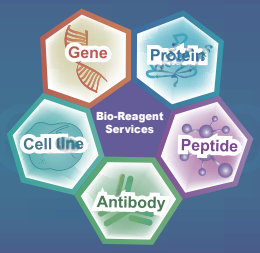
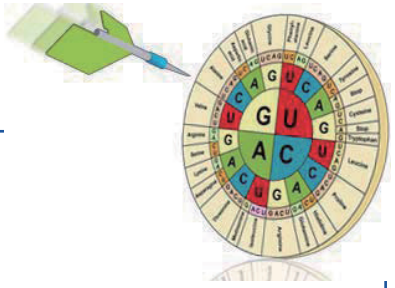


GenScript
Make Research Easy



Quotation #:

Gene name: **sp_Q899R2_SSB_CLOTE**

Customer:

Optimized for expression in: ***E. coli***

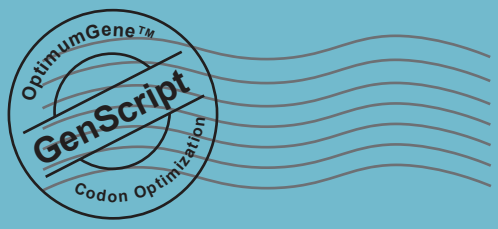
Gene length: **492**

Optimization region: **7-486**

Analysis conducted by: **Jason Zhou, Ph.D**

Analysis created: **07/17/2017 02:02:28**

QA: James



OptimumGene™ Codon Optimization Analysis

Optimization Parameters

OptimumGene™ algorithm optimizes a variety of parameters that are critical to the efficiency of gene expression, including but not limited to:

- Codon usage bias
- GC content
- CpG dinucleotides content
- mRNA secondary structure
- Cryptic splicing sites
- Premature PolyA sites
- Internal chi sites and ribosomal binding sites
- Negative CpG islands
- RNA instability motif (ARE)
- Repeat sequences (direct repeat, reverse repeat, and Dyad repeat)
- Restriction sites that may interfere with cloning

Additional sequences we propose to improve translational performance:

(1) To increase the efficiency of translational initiation

- Kozak sequence
- Shine-Dalgarno Sequence

(2) To increase the efficiency of translational termination

- Stop codon (TAA)

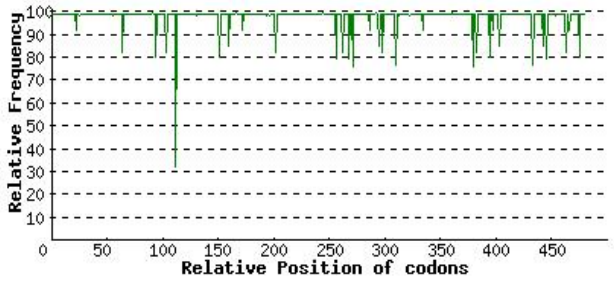
860 Centennial Ave., Piscataway, NJ 08854, USA
Tel: 1-732-885-9188 Fax: 1-732-210-0262 Email: gene@genscript.com Web: www.genscript.com

Results *E. coli*

1. Codon usage bias adjustment

Codon Adaptation Index (CAI)

After OptimumGene™ Optimization



CAI: 0.96

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.

Frequency of Optimal Codons (FOP)

After OptimumGene™ Optimization

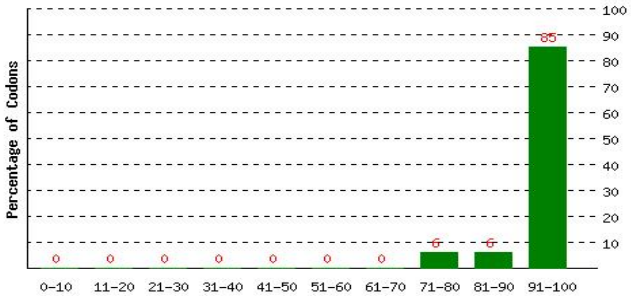
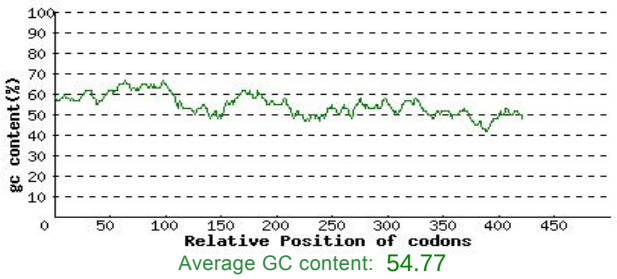


Figure 1b. The percentage distribution of codons in computed codon quality groups. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism.

2. GC Content Adjustment

GC Content Adjustment

After OptimumGene™ Optimization



Average GC content: 54.77

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.

3. Restriction Enzymes and CIS-Acting Elements

Restriction Enzymes	Optimized
** Green: filtered sites; Blue: checked sites (not filtered); Red: kept sites.*	
BtgI(CCRYGG)	0
NcoI(CCATGG)	0
BamHI(GGATCC)	0
EcoRI(GAATTC)	0
SacI(GAGCTC)	0
PstI(CTGCAG)	0
SalI(GTCGAC)	0
HindIII(AAGCTT)	0
NotI(GCGGCCGC)	0
AflII(CTTAAG)	0
NdeI(CATATG)	1(1)
BglII(AGATCT)	0
EcoRV(GATATC)	0
AatII(GACGTC)	0
KpnI(GGTACC)	0
XhoI(CTCGAG)	1(487)
AvrII(CCTAGG)	0
Polymerase slippage site 1	0
Polymerase slippage site 2	0
Frameshift element	0
Ribosome binding site	0

CIS-Acting Elements	Optimized
E.coli_RBS(AGGAGG)	0
PolyT(TTTTTT)	0
PolyA(AAAAAAA)	0
Chi_sites(GCTGGTGG)	0
T7Cis(ATCTGTT)	0
SD_like(GGRGGT)	0

4. Remove Repeat Sequences

After Optimization

Max Direct Repeat: Size:16 Distance:3 Frequency:2
Max Inverted Repeat: None
Max Dyad Repeat: None

5. Optimized Sequence(Optimized Sequence Length:492, GC%:54.77)

CATATG
CACCACCACCACCACCACCTGGAAGTTCTGTTCCAGGGTCCGATGAACCGTGTGGTTCTGATCGGCCGTCTGACC
AAGGACCCGGAGCTGAAATTTACCCCGGGCACCGGTACGGCGGTGACCACCTTCGTTCTGGCGGTGGACCGTCGT
TTTAGCAAGGATGGTAAAAACGAAGCGGACTTCATCCCGGTGGTTGTTTGGGGCAAGCAGGCGGAGAGCACCGCG
AACTACATGAGCAAGGGTAAACTGATCGGTATTAGCGGCCGTATTCAAACCCGTAGCTACGAAGCGAAAGATGGC
ACCCGTCGTTATGTTACCGAGGTTGTGGCGGACGAAGTGAAGTTCCTGGAGTGGGGTAACAAACAGAGCAGCGGT
AGCCAAGGCTTCAACAACTTTGAAAGCGATCCGCTGAGCTACAACAACGAGGACAACTATAACGACGATATTACC
CCGGTTGACGAGGGCGAAGTGCCGTTTTAA
CTCGAG

Conclusion

A wide variety of factors regulate and influence gene expression levels, and our OptimumGene™ algorithm takes into consideration as many of them as possible, producing the single gene that can reach the highest possible level of expression.

In this case, the native gene employs tandem rare codons that can reduce the efficiency of translation or even disengage the translational machinery. We increased the codon usage bias in E. coli by upgrading the CAI to 0.96 . GC content and unfavorable peaks have been optimized to prolong the half-life of the mRNA. The Stem-Loop structures, which impact ribosomal binding and stability of mRNA, were broken. In addition, our optimization process has screened and successfully modified those negative cis-acting sites as listed in the introduction.

We are honored to deliver the analysis that you requested. We hope that you are pleased with your GenScript OptimumGene™ results.

Supplementary

1. Protein Sequence

HHHHHHLEVLFQGPMNRVVLIGRLTKDPELKFTPGTGTAVTTFVLAVDRRFSKDGKNEAD
FIPVVVWGKQAESTANYMSKGKLIGISGRIQTRSYEAKDGTRRYVTEVVADEVKFLEWGN
KQSSGSQGFNNFESDPLSYNNEDNYNDDITPVDEGEVPF*