elixir CZECH REPUBLIC

ELIXIR CZ Annual Conference 2021

25 – 26 November 2021, Prague, Czech Republic

Organised by ELIXIR CZ

ELIXIR CZ Annual Conference 2021

Organising committee

Jiří Vondrášek, Anna Strachotová,
Natália Pižemová

www.elixir-czech.cz
ISBN 978-80-86241-67-8

No professional language editing of the abstract content was performed.
Scientific and legal responsibility for the abstracts belongs to the authors.

ISBN 978-80-86241-67-8

Summary of content

About ELIXIR CZ infrastructure

The Czech National Infrastructure for Biological Data, abbreviated ELIXIR Czech Republic or ELIXIR CZ, is a distributed research infrastructure for bioinformatics that has arisen from an advanced computational environment. We are dedicated to organisation, storage, sharing and facilitation of interoperability of life science data for further processing and analysis. We respond to the needs of national scientific community, but we are also a proud member of the pan-European infrastructure for biological data ELIXIR, which brings together life science resources from throughout Europe.

ELIXIR CZ is comprised of 14 research performing organisations across the Czech Republic, with its headquarters in the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences.

Institute of Organic Chemistry and Biochemistry of the CAS (IOCB)
Coordinating body of ELIXIR CZ and administrator of computational resources for bioinformatics research. IOCB develops proteomics resources and a database of small molecules, which are the flagships of ELIXIR CZ.

CESNET
CESNTE is one of providers of a large national e-infrastructure for research and development, more specifically, provides communication, computing and storage facilities. CESNET acts as an ambassador of the Czech Republic in GÉANT Project, EGI Federation, and TERENA Association.

Masaryk University: CEITEC, CERIT-SC
CEITEC dedicates its' services to molecular medicine and structural biology, it is also a member of INSTRUCT infrastructure. CERIT-SC is one of providers of an e-infrastructure that provides advanced IT services.

Palacký University Olomouc (UP)
Provider of structural bioinformatics tools. UPOL acts as a liaison point to infrastructure EATRIS.

Charles University (CU)
Developer of tools for diagnostics and prognosis in medicine. UK also provides tools for high-throughput analysis of genomic, proteomic and structural data. UK is active in education and training on aspects of work with biological data.

Institute of Molecular Genetics of the CAS (IMG)

UMG provides DNA and RNA sequence analysis and tools. UMG represents the liaison point to infrastructures INFRAFRONTIER and EU-OPENSCREEN.

Institute of Microbiology of the CAS (IMIC)

MBÚ provides tools for computational biology and bioinformatics as well as models of biological networks.

Institute of Biotechnology of the CAS (IBT)

Provider of bioinformatics tools for structural biology. IBT provides database of DNA structural families.

Biology Centre of the CAS (BC)

BC is dedicated to sequence composition, molecular organisation and evolution of plant genomes and chromosomes.

University of Chemistry and Technology, Prague (UCT)

UCT provides training in the use of tools in cheminformatics and bioinformatics. UCT also provides structural bioinformatics computing tools.

Czech Technical University in Prague – Faculty of Information Technology (CTU)

CTU is dedicated to conceptual modelling and software implementation of conceptual models and development of modelling tools.

University of West Bohemia (UWB)

IT provider for marrow donor analysis and search applications. UWB operates a synthetic biology laboratory that supports tools for efficient assembly protocols and tools for hybrid biochemical reaction simulation.

University of South Bohemia in České Budějovice (USB)

USB represents a genomic centre for plants and microorganisms and applied informatics.

International Clinical Research Center of St. Anne's University Hospital in Brno (FNUSA ICRC)

Developer and provider of novel bioinformatics tools for protein structure analysis and prediction of the effect of mutations on human health.

Scientific Programme

Thursday, 25 November

9:00 - 12:00 Opening get-together, Bowling & Billiard Dejvice

12:00 - 13:00 Registration and lunch, Vila Lanna

13:00 - 13:10 Welcome word by Jiří Vondrášek, Head of ELIXIR CZ

13:10 - 13:40 ELIXIR Czech Republic – achievements and future development
Jiří Vondrášek - ELIXIR Czech Republic, Director

Structural bioinformatics

Chair: Bohdan Schneider, Institute of Biotechnology of the CAS

13:50 - 14:20 Structural bioinformatics in 2021: a few subjective reflections
Bohdan Schneider, Institute of Biotechnology of the CAS

14:20 - 14:30 Visualization of protein structures in 3D, 2D and 1D
Radka Svobodová, Masaryk University

14:30 - 14:40 Simulation of oligosaccharide binding to HEV32 domain
Jan Beránek, University of Chemistry and Technology, Prague

14:40 - 14:50 Machine Learning for Annotating Cavities
Faraneh Haddadi, Masaryk University

14:50 - 15:00 A novel protein stabilization method based on molecular dynamics and force-field-based energy evaluation
Jana Horáčková, St. Anne’s University Hospital Brno – ICRC

15:00 - 15:10 Use of diNucleotide conformational classes, NtC, for refinement of DNA crystal structures
Jakub Svoboda Institute of Biotechnology of the CAS

15:10 - 15:20 WATlas (watlas.datmos.org), an online atlas of biomolecular hydration
Lada Biedermannová, Institute of Biotechnology of the CAS

15:20 - 15:40 20´ discussion

Thursday, 25 November

Genomics

Chair: Jan Pačes, Institute of Molecular Genetics of the Czech Academy of Sciences

16:00 - 16:30 Molecular surveillance of SARS-CoV-2 in CR
Jan Pačes, Institute of Molecular Genetics of the CAS

16:30 - 16:45 Is our model wrong? Identifying outliers in differential expression analysis
Martin Modrák, Institute of Microbiology of the CAS

16:45 - 17:00 Scdrake: a highly scalable and reproducible pipeline for scRNA-seq data
Jiří Novotný, Institute of Molecular Genetics of the CAS

17:00 - 17:15 Paperfly: a tool for analysis of ChIP-seq or similar sequencing data without a reference genome
Kateřina Faltejsková, Institute of Organic Chemistry and Biochemistry of the CAS

17:15 - 17:30 Sequencing and analysis of the supernumerary maize chromosome
Jan Bartoš, Institute of Experimental Botany of the CAS

17:30 - 17:50 20´ discussion

Parallel breakout sessions

18:00 - 18:45 ELIXIR Community 3D-Bioinfo
Bohdan Schneider, Institute of Biotechnology of the CAS

18:00 - 18:45 Building an effective and sustainable ELIXIR Node: the role of communications, collaborations, impact and funding
Andrew Smith, ELIXIR Hub

18:45 - 19:30 Dinner

19:30 - 21:00 Poster session

Friday, 26 November

Chemical biology

Chair: Jiří Vondrášek, Institute of Organic Chemistry and Biochemistry
of the Czech Academy of Sciences

9:00 - 9:30	Chemical biology in ELIXIR CZ Karel Berka, Palacký University Olomouc
9:30 - 9:45	ASAFind 2.0: Multi-class predictions of intracellular locations of proteins in organisms with complex plastids Ansgar Gruber, Biology Centre of the CAS
9:45 - 10:00	Charge Transport on Biomolecular Interfaces with Metal Electrodes Zdeněk Futera, University of South Bohemia in České Budějovice
10:00 - 10:15	In silico design of synthetically feasible compounds Pavel Polishchuk, Palacký University Olomouc
10:15 - 10:30	Integrated Database of Small Molecules Jakub Galgonek, Institute of Organic Chemistry and Biochemistry of the CAS
10:30 - 10:50	20´ discussion

Friday, 26 November

Bioinformatics Data Management and Access Control
Chair: Robert Pergl, Czech Technical University in Prague

11:10 - 11:40 Requirements and Interactions in Life Science Research Data Management
Korbinian Bösl, ELIXIR Norway

11:40 - 11:55 Data Stewardship Wizard: Towards the Cutting-Edge Collaborative Online Platform for Data Management Plans
Jan Slifka, Czech Technical University in Prague / Data Stewardship Wizard

11:55 - 12:10 DSW as a FAIR Data Entry Tool
Marek Suchánek, Czech Technical University in Prague / Data Stewardship Wizard

12:10 - 12:25 Distributed authorization using GA4GH passports
Dominik F. Bučík, Masaryk University

12:25 - 12:40 FAIR data portal
Milos Prokysek, University of South Bohemia in České Budějovice

12:40 - 13:00 20´ discussion

13:00 Farewell

Lectures

ELIXIR Czech Republic – achievements and future development

Jiří Vondrášek - ELIXIR Czech Republic, Director

ELIXIR Czech Republic is an internationally recognised infrastructure for biological data, which has an indisputable impact on the life science disciplines in these times of “Data deluge”. During its 6 years of existence, ELIXIR CZ has become an important player in shaping the biological data environment and its development at national as well as at international level. ELIXIR CZ is proud to have recently received an excellent evaluation result given by an international evaluation panel, which reflects the continuous efforts to provide top-level solutions for life science research. ELIXIR CZ follows its' scientific strategy and scientific programme which ensures further development and growth in delineated areas of excellence. On the other hand, there are still data management issues that the infrastructure should address as a matter of priority in the following years. How the data management will be implemented in heterogeneous environment to guarantee the interoperability of data and resources across health, medicine, and biology is a major challenge currently being addressed by ELIXIR CZ. To follow technological as well as computational progress, the infrastructure is ready to invest in human resources and education to able to respond the demands of the scientific community.

Structural bioinformatics in 2021: a few subjective reflections

Bohdan Schneider - Institute of Biotechnology of the Czech Academy of Sciences

Structural bioinformatics strives to discover general laws governing the structural behavior of biomolecular. To achieve the goal it deals with large volumes of structural data. The subject and methods of 3D bioinformatics are distinct from (general/sequence) bioinformatics, but both fields overlap and develop in synergy. 3D bioinformatics is also distinct from molecular modeling but it contributes to its development. In 3D bioinformatics, most attention has been historically paid to rules governing the behavior of protein molecules and their interactions with other molecules, “ligands”. The recent emergence of the AlphaFold prediction software based on machine learning has changed the landscape of protein 3D bioinformatics. This fact was reflected during a recent annual conference of the 3D-Bioinfo, an ELIXIR community, held November 2-4, 2021. Most talks in four of the five sessions of the conference dealt with AlphaFold and its consequences for database archives, discussed ways how to predict biologically relevant interactions, and examined the service AlphaFold can pay to protein engineering. Has AlphaFold changed “the paradigm” of 3D bioinformatics as some suggest? In my opinion not, it has just shifted it. We should in no case forget that scientists, not software, must determine the ultimate goals of the field, set the benchmarks, ontologies, database content and schemas, and most of all, critically evaluate as many predicted structures by their experimental mates as possible.

My research concentrates on nucleic acid structures and a part of my talk will be devoted to 3D bioinformatics of these molecules. Methods of their annotation, modeling, refinement, and validation attract much less attention than these methods for proteins but important new tools are being developed. We have developed a general annotation and validation protocol for nucleic acid structures; worldwide, several groups attempt to predict nucleic acid, mostly RNA, structures. My impression is that 3D-Bioinfo as an ELIXIR community should serve as a natural hub for at least some of these activities and together with RNA Puzzles, important community-wide competition, to integrate the field of RNA structural bioinformatics.

Visualization of protein structures in 3D, 2D and 1D

Radka Svobodová, Karel Berka, David Sehnal, Adam Midlik, Vladimír Horský, Aliaksei Chareshneu, Jaroslav Koča - NCBR and CEITEC, Masaryk University Department of Physical chemistry, Palacky University Olomouc

Biomacromolecular structural data is a highly valuable and scientifically vital resource that provides a mechanistic understanding of biological systems. Currently, more than 180,000 experimentally determined three-dimensional structures of biological macromolecules are available from the open-access Protein Data Bank (PDB). The PDB archive continues to grow both in terms of the number of new entries and also in the complexity and size of the structures deposited in the PDB. Furthermore, these data are accompanied by large and increasing amounts of experimental data. Additionally, the macromolecular data are enriched with value-added annotations describing their biological, physicochemical and structural properties. Last but not least, the structural data are divided into protein families and most of them are described by a rich dataset of structures originating from various organisms, having different mutations and binding various ligands.

Today, the scientific community requires fast and fully interactive web visualizations to exploit this complex structural information. Moreover, it is essential to show also experimental data and annotations. In parallel, a very useful insight into the protein structure can be obtained by protein 2D diagrams, showing a composition of secondary structure elements. These 2D diagrams can also depict a secondary structure arrangement in a whole protein family. Additionally, 1D view of proteins and protein families allows a researcher to get an illustrative overview of the protein or protein family structure.

We developed web visualization tools MolStar [1], 2DProts [2, 3] and OverProt [4], which provide 3D, 2D and 1D view of the proteins and protein families. We will show these tools and also their mutual cooperation.

References:

[1] Sehnal D., Bittrich S., Deshpande M., Svobodová R., Berka K., Bazgier V., Velankar S., Burley S.K., Koča J. and Rose A.S., 2021. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. Nucleic Acids Research.

[2] Sillitoe I. et al., 2021. CATH: increased structural coverage of functional space. Nucleic acids research, 49(D1), pp.D266-D273.

[3] Hutařová I, Hutař J., Midlik A., Horský V., Hladká E., Svobodová R. and Berka K., 2021. 2DProts: Database of Family-Wide Protein Secondary Structure Diagrams. Bioinformatics.

[4] https://overprot.ncbr.muni.cz/home

Simulation of oligosaccharide binding to HEV32 domain

Jan Beránek - University of Chemistry and Technology, Prague

HEV32 is a 32-residue domain of protein hevein known for its ability to bind chitin-based polysaccharides. HEV32 is a suitable model system for studying protein-saccharide interactions. We studied interaction of HEV32 domain with mono-, di- and trisaccharides derived from chitin using well-tempered funnel metadynamics. For all three systems, 2 µs long simulation was carried out and free energy surface was obtained. Standard binding free energy of all saccharide molecules to HEV32 were determined, notably the obtained binding energy of the trisaccharide was ≈-24 kJ/mol, corresponding with the value deterimined experimentally.

Machine Learning for Annotating Cavities

Faraneh Haddadi, Ondrej Vavra, David Bednar, Stanislav Mazurenko - Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic, International Clinical Research Center, St. Anne’s University Hospital Brno, Czech Republic

Certain structural elements play a critical role in binding and unbinding ligands. In this study, we aim to predict the types of these structural elements in proteins that could play a role in binding and unbinding ligands. The main three types of these features are surface binding interface, tunnel with one opening that connects the active site with the outside environment, or channel open on both sides. There are several methods for tunnel and channel calculation, but they require customized settings to produce high-quality results. In particular, the discrimination between surface and buried binding pockets is critical for tunnel calculation but is an open problem. Our study aims to solve this problem using FPOCKET features and new features, e.g., exposed ratio and ligand coverage. We utilized the machine learning algorithms (SVM, KNN, ANN, and Random Forest) for the binary prediction or three-class prediction. Our results show the cross-validation accuracy of 72% and 60%, respectively, indicating the promising potential of machine learning algorithms for solving this task. We plan to leverage this potential with the data extracted from molecular dynamics simulations via advanced Artificial Intelligence methods in the future.

Acknowledgements: We would like to express our thanks to ELIXIR (LM2018131)
and e-INFRA (LM2018140) infrastructures for the financial support.

A novel protein stabilization method based on molecular dynamics and force-field-based energy evaluation

Jana Horáčková - FNUSA-ICRC and Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Brno

Protein stabilization is crucial for both basic research and the applicability of proteins in biotechnology and biomedicine. While experimental stabilization techniques are costly and time-consuming, it is more suitable to design stable protein variants *in silico* and characterize only the top candidates. Current state-of-the-art stability prediction tools (e.g., Rosetta and FoldX) reach a maximum precision of 0.76 and 0.67, respectively[1],leaving plenty of room for improvement.

Recently, Caldararu et al.2 have demonstrated that stability prediction tools tend to be overly influenced by the selected structure, leading to substantial errors in the prediction. As a solution, they suggested using several slightly different structures of the same protein and averaging the results.

Therefore, we developed a method that addresses the sensitivity to the choice of the structure by introducing protein dynamics into the prediction. According to preliminary results, this method exceeds the reliability of Rosetta and FoldX with a significantly improved precision of 0.85. The workflow involves performing molecular dynamics simulation of the protein, extracting an ensemble of protein conformations from the simulation, evaluating stability by FoldX on each snapshot, and finally, obtaining the final prediction by statistical analysis of the calculated values.

The ultimate goal is to fully automate this method and implement it in a new version of FireProt[3], a web server for automated design of thermostable proteins, available at https://loschmidt.chemi.muni.cz/fireprotweb/.

References

1. Bednar D, Beerens K, Sebestova E, et al. FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. PLOS Computational Biology. 2015;11(11):e1004556. doi:10.1371/journal.pcbi.1004556
2. Caldararu O, Blundell TL, Kepp KP. A base measure of precision for protein stability predictors: structural sensitivity. BMC Bioinformatics. 2021;22(1):88. doi:10.1186/s12859-021-04030-w
3. Musil M, Stourac J, Bendl J, et al. FireProt: web server for automated design of thermostable proteins. Nucleic Acids Res. 2017;45(Web Server issue):W393-W399.doi:10.1093/nar/gkx285

Acknowledgment

Computational resources were provided by the ELIXIR-CZ project (LM2018131), part of the international ELIXIR infrastructure, and by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

Use of diNucleotide conformational classes, NtC, for refinement of DNA crystal structures

Jakub Svoboda - Institute of Biotechnology of the Czech Academy of Sciences

DNA crystal structure refinement benefits greatly from the inhouse developed diNucleotide conformational classes (NtCs). In the past, the usefulness of NtCs was described to potentially improve structures from the PDB archive (Černý et al. 2020) as well as during the refinement stage of new structures (Kolenko et al. 2020).

Structure is uploaded to the DNATCO (https://dnatco.datmos.org) website and automated protocol generates the restraint file. The file describes the target geometry based on the NtC standards. A well-chosen target leads to a better convergence to the best agreement with the experimental data. Parts of nucleic acid structures that are not assigned to any defined NtC class can be still fitted into defined conformation.

Herein, we present a practical application of the above described methodology to refinement of successfully crystallized 18-mer DNA oligonucleotides. They serve as a scaffold for various base pairing combinations in the middle of the sequence. Application of the NtC-driven refinement restraints lead to an improvement of several crystallographic indicators. Therefore, we demonstrated the flexibility and usability of NtC classes to solve nucleic acid crystal structures.

WATlas (watlas.datmos.org), an online atlas of biomolecular hydration

Lada Biedermannová, Ph.D. - Institute of Biotechnology, CAS

We will present WATlas (watlas.datmos.org), an online atlas of biomolecular hydration. We will focus on a new feature of the atlas, the recently developed hydrated dinucleotide blocs, obtained through a detailed analysis of over 5000 high resolution X-ray structures of naked DNA and protein/DNA complexes. The analysis is based on our formulation of a universal set of dinucleotide conformer classes, termed NtC (nucleotide conformer). The hydrated NtC blocks reveal differences in hydration of different DNA conformations and sequences, and can be applied for the prediction of hydration structure of DNA molecules of A-, B-, Z- and mixed forms.

Molecular surveillance of SARS-CoV-2 in CR

Jan Pačes - Institute of Molecular Genetics of the ASCR, v. v. i.

The novel coronavius, SARS-CoV-2, has infected more than 250 million people worldwide within first two years of pandemic. The disease it causes, COVID-19, is a systemic inflammation involving multiple organs, affecting all age groups, with high mortality rate, severe adverse outcomes, and high economic burden. With death toll surpassing 5 million people makes COVID-19 one of the biggest thread in our times. During these two years scientists gathered enormous amount of data about the virus, mechanism of infection and pandemic dynamics. One of the big challenges is how to store, share and make available all data in relatively short time, so the information can be used to defend the pandemy. Also, SARS-CoV-2, like other viruses, is mutating and optimizing to its new host, human. In recent months, we have seen the emergence and rapid spread of new SARS-CoV-2 variants with increased transmissibility.

We established informal group of scientists from many fields and from many academic institutions and hospitals called COG-CZ (COronavirus Genomics in CZ) in order to introduce regular monitoring of SARS-CoV-2 variants on territory of the Czech Republic. In this talk we are going to present our effort, reports for the scientists we generate and overall situation with coronavirus and its variants in CR and how ELIXIR-CZ infrastructure is involved.

Is our model wrong? Identifying outliers in differential expression analysis

Martin Modrák - Institute of Microbiology of the Czech Academy of Sciences

Most current tools to analyze differential expression assume that the biological and technical variability within an experimental condition is well described by a negative binomial distribution, but this assumption is rarely actually tested. The `ppcseq` package aims to detect outliers, i.e. read counts that are highly implausible under the negative binomial assumption and thus can have outsized influence on the results of a differential expression analysis. The talk will use `ppcseq` primarily as an example of *posterior predictive checks* - a general approach for validating the fit between a statistical model and a dataset. While posterior predictive checks are naturally connected with the Bayesian paradigm for data analysis, I will also describe how it can be adapted to a useful heuristic in the frequentist context.

The package can be found at https://github.com/stemangiola/ppcSeq

scdrake: a highly scalable and reproducible pipeline for scRNA-seq data

Jiří Novotný - Institute of Molecular Genetics of the Czech Academy of Sciences

Single-cell RNA-seq (scRNA-seq) is an emerging technology that is able to capture the transcriptional profiles of thousands of individual cells. This gives an excellent opportunity to study complex cellular subpopulations, states, or lineages.

However, scRNA-seq tends to produce noisy data requiring more complex analyses compared to traditional bulk RNA-seq. Although there are several pipelines with a graphical user interface, to our best knowledge, there is currently no configurable pipeline with a command-line interface, suitable for bioinformaticians.

Here, we present scdrake, a highly scalable, reproducible and configurable pipeline for scRNA-seq data prepared by a popular 10x Genomics droplet-based technology. Scdrake is implemented in the R language and is built on top of the drake package, a Make-like pipeline toolkit. Scdrake currently provides common steps of scRNA-seq data analysis: quality control and filtering of cells and genes, normalization, dimensionality reduction, clustering, finding of cluster markers and differentially expressed genes between clusters, and integration of multiple datasets. All pipeline steps are accompanied by rich graphical outputs and reports in HTML format.

Thanks to the drake package, all intermediate results can be reused, and the pipeline can be easily extended by users to incorporate custom analyses. Also, drake analyzes which parts of the pipeline are already done or haven't changed since the last run, and which can be run in parallel, resulting in great execution speed.

Scdrake's code and extensive documentation can be found
at github.com/bioinfocz/scdrake

Paperfly: a tool for analysis of ChIP-seq or similar sequencing data without a reference genome

Kateřina Faltejsková - Institute of Organic Chemistry and Biochemistry of the CAS

The available methods of studying gene expression regulation by experiments based on sequencing are limited only to certain organisms by the need to know its reference genome. So far, the reference genome was essential to be able to perform the subsequent computational analysis. To overcome this and therefore to extended the array of organisms that can be studied, we present Paperfly. This tool uses genome assembly algorithms to reconstruct the areas captured during the experiment and constructs an alignment of the seen sequences while taking into account the number of seen sequence segments.

Sequencing and analysis of the supernumerary maize chromosome

Jan Bartoš - Institute of Experimental Botany of the Czech Academy of Sciences

B chromosomes are dispensable, supernumerary and “selfish” genetic elements found in representatives of plants, animals and fungi. In a population, they are present only in some individuals and in several species they are selectively eliminated from specific tissues and organs. Interestingly, they do not follow rules of mendelian inheritance and can accumulate through a process called non-disjunction. The B chromosome in maize is one of the first discovered (a hundred years ago) and most studied.

We combined available genomics tools to sequence and assemble the maize B chromosome. With a pseudomolecule of 107 Mb, it is the first plant B chromosome ever assembled to chromosome-scale. A high-quality sequence enabled a detailed analysis of repetitive elements as well as its gene content. 15,145 distinct transposable elements were equally dispersed along the B chromosome with the exception of the region close to the centromere, where tandem repeats predominate.

While it had been believed for decades that B chromosomes lack gene sequences, annotation of the maize B chromosome with the Maker pipeline identified more than seven hundred protein-coding genes. The analysis of the closest orthologs/paralogs of B-localized genes in the genomes of maize and *Sorghum bicolor* indicates a relaxation of purifying selection on the B chromosome. Further, any synteny of genes from a potential progenitor chromosome cannot be recognized. Gene Ontology analysis indicated a prevalence of GO-terms linked to B-chromosome behavior among those over-represented on the B chromosome compared to the A-chromosome complement. Finally, sequencing of B-A translocations allowed a narrowing down number of candidates for a “distal element” involved in non-disjunction to 34 genes.

ELIXIR Community 3D-Bioinfo

Bohdan Schneider - Institute of Biotechnology of the Czech Academy of Sciences

3D-Bionfo is a thriving ELIXIR community. Its annual meeting, which occurred November 2-4, 2021, had over 300 registered participants. Each of the five 3D-Bioinfo activities had its own two-hour session organized by the activity coordinators. 3D-Bioinfo also organizes webinars, the next one is scheduled for December 14 at 3 PM. People interested in active participation in 3D-Bioinfo can contact me, any of the other four coordinators, or the community Head, Christine Orengo; very effective communication is with new ELIXIR representative, Katharine Heil. Contacts and other information about 3D-Bioinfo actions can be found at the ELIXIR website,
https://elixir-europe.org/communities/3d-bioinfo.

Building an effective and sustainable ELIXIR Node: the role of communications, collaborations, impact and funding

Andrew Smith, ELIXIR Hub

Building an effective ELIXIR Nodes that meets the needs of of various stakeholders is one of the most important requirements in ensuring long-term sustainability. In this breakout session, Andrew Smith will present his perspective and suggestions of how ELIXIR Nodes can best meet the needs of their various stakeholders: users, partners and funders. The interactive session will allow for discussions on a range of aspects including funding opportunities, business models, demonstrating impact, building effective collaborations and communications.

Chemical biology in ELIXIR CZ

Karel Berka - Palacký University Olomouc

Chemical entities of biological interest – such as small molecule metabolites – are essential building blocks of biological systems with crucial roles in human health and disease. The chemoinformatics characterization of small molecules (e.g., prediction of their key properties and calculation of their descriptors) enhances their utilization in other molecular biology fields.

ELIXIR CZ strategy in this field focuses on (i) the development of federated services at a global scale; (ii) the classification of chemical entities; and (iii) the integration of chemical biology resources.

In the lecture, we will walk through the key domains and services ELIXIR CZ developed and is offering in this area.

ASAFind 2.0: Multi-class predictions of intracellular locations of proteins in organisms with complex plastids

Ansgar Gruber - Biology Centre, Institute of Parasitology, Czech Academy of Sciences

Diatom plastids evolved by eukaryote-eukaryote endosymbiosis. This process led to a complex plastid ultrastructure, with a total of four membranes surrounding the stroma. The two innermost membranes correspond to the outer and inner envelope of primary plastids found in Archaeplastida. The second membrane from the outside (third from the inside) is considered to correspond to the former plasma membrane of the endosymbiont. Hence, the space between this second and third plastid membranes, the periplastidic compartment (PPC), is a remnant of the cytosol of the former endosymbiont.

Cell biological processes as well as metabolic reactions have been shown to take place in this compartment, however, genome wide predictions of the proteins targeted to this compartment were so far based on manual annotation work exclusively.

With the increase of published experimental data, this situation has changed. Using published experimental protein localizations as reference data, at least a subset of the PPC proteins can be predicted from genome data with high specificity. This method of PPC protein prediction, was included as a new feature in an updated version of the plastid protein predictor ASAFind. The new ASAFind version also accepts the output of the most recent versions of SignalP (5.0) and TargetP (2.0) as input data. Furthermore, a script to calculate custom scoring matrices, that can be used for predictions in a simplified score cut-off mode is included. This allows for adjustments of the method to other groups of algae.

Charge Transport on Biomolecular Interfaces with Metal Electrodes

Zdenek Futera - Faculty of Science, University of South Bohemia in Ceske Budejovice

The rapid development of nanobiotechnologies and experimental techniques able to probe the conductive properties of single-molecular junctions recently brought attention to the utilization of biomolecules in such devices. The conductance of the whole proteins connected to metallic contacts has been measured, their temperature dependencies investigated, and the obtained peculiar results raised many questions about undergoing charge transport mechanism.

While the proteins typically transfer charges by incoherent hopping mechanism in their native environment, coherent tunneling seems to be responsible for charge transport through the protein junctions between metallic contacts. However, how can the electron delocalize over a several-nanometer-long soft biomatter? How important is the contact with the metal? Are the metallic cations incorporated in metalloproteins affecting the transport? And what role plays the protein matrix in such processes?

We study the protein interactions with metals and their electronic properties employing computational simulations based on classical molecular dynamics (MD) and density functional theory (DFT) techniques. Structural and electronic aspects of blue-copper protein Azurin and small tetraheme cytochrome (STC), their interactions with gold electrodes, and conductivities will be discussed together with the effects of external electric fields.

In silico design of synthetically feasible compounds

Pavel Polishchuk - Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc

The chemical universe is extremely large that makes systematic enumeration of compounds and their virtual screening not efficient. De novo design proposes a reasonable alternative. These approaches adaptively explore chemical space by generating compounds which satisfy given activity/property constrains. The main issue of computationally generated compounds is their synthetic accessibility. Reaction-based generation approaches explicitly address this issue, but they have limited coverage of chemical space and are less flexible in making moves within the search space. Fragment-based approaches provide greater flexibility to structural modifications (growing, mutation or linking of molecules and fragments) but suffer from difficulties to control synthetic feasibility of generated compounds. We developed and implemented the framework of chemically reasonable mutations (CReM) which makes structural changes taking into account chemical context of fragments. This results in always chemically valid structures and greatly increases control over synthetic feasibility of generated compounds. We will demonstrate applicability of CReM to enumeration of analog series and scaffold decoration to support exploration of local SAR, hit expansion and lead optimization, de novo design using molecular docking and pharmacophores.

Acknowledgements: this research was funded by the Ministry of Education, Youth and Sports of the Czech Republic within the INTER-EXCELLENCE LTARF18013 project, the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868) and ELIXIR CZ research infrastructure project (MEYS Grant No: LM2018131).

Integrated Database of Small Molecules

Jakub Galgonek - Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences

Exporting data in the Resource Description Framework (RDF) significantly increases their interoperability and usability. To query these data, the SPARQL query language was introduced. Among other useful features, SPARQL supports federated queries that combine multiple independent data source endpoints. This allows users to obtain insights that are not possible using only a single data source. For these reasons, many biological and chemical databases present their data in RDF, and support SPARQL querying. In our project, we primary focused on PubChem, ChEMBL and ChEBI small-molecule datasets. These datasets are already being exported to RDF by their creators. However, none of them has an official and currently supported SPARQL endpoint. This omission makes it difficult to construct complex or federated queries that could access all of the datasets, thus underutilising the main advantage of the availability of RDF data. Our project addresses this gap by integrating the datasets into one database called the Integrated Database of Small Molecules (IDSM) that is accessible through a SPARQL endpoint. In the presentation, it will be demonstrated how the service can be utilized by other services and how federated queries can be used to solve complex tasks.

Requirements and Interactions in Life Science Research Data Management

Korbinian Bösl - ELIXIR Norway

Research data management (RDM) must fulfill requirements by funders and institutions and at the same time meet the needs of the scientists. It is important that this wide span of interests is accommodated by the tools and infrastructure framework provided for research data management.

Well planned management of research data will help enable researchers to identify suitable instruments at each stage of the research data lifecycle, from project planning, over data generation or reuse of data and data analysis, to sharing of FAIR data. To add complexity, the different tasks in a project may be fulfilled by different contributors with distinct expertises, roles, rights, and responsibilities. RDM tools for wider audiences need to model these interactions and have to be sufficiently flexible to accommodate different life science subdomains and their (meta)data standards.

The toolkit around the Norwegian e-Infrastructure for Life Sciences (NeLS) provides an example for an integrated solution for research data management of life science projects. It combines the Data Stewardship Wizard together with FAIRDOM SEEK, Galaxy and its own storage components. The potential of such an integrated solution and possible future developments will be discussed.

Data Stewardship Wizard: Towards the Cutting-Edge Collaborative Online Platform for Data Management Plans

Jan Slifka - Czech Technical University in Prague / Data Stewardship Wizard

Data Stewardship Wizard (DSW) is a tool for creating and collaborating on the data management plans (DMPs) used worldwide by thousands of people of various backgrounds and expertise. We understand that working on scientific experiments and related DMPs is teamwork. Therefore, we developed a full range of features transforming DSW into a cutting-edge collaborative online platform during the last year. Besides other features, we introduced: (i) A live online collaboration on DMPs so that multiple researchers can work together at the same time. (ii) Complex sharing optionsto set up who to share the project with and how. (iii) Comments and editor notes to discuss the questions with the team or to collect feedback. (iv) Version history that allows seeing who and when made which change, naming the specific versions, or coming back to any point in the project history. In this flash talk, we present the new collaborative DSW features and ideas on how to use them to transform collaboration on DMPs to the next level.

DSW as a FAIR Data Entry Tool

Marek Suchánek - Czech Technical University in Prague / Data Stewardship Wizard

Data Stewardship Wizard (DSW) is widely used as a tool for data management planning; however, thanks to its versatility, DSW can be turned into a data entry tool and easily integrated with various workflows. It has already proven its capabilities as a FAIR Data Entry tool in two scenarios: (i) As an integral part of the VODAN-in-a-Box solution, it is used as an entry tool for electronic Case Report Forms with the ability to transform to RDF based on a semantic model, store them in a triple store, and store related metadata in a designated FAIR Data Point. (ii) It provides a way to create FAIR Implementation Profiles (so-called FIP Wizard), where FAIR Communities can capture their implementation choices and even publish their FIP as a nanopublication. DSW is continually enhanced and improves user experience and versatility (template development, submission services, UI translations, etc.). We foresee further non-DMP use cases for DSW to provide an efficient way as a (FAIR) data entry tool.

Distributed authorization using GA4GH passports

Dominik F. Bučík - Masaryk University

ELIXIR AAI (Authentication and authorization infrastructure) delivers a complex solution for authentication, authorization, and identity management. Among other things, we also explore cutting-edge solutions and participate in the standardization processes. One of such examples is the work on a distributed authorization, where the collaboration with the Global Alliance for Genomics and Health (GA4GH) led to a new standard in the form of GA4GH Passports and Visas.

The GA4GH Passports and Passport Visas provide a convenient and standardized way of communicating users' data access authorizations based on either their role (e.g. being a researcher), affiliation, or access status. Each piece of such information is represented as a separate Passport Visa, while their distributed nature allows gathering them from multiple Assertion Sources and combining them into a single GA4GH Passport by a Passport Broker. Services consuming the Passports are called the Passport Clearinghouses. They are able to verify signatures on individual Passport Visas and decide on an individual basis which ones are trusted for the given user. As a result, services have fine-grained control over the data access.

ELIXIR AAI has been one of the first implementers of the GA4GH passports, displaying the capabilities of the research community among giants like Google Cloud. From a technical perspective, it acts as both Passport Broker and Assertion source service, which means it collects Passports and Visas from different sources, enriches the set with additional Visas derived from ELIXIR AAI internal data, and then releases all this information to the downstream relying parties as a single GA4GH Passport. This whole process happens in a well-established AAI environment, taking advantage of the capabilities of the OpenID Connect protocol, utilizing a federated environment.

While the standard is finished and the features are available to be used, it has not been left to become outdated. The upcoming update should reflect the new requirements coming from the community. This talk should encourage you in consideration of adopting the standard, as well as providing a place for expressing rising needs in further development.

FAIR data portal

Milos Prokysek - University of South Bohemia in České Budějovice

The FAIR data portal is a tool for publishing research data in FAIR-way, including machine- and human-readable interface, fully based on RDF data description in JSON-LD format. The tool was developed as a natural extension of the UniCatDB database developed at the University of South Bohemia. The tool is able to publish any type of data with proper metadata descriptions. The tool is still under development.

Posters

List of posters

1. **Cytogenetic verification of the scaffolds of P. sativum centromere 6 and nuclear architecture of its components** :
   Laura Ávila Robledillo - Biology Centre CAS
2. **CoverView: straightforward visualization of electron density coverage of PDB structures at the level of atoms to help validation** :
   Aliaksei Chareshneu - Masaryk University
3. **Ancient mtDNA database (AmtDB.org): an update** :
   Edvard Ehler - Institute of Molecular Genetics of the ASCR, v. v. i.
4. **Predicting ion binding sites in proteins using machine learning** :
   Christos Feidakis - Charles University
5. **Phylogeny of bacterial Hsp70** :
   Michal Gala - Pavol Jozef Šafárik University in Košice
6. **Machine Learning for Annotating Cavities** :
   Faraneh Haddadi - Masaryk University
7. **Regulatory function discovery using combined omics and sequence analysis** :
   Tomáš Honzík - University of West Bohemia
8. **FireProtASR - FULLY AUTOMATED ANCESTRAL SEQUENCE RECONSTRUCTION** :
   Rayyan Tariq Khan - University Hospital at Saint Anna in Brno, FNUSA – ICRC
9. **Increasing Effectiveness of Data Management Planning: DSW State-of-the-Art Features** :
   Vojtech Knaisl - Czech Technical University in Prague
10. **CaverWeb 2.0 – Identification of the protein tunnels in trajectories from molecular dynamics** :
    Petr Kohout - Loschmidt Laboratories
11. **Data analysis for MICA / MICB genes identification** :
    Kateřina Kratochvílová - University of West Bohemia

12. **The Chicken-and-Egg Problem of Landmark-Driven Molecular Dynamics: Are Random Landmarks Useful?** :
    Aleš Křenek - Masaryk University
13. **Signal-to-noise limitations of genetic toggle switches** :
    Lukas Kuhajda - University of West Bohemia
14. **OverProt: Overview of Secondary Structure Consensus in Protein Families** :
    Adam Midlik - Masaryk University
15. **Exosomes produced by melanoma cells significantly influence the biological properties of normal and cancer-associated fibroblasts** :
    Lucie Pfeiferova - Institute of Molecular Genetics of the ASCR, v. v. i.
16. **QM-like partial atomic charges for proteins available online** :
    Ondřej Schindler - Masaryk University
17. **Combining and visualizing experimental data and annotation features on S. coelicolor genome for biologists purposes** :
    Marek Schwarz - Institute of Microbiology of the CAS, v. v. i.
18. **MolMeDB – Molecules on Membranes Database** :
    Kateřina Storchmannová - Palacký University Olomouc
19. **Analysis of sequencing data from reprogramming of immortalized cell line** :
    Petra Svatoňová - Institute of Molecular Genetics of the ASCR, v. v. i.
20. **GlobalFungi – a global database of fungal occurrences** :
    Tomáš Větrovský - Institute of Microbiology of the CAS, v. v. i.
21. **Sequence organization of CENH3-associated heterochromatic loci on the holocentric chromosomes of Cuscuta europaea** :
    Tihana Vondrak - Biology Centre CAS

Cytogenetic verification of the scaffolds of *P. sativum* centromere 6 and nuclear architecture of its components.

Laura Avila Robledillo - Biology Centre, Czech Academy of Sciences, České Budějovice, Czech Republic

In most plant species, functional centromeres are marked by the presence of the histone variant CENH3. *Pisum sativum* chromosomes have an exceptional, meta-polycentric morphology characterized by multiple CENH3 domains located along extended primary constrictions. These primary constrictions are rich in satDNA families, and some of them are associated with CENH3 domains, while other satellite sequences are located within the primary constrictions but apart from the CENH3 loci. Using scaffolds for the assembly of chromosome 6 centromere as a reference, we characterized the spatial distribution of several selected satellites and CENH3 during interphase and individual stages of mitotic and meiotic division. We verified the location of each satDNA observed on scaffolds, demonstrating that an efficient approach for centromere assembly lies in the combination of genomic data with cytogenetics. In addition, we observed that the spatial conformation of CENH3 chromatin significantly differs from its surrounding regions constituting the extended primary constrictions. While CENH3 chromatin forms multiple separate domains on the metaphase chromosomes, it gets condensed into a single spot per chromosome during the interphase. Conversely, the regions of the primary constrictions that are not associated with CENH3 get de-condensed in the interphase.

CoverView: straightforward visualization of electron density coverage of PDB structures at the level of atoms to help validation

Aliaksei Chareshneu - CEITEC, Masaryk University, Czech Republic; National Centre for Biomolecular Research, Masaryk University, Czech Republic.

Protein Data Bank (PDB) contains huge amounts of biomacromolecular 3D structures. This extensive collection of structural data is rapidly growing. However, though the global quality of models (e.g., resolution) is slowly but continuously improving, there is still a large number of local quality issues in many structures. An important part of structure quality validation, both for X-ray and Cryo-EM structures, is related to the comparison of the model with the underlying experimental data, i.e., electron density. However, this procedure is time-consuming and requires a sufficient experience in structural biology and therefore cannot be performed by the wider audience. In order to fill this gap, we are developing a web server that would allow the users to straightforwardly visualize the electron density coverage of a protein structure at the level of atoms based on the provided isosurface threshold. The implementation is based on a Python library called *gemmi* (https://github.com/project-gemmi/gemmi).

To obtain feedback from the community on the early development stages, we prepared a prototype, freely available at: https://ncbr.muni.cz/CoverView. It consists of a sortable table with relevant information (title, assembly weight, total atoms, fraction of atoms covered, resolution high, R factor, R free) on a set of selected structures linked to pre-created Mol* sessions, where each structure is colored according to the electron density coverage of atoms (green – covered, red – not covered). The isosurface threshold for the majority of structures is set to 1.5 sigma. For another two structures we provide visualization for a range of thresholds from 1.0 to 2.0 so that the user can directly check how the threshold affects the resulting coverage. The visualization provided by CoverView, supplemented by the information in the table, helps to quickly grasp the quality of the model without the need to have in-depth understanding of crystallography field. We hope that the implementation of a web server based on the suggested concept will help not only to simplify the validation procedure, but also to make the quality information accessible to researchers examining the structures, and allow them to recognize, if the residue or ligand of their interest is represented well in the structure.

Ancient mtDNA database (AmtDB.org): an update

Edvard Ehler - Institute of Molecular Genetics of the Czech Academy of Sciences

Ancient human mitochondrial database (amtdb.org) has matured into its third year of operation. Currently, the version of the database is v1.008, which contains 2548 manually curated samples from 46 countries and 4 continents together with all the important metadata (dating, geo. location, archaeological background, etc.) and full mtDNA sequencies. During the last year we have also launched new tools and functionality on our webpage.

Firstly, it is the tool for annotation of mitochondrial pathologies. The group of mitochondrial diseases is very heterogeneous, and they can manifest differently in different tissue, cell types and age of patient. They are caused by inherited or spontaneous mutations in mtDNA (or even nDNA). Human mtDNA is well described and at least 90 mutations have been confirmed as causative for mitochondrial diseases, with as many as 750 variants were reported at least in some population as pathological. The abovementioned tool is available both online, integrated into our AmtDB, and in offline, stand-alone, version. It will annotate both modern and ancient mtDNA molecules and was used to update our database with the information about the mitochondrial disease mutations.

Secondly, we provide the certificate of ancient maternal genetic lineage (mtDNA) origin. We allow user, after providing his credentials and mitochondrial haplogroup, to search the database and return a list of the most closely related samples arranged according to their location and archaeological epoch. We also display an adjustable map of the returned samples. All this comes in printable version for the users to save or download their results.

As more and more aDNA samples are being published every month, we are planning further development of our database and services and aim for more community curated data approach.

Predicting ion binding sites in proteins using machine learning

Christos Feidakis, Radoslav Krivak, David Hoksza, Marian Novotný - Charles University Structural Bioinformatics Group; Faculty of Science, Charles University, Viničná 7, 12800 Praha 2, Czech Republic

About half of all known proteins have binding interactions with small acid radicals and metal ions in order to stabilize their structure and regulate their biological functions. The information within these interactions is key to understanding the underlying biological mechanisms that are involved in health and disease, and by extension, understanding newly discovered protein structures by assigning functional annotations to them.

The advances in protein structure prediction are expected to soon expand the size of our protein structure repositories by three orders of magnitude, highlighting the importance and utility of being able to understand and interpret new structures.

Here we focus on interactions between proteins and ions, and we are developing a software tool that allows the prediction of ion binding sites in proteins, by learning upon the structural information in known interactions. Our workflow consists of building datasets for each ion using the PDB, choosing appropriate features for learning, and optimizing a random forest-based, machine learning algorithm to get accurate predictions.

We have so far built datasets for a series of ion ligands (Cl^-, Zn^{2+}, Mg^{2+}, Na^+, Ca^{2+}, I^-, Br^-, Cd^{2+}, Mn^{2+}, Ni^{2+}, Hg^{2+}, Fe^{3+}, Co^{2+}, Cu^{2+}, Pt^{2+}, Cs^+, Fe^{2+}) and generated promising predictions for Zn^{2+} and Mg^{2+}.

Phylogeny of bacterial Hsp70

Michal Gala, Peter Pristaš, Gabriel Žoldák - Pavol Jozef Šafárik University in Košice

Heat shock proteins 70 (Hsp70) are ubiquitous ATP-dependent chaperones (folding helpers). Structurally, canonical Hsp70 consists of two folded domains and C-terminal intrinsically disordered part. A nucleotide binding domain (NBD) binds ATP. A substrate binding domain (SBD) binds protein clients, and the binding affinity depends on ATP/ADP status of the NBD. The function of the C-terminal disordered part is less understood.

Recently, single-molecule force spectroscopy Hsp70 studies pointed out a strong interplay between mechanical vulnerable/stable regions. Here we asked how such mechanical constraints are affected during the evolution of Hsp70. In particular, insertions and deletions (indels) can be highly disrupted for the mechanical interplay since such changes introduce a large shift in the register of the interactions.

In our collection, we included all bacterial Hsp70 from Swiss-Prot database. Our analyzes focused on conserved residues/motifs, sequence variability or indels occurrence. Multiple sequence alignment of Hsp70s shows that the large part of the SBD contain very low number of indels. We speculate that the absence of the indels in this region is the consequence of requirements for precise structural/geometrical mechanics of domains during Hsp70 allostery and hence indels in the SBD can cause loss-of-function of such precisely balanced molecular machine.

Machine Learning for Annotating Cavities

Faraneh Haddadi, Ondrej Vavra, David Bednar, Stanislav Mazurenko - Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Brno, Czech Republic, International Clinical Research Center, St. Anne’s University Hospital Brno, Czech Republic

Certain structural elements play a critical role in binding and unbinding ligands. In this study, we aim to predict the types of these structural elements in proteins that could play a role in binding and unbinding ligands. The main three types of these features are surface binding interface, tunnel with one opening that connects the active site with the outside environment, or channel open on both sides. There are several methods for tunnel and channel calculation, but they require customized settings to produce high-quality results. In particular, the discrimination between surface and buried binding pockets is critical for tunnel calculation but is an open problem. Our study aims to solve this problem using FPOCKET features and new features, e.g., exposed ratio and ligand coverage. We utilized the machine learning algorithms (SVM, KNN, ANN, and Random Forest) for the binary prediction or three-class prediction. Our results show the cross-validation accuracy of 72% and 60%, respectively, indicating the promising potential of machine learning algorithms for solving this task. We plan to leverage this potential with the data extracted from molecular dynamics simulations via advanced Artificial Intelligence methods in the future.

Acknowledgements: We would like to express our thanks to ELIXIR (LM2018131) and e-INFRA (LM2018140) infrastructures for the financial support.

Regulatory function discovery using combined omics and sequence analysis

Tomáš Honzík - University of West Bohemia

Purpose is to create a search tool for biological parts with desired regulatory function. Information from omics data obtained from specific experiments and affinity distribution of known transcription factors across the DNA sequence was used to design a novel method to solve the problem. Performance of the method was tested by genetic engineering of the yeast species Saccharomyces cerevisiae.

FireProt^{ASR} - FULLY AUTOMATED ANCESTRAL SEQUENCE RECONSTRUCTION

Rayyan Tariq Khan[*1,2], Milos Musil[1,2,3], Jan Stourac[1,2], David Bednar[1,2] and Jiri Damborsky[1,2] - [1] Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic; [2] International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic; [3] Department of Information Systems, Faculty of Information Technology, Brno University of Technology, 612 66 Brno, Czech Republic; Rayyan.Tariq.Khan@gmail.com

Robust and stable enzymes accelerating chemical reactions with high specificity represent one of the keystones of metabolic engineering. During the course of evolution, nature produced a large number of diverse solutions to enzymatic catalysis by evolving natural proteins. While most of these ancestral proteins have been lost due to extinction, or to the march of evolution, it is possible to statistically infer their sequences using Ancestral Sequence Reconstruction (ASR)[1]. ASR approach has been used to resurrect ancestral proteins, to identify key amino acid residues in metabolic enzyme complexes[2], to engineer thermostable enzymes[3], to deduce information about evolutionary events and natural history[4], as well as, to engineer enzymes with dual, catalytically distinct activities[5]. The latter is possible by reconstructing ancestors of enzymes that are catalytically differentiated and evolved along separate trajectories.

ASR is generally not accessible to scientists outside communities of evolutionary biologists. The purpose of this research is to construct a fully automated pipeline that allows anyone, including those who lack specialist knowledge, to perform ASR thus reducing the academic barrier to entry. The pipeline was verified against work that was previously done in the lab; on the ancestral sequence reconstruction of haloalkane dehalogenase family[5]. The pipeline is accessible online through the FireProt^{ASR} webserver[6] at: https://loschmidt.chemi.muni.cz/fireprotasr/. The server can be used for designing stable, highly expressible and catalytically active enzymes for the assembly of metabolic pathways for the construction of cell factories producing high-value chemicals.

The author would like to express his gratitude for the amazing support provided by the entire team of Loschmidt Laboratories as well as to ELIXIR and eINFRA for grant and financial support.

REFERENCES

1. Pauling, L., & Zuckerkandl, E., *Acta Chemica Scandinavica Supplements* 1963, 17, S9–S16.
2. Holinski, A., et al., *Proteins: Structure, Function, and Bioinformatics* 2017, 85.2, 312-321.
3. Babkova, P., et al., *ChemBioChem* 2017, 18, 1448.
4. Gaucher, E., et al., *Nature* 2008, 451, 704.
5. Musil, M., et al., *Briefings in Bioinformatics* 2021, 2020bbaa337.
6. Khan, R. T., et al., *Current Protocols* 2021, 1.2, e30.

Increasing Effectiveness of Data Management Planning: DSW State-of-the-Art Features

Freeman Jana, Knaisl Vojtěch, Machačová Tereza, Pergl Robert, Slifka Jan, Suchánek Marek - Czech Technical University in Prague, Faculty of Information Technology

Data Stewardship Wizard (DSW) as a cutting-edge data management planning (DMP) tool puts a lot of emphasis on the development of state-of-the-art features that are driven by the most recent technologies, as well as by the DMP community needs and ideas. In the past year we have focused on increasing the effectiveness of the DMP by introducing numerous innovative and by the community highly demanded features, such as i) online collaboration, ii) various sharing options including anonymous project, iii) version history, iv) comments and editor notes, v) multi-choice type of questions, or vi) template development. The effectiveness of the DMP has and always will be one of the main objectives of the DSW development. The aim of this poster is to bring an overview of these developments to help researchers reach peak effectiveness in DM planning with DSW.

CaverWeb 2.0 – Identification of the protein tunnels in trajectories from molecular dynamics

Petr Kohout - Loschmidt Laboratories and International Clinical Research Center, St. Anne’s University Hospital Brno, Brno, Czech Republic

Caver Web is an unique web server suitable for identification of protein tunnels and channels with the possibility of subsequent analysis of ligand transport process. The characteristic of this program is a straightforward simple workflow and synoptical user-friendly interface with a minimum required input from the user which makes the server suitable even to researchers without advanced bioinformatics or technical knowledge. Its current version is already highly used and well-established within the scientific community.

The biggest limitation of the current version is the possibility of analyzing only one static structure which is not sufficient today, so we looked for a way how to give more relevant results to users and how to minimize bad results because of artifacts in static structures. After increasing our computational resources, we decided to integrate new version of the Caver tool which enables tunnels analysis within the MDs and release new version of this very popular application.

The most significant extension will be calculation of molecular dynamics which will cover the dynamic tunnel nature by enabling the generation of a large number of different protein conformations for better reflection of the real protein states. By this feature, the application will become the only one that will provide tunnel analysis without need of manual calculations of molecular dynamics. Users will be able to analyze the individual tunnels statistically even one by one in a new interactive web interface that will be designed specifically for the scientific community with the support of simple export and with preserving the simplicity and clarity from the previous version. The tool will be available for free to the whole scientific society.

Data analysis for MICA / MICB genes identification

Katerina Kratochvilova [1], Alena Machuldova [2], Martin Leba [1], Pavel Jindra [3], Pavel Ostasov [2], Diana Maceckova [2], Robin Klieber [2], Hana Gmucova [3], Monika Holubova [2,3] and Lucie Houdova [1] - [1] Faculty of Applied Sciences, University of West Bohemia, [2] Faculty of Medicine in Pilsen, Charles University, [3] Department of Haematology and Oncology, University Hospital Pilsen

Natural killer (NK) cells play a key role in immune response. NK cells are the first reconstituted cells after the allogeneic hematopoietic stem cell transplantation. Their activity can be mediated through MICA and MICB ligands binding the activating receptor NKG2D. That is why MICA and MICB polymorphisms may have significant role in treatment outcome of acute myeloid leukemia (e.g., presence of specific amino acids could be associated with shorter survival or early relapse of disease).

Our approach to identification of MICA and MICB alleles is presented. Introduced pipeline combines data of MICA exons 2, 3, 4 and MICB exons 2, 3, 4 and 5 from Sanger sequencing together with available reference sequences from IPD database. Specific issues encountered during data processing such as low data quality, sequencing errors, consensus creation or obtaining final result from partial results of mentioned exons are presented as well as proposed solutions with an emphasis on possibility of future clinical usage.

The study was funded by the Ministry of Health of the Czech Republic - Czech Health Research Council (project no. NV18- 03-00277).

The Chicken-and-Egg Problem of Landmark-Driven Molecular Dynamics: Are Random Landmarks Useful?

Aleš Křenek, Jana Hozzová, Jaroslav Olha, Martin Kurečka, Dalibor Trapl, Vojtěch Spiwok - Masaryk University, University of Chemistry and Technology Prague

Molecular dynamics of proteins can be guided to explore wider range of configurational space (folding paths in particular) with biased potential built on path collective variables, which are derived from a set of landmark structures which approximate the desired trajectory [1,2]. The technique can reduce the time of the MD simulation dramatically. On the other hand, the choice of landmarks is the core of the “chicken and egg” problem – if we know the landmarks along the trajectory, what would be the reason of recomputing the same trajectory? (Well, things are not so simple, but this is the core.)

In this study, we aim at computing the trajectories de novo, without prior knowledge. First, a set of several hundreds to thousands of barely feasible landmarks is generated by random twisting of peptide bonds in the subject protein. Many such structures contain steric clashes and other unrealistic properties. Those are filtered by steepest descent energy minimization in vacuo, using simple force field (Amber99); if the minimization fails or it stops at too high energy, the structure is discarded. Typically, sufficiently high number (e.g. more than 80 %) of the structures pass. Those remaining landmarks can be used directly to define “propertymap” [2] collective variables, or they can be used to train a simple neural network to estimate such collective variables [3]. The latter approach usually yields slightly better results (wider explored space); we hypothesize that propertymap is designed to work efficiently in near vicinity of the landmarks, which is not the case with random ones, and it falls back to unbiased MD in the farther areas. On the contrary, the neural network can extract “feasible” features of the structures anyway.

We demonstrate results of the method with MD simulations of several small to medium sized proteins, known to fold in few microseconds with unbiased MD. With all of them we can show that a 200 ns MD trajectory is sufficient to explore the folding trajectory, while the protein does not leave its native state with unbiased simulation of the same length.

[1] https://doi.org/10.1063/1.2432340
[2] https://doi.org/10.1063/1.3660208
[3] https://doi.org/10.3389/fmolb.2019.00025

Signal-to-noise limitations of genetic toggle switches

Lukas Kuhajda - University of West Bohemia

Purpose is to create a genetic toggle switch by which gene expression can be turned ON or OFF.

A novel neural network architecture is designed and applied to solve the problem.

Performance of the method was tested by genetic engineering of the yeast species Saccharomyces cerevisiae.

OverProt: Overview of Secondary Structure Consensus in Protein Families

Adam Midlik[1], Ivana Hutařová Vařeková[1], Jan Hutař[1], Radka Svobodová[1], Karel Berka[2] - [1] Masaryk University, Brno, [2] Palacký University, Olomouc

Secondary structure elements (SSEs) provide a deep insight into the architecture of a protein. Therefore, a figure depicting a sequence of SSEs for individual proteins is used in key structural databases (PDBe, RCSB PDB). Similarly, for a whole protein family, a consensus of SSEs can be constructed. This consensus shows the general protein fold of the family and its structural variation. In order to construct the SSE consensus, the SSEs from individual protein family members must first be mapped together. Unfortunately, some proteins within a family often miss some SSEs, therefore the mapping is not straightforward. Previously, the only automated solution for the mapping of the SSEs was SecStrAnnotator, which finds and annotates structurally equivalent SSEs and enables their mutual mapping within the whole family. However, SecStrAnnotator requires a manually prepared annotation template for each family. The new OverProt algorithm overcomes this need and thus allows fully automated construction of the SSE consensus. The OverProt database (https://overprot.ncbr.muni.cz/) currently offers the precomputed SSE consensus for each CATH protein family. These results are visualized by the interactive viewer, which shows information about the SSE type, length, frequency of occurrence, spatial variability, and beta connectivity.

Exosomes produced by melanoma cells significantly influence the biological properties of normal and cancer-associated fibroblasts

Lucie Pfeiferova - Laboratory of Genomics and Bioinformatics,Institute of Molecular Genetics of the Czech Academy of Sciences; Department of Informatics and Chemistry,Faculty of Chemical Technology, University of Chemistry and Technology Prague

The incidence of cutaneous malignant melanoma is increasing worldwide. While the treatment of the initial stages of the disease is simple, the advanced disease frequently remains fatal despite novel therapeutic options. This urges for identification of novel therapeutic targets in melanoma. Similar to other types of tumors, the cancer microenvironment plays a prominent role and determines the biological properties of melanoma. Importantly, melanoma cell-produced exosomes represent an important tool of intercellular communication within this cancer ecosystem. We have focused on potential differences in the activity of exosomes produced by melanoma cells towards melanoma-associated fibroblasts and normal dermal fibroblasts. Cancer-associated fibroblasts were activated by the melanoma cell- produced exosomes significantly more than their normal counterparts, as assessed by increased transcription of genes for inflammation- supporting cytokines and chemokines, namely IL-6 or IL-8. We have observed that the response is dependent on the duration of the stimulus via exosomes and also on the quantity of exosomes. Our study demonstrates that melanoma-produced exosomes significantly stimulate the tumor-promoting proinflammatory activity of cancer-associated fibroblasts. This may represent a potential new target of oncologic therapy.

QM-like partial atomic charges for proteins available online

Ondřej Schindler, Tomáš Raček, Radka Svobodová - National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic, CEITEC-Central European Institute of Technology, Masaryk University, Kamenice 5, 602 00, Brno, Czech Republic

Partial atomic charges are real numbers assigned to individual atoms of a molecule that approximate the distribution of electron density among these atoms. Partial atomic charges find many applications in computational chemistry, chemoinformatics, bioinformatics, and nanoscience. In general, there are two approaches to calculate partial atomic charges. The partial atomic charges approximate the electron density and therefore the most reliable way is to obtain them directly from the electron density by some quantum mechanical method (QM). A substantial disadvantage of QM approaches is their high computational complexity, and therefore a long computational time. For this reason, QM methods are inapplicable for calculating the partial atomic charges of proteins. Empirical charge calculation methods are faster alternatives to QM methods. Empirical methods calculate partial atomic charges from information like about the position of the atoms and possibly the bonds between them. Empirical methods use only the positions of atoms, some of their characteristics (e.g. electronegativity, chemical hardness, radius, etc.) and possibly the bonds between them to calculate partial atomic charges. Many empirical methods have already been developed. However, these methods have their limitations—e.g., their application for peptides, proteins, and other homogeneous macromolecular systems (i.e., systems composed from just several types of residues) is problematic.

In this work, we introduce Split-charge Equilibration with Parameterized Initial Charges (SQE+qp) [1], adapted for peptides and proteins. Our method can reproduce QM partial atomic charges with high accuracy. We also present an implementation of SQE+qp via a web application Atomic Charge Calculator II (https://acc2.ncbr.muni.cz/) [2]. We provide the scientific community a freely available online tool for the accurate calculation of QM-like partial atomic charges.

[1] Schindler, O., Raček, T., Maršavelski, A., Koča, J., Berka, K. and Svobodová, R., 2021. Optimized SQE atomic charges for peptides accessible via a web application. *Journal of cheminformatics*, 13(1), pp.1-11.

[2] Raček, T., Schindler, O., Toušek, D., Horský, V., Berka, K., Koča, J. and Svobodová, R., 2020. Atomic Charge Calculator II: web-based tool for the calculation of partial atomic charges. *Nucleic acids research*, 48(W1), pp.W591-W596.

Combining and visualizing experimental data and annotation features on *S. coelicolor* genome for biologists purposes

Marek Schwarz - Institute of Microbiology of the Czech Academy of Sciences

The *Streptomyces coelicolor* has large number of σ factors, many of which do not have defined regulon. We were interested in σE and we wanted to find out which genes transcription it regulates. As the σ factors are required for specific recognition of gene promoter sequence we are naturally also interested in identification of binding sites for individual genes. As a basis for the analysis we performed a Chip-Seq experiment for the σE.

To identify the actual binding sites we combined the motif discovery analysis on Chip-Seq data with published data about TSS location [DOI:10.1038/ncomms11605], RNAseq data [DOI:10.1186/1471-2164-14-558], σE binding sites and peak location from [DOI:10.1111/mmi.14250] and HrdB Chip-Seq data from our previous work [DOI:10.1093/nar/gky1018]. Combining and visualization of such data to be easily usable by biologist was not straightforward as the view of the whole peak loci with its genome neighborhood and single nucleotide resolution was required in one visualization.

Without making the data publicly available from the web, options to make interactive visualization that can be easily transferred from bioinformatician to biologists are limited. We used NCBI’s Sequence Viewer to handle the visualization of genome annotations, TSS, binding sites and Chip-Seq peak locations; the RNAseq data and motif sites location were plotted with python in Jupyter notebook and whole visualization was exported as single HTML file that could be sent e.g. by e-mail. The visualization was then used to curate the candidate binding sites of σE.

MolMeDB – Molecules on Membranes Database

Jakub Juračka, Martin Šrejber, Michaela Jaroměřská, Václav Bazgier, Kateřina Storchmannová, Dominik Martinát, Karel Berka - Palacký University Olomouc

Biological membranes are natural barriers of cells. The membranes play a key role in cell life and also in the pharmacokinetics of drug-like small molecules. There are several ways how a small molecule can get through the membranes. Passive diffusion, active or passive transport via membrane transporters are the most relevant ways how the small molecules can get through the membranes. There is an available huge amount of data about interactions among the small molecules and the membranes also about interaction among the small molecules and the transporters. MolMeDB[1] (https://molmedb.upol.cz/detail/intro) is a comprehensive and interactive database. In the past, we have collected interactions of small molecules with the membranes such as partitioning, penetration, and free energy profiles of the small molecules especially drugs crossing the membranes. Recently, we have expanded our area of interest about the interactions of small molecules with transporters. Nowadays, data is available from 52 various methods for 40 biological or artificial membranes and for 184 transporters in MolMeDB. The data within the MolMeDB is collected from scientific papers, our in-house calculations (COSMOmic[2] and PerMM[3]) and obtained by data mining from several databases. Data in the MolMeDB are fully searchable and browsable by means of name, SMILES, membrane, method, transporter or dataset and we offer collected data openly for further reuse.

References

[1] Juračka, J., Šrejber, M., Melíková, M., Bazgier, V. & Berka, K. MolMeDB: Molecules on Membranes Database. Database 2019, (2019).

[2] Klamt, A., Huniar, U., Spycher, S. & Keldenich, J. COSMOmic: A mechanistic approach to the calculation of membrane-water partition coefficients and internal distributions within membranes and micelles. J. Phys. Chem. B 112, 12148–12157 (2008).

[3] Lomize, A. L. et al. PerMM: A Web Tool and Database for Analysis of Passive Membrane Permeability and Translocation Pathways of Bioactive Molecules. J. Chem. Inf. Model. 59, 3094–3099 (2019).

Analysis of sequencing data from reprogramming of immortalized cell line

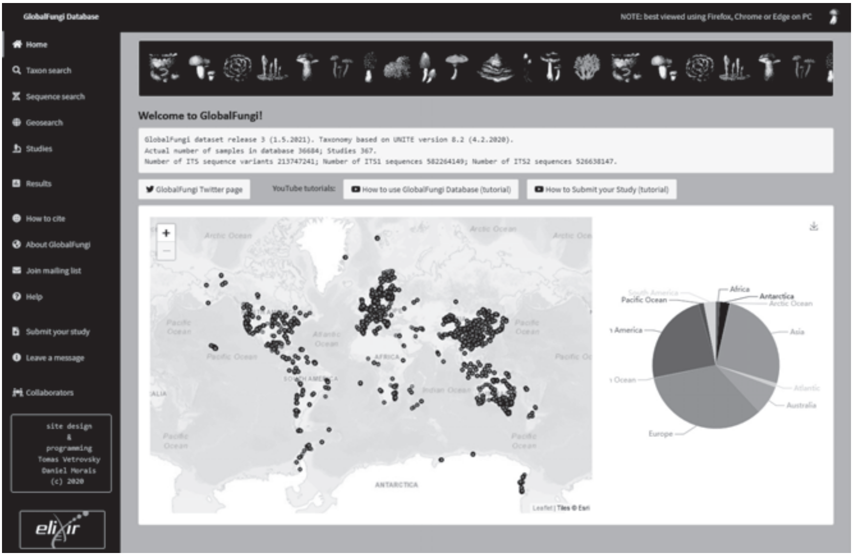
Petra Svatoňová - Institute of Molecular Genetics of the Czech Academy of Sciences

For exploring erythroid progenitors reprogramming of zebrafish immortalized cell line, bulk ATAC-Seq together with RNA-Seq was performed and analysed. After preprocessing part, where quality control (FastQC), trimming (Cutadapt/Trimmomatic), mapping (HISAT2/ Salmon) and filtering (Picards MarkDuplicates, Samtools/SortMeRNA) were included, regions of interest were detected for ATAC-Seq via peak calling (Macs2). Then quantification (FeatureCounts/Salmon quant, Tximport) were performed, followed by differential analysis (DESeq2). Reprogramming of cell type involves changes on many levels. Indeed, thousands of genes were significantly deregulated, tens of thousands peaks marked differentially abundant regions on genome. To reduce these huge numbers and gain more biological view, genes in defined proximities from peaks were considered. Not surprisingly, hemoglobins with other heme/oxygen binding/carrying genes occur among top10 downregulated genes. In contrast, most upregulated genes are markers of myeloid cells. That supports observed changes and proves successful reprogramming.

GlobalFungi – a global database of fungal occurrences

Tomáš Větrovský - Institute of Microbiology of the CAS, v. v. i.

Thanks to the recent advance of high-throughput-sequencing methods we are facing an accumulating wealth of fungal sequencing data from various geographical regions, ecosystems and habitats. Although the application of NGS methods revolutionized our understanding of fungal ecology, the accumulating raw fungal NGS data in sequence repositories did not bring much extra value so far. The idea behind this **GlobalFungi Database (https://globalfungi.com/)** is to provide everyone the access to published data on fungal community composition obtained by next-generation-sequencing through a web-based interface that allows various queries of the database and visualization of the results. To date, we have collected more than 1 billion observations of the fungal ITS1 and ITS2 marker sequences from next-generation sequencing data published in 367 studies and containing more than 36,000 samples from around the world. Our database covers data from all terrestrial habitats except those subject to experimental manipulation, containing information on fungal communities from soil, litter, dead plant material, living plant tissues, water, air, dust and others. GlobalFungi invites participation of the scientific community in that it encourages submission of data by the authors of studies that are not yet covered.



```
GlobalFungi dataset release 3 (1.5.2021). Taxonomy based on UNITE version 8.2 (4.2.2020).
Actual number of samples in database 36684; Studies 367.
Number of ITS sequence variants 213747241; Number of ITS1 sequences 582264149; Number of ITS2 sequences 526638147.
```

Sequence organization of CENH3-associated heterochromatic loci on the holocentric chromosomes of *Cuscuta europaea*

Tihana Vondrak - Biology Centre CAS, Institute of Plant Molecular Biology

The holocentric plant Cuscuta europaea displays unique distribution pattern of CENH3 chromatin which is associated with a subset of DAPI-positive heterochromatic bands that are unevenly distributed on its mitotic chromosomes. Since the mitotic spindle gets attached along entire chromosomes, it is likely that CENH3 has lost its function of the centromere determinant in this species. This rises the questions why is the CENH3 protein still present on chromosomes and what determines its targeting to the heterochomatic bands. As the first step to elucidate these questions, we investigated the long-range structure of the CENH3-binding chromosome regions employing ultra-long nanopore sequencing combined with bioinformatic analysis of the reads and FISH mapping of selected repeats. We revealed a complex structure of these regions, which are composed of the short arrays of CUS-TR24 satellite interrupted frequently by emerging simple sequence repeats and targeted insertions of a specific lineage of LINE retrotransposons. Interestingly, our preliminary ChIP-seq experiments revealed that despite this interspersion of three different repeats in the CENH3-associated chromosome bands, the CENH3 protein is enriched at CUS-TR24 repeats only. This suggests sequence-specific targeting of CENH3 to these loci, which contrasts with the epigenetically-determined deposition of CENH3 that is common in plants.

List of participants

Zahra Aliakbartehrani
Zahra.Aliakbartehrani@ibt.cas.cz
Institute of Biotechnology of the CAS

Laura Ávila Robledillo
l.avila.robledillo@gmail.com
Biology Centre of the CAS

Petr Baldrian
baldrian@biomed.cas.cz
Institute of Microbiology of the CAS

Jan Bartoš
bartos@ueb.cas.cz
Institute of Experimental Botany of the CAS

Jan Beránek
beranekj@vscht.cz
University of Chemistry and Technology,
Prague

Karel Berka
karel.berka@upol.cz
Palacký University Olomouc

Lada Biedermannová
lada.biedermannova@gmail.com
Institute of Biotechnology of the CAS

Korbinian Bösl
korbinian.bosl@uib.no
ELIXIR Norway

Dominik František Bučík
bucik@ics.muni.cz
Masaryk University

Tereza Čalounová
tereza.calounova@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Jiří Černý
jiri.cerny@ibt.cas.cz
Institute of Biotechnology of the CAS

Aliaksei Chareshneu
479052@mail.muni.cz
Masaryk University

Edvard Ehlera
edvard.ehler@img.cas.cz
Institute of Molecular Genetics of the CAS

Kateřina Faltejsková
katerina.faltejskova@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Christos Feidakis
christos.feidakis@natur.cuni.cz
Charles University

Zdeněk Futera
zfutera@prf.jcu.cz
University of South Bohemia in České
Budějovice

Michal Gala
michal.gala@student.upjs.sk
Pavol Jozef Šafárik University in Košice

Jakub Galgonek
jakub.galgonek@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Vijaya Geetha Gonepogu
geetha.gonepogu@paru.cas.cz
Biology Centre of the CAS

Ansgar Gruber
ansgar.gruber@paru.cas.cz
Biology Centre of the CAS

Faraneh Haddadi
518408@mail.muni.cz
Masaryk University

Mubasher Hassan
mubashirhassan_gcul@yahoo.com
University of Lahore

David Hoksza
david.hoksza@matfyz.cuni.cz
Charles University

Tomáš Honzík
tomas.honziik@seznam.cz
University of West Bohemia

Jana Horáčková
jana.horackova@recetox.muni.cz
International Clinical Research Center of St.
Anne's University Hospital in Brno

Lucie Houdová
houdina@ntis.zcu.cz
University of West Bohemia

Mariia Hrysiuk
mariagrisyuk@gmail.com
Institute of Biotechnology of the CAS

Jan Jelínek
jan.jelinek@biomed.cas.cz
Institute of Microbiology of the CAS

Jakub Juračka
jakub.juracka@upol.cz
Palacký University Olomouc

Rayyan Tariq Khan
Rayyan.Tariq.Khan@gmail.com
International Clinical Research Center of St.
Anne's University Hospital in Brno

Vojtěch Knaisl
knaisvoj@fit.cvut.cz
Czech Technical University in Prague

Petr Kohout
xkohou14@stud.fit.vutbr.cz
Loschmidt Laboratories

Kateřina Kratochvílová
kkratoch@ntis.zcu.cz
University of West Bohemia

Aleš Křenek
ljocha@ics.muni.cz
Masaryk University

Ivana Křenková
krenkova@ics.muni.cz
CESNET

Radoslav Krivák
rkrivak@gmail.com
Charles University

Lukáš Kuhajda
kuhajda.l@gmail.com
University of West Bohemia

Jiří Macas
macas@umbr.cas.cz
Biology Centre of the CAS

Simona Macháčova
machacova.simona@gmail.com
Charles University

Carlos Eduardo Madureira Trufen
trufenc@img.cas.cz
Institute of Molecular Genetics of the CAS

Dominik Martinát
dominik.martinat@gmail.com
Palacký University Olomouc

Luděk Matyska
ludek@ics.muni.cz
Masaryk University

Adam Midlik
midlik@mail.muni.cz
Masaryk University

Martin Modrák
martin.modrak@biomed.cas.cz
Institute of Microbiology of the CAS

Martin Mokrejš
martin.mokrejs@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Petr Novák
petr@umbr.cas.cz
Biology Centre of the CAS

Marian Novotný
marian@natur.cuni.cz
Charles University

Jiří Novotný
jiri.novotny@img.cas.cz
Institute of Molecular Genetics of the CAS

Miroslav Oborník
obornik@paru.cas.cz
University of South Bohemia in České
Budějovice

Jan Pačes
hpaces@img.cas.cz
Institute of Molecular Genetics of the CAS

Robert Pergl
perglr@fit.cvut.cz
Czech Technical University in Prague

Lucie Pfeiferová
lucie.pfeiferova@img.cas.cz
Institute of Molecular Genetics of the CAS

Natália Pižemová
natalia.pizemova@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Pavel Polishchuk
pavlo.polishchuk@upol.cz
Palacký University Olomouc

Miloš Prokýšek
prokysek@prf.jcu.cz
University of South Bohemia in České
Budějovice

Miroslav Ruda
ruda@ics.muni.cz
CESNET

Petr Ryšavý
petr.rysavy@fel.cvut.cz
Czech Technical University in Prague

Ondřej Schindler
ondrej.schindler@mail.muni.cz
Masaryk University

Bohdan Schneider
bohdan.schneider@gmail.com
Institute of Biotechnology of the CAS

Marek Schwarz
marek.schwarz@biomed.cas.cz
Institute of Microbiology of the CAS

Ayush Sharma
ayush.sharma@paru.cas.cz
Biology Centre of the CAS

Jan Slifka
slifkjan@fit.cvut.cz
Czech Technical University in Prague

Andrew Smith
asmith@ebi.ac.uk
ELIXIR Hub

Mykola Snisar
snisarm@natur.cuni.cz
Charles University

Vojtěch Spiwok
spiwokv@vscht.cz
University of Chemistry and Technology,
Prague

Kateřina Storchmannová
storchmannova.katerina@gmail.com
Palacký University Olomouc

Anna Strachotová
anna.strachotova@uochb.cas.cz
Institute of Organic Chemistry and
Biochemistry of the CAS

Michal Strejček
michal.strejcek@vscht.cz
University of Chemistry and Technology,
Prague

Marek Suchánek
marek.suchanek@fit.cvut.cz
Czech Technical University in Prague

Petra Svatoňová
svatonp@img.cas.cz
Institute of Molecular Genetics of the CAS

Jakub Svoboda
Jakub.Svoboda@ibt.cas.cz
Institute of Biotechnology of the CAS

Radka Svobodová
radka.svobodova@ceitec.muni.cz
Masaryk University

Ondřej Uhlík
ondrej.uhlik@vscht.cz
University of Chemistry and Technology, Prague

Tomáš Větrovský
kostelecke.uzeniny@seznam.cz
Institute of Microbiology of the CAS

Rudolf Vohnout
rudolf.vohnout@prf.jcu.cz
University of South Bohemia in České Budějovice

Marta Vohnoutová
mvohnoutova@jcu.cz
University of South Bohemia in České Budějovice

Jiří Vohradský
vohr@biomed.cas.cz
Institute of Microbiology of the CAS

Tihana Vondrak
tihana.vondrak12@gmail.com
Biology Centre of the CAS

Jiří Vondrášek
jiri.vondrasek@uochb.cas.cz
ELIXIR Czech Republic, Director

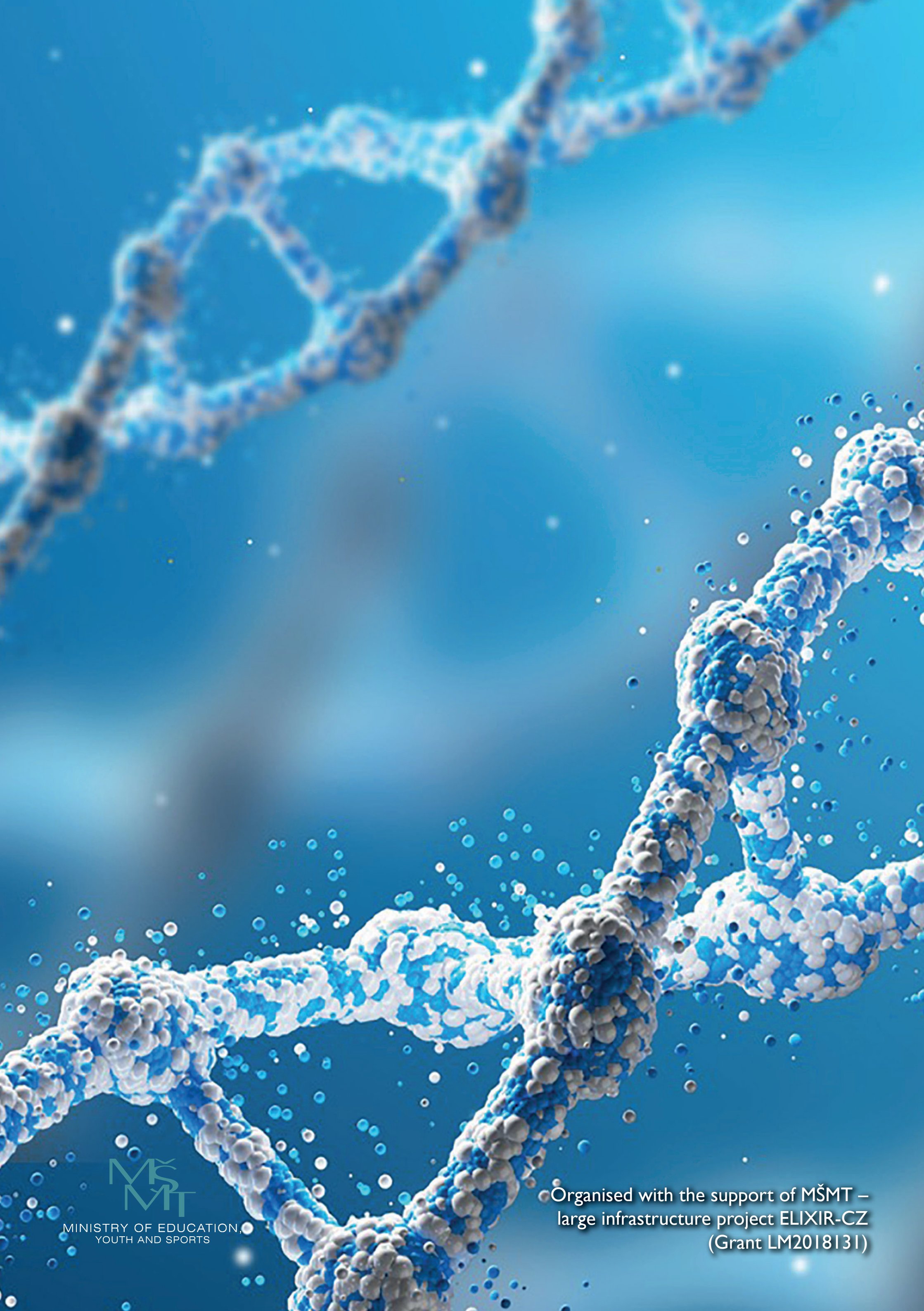
Jiří Vorel
vorel@cesnet.cz
CESNET

Shun-Min Yang
shun-min.yang@paru.cas.cz
University of South Bohemia in České Budějovice

Notes

Published by Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences and ELIXIR CZ.

Editors: Jiří Vondrášek, Anna Strachotová, Natália Pižemová
Print: VENICE Praha s.r.o.
Year: 2021
Pages: 70



MINISTRY OF EDUCATION,
YOUTH AND SPORTS

Organised with the support of MŠMT –
large infrastructure project ELIXIR-CZ
(Grant LM2018131)