Computational Tools for Synthetic Gene Optimization

Numaan Cheema - Computer Science, 21
Mentor: Dr. Dimitris Papamichail

Introduction

- Last two decades have seen a rise of computational tools developed by software developers and biologists for synthetic biology
- Many types:
  - Optimizing constructs for manufacturability
  - Enabling design of genetic circuits
  - Redesign of protein-coding genes for expression

- My focus: To examine gene optimization computation tools and to analyze their efficacy in performing synthetic gene optimization

Gene Editing

Genome editing (also called gene editing) is a group of technologies that give scientists the ability to change an organism's DNA. These technologies allow genetic material to be added, removed, or altered at particular locations in the genome

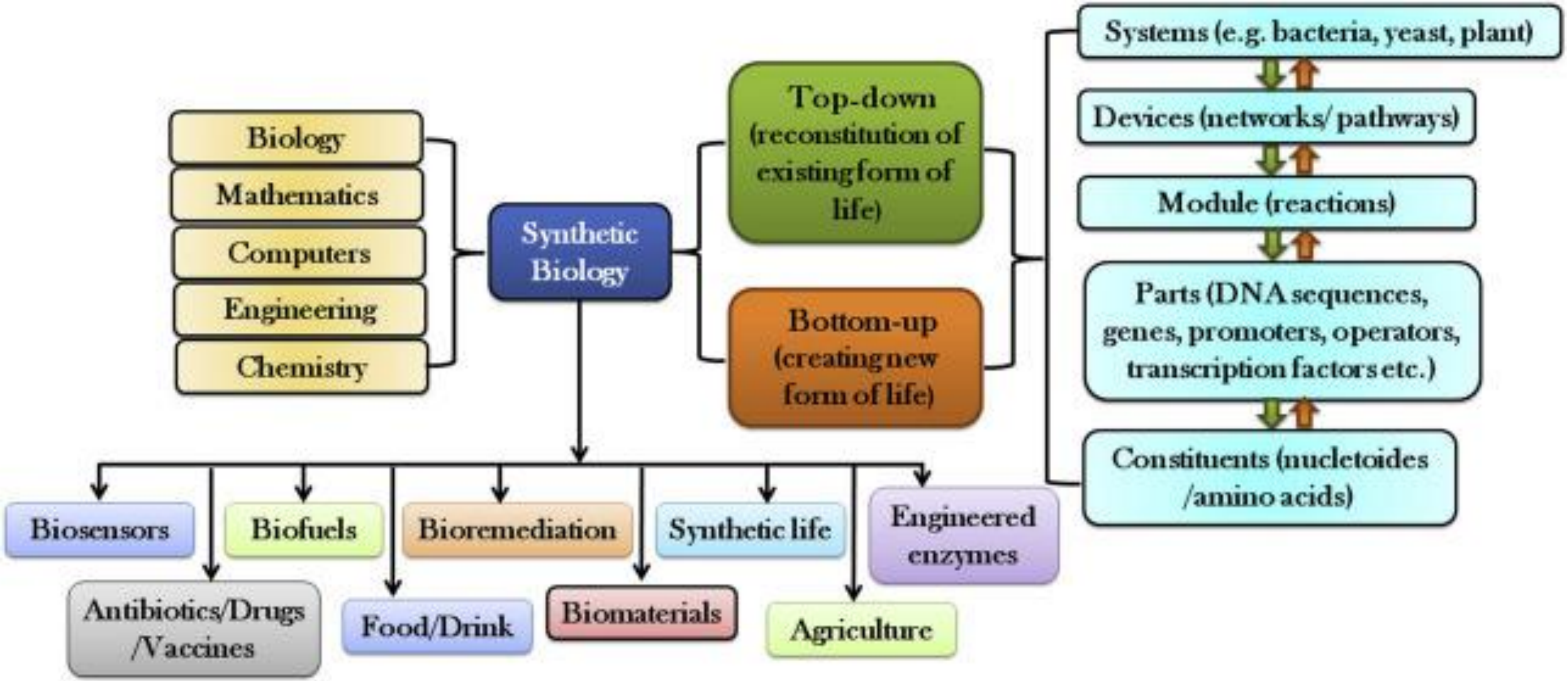

Biology
Mathematics
Computers
Engineering
Chemistry
Synthetic Biology
Top-down (reconstitution of existing form of life)
Bottom-up (creating new form of life)
Systems (e.g. bacteria, yeast, plant)
Devices (networks/ pathways)
Module (reactions)
Parts (DNA sequences, genes, promoters, operators, transcription factors etc.)
Constituents (nucleotoides /amino acids)
Biosensors
Antibiotics/Drugs /Vaccines
Biofuels
Food/Drink
Bioremediation
Biomaterials
Synthetic life
Agriculture
Engineered enzymes

Figure 2. Top down view of why gene optimization (part of synthetic biology) is important. (Microbial Electrochemical and Fuel Cells, 2016)

How does redesign work with tools?

1. The first step is to retrieve the natural DNA sequence of letters across a chromosome
2. Choose an appropriate tool that performs the wanted optimization features

   Examples of these include:

   a. Everytime it sees the letter series TAG change it to TAA
   b. Delete a class of genes referred to as tRNA genes from their positions
   c. Shuffle the DNA sequence according the researchers desired order
   d. Insert or Delete portions of DNA to enhance or minimize particular genes

3. Once the alterations and optimization is completed, then researchers can purchase DNA chunks from companies that can be easily manipulated in a lab

Common Optimization Objectives

1. Codon Usage Bias optimization
   Altering rare codons in the target gene so they more closely reflect the codon usage of host without modifying the amino acid sequence

2. Ribosomal binding site
   Modifying the RBS sequence can improve the translation efficiency and expression levels of protein and the final yield

3. Restriction enzyme site and motif avoidance
   Optimization in the restriction enzyme site is essentially splicing DNA sequences at specific patterns

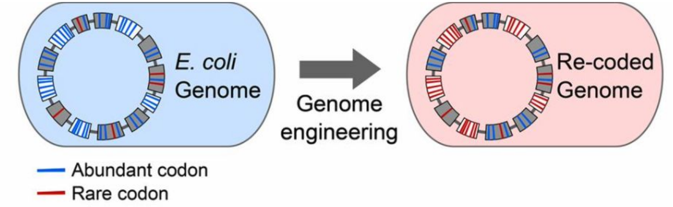


Figure 1. Simplification of codon usage bias optimization

Optimization tools

1. Simple or Single Optimization Objective (One click optimization or single purpose tools)
   a. MOOLTi
   b. Codon Optimization Calculator

2. Multiple Optimization Objectives (Allows multiple objectives to be optimized according to)
   a. DNA Chisel
   b. GENEius
   c. Eugene

How We Tested

Our testing methodology included doing the following:

1. First and foremost a list of “alleged” feature of each program was compiled
2. Then, using a singular consistent gene sequence, every program was thoroughly tested through the claimed features
3. Lastly, a list of features that was successfully tested was listed down

Other factors:

1. **Ease of Use**: Consulted non-technical academics to see their comfort level with each tool
2. **Program Operating System**: Programs on multiple operating systems were testing on each OS separately. Only the functional OS were listed
3. **Maintenance/Updates**: Whether or not these programs were still being updated was noted as well through documentation

Example 1: ATGme

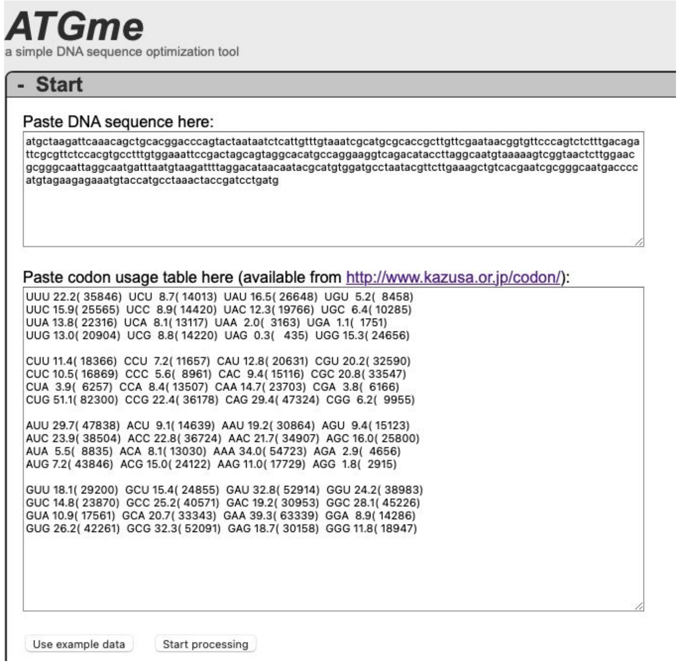
A simple, one-click web-based optimization tool that allows the researcher to perform DNA redesign with the following objectives with a single click:

· Codon usage optimization

· G/C content optimization

· A/T content optimization

· Motif avoidance



ATGme

a simple DNA sequence optimization tool

- Start

Paste DNA sequence here:

```
atgctaagattcaaacagctgcacggacccagtactaataatctcattgtttgtaaatcgcatgcgcaccgcttgttcgaataacggtgttcccagtctctttgacaga
ttcgcgttctccacgtgcctttgtggaaattccgactagcagtaggcacatgccaggaaggtcagacataccttaggcaatgtaaaaagtcggtaactcttggaac
gcgggcaattaggcaatgatttaatgtaagattttaggacataacaatacgcatgtggatgcctaatacgttcttgaaagctgtcacgaatcgcgggcaatgacccc
atgtagaagagaaatgtaccatgcctaaactaccgatcctgatg
```

Paste codon usage table here (available from http://www.kazusa.or.jp/codon/):

```
UUU 22.2( 35846)  UCU  8.7( 14013)  UAU 16.5( 26648)  UGU  5.2(  8458)
UUC 15.9( 25565)  UCC  8.9( 14420)  UAC 12.3( 19766)  UGC  6.4( 10285)
UUA 13.8( 22316)  UCA  8.1( 13117)  UAA  2.0(  3163)  UGA  1.1(  1751)
UUG 13.0( 20904)  UCG  8.8( 14220)  UAG  0.3(   435)  UGG 15.3( 24656)

CUU 11.4( 18366)  CCU  7.2( 11657)  CAU 12.8( 20631)  CGU 20.2( 32590)
CUC 10.5( 16869)  CCC  5.6(  8961)  CAC  9.4( 15116)  CGC 20.8( 33547)
CUA  3.9(  6257)  CCA  8.4( 13507)  CAA 14.7( 23703)  CGA  3.8(  6166)
CUG 51.1( 82300)  CCG 22.4( 36178)  CAG 29.4( 47324)  CGG  6.2(  9955)

AUU 29.7( 47838)  ACU  9.1( 14639)  AAU 19.2( 30864)  AGU  9.4( 15123)
AUC 23.9( 38504)  ACC 22.8( 36724)  AAC 21.7( 34907)  AGC 16.0( 25800)
AUA  5.5(  8835)  ACA  8.1( 13030)  AAA 34.0( 54723)  AGA  2.9(  4656)
AUG 7.2( 43846)  ACG 15.0( 24122)  AAG 11.0( 17729)  AGG  1.8(  2915)

GUU 18.1( 29200)  GCU 15.4( 24855)  GAU 32.8( 52914)  GGU 24.2( 38983)
GUC 14.8( 23870)  GCC 25.2( 40571)  GAC 19.2( 30953)  GGC 28.1( 45226)
GUA 10.9( 17561)  GCA 20.7( 33343)  GAA 39.3( 63339)  GGA  8.9( 14286)
GUG 26.2( 42261)  GCG 32.3( 52091)  GAG 18.7( 30158)  GGG 11.8( 18947)
```

Example 2: DNA Chisel

A powerful multi-objective optimization tool written in python that requires a firm grip on the python language to perform optimizations.

Example Methods used to perform certain optimizations:

· AvoidPattern: Removes unwanted patterns from a sequence

· AvoidChanges: Prevents changes in a specified sequence region

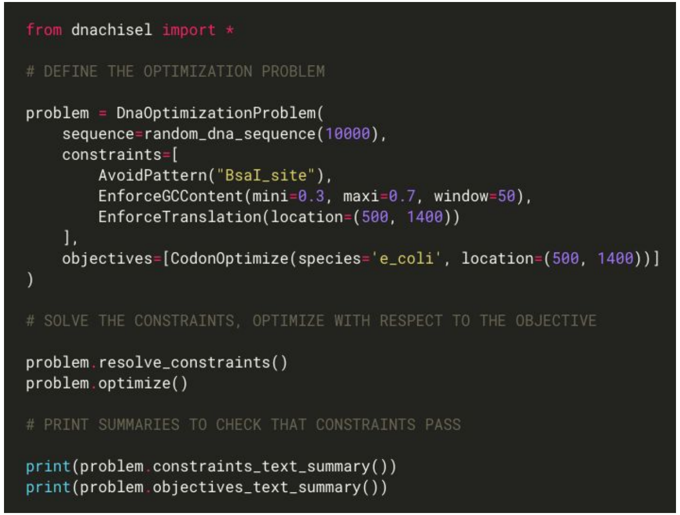
· AvoidMatches: Removes homologies with a genome

· CodonOptimize: Codon optimizes a sequence

· AvoidRareCodons: Replaces rare codons with common codons

· EnforceChanges: Enforces a change between the input and output sequence

· EnforceGCContent: Enforces output sequence GC content within given bounds



```
from dnachisel import *

# DEFINE THE OPTIMIZATION PROBLEM

problem = DnaOptimizationProblem(
    sequence=random_dna_sequence(10000),
    constraints=[
        AvoidPattern("BsaI_site"),
        EnforceGCContent(mini=0.3, maxi=0.7, window=50),
        EnforceTranslation(location=(500, 1400))
    ],
    objectives=[CodonOptimize(species='e_coli', location=(500, 1400))]
)

# SOLVE THE CONSTRAINTS, OPTIMIZE WITH RESPECT TO THE OBJECTIVE

problem.resolve_constraints()
problem.optimize()

# PRINT SUMMARIES TO CHECK THAT CONSTRAINTS PASS

print(problem.constraints_text_summary())
print(problem.objectives_text_summary())
```

Tools Tested

Gene Design Tool	Web URL	Reference
DNAWorks	https://hpcwebapps.cit.nih.gov/dnaworks/	(Hoover and Lubkowski, 2002)
Jcat	http://www.jcat.de/	(Grote et al., 2005)
Synthetic Gene Designer	http://userpages.umbc.edu/~wug1/codon/sgd/	(Wu et al., 2005)
GeneDesign	http://genedesign.org/	(Richardson et al., 2006)
OPTIMIZER	http://genomes.urv.es/OPTIMIZER	(Puigbò et al., 2007)
Visual Gene Developer	http://www.visualgenedeveloper.net/	(Jung and McDonald, 2011)
Eugene	http://bioinformatics.ua.pt/eugene	(Gaspar et al., 2012)
D-Tailor	https://sourceforge.net/projects/dtailor/	(Guimaraes et al., 2014)
GENEius	*http://www.geneius.de/GENEius/*	
GenSmart Codon Optimization	https://www.genscript.com/gensmart-free-gene-codon-optimization.html	
ATGme	http://ATGme.org	(Daniel et al., 2015)
IDT Codon Optimization Tool	https://www.idtdna.com/CodonOpt	
MOOLTi	http://2018.igem.org/Team:HebrewU/Software	(Cohen Tannoudji et al., 2018)
Codon Optimization Calculator by EnCor	http://www.encorbio.com/protocols/Codon.htm	
DNA Chisel	https://edinburgh-genome-foundry.github.io/DnaChisel/	(Zulkower and Rosser, 2020)

Credits

- Dr. Dimitris Papamichail (TCNJ) - Mentor, Teacher, Advisor
- Dr. Adnan Cheema (UPENN) - Non-Technical Academic
- Dr. Amera Cheema (RIDER) - Non-Technical Academic