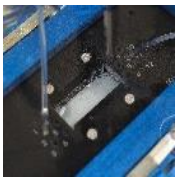
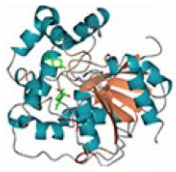
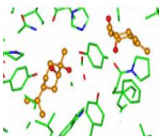
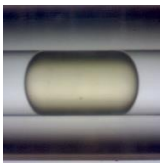

ESAB Webinar

Machine Learning, Computational, AI, IT Tools and Methodologies in Biocatalysis

November 19th 2021, 14.00-16.00 CET

Chairs: Jennifer Littlechild (University of Exeter)
Roland Wohlgemuth (Lodz University of Technology)

PROGRAMME

14.00 Prof. Dame Janet Thornton, European Bioinformatics Institute (EMBL-EBI)
Wellcome Genome Campus Hinxton, Cambridge, CB10 1SD, UK

Implications and Applications of AlphaFold

JM Thornton, JD Tyzack, AJM Ribeiro, Ioannis Riziotis, Neera Borkakoti, Roman Laskowski

At the end of 2020, a significant advance in predicting the structure of a protein from its amino acid sequence was announced by DeepMind, a London-based AI company now part of Google's parent company, Alphabet Inc. DeepMind's AlphaFold 2 program had significantly outperformed other methods in the biennial Critical Assessment of protein Structure Prediction (CASP)[1], producing models of a quality approaching that of experimental determination. AlphaFold 2 has since been published[2] and, more recently, they have released their source code[3]. Over 350,000 protein models from 21 species have been made available on the AlphaFold Protein Structure Database at EMBL-EBI, with tools to view and interrogate the structures[3]. Before the end of the year, DeepMind expect to release models covering UniRef90 – a unique sample of all known protein sequences comprising 130 million proteins. This trove of protein structures has implications for both experimental and computational structural biology and applications for biocatalysis.

Whilst protein structures do not of themselves reveal protein function, they often provide a better understanding of the molecular mechanisms of a protein and in so doing offer insights into how the protein works and how its modulation might lead to a disease or a therapy. Over the past 50 years, protein structures have been an integral part of drug design efforts.

In this talk I will focus on three aspects:

- Overview of AlphaFold predictions – the AlphaFold Protein Structure Database @EMBL-EBI https://alphafold.ebi.ac.uk/
- Predictions for Enzymes
- Potential Applications of the AF structural models for biocatalysis

References

1. CASP14 - 14th Critical Assessment of Techniques for Protein Structure Prediction. https://predictioncenter.org/casp14/

2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

3. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, Bridgland A, Cowie A, Meyer C, Laydon A, Velankar S, Kleywegt GJ, Bateman A, Evans R, Pritzel A, Figurnov M, Ronneberger O, Bates R, Kohl SAA, Potapenko A, Ballard AJ, Romera-Paredes B, Nikolov S, Jain R, Clancy E, Reiman D, Petersen S, Senior AW, Kavukcuoglu K, Birney E, Kohli P, Jumper J, Hassabis D. Highly accurate protein structure prediction for the human proteome. Nature 595, 590-596 (2021).

14.30 Dr. Bas Vroling, Bio-Prodict, Nijmegen, The Netherlands

Advanced sequence representations for machine learning-based protein engineering

Protein engineering is hampered by our limited ability to search through sequence space. High-throughput assays are often not ideal or simply do not exist for certain classes of proteins. Computational approaches that are capable of learning accurate representations of protein fitness landscapes from small training datasets are much sought after. Machine learning algorithms for protein engineering require the conversion of protein sequences into numerical feature vectors. The recent development of protein language models such as UniRep and ProtTrans provide promising ways of capturing general protein features that are suitable for downstream processing. In this talk I will discuss how such pre-trained language models can be applied to convert protein sequences into contextualized feature vectors, and how their performance on subsequent prediction tasks compares to more traditional static sequence representation methods with respect to hit rates and their ability to explore unseen sequence space. In addition I will address how strategies like evotuning and knowledge based hotspot selections can help to increase in silico yields.

15.00 Prof. Dr. Zbynek Prokop, Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, and International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic

Discovery and engineering of biocatalysts powered by microfluidics, molecular modelling and artificial intelligence

The trend in miniaturization and automation of laboratory procedures has emerged in life sciences and biomedicine and led to the development of lab-on-a-chip microfluidic systems. Microfluidic devices can perform thousands of analyzes per second, and in a combinatorial manner, they can analyze reaction volumes ranging from nano- to femtoliters. While the potential of microfluidic methods is already well exploited in nucleic acid analysis and screening of large metagenomic and directed evolution libraries, there is still room for broader applications in high-throughput biochemical characterization. This lecture will focus on droplet microfluidics methods and their utilization in steady-state[1] and transient kinetic analysis providing detailed insight into the mechanisms of enzyme catalysis[2] and stability[3]. The presented examples will show that synergy in the combination of high-throughput microfluidic analysis, modern numerical methods for global data analysis, and molecular modelling provide a deep understanding of the specific biocatalyst property essential for its rational engineering. The perspective potential of combining automated data collection by microfluidics with machine learning methods will also be discussed.[4]

References

1. Vasina M. *et al.*, Functional and Mechanistic Characterization of an Enzyme Family by Integrated Strategy Combining Bioinformatics and Microfluidics, under review.

2. Hess D, Dockalova V, Kokkonen P, Bednar D, Damborsky J, DeMello A, Prokop Z, Stavrakis S. Exploring Mechanism of Enzyme Catalysis by On-Chip Transient Kinetics Coupled with Global Data Analysis and Molecular Modelling. *Chem* 7, 1066-1079 (2021).

3. Yang T. *et al.*, A Microfluidic Temperature Jump Droplet-based Microfluidic Reactor for Rapid Biomolecular Kinetics, under review.

4. Mazurenko S, Prokop Z, Damborsky J. Machine Learning in Enzyme Engineering. ACS Catalysis 10, 1210-1223 (2020).

17.30 Prof. Dr. Jean-Louis Reymond, Department of Chemistry, Biochemistry and Pharmaceutical Sciences, University of Bern, Switzerland https://gdb.unibe.ch

A molecular transformer approach for predicting biotransformations

The use of enzymes for organic synthesis allows for simplified, more economical and selective synthetic routes not accessible to conventional reagents. However, predicting whether a particular molecule might undergo a specific enzyme transformation is very difficult. I will show how we used multi-task transfer learning to specialize the general Molecular Transformer, a sequence-to-sequence machine learning model for general chemistry, to the specific area of biotransformations. Our Enzymatic Transformer model predicts the structure and stereochemistry of enzyme-catalyzed reaction products with remarkable accuracy. One of the key novelties is that we combined the reaction SMILES language with human language tokens describing the enzymes, such that our Enzymatic Transformer not only learned to interpret SMILES, but also the natural language as used by human experts to describe enzymes and their mutations. The Enzymatic Transformer model can be used to predict which product might be formed for any substrate-enzyme pair, or to select a specific enzyme for a biotransformation of choice.

ABOUT THE SPEAKERS

Professor Dame Janet Thornton is a computational biologist, whose early career was rooted in structural biology. She graduated in physics before completing her PhD in Biophysics at the UK National Institute for Medical Research. After postdoctoral research at Oxford University, Janet moved to a fellowship at Birkbeck College, later taking up a joint professorship with University College London. She was Director of EMBL's European Bioinformatics Institute in Hinxton, Cambridge from 2001 - 2015, was awarded a CBE in 2000 and a DBE for services to bio-informatics in 2012.

Janet is a Fellow of the Royal Society, a Fellow of the Academy of Medical Sciences and a member of the European Molecular Biology Organisation as well as a foreign associate of the US National Academy of Sciences.

Today Janet is a senior scientist at EMBL-EBI. Her research group is Focused on proteins, especially their structure, function and evolution. She is a computational biologist, working at the interface of biology with physics, chemistry, and computing. This includes a detailed analysis of enzyme biocatalysts, their mechanisms and the evolution of new catalytic functions; the impacts of genetic variation on protein structure and function and studies on the molecular basis of ageing.

Dr. Bas Vroling is Head of Data Science at Bio-Prodict, Nijmegen, The Netherlands (www.bio-prodict.com). He obtained his Ph.D. at the Center of Molecular and Biomolecular Informatics in Nijmegen, The Netherlands under the guidance of Prof. Dr. Gert Vriend, where he worked on the integration of heterogeneous data in the life sciences. After obtaining his Ph.D. he joined Bio-Prodict and worked on the development of the 3DM protein data integration platform.

Currently, his main focus lies with combining complex protein data with novel machine learning approaches. He leads the development of Helix (helixlabs.ai), a best-in-class pathogenicity predictor for human missense variants, as well as the development of novel tooling to increase the effectiveness of machine learning assisted protein engineering.

ABOUT THE SPEAKERS

Prof Zbynek Prokop is Professor in the Loschmidt Laboratories at Masaryk University, where he leads a research team engaged in the study of fundamental principles of protein chemistry, enzyme mechanism and kinetics. He received his Ph.D. degree in Environmental Chemistry from Masaryk University and extended his expertise during his post-doc and research stays at University of Cambridge, ETH Zurich, TU Wien and the University of Groningen.

His current research interests are the development of advanced biophysical methods for the structural and functional study of proteins, enzyme technology development and applications of microfluidics in life and biomedical sciences. Among the awards and distinctions, he received the Werner von Siemens Award for Excellence in Innovation in 2015 and the Alfred Bader Prize for Bioorganic Chemistry in 2005. He is a co-founder of Enantis Ltd, the first biotechnology spin-out company from Masaryk University operating in the field of biotechnology.

Prof. Jean-Louis Reymond is a Professor of Chemistry at the University of Bern, Switzerland. He studied Chemistry and Biochemistry at the ETH Zürich and obtained his Ph.D. at the University of Lausanne on natural products synthesis (1989). After a Post-Doc and Assistant Professorship at the Scripps Research Institute, he joined the University of Bern (1997). His current research focuses on chemical space exploration methods to design, synthesize, and test bioactive small molecules and peptides, with applications in membrane transporters modulators, kinase inhibitors and antimicrobials (https://gdb.unibe.ch).

He is currently president of the division of medicinal chemistry and chemical biology of the Swiss Chemical Society and serves on the editorial board of RSC Medicinal Chemistry. He is the author of > 300 scientific publications and reviews.

https://orcid.org/0000-0003-2724-2942

NEXT ESAB WEBINARS

ESAB aims to promote the development of Applied Biocatalysis and takes initiatives in areas of growing scientific & industrial interest in the field.

Schedule and Topics of next ESAB webinar:
10th December 2021 Advances in Experimental
14.00-16.00 CET Tools and Methodologies for Screening Enzyme Functions, organized by Roland Wohlgemuth

HOW TO JOIN ESAB

You are cordially invited to join ESAB online *via* https://esabweb.org/Join+us/Application+form.html
Personal membership is free. ESAB, founded in 1980, has the mission of promoting the development of Applied Biocatalysis throughout Europe. The aims of ESAB are to promote initiatives in areas of growing scientific and industrial interest of importance within the field of Applied Biocatalysis. Further information can be found on the ESAB website www.esabweb.org
ESAB - European Society of Applied Biocatalysis (esabweb.org)