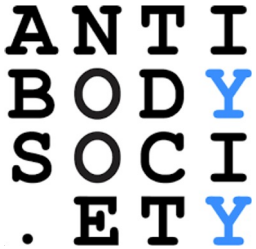


How structure prediction can enhance antibody repertoire sequence analysis

Prof. Charlotte Deane, Dr. Fergus Boyles, Dr. Matthew Raybould

Oxford Protein Informatics Group, University of Oxford

9th June, 2022

Selective pressure acts on the phenotype

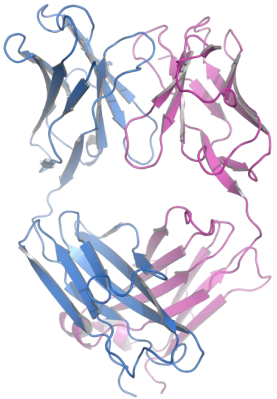
Less conserved
Information poor

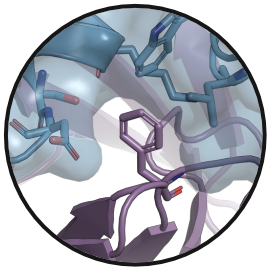
More conserved
Information rich

DNA seq → Protein Seq → Structure → Function

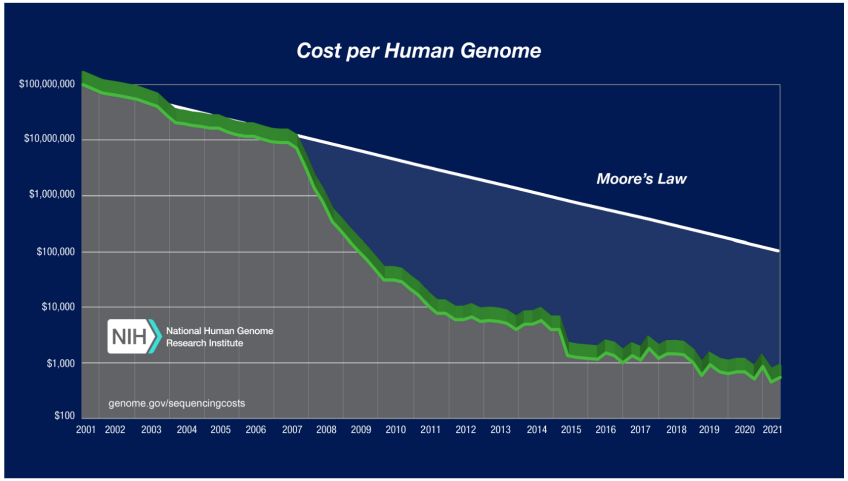
ACAGTTACAC
CGGCTATGTA
CTATACTTT……

EVTLRESGPA
LVKPTQTLTLT
CTFSGFSLS……



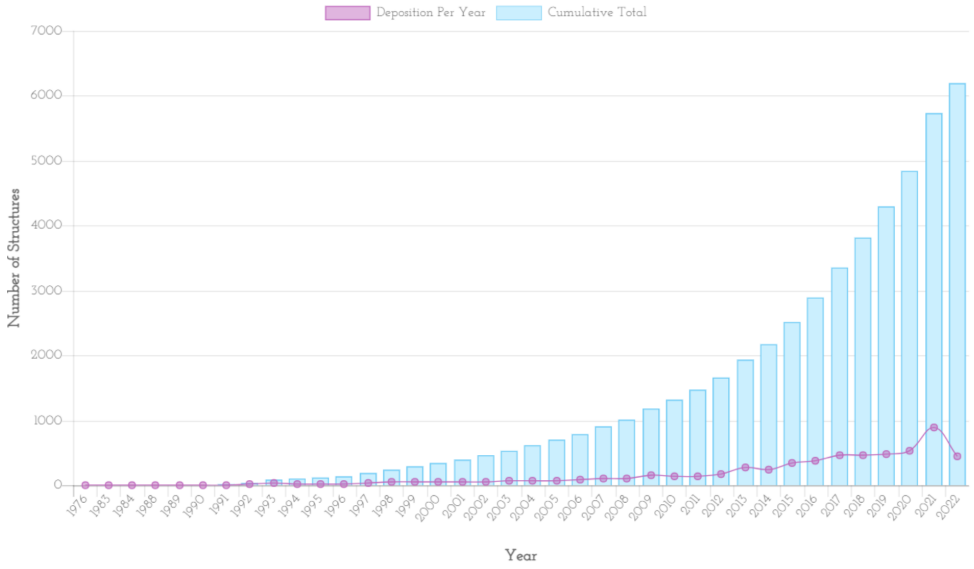


Experimental protein structure determination vs sequence acquisition

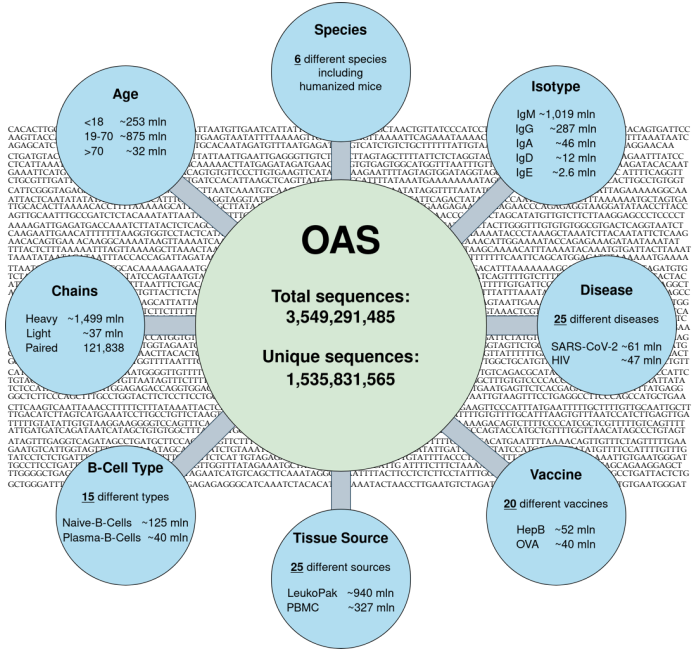


Structure vs sequence space for antibodies

Growth of Antibody Structures in SAbDab



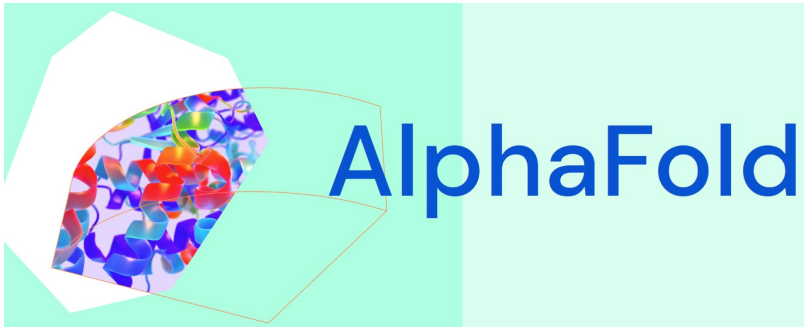
Sequences in OAS



SAbDab
The Structural Antibody Database

OAS
Observed Antibody Space

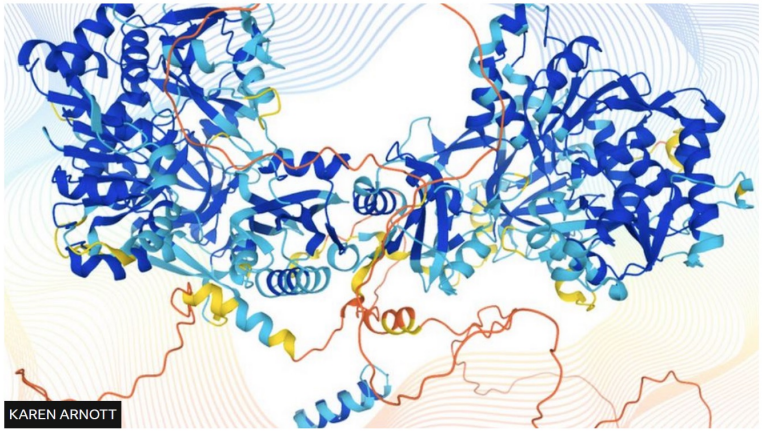
Structure prediction to close the gap



AI breakthrough could spark medical revolution

By Paul Rincon
Science editor, BBC News website

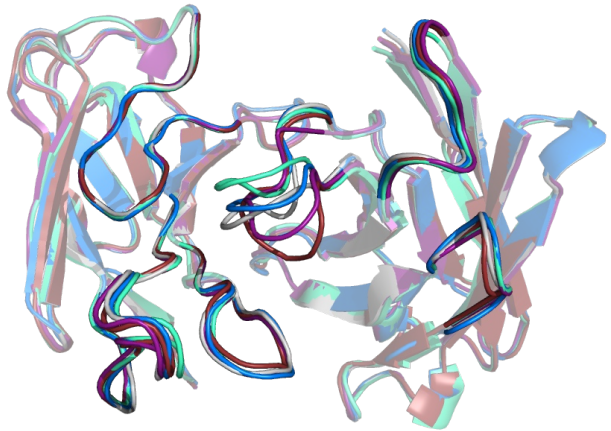
22 July 2021 | Comments



Only a fraction of proteins made by the human genome have confirmed structures

Structure prediction for antibodies

	Rosetta Antibody Benchmark
AlphaFold2	2.87*
ABodyBuilder	2.77
DeepAb	2.44
ABlooper	2.49
ABlooper (unrelaxed)	2.45

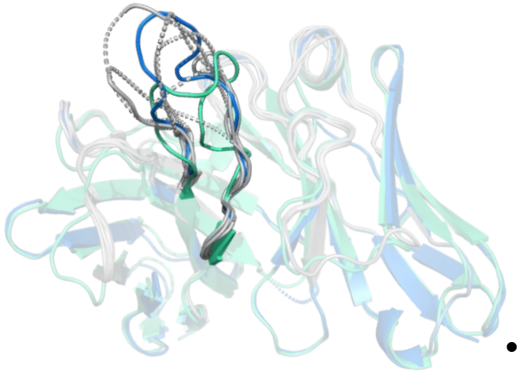


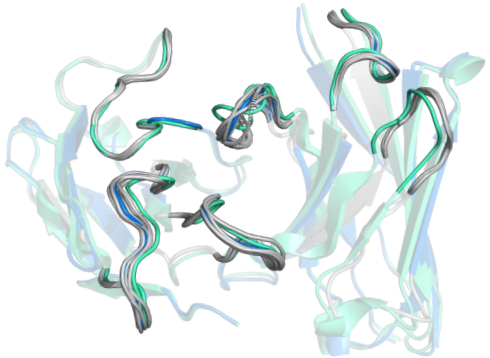
RMSD across backbone atoms of the CDR-H3 to the correct structure

*potentially these structures were contained in the training of this method

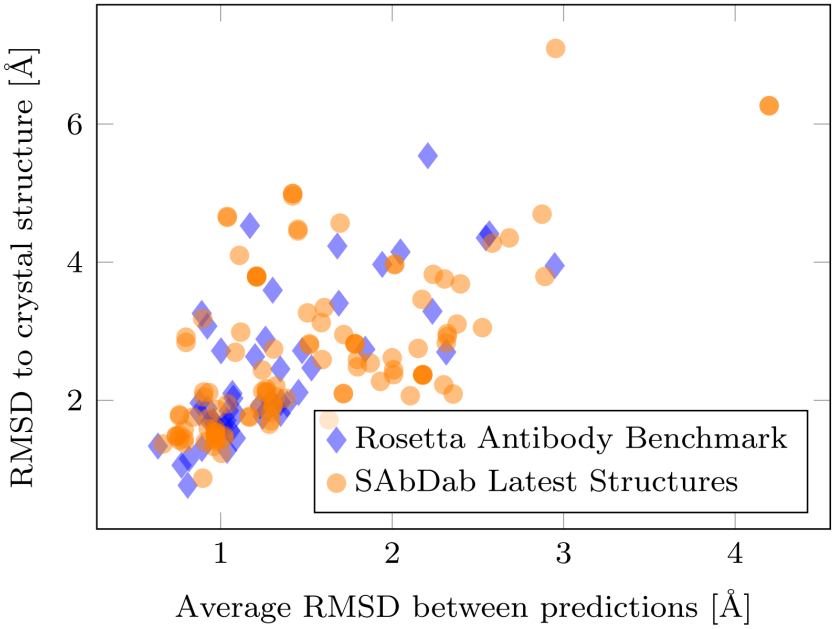
Abanades et al. (2022) Bioinformatics 10.1093/bioinformatics/btac016

ABlooper prediction diversity reveals prediction quality



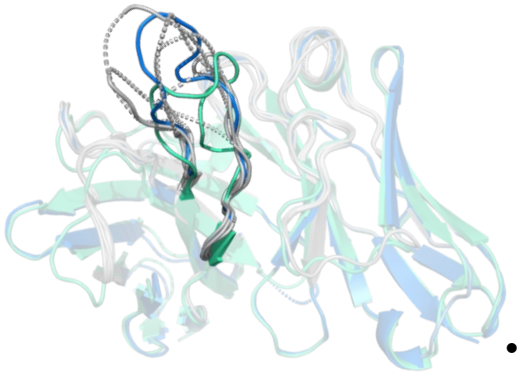


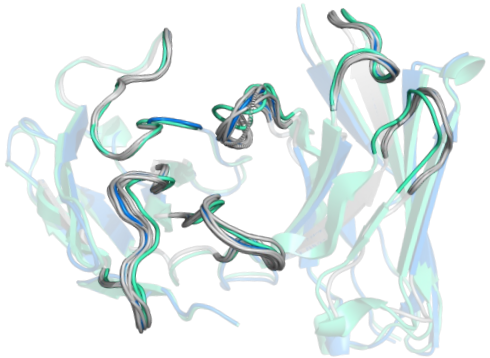
- **Crystal**
- **Decoys**
- **Prediction**



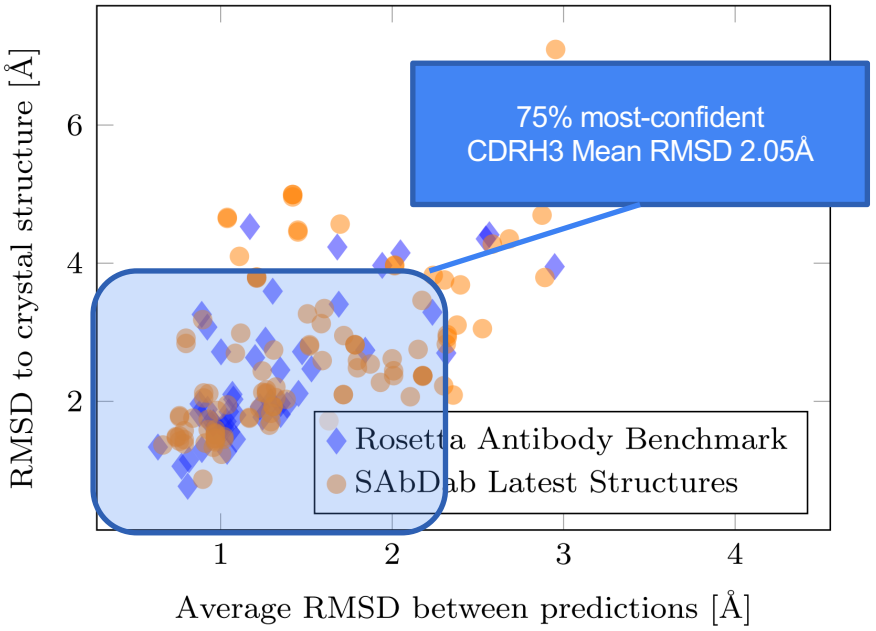
Abanades et al. (2022) Bioinformatics 10.1093/bioinformatics/btac016

ABlooper prediction diversity reveals prediction quality





- **Crystal**
- **Decoys**
- **Prediction**

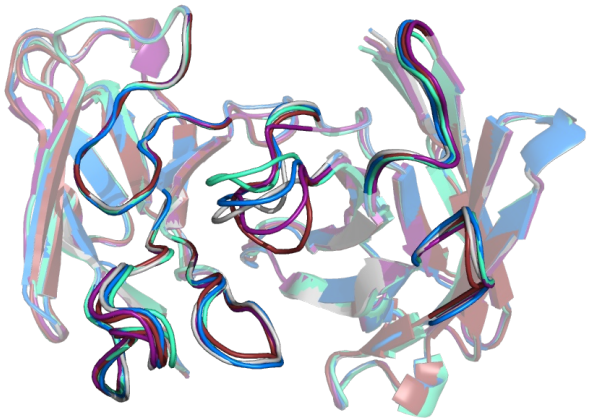


Abanades et al. (2022) Bioinformatics 10.1093/bioinformatics/btac016

ABlooper – rapid accurate structure prediction for antibodies

Overall similar levels of accuracy to other deep learning methods

ABlooper is **faster**
-> Can predict the CDR backbone conformation for one hundred antibodies in under five seconds.
-> ABlooper contains a useful accuracy estimate



- **Crystal**
- **ABodyBuilder**
- **ABlooper**
- **AlphaFold**
- **DeepAb**

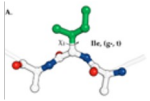
Abanades et al. (2022) Bioinformatics 10.1093/bioinformatics/btac016

EVQ../DIV..
QHQ../SYV..
(…)

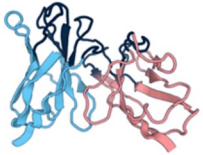
SAbPred

Antibody Structure Prediction/Featurisation

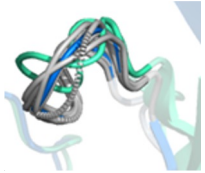
SCALOP/FREAD
Wong *et al*. 2018
Choi *et al*. 2011



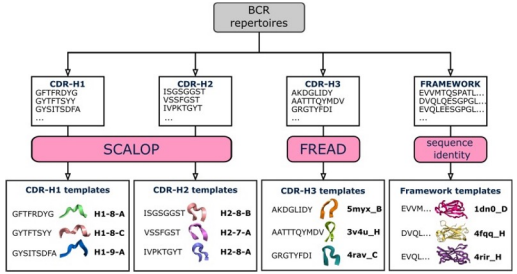
PEARS
Leem *et al*. 2018



ABodyBuilder
Leem *et al*. 2016

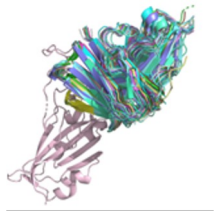


ABlooper
Abandes *et al*. 2022

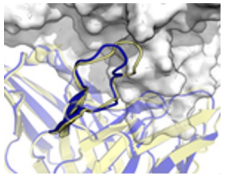


SAAB+, Repertoire Structural Profiling

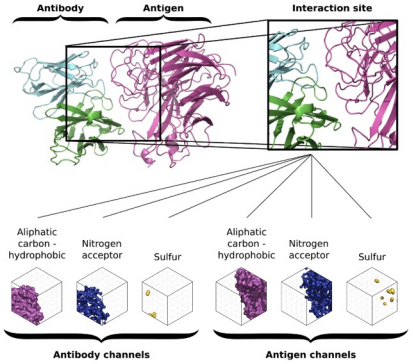
Kovaltsuk *et al*. 2021, Raybould *et al* 2021



SPACE
Robinson and Raybould *et al*. 2021



Paratyping, **Ab-Ligity**
Richardson *et al*. 2021
Wong *et al*. 2021



DLAB-VS
Schneider *et al*. 2022

How structure prediction can enhance antibody repertoire sequence analysis

Webinar Outline

- Recap on BCR/Antibody Repertoires & Publicly-available Sequence Data
- Sequence/Genetics-based Repertoire Analysis
- Structure-aware Repertoire Analysis
  1. Predict the paratope + cluster (paratyping)
  2. Predict structural features + cluster (SAAB+)
  3. Predict the full Fv structure + cluster (Ab-Ligity, SPACE, Repertoire Structural Profiling)
- Wrap-up & Questions

BCR/Antibody Repertoires

To ensure a spread of antigen complementarities, the body maintains a repertoire of chemically diverse BCRs

Estimated diversities range from 10^{10}-10^{18}

Many different “compartments” that reflect different B-cell roles/maturation stages

- Naive, Memory, Plasmablast
- CD proteins, isotypes can be diagnostic
- Different abundances in different physical locations

Repertoires are constantly **dynamic**:

- Turned over as B-cells die and new ones are made 
- Converging and expanding populations of BCRs that recognise pathogens/immunogens post-exposure

BCR/Antibody Repertoires

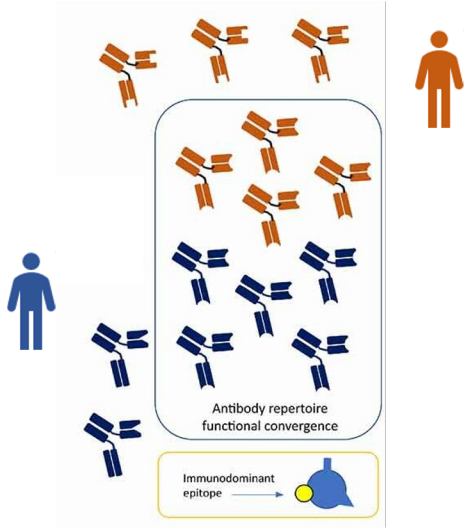
Therefore BCR repertoire functional dynamics often holds the key to understanding how an individual overcomes an external threat.

Antibody -> antigen specificity mappings can elucidate underlying causes of disease

This principle can be extended to:

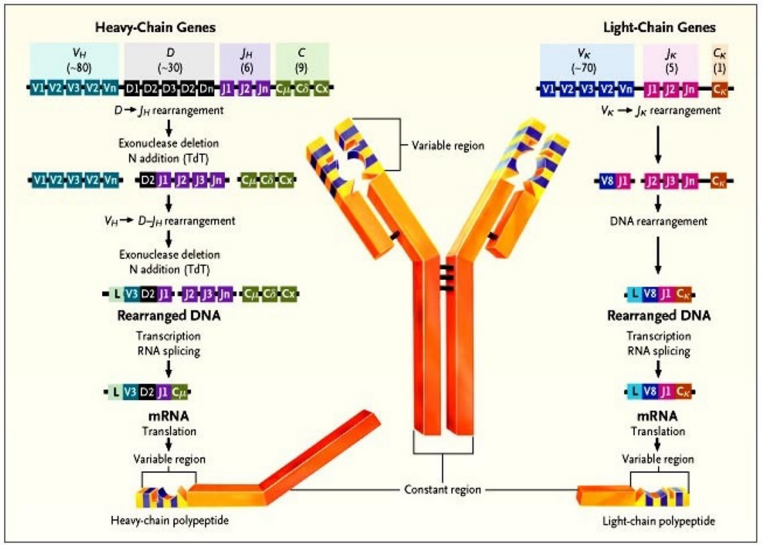
- Study epitope immunodominance through “public responses”
- Study why certain populations are resistant/susceptible to disease
- Find new, more “natural” antibody therapies

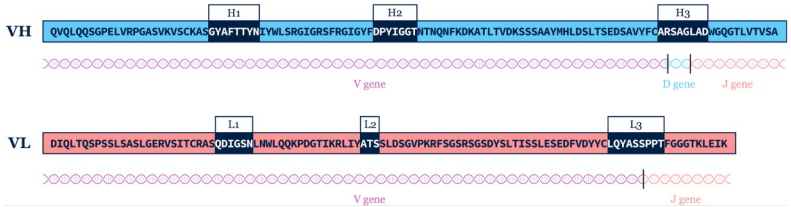
Immunoinformatics



Raybould *et al*. (2021) mAbs 13(1):1996732

Sources of Diversity in BCRs/antibodies





Naive B-cell Receptor (BCR) diversity results from
(a) diverse gene recombination
(b) junctional diversity at gene recombination regions
(c) heavy-light chain pairing

Antigen-experienced B-cells can migrate to the lymph nodes for further diversification through **somatic hypermutation**

Antibodies are soluble BCRs so bear identical diversification to their parent B-cell.

Schwarz, Robert (2003) N Engl J Med. 348:1017-1026

Sampling BCR Repertoires

Due to their enormous size, repertoires can only be **sampled:**

<u>Illumina MiSeq</u>
Ultra-deep **unpaired** V(D)J sequencing
Samples up to 10^9 VH [or VL] chains

<u>10X Chromium</u>
High-throughput paired VDJ+VJ chain sequencing
Samples up to 10^5 Fvs

Genomic sequencing is higher-throughput than proteomic sequencing -> we tend to study **BCR** not antibody repertoires

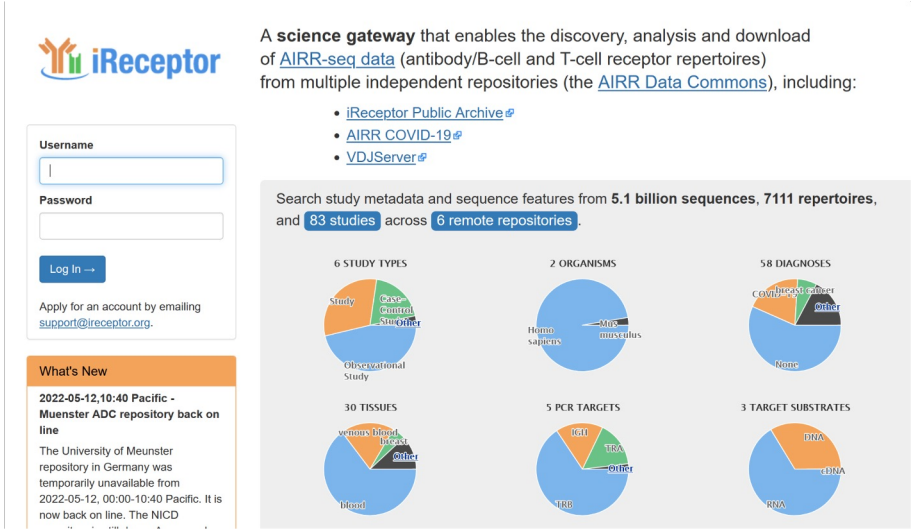
Particular B-cell populations of interest can be enhanced by pre-sorting (e.g. Fluorescence-activated mechanisms)

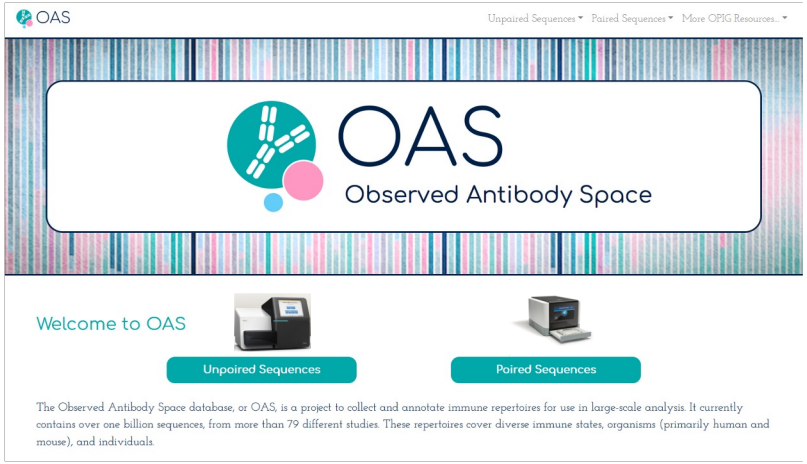
Community BCR-seq Databases

100's of BCR-seq datasets are stored in community resources such as iReceptor/OAS

Consistent data structure to facilitate "plug-and-play" analysis (MiAIRR Standards)

OAS: Structured data units, track repertoires by individual, disease status *etc*.





Corrie *et al*. (2018) Immunol Rev. 284(1):24-41

Kovaltsuk *et al*. (2018) J Immunol. 201(8):2502-2509
Olsen *et al*. (2022) Protein Sci. 31(1):141-146

Antigen-labelled Antibody Data

CoV-AbDab
The Coronavirus Antibody Database

-E.g. Fv sequences + structures against coronavirus antigens

> Search Database by Attribute

To view all entries, leave all search fields as 'All' and click 'Search'.

Origin: B-cells (SARS-CoV2 Human Patient)

Search Results

Back to search form

Your search returned 2405 results.

Key: Ab = Antibody, Nb = Nanobody, ND = Not Disclosed, Unk = Unknown, RBD = Receptor Binding Domain, NTD = N-terminal Domain, S = Spike Protein E = Envelope Protein, M = Membrane Protein, N = Nucleocapsid Protein, VH/VHH/VL = Variable Heavy Antibody/Heavy Nanobody/Light Antibody Domain, ABB = ABodyBuilder, SARS = Severe Acute Respiratory Syndrome, MERS = Middle-Eastern Respiratory Syndrome, TGEV = Transmissible Gastroenteritis Virus.

Downloads: CSV Excel

Show 10 entries Search:

Name	Ab or Nb	Binds to	Doesn't Bind to	Neutralising Vs	Not Neutralising Vs	Protein + Epitope	Origin
A19-46-1	Ab	SARS-CoV2_WT;SARS-CoV2_Alpha;SARS-CoV2_Beta;SARS-CoV2_Gamma;SARS-CoV2_Omicron-BA1		SARS-CoV2_WT;SARS-CoV2_Alpha;SARS-CoV2_Beta;SARS-CoV2_Gamma;SARS-CoV2_Omicron-BA1	SARS-CoV2_Delta	S; RBD	B-cells Patient
A19-61-1	Ab	SARS-CoV2_WT;SARS-CoV2_Alpha;SRAS-CoV2_Beta;SARS-CoV2_Gamma;SARS-CoV2_Delta	SARS-CoV2_Omicron-BA1	SARS-CoV2_WT;SARS-CoV2_Alpha;SRAS-CoV2_Beta;SARS-CoV2_Gamma;SARS-CoV2_Delta	SARS-CoV2_Omicron-BA1	S; RBD	B-cells Patient

Raybould et al. *Bioinformatics*. 37(5):734-735.

Classical BCR repertoire analysis pipelines

Repertoire Dataset of Interest
[+ reference repertoire(s)]

-nucleotides
-amino acids

Fv seqs

Map onto predicted clones

Clustering
(a.k.a. Clonotyping, building lineage trees)

Repertoire clonal properties

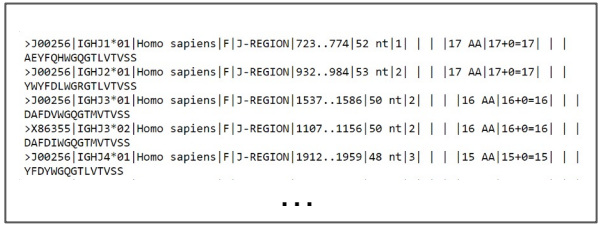
Constant domain sequence features
(e.g. isotype)

Clonal Mapping

CDRH3
(VDJ recombination region)

EVQLVESGGGLVKPGGSLRLSCAASGFTFRDVWMSWVRQAPGKGLEWVGRIKSKIDGGTTDYAAPVKGRFTISRDDSKNTLYLQMNSLKTEDTAVYYCTTAGSYYYDTVGPGLPEGKFDYWGQGTLVTVSS

V gene assignment



J gene assignment

Clonal Assignment
IGHV3-15, IGHJ4
TTAGSYYYDTVGPGLPEGKFDY

-can be done for VH/VL pairs too

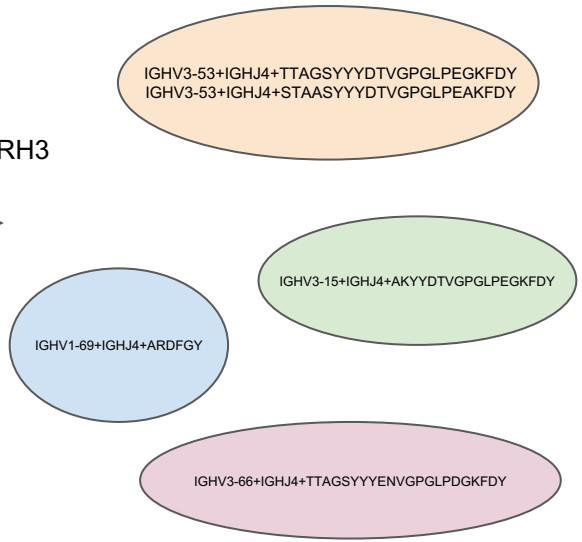
Clonal Clustering (Clonotyping)

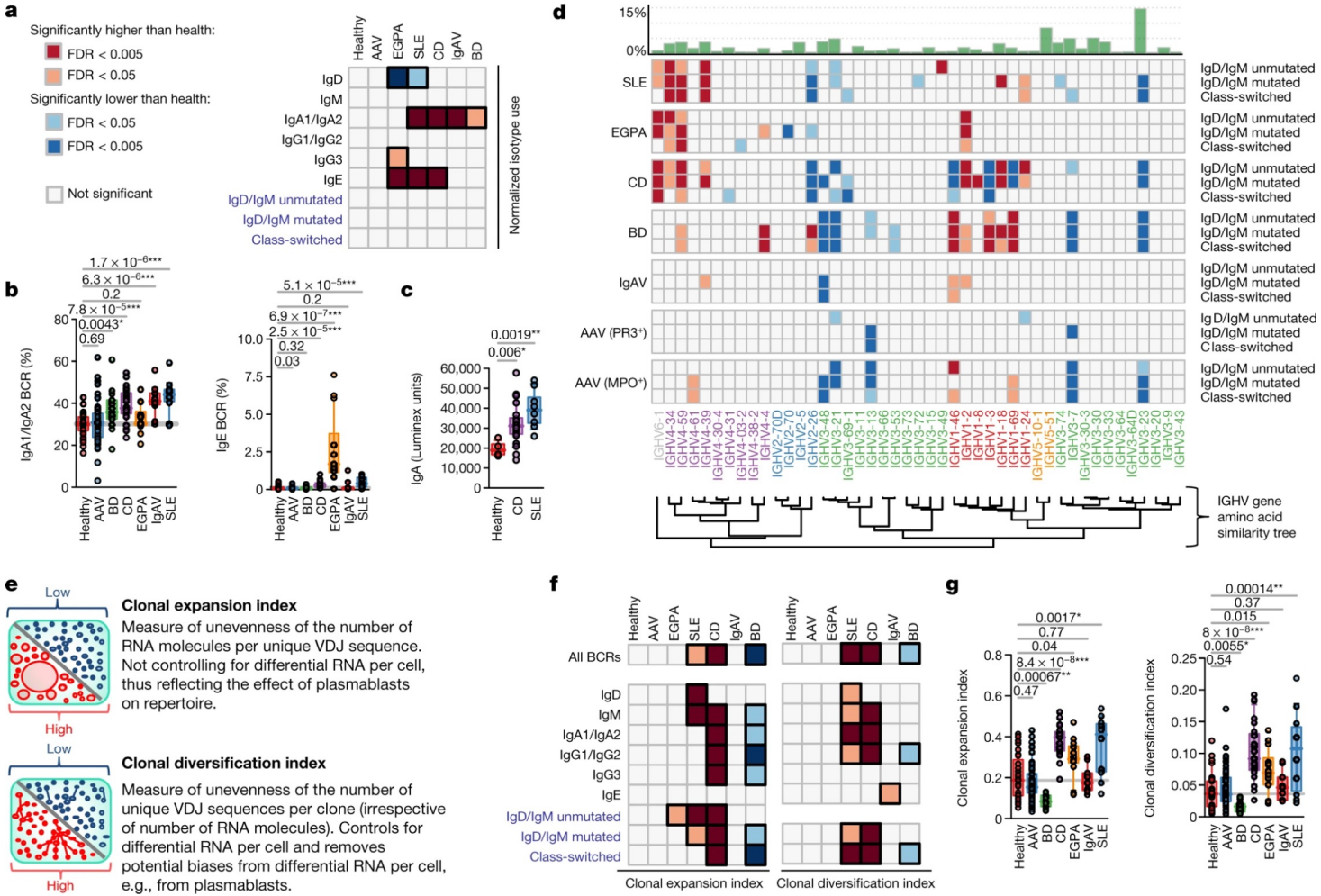
Clonotyped Repertoire

Clonal Representation of a Repertoire

IGHV3-53+IGHJ4+TTAGSYYYDTVGPGLPEGKFDY
IGHV1-69+IGHJ4+ARDFGY
IGHV3-66+IGHJ4+TTAGSYYYENVGPGLPDGKFDY
IGHV3-15+IGHJ4+AKYYDTVGPGLPEGKFDY
IGHV3-53+IGHJ4+STAASYYYDTVGPGLPEAKFDY
…

Cluster
same V, same J, same length CDRH3
80% CDRH3 ID





Bashford-Rogers *et al*. (2019) Nature 574(7776):122-126

Clustering repertoire clones with antibodies of known specificity

Key

Blue Antibodies
Clones with labelled specificity from other studies

ALC_#
Clonal clusters seen in multiple COVID-19 patients, no healthy individuals

Name	CDRH3	IGHV	IGHJ	Binds	Neutralizes	Reference
C154	AKQAGPYCSGGSCYSAPFDY	3-30	4	CoV-1, CoV-2	CoV-2	Robbiani et al. (14)
ALC_3983948	AKVSGPYCSGGSCYSFYFDY	3-30	4			
COV2-2068	ARSYDILTGYRDAFDI	3-53	3	CoV-2	CoV-2	Zost et al. (38)
ALC_2318471	VRNYDILTGYSDAFDI	3-53	3			
COV2-2007	AKVSATYYYYYGMDV	3-30	6	CoV-1, CoV-2		Zost et al. (38)
CC12.17	AKSSGSYYYYYGMDV	3-30	6	CoV-2	CoV-2	Rogers et al. (39)
ALC_2318830	AKVMTTYYYYYGMDV	3-30	6			
COVA2-14	ARVRYYDSSGYYEDY	1-69	4	CoV-1, CoV-2		Brouwer et al. (12)
ALC_1780442	ARYDYYDSSGYYLDY	1-69	4			
COV2-2270	AITYYYDSSGYWWDD	1-69	4	CoV-1, CoV-2		Zost et al. (38)
ALC_1781971	ASTYYYDSSGYWFDY	1-69	4			
COVA2-40	AGRYCSGGRCGWFDP	4-4	5	CoV-2		Brouwer et al. (12)
ALC_1784026	ESRYCSGGSCGWFDP	4-4	5			
COV2-2147	ARSTSGSYYYGMDV	3-30-3	6	CoV-1, CoV-2		Zost et al. (38)
CV34	ARSYGGSYYYGMDV	3-30-3	6	CoV-2		Seydoux et al. (40)
ALC_1249094	ARGTRGSYYYGMDV	3-30-3	6			
COV2-2006	ARPQSGGYYAPLDY	3-30-3	4	CoV-1, CoV-2		Zost et al. (38)
ALC_1246650	ARPYSGSYYAPLDY	3-30-3	4			
S304	ARGDSSGYYYYFDY	3-13	4	CoV-1, CoV-2	CoV-1, CoV-2	Pinto et al. (41)
ALC_1245048	ARGYSSGYYYYFDY	3-13	4			
COV2-2027	AIYGYYYYGLDV	3-30	6	CoV-2		Zost et al. (38)
ALC_480504	AVYGYYYYGMDV	3-30	6			

“Overlapping clonotypes” -> Predict the complementarities of the “public” antibodies in the immune response

Galson *et al*. (2020) *Front Immunol.* 3283

Limits to Classical Clonal Clustering

Clonotyping: all residues in the CDRH3 are equally weighted

But antigen-contacting CDRH3 residues (green) are often more conserved than the residues outside the paratope

IGHV3-53
IGHV3-66

Clonotyping: V genes should match

99% sequence identity despite coming from different loci, probably very similar V-encoded binding site

IGHV3-53 (100%)
IGHV3-53 (85%)

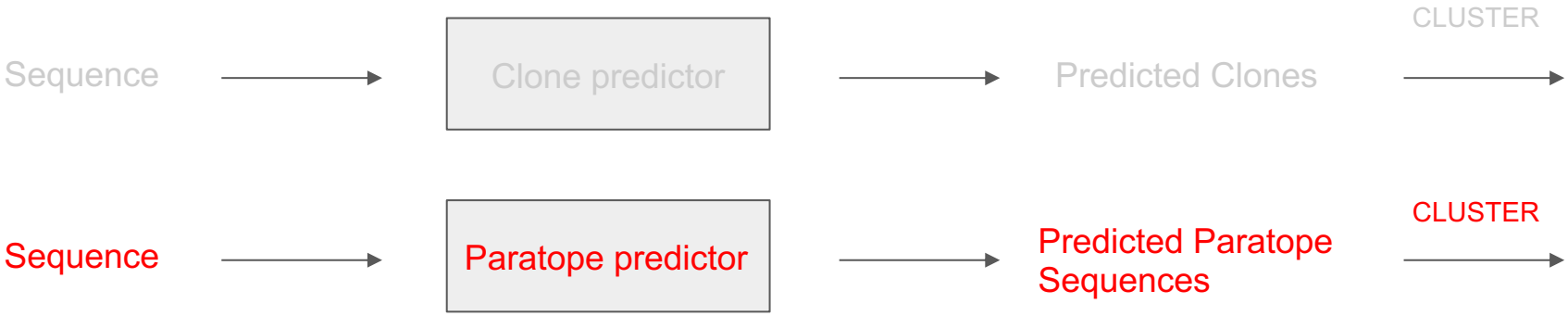
Clonotyping: Just label by closest V gene

Clonotyping does not account for somatic hypermutations - these can be very important for binding if in CDR loops

If we could predict the residues likely to be involved in binding, this may lead to new functionally-meaningful clusters

How structure can help (1)

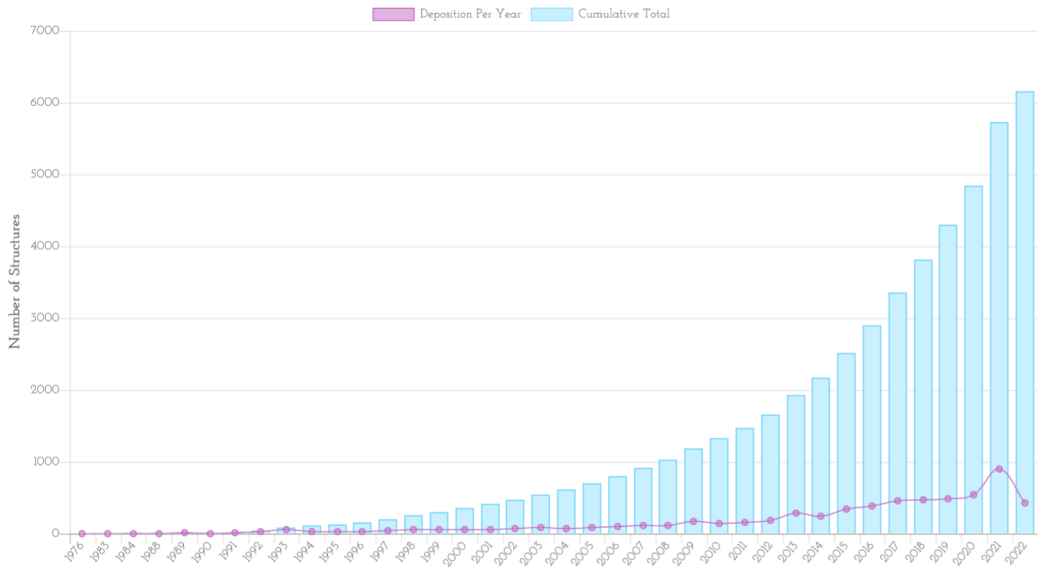
Use structures of antibody:antigen complexes to learn which residues tend to be in the paratope, then cluster on predicted paratope



SAbDab

The Structural Antibody Database

Antibody Depositions in the PDB



-records all PDB structures containing antibodies/nanobodies

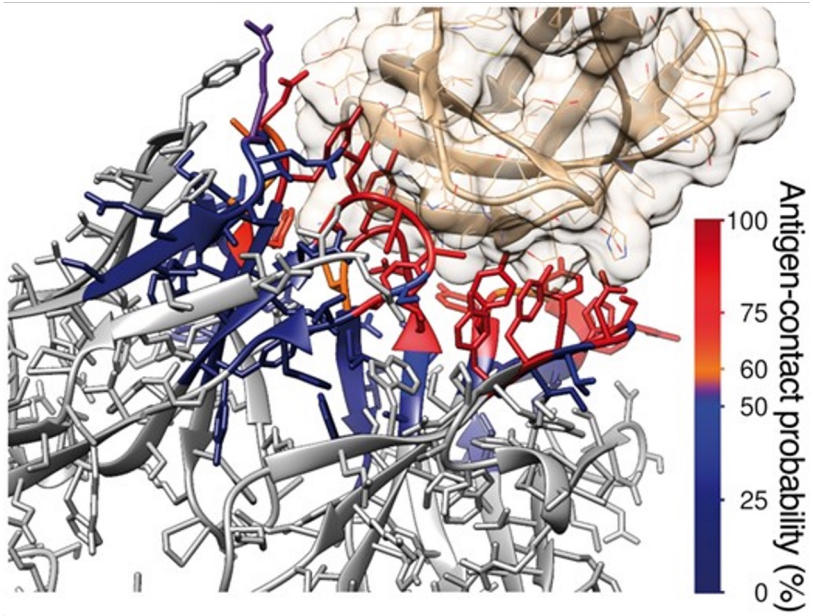
-75% in complex with antigen

-data labelled in consistent numbering schemes

-bespoke antibody data type inspired by Bio module. handy API for dataset curation.

Dunbar *et al*. (2014) Nucleic Acids Res. 42(D1):D1140-1146
Schneider *et al*. (2022) Nucleic Acids Res. 50 (D1):D1368-1372

Learning general paratopes from solved structures



Current field-standard: parapred, AG-Fast-parapred

One-hot encoded residues + 7 expressive chemical features

Labelled per-residue paratope/not paratope based on SAbDab structures

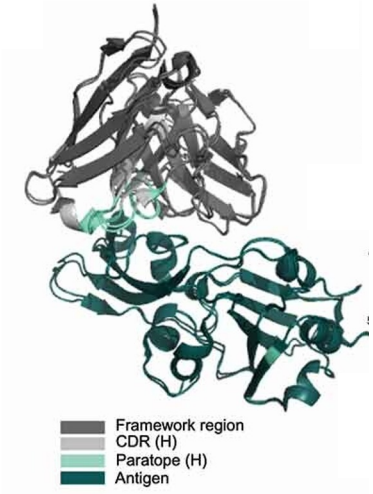
Deep neural network architectures applied to the sequence representation to predict whether part of the paratope.

Area under PR curve > 0.87

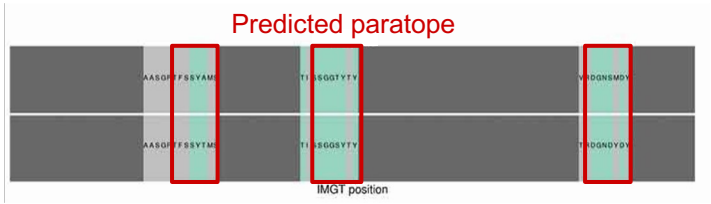
OPIG ML tool to do this coming soon…

Liberis *et al*. (2018) Bioinformatics. 34(17):2944-2950.

Cluster over the predicted paratope residues: “paratyping”



	VH	VL
4zpv_HL	IGHV9-5-3/IGHJ4/VRDGNSMDY	IGKV10-96/IGKJ5/QQANTLPPT
5do2_HL	IGHV5-6-4/IGHJ2/TRDGNDYDY	IGKV10-96/IGKJ1/QQGNTLPRT



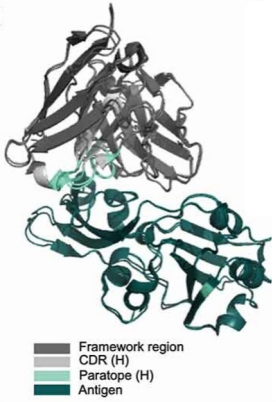
Don’t belong to the same clonotype
(Different V-genes, J-genes, CDRH3 identity < 80%)

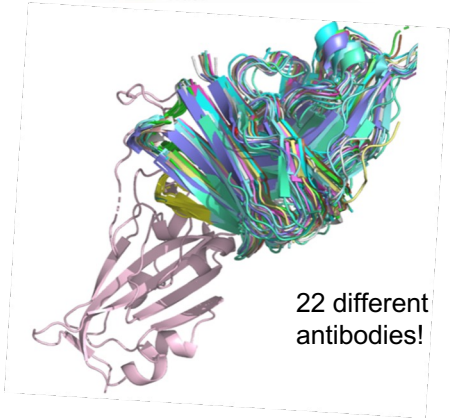
Do belong to the same paratope
Sequence identity across predicted VH paratope > 80%

-> Paratyping can group together antibodies that bind to the same epitope, despite them belonging to different clones. It can also reaffirm clonotype clusters.
-> Very simple to get started: http://opig.stats.ox.ac.uk/data/downloads/Paratyping_code.tar.gz

Richardson *et al*. (2021) mAbs 13(1):1869406

How structure can help (2): Structure-Function in Antibodies





Beyond paratope conservation, another common observation is that antibodies that bind to the same epitopes **tend to have very similar structures**

Incorporate **structure-awareness** into the way we cluster antibodies

Solving the structure of a single antibody (let alone a repertoire!) highly laborious, but we can computationally predict from sequence what an antibody will look like in 3D…

1. Structurally annotating VH repertoires with predictions of individual loop conformations, use this to study repertoire dynamics/complementarities

2. Explicitly predict the entire 3D structure, use this in epitope profiling and *de novo* complementary assessment

Predicting CDR Structural Features

Canonical Forms (SCALOP)

CDR loops, except H3, tend to adopt one of a small set of backbone structures, or 'canonical forms'.

By clustering solved structures, we can compare sequences to representatives of each cluster to rapidly assign a canonical form.

Homologous Loops (FREAD)

We can also use solved structures to perform homology modelling of loops.

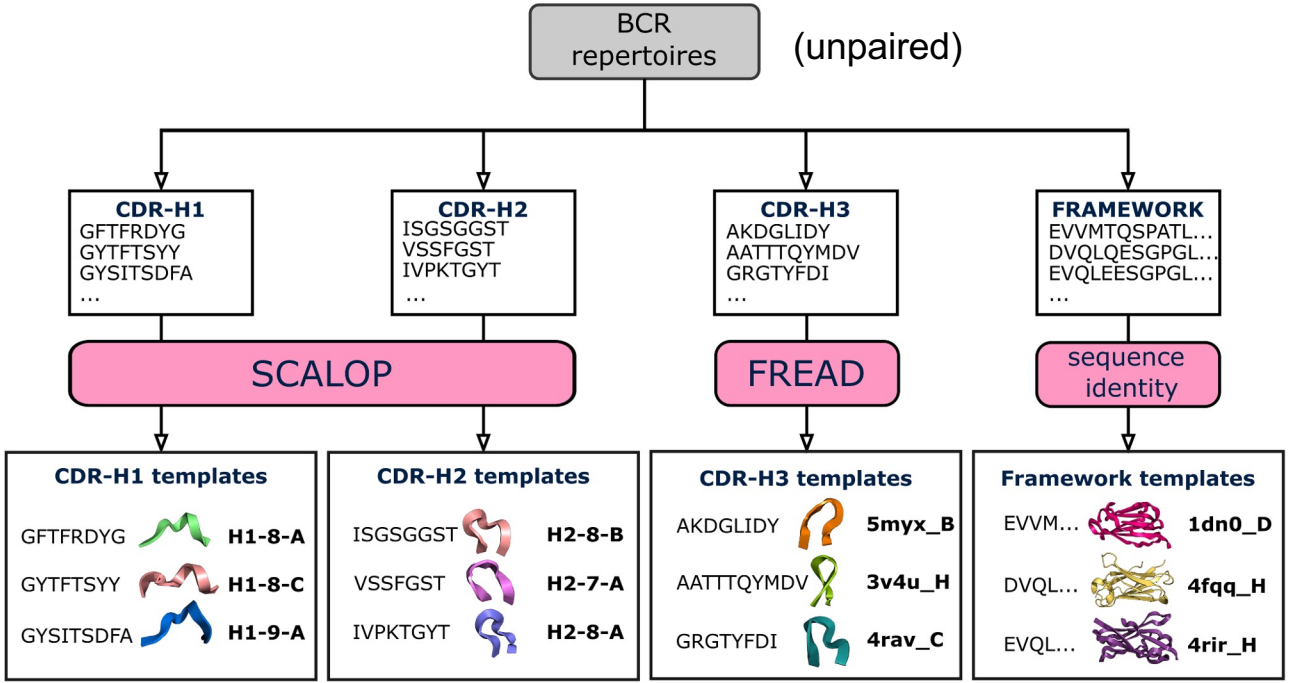
FREAD: dihedral-aware loop template assignment

So even without performing full modelling, assigning structural templates to sequences lets us group them by likely structural similarities…

Wong *et al.* (2019) Bioinformatics 10.1093/bioinformatics/bty877
Choi and Deane (2009) Proteins: Struct Funct Bioinf 10.1002/prot.22658

Structural annotation of BCR repertoires (SAAB+)

Can we say something useful about structural features without a model of the structure?



Kovaltsuk *et al*. (2020) PLOS Comput Biol. 10.1371/journal.pcbi.1007636

Benefits of Structural Annotation

Simpler/faster than full antibody modelling, can be readily to large repertoire datasets

Single-sequence, so useful even if we only have e.g. heavy chains

Structural profiling can tell us more about variations between e.g. repertoires from different species, or the B-cell differentiation axis.

CDRH3 Template Usage Clustering

a CDR-H3 Clusters

b Observed Naive BCR Repertoires

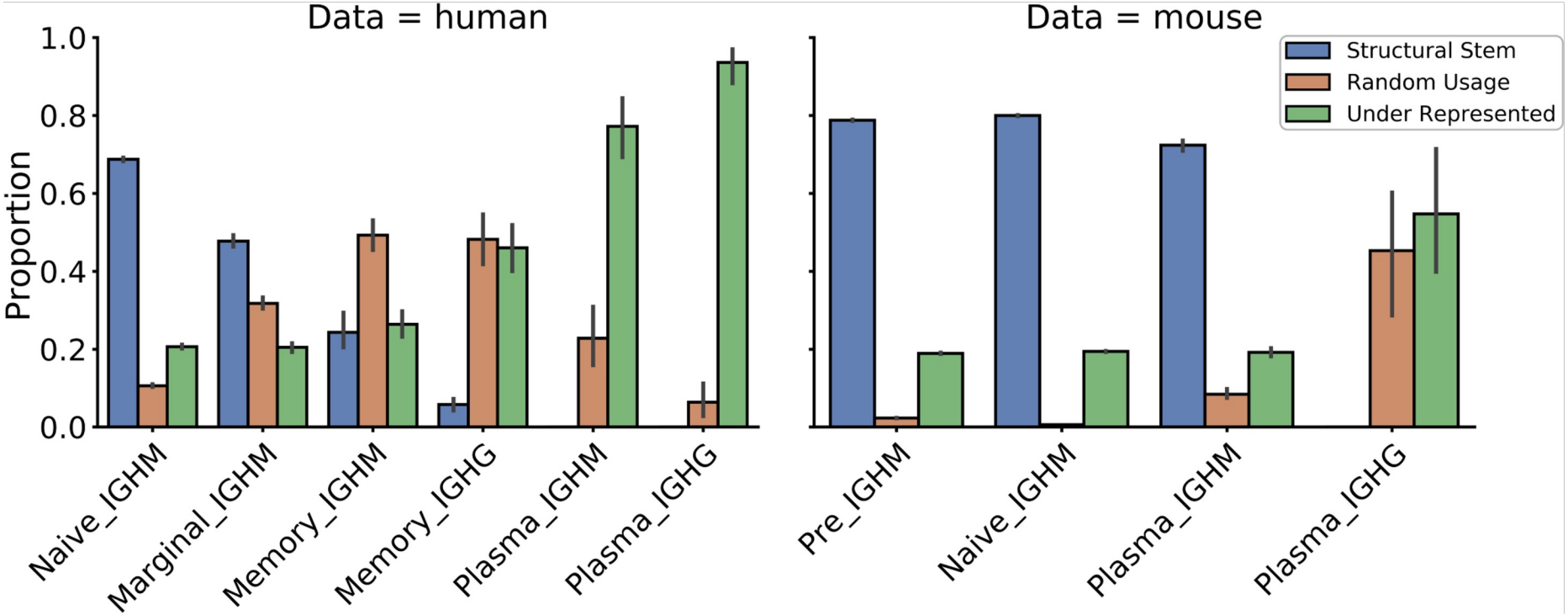
c Randomly Sampled (RS) Naive BCR Repertoires

d Cluster Usage

Cluster ID	RS	Observed				Classification
#1	25	41	33	30	45	Structural Stem
#2	25	30	29	30	37	Structural Stem
#3	25	27	27	16	8	Random-Usage
#4	25	2	11	24	10	Under-Represented

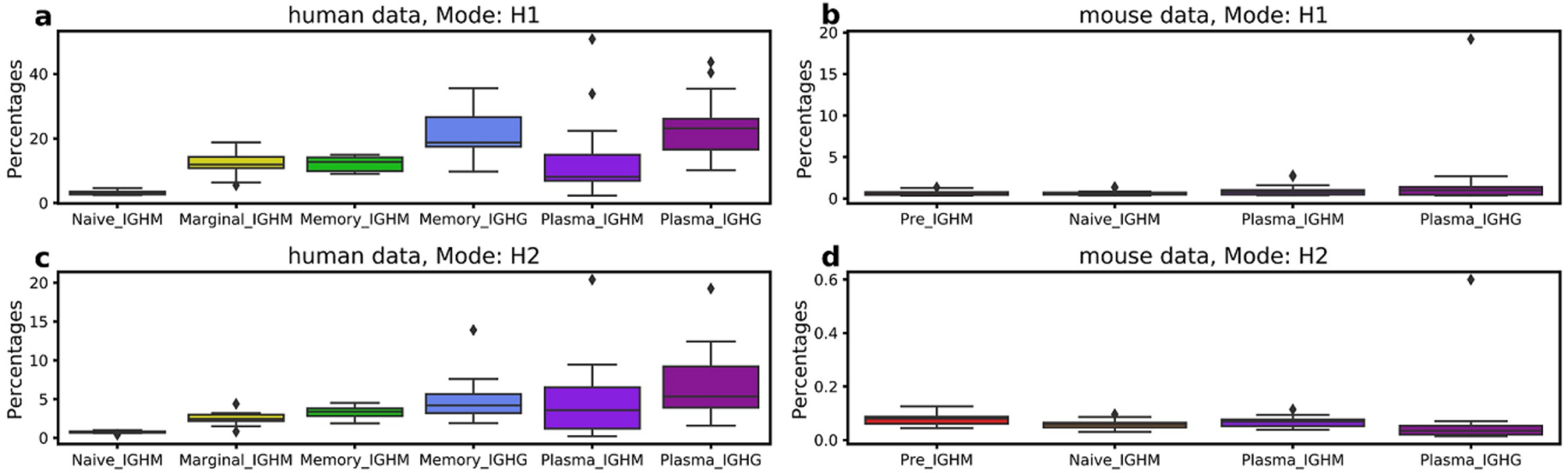
Kovaltsuk *et al*. (2020) PLOS Comput Biol. 10.1371/journal.pcbi.1007636

CDRH3 Template Usage Clustering: Differentiation



Kovaltsuk *et al*. (2020) PLOS Comput Biol. 10.1371/journal.pcbi.1007636

CDRH1/CDRH2 Canonical Form Usages: Germline Deviation



Kovaltsuk *et al*. (2020) PLOS Comput Biol. 10.1371/journal.pcbi.1007636

Using SAAB+ Annotations to Infer Antigen Binding

Ab Name	Batch	$CDR3_H$	$CDR3_L$
m396	/	ARDTVMGGMDV	QVWDSSSDYV
BD-494	1	ARDLVVYGMDV	QQLNSYPFT
BD-495	1	ARDPIRNGMDV	QQYDNLPRT
BD-498	2	ARDLVVYGMDV	QQLNSYPLT
BD-500	3	ARDAMSYGMDV	QQSYSTPPDT
BD-501	3	ARDRVVYGMDV	QQYDNLPPT
BD-503	5	ARDAAVYGIDV	QQSYTTPLFT
BD-504	5	ARDLISRGMDV	QQSYTTPLFT
BD-505	5	ARDRVVYGMDV	HQYDNLPPT
BD-506	6	ARDLVSYGMDV	QQLNSYPLT
BD-507	6	ARDLVVYGMDV	QQLNSNPPIT
BD-508	6	ARDAQNYGMDV	QQSYSTPPYT

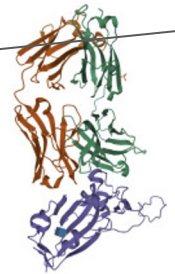
Paired sequences from SARS-CoV-2 response repertoires

Rel. low sequence similarity to each other and to known SARS-CoV-2 binders at the time (July 2020)

Characterization of the mAbs identified by $CDR3_H$ structure similarity

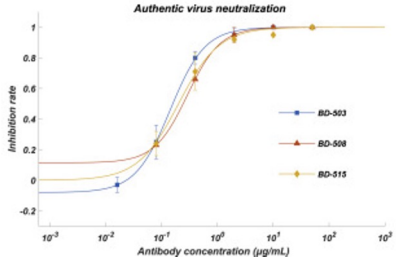
		Viral protein binding		Pseudovirus neutralizition		VDJ combination				$CDR3_H$ similair structure
mAbs	Batch	Enrichment antigen	K_D (RBD) (nM)	IC_{50} (μg/mL)	IC_{90} (μg/mL)	V_H	J_H	V_L	J_L	PDB ID
BD-494	1	Spike	0.69	0.024	0.12	IGHV3-53	IGHJ6	IGKV1-9	IGKJ3	2dd8H
BD-495	1	Spike	> 50	18	> 30	IGHV3-66	IGHJ6	IGKV1D-33	IGKJ1	2dd8H
BD-498	2	Spike	2.5	0.092	0.32	IGHV3-66	IGHJ6	IGKV1-9	IGKJ5	2dd8H
BD-500	3	RBD	2.3	0.018	0.13	IGHV3-53	IGHJ6	IGKV1D-39	IGKJ5	2dd8H
BD-501	3	RBD	> 50	> 30	> 30	IGHV3-53	IGHJ6	IGKV1D-33	IGKJ5	2dd8H
BD-503	5	RBD	0.24	0.016	0.12	IGHV3-53	IGHJ6	IGKV1D-39	IGKJ3	2dd8H
BD-504	5	RBD	0.32	0.033	0.27	IGHV3-66	IGHJ6	IGKV1-9	IGKJ3	2dd8H
BD-505	5	RBD	1.2	0.012	0.10	IGHV3-53	IGHJ6	IGKV1D-33	IGKJ5	2dd8H
BD-506	6	RBD	2.7	0.55	3.4	IGHV3-53	IGHJ6	IGKV1-9	IGKJ4	2dd8H
BD-507	6	RBD	1.3	0.070	0.18	IGHV3-53	IGHJ6	IGKV1-9	IGKJ3	2dd8H
BD-508	6	RBD	1.9	0.015	0.069	IGHV3-53	IGHJ6	IGKV1D-39	IGKJ2	2dd8H
BD-515	3	RBD	0.041	0.022	0.15	IGHV3-66	IGHJ4	IGKV1D-33	IGKJ5	2ghwD

m396-Fab/RBD



But formed a VH SAAB+ structural cluster that commonly exploited <u>loop templates from 2dd8, a SARS-CoV-1 binder</u>

Authentic virus neutralization



Authentic virus PRNT

mAbs	Batch	IC_{50} (μg/mL)	IC_{90} (μg/mL)
BD-503	5	0.14	0.39
BD-508	6	0.29	0.74
BD-515	3	0.20	0.64

-> 7/12 antibodies from this SAAB+ cluster were able to neutralise SARS-CoV-2 *via* a common epitope

Cao *et al*. (2020) Cell. 182(1):73-84

As we increasingly have **paired Fv** repertoire data, we can model the complete Fv structures

Whole-protein structure prediction in general has been steadily improving over recent years and is now at the stage where rapidly-modeled structures can provide valuable functional inference

Article | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, [...]Demis Hassabis

Nature (2021) | Cite this article

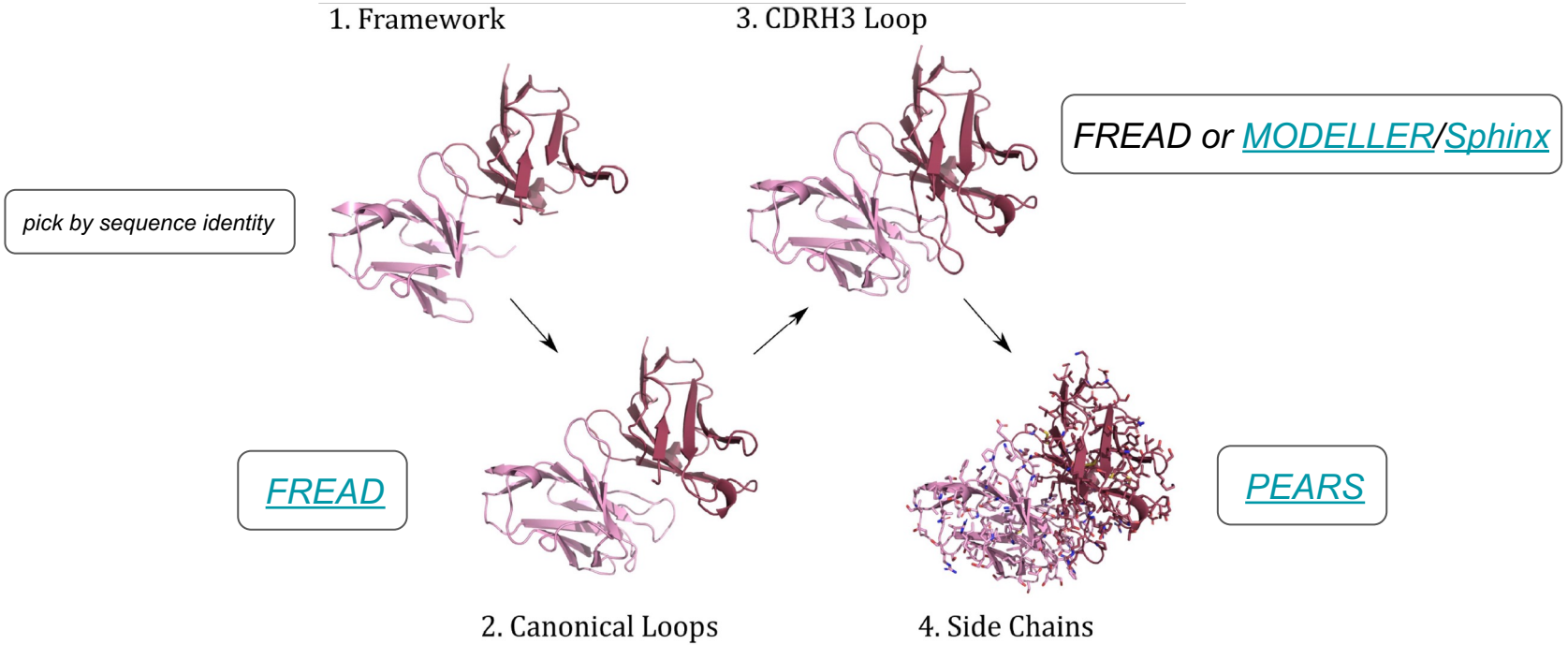
110k Accesses | **1** Citations | **2017** Altmetric | Metrics

RESEARCH ARTICLE

Accurate prediction of protein structures and interactions using a three-track neural network

Minkyung Baek[1,2], Frank DiMaio[1,2], Ivan Anishchenko[1,2], Justas Dauparas[1,2], Sergey Ovchinnikov[3,4], Gyu Rie Lee[1,2], Jue Wang[1,2], Qian Cong[5,6], Lisa N. Kinch[7], R. Dustin Schaeffer[6], Claudia Millán[8], Hahnbeom Park[1,2], Carson Adams[1,2], Caleb R. Glassman[9,10], Andy DeGiovanni[12], Jose H. Pereira[12], Andria V. Rodrigues[12], Alberdina A. van Dijk[13], Ana C. Ebrecht[13], Diederik J. Opperman[14], Theo Sagmeister[15], Christoph Buhlheller[15,16], Tea Pavkov-Keller[15,17], Manoj K. Rathinaswamy[18], Udit Dalwadi[19], Calvin K. Yip[19], John E. Burke[18], K. Christopher Garcia[9,10,11,20], Nick V. Grishin[6,21,7], Paul D. Adams[12,22], Randy J. Read[8], David Baker[1,2,23,*]

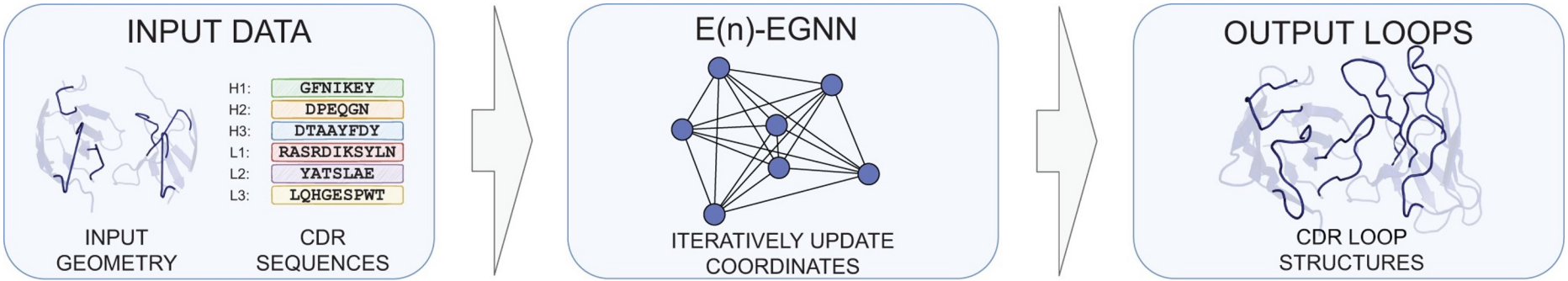
Modular Fv Structure Prediction: ABodyBuilder



c. 1A accuracy of canonical CDR loops, around 3A accuracy of CDRH3
Overall: ~20s/complete model on 1 CPU

ABLooper: loop modelling with deep learning

- AlphaFold2 - do EGNNs offer a way of rapidly predicting structure?
- ABLooper - EGNN predicts loop structures starting from sequence and anchor residues
- Energy minimisation required to ensure correct geometry
- Diversity of predicted structures -> confidence estimate



Abanades et al. (2022) Bioinformatics [10.1093/bioinformatics/btac016](https://doi.org/10.1093/bioinformatics/btac016)

ABlooper Benefits

Improved accuracy: reliably sub-Angstrom on canonical loops, 2.5Å on CDRH3

Confidence estimates are correlated with prediction accuracy, meaning you know more often when your models are reliable
-scores < 1 for canonical loops: canonical form is almost certainly correct
-score < 1.5 for CDRH3, within 2Å over 90% of the time

Extremely fast at predicting loop backbone conformations: 100 in 5 seconds on a single GPU
>10,000 sequences from a 10X run could be modelled in *c*. 8 minutes

GitHub

SAbPred

Web application coming soon

Applications of Full Fv Models: Epitope Profiling (SPACE)

With antigen-labelled data (e.g. antigen-sorted paired Fv repertoire), predict which antibodies bind to the same epitopes

Common approach: clonotype the data, overlapping clonotypes likely bind the same epitope

The approach in SPACE: cluster their model structures, overlapping structures likely bind the same epitope

Robinson and Raybould et al. (2021) PLoS Comput Biol. 17(12):e1009675

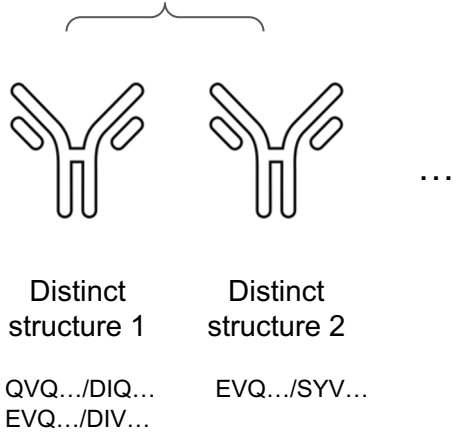
Applications of Full Fv Models: Epitope Profiling (SPACE)

CoV-AbDab
The Coronavirus Antibody Database

c. 2,000 Coronavirus-Binding Antibody Sequences [Mar' 21]

[now have closer to 4,500]

→ Structural Modelling → Filter for confidence → Structural Clustering

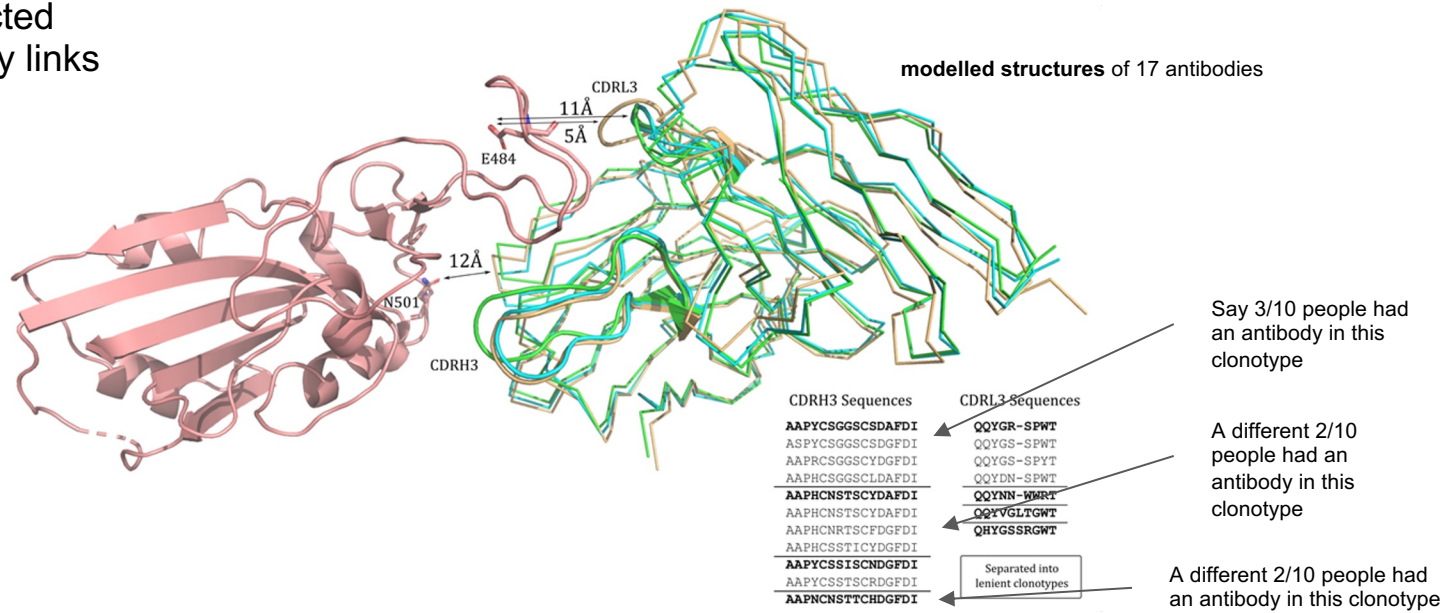


- Led to 200 multiple-occupancy distinct structures
- Of these clusters, the binding sites of antibodies in 92% were entirely domain-consistent
- Some that were inconsistent could be linked to experimental uncertainty

Robinson and Raybould et al. (2021) PLoS Comput Biol. 17(12):e1009675

Applications of Full Fv Models: Epitope Profiling (SPACE)

Clustering by predicted structure functionally links different clones…

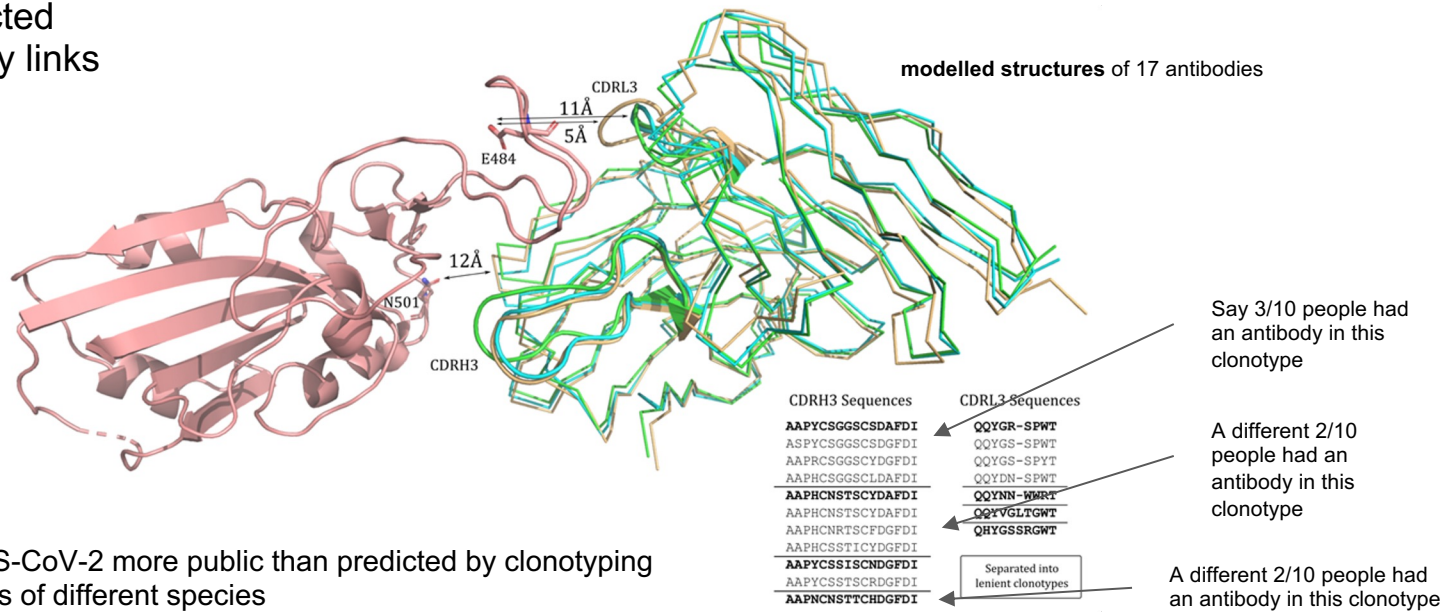


Robinson and Raybould et al. (2021) PLoS Comput Biol. 17(12):e1009675

-> up to 70% of people responded to this epitope

Applications of Full Fv Models: Epitope Profiling (SPACE)

Clustering by predicted structure functionally links different clones…

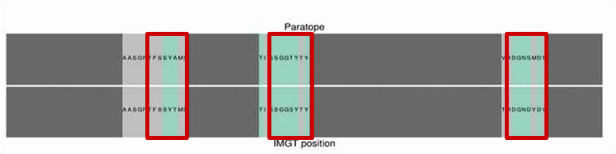


1) Response to SARS-CoV-2 more public than predicted by clonotyping
2) Can link repertoires of different species

Robinson and Raybould et al. (2021) PLoS Comput Biol. 17(12):e1009675

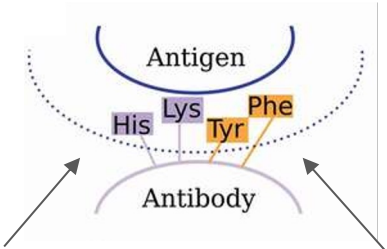
-> up to 70% of people responded to this epitope

Applications of Full Fv Models: Functional Clustering (Ab-Ligity)



Paratyping: purely sequence-based clustering over the predicted paratope

Richardson *et al*. (2021) mAbs. 13(1):1869406



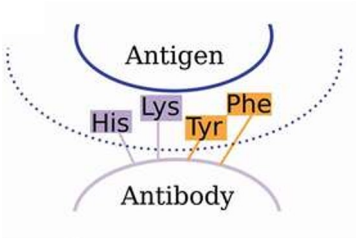
Ab-Ligity: sequence and structure-based clustering over the predicted paratope

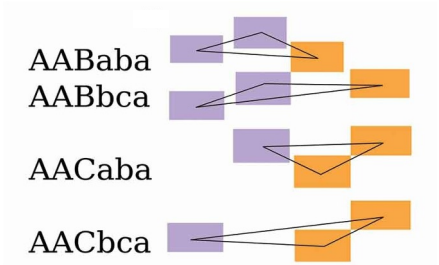
Wong *et al*. (2021) mAbs. 13(1):1873478

Relative spatial positioning of paratope residues imparted by Fv model

Predicted paratope residues in VH and VL. ‘Colour’ by properties

Ab-Ligity



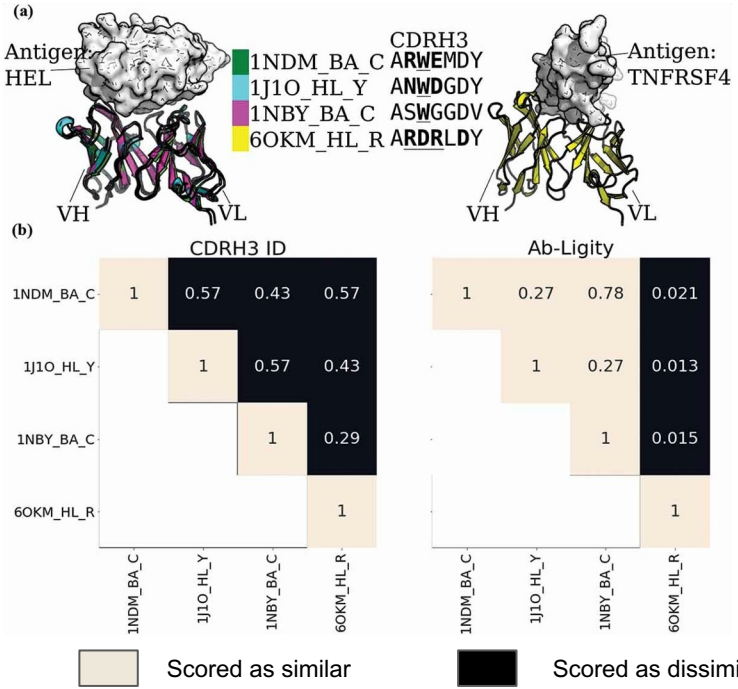


3D model of the predicted paratope

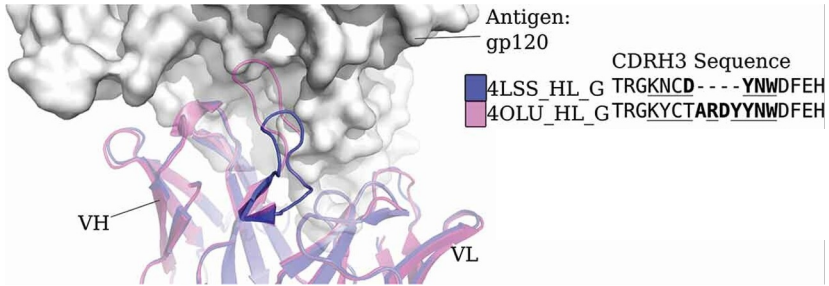
1. Convert predicted paratope into hash tables
2. Measure the Tversky similarity of different antibodies' hash tables.
   -threshold set based on solved structures
   -if threshold similarity met, classify as same-epitope binders

Wong *et al*. (2021) mAbs. 13(1):1873478

Ab-Ligity



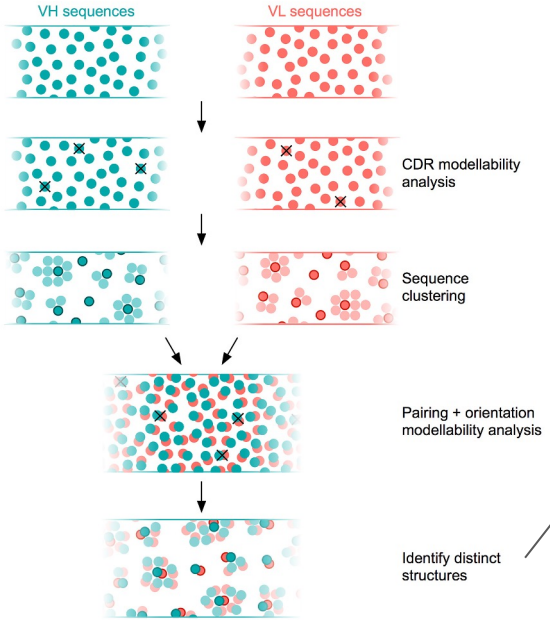
Ab-Ligity identifies same-epitope binders with ignoring a structurally similar binder to a different epitope



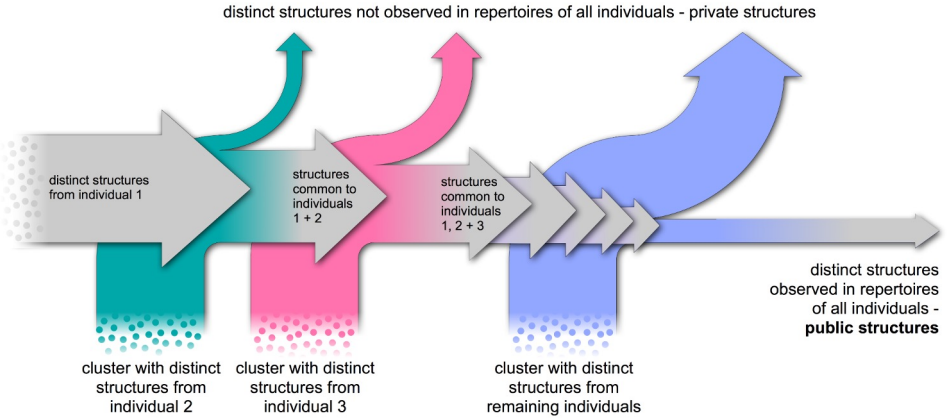
Despite CDRH3 length difference, Ab-Ligity still identifies these two antibodies as same-epitope binders

Wong *et al*. (2021) mAbs. 13(1):1873478

Applications of Full Fv Models: Repertoire Structural Profiling

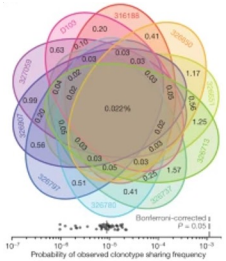


Snapshot of Fv structures in a repertoire



Raybould *et al*. (2021) PLoS Comput Biol 17(3):e1008781

Applications of Full Fv Models: Repertoire Structural Profiling



clonal sharing across healthy individuals is quite minimal

All 10 individuals shared just 0.002% of clonotypes

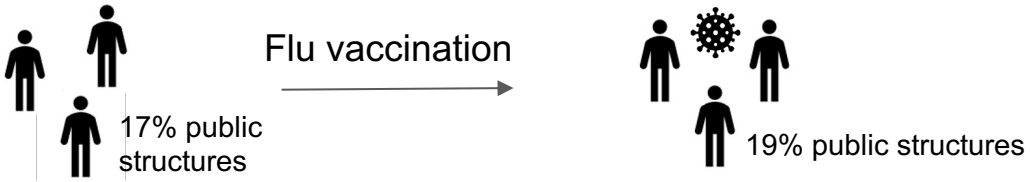
Briney *et al*. (2019) Nature **566**, 393–397

DS: Distinct Structures

# Repertoires Added	Fvs Added	Cumulative DSs	Public DSs (% Public)	Expected Public DSs (% Public)
1 (S64)	6,420,211	209,394	209,394	209,394
2 (+S57)	7,225,630	340,915	100,824 (29.57%)	12,307 (3.10%)
3 (+S5)	6,827,419	445,045	71,743 (16.12%)	1,600 (0.28%)
4 (+S56)	6,628,683	527,668	58,043 (11.00%)	322 (0.06%)
5 (+S83)	6,170,373	604,124	48,703 (8.06%)	86 (< 0.01%)
6 (+S67)	5,544,952	670,833	42,277 (6.30%)	31 (< 0.01%)
7 (+S84)	5,624,598	734,374	37,151 (5.06%)	17 (< 0.01%)
8 (+S76)	5,856,150	793,831	33,572 (4.23%)	9 (< 0.01%)
9 (+S54)	5,074,822	846,670	30,474 (3.60%)	6 (< 0.01%)
10 (+S89)	5,414,820	896,328	27,389 (3.06%)	4 (< 0.01%)

https://doi.org/10.1371/journal.pcbi.1008781.t002

However Fv structure sharing across healthy individuals is more significant at c. 3%. We propose this as a structural ‘basis set’.



Raybould *et al*. (2021) PLoS Comput Biol 17(3):e1008781

De novo antibody complementarity assessment

So far, we've used structural information to profile a repertoire, to cluster sequence-dissimilar antibodies by function, and to determine whether two antibodies bind to the same epitope.

One further question we might ask is: "we have a repertoire associated with a disease; does it contain anything that will bind to an antigen we're interested in?"

Having access to 3D structures for representative antibodies (e.g. the public basis set from Repertoire Structural Profiling) allows us to answer this question directly by assessing binding complementarity between the antibodies and antigen

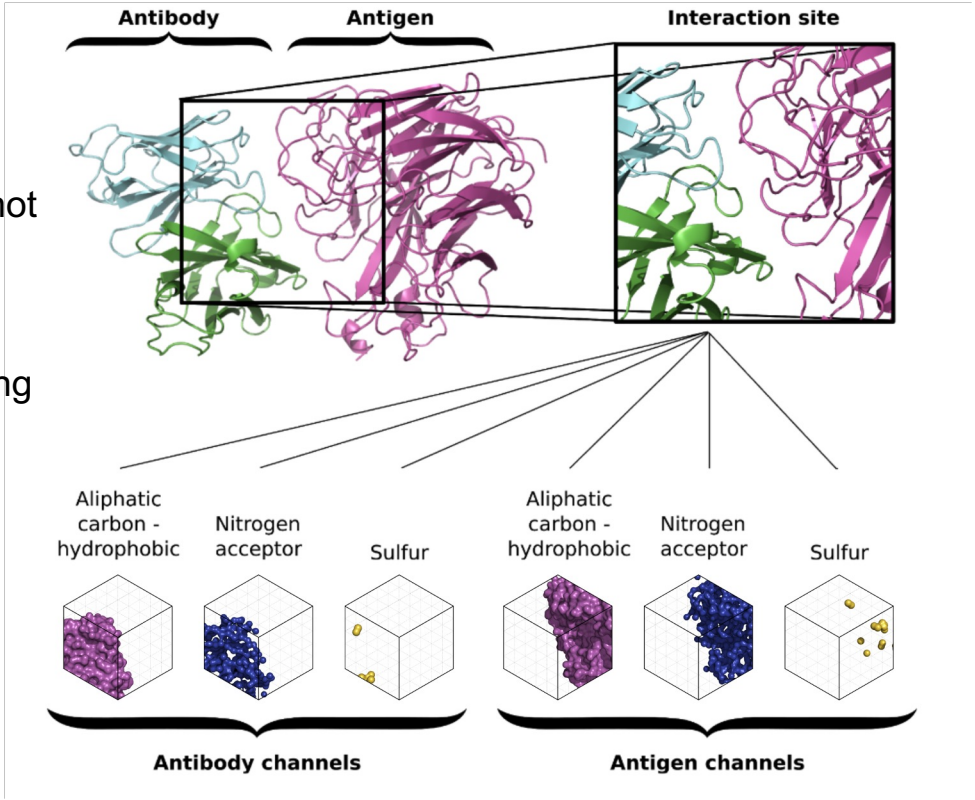
Often your only option if there are not a lot of other examples of antibodies that engage your antigen of interest

Antibody Virtual Screening - DLAB

If we have…
(a) a 3D model of an antibody, and
(b) a 3D structure of the antigen…
…we can train 3D CNNs to predict whether or not they bind one another

Requires a bound pose for the complex - use protein-protein docking to predict putative binding conformation

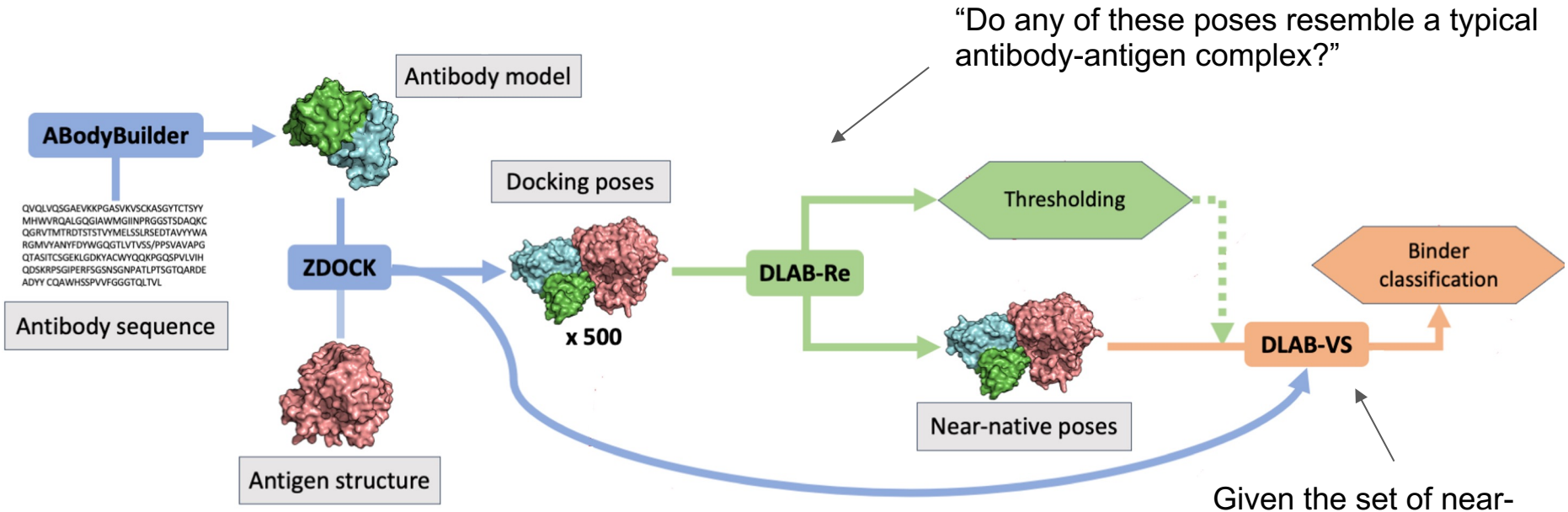
SAbDab



Schneider *et al.* (2022) Bioinformatics. https://doi.org/10.1093/bioinformatics/btab660

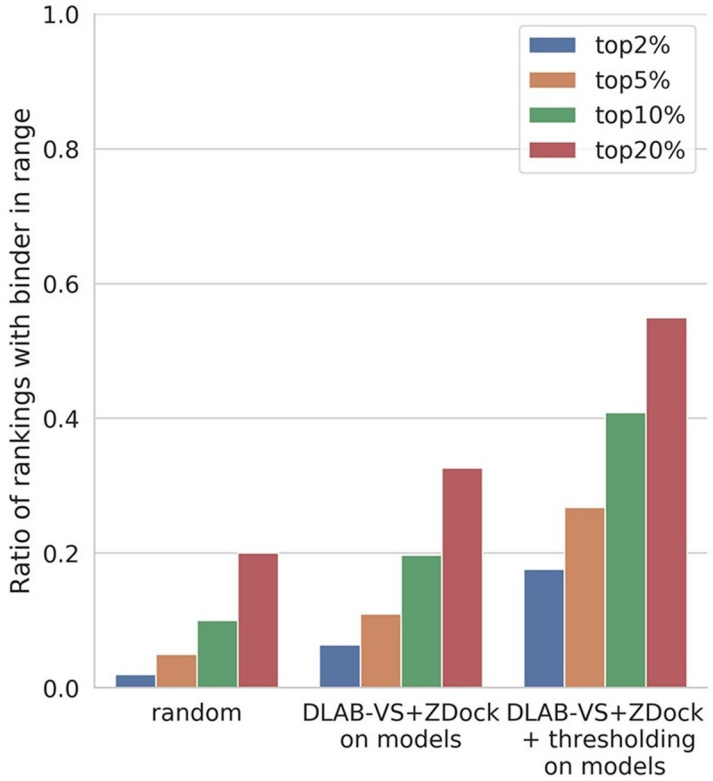
Antibody Virtual Screening - DLAB

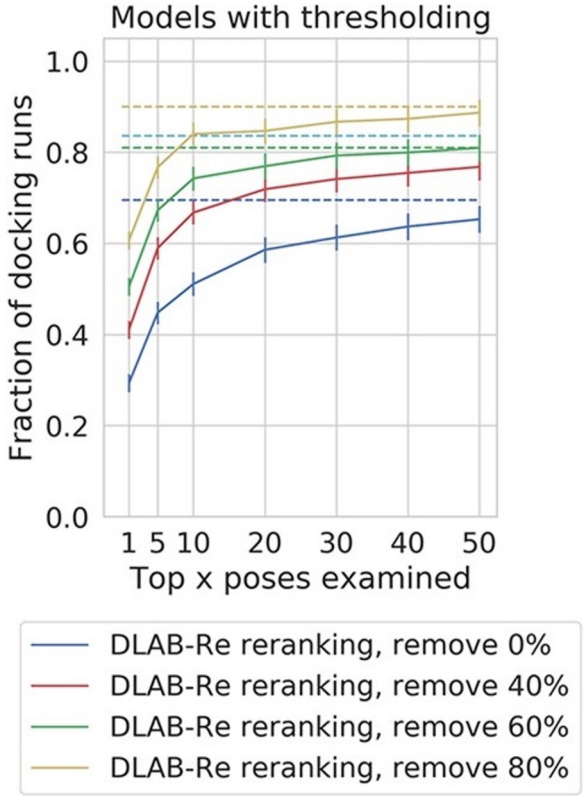
DLAB pipeline uses CNNs to accomplish both pose and binding prediction



Schneider *et al.* (2022) Bioinformatics. 10.1093/bioinformatics/btab660

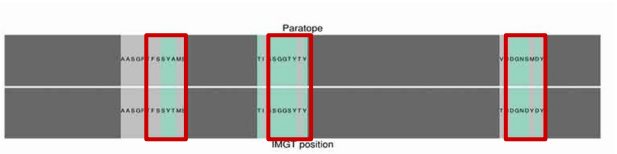
Antibody Virtual Screening - DLAB





Schneider *et al.* (2022) Bioinformatics. 10.1093/bioinformatics/btab660

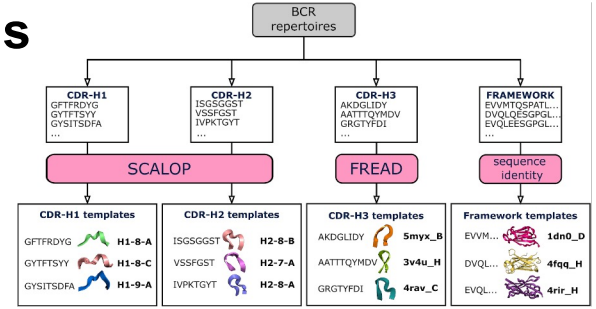
Increasing role for structure in repertoire analysis



Paratyping
(2021)

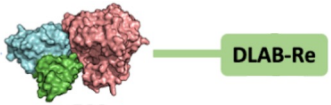
Sequence paratope clustering

Repertoire Structural Annotation + clustering



SAAB+
(2020)

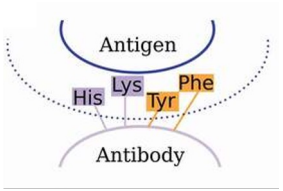
de novo antibody-antigen complementarity prediction



DLAB
(2021)

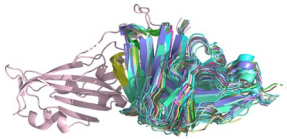
Antibody Repertoire Sequences

Spatial paratope clustering



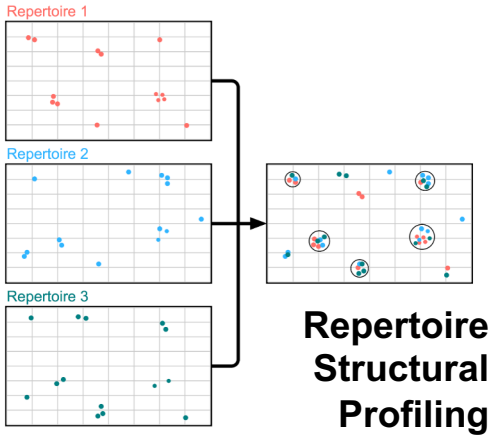
Ab-Ligity
(2021)

Epitope Profiling via structural modelling + clustering



SPACE
(2021)

Repertoire Structural Modelling + clustering



Repertoire Structural Profiling
(2021)

Recent review summarising the roles of these tools: Raybould *et al*. (2021) mAbs. 13(1):1996732

Software Availability

@oxpig

Paratyping, SAAB+, SCALOP, Ab-Ligity, DLAB

SAbBox

Full SAbDab and SAbPred

Virtual machine:
Academic Licence | Commercial Licence

Singularity container:
Academic Licence | Commercial Licence

Acknowledgements

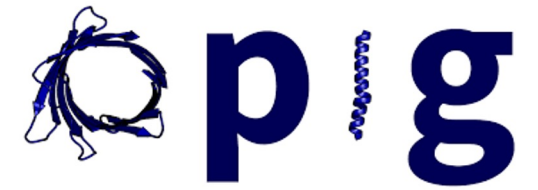
GSK
Alan Lewis
Matthew Bottomley
Newton Wahome
Ian Wall
UCB
Jiye Shi
James Snowden
Roche
Guy Georges
Alexander Bujotzek
AstraZeneca
Maria Flocco
Andrew Buchanan

Ex-members of OPIG,
in particular:
-Aleksandr Kovaltsuk [SAAB+]
-Wing Ki Wong [Ab-Ligity]
-Constantin Schneider [DLAB]
-Claire Marks [Several tools!]
-Jinwoo Leem [ABodyBuilder]

Current members of OPIG,
in particular:
-Eve Richardson [Paratyping]
-Brennan Abandes Kenyon [ABlooper]
-Sarah Robinson [SPACE]

kymab
NOVARTIS
NOVARTIS INSTITUTES FOR BIOMEDICAL RESEARCH
AstraZeneca
diamond
Boehringer Ingelheim
UKRI Medical Research Council
BenevolentAI
Roche
ucb
e-Therapeutics plc
gsk GlaxoSmithKline
UKRI Engineering and Physical Sciences Research Council
Exscientia
DRIVEN BY KNOWLEDGE
UKRI Biotechnology and Biological Sciences Research Council

Thanks for Listening!



opig@stats.ox.ac.uk