Adaptyv
Biosystems

Lemanic Life Sciences Hackathon 2024
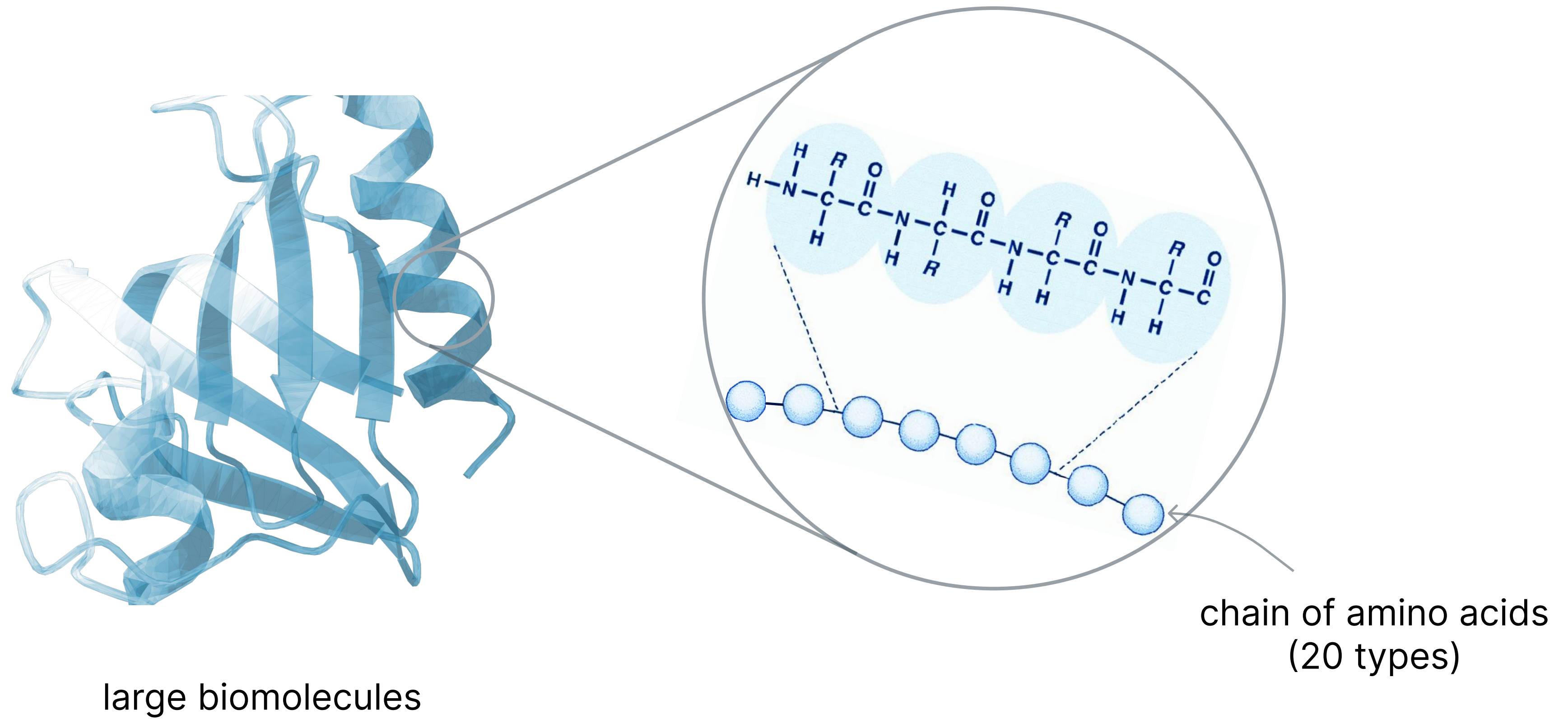
Predicting PPI on imbalanced data

2024-04

Proteins
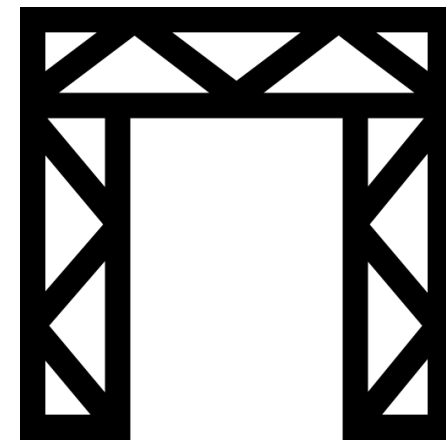## Definition
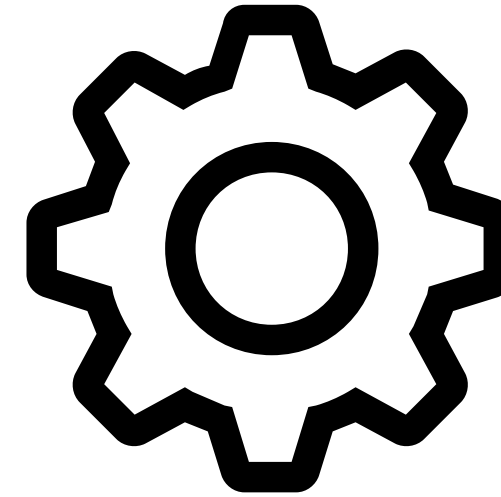

large biomolecules

chain of amino acids
(20 types)

Proteins
Function

structure
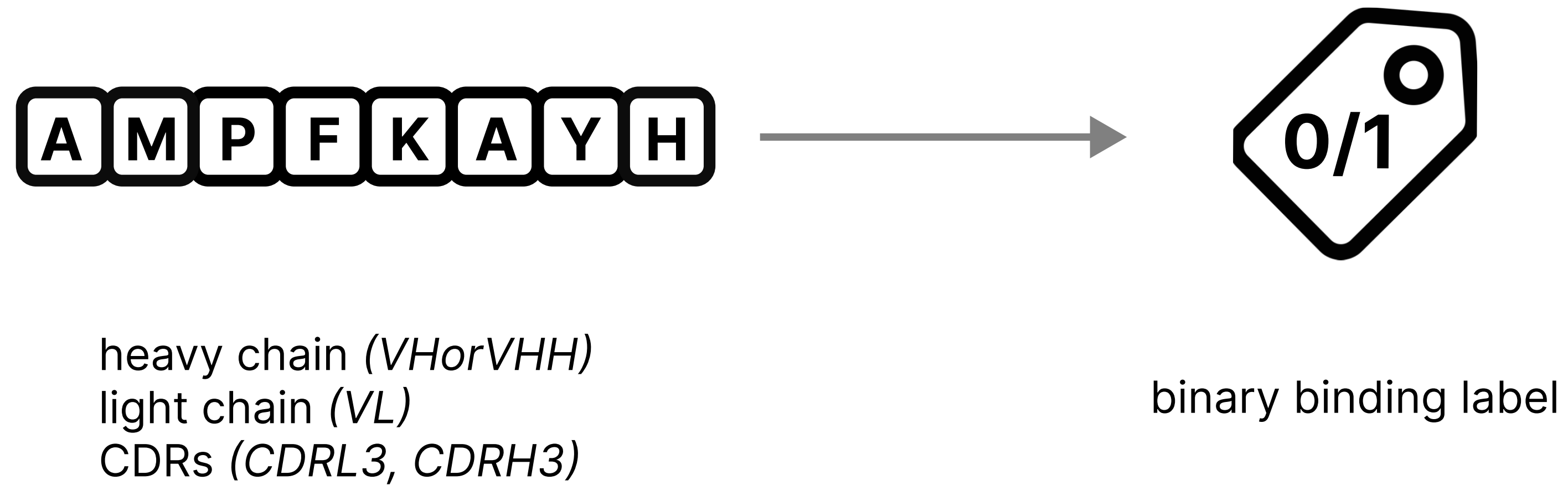
collagen

function

antibodies
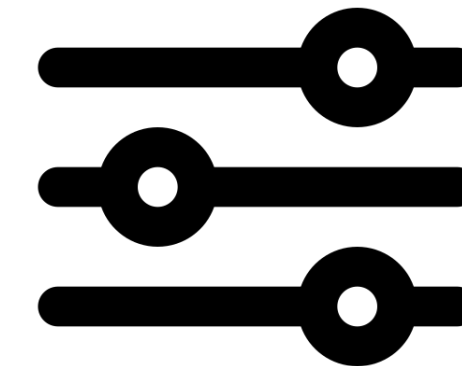
regulation

adrenalin

Adaptyv Bio

Proteins
Antibodies

CDR (complementarity determining regions)

antigen

(e.g. virus surface)

antibody / nanobody

Adaptyv Bio
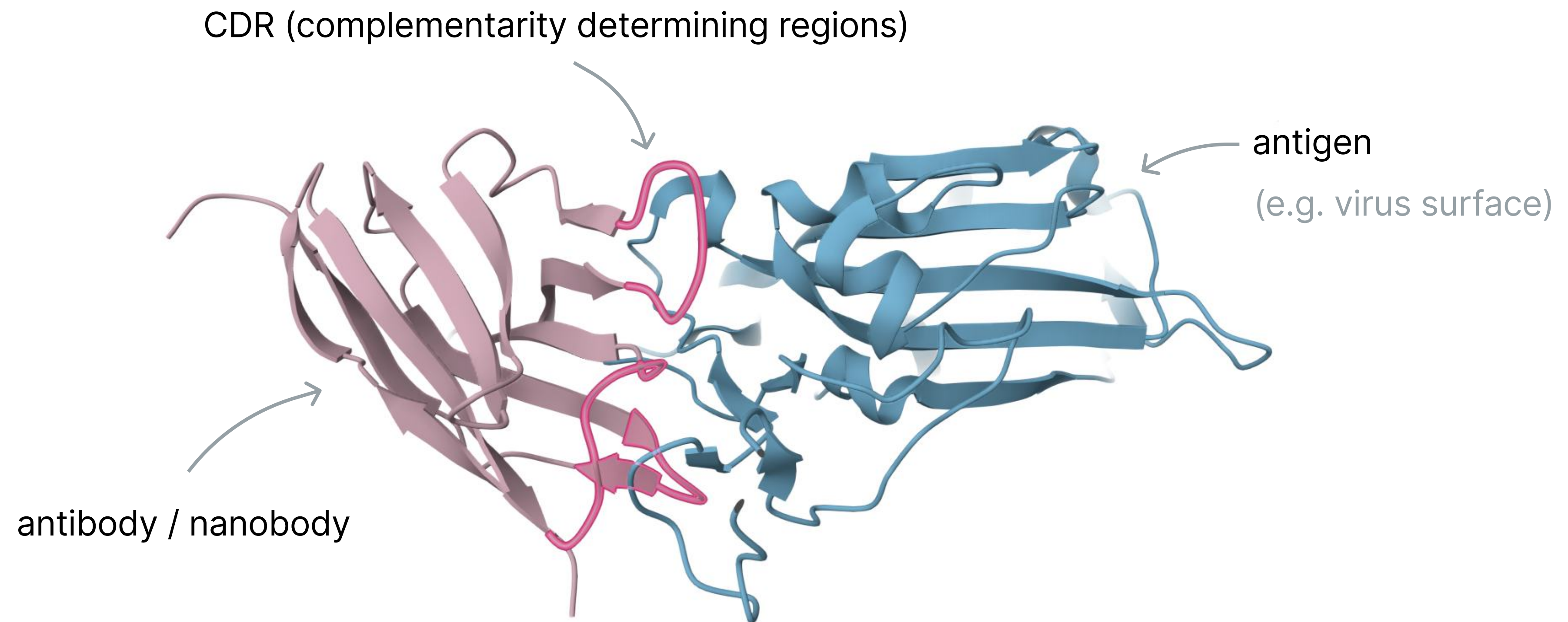
Task

Binary interaction prediction

heavy chain *(VHorVHH)*
light chain *(VL)*
CDRs *(CDRL3, CDRH3)*
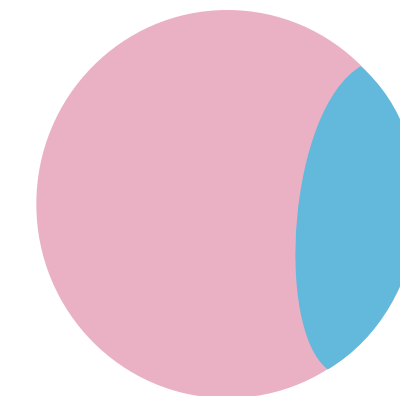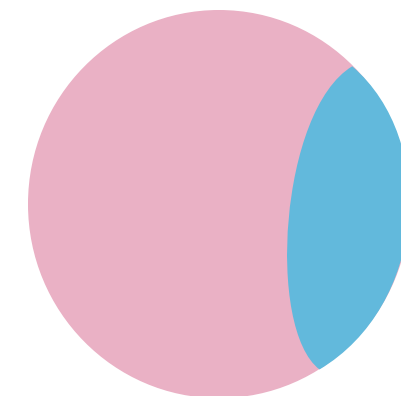
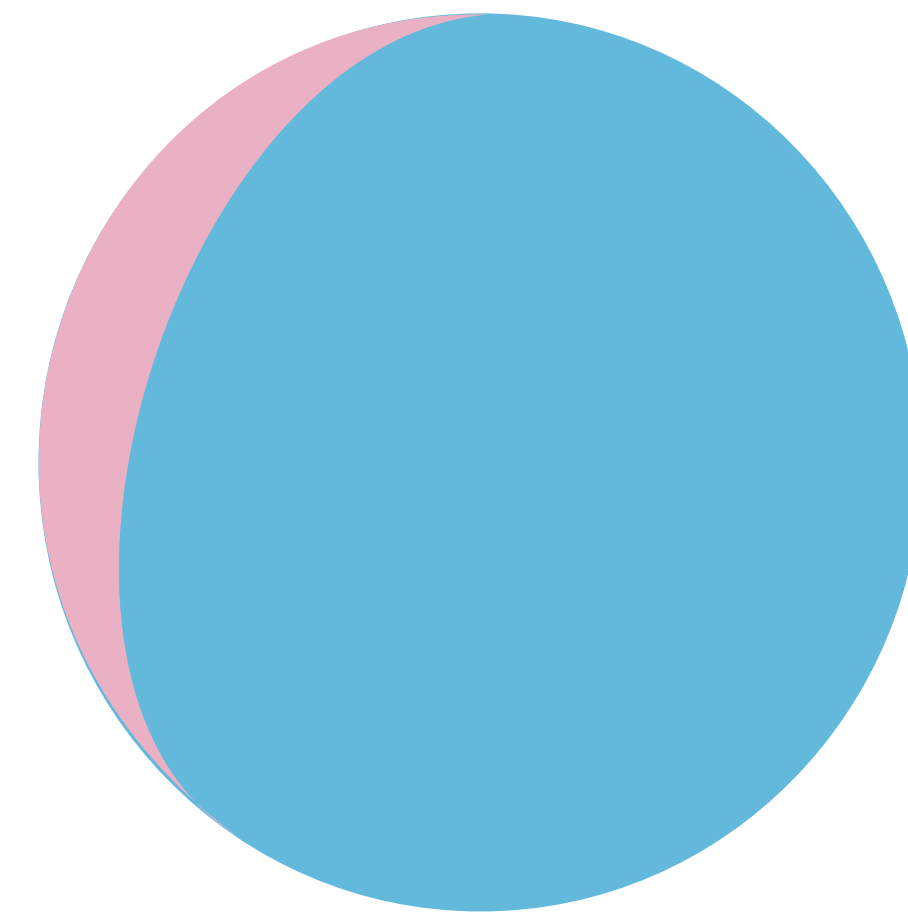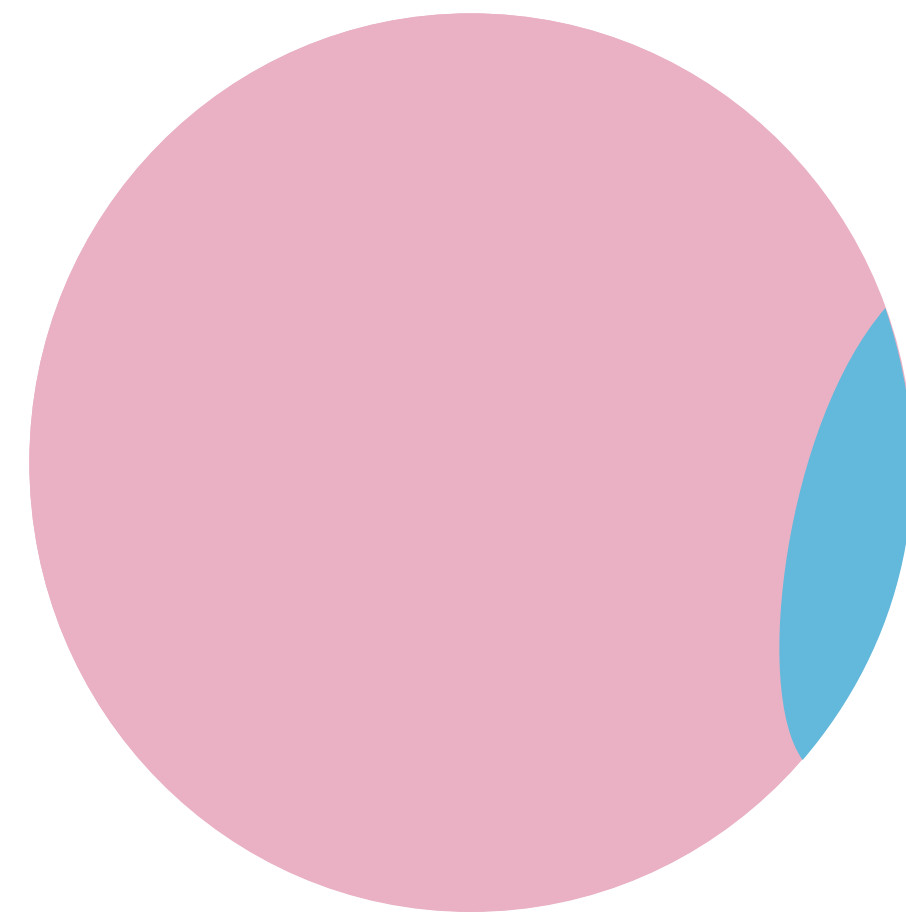binary binding label

Adaptyv Bio

Task

Two datasets

experiment (HCoV-HKU1)   literature (SARS-CoV1)

more negative data   more positive data

Adaptyv Bio

Task
Some directions

- language model embeddings: <u>ESM-2</u>, <u>ProteinBERT</u>, <u>ProtTrans</u>, ...
- <u>adding negative samples</u> to the literature dataset
- <u>methods for handling class imbalance</u>

always <u>make sure</u> that the results are not just <u>obtained by chance</u>
(easy check: shuffle the predictions and see if the score drops)

Adaptyv Bio