



(51) International Patent Classification:
C12N 9/22 (2006.01) C12N 15/66 (2006.01)

(21) International Application Number:
PCT/EP2010/052892

(22) International Filing Date:
8 March 2010 (08.03.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09154567.3 6 March 2009 (06.03.2009) EP

(71) Applicant (for all designated States except US): EURO-
PÄISCHES LABORATORIUM FÜR MOLEKU-
LAR BIOLOGIE (EMBL) [DE/DE]; Meyerhofstrasse 1,
69117 Heidelberg (DE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): BERGER, Imre [DE/
FR]; 2bis Rue de Bellevue, Parc Symphonie V. No. 3,
F-38120 St. Egreve (FR).

(74) Agent: HABERMANN, HRUSCHKA & SCHNABEL;
Montgelastrasse 2, 81679 München (DE).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP,
KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI,
NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,
SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,
MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- as to applicant's entitlement to apply for and be granted
a patent (Rule 4.17(ii))
- of inventorship (Rule 4.17(iv))

Published:

- without international search report and to be republished
upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: NUCLEIC ACIDS FOR CLONING AND EXPRESSING MULTIPROTEIN COMPLEXES

(57) Abstract: The present invention relates to a nucleic acid containing at least one homing endonuclease site (HE) and at least one restriction enzyme site (X) wherein the HE and X sites are selected such that HE and X result in compatible cohesive ends when cut by the homing endonuclease and restriction enzyme, respectively, and the ligation product of HE and X cohesive ends can neither be cleaved by the homing endonuclease nor by the restriction enzyme. Further subject-matter of the present invention relates to a vector comprising the nucleic acid of the present invention, host cells containing the nucleic acid and/or the vector, a kit for cloning and/or expression of multiprotein complexes making use of the vector and the host cells, a method for producing a vector containing multiple expression cassettes, and a method for producing multiprotein complexes. The invention also relates to a methods of assembling multiple single vectors ("vector entities") into fusion vectors and to method of disassembling a fusion vector containing multiple of such vector entities into single vectors. The invention is also directed to fusion vectors containing multiple vector entities.



Nucleic acids for cloning and expressing multiprotein complexes

The present invention relates to a nucleic acid containing at least one homing endonuclease site (HE) and at least one restriction enzyme site (X) wherein the HE and X sites are selected such that HE and X result in compatible cohesive ends when cut by the homing endonuclease and restriction enzyme, respectively, and the ligation product of HE and X cohesive ends can neither be cleaved by the homing endonuclease nor by the restriction enzyme. Further subject-matter of the present invention relates to a vector comprising the nucleic acid of the present invention, host cells containing the nucleic acid and/or the vector, a kit for cloning and/or expression of multiprotein complexes making use of the vector and the host cells, a method for producing a vector containing multiple expression cassettes, and a method for producing multiprotein complexes. The invention also relates to a method for assembling multiple single vectors ("vector entities") into fusion vectors and to a method for disassembling a fusion vector containing multiple of such vector entities into lower order fusion vectors and/or into single vectors. The invention is also directed to fusion vectors containing multiple vector entities.

Many vital processes in cells are controlled by proteins associating into interlocking molecular machines, in higher eukaryotes often containing 10 and more subunits (Rual, J.F. *et al. Nature* **437**, 1173-1178 (2005); Charbonnier, S., Gallego, O. and Gavin, A.C. *Biotechnol. Annu. Rev.* **14**, 1-28 (2008)). This has profound consequences for functional and structural studies that now aim to decipher physiologically relevant molecular mechanisms. Consequently, work on complexes is increasingly becoming imperative in contemporary biology. The low abundance and frequently heterogeneous nature of many multisubunit complexes, however, often preclude extraction from source.

Recombinant production methods certainly have had a decisive impact on life science research. In particular *E. coli*, as an expression host, is commonplace. Successful functional analysis of proteins and elucidation of their molecular architecture often crucially depends on introducing alterations, such as truncations, mutations and extension with purification tags, or with particular promoter/terminator elements. The ensuing requirements in terms of experimental throughput are already considerable for diversifying single open reading frames (ORFs). In particular structural genomics consortia demand the standardization of subcloning routines and implementation of automation for this. The exponential increase in workload when many ORFs have to be rapidly diversified and assembled in the context of a multisubunit complex is daunting, and an unresolved challenge to date.

A number of systems have been introduced in recent years for expression of several genes in eukaryotic and prokaryotic hosts; see, e.g. Fitzgerald et al. (2006) *Nat. Methods* **3**, 1021-1032; Tan et al. (2005) *Protein Expr. Purif.* **40**, 385-395 (2005); Tolia, N.H. and Joshua-Tor (2006). *Nat. Methods* **3**, 55-64; Chanda et al. (2006) *Protein Expr. Purif.* **47**, 217-224; Scheich et al. (2007). *Nucleic Acids Res.* **35**, e43 (2007). In spite of considerable improvements of eukaryotic expression systems, in particular the baculovirus/insect cell expression (Fitzgerald et al. (2006), *supra*), *E. coli* still remains to date the dominant work-horse in most laboratories, for many good reasons such as low-cost and availability of a multitude of specialized expression strains. The current co-expression systems for *E.coli* rely essentially on serial, mostly conventional (*i.e.* restriction/ligation) subcloning of encoding genes either as single expression cassettes (Tolia et al. (2006), *supra*; Chanda et al. (2006), *supra*) or as polycistrons constituting several genes under the control of the same promoter (Tan et al. (2005), *supra*). This considerably limits the applicability of these co-expression techniques for production of protein complexes with many subunits, in particular at the throughput typically required for structural molecular biology.

A major impediment of such largely serial (one gene at a time) constructions stems from the inherent inflexibility with regards to rapidly revising an expression experiment once the multiprotein complex has been produced, purified and characterized. However, such revisions, including variations of the protein subunits, are a *sine qua non* in contemporary functional and structural research.

Fitzgerald et al. (2006), *supra*, and WO-A-2005/085456 describe polynucleotides having a so-called multiplication module wherein two expression cassettes in head-to-head, head-to-tail or tail-to-tail orientation are flanked by specifically designed
5 pairs of restriction enzyme sites allowing iterative cloning of multiple genes into the expression cassettes.

In view of the draw backs of prior art constructs it is therefore the technical problem underlying the present invention to provide versatile systems for cloning and
10 expression of multiprotein complexes.

The solution to the above technical problem is achieved by the provision of the embodiments of the present invention as defined in the claims.

15 In particular, the present invention relates to a nucleic acid (or polynucleotide) containing at least one homing endonuclease site (HE) and at least one restriction enzyme site (X) wherein the HE and the X sites are selected such that HE and X result in compatible cohesive ends when cut by the homing endonuclease and restriction enzyme, respectively, and the ligation product of HE and X cohesive ends
20 can neither be cleaved by the homing endonuclease nor the restriction enzyme.

According to the present invention, the terms "nucleic acid" and "polynucleotide" are used interchangeably and refer to DNA, RNA or species containing one or more nucleotide analogues. Preferred nucleic acids or polynucleotides according to the
25 present invention are DNA, most preferred double-stranded (ds)DNA.

Preferably, the nucleic acid of the present invention has the following sequence elements:

30 HE-Prom-MCS-Term-X or HE-Prom-MCS-X

wherein

Prom: represents a promoter;

MCS: represent a multiple cloning site; and

Term: represents a terminator.

The above arrangement is hereinafter often referred to as “multiple integration element” (MIE).

5 Promoters useful in the present invention include, but are not limited to, promoters of prokaryotic, viral, mammalian, or insect cell origin or a combination thereof. Likewise, terminators useful in a nucleic acid according to the invention include, but are not limited to, terminators of prokaryotic, viral, mammalian, insect cell origin or a combination thereof. The term “multiple cloning site” according to the present
10 invention means a sequence having at least one restriction enzyme site different from the site X as defined above. The MCS according to the present invention may, e.g. be derived from the multiple cloning sites of any commercially available plasmid.

Preferred prokaryotic promoters are Lac, T7, arabinose and trc promoters. Further
15 promoters useful in the context of the present invention are viral promoters, in particular baculoviral promoters such as polh, p10 and p_{XIV} very late baculoviral promoters, vp39 baculoviral late promoter, vp39 polh baculoviral late/very late hybrid promoter, P_{cap/polh}, *pcna*, *etl*, *p35*, *egt*, *da26* baculoviral early promoters. Further promoters useful in the context of the present invention are the promoter sequences
20 CMV, SV40, UbC, EF-1 α , RSVLTR, MT, P_{DS47}, Ac5, P_{GAL} and P_{ADH}.

Examples of terminator sequences useful in the context of the present invention are T7, SV40, HSVtk or BGH.

25 The multiple cloning site according to the present invention may contain, in addition to the at least one restriction enzyme site (other than X), one or more, especially 1 to 4 homology regions. The restriction enzymes sites contained in the MCS can easily be chosen by the skilled person and examples of such sites together with their recognition sequences can be taken from the latest product catalogue of New
30 England Biolabs, Ipswich, MA, USA.

A “homing endonuclease” according to the present invention is a DNase specific for double-stranded DNA having a large, isometric recognition site of e.g. 12-40 base pairs or even more, preferably 20 to 30 base pairs. For a recent review with regard to

homing endonucleases, see Stoddard B.L. (2005) Q. Rev. Biophys. 38, 49-95. Due to the length of HE recognition sequences it is highly unlikely that a corresponding site occurs in the nucleotide sequence of a gene or polygene (or any other nucleotide sequence of any origin) to be inserted into the constructs according to the present invention making this strategy particularly useful for cloning larger and/or many genes of interest ("GOI").

A preferred HE site according to the present invention is a recognition sequence of a homing endonuclease that results in a 4 nucleotide overhang when cut by the respective homing endonuclease.

Examples of such HE sites include, but are not limited to, recognition sequences of PI-SceI, I-CeuI, I-PpoI, I-HmuI I-CreI, I-DmoI, PI-PfuI and I-MsoI, PI-PspI, I-SceI, other LAGLIDAG group members and variants thereof, SegH and Hef or other GIY-YIG homing endonucleases, I-ApeI, I-AniI, Cytochrome b mRNA maturase bl3, PI-TliI and PI-TfuI, PI-ThyI and others; see also Stoddard (2005), *supra*.

A preferred restriction enzyme site X according to the present invention compatible with HE sites producing a 4 bp overhang (examples are given above) is a BstXI site.

Corresponding enzymes are commercially available, e. g. from New England Biolabs Inc., Ipswich, MA, USA.

Especially preferred MIEs of the invention containing prokaryotic promoters/terminators have one of the following structures:

I-CeuI – T7 Prom – MCS - T7 Term - BstXI

PI-SceI – T7 Prom – MCS - T7 Term – BstXI

Especially preferred MIEs of the invention containing baculoviral promoters have one of the following structures:

I-CeuI – p10 – MCS - BstXI

PI-SceI – p10 – MCS – BstXI

I-CeuI – polh – MCS – BstXI

PI-SceI – polh – MCS – BstXI

Particularly preferred examples of nucleic acids according to the present invention comprise the sequence according to SEQ ID NO: 1 (for a detailed map see Fig. 13A and B; the sequence antisense to SEQ ID NO: 1 is outlined in SEQ ID NO: 54), SEQ ID NO: 50 (restriction map: Fig. 42), SEQ ID NO: 51 (restriction map: Fig. 43), SEQ ID NO: 52 (restriction map: Fig. 44) or SEQ ID NO: 53 (restriction map: Fig. 45).

In preferred embodiments of the present invention, the above-defined nucleic acid additionally comprises at least one site for integration of the nucleic acid into a vector or host cell. The integration site may allow for a transient or genomic incorporation.

With respect to the integration into a vector, in particular into a plasmid or virus, the integration site is preferably compatible for integration of the nucleic acid into an adenovirus, adeno-associated virus (AAV), autonomous parvovirus, herpes simplex virus (HSV), retrovirus, rhadinovirus, Epstein-Barr virus, lentivirus, semliki forest virus or baculovirus.

Particularly preferred integration sites that may be incorporated into the nucleic acid of the present invention can be selected from the transposon element of Tn7, λ -integrase specific attachment sites and site-specific recombinases (SSRs), in particular LoxP site or FLP recombinase specific recombination (FRT) site. Further preferred mechanisms for integration of the nucleic acid according to the invention are specific homologous recombination sequences such as lef2-603/Orf1629.

In further preferred embodiments of the present invention, the nucleic acid as described herein additionally contains one or more resistance markers for selecting against otherwise toxic substances. Preferred examples of resistance markers useful in the context of the present invention include, but are not limited to, antibiotics such as ampicillin, chloramphenicol, gentamycin, spectinomycin, and kanamycin resistance markers.

The nucleic acid of the present invention may also contain one or more ribosome binding site(s) (RBS), preferably integrated into an MIE as defined above.

Further subject-matter of the present invention relates to a vector comprising a nucleic acid as defined above.

Preferred vectors of the present invention are plasmids, expression vectors, transfer
5 vectors, more preferred eukaryotic gene transfer vectors, transient or viral vector-mediated gene transfer vectors. Other vectors according to the invention are viruses such as adenovirus vectors, adeno-associated virus (AAV) vectors, autonomous parvovirus vectors, herpes simplex virus (HSV) vectors, retrovirus vectors, rhadinovirus vectors, Epstein-Barr virus vectors, lentivirus vectors, semliki forest virus
10 vectors and baculovirus vectors.

Baculovirus vectors suitable for integrating a nucleic acid according to the invention (e.g. present on a suitable plasmid such as a transfer vector) are also subject matter of the present invention and preferably contain site-specific integration sites such as
15 a Tn7 attachment site (which may be embedded in a lacZ gene for blue/white screening of productive integration) and/or a LoxP site. Further preferred baculovirus according to the invention contain (alternative to or in addition to the above-described integration sites) a gene for expressing a substance toxic for host flanked by sequences for homologous recombination. An example for a gene for expressing a
20 toxic substance is the diphtheria toxin A gene. A preferred pair of sequences for homologous recombination is e.g. Isf2-603/Orf1629. The baculovirus can also contain further marker gene(s) as described above, including also fluorescent markers such as GFP, YFP and so on. Specific examples of corresponding baculovirus of the invention have the structure of EMBac, EMBacY, EMBac_Direct
25 and EMBacY_Direct as disclosed in the schemes according to Figs. 38, 39, 40 and 41, respectively.

Vectors useful in prokaryotic host cells comprise, preferably besides the above-exemplified marker genes (one or more thereof), an origin of replication (ori).
30 Examples are BR322, ColE1, and conditional origins of replication such as OriV and R6Ky, the latter being a preferred conditional origin of replication which makes the propagation of the vector of the present application dependent on the *pir* gene in a prokaryotic host. OriV makes the propagation of the vector of the present application dependent on the *trfA* gene in a prokaryotic host.

Furthermore, the present invention is directed to a host cell containing the nucleic acid of the invention and/or the vector of the present invention.

5 The host cells may be prokaryotic or eukaryotic. Eukaryotic host cells may for example be mammalian cells, preferably human cells. Examples of human host cells include, but are not limited to, HeLa, Huh7, HEK293, HepG2, KATO-III, IMR32, MT-2, pancreatic β -cells, keratinocytes, bone-marrow fibroblasts, CHP212, primary neural cells, W12, SK-N-MC, Saos-2, WI38, primary hepatocytes, FLC4, 143TK, 10 DLD-1, embryonic lung fibroblasts, primary foreskin fibroblasts, MRC5, and MG63 cells. Further preferred host cells of the present invention are porcine cells, preferably CPK, FS-13, PK-15 cells, bovine cells, preferably MDB, BT cells, bovine cells, such as FLL-YFT cells. Other eukaryotic cells useful in the context of the present invention are *C. elegans* cells. Further eukaryotic cells include yeast cells 15 such as *S. cerevisiae*, *S. pombe*, *C. albicans* and *P. pastoris*. Furthermore, the present invention is directed to insect cells as host cells which include cells from *S. frugiperda*, more preferably Sf9, Sf21, Express Sf+, High Five H5 cells, and cells from *D. melanogaster*, particularly S2 Schneider cells. Further host cells include *Dictyostelium discoideum* cells and cells from parasites such as *Leishmania spec.*

20 Prokaryotic hosts according to the present invention include bacteria, in particular *E. coli* such as commercially available strains like TOP10, DH5 α , HB101 etc.

25 The person skilled in the art is readily able to select appropriate vector construct/host cell pairs for appropriate propagation and/or transfer of the nucleic acid elements according to the present invention into a suitable host. Specific methods for introducing appropriate vector elements and vectors into appropriate host cells are equally known to the art and methods can be found in the latest edition of Ausubel et al. (ed.) Current Protocols In Molecular Biology, John Wiley & Sons, New York, USA.

30 In preferred embodiments of the present invention, the vector as defined above additionally comprises a site for site specific recombinases (SSRs), preferably one or more LoxP sites for Cre-lox specific recombination. In further preferred embodiments,

the vector according to the present invention comprises a transposon element, preferably a Tn7 attachment site.

It is further preferred that the attachment site as defined above is located within a marker gene. This arrangement makes it feasible to select for successfully integrated sequences into the attachment site by transposition. According to preferred embodiments, such a marker gene is selected from luciferase, β -GAL, CAT, fluorescent encoding protein genes, preferably GFP, BFP, YFP, CFP and their variants, and the lacZ α gene.

Particularly preferred embodiments of the vector according to the present invention have a sequence selected from the group consisting of SEQ ID NO: 2 to SEQ ID NO: 17.

Further preferred embodiments of the present invention are vectors containing more than one of the sequence elements of the nucleic acids of the present invention as defined above and, optionally, additionally containing more than one recombination sequence for a site specific recombinase, e.g. 2 to 6, more preferred 2, 3 or 4 of such recognition sequences, preferably 2 to 6, especially preferred 1 to 4 loxP sites.

A particularly preferred example of such a vector has the sequence of SEQ ID NO. 18.

It is to be understood that, if the vector of the present invention contains more than one recombination sequences, these can be recognition sequences of the same or different site-specific recombinases.

Further subject-matter of the present invention is a kit for cloning and/or expression of multiprotein complexes containing at least one vector as defined above together with at least host cell suitable for the propagation of said vector(s). Preferred host cells have been already described above. Preferably, the kit of this aspect of the present invention additionally contains a site-specific recombinase such as Cre.

The present invention also relates to a method for producing a vector containing multiple expression cassettes comprising the steps of:

- (a) inserting one or more genes between the HE and the X site of a first vector of the present invention;
- 5 (b) inserting one or more genes between the HE and the X site of a second vector as defined herein;
- (c) cleaving the first vector with a homing endonuclease specific for site HE and with a restriction enzyme specific for site X yielding a fragment containing the at least one gene flanked by the cleaved HE and X sites;
- 10 (d) cleaving the second vector with a homing endonuclease specific for site HE;
- (e) ligating the fragment obtained in step (c) into the cleaved second vector obtained in step (d) generating a third vector; and optionally
- (f) repeating steps (a) to (e) with one or more vector(s) generating a vector containing multiple genes.

15

According to preferred embodiments of the present invention it is possible to insert one or more genes into the vectors of the invention by methods known to the skilled person, e.g. by restriction enzyme digestion/ligation via compatible sites within the MCS or by recombination, preferably using the optionally present homology
20 region(s), preferably using the SLIC method. If more than one gene is inserted, these can be provided as single expression cassettes. However, it is clear for the skilled person that the (several or multiple) genes can be present as a polygene within in one ORF.

25 The present invention is further directed to a method for producing multiple protein complexes comprising the steps of

- (i) producing a vector containing multiple expression cassettes by the method as defined above;
- (ii) introducing the vector obtained in step (i) into a suitable host cells such as the
30 host cells described above; and
- (iii) incubating the host cell under conditions allowing the simultaneous expression of the genes present in the vector.

The introduction of the vector into suitable host cells (as exemplified above) is carried out by methods known to the skilled person (see, e.g. Ausubel et al. (ed.), *supra*).

A further aspect of the present application is a fusion vector comprising n vector entities separated from each other by n of the same site-specific recombination site wherein each vector entity contains an individual resistance marker gene different from the resistance marker genes of the other vector entities, wherein n is an integer of at least 3.

A “single vector” or “vector entity” according to the present aspect of the invention is generally a nucleic acid suitable for integration of foreign genetic elements (in particular, one or more genes of interest) into host cells and which are suited for amplification. Typical examples are plasmids, bacmids, viruses, lambda vectors, cosmids etc. Preferred examples of one or more of the above vector categories are outlined in more detail above with respect to the HE/X site containing vector which definitions are also valid for this aspect of the present invention.

It is clear for the skilled person that the number of vector entities to be assembled into a fusion vector according to the present invention (or disassembled from such a fusion vector; with respect to methods of assembly/disassembly see below) is generally not specifically limited as long as a corresponding number of resistance markers is available. With respect to practical considerations, the number n in the context of the present invention is preferably 3, 4, 5 or 6, (but may be more) which in part depends on the size of constructs that can be propagated in the host.

The present invention furthermore relates to a kit for assembly and/or disassembly of n vectors comprising

a fusion vector comprising n vector entities separated from each other by n of the same site-specific recombination site wherein each vector entity contains an individual resistance marker gene different from the resistance marker genes of the other vector entities; and/or

n vectors (vector entities) each containing a site-specific recombination site and an individual resistance marker gene different from the resistance marker genes of the other vectors,

wherein n is an integer of at least 3; and

a recombinase specific for said site-specific recombination site and/or cells for the propagation of said fusion vector and/or said n vectors.

- 5 Preferred embodiments of the above fusion vector and vector kits are or contain, respectively, fusion vector(s) and/or vector entities comprising LoxP sites and Cre as the corresponding recombinase enzyme. Other examples of site-specific recombination sites/recombinases are FRT sites and the corresponding enzyme (FLP recombinase).

10

According to a preferred embodiment the above-defined n vectors or vector entities, respectively, each contain one or more expression cassettes of the form Prom-MCS-Term or Prom-MCS-Term (definitions are as defined above, preferably between a HE and restriction enzyme site X as defined above). It is further preferred that the
15 expression cassette preferably present in the vectors or vector entities, respectively, contains one or more genes of interest ("GOI").

Examples of resistance marker genes (or simply "resistance markers") useful in the context of this aspect of the present invention are as already defined above.

20

An especially preferred example of the fusion vector as defined above is vector pACKS (SEQ ID NO: 18) described in more detail below.

- Preferred examples of the vector entities are pACE (SEQ ID NO: 2), pACE2 (SEQ ID
25 NO: 3), pDC (SEQ ID NO: 4), pDK (SEQ ID NO: 5) and pDS (SEQ ID NO: 6), which are all adapted for expression in prokaryotic hosts, and pIDC (SEQ ID NO: 7), pIDK (SEQ ID NO: 8), pIDS (SEQ ID NO: 9), pACEBac1 (SEQ ID NO: 10), pACEBac2 (SEQ ID NO: 11), pACEBac3 (SEQ ID NO: 12), pACEBac4 (SEQ ID NO: 13), pOmniBac1 (SEQ ID NO: 14), pOmniBac2 (SEQ ID NO: 15, pOmniBac3 (SEQ ID
30 NO: 16) and pOmniBac4 (SEQ ID NO: 17), which are tailored for expression in insect cells using baculovirus. The above preferred examples of vector entities are described in more detail below.

It is further preferred that at least one of the vector entities (and/or of the individual vectors in the above kit) contains a further selectable marker different from the resistance marker genes. An example is a conditional origin of replication making the propagation of the respective vector entity dependent on a specific genetic background in a host. An example is an Ori derived from (or being) R6ky making the propagation of the vector dependent on the *pir* gene.

The present invention further provides a method for assembling n vector entities into 1 to $(n-1)$ fusion vectors wherein said fusion vector(s) contain(s) 2 to n of said vector entities comprising the steps of:

- (1) contacting n vector entities each containing a site-specific recombination site and an individual resistance marker different from the resistance markers of the other vector entities with a recombinase specific for said site-specific recombination site so as to generate a mixture of fusions of the vector entities comprising 2 to n of said vector entities,
- (2) transforming said mixture into host cells;
- (3) culturing one or more sample(s) of the transformed cells in the presence of the appropriate combination of antibiotics for selecting a desired fusion vector containing 2 to n vector entities.
- (4) obtaining n single clones of transformed cells from the culture obtained in step (3) in which these were viable in the presence of the respective combination of antibiotics; and
- (5) culturing n samples of each of said n single clones in the presence of each of n antibiotics specific for the n individual resistance markers present in said n vector entities;

wherein n is as defined above.

If it is desired to select for more than one desired vector fusions, the transformed cells obtained in above step (2) are divided into the appropriate number of aliquots or samples. For example, if it is desired to select all possible $(n! - n)$ vector fusions (i.e. the single vector entities as educts of the above method are not selected for), the transformed host cells are divided into $(n! - n)$ aliquots (or samples) and each aliquot is cultured in the presence of the appropriate antibiotics.

In the context of the present invention, the term "aliquot" as used herein does not necessarily mean that the aliquots have the same volume or number of cells. Rather, each of the aliquots or samples may have the same or different volumes or number of cells.

5

The term "culturing" the transformed cells or the aliquot (or sample) means that the transformed cells are incubated under the appropriate conditions for viability of the host cells. For example, the transformed host cells may be used to inoculate a (e.g. larger) volume of liquid culture medium or the aliquot may be plated out on an appropriate solid medium.

10

If the vector assembly method as defined above is used to select for more than one desired vector fusion, e.g. if all possible fusions are desired, the selection step (3) is preferably carried out using typical well plate formats such as 96-well plates.

15

According to a preferred embodiment of the present vector assembly method (n-1) of the vector entities to be fused each contains a further selectable marker different from the resistance marker. Such vector entities are hereinafter referred to as "Donor" vectors, since, when fused to a vector entity which does not contain said selectable marker different from the resistance marker (hereinafter referred to as "Acceptor"), in a fusion between the Donor(s) and the Acceptor, said Donor(s) provide host cells with a phenotype that allows only the propagation of Acceptor-Donor fusions but not Donor-Donor fusions. Preferred examples of such a selectable marker are conditional origins of replication making the propagation of the Donor dependent on a specific genetic background. A specific example of such a selectable marker is R6K γ Ori making the propagation of the Donor dependent on the presence of the *pir* gene in a bacterial host such as *E. coli*. In this case, the mixture obtained in step (i) of the above vector assembly method is transformed into bacterial cells lacking the *pir* gene (such *E. coli* strains TOP10, DH5 α , HB101 or other commercially available *pir*⁻ cells).

20

25

30

A preferred embodiment of the above-defined vector assembly method is described in more detail below (ACEMBL system; Section C.2.1)

According to a preferred embodiment of the above-defined method, the n vector entities, respectively, each contain one or more expression cassettes of the form Prom-MCS-Term or Prom-MCS (as defined above, preferably between a HE and restriction enzyme site X as defined above). It is further preferred that the expression cassette preferably present in the vectors or vector entities, respectively, contains one or more genes of interest ("GOI") to be expressed in a suitable host.

Another method for providing fusion vectors according to the present invention is a sequential assembly process wherein in the first step two of the vector entities are recombined, transformed into host cells and the host cells cultured in the presence of two antibiotics. The second round comprises the isolation of the double fusion vector ($n = 2$) from a viable clone, contacting with a third vector entity in the presence of the respective recombinase, transformation into host cells and selection for the three resistance markers present in the triple fusion vector ($n = 3$) and so on until the desired multifusion vector is reached.

Of course, it is also possible to provide fusion vectors according to the invention, in particular fusion vectors of higher order (i.e. $n > 3$) by a combined approach using the vector assembly method of steps (1) to (5) as defined above (e.g. for assembling a fusion vector with $n = 3, 4$ or 5) and then adding one or more further vector entities sequentially as described in the previous paragraph.

The principle underlying the above-described method for assembling a fusion vector, i.e. the equilibrium of educts and products in recombination reactions, can equally be applied to the disassembly of fusion vectors.

Therefore, the invention further provides a method of disassembling a fusion vector containing n vector entities into one or more desired fusion vectors selected from the group consisting of fusion vectors containing 2 to $(n-1)$ vector entities or into one or more desired single vector entities, wherein in said fusion vector containing n vector entities said n vector entities are separated from each other by n site-specific recombination sites and each vector entity contains an individual resistance marker different from the resistance markers of the other vector entities, comprising the steps of:

- (A) contacting the fusion vector containing n vector entities with a recombinase specific for said site-specific recombination sites in order to generate a mixture of fusion vectors comprising 2 to $(n-1)$ of said vector entities and single vector entities;
- 5 (B) transforming said mixture into host cells; and
- (C) culturing one or more sample(s) of the transformed cells in the presence of
- (C1) an appropriate combination of antibiotics for selecting one or more desired fusion vector(s) containing 2 to $(n-1)$ vector entities; and/or
- (C2) a single appropriate antibiotic for selecting a desired single vector entity;
- 10 (D) obtaining n single clones of transformed cells from the sample of the transformed cells in which these were viable in the presence of the respective antibiotic or combination of antibiotics, respectively; and
- (E) culturing n samples of each of said n single clones in the presence of each of n antibiotics specific for the n individual resistance markers present in said n
- 15 vector entities;

wherein n is as defined above.

If it is desired to select for single vectors using the above fusion vector disassembly method, it is preferred that steps (A), (B) and (C1) to (E) are carried out for selecting

20 an appropriate fusion vector containing 2 to $(n-1)$ vector entities and then to perform steps (A), (B) and (C2) to (E) are carried out with said selected fusion vector containing 2 to $(n-1)$ vector entities. It is understood that this sequential approach can be repeated which is especially preferred when starting from a fusion vector containing a higher number of vector entities, i.e. one can select for a $(n-1)$ fusion

25 vector in the first, then for a $(n-2)$ construct in the second round and so on, e.g. until reaching a fusion vector with $n = 3$ or 2 such that the presence of the single vector entities in the recombinase reaction equilibrium makes the selection of respective clones containing said single vector entities according to the selection steps (C2) to (E) more likely.

30

Furthermore, in analogy to the above-defined vector assembly method, it is preferred in the fusion vector disassembly method of the present invention that $(n-1)$ of the vector entities in said fusion vector containing n vector entities each contains a further selectable marker different from the resistance markers such that only host

cells transformed with fusions between a vector entity not containing the further selectable marker and one or more of the vector entities containing the selectable marker are viable in step (C1).

- 5 With respect to preferred selectable markers (conditional Ori etc.), host cells, the use of multi well test plates etc. it is referred to the preferred embodiments of the vector assembly method outlined above.

10 The fusion vector disassembly method of the present application is further elaborated below with respect to a preferred embodiment (ACEMBL system; Section C.2.2).

The nucleic acids and vectors (including fusion vectors and single vectors (i.e. vector entities)) of the present invention may contain further typical sequence elements, e.g. elements that enable or simplify the detection and/or purification of the (multiple)
15 proteins expressed from the one or more genes of interest. Typical examples of such elements are sequences coding for GFP and its derivatives, His-tags, GST etc.

Fusion vectors according to the present invention are advantageously used for the expression of mutliprotein complexes in a suitable host. Thus, the present invention
20 further provides a corresponding process comprising transforming a fusion vector of the invention (containing vector entities having inserted one or more genes of interest, e.g. in form of multiple or single expression cassettes, or in the form of polygenes as appropriate) into a suitable host and culturing the transformed host under conditions allowing simultaneous expression of the genes of interest.

25

From the disclosure of the various aspects of the present invention the skilled person readily understands that the HE/X site polynucleotide (in particular corresponding vectors), preferably used for iterative cloning of multiple expression cassettes, can be combined with the assembly (or disassembly) methods as defined above for creating
30 multigene constructs: For example, one or more of a single gene or multigene vector(s) can be prepared using the HE/X site elements as described which may then be assembled into fusion vectors of choice (e.g. triple, quadruple or higher order fusion vectors) using the recombination-based assembly methods defined herein. Such fusion vectors may then be (partly or completely) disassembled as disclosed

herein and different constructs can be assembled in turn as appropriate for the respective multiprotein application envisaged by the skilled person. Thus, the aspects of the present invention represent a building block system which provides the person skilled in the art with a hitherto unknown freedom of combining multiple genes (or polygenes) of interest for multiprotein applications.

The figures show:

Fig. 1 shows a schematic overview of preferred vectors according to the present invention for expression of multiprotein complexes in prokaryotic hosts contained in a preferred kit called "ACEMBL".

Fig. 2 is a graphic representation of a preferred embodiment of the nucleic acid of the present invention called "multiple integration element" (MIE).

Fig. 3 shows a schematic overview of a preferred method for inserting a gene of interest ("GOI") into a vector of the present invention by sequence and ligation independent cloning (SLIC; see Tan, S. et al. *Protein Expr. Purif.* **40**, 385 (2005)). A gene of interest (GOI 1) is PCR amplified with specific primers and integrated into a vector (Acceptor, Donor) linearized by PCR with complementary primers (complementary regions are shaded in light gray or dark grey, respectively). Resulting PCR fragments contain homology regions at the ends. T4 DNA polymerase acts as an exonuclease in the absence of dNTP and produces long sticky overhangs. Mixing (optionally annealing) of T4 DNA polymerase exonuclease treated insert and vector is followed by transformation, yielding a single gene expression cassette.

Fig. 4 shows a schematic overview of a preferred method for inserting a polycistron into a vector of the present invention by SLIC. Genes of interest (GOI 1, 2, 3) are PCR amplified with specific primers and integrated into a vector (Acceptor, Donor) linearized by PCR with primers complementary to the ends of the forward primer of the first (GOI 1) and the reverse primer of the last (GOI 3) gene to be assembled in the polycistron (complementary regions are shaded in light gray or dark grey, respectively). Resulting PCR

fragments contain homology regions at the ends. T4 DNA polymerase acts as an exonuclease in the absence of dNTP and produces long sticky overhangs. Mixing (optionally annealing) of T4 DNA polymerase exonuclease treated insert and vector is followed by transformation, yielding a polycistronic expression cassette.

Fig. 5 shows the sequence of a LoxP imperfect inverted repeat (SEQ ID NO: 19).

Fig. 6 (left panel) shows a schematic representation in form of a pyramid illustrating Cre-mediated assembly and disassembly of preferred embodiments of the vector of the present invention (pACE, pDK and pDS vectors). LoxP sites are shown as red circles, resistance markers and origins are labelled. White arrows stand for the entire expression cassette (including promoter, terminator and multiple integration elements) in the ACEMBL vectors. Not all possible fusion products are shown for clarity. Levels of multiresistance are indicated in the right panel.

Fig. 7 is a schematic representation of a multiresistance analysis of bacterial colonies carrying vector constructs resulting from Cre-deCre assembly/disassembly according to the invention (cf. also Fig. 12).

Fig. 8 shows a schematic representation of the strategy for cloning of human RAP74 and human RAP30 into vectors of the present invention for expression of human TFIIIF (left panel). hRAP74 was cloned by SLIC into pDC. hRAP30 was cloned by SLIC into pACE. Cre-Lox recombination of pDC-RAP74 (donor) and pACE-RAP30 results in vector pACEMBL-hTFIIIF. Results from restriction mapping by BstZ17I/BamHI double digestion of 11 double resistant (Cm, Ap) colonies are shown by a gel section from 1% E-gel electrophoresis (M: NEB 1kb DNA marker). All clones tested showed the expected pattern (5.0 + 2.8 kb) (left panel).

Fig. 9 illustrates the strategy for cloning of human VHL/elongin b/elongin c complex (VHLbc) (tracistron) into vector pACE by multifragment SLIC.

Fig 10 shows a schematic representation of the strategy for iterative cloning of the components of yeast RES complex (Pml1p, Snu17p, Bud13p) using a preferred homing endonuclease site (HE)/restriction enzyme site (X) module (PI-SceI/BstXI) according to the present invention.

Fig. 11 shows a schematic representation of the generation of single vectors from multifusion vector pACKS (SEQ ID NO: 18).

Fig. 12 shows schematic representations and photographs illustrating a 96 well microtiter analysis of pACKS De-Cre reaction.

Fig. 13 shows the sequence and map of a preferred nucleic acid ("multiple integration element", MIE) according to the present invention (SEQ ID NO: 1). Forward and reverse primers for sequencing can be standard vector primers for T7 and lac. Adaptor primer sequences (see Table I) are indicated. DNA sequences in these homology regions contain tried-and-tested sequencing primers (Tan et al. (2005), *supra*). Sites of insertion (I1-I4) are shown. The adaptor sequences, and probably any sequence in the homology regions, can be used as adaptors for multifragment insertions. The ribosome binding site present in the MIE (rbs) is boxed in red.

Fig. 14 shows a plasmid map of Acceptor vector pACE.

Fig. 15 shows a plasmid map of Acceptor vector pACE2.

Fig. 16 shows a plasmid map of Donor vector pDC.

Fig. 17 shows a plasmid map of Donor vector pDK.

Fig. 18 shows a plasmid map of Donor vector pDS.

As can be seen in the above plasmid maps, Acceptor vectors pACE (Fig. 14) and pACE2 (Fig. 15) contain a T7 promoter and terminator. Donor vectors pDC (Fig. 16), pDK (Fig. 17) and pDS (Fig. 18) contain conditional origins of replication. pDS (Fig.

18) and pDK (Fig. 17) have a lac promoter. pDC (Fig. 16) has a T7 promoter. Resistance markers and origins of replication are shown. LoxP imperfect inverted repeat sequences are shown as circles. Homing endonuclease sites and corresponding BstXI sites are boxed. The restriction enzyme sites in the multiple
5 integration element (MIE) are indicated. All MIEs have the same DNA sequence between ClaI and PmeI. Differences in unique restriction site composition stem from differences in the plasmid backbone sequences.

Fig. 19 shows the results of a restriction mapping of preferred vectors according to
10 the invention. Both undigested Acceptor (pACE, pACE2) and Donor vectors (pDC, pDK, pDS) are shown as well as the same vectors digested with BamHI. All restriction reactions yield the expected sizes. Lanes 1-5 show uncut pACE, pACE2, pDC, pDK, and pDS vectors; lane M shows λ Styl marker; lanes A-E show BamHI digested pACE, pACE2, pDC, pDK, and
15 pDS vectors.

Fig. 20 shows the strategy for Acceptor/Donor recombineering according to the invention exemplified for genes coding for Von Hippel-Lindau/elonginB/elonginC (VHLbc) complex (Tan et al. (2005), *supra*; see
20 also Fig. 9 above), FtsH soluble domain (Bieniossek et al. (2006) *Proc. Natl. Acad. Sci. USA* **103**, 3066-3071), blue fluorescent protein (BFP), and green fluorescent protein (mGFP) with a coiled-coil domain (Berger et al. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 12177-12182) were inserted into pACE, pDC, pDK and pDS, respectively. Cre-fusion was followed by transformation
25 into *pir*⁻ cells (TOP10). Aliquots were plated on agar with two (Ap/Kn; Ap/Cm; Ap/Sp), three (Ap/Kn/Sp) and four (Ap/Kn/Sp/Cm) antibiotics. Four colonies from each plate were challenged in a 96 well microtiter plate. Labels left of the plate image denote antibiotics contained in media aliquots in horizontal rows. Wells in the bottom two rows were charged differently (labels below
30 the plate image). Those inoculated with four colonies each from one agar plate are boxed in black, and flagged with antibiotics contained in the agar plate. Four vertical rows in each such 16-well box were inoculated with the same colony. In the bottom two rows, four wells in a row were inoculated with the same colony. Expected vector architecture of the double, triple

(ADD) and quadruple (ADDD) fusions is shown left or right (16 well boxes), respectively, or below (bottom two rows) of the plate image. Red dye is used as positional marker. Deconstruction of the ADDD fusion was carried out successfully in the reverse approach.

5

Fig. 21 shows the results of multiprotein complex expression of human TFIIIF (Fig. 21A), the Von Hippel-Lindau/elonginB/elonginC (VHLbc) complex (Fig. 21B) and the prokaryotic transmembrane holotranslocon (HTL) YidC-SecYEGDF (Fig. 21C). (A) Human TFIIIF was assembled and purified using a TECAN Freedom EvoII 200 workstation, and analyzed by SDS-PAGE. Uninduced and induced whole cell extracts and purified hTFIIIF are shown, with subunits (RAP74, RAP30) marked. RAP74 contained a C-terminal oligohistidine tag. (B) All multigene constructs from Fig. 20 were assembled, expressed and cell lysates analyzed in parallel following the same routine as for hTFIIIF (labels as in Fig. 20). The VHLbc complex was captured by an oligohistidine-thioredoxin-fusion tag on the VHL subunit (Tan et al. (2005), *supra*). FtsH contained an oligohistidine tag at its C-terminus (Bieniossek et al. (2006, *supra*). Fluorescent proteins were identified in lysates by Western blot with antibody Roche 1814460 (1:1000 in TBST/3%BSA). (C) Production of the entire prokaryotic transmembrane holotranslocon (HTL) YidC-SecYEGDF. Membrane vesicle preparation, detergent solubilization, Ni²⁺ affinity capture and size exclusion chromatography resulted in purified holotranslocon complex (right). Subunits are labeled. A breakdown product of SecY is marked with an asterisk. In all panels, M stands for Biorad broad range marker (sizes in kDa).

10

15

20

25

Fig. 22 shows a schematic workflow for an automated SLIC process.

Fig. 23 shows a schematic workflow for an automated Cre fusion process.

30

Fig. 24 shows a plasmid map of a preferred vector (Donor vector) of the invention called pIDC (SEQ ID NO: 7).

Fig. 25 shows the plasmid map of a preferred vector (Donor vector) of the invention called pIDK (SEQ ID NO: 8).

Fig. 26 shows the plasmid map of a preferred vector (Donor vector) of the invention called pIDS (SEQ ID NO: 9).

Fig. 27 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pACEBac1 (SEQ ID NO: 10).

Fig. 28 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pACEBac2 (SEQ ID NO: 11).

Fig. 29 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pACEBac3 (SEQ ID NO: 12).

Fig. 30 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pACEBac4 (SEQ ID NO: 13).

Fig. 31 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pOmniBac1 (SEQ ID NO: 14).

Fig. 32 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pOmniBac2 (SEQ ID NO: 15).

Fig. 33 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pOmniBac3 (SEQ ID NO: 16).

Fig. 34 shows the plasmid map of a preferred vector (Acceptor vector) of the invention called pOmniBac4 (SEQ ID NO: 17).

Fig. 35 shows a scheme for multiprotein expression in insect cells by generating composite baculovirus using Acceptor vectors of the present invention carrying a ColE1 origin (pACEBac1, pACEBac2, pOmniBac1, pOmniBac2). Multigene fusions are generated by Cre-LoxP fusion of the desired

Donor/Acceptor combinations (multigene construction). The fusion vector is transformed in bacteria carrying a baculovirus genome (such as baculovirus EMBac or EMBacY) as a bacterial artificial chromosome (BAC). The vector fusion is integrated into the baculovirus genome by Tn7 based transposition. Productive composite viruses are selected by blue/white screening (integration of the vector fusion into the T7 attachment site of the virus destroys a lacZ gene present on the virus). Composite viruses are prepared and suitable insect cells are transfected for protein production.

Fig. 36 shows a scheme for multiprotein expression in insect cells by generating composite baculovirus using Acceptor vectors of the present invention carrying an OriV origin (pACEBac3, pACEBac4, pOmniBac3, pOmniBac4). Multigene fusions are generated by Cre-LoxP fusion of the desired Donor/Acceptor combinations. The fusion vector is transformed in bacteria carrying a baculovirus genome (such as baculovirus EMBac or EMBacY) as a bacterial artificial chromosome (BAC). The vector fusion is integrated into the baculovirus genome by Tn7 based transposition. Since the Acceptor vectors carrying an OriV can only be propagated, if a trfA gene is provided in trans, unproductive integration events in bacteria not containing the trfA gene leads to elimination of such transformants upon exposure to the appropriate antibiotic (here: gentamycin). Thus, blue/white screening is not necessary in this case. Composite viruses are then prepared and suitable insect cells are transfected for protein production.

Fig. 37 shows a scheme for multiprotein expression in insect cells by generating composite baculovirus using Acceptor vectors of the present invention carrying lef2-603 and Orf1629 homology sequences (pOmniBac1, pOmniBac2, pOmniBac3, pOmniBac4). Multigene fusions are generated by Cre-LoxP fusion of the desired Donor/Acceptor combinations (multigene construction). The multigene construct and genomic baculovirus DNA carrying a diphtheria toxin A gene flanked by the lef2-603/Orf1629 homology sequences can be directly co-transfected into suitable insect cells for protein production. Transformation of transfer vector into bacteria containing the

baculovirus genome, blue/white screening for composite viruses and preparation of composite viruses from the bacteria is no longer necessary.

Fig. 38 shows a schematic representation of a baculovirus vector according to the invention called EMBac.

Fig. 39 shows a schematic representation of a baculovirus vector according to the invention called EMBacY.

Fig. 40 shows a schematic representation of a baculovirus vector according to the invention called EMBac_Direct.

Fig. 41 shows a schematic representation of a baculovirus vector according to the invention called EMBac_DirectY.

Fig. 42 shows a schematic representation of an MIE according to the invention having the general structure I-CeuI – p10 – MCS – BstXI present, for example in Acceptor vectors such as pACEBac 2.

Fig. 43 shows a schematic representation of an MIE according to the invention having the general structure PI-SceI – p10 – MCS – BstXI present, for example in Donor vectors such as pIDS.

Fig. 44 shows a schematic representation of an MIE according to the invention having the general structure I-CeuI – polh – MCS – BstXI present, for example in Acceptor vectors such as pACEBac 1.

Fig. 45 shows a schematic representation of an MIE according to the invention having the general structure PI-SceI – polh – MCS – BstXI present, for example in Donor vectors such as pIDC.

Fig. 46 shows a schematic representation of vector pACEBac1-HisIKK1.

Fig. 47 shows a schematic representation of vector pIDC-CSIKK2.

Fig. 48 shows a schematic representation of vector pIDS-IKK3.

Fig. 49 shows a schematic representation of vector pACEBac-HA-NA.

Fig. 50 shows a schematic representation of vector pIDC-M1-M2.

The present invention is in the following further described in detail with reference to preferred embodiments designated as "ACEMBL" system.

A. Synopsis

The preferred embodiments according to the present invention denoted as "ACEMBL" provide a multi-expression system for multigene expression in *E. coli* and insect cells using the baculovirus system. ACEMBL can be used both manually and also in an automated setup by using a liquid handling workstation. ACEMBL applies tandem recombination steps for rapidly assembling many genes into multigene expression cassettes. These can be polycistronic or multiple expression modules, or a combination of these elements. ACEMBL also offers the option to employ conventional approaches involving restriction enzymes and ligase, if desired.

The following strategies for multigene assembly and expression are provided for in the ACEMBL system:

- (1) Single gene insertions into vectors (recombination or restriction/ligation)
- (2) Multigene assembly into a polycistron (recombination or restriction/ligation)
- (3) Multigene assembly using homing endonucleases
- (4) Multigene plasmid fusion by Cre-LoxP reaction
- (5) Multigene expression by cotransformation in *E. coli*
- (6) Multigene expression in insect cells using the baculovirus system

These strategies can be used individually or in conjunction, depending on the project and user.

In the following Section C, step-by-step protocols are provided for each of these methods for multigene cassette assembly that can be used in the ACEMBL system.

B. ACEMBL System

B.1 ACEMBL vectors

The present invention provides as preferred exemplary embodiments small *de novo* designed vectors which are called “Acceptor” and “Donor” vectors (Figs. 1 and 20; for plasmid maps, see Figs. 14 to 18 and Figs. 21 to 31). Acceptor vectors for expression of proteins in prokaryotic hosts (e.g. pACE, pACE2) contain origins of replication derived from ColE1 and resistance markers (ampicillin or tetracycline). Donor vectors contain conditional origins of replication (derived from R6Kγ), which make their propagation dependent on hosts expressing the *pir* gene. Donor vectors contain resistance markers kanamycin, chloramphenicol, and spectinomycin. Preferably, three Donor vectors are used in conjunction with one Acceptor vector.

All Donor and Acceptor vectors according the present example contain a LoxP imperfect inverted repeat and in addition, a multiple integration element (MIE). The preferred MIE of the invention comprises an expression cassette with a promoter of choice (prokaryotic, mammalian, insect cell specific or a combination thereof) and a terminator (prokaryotic, mammalian, insect cell specific or a combination thereof). In between is a DNA segment which contains a number of restriction sites that can be used for conventional cloning approaches or also for generating double-strand breaks for the integration of expression elements of choice (further promoters, ribosomal binding sites, terminators and genes). The MIE is completed by a homing endonuclease site and a specifically designed restriction enzyme site (BstXI) flanking the promoter and the terminator (see B.2.)

The sequences of ACEMBL vectors for expression in prokaryotic hosts are outlined in the sequence listing (pACE: SEQ ID NO: 2, pACE2: SEQ ID NO. 3; pDC: SEQ ID

NO: 4; pDK: SEQ ID NO: 5; pDS: SEQ ID NO: 6; pACKS: SEQ ID NO: 18). Maps of the vectors pACE, pACE2, pDC, pDK and pDS are shown in Figs. 14 to 18.

The ACEMBL system according to the present invention also provides Donor and
5 Acceptor vectors adapted for expression of multiprotein complexes in insect cells using baculovirus (pIDC (SEQ ID NO: 7), pIDK (SEQ ID NO: 8), pIDS (SEQ ID NO: 9), pACEBac1 (SEQ ID NO: 10), pACEBac2 (SEQ ID NO: 11), pACEBac3 (SEQ ID NO: 12), pACEBac4 (SEQ ID NO: 13), pOmniBac1 (SEQ ID NO: 14), pOmniBac2 (SEQ ID NO: 15), pOmniBac3 (SEQ ID NO: 16) and pOmniBac4 (SEQ ID NO: 17)).

10 Plasmid maps of the vectors are shown in Figs. 24 to 34.

Donor vectors pIDS, pIDK and pIDS contain a conditional origin of replication (from R6Kgamma phage), a homing endonuclease (HE) site (PI-SceI) and a complementary BstXI site (see the corresponding E.coli vectors pDC, pDK, pDS).

15 Donors are propagated in cell strains containing the *pir* gene.

In contrast to the versions adapted for expression in bacteria, the vectors for expression of proteins in insect cells do not contain prokaryotic promoter and terminator structures. Instead, they have either a polh expression cassette (polh EC)
20 or a p10 expression cassette (p10 EC). These expression cassettes contain common polyhedron or p10 promoters from AcMNPV, an oligonucleotide encoding for restriction sites (different from the MIE in the prokaryotic ACEMBL version) and either SV40 or HSVtk polyadenylation signal sequences.

25 Obviously, due to the HE and BstXI sites, the expression cassettes can be freely exchanged in between the vectors, also if they contain an inserted gene. This can be done by restriction ligation or by restriction enzyme/ligase independent methods (e.g. SLIC). Therefore, versions can be created at ease which contain a p10 or polh marker in combination with any one of the resistance markers (spectinomycin,
30 kanamycin, chloramphenicol, or others).

The HE/BstXI site combinations can be used to multiply expression cassettes or also to fit the vectors with combinations of p10 and polh expression cassettes.

All Donors contain a LoxP inverted imperfect repeat. This can be used for LoxP mediated constructions and deconstructions of Acceptor/Donor multifusions as described for the bacterial ACEMBL vectors.

- 5 The present embodiment of the invention relating to vectors adapted for protein expression in insect cells provides a number of Acceptor vectors in the baculovirus-version of ACEMBL. These share common features: all contain a LoxP site, a resistance marker (gentamycin) and again either a p10 or a polh expression cassette (identical to the ones present in the Donors).

10

The expression cassettes of the Acceptors are flanked by a homing endonuclease site (I-CeuI) and a corresponding BstXI site.

15

The expression cassettes can be exchanged in between the Acceptors and also multiplied or combined using the HE/BstXI combination as described for the bacterial ACEMBL vectors.

20

There are two families of Acceptors in terms of the origin used:

pACEBac1, pACEBac2, pOmniBac1 and pOmniBac2 contain all a ColEI origin of replication which allows propagation in all common E.coli cloning cell strains.

25

All Acceptor vectors contain the Tn7L and Tn7R sequences which enable integration of the region in between into a Tn7 attachment site by using the Tn7 transposition procedure.

30

pACEBac3, pACEBac4, pOmniBac3 and pOmniBac4 contain a conditional origin of replication (OriV) from *V. Cholerae* which is dependent on the trfA gene that needs to be provided in trans in the cloning strains used. The function of this OriV is to eliminate the background (blue colonies) when these Acceptors, fitted with genes and if required fused with Donors, are transformed into cells that contain the baculovirus genome in form of a bacterial artificial chromosome (i.e. DH10Bac from Invitrogen and similar). Here, the Tn7 transposition system is used to integrate the regions in between Tn7L and Tn7R of the DNA transformed into the cells into a Tn7 attachment site on the viral genome of choice. Normally, unproductive integration

events would result in blue colonies (if the Tn7 attachment site is embedded in a LacZalpha gene on the baculovirus genome). These blue colonies propagate the plasmid transformed outside of the baculovirus genome. With these four OriV containing plasmids, the blue colonies cannot survive upon exposure to Gentamycin (since the DH10Bac or other cells do not contain trfA) and only white colonies are produced, which all contain productively integrated composite bacmid carrying the heterologous genes provided on the plasmid transformed; see also the scheme in Fig. 36.

The Acceptor vectors pOminBac1-4 contain, in addition to the Tn7L and Tn7R regions, also the lef2-603 and Orf1629 homology sequences. These are used for homologous recombination procedures for generating composite baculovirus as used by the Novagen Bacvector series, the Baculogold system from Pharmingen, FlashBac from OET and others. Thus, these Acceptor vectors can be used for every baculovirus system that is currently available, including the Tn7 based baculoviruses and all viruses relying on lef2,603/1629 homologous recombination procedures, for expressing heterologous genes in insect cell cultures; see also the scheme in Fig. 37.

B.2 Multiple integration element (MIE)

A preferred multiple integration element (MIE) according to the invention was derived from a polylinker (see Tan et al. (2005) *supra*) and allows for several approaches for multigene assembly (see Section C below). Multiple genes can be inserted into the MIE of any one of the vectors by a variety of methods, for example BD-In-Fusion recombination (see ClonTech TaKaRa Bio Europe, www.clontech.com) or SLIC (sequence and ligation independent cloning; see Li et al. (2007) *Nat. Methods* **4**, 251). For this, the vector needs to be linearized, which can also be carried out efficiently by PCR reaction with appropriate primers, since the vectors are all small (2-3.0 kb). Use of ultrahigh-fidelity polymerases such as Phusion (Finnzymes/New England BioLabs, www.neb.com) is preferred. Alternatively, if more conventional approaches shall be used, e.g. in an ordinary wet lab setting without robotics, the vectors can also be linearized by restriction digestion, and a gene of interest can be integrated by restriction / ligation (see below Section C of the present embodiment). The DNA sequence (SEQ ID NO: 1) and map of the present MIE is shown in Fig. 13.

B.3. Tags, promoters, terminators

For expression of proteins in prokaryotic hosts, the vectors of the ACEMBL system contain per default promoters T7 and Lac, as well as the T7 terminator element (Figs.1, 14). The T7 system requires bacterial strains which contain a T7 polymerase gene, e.g. in the *E. coli* genome. The Lac promoter is a strong endogenous promoter which can be utilized in most strains. The present ACEMBL vectors contain the lac operator element for repression of heterologous expression.

Evidently, all promoters and terminators present in ACEMBL Donor and Acceptor vectors, and in fact the entire multiple integration element (MIE), excluding the HE and X site, respectively, can be exchanged with an expression cassette of choice by using restriction/ligation cloning with appropriate enzymes (for example ClaI/PmeI, Fig. 2) or insertion into linearized ACEMBL vectors where the MIE was removed by sequence and ligation independent approaches such as SLIC. For example, the T7 promoter in pDC can be substituted with a trc promoter (pDC^{trc}), and the T7 promoter in pACE with an arabinose promoter (pACE^{ara}). Such variants can be used successfully in coexpression experiments by inducing with arabinose and IPTG.

In contrast to the ACEMBL vectors for expression in prokaryotic hosts, the vectors for expression in insect cells do not contain prokaryotic promoter and terminator structures. As already mentioned above, they have either a polh expression cassette (polh EC) or a p10 expression cassette (p10 EC). These expression cassettes contain common polyhedron or p10 promoters from AcMNPV, a sequence of restriction sites and either SV40 or HSVtk polyadenylation signal sequences.

The ACEMBL system vectors of the present example do not contain DNA sequences encoding for affinity tags to facilitate purification or solubilization of the protein(s) of interest. However, typically used C- or N-terminal oligohistidine tags, with or without protease sites for tag removal can be introduced by means of the respective PCR primers used for amplification of the genes of interest prior to insertion into the MIE, e.g. by SLIC-mediated insertion. Thus, Donor and Acceptor vectors of the present invention may be equipped by the array of custom tags prior to inserting recombinant

genes of interest. This is best done by a design which will, after tag insertion, still be compatible with the recombination based principles of ACEMBL system usage.

B.4 Complex expression

5

For expression in *E.coli*, the ACEMBL multigene expression vector fusions with appropriate promoters or terminators are transformed into the appropriate expression host of choice. With respect to the present exemplary vectors (T7 and lac promoter elements), most of the wide array of currently available expression strains can be utilized. If particular expression strains already contain helper plasmids with DNA encoding for chaperones, lysozyme or else, the design of the multigene fusion is preferably such that the ACEMBL vector containing the resistance marker that is also present on the helper plasmid is not included in multigene vector construction.

10

15 Alternatively, if further vectors are required for complex production in an experiment, the issue can be resolved by creating alternative versions of the ACEMBL vectors containing resistance markers that circumvent the conflict. This can be easily performed by PCR amplifying the vectors minus the resistance marker, and combine the resulting fragments with a PCR amplified resistance marker by recombination
20 (SLIC) or blunt-end ligation (using 5'phosphorylated primers).

25

Donor vectors of the present example depend on expression by the host of the *pir* gene product, due to the R6Ky conditional origin of replication. In regular expression strains, they rely on fusion with an Acceptor for productive replication. Donors or Donor-Donor fusions can nonetheless be used even for expression when not fused with an Acceptor, by using expression strains carrying a genomic insertion of the *pir* gene. Such strains are commercially available (Novagen Inc., Madison WI, USA).

30

Cotransformation of two ACEMBL plasmids adapted for expression in bacteria can lead to a successful protein complex expression. The present ACEMBL system for expression in prokaryotic hosts contains two Acceptor vectors, pACE and pACE2, which are identical except for the resistance marker (Figs. 1, 14). These can be used to express genes present on pACE or pACE2, respectively, by cotransformation and

exposure to both antibiotics simultaneously. In fact, entire Acceptor-Donor fusions containing several genes, based on pACE or pACE2 as Acceptors, can in principle be cotransformed for multi-expression, if needed.

- 5 For expression in insect cells (such as Sf9, Sf21, Hi5 etc.) using the baculovirus system, suitable ACEMBL vectors of the present invention need to be integrated into a baculovirus genome (composite virus generation). This is typically carried out by transformation of the desired Cre-LoxP fusion into bacterial cells containing the desired virus genome as a bacterial artificial chromosome. Using the vector system
10 of the present invention adapted for baculovirus integration is used, three approaches are possible as outlined in Fig. 35, 36 and 37, respectively.

C. Procedures

15 *C.1. Cloning into ACEMBL vectors*

All Donors and Acceptors of the preferred embodiment for expression prokaryotic hosts contain an identical MIE with exception of the homing endonuclease site / BstXI tandem encompassing the MIE (Figs. 1 and 14). The MIE is tailored for
20 sequence and ligation independent gene insertion methods. In addition, the MIE also contains a series of unique restriction sites, and therefore can be used as a classical polylinker for conventional gene insertion by restriction/ligation. For automated applications insertion of genes of interest is preferably carried out by recombination approaches such as SLIC.

25

The Donor vectors for expression in insect cells according to the present preferred embodiment also contain an MIE which is, however, different for each vector (see plasmid maps of vectors pIDC, pIDK and pIDS in Figs. 21, 22 and 23, respectively).

30 *C1.1. Single gene insertion into the MIE by SLIC*

Several procedures for restriction/ligation independent insertion of genes into vectors have been published or commercialized (e.g. Novagen LIC, Becton-Dickinson BD In-Fusion etc.). These systems share in common that they rely on the exonuclease

activity of DNA polymerases. In the absence of dNTPs, 5' extensions are created from blunt ends or overhangs by digestion from the 3' end. If two DNA fragments contain the same ~20-30 bp sequence at their termini at opposite ends, this results in overhangs that share complementary sequences capable of annealing. This can be exploited for ligation independent combination of two or several DNA fragments containing homologous sequences.

If T4 DNA polymerase is used, this can be carried out in a manner that is independent of the sequences of the homology regions (Sequence and Ligation Independent Cloning, SLIC) and detailed protocols are available for the skilled person. In the context of multiprotein expression, this is particularly useful, as this approach is independent of the presence of unique restriction sites, or of their creation by mutagenesis, in the ensemble of encoding DNAs.

For use in the context of the present invention, the SLIC process was adapted for inserting encoding DNAs amplified by Phusion polymerase into the ACEMBL Acceptor and Donor vectors. In this way, not only seamless integration of genes into the expression cassettes, but also concatamerization of expression cassettes to multigene constructs can be achieved by applying the same, simple routine that can be readily automated.

The following Protocol 1 represents an improved process based on the method described in Li et al. (2007, *supra*). Protocol 1 is designed for manual operation. Other systems may be used (e.g. BD-InFusion etc.), and if so, the manufacturers' recommendations should be followed. The present protocol may be adopted for robotics applications. Corresponding modifications of the protocol are outlined in Section D).

Protocol 1: Single gene insertion by SLIC.

Reagents required:

Phusion Polymerase

5x HF Buffer for Phusion Polymerase

dNTP mix (10 mM)

T4 DNA polymerase (and 10x Buffer)

DpnI enzyme
E. coli competent cells
100mM DTT, 2M Urea, 500 mM EDTA
Antibiotics

5

Step 1: Primer design

Primers for the SLIC procedure are designed to provide the regions of homology which result in the long sticky ends upon treatment with T4 DNA polymerase in the absence of dNTP:

10

Primers for the insert contain a DNA sequence corresponding to this region of homology ("Adaptor sequence" in Fig. 3, inset), followed by a sequence which specifically anneals to the insert to be amplified (Fig. 3, inset). Useful examples of adaptor sequences for SLIC are listed below (Table I).

15

This "insert specific sequence" can be located upstream of a ribosome binding site (rbs), for example if the gene of interest (GOI) is amplified from a vector already containing expression elements (e.g. the pET vector series). Otherwise, the forward primer needs to be designed such that a ribosome binding site is also provided in the final construct (Fig. 3, inset).

20

Primers for PCR linearization of the vector backbone are simply complementary to the two adaptor sequences present in the primer pair chosen for insert amplification (Fig. 3).

25

Step 2: PCR amplification of insert and vector

Identical reactions are prepared in 100 µl volume for DNA insert to be cloned and vector to be linearized by PCR:

30

ddH ₂ O	75 µl
5× Phusion HF Reaction buffer	20 µl
dNTPs (10 mM stock)	2 µl

Template DNA (100 ng/μl)	1 μl
5 SLICprimer (100 μM stock)	1 μl
3 SLICprimer (100 μM stock)	1 μl
Phusion polymerase (2 U/μl)	0.5 μl

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, then double extension time):

1 x 98° C for 2 min

5 30 x [98° C for 20 sec. -> 50°C for 30 sec. -> 72°C for 3 min]

Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

10 **Step 3:** DpnI treatment of PCR products (optional)

PCR reactions are then supplied with 1 μl DpnI enzyme which cleaves parental plasmids (that are methylated). For insert PCR reactions, DpnI treatment is not required if the resistance marker of the template plasmid differs from the destination vector.

15 Reactions are then carried out as follows:

Incubation: 37°C for 1-4h

Inactivation: 80°C for 20 min

Step 4: Purification of PCR products

20

PCR products should be cleaned of residual dNTPs. Otherwise, the T4 DNA polymerase reaction (Step 5) is compromised. Product purification is preferably performed by using commercial PCR Purification Kits or NucleoSpin Kits (e.g. from Qiagen, Macherey-Nagel etc.). It is recommended to perform elution in the minimal possible volume indicated by the respective manufacturer.

25

Step 5: T4 DNA polymerase exonuclease treatment

Identical reactions are prepared in 20 μl volume for insert and for vector (eluted in Step 4):

30

10x T4 DNA polymerase buffer 2 μl

100mM DTT	1 μ l
2M Urea	2 μ l
DNA eluate from Step 3 (vector or insert)	14 μ l
T4 DNA polymerase	1 μ l

Reactions are then carried out as follows:

Incubation: 23°C for 20 min

Arrest: Addition of 1 μ l 500 mM EDTA

5 Inactivation: 75°C for 20 min

Step 6: Mixing and Annealing

10 T4 DNA polymerase exonuclease-treated insert and vector are then mixed, followed by an (optional) annealing step which enhances the efficiency:

T4 DNA pol-treated insert:	10 μ l
T4 DNA pol-treated vector:	10 μ l

Annealing: 65°C for 10 min

Cooling: Slowly (in heat block) to RT (room temperature)

15

Step 7: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

20

Reactions for pACE and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery (2 - 4 h) plated on agar containing ampicillin (100 μ g/ml) or tetracycline (25 μ g/ml), respectively.

25 Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 8: Plasmid analysis

Plasmids are cultured in small-scale in media containing the corresponding antibiotic,
5 and analyzed by sequencing and (optionally) restriction mapping with an appropriate restriction enzyme.

C1.2. Polycistron assembly in MIE by SLIC

10 The multiple integration element according to the present invention can also be used to integrate genes of interest by using multi-fragment SLIC recombination as shown in Fig. 4. Genes preceded by ribosome binding sites (rbs) can be assembled in this way into polycistrons.

15 A detailed protocol is outlined in the following Protocol 2:

Protocol 2. Polycistron assembly by SLIC.

Reagents required:

Phusion Polymerase
20 5x HF Buffer for Phusion Polymerase
dNTP mix (10 mM)
T4 DNA polymerase (and 10x Buffer)
E. coli competent cells
100mM DTT, 2M Urea, 500 mM EDTA
25 Antibiotics

Step 1: Primer design

The MIE element according to the present embodiment is composed of tried-and-
30 tested primer sequences. These constitute the “Adaptor” sequences that can be used for inserting single genes or multigene constructs. Examples of useful adaptor sequences are listed below (see Table I).

Adaptor sequences form the 5' segments of the primers used to amplify DNA fragments to be inserted into the MIE. Insert specific sequences are added at 3', DNA coding for a ribosome binding sites can be inserted optionally, if not already present on the PCR template.

5

Step 2: PCR amplification of insert and primer

Identical reactions are prepared in 100 µl volume for all DNA insert (GOI 1, 2, 3) to be cloned and the vector to be linearized by PCR:

10

ddH ₂ O	75 µl
5× Phusion HF Reaction buffer	20 µl
dNTPs (10 mM stock)	2 µl
Template DNA (100 ng/µl)	1 µl
5' SLIC primer (100 µM stock)	1 µl
3' SLIC primer (100 µM stock)	1 µl
Phusion polymerase (2 U/µl)	0.5 µl

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, then double extension time):

1 x 98° C for 2 min

30 x [98° C for 20 sec. -> 50°C for 30 sec. -> 72°C for 3 min]

15

Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

Step 3: DpnI treatment of PCR products (optional)

20 PCR reactions are then supplied with 1 µl DpnI enzyme which cleaves parental plasmids (that are methylated). For insert PCR reactions, DpnI treatment is not required if the resistance marker of the template plasmids differs from the destination vector.

Reactions are then carried out as follows:

25

Incubation: 37°C for 1-4h

Inactivation: 80°C for 20 min

Step 4: Purification of PCR products

PCR products should be cleaned of residual dNTPs. Otherwise, the T4 DNA polymerase reaction (Step 5) is compromised.

5

Product purification is preferably performed by using commercial PCR Purification Kits or NucleoSpin Kits (Qiagen, Macherey-Nagel or others). It is recommended to perform elution in the minimal possible volume indicated by the respective manufacturer.

10 **Step 5:** T4 DNA polymerase exonuclease treatment

Identical reactions are prepared in 20 µl volume for each insert (GOI 1, 2, 3) and for the vector (eluted in Step 4):

10x T4 DNA polymerase buffer	2 µl
100mM DTT	1 µl
2M Urea	2 µl
DNA eluate from Step 3 (vector or insert)	14 µl
T4 DNA polymerase	1 µl

15

Reactions are then carried out as follows:

Incubation: 23°C for 20 min
Arrest: Addition of 1 µl 500 mM EDTA
Inactivation: 75°C for 20 min

20

Step 6: Mixing and Annealing

T4 DNA polymerase exonuclease-treated insert and vector are then mixed, followed by an (optional) annealing step which enhances efficiency.

25

T4 DNA pol-treated insert 1 (GOI 1):	5 µl
T4 DNA pol-treated insert 2 (GOI 2):	5 µl
T4 DNA pol-treated insert 3 (GOI 3):	5 µl
T4 DNA pol-treated vector:	5 µl

Annealing: 65°C for 10 min

Cooling: Slowly (in heat block) to RT

Step 7: Transformation

5

Mixtures are next transformed into competent cells following standard transformation procedures.

10

Reactions for pACE and pACE2 derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery plated on agar containing ampicillin (100 μ g/ml) or tetracycline (25 μ g/ml), respectively.

15

Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, available from Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

Step 8: Plasmid analysis

20

Plasmids are cultured and correct clones are selected based on specific restriction digestion and DNA sequencing of the inserts.

Table I. Adaptor DNA sequences.

For single gene or multigene insertions into ACEMBL vectors by SLIC.

Adaptor *	Sequence	Description
(a) Adaptors for cloning into ACEMBL vectors for expression in prokaryotic hosts		
T7InsFor	TCCCGCGAAATTAATA CGACTCACTATAGGG (SEQ ID NO: 20)	Forward primer for <u>insert</u> amplification, if gene of interest (GOI) is present in a T7 system vector (i.e. pET series). No further extension (rbs, insert specific overlap) required.
T7InsRev	CCTCAAGACCCGTTTA GAGGCCCAAGGGGT TATGCTAG	Reverse primer for <u>insert</u> amplification, if GOI is present in a T7 system vector (i.e. pET series). No further extension (stop codon, insert

	(SEQ ID NO: 21)	specific overlap) required.
T7VecFor	CTAGCATAACCCCTTG GGGCCTCTAAACGGG TCTTGAGG (SEQ ID NO: 22)	Forward primer for <u>vector</u> amplification, reverse complement of T7InsRev. No further extension required.
T7VecRev	CCCTATAGTGAGTCGT ATTAATTTTCGCGGGA (SEQ ID NO: 23)	Reverse primer for <u>vector</u> amplification, reverse complement of T7InsFor. No further extension required.
NdeInsFor	GTTTAACTTTAAGAAG GAGATATACATATG (SEQ ID NO: 24)	Forward primer for <u>insert</u> amplification for insertion into MIE site I1 (Fig. 2). Further extension at 3' (insert specific overlap) required. Can be used with adaptor XhoInsRev in case of single fragment SLIC (Fig. 3).
XhoInsRev	GGGTTTAAACGGAACT AGTCTCGAG (SEQ ID NO: 25)	Reverse primer for <u>insert</u> amplification for insertion into MIE site I4 (Fig. 2). Further extension at 3' (stop codon, insert specific overlap) required. Can be used with adaptor NdeInsFor in case of single fragment SLIC (Fig. 3).
XhoVecFor	CTCGAGACTAGTTCCG TTTAAACCC (SEQ ID NO: 26)	Forward primer for <u>vector</u> amplification, reverse complement of XhoInsRev. No further extension required.
NdeVecRev	CATATGTATATCTCCTT CTTAAAGTTAAAC (SEQ ID NO: 27)	Reverse primer for <u>vector</u> amplification, reverse complement of NdeInsFor. No further extension required.
SmaBam	GAATTCAGTGGCCGTC GTTTTACAGGATCC (SEQ ID NO: 28)	Reverse primer for <u>insert</u> amplification (GO11) for insertion into MIE site I1 (Fig. 2). Further extension at 3' (stop codon, insert specific overlap) required. Use with adaptor NdeInsFor.
BamSma	GGATCCTGTAAAACGA CGGCCAGTGAATTC (SEQ ID NO: 29)	Forward primer for <u>insert</u> amplification (GO12) for insertion into site I2 (Fig. 2, 4). Further extension at 3' (rbs, insert specific over-lap) required. Use with adaptor SacHind.(multifragment SLIC, Fig. 4)
SacHind	GCTCGACTGGGAAAA CCCTGGCGAAGCTT (SEQ ID NO: 30)	Reverse primer for <u>insert</u> amplification (GO12) insertion into MIE site I2 (Fig. 2, 4). Further extension at 3' (stop codon, insert specific overlap) required. Use with adaptor BamSma.(multifragment SLIC, Fig. 4)

HindSac	AAGCTTCGCCAGGGTT TTCCCAGTCGAGC (SEQ ID NO: 31)	Forward primer for <u>insert</u> amplification (GOI3) for insertion into site I3 (Fig. 2, 4). Further extension at 3' (rbs, insert specific over-lap) required. Use with adaptor BspEco.(multifragment SLIC, Fig. 4)
BspEco5	GATCCGGATGTGAAAT TGTTATCCGCTGGTAC C (SEQ ID NO: 32)	Reverse primer for <u>insert</u> amplification (GOI3) insertion into MIE site I3 (Fig. 2, 4). Further extension at 3' (stop codon, insert specific overlap) required. Use with adaptor HindSac.(multifragment SLIC, Fig. 4)
Eco5Bsp	GGTACCAGCGGATAA CAATTTACATCCGGA TC (SEQ ID NO: 33)	Forward primer for <u>insert</u> amplification (GOI3) for insertion into site I4 (Fig. 2,4). Further extension at 3' (rbs, insert specific over-lap) required. Use with adaptor XhoInsRev .(multifragment SLIC, Fig. 4)
(b) Adaptors for cloning into ACEMBL vectors for expression in insect cells		
PolhInsFor	CCCACCATCGGGCGC GGATCCCG (SEQ ID NO: 34)	Forward primer for <u>insert</u> amplification, needs to be followed by insert specific sequence (ca. 20 bp)
PolhInsRev	CGAGACTGCAGGCTC TAGATTCG (SEQ ID NO: 35)	Reverse primer for <u>insert</u> amplification, needs to be followed by insert specific sequence (ca. 20 bp)
PolhVecFor	CGGGATCCGCGCCCG ATGGTGGG (SEQ ID NO: 36)	Forward primer for <u>vector</u> amplification, reverse complement of PolhInsRev. No further extension required.
PolhVecRev	CGAATCTAGAGCCTGC AGTCTCG (SEQ ID NO: 37)	Reverse primer for <u>vector</u> amplification, reverse complement of PolhInsFor. No further extension required.
P10hInsFor	CTCCCGGTACCGCAT GCTATGCATCAGC (SEQ ID NO: 38)	Forward primer for insert amplification, needs to be followed by insert specific sequence (ca. 20 bp)
P10InsRev	AATCACTCGACGAAGA CTTGATCACC (SEQ ID NO: 39)	Reverse primer for insert amplification, needs to be followed by insert specific sequence (ca. 20 bp)
P10VecFor	GCTGATGCATAGCATG CGGTACCGGGAG (SEQ ID NO: 40)	Forward primer for vector amplification, reverse complement of P10InsRev. No further extension required.
P10VecRev	GGTGATCAAGTCTTCG TCGAGTGATT	Reverse primer for vector amplification, reverse complement of P10InsFor.

	(SEQ ID NO: 41)	No further extension required.
--	-----------------	--------------------------------

* All Adaptor primers (without extension) can be used as sequencing primers for genes of interest that were inserted into the MIE according to the present embodiment.

5

C.1.3. Gene insertion by restriction/ligation

The MIEs of the present invention can also be used as a multiple cloning site with a series of unique restriction sites. Preferably, the MIE described herein for expression of proteins in prokaryotic hosts is preceded by a promoter and a ribosome binding site, and followed by a terminator. The MIEs of the preferred embodiments described herein for expression of proteins in insect cells contain a polh expression cassette or a p10 expression cassette as already mentioned above. Therefore, cloning into the MIE by classical restriction/ligation also yields functional expression cassettes.

15

Genes of interest (GOI) can be subcloned by using standard cloning procedures into the multiple integration element (MIE) (see, for example, Fig. 13) of ACEMBL vectors.

20 **Protocol 3.** Restriction/ligation cloning into an MIE.

Reagents required:

Phusion Polymerase
 5x HF Buffer for Phusion Polymerase
 25 dNTP mix (10 mM)
 10 mM BSA
 Restriction endonucleases (and 10x Buffer)
 T4 DNA ligase (and 10x Buffer)
 Calf or Shrimp intestinal alkaline phosphatase
 30 *E. coli* competent cells
 Antibiotics

Step 1: Primer design

For conventional cloning, PCR primers are designed containing chosen restriction sites, preceded by appropriate overhangs for efficient cutting (see, e.g. New England Biolabs catalogue), and followed by ≥ 20 nucleotides overlapping with the gene of interest that is to be inserted.

In the case of the ACEMBL system for expression in bacteria, the MIE of the present embodiment is identical for all ACEMBL vectors. They contain a ribosome binding site preceding the NdeI site. For single gene insertions, therefore, an rbs need not be included in the primer.

If multigene insertions are needed (for example in insertion sites I1-I4 of the MIE), primers should be designed such that an rbs preceding the gene and a stop codon at its 3' end are provided.

In particular for polycistron cloning by restriction/ligation, it is recommended to construct templates by custom gene synthesis. In the process, the restriction sites present in the MIE can be eliminated from the encoding DNAs.

Step 2: Insert preparation

PCR of insert(s):

Identical PCR reactions are prepared in 100 μ l volume for genes of interest to be inserted into the MIE:

ddH ₂ O	75 μ l
5 $\times$ Phusion HF Reaction buffer	20 μ l
dNTPs (10 mM stock)	2 μ l
Template DNA (100 ng/ μ l)	1 μ l
5' primer (100 μ M stock)	1 μ l
3' primer (100 μ M stock)	1 μ l
Phusion polymerase (2 U/ μ l)	0.5 μ l

PCR reactions are then carried out with a standard PCR program (unless very long DNAs are amplified, then double extension time):

1 $\times$ 98° C for 2 min

30 $\times$ [98° C for 20 sec. $\rightarrow$ 50° C for 30 sec. $\rightarrow$ 72° C for 3 min]

Hold at 10°C

Analysis of the PCR reactions by agarose gel electrophoresis and ethidium bromide staining is recommended.

5

Product purification is preferably performed by using commercial PCR Purification Kits or NucleoSpin Kits (available from Qiagen, Macherey-Nagel and other manufacturers). It is recommended to perform elution in the minimal possible volume indicated by the manufacturer.

10

Restriction digestion of insert(s):

Restriction reactions are carried out in 40 µl reaction volumes, using specific restriction enzymes as specified by manufacturer's recommendations (c.f. New England Biolabs catalogue and others).

15

PCR Kit eluate (≥ 1 µg)	30 µl
10x Restriction enzyme buffer	4 µl
10 mM BSA	2 µl
Restriction enzyme for 5'	2 µl
Restriction enzyme for 3'	2 µl (in case of double digestion, otherwise ddH ₂ O)

Restriction digestions are performed in a single reaction with both enzymes (double digestion), or, alternatively, sequentially (two single digestions) if the buffer conditions required are incompatible.

20

Gel extraction of insert(s):

Processed insert is then purified by agarose gel extraction using commercial kits (Qiagen, Macherey-Nagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the respective manufacturer.

25

Step 3: Vector preparation

Restriction digestion of ACEMBL plasmid(s):

Restriction reactions are carried out in 40 µl reaction volumes, using specific restriction enzymes as specified by manufacturer's recommendations (see, e.g. New England Biolabs catalogue and others).

ACEMBL plasmid (≥ 0.5 µg) in ddH ₂ O	30 µl
10x Restriction enzyme buffer	4 µl
10 mM BSA	2 µl
Restriction enzyme for 5'	2 µl
Restriction enzyme for 3'	2 µl (in case of double digestion, otherwise ddH ₂ O)

5

Restriction digestions are performed in a single reaction with two enzymes (double digestion), or, alternatively, sequentially (two single digestions), if the buffer conditions required are incompatible.

10 Gel extraction of vector(s):

The processed vector is then purified by agarose gel extraction using commercial kits (Qiagen, Macherey-Nagel etc.). It is recommended to elute the extracted DNA in the minimal volume defined by the respective manufacturer.

15 **Step 4:** Ligation

Ligation reactions are carried out in 20 µl reaction volumes according to the recommendations of the supplier of the T4 DNA ligase:

ACEMBL plasmid (gel extracted)	8 µl
Insert (gel extracted)	10 µl
10x T4 DNA Ligase buffer	2 µl
T4 DNA Ligase	0.5 µl

20

Ligation reactions are performed at 25°C (sticky end) for 1h or at 16°C (blunt end) overnight.

Step 5: Transformation

25

Mixtures are next transformed into competent cells following standard transformation procedures.

Reactions for Acceptor derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5 α , HB101) and after recovery plated on agar containing ampicillin (100 μ g/ml) or tetracycline (25 μ g/ml), respectively.

- 5 Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on agar containing chloramphenicol (25 μ g/ml, pDC), kanamycin (50 μ g/ml, pDK), and spectinomycin (50 μ g/ml, pDS).

10 **Step 6: Plasmid analysis**

Plasmids are cultured and correct clones are selected based on specific restriction digestion and DNA sequencing of the inserts.

15 C.1.4. Multiplication by using the HE and BstXI sites

The ACEMBL system vectors according to the present invention contain a homing endonuclease (HE) site and a designed BstXI site that envelop the multiple integration element (MIE). The homing endonuclease site can be used to insert entire
20 expression cassettes, containing single genes or polycistrons, into a vector already containing one gene or several genes of interest. Homing endonucleases have long recognition sites (12 to 40 base pairs or more, preferably 20-30 base pairs). Although not all equally stringent, homing endonuclease sites are most probably unique in the context of even large plasmids, or, in fact, entire genomes.

25

In the ACEMBL system of the present embodiment, Donor vectors contain a recognition site for homing endonuclease PI-SceI (Fig. 2). This HE site yields upon cleavage a 3' overhang with the sequence -CTGC. Acceptor vectors contain the homing endonuclease site I-CeuI, which upon cleavage will result in a 3' overhang of
30 -CTAA. On Acceptors and Donors, the respective HE site is preceding the MIE. The 3' end of the MIE contains a specifically designed BstXI site, which upon cleavage will generate a matching overhang. The basis of this is the specificity of cleavage by BstXI. The recognition sequence of BstXI is defined as CCANNNNN'NTGG (SEQ ID NO: 42) (apostrophe marks position of phosphodiester link cleavage). The residues

denoted as N can be chosen freely. Donor vectors thus contain a BstXI recognition site of the sequence CCATGTGC'CTGG (SEQ ID NO: 43), and Acceptor vectors contain CCATCTAA'TTGG (SEQ ID NO: 44). The overhangs generated by BstXI cleavage in each case will match the overhangs generated by HE cleavage. Note
5 that Acceptors and Donors have different HE sites.

The recognition sites are not symmetric. Therefore, ligation of a HE/BstXI digested fragment into a HE site of an ACEMBL vector will be (1) directional and (2) result in a hybrid DNA sequence where a HE half site is combined with a BstXI half site. This
10 site will be cut by neither HE nor BstXI. Therefore, in a construct that had been digested with a HE, insertion by ligation of HE/BstXI digested DNA fragment containing an expression cassette with one or several genes will result in a construct which contains all heterologous genes of interest, enveloped by an intact HE site in front, and a BstXI site at the end. Therefore, the process of integrating entire
15 expression cassettes by means of HE/BstXI digestion and ligation into a HE site can be repeated iteratively.

Protocol 4. Multiplication by using homing endonuclease/BstXI.

20 Reagents required:

- Homing endonucleases PI-SceI, I-CeuI
- 10x Buffers for homing endonucleases
- Restriction enzyme BstXI (and 10x Buffer)
- T4 DNA ligase (and 10x Buffer)
- 25 *E. coli* competent cells
- Antibiotics

Step 2: Insert preparation

30 Restriction reactions are carried out in 40 µl reaction volumes, using homing endonucleases PI-SceI (Donors) or I-CeuI (Acceptors) as recommended by the supplier (e.g. New England Biolabs or others).

ACEMBL plasmid ($\geq 0.5 \mu\text{g}$) in ddH ₂ O	32 μl
10x Restriction enzyme buffer	4 μl
10 mM BSA	2 μl
PI-SceI (Donors) or I-CeuI (acceptors)	2 μl

Reactions are then purified by PCR extraction kit or acidic ethanol precipitation, and next digested by BstXI according to the recommendations of the supplier.

5

HE digested DNA in ddH ₂ O	32 μl
10x Restriction enzyme buffer	4 μl
10 mM BSA	2 μl
BstXI	2 μl

Gel extraction of insert(s):

Processed insert is then purified by agarose gel extraction using commercial kits (Qiagen, Macherey-Nagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the respective manufacturer.

10

Step 3: Vector preparation

Restriction reactions are carried out in 40 μl reaction volumes, using homing endonucleases PI-SceI (Donors) or I-CeuI (Acceptors) as recommended by the supplier (e.g. New England Biolabs catalogue or others).

15

ACEMBL plasmid ($\geq 0.5 \mu\text{g}$) in ddH ₂ O	33 μl
10x Restriction enzyme buffer	4 μl
10 mM BSA	2 μl
PI-SceI (Donors) or I-CeuI (acceptors)	1 μl

Reactions are then purified by PCR extraction kit or acidic ethanol precipitation, and next treated with intestinal alkaline phosphatase according to the recommendations of the respective supplier.

20

HE digested DNA in ddH ₂ O	17 µl
10x Alkaline phosphatase buffer	2 µl
Alkaline phosphatase	1 µl

Gel extraction of vector:

- 5 Processed vector is then purified by agarose gel extraction using commercial kits (Qiagen, Macherey-Nagel etc). It is recommended to elute the extracted DNA in the minimal volume defined by the respective manufacturer.

Step 4: Ligation

10

Ligation reactions are carried out in 20 µl reaction volumes:

HE/Phosphatase treated vector (gel extracted)	4 µl
HE/BstXI treated insert (gel extracted)	14 µl
10x T4 DNA Ligase buffer	2 µl
T4 DNA Ligase	0.5 µl

Ligation reactions are performed at 25°C for 1h or at 16°C overnight.

15

Step 5: Transformation

Mixtures are next transformed into competent cells following standard transformation procedures.

20

Reactions for Acceptor derivatives are transformed into standard *E. coli* cells for cloning (such as TOP10, DH5α, HB101) and after recovery plated on agar containing ampicillin (100 µg/ml) or tetracycline (25 µg/ml), respectively.

- 25 Reactions for Donor derivatives are transformed into *E. coli* cells expressing the *pir* gene (such as BW23473, BW23474, or PIR1 and PIR2, Invitrogen) and plated on

agar containing chloramphenicol (25 µg/ml, pDC, pIDC), kanamycin (50 µg/ml, pDK, pIDK), and spectinomycin (50 µg/ml, pDS, pIDS).

Step 6: Plasmid analysis

Plasmids are cultured and correct clones selected based on specific restriction digestion and DNA sequencing of the inserts.

C.2. Cre-LoxP reaction of Acceptors and Donors

Cre recombinase is a member of the integrase family (Type I topoisomerase from bacteriophage P1). It recombines a 34 bp loxP site (SEQ ID NO: 19; see Fig. 5) in the absence of accessory protein or auxiliary DNA sequence. The loxP site is comprised of two 13 bp recombinase-binding elements arranged as inverted repeats which flank an 8 bp central region where cleavage and ligation reaction occur.

The site-specific recombination mediated by Cre recombinase involves the formation of a Holliday junction (HJ). The recombination events catalyzed by Cre recombinase are dependent on the location and relative orientation of the LoxP sites. Two DNA molecules, for example an Acceptor and a Donor plasmid, containing single LoxP sites will be fused. Furthermore, the Cre recombination is an equilibrium reaction with 15 - 20 % efficiency in recombination. This creates useful options for multigene combinations for multiprotein complex expressions.

In a reaction where several DNA molecules such as Donors and Acceptors are incubated with Cre recombinase, the fusion/excision activity of the enzyme will result in an equilibrium state where single vectors (educt vectors) and all possible fusions coexist. Donor vectors can be used with Acceptors and/or Donors, likewise for Acceptor vectors. Higher order fusions are also generated where more than two vectors are fused. This is shown schematically in Fig. 6.

The fact that Donors of the present example contain a conditional origin of replication that depends on a *pir*⁺ (*pir* positive) background now allows for selecting out from this reaction mix all desired Acceptor-Donor(s) combinations. For this, the reaction mix is

used to transform to *pir* negative strains (TOP10, DH5 α , HB101 or other common laboratory cloning strains). Then, Donor vectors will act as suicide vectors when plated out on agar containing the antibiotic corresponding to the Donor encoded resistance marker, unless fused with an Acceptor. By using agar with the appropriate combinations of antibiotics, all desired Acceptor-Donor fusions can be selected for.

In this way, fusion vectors of 25 kb and larger can be generated. In stability tests (serial passaging for more than 60 generations), even such large plasmids are stable as checked by restriction mapping, even if only one of the antibiotics corresponding to the encoded resistance markers was provided in the growth medium.

C.2.1. Cre-LoxP fusion of Acceptors and Donors

The following protocol is designed for generating multigene fusions from Donors and Acceptors by Cre-LoxP reaction.

Reagents:

Cre recombinase

Standard *E. coli* competent cells (*pir*⁻ strain)

Antibiotics

96 well microtiter plates

12 well tissue-culture plates (or petri dishes) w. agar/antibiotics

LB media

1. For a 20 μ l Cre reaction, mix 1~2 μ g of each educts in approximate equal amounts (5' DNA termini). Add ddH₂O to adjust the total volume to 16~17 μ l, then add 2 μ l 10x Cre buffer and 1~2 μ l Cre recombinase.

2. Incubate the Cre reaction at 37°C (or 30°C) for 1 hour.

3. Optional: load 2-5 μ l Cre reaction on an analytical agarose gel for examination. Heat inactivation at 70°C for 10 minutes before the gel loading is strongly recommended.

4. For chemical transformation, mix 10-15 µl Cre reaction with 200 µl chemical competent cells. Incubate the mixture on ice for 15-30 minutes. Then perform heat shock at 42°C for 45-60 s.

Up to 20 µl Cre reaction (0.1 volumes of the chemical competent cell suspension) can be directly transformed into 200 µl chemical competent cells.

For electrotransformation, up to 2 µl Cre reaction could be directly mixed with 100 µl electrocompetent cells, and transformed by using an electroporator (e.g. BIORAD E. coli Pulser) at 1.8-2.0 kV.

Larger volumes of Cre reactions should be desalted by ethanol precipitation or PCR purification column before electrotransformation. The desalted Cre reaction mix does preferably not exceed 0.1 volumes of the electrocompetent cell suspension.

The cell/DNA mixture could be immediately used for electrotransformation without prolonged incubation on ice.

5. Add up to 400 µl of LB media (or SOC media) per 100 µl of cell/DNA suspension immediately after the transformation (heat shock or electroporation).

6. Incubate the suspension in a 37°C shaking incubator overnight or for at least 4 hours.

For recovering multifusion plasmid containing more than 2 resistance markers, it is strongly recommended to incubate the suspension at 37°C overnight.

7. Plate out the recovered cell suspension on agar containing desired combination of antibiotics. Incubate at 37°C overnight.

8. Emerged colonies after overnight incubation might be verified directly by restriction digestion at this stage, by referring to steps 12-16, *supra*. Especially in the case that only one multifusion plasmid is desired.

For further selection by single antibiotic challenges on a 96 well microtiter plate, continue to step 9.

Several various multifusion plasmids can be processed and selected on one 96 well microtiter plate in parallel.

9. For 96 well antibiotic tests, inoculate four colonies from each antibiotic agar plate into ~500 µl LB media without antibiotics. Incubate the cell cultures in a 37°C shaking incubator for 1-2 hours.

10. During the incubation of colonies, fill a 96 well microtiter plate with 150 µl antibiotic-containing LB media or colourful dye (positional marker) in corresponding wells.

A typical arrangement of the solutions, which is used for parallel selections of multifusion plasmids, is shown in Fig. 7. The basic principle underlying this aspect of the present invention is that every cell suspension from single colonies needs to be challenged by all four single antibiotics.

11. Add 1 µl aliquots of pre-incubated cell cultures to the corresponding wells. Then incubate the inoculated 96 well microtiter plate in a 37°C shaking incubator overnight at 180-200 rpm.

It is recommendable to use parafilm to wrap the plate.

The rest pre-incubated cell cultures could be kept in 4°C fridge for further inoculations.

12. Select transformants containing desired multifusion plasmids according to the combination of dense and clear cell cultures from each colony. Inoculate 10-20 µl cell cultures into 10 ml LB media with corresponding antibiotics. Incubate in a 37°C shaking incubator overnight.

13. Centrifuge the overnight cell cultures at 4000g for 5-10 minutes. Purify cell pellets with plasmid miniprep kit according to manufacturers' information.

14. Determine the concentrations of purified plasmid solutions by using UV absorption (e.g. NanoDrop™ 1000).

15. Digest 0.5~1 µg of the purified plasmid solution in a 20 µl restriction digestion (with 5-10 unit endonuclease). Incubate under recommended reaction condition for ~2 hours.

16. Use 5-10 µl of the digestion for analytical agarose gel (0.8-1.2%) electrophoresis. Verify the plasmid integrity by comparing the actual restriction pattern to the predicted restriction pattern *in silico* (e.g. by using VectorNTI).

C.2.2. Deconstruction of fusion vectors by Cre

The following protocol can be used, for instance for the recovery of four single ACEMBL vectors (pACE, pDC, pDK, pDS) by deconstructing tetra-fused pACKS plasmid (pACE-pDC-pDK-pDS) which preferably forms part of the ACEMBL System kit (see below Section E of the present embodiment). Likewise, the protocol is suitable for releasing single educts from multifusion constructs. This is achieved by Cre-LoxP reaction, transformation and plating on agar with appropriately reduced antibiotic resistance level (Fig. 6). For the liberated educt, encoding genes can be modified and diversified. Then, the diversified construct is resupplied by Cre-LoxP reaction.

Reagents:

Cre recombinase (and 10x Buffer)

E. coli competent cells

(*pir*⁺ strains, *pir*⁻ strains could be used only when partially deconstructed Acceptor-Donor fusions are desired).

Antibiotics

1. For a 20 µl De-Cre reaction, Incubate ~1 µg multifusion plasmid with 2 µl 10x Cre buffer, 1~2 µl Cre recombinase, add ddH₂O to adjust the total reaction volume to 20 µl.

2. Incubate the De-Cre reaction at 30°C for 1-4 hour.
3. Optional: load 2-5 µl De-Cre reaction on an analytical agarose gel for examination.

5

Heat inactivation at 70°C for 10 minutes before the gel loading is recommended.

4. For chemical transformation, mix 10-15µl De-Cre reaction with 200 µl chemical competent cells. Incubate the mixture on ice for 15-30 minutes. Then perform heat shock at 42°C for 45-60 s.

10

Up to 20 µl De-Cre reaction (0.1 volumes of the chemical competent cell suspension) can be directly transformed into 200 µl chemical competent cells.

- 15 For electrotransformation, up to 2 µl De-Cre reaction could be directly mixed with 100 µl electrocompetent cells, and transformed by using an electroporator (e.g. BIORAD E. coli Pulser) at 1.8-2.0 kV.

- 20 Larger volumes of De-Cre reaction should be desalted by ethanol precipitation or PCR purification column before electrotransformation. The desalted De-Cre reaction mix does preferably not exceed 0.1 volumes of the electrocompetent cell suspension.

- 25 The cell/DNA mixture could be immediately used for electrotransformation without prolonged incubation on ice.

5. Add up to 400 µl of LB media (or SOC media) per 100 µl of cell/DNA suspension immediately after the transformation (heat shock or electroporation).

- 30 6. Incubate the suspension in a 37°C shaking incubator.

For recovery of partially deconstructed double/triple fusions, incubate the suspension in a 37°C shaking incubator overnight or for at least 4 hours.

For recovery of individual educts such as single ACEMBL vectors from pACKS plasmid, incubate the suspension in a 37°C incubator for 1~2 hours.

7. Plate out the recovered cell suspension on agar containing desired combination of antibiotics. Incubate at 37°C overnight.

8. Colonies after overnight incubation can be verified directly by restriction digestion at this stage, by referring to steps 12-16.

This is especially recommended if only one single educt or partially deconstructed multifusion plasmid is desired.

For further selection by single antibiotic challenges on a 96 well microtiter plate, continue to step 9.

Several various single educts/partially deconstructed multifusion plasmids can be processed and selected on one 96 well microtiter plate in parallel.

9. For 96 well antibiotic tests, inoculate four colonies from each antibiotic agar plate into ~500 µl LB media without antibiotics. Incubate the cell cultures in a 37°C shaking incubator for 1-2 hours.

10. During the incubation of colonies, fill a 96 well microtiter plate with 150 µl antibiotic-containing LB media or colourful dye (positional marker) in corresponding wells.

Referring to Fig. 7 it is possible to provide a similar arrangement of the solutions, which is used for parallel selections of four various single educts/partially deconstructed multifusion plasmids. The underlying principle of the present aspect of the invention is that every cell suspension from single colonies is to be challenged by all four antibiotics separately.

11. Add 1 μ l aliquots of pre-incubated cell cultures to the corresponding wells. Then incubate the inoculated 96 well microtiter plate in a 37°C shaking incubator overnight at 180-200 rpm.
- 5 It is recommendable to use parafilm to wrap the plate.
The remaining pre-incubated cell cultures could be kept in 4°C fridge for further inoculations.
12. Select transformants containing desired single educts/partially deconstructed multifusion plasmids according to the combination of dense and clear cell cultures from each colony. Inoculate 10-20 μ l cell cultures into 10 ml LB media with corresponding antibiotic(s). Incubate in a 37°C shaking incubator overnight.
- 10
13. Centrifuge the overnight cell cultures at 4000g for 5-10 minutes. Purify cell pellets with plasmid miniprep kit according to manufacturers' information.
- 15
14. Determine the concentrations of purified plasmid solutions by using UV absorption (e.g. NanoDrop™ 1000).
- 20
15. Digest 0.5~1 μ g of the purified plasmid solution in a 20 μ l restriction digestion (with 5-10 unit endonuclease). Incubate under recommended reaction condition for ~2 hours.
- 25
16. Use 5-10 μ l of the digestion for analytical agarose gel (0.8-1.2%) electrophoresis. Verify the plasmid integrity by comparing the actual restriction pattern to predicted restriction pattern in silico (e.g. by using VectorNTI).
- 30
17. Optional: during recovery of all four single ACEMBL vectors from pACKS plasmid, in case one or more single ACEMBL vectors fail to be liberated from one De-Cre reaction. One can just pick partially deconstructed double/triple fusions containing desired single ACEMBL vector(s), and perform a second De-Cre reaction (repeat steps 1-8).

Typically, up to 2 sequential De-Cre reactions are sufficient to recover all four single ACEMBL vectors from pACKS plasmid, and the liberation of single educts from double/triple fusions could be much more efficient than from pACKS plasmid (quadruple fusion). The same principle also applies to the deconstruction of any other multifusion plasmid based on the ACEMBL system according to the present invention.

C.3. Coexpression in bacteria by Cotransformation

Protein complexes can be expressed also from two separate vectors that were cotransformed in expression strains. The cotransformed vectors can have the same or different origins of replication, however, they must encode for different resistance markers. Plasmids pACE (ampicillin resistance marker) and pACE2 (tetracycline resistance marker) have both a ColE1 derived replicon and can therefore be used with all common expression strains. pACE and pACE2 derivatives (also including fused Donors if needed) can be cotransformed into expression strains, and double transformants selected for by plating on agar plates containing both ampicillin and tetracycline antibiotics.

Transformations are carried out using standard transformation protocols (see, e.g. the latest edition of Ausubel et al. (ed.), *supra*).

D. Automation

As already outlined above, cloning and expression of multiple protein complexes using the nucleic acids, vectors and methods of the present invention is highly suited for automation equipment employing current robotic techniques.

In the following general protocols as exemplified for a Tecan Freedom EvoII 200 pipetting device are provided. The pipeting device is typically equipped with liquid handling arm1 (LiHa1), 4 fixed tips (steel needles), 4 disposable tips coni (Diti's), 250µl syringes, liquid handling arm2 (LiHa2), 8 fixed tips (steel needles), 2.5ml syringes, robotic manipulator arm (RoMa / transportation of plates), version long. The work station usually contains the following integrated devices: thermocycler PTC-200

(Biorad), Te-Shake, heatable plate shaker (Tecan), Variomag Thermoshaker, heat- and coolable plate shaker (Inheco), Te-Vacs, dual vacuum station for filter plates (Tecan), SafireII, UV VIS plate reader (Tecan) and cooling unit 400W (FRYKA multistar).

5

D.1. Automated SLIC process

A schematic representation of a workflow for automated SLIC is shown in Fig. 22.

10 **Step 1: Initial PCR**

Source plate: 96 well standard microtiter plate containing the PCR templates (cDNA approx. 0.2 µg/µl)

15 **Reaction plate:** 96 well PCR plate (Eppendorf)

Material: Sample mix plate (96 well PCR plate; Eppendorf), 1% agarose E-Gel® (Invitrogen), Phusion® DNA Polymerase master mix, oligonucleotide primers at 20µM, 2x DNA loading dye (2xDLD) (Fermentas), E-Gel® Low Range quantitative DNA Ladder (Invitrogen), 10x Buffer Tango® with BSA (Fermentas), DpnI (Fermentas)

20

PCR program:

11x [98°C for 20 sec. → 60-50°C for 30 sec.(step down every 2nd cycle 1°C)
25 → 72°C for 3 min.]
19x [98°C for 20 sec. → 50°C for 30 sec. → 72°C for 3 min.]
72°C for 3 min.
Hold at 10°C

30 **DpnI digest program:**

37°C for 3 h
10°C for 1 min

Procedure:

Wash tips → Pipet 89 µl PCR master-mix into reaction plate

Wash tips → Pipet 1 µl template DNA according to worklist

Wash tips → Pipet 5 µl primer each to reaction plate

Wash tips → Run PCR program

5 Wash tips → Pipet 10 µl 10x Buffer Tango® with BSA to reaction plate

Wash tips → Pipet 5 µl DpnI to reaction plate

Wash tips → Run DpnI digest program

Wash tips → Pipet 10 µl 2xDLD to each well of sample mix plate

Wash tips → Pipet 15 µl DNA marker each to the E-gel marker slots

10 Wash tips → Pipet 10 µl PCR product to 2xDLD on sample mix plate

Wash tips → Pipet 15 µl sample mix to the E-Gel sample slots

Wash tips → Run E-Gel® for 25 min.

Assess results

15 **Step 2: PCR Purification**

Source plate: 96 well PCR plate (Eppendorf) with PCR samples

Target plate: 96 well microtiter elution plate (Macherey-Nagel)

20

Material: PCR purification kit, NucleoSpin 96 Extract II Kit (Macherey-Nagel)

Procedure: According to manufacturer's information (<http://www.macherey-nagel.com/tabid/10887/default.aspx>)

25

Step 3: T4 DNA Polymerase Reaction

Source plate: 96 well microtiter elution plate (Macherey-Nagel)

30 **Reaction plate:** 96 well PCR plate (Eppendorf)

Material: bidest. water, 10x T4 DNA polymerase reaction buffer (Novagen), 100mM DTT, 2M Urea, T4 DNA polymerase (Novagen LIC qualified), 500 mM EDTA

Incubation program: 23°C for 10 min. (program 1)

75°C for 20 min. (program 2)

5 **Procedure:**

Wash tips → Pipet 6 µl water in to reaction plate

Wash tips → Pipet 2 µl 10x reaction buffer into reaction plate

Wash tips → Pipet 1 µl 100mM DTT into reaction plate

10 Wash tips → Pipet 2 µl 2M Urea into reaction plate

Wash tips → Pipet 8 µl DNA sample from prev. PCR into reaction plate

Wash tips → Pipet 0.5 µl T4 DNA polymerase into reaction plate

Wash tips → Run incubation program 1

Wash tips → Pipet 1 µl 500 mM EDTA into reaction plate

15 Wash tips → Run incubation program 2

Step 4: Annealing

Source plate: Reaction plate from T4 DNA polymerase reaction

20

Reaction plate: 96 well PCR plate (Eppendorf)

Material: bidest. water, 10x DNA Ligase Reaction Buffer (NEB), linearised vector

25 **Incubation program:** 65°C for 8 min. → ramp down 0.4°C/min. to 35°C → 10°C for 1 min.

Procedure:

30 Wash tips → Pipet 150 ng T4 DNA polymerase treated insert DNA according to worklist into reaction plate

Wash tips → Pipet 150 ng linearised vector DNA according to worklist into reaction plate

Wash tips → Run incubation program

Step 5: Transformation in *E. coli*

Source plate: Reaction plate from the annealing step

5

Reaction plate: 96 well PCR plate (Eppendorf)

Culture plate: 2 ml 96 well plate (Nunc)

10 **Target plates:** 12 well cell culture plates containing 2ml of LB-agar with appropriate antibiotics (standard concentrations used: Ampicillin 100 µg/ml, Kanamycin 50 µg/ml, Spectinomycin 50 µg/ml, Chloramphenicol 30 µg/ml)

15 **Material:** *E. coli* cells (Xl1blue) that are chemically competent for transformation , SOC-medium

Transformation program: Heat thermocycler to 42°C
Incubate at 42°C for 30sec.
20 Transfer immediately to cooled (0°C) pipetting carrier

Procedure:

25 Wash tips → Pipet 100 µl competent *E. coli* cells into reaction plate
Wash tips → Pipet 10 µl DNA sample from annealing step into reaction plate
Wash tips → Incubate at 0°C for 30 min.
Run transformation program
Incubate at 0°C for 5 min.
30 Wash tips → Pipet 250 µl SOC-medium into culture plate
Wash tips → Transfer transformation mix into culture plate
Incubate at 37°C and 720 rpm. (Te-Shake Shaker) for 2 h
Wash tips → Pipet 50 µl culture into target plate (agar plate)
Wash tips → Shake target plate at 12 Hz for 1 min. (plating out)

Incubate target plates over night at 37°C

Step 6: Picking clones and setting up over night cultures (manual step)

5 **Source plate:** 12 well cell culture plates containing *E.coli* colonies

Target plate: 24 well culture plate

Material: 2xTY culture medium, incubator which carries culture plates

10

Procedure: Pick 4 colonies per reaction and transfer to 3 ml 2xTY medium in a 24 well culture plate. Incubate at 37°C and approx. 220 rpm over night.

Step 7: Plasmid Extraction (Miniprep)

15

Source plate: 24 well culture plate (usually 3 ml culture)

Target plate: 96 well microtiter elution plate (Macherey-Nagel)

20 **Material:** Plasmid extraction kit, NucleoSpin Robot 96 Plasmid Kit (Macherey-Nagel)

Procedure: According to manufacturer (see
<http://www.machereynagel.com/tabid/10885/default.aspx>)

25

Step 8: Assessment

Plasmid yield is quantified by measuring UV absorbance with a Thermo Scientific NanoDrop™ 1000 Spectrophotometer according to manufacturer. Plasmid integrity
30 was assessed by E-gel (Invitrogen)

The efficacy of the SLIC protocol is assessed in manual and robotics mode. The results of the comparison are shown in Table II. Results are based on a set of 25 different Donor/Acceptor constructions prepared.

Table II: Comparison Manual versus Robotic SLIC procedure (based on 25 constructs each)		
	Manual	EvoII
DNA used for T4 reaction:	200-400 ng insert	400-800 ng insert
	200-400 ng vector	400-800 ng vector
T4 reaction volume for transformation:	5 µl: 2.5 µl (insert) +2.5 µl (vector)	5ul: 2.5 µl (insert) +2.5 µl (vector)
Volume comp. cells (Xl1Blue, chem. comp):	100 µl (+300 µl SOC)	100 µl (+300 µl SOC)
Volume plated	200 µl (Petri dish)	50 µl/well (12 well plate) 200 µl (petri dish)
Clones obtained:	200->2000 (Petri dish)	25-250 (12 well plate) 70-5300 (petri dish)

D.2 Automated Cre fusion process

- 5 A schematic representation of a workflow for automated Cre fusion is shown in Fig. 23.

Step 1: Cre-LoxP Plasmid Fusion Reaction

- 10 **Source plate:** 96 well microtiter elution plate from the plasmid extraction process containing plasmids suitable for Cre-Lox fusion

Reaction plate: 96 well PCR plate (Eppendorf)

- 15 **Material:** bidest. water, 10x Cre reaction buffer (NEB), Cre recombinase (NEB)

Incubation program: 37°C for 1 h → 10°C for 1 min.

Procedure:

- 20 Wash tips → Pipet 6 µl bidest. water into reaction plate
Wash tips → Pipet 2 µl 10x cre reaction buffer into reaction plate

Wash tips → Pipet plasmid DNA suitable for Cre recombination according to
worklist into reaction plate

Wash tips → Pipet 2 µl Cre recombinase into reaction plate

Wash tips → Run incubation program

5 Total reaction volume: 20 µl

Step 2, 3 and 4: Transformation in *E. coli* and Plasmid Extraction:

10 Identical to the method described in above Section D.1., with the exception that
reaction plate from Cre recombination step is used as source plate and recovery time
in SOC-medium is prolonged to a total of 4h. Chemically competent Mach1 cells are
used for transformation. For Cre reaction with 3 and 4 vectors agar-plates with half of
the antibiotic concentration (standard concentrations used: Ampicillin 100 µg/ml,
Kanamycin 50 µg/ml, Spectinomycin 50 µg/ml, Chloramphenicol 30 g/ml) are used.

15

Step 5: Assessment

Plasmid fusion yield is quantified by measuring UV absorbance with a Thermo
Scientific NanoDrop™ 1000 Spectrophotometer according to the manufacturer's
20 instructions. Plasmid integrity is assessed by E-gel (Invitrogen) of undigested and
digested samples. Suitable restriction sites that yield a digestion pattern
characteristic for the respective fusions are identified by using Vector NTI (Invitrogen)
and used for restriction mapping.

25 The efficacy of the Cre reaction is tested by performing a series of fusion reactions,
each in triplicate, by using the Evoll liquid handling workstation. The results are
summarized in Table III.

Table III: Efficiency of Cre-LoxP Reactions on Evoll (assessed in triplicate for each reaction)	
Volume Cre-reaction used for transformation (all reactions):	10 µl
Volume chem. comp. cells (XL1Blue, Mach1) per transformation (all reactions):	100 µl (+300 µl SOC)
Volume transformation reaction plated:	50 µl/well (12 well plate) 200 µl (petri dish)
Clones obtained: (a) Double vector fusion reaction (AD, one Acceptor, one Donor) >1000 fused functional AD plasmids plated on a standard petri dish containing the respective two antibiotics (b) Triple vector fusion reaction (ADD, one Acceptor, two Donors) 12-80 fused functional ADD plasmids plated on a standard petri dish containing the respective three antibiotics (c) Quadruple vector fusion reaction (ADDD, one Acceptor, three Donors) For quadruple vector fusions (ADDD, one Acceptor and three Donors), two possibilities exist. (1) Single reaction ADDD (four vector Cre-Lox fusion, low efficiency) (2) Two step reaction ADD+D: Triple fusion as in (b) , then addition of a further Donor. Option 2 (ADD + D) is preferred for routine robot use as it represents a more robust approach, resulting in example experiments in 20-100 fused functional ADDD plasmids when plated on a standard petri dish containing all four antibiotics.	

D.3. High-throughput micro batch I.MAC

5 **Source plate:** 2 ml deepwell plate (Eppendorf)

Filter plate: Glas filter plate (Novagen)

Target plate: standard microtiter plate (Greiner)

Material: Ni-NTA bulk beads 50% in 20% ethanol (Ge-Healthcare), freezer at -20°C, tabletop centrifuge suitable for microtiter plates, sonication device with microtip, IMAC binding and elution buffer suitable for the specific protein (Berrow et al., Acta Cryst. (2006). D62, 1218 – 1226).

Procedure:

10 **Sample Preparation (off line)**

Harvest *E. coli* cells expressing the desired protein by centrifugation at 3000 g (4°C) directly in the source plate

Freeze cell pellets for 30 min. at -20°C

Thaw cell pellets 15 min. at room temperature

15 **Preparation of the filter plate**

Wash tips → Resuspend Ni-NTA bead suspension by pipetting up and down 20 times 200 µl → Transfer 200 µl bead suspension to filter plate

Wash tips → Apply vacuum 550 mbar for 30 sec. (remove 20% ethanol)

Wash tips → Pipet 1 ml equilibration buffer (e.g. binding buffer) to resin

20 Wash tips → Apply vacuum 300 mbar for 60 sec. (equilibration)

IMAC purification, preparation

Wash tips → Pipet 1 ml binding buffer to the samples in the source plate

Wash tips → Resuspend cell pellets by pipetting up and down 10 times 750 µl

25 Wash tips

Sonication of samples (off line)

Sonication of the samples to insure complete lysis of the cells

30 **IMAC purification, loading and elution**

Wash tips → Transfer whole lysate to filter plate

Wash tips → Apply vacuum 300 mbar for 90 sec. (binding step)

Wash tips → Pipet 1 ml wash buffer to the samples

Wash tips → Apply vacuum 300 mbar for 90 sec. (wash step)

Repeat wash step 3 times

Wash tips → Pipet 100 µl elution buffer to the samples

Wash tips → Incubate 3 min. at room temperature

Apply vacuum 650 mbar for 90 sec. (elution step)

5

Assessment

Eluted samples (10 µl - 12 µl) are loaded manually on 12% denaturing gels using a Biorad Minigel System, pre-run at 135 V for 25 min, and then run for 65-70 min. at 185 V. Gels are stained with Coomassie Brilliant Blue according to standard procedures.

10

E. ACEMBL kit for expression of proteins in prokaryotic hosts

A kit according to a preferred embodiment for expression in prokaryotic hosts contains:

15

- BW23473, BW23474 cells¹ and/or Cre recombinase
- pACKS quadruple fusion vector²
made of pACE (Acceptor), and pDC, pDK, pDS (Donors)
- pACE2 vector
- 20 - pACE-[VHLbc/BFP/mGFP] control plasmid
triple fusion vector made of pACE-VHLbc, pDK-BFP, pDS-mGFP³

¹ *E. coli* strains expressing the *pir* gene for propagation of Donor derivatives (any other strain with *pir*⁺ background can be used).

² This fusion vector was created by Cre-LoxP reaction of pACE, pDC, pDK and pDS.

25

It is resistant to ampicillin, kanamycin, chloramphenicol and spectinomycin. Individual ACEMBL vectors can be liberated from this quadruple fusion by Cre-LoxP mediated deconstruction as described above in protocol C.2.2. Sequences for single ACEMBL vectors according to the present embodiment and pACKS quadruple fusion are provided in SEQ ID NO: 2 to 7.

30

³ pDS-mGFP contains a coiled-coil fused to the N-terminus of eGFP (see Berger et al. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 12177-82).

Optional components:

- Antibiotics: ampicillin, chloramphenicol, kanamycin, spectinomycin, tetracycline
- Enzymes:
 - 5 T4 DNA polymerase (for recombination insertion of genes)
 - Phusion polymerase (for PCR amplification of DNA)
 - Restriction enzymes and T4 DNA ligase (for conventional cloning, if desired)

The present invention is further illustrated by the following non-limiting examples.

EXAMPLES

Examples of multiprotein expressions by using the above-described ACEMBL system are shown in the following illustrating the gene combination procedures outlined above. Reactions presented were either carried out manually following the protocols provided in above Section C. or on a Tecan Freedom Evoll 200 robot with adapted protocols according above Section D.

Example 1: SLIC cloning into ACEMBL vectors: human TFIIIF

Genes coding for full-length human RAP74 with a C-terminal oligo-histidine tag and full-length human RAP30 were amplified from pET-based plasmid template (Gaiser et al. (2000) *J. Mol. Biol.* **302**, 1119-1127) by using the primer pair T7InsFor (5'-TCCCGCGAAATTAATACGACTCACTATAGGG-3'; SEQ ID NO: 20) and T7Insrev (5'-CCTCAAGACCCGTTTAGAGGCCCAAGGGGTTATGCTAG-3'; SEQ ID NO: 21) following the protocols described above. Linearized vector backbones were generated by PCR amplification from pACE and pDC by using primer pair T7VecFor (5'-CTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG-3'; SEQ ID NO: 22) and T7VecRev (5'-CCCTATAGTGAGTCGTATTAATTTTCGCGGGA-3'; SEQ ID NO: 23) in both cases. Above Protocol 1 (Section C) was followed, resulting in pACE-RAP30 and pDC-RAP74his (Fig 8). These plasmids were fused by Cre-LoxP reaction (see above Section C). Results from restriction mapping by BstZ17I/BamHI double digestion of 11 double resistant (Cm, Ap) colonies is shown by a gel section from 1% E-gel electrophoresis (M: NEB 1kb DNA marker) in Fig. 8. All clones tested showed

the expected pattern (5.0 + 2.8 kb). One clone was transformed in BL21(DE3) cells. Expression and purification by Ni²⁺-capture and S200 chromatography resulted in human TFIIIF complex (Fig. 21A).

- 5 The high-level soluble expression of full-length human TFIIIF (Fig. 21A) is noteworthy, as individual expression of the subunits invariably leads to insoluble material. In the past, crystal structure analysis of human TFIIIF dimerization domain had necessitated many iterative cycles of limited proteolysis, recloning, insoluble expression of the designed fragments and co-refolding (Gaiser et al. (2000), *supra*). Similar laborious
10 situations are commonplace in prior art protein complex research. It is conceivable that the large investment of labor involved can now be significantly reduced using the nucleic acids and vectors of the present invention, in particular the ACEMBL system.

15 **Example 2: Polycistron insertion by SLIC: human VHL/elongin b/elongin c complex.**

The gene encoding for Von Hippel Lindau protein (amino acids 54-213), fused at its N-terminus to a six-histidine-thioredoxin fusion tag, was PCR amplified from plasmid pET3-HisTrxVHL by using primers T7InsFor (see above Table I) and SmaBamVHL
20 (5'-GAATTCACCTGGCCGTCGTTTTACAGGATCCTTAATCTCCCATCCGTTGATG TGCAATG-3'; SEQ ID NO: 45). SmaBamVHL primer is a derivative of the SmaBam adaptor sequence (Table I; SEQ ID NO: 17) elongated at its 3' by the insert specific sequence at the 3' end of the VHL gene (including a stop codon). The gene encoding for full-length elongin b was PCR amplified from pET3-ElonginB by using
25 primers BamSmaEB (5'-GGATCCTGTAAAACGACGGCCAGTGAATTCG CTAGCTCTAGAAATAATTTTGTTTAAC-3'; SEQ ID NO: 46) and SacHindEB (5'-GAGCTCGACTGGGAAAACCCTGGCGAAGCTTAGATCTGGATCCTTACTGCACG GCTTGTTCAATTGG-3'; SEQ ID NO: 47), which are derivatives of the corresponding adaptors (Table I). The gene for elongin c (amino acids 17-112) was amplified from
30 pET3-ElonginC by using primers HindSacEC (5'-AAGCTTCGCCAGGGTTTTCCCA GTCGAGCTCCAATTGGAATTCGCTAGCTCTAG-3'; SEQ ID NO: 48) and BspEco5EC (5'-GATCCGGA TGTGAAATTGTTATCCGCTGGTACCAAGCTTAGAT CTGGATCCTTAACAATCTAAGAAG-3'; SEQ ID NO: 49), which are derivatives of the corresponding adaptors (Table I). Vector backbone was PCR amplified by using

primers Tn7VecRev and Eco5Bsp, and pACE as a template (Fig. 9). Multifragment SLIC was carried out according to above Protocol 2 (Section C) resulting in pACE-VHLbc which contains a tricistron. Clones were plated on agar plates containing ampicillin. A positive clone, verified by sequencing, was used in the coexpression experiment described below (Example 5).

Example 3: The Homing endonuclease/BstXI module: yeast RES complex

Plasmids pCDFDuet-Pml1p, pRSFDuet-Snu17p-NHis and pETDuet-Bud13p, coding for yeast proteins (all full-length) Pml1p, Snu17p and Bud13p, respectively, were provided by Dr. Simon Trowitzsch and Dr. Markus Wahl (Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany). Snu17p contains a six-histidine tag fused to its N-terminus. The gene encoding for His6-tagged Snu17p was excised from pRSFDuet-Snu17p-NHis by using restriction enzymes NcoI and XhoI, and ligated into a NcoI/XhoI digested pACE construct (containing an unrelated gene between NcoI and XhoI sites) resulting in pACE-Snu17. The gene encoding for Bud13p was liberated from pETDuet-Bud13p by restriction digestion with XbaI and EcoRV, and placed into XbaI/PmeI digested pDC resulting in pDC-Bud13. The gene encoding for Pml1p was liberated from pCDFDuet-Pml1p by restriction digestion with NdeI and XhoI, and placed into NdeI/XhoI digested pDC resulting in pDC-Pml1. Next, the expression cassette for Bud13p was liberated from pDC-Bud13 by digestion with PI-SceI and BstXI. The liberated fragment was inserted into PI-SceI digested and alkaline phosphatase treated pDC-Pml1p resulting in pDC-Bud13p-Pml1p.

pACE-Snu17 and pDC-BudPml were fused by Cre-LoxP reaction and selected for by plating on agar plates containing ampicillin and chloramphenicol. Fusion plasmids were transformed into BI21(DE3) cells. Expression and purification by Ni²⁺-capture and S200 size exclusion chromatography resulted in the trimeric RES complex.

The strategy for cloning the yeast RES complex according to the method of the present invention is schematically illustrated in Fig. 10.

Example 4: Coexpression by cotransformation: human NYB/NYC

Genes encoding for protein NYB (amino acids 49-141) and NYC (amino acids 27-12)
5 were excised from vectors pACYC18411-NYB and pET15-NYC, respectively (Romier
et al. (2003) *J. Biol. Chem.* **278**, 1336-1345). NdeI and BamHI were used for NFYB.
XbaI and BamHI were used for NYC, thus importing a six-histidine tag at the N-
terminus of the protein. The NYB insert was ligated into pACE digested with NdeI
and BamHI. The NYC insert was ligated into pACE2 digested by XbaI and BamHI.
10 pACE-NFYB and pACE2-NFYC were transformed into BL21(DE3) cells containing
the pLysS plasmid. Selection on agar plates containing ampicillin, tetracycline and
chloramphenicol resulted in triple resistant colonies. The complex was expressed
and purified by Ni²⁺ capture (IMAC) and S75HR (Pharmacia) size exclusion
chromatography.

Example 5: Coexpression from Acceptor-Donor fusions

Six heterologous genes coding for a trimeric protein complex (VHLbc: VonHippel-
Lindau protein amino acids 54-213 / full-length elonginB / elonginC amino acids 17-
20 112) (Stebbins et al. (1999) *Science* **284**, 455-61), a gene encoding for the AAA
ATPase FtsH (amino acids 147-610), and two genes encoding for fluorescent
markers (BFP and GFP) were assembled as illustrated in Fig. 20. In a single Cre
reaction, all combinations of one Acceptor (pACE-VHLbc) and three Donors (pDC-
FtsH, pDK-BFP, pDS-mGFP) were obtained and selected, including a quadruple
25 fusion containing all six heterologous genes; see Fig. 20). Clones were verified by 96
well microtiter assay as described above for the ACEMBL system, Section C.
Expression and Ni²⁺ affinity capture, combined with immunostaining of the untagged
fluorescent markers, confirmed successful multiprotein expression (Figs. 16 and
17B). Proteins were expressed overnight in BL21(DE3) cells in 24 well deep-well
30 plates in small scale using autoinduction media (Studier (2005) *Protein Expr. Purif.*
41, 207-34). Restriction mapping revealed that even large fusion plasmids were
stable over many (more than 60) generations, even if challenged by a single
antibiotic in the medium only.

Example 6: Expression of the YidC-SecYEGDF holotranslocon

As illustrated in Fig. 21C, the ACEMBL system was used to produce a large multiprotein complex, the YidC-SecYEGDF holotranslocon that contains in total 33 transmembrane helices. This machinery is used to transport unfolded polypeptides into the cell membrane or for translocation into the periplasm of bacteria (Duong et al. (1997) *EMBO J.* **16**, 2757-68).

Example 7: Expression of human IKK complex in insect cells

Following the protocols for single gene insertion into ACEMBL vectors as outlined above in Section C.1., the genes for IKK1 (also called IKKalpha), IKK2 (also called IKKbeta) and IKK3 (also called Nemo) were cloned into pACEBac1, pIDC and pIDS respectively (maps of the resulting plasmids pACEBac1-HisIKK1, pIDC-CSIKK2 and pIDS-IKK3 are shown in Fig. 46, 47 and 48, respectively). IKK1-2 double fusion (pACEBac1-HisIKK1 with pIDC-CSIKK2) and IKK1-2-3 triple fusions (all three vectors) were created by Cre-LoxP fusions as outlined above in Section C.2. The fusions were introduced into suitable host cells carrying a baculovirus genome (EMBac) as a bacterial artificial chromosome. The vector fusions were integrated into the baculoviral genome via Tn7 transposition. Productive integration was assessed by blue/white screening. DNA of composite virus was prepared from white clones and transfected into Sf21 cells.

Example 8: Expression of a H1N1-influenza virus-like particle

A virus-like particle (VLP) of the swine-flu virus (influenza virus of type H1N1) comprising the proteins HA, NA, M1 and M2 was expressed in insect cells (Sf21) by the following strategy: genes coding for HA and NA were cloned into pACEBac1 by single gene insertion as outlined above in Section C.1. The same procedure was followed for cloning the genes coding for M1 and M2 into pIDC. Double expression cassettes for HA-NA and M1-M2, respectively, were generated by using the HE-BstXI sites in the respective MIE (see above Section C.1.4.) resulting in plasmids pACEBac-HA-NA (plasmid map see Fig. 49) and pIDC-M1-M2 (plasmid map see Fig.

50). The vector for coding the complete H1N1-influenza-VLP was generated by CreLoxP fusion of pACEBac-HA-NA with pIDC-M1-M2 following the protocol in above Section C.2. The fusion vector was introduced into suitable host cells carrying a baculovirus genome (EMBac) as a bacterial artificial chromosome. The vector
5 fusions were integrated into the baculoviral genome via Tn7 transposition. Productive integration was assessed by blue/white screening. DNA of composite virus was prepared from white clones and transfected into Sf21 cells.

Claims

- 5 1. A nucleic acid containing at least one homing endonuclease site (HE) and at least one restriction enzyme site (X) wherein the HE and X sites are selected such that HE and X result in compatible cohesive ends when cut by the homing endonuclease and restriction enzyme, respectively, and the ligation product of HE and X cohesive ends can neither be cleaved by the homing endonuclease nor the restriction enzyme.
- 10 2. The nucleic acid of claim 1 having the sequence elements HE-Prom-MCS-Term-X or HE-Prom-MCS-X wherein
- 15 Prom: represents a promoter;
MCS: represent a multiple cloning site; and
Term: represents a terminator.
- 20 3. The nucleic acid of claim 1 or 2 wherein the HE site is a recognition sequence that results in a 4 nucleotide overhang when cut by the respective homing endonuclease.
- 25 4. The nucleic acid of claim 3 wherein the HE site is a recognition sequence of a homing endonuclease selected from the group consisting of the group consisting of PI-SceI, I-CeuI, I-PpoI, I-HmuI I-CreI, I-DmoI, PI-PfuI, I-MsoI, PI-PspI, I-SceI, SegH, Hef, I-ApeII, I-AniI, Cytochrome b mRNA maturase bl3, PI-TiiI, PI-Tfull and PI-Thyl .
- 30 5. The nucleic acid of claim 3 or 4 wherein the X site is a BstXI site.
6. The nucleic acid according to any one of claims 2 to 5 wherein the MCS contains one or more homology regions.
- 35 7. The nucleic acid according to any one of the preceding claims comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52 and SEQ ID NO: 53.

8. The nucleic acid according to any one of the preceding claims additionally comprising at least one site for integration of the nucleic acid into a vector or host cell.

9. A vector comprising the nucleic acid according to any one of the preceding claims.

10. The vector of claim 9 further containing at least one recognition sequence for a site-specific recombinase, preferably a loxP imperfect inverted repeat or a Tn7 attachment site.

11. The vector of claim 9 or 10 having a sequence selected from the group consisting of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16 and SEQ ID NO: 17.

12. The vector of claim 11 containing more than one of the sequence elements of the nucleic acid as defined in any one of claims 1 to 7 and containing more than one recognition sequence for a site-specific recombinase.

13. The vector of claim 12 having the sequence of SEQ ID NO: 18.

14. The vector of any one of claims 9 to 12 wherein the vector is a virus, preferably a baculovirus.

15. A host cell containing the nucleic acid according to any one of claims 1 to 8 and/or the vector according to any one of claims 9 to 14.

16. A kit for cloning and/or expression of multiprotein complexes containing at least one vector according to any one of claims 9 to 14 together with at least one host cell suitable for the propagation of said vector(s).

17.A method for producing a vector containing multiple expression cassettes comprising the steps of:

- (a) inserting one or more genes between the HE and the X site of a first vector according to any one of claims 9 to 14;
- 5 (b) inserting one or more genes between the HE and the X site of a second vector according to any one of claims 9 to 14;
- (c) cleaving the first vector with a homing endonuclease specific for site HE and with a restriction enzyme specific for site X yielding a fragment containing the at least one gene flanked by the cleaved HE and X site;
- 10 (d) cleaving the second vector with a homing endonuclease specific for site HE;
- (e) ligating the fragment obtained in step (c) into the cleaved second vector obtained in step (d) generating a third vector; and optionally
- (f) repeating steps (a) to (e) with one or more vector(s) generating a vector
- 15 containing multiple genes.

18.A method for producing multiprotein complexes comprising the steps of

- (i) producing a vector containing multiple expression cassettes by the method of claim 17;
- 20 (ii) introducing the vector obtained in step (i) into a host cell; and
- (iii) incubating the host cell under conditions allowing the simultaneous expression of the genes present in the vector.

19. Use of the nucleic acid according to any one of claims 1 to 8 and/or the vector

25 of any one of claims 9 to 14 for the cloning and/or expression of multiple expression cassettes.

20.A method for assembling n vector entities into 1 to (n-1) fusion vectors wherein said fusion vector(s) contain(s) 2 to n of said vector entities

30 comprising the steps of:

- (1) contacting n vector entities each containing a site-specific recombination site and an individual resistance marker different from the resistance markers of the other vector entities with a recombinase specific for said site-specific recombination site so as to generate a

mixture of fusions of the vector entities comprising 2 to n of said vector entities,

- (2) transforming said mixture into host cells;
- (3) culturing one or more sample(s) of the transformed cells in the presence of an appropriate combination of antibiotics for selecting one or more desired fusion vector(s) containing 2 to n vector entities;
- (4) obtaining n single clones of transformed cells from the culture obtained in step (3) in which these were viable in the presence of the respective combination of antibiotics; and
- (5) culturing n samples of each of said n single clones in the presence of each of n antibiotics specific for the n individual resistance markers present in said n vector entities;

wherein n is an integer of at least 3, preferably 3 to 5.

21. The method of claim 20 wherein (n-1) of the vector entities to be fused each contains a further selectable marker different from the resistance markers such that only host cells transformed with fusions between the vector entity not containing the further selectable marker and one or more of the vector entities containing the selectable marker are viable in step (3).

22. The method of claim 21 wherein (n-1) of the vector entities contain a conditional origin of replication making the propagation of said vector entities dependent on the presence or absence of a specific gene in the host cells.

23. The method of claim 22 wherein the host cells are bacteria, preferably *E. coli*, the origin of replication is R6K γ or a derivative thereof and the bacteria are *pir*⁻.

24. The method according to any one of claims 20 to 23 wherein each of the n vector entities contains one or more genes of interest, preferably within an expression cassette.

25. A method of disassembling a fusion vector containing n vector entities into one or more desired fusion vectors selected from the group consisting of fusion vectors containing 2 to (n-1) vector entities or into one or more of said single

vector entities, wherein in said fusion vector containing n vector entities said n vector entities are separated from each other by n site-specific recombination sites and each vector entity contains an individual resistance marker different from the resistance markers of the other vector entities, comprising the steps of:

- (A) contacting the fusion vector containing n vector entities with a recombinase specific for said site-specific recombination sites in order to generate a mixture of fusions of the vector entities comprising 2 to (n-1) of said vector entities and single vector entities;
- (B) transforming said mixture into host cells; and
- (C) culturing one or more sample(s) of the transformed cells in the presence of
 - (C1) an appropriate combination of antibiotics for selecting one or more desired fusion vector(s) containing 2 to (n-1) vector entities; and/or
 - (C2) a single appropriate antibiotic for selecting a desired single vector entity;
- (D) obtaining n single clones of transformed cells from the sample of the transformed cells in which these were viable in the presence of the respective antibiotic or combination of antibiotics, respectively; and
- (E) culturing n samples of each of said n single clones in the presence of each of n antibiotics specific for the n individual resistance markers present in said n vector entities;

wherein n is an integer of at least 3, preferably 3 to 5.

26. The method of claim 25 wherein for disassembling the fusion vector containing n vector entities into single vector entities steps (A), (B) and (C1) are carried out for selecting an appropriate fusion vector containing 2 to (n-1) vector entities and steps (A), (B) and (C2) to (E) are carried out with said selected fusion vector containing 2 to (n-1) vector entities.

27. The method of claim 26 wherein (n-1) of the vector entities in said fusion vector containing n vector entities each contains a further selectable marker different from the resistance markers such that only host cells transformed with fusions between a vector entity not containing the further selectable

marker and one or more of the vector entities containing the selectable marker are viable in step (C1).

5 28. The method of claim 27 wherein (n-1) of the vector entities contain a conditional origin of replication making the propagation of said vector entities dependent on the presence or absence of a specific gene in the host cells.

29. The method of claim 28 wherein the host cells are bacteria, preferably *E. coli*, the origin of replication is R6Ky or a derivative thereof and the bacteria are *pir*⁻.

10 30. The method according to any one of claims 25 to 29 wherein each of the n vector entities contains one or more genes of interest, preferably within an expression cassette.

15 31. A fusion vector comprising n vector entities separated from each other by n of the same site-specific recombination site wherein each vector entity contains an individual resistance marker gene different from the resistance marker genes of the other vector entities, wherein n is an integer of at least 3, preferably 3 to 5

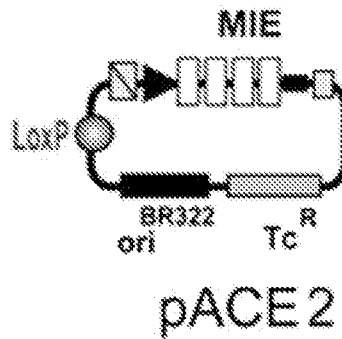
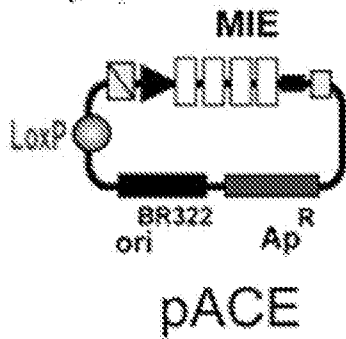
20 32. A kit for assembly and/or disassembly of n vectors comprising

- a fusion vector comprising n vector entities separated from each other by n of the same site-specific recombination site wherein each vector entity contains an individual resistance marker gene different from the resistance marker genes of the other vector entities; and/or
- 25 - n vectors (vector entities) each containing a site-specific recombination site and an individual resistance marker gene different from the resistance marker genes of the other vectors, and
- a recombinase specific for said site-specific recombination site and/or
- 30 cells for the propagation of said fusion vector and/or said n vectors;

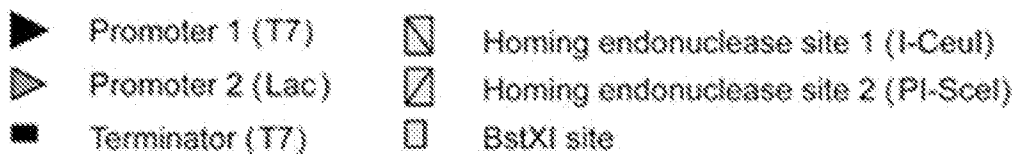
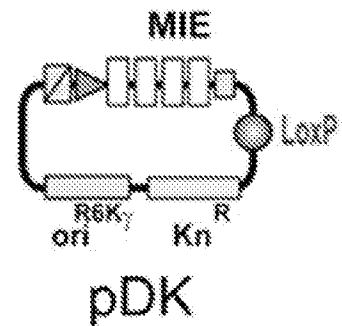
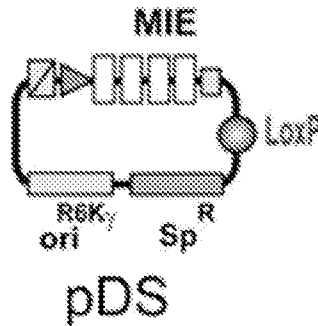
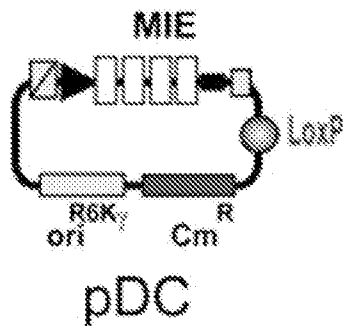
wherein n is an integer of at least 3, preferably 3 to 5.

1/45

Acceptor



Donor



MIE Multiple Integration Element

LoxP 31bp imperfect inverted repeat for Cre mediated fusion

R6K_γ ori conditional origin of replication from phage R6K_γ

BR322 ori regular origin of replication

Resistance markers Ampicillin (Ap), Tetracycline (Tc), Chloramphenicol (Cm), Spectinomycin (Sp), Kanamycin (Kn)

Fig. 1

2/45

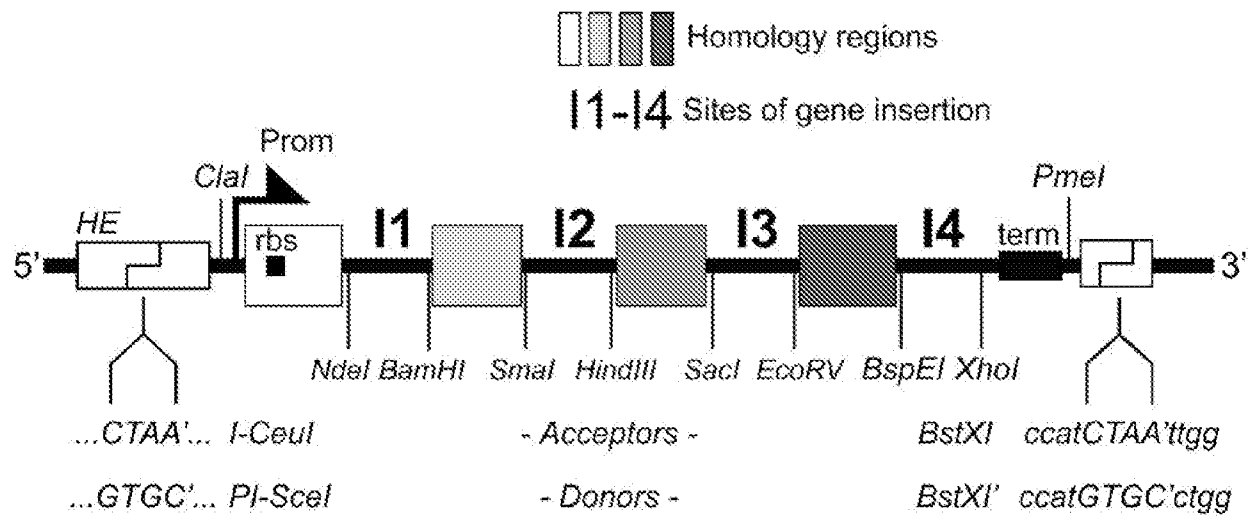


Fig. 2

3/45

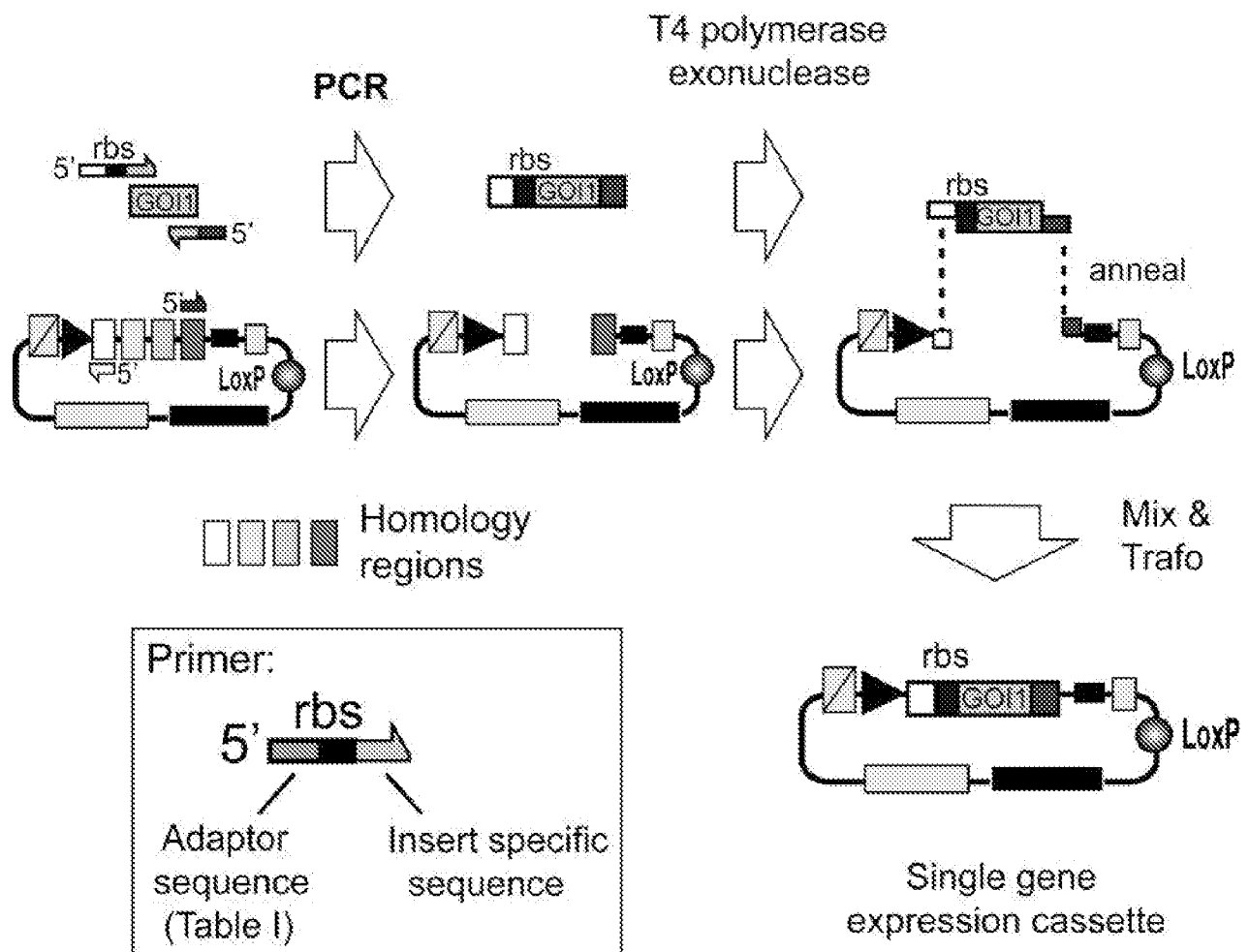


Fig. 3

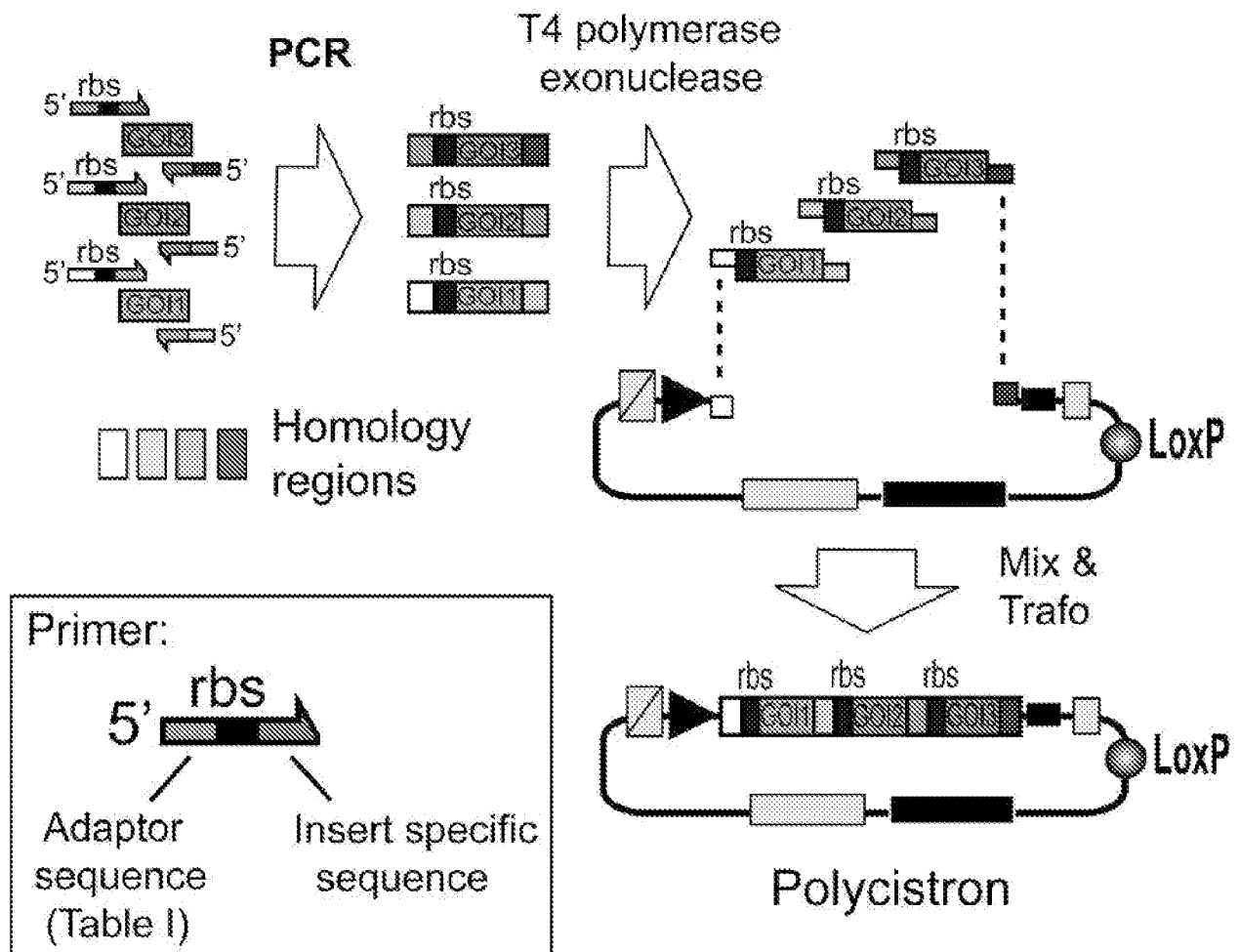


Fig. 4

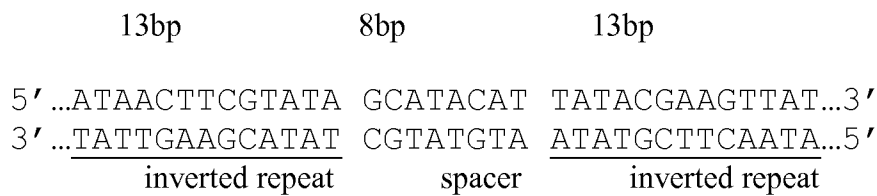


Fig. 5

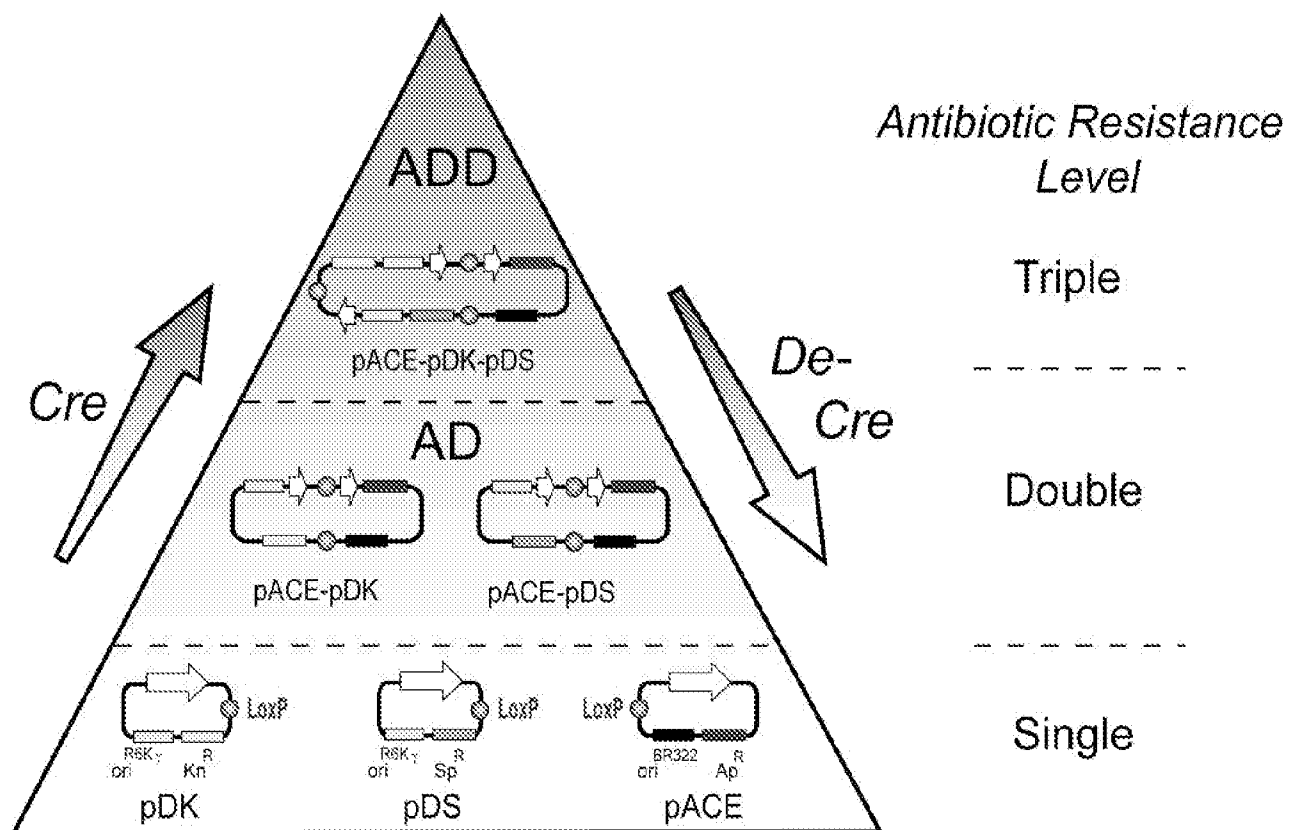


Fig. 6

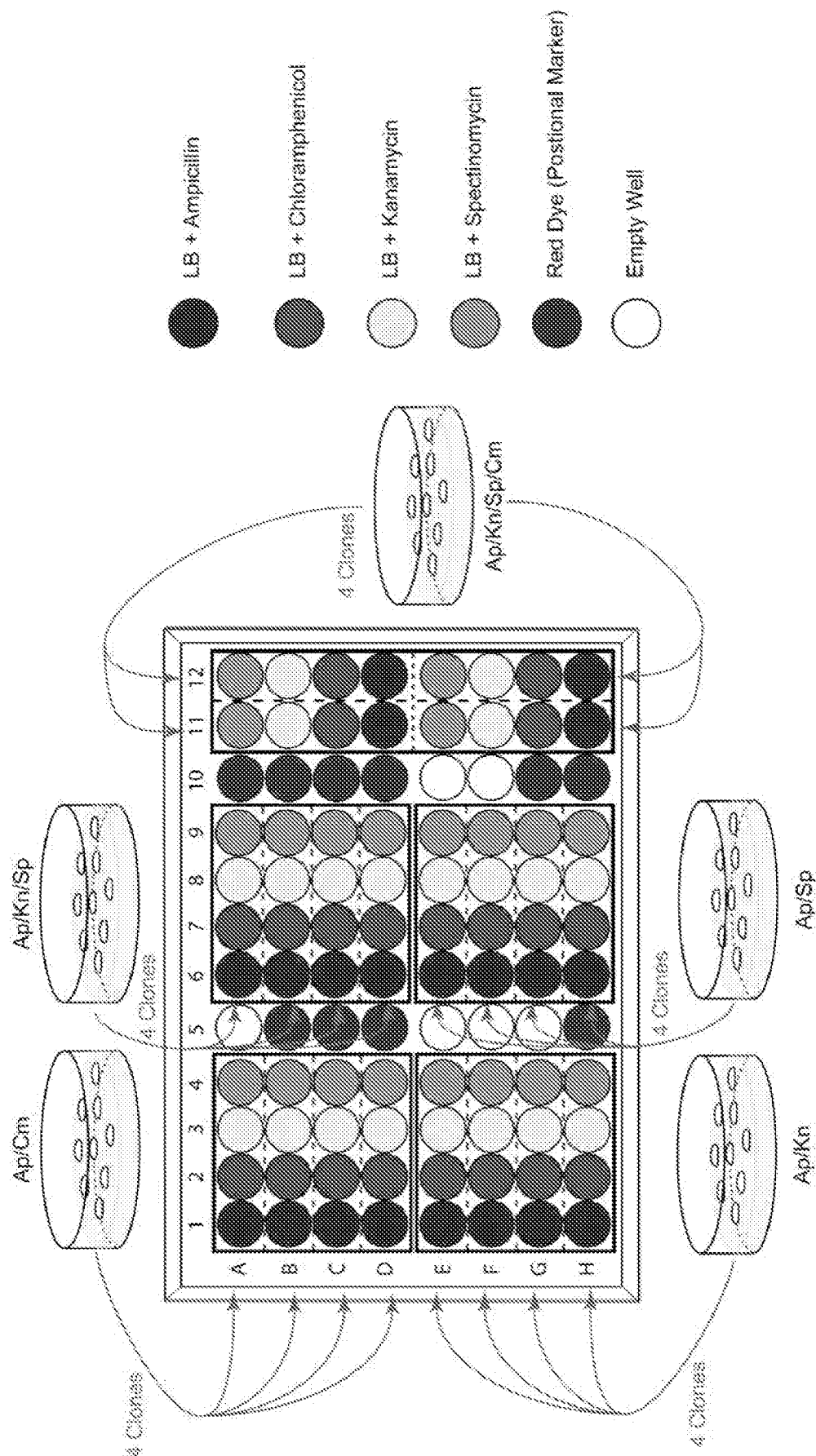


Fig. 7

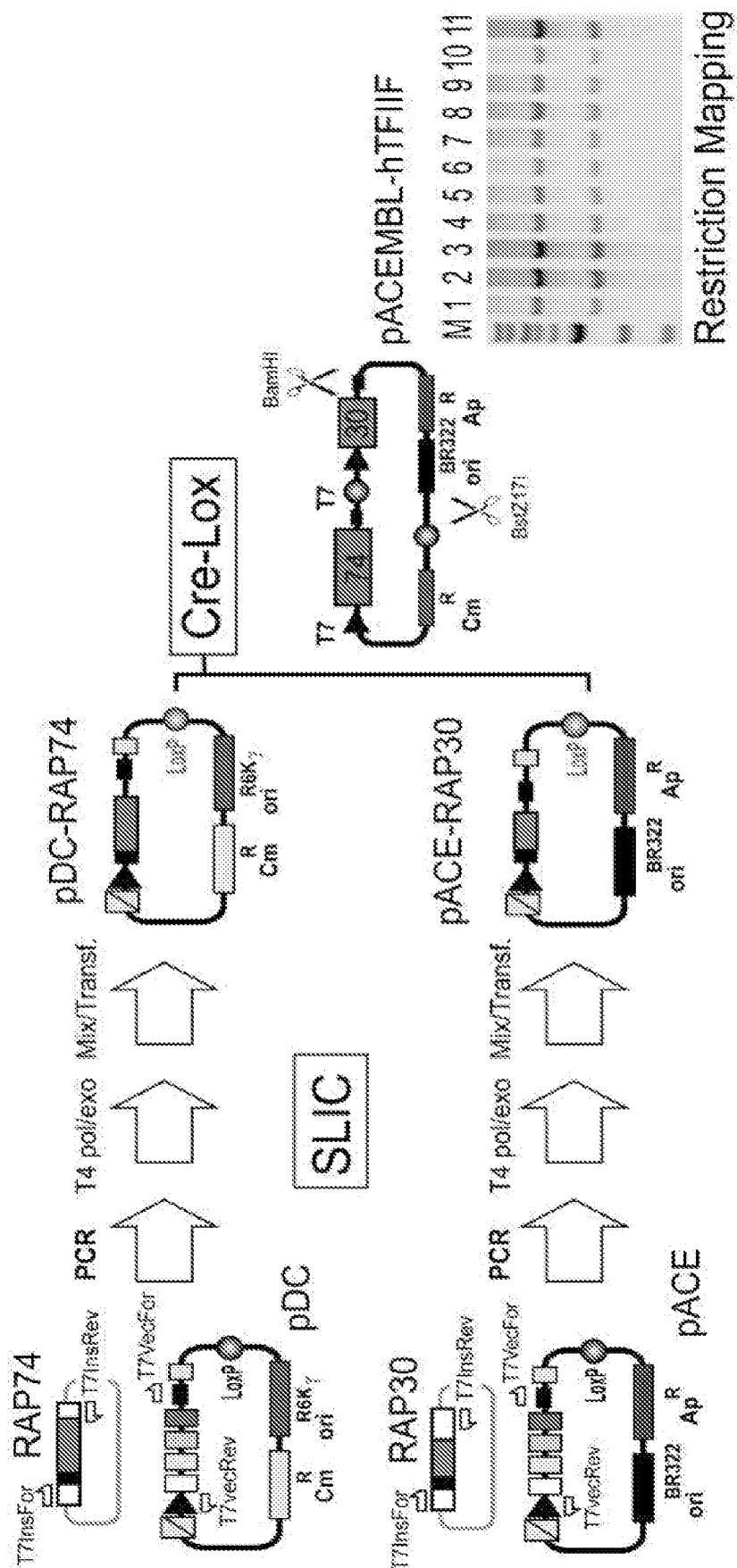


Fig. 8

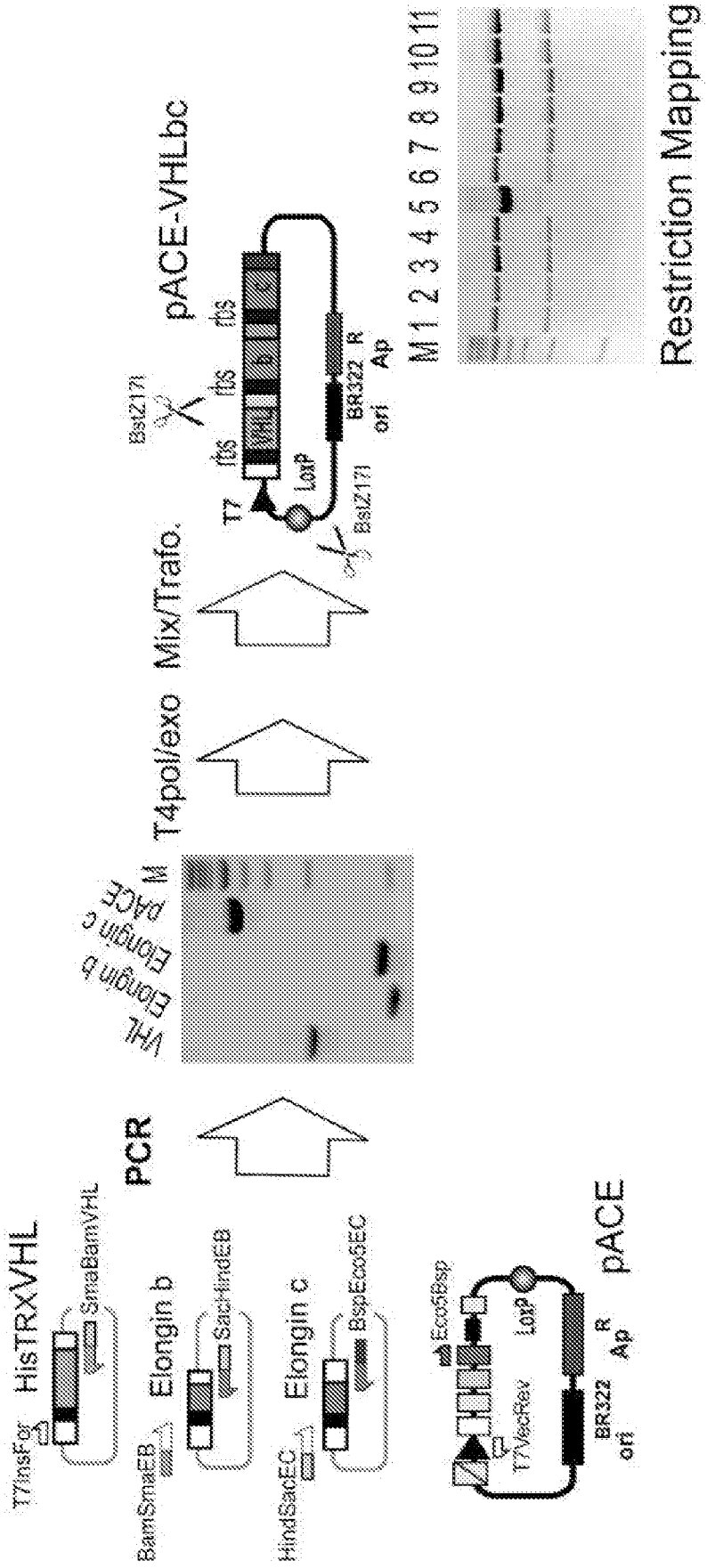


Fig. 9

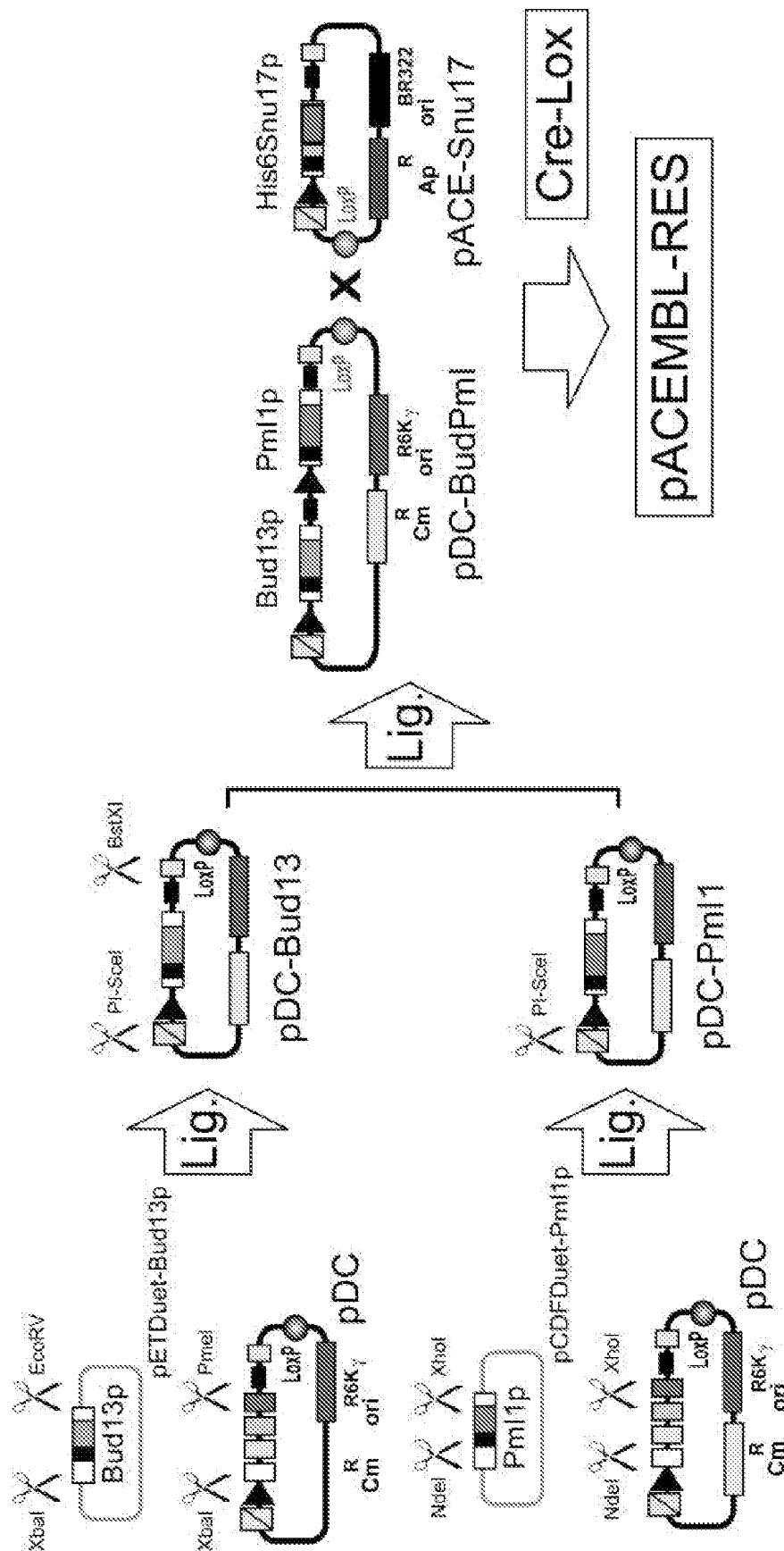


Fig. 10

10/45

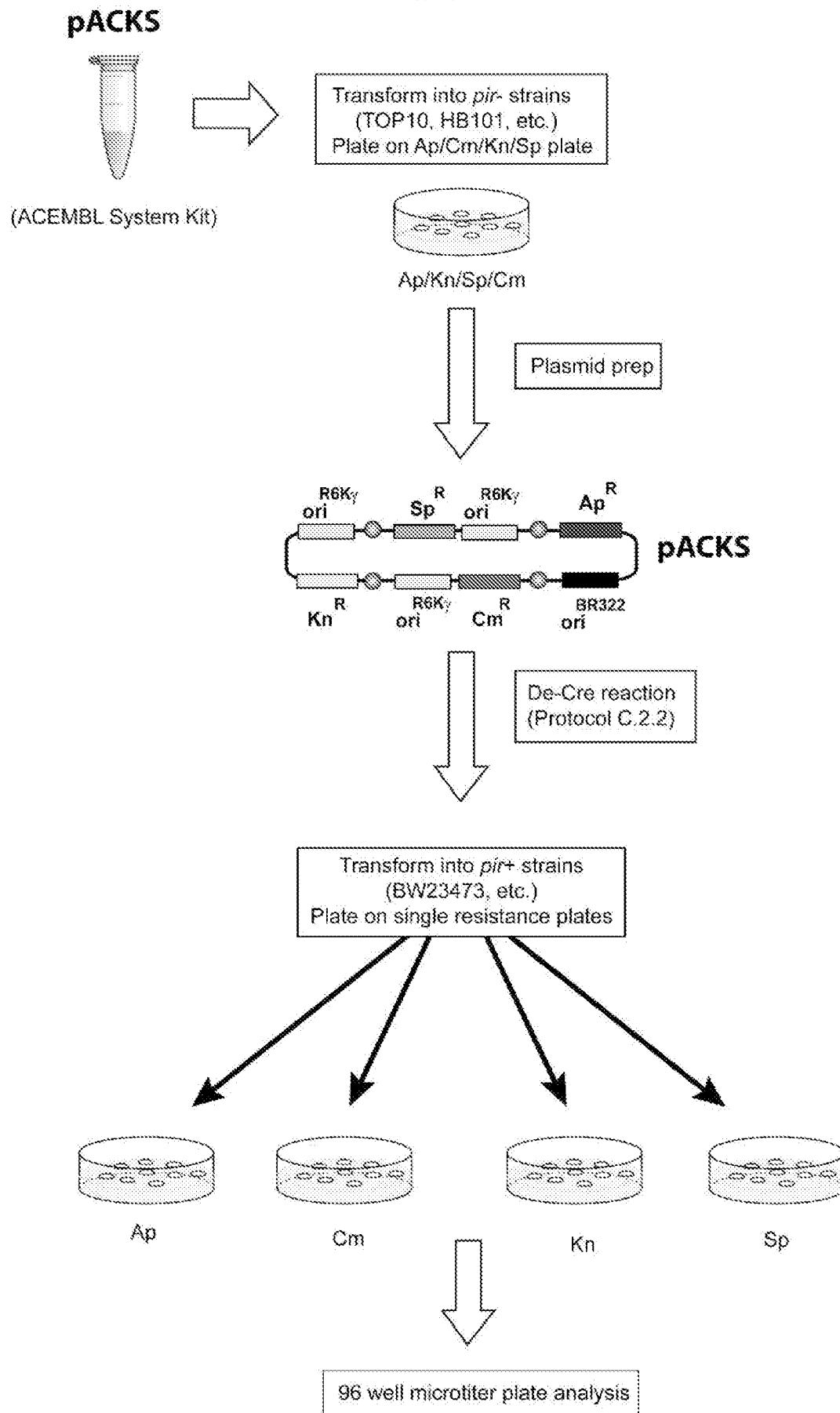
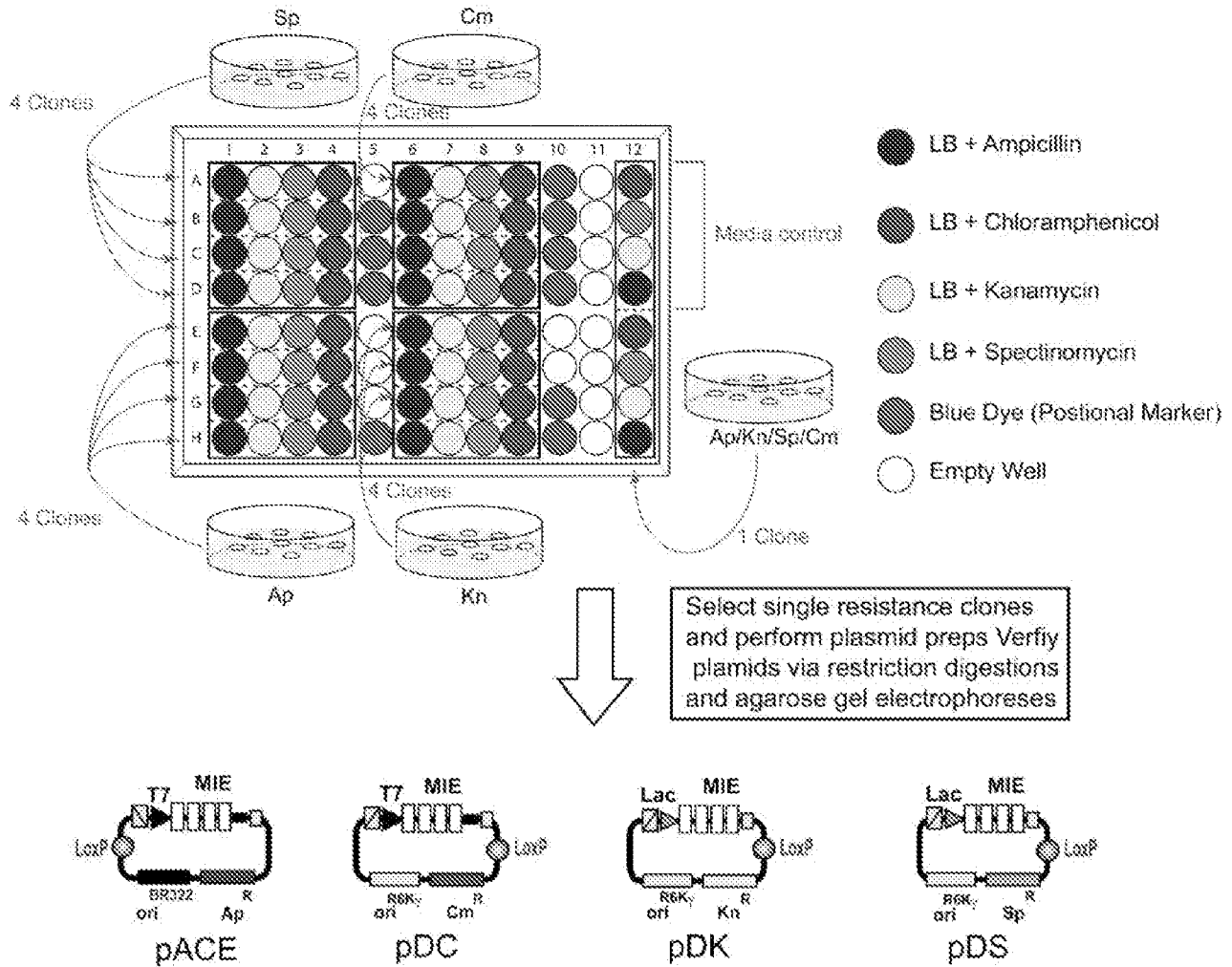
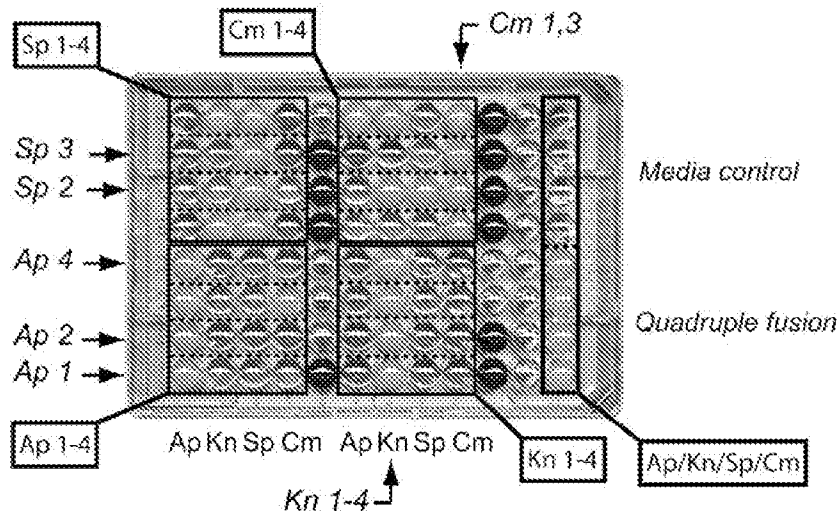


Fig. 11

11/45



Example for De-Cre 96 well microtiter plate analysis



Deconstruction of a quadruple fusion vector (example). De-Cre reaction of the fusion resulted in single vectors (i.e. total deconstruction) in more than 50% of the clones tested (marked by arrows). Double (Ap3, Sp1, Sp4) and even triple fusions (Cm2, Cm4) are also present. These can be deconstructed into single vectors by a second De-Cre reaction.

Fig. 12

12/45

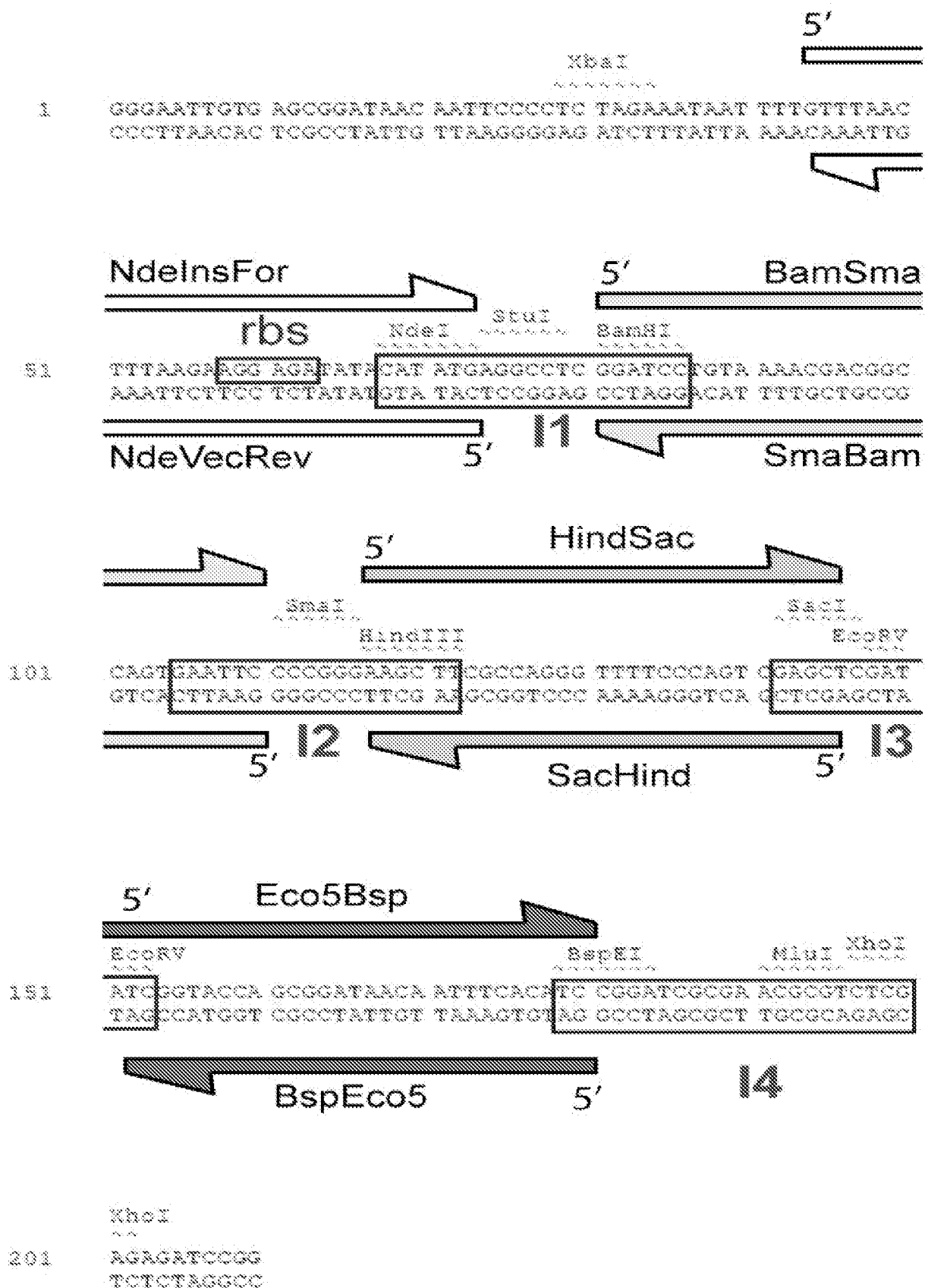


Fig. 13A

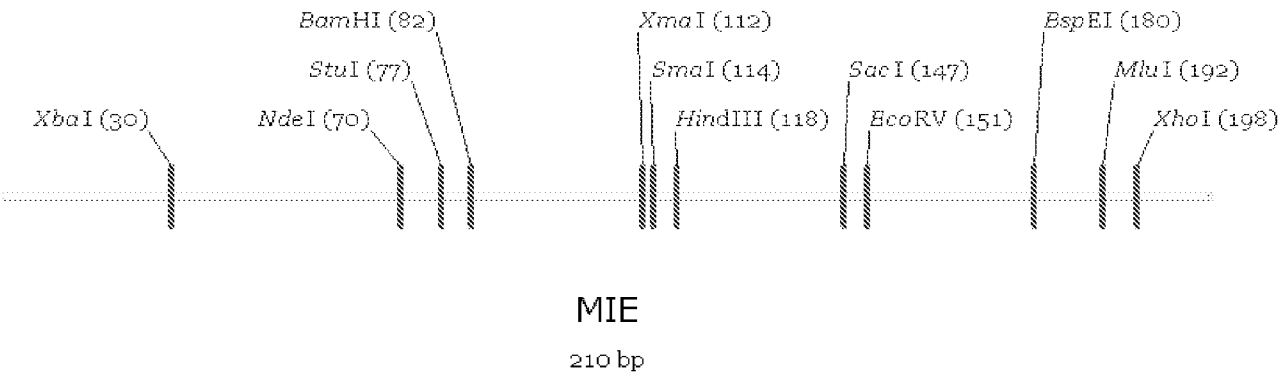


Fig. 13B

14/45

pACE

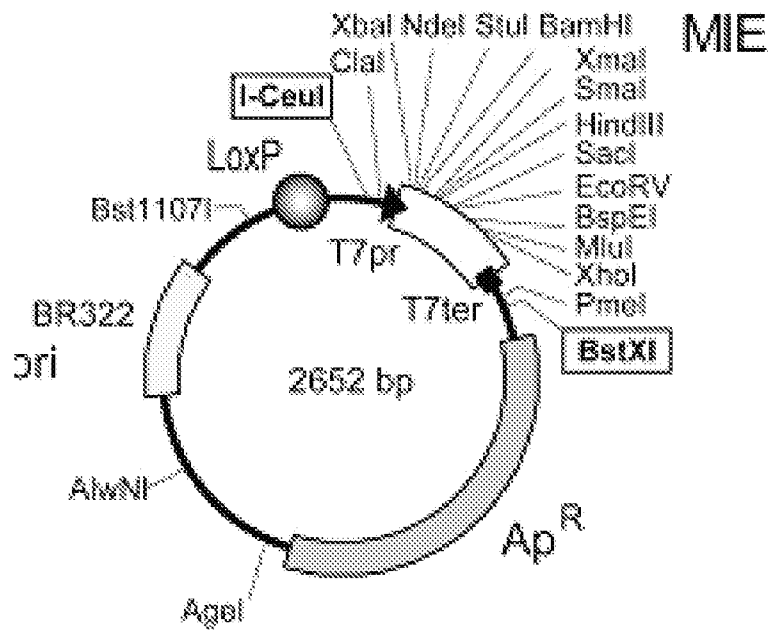


Fig. 14

pACE2

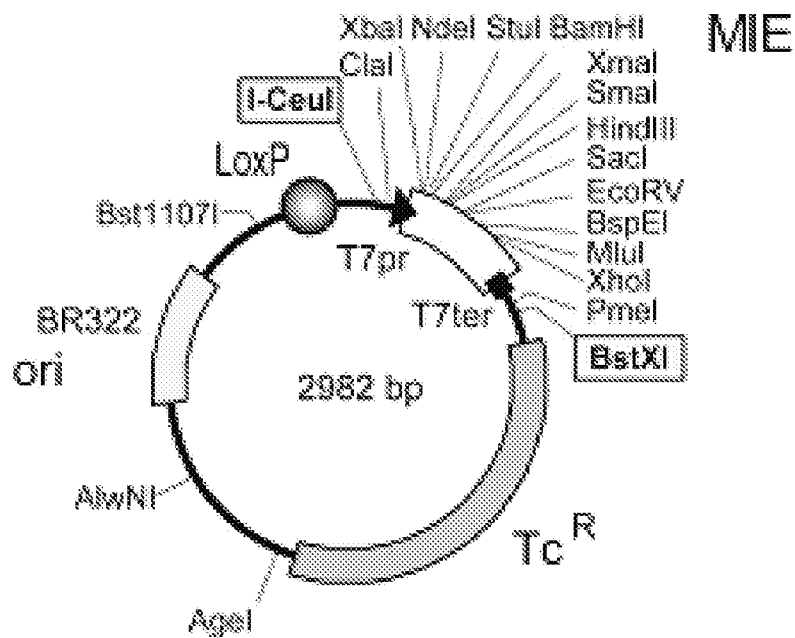


Fig. 15

15/45

pDC

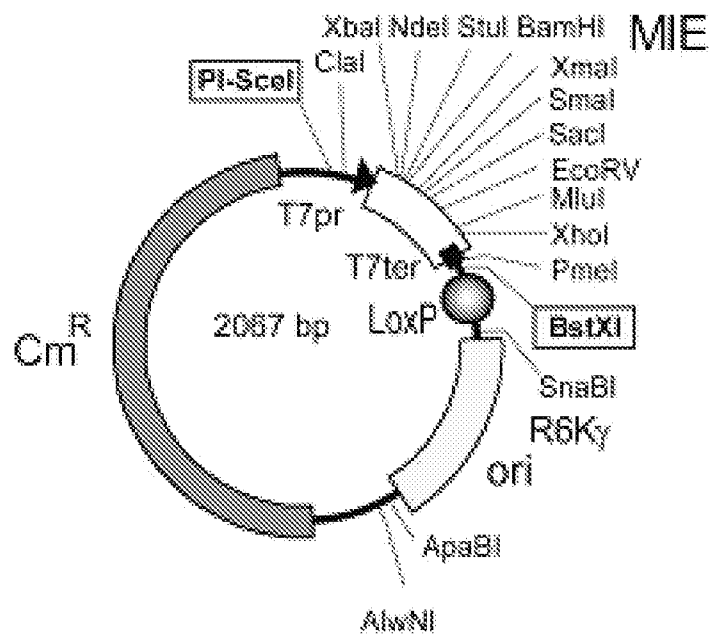


Fig. 16

pDK

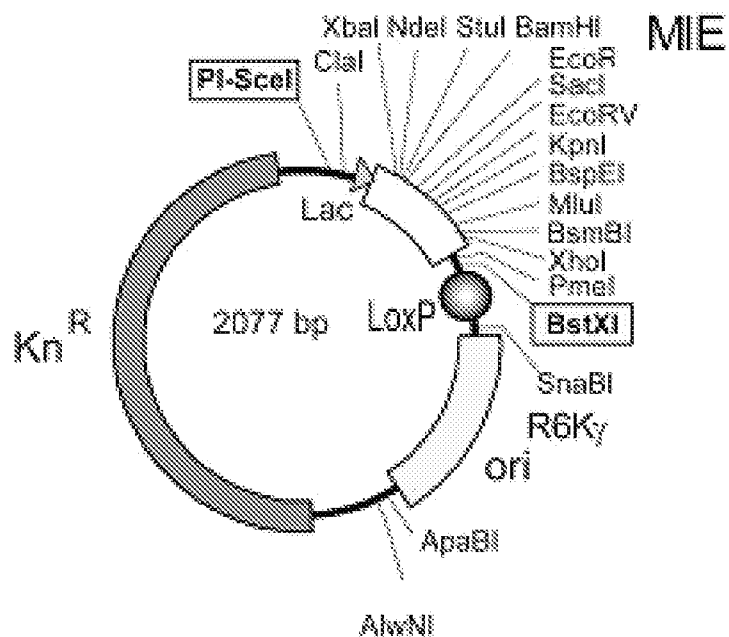


Fig. 17

16/45

pDS

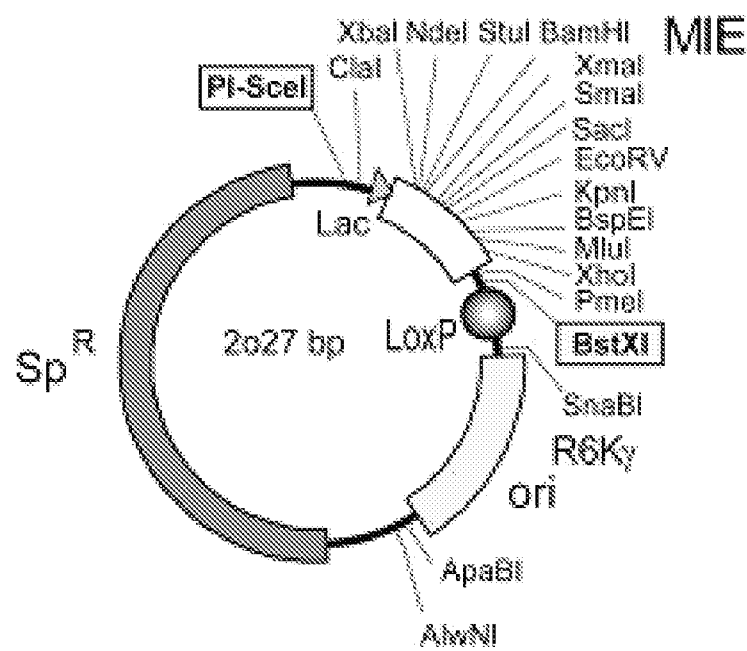


Fig. 18

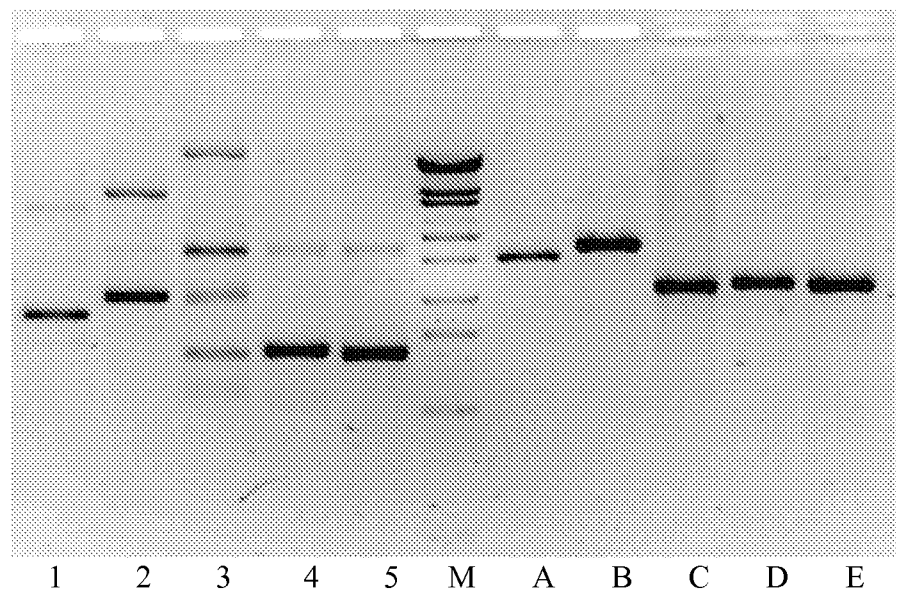


Fig. 19

17/45

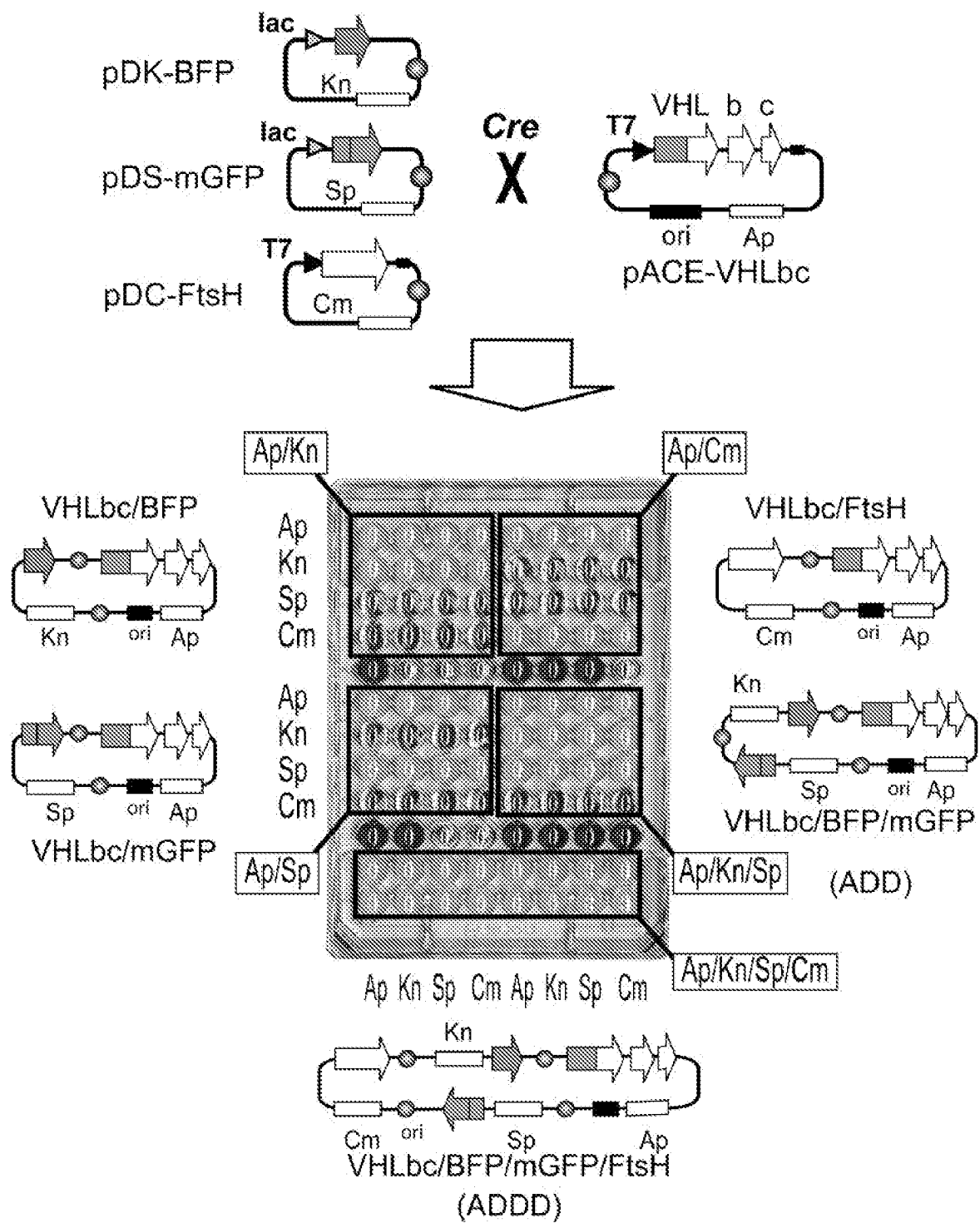


Fig. 20

18/45

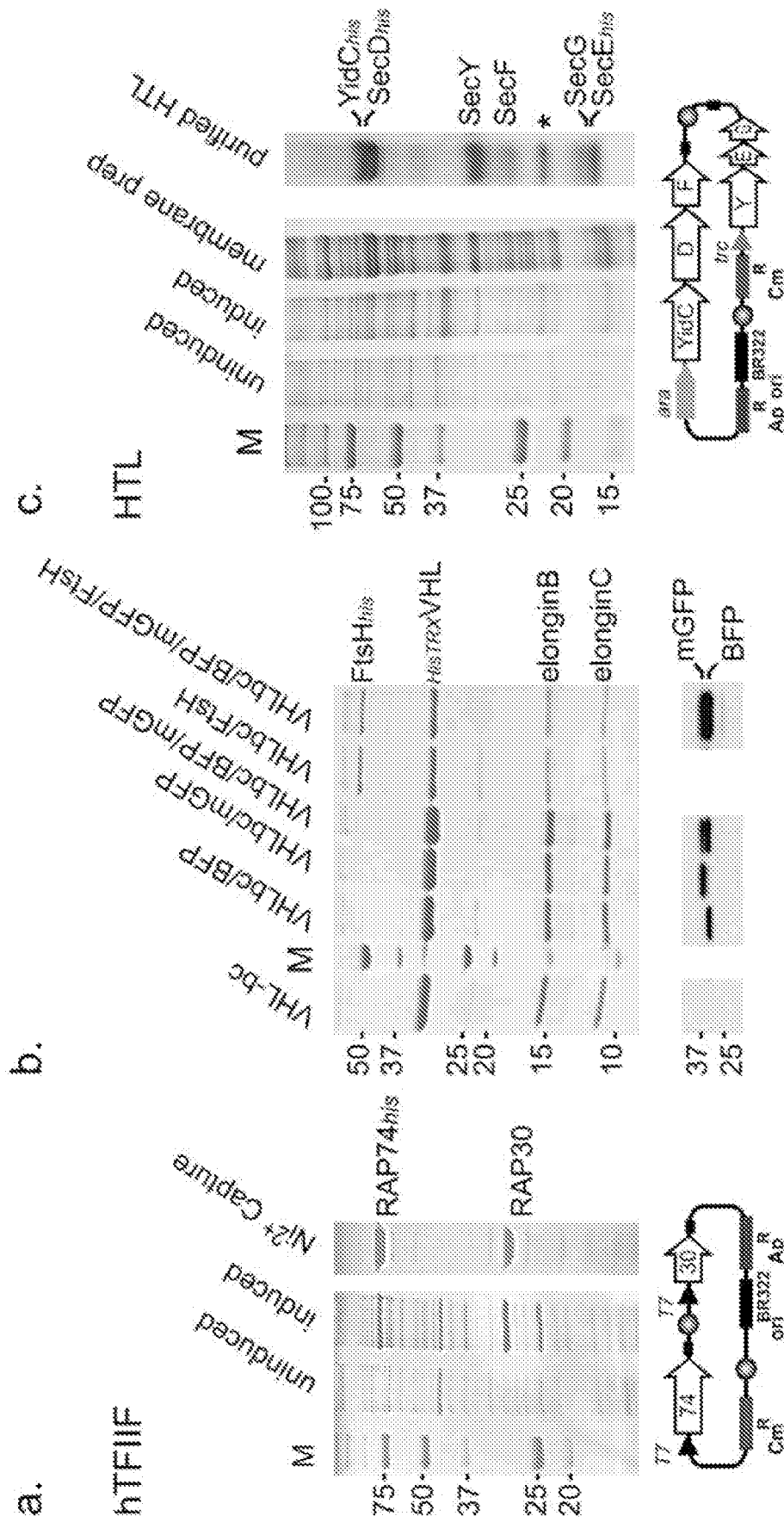


Fig. 21

19/45

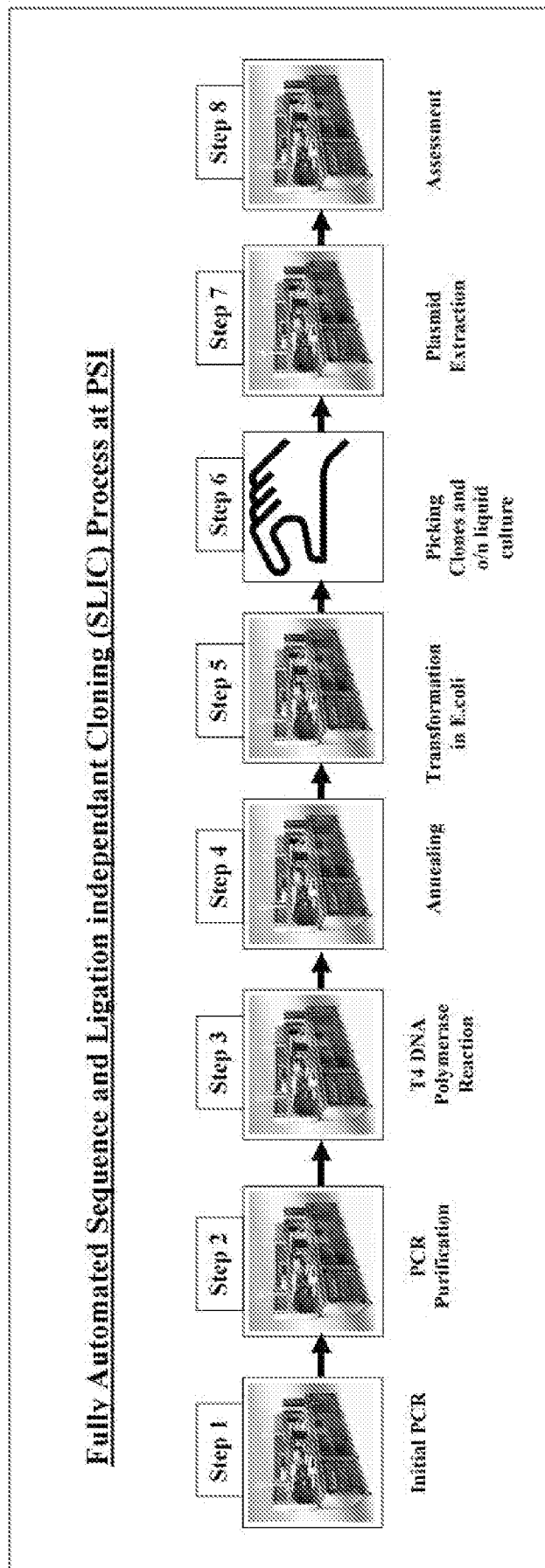


Fig. 22

20/45

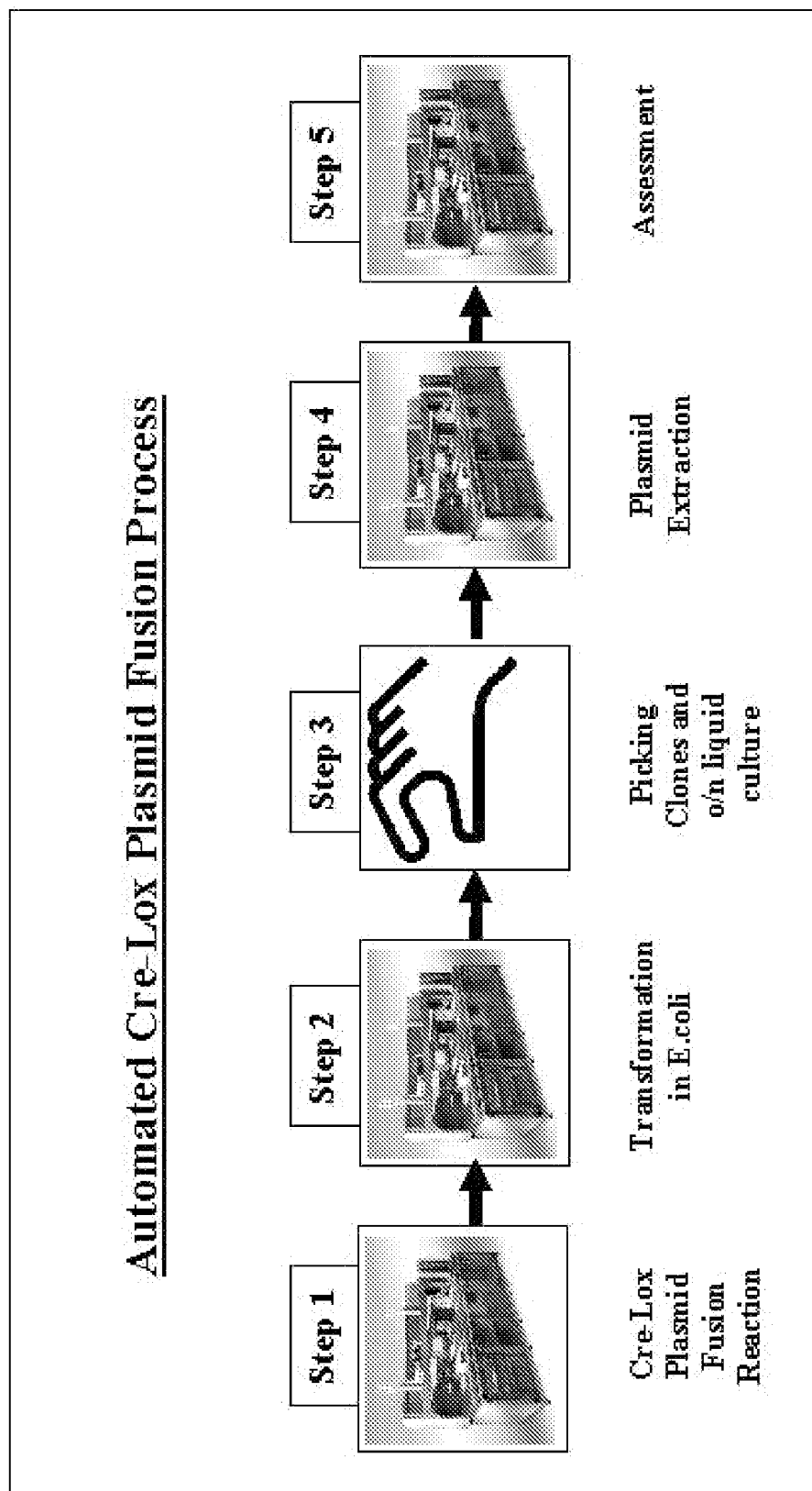


Fig. 23

21/45

pIDC

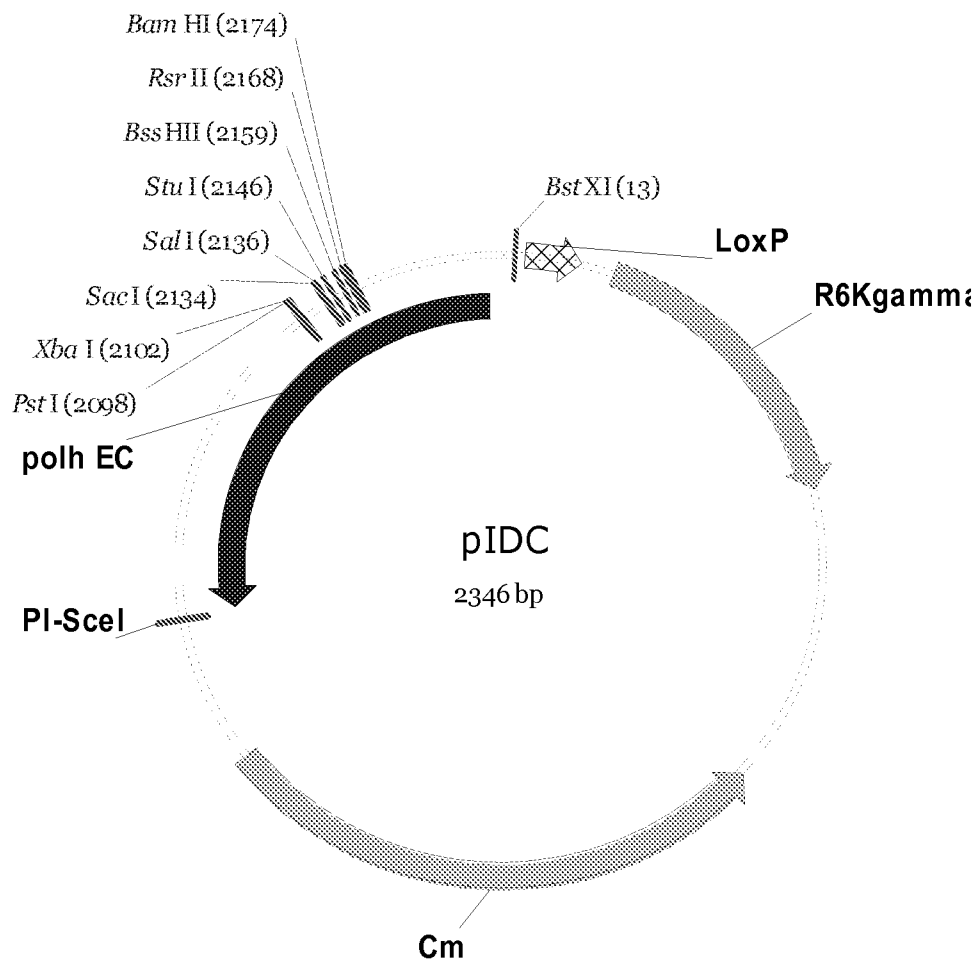


Fig. 24

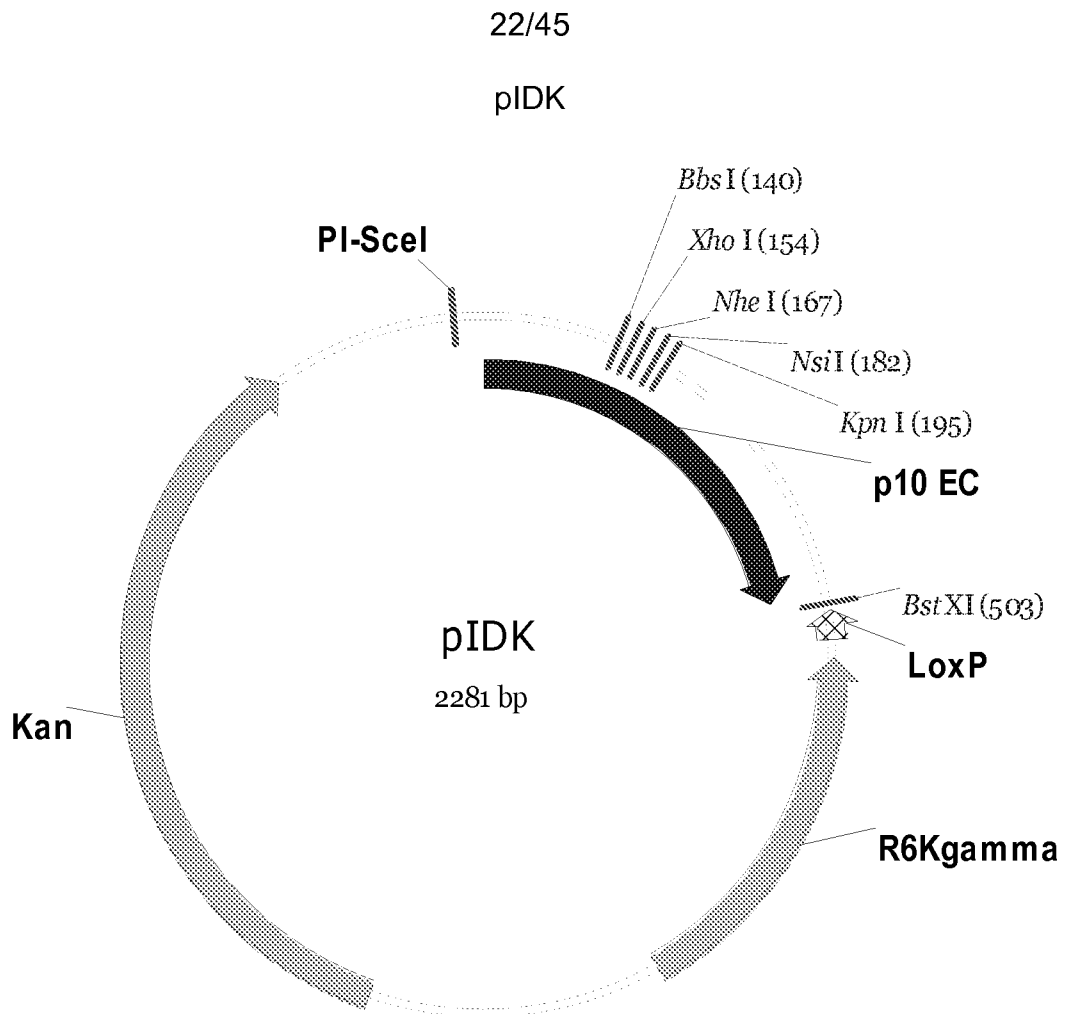


Fig. 25

23/45

pIDS

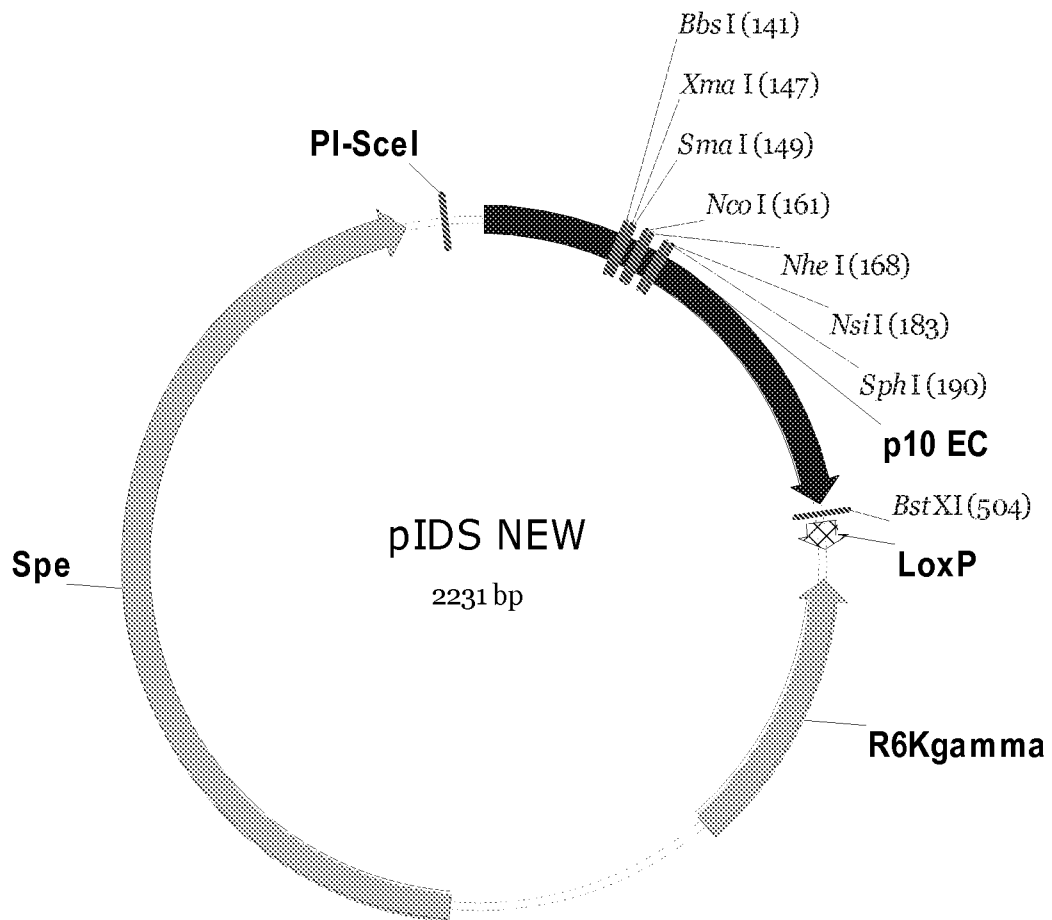


Fig. 26

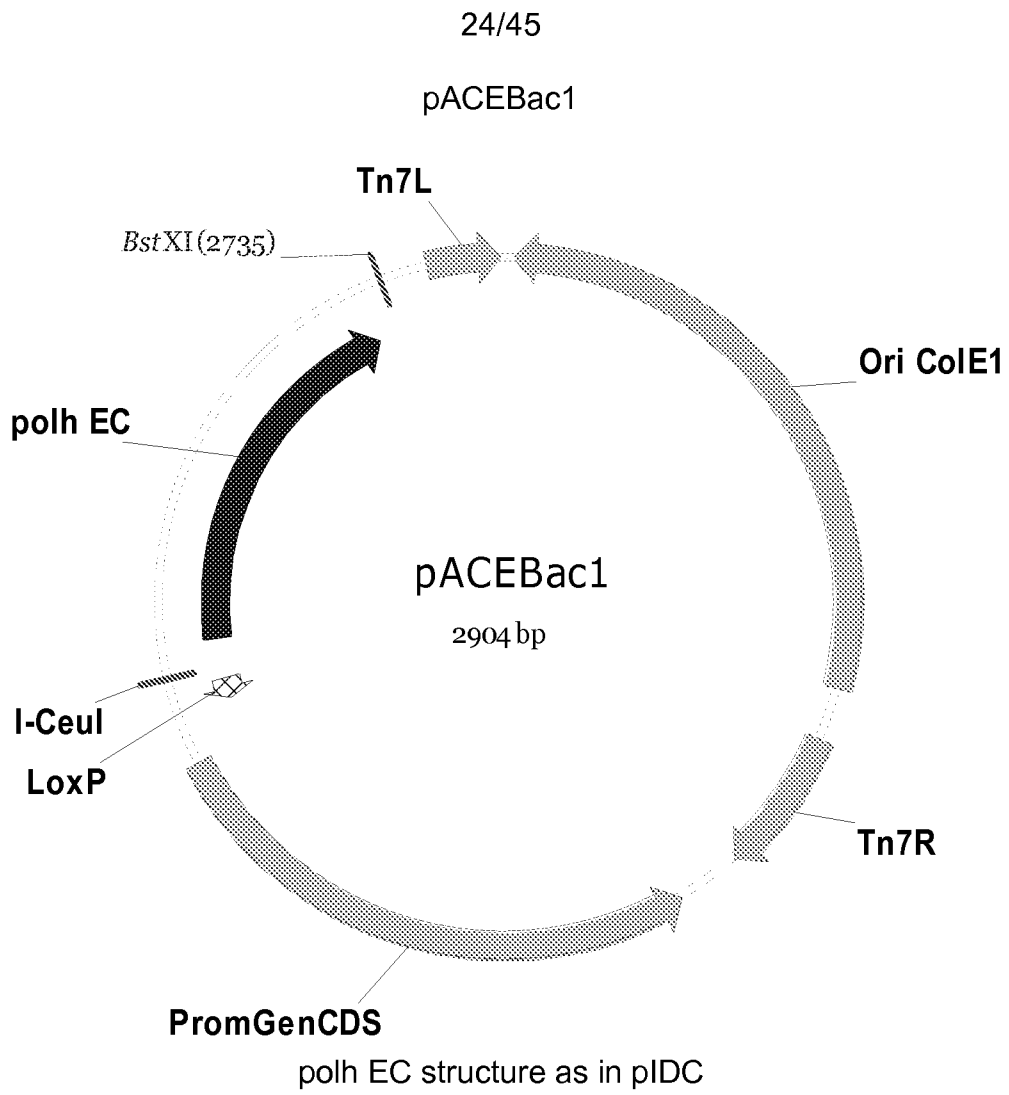


Fig. 27

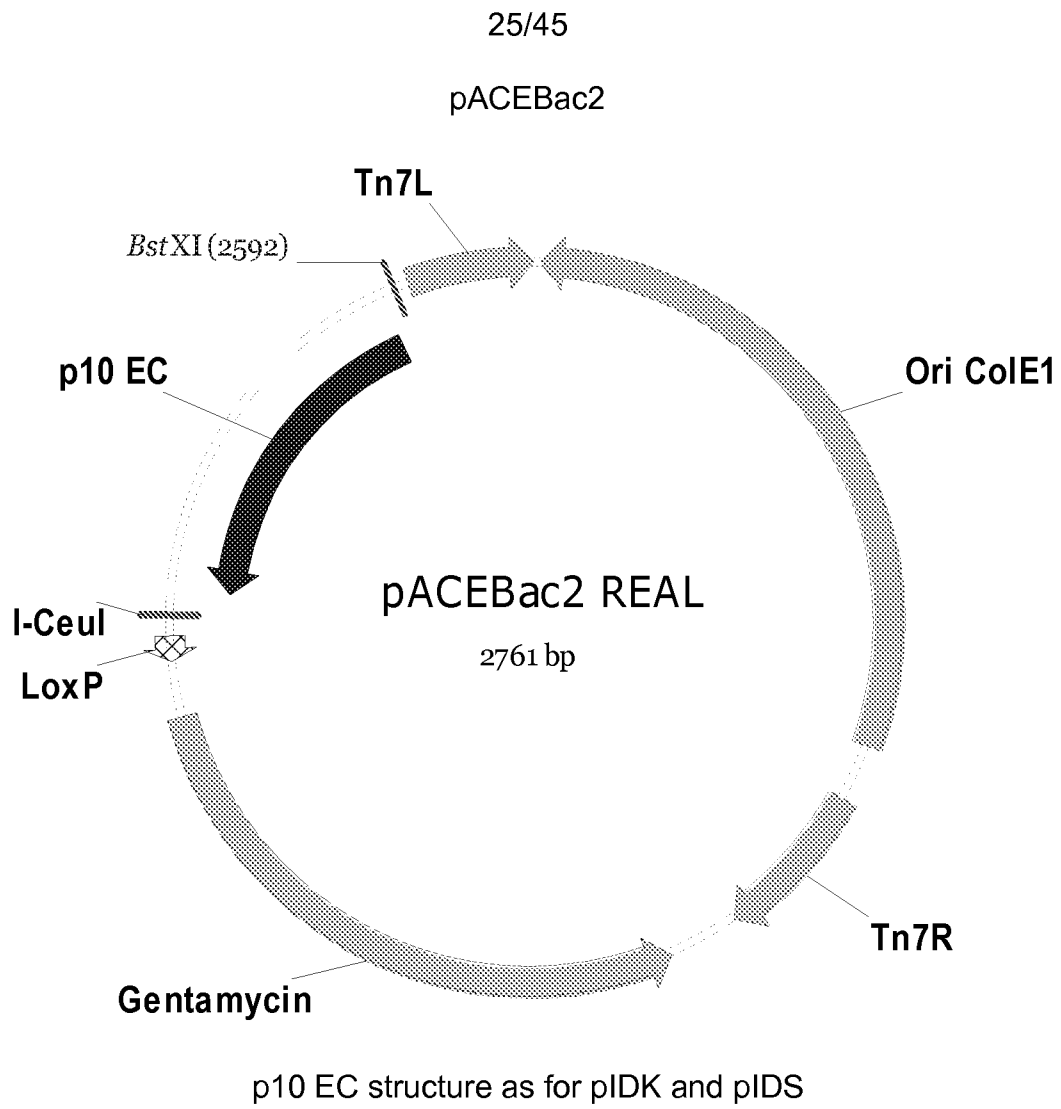


Fig. 28

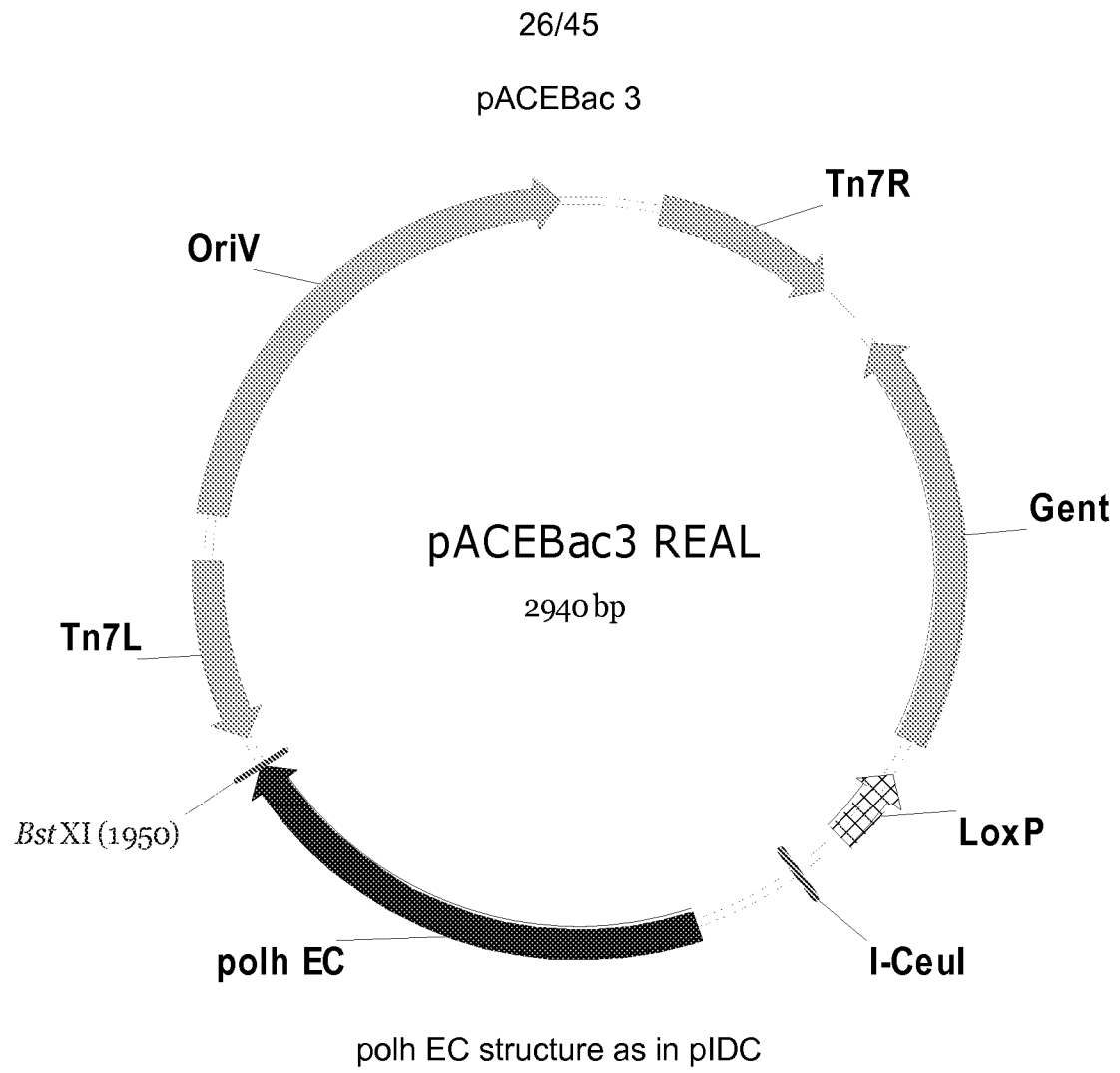


Fig. 29

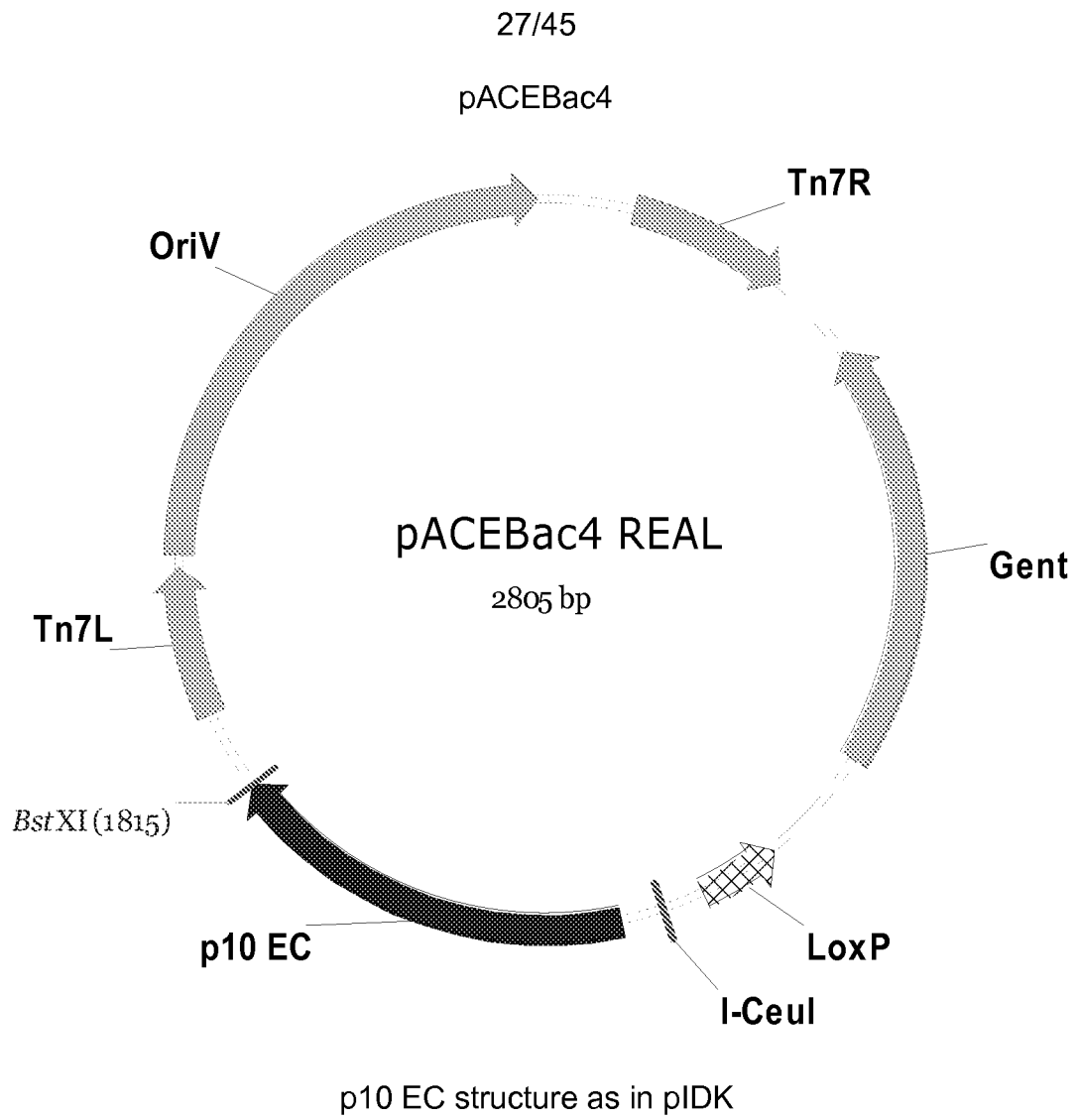


Fig. 30

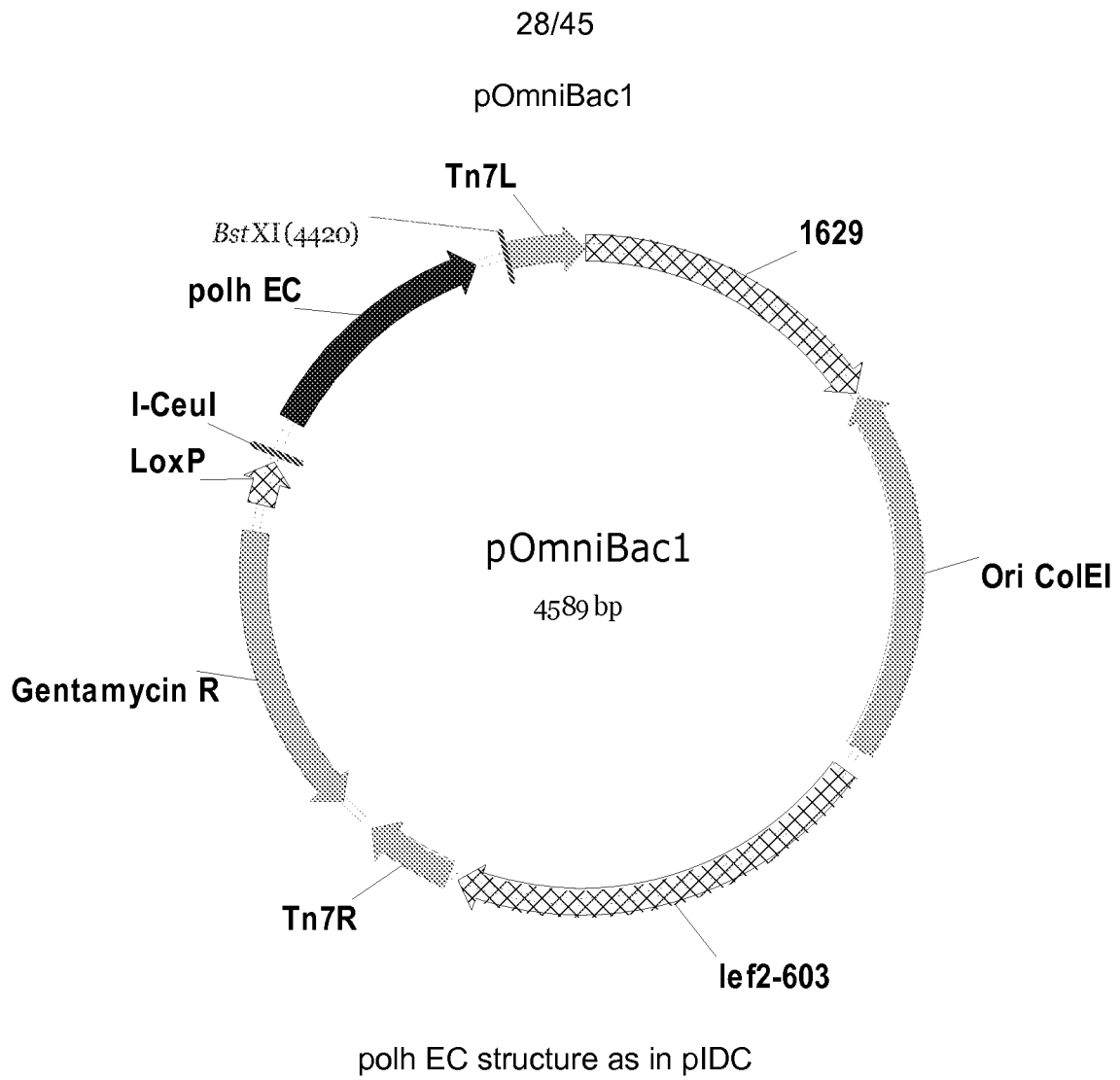


Fig. 31

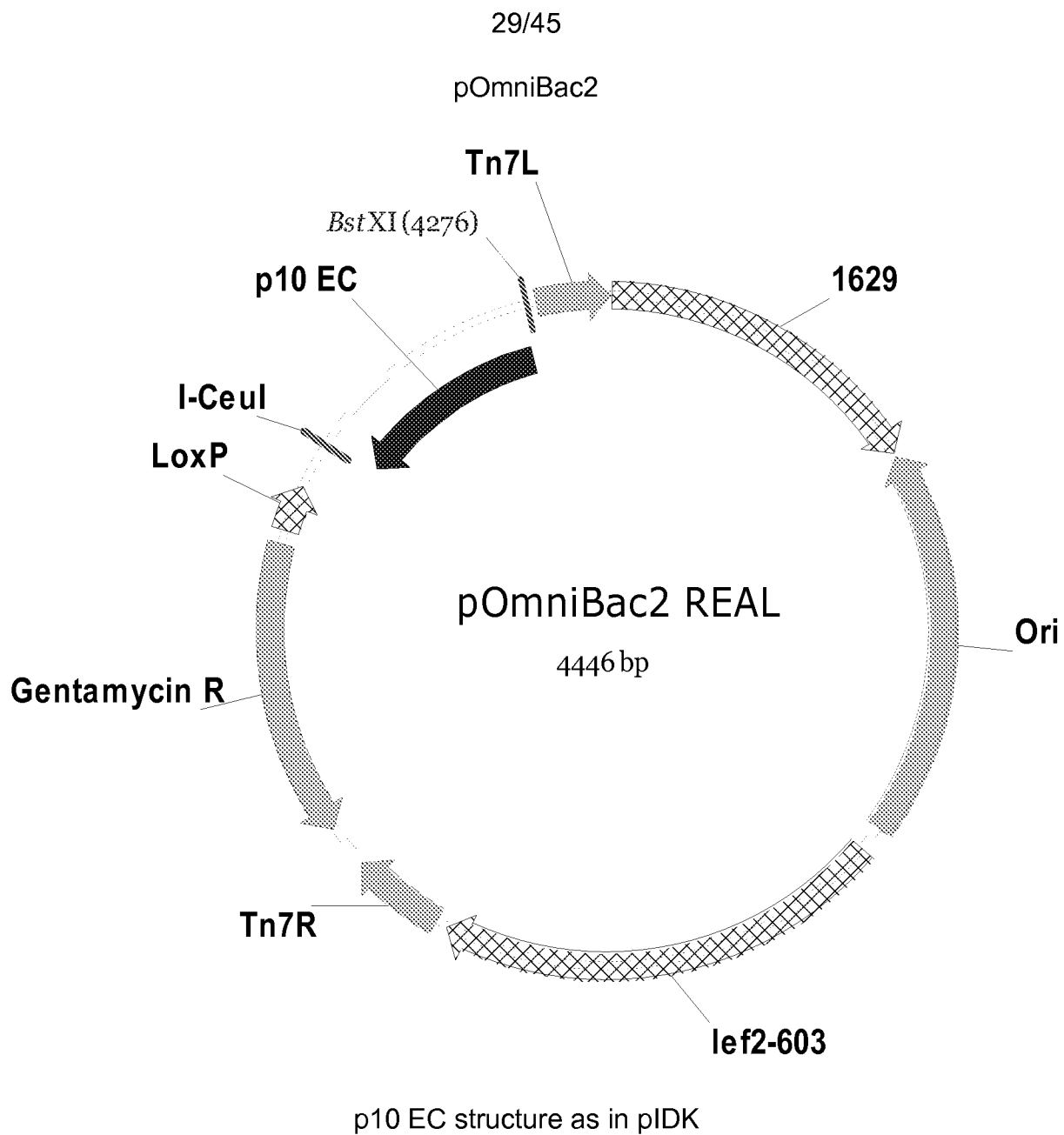


Fig. 32

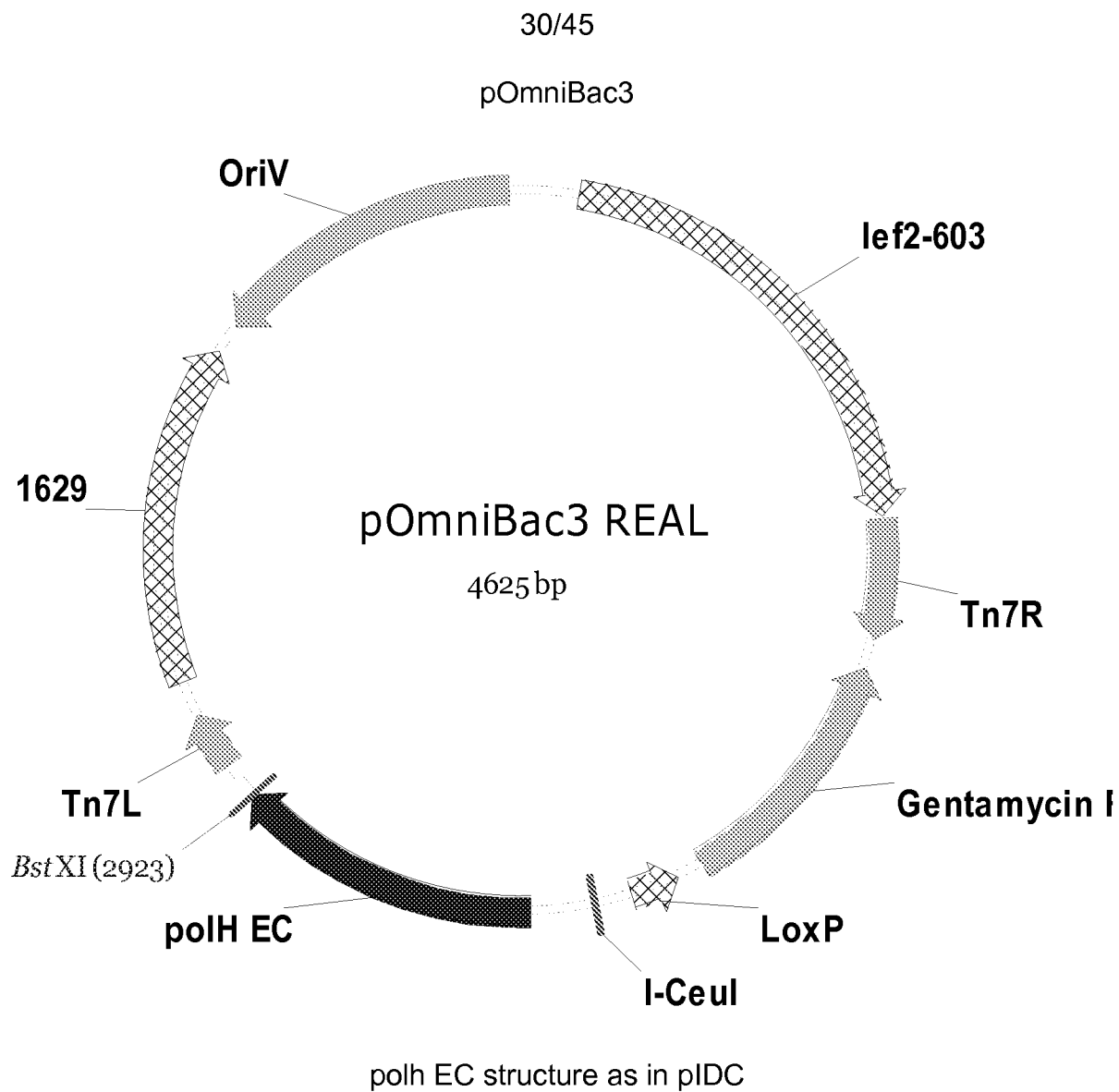
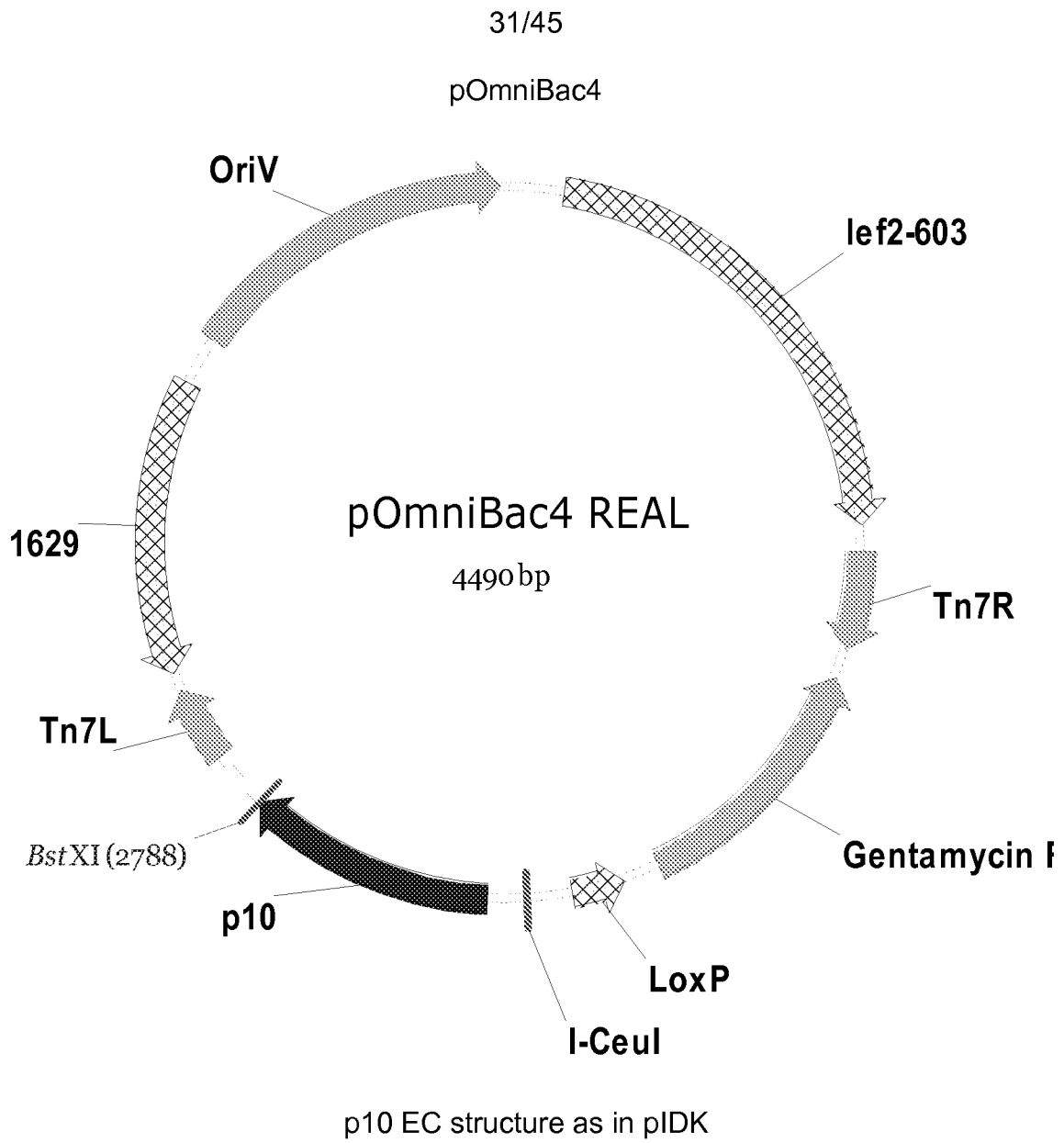


Fig. 33

**Fig. 34**

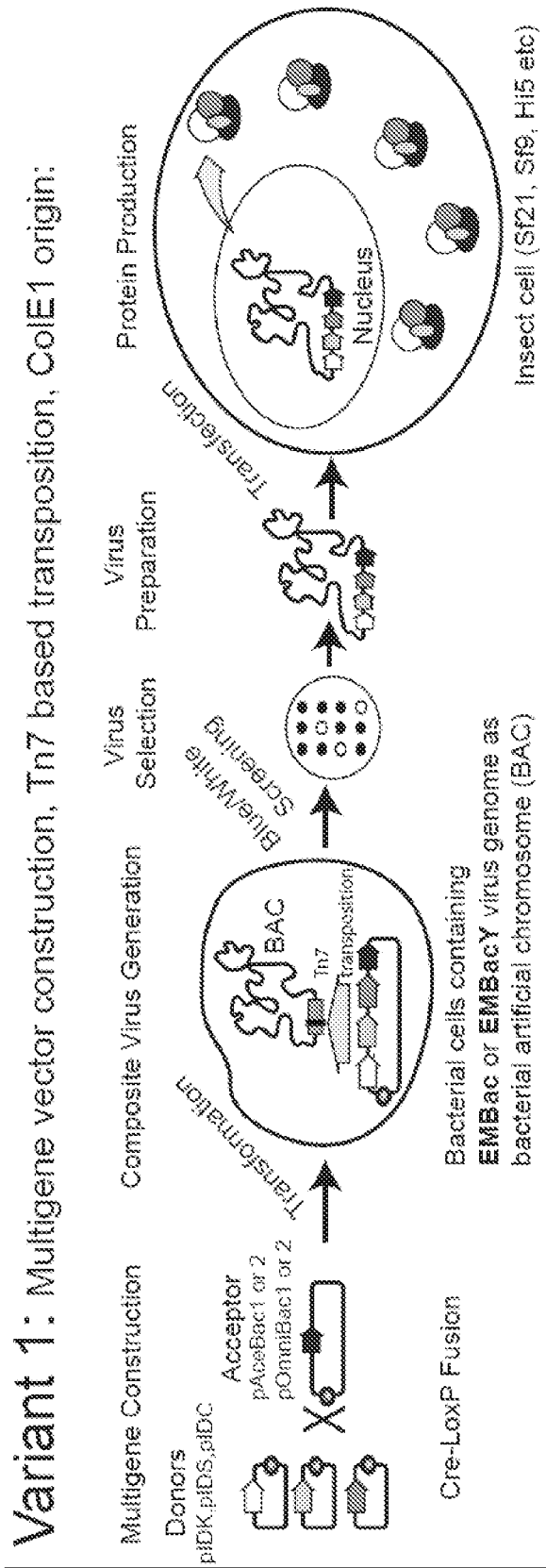


Fig. 35

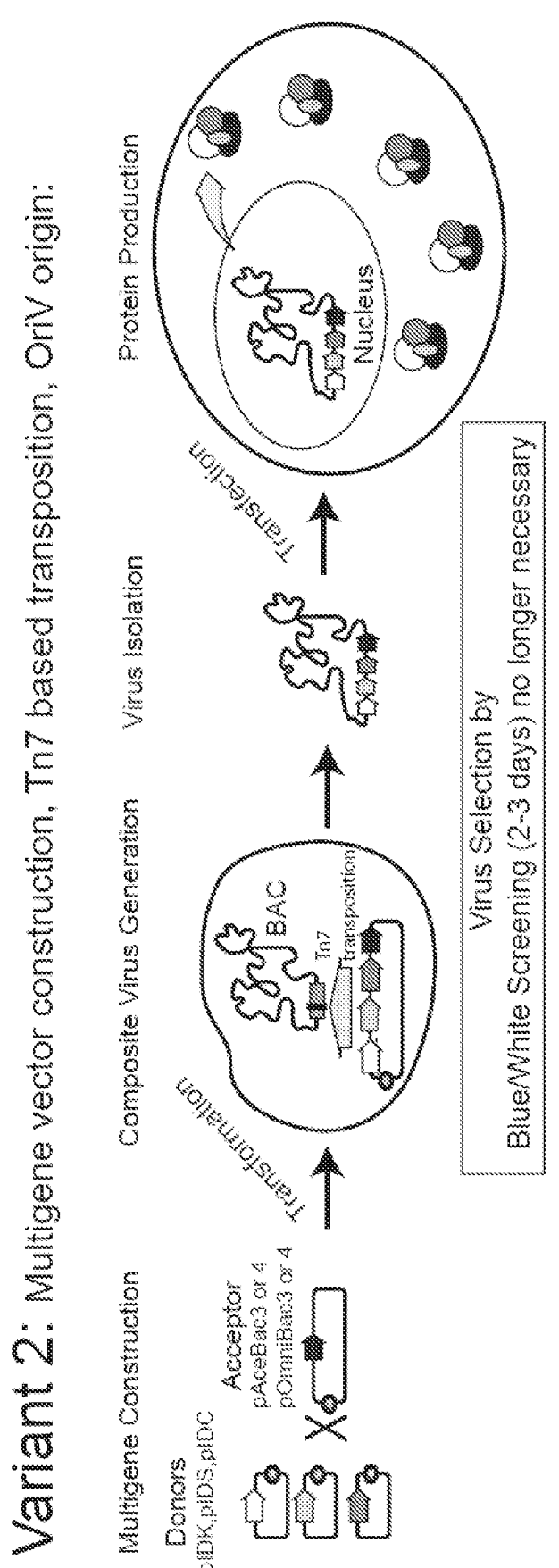


Fig. 36

Variant 3: Multigene vector construction, Direct transposition:

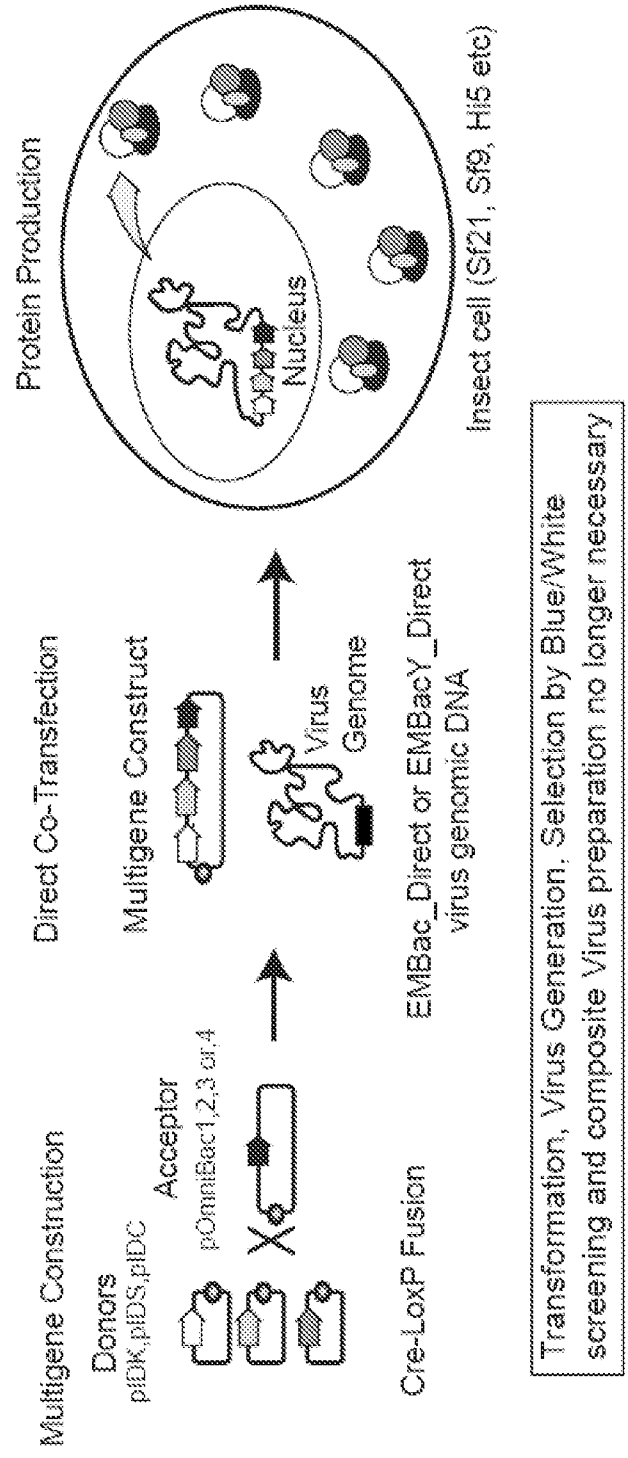


Fig. 37

35/45

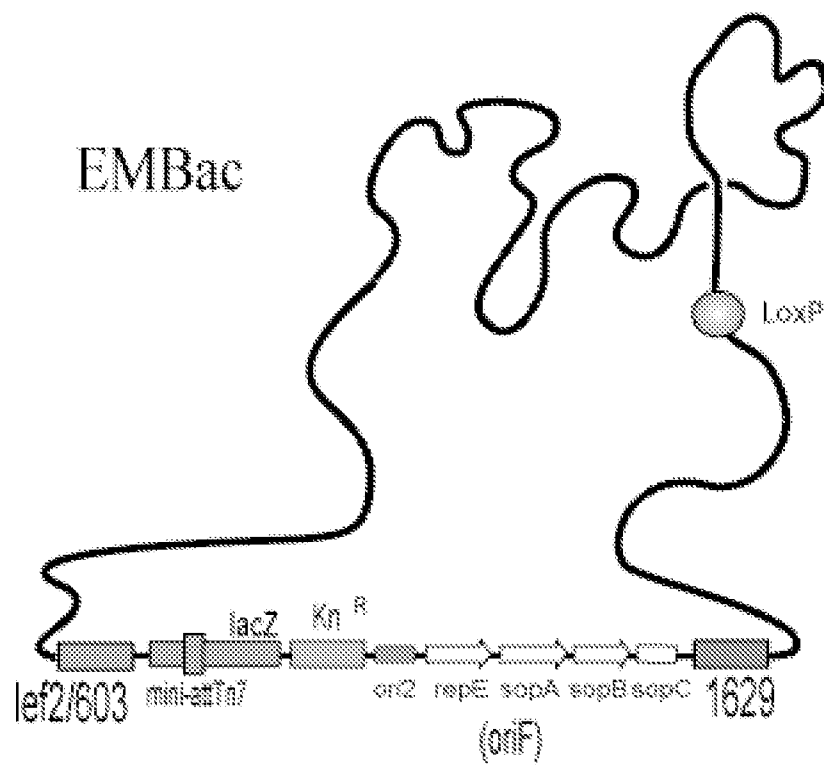


Fig. 38

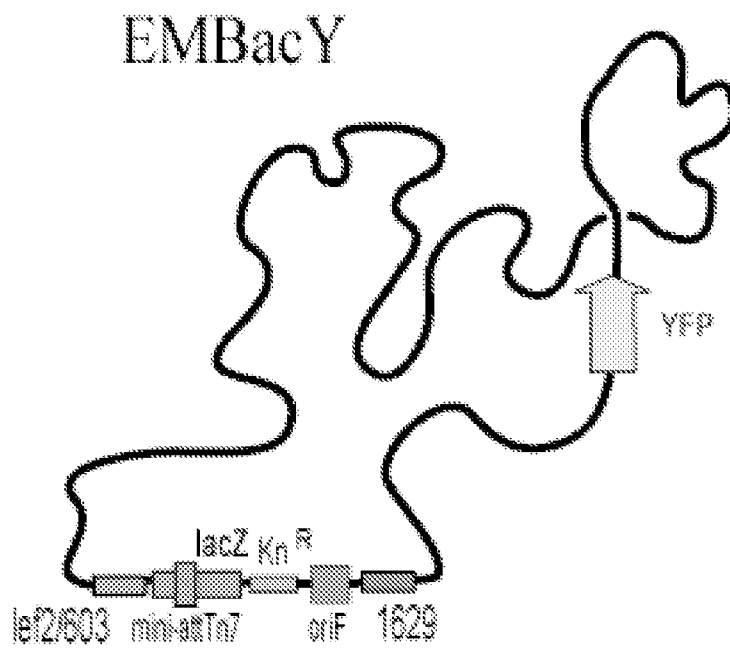


Fig. 39

36/45

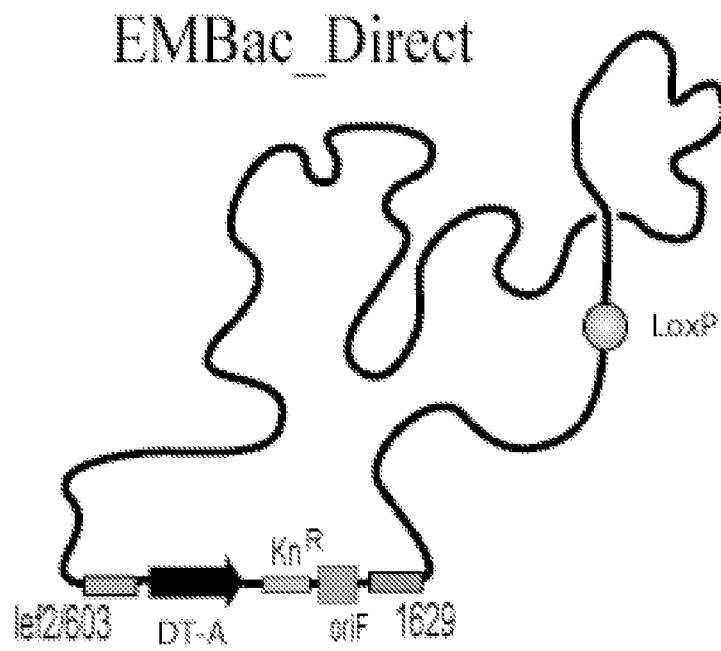


Fig. 40

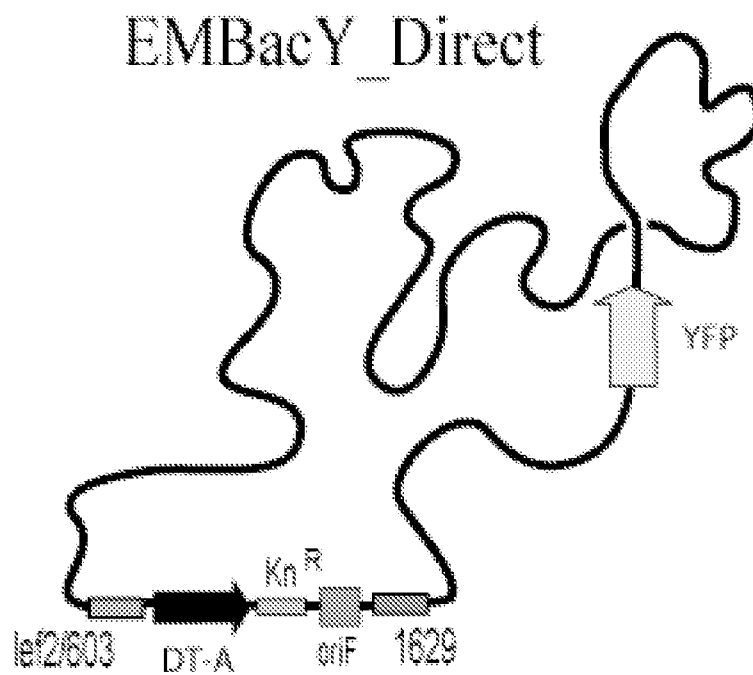


Fig. 41

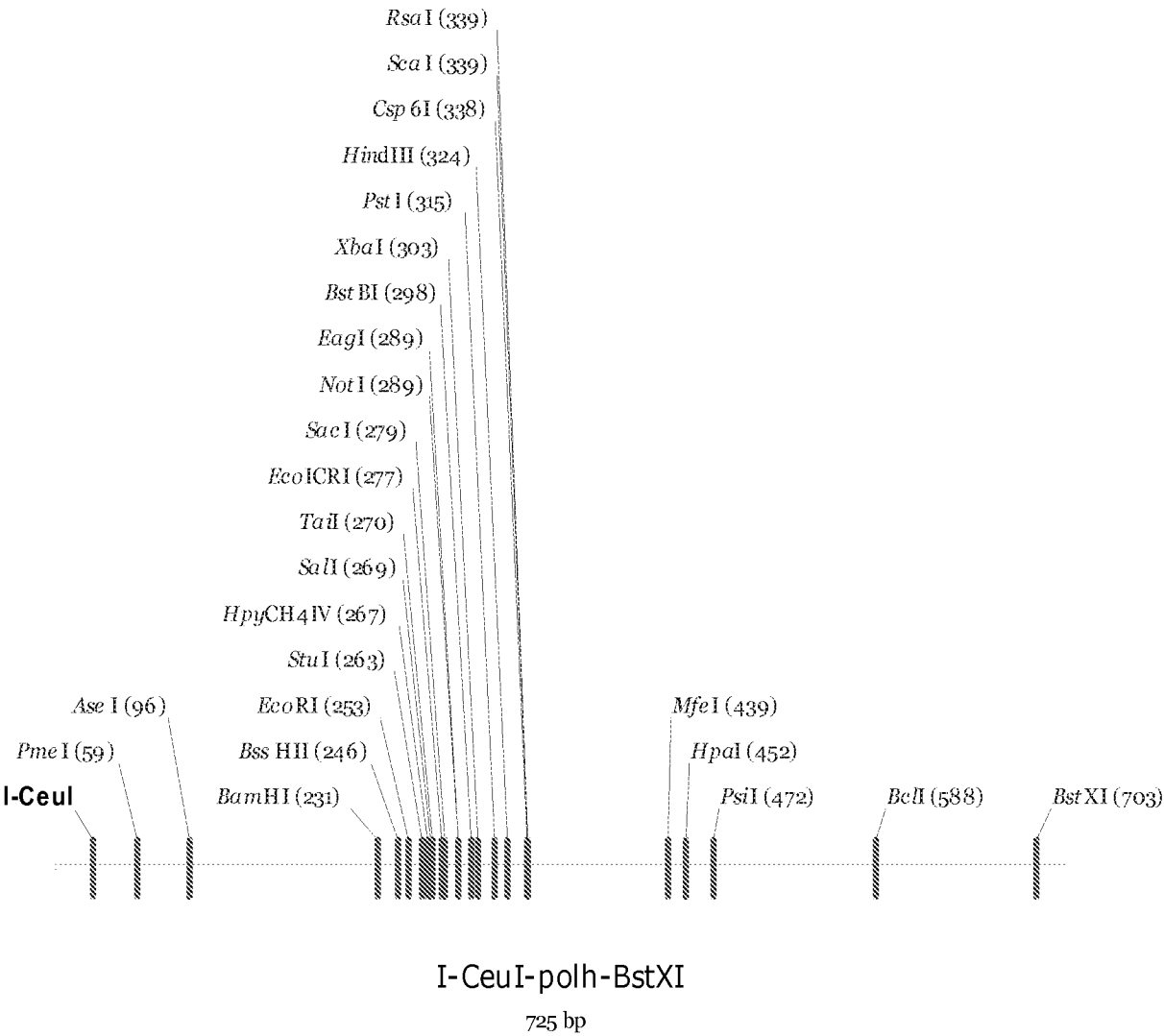


Fig. 42

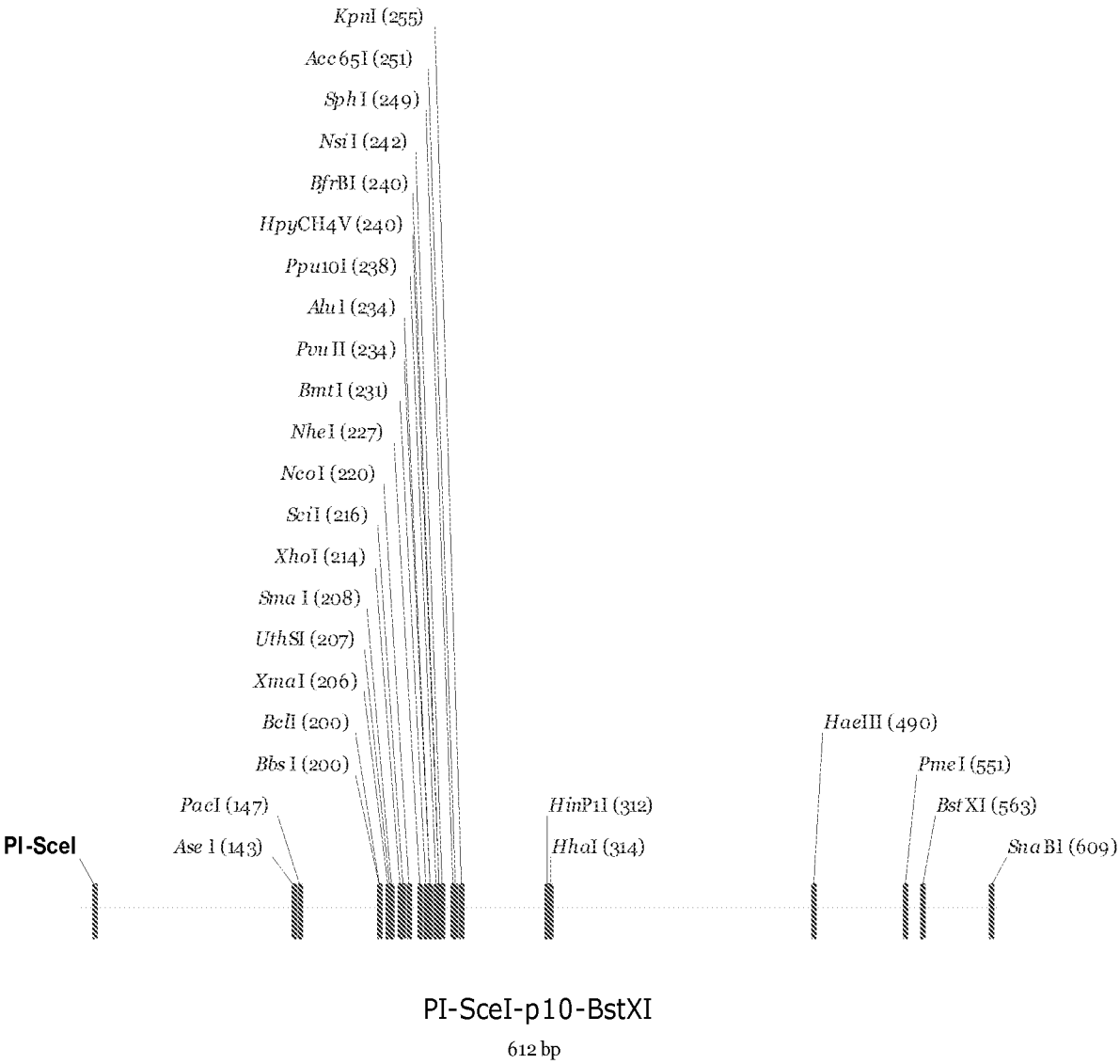


Fig. 43

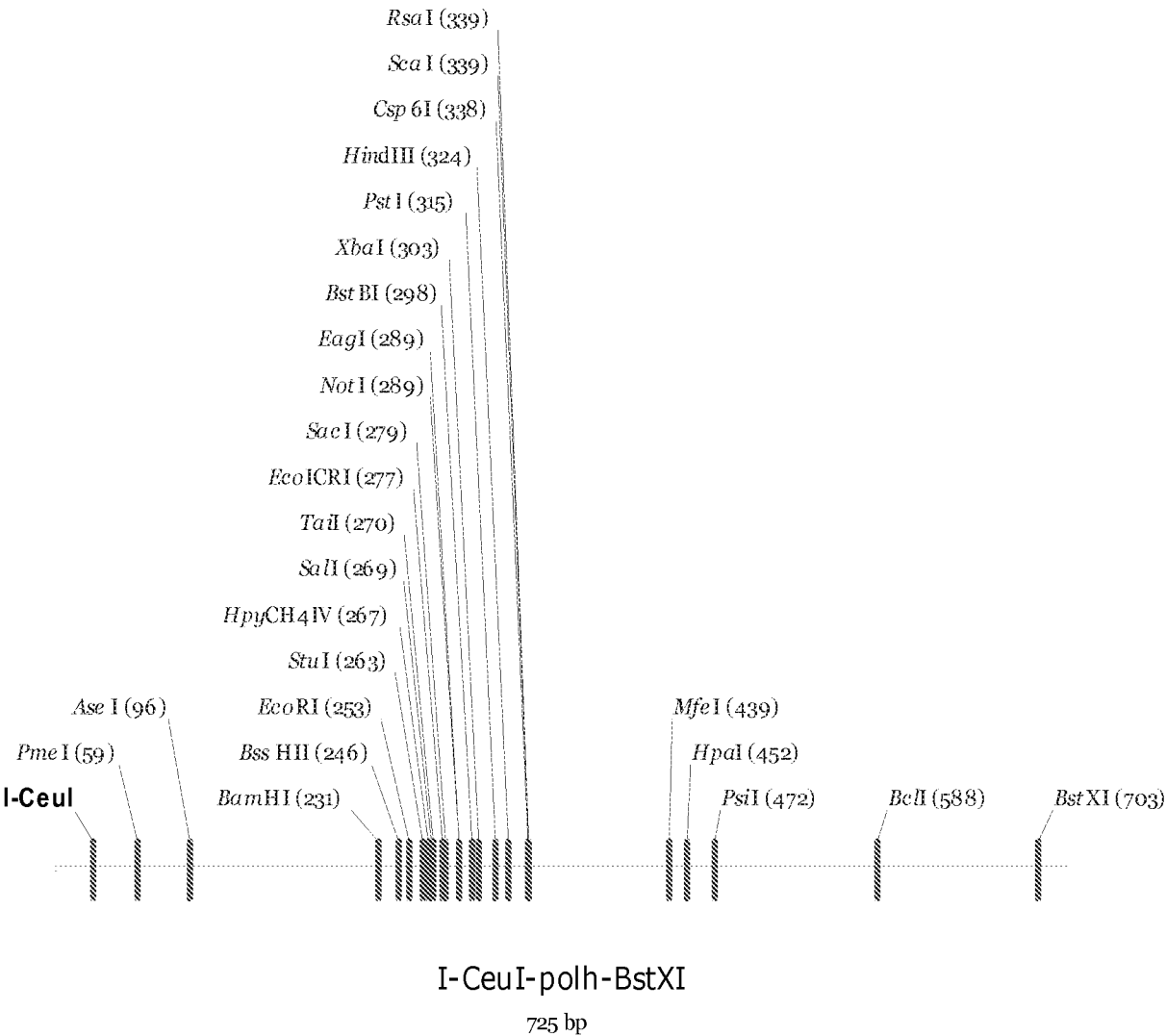


Fig. 44

40/45

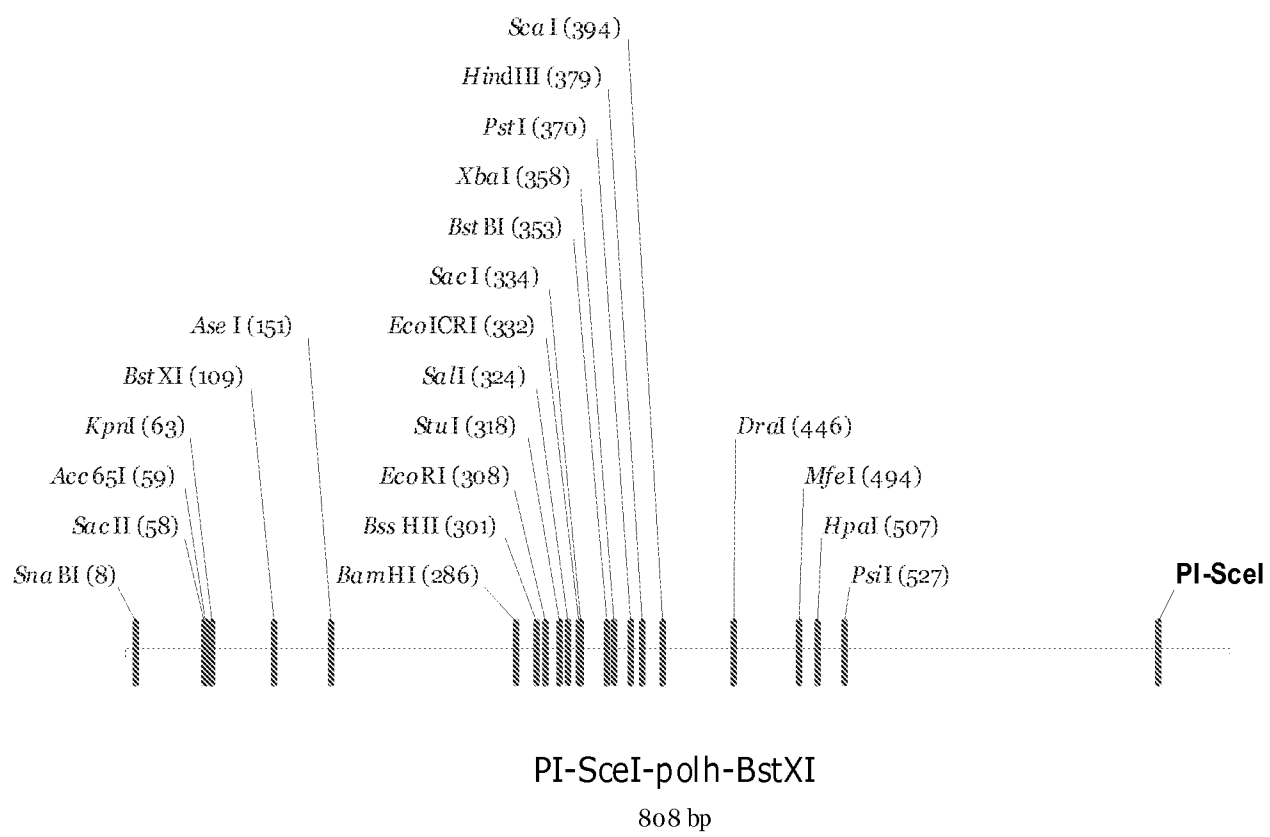


Fig. 45

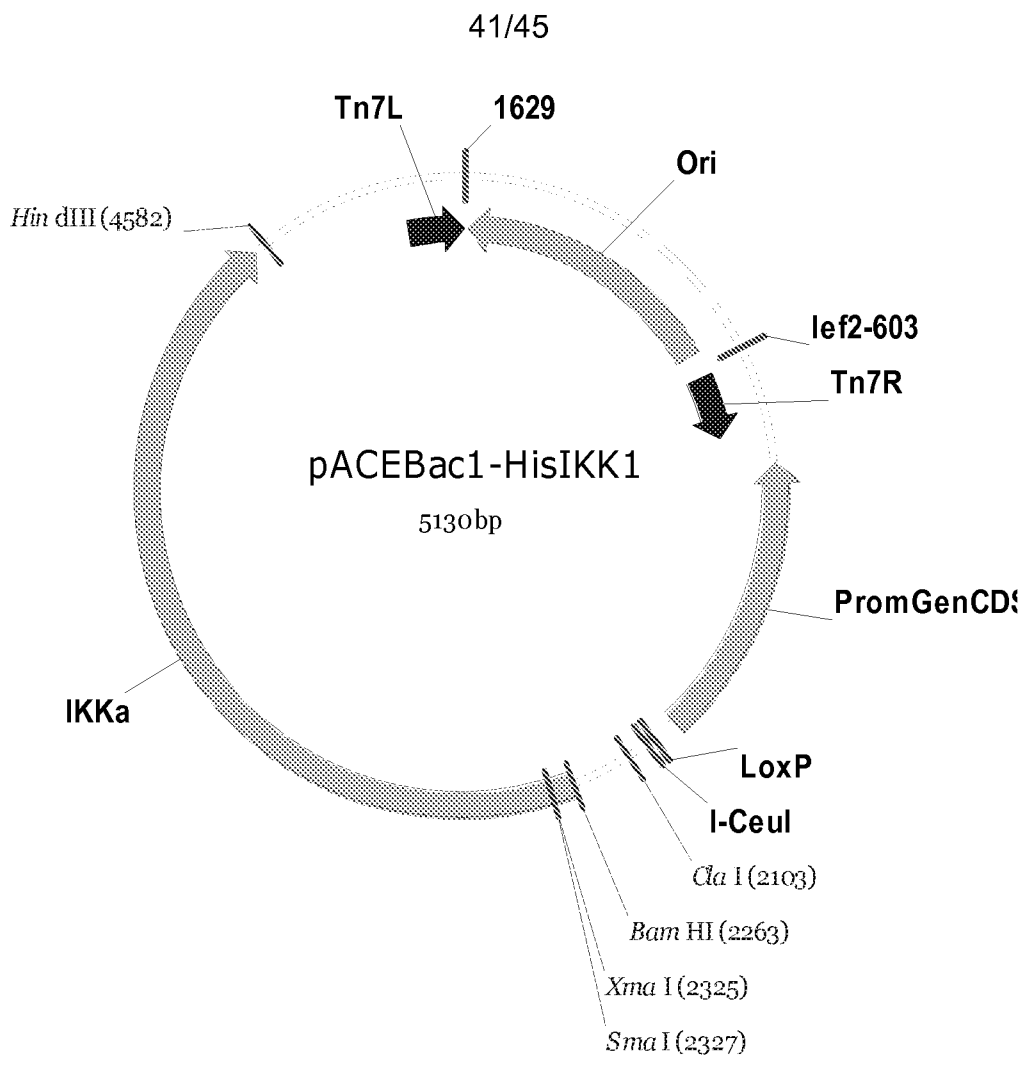


Fig. 46

42/45

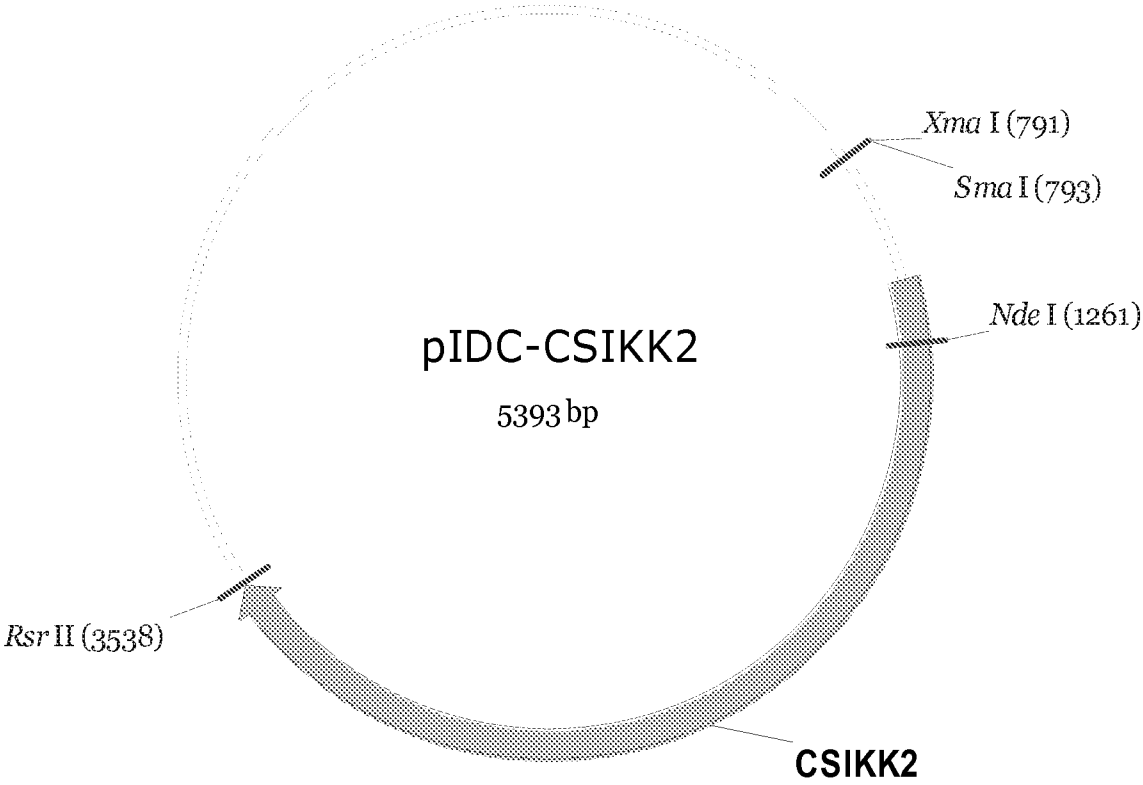


Fig. 47

43/45

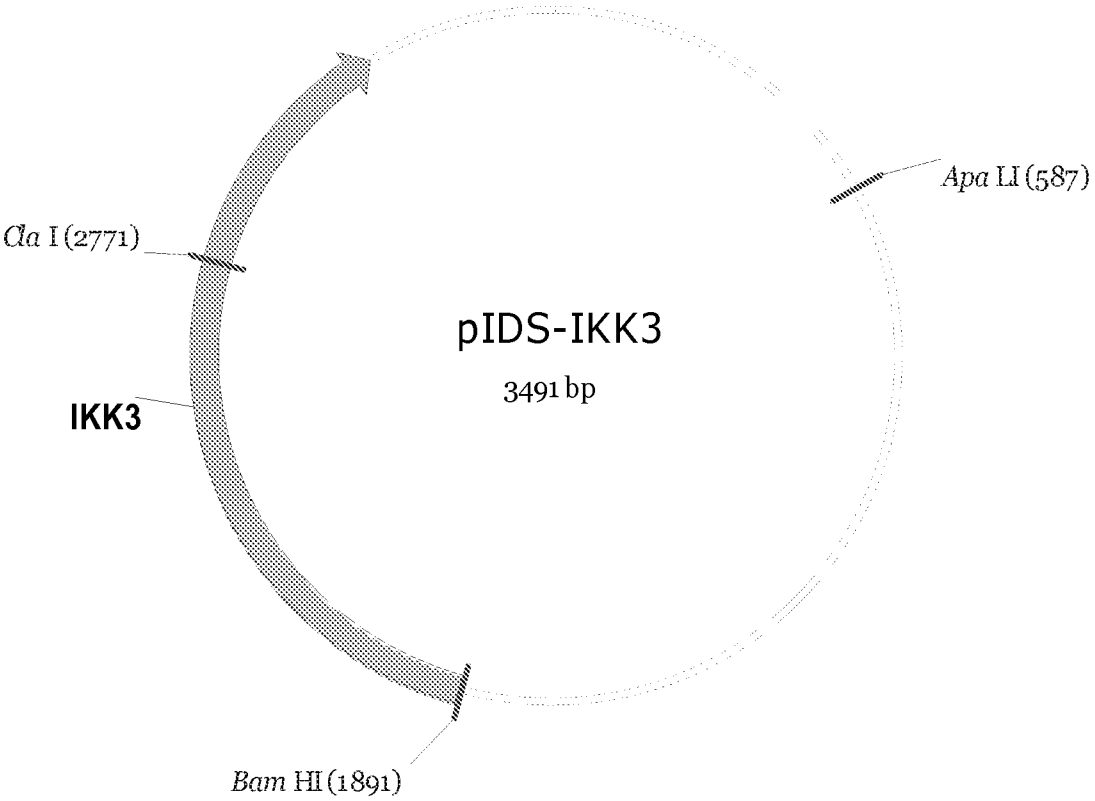


Fig. 48

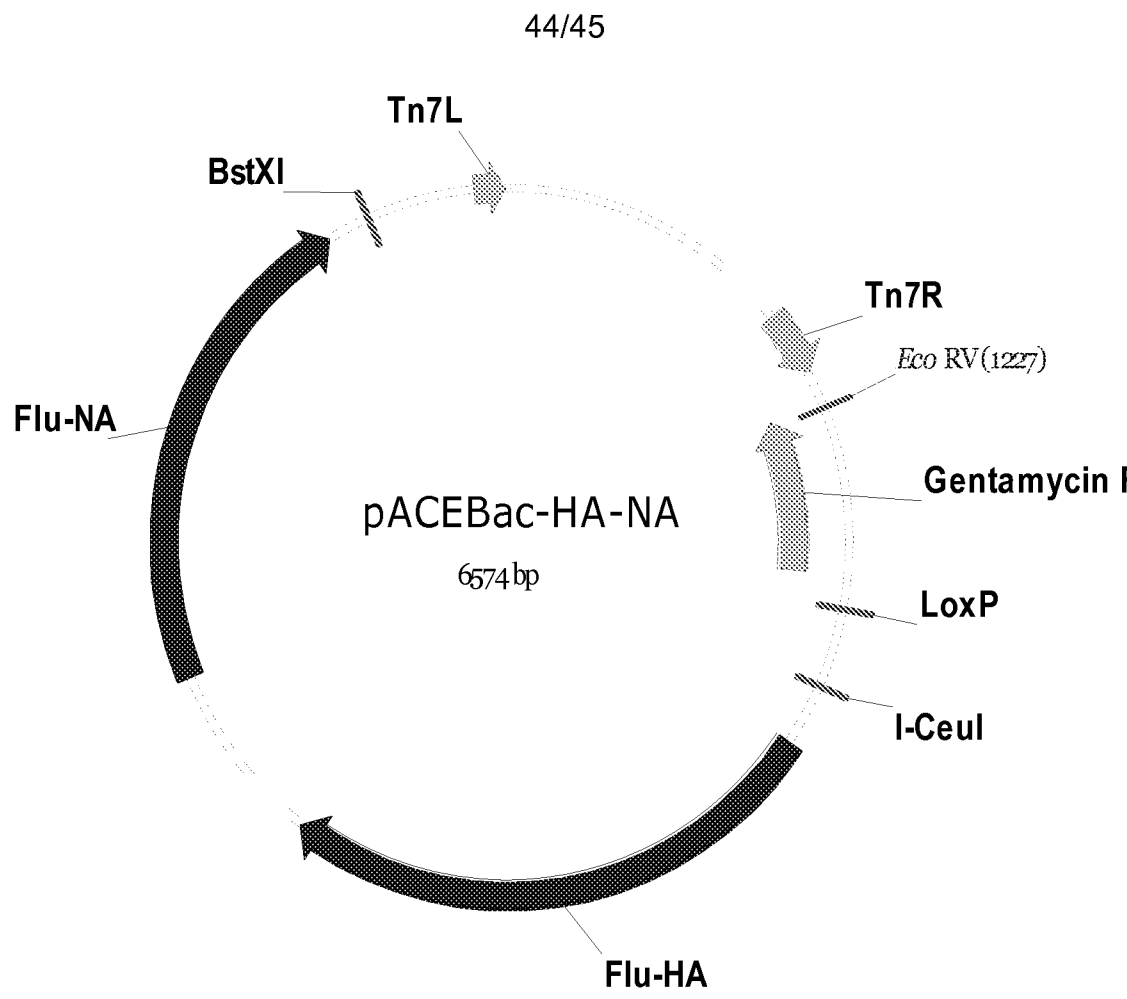


Fig. 49

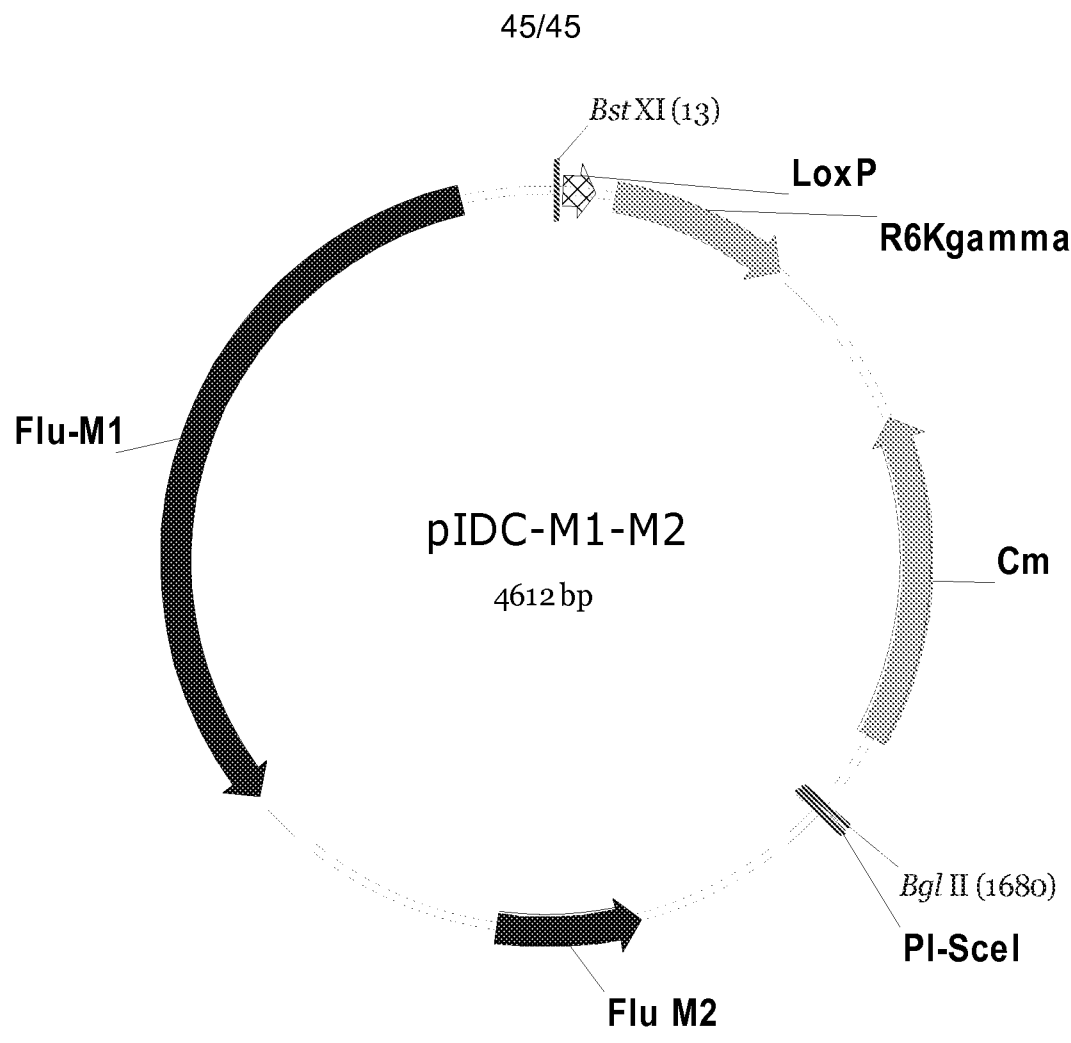


Fig. 50