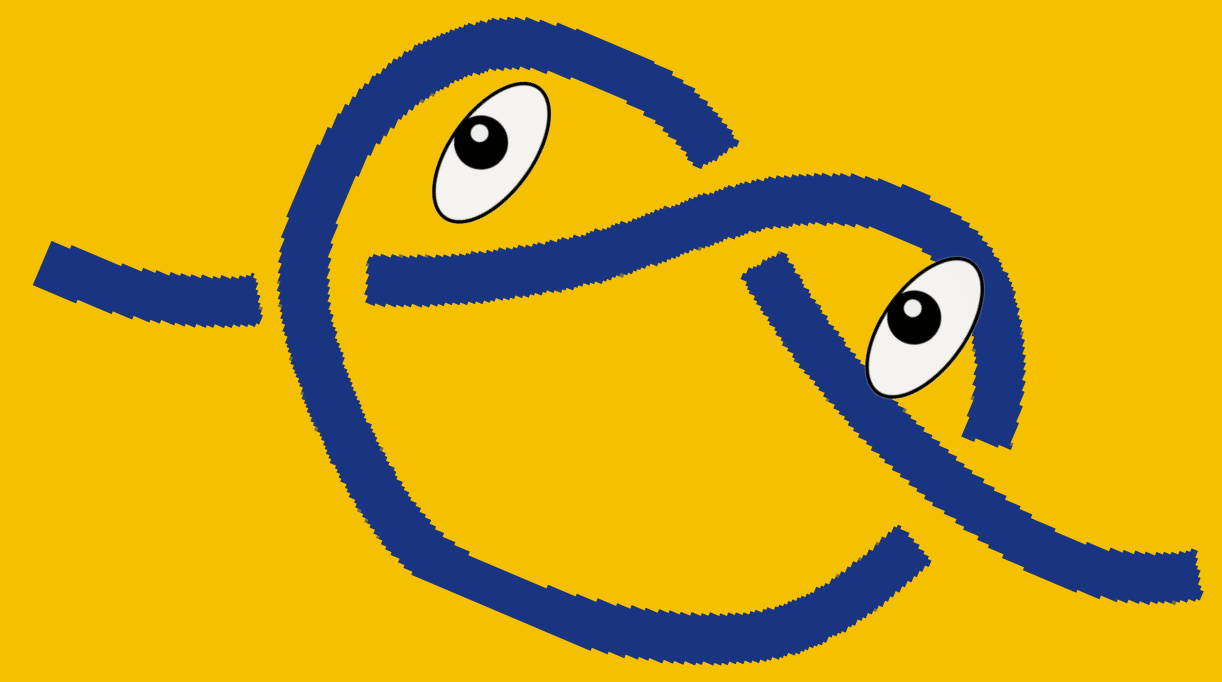


KNOTTING PATTERNS IN PROTEINS: INSIGHTS FROM RFDIFFUSION AND EvoDIFF



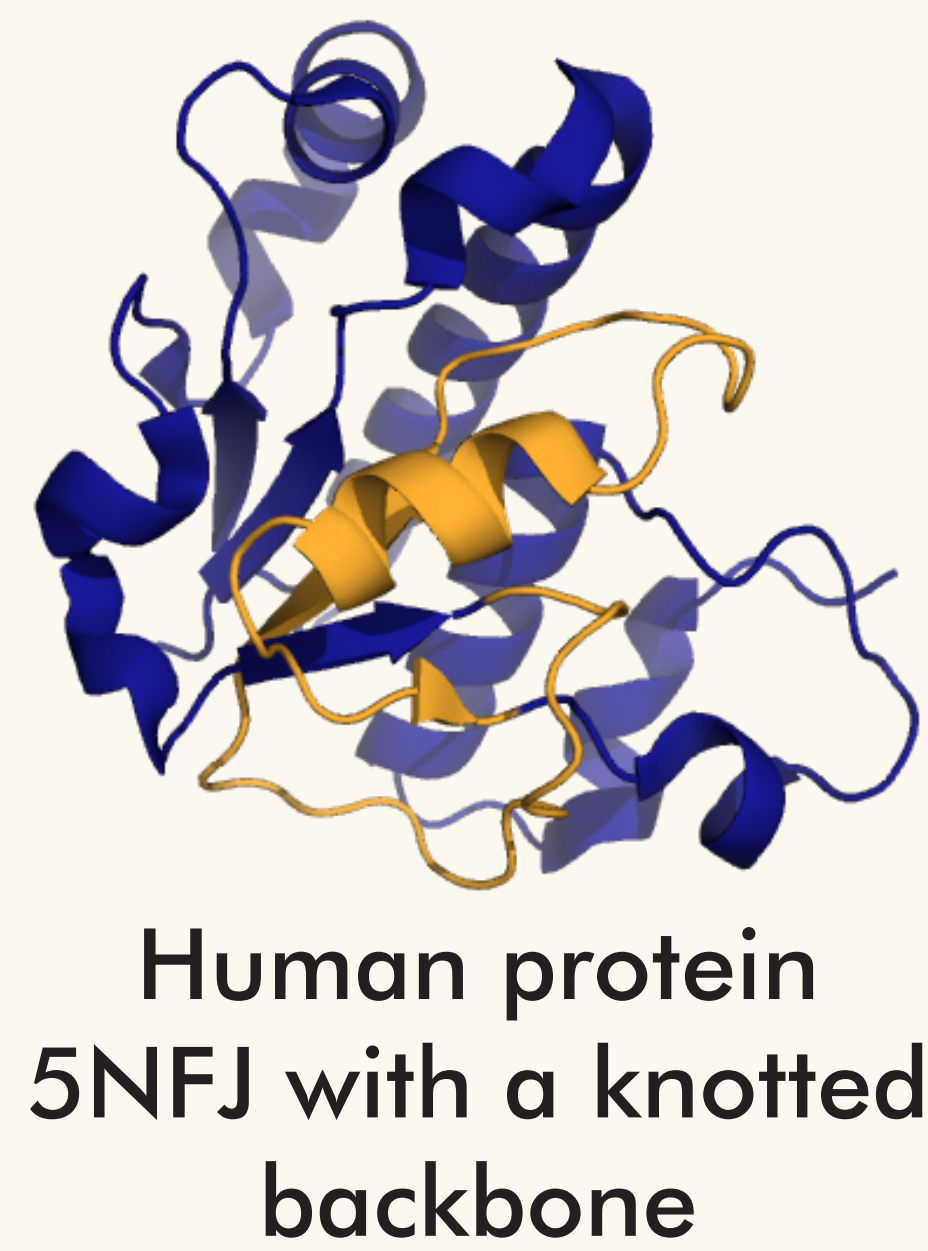
Eva Klimentová^{1,2}, Petr Šimeček²

[1] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, CZ

[2] Central European Institute of Technology, Masaryk University, Brno, CZ

WHY?

- proteins with a knotted backbone are a very rare phenomenon (0.35% of AFdb) [1]
- amino acid patterns responsible for the entanglement, or the role of the knot is still unknown [2]
- some knotted protein families hold potential for industrial and pharmaceutical applications (eg. the design of new antimicrobial drugs) [3]
- knotted proteins are clustered into only a few protein families [4]

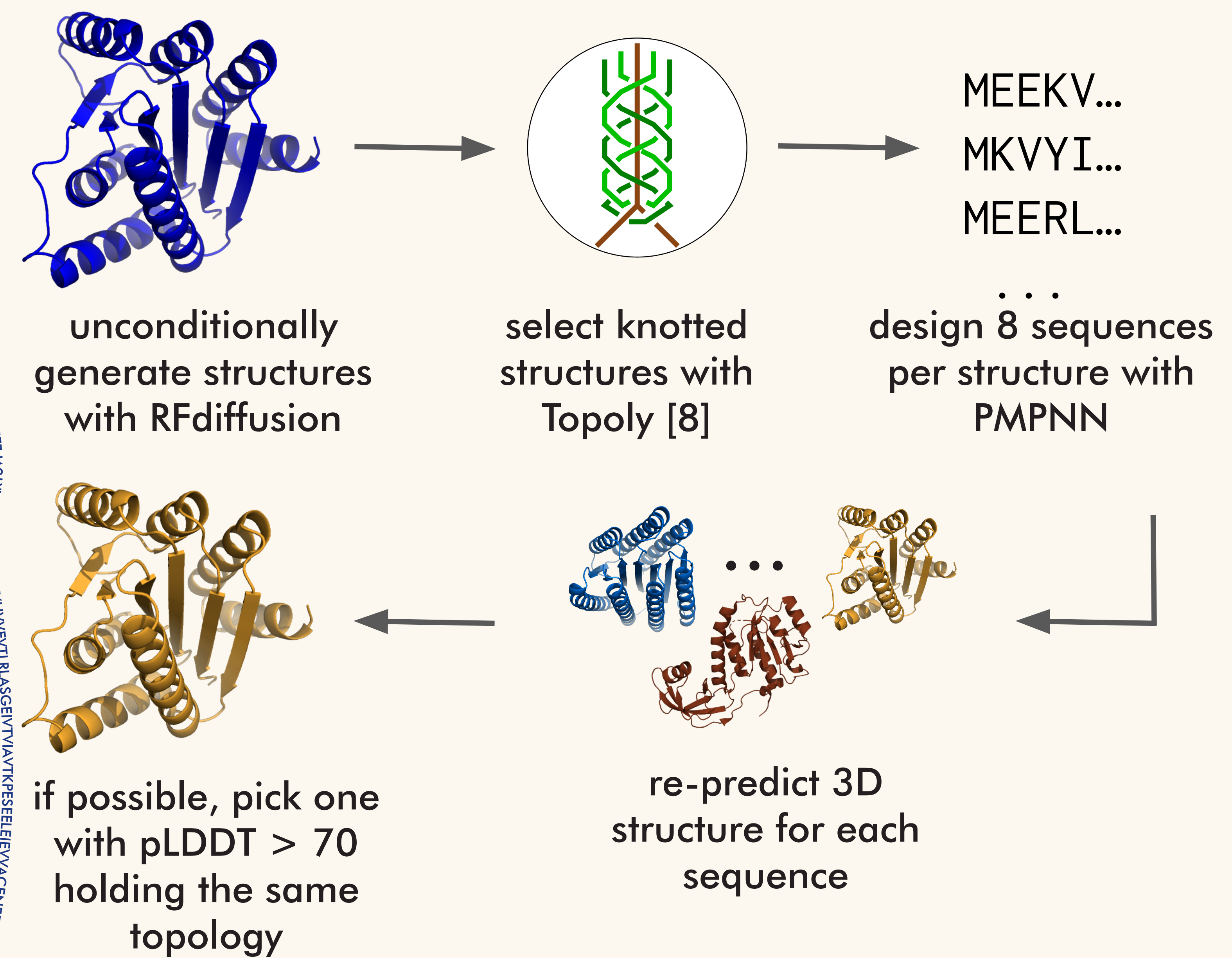


WHAT?

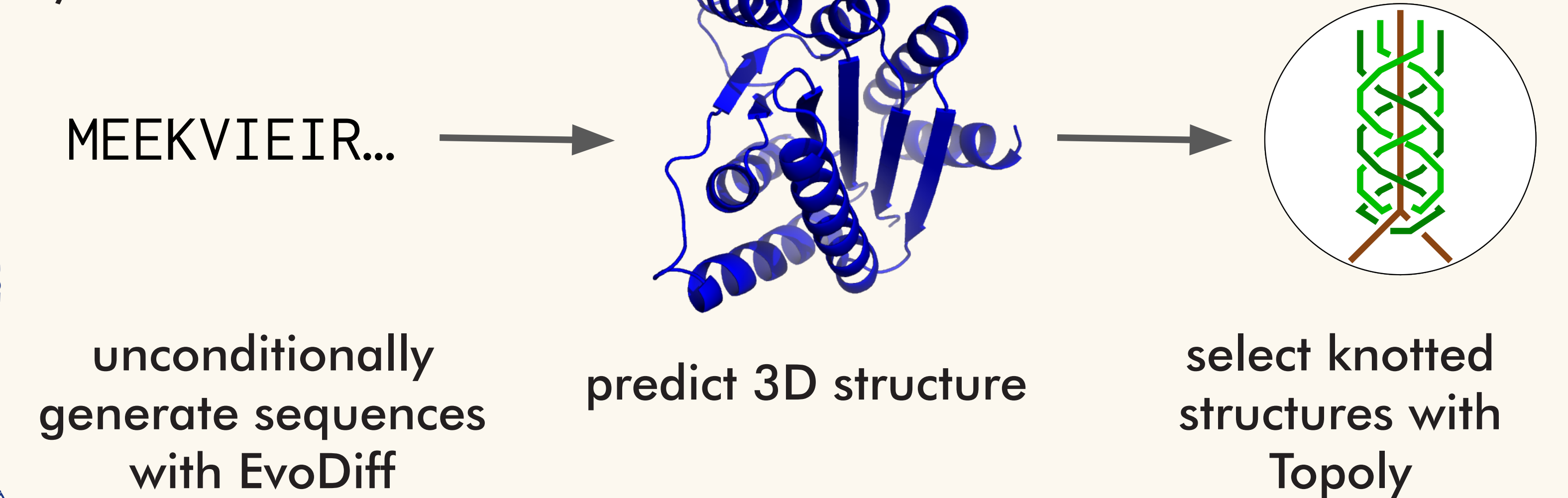
- we created a dataset of artificial knotted proteins by unconditionally generating proteins with RFDiffusion [5] + ProteinMPNN [6] and EvoDiff [7] and analyzing their knotting status
- approximately 0.5% (2 016) of the freely generated proteins (425 362) are knotted, which corresponds to the ratio in nature [4]
- most of the generated knotted proteins are 3_1 (the simplest knot), but we also have rare cases: $3_1\#3_1$ (composite knot), 5_1 , 7_4 and 8_{19}
- generated proteins are far from the natural sequence space
- structures of proteins from EvoDiff are different from the natural ones, but RFDiffusion+PMPNN generate similar protein structures

How?

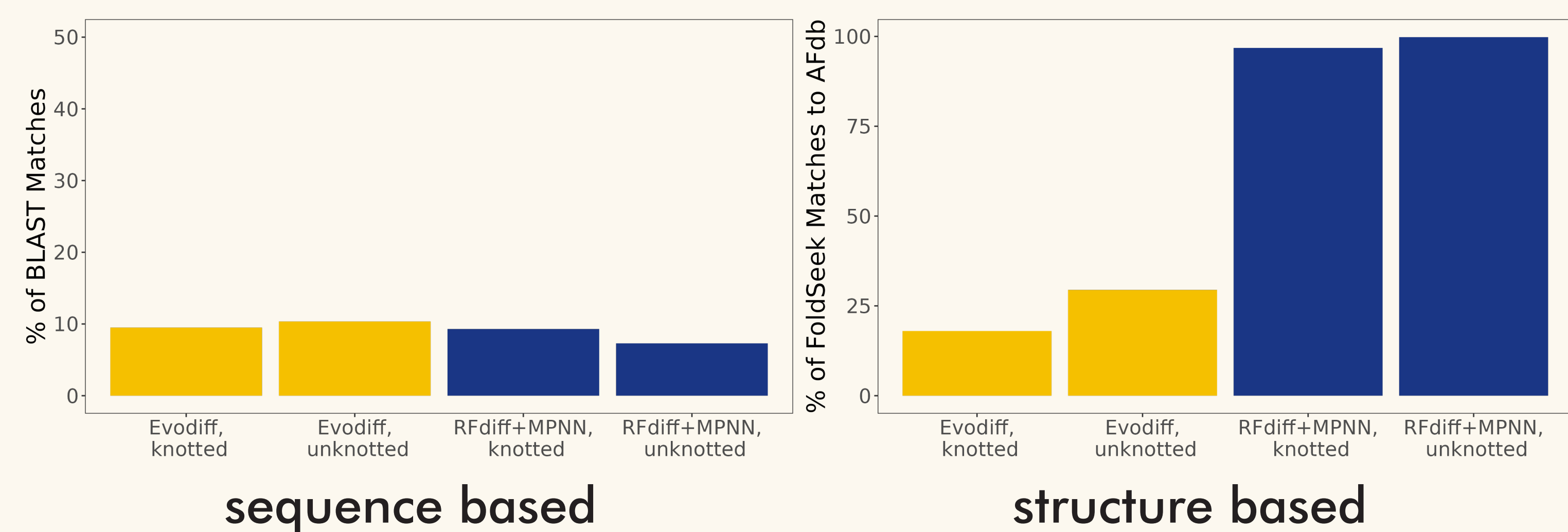
A) RFDiffusion + ProteinMPNN



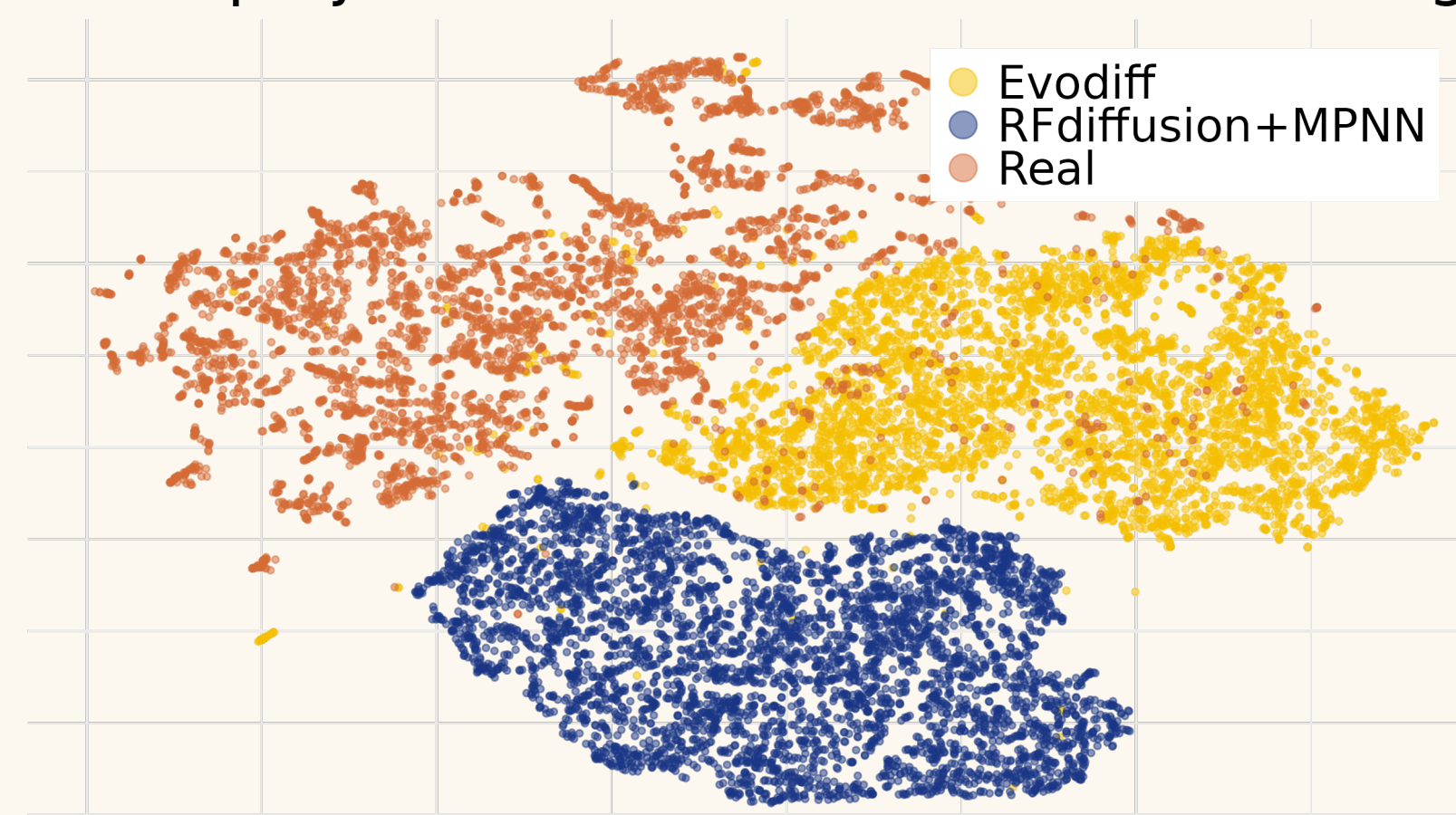
B) EvoDiff



Comparison of artificially generated proteins to real-world proteins



t-SNE projection of ProtBert-BFD embeddings



Mixture of knotted and unknotted proteins from EvoDiff, RFDiffusion+PMPNN and fraction of real proteins from UniProt

WHAT NEXT?

- we will attempt to crystallize one of the artificial knotted proteins
- we plan to enrich the dataset of real knotted proteins with those generated artificially and look for knotting patterns

REFERENCES

- [1] Virnau, P., et al. (2006). Intricate Knots in Proteins: Function and Evolution. PLOS Computational Biology.
- [2] Jackson, S. E.; Suma, A.; Micheletti, C. (2017). How to fold intricately: using theory and experiments to unravel the properties of knotted proteins. Current Opinion in Structural Biology.
- [3] Strassler, S., et al. (2022). Tied up in knots: Untangling substrate recognition by the SPOUT methyltransferases. Journal of Biological Chemistry.
- [4] Šrámková, D. et al. (2023) Knot or not? sequence-based identification of knotted proteins with machine learning. bioRxiv.
- [5] Watson J. L., et al. (2023). De novo design of protein structure and function with RFDiffusion. Nature.
- [6] Dauparas, J., et al. (2022). Robust deep learning-based protein sequence design using proteinmpnn. Science.
- [7] Alamdari, S., et al. (2023). Protein generation with evolutionary diffusion: sequence is all you need. bioRxiv.
- [8] Dabrowski-Tumanski, P., et al. (2020). Topoly: Python package to analyze topology of polymers. Briefings in Bioinformatics.

ACKNOWLEDGEMENT

We extend our profound gratitude to Joanna Sulkowska, Agata Perlinska and Maciej Sikora for their invaluable feedback and insights. This work has received funding from Czech Science Foundation, project no. 23-04260L. Computational resources were provided by the e-INFRA CZ project (ID:90140), supported by the Ministry of Education, Youth and Sports of the Czech Republic.



GitHub

MUNI

CEITEC

eva.klimentova@ceitec.muni.cz

rb.gy/hpep4r