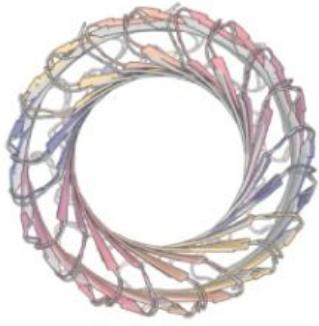
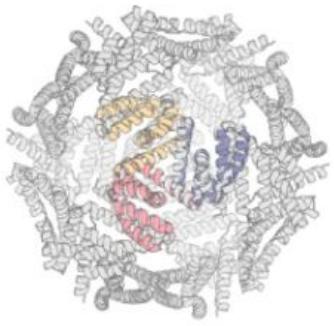


Unveiling the Entangled Landscape of Artificial Knotted Proteins

Eva Klimentová
14. 2. 2024



- How does diffusion work in general?
- What do we have for proteins?
- Can it make knots?
- How do the knots look like?

Diffusion ~ adding and removing noise

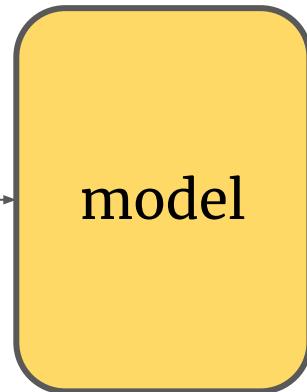
Original image



add noise

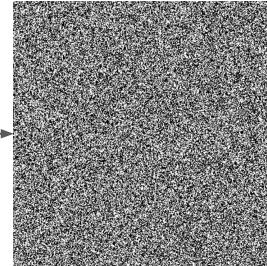


input



model

predict noise

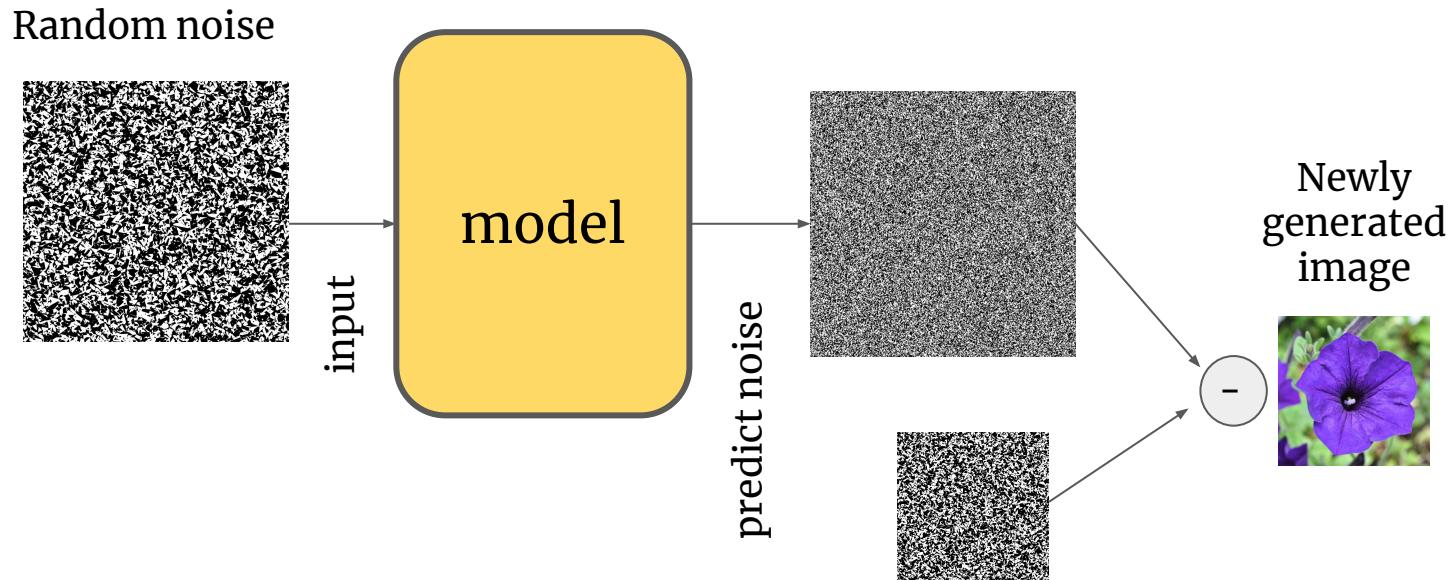


Reconstructed
image



-

Using diffusion to generate new images

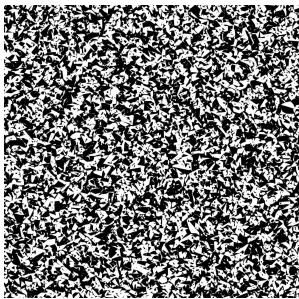


Generate images based on some description

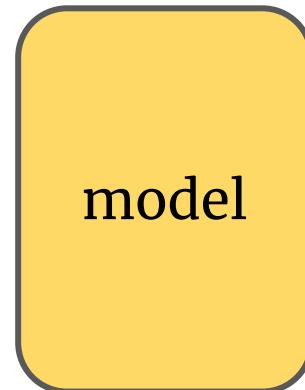


+

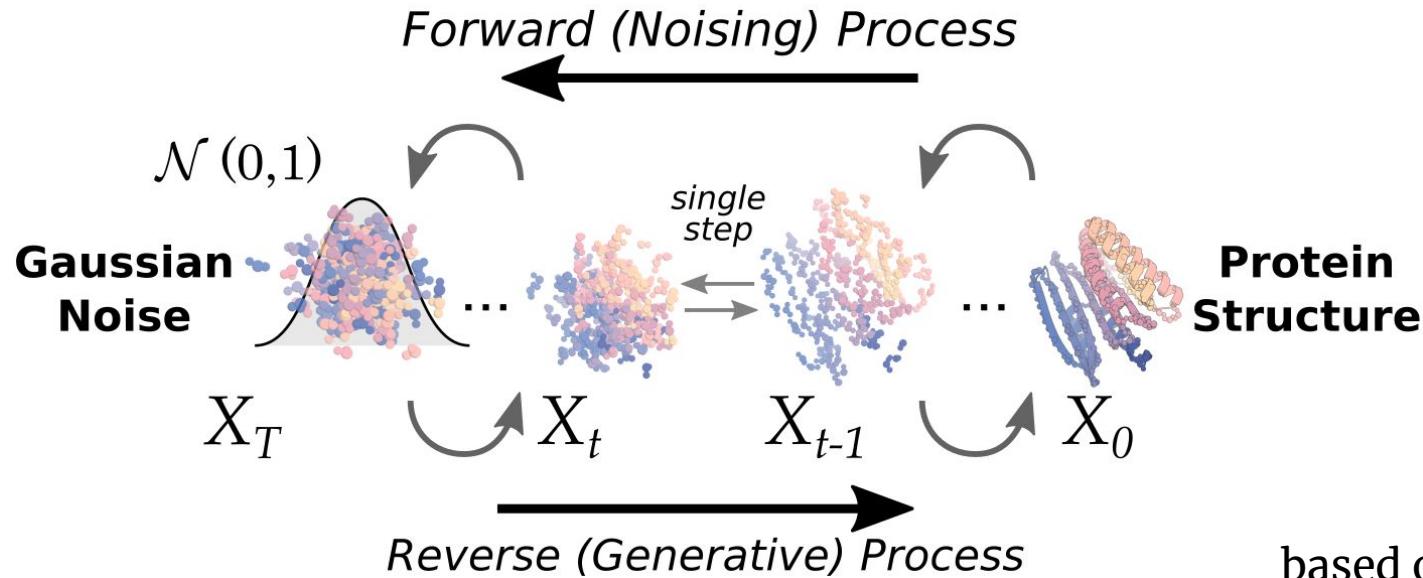
curly girl with glasses
wearing a sweater
cuddling with a rat



curly girl with glasses
wearing a sweater
cuddling with a rat



Diffusion for protein backbone: RFdiffusion

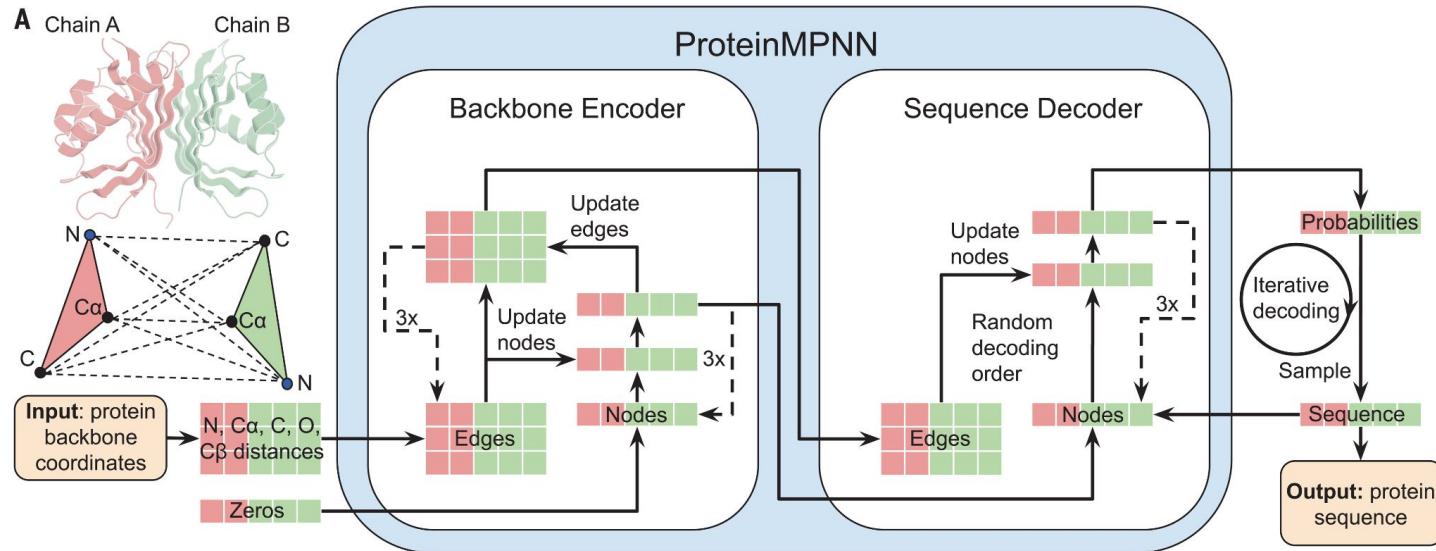


based on
RoseTTAFold
model

<https://github.com/RosettaCommons/RFdiffusion>

<https://www.nature.com/articles/s41586-023-06415-8>

Diffusion for protein sequence based on structure: ProteinMPNN

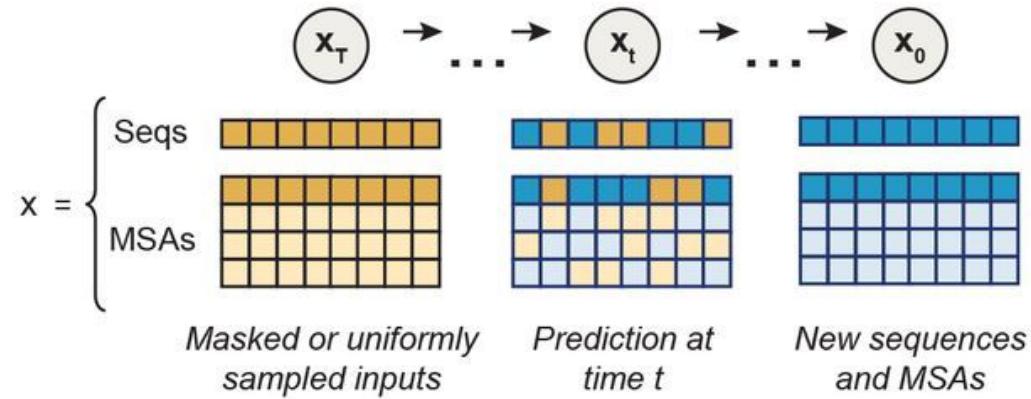
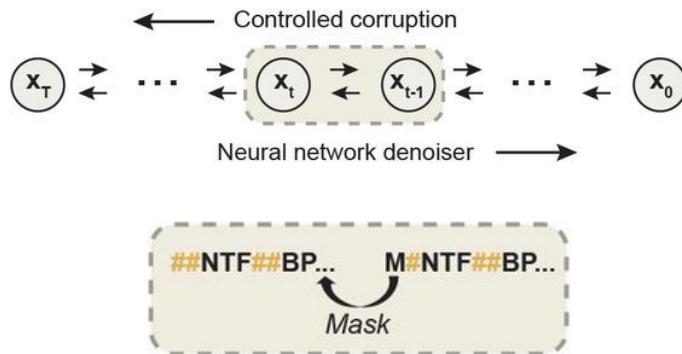


<https://github.com/dauparas/ProteinMPNN/tree/main>

<https://www.science.org/doi/10.1126/science.add2187>

<https://huggingface.co/spaces/simondue/ProteinMPNN>

EvoDiff - directly generate sequence

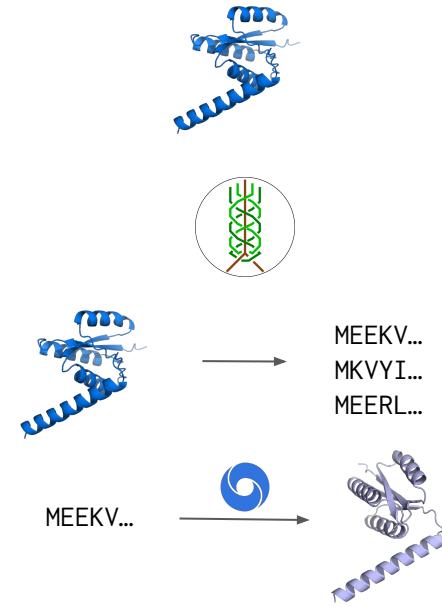


<https://github.com/microsoft/evodiff/tree/main>

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 EvoDiff

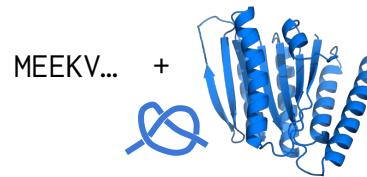
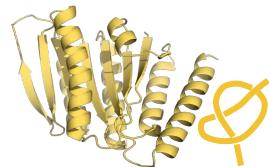
Workflow:

1. design random protein sequence with EvoDiff
2. predict 3D structure with OmegaFold
3. check topology of predicted structure



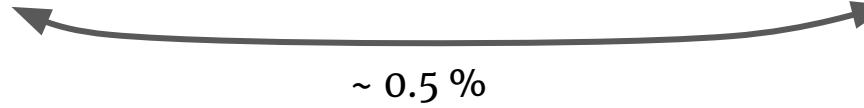
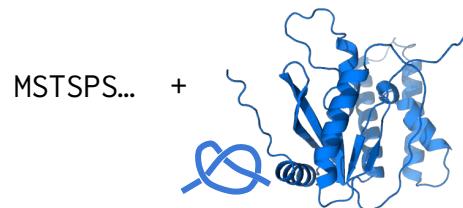
RFdiffusion
+MPNN

1. # 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



EvoDiff

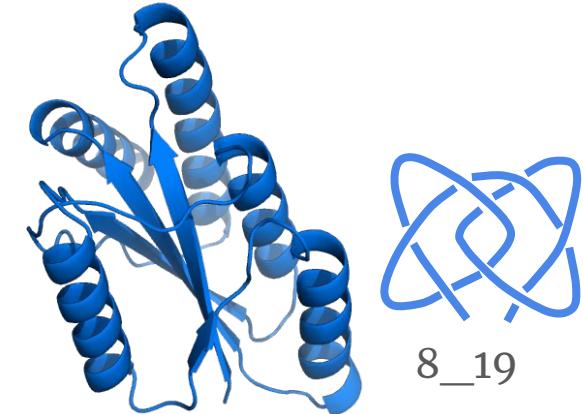
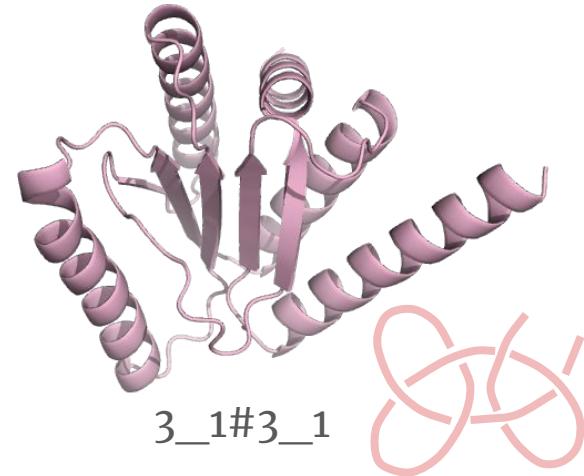
1. # of designed sequences		3. 3D structure of sequence with non-trivial topology
212 681		979



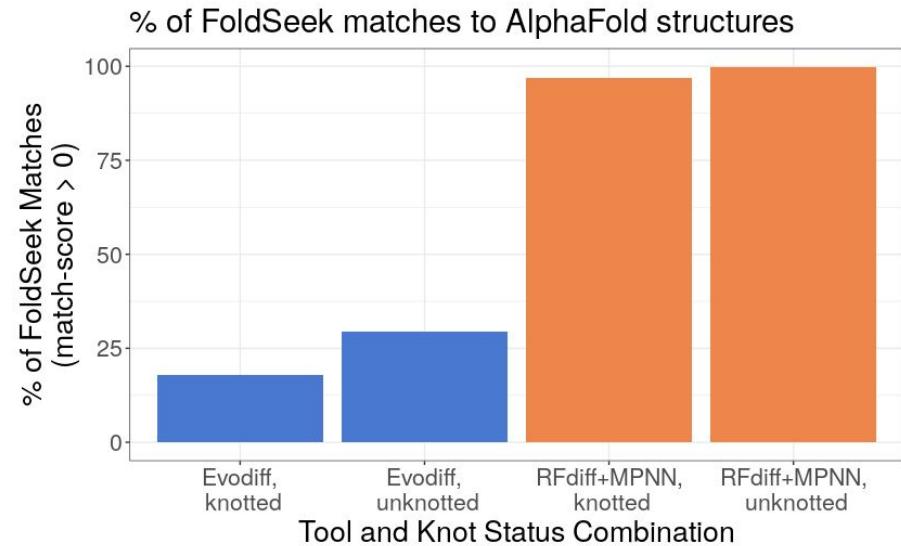
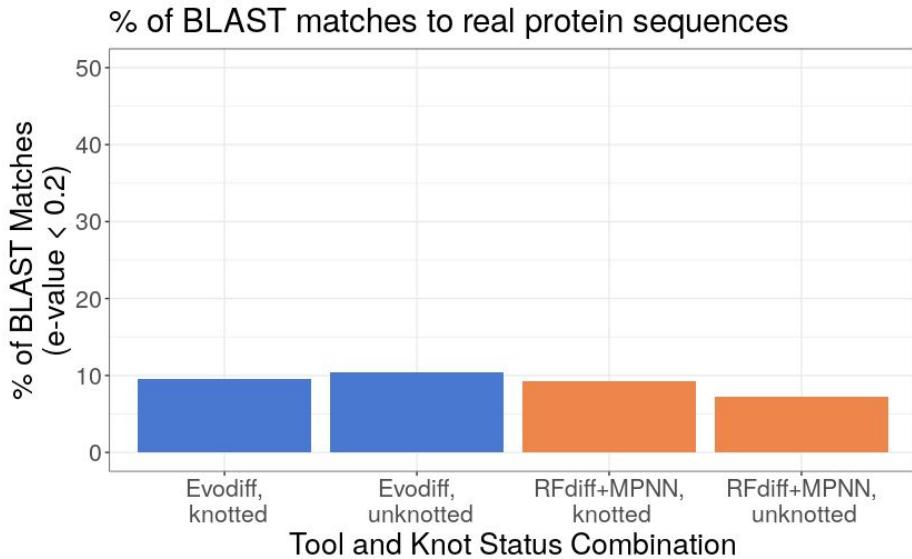
~ 0.5 %

Generated knotted proteins

Major knot type	# of RF+MPNN designs	# of EvoDiff designs
3_1	950	866
3_1#3_1	5	2
4_1	7	44
5_1	26	15
5_2	1	11
8_19	3	0
and others		

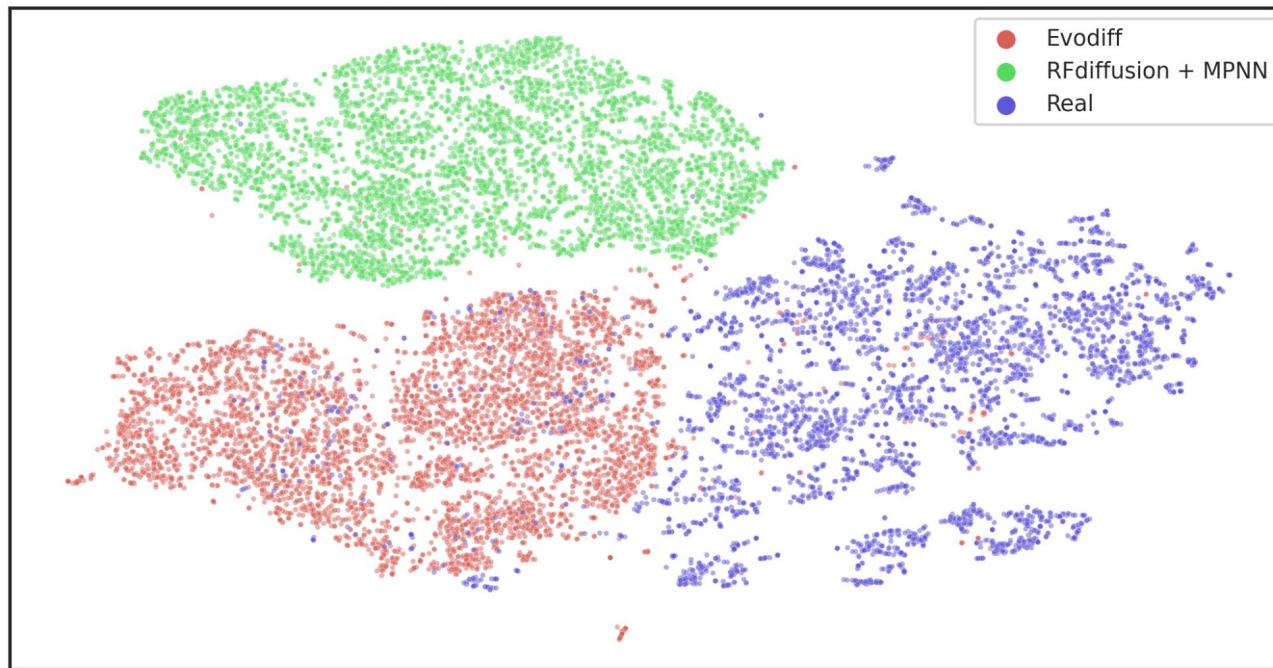


Insights into generated proteins



Insights into generated proteins

T-SNE projection of ProtBert-BFD embeddings



What can we do next?

- look for knotted protein suitable for crystallization
- try to expand the set of knotted proteins

Takeaway

- EvoDiff and RFdiffusion+ProteinMPNN for creating artificial proteins
- it can make artificial knotted proteins
- the artificial proteins can be quite different from real world

https://huggingface.co/datasets/EvaKlimentova/Diffusion-all_knots



Thank you

?

