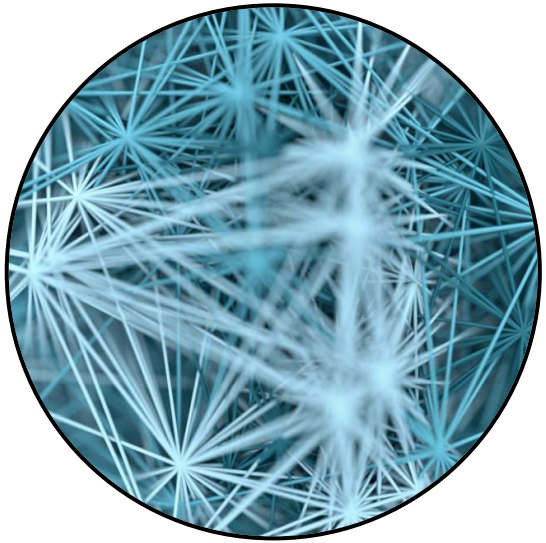
Can we create artificial knots with protein diffusion models?

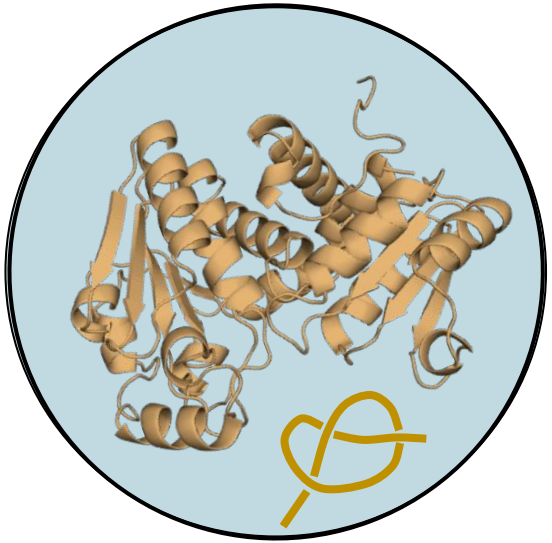
Eva Klimentová
18. 10. 2023

Outline

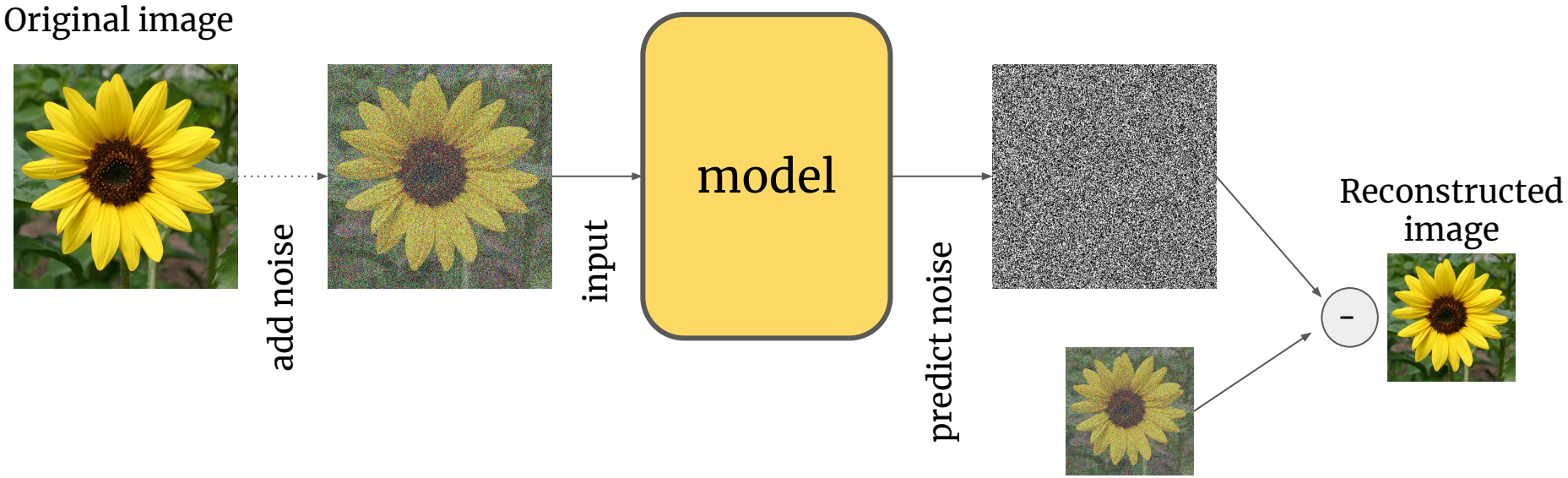
- How does diffusion work in general?
- What do we have for proteins?
- Can it make knots?
- Some other ideas and recent directions

Why do we need diffusion for proteins?

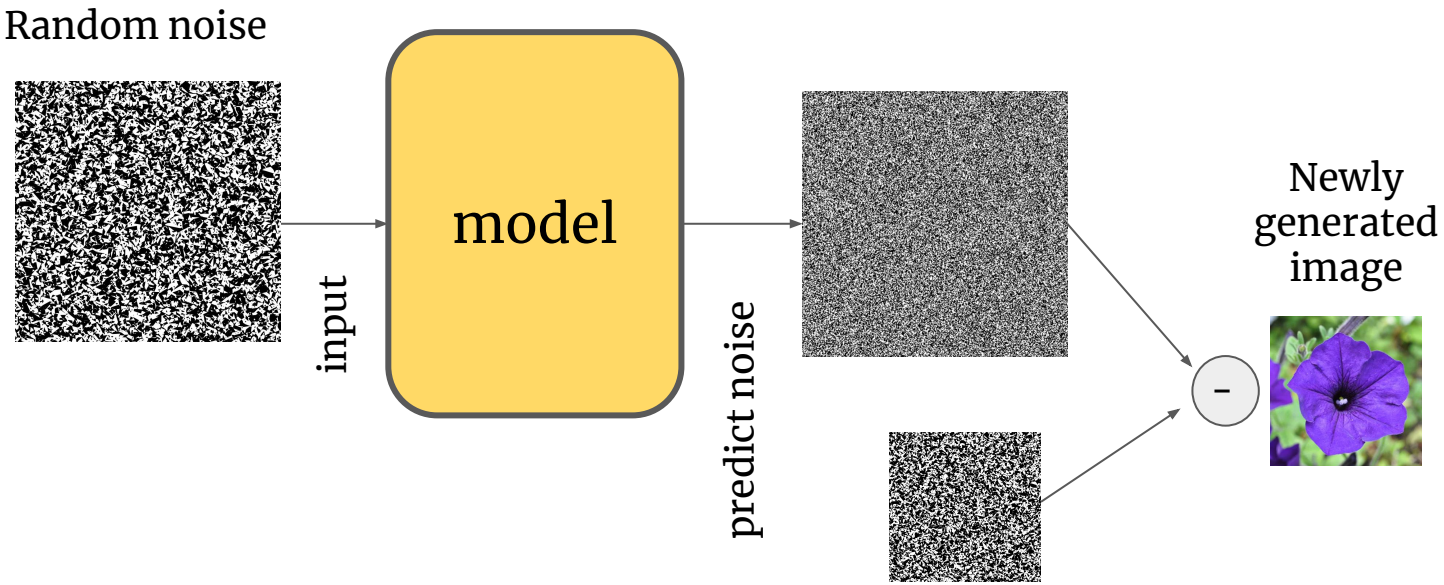




Diffusion ~ adding and removing noise

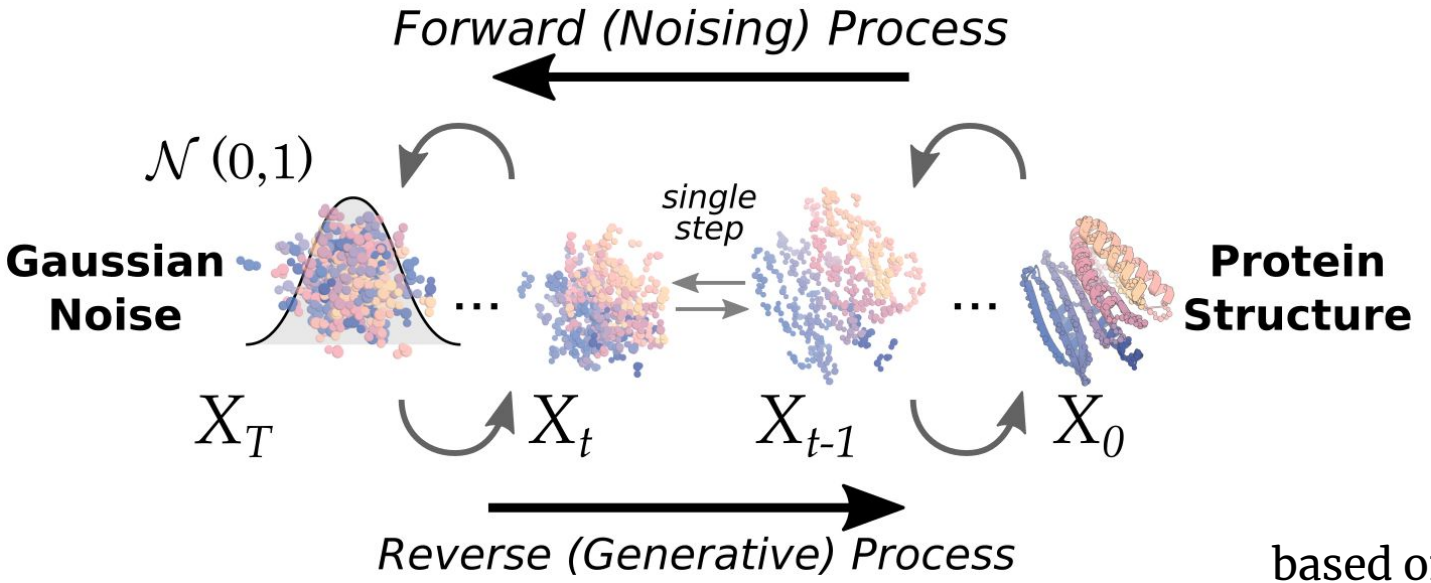


Using diffusion to generate new images



https://jalammar.github.io/illustrated-stable-diffusion/

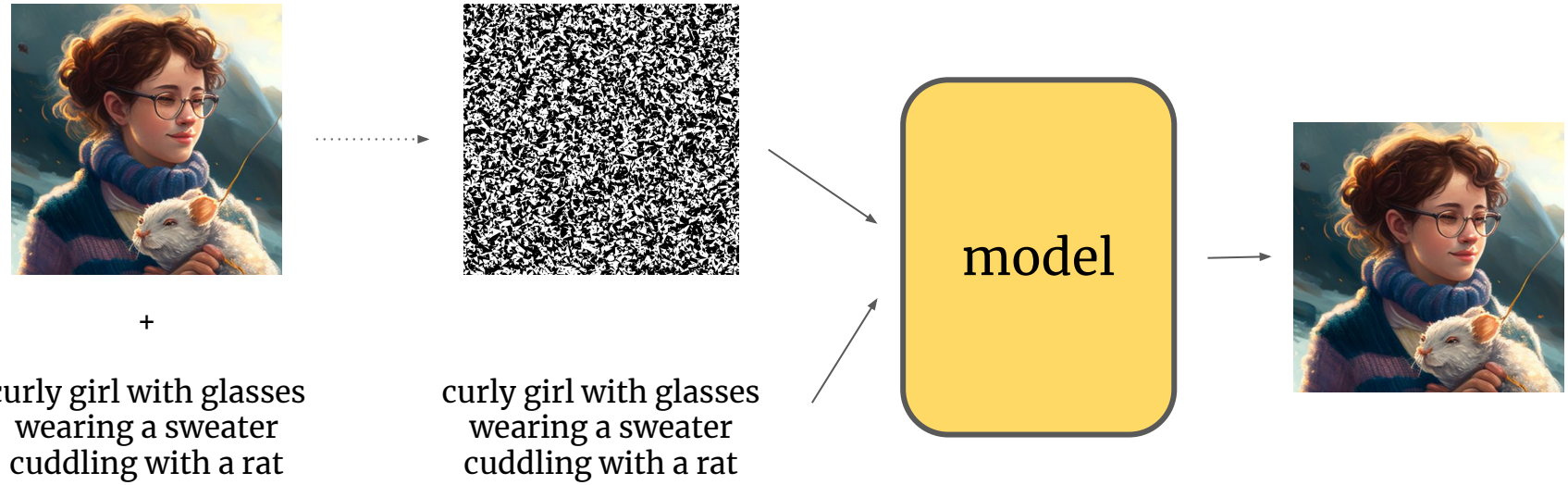
Diffusion for protein backbone: **RFdiffusion**



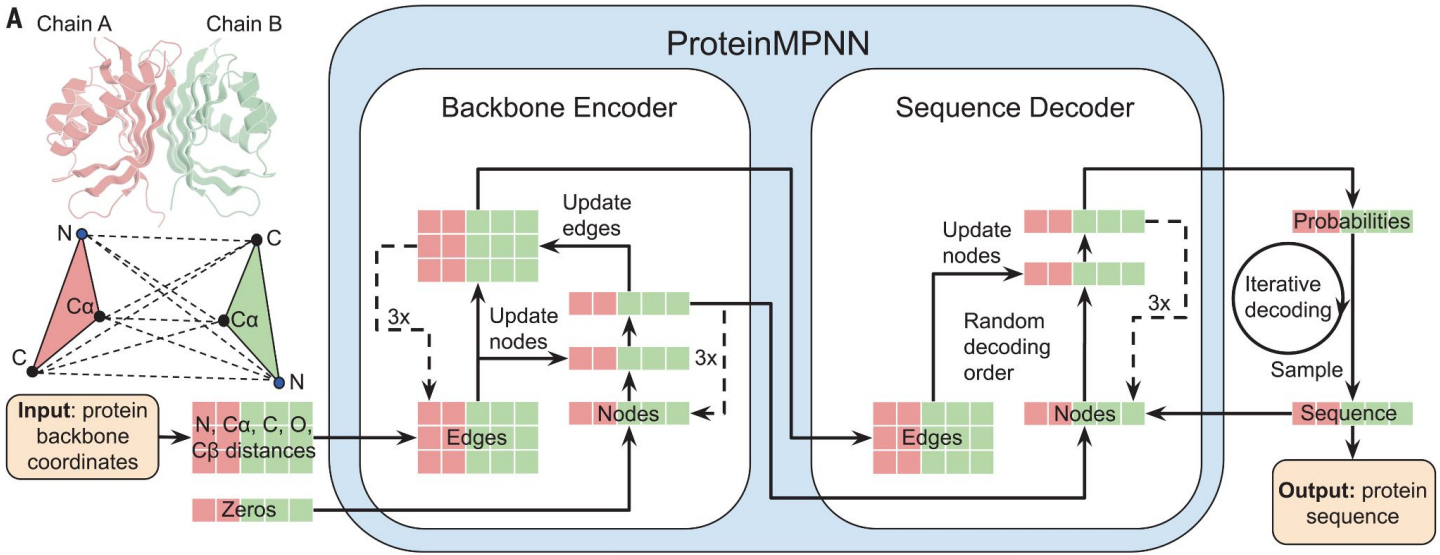
based on RoseTTAFold model

https://github.com/RosettaCommons/RFdiffusion
https://www.nature.com/articles/s41586-023-06415-8

Diffusion for protein sequence based on structure: **ProteinMPNN**



Diffusion for protein sequence based on structure: **ProteinMPNN**

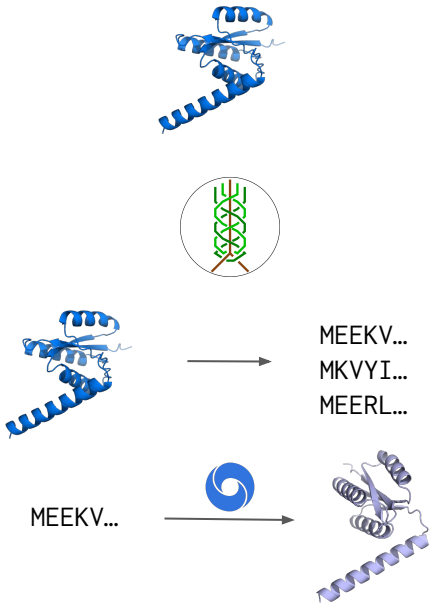


https://github.com/dauparas/ProteinMPNN/tree/main
https://www.science.org/doi/10.1126/science.add2187
https://huggingface.co/spaces/simonduerr/ProteinMPNN

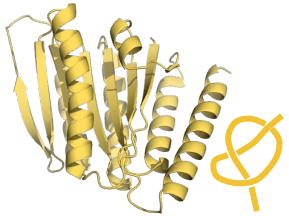
Generating random proteins with RFdiffusion + MPNN

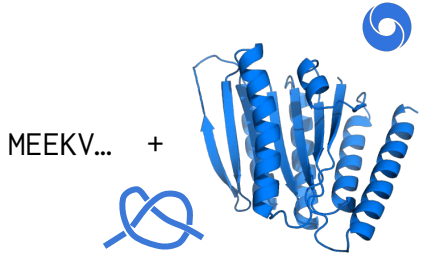
Workflow:

1. design random structure with RFdiffusion
2. check it’s topology
3. design multiple sequences with MPNN based on structure
4. predict structure from sequences
    - check pLDDT
    - check topology again



Generating random proteins with RFdiffusion + MPNN



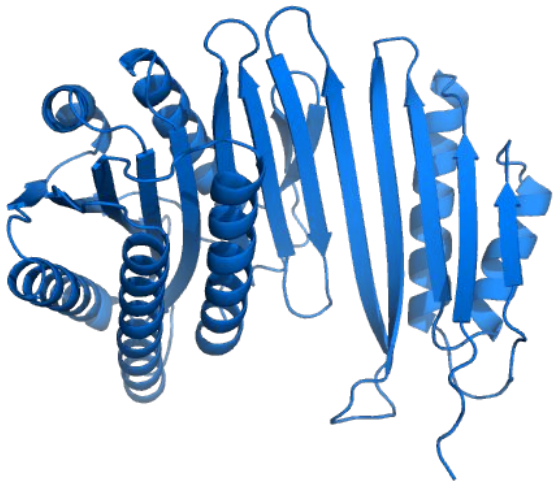


1. Total number of designed structures	2. Non-trivial topology in designed structure	4. Successful design of sequence with non-trivial topology
212 681	2 814	1 037

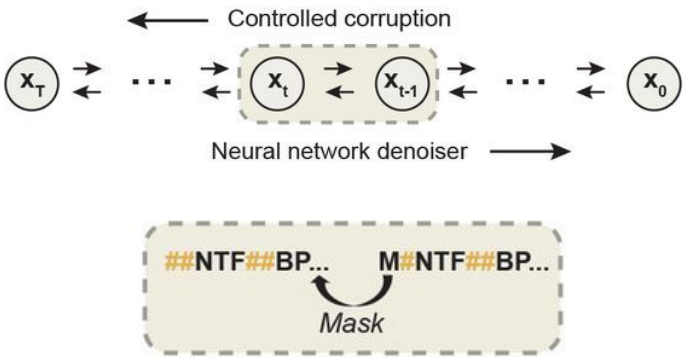
~ 1 %

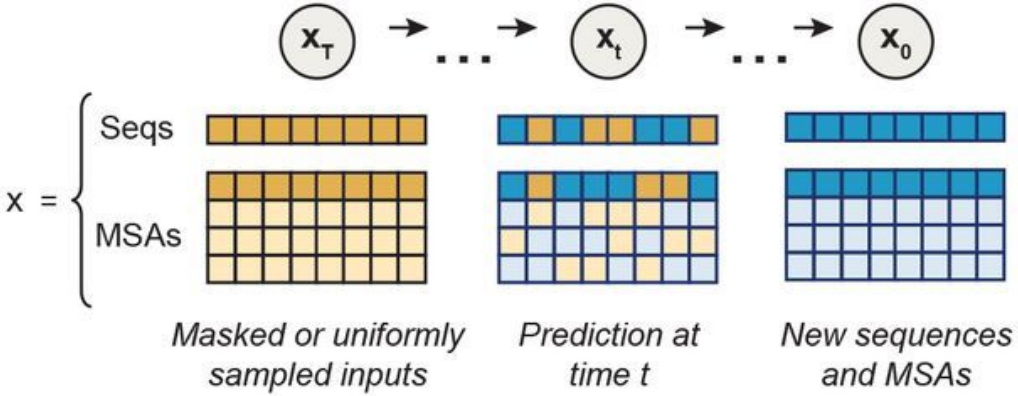
Generated knotted proteins

Major knot type	Number of designs
3_1	985
5_1	28
3_1#3_1	5
8_19	3 👀
and others	



EvoDiff – directly generate sequence

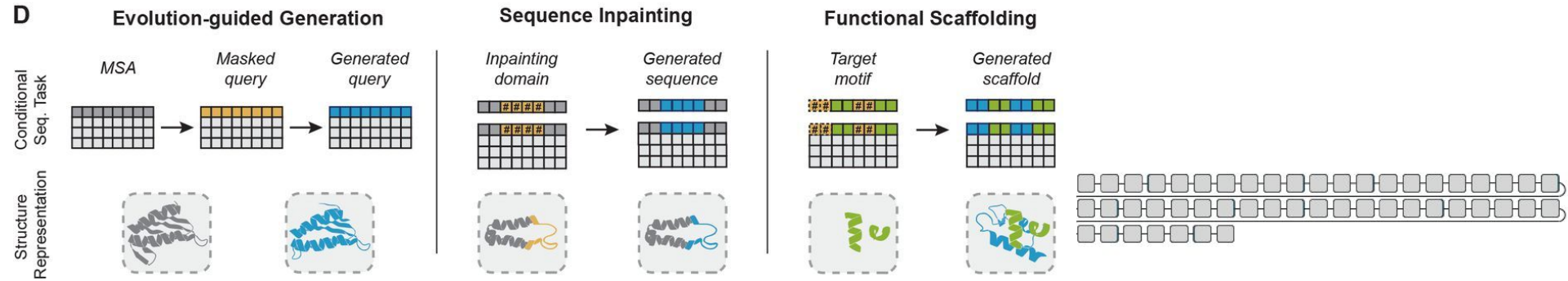


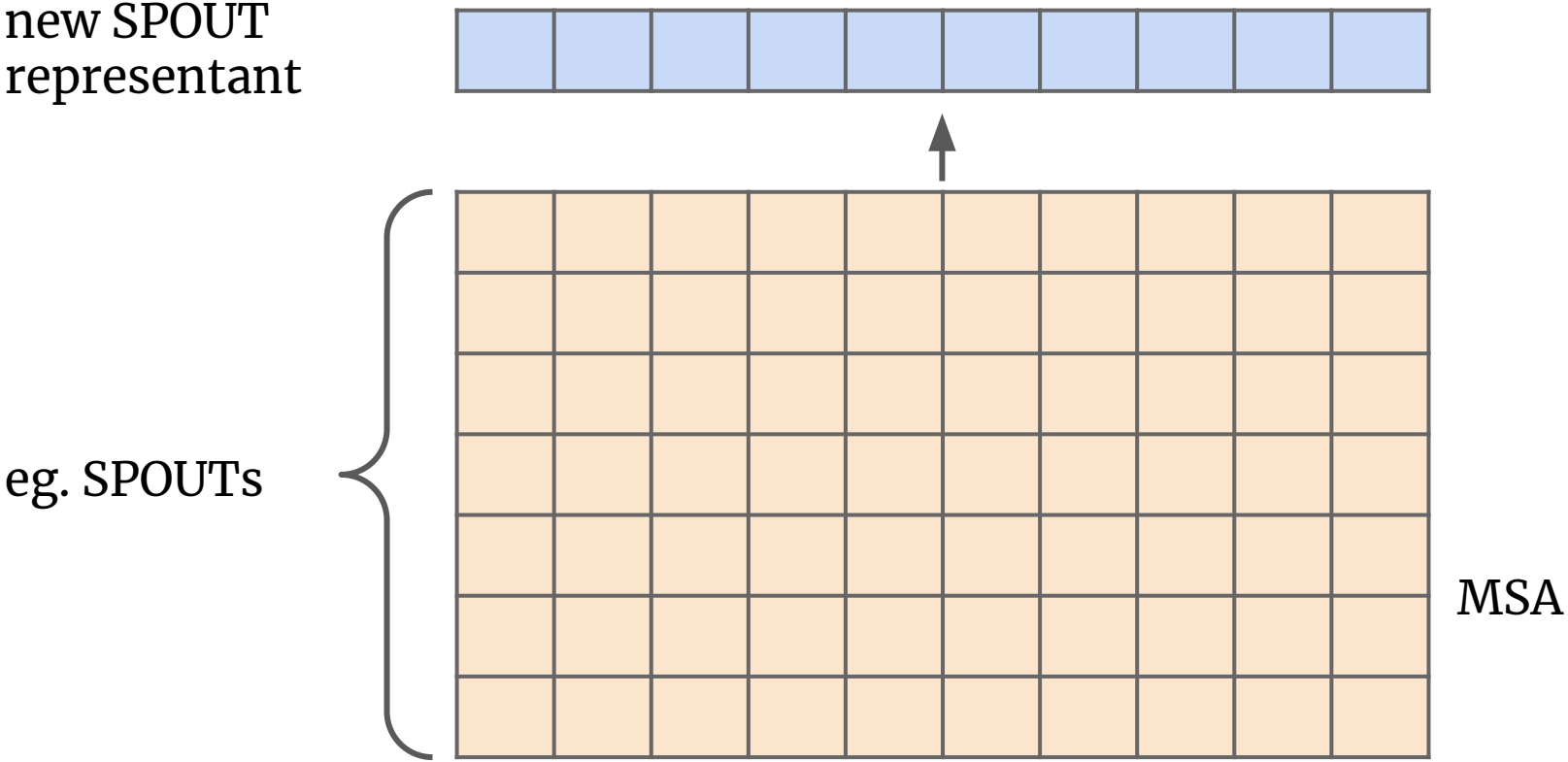


https://github.com/microsoft/evodiff/tree/main

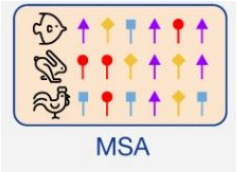
EvoDiff - directly generate sequence

- ➔ random sequence
- ➔ full sequence containing chosen sequence motif
- ➔ new representant from MSA

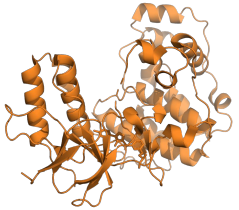



new SPOUT
representant
eg. SPOUTs
MSA

MVHLTPEEKSAVT...

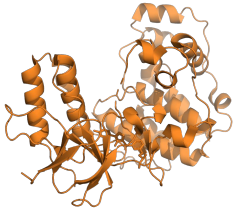


AlphaFold



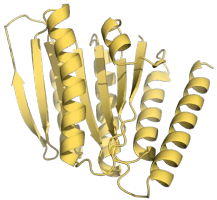
MVHLTPEEKSAVT...

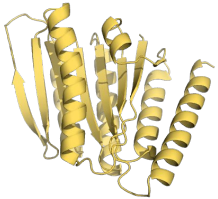
ESM fold



(eg. structural motif)

RFdiffusion





MPNN

MNRIYSLRYSAVA...

(eg. sequence motif or MSA)

EvoDiff

MNRIYSLRYSAVA...