

Your GeneOptimizer® Assisted Sequence Analysis



Customer: **Universität Regensburg - Jennifer Flechsler**
Gene name: **Igni_0475**
Scientist: **Dr. Andre Rusch**
Optimized for expression in: ***Escherichia coli***
Analysis created: **6/10/2013 (3:37 PM)**

Construct No: **13ABGMWC**
Gene length: **1494 bp**
Ref. No: **1363429**



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

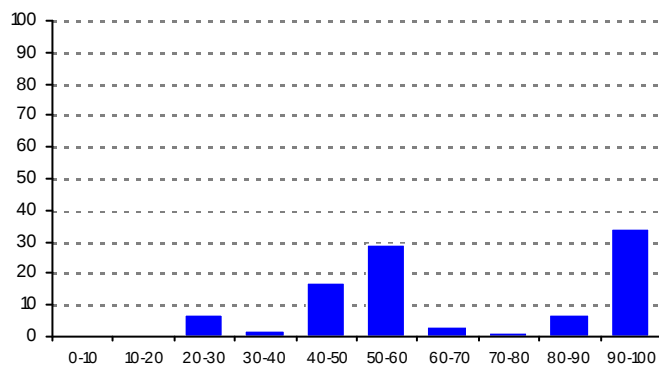


Fig. 1a

Codon Adaptation Index : **0.62**

Optimized with GeneOptimizer®

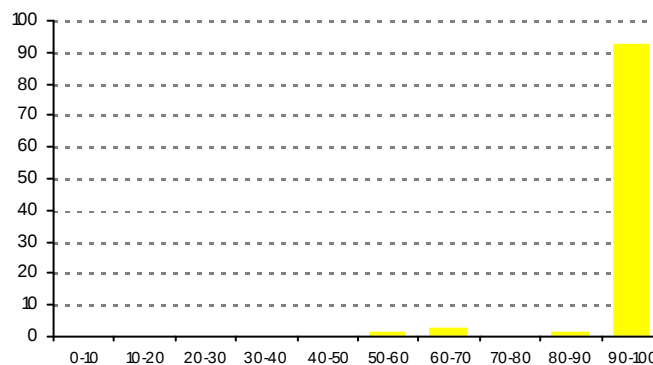


Fig. 1b

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

Optimized with GeneOptimizer®

Codon Quality Plot

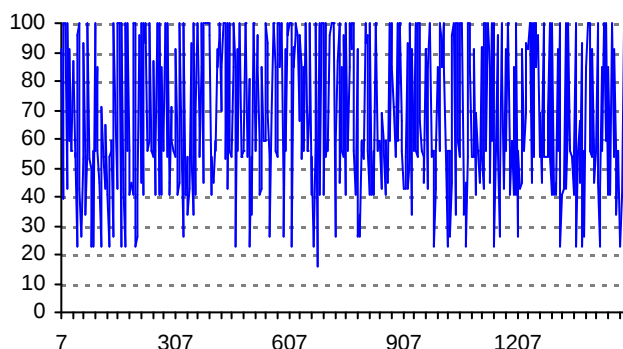


Fig. 2a

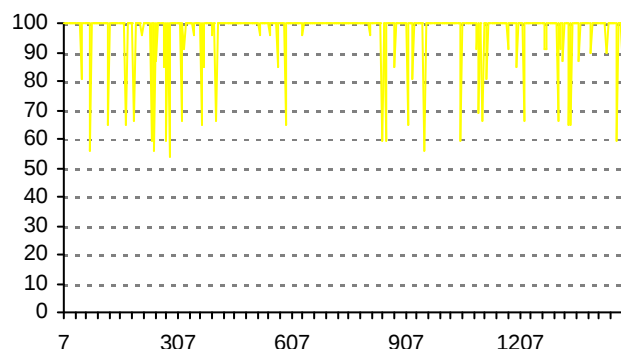


Fig. 2b

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

GC Content

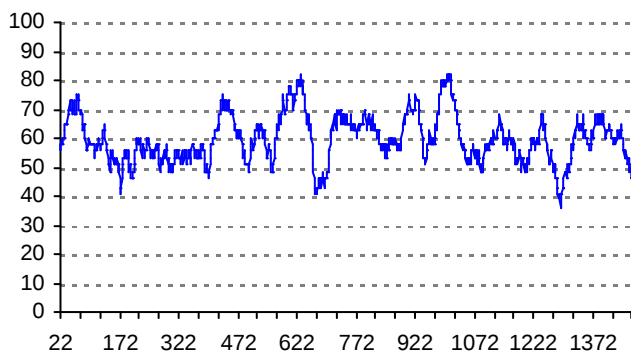


Fig. 3a

Average GC content: **60 %**

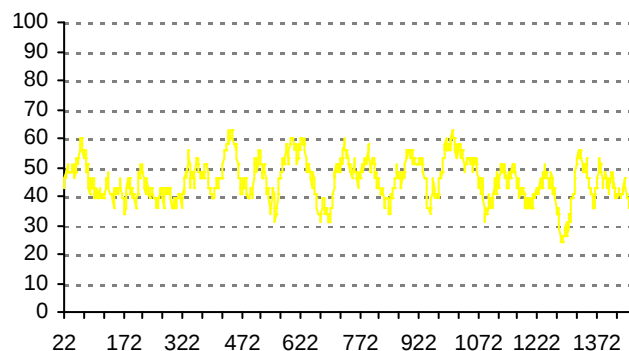


Fig. 3b

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 13ABGMWC 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_0475**

Construct No: **13ABGMWC**

Gene length: **1494 bp**

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

Analysis created: **6/10/2013 (3:37 PM)**

*Bam*HI

GGATCCGTTCTGAGCGGTCGTGCAGTTGTTTTGAACGTTATAGCCGTGCAACACCGCTG
1 -----+-----+-----+-----+-----+-----+-----+
CCTAGGCAAGACTCGCCAGCACGTCAACAAAACTTGCAATATCGGCACGTTGTGGCGAC
V L S G R A V V F E R Y S R A T P L
1 3 5 7 9 11 13 15 17

GGTGCATATCTGGAAGAGGGTCGTCGTTTTCTGTAAGAAGGTTTTATTCGTACCGATTAT
61 -----+-----+-----+-----+-----+-----+-----+
CCACGTATAGACCTTCTCCAGCAGCAAAAGCACTTCTTCCAAAATAAGCATGGCTAATA
G A Y L E E G R R F R E E G F I R T D Y
19 21 23 25 27 29 31 33 35 37

CCGCCTAAACGTGATGAAGTTCTGTGGAATCCGAAAATTATGCGTATGCCTCGTGAAGAA
121 -----+-----+-----+-----+-----+-----+-----+
GGCGGATTTGCACTACTTCAAGACACCTTAGGCTTTTAATACGCATACGGAGCACTTCTT
P P K R D E V L W N P K I M R M P R E E
39 41 43 45 47 49 51 53 55 57

CTGGAAAAAATCAAAACCCAGCGTCTGAAAGCCATTGTAAATGGGCATGGGAACACAGC
181 -----+-----+-----+-----+-----+-----+-----+
GACCTTTTTTAGTTTTGGGTGCGAGACTTTTCGTAACAATTTACCCGTACCCTTGTGTCG
L E K I K T Q R L K A I V K W A W E H S
59 61 63 65 67 69 71 73 75 77

*Eco*RI

GAATTCTATCGCAAATTTTGGAAAAGCAAAGGCTTTCACCCGGACATGATTAAACCCAT
241 -----+-----+-----+-----+-----+-----+-----+
CTTAAGATAGCGTTTAAACCTTTTCGTTTCCGAAAGTGGGCCTGTACTAATTTTGGGTA
E F Y R K F W K S K G F H P D M I K T H
79 81 83 85 87 89 91 93 95 97

*Bgl*II

*Pst*I

GAAGATGTTGTTAAAATCCCGGTGCTGACCAAAAAAGATCTGCGTGCCGATCTGCAGAAA
301 -----+-----+-----+-----+-----+-----+-----+
CTTCTACAACAATTTTAGGGCCACGACTGGTTTTTTCTAGACGCACGGCTAGACGTCTTT
E D V V K I P V L T K K D L R A D L Q K
99 101 103 105 107 109 111 113 115 117

361 AATCCGCCTTATGGCACCATTATGGTTCGGAACTGGCCAAATATATCCATTTTGTGGT
 -----+-----+-----+-----+-----+-----+-----+
 TTAGGCGGAATACCGTGGTAATACCAAGGCCTTGACCGGTTTATATAGGTAAAACAACCA
N P P Y G T I M V P E L A K Y I H F V G
 119 121 123 125 127 129 131 133 135 137

AgeI

421 GCAACCAGCGGTAGCACCGGTATGCCGACCTTTCAGGGTTGGGGTCGTGTTGAACTGGAT
 -----+-----+-----+-----+-----+-----+-----+
 CGTTGGTCGCCATCGTGGCCATACGGCTGGAAAGTCCCAACCCAGCACAACTTGACCTA
A T S G S T G M P T F Q G W G R V E L D
 139 141 143 145 147 149 151 153 155 157

481 TATTTTGAAGAACAGCAGGCACGTTATCTGTGGACCTTTGCCGGTGTTAAACCGGGTGT
 -----+-----+-----+-----+-----+-----+-----+
 ATAAACTTCTTGTCTCGTCCGTGCAATAGACACCTGGAAACGGCCACAATTTGGCCACAA
Y F E E Q Q A R Y L W T F A G V K P G V
 159 161 163 165 167 169 171 173 175 177

541 GTTTATGCCAATTATCTGAATATGAGCGGCTTTTATAGCTGGGGTCCGCCTCTGGTTGAA
 -----+-----+-----+-----+-----+-----+-----+
 CAAATACGGTTAATAGACTTATACTCGCCGAAAATATCGACCCAGGCGGAGACCAACTT
V Y A N Y L N M S G F Y S W G P P L V E
 179 181 183 185 187 189 191 193 195 197

601 ACCGCAATGTGGCGTTGTGGTGCAACCGCAATTGCCGGTGGTGGTGAACCTTTTTTACC
 -----+-----+-----+-----+-----+-----+-----+
 TGGCGTTACACCGCAACACCACGTTGGCGTTAACGGCCACCACCACTTTGGAAAAAATGG
T A M W R C G A T A I A G G G E T F F T
 199 201 203 205 207 209 211 213 215 217

NdeI *Pf1MI*

661 TGGAAGATCGTCATATGCTGATTATGAACTGTGGAAAGTTGATGTTTTTGCAACCACC
 -----+-----+-----+-----+-----+-----+-----+
 ACCTTTCTAGCAGTATACGACTAATACTTTGACACCTTTCAACTACAAAAACGTTGGTGG
W K D R H M L I M K L W K V D V F A T T
 219 221 223 225 227 229 231 233 235 237

BstEII

721 CCGTGGCTGCATCGTTGGATTGGTGAACAGGCAAACTGGAAGTTGGGTTACCCGTTT
 -----+-----+-----+-----+-----+-----+-----+
 GGCACCGACGTAGCAACCTAACCACCTTGTCGTTTTGACCTTCCAACCCAATGGGGCAAA
P W L H R W I G E Q A K L E G W V T P F
 239 241 243 245 247 249 251 253 255 257

781 AAAGTTCTGCTGCTGCATGGTGGTGCAGCAGCCGAAAAACCAAAGAAAACTGATGGAA
 -----+-----+-----+-----+-----+-----+-----+
 TTTCAAGACGACGACGTACCACCACGTCGTCGGCTTTTTTGGTTTCTTTTGAACCTT
K V L L L H G G A A A E K T K E K L M E
 259 261 263 265 267 269 271 273 275 277

NdeI

841 GTTCACCCGAATGTCGAAAAAGTTATTAGCGTTTGGGGCACCACCGATGGTCATATGGCA
 -----+-----+-----+-----+-----+-----+-----+
 CAAGTGGGCTTACAGCTTTTTCAATAATCGCAAACCCGTTGGTGGCTACCAGTATACCGT
V H P N V E K V I S V W G T T D G H M A
 279 281 283 285 287 289 291 293 295 297

901 ATTGAACCGCTGATGCACCGGGTACACTGGTTTTTTGGGAAGATACCCAGATTTTCGAT
 -----+-----+-----+-----+-----+-----+-----+
 TAAGTGGCGGACTACGTGGCCCATGTGACCAAAAAACCTTCTATGGGTCTAAAAGCTA
I E P P D A P G T L V F W E D T Q I F D
 299 301 303 305 307 309 311 313 315 317

BspMI

961 ATTGTTGTTCCGGGTAGCTATGAAGAACGTCCGGAAGGTGAACCGGCAGGTGAGGGTGA
 -----+-----+-----+-----+-----+-----+-----+
 TAACAACAAGGCCCATCGATACTTCTTGACGGCCTTCCACTTGGCCGTCCAGTCCCACTT
I V V P G S Y E E R P E G E P A G Q G E
 319 321 323 325 327 329 331 333 335 337

1021 CGTGGTGAACGATTGCAACCCTGCTGACCCACTATACCATGCCGCTGATTCGTTATAGC
 -----+-----+-----+-----+-----+-----+-----+
 GCACCACTTGACTAACGTTGGGACGACTGGGTGATATGGTACGGCGACTAAGCAATATCG
R G E L I A T L L T H Y T M P L I R Y S
 339 341 343 345 347 349 351 353 355 357

1081 CTGGGTGATTATGTGAAAAACGAATTTATCTGTGATCCGACACCGGAACTGGGTATTACC
 -----+-----+-----+-----+-----+-----+-----+
 GACCCACTAATACTTTTTGCTTAAATAGACACTAGGCTGTGGCCTTGACCCATAATGG
L G D Y V K N E F I C D P T P E L G I T
 359 361 363 365 367 369 371 373 375 377

1141 CATTGTCGTTTTGCAGAACTGATTCCGGGTCGTGTGGAATGGATGTTTAAAGTTAAAGGC
 -----+-----+-----+-----+-----+-----+-----+
 GTAACAGCAAAACGTCTTGACTAAGGCCAGCACACCTTACCTACAAATTTCAATTTCCG
H C R F A E L I P G R V E W M F K V K G
 379 381 383 385 387 389 391 393 395 397

PvuI *AgeI*

1201 AAACGCTGTTTCCGATCGTTGTTGAAAATGCAGTTAGCCAGATTCCGGATACCACCGGT
 -----+-----+-----+-----+-----+-----+-----+
 TTTGACGACAAAGGCTAGCAACAACTTTACGTCAATCGGTCTAAGGCCTATGGTGGCCA
K L L F P I V V E N A V S Q I P D T T G
 399 401 403 405 407 409 411 413 415 417

1261 GCATATAATATTGTGGTGTATGATGATAATATGGATAAACTGAAAATCAAAGTGGAAC
 -----+-----+-----+-----+-----+-----+-----+
 CGTATATTATAACACCACATACTACTATTATACCTATTTGACTTTTAGTTTCACCTTTGG
A Y N I V V Y D D N M D K L K I K V E T
 419 421 423 425 427 429 431 433 435 437

1321 CGCAAACCGATTCCGCCTCCGCCTGAATATGATAAAGAAGCACGCGAAATTATTGCACGT
 -----+-----+-----+-----+-----+-----+-----+
 GCGTTTGGCTAAGGCGGAGGCGGACTTATACTATTTCTTCGTGCGCTTTAATAACGTGCA
R K P I P P P P E Y D K E A R E I I A R
 439 441 443 445 447 449 451 453 455 457

1381 GAACTGGGTCTGAGTCCGGAAGATGTTGAAATTGAATGGATTGTCGCGGTGAATCAGTT
 -----+-----+-----+-----+-----+-----+-----+
 CTTGACCCAGACTCAGGCCTTCTACAACCTTAACTTACCTAAGCAGGCCCACTTAGTCAA
E L G L S P E D V E I E W I R P G E S V
 459 461 463 465 467 469 471 473 475 477

PstI *NcoI*

1441 TGGAAAGGTTATAAACTGCAGGTCTTTCTGGATGAACGCAAAAAATGACCATGG
 -----+-----+-----+-----+-----+-----+-----+
 ACCTTTCCAATATTTGACGTCCAGAAAGACCTACTTGCGTTTTTTACTGGTACC
W K G Y K L Q V F L D E R K K * _
 479 481 483 485 487 489 491 493

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>