Thursday, November 16, S4 - Activity 4

To develop tools to Describe, Analyse, Annotate, and Predict Nucleic Acid Structures

L14

RNA-PUZZLES : BLIND ASSESSMENTS OF (SEMI)-AUTOMATIC 3D RNA MODELING

Eric Westhof[1], Zhichao Miao[2]

[1]Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS, 67000 Strasbourg France
[2]GMU-GIBH Joint School of Life Sciences, Guangzhou Laboratory, Guangzhou Medical University, Guangzhou, China;
e.westhof@unistra.fr

RNA 3D structure modeling dates to the late 1960s and several computer programs for predicting RNA 3D structures have been proposed since then. RNA-Puzzles is a collaborative effort dedicated to advancing and improving RNA 3D structure prediction. With the agreement of crystallographers, RNA structures are predicted by different groups before the publication of crystal structures. Since the success of AlphaFold in protein structure prediction, artificial intelligence approaches are continuously designed to solve the problem of RNA 3D structure prediction with strategies like AlphaFold. However, eliminating redundancy between training and test data is not trivial and some programs have shown overfitting results. Therefore, blind, unbiased evaluations (based on equivalence of comparison metrics) of all prediction tools are a necessary requirement.

A dedicated website (http://www.rnapuzzles.org/) gathers the systematic protocols and parameters used for comparing models and crystal structures, all the data, analysis of the assessments, and related publications. Up to now, 40 RNA sequences with experimentally determined structures (X-ray or cryo-EM) have been predicted by many groups from several countries. Many of the predictions have achieved high accuracy after comparison with the solved structures.

L15

Rfam, RNA 3D STRUCTURES, AND ISSUES FACING RNA 3D STRUCTURE PREDICTION

Blake Sweeney

EMBL-EBI, United Kingdom
bsweeney@ebi.ac.uk

Rfam is a database of over 4,000 non-coding RNA (ncRNA) families. Each family is composed of a sequence alignment called the seed, often manually curated, a consensus secondary structure and a covariance model. Rfam was originally developed 20 years ago to annotate genomes with ncRNAs using the covariance models. However, it has become the de-facto reference database for known ncRNAs and their alignments. This has led to it being used in new contexts including, RNA 3D structure prediction. This has pushed Rfam in new directions. Recently, Rfam has been improved by aligning sequences and base pair annotations from 3D structures into seed alignments. This connects Rfam alignments with 3D structures directly and allows improvements of families. We have used this to improve over 30 families and have started annotating pseudoknots. However despite these improvements, Rfam still has several limitations that make the prediction of RNA 3D structures challenging. Briefly, they are that ncRNA data is limited, biassed and incomplete. In this talk we will discuss some of these issues, suggest possible improvements, and challenge the community to solve them.

L16

UNRAVELING THE RNA WEB: DETECTING AND DECIPHERING ENTANGLEMENTS IN 3D STRUCTURES

M. Szachniuk[1,2], M. Antczak[1,2], M. Popenda[2], J. Sarzynska[2], T. Zok[1]

[1]Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry PAS, Poznan,
Polandmszachniuk@cs.put.poznan.pl

RNA molecules, essential players in the intricate machinery of cellular processes, exhibit a remarkable level of complexity in their three-dimensional structures. For many years, the primary focus in RNA structure study has traditionally been on base-pairing interactions and simple structure motifs. However, recent advances have unveiled another dimension of complexity – the presence of entanglements within RNA 3D structures [1]. These structural intricacies, reminiscent of topological puzzles, may have profound implications for RNA function and dynamics [2]. On the other hand, some of their types may be bugs injected into the structure, during its determination or in silico modelling process.

In this presentation, we will explore the diverse range of entangled motifs that can be found within RNA molecules [3]. We will delve into the computational algorithms that have been developed to detect and analyse these unusual topological configurations in RNA structures [4, 5]. Finally, we will take a look at entanglements in experimental and simulated models of RNA 3D structure [6] and we will learn if they can be untangled with any existing methods.

1. A. Burton, M. Di Stefano, N. Lehman, H. Orland, C. Micheletti, *RNA Biol.*, **13**, (2016), 134.
2. T. Schlick, Q. Zhu, A. Dey, S. Jain, S. Yan, A. Laederach, *J. Am. Chem. Soc*., **143**, (2021), 11404.
3. M. Popenda, T. Zok, J. Sarzynska, A. Korpeta, R. W. Adamiak, M. Antczak, M. Szachniuk, *Nucleic Acids Res*., **49**, (2021), 9625.
4. A. Jarmolinska, A. Gambin, J. Sulkowska, *Bioinformatics*, **36**, (2019), 953.
5. K. Luwanski, V. Hlushchenko, M. Popenda, T. Zok, J. Sarzynska, D. Martsich, M. Szachniuk, M. Antczak, *Nucleic Acids Res*., **50**, (2022), W663.
6. J. Sarzynska, M. Popenda, M. Antczak, M. Szachniuk, *PROTEINS: Struct., Funct., Bioinf.*, **91**, (2023), 1.

This work was supported by the National Science Center, Poland [2019/35/B/ST6/03074 to M.S.], statutory funds of Poznan University of Technology and the Institute of Bioorganic Chemistry, PAS.

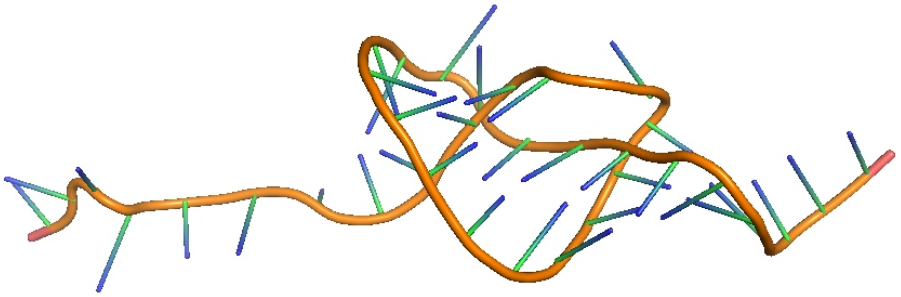

Figure 1. Entangled model generated by RNAComposer.

L17

REFERENCE-FREE RANKING METHOD FOR RNA 3D MODELS

Tomasz Zok[1], Jan Pielesiak[1], Maciej Antczak[1,2], Marta Szachniuk[1,2]

[1]Poznan University of Technology, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poland
tomasz.zok@cs.put.poznan.pl

There has been a surge of interest in predicting RNA 3D structures lately, with more researchers recognizing the significance of understanding the structure and function of RNA. As our knowledge of RNA molecules expands, we can leverage the advancements made in protein structure prediction to improve our predictions in the RNA field.

However, a significant challenge when predicting novel structural folds is assessing the quality of the models produced. Modeling software often generates multiple models per input, sometimes even thousands, making selecting the most promising ones crucial. Traditionally, researchers determine the quality of the model based on energy terms calculated using force fields or coarse-grained statistical potentials. The lower the energy calculated, the more likely the RNA structure is considered to be. However, the energy landscape usually contains many local minima, leading to inconclusive results.

Therefore, we propose a different approach for ranking multiple 3D models created from the same sequence by analyzing the base pairs and stacking interactions within them. We build a consensus secondary structure from the extracted data and rank each model's interaction network against that consensus to provide a final ranking.

We benchmarked our proposed method on public RNA 3D modeling datasets to verify its usefulness, comparing its performance against state-of-the-art energy-based evaluations.

L18

POSTTRANSCRIPTIONAL MODIFICATIONS IN RNA EXPERIMENTAL 3D STRUCTURES: OCCURRENCES AND EFFECT ON INTERBASE HYDROGEN BONDING

M. Chawla[1], L. Cavallo[1], R. Oliva[2]

[1]King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering Division, Kaust Catalysis Center, Thuwal 23955-6900, Saudi Arabia
[2]Department of Sciences and Technologies, University Parthenope of Naples, Centro Direzionale Isola C4, I-80143 Naples, Italy
romina.oliva@uniparthenope.it

The physicochemical information of RNA molecules is greatly enhanced by posttranscriptional modifications, contributing to explain the diversity of their structures and functions.

To date, over 150 natural modifications have been characterized in all major classes of RNAs, ranging from isomerization or methylation, to the addition of bulky and complex chemical groups [1-2]. Modifications can change the folding landscape of RNA, resulting at times in alternative conformations [2-4]. This occurs by altering the interactions between nucleotides. Especially the H-bonding between nucleobases, both the regular Watson–Crick pairs enclosed in the RNA stems and the non Watson–Crick pairs outside the stems [5] - also known as tertiary interactions -, can be affected by modifications due to steric and energetic effects.

In order to investigate the impact of modifications on the interbase H-bonding, we have set up an approach combining structural bioinformatics with quantum mechanics (QM) calculations. Specifically, occurrences and structural context of modified base pairs (MBPs), i.e. base pairs featuring posttranscriptional modifications, are collected from the RNA structures in the PDB and classified by bioinformatics tools. Then, QM calculations are performed to clarify the effect of the modification on the geometry and stability of the corresponding base pair. We have applied this approach over time to both natural and non-natural (synthetic) modifications (see for instance [6-7]) and, in 2015, we have presented an atlas of MBPs, i.e. a systematic study of all the MBPs in RNA experimental structures [8]. At the time, we could identify a total of 900 occurrences for 11 natural modifications, with roughly half of them involved in base pairing. Our atlas 1.0 consisted of 27 MBPs, unique in terms of identity of H-bonded bases and/or geometry classification.

Herein, to extend our understanding of how posttranscriptional modifications act on the structure of RNA molecules to influence their function, we present an updated atlas, derived from an over doubled structures dataset. It consists overall of almost 100 unique MBPs, featuring 35 different posttranscriptional modifications, located in a variety of different RNA molecules and structural motifs. Consistently with our previous findings, most of the MBPs are non Watson–Crick like and are in-

volved in RNA tertiary structure motifs. Results of the structural analyses, along with insight from QM calculations into the impact of the different modifications on the geometry and stability of the corresponding base pairs, will be presented and discussed.

1. P. Boccaletto, F. Stefaniak, A. Ray, A. Cappannini, S. Mukherjee, *et al*., *Nucleic Acids Res.*, **50**, (2022), D231.
2. P. F. Agris, *RNA*, **21**, (2015), 552.
3. M. Helm, *Nucleic Acids Res.*, **34**, (2006), 721.
4. K. I. Zhou, M. Parisien, Q. Dai, N. Liu, L. Diatchenko, J. R. Sachleben, T. Pan, *J. Mol. Biol*, **428**, (2016), 822.
5. N. B. Leontis, J. Stombaugh, E. Westhof, *Nucleic Acids Res.*, **30**, (2002), 3497.
6. R. Oliva, L. Cavallo, A. Tramontano, *Nucleic Acids Res.*, **34**, (2006), 865.
7. M. Chawla, S. Gorle, A. R. Shaikh, R. Oliva, L. Cavallo, *Comput. Struct. Biotechnol. J.,* **19**, (2021), 1312.
8. M. Chawla, R. Oliva, J. M. Bujnicki, L. Cavallo, *Nucleic Acids Res.*, **43**, (2015), 6714.

L19

RNADVISOR: EVALUATION OF RNA 3D STRUCTURES WITH METRICS AND ENERGIES

C. Bernard[1], G. Postic[1], S. Ghannay[2], F. Tahi[1]

[1]Université Paris-Saclay, Univ. Evry, IBISC, 91200, Evry-Courcouronnes, France
[2]Université Paris-Saclay, CNRS, LISN, 91405, France
fariza.tahi@univ-evry.fr

RNA adopts three-dimensional structures that play a crucial and direct role in its biological function. Understanding these diverse functions is necessary for the development of RNA-based therapies, but the complex structure of RNA molecules remains a major challenge. Computational methods have been developed throughout the years to fill the gap between the huge amount of known RNA sequences and their structures. With the increased number of RNA structures that are still to be discovered, predictive methods need to be robust and to be able to generalize to unseen new RNA families.

While structure predictions are a vast and complex problem, the evaluation and assessment of structure nativity is also at stake. RNA structure is a 3D object where the evaluation of a prediction has been discussed for years. Current methods rely on the comparison of a reference solved structure with a prediction, categorised as metrics. It can compare deviation on atoms like RMSD or RMSD [1], or overlaps between them like CLASH score [2]. Other metrics are inspired by protein 3D evaluation metrics from the CASP competition. Indeed, RNA and protein 3D structures share common properties as 3D objects and adaptation of the known protein's metrics like TM-score [3] can be done to RNA. It remains structural differences between protein and RNA molecules that hamper the full efficiency of structural evaluation metrics. RNA-oriented metrics have been developed to take full advantage of structural specificities like INF [4] or MCQ [5] scores.

Nonetheless, the metrics rely on a known solved structure, which in practice is not available. Predictive models are also based on the generation of multiple structures before selecting the best ones. Common approaches are thus to replicate the molecule free energy, where a minimum of energy would mean a stabilisation in the structure. This adaptation of the free energy of the structure has become a standard in the ranking, filtering and confidence assessment of structures. It often uses knowledge-based statistical potentials, with the requirement of a reference state to simulate structures without native interactions. This is the case for NAST [6], 3dRNAScore [7], DFIRE-RNA [8] and rsRNASP [9]. Recent advances tend to use deep learning to prevent manual pre-processing of RNA features like RNA3DCNN [10] or ARES [11].

RNA 3D structures remain of high complexity, and there is not a single existing metric or energy that could evaluate correctly all the available structures. Metrics and energies can be redundant between each other, while also complementary for structure assessment. The different existing metrics can be required to develop and understand predictive models' weaknesses, while the diverse energies could help improve models' generation such as the filtering process.

The current metrics and energies are the results of years of research by various groups. Each work has been developed in different programming languages, with different installation procedures and library versions that have evolved over the years. The installation process can be laborious for the community and is multiplied by the number of different metrics and energies. Efforts should be made on developing predictive models while engineering aspects for structures assessment should not be a bottleneck. Works have been done by the community with the development of RNAPuzzles [12], a CASP-like competition for RNA 3D structure assessment. It comes with RNA-tools [13], a centralised platform that tries to include the available RNA 3D structure related works. Nonetheless, it is limited in practice with the need to manually include binary files; that depend on the operating system of the user. There are also web servers available for some metrics and energies, which are useful for non-coder users. However, it limits the automation procedure, which should be considered due to the increasing number of solved 3D structures.

To help the development and the automation of RNA 3D structures evaluation, we have developed RNAdvisor: a software usable with one command line and that can compute both metrics and energies for given RNA. It uses eight

existing codes written in C++, Java or Python and gathers them into a single interface. All the laborious installations are done in different stages of the Dockerfile. It leverages Docker containers for easy installation across diverse operating systems, simplifying accessibility for all researchers. It enables researchers to access both metrics and energies in one line of code, with customizable parameters to suit individual preferences.

RNAdvisor represents a significant advancement for the automation of RNA 3D structure evaluation. It offers a unified tool that enhances the accessibility of existing metrics and energies. It helps accelerate investigation in RNA 3D structure predictions.

The source code is available at:
https://github.com/EvryRNA/rnadvisor.

1. Bottaro S, Di Palma F, and Bussi G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic acids research 2014*;42.
2. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research* 2007; 35: W375–W383.
3. Zhang Y and Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins* 2004;57:702–10.
4. Parisien M, Cruz J, Westhof E, and Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* (New York, N.Y.) 2009;15:1875–85.
5. Zok T, Popenda M, and Szachniuk M. MCQ4Structures to compute similarity of molecule structures. *Central European Journal of Operations Research* 2013;22.
6. Jonikas MA, Radmer RJ, Laederach A, et al. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA 2* 2009;15:189–99.
7. Wang J, Zhao Y, Zhu C, and Xiao Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Research 10* 2015;43:e63–e63.
8. Capriotti E, Norambuena T, Marti-Renom MA, and Melo F. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* 2011; 27:1086–93.
9. Tan YL, Wang X, Shi YZ, Zhang W, and Tan ZJ. rsRNASP: A residue-separation based statistical potential for RNA 3D structure evaluation. *Biophysical Journal 1* 2022;121:142–56.
10. Li J, Zhu W, Wang J, et al. RNA3DCNN: Local and global quality assessments of RNA 3D structures using 3D deep convolutional neural networks. *PLOS Computational Biology* 2018;14:1–18.
11. Townshend RJL, Eismann S, Watkins AM, et al. Geometric deep learning of RNA structure. *Science 6558* 2021;373:1047–51.
12. Cruz J, Blanchet MF, Boniecki M, et al. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. *RNA (New York, N.Y.)* 2012;18:610–25.
13. Magnus M. rna-tools.online: a Swiss army knife for RNA 3D structure modeling workflow. *Nucleic Acids Research 2022*;50:W657–W662.

L20

PREDICTION OF SECONDARY STRUCTURE FOR LONG NON-CODING RNAS USING A RECURSIVE CUTTING METHOD BASED ON DEEP LEARNING

L. Omnes[1], E. Angel[1], P. Bartet[3], F. Radvanyi[2], F. Tahi[1]

[1]Université Paris-Saclay, Univ Evry, IBISC, 91120 Evry-Courcouronnes, France
[2]CNRS – Institut Curie, 75000 Paris, France
[3]ADLIN Science, 75000 Paris, France
fariza.tahi@univ-evry.fr

While the primary structure of RNA is defined by its sequence of nucleotides, the secondary structure refers to the pairings that occur between the nucleotides. The secondary structure emerges from the interactions between complementary bases. The secondary structure is crucial as it determines the overall shape and stability of the RNA molecule, which, in turn, influences its function.

The accurate prediction of RNA secondary structure, particularly for long non-coding RNAs (lncRNAs), holds immense potential in healthcare. It can be used for diagnostic, therapeutic, and drug discovery purposes, enabling precision medicine approaches, and developing targeted therapies for various diseases. A better understanding of their structures could improve disease diagnosis and treatment, notably for cancer, and allow for novel therapeutic interventions in the future. However, the majority of previous approaches [5-9] have focused on short RNAs and are too costly in terms of computation budget to cope with the increasing complexity of long RNAs. Plus, the ones that can scale to long RNAs lack accuracy to reliably predict their structures.

We propose a new approach combining recursive cutting and machine learning. By leveraging existing successful methods for small RNAs and introducing innovative cut point selection, this approach aims to improve the accuracy and efficiency of long RNA structure prediction. Our model uses as input only the RNA sequence, without the need for homologous sequences which are often not available for long RNAs. Our method proves to be computationally efficient by recursively partitioning a se-

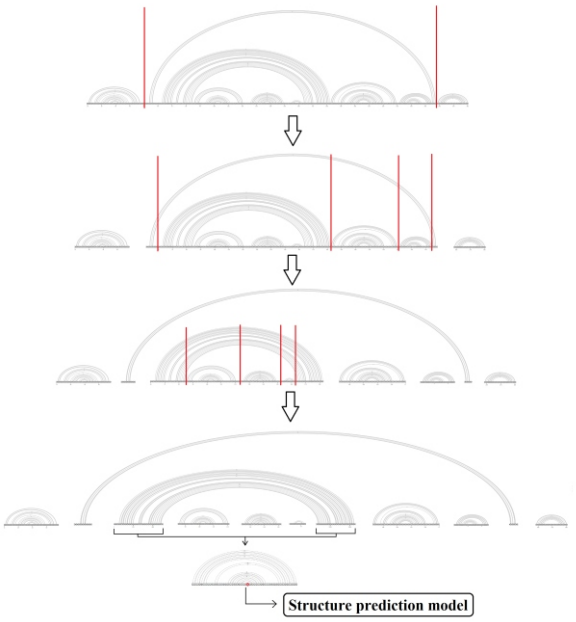

Figure 1. Successive iterations of the recursive cutting strategy. Successive steps occur until all fragments are small enough to be sent to the structure prediction model. At each step, left-most and right-most parts are combined to form a single fragment.

quence into smaller fragments until they can be easily managed by an existing model. We use deep learning to search for cut points in a linear time complexity. A visual example of the successive iterations of the partitioning strategy can be seen in Fig. 1. We used the bpRNA_1m [11] dataset for training and for our experiments.

We perform a benchmark of MXfold2 [7], Linearfold [3] and our proposed approach on the bpRNA_1m [11] database and show that our approach indeed demonstrates better performance for long RNAs and a potential to bring significant improvements in the future, as well as interesting enhancing properties, which we discuss.

1. M. Zuker and P. Stiegler, “Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information,” Nucleic Acids Research, vol. 9, pp. 133–148, Jan. 1981.2. P. J. Chupas, M. F. Ciraolo, J. C. Hanson, C. P. Grey, *J. Am. Chem. Soc.*, **123**, (2001), 1694.
2. S. McCaskill, “The equilibrium partition function and base pair binding probabilities for RNA secondary structure,” Biopolymers, vol. 29, no. 6-7, pp. 1105–1119, 1990.4. D. Balzar & N. C. Popa, in *Diffraction Analysis of the Microstructure of Materials*, edited by E.J. Mittemeijer & P. Scardi (Berlin: Springer), 2004, pp. 125-145. 3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
3. L. Huang, H. Zhang, D. Deng, K. Zhao, K. Liu, D. A. Hendrix, and D. H. Mathews, “LinearFold: linear-time approximate RNA folding by 5’-to-3’ dynamic programming and beam search,” Bioinformatics, vol. 35, pp. i295–i304, July 2019.3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
4. S. Engelen and F. Tahi, “Tfold: efficient in silico prediction of non-coding RNA secondary structures,” Nucleic Acids Research, vol. 38, pp. 2453–2466, Apr. 2010.3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
5. L. Fu, Y. Cao, J. Wu, Q. Peng, Q. Nie, and X. Xie, “UFold: fast and accurate RNA secondary structure prediction with deep learning,” Nucleic Acids Research, vol. 50, p. e14, Feb. 2022. 3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
6. J. Singh, K. Paliwal, T. Zhang, J. Singh, T. Litfin, and Y. Zhou, “Improved RNA secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning,” Bioinformatics (Oxford, England), vol. 37, pp. 2589–2600, Sept. 2021.3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
7. K. Sato, M. Akiyama, and Y. Sakakibara, “RNA secondary structure prediction using deep learning with thermodynamic integration,” Nature Communications, vol. 12, p. 941, Feb. 2021. Number: 1 Publisher: Nature Publishing Group. 3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.
8. M. Hamada, H. Kiryu, K. Sato, T. Mituyama, and K. Asai, “Prediction of RNA secondary structure using generalized centroid estimators,” Bioinformatics (Oxford, England), vol. 25, pp. 465–473, Feb. 2009. 4. D. Balzar & N. C. Popa, in *Diffraction Analysis of the Microstructure of Materials*, edited by E.J. Mittemeijer & P. Scardi (Berlin: Springer), 2004, pp. 125-145.
9. M. Saman Booy, A. Ilin, and P. Orponen, “RNA secondary structure prediction with convolutional neural networks,” BMC Bioinformatics, vol. 23, p. 58, Feb. 2022.
10. A. Legendre, E. Angel, and F. Tahi, “Bi-objective integer programmingfor RNA secondary structure prediction with pseudoknots,” BMC bioinformatics, vol. 19, p. 13, Jan. 2018.
11. P. Danaee, M. Rouches, M. Wiley, D. Deng, L. Huang, and D. Hendrix, “bpRNA: large-scale automated annotation and analysis of RNA secondary structure,” Nucleic Acids Research, vol. 46, pp. 5381–5394, June 2018. 3. H. J. Bunge, *Texture Analysis in Materials Science*. London: Butterworth. 1982.

Thursday, November 16, S5 - Activity 2, 5

Additional talks

L21

PROTEIN QUATERNARY STRUCTURES IN SOLUTION ARE A MIXTURE OF MULTIPLE FORMS

Gideon Schreiber[1], Shir Marciano[1], Debabrata Dey[1], Dina Listov[1], Sarel J Fleishman[1], Adar Sonn-Segev[2], Haydyn Mertens[3], Florian Busch[4], Yongseok Kim[4], Sophie R. Harvey[4], Vicki H. Wysocki[4]

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel
[2]Refeyn Ltd, 1 Electric Avenue, Ferry Hinksey Road, Oxford OX2 0BY, UK
[3]Hamburg Outstation, European Molecular Biology Laboratory, Notkestrasse 85, Hamburg, 22607, Germany
[4]Department of Chemistry and Biochemistry and Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH, 43210, USA

Over half the proteins in the *E.coli* cytoplasm form homo or hetero-oligomeric structures. Experimentally determined structures are often considered in determining a protein's oligomeric state, but static structures miss the dynamic equilibrium between different quaternary forms. The problem is exacerbated in homo-oligomers, where the oligomeric states are challenging to characterize. Here, we re-evaluated the oligomeric state of 17 different bacterial proteins across a broad range of protein concentrations and solutions by native mass-spectrometry (MS), mass photometry (MP), size exclusion chromatography (SEC), and small-angle X-ray scattering (SAXS), finding that most exhibit several oligomeric states. Surprisingly, many proteins did not show mass-action driven equilibrium between the oligomeric states. For approximately half the proteins, the predicted oligomeric forms described in publicly available databases underestimated the complexity of protein quaternary structures in solution. Conversely, AlphaFold Multimer provided an accurate description of the potential multimeric states for most proteins, suggesting that it could help resolve uncertainties on the solution state of many proteins.

1. Marciano S, Dey D, Listov D, Fleishman SJ, Sonn-Segev A, Mertens H, Busch F, Kim Y, Harvey SR, Wysocki VH, Schreiber G. Protein quaternary structures in solution are a mixture of multiple forms. *Chem Sci*, **13** (2022), 11680.

L22

STRUCTURAL PLASTICITY IN THE LOOP REGION OF ENGINEERED LIPOCALINS WITH NOVEL LIGAND SPECIFICITIES – ANTICALINS

A. Skerra

Lehrstuhl für Biologische Chemie, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany
skerra@tum.de

Anticalins are generated via combinatorial protein design on the basis of the lipocalin protein scaffold and constitute a novel class of small and robust binding proteins. These engineered lipocalins offer prospects as an alternative to antibodies for applications in medical therapy as well as *in vivo* diagnostics. The lipocalins are natural binding proteins with diverse ligand specificities which share a simple architecture with a central eight-stranded antiparallel -barrel and an -helix attached to its side. At the open end of the -barrel, four structurally variable loops connect the -strands in a pair-wise manner and, together, shape the ligand pocket. Using targeted random mutagenesis in combination with molecular selection techniques, this loop region can be reshaped to generate pockets for the tight binding of various ligands ranging from small molecules over peptides to proteins. While such Anticalin proteins can be derived from different natural lipocalins, the human lipocalin 2 (Lcn2) scaffold proved particularly successful for the design of binding proteins with novel specificities and, over the years, more than 20 crystal structures of Lcn2-based Anticalins have been elucidated. Using a novel way of graphical representation, the conformational variability that emerged in the loop region of these functionally diverse artificial binding proteins can be illustrated in comparison with the natural scaffold. This analysis has provided picturesque evidence of the high structural plasticity around the binding site of the lipocalins which explains their proven tolerance toward excessive mutagenesis. Furthermore, apart from a simple lock-and-key mode of ligand recognition, structural evidence suggests two distinct

mechanisms of spatial adaptation during the formation of Anticalin-ligand complexes: (i) induced fit, in which conformational alteration follows ligand binding, and (ii) conformational selection, which is based on a pre-existing mixture of conformational states. Taken together, these molecular mechanisms demonstrate remarkable resemblance between the binding site of lipocalins (natural or engineered) and the well characterized complementarity-determining region of immunoglobulins (antibodies), which represent two structurally and functionally different types of mammalian plasma proteins.

1. Richter, A., Eggenstein, E. & Skerra, A. (2014) Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* 588, 213-218.
2. Deuschle, F. C., Ilyukhina, E. & Skerra, A. (2021) Anticalin® proteins: from bench to bedside. *Expert Opin. Biol. Ther.* 21, 509-518.
3. Achatz, S., Jarasch, A. & Skerra, A. (2021) Structural plasticity in the loop region of engineered lipocalins with novel ligand specificities, so-called Anticalins. *J. Struct. Biol. X.* 6, 100054.
4. Jerschke, E., Eichinger, A. & Skerra, A. (2023) Drastic alterations in the loop structure around colchicine upon complex formation with an engineered lipocalin indicate a conformational selection mechanism. *Acta Crystallogr. F Struct. Biol. Commun.* 79, 231-239.

Thursday, November 16, S5 - Activity 5

To establish a Biostudies database of protein engineering results

L23

NOVEL STRATEGIES AND WEB-BASED TOOLS FOR PROTEIN ENGINEERING

Jiri Damborsky[1,2], David Bednar[1,2], Stanislav Mazurenko[1,2], Zbynek Prokop[1,2], Martin Marek[1,2], Simeon Borko[1,2], Jan Stourac[1,2], Milos Musil[1,2], Jan Mican[1,2], Sergio M. Marques[1,2], Joan Planas-Iglesias[1,2], Petr Kabourek[1,2], Petr Kouba[1,3], Anton Bushiev[3], Jiri Sedlar[3], Josef Sivic[3]

[1]Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
[2]International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic
[3]Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Jugoslávských partyzánů 1580/3, 160 00 Prague, Czech Republic
jiri@chemi.muni.cz

We develop novel strategies and web-based protein engineering tools under the ELIXIR Czech Republic umbrella. These are fully automated computational workflows which can operated using the intuitive graphical user interface [1]. Protein sequence or structure is typically the only input required for the calculation. The tools can be accessed freely via the Protein Engineering Portal (**Figure 1**). The tools are particularly suitable for experimentalists without prior structural biology or bioinformatics knowledge. The National Supercomputing Centre IT4Innovations provides the infrastructure for high-performance computing. This talk will introduce some of our web tools and illustrate their use for engineering proteins for biotechnological and biomedical applications [2].

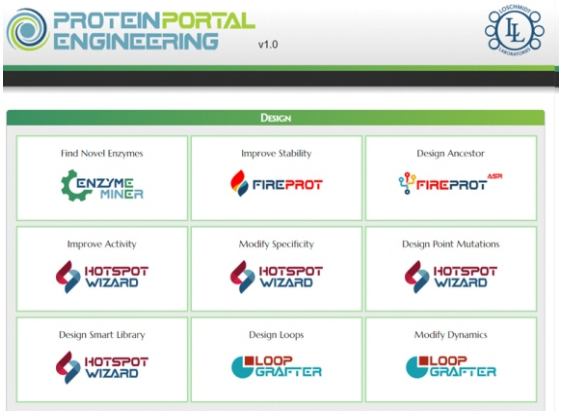

Figure 1. The graphical user interface of the Protein Engineering Portal provides unified access to the software tools and databases developed by the Loschmidt Laboratories and partners: https://loschmidt.chemi.muni.cz/portal/.

1. Marques, S. M., Planas-Iglesias, J., Damborsky, J., 2021: Web-based Tools for Computational Enzyme Design. *Current Opinion in Structural Biology* **69**: 19-34.
2. Vasina, M., Velecky, J., Planas-Iglesias, J., Marques, S. M., Skarupova, J., Damborsky, J., Bednar, D., Mazurenko, S., Prokop, Z., 2022: Tools for Computational Design and High-Throughput Screening of Therapeutic Enzymes. Advanced *Drug Delivery Reviews* **183**: 114143.