Protein Engineering

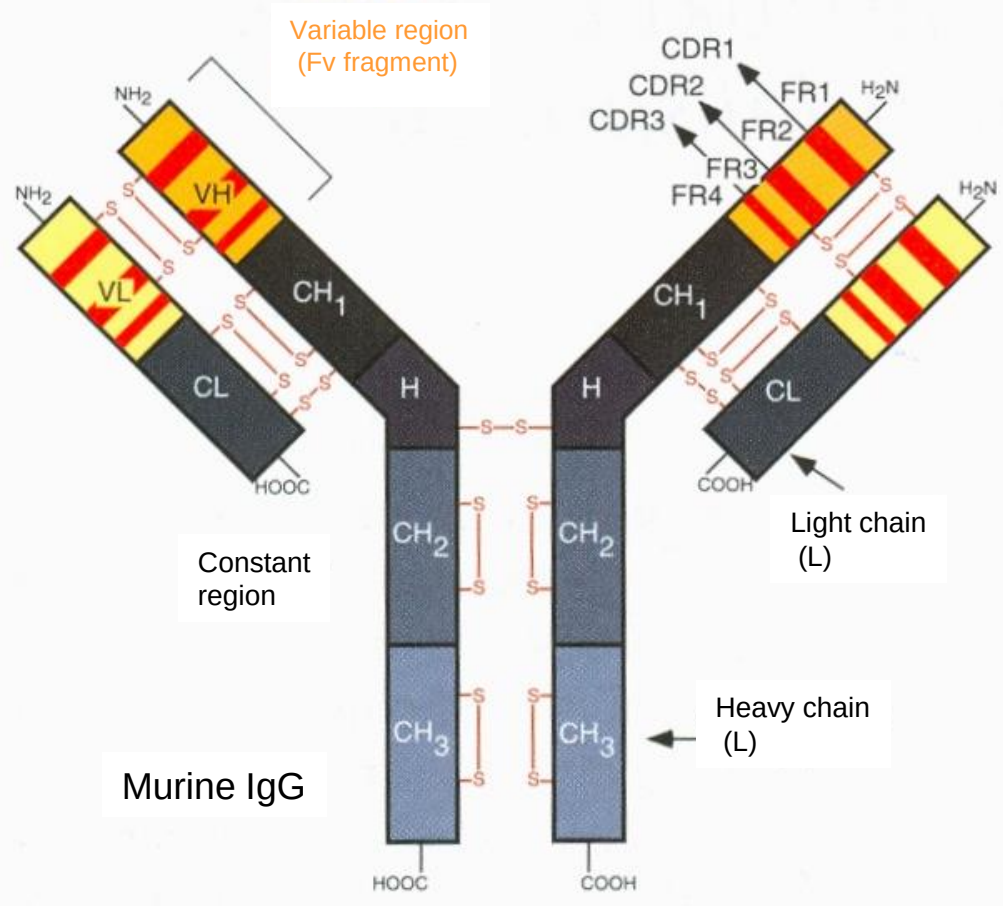
Antibodies

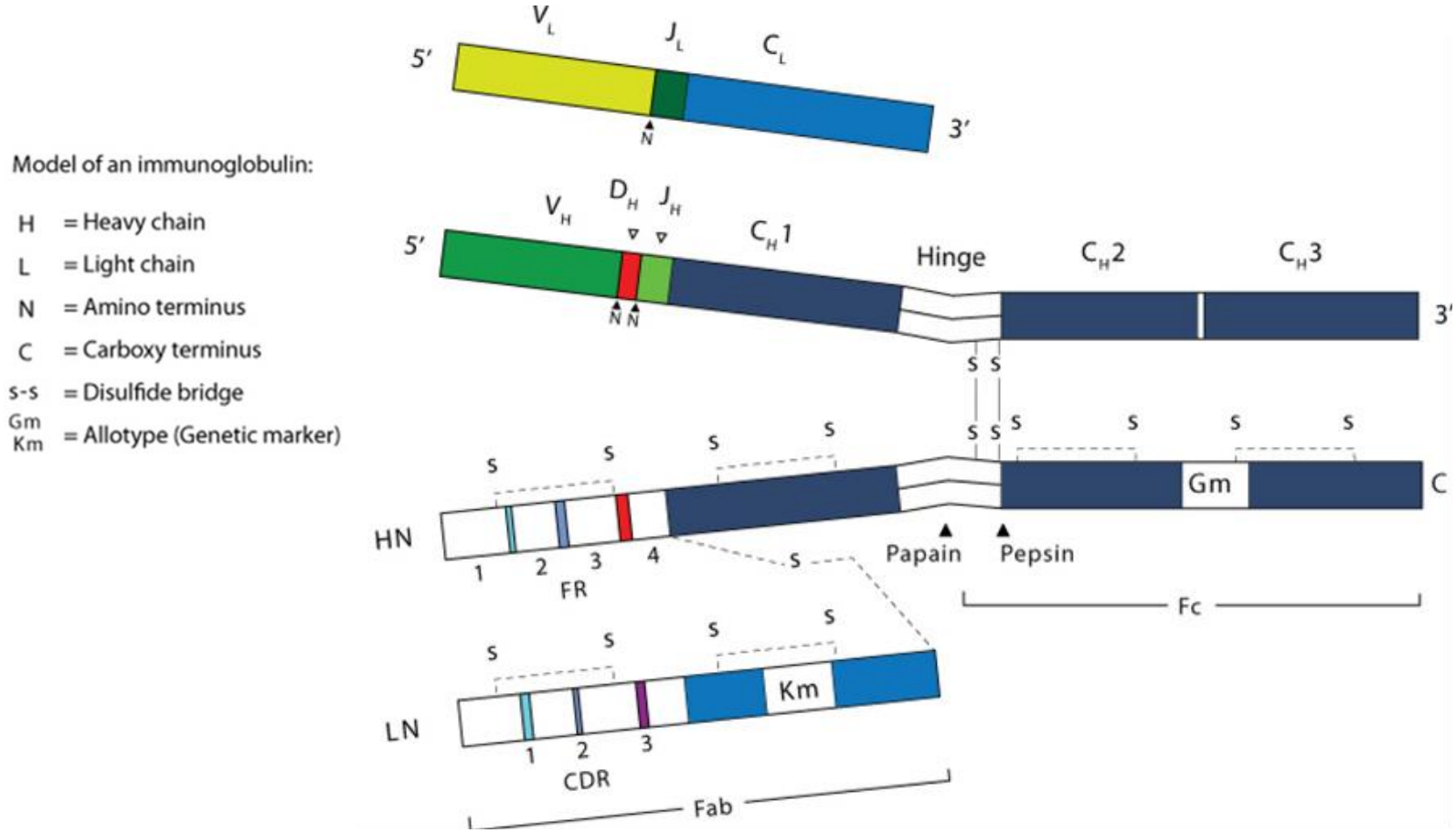
Antibody Engineering

Target: Specific Affinity

General structure of IgG antibodies

IgG antibodies consist of two identical light and two identical heavy chains. These four chains are arranged in parallel and linked with different disulfide bonds to form a molecule. The antigen binding regions (VH = variable region of the heavy chain; VL = variable region of the light chain) are located at the N-terminus of each protein chain. In this region the antibodies can vary tremendously. “Framework regions” (FR1-FR4) are distinguished from “complementary determining regions” (CDR1-CDR4). This part of the antibody is named Fv-region. CL and OH are domains in the constant region of heavy and light chains. The region marked with “H” is the co-called “Hinge” region.

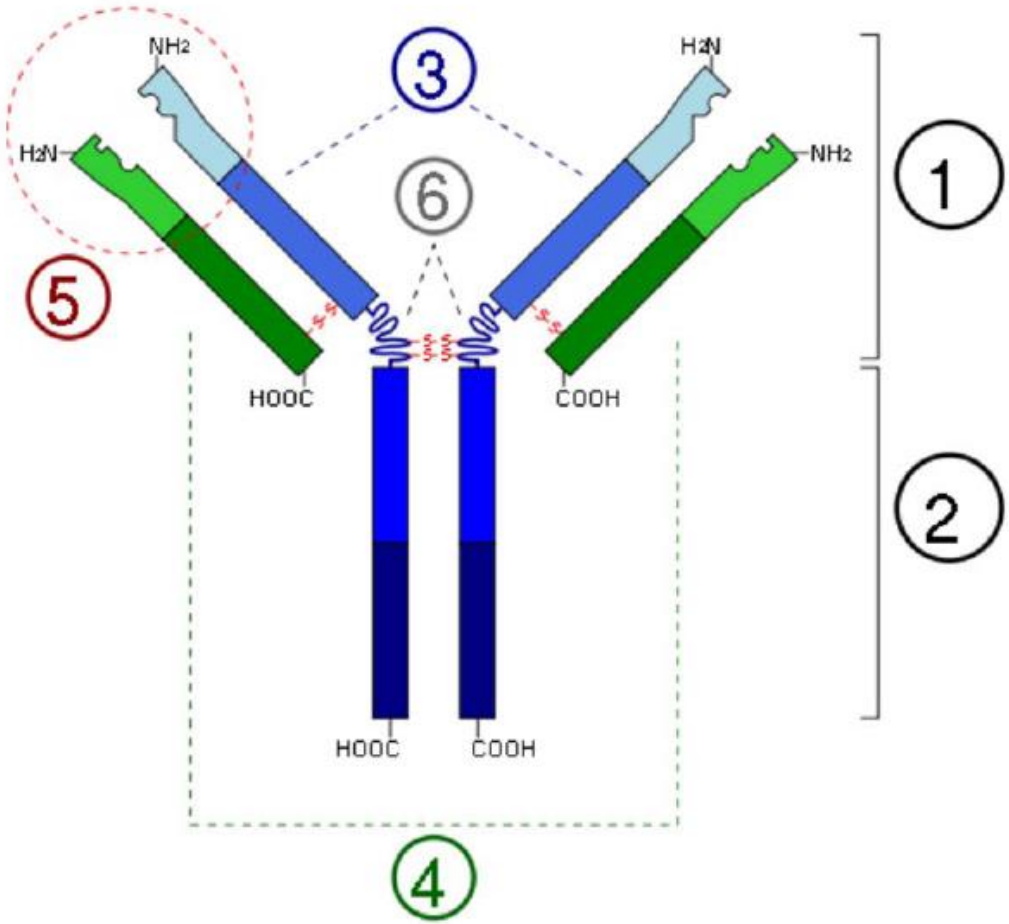




IgG antibodies consist of 2 heavy (H) and 2 light (L) chains where the L chain can consist of either a κ or a λ chain. Each component chain contains one NH_2-terminal variable (V) IgSF domain and 1 or more COOH-terminal constant (C) IgSF domains, each of which consists of 2 sandwiched b-pleated sheets pinned together by a disulfide bridge between 2 conserved cysteine residues. Papain digests IgG into 2 Fab fragments, each of which can bind antigen, and a single Fc fragment. Pepsin splits IgG into an Fc fragment and a single dimeric F(ab)2 that can cross-link, as well as bind, antigens. The Fab contains 1 complete L chain in its entirety and the V and C_H1 portion of 1 H chain. The Fab can be further divided into a variable fragment (Fv) composed of the V_H and V_L domains, and a constant fragment composed of the C_L and C_H 1 domains. The primary sequence of the V domain is functionally divided into 3 hypervariable intervals termed complementarity-determining regions (CDRs) that are situated between 4 regions of stable sequence termed framework regions (FRs).

J Allergy Clin Immunol. 2010 Feb; 125(2 0 2): S41–S52. http://dx.doi.org/10.1016/j.jaci.2009.09.046

Antibody Engineering

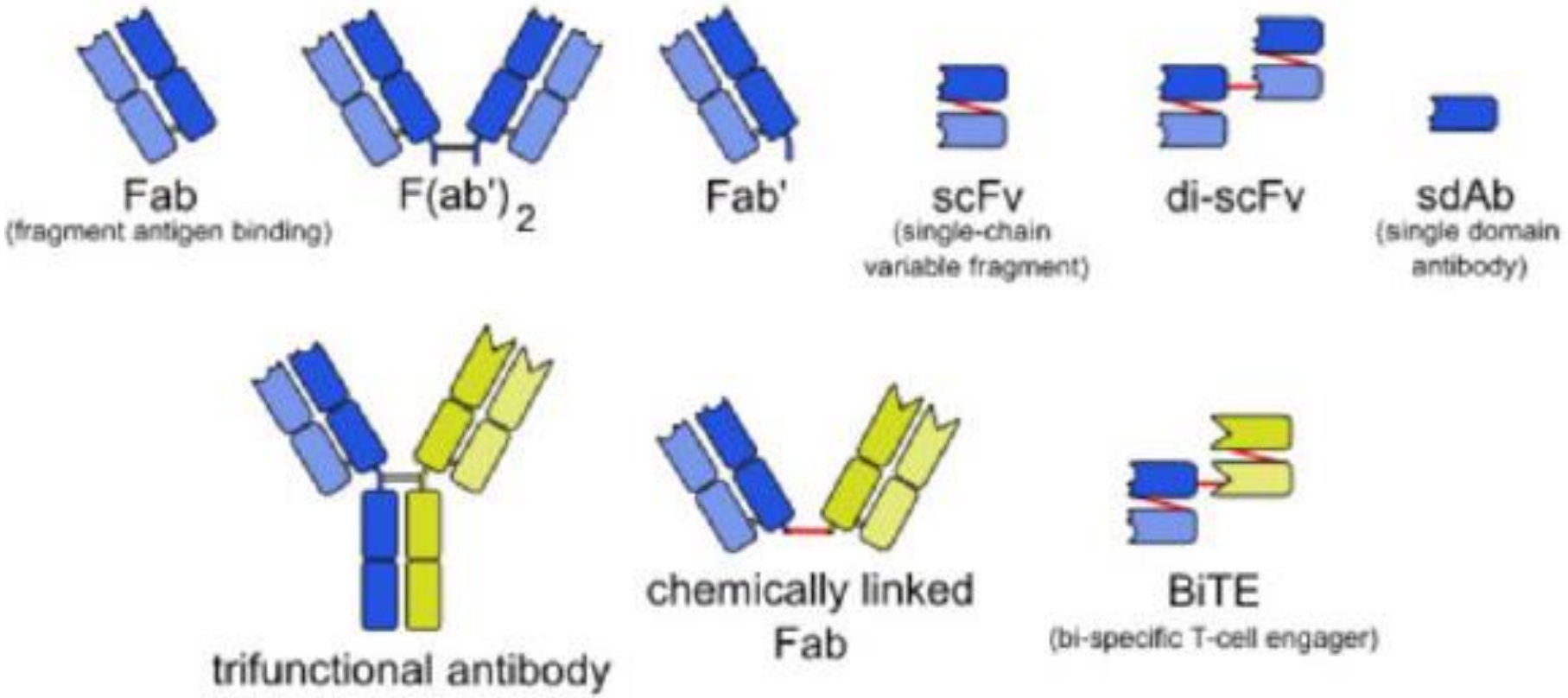


1. Fab region
2. Fc region
3. Heavy chain with one variable (VH) domain followed by a constant domain (CH1), a hinge region, and two more constant (CH2 and CH3) domains.
4. Light chain with one variable (VL) and one constant (CL) domain
5. Antigen binding site (paratope)
6. Hinge regions.

http://en.wikipedia.org

Antibody Engineering

Synthetic Antibody fragments



http://en.wikipedia.org

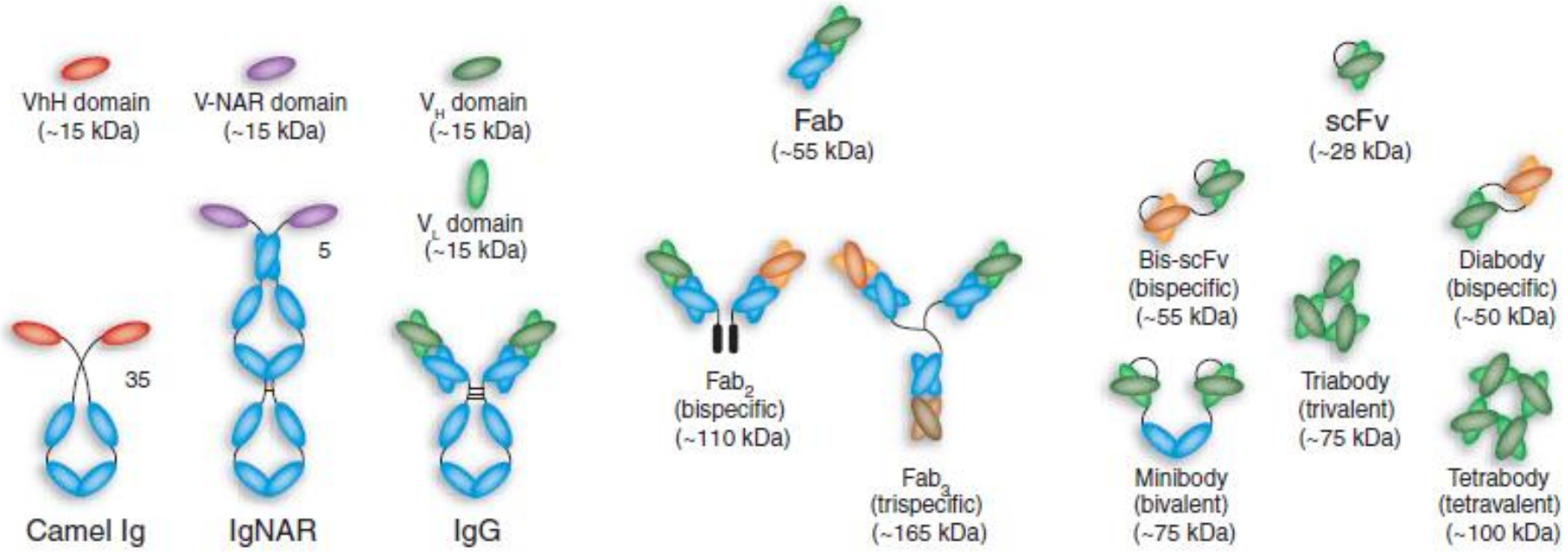
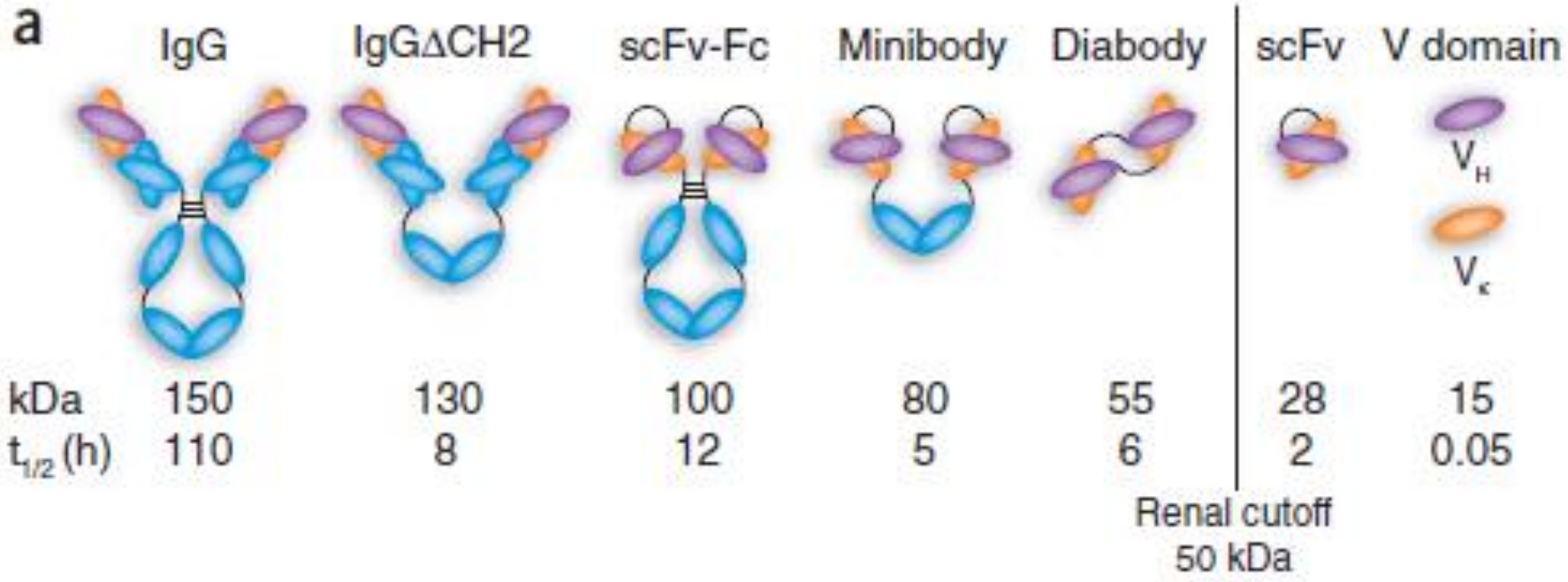


Figure 1 Schematic representation of different antibody formats, showing intact 'classic' IgG molecules alongside camelid VhH-Ig and shark Ig-NAR immunoglobulins. Camelid VhH-Ig and shark Ig-NARs are unusual immunoglobulin - like structures comprising a homodimeric pair of two chains of V-like and C-like domains (neither has a light chain), in which the displayed V domains bind target independently. Shark Ig-NARs comprise a homodimer of one variable domain (V-NAR) and five C-like constant domains (C-NAR). A variety of antibody fragments are depicted, including Fab, scFv, single-domain VH, VhH and V-NAR and multimeric formats, such as minibodies, bis-scFv, diabodies, triabodies, tetrabodies and chemically conjugated Fab´ multimers (sizes given in kilodaltons are approximate).

Nature Biotechnology 23, 1126 - 1136 (2005); doi:10.1038/nbt1142



A cartoon representation of different antibody formats used for *in vivo* imaging, together with their respective molecular weights (kDa) and serum half-life (phase).

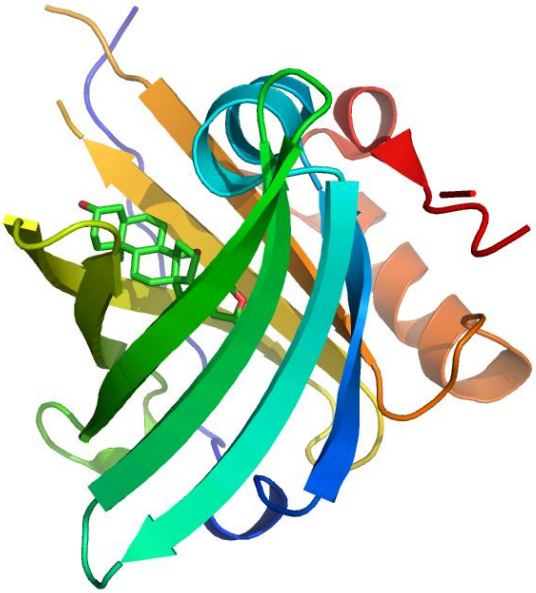
Nature Biotechnology 23, 1126 - 1136 (2005); doi:10.1038/nbt1142

Antibody mimetics

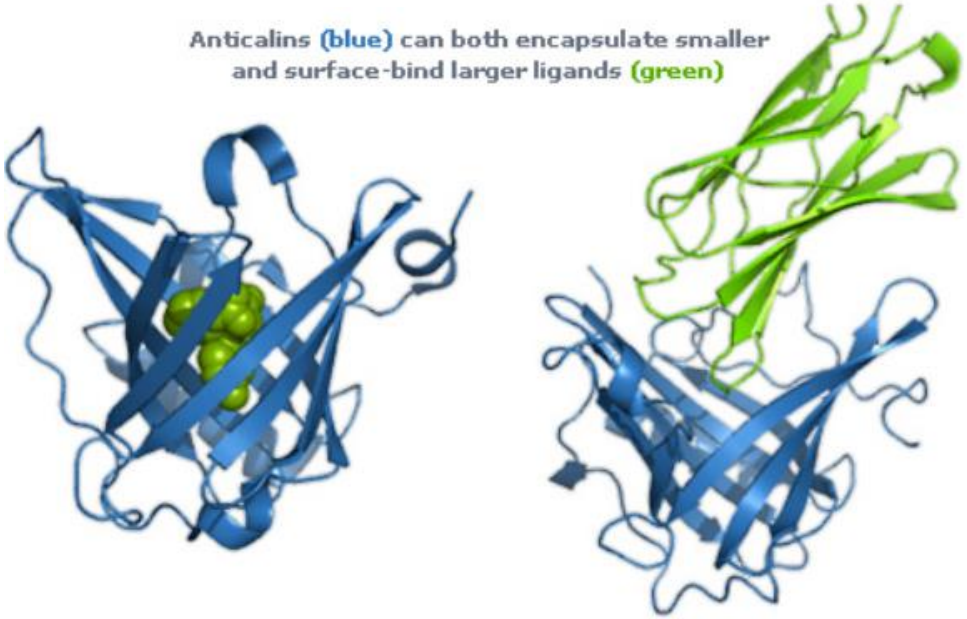
Antibody mimetic	Framework	Molar mass
Affibodies	Staphylococcus ProteinA, Z domain	6 kDa
Affilins	Gamma b crystallin Ubiquitin	20 kDa 10 kDa
Affimers	Cystatin	12-14 kDa
Affitins	Sac7d (from *Sulfobolus acidocaldarius*)	7 kDa
Alphabodies	Triple helix coiled coil	10 kDa
Anticalins	Human Lipocalins	20 kDa
Avimers	A domains of various membrane receptors	9–18 kDa
DARPins	Ankyrin repeat motif	12–19 kDa
Fynomers	Fyn, SH3 domain	7 kDa
Kunitz domain peptides	Kunitz domains of various protease inhibitors	6 kDa
Monobodies	Fibronectin, 10th Type III domain	10 kDa

http://en.wikipedia.org

Anticalins are artificial proteins that are able to bind to antigens, either to proteins or to small molecules. They are not structurally related to antibodies, which makes them a type of antibody mimetic. Instead, they are derived from human lipocalins which are a family of naturally binding proteins. Anticalins are being used in lieu of monoclonal antibodies, but are about eight times smaller with a size of about 180 amino acids and a mass of about 20 kDa.



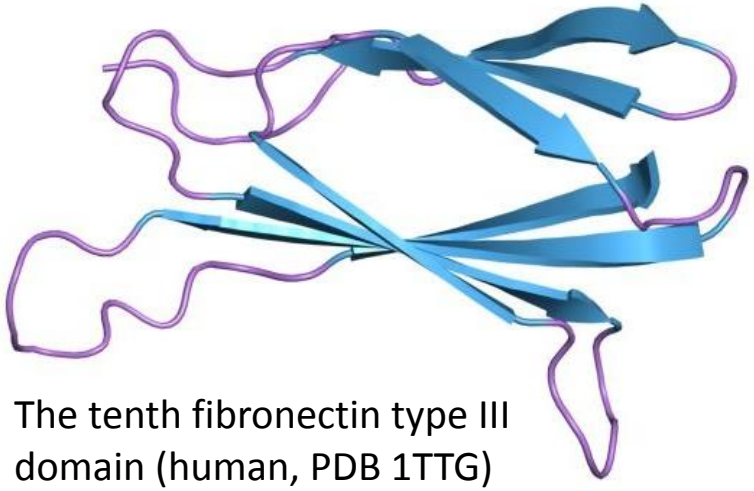
3D structure model of an anticalin (ribbon) in complex with digitoxigenin (sticks)



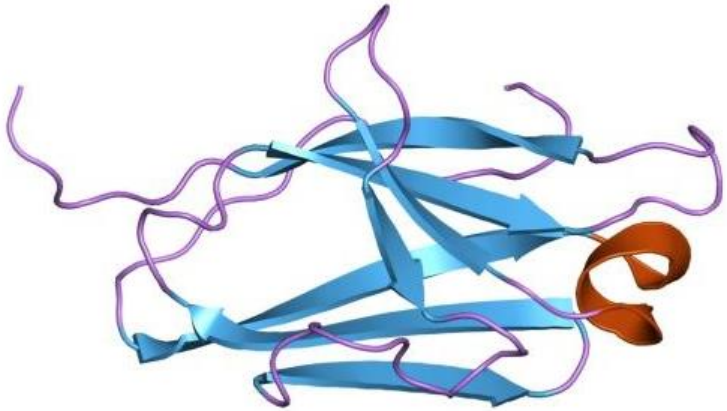
http://en.wikipedia.org

http://www.pieris.com/anticalin-technology/overview

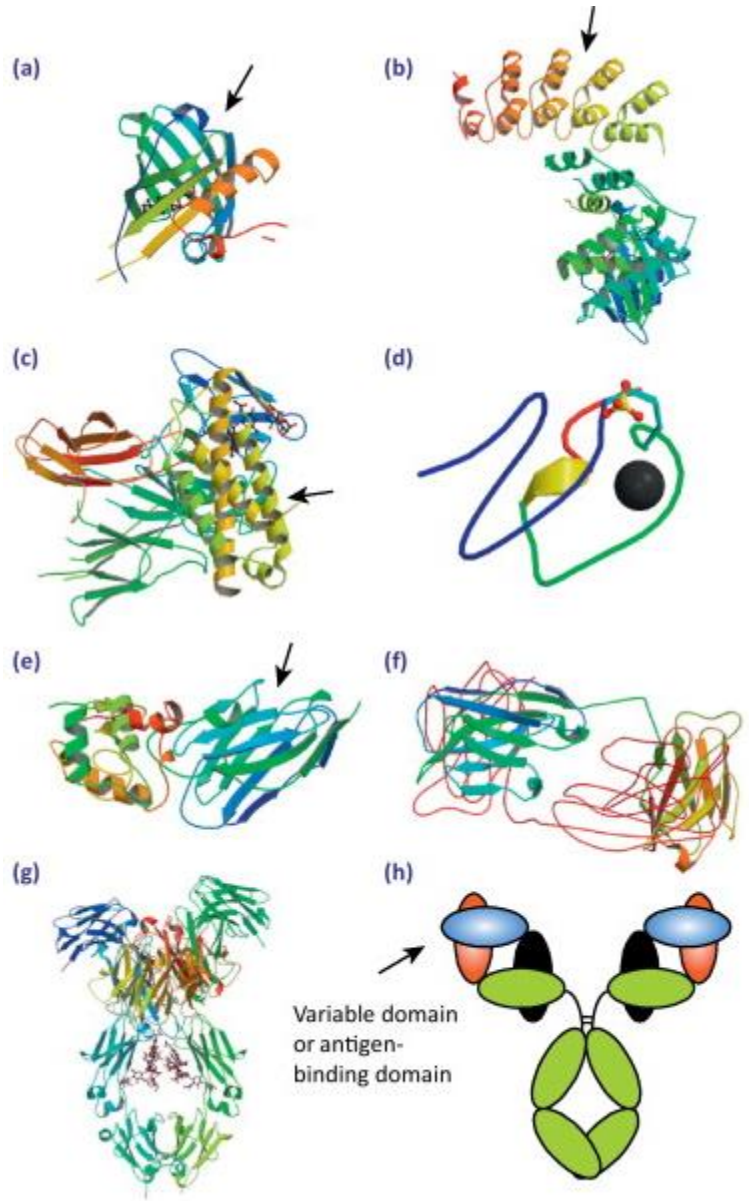
Monobodies consist of 94 amino acids and have a molecular mass of about 10 kDa, fifteen times smaller than an IgG type antibody and comparable to the size of a single variable domain of an antibody. They are based on the structure of human fibronectin, more specifically on its tenth extracellular type III domain. This domain has a structure similar to antibody variable domains, with seven beta sheets forming a barrel and three exposed loops on each side corresponding to the three complementarity determining regions. Monobodies lack binding sites for metal ions and the central disulfide bond.
Monobodies with specificity for different proteins can be tailored by modifying the loops BC (between the second and third beta sheets) and FG (between the sixth and seventh sheets).



The tenth fibronectin type III domain (human, PDB 1TTG)



Variable domain of an antibody's lambda light chain (human, PDB 2RHE)



3D representations of non-Ig-based and Ig-like protein scaffolds

Structures are extracted from the PDB protein database and are not drawn to scale. Arrows indicate the scaffold of interest if complexed. **(a)** Anticalin in complex with digitoxigenin (PDB ID: 1LNM); **(b)** Designed Ankyrin Repeat ProteIN (DARPin) in complex with aminoglycoside 3' phosphotransferase (PDB ID: 2BKK); **(c)** adnectin in complex with interleukin-23 (IL-23, PDB ID: 3QWR); **(d)** prototypical A-domain structure, as part of the avimer scaffold (PDB ID: 1AJJ); **(e)** VHH nanobody in complex with lysozyme (PDB ID: 1MEL); **(f)** bivalent diabody [Bispecific T cell Engager (BiTE)-like structure, PDB ID: 1LMK]; **(g)** whole human IgG1 anti-HIV antibody (PDB ID: 1HZH); **(h)** schematic drawing of a whole IgG1, indicating the antigen-binding domain or variable domain.

Trends in Biotechnology,Volume 30, Issue 11, November 2012, Pages 575–582; doi:10.1016/j.tibtech.2012.07.006

Affibodies

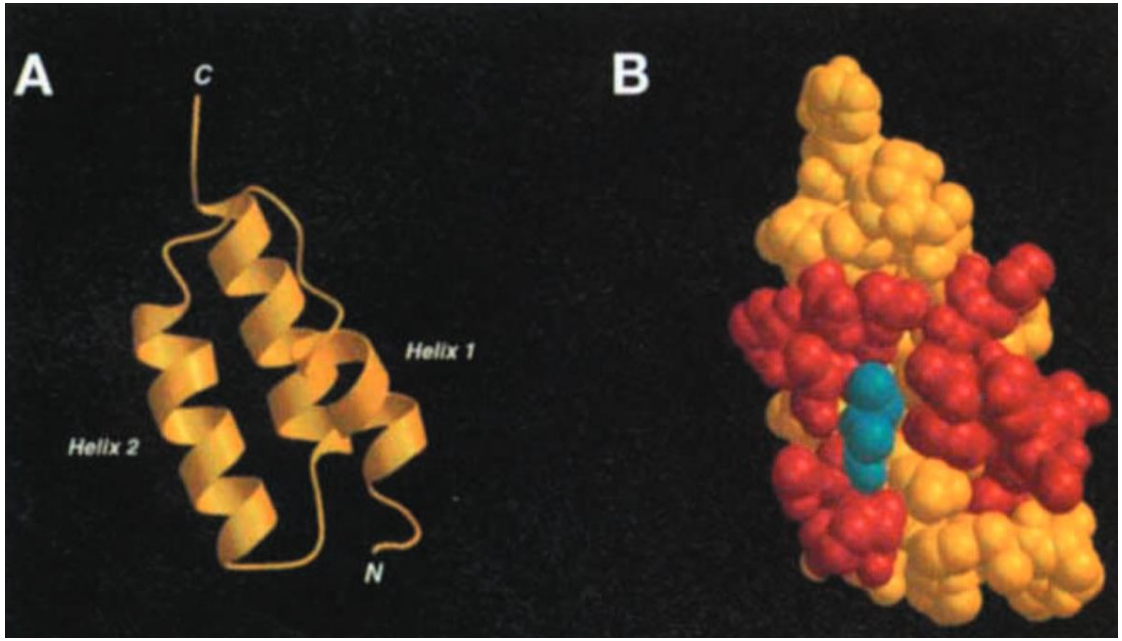


Figure 1. Graphic representation of the wild-type Z domain (residues 5-58) based on the NMR structure (A). Main-chain trace ribbon diagram showing the three helix-bundle structure. (B) Space-filling representation of the domain showing the positions of the 13 amino acids located in helices 1 and 2 subjected to the randomization (red). The position of Ile31 (blue) is also shown.

	1 Helix 1	20	Helix 2	40	Helix 3	
Z wt	VDNKFNK	EQQNAFYEILH	LPNLNE	EQRNAFIQSLKD	DPSQ	SANLLAEAKKLNDA	QAPK
α-Taq 4:1	-------	-KGE-VV--FR	-----G	R-VK---A--Y-	----	----------	----
α-Taq 4:8	-------	-LGW-TW--FN	-----G	F-AA---R--R-	--R-	----------	----
α-Insulin B6	-------	-AGV-WR--RT	-----D	E-LM---L--V-	----	----------	----
α-Apolipo 14:2	-------	-KAV-GD--VL	-----L	K-VC---S--V-	----	----------	----
α-Apolipo 14:5	-------	-NIV-NG--VP	-----L	S-SC---M--L-	----	----------	----
α-Apolipo 24:4	-------	DKNN-AV--MQ	-----A	G-IL---I--P-	----	----------	----

Figure 2. Amino acid sequence corresponding to the wild-type Z domain aligned to deduced amino acid sequences of different phagemid clones selected against various target proteins. Residues Q9, Q10, N11, F13, Y14, L17, H18, E24, R27, N28, Q32 and K35 included in the randomization are indicated. Helices in the wild type Z domain are boxed. Horizontal bars indicate amino-acid identities. Two of the clones showed additional substitutions outside the variegated regions (α-Taq 4.8 and α-Apolipo 24:4).

Nat Biotechnol. 1997 Aug;15(8):772-7

Antibody Engineering Variable Regions of heavy and light chains

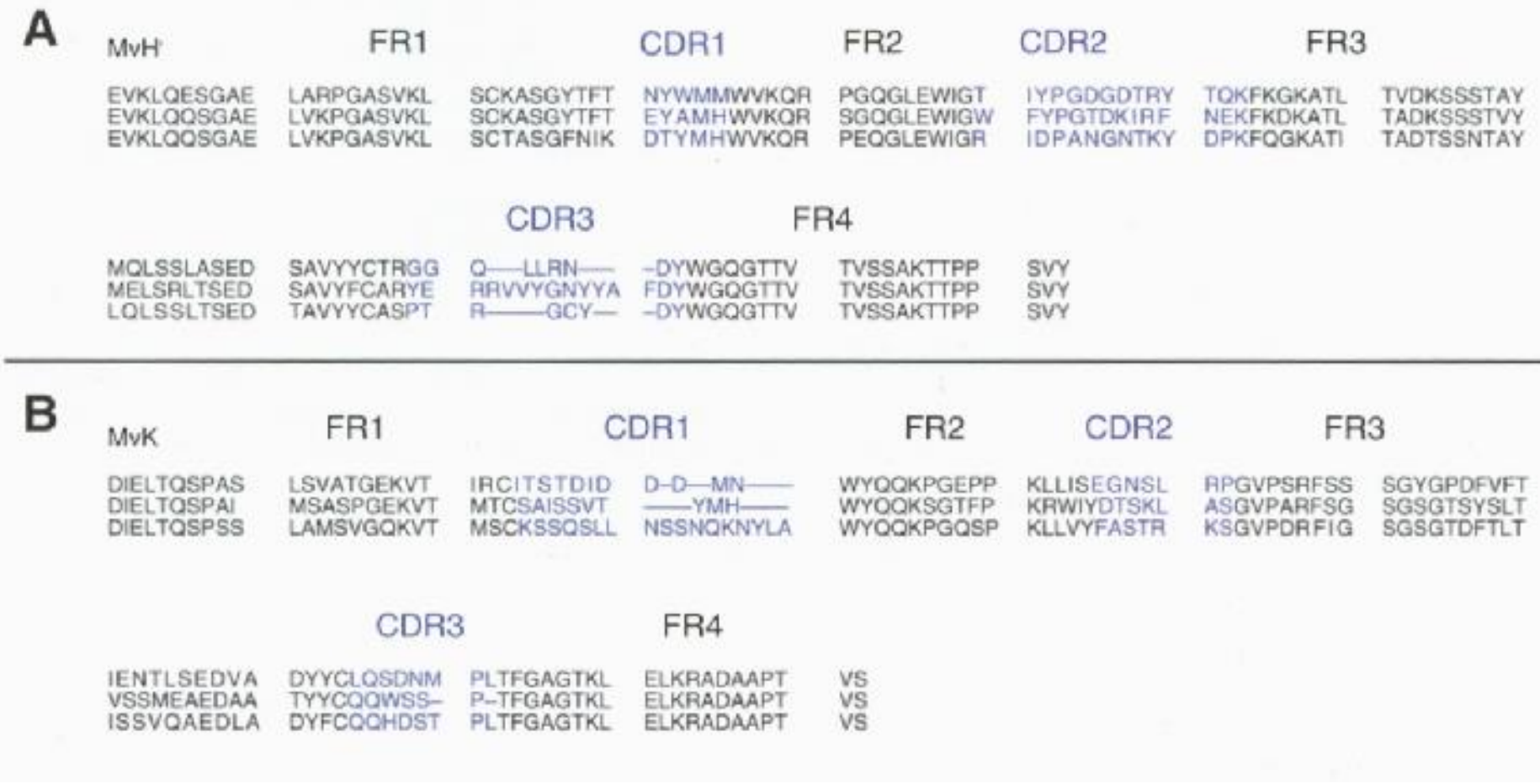


Fig. 8.2.2 Three detailed sequences of variable regions of heavy (A) and light (B) chains of three antibodies. Amino acid sequences are shown in one-letter-code. FR = framework regions (black letters), CDR = complementarity determining regions (blue letters). Although differences within the FR-domains are visible, differences in the CDR-domains are substantially more distinctive.

Antibody Engineering

Variability increase via degenerate oligos

Single-letter Code	Nucleotide/s	Explanation
A	A	**A**denine
C	C	**C**ytosine
G	G	**G**uanine
T	T	**T**hymine
I	I	**I**nosine
R	A or G	pu**R**ine
Y	C or T	p**Y**rimidine
M	A or C	a**M**ino
K	G or T	**K**eto
S	C or G	**S**trong interaction
W	A or T	**W**eak interaction
H	A or C or T	not G, **H** follows G in alphabet
B	C or G or T	not A, **B** follows A in alphabet
V	A or C or G	not T/U, **V** follows U in alphabet
D	A or G or T	not C, **D** follows C in alphabet
N	A or C or G or T	a**N**y

Table 1. IUPAC system for nucleotide nomenclature

A

MvH 1-25:

MvH-1	5'– GAG GTR MAG CTT CAG GAG TCA GGA C	–3'
MvH-2	5'– GAG GTS CAG CTK CAG CAG TCA GGA C	–3'
MvH-3	5'– CAG GTG CAG CTG AAG SAS TCA GG	–3'
MvH-4	5'– GAG GTG CAG CTT CAG GAG TCS GGA C	–3'
MvH-5	5'– GAR GTC CAG CTG CAA CAG TCY GGA C	–3'
MvH-6	5'– CAG GTC CAG CTK CAG CAA TCT GG	–3'
MvH-7	5'– CAG STB CAG CTG CAG CAG TCT GG	–3'
MvH-8	5'– CAG GTY CAG CTG CAG CAG TCT GGR C	–3'
MvH-9	5'– GAG GTY CAG CTY CAG CAG TCT GG	–3'
MvH-10	5'– GAG GTC CAR CTG CAA CAA TCT GGA CC	–3'
MvH-11	5'– CAG GTC CAC GTG AAG CAG TCT GGG	–3'
MvH-12	5'– GAG GTG AAS STG GTG GAA TCT G	–3'
MvH-13	5'– GAV GTG AAG YTG GTG GAG TCT G	–3'
MvH-14	5'– GAG GTG CAG SKG GTG GAG TCT GGG G	–3'
MvH-15	5'– GAK GTG CAM CTG GTG GAG TCT GGG	–3'
MvH-16	5'– GAG GTG AAG CTG ATG GAR TCT GG	–3'
MvH-17	5'– GAG GTG CAR CTT GTT GAG TCT GGT G	–3'
MvH-18	5'– GAR GTR AAG CTT CTC GAG TCT GGA	–3'
MvH-19	5'– GAA GTG AAR STT GAG GAG TCT GG	–3'
MvH-20	5'– GAA GTG ATG CTG GTG GAG TCT GGG	–3'
MvH-21	5'– CAG GTT ACT CTR AAA GWGTST GGC C	–3'
MvH-22	5'– CAG GTC CAA CTV CAG CAR CCT GG	–3'
MvH-23	5'– CAG GTY CAR CTG CAG CAG TCT G	–3'
MvH-24	5'– GAT GTG AAC TTG GAA GTG TCT GG	–3'
MvH-25	5'– GAG GTG AAG GTC ATC GAG TCT GG	–3'

B

MvK 1-25:

MvK-1	5'– GAC ATT GTT CTC ACC CAG TCT CC	–3'
MvK-2	5'– GAC ATT GTG CTS ACC CAG TCT CC	–3'
MvK-3	5'– GAC ATT GTG ATG ACT CAG TCT CC	–3'
MvK-4	5'– GAC ATT GTG CTM ACT CAG TCT CC	–3'
MvK-5	5'– GAC ATT GTG YTR ACA CAG TCT CC	–3'
MvK-6	5'– GAC ATT GTR ATG ACA CAG TCT CC	–3'
MvK-7	5'– GAC ATT MAG ATR ACC CAG TCT CC	–3'
MvK-8	5'– GAC ATT CAG ATG AMC CAG TCT CC	–3'
MvK-9	5'– GAC ATT CAG ATG ACD CAG TCT CC	–3'
MvK-10	5'– GAC ATT CAG ATG ACA CAG ACT AC	–3'
MvK-11	5'– GAC ATT CAG ATG ATT CAG TCT CC	–3'
MvK-12	5'– GAC ATT GTT CTC AWC CAG TCT CC	–3'
MvK-13	5'– GAC ATT GTT CTC TCC CAG TCT CC	–3'
MvK-14	5'– GAC ATT GWG CTS ACC CAA TCT CC	–3'
MvK-15	5'– GAC ATT STG ATG ACC CAR TCT C	–3'
MvK-16	5'– GAC ATT KTG ATG ACC CAR ACT CC	–3'
MvK-17	5'– GAC ATT GTG ATG ACT CAG GCT AC	–3'
MvK-18	5'– GAC ATT GTG ATG ACB CAG GCT GC	–3'
MvK-19	5'– GAC ATT GTG ATA ACY CAG GAT G	–3'
MvK-20	5'– GAC ATT GTG ATG ACC CAG TTT GC	–3'
MvK-21	5'– GAC ATT GTG ATG ACA CAA CCT GC	–3'
MvK-22	5'– GAC ATT GTG ATG ACC CAG ATT CC	–3'
MvK-23	5'– GAC ATT TTG CTG ACT CAG TCT CC	–3'
MvK-24	5'– GAC ATT GTA ATG ACC CAA TCT CC	–3'
MvK-25	5'– GAC ATT GTG ATG ACC CAC ACT CC	–3'

Fig. 8.2.3 Collection of diverse degenerate primers for the amplification of cDNA sequences of heavy (A) and Light (B) antibody chains. Primers bind at the N-terminal region and recognize the FR domain. These primers are named „backward-primer“. Degenerate positions are marked with orange letters.

Antibody Engineering

B

Linker sequence	length
GGGGS GGGGS GGGGS	15
$(GGGGS)_n$, n = 2 – 5	10-25
EGKSS GSGSE SKST	14
GSTSG SGKSS EKG	14
KESGS VSSEQ LAQRF SLD	18
LTVSS ALTTP PSVYP L	16
SSPSV TLFPP SSNG	14
IKRAG QGSSV	10
SSADD AKKDA AKKDD AKKDD AKKDDG	25
PGGNR GTTTT RRPAT TTGSS PGPTQ SH	27
GSASA PKLEE GEFSE ARE	18

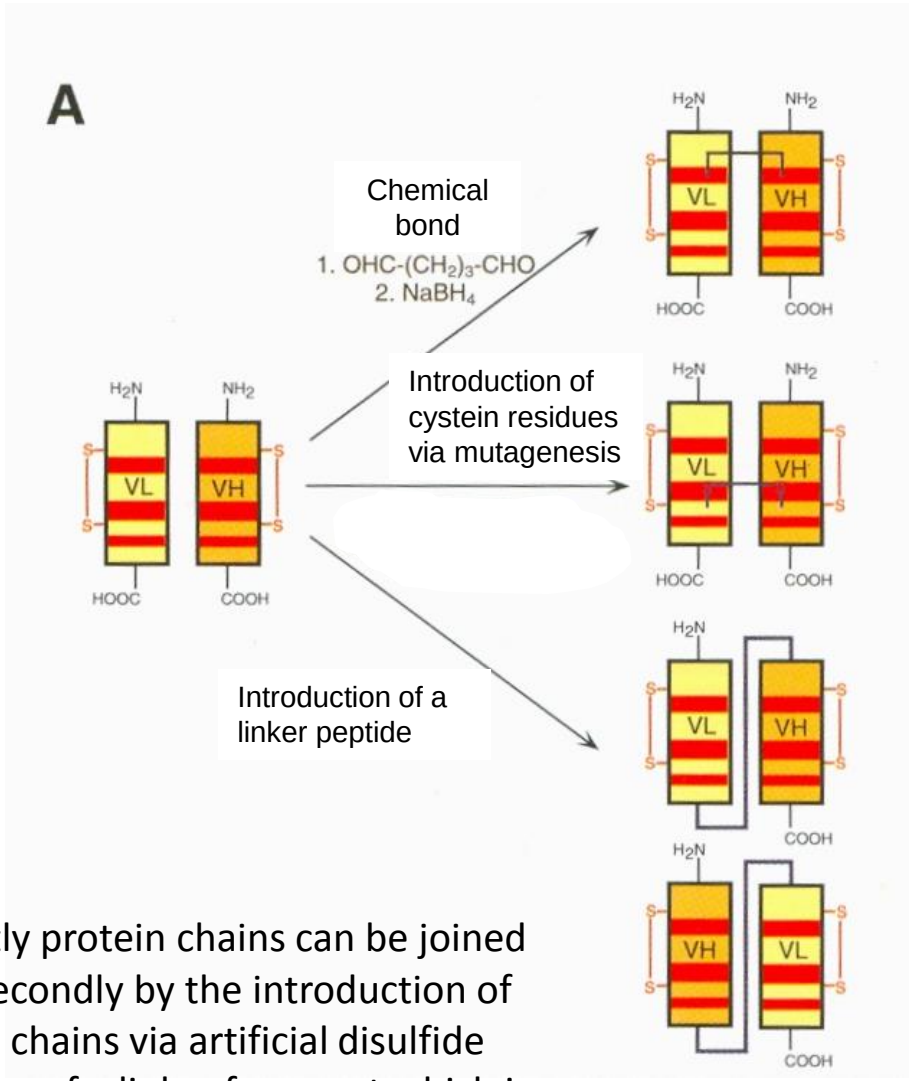


Fig. 8.2.5 Linkage of a VL and a VH domain

A Both domains can be linked in different ways: firstly protein chains can be joined chemically with the help of a bifunctional reagent, secondly by the introduction of cysteine codons on gene level and connection of the chains via artificial disulfide bonds. The third and most common method is the use of a linker fragment which is introduced by PCR strategy.

B Possible linker sequences for the chain connection

Antibody Engineering

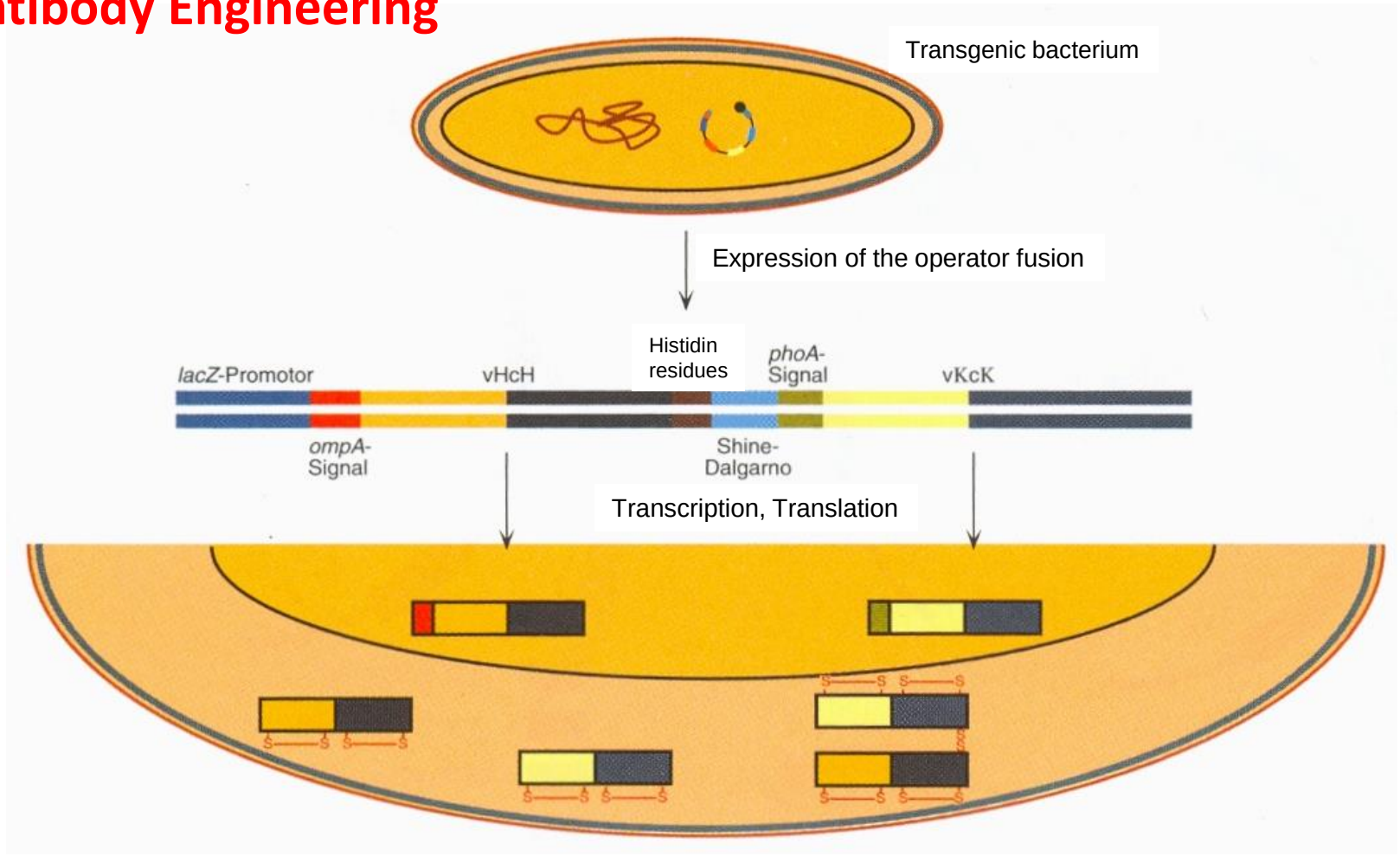
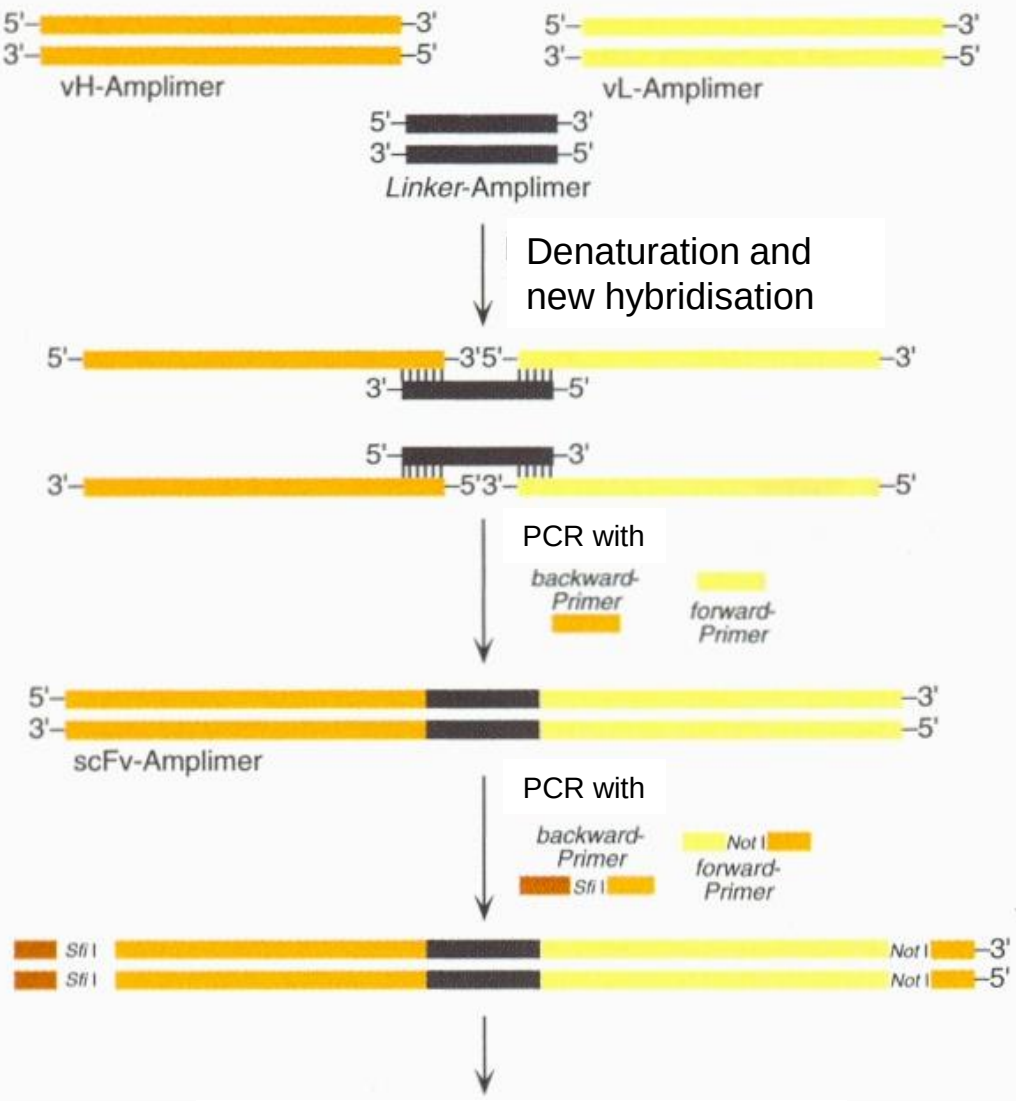


Fig. 8.2.12 Biosynthesis of a recombinant Fab-fragment.
The pASK84 expression cassette is an “operator fusion”. The whole cassette is expressed under the control of the lacZ promoter. Two protein chains are made simultaneously based on the mRNA: one for the heavy and one for the light chain. The formed proteins are directed to the periplasm of the bacterial cell with the help of the signal peptides OpmA and PhoA. Signal peptides are hydrolysed while passing through the cell membrane. The reducing environment in the periplasm is sufficient for disulfide bond formation between heavy and light chains.

Antibody Engineering

Fig. 8.2.6 **Splicing by overlap extension**. For the connection of DNA fragments for heavy and light chains, a linker is synthesized not only bearing the information for $(Gly_4Ser)_3$ but also being complementary to the 24 3´ terminal nucleotides of vH amplimer and the 24 5´-terminal nucleotides of vL amplimer, respectively. All three amplimers are put into the same PCR tube. In the next step the backward primer (heavy chain amplification) and forward primer (light chain amplification) are added. This experimental setup facilitates the combination of all three fragments and the amplification thereof. Another PCR step provides a terminal extension of the amplimers adding recognition sequences for the restriction endonucleases *Not*I and *Sfi*I.

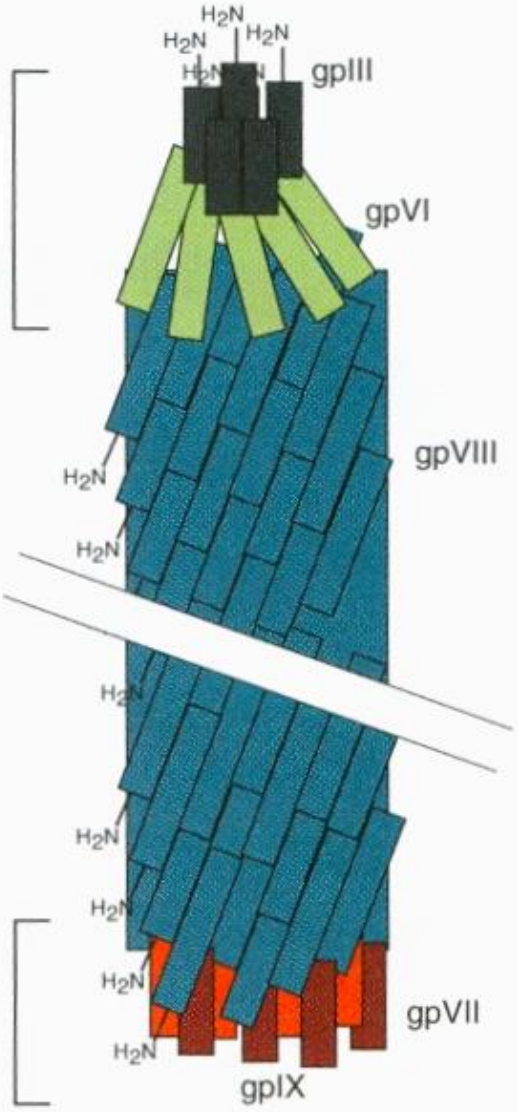


Antibody Engineering

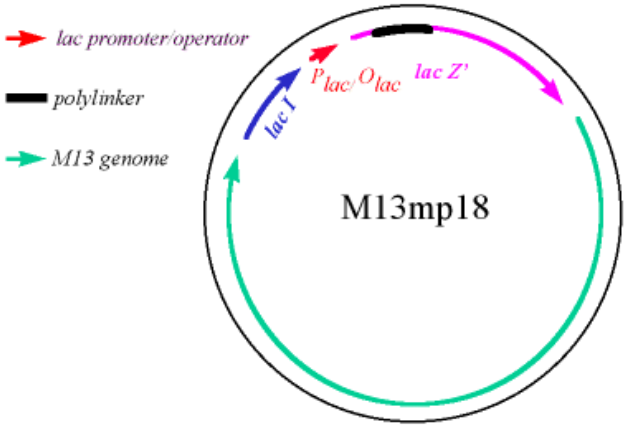
Phage display

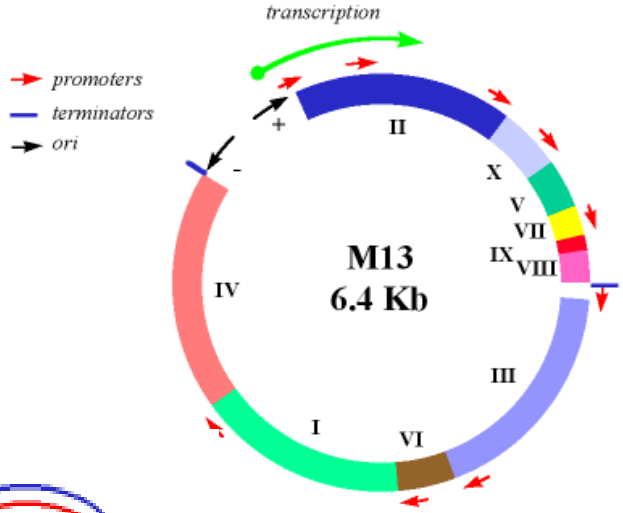
Fig. 8.2.7 **Composition of M13 phages**.
The capsid of a M13 phage is basically composed of five different proteins coded by the genes gpIII, gpVI, gpVII, gpVIII and gpIX. N-termini of two proteins protrude from the capsid the way that a fusion protein could be easily added without disrupting the composition of the phage capsid too much. This has already been realized for both proteins. Copy numbers of these two proteins differ considerably – a phage capsid has ca. 2500 gpVIII gene products, but only three to five gpIII.

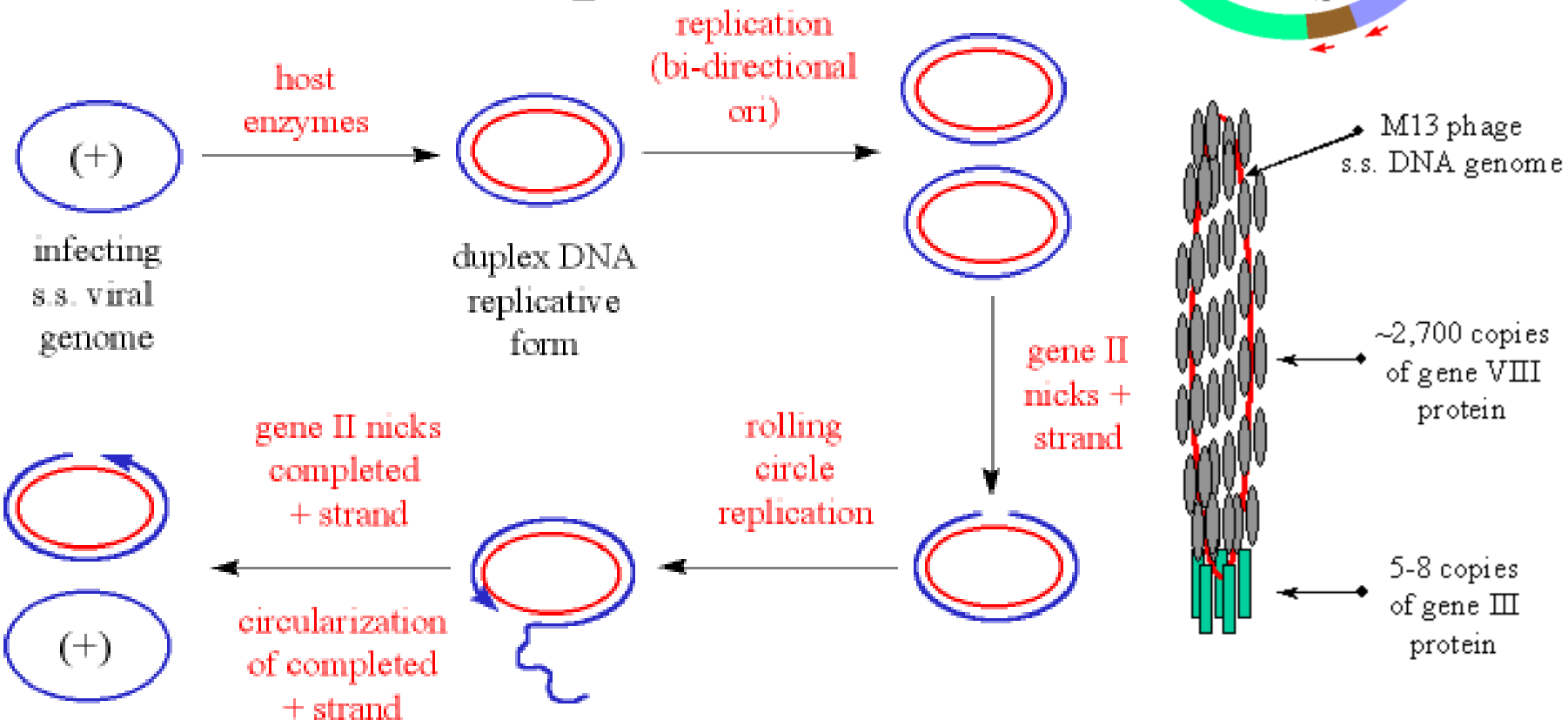
Abb. 8.2.8 Recombinant phage with a modified gpIII protein on the top



Antibody Engineering







Antibody Engineering

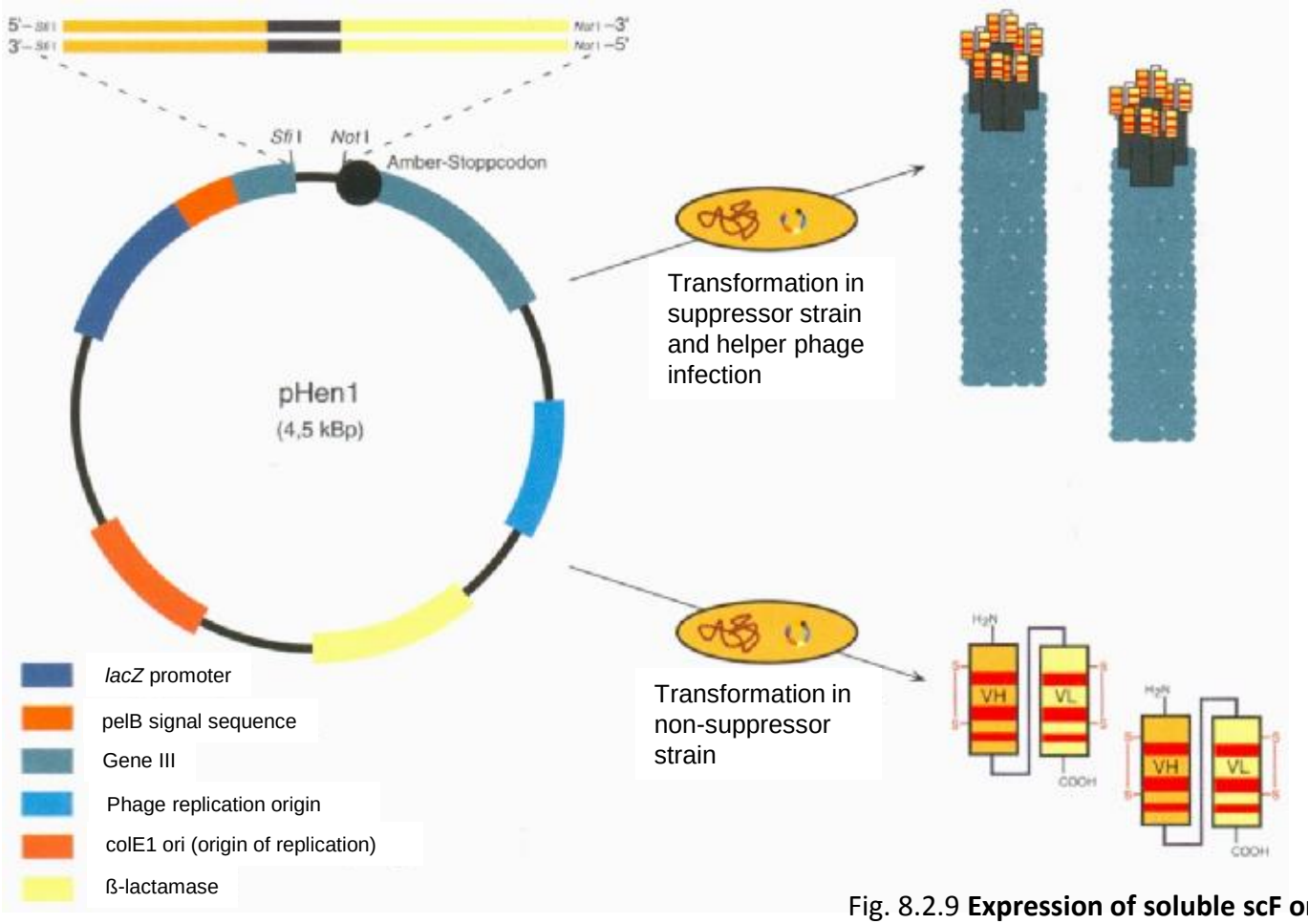


Fig. 8.2.9 **Expression of soluble scF or recombinant phages**

For expression of soluble scF-fragments or recombinant phages plasmid pHen1 is used. The plasmid consists of a colE1 replication origin, a phage replication origin, a ß-lactamase gene and an expression cassette under the control of a bacterial ß-galactosidase promoter (lacZ promoter) followed by a signal sequence for the transport of soluble antibody fragments to bacterial periplasm where disulfide bonds are formed. Subsequently gpIII gene sequence is located which is interrupted by a short sequence harboring *Sfi*I and *Not*I restriction sites for in frame cloning of scF fragment coding sequence. The black dot stands for an amber stop codon termination of protein biosynthesis if the plasmid is transformed into a wild type *E.coli* strain. Soluble scF fragments are produced. However, if the plasmid is transformed into an E.coli strain, which harbors a tRNA for amber stop codon reading additionally to the general tRNPt setup, the rest of gpIII is translated. If those cells are infected with “helper phages” providing the residuary proteins for a correct phage assembly, modified gpIII gene products are incorporated into the phages resulting in recombinant phages with scfV fragment at the top.

Antibody Engineering

Fig. 8.2.9 **Expression of soluble scF or recombinant phages**
For expression of soluble scF-fragments or recombinant phages plasmid pHen1 is used. The plasmid consists of a colE1 replication origin, a phage replication origin, a ß-lactamase gene and an expression cassette under the control of a bacterial ß-galactosidase promoter (*lacZ* promoter) followed by a signal sequence for the transport of soluble antibody fragments to bacterial periplasm where disulfide bonds are formed. Subsequently gpIII gene sequence is located which is interrupted by a short sequence harboring *Sfi*I and *Not*I restriction sites for in frame cloning of scF fragment coding sequence. The black dot stands for an amber stop codon termination of protein biosynthesis if the plasmid is transformed into a wild type *E.coli* strain. Soluble scF fragments are produced. However, if the plasmid is transformed into an E.coli strain, which harbors a tRNA for amber stop codon reading additionally to the general tRNPt setup, the rest of gpIII is translated. If those cells are infected with “helper phages” providing the residuary proteins for a correct phage assembly, modified gpIII gene products are incorporated into the phages resulting in recombinant phages with scfV fragment at the top.

Antibody Engineering

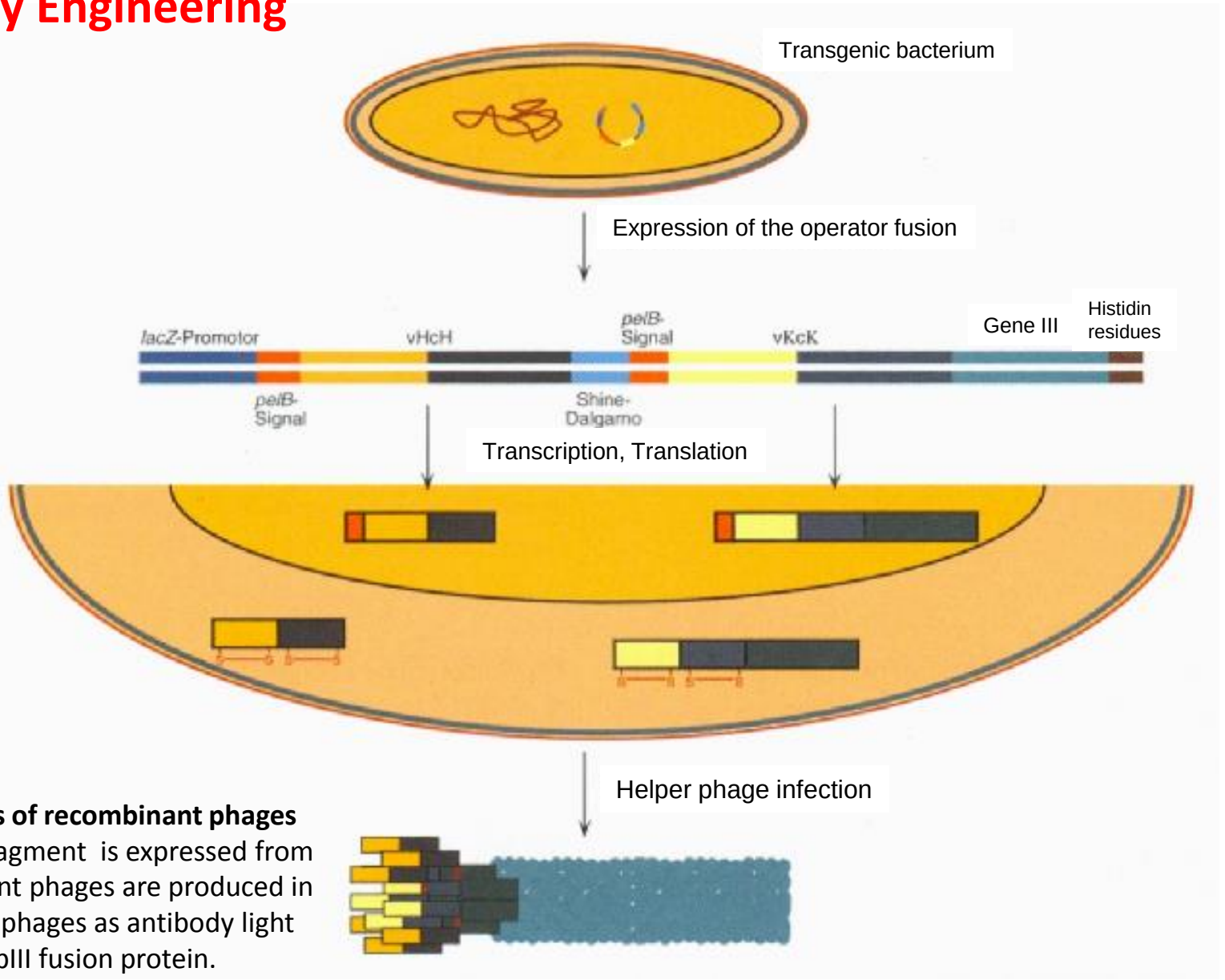


Fig. 8.2.13 **Biosynthesis of recombinant phages**
If a recombinant Fab fragment is expressed from pFAB5c.His, recombinant phages are produced in the presence of helper phages as antibody light chain is translated as gpIII fusion protein.

Antibody Engineering

Screening by panning

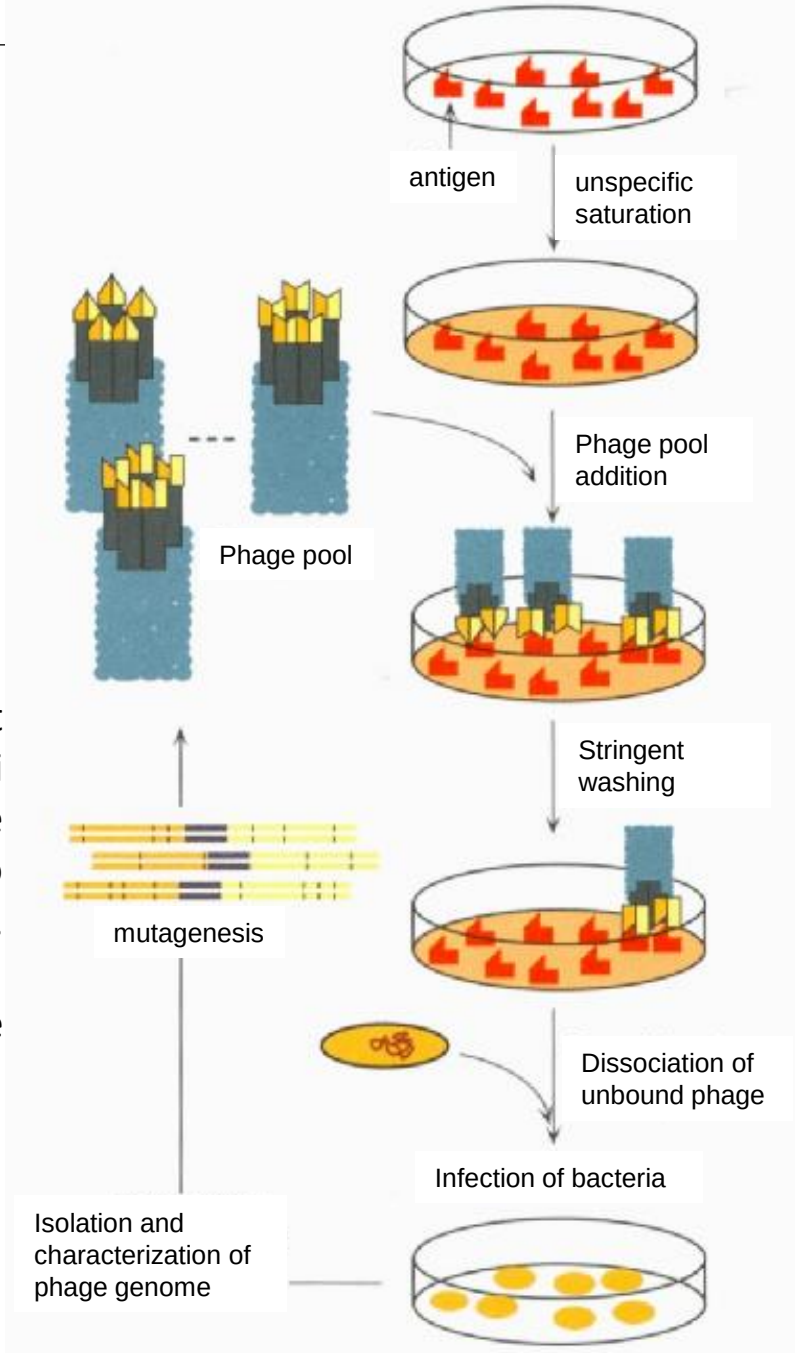
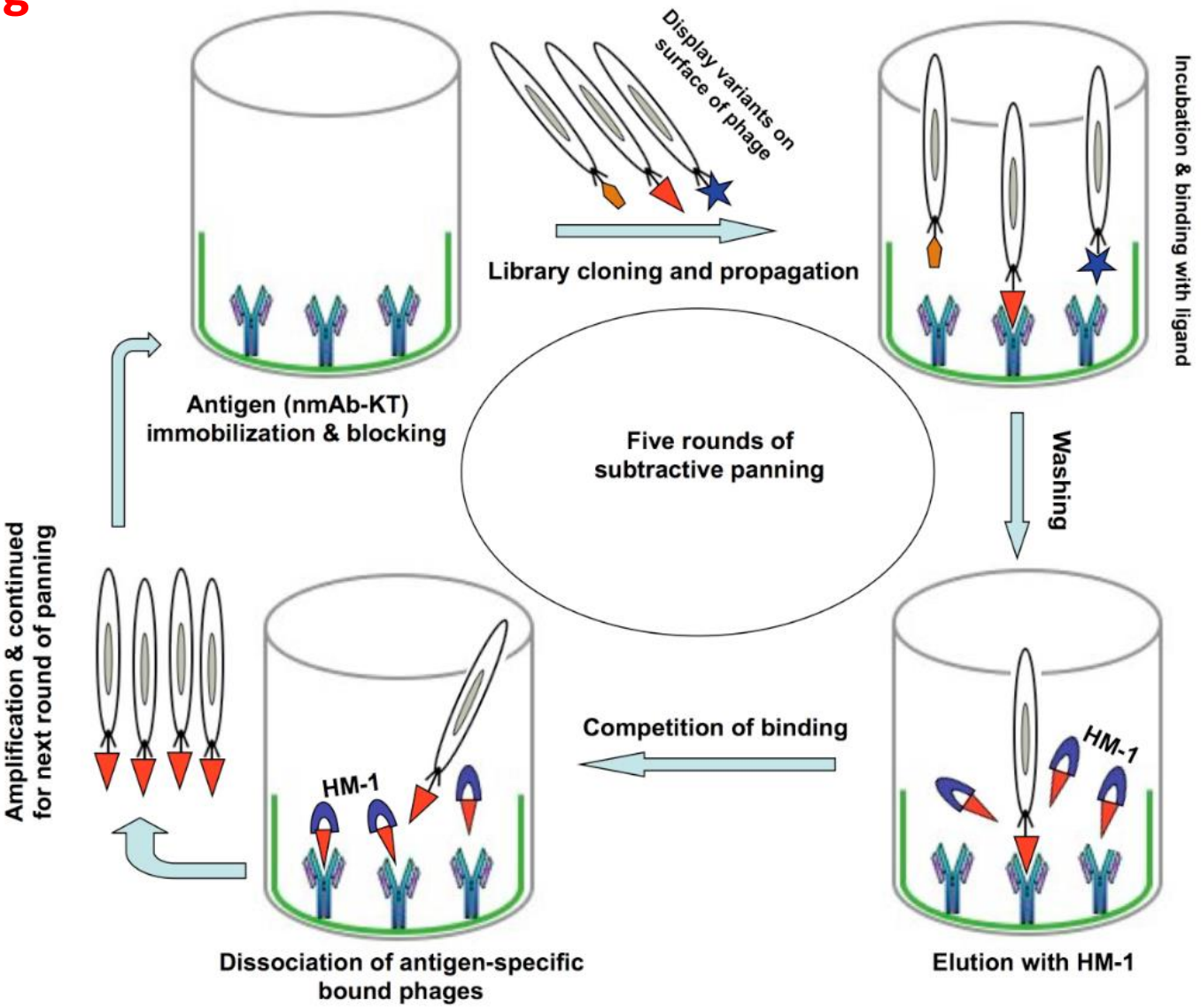


Fig. 8.2.16 **„phage panning“ for antibody fragment selection**
For the identification and isolation of the antibody fragment that reacts best with the antigen, the antigen is fixed in a petri dish. After saturation of free valences with inert protein the phages are added. Stringent washing steps are performed to enrich the phages with highest affinity to the fixed antigen. E.coli cells, which enable the phages to replicate, are added for phage amplification. After several selection rounds, phage DNA is isolated and characterized.

Phage panning

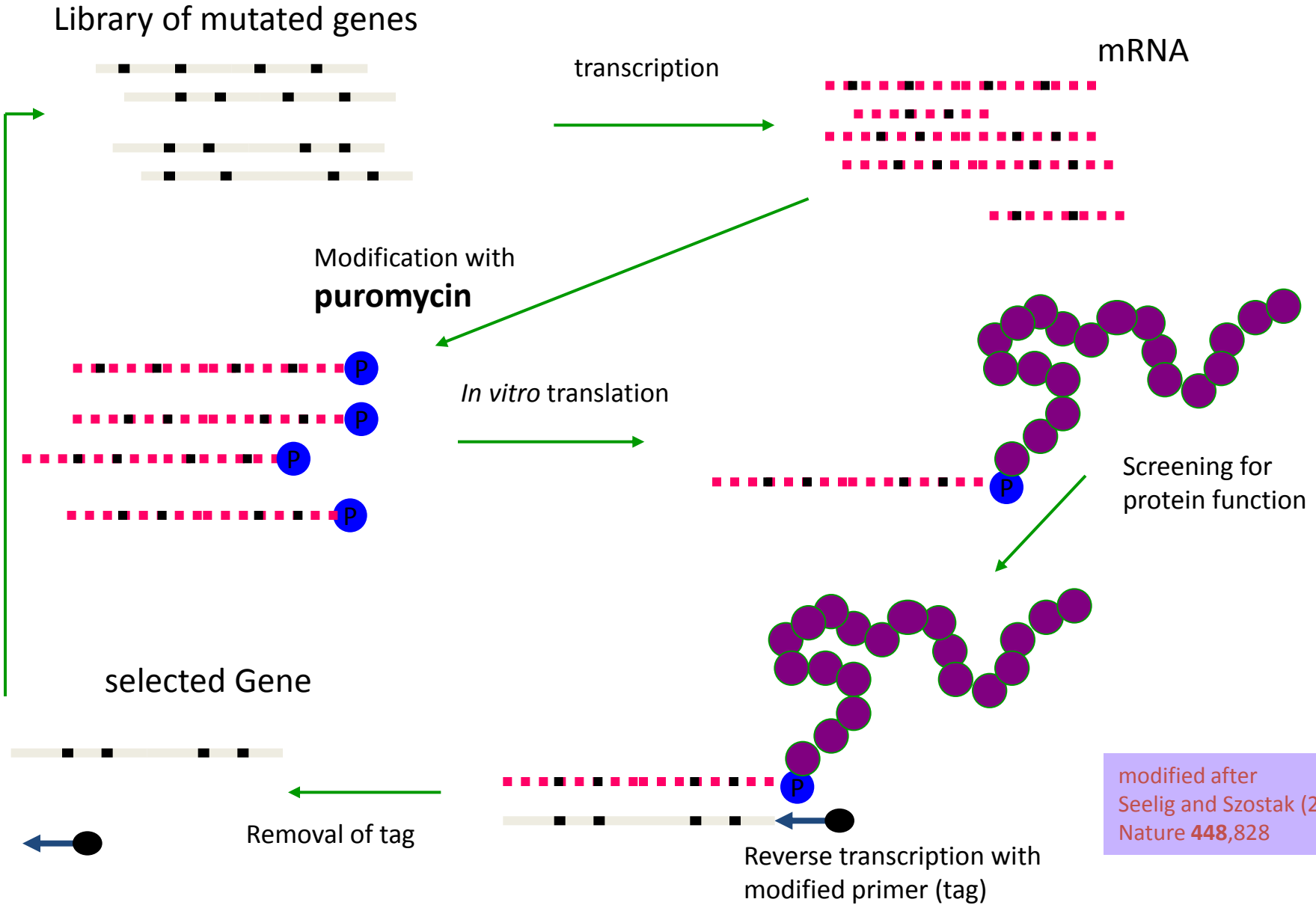
Schematic representation of the phage-display selection procedure: modified subtractive panning. Phage particles displaying the members of the library were produced. The specific binder (nmAb-KT) was allowed to bind with the target, and other variants were removed by washing. Molecular variants with specificity for the target were retrieved after multiple cycles of selection and were characterized in detail. Bound phages were eluted with HM-1 containing phosphate buffer (pH 7.0). HM-1 had high binding affinity to the immobilized nmAb-KT. This competition of binding favored quick dissociation of scFv-containing phages from the bound nmAb-KT and consequently increased the elution stringency of the infected phages.



Kabir et al. BMC Biotechnology 2009 9:99 doi:10.1186/1472-6750-9-99

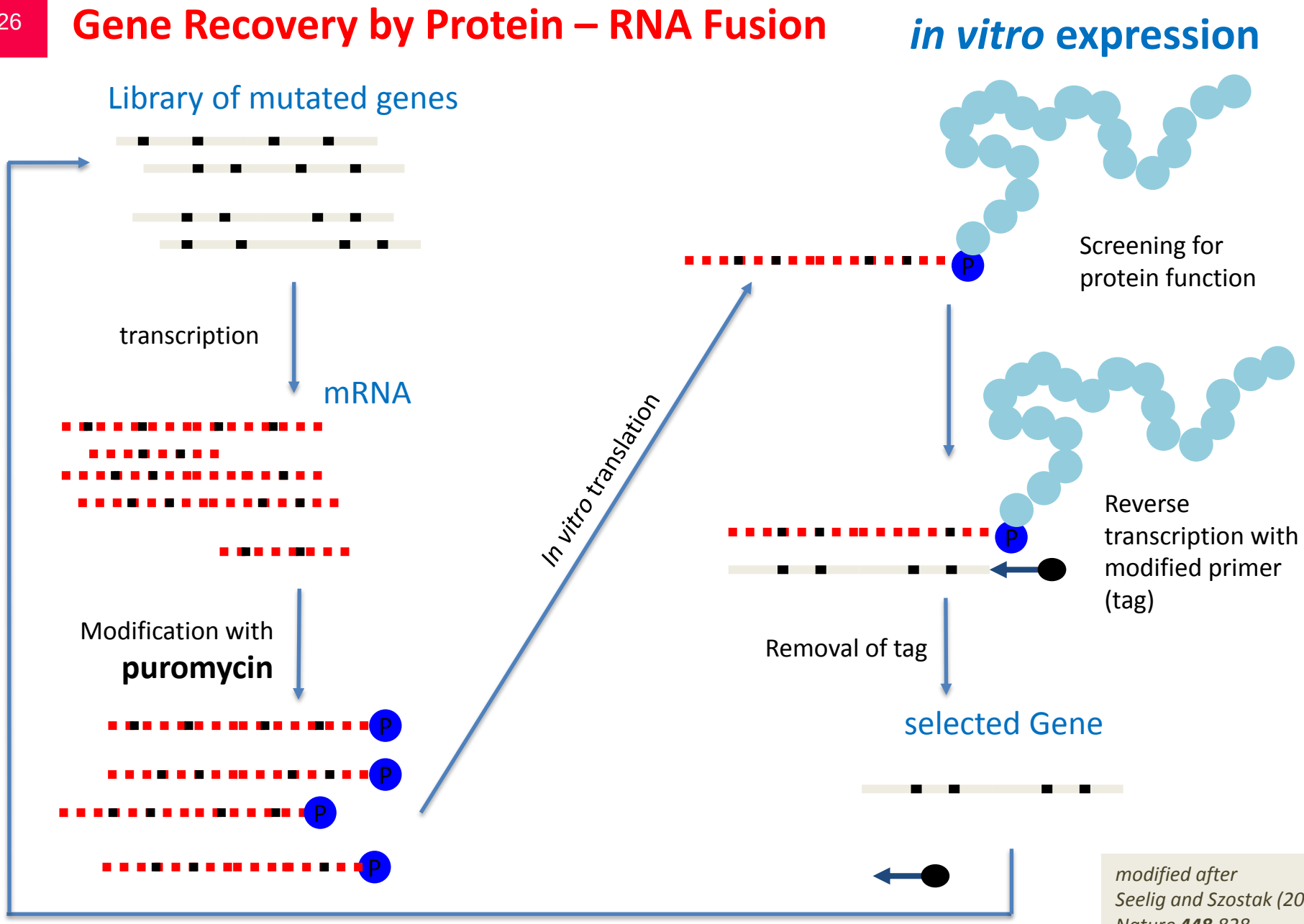
Gene Recovery by Protein – RNA Fusion

in vitro expression



modified after
Seelig and Szostak (2007)
Nature **448**,828

Gene Recovery by Protein – RNA Fusion



modified after
Seelig and Szostak (2007)
*Nature **448**,828*