Recent applications of bioinformatics
to antibody engineering

Troubleshooting Problematic Antibody Variable Domain Sequences

Antibody Troubleshooter and Graft Design Templates

Annemarie Honegger,
PEGS Europe Protein & Antibody Engineering Summit 7/11/2013

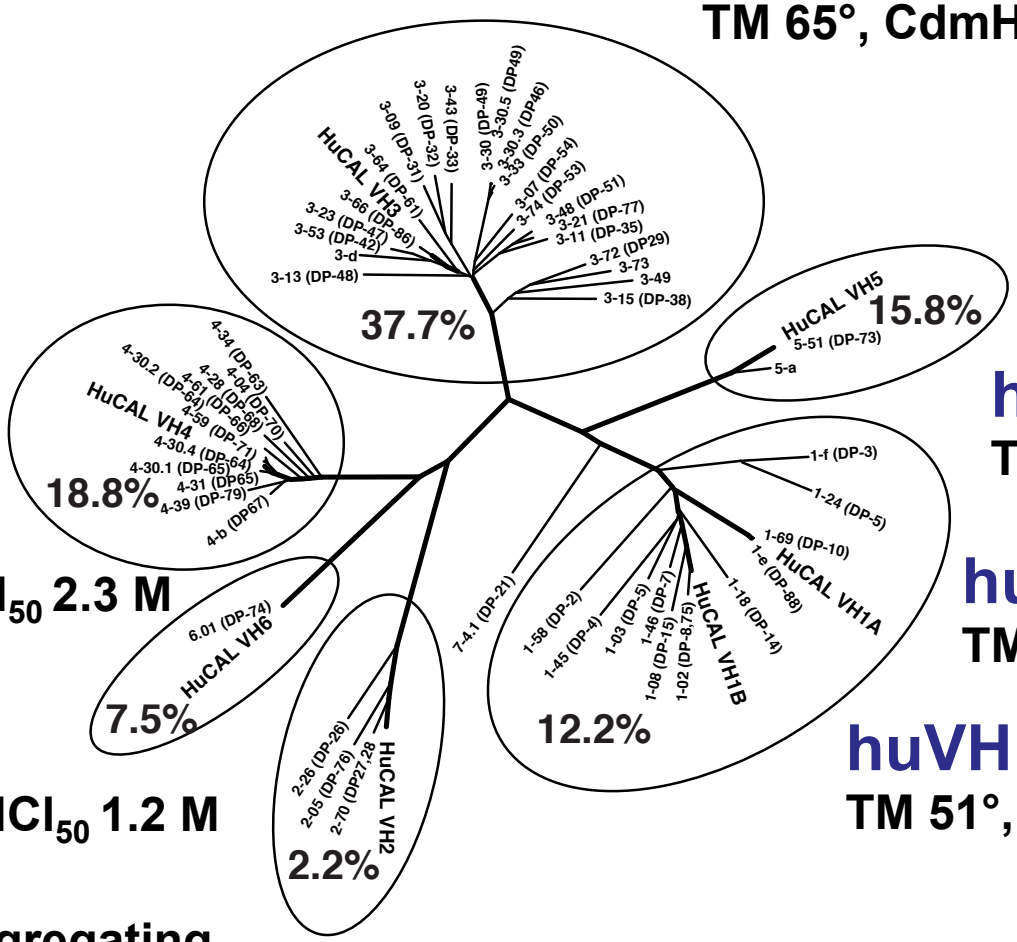
FAQ

(Frequently Asked Question)

Do well expressed scFv have any features in common that distinguish them from poorly expressed, aggregation-prone scFv?

Not All Antibody Variable Domains Are Created Equal

huVH3: monomeric
TM 65°, $CdmHCl_{50}$ 3.0 M



huVH4:
aggregating,
TM 44°, $GdmHCl_{50}$ 2.3 M

huVH6:
aggregating,
TM 39°, $GdmHCl_{50}$ 1.2 M

huVH2: aggregating
TM 45°, $GdmHCl_{50}$ 1.4 M

huVH5: monomeric
TM 44°, $GdmHCl_{50}$ 2.2 M

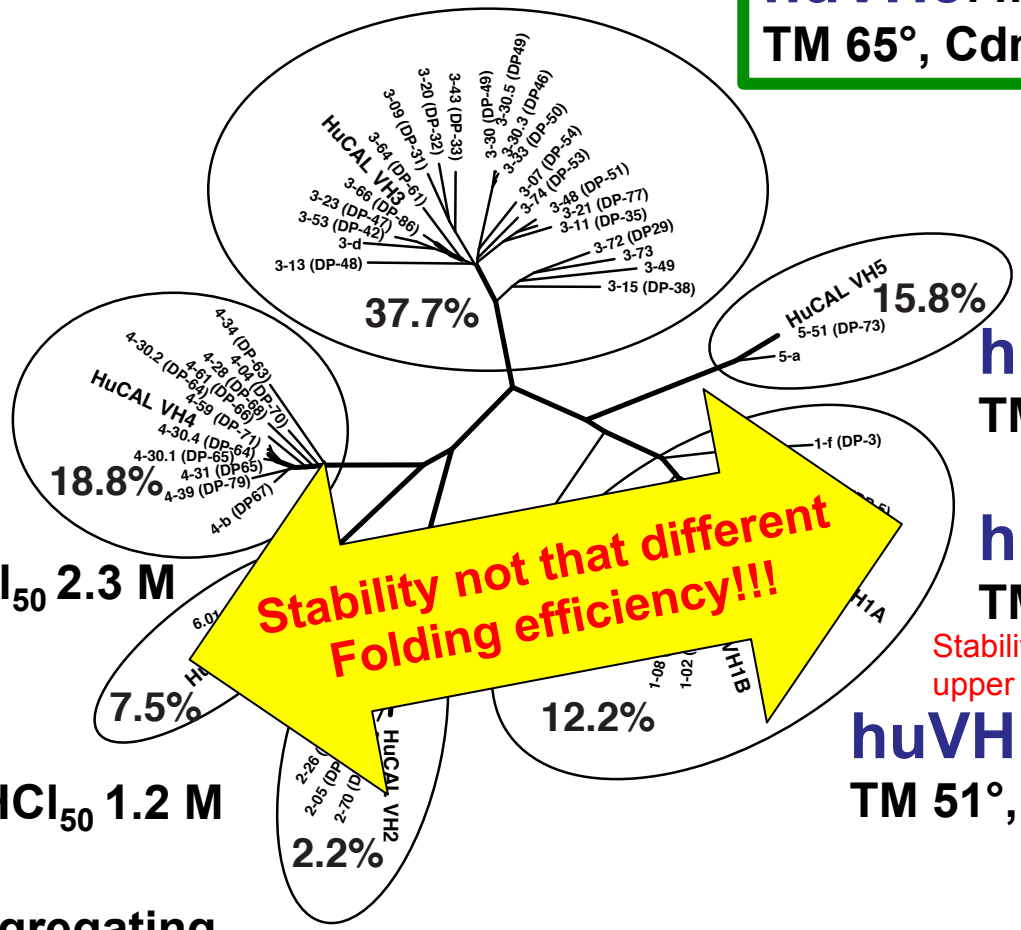
huVH1a: monomeric
TM 41°, $GdmHCl_{50}$ 1.5 M

huVH1b: monomeric
TM 51°, $GdmHCl_{50}$ 2.1 M

Ewert et al., J.Mol. Biol. 325 (2003) 531-553
Knappik et al. J. Mol. Biol. 296 (2000) 57-86

Not All Antibody Variable Domains Are Created Equal

huVH3: monomeric
TM 65°, $CdmHCl_{50}$ 3.0 M



huVH5: monomeric
TM 44°, $GdmHCl_{50}$ 2.2 M

huVH4:
aggregating,
TM 44°, $GdmHCl_{50}$ 2.3 M

huVH1a: monomeric
TM 41°, $GdmHCl_{50}$ 1.5 M
Stability difference due to very poor upper core and dimer interface packing

huVH6:
aggregating,
TM 39°, $GdmHCl_{50}$ 1.2 M

huVH1b: monomeric
TM 51°, $GdmHCl_{50}$ 2.1 M

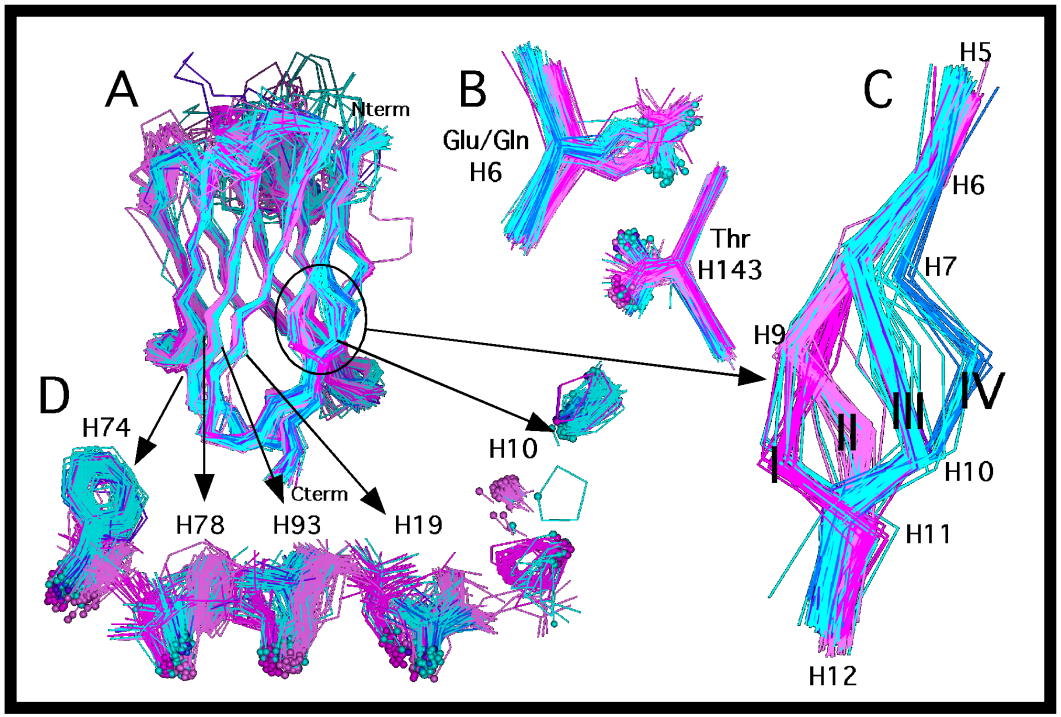
huVH2: aggregating
TM 45°, $GdmHCl_{50}$ 1.4 M

Ewert et al., J.Mol. Biol. 325 (2003) 531-553
Knappik et al. J. Mol. Biol. 296 (2000) 57-86

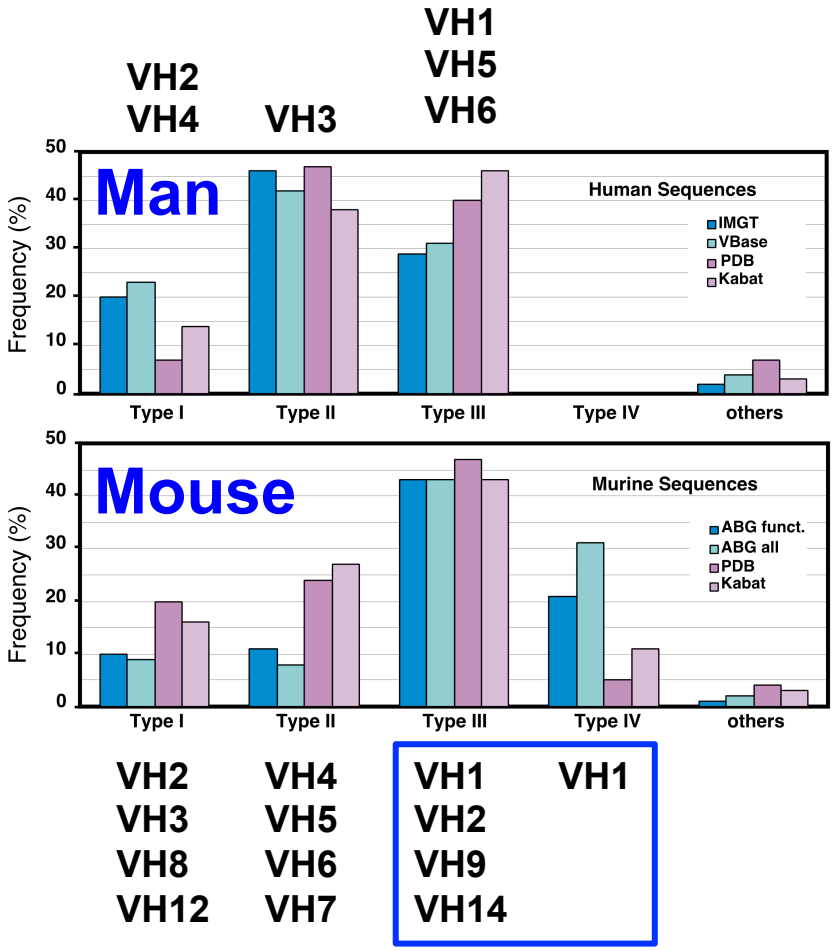
Stability of human consensus domains

Domain	yield mg/L_{OD10}	oligomeric state	T_m °C	$[GdmHCl]_{50}$ M	$\Delta G(H_2O)$ kJ/mol	m kJ L/mol^2
$hV_L\kappa1$	4.5	monomer	64	2.1	29	14
$hV_L\kappa2$	14.2	monomer	63	1.5	25	16
$hV_L\kappa3$	17.1	monomer	73	2.3	35	15
$hV_L\kappa4$	9.6	mono+dimer	58	1.5	*n.d.*	*n.d.*
$hV_L\lambda1$	0.3	monomer	64	2.1	24	11
$hV_L\lambda2$	1.9	monomer	50	1.0	16	16
$hV_L\lambda3$	0.8	mono+dimer	49	0.9	15	16
Domain	**yield mg/L_{OD10}**	**oligomeric state**	**T_a °C**	**$[GdmHCl]_{50}$ M**	**$\Delta G(H_2O)$ kJ/mol**	**m kJ L/mol^2**
hV_H1a	1.0	monomer	41	1.5	14	10
hV_H1b	1.2	monomer	51	2.1	26	13
hV_H2	refolded	*n.d.*	45	1.4	*n.d.*	*n.d.*
hV_H3	2.4	monomer	65	3.0	53	18
hV_H4	refolded	*n.d.*	44	2.3	*n.d.*	*n.d.*
hV_H5	refolded	monomer	44	2.2	17	7
hV_H6	refolded	*n.d*	39	1.2	*n.d.*	*n.d.*

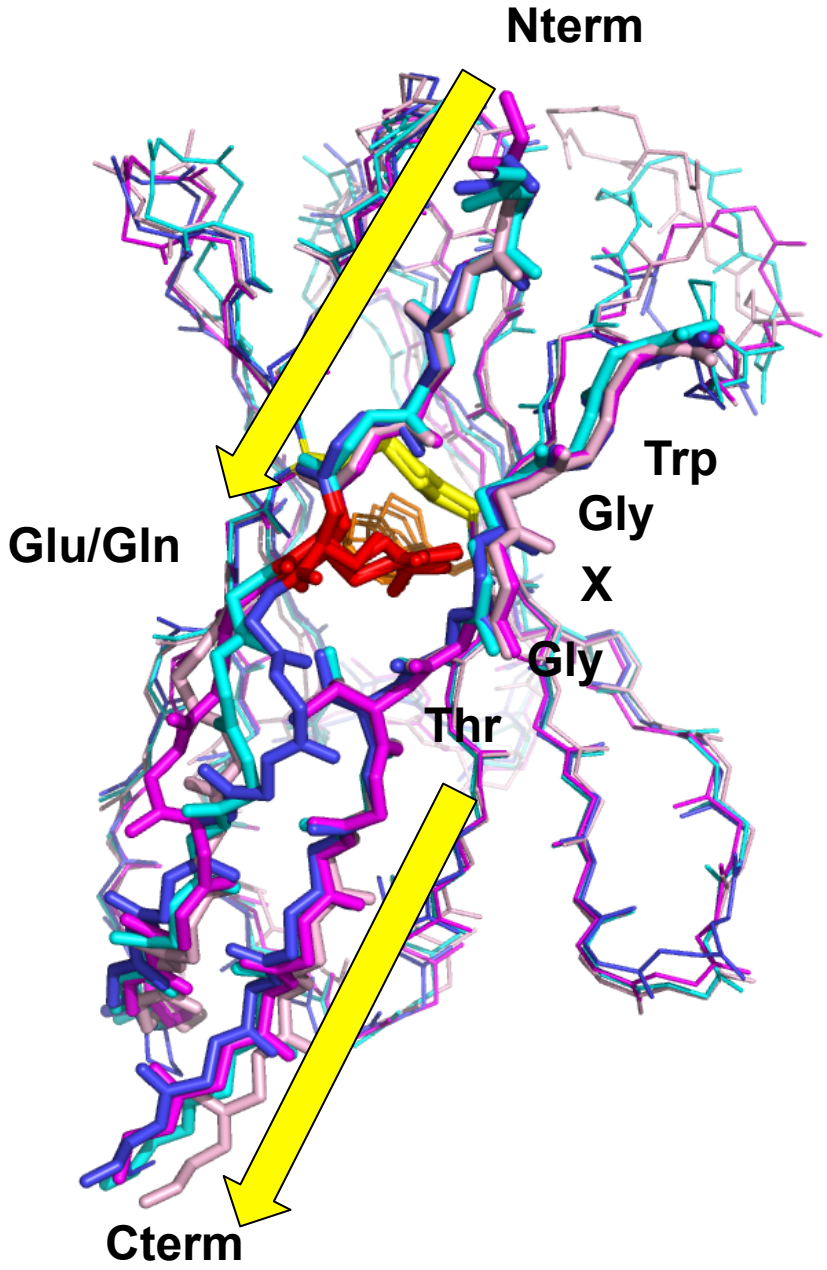
Distinct Structural Subclasses

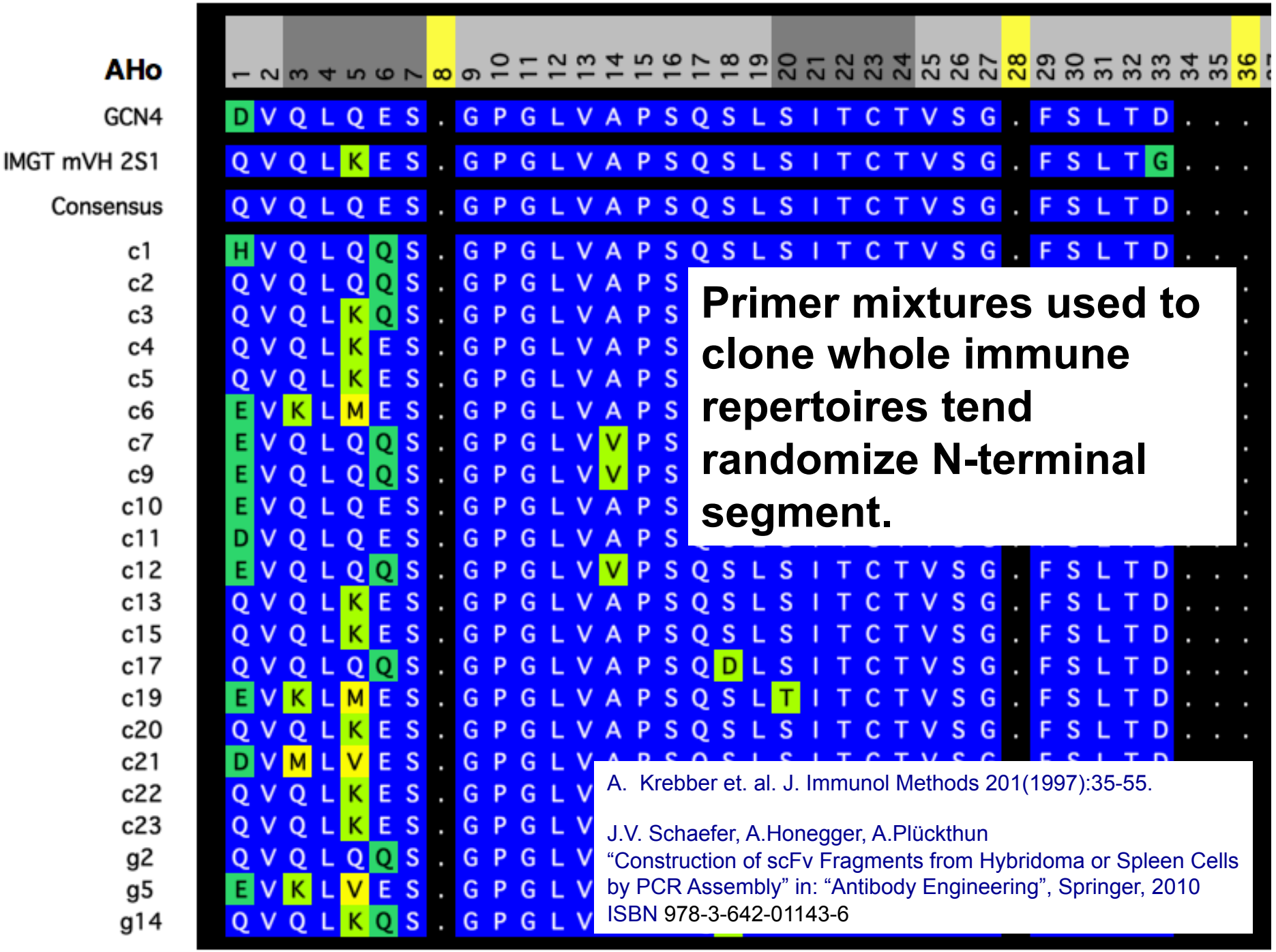


Framework conformation determined by H6, H7 and H10

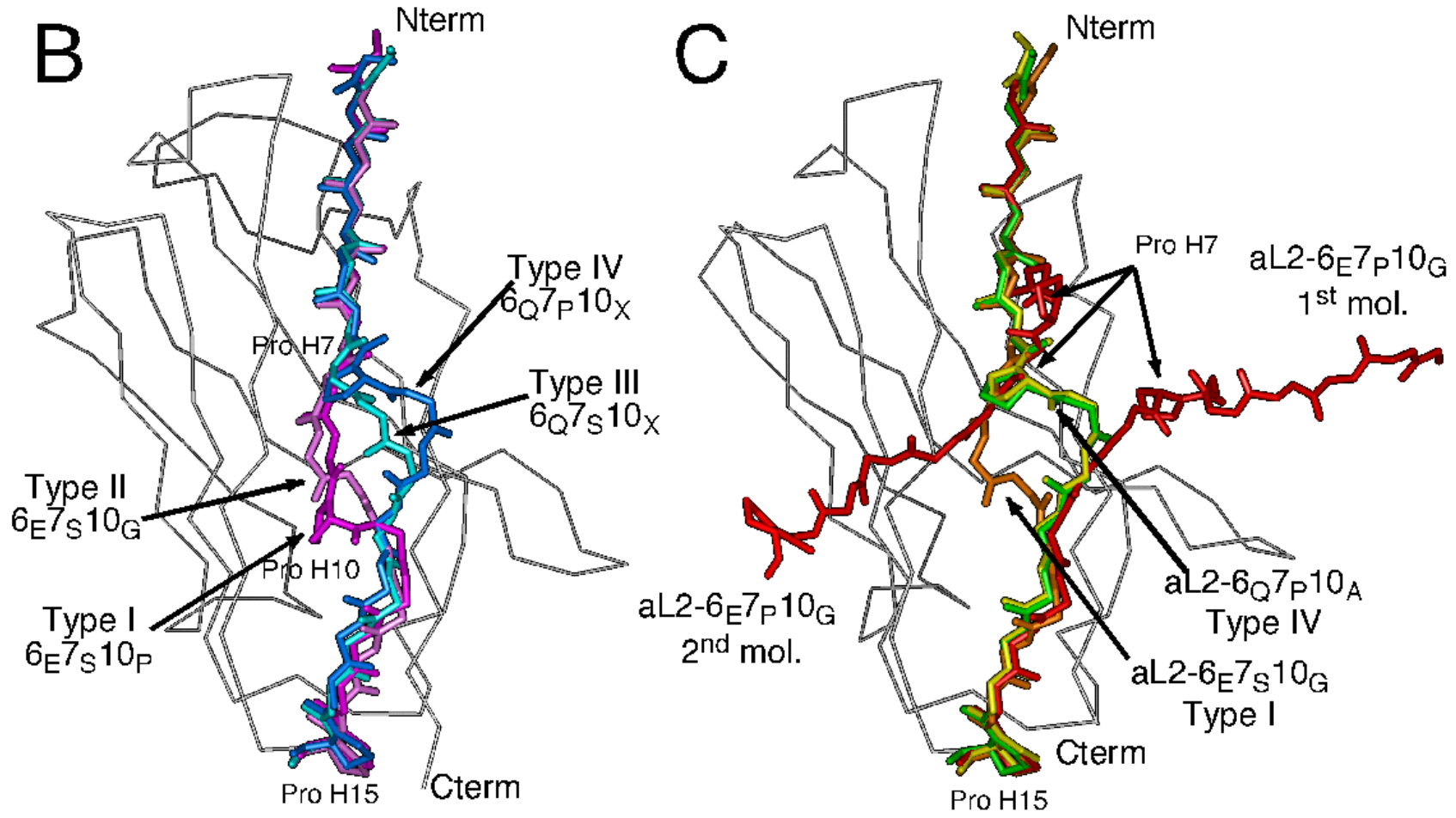


Honegger & Plückthun, J.Mol. Biol. 309 (2001) 687-699


Nterm
Trp
Gly
X
Gly
Thr
Glu/Gln
Cterm


AHo
GCN4
IMGT mVH 2S1
Consensus
c1
c2
c3
c4
c5
c6
c7
c9
c10
c11
c12
c13
c15
c17
c19
c20
c21
c22
c23
g2
g5
g14
Primer mixtures used to clone whole immune repertoires tend randomize N-terminal segment.
A. Krebber et. al. J. Immunol Methods 201(1997):35-55.
J.V. Schaefer, A.Honegger, A.Plückthun
“Construction of scFv Fragments from Hybridoma or Spleen Cells by PCR Assembly” in: “Antibody Engineering”, Springer, 2010
ISBN 978-3-642-01143-6

Out-of-context exchanges of H6,H7 and H10 are highly destabilizing



Check your sequence against the GL family!

Jung et al., J.Mol. Biol. 309 (2001) 701-716

DEMO

Compare your variable domain sequences to the closest germline sequences, and the germline familiy consensus in the species of origin

Compare your variable domain sequences to the human germline sequences and family consensus sequences if humanization is planned

FAQ

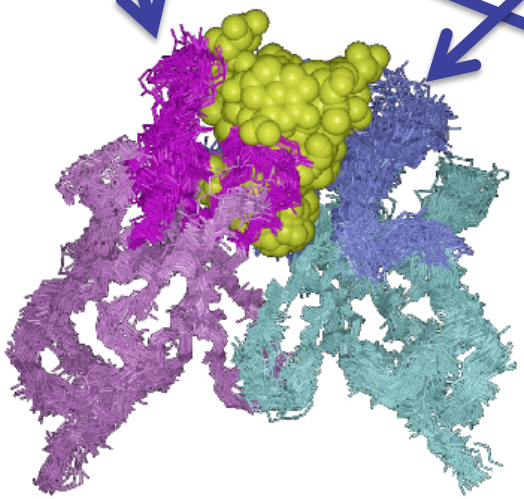
Why bother about poorly folding frameworks? Can’t we just eliminate those from consideration?

Long CDR L1
=> huVκ2/huVκ4

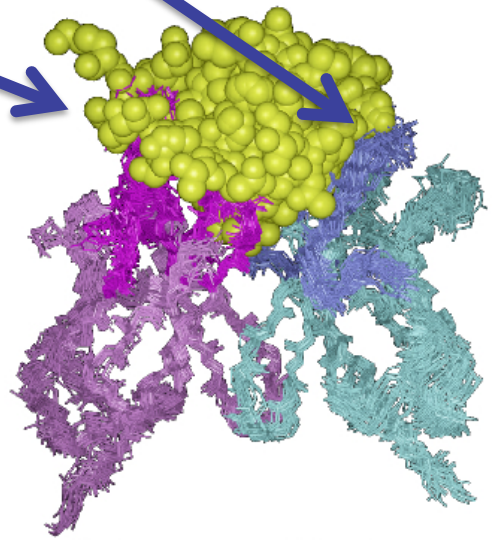
Long CDR H1
=> huV_H2/huV_H6

Long CDR H2
=> huV_H3/huV_H6

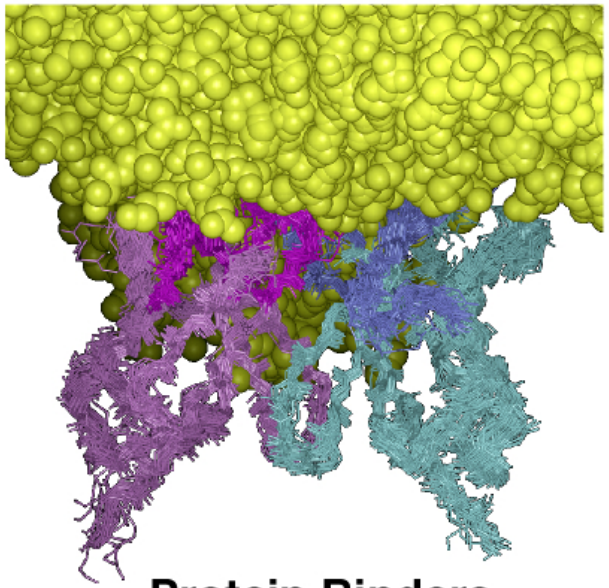
Short CDR L1/H1
Short CDR H2



Hapten Binders
(52 structures)



Oligomer Binders
(30 structures)

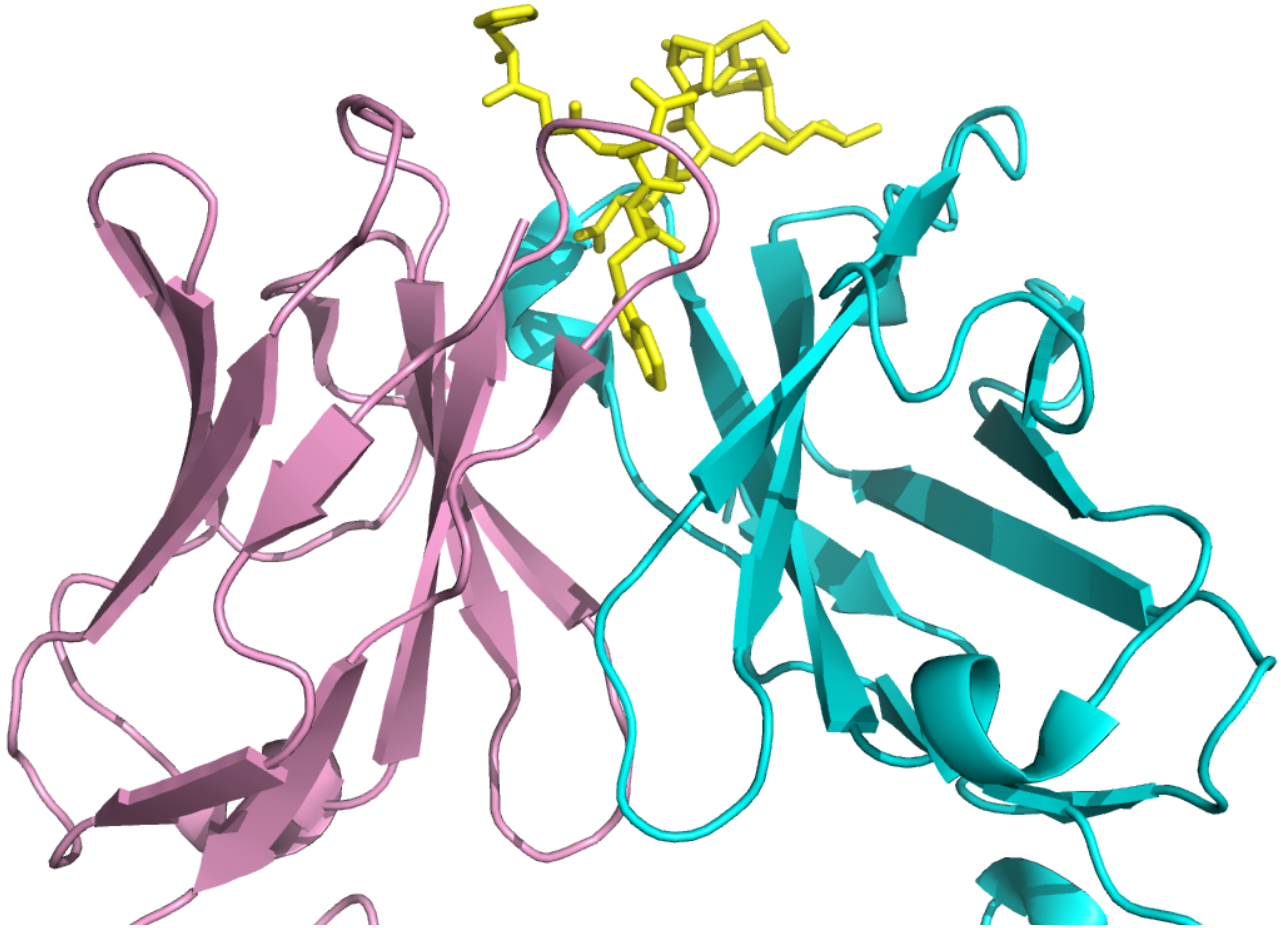


Protein Binders
(45 structures)

Short CDR H3
"open" CDR H3
conformation

Long CDR H3,
"closed" CDR H3
conformation

A typical “Hapten” Binder

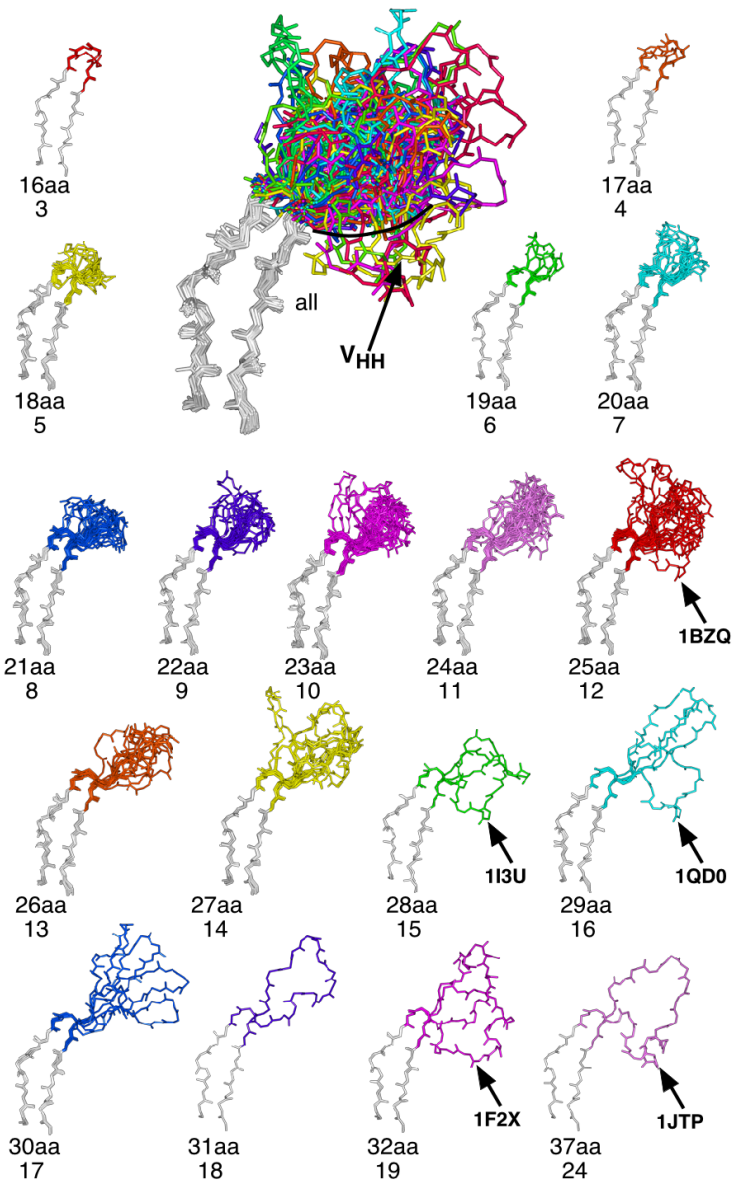


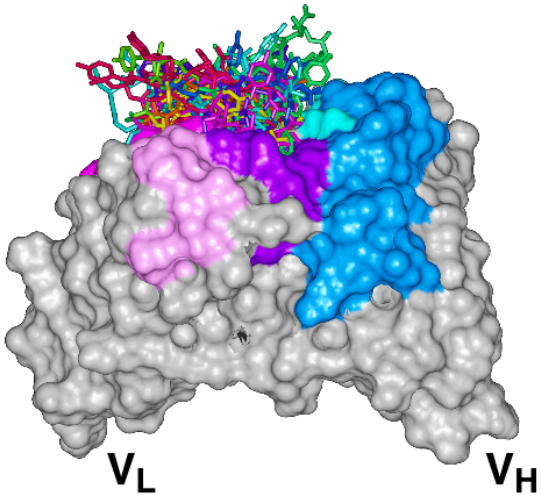
B.Luginbühl et al. *J Mol Biol* **363**, 75-97.

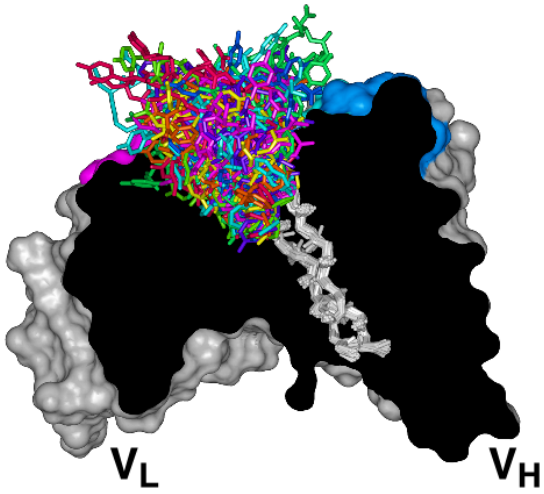
PDB entry 2HHO

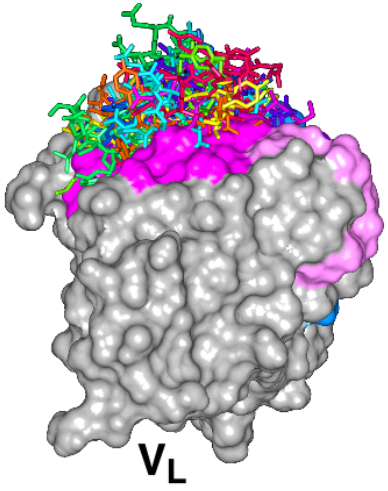
= low pM binder to a protruding loop of a protein

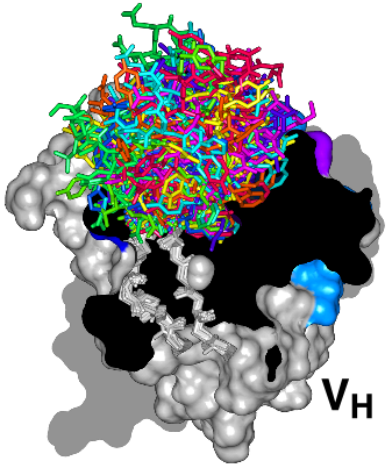
CDR H3

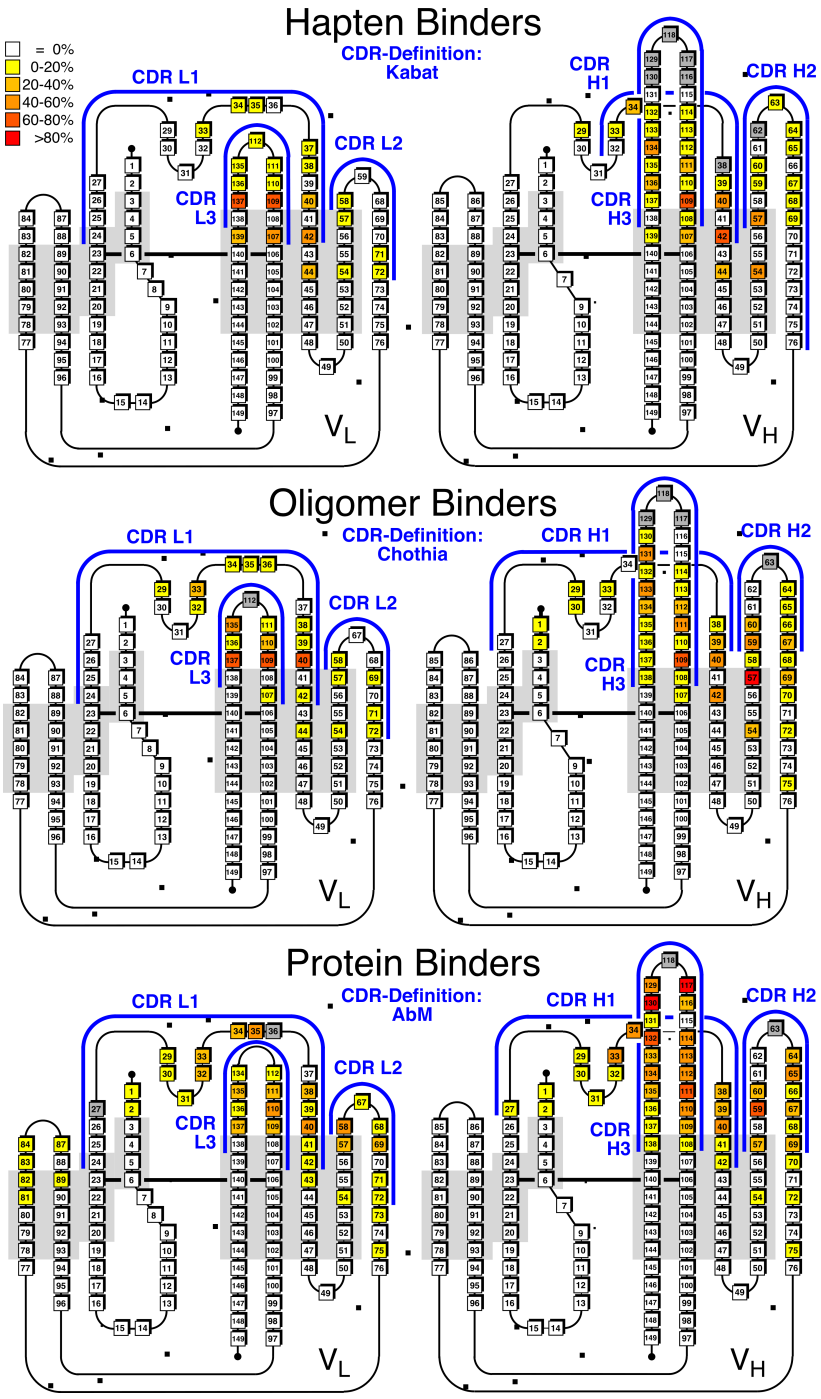

16aa
3
all
V_{HH}
17aa
4
18aa
5
19aa
6
20aa
7
21aa
8
22aa
9
23aa
10
24aa
11
25aa
12
1BZQ
26aa
13
27aa
14
28aa
15
1I3U
29aa
16
1QD0
30aa
17
31aa
18
32aa
19
1F2X
37aa
24
1JTP


V_L
V_H

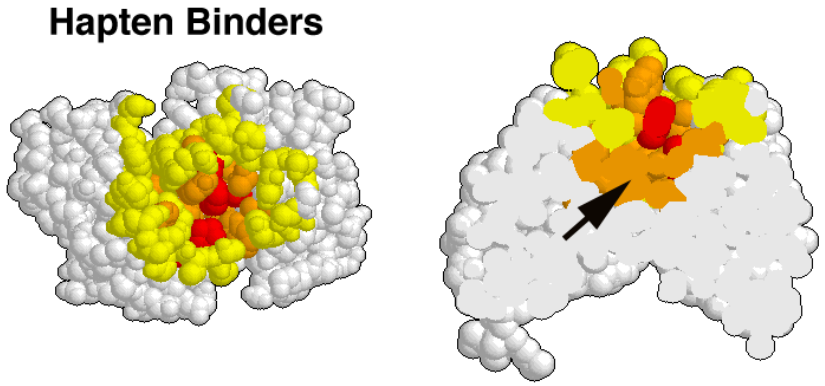

V_L
V_H

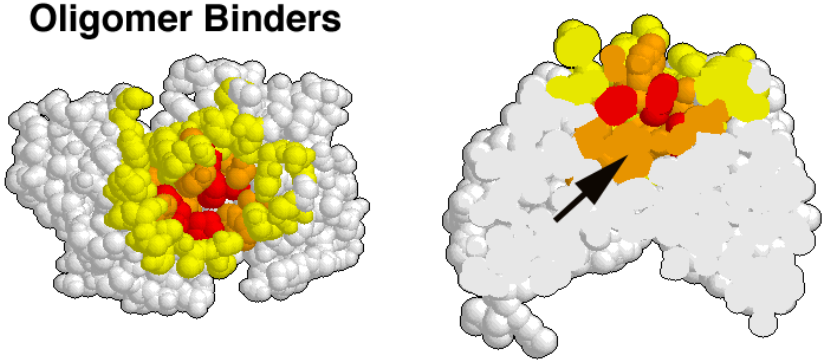

V_L

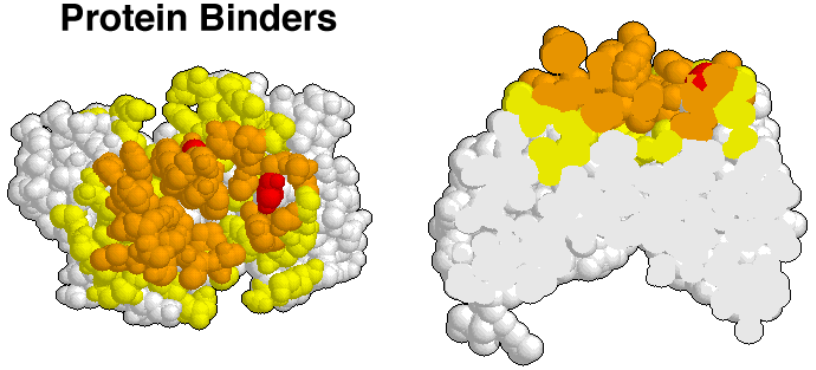

V_H



Antigen Contacts







FAQ

What can be done to improve poorly expressing variable domains?

Checklist for potential problems

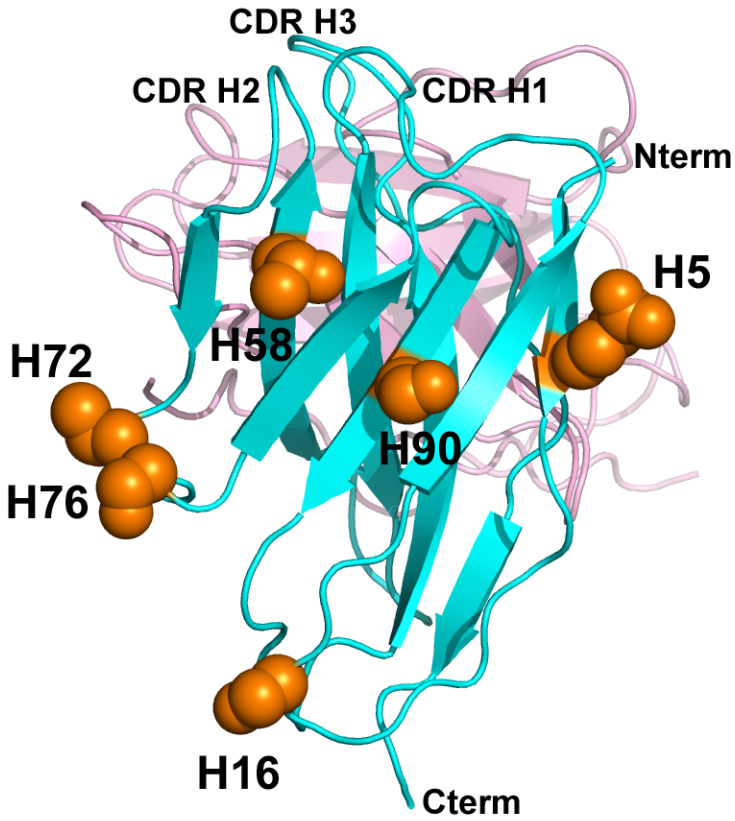
- **Hydrophobic core packing**
  steric clashes and cavities destabilize the domain,
  mutations to hydrophilic residues are destabilizing
- **Hydropathic contrast** between core and surface
  hydrophobic surface residue can decrease folding efficiency
- **Conserved hydrogen bonding interactions**
  core hydrogen bonding network (Glu/Gln 6, Thr 143, Tyr 104)
- **Conserved charge interactions**
  buried charge cluster (Arg 77, Asp 100, Glu 99, Arg/Gln 45, Glu/Arg 53)
- **Conserved unusual main-chain torsion angles**
  positions which enforce a positive Φ torsion angle,
  conserved Gly residues
- **Conserved Pro positions**
  cis-Pro L8 and and L136 of $V_L\kappa$,
  conserved *trans*-Pro in various positions
- **Secondary structure propensity and torsional preference**

Honegger Handbook Exp. Pharmacology 02/2008
Ewert et al. Methods 34(2004)184-199

Two scFv: 2C2 (V_K3-V_H6), 6B3 (V_λ3-V_H6)

	yield	stability
Gln H5 Val (secondary structure propensity	+	+
Ser H16 Gly (pos. Φ, conformational strain)	+	+
Thr H58 Ile (hydrophobic packing, to V_H consensus)	0	+
Val H72 Asp (exposed hydrophobic residue)	+	0
Ser H76 Gly (pos. Φ, conformational strain)	+	+
Ser H90 Tyr (semiexposed hydrophobic, to V_H cons.)	+	0

$[GdmHCl]_{50}$ shifted from 2.0 to 2.8 M
and from 0.7 to 2.5 M *
Total stabilisation by 21 and 25 kJ/mol
from 51 to 72 kJ/mol
and from 42 to 67 kJ/mol *
Total increase in yield 4.3 and 4.2-fold,
from 1.2 mg/L to 5 mg/L
and from 0.4 mg/L to 1.7 mg/L



Ewert et al. Biochemistry 42(2003)1517-1528
Ewert et al. J.Biol.Chem. 347 (2005) 773-789
Schaefer et al. Prot.Eng.Des.Sel. 25 (2012) 485-506

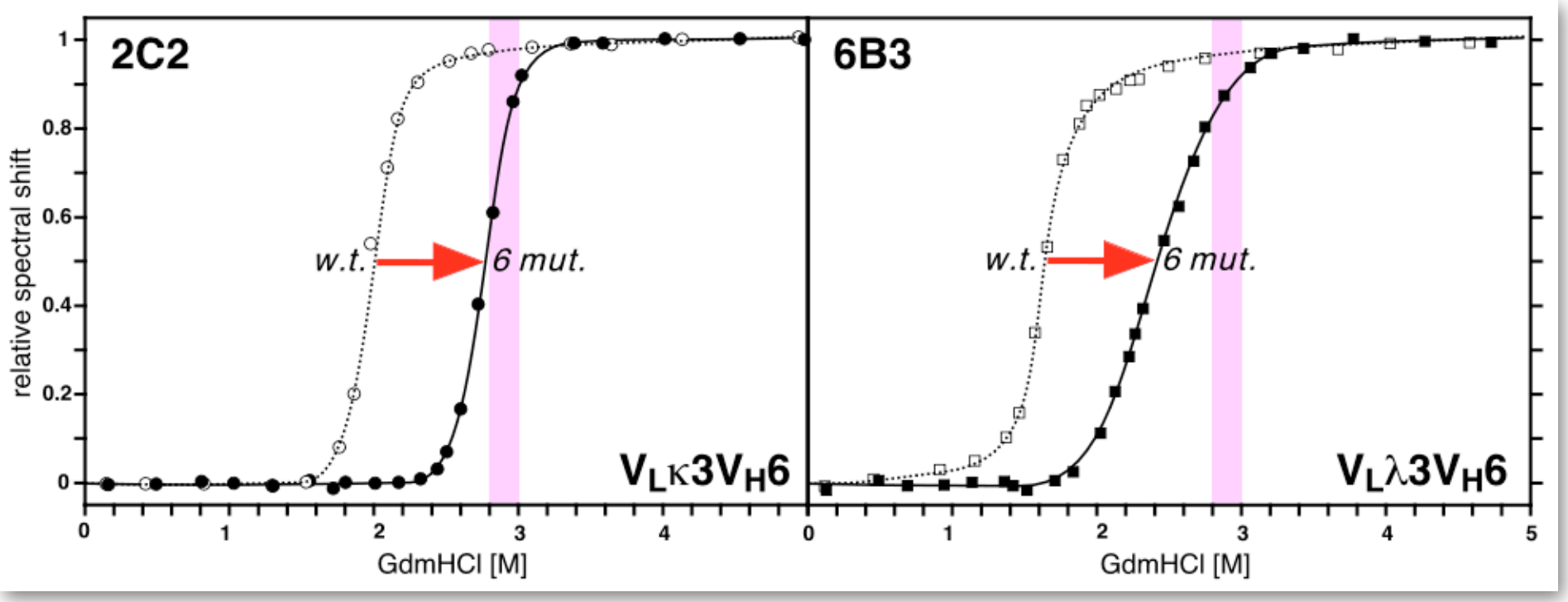
Improving the hV_H6 HuCAL framework

Six mutations were needed, five of them common to hV_H2, hV_H4 and hV_H6:

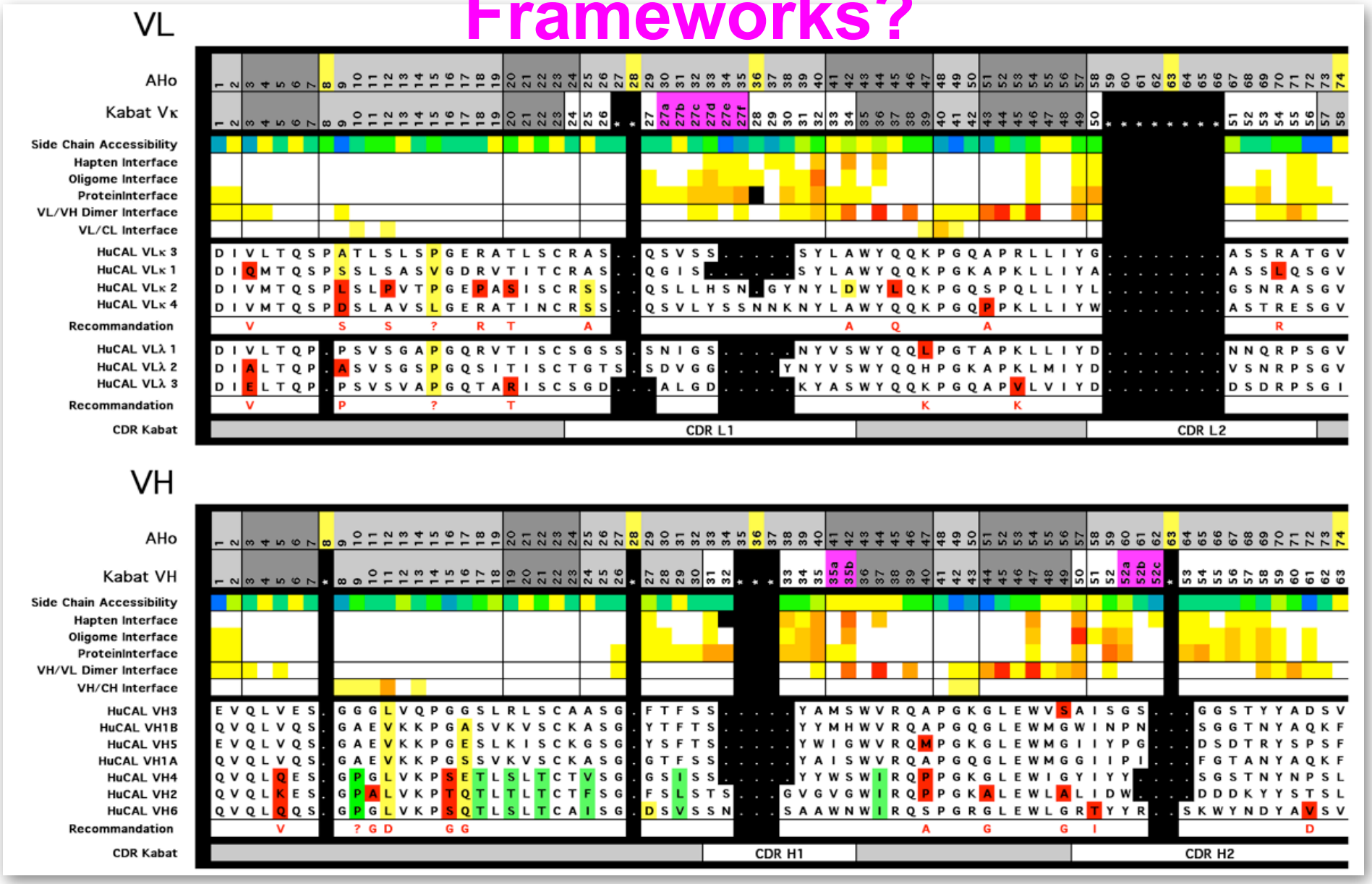
Three mutations improve both stability and yield

Two improve the folding yield, but have no measurable effect on thermodynamic stability

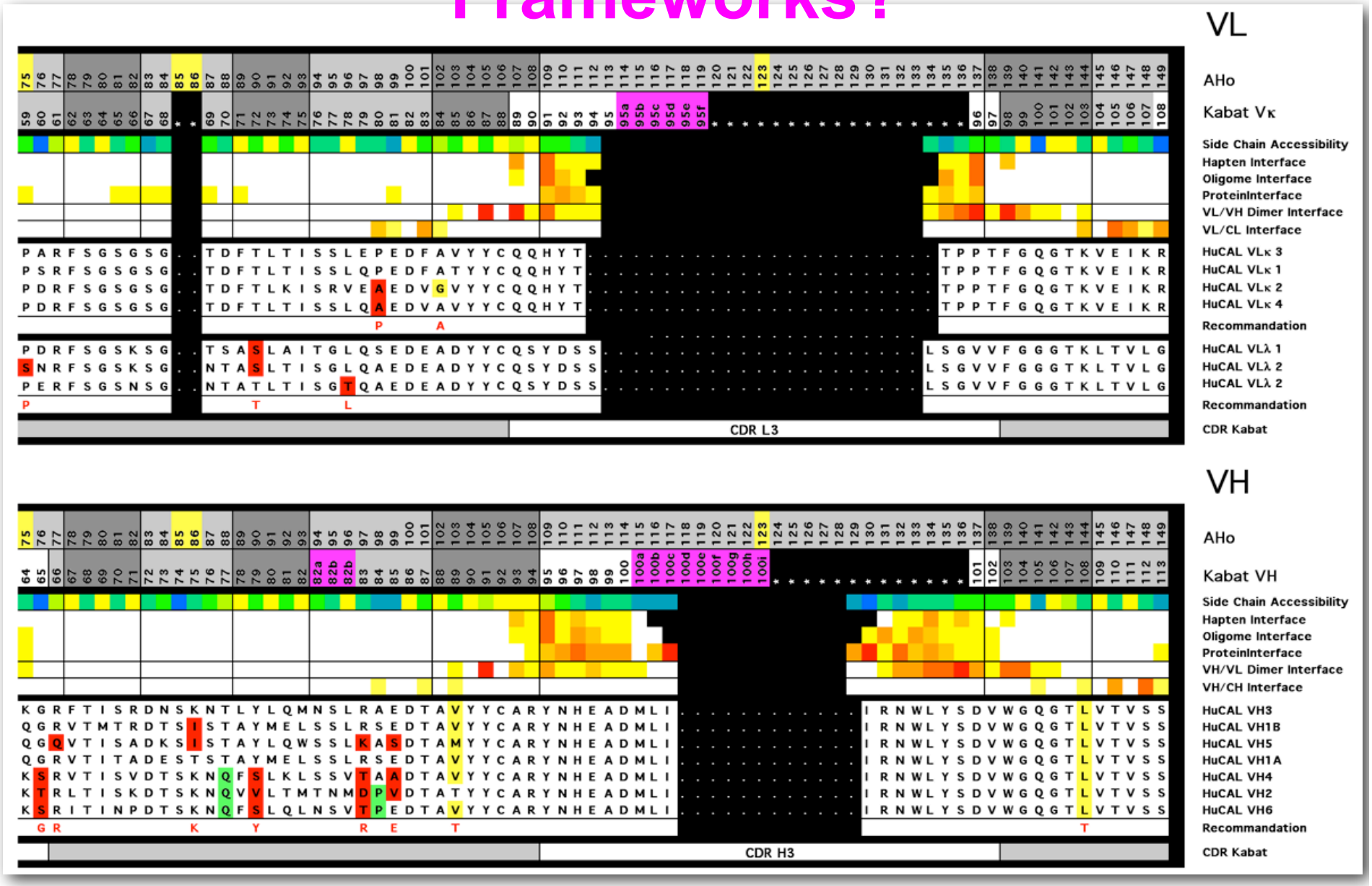
One significantly improves stability without affecting the yield



How about the other Consensus Frameworks?



How about the other Consensus Frameworks?



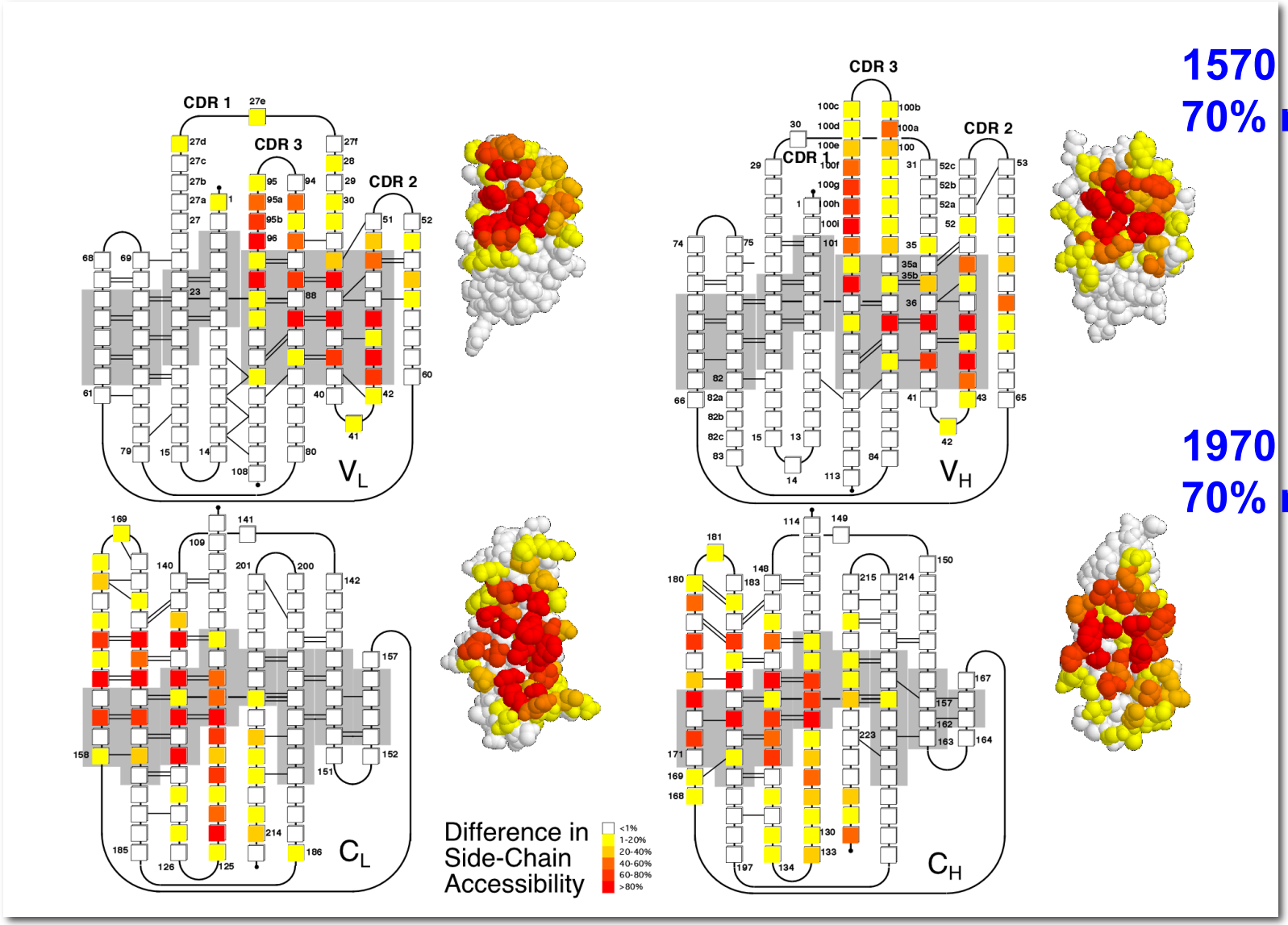
DEMO

Compare your variable domain sequences to the troubleshooting checklist

FAQ:

Does stability and folding efficiency of an individual variable domain matter for Fab or whole IgG expressed in mammalian cells?

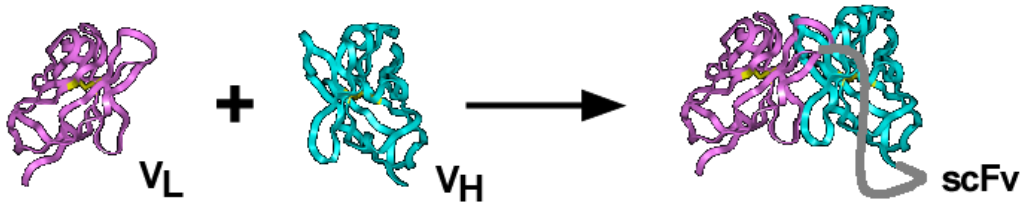
V_L/V_H and C_L/C_H Interface

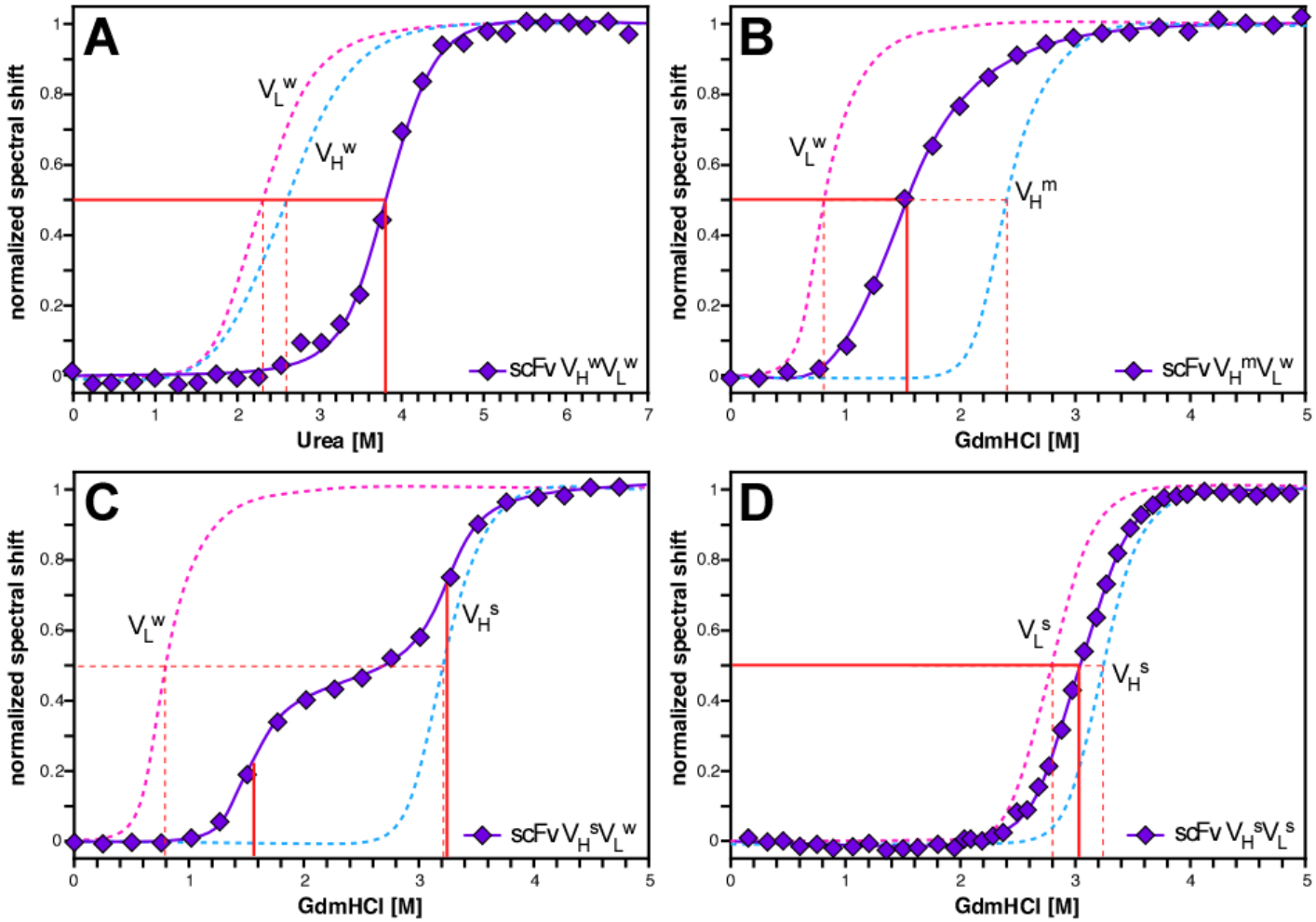


1570 +/- 160 Å^2
70% non-polar

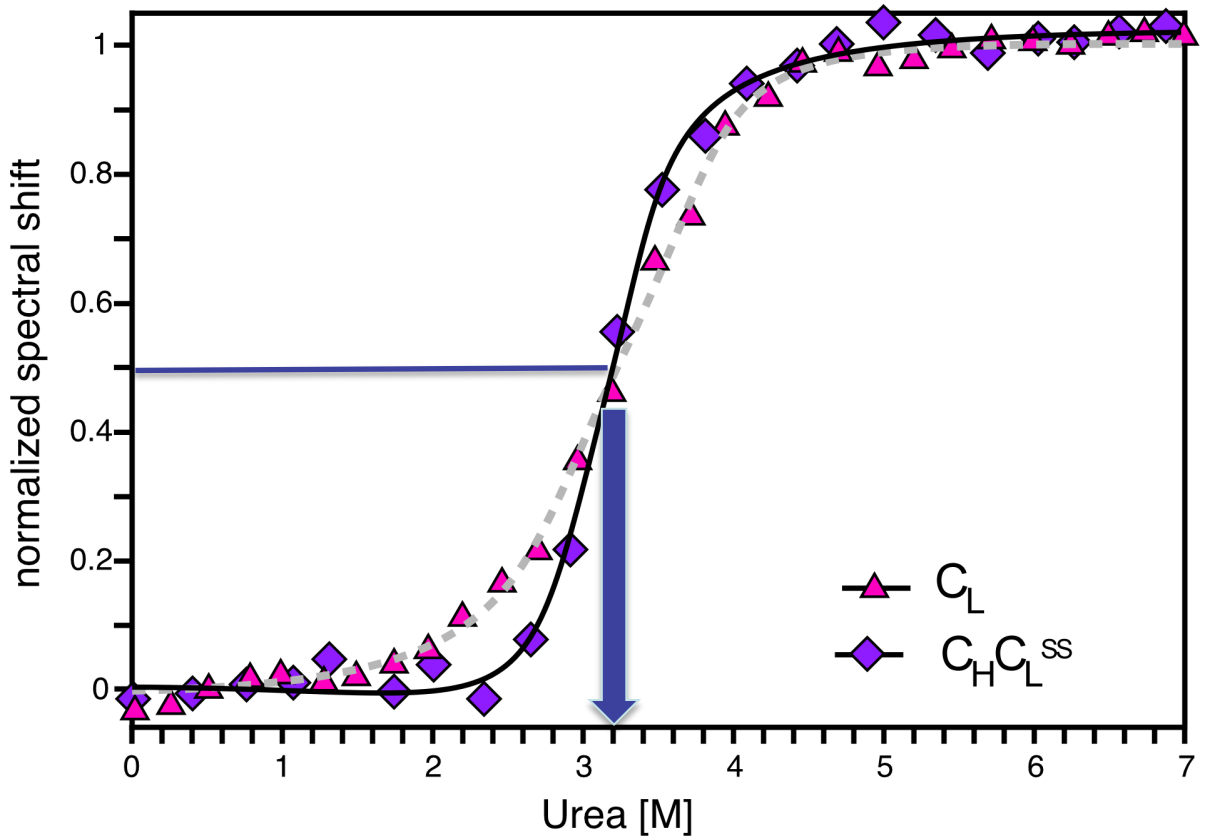
1970 +/- 160 Å^2
70% non-polar

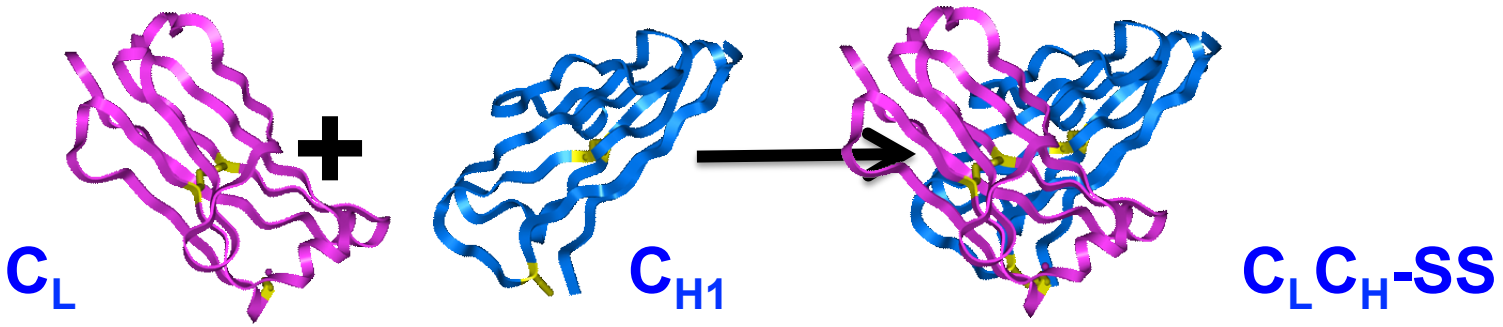
Mutual Stabilization V_L/V_H



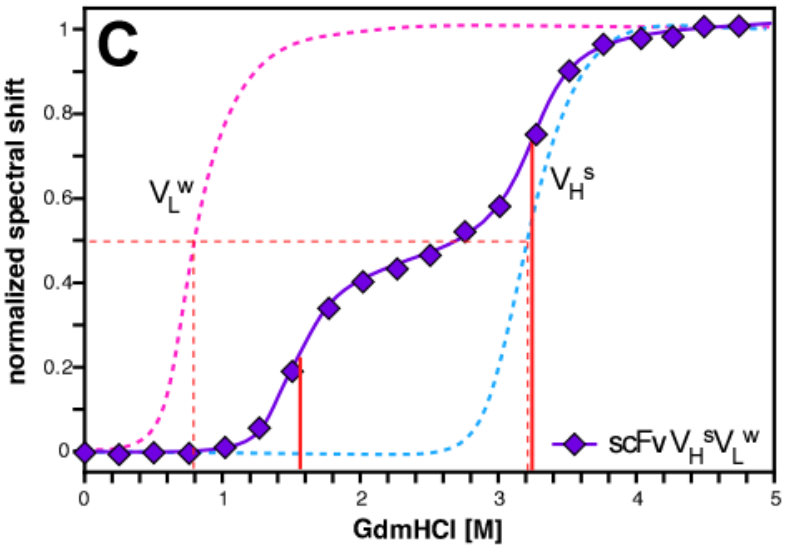


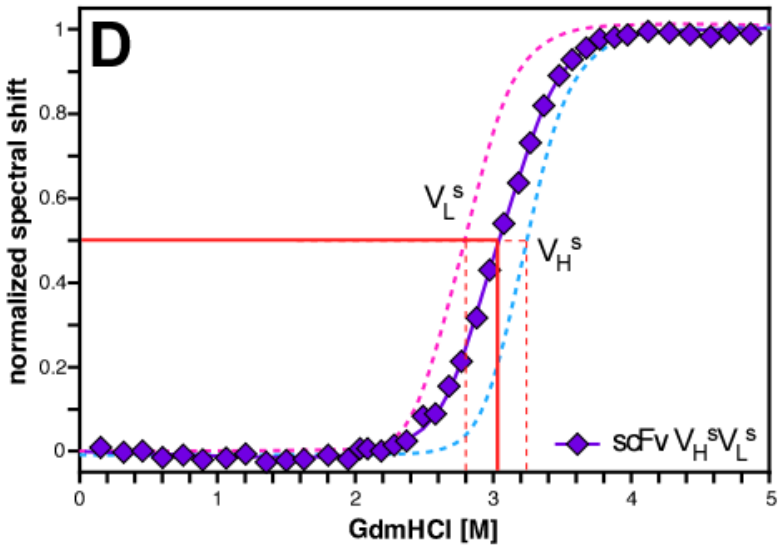
Little Mutual Stabilization of C_L and C_{H1}

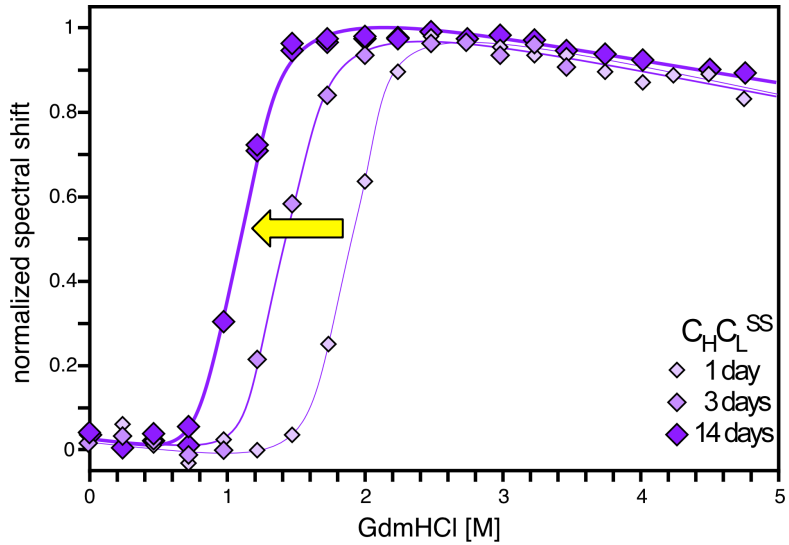




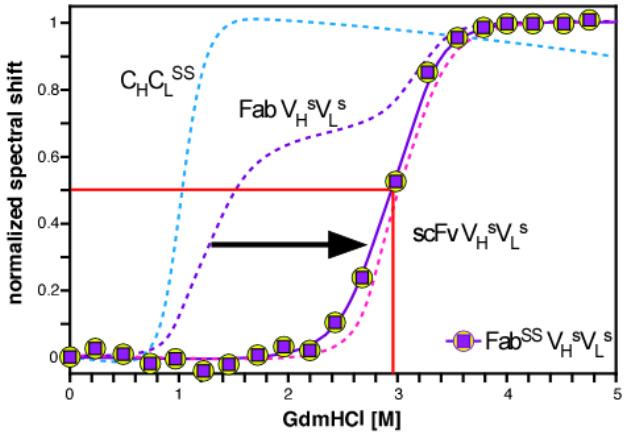
V_L/V_H compared to C_L/C_H^{SS}



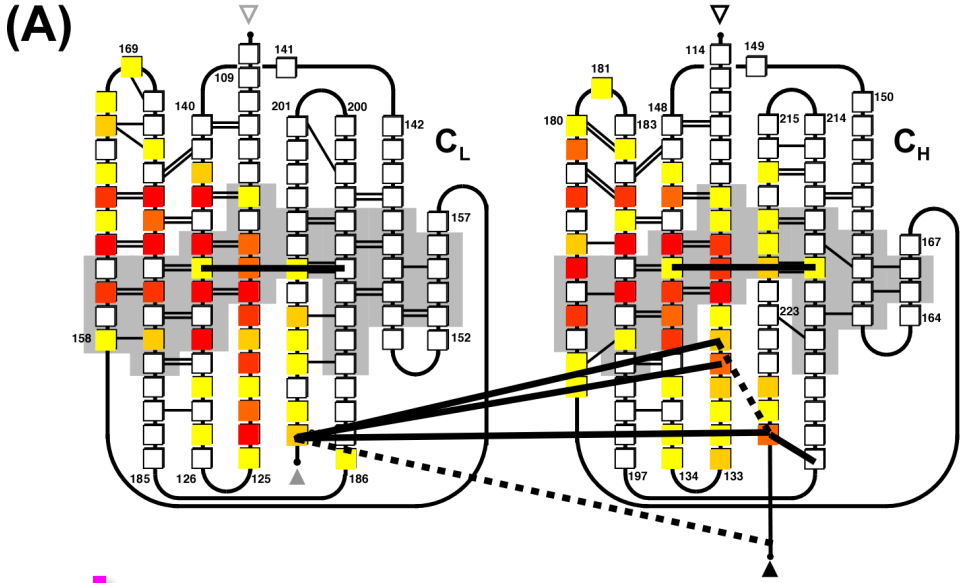




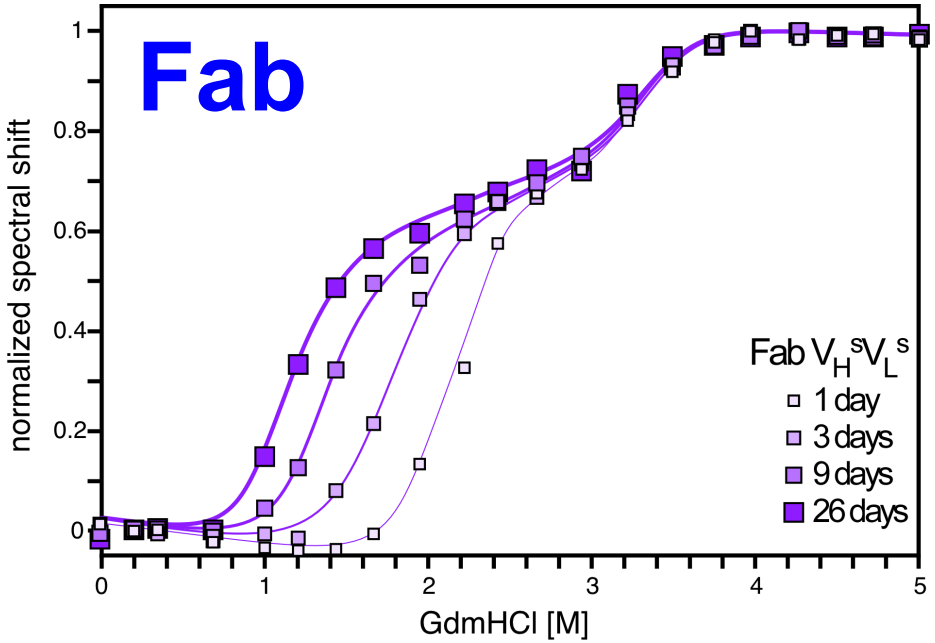
C_L/C_H^{SS} thermodynamically not very stable, but slow to unfold (kinetic stabilization)

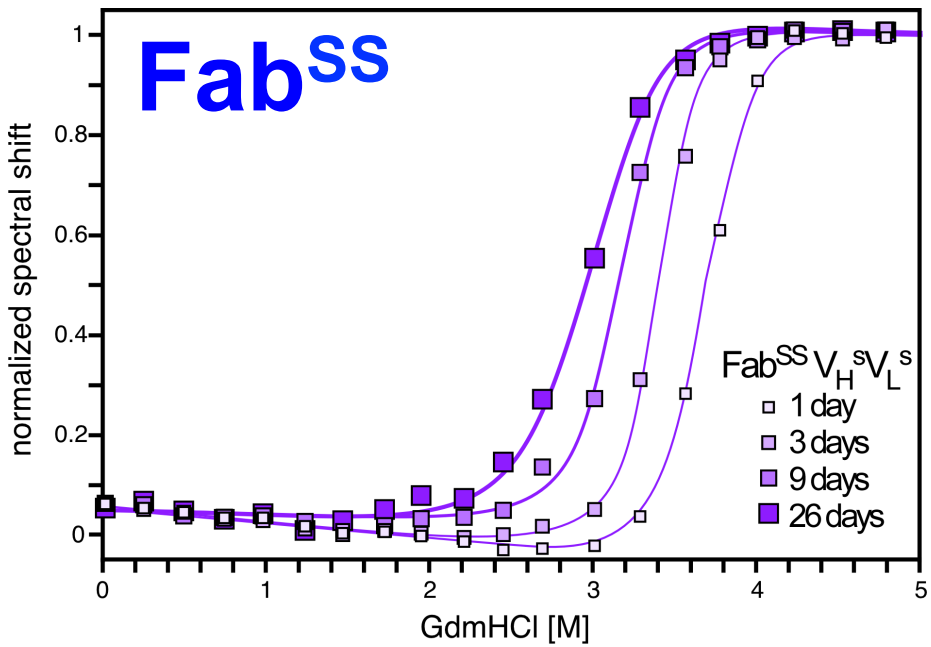


Röthlisberger et al. J.Biol.Chem. 347 (2005) 773-789



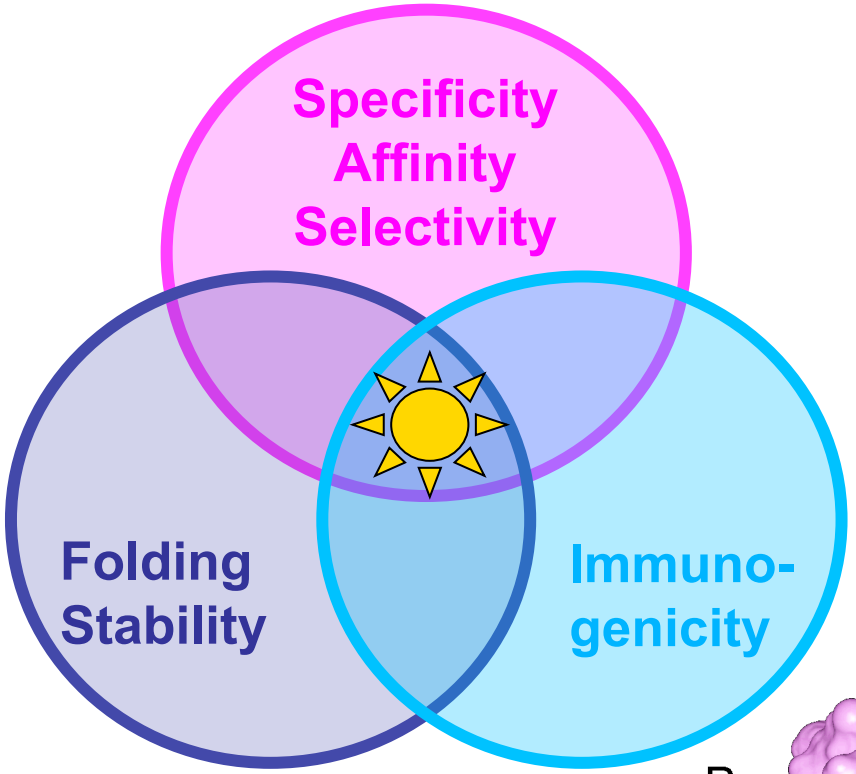
Effect of Interchain SS-Bond



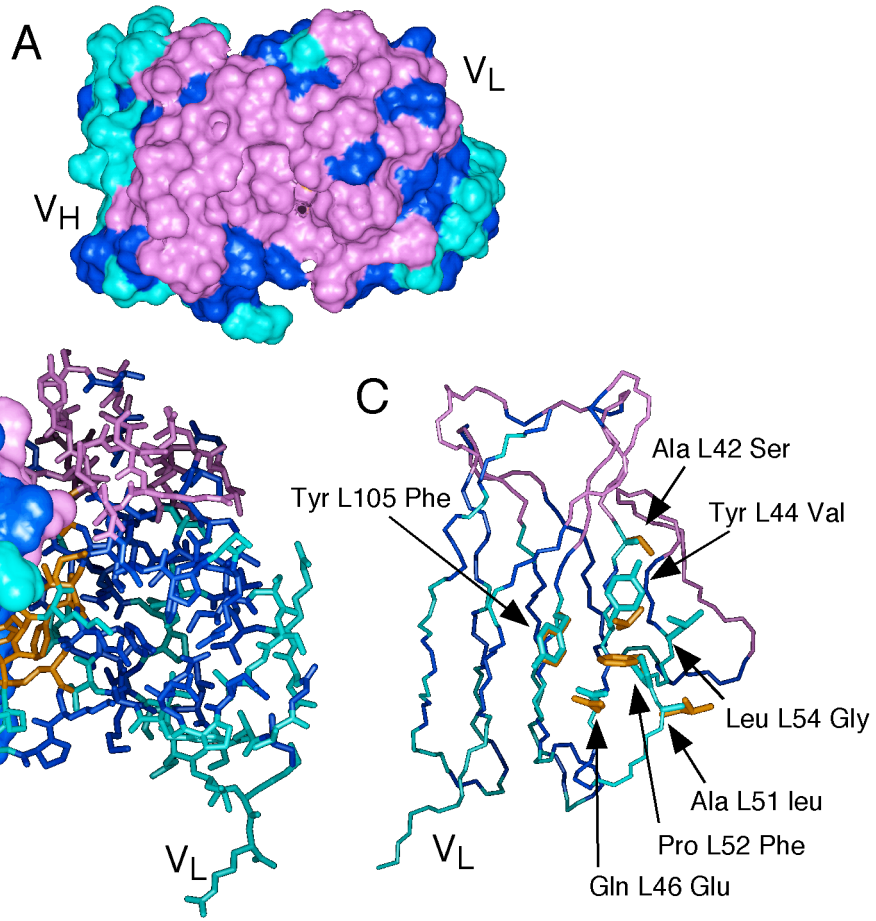


Stabilization and Humanization by CDR Graft

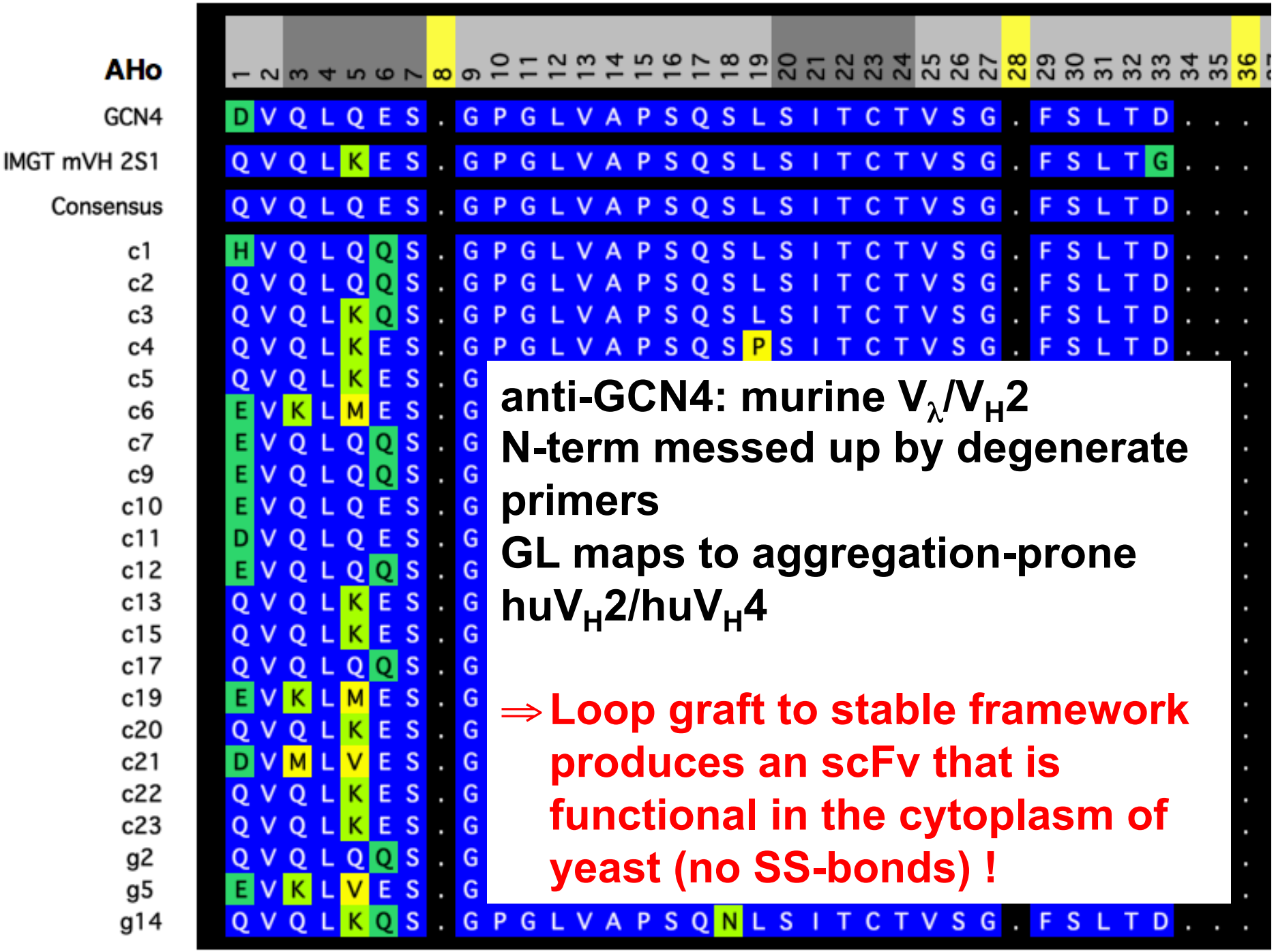
CDR Grafts



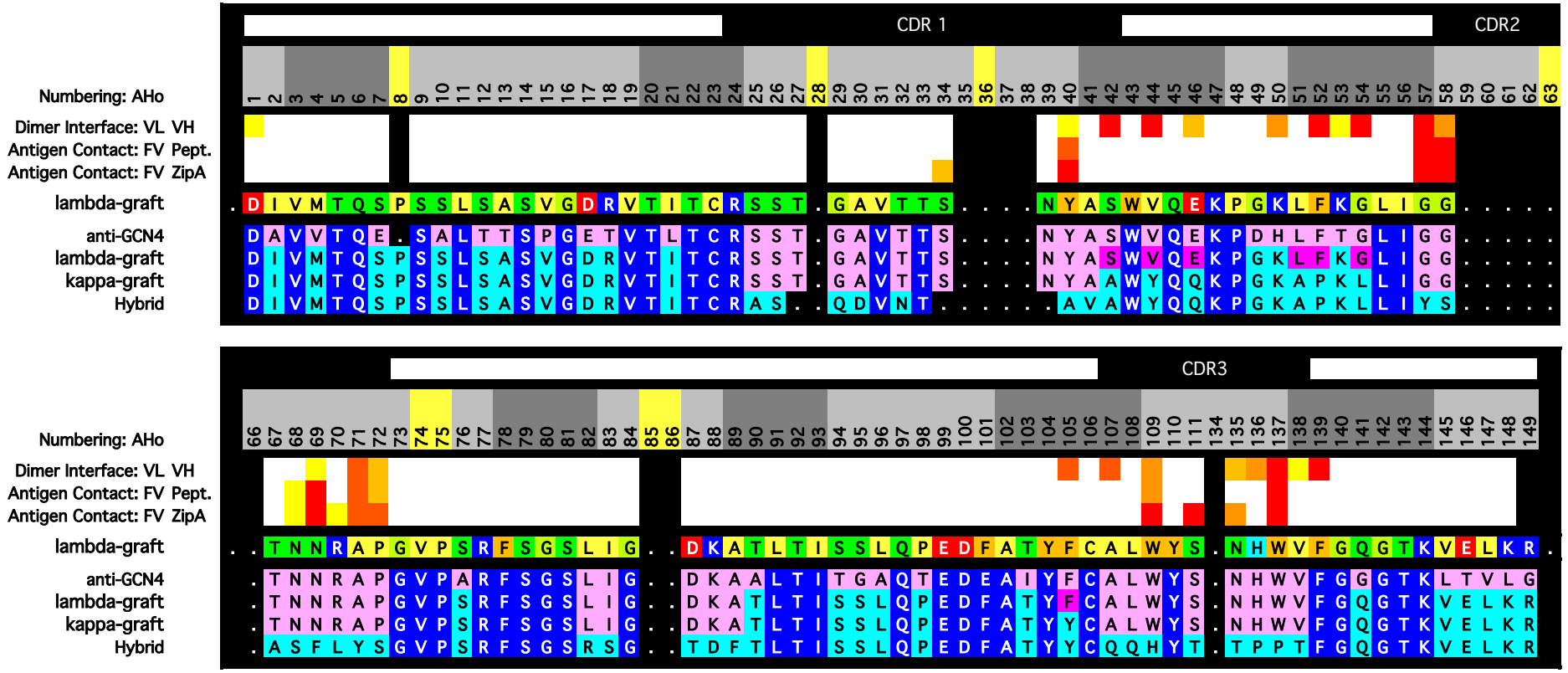
Don't forget the V_L/V_H interface residues!



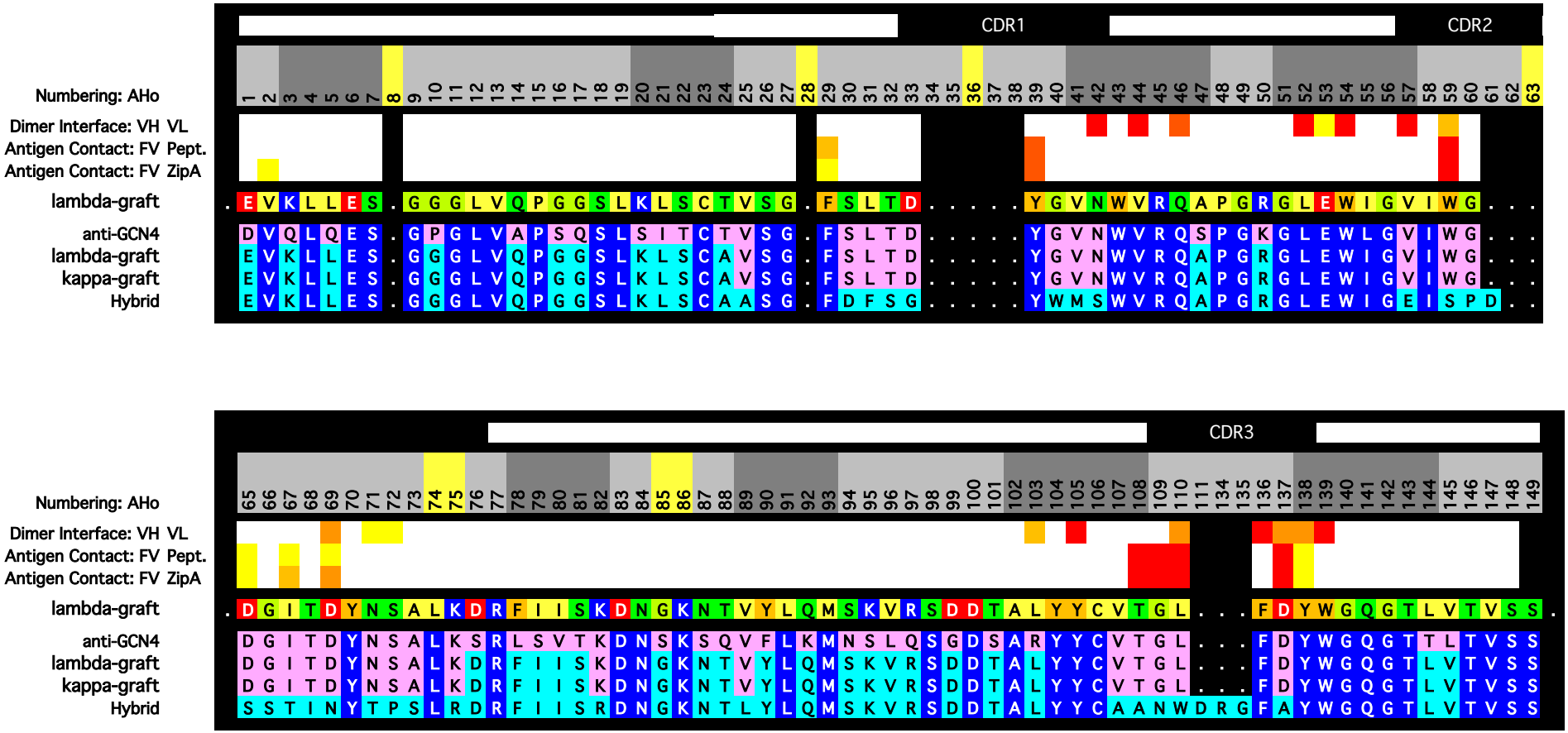
Wörn et al. J.Biol.Chem. 275 (2000) 2795-2803


AHo
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
GCN4 D V Q L Q E S . G P G L V A P S Q S L S I T C T V S G . F S L T D . . .
IMGT mVH 2S1 Q V Q L K E S . G P G L V A P S Q S L S I T C T V S G . F S L T G . . .
Consensus Q V Q L Q E S . G P G L V A P S Q S L S I T C T V S G . F S L T D . . .
c1 H V Q L Q Q S . G P G L V A P S Q S L S I T C T V S G . F S L T D . . .
c2 Q V Q L Q Q S . G P G L V A P S Q S L S I T C T V S G . F S L T D . . .
c3 Q V Q L K Q S . G P G L V A P S Q S L S I T C T V S G . F S L T D . . .
c4 Q V Q L K E S . G P G L V A P S Q S P S I T C T V S G . F S L T D . . .
c5 Q V Q L K E S . G
c6 E V K L M E S . G
c7 E V Q L Q Q S . G
c9 E V Q L Q Q S . G
c10 E V Q L Q E S . G
c11 D V Q L Q E S . G
c12 E V Q L Q Q S . G
c13 Q V Q L K E S . G
c15 Q V Q L K E S . G
c17 Q V Q L Q Q S . G
c19 E V K L M E S . G
c20 Q V Q L K E S . G
c21 D V M L V E S . G
c22 Q V Q L K E S . G
c23 Q V Q L K E S . G
g2 Q V Q L Q Q S . G
g5 E V K L V E S . G
g14 Q V Q L K Q S . G P G L V A P S Q N L S I T C T V S G . F S L T D . . .
anti-GCN4: murine V_λ/V_H2
N-term messed up by degenerate primers
GL maps to aggregation-prone huV_H2/huV_H4
⇒ Loop graft to stable framework produces an scFv that is functional in the cytoplasm of yeast (no SS-bonds) !

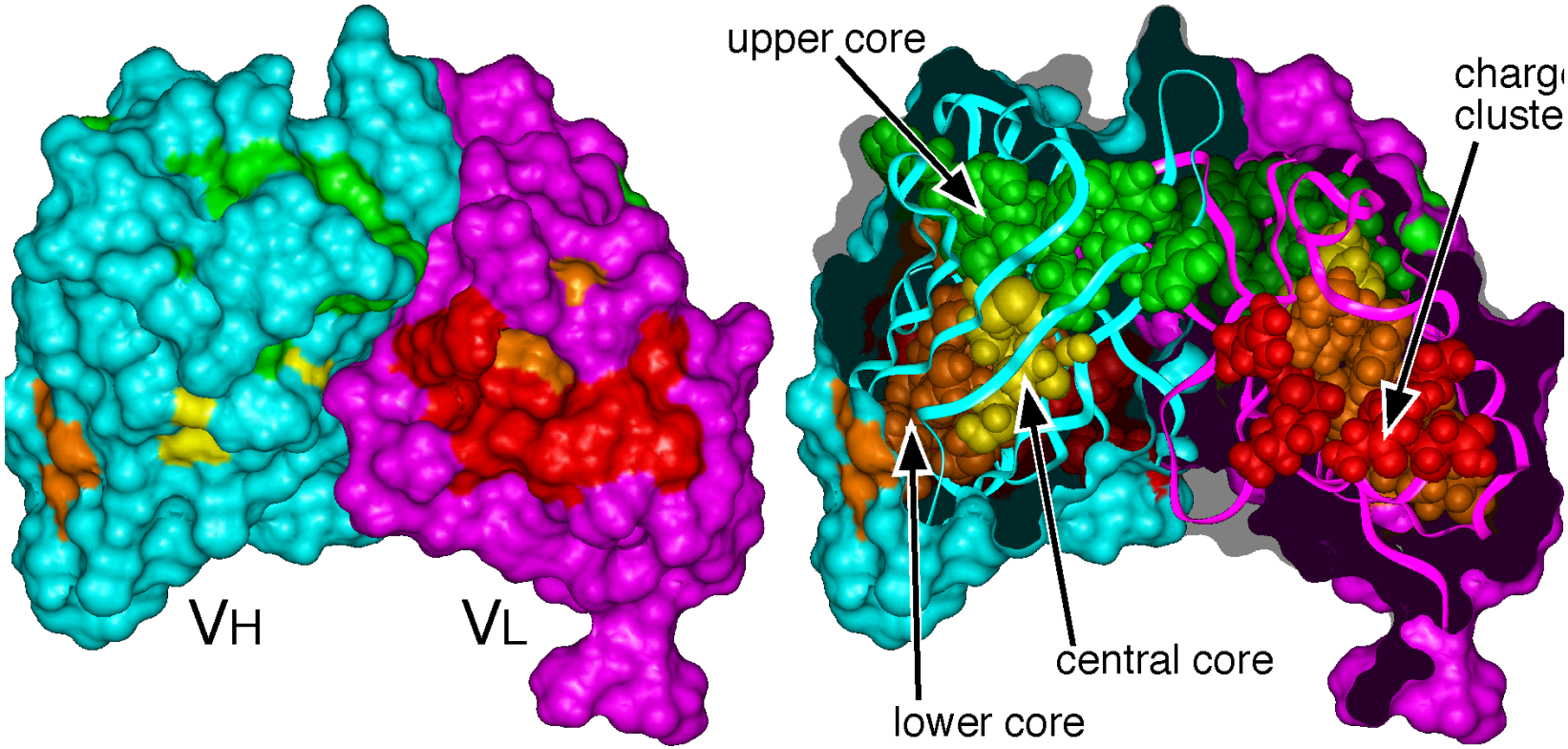
GCN4-Graft VL



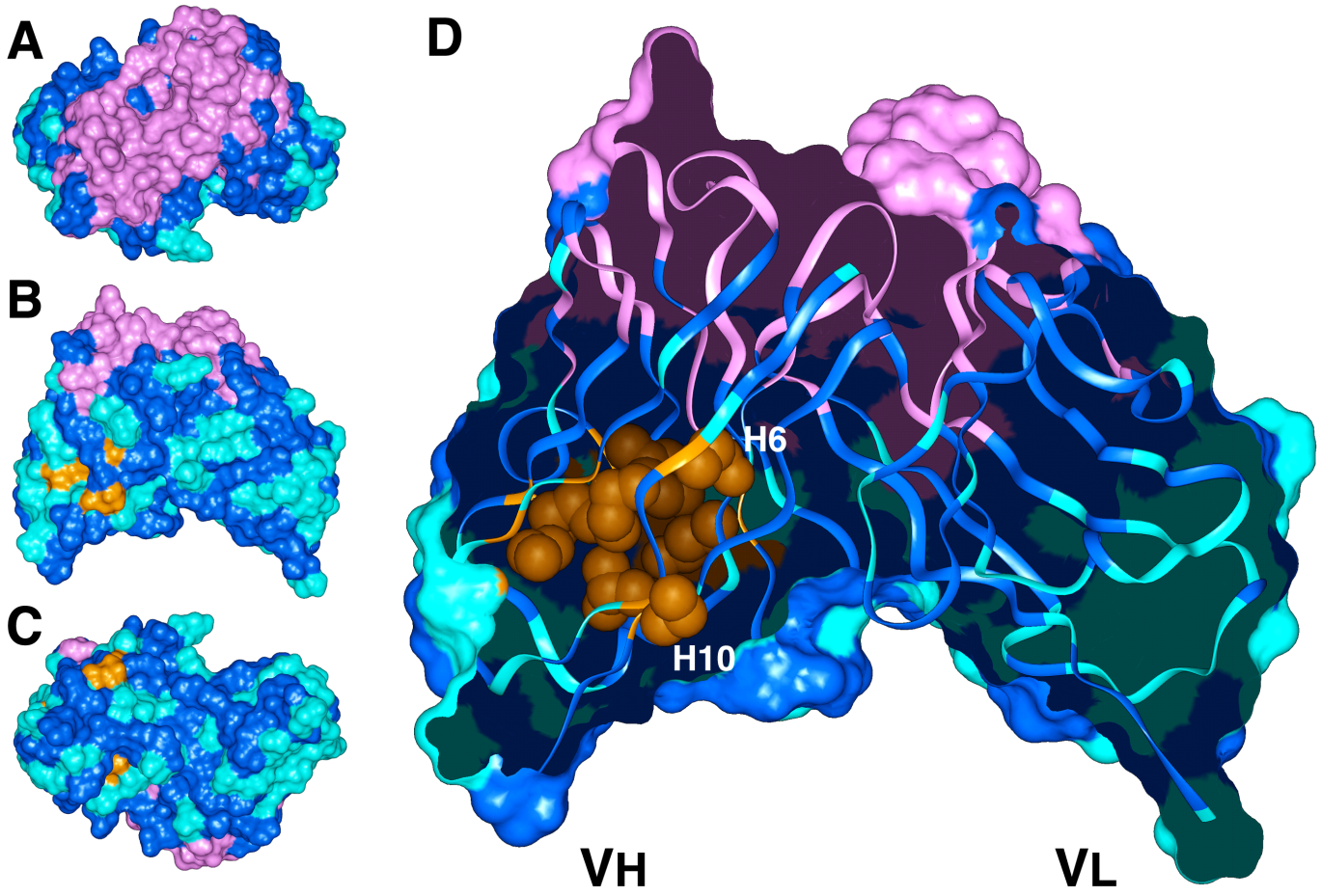
GCN4-Graft V_H



Upper core packing



Framework selection: Why huVH3 is not Always the Best Acceptor Framework

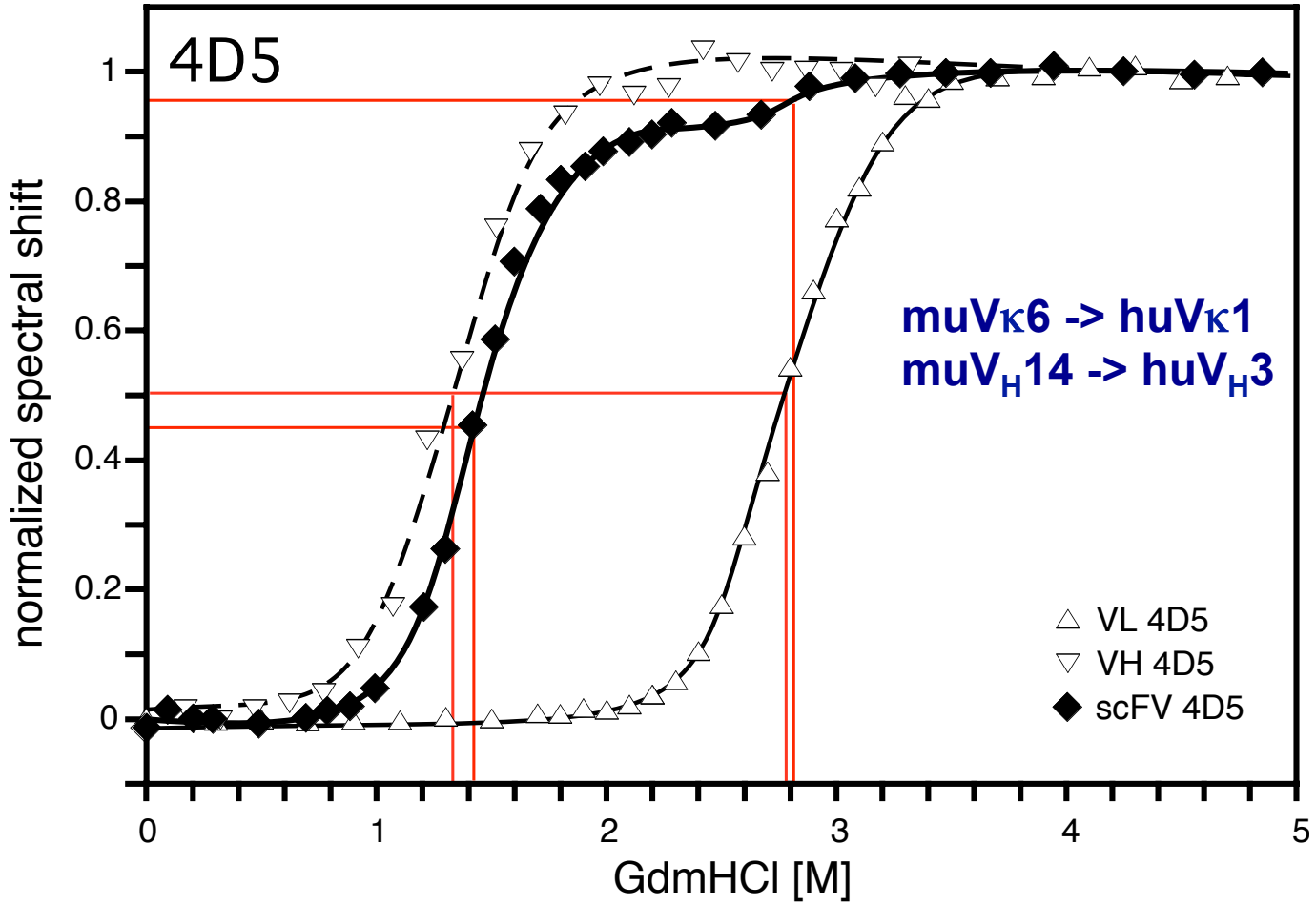


Kügler et. al. Prot. Eng. Des. Sel. 22 (2009)135-137
Honegger et. al. Prot. Eng. Des. Sel. 22 (2009)121-134
Willuda et al. Cancer Res. 59 (1999) 5758-5767

scFv 4d5mocB

- Monoclonal antibody Moc31, raised against small cell lung carcinoma (*Lung Cancer 4 (1988) 1-114*), recognizing Ep-CAM / EGP-2 (Epithelial Glycoprotein 2)
- Murine Moc31 scFv, K_D $3.9*10^{-9}$ M, poor expression, very poor thermal stability, pseudomonas exotoxin A fusion protein selectively cytotoxic to EGP-2 positive SCLC and adenocarcinoma cell lines. (*Cancer Immunol. Immunother. 44 (1997) 1-9*). No tumor enrichment in mouse xenografts
- scFv 4D5mocA: classical loop graft, no loss of antigen affinity, improved expression and thermal stability, poor tumor enrichment in mouse xenografts.
- scFv 4D5mocB: V_L=classical loop graft, V_H: Residues H6, H7, H10 and hydrophobic core of V_H retained from murine Moc31, resulted into further improved thermal stability, good tumor enrichment in mouse xenografts. (*Cancer Research 59(1999) 5758-5767*).

The Gold Standard: hu4D5-V8 scFv



HuCAL Vκ1: $[GdmCL]_{50}$ = 2.3 M
HuCAL V_H3: $[GdmCL]_{50}$ = 2.7 M[a] (3.0 M[b])
[a] with CDR-H3 of 4D5, [b] with long, stabilizing CDR-H3
huVκ1/huV$_H$3 scFv: $[GdmCL]_{50}$ = 2.8 M

4D5 Vκ1: $[GdmCL]_{50}$ = 2.8 M
4D5 V_H3: $[GdmCL]_{50}$ = 1.3 M

hu4D5-V8: $[GdmCL]_{50}$ = 1.4 M

DEMO

CDR Graft Designer

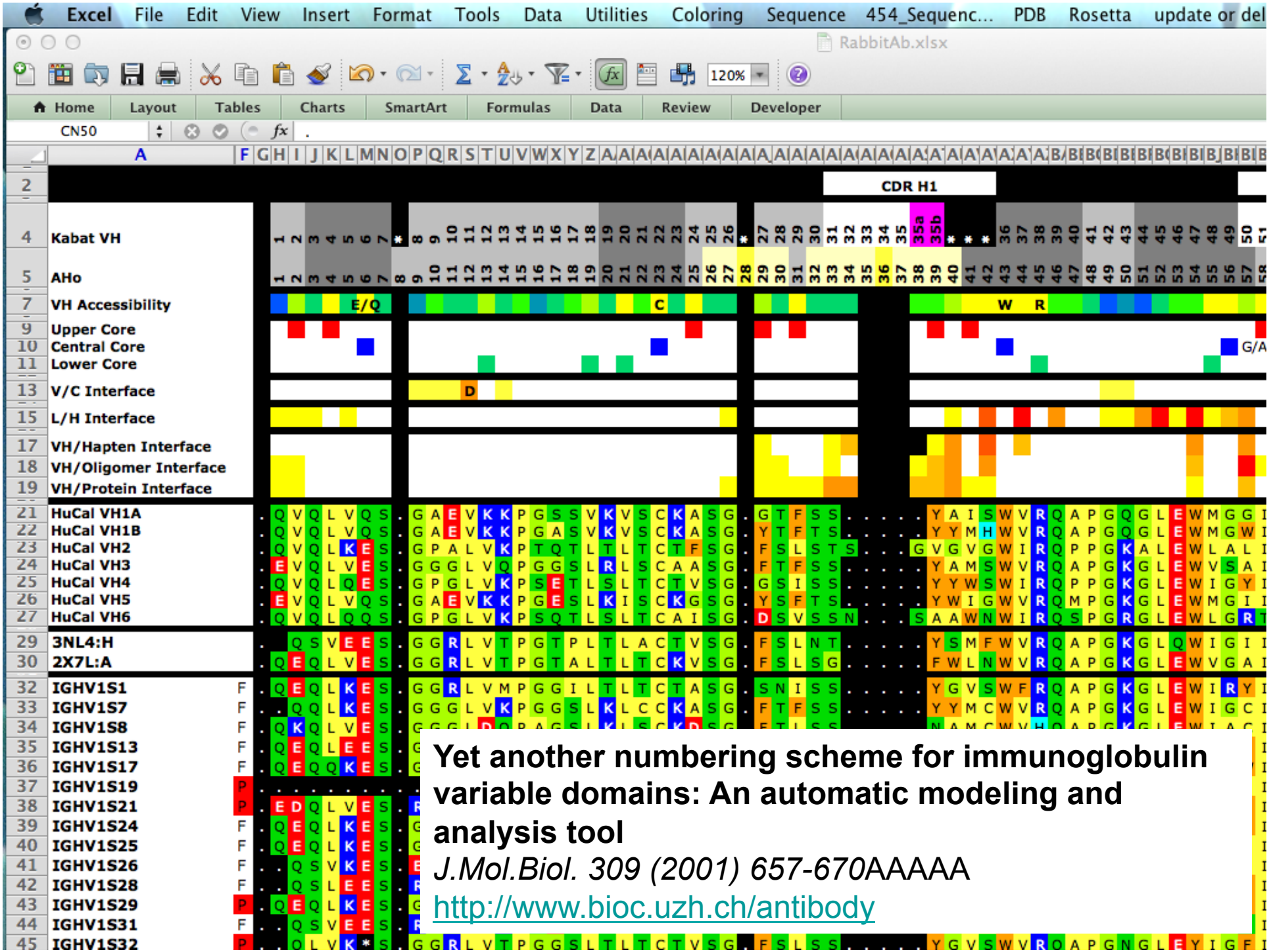
FAQ:

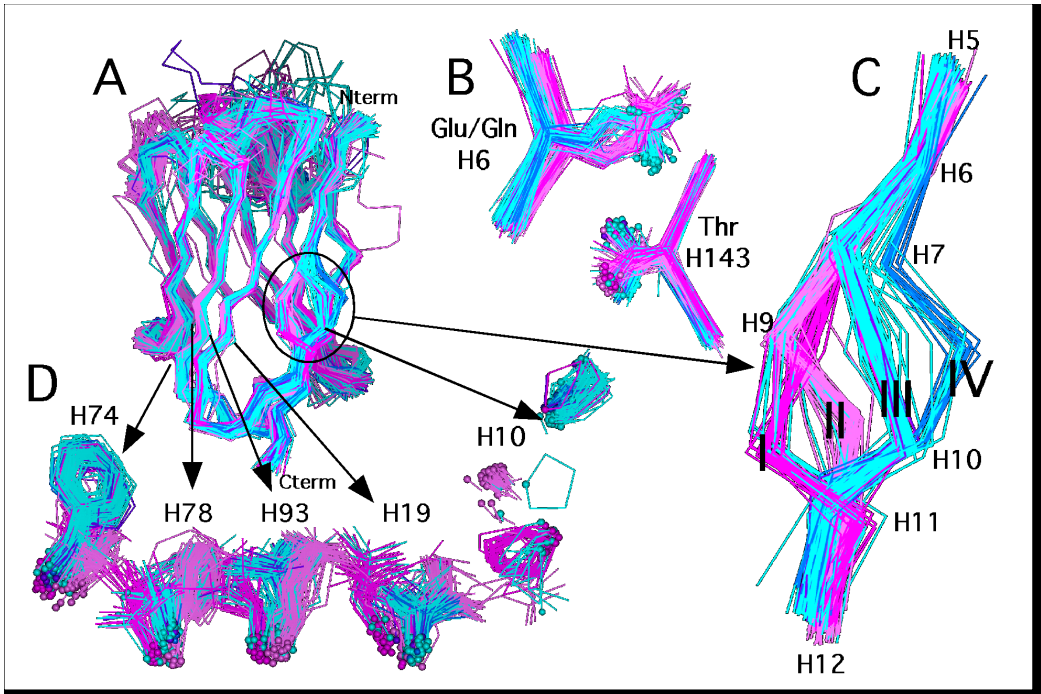
Are the rules derived from work with murine and human antibodies transferrable to antibodies from other species?

Extreme Example:
CDR Grafts from Rabbit Antibodies

- Germline sequence alignment retrieved from IMGT
- ~1000 rabbit V_H, ~500 Vκ and ~30 Vλ retrieved from NCBI
- Less sequence variability than human and murine antibodies,
- frameworks huVκ1 and huV_H3-like
- Several sequence features found that are not seen in human and murine antibody variable domains
- Only two rabbit antibody structures found in the pdb: **3NL4** (1.54 Å res.) is annotated as such, **2X7L** isn't.

  (3NL4 superseeded by 4HBC, Newly added 4HTI, 4JO1, 4JO2,4JO3, 4JO4)


Excel File Edit View Insert Format Tools Data Utilities Coloring Sequence 454_Sequenc... PDB Rosetta update or del
RabbitAb.xlsx
120%
Home Layout Tables Charts SmartArt Formulas Data Review Developer
CN50
CDR H1
Kabat VH
AHo
VH Accessibility
Upper Core
Central Core
Lower Core
V/C Interface
L/H Interface
VH/Hapten Interface
VH/Oligomer Interface
VH/Protein Interface
HuCal VH1A
HuCal VH1B
HuCal VH2
HuCal VH3
HuCal VH4
HuCal VH5
HuCal VH6
3NL4:H
2X7L:A
IGHV1S1
IGHV1S7
IGHV1S8
IGHV1S13
IGHV1S17
IGHV1S19
IGHV1S21
IGHV1S24
IGHV1S25
IGHV1S26
IGHV1S28
IGHV1S29
IGHV1S31
IGHV1S32
Yet another numbering scheme for immunoglobulin variable domains: An automatic modeling and analysis tool
J.Mol.Biol. 309 (2001) 657-670AAAAA
http://www.bioc.uzh.ch/antibody



H6

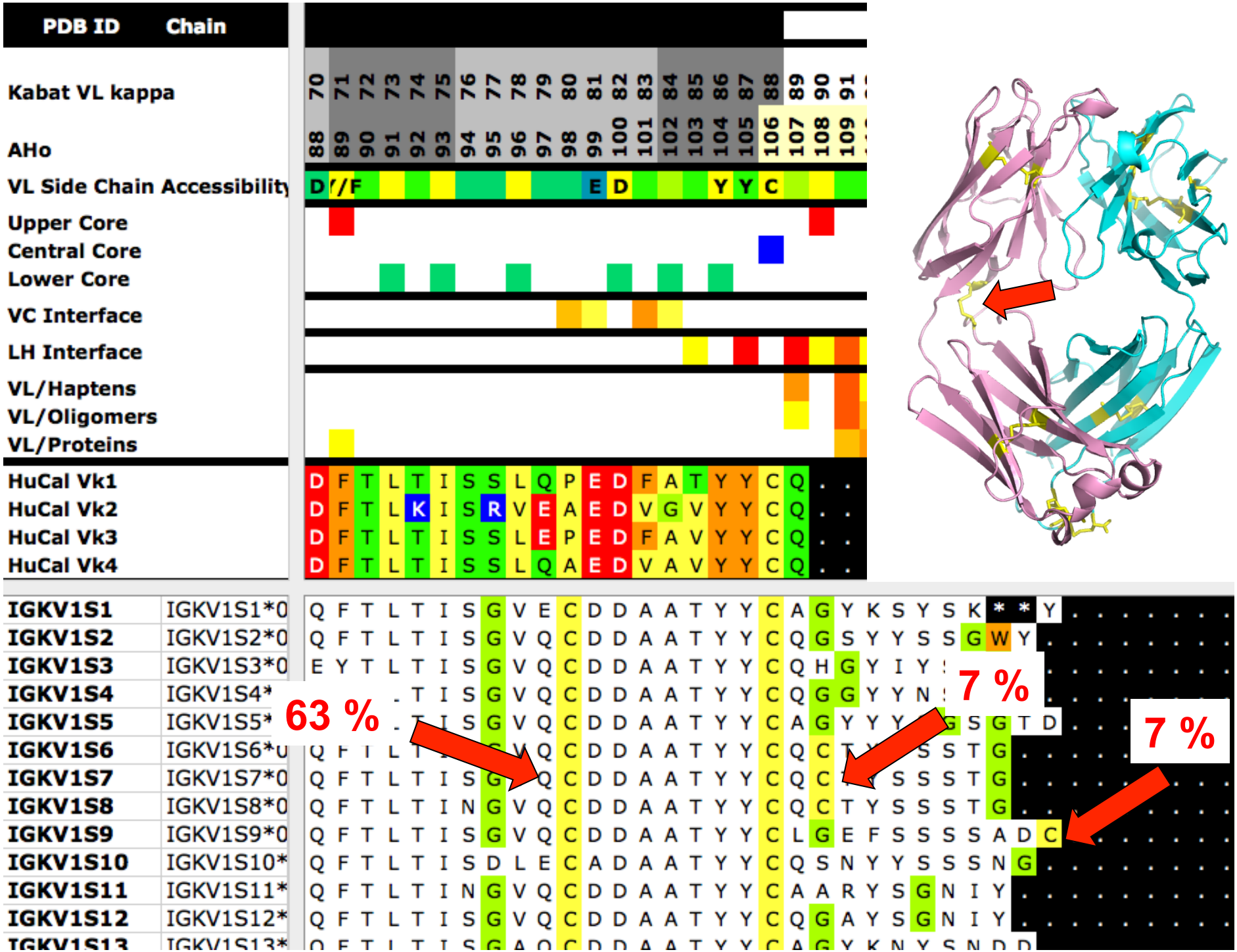
>gi	2920429	IGHV1S7	80.8
>gi	9799096	IGHV1S7	80.2
>gi	2920459	IGHV1S7	80.8
>gi	165196	IGHV1S7	84.0
>gi	2920467	IGHV1S7	86.6
>gi	9799316	IGHV1S7	81.1
>gi	9799304	IGHV1S7	81.1
>gi	9799249	IGHV1S7	82.1
>gi	2920496	IGHV1S7	74.2
>gi	2407786	IGHV1S7	75.0
>gi	2920469	IGHV1S7	86.6
>gi	2920490	IGHV1S7	85.6
>gi	3930276	IGHV1S7	80.4
>gi	27373563	IGHV1S7	74.2
>gi	2920498	IGHV1S7	70.1
>gi	2920471	IGHV1S7	83.5
>gi	165250	IGHV1S7	73.0
>gi	9799143	IGHV1S7	84.7

M E T G L R W L L L V A V L K G V Q C

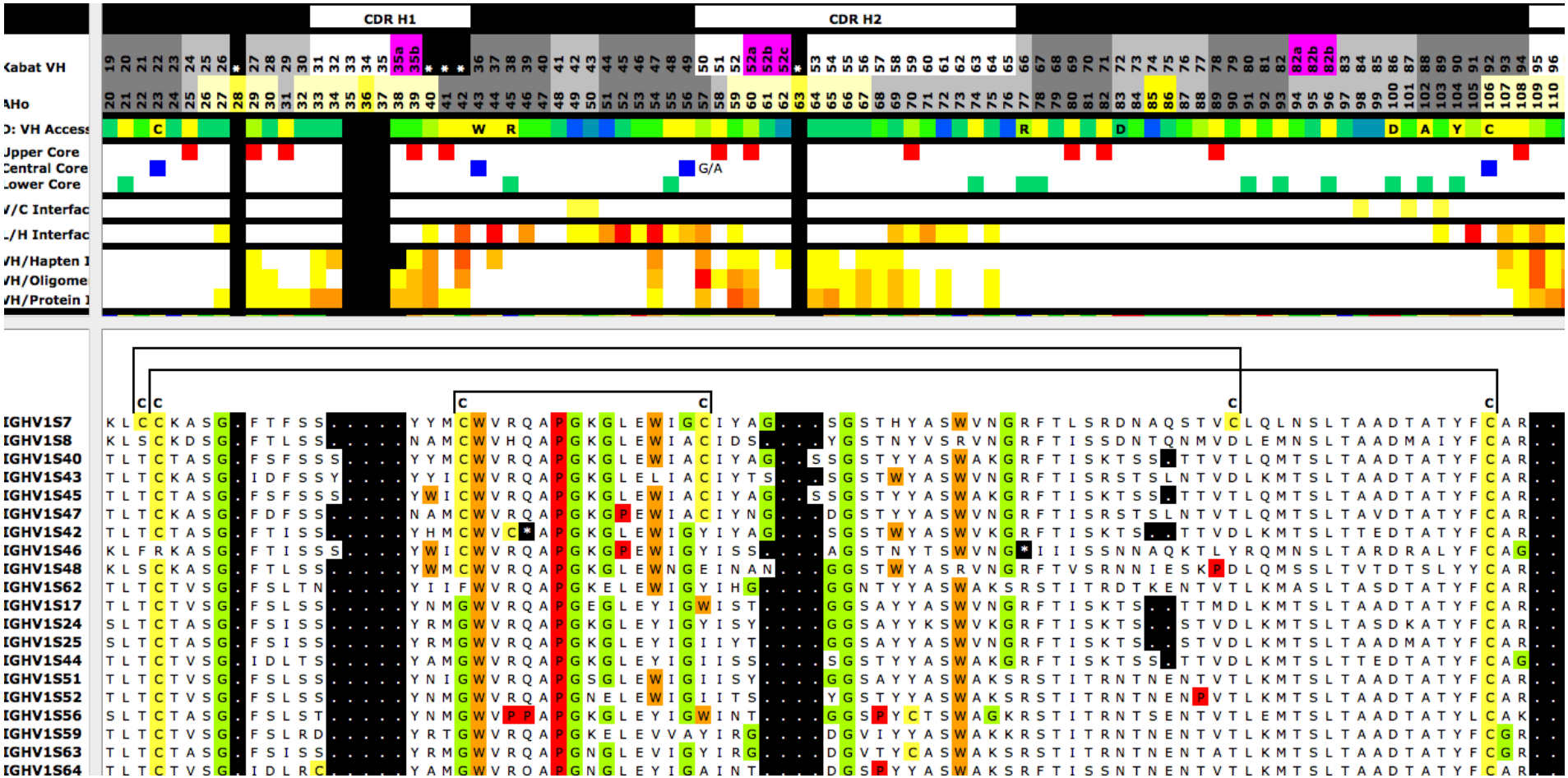
V Q C

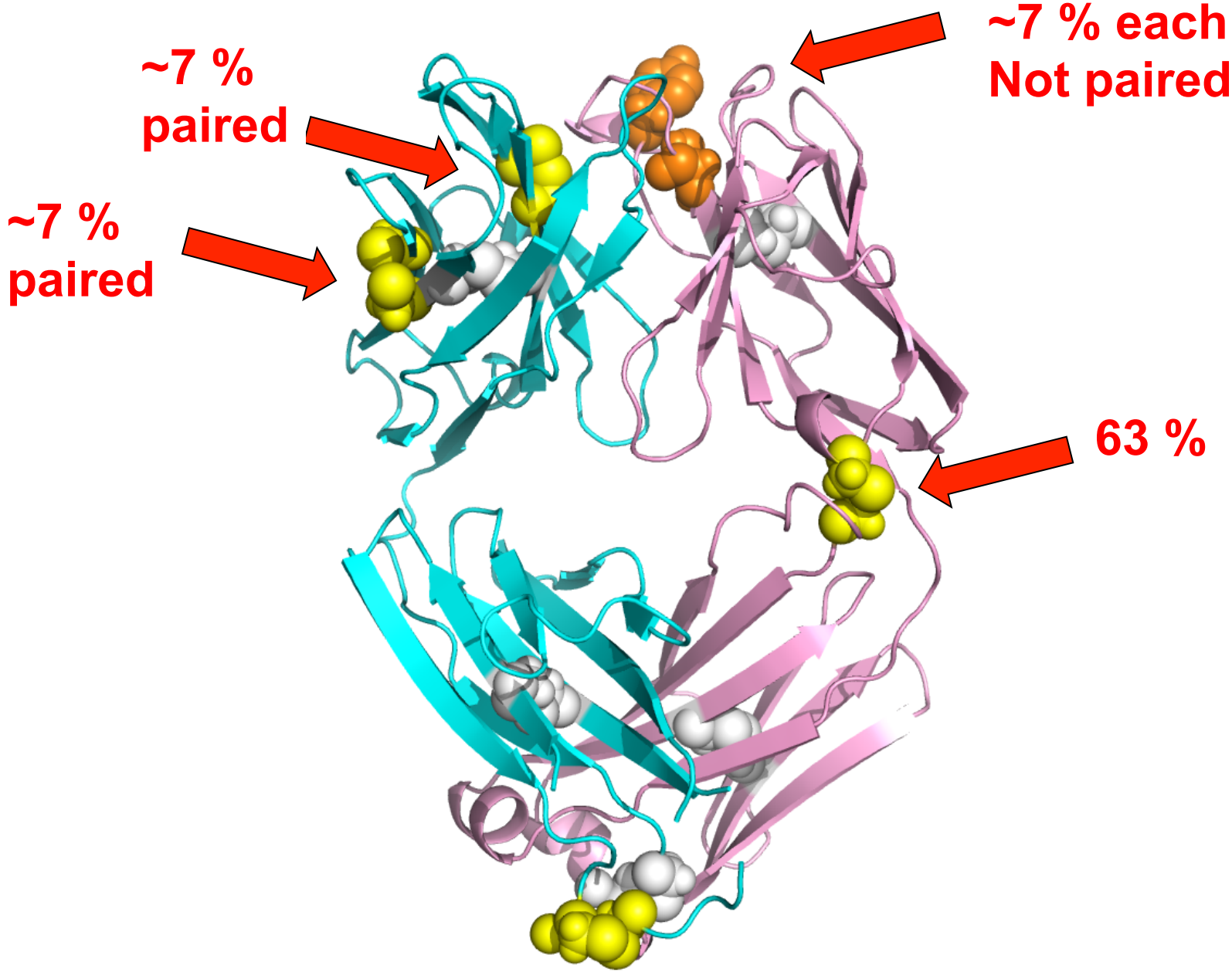
Q C

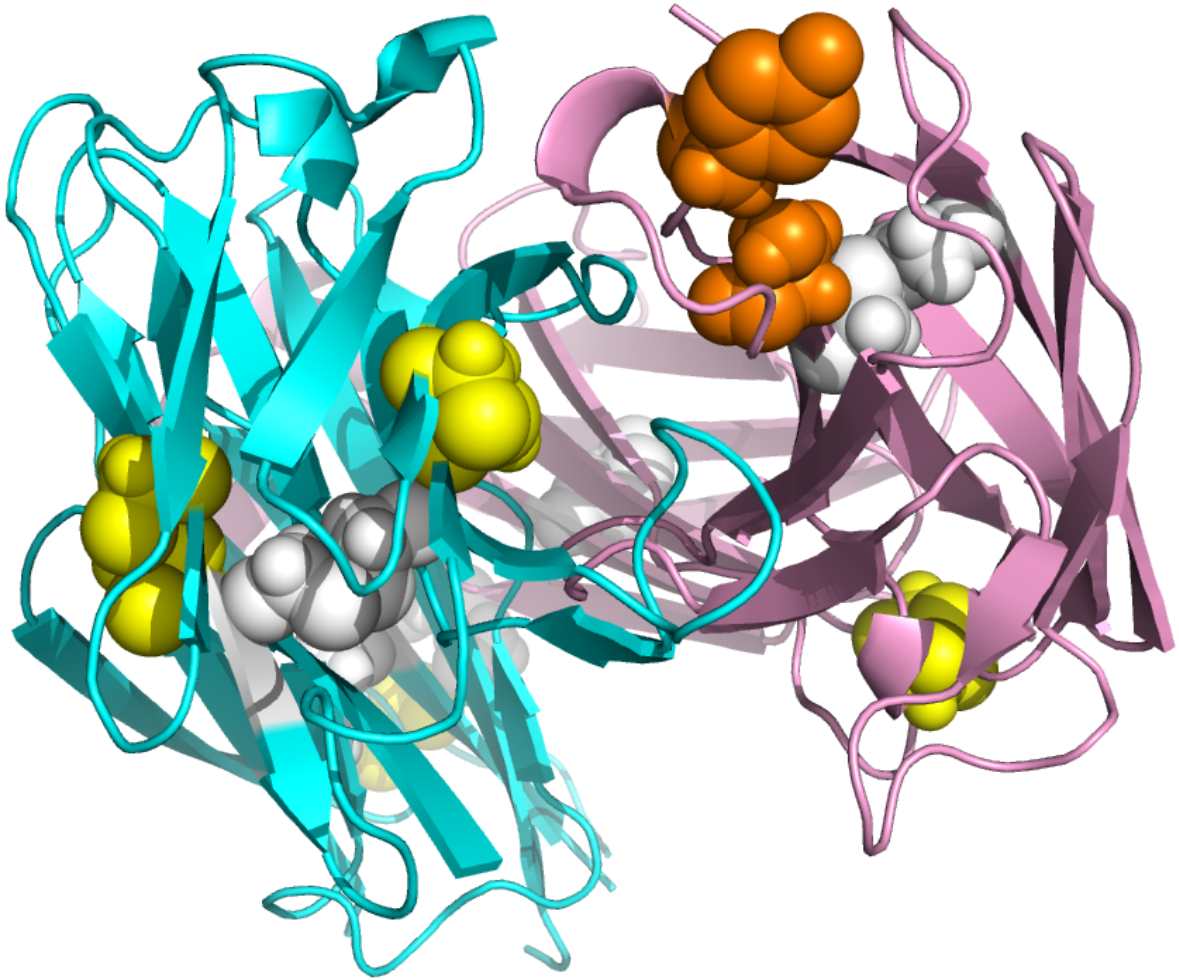
Q C

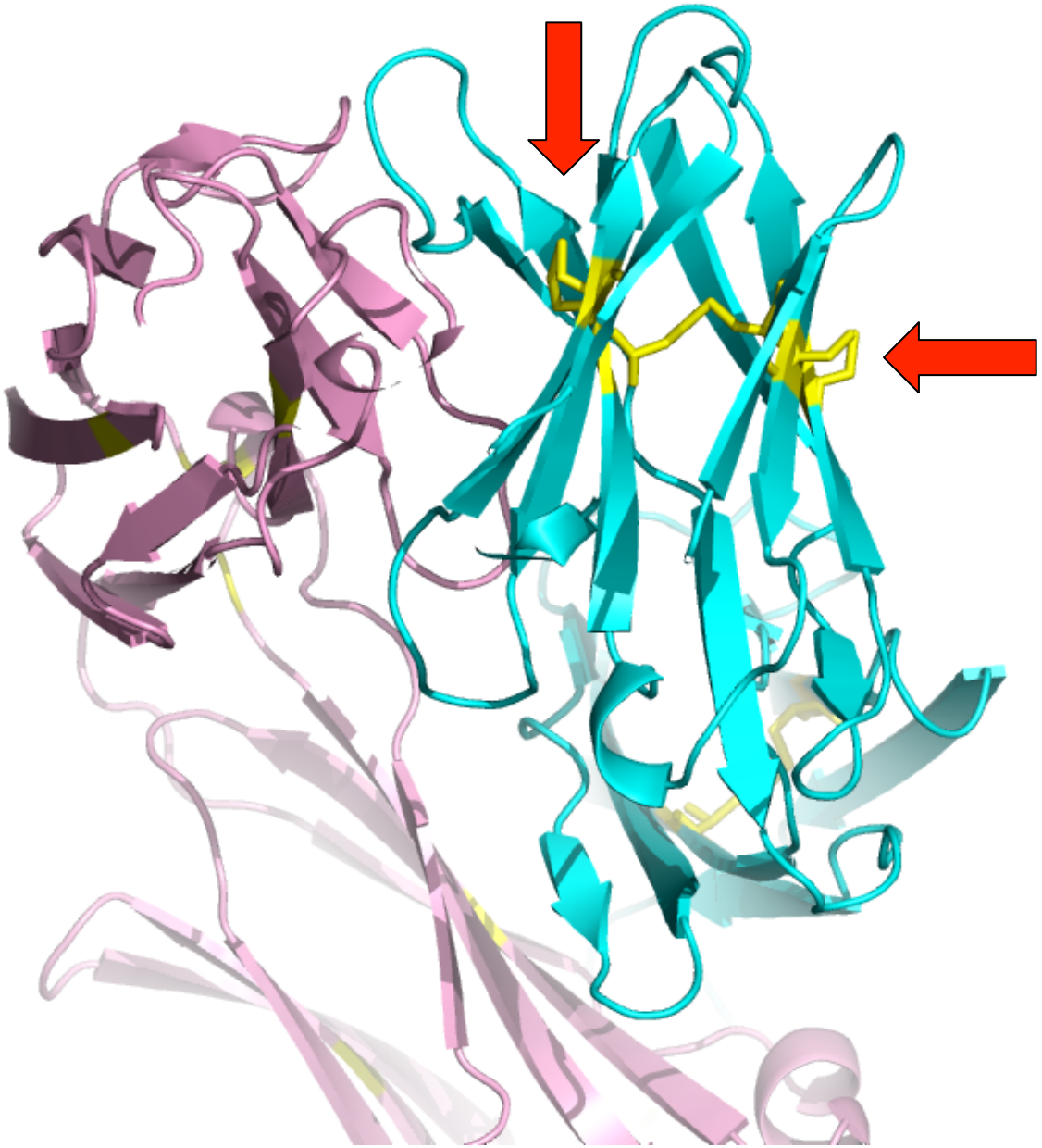

PDB ID
Chain
Kabat VL kappa
AHo
VL Side Chain Accessibility
Upper Core
Central Core
Lower Core
VC Interface
LH Interface
VL/Haptens
VL/Oligomers
VL/Proteins
HuCal Vk1
HuCal Vk2
HuCal Vk3
HuCal Vk4
IGKV1S1
IGKV1S2
IGKV1S3
IGKV1S4
IGKV1S5
IGKV1S6
IGKV1S7
IGKV1S8
IGKV1S9
IGKV1S10
IGKV1S11
IGKV1S12
63 %
7 %
7 %

Additional S-S bridges in V_H

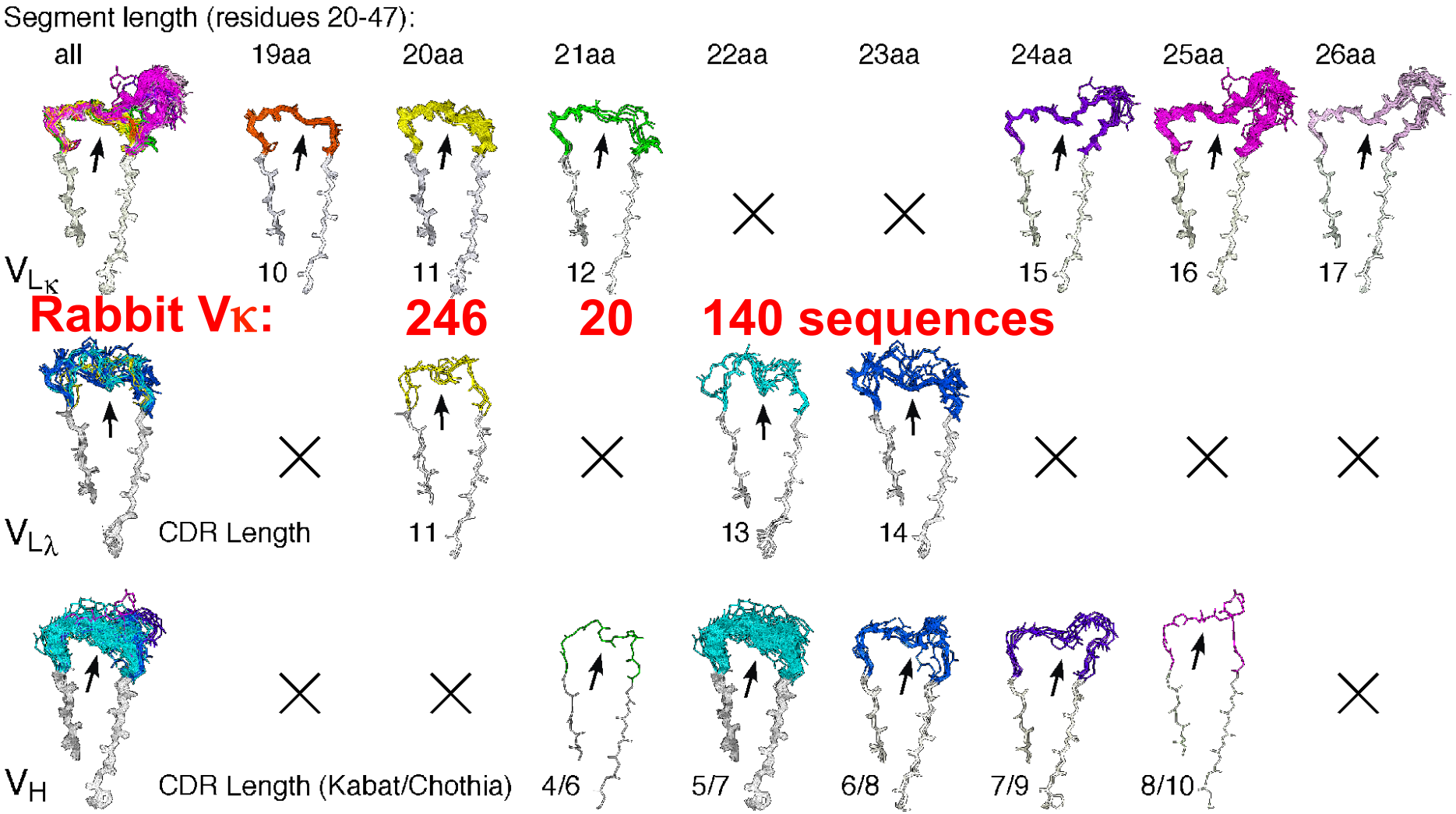



~7 % each
Not paired
~7 %
paired
~7 %
paired
63 %

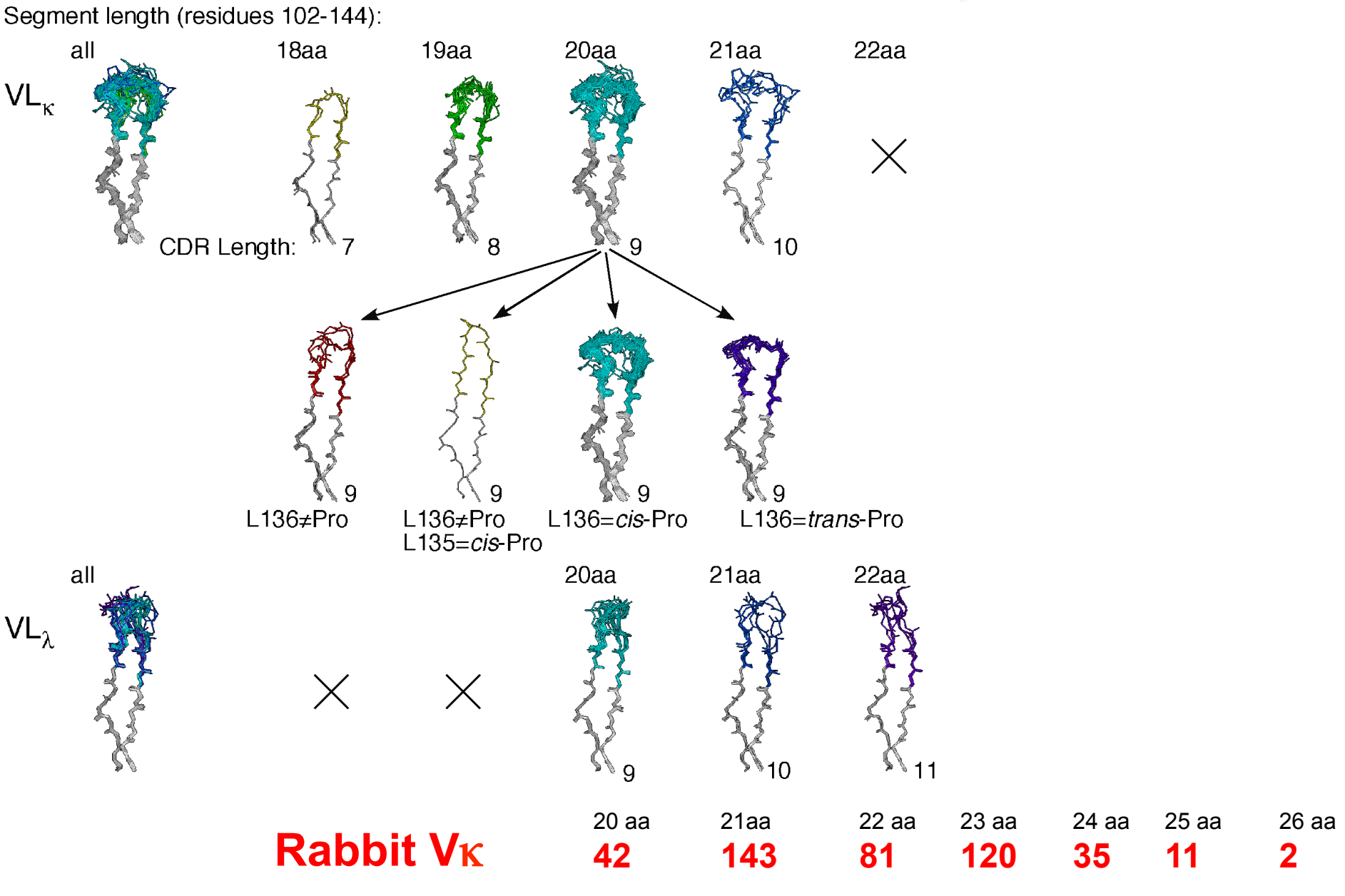




CDR 1 in murine/human antibodies



CDR L3 in murine/human light chains



	20 aa	21aa	22 aa	23 aa	24 aa	25 aa	26 aa
Rabbit Vκ	**42**	**143**	**81**	**120**	**35**	**11**	**2**

CDR-Donors: Rabbit Antibodies

Murine and human antibody repertoires are quite similar – the rabbit repertoire is different:

Vκ-Domains

- Rabbit kappa light chains contain an additional Cys in position L98, which can form a disulfide bond with a Cys in C_L
- (Chothia canonical rules do not recognize most rabbit CDR L1s, although there is no reason why they should not assume the conformation appropriate to their length) not surprising!
- There is less length variability in CDR-L1 of rabbit Vκ domains than in human and murine kappa domains
- CDR L3 in rabbit Vκ lack Gln L108 (L90) and *cis*-Pro L136 (L?), which in human and murine Vκ domains produce the typical Ω-loop conformation. This produces a lambda-like CDR-L3 which might increase the flexibility of the V_L/V_H interface.
- Germline-encoded Cys in CDR L3
- CDR-L3 frequently longer than longest in mouse or man

CDR-Donors: Rabbit Antibodies

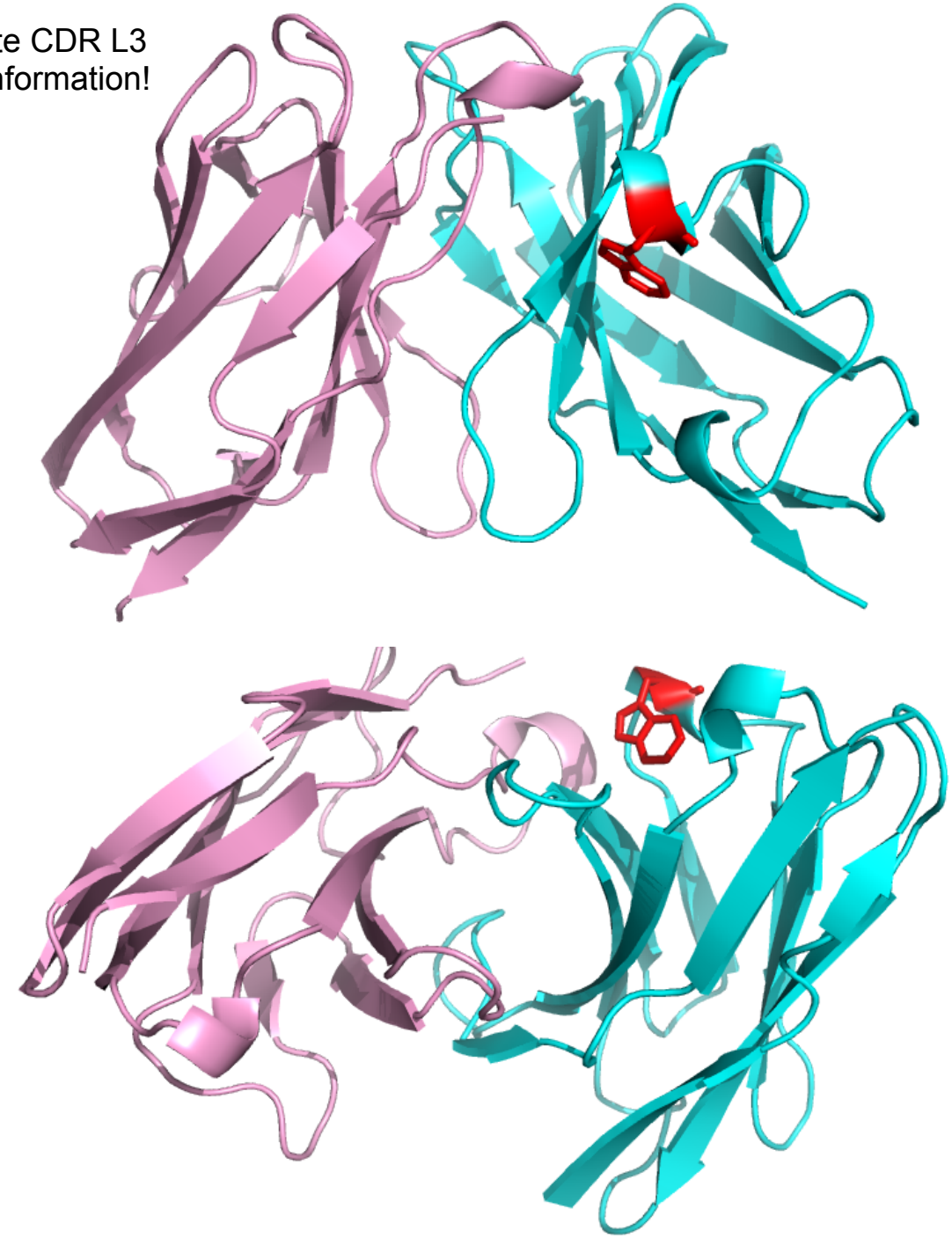
VH-Domains

- A'-strand (N-terminus) is frequently shortened by one residue
- Upper core residue H2 is hydrophilic
- Some V_H domains have a flexible insertion (4 or 5 residues) in the kink between strands A' and A''.

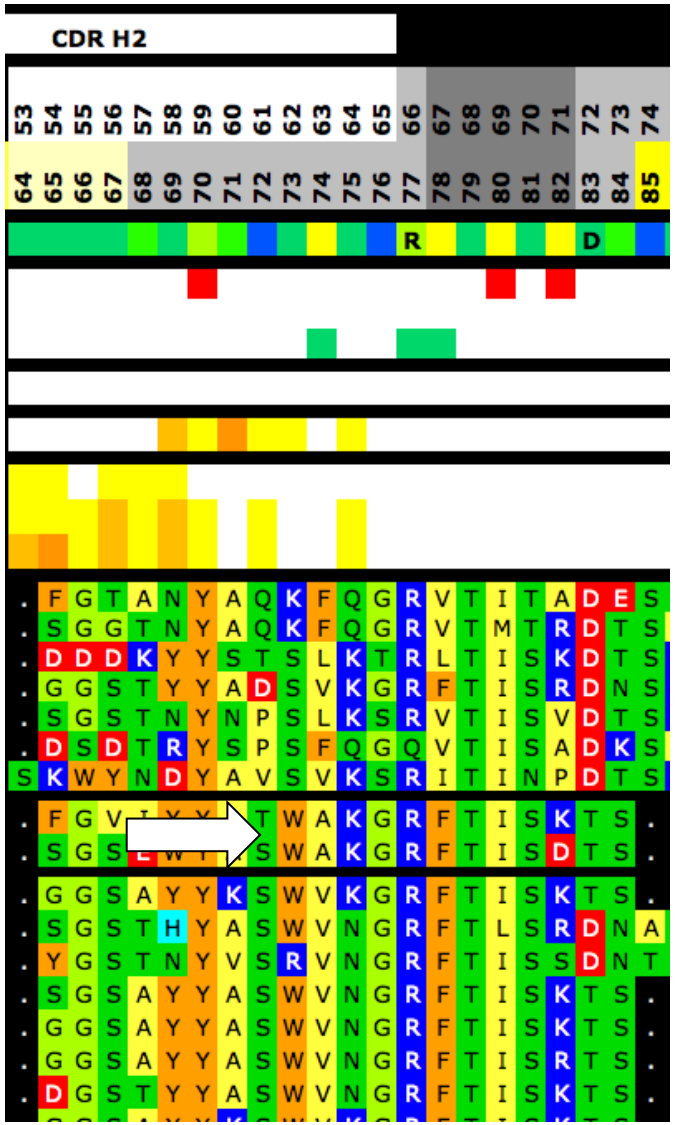
  => Destabilization of strand A'
- These V_H domains have additional Cys residues in positions H22 (H21) and H90 (H79) that can form a disulfide bond connecting strand B to strand F.
- Others have additional Cys in positions H42 (H?) and H57 (H50), allowing a disulfide bond that connects strands C and D.
- Some combine both additional disulfide bonds

  => May compensate for destabilization of N-term
- Rabbit V_H domains have a highly conserved additional Trp at the base of CDR H2
- Rabbit V_H domains show length variability in the outer loop
- Cys relatively frequent in CDR-H3 (av. 2.9% per position)

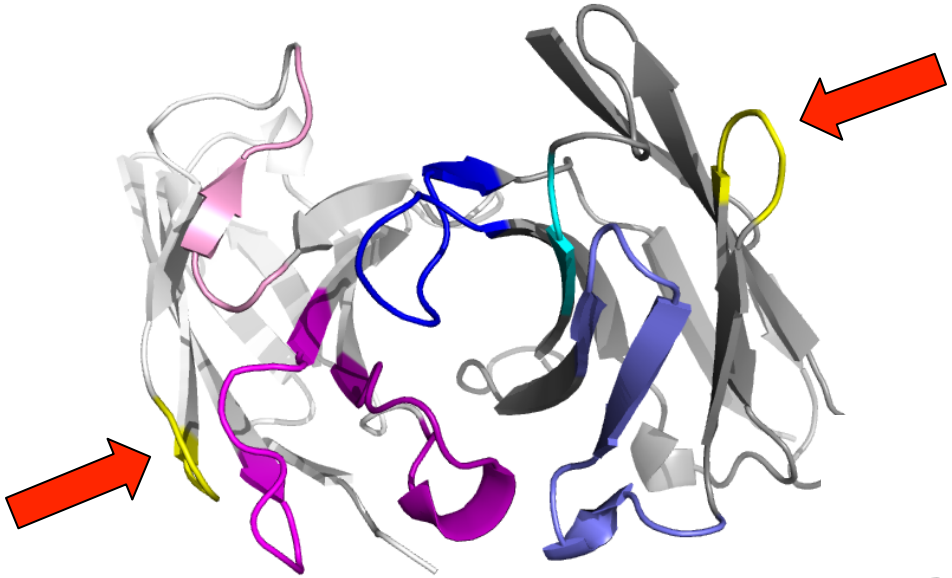
Note CDR L3 conformation!

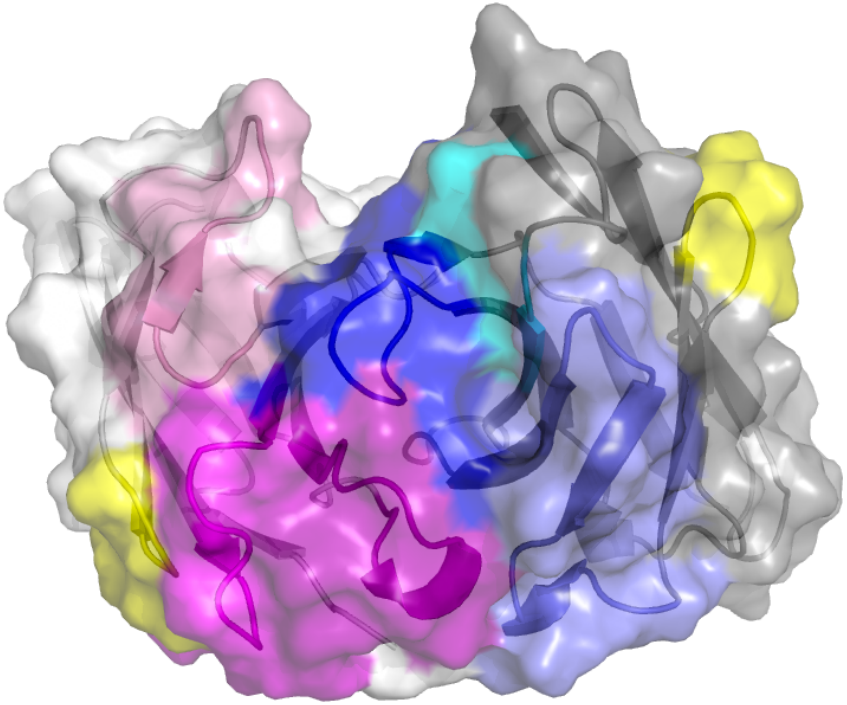


Trp H73(H62)



Outer Loops





Dinner
and
Discussion