UNIVERSITY OF OXFORD

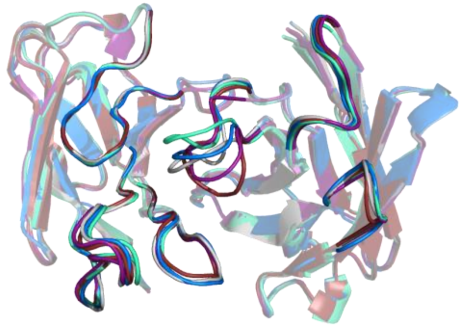
Computationally designing therapeutic antibodies - combining immune repertoire data and structural information

Charlotte Deane

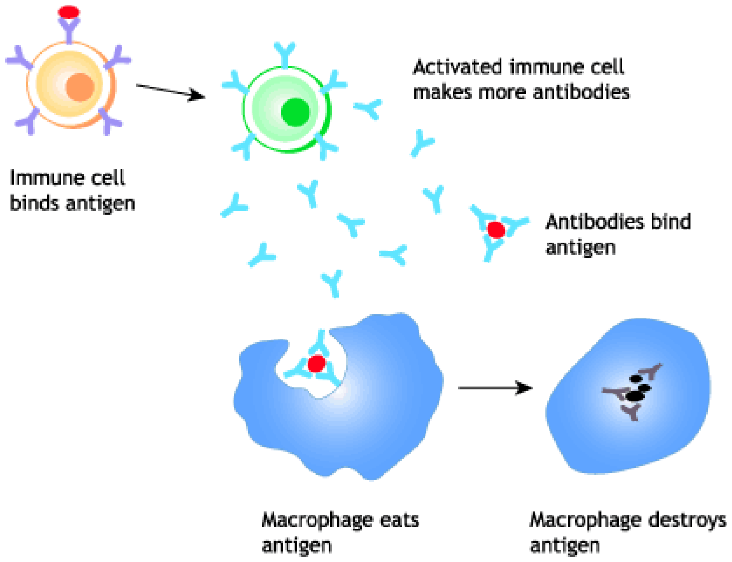
Department of Statistics

Oxford University

```
GGTCCCTGAGACTCTCCTGTGCAGCCTCT
GGATTCACCTTTGATGATTATGCCATGCAC
TGGGTCCGCCAAGCTCCAGGGAAGGGCT
GGAGTGGGTCTCAGGTACTAGTTGGAGTA
GTAGTTCCATAGGCTATGTGGACTCTGTGA
AGGGCCGATTCACCATCTCCAGAGACAAC
GCCAAGAACTCCCTGTATCTGCAAATGAAC
AGTCTGAGAGTTGAGGACACGGCCTTATAT
TACTGTGCAAAAGATGTTCTTAGCCGCAGC
TGGCGATATCTTGACCCCTGGGGCCATGGA
ACCCTGGTCACCGTCTCCTCAGCATCCCCG
ACCAGCCCCAAGGTCTTCC
```



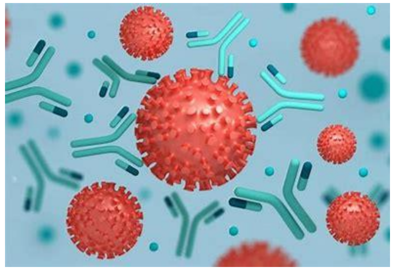
Antibodies



- It has been estimated that a typical human is capable of producing more than 10^{12} different antibodies, each capable of binding a distinct epitope

- Recognise and bind to potentially harmful molecules (antigens)

- Either inhibit the antigen themselves or recruit other parts of the immune system to deal with them



Covid-19

Vaccine efficacy
Diagnosing exposure
Effective biotherapeutics

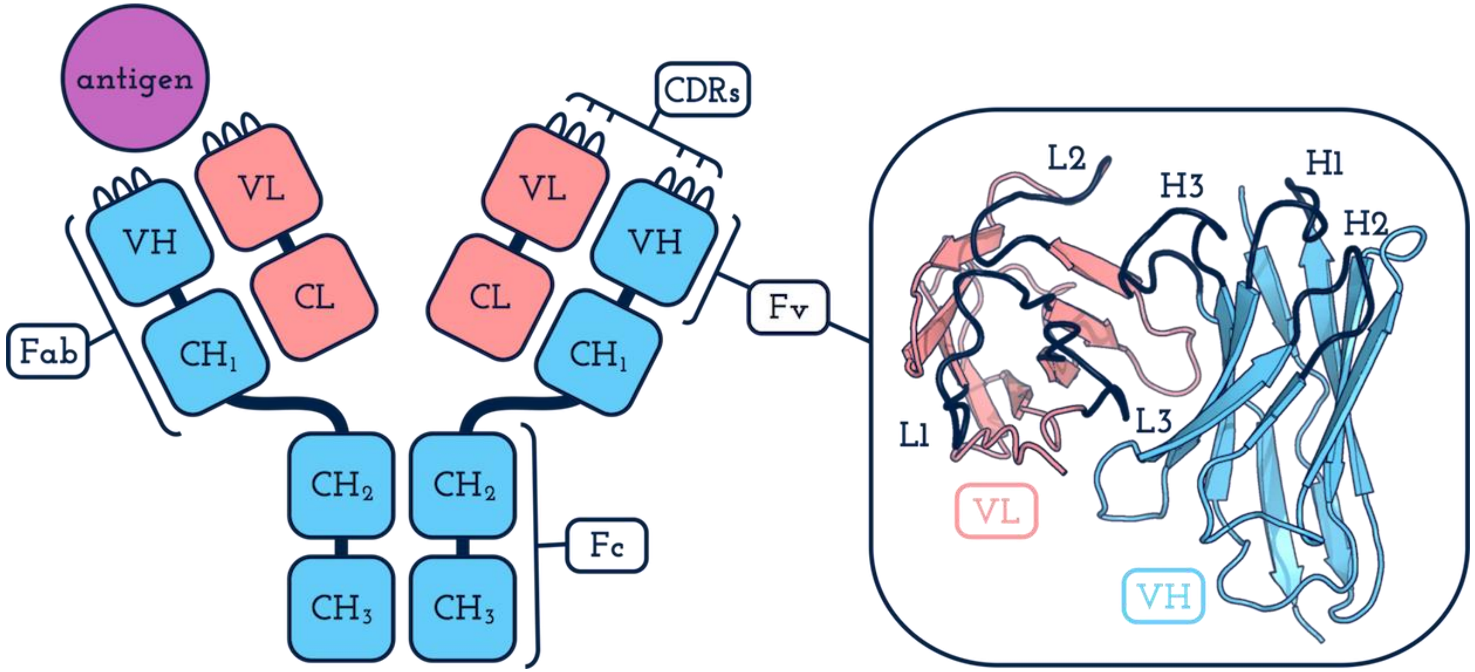
- Target specifically and with high affinity

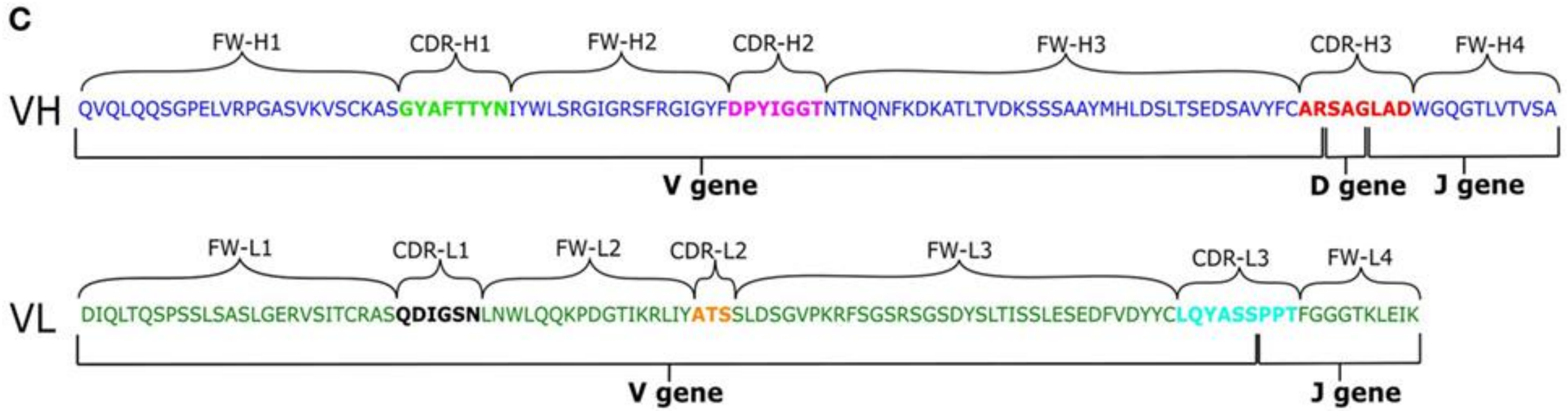
- Can be raised against almost any antigen

- Currently >100 approved antibody “drugs”

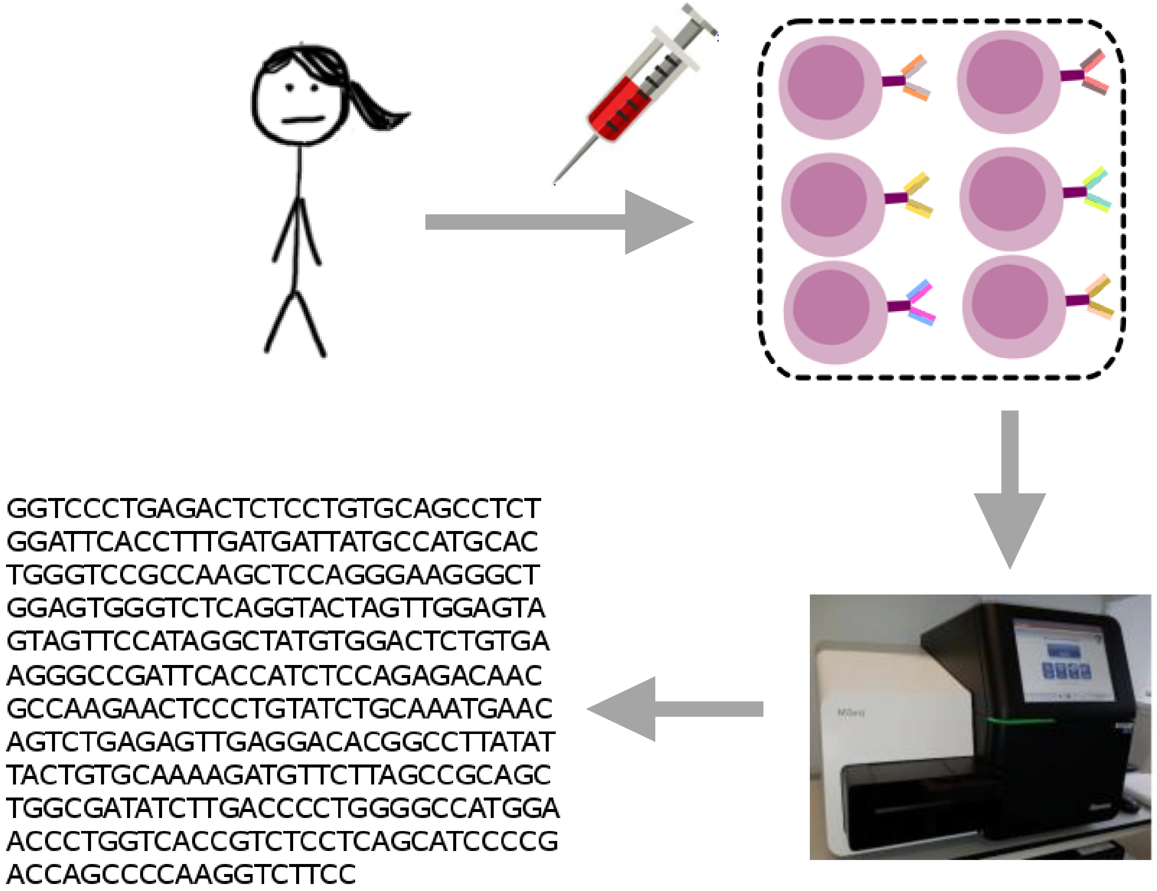
- Antibody-based therapeutics are entering clinical study at a rapid rate

Antibody Sequence/Structure - Orientation




C
FW-H1
CDR-H1
FW-H2
CDR-H2
FW-H3
CDR-H3
FW-H4
VH
QVQLQQSGPELVRPGASVKVSCKASGYAFTTYNIYWLSRGIGRSFRGIGYFDPYIGGTNTNQNFKDKATLTVDKSSSAAYMHLDSLTSEDSAVYFCARSAGLADWGQGTLVTVSA
V gene
D gene
J gene
FW-L1
CDR-L1
FW-L2
CDR-L2
FW-L3
CDR-L3
FW-L4
VL
DIQLTQSPSSLSASLGERVSITCRASQDIGSNLNWLQQKPDGTIKRLIYATSSLDSGVPKRFSGSRSGSDYSLTISSLESEDFVDYYCLQYASSPPTFGGGTKLEIK
V gene
J gene

Antibody Next-Generation Sequencing (immune repertoire sequencing)



```
GGTCCCTGAGACTCTCCTGTGCAGCCTCT
GGATTCACCTTTGATGATTATGCCATGCAC
TGGGTCCGCCAAGCTCCAGGGAAGGGCT
GGAGTGGGTCTCAGGTACTAGTTGGAGTA
GTAGTTCCATAGGCTATGTGGACTCTGTGA
AGGGCCGATTCACCATCTCCAGAGACAAC
GCCAAGAACTCCCTGTATCTGCAAATGAAC
AGTCTGAGAGTTGAGGACACGGCCTTATAT
TACTGTGCAAAAGATGTTCTTAGCCGCAGC
TGGCGATATCTTGACCCCTGGGGCCATGGA
ACCCTGGTCACCGTCTCCTCAGCATCCCCG
ACCAGCCCCAAGGTCTTCC
```

- Snapshots between 10^4 and 10^7 antibody sequences
- Theoretical antibody repertoire in humans $>10^{12} – 10^{15}$
- Circulating diversity ~ 10^9

- Naïve human antibody repertoire
- Pre and post immunisation datasets
- Sequences repertoires from different species

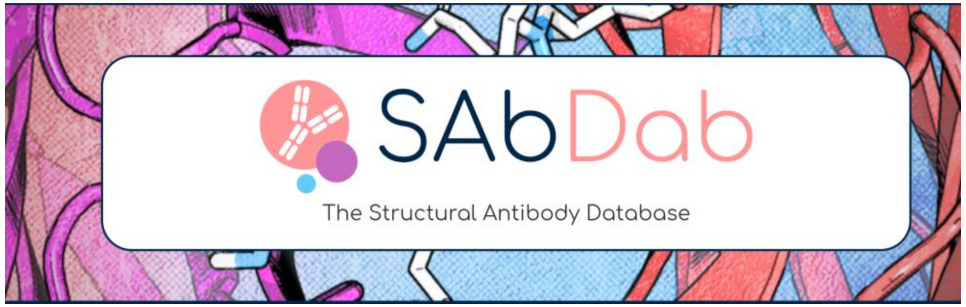
Olsen *et al* (2021), Kovaltsuk *et al*. (2018).

Observed Antibody Space

Over 80 BCR repertoire studies covering ~ 1.5 billion antibody sequences across diverse immune states, organisms and individuals.

Contains Paired and unpaired data

Sorted, cleaned, annotated, translated and numbered



Schneider *et al* (2021), Dunbar *et al.* (2014)

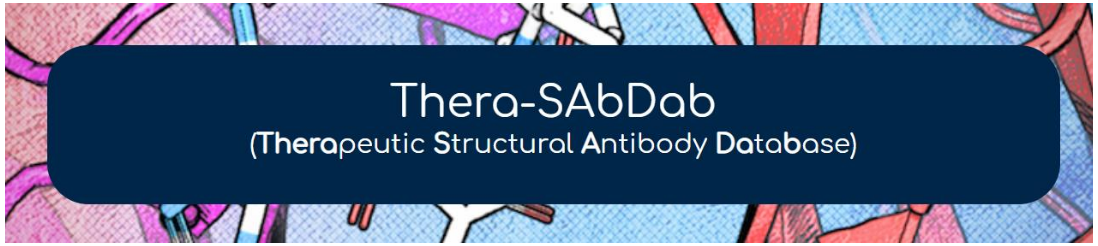
Structural Antibody Database

Fully automated updating collection of all publicly available antibody structure data

As of 31st October 2021 contains 5534 structures

4575 antibody antigen complexes

Collect, curate and present structural data consistently.
Contains antibodies and nanobodies (849)



Raybould *et al.* (2020)

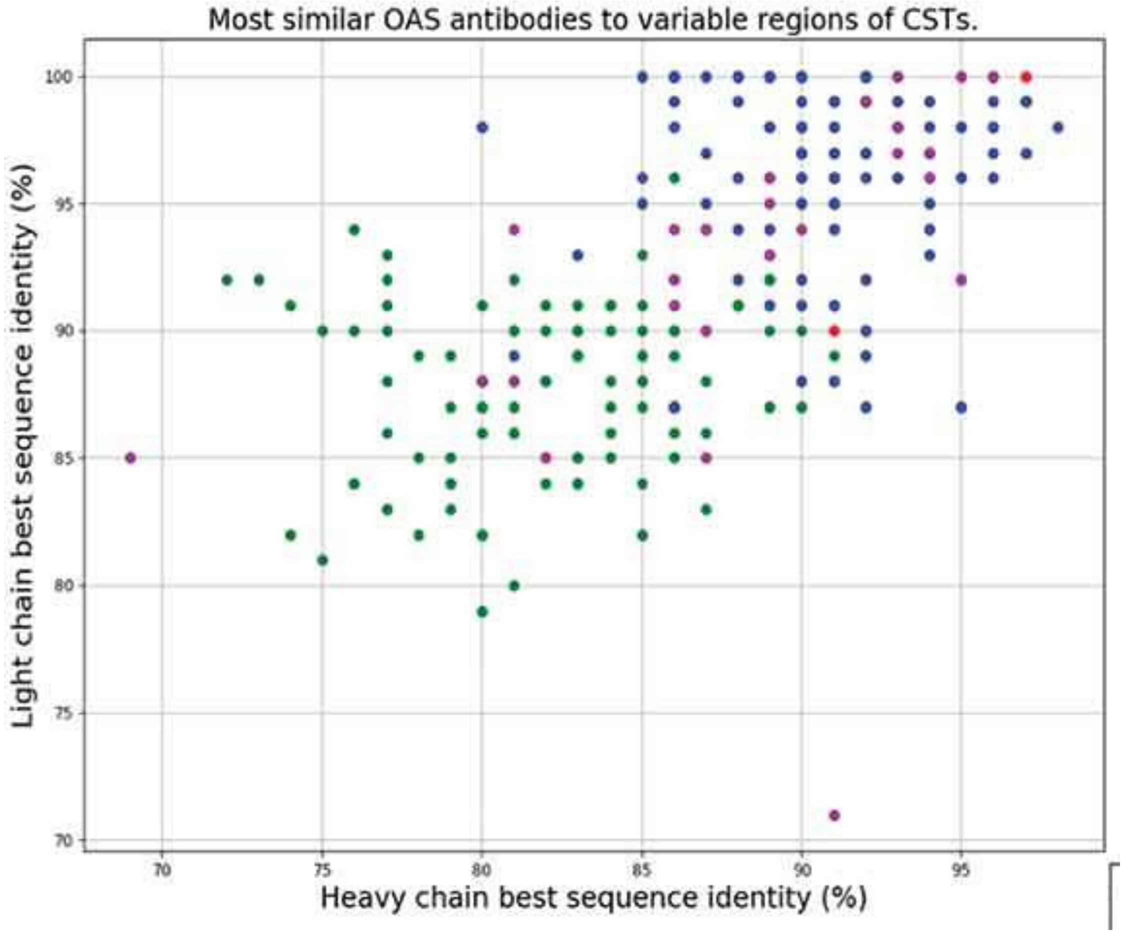
Thera-SAbDab

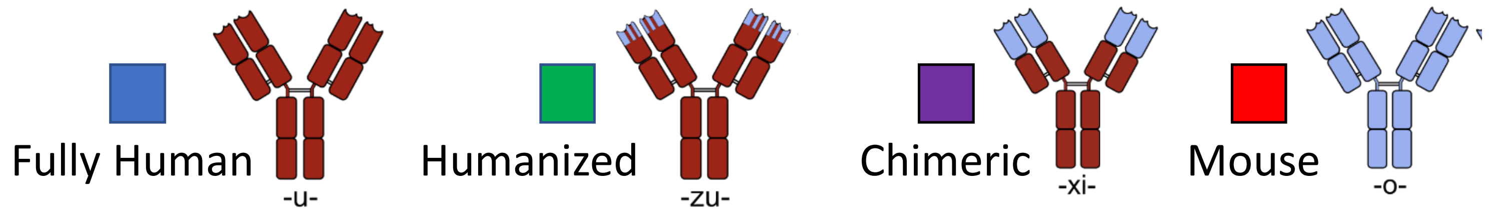
Self updating database of immunotherapeutic variable domain sequences and their corresponding structural representatives in SAbDab

Harvests therapeutic sequences as they are released by the World Health Organisation

October 2021 (696 entries)

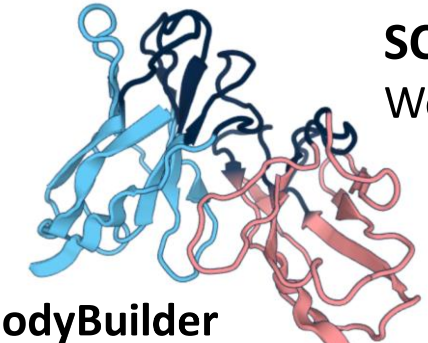
Looking for Therapeutic Antibodies in OAS





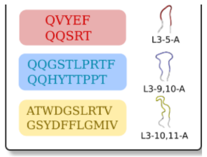
- 242 post phase-1 antibodies
- Unexpected high percentages of sequence overlap with therapeutics
- Many can be found in OAS with sequence identities >95%
- Enfortumab, heavy and light chain have 98% seqID
  - differences H38:N-S, H88:S-Y, L37:G-S, L52:F-L
- 54 have a perfect CDRH3 match
  - 22 of these found in more than one dataset

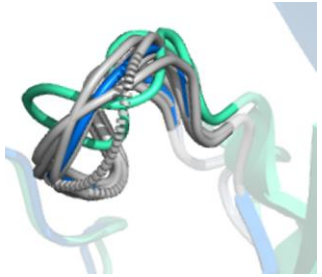
Krawczyk *et al*. (2019). *mAbs*



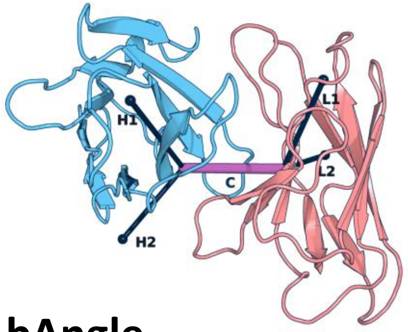
AbodyBuilder
Leem *et al* (2016)

SCALOP
Wong *et al.* (2018)

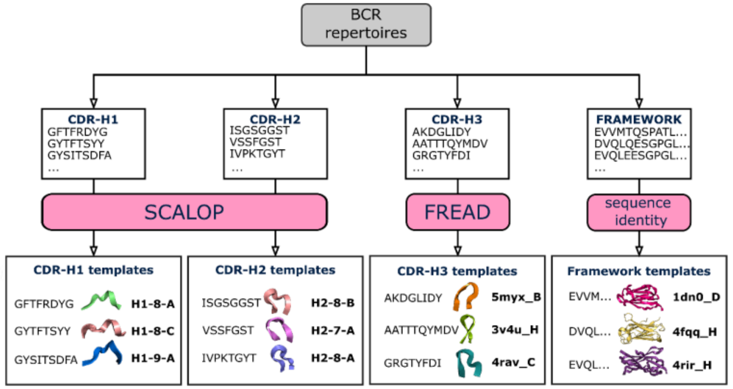




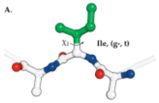
ABlooper
Abanades *et al.* (2021)

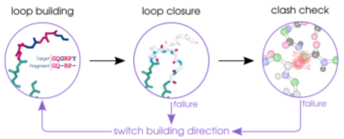


AbAngle
Dunbar *et al* (2013)



SAAB+
Kovaltsuk *et al.* (2020)

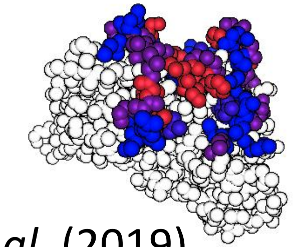




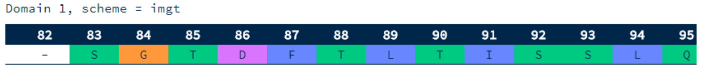
PEARS
Leem *et al* (2018)

SPHINX
Marks *et al* (2017)

SAbPred
The Antibody Prediction Toolbox

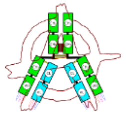


TAP
Raybould *et al.* (2019)

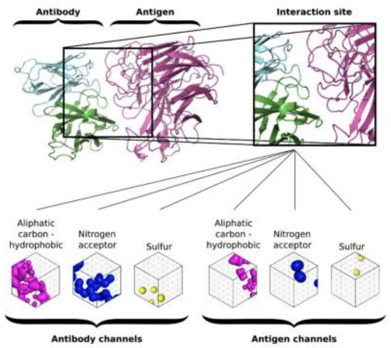


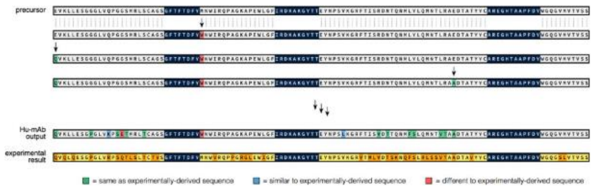
```
Domain 1, scheme = imgt
```

82	83	84	85	86	87	88	89	90	91	92	93	94	95
-	S	G	T	D	F	T	L	T	I	S	S	L	Q



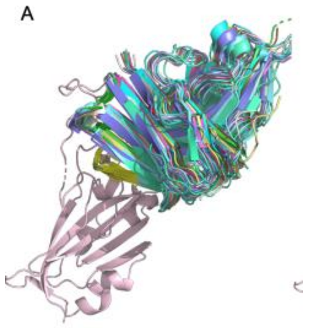
ANARCI
Dunbar *et al* (2016)



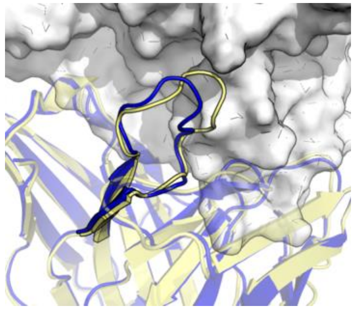

precursor
Hu-mAb output
experimental result
= same as experimentally-derived sequence
= similar to experimentally-derived sequence
= different to experimentally-derived sequence

Dlab
Schneider *et al* (2021)

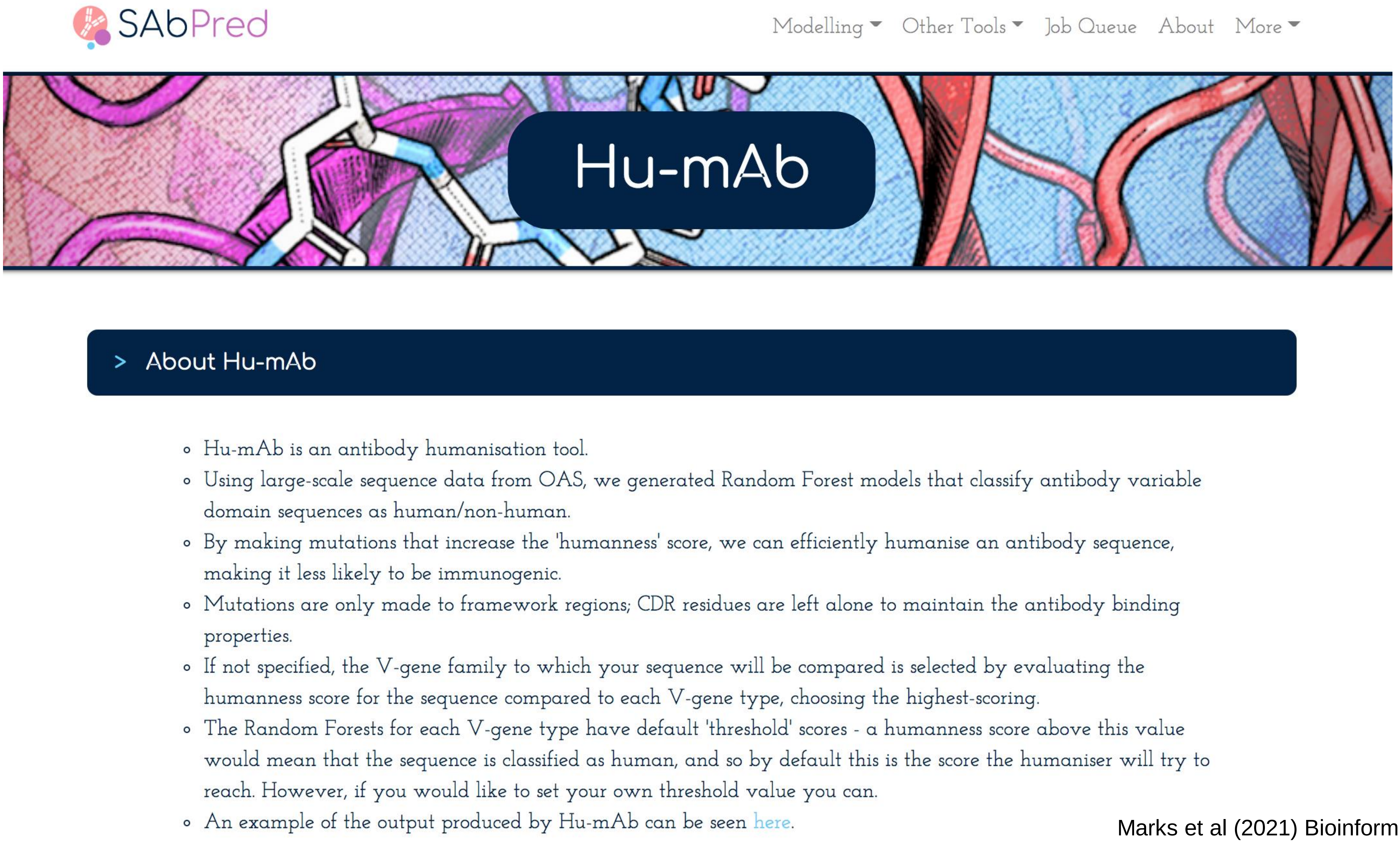
Hu-mAb
Marks *et al.* (2021)



Epitope Profiling - SPACE
Robinson *et al.* (2021)



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



Hu-mAb

> About Hu-mAb

- Hu-mAb is an antibody humanisation tool.
- Using large-scale sequence data from OAS, we generated Random Forest models that classify antibody variable domain sequences as human/non-human.
- By making mutations that increase the 'humanness' score, we can efficiently humanise an antibody sequence, making it less likely to be immunogenic.
- Mutations are only made to framework regions; CDR residues are left alone to maintain the antibody binding properties.
- If not specified, the V-gene family to which your sequence will be compared is selected by evaluating the humanness score for the sequence compared to each V-gene type, choosing the highest-scoring.
- The Random Forests for each V-gene type have default 'threshold' scores - a humanness score above this value would mean that the sequence is classified as human, and so by default this is the score the humaniser will try to reach. However, if you would like to set your own threshold value you can.
- An example of the output produced by Hu-mAb can be seen here.

Marks et al (2021) Bioinformatics

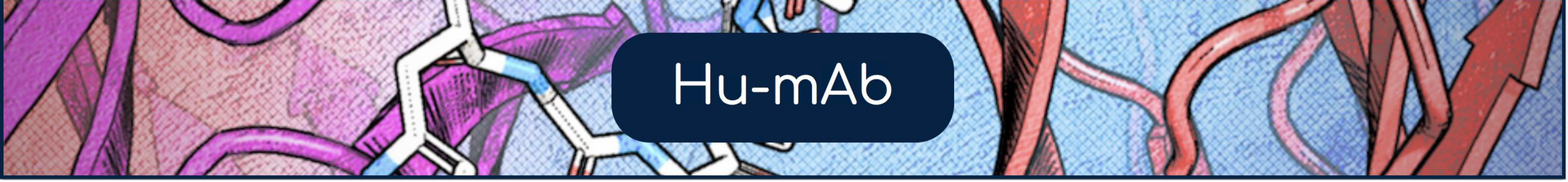
Humanization of antibodies using a machine learning approach Hu-mAb

- Many antibody therapeutics derived from non-human sources
  - ~50% of those currently in development

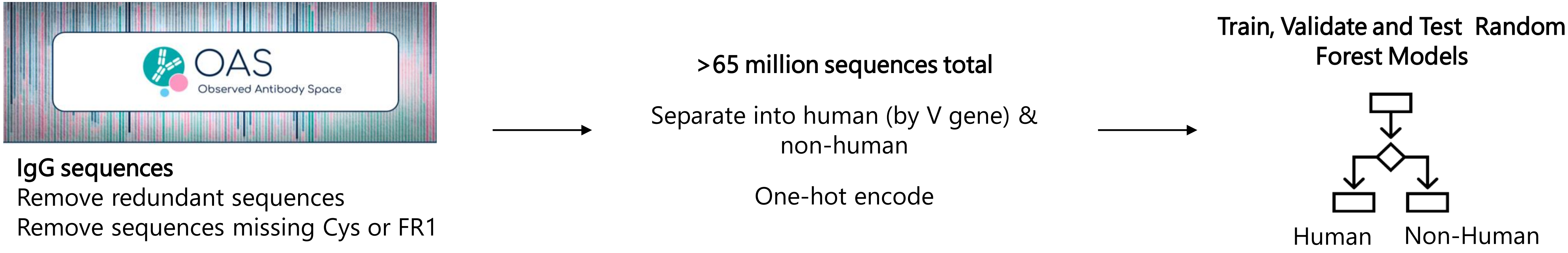
- ‘Non-human’ antibodies can result in a potentially harmful immune response in patients (immunogenicity)
  - Important to ‘humanize’ antibody therapeutics for safety & efficacy

- Currently, humanization is normally carried out experimentally, in a largely trial-and-error process.

Hu-mAb



Random Forest (RF) machine learning models built with >65 million human and non-human sequences from the OAS database



IgG sequences
Remove redundant sequences
Remove sequences missing Cys or FR1

>65 million sequences total

Separate into human (by V gene) & non-human

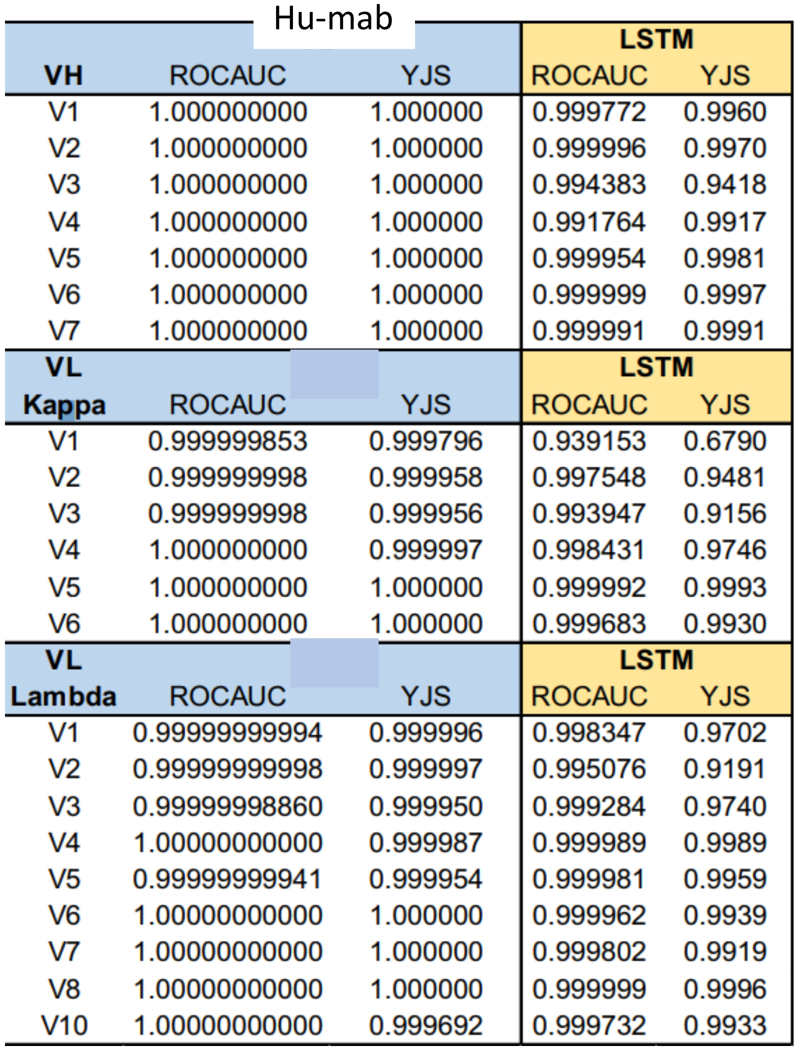
One-hot encode

Train, Validate and Test Random Forest Models

Human Non-Human

- Separate models for each human V gene type
- Human / non-human classification threshold set to maximize Youden’s J statistic (model performance)

Classification performance on OAS held out sets

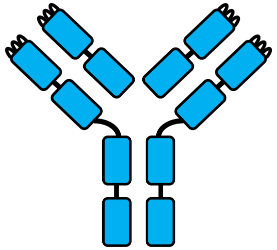


	Hu-mab		LSTM	
VH	ROCAUC	YJS	ROCAUC	YJS
V1	1.000000000	1.000000	0.999772	0.9960
V2	1.000000000	1.000000	0.999996	0.9970
V3	1.000000000	1.000000	0.994383	0.9418
V4	1.000000000	1.000000	0.991764	0.9917
V5	1.000000000	1.000000	0.999954	0.9981
V6	1.000000000	1.000000	0.999999	0.9997
V7	1.000000000	1.000000	0.999991	0.9991
VL			LSTM	
Kappa	ROCAUC	YJS	ROCAUC	YJS
V1	0.999999853	0.999796	0.939153	0.6790
V2	0.999999998	0.999958	0.997548	0.9481
V3	0.999999998	0.999956	0.993947	0.9156
V4	1.000000000	0.999997	0.998431	0.9746
V5	1.000000000	1.000000	0.999992	0.9993
V6	1.000000000	1.000000	0.999683	0.9930
VL			LSTM	
Lambda	ROCAUC	YJS	ROCAUC	YJS
V1	0.99999999994	0.999996	0.998347	0.9702
V2	0.99999999998	0.999997	0.995076	0.9191
V3	0.99999998860	0.999950	0.999284	0.9740
V4	1.00000000000	0.999987	0.999989	0.9989
V5	0.99999999941	0.999954	0.999981	0.9959
V6	1.00000000000	1.000000	0.999962	0.9939
V7	1.00000000000	1.000000	0.999802	0.9919
V8	1.00000000000	1.000000	0.999999	0.9996
V10	1.00000000000	0.999692	0.999732	0.9933

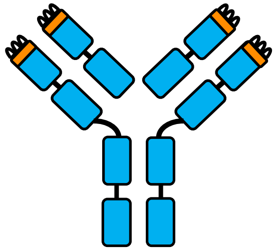
- Separate models for each HV, KV, LV genes
- Achieve very high ROC AUCs all over 0.99

- Hu-mAb outperforms previous LSTM in both AUC and YJS scores (Wollacott et al 2019)
- Also outperforms more recent humanness scorer BioPhi OASis (Prihoda et al 2021)

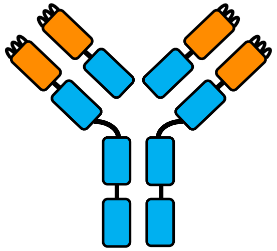
Testing Hu-mAb on known therapeutics



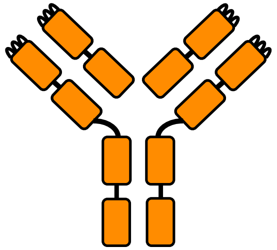
Human



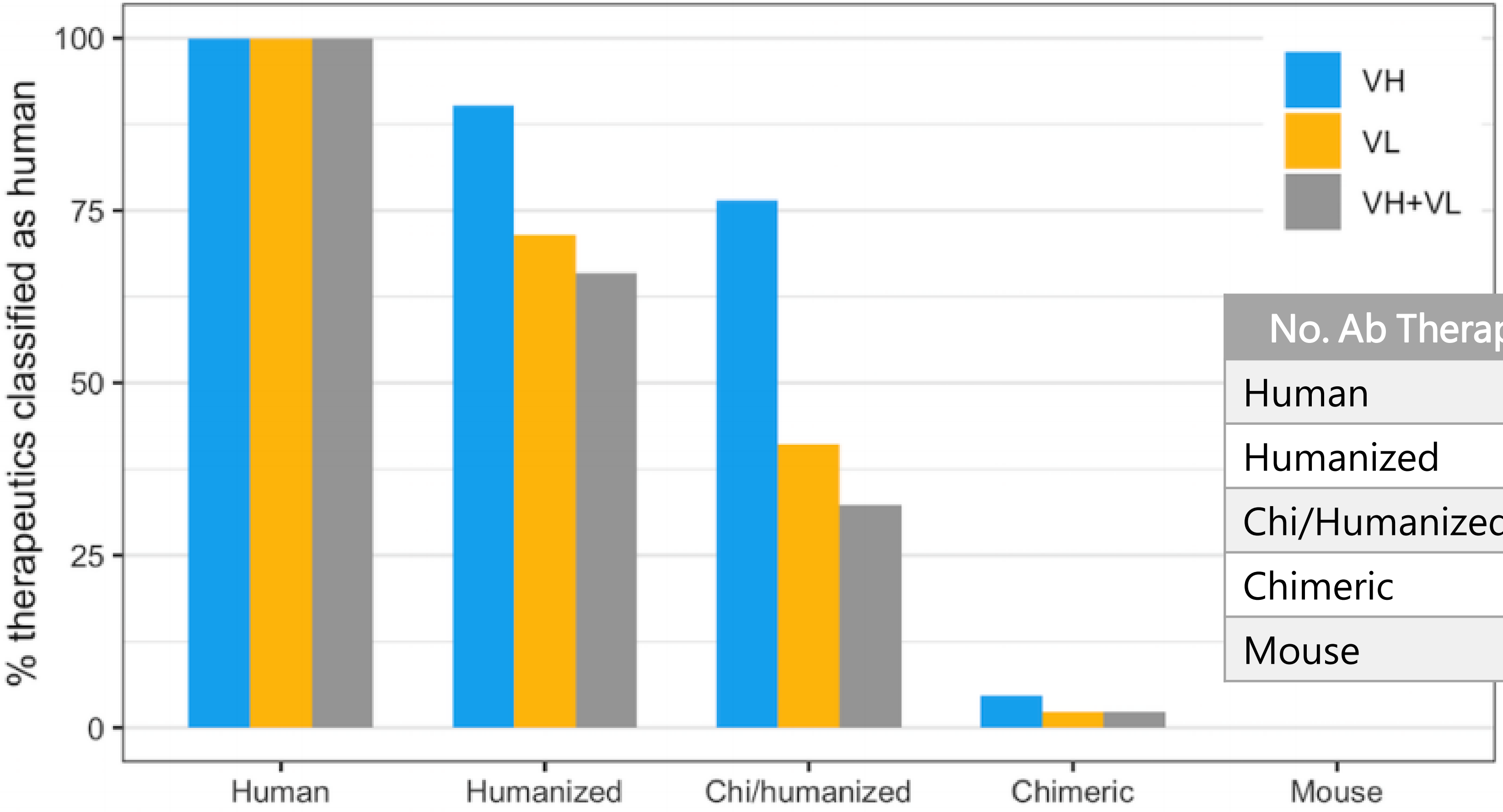
Humanized



Chimeric



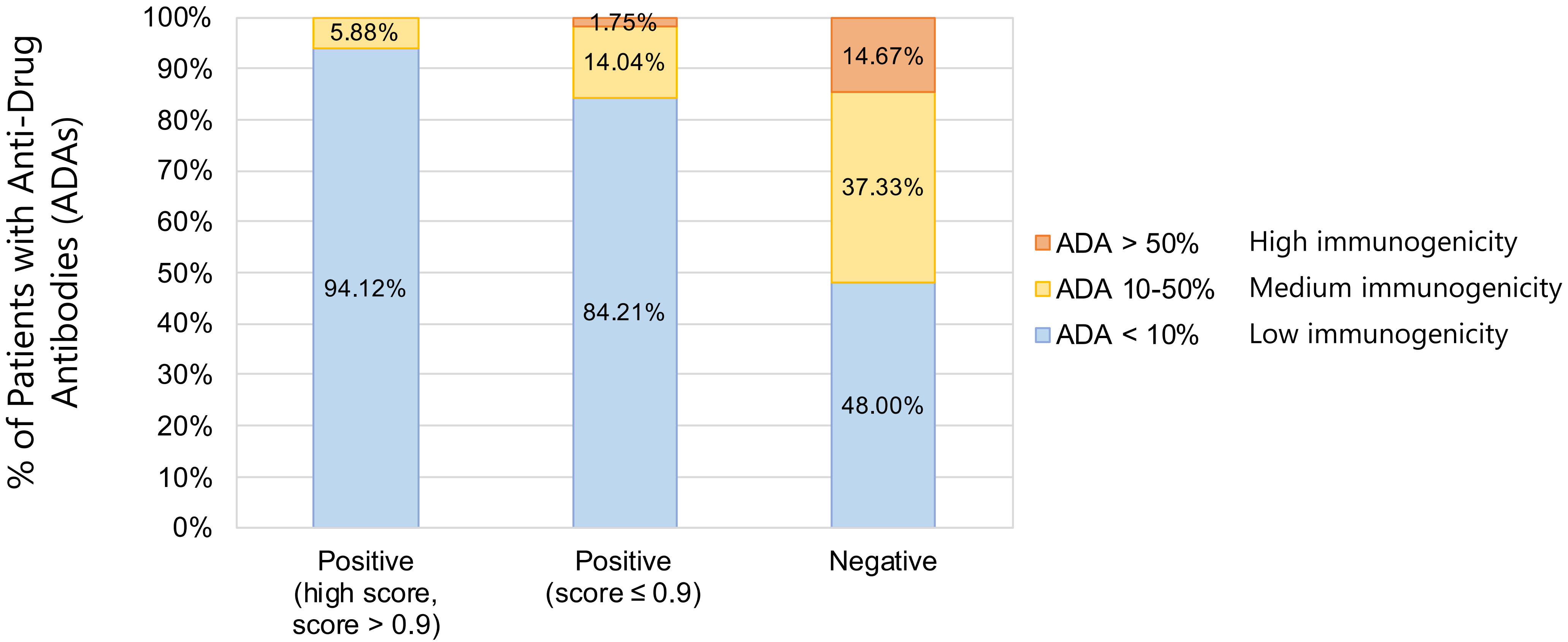
Mouse



No. Ab Therapeutics	
Human	176
Humanized	214
Chi/Humanized	34
Chimeric	43
Mouse	14

Tested on known therapeutics (Phase1-3 & Approved) from Thera-SabDab

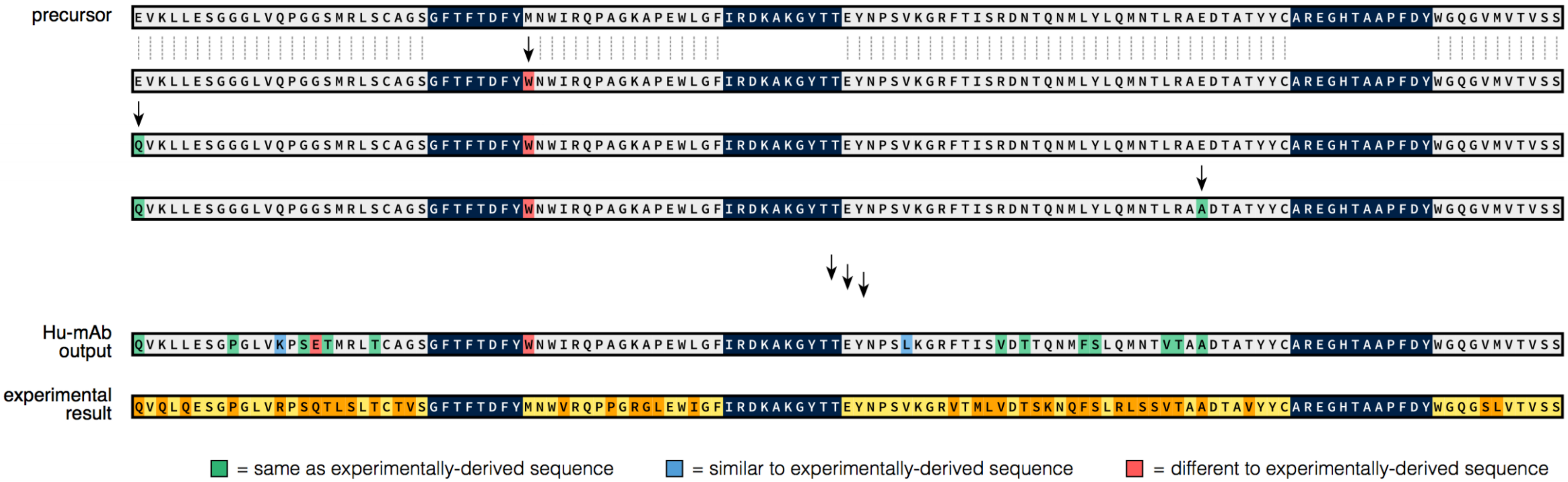
Relationship between Hu-mAb scores and experimental immunogenicity.



Therapeutic sequences classified as human by our model tend to have low immunogenicity levels, while sequences classified as not human are more immunogenic

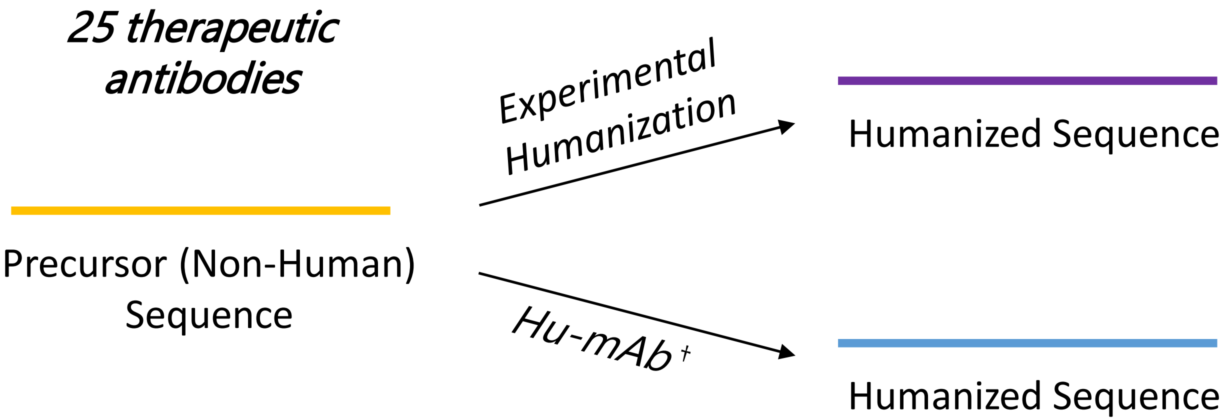
The Hu-mAb humanization procedure

- Computationally suggest the optimal mutations that would lower immunogenicity.


precursor
EVKLLESGGGLVQPGGSMRLSCAGSGFTFTDFYMNWIRQPAGKAPEWLGFIRDKAKGYTTEYNPSVKGRFTISRDNTQNMLYLQMNTLRAEDTATYYCAREGHTAAPFDYWGQGVMVTVSS
EVKLLESGGGLVQPGGSMRLSCAGSGFTFTDFYWNWIRQPAGKAPEWLGFIRDKAKGYTTEYNPSVKGRFTISRDNTQNMLYLQMNTLRAEDTATYYCAREGHTAAPFDYWGQGVMVTVSS
QVKLLESGGGLVQPGGSMRLSCAGSGFTFTDFYWNWIRQPAGKAPEWLGFIRDKAKGYTTEYNPSVKGRFTISRDNTQNMLYLQMNTLRAEDTATYYCAREGHTAAPFDYWGQGVMVTVSS
QVKLLESGGGLVQPGGSMRLSCAGSGFTFTDFYWNWIRQPAGKAPEWLGFIRDKAKGYTTEYNPSVKGRFTISRDNTQNMLYLQMNTLRAADTATYYCAREGHTAAPFDYWGQGVMVTVSS
Hu-mAb output
QVKLLESGPGLVKPSETMRLTCAGSGFTFTDFYWNWIRQPAGKAPEWLGFIRDKAKGYTTEYNPSLKGRFTISVDTTQNMFSLQMNTVTAADTATYYCAREGHTAAPFDYWGQGVMVTVSS
experimental result
QVQLQESGPGLVRPSQTLSLTCTVSGFTFTDFYMNWVRQPPGRGLEWIGFIRDKAKGYTTEYNPSVKGRVTMLVDTSKNQFSLRLSSVTAADTAVYYCAREGHTAAPFDYWGQGSLVTVSS
🟩 = same as experimentally-derived sequence
🟦 = similar to experimentally-derived sequence
🟥 = different to experimentally-derived sequence

- Tested Hu-Mab on 25 humanized sequences that demonstrated low immunogenicity and for which the precursor sequences were available.
  - Precursors were of murine, rat or rabbit origin

Evaluation of Humanization by Hu-mAb



25 therapeutic antibodies

Precursor (Non-Human) Sequence

Experimental Humanization

Humanized Sequence

Hu-mAb†

Humanized Sequence

77/85%

of mutations suggested are similar to those made experimentally

VH/VL

59/58%

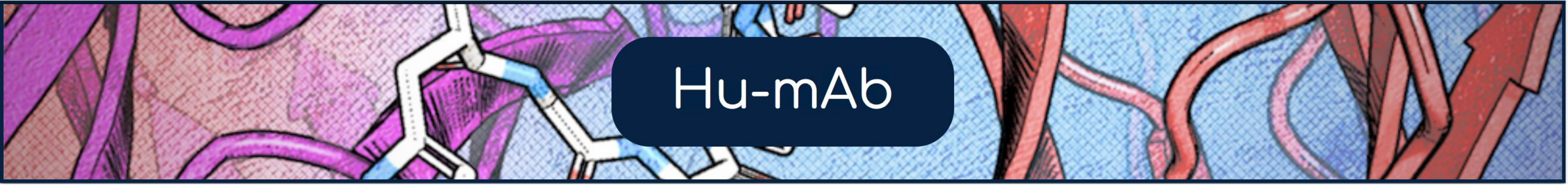
as many suggested by Hu-mAb, as experimentally

Comparison of Hu-mAb results with experimental humanization demonstrates **good agreement** but **greater efficiency** –

Hu-mAb proposes fewer mutations to the VH-VL interface making the orientation and therefore binding properties more likely to be preserved.

-> greater likelihood of preserving antibody structure & function

† Humanness threshold set to Hu-mAb humanness score of the experimentally humanized sequence



Hu-mAb

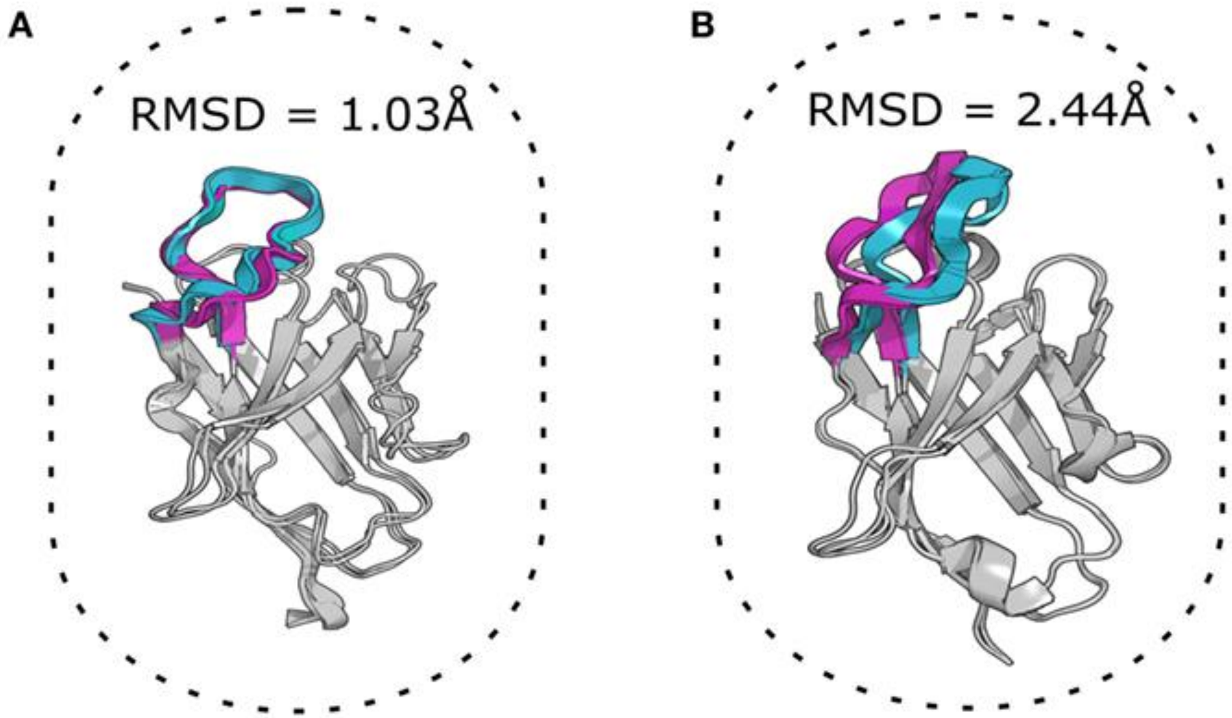
- Accurately evaluate whether an antibody is ‘human’ or not (humanness)
- Predict whether an antibody is immunogenic
- Be used to improve the humanness of a sequence

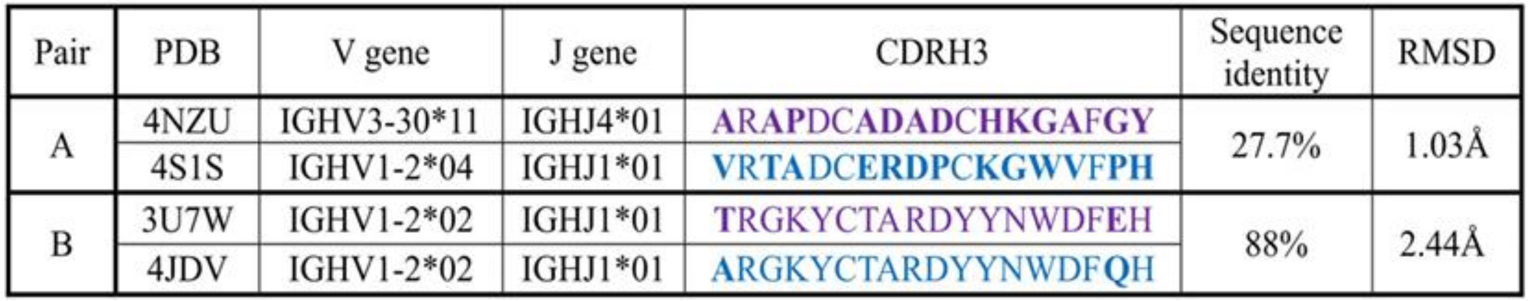
Available as a webserver at: opig.stats.ox.ac.uk/webapps/humab

Marks, C., Hummer, A.M., Chin, M. and Deane, C.M., *Bioinformatics*, 2021

Structurally annotating Immune repertoire data

Similar structure/similar sequence

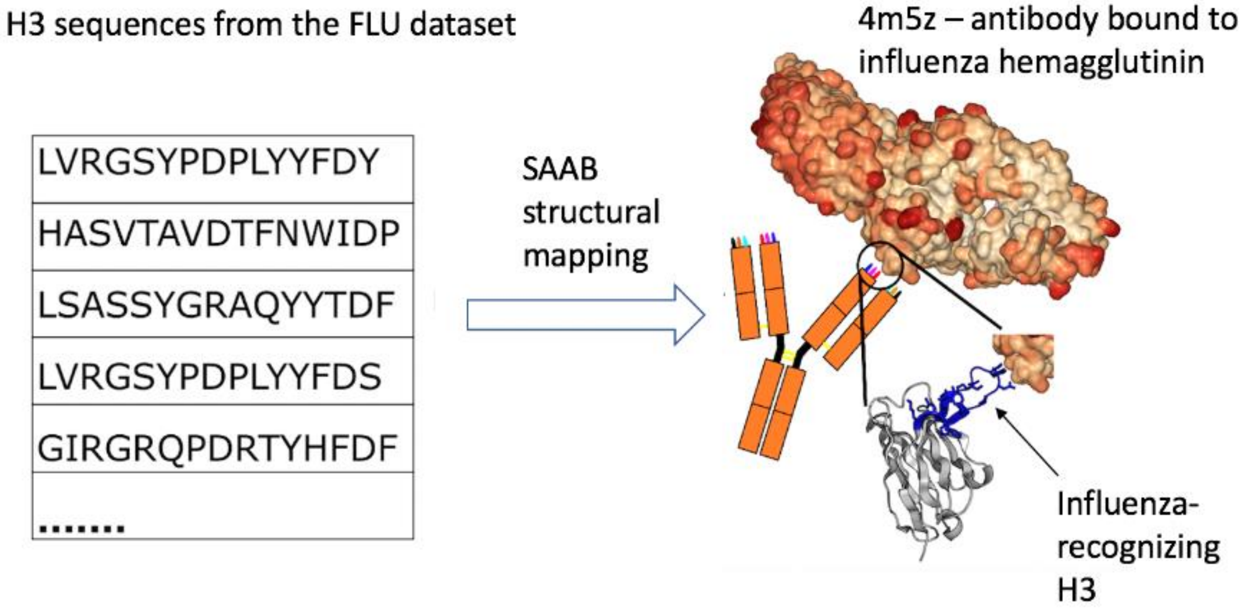




Pair	PDB	V gene	J gene	CDRH3	Sequence identity	RMSD
A	4NZU	IGHV3-30*11	IGHJ4*01	**A**R**AP**DC**ADAD**C**HKGA**F**GY**	27.7%	1.03Å
	4S1S	IGHV1-2*04	IGHJ1*01	**V**R**TA**DC**ERDP**C**KGWV**F**PH**		
B	3U7W	IGHV1-2*02	IGHJ1*01	**T**RGKYCTARDYYNWDF**E**H	88%	2.44Å
	4JDV	IGHV1-2*02	IGHJ1*01	**A**RGKYCTARDYYNWDF**Q**H		

Kovaltsuk *et al*. (2017). *Front. Immunol.*

Structural information on BCR repertoire antigen specificity



H3 sequences from the FLU dataset

LVRGSYPDPLYYFDY
HASVTAVDTFNWIDP
LSASSYGRAQYYTDF
LVRGSYPDPLYYFDS
GIRGRQPDRTYHFDF
....

SAAB structural mapping

4m5z – antibody bound to influenza hemagglutinin

Influenza-recognizing H3

Example: Post FLU challenge Ig-seq
Many H3 sequences (>7000) are structurally the same as those in 4m5z – complex of an antibody with influenza hemagglutinin

The similarity of these H3s could not be identified by sequence alone

Krawczyk *et al* (2018) *Front. Immunol*

The Therapeutic Antibody Profiler (TAP)
Five Computational Developability Guidelines

- Therapeutic antibodies must not only bind to their target but must also be free from ‘developability issues’ such as poor stability or high levels of aggregation.

- TAP is an *in-silico* antibody design analog of the Lipinski’s rule of five for small molecules
    - to guide the selection of antibodies with appropriate biophysical properties

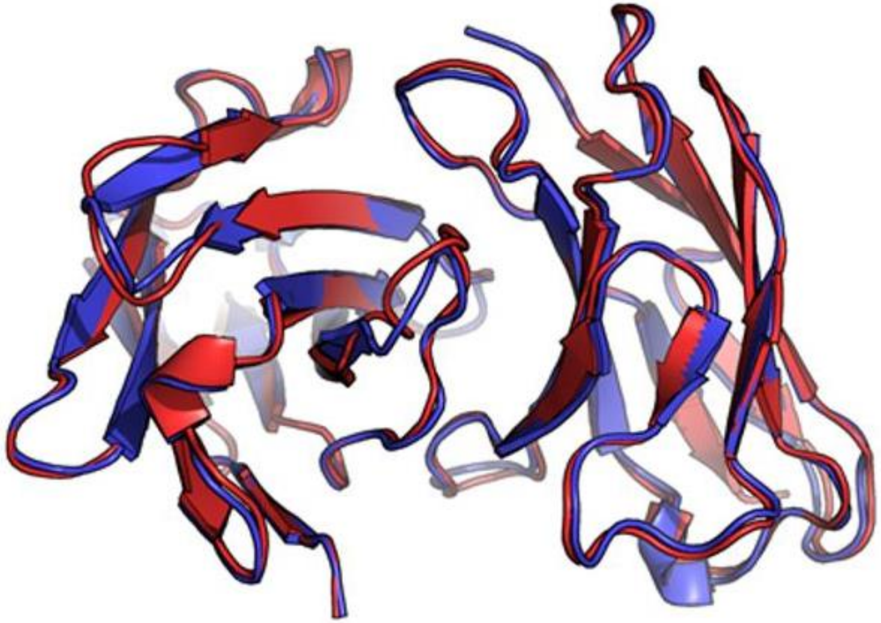
- Derive distributions of metrics for clinical stage therapeutics and assume that these indicate the allowed values of these properties.
    - Calculate these metrics on models so can run against potential therapeutics where crystal structures are unavailable

- These metrics don’t have to correlate with a particular experiment that tests for developability rather they indicate that a potential therapeutic has outlying values.

Raybould et al (2019) PNAS

Datasets – structural models

- Models of the variable domain structures of 137 post-Phase I clinical-stage antibody therapeutics (CSTs)*
    - Models are accurate enough for our metrics (tested with the 56 CSTs with known structure)
    - Average RMSD of framework < 1A
    - Less than 4% of residues are wrongly annotated exposed/buried



*Jain et al (2017) PNAS

Five properties:

1. CDRH3 or Total CDR length [aggregation, flexibility, topology]
2. Patches of Surface Hydrophobicity (PSH) across the CDR Vicinity [aggregation, viscosity]
3. Patches of Surface Positive Charge (PPC) across the CDR Vicinity [poor expression, aggregation, viscosity, polyspecificity]
4. Patches of Surface Negative Charge (PNC) across the CDR Vicinity [poor expression, aggregation, viscosity, polyspecificity]
5. Structural Fv Charge Symmetry Parameter [aggregation, viscosity]

Datasets:

137 Post-Phase I Therapeutic Models

Sets the **acceptable bounds** of the five properties

14k Representative Human Antibody Models[2,3]

Provides a "**natural antibody comparison**"

2 Datasets of MedImmune Developability Failures

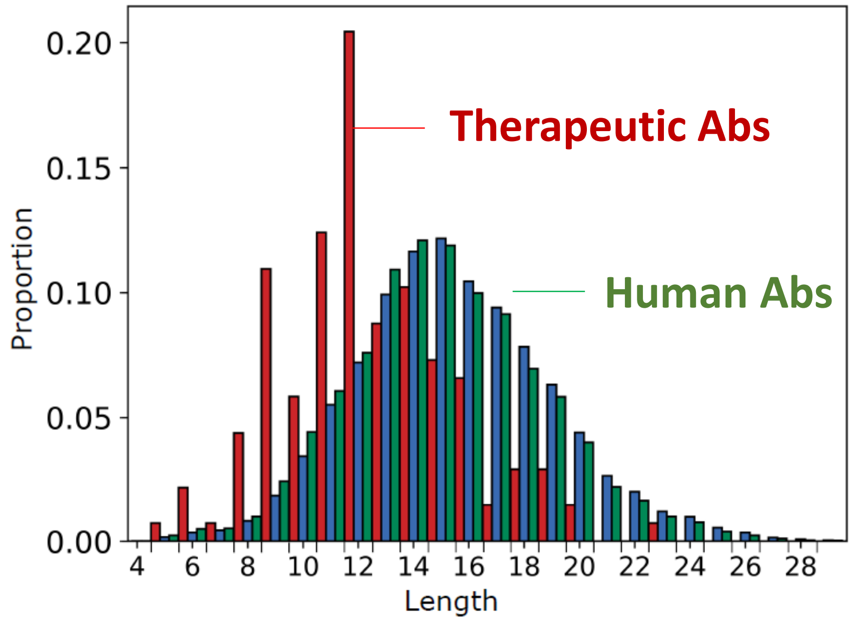
Used to **validate** that we can selectively highlight mAbs with developability issues

[2] Vander Heiden JA, *et al*. (2017) Dysregulation of B cell repertoire formation in myasthenia gravis patients revealed through deep sequencing. *J. Immunol*. 198:1460–1473.

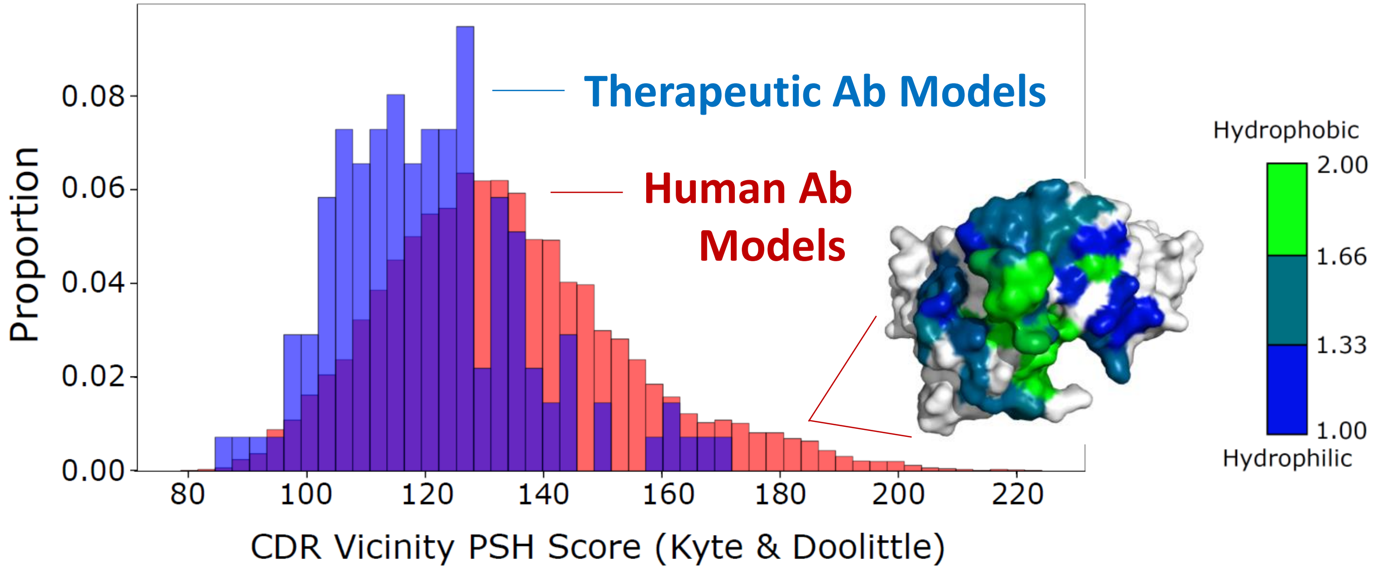
[3] Raybould, MIJ *et al*. (2019) Five computational developability guidelines for therapeutic antibody profiling. *Proc Natl Acad Sci USA* 116(10):4025-4030.

Comparisons: Therapeutics *vs.* Human Antibodies

CDRH3 Length



Patches of Surface Hydrophobicity (PSH)

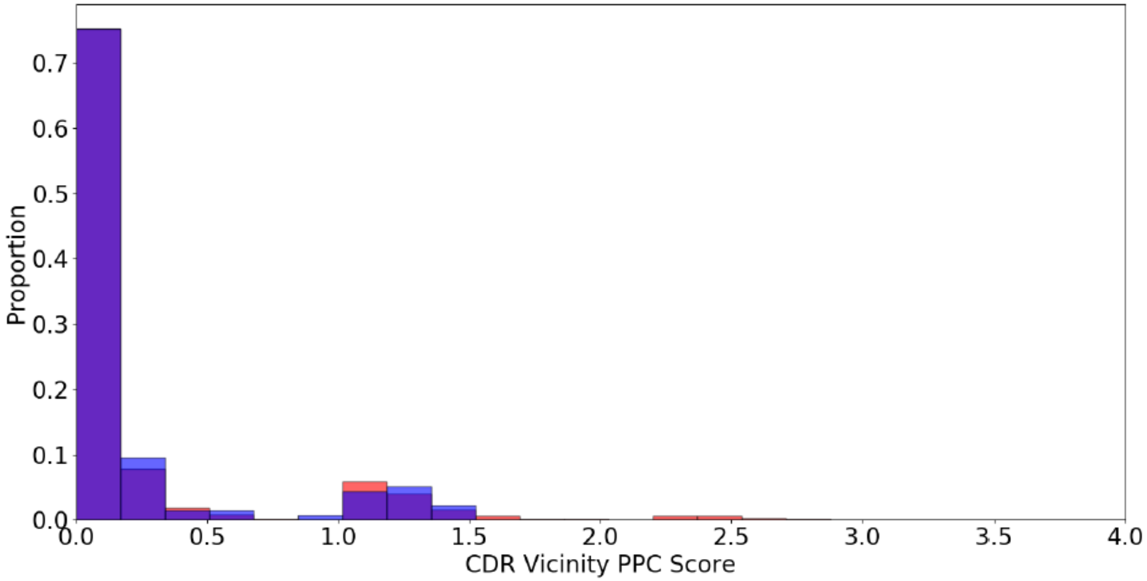


$$\sum_{R1\,R2} \frac{H(R1,S)\,H(R2,S)}{r_{12}^{2}}$$

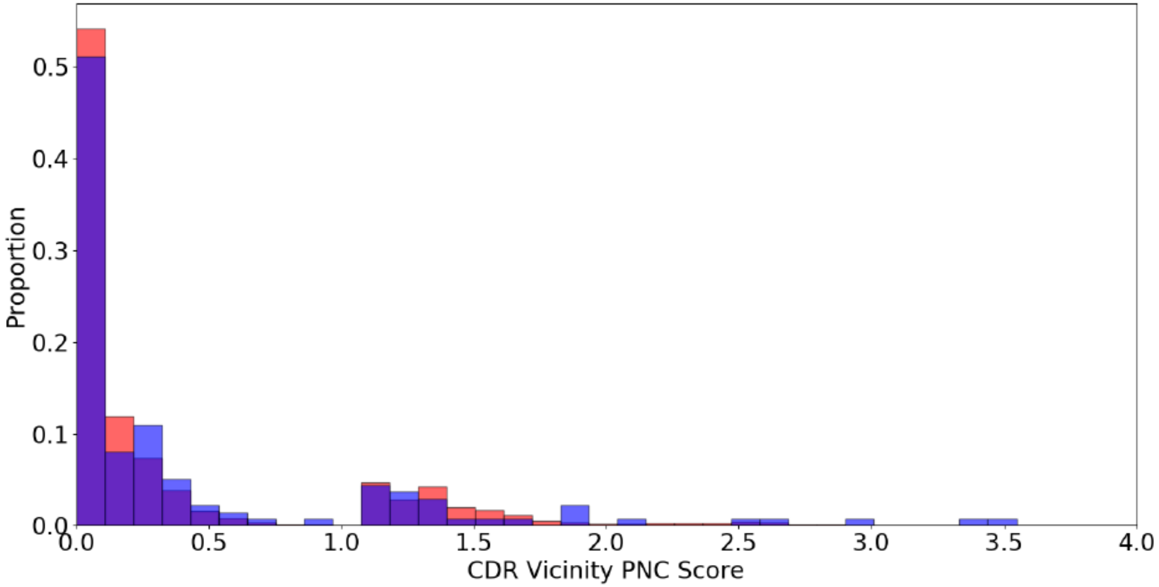
- Therapeutics tend to have shorter CDRH3s and smaller patches of surface hydrophobicity than human antibodies

Comparisons: Therapeutics *vs.* Human Antibodies

Patches of Surface Positive Charge (PPC)



Patches of Surface Negative Charge (PNC)



$$\sum_{R1\,R2} \frac{|Q(R1)|\,|Q(R2)|}{r_{12}^2}$$

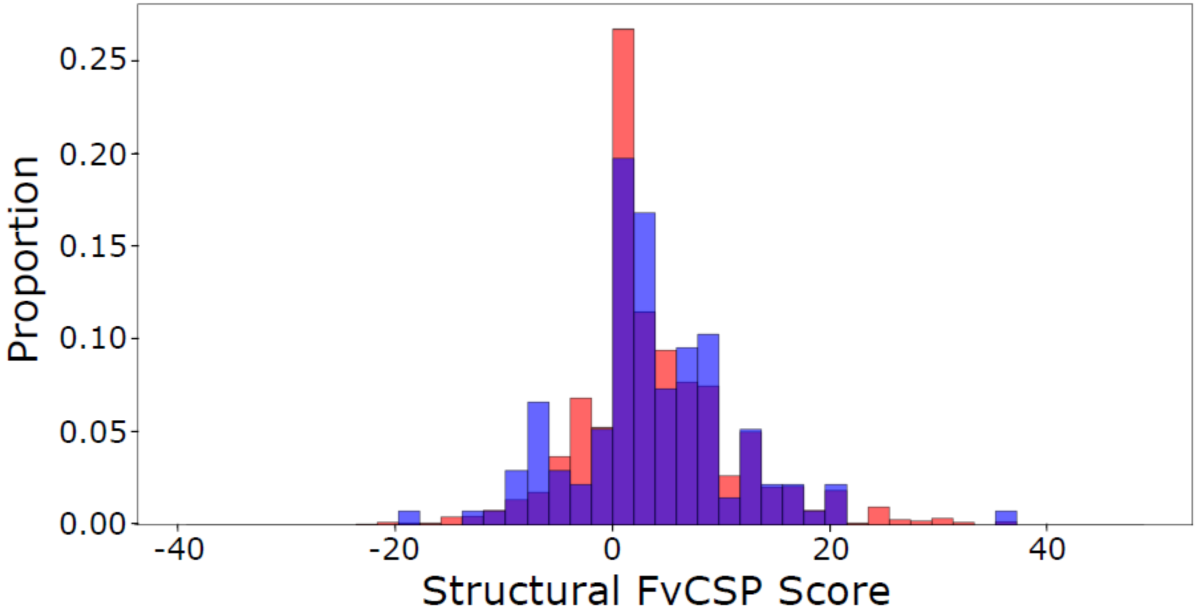
Blue: Therapeutic Antibody Models

Red: Human Antibody Models

- Therapeutics and human Abs have similar sizes of positive charge and negative charge patches

Comparisons: Therapeutics *vs*. Human Antibodies

Structural Fv Charge Symmetry Parameter (SFvCSP)



$$\left[\sum_{R_H} Q(R_H)\right]\left[\sum_{R_L} Q(R_L)\right]$$

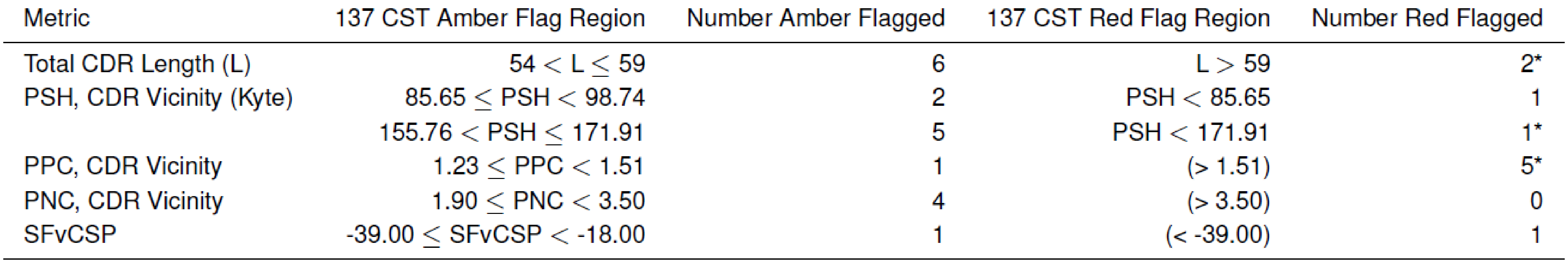
Blue: Therapeutic Antibody Models

Red: Human Antibody Models

- Both therapeutic and human antibodies have an aversion to strongly oppositely-charged VH and VL chains

Validation: Things TAP shouldn’t flag

- Tested against 105 extra post-Phase I therapeutics



Metric	137 CST Amber Flag Region	Number Amber Flagged	137 CST Red Flag Region	Number Red Flagged
Total CDR Length (L)	$54 < L \leq 59$	6	$L > 59$	2*
PSH, CDR Vicinity (Kyte)	$85.65 \leq PSH < 98.74$	2	$PSH < 85.65$	1
	$155.76 < PSH \leq 171.91$	5	$PSH < 171.91$	1*
PPC, CDR Vicinity	$1.23 \leq PPC < 1.51$	1	(> 1.51)	5*
PNC, CDR Vicinity	$1.90 \leq PNC < 3.50$	4	(> 3.50)	0
SFvCSP	$-39.00 \leq SFvCSP < -18.00$	1	(< -39.00)	1

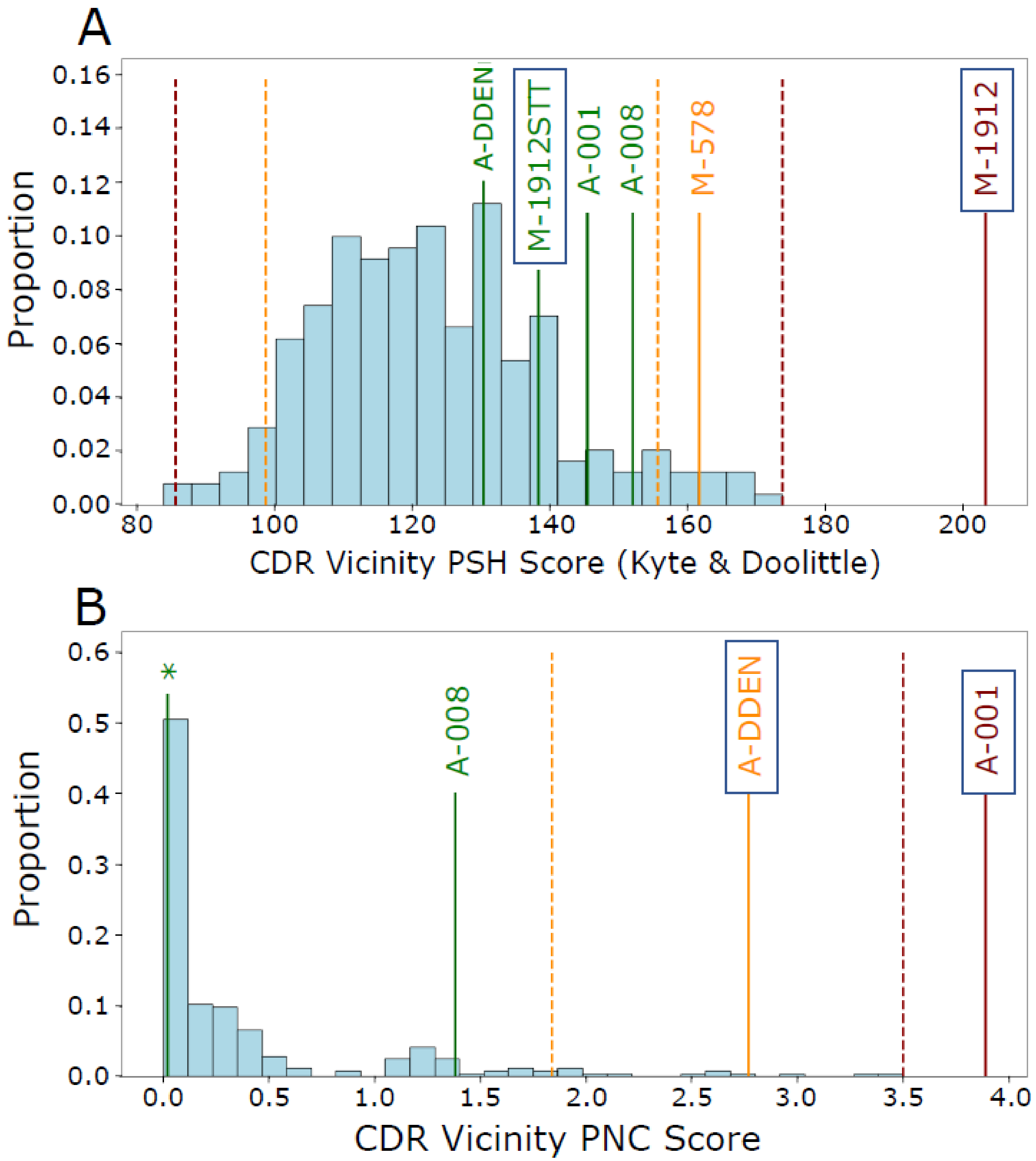
*Erenumab flagged for each of these properties

- Low red-flagging rate (8 of 105), implies won’t pick out genuine therapeutics as having issues very often.

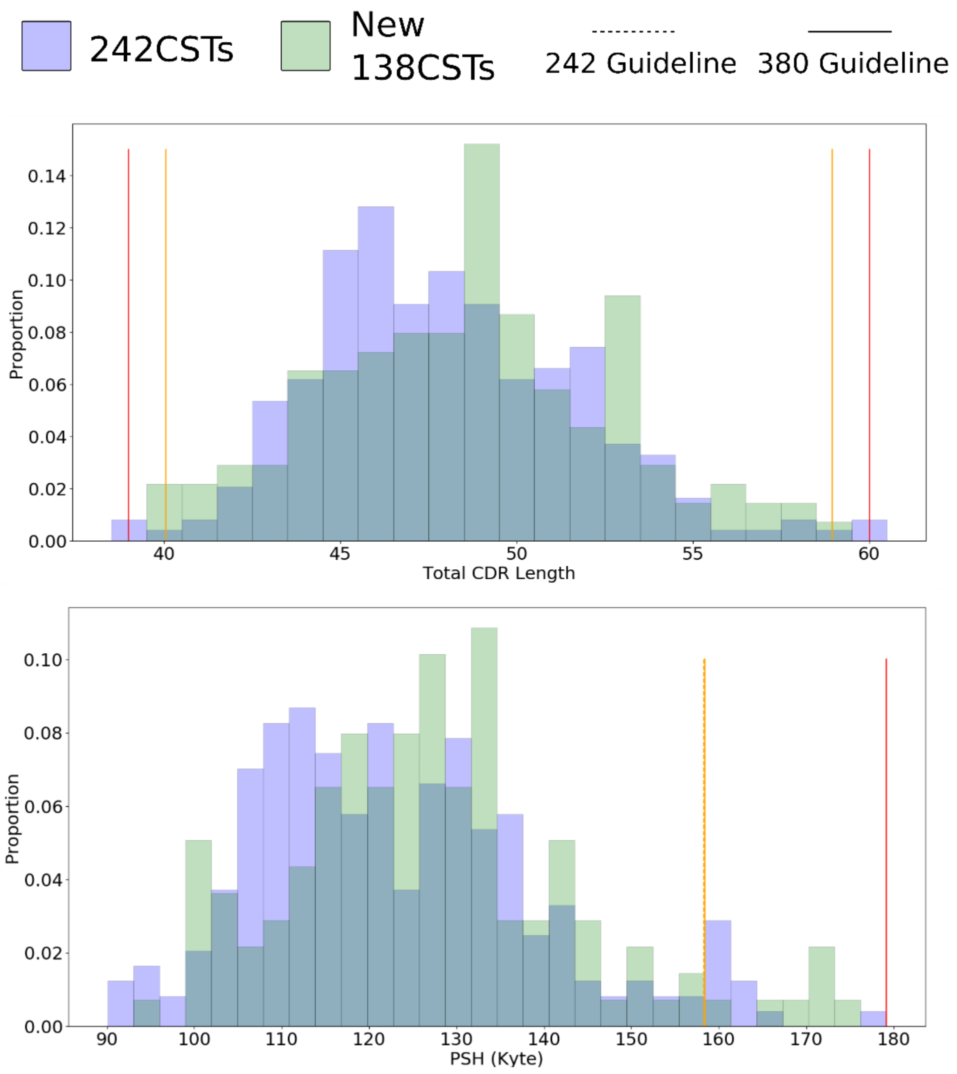
Validation

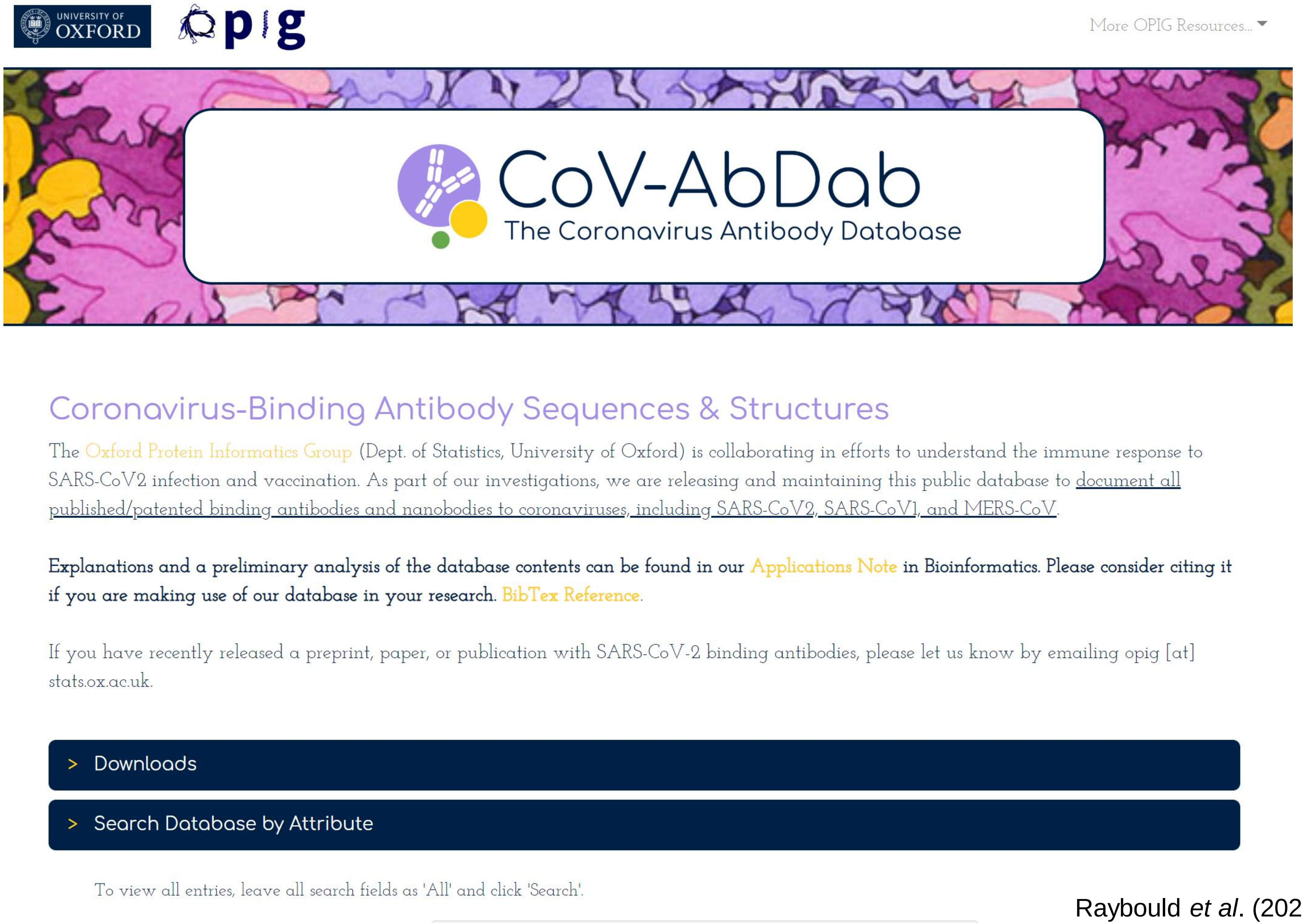
M-1912 aggregated uncontrollably during development, and exhibited extremely high values in our CDR Vicinity PSH metric.
M-1912STT resolved the issue.

A001 had prohibitively poor expression levels, and exhibited extremely high values in our CDR Vicinity PNC metric.
A-DDEN fixed the issue (backbone engineering)



TAP guidelines auto-updating





Coronavirus-Binding Antibody Sequences & Structures

The Oxford Protein Informatics Group (Dept. of Statistics, University of Oxford) is collaborating in efforts to understand the immune response to SARS-CoV2 infection and vaccination. As part of our investigations, we are releasing and maintaining this public database to document all published/patented binding antibodies and nanobodies to coronaviruses, including SARS-CoV2, SARS-CoV1, and MERS-CoV.

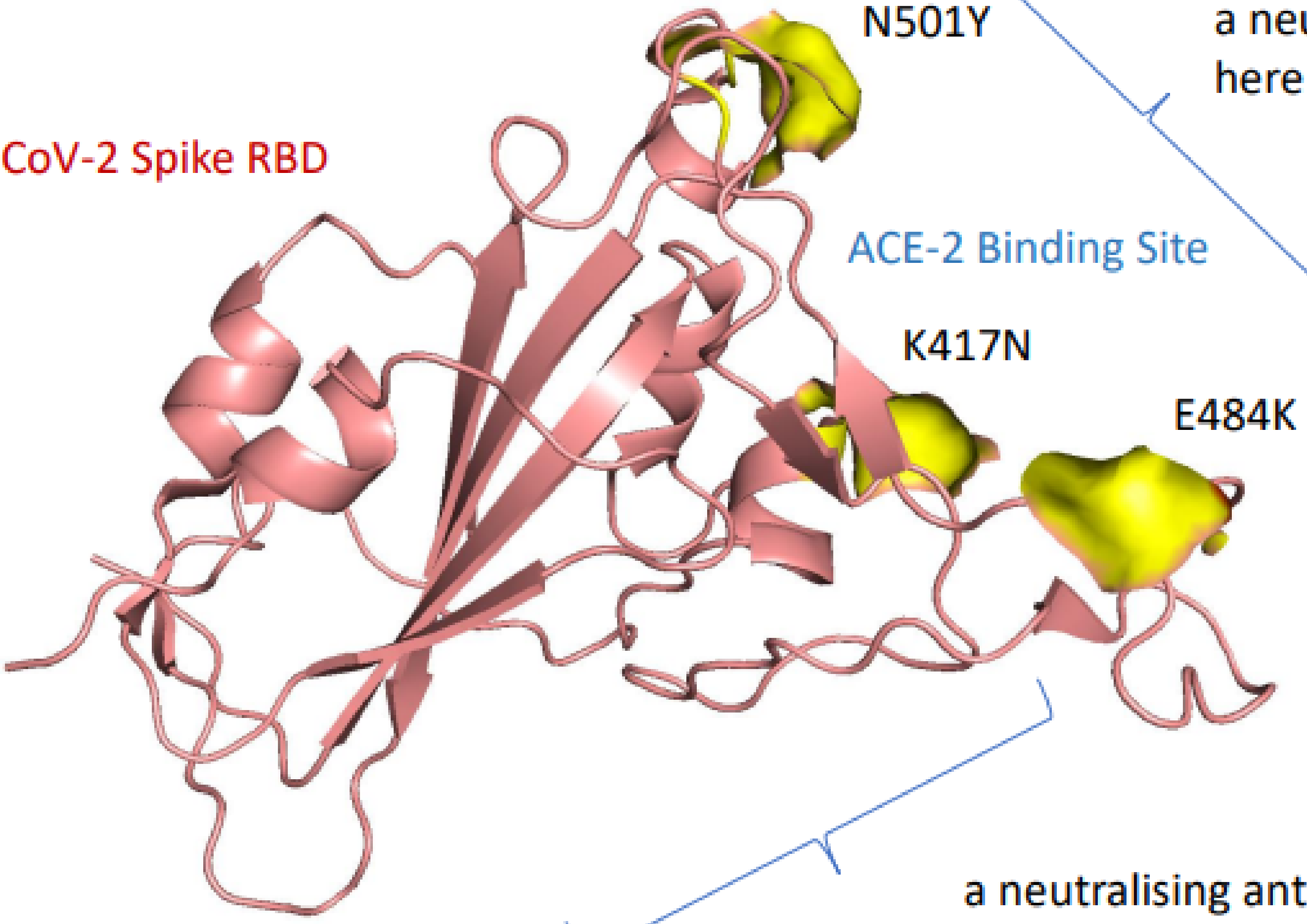
Explanations and a preliminary analysis of the database contents can be found in our Applications Note in Bioinformatics. Please consider citing it if you are making use of our database in your research. BibTex Reference.

If you have recently released a preprint, paper, or publication with SARS-CoV-2 binding antibodies, please let us know by emailing opig [at] stats.ox.ac.uk.

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

Raybould *et al*. (2020). *Bioinformatics.*

Epitope profiling: it’s really important to know **where** pathogen response antibodies bind…



a neutralising antibody that binds wildtype SARS-CoV-2 here is much more likely to be SARS-CoV-2 variant-specific

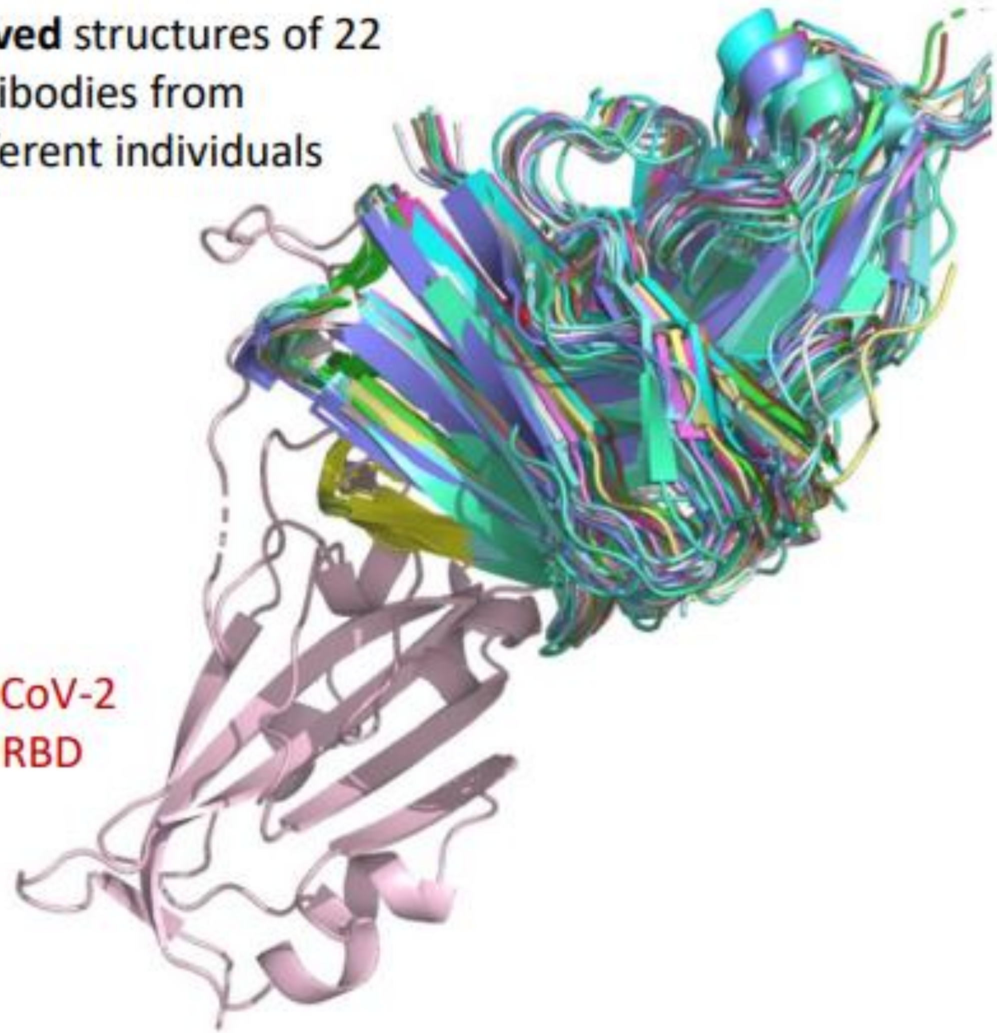
Better epitope profiling allows us to gain improved understanding of which binding modes give each individual B-cell immunity

-> evaluate susceptibility to new variants
-> possibility of targeting “sub-dominant” sites through monoclonal antibody design

a neutralising antibody that binds wildtype SARS-CoV-2 here would be more likely to neutralise the variants

Computational Epitope Profiling using **solved structures**

solved structures of 22 antibodies from different individuals



CDRH3 Sequences

AREAYGMDV

ARSPYGGNS

AREVAGTYDY

ARDVADAFDI

ARDFYEGSFDI

ARDLGPYGMDV

ARDFGDFYFDY
ARDYGDYYFDY
ARDYGDYYFDY

ARDLDVYGLDV
ARDLMVYGIDV

ARDLGSGDMDV

ARDLVVYGMDV

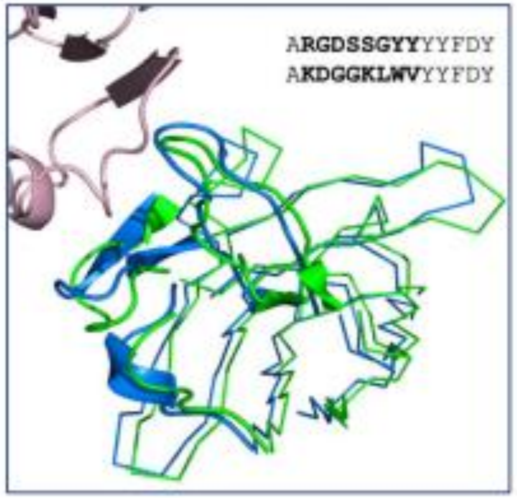
ARDLERAGGMDV
ARDLGEAGGMDV

ARDLDVSGGMDV

ARDLQELGSLDY

ARVLPMYGDYLDY

ARGDVSGYRYGLDY
ARGDVSGYRYGLDY
ARGDVSGYRYGLDY
ARGDVSGYRYGLDY



Antibody response to SARS-CoV-2 can be **functionally public** even if the sequences are dissimilar

We can use the structures of the antibodies as another way to functionally group them as binding the same epitope

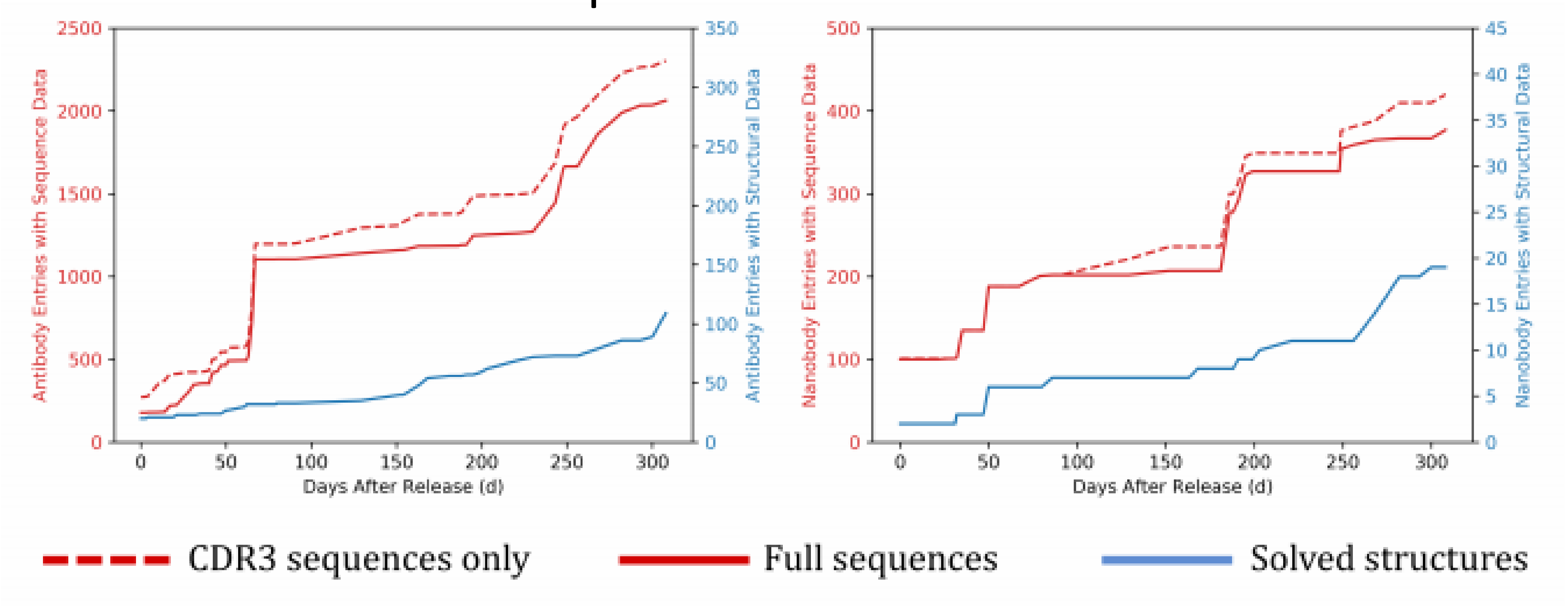
But most antibodies don't have solved structures…

Epitope profiling of coronavirus-binding antibodies using computational structural modelling

- Antibodies from markedly different lineages but with similar structures can engage the same epitope with near-identical binding modes.
- Identify sequence-dissimilar antibodies that engage the same epitope
  - Input a large dataset of antibodies known to bind to a single antigen some with known epitopes
  - Use a novel computational method to epitope profile the dataset based on structural modelling and clustering
- Show this on CovAbDab

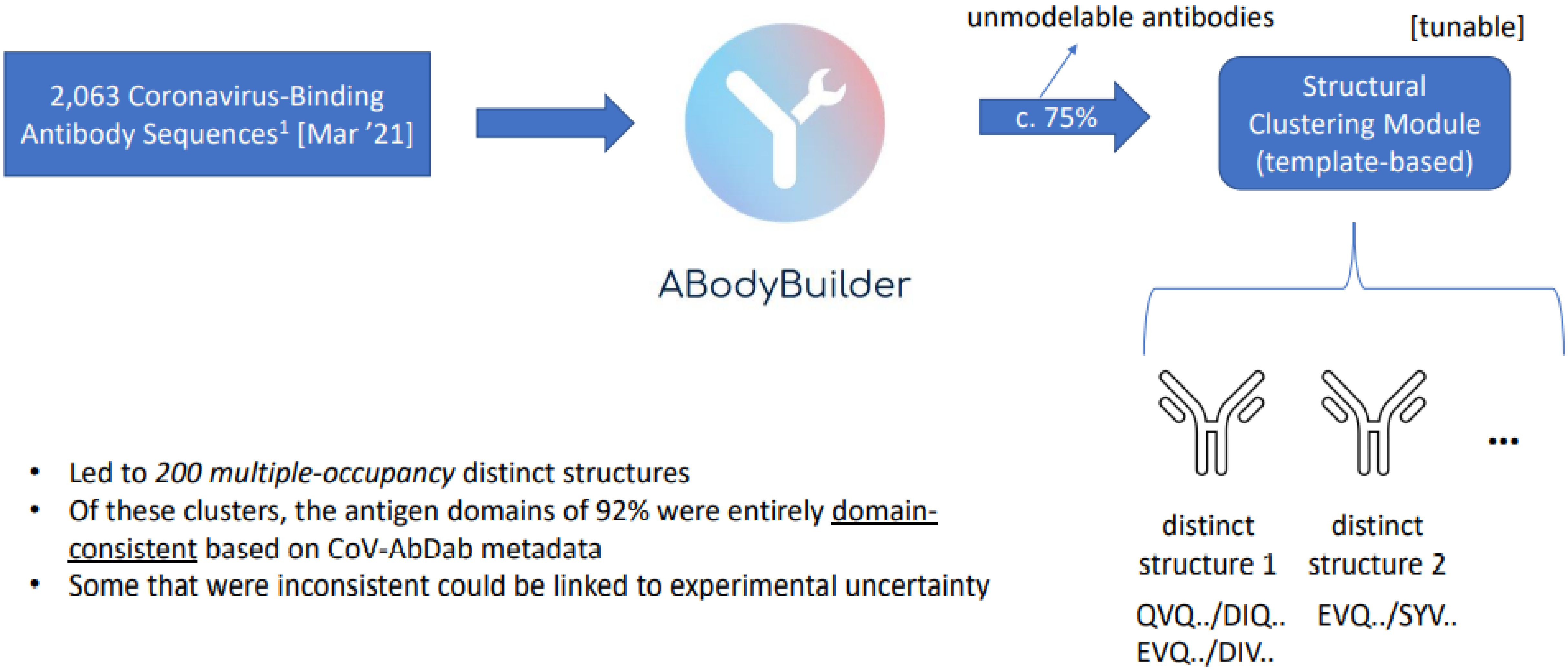
Robinson *et al*. (2021). *bioRxiv.*

CoVAbDab in sequences and structures



As of 11 March, just ~5% (113/2,304) of the antibodies in CoV-AbDab had at least one solved X-ray or cryo-EM structure, while ~90% (2,063/2,304) of the antibodies had full Fv amino acid sequences

Computational Epitope Profiling using **predicted structures**



2,063 Coronavirus-Binding Antibody Sequences[1] [Mar '21]

Structural Clustering Module (template-based)

- Led to *200 multiple-occupancy* distinct structures
- Of these clusters, the antigen domains of 92% were entirely domain-consistent based on CoV-AbDab metadata
- Some that were inconsistent could be linked to experimental uncertainty

[1] Raybould MIJ, Kovaltsuk A, Marks C, Deane CM (2021) CoV-AbDab: the Coronavirus Antibody Database. *Bioinformatics. 37(5):734-735.*

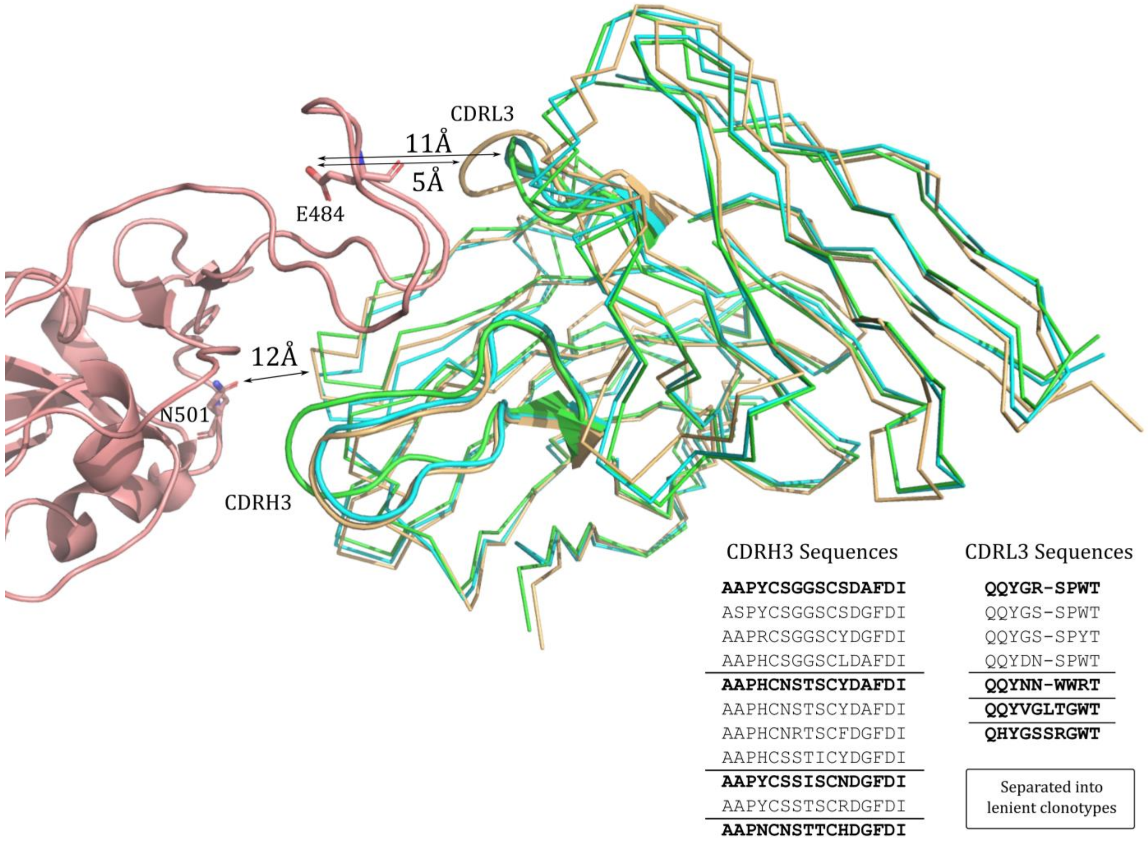
Predicting structure to predict epitopes

- Use Abodybuilder to model the 2063 sequences
  - “accurate models for 1500”
- Structurally cluster the models
  - 1,159 clusters
  - 541 sequences belonged to the 200 clusters that had > 1 sequence in
- For 184 of the 200 clusters the antibodies engage the same epitope based on available data.→ 92% accuracy
- The 16 false positives
  - poor expt labelling
  - poor modelling
- Structural clusters frequently span multiple clonal lineages.

Predicting structure to predict epitopes

- The functional properties of the less well-characterised antibodies can be inferred from other antibodies predicted to adopt the same structure.

- One experiment could reveal the binding side of whole un-annotated clusters



CDRH3 Sequences

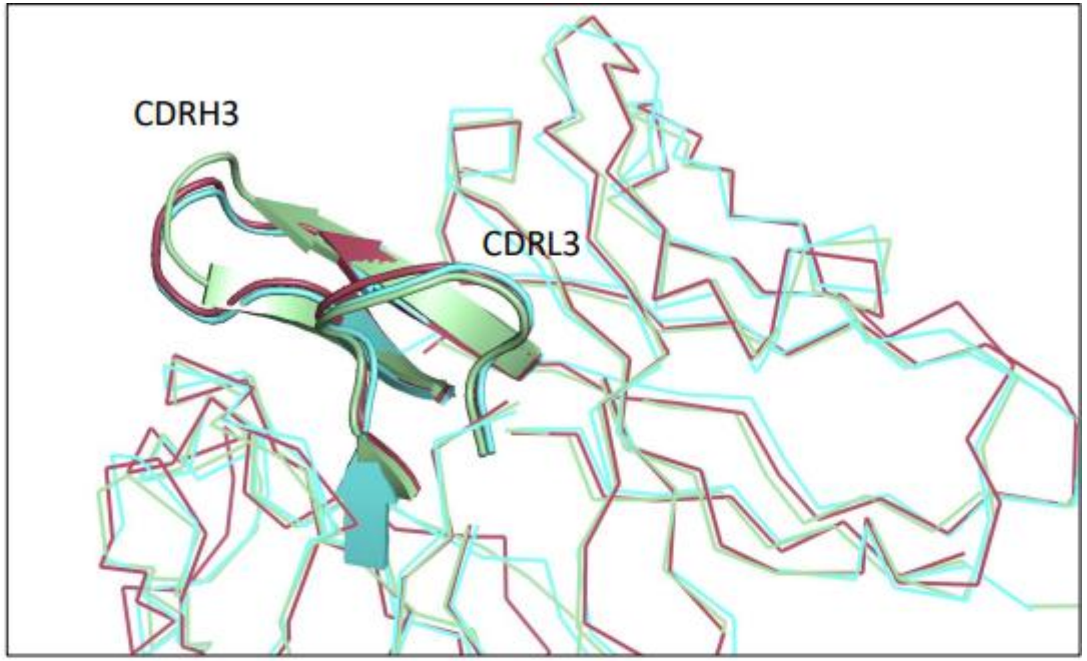
AAPYCSGGSCSDAFDI
ASPYCSGGSCSDGFDI
AAPRCSGGSCYDGFDI
AAPHCSGGSCLDAFDI
AAPHCNSTSCYDAFDI
AAPHCNSTSCYDAFDI
AAPHCNRTSCFDGFDI
AAPHCSSTICYDGFDI
AAPYCSSISCNDGFDI
AAPYCSSTSCRDGFDI
AAPNCNSTTCHDGFDI

CDRL3 Sequences

QQYGR-SPWT
QQYGS-SPWT
QQYGS-SPYT
QQYDN-SPWT
QQYNN-WWRT
QQYVGLTGWT
QHYGSSRGWT

Separated into lenient clonotypes

Clustering by predicted structure functionally links coronavirus-binding antibodies across the species barrier



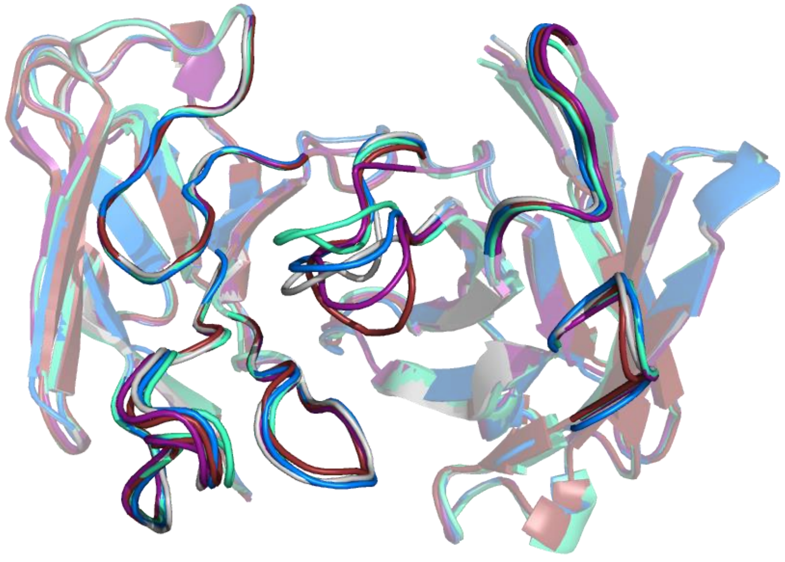
- Mice (maroon and green) and humans (cyan) create sequence dissimilar since they have distinct germlines
- Example of a human and two murine RBD binders with very high structural similarity
- Allows us to understand which coronavirus binding sites are targetable by different gene loci
- **Compare immune functions of different organisms**

Structural Profiling of Antibodies to Cluster by Epitope, “SPACE”

- 92% prediction accuracy
- Functionally links antibodies with distinct genetic lineages, species origins, and coronavirus specificities
- Greater convergence exists in the immune responses to coronaviruses than would be suggested by sequence-based approaches.
- Applying structural analytics to large class-specific antibody databases will enable high confidence structure-function relationships to be drawn

- Will not only be useful for early-stage drug discovery but also for understanding epitope immunodominance, and therefore vaccine design.

Robinson *et al*. (2021). *bioRxiv.*

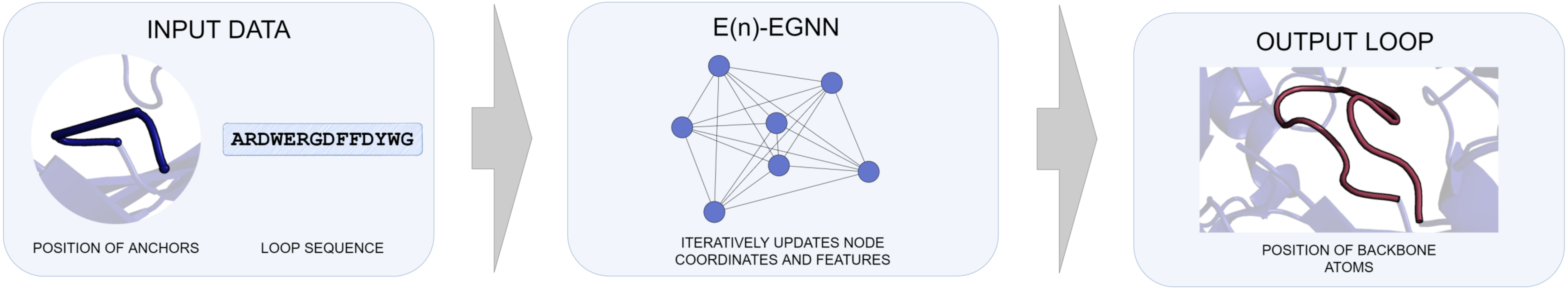


ABlooper

Improving the speed and quality of structural models of antibodies

Abanades *et al.* (2021)

ABlooper pipeline



INPUT DATA

ARDWERGDFFDYWG

POSITION OF ANCHORS

LOOP SEQUENCE

E(n)-EGNN

ITERATIVELY UPDATES NODE COORDINATES AND FEATURES

OUTPUT LOOP

POSITION OF BACKBONE ATOMS

- Use 5 E(n)-Equivariant Graph Neural Networks (E(n)-EGNN) to give 5 predictions of all of the CDRs
    - Average the 5 to create the final prediction

- End to end predictor – small energy minimisation useful.

- Gives an estimate of the acuraccy of prediction.

Abanades *et al.* (2021)

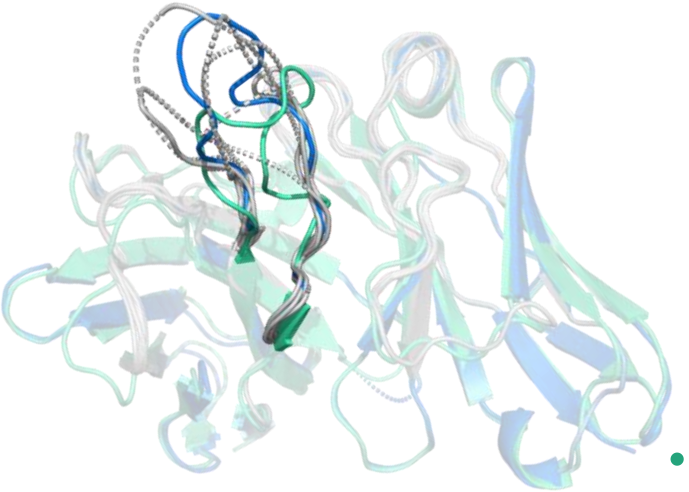
Predicting CDR-H3 on modelled structures

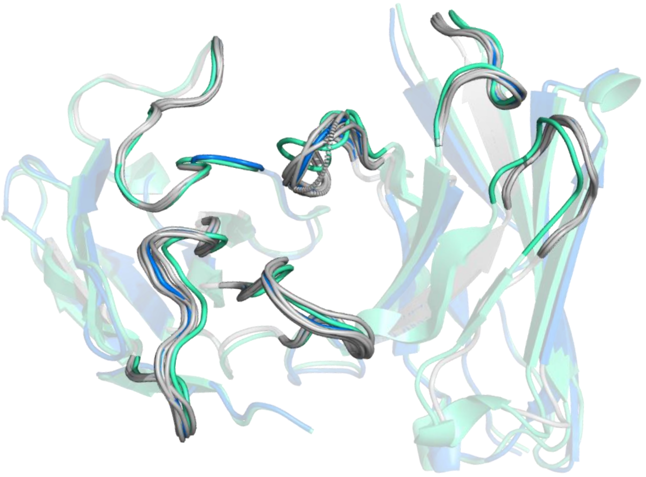
	Rosetta Antibody Benchmark	**SABDab Latest Structures**
AlphaFold2	2.87*	
ABodyBuilder	2.77	3.25
DeepAb	2.44	2.49*
ABlooper	2.49	2.72
Ablooper Unrelaxed	2.45	2.66

RMSD across backbone atoms to the correct structure
*potentially these structures were contained in the training of these methods

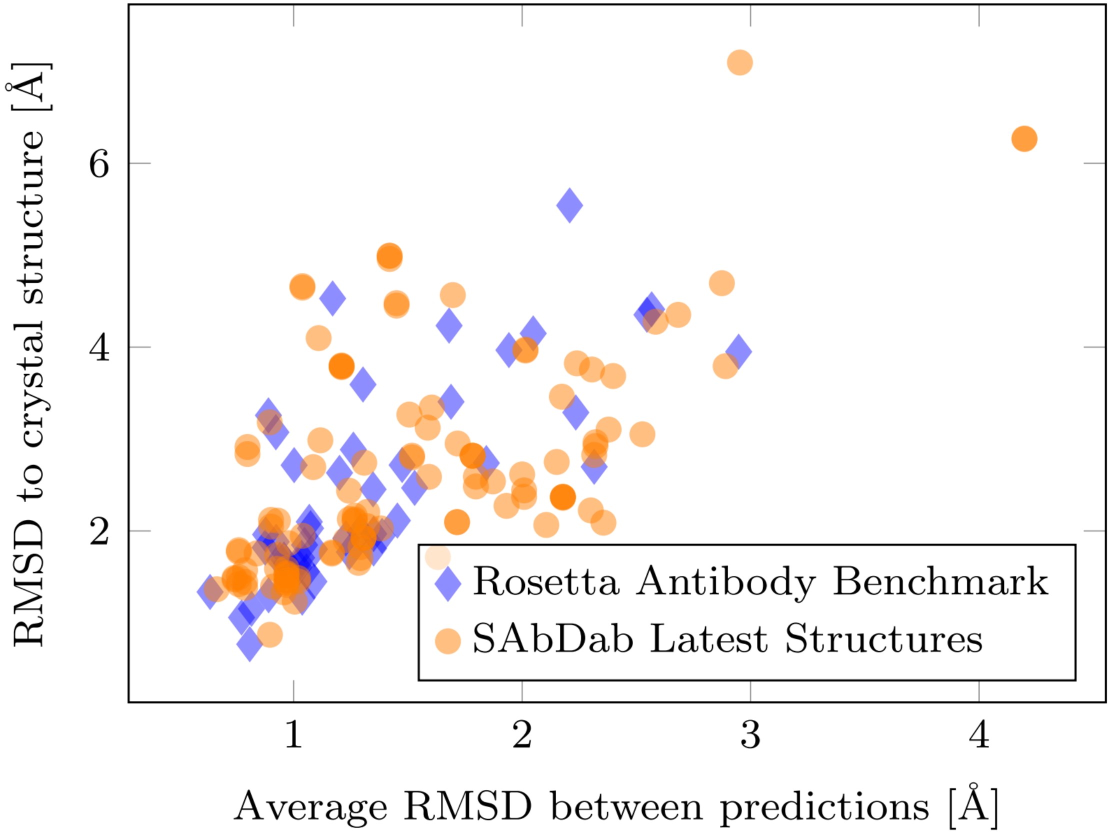
Abanades *et al.* (2021)

Prediction diversity reveals prediction quality





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



Abanades *et al.* (2021)

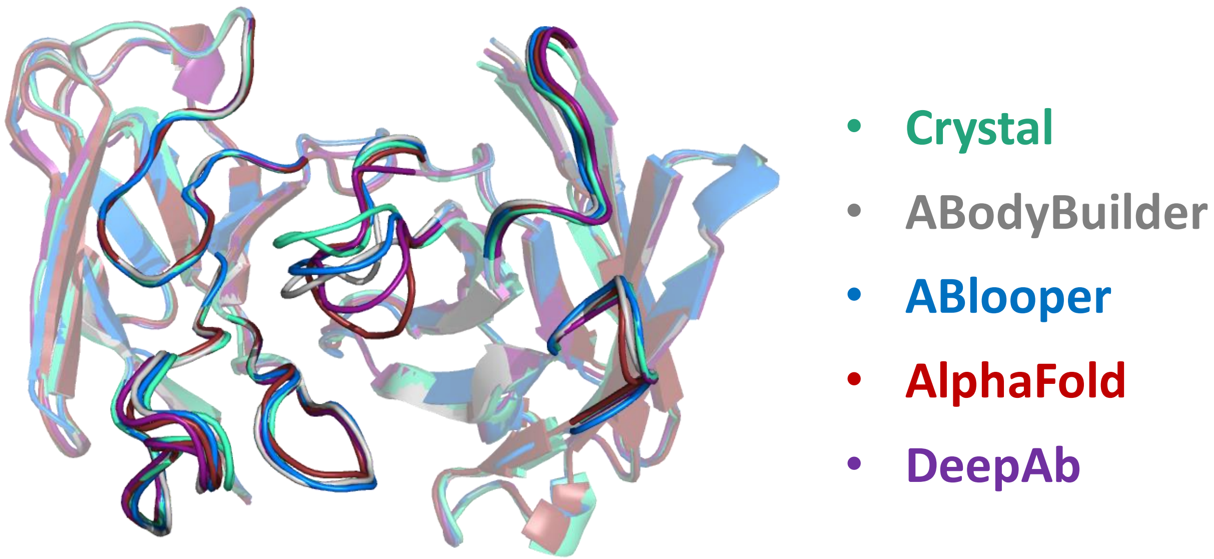
ABlooper – rapid accurate structure prediction for antibodies.

Overall similar levels of accuracy to other deep learning methods

ABlooper is faster

- Can predict the CDRs for one hundred structures in under five seconds.

ABlooper contains an accuracy estimate



Abanades *et al.* (2021)

Acknowledgements

Kymab
Jacob Galson
Paul Kellam

GSK
Alan Lewis
Matthew Bottomley
Newton Wahome
Ian Wall

UCB
Jiye Shi
James Snowden

Roche
Guy Georges
Alexander Bujotzek

Astra Zeneca
Maria Flocco
Andrew Buchanan

All the ex members of OPIG

Particularly
Claire Marks
Wing Ki Wong
Aleksander Kovaltsuk
Mark Chin
Jinwoo Leem
Konrad Krawczyk

All the current members of OPIG

Particularly
Matthew Raybould
Eve Richardson
Sarah Robinson
Brennan Abanades Kenyon
Constantin Schneider

opig

kymab
Vernalis RESEARCH
AstraZeneca
diamond
LifeArc
UKRI Medical Research Council
BenevolentAI
Roche
ucb
gsk GlaxoSmithKline
Boehringer Ingelheim
UKRI Engineering and Physical Sciences Research Council
Exscientia DRIVEN BY KNOWLEDGE
IBM
Global Phasing
UKRI Biotechnology and Biological Sciences Research Council

Software Availability

- Free OPIG Webserver and GitHub.

http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/

SAbDab SAbPred STCRDab

OAS CoV-AbDab GitHub

If data is IP-sensitive or as an academic you want to run large batches

- Vagrant VirtualBox
- Singularity Virtual Machine

SAbBox

enquiries: opig@stats.ox.ac.uk