

Your GeneOptimizer® Assisted Sequence Analysis



Customer: **Universität Regensburg - Jennifer Flechsler**
Gene name: **Igni_0379**
Scientist: **HaiQuyen Arenskoetter**
Optimized for expression in: ***Escherichia coli***
Analysis created: **8/28/2014 (8:36 PM)**

Construct No: **14AGDG6C**
Gene length: **1437 bp**
Ref. No: **1574578**



Gene design according to DIN EN ISO 9001:2008

The codon usage was adapted to the codon bias of *Escherichia coli* genes.

In addition, regions of very high (> 80 %) or very low (< 30 %) GC content have been avoided where possible.

During the optimization process the following cis-acting sequence motifs were avoided where applicable:

- internal TATA-boxes, chi-sites and ribosomal entry sites
- AT-rich or GC-rich sequence stretches
- RNA instability motifs
- repeat sequences and RNA secondary structures
- (cryptic) splice donor and acceptor sites in higher eucaryotes

Non-optimized

Codon Quality Distribution

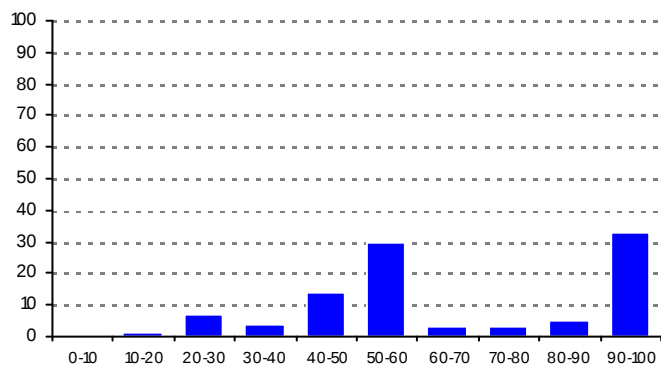


Fig. 1a

Codon Adaptation Index : **0.61**

Optimized with GeneOptimizer®

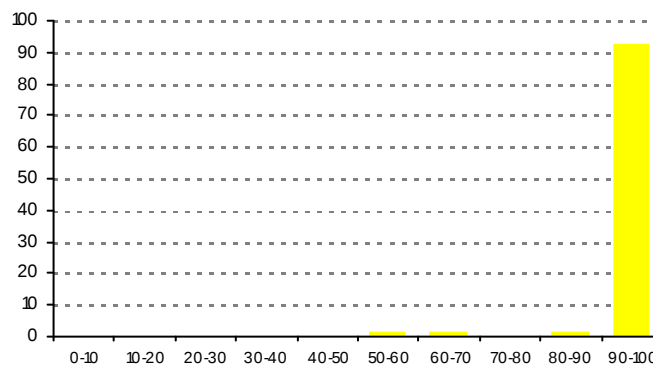


Fig. 1b

Codon Adaptation Index : **0.97**

The histograms show the percentage of sequence codons which fall into a certain quality class. The quality value of the most frequently used codon for a given amino acid in the desired expression system is set to 100, the remaining codons are scaled accordingly (see also Sharp, P.M., Li, W.H., *Nucleic Acids Res.* 15 (3), 1987).

Non-optimized

Codon Quality Plot

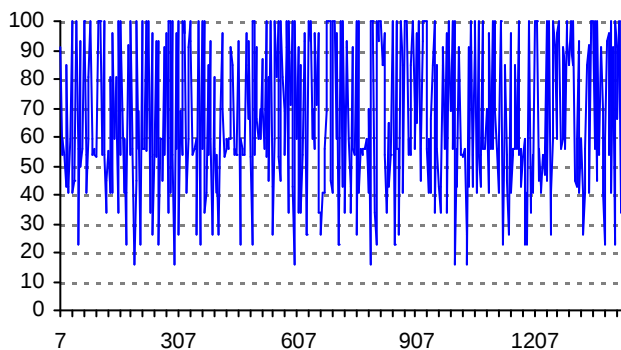


Fig. 2a

Optimized with GeneOptimizer®

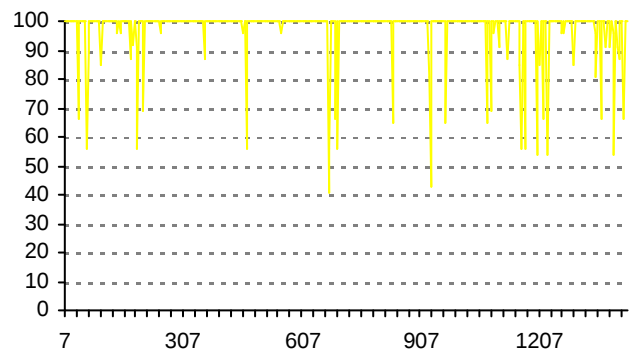


Fig. 2b

The plots show the quality of the used codon at the indicated codon position.

GC Content

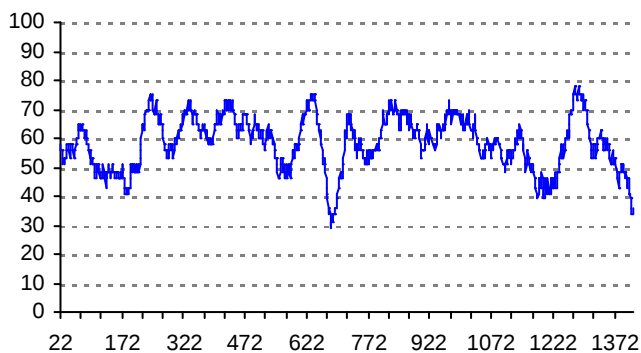


Fig. 3a

Average GC content: **58 %**

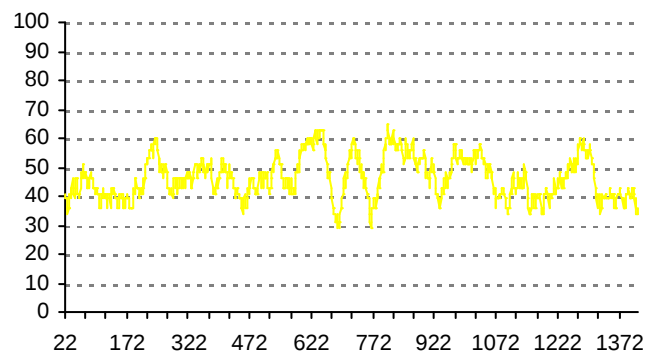


Fig. 3b

Average GC content: **46 %**

The plots show the GC content in a 40 bp window centered at the indicated nucleotide position.

Summary

The optimization was successful: Negative cis-acting sites (such as splice sites, TATA-boxes, etc.) which may negatively influence expression were eliminated wherever possible. GC content was adjusted to prolong mRNA half life. Codon usage was adapted to the bias of *Escherichia coli* resulting in a CAI* value of **0.97**. The optimized gene should therefore allow high and stable expression rates in *Escherichia coli*.

Please note: for mammalian, insect and yeast expression, we recommend to add a KOZAK sequence upstream of the construct.

* CAI: The parameter CAI (codon adaptation index) describes how well the codons match the codon usage preference of the target organism. Thus, a CAI of 1.0 would be perfect. However, a CAI of > 0.9 is considered as very good (i.e. allowing high expression).

** In some rare cases negatively cis acting motifs cannot be removed due to the particular amino acid sequence or restriction site and codon quality constraints. In all of these cases the optimization scientist chooses an optimal compromise for each individual site.

GeneArt®'s gene optimization process is certified according to the standards of **ISO 9001** (registration no. 12 100 24212). Our internal quality management system covers all aspects from gene optimization to gene synthesis, DNA sequencing as well as sales and service. I certify to the best of my knowledge that: (1) no special safety precautions are needed for the production of the below mentioned DNA and (2) the below mentioned DNA sequence will be used in accordance with applicable laws and regulations.

I agree that GeneArt® shall synthesize the following sequence 14AGDG6C according to the conditions described in the quotation, including the presence or absence of:

- stop codons, as per my request or lack thereof
- 5', 3' and internal restriction sites for downstream cloning and processing, as per my request or lack thereof

date

signature

Optimized Sequence

Customer: **Universität Regensburg - Jennifer Flechsler**

Gene name: **Igni_0379**

Construct No: **14AGDG6C**

Gene length: **1437 bp**

Optimized for expression in: ***Escherichia coli***

Analysis created: **8/28/2014 (8:36 PM)**

```

      NheI                      AgeI
1  GCTAGCGTTTATCATGATCCGGGTAAACCGGTTCAAAAATCATTGAAGAACGTGCACGT
   -----+-----+-----+-----+-----+-----+-----+
   CGATCGCAAATAGTACTAGGCCCATTTGGCCAAGTATTTTAGTAACCTCTTGCACGTGCA
      V   Y   H   D   P   G   K   P   V   H   K   I   I   E   E   R   A   R
      1     3     5     7     9    11    13    15    17

                                     BglII
61  GAGAGTCCGGATAAAGTTGCACTGGTTTATGAAGATGGCACC GAACTGACCTATAAAGAT
   -----+-----+-----+-----+-----+-----+-----+
   CTCTCAGGCCTATTTCAACGTGACCAAATACTTCTACCGTGGCTTGACTGGATATTTCTA
   E   S   P   D   K   V   A   L   V   Y   E   D   G   T   E   L   T   Y   K   D
   19    21    23    25    27    29    31    33    35    37

121  CTGATTAGCAAAGCAAAGCCGTTGCCAATCTGCTGTATGAAAATAAAGTTCGCAAATGC
   -----+-----+-----+-----+-----+-----+-----+
   GACTAATCGTTTTCGTTTTCGGCAACGGTTAGACGACATACTTTTATTTCAAGCGTTTACG
   L   I   S   K   A   K   A   V   A   N   L   L   Y   E   N   K   V   R   K   C
   39    41    43    45    47    49    51    53    55    57

181  GATACCGTGTTCTGATTATGTTTAACCGTCCGGAATTTGTTACCAGCCTGCTGGGTGCA
   -----+-----+-----+-----+-----+-----+-----+
   CTATGGCACAAGGACTAATACAAATTGGCAGGCCTTAAACAATGGTCGGACGACCCACGT
   D   T   V   F   L   I   M   F   N   R   P   E   F   V   T   S   L   L   G   A
   59    61    63    65    67    69    71    73    75    77

                                     BglII
241  CTGTATGCCGGTGCACGTGTTGTTATTGTTGATGCACTGACCCAGAAAGAAGATCTGGCA
   -----+-----+-----+-----+-----+-----+-----+
   GACATACGGCCACGTGCACAACAATAACAAC TACGTGACTGGGTCTTTCTTCTAGACCGT
   L   Y   A   G   A   R   V   V   I   V   D   A   L   T   Q   K   E   D   L   A
   79    81    83    85    87    89    91    93    95    97

      PvuII                      PstI
301  TTTTCAGCTGAATGATAGCGCACCGAAAGTTGTTCTGGCAGATGAAGAAGTTCTGCAGCGC
   -----+-----+-----+-----+-----+-----+-----+
   AAAGTCGACTTACTATCGCGTGGCTTTCAACAAGACCGTCTACTTCTTCAAGACGTGCGG
   F   Q   L   N   D   S   A   P   K   V   V   L   A   D   E   E   V   L   Q   R
   99    101   103   105   107   109   111   113   115   117
```

BglII

361 GAAAGCGAAACCTGAAAGGTTATGCAGTTCTGACCGAAAAAGATCTGAGCAGCGCAAGC
 -----+-----+-----+-----+-----+-----+-----+
 CTTTCGCTTTGGGACTTTCCAATACGTCAAGACTGGCTTTTTCTAGACTCGTCGCGTTGC
E S E T L K G Y A V L T E K D L S S A S
 119 121 123 125 127 129 131 133 135 137

BspMI

421 GGTGAACATGAAGTTGTTGTTGGTTATGAAGATGATGCCCGTGTCTTCTATTATGCAGGT
 -----+-----+-----+-----+-----+-----+-----+
 CCACTTGTACTTCAACAACAACCAATACTTCTACTACGGGCACAAAAGATAATACGTCCA
G E H E V V V G Y E D D A R V F Y Y A G
 139 141 143 145 147 149 151 153 155 157

BspMI *BspMI* *AgeI*

481 ATTGCAGGTCGTACCATGCAGGTTATTCATAGCCATCGTAGCTTTACCGGTGCAATTATG
 -----+-----+-----+-----+-----+-----+-----+
 TAACGTCCAGCATGGTACGTCCAATAAGTATCGGTAGCATCGAAATGGCCACGTTAATAC
I A G R T M Q V I H S H R S F T G A I M
 159 161 163 165 167 169 171 173 175 177

BstEII

541 CCGCTGGTTCAGGCCGAAGGTATTAATGCAAATGATGTTAGCCTGGTTACCGTTCCGCTG
 -----+-----+-----+-----+-----+-----+-----+
 GGCGACCAAGTCCGGCTTCCATAATTACGTTTACTACAATCGGACCAATGGCAAGGCGAC
P L V Q A E G I N A N D V S L V T V P L
 179 181 183 185 187 189 191 193 195 197

PflMI

601 ACCCATGTTCTGGGTCTGGATGCAGCACTGCTGAGCGCACTGGTTAGCGGTGCAACCGCA
 -----+-----+-----+-----+-----+-----+-----+
 TGGGTACAAGACCCAGACCTACGTCGTGACGACTCGCGTGACCAATCGCCACGTTGGCGT
T H V L G L D A A L L S A L V S G A T A
 199 201 203 205 207 209 211 213 215 217

661 CTGCTGCTGAAAAAGTTCAATATGGATAAAATCAAAGAGATGGTTGCACGTCGTAGCCCG
 -----+-----+-----+-----+-----+-----+-----+
 GACGACGACTTTTTCAAGTTATACCTATTTTAGTTTCTCTACCAACGTGCAGCATCGGGC
L L L K K F N M D K I K E M V A R R S P
 219 221 223 225 227 229 231 233 235 237

721 ACCTATCTGGTTGCAGTTCCGCTGGTTTTTTCAGACCCTGATGAAAGAAGATGAAGAATTT
 -----+-----+-----+-----+-----+-----+-----+
 TGGATAGACCAACGTCAAGGCGACCAAAAAGTCTGGGACTACTTTCTTCTACTTCTTAAA
T Y L V A V P L V F Q T L M K E D E E F
 239 241 243 245 247 249 251 253 255 257

781 GTTAAAAGCCTGGGTCTAGAGCCTGCGTTGGGCAATGAGCGGTGGTGCATATCTGCCTCCG
 -----+-----+-----+-----+-----+-----+-----+
 CAATTTTCGGACCCAGTCTCGGACGCAACCCGTTACTCGCCACCACGTATAGACGGAGGC
V K S L G Q S L R W A M S G G A Y L P P
 259 261 263 265 267 269 271 273 275 277

BclI *AgeI* *PstI* *BspMI*

841 GATGATCAGCGTCGTTGGGAAGAACTGACCGGTAAACCGCTGCTGCAGGTTTATGGTATG
 -----+-----+-----+-----+-----+-----+-----+
 CTACTAGTCGCAGCAACCCTTCTTGACTGGCCATTTGGCGACGACGTCCAAATACCATAC
D D Q R R W E E L T G K P L L Q V Y G M
 279 281 283 285 287 289 291 293 295 297

ACCGAAGCACCGCAGATTTTTGCAACCACACCCGAAAAACATAAAATTGGTAGCCTGGGT
 901 -----+-----+-----+-----+-----+-----+
 TGGCTTCGTGGCGTCTAAAAACGTTGGTGTGGGCTTTTTGTATTTAACCATCGGACCCA
T E A P Q I F A T T P E K H K I G S L G
 299 301 303 305 307 309 311 313 315 317

TTTCCGCTGCCTGGTGTGAAGCACTGCTGGTTAATCCGGAAACCCTGGAACCGGTTGAA
 961 -----+-----+-----+-----+-----+-----+
 AAAGGCGACGGACCACAACCTTCGTGACGACCAATTAGGCCTTTGGGACCTTGGCCAACCT
F P L P G V E A L L V N P E T L E P V E
 319 321 323 325 327 329 331 333 335 337

GATCAGGGTGAAGTCTGGTTCGTGGTCCGCAGGTTATGAAATGTTATCCGGATCCTGAA
 1021 -----+-----+-----+-----+-----+-----+
 CTAGTCCCACTTGACGACCAAGCACCAGGCGTCCAATACTTTACAATAGGCCTAGGACTT
D Q G E L L V R G P Q V M K C Y P D P E
 339 341 343 345 347 349 351 353 355 357

GAAAACAGCAAAGCCTTTGTTGAAGTGAATGGTAAAGTTTGGATGCGCACCGGTGATATT
 1081 -----+-----+-----+-----+-----+-----+
 CTTTTGTCGTTTCGGAAACAACCTTCACTTACCATTTCAAACCTACGCGTGGCCACTATAA
E N S K A F V E V N G K V W M R T G D I
 359 361 363 365 367 369 371 373 375 377

CTGGCAAAAGATGAAGAAGGCTTCTATTATTTCCGTGGTGTTCGTAAACGTATGCTGAAA
 1141 -----+-----+-----+-----+-----+-----+
 GACCGTTTTCTACTTCTTCCGAAGATAATAAAGGCACCACAAGCATTTGCATACGACTTT
L A K D E E G F Y Y F R G V R K R M L K
 379 381 383 385 387 389 391 393 395 397

TACAAAGGCTATCCGATCTTTCCGAAAGACCTGGAAGTCTGCTGAAACATCCGTGT
 1201 -----+-----+-----+-----+-----+-----+
 ATGTTTCCGATAGGCTAGAAAGGCTTTCTGGACCTTGACTAAGACGACTTTGTAGGCACA
Y K G Y P I F P K D L E L I L L K H P C
 399 401 403 405 407 409 411 413 415 417

GTTGCCGAAGCCGAAGTTGTTGGTGAACCGGCAGGCGAAGTTGGTCAGATTCCGGTTGCA
 1261 -----+-----+-----+-----+-----+-----+
 CAACGGCTTCGGCTTCAACAACCACTTGGCCGTCCGCTTCAACAGTCTAAGGCCAACGT
V A E A E V V G E P A G E V G Q I P V A
 419 421 423 425 427 429 431 433 435 437

AAAGTTAAACTGAAAGAAGGTTGTAAAGCCACACCGGAAGAAATCATGGATTTTGTGAAT
 1321 -----+-----+-----+-----+-----+-----+
 TTTCAATTTGACTTTCTTCCAACATTTCCGTGTGGCCTTCTTTAGTACCTAAAACACTTA
K V K L K E G C K A T P E E I M D F V N
 439 441 443 445 447 449 451 453 455 457

AAACGTGTGGCAGCCTACAAAAAAGTGCGCAAAGTTATTATCCTGGAATAACCATGG
 1381 -----+-----+-----+-----+-----+-----+
 TTTGCACACCGTCCGATGTTTTTTCACGCGTTTCAATAATAGGACCTTATTGGTACC
K R V A A Y K K V R K V I I L E *
 459 461 463 465 467 469 471 473 475

Biosecurity Compliance

GeneArt® routinely screens all its incoming orders for biosecurity according to international guidelines. We therefore reserve the right to refuse any order or to involve government authorities, if a DNA sequence may be of dual-use nature.

Further reading: Buegl *et al.* (2007), DNA synthesis and biological security, Nature Biotechnology

Free pdf available at <http://dx.doi.org/doi:10.1038/nbt0607-627>