

## THE BASIC GENE DESIGN PROCESS FOR EXPRESSION OF HETEROLOGOUS PROTEINS IN TRANSGENIC PLANTS

LALIT LATA JHA<sup>1\*</sup>, RAJESH KS<sup>1</sup>

<sup>1</sup>Parul Institute of Pharmacy, Limda, Vadodara, India. Email: lalit\_lata@hotmail.com

Received: 29 Mar 2012, Revised and Accepted: 12 Jan 2013

### ABSTRACT

Transgenic plants are one of the promising expression systems for the production of heterologous proteins, especially therapeutic proteins and peptides. The other expression systems are bacteria, mammalian cell culture, transgenic animals, yeast and fungi. Transgenic plants as bioreactors for the production of useful proteins based pharmaceuticals is not very popular, primarily because of low expression of transgene in plants. With recent development of some gene design strategies using different softwares for best codon usage, study of codon pairs, extreme GC content, eliminating repetitive sequences, avoiding unfavourable mRNA secondary structures and study of restriction sites, it is possible to design a best gene sequence expressing highly in the transgenic plants. The aim of the present study is to understand the factors affecting heterologous protein expression in transgenic plants and to develop a basic gene design process using different computer programs and softwares available in literature leading to high level expression of therapeutic proteins and peptides in plants.

### INTRODUCTION

Expression of recombinant therapeutic proteins and peptides outside its natural host is often a challenge for the plant biotechnologist. The techniques often are simple for the soluble proteins but major obstacles are encountered in the expression of heterologous proteins<sup>1</sup>. Heterologous proteins are the therapeutic proteins and peptides, biochemical marker proteins to study metabolic pathway, protein for disease diagnosis and various industrial proteins which are expressed artificially outside its natural host system<sup>2</sup>. It is often difficult to express the heterologous protein outside its host because of low codons usage and post translational modifications of the proteins<sup>3</sup>. Therefore studies have been started to redesign the gene of the target protein for maximising the high expression of heterologous proteins in transgenic plants. A gene design process can be defined as different strategies used for optimum and sometimes enhanced expression of therapeutic proteins and peptides outside its natural host system. Different softwares are available to study best codon usage, extreme GC content, repetitive sequences, restriction sites, unfavourable mRNA structures by which one can modify the desired protein sequence for biologically active, stable and optimum expression in transgenic plant. Transgenic plants are the plants that are produced in the lab using rDNA technology in order to create plants with special characteristics by artificial insertion of genes from other species, and sometimes entirely from different kingdom[4,5]. The plants that are successfully used for making transgenics are rice, corn, tobacco, soyabean and tomato, etc.[6,7]. The special attributes that can be developed inside the transgenic plants are tolerance towards various abiotic and biotic stresses, plants with increased yields<sup>8</sup> and transgenic plants as bioreactor for production of therapeutic peptides and proteins<sup>9</sup>. There are specific advantage of using transgenic plants for production of therapeutic peptides and proteins particularly high yield of protein in natural active conformation, Low production cost, lower capital investment and easy scale up and safe to produce[10]. Once the gene will be designed, there are different methods available for the expression of the heterologous proteins in transgenic plants, for example, nuclear or plastid transformation, transient or stable expression, callus, whole plant or plant cell suspension culture as expression system[11]. Out of these nuclear transformation via *Agrobacterium* is one of the important tool which is a simple, precise transfer and integration of gene of interest leading to stable transformation with single copy insertion[12,13,14]. Apart from *Agrobacterium*, gene gun or biolistic method and is also used to transfer the gene into the nucleus and leads to stable transformants[15,16].

### METHODS

#### By using Codon bias table

Codon bias table are used to alter the rare codon in the target gene so that they reflect the codon usage of the host more closely[17,18]. There are different websites which deals with codon bias analysis

and one of them is <http://gcu.schoedl.de>[21]. This particularly describes the graphical codon usage analyser (GCUA). The GCUA displays the codon quality either in codon usage frequency values or relative adaptiveness value. Here there is comparison of two usage tables. The fraction of usage of each codon in the submitted usage tables will be compared graphically. Also there are options for each triplet position v/s usage table and each codon v/s usage table. Another very useful website for codon bias table is <http://www.kazusa.or.jp/codon/>[22] which deals with codon usage database. It contains 35,799 organisms and 3,027,973 complete protein coding genes.

#### Choice of expression vector

Expression of recombinant therapeutic peptide and protein outside its natural host is often a difficult challenge especially for heterologous protein which is expressed outside its natural host. For expression of heterologous proteins, a good expression vector is required as a carrier to integrate the foreign DNA into the plant genome[19]. These plasmid vectors are often designed in such a way that it has compatible genetic background. The genetic element of the expression plasmid include ORI, an antibiotic resistance marker, transcriptional promoter, translation initiation regions (TIRs) as well as transcriptional and translational terminators. Some of the useful vectors which are available for use and shown to be promising results are pGreen[29], pBECKS<sub>2000</sub>, pCAMBIA, pBIN19, pC22, pGA482, pGPVT, pJ, pMON, etc.

#### By using softwares available in different websites

Apart from codon bias table and plant expression vectors, designing of the gene can also be done by various softwares which are available worldwide and study the extreme GC content, eliminating repetitive sequences, avoiding unfavourable mRNA secondary structures study of restriction sites, primer designing etc. By studying all these parameters, it is possible to design a best gene sequence expressing highly in the transgenic plants. <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>[23] is the website which is very useful for ORF finder. This is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a users sequence or in a sequence already in the database. This tool identifies all ORF using the standard or alternative genetic codes. The deduced amino acid can be saved in various formats and searched against the sequence database using [www.blast server](http://www.blast.ncbi.nlm.nih.gov/)[24]. It is helpful in preparing complete and accurate sequence submission. <http://www.bioinformatics.org/seqext> is giving a sequence extractor software which generates restriction map, PCR primer map of a DNA sequence, Protein translations and introns exon boundaries[25]. <http://cellbiol.com/tools.html> are important websites to find a sequence, protein structure, and featured prediction, search homologies, alignments, working with expressed

sequence tags, genomes and mapping[26]. It also helps to pick and analyse PCR oligos and other tools like restriction enzymes and mapping, ORFs and protein translation. [www.ebi.ac.uk](http://www.ebi.ac.uk) deals with many tools for analysis and comparison of nucleotide and protein sequences, data from functional genomics experiments, tools for determination and visualisation of macromolecular structures[27]. [www.fermentas.com](http://www.fermentas.com) gives various tools like Research™, Restrict enzyme, to perform double digestion and DNA sequence analysis, plasmid map creation and biochemical calculations[28]. [www.lablife.org](http://www.lablife.org) and [www.pgreen.ac.uk](http://www.pgreen.ac.uk)[29] are very helpful for vector database. <http://web.expasy.org/translate> is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence[30]. [www.neb.com](http://www.neb.com) deals with various molecular biology tools for example restriction enzyme finder, NEB cutter, double digest finder, T<sub>m</sub> calculator, DNA sequence and maps, PCR selection temperature, cleavage close to the end of the DNA fragments and many more[31]. There are other websites also which deals with similar kind of exercises[32-34]. By looking over all these and analysing our gene sequence by using all these softwares best gene sequence can be constructed which will show high expression of heterologous protein in transgenic plant.

### CONCLUSIONS

Expression of heterologous proteins, especially therapeutic peptides and proteins in transgenic plants are promising area. To overcome various difficulties during expression and also for the high level expression of the heterologous proteins, various gene design strategies can be used. Factors like codon optimisation and choice of expression vectors are very important tools apart from others like method of transformation, integration & stability of transgene, transcription and roles of introns and post translational modification of the protein.

### REFERENCES

- Richard MT, Eva S, Stefan S, Paul C, Rainer F: Molecular farming in plants: host systems and expression technology. *Trends in Biotechnology* 2003; 21: 570-578.
- Sorensen HP, Eva S, Mortenson KK: Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology* 2005; 115: 113-118.
- Claes G, Sridhar G, Jeremy M: Codon bias and heterologous protein expression. *Trends in Biotechnology* 2004; 22: 346-353.
- Eva F, Ute T and Rtidiger H: Recombinant proteins from transgenic plants. *Current Opinion in Biotechnology* 1997; 8: 411-416.
- Tian XC and Yang XJ: Life on the bio-pharm: therapeutic proteins from transgenic organisms. *Molecular Medicine Today* 1998; 98: 424-425.
- Karin H, Uwe S: Production of new/modified proteins in transgenic plants. *Current Opinion in Biotechnology* 1999; 10: 163-168.
- Rainer F, Eva S, Stefan S, Paul C: Plant-based production of biopharmaceuticals. *Current Opinion in Plant Biology* 2004; 7:152-158.
- Grover A, Gowthaman R: Strategies for development of fungus-resistant transgenic plants. *Current Science* 2003; 84: 330-340.
- Carla M, Francesca DA, Barbara S, Margherita Z, Stefano M: Production of a recombinant antimicrobial peptide in transgenic plants using a modified VMA intein expression system. *FEBS Letters* 2002; 519: 141-146.
- Francesco S, Rigano MM, Barbante A, Barbara B, Amanda M W, Stefano C: Vaccine antigen production in transgenic plants: strategies, gene constructs and perspectives. *Vaccine* 2003; 21: 803-808.
- Loic SG, Marzena JF, Maryse B, Didier C, Vincent P, Hilary K: Expression of a human anti-rabies virus monoclonal antibody in tobacco cell culture. *Biochemical and Biophysical Research Communications* 2006; 345: 602-607.
- Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R et al: Hydroxylated human homotrimeric collagen I in *Agrobacterium tumefaciens*-mediated transient expression and in transgenic tobacco plant. *FEBS Letters* 2002; 515:114-118.
- Roy M, Pathak A, Riyad YH: An effective method of sonication-assisted *Agrobacterium*-mediated transformation of chickpeas. *Plant Cell Tissue Organ Cult* 2008; 93: 65-71.
- Scott B, Fabricio MB, Qiang C, Kevin VC, Chenming Z: Expression of Porcine Prolaxin in Transgenic Tobacco. *Ann. N.Y. Acaemic Scienc.* 2005; 1041: 77-81.
- Zhong H, Kate L, Galina E, Yasmin T, Hugh SM: High-yield rapid production of hepatitis B surface antigen in plant leaf by a viral expression system. *Plant Biotechnology Journal* 2008; 6: 202-209.
- Kuan-Hung L, Hongyong F, Cheng-Han C, Hsiao-Feng L, Ming-Chih S, You-Ming C, Long-Fang C: Generation and analyses of the transgenic potatoes expressing heterologous thermostable  $\alpha$ -amylase. *Plant Science* 2008; 174: 649-657.
- Kurland CG: Codon bias and gene expression. *FEBS 09993* 1991; 285:165-169.
- James FK: Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology* 1995; 6: 494-500.
- Desai PN, Srivastava N, Padh H: Production of heterologous proteins in plants: Strategies for optimal expression. *Biotechnology Advances* 2010; 28: 427-435.
- Markus H. Almut E, Markus F and Christoph G: Influence of Codon Bias on the expression of foreign genes in microalgae. In: Rosa Leon, Emilio F, Aurora G, eds. *Transgenic microalgae as green cell factories* 1995; 46-53.
- Graphical codon usage analyser <http://gcua.schoedl.de>.
- Codon databases <http://www.kazusa.or.jp/codon/>.
- ORF finder <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>.
- Sequence homology <http://www.blast>.
- Sequence extractor <http://www.bioinformatics.org/seqext>.
- bioinformatics tools <http://cellbiol.com/tools.html>.
- Sequence analysis <http://www.ebi.ac.uk>.
- Gene analysis softwares [www.fermentas.com](http://www.fermentas.com).
- Plant Vector details [www.pgreen.ac.uk](http://www.pgreen.ac.uk).
- Protein translation information <http://web.expasy.org/translate>.
- DNA tools <http://www.neb.com>.
- DNA Design software <http://www.dna2.com>.
- DNA Design software <http://www.lablife.org>
- DNA, protein and all related informations <http://www.ncbi.nlm.nih.gov>.