Massively Multiplexed Zinc Finger Protein Engineering

Harvard iGEM 2011

In every cell, life is governed by 3 major interactions
Protein-Small Molecule
HIV Protease and Saquinovir

Protein-Protein
B12 Antibody and HIV gp120

Protein-DNA
Zinc finger protein and DNA
How do we engineer interactions?
Universal Challenge

Designing new interactions is hard!

No set rules, only guidelines
Traditional: Two Extremes

- Small number of highly educated guesses (structural & biochemical)
  - Higher probability of success
  - Fewer interactions tested
- Vast number of random guesses
  - Lower probability of success
  - More interactions tested
A Middle Approach
Introduction to Zinc Finger Proteins

- Naturally evolved proteins that bind to DNA
- Regulate genes
- Customizable
- Structure coordinated by a zinc ion
**Structure**

**Helix:** Responsible for **binding to a DNA triplet.** Helices are made up of 7 amino acids.

**Backbone:** Gives protein its structure

**Finger:** contains a backbone and a helix, each **binds to a single 3-base DNA triplet**

**Zinc finger protein:** array of three fingers that binds to 9 bases (3 triplets) of DNA.

Example finger sequence: FQCRICMRNFSRSDHHLTTHIRTH
Why Zinc Finger Proteins?

- Binds directly to DNA with high specificity
- Promising applications for gene therapy
- Relatively small protein
- Found naturally in many organisms
Our Project

1. **Design:** use a bioinformatics approach to predict 55,000 zinc finger sequences
   - Targeted against six DNA sequences for three diseases

2. **Synthesize:** use chip-based DNA synthesis to make all 55,000 sequences in one tube

3. **Test:** use a metabolic selection system to test which zinc finger sequences successfully bind DNA
Step 1: **Design**

Determine the most suitable amino acid sequences for binding specific target nucleotide sequences of our choosing.
Problem:

With 7 amino acids per binding helix, there are \(20^7 = 1,300,000,000\) possible helix sequences.

How do we know which ones are likely to bind?
Plan:
Create an algorithm that generates zinc fingers with high probability of binding target sequences:

1. **Analyze data** from previous studies of zinc fingers

2. **Expand** the pool of zinc fingers by including homologous backbones

3. **Make predictions** using known models of zinc finger-DNA binding

4. **Add randomness** to discover even more possible solutions

Design  Synthesize  Test
Data Analysis: Novel Helices

CTG

CNN

NTN

NNG

CTG Helices
Current Binding Models

Relate the change of a letter in a DNA triplet to a change in an amino acid on the helix.
Results

Verifying Our Generator

To make sure our program is creating valid results, we compared our database’s helices for the DNA triplets ANN to ones we generated:

Database Frequencies

Our Generated Frequencies
Step 2: Synthesize

We generated 55,000 predictions, but how do we synthesize that many oligos?
Chip Synthesis

- New technology that synthesizes DNA sequences on a microarray chip
- More cost effective than traditional methods
- 55,000 200-mer sequences per chip
  - Allows us to test a large library to find zinc finger binders

Kosuri et al. 2011
DNA Pool to Zinc Finger Library

1. qPCR

Each prediction is a **DNA sequence**.

Each **DNA sequence** enters **one cell**.

These **cells** become a **living library**.

2. Digestion and ligation

3. Transformation

**Zinc Finger Expression Plasmid with Finger 1 Insert**
Chip Synthesis: Sequencing Results

- Perfect sequence: 57.1%
- 2+ point mutations: 18.2%
- 1 point mutation: 2.6%
- Frameshift: 22.1%
Step 3: Test

Now we have a library of 55,000 variants, but how do we test which ones work?
One-Hybrid Selection System

• His3: positive metabolic selection

• URA3: negative selection

• 3-AT and 5-FOA to fine-tune

Advantages of genome-based parts:
• Stability
• One copy per cell
• Easy!
  – Protocols available on Harvard iGEM 2011 wiki
  – Strains submitted to the Registry
MAGE: Multiplex Automated Genome Engineering

How it works:

- Lagging strand incorporation
- Make small alterations to existing genes
- Perform multiple changes simultaneously and screen

Wang et al, Nature 2009
Lambda Red

- Homologous recombination
- Introduce new sequences into genome
- Antibiotic resistance selection
Building the Selection Strain

- **HisB**: endogenous E. coli version of His3, histidine production
- **PyrF**: endogenous E. coli version of URA3
- **rpoZ**: omega subunit of RNA polymerase
Results

Growth Phenotype: Incomplete Media

![Graph showing growth phenotype with OD 600nm over time for different conditions.](image-url)
Results

Growth Phenotype in Incomplete Media Supplemented with Histidine

![Graph showing growth phenotype over time with and without zinc fingers. The graph plots OD 600nm against time.]

- **Selection Strain**
- **Selection Strain + Zinc Fingers**
Results
Fine-tuning selection

• 3-AT increases stringency of selection
Results

Sensitivity

Recognizes control zinc fingers diluted one in one million with negative control zinc fingers.
Results

Novel Zinc Fingers

- Transformed zinc fingers for the colorblindness target into selection strain and grew in minimal media
- Colonies grew in various 3-AT concentrations
- So far 15 novel zinc fingers have been sequenced
Accomplishments

Design
• Programmed an **algorithm** to generate thousands of potential proteins to bind to specific DNA triplets

Synthesize
• Created a **living library** of our 55,000 sequences targeted to our six target DNA sequences

Test
• Built a **metabolic genomic selection system** sensitive enough to detect the binding of 1:1,000,000 proteins

Parts
• Submitted characterized chassis strains and Biobricks to the Registry
• All protocols available on wiki

Overall
• Engineered **15 potential novel zinc fingers** to bind the triplet TGG, with more currently being characterized
# Acknowledgments

## Team Members
- Kristin Barclay
- Justin Chew
- William Clerx
- Sarah Choudhury
- Naomi Genuth
- Brandon Gerberich
- Mark Kopelman
- Matt Lunati
- Nida Naushad

## Teaching Fellows
- Dan Goodman
- Srivatsan Raman, PhD
- Joyce Yang
- Jun Li
- Noah Taylor
- Jamie Rodgers

## Special Thanks to:
- Jagesh Shah, PhD
- Alain Viel, PhD
- George Church, PhD
- Sri Kosuri, PhD
Accomplishments

Design
• Programmed an algorithm to generate thousands of potential proteins to bind to specific DNA triplets

Synthesize
• Created a living library of our 55,000 sequences targeted to our six target DNA sequences

Test
• Built a metabolic genomic selection system sensitive enough to detect the binding of 1:1,000,000 proteins

Parts
• Submitted characterized chassis strains and Biobricks to the Registry
• All protocols available on wiki

Overall
• Engineered 15 novel zinc fingers to bind the triplet TGG, with more currently being characterized
Supplementary slides
Lambda Red Recombination
Zeocin cassette being swapped into place for rpoZ by lambda red recombination
How Lambda Red works

• Lambda bacteriophage P1 transduction

• Lambda red machinery, exo, beta, gam proteins
ECNR2: designed for lambda red recombination

• Has the lambda red machinery
• Is temperature inducible, 42°C for 15 min
• MutS knockout- reduction in DNA mismatch repair activity, insert less likely to be excised.