

USING COMBINED SEQUENCE AND STRUCTURAL FEATURES TO PREDICT PROTEIN-LIGAND BINDING SITES

Matyáš Brabec

PROTEINS

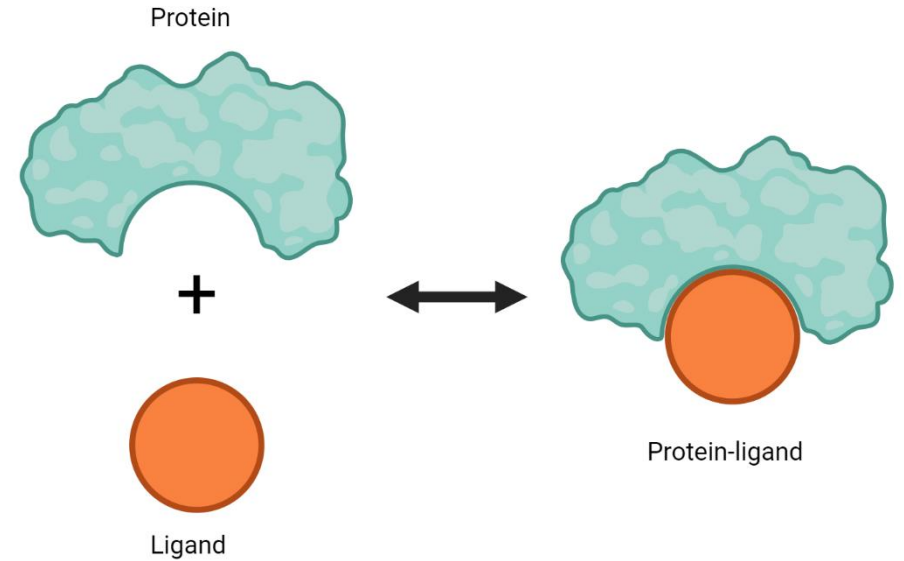
...and Ligands

Basic building blocks of organisms

- Immune system
- Transportation
- Cell replication
- ... many more

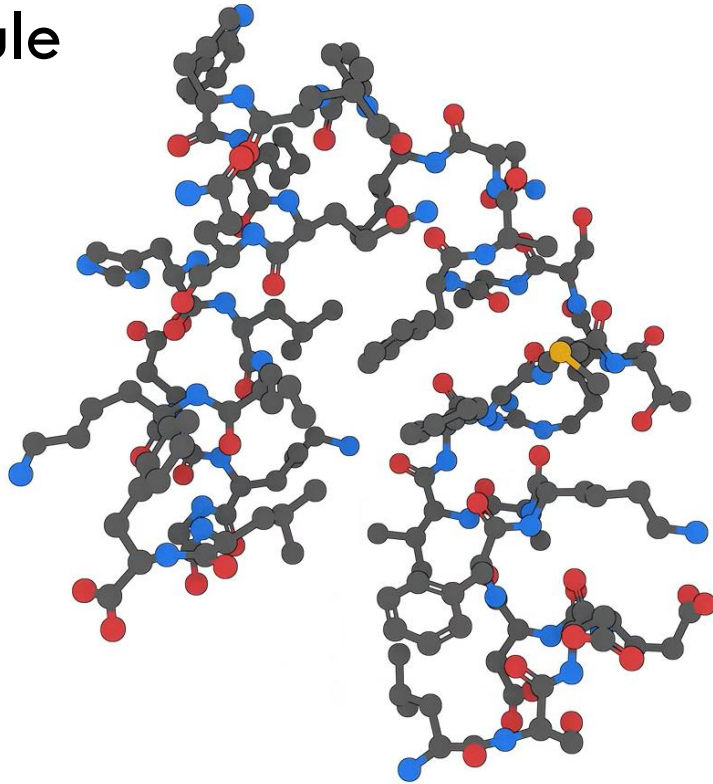
Ligands

- Binds to proteins

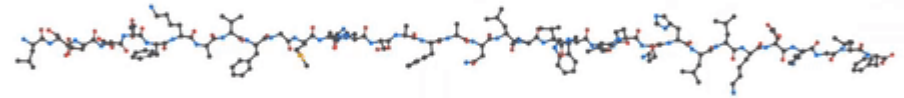


WHAT IS A PROTEIN ?

...Molecule

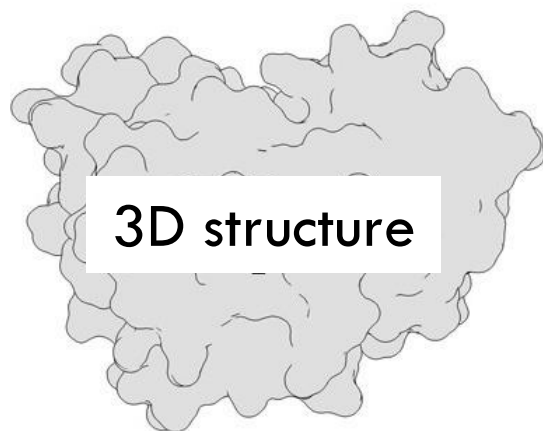
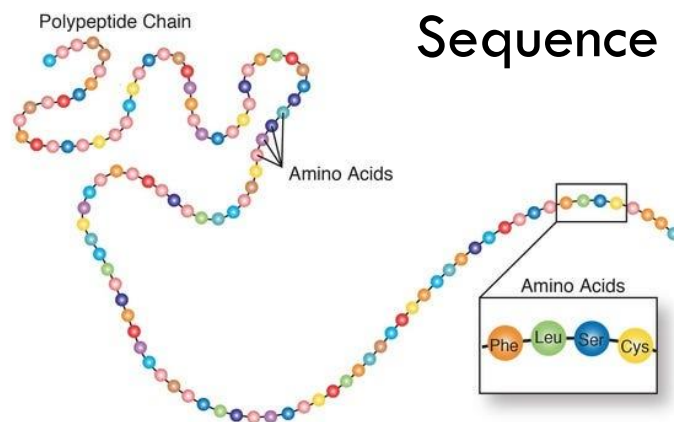


...Composed of amino acids

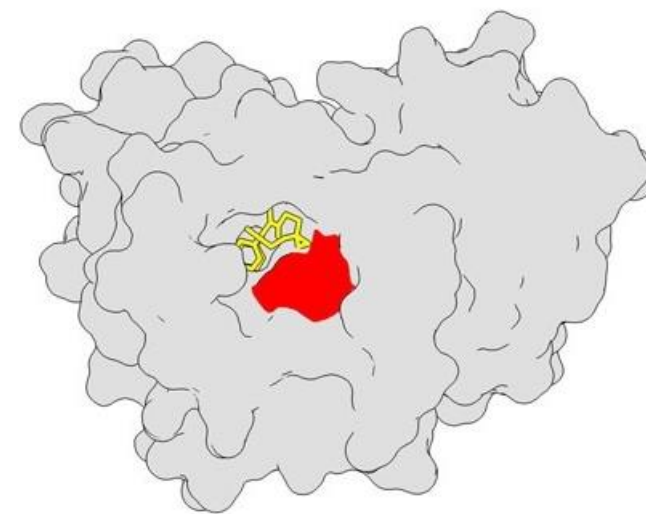


makeagif.com

THESIS TOPIC

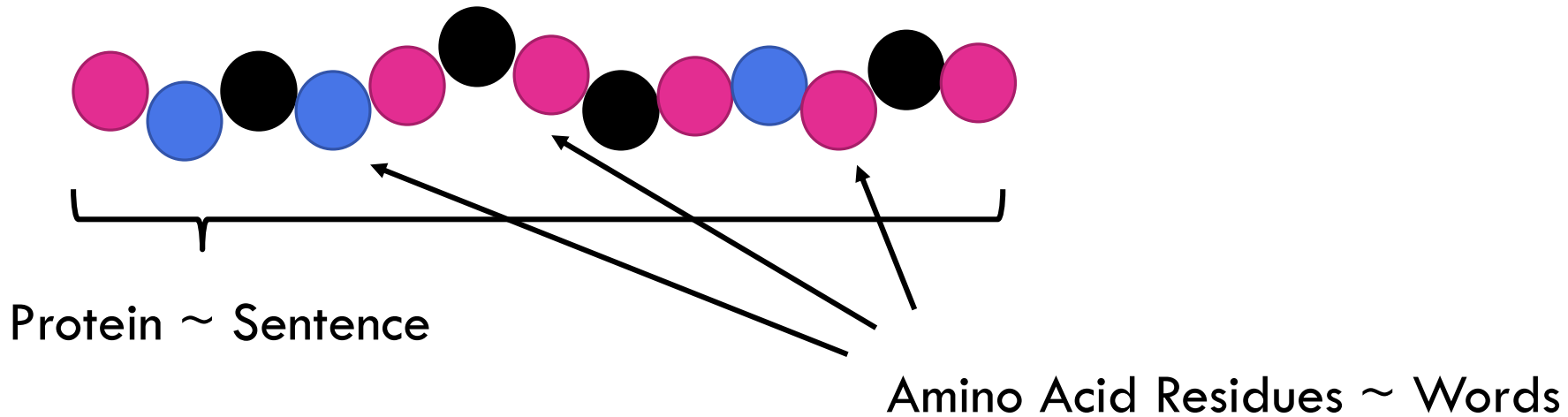


Model



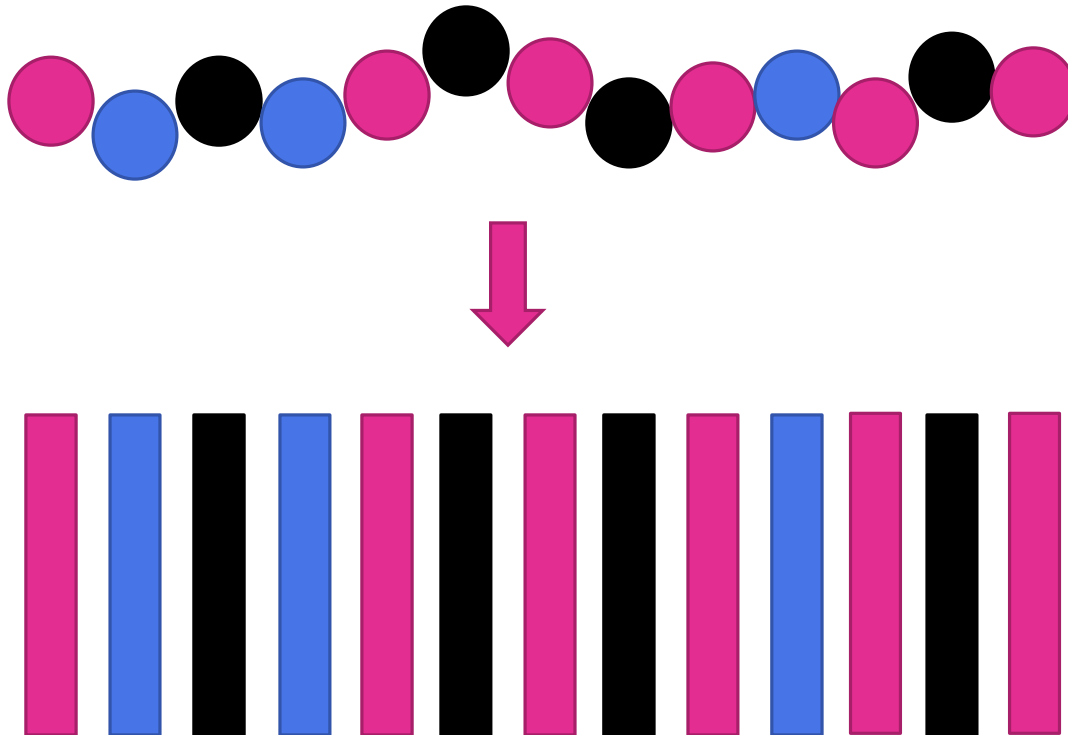
WHAT TO DO WITH SEQUENCE ?

Protein language models



WHAT TO DO WITH SEQUENCE ?

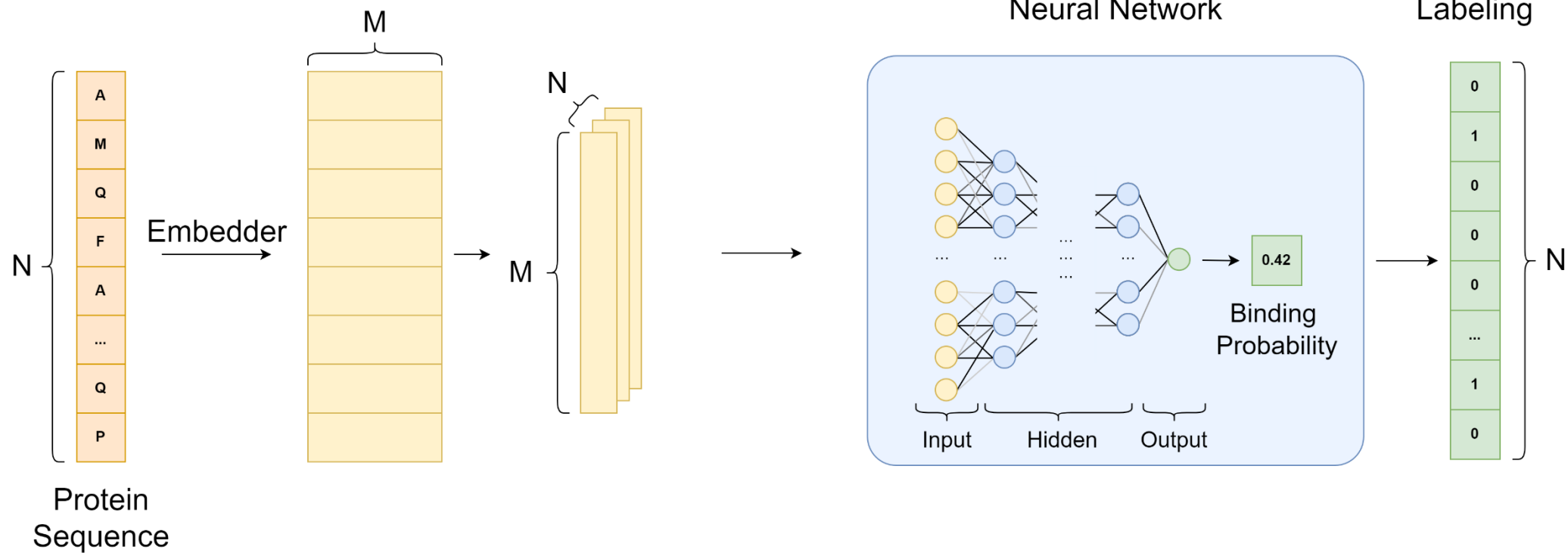
Protein language models



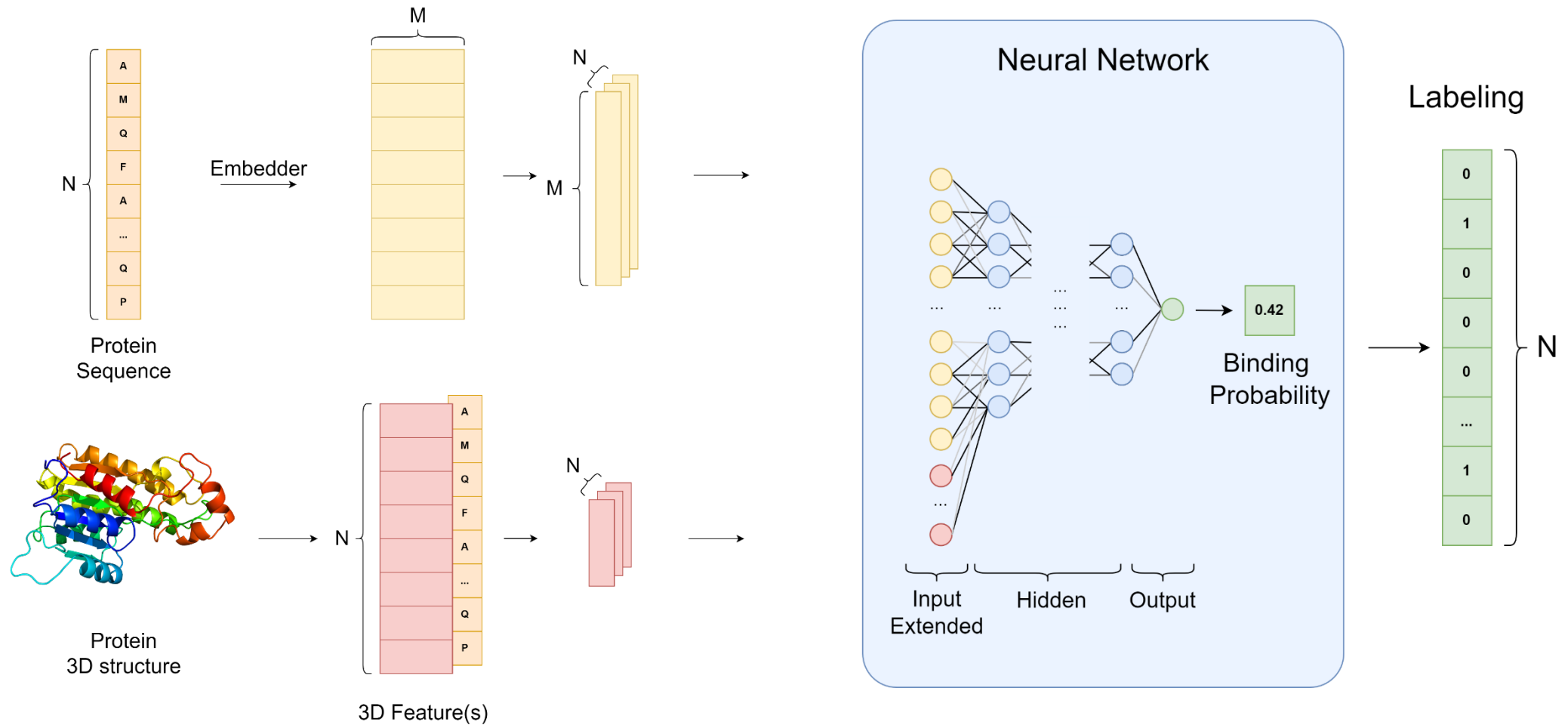
Sequence of amino
acid residues

Sequence of vectors
(embeddings)

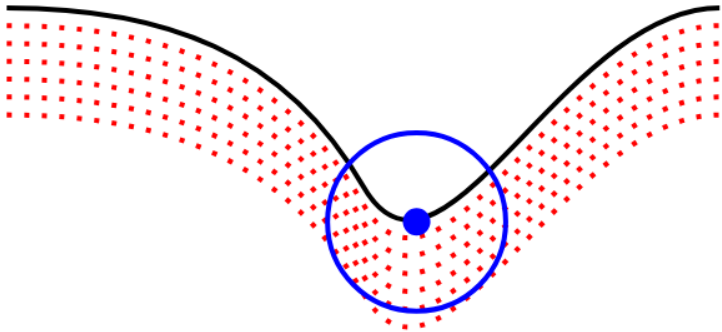
BASELINE MODEL



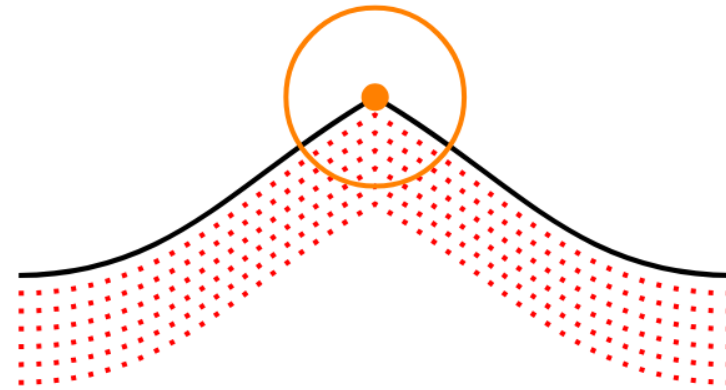
3D FEATURES MODELS



PROTRUSION



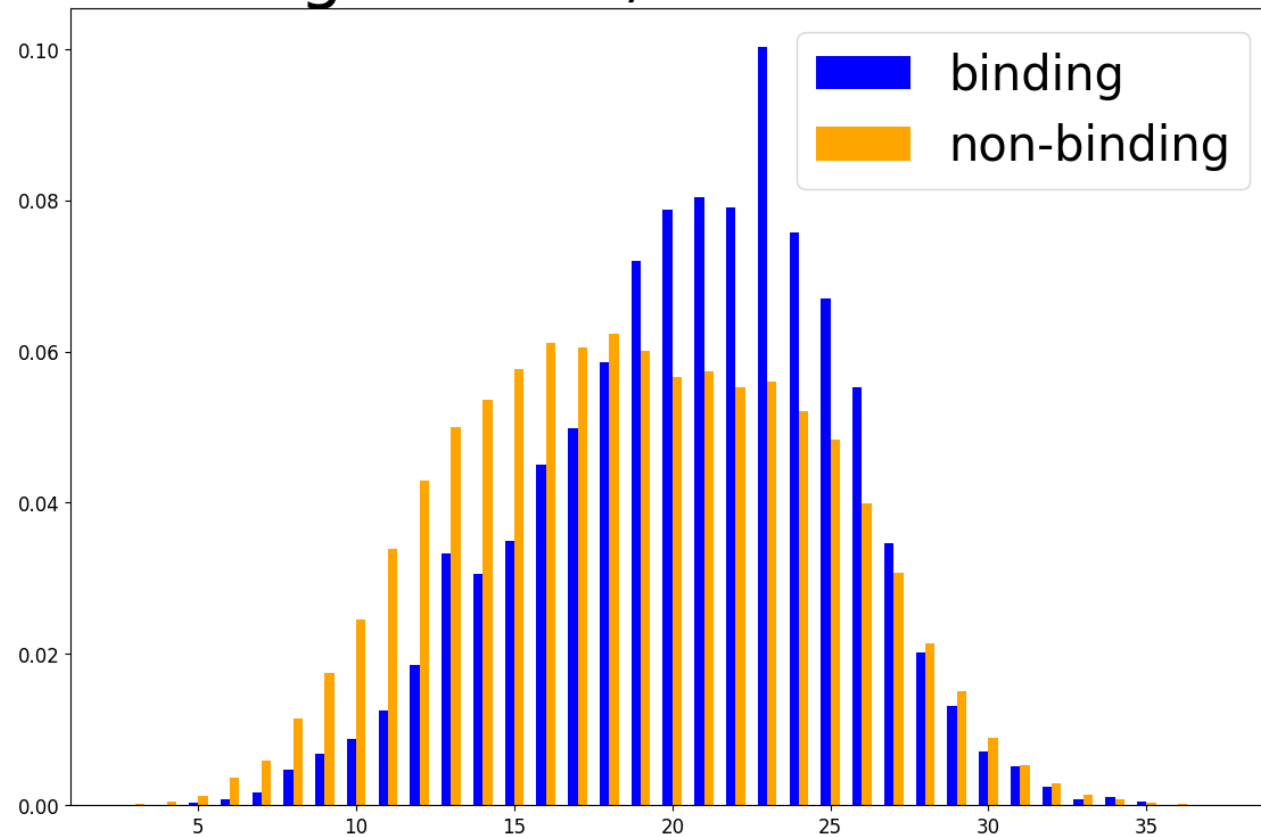
(a) Ligand binding site



(b) Non-binding site

PROTRUSION

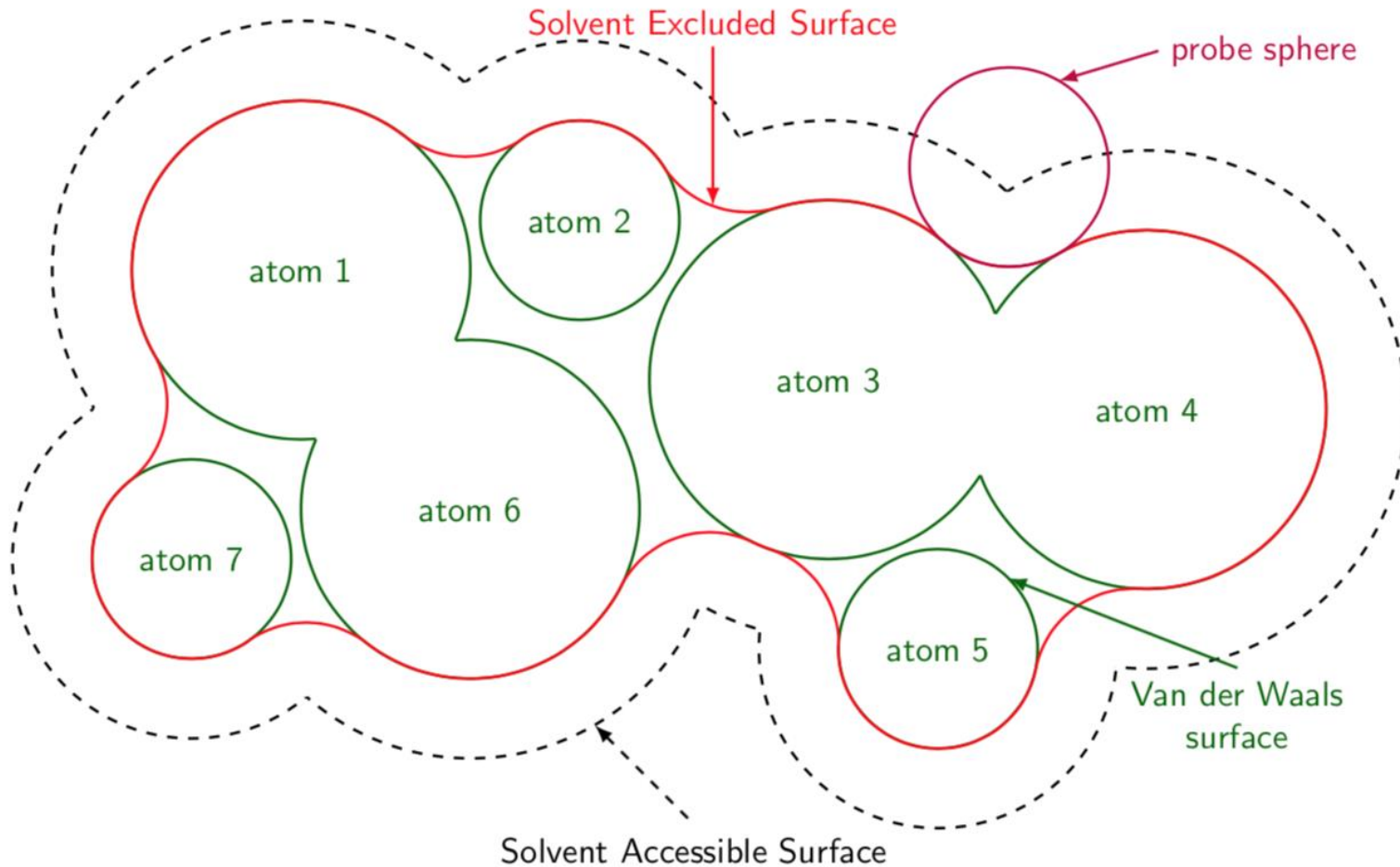
Ligand ATP, Radius 8.5 Å



PROTRUSION

Ligand	Baseline		One Protrusion - 1. layer		
	MCC	Average MCC	MCC	Average MCC	p-value
ADP	0.548	0.636 (± 0.010)	0.580	0.631 (± 0.013)	0.312
AMP	0.492	0.462 (± 0.005)	0.482	0.502 (± 0.008)	0.000
ATP	0.557	0.570 (± 0.007)	0.563	0.585 (± 0.007)	0.000
CA	0.392	0.463 (± 0.009)	0.399	0.470 (± 0.012)	0.129
DNA	0.432	0.529 (± 0.010)	0.455	0.545 (± 0.011)	0.005
FE	0.606	0.691 (± 0.011)	0.648	0.712 (± 0.006)	0.000
GDP	0.752	0.747 (± 0.013)	0.744	0.747 (± 0.009)	0.957
GTP	0.656	0.617 (± 0.025)	0.604	0.622 (± 0.030)	0.688
HEME	0.705	0.652 (± 0.013)	0.722	0.671 (± 0.006)	0.000
MG	0.341	0.404 (± 0.016)	0.281	0.405 (± 0.020)	0.890
MN	0.601	0.588 (± 0.018)	0.594	0.604 (± 0.010)	0.026
ZN	0.665	0.660 (± 0.014)	0.672	0.667 (± 0.004)	0.151

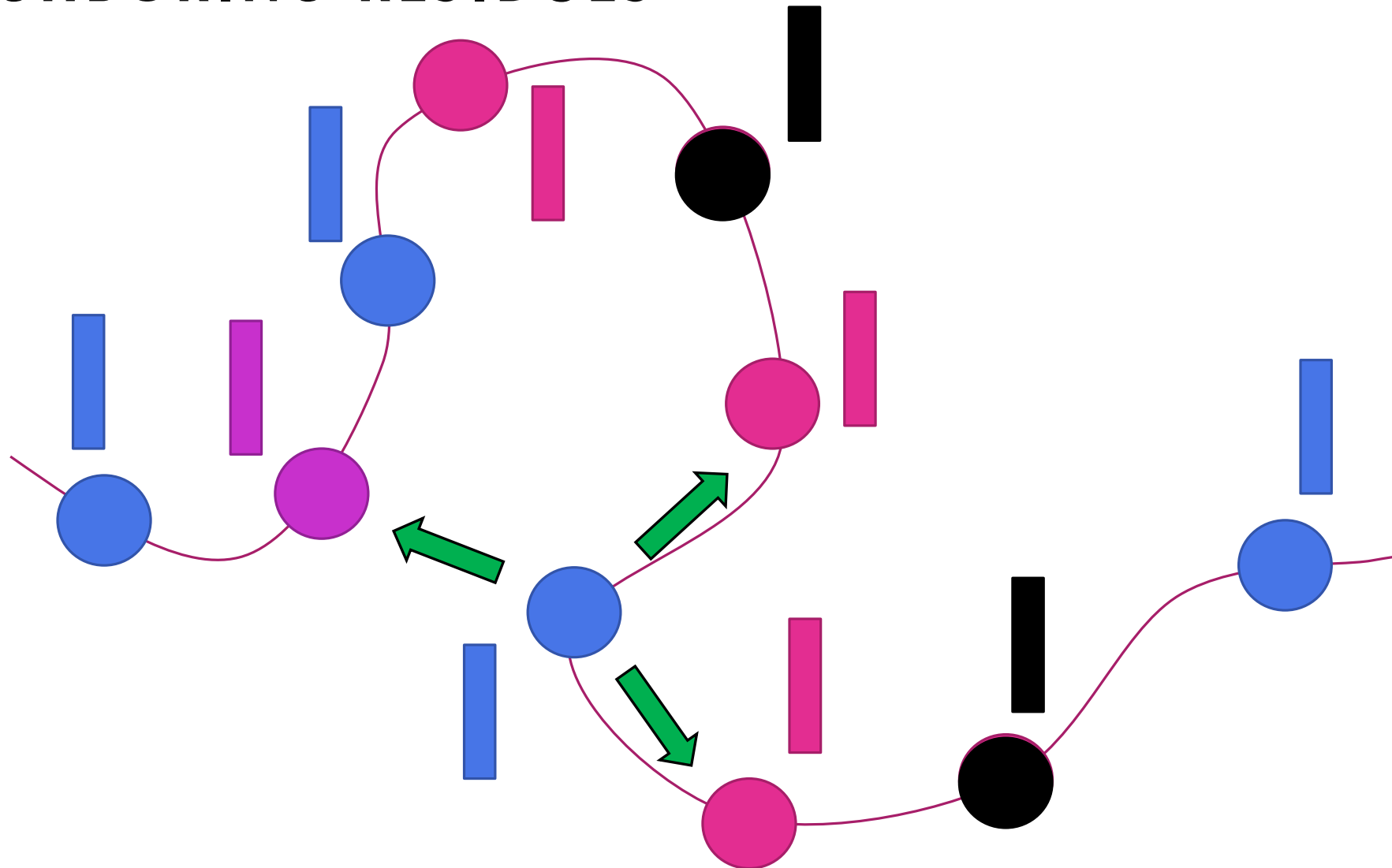
SOLVENT ACCESSIBLE SURFACE AREA (SASA)



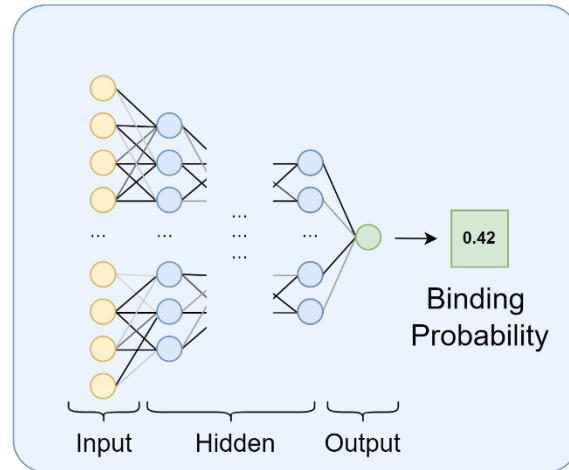
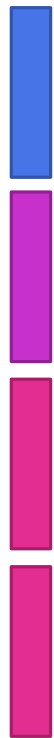
SOLVENT ACCESSIBLE SURFACE AREA (SASA)

Ligand	Baseline		SASA in First Layer		
	MCC	Average MCC	MCC	Average MCC	p-value
ADP	0.548	0.636 (± 0.010)	0.568	0.633 (± 0.011)	0.553
AMP	0.492	0.462 (± 0.005)	0.475	0.467 (± 0.011)	0.257
ATP	0.557	0.570 (± 0.007)	0.576	0.586 (± 0.007)	0.000
CA	0.392	0.463 (± 0.009)	0.400	0.461 (± 0.010)	0.669
DNA	0.432	0.529 (± 0.010)	0.467	0.552 (± 0.008)	0.000
FE	0.606	0.691 (± 0.011)	0.655	0.694 (± 0.007)	0.489
GDP	0.752	0.747 (± 0.013)	0.753	0.735 (± 0.014)	0.061
GTP	0.656	0.617 (± 0.025)	0.569	0.594 (± 0.022)	0.046
HEME	0.705	0.652 (± 0.013)	0.704	0.651 (± 0.009)	0.788
MG	0.341	0.404 (± 0.016)	0.356	0.408 (± 0.007)	0.478
MN	0.601	0.588 (± 0.018)	0.596	0.604 (± 0.010)	0.030
ZN	0.665	0.660 (± 0.014)	0.674	0.664 (± 0.008)	0.412

NEIGHBORING RESIDUES

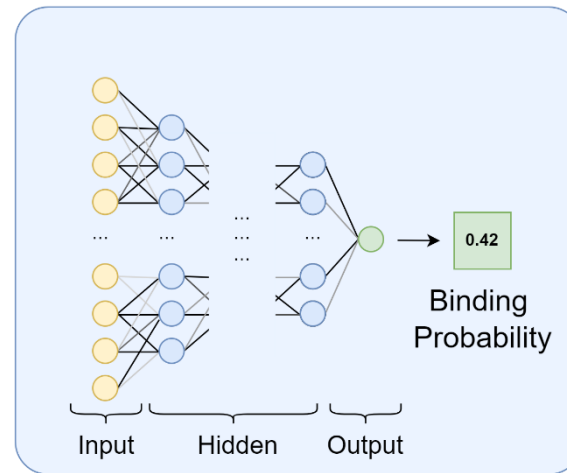
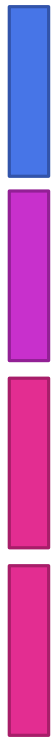


NEIGHBORING RESIDUES



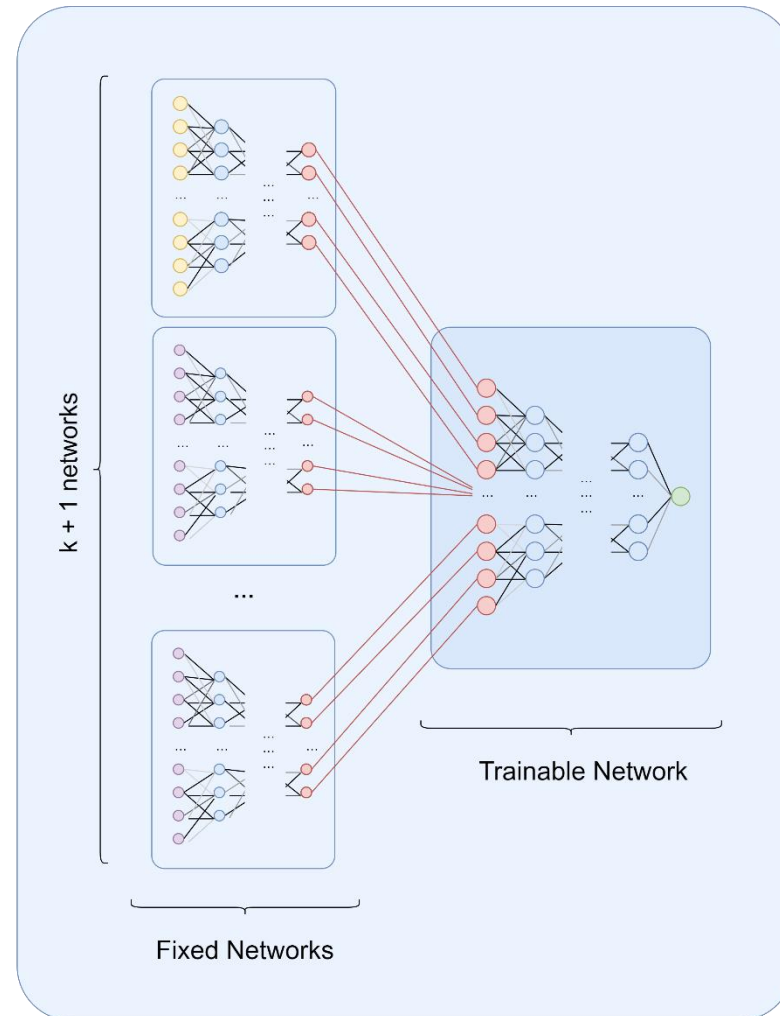
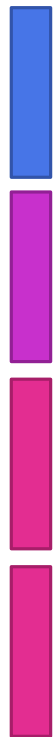
Binding / Not binding

NEIGHBORING RESIDUES



Binding / Not binding

NEIGHBORING RESIDUES



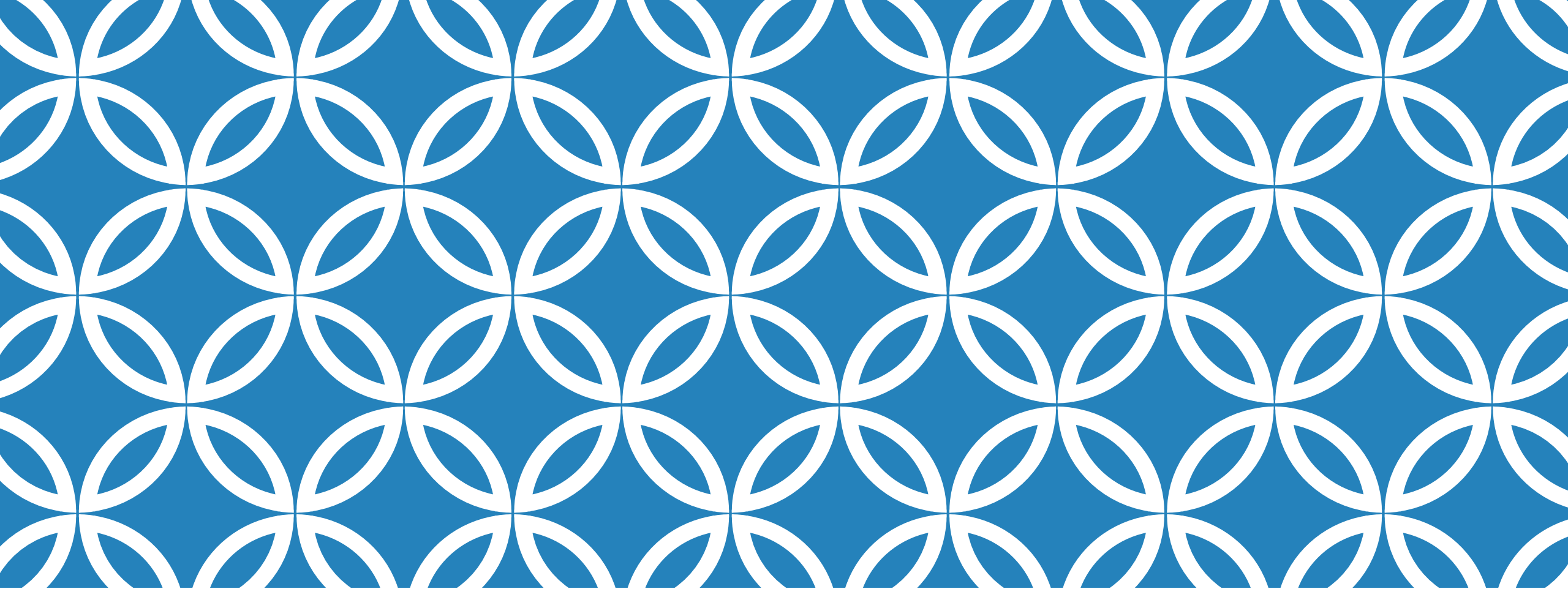
Binding / Not binding

P2RANK COMPARISON

Ligand	ESM-2	P2Rank
	<i>MCC</i>	<i>MCC</i>
ADP	0.573	0.454
AMP	0.440	0.424
ATP	0.569	0.484
CA	0.388	0.235
DNA	0.457	0.261
FE	0.635	0.354
GDP	0.724	0.531
GTP	0.645	0.591
HEME	0.708	0.503
MG	0.339	0.272
MN	0.585	0.468
ZN	0.672	0.348



CONCLUSION



NOTES

Q&A

