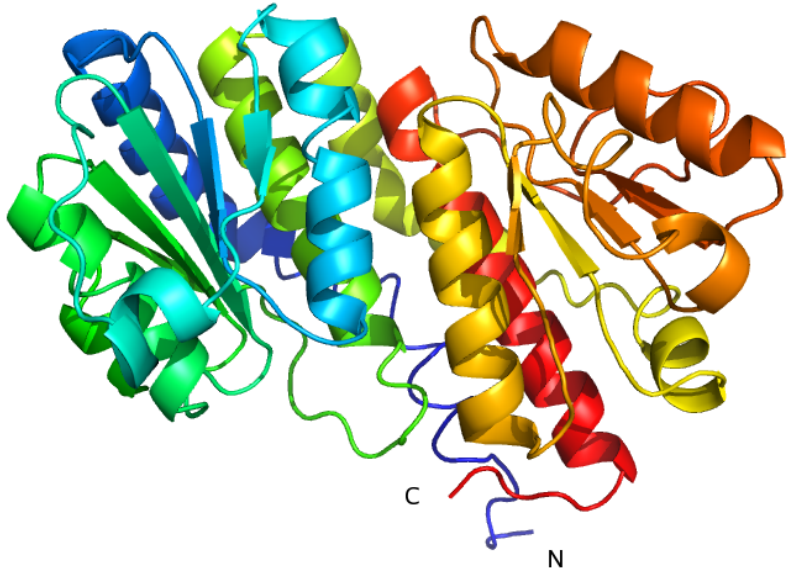
Recent Advances in Directed Protein Evolution



Hui-Wen Shih

October 19, 2011

Lin, H.; Cornish, V. W. *Angew. Chem. Int. Ed.* **2002**, *41*, 4402.

Yuan, L.; Kurek, I.; English, J.; Keenan, R. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 373.

Turner, N. J. *Nat. Chem. Biol.* **2009**, *5*, 567.

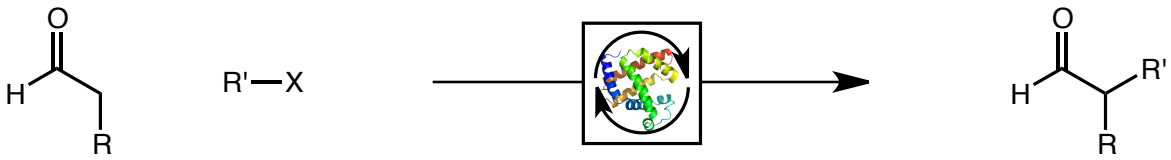
If you could design a protein...

...why would you do it?

...what would it do?

Protein Design

Develop novel catalysts to access unknown transformations



Improve upon existing catalysts

Thermostability

Solvent tolerance

pH tolerance

Increased activity/selectivity

Expand substrate scope

Develop novel biological tools

Probe mechanism and structure

Improve upon rational design

Understand natural protein evolution

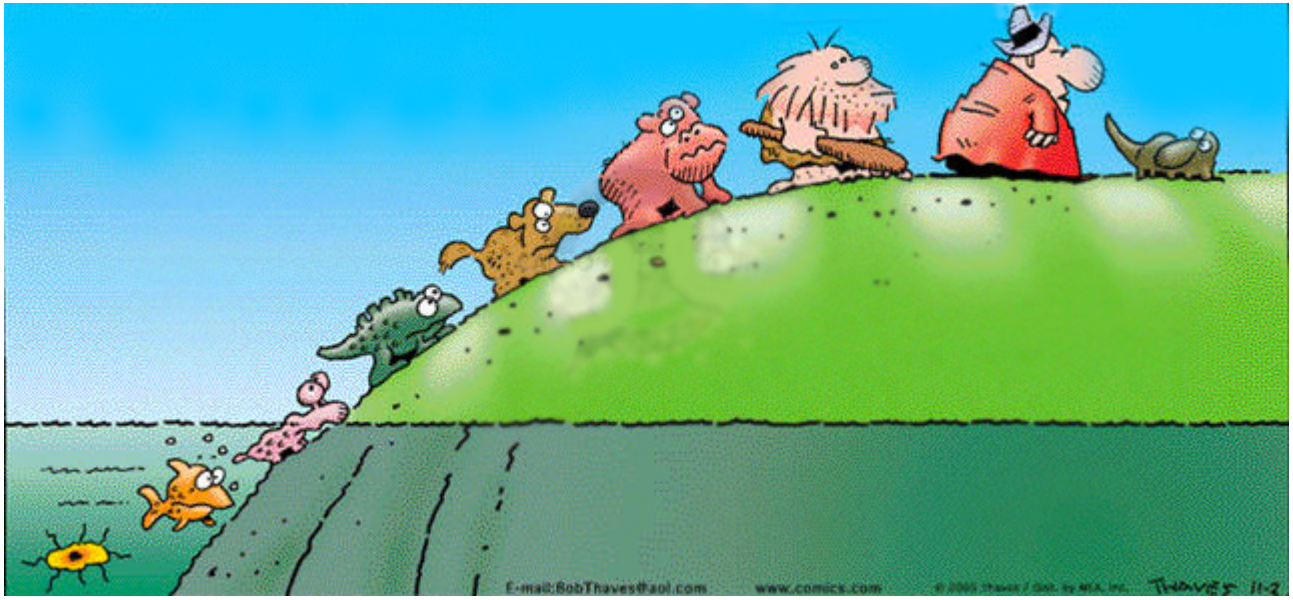
If you could design a protein...

...why would you do it?

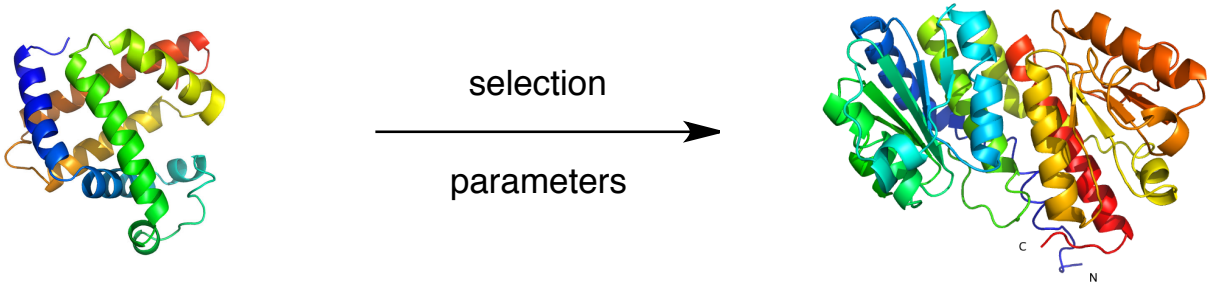
...what would it do?

...how would you do it?

Evolution: A Nature-Inspired Strategy

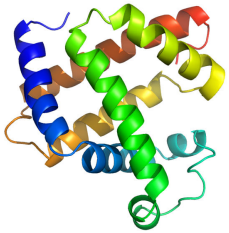


- Directed protein evolution is a protein engineering strategy that harnesses the power of natural selection to evolve proteins with desirable functions.

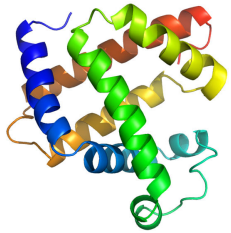


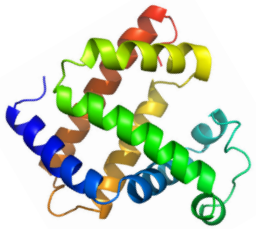
Why Directed Evolution?

- Currently, we cannot accurately tailor proteins for a specific purpose using rational design
- There are too many possibilities to generate and search all of them



~ 20 possible amino acids

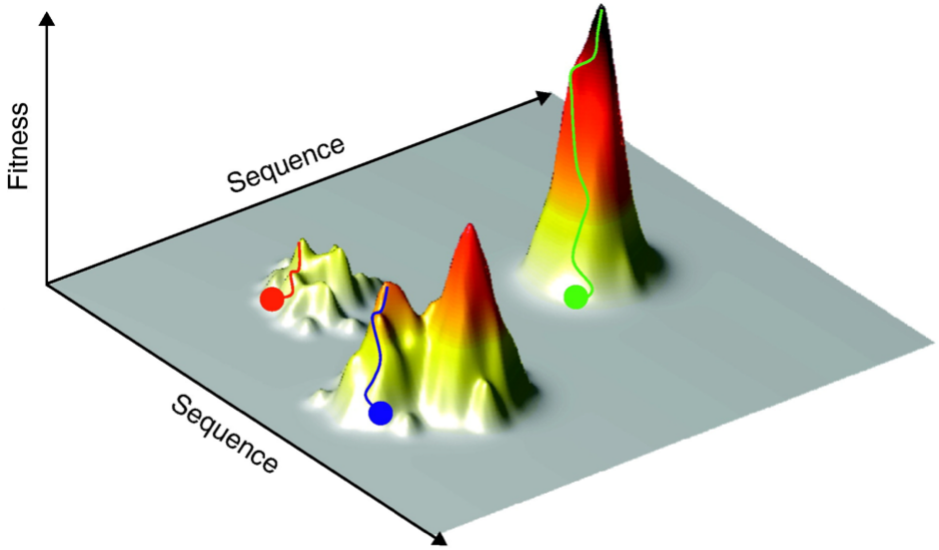




100 amino acids long

20^{100} possibilities

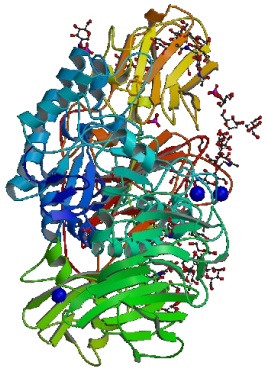
- By imposing selection pressures, iterative rounds of evolution will naturally select for best players



Brustad, E. M.; Arnold, F. H. *Curr. Opin. Chem. Biol.* **2011**, *15*, 201.

A Bit of History

- 1973 - Campbell, Lengyel and Langridge uses directed evolution to discover a novel β-galactosidase

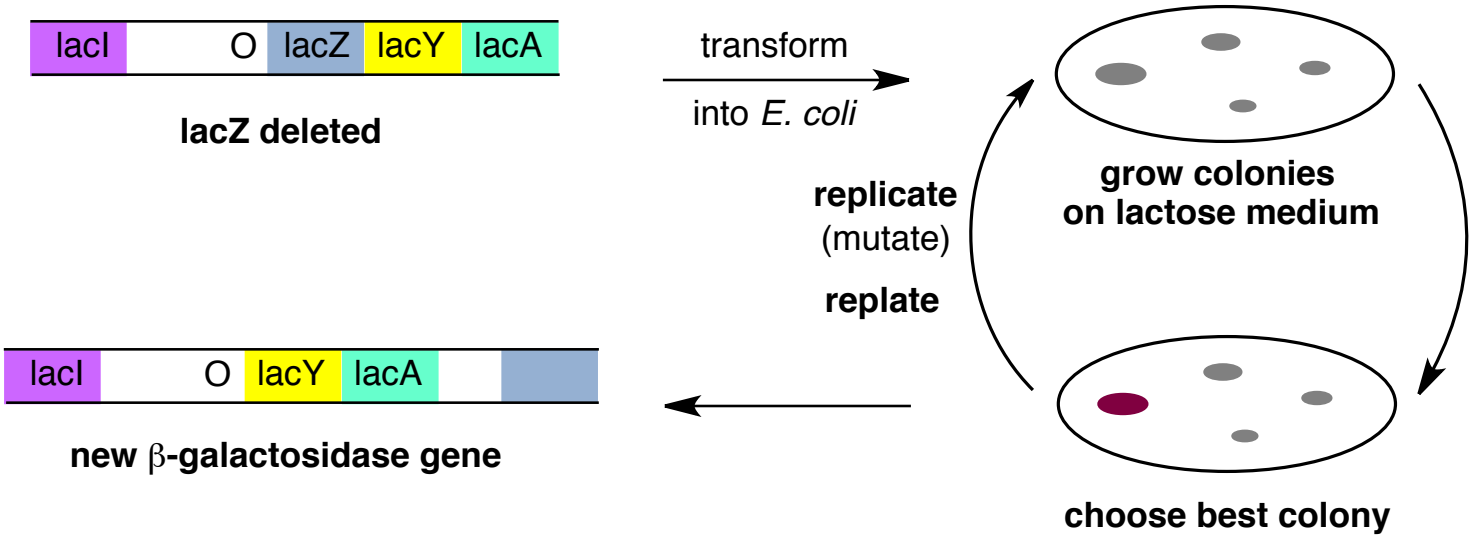


encoded by *lacZ* in *E. coli*

metabolizes lactose

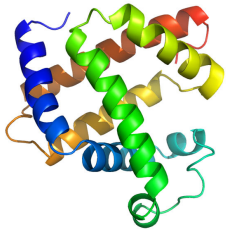
allows for growth on lactose medium

- After *lacZ* deletion, a new β-galactosidase is evolved

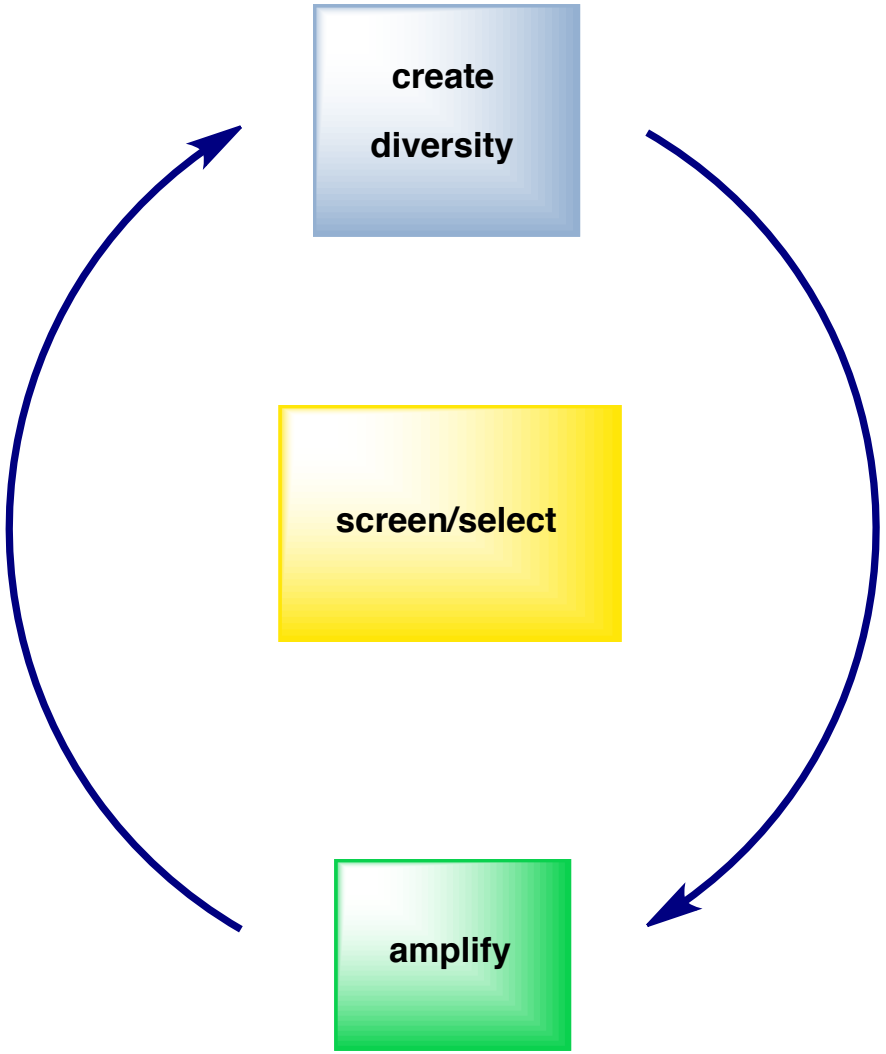


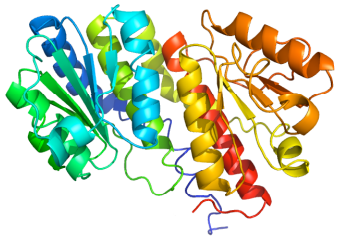
Campbell, J. H.; Lengyel, J. A.; Langridge, J. *Proc. Nat. Acad. Sci. USA* **1973**, *70*, 1841.

How It Works



choose starting molecule



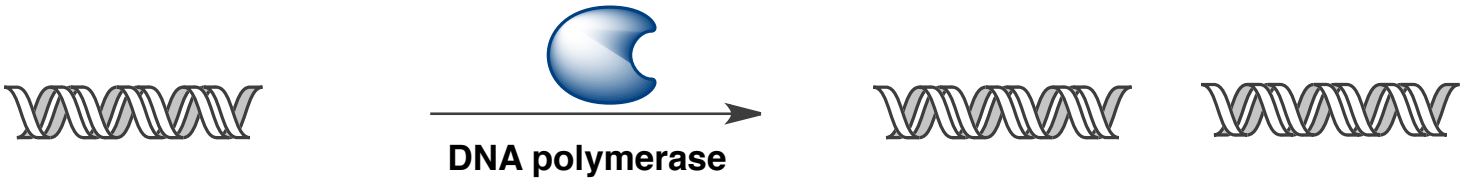


analyze/characterize

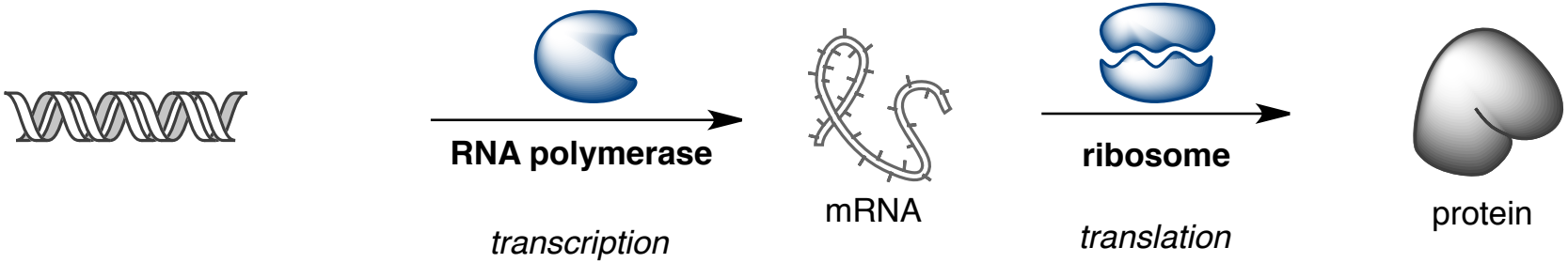
Jackel, C.; Kast, P.; Hilvert, D. *Annu. Rev. Biophys.* **2008**, *37*, 153.

Biochemistry: A Few Notes

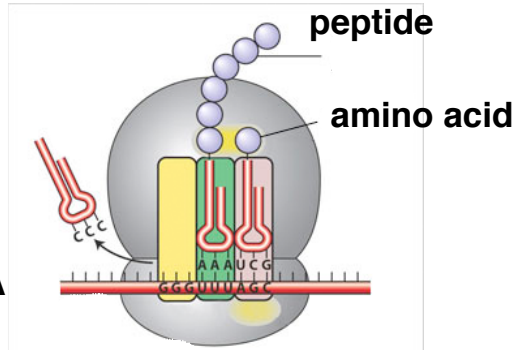
- DNA replication



- Protein biosynthesis



- A closer look at translation



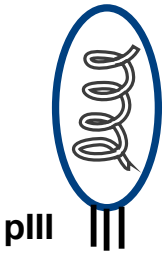
amino acids are encoded by unique triplet codons

tRNAs recognize triplet codons and are attached to corresponding amino acid

Biology: A Few Notes

- What you need to know about phages

bacteriophage



phages display pIII
on the surface

pIII is required
to infect bacteria

- What you need to know about bacteria

bacteria can carry
multiple plasmids

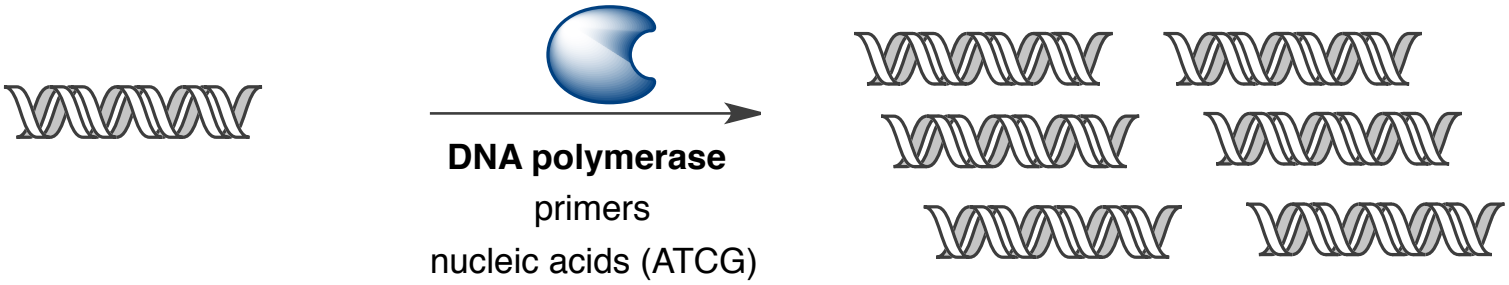
plasmids contain genetic material

plasmids are widely used to introduce
new genes in bacteria (transform)

engineered to confer specific traits

Creating Diversity

- Polymerase Chain Reaction (PCR) allows for rapid genetic amplification



- Mutations during PCR generates a genetic library

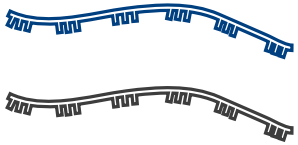
Error-prone PCR

non-ideal conditions cause polymerase to make random mistakes

Site-directed mutagenesis

primers encoding a mismatch or variability at given position lead to change during copy

Saturation mutagenesis

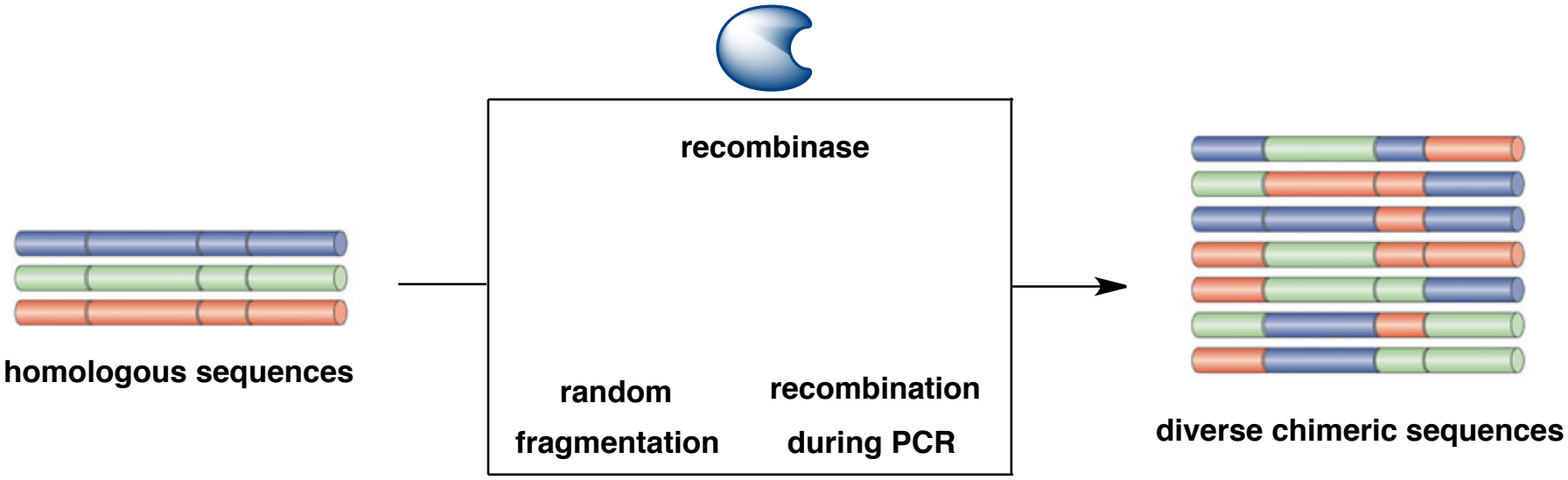


primers encoding mismatchs to generate all possible mutations at site of interest

Yuan, L.; Kurek, I.; English, J.; Keenan, R. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 373.

Creating Diversity

- Recombination/DNA shuffling generates more libraries with more working proteins



- Mutations and recombinations also occur in vivo

Yuan, L.; Kurek, I.; English, J.; Keenan, R. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 373.

Screening and Selecting for Desired Function

- Assays must accomplish the following

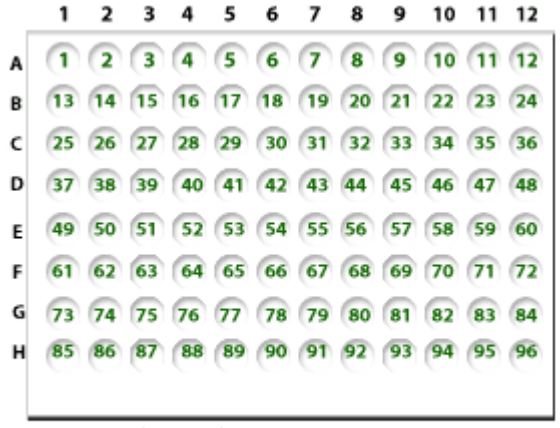
Link genotype with phenotype

Allow for genetic amplification

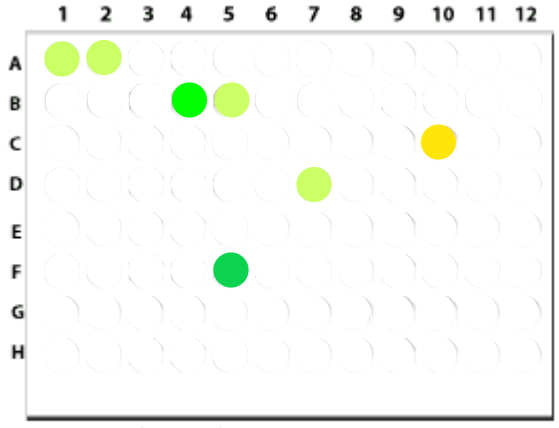
Screen individuals rapidly

Caution: you get what you screen for!

- Microtiter plates



each well contains a different gene mutation



each well contains a corresponding protein

Typical assays

GC/HPLC

UV-Vis

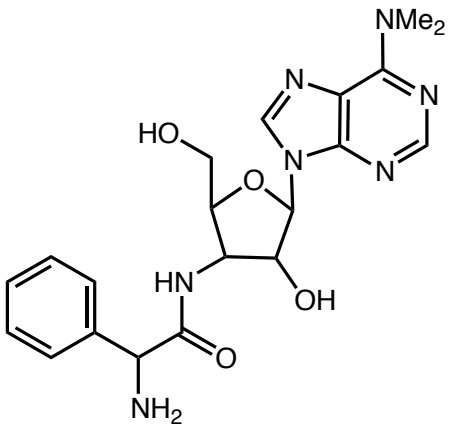
radioactivity

fluorescence

Lin, H.; Cornish, V. W. *Angew. Chem. Int. Ed.* **2002**, *41*, 4402.

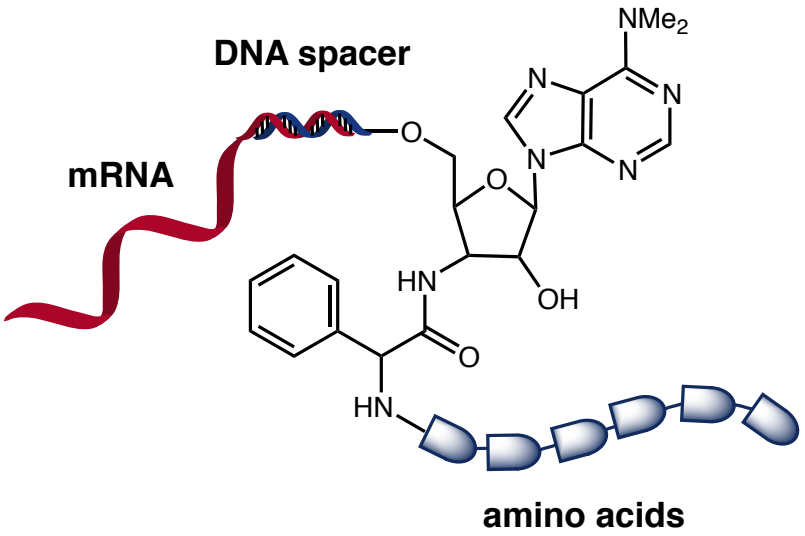
Screening and Selecting for Desired Function

- mRNA display physically links phenotype and genotype using puromycin



puromycin mimics the aminoacyl end of tRNA

forms a covalent linkage between mRNA and peptide



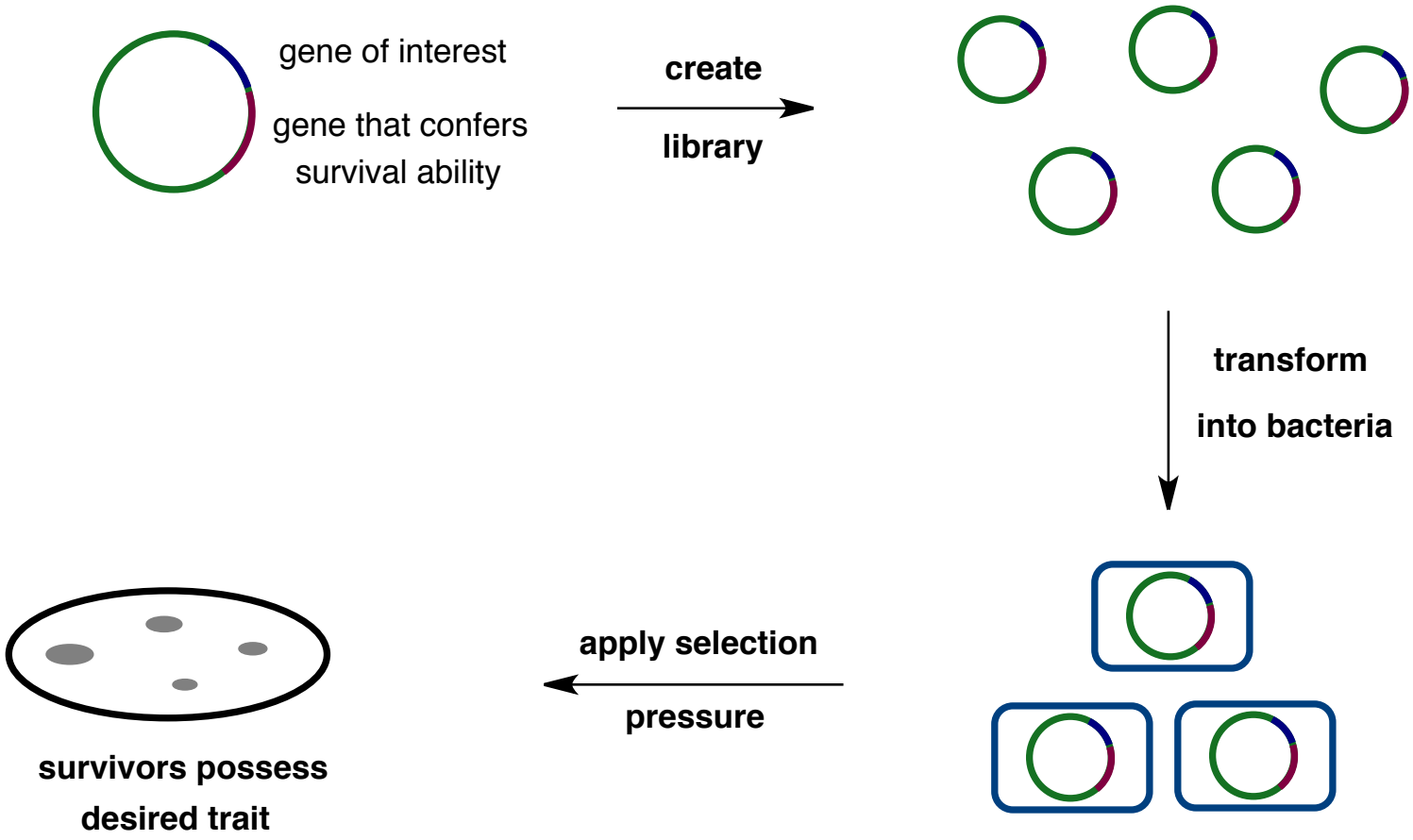
During translation, ribosome pauses at the DNA spacer, allowing puromycin to react with peptide chain

Lin, H.; Cornish, V. W. *Angew. Chem. Int. Ed.* **2002**, *41*, 4402.

Screening and Selecting for Desired Function

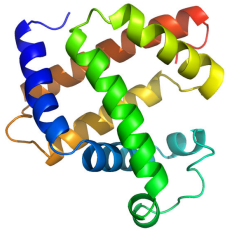
- In vivo selection - desired function is linked to survival

eg. antibiotic resistance, replication ability, metabolism ability...



Lin, H.; Cornish, V. W. *Angew. Chem. Int. Ed.* **2002**, *41*, 4402.

How It Works



choose starting molecule

In vitro

In vivo

error-prone PCR
DNA shuffling
degenerate codons

create diversity

mutator strains
error-prone polymerase
recombination

in vitro
gene expression

transcription
translation

display/sorting
techniques
IVC

screen/select

complementation
transcription activation
detoxification

mRNA/ribosome
display, FACS

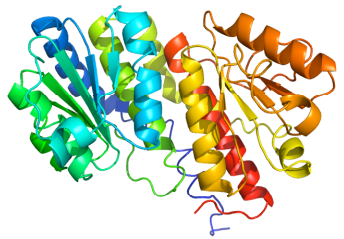
Phage/cell surface
display, FACS

growth of
survivors

RT-PCR
PCR

amplify

replication

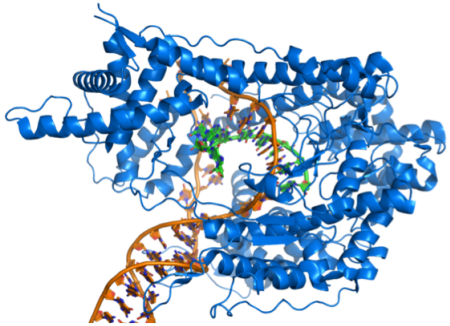


analyze/characterize

Jackel, C.; Kast, P.; Hilvert, D. *Annu. Rev. Biophys.* **2008**, *37*, 153.

PACE: A System for Continuous Directed Evolution

- Design T7 RNA polymerase to recognize a new promoter sequence



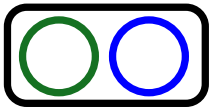
Bacteriophage RNA polymerase

Widely used to transcribe RNA *in vitro* and *in vivo*

Very specific for promoter sequence

- Strategy: link phage survival (pIII expression) with activity of T7 RNA polymerase

host cells



bacteriophage

pIII

mutagensis plasmid

supresses proofreading
enhances error-prone bypass

evolving gene

gene encodes desired activity that induces pIII production

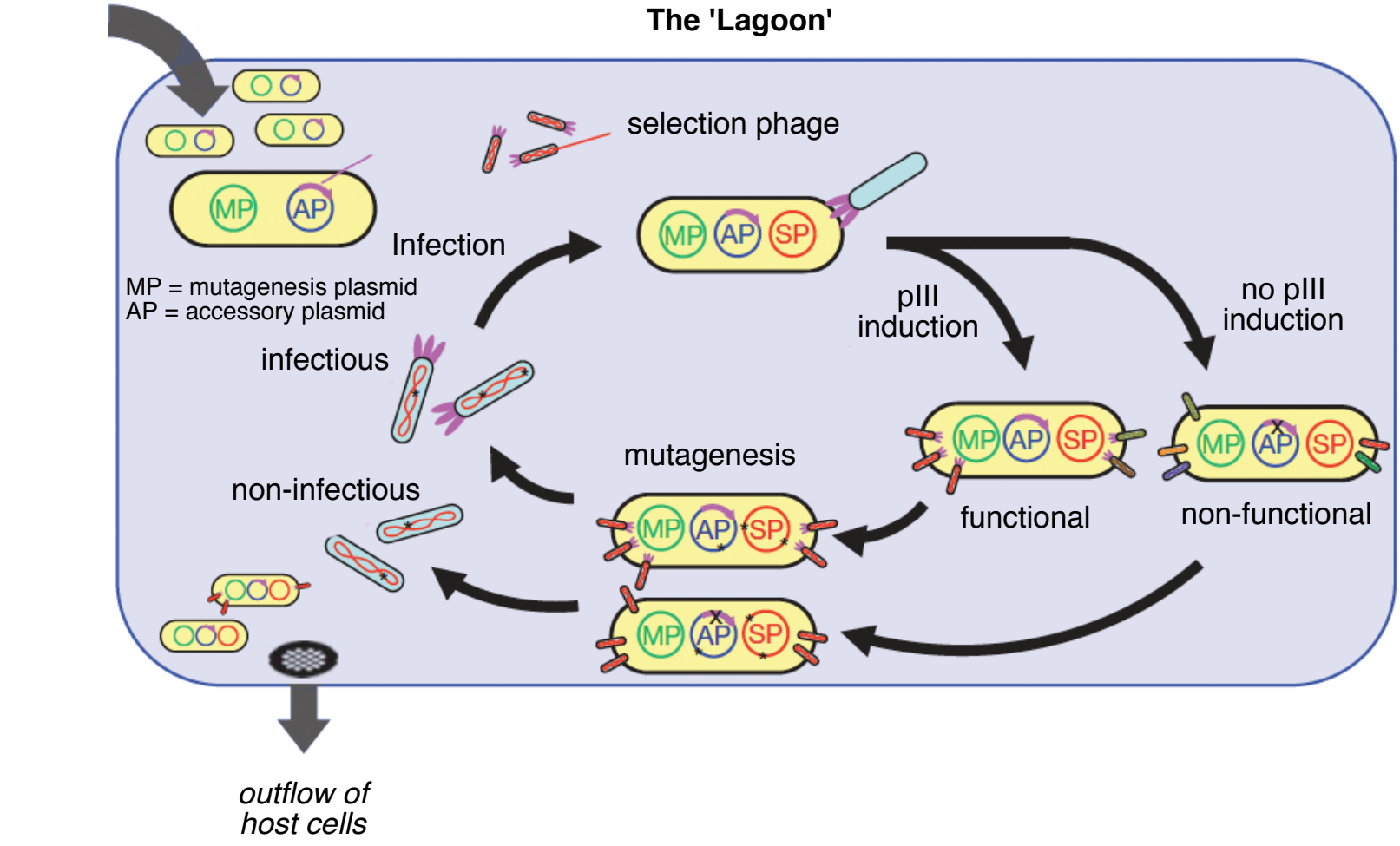
accessory plasmid

contains gene for pIII production

Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472*, 499.

PACE: A System for Continuous Directed Evolution

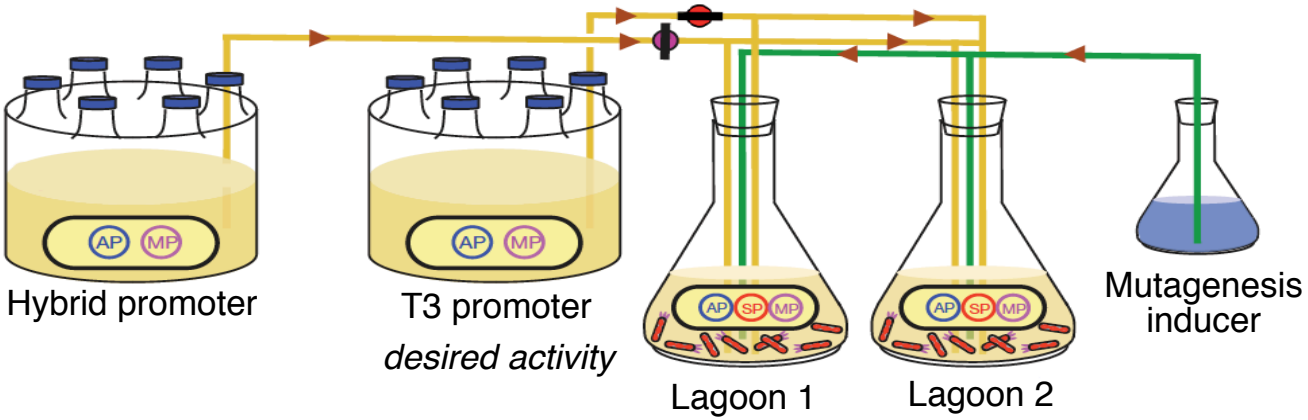
- Host cells continuously flow through lagoon faster than they can replicate

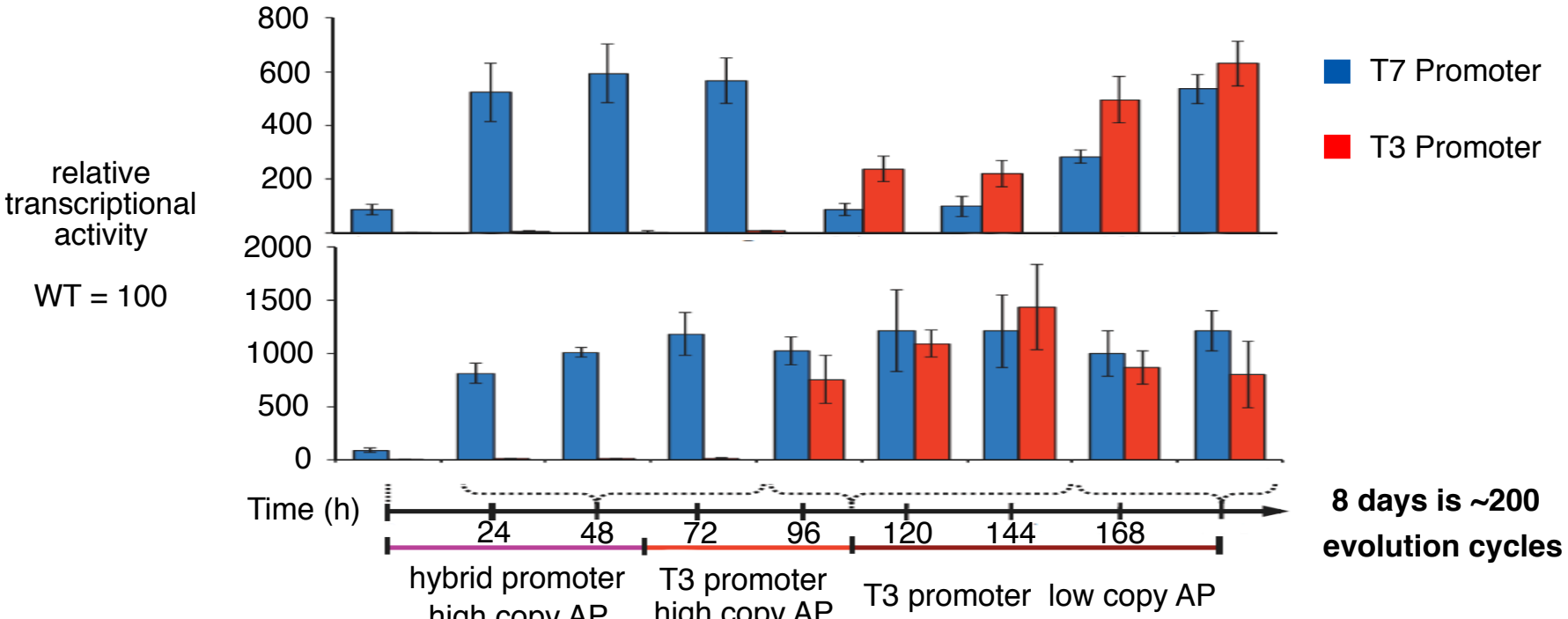


Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472*, 499.

PACE: A System for Continuous Directed Evolution

- Selection pressure becomes more stringent over time

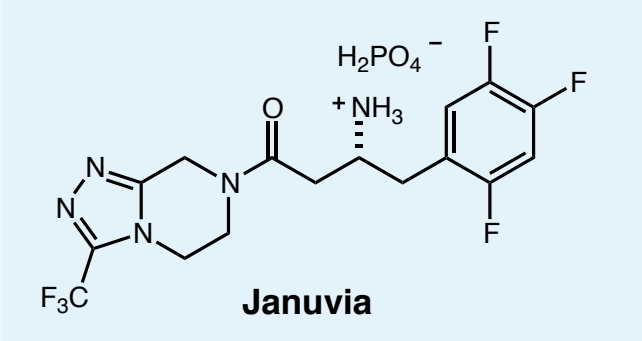




Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472*, 499.

Towards the Synthesis of Sitagliptin

- Januvia (sitagliptin phosphate) is a blockbuster drug

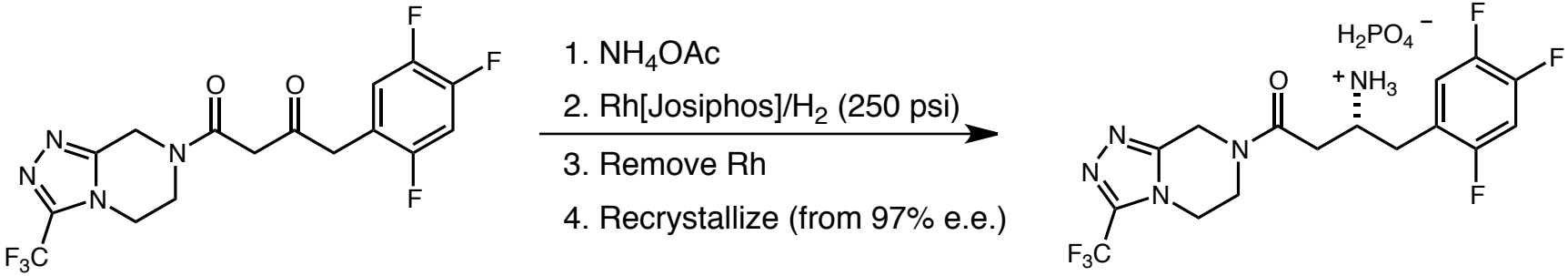


Januvia

Treatment of diabetes
DPP-IV inhibitor

#24 brand name drug in 2010
$1,294M

- Previous manufacture route to set amine stereocenter

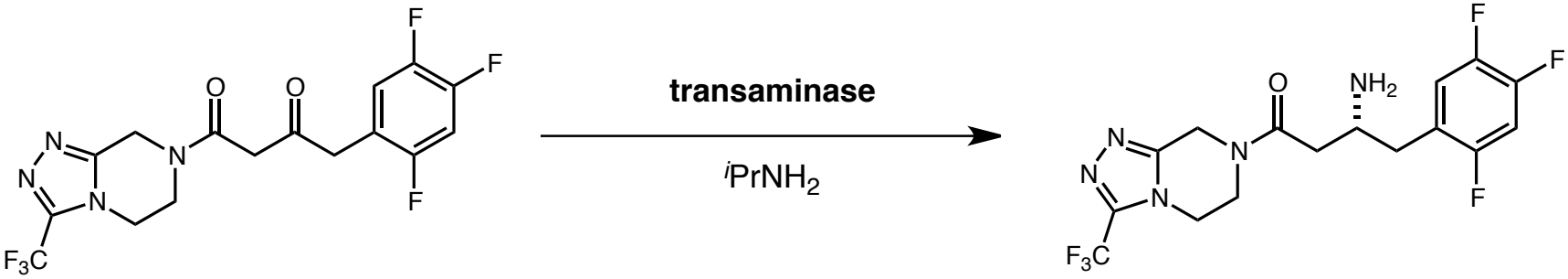


Njardson Group, http://cbc.arizon.edu/njardarson/group/top-pharmaceuticals-poster

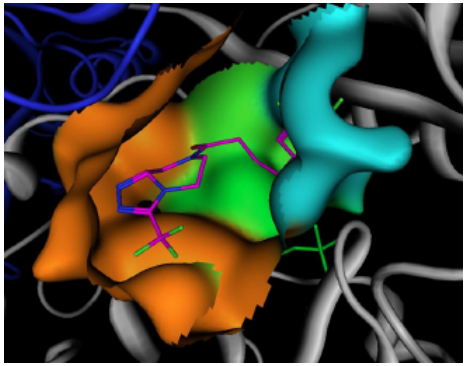
Savile, C. K. *et. al. Science* **2010**, *329*, 305.

Towards the Synthesis of Sitagliptin

- Desired reactivity



- A computational approach in silico rational design



Transaminase ATA-117 (dimer)

Small and large binding pockets identified

Docking studies shows binding pocket is too small

Residues in active site identified for mutation

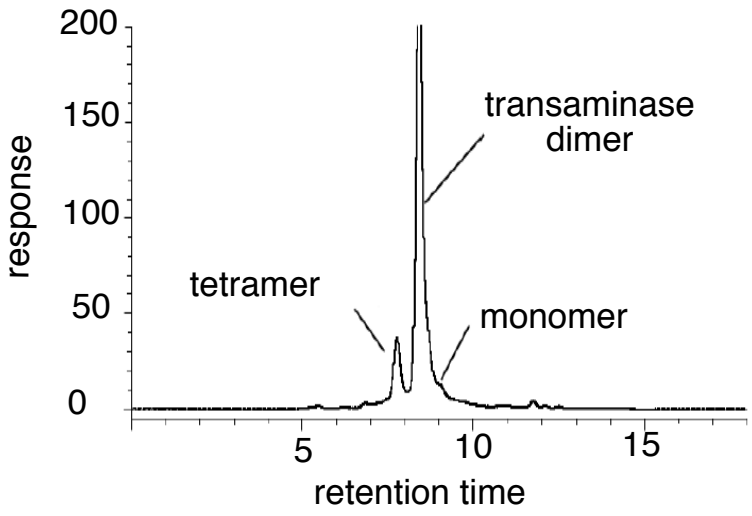
Savile, C. K. *et. al. Science* **2010**, *329*, 305.

Towards the Synthesis of Sitagliptin

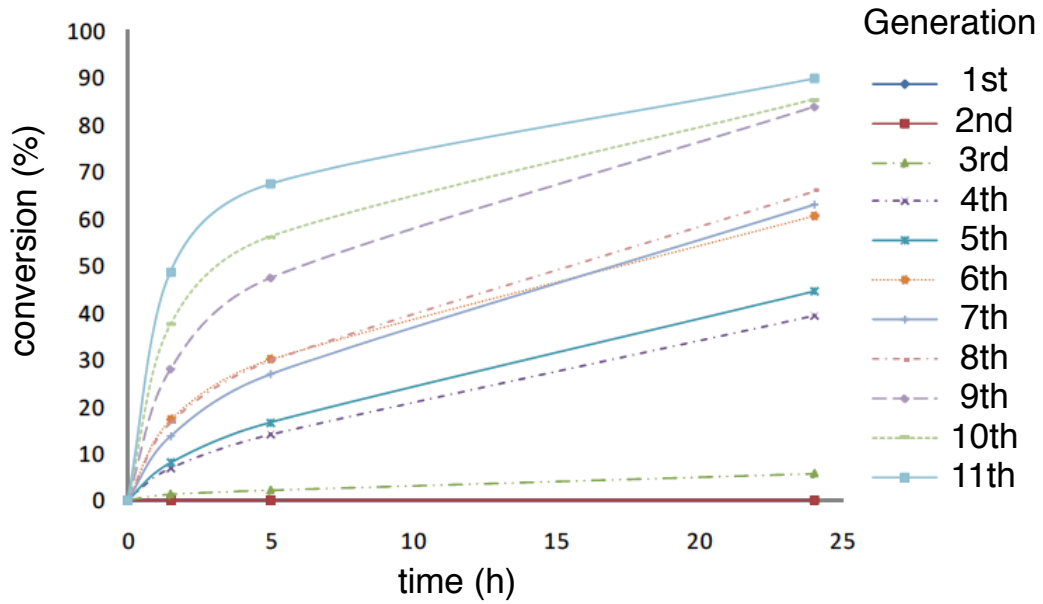
- To qualify for use in process, very specific parameters had to be met

Generation	a-d	1-5	6-11
	enlarge binding pocket	**optimize towards process** [^{i}PrNH$_2$] RT to 45 ^{o}C organic solvent pH	**optimize towards process** organic solvent acetone tolerance

- Protein must be expressed in E. coli
- Protein must be easy to purify



- Comparison of top variant under process-like conditions

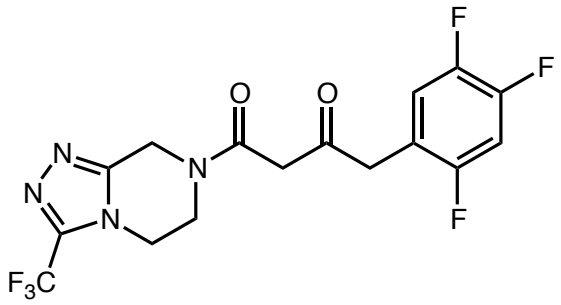


Savile, C. K. *et. al. Science* **2010**, *329*, 305.

Towards the Synthesis of Sitagliptin

- Improvements on original process route

1. NH_4OAc
2. Rh[Josiphos]/H_2 (250 psi)
3. Remove Rh
4. Recrystallize (from 97% e.e.)



transaminase

$^{i}PrNH_2$

92%, >99.95% ee

Improvements

10-13% yield increase

53% productivity increase (kg/l/day)

19% total waste reduction

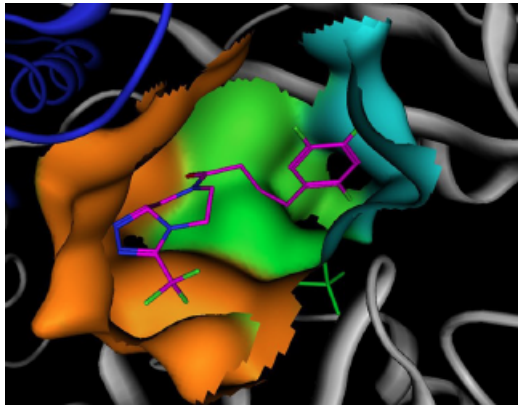
no heavy metals

no high-pressure equipment

Savile, C. K. *et. al. Science* **2010**, *329*, 305.

Towards the Synthesis of Sitagliptin

- Mutations reflect contribution from in silico modeling



Final catalyst has 27 mutations

10 mutations: noncatalytic AA interacting with substrate

4 mutations: design of small binding pocket

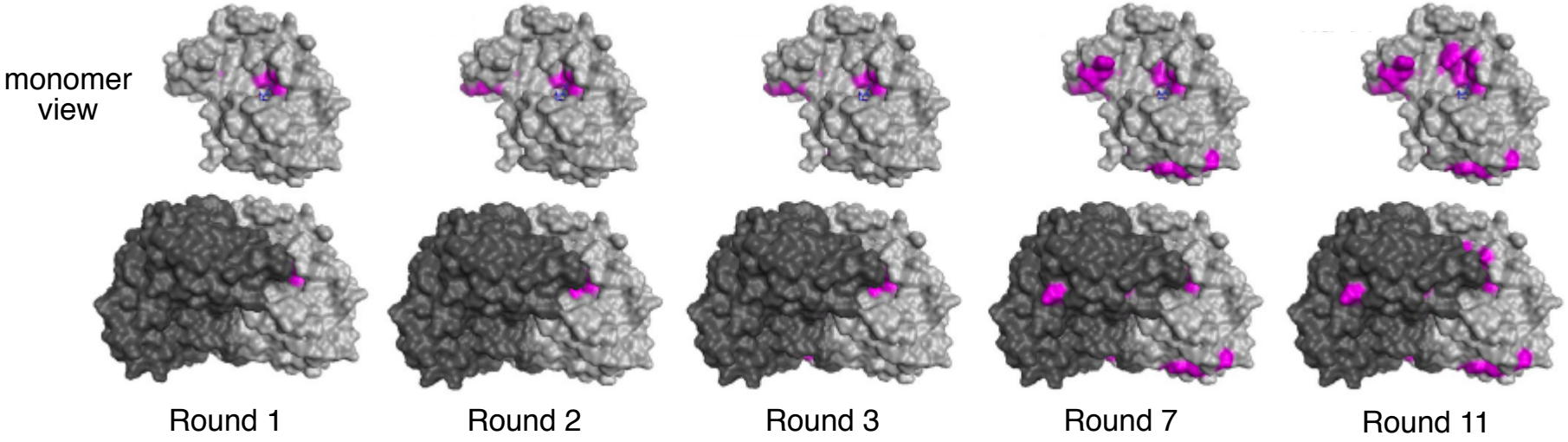
5 mutations: evolution of large binding pocket

1 mutation: evolution of small binding pocket

10 mutations: homology libraries

5 random mutations

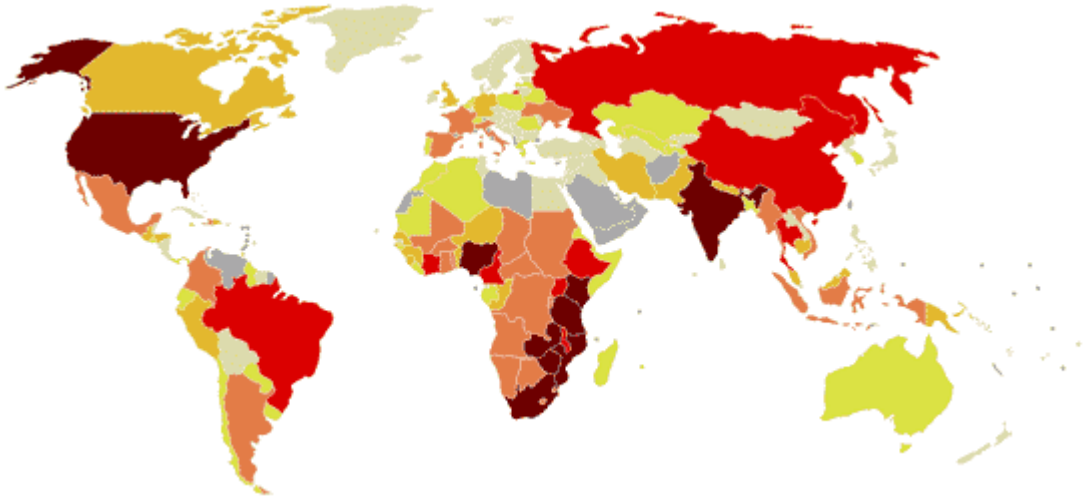
- Later improvements modify the dimer interfacial region, persumably leading to more stable active dimer



Savile, C. K. *et. al. Science* **2010**, *329*, 305.

Potential Gene Therapies

- The World Health Organization: HIV is a pandemic



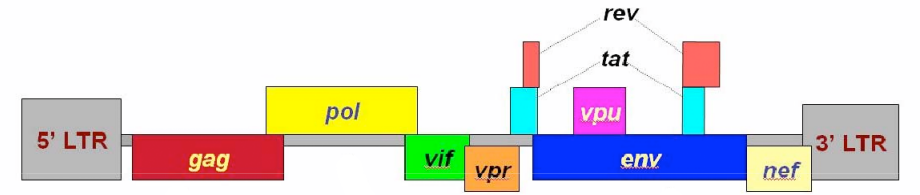
- 0.6% of the world population is infected
- 13 of the best 200 drugs (2010) are for HIV therapies
- Combined sales of $4.7 billion
- No cure

Njardson Group, http://cbc.arizon.edu/njardarson/group/top-pharmaceuticals-poster

Sarkar, I.; Hauber, I.; Hauber, J.; Buchholz, F. *Science* **2007**, *316*, 1912.

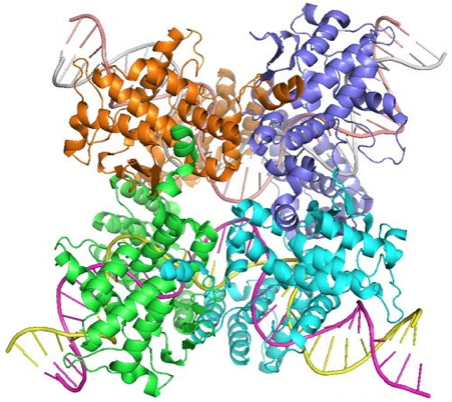
Potential Gene Therapies

- Evolution of a HIV-1 DNA exicision enzyme



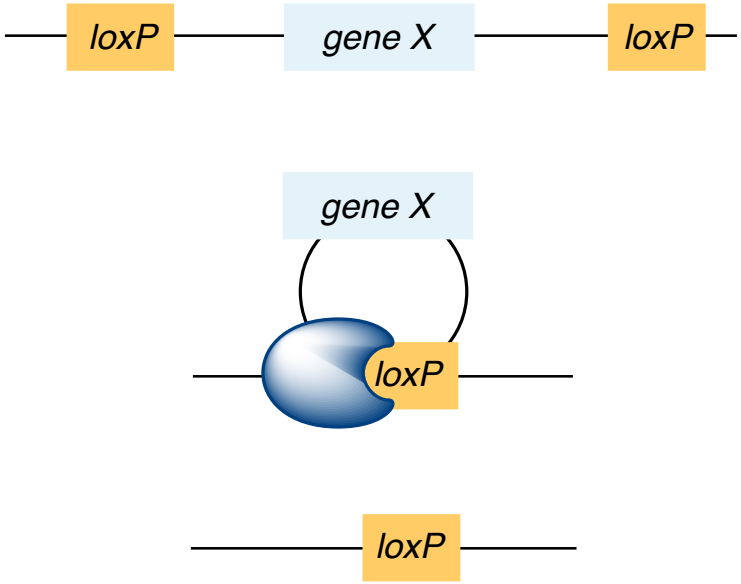
HIV-1 proviral DNA is flanked by long terminal repeats (LTRs)

- Cre recombinase is a bacteriophage topoisomerase



recognizes 34 bp loxP site

removes sequences flanked by two loxP sites



Sarkar, I.; Hauber, I.; Hauber, J.; Buchholz, F. *Science* **2007**, *316*, 1912.

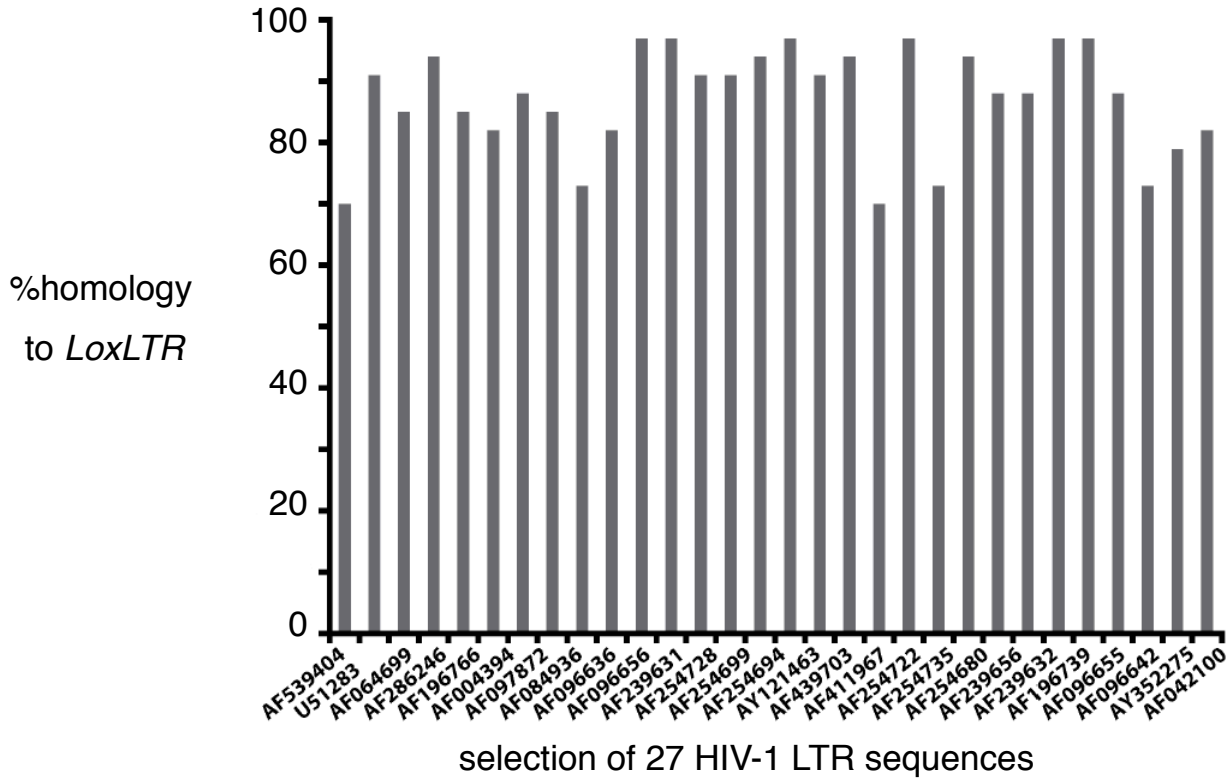
Potential Gene Therapies

- loxLTR sequence chosen as recognition target has 50% sequence similarity with *loxP*

ATAACTTCGTATA ATGTATGC TATACGAAGTTAT *LoxP*

ACAACATCCTATT ACACCCTA TATGCCAACATGG *LoxLTR*

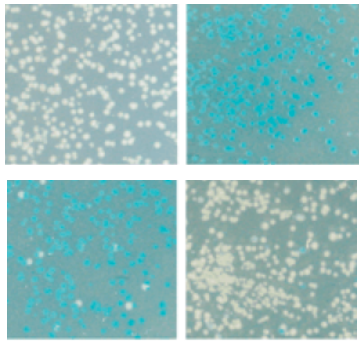
- loxLTR sequence shows high sequence similarity to other HIV-1 LTRs



Sarkar, I.; Hauber, I.; Hauber, J.; Buchholz, F. *Science* **2007**, *316*, 1912.

Potential Gene Therapies

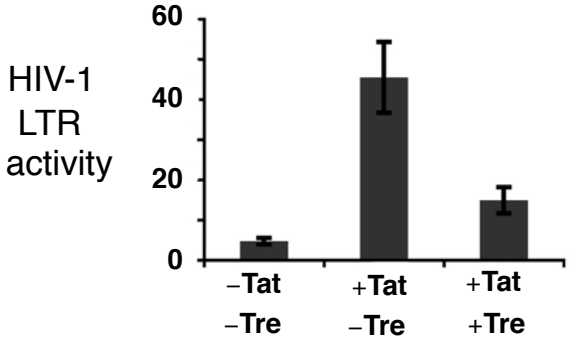
- After 126 rounds of evolution, Tre is identified as the most active recombinase with 19 amino acid changes



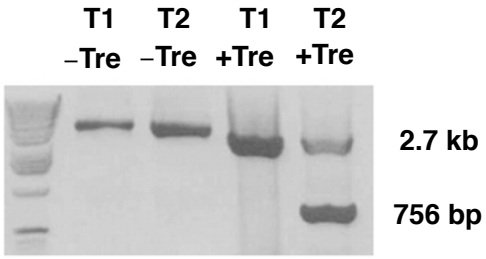
Blue cells = no recombination

White cells = recombination

- Tre is inserted into mammalian HeLa cells with HIV promoter (Tat)



activity decreases in presence of Tre



Tre works only if two loxLTR sites are present

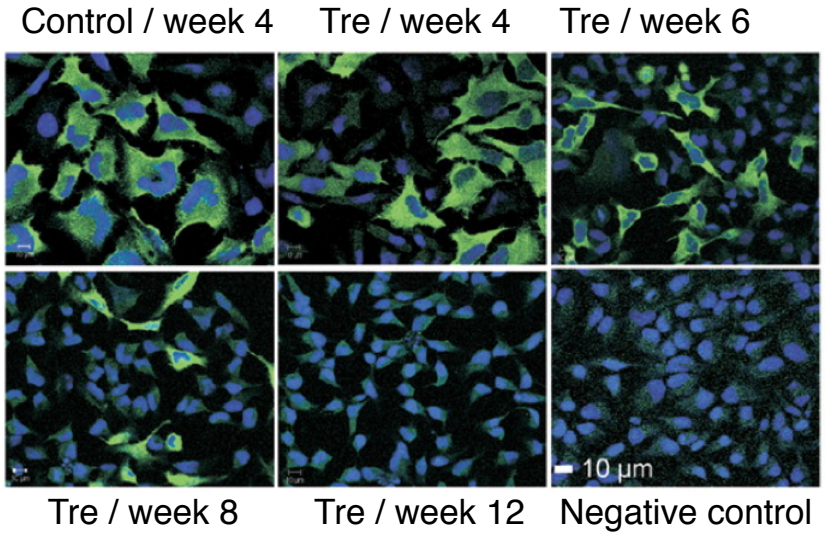
T1 = 1 loxLTR site

T2 = 2 lox LTR sites

Sarkar, I.; Hauber, I.; Hauber, J.; Buchholz, F. *Science* **2007**, *316*, 1912.

Potential Gene Therapies

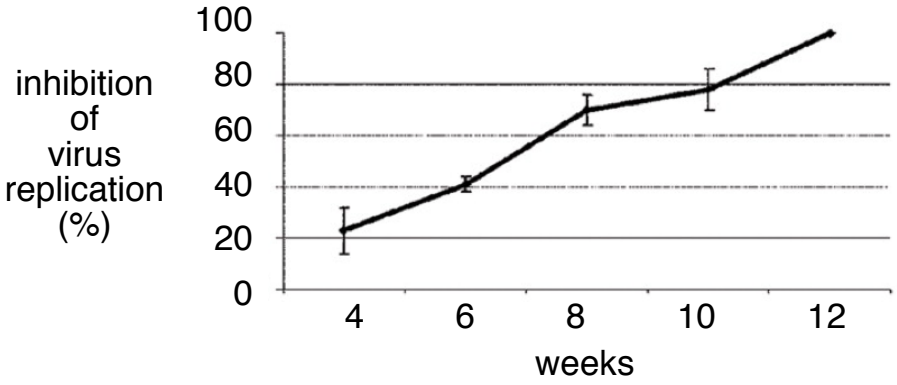
- Tre is inserted into HeLa cells and cells are infected with HIV



Green = gag expressing cells (infected)

Blue = cell nucleus

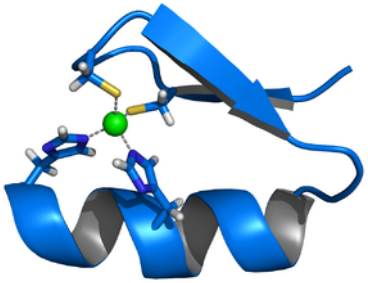
- Provirus is deleted from infected cells without obvious cytotoxicity



Sarkar, I.; Hauber, I.; Hauber, J.; Buchholz, F. *Science* **2007**, *316*, 1912.

Evolving Activity: Zero to Hero

- You get what you screen for!

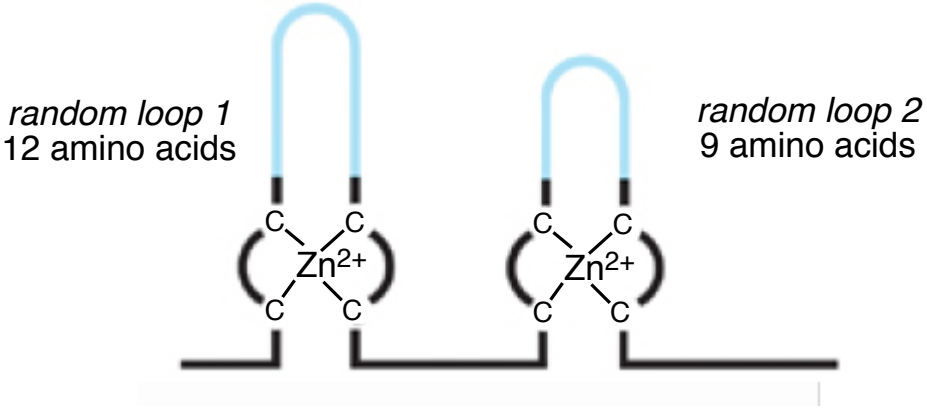


Zinc fingers bind DNA

evolve enzymatic activity

RNA ligase

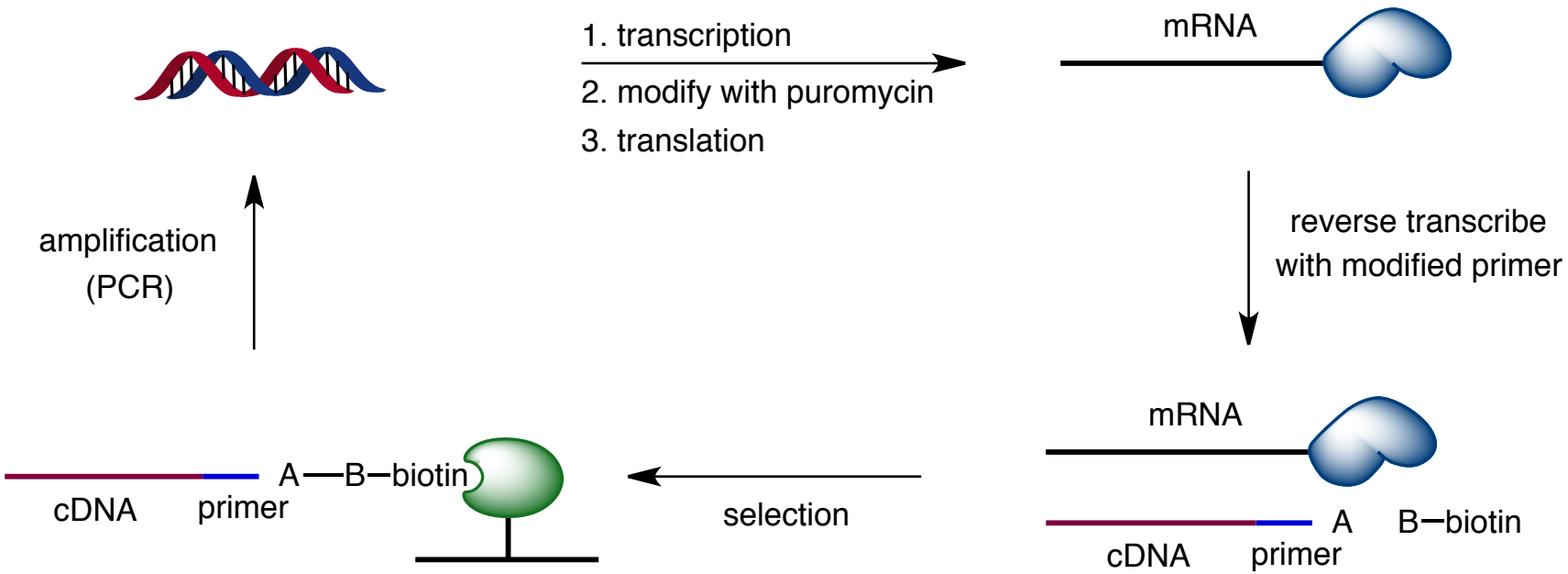
- Library is generated by mutating two random loop regions



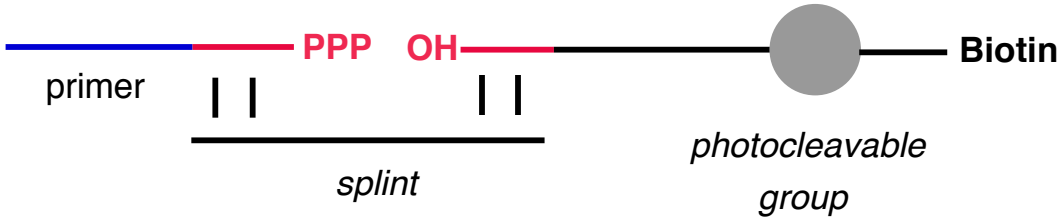
Seelig, B.; Szostak, J. W. *Nature* **2007**, *448*, 828.

Evolving Activity: Zero to Hero

■ Evolution



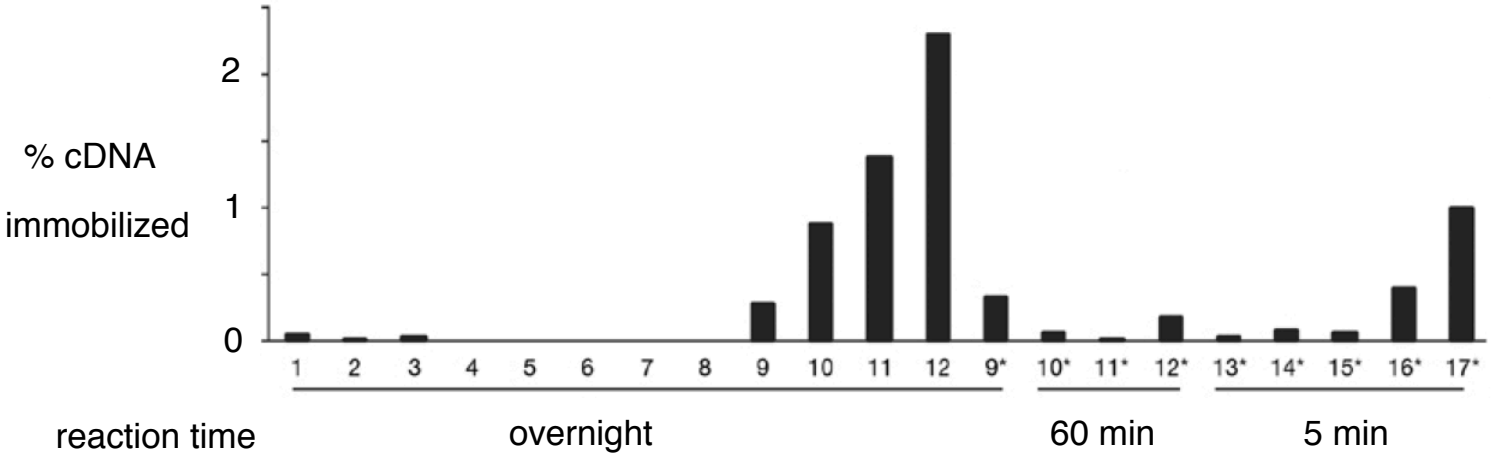
■ Assay



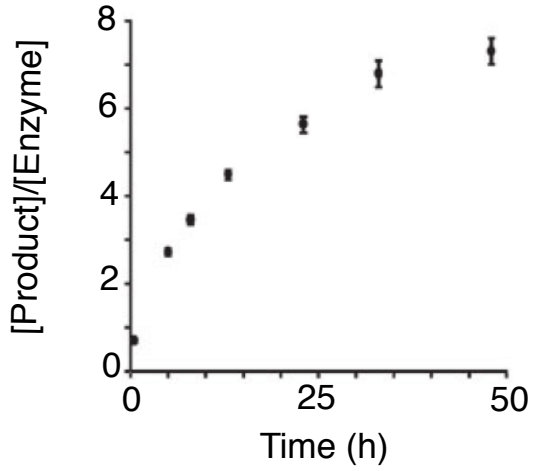
Seelig, B.; Szostak, J. W. *Nature* **2007**, *448*, 828.

Evolving Activity: Zero to Hero

- After 17 rounds of mutation, competent RNA ligase was evolved



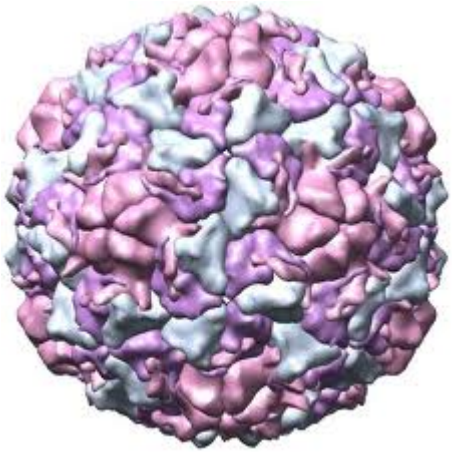
- Enzyme k_{obs} is 10^6 times faster than uncatalyzed reaction



Seelig, B.; Szostak, J. W. *Nature* **2007**, *448*, 828.

Modifying Nanoparticles

- Capsids readily undergo modification for a variety of applications



gene therapy

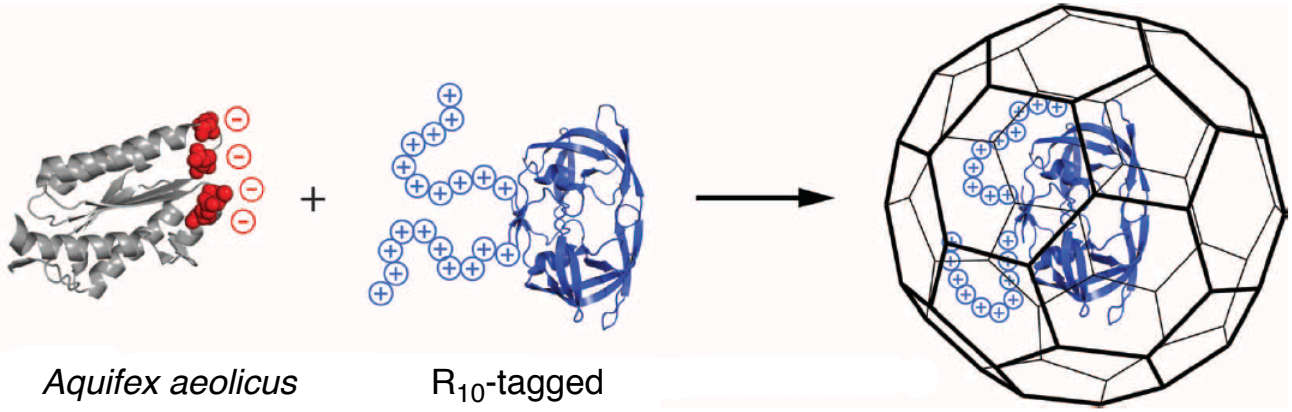
drug delivery

bioimaging

catalysis

controlled synthesis

- Protein capsid can be engineered to contain toxic materials



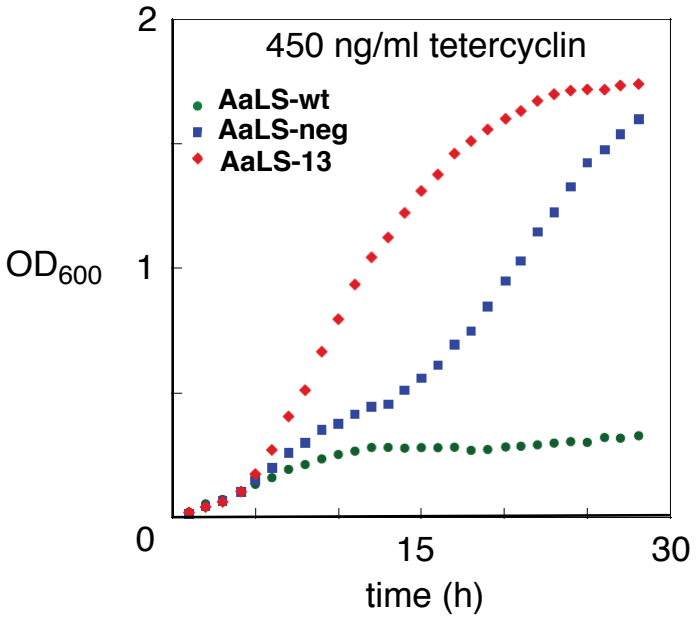
Aquifex aeolicus
Lumazine synthase
AaLS

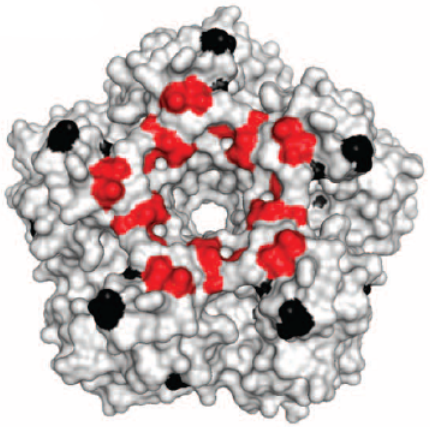
R_{10}-tagged
RNA protease

Worsdorfer, B.; Woycechowsky, K. J.; Hivert, D. *Science*, **2011**, *311*, 589.

Modifying Nanoparticles

- HIV expression is promoted by tetracycline

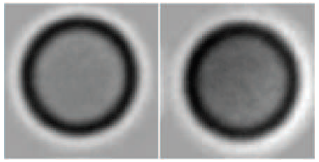




AaLS-13 has 7 point mutations
No exterior modifications
Addition of 3 anionic residues
Loss of three cationic residues

- On average, each capsid can bind seven HIV protease-R_{10} dimers
- 5-10 fold improvement results in cell survival even at 1400 ng/ml tetracyclin

–HIV prot.-R_{10} +HIV prot.-R_{10}



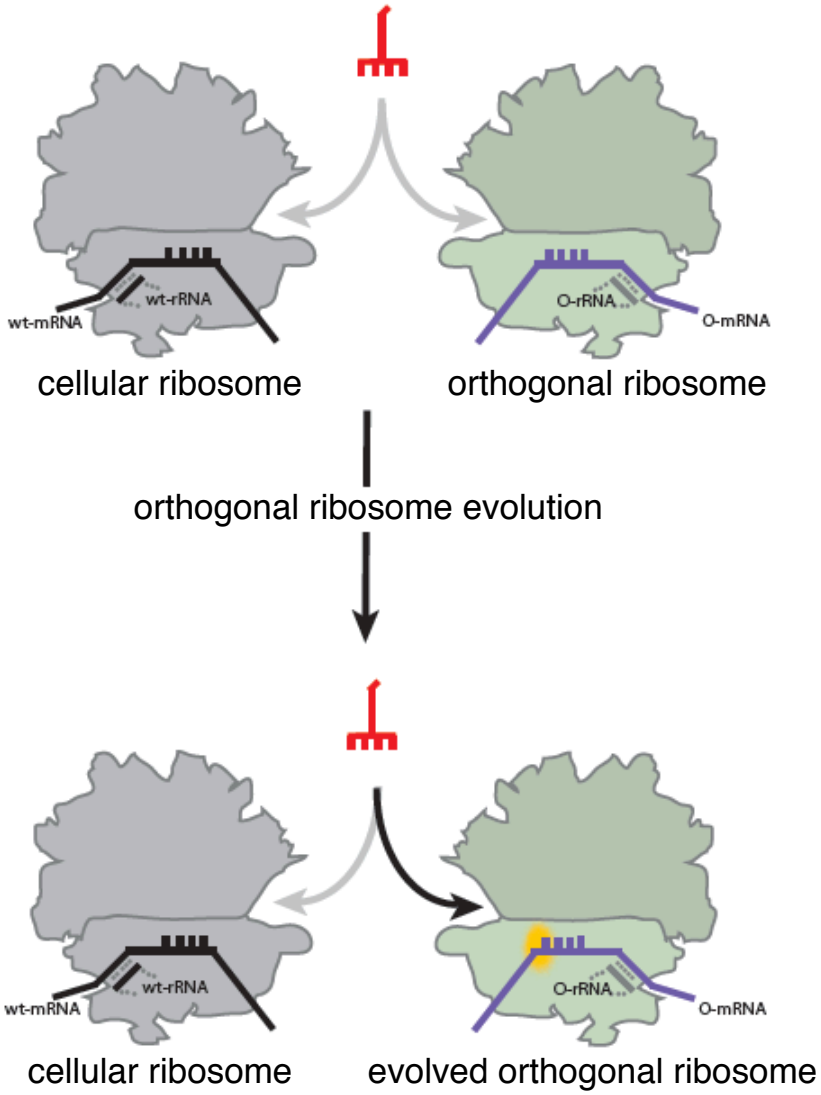
dark interiors indicate prescence of protein

cryogenic electron micrograph

Worsdorfer, B.; Woycechowsky, K. J.; Hivert, D. *Science*, **2011,** *331*, 589.

Evolving Tools for Biology

- Goal: Evolve orthogonal quadruplet-decoding ribosome



Natural amino acids are encoded with triplet codons

To encode unnatural amino acids, amber stop codon (UAG) is not read as stops in amber mutations

Artificial tRNAs recognize amber codon

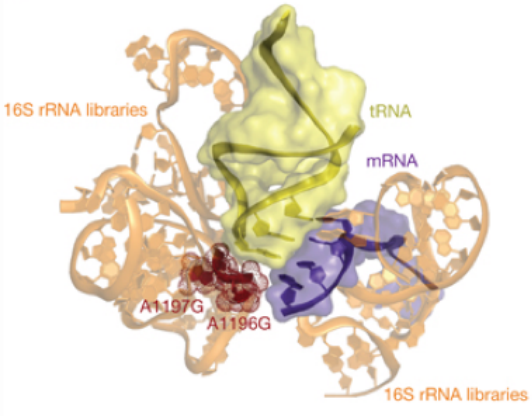
Unnatural amino acid incorporation is usually low (~20%)

Limited number of UAAs can be incorporated

Neumann, H.; Wang, K.; Davis, L. Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464,* 441.

Evolving Tools for Biology

- Goal: Evolve orthogonal quadruplet-decoding ribosome



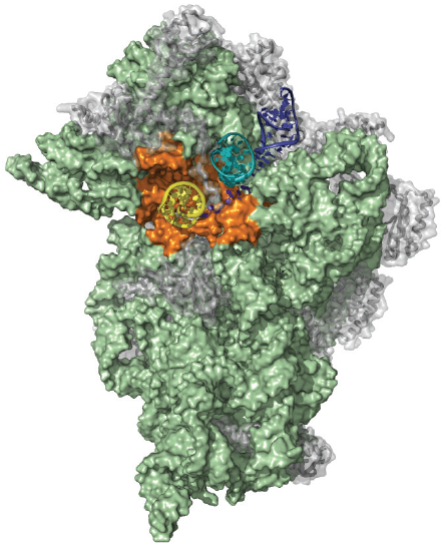
quadruplet codon recognition would

allow multiple UAAs to be incorporated

in silico design combined with evolution

starting from evolved orthogonal ribosome

- Orthogonal ribosome (Ribo-x) recognizes orthogonal tRNA



ribo-X improves unnatural amino acid

incorporation to >60%

127 sequences identified for mutation

are highlighted in orange

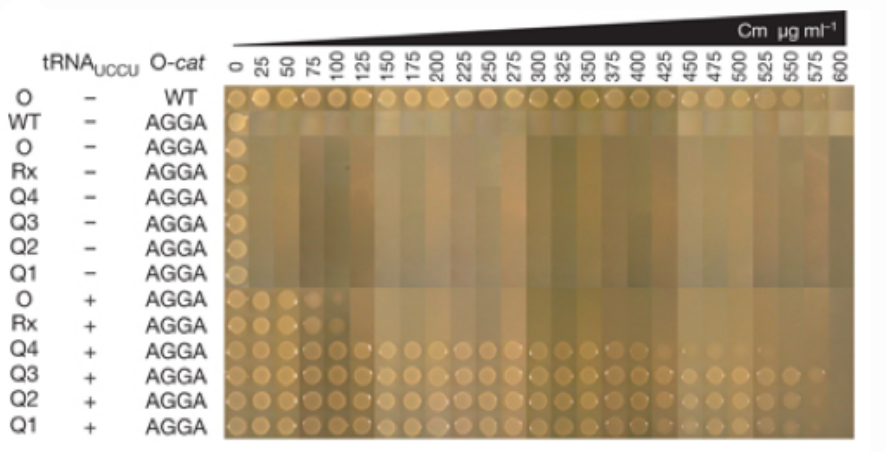
11 saturation mutagenesis libraries created

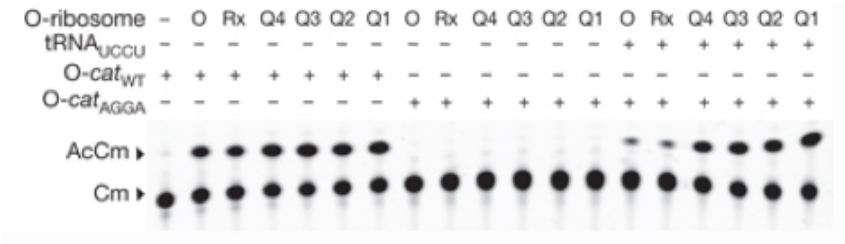
selection is linked to chloramphenicol resistance

Neumann, H.; Wang, K.; Davis, L. Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464,* 441.

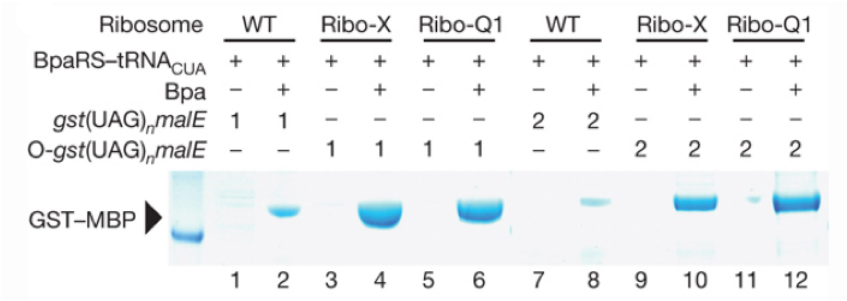
Evolving Tools for Biology

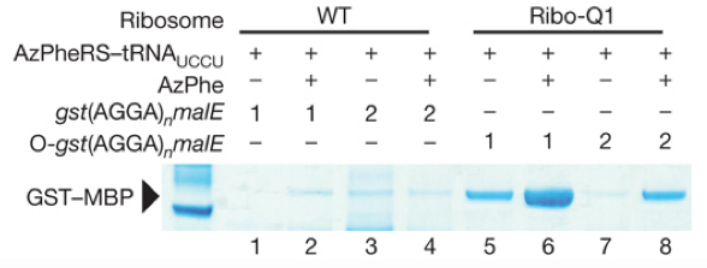
- ribo-Q1 demonstrates best chloramphenicol resistance





- Evolved ribosome recognizes both amber codon and AGGA quadruple codon





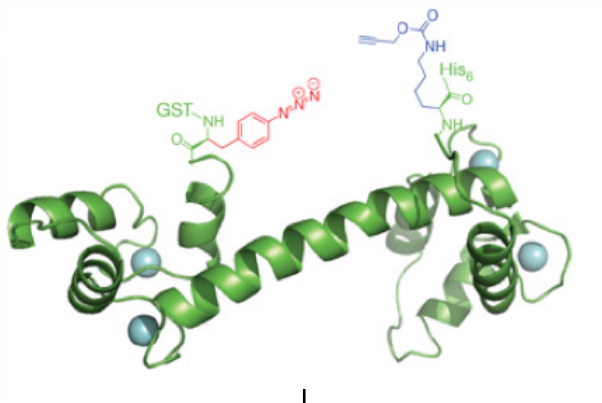
Glutathione-S-transferase (GST) and maltose-binding protein (MBP) fusion protein linked by codon for UAA

Neumann, H.; Wang, K.; Davis, L. Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464,* 441.

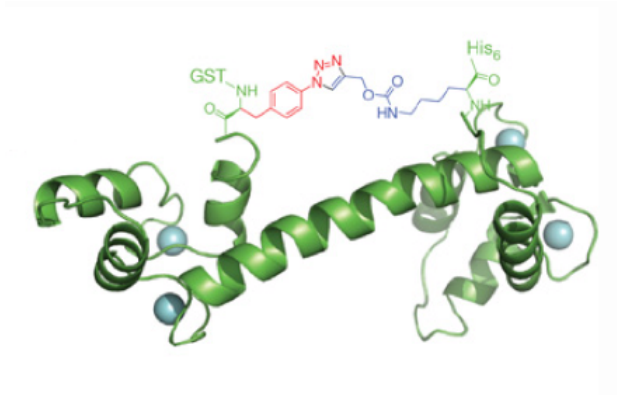
Evolving Tools for Biology

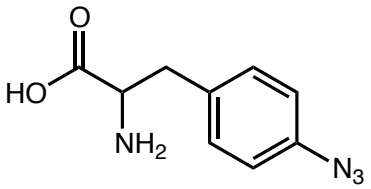
- ribo-Q1 applied to novel protein synthesis

Glutathione-*S*-transferase-calmodulin fusion protein

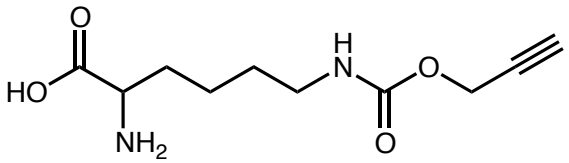


Cu(I) [3+2]

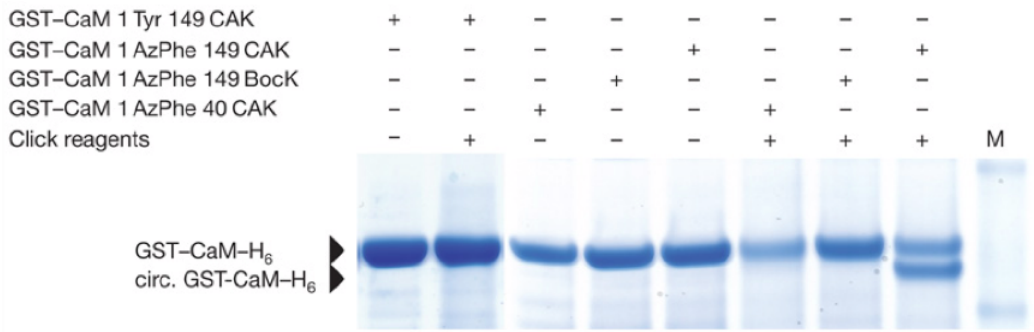




AzPh
AGGA codon



CAK
UAG codon



Neumann, H.; Wang, K.; Davis, L. Garcia-Alai, M.; Chin, J. W. *Nature* **2010**, *464,* 441.

If you could design a protein...

...why would you do it?

...what would it do?