

(51) International Patent Classification:
C12N 15/00 (2006.01) *C12N 15/62* (2006.01)(21) International Application Number:
PCT/DK2010/050043(22) International Filing Date:
18 February 2010 (18.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PA 2009 00239 20 February 2009 (20.02.2009) DK
PA 2009 70302 28 December 2009 (28.12.2009) DK(71) Applicant (for all designated States except US):
AARHUS UNIVERSITET [DK/DK]; Nordre Ringgade
1, DK-8000 Århus C (DK).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MORTENSEN, Kim Kusk** [DK/DK]; Farrevej 17, DK-8464 Galten (DK). **HANSTED, Jon Gade** [DK/DK]; Carl Blochs Gade 27, 2. th., DK-8000 Århus C (DK). **SPERLING-PETERSEN, Hans Uffe** [DK/DK]; Parkvænget 6, DK-8700 Horsens (DK). **HÖG, Friederike** [DE/DE]; Zur Gartenstadt 232, D-12526 Berlin (DE). **PIETIKÄINEN, Laura** [FI/FI]; Puulinnankatu 4 E 43, FI-90570 Oulu (FI).(74) Agent: **PLOUGMANN & VINGTOFT A/S**; Sundkrogs-
gade 9, P.O. Box 831, DK-2100 Copenhagen Ø (DK).(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: EXPRESSIVITY TAG AND USE THEREOF

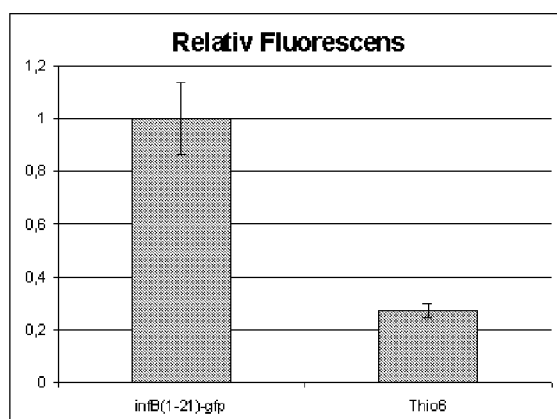


Fig. 27

(57) **Abstract:** The present invention relates to tags which increase the yield of protein expression. In particular the present invention relates to hybrid nucleic acid encoding a hybrid polypeptide comprising an expressivity tag, which increases the yield of the hybrid polypeptide during the manufacturing of the hybrid polypeptide. The invention furthermore relates to expressed polypeptides, vectors, host cells and methods for using the hybrid nucleic acid sequences for expression purposes.

Expressivity tag and use thereof

Technical field of the invention

The present invention relates to tags which increase the yield of protein

- 5 expression. In particular the present invention relates to hybrid nucleic acid encoding a hybrid polypeptide comprising an expressivity tag, which increases the yield of the hybrid polypeptide during the manufacturing of the hybrid polypeptide.

10 Background of the invention

The gram-negative bacterium *Escherichia coli* is one of biotechnologies most commonly used organisms for heterologous protein expression. It has many advantages justifying its use. It can grow on inexpensive substrates and still rapidly reach high cell densities. In the bacterial cell, production of the protein of

- 15 interest can reach sixty to seventy percent of total protein synthesized, giving a much wanted high product to cost ratio. The genome is well characterized and numerous different strains with specific properties are commercially available. Many biotechnological tools are also compatible with the organism, making it easy to manipulate for individual use of interest. In conclusion it is an ideal protein
- 20 factory for small to medium scaled productions. Being such a widely used expression host several issues also arise in using it for routine applications. Besides its limitations in not being able to make post-translational modifications and large complex proteins with multiple disulfide bonds, a remarkable difference exist in the efficiency in which both heterologous and homologous proteins are
- 25 produced. The bias ranges from super efficient expression of some proteins, to a total shut-down in others. The development of new and improved methods which secures a stable expression of any homologous and heterologous gene *E.coli* is therefore much wanted.

- 30 Protein synthesis is immensely complicated and a vast amount of variables are to be taken into account when evaluating expression profiles. Four main areas influence the absolute level of a protein: Transcription, mRNA stability, translation and protein half-life.

Expressivity tags are known to be able to enhance the yield of heterologous proteins expressed in *E. coli*. The tag, often placed 5' of the gene of interest, is usually a natural *E. coli* gene which expresses well, and for unknown reasons
5 promotes a higher level of translation of its 3' carrier gene (Wolfgang Peti a and Rebecca Page, 2007).

The first 471 nucleotides of the *InfB* gene termed *InfB*(1-471) encodes the domain I of *E.coli* Initiation Factor II (IF2). When fused to the aggregation prone
10 streptavidin protein and expressed in *E. coli*, domain I of IF2 is known to enhance its solubility (Sørensen et. al, 2003).

Improved expressivity tags which can increase expression of most proteins would be advantageous.
15

Summary of the invention

An object of the present invention relates to manufacturing an expression system providing a high yield of protein. In particular, it is an object of the present invention to provide a short expressivity tag, which may incorporated into a
20 heterologous polypeptide sequence of interest in order to increase the yield in a manufacturing process.

Expressivity tags are generally long (above 150 nucleotides) which is generally not favourable when you want to have a specific polypeptide expressed in high
25 amounts. 1) Long expressivity tags have a higher probability of interfering with the polypeptide of interest. 2) Long expressivity tags consume a larger proportion of the overall amino acids available.

Recombinant hybrid nucleic acid sequence

30 Thus, one aspect of the invention relates to a recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide, said nucleic acid comprising a first nucleic acid sequence of maximum 99 nt encoding a polypeptide operably linked to the a second heterologous nucleic acid sequence encoding a polypeptide,

wherein said first nucleic acid sequence comprises a nucleic acid sequence selected from the group consisting of

- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
- b) a nucleic acid sequence having at least 75% sequence identity to a),
- 5 and
- c) a nucleic acid sequence, which is a sub-sequence of a) and b).

Operably linked to a suitable promoter, the above mentioned hybrid nucleic acid sequence encoding a hybrid polypeptide sequence may be transcribed and
10 translated into said hybrid polypeptide in a suitable expression system.

Recombinant hybrid polypeptide

In another aspect the present invention relates to a recombinant hybrid polypeptide, said polypeptide comprising a first polypeptide amino acid sequence
15 and a second heterologous amino acid sequence, wherein said first amino acid sequence of maximum 33 amino acids is selected from the group consisting of

- a) an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12, and
- 20 b) an amino acid sequence having at least 75% sequence identity to a), and
- c) an amino acid sequence which is a sub-sequence of a) and b).

In order to express the hybrid polypeptide of the invention in a suitable host a
25 vector encoding the nucleic acid sequence of the invention incorporated in a vector. In particular such vector could be an expression vector.

Vector

Therefore, a third aspect of the present invention is to provide a vector comprising
30 a first nucleic acid sequence of maximum 99 nt, said first sequence comprising a nucleic acid sequence selected from the group consisting of

- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
- b) a nucleic acid sequence having at least 75% sequence identity to a), and

- c) a nucleic acid sequence, which is a sub-sequence of a) and b).

Host cell

An optimized vector can be introduced into a suitable host cell, which can be used
5 for expression the hybrid polypeptide.

Thus, a fourth aspect of the present invention is to provide a host cell comprising a vector according to the invention.

- 10 Apart from increasing the yield during manufacturing of the hybrid polypeptide, the expressivity tag be used as an affinity tag for an antibody binding.

Antibody

Thus, a fifth aspect of the present invention is to provide an isolated antibody or
15 antigen-binding fragment thereof which selectively binds to the first polypeptide amino acid sequence of the invention.

The incorporation of the expressivity tag of the invention into and operably linking the tag sequences to the nucleic acid sequences encoding a protein of interest
20 may increase the yield of the polypeptide in a manufacturing process.

Use of a recombinant hybrid nucleic acid sequence

Therefore, in a sixth aspect the present invention relates to the use of a recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide according
25 to the invention for the expression of a said polypeptide.

Method for expressing a recombinant polypeptide

An expression system capable of expressing a polypeptide in high amounts by the use of an expressivity tag would be advantageous.
30

Thus, in a seventh aspect the present invention relates to a method for expressing a recombinant polypeptide comprising

- a) introducing into an expression system the hybrid nucleic acid sequence according to the invention encoding a recombinant hybrid polypeptide,

- b) expressing said nucleic acid sequence in the expression system, and optionally
- c) purifying the expressed recombinant hybrid polypeptide .

5 Kit

A final aspect relates to a kit comprising a vector according of the invention.

Brief description of the figures

Figure 1

10 Figure 1 is a graphical representation of the protein secondary structure of domain I of IF2 and deletions in the *InfB*(1-471) gene. Deletions were placed in loop or random coil regions. Structure was obtained from PDB entry 1ND9 and tools at expasy.org.

15 Figure 2

Figure 2 shows *InfB*(1-x)-*InfA* constructs. Numbering refers to nucleotide position in coding region of the genes. *InfA* gene encodes the Initiation Factor I protein from *E.coli*. *InfB*(1-471) encodes the first domain of Initiation Factor II from *E.coli*. The six consecutive CAC codons encode a polyhistidine tag. The *SacI* site
20 codes for Glutamic acid-Leucine amino acids. GenBank. *InfB* GeneID: 5586604. *InfA* GeneID: 945500.

Figure 3

Figure 3 shows *InfB*(1-x)-*gfp* deletion constructs. Numbering refers to nucleotide
25 position in coding region of the genes. The *gfp* gene encodes the Green Fluorescence Protein (GFP) from *Aequorea victoria*. The six consecutive CAC codons encode a polyhistidine tag. The *SacI* site codes for Glutamic acid-Leucine amino acids. GenBank. *InfB* GeneID: 5586604. *gfp* GeneID 193239069.

Figure 4

Figure 4 is a representation of SDS-PAGE of expression of *InfB*(1-x)-*InfA* by BL21(DE3) pLysS pET15b cells. Samples were taken after 3.5 hours of 37° C growth with 0.8 mM IPTG. Commassie stained. Lanes: 1. *InfA*, 2. *InfB*(1-9)-*InfA*, 3. *InfB*(1-12)-*InfA*, 4. *InfB*(1-15)-*InfA*, 5. *InfB*(1-18)-*InfA*, 6. *InfB*(1-21)-*InfA*, 7. *InfB*(1-90)-*InfA*, 8. *InfB*(1-120)-*InfA*, 9. *InfB*(1-177)-*InfA*, 10. *InfB*(1-282)-*InfA*, and 11. *InfB*(1-471)-*InfA*. 12. Protein molecular mass marker.

Figure 5

Figure 5 is a representation of a Western blot of expression of *InfB*(1-x)-*InfA* by BL21(DE3) pLysS pET15b cells. Primary antibody used was targeted against IF1. Lanes: 1. *InfA*, 2. *InfB*(1-9)-*InfA*, 3. *InfB*(1-12)-*InfA*, 4. *InfB*(1-15)-*InfA*, 5. *InfB*(1-18)-*InfA*, 6. *InfB*(1-21)-*InfA*, 7. *InfB*(1-90)-*InfA*, 8. *InfB*(1-120)-*InfA*, 9. *InfB*(1-177)-*InfA*, 10. *InfB*(1-282)-*InfA*, 11. *InfB*(1-471)-*InfA*, and 12. IF1 marker protein.

15

Figure 6

Figure 6 is a representation of SDS-PAGE of expression of *InfB*(1-x)-*gfp* by (A) BL21 pBAD and (B) BL21(DE3) pLysS pET24d. Samples were taken after 3.5 hours of 37° C growth with (A) 0.02 % L-arabinose and (B) 0.8 mM IPTG. Commassie stained. Lanes: 1. Un-induced cells, 2. *gfp*, 3. *InfB*(1-9)-*gfp*, 4. *InfB*(1-12)-*gfp*, 5. *InfB*(1-15)-*gfp*, 6. *InfB*(1-18)-*gfp*, 7. *InfB*(1-21)-*gfp*, 8. *InfB*(1-90)-*gfp*, 9. *InfB*(1-120)-*gfp*, 10. *InfB*(1-177)-*gfp*, 11. *InfB*(1-282)-*gfp*, and 12. *InfB*(1-471)-*gfp*.

25 Figure 7

Figure 7 shows fluorescence from GFP moiety of fusion proteins expressed from *InfB*(1-x)-*gfp* in (A) BL21 pBAD and (B) BL21(DE3) pLysS pET24d. "÷" represents background. The fluorescence count of the *gfp* construct for each expression system is set to 1. A and B are therefore not directly comparable.

30

Figure 8

Figure 8 shows insoluble protein fractions of expression of *InfB*(1-x)-*gfp* from BL21 pBAD grown with 0.02 % L-arabinose at (A) 37° C and (B) 30° C. The bands observed in all lanes are chromosomal gene products. Insoluble proteins of

- 5 interest are marked with arrows. Lanes: 1. *gfp*, 2. *InfB*(1-9)-*gfp*, 3. *InfB*(1-12)-*gfp*, 4. *InfB*(1-15)-*gfp*, 5. *InfB*(1-18)-*gfp*, 6. *InfB*(1-21)-*gfp*, 7. *InfB*(1-90)-*gfp*, 8. *InfB*(1-120)-*gfp*, 9. *InfB*(1-177)-*gfp*, 10. *InfB*(1-282)-*gfp*, 11. *InfB*(1-471)-*gfp*.

Figure 9

- 10 Figure 9 shows insoluble protein fractions from expression of *InfB*(1-x)-*gfp* by BL21(DE3) pET24d pLysS with 0.8 mM IPTG at (A) 37° C, (B) 30° C and (C) 20° C. The bands observed in all lanes are chromosomal gene products. Lanes: 1. *gfp*, 2. *InfB*(1-9)-*gfp*, 3. *InfB*(1-12)-*gfp*, 4. *InfB*(1-15)-*gfp*, 5. *InfB*(1-18)-*gfp*, 6. *InfB*(1-21)-*gfp*, 7. *InfB*(1-90)-*gfp*, 8. *InfB*(1-120)-*gfp*, 9. *InfB*(1-177)-*gfp*, 10. *InfB*(1-282)-*gfp*, 11. *InfB*(1-471)-*gfp*.

Figure 10

- Figure 10 is a representation of SDS-PAGE of expression of *InfB*(1-x)-*gfp* by (A) BL21 pBAD and (B) BL21(DE3) pLysS pET24d. Samples were taken after 3.5 hours of growth with (A) 0.02 % L-arabinose at 30° C and (B) 0.8 mM IPTG at 20° C. Commassie stained. Lanes: 1. *gfp*, 2. *InfB*(1-9)-*gfp*, 3. *InfB*(1-12)-*gfp*, 4. *InfB*(1-15)-*gfp*, 5. *InfB*(1-18)-*gfp*, 6. *InfB*(1-21)-*gfp*, 7. *InfB*(1-90)-*gfp*, 8. *InfB*(1-120)-*gfp*, 9. *InfB*(1-177)-*gfp*, 10. *InfB*(1-282)-*gfp*, 11. *InfB*(1-471)-*gfp*.

25 Figure 11

- Figure 11 shows fluorescence from GFP moiety of fusion proteins expressed from *InfB*(1-x)-*gfp* in (A) BL21 pBAD at 30 ° C and (B) BL21(DE3) pLysS pET24d at 20 ° C. "÷" represents the background fluorescence. The fluorescence count of the *gfp* construct for each expression system is set to 1. A and B are therefore not directly comparable.

Figure 12

Figure 12 shows the purification of protein expressed from (1) *InfB*(1-9)-*gfp* (2) *InfB*(1-12)-*gfp* (3) *InfB*(1-15)-*gfp* (4) *InfB*(1-18)-*gfp* (5) *InfB*(1-21)-*gfp* constructs by IMAC on a 8 mL column of chelating sepharose, charged with Ni²⁺.

5 Relevant fractions were pooled and are shown in the SDS-PAGE.

Figure 13

Figure 13 is a representation of the MALDI-TOF spectra of purified protein from (A) *InfB*(1-9)-*gfp* (B) *InfB*(1-12)-*gfp* (C) *InfB*(1-15)-*gfp* (D) *InfB*(1-18)-*gfp* (E)

10 *InfB*(1-21)-*gfp* constructs. Mass spectra were acquired by a Voyager-DE PRO instrument. Final spectra obtained by averaging 150-200 single shot spectra which were externally calibrated with Bovine Insulin (5734,6 Da) and E.coli Thioredoxin (11674,5 Da).

15 Figure 14

Figure 14 is a representation of the values obtained from all spectra. The theoretical mass is calculated from the assumption that the N-formyl-methionine was cleaved off.

20 Figure 15

Figure 15 is a representation of the MALDI-TOF zoomed picture of [M+H⁺] ion of protein expressed from (A) *InfB*(1-9)-*gfp* and (B) *InfB*(1-21)-*gfp*. The peak with the lowest molecular weight (1) is the analyzed protein without an N-formyl-methionine group. The middle peak (2), seen only clearly in A, is the protein with the methionine residue still attached. The highest molecular weight peak (3) is the protein with the formyl-methionine group attached.

Figure 16

Figure 16 is a graphical representation of the folded mRNA sequences of 5'-end of 30 pET15b vector transcript (87 nt) followed by the first 50 nucleotides of *InfB*(1-x)-*InfA*. The online available MFOLD program was used. SD = Shine-Dalgarno

sequence. Start = Initiation start codon. dG is in units of kcal/mol. The table summarizes the dG values.

Figure 17

- 5 Figure 17 is a graphical representation of the folded mRNA sequences of 5'-end of pBAD vector followed by 50 first nucleotides of *InfB*(1-x)-gfp. Folding was done by the MFOLD program. dG is in units of kcal/mol. The table summarizes the dG values.

10 Figure 18

Figure 18 shows the coding sequences of *Trx* and *InfB* having expressivity tag properties. Matching bases are marked in bold.

Figure 19

- 15 Figure 19 shows the fluorescence of GFP moiety of fusion proteins expressed from *InfB*(1-21)-gfp with point mutations at position (A) 7 (B) 13 and (C) 16. Fluorescence is plotted with the frequency of the codon which the mutation caused. Wildtype bases are indicated with an asterix and set to a fluorescence count of 1. The frequency of the codon plotted at the 2nd y-axis is obtained from
- 20 a codon table of 8087 coding sequences in the *E.coli* genome. The fluorescence of gfp is represented with two dotted lines, each being the limits of the standard deviation on the dataset. Constructs were expressed in BL21 pBAD for 3.5 hours at 30° C with 0.02 % L-arabinose.

25 Figure 20

Figure 20 shows the numbers of A/U residues and CA repeat bases have been reported to be of importance for the efficiency of translation, and are shown for sequences of *InfB*, *InfA* and *gfp*.

Figure 21

Figure 21 shows *Thermus thermophilus* 70S ribosome, fMet-tRNA and mRNA in complex. pdb entry: 2HGR . A sphere of rRNA bases is shown around the mRNA at a 16 Å radius in the downstream tunnel.

5

Figure 22

Figure 22 shows the possible interaction between rRNA position C1397 and mRNA position +7.

10 Figure 23

Figure 23 shows the mRNA sequences of *InfB*(1-21), *InfA*(1-21) and *gfp*(1-21). Possible Watson-Crick base pair interactions between mRNA position +7 to rRNA C1397 are marked in bold.

15 Figure 24

Figure 24 shows the expression of LacZ gene as amount of β -galactosidase fusion protein in percentage of total protein synthesized. The codon producing the most protein is defined as 1, in both studies this was AAA (Lys).

20 Figure 25

Figure 25 shows primers used for making InfB(1-x)-InfA and InfB(1-x)-gfp and InfB(1-21)-gfp point mutants, constructs are listed in table I and II. Annealing temperatures for each PCR reaction is listed in the c°-column.

25 Figure 26

Figure 26 shows primers used for making InfB(1-21)-gfp (Table III). Screening primers are listed in table IV. Annealing temperatures for each PCR reaction is listed in the c°-column.

Figure 27

Figure 27 shows the relative expression of the Thio6-GFP system compared to the infB(1-21)-gfp system, measured by the fluorescence of the expressed reporter protein.

5

Figure 28

Figure 28 shows the expression levels of point mutations of infB(1-21)-GFP relative to infB(1-21)-GFP. The mutated nucleic acids and the resulting new codons are indicated below each graph.

10

Figure 29

Figure 29 shows an increased expression of proteins when these proteins comprise the infB(1-21)-tag compared to control proteins. (A) Streptavidin +/- infB(1-21)-tag. (B) L32 +/- infB(1-21)-tag. For more information see example 9.

15

Figure 30

Figure 30 shows sequences and characteristics of the used primers for amplification of infB(1-21)-l32, InfB(1-21)-streptavidin, l32 and streptavidin. These constructs are used in example 9.

20

Primers listed in figures 25, 26 and 30 are also listed in the sequence list as SEQ ID NO: 92-161.

The present invention will now be described in more detail in the following.

25

Detailed description of the invention

Definitions

Prior to discussing the present invention in further details, the following terms and
30 conventions will first be defined:

Expressivity tag

The term "expressivity tag" refers to nucleic acid sequences or amino acid sequences which increase the expression and thus the yield of a polypeptide to which the tag is operably linked.

5

Operably linked

The term "operably linked" refers to the connection of elements being a part of a functional unit such as a gene or an open reading frame. Accordingly, by operably linking a promoter to a nucleic acid sequence encoding a polypeptide the two
10 elements becomes part of the functional unit – a gene. The linking of the promoter to the nucleic acid sequence enables the transcription of the nucleic acid sequence directed by the promoter. By operably linking two heterologous nucleic acid sequences encoding a polypeptide the sequences becomes part of the functional unit – an open reading frame encoding a fusion protein comprising the
15 amino acid sequences encoding by the by the heterologous nucleic acid sequences. By operably linking two amino acids sequences, the sequences become part of the same functional unit – a polypeptide. Operably linking two heterologous amino acid sequences generates a hybrid (fusion) polypeptide.

20 *Recombinant*

Recombinant gene, promoter, nucleic acid sequence (DNA), amino acid sequences (polypeptide) refers to the products generated by genetic engineering such as the combination or insertion of one or more nucleic acid sequences, thereby combining nucleic acid sequences that would not normally occur together
25 (recombinant nucleic acid sequence, recombinant polynucleotide). Accordingly, recombinant expression refers to the expression of a RNA transcript and/or polypeptide from a coding region directed by a heterologous promoter or a promoter comprising heterologous promoter/enhancer elements.

30 *Heterologous sequences*

The term "heterologous" sequences refer to sequence elements of different sequences species. Thus, the fusion of nucleic acid sequences of one species of nucleic acids sequence to nucleic acid sequences of another species of nucleic acids generates a recombinant fusion nucleic acid sequence comprising

heterologous sequences. Likewise, the fusion of amino acid sequences of one species of polypeptide to amino acid sequences of another species of amino acid sequences generates a recombinant fusion amino acid sequences comprising heterologous sequences. Accordingly, elements originating from the same species
 5 of nucleic acids sequence or polypeptide are not considered heterologous according to the invention. Thus, according to the invention a nucleic acid sequence originating from *InfB* encoding an expressivity tag and a 5' subsequence of *InfB* encoding a amino acid sequences are not heterologous sequences according to the invention. The molecule formed by operably linking
 10 genetic/protein elements of heterologous genetic/protein species is referred to as a "chimera", "chimeric", fusion, or hybrid molecule.

Hybrid nucleic acid sequence

"Hybrid nucleic acid sequence" refers to a nucleic acid sequence (polynucleotide)
 15 comprising at least two heterologous nucleic acid sequences. The molecule is also referred to as "chimeric" nucleic acid sequence.

Hybrid polypeptide

"Hybrid polypeptide" or "hybrid amino acid sequences" refer to an amino acid
 20 sequence (polypeptide) comprising at least two heterologous amino acid sequences. The molecule is also referred to as "chimeric" polypeptide.

Sequence identity

The term "sequence identity" indicates a quantitative measure of the degree of
 25 homology between two amino acid sequences or between two nucleic acid sequences of equal length. If the two sequences to be compared are not of equal length, they must be aligned to give the best possible fit, allowing the insertion of gaps or, alternatively, truncation at the ends of the polypeptide sequences or

nucleotide sequences. The sequence identity can be calculated as $\frac{(N_{ref} - N_{dif})100}{N_{ref}}$,
 30 wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC ($N_{dif}=2$ and $N_{ref}=8$). A gap is counted as non-identity of

the specific residue(s), i.e. the DNA sequence AGTGTC will have a sequence identity of 75% with the DNA sequence AGTCAGTC (Ndif=2 and Nref=8).

With respect all embodiments of the invention relating to nucleotide sequences, the percentage of sequence identity between one or more sequences may also be

5 based on alignments using the clustalW software

(<http://www.ebi.ac.uk/clustalW/index.html>) with default settings. For nucleotide sequence alignments these settings are: Alignment=3Dfull, Gap Open 10.00, Gap Ext. 0.20, Gap separation Dist. 4, DNA weight matrix: identity (IUB).

Alternatively, and as illustrated in the examples, nucleotide sequences may be

10 analysed using programme DNASIS Max and the comparison of the sequences may be done at <http://www.paralign.org/>. This service is based on the two

comparison algorithms called Smith-Waterman (SW) and ParAlign. The first algorithm was published by Smith and Waterman (1981) and is a well established method that finds the optimal local alignment of two sequences. The other

15 algorithm, ParAlign, is a heuristic method for sequence alignment. Default settings for score matrix and Gap penalties as well as E-values were used.

Vector

The term "vector" refers to a DNA molecule used as a vehicle to transfer

20 recombinant genetic material into a host cell. The four major types of vectors are plasmids, bacteriophages and other viruses, cosmids, and artificial chromosomes.

The vector itself is generally a DNA sequence that consists of an insert (a heterologous nucleic acid sequence, transgene) and a larger sequence that serves as the "backbone" of the vector. The purpose of a vector which transfers genetic

25 information to the host is typically to isolate, multiply, or express the insert in the target cell. Vectors called expression vectors (expression constructs) are

specifically adapted for the expression of the heterologous sequences in the target cell, and generally have a promoter sequence that drives expression of the heterologous sequences. Simpler vectors called transcription vectors are only

30 capable of being transcribed but not translated: they can be replicated in a target cell but not expressed, unlike expression vectors. Transcription vectors are used to amplify the inserted heterologous sequences. The transcripts may subsequently be isolated and used in as templates suitable in vitro translations systems.

Functional equivalent

The term "functional equivalent" relates in the present context to a nucleic acid sequence or polypeptide sequence which pertains substantially the same activity as nucleic acid sequence or polypeptide sequence to which it is compared. The same functional activity (or functional equivalent) may be an activity of at least 25%, such as at least 40%, such as at least 60 %, such as at least 80% of the sequence to which it is compared. It is of course to be understood that a functional equivalent may have a higher activity such as at the most 400%, such as at the most 350%, such as at the most 300%, such as at the most 200%, such as at the most 175%, such as at the most 150%, or such as at the most 125% when compared to a sequence. In the present context the same functional activity as an expressivity tag can be measured by comparing the expression levels of the protein or peptides to which the tags are operably linked. An example of such measurement can be seen in Figure 28 and the corresponding text.

15

Single chain antibody

In the present context "single chain antibody", "single domain antibody", "sdAb" (also called Nanobody) is an antibody fragment consisting of a single monomeric variable antibody domain. Like a whole antibody, it is able to bind selectively to a specific antigen. Single domain antibodies are much smaller than common antibodies.

20

Hybrid nucleic acid sequence

In one aspect the invention relates to a recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide, said nucleic acid comprising a first nucleic acid sequence of maximum 99 nt encoding a polypeptide operably linked to the a second heterologous nucleic acid sequence encoding a polypeptide, wherein said first nucleic acid sequence comprises a nucleic acid sequence selected from the group consisting of

30

- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
- b) a nucleic acid sequence having at least 75% sequence identity to a), and
- c) a nucleic acid sequence, which is a sub-sequence of a) and b).

Length of the first nucleic acid sequence

The size of the first nucleic acid sequence is preferably shorter while maintaining an effect as an expressivity tag.

Thus, in one embodiment the maximum length of the first nucleic acid sequence is
5 90 nucleotides, such as 80 nucleotides, 70 nucleotides, 60 nucleotides, 50 nucleotides, 40 nucleotides, 30 nucleotides, or such as 24 nucleotides,

Substitution may be introduced in the nucleic acid sequences of the invention provided that the modified nucleic acid sequence is a functional equivalent of the
10 nucleic acid sequences from which it originates. Thus, the modified nucleic acid sequence encodes an expressivity tag. The modification of the coding sequences may be advantageous in order to optimize codons of the coding region for optimal translation in the host of production. Moreover, minor modification of the nucleic acid sequences may generate restrictions site, which facilitate the cloning of the
15 sequences.

Thus, in another embodiment the first nucleic acid sequence has a sequence identity of at least 80% such as at least 85%, at least 90%, at least 95% or such as 100% to a amino acid sequence selected from the group consisting of SEQ ID
20 NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.

Different nucleic acid sequences which are subsequences of one or more of the sequences of the invention may also be used to obtain an effect similar to the ones disclosed in this invention. Especially sequences having a length between 15 and 99 nucleotides are likely to have an effect similar to the claimed sequences.
25

Therefore, in yet another embodiment the sub-sequences of SEQ ID NO:1 have a length of at least 23 nucleotides, such as 33 nucleotides, 43 nucleotides, 53 nucleotides, 63 nucleotides, 73 nucleotides, or such as 83 nucleotides.

30 It is to be understood that these subsequences are functional equivalent to SEQ ID NO:1.

Examples of second heterologous nucleic acid sequences encoding polypeptides which could be operably linked to the first nucleic acid sequences are GFP,
35 streptavidin, single chain antibodies such as L32 and MMP's. See e.g. examples 3,

9 and 10 and corresponding figures. By having increased expression of GFP more sensitive reporter systems may be constructed. Similar, streptavidin is used in many assays, especially in vitro assays, thus by having a high expression level cheaper assays may be constructed. Thus, in an embodiment according to the invention the second heterologous nucleic acid sequence is selected from the group consisting of reporter constructs such as GFP, streptavidin, MMP's and single chain antibodies such as L32. These are mere examples of nucleic acids (or proteins) which could be used in connection with the present invention. The person skilled in the art could come up with other relevant proteins.

10

Expressivity tag

To increase the expression of the hybrid polypeptide it is advantageous if the first nucleic acid sequence is coding for an expressivity tag.

15 Therefore, in an embodiment of the invention, the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, wherein said first nucleic acid sequence encodes an expressivity tag.

Length of expressivity tag

20 A minimum length is required of the first nucleic acid sequence of the expressivity tag to obtain an increased expression of the hybrid nucleic acid sequence, compared to when an expressivity tag is not operably linked to a second nucleic acid sequence.

Therefore, in another embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, wherein said first nucleic acid sequence of c) has a minimum length of 15 nucleotides. This minimum length could e.g. also be 16 nucleotides, such as 17 nucleotides, or such as 18 nucleotides.

30 Position of Expressivity tag

To maximize the effect of the expressivity tag of the invention, the tag can be positioned in different positions in relation to the rest of the expressed polypeptide.

Thus, in a further embodiment the invention relates to a recombinant hybrid nucleic acid sequence of the invention, wherein said first nucleic acid sequence is positioned 5' to the second heterologous nucleic acid sequence.

- 5 Preferably the first nucleic acid sequence is positioned in such a way that the expressivity tag will be positioned in the N-terminal end of the expressed fusion polypeptide. Similar, this means that preferably the first nucleic acid sequence is positioned at the 5'-end of the hybrid nucleic acid sequence.

10 Codon preference

- It has been shown, that proteins that are expressed in high amounts within the cell, have a statistical bias towards the codons that are overall more frequently used within a genome. The different tRNA species does not exist in equal concentrations within the cell. Actually, the concentration of each tRNA species
- 15 has been shown to correlate with its cognate codon usage. This relationship is thought to be a result of a optimized translation machinery utilizing a balanced codon to tRNA ratio to maximize cell growth. It is also thought to serve to regulate the rate at which different transcripts are translated. When a gene is introduced into E. coli from a foreign organism, its codon context might be
- 20 unfavorable for high level expression. In general, the more codons a gene contains that are rarely used in the expression host, the less likely it is that the protein will be expressed at reasonable levels. Since optimal expression of a protein can depend on the codon preference of the host organism selected to express the hybrid nucleic acid sequence, it is advantageous to optimize the
- 25 nucleic sequences to the host.

- Thus, in yet a further embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, wherein said second heterologous nucleic acid sequence has been modified to reflect the codon
- 30 preference of a host organism selected to express the nucleic acid molecule.

Nucleic acid comprising tag

- The hybrid nucleic acid sequence may further comprise one or more additional elements which could be advantageous for detection, solubility or affinity
- 35 purification of the expressed hybrid polypeptide.

Thus, in an additional embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, further comprising a nucleic acid sequence encoding a polypeptide tag for detection, solubility or affinity

5 purification.

In a further embodiment, this/these one or more tag-elements is positioned in the hybrid nucleic acid sequence depending on the precise use and functionality of the element. It/they are e.g. positioned 3' to the first nucleic acid sequence, between
10 the first and second nucleic acid sequences or 3' to the second nucleic acid sequence. In special cases it/they are positioned 5' to the first nucleic acid sequence.

A wide-range of different elements may be employed for detection, solubility or
15 affinity purification. A skilled person will be able to choose the tag which is optimal for his/her preferred use. The skilled person will furthermore, be able to choose both primary and secondary antibodies which can be used for either purification or detection.

20 Thus, in yet an additional embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, wherein said tag for affinity purification, solubility or detection is selected from the group consisting of BCCP-tag, c-myc-tag, calmodulin-tag (SEQ ID NO:33), FLAG-tag (SEQ ID NO:45), HA-tag (SEQ ID NO:50), His-tag (SEQ ID NO:29), Maltose binding protein-tag
25 (SEQ ID NO:27), Nus-tag (SEQ ID NO:60), Glutathione-S-transferase-tag (SEQ ID NO:31), Green fluorescent protein-tag (SEQ ID NO:62), Thioredoxin-tag (SEQ ID NO:64), S-tag (SEQ ID NO:41), Strep-tag (SEQ ID NO:35), human protein C tag (SEQ ID NO:37), Chitin binding protein tag (SEQ ID NO:39), T7-tag (SEQ ID NO:43), Myc-tag (SEQ ID NO:52), V5-tag, VSV-tag, Avi.tag (SEQ ID NO:56),
30 BioEase-tag (SEQ ID NO:58), SNAP-tag, FlaSH-tag, Nus A-tag, DsbA-tag (SEQ ID NO:66) and the recombinant nucleic acid sequence itself.

Spacer sequence

In certain applications it is advantageous to position a spacer sequence
35 somewhere in the hybrid nucleic acid sequence.

Thus, in yet an embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, further comprising a nucleic acid sequence encoding a spacer positioned between said first and said second nucleic acid sequence.

In special cases the first nucleic acid sequence may be divided into two or more sections using one or more spacer sequences.

Thus, in yet an embodiment the invention relates to a recombinant hybrid nucleic acid sequence according to the invention, further comprising a nucleic acid sequence encoding a spacer positioned between the first and second codon of said first nucleic acid sequence and operably linking said codons.

Introduction of one or more spacer sequences is preferred for example if there is a need for a specific restriction site in either the nucleic acid sequence or the polypeptide sequence. A restriction at the DNA level may be useful for cloning purposes. Restriction sites can e.g. be cleaved using either restriction or nicking enzymes. Cleavage at the peptide level may be useful if part of the hybrid amino acid sequence is not desired in the final peptide. For example a sequence encoding for specific proteases can be introduced between the first and second nucleic acid sequences. In this way the encoded first amino acid sequence can be removed from the final peptide. This removal can take place both before or after a purification step. Examples of useful proteases are Factor Xa (fXa), TEV, PreScission Protease.

Introduction of one or more spacer sequences may also be favourable to avoid structural interference between the first and second amino acid sequences in the hybrid polypeptide.

30 Recombinant hybrid polypeptide

In a preferred embodiment, the hybrid nucleic acid sequence of the invention is transcribed generating a transcript from which the hybrid polypeptide is translated.

Accordingly, one aspect the invention relates to a recombinant hybrid polypeptide, said polypeptide comprising a first polypeptide amino acid sequence and a second heterologous amino acid sequence, wherein said first amino acid sequence of maximum 33 amino acids is selected from the group consisting of

- 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12, and
- b) an amino acid sequence having at least 75% sequence identity to a), and
- 10 c) an amino acid sequence which is a sub-sequence of a) and b).

The maximum length of the first amino acid sequence could be shorter while maintaining an effect as an expressivity tag.

15 Length of first amino acid sequence

Thus, in an embodiment the invention relates to a maximum length of the first amino acid sequence being 30 amino acids, such as 26 amino acids, 23 amino acids, 20 amino acids, 17 amino acids, 14 amino acids, 11 amino acids, or such as 9 amino acids.

20

Sequence identity of amino acid sequence

It may also be possible to change some of the amino acids in the sequences of the invention and still maintain an effect as an expressivity tag. Substitution may be introduced in the amino acids in the sequences of the invention provided that the
25 modified amino acid the sequence is a functional equivalent of the amino acid the sequence from which it originates. Thus, the modified amino acid the sequence is an expressivity tag.

Thus, one embodiment the invention relates a first amino acid sequence having a
30 sequence identity of at least 80% such as at least 85%, at least 90%, at least 95% or such as 100% to a amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12.

Subsequences

Different amino acid sequences which are subsequences of one or more of the sequences of the invention may also be used to obtain an effect similar to the ones disclosed in this invention. Especially sequences having a length between 4
5 and 33 amino acids are likely to have an effect similar to the claimed sequences.

Therefore, in yet a further embodiment the sub-sequences of SEQ ID NO:5 and SEQ ID NO:6, are at least 9 amino acids, such as 15 amino acids, or such as 21 amino acids. or such as 27 amino acids.

10

It is to be understood that these subsequences are functional equivalent to SEQ ID NO:5 and SEQ ID NO:6.

Removal of methionine

15 In bacteria, protein synthesis initiates with a formyl-methionine. The formyl group is removed post-translationally by a peptide deformylase (PDF). The methionine residue is, depending on the identity of the second amino acid in the peptide chain, removed by a methionine aminopeptidase (MAP).

20 Accordingly, one embodiment of the present invention relates to an isolated hybrid polypeptide comprising a first polypeptide amino acid sequence selected from the group consisting of SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12. The first polypeptides of the embodiment reflect the removal of the N-terminal methionine from the corresponding first polypeptides SEQ ID NO:5, SEQ
25 ID NO:7, SEQ ID NO:9, SEQ ID NO:11 respectively.

Expressivity tag

To increase the expression of the hybrid polypeptide it is advantageous if the first amino acid sequence is coding for an expressivity tag.

30 Therefore, in an embodiment of the invention, the invention relates to a recombinant hybrid polypeptide according to the invention, wherein said first amino acid sequence is an expressivity tag.

Minimum length of expressivity tag

A minimum length is naturally required of the amino acid sequence of the expressivity tag to obtain an increased expression of the hybrid polypeptide sequence, compared to when an expressivity tag is not operably linked to a

5 second nucleic amino acid sequence.

Therefore, in another embodiment the invention relates to a recombinant hybrid polypeptide according to the invention, wherein said first amino acid sequence of c) has a minimum length of 4 amino acids. This minimum length could e.g. also be 5 amino acids, such as 6 amino acids, such as 7 amino acids, such as 8 amino
10 acids, such as 9 amino acids, or such as 10 amino acids. It is to be understood that the minimum lengths cited above relates to said first amino acid sequence when the N-terminal methionine has been removed, as described above.

Positioning of expressivity tag

15 To obtain the full effect of the expressivity tag of the invention, correct positioning of the tag in relation to the rest of the polypeptide to be expressed may be relevant.

Thus, in a further embodiment the invention relates to a recombinant hybrid
20 polypeptide according to the invention, wherein said first amino acid sequence is positioned N-terminal to the second heterologous amino acid sequence.

Preferably the first amino acid sequence is positioned in such a way that the expressivity tag will be positioned at the N-terminal end of the expressed fusion
25 polypeptide.

Tag on amino acid sequence

The hybrid amino acid sequence may further comprise one or more additional elements which could advantageous for detection, solubility or affinity purification
30 of the expressed hybrid polypeptide.

Thus, in an additional embodiment the invention relates to a recombinant hybrid polypeptide according to the invention, further comprising a polypeptide tag for detection, solubility or affinity purification.

This/these one or more tag-elements can be positioned in the hybrid amino acid sequence depending on the precise use and functionality of the element.

Accordingly, the tags for detection, solubility or affinity purification may be

- 5 positioned C-terminal to the first amino acid sequence, between the first and second amino acid sequences or C-terminal to the second nucleic acid sequence. In particular embodiments the one or more tag-elements are positioned N-terminal to the first amino acid sequence

- 10 Many different elements can be used for detection, solubility or affinity purification. A skilled person will be able to choose the tag which is optimal for his/her preferred use. The skilled person will furthermore, be able to choose both primary and secondary antibodies which can be used for either purification or detection.

15

- Thus, in an additional embodiment relates to a recombinant hybrid polypeptide sequence according to the invention, wherein said tag for affinity purification, solubility or detection is selected from the group consisting of BCCP-tag, c-myc-tag, calmodulin-tag (SEQ ID NO:34), FLAG-tag (SEQ ID NO:46, SEQ ID NO:47, 20 SEQ ID NO:48 and SEQ ID NO:49), HA-tag (SEQ ID NO:51), His-tag (SEQ ID NO:30), Maltose binding protein-tag (SEQ ID NO:28), Nus-tag (SEQ ID NO:61), Glutathione-S-transferase-tag (SEQ ID NO:32), Green fluorescent protein-tag (SEQ ID NO:63), Thioredoxin-tag (SEQ ID NO:65), S-tag (SEQ ID NO:42), Strep-tag (SEQ ID NO:36), human protein C tag (SEQ ID NO:38), Chitin binding protein 25 tag (SEQ ID NO:40), T7-tag (SEQ ID NO:44), Myc-tag (SEQ ID NO:53), V5-tag (SEQ ID NO:54), VSV-tag (SEQ ID NO:55), Avi.tag (SEQ ID NO:57), BioEase-tag (SEQ ID NO:59), SNAP-tag, FlaSH-tag, Nus A-tag, DsbA-tag (SEQ ID NO:67) and the recombinant polypeptide sequence itself.

30 Vector

For recombinant expression of the hybrid polypeptide the recombinant hybrid nucleic acid of the invention may be incorporated in a vector for efficient transfer to and expression of the recombinant gene in a suitable host cell.

Therefore in a further aspect the invention relates to a vector comprising a first nucleic acid sequence of maximum 99 nt, said first sequence comprising a nucleic acid sequence selected from the group consisting of

- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
- 5 b) a nucleic acid sequence having at least 75% sequence identity to a),
and
- c) a nucleic acid sequence, which is a sub-sequence of a) and b), said
vector.

- 10 In an embodiment the maximum length of the first nucleic acid sequence
comprised in a vector is 90 nucleotides, such as 80 nucleotides, 70 nucleotides,
60 nucleotides, 50 nucleotides, 40 nucleotides, 30 nucleotides, or such as 24
nucleotides,
- 15 Thus, in another embodiment the first nucleic acid sequence has a sequence
identity of at least 80% such as at least 85%, at least 90%, at least 95% or such
as 100% to a amino acid sequence selected from the group consisting of SEQ ID
NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4.
- 20 In a particular embodiment the sub-sequences of SEQ ID NO:1 comprised in a
vector have a length of at least 23 nucleotides, such as 33 nucleotides, 43
nucleotides, 53 nucleotides, 63 nucleotides, 73 nucleotides, or such as 83
nucleotides.
- 25 It is to be understood that these subsequences are functional equivalent to SEQ
ID NO:1.

- The first nucleic acid sequences of the invention encoding the expressivity tag
may be cloned into a vector: The vector comprising the first nucleic acid
- 30 sequences may serve as a cloning vector for subsequent cloning of the second
heterologous nucleic acid sequence encoding the polypeptide of interest. The
second heterologous nucleic acid sequence is inserted in a position that operably
links the sequence to the first nucleic acid sequences encoding the expression tag.
In one embodiment, the cloning vector further comprises a least one promoter

element or a complete promoter operably linked to said first nucleic acid sequence.

Thus, in a further embodiment the invention relates to a vector according to the invention, wherein said vector is further adapted for insertion of a second heterologous nucleic acid sequence encoding a polypeptide operably linked to said first nucleic acid sequence.

Such second heterologous nucleic acid sequence could be incorporated into the vector by standard cloning techniques known to the skilled person such as PCR-based techniques.

A vector comprising the second nucleic acid sequence operably linked to the first nucleic acid sequence (the expression tag) encodes a hybrid polypeptide for efficient expression in a suitable host.

Therefore, in a further embodiment the invention relates to a vector comprising a recombinant hybrid nucleic acid sequence according to the invention.

The vector may be further adapted for expression of the hybrid polypeptide by insertion of transcriptional control elements such as a promoter or a promoter including additional promoter/enhancer elements.

Thus, to direct the expression (transcription) of the hybrid nucleic acid sequence inserted in a vector transcriptional control may be inserted thereby generating an expression vector. The use transcriptional control elements depend on the expression system and host. The promoter of the expression vector may comprise several independent sequence elements. These elements may both be elements which can promote expression, elements which can repress expression or a combination of both. In one embodiment, the expression control sequence comprises transcriptional control elements suitable for conditional expression of the recombinant hybrid nucleic acid sequence.

Thus, in a further embodiment the invention relates to a vector according to the invention, wherein said recombinant hybrid nucleic acid sequence is operably linked to an expression control sequence, such as the araBAD promoter system

such as the Ara-promoter (SEQ ID NO:13), Ara-operator 1 (SEQ ID NO:14), Ara-operator 2 (SEQ ID NO:15) and the AraC repressor (SEQ ID NO:16), the lac promoter system such as the Lac Promoter (SEQ ID NO:17), the Lac-operator (SEQ ID NO:18) and the LacI repressor (SEQ ID NO:19), Tac/Trc-promoter system (SEQ ID NO:20), Lambda PL – promoter (SEQ ID NO:21), Lambda PR promoter (SEQ ID NO:22), T7 promoter (SEQ ID NO:23), T3 promoter (SEQ ID NO:24), SP6 promoter (SEQ ID NO:25) and the AmpR promoter (SEQ ID NO:26).

Different types of vectors may be used for incorporation of a tag sequence according to the invention. The choice of tag depends on the specific use of the vector, how to clone the vector and how to express the inserted sequence. Therefore, in yet an embodiment the invention relates to a vector according to the invention, wherein the vector is selected from the group consisting of plasmids, cosmids, phages, bacterial artificial chromosomes (BAC), Phagemids and P1-derived artificial chromosomes.

Furthermore, the nucleic acid sequences of the invention may also be positioned in oligonucleotides for *in vitro* translation and/or transcription purposes. The nucleic acid sequences of the invention may also be positioned in PCR-products for cloning purposes and for *in vitro* transcription and/or translation purposes.

The vectors may both be single stranded or double stranded. Furthermore, they may be linear circular or nicked. Especially when using *in vitro* translation systems it may be advantageous to use nicked or linear plasmids.

In vitro translation vector

Thus, in yet a further embodiment the invention relates to a vector according to the invention, wherein the vector is an *in vitro* translation vector.

In vitro translations systems are compatible with most *in vivo* expression system and vice versa. Therefore it is to be understood that the *in vitro* vectors can also be used for *in vivo* expression, and that *in vivo* expression vectors can be used for *in vitro* expression.

In a particular embodiment the *in vitro* translation system and or PCR fragments and vectors are selected from the list consisting of:

- Expressway™ Cell-Free Expression Systems from Invitrogen using vectors such as pEXP5-NT/TOPO®, pEXP5-CT/TOPO®, pEXP1-DEST, pEXP2-DEST, pEXP3-DEST, and pEXP4-DEST.
- The EasyXpress Linear Template system from Qiagen where transcription is directly from PCR-fragments followed by translation.
- The EasyXpress system, where no expression vector is specified
- EcoPro™ T7 System from Novagen where circular or linear DNA can be used under control of *E. coli* or T7 promoters.
- Single Tube Protein® System 3 (STP3) using plasmid DNA or linear DNA under T7 or SP6 promoter control.
- S30 T7 High-Yield Protein Expression System from Promega where all vectors with a T7 promoter or strong *E. coli* promoter can be used.
- PROTEINscript™ II from Applied Biosystems where PCR fragments, supercoiled DNA or linear DNA templates with T7 polymerase promoters are used.

Plasmid vector

- As described above many different vectors can be used to carry out different aspects of the invention. In some applications a plasmid vector is preferred. Thus, in a specific embodiment the invention relates to a vector according to the invention, wherein said plasmid is selected from the group consisting of TA cloning vectors, Gateway cloning vectors, restriction cloning vectors, Topo cloning vectors, pET vector system and pBAD vector systems.

- In one particular embodiment the vector is selected from the group consisting of Charon 4, EZ-Tn5 pMOD-2, EZ-Tn5 pMOD-3, EZ-Tn5 pMOD-4, EZ-Tn5 pMOD-5, LITMUS28i, LITMUS38i, PinPoint Control, PinPoint Xa Control, PinPoint Xa-1, PinPoint Xa-2, PinPoint Xa-3, pACYC177, pACYC184, pACYCDuet-1, pALTER-EX1, pALTER-EX2, pBAC-2cp, pBACgus-2cp, pBAD-DEST49, pBAD-TOPO, pBAD-TOPO/LacZ, pBAD-TOPO/LacZ/V5-His, pBAD/His A, pBAD/His B, pBAD/His C, pBAD/His/LacZ, pBAD/Myc-His/lacZ, pBAD/Thio, pBAD/Thio-E, pBAD/Thio-TOPO, pBAD/gIII A, pBAD/gIII B, pBAD/gIII C, pBAD/gIII Calmodulin, pBAD/myc-His A, pBAD/myc-His B, pBAD/myc-His C, pBAD102/D-TOPO, pBAD102/D/LacZ, pBAD18, pBAD202/D-TOPO, pBAD202/D/LacZ, pBEc-Q, pBEc-SBP, pBEc-SBP-Q,

pBEc-SBP-SET1, pBEc-SBP-SET1-Q, pBEc-SBP-SET2, pBEc-SBP-SET2-Q, pBEc-SBP-SET3, pBEc-SBP-SET3-Q, pBEc-SET1, pBEc-SET2, pBEc-SET3, pBEn-SBP-SET1a, pBEn-SBP-SET1b, pBEn-SBP-SET1c, pBEn-SBP-SET2a, pBEn-SBP-SET2b, pBEn-SBP-SET2c, pBEn-SBP-SET3a, pBEn-SBP-SET3b, pBEn-SBP-SET3c, pBEn-
 5 SBPa, pBEn-SBPb, pBEn-SBPc, pBEn-SET1a, pBEn-SET1b, pBEn-SET1c, pBEn-SET2a, pBEn-SET2b, pBEn-SET2c, pBEn-SET3a, pBEn-SET3b, pBEn-SET3c,
 pBR322, pBiEx-1, pBiEx-2, pBiEx-3, pCAL-c, pCAL-kc, pCAL-n, pCAL-n-EK, pCAL-n-FLAG, pCC1BAC, pCDF-2 Ek/LIC, pCDFDuet-1, pCMV-GLuc, pCOLADuet-1, pCR2.1, pCRT7-E, pCX, pCX-TOPO, pDEST14, pDEST15, pDEST17, pDEST24,
 10 pDONR201, pET-11a, pET-11b, pET-11c, pET-11d, pET-14b, pET-15b, pET-16b, pET-17b, pET-19b, pET-20b(+), pET-21(+), pET-21a(+), pET-21b(+), pET-21c(+), pET-21d(+), pET-22b(+), pET-23a(+), pET-23b(+), pET-23c(+), pET-23d(+), pET-24(+), pET-24a(+), pET-24b(+), pET-24c(+), pET-24d(+), pET-25b(+), pET-26b(+), pET-27b(+), pET-28a(+), pET-28b(+), pET-28c(+), pET-
 15 29a(+), pET-29b(+), pET-29c(+), pET-30 Ek/LIC, pET-30 Xa/LIC, pET-30a(+), pET-30b(+), pET-30c(+), pET-31b(+) DNA, pET-32 Ek/LIC, pET-32 Xa/LIC, pET-32a(+) DNA, pET-32b(+) DNA, pET-32c(+) DNA, pET-33b(+), pET-39b(+), pET-3a, pET-3b, pET-3c, pET-3d, pET-40b(+), pET-41 Ek/LIC, pET-41a(+), pET-41b(+), pET-41c(+), pET-42a(+), pET-42b(+), pET-42c(+), pET-43.1 Ek/LIC,
 20 pET-43.1a(+), pET-43.1b(+), pET-43.1c(+), pET-44 Ek/LIC, pET-44a(+), pET-44b(+), pET-44c(+), pET-45b(+), pET-46 Ek/LIC, pET-47b(+), pET-48b(+), pET-49b(+), pET-50b(+), pET-51 Ek/LIC, pET-51b(+), pET-52 3c/LIC, pET-52b(+), pET-5a, pET-5b, pET-5c, pET-9a, pET-9b, pET-9c, pET-9d, pET-DEST42,
 pET100/D-TOPO, pET100/D/LacZ, pET101/D-TOPO, pET101/D/LacZ, pET102/D-TOPO, pET102/D/LacZ, pET151/D-TOPO, pET151/D/LacZ, pET160-GW/CAT, pET160/GW/D-TOPO, pET161-GW/CAT, pET161/GW/D-TOPO, pET200/D-TOPO, pET200/D/LacZ, pETBlue-1, pETBlue-2, pETDuet-1, pETcoco-1, pETcoco-2,
 pEpiFOS-5, pF1A T7, pF1K T7, pF3A WG (BYDV), pF3K WG (BYDV), pF4A CMV, pF4K CMV, pF5A CMV-neo, pF5K CMV-neo, pF9A CMV hRluc-neo, pFC7A (HQ),
 30 pFC7K (HQ), pFC8A (HaloTag), pFC8K (HaloTag), pFN10A (ACT), pFN11A (BIND), pFN2A (GST), pFN2K (GST), pFN6A (HQ), pFN6K (HQ), pG5luc, pGEM-1, pGEM-11Zf(+), pGEM-11Zf(-), pGEM-13Zf(+), pGEM-15Zf(-), pGEM-2, pGEM-3, pGEM-3Z, pGEM-3Zf(+), pGEM-3Zf(-), pGEM-4 Vector, pGEM-4Z, pGEM-5Zf(+), pGEM-5Zf(-), pGEM-7Zf(+), pGEM-7Zf(-), pGEM-9Zf(-), pGEMEX-1, pGEMEX-2, pGEX-
 35 2T, pGEX-2TK, pGEX-3X, pGEX-4T-1, pGEX-4T-2, pGEX-4T-3, pGEX-5X-1, pGEX-

5X-2, pGEX-5X-3, pGEX-6P-1, pGL2-Basic, pGL2-Control, pGL2-Enhancer, pGL2-Promoter, pGLuc-Basic Vector, pGPS1.1, pGPS2.1, pGPS3, pGPS4, pGPS5, pIndigoBAC-5, pKLAC1, pKLAC1-malE, pKO500, pLacI, pLysE, pLysS, pMAL-c4E, pMAL-c4G, pMAL-c4X, pMAL-p4E, pMAL-p4G, pMAL-p4X, pMAL-pIII, pNEB193, pNEB206A, pNEBR-R1, pNEBR-X1, pNEBR-X1GLuc, pNEBR-X1Hygro, pRS1274, pRS308, pRS414, pRS415, pRS475, pRS550, pRS551, pRS552, pRS577, pRS591, pRSET A, pRSET B, pRSET C, pRSET T7, pRSET-E, pRSET/LacZ, pRSF-2 Ek/LIC, pRSFDuet-1, pSP64, pSP64 Poly(A), pSP65, pSP70, pSP72, pSP73, pTWIN1, pTXB1, pTXB3, pTYB1, pTYB11, pThioHis A, pThioHis B, pThioHis C, pTrcHis A, pTrcHis B, pTrcHis C, pTrcHis-TOPO, pTrcHis-TOPO/lacZ, pTrcHis/CAT, pTrcHis2 A, pTrcHis2 B, pTrcHis2 C, pTrcHis2-TOPO, pTrcHis2-TOPO/LacZ, pTrcHis2/LacZ, pUC19, pWEB, and pWEB-TNC.

Host cell

15 For cloning and expression purposes it may be useful to have the vector cloned in a host cell. In this way the vector can be maintained, cloned and purified and thus provide a source for almost unlimited numbers of the vectors. Moreover, stable producer cell lines expressing the hybrid polypeptide of interest may also be advantageous. In order to maintain a producer cell selection marker may be
20 incorporated in the expression vector.

Thus in an embodiment the invention relates to a host cell comprising a vector according to the invention.

25 As described above bacteria are in many instances optimal host cells for e.g. expression purposes.

Bacteria

Thus, in yet an embodiment the invention relates a host cell according to the
30 invention, wherein said host cell is a bacterium.

Examples of bacterial strains which may be suitable as host cells are *Bacillus licheniformis*, *Bacillus subtilis* and *E.coli*.

Bacillus licheniformis and *Bacillus subtilis* is e.g. cultured in order to obtain enzymes, such as proteases, for use in biological washing powder, and it may increase production if the expressivity tag according to the invention is positioned in frame in front of such proteases. Enzymes produced by *B. subtilis* and *B.*

5 *licheniformis* are widely used as additives in laundry detergents.

It may be preferred that the bacterium used for hosting a vector is an *E. coli* strain. Thus, in yet an embodiment the invention relates a host cell according to the invention, wherein said host cell is *E. coli*.

10

In an embodiment of the invention the host cell is selected among AG1, AB1157, BL21(AI), BL21(DE3), BL21 (DE3) pLysS, BNN93, BNN93 (λ gt11), BNN97, BW26434 (CGSC Strain # 7658), B5615, B834, B834 (DE3), B834(DE3) pLysS, BLR(DE3), BLR(DE3) pLysS, C600, C600 hflA150 , Y1073, BNN102, CSH50,

15 D1210, DB3.1, DH1, DH5 α , DH10B, DH12S, DM1ER2566, ER2267, HB101, HMS174(DE3), HT96 NovaBlue, HT96 NovaBlueF-, IJ1126, IJ1127, JM83, JM101, JM103, JM105, JM106, JM107, JM108, JM109, JM109(DE3), JM110, JM2.300, LE392, Mach1, MC1061, MC4100, MG1655, OmniMAX2, Origami , Origami 2 , Origami B, Rosetta, RosettaBlue, Rosetta-gami, Rosetta-gami 2, Rosetta-gami B,
20 Rosetta(DE3)pLysS, Rosetta-gami(DE3)pLysS, RR1, STBL2, STBL3, STBL4, SURE, SURE2, TOP10, Top10F', Tuner, UT5600, W3110, XL1-Blue, XL2-Blue, XL2-Blue MRF', XL1-Red, XL10-Gold, and XL10-Gold KanR.

In a specific embodiment the host cell is *E. coli* JM109 and *E. coli* BL21(DE3).

25

Fungi and molds

In another embodiment the host cells are selected from non-bacterial origin such as fungi or molds. In a more specific embodiment the host cells are selected among *Aspergillus niger* and *Aspergillus oryzae*.

30

Many useful enzymes are produced using industrial fermentation of *A. niger*. For example, *A. niger* glucoamylase is used in the production of high fructose corn syrup, and pectinases are used in cider and wine clarification. α -galactosidase, an enzyme that breaks down certain complex sugars, is a component of Beano and

35 other medications which the manufacturers claim can decrease flatulence. Another

use for *A. niger* within the biotechnology industry is in the production of magnetic isotope-containing variants of biological macromolecules for NMR analysis.

A. niger is also cultured for the extraction of the enzymes glucose oxidase (GO) and Alpha-galactosidase (AGS). Glucose oxidase is used in the design of glucose biosensors, due to its high affinity for β -D-glucose. Alpha-galactosidase can be produced by *A. niger* fermentation; it is used to hydrolyze alpha 1-6 bonds found in melibiose, raffinose, and stachyose. Research published in 2006-2008 investigated *A. niger* prolyl endoprotease (AN-PEP), a microbial-derived prolyl endoprotease which cleaves gluten. This has strong implications in the treatment of Coeliac (Celiac) disease or other metabolic gluten sensitivity disease processes.

Aspergillus oryzae has strong secretion of amylases (α -Amylase and glucoamylase), some carboxypeptidases and tyrosinases.

Thus, by using the expressivity tag of the invention it may also be possible to increase production of enzymes in host cells of non-bacterial origin.

Antibody

The first polypeptide amino acid sequence disclosed in the invention may be used as a target for one or more antibodies. In this way the expressivity tag may also be used for both detection and purification of the hybrid polypeptide using standard molecular techniques, such as affinity purification, and immuno detection.

Therefore, in an embodiment the invention relates to an isolated antibody or antigen-binding fragment thereof which selectively binds to the first polypeptide amino acid sequence of the invention.

Similarly, the invention relates to an antibody, isolated antibody or antigen-binding fragment thereof according to the invention, wherein the antibody is a purification tag or an epitope tag.

Use of a recombinant hybrid nucleic acid sequence

In an aspect of the invention, the recombinant hybrid nucleic acid sequences of the invention are used to express recombinant hybrid polypeptides. This may be especially relevant where large amount of protein is required, e.g. for the
5 production of commercial enzymes such as restriction enzymes, therapeutic proteins, drugs, etc. The presence of the expressivity tag increases the yield of the polypeptide of interest during the fermentation process.

Thus, one aspect of the invention relates to the use of a recombinant hybrid
10 nucleic acid sequence encoding a hybrid polypeptide according to the invention for the expression of a recombinant polypeptide.

Similarly, the invention relates to the use of a recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide according to the invention, wherein said
15 recombinant hybrid nucleic acid sequence is comprised in a vector according to the invention adapted for expression of said recombinant polypeptide.

Method for expressing a recombinant polypeptide

Furthermore, in yet an aspect the invention relates to a method for expressing a
20 recombinant polypeptide comprising

- a) introducing into an expression system the hybrid nucleic acid sequence according to the invention encoding a recombinant polypeptide,
- b) expressing said nucleic acid sequence in the expression system, and optionally
- 25 c) purifying the expressed recombinant polypeptide.

The above disclosed method may be performed by a method, wherein a vector disclosed above is used. Thus, in an embodiment the invention relates to a method according to the invention, wherein the hybrid nucleic acid sequence is
30 comprised in a vector according to the invention is used.

Furthermore, in a further embodiment the invention relates to a method according to the invention, wherein the expression system is a host according to the invention.

In addition, the invention relates to a method according to the invention, wherein said expression system is an *in vitro* expression system.

When performing *in vitro* translation the nucleic acid sequences of the invention
5 are at some stage RNA. Therefore, it is to be understood that the nucleic acid sequences of the invention can also be RNA sequences.

When performing *in vitro* translation one or more components are normally obtained from a cell extract. Different bacterial strains can be used as the source,
10 but often the extract composition is obtained from *E. coli*.

Therefore, in another embodiment, the invention relates to a method according to the invention, wherein the *in vitro* translation system comprises extract from *E. coli*.

15

As means of purification or detection an antibody which binds to the recombinant polypeptide of the invention may be employed. Thus, in an embodiment, the invention relates to a method according to the invention, wherein an antibody according to claims the invention is used for purification of said recombinant
20 polypeptide.

Similarly, in an embodiment, the invention relates to a method according to the invention, wherein an antibody according to the invention is used for detecting said recombinant polypeptide.

25 As also mentioned earlier, cleavage at the peptide level may be useful if part of the hybrid amino acid sequence is not desired in the final peptide. For example a sequence encoding for specific proteases can be introduced between the first and second nucleic acid sequences. In this way the encoded first amino acid sequence can be removed from the final peptide. This removal can take place both before or
30 after a purification step. Examples of useful proteases are Factor Xa (fXa), TEV and PreScission Protease.

Thus, in yet an embodiment the invention relates to method according to the invention, wherein said first polypeptide amino acid sequence is subsequently
35 removed from the expressed target nucleic acid sequence by proteolytic cleavage.

Kit

A vector, as disclosed above, may be very useful in a kit, such as an *in vivo* translation kit, an *in vitro* translation kit, and a PCR-kit, a transcription kit.

5

Thus, in a further embodiment the invention relates to a kit comprising a vector according to the invention.

As means of further optimization of the expressivity tag of the present invention
10 directed evolution using phages may be used. An expressivity tag could be cloned
5' of an essential chromosomal gene of a phage. The gene should be vital for the
fitness of the phage. Optimal expressivity tags would be evolved and the fittest
phages could be harvested as they would be most abundant. Alternatively, phage
display could be used if a selection mechanism could be constructed favoring high
15 expression of the binding protein. Depending on similarity of sequences produced
a possible expressivity tag motif could hopefully be defined from such
experiments.

Thus, in a further embodiment the invention relates to a method for producing
20 and or optimizing expressivity tags, wherein

- 1) the nucleic acid sequences encoding an expressivity tag of the invention
is inserted upstream of operably linked to an essential gene in a phage
- 2) the phages are grown under a selection pressure
- 3) phages with high fitness is selected
- 25 optionally
- 4) steps 1) to 3) are repeated

A person skilled in the art of e.g. phage display would know the different
important parameters to adjust to provide a selection pressure.

30

It should be noted that embodiments and features described in the context of one
of the aspects of the present invention also apply to the other aspects of the
invention.

All patent and non-patent references cited in the present application, are hereby incorporated by reference in their entirety.

The invention will now be described in further details in the following non-limiting
5 examples.

Examples

The inventors fused *InfB*(1-471) 5' of two different reporter genes; *InfA* and *gfp*. Deletions were next made from the 3' end of *InfB*(1-471) and amounts of protein evaluated by SDS-PAGE, western immunoblotting and quantitative fluorescence
10 measurements. The deletions in *InfB*(1-471) were placed in regions without protein secondary structure. If the functional part of the expressivity tag was a protein structure, it would by this design not be compromised. Figure 1 shows the secondary structure elements of domain I of IF2 and the 3' *InfB*(1-471) deletions. Results showed that the expressivity tag region was located on *InfB*(1-21). Next,
15 3' deletions in *InfB*(1-21) was constructed, removing one codon at the time, and the expression analyzed. The results showed that the intact *InfB*(1-21) sequence contained the expressivity tag, as expression dropped when it was compromised. The expression of *InfB*(1-21) point mutants fused to *gfp* was next analyzed to define important positions in *InfB*(1-21). All the mutants produced compromised
20 the efficient expression of *InfB*(1-21). No position of specific importance was therefore defined. It could be shown that there was no correlation between expression and codon frequency or number of downstream A/U bases. The mRNA structure in the TIR and possible interactions between mRNA and the ribosome in initiation of translation could be of importance for *InfB*(1-21).

25

Materials

Bacterial strains

E. coli JM109 was from New England Biolabs and used for propagation of
30 plasmids.

Genotype: F' *traD36 proA⁺B⁺ lacI^q Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14⁻ gyrA96 recA1 relA1 endA1 thi hsdR17*

E. coli BL21(DE3) was from Novagen and used for protein expression combined with the pET vector system. The strain is deficient in the Lon protease and lacks

the OmpT outer membrane protease which improves the yield of intact recombinant protein. It contains a copy of the T7 polymerase inserted chromosomally with the DE3 lysogen.

Genotype: F^- , *ompT*, *hsdSB*($rB^- mB^-$), *gal*, *dcm*, (DE3)

5

E. coli BL21 was from Novagen and used for protein expression with the pBAD vector system. This strain is like BL21(DE3) also deficient in the Lon protease and lacks the OmpT outer membrane protease.

Genotype: F^- , *ompT*, *hsdSB*($rB^- mB^-$), *gal*, *dcm*

10

Plasmids

pET15b was from Novagen and used for both cloning and expression. It contains a *bla* gene coding for the β -lactamase enzyme giving resistance to ampicillin.

Transcription from its multiple cloning site is under control of the T7 promoter,

15 requiring the T7 polymerase.

pET24d was from Novagen and used for both cloning and expression. It contains a kanamycin resistance gene and transcription from its multiple cloning site is controlled by the T7 promoter.

20

pBAD was from Invitrogen and used for both cloning and expression. It contains an ampicillin resistance gene and transcription from its multiple cloning site is controlled by the araBAD promoter.

25 pLysS) was from Novagen and used to avoid leaky expression from pET15b and pET24d plasmids. pLysS has a weak constitutive expression of T7 lysozyme which binds to the T7 polymerase inhibiting its transcription activity.

pET15b[IF2-I] was provided by H.P. Sørensen and was used as a template to
30 create the *InfB*(1-471) fragments.

pGLO was from Bio-Rad and was used as a template to obtain *gfp*.

pET15b[IF2domI-IF1] was provided by T.B Kallehauge and was used as a
35 template to obtain *InfA*.

Relevant enzymes, antibodies and reagents

Restriction enzymes (*Bsp*HI, *Sac*I, *Bam*HI, *Eco*RI, *Nco*I, *sph*I) was supplied by New England Biolabs® with its relevant reaction buffer. T7 ligase and reaction buffer was from Pharmacia. Plasmids were purified by QIAGENs Qiaprep Spin
 5 Miniprep Kit. Restriction digested DNA fragments were purified by the Wizard® SV Gel and PCR clean-up system from Omega. Polyclonal rabbit anti-mouse immunoglobulin HRP conjugated antibody was from DAKO. Monoclonal mouse anti-*Escherichia coli* IF1 antibody no. 1 was produced in the laboratory of biodesign.

10 A complete list of buffer and solutions can be found in Appendix B.

Oligonucleotides and sequencing

All oligonucleotides and sequencings were made by DNA-technology A/S. A complete list of primers is to be found in Figures 25 and 26.

15

Cloning

Detailed information on PCR amplification, restriction digestion, ligation, transformation, DNA purification and sequence verification can be found in

20 Appendix A. A summary of the cloning procedure will be described.

The various *InfB*(1-471) 3' deletions fused to *InfA* were prepared using the primers listed in Figure 25 table I. Each PCR fragment produced is listed with its forward and reverse primer, and is in this cloning description referred to by its PCR fragment name. Figure 2 shows all the constructs cloned. The *InfB*(1-471) 3'
 25 deletion fragments and *InfA* are linked by a *Sac*I restriction site coding for Glu-Leu amino acids. The initiator ATG of *InfA* was removed in all constructs where it was fused to any fragment of *InfB*(1-471). When all eleven constructs are mentioned at the same time in the text they will be referred to as *InfB*(1-x)-*InfA*.

30 *InfA* was first constructed by one PCR fragment which was ligated into a pET15b vector. PCR *InfA* I was made with a pET15b[IF2-IF1] template which introduced restrictions sites *Bsp*HI and *Bam*HI and six consecutive histidines. The *Bsp*HI/*Bam*HI digested *InfA* I PCR fragment was ligated into an *Nco*I/*Bam*HI digested pET15b plasmid to make *InfA*.

InfB(1-471)-*InfA*, *InfB*(1-282)-*InfA*, *InfB*(1-177)-*InfA* and *InfB*(1-120)-*InfA* was next constructed by two PCR fragments, a *InfB* part and a *InfA* named *InfA* II, which was ligated into pET15b.

PCR fragments *InfB*(1-471), *InfB*(1-282), *InfB*(1-177) and *InfB*(1-120) were
 5 prepared with a pET15b[IF2-I] template. The primers introduced restriction *Bsp*HI and *Sac*I. An *InfA* II PCR was made with a pET15b[IF2-IF1] template and introduced restriction sites *Sac*I and *Bam*HI and six consecutive histidines. *Bsp*HI/*Sac*I digested *InfB*(1-471), *InfB*(1-282), *InfB*(1-177) and *InfB*(1-120) was ligated with *Sac*I/*Bam*HI digested *InfA* II into a *Nco*I/*Bam*HI digested pET15b
 10 plasmid to make *InfB*(1-471)-*InfA*, *InfB*(1-282)-*InfA*, *InfB*(1-177)-*InfA* and *InfB*(1-120)-*InfA*.

InfB(1-9)-*InfA*, *InfB*(1-12)-*InfA*, *InfB*(1-15)-*InfA*, *InfB*(1-18)-*InfA*, *InfB*(1-21)-*InfA* and *InfB*(1-90)-*InfA* were constructed each by one PCR fragment which was ligated into pET15b.

15 PCR fragments *InfB*(1-9), *InfB*(1-12), *InfB*(1-15), *InfB*(1-18), *InfB*(1-21) and *InfB*(1-90) were prepared with a pET15b[*InfB* (1-471)-*InfA*] template and introduced restriction sites *Sph*I and *Sac*I. *Sph*I/*Sac*I digested *InfB*(1-9), *InfB*(1-12), *InfB*(1-15), *InfB*(1-18), *InfB*(1-21) and *InfB*(1-90) were ligated with *Sac*I/*Bam*HI digested *InfA* II into a *Sph*I/*Bam*HI digested pET15b plasmid to make
 20 *InfB*(1-9)-*InfA*, *InfB*(1-12)-*InfA*, *InfB*(1-15)-*InfA*, *InfB*(1-18)-*InfA*, *InfB*(1-21)-*InfA* and *InfB*(1-90)-*InfA*.

The various 3' *InfB*(1-471) deletions fused to *gfp* were prepared using the primers listed in Figure 25 table II. Figure 3 shows all the constructs made. The initiator
 25 ATG of *gfp* was removed in all constructs where it is fused to any fragment of *InfB*(1-471). When all eleven constructs are mentioned at the same time in the text they will be referred to as *InfB*(1-x)-*gfp*.

gfp was first constructed by one PCR which was ligated into pBAD.

30 PCR *gfp* I was made with a template pGLO plasmid. This introduced restriction sites *Bsp*HI and *Eco*RI along with six consecutive histidines. The *Bsp*HI/*Eco*RI digested PCR fragment was ligated into an *Nco*I/*Eco*RI digested pBAD plasmid to make *gfp*.

Next *InfB*(1-471)-*gfp*, *InfB*(1-282)-*gfp*, *InfB*(1-177)-*gfp* and *InfB*(1-120)-*gfp* were constructed by two PCR fragments, a *InfB* part and a *gfp* part named *gfp* II, which was ligated into pBAD.

PCR fragments *InfB*(1-471), *InfB*(1-282), *InfB*(1-177), *InfB*(1-120) were prepared
 5 with a pET15b[IF2-I] template. The primers introduced restriction *Bsp*HI and *Sac*I. A PCR of *gfp* II was made with pGLO as template which introduced restriction sites *Sac*I and *Eco*RI and six consecutive histidines. *Sac*I/*Eco*RI digested *gfp* II was ligated with *Bsp*HI/*Sac*I digested *InfB*(1-471), *InfB*(1-282), *InfB*(1-177) or *InfB*(1-120) into a *Nco*I/*Eco*RI digested pBAD plasmid to make *InfB*(1-471)-*gfp*, *InfB*(1-282)-*gfp*, *InfB*(1-177)-*gfp* and *InfB*(1-120)-*gfp*.
 10

InfB(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp* were next constructed by one PCR fragment each which was ligated into pBAD. Primers contained a 5' non-annealing overhang of the *InfB* part along with an annealing *gfp* part. PCR fragments *InfB*(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*,
 15 *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp* were made using pGLO as template. It introduced restriction sites *Bsp*HI, *Eco*RI and six consecutive histidines. The five *Bsp*HI/*Eco*RI digested PCR fragments were ligated into *Nco*I/*Eco*RI digested pBAD plasmids to make *InfB*(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp*.

20 The *InfB*(1-90)-*gfp* construct was made by 3 PCR reactions. First PCR produced a fragment of *InfB*(1-90) with a *gfp* 3' overhang named *InfB*(1-90) I. Second PCR produced a fragment of *gfp* named *InfB*(1-90) II. Finally a PCR named *InfB*(1-90) III was made using the two PCR fragments as primers, as they annealed to each other. The *InfB*(1-90) III PCR product was ligated into pBAD. First PCR *InfB*(1-90) I was made with pET15b[IF2-I] as template which introduced restriction site *Bsp*HI and a 3' *gfp* overhang. Next PCR *InfB*(1-90) II was made with a pGLO template which introduced restriction site *Eco*RI and six consecutive histidines. A PCR using *InfB*(1-90) I and *InfB*(1-90) II as primers was made which produced a *InfB*(1-90)-*gfp* PCR fragment with *Bsp*HI and *Eco*RI restriction sites. The
 25 *Bsp*HI/*Eco*RI digested fragment was ligated into an *Nco*I/*Eco*RI digested pBAD plasmid.
 30

The procedure described for making each construct was followed in cloning them into the pET24d plasmid.

A point mutation analysis was made of *InfB*(1-21)-*gfp* changing the base pairs at downstream positions +7, +13 and +16 in the *InfB*(1-21) construct. This was done with the primers in Figure 26 table III. The primers were constructed with an *InfB* overhang and a *gfp* annealing part. The constructs were therefore made by one PCR each which was ligated into pBAD. Eight PCRs named *InfB*(1-21)-*gfp* 7A, 7T, 7C, 13T, 13G, 16T, 16G, 16C was made with a pGLO template. This introduced restriction sites *Bsp*HI and *Eco*RI along with six consecutive histidines. The *Bsp*HI/*Eco*RI digested PCR fragment was ligated into an *Nco*I/*Eco*RI digested pBAD plasmid to make the constructs *InfB*(1-21)-*gfp* 7A, 7T, 7C, 13T, 13G, 16T, 16G, 16C.

All ligation reactions were transformed into *E.coli* JM109 for further propagation. Clones were picked and grown separately for subsequent plasmid purification. A PCR screen was performed with either T7 screen primers or pBAD screen primers (Figure 26 table IV) for verification of the approximate size of an insert. Plasmids with correctly sized inserts were sequenced by DNA-technology A/S using the screening primers.

pET15b and pET24d constructs were transformed into BL21(DE3) pLysS cells while pBAD constructs were transformed into BL21 for subsequent expression.

Small-scale expression

BL21(DE3) pLysS harboring pET15b or pET24d constructs and BL21 harboring pBAD constructs were grown in 3 mL 2xTY media with appropriate antibiotics (conc. 100 µg/mL ampicillin, 34 µg/mL chloramphenicol, 50 µg/mL kanamycin) overnight. 500 µL overnight culture was centrifuged (15.000 x g, 1 min), cells resuspended in 150 µL 2xTY and inoculated into 50 mL 2xTY media with appropriate antibiotics. Cells were grown at 37° C until OD₆₀₀ reached 0.8.

BL21(DE3) pLysS pET15b and pET24d was induced with a final conc. of 0.8 mM IPTG and BL21 pBAD with 0.02 % L-Arabinose. Different growth temperatures were tested to optimize soluble protein expression. Cells were grown at 37°C and 30°C for 3.5 hours and at 20°C for 16 hours. A sample was taken for SDS-PAGE by the following procedure. 1 mL cell culture was centrifuged (15.000 x g, 4 min) and resuspended in OD₆₀₀·50µL disintegration buffer, boiled for 5 min and stored at -20°C. The rest of the cell culture was recovered by centrifugation (3.700 x g, 25 min, 4°C) and cell pellets stored at -20°C.

Western Immunoblotting

Expression of *InfB*(1-x)-*InfA* was analyzed by a western blot with anti-IF1 antibodies.

- 5 A sandwich western blot was constructed. On the lower plate (anode) the sandwich was built as following: Filter paper, nitrocellulose paper, SDS-PAGE run with total cell extracts from small-scale overexpression, filter paper. Filter papers were soaked in blotting buffer, nitrocellulose paper in distilled water. Blotting was carried out at 200 mA for 1 hour in a Kem Tec apparatus. After blotting the
- 10 nitrocellulose membrane was transferred to a rolling glass and washed with 50 mL TBS-T for 5 min. 15 mL primary mouse anti-*E.coli* IF1 monoclonal antibody no. 1 (1µg/mL in TBS-T) was added and incubated for 1 hour. The membrane was washed 3x5 min with 50 mL TBS-T. 10 mL secondary antibody (rabbit anti-mouse HRP-conjugated anti-immunoglobulin, 2.3 µg/mL in TBS-T) was added and
- 15 incubated 1 hour. The membrane was next washed 6x5 min with 50 mL TBS-T. A mixture of 4 mL solution 1 and 100 µL solution 2 of the ECL western blotting kit was incubated with the membrane for 5 min at constant shaking. The membrane was wrapped in Vita Wrap and exposed to an X-ray film for only a few seconds, which gave the clearest picture.

20

Quantitative fluorescence measurements

Amount of protein expressed from *InfB*(1-x)-*gfp* constructs was analyzed, by measuring the fluorescence from a fixed amount of cells.

- Frozen BL21(DE3) pLysS pET24d[*InfB*(1-x)-*gfp*] and BL21 pBAD[*InfB*(1-x)-*gfp*]
- 25 cell pellets, previously grown for small-scale expression, were thawed on ice and resuspended in resuspension buffer. The optical density at 600 nm was measured, and cell densities adjusted according to the optimum range of measurement, for the fluorescence reader. Fluorescence counts were later normalized according to cell density. Cell suspensions were pipette in a 96-well tray in 10 % dilution
- 30 series. The fluorescence was measured in a Synergy HTTR-I reader (Bio-tek), with an excitation wavelength of 400 nm (bandwidth 30) and an emission wavelength of 508 nm (bandwidth 20).

Solubility Test

For the GFP protein to be active, it needs to be correctly folded. The amount of GFP inclusion bodies produced was therefore checked. Frozen BL21(DE3) pLysS pET24d[*InfB*(1-x)-*gfp*] and BL21 pBAD[*InfB*(1-x)-*gfp*] cell pellets, previously
5 grown for small-scale expression, were thawed on ice and resuspended in lysis buffer. The optical density at 600 nm was measured and adjusted to the same numerical value for all cell suspensions. A volume of 3 mL cells were sonicated in a 13 mL Falcon tube on ice, for a total of 15 min (0.7 cycles, 100 amplitude). Prior
10 optimization of the sonication process, had shown that these conditions gave optimum cell disruption. The temperature of the sample and sonicator tip was checked regularly, to avoid overheating the sample. 750 μ L disrupted cells were centrifuged (15.000 x g, 15 min, 4°C), and pellets resuspended in 750 μ L water, boiled for 5 min and saved for SDS-PAGE analysis.

15 Large-scale expression

An odd SDS-PAGE migration pattern was observed for some of the constructs, and so protein was expressed, purified by chromatography and checked by mass spectrometry.

BL21(DE3) pLysS pET24d[*InfB*(1-9)-*gfp*], pET24d[*InfB*(1-12)-*gfp*],
20 pET24d[*InfB*(1-15)-*gfp*], pET24d[*InfB*(1-18)-*gfp*] and pET24d[*InfB*(1-21)-*gfp*] cells were grown in 50 mL 2xTY over night with antibiotics (34 μ g/mL chloramphenicol, 50 μ g/mL kanamycin). 16 mL culture was centrifuged (3.700 x g, 10 min, 4°C), resuspended in 1 mL 2xTY and inoculated into 1,6 L 2xTY with antibiotics. Cells were grown at 37° C, until the optical density at 600 nm reached
25 0.8, at which they were induced with 0.8 mM IPTG. Growth was continued 3.5 hours at 30° C. Cells were harvested by centrifugation (9.500 x g, 10 min, 4°C), resuspended in 0.9 % NaCl and centrifuged again (3.700 x g, 25 min, 4°C). Cell pellets were stored at -20° C.

30 Protein purification for mass spectrometry

Fusion proteins expressed from *InfB*(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp* and *InfB*(1-21)-*gfp* contained a C-terminal His-tag, and were purified by Immobilized Metal Affinity Chromatography (IMAC).

Cells previously grown for expression were resuspended in 2 mL buffer A per
35 gram cells. Cell suspensions were sonicated in a 50 mL Nunc tube on ice, for a

total of 15 min (0,5 cycle, 100 amplitude). Lysed cells were ultracentrifuged (184.000 x g, 75 min, 4° C) to sediment cell debris and insoluble protein. Soluble protein was 0,2 µm filtered and loaded onto a 8 mL chelating sepharose column, pre-loaded with Ni²⁺ and calibrated in Buffer A. Bound protein was eluted by a
5 stepwise 9-15-100 % gradient of Buffer C against Buffer A. Relevant fractions were analysis by SDS-PAGE. Identified fractions were pooled in a dialysis sack (MWCO: 6-8 kDa), and dialysed against 1 L of storage buffer over night at 4°C. Protein was stored at -20°C.

10 **Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)**

Fusion proteins expressed from *InfB*(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp* and *InfB*(1-21)-*gfp* constructs and purified as described were analyzed by MALDI-TOF, to confirm their masses. All five types of protein were dialysed against 2 L 50 mM Tris-HCl pH 7.5 overnight, 4°C. 0.5 µL of protein (20
15 pmol/µL) was mixed with 0.5 µL sinapinic acid (10mg/mL) on a MALDI plate. Samples were left to crystallize for 30 min. at room temperature. Mass spectra were acquired by a Voyager-DE PRO (MALDI-TOF) instrument (Applied Biosystems). Final spectra were obtained by averaging 150-200 single shot spectra. Spectra were externally calibrated with bovine insulin (5734.59 Da) and
20 *E.coli* thioredoxin (11674,48 Da).

N-terminal sequencing

Protein expressed from *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp* and *InfB*(1-90)-*gfp* was analyzed by N-terminal sequencing by Prof. Claus Oxvig. Protein was excised from
25 SDS-PAGE bands run with total cellular extracts from the small-scale expression.

mRNA folding by MFOLD

In order to investigate the base pairing in the TIR of mRNA sequences for *InfB*(1-x)-*InfA* and *InfB*(1-x)-*gfp* constructs, mRNA sequences were folded by the online
30 available program MFOLD.

The following standard settings were applied.

Nucleic acid type: RNA

Sequence type: Linear

35 Temperature: 37° C

Percent: 5

Na²⁺ conc: 1.0

Mg²⁺ conc: 1.0

Maximum bulge/interior loop size: 30

5 Maximum asymmetry of a bulge/interior loop: 30

Maximum number of foldings to be computed: 50

Codon frequency

Codon frequencies are obtained from

10 kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=37762

Example 1

Analysis of the expression of *InfB*(1-x)-*InfA*

Total cellular extracts of induced cells expressing *InfB*(1-x)-*InfA* (figure 2) were
15 analyzed by SDS-PAGE (figure 4).

It was evident, that *InfA* was expressed at lower levels than either of the *InfB*(1-x)-*InfA* constructs. It was not possible to see much difference between any of the *InfB* deletions, except *InfB*(1-177)-*InfA* which seemed to have a slightly higher expression. The molecular weight of the proteins expressed from the *InfB*(1-9)-
20 *InfA*, *InfB*(1-12)-*InfA*, *InfB*(1-15)-*InfA*, *InfB*(1-18)-*InfA* did not correspond to the molecular weight, of the surrounding proteins in the SDS-PAGE. All four proteins migrated slower than the protein expressed from *InfB*(1-21)-*InfA* and *InfB*(1-90)-*InfA*, which should not be the case. All constructs were verified by DNA
sequencing, and it thus seemed like an odd migration pattern that had to be
25 addressed. To analyze the integrity of the *InfA* moiety, and further evaluate amounts of protein expressed, a western blot with anti-IF1 antibodies was made of the total cellular extracts (figure 5).

IF1 protein was clearly recognized by the primary antibody in all constructs,
30 verifying the integrity of the *InfA* part. It was quite obvious from the western blot, that *InfA* expressed less than either of the fusion constructs. *InfB*(1-177)-*InfA* produced the most intensive band, as in the SDS-PAGE, but it was not much more intense than the band produced by *InfB*(1-90)-*InfA*. It seemed that a fragment as small as *InfB*(1-90) or possibly even *InfB*(1-21), could contain the expressivity tag

region. Reliably quantifying the amount of protein is not feasible from either western blot or SDS-PAGE. Proteins of high molecular weight stain more intensively than low molecular weight proteins in SDS-PAGE. Western blots also have some quantitative discrepancies, as smaller proteins transfers more easily from the SDS-PAGE to the nitrocellulose filter in the blotting procedure. Definitive conclusions were therefore troublesome, as proteins of differing molecular weights were compared. A fundamental problem realized with the *InfB*(1-x)-*InfA* constructs, was that the *SacI* restriction site (figure 2), which could interfere with results as the *InfA* reference did not have this site. Also when analyzing fragments as small as *InfB*(1-9), the *SacI* site could also have just as much influence on expression as the actual *InfB* part. As results indicated, that the area of interest could be a region on *InfB*(1-90), it was decided to leave out the *SacI* site for fragments of *InfB*(1-90) and smaller for *InfB*(1-x)-*gfp* constructs (figure 4).

15 Example 2

Analysis of expression of *InfB*(1-x)-*gfp*

Total cellular extracts of cells expressing pBAD[*InfB*(1-x)-*gfp*] and pET24d[*InfB*(1-x)-*gfp*] were analyzed by SDS-PAGE (figure 6).

The odd migration pattern observed in figure 4 was still evident in both figure 6a and 6b. The reason had to stem from the common *InfB* moiety, even though this made little sense. Protein samples were boiled in sample buffer containing SDS and β -mercaptoethanol. This left out the possibility of only partial denaturation or persisting di-sulfidebridges, which could interfere with migration. Increasing the concentration of SDS and β -mercaptoethanol in the sample buffer, did not change the migration pattern. Figure 6a showed gradually increasing amounts of protein expressed from *gfp* to *InfB*(1-21)-*gfp*. *InfB*(1-9)-*gfp* and *InfB*(1-12)-*gfp* were comparable to *gfp*, while *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp* and *InfB*(1-21) produced more protein. *InfB*(1-90)-*gfp* and larger constructs seemed comparable to *InfB*(1-21)-*gfp*, except *InfB*(1-471)-*gfp* which produced a slightly bigger band. This could however be attributed to more intensive staining, because of a higher molecular weight. The results in figure 6b showed a less pronounced enhancement of expression. *InfB*(1-9)-*gfp* expressed less than *gfp*. *InfB*(1-12)-*gfp* and *InfB*(1-15)-*gfp* expressed approximately the same amounts as *gfp*. *InfB*(1-18)-*gfp* and larger

fragments expressed more than *gfp*, even though this was not as pronounced as previously seen. The observation from the *InfA* experiments, that a fragment as small as *InfB*(1-90) or possibly *InfB*(1-21) contained the expressivity was confirmed by the results from *gfp*.

5 Example 3

The amount of protein expressed from *InfB*(1-x)-*gfp*, was measured by the fluorescence from a volume of induced cells having a fixed cell density (figure 7).

The results seen in figure 7a deviated from the corresponding results in the SDS-PAGE (figure 6a).

InfB(1-90)-*gfp*, *InfB*(1-120)-*gfp* and *InfB*(1-471)-*gfp* had 2-3 fold lower fluorescence counts than *InfB*(1-177)-*gfp* and *InfB*(1-282)-*gfp*, which did not match with the intensity of their corresponding protein bands in the SDS-PAGE. Results in figure 7b also showed a deviation from the corresponding results in the SDS-PAGE (figure 6b). *gfp* had the highest fluorescence count, while *InfB*(1-90)-*gfp*, *InfB*(1-120)-*gfp* and *InfB*(1-471)-*gfp* had very low counts as in figure 6a. Cells were grown at 37° C, a temperature unfavorable for high level soluble protein expression. It was therefore suspected, that a part of the total amount of protein seen in the SDS-PAGEs (figure 6), was expressed as inactive inclusion bodies. The amount of insoluble protein produced at different growth temperatures (20° C, 30° C, 37° C) was analyzed (figure 8 and 9).

From figure 8a it can be seen, that a fraction of the protein expressed from *InfB*(1-471)-*gfp*, *InfB*(1-120)-*gfp*, *InfB*(1-90)-*gfp* at 37° C, was insoluble protein. This explains the deviation between the fluorescence measurements (figure 7a) and SDS-PAGE (figure 6a). *InfB*(1-177)-*gfp*, *InfB*(1-21)-*gfp* and *InfB*(1-18)-*gfp* also expressed some degree of insoluble protein. At a growth temperature of 30° C (figure 8b) none of the constructs produced an observable insoluble fraction, and this condition was concluded to be optimal for measuring the fluorescence. In figure 9 the insoluble fractions of protein expressed from BL21(DE3) pLysS pET24d[*InfB*(1-x)-*gfp*] is seen.

In figure 9a it is seen how almost all constructs expressed at 37° C produced a considerable fraction of insoluble protein. This explains the deviation between the

fluorescence measurements (figure 7b) and the SDS-PAGE (figure 6b). At a temperature of 30° C (figure 9b), insoluble proteins were still evident for some of the constructs. Lanes 1-3 had more protein loaded, which can be explained by incomplete sonication of cells. It does however not change the conclusion, that this growth temperature could not be used. At 20° C (figure 9c) the amount of insoluble protein was minimized even though bands could still be seen for protein expressed from *InfB*(1-471)-*gfp*, *InfB*(1-282)-*gfp* and *InfB*(1-177)-*gfp* constructs. It was decided to measure the fluorescence, for the constructs expressed from the BL21(DE3) pLysS pET24d cells at 20° C, keeping in mind that *InfB*(1-471)-*gfp*, *InfB*(1-282)-*gfp* and *InfB*(1-177)-*gfp* would produce lower readings, because of low solubility. The difference in the amount of insoluble protein fractions produced for pBAD and pET24 was caused by a difference in transcriptional levels, and thereby overall protein expressed. The pET system produced more protein than pBAD, and therefore also had higher amounts of inclusion bodies.

BL21 pBAD[*InfB*(1-x)-*gfp*] and BL21(DE3) pLysS pET24d[*InfB*(1-x)-*gfp*] cells were grown at 30° C and 20° C and analyzed by SDS-PAGE (figure 10).

In both figure 10a and 10b it is seen that the overall expression of the constructs were lower than at 37° C (figure 6). The expression profile however, was the same and the fluorescence of the constructs measured (figure 11).

From figure 11a it is seen that the *InfB*(1-21) promoted the highest expression. As one codon is removed in *InfB*(1-18)-*gfp* to *InfB*(1-15)-*gfp*, the expression drops until *InfB*(1-12)-*gfp*, where it reaches the level of *gfp*. The fluorescence measurements correspond well with the SDS-PAGE result (figure 10a). The results in figure 11b showed that *InfB*(1-21) promoted the highest expression, correlating well the results in figure 11a. The expression drops a bit with *InfB*(1-18)-*gfp* while *InfB*(1-15)-*gfp* was at the level of *gfp*. The expression of *InfB*(1-15)-*gfp* deviated from the results in figure 11a, where it was expressed 2.25 fold compared to *gfp*. The expression of *InfB*(1-120)-*gfp* in figure 11b was lower than its surrounding constructs. If the SDS-PAGE in figure 10b is inspected, the *InfB*(1-120)-*gfp* band is less intensive than its surrounding bands, corresponding with the fluorescence results. In conclusion, a significant difference in expression was shown between *gfp* and *InfB*(1-21)-*gfp* of 3.5 and 5 fold in the two expression systems. It is

confirmed by two independent systems, that *InfB*(1-21) is the expressivity tag region in *InfB*(1-471).

Example 4

5 Protein analysis by N-terminal sequencing

In bacteria, protein synthesis initiates with a formyl-methionine. The formyl group is removed post-translationally by a peptide deformylase (PDF). The methionine residue is, depending on the identity of the second amino acid in the peptide chain, removed by a methionine aminopeptidase (MAP). (Spector S. et. al, 2003.

- 10 Expression of N-formylated protein in *Escherichia coli*. Protein expr purif. 32:317-322). The N-terminal part of protein from *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp* and *InfB*(1-90)-*gfp* was analyzed by N-terminal sequencing. Peptides and proteins with blocked N-terminal amino groups (e.g. acetylated, formylated) were not sequenced. Total cellular extracts was run on a SDS-PAGE and protein bands
- 15 excised. The extracted protein was analyzed and eight cycles of sequencing showed that the eight N-terminal amino acids of protein from *InfB*(1-18)-*gfp*, *InfB*(1-21)-*gfp* and *InfB*(1-90)-*gfp* was correct. The methionine group was removed to great extent but a fraction of protein still had the residue attached. *InfB*(1-18)-*gfp* (10.6 %), *InfB*(1-21)-*gfp* (6.8 %) and *InfB*(1-90)-*gfp* (10.4 %),
- 20 respectively. Differences were evident in the methionine removal; this could be of influence to the half-life of the proteins.

Example 5

25 Protein analysis by MALDI-TOF

- The odd SDS-PAGE migration pattern observed for *InfB*(1-x)-*InfA* (figure 4) and *InfB*(1-x)-*gfp* (figure 6), protein from five constructs were analyzed by MALDI-TOF. *InfB*(1-9)-*gfp*, *InfB*(1-12)-*gfp*, *InfB*(1-15)-*gfp*, *InfB*(1-18)-*gfp* and *InfB*(1-21)-*gfp* constructs were expressed by BL21(DE3) pLysS pET24d cells, and protein
- 30 purified by immobilized metal affinity chromatography (IMAC) with the aid of their C-terminal his-tags (figure 12).

All five proteins showed the same elution profile thus only one chromatogram is shown. Fractions of purified protein were very pure, with only a minor contaminant just below the 26 kDa marker band, seen on the SDS-PAGE especially in lane 4 and 5 (figure 12). It should also be noted, that the odd migration pattern persisted after the purification process. The mass of each protein were analyzed by MALDI-TOF (figure 13).

All five MALDI-TOF spectra gave a scheme of four major peaks. Three of these peaks corresponded to the analyzed protein with 1-3 charges (z), respectively. A fourth peak was seen on all spectra, which did not immediately have a mass (m) that corresponded to any of the analyzed proteins. It was noted that the m/z value for this peak, varied by approximately 100-120 Da between spectra, corresponding to the approximate mass of an amino acid. By closer inspections it was found that the peak matched an N-terminal fragment, of each construct broken at the fluorophore site of GFP. The fluorophore site is made up of an internal Ser-Tyr-Gly sequence, and the fragment matched a breakage between the Tyr-Gly residues. The C-terminal breakage fragment did not produce any clear peak on the spectra. It had a size of approximately 20 kDa, and is marked on all spectra even though only a minor peak was seen. The fragmentation relates to the nature of the MALDI-TOF method. The crystallized protein was excited by laser photons which fragmented the GFP protein, as a result of the high intensity. Figure 14 summarizes the values obtained from all spectra. The theoretical mass is calculated from the assumption that the N-formyl-methionine was cleaved off. It can be seen, that the measured mass for each construct was close to its theoretical value. Any deviations between the measured data and theoretical values could relate to the fact, that calibration was done externally with only two proteins, Insulin (5734,6 Da) and Thioredoxin (11674,5 Da) which both had lower masses than the analyzed proteins. Deviations between measured and theoretical values are however acceptable. The odd migration pattern seen in the SDS-PAGE could not be attributed to a false nature of any of the analyzed proteins.

Each peak on the spectra had "shoulder" peaks. The shoulders were two different versions of the analyzed protein. One version contained the N-terminal methionine and the other contained the N-formyl-methionine. The degree of

deformylation was quite similar for all constructs analyzed. There was a difference in the efficiency with which the methionine was cleaved off between constructs. For protein of *InfB*(1-21)-*gfp* it was efficiently removed while for *InfB*(1-9)-*gfp* it was not. The last three constructs had intermediate removal of the methionine.

- 5 Figure.15 illustrates the difference between protein from *InfB*(1-21)-*gfp* and *InfB*(1-9)-*gfp* constructs.

The fact that some of the constructs had their N-terminal methionine more efficiently cleaved off could contribute to the odd migration pattern seen in the
10 SDS-PAGEs, however in no way explain it. The migrational difference of the bands were of multiple kDa while the mass of a methionine is 149,2 Da. The MALDI-TOF results corresponded well with the N-terminal sequencing. Protein expressed from *InfB*(1-18)-*gfp* had a higher fraction of N-terminal methionine residues than *InfB*(1-21)-*gfp*. It was noted that this could be of influence for the half-life of the
15 proteins.

Example 6

Structure of mRNA folded by MFOLD

In order to try and explain the optimal expressivity tag properties of *InfB*(1-21),
20 folding schemes of mRNA sequences were analyzed. Any difference in structure in the TIR was noted. A loose structure in the TIR could be connected to a higher expression than a highly base paired TIR. In order to be able to compare the free energy between folded sequences, they had to be of the same length. It was therefore decided to minimize transcripts to nucleotides from the 5' end of the
25 mRNA, and 50 nucleotides into the coding sequence of each construct. Due to these settings the constructs *InfB*(1-90)-*InfA*, *InfB*(1-120)-*InfA*, *InfB*(1-177)-*InfA*, *InfB*(1-282)-*InfA*, *InfB*(1-471)-*InfA* were not distinguished.

The 5'-end of the pET15b vector transcript (87 nt) and 50 nts into each *InfB*(1-471)-*InfA* construct were folded by the MFOLD program (figure 16).

- 30 An array of folding schemes was obtained for each construct, and the ones with the lowest energy analyzed. A problem with comparing the schematic results was that a common scaffold was not applied, and similar sequences therefore ended up having quite different overall structure. This disturbed the interpretation of base pairs formed by the SD-sequence and initiation start codon. Due to parts of

each sequence being removed, the results were only indicative. It was evident from figure 16, that no major difference in free energy was seen between constructs. If the structure of the mRNA should be attributed to the property of *InfB*(1-21), it was expected that *InfA* had the most base paired structure and *InfB*(1-21) and *InfB*(50) the least base paired. This was not the case, as *InfB*(1-50) and *InfB*(1-15)-*InfA* actually had a more base paired structure, looking at the free energy, than *InfA*. *InfB*(1-9)-*InfA*, *InfB*(1-12)-*InfA* and *InfB*(1-18)-*InfA* furthermore had comparable free energies of only a marginal higher value than *InfA*. *InfB*(1-21)-*InfA* had the highest free energy. The foldings did not correspond well with the amount of each protein produced (figure 4 and 5). Interpreting how base paired the SD-sequence and initiation start codon was in each construct, was not found feasible.

The *InfB*(1-x)-*gfp* constructs were also analyzed by the MFOLD program (figure 17).

From these folding schemes a tendency was evident. *gfp* had the lowest free energy while *InfB*(1-21)-*gfp* had the highest. This corresponded well with the relative amount of protein observed by the fluorescent measurements (figure 11). The rest of the constructs had intermediate free energies. If the folding schemes was to match the fluorescence results, *InfB*(1-9)-*gfp* and *InfB*(1-12)-*gfp* should have comparable free energies to *gfp*, which was not seen. Nevertheless it seemed that some degree of correspondence was seen, between structure in the TIR and the amount of protein expressed. It could in no way be ruled out, that a low level of structure in the TIR caused the property of the expressivity tag.

25

Example 7

Analysis of the expression of point mutants of *InfB*(1-21)-*gfp*

An intensive literature survey was conducted, in order see if any sequences were reported having similar properties of *InfB*(1-21). Only one sequence of a comparable length was found, being the first 18 nucleotides of the *E. coli Trx* gene. (Peti W. and Page R. 2007). expression in *Escherichia coli* with minimal cost. Protein Expr Purif. 51:1-10). *Trx* encodes the Thioredoxin Reductase protein and its sequence is seen in figure 18.

To compare the effect of expression between the *infB*(1-21)-tag and the Thiredoxin-tag disclosed in figure 18, the expressional levels of GFP positioned in frame behind each of the tags were measured (figure 27). With a relative expressional level of on 27% of Thioredoxin-GFP compared to *infB*(1-21)-GFP, the tags of the invention can be considered better expressional tags than thioredoxin. The expressional levels were measured using fluorescence.

Three codons out of six and twelve out of eighteen bases were the same.

In order to check these possibly conserved positions were of particular importance to the expressivity of *InfB*(1-21), point mutations were made at position G7, A13 and A16. The mutants were expressed in BL21 pBAD and the fluorescence measured and compared to the wildtype *InfB*(1-21)-*gfp* construct. Fluorescence count is plotted in figure 19 together with the frequency of the specific codon which the point mutation caused, to see if codon frequency could be correlated with expression.

It could be seen that all mutations to some extent compromised the optimal expressivity of the wildtype *InfB*(1-21).

The fact that point mutations at position 7 (figure 19a) had an influence on expression was interesting. The fluorescence counts in figure 11 showed, that *InfB*(1-9)-*gfp* expressed a little less than *gfp*. I might thus seem like a possible discrepancy between results, that position 7 in figure 19 influenced the expression. This was however not necessarily the case because of the complexity of the system. As deletions were from the 3'-end, the importance of the 5'-end had not been evaluated. The fact that *InfB*(1-9) did not have an expressivity tag property on its own, did not mean that the first 9 nucleotides of *InfB*(1-21) had no influence on its property. The results in figure 19 thus indicated, that it was the entire *InfB*(1-21) sequence, and not only the actual positions of bases 15-21 (showing a high expression in figure 11) that was the expressivity tag region. Quite a drastic drop in expression was seen in figure 19c for the *InfB*(1-21) 16G mutation, changing the codon from ACG (Thr) to GCG (Ala). The expressivity of *InfB*(1-21) could, according to this result be abolished by a single base pair substitution. It was noted, that many mutations caused a drop in expression to 0.4-0.6, compared to the wildtype. This drop was not below the expression of *gfp* and so some effect of the expressivity tag was still intact. No correlation between expression and codon frequency was found. For instance in figure 19b, the G mutation made a codon frequency of above 2, but still produced lower

fluorescence counts than a mutation to T, which only had a codon frequency of 1. Contrary in figure 19c, the wildtype A both had a higher fluorescence count and codon frequency than the mutant T.

The three positions analyzed all had an importance for the property of *InfB*(1-21).

5 Depending on which wildtype base was manipulated and which base was inserted, the expression could decrease to levels below *gfp*, while most of them caused a intermediate drop to 0.4-0.6.

A comprehensive point mutation analysis of every position in *InfB*(1-21) from position +7 to +21 was initiated, in order to understand the importance of each
 10 individual base position. In addition a study of positions +4 to +6 was initiated. The result of this study is shown in figure 28. Figure 28 shows that a number of mutations result in expression levels comparable to *InfB*(1-21) (seq ID NO: 2). These mutations are position 5C→5A, 7G→7A, 7G→7T, 7G→7C, 8A→8C, 8A→8G, 8A→8T, 9T→9C, 12A→12T, 13A→13T, 14C→14A, 15G→15A, 17T→17C, 17T→17G,
 15 17T→17A, 20A→20C, 20A→20G, 20A→20T, 21A→21G, and 21A→21T.

It is to be understood that the above list of mutations are also applicable to other versions of the expressivity tags of the invention (such as SEQ ID: NO 1, 3, 4 and sub-sequences of these sequences), wherein the specific positions are present. It
 20 is also to be understood that the different mutations may be combined to create expressivity tags with multiple mutations compared to *InfB*(1-21) (and the other versions such as e.g. SEQ ID: NO 1, 3, 4 and sub-sequences of these sequences).

Similar, it is to be understood that the peptides corresponding to the above list of
 25 mutation also form part of the present invention.

Thus, in an embodiment of the invention, the first nucleic acid sequence of the invention is mutated in at least one the following positions, as indicated, selected from the group consisting of positions 5C→5A (SEQ ID: NO 68), 7G→7A (SEQ ID: NO
 30 69), 7G→7C (SEQ ID: NO 70), 7G→7T (SEQ ID: NO 71), 8A→8C (SEQ ID: NO 72), 8A→8T (SEQ ID: NO 73), 8A→8G (SEQ ID: NO 74), 9T→9C (SEQ ID: NO 75), 12A→12T (SEQ ID: NO 76), 13A→13T (SEQ ID: NO 77), 14C→14A (SEQ ID: NO 78), 15G→15A (SEQ ID: NO 79), 17T→17C (SEQ ID: NO 80), 17T→17A (SEQ ID: NO 81), 17T→17G (SEQ ID: NO 82), 20A→20C (SEQ ID: NO 83), 20A→20G (SEQ ID: NO 84), 20A→20T
 35 (SEQ ID: NO 85), 21A→21G (SEQ ID: NO 86), and 21A→21T (SEQ ID: NO 87).

It is to be understood that the specific mutations present in SEQ ID: NO 68-87 are also applicable to SEQ ID: NO 1, 3, 4 and sub-sequences of SEQ ID: NO 1, 3, 4. Of course this requires that the specific position is present in said sequences.

5

Sequence analysis

The numbers of A/U residues and CA repeat bases have been reported to be of importance for the efficiency of translation, and are shown for sequences of *InfB*, *InfA* and *gfp* in figure 20.

- 10 *InfB*(1-21) contained one more A/U residue than *InfA*(1-21) and *gfp*(1-21). This seems like a small difference. The point mutants of *InfB*(1-21) 7A and 7U increased the amount of A/U bases but decreased expression (figure 19a). There were no repeats of CA bases, but one and two in the entire sequence of *InfB*(1-21) and *InfA*(1-21), respectively. No correlation was therefore found between the
- 15 efficiency of expression of *InfB*(1-21) and A/U or CA bases.

Example 8

Downstream interactions between mRNA and 70S ribosome

- In an attempt to explain the properties of *InfB*(1-21), interactions between the
- 20 mRNA and ribosome in initiation of translation was analyzed. Interactions were visualized by crystal structures and the PyMol program. In figure 21, mRNA is seen with the fMet-tRNA bound to the AUG initiation codon. A sphere of rRNA base residues, spaced within a 12 Ångström radius around position +4 to +12 position on the mRNA is shown, to visualize the downstream tunnel.

25

- Approximately 15 nucleotides of the mRNA are placed in the downstream tunnel. Position +11 to +15 of the mRNA is surrounded by a layer of protein consisting of S3, S4 and S5. A dense array of basic Arg amino acids is projected into the tunnel from these proteins. This is thought to interact with backbone phosphates of the
- 30 mRNA. From position +7 to +10 the mRNA passes through a layer of rRNA. Cytosine 1397 and Uracil 1196 from this layer is oriented towards the mRNA around position +7 and +9 on the mRNA. It is though that these bases might help to position the mRNA immediately downstream of the A codon position. Base C1397 is shown in figure 22 in interaction with position +7 of the mRNA. The base

U1196 is spaced from the mRNA in an intermediate position between +8 and +9. It is placed on the internal side of the downstream tunnel but not oriented directly towards the +9 mRNA base

When the coordination of C1397 of is linked to the sequence of *InfB*(1-21) an
5 interesting notion can be made. At position +7 *InfB* contains a G. The possible interactions of *InfB*(1-21), *InfA*(1-21) and *gfp*(1-21) to C1393 is shown for each codon frame in figure 23.

The +1 codon frame is where the AUG codon is in the P-site. It can be seen that
10 in this frame *InfB*(1-21) contains a G at the first codon position, which neither *InfA* or *gfp* does. In the next codon frames, *InfA* and *gfp* contains more G's at the first codon position than *InfB*. It could be speculated, that a G at position +7 in the mRNA is of particular importance for the coordination of the transcript in the tunnel. As the first and second peptide bond is formed, it is no longer of such
15 importance. This is however only speculations, trying to fit the expressivity of *InfB*(1-21) to the structural arrangement. The analysis does also not explain why the entire sequence of *InfB*(1-21), is of importance to its expressivity.

Discussion

20 The main aim of this project was to determine a expressivity tag region on the 471 nucleotide transcript of *InfB*(1-471) and an expressivity tag motif should be defined which was general for the *Escherichia coli* organism.

Odd migration pattern

25 At first, the puzzle of the odd migration pattern observed in the SDS-PAGEs in figures 4 and 6 needs to be addressed. A lot of the results confirmed the integrity of the constructs produced. All constructs were sequenced correctly; an anti-IF1 monoclonal antibody clearly recognized IF1 protein expressed from *InfB*(1-x)-*InfA* constructs (figure 5). Fluorescence was measured from protein expressed from
30 *InfB*(1-x)-*gfp* constructs (figure 11), and MALDI-TOF measurements of a selection of proteins involved in the odd migration pattern verified their mass within a very small margin of error (figure 13). The idea that an error in the primary sequence was the cause of the migration pattern was ruled out. The possibility that some of the constructs were modified post-translationally was considered. The number of
35 post-translational modifications seen in *E.coli* is however limited. Any such

modification of a covalent nature should also have been observed in the MALDI-TOF spectra, which it was not. A possible cleavage by a protease would also have been seen in the MALDI-TOF spectra. The possibility that another protein, could bind to some of the constructs, only after heat denaturation which was done
5 before SDS-PAGE analysis, was considered. The presence of the anionic surfactant SDS should however disrupt any non-covalent interactions, and thereby exclude this possibility. Increasing the concentration of SDS and β -mercaptoethanol in the sample buffer did not change the migration pattern. No logical explanation could be reasoned for the odd migration pattern, and it can therefore for the moment
10 only be characterized as an SDS-PAGE artifact.

Expressivity tag region on *InfB*(1-471)

At lot of work was put into analyzing and proving the authenticity of the expression promoted by the *InfB*(1-471) 3' deletions. Three different vector
15 systems were made, and two different reporter genes analyzed in order to thoroughly support all results produced.

The analysis of the *InfB*(1-x)-*InfA* constructs revealed, that every 3' deletion produced could increase the expression of *InfA*. Differences in expression were
20 seen as *InfB*(1-471) was shortened. According to SDS-PAGE (figure 4) and especially western blot (figure 5), the shortest fragment containing the region of interest was *InfB*(1-90). The *InfB*(1-21) fragment induced an expression almost as high as *InfB*(1-90), but it could not be definitively concluded if the region could be narrowed down to this small fragment, because of the limited sensitivity of
25 SDS-PAGE and western blotting. It was interesting to see that *InfA*, a homologous gene, could have its expressivity enhanced by a part of another *E.coli* gene *InfB*.

Since GFP could be analyzed directly from cellular suspensions by fluorescence emission, protein amounts could be quantified in a more sensitive way. The
30 results from the pBAD and pET24d expression system both showed, that *InfB*(1-21) indeed was the minimum fragment of the ones produced, inducing optimal expressivity. Few differences between the expression from pBAD and pET24d were seen. A drop in expression was seen for constructs smaller than *InfB*(1-21)-*gfp* in both expression systems (figure 11). For the constructs expressed from
35 pBAD, it was a linear drop per codon removed. In the pET24d system, a steep

drop was seen for *InfB*(1-15)-*gfp* which only expressed at the level of *gfp*. This was a discrepancy between results not readily explained. It can therefore not from the results in figure 11 be concluded, if codon no. 5 is of higher importance as indicated by the pET24d system, or if every codon is equally important as indicated by the pBAD results. A comparison with the *InfB*(1-x)-*InfA* constructs is dubious, as this system contained the interfering *SacI* restriction site. Some of the construct with a *InfB* fragment longer than *InfB*(1-21) expressed lower amounts of protein than *InfB*(1-21)-*gfp*, which might seem puzzling as they contain the expressivity tag region of *InfB*(1-21). When the deviation on the dataset is taken into account, *InfB*(1-471)-*gfp* and *InfB*(1-120)-*gfp* for pET24d and *InfB*(1-90)-*gfp*, *InfB*(1-120)-*gfp*, *InfB*(1-177)-*gfp* and *InfB*(1-471)-*gfp* for pBAD expressed less than *InfB*(1-21)-*gfp*. The lower *InfB*(1-471)-*gfp* expression from pET24d can be explained by the fraction of insoluble protein produced at 20° C (figure 9c). Also, when the SDS-PAGE at figure 10 is inspected, it is seen that *InfB*(1-120)-*gfp* expresses less than its surrounding constructs indicating that the fluorescence measurements are reliable. The lower expressivity seen for the constructs from pBAD, could be a result of the crudeness of the fluorescence measurements, however also easily attributed to the complexity of the system as a whole. As reviewed in the introduction, small changes in the coding sequence of a gene can have a huge impact on its expression, e.g. because of a change in mRNA stability, structure etc. Observations might be attributed to this complexity, even though this is only speculations. It should be noted, that samples have been measured more than once and no difference in the expression has been seen, outside the standard deviation reported on the dataset in figure 11. Testing the expressivity of a *InfB* fragment of an intermediate length between *InfB*(1-21) and *InfB*(1-90) might be a very relevant case, as the expressivity for the 3' deletions were not constant until the point of the smallest fragment with maximum expressivity; *InfB*(1-21). A fragment inducing an higher expressivity than *InfB*(1-21) could possibly be found when including an extra region. In conclusion, from the results presented it was shown that *InfB*(1-21) was the expressivity tag region on *InfB*(1-471). Furthermore, this region was shown to be able to enhance the expressivity of a heterologous gene, *gfp*, along with a homologous gene, *InfA*.

Explanation for *InfB*(1-21) properties

Due to the nature of the vector systems, cellular regulation of expression at the DNA level e.g. involving promoters of different strength was eliminated. Reasons for the efficient expressivity of *InfB*(1-21) was to be found in some of the factors presented in the introduction. The fact that the region of interest had been narrowed down to quite a small area on the transcript, eliminated some possibilities.

Neither *InfA* nor *gfp* contained a NGG codon at any of the position +2 to +5, which could cause abnormally low amounts of protein (Gonzales de Valdivia E. et. al, 2004. A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in Escherichia coli. Nucleic acids res. 32:17:5198-5205). Also, none of the point mutations caused a NGG codon, so falsely low protein level can be excluded for any of the constructs.

The 2nd codon position of a gene has been shown to have a great impact on protein synthesis. From the results in figure 11, it could be excluded that the 2nd codon position of *InfB*(1-21) was of particular importance for its expressivity as *InfB*(1-9) and *InfB*(1-12) did not increase expression of *gfp*. It could however be speculated, that the second codon position of *InfB* caused an increased of *InfA*.

This would be a result of *InfA* having a poorly expressed second codon. It is speculated because *InfB*(1-9) and *InfB*(1-12) had an enhancing effect on *InfA* as seen in figure 5. An evaluation of the efficiency of the second codon according to the literature is noted in figure 24. The effect of the second codon on the expression of the *lacZ* gene in *E.coli* is shown for two different studies (Looman A.C. et. al and Stenström C.M. et al.). In both studies the expression is shown relative to the codon which mediated the highest expression in the studies, AAA (Lys), which is set to 1. As the expression is related to a fixed maximum in both studies, the relation between codons in one study should be comparable to the relation of second study.

30

There is a deviation between the two reports. This is explained by Looman et al. using an expression system with a canonical SD sequence while Stenström et al. uses a system without. The study by Looman et al. is most comparable to our system as it also contained a SD sequence and will therefore be considered.

The second codon of *InfB* could not have a positive effect on expression relative to *InfA* or *gfp* as seen in figure 20. It should actually decrease their expression. This correlated well with the results for *gfp* where e.g. *InfB*(1-9)-*gfp* expressed less than *gfp*. The slight increase in expression of *InfB*(1-9)-*InfA* relative to *InfA* could
5 not be attributed to the second codon.

In conclusion, the 2nd codon position of *InfB*(1-21) is not a determining factor for the expressivity tag.

The structure of mRNA for all constructs produced was investigated in two folding
10 schemes figures 16 and 17. For *InfB*(1-x)-*InfA* no convincing correlation could be found between expression and the structure of mRNA. It should however be noted, that *InfB*(1-21)-*InfA* had the highest free energy of the constructs, which should promote a higher expression. For *InfB*(1-x)-*gfp* a noteworthy correlation was seen, and again *InfB*(1-21)-*gfp* had the highest free energy. The free energy
15 for *InfB*(1-9)-*gfp* and *InfB*(1-12)-*gfp* did not match their expression. If only *InfB*(1-21)-*InfA/gfp* were compared to *InfA/gfp*, a nice correlation between expression and free energy would be seen, however as it did not match for the other constructs, the results are questionable.

It is difficult to draw definitive conclusions from MFOLD folding, because of
20 questionable accuracies in its predictions. An evaluation of the 3.1 version of MFOLD, found an average prediction accuracy of 41 % for the 16S and 23S rRNA compared to crystal structures. Smaller RNA structures like 5S rRNA and tRNA, both of which are more comparable in length to the sequences folded in this thesis, all had prediction accuracies greater than 60 %.

25 Noting the inaccuracy of MFOLD, it seems that *InfB*(1-21) could possibly promote a loose structure in the TIR, which could be of importance for its efficient expressivity.

The correlation between codon frequency and expression was analyzed by point
30 mutations in *InfB*(1-21) (figure 19). It was speculated that a favorable codon composition in the immediate downstream region provided by *InfB*(1-21), could mediate a more efficient translation by eg. minimizing peptidyl tRNA drop-off. No convincing relation could however be seen between the expression of *InfB*(1-21) point mutants and codon frequency. The fact that codon frequency is not the
35 determining factor is not a unique, as the second codon of a gene is not dictated

by codon frequency. Additional information was gained from the point mutants. It was shown in figure 11 that *InfB*(1-9) and *InfB*(1-12) did not mediate an increased expressivity. It could therefore be speculated, that it was only the region *InfB*(15-21) that was the region of interest. It was in this regard

5 interesting to note, that the positions +7 and +13 both was of influence to *InfB*(1-21). These results indicated, that it was the entire 21 nucleotide sequence and not only the actual positions of nucleotides *InfB*(15-21) which were important to its expressivity. In conclusion, codon frequency is not a determining factor for the efficient expressivity of *InfB*(1-21).

10

When the ribosome is anchored at the Shine-Dalgarno sequence, it physically covers approximately 15 nt downstream of the initiation start codon. It is quite interesting how *InfB*(1-21) is close to this region also in length. Intuitively, this leads one to think of a mechanism explaining this coincidence. It could be

15 speculated that the sequence of *InfB*(1-21) has motifs favorably interacting with one or two of the ribosomal subunits. A perfect match in the downstream tunnel involving multiple favorable interactions without steric clashes could promote an efficient initial expression. This could be of special importance in the initiation of translation, as a peptide chain has not made its way into the ribosomal exit tunnel
20 stabilizing the complex. It could be speculated that once the peptide chain reaches a critical length, the amount of peptidyl tRNA drop-off is minimized. The fact that NGG codons at certain positions are known to cause peptidyl tRNA drop off, could be an example of the contrary to *InfB*(1-21), NGG codons causing unfavorable interactions. Since they cause peptidyl-tRNA drop off if from codon position +2 to
25 +5 it seems reasonable that the A-site is of interest. If NGG is placed e.g. position +5, after four peptide bonds are formed it is in the A-site and could cause drop off. More work should be put into analysis of the structural movements in the decoding process between ribosome, mRNA and tRNA.

30 An interpretation of the interactions between mRNA and the downstream tunnel is complex. The crystal structure analyzed (figures 21 and 22) has captured the complex in one state, while it *in vivo* is moving, constantly breaking and forming bonds. It is not known how much the mRNA can wiggle inside the tunnel which is of importance for the dynamics of its interactions. It was e.g. noted from two
35 different crystal structures (pdb: 2HGR, 2HGP) that when the complex went from

initiation to elongation, the residues spaced within a fixed Ångström radius of the mRNA changed. Another problem was how it was not evident which interactions were acting favorably and unfavorably. It is e.g. likely, that a too strong interaction between mRNA and the ribosome could slow down translation. No clear motif in the downstream tunnel was found, which on some level was expected. The ribosome should be able to translate a broad combination of transcripts, and therefore primarily interacts with the mRNA backbone in the downstream tunnel. From crystal structures it can be seen that C1397 could coordinate G7 in *InfB*(1-21) in the first codon frame position (figure 23). It could be of importance in positioning the mRNA as a supplement to the SD sequence. The coordination made by C1397 is not a new report. There is a statistical preference for guanosine at first codon position in highly expressed *E.coli* genes. It has been suggested, that a GNN pattern in these genes may be responsible for monitoring the correct reading frame during translation by complementary to C periodical sites in the *E.coli* 16 S rRNA. The point mutants for position +7 of *InfB*(1-21) (figure 8a) did not show that this residues was of extra importance. The expression of the two dropped to 0.5-0.6 compared to the wildtype which was a general pattern for the mutants.

The structural explanation still provides a theory why position +7 is of importance to *InfB*(1-21).

To conclude, the structural arrangement of *InfB*(1-21) mRNA on the ribosome was analyzed and favorable interactions between the two in the downstream tunnel could be of importance to the efficient expressivity of *InfB*(1-21). Analysis of a wider repertoire of structures would be preferred before conclusions are drawn for this subject.

Appendix A

PCR amplification

PCR reactions were made in 100 µL aliquots by the following buffer concentrations and reaction conditions. Annealing temperatures for each PCR reactions along with primers can be found in figures 25 and 26.

0.75 µL Template (50-100 ng/µL)

- 0.6 μ M Forward primer
- 0.6 μ M Reverse primer
- 3 μ L Pfu polymerase (conc. N.D.)
- 0.2 mM dNTP
- 5 2 mM MgCl₂
- 75 mM Tris-HCl (pH 9.0)
- 20 mM (NH₄)₂SO₄
- 0.01% Tween
- 10 30 cycles
- 94° C – 1 min
- See figures 25-26 for annealing temperatures – 1 min.
- 72° C – 1-2 min
- 1 cycle
- 15 72° C – 5 min

Plasmid purification

- Plasmids were purified by QIAGENs Qiaprep Spin Miniprep Kit by the supplied protocol. 3 mL of bacterial cell culture grown over night at 37° with appropriate
- 20 antibiotics was used per spin column and DNA was eluted with 60 μ L H₂O per spin column.

DNA fragment purification

- PCR fragments and digested plasmids were run on a 1 % agarose gel using
- 25 marker lanes to avoid UV-exposure of the DNA. Bands were excised and the DNA extracted from the gel by the Wizard® SV Gel and PCR clean-up system according to the supplied protocol. Digested PCR fragments were directly purified by the Wizard® SV Gel and PCR clean-up system.

30 *Digestion of DNA*

- Digestion reactions were performed by protocols from New England Biolabs using relevant restriction enzymes and buffers. A standard 75 μ L scheme was used as shown below. Restriction reactions were incubated 90 minutes at 37°C. If double digest reactions were not possible single digests were performed with an
- 35 intermediate DNA purification step as already described.

- 55 μ L purified plasmid DNA/PCR fragment (100-200 ng/ μ L)
- 6.6 μ L NEB buffer (10x)
- 5 2 μ L Enzyme I (10U/ μ L)
- 2 μ L Enzyme II (20U/ μ L)
- 1 μ L BSA (10mg/mL)

Ligation

- 10 Digested inserts and plasmids were ligated by one of two schemes below depending on whether one or two inserts were needed. Ligation reactions were incubated at room temperature for 45 minutes.

- 10 μ L Plasmid (200 ng/ μ L)
- 15 7 μ L Insert (400 ng/ μ L)
- 2 μ L T7 ligase buffer (10x)
- 1 μ L T7 ligase (400.000 U/ μ L)

- 10 μ L Plasmid (200 ng/ μ L)
- 20 6 μ L Insert I (400 ng/ μ L)
- 6 μ L Insert II (400 ng/ μ L)
- 2.4 μ L T7 ligase buffer (10x)
- 1 μ L T7 ligase (400.000 U/ μ L)

25 *Transformation*

- JM109, BL21 or BL21(DE3) pLysS cells were first made competent for transformation. Cells were grown in 3 mL 2xTY media overnight at 37° C. 600 μ L overnight culture was collected by centrifugation (15.000 x g, 1 min) and redissolved in 100 μ L fresh 2xTY. The cell suspension was inoculated into 60 mL fresh 2xTY media and grown at 37°C until OD600 reached 0,5. Cells were collected by centrifugation (3.700 x g, 10 min, 4°C), resuspended in 10 mL TF1 solution and incubated on ice for 30 min. Cells were again collected by centrifugation (3.700 x g, 10 min, 4°C) and resuspended in 1 mL TF2 solution. Competent cells not used immediately were frozen in liquid nitrogen and stored at
- 35 -80° C.

Transformation reactions were carried out as follows. Ligation reaction of 20-24 µL or 5 µL purified plasmid was added to 100 µL aliquotes of competent cells and incubated on ice for 5 min. The reaction was given a 30 sec. heat shock at 42° C and incubated on ice for 1 min. 500 µL SOC media was added and the culture was
 5 incubated at 37°C for 1 hour. Cells were collected by centrifugation (15.000 x g, 1 min) redissolved in 150 µL 2xTY media and streaked on an agar plate with appropriate antibiotics (conc. 100 µg/mL ampicillin, 34 µg/mL chloramphenicol, 50 µg/mL kanamycin). Agar plates were incubated 12 hours at 37° C before clones were picked.

10

DNA sequence verification

Clones were picked from transformation plates and grown in 3 mL 2xTY at 37° C over night with appropriate antibiotics. Plasmid from each clone was purified by the QIAGENs Qiaprep Spin Miniprep Kit and eluted into 60 µL H₂O per spin
 15 column. A PCR screen was performed with T7 screen primers or pBAD screen primers according to the general PCR amplification scheme above with the primers listed in figure 26. Clones with right sized inserts were sent to sequencing at DNA-technology A/S. Sequencing primers used were T7 screen primers and pBAD screen primers.

20

Appendix B

Buffers and solutions

All buffers and solutions are in ddH₂O unless otherwise stated.

Fractional ratios (%) are indicated in weight per volume (w/v) for all buffers and
 25 solutions unless otherwise stated.

Agarose gel electrophoresis

Ethidium bromide 10 mg/mL ethidium bromide

Loading buffer (3x) 0.13 % bromophenol blue (v/v), 0.13 % xylene cyanol (v/v) ,
 30 30 % glycerol (v/v) TAE-buffer (50x) 2 M Tris-HCl (pH 8.0), 2 M acetic acid, 50 mM EDTA

Molecular weight marker λ-DNA, BstEII digested (Bilabs)

Calcium competent cells

TF1	10 mM NaAc pH 5.8, 50 mM MnCl ₂ , 5 mM NaCl
TF2	10 mM NaAc pH 5.8, 5 mM MnCl ₂ , 70 mM CaCl ₂ , 5 % glycerol

Cell growth and lysis

5	2xTY medium	1.6 % peptone from casein, 1 % yeast extract, 0.5 % NaCl (autoclaved)
	2xTY agar	2xTY medium with 1.5 % bacto agar (autoclaved)
	Ampicillin	25 mg/mL, 0.2 µM filtered
	Chloramphenicol	34 mg/mL in ethanol and 0.2 µM filtered
10	Kanamycin	25 mg/mL, 0.2 µM filtered
	L-arabinose	20 % L-arabinose, 0.2 µM filtered
	IPTG	200 mg/mL IPTG (0.8 M), 0.2 µM filtered
	0.9 % NaCl	0.9 % NaCl (autoclaved)
	Lysis buffer	50 mM Tris-HCl pH 8, 150 mM NaCl, 1 mM EDTA, 0.1 mM PMSF
15		
	Redissolvment buffer	10 mM Tris-HCl pH 8, 0.1 mM PMSF

DNA amplification

	PCR buffer (10x)	750 mM Tris-HCl (pH 9.0), 200 mM (NH ₄) ₂ SO ₄ , 0.1 % tween (v/v)
20	MgCl ₂	25 mM MgCl ₂
	dNTP-mix	10 mM dATP, 10 mM dTTP, 10 mM dCTP, 10 mM dGTP
	Pfu DNA polymerase	N.D
	All primers are listed in Figures 25 and 26	

25

DNA digestion

	BamHI	20 U/µL
	BspHI	10 U/µL
	BSA (10x)	10 mg/mL
30	EcoRI	20 U/µL
	NcoI	10 U/µL
	NEB buffer 1 (10x)	100 mM Bis-Tris-Propane-HCl, 100 mM MgCl ₂ , 10 mM dithiothreitol, pH 7
	NEB buffer 3 (10x)	1 M NaCl, 500 mM Tris-HCl, 100 mM MgCl ₂ , 10 mM dithiothreitol, pH 7.9
35		

NEB buffer 4 (10x) 500 mM potassium acetate, 200 mM tris-acetate, 100 mM magnesium acetate, 10 mM dithiothreitol, pH 7.9

SacI 20 U/μL

SphI 40 U/μL

5

DNA Ligation

T4 ligase buffer (10x) 0.5 M Tris-HCl (pH 7.5), 0.1 MgCl₂, 0.01 M ATP, 0.1 M DTT

T4 DNA ligase 400.000 U/mL

10

DNA purification

Qiaprep® Spin miniprep Kit (Qiagen)

Wizard® SV Gel and PCR Clean-Up system (Omega)

15

Protein purification

Buffer A 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 5 mM Imidazole

Buffer C 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 500 mM Imidazole.

H(0) 50 mM HEPES (pH 7.6), 10 mM MgCl₂, 1 mM DTT, 0.1 mM

20

PMSF, 15 mM NaN₃

Storage Buffer 10 % 10x H(0), 100 mM NaCl, 50 % glycerol

SDS-PAGE

Cracking buffer 50 mM Tris-HCl (pH 6.8), 15 % glycerol, 10 % SDS,

25

140 mM β-mercaptoethanol, 0.1 % bromophenol blue, 2 mM EDTA

Sample buffer (5x) 125 mM Tris-HCl (pH 6.8), 50 % glycerol, 10 % SDS, 760 mM β-mercaptoethanol, 0.13 % bromophenol blue

Separation buffer 0.75 M Tris-HCl (pH 8.8), 0.2 % SDS

30

Stacking buffer 0.75 M Tris-HCl (pH 6.8), 0.2 % SDS

Electrophoresis buffer (10x) 0.25 M Tris-HCl (pH 8.3), 1.92 M glycine, 1% SDS

Molecular weight marker PageRuler™ Prestained protein ladder (fermentas). (170, 130, 95, 72, 55, 43, 34, 26, 17, 11 kDa)

Acrylamide-bis solution 30 % 37.5:1 of acrylamide:N,N'-Methylene Bis-acrylamide

35

APS	10 % ammoniumpersulfate
Staining solution	Gelcode ® Blue stain reagent

Western Blotting

5	Blocking buffer	10 % skim milk in TBS-T
	Electrode blotting buffer	48 mM Tris-HCl (pH 8.9), 39 mM glycine, 0.0375 % SDS, 20 % ethanol (v/v)
	Primary antibody	0.74 mg/mL monoclonal mouse anti-E.coli IF1 antibody no. 1 produced in Biodesign lab
10	Secondary antibody	1.3 mg/mL polyclonal rabbit anti-mouse immunoglobulin antibody HRP-conjugated
	TBS-T	0.9 % NaCl, 50 mM Tris-HCl (pH 7.4), 1 % tween (v/v)

The ECL Plus Western Blotting Detection System (GE healthcare) was used for
 15 westernblotting. The kit consists of two solutions ECLA and ECLB, compositions of these is unknown.

Example 9

Aim: To test whether the *InfB*(1-21) expressivity tag can increase the expression of other genes than *gfp* and *InfA*. Two poorly expressed genes *streptavidin* (SEQ
 20 ID NO: 88 (without tag) and SEQ ID NO: 89 (with tag)) and *I32* (SEQ ID NO: 90 (without tag) and SEQ ID NO: 91 (with tag)) was chosen as reporter proteins.

Intro: Streptavidin (Genbank: X03591.1) is from the organism *Streptomyces avidinii*. Its 24 amino acid signal peptide was excluded (Argarana, C.E., et al.
 25 1986, Miksch G, et al. 2008).

L32 is a single chain antibody fragment (scFv) of human origin (Sanz, L., et al. 2001 and Sanz, L., et al. 2002).

Streptavidin has been expressed in *E. coli* in several vector contexts, and is
 30 included in this example, since it is a very important biotechnological protein product. In most of the reported expression systems for streptavidin, the system has been optimally tuned for high expression. Many of these optimized systems for heterologous expression often results in premature translational abortion and

protein folding problems. In order to avoid these problems it is often a necessity to employ sub-optimal expression conditions.

Expression of murin or humanized single chain antibodies are notoriously difficult to express in *E. coli* (Arbabi-Ghahroudi M et. Al. 2005). In this example, the L32 single chain antibody raised against human laminin through phagedisplay technology, has been expressed with and without the 21nt expressivity tag.

This example is carried out under such sub-optimal conditions, e.g. less strong promotor and low expression temperature. Under these conditions streptavidin and L32 is expressed at levels that is barely or not detectable on a SDS-PAGE.

Methods:

Cloning:

InfB(1-21)-*I32*, *InfB*(1-21)-*streptavidin*, *I32* and *streptavidin* were amplified from a template containing the appropriate gene by standard PCR. The *infB*(1-21) sequence was introduced through a non-annealing part of the forward primer (see Figure 30). Forward primer contained a *Bsp*HI restriction site, followed by the *infB*(1-21) sequence and the annealing region of the appropriate gene. The forward primer for genes without an *infB*(1-21) sequence contained a *Nco*I restriction site, followed by the annealing region of the appropriate gene (*I32* or *streptavidin*). Reverse primer contained an *Eco*RI restriction site followed by the annealing region of the appropriate gene (*I32* or *strep*).

The PCR amplification included 30 cycles of 1 min at 94 °C (DNA denaturation), 1 min at 56 °C (primer hybridization) and 1:10 min at 72 °C (DNA elongation), followed by a single step for 5 min at 72 °C (final elongation).

The PCR products were digested with *Bsp*HI or *Nco*I, respectively, and *Eco*RI and ligated into a *Nco*I and *Eco*RI treated pBAD plasmid, followed by transformation of ligation mixture into chemical competent *E. coli* JM109 cells.

Recombinant transformants on agar plate were identified using PCR screen.

Identified plasmids with a correct insertion were verified by sequencing (DNA sequencing facility, MBI, Aarhus University, Denmark). Verified plasmids were transformed into chemical competent *E. coli* BL21 cells for expression.

In figure 30 the restriction site for BspHI and EcoRI in the forward-primer and the reverse-primer, respectively, is underlined. The hybridization region of the primers is shown by italic and a bold for the *infB*(1-21) sequence and the annealing sequence with the appropriate gene, respectively.

5

Protein expression:

pBAD*InfB*(1-21)-*l32*, *InfB*(1-21)-*strep* and pBAD-*l32*, -*streptavidin*) was transformed into chemical competent *E. coli* BL21 cells. For each construct, two 50 mL cultures were grown in 2x TY (containing 100 µg/ml ampicillin) at 37 °C. At an OD₆₀₀ of 0.8, the gene expression was induced by adding L-arabinose to a final concentration of 0,04 %. For each construct, one culture was incubated at 30 °C and another at 37 °C for 3,5 h post induction. Analysis of expression was performed by a SDS-PAGE.

15 **Results:**

Streptavidin expression

The data presented in figure 29 clearly shows that an increase in the expression of streptavidin is observed when the tag is present. Arrows indicate the expected size of the proteins in question. This example is carried out under such sub-optimal conditions, e.g. less strong promotor and low expression temperature. Under these conditions streptavidin (no tag) is expressed at a level that is not detectable on a SDS-PAGE. Under the same conditions, but using the 5' 21nt expressivity tag a significant expression level is observed (more than 20 fold higher than with no tag, estimated using gel-scanning technology).

25

L32 expression

The data presented in figure 29 clearly shows that an increase in the expression of L32 takes place when the tag is present. Arrows indicate the expected size of the proteins in question. This example is carried out under such sub-optimal conditions, e.g. less strong promotor and low expression temperature. Under these conditions L32 (no tag) is expressed at a level that is barely detectable on a SDS-PAGE. Under the same conditions, but using the 5' 21nt expressivity tag a significant expression level is observed (more than 10 fold higher than with no tag, estimated using gel-scanning technology).

35

Example 10

In a pilot study expression of subset of matrix metalloproteases were tested for expression using the 21 nt tag according to the invention. (MMP: Matrix metalloprotease; MT: Membran-type)

5

Tested constructs:

Human MT1-MMP catalytic domain

Human MT1-MMP catalytic domain (E250A)

Human MT1-MMP hemopexin domain

10

Human MT1-MMP hemopexin domain+linker

Human MT1-MMP catalytic and hemopexin domain+linker

Human MMP-9 catalytic domain

Human MMP-2 catalytic domain

Human MMP-3 catalytic domain

15

Human MMP-12 catalytic domain

Human MMP-8 catalytic domain

Human MMP-1 catalytic domain

Human MMP-7 catalytic domain

20

None of these constructs gave any detectable expression in *E. coli* without the tag. By including the 21 nt expressivity tag they were expressed with up to 50 mg protein/L culture. Further information on these protein can be found in (Maeda H, 1998)

25 **References**

Wolfgang Peti a and Rebecca Page. Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost. Protein Expression and Purification 51 (2007) 1–10.

30

Sørensen et. al, 2003. A favorable solubility partner for the recombinant expression of streptavidin. Protein Expr Purif. **32**:252-259

Looman A.C. et. al, 1987. Influence of the codon following the AUG initiation codon on the expression of a modified lacZ gene in Escherichia coli. The EMBO jour. 6:8:2489-2492.

- 5 Stenström C.M. et al., 2000. Codon bias at the 3'-side of the initiation codon is correlated with translation initiation in Escherichia coli. Gene. 263:273-284.

Argarana, C.E., et al., *Molecular cloning and nucleotide sequence of the streptavidin gene*. Nucleic acids research, 1986. **14**(4): p. 1871.

10

Sanz, L., et al., *Generation and characterization of recombinant human antibodies specific for native laminin epitopes: potential application in cancer therapy*. Cancer Immunology, Immunotherapy, 2001. **50**(10): p. 557-565.

- 15 Sanz, L., et al., *Single-chain antibody-based gene therapy: inhibition of tumor growth by in situ production of phage-derived human antibody fragments blocking functionally active sites of cell-associated matrices*. Gene therapy, 2002. **9**(15): p. 1049.

- 20 Miksch G, Ryu S, Risse JM, Flaschel E. Factors that influence the extracellular expression of streptavidin in Escherichia coli using a bacteriocin release protein. Appl Microbiol Biotechnol. 2008; 81(2):pp 319-26.

Arbabi-Ghahroudi M, Tanha J, MacKenzie R. Prokaryotic expression of antibodies.

- 25 Cancer Metastasis Rev. 2005;24(4):pp 501-19.

Maeda H, Okamoto T, Akaike T. Human matrix metalloprotease activation by insults of bacterial infection involving proteases and free radicals. Biol Chem. 1998 Feb;379(2):193-200.

Claims

1. A recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide, said nucleic acid comprising a first nucleic acid sequence of maximum 99 nt encoding a polypeptide operably linked to the a second heterologous nucleic acid sequence encoding a polypeptide, wherein said first nucleic acid sequence comprises a nucleic acid sequence selected from the group consisting of
- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
 - b) a nucleic acid sequence having at least 75% sequence identity to a),
 - and
 - c) a nucleic acid sequence, which is a sub-sequence of a) and b).
2. The recombinant hybrid nucleic acid sequence according to claim 1, wherein said first nucleic acid sequence encodes an expressivity tag.
3. The recombinant hybrid nucleic acid sequence according to any of the claims 1-2, wherein said first nucleic acid sequence of c) has a minimum length of 15 nt nucleotides.
4. The recombinant hybrid nucleic acid sequence according to any of the preceding claims 1-3, wherein said first nucleic acid sequence is positioned 5' to the second heterologous nucleic acid sequence.
5. The recombinant hybrid nucleic acid sequence according to any of the preceding claims 1-4, wherein said second heterologous nucleic acid sequence has been modified to reflect the codon preference of a host organism selected to express the nucleic acid molecule.
6. The recombinant hybrid nucleic acid sequence according to any of the preceding claims, further comprising a nucleic acid sequence encoding a polypeptide tag for detection, solubility or affinity purification.

7. The recombinant hybrid nucleic acid sequence according to any of the preceding claims, wherein said tag for affinity purification, solubility or detection is selected from the group consisting of
BCCP-tag, c-myc-tag, calmodulin-tag (SEQ ID NO:33), FLAG-tag (SEQ ID NO:45),
5 HA-tag (SEQ ID NO:50), His-tag (SEQ ID NO:29), Maltose binding protein-tag (SEQ ID NO:27), Nus-tag (SEQ ID NO:60), Glutathione-S-transferase-tag (SEQ ID NO:31), Green fluorescent protein-tag (SEQ ID NO:62), Thioredoxin-tag (SEQ ID NO:64), S-tag (SEQ ID NO:41), Strep-tag (SEQ ID NO:35), human protein C tag (SEQ ID NO:37), Chitin binding protein tag (SEQ ID NO:39), T7-tag (SEQ ID
10 NO:43), Myc-tag (SEQ ID NO:52), V5-tag, VSV-tag, Avi.tag (SEQ ID NO:56), BioEase-tag (SEQ ID NO:58), SNAP-tag, FlaSH-tag, Nus A-tag, DsbA-tag (SEQ ID NO:66) and the recombinant nucleic acid sequence itself.
8. The recombinant hybrid nucleic acid sequence according to any of the
15 preceding claims, further comprising a nucleic acid sequence encoding a spacer positioned between said first and said second nucleic acid sequence.
9. The recombinant hybrid nucleic acid sequence according to any of the preceding claims 1-8, further comprising a nucleic acid sequence encoding a
20 spacer positioned between the first and second codon of said first nucleic acid sequence and operably linking said codons.
10. A recombinant hybrid polypeptide, said polypeptide comprising a first polypeptide amino acid sequence and a second heterologous amino acid
25 sequence, wherein said first amino acid sequence of maximum 33 amino acids is selected from the group consisting of
- a) an amino acid sequence selected from the group consisting of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11 and SEQ ID NO:12, and
 - 30 b) an amino acid sequence having at least 75% sequence identity to a), and
 - c) an amino acid sequence which is a sub-sequence of a) and b).
11. The recombinant hybrid polypeptide according to claim 10, wherein said first
35 amino acid sequence is an expressivity tag.

12. The recombinant hybrid polypeptide according to any of the claims 10-11, wherein said first amino acid sequence of c) has a minimum length of 4 amino acids.

5

13. The recombinant hybrid polypeptide according to any of the claims 10-12, wherein said first amino acid sequence is positioned N-terminal to the second heterologous amino acid sequence.

10 14. The recombinant hybrid polypeptide according to any of the claims 10-13, further comprising a polypeptide tag for detection, solubility or affinity purification.

15 15. The recombinant hybrid polypeptide according to any of the claims 10-14, wherein said tag for affinity purification, solubility or detection is selected from the group consisting of

BCCP-tag, c-myc-tag, calmodulin-tag (SEQ ID NO:34), FLAG-tag (SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48 and SEQ ID NO:49), HA-tag (SEQ ID NO:51), His-tag (SEQ ID NO:30), Maltose binding protein-tag (SEQ ID NO:28), Nus-tag (SEQ ID NO:61), Glutathione-S-transferase-tag (SEQ ID NO:32), Green fluorescent protein-tag (SEQ ID NO:63), Thioredoxin-tag (SEQ ID NO:65), S-tag (SEQ ID NO:42), Strep-tag (SEQ ID NO:36), human protein C tag (SEQ ID NO:38), Chitin binding protein tag (SEQ ID NO:40), T7-tag (SEQ ID NO:44), Myc-tag (SEQ ID NO:53), V5-tag (SEQ ID NO:54), VSV-tag (SEQ ID NO:55), Avi.tag (SEQ ID NO:57), BioEase-tag (SEQ ID NO:59), SNAP-tag, FlaSH-tag, Nus A-tag, DsbA-tag (SEQ ID NO:67) and the recombinant polypeptide sequence itself.

16. A vector comprising a first nucleic acid sequence of maximum 99 nt, said first sequence comprising a nucleic acid sequence selected from the group consisting
30 of

- a) SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3 and SEQ ID NO:4, and
- b) a nucleic acid sequence having at least 75% sequence identity to a), and
- c) a nucleic acid sequence, which is a sub-sequence of a) and b).

17. The vector according to claim 16, wherein said vector is further adapted for insertion of a second heterologous nucleic acid sequence encoding a polypeptide operably linked to said first nucleic acid sequence.

5 18. The vector comprising a recombinant hybrid nucleic acid sequence according to any of the claims 1-9.

19. The vector according to any of claims 16-18, wherein said recombinant hybrid nucleic acid sequence is operably linked to an expression control sequence, as the
10 araBAD promoter system such as the Ara-promoter (SEQ ID NO:13), Ara-operator 1 (SEQ ID NO:14), Ara-operator 2 (SEQ ID NO:15) and the AraC repressor (SEQ ID NO:16), the lac promoter system such as the Lac Promoter (SEQ ID NO:17), the Lac-operator (SEQ ID NO:18) and the LacI repressor (SEQ ID NO:19), Tac/Trc-promoter system (SEQ ID NO:20), Lambda PL – promoter (SEQ ID
15 NO:21), Lambda PR promoter (SEQ ID NO:22), T7 promoter (SEQ ID NO:23), T3 promoter (SEQ ID NO:24), SP6 promoter (SEQ ID NO:25) and the AmpR promoter (SEQ ID NO:26).

20. The vector of any of claims 16-19, wherein the vector is selected from the
20 group consisting of plasmids, cosmids, phages, bacterial artificial chromosomes (BAC), Phagemids and P1-derived artificial chromosomes.

21. The vector according to any of the claims 16-20, wherein the vector is an in vitro translation vector.

25

22. The vector according to any of claims 16-21 , wherein said plasmid is selected from the group consisting of TA cloning vectors, Gateway cloning vectors, restriction cloning vectors, Topo cloning vectors, pET vector system and pBAD vector systems.

30

23. A host cell comprising a vector according to any of claims 16-22.

24. The host cell according to claim 23, wherein said host cell is a bacterium or fungus.

35

25. The host cell according to claim 23 or 24, wherein said host cell is *E.coli*.

26. The host cell according to any of claims 23-25, wherein the host cell is *E. coli* JM109 and *E. coli* BL21(DE3).

5

27. An isolated antibody or antigen-binding fragment thereof which selectively binds to the first polypeptide amino acid sequence of any of the claims 10-15.

28. The antibody, isolated antibody or antigen-binding fragment thereof
10 according to claim 27, wherein the antibody is a purification tag or an epitope tag.

29. Use of a recombinant hybrid nucleic acid sequence encoding a hybrid polypeptide according to the preceding claims for the expression of a said polypeptide.

15

30. The use according to claim 29, wherein said recombinant hybrid nucleic acid sequence is comprised in a vector according to any of the claims 16-22 adapted for expression of said recombinant polypeptide.

20 31. A method for expressing a recombinant polypeptide comprising
a) introducing into an expression system the hybrid nucleic acid sequence according to any of the claims 1-9 encoding a recombinant hybrid polypeptide,
b) expressing said nucleic acid sequence in the expression system, and
25 optionally
c) purifying the expressed recombinant polypeptide .

32. A method according to claim 31, wherein the hybrid nucleic acid sequence is comprised in a vector according to any of claims 16-22 is used.

30

33. A method according to claims 31 or 32, wherein the expression system is a host according to any of claims 23-26.

34. A method according to claims 31 or 32, wherein said expression system is an
35 in vitro expression system.

35. A method according to claim 34, wherein the in vitro translation system comprises extract from E.coli.

5 36. A method according to any of the claims 31-35, wherein an antibody according to claims 27 or 28 is used for purification of said recombinant polypeptide.

37. A method according to any of the claims 31-36, wherein an antibody
10 according to claims 18 or 19 is used for detecting said recombinant polypeptide.

38. A method according to any of the claims 31-37, wherein said first polypeptide amino acid sequence is subsequently removed from the expressed target nucleic acid sequence by proteolytic cleavage.

15

39. A kit comprising a vector according to any of the claims 16-22.

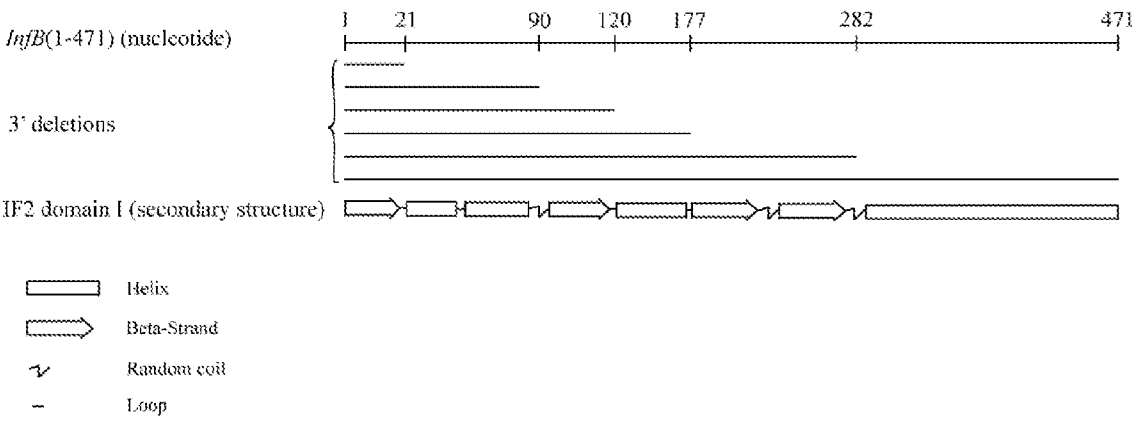


Fig. 1

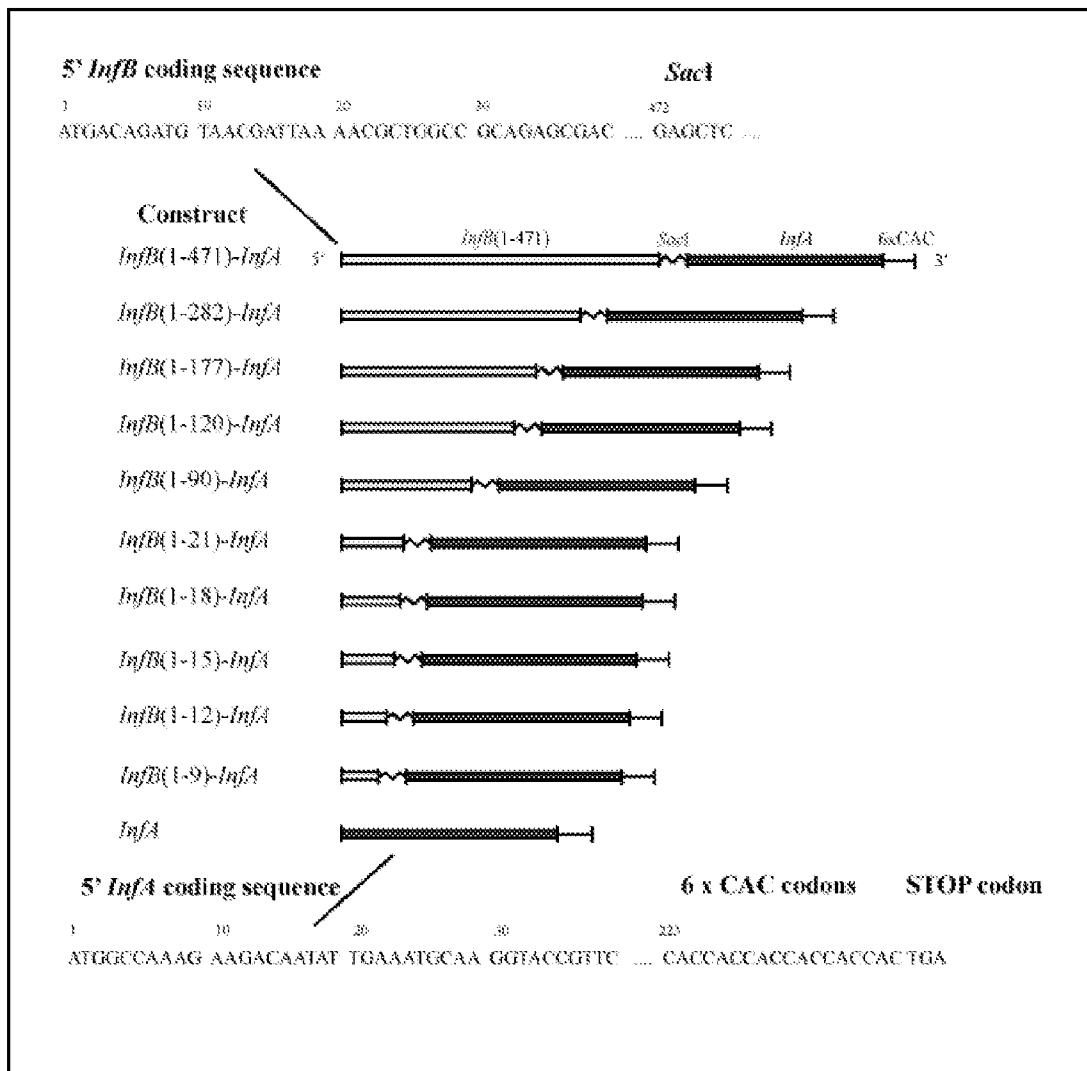


Fig. 2

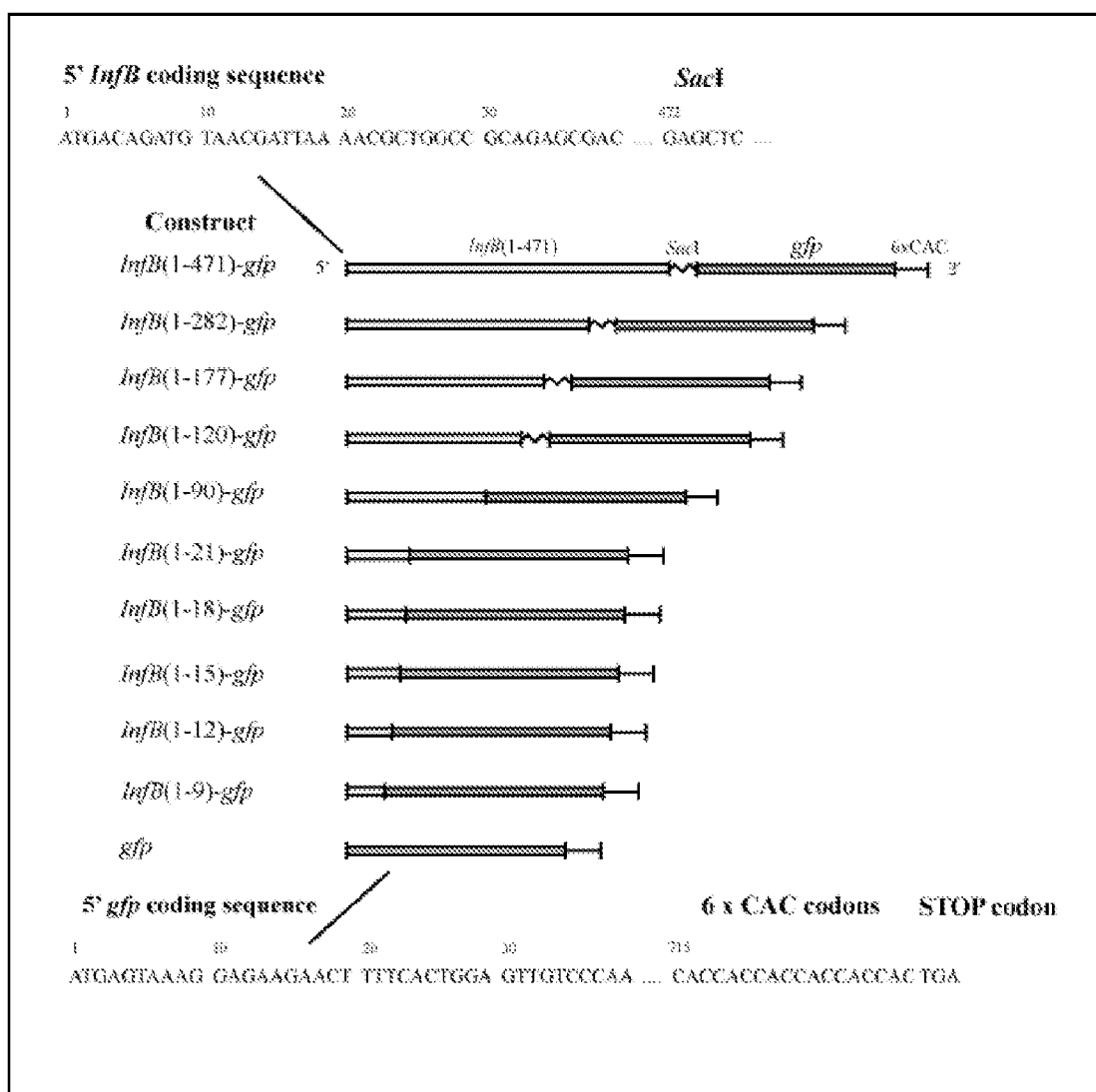


Fig. 3

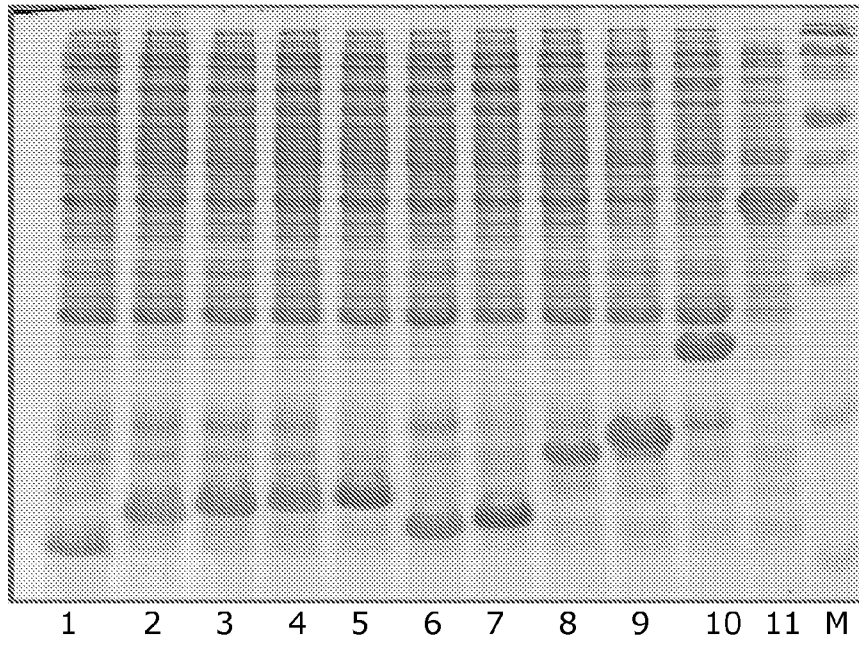


Fig. 4

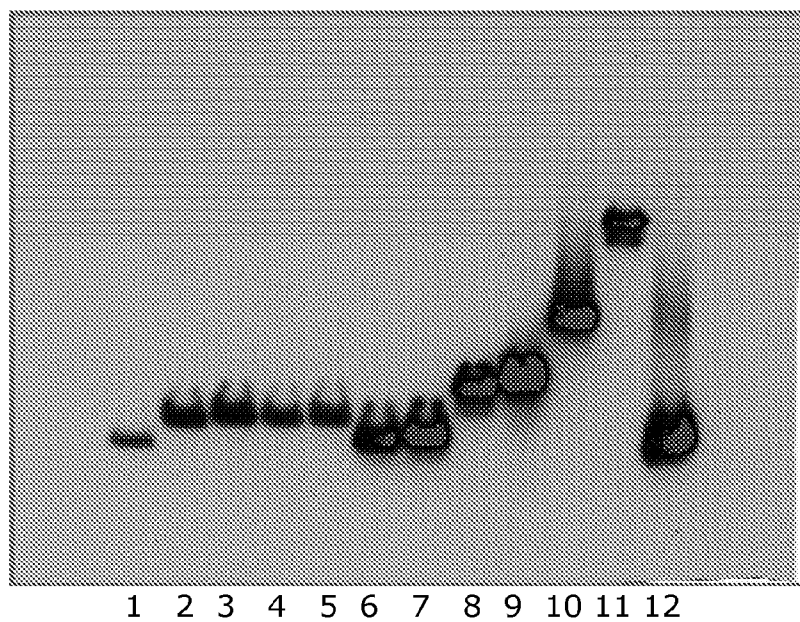


Fig. 5

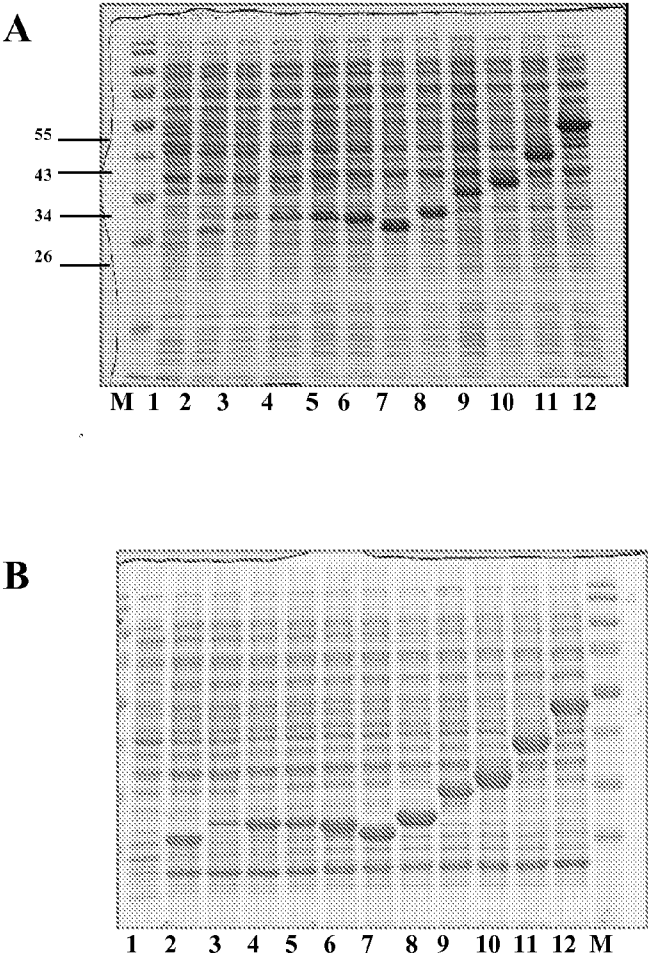


Fig. 6

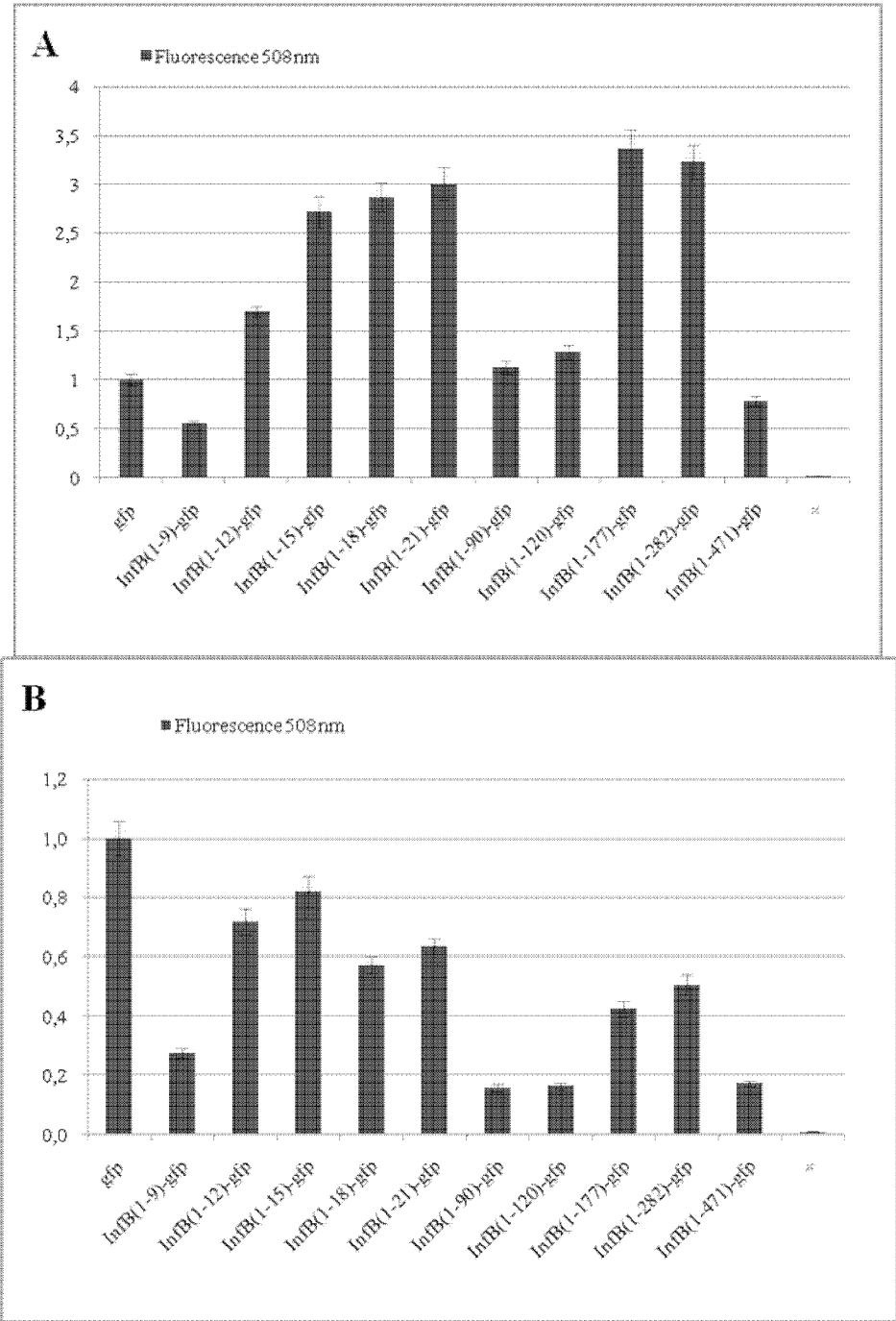


Fig. 7

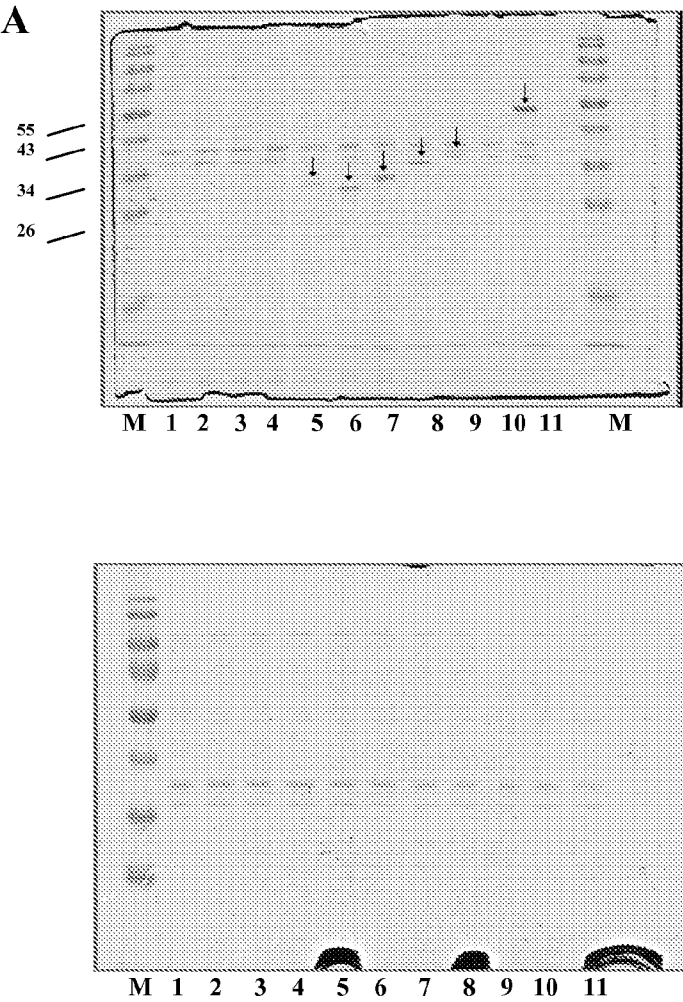


Fig. 8

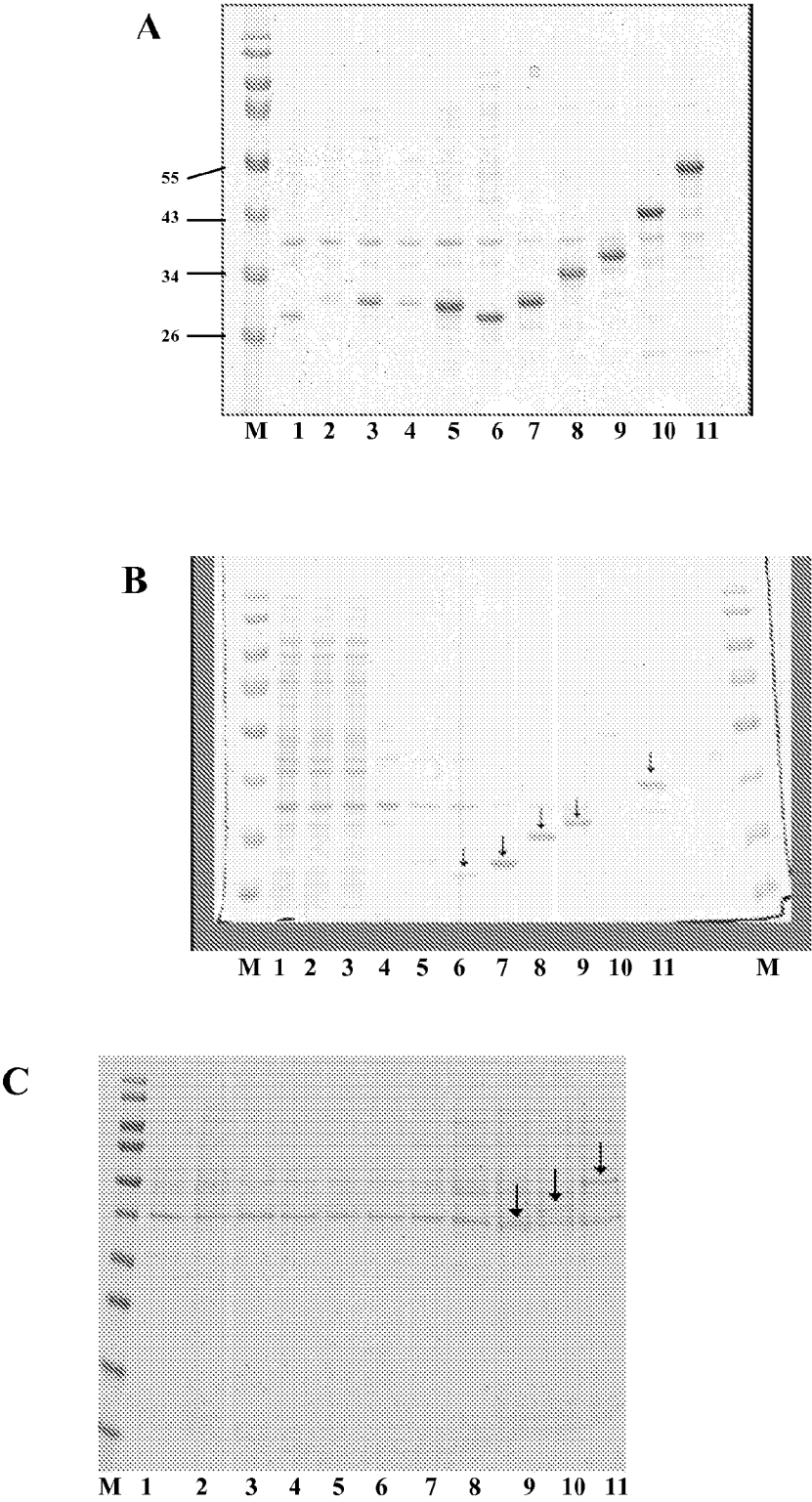


Fig. 9

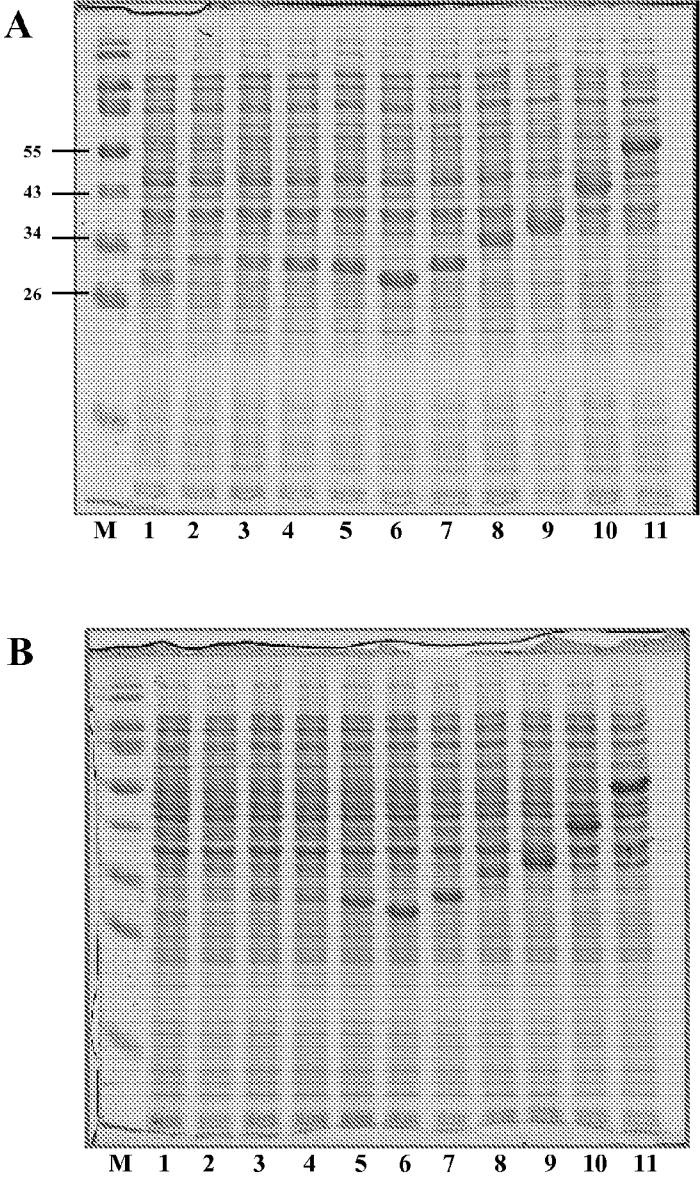


Fig. 10

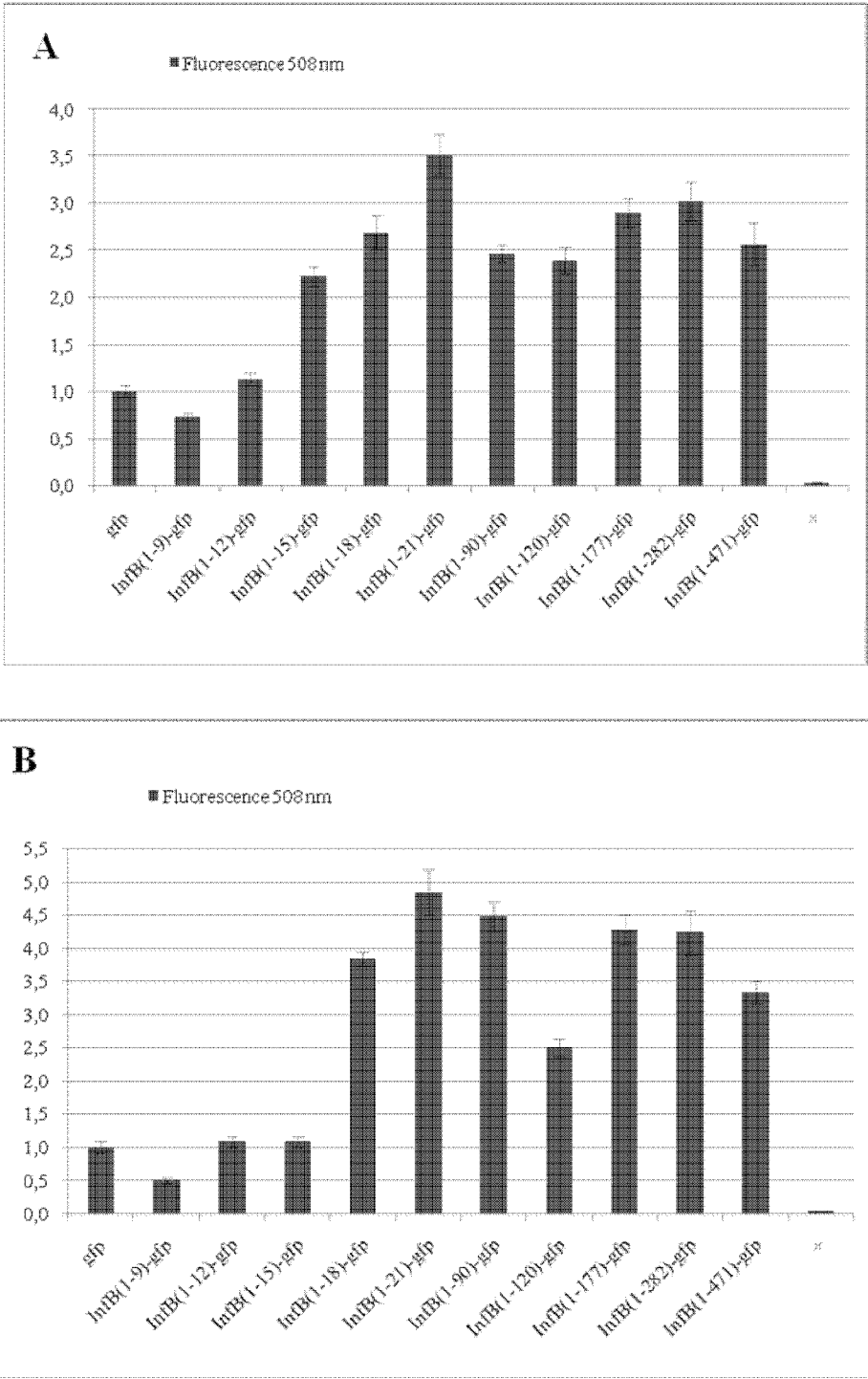


Fig. 11

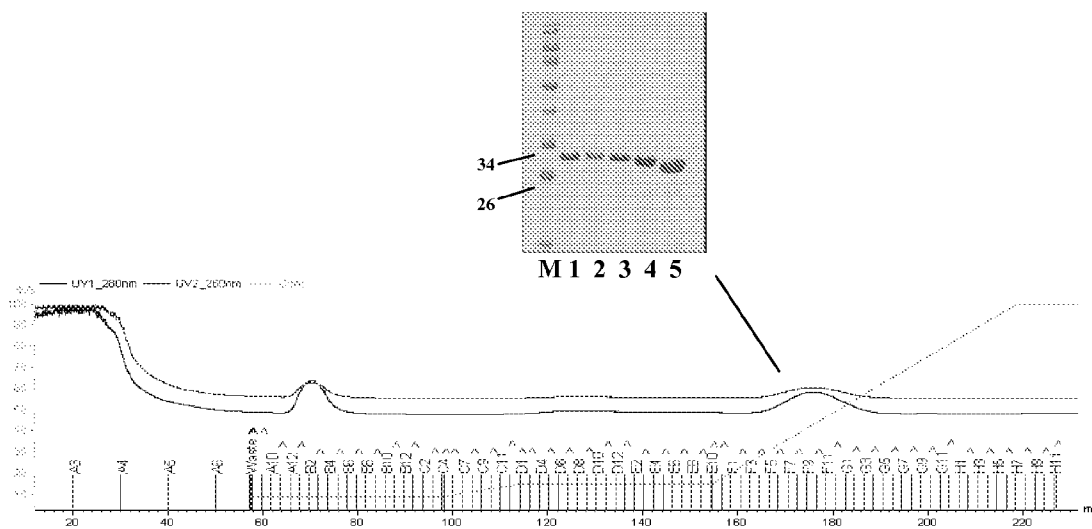


Fig. 12

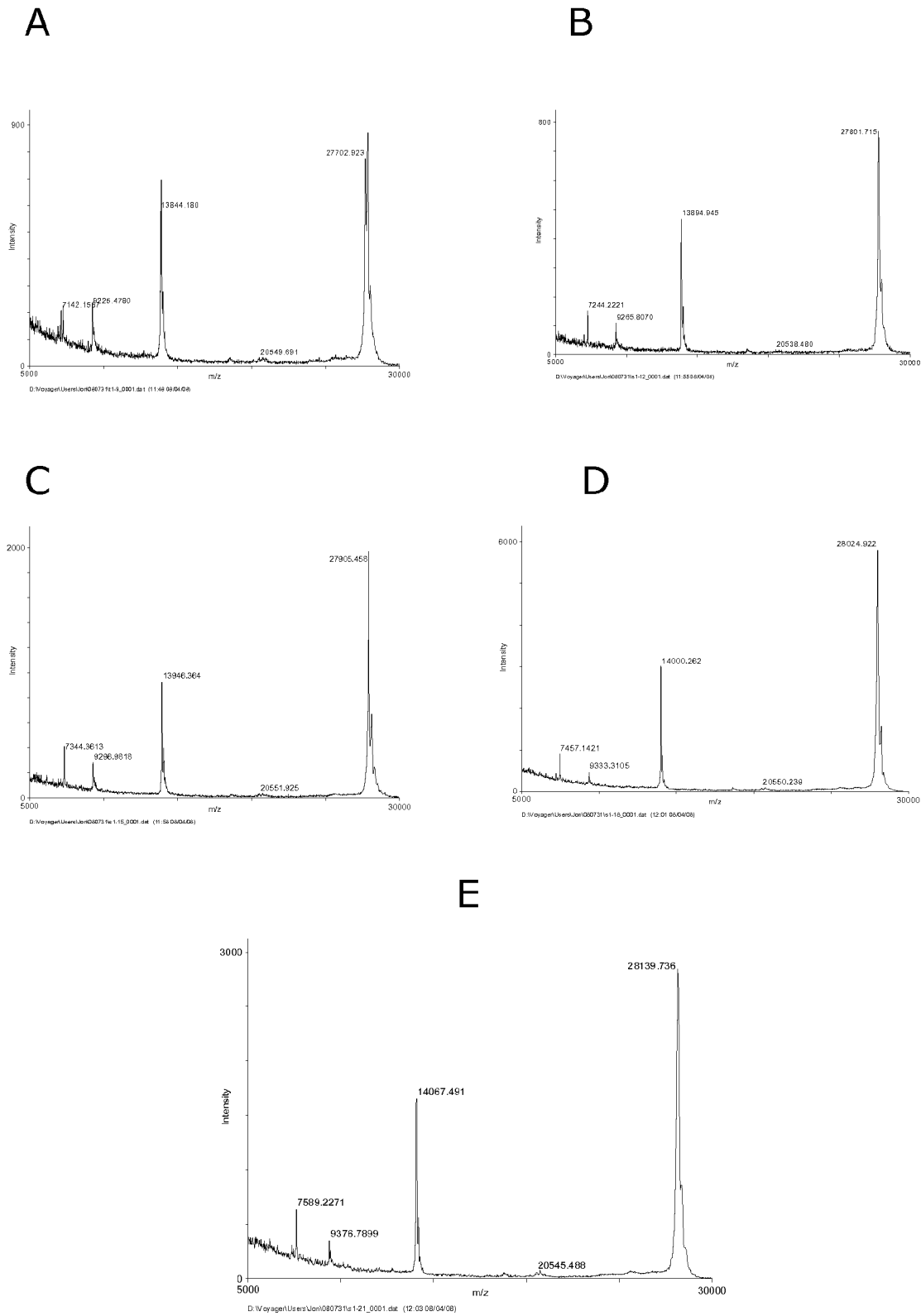


Fig. 13

Construct	z	m (Da)	Mean m/z (Da)	Theoretical (Da)
Fragment <i>InfB</i> (1-9)- <i>gfp</i>	1	7142,2	7142,2	7146,1
	1	27702,9		
<i>InfB</i> (1-9)- <i>gfp</i>	2	13844,2	27689,2	27703,9
	3	9225,5		
Fragment <i>InfB</i> (1-12)- <i>gfp</i>	1	7244,2	7244,2	7245,2
	1	27801,7		
<i>InfB</i> (1-12)- <i>gfp</i>	2	13894,9	27796,3	27803,1
	3	9265,8		
Fragment <i>InfB</i> (1-15)- <i>gfp</i>	1	7344,4	7344,4	7346,3
	1	27905,5		
<i>InfB</i> (1-15)- <i>gfp</i>	2	13946,4	27896,4	27904,2
	3	9297,0		
Fragment <i>InfB</i> (1-18)- <i>gfp</i>	1	7457,1	7457,1	7461,5
	1	28024,9		
<i>InfB</i> (1-18)- <i>gfp</i>	2	14000,3	28008,5	28017,3
	3	9333,3		
Fragment <i>InfB</i> (1-21)- <i>gfp</i>	1	7589,2	7589,2	7587,6
	1	28139,7		
<i>InfB</i> (1-21)- <i>gfp</i>	2	14067,5	28135,0	28145,5
	3	9376,8		

Fig. 14

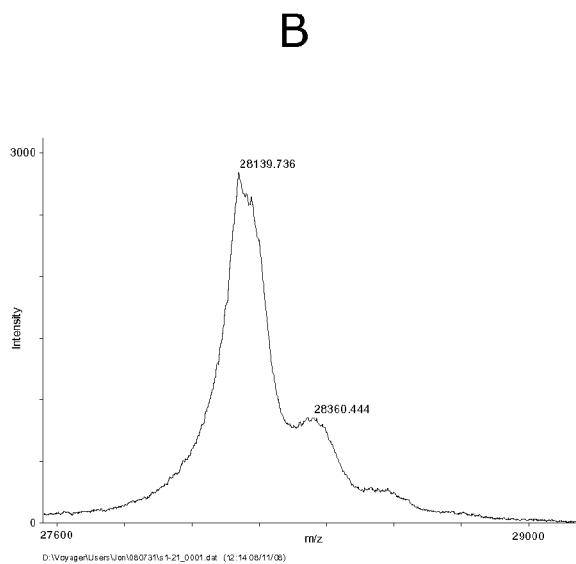
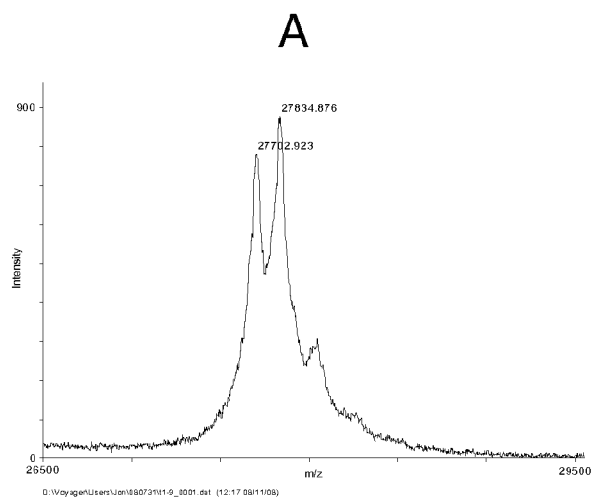


Fig. 15

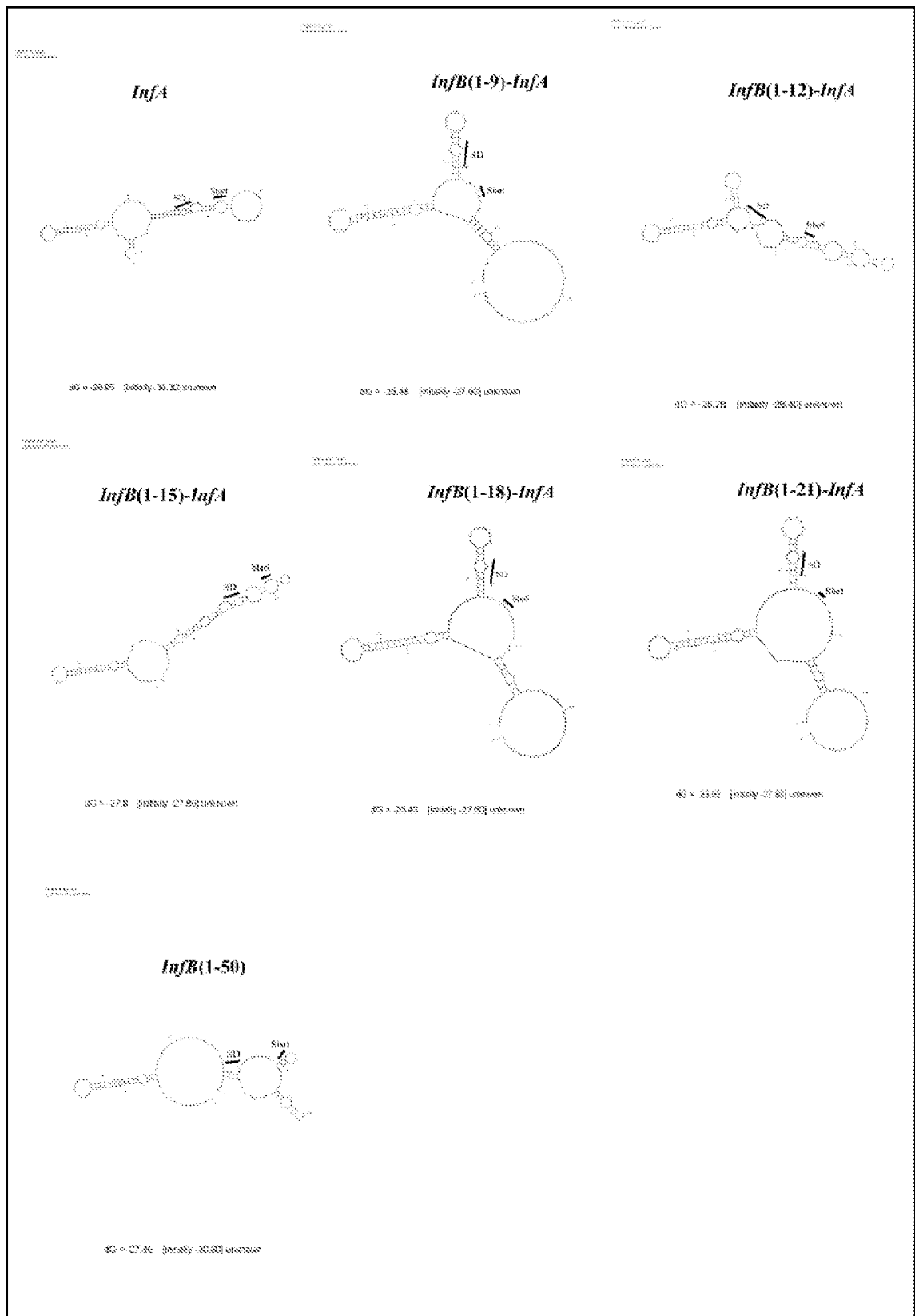


Fig. 16

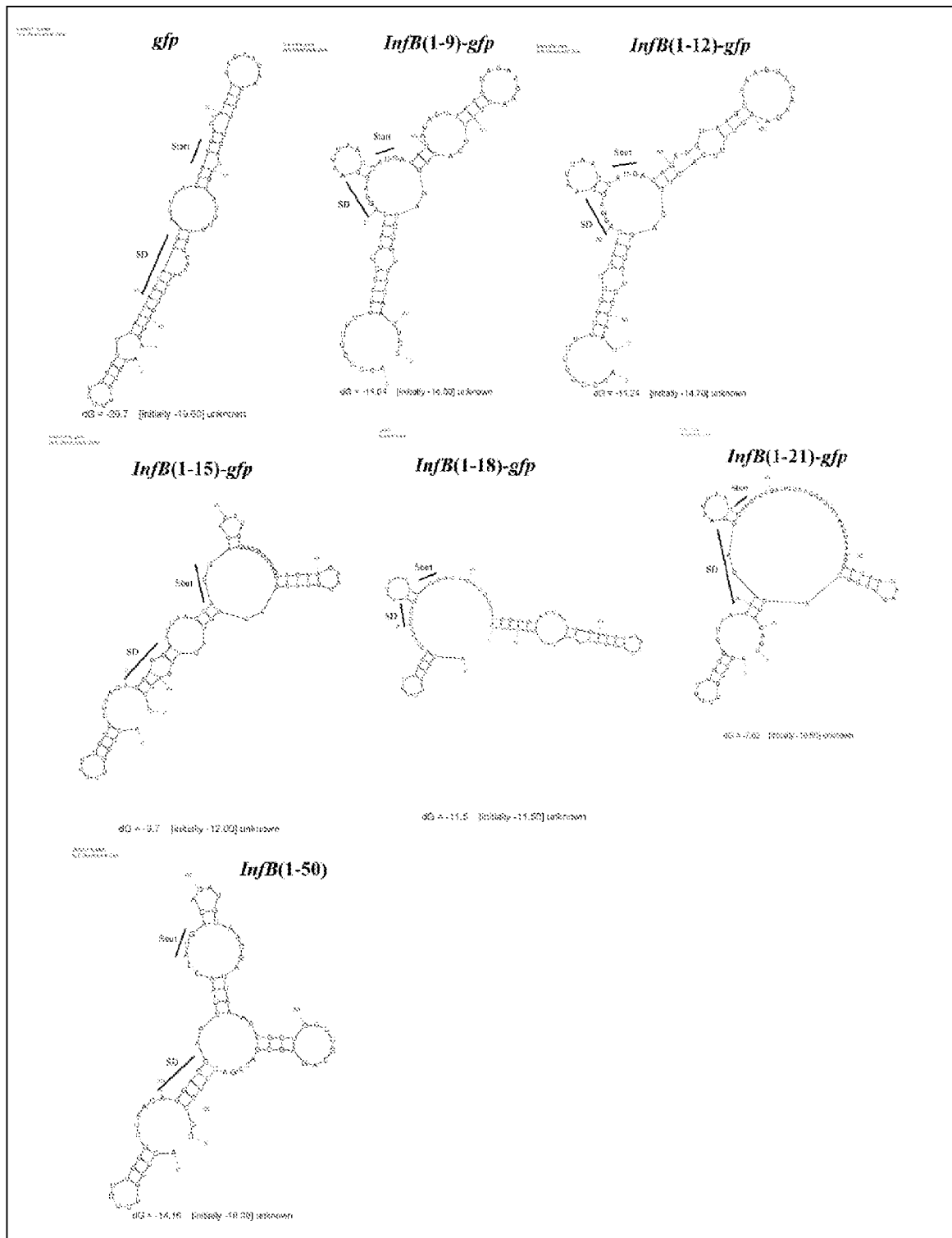
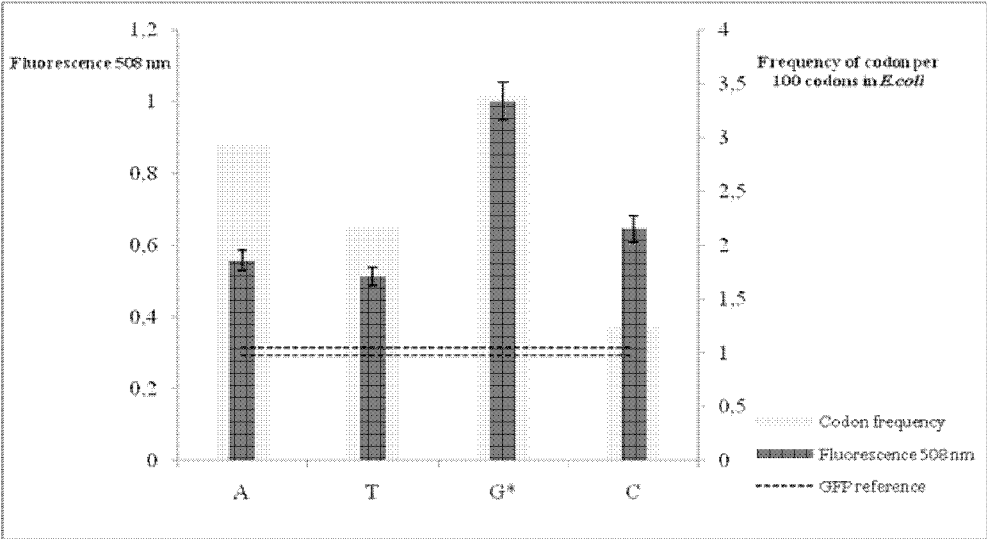


Fig. 17

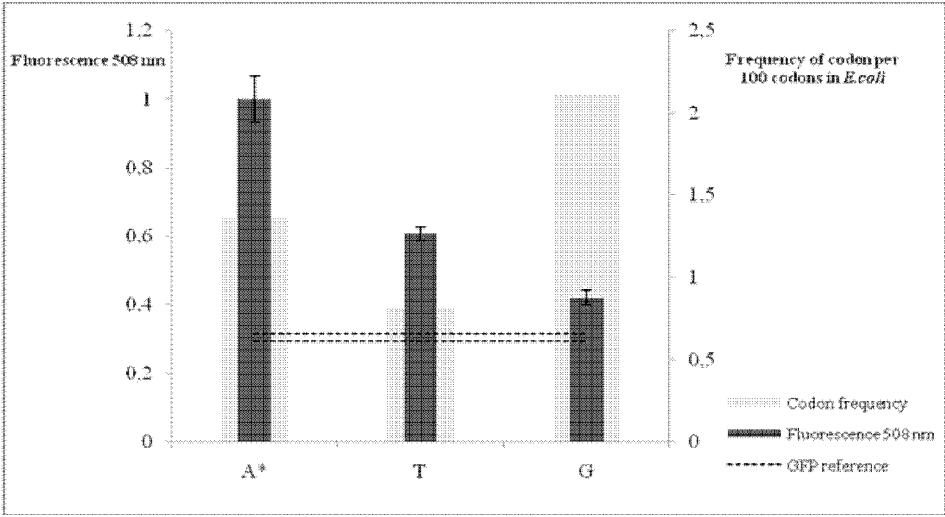
+1
InfB(1-21) **ATG** **A**_{CA} **GAT** **G**_T**A** **A**_{CG} **ATT** **A**_{AA}
Trx(1-18) **ATG** **A**_{GC} **GAT** **A**_{AA} **ATT** **ATT**

Fig. 18

A



B



C

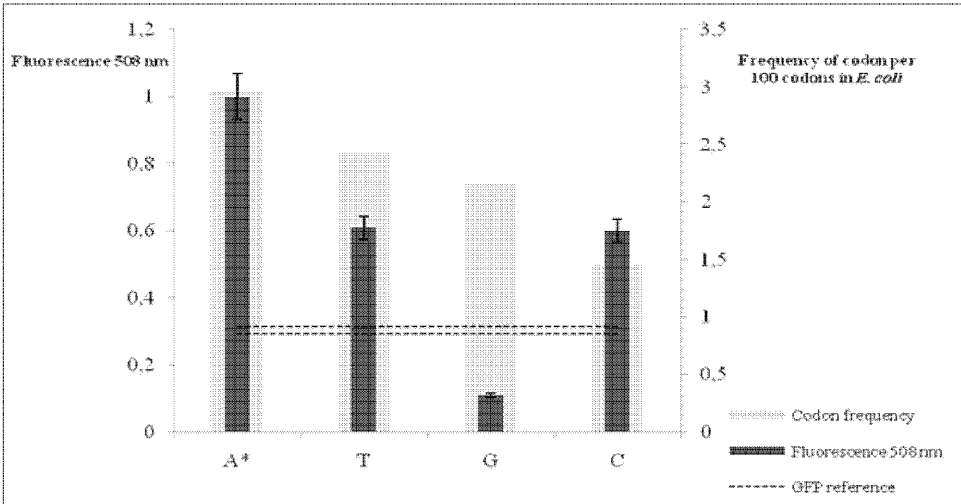


Fig. 19

Construct	Sequence	A/U	CA
<i>InfB</i> (1-21)	AUGACAGAUUGUAACGAUUAAA	15/21	1
<i>InfA</i> (1-21)	AUGGCCAAAGAAGACAAUAUU	14/21	2
<i>gfp</i> (1-21)	AUGAGUAAAGGAGAAGAACUU	14/21	0

Fig. 20

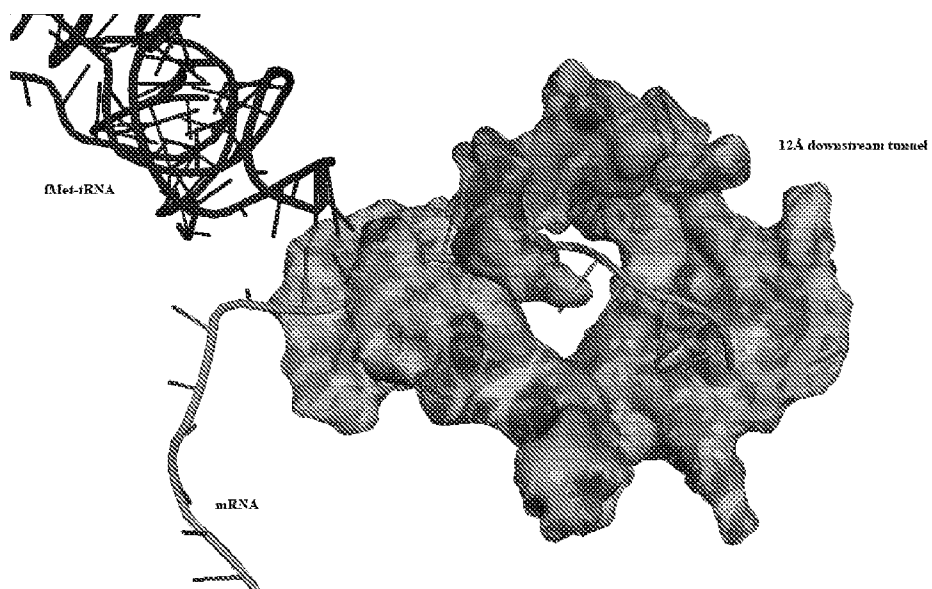


Fig. 21

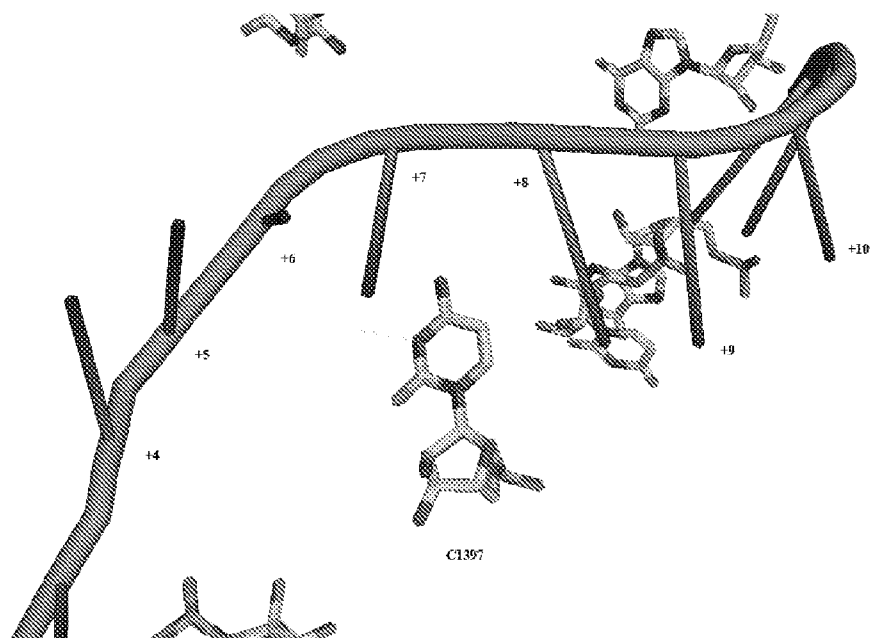


Fig. 22

Codon frame		+1	+2	+3	+4	+5
<i>InfB</i> (1-21)	A U G A C A	G A U	G U A	A C G	A U U	A A A
<i>InfA</i>	A U G G C C	A A A	G A A	G A C	A A U	A U U
<i>gfp</i>	A U G A G U	A A A	G G A	G A A	G A A	C U U

Fig. 23

2nd codon	<i>InfB</i> (ACA)	<i>InfA</i> (GCC)	<i>gfp</i> (AGU)
Looman et al.	0,10	0,70	0,50
Steenström et al.	0,58	0,11	0,44

Fig. 24

Table I

PCR	Primer sequence	Introduced restriction sites	°C
<i>InfA</i> I	F CCCC <u>TCATGA</u> GCCAAAGAAGACAATATTGAA	<i>Bsp</i> HI	55
	R CGGGATCCTCATCAGTGGTGGTGGTGGTGGTGGCGACTACGGAAGACAATGCGGCCTT	<i>Bam</i> HI	
<i>InfA</i> II	F CCGAGCTCGCCAAAGAAGACAATATTGAA	<i>Sac</i> I	55
	R CGGGATCCTCATCAGTGGTGGTGGTGGTGGTGGCGACTACGGAAGACAATGCGGCCTT	<i>Bam</i> HI	
<i>InfB</i> (1-9)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	58
	R CCCGAGCTCATCTGTTCATGGTATATCTCCT	<i>Sac</i> I	
<i>InfB</i> (1-12)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	58
	R CCCGAGCTCTACATCTGTTCATGGTATATCT	<i>Sac</i> I	
<i>InfB</i> (1-15)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	58
	R CCCGAGCTCCGTTACATCTGTTCATGGTATA	<i>Sac</i> I	
<i>InfB</i> (1-18)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	58
	R CCCGAGCTCAATCGTTACATCTGTTCATGGT	<i>Sac</i> I	
<i>InfB</i> (1-21)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	58
	R CCCGAGCTCTTTAATCGTTACATCTGTTCAT	<i>Sac</i> I	
<i>InfB</i> (1-90)	F CCCCCCGCATGCAAGGAGATGGCGCCCA	<i>Sph</i> I	54
	R CCCGAGCTCCCGGATACCTGCATAGCAAA	<i>Sac</i> I	
<i>InfB</i> (1-120)	F ACGTTTCATGACAGATGTAACGATTAAAAACG	<i>Bsp</i> HI	57
	R CCCGAGCTCTTGTGCAGACACAGTCGTCA	<i>Sac</i> I	
<i>InfB</i> (1-177)	F ACGTTTCATGACAGATGTAACGATTAAAAACG	<i>Bsp</i> HI	52
	R CCCGAGCTCCAATTGTTCGGGGCCTGAATT	<i>Sac</i> I	
<i>InfB</i> (1-282)	F ACGTTTCATGACAGATGTAACGATTAAAAACG	<i>Bsp</i> HI	52
	R CCCGAGCTCCGCGTTTCCAAAGGTGCGTTT	<i>Sac</i> I	
<i>InfB</i> (1-471)	F ACGTTTCATGACAGATGTAACGATTAAAAACG	<i>Bsp</i> HI	64
	R CCCGAGCTCTTGTCTTTTCCGACGCTICA	<i>Sac</i> I	

Table II

PCR		Primer sequence	Introduced restriction sites	°C
<i>gfp</i> I	F	CCCTCATGAGTAAAGGAGAAGAACTTTCA	BspHI	52
	R	CGGAAATTCATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>gfp</i> II	F	CCC <u>GAGCTC</u> AGTA AAGGAGAAGAAC TTTCA	SacI	52
	R	CGGAAATTCATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-9)-gfp</i>	F	CCCC <u>T CATGA</u> CAGATAGTAAAGGAGAAGAA	BspHI	52
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-12)-gfp</i>	F	CCCC <u>T CATGA</u> CAGATGTAAGTAAAGGAGAAGAA	BspHI	52
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-15)-gfp</i>	F	CCCC <u>T CATGA</u> CAGATGTAACGAGTAAAGGAGAAGAA	BspHI	52
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-18)-gfp</i>	F	CCCC <u>T CATGA</u> CAGATGTAACGATTAGTAAAGGAGAAGAA	BspHI	52
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-21)-gfp</i>	F	CCCC <u>T CATGA</u> CAGATGTAACGATTAAAGTAAAGGAGAAGAA	BspHI	54
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-90) I</i>	F	ACGT <u>T CATGA</u> CAGATGTAACGATTA AAAACG	BspHI	54
	R	TTCTTCCTCTTACTCCGGATACCTGCATCAGCAA		
<i>InfB(1-90) II</i>	F	GATGCAGGTATCCCGAGTAAAGGAGAAGAACTTTT		54
	R	C <u>GGAATTC</u> ATCATCAGTGGTGGTGGTGGTGGTTTTGTATAGTTTCATCCATGCC	EcoRI	
<i>InfB(1-90)-gfp</i>	F	F <i>InfB(1-90)</i> I	BspHI	55
	R	R <i>InfB(1-90)</i> II	EcoRI	
<i>InfB(1-120)</i>	F	ACGT <u>T CATGA</u> CAGATGTAACGATTA AAAACG	BspHI	55
	R	CCC <u>GAGCTCT</u> TGTGCAGACACAGAGTCGTCA	SacI	
<i>InfB(1-177)</i>	F	ACGT <u>T CATGA</u> CAGATGTAACGATTA AAAACG	BspHI	52
	R	CCC <u>GAGCTCCA</u> ATT TGTCCGGGCC TGAATT	SacI	
<i>InfB(1-282)</i>	F	ACGT <u>T CATGA</u> CAGATGTAACGATTA AAAACG	BspHI	52
	R	CCC <u>GAGCTCG</u> CGTTT CACA AAGGTGCGTTT	SacI	
<i>InfB(1-471)</i>	F	ACGT <u>T CATGA</u> CAGATGTAACGATTA AAAACG	BspHI	52
	R	CCCAGGCTCTTGTCAGACACAGAGTCGTCA	SacI	

PCR		Primer sequence	°C
T7 Screen	F	TAATACGACTCACTAT	55
	R	CCCAAGGGGTATGCTAG	
pBAD Screen	F	CTTTGCTATGCCATAGCATTT TTAT	55
	R	GAT TTAACTGTATCAGGCTG	

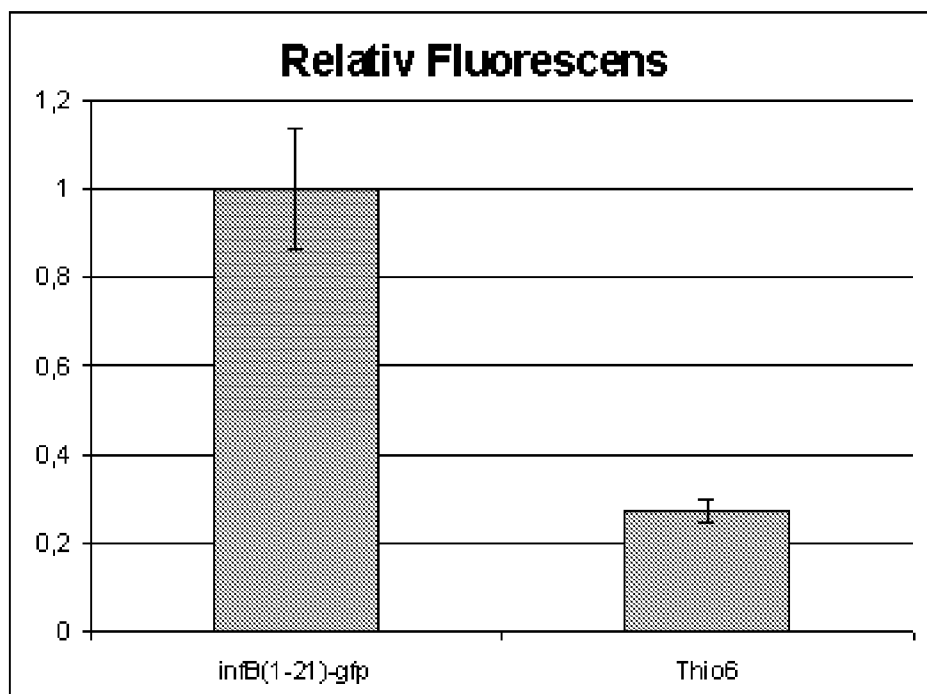


Fig. 27

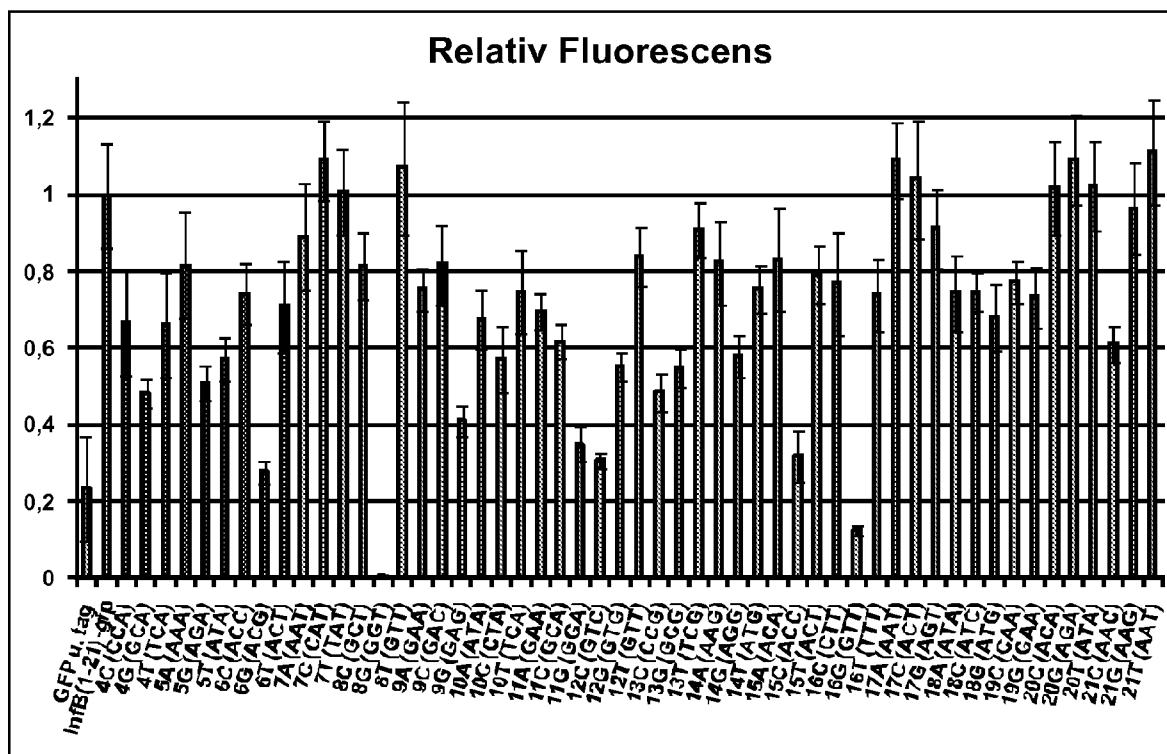
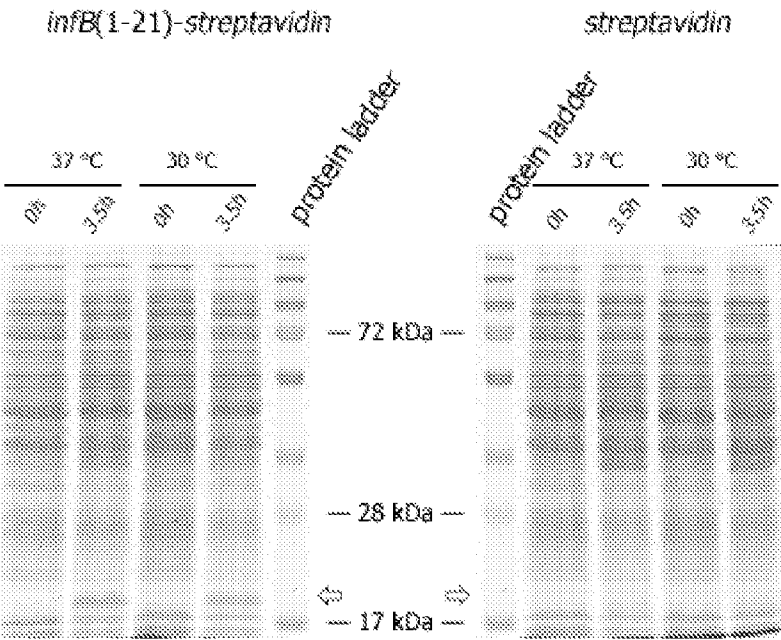


Fig. 28

A



B

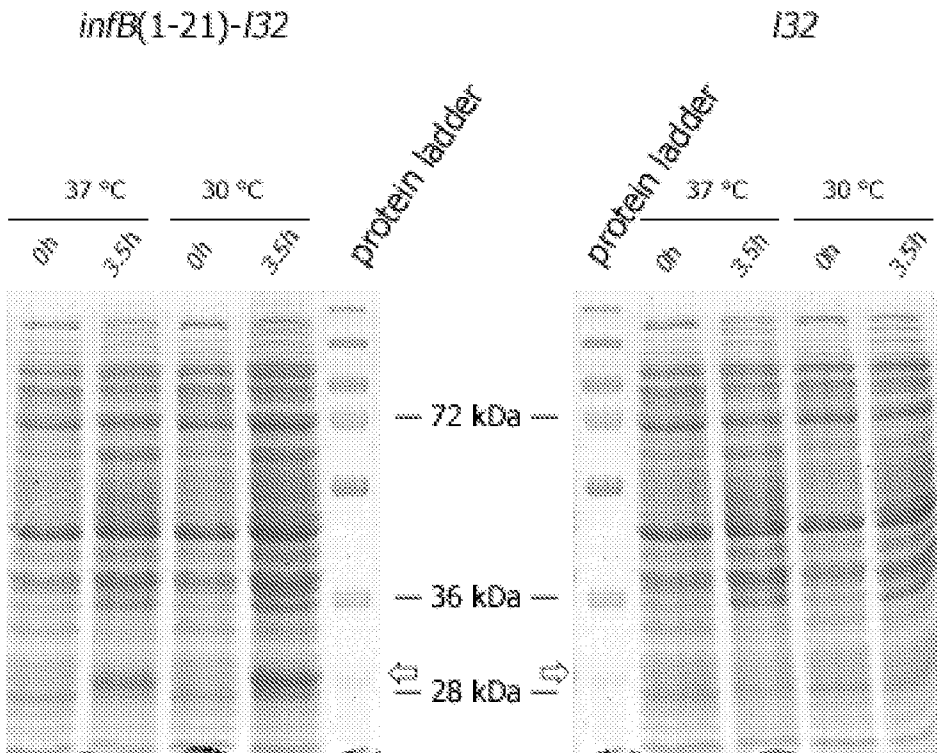


Fig. 29

Primer (restriction site)	Sequence	Use
Forward-primers		
F <i>infB</i>(1-21)-<i>l32</i> (<i>Bsp</i>HI)	5' CCCTCATGACAGATGTAACGATT AAAGCCGAGGTGCAGCTGGTGGAG3'	Amplification of <i>infB</i> (1-21)- <i>l32</i>
F <i>infB</i>(1-21)- <i>streptavidin</i> (<i>Bsp</i>HI)	5' GGGGGTCATGACAGATGTACCGATT AAAGACCCCTCCAAGGACTCGAAG 3'	Amplification of <i>infB</i> (1-21)- <i>streptavidin</i>
F <i>l32</i> (<i>Nco</i>I)	5' CCCC <u>CATG</u> GCCGAGGTGCAGCTGGTG 3'	Amplification of <i>l32</i>
F <i>streptavidin</i> (<i>Nco</i>I)	5' GGGGGCCATGGACCCCTCCAAGGA CTCGAAG 3'	Amplification of <i>streptavidin</i>
Reverse-primer (<i>Eco</i>RI)		
R <i>l32</i> (<i>Eco</i>RI)	5' CCCGAATTCTCATCAGTGGTGGTGGTG GTGGTGACCCAGTACTGTCAGTTTCGT 3'	Amplification of <i>l32</i> and <i>infB</i> (1-21)- <i>l32</i>
R <i>streptavidin</i> (<i>Eco</i>RI)	5' CCCGAATTCTCATCAGTGGTGGTGGTG GTGGTGCTGCTGAACGGCGTCGAGCGG 3'	Amplification of <i>streptavidin</i> and <i>infB</i> (1-21)- <i>streptavidin</i>

Fig. 30

INTERNATIONAL SEARCH REPORT

International application No
PCT/DK2010/050043

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/00 C12N15/62
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, Sequence Search, BIOSIS, EMBASE, FSTA, WPI Data, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SORENSEN H P ET AL: "A favorable solubility partner for the recombinant expression of streptavidin" PROTEIN EXPRESSION AND PURIFICATION, ACADEMIC PRESS, SAN DIEGO, CA LNKD-DOI:10.1016/J.PEP.2003.07.001, vol. 32, no. 2, 1 December 2003 (2003-12-01), pages 252-259, XP004475036 ISSN: 1046-5928	1-4,6-39
Y	the whole document ----- -/--	5

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

12 April 2010

Date of mailing of the international search report

25/05/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Voigt-Ritzer, Heike

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2010/050043

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JANA S ET AL: "Strategies for efficient production of heterologous proteins in Escherichia coli" APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, SPRINGER, BERLIN, DE LNKD- DOI:10.1007/S00253-004-1814-0, vol. 67, no. 3, 1 May 2005 (2005-05-01), pages 289-298, XP019331816 ISSN: 1432-0614	5
A	table 1	1-4,6-39
X	PETI ET AL: "Strategies to maximize heterologous protein expression in Escherichia coli with minimal cost" PROTEIN EXPRESSION AND PURIFICATION, ACADEMIC PRESS, SAN DIEGO, CA LNKD- DOI:10.1016/J.PEP.2006.06.024, vol. 51, no. 1, 30 November 2006 (2006-11-30), pages 1-10, XP005727958 ISSN: 1046-5928	1-3, 16-18, 20, 22-25, 29-34,39
A	the whole document	4-15,19, 21, 26-28, 35-38
A	DATABASE UniProt [Online] 10 February 2009 (2009-02-10), "SubName: Full=Translation initiation factor IF-2;" XP002577094 retrieved from EBI accession no. UNIPROT:B6ZXX7 Database accession no. B6ZXX7 the whole document	1-39