(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 July 2005 (21.07.2005)

PCT

(10) International Publication Number
WO 2005/066344 A2

(51) International Patent Classification[7]: C12N 15/10

(21) International Application Number:
PCT/PL2005/000004

(22) International Filing Date: 10 January 2005 (10.01.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
P.364295 9 January 2004 (09.01.2004) PL

(71) Applicant *(for all designated States except US)*: **INSTYTUT BIOTECHNOLOGII I ANTYBIOTYKOW** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PLUCIENNICZAK, Andrzej** [PL/PL]; ul.Prochowa 14/5, PL-04-360 Warszawa (PL). **SMORAWINSKA, Maria, Ludwika** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **WOLINOWSKA, Renata** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **MIKIEWICZ-SYGULA, Diana** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **JAGIELLO, Agata** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **KACZANOWSKI, Radoslaw** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **CHOJNACKA, Luiza** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **PLUCIENNICZAK, Grazyna** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **ZIELINSKA, Ewa** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **STRZEZEK, Krystyna** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL). **MARCINIAK-RUSEK, Alina** [PL/PL]; ul. Staroscinska 5, PL-02-516 Warszawa (PL).

(74) Agent: **WITEK, Rafal**; ul. Rudolfa Weigla 12, PL-53-114 Wroclaw (PL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NEW PLASMIDS, THEIR DERIVATIVES AND FRAGMENTS, THEIR METHODS OF MANUFACTURE AND APPLICATION

(57) Abstract: A method is described of obtaining a plasmid contained in a difficult to separate, heterogenous plasmid DNA fraction, particularly one containing various plasmids of similar sizes. Another aspect of the present invention are new plasmids obtained via the method according to the present invention, their derivatives and/or fragments, as well as the application of these products in biotechnology and medicine, particularly gene therapy.

WO 2005/066344 A2

New plasmids, their derivatives and fragments, their methods of manufacture and application

The present invention relates to the method of obtaining a plasmid contained in a difficult to separate heterogeneous plasmid DNA fraction, particularly one containing plasmids of similar sizes. Another aspect of the present invention are new plasmids obtained using the method according to the present invention, their derivatives and fragments, and the application of these products in biotechnology and medicine, particularly in gene therapy.

In modern biotechnology, we assume that a DNA fragment (gene) we introduce into a cell, eg. *Escherichia coli*, will be expressed, but an unprotected foreign DNA fragment would shortly be degraded into nucleotides. DNA fragments coding sequences of biologically significant proteins such as insulin, interferon or growth hormone are introduced into host cells via a vector. Vectors are DNA molecules which are able to replicate autonomously in certain types of cells, and which ensure the amplification of the introduced DNA fragment, and in many cases the efficient expression of genes.

Most often, vectors are derivatives of naturally occurring plasmids. However, to turn the latter into useful plasmid vectors, a series of modifications need to be introduced into them. It is necessary to equip them with a marker, such as a gene or genes responsible for easily discerned phenotypic characteristics such as antibiotic resistance, and also to maximally reduce its molecular mass (the smaller the vector, the larger its "capacity" and ease of manipulation). It is also necessary to introduce a single restriction site of a given type which will be used for cloning, or to remove excess ones (two or more).

The nucleotide sequence should be known in its entirety, so that it may be discerned whether it contains known, or related, genes which may pose a threat to the health or even lives of people, plants or animals. Knowing the sequence also allows one to remove undesirable nucleotide sequences or add desirable ones such as immunogenic sequences in DNA vaccines . It has been determined that plasmid DNA causes the strongest immune response when CG sequences abut two purine bases (adenine or guanine) at the "C" side and two pyrimidines (thymine or cytosine) on the "G" side as described by Roman et al., Nat.Med. 8:849-854, schematically represented by RRCGYY, and herewith indicated as

imm1. A particular case of this sequence which has a particularly strong immunogenic effect is imm2, which has two thymine bases at the "G" end.

Furthermore, CG units in bacterial plasmids are not methylated, whereas in vertebrates they usually are. It has been hypothesized that vertebrate organisms recognize a large frequency of unmethylated CG units as a danger signal, hence the amplified immune response. In gene therapy such an amplified immune response is undesirable, because it may lead to the destruction of the therapeutic protein and the plasmid vector. This is why it is preferable to use plasmids with low GC pair content in gene therapy.

In biotechnology, the most useful vectors are the so-called expression vectors, which facilitate efficient synthesis of the proteins encoded by genes contained on the vector. Such vectors bear promoter sequences which facilitate transcription and translation, and sequences ensuring the stability of the synthesized protein. There are expression vectors known under the control of strong promoters, whose synthesis can lead to accumulations of a given protein totalling 30% or even more of total cellular protein. Such vectors have been used for years in the production of many well known and useful proteins, particularly ones with desirable pharmacological properties.

It is known that certain compound segments of DNA, called transposons, are able to place themselves in many portions of the host genome. This means that certain large DNA segments, known as insertion sequences (IS's), can, if they are located nearby, transpose themselves as a larger unit encompassing genes located between them. Complex units of this type form the transposon. They are found in *Prokaryota*, such as bacteria, but also in *Eukaryota*.

Recently, many useful bacterial transposons have been discovered and characterised, among them one called Tn5. It is a 5.8 kilobase pair (kbp) segment of bacterial DNA, which can undergo insertion in many places in the chromosome, in plasmids, as well as in "latent" phages of gram-negative bacteria. It codes a bacterial resistance gene to aminoglycoside antibiotics, kanamycin and neomycin, as well as gentamycin resistance (G418) in eukaryotic cells. A restriction map of Tn5 was presented by Berg et al., Genetics 105, 813-828 (1983). Further information regarding the technology of Tn5 is contained in the review articles according to Berg and Berg, Bio/Technology 1, 417-435 (1983); and

Berg and Berg, in Neidhardt et al., (ed.), "*Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology" ASM, Washington, D.C., Chapter 63, p. 1,071-1,109 (1987).

Starting with naturally occurring transposons, a series of artificial constructs have been made, meant for use in genetic engineering. For example, patent description US5137829 presents a new DNA transposone derived from transposone Tn5 useful in the production of mutants and rapid screening of long DNA sequences. The transposone described encompasses a partial Tn5 transposone sequence with oligonucleotide SP6 and T7 phage primers, conveniently placed near opposite ends of the Tn5 transposone in question.

Classical methods used to isolate plasmid DNA, commonly used in laboratory practice are usually based on electrophoretic techniques. Despite their many pluses, which made them so popular, these techniques also do have certain limitations, particularly concerning their resolution. They are not suitable for separating DNA fragments of similar sizes, especially when the mixture being separated contains decidedly varied amounts of the individual fragments. In this case, both fragments migrate through the electrophoretic gel as one band, and the presence of the less numerous fragment is masked by the more numerous fragment. In the light of the above described in the state of the art, several technical problems may objectively be described, which are still awaiting resolution.

It is worthwhile then to seek out natural plasmids with specific and useful properties from bacterial strains isolated from various sources. To realise this goal, it is desirable to obtain an easy method of isolating a plasmid contained in a heterogeneous, not easily separable fraction of plasmid DNA, particularly one containing various plasmids of similar sizes.

It is particularly desirable to obtain new plasmids which could be used to produce new constructs useful in biotechnology, especially ones facilitating stable or regulated expression of desired proteins. In this context it is particularly desirable to obtain autonomic functional elements which could be used in the production of other, useful constructs. For example, it is still desirable to produce transcription regulatory elements, like strong transcriptional promoters.

Furthermore, it is also desirable to obtain new plasmids, which would contain a decreased number of immunogenic sequences. Such plasmids could be a convenient source of various constructs designed for medical use, particularly gene therapy.

The above described problems have been solved in the present invention.

The subject of the present invention is a method for obtaining a plasmid contained in a difficult to separate, heterogeneous fraction of plasmid DNA, particularly one containing various plasmids of similar sizes characterised in that the fraction of plasmid DNA contained in the total DNA of the source organism comes into contact with the enzyme transposase and the DNA of a transposone containing a selection marker gene; a microbiological host cell is transformed with a plasmid containing the transposone; the transformants obtained is grown under selection conditions matched to the selection marker used, and the plasmid containing the transposone is isolated from the amplified host cells.

Preferably, the structure of the DNA containing the transposone is additionally analysed in order to establish the characteristics of the obtained plasmid. Preferably, the plasmid DNA fraction contains plasmids of similar sizes, present in significantly different quantities. Preferably, plasmid DNA containing the transposone is further modified. Preferentially, at least one of the following modifications is made: a restriction site is deleted or added, a selection marker gene is introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced. Equally preferentially, the multi-copy plasmid masking the presence of another plasmid is the plasmid pIGRK. Preferably, the plasmid present in lower quantities, masked by the presence of another plasmid is plasmid pIGMS31.

The next subject of the present invention is new plasmid isolated from the plasmid DNA fraction using the method according to invention, as defined above, its derivative or fragment.

The next subject of the present invention is plasmid pIGRK, its derivative or fragment.

Preferably, a plasmid according to invention contains at least one of the following modification: a restriction site is deleted or added, a selection marker gene is introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced, or any other arbitrary mutation. Equally preferentially, it is a plasmid selected from the

following group: pIGRKKAN, pIGRKKANde, pIGRKCM, pIGRKKANT7, pIGRKKhGH or one of their derivatives. Preferably, a fragment of the plasmid according to invention is an autonomous functional element, preferentially selected from the following group: transcription regulatory sequences, replication origins, and sequences coding reading frames. Preferentially, it is a promoter. Preferably, it contains the nucleotide sequence, positions 1240 to 1367 of the plasmid pIGRKKAN sequence, its derivative or fragment.

The next subject of the present invention is plasmid pIGMS31 its derivative or fragment. Preferentially, it contains at least one of the following modifications: a restriction site is deleted or added, a selection marker gene is introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced, or any other known mutation. Equally preferentially, it is a plasmid selected from the following group: pIGMS31KAN, pIGMS31KANT7, pIGMS31PR, pIGMS31PRH or one of their derivatives.

The next subject of the present invention is an application of the pIGRK or pIGMS31 plasmids or one of their derivatives or fragments according to invention, as defined above, in the expression of a sequence coding a polypeptide, or the regulation of such a process. Preferentially, the sequence coding the polypeptide codes a heterologous protein. Preferentially, plasmid selected from the group containing pIGMS31PR, pIGMS31PRH, pIGRKKhGH or one of their derivatives is used to produce human growth hormone or any heterologous protein.

The next subject of the present invention is an application of the nucleotide sequence containing the sequence from between positions 1240 to 1367 of the pIGRKKAN plasmid sequence, its derivative or fragment, to obtain promoter sequences.

The next subject of the present invention is applications of plasmids pIGRK or pIGMS31 or one of their derivatives or fragments according to invention, as defined above, to obtain plasmids meant for use in medicine, particularly gene therapy. Preferentially, the plasmids obtained possess lowered numbers of immunogenic sequences. Preferentially, the plasmids obtained possess less than 20 immunogenic sequences, preferentially less than 10, and most preferentially no more than 5 such sequences.

To better illustrate the nature of the present invention, its description has been supplemented with a detailed discussion of example embodiments and examples located

further in the description. It is not, however, the intention of the applicant to limit the scope of the present invention solely to the presented embodiments nor to the content of the examples below. Based on the revelation of the nature of the present invention stemming from this description in conjunction with commonly accessible knowledge, a specialist will be able to prepare other variants encompassed by the protection defined in the claims.

A series of bacterial strains were analysed, originating from strains collected from patients at The Children's Memorial Health Institute (CZD-Centrum Zdrowia Dziecka), during the search for new plasmids potentially useful in the construction of industrial or therapeutic expression vectors.

Using agarose gel electrophoresis, the presence of a plasmid DNA fraction was observed in a clinical strain of *Klebsiella pneumoniae*, No. 287-w in the IBA system (No. 2324 in the CZD numeration). The electrophoresis image indicated the presence of one, multicopy plasmid (ca. 2500bp) as well as a few-copy one, four times larger than the former (ca. 10000bp). Unexpectedly, application of the method according to the present invention facilitated the discernment of two plasmids of similar size (ca. 2500 base pairs) in a strain 287-w of *K. pneumoniae*, namely plasmids pIGRK and pIGMS31. The present invention thus provides an easy method of isolating a plasmid contained in a difficult to separate, heterogenous fraction of plasmid DNA, particularly one containing plasmids of similar sizes. This is testament to its superiority over other known methods of isolating plasmids, in particular electrophoretic methods.

Furthermore, the present invention provides new plasmids, obtained using the method according to the present invention, their derivatives and useful fragments.

The new plasmids and their derivatives may be used in the production of various constructs meant for biotechnology. In one embodiment, they may facilitate the stable and/or regulated, efficient expression of desired proteins. Plasmids according to the present invention may also be used as regulatory elements in more complex expression systems. For example, a plasmid according to the present invention or its derivative may contain regulatory elements able to influence the expression of proteins coded by another plasmid (eg. a vector other than pIG), but located within the same cell.

Plasmids according to the present invention, and their preferential derivatives, contain far fewer immunogenic sequences, which makes them better vectors or vector fragments for use in medicine, especially gene therapy. For example, when compared to other popular plasmids used as vectors, pIGRKKAN and pIGMS31KAN contain 5 such undesirable immunogenic sequences each, whereas the popular plasmid, pBR322 contains 38 of these sequences, pACYC184 contains 33, and pUC18 contains 20 undesirable immunogenic sequences.

Plasmids obtained using a method according to the present invention may serve as sources of other, useful functional elements, useful in the production of other constructs. For example, a new transcription regulation element is described, with the characteristics of a strong transcriptional promoter.

To better illustrate the content of the description, it has been supplemented with figures.

Figure 1 and figure 2 represent the restriction map and nucleotide sequence respectively (sequence No. 1 in the sequence list) of the plasmid vector pIGRKKAN. The immunogenic sequences *imm*1 and *imm*2 have been indicated on the map. Sequence from 1169 to 1199 encompass a multicloning site. Arrows indicate open reading frames for proteins coded by the plasmid. The frame located between nucleotides 87 and 959 codes for an aminoglycoside phosphotransferase, a protein involved in kanamycin resistance. The functions of the remaining open reading frames are unknown.

Figure 3 and figure 4 represent the restriction map and nucleotide sequence respectively (sequence No. 2 in the sequence list) of the plasmid vector pIGMS31KAN. The immunogenic sequences *imm*1 and *imm2* have been indicated on the map. Sequence from 3651 to 3723 encompass a multicloning site. Arrows indicate open reading frames for proteins coded by the plasmid. The reading frame located between nucleotides 2667 do 3479 codes for an aminoglycoside phosphotransferase, a protein involved in kanamycin resistance. The functions of the remaining open reading frames are unknown.

Figure 5 represents a comparison of the sequence of the vector pIGRKKAN containing a nucleotide sequence recognized by the restriction endonuclease *Nde*I with the same region in the vector pIGRKANde in which it has been removed. Two nucleotide bases, T and A,

were removed from the plasmid vector which formed the so-called sticky ends of the sequence recognized by the restrictase *Nde*I.

Figure 6 and figure 7 represent the restriction map and nucleotide sequence respectively (sequence No. 3 in the sequence list) of the plasmid pIGRKCM. The immunogenic sequences imm1 and imm2 have been indicated on the map. Arrows indicate open reading frames for proteins coded by the plasmid. The reading frame located between nucleotides 1639 and 2236 codes for chloramphenicol acetyltransferase (CAT) warranting chloramphenicol resistance. The functions of the remaining open reading frames are unknown.

Figure 8 and figure 9 represent the restriction map and nucleotide sequence respectively (sequence No. 4 in the sequence list) of the expression vector pIGRKANT7. Nucleotides between the *Xba*I and *Pst*I restriction sites form the „promoter-terminator" region containing the transcription promoter and terminator for a gene coding the RNA polymerase in the bacteriophage T7 genome. The frame located between nucleotides 147 and 960 codes for an aminoglycoside phosphotransferase, a protein involved in kanamycin resistance. The functions of the remaining open reading frames are unknown.

Figure 10 and figure 11 represent the restriction map and nucleotide sequence respectively (sequence No. 5 in the sequence list) of the expression vector pIGRKKhGH. Nucleotides between the *Bam*HI (1169) and *Nde*I (2112) contain the promoter regions P1 and P2 of the *E. coli deo* operon (Fischer and Short Gene. 17, 291-298(1982) . The arrow from nucleotide 2113 (ATG) to 2916 (TAA) encompasses the sequence of a synthetic ubiquitin gene and human growth hormone. The arrow from nucleotide 147 to nucleotide 960 encompasses a kanamycin resistance gene.

Figure 12 and figure 13 represent the restriction map and nucleotide sequence respectively (sequence No. 6 in the sequence list) of the plasmid pIGMS31PRH. The promoter sequence is found between nucleotides 3699-3898. The arrow from nucleotide 3918 (ATG) to 4721 (TAA) encompasses the sequence of a synthetic ubiquitin gene and human growth hormone.

Figure 14 represents the electrophoretic analysis of cell lysates in a 15% polyacrylamide gel. 1. molecular mass marker (97,0, 66,0, 45,0, 30,0 20,1 14,4 kDa). 2. Host strain *E.coli*

DH5α. 3. The *E.coli* DH5α strain transformed with plasmid pIGMS31PR. 4. The *E.coli* DH5α strain transformed with plasmid pIGMS31PRH. The arrow indicates the location of the fusion protein.

Example 1. Insertion of the EZ::TN™<kan-2> transposone into plasmids from *K. pneumoniae* strain 287-w.

Strain 287-w is resistant to several antibiotics. It has proved impossible to ascertain whether the genes responsible for antibiotic resistance are found on one of the plasmids. Thus the plasmids isolated from the *K. pneumoniae* strain did not possess a selection marker, one of the basic features of every vector. The modified transposone Tn5, EZ::TN™<kan-2>, bearing a kanamycin resistance gene was inserted into plasmids isolated from the *K. pneumoniae* strain 287-w with the aid of the enzyme transposase. In addition to the kanamycin resistance gene, the modified transposone EZ::TN™<kan-2> (Epicentre Technologies, Madison WI, USA) contains a grouping of sites recognized by enzymes often used in cloning, the so-called multicloning site. In addition to the transposon, the kit from Epicentre Technologies contains the transposase enzyme and primers for cloning the plasmid with the inserted transposon. The in vitro conditions have been selected by the producer so as to maximize the frequency of single insertions and to minimize multiple insertions transposone molecules into individual plasmid molecules. *E.coli* DH5α cells with the kanamycin transposone were selected with LB medium containing (50μg/ml).

Example 2. Nucleotide sequences of plasmids from *K. pneumoniae* strain 287-w.

Sequencing of the plasmids surrounding the inserted kanamycin transposone showed that there were two plasmids of a similar size present (ca. 2500 base pairs) in *K.pneumoniae* strain 287-w. Thus the insertion of the kanamycin transposone allows one to identify different plasmids which are not identifiable electrophoretically due to an insignificant size difference and the masking effect of the multi-copy plasmid over the low copy number plasmid. Plasmids containing the kanamycin plasmid can be transformed into known strains of *E. coli* using kanamycin as a selection factor. In this way, *E. coli* strains are obtained containing single plasmids, which facilitates the amplification of their DNA for use in broadly understood cloning.

Both plasmids have been fully sequenced.

The restriction map and nucleotide sequence of the smaller plasmid, symbol pIGRKKAN, are presented in figures 1 and 2.

The restriction map and nucleotide sequence of the larger plasmid, symbol pIGMS31KAN are presented in figures 3 and 4.

The sequence of the plasmid pIGRKKAN contains 3578 base pairs with a GC content of 33,4%. The plasmid sequence contains two *imm*1 and three *imm2* sequences. 2 sequences recognized by the restriction enzyme *SwaI* are found in the plasmid, which recognizes an 8-nucleotide sequence statistically occurring once per 40 000 base pairs in the *E. coli* genome. Nucleotides 1-9 and 1241-1249 are so-called direct repeat sequences, of which one is additionally formed at the transposone insertion site. Nucleotides from 10 to 1230 correspond to the 1221 base pairs of the inserted kanamycin transposone EZ::TN™<KAN-2>. Nucleotides from 1240 to 3578 correspond to the sequence of the naturally occurring plasmid in *K. pneumoniae* strain 287-w. The sequence of the naturally occurring plasmid pIGRK is 2338 base pairs long.

Using BLASTN software, the sequence of the pIGRK plasmid was compared to the NCBI database. The homology of pIGRK to sequences found in the said database pertains to short DNA lengths, which allows one to state that this is a new, undescribed plasmid basing on a classification based on replicon sequence probability as described by as described by Couturier et al., Microbiol. Rev.52, 375-395 (1988).

. Using the BLASTX program, the translated amino acid sequence of the pIGRK plasmid was compared with a protein sequence database. The post-translation amino acid sequences from the plasmid open reading frames of plasmid pIGRK have very little homology to proteins found in the SwissProt databases. These homologies are so low that it may be stated that the proteins coded for by pIGRK are unlike any other proteins.

The sequence of plasmid pIGMS31KAN is 3750 base pairs long with a GC content of 36,8%. The plasmid contains four *imm*1 and one *imm*2 sequences. 3 *Swa*I restriction enzyme recognition sequences are found in the plasmid, which recognizes an 8-nucleotide sequence statistically occurring once per 40 000 base pairs in the *E. coli* genome. Nucleotides 1-9 and 2521-2529 are so-called direct repeat sequences of which one is

additionally formed through a 9 base pair repeat at the transposone insertion site. Nucleotides 2530-3750 correspond to the 1221 base pairs of the inserted kanamycin transposone EZ::TN™<KAN-2>. The sequence of the naturally occuring *K. pneumoniae* strain 287-w plasmid is contained between nucleotides 1 and 2520. Nucleotides 2521 to 2529 are a simple repeat of the sequence 1-9, which is formed at the transposone insertion site. The sequence of the naturally occurring plasmid pIGMS31 is 2520 base pairs long.

Using BLASTN software, the sequence of the pIGMS31 plasmid was compared to the NCBI database. The homology of pIGMS31 to sequences found in the said database pertains to short DNA lengths, which allows one to state that this is a new, undescribed plasmid basing on a classification based on replicon sequence probability as described by Couturier et al., Microbiol. Rev.52, 375-395 (1988).

Using the BLASTX program, the translated amino acid sequence of the pIGMS31 plasmid was compared with a protein sequence database. The largest homology, 36%, of the amino acid sequence pertained solely to short lengths of the translated amino acid sequence. These insignificant homologies pertain to proteins taking part in the recombination process and replication initiation. These homologies are small enough that they also allow one to state that pIGMS31 is a new, undescribed plasmid.

The cryptic plasmids from *K. pneumoniae* strain 287-w were used in a series of transformations and modifications in order to obtain new vectors.

Both plasmids pIGRKKAN and pIGMS31KAN may be used as vectors because they contain the so-called multicloning site, meaning a site with sequences recognized by restriction enzymes most commonly used in cloning. The kanamycin resistance gene, isolated from transposone Tn903 also gives neomycin resistance in *E. coli*. The examples below are of modifications of the plasmids pIGRKKAN and pIGMS31KAN.

Example 3. Deletion of the *Nde*I restriction site.

Plasmid pIGRKAN contains one *Nde*I restriction site. It is a commonly used restriction site in cloning, because it contains an ATG sequence, the methionine codon which is the translation origin for most proteins. The restriction site was removed using limited digestion with Mung Bean Nuclease. The effectiveness of the nuclease activity was assayed using *Nde*I digestion and through sequencing of the region which contained the

recognition site for this enzyme. The sequence of the portion of plasmid pIGRKKAN from which the NdeI restriction site was removed and designated as pIGRKKANde is presented in figure 5. Two T and A bases were removed, which formed the so-called sticky ends following *NdeI* digestion, thus giving the plasmid resistance to this enzyme. The hypothetical reading frame was altered, which did not visibly influence the plasmid's level of replication.

Example 4. Modification of the plasmid pIGRK through the insertion of a DNA fragment bearing chloramphenicol resistance.

A DNA fragment containing the chloramphenicol acetyltransferase (CAT) gene was inserted into plasmid pIGRK. Plasmid pIGRK, digested with the restriction enzyme *AsuII*, was ligated with a 952 base pair DNA fragment containing the CAT gene also digested with *AsuII*, from the pBW4 plasmid as described by Mikiewicz at al., Plasmid 38, 210-219(1997). Plasmid pIGRKCM, 3298 base pairs long, bearing the chloramphenicol resistance gene is represented in figures 6 and 7. Such a plasmid can be used in the production of an expression vector.

Example 5. Construction of an expression vector containing the T7 bacteriophage transcription promoter and terminator.

Two expression vectors were constructed containing a transcription promoter and terminator for the gene coding the RNA polymerase in the T7 bacteriophage genome, based on the plasmids pIGMS31KAN and pIGRKKAN. The restriction maps are represented in figures 8 and 9, respectively. Plasmid pIGMS31KAN and its derivative pIGMS31KANT7 are low-copy number plasmids, in contrast with plasmids pIGRKKAN and pIGRKKANT7, which occur in large numbers of copies in a bacterial cell.

Example 6. Using plasmid pIGMS31KAN to express the human growth hormone gene.

The promoter sequence of retron Ec86 described by Lim and Maas Cell **56**: 891-904 and its following transcription terminator sequence were inserted into plasmid pIGMS31KAN. The restriction sites for *BamHI* and *HindIII* were used for this. The plasmid formed was designated pIGMS31PR. Plasmid pIGMS31PR was used to clone a gene coding a fusion protein composed of yeast ubiquitin and human growth hormone. The cloning made use of *NdeI* and *Sal*I restriction sites. The full sequence of the derivative plasmid, pIGMS31PRH

is presented in figures 12 and 13. The plasmid was used to transform cells of *E. coli* strain DH5α. Electrophoretic analysis of cell lysates showed the presence of a protein of a size corresponding to the fusion protein ubiquitin-growth hormone (figure 14).

Example 7. Using plasmid pIGRKKAN to express the gene coding human growth hormone.

DNA fragments containing the promoters P1 and P2 from the *deo* operon of *E. coli* strain K-12 as described by Fischer and Short Gene **17**:291-298, the growth hormone and modified ubiquitin gene fusion and the transcription terminator sequence were cloned into plasmid pIGRKKAN. The restriction map of the recombined plasmid designated pIGRKKhGH is presented in figures 10 and 11. The selection factor in the expression vector is the kanamycin resistance gene from the commercial transposone EZ::TN™<KAN-2>. The plasmid was used to transform cells of *E. coli* strain DH5α. Electrophoretic analysis of lysates indicated the presence of a protein nearly identical in size with a UBI-hGH marker.

Example 8. Using the pIGRKKAN plasmid promoter in the expression of the gene coding human growth factor in the plasmid pIGMS31KAN.

The region of plasmid pIGRKKAN contained between nucleotides 1240 and 1367 was transferred to the polylinker sequence of plasmid pIGMS31KAN. The gene coding the ubiquitin and human growth factor fusion protein and transcription terminator was placed downstream of this sequence. Electrophoretic analysis of cell lysates of *E. coli* DH5α cells transformed with this plasmid indicated the presence of considerable quantities of a protein corresponding in size to the fusion protein. This means that the region of plasmid pIGRKKAN contained between nucleotides 1240 and 1367 contains a sequence which functions as a very efficient transcription promoter. The region described contains the polypurine (AGGAGG) Shine-Dalgarno sequence between nucleotides 1356-1361 in close proximity to the ATG codon at the start of one of the reading frames in plasmid pIGRKAN.

Literature

1.Berg,D.E., and Berg,C.M.(1983). The prokaryotic transposable element Tn5. Biotechnology 1, 417-435.

2.Berg,D.E., Schmandt, M.A., and Lowe J.B.(1983) Specificity of transposon Tn5 Insertion.Genetics 105(4), 813-828

3. Berg and Berg, in Neidhardt et al., (ed.), "*Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology" ASM, Washington, D.C., Chapter 63, p. 1,071-1,109 (1987).

3. Couturier, M., Bex, F., Bergquist, P. L., and Maas, W. K., (1988) Identification and classification of bacterial plasmids. Microbiol. Rev., 52, 375-395.

4. Fischer M., and Short S.A. 1982. The cloning of the *Escherichia coli* deoxyribonucleoside operon. Gene. 17 291-298).

5. Lim D., Maas W.K. (1989) Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E.coli* B. Cell 56, 891-904.

6. Mikiewicz, D., Wróbel, B., Węgrzyn, G., and Płucienniczak, A. (1997) Isolation and and characterization of a ColE1-like plasmid from *Enterobacter agglomerans* with a novel variant of *rom* gene. Plasmid, 38, 210-219.

7. Roman, M., Martin-Orozoco, E., Goodman,J.S., Nguyen, M.D., Sato, Y., Ronaghy, A., Kornbluth, R.S., Richman, D. D., Carson, D. A., and Raz, E. (1997). Immunostimulatory DNA sequences function as T helper-1-promoting adjuvants. Nat. Med. 8, 849-854.

Claims

1. A Method for obtaining a plasmid contained in a difficult to separate, heterogenous fraction of plasmid DNA, particularly one containing various plasmids of similar sizes characterised in that the fraction of plasmid DNA contained in the total DNA of the source organism comes into contact with the enzyme transposase and the DNA of a transposone containing a selection marker gene; a microbiological host cell is transfected with a plasmid containing the transposon; the transformant obtained is grown under selection conditions matched to the selection marker used, and the plasmid containing the transposone is isolated from the amplified host cells.

2. A method according to Claim 1, characterised in that the structure of the DNA containing the transposone is additionally analysed in order to establish the characteristics of the obtained plasmid.

3. A method according to Claim 1, characterised in that the plasmid DNA fraction contains plasmids of similar sizes, present in significantly different quantities.

4. A method according to Claim 1, characterised in that DNA isolated from the plasmid containing the transposone is further modified.

5. A method according to Claim 4, characterised in that at least one of the following modifications is made: a restriction site is deleted or added, a selection marker gene is introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced.

6. A method according to Claim 3, characterised in that the multi-copy plasmid masking the presence of another plasmid is the plasmid pIGRK.

7. A method according to Claim 3 or 6, characterised in that the plasmid present in lower quantities, masked by the presence of another plasmid is plasmid pIGRMS31.

8. A new plasmid isolated from the plasmid DNA fraction using the method according to Claims 1 to 7, its derivative or fragment.

9. The plasmid pIGRK, its derivative or fragment.

10. A plasmid according to Claim 9, characterised in that it contains at least one of the following modification: a restriction site is deleted or added, a selection marker gene is

introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced, or any other arbitrary mutation.

11. A plasmid according to Claim 9 or 10, characterised in that it is a plasmid selected from the following group: pIGRKKAN, pIGRKKANde, pIGRKCM, pIGRKKANT7, pIGRKKhGH or one of their derivatives.

12. A fragment according to Claim 9, characterised in that it is an autonomous functional element, preferentially selected from the following group: transcription regulatory sequences, replication origins, and sequences coding reading frames.

13. A fragment according to Claim 9 or 12, characterised in that it is a promoter.

14. A fragment according to Claim 13, characterised in that it contains the nucleotide sequence, positions 1240 to 1367 of the plasmid pIGRKKAN sequence, its derivative or fragment.

15. Plasmid pIGMS31 its derivative or fragment.

16. A plasmid according to Claim 15, characterised in that it contains at least one of the following modifications: a restriction site is deleted or added, a selection marker gene is introduced, a gene coding a foreign protein is introduced, and/or a regulatory sequence is introduced, or any other known mutation.

17. A plasmid according to Claim 15 or 16, characterised in that it is a plasmid selected from the following group: pIGMS31KAN, pIGMS31KANT7, pIGMS31PR, pIGMS31PRH or one of their derivatives.

18. An application of the pIGRK or pIGMS31 plasmids or one of their derivatives or fragments according to one of Claims 8 to 17 in the expression of a sequence coding a polypeptide, or the regulation of such a process.

19. An application according to Claim 18, characterised in that the sequence coding the polypeptide codes a heterologous protein.

20. An application according to Claim 18 or 19, characterised in that the plasmid selected from the group containing pIGMS31PR, pIGMS31PRH, pIGRKKhGH or one of their derivatives is used to produce human growth hormone or any heterologous protein.

21. An application of the nucleotide sequence containing the sequence from between positions 1240 to 1367 of the pIGRKKAN plasmid sequence, its derivative or fragment, to obtain promoter sequences.

22. Applications of plasmids pIGRK or pIGMS31 or one of their derivatives or fragments according to one of Claims 8 to 17 to obtain plasmids meant for use in medicine, particularly gene therapy.

23. An application according to Claim 22, characterised in that the plasmids obtained possess lowered numbers of immunogenic sequences.

24. An application according to Claim 23, characterised in that the plasmids obtained possess less than 20 immunogenic sequences, preferentially less than 10, and most preferentially no more than 5 such sequences.

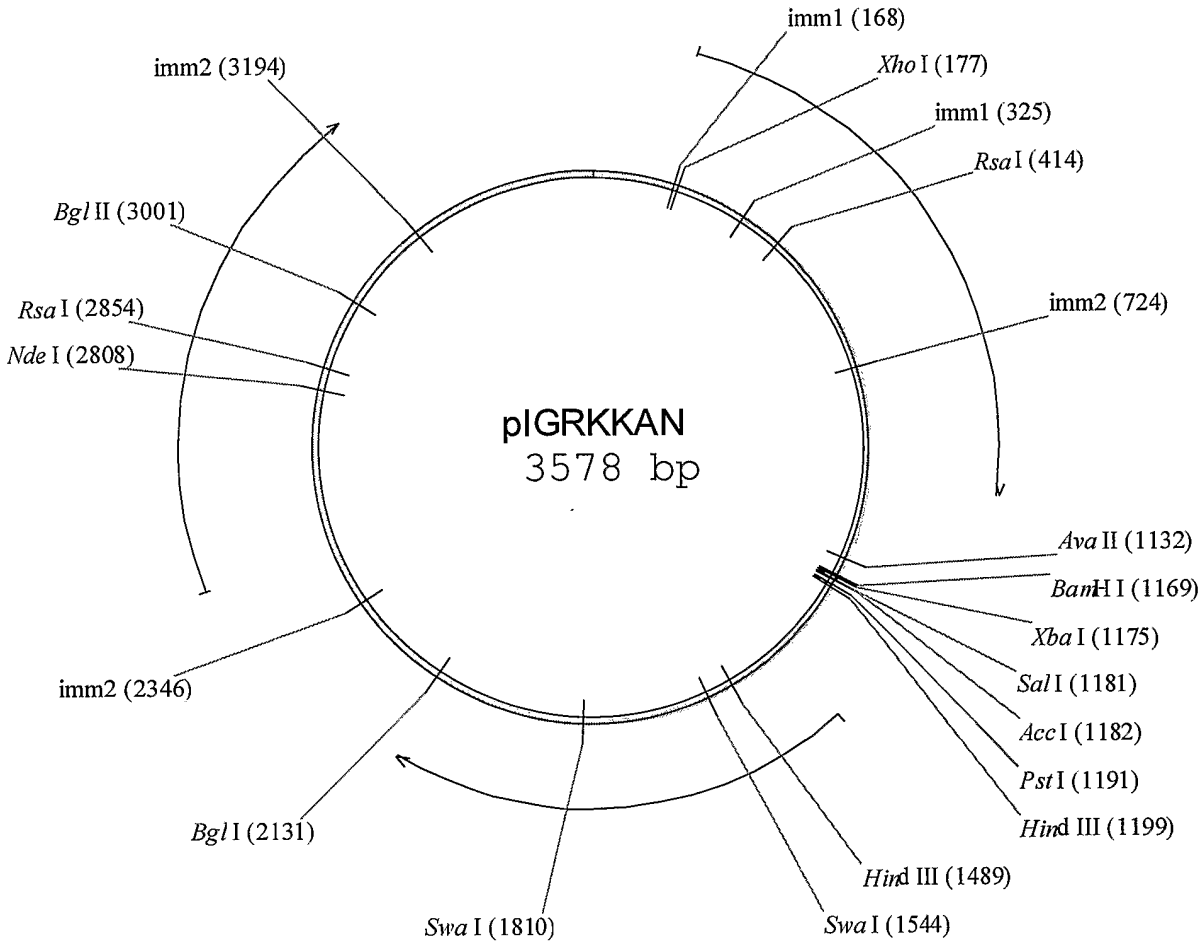

imm1 (168)
Xho I (177)
imm1 (325)
Rsa I (414)
imm2 (724)
Ava II (1132)
BamH I (1169)
Xba I (1175)
Sal I (1181)
Acc I (1182)
Pst I (1191)
Hind III (1199)
Hind III (1489)
Swa I (1544)
Swa I (1810)
Bgl I (2131)
imm2 (2346)
Nde I (2808)
Rsa I (2854)
Bgl II (3001)
imm2 (3194)
pIGRKKAN
3578 bp

Fig. 1

actctagccctgtctcttatacacatctcaaccatcatcgatgaattgtgtct
caaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaat
aaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattc
aacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgat
ttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaat
ctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggca
aaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctg
acggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatga
tgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattag
aagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctg
cgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgt
atttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcga
gtgatttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaa
atgcataaacttttgccattctcaccggattcagtcgtcactcatggtgattt
ctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatg
ttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaac
tgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatgg
tattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagt
ttttctaatcagaattggttaattggttgtaacactggcagagcattacgctg
acttgacgggacggcggctttgttgaataaatcgaacttttgctgagttgaag
gatcagatcacgcatcttcccgacaacgcagaccgttccgtggcaaagcaaaa
gttcaaaatcaccaactggtccacctacaacaaagctctcatcaaccgtggcg
gggatcctctagagtcgacctgcaggcatgcaagcttcagggttgagatgtgt
ataagagacagactctagccagtttccaagtagaaactacagtttctaaactg
caacttttttctacttttttgcaacttaatctattgactagtcctttataaatgt
taaaacatatatatagaaataaataaaaagaggaggtttctatggatattgga
aatatattaaatgagagtttaagtattgattacgaaaaattagatttgttttt
ggaaaaatatgatttaacaccagaacaaaaagttgcagtttatgaatttcacg
caaaagcttataaaaaaaataaaactttagttatttctgaaacaaaagaaaat
aaatttaaatctatttccgaaggtgttgaatacgtgcatttattcccaaaaaa
tttaaaaatttttaattaaaaaatatggtttaaatacaaacgaattattggttt
taacggaaataatggagtcaatgctttcacacggaaatttattaattaatttt
tcgcaaaaggcactttgcgaattaacaggaattaataaatctacaatgtgtaa
aacatttaaaaccctcaaacaaaagcagtgtttaattgagaaaaacggacata
tttatttaaattctgtgatatttatgaaagggttacctcataaattgtttatg
caatttagagatcatttttttaaattctatctcatataaattagatgatgaaga
agaatttgaaaaagtcttcgacgataatttttattaaagcatacgaaaaaaatc
tcaaagagattaaaaagaaaaagcaacaaataaaagaaaagaaaatatcaaaa
gcattagataatttttgaaaaagaaatctcgaaagaatggaaggaaaagtttaa
agacgaagaggaaaatttcgaatttggttttgaatcggaaatataaaaccgcc
ctcgccgggcaggcgaatcccttattgaaatagaataaattctattccactaa
gggatttttttattcattgtttctccacatttgcaatattgacattaacttc

Fig. 2a

cacccggatataacagtagtataagttgttgtttcaacccgtcttttgggtg
gaacaacaaggcattttagggatagagcaaagcgaaggccataaaattgccac
ccccaaccgggggtcgttgttcgatttgagcgatagcgaaaaattgaacataa
ggggggagggtttgggttttacggtatttcaaatttgagcaaagcgaattttt
gaaatttccggttcttttaatttgcaatgaggaaaaatcaatatgggtaattc
aaaaagaaatataaaaaaaactaaatgataattttagagaggatattttagatt
atgcgatcgcgcacaatctaaaatgtgctaacgcacttgctattttatacgca
acgggttgccgtccggacgaactccaaaccggagttactgtaaactatgacag
taaaaaaaatgaaattgaatttagaataattggatcaaaactaaatagaagaa
tgagaagaggcataggggttagaaaaataaaagtaaaaatcaataatgaaaat
gccaggtttttaaaaacattgttgataaatttattgaaaacccaatgtcata
tgatcacaaaatcaaaattgaaagtgccaaagcattttccgggtacataacaa
aaatatcgaaaaagctatggcccaggaaaacctatcatgcttctgcatattct
tttagacatgcaaaagcaacggaattaaaaaattccgattatgataaaatcga
aatagctcagattatgggccatgcctcagttagatctcagcagagttacggaa
gaaagagcaaaaaaagcaaaggtggatttgatgacatcgcagatgtcgaaacc
aatgttaaaccccgtggcggtgatagattattgagatttaagatcgcaaataa
aaacaaagcagcggcaaaaattgccgatacttccacccccagcagtcctccac
cggctcccgttcgtcgcttcaaaatgtgaaccgtgagcagttcaggaggttcc
ctcctggactgtgaagggttggcccgtccggtcaggacggttttacagcaaaa
tcctccatagcgaagcagaagcccggaacgggtaactggatggttttccccccg
tgggggattgatctgttacttgaaaaccaatgatcttaaaagccatctcaaaa
gttgaaaatttcacccccttagtgttcttaaaattcttagatgttcttaggag
ttaaaaaactactctctaaccattgatattactggatttttaaaaaaggcagt
tgtcaaaaacttcaaccgtagttgtcaaattcgtcaactccagttgtcaaatt
cgtcaactgaggttgtcaaatccgaca

Fig. 2b

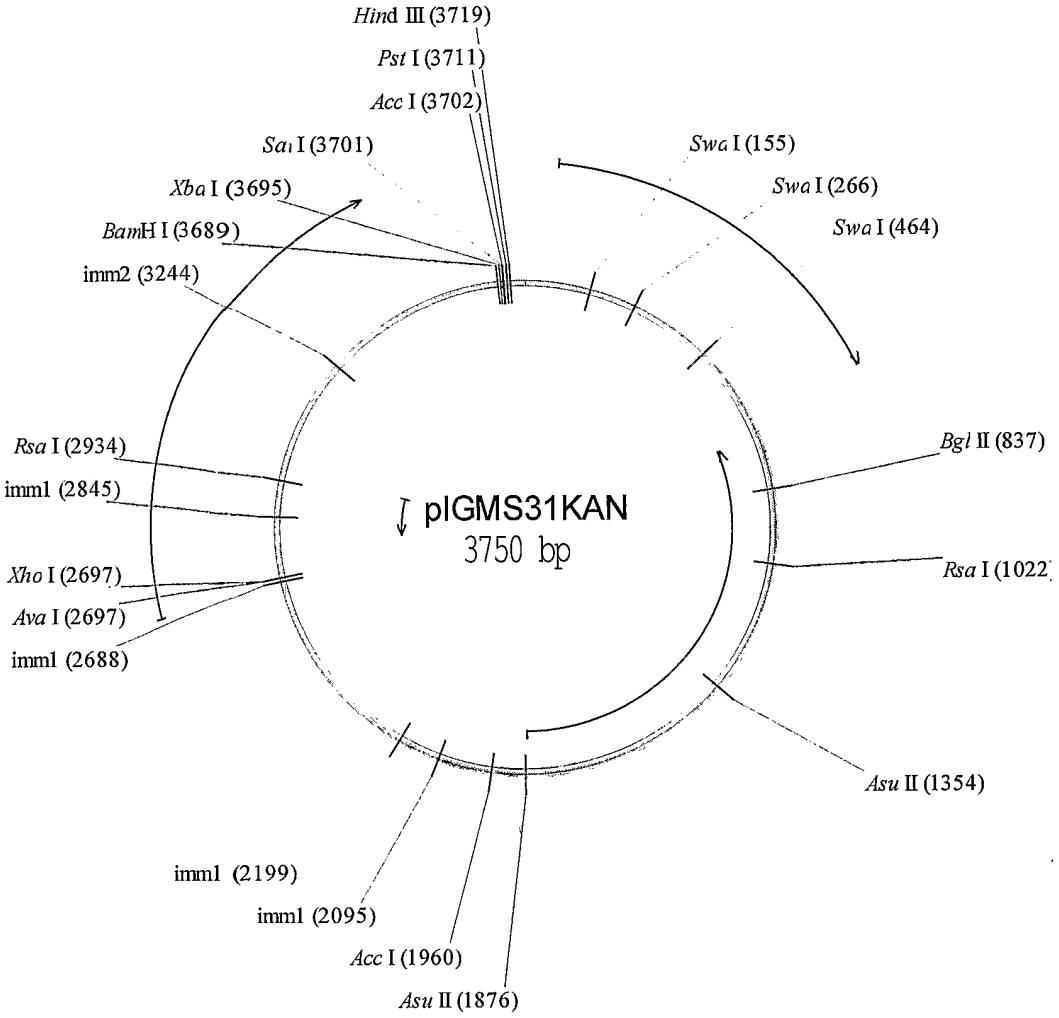

Hind III (3719)
Pst I (3711)
Acc I (3702)
Sal I (3701)
Xba I (3695)
BamH I (3689)
imm2 (3244)
Swa I (155)
Swa I (266)
Swa I (464)
Bgl II (837)
Rsa I (1022)
Asu II (1354)
Rsa I (2934)
imm1 (2845)
Xho I (2697)
Ava I (2697)
imm1 (2688)
pIGMS31KAN
3750 bp
imm1 (2199)
imm1 (2095)
Acc I (1960)
Asu II (1876)

Fig. 3

gtttataactgagttataaatacttataacttaattattaatggggttttaat
atgaaaaaaaataaattagtaaataaagaaaattactcaatattagagacttt
gccggaagatccattatttgaaaataaatcgactttagaaattgatttaaatc
aattcgatttatttaatagaattgcaaacgaaactgtagaagaacttataata
aaagaagttaacgatcctaatgaccgaagcgataaaagcaatggtgttaattt
aaatgcaaaagtttatgtagaaaaagaaaaaaagacttcattaaaaaaagatt
ttgttattacatttgtagataatcttgaggctttagcaaaattaaatttaaaa
cctaatgagtttagaattatcgtcgagattgtaaaggttatggaatacggaaa
tctaattaacctttcacaatctacaattgcgaaaaatttaaatcttgcaaaat
caaatgtaagttattattttaaaaaccttaaaaagaaaaatatattagtagaa
aaagacggacacgtctttatgaatagtaatatttttttctaaaggattagccca
tcgtttggacgaagaaaaaagaaaaaatttgaaatccgcacaagtcgaagacg
ataatttttaaaaactcattttaaaacccaacgggaaatttttcactgtttccc
gttccgggctttataattttaaagcctttggcttattctggggtgtgtagtta
ttattttgctgtttctgtgaatattcggcatctgctgctgcaatagcagcatt
gaagagttgtttaaattctgccggtttatgctcttgtattagatctaaaacat
cactgttaattttatatttttcatatctctgagaaattgaagcattttccttt
aactgccttttgtattctattatttccaactccatatcttttatttttctgtt
tttattttgaatgatttgttcttgctcagccgtggctctggccactgctaact
tttgctcctgttgtacccgttcctggactcttctgttgaagttctgacggaga
tcagacaactcagcctctgaatgctctaaacggctcttaaacccgtttagacg
agctataaggggctttcgttcgttcttgtaaggttttagagcctttcgtatgt
aacgattaacctcagctcctgtgtagtgcttaggcactacctccggatctcta
aataatttttttcttttcttctgtttcaacaactctaatttttttcaattttatt
attttctatttctaattcggtttcattcttaatgtttttaatatcgctgtaaa
attctttaacttccttatattctgcattcgaaacctctttttttaatcccacgg
attaaacctaatggcttattgtatttaaaatatatatcctgcattgtttcgta
ttttttcatatacgatttattgtttagtttatatattccgtttttatttttcga
ttggtgttataaaggcgtgaatatgtggagtttgctcgtccaagtgtaaaact
gcattaattgcattttccccatattcactttgcaaatattccatttgaacttt
aatccaatcctccaatttttttgttatttgcaaaaaattctggggaagctgtta
aaactaattcattacaaataacagaagtagaattacgagctttaacattggtt
gcttgaaaccttgcattaatatcagtccttaaatcaccagaaccgattaaaat
tcggttttgggatttaaggttaggatctgcattatgtgttttcctcaatcgca
tattgtgagaatttttcccggcgattgaagtatttttttgtttttttccactctt
aaaattgcataagccatattcgaaaacctcccgttaaaagcagtaaggttttt
ttctttttggcccctgccaggctcacaccgagatttctcggtatagtgagtat
acctttttctgcaatattgaaaatctataaatacatctacaataaaaaaagcaa
aagtcaacggctcaatccctcgcaagggaaaattaaaatttccccttactcac
gatttccaataaaagaaaaaagacagaacgctgagcaagtcaaaattttaatt
ttggcttgtgaagggttgaccaagcgaagcgcggtagggaaatctgcgcagat
gcttatgtattgcccggaacgggaaacgtctgttgtagcggtagcgaaaacac

Fig. 4a

atctcccggaacgggggttttcttttgcgtagcctggcaagttctgctcgatc
tggaggtttgcaccgtttactctcttactttcttattgttttaaatcttacat
acccctccagcccttgctattactgacttaaatcaaaaaaaagttatagattc
ctataacctaaaagttatagatttctataaccccagttatagattcctataac
ccccctaagttggtcattcgaccaacttcttataactaagttataaaaagttg
taatcatgtattgactagttgtatattttgtttataacctgtctcttatacac
atctcaaccatcatcgatgaattgtgtctcaaaatctctgatgttacattgca
caagataaaaatatatcatcatgaacaataaaactgtctgcttacataaacag
taatacaaggggtgttatgagccatattcaacgggaaacgtcttgctcgaggc
cgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgc
gataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccga
tgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgtta
cagatgagatggtcagactaaactggctgacggaatttatgcctcttccgacc
atcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgat
ccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtgaaa
atattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctgtt
tgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatc
acgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaatg
gctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattctca
ccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttga
cgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcagacc
gataccaggatcttgccatcctatggaactgcctcggtgagttttctccttca
ttacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataa
attgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaatt
ggttgtaacactggcagagcattacgctgacttgacgggacggcggctttgtt
gaataaatcgaacttttgctgagttgaaggatcagatcacgcatcttcccgac
aacgcagaccgttccgtggcaaagcaaaagttcaaaatcaccaactggtccac
ctacaacaaagctctcatcaaccgtggcggggatcctctagagtcgacctgca
ggcatgcaagcttcagggttgagatgtgtataagagacag

Fig. 4b

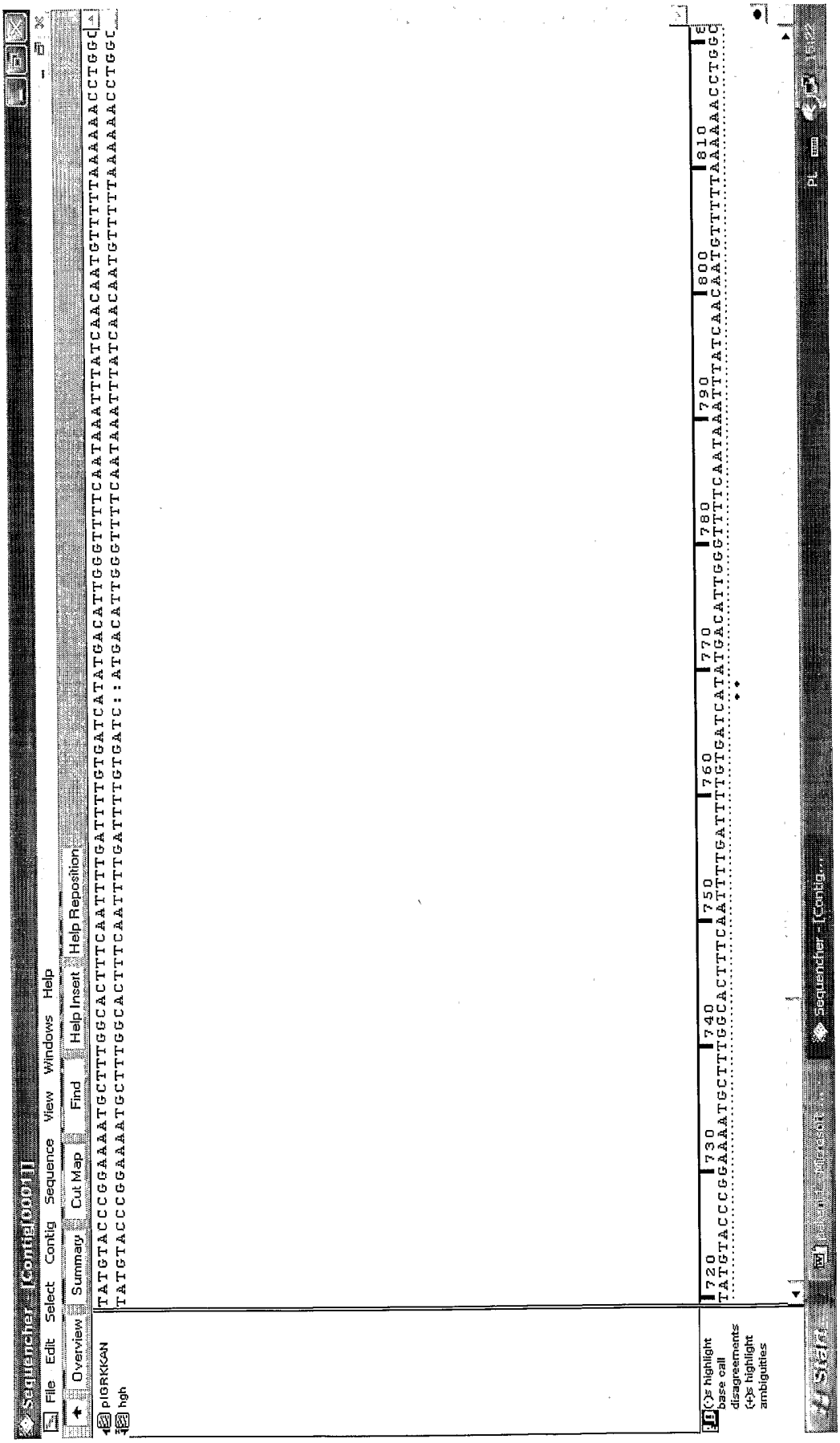

Sequencher - [Contig1000]
File Edit Select Contig Sequence View Windows Help
Overview Summary Cut Map Find Help Insert Help Reposition
pIGRKIAN
hgh
TATGTACCCGGAAAATGCTTTGGCACTTTCAATTTTGATTTTGTGATCATATGACATTGGGTTTTCAATAAATTTATCAACAATGTTTTTAAAAACCTGGC
TATGTACCCGGAAAATGCTTTGGCACTTTCAATTTTGATTTTGTGATC::ATGACATTGGGTTTTCAATAAATTTATCAACAATGTTTTTAAAAACCTGGC
720 730 740 750 760 770 780 790 800 810
TATGTACCCGGAAAATGCTTTGGCACTTTCAATTTTGATTTTGTGATCATATGACATTGGGTTTTCAATAAATTTATCAACAATGTTTTTAAAAACCTGGC
(-)s highlight base call disagreements
(+)s highlight ambiguities
Start
Sequencher - [Contig...
PL

Fig. 5

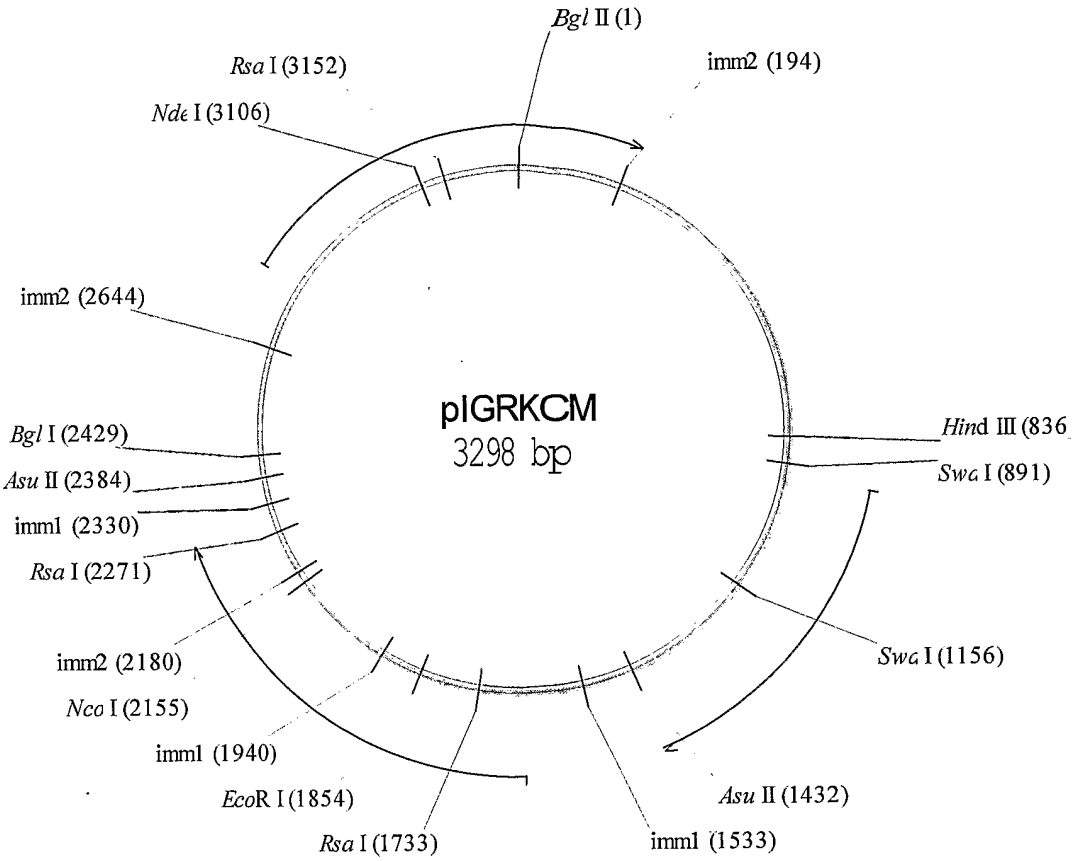

Bgl II (1)
imm2 (194)
Rsa I (3152)
Nde I (3106)
imm2 (2644)
pIGRKCM
3298 bp
Bgl I (2429)
Asu II (2384)
imm1 (2330)
Rsa I (2271)
Hind III (836
Swa I (891)
Swa I (1156)
imm2 (2180)
Nco I (2155)
imm1 (1940)
EcoR I (1854)
Rsa I (1733)
Asu II (1432)
imm1 (1533)

Fig. 6

gatctcagcagagttacggaagaaagagcaaaaaaagcaaaggtggatttgat
gacatcgcagatgtcgaaaccaatgttaaaccccgtggcggtgatagattatt
gagatttaagatcgcaaataaaaacaaagcagcggcaaaaattgccgatactt
ccacccccagcagtcctccaccggctcccgttcgtcgcttcaaaatgtgaacc
gtgagcagttcaggaggttccctcctggactgtgaagggttggcccgtccggt
caggacggttttacagcaaaatcctccatagcgaagcagaagcccggaacggt
aactggatggttttcccccgtgggggattgatctgttacttgaaaaccaatga
tcttaaaagccatctcaaaagttgaaaatttcacccccttagtgttcttaaaa
ttcttagatgttcttaggagttaaaaaactactctctaaccattgatattact
ggatttttaaaaaaggcagttgtcaaaaacttcaaccgtagttgtcaaattcg
tcaactccagttgtcaaattcgtcaactgaggttgtcaaatccgacaactcta
gccagtttccaagtagaaactacagtttctaaactgcaacttttttctacttttt
tgcaacttaatctattgactagtcctttataaatgttaaaacatatatatagaa
aataaataaaaagaggaggtttctatggatattggaaatatattaaatgagag
tttaagtattgattacgaaaaattagatttgtttttggaaaaatatgatttaa
caccagaacaaaaagttgcagtttatgaatttcacgcaaaagcttataaaaaa
aataaaactttagttatttctgaaacaaaagaaaataaatttaaatctatttc
cgaagtgttgaatacgtgcatttattcccaaaaaatttaaaaattttaattaa
aaaatatggtttaaatacaaacgaattattggttttaacggaaataatggagt
caatgctttcacacggaaatttattaattaatttttcgcaaaaggcactttgc
gaattaacaggaattaataaatctacaatgtgtaaaacatttaaaaccctcaa
acaaaagcagtgtttaattgagaaaaacggacatatttatttaaattctgtga
tatttatgaaagggttacctcataaattgtttatgcaatttagagatcatttt
ttaaattctatctcatataaattagatgatgaagaagaatttgaaaaagtctt
cgacgataattttattaaagcatacgaaaaaaatctcaaagagattaaaaaga
aaaagcaacaaataaaagaaaagaaaatatcaaaagcattagataattttgaa
aaagaaatctcgaaagaatggaaggaaaagtttaaagacgaagaggaaaattt
cgaataaatacctgtgacggaagatcacttcgcagaataaataaatcctggtg
tccctgttgataccgggaagccctgggccaacttttggcgaaaatgagacgtt
gatcggcacgtaagaggttccaactttcaccataatgaaataagatcactacc
gggcgtatttttgagttatcgagattttcaggagctaaggaagctaaaatgg
agaaaaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaa
gaacattttgaggcatttcagtcagttgctcaatgtacctataaccagaccgt
tcagctggatattacggccttttaaagaccgtaaagaaaaataagcacaagt
tttatccggcctttattcacattcttgcccgcctgatgaatgctcatccggaa
ttccgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcaccc
ttgttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtg
aataccacgacgatttccggcagtttctacacatatattcgcaagatgtggcg
tgttacggtgaaaacctggcctatttccctaaagggtttattgagaatatgtt
tttcgtctcagccaatccctgggtgagtttcaccagttttgatttaaacgtgg
ccaatatggacaacttcttcgcccccgttttcaccatgggcaaatattatacg
caaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtttg

Fig. 7a

tgatggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatg
agtggcagggcggggcgtaatttttttaaggcagttattggtgcccttaaacg
cctggtgctacgcctgaataagtgataataagcggatgaatggcagaaattcg
aatttggttttgaatcggaaatataaaaccgccctcgccgggcaggcgaatcc
cttattgaaatagaataaattctattccactaagggatttttttattcattg
tttctccacatttgcaatattgacattaacttccacccggatataacagtagt
ataagttgttgtttcaacccgtcttttgggtggaacaacaaggcattttagg
gatagagcaaagcgaaggccataaaattgccacccccaaccgggggtcgttgt
tcgatttgagcgatagcgaaaaattgaacataaggggggagggtttgggtttt
acggtatttcaaatttgagcaaagcgaatttttgaaatttccggttcttttaa
tttgcaatgaggaaaaatcaatatgggtaattcaaaaagaaatataaaaaaac
taaatgataatttagagaggatattttagattatgcgatcgcgcacaatcta
aaatgtgctaacgcacttgctattttatacgcaacgggttgccgtccggacga
actccaaaccggagttactgtaaactatgacagtaaaaaaaatgaaattgaat
ttagaataattggatcaaaactaaatagaagaatgagaagaggcataggggtt
agaaaaataaaagtaaaaatcaataatgaaaatgccaggtttttaaaaacat
tgttgataaatttattgaaaacccaatgtcatatgatcacaaaatcaaaattg
aaagtgccaaagcattttccgggtacataacaaaaatatcgaaaaagctatgg
cccaggaaaacctatcatgcttctgcatattcttttagacatgcaaaagcaac
ggaattaaaaaattccgattatgataaaatcgaaatagctcagattatgggcc
atgcctcag

Fig. 7b

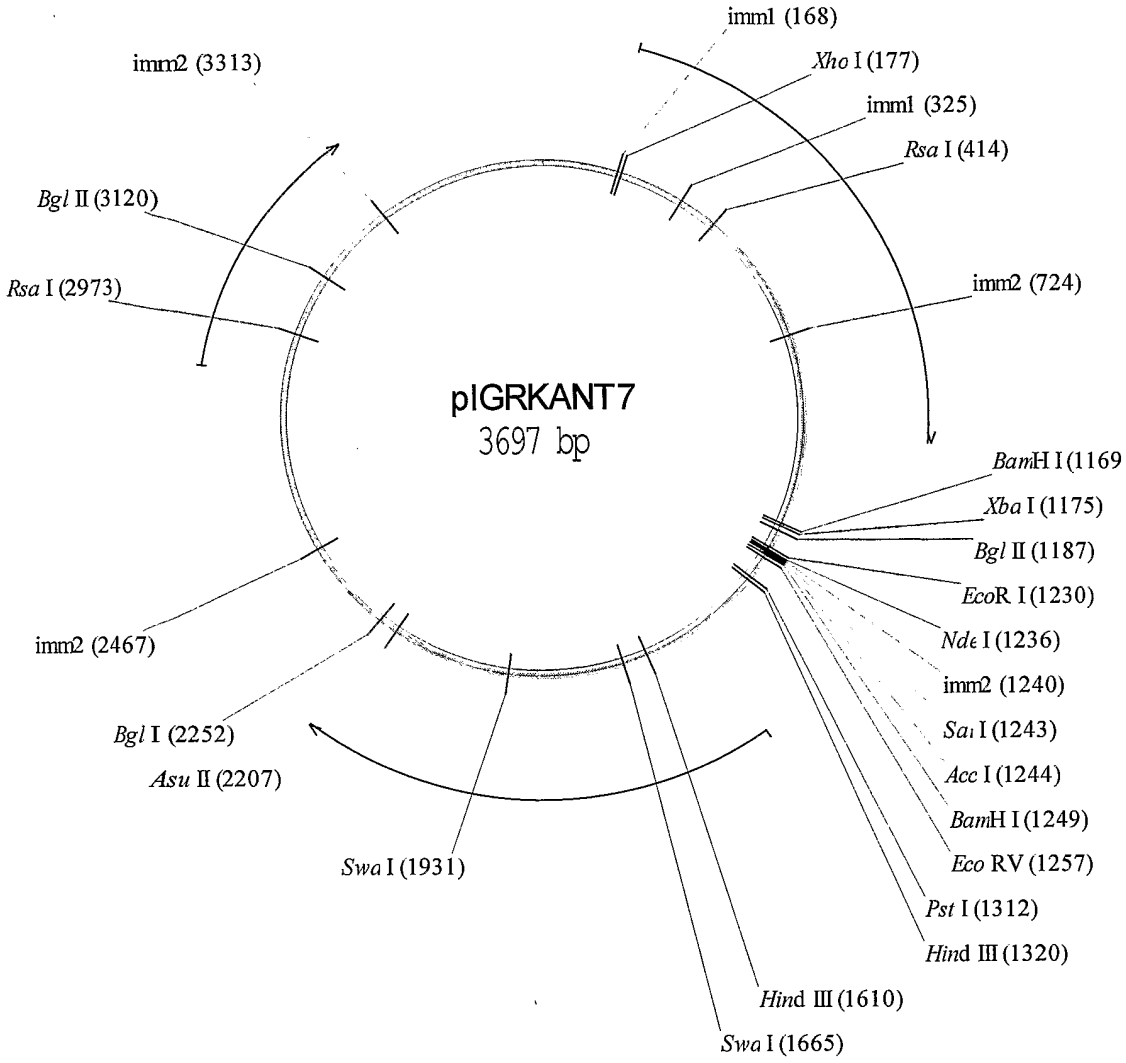

imm1 (168)
Xho I (177)
imm1 (325)
Rsa I (414)
imm2 (724)
BamH I (1169
Xba I (1175)
Bgl II (1187)
EcoR I (1230)
Nde I (1236)
imm2 (1240)
Sal I (1243)
Acc I (1244)
BamH I (1249)
Eco RV (1257)
Pst I (1312)
Hind III (1320)
Hind III (1610)
Swa I (1665)
Swa I (1931)
Asu II (2207)
Bgl I (2252)
imm2 (2467)
Rsa I (2973)
Bgl II (3120)
imm2 (3313)
pIGRKANT7
3697 bp

Fig. 8

actctagccctgtctcttatacacatctcaaccatcatcgatgaattgtgtct
caaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaat
aaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattc
aacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgat
ttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaat
ctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggca
aaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctg
acggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatga
tgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattag
aagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctg
cgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgt
atttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcga
gtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaa
atgcataaacttttgccattctcaccggattcagtcgtcactcatggtgattt
ctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatg
ttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaac
tgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatgg
tattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagt
ttttctaatcagaattggttaattggttgtaacactggcagagcattacgctg
acttgacgggacggcggctttgttgaataaatcgaacttttgctgagttgaag
gatcagatcacgcatcttcccgacaacgcagaccgttccgtggcaaagcaaaa
gttcaaaatcaccaactggtccacctacaacaaagctctcatcaaccgtggcg
gggatcctctagagtcgagagatctaattaatacgactcactatagggagacc
aagaaggaagaattcatatgtcgtcgacggatccgatatctagcataacccct
tggggcctctaaacgggtcttgaggggttttttgctgcaggcatgcaagcttc
agggttgagatgtgtataagagacagactctagccagtttccaagtagaaact
acagtttctaaactgcaacttttttctacttttttgcaacttaatctattgacta
gtcctttataaatgttaaaacatatatatagaaataaataaaaagaggaggtt
tctatggatattggaaatatattaaatgagagtttaagtattgattacgaaaa
attagatttgtttttggaaaaatatgatttaacaccagaacaaaaagttgcag
tttatgaatttcacgcaaaagcttataaaaaaaataaaactttagttatttct
gaaacaaaagaaaataaatttaaatctatttccgaaggtgttgaatacgtgca
tttattcccaaaaaatttaaaaattttaattaaaaaatatggtttaaatacaa
acgaattattggttttaacggaaataatggagtcaatgctttcacacggaaat
ttattaattaatttttcgcaaaaggcactttgcgaattaacaggaattaataa
atctacaatgtgtaaaacatttaaaaccctcaaacaaaagcagtgtttaattg
agaaaaacggacatatttatttaaattctgtgatatttatgaaagggttacct
cataaattgtttatgcaatttagagatcatttttaaattctatctcatataa
attagatgatgaagaagaatttgaaaaagtcttcgacgataattttattaaag
catacgaaaaaaatctcaaagagattaaaaagaaaaagcaacaaataaaagaa
aagaaaatatcaaaagcattagataattttgaaaaagaaatctcgaaagaatg
gaaggaaaagtttaaagacgaagaggaaaatttcgaatttggttttgaatcgg

Fig. 9a

aaatataaaaccgccctcgccgggcaggcgaatcccttattgaaatagaataa
attctattccactaagggatttttttattcattgtttctccacatttgcaat
attgacattaacttccacccggatataacagtagtataagttgttgtttcaac
ccgtcttttgggtggaacaacaaggcattttagggatagagcaaagcgaagg
ccataaaattgccacccccaaccgggggtcgttgttcgatttgagcgatagcg
aaaaattgaacataaggggggagggtttgggttttacggtatttcaaatttga
gcaaagcgaatttttgaaatttccggttcttttaatttgcaatgaggaaaaat
caatatgggtaattcaaaaagaaatataaaaaaactaaatgataattttagag
aggatattttagattatgcgatcgcgcacaatctaaaatgtgctaacgcactt
gctattttatacgcaacgggttgccgtccggacgaactccaaaccggagttac
tgtaaactatgacagtaaaaaaaatgaaattgaatttagaataattggatcaa
aactaaatagaagaatgagaagaggcataggggttagaaaaataaaagtaaaa
atcaataatgaaaatgccaggtttttttaaaaacattgttgataaatttattga
aaacccaatgtcatgatcacaaaatcaaaattgaaagtgccaaagcattttcc
gggtacataacaaaaatatcgaaaaagctatggcccaggaaaacctatcatgc
ttctgcatattcttttagacatgcaaaagcaacggaattaaaaaattccgatt
atgataaaatcgaaatagctcagattatgggccatgcctcagttagatctcag
cagagttacggaagaaagagcaaaaaaagcaaaggtggatttgatgacatcgc
agatgtcgaaaccaatgttaaaccccgtggcggtgatagattattgagattta
agatcgcaaataaaaacaaagcagcggcaaaaattgccgatacttccaccccc
agcagtcctccaccggctcccgttcgtcgcttcaaaatgtgaaccgtgagcag
ttcaggaggttccctcctggactgtgaagggttggcccgtccggtcaggacgg
ttttacagcaaaatcctccatagcgaagcagaagcccggaacgggtaactgga
tggttttcccccgtgggggattgatctgttacttgaaaaccaatgatcttaaa
agccatctcaaaagttgaaaatttcacccccttagtgttcttaaaattcttag
atgttcttaggagttaaaaaactactctctaaccattgatattactggatttt
taaaaaaggcagttgtcaaaaacttcaaccgtagttgtcaaattcgtcaactc
cagttgtcaaattcgtcaactgaggttgtcaaataca

Fig. 9b

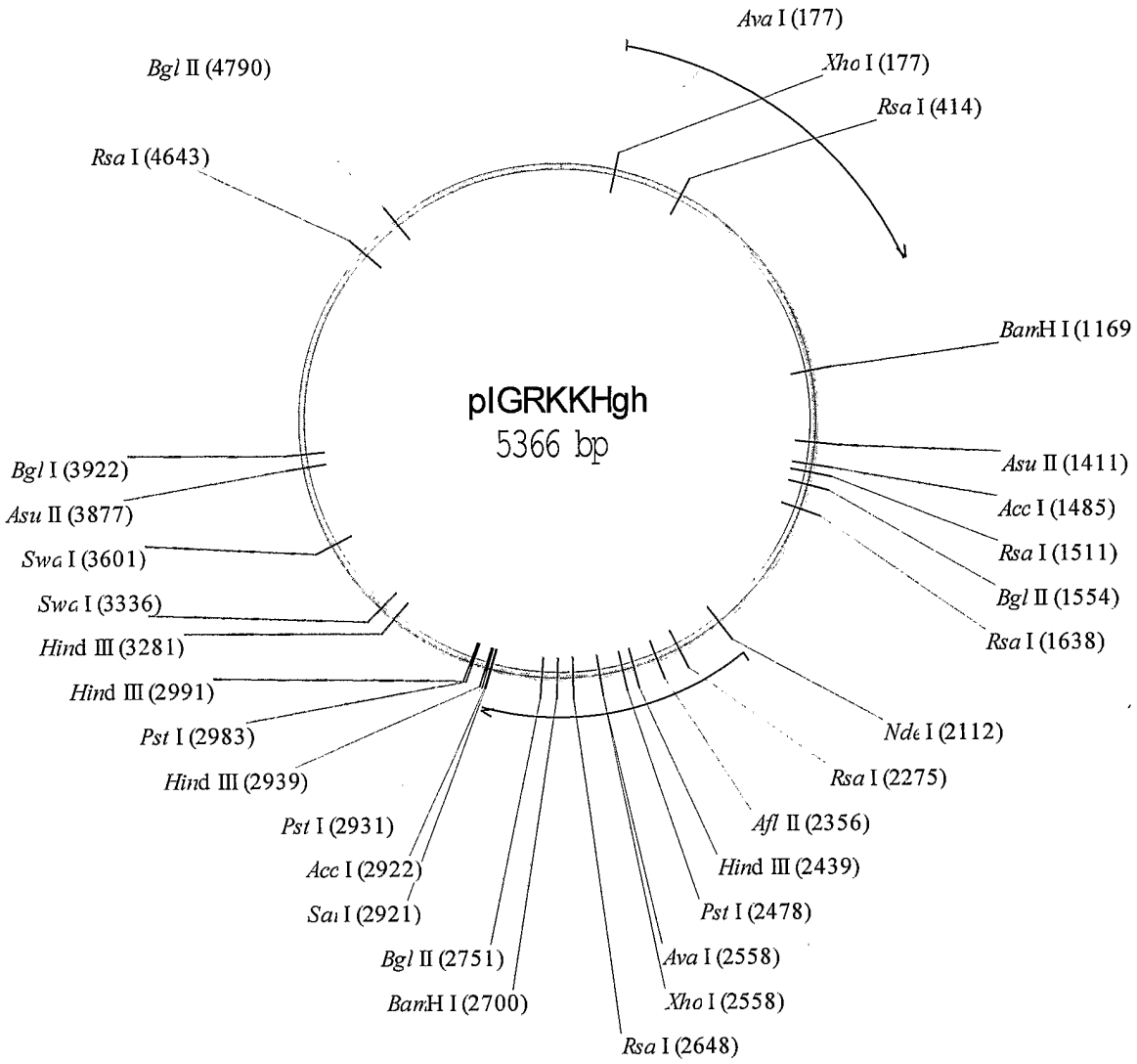

pIGRKKHgh
5366 bp
Ava I (177)
Xho I (177)
Rsa I (414)
BamH I (1169
Asu II (1411)
Acc I (1485)
Rsa I (1511)
Bgl II (1554)
Rsa I (1638)
Nde I (2112)
Rsa I (2275)
Afl II (2356)
Hind III (2439)
Pst I (2478)
Ava I (2558)
Xho I (2558)
Rsa I (2648)
BamH I (2700)
Bgl II (2751)
Sal I (2921)
Acc I (2922)
Pst I (2931)
Hind III (2939)
Pst I (2983)
Hind III (2991)
Hind III (3281)
Swa I (3336)
Swa I (3601)
Asu II (3877)
Bgl I (3922)
Rsa I (4643)
Bgl II (4790)

Fig. 10

actctagccctgtctcttatacacatctcaaccatcatcgatgaattgtgtct
caaaatctctgatgttacattgcacaagataaaaatatatcatcatgaacaat
aaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattc
aacgggaaacgtcttgctcgaggccgcgattaaattccaacatggatgctgat
ttatatgggtataaatgggctcgcgataatgtcgggcaatcaggtgcgacaat
ctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggca
aaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctg
acggaatttatgcctcttccgaccatcaagcattttatccgtactcctgatga
tgcatggttactcaccactgcgatccccggaaaaacagcattccaggtattag
aagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctg
cgccggttgcattcgattcctgtttgtaattgtccttttaacagcgatcgcgt
atttcgtctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcga
gtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaa
atgcataaacttttgccattctcaccggattcagtcgtcactcatggtgattt
ctcacttgataaccttatttttgacgaggggaaattaataggttgtattgatg
ttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaac
tgcctcggtgagttttctccttcattacagaaacggctttttcaaaaatatgg
tattgataatcctgatatgaataaattgcagtttcatttgatgctcgatgagt
ttttctaatcagaattggttaattggttgtaacactggcagagcattacgctg
acttgacgggacggcggctttgttgaataaatcgaacttttgctgagttgaag
gatcagatcacgcatcttcccgacaacgcagaccgttccgtggcaaagcaaaa
gttcaaaatcaccaactggtccacctacaacaaagctctcatcaaccgtggcg
gggatccggaccgttggcgatgtgcggtttgctacattcacagatgttcttcg
ccacttccagcagcaggtcatcaggggtgatttcaggatcgtagataaaggtc
aggttcggtgaaacctgcttcaactctgcatctgcacgtaagatcgcgcgggt
aatgggcgaatcagacgggccgatattggcgtgcataaaggcgtctggcaggg
ttctgtcgaggtaacgccagaaacgttttattcgaacatcgatctcgtcttgt
gttagaattaattctaacatacggttgcaacaacgcatccagttgccccaggt
agaccggcatcgatgtgaccgacggtacgtggtggtaaagaatggtcagcaga
gagagtgcgtcatcaagatctttcgcgccttccagctccagccattcggaacc
gttcgccagaaaacgggcgtaatcgggtaagacatagcgcggtttgtacggcg
catgaccttcaaacatatcgcagattacaccttcatccagcgcgcggcgggct
tcggcaggaagctgtgggtaaggcagattgttttctgcttccagtgccagaaa
atggcgcttctgctccgggctaagcactgggctggtgacaatttgctggcaac
gttgttgcagtgcattttcatgagaagtgggcatcttcttttccttttatgcc
gaaggtgatgcgccattgtaagaagtttcgtgatgttcactttgatcctgatg
cgtttgccaccactgacgcattcatttgaaagtgaattatttgaaccagatcg
cattacagtgatgcaaacttgtaagtagatttccttaattgtgatgtgtatcg
aagtgtgttgcggagtagatgttagaatactaacaaactcgcaaggtgaattt
tattggcgacaagcctaggtttgtttaactttaaggagaaatcatatgcaaat
ttttgttaaaactttaactggtaaaaccattaccttagaagttgaatcttcag
ataccattgataatgttaaatctaaaattcaagataaagaaggtattcctcca

Fig. 11a

gatcaacaacgtctaatatttgcaggtaaacagttagaagatggtcgtaccct
gtctgattataacattcagaaagaatctaccttacatctggtcttacgtctcc
gcggtggtttcccaaccattcccttaagtaggcttttgacaacgctatgctc
cgcgcccatcgtctgcaccagctggcctttgacacctaccaggagtttgaaga
agcttatatcccaaaggaacagaagtattcattcctgcagaaccccagacct
ccctctgtttctcagagtctattccgacaccctccaacagggaggaaacacaa
cagaaatccaacctcgagctgctccgcatctccctgctgctcatccagtcgtg
gctggagcccgtgcagttcctcaggagtgtcttcgccaacagcctggtgtacg
gcgcctctgacagcaacgtctatgacctcctaaaggacctagaggaagggatc
caaacgctgatggggaggctggaagatggcagcccccggactgggcagatctt
caagcagacctacagcaagttcgacacaaactcacacaacgatgacgcactac
tcaagaactacgggctgctctactgcttcaggaaggacatggacaaggtcgag
acattcctgcgcatcgtgcagtgccgctctgtggagggcagctgtggcttcta
aaaagtcgacctgcaggcatgcaagcttagcccgcttaatgagcgggcttttt
tttctcgacctgcaggcatgcaagcttcagggttgagatgtgtataagagaca
gactctagccagtttccaagtagaaactacagtttctaaactgcaacttttтc
tacttttgcaacttaatctattgactagtcctttataaatgttaaaacatat
atatagaaataaataaaaagaggaggtttctatggatattggaaatatattaa
atgagagtttaagtattgattacgaaaaattagatttgtttttggaaaaatat
gatttaacaccagaacaaaaagttgcagtttatgaatttcacgcaaaagctta
taaaaaaaataaaactttagttatttctgaaacaaaagaaaataaatttaaat
ctatttccgaagtgttgaatacgtgcatttattcccaaaaaatttaaaaattt
taattaaaaaatatggtttaaatacaaacgaattattggttttaacggaaata
atggagtcaatgctttcacacggaaatttattaattaatttttcgcaaaaggc
actttgcgaattaacaggaattaataaatctacaatgtgtaaaacatttaaaa
ccctcaaacaaaagcagtgtttaattgagaaaaacggacatatttatttaaat
tctgtgatatttatgaaagggttacctcataaattgtttatgcaatttagaga
tcatttttaaattctatctcatataaattagatgatgaagaagaatttgaaa
aagtcttcgacgataattttattaaagcatacgaaaaaaatctcaaagagatt
aaaaagaaaaagcaacaaataaaagaaaagaaaatatcaaaagcattagataa
ttttgaaaaagaaatctcgaaagaatggaaggaaaagtttaaagacgaagagg
aaaatttcgaatttggttttgaatcggaaatataaaaccgccctcgccgggca
ggcgaatcccttattgaaatagaataaattctattccactaagggatttttt
tattcattgtttctccacatttgcaatattgacattaacttccacccggatat
aacagtagtataagttgttgtttcaacccgtctttttgggtggaacaacaagg
cattttagggatagagcaaagcgaaggccataaaattgccacccccaaccggg
ggtcgttgttcgatttgagcgatagcgaaaaattgaacataaggggggagggt
ttgggttttacggtatttcaaatttgagcaaagcgaatttttgaaatttccgg
ttcttttaatttgcaatgaggaaaaatcaatatgggtaattcaaaaagaaata
taaaaaaactaaatgataattttagagaggatattttagattatgcgatcgcg
cacaatctaaaatgtgctaacgcacttgctattttatacgcaacgggttgccg
tccggacgaactccaaaccggagttactgtaaactatgacagtaaaaaaaatg

Fig. 11b

aaattgaatttagaataattggatcaaaactaaatagaagaatgagaagaggc
atagggttagaaaaataaaagtaaaaatcaataatgaaaatgccaggttttt
taaaaacattgttgataaatttattgaaaacccaatgtcatgatcacaaaatc
aaaattgaaagtgccaaagcattttccgggtacataacaaaaatatcgaaaaa
gctatggcccaggaaaacctatcatgcttctgcatattcttttagacatgcaa
aagcaacggaattaaaaaattccgattatgataaaatcgaaatagctcagatt
atgggccatgcctcagttagatctcagcagagttacggaagaaagagcaaaaa
aagcaaaggtggatttgatgacatcgcagatgtcgaaaccaatgttaaacccc
gtggcggtgatagattattgagatttaagatcgcaaataaaaacaaagcagcg
gcaaaaattgccgatacttccacccccagcagtcctccaccggctcccgttcg
tcgcttcaaaatgtgaaccgtgagcagttcaggaggttccctcctggactgtg
aagggttggcccgtccggtcaggacggttttacagcaaaatcctccatagcga
agcagaagcccggaacggtaactggatggttttcccccgtgggggattgatct
gttacttgaaaaccaatgatcttaaaagccatctcaaaagttgaaaatttcac
ccccttagtgttcttaaaattcttagatgttcttaggagttaaaaaactactc
tctaaccattgatattactggatttttaaaaaaggcagttgtcaaaaacttca
accgtagttgtcaaattcgtcaactccagttgtcaaattcgtcaactgaggtt
gtcaaatccgaca

Fig. 11c

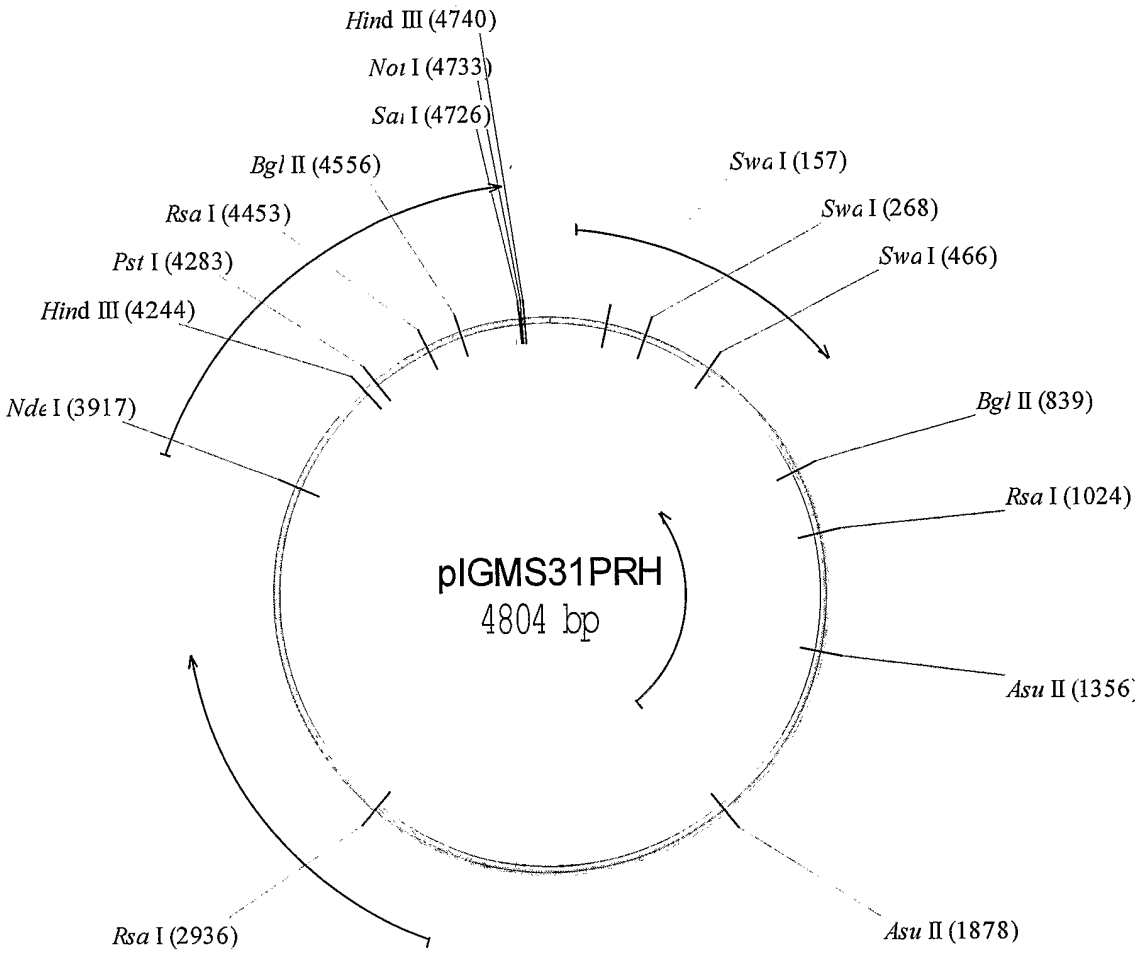

Hind III (4740)
Not I (4733)
Sal I (4726)
Bgl II (4556)
Rsa I (4453)
Pst I (4283)
Hind III (4244)
Nde I (3917)
Swa I (157)
Swa I (268)
Swa I (466)
Bgl II (839)
Rsa I (1024)
Asu II (1356)
Asu II (1878)
Rsa I (2936)
pIGMS31PRH
4804 bp

Fig. 12

aggtttataactgagttataaatacttataacttaattattaatggggtttta
atatgaaaaaaaataaattagtaaataaagaaaattactcaatattagagact
ttgccggaagatccattatttgaaaataaatcgactttagaaattgatttaaa
tcaattcgatttatttaatagaattgcaaacgaaactgtagaagaacttataa
taaaagaagttaacgatcctaatgaccgaagcgataaaagcaatggtgttaat
ttaaatgcaaaagtttatgtagaaaaagaaaaaaagacttcattaaaaaaaga
ttttgttattacatttgtagataatcttgaggctttagcaaaattaaatttaa
aacctaatgagtttagaattatcgtcgagattgtaaaggttatggaatacgga
aatctaattaacctttcacaatctacaattgcgaaaaatttaaatcttgcaaa
atcaaatgtaagttattattttaaaaaccttaaaaagaaaaatatattagtag
aaaaagacggacacgtctttatgaatagtaatatttttttctaaaggattagcc
catcgtttggacgaagaaaaaagaaaaaatttgaaatccgcacaagtcgaaga
cgataatttaaaaactcattttaaaacccaacgggaaatttttcactgtttc
ccgttccgggctttataattttaaagcctttggcttattctggggtgtgtagt
tattattttgctgtttctgtgaatattcggcatctgctgctgcaatagcagca
ttgaagagttgtttaaattctgccggtttatgctcttgtattagatctaaaac
atcactgttaatttatatttttcatatctctgagaaattgaagcattttcct
ttaactgccttttgtattctattatttccaactccatatcttttatttttctg
tttttatttgaatgatttgttcttgctcagccgtggctctggccactgctaa
cttttgctcctgttgtacccgttcctggactcttctgttgaagttctgacgga
gatcagacaactcagcctctgaatgctctaaacggctcttaaacccgtttaga
cgagctataaggggctttcgttcgttcttgtaaggttttagagcctttcgtat
gtaacgattaacctcagctcctgtgtagtgcttaggcactacctccggatctc
taaataatttttttcttttcttctgtttcaacaactctaatttttttcaatttta
ttattttctatttctaattcggtttcattcttaatgtttttaatatcgctgta
aaattctttaacttccttatattctgcattcgaaacctctttttttaatcccac
ggattaaacctaatggcttattgtatttaaaatatatatcctgcattgtttcg
tatttttttcatatacgatttattgtttagtttatatattccgtttttattttc
gattggtgttataaaggcgtgaatatgtggagtttgctcgtccaagtgtaaaa
ctgcattaattgcattttccccatattcactttgcaaatattccatttgaact
ttaatccaatcctccaatttttttgttatttgcaaaaaattctggggaagctgt
taaaactaattcattacaaataacagaagtagaattacgagctttaacattgg
ttgcttgaaaccttgcattaatatcagtccttaaatcaccagaaccgattaaa
attcggttttgggatttaaggttaggatctgcattatgtgttttcctcaatcg
catattgtgagaatttttcccggcgattgaagtatttttttgtttttttccactc
ttaaaattgcataagccatattcgaaaacctcccgttaaaagcagtaaggttt
ttttctttttggcccctgccaggctcacaccgagatttctcggtatagtgagt
atacctttttctgcaatattgaaaatctataaatacatctacaataaaaaaagc
aaaagtcaacggctcaatccctcgcaagggaaaattaaaatttccccttactc
acgatttccaataaaagaaaaaagacagaacgctgagcaagtcaaaattttaa
ttttggcttgtgaagggttgaccaagcgaagcgcggtagggaaatctgcgcag
atgcttatgtattgcccggaacgggaaacgtctgttgtagcggtagcgaaaac

Fig. 13a

acatctcccggaacgggggttttcttttgcgtagcctggcaagttctgctcga
tctggaggtttgcaccgtttactctcttactttcttattgttttaaatcttac
atacccctccagcccttgctattactgacttaaatcaaaaaaaagttatagat
tcctataacctaaaagttatagatttctataaccccagttatagattcctata
acccccctaagttggtcattcgaccaacttcttataactaagttataaaaagt
tgtaatcatgtattgactagttgtatattttgtttataacctgtctcttatac
acatctcaaccatcatcgatgaattgtgtctcaaaatctctgatgttacattg
cacaagataaaaatatatcatcatgaacaataaaactgtctgcttacataaac
agtaatacaaggggtgttatgagccatattcaacgggaaacgtcttgctcgag
gccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctc
gcgataatgtcgggcaatcaggtgcgacaatctatcgattgtatgggaagccc
gatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgt
tacagatgagatggtcagactaaactggctgacggaatttatgcctcttccga
ccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcg
atccccggaaaaacagcattccaggtattagaagaatatcctgattcaggtga
aaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgattcctg
tttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaa
tcacgaatgaataacggtttggttgatgcgagtgattttgatgacgagcgtaa
tggctggcctgttgaacaagtctggaaagaaatgcataaacttttgccattct
caccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttt
gacgaggggaaattaataggttgtattgatgttggacgagtcggaatcgcaga
ccgataccaggatcttgccatcctatggaactgcctcggtgagttttctcctt
cattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaat
aaattgcagtttcatttgatgctcgatgagtttttctaatcagaattggttaa
ttggttgtaacactggcagagcattacgctgacttgacgggacggcggctttg
ttgaataaatcgaacttttgctgagttgaaggatcagatcacgcatcttcccg
acaacgcagaccgttccgtggcaaagcaaaagttcaaaatcaccaactggtcc
acctacaacaaagctctcatcaaccgtggcggggatcgatcttagatccgttg
tttctcgtctaataaatgaacgaaaaatacttcaaatgactgatggttatcag
gtcactgctttgggggctagctatgttaggagcgtctttgatagaaagacact
tgaccgattgcggcttgagattatgaattttgaaaaccgtagaaaatcaacat
ttaactatgataagattccgtatgcgcaccaagaaggaagaattccatatgca
gattttcgtcaaaactttgaccggtaaaaccataacattggaagttgaatctt
ccgataccatcgacaacgttaagtcgaaaattcaagacaaggaaggtatccct
ccagatcaacaaagattgatctttgccggtaagcagctagaagacggtagaac
gctgtctgattacaacattcagaaggagtccaccttacatcttgtcttaagac
tccgcggtggtttcccaaccattcccttatccaggcttttgacaacgctagt
ctccgcgcccatcgtctgcaccagctggcctttgacacctaccaggagtttga
agaagcttatatcccaaaggaacagaagtattcattcctgcagaacccccaga
cctccctctgtttctcagagtctattccgacaccctccaacagggaggaaaca
caacagaaatccaacctcgagctgctccgcatctccctgctgctcatccagtc
gtggctggagcccgtgcagttcctcaggagtgtcttcgccaacagcctggtgt

Fig. 13b

acggcgcctctgacagcaacgtctatgacctcctaaaggacctagaggaaggg
atccaaacgctgatggggaggctggaagatggcagcccccggactgggcagat
cttcaagcagacctacagcaagttcgacacaaactcacacaacgatgacgcac
tactcaagaactacgggctgctctactgcttcaggaaggacatggacaaggtc
gagacattcctgcgcatcgtgcagtgccgctctgtggagggcagctgtggctt
ctaaaaagtcgacgcggccgcaagcttagcccgcttaatgagcgggctttttt
ttagcttcagggttgagatgtgtataagagacag

Fig. 13c

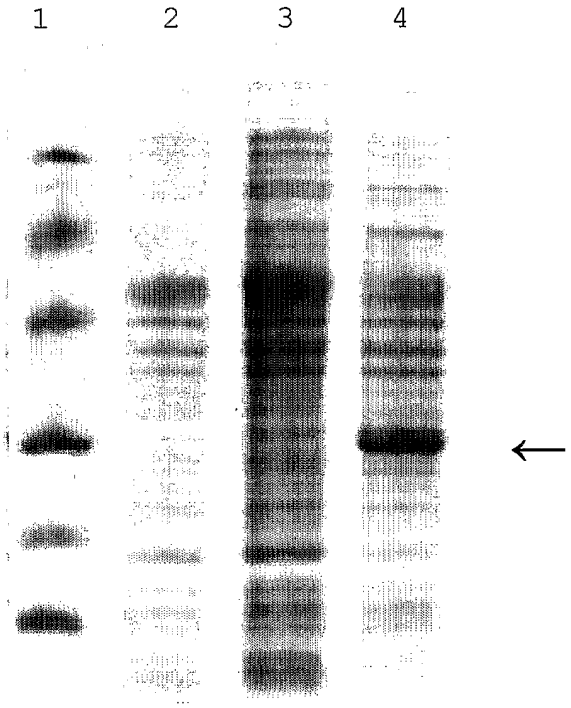


Fig. 14