Analyzing antibody sequence for recombinant antibody expression

Hangxing Yu, Ph.D
Senior Scientist, GenScript
May 20, 2015

GenScript
Make Research Easy

Presentation Outline

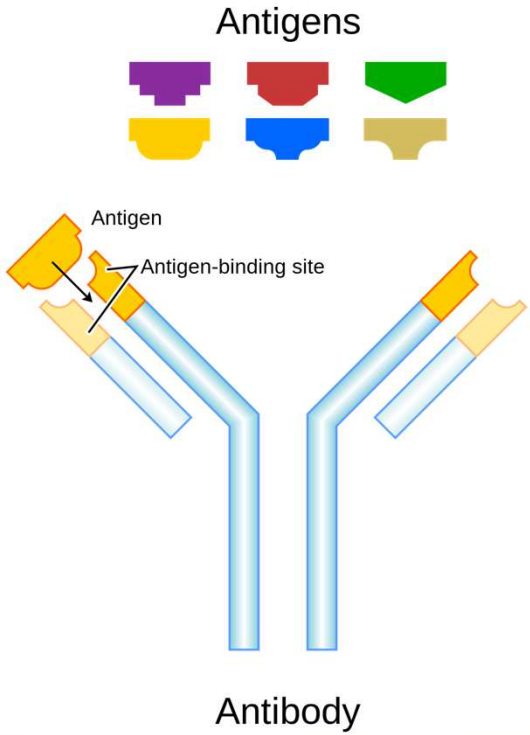
1. Antibody basics, structure and function
2. Antibody production *in vivo*
3. Antibody database and analysis
4. Optimization of antibody production

1. Antibody basics, structure and function
2. Antibody production *in vivo*
3. Antibody database and analysis
4. Optimization of antibody production

What is an Antibody?

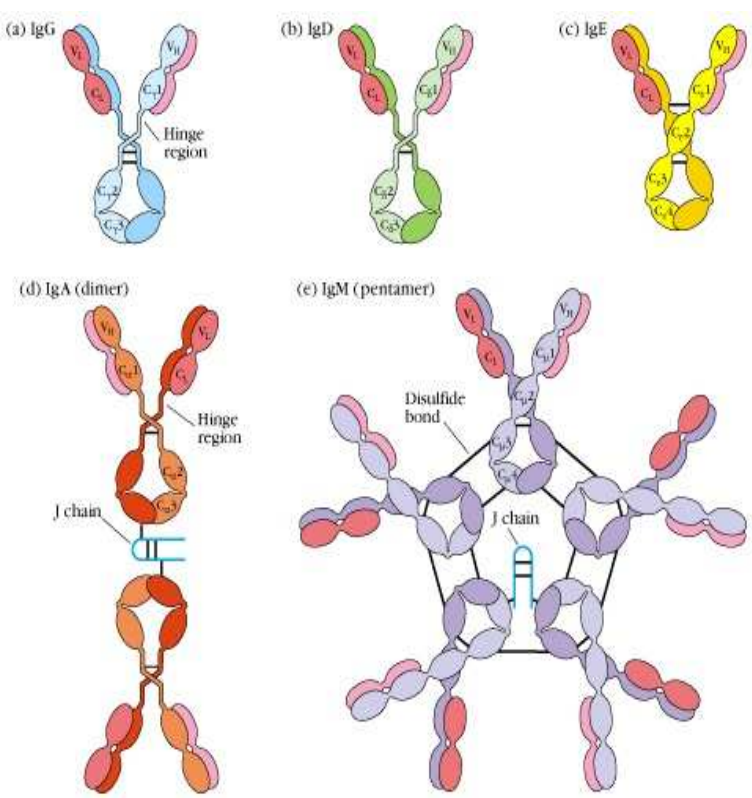
- An antibody (aka Immunoglobulin) is a large Y-shaped protein produced by B cells.

- Used by the immune system to neutralize foreign objects.

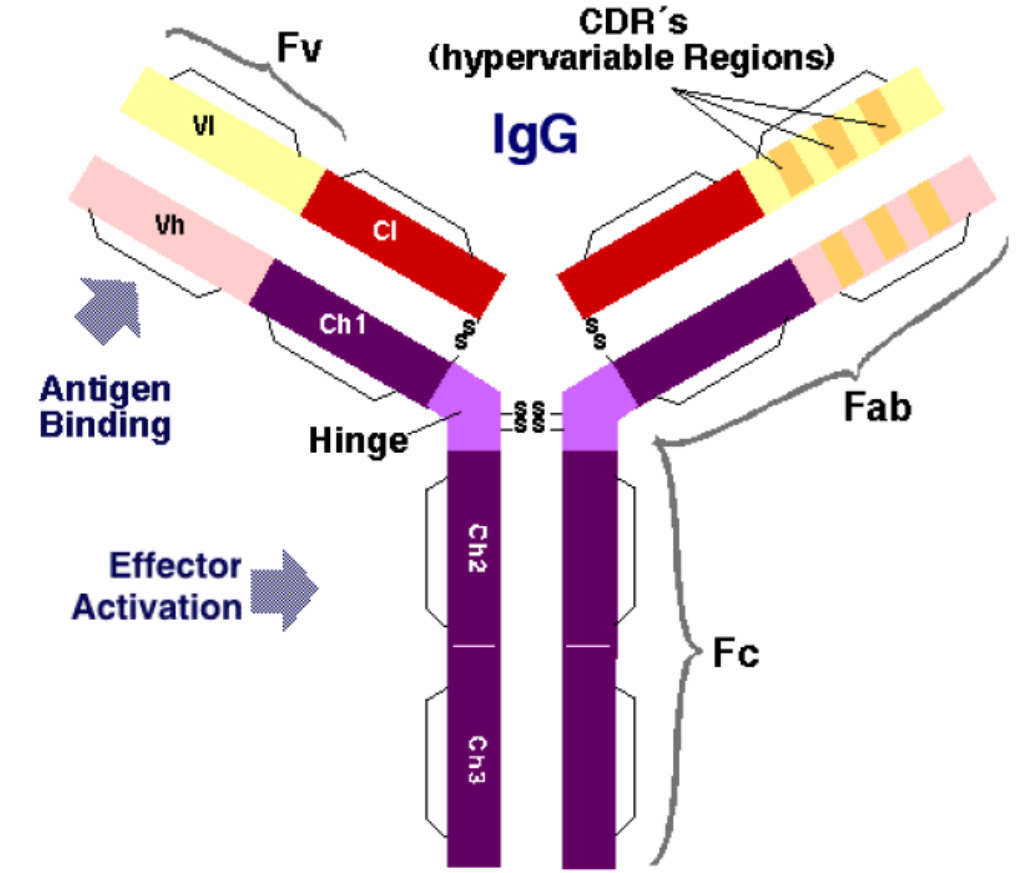


Types of Antibody *in vivo*

Name	Function	Light Chain	Heavy Chain	
IgA	Found in mucosal areas	κ or λ κ or λ	α1 α2	
IgD	Antigen receptors on B cells	κ or λ	δ	
IgE	Binds to allergen and triggers histamine release	κ or λ	ε	
IgM	Secreted from B cells with very high avidity	κ or λ	μ	
IgG	Majority of antibody-based immunity	κ or λ	γ1 γ2 γ3 γ4	γ1 γ2a γ2b γ3



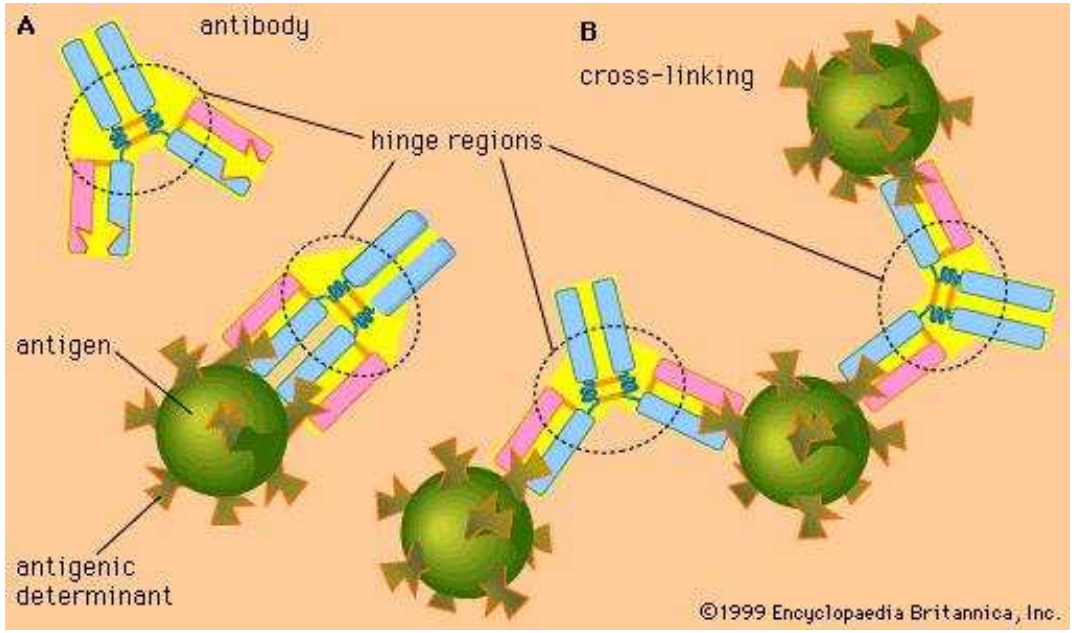
Structural and Functional Units of IgG



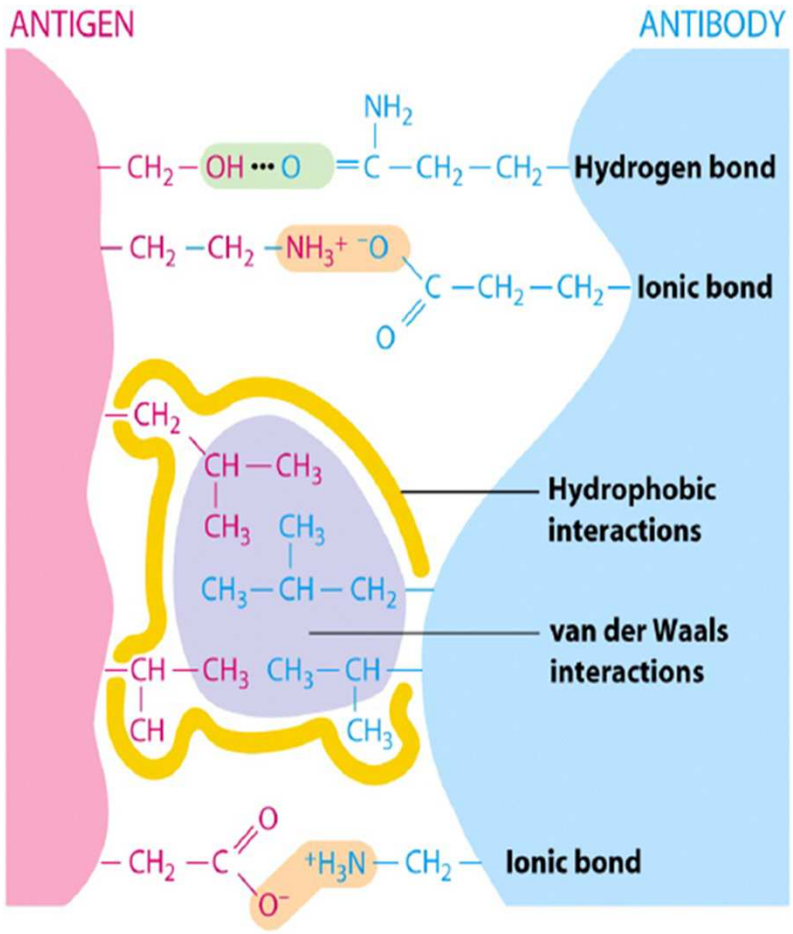
- **"Y"-shaped** molecule
- **Fab region** determines its antigen specificity. **(Fv region** and **CDR)**
- **Fc region** determines the effects.
- **Hinge region.**

Antibody-Antigen Interaction

- Each antibody binds only to a **specific** antigen.
- Ab-Ag interaction are based on **non-covalent binding** between Ab and Ag.
- The **hinge region** allow better binding.



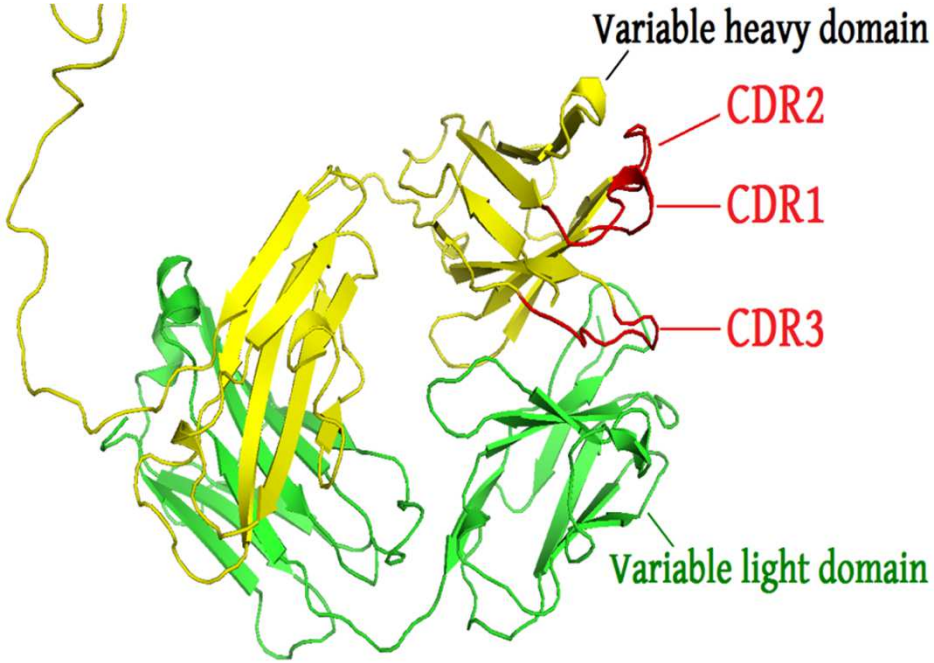
Antibody-Antigen Interaction



- Antibody and antigen interact by **spatial complementarity**.
- Weak and **non-covalent bonds.**
  electrostatic interactions
  hydrogen bonds
  Van der Waals forces
  hydrophobic interactions
- Binding between antibody and antigen is **reversible.**

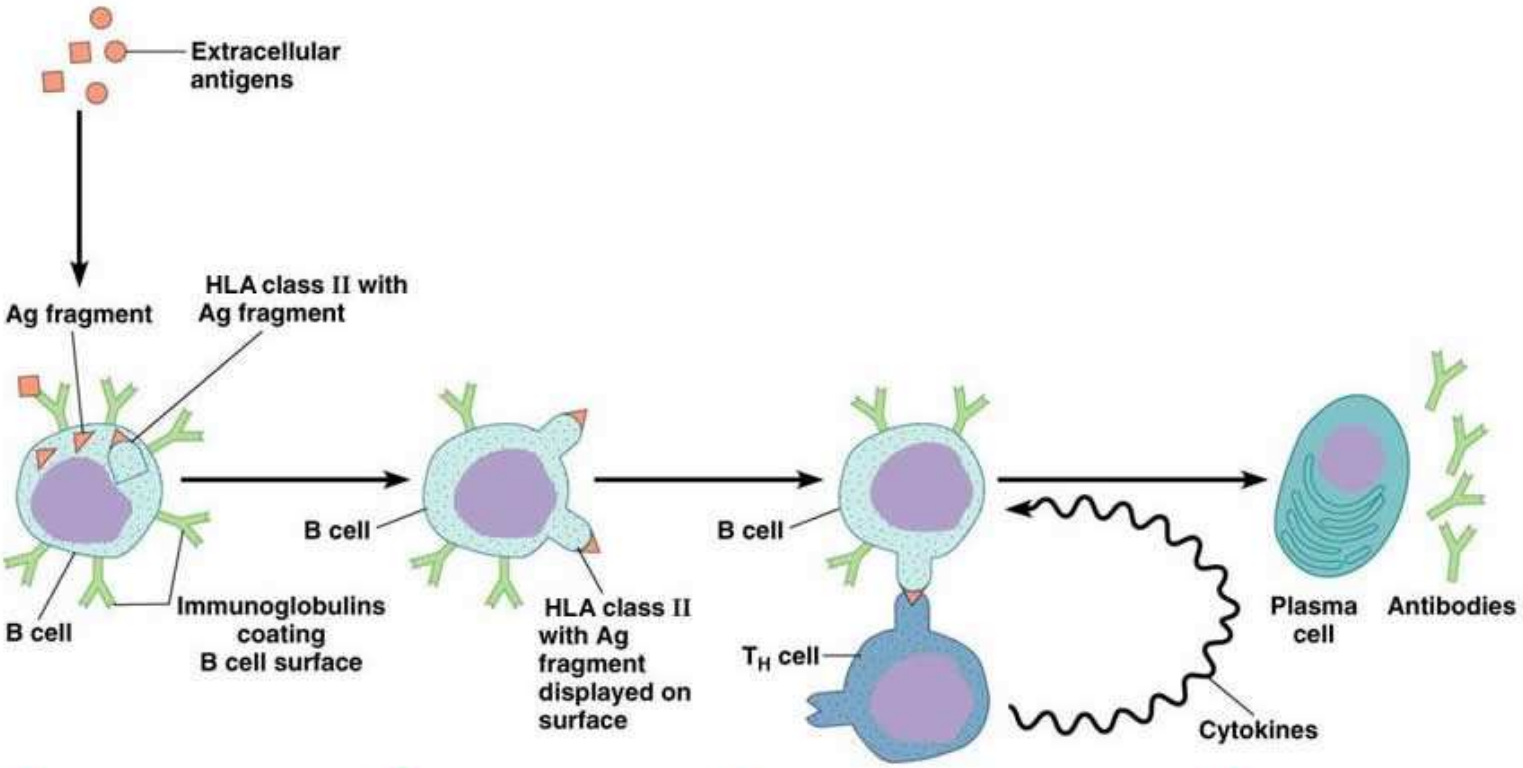
Complementarity Determining Regions (CDRs)

- Ab binding residues are mainly located in **CDRs.**
- **6 CDRs** form the combining site: L1, L2, L3, H1, H2, H3.
- **CDR-H3 and L3** have a distinctive role in antigen recognition.



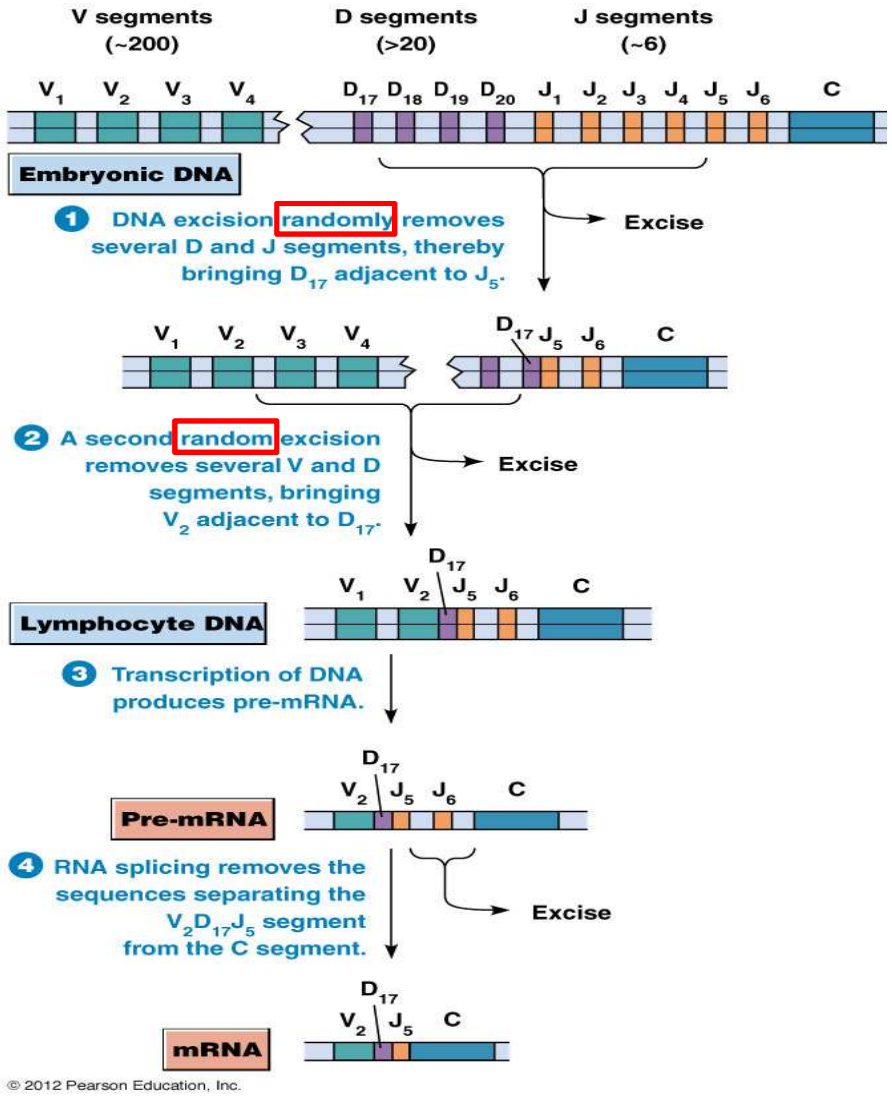
1. Antibody basics, structure and function
2. Antibody production *in vivo*
3. Antibody database and analysis
4. Optimization of antibody production

Activation of B cells to Produce Abs

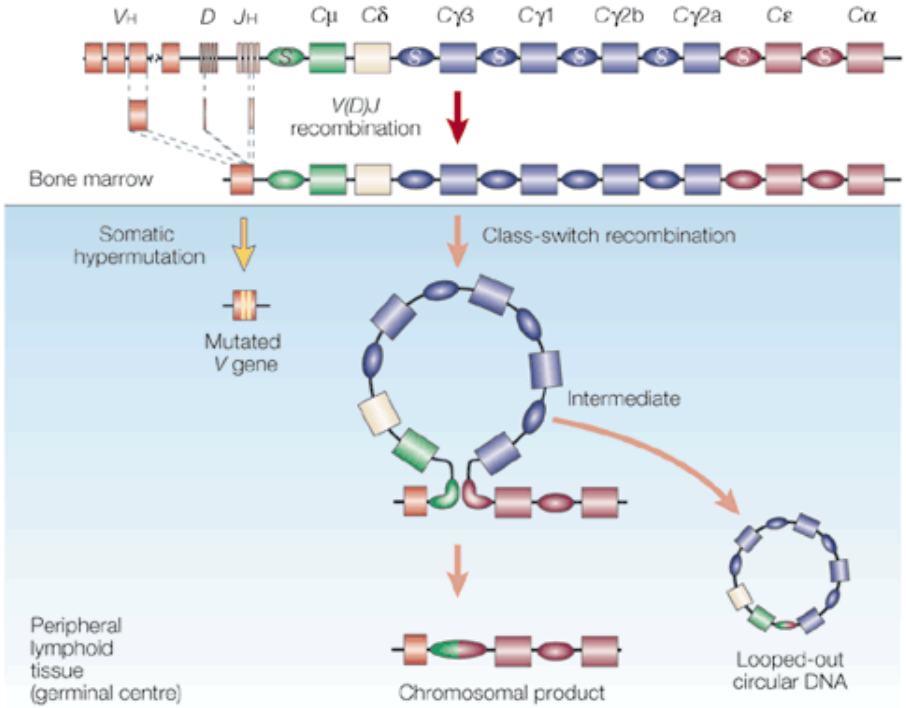


http://classes.midlandstech.edu/carterp/Courses/bio225/chap17/lecture3.htm

Antibody Gene Rearrangement



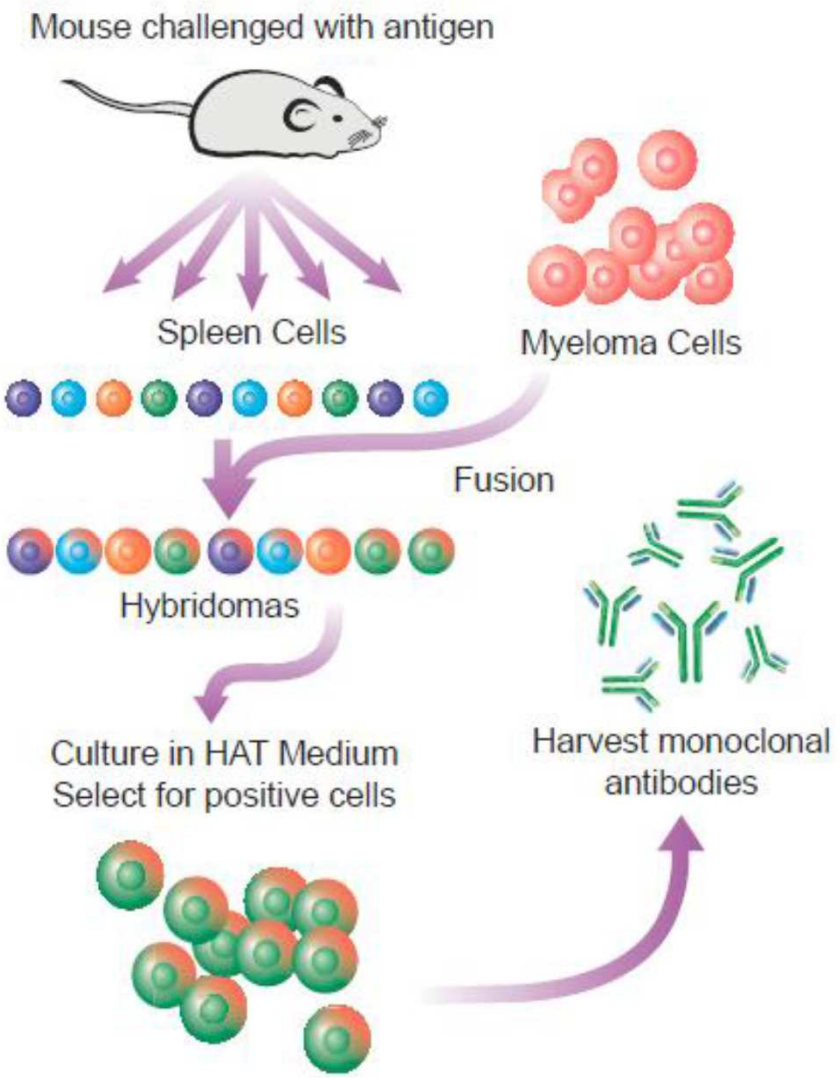
Antibody Gene Rearrangement



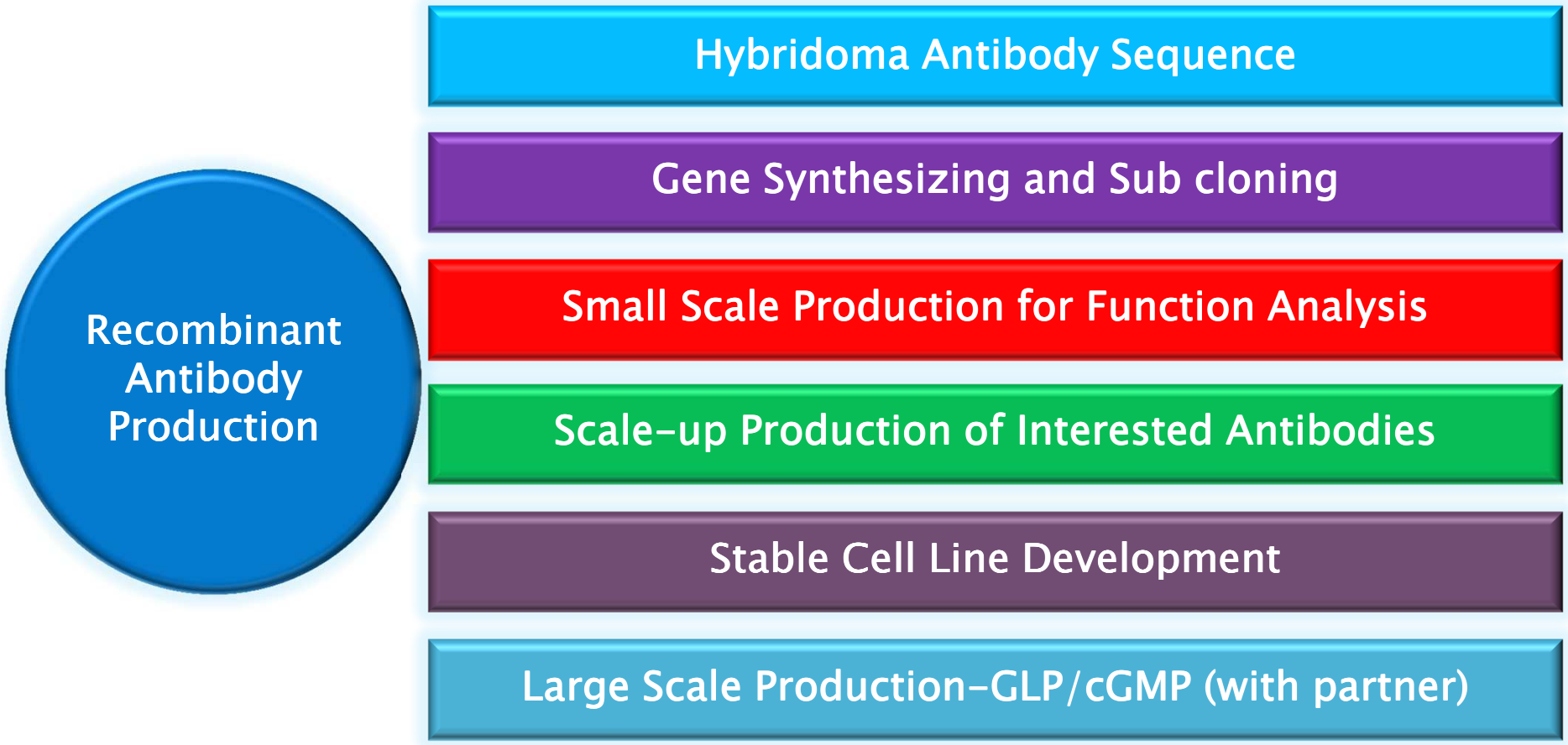
- **Somatic Hypermutation** allows slight amino acid differences in the variable domains.
- **Class-switch recombination** produce antibodies of different isotypes or subtype.

Production of mIgG by Hybridoma

- **Hybridoma:** fusion cell of specific antibody producing B cells with myeloma.
- Used for production of **monoclonal antibody.**
- Inventor: Milstein and Köhler.
- Nobel Prize in Physiology or Medicine in 1984.



Recombinant Antibody Production



1. Antibody basics, structure and function
2. Antibody production *in vivo*
3. Antibody database and analysis
4. Optimization of antibody production

Online Antibody Analysis Tool

- **Bioinf**

http://www.bioinf.org.uk/abs/labmanual/

- **IMGT**

http://www.imgt.org/IMGTrepertoire/

- **AbCheck - Antibody Sequence Test**

http://www.bioinf.org.uk/abs/seqtest.html

- **IgBlast**

http://www.ncbi.nlm.nih.gov/igblast/

- **Sequence alignment tools**

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://www.ebi.ac.uk/Tools/msa/clustalo/

Kabat Numbering

Light chain

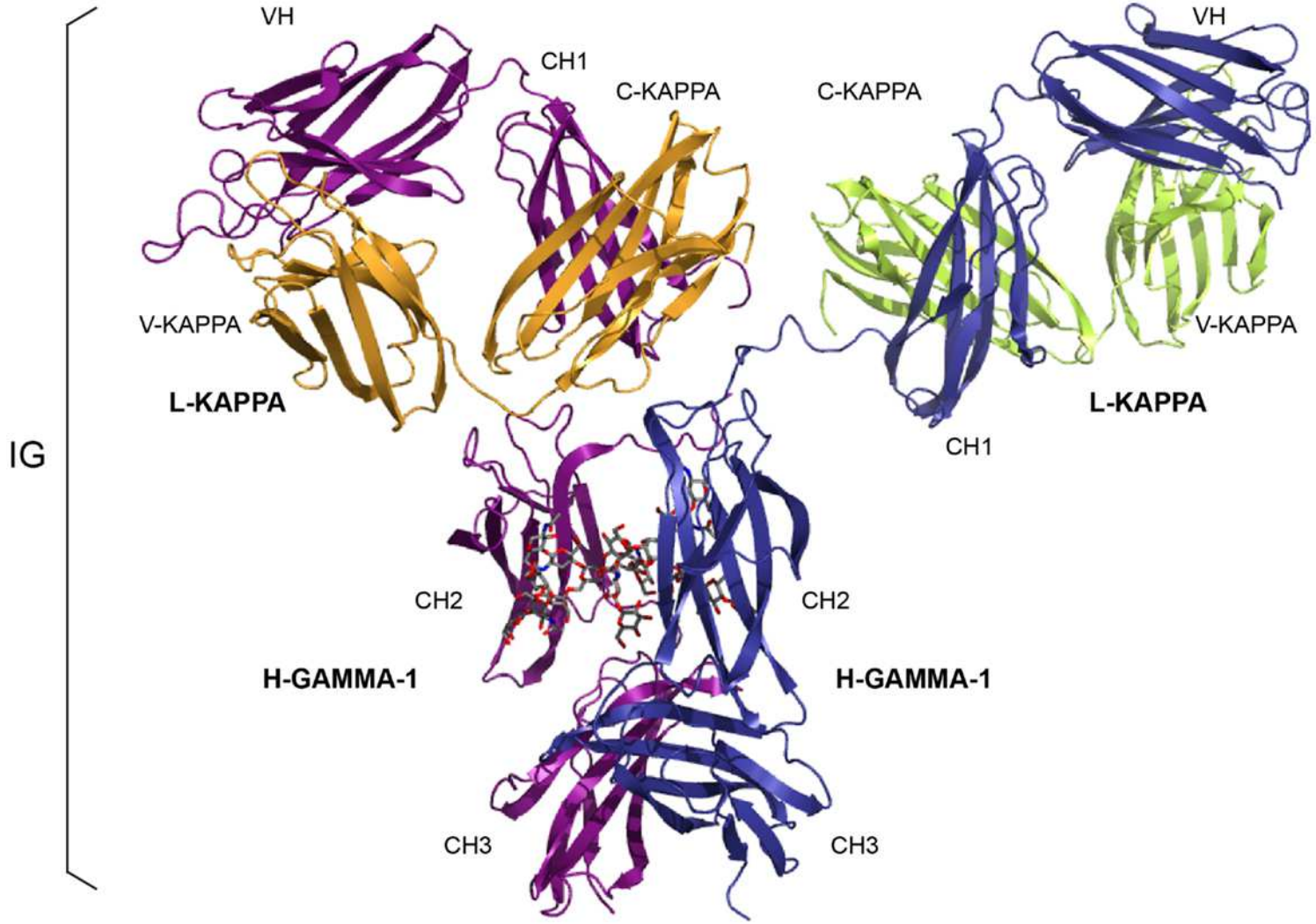
0	1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27		
27A	27B	27C	27D	27E	27F			28	29
30	31	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	58	59
60	61	62	63	64	65	66	67	68	69
70	71	72	73	74	75	76	77	78	79
80	81	82	83	84	85	86	87	88	89
90	91	92	93	94	95				
95A	95B	95C	95D	95E	95F	96	97	98	99
100	101	102	103	104	105	106			
106A							107	108	109

Heavy chain

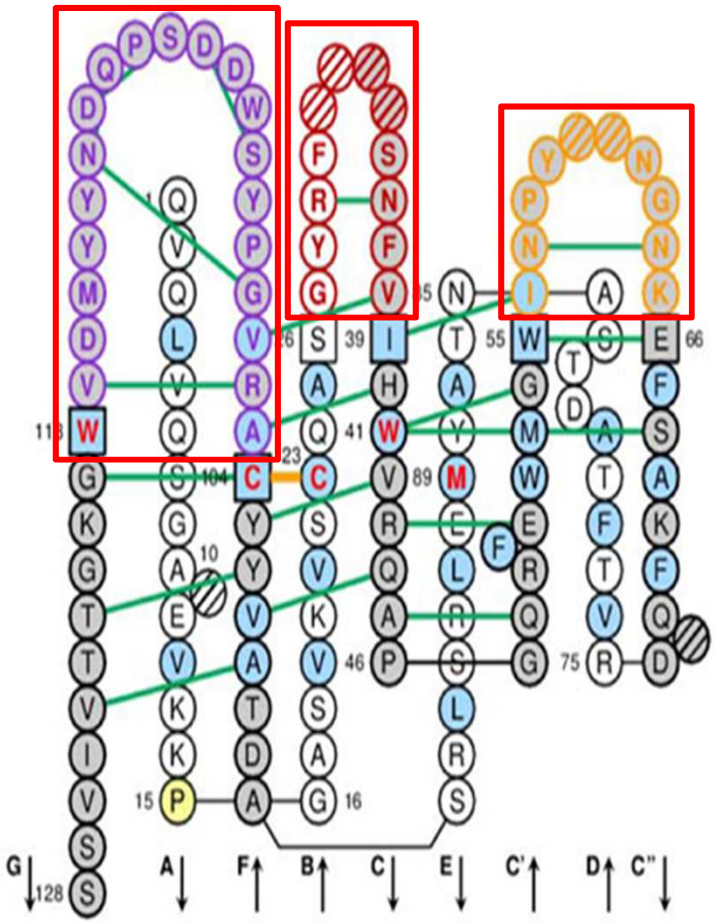
0	1	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	28	29
30	31	32	33	34	35				
35A	35B					36	37	38	39
40	41	42	43	44	45	46	47	48	49
50	51	52							
52A	52B	52C	53	54	55	56	57	58	59
60	61	62	63	64	65	66	67	68	69
70	71	72	73	74	75	76	77	78	79
80	81	82							
82A	82B	82C	83	84	85	86	87	88	89
90	91	92	93	94	95	96	97	98	99
100									
100A	100B	100C	100D	100E	100F	100G	100H	100I	100J
100K	101	102	103	104	105	106	107	108	109
110	111	112	113						

http://www.bioinf.org.uk/abs/

IMGT Analysis



IMGT Analysis – V domain



V domain strands and loops[a]	IMGT positions	Lengths[b]	Characteristic Residue@Position[c]
A-STRAND	1–15	15 (14 if gap at 10)	
B-STRAND	16–26	11	1st-CYS 23
BC-LOOP	27–38	12 (or less)	
C-STRAND	39–46	8	CONSERVED-TRP 41
C′-STRAND	47–55	9	
C′C″-LOOP	56–65	10 (or less)	
C″-STRAND	66–74	9 (or 8 if gap at 73)	
D-STRAND	75–84	10 (or 8 if gaps at 81, 82)	
E-STRAND	85–96	12	hydrophobic 89
F-STRAND	97–104	8	2nd-CYS 104
FG-LOOP	105–117	13 (or less, or more)	
G-STRAND	118–128	11 (or 10)	V-DOMAIN J-PHE 118 or J-TRP 118[d]

Identifying CDRs by sequence

CDR-L1

Start	Approx residue 24
Residue before	**always** a Cys
Residue after	**always** a Trp. Typically Trp-Tyr-Gln, but also, Trp-Leu-Gln, Trp-Phe-Gln, Trp-Tyr-Leu
Length	10 to 17 residues

CDR-L2

Start	**always** 16 residues after the end of L1
Residues before	generally Ile-Tyr, but also, Val-Tyr, Ile-Lys, Ile-Phe
Length	**always** 7 residues (except NEW (7FAB) which has a deletion in this region)

CDR-L3

Start	**always** 33 residues after end of L2 (except NEW (7FAB) which has the deletion at the end of CDR-L2)
Residue before	**always** Cys
Residues after	**always** Phe-Gly-XXX-Gly
Length	7 to 11 residues

Identifying CDRs by sequence

CDR-H1

Start	Approx residue 26 (**always** 4 after a Cys) [Chothia / AbM defintion]; Kabat definition starts 5 residues later
Residues before	**always** Cys-XXX-XXX-XXX
Residues after	**always** a Trp. Typically Trp-Val, but also, Trp-Ile, Trp-Ala
Length	10 to 12 residues [AbM definition]; Chothia definition excludes the last 4 residues

CDR-H2

Start	**always** 15 residues after the end of Kabat / AbM definition) of CDR-H1
Residues before	typically Leu-Glu-Trp-Ile-Gly, but a number of variations
Residues after	Lys/Arg-Leu/Ile/Val/Phe/Thr/Ala-Thr/Ser/Ile/Ala
Length	Kabat definition 16 to 19 residues; AbM (and recent Chothia) definition ends 7 residues earlier

CDR-H3

Start	**always** 33 residues after end of CDR-H2 (**always** 2 after a Cys)
Residues before	**always** Cys-XXX-XXX (typically Cys-Ala-Arg)
Residues after	**always** Trp-Gly-XXX-Gly
Length	3 to 25(!) residues

Identifying CDRs by sequence

Region	Position	Sequence Characterization	Length
CDR-L1	24-34	C----------W/WYQ/WLQ/WFQ/WYL	10-17
CDR-L2	50-56	IY/VY/IF-------	7
CDR-L3	89-97	C-------FG*G	7-11

Light chain: Amino acids sequence (237 AA)

Leader sequence-FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4-Constant region-Stop codon

MKLPVRLLVLMFWIPASSSDVVMTQIPLSLPVSLGDQASIS**C**RSSQSLVHSNGNTYLH**WYL**QKPGQ
SPKLL**IY**KVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYF**C**SQSTHVPT**FGGG**TKLEIKRADAA
PTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVKWKIDGSERQNGVLNSWTDQDSKDSTYSMSST
LTLTKDEYERHNSYTCEATHKTSTSPIVKSFNRNEC

Identifying CDRs by sequence

Region	Position	Sequence Characterization	Length
CDR-H1	31-35B	C****----------W/WV/WI/WA	10-12
CDR-H2	50-65	LQWIG-------K/R|L/I/V/F/T/A|T/S/I/A	16-19
CDR-H3	95-102	C**-------WG*G	3-25

Heavy chain: Amino acids sequence (458 AA)

Leader sequence-FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4-Constant region-Stop codon

MGWSCIFLFLLSVTVGVFSEVQLQQSGPELVKPGASVKIS**CKASG**YSFTGYFMN**WV**KQSHGKS**LE**
WIGRINPNNGDSLYNQKFKG**KAT**LTVDKSSTTAHMELLSLTSEDSAVYY**CGR**DYFFDY**WGQG**TTL
TVSSAKTTPPSVYPLAPGSAAQTNSMVTLGCLVKGYFPEPVTVTWNSGSLSSGVHTFPAVLQSDLYT
LSSSVTVPSSTWPSETVTCNVAHPASSTKVDKKIVPRDCGCKPCICTVPEVSSVFIFPPKPKDVLTITLT
PKVTCVVVDISKDDPEVQFSWFVDDVEVHTAQTQPREEQFNSTFRSVSELPIMHQDWLNGKEFKC
RVNSAAFPAPIEKTISKTKGRPKAPQVYTIPPPKEQMAKDKVSLTCMITDFFPEDITVEWQWNGQPA
ENYKNTQPIMDTDGSYFVYSKLNVQKSNWEAGNTFTCSVLHEGLHNHHTEKSLSHSPGK

Ab Check - Sequence Analysis

Info

Sequences

Structures

Software

Humanization patents

Links

Awards

AbCheck - Antibody Sequence Test

This page allows you to test an antibody sequence against the Kabat sequence database. Any unusual residues (occurring in < 1% of chains in the database) will be reported to you. This allows the identification of potential cloning artifacts and sequencing errors. The current Kabat database contains 6014 light chains and 7895 heavy chains.

Click here for details of the method used for checking your antibody sequence.

If you get any error or warning messages, please check you have entered your sequence correctly. Strange sequence features may cause the alignment stage to fail. Loops longer than anything observed in the current Kabat database will also cause the alignment to fail. If, after checking your sequence, you *still* get errors or warnings, please send me EMail: andrew@bioinf.org.uk and I'll see if the programs can be modified to accomodate your sequence. Alternatively, your antibody may just be very strange!

Identify Chothia canonical classes for your antibody sequence.

Enter the **amino acid** sequence (1-letter code) of your Fv fragment (optionally you may include the whole Fab fragment, but only the Fv portion will be tested).

Enter the light chain **amino acid sequence** here:

```
DVVMTQIPLSLPVSLGDQASISCRSSQSLVHSNGNTYLHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKI
SRVEAEDLGVYFCSQSTHVPTFGGGTKLEIK
```

Enter the heavy chain **amino acid sequence** here:

```
EVQLQQSGPELVKPGASVKISCKASGYSFTGYFMNWVKQSHGKSLEWIGRINPNNGDSLYNQKFKGKATLTVDKSSTTAH
MELLSLTSEDSAVYYCGRDYFFDYWGQGTTLTVSS
```

Ab Check - Sequence Analysis

Antibody Sequence Test Results

If no results appear below, please send me EMail with your sequence as an error has occurred!

If you get lots of errors, please make sure you have used the AMINO ACID sequence (not DNA) and don't have any leader sequence residues

Kabat alignment results

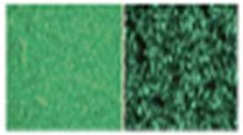
```
Alignment with consensus:
DVVMTQIPLSLPVSLGDQASISCRSSQSLVHSNGNTY-LHWYLQKPGQSPKLLIYKVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDLGVYFCSQSTHVPT-------FGGGTKLEIK---
~AVLTQPPXS!%!S!GXXVTI%CXXSXXXXXXXXXXXX!XWYQQKXGXXPK!LIYXX%XXXSGVPXRFSGS!SGTXX%LXISX!XXEDX!XY#CXXXXXXXXXXXXXXXFGXGTKLEIXKRA

EVQLQQSGPELVKPGASVKISCKASGYSFTGYFMN--WVKQSHGKSLEWIGRINPNN--GDSLYNQKFKGKATLTVDKSSTTAHMELLSLTSEDSAVYYCGRDYFFDY-------------------WGQGTTLTVSS
XVQLXXSGXXL!XPGXS!$!SCX!SG#%F%XXXXXXXWV$QXPG$XLEW!!XIXXXXXXGXXXYXXXXK!$XX!%XDXSXX%!YXXXXSLXXED%AXYYCXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXWGQGTXVTVSS

L1 D
L2 V
L3 V
L4 M
L5 T
L6 Q
L7 I
L8 P
L9 L
L10 S
L11 L
L12 P
L13 V
L14 S
L15 L
L16 G
L17 D
L18 Q
```

The following unusual sequence features have been identified:

- L7 = I in 0.542% of light chains (17 examples)
- L106A = K in 0.542% of light chains (17 examples)

1. Antibody basics, structure and function
2. Antibody production *in vivo*
3. Antibody database and analysis
4. Optimization of antibody production

Rapid Transient Gene Expression



Analyzed over 1000 sequences for Ab production

GenScript is the largest gene supplier in the world

Success rate of ≥ 95% with strategy to optimize transfection ratio of heavy and light chains

High success rate ≥ 95% with rigorous optimization strategies

Sequence Design → Codon Optimization → Gene Synthesis → Cloning → Transfection → Expression Evaluation → Large-scale Production

OptimumGene™ codon optimization technology to maximize chances of successful Ab binding

Validated fermentation platform with 100% success rate in the past 5 years

100ml ~ 10L in shake flask

10L~200L in Wave/SUB bioreactor

Advantages of GenScript Recombinant Abs

- **Advantages:**
  - Prominent 293 and CHO suspension systems promote large-scale and fast-delivery of recombinant protein and monoclonal antibody up to gram level.
  - A strong commitment of scientific teams with many years of experience in developing highly productive stable cell lines with strict expression stability evaluation.
  - Comprehensive professional process development enhances the protein expression capacity of cell lines.
  - GenScript has successfully delivered >500 proteins in mammalian expression system with >100 stable cell lines.

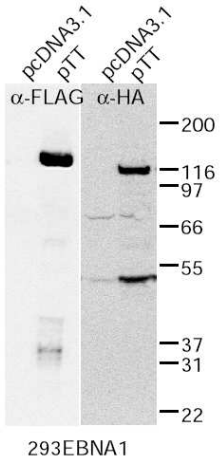
Licensed NRC Technologies

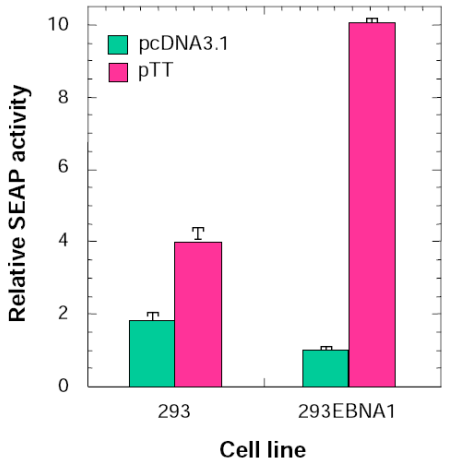
HEK293

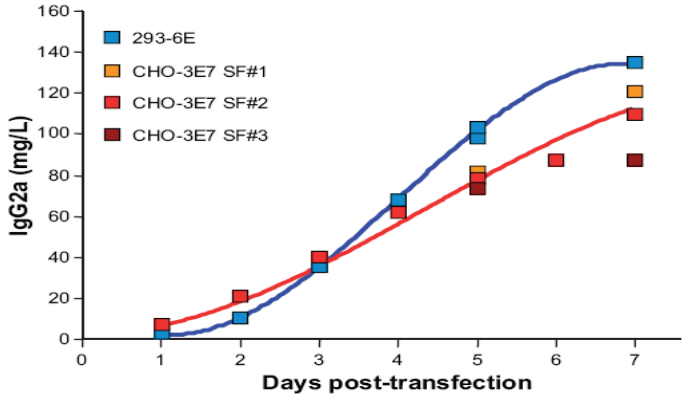
- Highly stable 293EBNA1 clone
- Increase recombinant protein yields through co-expression
- Afford the advantages of serum-free production
- Rapid and high yield recombinant protein production
- Availability of a Master Cell Bank (MCB) for cGMP production

CHO-pTT

- Provides comparable protein expression level to those of HEK293 system
- Rapid production of recombinant proteins that require post-translation modification for research or pharmaceutical use.







Transient gene expression in CHO-3E7 cells using an optimized process and pTT vectors yield antibody titers closely matching those obtained in 293 cells (LODs 11266 & 11565)

National Research Council Canada | Conseil national de recherches Canada

MamPower™ Guaranteed Services

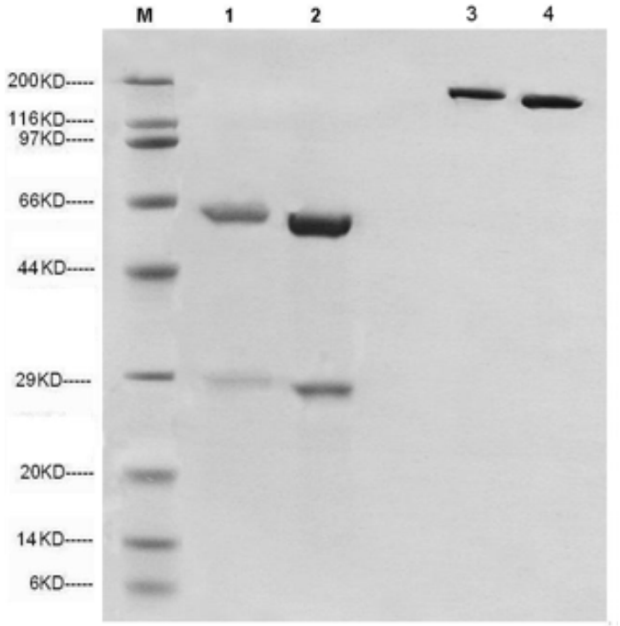
Customer provides	Amount	Purity & endotoxin level options	Timeline	Deliverables	Price starting from
Antibody sequence	5mg	• ≥95% by SDS-PAGE	5-6 weeks	• LC and HC in pUC57 vector • Optimized gene sequence report • Purified antibody at listed amount and purity • QC data	$2,999
	10mg 25mg 50mg 100mg	• ≥95% by SDS-PAGE • Endotoxin ≤10EU/mg • Concentration ≥1mg/ml	6-8 weeks		
	250mg 500mg	• ≥95% by HPLC • Endotoxin ≤1EU/mg	8-10 weeks		

➢ Key features of MamPower™ guaranteed recombinant antibody production services

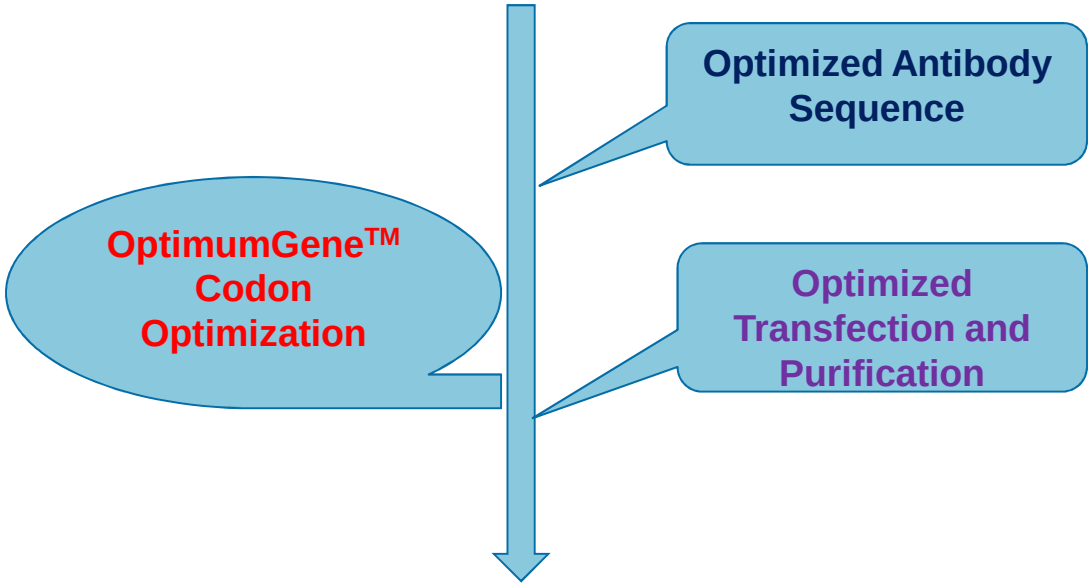
- Gene synthesis is included in the package – no additional cost to you.
- Guarantee recombinant antibody amount, purity and desired endotoxin level – no cost to you if we don't deliver as promised.
- Flexible scale from 5mg to 500mg.
- Competitive price.
- Fast turnaround time - as little as 5 weeks.

Case Study

384390



- Expression level was low (2.5 mg/L)

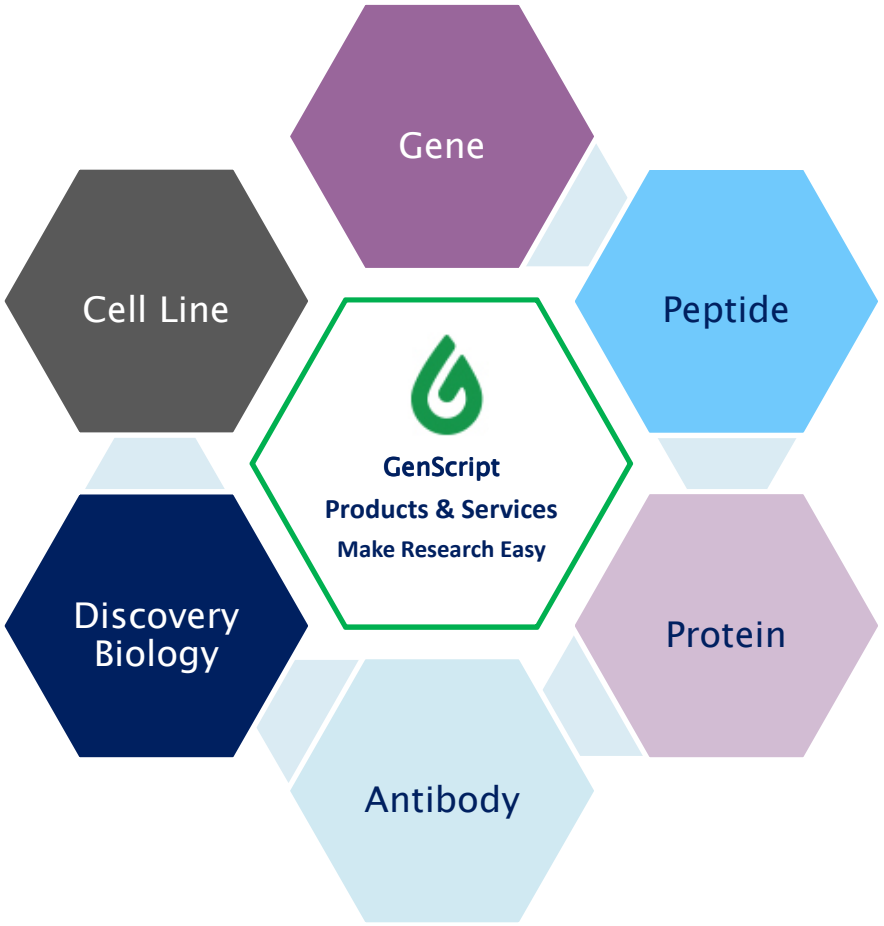


- Expression level was significantly increased (25 mg/L)

Summary

- Antibody is used by the immune system to neutralize foreign objects (antigens).
- Antibody-antigen binding residues are mainly located in CDRs.
- Analysis of sequence could assist to evaluate the expression level of a certain antibody.
- GenScript could offer you the professional service for optimized recombinant antibody production.

About GenScript



Thank you

Upcoming webinar

Expression vectors: how to choose, or customize, vectors for gene & protein expression
Presented by Rachel Speer, Ph.D, Technical Writer, GenScript
June 3, 2015/11:00 AM EST

Visit our website for archived webinars: http://www.genscript.com/webinars.html

Contact today's presenter: hangxing.yu@genscript.com

Thank you