affinity estimation of protein–ligand complexes: progress, opportunities, and challenges. *WIREs Comp. Mol. Sci.* 14: e1716.

Hayes, T., Rao, R., Akin, H., Sofroniew, N.J., Oktay, D., Lin, Z., Verkuil, R., Tran, V.Q., Deaton, J., Wiggert, M., et al. (2025). Simulating 500 million years of evolution with a language model. *Science* 387: 850–858.

Heinzinger, M. and Rost, B. (2025). Teaching AI to speak protein. *Curr. Opin. Struct. Biol.* 91: 102986.

Hennig, J. (2025). Structural biology of RNA and protein-RNA complexes after AlphaFold3. *Chembiochem* 26: e202401047.

Horos, R., Buscher, M., Kleinendorst, R., Alleaume, A.M., Tarafder, A.K., Schwarzl, T., Dziuba, D., Tischer, C., Zielonka, E.M., Adak, A., et al. (2019). The small non-coding vault RNA1-1 acts as a riboregulator of autophagy. *Cell* 176: 1054–1067.

Huppertz, I., Perez-Perri, J.I., Mantas, P., Sekaran, T., Schwarzl, T., Russo, F., Ferring-Appel, D., Koskova, Z., Dimitrova-Paternoga, L., Kafkia, E., et al. (2022). Riboregulation of Enolase 1 activity controls glycolysis and embryonic stem cell differentiation. *Mol. Cell* 82: 2666–2680.

Ingraham, J.B., Baranov, M., Costello, Z., Barber, K.W., Wang, W., Ismail, A., Frappier, V., Lord, D.M., Ng-Thow-Hing, C., Van Vlack, E.R., et al. (2023). Illuminating protein space with a programmable generative model. *Nature* 623: 1070–1078.

Jendrusch, M.A., Yang, A.L.J., Cacace, E., Bobonis, J., Voogdt, C.G.P., Kaspar, S., Schweimer, K., Perez-Borrajero, C., Lapouge, K., Scheurich, J., et al. (2025). Alpha Design: a de novo protein design framework based on Alpha Fold. *Mol. Syst. Biol.* 21: 1166–1189.

Kaden, F., Koch, I., and Selbig, J. (1990). Knowledge-based prediction of protein structures. *J. Theor. Biol.* 147: 85–100.

Khakzad, H., Igashov, I., Schneuing, A., Goverde, C., Bronstein, M., and Correia, B. (2023). A new age in protein design empowered by deep learning. *Cell Syst.* 14: 925–939.

Kim, W., Mirdita, M., Levy Karin, E., Gilchrist, C.L.M., Schweke, H., Söding, J., Levy, E.D., and Steinegger, M. (2025). Rapid and sensitive protein complex alignment with Foldseek-Multimer. *Nat. Methods* 22: 469–472.

Korbeld, K.T. and Fürst, M.J.L.J. (2025). Enriching stabilizing mutations through automated analysis of molecular dynamics simulations using BoostMut, *BioRxiv*, https://doi.org/10.1101/2025.04.29.651183.

Kosinski, J. (2024). Improving Alpha Fold 3 structural modeling by incorporating explicit crosslinks, *BioRxiv*, https://doi.org/10.1101/2024.12.03.626671.

Lau, A.M., Bordin, N., Kandathil, S.M., Sillitoe, I., Waman, V.P., Wells, J., Orengo, C.A., and Jones, D.T. (2024). Exploring structural diversity across the protein universe with the Encyclopedia of Domains. *Science* 386: eadq4946.

Lu, T., Liu, M., Chen, Y., Kim, J., and Huang, P.S. (2025). Assessing generative model coverage of protein structures with SHAPES. *Cell Syst.* 16: 101347.

Madaj, R., Martinez-Goikoetxea, M., Kaminski, K., Ludwiczak, J., and Dunin-Horkawicz, S. (2025). Applicability of AlphaFold2 in the modeling of dimeric, trimeric, and tetrameric coiled-coil domains. *Protein Sci.* 34: e5244.

Mosalaganti, S., Obarska-Kosinska, A., Siggel, M., Taniguchi, R., Turoňová, B., Zimmerli, C.E., Buczak, K., Schmidt, F.H., Margiotta, E., Mackmull, M.T., et al. (2022). AI-based structure prediction empowers integrative structural analysis of human nuclear pores. *Science* 376: eabm9506.

Stocco, F., Artigues-Lleixa, M., Hunklinger, A., Widatalla, T., Guell, M., and Ferruz, N. (2025). Guiding generative protein Language Models with reinforcement learning. *ArXiv*, https://doi.org/10.48550/arXiv.2412.12979.

Tang, Y., Moretti, R., and Meiler, J. (2025). Rosetta AMRLD: a reaction-driven approach for structure-based drug design from combinatorial libraries with Monte Carlo metropolis algorithms. *J. Chem. Inf. Model.* 65: 5945–5959.

Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RF diffusion. *Nature* 620: 1089–1100.

Wayment-Steele, H.K., El Nesr, G., Hettiarachchi, R., Kariyawasam, H., Ovchinnikov, S., and Kern, D. (2025). Learning millisecond protein dynamics from what is missing in NMR spectra, *BioRxiv*, https://doi.org/10.1101/2025.03.19.642801.

Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* 625: 832–839.

Zhang, Z., Ou, C., Cho, Y., Akiyama, Y., and Ovchinnikov, S. (2025). Artificial intelligencemethods for protein folding and design. *Curr. Opin. Struct. Biol.* 93: 103066.

Zheng, W., Wuyun, Q., Li, Y., Liu, Q., Zhou, X., Peng, C., Zhu, Y., Freddolino, L., and Zhang, Y. (2025). Deep-learning-based single-domain and multidomain protein structure prediction with D-I-TASSER. *Nat. Biotechnol.*, https://doi.org/10.1038/s41587-025-02654-4 (Epub ahead of print).