OptGene™

- An accurate systems scale gene optimization tool

Introduction: Genetic Modification And Optimization

Biotechnology industry is now the fastest growing sector associated with both health and food systems. It involves making molecular changes to living or almost-living things. Scientists have now taken the idea of modifying organisms at the genetic level. Genetic modification involves the transfer of specific genes to new host organisms, so that these hosts are able to do new things, which is not possible with their own set of genes. It has the potential to revolutionize both medical/health and food/agriculture industries, and so change the lives of billions of people.

Production of genetically modified organisms (GMOs) to achieve higher productivity, disease resistance and other desirable properties is still based on naturally occurring gene sequences. Naturally occurring sequences prove futile in modern biotechnology, as there is increased focus on safety requirement for recombinant products and at the same time higher flexibility in protein design is needed. These gene sequences seldom meet the ever-growing demand for optimized yields in heterogeneous systems. The goal of the Gene Optimization is to create a gene with optimized features in order to have maximum expression in the intended organism.

In many organisms, codon usage has a great impact on expression efficiency. This is especially true for genes, which are to be expressed in organisms that are not related to the source organisms. To enhance the expression level of a foreign protein in a particular expression system (E. coli, Yeast, Insect, or Mammalian cell), it is very important to adjust the codon frequency of the foreign protein to match that of the host expression system. One classic example is GFP (Green Fluorescent Protein), which was optimized to achieve high-level of expression in mammalian cells. Codon optimization allows for maximum protein expression by increasing the translational efficiency of your gene. An optimized gene allows one to engineer desired properties such as codon usage for an expression host of choice and insert convenient restriction enzyme sites. This eliminates nasty signals, thereby doing away with the need to clone the existing gene.

The Problem

Each organism has its preferred choice of nucleotide usage to encode any particular amino acid, called as codon usage. Codon usage may influence the expression of genes at the level of translation, transcription or mRNA processing. Humans and plants vary in their codon preferences for translating mRNA into proteins. Plants, in general, prefer G and C rich codons as compared to mammals rich in A and T. Transformation of heterologous genes with low G and C content in plants often leads to very low yield. Researchers, therefore, are spending considerable amount of time and resources in the area of gene synthesis and its optimization.

Earlier, gene optimization was done manually. The procedure followed in general was tedious and time-consuming. DNA sequences were scanned (using GCG and MacVector) and the scanned data was imported into Excel sheets. The sequence was then modified as per codon usage tables. The changed sequence was again imported for scanning. The entire optimization process is carried out using GCG/MacVector and Excel sheets, back and forth.

The Need

A one-stop platform system that would import a sequence into the system, perform optimization and the optimized sequence can further be reviewed for restriction enzyme analysis, six frame and ORF analysis.

A Proficient Solution To The Problem Of Gene Optimization

Ocimum Biosolutions' OptGene™ aggregates the latest technology and information pertaining to gene optimization. OptGene™ can optimize sequences for protein expression using either local codon usage table or those from publicly available codon usage database. It can convert your amino acid sequence into a DNA sequence with overall codon usage similar to a specified organism, and also optimizes the RNA secondary structure, GC content, repetitive codons etc.

OptGene™ achieves optimization through:

- Adaptation of codon usage to that of the host

Ocimum Biosolutions
...enabling R&D™
www.ocimumbio.com

- Extensive mutagenesis insertion of restriction sites
- Removal of cryptic splice sites, RNA destabilizing patterns and other undesirable signals
- Removal of secondary structures in RNA
- Reduction of transcription regions in the unused frames

What OptGene™ Extends

Sequence Import

The user can create an experiment to work on a sequence. The experiment can contain any number of sequences. The sequence can be imported in the following ways:

- Type/Paste into the Sequence Editor
- Upload from a local file. The file can be in GenBank, EMBL, FASTA format or a plain text file with just the sequence
- Import from the database

Sequence Editor

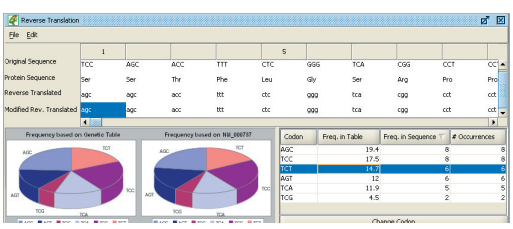
The sequence imported into the system can be modified using the sequence editor. This editor will provide options for find/replace, format and print. The sequence editor will allow valid characters, depending on the sequence type (DNA/Protein).

Translation

The sequence will be translated into the corresponding amino acid sequence using the selected Genetic Code table. The translated sequence will be displayed in either single letter or three-letter amino acid sequence.

Reverse Translation

This module reverse translates protein sequence(s) using different codon usage tables and genetic codon tables. Distribution of codons for all the amino acids may be the same as in the frequency tables, the highest frequency or as specified by the user. The reverse translated sequence can be compared with the original sequence and if necessary, be further optimized by substituting codons with their alternatives.

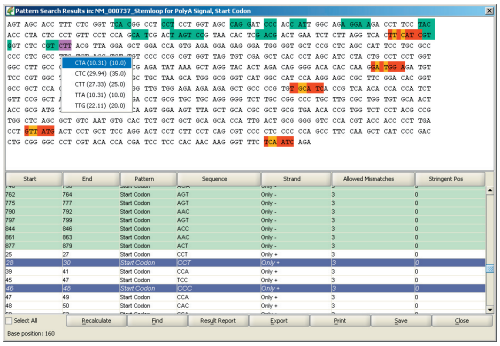


Destabilizing Signal Analyzer

This module allows researcher to search for and remove sequence patterns or signals that slow down the rate of transcription. Some 'nasty' sequences are:

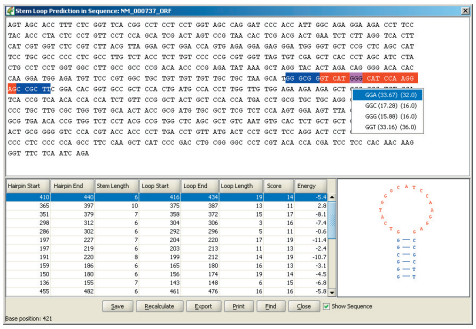
- Poly A sites
- Upstream/downstream Poly A sequences
- mRNA destabilization signals
- Hairpin structures
- Splice signals
- AT runs
- Potential ATGs
- TA and CG doublets
- TG and CT doublets

Researchers can create their own sequence patterns, choose different colors for highlighting patterns, and store them in an integrated relational database.



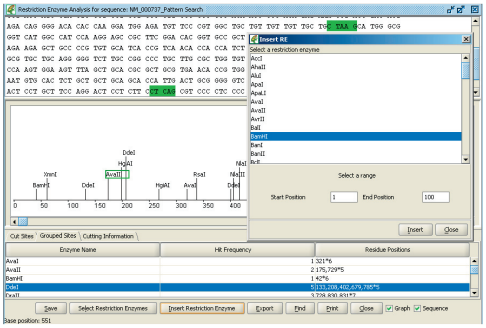
Hairpin Structure Analysis

Inverted repeats in the sequences may create stem-loop structures that may lower the translation efficiency of m-RNA. This module reads through a DNA sequence to find and highlight potential hairpin structures for a given setting of minimum stem size, minimum loop size, maximum loop size and maximum allowed mismatches. These identified stem loops can be removed by choosing alternate codons.



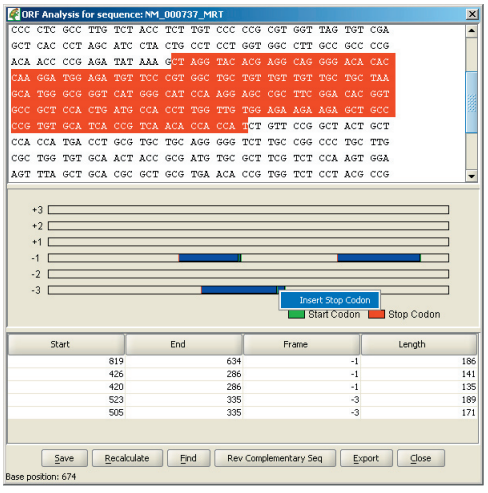
Restriction Site Insertion and Elimination

OptGene™ has an integrated restriction enzyme analysis tool that identifies existing restriction sites, facilitates insertion and elimination of restriction sites, without compromising on the amino acid sequence. Information such as recognition sequence for a particular enzyme, length of recognition sequence and position of the restriction site are also displayed in both graphical and tabular formats. This module also displays a summarized and categorized view (by enzyme) of restriction sites.



ORF Analysis and Reduction of Transcription Regions

This module gives a graphical view of all open reading frames of length specified by the user. The tool also allows insertion of stop codons to reduce transcription regions in unused frames without altering the native amino acid sequence.

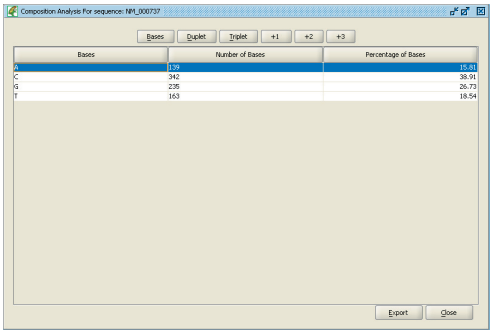


Six Frame Analysis

This module allows researcher to visualize and analyze the sequence in all of its six frames (-3 to +3).

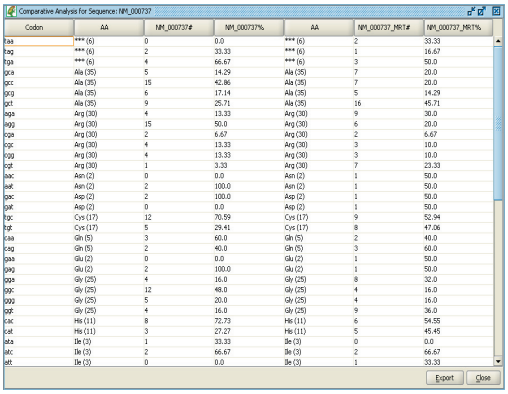
Composition Analysis

Composition Analysis module of OptGene™ analyzes the DNA sequences as bases, duplets, triplets in all the frames.



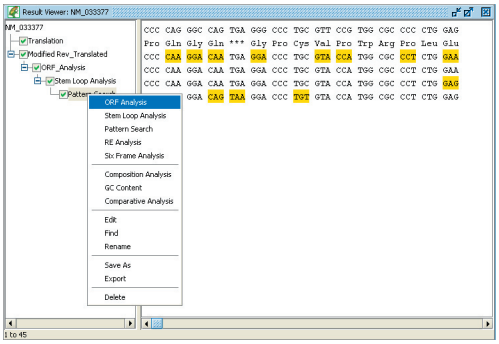
Comparative Analysis

Two or more sequences are compared for the occurrence of alternate codons of every amino acid. Calculation of percentage composition of the alternate codons of an amino acid is an additional feature of the tool



Result Viewer

Result viewer is a one-stop window where all appropriate sequence optimizations can be carried out and viewed. Each of the optimization steps is depicted as a node of the optimization tree. The tool also provides a 'Search' feature for locating codons and amino acids.



Technology Used To Make All This Possible

OptGene™ has been developed using advanced Java Technology and is platform independent.

Ocimum Biosolutions
...enabling R&D™
www.ocimumbio.com

This allows it to be used on varied operating systems like Windows, Linux, Unix or Macintosh. OptGene™ comes with an integral database that supports any of the several database systems such as Oracle, MySQL etc. However, the front end for a user is a friendly Graphical User Interface that does not require the user to be familiar with a database Platform Language (PL) or Structured Query Language (SQL).

The Last Word

In light of the recent elucidation of various genomes, the need to optimize potentially therapeutic genes is now more vital than ever. A gene optimization software's primary objective is to perform optimization and further analysis on the optimized sequence as restriction enzyme analysis, six-frame and ORF analysis.

OptGene™ is multi-platform bioinformatics software that enables the design of genes with features optimized for expression in any organism of choice. The software exploits redundancy in codon usage to engineer desired features into the targeted gene sequence. OptGene™ comes with a set of integrated, user-friendly and easy to use tools, which manipulate gene sequences to contain all the desired structural features.

OptGene™ is a novel tool that optimizes naturally occurring genes to achieve higher productivity, at the same time giving higher flexibility in protein designing. The software optimizes genes using only sequence information and choice of expression system. OptGene™ has been conceived, designed and developed by highly qualified dedicated biological researchers and technical developers to allow researchers to precisely adapt genes and gene products to suit specific requirements.

United States: Ocimum Biosolutions Inc. Fortune Park VI, 8765, Guion Road, Suite G, Indianapolis, IN, 46268, USA
Phone: +1 317 228 0600, Fax: +1 317 228 0700, Email: us@ocimumbio.com

India: Ocimum Biosolutions Ltd, 6th Floor, Reliance Classic, Road No. 1, Banjara Hills, Hyderabad 500 034, A.P., India
Phone: +91 40 666 27200, Fax: +91 40 666 27205, Email: india@ocimumbio.com

The Netherlands: Isogen Biosolutions, (An Ocimum Biosolutions Company), Lagedijk Noord 18, Postbus 220, NL -3400 AE IJsselstein, The Netherlands
Phone: +31 (0) 30 68 78 788, Fax: +31 (0) 30 68 88 009, Email: europe@ocimumbio.com

Ocimum Biosolutions ...enabling R&D™

www.ocimumbio.com