Codon optimization:
Why & how to design DNA sequences for optimal soluble protein expression

Rachel Speer, Ph.D.

GenScript
Make Research Easy

1. Why do codons matter?
2. OptimumGene principles & performance
3. Case Studies
4. How to get better protein expression

Protein Expression Overview

Select/Design the end product (amino acid sequence)

↓

Choose expression system

↓

Design expression clone (DNA construct)

↓

Express the protein

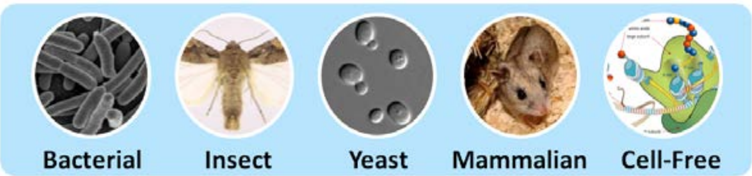
↓

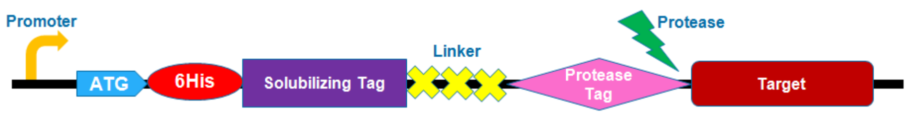
Purify the protein

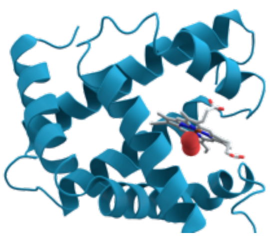
↓

Characterize the protein

MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLIC…







Protein Expression Overview

Select/Design the end product (amino acid sequence)

↓

Choose expression system

↓

Design expression clone (DNA construct)

↓

Express the protein

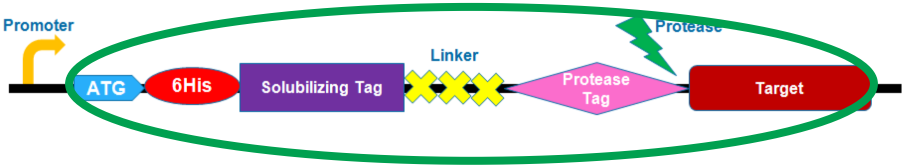
↓

Purify the protein

↓

Characterize the protein

Codon Optimization



Why do Codons Matter? The Facts

- Redundancy in the genetic code
- Synonymous mutations affect protein expression rates up to 1000-fold.
- Synonymous mutations can also alter protein conformation, PTM, stability, and function.

Second Letter

1st letter	U		C		A		G		3rd letter
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Codon Optimization:

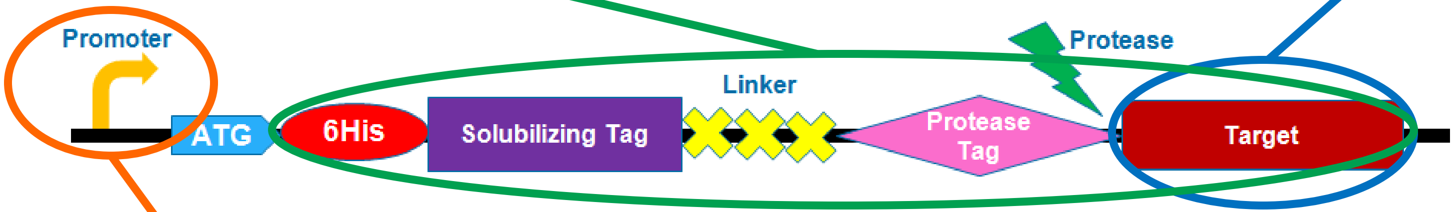
Introducing synonymous mutations that favor efficient soluble protein expression

```
Optimized AGTTTTCCAGGTTGAGGTCCGCCCGTT
          || || || || ||||| || || |||
Original  AGCTTCCCGGGATGAGGGCCCCCGGTT
```

What Codon Optimization is – and isn't

Codon Optimization:
Introducing synonymous mutations that favor efficient soluble protein expression

Protein Design:
changing the amino acid sequence



Expression strategy:
Selecting promoter, tags, etc.

Codon Bias

Observations:

- Species-specific bias in codon use and tRNA abundance
- Heterologous protein expression is often inefficient

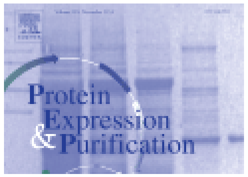
Theory:

- Rare codons reduce protein expression

Solutions:

- Express tRNA to remove bias in the host cells
- Alter the gene to replace rare codons with preferred ones:
  - site-directed mutagenesis
  - *de novo* gene synthesis with codon optimization

Codon Optimization Significantly Increases Protein Expression Levels



Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study.

Burgess-Brown NA *et al. Protein Expr Purif*. May 2008; 59(1): 94-102

Gene Name	Native			Synthetic			Expression	Solubility
	1	2	3	1	2	3	Syn vs Nat	Syn vs Nat
CBR1	Blue	Blue	Blue	Blue	Blue	Blue	▲	▲
CBR3	Blue	Blue	Blue	Blue	Blue	Blue	▲	▲
GMDS					Blue	Blue	▲	▲
HADH2		Green	Green		Blue	Green	▲	▲
HSD1 7B2	Green	Green		Green	Green		▲	
HSD17B4		Blue	Blue	Blue	Blue	Blue	▲	▲
MGC4172	Blue	Blue	Blue	Blue	Blue	Blue		
PECR				Blue	Blue	Green	▲	▲
RETSDR2	Green	Blue	Green	Green	Green	Green	▲	
SPR	Blue	Blue	Blue	Blue	Blue	Blue		

- Green: Expressed
- Blue: Expressed, Soluble and Purified
- White: Not Expressed

▲ Targets shown improvement of expression and/or solubility with synthetic gene after codon optimization

1. Total Cellular Protein
2. Soluble Fraction
3. Eluted Fraction

Codon Adaptation is not the most important factor for protein yield

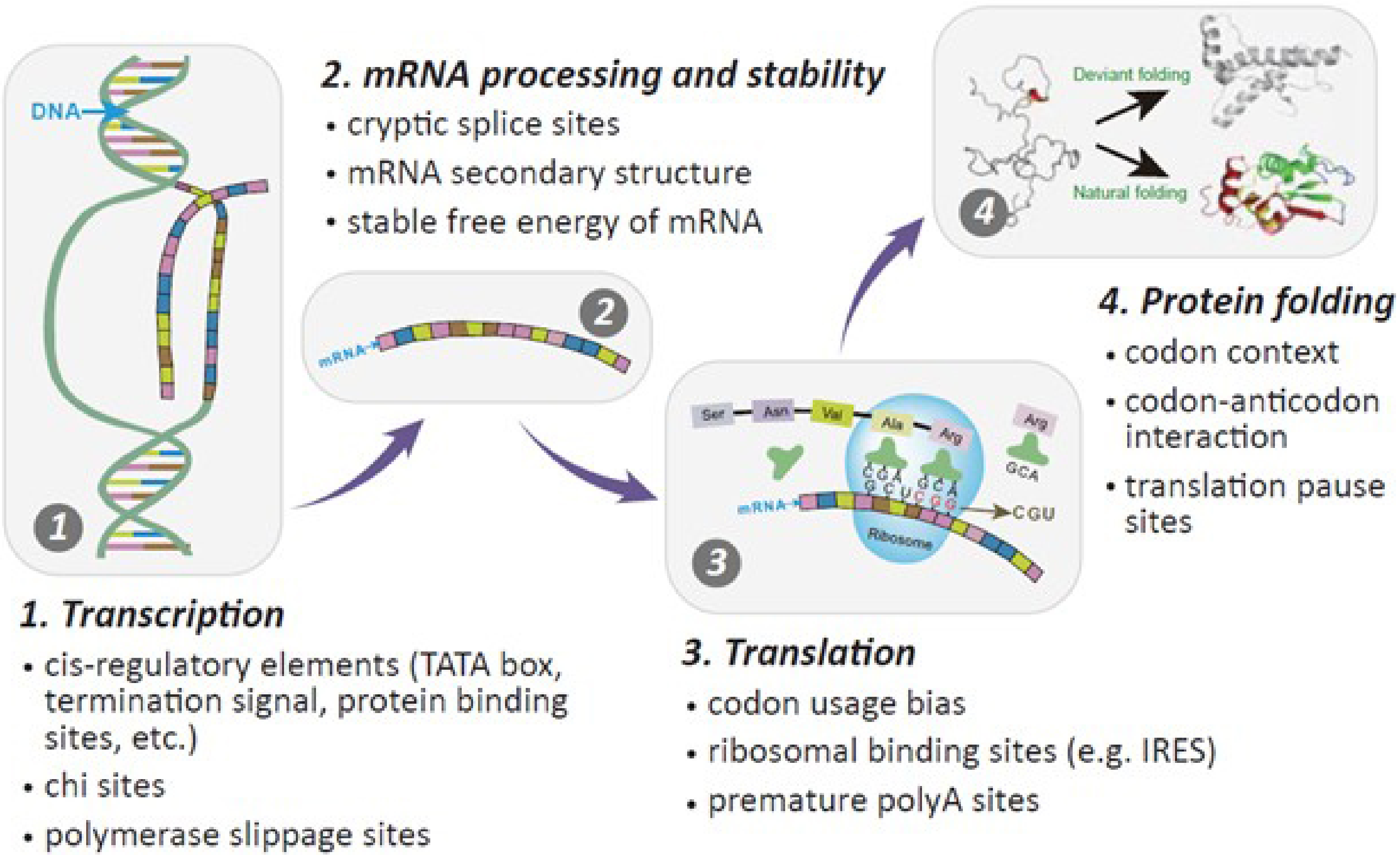
Coding-sequence determinants of gene expression in Escherichia coli.
Kudla G, Murray AW, Tollervey D, Plotkin JB. *Science.* 2009 Apr 10;324(5924):255-8.

- 154 synthetic GFP genes with random synonymous mutations
- 250-fold variation in fluorescence
- 44% of variation explained by 5' mRNA free energy (nt −4 to +37)

The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria.
Li GW, Oh E, Weissman JS. *Nature.* 2012 Mar 28;484(7395):538-41.

- Variation in Translation Rate does not correlate with rare codon use
- Orthogonal ribosomes with altered anti-SD sequences: pausing results from hybridization between 16s rRNA and SD-like sequences in mRNA

Many sequence features influence protein expression



Evidence-Based Codon Optimization: OptimumGene

Transcriptional Efficacy:

- GC content
- CpG dinucleotides content
- Cryptic splicing sites
- Negative CpG islands
- SD sequence
- TATA boxes
- Terminal signal

Translation Efficiency:

- Codon usage bias
- GC content
- mRNA secondary structure
- Premature PolyA sites
- RNA instability motif (ARE)
- Stable free energy of mRNA
- Internal chi sites and ribosomal binding sites

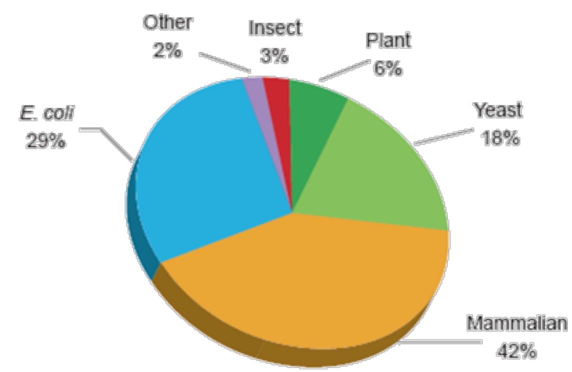
Protein Refolding:

- Codon usage bias
- Interaction of codon and anti-codon
- Codon-context
- RNA secondary structures

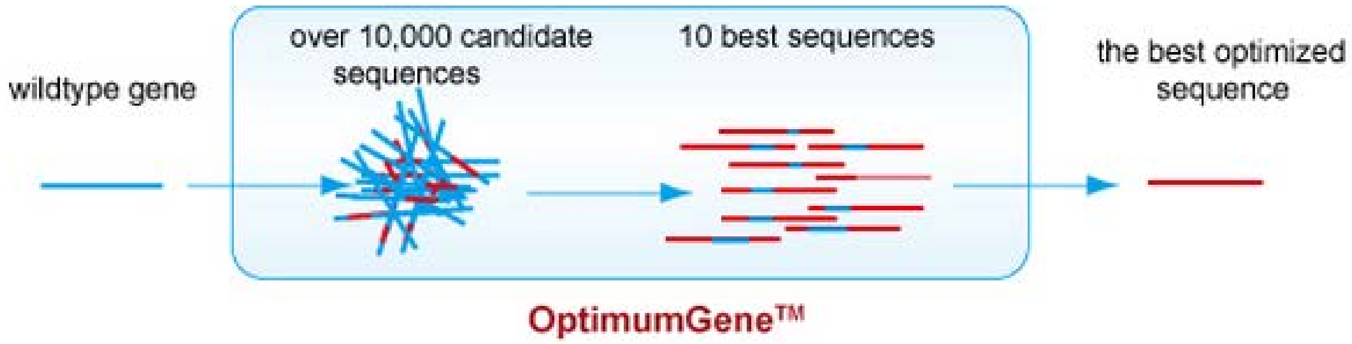
Flexibility to adjust the weight of different factors or add customized constraints:

- Filter out restriction sites
- Reduce similarity between library members
- Alternative codon tables / condition-specific codon preferences

GenScript has optimized over 50,000 sequences in all major expression systems.



Patented Bioinformatic Algorithm powers OptimumGene



Liu *et al.* **Method of sequence optimization for improved recombinant protein expression using a particle swarm optimization algorithm**. US Patent 8,326,547, issued December 4, 2012.

OptimumGene™ Improves Protein Expression Better that Competitors' Optimization

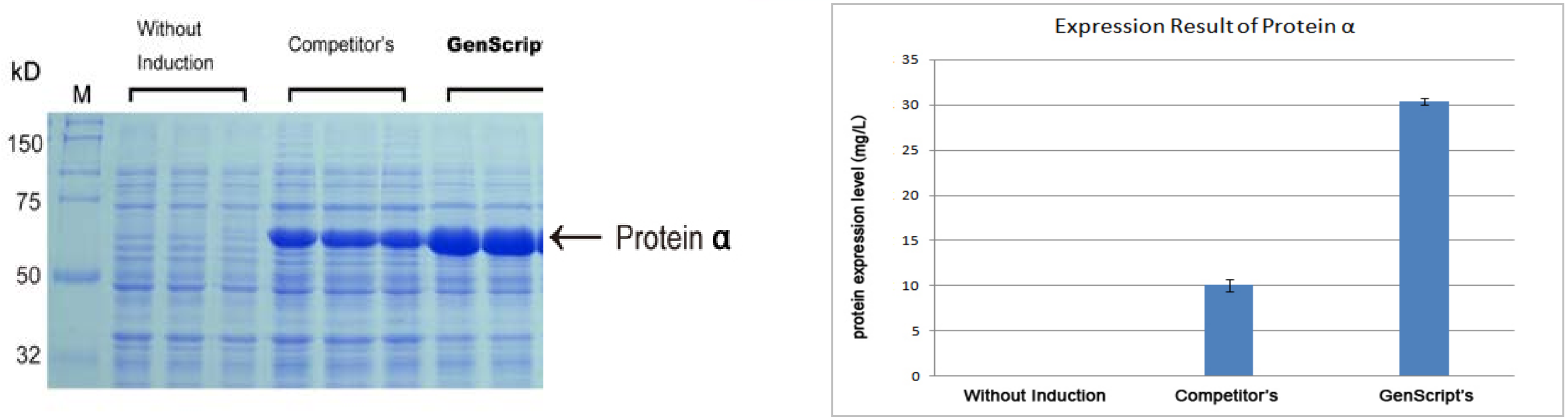


Fig. 1: Expression Result of Protein α after Codon Optimization. The expression level of Protein α using GenScript's OptimumGene™ Codon Optimization is **3** times more than that of competitor's.

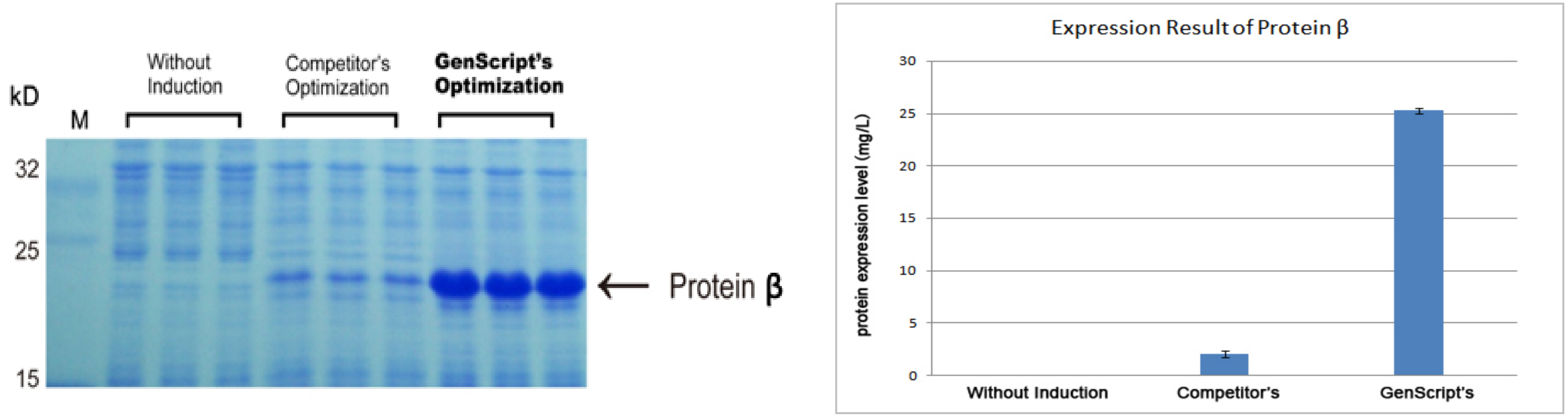
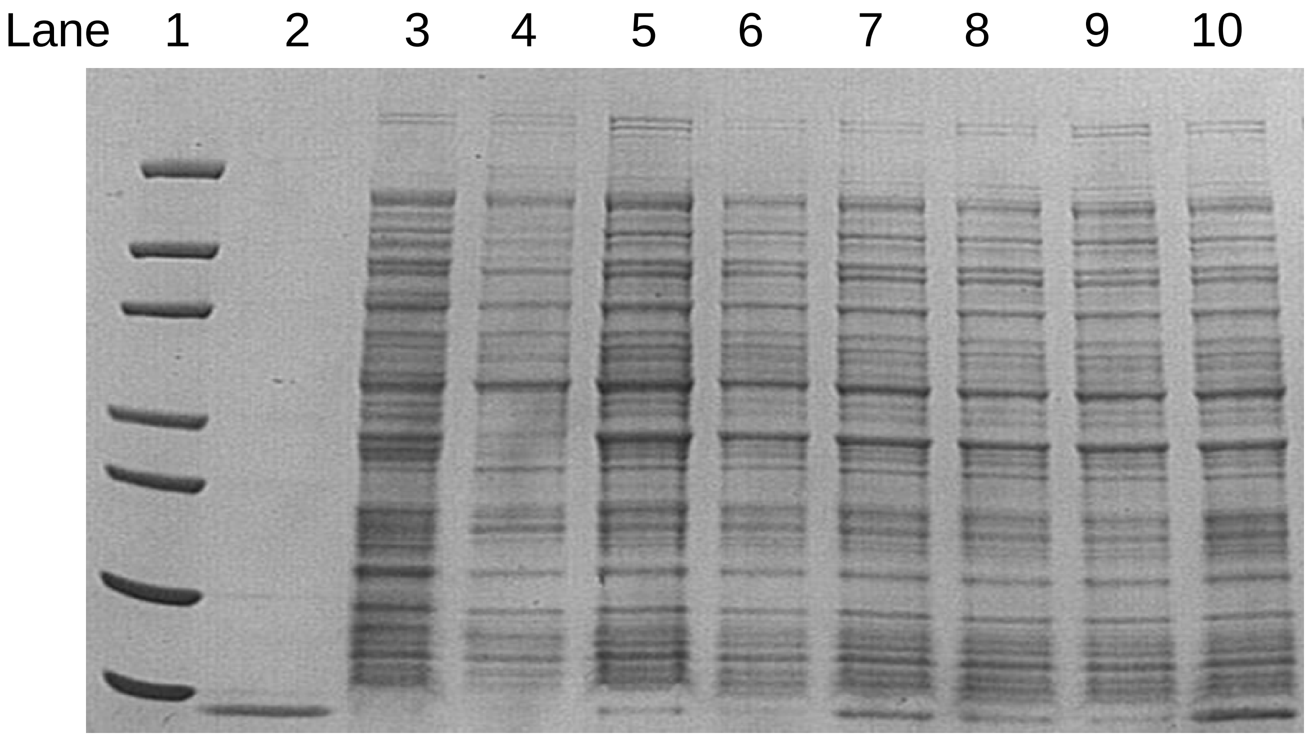


Fig. 2: Expression Result of Protein β after Codon Optimization. The expression level of Protein β using GenScript's OptimumGene™ Codon Optimization is **13** times more than that of competitor's.

We routinely test and improve our algorithm

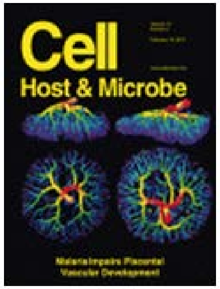


Lane	1	2	3	4	5	6	7	8	9	10
Sequence	MW marker	Purified hIGF-1 (PC)	BL21 cell lysate (NC)	WT (non-optimized) hIGF-1	Opti-0	Opti-1	Opti-2	Opti-3	Opti-4	Opti-5
Yield (mg/L)	--	--	--	Not detectable	7.7	3.1	18.5	11.4	5.4	28.5

Hundreds of papers cite GenScript for codon-optimized gene synthesis

"Humanization and optimization of codon usage was performed (GenScript) owing to **poor expression of the original zebrafish lyn in HEK293 cells**."

"…IFP1.4 gene was de novo synthesized by **GenScript** Company, based on the available protein sequence. The DNA sequence was **optimized with proprietary OptimumGene algorithm (GenScript)**..."



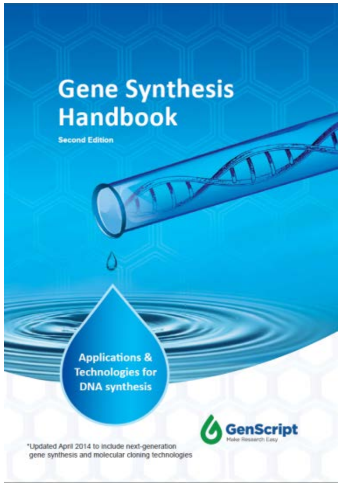
"...The following genes were **codon optimized and synthesized (Genscript):…"**

Resources » Reference Databases » Citations Database

By Category

- ▼Peptide Services(*1862*)
- ▼Antibody Engineering(*1*)
- ▼Animal Model Services(*4*)
- ▼Bio-Assay Center(*3*)
- ▼Cell Line Services(*6*)
- ▼Antibody Services(*628*)
- ▼Gene Services(*3169*)
- ▼Protein Services(*160*)
- ▼Bioinformatics Tools(*332*)
- ▼Catalog Products(*3042*)

View all (*10456*)



Case Study 1: Neurospora Circadian Clock

Non-optimal codon usage affects expression, structure and function of clock protein FRQ

Zhou M. *et al. Nature.* 2013 Mar;495 (7439); 111 – 5

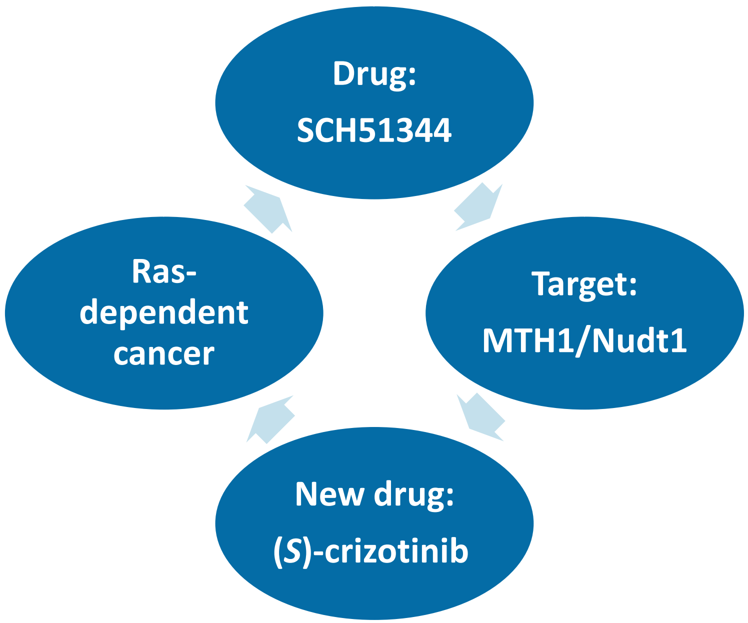
Codon Optimization performed in only specific regions of the gene:

"…**Optimized frq sequences (synthesized by Genscript)** …In the m1-frq construct, **only the codons upstream of the predicted intron branch point were optimized** as m-frq. For the m2-frq construct, **only the codons downstream of the intron 3' end were optimized** as m-frq..."

Case Study 2: Crystal Structure & PK/PD for Cancer Drug Discovery

Stereospecific targeting of MTH1 by (*S*)-crizotinib as anticancer strategy

Huber KV, et al. *Nature.* 2014 Apr 10;508(7495):222-7.



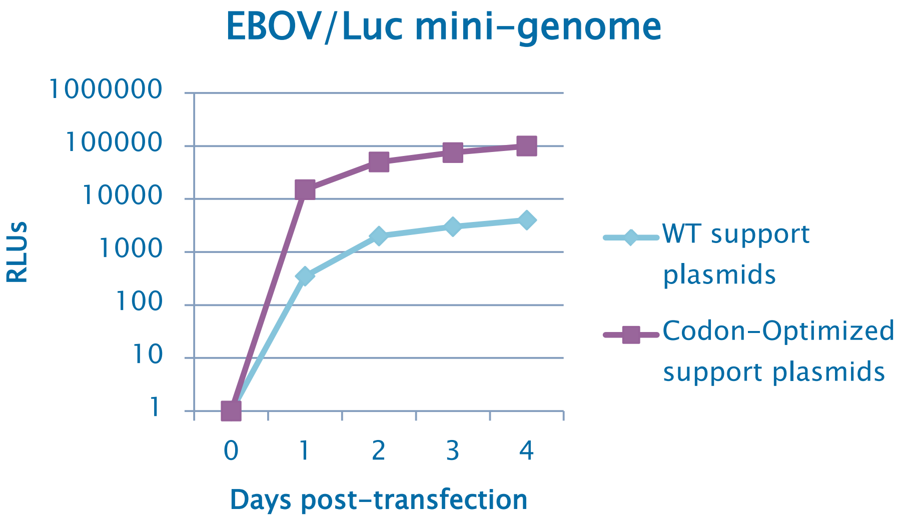
- **Codon-optimized gene synthesis from GenScript** was used to express Nudt1 for **enzymatic assays** and **crystallization studies.**

Case Study 3: Antiviral Drug Discovery for Ebola Virus

High-throughput, luciferase-based reverse genetics systems for identifying inhibitors of Marburg and Ebola viruses.

Uebelhoer *et al. Antiviral Res*. 2014 Jun;106:86-94.

- **Codon-optimized EBOV gene provided by GenScript**
- **Codon-optimized support plasmids increased signal 2000-fold**



Case Study 4: Combinatorial Library for Human Genome Editing

nature biotechnology

A library of TAL effector nucleases spanning the human genome.

Kim Y, et al. *Nat. Biotechnol.* **31,** 251–258 (2013).

Table I. TALE specificity code

RVD	Nucleotide
NI	adenine
HD	cytosine
NG	thymine
NN	guanine

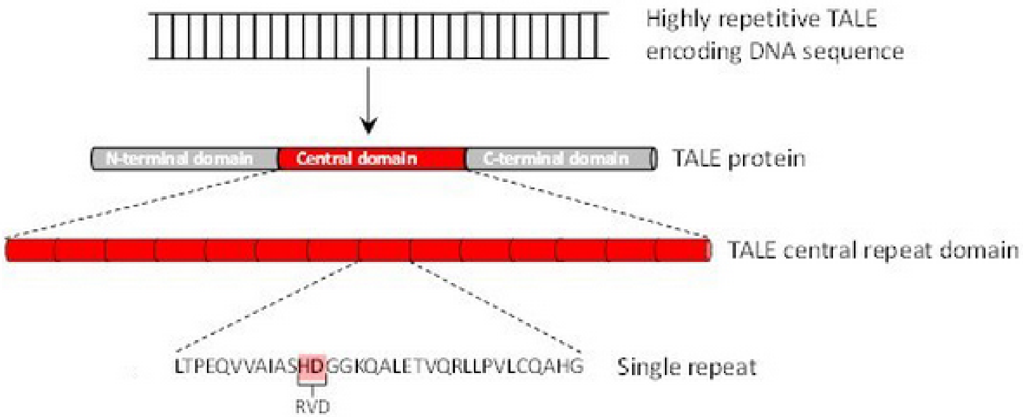


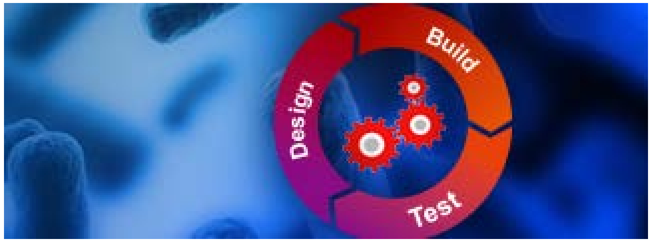
Fig 2. TALE protein organization

•**Codon Optimized Gene Synthesis from GenScript** was used to

1. Limit sequence similarity
2. Exclude rare codons
3. Guarantee accuracy of highly-repeated sequences

Case Study 5: Metabolic Engineering

Substantial improvements in methyl ketone production in E. coli and insights on the pathway from in vitro studies.
Goh EB *et al. Metab Eng.* 2014 Sep 18;26C:67-76.



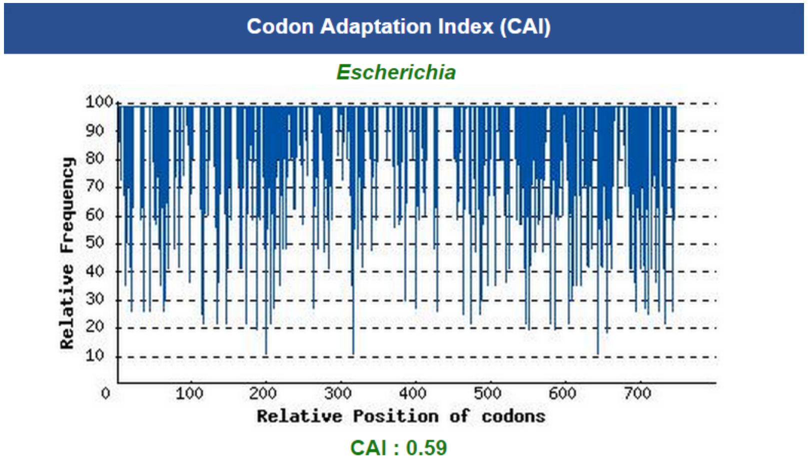
•**Codon Optimized Gene Synthesis from GenScript** was used to

1. improve metabolic pathway efficiency (↑substrate influx, ↓diversion)
2. improve GC content of gene from *M. luteus*, whose genome is 73% GC

How to get codon-optimized genes

- Online Tools to identify rare codons

- Rare Codon Analysis Tool
- Codon Frequency Tables

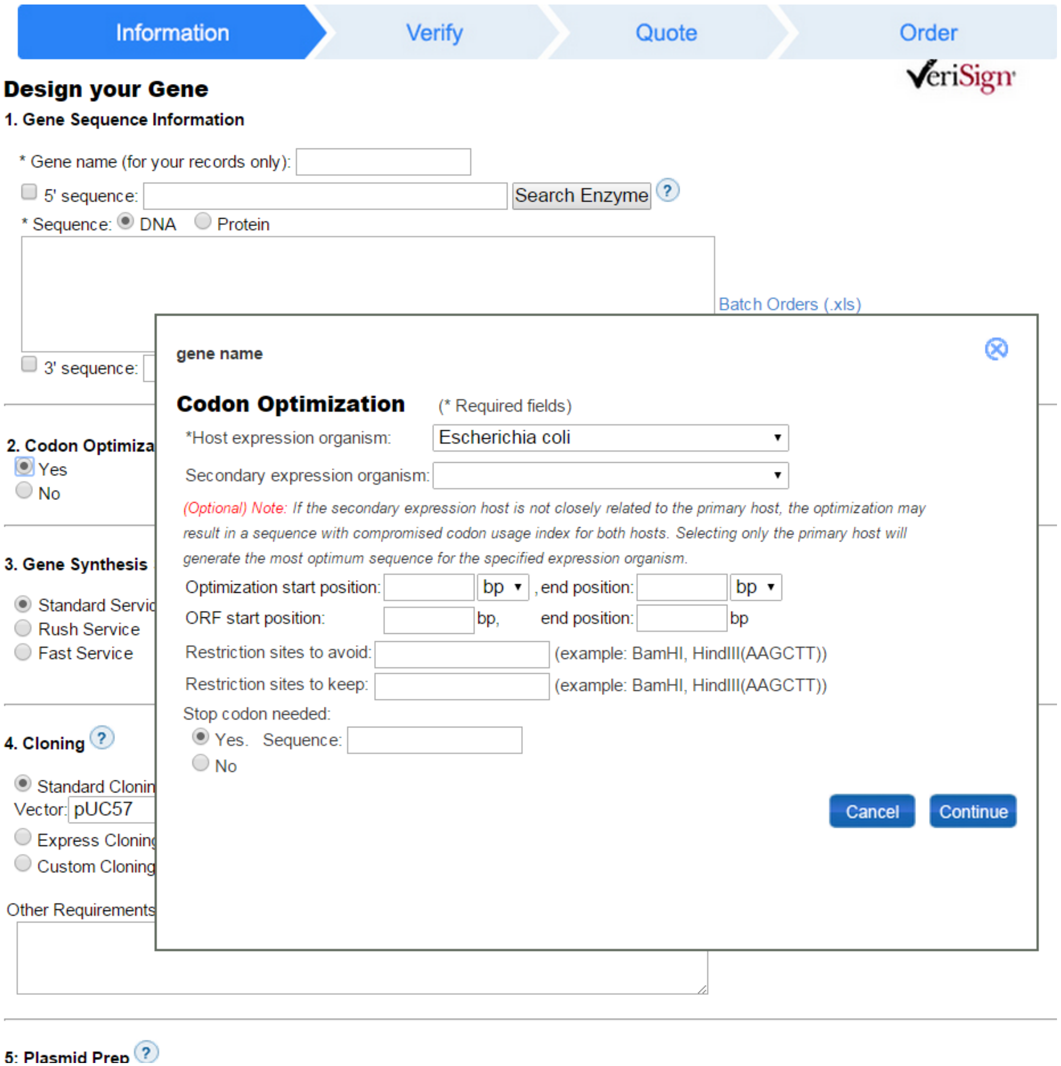


•Request Free Codon Optimization using OptimumGene

Request Free Codon Optimization

Get a Quote Now
Via Secured Online Quotation

- Quick & easy online form



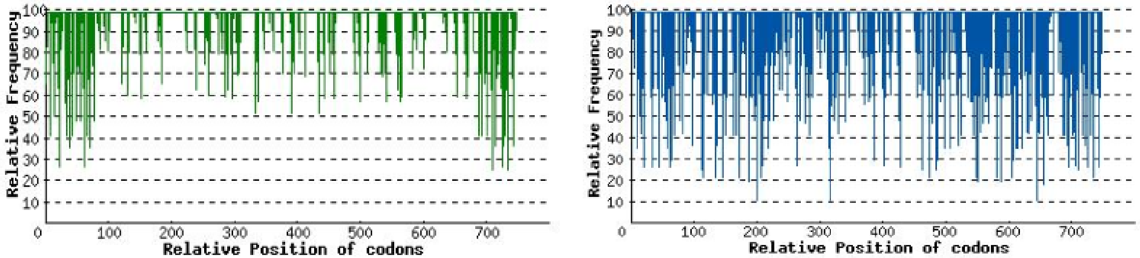
Email: gene@genscript.com

Phone: 1-877-436-7274 (Toll-Free)

Review your Free Optimization Report

Codon Adaptation Index (CAI)

After OptimumGene™ Optimization

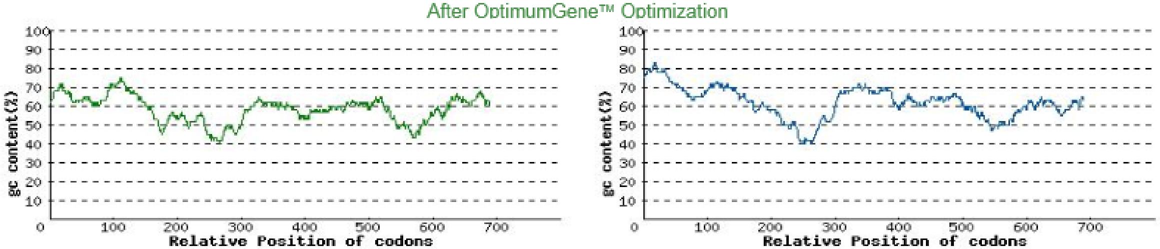


0.83 CAI: 0.59

Figure 1a. The distribution of codon usage frequency along the length of the gene sequence. A CAI of 1.0 is considered to be perfect in the desired expression organism, and a CAI of > 0.8 is regarded as good, in terms of high gene expression level.

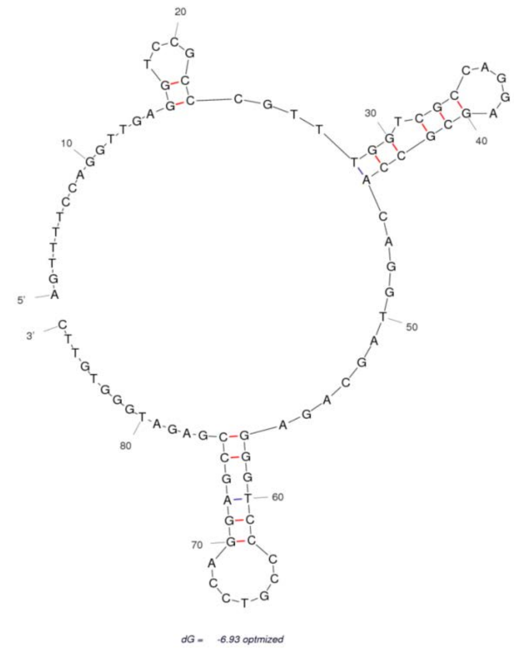
GC Content Adjustment

After OptimumGene™ Optimization



58.21 Average GC content: 61.82

Figure 2. The ideal percentage range of GC content is between 30-70 %. Peaks of %GC content in a 60 bp window have been removed.



Helices in structure (all)

Helix	ΔG (kcal/mol)	Length	Position
1	-4.74	5	64-->68 ; 85<--81
2	-4.08	3	16-->18 ; 42<--40
3	-3.12	3	19-->21 ; 38<--36
4	-2.17	2	69-->70 ; 76<--75
5	-2.17	2	24-->25 ; 33<--32
6	-1.84	2	53-->54 ; 59<--58
7	-1.84	2	43-->44 ; 49<--48

Hairpins in structure (all)

Hairpin	ΔG (kcal/mol)	Length	Position
1	3.10	10	24-->...<--33
2	2.50	8	69-->...<--76
3	1.50	7	53-->...<--59
4	1.50	7	43-->...<--49

Order Gene Synthesis for your Codon-Optimized Gene

Recommended Services for your needs:	Low Price	Fast Turnaround	High-Volume	Long Genes
Custom Gene Synthesis Cat No. SC1010	✓ $0.35/bp	✓ 9 business days	No min / max	≤8 kb
Rush Gene Synthesis Cat No. SC1575	Request a quote	✓ 4 business days	No min / max	≤2 kb
GenPlus™ High-Throughput Gene Synthesis Cat No. SC 1645	✓ $0.23/bp	15 business days	✓ ≥25 genes	✓ ≤10 kb
GenPlus™ Economy Gene Synthesis Cat No. SC1681	✓ $0.23/bp	25 business days	No min / max	✓ ≤10 kb
GenBrick™ Synthesis Cat No. SC1584	$0.45/bp	23 business days	No min / max	✓ 8 - 15kb or more

GenScript Toolkit For Improving Protein Expression

Select/Design the end product (amino acid sequence)	GenPlus™ high-throughput gene synthesis Gene Variant Library services
↓	
Choose expression system	PROTential™ protein expression evaluation service
↓	
Design expression clone (DNA construct)	Codon-Optimized Gene Synthesis
↓	
Express the protein	BacPower™ YeastHIGH™ MamPower™ InsectPower™ FragPower™ Recombinant Antibody
↓	
Purify the protein	FoldArt™ Refolding ToxinEraser™ Endotoxin Removal
↓	
Characterize the protein	Protein Characterization Services

Related GenScript Webinars

Select/Design the end product (amino acid sequence)

↓

Choose expression system

↓

Design expression clone (DNA construct)

↓

Express the protein

↓

Purify the protein

↓

Characterize the protein

Mutant library for protein engineering?
Combinatorial Library?
Truncation Variants?

Gene variant libraries: design, construction, and research applications

Presented by: Rachel Speer, Ph.D.
Originally aired May 21st and June 18th, 2014

On Demand

View now

Related GenScript Webinars

Select/Design the end product
(amino acid sequence)

↓

Choose expression system

↓

Design expression clone
(DNA construct)

↓

Express the protein

↓

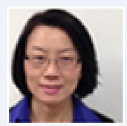
Purify the protein

↓

Characterize the protein

- Pros and Cons of different expression hosts
- Techniques for protein re-folding, protection from rapid degradation

Recombinant protein expression & purification: challenges and solutions



Presented by: Liyan Pang, Ph.D.
Originally aired June 11th and June 12th, 2014

On Demand

View now

Related GenScript Webinars

Select/Design the end product (amino acid sequence)

↓

Choose expression system

↓

Design expression clone (DNA construct)

↓

Express the protein

↓

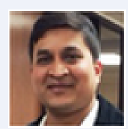
Purify the protein

↓

Characterize the protein

- Fusion partners/epitope tags
- e. coli strain, induction conditions, etc

Optimizing conditions for recombinant soluble protein production in *E. coli*



Presented by: Keshav Vasanthavada
Originally aired May 8th and June 24th, 2014

On Demand

View now

Related GenScript Webinars

Select/Design the end product (amino acid sequence)

↓

Choose expression system

↓

Design expression clone (DNA construct)

↓

Express the protein

↓

Purify the protein

↓

Characterize the protein

- maximizing purity and yield

Identify the optimal protein purification strategy for your recombinant protein production

Presented by: Liyan Pang, Ph.D.

November 12, 2014 8:00 am	November 12, 2014 2:00 pm
Register now	Register now

GenScript – The most cited biology CRO

Gene Services

Peptide Services

Protein Services

Antibody Services

Discovery Biology Services

Catalog Products

GenScript
Make Research Easy

References

Belfield EJ, Hughes RK, Tsesmetzis N, Naldrett MJ, Casey R **The gateway pDEST17 expression vector encodes a -1 ribosomal frameshifting sequence.** *Nucleic Acids Res.* 2007;35(4):1322-32.

Chu D *et al.* **Translation elongation can control translation initiation on eukaryotic mRNAs**. *EMBO J.* 2014 Jan 7;33(1):21-34.

Goh EB et al. **Substantial improvements in methyl ketone production in E. coli and insights on the pathway from in vitro studies.** *Metab Eng.* 2014 Sep 18;26C:67-76.

Huber KV, et al. **Stereospecific targeting of MTH1 by (*S*)-crizotinib as anticancer strategy.** *Nature.* 2014 Apr 10;508(7495):222-7

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

Kramer G, Boehringer D, Ban N, Bukau B. **The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins.** *Nat Struct Mol Biol.* 2009;16:589–597.

Kudla G, Murray AW, Tollervey D, Plotkin JB. **Coding-sequence determinants of gene expression in Escherichia coli.** *Science.* 2009 Apr 10;324(5924):255-8. Free Full Text

Li GW, Oh E, Weissman JS. **The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria.** *Nature.* 2012 Mar 28;484(7395):538-41. Free Full Text

Mellitzer A. *et al.* **Synergistic modular promoter and gene optimization to push cellulase secretion by Pichia pastoris beyond existing benchmarks**. *J. Biotechnol.* (2014), http://dx.doi.org/10.1016/j.jbiotec.2014.08.035

Plotkin JB, Kudla G. **Synonymous but not the same: the causes and consequences of codon bias.** *Nat Rev Genet.* 2011;12:32–42. Free Full Text

Shine J, Dalgarno L. **The 3′-terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites.** *Proc. Natl Acad. Sci. USA* 1974;71:1342-1346. Free Full Text

Uebelhoer *et al.* **High-throughput, luciferase-based reverse genetics systems for identifying inhibitors of Marburg and Ebola viruses.** *Antiviral Res*. 2014 Jun;106:86-94.

Welch M et al. **Design parameters to control synthetic gene expression in Escherichia coli**. *PLoS One*. 2009 Sep 14;4(9):e7002. doi: 10.1371/journal.pone.0007002.

Thank you!

- Please complete the survey
- Questions/feedback: rachel.speer@genscript.com
- Webinar Archives: www.genscript.com/webinars.html
- Request codon optimization: gene@genscript.com