

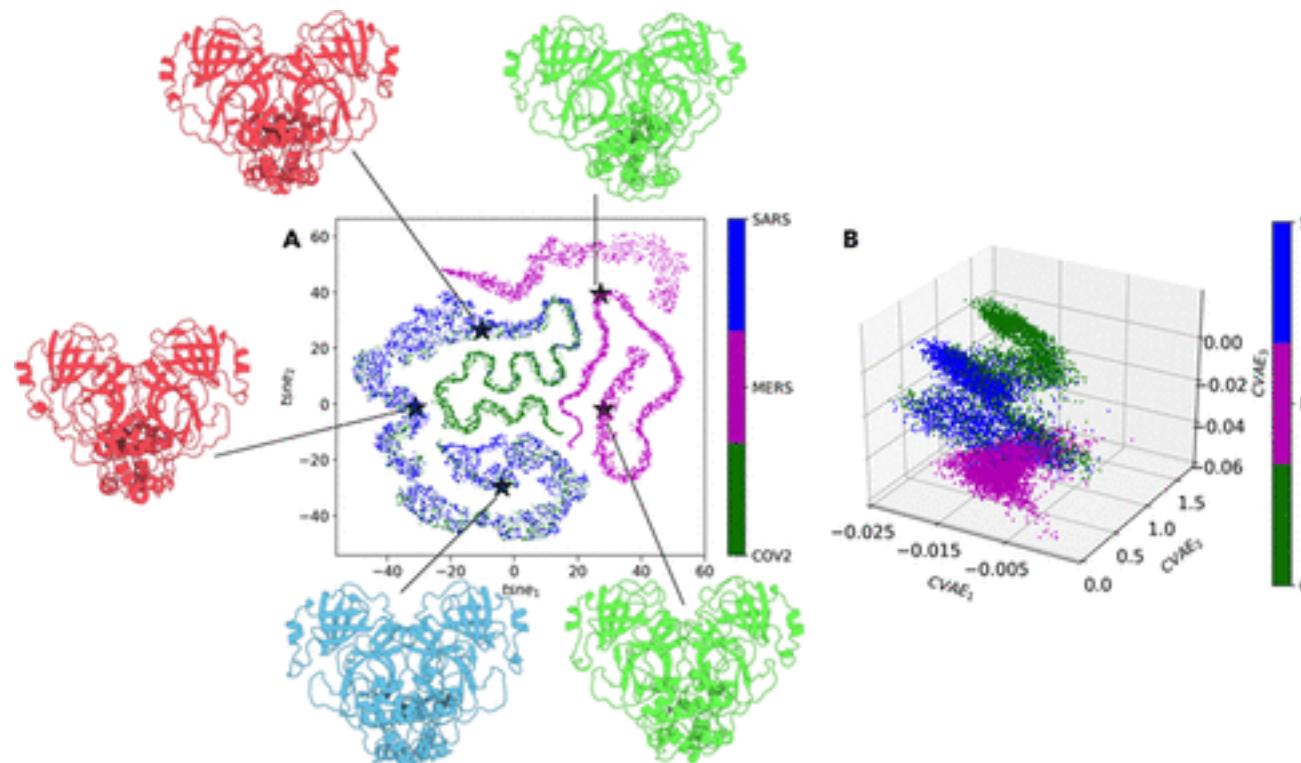
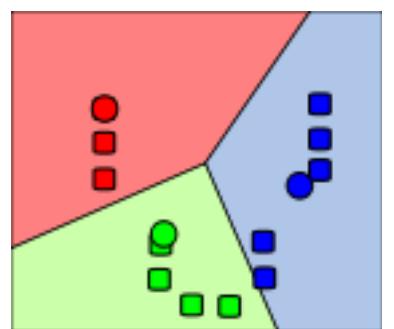
From (chemical) data to information

Introduction to Machine Learning Lecture 5

Dr Antonia (Toni) Mey

✉ antonia.mey@ed.ac.uk

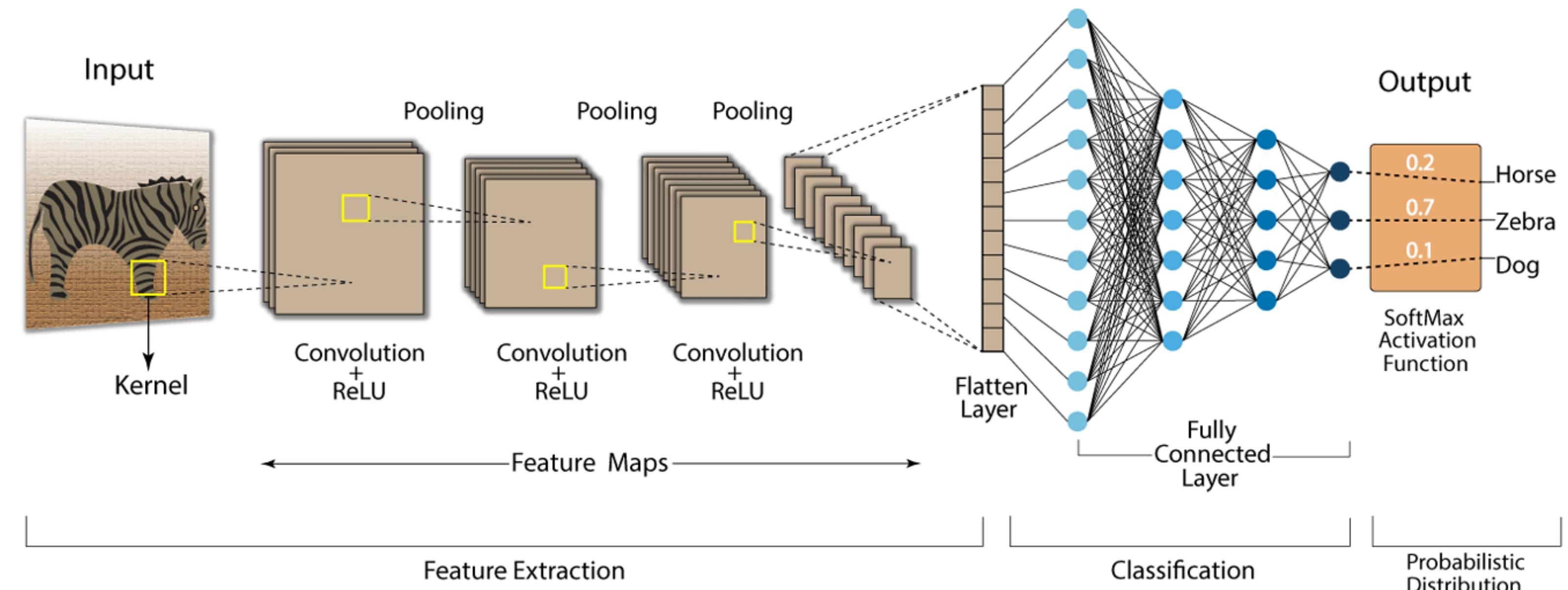
📍 Room B23 JBB



More complex architectures now exists

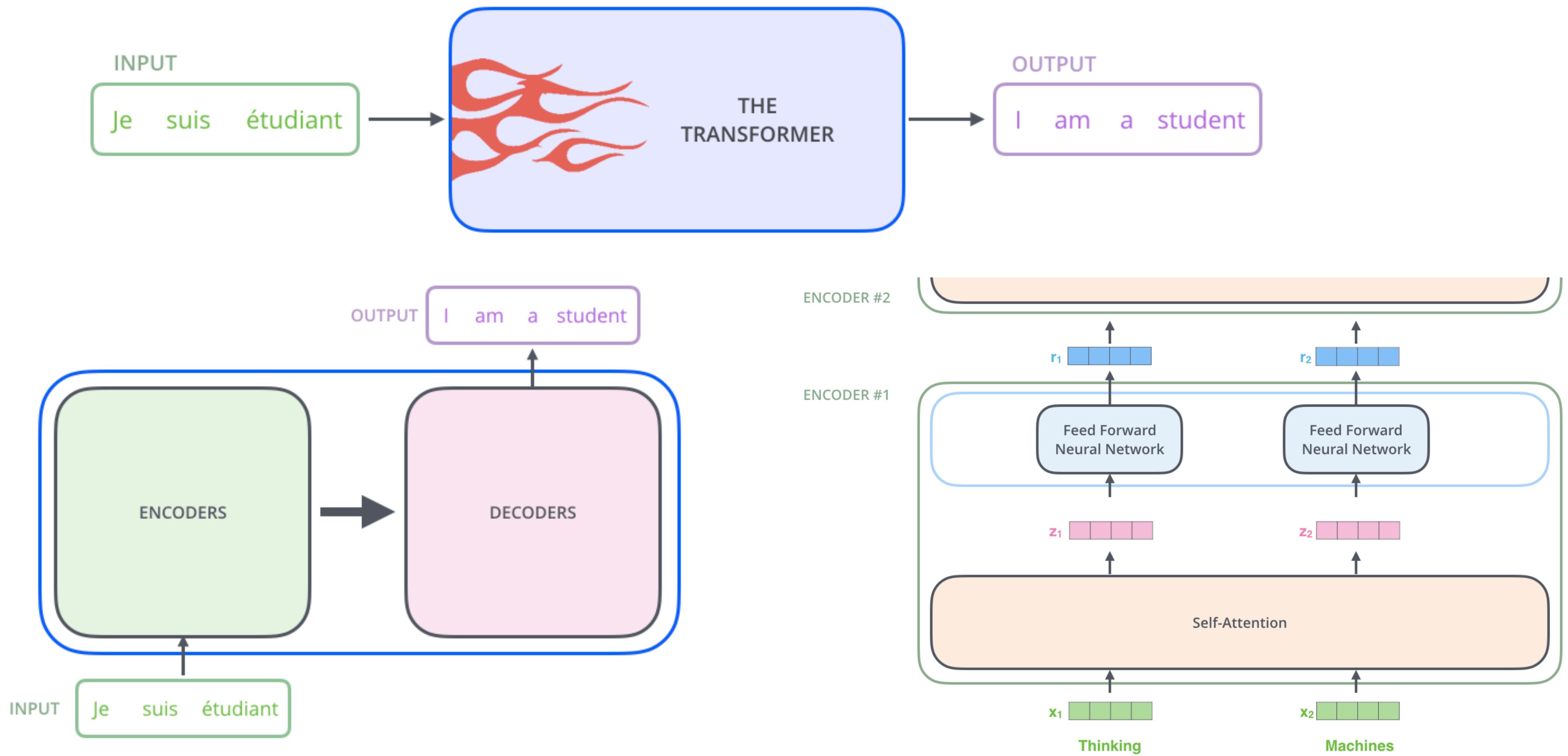
1. Convolutional neural networks

Many different complicated architectures are possible



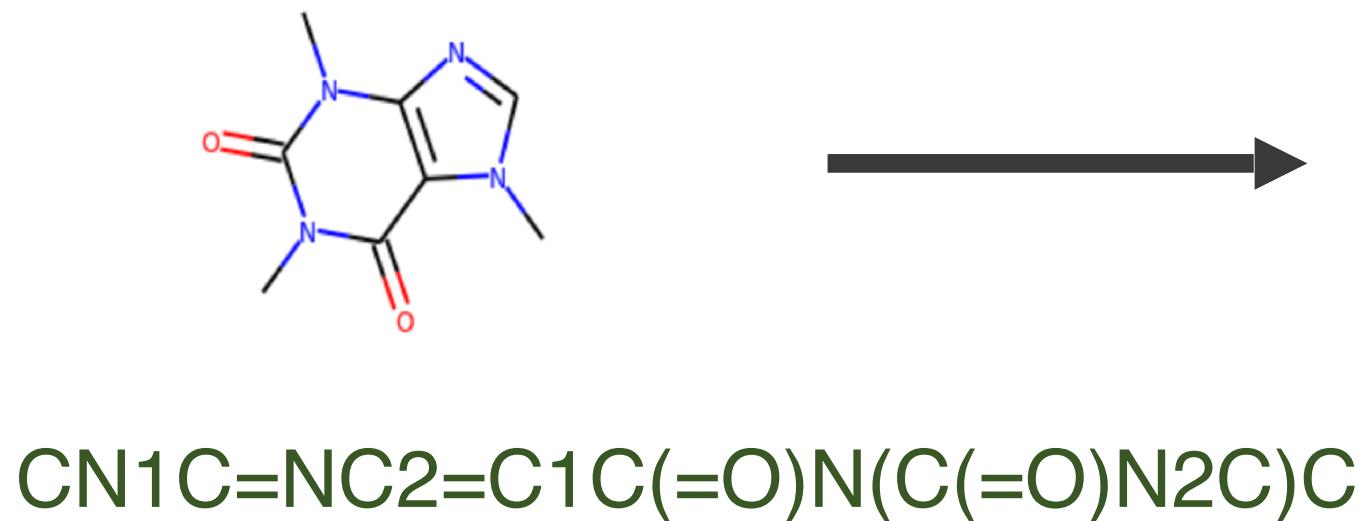
The transformer is used in machine translation

2. Transformers

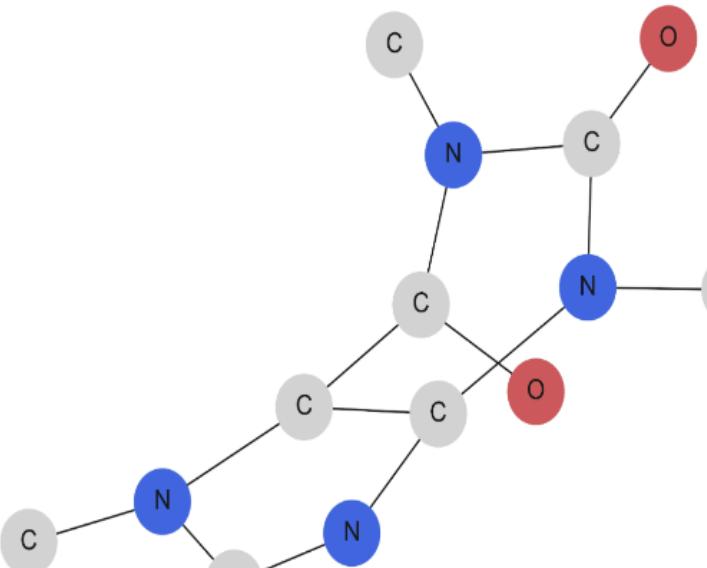


How can you represent a small molecule?

3. Graph neural networks

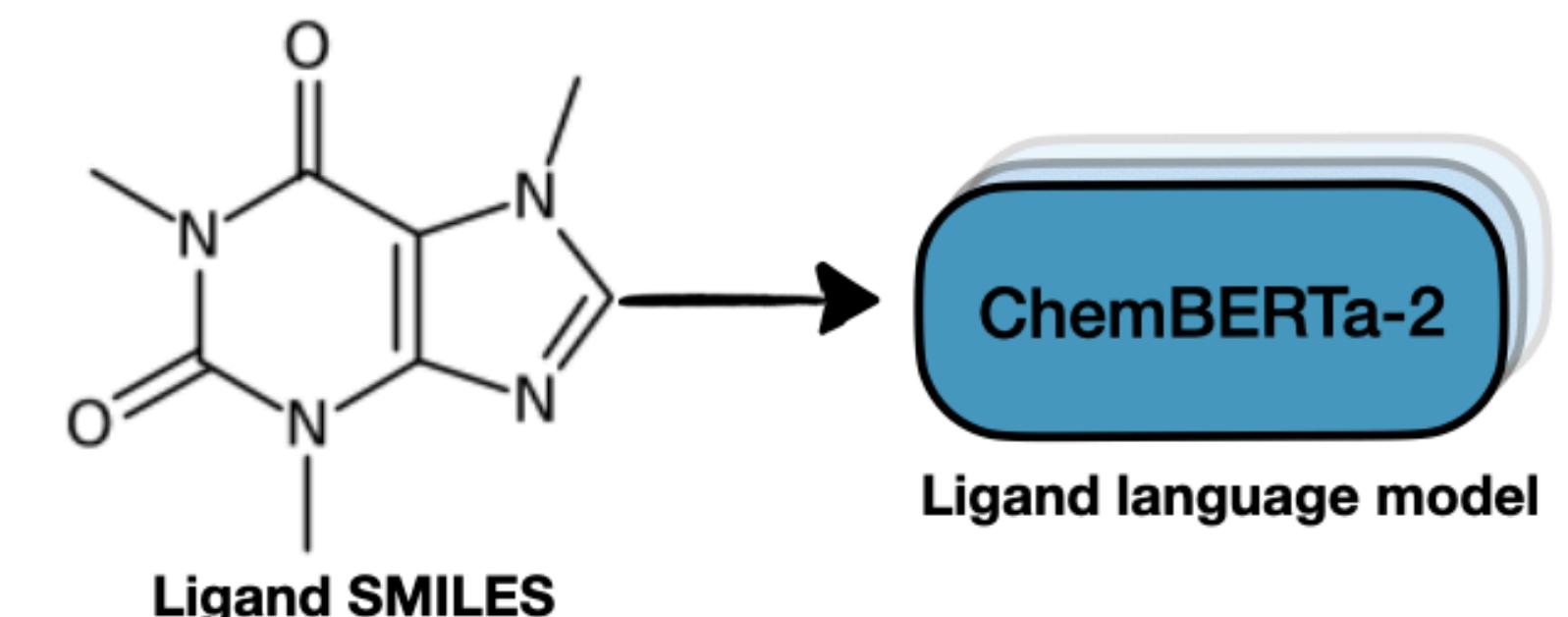
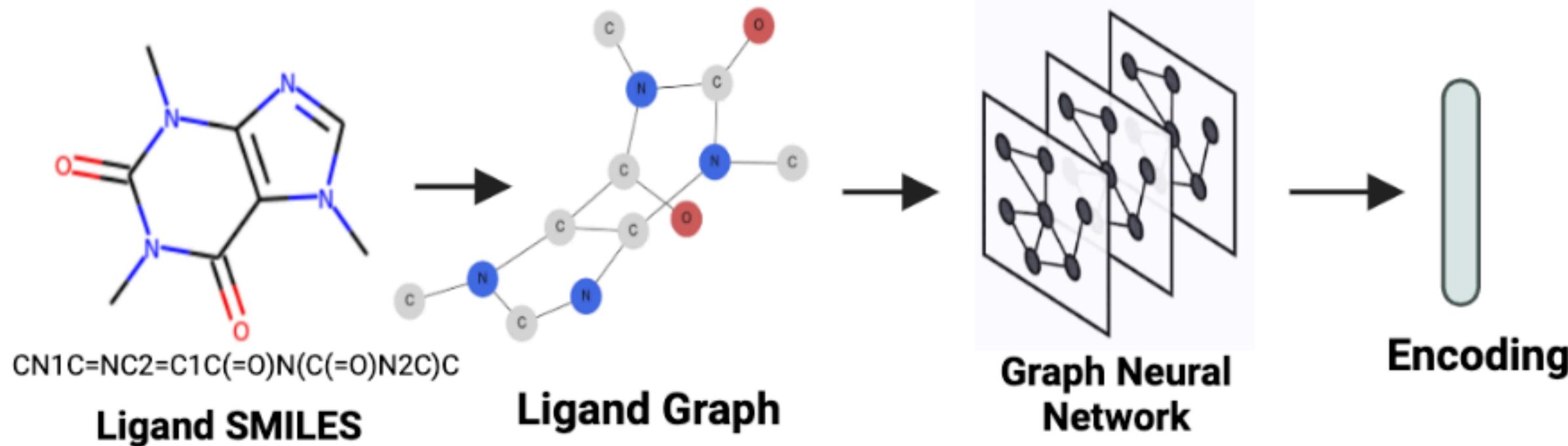


Ligand graph, each atom has many ‘node features’



Trained on 77 million Compounds

Aromaticity, degree of atom, number of hydrogen bonds, etc.



How can we represent a protein?

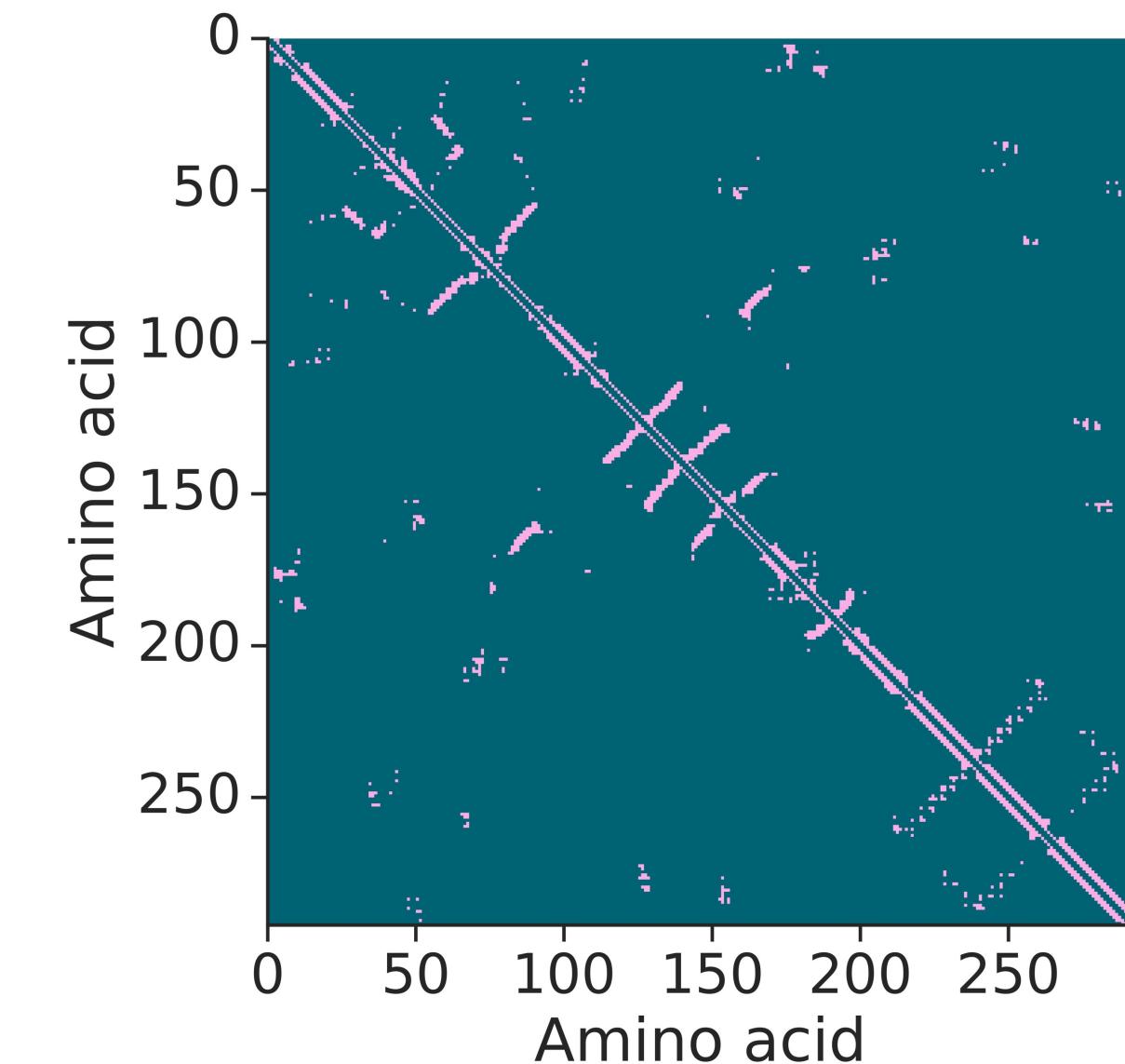
Protein Structure



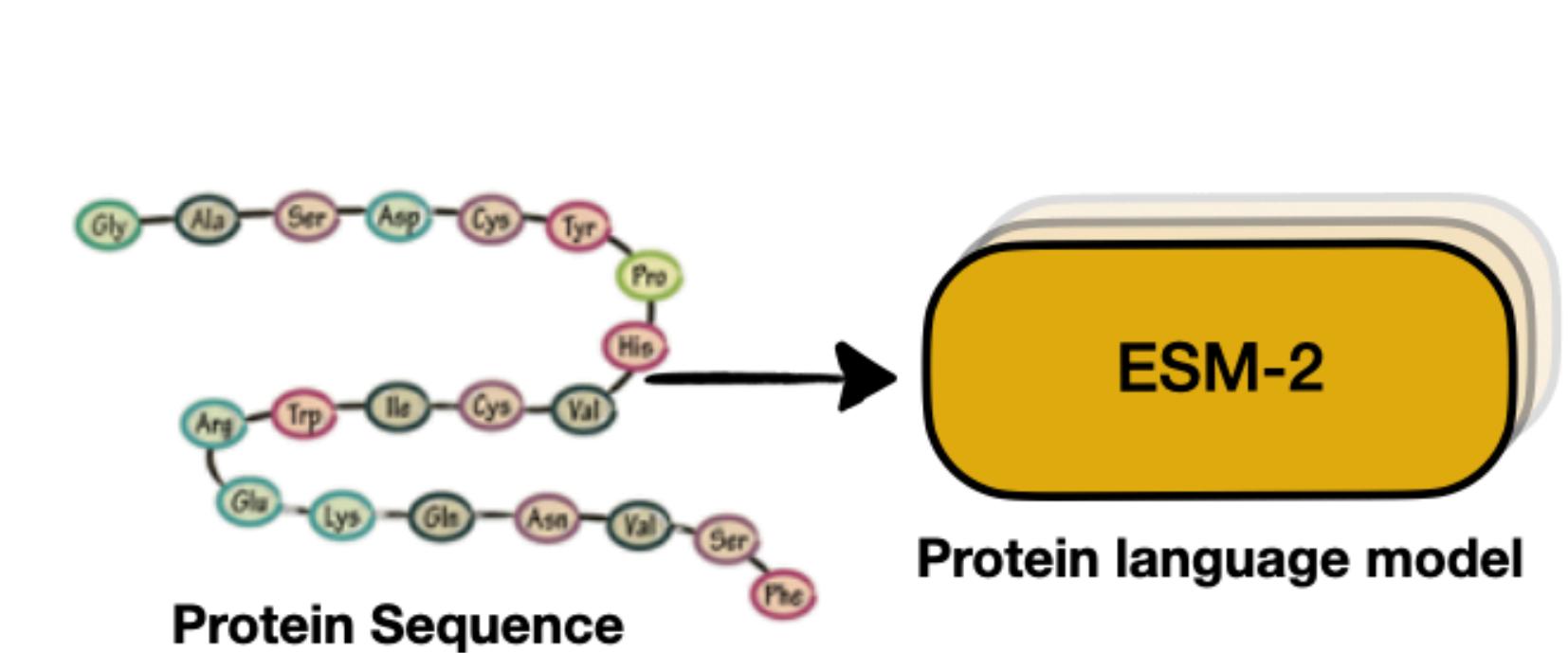
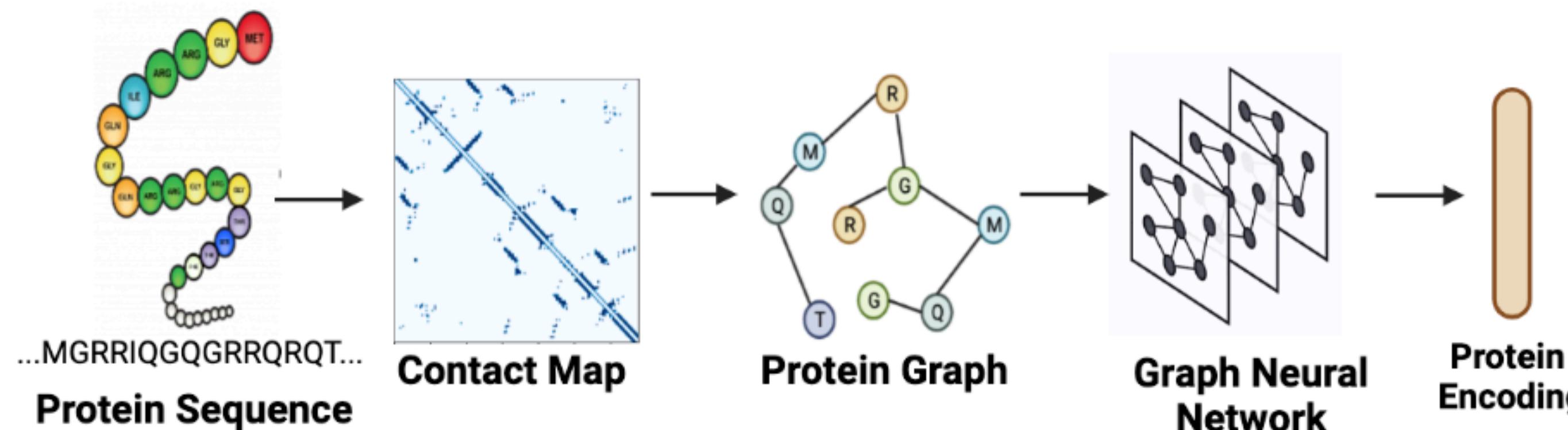
Protein Graph



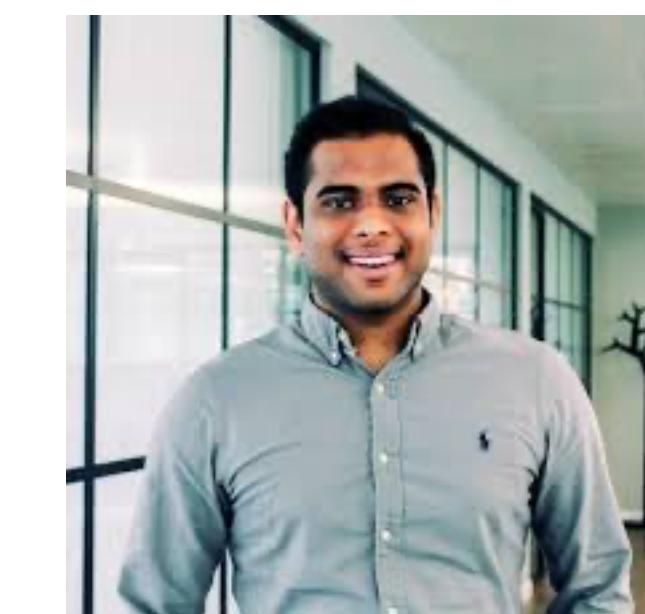
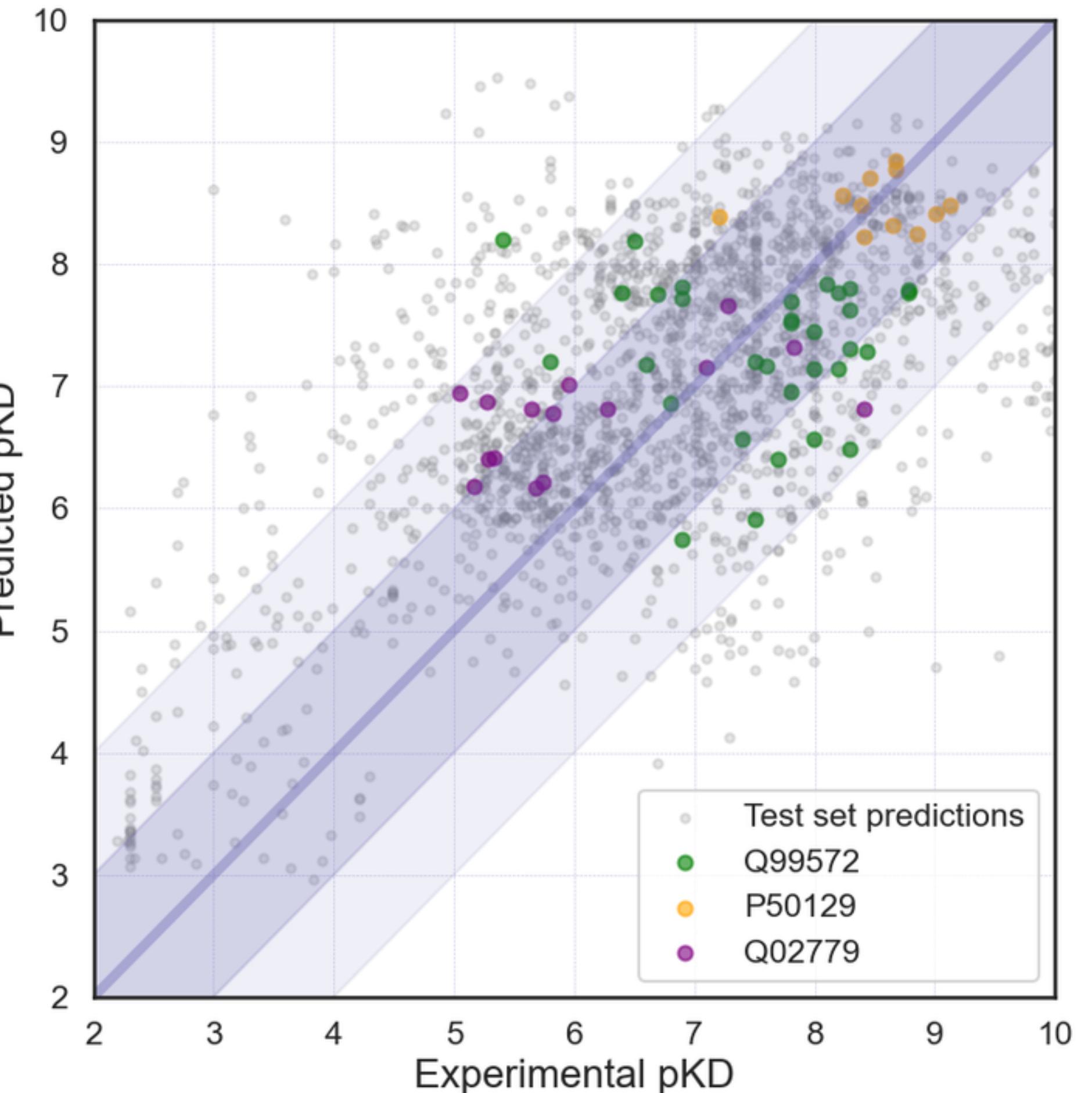
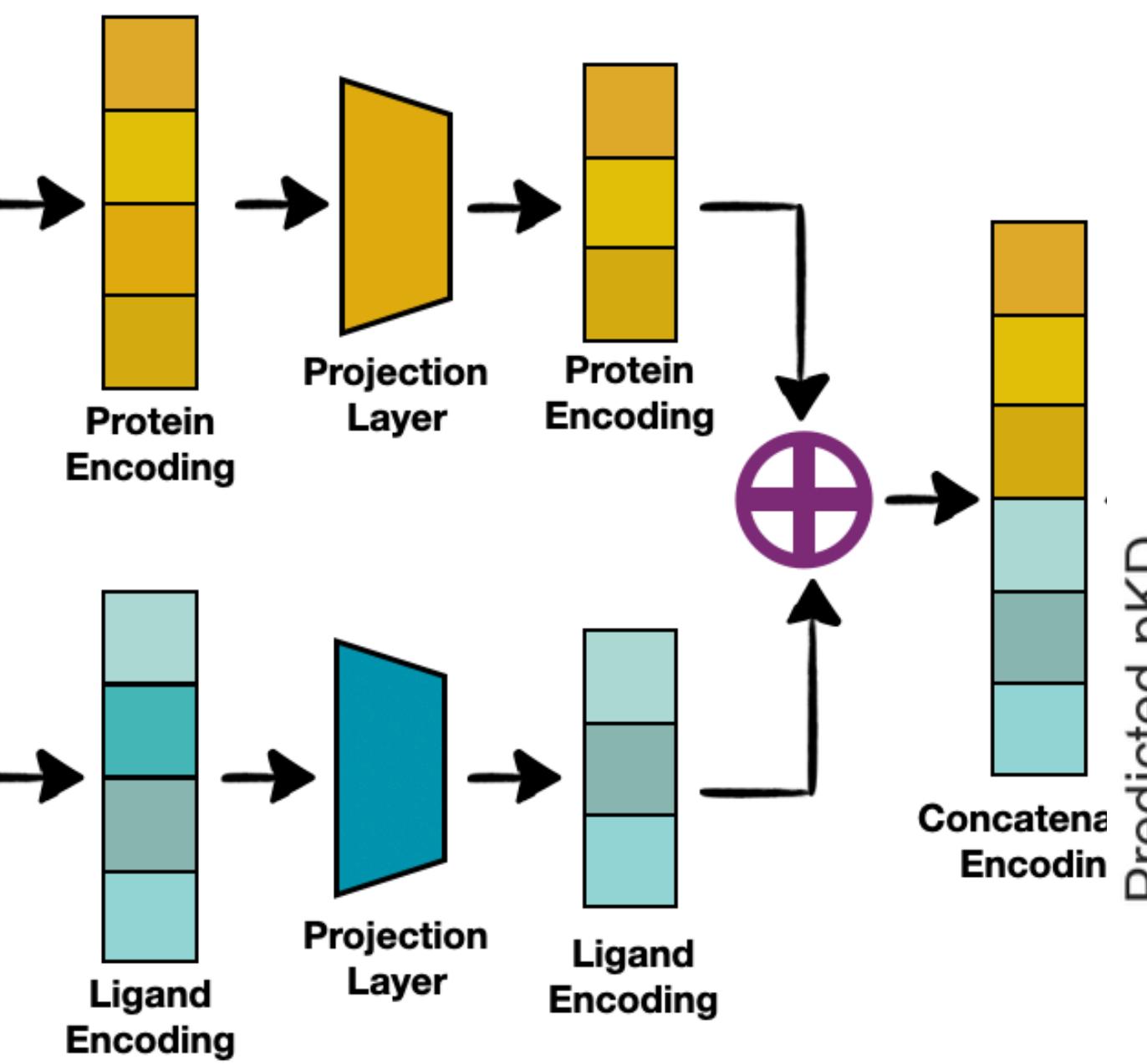
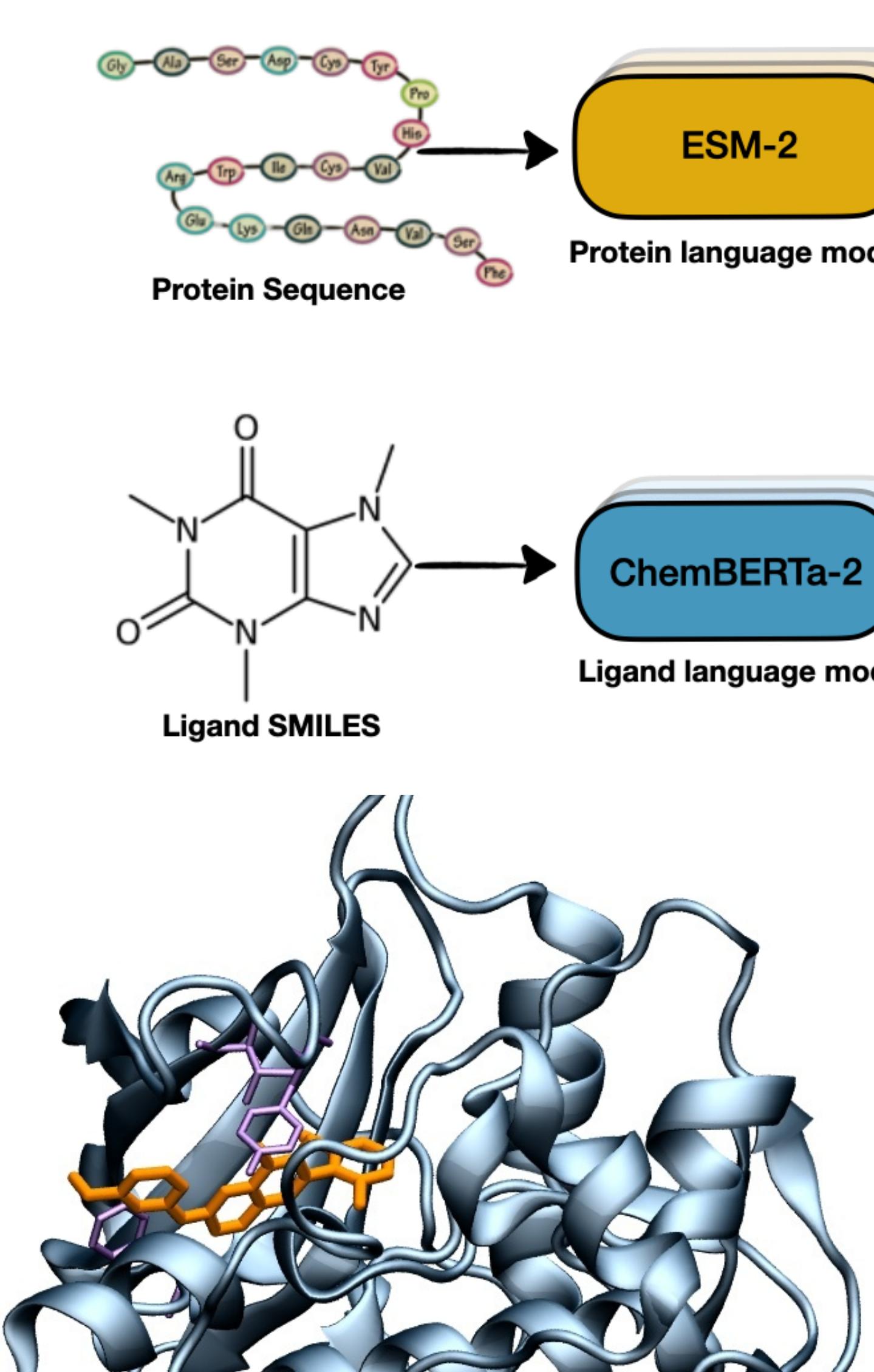
Adjacency matrix / contact map



Protein encoding from contact maps:

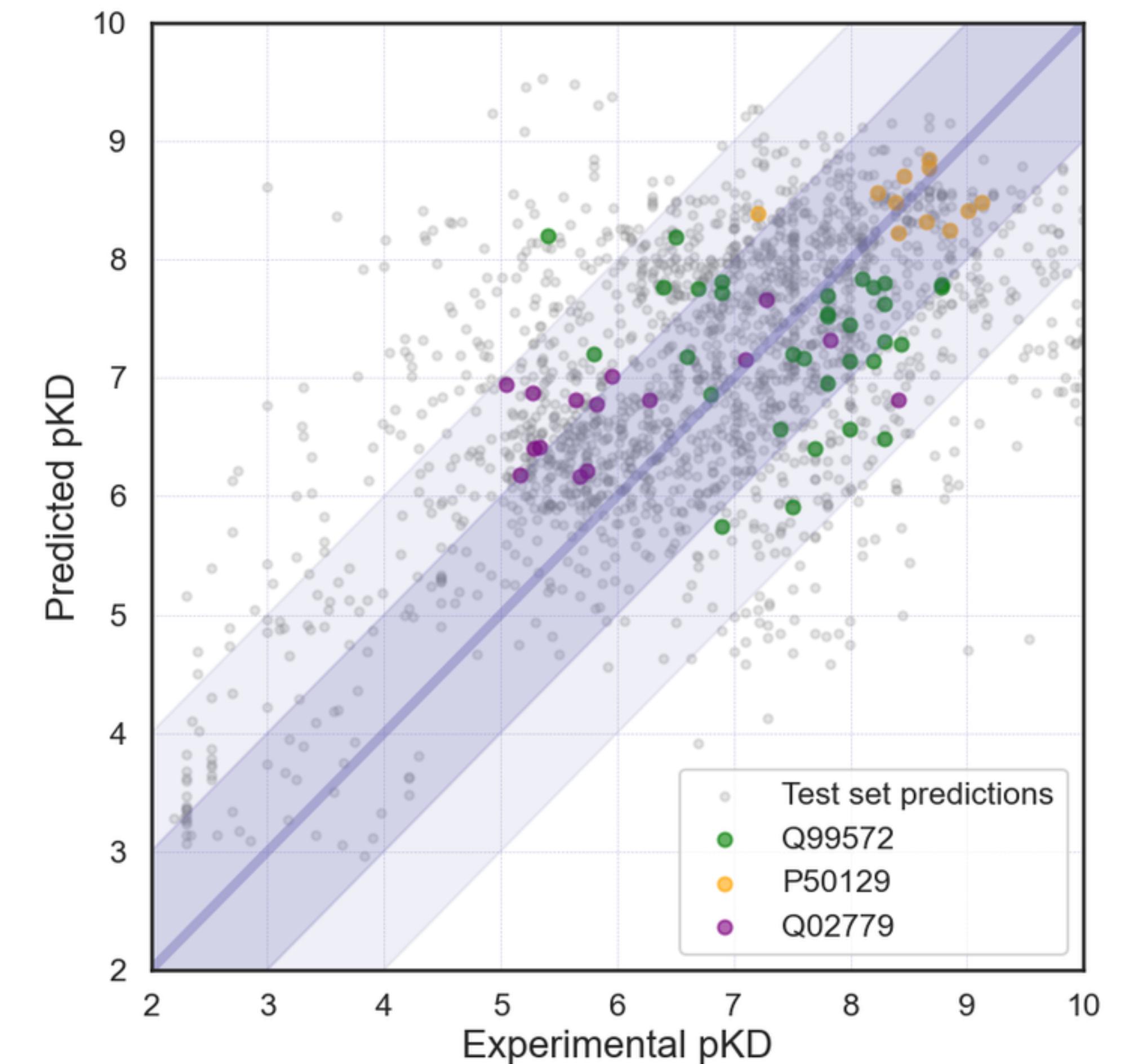
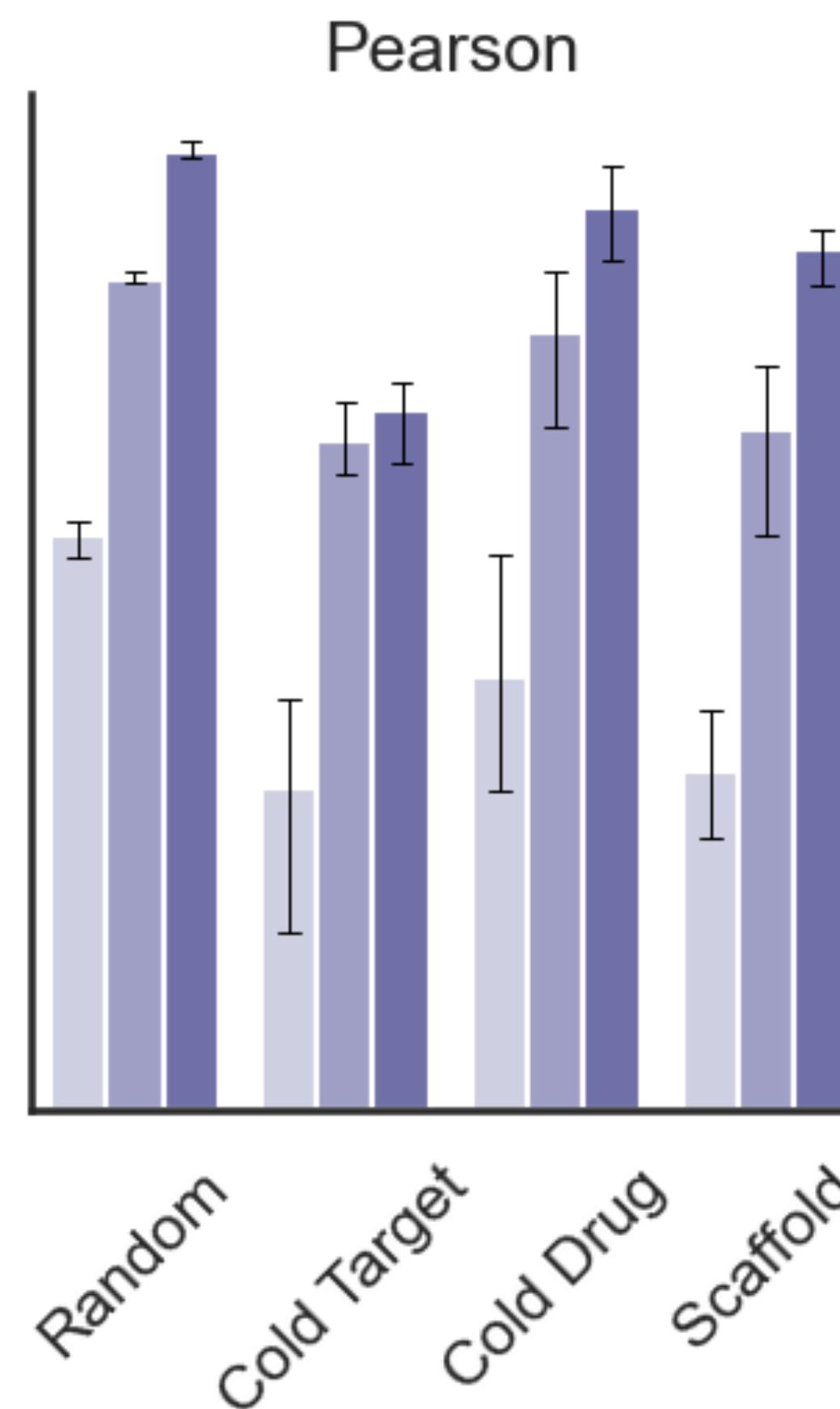
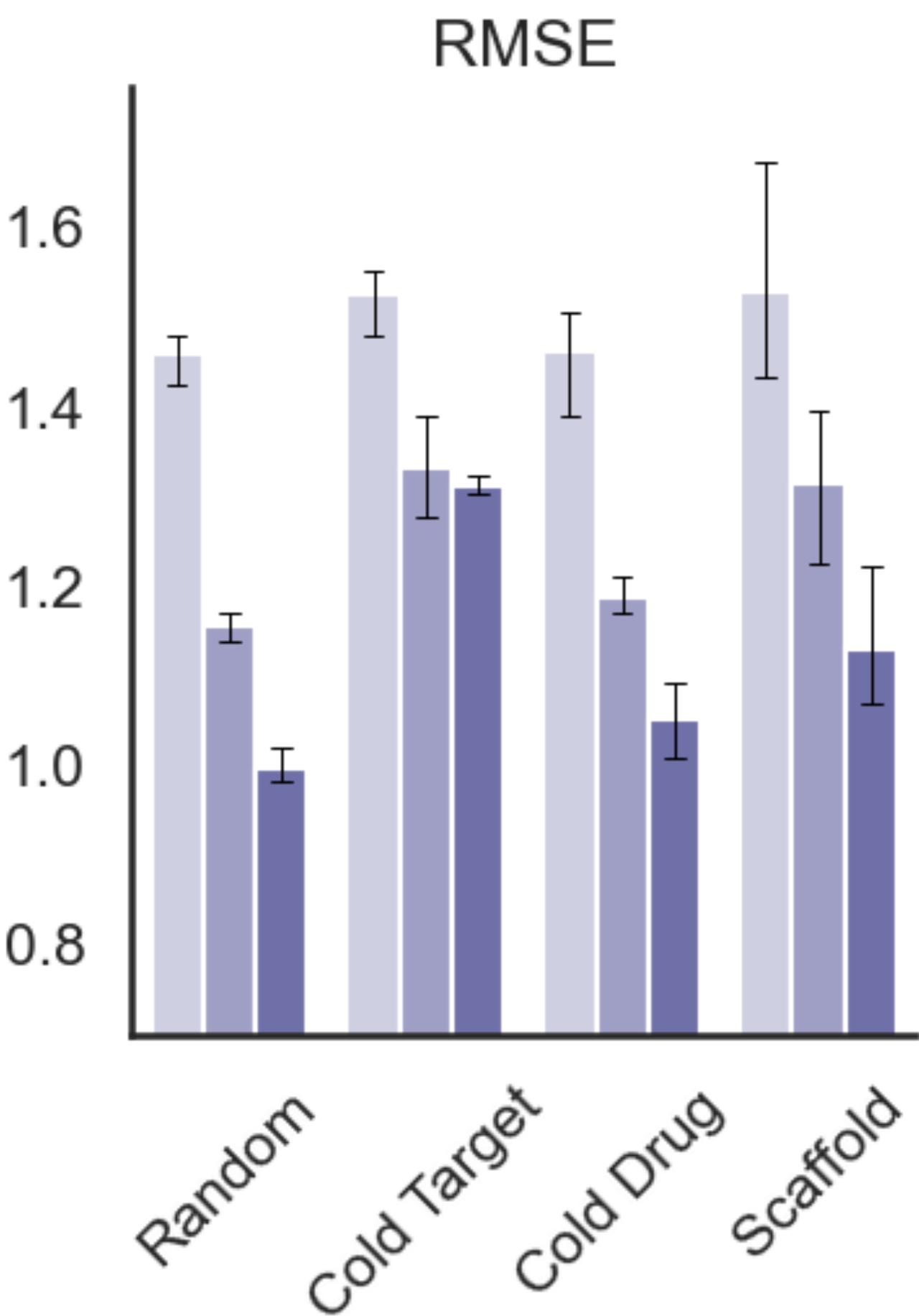


Modern architectures become very complicated



Rohan Gorantla

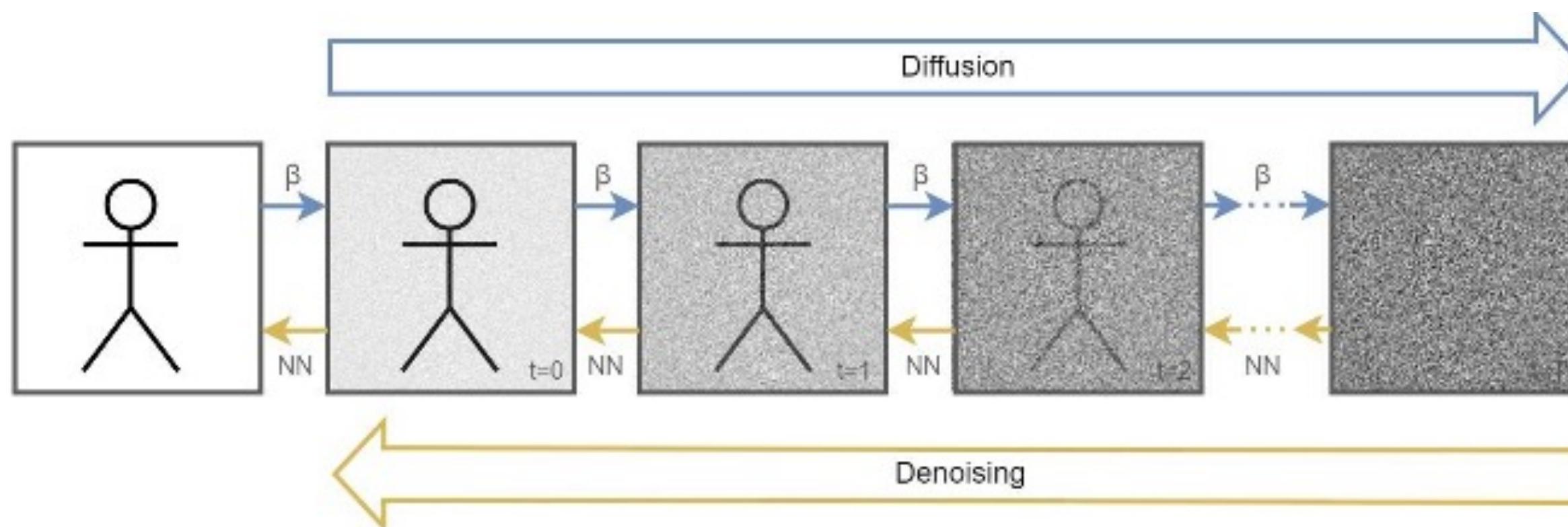
Errors and correlation statistics can be deceiving



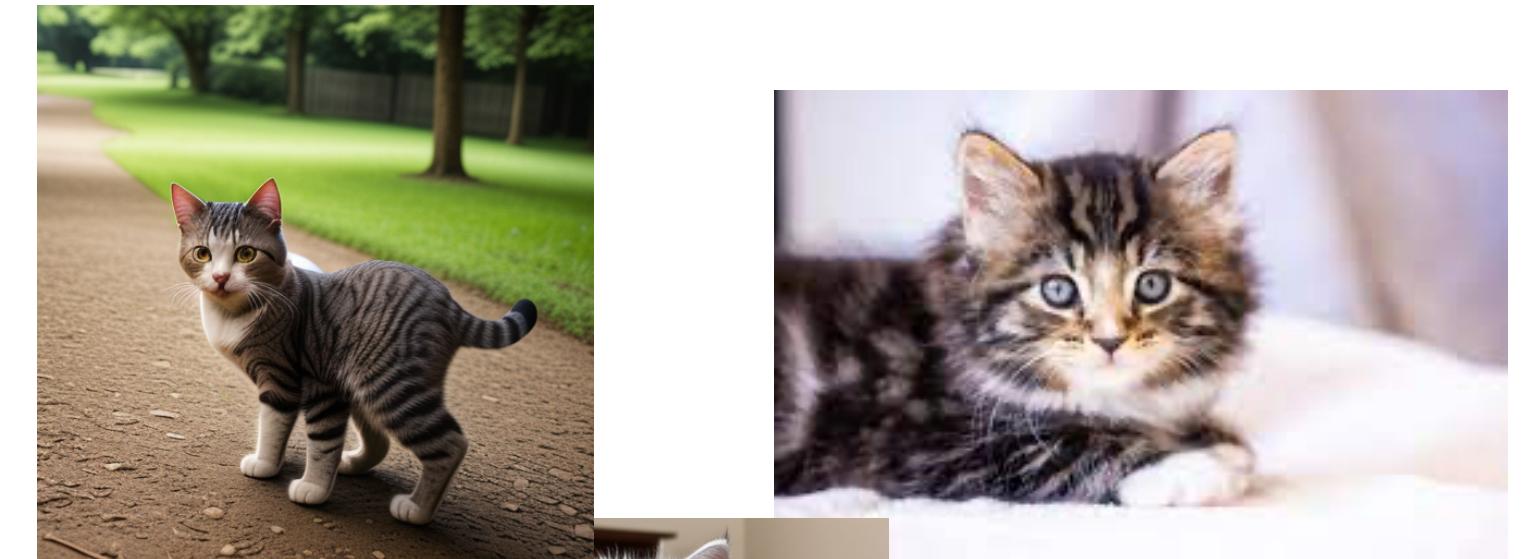
Diffusion models are a class of generative models

4. Diffusion models

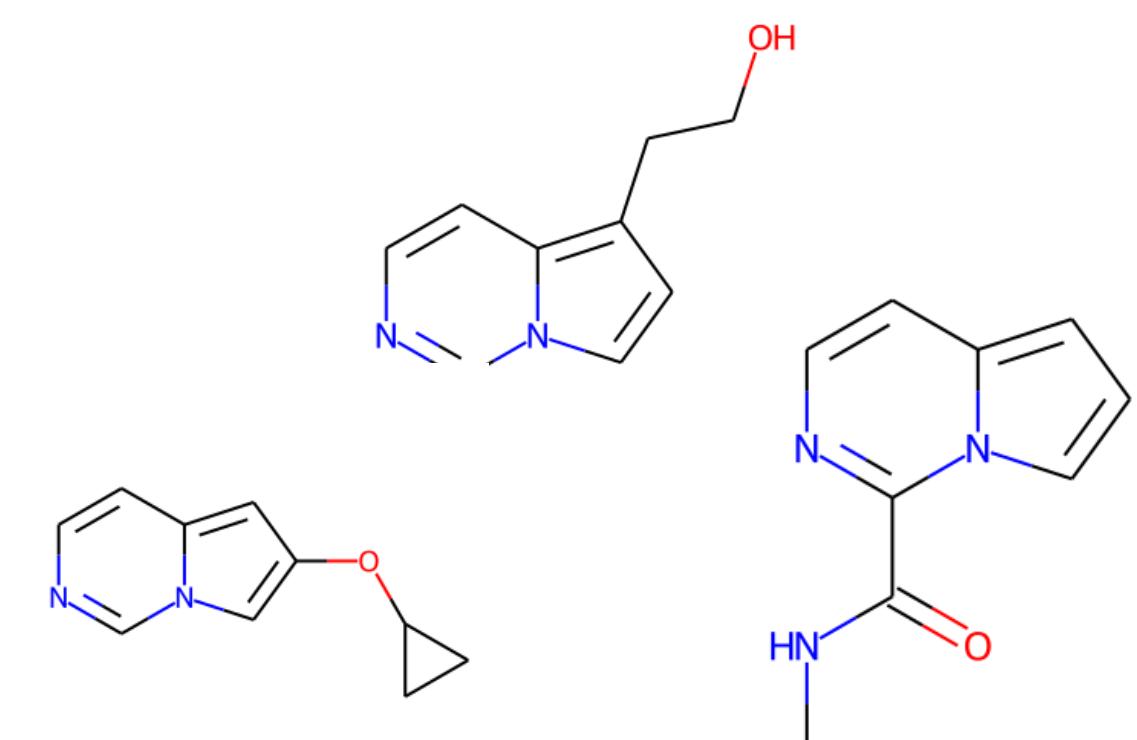
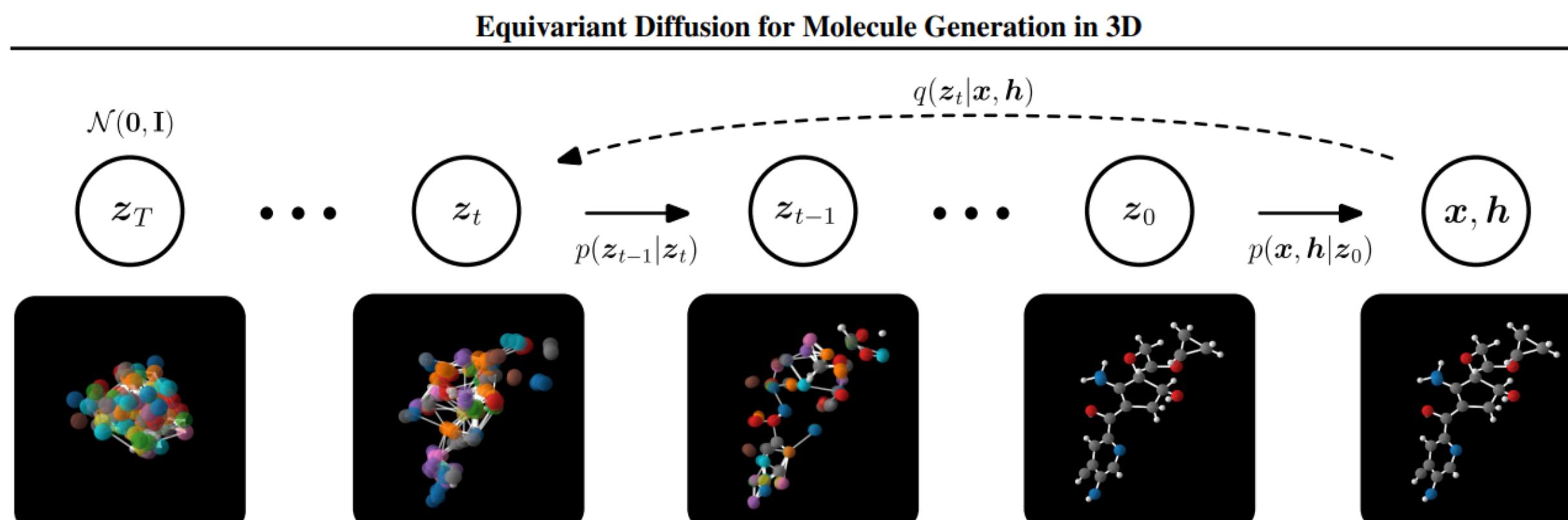
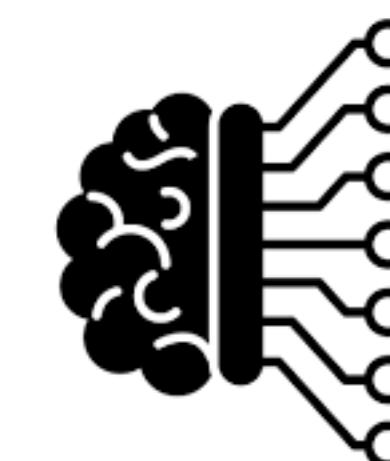
Train



Sample

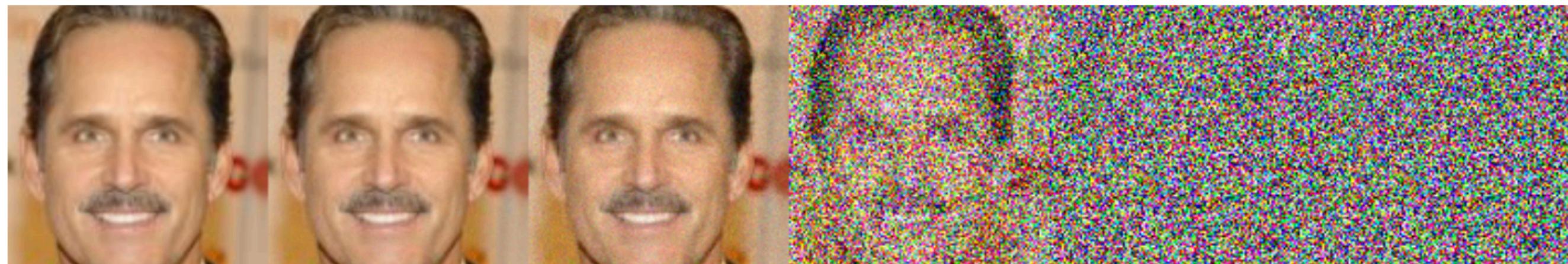


Model



Diffusion models diffuse signal to noise

Original → Degraded



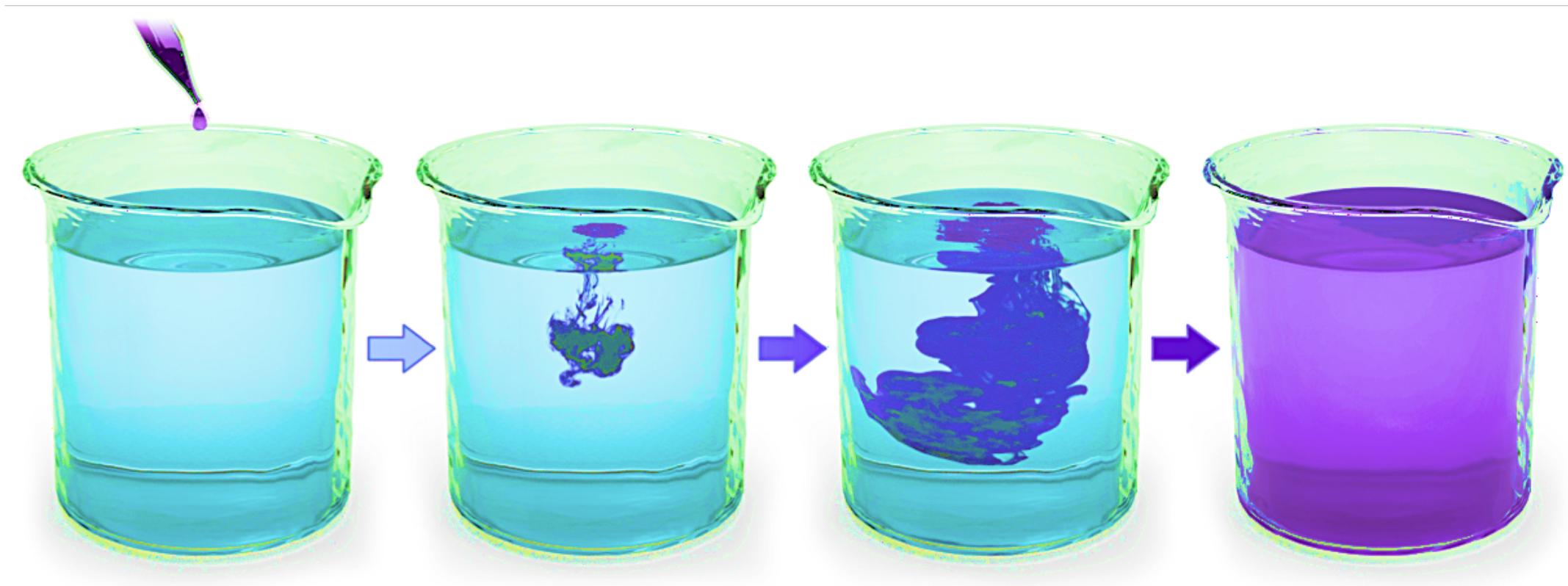
x_0 x_1 x_{t-1} x_t

Forward process

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t, \sqrt{1 - \beta_t}x_{t-1}, \beta_t I)$$

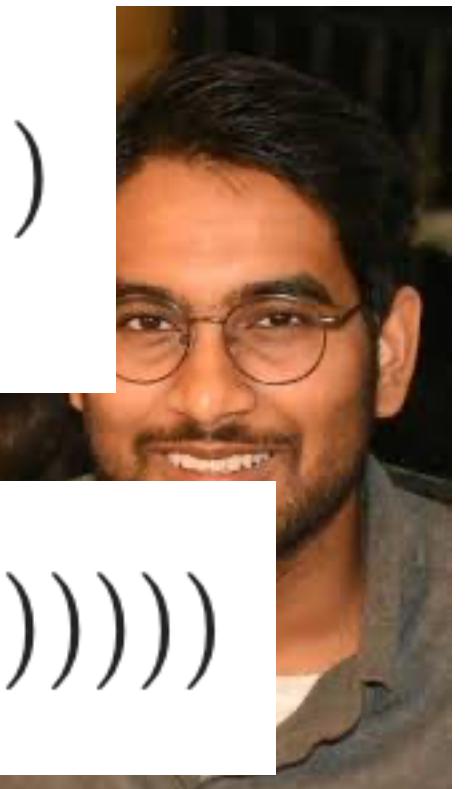
- ϵ is from $\mathcal{N}(0, 1)$

$$q(x_t|x_{t-1}) = \sqrt{1 - \beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon$$



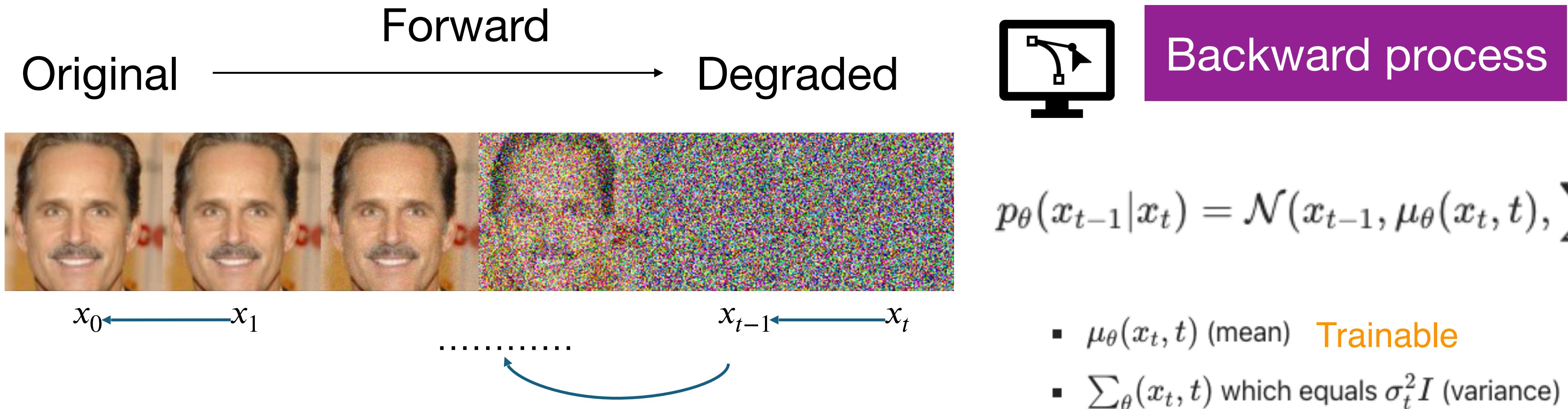
$$q(x_{1:T}|x_0) := \prod_{t=1}^T q(x_t|x_{t-1})$$

$$q_t(q_{t-1}(q_{t-2}(q_{t-3}(\cdots q_1(x_0))))))$$



Auro Patnaik

Neural network learns to predict signal from noise



- The model predicts noise between arbitrary steps, unaware of specific timesteps.
- It treats x_t to x_{t-1} the same across all steps.
- Trained to denoise from one step to the previous, regardless of position in the process.

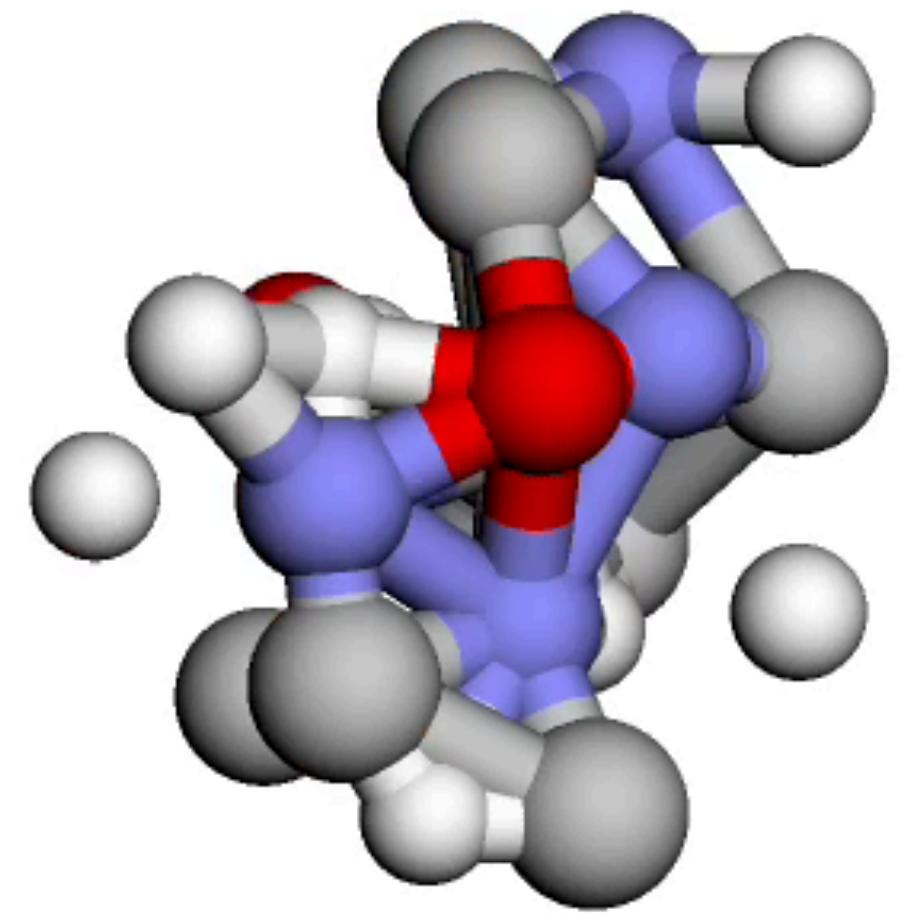
Diffusion model can be used to bias towards molecules we want



$$p_{\theta}(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_{\theta}(x_t, t), \sum_{\theta}(x_t, t))$$

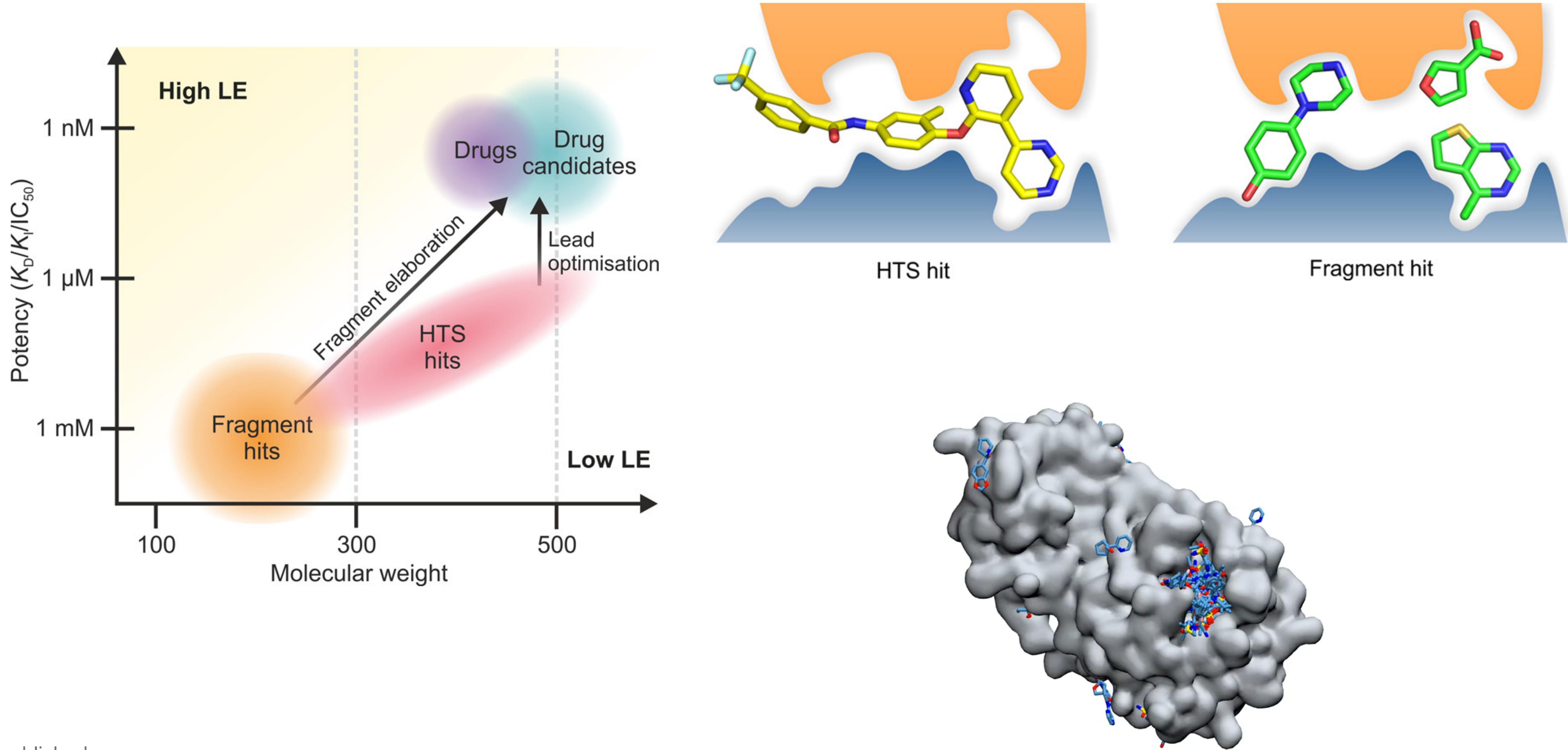
We can refine without training

$$x_t(t, r) = (1 - \phi_t) \bullet x_t + \phi_t \bullet x_R$$

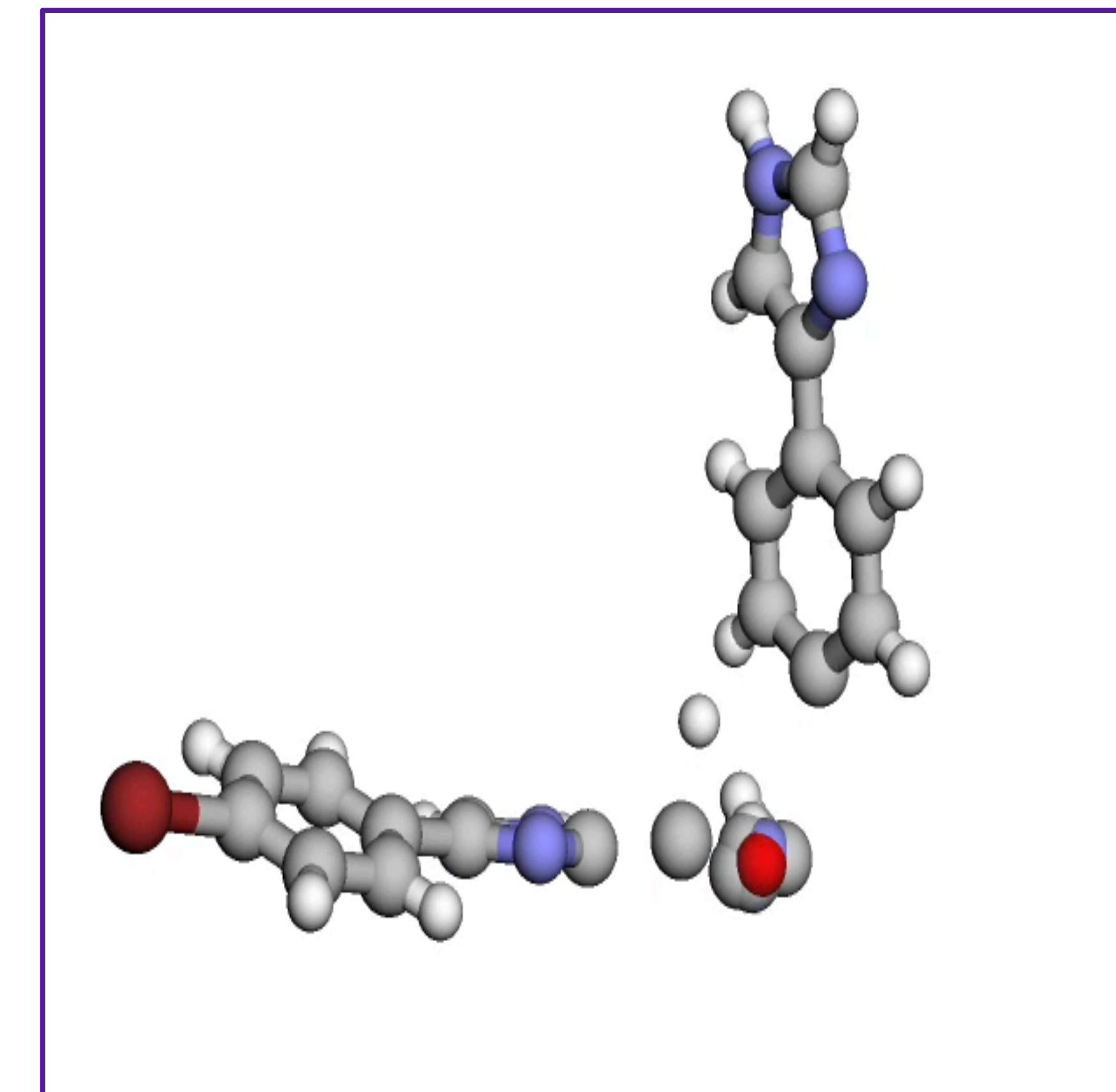
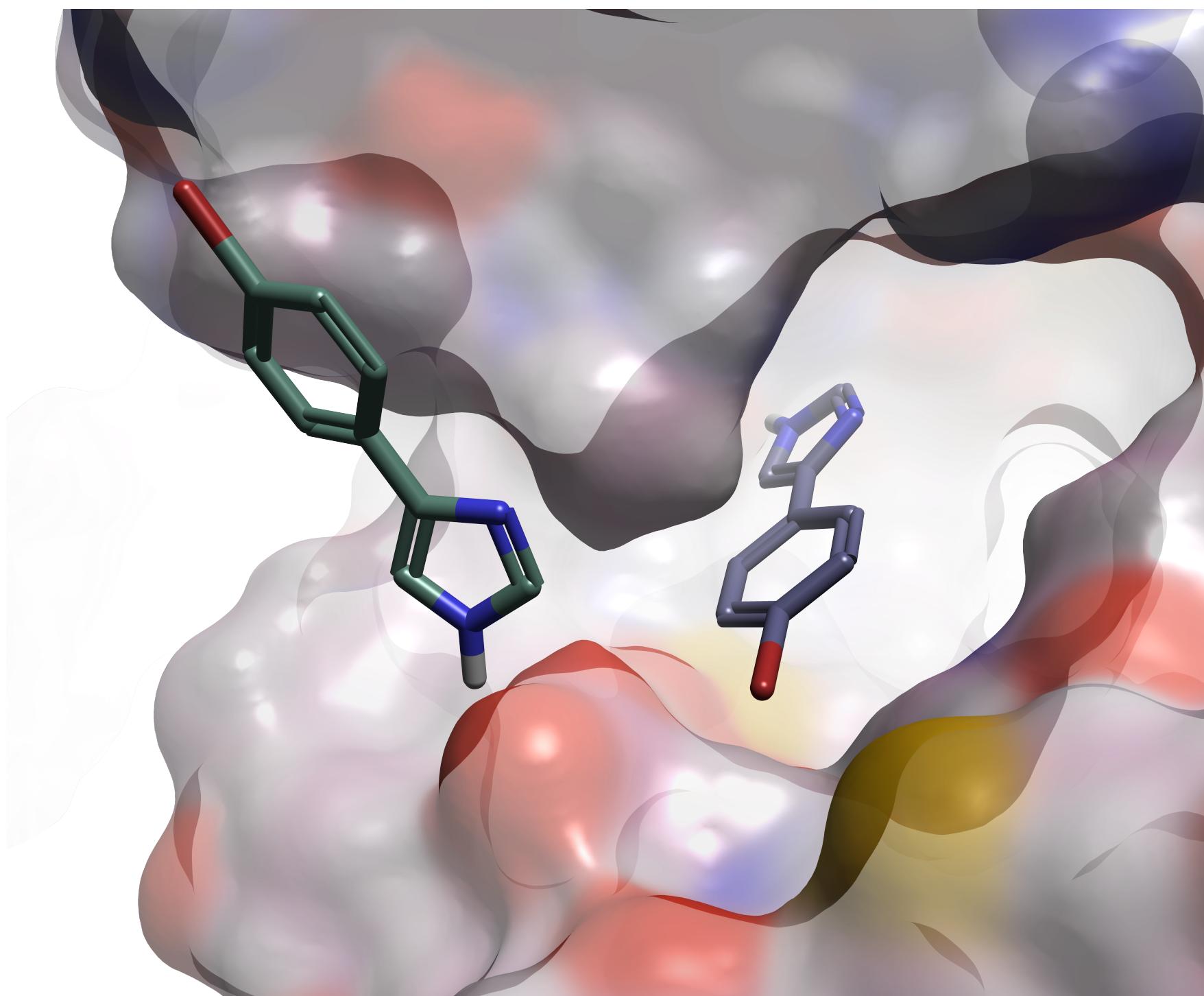


Generating a molecule of caffeine by
biasing a drug diffusion model

Biasing a diffusion model can be useful in drug discovery



It is possible to use diffusion to make new molecules

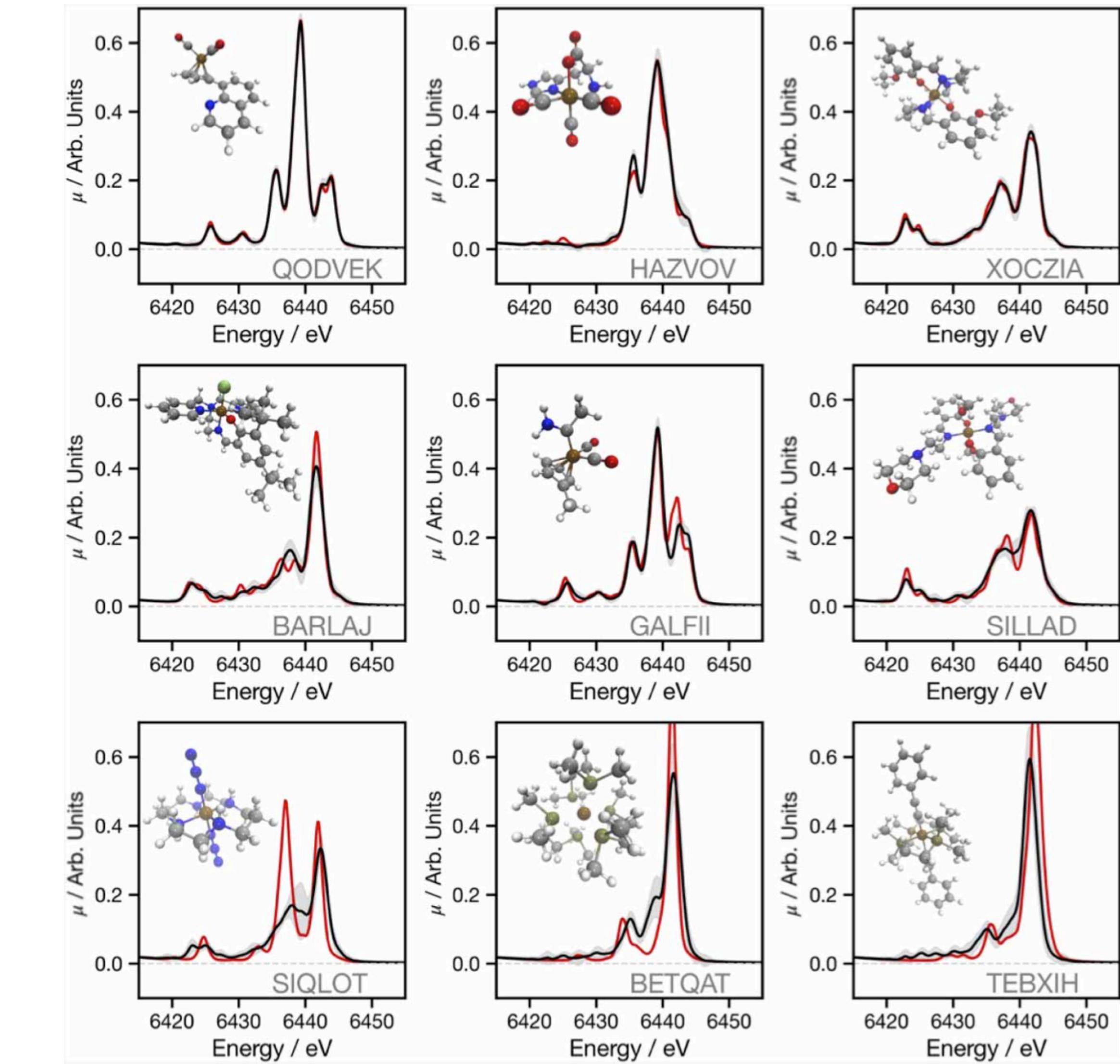


Valence to core X-ray emission spectra

Neural networks can trained to predict X-ray emission spectra

Red - reference spectrum

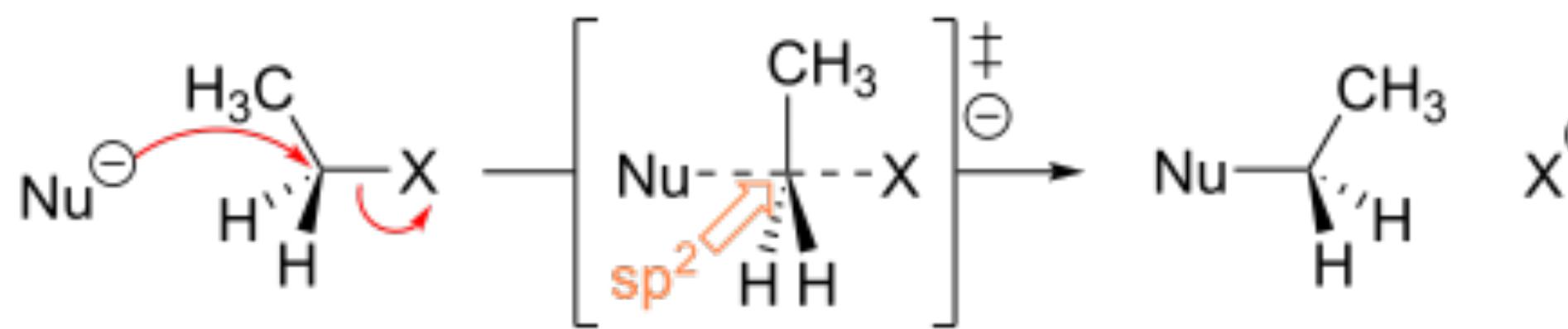
Black - spectrum from NN



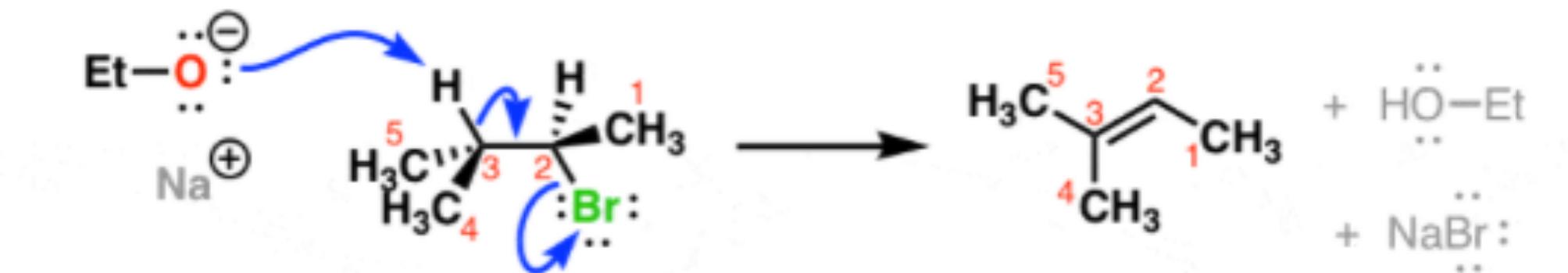
Neural networks can be used to classify geometries



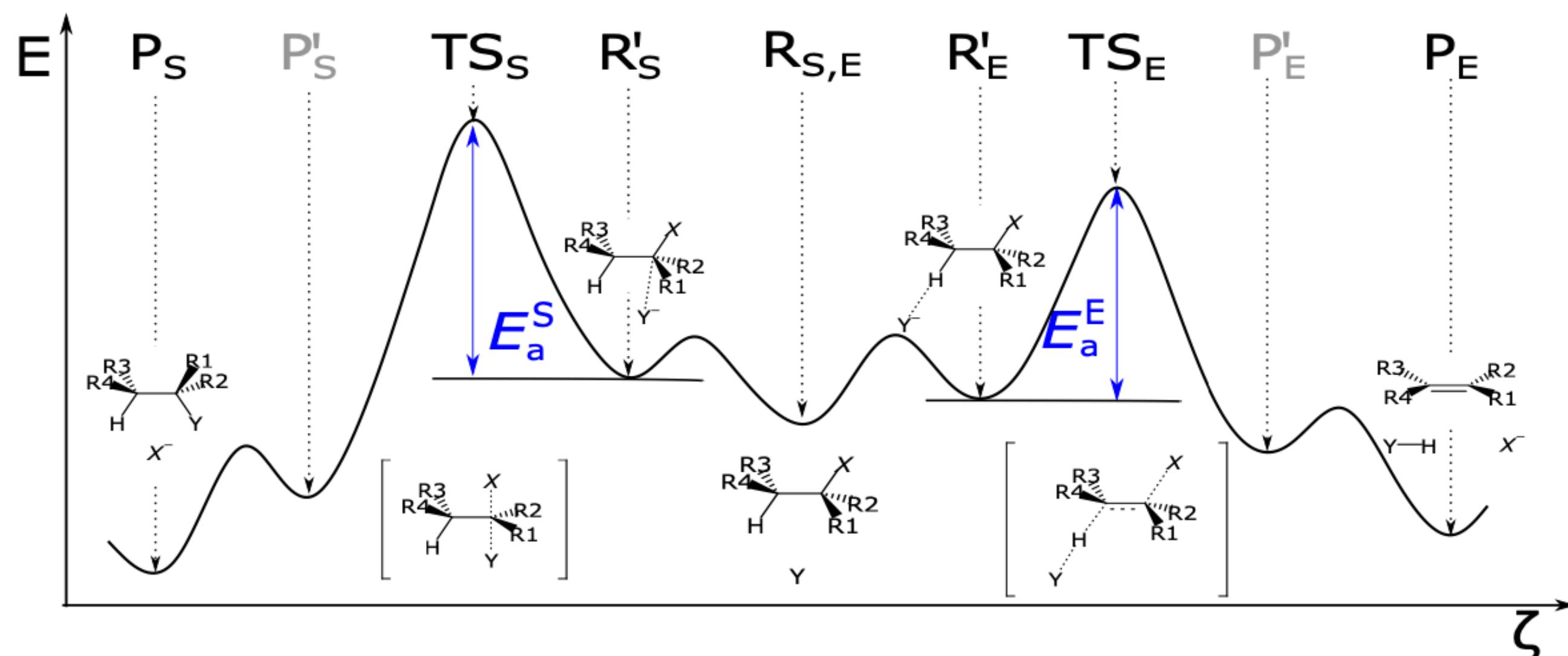
SN2 reaction



E2 reaction



One step! Hydrogen removed must be "anti" to the leaving group (dihedral angle 180°)





*

For questions and comments,
Please send me an email:

antonia.mey@ed.ac.uk

*Generate a picture in the style of van Gough to say thank you to students for listening to the introduction to machine learning