GRAMMR® Gene and Protein Engineering Technology

Overview:

Directed evolution is an effective tool for improving genes with little or no *a priori* knowledge of the target's structure or function, especially when efficient gene shuffling technology is coupled with an effective high throughput screen. Novici's gene shuffling technology, called Genetic ReAssortment by MisMatch Resolution (GRAMMR®), is a synthetic biology power tool that overcomes the limitations of conventional gene variant library creation approaches by producing large, high quality shuffled gene libraries with exceptional crossover frequencies, high crossover resolution, and low levels of unwanted random mutagenesis. This patented technology* streamlines directed evolution workflows with rapid iteration hits from initial synthetic DNA libraries to re-shuffle hits at much lower per-gene costs and higher diversity payloads than can be achieved by gene synthesis alone.

*Ready-to-screen genes from GRAMMR libraries cost **as little as $0.01 per gene**, whatever the size of the gene.*

The GRAMMR technology has been successfully used to create proprietary proteins with improved functional properties and to increase recombinant protein yields through empirical codon optimization of genes in multiple expression hosts.

GRAMMR Technology – How It's Different:

The GRAMMR Technology is an *in vitro* enzymatic process that operates by resolving sequence mismatches in heteroduplex DNAs through the coordinated action of several enzyme activities. In contrast to other combinatorial library technologies, this process focuses on the differences between genes rather than on their similarities. Those differences, called mismatches, are resolved in high-sequence-complexity heteroduplex molecules by endonuclease nicking at individual mismatches, followed by excision and re-synthesis at those sites to transfer sequence information between strands. Independently resolved mismatch sites can be located very close to one-another, with as few as one or two intervening base paired nucleotides, and the high crossover frequency of the process ensures a thorough blending of diversity that is introduced in the library design process.

In the simplest sense, GRAMMR is a technology for breeding genes together at the molecular level, in a highly controlled fashion, to generate large and complex populations of chimeric genes. Because this technology operates *in vitro* at the level of DNA, it can be used with equal effectiveness for codon optimization and protein engineering.

Increasing Recombinant Protein Yields Through GRAMMR Codon Optimization

It is well understood that gene expression in many organisms can be improved by adapting the codon usage of the gene to match the presumed preferences of the specific host. Traditional

codon optimization strategies rely on statistical predictions of those preferences by computer algorithms entrained with data such as natural genomic codon usage frequencies.

In light of the many known factors that influence gene expression levels, such as transcription rate, transcript stability, localized translational speed, occurrences of spurious transcription factor binding sites, and demands on host resources such as tRNA pools, adequate gene design requires more insight than predictive models are capable of providing. Moreover, since the number of combinatorial possibilities for encoding an average-sized gene includes well over 10^{100} possible sequence combinations, predicting the sequences that can best accommodate these factors is an impossible task. GRAMMR helps to overcome this limitation.

The Empirical Codon Optimization process begins by generating a population of codon-variant genes in which all possible alternative codons are densely interspersed throughout the original gene sequence. This starting population of codon variant genes contains tens of thousands to even hundreds of thousands, or even millions, of shuffled genes that are ready for expression and screening. After these variant clones are assessed for their expression levels and health impacts on the host, iterative rounds of GRAMMR shuffling and screening are performed, revealing patterns of codon usage that are custom-evolved for high-performance expression and genetic stability (Fig. 1).

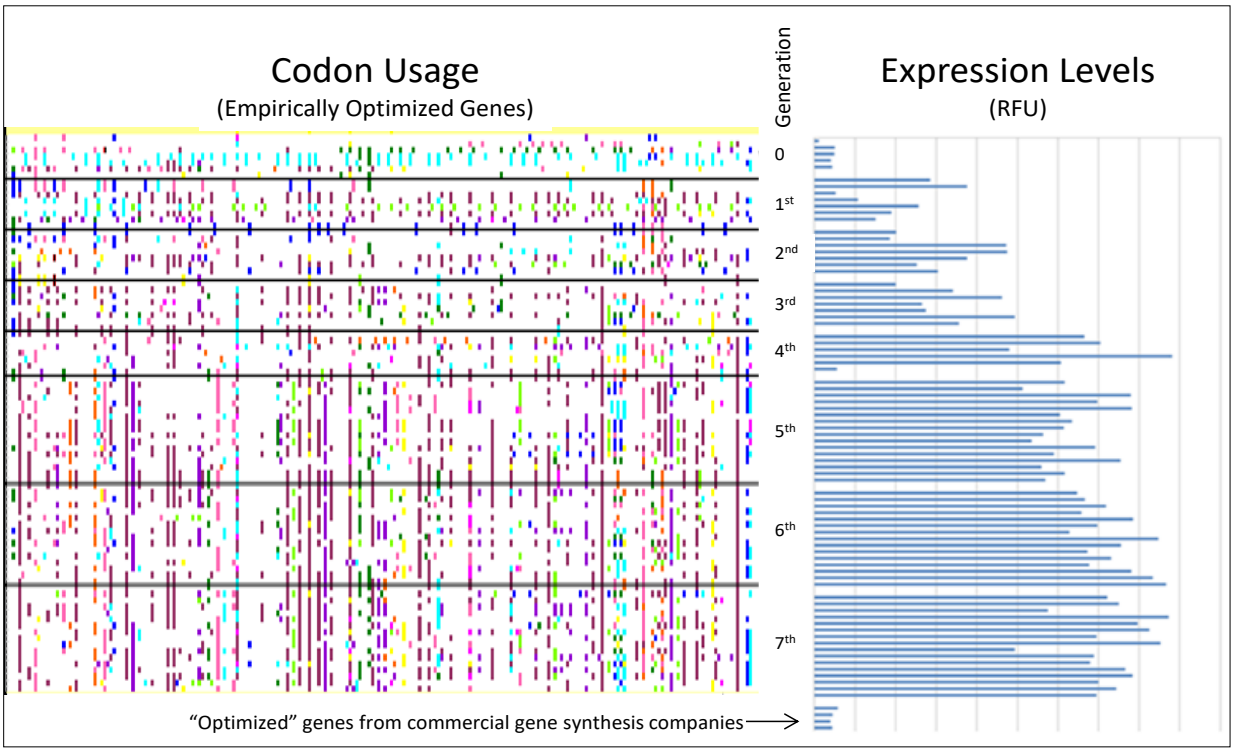

Fig 1. Empirical Codon Engineering of the GFP gene (238 codons) for maximal expression in *E. coli*. Alternative codons are shown as colored squares on white background denoting the starting gene sequence. Sample sequences of high-expressing clones from each generation are grouped with the initial library at the top and later generations below. The histogram at right indicates background-subtracted fluorescence levels measured in liquid culture. Similar patterns of codon usage emerge in many of the clones as maximal expression levels are reached. Note the positions at which particular codons appear to be favored over the original codon. Bottom rows of the

histogram show fluorescence levels of four commercially designed and synthesized genes that were computationally optimized by four gene synthesis companies for expression in *E. coli*.

Codon usage patterns that evolve in our process don't necessarily conform to canonical rules for algorithm-based gene optimization (Fig 2), and can yet be strongly preferred in the GRAMMR optimized genes, highlighting an important advantage of Novici's patented GRAMMR-driven process. Moreover, we now know that specific patterns of codon usage evolve with pronounced differences from one another when expressed and evolved in different contexts, such as with or without a leader peptide sequence, or when expression is carried out in different host strains or under different environmental conditions, suggesting a layer of design subtlety not immediately accessible with prediction programs.

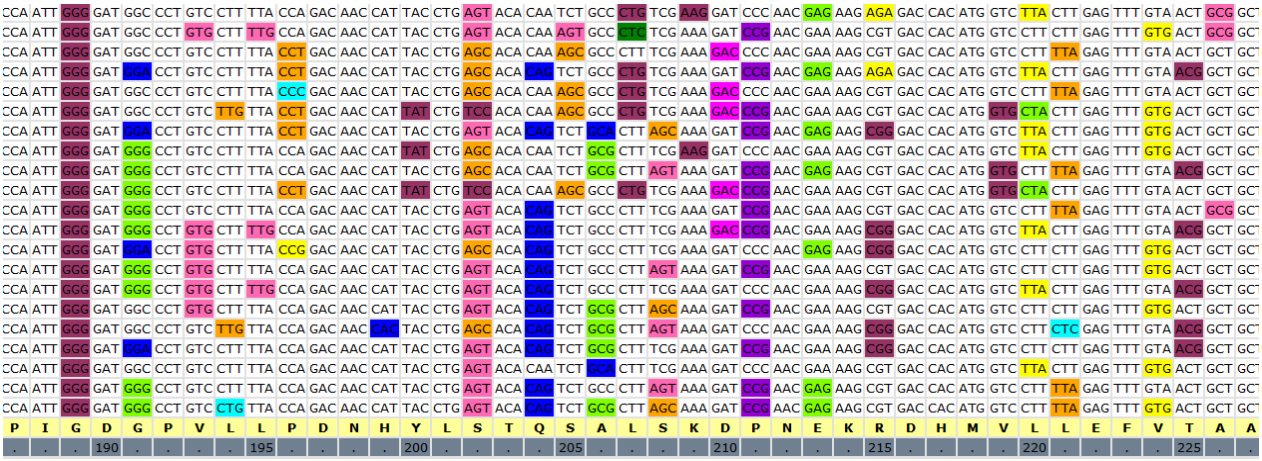

Fig 2. Detailed view of a 40 codon section of a group of empirically codon optimized genes from *E. coli* GFP example shown in Fig 1. Each row represents the linear sequence of an individual codon-evolved gene within the population of highest-expressing assay hits. Different colored cells represent alternative codons that are color coded according to the original 'codon parent' gene that we used to introduce codon diversity into the library. In all, the 10 'codon parent' genes were designed and synthesized to contain all possible codon substitutions at each position relative to the entire wild-type 'reference parent' gene. The high frequency and fine resolution of codon reassortment is evident in the diversity of differently-colored boxes displayed in the evolved genes.

The vast codon diversity content within initial GRAMMR libraries, combined with the power to iteratively reassort codons into new combinations, ensures a thorough search of sequence space to identify genes with superior performance. The empirical codon optimization process can both enhance expression yield and produce greater stability of microbial expression over time (Fig 3), possibly through the removal of codons or other more complex sequence elements that may conflict with housekeeping or other host functions.

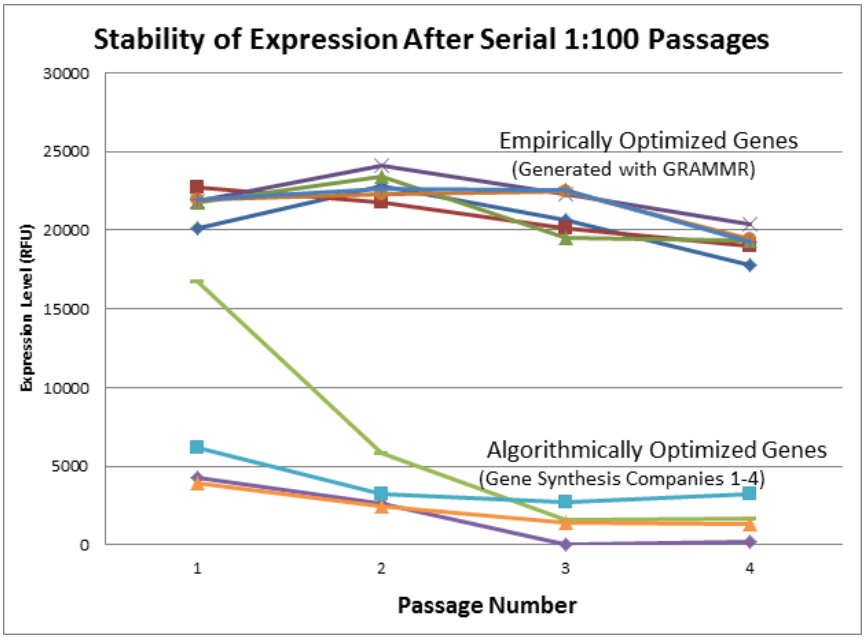

Fig 3. Genetic stability of optimized GFP in *E. coli*. Cultures with codon optimized GFP genes were serially cultured through four passages with GFP activity measured after each passage.

It is widely known that suboptimal patterns of codon usage can negatively impact the health of host cells in recombinant systems. Eliminating such usage patterns from the library is a natural consequence of Novici's empirical codon optimization process (Fig 4). Evolving codon usage patterns toward optimality without rational bias eliminates the need to pre-define which codon usage factors to address. This is especially important in expression hosts where little objective and empirical data exists for optimizing recombinant expression.

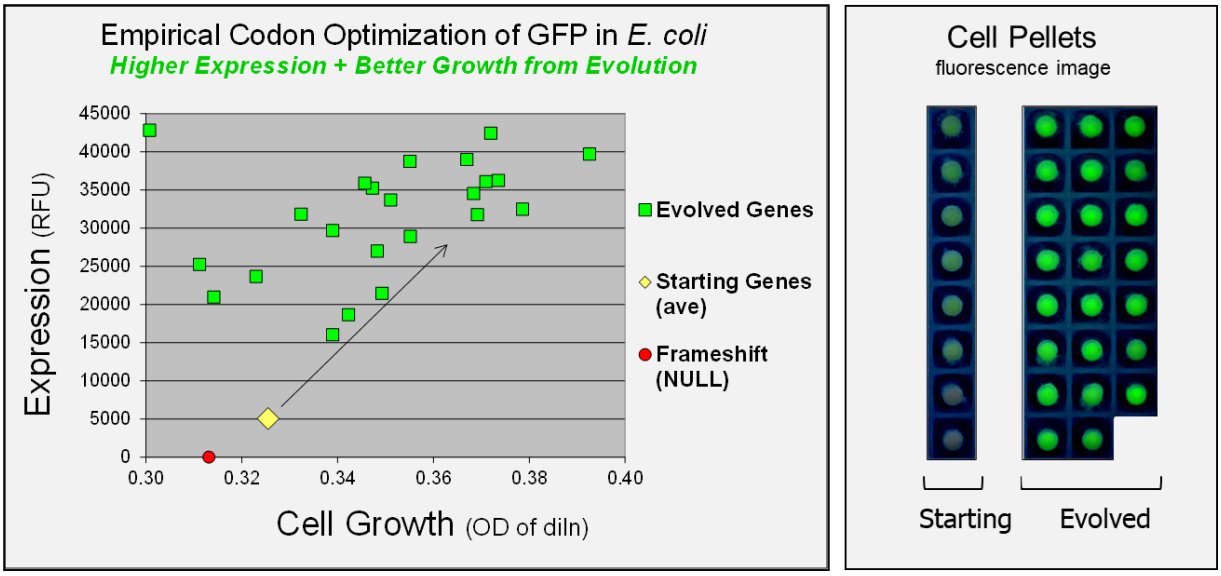

Fig 4. Simultaneous optimization of expression and culture density in *E. coli*. Codon-optimization can simultaneously optimize expression levels and cell densities.

Additional variation can be incorporated simultaneously into promoter regions or other vector elements to further broaden the search. Aside from enhancing yield, additional benefits from this process can include optimized translational kinetics by revealing appropriate translational pause sites to facilitate proper folding. This can be particularly important for secreted proteins, such as antibodies, that are cotranslationally folded and assembled as they transit the eukaryotic endomembrane system.

An example of empirical codon optimization is shown in Fig 5, in which a high-density αGalactosidase-A codon substitution library was built using GRAMMR and the shuffled variants were expressed in *Nicotiana benthamiana* using a viral vector and screened for yield. After only two iterative rounds of screening and library shuffling, multiple variants with greater than two-fold higher expression were identified, one of which is shown beside an αGalactosidase-A protein sample produced from the wild-type gene. The best-expressing genes at this stage still contained high degrees of codon diversity relative to one another, suggesting the evolutionary potential for additional gains with further rounds of shuffling and screening.

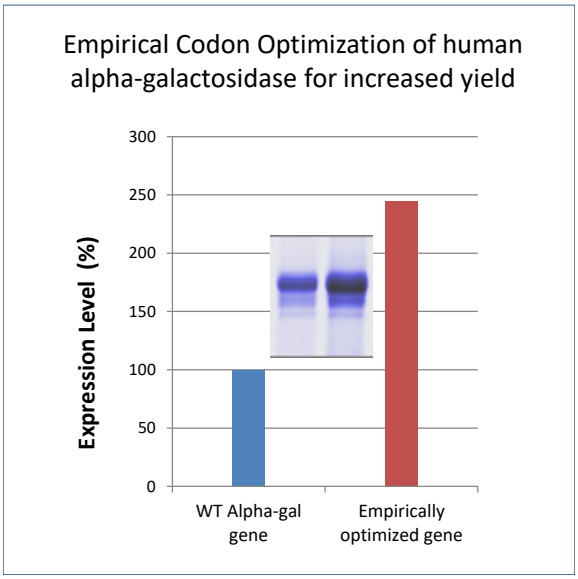

Fig 5. Coomassie-stained SDS-PAGE of *N. benthamiana* expressing either the wild-type gene or a codon-optimized version of the human α-galactosidase A gene. Although the genes differ in their codon usage patterns, the α-galactosidase protein sequences they encode are identical.

Protein and Vector Engineering Applications of GRAMMR Technology

Similar approaches using large libraries of sequence variants can be applied to improving desired properties of a target protein such as affinity maturation or other pharmacologic properties (see USAMRIID/Novici paper; Koehler et al., Antiviral Res 92:461-469. 2011).

Since the GRAMMR Technology operates on intact, full-length sequences that are not subjected to fragmentation or homology-driven reassembly, even large genes or multiple gene targets can remain within their expression vectors to be efficiently chimerized with minimal damage from the shuffling process itself. Libraries can be constructed to manage varying degrees of diversity input, as desired.

Additionally, since GRAMMR doesn't require knowledge of hit sequences, library iteration is a snap, and doesn't require any additional gene synthesis steps.

At the lower end of the diversity scale, defined substitutions can be introduced at specific sites (Fig 6a). Further library strategies can be employed to interpolate sequence space between naturally occurring genes, either in pairs, or as multi-genic libraries to populate the sequence space between many genes. Protein alignments often contain contiguous strings of dissimilar sequences that GRAMMR can separate into blocks of varying size and redistribute along with discrete amino acid substitutions to obtain maximal dispersion of diversity into the library (Fig 6b).

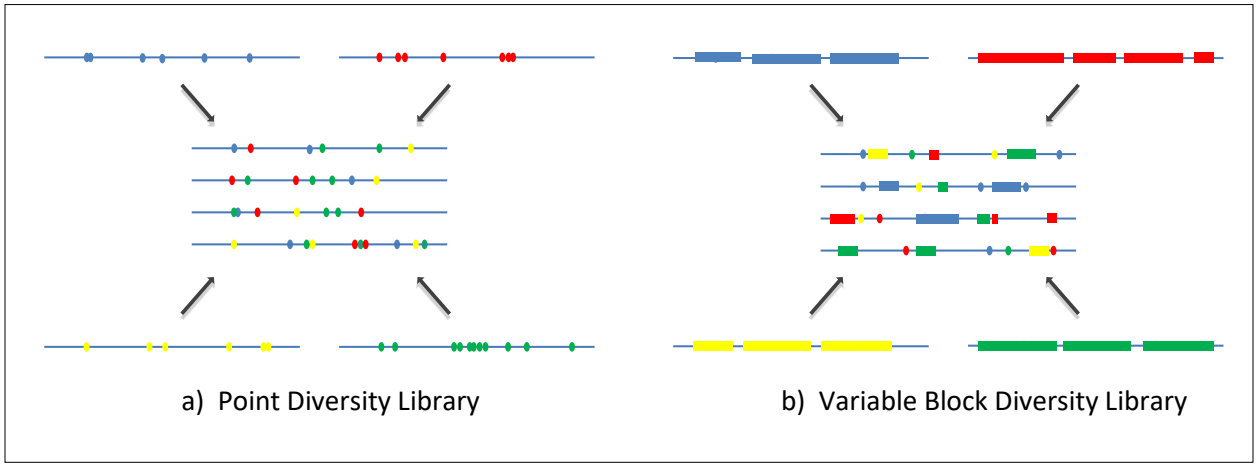

Fig 6. a) Low-complexity point diversity library comprised of targeted substitutions at defined sites. b) High-complexity variable block diversity library constructed to incorporate family diversity from related genes by breaking blocks of nonidentical sequences into individual amino acid substitutions and strings of various lengths.

Novici has demonstrated the effectiveness of the GRAMMR Technology in several applications. Fig 6 shows the evolution of an interferon-based antagonist of the B18R decoy receptor from *Vaccinia* virus. A two to three order of magnitude improvement in selectivity for binding to the poxvirus decoy is apparent over the four rounds of iterative shuffling and screening that were performed, effectively generating molecules that blocked the ability of the

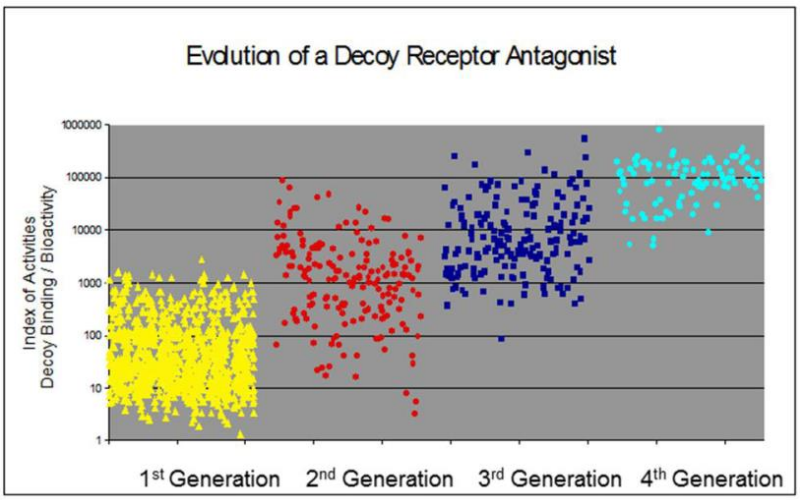

Fig 6. Stepwise optimization in serial rounds of the GRAMMR® shuffling. Following shuffling, expression, purification, and quantification, proteins were assayed for cellular activity and in competitive indirect ELISA. Shufflants with desired activity ratios in these assays were selected for iterative rounds of shuffling and screening.

B18R protein to disrupt extracellular interferon signaling (Unpublished, and US Patent 7,888,475). Other examples of successful use of GRAMMR include increasing enzyme thermal stability and engineering of more potent human cytokines, as well as optimization of a number of insecticidal toxin targets and herbicide tolerance traits for commercial agricultural trait developers.

About Novici:

Novici Biotech LLC is a private biotechnology company located in the San Francisco Bay Area. Founded in 2007, the Novici team offers the GRAMMR technology to partners and customers to help them quickly maximize gain in molecular performance and yield through directed evolution.

For more information, please contact:

Novici Biotech LLC
info@novicibiotech.com

* US Patent: 7,582,423 Population of polynucleotide sequence variants
US Patent: 7,217,514 Method of increasing complementarity in a heteroduplex