

Quotation #:

Gene name: **sp\_O69302\_SSB\_CAMJE**

Customer:

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

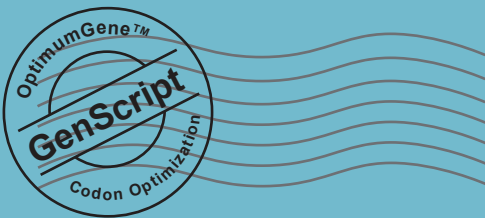
Gene length: **606**

Optimization region: **7-600**

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

Analysis created: **07/17/2017 02:02:09**

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

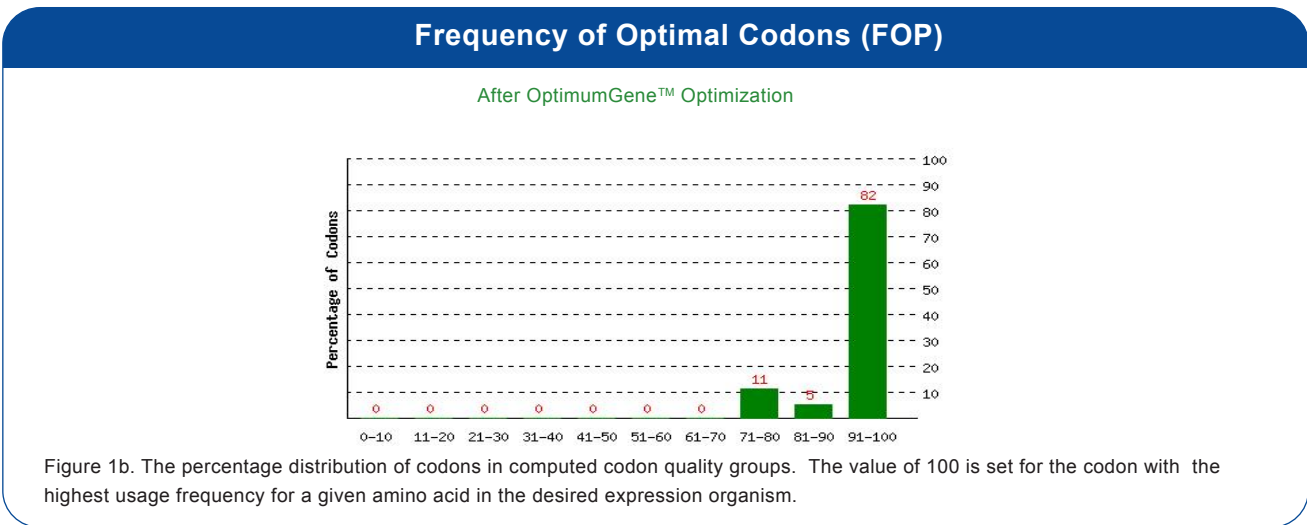
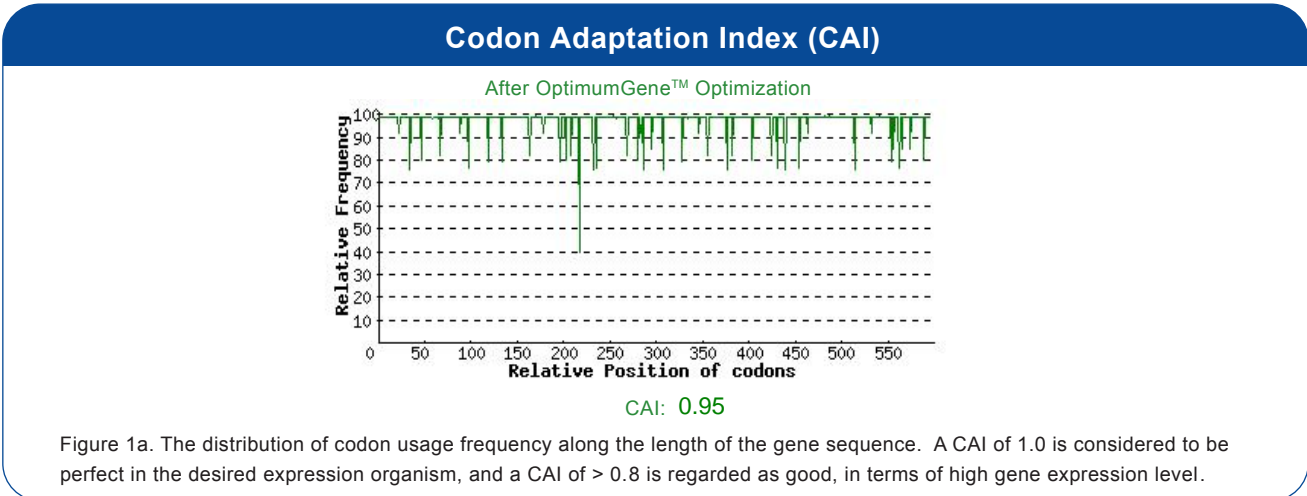
## Additional sequences we propose to improve translational performance:

- (1) To increase the efficiency of translational initiation
  - Kozak sequence
  - Shine-Dalgarno Sequence
- (2) To increase the efficiency of translational termination
  - Stop codon (TAA)

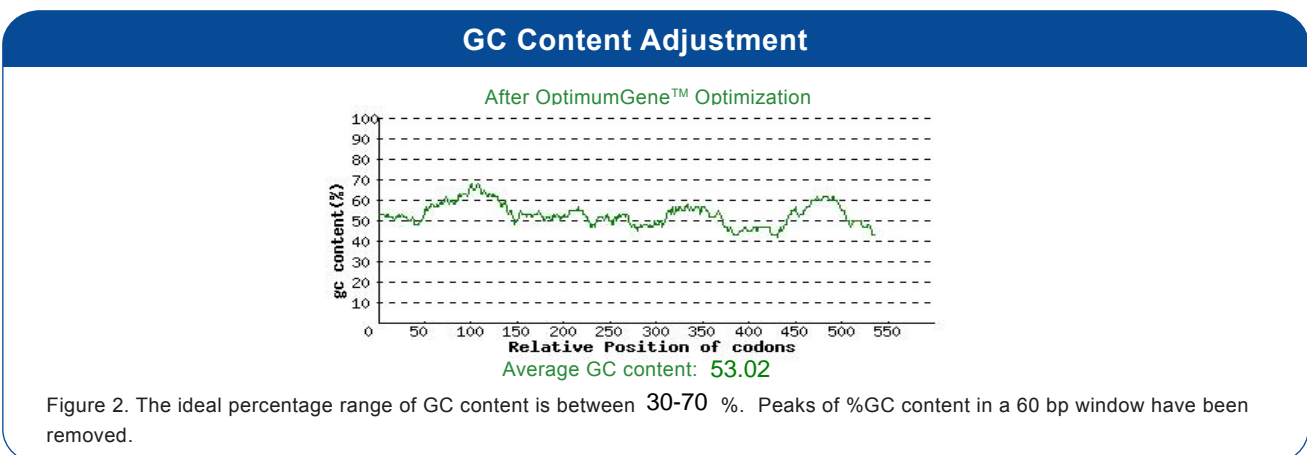


## Results *E. coli*

### 1. Codon usage bias adjustment



### 2. GC Content Adjustment



### 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
Sall(GTCGAC)	0
HindIII(AAGCTT)	0
NotI(GCGGCCGC)	0
AflIII(CTTAAG)	0
NdeI(CATATG)	1(1)
BglII(AGATCT)	0
EcoRV(GATATC)	0
AatII(GACGTC)	0
KpnI(GGTACC)	0
XhoI(CTCGAG)	1(601)
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(AAAAAA)	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:606, GC%:53.02)

CATATG

CACCACCACCACCACCACCTGGAAGTGCTGTTCCAAGGTCCGATGTTTAAACAAGGTGGTTCTGGTTGGCAACCTG  
ACCCGTGACATCGAAATGCGTTATGCGCAGAGCGGTAGCGCGATTGGTGCGAGCGCGATTGCGGTGACCCGTCGT  
TTCACCGGAACGGCGAGAAACGTGAGGAAACCTGCTTCATCGACATTAGCTTTTACGGCCGTACCGCTGAGGTG  
GCGAACCAATATCTGACCAAGGGTAGCAAAGTTCTGATCGAGGGCCGTCTGCGTTTGAACAATGGAGCGATCAG  
AACGGTCAAAACCGTAGCAAGCACAGCATTTCAGGTTGAGAACATGGAAATGCTGGGCAACAGCAACGCGCCGCAG  
CAAGGTGGCAACTTCGGTAACAACAGCTTTAGCAACAACAACACTACAGCGGCAACTATGAGAACCAAAGCTACGAC  
CCGTATATGAGCGAAAACCAGAACTTCAACAAAGCGAAAGCGAACC CGCGCCGCAGCGTAACCAAAGCCCGCAG  
CACGAGGAAAAGCTGAAAGAGATCGACATTGATGCGTATGATAGCGACGATACCAACCTGCCGTTTAA

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 **0.95**. 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.

**GenScript Recombinant Protein Expression Service (Bacteria, Mammalian, Insect, Yeast)**

*High quality recombinant protein for your research!*

## Supplementary

### 1. Protein Sequence

HHHHHHLEVLFGQPMFNKVVLVGNLTRDIEMRYAQSGSAIGASAI AVTRRF TANGEKREE  
TCFIDISFYGR TAEVANQYLTKGSKVLI EGRLRFEQWSDQNGQNRSKHSIQVENMEMLGN  
SNAPQQGGNFGNNSFSNNNYSNGNYENQSYDPYMSENQNFN KAKANPAPQRNQSPQHEEKL  
KEIDIDAYDSDDTNLPF\*