

EXPRESSION VECTOR FOR PRODUCTION OF RECOMBINANT PROTEINS IN PROKARYOTIC HOST CELLS

FIELD OF THE INVENTION

- 5 The present invention relates to production of recombinant proteins in a host cell, and more particularly to an expression vector for production of recombinant proteins or their analogues in prokaryotic cells.

DESCRIPTION OF THE RELATED ART

- 10 Recombinant DNA (rDNA) technology has been used to clone, express and purify several proteins of therapeutic or other economic value from prokaryotic cells e.g., bacterial cells. The major advantages of producing recombinant proteins in bacterial cells are shorter time to express proteins coupled with lower costs for production of them. The proteins may be produced in bacterial cells either intracellularly as soluble proteins or inclusion bodies, or
15 extracellularly by secretion into periplasm or nutrient media. Despite the wide applications in production of different types of recombinant proteins, the bacterial production of heterologous proteins continues to face major challenges pertaining to low yields or expression of the recombinant protein like Insulin, Granulocyte Colony Stimulating Factor (GCSF) etc.

- 20 There have been attempt in designing expression constructs or plasmid vectors that increase the expression of recombinant gene introduced in them. From the various strategies of increasing expression of recombinant gene in a host cell by way of increasing production of inclusion bodies include incorporating active promoters, optimising codons, including leader sequences or a combination of these and other strategies known in the art.

- 25 The inclusion of leader peptides in an expression construct finds favour since it directly leads to increase in production of inclusion bodies and may be attached to a purification or expression sequence tag for simplifying purification of recombinant protein during downstream processing. Further, the leader peptide may be cleaved using enzymatic methods.

30

Currently available leader peptides come with host of difficulties. One of them being overall incompatibility with large number of recombinant proteins and being very specific to a particular protein. There are fewer universal leader peptides and expression constructs based on them.

5

Accordingly, there is a need to develop expression constructs that are substantially universal in application with respect to expression of recombinant proteins in prokaryotic host cells and provide uniformly high expression for range of recombinant proteins of therapeutic and non-therapeutic value.

10

SUMMARY OF THE INVENTION

In view of the foregoing, the embodiments herein, provide an expression vector having a leader peptide sequence that results in higher production of inclusion bodies.

15 In an aspect, an expression vector for production of a recombinant protein in a host cell is provided. The expression vector includes a nucleotide sequence of Sequence ID No 2 encoding for a leader peptide of sequence ID No 3.

The expression vector expresses said recombinant protein as a fusion protein comprising fusion of said leader peptide of SEQ ID NO 3 and said recombinant protein and the host cell is
20 bacteria, preferably *E. coli*. The leader peptide has Methionine at N-terminus, followed by Glycine to impart stability to fusion of said recombinant protein and said leader peptide.

The expression vector further includes DNA sequence encoding for a cleavage site or Restriction Enzyme (RE) site ligated to DNA sequence of said leader peptide. The expression
25 vector further includes a DNA sequence encoding a multiple cloning site (MCS) in upstream region of said leader peptide, a DNA sequence of said heterologous protein is cloned in said MCS; a DNA sequence encoding ribosome binding site (RBS) ligated to N-terminus of said leader peptide, a DNA sequence encoding a promoter or operator in the downstream of said ribosome binding site and DNA sequence encoding an antibiotic selection marker in upstream
30 region of said promoter/operator sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the embodiments herein, reference should now be made to the embodiments illustrated in greater detail in the accompanying drawings and described below by way of examples:

5

Figure 1 illustrates an expression construct according to an embodiment herein;

Figure 2a and 2b illustrates comparison of SDS PAGE analysis and densitometry data of GCSF as expressed in the expression vector of Figure 1 and as expressed in a control vector; and

10

Figure 3a and 3b illustrates comparison of GCSF expression, as measured by MALDI-TOF, in a control vector and in vector of Figure 1.

DETAILED DESCRIPTION OF THE INVENTION

15 As required, detailed embodiments are disclosed herein; however, it is to be understood that the disclosed embodiments are merely exemplary, which can be embodied in various forms. Therefore, specific structural and functional details disclosed herein are not to be interpreted as limiting, but merely as a basis for the claims and as a representative basis for teaching one skilled in the art to variously employ the present invention in virtually any appropriately
20 detailed structure. Further, the terms and phrases used herein are not intended to be limiting but rather to provide an understandable description of the invention.

The terms “a” or “an”, as used herein, are defined as one or more than one. The term “plurality”, as used herein, is defined as two or more than two. The term “another”, as used
25 herein, is defined as at least a second or more. The terms “including” and/or “having”, as used herein, are defined as comprising (i.e., open language).

Vector Deposition

30 The vector pBGBactX is deposited for the patent purposes under Budapest Treaty at the MTCC (Microbial Type of Culture Collection) Chandigarh, India. The deposit was made on March

21, 2013 and accorded deposit number as MTCC 5818. The sequence was characterised using DNA Sequencer.

As mentioned, there is a need for universal plasmid vectors which lead to high yield of heterologous proteins through simple purification processes. The embodiments herein provide a plasmid vector having nucleotide sequence listed under SEQ ID NO. 1.

The expression construct of Figure 1 includes a DNA sequence, of SEQ ID NO 2 encoding for the leader peptide of SEQ ID NO. 3. The leader peptide of SEQ ID NO. 2 includes DNA sequence encoding for Methionine in its N-terminal end. The DNA sequence for Methionine is followed up by addition of DNA sequence encoding for Glycine. The addition of Glycine provides stability to the protein-leader peptide fusion.

The leader peptide of SEQ ID NO 2 is a neutral peptide with nearly as many hydrophobic amino acids as hydrophilic amino acids. In one embodiment, the leader peptide has 49% amino acids as hydrophobic. The neutrality of the leader peptide enables formation of stable inclusion bodies when the expression construct of Figure 1 is expressed in the bacterial cells.

The DNA sequence for the protein of interest is inserted in the Multiple Cloning Site (MCS) of the expression vector as shown in Figure 1. Multiple cloning site or polylinker constitutes a short segment of DNA which contains a number of (generally up to 20) Restriction Enzyme (RE) sites - a standard feature of engineered plasmids.

In a preferred embodiment, the leader peptide and the MCS are custom synthesised as single stranded oligonucleotides, which are used for synthesis of double stranded DNA fragment by PCR. In one embodiment, the overlapping PCR method is used to synthesise double stranded DNA. Optionally, the leader peptide and the MCS may be directly synthesised as double stranded DNA fragments.

Further, the RE sites were incorporated at 5' end and the 3' end of the synthesised DNA fragment. Furthermore, a Promoter/Operator region, a Ribosome Binding Site (RBS), an origin

of replication and a antibiotic resistant gene were ligated with the PCR amplified DNA sequence coding for leader peptide, followed by MCS containing unique restriction enzyme sites. In one embodiment, the leader peptide is cloned downstream of the RBS, between *NcoI* and *EcoRI* restriction sites in the MCS.

5

The protein of interest may include filgrastim, interferon, human growth hormone, trypsin, carboxypeptidase, transferrin and various such recombinant proteins and peptides of therapeutic and non-therapeutic significance. A cleavage site may be included between the leader peptide and the protein of interest to cleave off the leader peptide and purify recombinant protein from the inclusion bodies. The expression vector of the embodiments herein has a sequence of SEQ ID No 1. The gene of interest may be inserted in any of the cleavage sites in the MCS.

10

The embodiments above are further explained through way of examples as follows:

15 Examples

Example 1: Construction of vector

The nucleotide sequence coding the leader peptide and the multiple cloning sites (MCS) were custom synthesized as single stranded oligonucleotides. The single stranded oligonucleotides were utilized for the synthesis of double stranded DNA fragment by overlapping PCR method.

20

The restriction enzyme (RE) sites were incorporated at 5' end and 3' end of the synthesized DNA fragment. The Promoter/Operator region, Ribosome binding site (RBS), origin of replication and antibiotic resistant gene were cleaved and ligated with the PCR amplified leader peptide sequence in MCS region containing unique restriction enzyme sites. The DNA fragment was cloned downstream of RBS between the *Nco I* and *Xho I* restriction site.

25

Thereafter, the positive clones were screened by PCR method and the nucleotide sequence of the cloned Leader sequence and MCS were confirmed by DNA sequencing for the correctness of nucleotides. The construction of vector employs standard techniques, reagents and/or kits.

Example 2: SDS PAGE analysis of GCSF expressed from the vector described herein

30

The sequence encoding for GCSF was incorporated in the MCS of the expression vector described herein along with in a control vector devoid of any leader sequence. In the vector

described herein, an enzymatic site for Enterokinase is inserted between the leader sequence and the GCSF sequence. The expression vector was cloned in bacterial cells and the GCSF inclusion bodies were obtained. The leader peptide was cleaved off by enzymatic and/or chemical means and the expression of GCSF from both the vectors was analysed on SDS PAGE as shown in Figure 2a. Lane 1 shows Medium molecule weight marker, Lane 2 shows expression sample GCSF from control vector, Lane 3 shows GCSF expression sample from pBG-BactX vector. As may be observed, there is negligible expression of GCSF from the control vector. Figure 2b illustrates gel densitometry data comparison for expression of GCSF in control vector and in the vector as described herein. Lane 1 shows Medium molecule weight marker, Lane 2 shows densitometry data for GCSF expression in the control vector and Lane 3 shows densitometry data for GCSF expression in the vector as described herein.

Example 3: Comparison of expression levels of GCSF by MALDI TOF analysis

Figure 3a and 3b illustrates comparison of GCSF expression in a control vector with GCSF expression in vector of Figure 1, according to an embodiment herein. The expression level in control vector is negligible whereas the expression level in the vector described herein has expression level of 30%.

We claim:

1. An expression vector for production of a recombinant protein in a host cell comprising a nucleotide sequence of Sequence ID No 2 encoding for a leader peptide of sequence ID No 3.
- 5 2. The expression vector of claim 1, wherein said expression vector expresses said recombinant protein as a fusion protein comprising fusion of said leader peptide of SEQ ID NO 3 and said recombinant protein.
3. The expression vector of claim 1, wherein, said host cell is bacteria, preferably *E. coli*.
4. The expression vector of claim 1, wherein said leader peptide has Methionine at N-terminus, followed by Glycine to impart stability to fusion of said recombinant protein and said leader peptide.
- 10 5. The expression vector of claim 1 further comprises DNA sequence encoding for a cleavage site or Restriction Enzyme (RE) site ligated to DNA sequence of said leader peptide.
6. The expression vector of claim 1 further comprises DNA sequence encoding a multiple cloning site (MCS) in upstream region of said leader peptide, a DNA sequence of said heterologous protein is cloned in said MCS; a DNA sequence encoding ribosome binding site (RBS) ligated to N-terminus of said leader peptide, a DNA sequence encoding a promoter or operator in the downstream of said ribosome binding site and DNA sequence encoding an antibiotic selection marker in upstream region of said promoter/operator sequence.
- 15 7. The expression vector of claim 6 wherein, said antibiotic selection marker is kanamycin.
8. A process for production of recombinant proteins comprising expressing said recombinant protein in a host cell through said expression vector of claim 1.
9. The expression vector of claim 1 having sequence as SEQ ID No 1.
- 20 10. The expression vector of claim 1 further comprising a nucleotide sequence encoding for Granulocyte Colony Stimulating Factor (GCSF) separated by said nucleotide sequence for leader peptide by an enterokinase cleavage site.