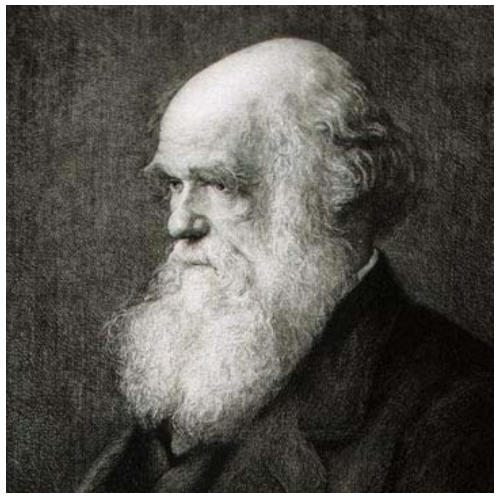
Directed Evolution

Natural evolution: Over time, random genetic mutations occur within an organism's genetic code, of which beneficial mutations are preserved because they are beneficial to survival

Directed evolution mimics natural evolution in the laboratory (in vitro), but operates on a molecular level (i.e. no new organisms are created) and focuses on specific molecular properties



Similarities between natural and directed evolution

- diversification: offsprings are different from the parents
- selection: survival of the fittest
- amplification: procreation

Benefits of directed evolution

Does not require forethought of what type of mutations are beneficial

Lack of detailed knowledge is compensated for by use of a powerful selection/screening method based on the concept of the **survival of the fittest**

Custom selection schemes can be designed to fit the needs of an engineer

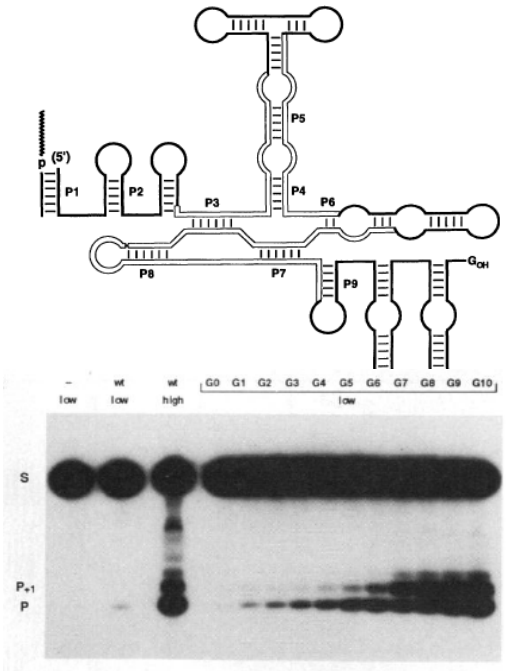
Various implementations of directed evolution

- platform ➔ choice of system
- diversification ➔ library construction
- selection ➔ assay development
- amplification ➔ depend on the previous three

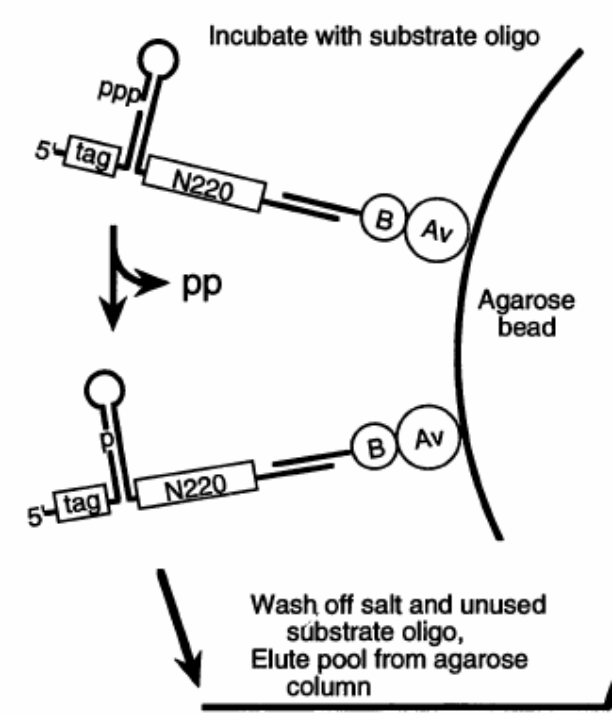
What to engineer in the lab

Engineering goals may vary and can be applied to many different molecules

- evolution of catalytic RNA/DNA
- evolution of RNA/DNA aptamer



after 10 generations of selection, a 100 fold increase in the ability to cut DNA



10^6 fold increase in the ability to ligate RNA backbone

Beaudry and Joyce, Science 257, 635 (1992)

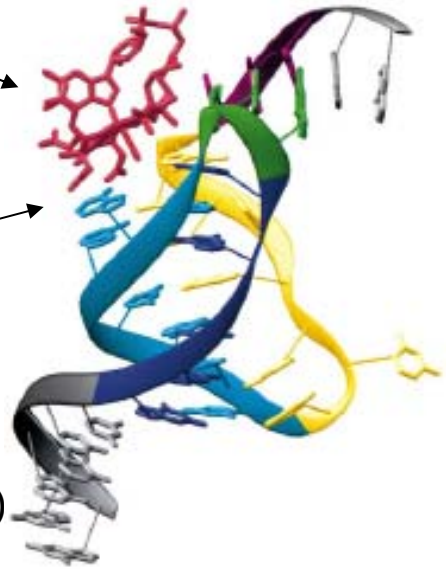
Bartel and Szostak, Science 261, 1411 (1993)

Aptamer

- aptamer = aptus (“to fit”) + mer
- Short fragment of RNA/DNA/peptide that binds a target molecule with high affinity

vitamin B12

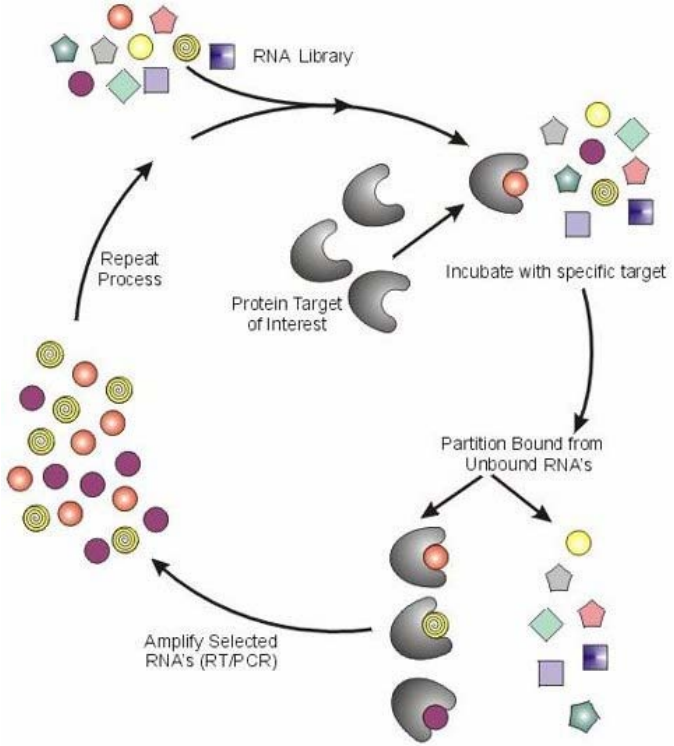
solvent accessible docking surface



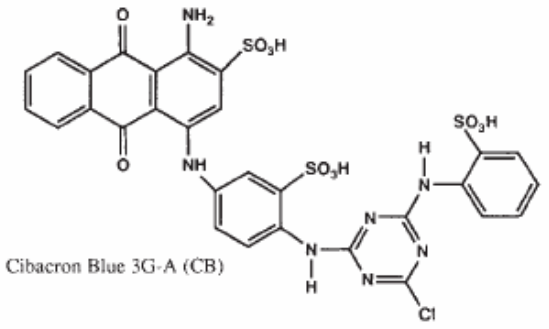
Kd = 90 nM

Sussman et al NSMB 7, 53 (2000)

SELEX



- 10^2 – 10^5 sequence out of 10^13 randomly generate RNA binds dye molecules



Ellington and Szostak, Nat 346, 818 (1990)

Directed evolution and protein engineering

Protein molecules with altered structural and functional properties

- increase thermal stability
- introduce a new functionality—e.g. engineer an enzyme
- change the topology or quaternary structure
- alter the details of existing properties—e.g. fluorescence

Unlike RNA/DNA, proteins cannot be amplified or propagated directly

Decoding the amino acid sequence requires sequencing the original DNA

Physical linkage between nucleic acid and protein is essential during protein engineering via directed evolution

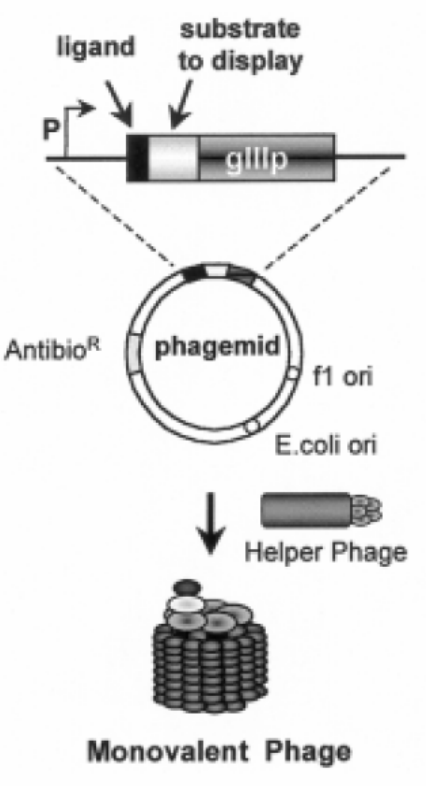
- genotype and phenotype need to be coupled
- key to high throughput screening
- protein needs to be “displayed” in order to be assayed (i.e. tested)
  - » different methods of coupling result in different **display** platforms
- cell-based functional assay

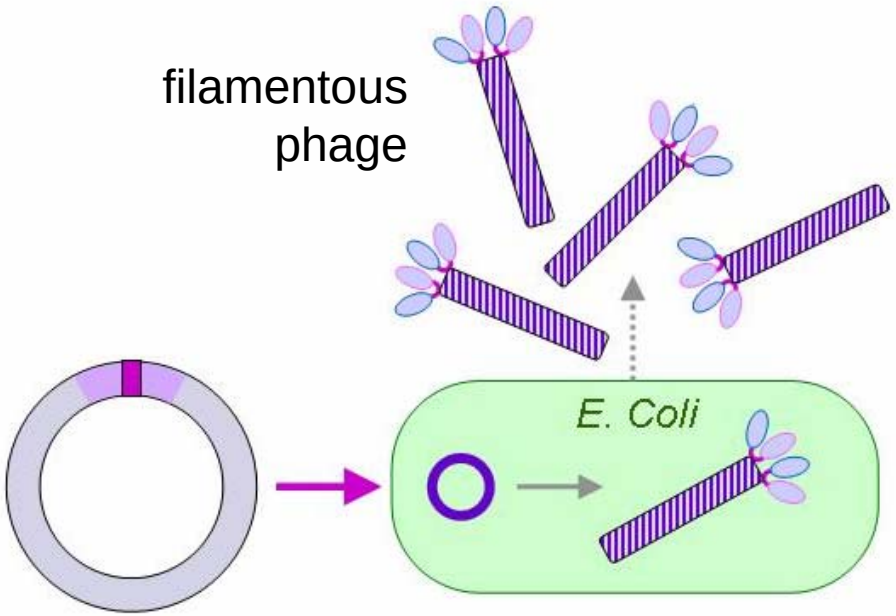
Display technologies

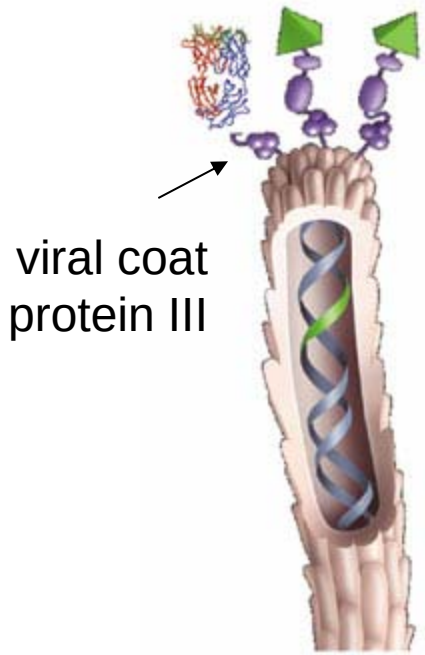
- Phage display
- Bacterial display
- Yeast display
- mRNA display
- *Ribosomal display*

Phage display

- filamentous phage is a virus that infects bacteria
- through recombinant technology, a protein of interest can be introduced into the viral genome (phagemid)
- virus expresses the foreign protein on the surface
- once the protein is displayed, we can test for its activity, e.g. binding affinity, catalysis, etc



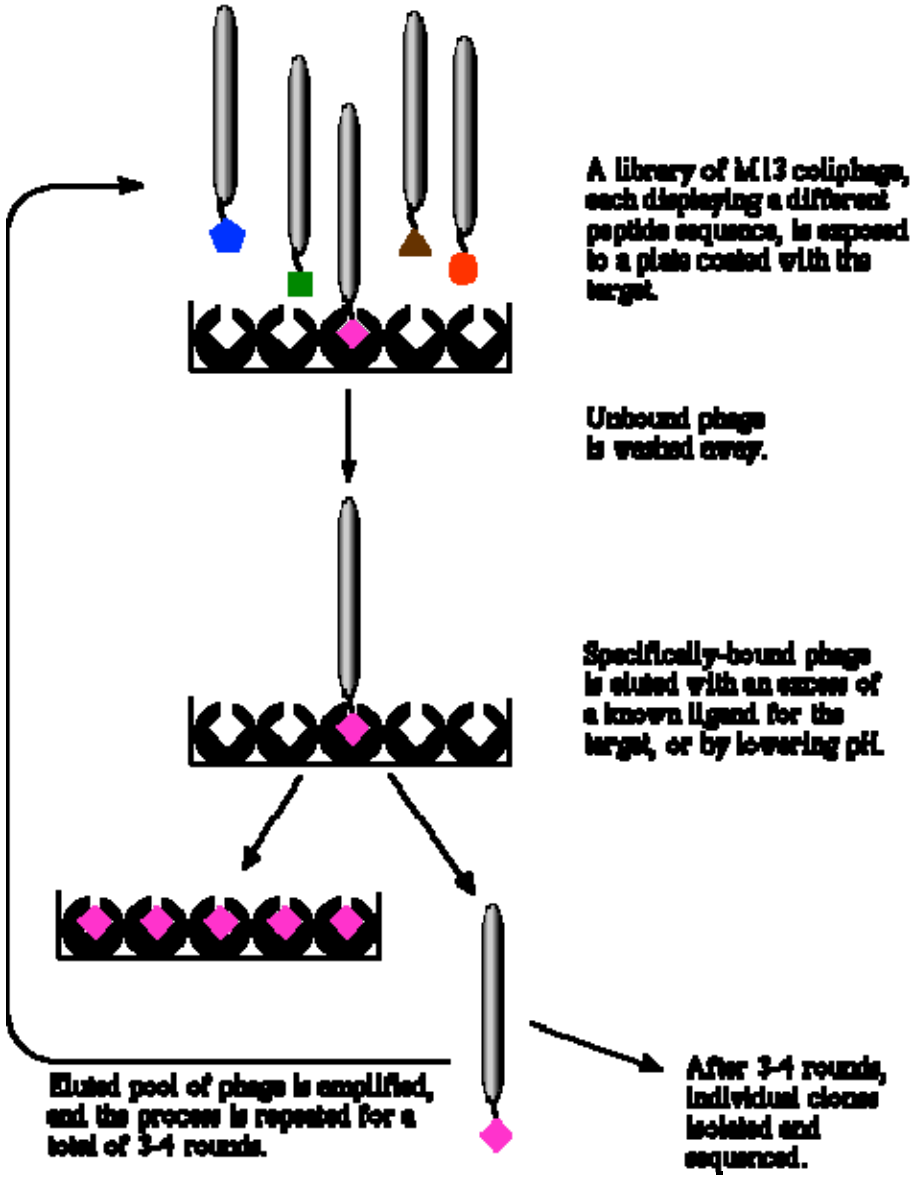


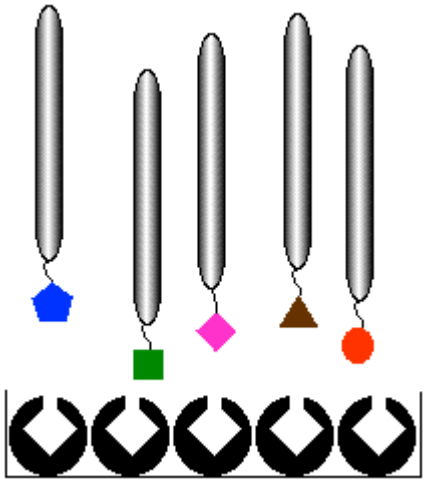


Depertes, Biol Chem 383, 1107 (2002)

most common fitness criterion is affinity for a target receptor

“biopanning” isolates phage particles with optimized binding affinity





Peptide and protein libraries

Peptides up to 40 residues

- not all clones are represented
- 15 residues ➔ 3 x 10^{19} possibilities
- typically < 10^{9} clones in a library

Proteins

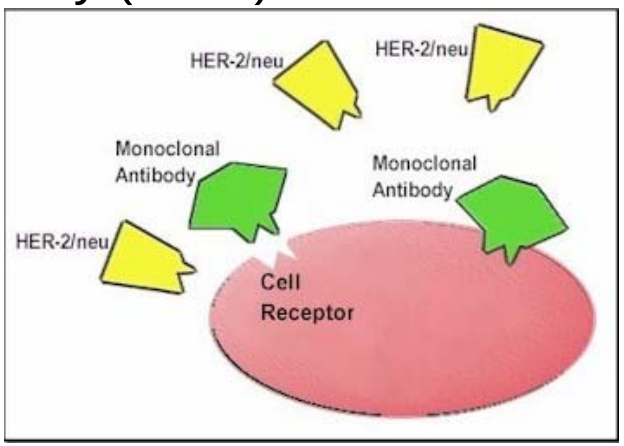
- enzymes—e.g. phosphatase, proteases, beta-lactamase
- hormones—e.g. human growth hormone, angiotensin
- inhibitors—e.g. BPTI, cystatin
- toxins—e.g. ricin, ribotoxin
- receptors—e.g. IgG binding domain of protein A and G, T cell receptor
- ligands—e.g. SH3
- DNA binding protein—e.g. zinc finger protein
- cytokines—e.g. IL3, IL6

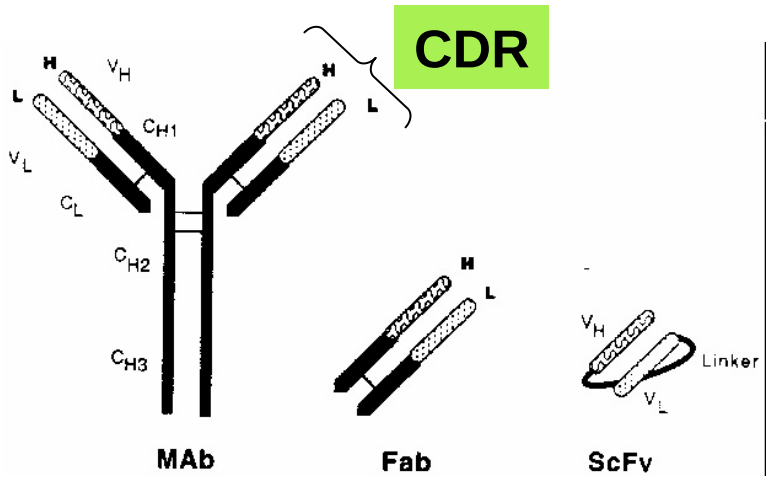
Smith and Petrenko, Chem Rev 97, 391 (1997)

Engineering antibody

- Monoclonal antibodies (mAb) have a huge therapeutic potential
  - cf. polyclonal antibodies
- Antibody with designed specificity finds applications in science and biotech
- Phage display can either engineer new specificity or fine-tune an existing one
- The functional part of an antibody is the antigen binding fragment (Fab), and the variable domains of light and heavy chains may be covalently linked in a single chain antibody (**scFv**)

	receptors
monoclonal antibody antibodies against	tumor suppressor protein p53
	HIV gp120
	fibroblast growth factor
	carbohydrate Lewis^{Y} antigen
	acetylcholine receptor
	angiotensin II
	glycoprotein D of herpes simplex virus type I
	oncoprotein p185^{HER2}
	keratin
	plasminogen activator inhibitor type-1
	bluetongue virus VP7
	FLAG octapeptide
	Na^+/K^+-ATPase β-subunit
	hepatitis B virus surface antigen
	dengue virus
	dystrophin
	von Willebrand factor





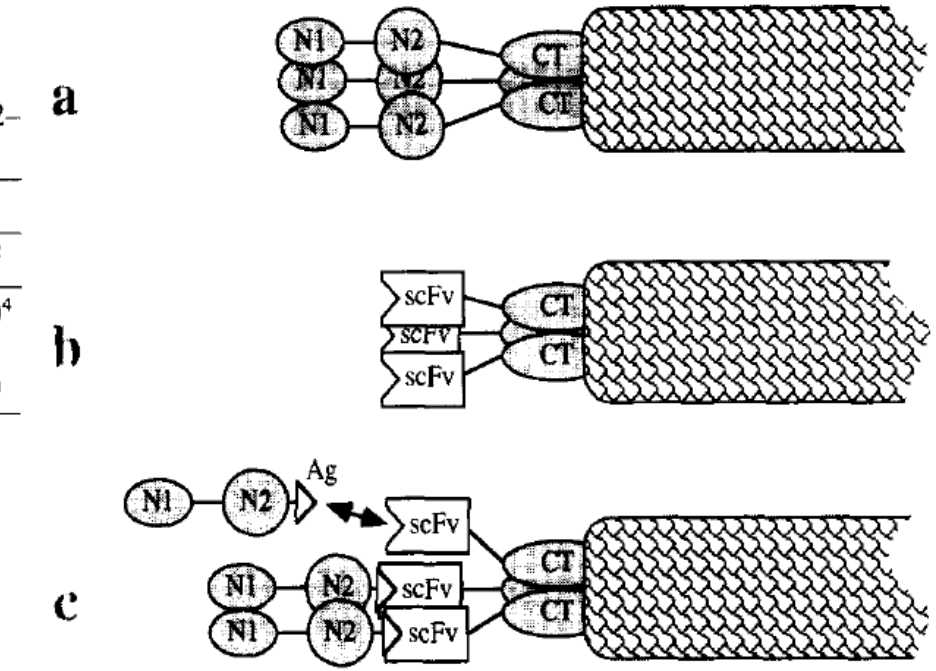
Simultaneous engineering of Ab and antigen

- pIII coat protein is required for infectivity, which is the physical basis of selection
- Apply structural complementation assay to phage display by fusing antibody to the C-terminal fragment of pIII and antigen to the N-terminal fragment
- Co-expression of the two will make the phage infective (selectively infective phage, **SIP**) only when there is strong binding between antibody and antigen

Table 2
Dependence of infectivity on the combination of scFv-CT with N1-N2–Ag

Phage			Input phage particles			
fscFv-Ag	scFv-CT	N1-N2-Ag	10^9	10^{10}	10^{11}	10^{12}
f17/9-hag	sc17/9-CT	N1-N2-hag	13	228	~10^3	~10^4
f4D5-hag	sc4D5-CT	N1-N2-hag	0	0	0	5
f17/9 no Ag	sc17/9-CT	N1-N2	0	0	0	0

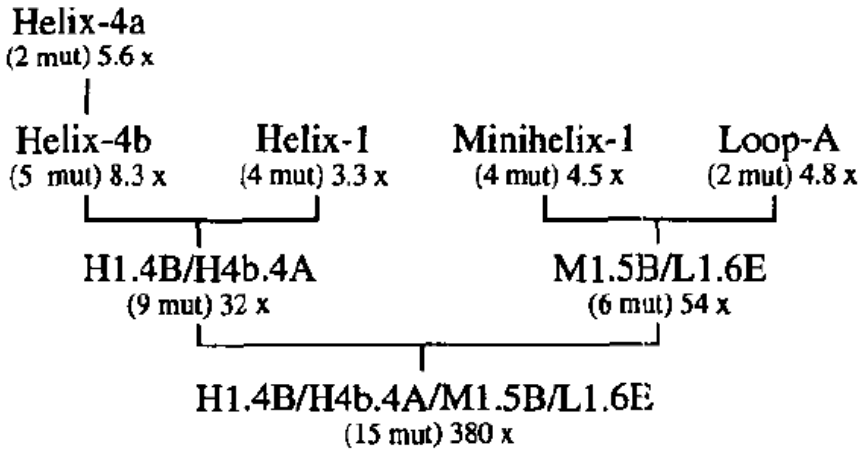
Number of colonies as a function of input phage particles.

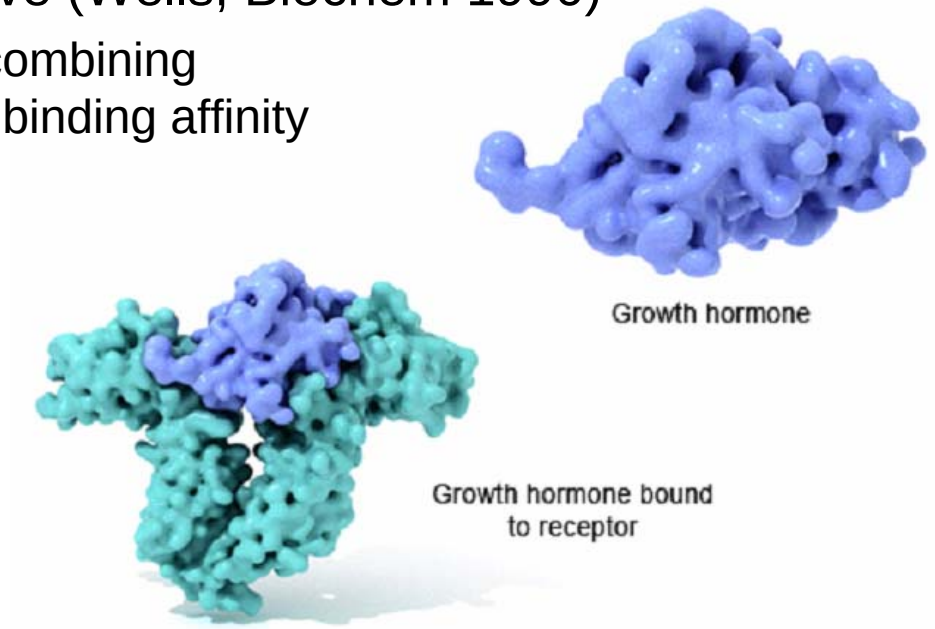


Krebber et al, FEBS Lett 377, 227 (1995)

Affinity maturation

- Phage display is well-suited to optimize binding affinity
- Human growth hormone binds hGH receptor (hGHbp) on cell surface using two independent sites (1 and 2)
  - Cunningham and Wells, Science 24, 1081 (1989)
  - Mutations on hGH that increases affinity at site 1 and decreases affinity at site 2 is an effective hGHbp antagonist
- Mutational effects are often additive (Wells, Biochem 1990)
  - engineer high affinity mutant by combining mutations selected for increased binding affinity

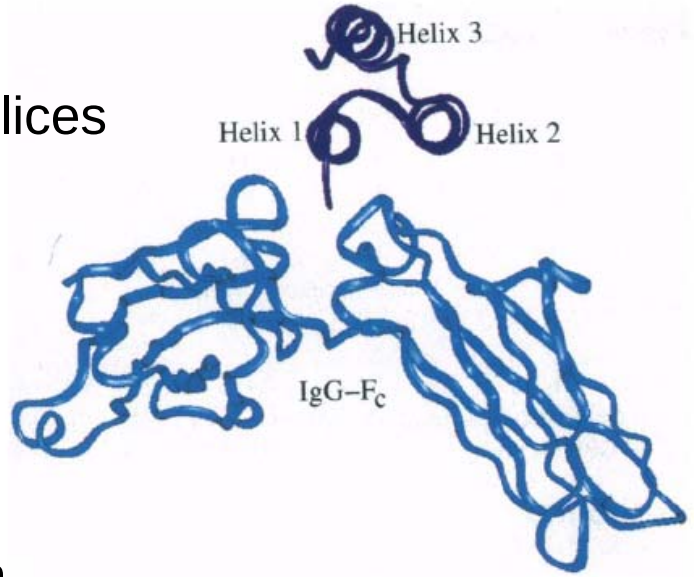




Lowman and Wells, JMB 234, 564 (1993)

Minimal IgG binding peptide

The B-domain of protein A from *Staphlococcus aureus* binds to the hinge region of the constant domain of IgG (Fc) with Kd = 10 – 50 nM



Z-domain is part of B-domain consisting of 3 helices

Crystal structure shows that only two helices contact IgG but the third helix is required to provide stability

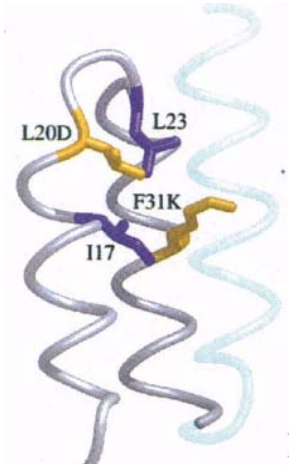
<u>Engineering a two helix version of the Z domain</u>

Systematic optimization of the exoface, intraface, and interface

i) exoface: between helix 1 and 2 with helix 3
   - degenerate library using “NNS” codons
   - randomize four residues from H1 and H2 ➔ L20D, F31K

Braisted and Wells, PNAS 93, 5688 (1996)

Wild-type residue	Selected residues	P_e	P_f	$(P_f-P_e)/\sigma$
		*Exoface 1 library**		
I17	I	0.031	0.47	10.5
	A	0.062	0.53	8.1
L20	D	0.031	0.67	15.6
	N	0.031	0.17	3.4
L23	L	0.094	0.94	12.3
F31	K	0.031	0.47	10.5
	F	0.031	0.18	3.5

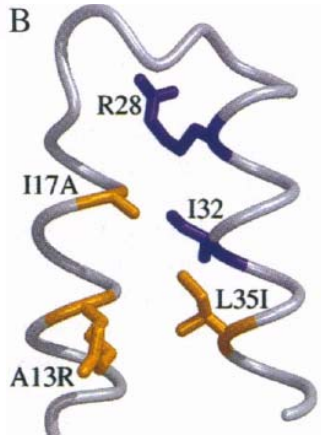


ii) intraface: between helix 1 and 2

- incorporate the residues identified from exoface selection
- randomize five residues at the interface between H1 and H2
- three new residues together repack the core

iii) interface: between helix 1 and 2 with IgG

- mutate 19 residues facing IgG Fc in groups of four

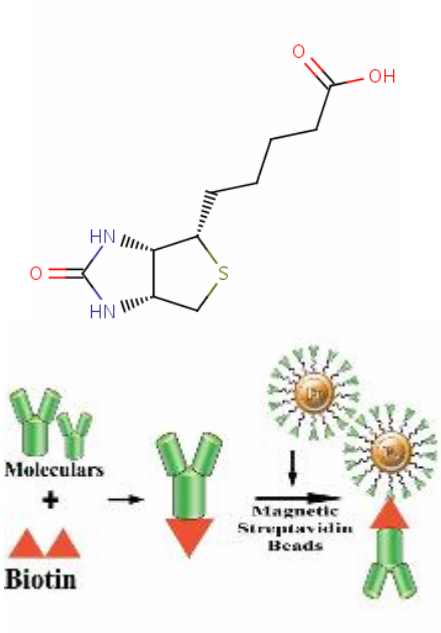


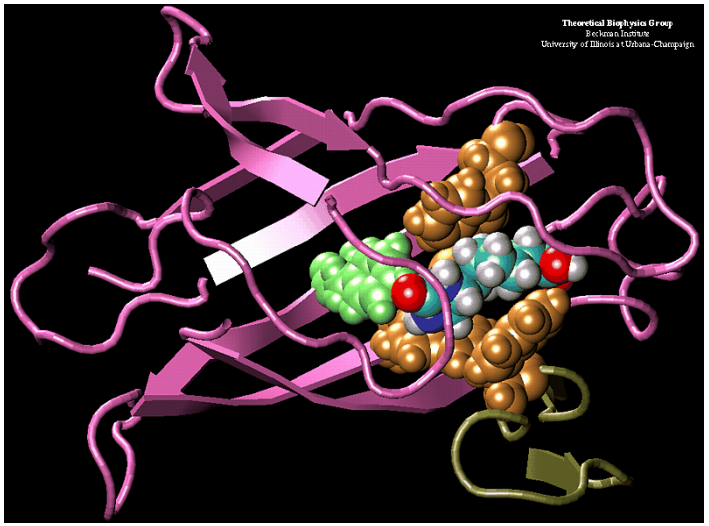
C = exo and intraface mutant

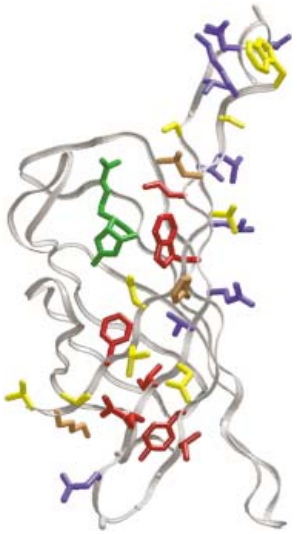
Exoface 1 plus intraface 2 plus interface 3 variants				
D = C + D3R/K5G	2.06	0.091	440	230
E = C + D3A/N4Q/K5S	1.61	0.091	570	140
F = C + K8M/E9Q	1.48	0.135	910	300
G = C + F6G/N7W/K8M/E9R	2.97	0.099	333	150
H = C + N12R	3.08	0.094	310	180
I = C + N12R/R13A	1.97	0.125	630	260
J = C + Q33K/K36R	2.00	0.073	370	140
Exoface 1 plus intraface 2 plus combined interface 3 variants				
K = D + F + H + J	5.04	0.030	60	180
L = E + F + H + J	4.87	0.030	62	60
M = F6-D38 of L	4.60	0.020	43	ND

Shotgun scanning

- Shotgun scanning refers to large scale alanine scanning in phage display format
- Offers a high throughput method of analyzing the importance of protein side chain
- Each residue included in the study is represented as a 50-50 mix of wild type and alanine in the library
- Ratio of wt to ala is used to assess the role of the side chain



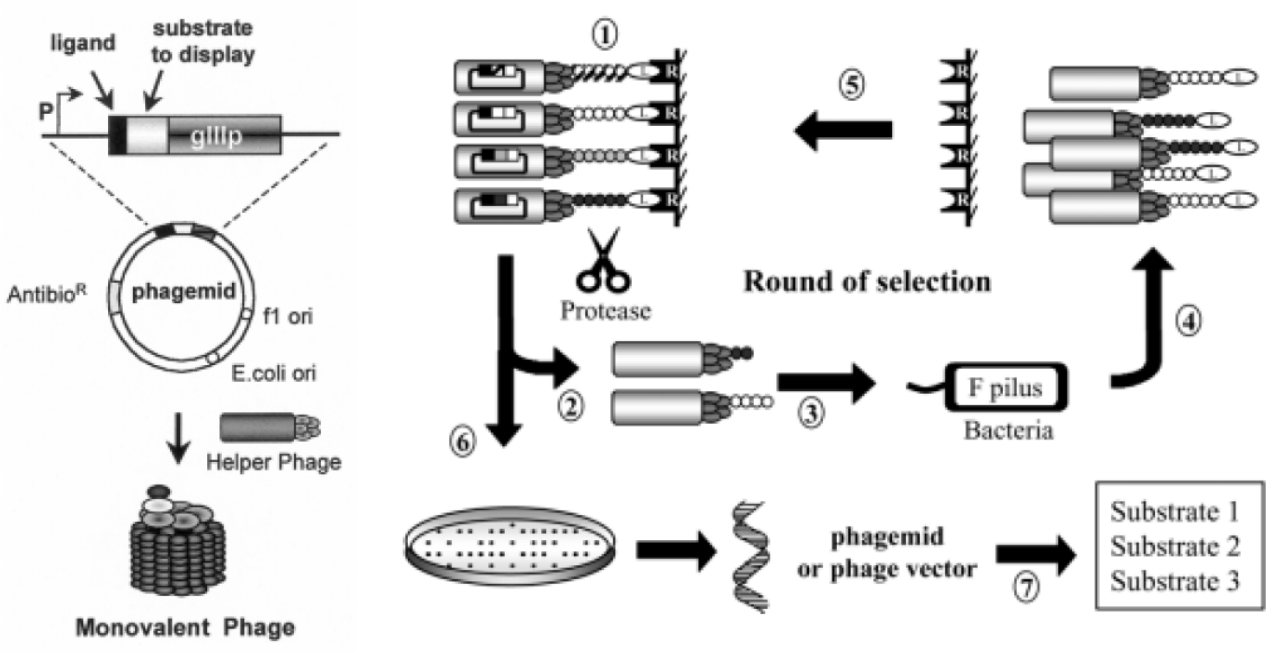




Avrantinis et al, ChemBioChem 3, 1229 (2002)

Determining protease specificity

- Selection made based on criteria other than binding affinity
- Express a peptide library on phage and bind to immobilized receptor
- Addition of a protease will cleave phage particles expressing peptides that are preferentially cleaved by the protease
- Released phage can be amplified and selection can be repeated
- Sequencing yields the consensus protease recognition sequence



Deperthes, Biol Chem 383, 1107 (2002)

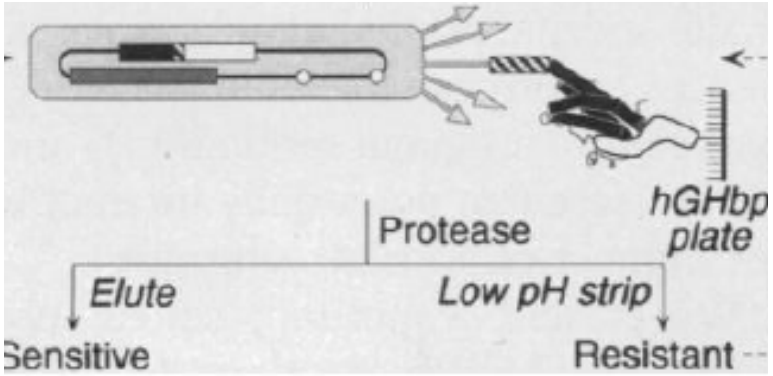
Protease Substrate

- Knowledge of protease specificity helps design inhibitors

- Substrate specificity of subtilisin H54A mutant

Protocol

- Construct a randomized peptide phage library
  - $GPGGX_5GGPG$ or $GPAAX_5AAPG$ ($3.2x10^6$)
- Bind to a matrix through an epitope (hGH)
- Add a protease to cleave
  - phage expressing a consensus protease sequence is released
  - others remain bound
- Known substrate: AAHYTRQ
- Substrate mediated catalysis



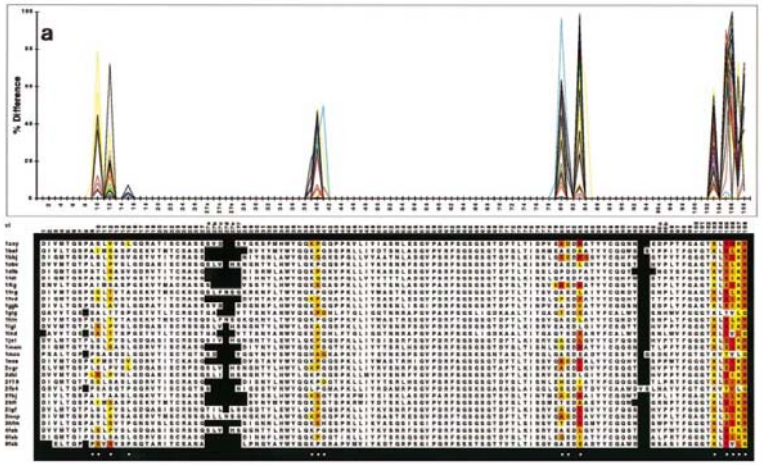
Sequence	AP released (ng/ml)	His position
NHY ↓ TL	401 ± 22	P2
TSM ↓ HT	221 ± 18	P1′
Y ↓ HLKM	89 ± 13	P1′
FHM ↓ NV	56 ± 14	P2
DGY ↓ HY	47 ± 14	P1′
THY ↓ FL	35 ± 13	P2
TSN ↓ HI	24 ± 11	P1′
HPSEP	0	

Matthews and Wells, Science 260, 1113 (1993)

Stability Optimization

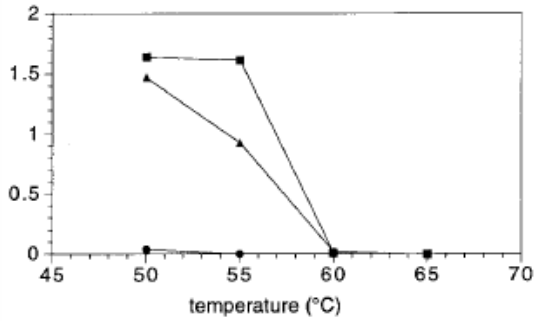
- Improve stability of scFv by phage display
- Incubate phage at high temperature or in the presence of GdnHCl
- Replace solvent exposed hydrophobic residues to polar residues
- In vivo, in vitro folding, affinity, solubility
- V48D mutation in 4-4-20 scFv improves expression by 25 fold, affinity unaffected

Nieba et al, Protein Eng 10, 435 (1997)



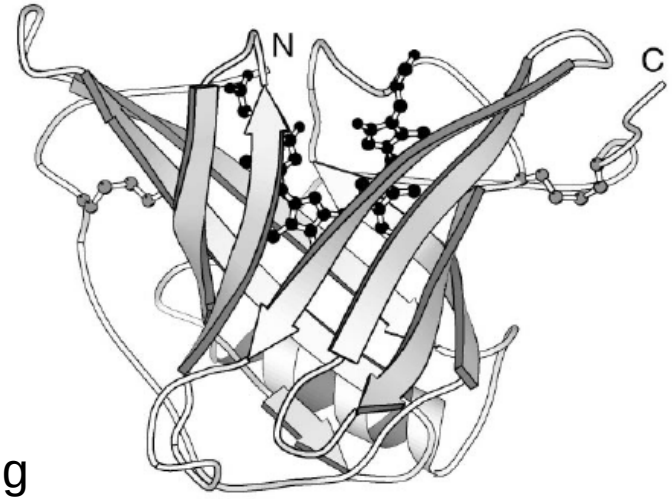
Position	w.t. amino acid	Intended mutations[a]	Allowed mutations in randomization[b]	Rationale
V_L 15	V	S,T,L,P	V,S,T,L,P,M	Residues at former V/C interface and involved in core packing: increase hydrophilicity and optimize core-packing (Forsberg *et al.*, 1997; Nieba *et al.*, 1997)
V_L 78	L	T,V,A	L,T,V,A,P,M	
V_L 80	P	S,A	P,S,A	
V_L 83	F	S,T,A	F,S,T,A,I,V	
V_L 106	L	S,T	L,S,T	
V_H 5	V	Q	V,Q,L,E	Q is highly conserved/exposed
V_H 6	E	Q	E,Q	Subgroup change: charged residue buried in core
V_H 9	G	A,P	G,A,P,R	
V_H 11	L	S,D,N	L,S,D,N,F,Y,P,H,I,T,V,A	Increase hydrophilicity at former V/C interface
V_H 89	V	S		
V_H 108	L	S		
V_H 49	A	G	A,G	Fairly conserved G with positive ϕ-angle

[a] Mutations intended for the reasons given in the column Rationale.
[b] The mutations were encoded on oligonucleotides by randomized bases. The oligonucleotides were introduced into the DNA shuffling experiment, shuffling the 4D5Flu w.t. and the previously constructed single point mutants L(H11)S, V(H89)S, L(H108)S.



Jung et al, JMB 294, 163 (1999)

Lipocalin

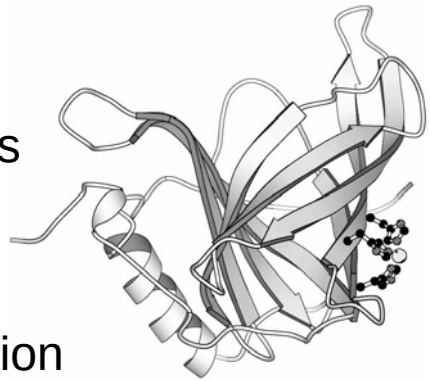


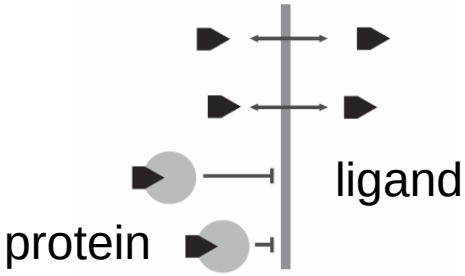
- Lipocalin is a family of beta-barrel proteins often involved in molecular transport
  - can bind hydrophobic molecules
  - include retinol binding protein, fatty acid binding protein, bilin binding protein
  - implicated in pheromone transport, olfaction, inflammation
  - low sequence homology, high structural similarity
- Some topologies are better suited to serve as a scaffold
  - small yet stable, soluble, good expression
  - tolerant to amino acid substitution and insertion
  - separation of a tertiary structure into a part responsible for stability, and another part that allows structural variability
  - e.g. immunoglobulin, TIM barrel, lipocalin
- Structural stability has enabled introduction of affinity for novel compounds—these are called “anticalin” in analogy to antibody

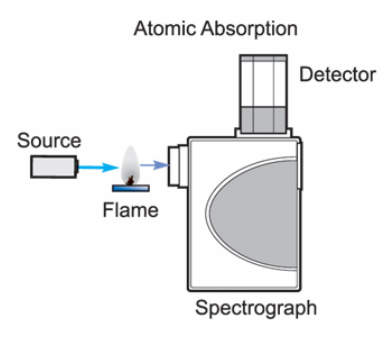
Skerra, Biochim Biophy Acta 1482, 337 (2000)

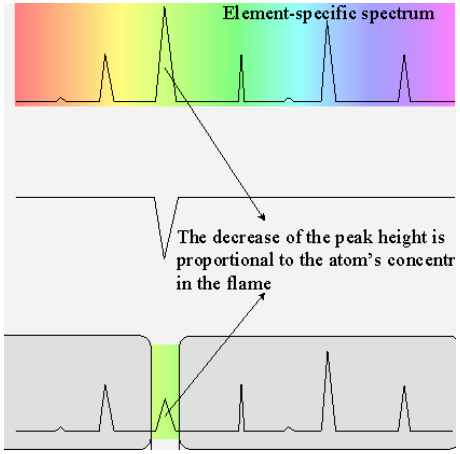
Metal binding site on the outer surface

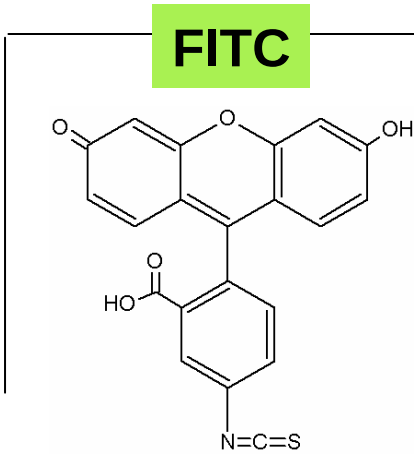
- Zn binding in human carbonic anhydrase II comprises 3 His
- mutate residues i) 46, 54, 56 or ii) 76, 78, 79 to His
- equilibrium dialysis to test zinc binding
- atomic absorption spectrometry to measure the concentration of free and bound metal ion
- Kd ~ i) 36 nM and ii) 440 nM









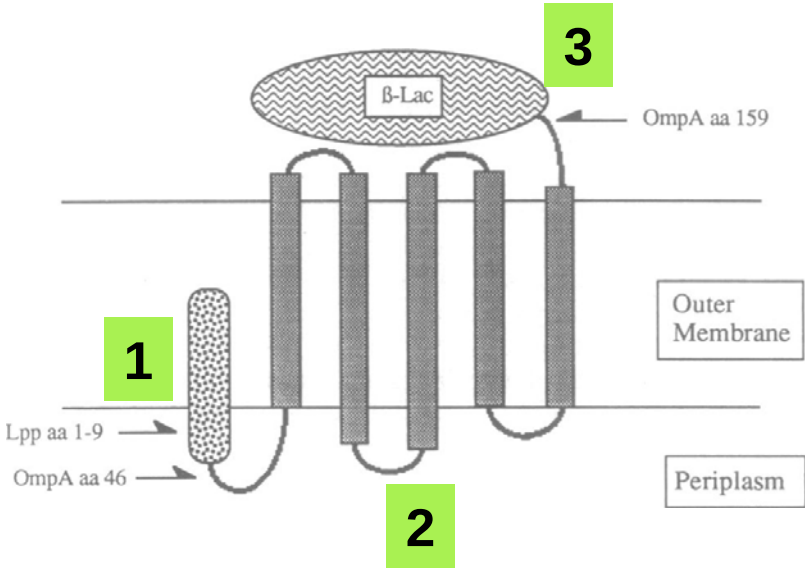


Randomize 16 residues within the four loops of BBP

- display randomized proteins on phage and look for binding to fluorescein
- Kd ~ 35 nM
- deep ligand binding pocket can be reshaped
- same library can be screened for other molecular targets

Bacterial surface display

- Proteins can be similarly displayed on bacteria
- Targeting a protein for expression on the surface of bacteria requires a signaling peptide—e.g. first nine residues of E coli lipoprotein (Lpp)
- OmpA provides anchoring in the membrane
- beta-lactamase expressed on the surface is susceptible to proteolysis and can hydrolyze penicillin in the solution



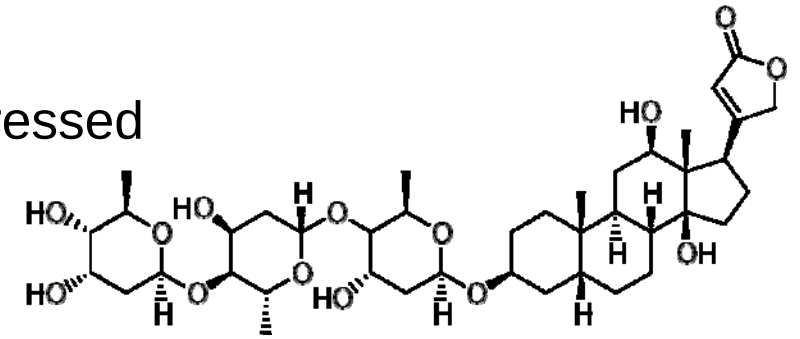
Plasmid	Growth temperature, °C	% decrease in penicillin G hydrolysis after incubation* With proteinase K	With trypsin	Nitrocefin or penicillin G hydrolysis activity in intact cells, % of total lysate activity Nitrocefin	Penicillin†
pJG311	37	3	3	1	1 (0.03/3.45)
pJG311	24	—	—	6	3 (0.10/3.74)
pTX101	37	23	18	33	56 (0.67/1.19)
pTX101	24	—	89	93	100 (1.85/1.84)

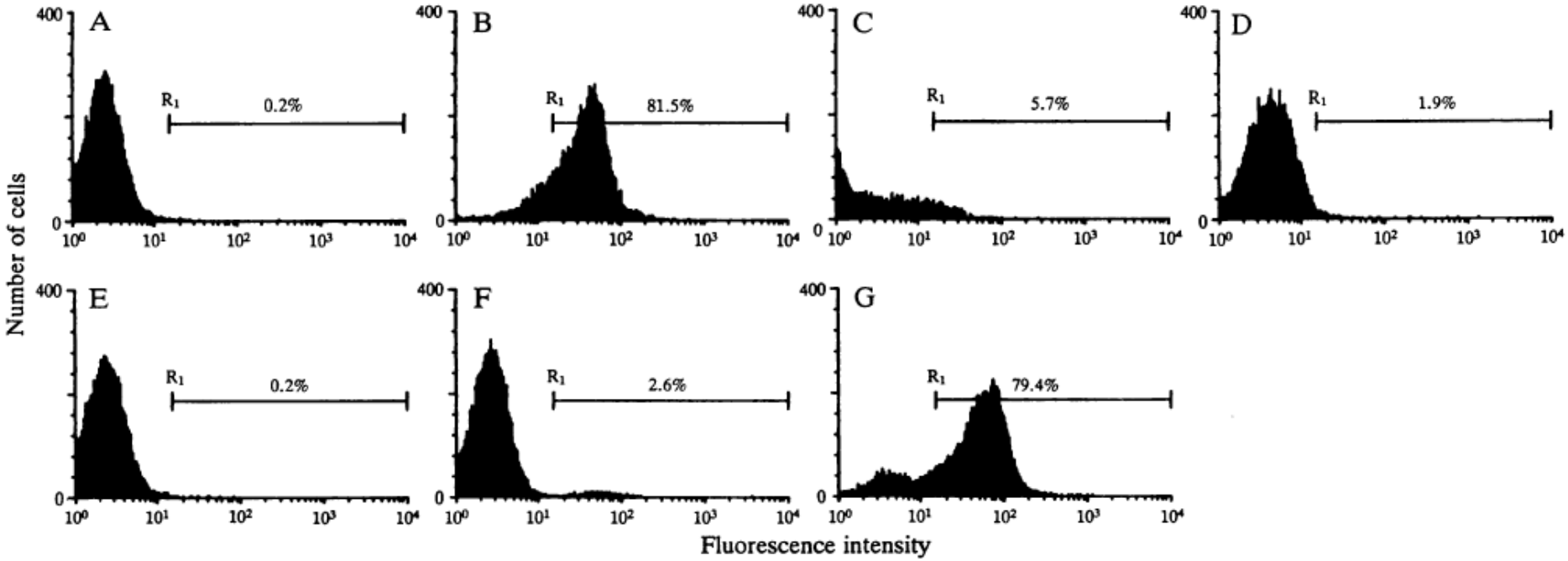
Francisco et al, PNAS 89, 2713 (1992)

Functional single chain antibody was expressed

- antibody known to bind digoxin

Use digoxin-FITC to visualize binding



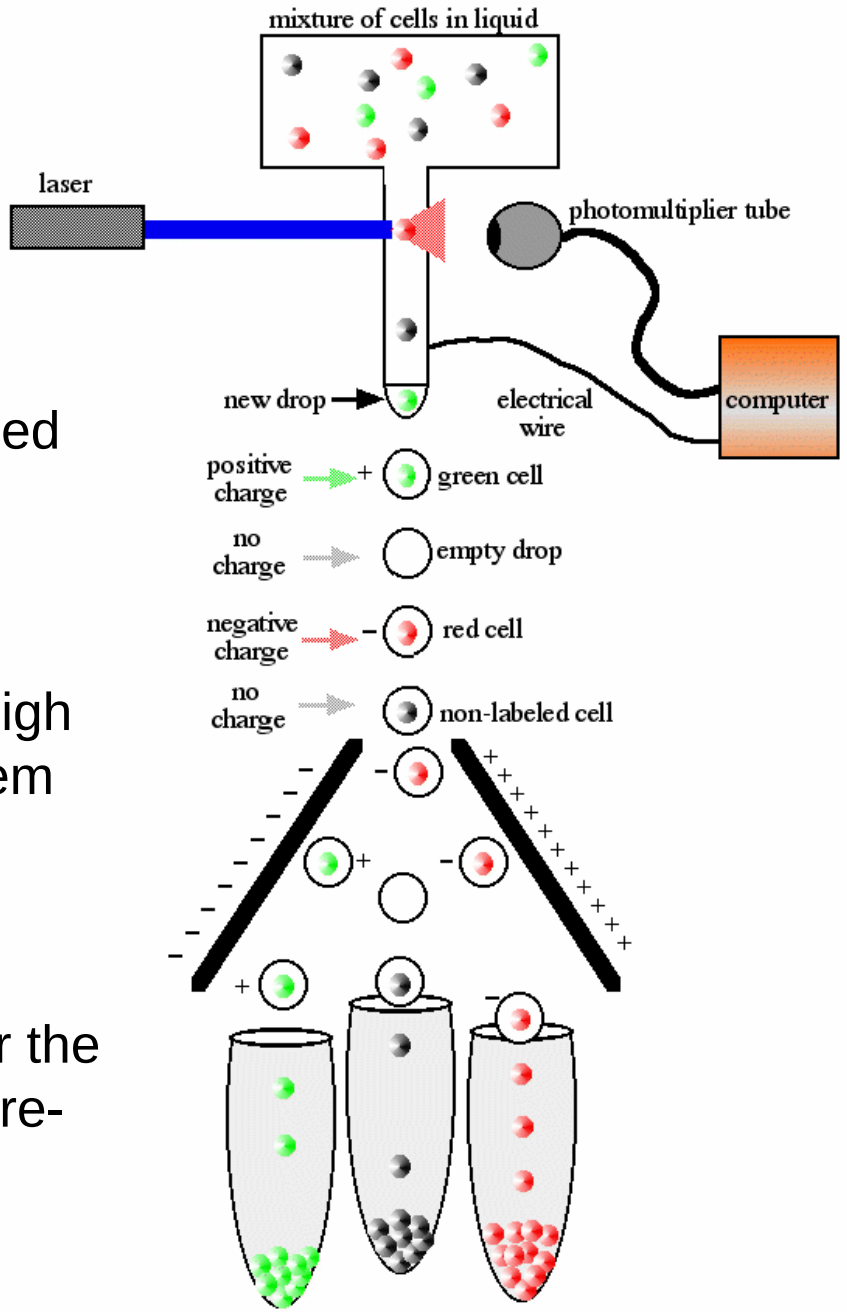


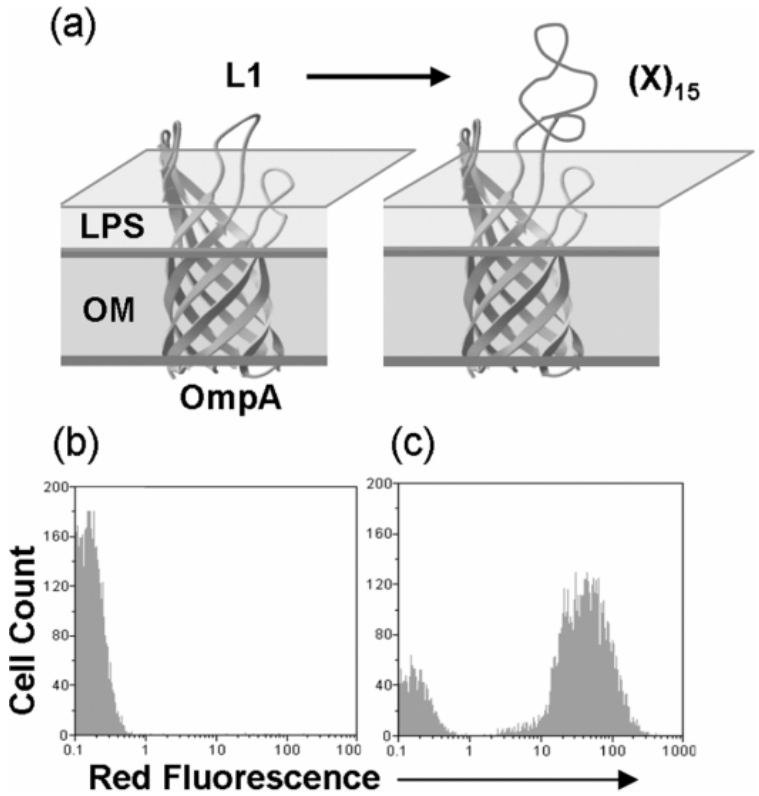
a. negative control
b. positive control
c. trypsin treatment prior to binding
d. pre-treatment with digoxin
e. diluted with 10^5 excess control E coli
f. after first round of sorting
g. after second round of sorting

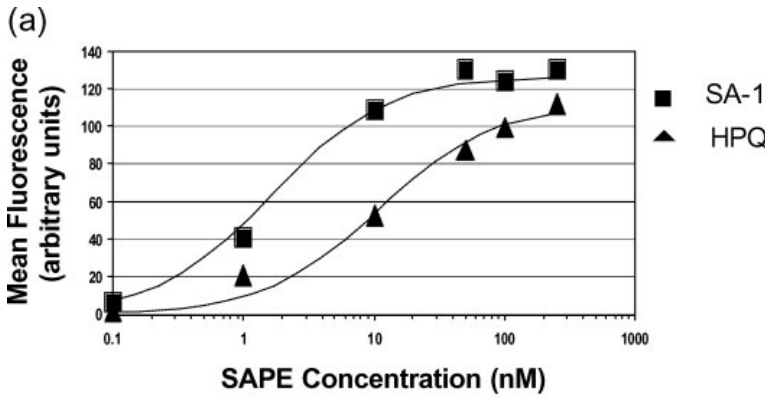
Francisco et al, PNAS 90, 10444 (1993)

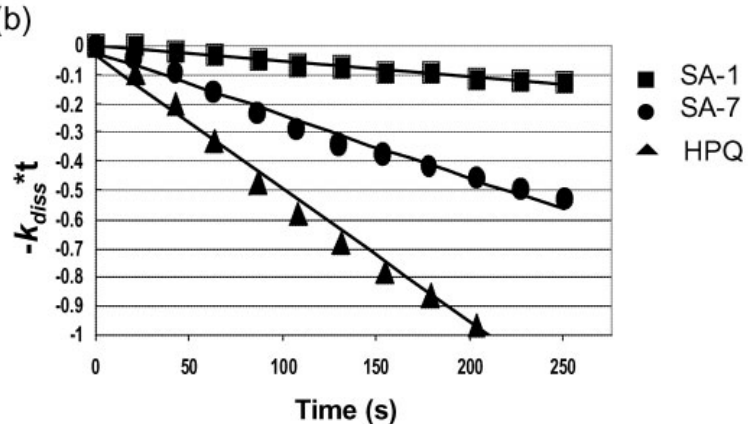
FACS

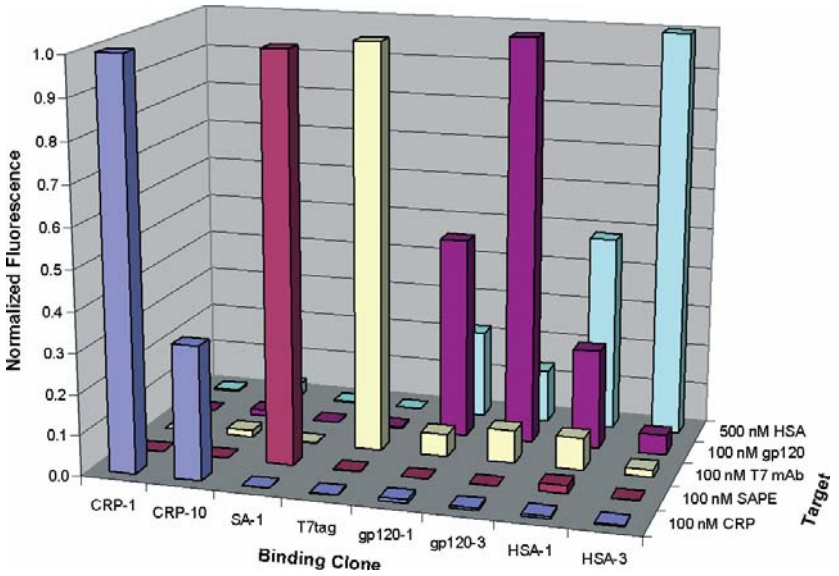
- Fluorescence assisted cell sorting
- Cells or beads with proteins on the surface
- Individual cells are fluorescently labeled using antibody (or streptavidin)
- Substrate binding correlates with increased fluorescence
- Laser can inspect individual cells at high speed (> 1,000 cells/sec) and sort them based on a combination of color and intensity
- Sorted cells represent an “enriched” population and the average affinity for the substrate is higher compared to the pre-sort population
- Works with bacteria, yeast, and mammalian cells, but not with phage











Bessette et al, PEDS 17, 731 (2004)

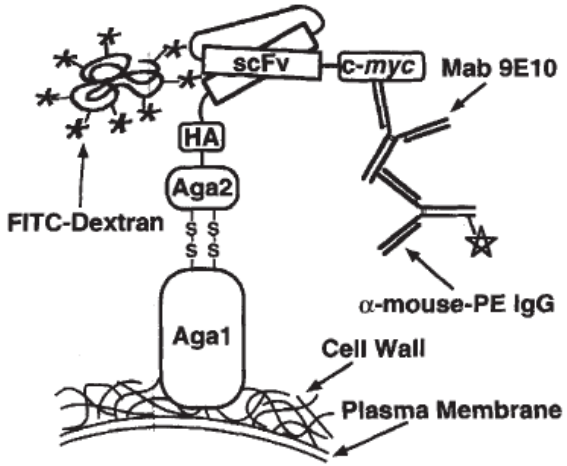
Yeast surface display

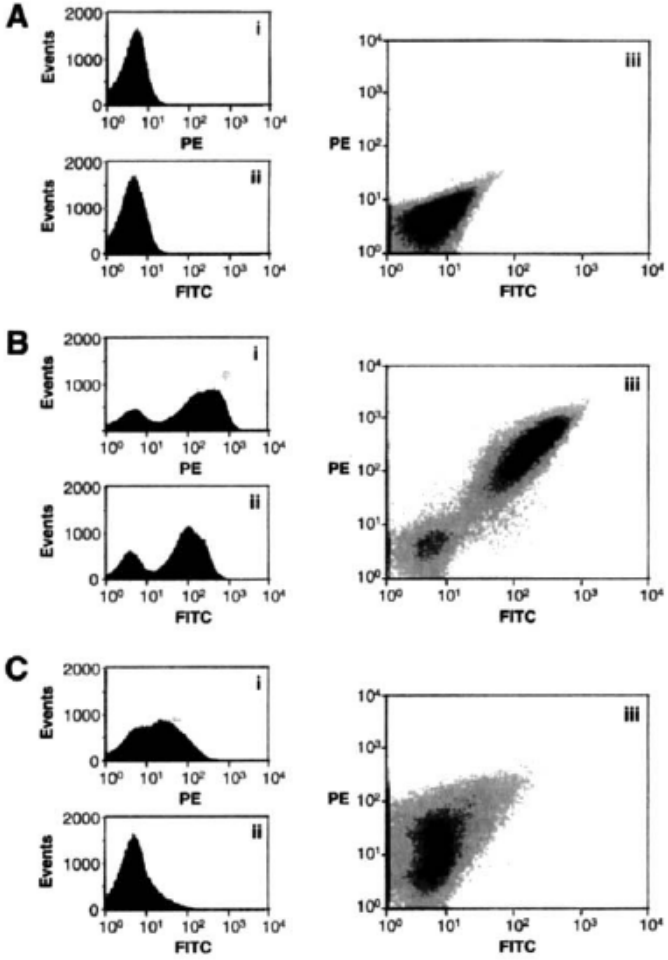
Similar to bacterial surface display but takes advantage of the eukaryotic expression system available in yeast

- post-translational modification, including glycosylation and disulfide formation

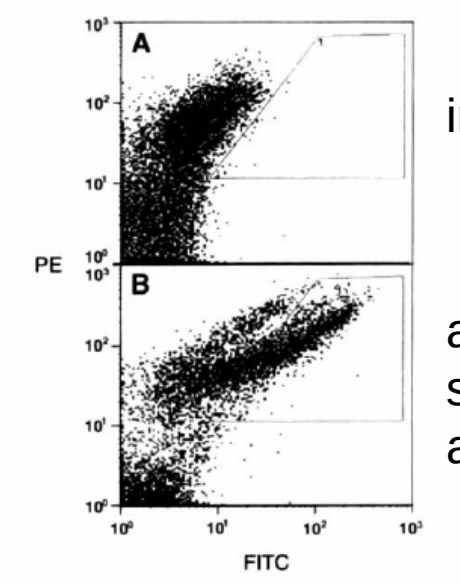
Ease of genetic manipulation in yeast

e.g. ability to maintain multiple plasmids





Boder and Wittrup, NSB 15, 553 (1997)

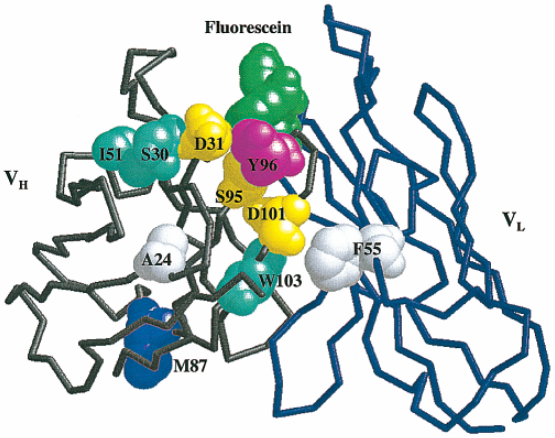


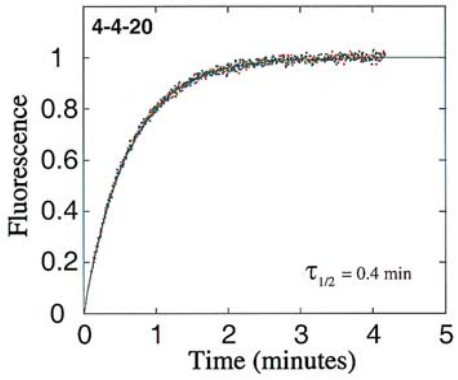
initial library

after three rounds of sorting and amplification

Table 1. Sorting enrichments of scFv-displaying yeast.

Initial purity	Sorted purity	Enrichment ratio
14%	83%	6.0
0.5%	95%	200
0.1%	59%	600





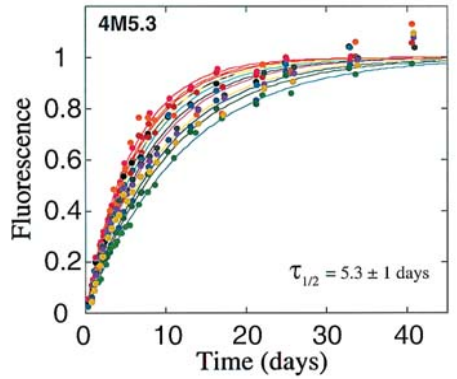
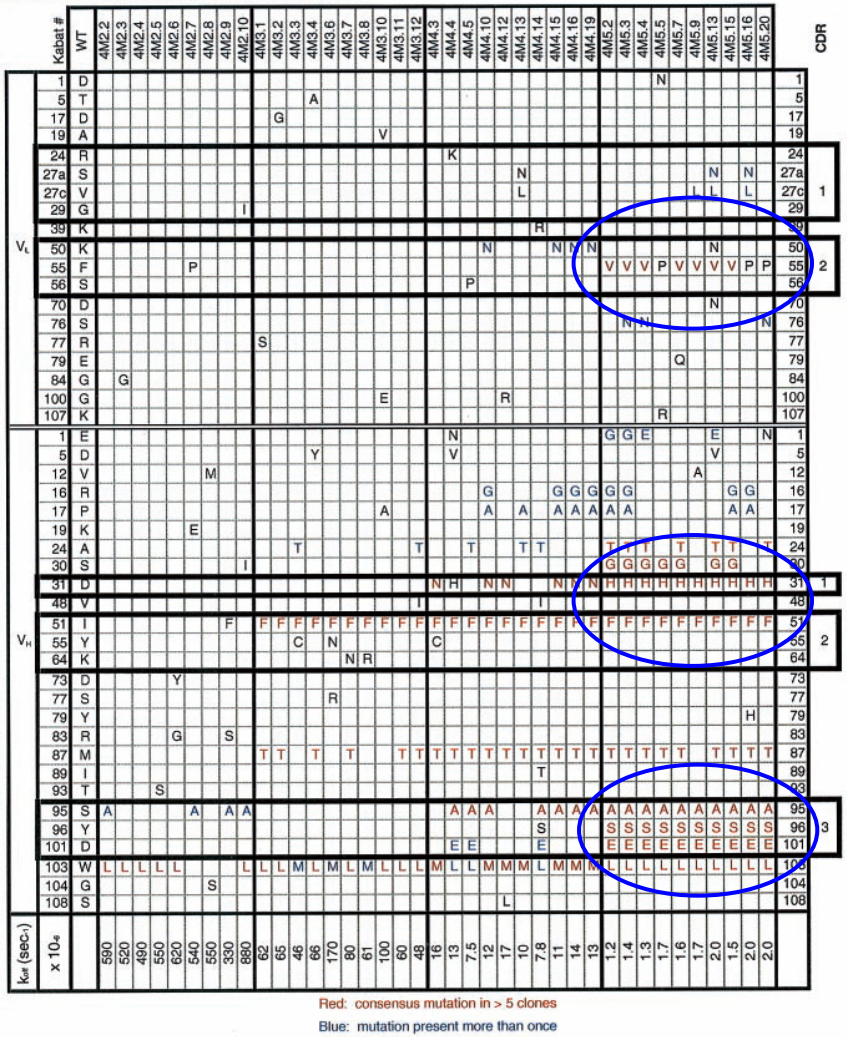


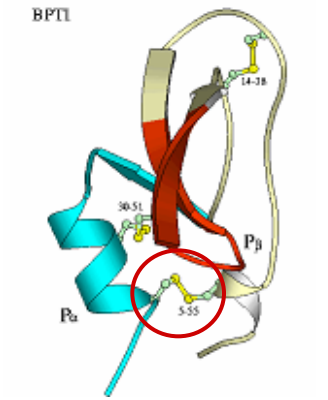
Table 2. Mutations in selected scFvs

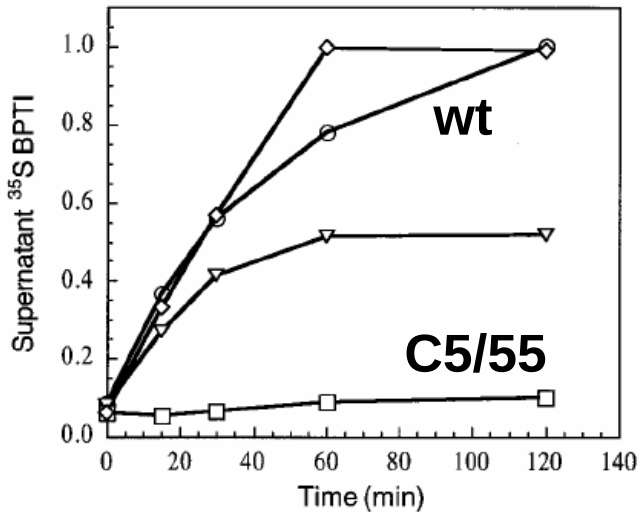


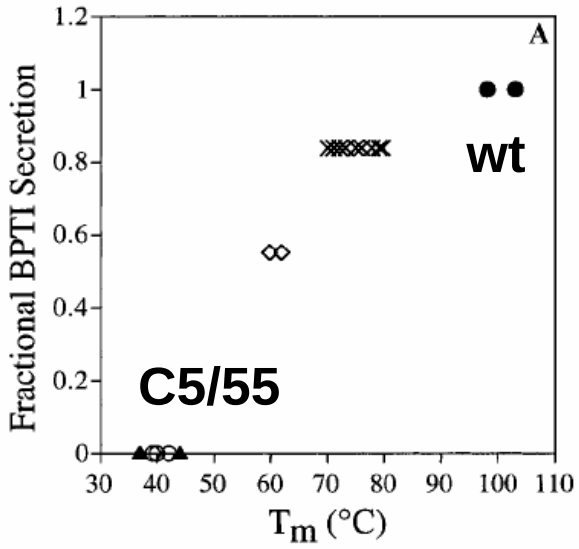
Boder et al, PNAS 97, 10701 (2000)

Engineering by Secretion

Secretion efficiency from yeast correlates with stability
- Kowalski et al, Biochemistry 36, 1264 (1998)

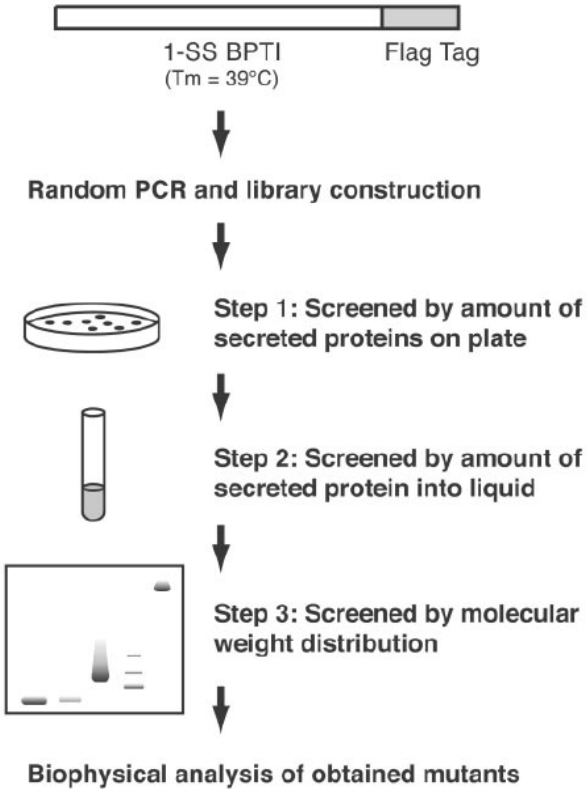






mutants lacking various disulfide bonds

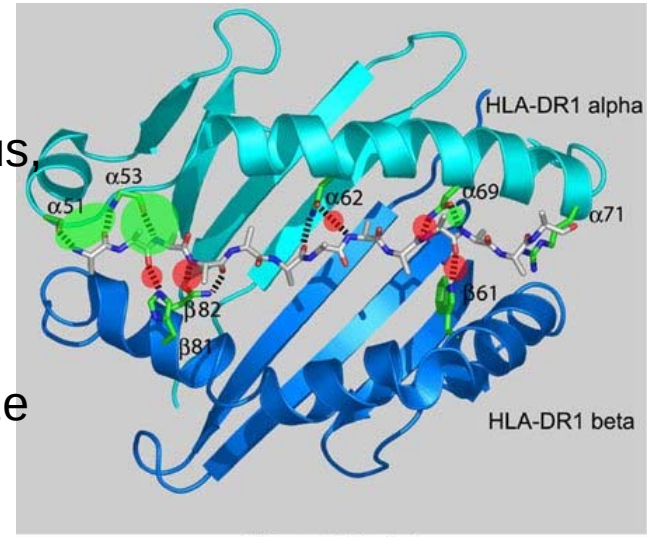
Protein	Mutations	Number of disulfide bonds	Tm, °C	Ref.
Wild-type BPTI	None	3	>90	22
2-SS BPTI	C30A, C51A	2	66	26
1-SS BPTI	C14A, C30A, C38A, C51A	1	39	25
1-SS BPTI (Y35A)	C14A, C30A, C38A, Y35A, C51A	1	31	19
1-SS BPTI (N44A)	C14A, C30A, C38A, N44A, C51A	1	15	19
1-SS BPTI (N43A)	C14A, C30A, C38A, N43A, C51A	1	<0	19

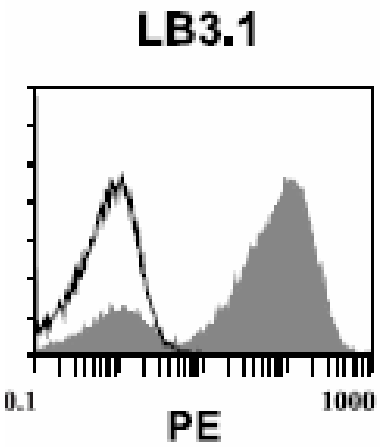


Hagihara and Kim, PNAS 99, 6619 (2002)

Single Chain MHC II

- Membrane protein that binds small foreign peptides generated by degrading invading pathogens, e.g. virus, bacteria
- Recognized by the T cell receptor (TCR) to initiate a chain of events that constitute acquired immunity
- MHC without peptide is unstable, difficult to synthesize
- Engineer a single chain version in order to study structure-function relationship better, e.g. specificity determinants: scDR1$\alpha\beta$



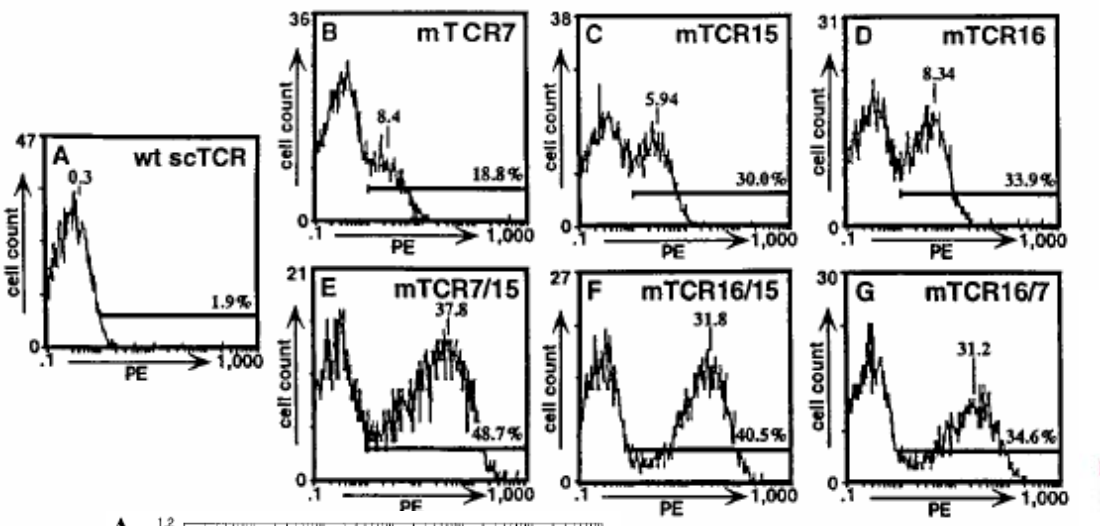


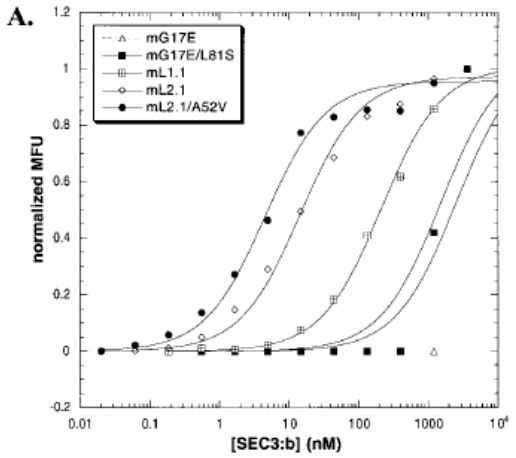
- Improved folding rate and expression

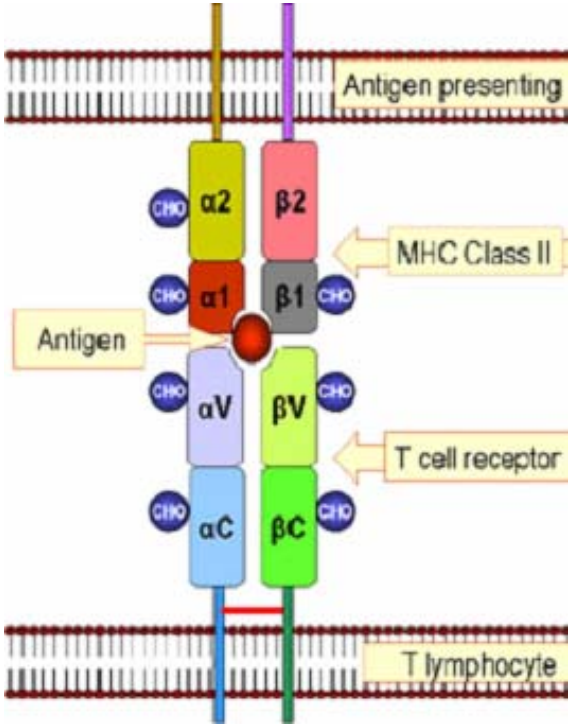
Esteban and Zhao, JMB 340, 81 (2004)

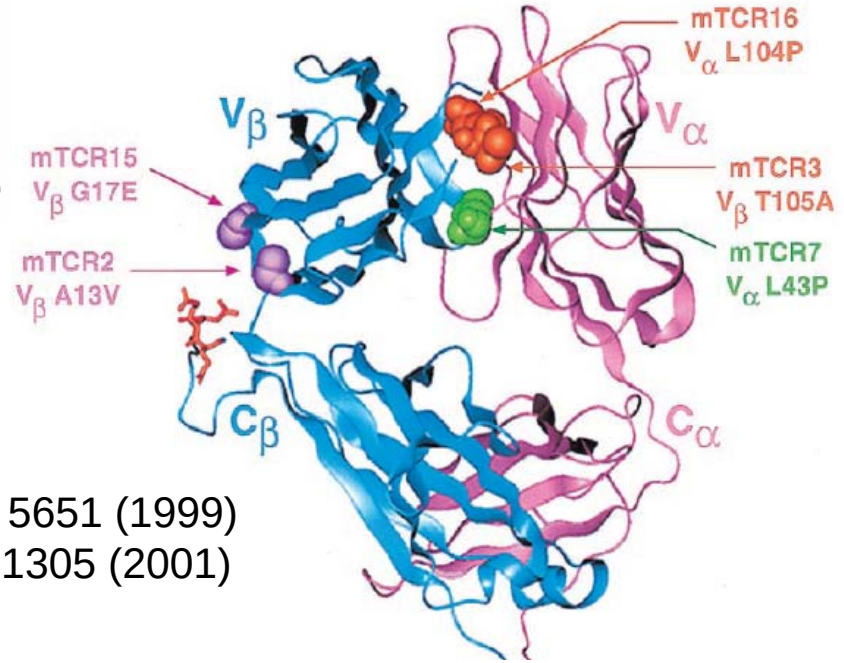
Engineering T Cell Receptor

- TCR recognizes each MHC-peptide combination
- Engineering stable single chain TCR not successful by other methods
- Yeast, being eukaryote, can synthesize large molecules









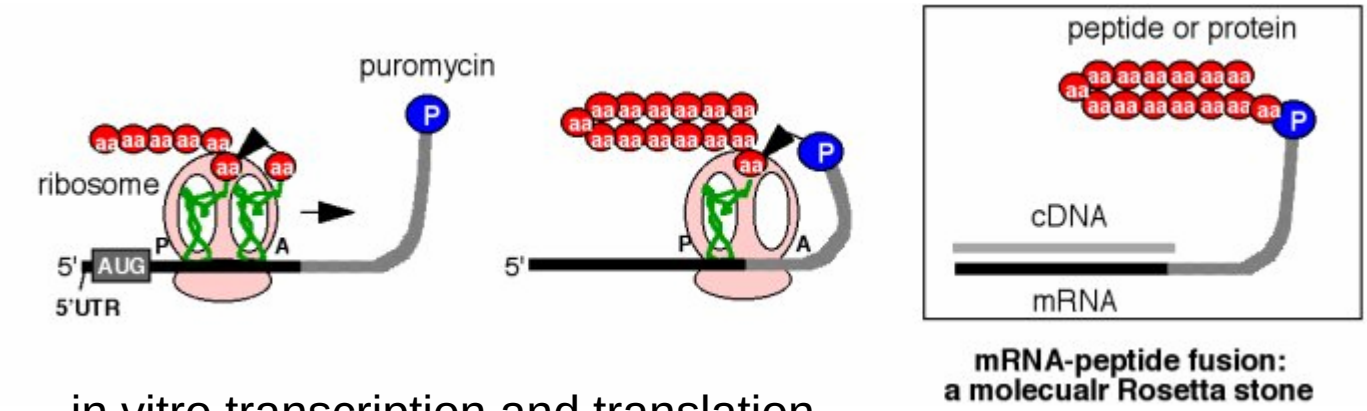
Kieke et al, PNAS 96, 5651 (1999)
Kieke et al, JMB 307, 1305 (2001)

mRNA display

- Key to doing high throughput screening is the physical association between the genomic and phenotypic data
- Genomic information can also be RNA

<u>Protocol</u>

- Covalently link mRNA to protein
- Protein can be assayed and sorted for function (e.g. binding)
- mRNA is amplified by PCR and/or sequenced

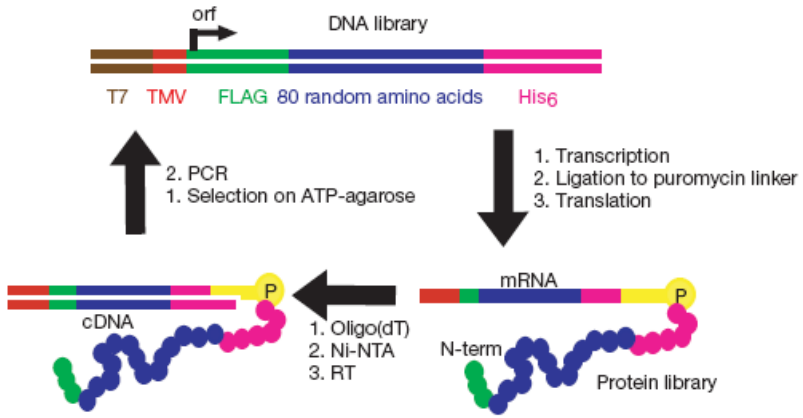


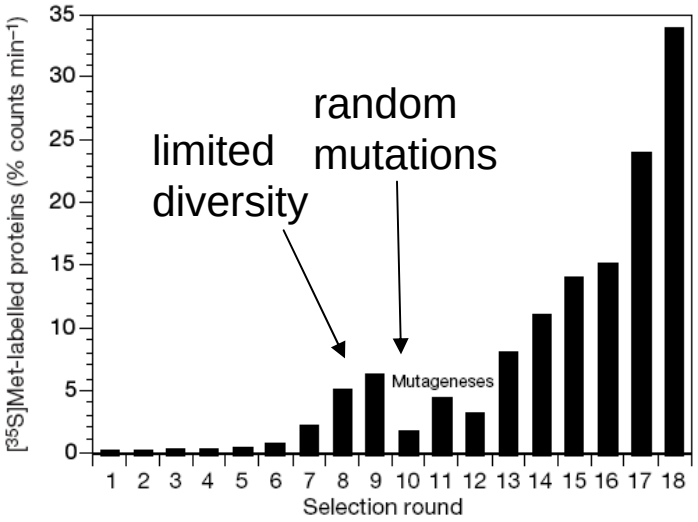
mRNA-peptide fusion:
a molecualr Rosetta stone

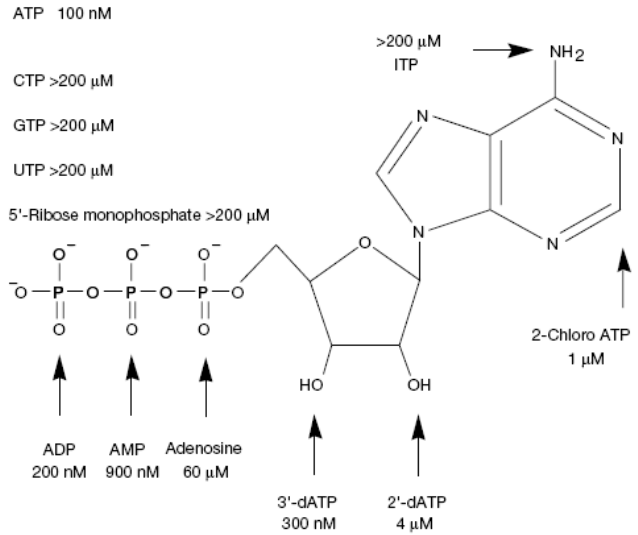
in vitro transcription and translation

Roberts and Szostak, PNAS 94, 12297 (1997)

- Randomized 80 amino acids
- Diversity of 6 x 10^12







- Enriched residues spread out throughout the protein—amino acids throughout the region contribute to the formation of folded structure

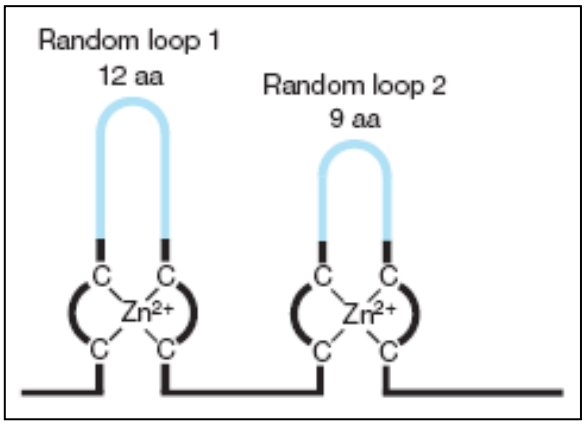
b Full-length sequence data of mutagenized re-selected proteins (round 18), free protein

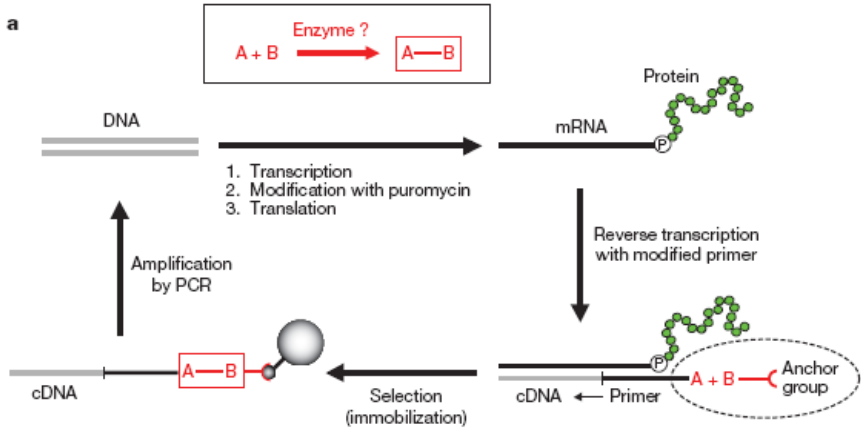
		Immobilized ATP binding (%)	Solution ATP K_d (µM)
Family B	MDYKDDDDKKTNWQKRIYRVKPCVICKVAPRDWWVENRHLRIYTMCKTCFSNCINYGDDTYYGHDDWLMYTDCKEFSNTYHNLGRLPDEDRHWSASCHHHHHHMGMSGS	5	-
18predom	--------------------R------------R---------N--------S----------	11	30
18select	------------RL------R---V--E-----R-K-G-----N-----Y--SVKN----HH--E---L---SE----------	-	-
18-19	-------------L------R---K------N-K-K-K-----N------N-S-DI-----H--------A-S--I----------N--K----------	28	0.1

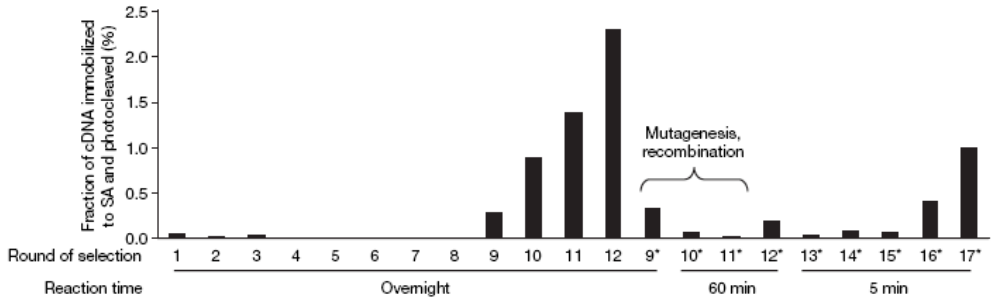
Keefer and Szostak, Nature 410, 715 (2001)

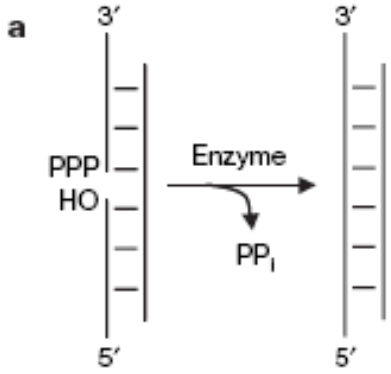
RNA ligase

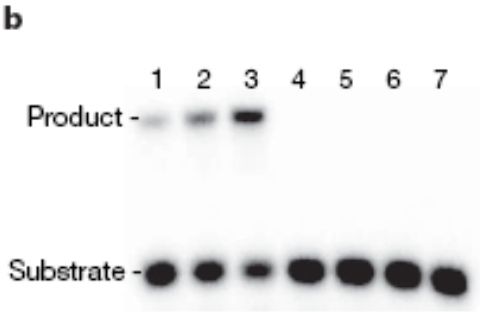
- Find a protein to ligate two pieces of RNA
- Engineer on the zinc finger motif

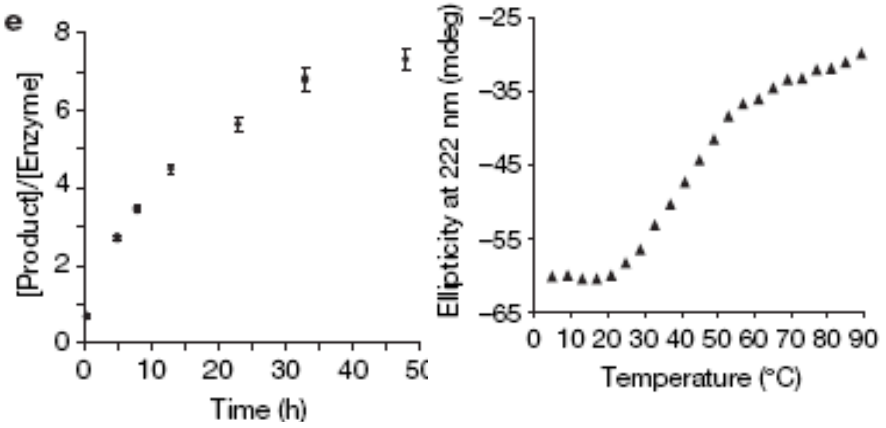












Seelig and Szostak, Nature 447, 828 (2007)

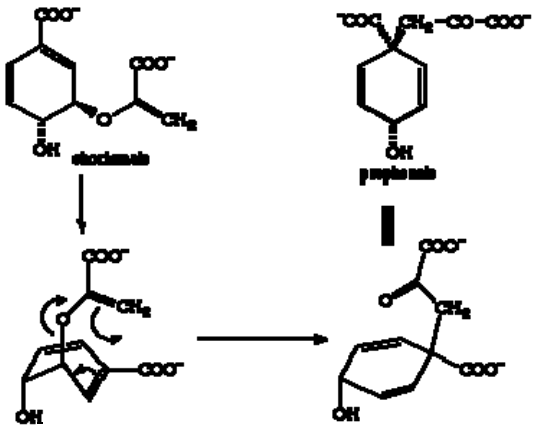
Screen vs. Selection

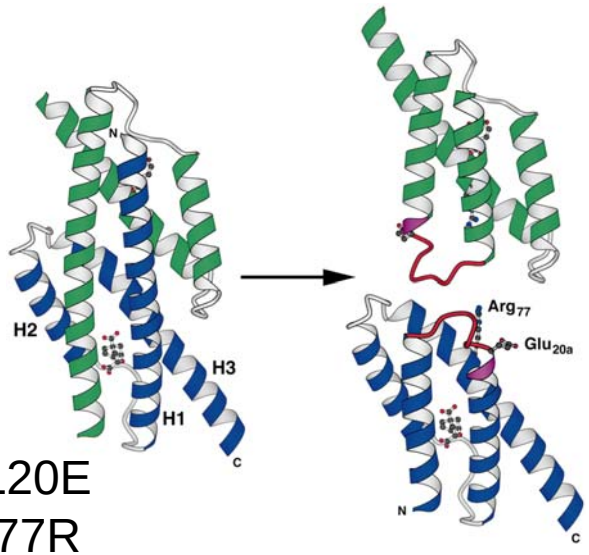
- There are two types of high throughput assays: screen and selection

- Screen requires inspecting each member of the library one at a time
  - antibody library expressed on bacterial or yeast surface needs to be screened by flow cytometry
  - enzymatic activity may be screened in 96 well plates
  - need a quantitative mechanism, e.g. fluorescence, catalysis
- Selection relies on a property that is essential for survival
  - antibody library on phage is used to select the particles with high affinity to an immobilized substrate
  - selectively infective phage
  - active engineered enzyme may be required for survival
  - may be used against a library that may be too large to screen
  - mRNA-peptide library can be as large as 10^{15}➔at the rate of 10^{4} mol/sec, would take 4,000 yrs.

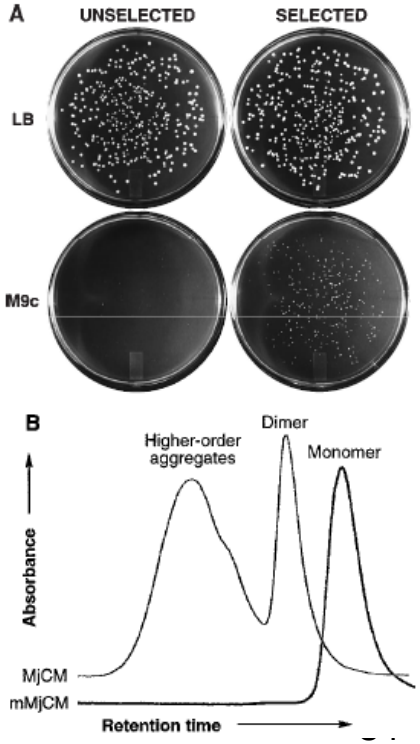
Designing a new topology

- Chorismate mutase is an essential enzyme required for the biosynthesis of aromatic amino acids Tyr and Phe
- Bacterial CM is a domain-exchanged homodimer; the active site consists of residues from both monomers
- Engineer a functional monomeric CM and thus change the enzyme topology
  - how important are the turn residues?





- L20E
- I77R
- 6 randomized residues



MacBeath et al, Science 279, 1958 (1998)

Diversity generation

- In nature, random mutations and recombination lead to genetic diversification
- Direction evolution requires a mechanism for introducing genetic variations

Random mutation

- error prone PCR
- nucleoside analogs
- degenerate oligonucleotides
- propagation in strains lacking DNA repair capabilities: mutS, mutD, mutT
- growth in the presence of chemical mutagens: deamination, alkylation

Recombination

- DNA shuffling

Error prone PCR

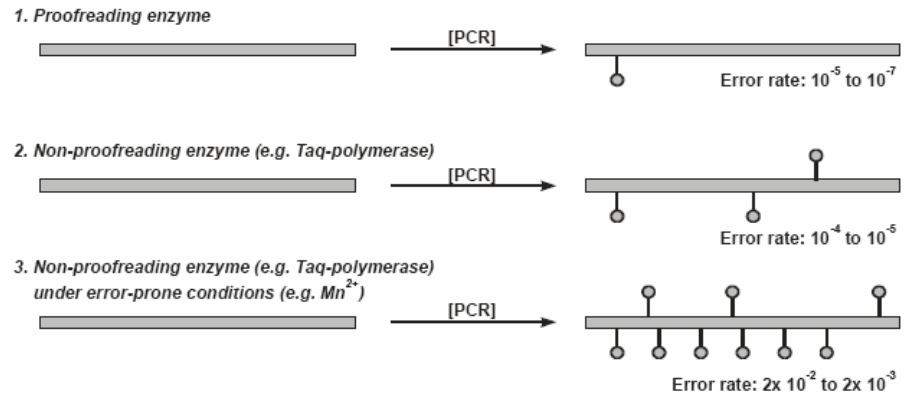
DNA polymerases are naturally engineered to achieve high fidelity

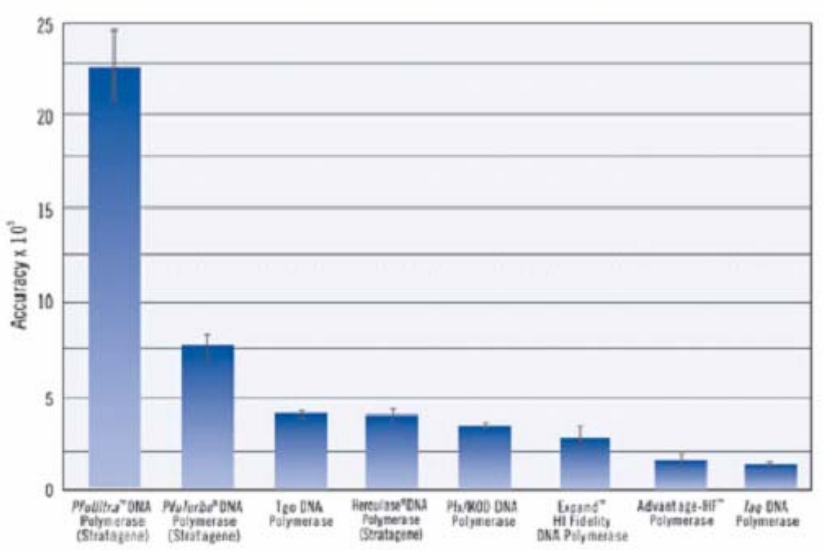
- pfu – 1 error in 10^6 to 10^7
- Taq – 1 error in 10^4 to 10^5

Create an artificial condition that is conducive of base misparing

- use polymerase lacking proofreading activity—e.g. Taq
- high Mg++ concentration
- substitution of Mn++ for Mg++
- uneven concentrations of dATP, dCTP, dTTP, and dGTP
- very low template concentration
- organic solvent—DMSO, alcohol

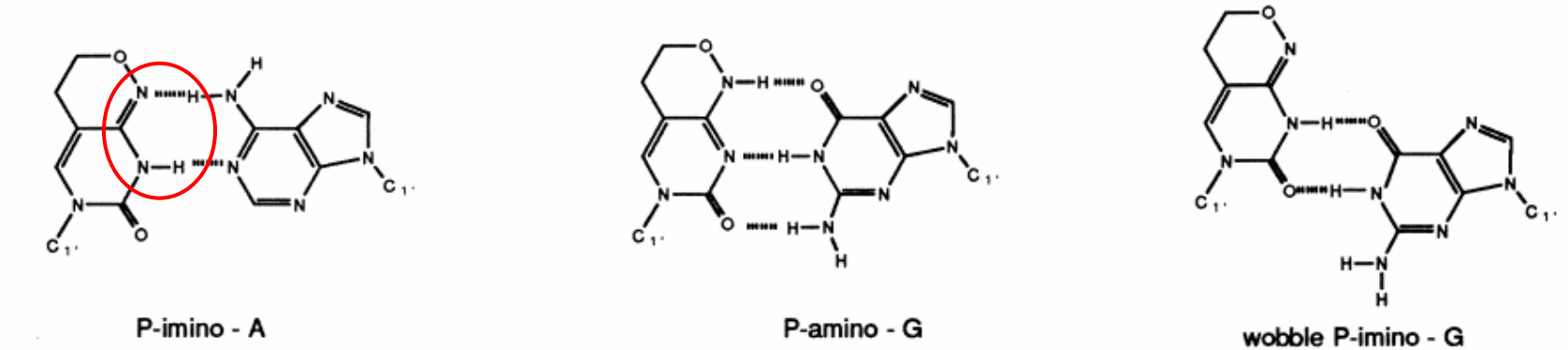
Used to create mutations at random locations—cf. site-directed mutagenesis





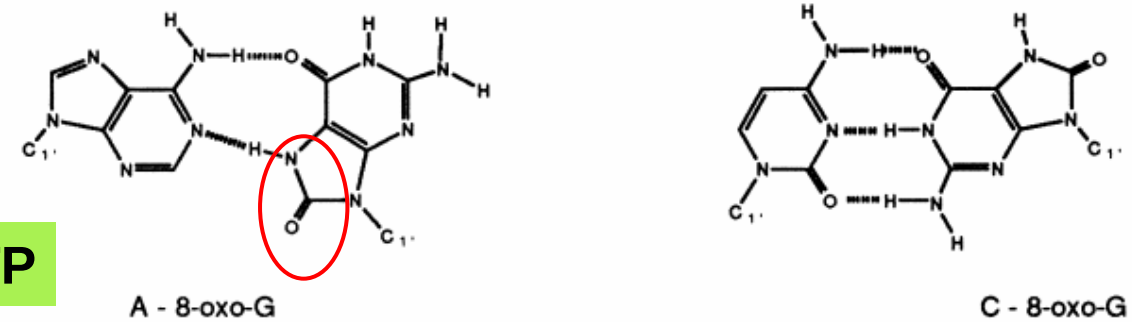
Nucleoside analog

Adding nonnatural dNTP during PCR leads to mutations because some of them base-pair promiscuously



dPTP 6-(2-deoxy-b-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C][1,2]oxazin-7-one(dP)

d8-oxo-GTP



Zaccolo et al, JMB 255, 589 (1996)

Degenerate oligonucleotides

Targeting mutations to a fixed location

5′-ACT GGC GAT **ATA AGT GAC** GGA TTA CGT-3′

5′-ACT GGC GAT **NNN NNN NNN** GGA TTA CGT-3′

Deletion and insertion of amino acids

5′-ACT GGC GAT **ATA AGT GAC** GGA TTA CGT-3′

5′-ACT GGC GAT **NNN --- NNN** GGA TTA CGT-3′

Use degenerate codons to control the type of amino acids

NTN: M, L, I, V, F

VAN: K, H, Q, E, N, D

N	Any	S W	G, C A, T
K M R Y	G, T A, C A, G C, T	H B V D	Not G Not A Not T Not C

Table II. Degenerate codons computed by LibDesign at each position, from most-inclusive to least-inclusive

Pos1	
NNK	ACDEFGHIKLMNPQRSTVWXY
NHK	ADEFHIKLMNPQSTVXY
DYK	AFILMSTV
WTK	FILM
TTC	F
Pos2	
NDK	CDEFGHIKLMNQRSVWXY
HDK	CFHIKLMNQRSWXY
MWK	HIKLMNQ
CAC	H
Pos3	
NHK	ADEFHIKLMNPQSTVXY
MHK	HIKLMNPQT
MMC	HNPT
AMC	NT
ACA	T
Pos4	
VNK	ADEGHIKLMNPQRSTV
NHK	ADEFHIKLMNPQSTVXY
RNK	ADEGIKMNRSTV
WYG	LMST
ASC	ST
AGC	S
Pos5	
VNG	AEGKLMPQRTV
DBG	AGLMRSTVW
DYG	ALMSTV
DTG	LMV
RTG	MV
GTA	V
Pos6	
VWK	DEHIKLMNQV
VWC	DHILNV
KTA	LV
GTA	V
Pos7	
TTC	F
Pos8	
MDK	HIKLMNQRS
MWK	HIKLMNQ
AWK	IKMN
ATK	IM
ATA	I
Pos9	
MWG	KLMQ
MTG	LM
CTA	L
Pos10	
RNK	ADEGIKMNRSTV
RBG	AGMRTV

Mena & Daugherty, PEDS 18, 559 (2005)

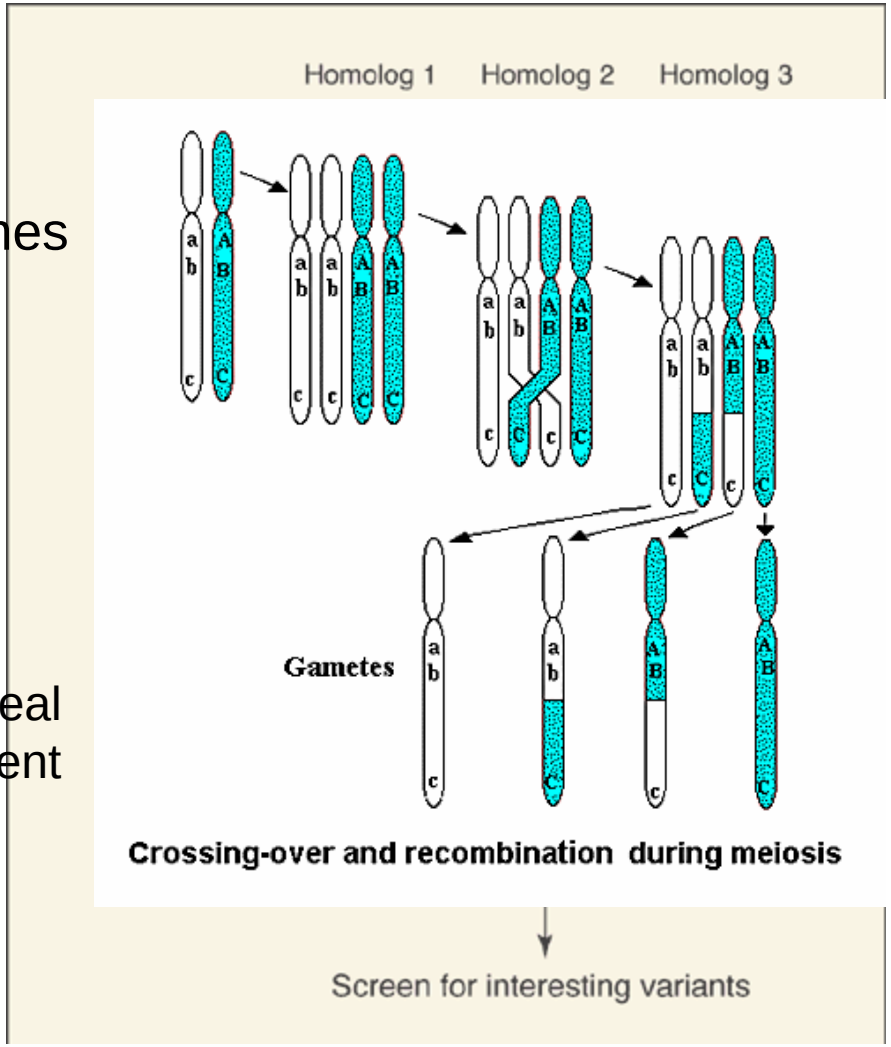
DNA shuffling

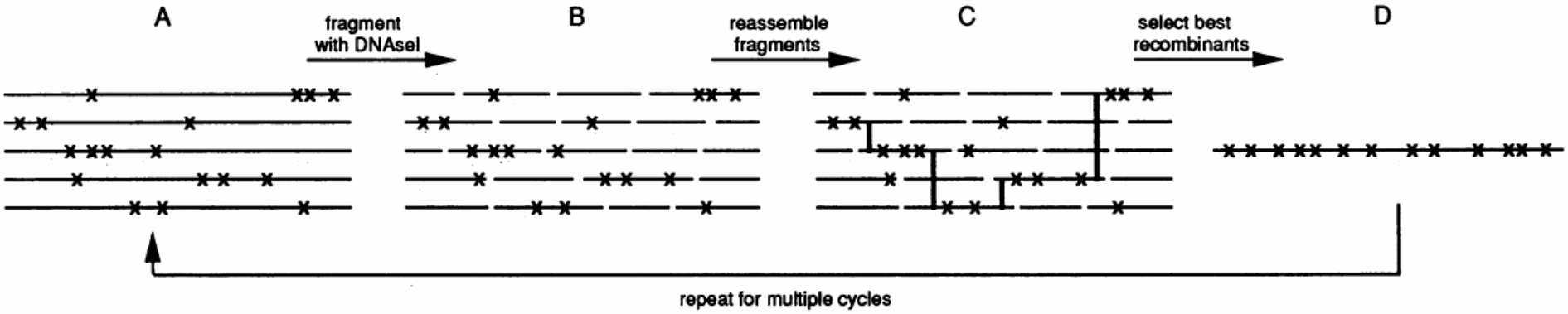
Method of recombining homologous genes

Protocol

- random digest with a nuclease
- annealing to pair up fragments from different parents
- PCR extension to reassemble full gene
- limited to genes of high sequence similarity—otherwise the fragments anneal with other fragments from the same parent

Stemmer, Nature 370, 389 (1994)





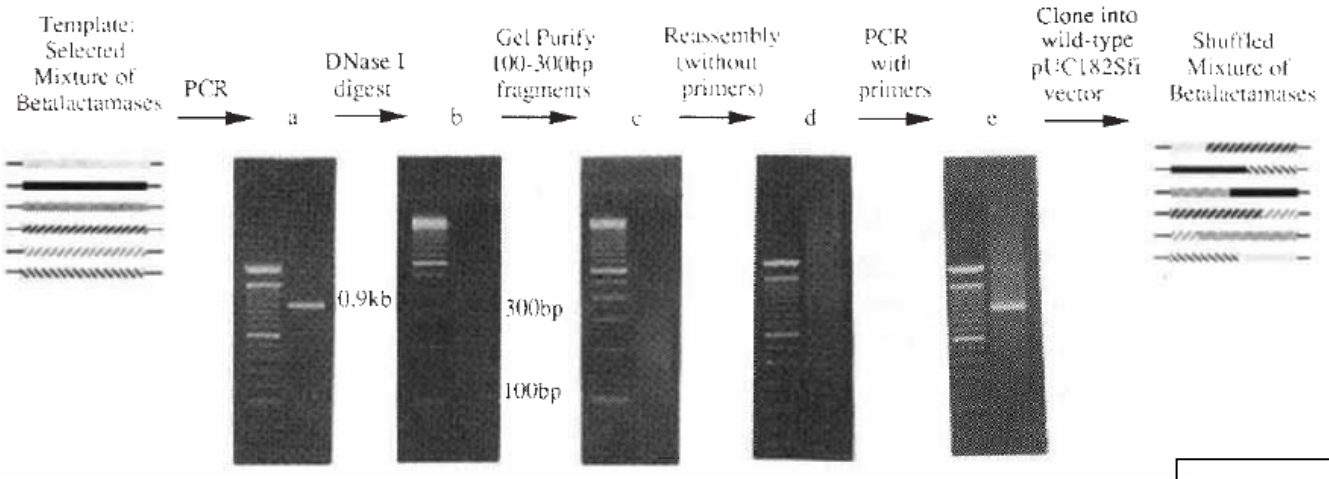
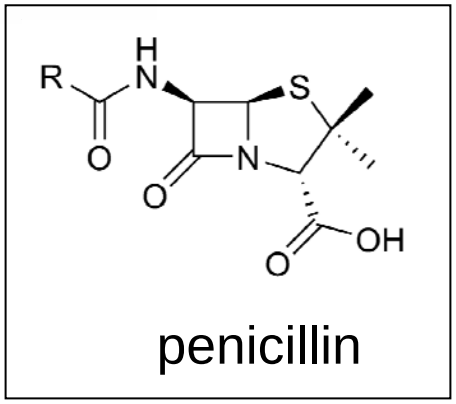


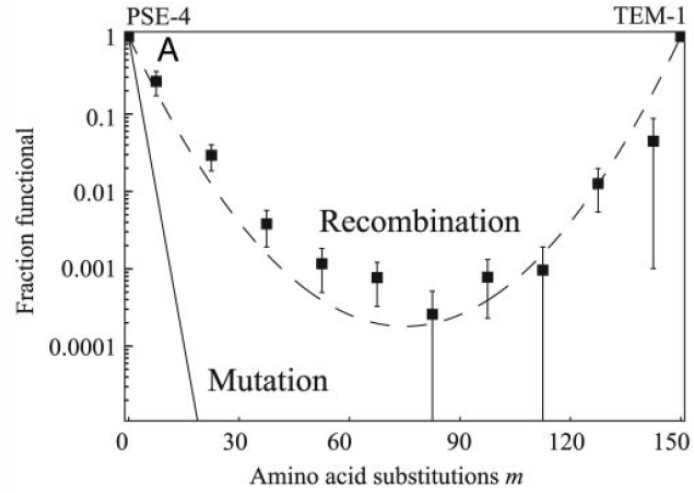
TABLE 1 Characterization of cefotaxime resistance of different combinations of mutations

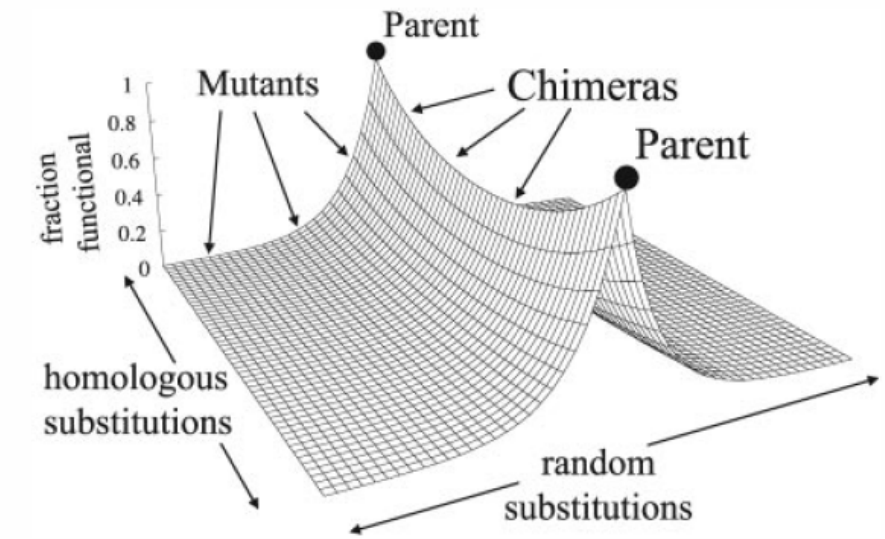
Name	Genotype	MIC	Source of MIC
TEM-1	Wild-type	0.02	This study
—	E104K	0.08	Ref. 9
—	G238S	0.16	Ref. 9
TEM-15	E104K/G238S*	10	This study
TEM-3	E104K/G238S/Q39K	10*	This study
		2–32	Refs 19, 20
ST-4	E104K/G238S/M182T*	10	This study
ST-1	E104K/G238S/M182T/A18V/t3959a/g3713a/g3934a/a3689g*	320	This study
ST-2	E104K/G238S/M182T/A42G/G92S/R241H/t3842c/a3767g*	640	This study
ST-3	E104K/G238S/M182T/A42G/G92S/R241H*	640	This study

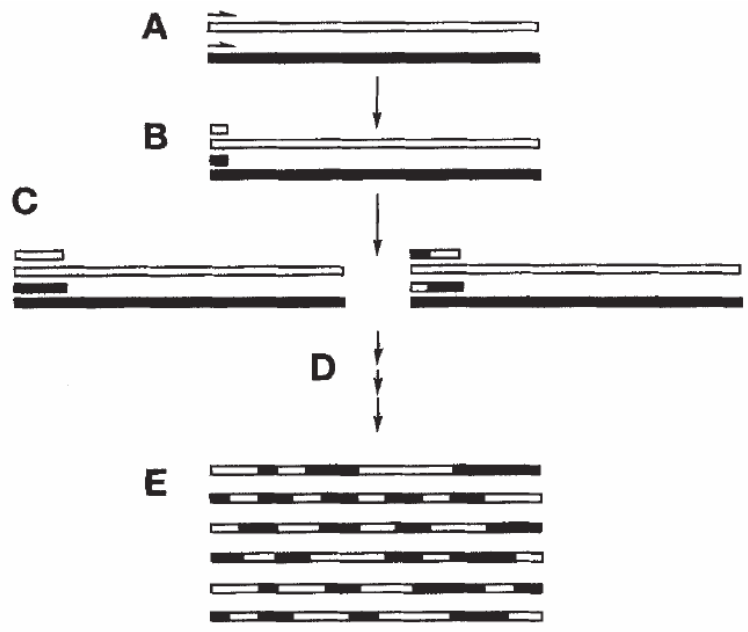


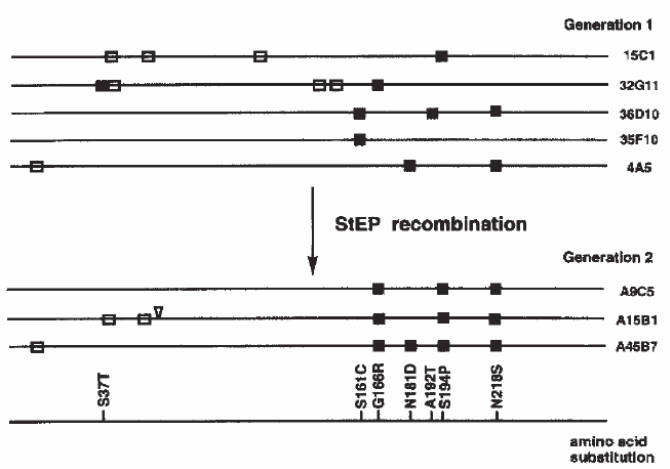
penicillin

DNA shuffling between homologous proteins can rapidly evolve new function





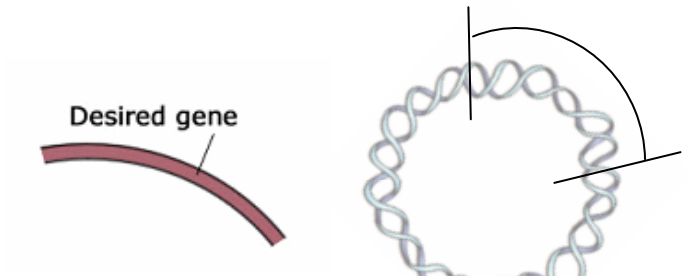


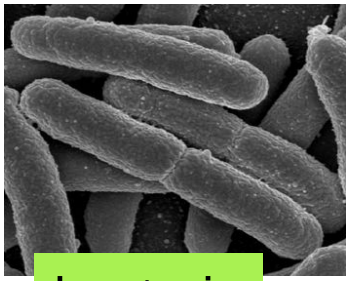


Drummond et al, PNAS 102, 5380 (2005)

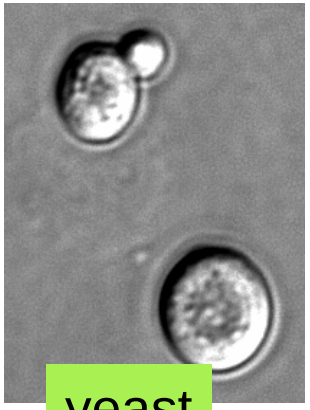
Zhao et al, Nat Biotech 16, 258 (1998)

Summary of Directed Evolution



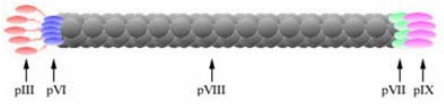


bacteria

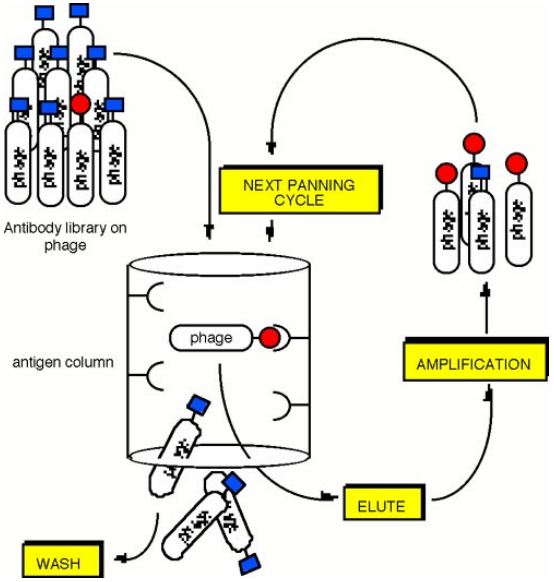


yeast

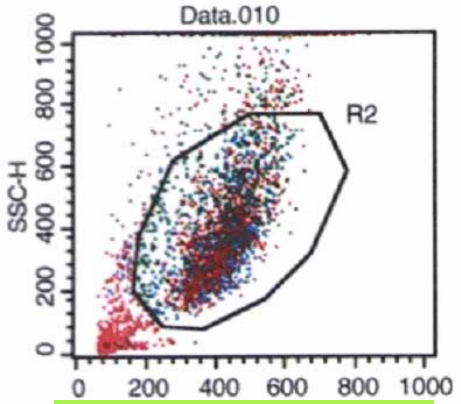
phage



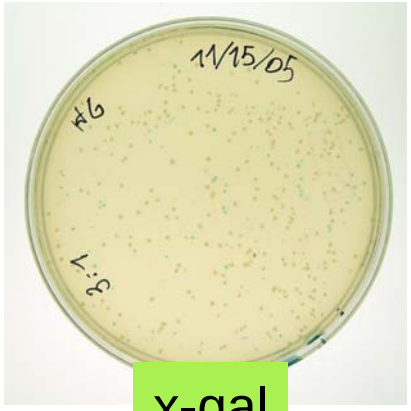
. . .



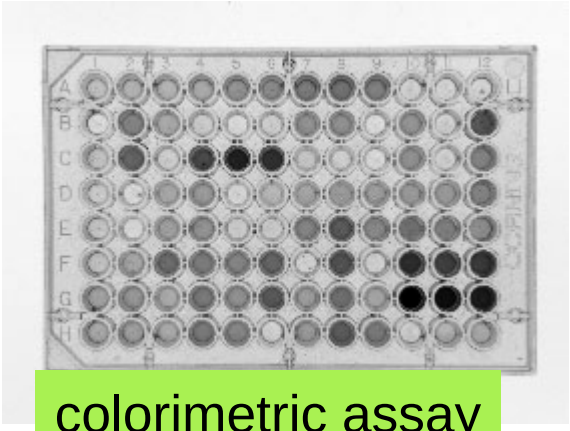
bio-panning



flow cytometry



x-gal



colorimetric assay