Molecular Diversity I
Biomolecular Diversity

- A lesson from nature; Antibodies;
- Catalytic antibodies (Abzymes)
- Aptamers
- Synthetic Ribozymes

A hypothetical example.

- Lysozyme contains approximately 100 amino acids.
- If you make a random library of 100 amino acids would you expect to find a copy of lysozyme?
- There is $20^{100}$ possible sequences!
  - Each weighing 13.5 kD
  - The universe is only a mere $3 \times 10^{55}$g
- How many of the $20^{100}$ sequences have lysozyme like activity? No body knows.
$10^{12}$ Never looked so small!

How do you use Tweezers?
Selection is easy if you can easily get rid of the items you don’t want!

Note: In the haystack of possible molecules there could be more than one needle
General Concepts for Selections:

- Most often there are many more possible molecular structures than can be made.
- Selection of peptides (or nucleic acids) to bind TS analogs is not as efficient as selecting for reaction catalysis directly.
- The solution identified by selection must be amplifiable. There must be some way to connect the property of the molecule to its encoded sequence. (DNA in B-cells by reverse transcription of RNA)

PROBLEM: translated sequences (proteins) cannot be converted back to their genetic code
Molecular Evolution involves multiple rounds of selection (or screening) while introducing new variations to improve fitness.

What to do when the numbers get too large?

- Genetic selection, phage display, PEGE, selectively infectious phage
- Screen on agar plates: colour / halo formation
- Screen in microtitre plates: chromo/fluoro/colour
- Low throughput high-recov assay: HPLC, mass spectrometry
- Positive for next round of screening
Most functional proteins need to have a stable structure.

There are many proteins that have different functions but related protein folds.

Can an enzyme with a specific catalytic function evolve into a new enzyme with a new catalytic function?

Ways to introduce mutations:
- "Dope" DNA synthesizer with mix of phosphoramidites
- Error prone PCR
- Grown organisms in presence of chemical mutagens
Ways to encode molecular diversity.

Cassette Mutagenesis: Randomize only a small segment of the gene.

Random DNA made on synthesizer

Ligate into vector expressing gene

Error prone PCR can introduce a few changes throughout the gene.

Many polymerases do not replicate DNA with high fidelity.

Original:

The Doping Game

Figure 3. Approaches to randomizing synthetic DNA. Examples show randomization of one codon with mixed nucleotides (NNN, NNTC, NNGT or NNTC) and with trinucleotide phosphoramidites. Synthesis in all three cases commences conventionally 5' of the randomized codon. At the 3' end of the randomized codon (a) all four nucleotides, (b) a mixture of T and C, (c) a mixture of G and T or (b) a mixture of T, G and C can be added. In each case a mixture of all four nucleotides is added at each of the remaining two positions. Having a mixture of G and C in the 3' end of the codon will provide 32 codons, all 20 amino acids and one stop codon. (b) Conversely, the codon can be synthesized by the direct addition of a mixture of 20 trinucleotide phosphoramidites in one step. ALA-TRP represent 20 pre-synthesized 3-m codons, one to code for each amino acid.
* Gene Sequence Encoding Enzyme with Function 1.
  
ex. 50 nt long.

In this example it requires only two changes to the sequence to attain the desired activity, BUT we don't know which seven!

* Resynthesize gene with high error rate/ high doping

Occurrence of Bad mutations out ways the good ones.

* Re-synthesize with low error rate/ low doping

Chances of achieving all three required mutations is too low.

And excellent practical Guide to statistics of Doping: Rob Knight Nucleic Acids Res. 2003 31,

Too many mutations can build a big enough library to select this out of.

Where do you get the Monkey from?

![Graphs showing the relationship between number of mutations and relative performance for wild-type enzyme performing natural and new functions.](image)
How can you find multiple “fit” mutations?

These mutations are both “good” but the chances of finding them in the same molecule is extremely small.

How does Nature optimize phenotype?

Breeding

Desired phenotype

Stemmer developed a method to perform molecular breeding (homologous recombination) in vitro!
Stemmers method of gene reassembly creates mutations that can be recombined to evolve into new functions.

“Sexual PCR” was used to evolve GFP.
Molecular Cross-Breeding

Similar genes from different species contain mutations (variations) that are tolerated by the structure and function of the protein.

Difficult to do when genes are not very homologous or have regions of low homology.

Exploring the fitness landscape:

How do you jump from one fitness plateau to another...
... and you have to do it in the dark!
What if you want to combine (crossover) genes that are not very homologous?

Incremental truncation for the creation of hybrid enzymes (ITCHY). This technology is based on the ability to control both the rate and directionality of DNA digestion using Exonuclease III, enabling the combinatorial generation of protein hybrids.

Problem: only one in three ligations is "in frame"
How do you get more than one

S. J. Benkovic Penn. State

One last trick from the Benkovic Group:

The key features of this plasmid are a Tat export signal sequence and β-lactamase (TEM-1), which functions only in the periplasm.
A General Method for Scanning Unnatural Amino Acid Mutagenesis

Kelly A. Slagel, Mark Lopez, and T. Anthony Crooke

Figure 1. Schematic diagram of codon scanning mutagenesis. (a) A plasmid containing the gene (blue) to be mutated is first targeted with a modified transposon (green) that is inserted into the gene by the action of wild-type transposase, resulting in a library of plasmids that are recovered in E. coli. This insertion event results in the duplication of five nucleotides, 5′-AATTC-3′, at the site of insertion (shown in panel b). The library can be purified to only contain members where the transposase was inserted into the gene of interest by restriction digest. (b) The transposon fragment with recognition sequences, 5′ and 3′, is modified to have MspI restriction sites on either end. Three base pairs, N/K/N′, of the original plasmid sequence are then removed by digestion with MspI, and a new frame-selectable 163-bp linear segment of DNA (orange) is ligated into this digestion site, resulting in a plasmid library (c).
A Heritable Recombination System for Synthetic Darwinian Evolution in Yeast


Heritable Recombination system centered around a library cassette plasmid that enables inducible meiosis and subsequent homologous recombination and subsequent combination of beneficial mutations through sexual reproduction in Saccharomyces cerevisiae.