Gene Variant Libraries:
Design, Construction, and Research Applications

Rachel Speer, Ph.D.

GenScript
Make Research Easy

Gene Variant Libraries:
Design, Construction, and Research Applications

GenScript – The most cited biology CRO

Gene Services

Peptide Services

Protein Services

Antibody Services

Discovery Biology Services

Catalog Products

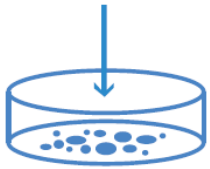
GenScript
Make Research Easy

Part 1: What are Gene Libraries?

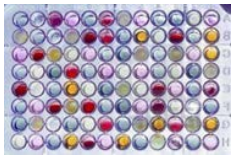
- a collection of many unique DNA sequences

- cloned into a vector

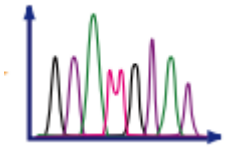
- propagated in micro-organisms



- screened for sequences of interest



- individually sequenced *before or after* assay



The source of your insert DNA depends on your goal

Inventory and study naturally-occurring DNA sequences

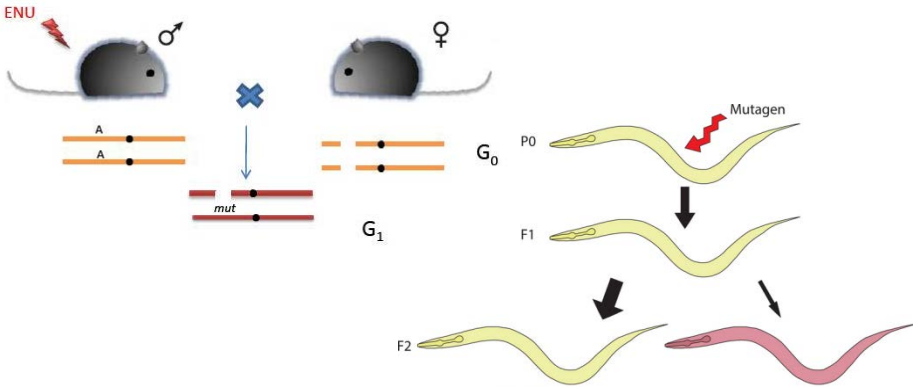
Purify & clone nucleic acids from biological samples.
e.g. cDNA libraries, genomic libraries

Create novel sequences (synthetic genes)

Construct desired sequences via mutagenesis, recombination, or gene synthesis.
e.g. Mutant Libraries

Mutant Libraries

in vivo libraries vs. *in vitro* libraries




```
                    50        60        70        80
                    ....|....|....|....|....|....|....|....|
wildtype            GVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
clone A01           GVVPILVELDGDVNGHKFSVPGEGEGDATYGKLTLKFICT
clone A02           GVVLILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIGT
clone A03           GVVPILVELDGDLTGHKFSVSGEGEGDATYGKLTLKFICT
clone A04           GVVPILVELDGDVNGHKFSVSAEGECDATYGKLTLKFICT
clone A05           GIVPILVELDGDVNGHKFPVSGEGEGDATYGKLTLKFICT
clone A06           GVVPILVKLDGDVNAHKFSVSGQGEGDATYGKLTLKFICT
clone A07           GVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKAICT
clone A08           GVVPILVELEGDVNGHKFSVSGEGEGDASYGKLTLKFICT
....                ....
clone H12           GVVPILVELDGDVNGHKFSVSGERGEGDATYGKLTLKFICT
```

Expose organisms to a mutagen

Screen for phenotypes of interest

Identify gene mutation

Create mutant DNA sequences

Express in organism

Screen for desired phenotypes

Synthetic Gene Libraries: Pools vs. Sequence-Verified Clones

dsDNA Fragments
Mixture of sequences
e.g. degenerated PCR products

Transformants
(colonies = unique clones?)

Sequence-verified clones
(purified plasmid DNA, glycerol stock)

- inexpensive & faster to create
- "quick & dirty" initial screening

vs.

- high-quality
- known identity
- reliable replication

When to use synthetic gene libraries

- When you want to create something that doesn’t already exist in nature.
- When you want to be systematic and unbiased
- When expressing genes in a different model organism
- When you know the sequences you want

When *not* to use synthetic libraries

- When you want to create a library from a biological sample
- When you don’t know what kind of variants you want

Part 2: Design Strategies for Synthetic Libraries

- Expression-Ready Gene Variant Libraries
- Mutant Libraries for Protein Engineering
- Synthetic Biology Libraries

Gene Variant Libraries

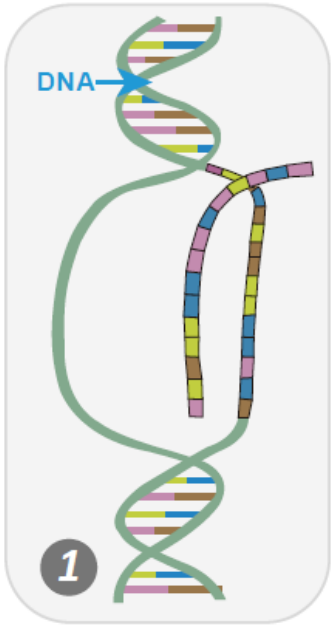
ORFs cloned into expression vectors allow functional characterization of gene variants

- family members
- isoforms / splice variants
- disease-related variants
- mutants

Why create synthetic libraries of naturally occurring genes?

- Ensure sequence accuracy
- Get expression-ready clones (optional: with tags)
- Improve expression levels through **codon optimization**

Codon Optimization improves protein expression

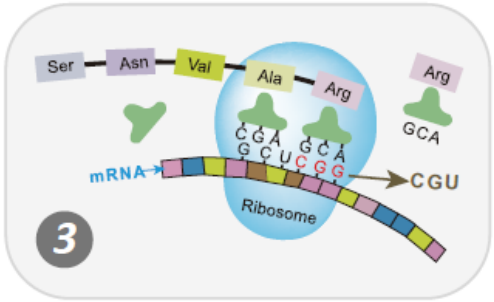


1. Transcription

- cis-regulatory elements (TATA box, termination signal, protein binding sites, etc.)
- chi sites
- polymerase slippage sites

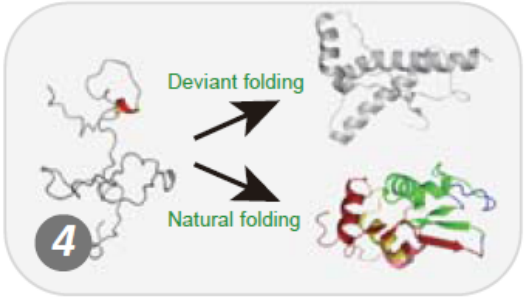
2. mRNA processing and stability

- cryptic splice sites
- mRNA secondary structure
- stable free energy of mRNA



3. Translation

- codon usage bias
- ribosomal binding sites (e.g. IRES)
- premature polyA sites

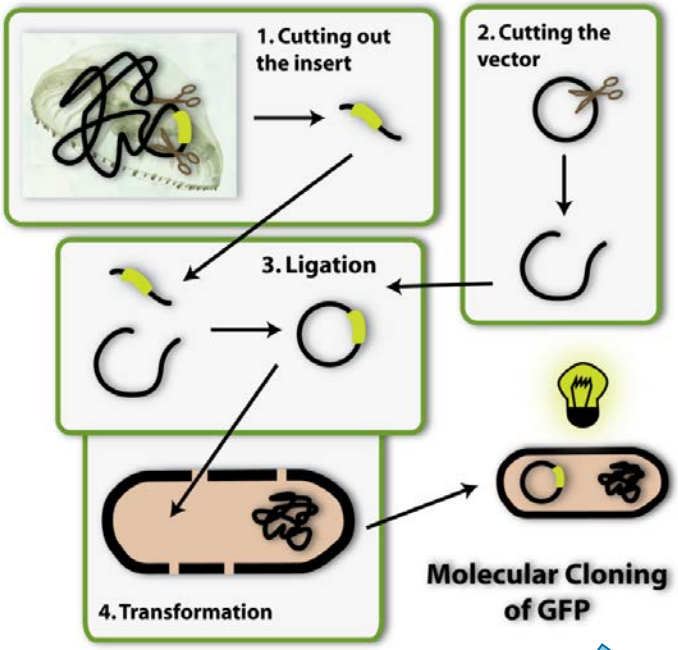


4. Protein folding

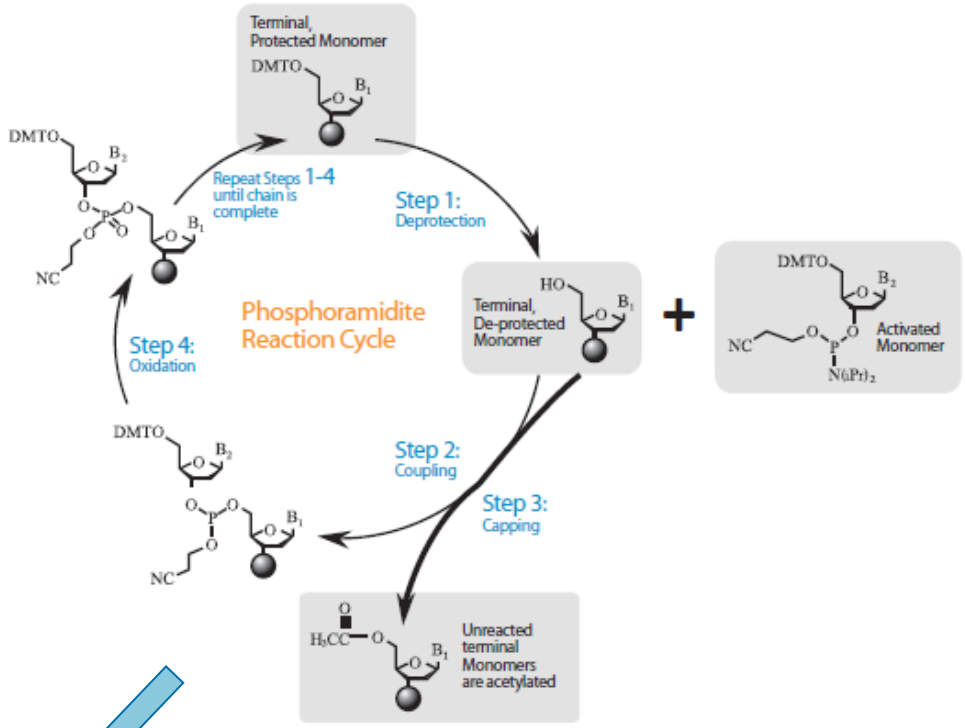
- codon context
- codon-anticodon interaction
- translation pause sites

Constructing Gene Libraries

Traditional Molecular Cloning:
Cut & Paste naturally occurring sequences, PCR-based SDM as needed



Gene Synthesis:
de novo Chemical Synthesis of DNA does not require a template



Clone into desired vector

Choosing the most appropriate vector

Type of Cloning Vector	Advantages	Disadvantages
Plasmid	Replicate most prolifically; 700 copies per cell; most popular for inserts <5kb	Insert size Limited to 15kb (often lose larger inserts)
Bacteriophage	5-15kb	
Cosmids	replicate well, hold inserts 30-45kb, rarely lose inserts	Somewhat unstable, susceptible to recombination if contain repeats
BAC – Bacterial Artificial Chromosome	Insert size up to 350kb, fewer chimeras than YACs,	Only 1 copy per cell
YAC – Yeast Artificial Chromosome	Insert size up to 1000kb	High rate of chimeras

Plasmids: cloning vectors vs. expression vectors

Amplify, manipulate, and store DNA sequences

Cloning Vector

- high copy number
- multiple cloning region
- antibiotic resistance / lacz for selection

Efficiently express a sequence in my cell/tissue of interest

Expression Vector

- promoter for transcription
- Kozak (eukaryotic) or Shine-Dalgarno (prokaryotic) sequences for translation
- Viral vectors for *in vivo* delivery

You can create a library directly using your choice of Expression Vector; no need to shuttle!

Plasmid-driven vs. endogenous expression

Extra-chromosomal plasmid DNA

- strong promoter → overexpression
- *ideal for protein purification, reporter assays*

Targeting Vector for integration into chromosomal DNA

- CRISPR/Cas9 is simpler and more efficient than ZFN, TALEN, Cre-lox
- *ideal to study gene function under endogenous promoter, normal stoichiometry*

Design Strategies for Synthetic Libraries

- Gene Variant Libraries
- Mutant Libraries
  - Rationally designed
  - Systematic / Saturated
  - Truncation variants
  - Random / degenerated
- Synthetic Biology Libraries

Protein Engineering Strategies

Rational Design

Synthesize Desired Sequences (up to 10^4)

↓

Functional Assay

↓

Identify Best Variants

Directed Evolution

Create Mutant Library (10^4 – 10^{11})

↓

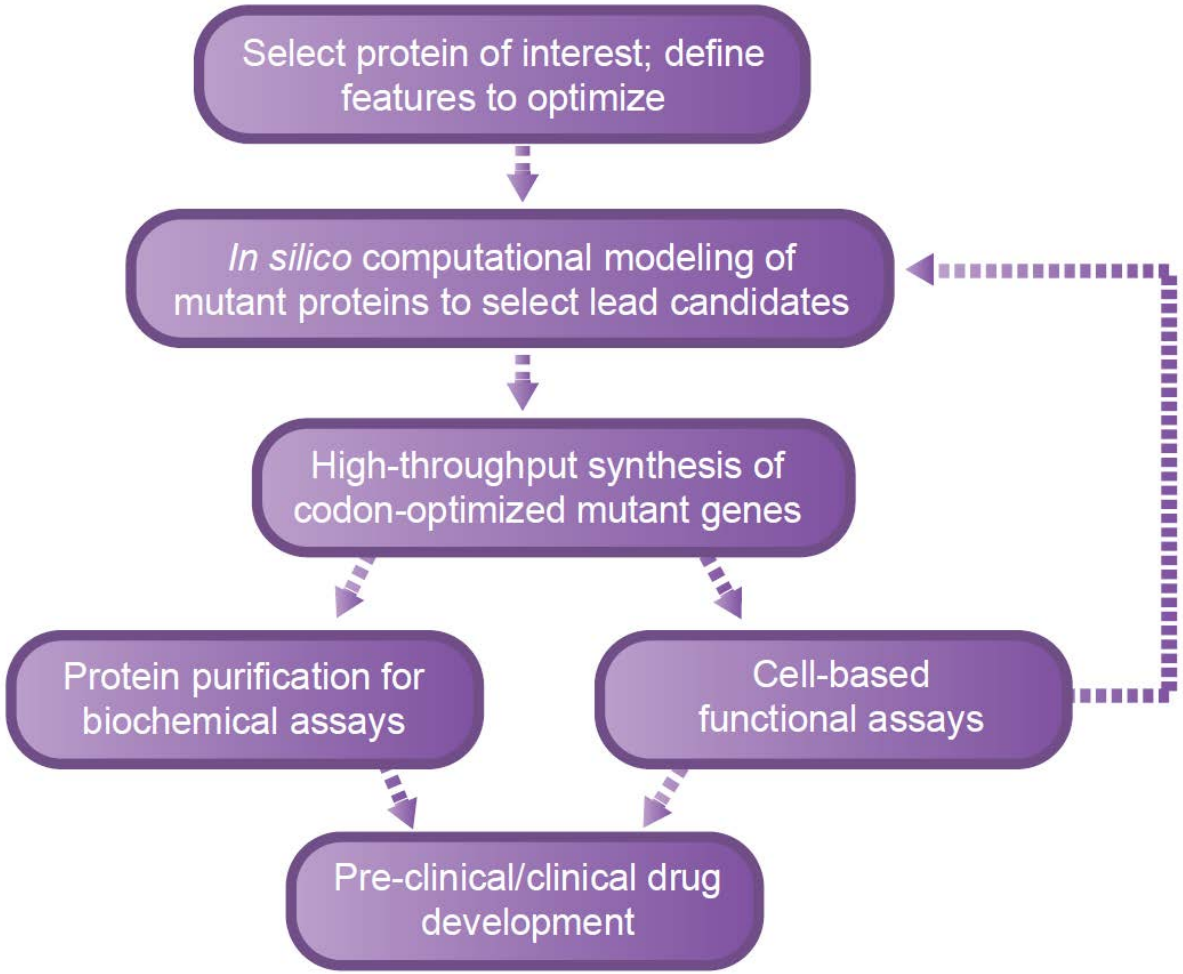
High-throughput Screening

↓

Select Best Hits after multiple rounds of selection

Use best hits as template for next round of mutagenesis

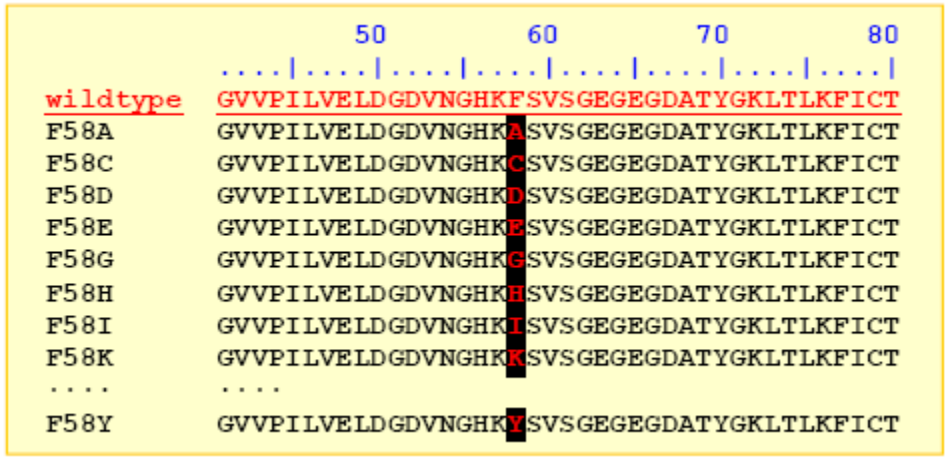
Case Study: rational design of new drugs based on computational modeling



Medical applications for protein engineering:

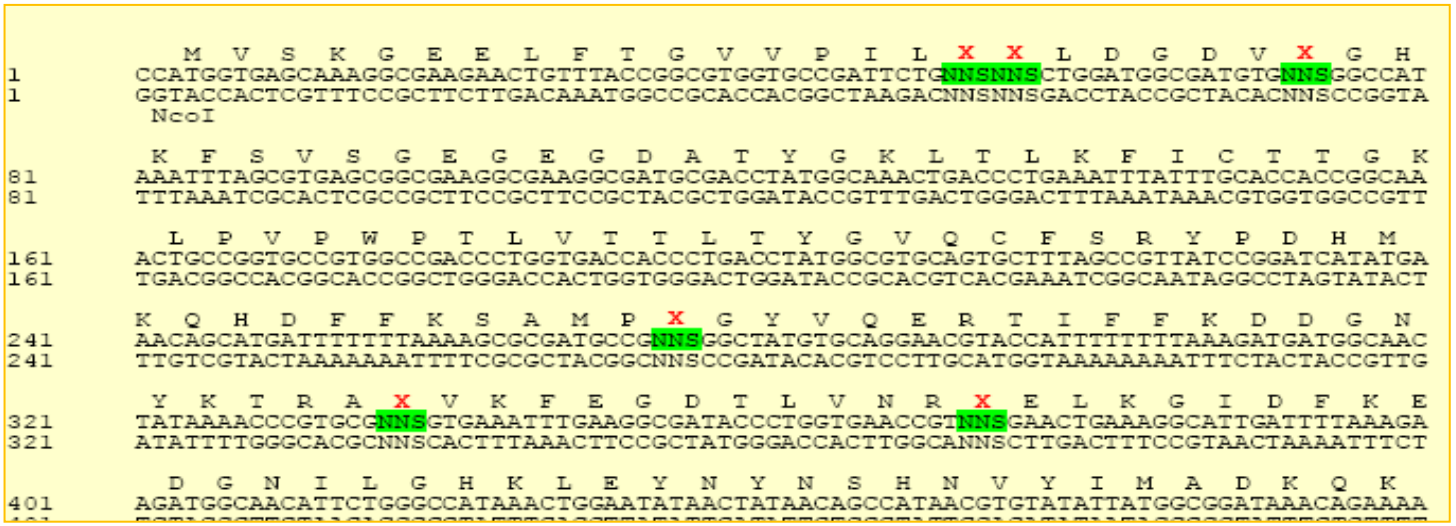
- vaccine design
- mAbs for immuno-oncology
- therapeutic enzymes

Systematic Point Mutation Libraries



Site-Saturation Mutagenesis

- a single amino acid
- multiple residues



Scanning Libraries and Sequential Permutation Libraries

GVVPILVELDGDVNGHKFSVSGEGEGDATYGK
GVVPILVELDGDVNG**A**KFSVSGEGEGDATYGK
GVVPILVELDGDVNGH**A**FSVSGEGEGDATYGK
GVVPILVELDGDVNGHK**A**SVSGEGEGDATYGK
GVVPILVELDGDVNGHKF**A**VSGEGEGDATYGK

Alanine scanning

GVVPILVELDGDVNGHKFSVSGEGEGDATYGK
GVVPILVELDGDVNG**X**KFSVSGEGEGDATYGK
GVVPILVELDGDVNGH**X**FSVSGEGEGDATYGK
GVVPILVELDGDVNGHK**X**SVSGEGEGDATYGK
GVVPILVELDGDVNGHKF**X**VSGEGEGDATYGK

Consecutive site-saturations

GVVPILVELDGDVNGHKFSVSGEGEGDATYGK
GVVPILVELDGDVNG**X**KFSVSGEGEGDATYGK
GVVPILVELDGDVNG**XX**FSVSGEGEGDATYGK
GVVPILVELDGDVNG**XXX**SVSGEGEGDATYGK
GVVPILVELDGDVNG**XXXX**VSGEGEGDATYGK

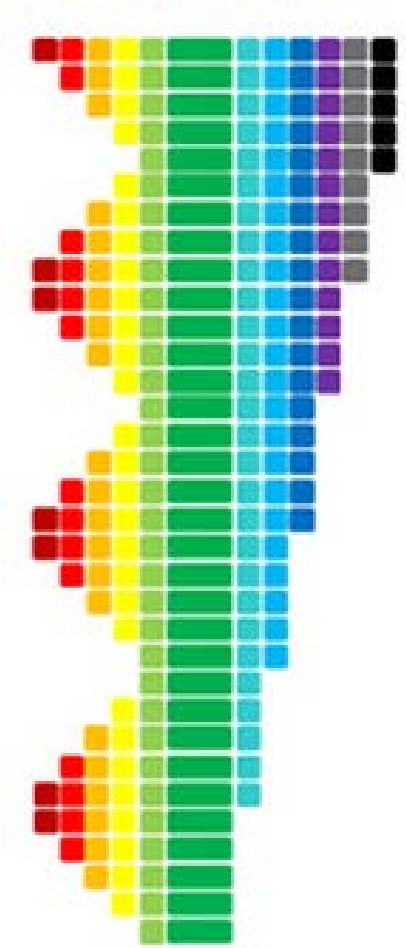
Sequential Permutation

Size Matters: Constraining your design will reduce screening burden

Mutating only 6 residues of interest yields ***many*** unique sequences

- Site-saturated: 64 codons — 64^6 =68,719,476,736
- NNS: — 32^6 = 1,073,741,824
- Trimer library: 20 a.a. — 20^6 = 64,000,000
- Rationally designed SDM — 2*2*8*4*3*3= 1152

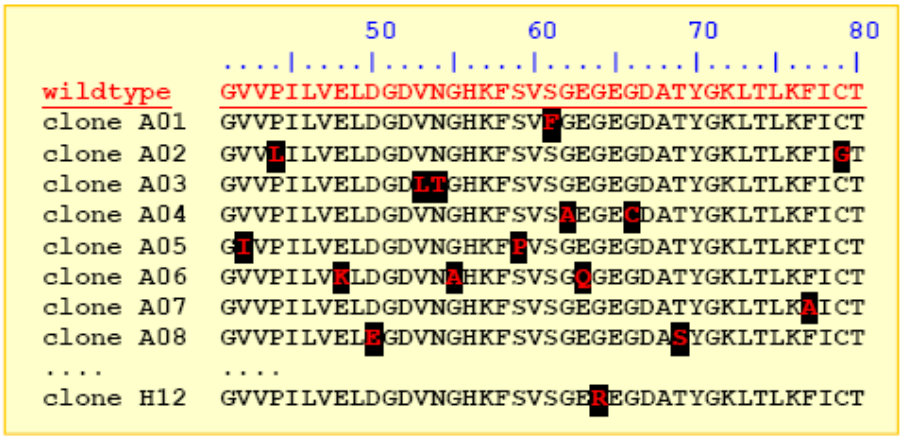
Truncation Variant Library



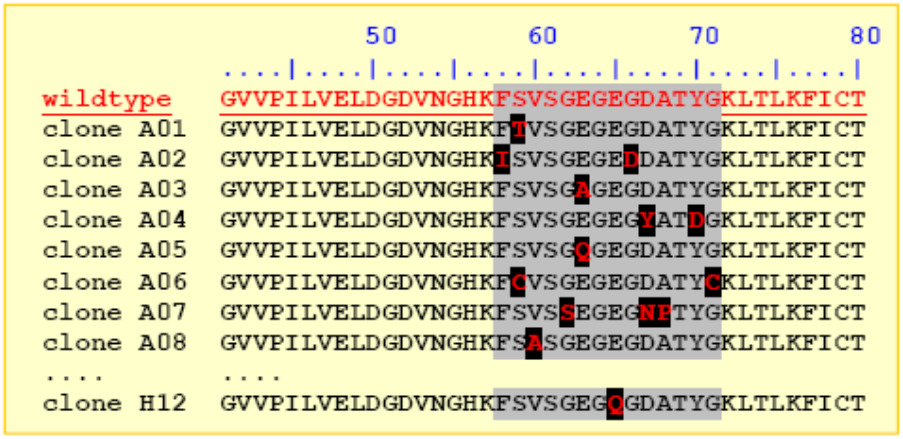
Valuable for Structural Biology:

- Optimize protein solubility and stability
- Identify minimal domains required for folding, conformational stability, protein-protein interactions, catalytic activity

Random Mutagenesis for Directed Evolution



Controlled randomization of complete reading frame



Controlled randomization of partial reading frame

Techniques for constructing large (systematic or random) mutant libraries

Error-prone PCR for random mutations

- Polymerase, cations, dNTPs, cycles

Degenerate Oligos for site-saturation / sequential permutation libraries

- Single or combinations

Trimers for efficient protein engineering

- Replace codons instead of single nucleotides

Which library type is best?

Depends on:

- Your starting knowledge
- Your goals
- Your screening capacity

Bigger isn't always better!

Curr Opin Struct Biol. 2013 Jun;23(3):403-8. doi: 10.1016/j.sbi.2013.03.004. Epub 2013 Apr 17.

Library methods for structural biology of challenging proteins and their complexes.

Hart DJ[1], Waldo GS.

Author information

Abstract

Genetic engineering of constructs to improve solubility or stability is a common approach, but it is ...
the domain composition of a target is poorly understood, or if there are insufficient ...
phases of subcloning or mutation and expression often prove ...
problem and involve construction of large libraries ...
improvements in construct ...
these experiment...

Curr Opin Biotechnol. 2013 Dec;24(6):1017-22. doi: 10.1016/j.copbio.2013.03.003. Epub 2013 Mar 26.

Accelerated protein engineering for chemical biotechnology via homologous recombination.

Nordwald EM[1], Garst A, Gill RT, Kaar JL.

Author information

Abstract

Protein engineering has traditionally reli... generate diverse libraries,
which require high-throughput screenin... natively,
approaches to semi-rational library co... nce the
efficiency by which improved mutants...
of relatively few variants due to the ...
biology, namely multiplexed DNA sy...
rapidly construct large, rational libr...
encoded features that permit effic...
protein library design could comp...
search complex fitness landscap...
also has the potential to expan...
whose concerted activities are ...

Copyright © 2013 Elsevier Ltd. All rights reser...

PMID: 23540421 [PubMed - indexed for MEDLINE]

Biotechnol J. 2007 Feb;2(2):180-91.

Better library design: data-driven protein engineering.

Chaparro-Riggers JF[1], Polizzi KM, Bommarius AS.

Author information

Abstract

Data-driven protein engineering is increasingly used as an alternative to rational ...
knowledge to limit library size, while still allowing for the identification of unpredi...
computational modeling and bioinformatics, as well as an increasing databank ...
choose particular amino acid residues to vary in order to increase the chances ...
limiting diversity at each position, design of small sub-libraries, and the performa...
automated, further reducing the library size.

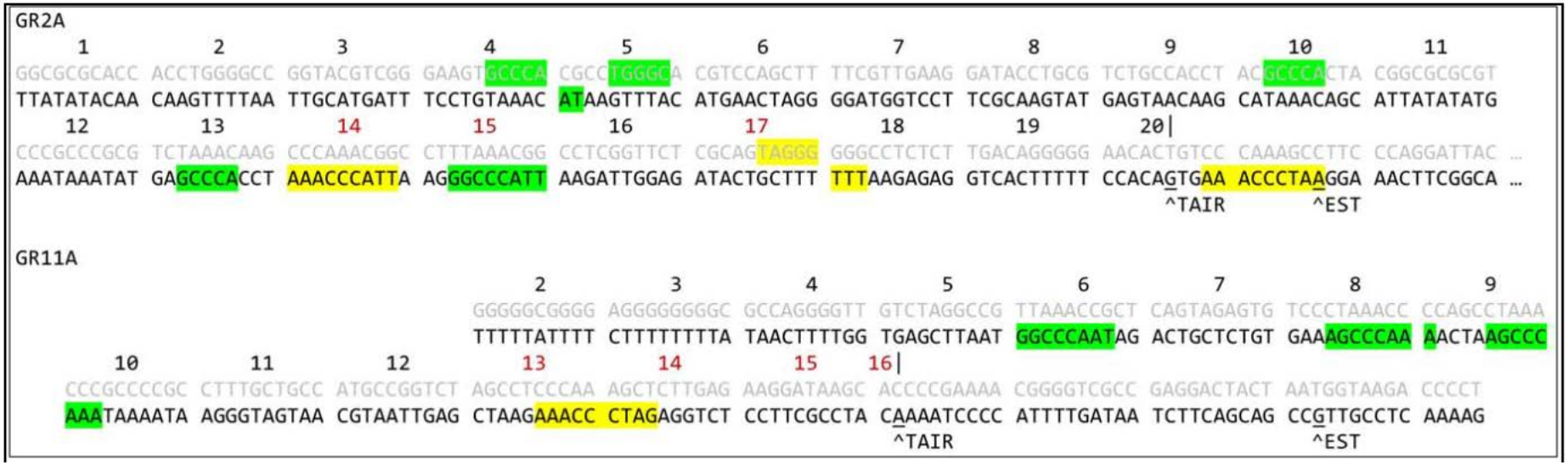
PMID: 17183506 [PubMed - indexed for MEDLINE]

Rationally designed Mutant library for Promoter-Bashing

Research Field: Agriculture/Plant Biology

Challenge: Validate a new computational algorithm for identifying novel gene regulatory sequence motifs by systematic mutation

Solution: Synthesize a small library of constructs harboring systematic mutations in the 5' intergenic sequence.



Davis IW, Benninger C, Benfey PN, and Elich T. POWRS: Position-Sensitive Motif Discovery PLoS One. 2012; 7(7): e40373. doi: 10.1371/journal.pone.0040373

Design Strategies for Synthetic Libraries

- Gene Variant Libraries
- Mutant Libraries
- Synthetic Biology Libraries
  - Combinatorial assembly variants
  - Synthetic Genomes / Synthetic Organisms

Combinatorial Libraries

If you want to use 5 variants of each component:

Promoter 1	RBS 1	ORF 1	Terminator 1
Promoter 2	RBS 2	ORF 2	Terminator 2
Promoter 3	RBS 3	ORF 3	Terminator 3
Promoter 4	RBS 4	ORF 4	Terminator 4
Promoter 5	RBS 5	ORF 5	Terminator 5

Then you will need a combinatorial assembly library containing
5×5×5×5=625 unique composite sequences

Just a few of those combinations:

Promoter 1	RBS 1	ORF 1	Terminator 1
Promoter 2	RBS 2	ORF 2	Terminator 2
Promoter 3	RBS 3	ORF 3	Terminator 3
Promoter 4	RBS 4	ORF 4	Terminator 4
Promoter 5	RBS 5	ORF 5	Terminator 5

iGEM Synthetic Biology
based on standard parts

Case Study: combinatorial assembly library for gene targeting / genome editing

Kim, Y. *et al.* A library of TAL effector nucleases spanning the human genome. *Nat. Biotechnol.* **31,** 251–258 (2013).

- created TALENs for 18,740 unique protein-coding human genes
- synthesized 84 TALE plasmids containing all possible RVD combinations

Table I. TALE specificity code

RVD	Nucleotide
NI	adenine
HD	cytosine
NG	thymine
NN	guanine

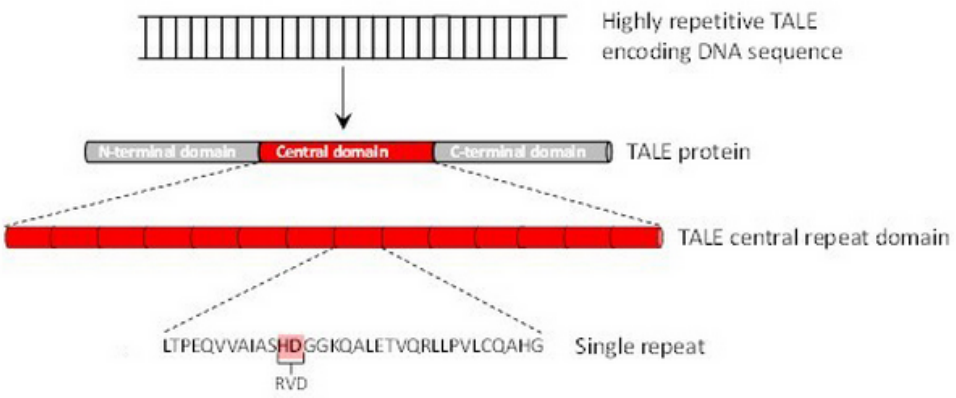


Fig 2. TALE protein organization

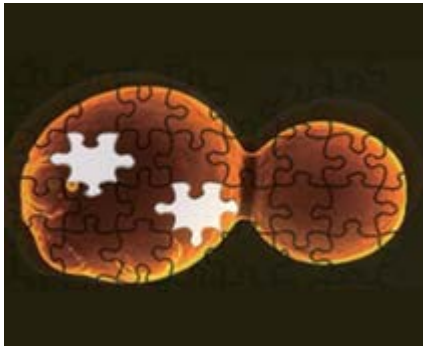
Codon-optimized gene synthesis solves problems for TALEN Combinatorial Library

1. Limit sequence similarity

2. Exclude rare codons to maximize translation efficiency

3. Guarantee accuracy of highly-repeated sequences

Synthetic Genomes: a new model for *in vivo* mutant libraries



GenScript is proud to be a contributing partner in the Sc2.0 International Consortium whose goal is to build a designer synthetic eukaryotic genome.

Sc2.0 – The Synthetic Yeast Genome Project

SCRaMbLE, a built-in inducible diversity generator

Goals:

- Deduce the limits of chromosome structure
- Accelerate new strain development for biofuels & medical applications

Building a Synthetic Eukaryotic Genome – Sc2.0

Presented by: Leslie Mitchell, Ph.D., NYU Langone Medical Center

June 25, 2014/ 2:00 pm EST

Register now

Case Study: combining multiple design strategies in an “incognito library”

King, S. R. F. *et al.* Phytophthora infestans RXLR Effector PexRD2 Interacts with Host MAPKKK{varepsilon} to Suppress Plant Immune Signaling. *Plant Cell* **26,** 1345–1359 (2014).

Design strategies:

1) **gene variants**: multiple PexRD2-like family members
2) **structure-led mutagenesis:** 5 specific PexRD2 mutants
3) **combinatorial assembly**: mutant genes fused with GFP, FLAG, or YN tags for different assays

Construction method:

A combination of *de novo* gene synthesis, SDM, and recombination

GenScript Toolkit

Library Type	Service	Advantages
Expression-Ready Clones	GenEZ ORF Cloning or Custom Cloning	•choose your vector •10µg sequence-verified plasmid DNA
Codon-Optimized Genes	OptimumGene + Gene Synthesis	•increases protein expression •optional: protein expression evaluation
Sequence-Verified Synthetic Library *25-10,000 variants*	GenPlus Next-Gen Gene Synthesis	•cost-effective HT platform •4µg sequence-verified plasmid DNA
Mutant Library – Rationally Designed *10^1-10^6 variants*	Site-Directed Mutagenesis Library, Scanning Point Mutation	•Choose sequence-verified clones or mixed library
Mutant Library – Systematic or Randomized *10^3 - 10^{11} variants*	Sequential Permutation Libraries or Randomized and Degenerated Libraries	•10µg of dsDNA up to 10^{11} variants •pooled clones in your choice of vector •pooled glycerol stock up to 10^9
Truncation Variants	Truncation Variant Library	•Up to 2000 variants within 4 weeks •4µg sequence-verified plasmid DNA
Combinatorial Variants	Combinatorial Assembly	
Genome / Chromosome / metabolic circuits	GeneBricks	•~10kb building blocks •100% guaranteed sequence accuracy

Q&A

- Thank you for attending!
- Please submit questions through chat.
- Please complete the survey you'll receive by email.
- Check upcoming and archived webinars at www.genscript.com/webinars.html
- Email me any time: rachel.speer@genscript.com