Constraints and variability of complementarity determining regions in antibodies

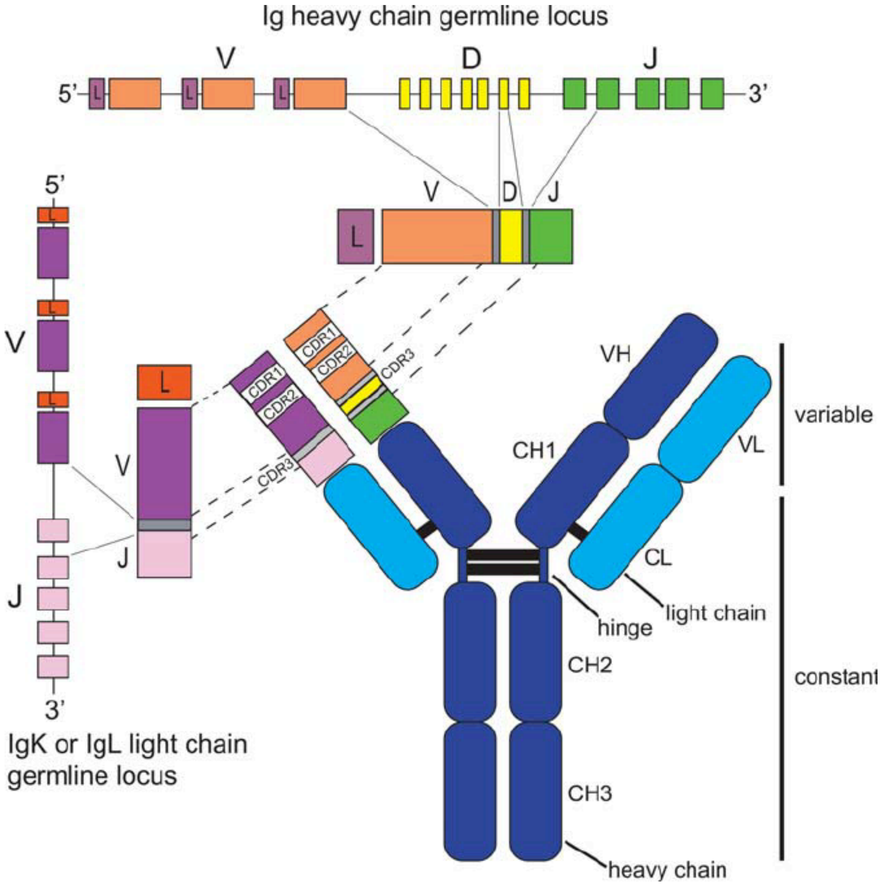
Rosalba Lepore
Barcelona - BSC 2018

Institut Pasteur

BIOZENTRUM
University of Basel
The Center for
Molecular Life Sciences

SIB
Swiss Institute of
Bioinformatics

Origin of antibody variability

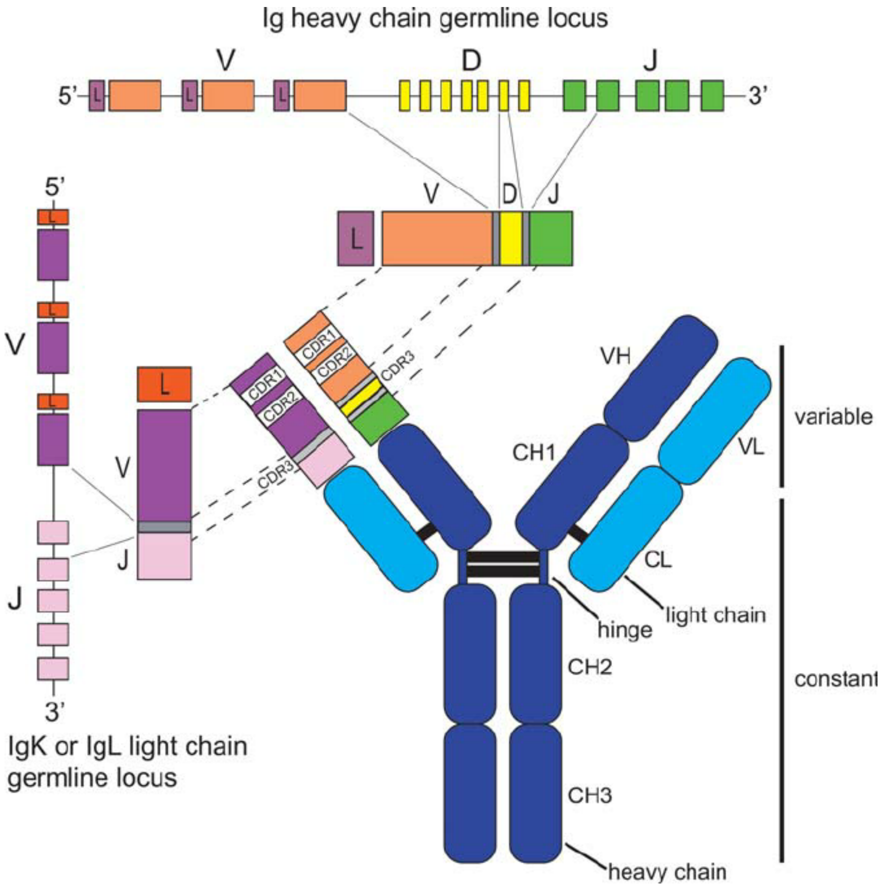


Boyd et al. 2014

V regions are encoded by a pool of gene segments (V, D, J)

1. Combinatorial diversification
2. Joining
3. VH/VL pairing
4. Somatic Hypermutation

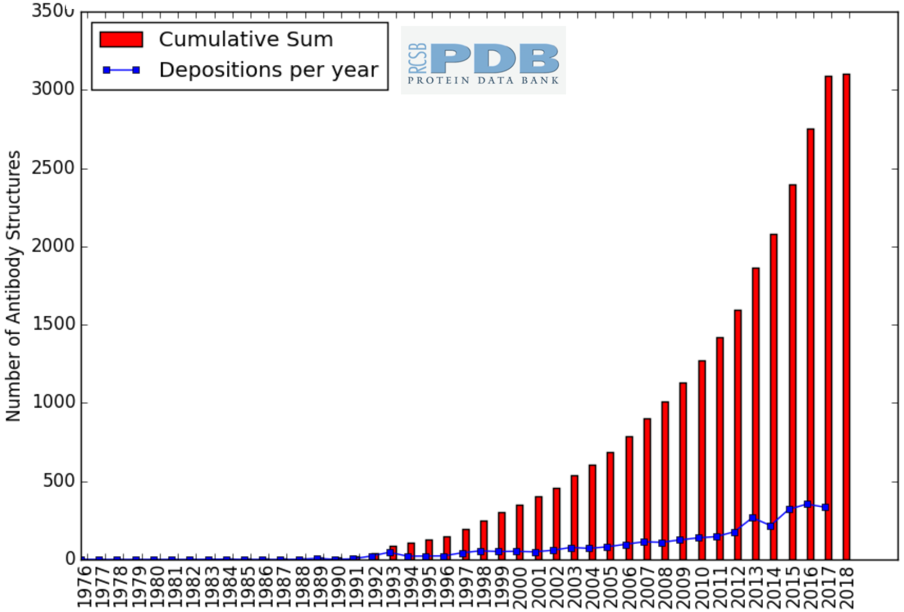
Origin of antibody variability



Boyd et al. 2014

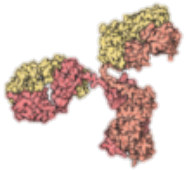
Human repertoire
> 10^{12} different antibodies

Available structural information



Number of Antibody Structures in PDB
(as of Wed 9 May 2018) from SabDab

Human repertoire
> 10^{12} different antibodies

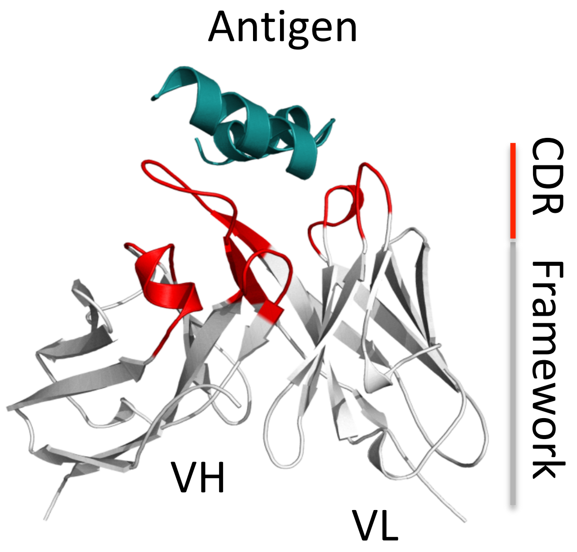


Total structures: **3125**

Paired VH/VL: **~2500**

In complex with antigens ~ **300**

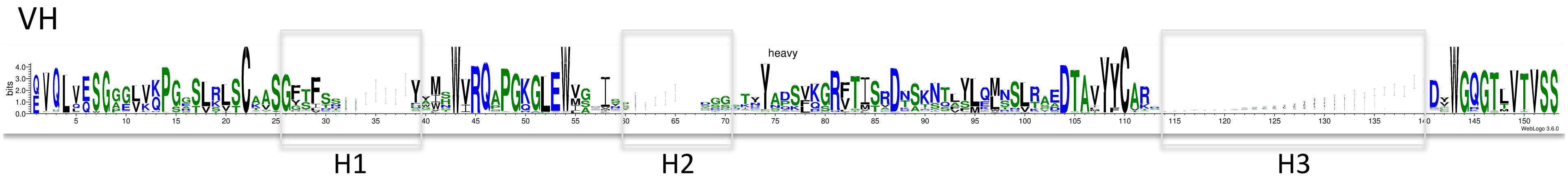
Variability clusters in CDR

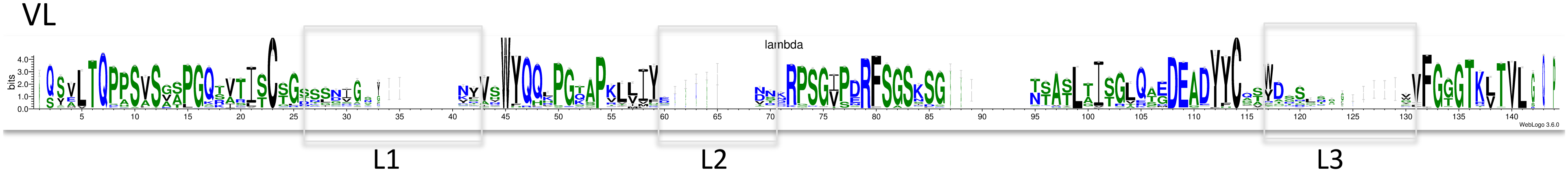


Three hypervariable loops for each chain or **complementarity determining regions** or **CDR**

Antibody specificity arises from **length and sequence variation** of these regions

Conserved regions: **framework**



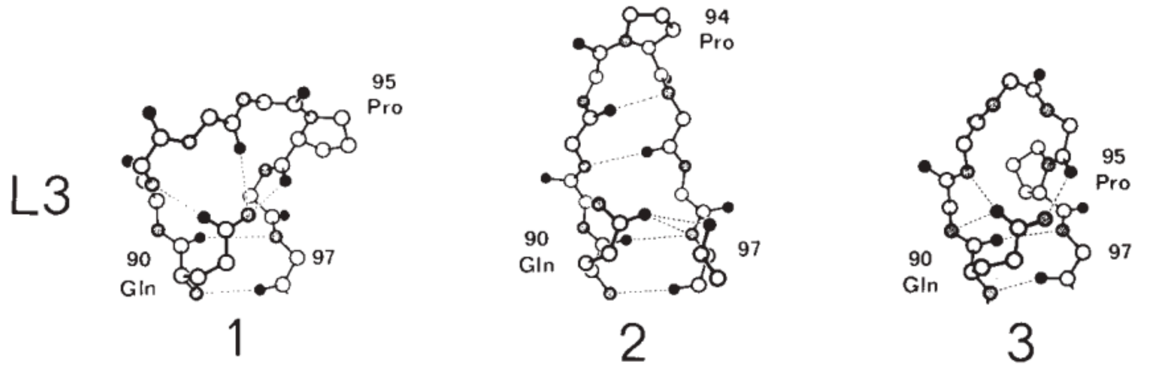


Canonical structure model

Hypervariable loops only exibit a limited number of conformations or **canonical structures**

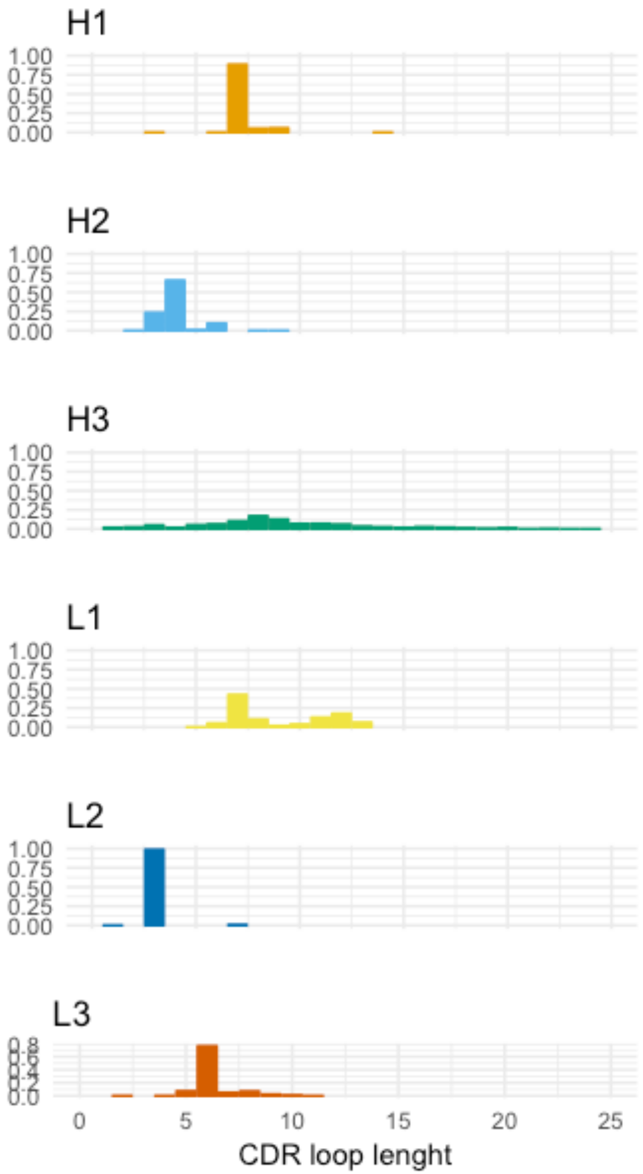
Key conserved residues at specific positions and the **lenght of the loop** determine the canonical structure

Hypervariable loop L3 canonical structure: Gln90, Pro94-95



Conformations of immunoglobulin hypervariable regions. Nature 1989

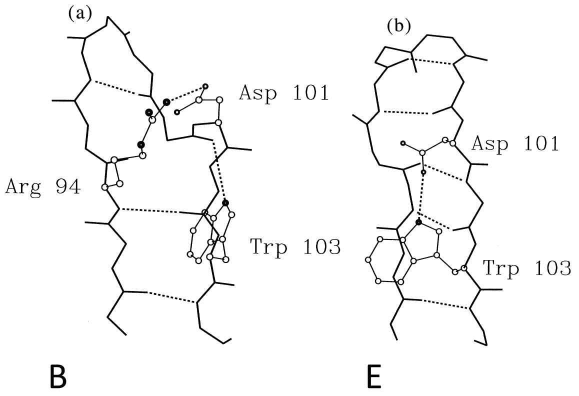
H3



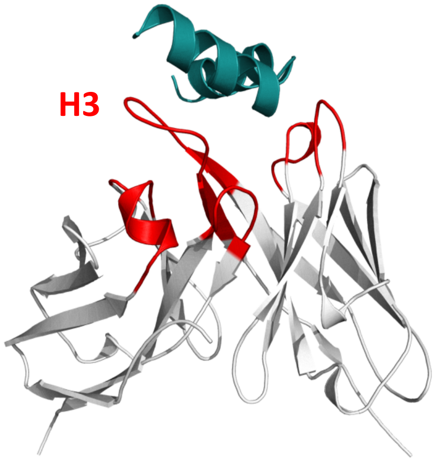
H3 is the most variable loop in terms of length, sequence composition and structure

Only a partial canonical structure identified:
Bulged / Extended

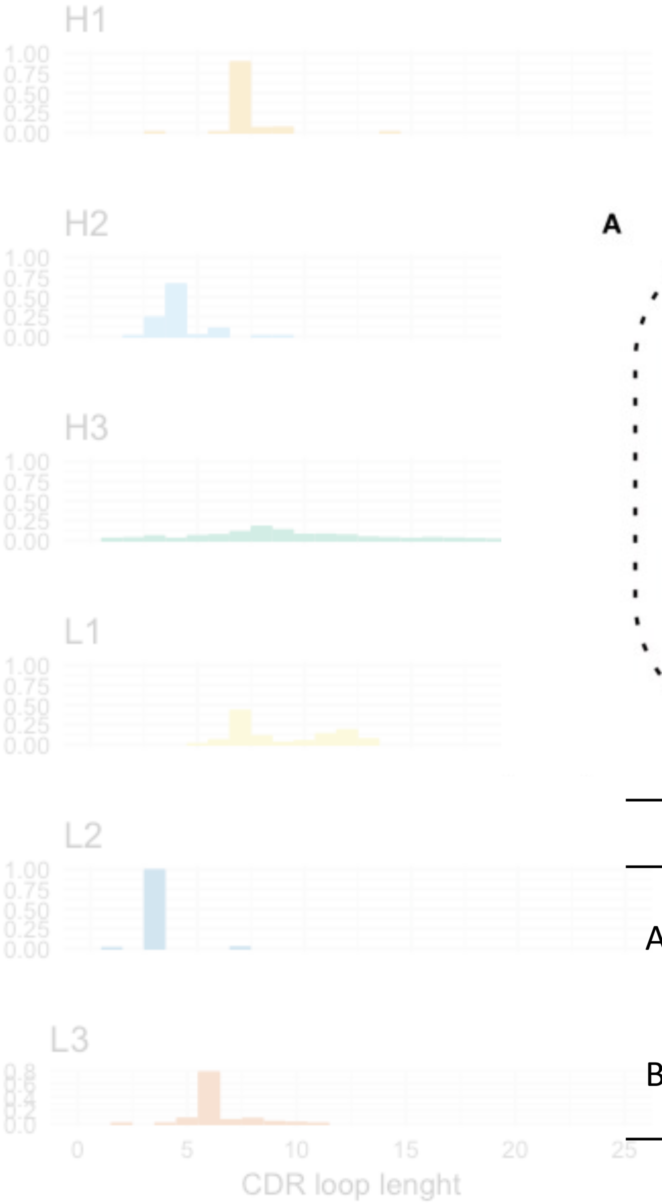
Head of the loop highly variable

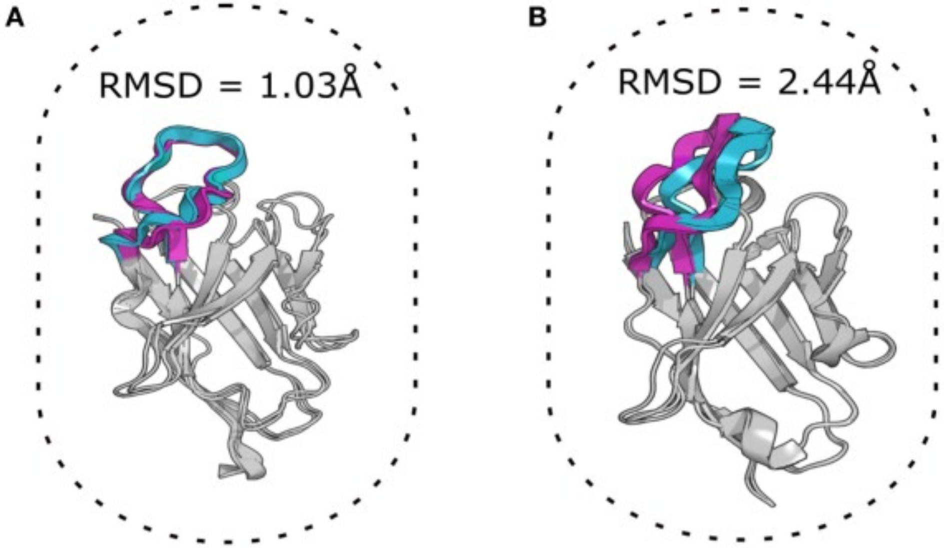


Morea et al. J Mol Bio 1998



H3

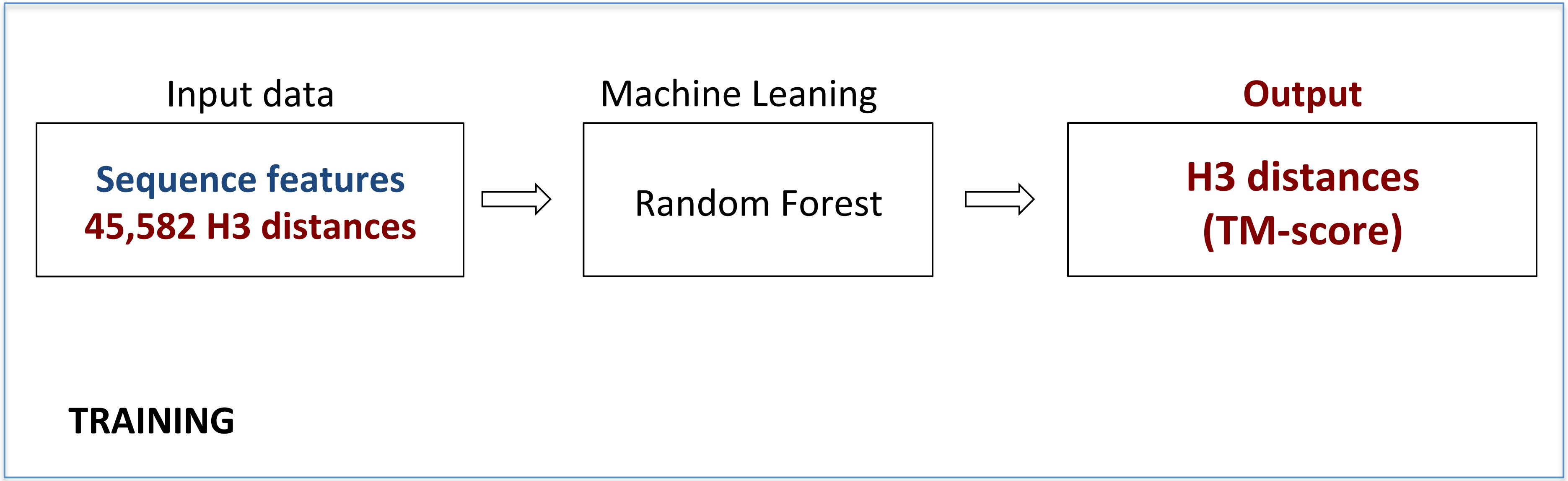




	PDB	CDR H3 SEQUENCE	SEQ ID %	RMSDÅ
A	4NZU	ARAPDVADADCHKGAFGY	27.7	1.03
	4SIS	VRTADCERDPCKGWVFPH		
B	3U7W	TRGKYCTARDYYNWDFEH	88	2.44
	4JDV	ARGKYCTARDYYNWDFQH		

from Kovaltsuk et al., Front Immunol 2017

Structure prediction of H3



Input data

Sequence features

45,582 H3 distances

Machine Leaning

Random Forest

Output

H3 distances

(TM-score)

TRAINING

Structure prediction of H3

Input data

Antibody sequence ⇨ H3Loopred ⇨

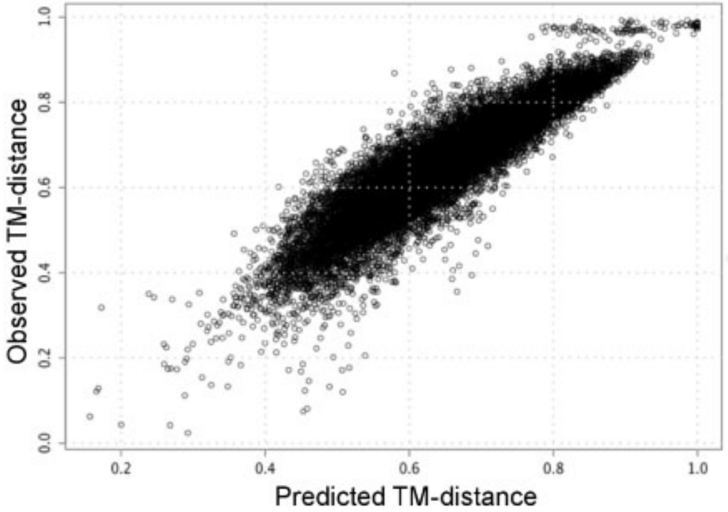
Output

H3 distances (TM-score)

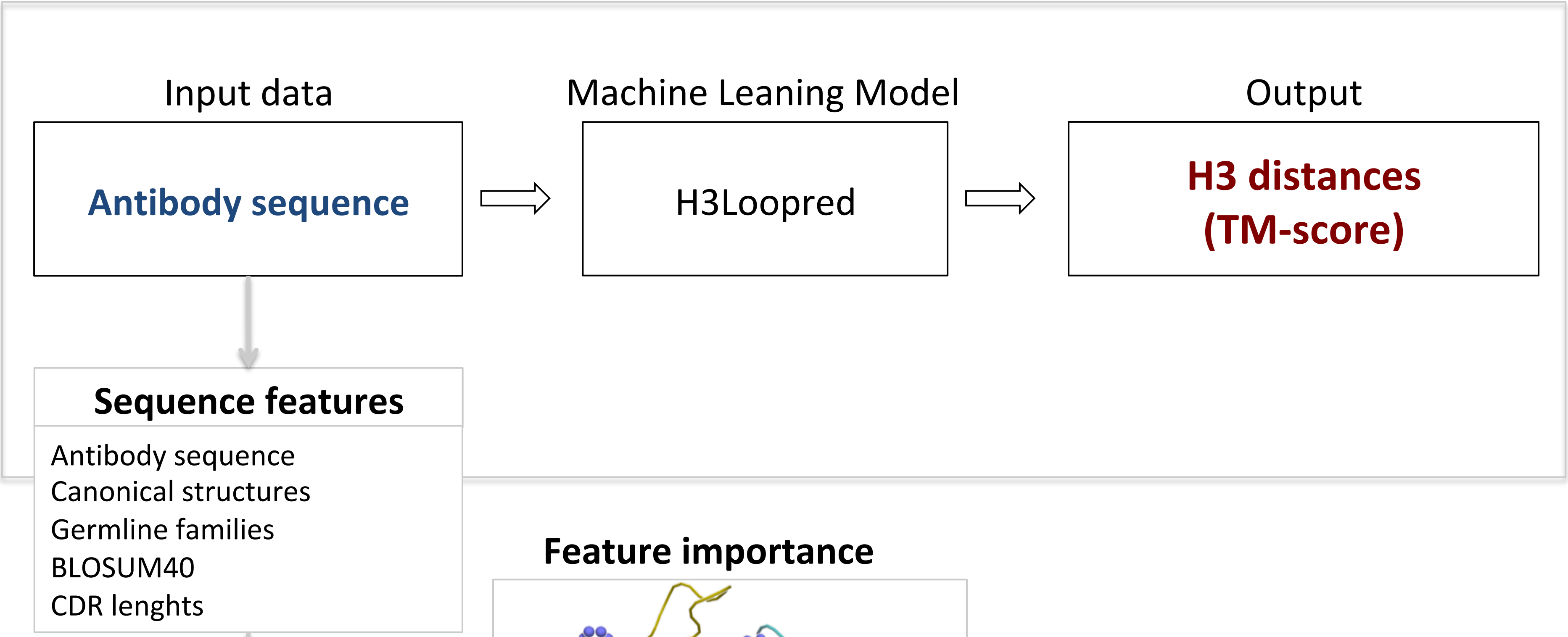
VALIDATION

5-fold cross validation

Cor=0.88 (45,582 loop pairs)



Structure prediction of H3



Input data

Antibody sequence

Machine Leaning Model

H3Loopred

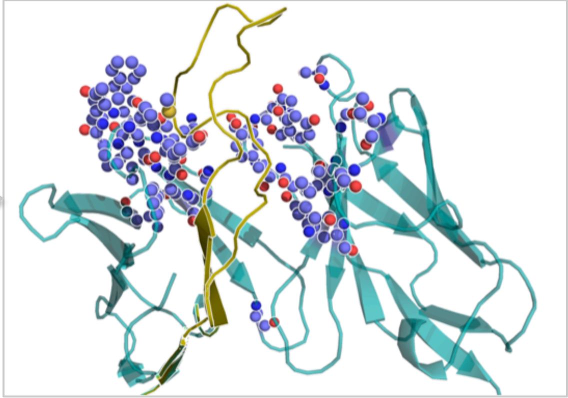
Output

H3 distances (TM-score)

Sequence features

Antibody sequence
Canonical structures
Germline families
BLOSUM40
CDR lenghts

Feature importance



Structure prediction of H3

Input data

Antibody sequence

Machine Leaning Model

H3Loopred

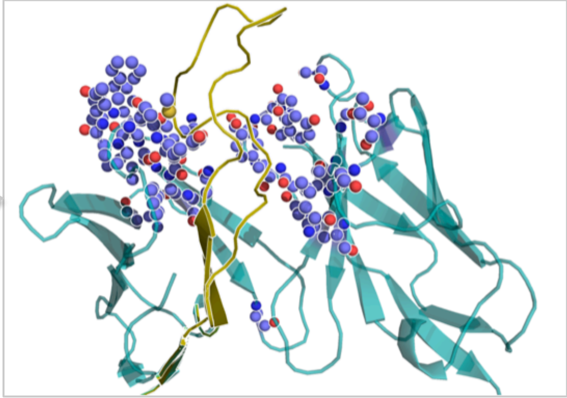
Output

H3 distances (TM-score)

Sequence features

Antibody sequence
Canonical structures
Germline families
BLOSUM40
CDR lenghts
Matching/non matching gaps

Feature importance

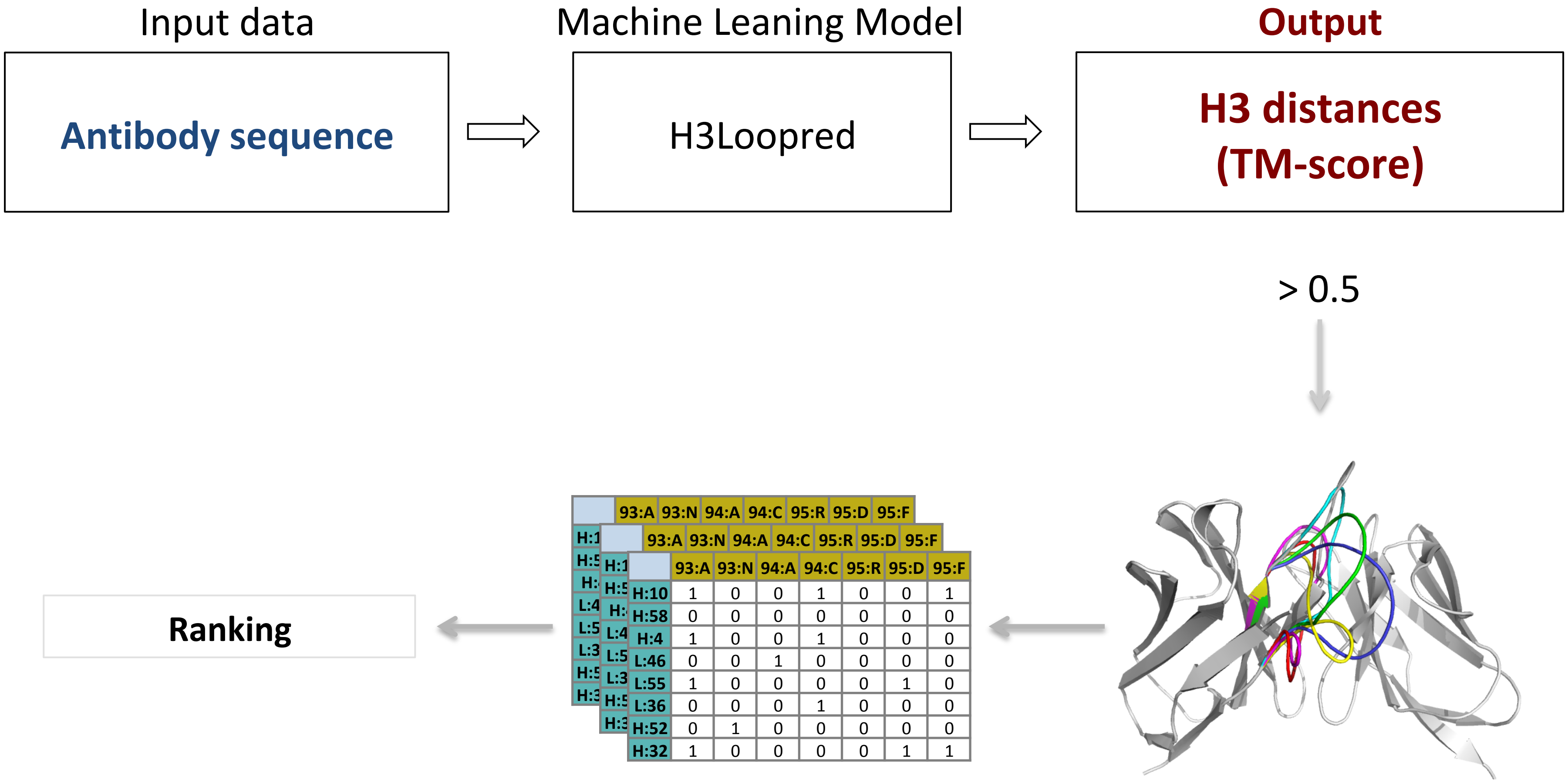


Ab structures

H3 loop residues and positions

Environment	93:A	93:N	94:A	94:C	95:R	95:D	95:F
H:10	1	0	0	1	0	0	1
H:58	0	0	0	0	0	0	0
H:4	1	0	0	1	0	0	0
L:46	0	0	1	0	0	0	0
L:55	1	0	0	0	0	1	0
L:36	0	0	0	1	0	0	0
H:52	0	1	0	0	0	0	0
H:32	1	0	0	0	0	1	1

Structure prediction of H3



Input data

Antibody sequence

Machine Leaning Model

H3Loopred

Output

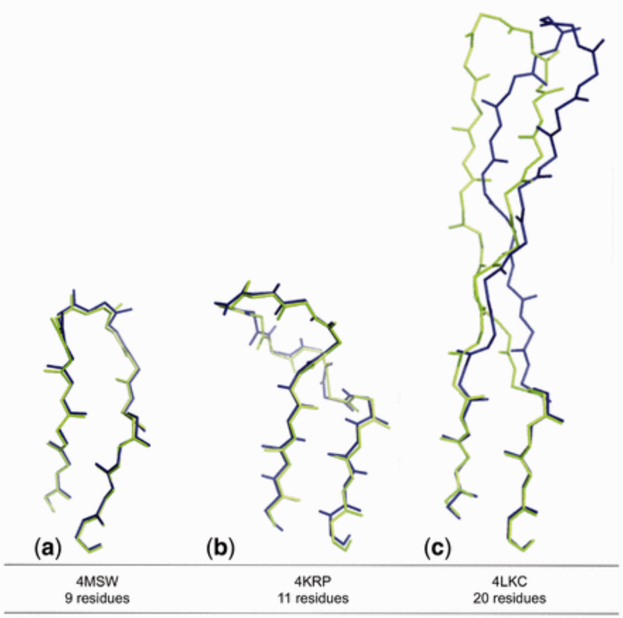
H3 distances (TM-score)

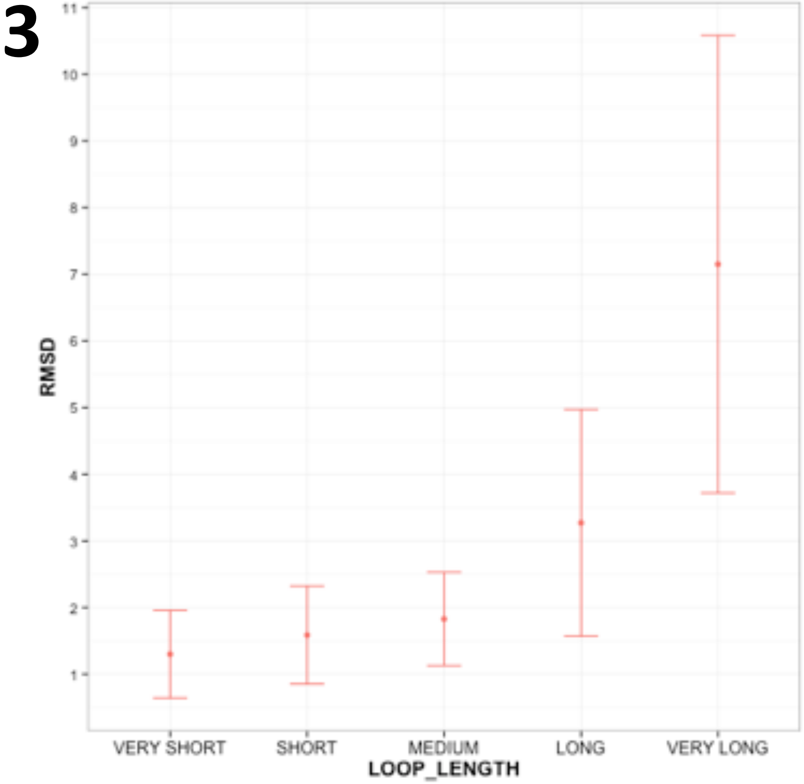
> 0.5

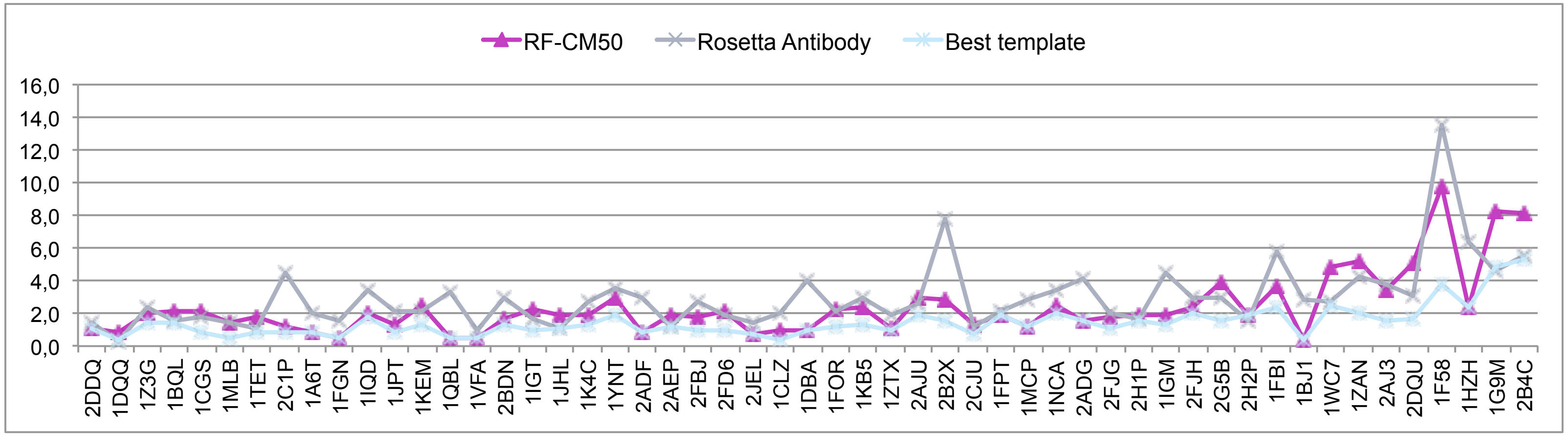
	93:A	93:N	94:A	94:C	95:R	95:D	95:F
H:10	1	0	0	1	0	0	1
H:58	0	0	0	0	0	0	0
H:4	1	0	0	1	0	0	0
L:46	0	0	1	0	0	0	0
L:55	1	0	0	0	0	1	0
L:36	0	0	0	1	0	0	0
H:52	0	1	0	0	0	0	0
H:32	1	0	0	0	0	1	1

Ranking

Structure prediction of H3







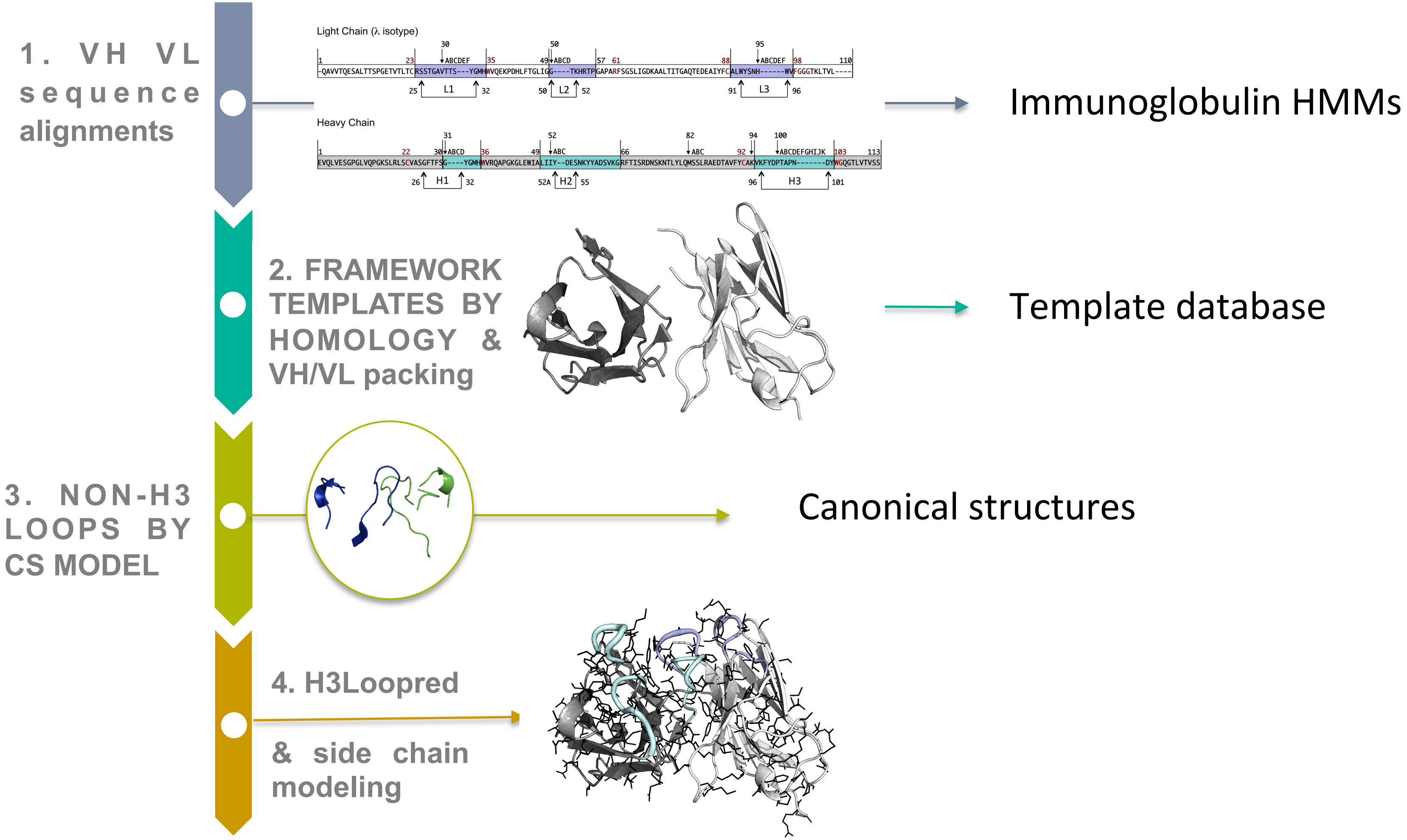
Lepore et al. Bioinformatics 2014

Prediction of Immunoglobulin Structures

PROTOCOL

Antibody modeling using the Prediction of ImmunoGlobulin Structure (PIGS) web server

Paolo Marcatili[1,3], Pier Paolo Olimpieri[1], Anna Chailyan[1,3] & Anna Tramontano[1,2]



1. VH VL sequence alignments

Immunoglobulin HMMs

2. FRAMEWORK TEMPLATES BY HOMOLOGY & VH/VL packing

Template database

3. NON-H3 LOOPS BY CS MODEL

Canonical structures

4. H3Loopred

& side chain modeling

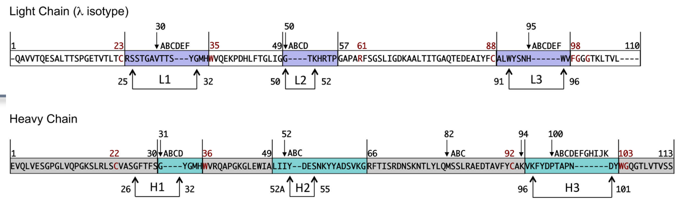
Prediction of Immunoglobulin Structures

PROTOCOL

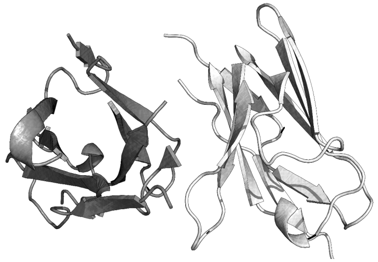
Antibody modeling using the Prediction of ImmunoGlobulin Structure (PIGS) web server

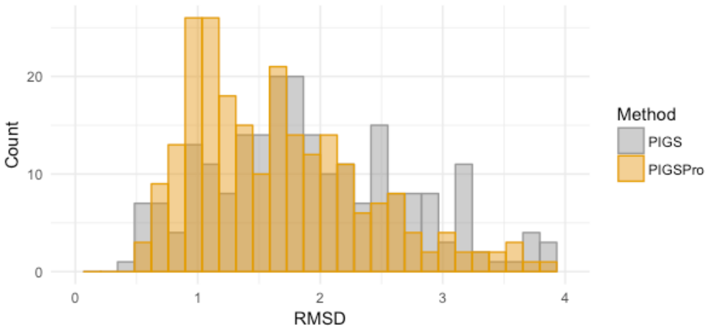
Paolo Marcatili[1,3], Pier Paolo Olimpieri[1], Anna Chailyan[1,3] & Anna Tramontano[1,2]

1. VH VL sequence alignments



2. FRAMEWORK TEMPLATES BY HOMOLOGY & VH/VL packing

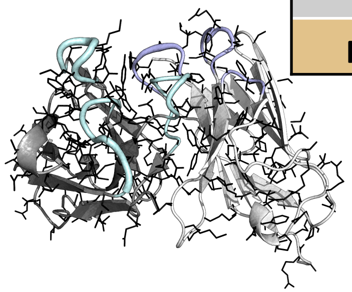




3. NON-H3 LOOPS BY CS MODEL

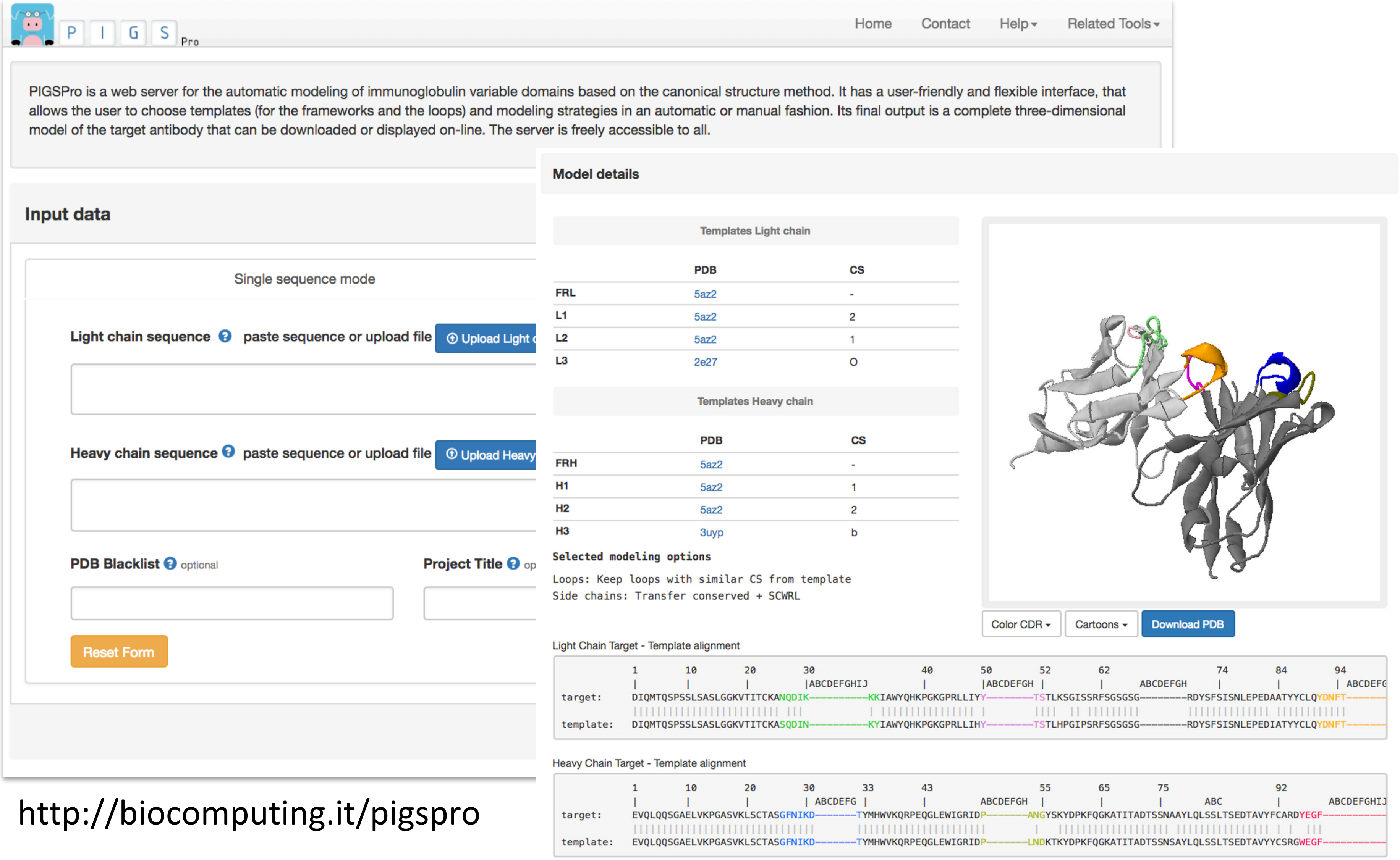
	CDRS	H3	Non H3 loops
PIGS	2.21 ± 1.48	3.59 ± 2.93	1.36 ± 0.61
PIGSPro	1.75 ± 0.95	2.41 ± 2.2	1.34 ± 0.6

4. H3Loopred & side chain modeling



PIGSPro: prediction of immunoGlobulin structures v2
Lepore et al. Nucl Acids Res 2017

Prediction of Immunoglobulin Structures



PIGSPro is a web server for the automatic modeling of immunoglobulin variable domains based on the canonical structure method. It has a user-friendly and flexible interface, that allows the user to choose templates (for the frameworks and the loops) and modeling strategies in an automatic or manual fashion. Its final output is a complete three-dimensional model of the target antibody that can be downloaded or displayed on-line. The server is freely accessible to all.

Templates Light chain

	PDB	CS
FRL	5az2	-
L1	5az2	2
L2	5az2	1
L3	2e27	O

Templates Heavy chain

	PDB	CS
FRH	5az2	-
H1	5az2	1
H2	5az2	2
H3	3uyp	b

```
Selected modeling options

Loops: Keep loops with similar CS from template
Side chains: Transfer conserved + SCWRL
```

Light Chain Target - Template alignment

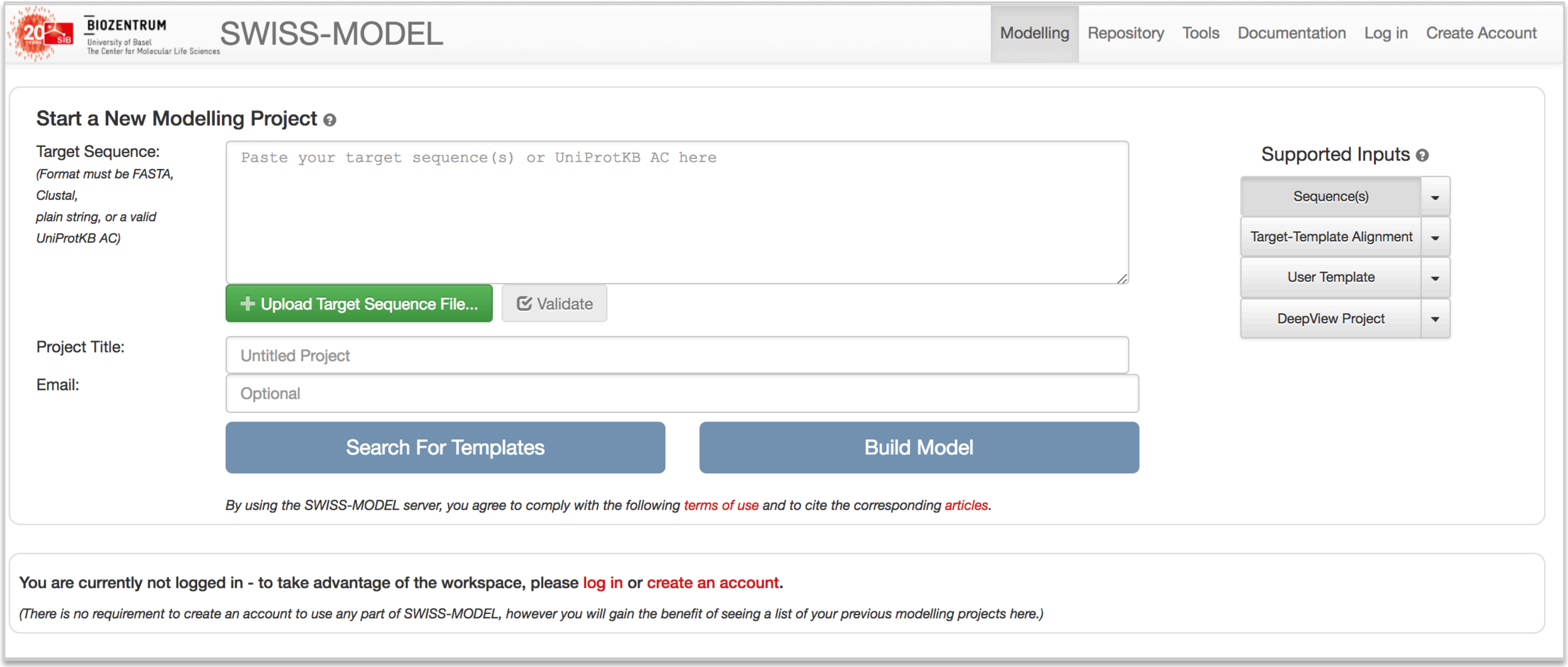
```
           1        10        20        30                  40        50        52        62                  74        84        94
           |        |         |         |ABCDEFGHIJ         |         |ABCDEFGH |         |      ABCDEFGH     |         |         | ABCDEFG
target:    DIQMTQSPSSLSASLGGKVTITCKANQDIK----------KKIAWYQHKPGKGPRLLIYY--------TSTLKSGISSRFSGSGSG--------RDYSFSISNLEPEDAATYYCLQYDNFT-------
           ||||||||||||||||||||||||| |||           | |||||||||||||||| |        ||||  || |||||||||        |||||||||||||| ||||||||||||
template:  DIQMTQSPSSLSASLGGKVTITCKASQDIN----------KYIAWYQHKPGKGPRLLIHY--------TSTLHPGIPSRFSGSGSG--------RDYSFSISNLEPEDIATYYCLQYDNFT-------
```

Heavy Chain Target - Template alignment

```
           1        10        20        30        33        43                  55        65        75                  92
           |        |         |         | ABCDEFG |         |          ABCDEFGH |         |         |      ABC          |        ABCDEFGHIJ
target:    EVQLQQSGAELVKPGASVKLSCTASGFNIKD-------TYMHWVKQRPEQGLEWIGRIDP-------ANGYSKYDPKFQGKATITADTSSNAAYLQLSSLTSEDTAVYFCARDYEGF-----------
           |||||||||||||||||||||||||||||||       ||||||||||||||||||||||        |   ||||||||||||||||||| |||||||||||||||| | |  |||
template:  EVQLQQSGAELVKPGASVKLSCTASGFNIKD-------TYMHWVKQRPEQGLEWIGRIDP-------LNDKTKYDPKFQGKATITADTSSNSAYLQLSSLTSEDTAVYYCSRGWEGF-----------
```

http://biocomputing.it/pigspro

SWISS-MODEL: comparative modelling of protein structures and complexes



Start a New Modelling Project

Target Sequence:
(Format must be FASTA, Clustal, plain string, or a valid UniProtKB AC)

`Paste your target sequence(s) or UniProtKB AC here`

Project Title: Untitled Project

Email: Optional

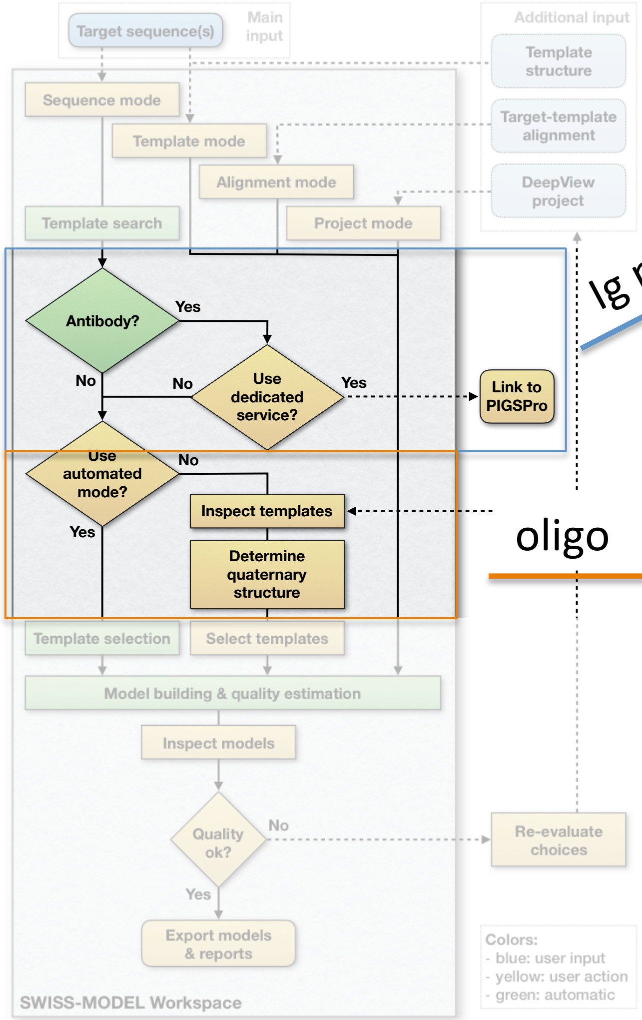
By using the SWISS-MODEL server, you agree to comply with the following terms of use and to cite the corresponding articles.

You are currently not logged in - to take advantage of the workspace, please log in or create an account.

(There is no requirement to create an account to use any part of SWISS-MODEL, however you will gain the benefit of seeing a list of your previous modelling projects here.)

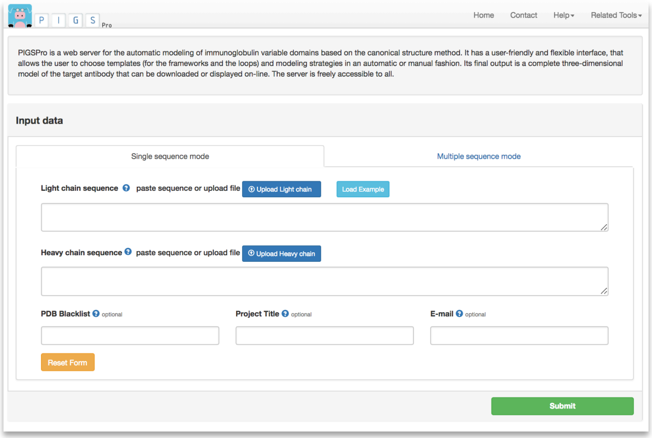
https://swissmodel.expasy.org/

SWISS-MODEL: comparative modelling of protein structures and complexes



Ig modeling

oligo



PIGSPro is a web server for the automatic modeling of immunoglobulin variable domains based on the canonical structure method. It has a user-friendly and flexible interface, that allows the user to choose templates (for the frameworks and the loops) and modeling strategies in an automatic or manual fashion. Its final output is a complete three-dimensional model of the target antibody that can be downloaded or displayed on-line. The server is freely accessible to all.

Input data

Single sequence mode | Multiple sequence mode

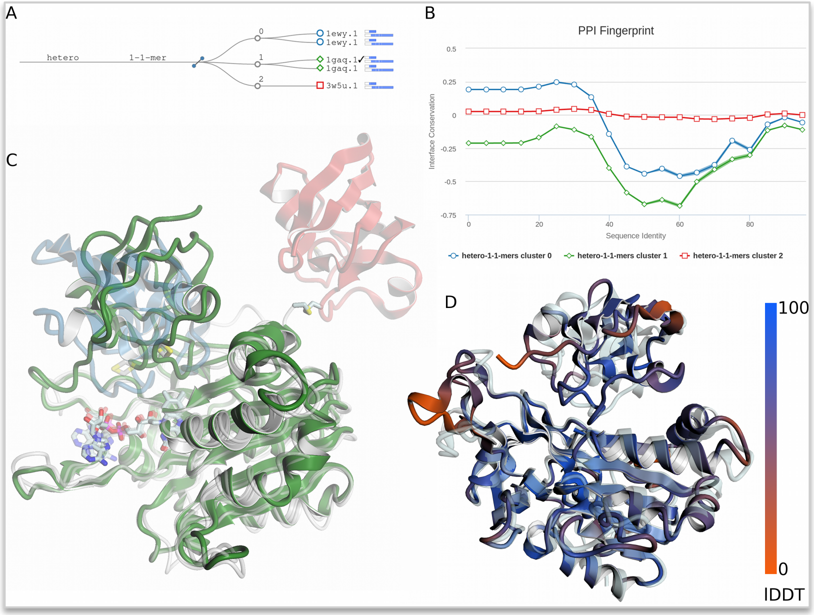
Light chain sequence paste sequence or upload file [Upload Light chain] [Load Example]

Heavy chain sequence paste sequence or upload file [Upload Heavy chain]

PDB Blacklist optional **Project Title** optional **E-mail** optional

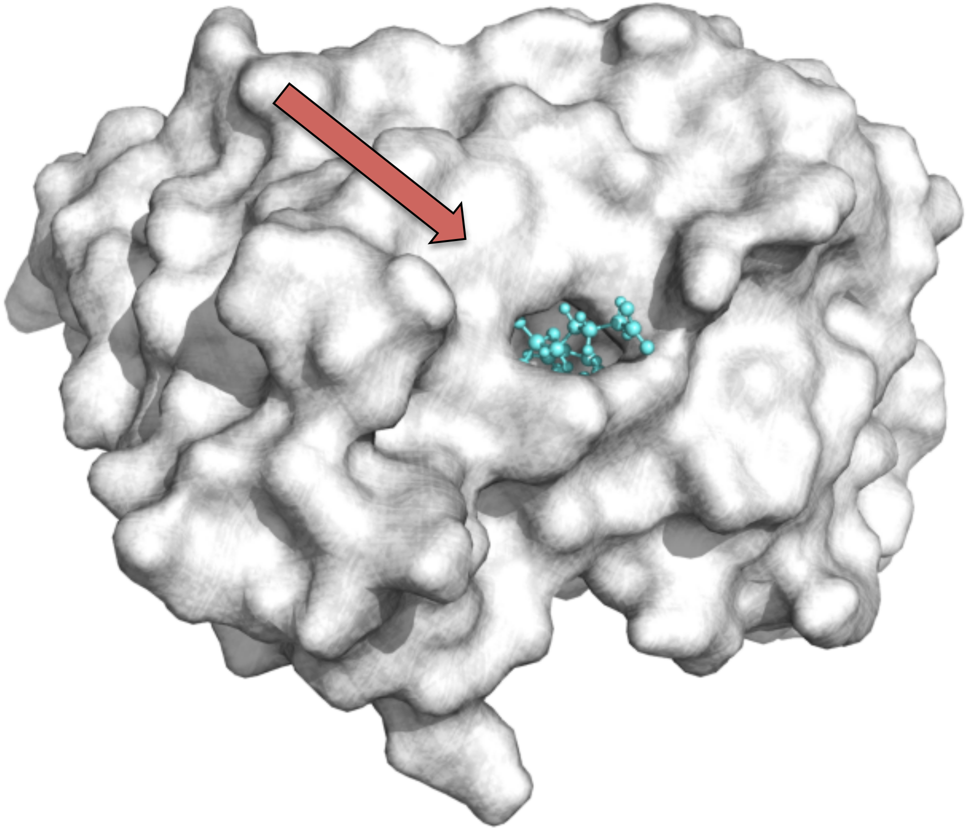
[Reset Form]

[Submit]

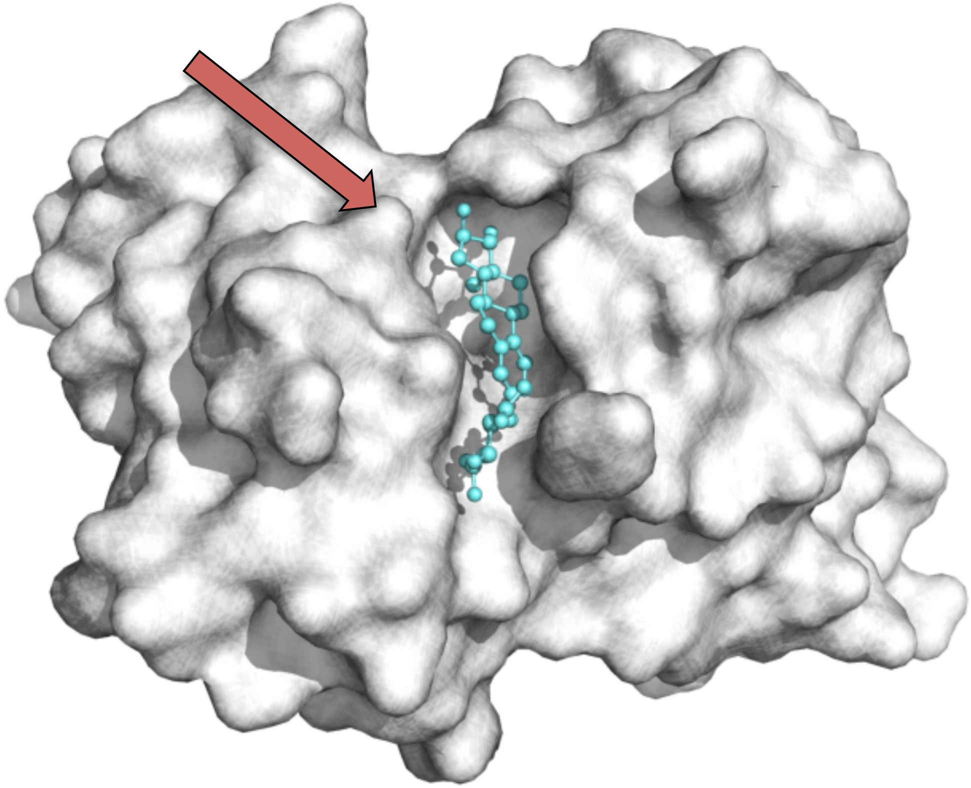


Waterhouse et al. Nucl Acids Res 2018

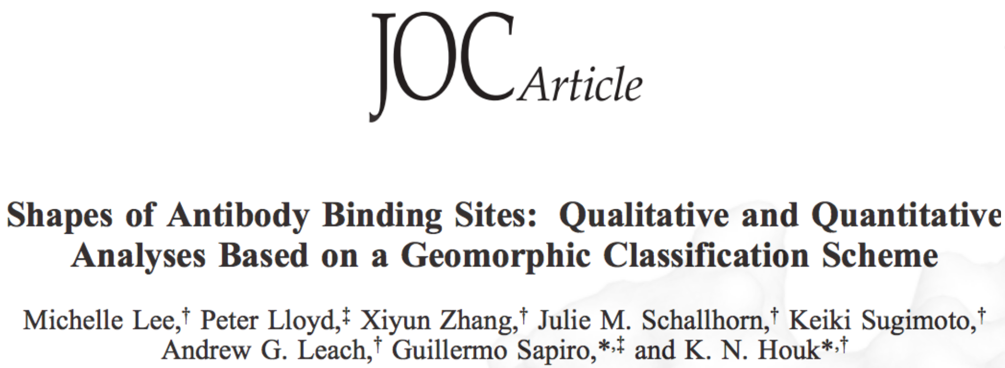
CDR structure vs function



CDR structure vs function



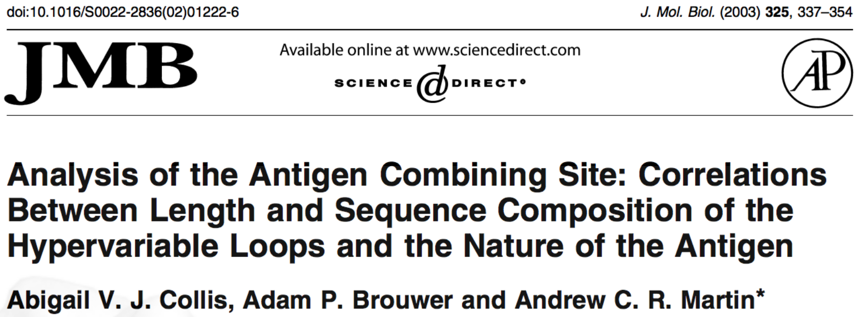
CDR structure vs function



JOC Article

Shapes of Antibody Binding Sites: Qualitative and Quantitative Analyses Based on a Geomorphic Classification Scheme

Michelle Lee,[†] Peter Lloyd,[‡] Xiyun Zhang,[†] Julie M. Schallhorn,[†] Keiki Sugimoto,[†] Andrew G. Leach,[†] Guillermo Sapiro,*[,‡] and K. N. Houk*[,†]

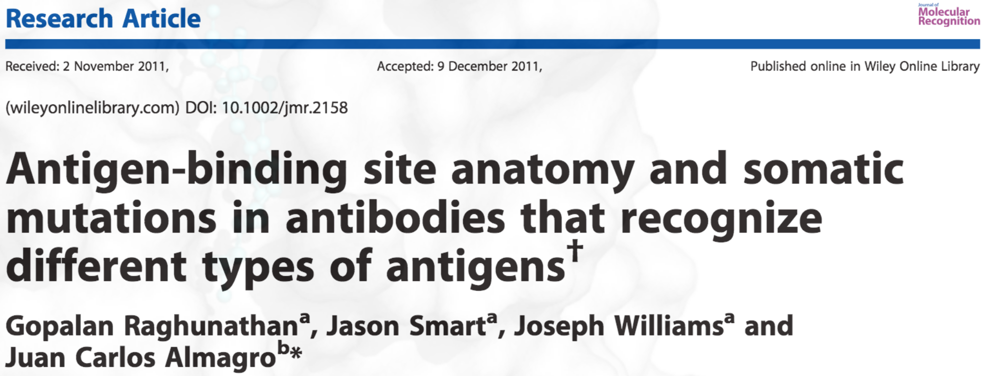


doi:10.1016/S0022-2836(02)01222-6 *J. Mol. Biol.* (2003) **325**, 337–354

JMB Available online at www.sciencedirect.com SCIENCE DIRECT®

Analysis of the Antigen Combining Site: Correlations Between Length and Sequence Composition of the Hypervariable Loops and the Nature of the Antigen

Abigail V. J. Collis, Adam P. Brouwer and Andrew C. R. Martin*



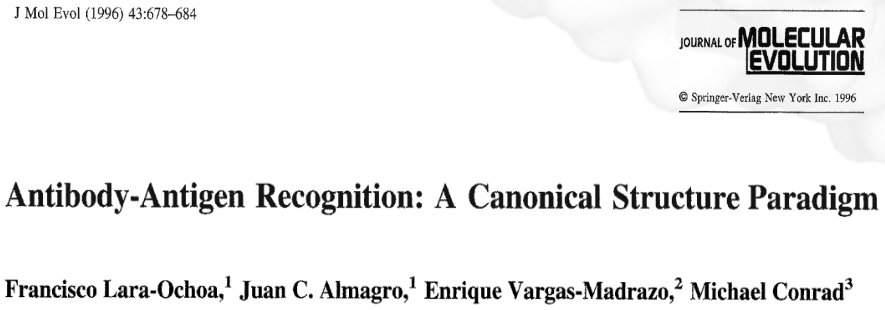
Research Article Molecular Recognition

Received: 2 November 2011, Accepted: 9 December 2011, Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/jmr.2158

Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens[†]

Gopalan Raghunathan[a], Jason Smart[a], Joseph Williams[a] and Juan Carlos Almagro[b]*



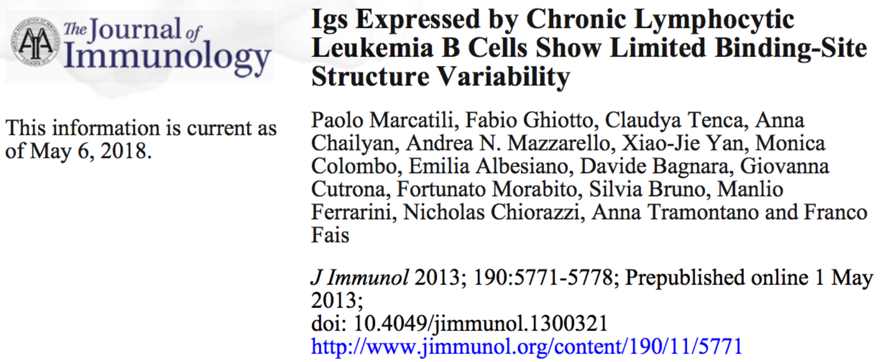
J Mol Evol (1996) 43:678–684

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1996

Antibody-Antigen Recognition: A Canonical Structure Paradigm

Francisco Lara-Ochoa,[1] Juan C. Almagro,[1] Enrique Vargas-Madrazo,[2] Michael Conrad[3]



The Journal of Immunology

This information is current as of May 6, 2018.

Igs Expressed by Chronic Lymphocytic Leukemia B Cells Show Limited Binding-Site Structure Variability

Paolo Marcatili, Fabio Ghiotto, Claudya Tenca, Anna Chailyan, Andrea N. Mazzarello, Xiao-Jie Yan, Monica Colombo, Emilia Albesiano, Davide Bagnara, Giovanna Cutrona, Fortunato Morabito, Silvia Bruno, Manlio Ferrarini, Nicholas Chiorazzi, Anna Tramontano and Franco Fais

J Immunol 2013; 190:5771-5778; Prepublished online 1 May 2013;
doi: 10.4049/jimmunol.1300321
http://www.jimmunol.org/content/190/11/5771

Shape analysis and specificity of CDR

Different loops forming the binding site can have **different lenght and structures** and yet result in a **similar binding site** surface

Aim is to **compare the actual surface** rather than the atomic coordinates of the CDRs

3D Zernike polynomials

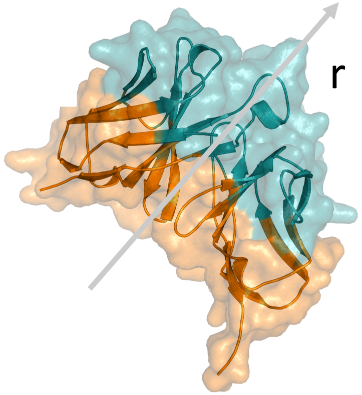
Image retrieval

Shape retrieval

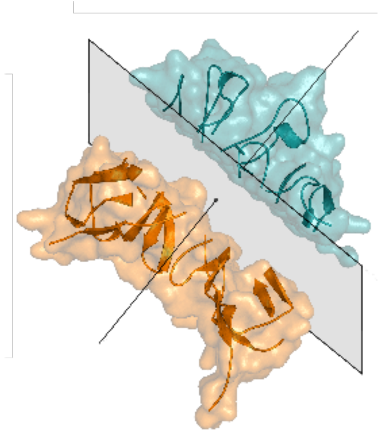
“descriptive power providing a basis for similarity measure between three-dimensional objects which is close to the human notion of resemblance”

Novotni and Klein. ACM Symposium on Solid Modeling and Applications, 2003

Shape analysis and specificity of CDR



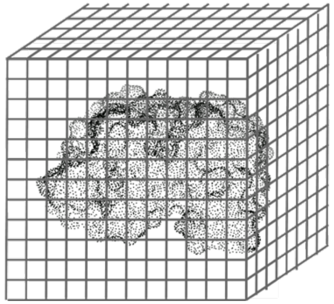
Centroids of conserved amino acids coordinates



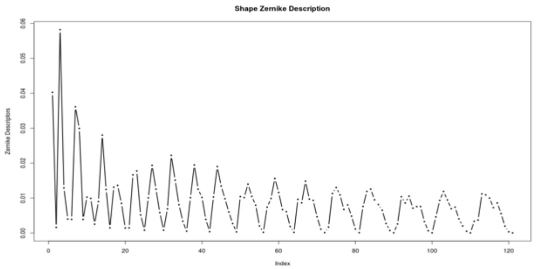
Plane perpendicular to r

CDR

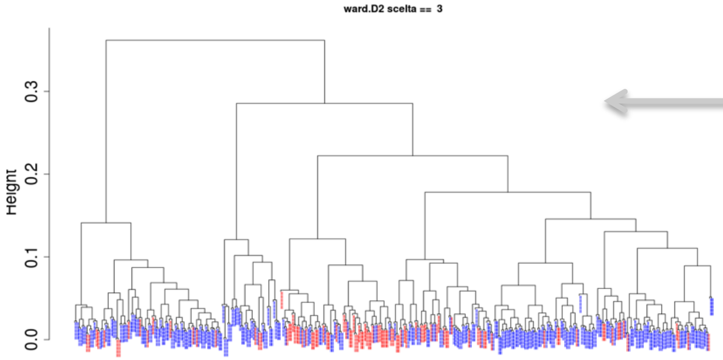
1 if d< 1.7 A
0 otherwise



Voxelization of the CDR surface



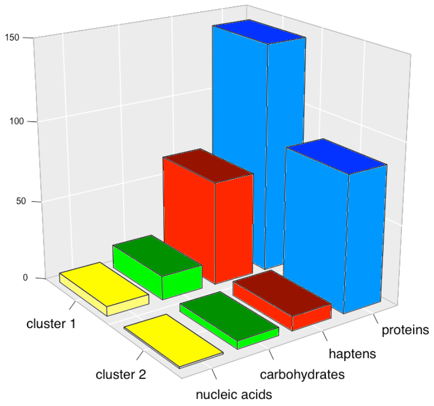
$$f(r, \theta, \varphi) = \sum_{n=0}^{\infty} \sum_{l=0}^{n} \sum_{m=-l}^{l} C_{nlm} Z_{nl}^{m}(r, \theta, \varphi)$$



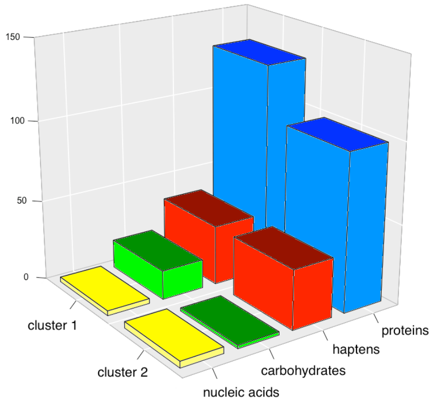
Shape describes CDR similarities better than TM

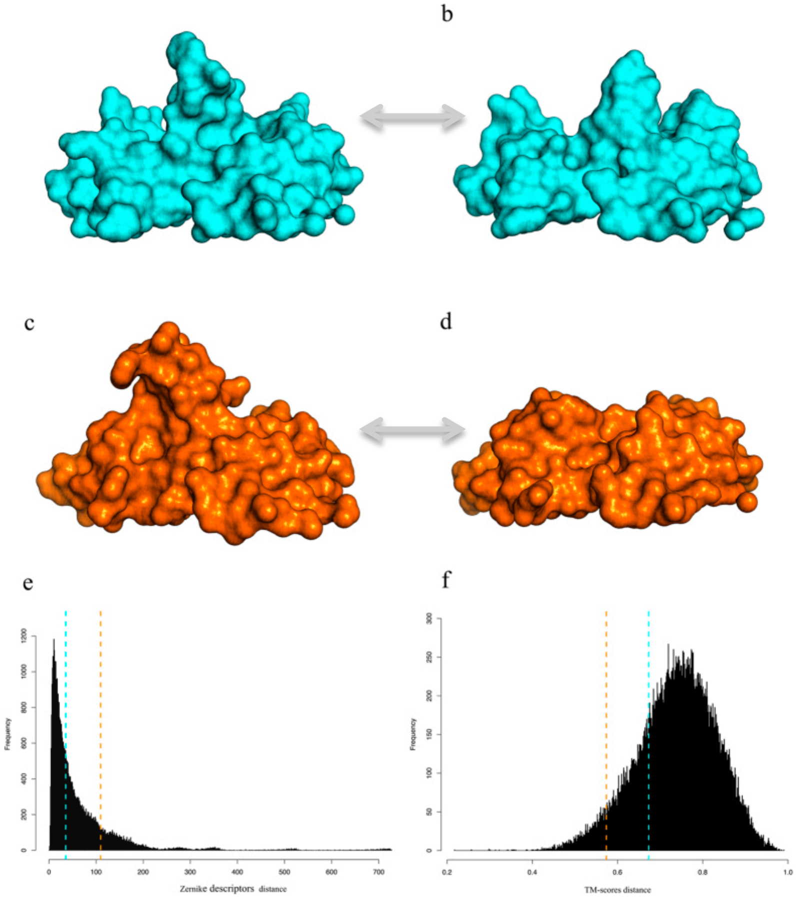
Dataset: 329 non redundant antibodies in complex with antigens

Shape similarity



Structure similarity (TM-score)

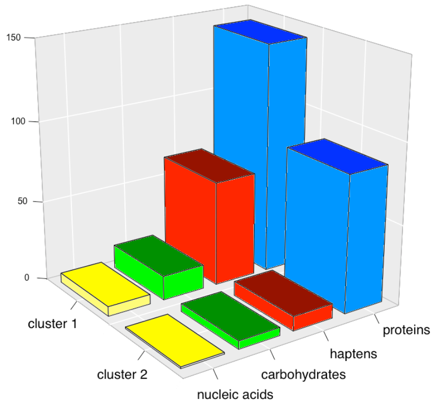




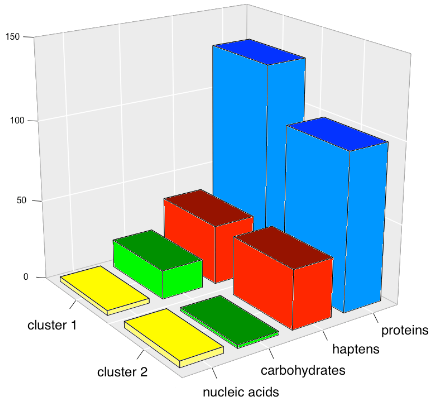
Shape analysis and specificity of CDR

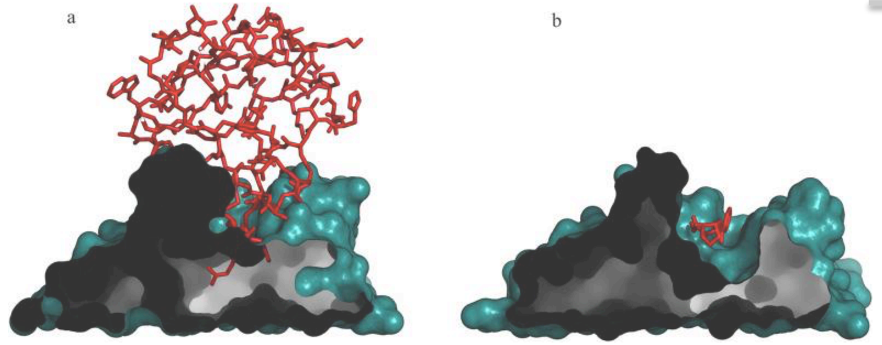
Dataset: 329 non redundant antibodies in complex with antigens

Shape similarity

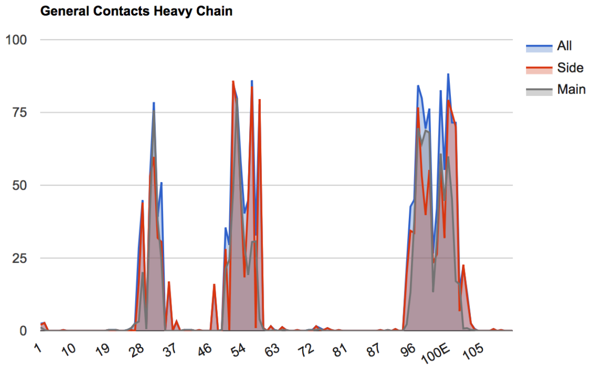


Structure similarity (TM-score)





Predicted Antibody contacts



Accuracy ~ 76%

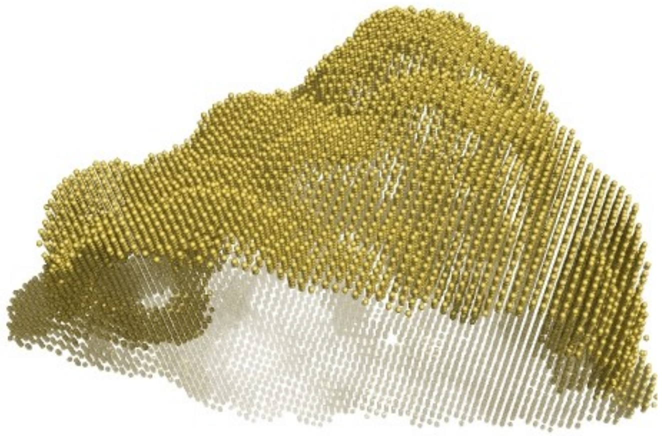
Di Rienzo et al, Scientific Reports 2017

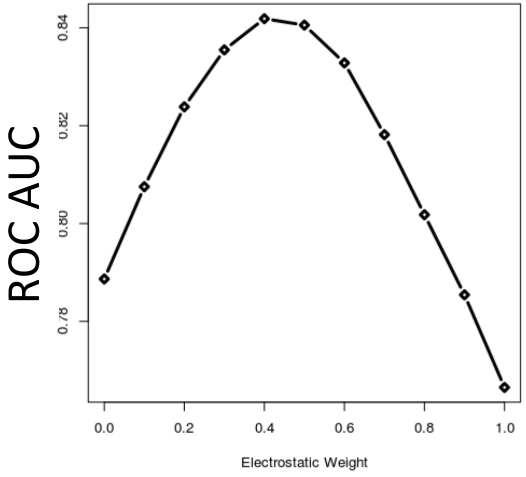
Shape analysis and specificity of CDR

Better definition of CDR shape: no artificial plane

~70% accuracy based on shape only

78% accuracy including electrostatic potential





Area under the curve