



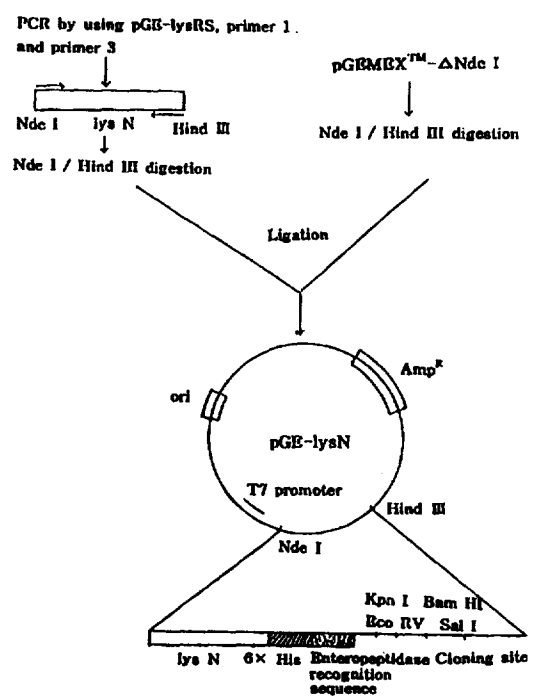
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : C12N 15/52, 15/70, 1/21 // (C12N 1/21, C12R 1:19)</p>	<p>A1</p>	<p>(11) International Publication Number: WO 98/14591 (43) International Publication Date: 9 April 1998 (09.04.98)</p>
<p>(21) International Application Number: PCT/KR97/00186 (22) International Filing Date: 4 October 1997 (04.10.97) (30) Priority Data: 1996/44010 4 October 1996 (04.10.96) KR (71) Applicant (for all designated States except US): HANIL SYNTHETIC FIBER CO., LTD. [KR/KR]; 222, Yangduk-dong, Hoiwon-ku, Masan-si, Kyongsangnam-do 630-490 (KR). (72) Inventors; and (75) Inventors/Applicants (for US only): CHOI, Seong, II [KR/KR]; 461-6, Chumin-dong, Yoosung-gu, Taejun 305-390 (KR). SEONG, Baik, Lin [KR/KR]; 461-6, Chumin-dong, Yoosung-gu, Taejun 305-390 (KR). (74) Agent: LEE, Won-Hee; Chunwoo Building, 5th floor, 736, Yoksam-dong, Kangnam-ku, Seoul 135-080 (KR).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published With international search report.</p>

(54) Title: NOVEL EXPRESSION VECTORS FOR PRODUCTION OF FOREIGN PROTEINS AS SOLUBLE FORMS

(57) Abstract

The present invention relates to novel expression vectors which can produce foreign proteins as soluble forms by using lysyl-tRNA synthetase and a process for preparing foreign proteins by using the expression vectors. Particularly, the present invention relates to the expression vectors which can provide foreign proteins as fused and soluble forms by exploiting the structure and expression pattern of lysyl-tRNA synthetase and the processes for preparing foreign proteins in *E. coli* effectively, which can be utilized industrially to produce active proteins in mass.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

NOVEL EXPRESSION VECTORS FOR PRODUCTION OF FOREIGN PROTEINS AS SOLUBLE FORMS

Field of the Invention

5

The present invention relates to novel expression vectors which can produce foreign proteins as soluble forms by using lysyl tRNA synthetase and a process for preparing foreign proteins by using the expression vectors.

More particularly, the present invention relates to the expression
10 vectors which can provide foreign proteins as fused and soluble forms by exploiting the structure and expression pattern of lysyl-tRNA synthetase and the process for preparing foreign proteins in *E. coli* effectively, which can be utilized industrially to produce active proteins in mass.

15 Background of Invention

With the advance of genetic engineering, heterologous proteins which are used industrially as medicine and the like, have been produced by utilizing animal cells, yeasts and prokaryotes such as *E. coli*. Especially
20 *E. coli* has been exploited as a popular host cell to produce foreign proteins since it grows fast and has been studied more thoroughly than any other

organisms.

Unfortunately, *E. coli* lacks cellular components necessary for posttranslational modification processes like glycosylation, disulfide-crosslinking or the like. And foreign proteins produced massively and excessively in *E. coli* are sequestered into inclusion bodies, which can be easily separated. But in order to obtain active proteins, these inclusion bodies should be solubilized to form primary structure by using high concentration of urea, guanidium HCl or the like and then refolded removing the above reagents.

Generally, the refolding process for preparing a active protein can not be always performed successfully since its result varies according to the cases. For example, proteins having high molecular weight, such as antibodies, tissue plasmingen activator, factor VIII and so on, are not refolded easily to become active proteins. And, it is difficult to produce a recombinant protein on a large scale.

Therefore, it is very important to express foreign proteins as soluble forms in *E. coli* for improving the problems caused in above cases.

Presently, following methods have been exploited to express foreign proteins as soluble forms effectively.

First, there is a method in which N-terminus of foreign protein is linked to signal peptide so as to secrete foreign protein into periplasm of *E.*

coli as a soluble form (Stader, J. A. and Silhavy, T. J., 1970, *Methods in Enzymol.*, 165 : 166-187). Since the foreign proteins are not expressed effectively by the process, this method is not useful industrially.

Second, there is a method in which foreign proteins are expressed
5 with chaperone genes such as groES, groEL, dnaK and the like to obtain soluble proteins (Goloubinoff, P., Gatenby, A. A. and Lorimer, G. H., 1989, *Nature*, 337 : 44-47). But this method is not general to prevent the formation of inclusion body since it is available on only specific proteins.

Third, there is a method in which target proteins are fused at the C-
10 terminus with fusion partner proteins which can be expressed highly in *E. coli*. Since the target proteins are linked at the C-terminus of fusion partners, translation initiation signal of the fusion partner protein can be exploited usefully. And the solubility of the fused foreign protein increases so that large amount of foreign proteins can be obtained as soluble
15 forms in *E. coli*.

Lac Z or Trp E protein have been utilized as a fusion partner protein in order to produce fusion proteins in *E. coli*. But active-form proteins can not be obtained easily since most fusion proteins were
20 expressed in the forms of inclusion body. Therefore, many researches have been accomplished to obtain novel fusion partner proteins which

facilitates the production of active-form proteins. Practically, some fusion partner proteins have been developed, such as glutathione-S-transferase (Smith, D. B. and Johnson, K. S., 1988, *Gene*, 67 : 31-40), maltose-binding protein (Bedouelle, H. and Duplay, P., 1988, *Euro. J. Biochem.*, 171 : 541-549), protein A (Nilsson, B. et al., 1985, *Nucleic Acid Res.*, 13 : 1151-1162), Z domain of protein A (Nilsson, B. et al., 1987, *Prot. Eng.*, 1 : 107-113), protein Z (Nygren, P. A. et al., 1988, *J. Mol. Recog.*, 1 : 69-74) and thioredoxin (Lavallie, E. R. et al., 1993, *Bio/Technology*, 11 : 187-193).

10 Although foreign proteins have been expressed by linking the fusion partner described above and prepared as soluble forms, some were expressed as inclusion body or partly as soluble proteins according to the fusion partner protein.

Particularly, thioredoxin has been known to be the most successful
15 protein as a fusion partner. However, in the case of thioredoxin *E. coli* transformant should be cultured at low temperature such as 15°C in order to express most fusion proteins as soluble forms. Since *E. coli* grows very slowly at that temperature, the process using the thioredoxin may be inefficient.

20

Lysyl-tRNA synthetase (hereinafter it refers to "Lys RS") and its

gene have been investigated as described below, which is preferable for the fusion partner protein and expressed highly in *E. coli*.

Although in *E. coli* aminoacylation is performed by using a specific aminoacyl-tRNA synthetase, two lysyl-tRNA synthetases which are encoded from *lys S* gene and *lys U* gene are involved in the aminoacylation independently. *lys S* gene is expressed constitutively in normal condition and *lys U* gene is induced by heat shock, low pH, anaerobiosis, L-alanine, L-leucine, L-leucyl dipeptide. And amino acid sequences derived from the two genes show 88% of homology.

In addition, the X-ray crystallographical structure of lysyl-tRNA synthetase which is expressed from *lys U* gene (hereinafter it refers to "Lys U") was elucidated at the 2.8 Å resolution level (Onesti, S., Miller, A. D. and Brick, P., 1995, *Structure*, 3 : 163-176). Lys U protein is composed of homodimer which has N-terminal domain contacting with tRNA and C-terminal domain of dimer interface showing the enzyme activity (see Fig. 1).

In addition, nuclear magnetic resonance (NMR) structure of N-terminal domain (31-149 amino acid residues) of lysyl-tRNA synthetase which is expressed from *lys S* gene (hereinafter it refers to "Lys S") was revealed by Frederic Dardel group (Stephane, C. et al., *J. Mol. Biol.*, 253 : 100-113). As Lys U protein and Lys S protein share a high degree of

identity in the amino acid sequences, the N-terminal structures of the two enzymes are identified to be very similar.

In detail, the N-terminal domain of lysyl-tRNA synthetase has secondary structure of five stranded antiparallel β barrel which is composed of α -helix (H4) located between 3rd and 4th β -sheet and contiguous 3 α -helices. The post-part of N-terminal domain corresponds to OB fold (A1A2A3H4A4A5) which is found in proteins binding with oligosaccharides or oligonucleotides commonly. It has been reported that OB fold was discovered in aspartyl-tRNA synthetase of yeast, β -subunit of heat labile enterotoxin, berotoxin and staphylococcal nuclease (Murzin, A. G., 1993, *EMBO J.*, 12 : 861-867).

The N-terminal domain of Lys RS protein of which the structure is described above shows the following characteristics as a fusion partner protein.

When *lys S* gene was expressed in *E. coli*, Lys S protein has accumulated to 80% of total soluble proteins. Since Lys S protein is composed of homodimer of which the contact region is located at the C-terminus of monomer, the fusion protein using intact Lys S protein or C-terminal domain of Lys S protein as a fusion partner makes heterodimer

with Lys S protein of *E. coli*.

But such a heterodimer is fatal to *E. coli*. Thus the C-terminal domain of Lys S protein is not appropriate as a fusion partner protein and only the N-terminal domain can be exploited as a fusion partner protein.

5 Practically, only N-terminal domain of Lys S protein (hereinafter it refers to "Lys N") can be used to express foreign proteins well, to approximately 40% of the total proteins and produced mostly as a soluble form.

As mentioned above, OB fold located in the N-terminal domain of Lys RS protein has a secondary structure which facilitates protein folding
10 and increases the solubility of fusion proteins expressed.

The present inventors have researched to develop a fusion partner protein which is useful to produce heterologous proteins by recombinant DNA technology. Thus we have demonstrated that the N-terminal domain
15 of lysyl-tRNA synthetase can be utilized as a fusion partner protein to produce foreign proteins massively in a soluble form. And by using the lysyl-tRNA synthetase, we have developed novel *E. coli* expression vectors and a process for preparing active foreign proteins effectively.

20 Summary of the Invention

The object of the present invention is to provide expression vectors containing total or part of aminoacyl-tRNA synthetase gene. The aminoacyl-tRNA synthetase gene can be obtained from all kinds of cells.

The expression vectors of the present invention are composed of
5 linker peptide sequence, tag sequence, protease recognition site, restriction enzyme recognition site for inserting foreign gene or the like, in addition to the aminoacyl-tRNA synthetase gene.

In addition, the object of the present invention is to provide the *E. coli* expression vectors containing total or part of lysyl-tRNA synthetase
10 gene. The lysyl-tRNA synthetase gene can be selected among *lys S* gene or *lys U* gene.

Particularly, the present invention provides the expression vector pGE-lysRS containing intact *lys S* gene.

In addition, the object of the present invention is to provide the
15 expression vectors containing the N-terminal domain gene of lysyl-tRNA synthetase.

The present invention provides the expression vectors containing the N-terminal domain of lysyl-tRNA synthetase which is deleted at the amino acid residues 1 to 13. And the present invention also provides the
20 expression vectors containing the N-terminal domain gene of lysyl-tRNA synthetase which is deleted at the amino acid residues 1 to 29.

In addition, the present invention provides the expression vector containing only OB fold gene of lysyl-tRNA synthetase. For the purpose, the expression vectors contain the N-terminal domain gene of lysyl-tRNA synthetase which is deleted at the amino acid residues 1 to 65.

5 Particularly, the present invention provides the *E. coli* expression vector pGE-lysN. *E. coli* HMS 174 strain was transformed by the expression vector pGE-lysN and the transformant has been deposited with Korea Research Institute of Bioscience and Biotechnology, Korea, on September 26, 1997 (accession number : KCTC 0388 BP).

10 The object of the present invention is to provide a process for preparing useful foreign proteins as soluble forms of fusion protein by inserting the foreign genes into the above expression vectors.

Particularly, the present invention provides the expression vector plysN-GMcsf by inserting GMcsf (human granulocyte and macrophage colony stimulating factor) gene into the expression vector pGE-lysN. Host cell was transformed with the expression vector and induced to express GMcsf protein as a fusion protein.

At that time, all kinds of *E. coli* strain can be used, which is appropriate for the expression of the fusion protein. Preferably, *E. coli* HMS 174 strain can be used as a host cell.

Particularly, the present invention provides the expression vector

plysN-Gcsf by inserting Gcsf (human granulocyte colony stimulating factor) gene into the expression vector. By using the above process, Gcsf protein is prepared.

Particularly, the present invention provides the expression vector
5 plysN-TIMP2 by inserting TIMP2 (human tissue inhibitor of
metalloprotease 2) gene into the expression vector. By using the above
process, TIMP2 protein is prepared.

Brief Description of the Drawings

10

Fig. 1 depicts the secondary structure of lysyl-tRNA synthetase (Lys U).

Stick is helix structure and arrow is β -sheet structure.

15 Fig. 2 depicts a strategy for constructing the expression vector pGE-lysRS into which *lys S* gene is inserted.

Fig. 3 depicts the expression of Lys S protein by performing SDS-polyacrylamide gel electrophoresis, which used *E. coli* HMS 174 strain
20 transformed with the expression vector pGE-lysRS of the present invention.

lane 1: standard protein marker;

- lane 2: total proteins of *E. coli* induced for the protein expression ;
- lane 3: total proteins of *E. coli* transformant;
- lane 4: total proteins of *E. coli* transformant induced for the protein
expression ;
- 5 lane 5: supernatant of disrupted *E. coli* induced ;
- lane 6: supernatant of disrupted *E. coli* transformant;
- lane 7: supernatant of disrupted *E. coli* transformant induced;
- lane 8: precipitate of disrupted *E. coli* induced;
- lane 9: precipitate of disrupted *E. coli* transformant;
- 10 lane 10: precipitate of disrupted *E. coli* transformant induced

Fig. 4 depicts a strategy for constructing the expression vector pGE-lysN which uses the N-terminal domain of Lys S protein as a fusion partner protein.

15

Fig. 5 depicts a strategy for constructing the *E. coli* expression vector pLysN-GMcsf which expresses GMcsf protein by using the expression vector pGE-lysN.

20

Fig. 6 depicts the expression of GMcsf protein by performing SDS-polyacrylamide gel electrophoresis, which used *E. coli* HMS 174

strain transformed with the expression vector pLysN-GMcsf of the present invention.

lane 1: standard protein marker;

lane 2: total proteins of *E. coli* transformant;

5 lane 3: total proteins of *E. coli* transformant induced for the expression;

lane 4: precipitate of disrupted *E. coli* transformant;

lane 5: precipitate of disrupted *E. coli* transformant induced;

lane 6: supernatant of disrupted *E. coli* transformant;

10 lane 7: supernatant of disrupted *E. coli* transformant induced;

Fig. 7 depicts the expression of GMcsf protein for comparison by performing SDS-polyacrylamide gel electrophoresis, which used thioredoxin as a fusion partner protein and *E. coli* GI724 strains transformed with the expression vector pTRXFUS-GMcsf and pTRXFUS
15 respectively.

lane 1: standard protein marker;

lane 2: supernatant of disrupted *E. coli*/pTRXFUS-GMcsf transformant induced for the protein expression;

20 lane 3: precipitate of disrupted *E. coli*/pTRXFUS-GMcsf transformant induced;

lane 4: supernatant of disrupted *E. coli*/pTRXFUS transformant
induced;

lane 5: precipitate of disrupted *E. coli*/pTRXFUS transformant
induced;

5

Fig. 8 depicts a strategy for constructing the *E. coli* expression vector pLysN-Gcsf which expresses Gcsf protein by using the expression vector pGE-lysN.

10

Fig. 9 depicts the expression of Gcsf protein by performing SDS-polyacrylamide gel electrophoresis, which used *E. coli* HMS 174 strain transformed with the expression vector pLysN-Gcsf.

lane 1: standard protein marker;

lane 2: total proteins of *E. coli* transformant;

15 lane 3: precipitate of *E. coli* transformant;

lane 4: supernatant of *E. coli* transformant;

lane 5: total proteins of *E. coli* transformant induced for the protein
expression;

lane 6: precipitate of disrupted *E. coli* transformant induced;

20 lane 7: supernatant of disrupted *E. coli* transformant induced

Fig. 10 depicts a strategy for constructing the *E. coli* expression vector pLysN-TIMP2 which expresses TIMP2 protein by using the expression vector pGE-lysN.

5 Fig. 11 depicts the expression of TIMP2 protein by performing SDS-polyacrylamide gel electrophoresis, which used *E. coli* HMS 174 strain transformed with the expression vector pLysN-TIMP2.

lane 1: standard protein marker;

lane 2: total proteins of *E. coli*/pGE-lysN transformant;

10 lane 3: total proteins of *E. coli*/pGE-lysN transformant induced for the protein expression;

lane 4: precipitate of disrupted *E. coli*/pGE-lysN transformant induced;

15 lane 5: supernatant of disrupted *E. coli*/pGE-lysN transformant induced;

lane 6: total proteins of *E. coli* transformant

lane 7 : total proteins of *E. coli* transformant induced;

lane 8 : precipitate of disrupted *E. coli* transformant induced;

lane 9 : supernatant of disrupted *E. coli* transformant induced

20

Description of The Preferred Embodiments

The present invention provides expression vectors which produce
5 useful foreign proteins as soluble forms by exploiting the structural
characteristics of aminoacyl-tRNA synthetase. All kinds of aminoacyl-
tRNA synthetase genes can be used to prepare expression vectors of the
present invention as fusion partner proteins.

The present invention provides expression vectors which use lysyl-
10 tRNA synthetase (Lys RS) which has been studied well as a fusion partner.
At that time, Lys RS protein gene can be selected among *lys S* gene and *lys*
U gene.

Lys RS protein gene can be obtained by performing polymerase
chain reaction (PCR) which utilized *E. coli* chromosomal DNA as a
15 template.

Particularly, *lys S* gene obtained by the above process has been
inserted into the plasmid vector such as pGEMEXTM-1 (Promega) so as to
construct the expression vector pGE-lysRS of the present invention (see Fig.
2). *E. coli* strains proper for the expression have been transformed with
20 the expression vector pGE-lys RS and induced to express Lys RS protein.
As a result, Lys RS protein was expressed well, to 80% of total soluble

proteins of the host cell. Generally *E. coli* transformants are cultured at 37°C in order to express Lys RS protein of the present invention. But soluble proteins are expressed efficiently at low temperature such as 15°C - 30°C which facilitates the increase of the soluble protein ratio.

5

The present invention provides expression vectors which uses the N-terminal domain of Lys RS protein as a fusion partner protein.

In order to produce useful foreign proteins effectively, the expression vector of the present invention contains linker peptide sequence, tag sequence, protease recognition site, restriction enzyme recognition site and so forth selectively, in addition to the N-terminal domain of Lys RS protein. Therefore, fusion proteins expressed by using the expression vectors can be produced as forms of soluble proteins in the host cells and separated easily and only the foreign proteins can be purified by digesting the fusion proteins with specific protease.

Particularly, the N-terminal domain gene of Lys RS protein can be obtained by performing polymerase chain reaction which utilizes the expression vector pGE-lysRS as a template. And the N-terminal domain gene obtained by the above process has been inserted into the plasmid vector pGEMEXTM- Δ NdeI to construct the expression vector pGE-lys N of

the present invention (see Fig. 4).

The *E. coli* HMS 174 strain was transformed by the expression vector pGE-lysN of the present invention and the transformant has been deposited with Korea Research Institute of Bioscience and Biotechnology, Korea, on September 26, 1997 (accession number : KCTC 0388 BP).

The expression vector constructed by the above process has the following characteristics. The expression vector of the present invention contains T7 promoter which regulates transcription of the fusion protein. In addition to T7 promoter, all kinds of promoters which can be used in *E. coli* strains, such as tac promoter, λ pL promoter and the like, is available for the expression vector of the present invention.

The expression vectors of the present invention have been constructed in order to exploit the N-terminal domain of Lys RS protein as a fusion partner protein effectively.

In the N-terminal domain of Lys RS protein, helix 1 structure exists. Since the helix 1 structure is very close to linker peptide, it may prevent enteropeptidase from digesting fusion protein and affect protein folding. In order to provide the suitable expression vector for the production of foreign proteins, helix 1 structure can be removed from the expression vector.

The present invention provides the expression vector removed at the helix 1 structure to prepare foreign proteins more efficiently.

Preferably, the expression vector of the present invention contains the N-terminal domain of Lys RS protein which is deleted at the amino acid residues 1 to 13. Preferably the expression vector also contains the N-terminal domain of LysRS protein which is deleted at the amino acid residues 1 to 29.

In addition, preferably the expression vector of the present invention contains OB fold gene which is involved in folding process of Lys RS protein. Particularly, the expression vector contains the N-terminal domain of Lys RS protein which is deleted at the amino acid residues 1 to 65 corresponding to helix structure 1, 2 and 3. The expression vectors above are suitable for the production of fusion proteins as soluble forms.

The expression vector of the present invention can also contain OB fold domain gene of other proteins in addition to the N-terminal domain gene of Lys RS protein. In detail, OB fold genes found in aspartyl-tRNA synthetase of yeast, B subunit of thermolabile enterotoxin, berotoxin and Staphylococcal nuclease can be utilized for the construction of the expression vector.

The expression vector of the present invention contains linker peptide connecting fusion partner protein and foreign protein. Particularly,

the amino acid residues 147 to 154 of Lys RS protein is used as a linker peptide. This linker peptide is very useful since it is protruded on the protein surface and the length of linker peptide can be controlled according to the foreign proteins expressed. The expression vector can also contain
5 useful linker peptides of other proteins in addition to Lys S protein described above.

The expression vector of the present invention also contains histidine tag of 6 histidine residues after the above linker peptide. This
10 histidine tag enables the fusion proteins expressed with the expression vector to be purified easily. Practically, histidine tagged fusion protein can be separated and purified easily by using nickel chelating column chromatography and the like.

In addition to hisitidine tag, polyarginine or consensus biotinylation
15 sequence can be inserted into the expression vector. Fusion proteins produced by using the above expression vector can be separated and purified from various affinity column chromatographies. The tag sequences described above can be located in any available region of C-terminus or N-terminus of the fusion protein.

20

The expression vector of the present invention contains protease

recognition site in order to separate only foreign protein from fusion protein expressed and purified. In detail, the expression vector of the present invention contains enteropeptidase recognition site (DDDDK sequence) after 6 histidine residues, which enables fusion protein to be separated into
5 fusion partner protein and foreign protein easily. At that time, enteropeptidase digests the C-terminus of the above enteropeptidase recognition site.

In addition, the above protease recognition site can be substituted with thrombin recognition site (LVPRGS sequence) or Xa factor
10 recognition site (IEGR sequence) in order to produce foreign proteins efficiently.

The expression vector of the present invention contains restriction enzyme sites after the above protease recognition site in order to insert
15 foreign protein genes conveniently. In detail, the expression vector pGE-lysN of the present invention contains restriction enzyme recognition sites *KpnI* - *BamHI* - *EcoRI* - *Sall* - *HindIII*. All kinds of restriction recognition sites which is used conveniently in cloning foreign genes can be inserted in addition to the above recognition sites.

20

Various foreign proteins which are expressed as inclusion bodies in

E. coli can be prepared as soluble forms efficiently by using the expression vectors of the present invention.

Particularly, the present invention provides the expression vectors which uses the N-terminal domain of lysyl-tRNA synthetase (Lys N) in order to produce human granulocyte and macrophage colony stimulating factor (GMcsf), human granulocyte colony stimulating factor (Gcsf) and human tissue inhibitor of metalloprotease (TIMP 2) and the like massively.

The present invention constructs the expression vector which produces GMcsf protein as a soluble form by using Lys N protein. In detail, GMcsf gene was obtained by performing polymerase chain reaction which utilized the expression vector pTRXFUS-GMcsf as a template. And the GMcsf gene obtained above has been inserted into the expression vector pGE-lysN to construct the expression vector plysN-GMcsf of the present invention (see Fig. 5).

In order to examine the availability of Lys N as a fusion partner protein, GMcsf protein fused with Lys N protein has been compared with GMcsf protein fused with thioredoxin according to their expression. For the previous comparison, the expression vector pTRXFUS-GMcsf which contains GMcsf gene and thioredoxin gene and produces their fusion protein has been constructed (see Fig. 7).

In addition, the present invention constructs the expression vector

which produces Gcsf protein as a soluble form. In detail, Gcsf gene was obtained by performing polymerase chain reaction which utilized the expression vector pTRXFUS-Gcsf as a template. And the Gcsf gene has been inserted into the expression vector pGE-lysN to construct the
5 expression vector plysN-Gcsf of the present invention (see Fig. 8).

In addition, the present invention constructs the expression vector which produces TIMP 2 protein as a soluble protein. In detail, TIMP 2 gene was obtained by performing polymerase chain reaction which utilized the vecor pGETIMP 2 as a template. And the TIMP 2 gene has been
10 inserted into the expression vector pGE-lys N to construct the expression vector pGElysN-TIMP 2 of the present invention (see Fig. 10).

The *E. coli* strains proper for the expression have been transformed with the above expression vectors. Transformants have been cultured at 37°C and as results foreign proteins fused with Lys N protien, namely, Lys
15 N-GMcsf protein, Lys N-Gcsf protein and Lys N-TIMP 2 protein as soluble forms were expressed at the ratio of 5 - 30% of total soluble proteins (see Fig. 6, Fig. 9 and Fig. 11). On the other hand, when thioredoxin was used as a fusion partner protein, fusion protein was expressed as an inclusion body (see Fig. 7). Therefore, Lys N protein of the present invention is
20 identified to be a more outstanding fusion partner protein than thioredoxin.

Practical and presently preferred embodiments of the present invention are illustrative as shown in the following Examples.

However, it will be appreciated that those skilled in the art, on consideration of this disclosure, may make modification and improvements
5 within the spirit and scope of the present invention.

Examples

<Example 1> Cloning of *lys S* gene and construction of the expression 10 vector pGE-lysRS

In order to clone *lys S* gene which is necessary to construct the expression vector of the present invention, polymerase chain reaction (PCR) was performed, which utilized primer 1 of SEQ ID. NO: 1, primer 2 of SEQ ID. NO: 2 and *E. coli* chromosomal DNA as a template (see Sequence
15 Listing). Amplified *lys S* gene was digested with restriction enzyme *NdeI* and *HindIII*.

For the convenience of cloning process, among two *NdeI* sites of plasmid pGEMEXTM-1 (Promega) *NdeI* site located on DNA sequence 3251 was removed from the plasmid as shown in Fig. 2 to construct the plasmid
20 pGEMEXTM- Δ NdeI. The plasmid pGEMEXTM- Δ NdeI was digested with restriction enzyme *NdeI* and *HindIII*. The plasmid and PCR product

digested above were electrophoresed on 1% agarose gel and the gel which contained the DNA fractions appearing at long wavelenth of UV was cut. Each DNA fraction was eluted from the gel by Jetsorb Kit (GENOMED) and was ligated.

5 As a result, the expression vector pGE-lysRS containing *lys S* gene was constructed.

<Example 2> Expression of Lys S protein

E. coli HMS 174 strain was transformed with the expression vector
10 pGE-lysRS constructed in Example 1. The *E. coli* transformant selected was inoculated into 1.5 ml of LB medium containing ampicillin 100 µg/ml, chloramphenicol 30 µg/ml. The transformant was cultured overnight at 37°C, and the growing culture was again inoculated into 50 ml of LB media. When the concentration of *E. coli* was 0.5 at OD₆₀₀, IPTG was added into
15 the *E. coli* culture in order to induce the expression of protein and again the *E. coli* culture was incubated for more 5 hours. The above culture broth was centrifuged for 10 minutes at 5,000g, and cell pellet was suspended in 10 ml of phosphate buffered saline (PBS) buffer. The cells were disrupted and the crude exrtact prepared in the above process was centrifuged for
20 minutes at 15,000g in order to separate supernatant from precipitate. This

precipitate was again suspended in 10 ml of PBS buffer. 28 μ l of above each sample was mixed with 7 μ l of 5 X SDS loading buffer, and boiled for 5 minutes. 10 μ l of the above mixture was loaded onto 12% SDS-polyacrylamide gel, electrophoresed at 120 V and identified with the protein band by using Coomassie blue dye.

As a result, as is shown in lane 7 of Fig. 3, the expression vector of the present invention expresses Lys S protein highly at the ratio of 80% of total soluble proteins (see Fig. 3).

10 <Example 3> Construction of the expression vector pGE-lysN

In order to construct the expression vector using the N-terminal domain of Lys S protein as a fusion partner protein, polymerase chain reaction was performed, which utilized primer 1 of SEQ ID. NO: 1, primer 3 of SEQ ID. NO: 3 and the expression vector pGE-lysRS constructed in Example 1 as a template (see Sequence Listing).

Amplified gene in the above reaction was digested with restriction enzyme *NdeI* and *HindIII* and the plasmid vector pGEMEXTM- Δ NdeI was also digested with *NdeI* and *HindIII*. And above products were ligated after elution (see Fig. 4).

20 As a result, the expression vector containing the N-terminal domain

gene of Lys S protein was constructed and named as the expression vector pGE-lysN (accession number : KCTC 0388 BP).

**<Example 4> Construction of the expression vector plysN-GMcsf and
5 expression of fusion protein LysN-GMcsf**

In order to express human GMcsf protein as a soluble protein in *E. coli*, which has been expressed independently as an inclusion body in *E. coli*, GMcsf gene was cloned into the expression vector pGE-lysN of the present invention (see Fig. 5).

10 In order to obtain GMcsf gene, PCR was performed by utilizing primer 4 of SEQ ID. NO: 4, primer 5 of SEQ ID. NO: 5 and the expression vector pTRXFUS-GMcsf as a template (see Sequence Listing).

Amplified gene by the above reaction was digested with restriction enzyme *KpnI* and *HindIII* and the expression vector pGE-lysN of the present invention was also digested with *KpnI* and *HindIII*. And the above
15 products were ligated after elution. As a result, the expression vector which produces GMcsf protein fused with LysN protein was constructed and named as the expression vector plysN-GMcsf.

In addition, *E. coli* was transformed with the expression vector
20 plysN-GMcsf. As a result, fusion protein was expressed as is shown in Fig. 6 and the size is 33kDa as is predicted. In addition, most LysN-

GMcsf fusion protein was expressed highly at the ratio of 10% of total soluble proteins (see Fig. 6).

**<Example 5> Construction of the expression vector pTRXFUS-GMcsf
5 and expression of thioredoxin-GMcsf**

In order to examine the availability of Lys N of the present invention as a fusion partner protein, as a control experiment the effect of fusion partner protein, thioredoxin on the expression of GMcsf fusion protein was examined.

10 The expression vector pTRXFUS-GMcsf which expresses GMcsf fusion protein was constructed by subcloning GMcsf gene into *KpnI* and *BamHI* site of the expression vector pTRXFUS using thioredoxin as a fusion partner protein (see Fig. 7).

When the *E. coli* transformed with the expression vector of the
15 present invention was cultured at 37°C, fusion proteins were expressed as inclusion bodies (see Fig. 7, lane 3).

As a result, thioredoxin was less effective than Lys N protein as a fusion partner protein.

20 <Example 6> Construction of the expression vector pLysN-Gcsf and expression of LysN-Gcsf fusion protein

In order to express human Gcsf (granulocyte colony stimulating factor) as a soluble protein, which has been expressed as an inclusion body independently in *E. coli*, Gcsf gene was cloned by performing the same method as Example 4.

5 Polymerase chain reaction was performed by utilizing primer 6 of SEQ ID. NO: 6, primer 7 of SEQ ID. NO : 7 and the plasmid vector pTRXFUS-Gcsf as a template (see Sequence Listing). Gcsf gene amplified by the above reaction was phosphorylated by T4 polynucleotide kinase, and the expression vector pGE-lysN was also digested with *EcoRV*,
10 and then treated by CIP (calf intestine phosphatase). The two resultants were ligated after elution by performing the same method of Example 1. As a result, the expression vector plysN-Gcsf was constructed which expresses Gcsf-LysN fusion protein (see Fig. 8).

In addition, fusion protein was expressed by transforming *E. coli*
15 with the expression vector plysN-Gcsf. As a result, fusion protein was expressed as is shown in Fig. 9, and the size of protein is 36kDa as is predicted. Particularly, Lys-Gcsf fusion protein was expressed as a soluble protein, and occupied 30% of total soluble proteins.

20 **<Example 7> Construction of the expression vector plysN-TIMP2 and expression of fusion protein LysN-TIMP2**

In order to express human TIMP2 (tissue inhibitor of metalloprotease 2) as a soluble protein, which has been expressed as an inclusion body in *E.coli*, TIMP2 gene was inserted into the expression vector pGE-lysN of the present invention.

5 In order to clone TIMP2 gene, polymerase chain reaction was performed by utilizing primer 8 of SEQ ID. NO: 8, primer 9 of SEQ ID. NO: 9 and the plasmid vector pGE-TIMP2 (see Sequence Listing). Amplified TIMP2 gene by the above reaction was digested with restriction enzyme *EcoRV* and *HindIII*, and the expression vector pGE-lysN of the
10 present invention was also digested by *EcoRV* and *HindIII*. Above two resultants were ligated after elution by performing the same method as Example 1. As a result, the expression vector plysN-TIMP2 which expresses fusion protein LysN-TIMP2 was constructed (see Fig. 10).

In addition, fusion protein was expressed by transforming *E. coli*
15 with the expression vector plysN-TIMP2. As a result, fusion protein was expressed as is shown in Fig. 11, and the size of protein is 41kDa as is predicted. Particularly, Lys N-TIMP2 fusion protein was expressed as a soluble protein, to 5% of total soluble proteins (see Fig. 11, lane 9).

20 The expression vectors of the present invention expresses lysyl-tRNA synthetase and foreign proteins fused with the N-terminal domain of

lysyl-tRNA synthetase as soluble forms, which makes their protein activities maintained. Thus the present invention is outstanding in view of recombinant DNA technology.

Practically, the expression vector of the present invention expresses
5 Lys RS protein highly at the ratio of 80% of total soluble proteins, and also expresses foreign proteins fused with Lys N protein highly at the ratio of 5-30%. In addition, Lys N protein is more effective than thioredoxin developed already.

Particularly, the expression vectors of the present invention can
10 produce foreign protein efficiently, for example GMcsf, Gcsf and TIMP2 proteins. In addition to the previous proteins, the expression vector of the present invention is useful to produce foreign proteins which are difficult or impossible to be obtained as active forms and have high molecular weights, such as antibodies, tissue plasminogen activator and factor VIII.

15 In addition, the expression vector of the present invention is constructed to make foreign proteins genes inserted, fusion proteins purified easily and protease recognition site digested specifically, which facilitates the production of intact target proteins. Thus the expression vector is very useful to produce various foreign proteins.

Sequence Listing

(1) General Information

(iii) Number of sequences : 9

5

(2) Information for SEQ ID NO: 1:

(i) Sequence Characteristics :

(A) Length : 36 nucleic acids

(B) Type : nucleic acid

10 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 1

GACTACCATA TGTCTGAACA ACACGCACAG GGCGCT

36

(2) Information for SEQ ID NO: 2:

(i) Sequence Characteristics :

(A) Length : 42 nucleic acids

(B) Type : nucleic acid

5 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 2

GACTACAAGC TTCTATTATT TTACCGGACG CATCGCCGGG AA

42

10

(2) Information for SEQ ID NO: 3:

(i) Sequence Characteristics :

(A) Length : 96 nucleic acids

(B) Type : nucleic acid

5 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 3

10 GACTACAAGC TTGTCGACGA TATCGGATCC GGTACCCTTG TCATCGTCAT CGTGGTGGTG

60

GTGGTGGTGC GGCAGCGGAC GCAGTGCTTT GGTCAG

96

15

20

25

(2) Information for SEQ ID NO: 4:

(i) Sequence Characteristics :

(A) Length : 33 nucleic acids

(B) Type : nucleic acid

5 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 4

GACAAGGGTA CCGCACCCCG CTCGCCAGC CCC

33

(2) Information for SEQ ID NO: 5:

(i) Sequence Characteristics :

(E) Length : 33 nucleic acids

(F) Type : nucleic acid

5 (G) Strandedness : single

(H) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 5

GAGCGCAAGC TTCACTCCT GGACTGGCTC CCAGCA

(2) Information for SEQ ID NO: 6:

(i) Sequence Characteristics :

(A) Length : 33 nucleic acids

(B) Type : nucleic acid

5 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 6

GACAAGGGTA CCAACCCCCT GGGCCCTGCC AGC

33

(2) Information for SEQ ID NO: 7:

(i) Sequence Characteristics :

(E) Length : 36 nucleic acids

(F) Type : nucleic acid

5 (G) Strandedness : single

(H) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 7

GACAAGAAGC TTTCATCAGG GCTGGGCAAG GTGGCG

36

(2) Information for SEQ ID NO: 8:

(i) Sequence Characteristics :

(I) Length : 33 nucleic acids

(J) Type : nucleic acid

5 (K) Strandedness : single

(L) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 8

GTCATCGATA TCTGCAGCTG CTCCTCCGGTG CAC

33

(2) Information for SEQ ID NO: 9:

(i) Sequence Characteristics :

(A) Length : 36 nucleic acids

(B) Type : nucleic acid

5 (C) Strandedness : single

(D) Topology : linear

(ii) Molecular type : oligonucleotide

(xi) Sequence Description : SEQ ID NO: 9

GTCATCAAGC TTTCATTATG GGTCCTCGAT GTCGAG

36

10

What is claimed is :

1. A expression vector which contains aminoacyl-tRNA synthetase gene and linker peptide sequence, protease recognition site, tag
5 sequence or restriction enzyme recognition site.
2. The expression vector according to Claim 1, wherein aminoacyl-tRNA synthetase gene is *E. coli* aminoacyl-tRNA synthetase gene.
- 10 3. The expression vector according to Claim 2, wherein aminoacyl-tRNA gene is *lys S* gene of lysyl-tRNA synthetase.
4. The expression vector pGE-lysRS according to Claim 3.
- 15 5. The expression vector according to Claim 3, which contains a part of *lys S* gene encoding the N-terminal domain of lysyl-tRNA synthetase.
6. The expression vector according to Claim 5, which contains the
20 N-terminal domain gene deleted at the amino acid residues 1 to 13.

7. The expression vector according to Claim 5, which contains the N-terminal domain gene deleted at the amino acid residues 1 to 29.
8. The expression vector according to Claim 5, wherein the N-terminal domain gene is OB fold gene.
9. The expression vector according to Claim 8, wherein the OB fold gene is the N-terminal domain gene deleted at the amino acids residues 1 to 65.
10. The expression vector pGE-lysN according to Claim 5.
11. A *E. coli* transformant which is prepared by transforming *E. coli* HMS 174 strain with the expression vector of Claim 10 (accession number : KCTC 0388 BP).
12. A expression vector which is prepared by inserting a foreign protein gene into the expression vector of Claim 1.
13. The expression vector plysN-GMcsf according to Claim 12, wherein the foreign protein is GMcsf protein.

14. The expression vector p_{lysN}-Gcsf according to Claim 12, wherein the foreign protein is Gcsf protein.
15. The expression vector p_{lysN}-TIMP2 according to Claim 12,
5 wherein the foreign protein is TIMP2 protein.
16. The transformant which is prepared by transforming host cell with the expression vector of Claim 12.
- 10 17. A process for preparing foreign proteins as soluble forms, wherein the transformant of Claim 16 is cultured and induced for the expression of protein.

Fig. 1

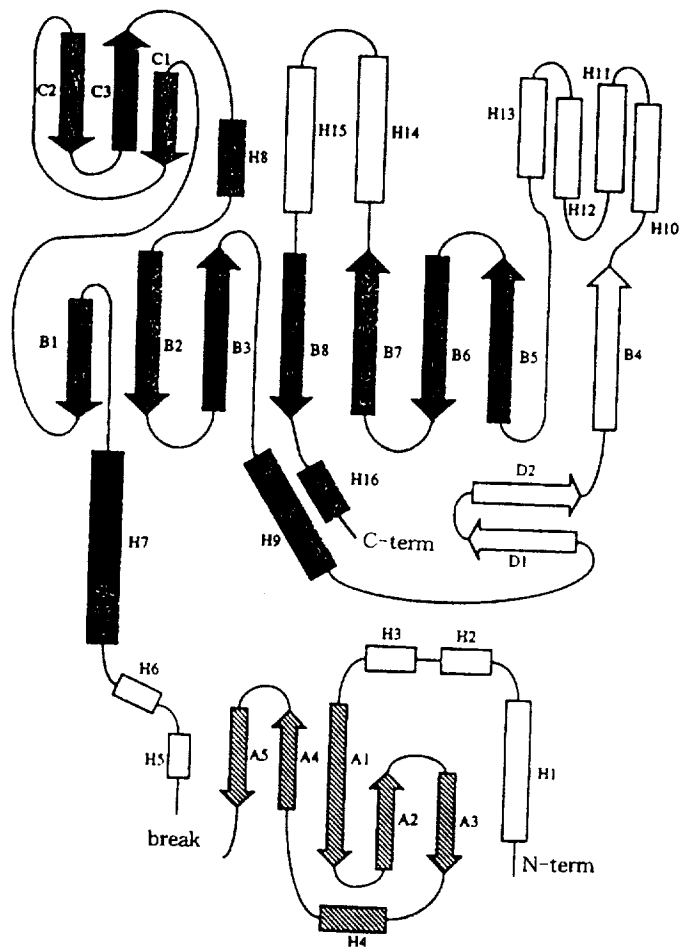


Fig. 2

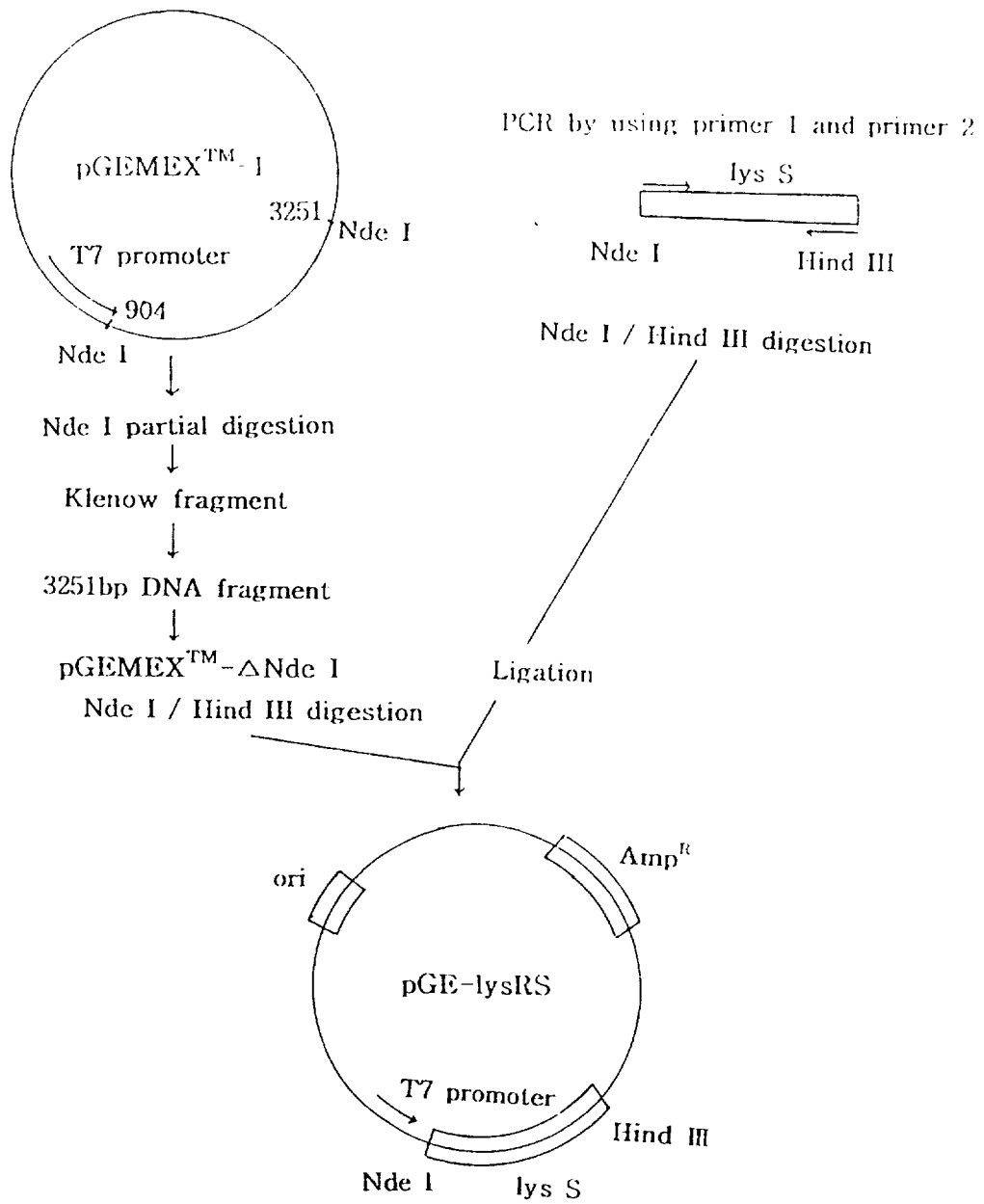


Fig. 3

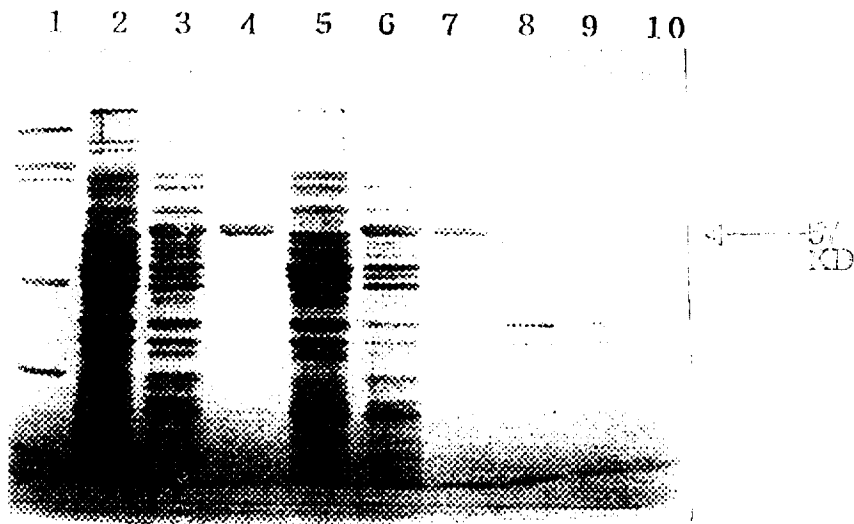


Fig. 4

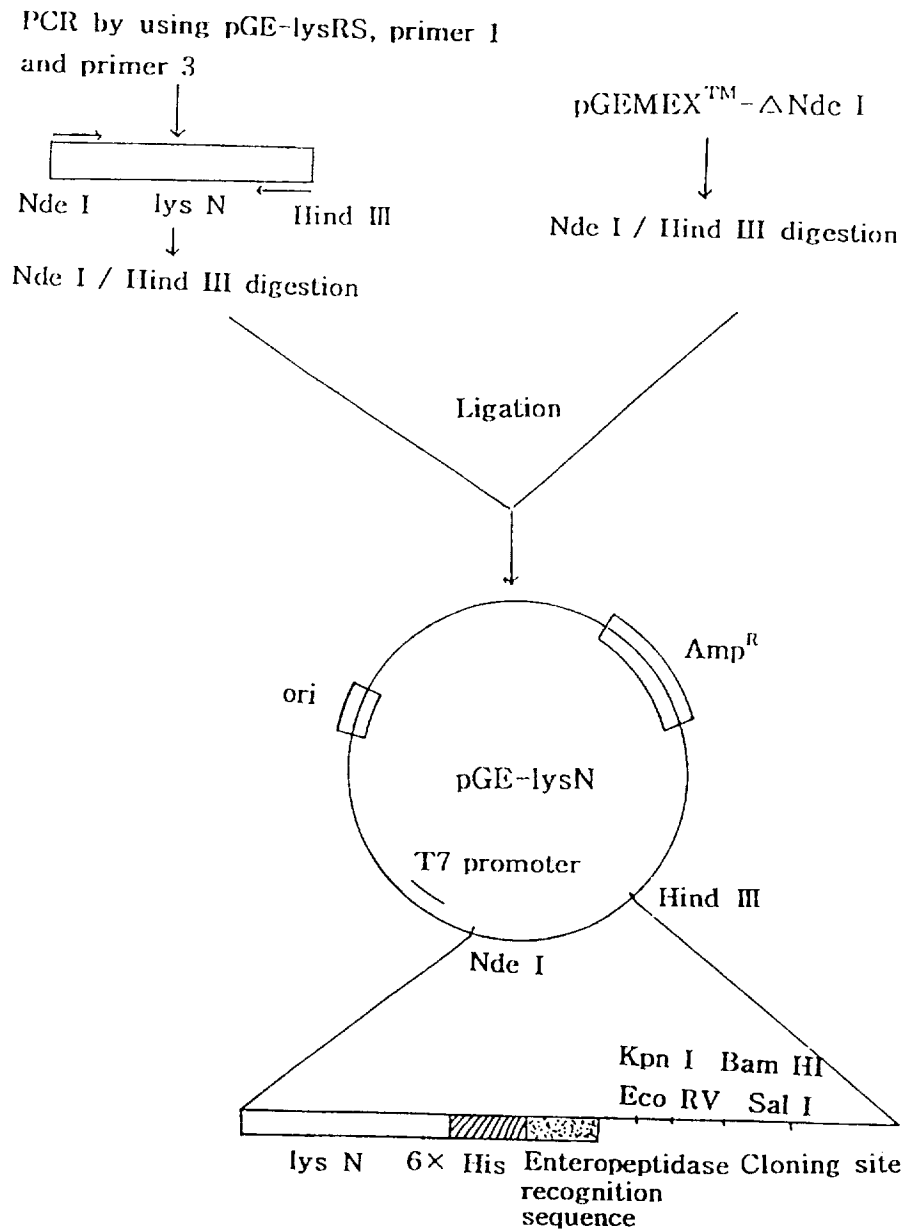


Fig. 5

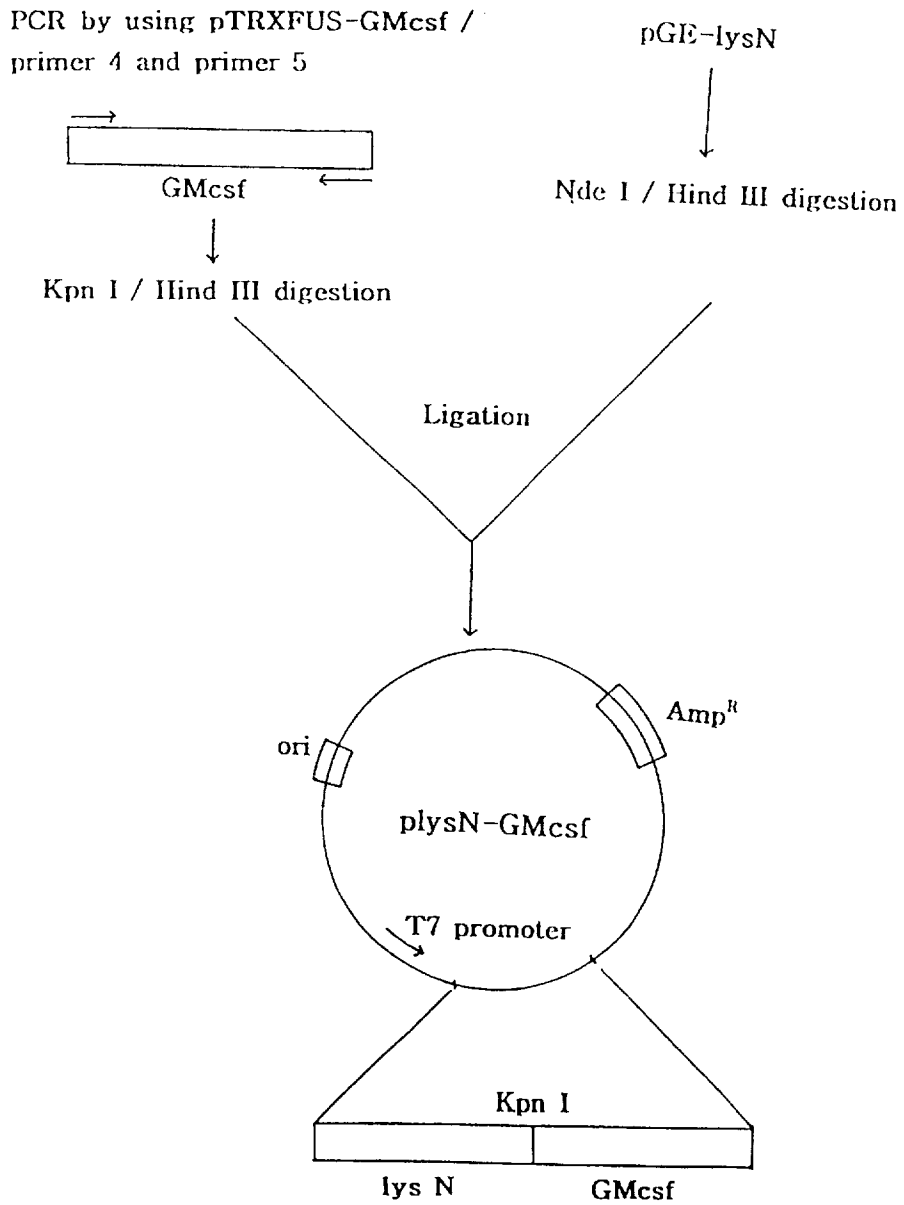


Fig. 6

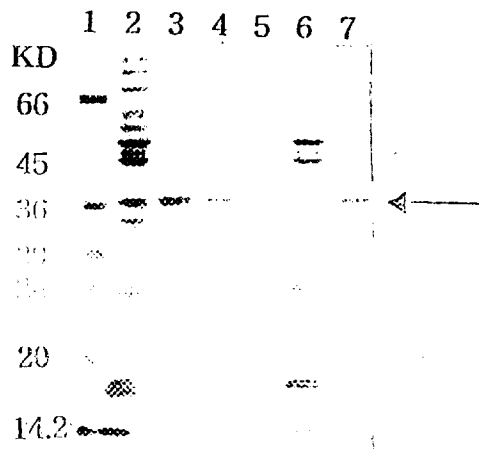


Fig. 7

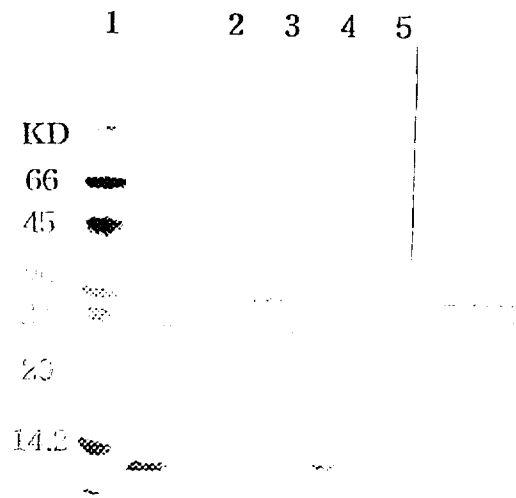


Fig. 8

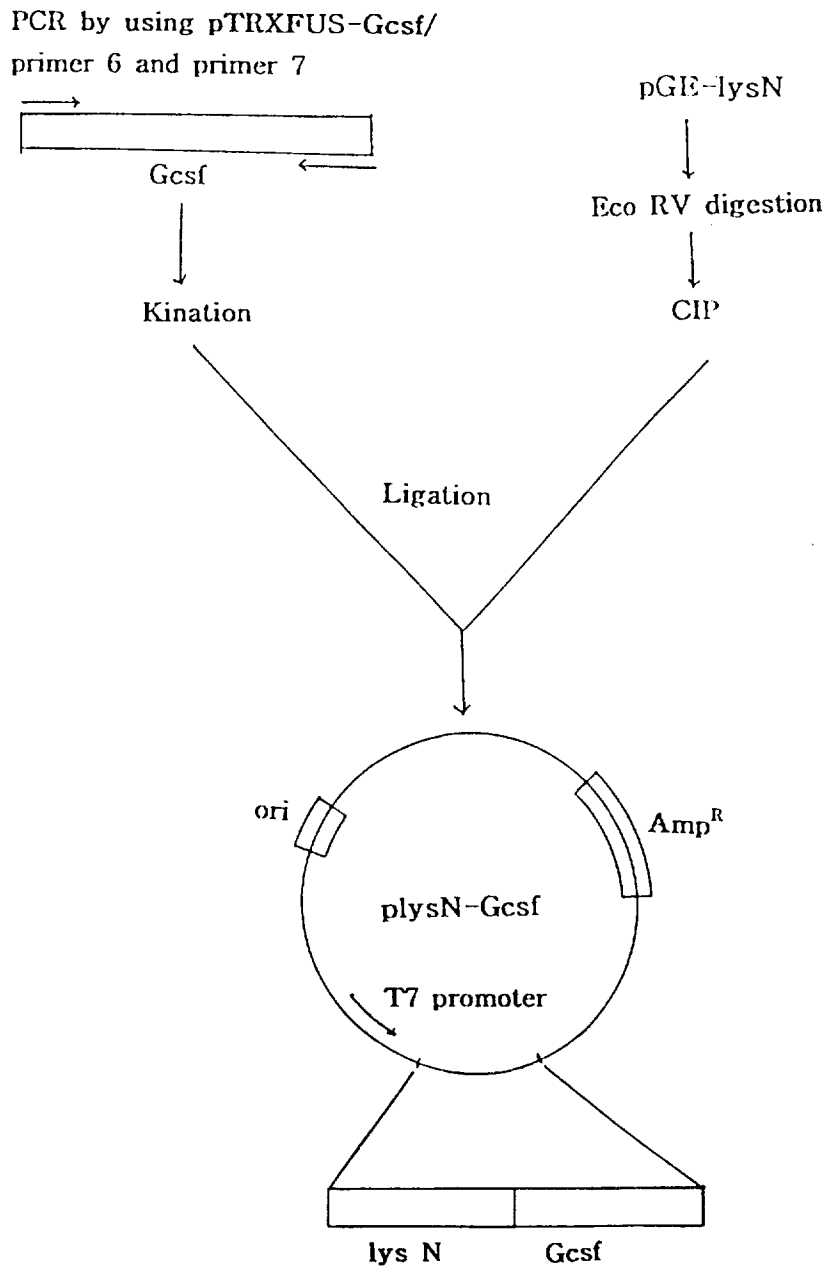


Fig. 9

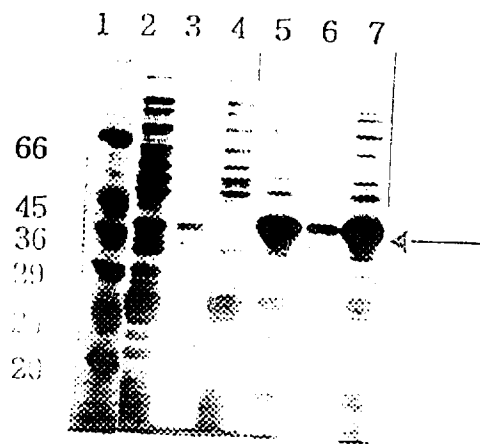


Fig. 10

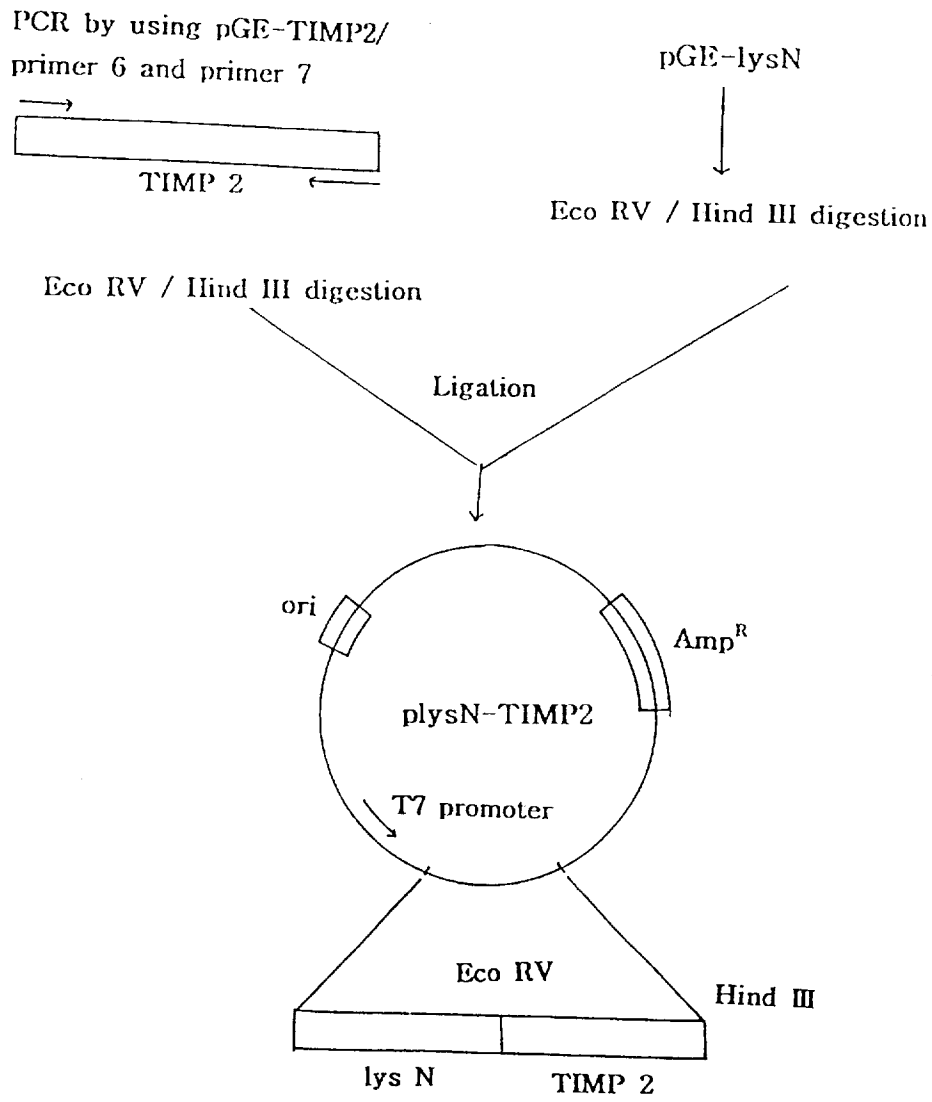
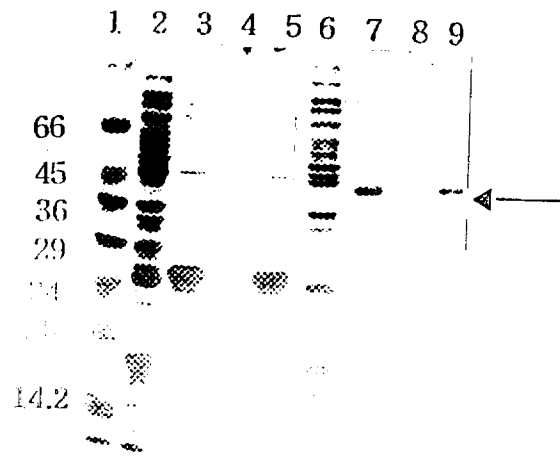


Fig. 11



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR 97/00186

A. CLASSIFICATION OF SUBJECT MATTER

IPC⁶: C 12 N 15/52, 15/70, 1/21 // (C 12 N 1/21; C 12 R 1:19)

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)

IPC⁶: C 12 N 15/52, 15/70, 1/21

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 practicable, search terms used)

WPIL, CAS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOURNAL OF MOLECULAR BIOLOGY, Vol.253, No.1, 13 October 1995 (13.10.95), S. COMMANS et al.: "Solution Structure of the Anticodon-binding Domain of Escherichia coli Lysyl-tRNA Synthetase and Studies of its Interaction with tRNA-Lys", pages 100-113; abstract.	1-3
A	EP 0 247 819 A2 (UNITIKA LTD.) 02 December 1987 (02.12.87), abstract.	1

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

16 December 1997 (16.12.97)

Date of mailing of the international search report

05 January 1998 (05.01.98)

Name and mailing address of the ISA/ AT

AUSTRIAN PATENT OFFICE
Kohlmarkt 8-10
A-1014 Vienna
Facsimile No. 1/53424/535

Authorized officer

Wolf

Telephone No. 1/53424/436

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/KR 97/00186

Im Recherchenbericht angeführtes Patentedokument Patent document cited in search report Document de brevet cité dans le rapport de recherche	Datum der Veröffentlichung Publication date Date de publication	Mitglied(er) der Patentfamilie Patent family member(s) Membre(s) de la famille de brevets	Datum der Veröffentlichung Publication date Date de publication
EP A2 247819	02-12-87	DE C0 3771577 EP A3 247819 EP B1 247819 JP A2 62278992 JP B4 5073392 US A 4886749	29-08-91 15-02-89 24-07-91 03-12-87 14-10-93 12-12-89