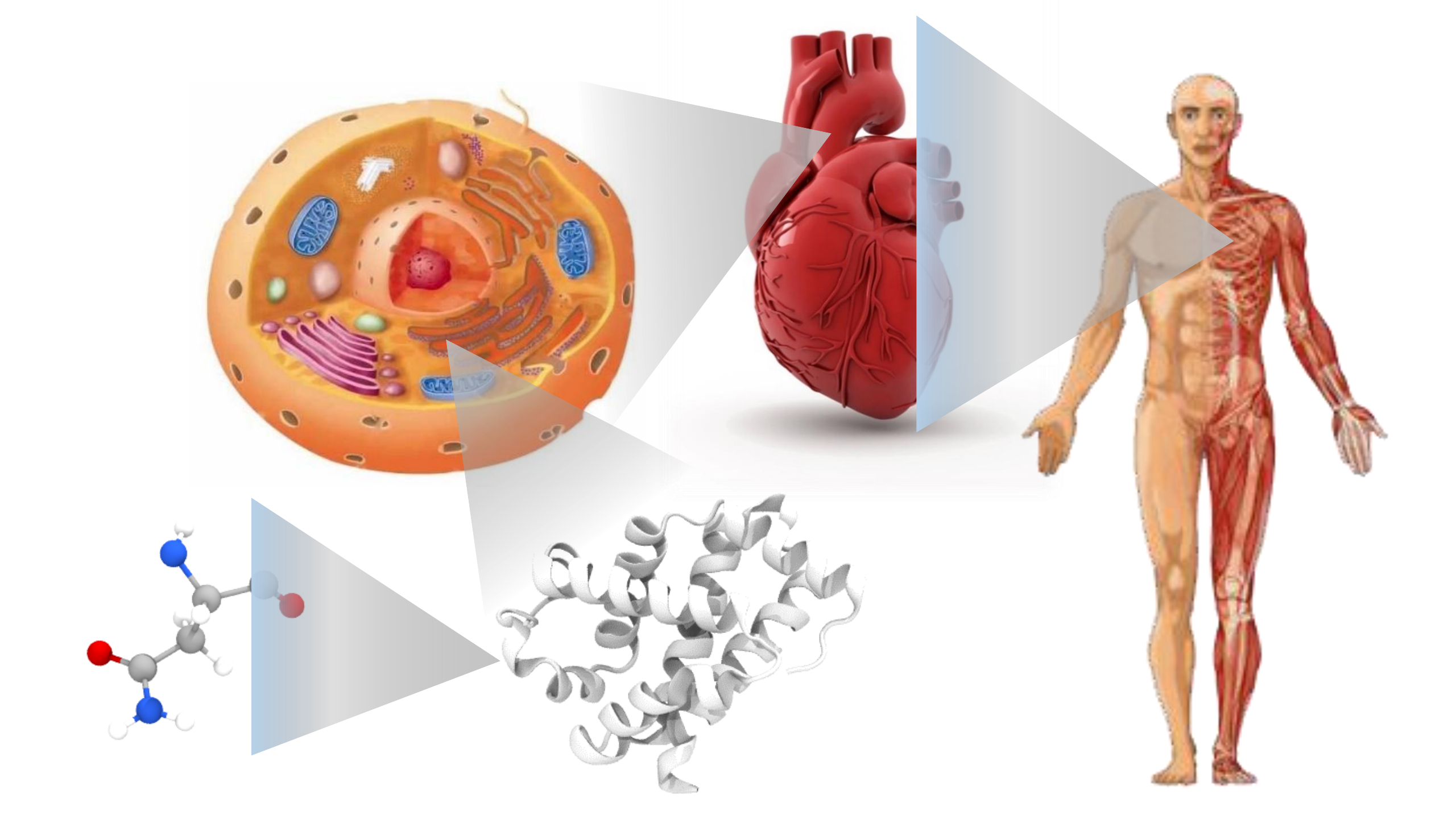
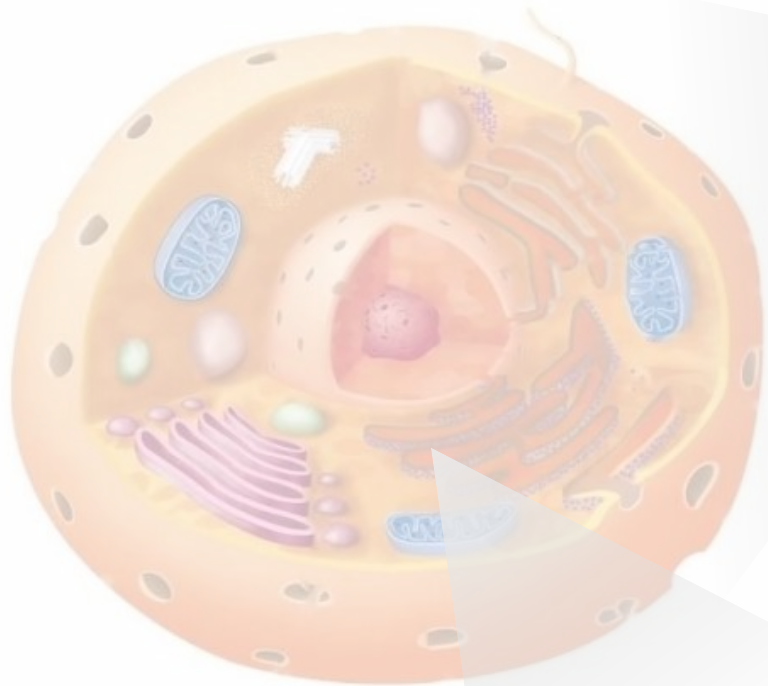
Learning (from) protein dynamics

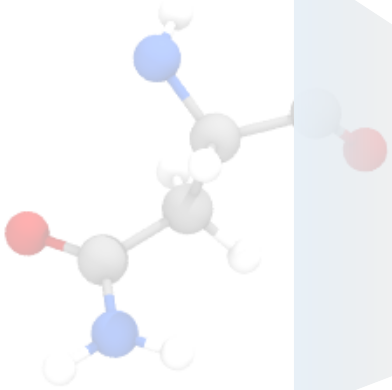
Matteo Degiacomi

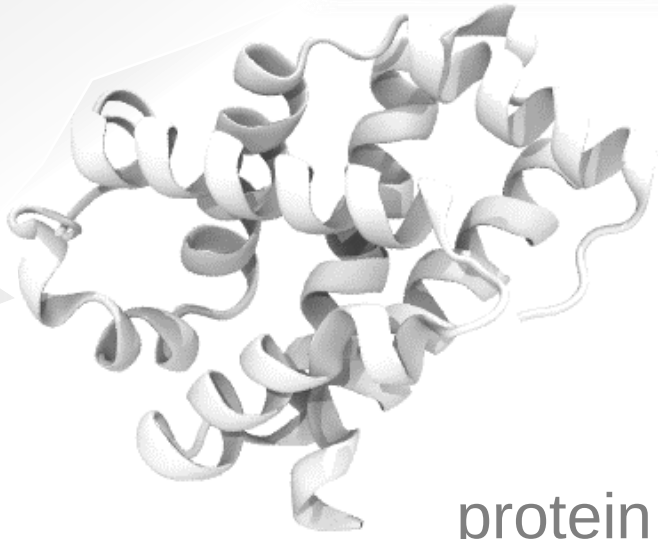
matteo.degiacomi@ed.ac.uk

www.degiacomi.org

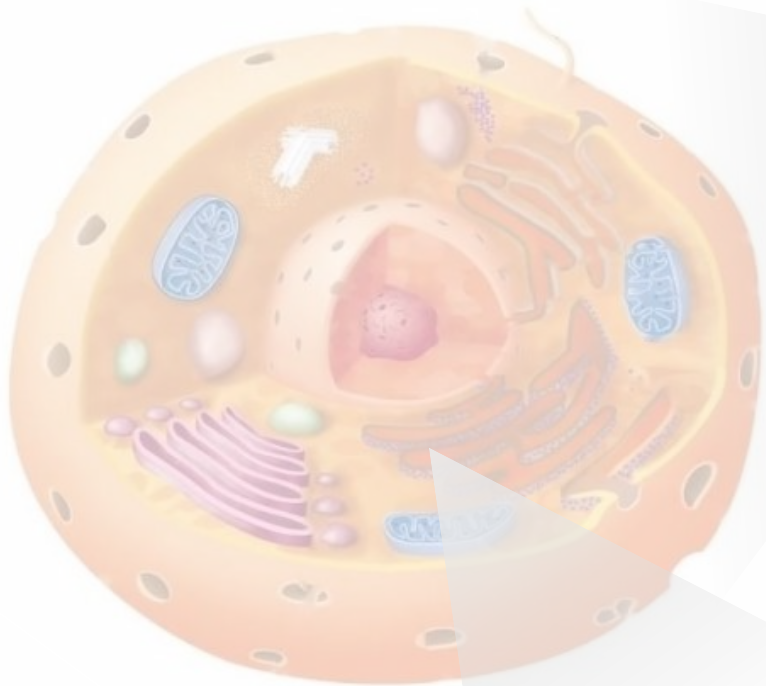


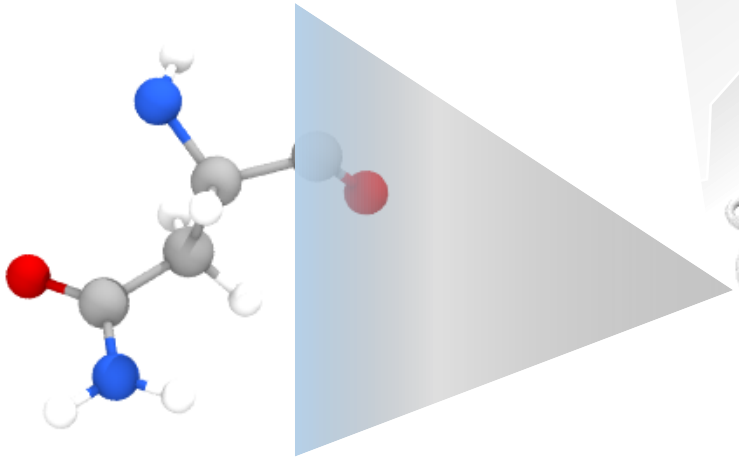


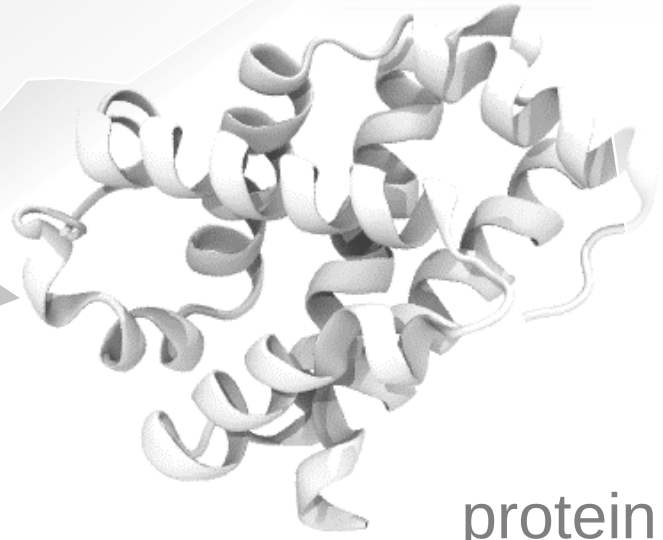




protein



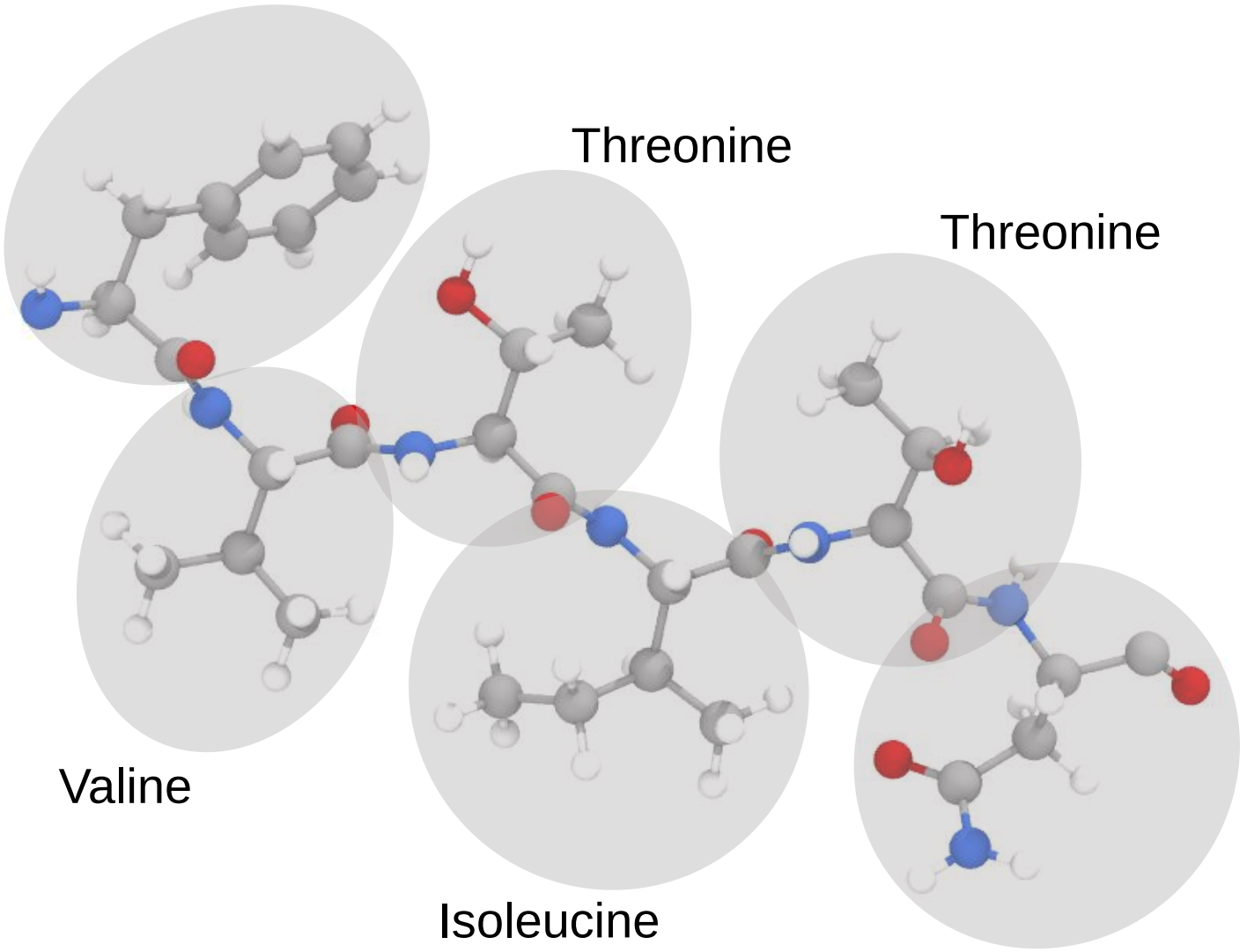




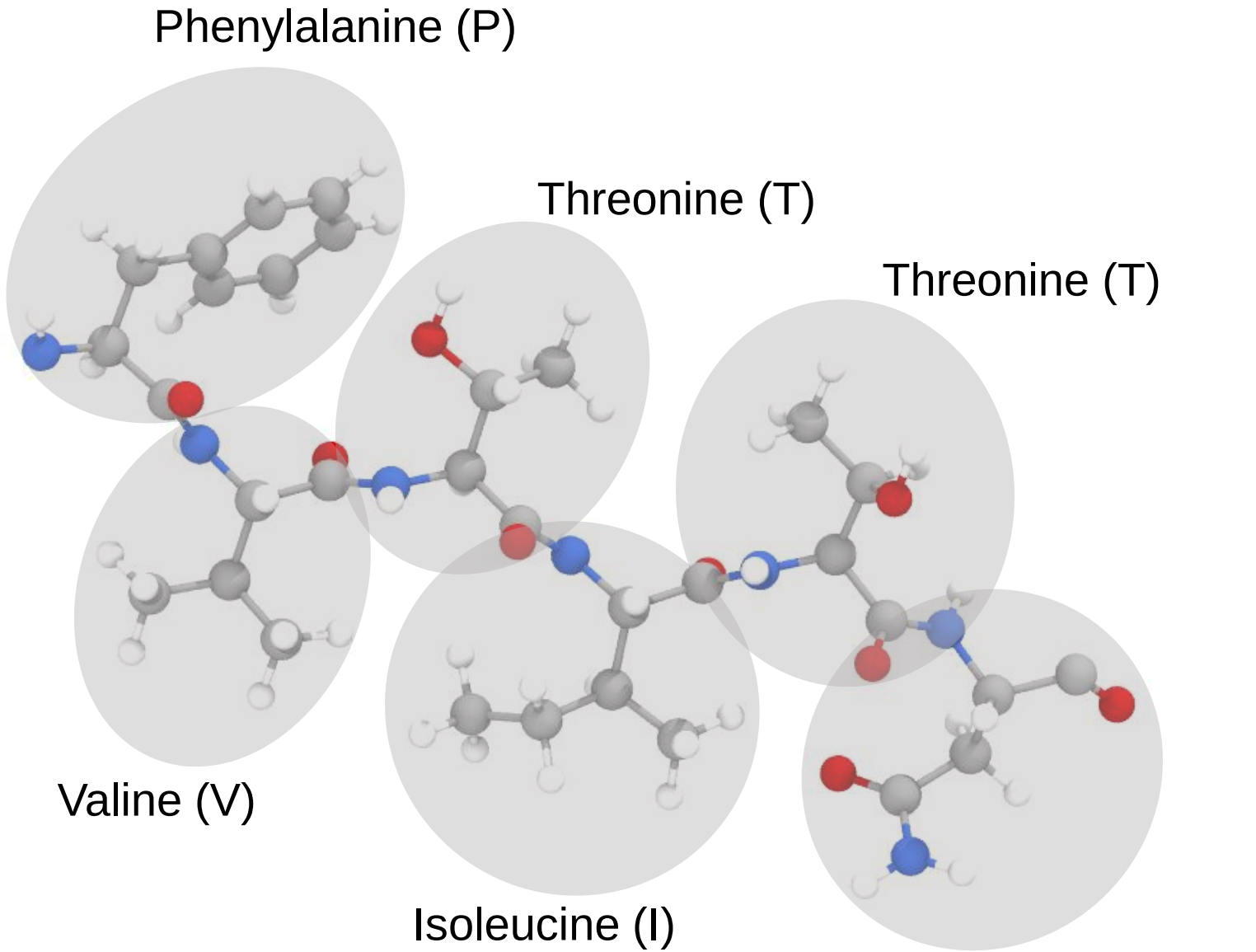
amino acid

protein

Amino acids form linear chains by connecting to each other via their backbone (*polymerise*)



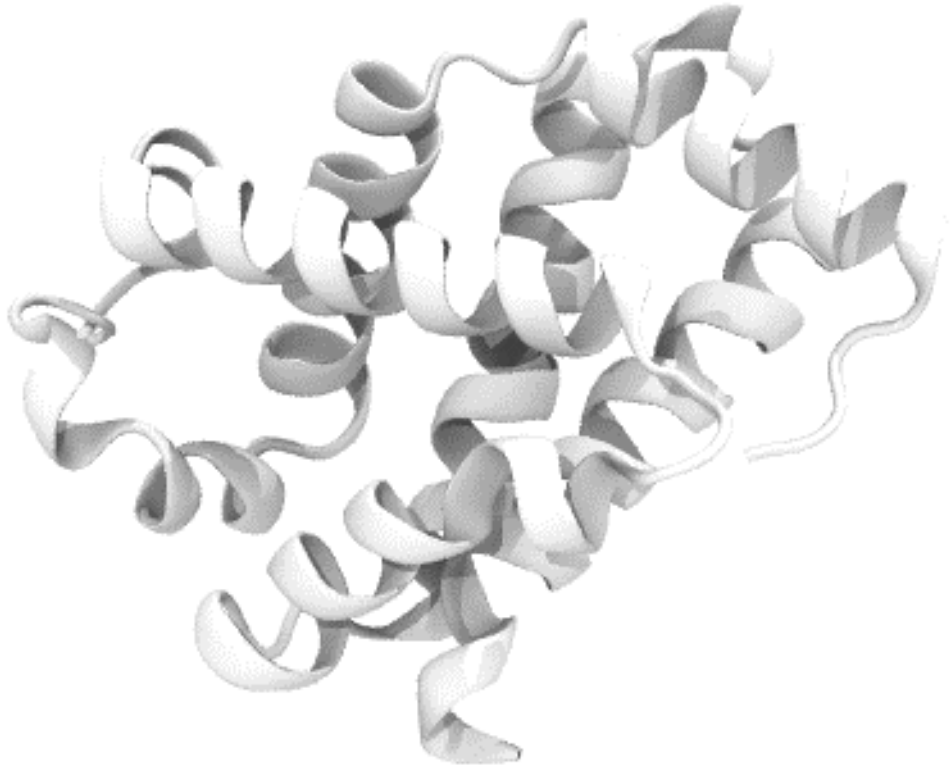
Amino acids form linear chains by connecting to each other via their backbone (*polymerise*)



Primary sequence → **PVTITN**

Proteins: ~300 amino acids on average

Similar sequence, similar fold (usually)



folding

AILGPATTLKMRSGGA…

Anfinsen's dogma

The three-dimensional structure of a protein in its native environment is solely determined by its amino acid sequence.

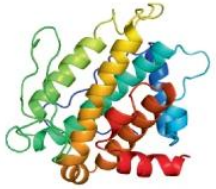
Christian Anfinsen. Principles that govern the folding of protein chains. *Science*, 1973

Scientific Background to the Nobel Prize in Chemistry 2024

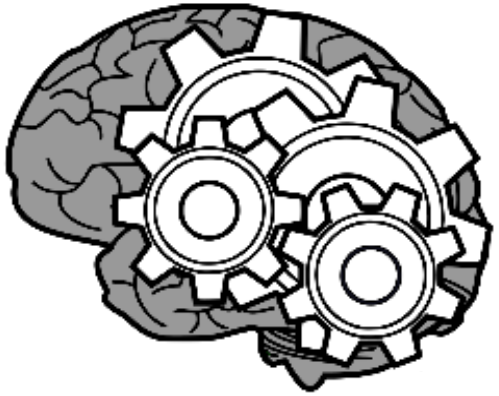
COMPUTATIONAL PROTEIN DESIGN AND PROTEIN STRUCTURE PREDICTION

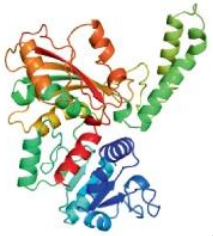
training

NITPLAKRTYNYRAVL…

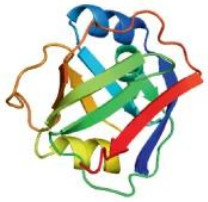


HTIWKLSRLWSLQ…





ALRIKAIVVPRILGPQ…



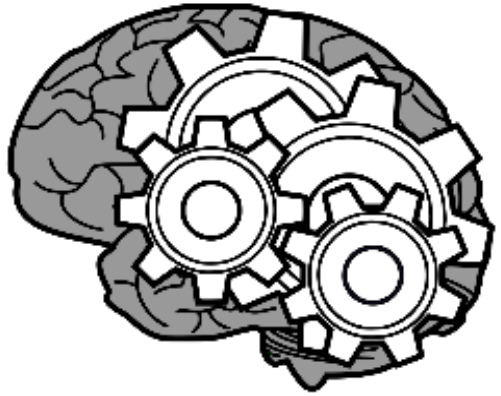
LIYRICMPGILCYEND…

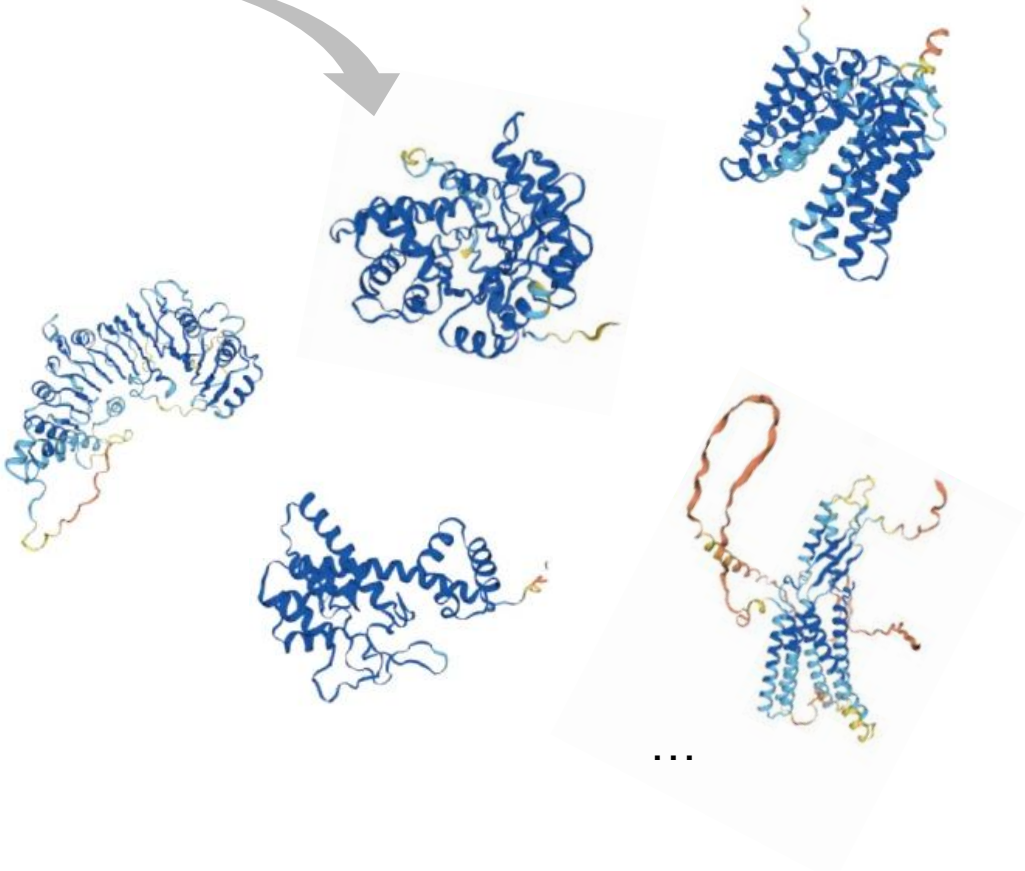
Scientific Background to the Nobel Prize in Chemistry 2024

COMPUTATIONAL PROTEIN DESIGN AND PROTEIN STRUCTURE PREDICTION

prediction

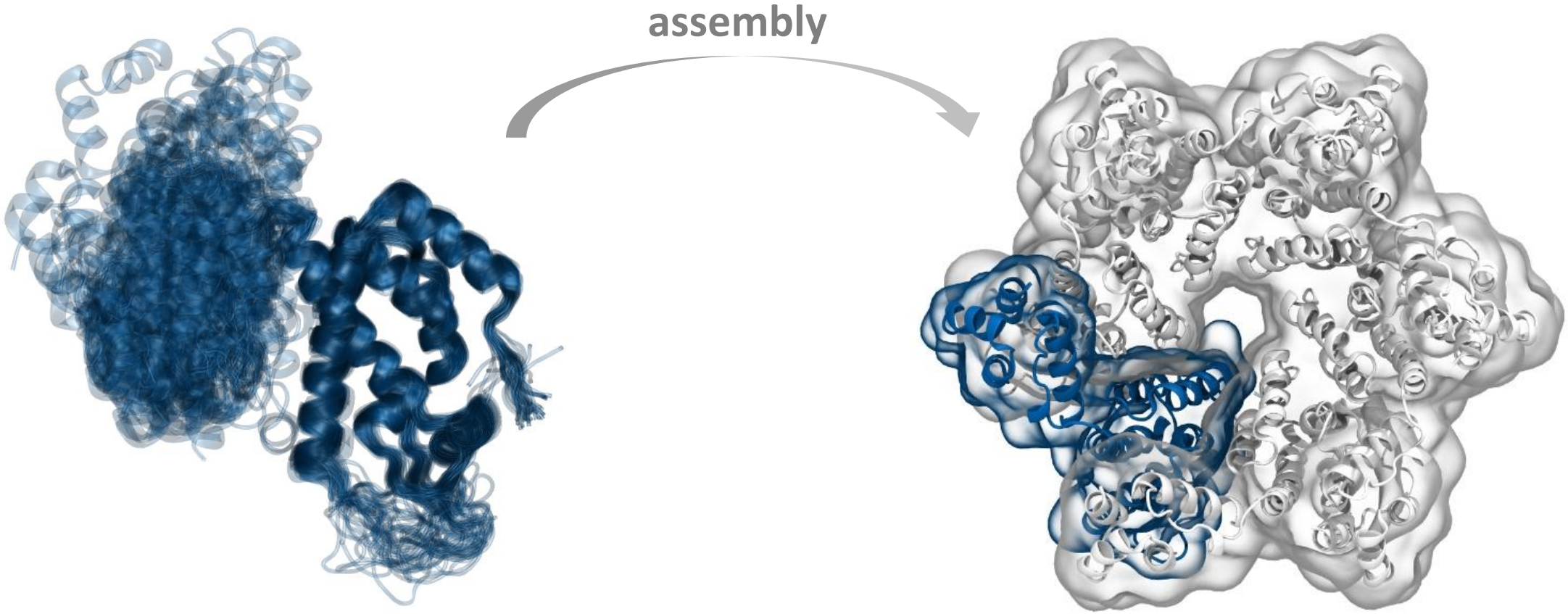
MAPVYEGMAS…
HVQVFSPHTL…
QSSAFCSVKK…
LKIEPSSNWD…
MTGYGSHSKV…
…



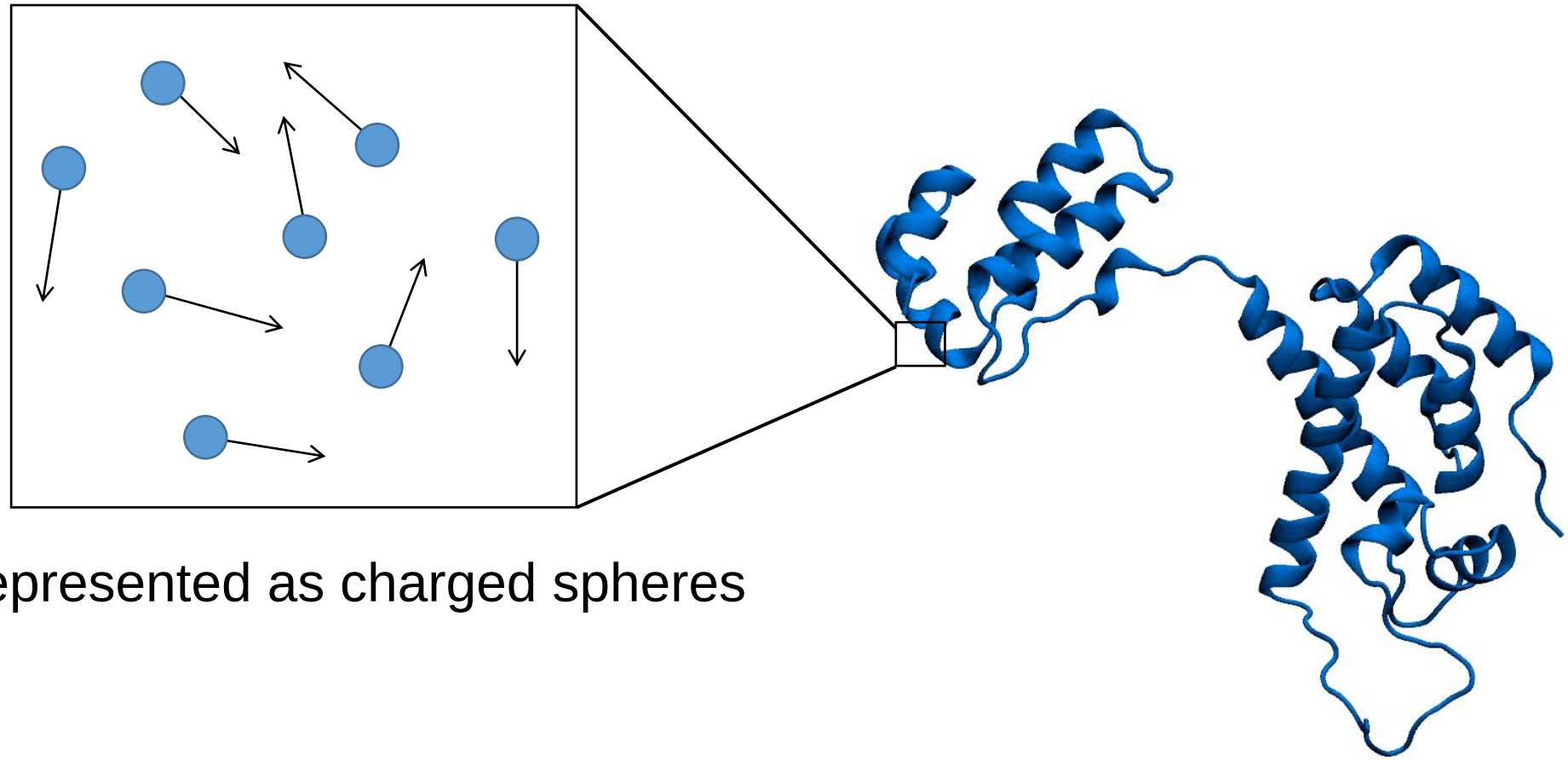


…

Structure *and dynamics* determine the function

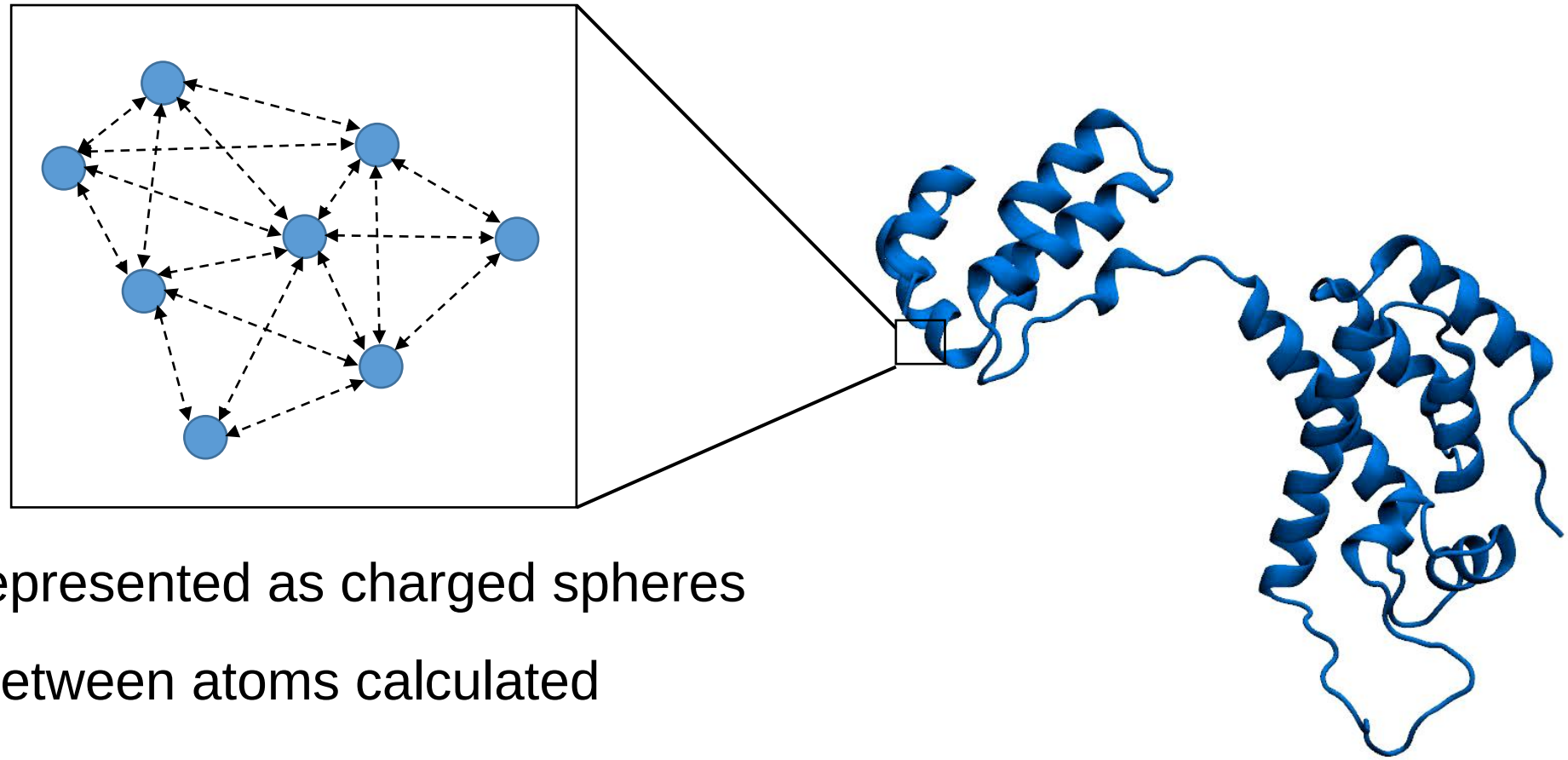


Molecular Dynamics (MD) simulations



Atoms represented as charged spheres

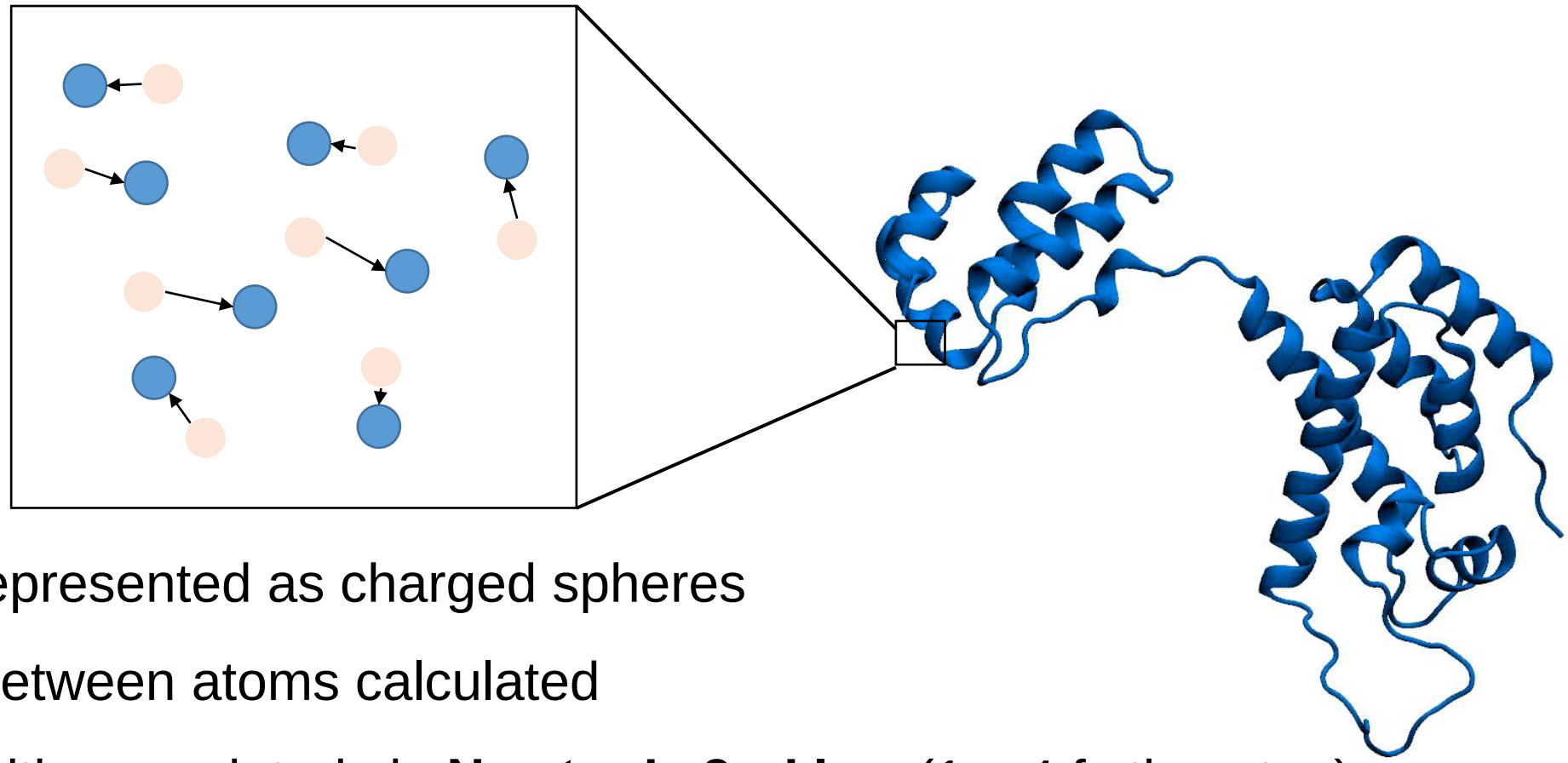
Molecular Dynamics (MD) simulations



Atoms represented as charged spheres

Forces between atoms calculated

Molecular Dynamics (MD) simulations



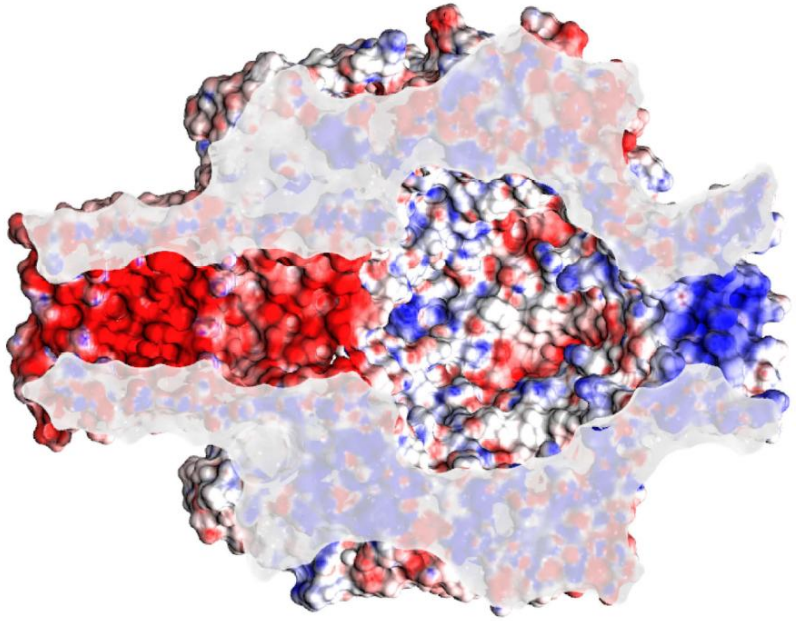
Atoms represented as charged spheres

Forces between atoms calculated

Atom positions updated via **Newton's 2nd law** (1 – 4 fs timestep)

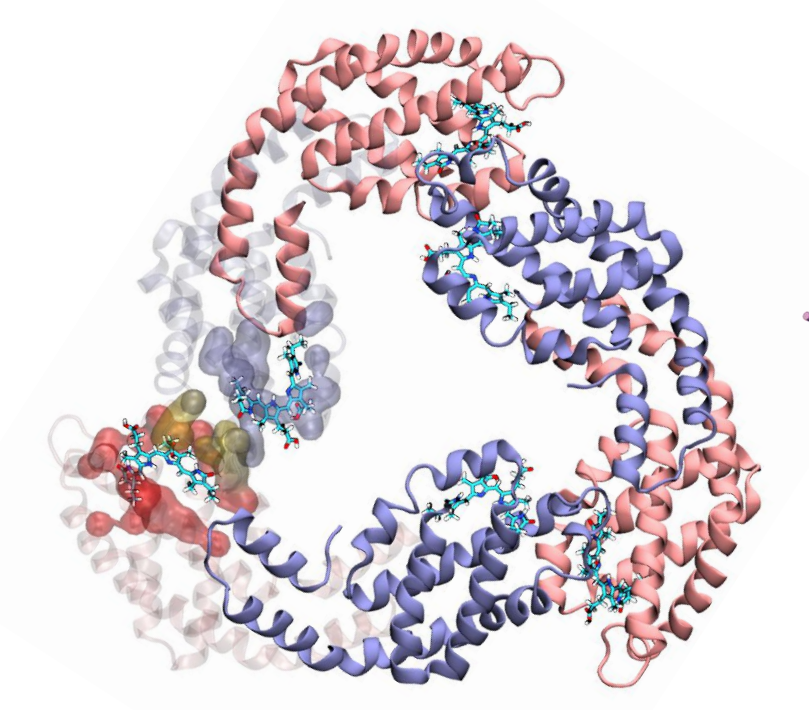
Understanding protein function via simulations

Membrane pore ion selectivity



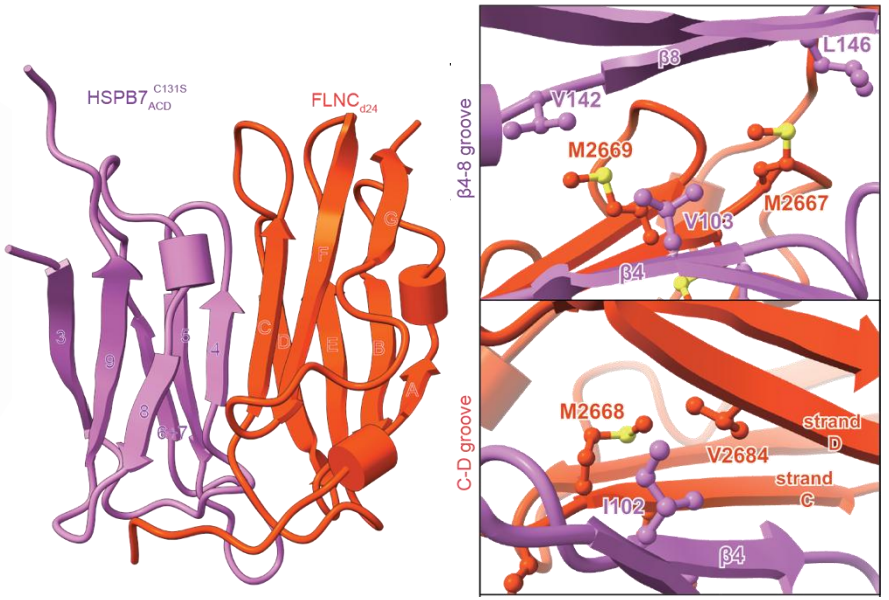
J. Bruggisser et al.,
***EMBO Reports**, 2022*

Light harvesting mechanics



A. Guillén-García et al.,
***Nature Comms.**, 2022*

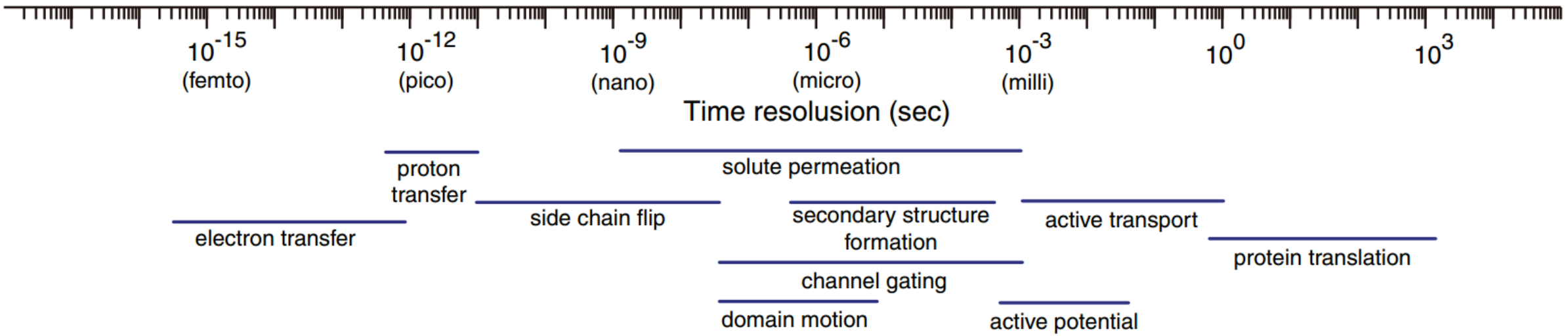
Post-translational modifications



Z. Wang et al.,
***Nature Comms.**, 2025*

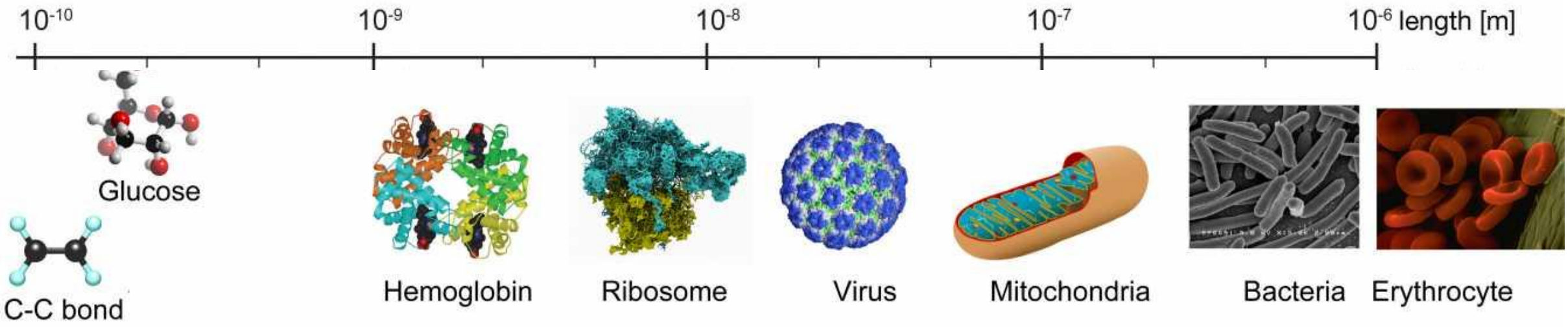
Timescales in biology

Molecular Dynamics simulations



Sizes in biology

Molecular Dynamics simulations

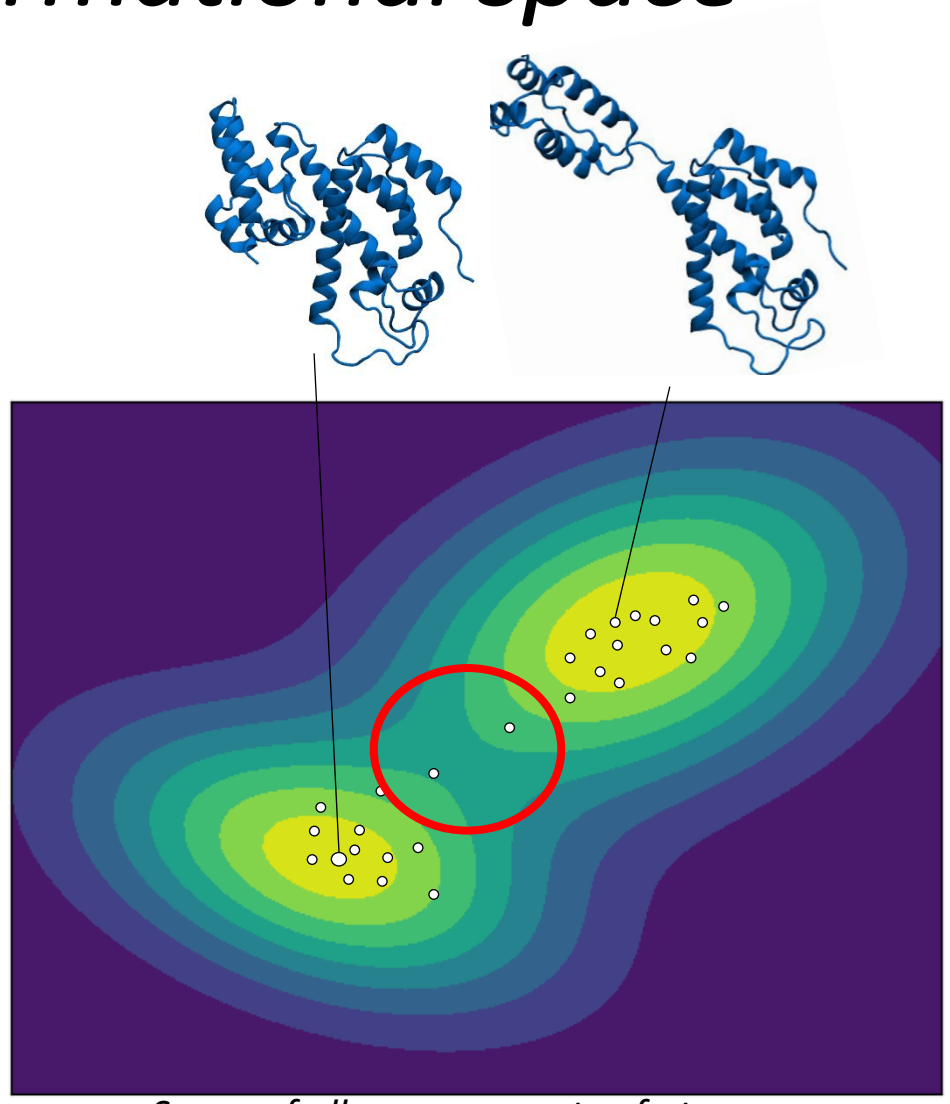


Proteins sample a *conformational space*

Question:
can we overcome MD sampling limitations using generative ML models?

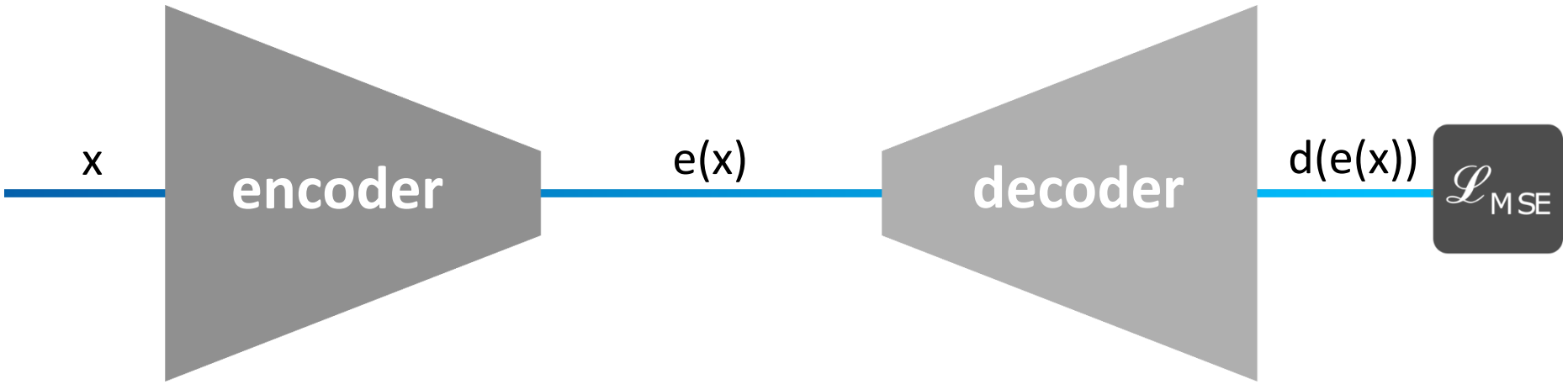
Hurdles:

- High-dimensional space (>1000D)
- Sparse and biased sampling
- Outliers matter
- Ground truth paucity



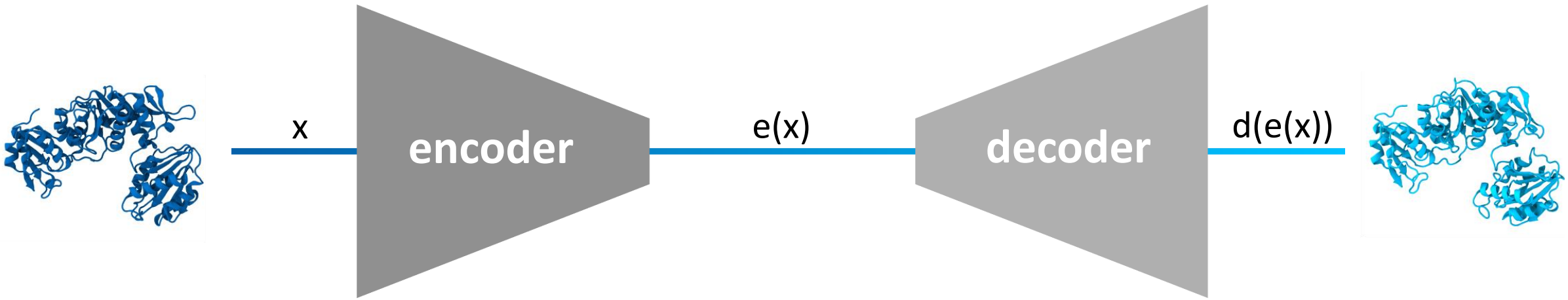
Space of all arrangements of atoms

Can a generative model learn a protein conformational space?



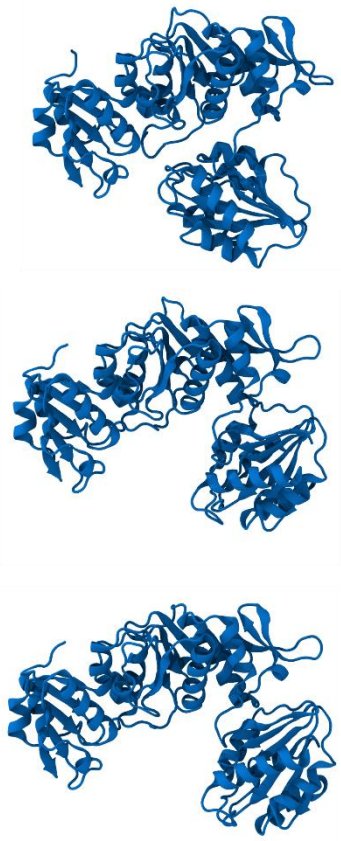
M.T. Degiacomi, ***Structure***, 2019

Can a generative model learn a protein conformational space?

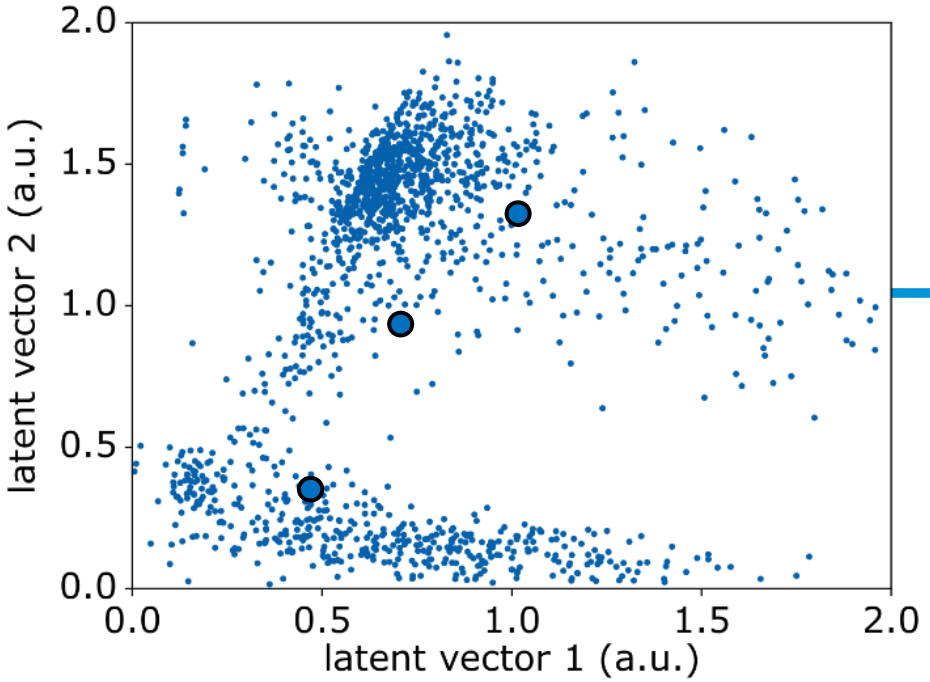


M.T. Degiacomi, ***Structure***, 2019

Can a generative model learn a protein conformational space?

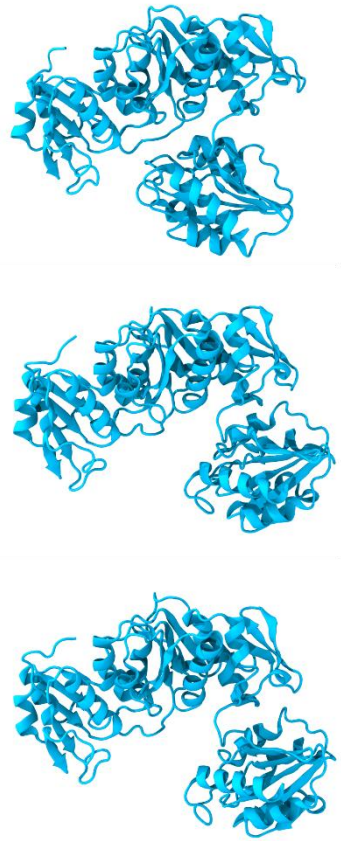


...



decoder

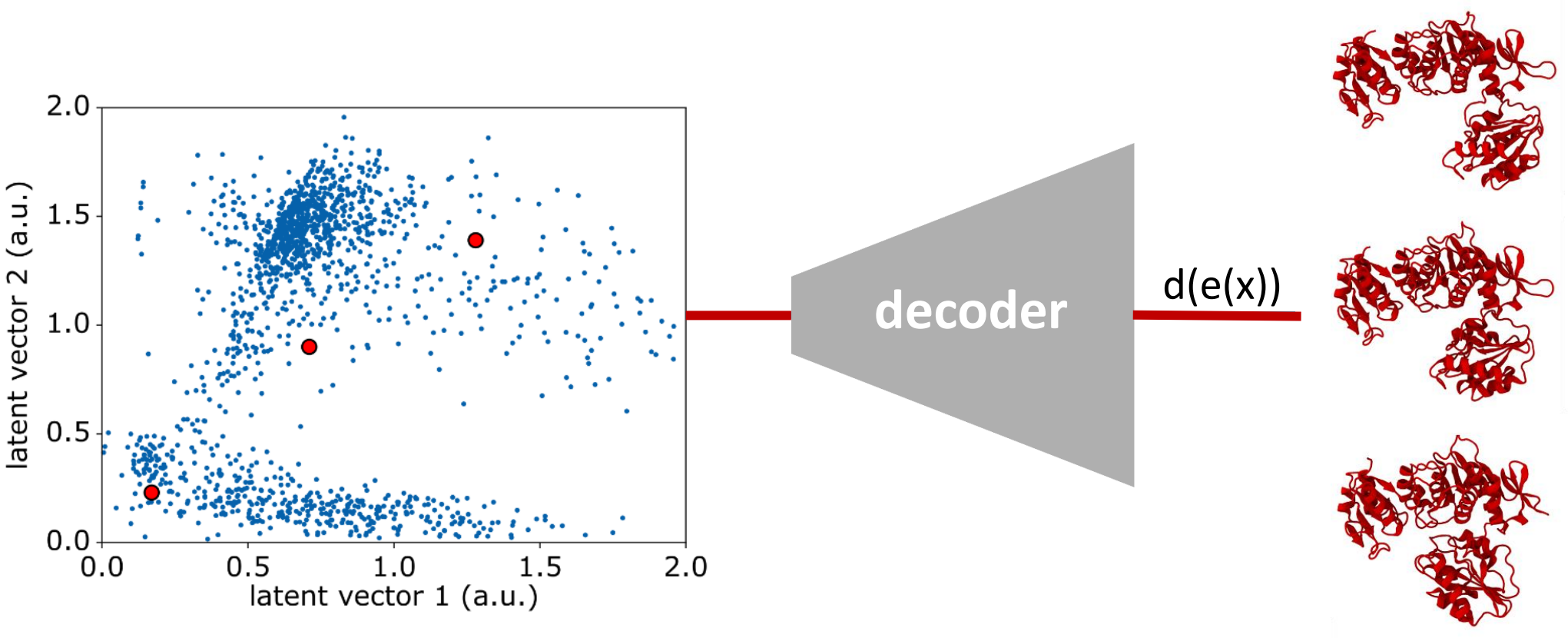
d(e(x))



...

M.T. Degiacomi, ***Structure***, 2019

Can a generative model learn a protein conformational space?



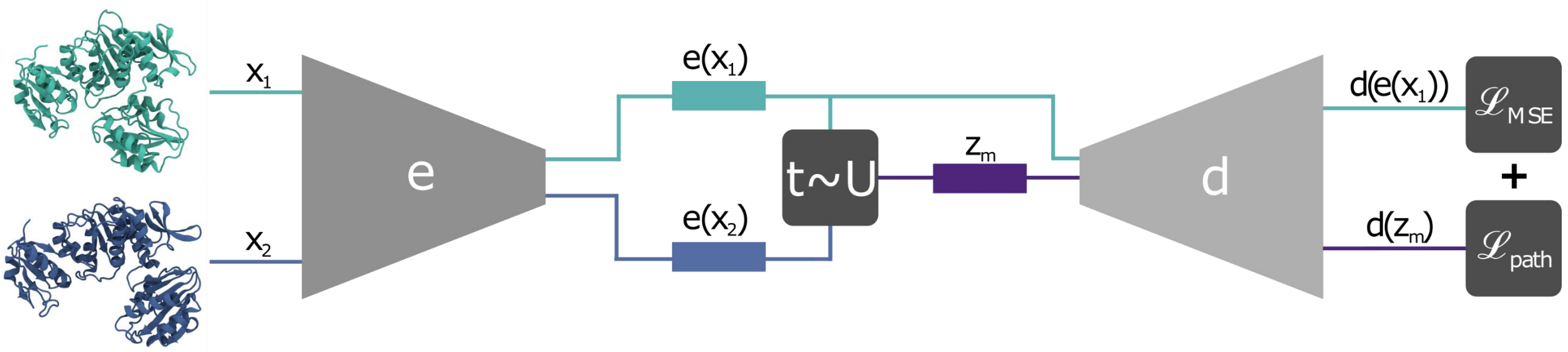
M.T. Degiacomi, ***Structure***, 2019

Integrating physical and topological knowledge into learning



Chris Willcocks *Venkata Ramaswamy* *Samuel Musson*

- Use an architecture inspired by image processing
- Training method encouraging model generalisation

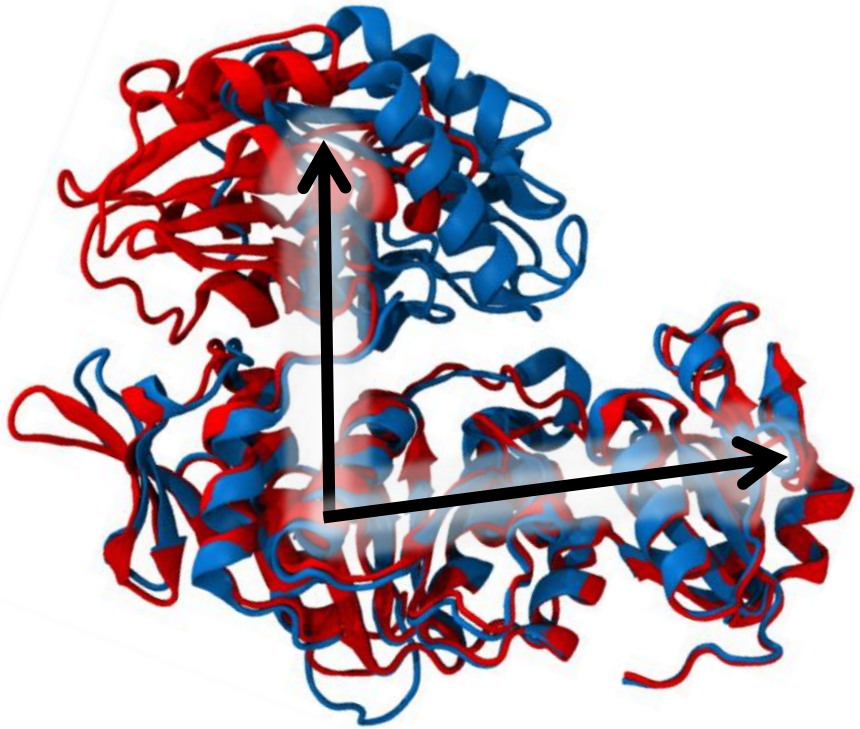


www.github.com/Degiacomi-Lab/molearn

V.K. Ramaswamy *et al.*, ***PRX***, 2021

Predicting transition paths

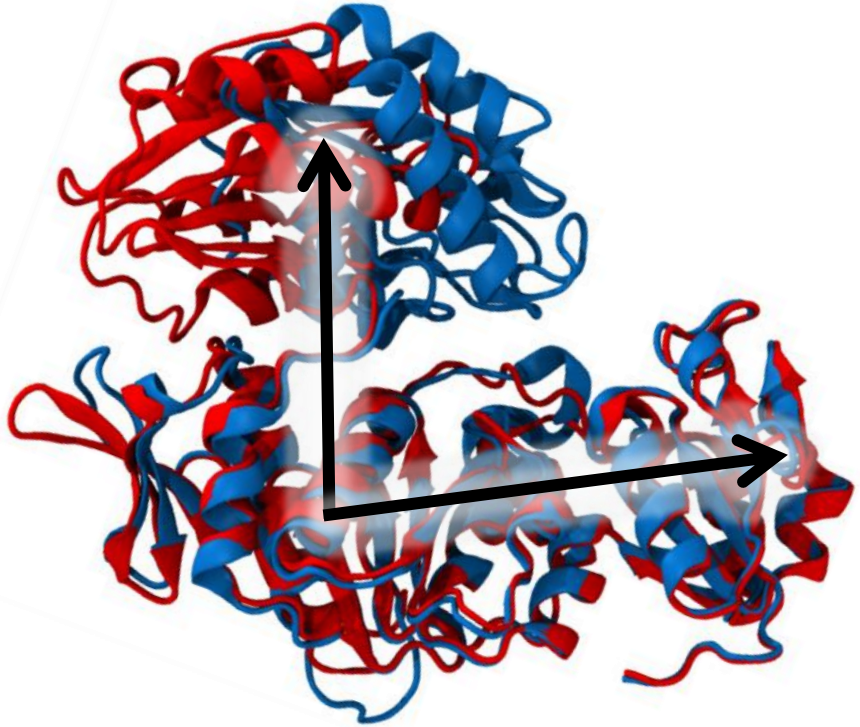
MurD features known **closed** (ligand bound) and **open** (unbound) states



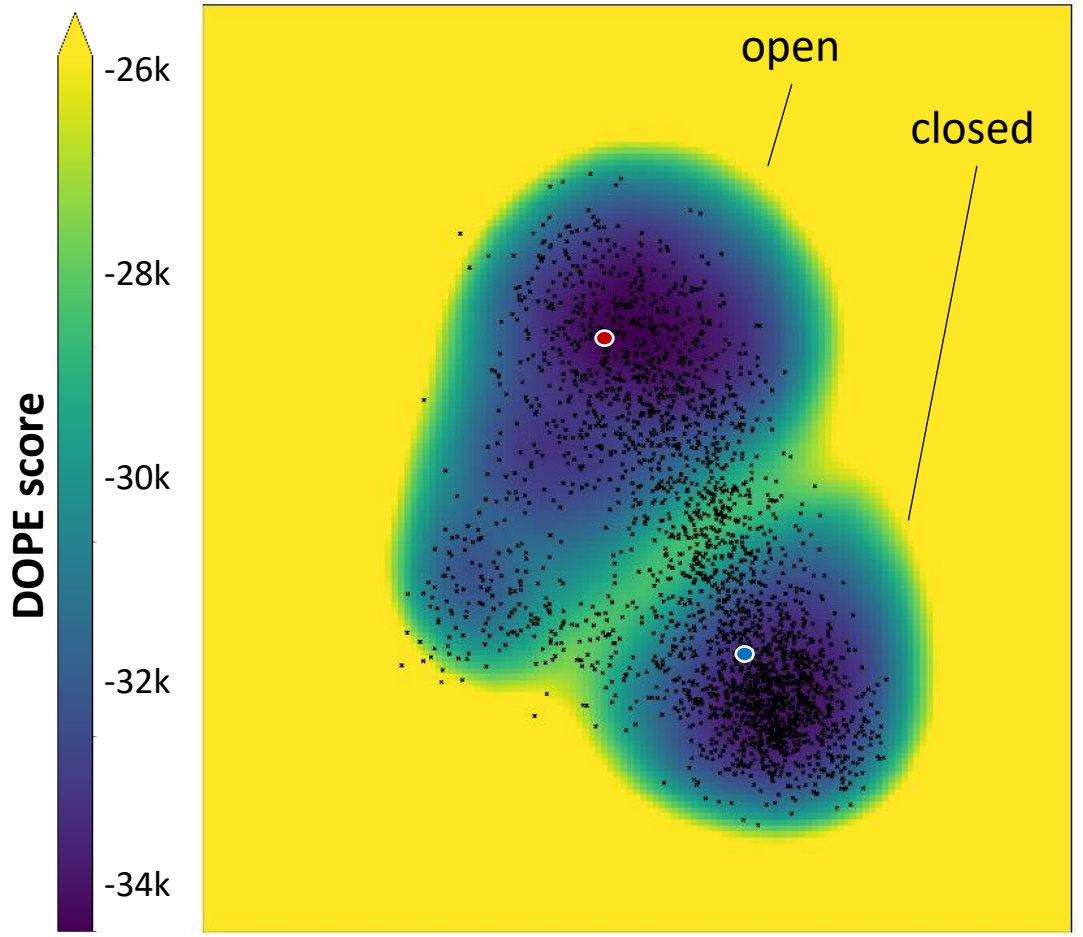
V.K. Ramaswamy *et al.*, ***PRX***, 2021

Predicting transition paths

MurD features known **closed** (ligand bound) and **open** (unbound) states

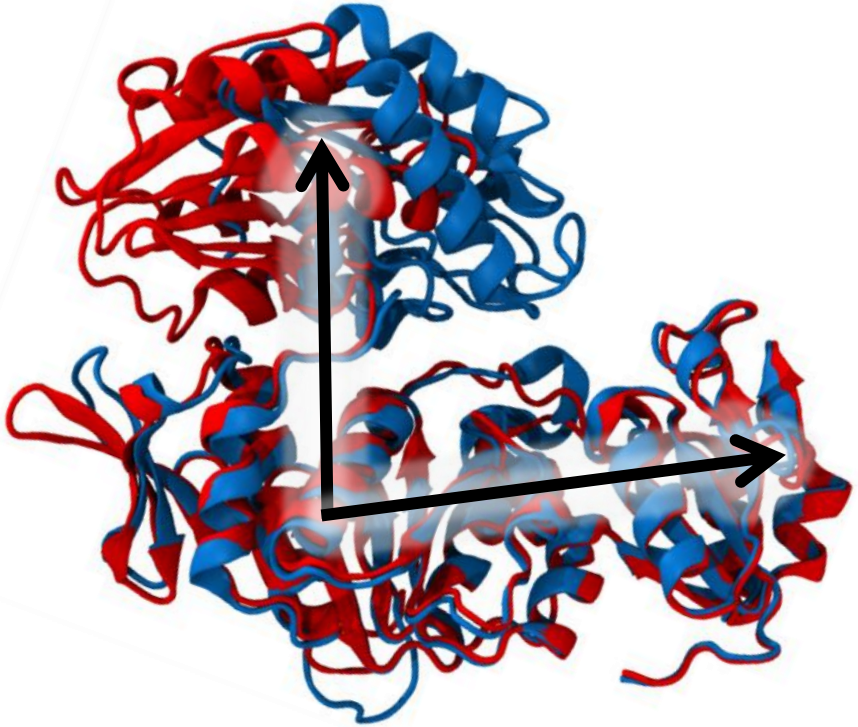


- Train *molearn* with simulation of each state

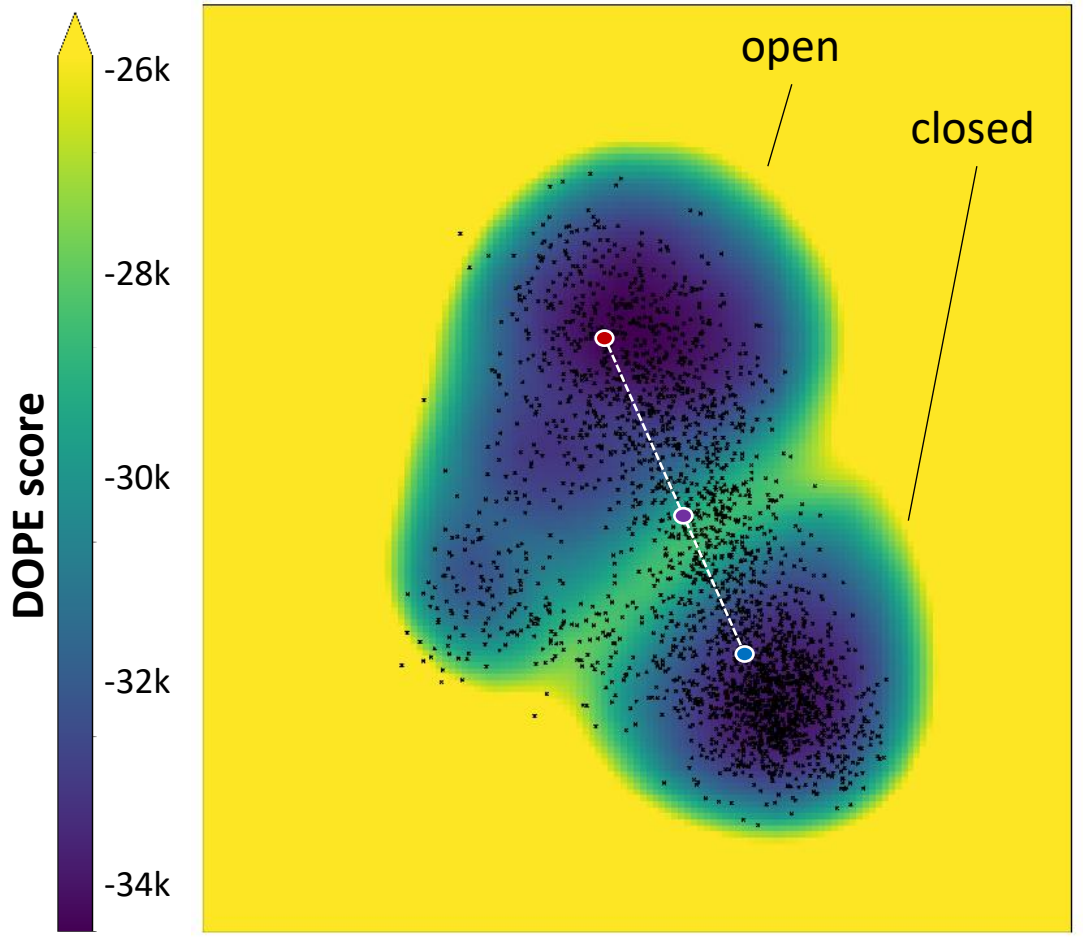


Predicting transition paths

MurD features known **closed** (ligand bound) and **open** (unbound) states

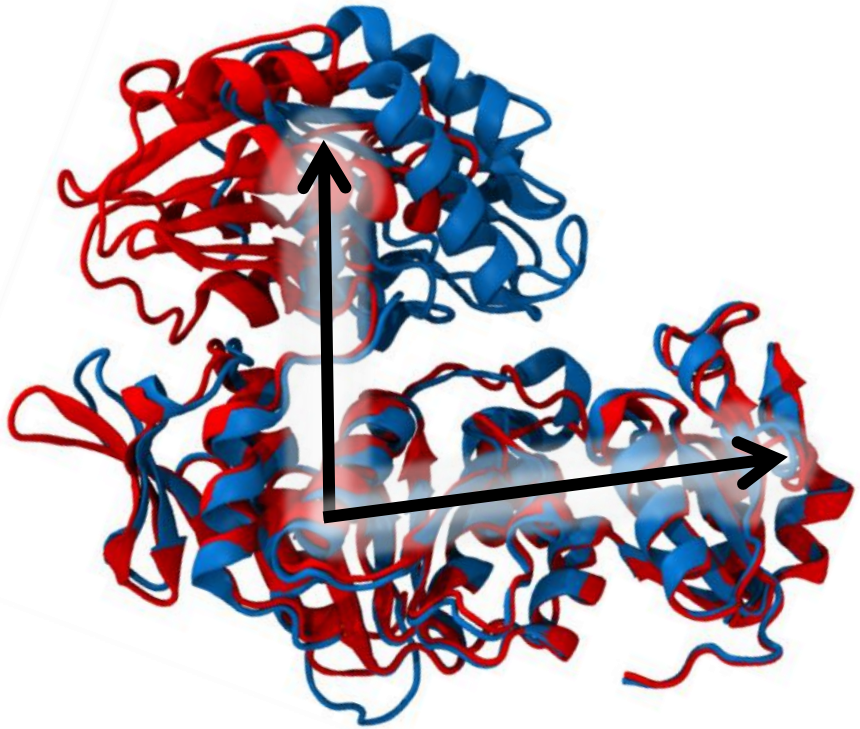


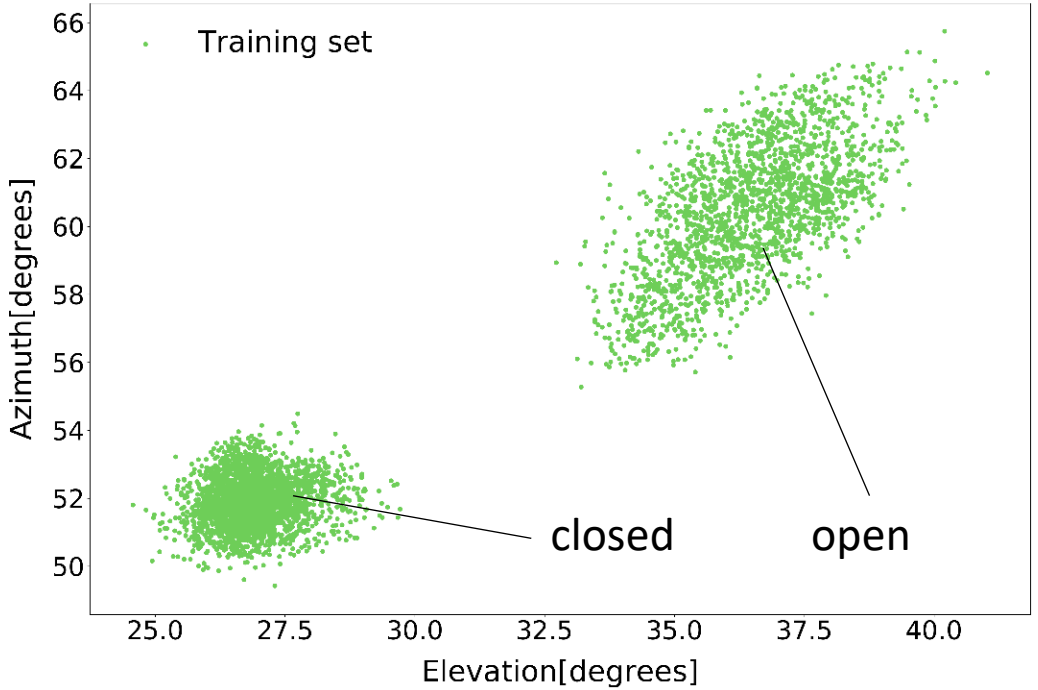
- Train *molearn* with simulation of each state
- Predict transition between two states



Predicting transition paths

MurD features known **closed** (ligand bound) and **open** (unbound) states



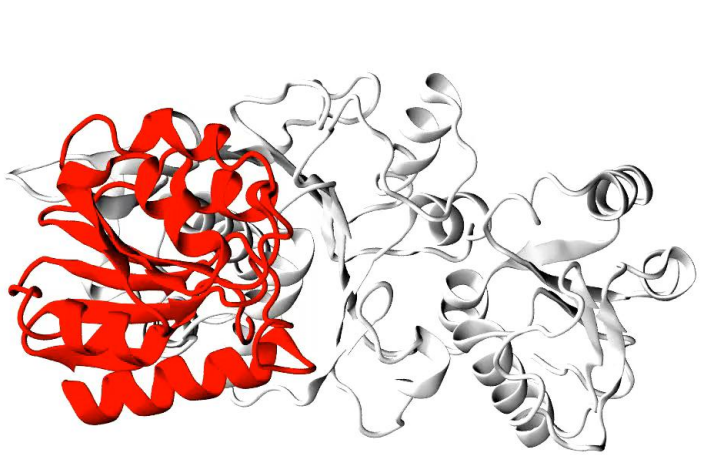


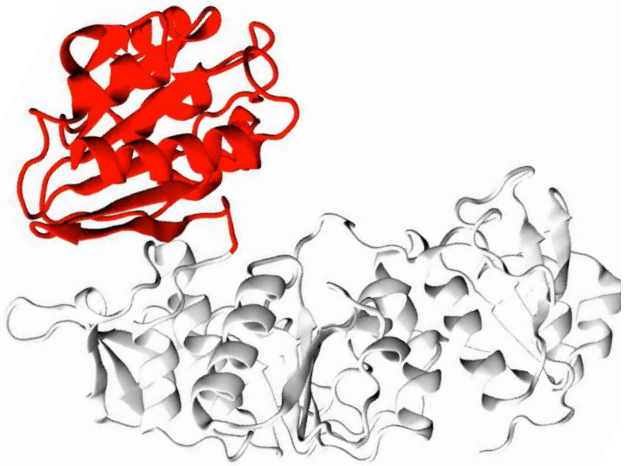
- Train *molearn* with MD of each state
- Predict transition between two states

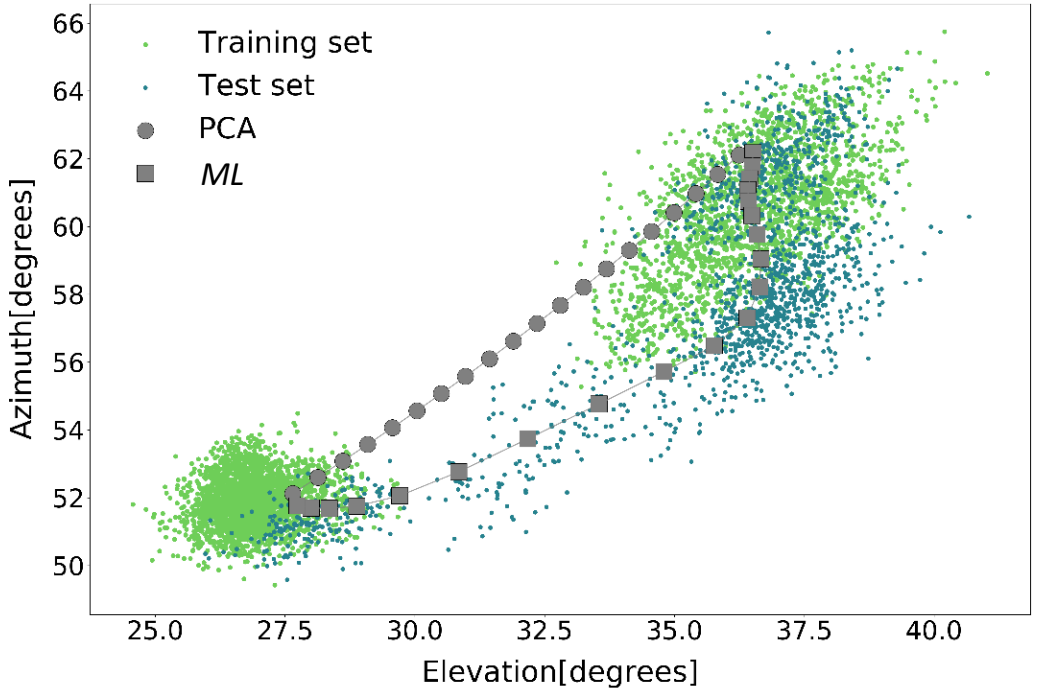
V.K. Ramaswamy *et al.*, ***PRX***, 2021

Predicting transition paths

MurD features known **closed** (ligand bound) and **open** (unbound) states





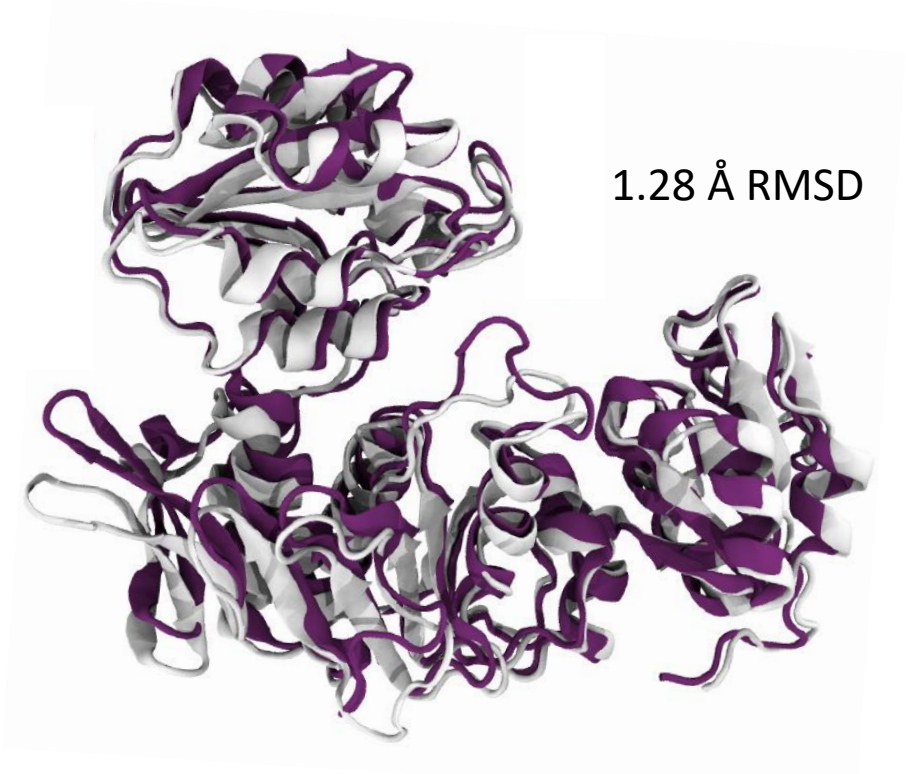


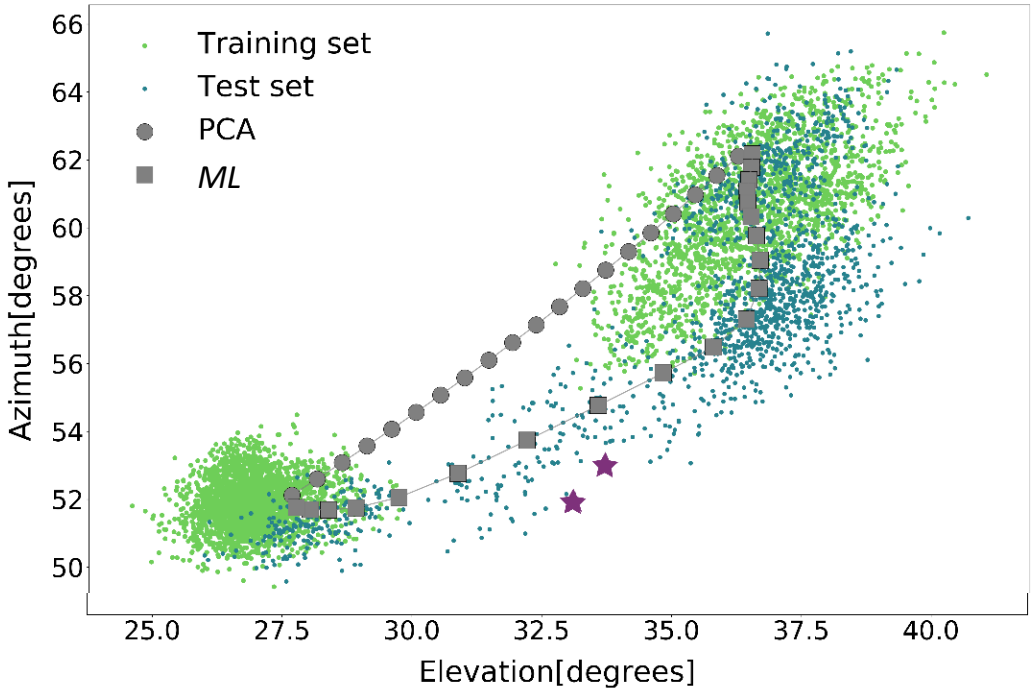
- Train *molearn* with MD of each state
- Predict transition between two states
- Prediction consistent with **closed apo** MD

V.K. Ramaswamy *et al.*, ***PRX***, 2021

Predicting transition paths

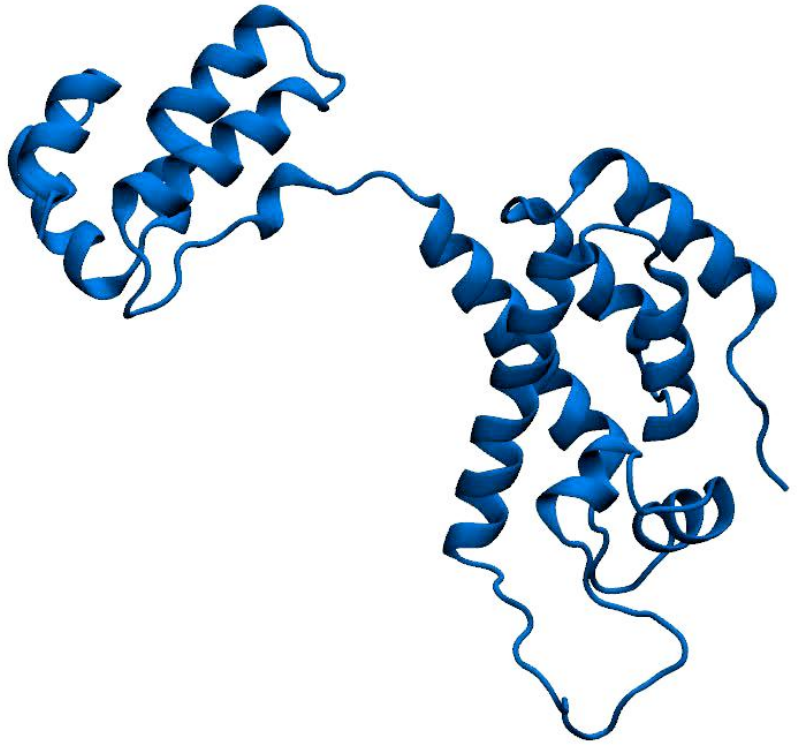
MurD features known **closed** (ligand bound) and **open** (unbound) states



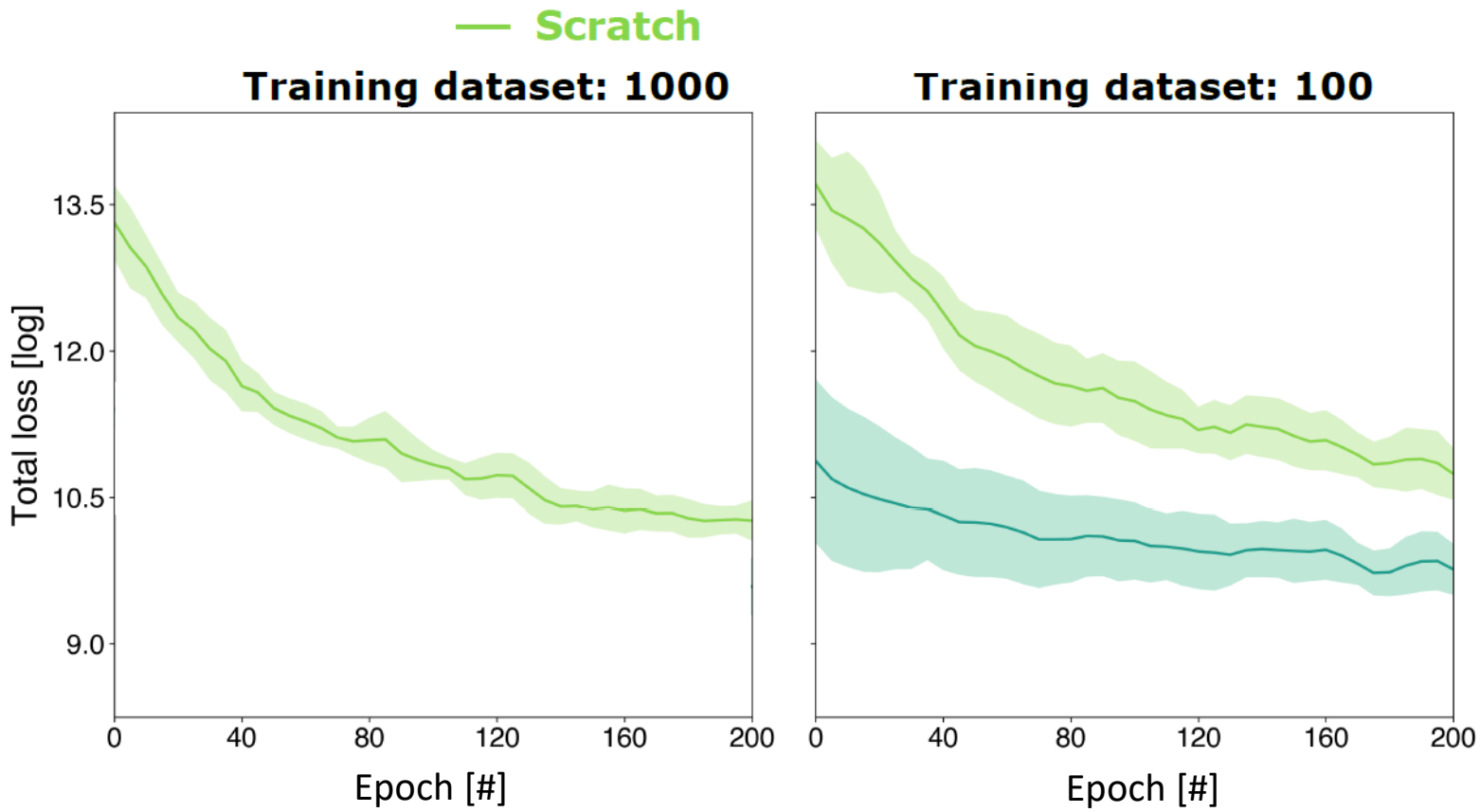


- Train *molearn* with MD of each state
- Predict transition between two states
- Prediction consistent with **closed apo** MD and intermediate **X-ray** structures

Towards a transferable network



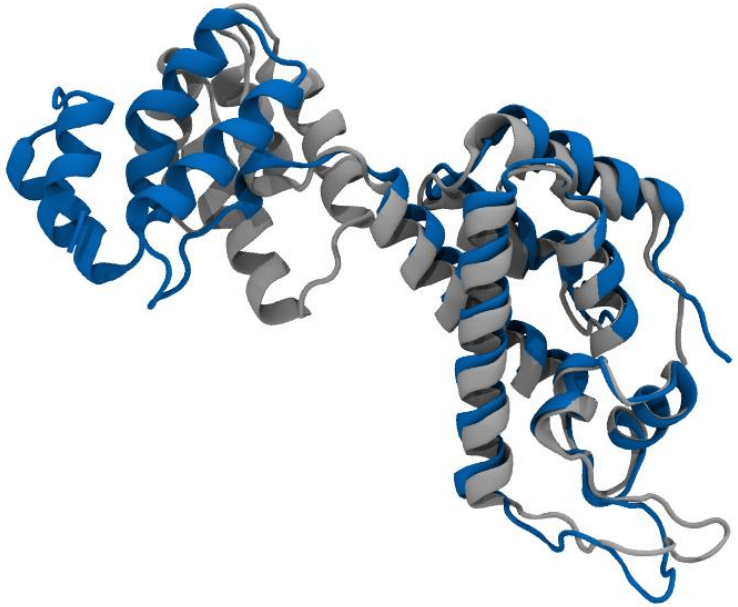
HIV-1 capsomer protein (p24)

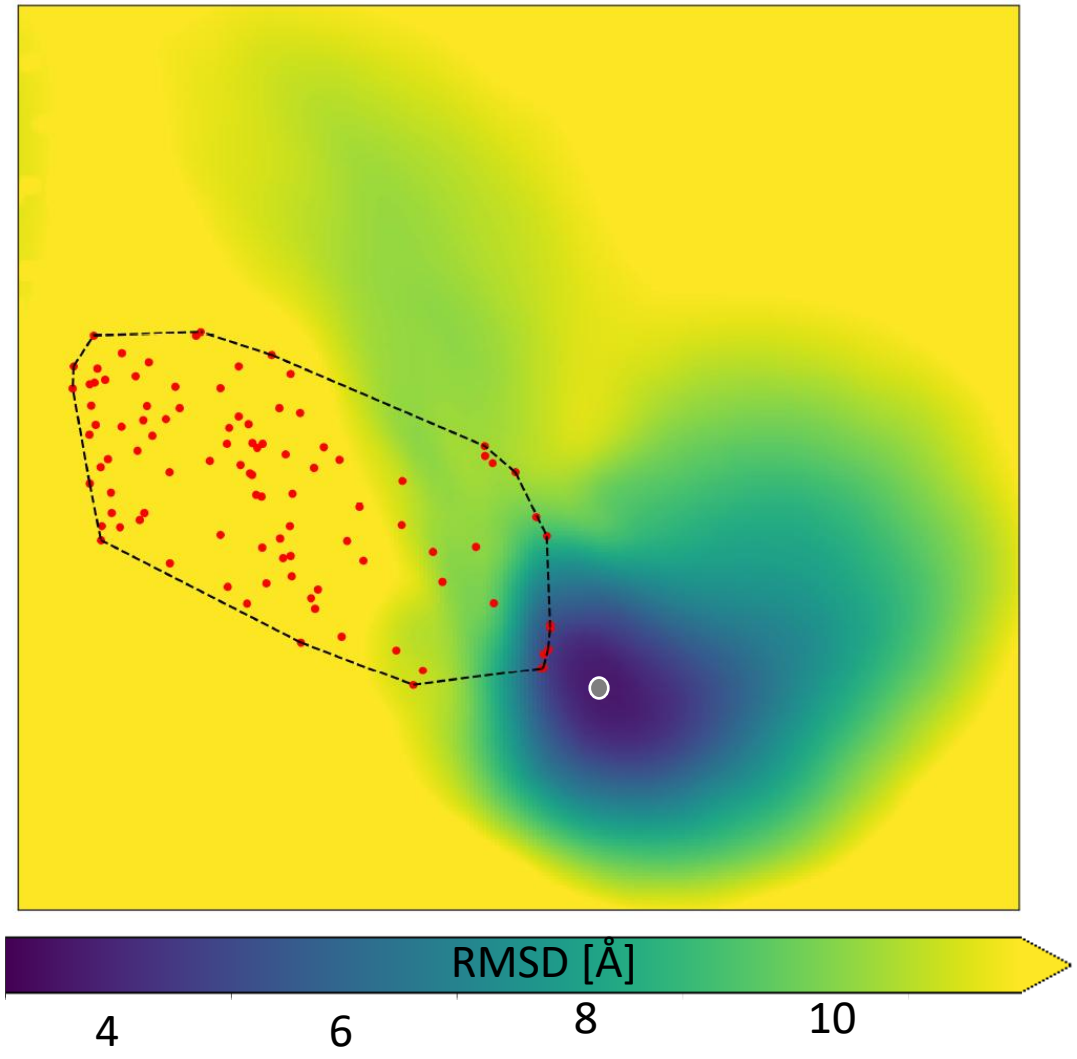


- A pre-trained network learns faster
- 100 protein structures are enough to train a network (4 test cases, 10 repeats)

V.K. Ramaswamy *et al.*, ***PRX***, 2021

HIV-1 capsomer bound state discovery





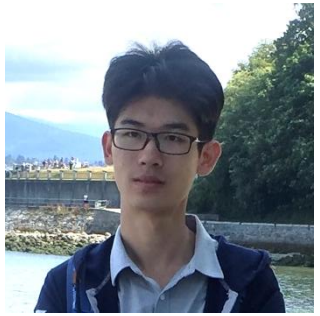
Best representative of bound state
in ML model (RMSD = 2.7 Å)
beyond the front of what sampled
by **MD** (lowest RMSD=3.2 Å)

V.K. Ramaswamy *et al.*, ***PRX***, 2021

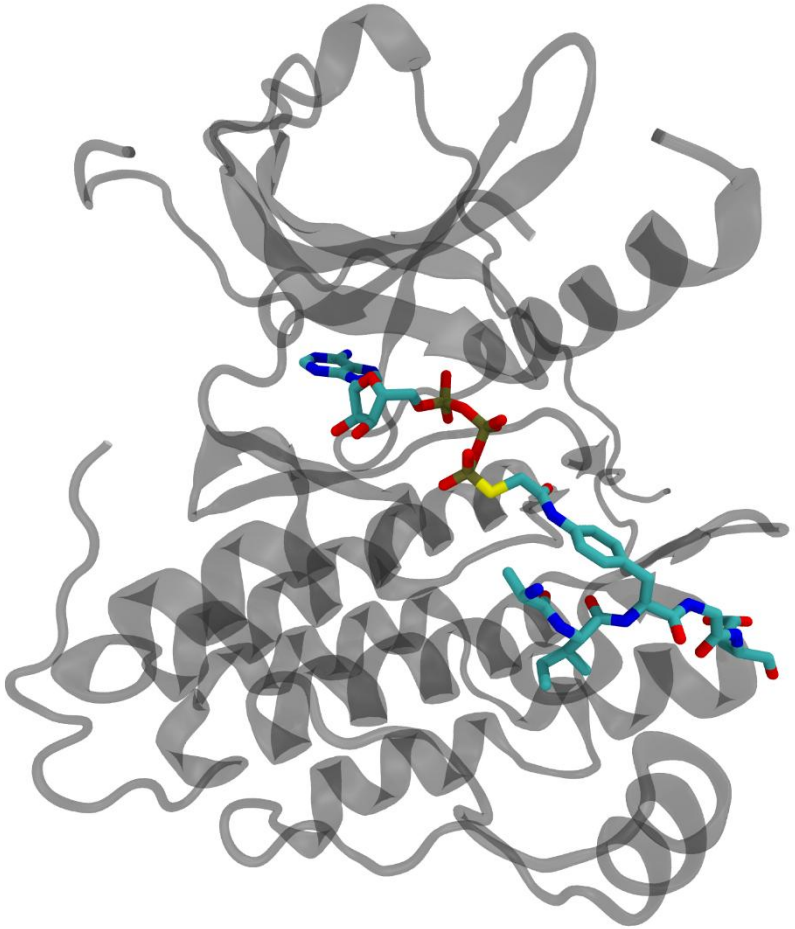
Modelling kinases dynamics

Sukrit Singh

Antonia Mey



Ryan Zhu

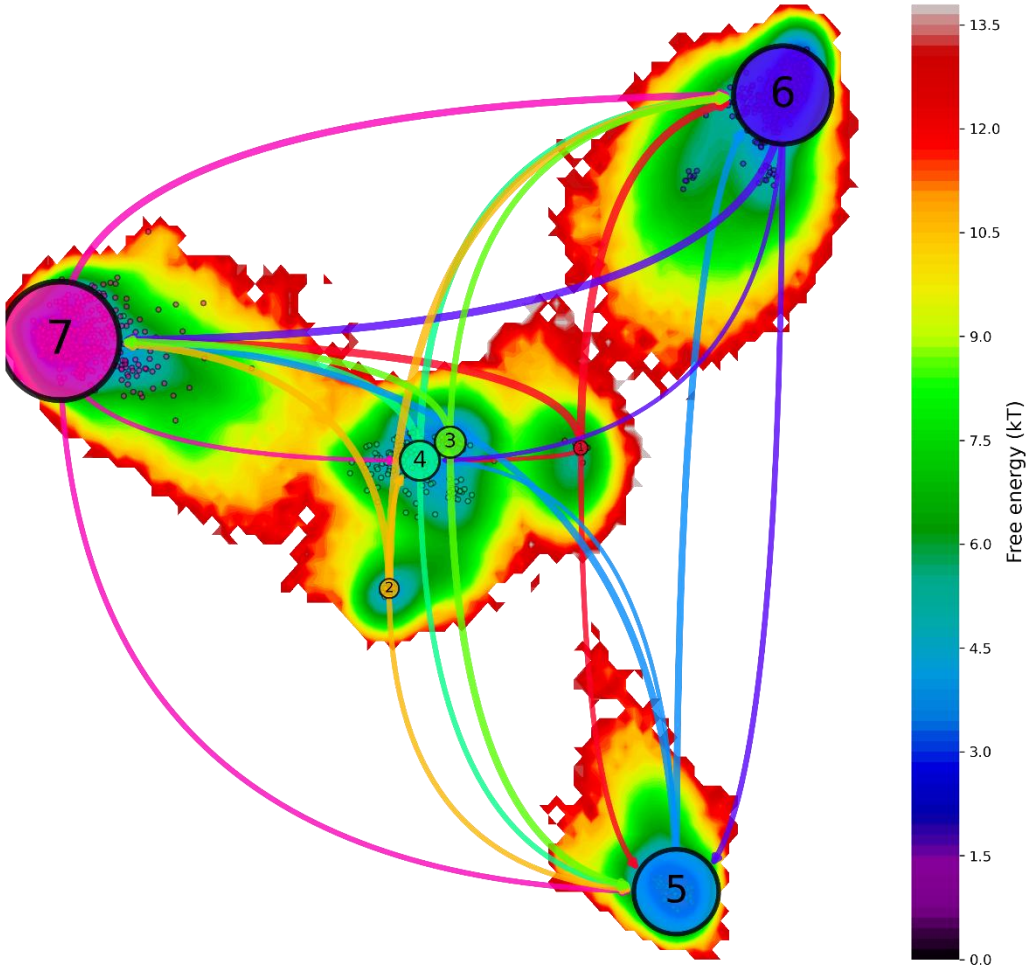


FOLDING @HOME

EGFR: 1.836 ms

Abl: 0.592 ms

c-Met: 0.419 ms



jupyter

MOLEARN

Dataset: training

cmap range: -43613.5 – -

Surf.: DOPE_unrefined

Path: A*

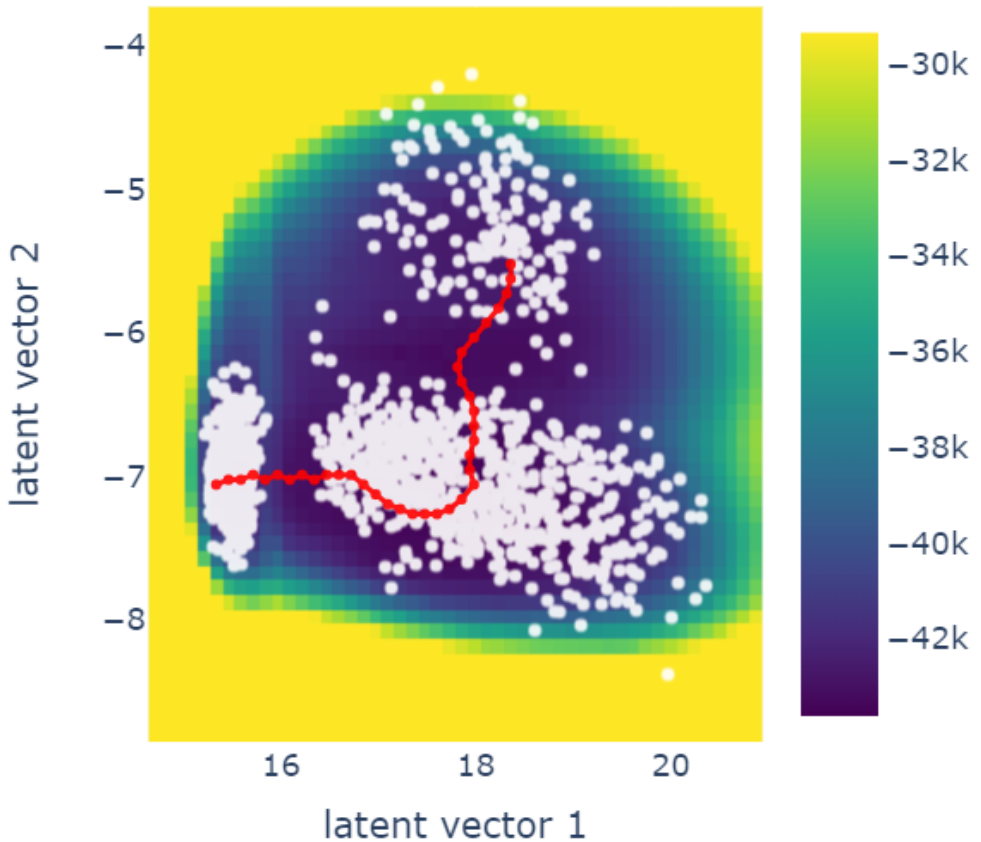
sampling: 10

crds: 17.9465 -7.0236
18.3564 -5.3936

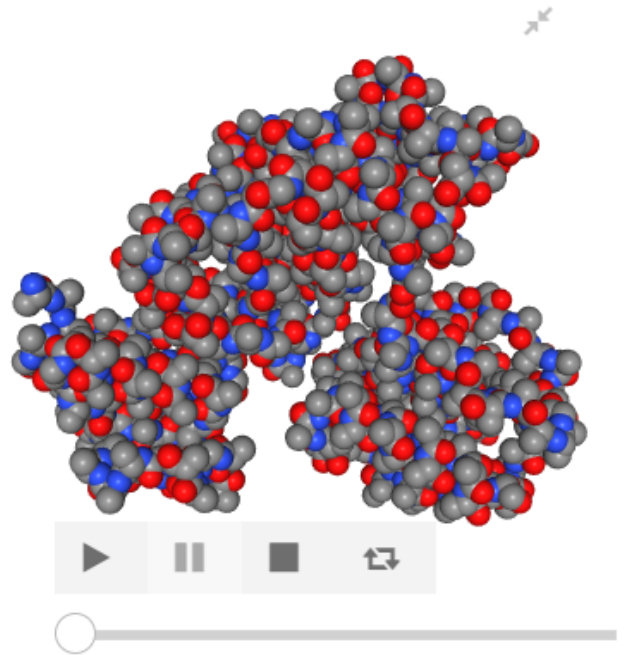
Save PDB

Save state

Load state



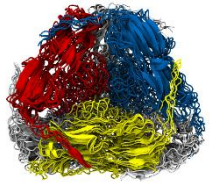
37 struct. in 4.4650 sec.



www.github.com/Degiacomi-Lab/molearn

S.C.Musson and M.T. Degiacomi, ***JOSS***, 2023

Thank You!



Dr Venkata Ramaswamy
Dr Samuel Musson
Ryan Zhu
Josh McKeown
Alexandros Angeli
Asal Azar
Lucy Vost
Mateusz Wiszniovski

Dr Antonia Mey

Dr Chris Willcocks
Dr Chris Prior

Gregor Wirnsberger

Dr Sukrit Singh

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

UKRI Engineering and Physical Sciences Research Council

SOFI² CDT
EPSRC-funded Centre for Doctoral Training in Soft Matter for Formulation and Industrial Innovation

MoSMed
Molecular Sciences for Medicine