Massively Multiplexed Zinc Finger Protein Engineering

Harvard iGEM 2011

A novel integrated system to make and test biological parts
Engineering Biological Parts

Designing new interactions is difficult

No set rules, only guidelines
Traditional: Two Extremes

- Small number of highly educated guesses, using structural and biochemical information
  - Higher probability of success
  - Fewer interactions tested
- Vast number of random guesses
  - Lower probability of success
  - More interactions tested
Our Method

• We reduced-to-practice a **middle approach**
  – Test many interactions
  – Higher probability of success

1. Design
2. Synthesize \[\text{Novel integration of technologies}\]
3. Test

• Applicable to many biological interactions and future iGEM teams
Introduction to Zinc Finger Proteins
• Naturally evolved DNA-binding protein

• Can be customized to target arbitrary DNA sequences

5' - G C G - 3'

Zinc Finger Binding to DNA Triplet
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: FQCRICMRNFSRSDHHLTTTHIRTH
Why Zinc Finger Proteins?

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

Zinc finger protein array bound to DNA
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 genomic metabolic selection system to test which zinc finger sequences successfully bind DNA

**Result:** 15 novel zinc fingers
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. **Make predictions** using **known models** of zinc finger-DNA binding

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

4. **Add randomness** to discover even **more possible** solutions
Data Analysis: Novel Helices
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
- Cost is **1000x cheaper** 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

Design  Synthesize  Test  Human Practices
Chip Synthesis: Sequencing Results

- Perfect sequence: 57.1%
- Frameshift: 22.1%
- 2+ point mutations: 18.2%
- 1 point mutation: 2.6%
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](image)

- **OD 600nm**
- **Time (hours)**

- **Selection Strain**
- **Selection Strain + Zinc Fingers**
Results

Growth Phenotype in Incomplete Media Supplemented with Histidine

![Graph showing growth phenotype](image)

- **Selection Strain**
- **Selection Strain + Zinc Fingers**

Design  Synthesize  **Test**  Human Practices
Results

Fine-tuning selection

3-AT increases stringency of selection

Design  Synthesize  Test  Human Practices
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
Human Practices
• **iGEM Goal**: make a difference in the world

• How do we bring technology to the world?
  – Commercialization
  – iGEM **Entrepreneurial** Division

“*The iGEM Foundation is dedicated to...the development of open community and collaboration*”
The present invention relates to:

1. A fusion protein having a dimerizing domain with or without a ligand-binding region and toxR DNA-binding and hydrophobic transmembrane regions;

2. Host cells comprising the fusion protein and a nucleic acid molecule having a reporter gene operatively linked to the ctx operon, wherein dimerization (ligand-depen-
iGEM: Open-Source and Commercialization

Conflict:

**Commercialization**: bring technology to the world, for a profit

**vs.**

**Open Source Model**: move technology ahead

What happens when iGEM technologies enter the commercial world?
Case Study: Zinc Finger Proteins

• We explored the impact of our project in the existing open-source and commercial context

• Sangamo Biosciences
  – Produce zinc finger proteins as commercial tools
  – $15,000 per zinc finger protein
  – Restricted usage and distribution
  – Control the patent landscape
Ownership (assignees) of US ZFP patents by institution, 1993-2007

- Sangamo (+ Gendaq)
- MIT
- Assorted inventors
- Scripps
- US DHHS
- General Hospital Corp.
- Rockefeller University
- ARCH Development Corp. (University of Chicago)
- Johns Hopkins University
- CalTech
- University of North Carolina at Chapel Hill
- Pioneer Hi-Bred International, Inc.
- Dana-Farber Cancer Institute
- Harvard College
- University of Massachusetts

What are the implications?
Open Source Technology  Intellectual Property Rights
Open Source Technology

- (+) Increases idea sharing
- (+) Lowers cost of using new technologies

Intellectual Property Rights
<table>
<thead>
<tr>
<th>Open Source Technology</th>
<th>Intellectual Property Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>(+) Increases idea sharing</td>
<td>(+) Incentivizes research</td>
</tr>
<tr>
<td>(+) Lowers cost of using new technologies</td>
<td>(+) Provides path to commercialization</td>
</tr>
</tbody>
</table>
Open Source Technology

- (+) Increases idea sharing
- (+) Lowers cost of using new technologies

- (-) Lack of incentives
- (-) Problems with commercialization

Intellectual Property Rights

- (+) Incentivizes research
- (+) Provides path to commercialization
Open Source Technology

• (+) Increases idea sharing
• (+) Lowers cost of using new technologies

• (-) Lack of incentives
• (-) Problems with commercialization

Intellectual Property Rights

• (+) Incentivizes research
• (+) Provides path to commercialization

• (-) Temporarily inhibits spread of technology
• (-) Increases costs of access
Proposal: Research Exemption

Allows academic research without high licensing costs.

(+) Opens the field to research
(+) Maintains incentive for research
(+ ) Can be implemented by policy-makers and scientists.
Conclusion
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

Human Practices
• Initiated discussion on the balance between intellectual property and open source technology

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
References


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

Human Practices
• Initiated discussion on the balance between intellectual property and open source technology

Overall
• Engineered 15 potential novel zinc fingers to bind the triplet TGG, with more currently being characterized
Current Binding Models

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

Persikov, 2011
Supplementary slides
Lambda Red Recombination
Zeocin cassette being swapped into place for rpoZ by lambda red recombination

upstream homology

downstream homology

rpoZ

Rest of the ECNR2 genome
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.