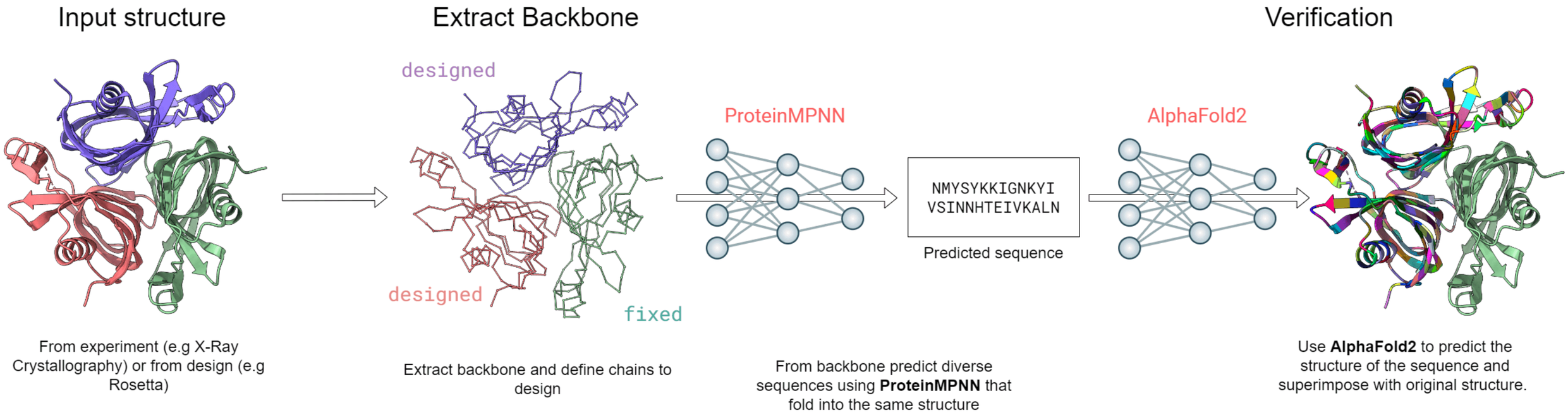
‘Generative AI’ for protein design

Chaitanya K. Joshi, PhD Student, University of Cambridge, UK — https://chaitjo.com

Structure-based protein design workflow

Assumption: Structure → Function



Not shown: protein Language Models (purely sequence-based)

Dauparas et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science. 2022.
Figure: Simon Duerr

Analogy to ChatGPT

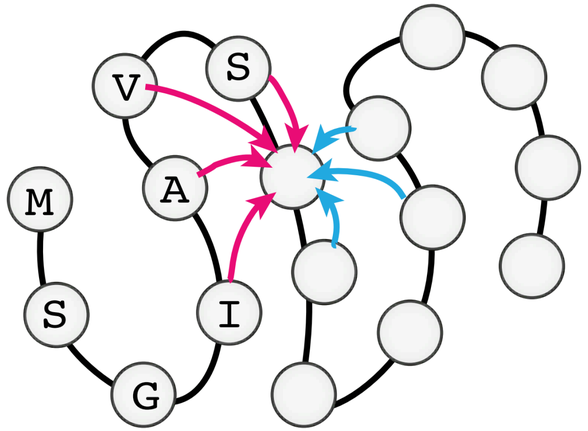
Natural language models

S = Where are we going

Previous words (Context)

Word being predicted

$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$



Trained on PDB structures:

Samples are biased towards thermal stability, expression.

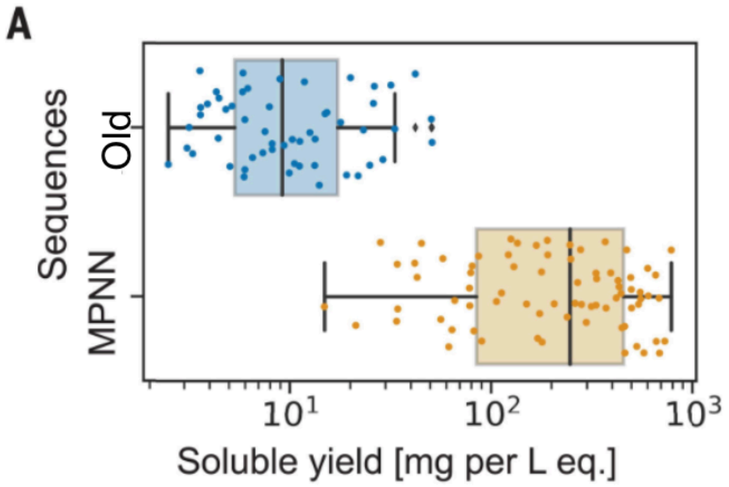
Sequence generation: **Language model**

Sequence generation conditioned on structure: **ProteinMPNN** (inverse folding)

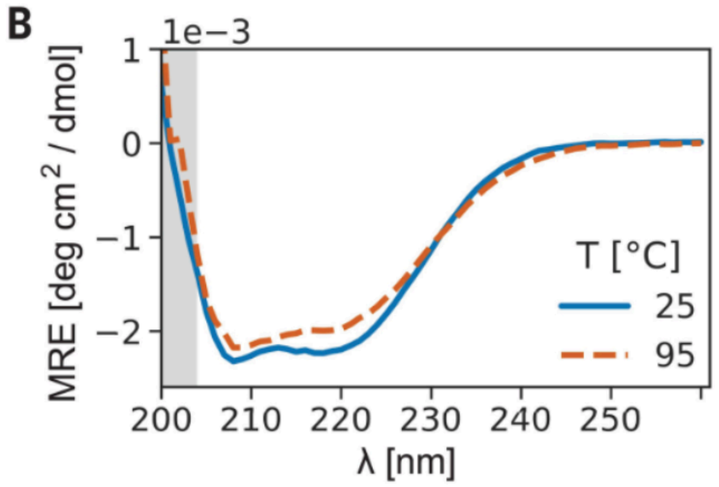
Ingraham et al. Generative Models for Graph-Based Protein Design. NeurIPS, 2019.

ProteinMPNN improves over physics-based tools

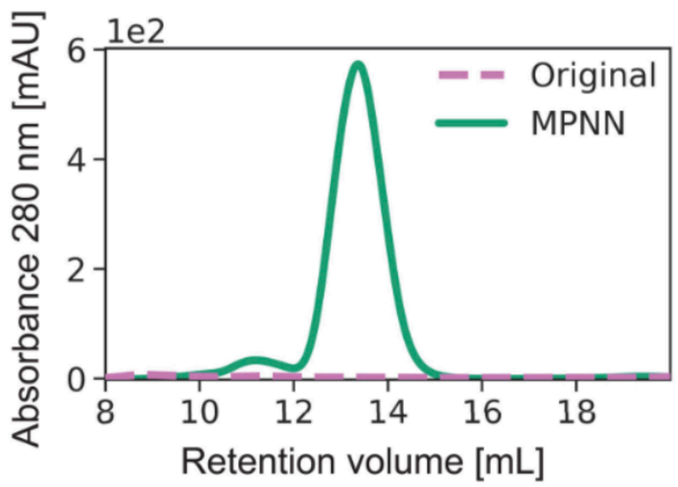
Fixed backbone re-design of structures from Rosetta



Soluble protein yield after expression in E.coli for old designs vs. re-designs (129 proteins)



A highly thermostable design, sec. structure maintained up to 95°C



Size exclusion chromatography profile of a failed design vs. re-design.

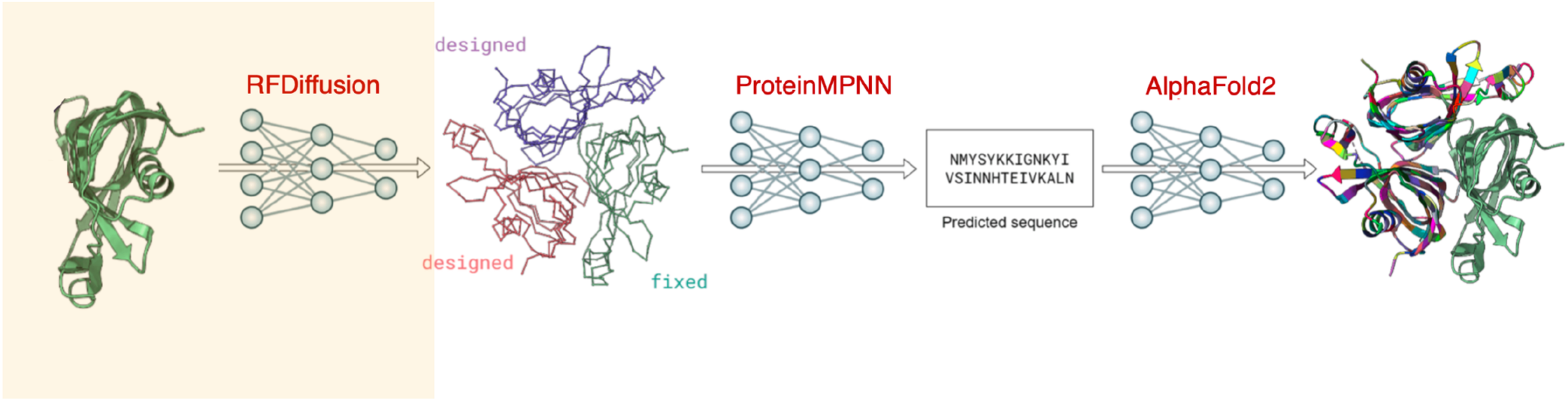
De-novo protein design workflow

Starting from scratch

Backbone design

Inverse folding

Verification



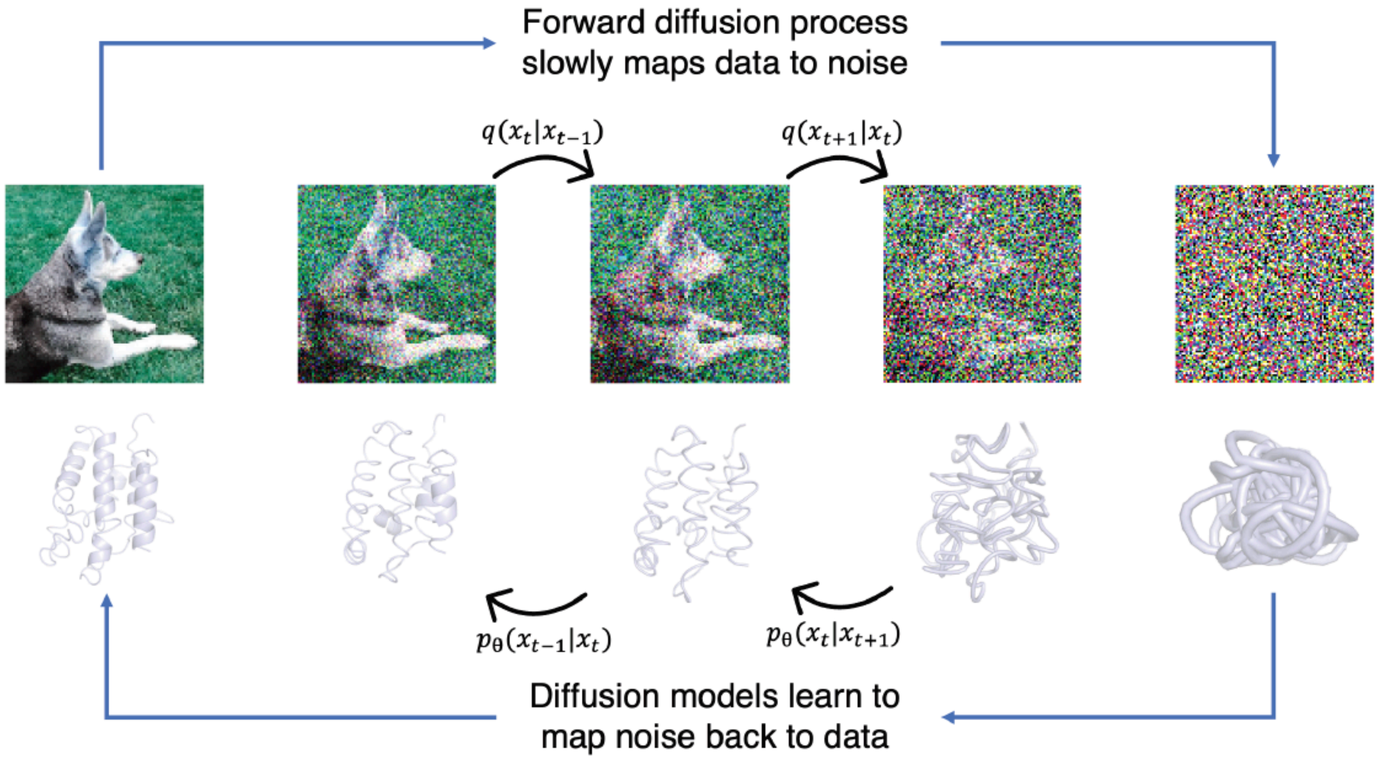
Specify target or
design constraint

Watson, Juergens et al. De novo design of protein structure and function with RFdiffusion. Nature. 2023.

Analogy to DALL-E

DALL·E

Image generation models



Trained on PDB structures:

Learn to mix and match real protein sub-structures.

Backbone design: **RFdiffusion**

Figure: Kieran Didi

RFdiffusion designs functional proteins

Task: scaffold a p53 helix for improved binding to MDM2

Among 96 designs, they find 0.5 nM binders. This is 3 orders of magnitude better compared to native 600nM affinity of p53 alone.