Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Critical reflections on synthetic gene design for recombinant protein expression

Annabel HA Parret[1], Hüseyin Besir[2] and Rob Meijers [1]

Gene synthesis enables the exploitation of the degeneracy of the genetic code to boost expression of recombinant protein targets for structural studies. This has created new opportunities to obtain structural information on proteins that are normally present in low abundance. Unfortunately, synthetic gene expression occasionally leads to insoluble or misfolded proteins. This could be remedied by recent insights in the effect of codon usage on translation initiation and elongation. In this review, we discuss the interplay between optimal gene and vector design to enhance expression in a particular host and highlight the benefits and potential pitfalls associated with protein expression from synthetic genes.

Addresses
[1] European Molecular Biology Laboratory, Hamburg Unit, Notkestrasse 85, 22607 Hamburg, Germany
[2] Protein Expression and Purification Core Facility, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Corresponding author: Meijers, Rob (r.meijers@embl-hamburg.de)

Current Opinion in Structural Biology 2016, **38**:155–162

This review comes from a themed issue on **New constructs and expression of proteins**

Edited by **Rob Meijers** and **Anastassis Perrakis**

http://dx.doi.org/10.1016/j.sbi.2016.07.004

0959-440/© 2016 Published by Elsevier Ltd.

Introduction

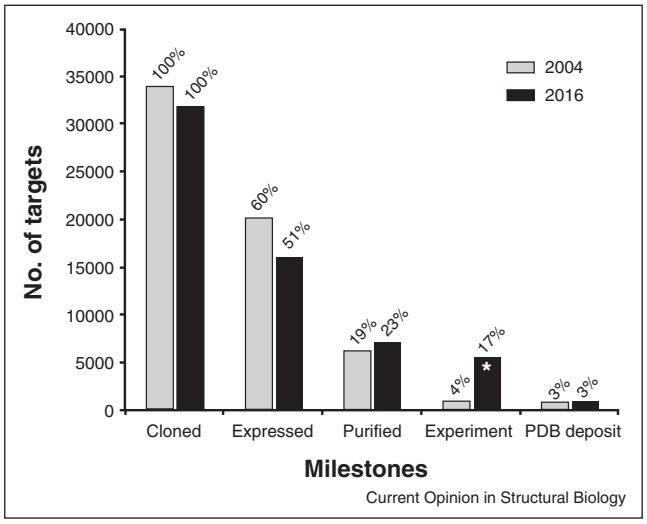
The efficient overexpression of proteins is a prerequisite for successful structure determination. Structural genomics studies on a wide range of soluble protein targets over more than a decade have shown that more than 40% of the constructs failed due to limited expression (Figure 1). The high failure rate is a result of the use of standardized approaches, and can serve as a baseline for a structure determination project on a specific target. To alleviate these ominous statistics, a number of approaches have been adopted which address recombinant protein expression. The use of parallel cloning strategies employing different construct lengths and purification tags can increase success rate to a certain extent [1]. Alternatively, larger amounts of protein can be obtained by scaling up expression using large-scale fermentation [2]. However, large-scale fermentation often requires optimization to address problems of plasmid loss, oxygen deprivation and pH fluctuation [3].

A more cost-effective approach is the optimization of the protein source, either by opting for a protein species of high natural abundance, or by using an efficient recombinant expression system. Traditionally, selecting protein homologues from species that can be purified from natural source has been rather successful [4–8]. Nevertheless, there is an increasing need to obtain the structure of a protein from a particular species, in order to assess species-specific functional aspects. Structures of human or murine proteins involved in the immune or nervous system can help to relate the effect of specific point mutants or posttranslational modifications. Similarly, structural analysis of virulent proteins from bacterial pathogens can reveal species-specific details important for drug design. In this review, we focus on the opportunities that gene synthesis provides to optimize recombinant protein expression of specific protein targets.

Application of synthetic genes for protein production

The use of synthetic genes for the production of proteins for structural characterization is widespread. In general, gene synthesis may expedite the production of a series of constructs with engineered restriction sites. It is often preferred to the use of cDNA, which can contain several isoforms or splice variants. For target proteins identified from metagenomes or extremophiles, gene synthesis may provide the only route to protein expression, because cDNA is not available. Protein expression from codon-optimized genes has been especially successful for proteins involved in cell signalling whose native codon sequence can be strongly regulated, such as Irisin [9], Parkin [10] and Netrin [11]. Codon optimization has also uncovered hidden levels of gene regulation that are encoded in codon bias. Codon optimization of the FRQ protein involved in the regulation of circadian rhythms resulted in a loss of function, although expression levels were increased [12•]. Codon optimization also revealed a hidden quality control mechanism encoded in the gene of the cystic fibrosis transmembrane conductance regulator (CFTR), an anion channel membrane protein [13]. Similarly, optimization of the CTP1L endolysin gene led to the discovery of a secondary translation site that produced a truncated protein. The truncated protein forms a 1:1 complex with the full length protein, and lysis activity is substantially reduced when the secondary translation site is removed by codon optimization

Figure 1



A comparison of the total number of targets achieved per milestone along a structure genomics pipeline as reported by TargetDB in 2004 [104] and as of 15th of March 2016 (http://targetdb.rcsb.org/metrics/). The recent data are compiled from soluble protein targets from three consortia (JCSG, NESG and NYSGRC). The attrition rate from clone to purified target has changed little over a decade and hovers around 75%, indicating that the early stages of a structure determination project are most crucial for successful completion. *The definition of the milestone 'Experiment' has changed from 'crystals' to 'X-ray/NMR/EM studies performed for structural studies'. As a result, there is a large discrepancy between crystals obtained (4% in 2004) and experiments done on the sample (17%), irrespective of the outcome.

[14]. Toll-like receptors (TLRs) contain leucine-rich domains, and it was shown through targeted leucine codon optimization that transcription of TLRs is regulated by leucine codons to balance the abundance of certain receptors [15]. It is clear that gene synthesis may provide even unexpected benefits, but there are also numerous reports were it leads to reduced protein yield [16,17].

Adapting codon usage of target genes using codon optimization algorithms

Exploiting gene synthesis for transgene expression enables researchers to influence the multitude of variables controlling protein yield. The biased usage of synonymous codons has a strong effect on protein expression [18••]. Species-specific differences in this so-called codon bias can result in tRNA depletion, which slows down protein translation. This effect can be in part compensated by using an expression host that expresses rare tRNAs [16], however this is often insufficient [19].

An extensive collection of software tools (Table 1) are available to analyze codon usage and identify unwanted features that have been identified to hamper expression (Table 2). Gene design tools enable researchers to optimize the sequence of their gene of interest (GOI) [20]. A common strategy to manipulate codon bias to maximize protein translation is based on the 'codon adaptation index' (CAI) [21]. CAI is a species-specific index for codon frequency based on a set of highly expressing genes. Individual genes can then be analyzed, comparing the actual codons used to a fully optimized gene that contains only the most frequently used codons. It should be noted that CAI maximization does not necessarily correlate with high protein yield [22–25]. In general, the best results using CAI optimization are obtained when a subset of less frequent codons is replaced, after which a secondary list of criteria should be optimized involving DNA and RNA sequence elements that can negatively influence expression of the GOI (Table 2). It appears some features of codon bias are universal for all expression systems, whereas others are specific for prokaryotic or eukaryotic expression.

Codon optimization algorithms will generally allow adjustment of several of these undesirable sequence properties simultaneously [20,26–28], although the weight given to each sequence parameter varies between different algorithms. It is important to note that proprietary CAI-based codon optimization algorithms adjust a subset of rare codons in a random manner. As a result, gene optimization is ambiguous, since the replacement of rare codons is not based on empirical data. Species-specific gene optimization routinely involves the use of the species-relevant CAI index, taking into account the codon frequency of the particular species. Other factors that affect for example eukaryotic protein expression are most often not considered, because the influence on expression levels is poorly understood.

Recent insights into the relationship between mRNA, protein expression and protein folding

It is obvious that codon-mediated translational control is not yet well understood, and further insight will benefit gene design for protein expression. Recent studies indicate that rare codons may play different roles, such as safeguarding mRNA structure [29], depending on their positioning within the gene. A large scale study on bacterial genes expression in *Escherichia coli* under a common promotor system revealed that mRNA stability correlates with expression levels [30•]. However, there is a difference in the effect between the 'head' of the gene (covering approximately 16 codons) and the rest of the gene. In the 'head', the mRNA structure is more important, whereas the rest of the gene is rather affected by codon-mediated translation elongation efficiency. This phenomenon is likely a universal property for both prokaryotic and eukaryotic protein expression [31]. Cell-free expression studies have shown that universal translation initation tags can be introduced upstream of the GOI that boost expression both in bacterial and eukaryotic systems [32,33]. To boost protein expression, it may be sufficient

Download English Version:

https://daneshyari.com/en/article/8319677

Download Persian Version:

https://daneshyari.com/article/8319677

Daneshyari.com