Title: TOOLS FOR MULTIPROTEIN COMPLEX EXPRESSION IN PICHIA PASTORIS

Abstract: The invention provides an isolated multiple integration element (MIE) comprising, in the specified order: a) a homing endonuclease site (HE); b) a promoter selected from the group consisting of AOX1, FLD1, FMD1, GAP and TEF1; c) a BstBI restriction site; d) a sequence encoding a secretion signal; e) a NcoI restriction site; f) a selection marker gene; g) a Pael restriction site; h) a DNA sequence encoding a detection and/or purification polypeptide; i) a terminator selected from the group consisting of AOX1, FLD1, FMD1, GAP and TEF1; and j) a BstXI restriction site, wherein the HE and BstXI sites are selected such that HE and BstXI result in compatible cohesive ends when cut by the homing endonuclease and the BstXI restriction enzyme, respectively, and the ligation product of HE and BstXI cohesive ends can neither be cleaved by the homing endonuclease nor the BstXI restriction enzyme.
Tools for multiprotein complex expression in *Pichia pastoris*

The present invention relates to the field of molecular biology, particularly to protein expression systems. The invention provides tools for multiprotein complex expression in *Pichia pastoris*.

BACKGROUND ART

The recombinant production of multiprotein complexes is a time-consuming, labor-intensive and costly endeavor that is nonetheless crucial for modern biology (from biotechnology to systems biology) owing to the fact that nearly all biological functions are performed by protein complexes with several subunits.

A limited number of systems have been introduced in recent years for simultaneous expression of several genes in eukaryotic and prokaryotic hosts (see, e.g. Fitzgerald et al., *Nat. Methods* 2006). In spite of considerable improvements of eukaryotic expression systems, in particular the baculovirus/insect cell expression (Fitzgerald et al. (2006), *infra*), *E. coli* still remains to date the dominant workhorse 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 or as polycistrons constituting several genes under the control of the same promoter. 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.

Berger *et al* have recently developed a new method for producing multiple protein complexes (*Nat Methods* 2009, WO2010100278, *infra*) in *E. coli* as well as in insect cells through baculovirus. The method makes use of donor and acceptor vectors, each containing a multiple integration element (MIE) flanked by a homing endonuclease recognition site (acceptor vectors: I-Ceul; donor vectors: PI-SceI) and a complementary BstXI site. Homing endonucleases are rare cutters with long (20–30 bp) recognition sequences that are unique even in very large DNAs. Digestion by homing endonuclease
gives rise to a specific overhang that matches a corresponding BstXI site. This can be used to generate multiple expression cassettes iteratively by insertion of expression cassettes liberated by homing endonuclease and BstXI digestion into constructs linearized at the homing endonuclease site.

5 Donor and acceptor vectors disclosed by Berger et al (supra) additionally comprise a loxP imperfect inverted repeat which allows for Cre recombinase-mediated fusion of said vectors. Interestingly, donor and acceptor vectors carrying multiple expression cassettes generated following the procedure explained above may be fused in the presence of Cre recombinase to rapidly create plasmids containing a large number of genes of interest. The authors demonstrate that this system provides for efficient expression of multiprotein complexes in *E. coli* and insect hosts.

15 However, the vectors described above are not suitable for expression in other hosts of interest. In particular, these vectors may not be used for multiprotein expression in yeasts. The use of yeasts, in particular, *Pichia pastoris*, as a cellular host for the expression of recombinant proteins has become increasingly popular. Being a eukaryote, *P. pastoris* has the advantage, when compared to *E. coli*, that it provides the potential for producing soluble, correctly folded eukaryotic recombinant proteins that have undergone post-translational modifications required for functionality. Additionally, *P. pastoris* may rapidly produce recombinant proteins at higher yields and lower costs when compared to other eukaryotic hosts and, besides, it is easier to genetically manipulate and culture than, for example, mammalian cells. *P. pastoris* is thus often the protein expression host of choice for the production of biopharmaceuticals. And yet, appropriate tools for expression of multiprotein complexes in *P. pastoris* are notably lacking.

30 In view of the above, there is a need to provide tools allowing for rapid, facile and versatile assembly of multigene constructs in *P. pastoris* for effective expression of multiprotein complexes.

SUMMARY OF THE INVENTION

35 The inventors have developed new tools that enable facile and efficient expression of multiprotein complexes in *Pichia pastoris*. Said tools are in the
form of nucleic acids that comprise a MIE specially designed to generate very complex multigene constructs for multiprotein expression in *P. pastoris* in a rapid and convenient manner.

5 In a first aspect the invention refers to an isolated multiple integration element (MIE) comprising, in the specified order: a) a homing endonuclease site (HE); b) a promoter sequence selected from the group consisting of *AOX1*, *FLD1*, *FMD1*, *GAP* and *TEF1*, wherein the promoter sequence is devoid of HE, BstBI, NcoI, Pmel and BstXI restriction sites if said restriction sites are present; c) a BstBI restriction site; d) a sequence encoding a secretion signal; e) an NcoI restriction site; f) a selection marker gene; g) a Pmel restriction site; h) a sequence encoding a detection and/or purification polypeptide; i) a terminator sequence selected from the group consisting of *AOX1*, *FLD1*, *FMD1*, *GAP* and *TEF1*, wherein the terminator sequence is devoid of HE, BstBI, NcoI, Pmel and BstXI restriction sites if said restriction sites are present; and j) a BstXI restriction site, wherein the HE and BstXI sites are selected such that HE and BstXI result in compatible cohesive ends when cut by the homing endonuclease and the BstXI restriction enzyme, respectively, and the ligation product of HE and BstXI cohesive ends can neither be cleaved by the homing endonuclease nor the BstXI restriction enzyme.

The MIE of the first aspect of the invention may be represented as disclosed in Figure 1.

25 This design provides several advantageous functionalities:

1- The MIE of the invention has been optimized for ligase-independent cloning strategies such as SLIC/Gibson/InFusion cloning. This provides for: (a) efficient and cost-effective parallel vector construction, (b) sequence-independent insertion, (c) rapid, single-step reaction, (d) precise engineering of constructs, with no additional vector or restriction site-derived amino acids added to the expressed protein, and (e) efficient insertion over a wide insert size and concentration range, all of which, together with the introduction of a selection marker, allows for high-throughput cloning of genes and rapid selection of clones containing the gene of interest for evaluating expression performance.
2- Rapid and facile addition of secretion signals and purification tags by using restriction sites especially reserved for this purpose. The versatility of the tool enabling the user to select between intracellular expression and secretion of the gene of interest is particularly advantageous. The state of the art in P. pastoris vectors is that two separate vectors are necessary, one with the alpha-MF leader sequence, which, when fused to the gene of interest leads to the secretion of the expressed protein product; and another vector, without that sequence, for intracellular expression. High-throughput cloning and efficiency is greatly enhanced by the specific design of the MIE of the invention, because one single vector can be used for both purposes. Hence, the MIE of the invention has a secretion signal sequence with a BstBI restriction site between the promoter and the secretion signal sequence, and an Ncol site right after the secretion signal sequence. Cloning a gene of interest for intracellular expression involves linearizing the vector which comprises the MIE using BstBI and a second restriction site, while cloning it for secretion expression involves digesting the vector with Ncol and a second restriction site.

3- Assembly of multigene constructs by concatamerization of genes using HE/BstXI-digested cassettes containing the gene of interest into HE/BstXI.

In order to arrive at a MIE which provides all the above advantageous functionalities, the inventors have overcome multiple practical difficulties. For example, appropriate sequence elements had to be selected, the most convenient orientation had to be devised for the various elements and adequate restriction sites needed to be selected and placed in the convenient order. Further, due care is required to maintain all sequence elements working together, which brings the difficulty that introduction/modification of one element might result in dysfunction of the system because of defective interaction with other sequence elements. This is illustrated, for instance, by the fact that the promoter and terminator need to be modified to eliminate the HE, BstBI, Ncol, Pmel and BstXI restriction sites, whenever said restriction sites are present in the promoter/terminator sequence. And all this keeping in mind the ultimate goal of the system, which is to allow for cloning of multigene constructs in P. pastoris while preserving the integrity and the functionality expected from the expression system.
The above MIE may be used to obtain an expression cassette by inserting one or more genes of interest (GOI) between the HE and the BstXI sites following any of the available cloning techniques. The resulting expression cassette, flanked by the HE and BstXI restriction sites, may be then used to generate multigene constructs containing multiple expression cassettes by any of the methods explained below. Thus, aspect of the invention refers to a method for obtaining an expression cassette that comprises inserting one or more GOI into a MIE as defined above and to the expression cassette directly obtained by said method. This aspect can also be formulated as an expression cassette derived from the MIE as defined above comprising one or more GOI.

The MIE or the expression cassette of the invention may be conveniently contained in a plasmid or vector. Thus, a third aspect of the invention refers to a plasmid or vector comprising the MIE according to the first aspect of the invention or the expression cassette according to the second aspect of the invention. The invention also provides a host cell containing the expression cassette or the vector according to the first and second aspects of the invention.

Another aspect of the invention relates to a kit for cloning and/or expression of multiprotein complexes containing at least one vector as defined above.

The present invention also relates to a method for producing a vector containing multiple expression cassettes comprising the steps of: (a) providing a first vector comprising an expression cassette according to the present invention; (b) providing a second vector comprising an expression cassette according to the present invention; (c) cleaving the first vector with a homing endonuclease specific for site HE and with BstXI yielding a first expression cassette flanked by the cleaved HE and BstXI sites; (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 expression cassettes.

The present invention is further directed to a method for producing multiple
protein complexes comprising the steps of: (i) producing a vector comprising from 1 to 10 expression cassettes as defined herein; (ii) introducing the vector obtained in step (i) into a P. pastoris host cell; and (iii) incubating the P. pastoris host cell under conditions allowing the simultaneous expression of the genes present in the vector.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG 1: Map of the MIE of the invention. HE (1), promoter (2), secretion signal (3), marker/selection cassette (4), detection and/or purification polypeptide (5), terminator (6), BstXI (7).

FIG 2: Map of a particular embodiment of the MIE of the invention. alpha-MF: S. cerevisiae alpha-mating factor; His 6-3C: N-terminal hexahistidine-3C tag; KH10: C-terminal lysine-decahistidine.

FIG 3: Schematic vector maps for pPACE (acceptor), pBiPACE (acceptor), pPDK (donor), pPDC (donor) and pPDS (donor). AOX1 P: AOX1 Promoter; AOX1 T: AOX1 Terminator.

FIG 4: Transformation approaches for pPACE and pBiPACE Acceptor vectors. 1: digestion with ScaI/HE and electroporation. 2: Direct transformation of circular plasmid. A: One or more copies of linearized plasmid become integrated into P. pastoris genome at AOX1 locus. B: Multiple copies of circular plasmid are maintained episomally.

FIG 5: pPACE vector containing HPV18L1 protein (B) and SDS-PAGE electrophoretic profile of the HPV18L1 protein expressed in P. pastoris and purified by cation-exchange capture and S200 size exclusion chromatography (A). M: molecular ladder. I: Input (cell-free extract before purification). 1, 2, 3 and 4 correspond to different purification fractions over a size-exclusion chromatography column (Superdex 200, S200). HPV18L1 protein is mainly obtained in fraction 3.

FIG 6: pBiPACE vector containing Sch. pombe PSAT gene (B) and SDS-PAGE electrophoretic profile of the PSAT protein expressed in P. pastoris and purified by Ni^{2+} IMAC capture and S200 size-exclusion chromatography


FIG 8: Construction of a vector containing multiple expression cassettes by Acceptor-donor fusion method using pPACE and pPDC for co-expression of anti-MUC1 IIB6 scFv and PpHAC1 (B). 1: Cre-mediated recombination. SDS-PAGE electrophoretic profile of the anti-MUC1 IIB6 protein co-expressed together with PpHAC1 in *P. pastoris* and purified by Ni²⁺ IMAC capture and S200 size exclusion chromatography (A). M: molecular ladder. 5-20 correspond to different purification fractions. anti-MUC1 IIB6 scFv protein is mainly obtained in fractions 15, 16 and 17. Fraction PNGF is fraction 16 treated with the deglycosylase PNGase F (NEB), and fraction neg represents the untreated control; the absence of a conspicuous shift in apparent molecular size indicates that overexpressed anti-MUC IIB6 scFv is very little glycosylated, if at all.

FIG 9: Construction of a vector containing multiple expression cassettes by Acceptor-donor fusion method using pPACE and pPDC for co-expression of human complement factor B, fB, and furin (B). 1: Cre-mediated recombination. SDS-PAGE electrophoretic profile of the fB protein co-expressed together with furin in *P. pastoris* and purified by anti-FLAG affinity capture and S200 size-exclusion chromatography (A). M: molecular ladder. Left-hand side SDS-PAGE gel: Inp, secreted recombinant fB; Q, fB purified by ionic exchange chromatography; GF, last purification step over a size-exclusion chromatography column (S200). Right-hand side SDS-PAGE gel: C3, complement factor C3, purified from plasma. C3+aB, C3 treated with fB obtained from plasma (aB stands for “authentic” fB). +fD, C3+aB+fD, where fD cleaves triggers the proteolytic cleavage of C3 into C3b and fB into fBa and
fBb. C3+rB, as previous fraction C3+aB but using the recombinant fB produced in *P. pastoris*. +fD, C3+rB+fD, as previous fraction C3+aB+fD. The last two lanes (C3+rB and +fD) are as the previous two lanes, using recombinant fB from a second, independent expression and purification experiment, to demonstrate the reproducible authentic-like activity obtained from recombinant fB.

FIG 10: Construction of a pBiPACE vector containing a single synthetic DNA sequence by restriction-ligation into BstBI and Pmel sites, comprising two independently transcribed units, a Cas9GFP unit and a guide RNA (gRNA) unit under the control of a second promoter (Prom). The Cas9GFP unit comprises the AOX1 promoter (red arrow), Cas9GFP coding sequence, and the AOX1 terminator (TT). The gRNA unit is composed of a promoter (Prom) that drives the transcription of the ensuing two elements, a 19-21 bp target region (1) and the rest of the gRNA (2).

DETAILED DESCRIPTION OF THE INVENTION

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 present invention are DNA, most preferred double-stranded (ds) DNA.

In the sense of the present invention a “multiple integration element” (MIE) is understood as a nucleic acid comprising several sequence elements as defined in the first aspect of the invention that allows for cloning of one or more GOI and may be assembled by Cloning HE/BstXI-digested cassettes containing the gene of interest into the HE/BstXI sites of the MIE. The MIE of the invention is optimized for multiple gene expression in *P. pastoris*.

The term “expression cassette” is defined as known in the art of molecular biology as a nucleic acid sequence composed of one or more genes and the sequences controlling their expression (promoter, open reading frame and terminator). Additional sequence elements such as secretion signals, marker genes or purification tags may also be comprised within the expression cassette.
The term "vector" 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 propagation and/or expression of the gene of interest. Appropriate vectors of the present invention are plasmids, expression vectors, transfer vectors, more preferred eukaryotic gene transfer vectors.

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.

The present invention provides a nucleic acid or polynucleotide comprising a MIE as defined in the first aspect of the invention (and Figure 1). The skilled person will understand that the MIE may comprise additional sequence elements, such as additional restriction sites, purification tags, etc. The MIE may also contain irrelevant sequences, spacers, etc.

The MIE contains a promoter and a terminator which are suitable for driving protein expression in *P. pastoris* hosts. Suitable promoters and terminators are selected from *AOX1, FLD1, FMD1, GAP* and *TEF1*. In order to satisfy the requirements of the multiprotein expression tools of the present invention, the promoter and terminator of the present MIE is modified with respect to the standard by destroying relevant restriction sites, namely, the HE, BstBI, Ncol, Pmel and BstXI restriction sites, if they are present in the sequence. The skilled person will understand that a suitable terminator sequence must accompany the selected promoter. The promoter contains a Kozak sequence functional in *P. pastoris*, preferably, the promoter has the following Kozak sequence: AAAACGATG.

In particular embodiments the promoter is AOX1 promoter and the terminator is AOX1 terminator. The AOX1 promoter is a strong, tightly regulated promoter which is strongly repressed when *P. pastoris* is grown on glucose, glycerol or ethanol. Upon depletion of these carbon sources, the promoter is de-repressed, but is fully induced only upon addition of methanol. This enables uncoupling the growth from the production phase, such that biomass is accumulated prior to protein expression. Therefore, cells are not stressed
by the accumulation of recombinant protein during growth phase, and even
the production of proteins that are toxic to *P. pastoris* is possible.
Furthermore, it may be desirable to co-express helper proteins like
chaperones at defined time points, for example, before the actual target
protein is formed.

The AOX1 promoter of the present MIE is modified with respect to standard
AOX1. In particular, Pmel and BstXI restriction sites are eliminated from the
wild-type AOX1 promoter sequence. In particular embodiments, the AOX1
promoter in the MIE may contain additional modifications, for instance, other
restriction sites may be destroyed. In a particular embodiment, the AOX1
promoter of the present MIE contains an additional modification consisting of
the destruction of Sacl site. The Sacl site is usually used for plasmid
linearization, and in this particular embodiment is deleted to allow for cassette
multimerization without causing the multiplication of Sacl sites.

In other embodiments, the AOX1 promoter sequence is further modified in
order to enhance protein expression in the *P. pastoris* host. While optimizing
AOX1 sequence, care needs to be taken not to generate undesired restriction
sites that would be incompatible with other advantageous functionalities of
the present multiprotein expression tool. In one particular embodiment, the
sequence of the AOX1 promoter is SEQ ID NO: 1.

The MIE contains a homing endonuclease (HE) site and a specifically
designed restriction enzyme site (BstXI) flanking the promoter and the
terminator. 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.

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-Scel, I-Ceul, I-
Ppol, I-Hmul I-Crel, I-Dmol, PI-Pful and I-Msol, PI-Pspl, I-Scel, other
LAGLIDAG group members and variants thereof, SegH and Hef or other GIY-
YIG homing endonucleases, I-Apel, I-Anil, Cytochrome b mRNA maturase bl3, PI-77/1 and PI-Tfull, PI-Thyl and others. Corresponding enzymes are commercially available, e.g. from New England Biolabs Inc., Ipswich, MA, USA.

In particular embodiments, the HE is selected from the group consisting of a I-Ceul site, a PI-SceI site, a I-SceI site, a I-Ppol site, a I-Crel site and a I-Dmol site.

The BstXI restriction site in the MIE is complementary to the specific HE site. The recognition sequence of BstXI is defined as CCANNNNNN'TTGG (SEQ ID NO: 2) (apostrophe marks position of phosphodiester link cleavage). The residues denoted as N are selected to match the overhangs generated by the HE cleavage. Table 1 shows HE/BstXI combinations appropriate for use in the MIE of the invention.

Table 1. HE/BstXI combinations appropriate for the MIE

<table>
<thead>
<tr>
<th>HE</th>
<th>HE site</th>
<th>Corresponding BstXI site</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Ceul</td>
<td>TAACTATAACCGTCCCTAAGGTAGCGAC (SEQ ID NO: 3)</td>
<td>ccatCTAAttgg (SEQ ID NO: 4)</td>
</tr>
<tr>
<td></td>
<td>CGCTACCTTAGGACCGTTATAGTTATG (SEQ ID NO: 5)</td>
<td>ggtaGATTaacc (SEQ ID NO: 6)</td>
</tr>
<tr>
<td>PI-SceI</td>
<td>ATCTATGCtCGGGTTGCCGAGAAAGAGGGTAATTGGA ATGG (SEQ ID NO: 7)</td>
<td>ccatGTGcctgg (SEQ ID NO: 8)</td>
</tr>
<tr>
<td></td>
<td>TAGATAcACGCCCCTCTTTCTCCATTATTTACC (SEQ ID NO: 9)</td>
<td>ggtaCAGGgacc (SEQ ID NO: 10)</td>
</tr>
<tr>
<td>PI-Ppol</td>
<td>TGGCAAAACAGCTATTTATGGGTATATGGGT (SEQ ID NO: 11)</td>
<td>ccatTTATcctgg (SEQ ID NO: 12)</td>
</tr>
<tr>
<td></td>
<td>ACCGTtTTGCTGATATAccCAAAATAACCA (SEQ ID NO: 13)</td>
<td>ggtaAATAgacc (SEQ ID NO: 14)</td>
</tr>
<tr>
<td>I-SceI</td>
<td>TAGGGATAACAGGTAAT (SEQ ID NO: 15)</td>
<td>ccatATAActgg (SEQ ID NO: 16)</td>
</tr>
<tr>
<td></td>
<td>ATCCCTATTGTCCTATT (SEQ ID NO: 17)</td>
<td>ggtaTATTgacc (SEQ ID NO: 18)</td>
</tr>
<tr>
<td>I-Ppol</td>
<td>CTCTCTTAAAGGTAGC (SEQ ID NO: 19)</td>
<td>ccatTTAActtg (SEQ ID NO: 20)</td>
</tr>
<tr>
<td></td>
<td>GAGAGAATTCCATCG (SEQ ID NO: 21)</td>
<td>ggtaAATTgacc (SEQ ID NO: 22)</td>
</tr>
<tr>
<td>I-Crel</td>
<td>CTGGGTTCAAAAACGTCTCGTGAGACAGTTGG (SEQ ID NO: 23)</td>
<td>ccatGTGActtg (SEQ ID NO: 24)</td>
</tr>
<tr>
<td></td>
<td>GACCCAAAGTTTGCAGCActcTGTCAAACC (SEQ ID NO: 25)</td>
<td>ggtaCACTgacc (SEQ ID NO: 26)</td>
</tr>
<tr>
<td></td>
<td>ATGCGCTTGGCGGGTGTAAGTTCCGGGCAGCAT (SEQ ID NO: 27)</td>
<td>ccatGTAActgg (SEQ ID NO: 28)</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td>TACGGGAACCGGCCCATTCAAGGGCCGCAGCGTA (SEQ ID NO: 29)</td>
<td>ggtacATTgacc (SEQ ID NO: 30)</td>
</tr>
</tbody>
</table>

In one embodiment the invention provides a MIE wherein the HE/BstXI restriction sites are selected from the following HE/BstXI pairs: SEQ ID NO: 3/SEQ ID NO: 4, SEQ ID NO: 7/SEQ ID NO: 8, SEQ ID NO: 11/SEQ ID NO: 12, SEQ ID NO: 15/SEQ ID NO: 16, SEQ ID NO: 19/SEQ ID NO: 20, SEQ ID NO: 23/SEQ ID NO: 24, SEQ ID NO: 27/SEQ ID NO: 28, or their complementary sequences.

One of the main advantages of using *P. pastoris* as a protein production host is its ability to secrete high titers of properly folded, post-translationally processed and active recombinant proteins into the culture media. As a rule of thumb, proteins secreted in their native hosts will also be secreted in *P. pastoris*. However, quite often it is convenient to achieve secretion of expressed proteins which are not naturally secreted. For this aim, the MIE of the invention includes a sequence encoding for a secretion signal peptide. In a particular embodiment, the secretion signal peptide is the *S. cerevisiae* a-mating factor secretion signal (alpha-MF). As explained above, the particular design of the MIE provides a dual functionality: by simply selecting the appropriate digestion strategy, the secretion signal may be fused with the gene of interest, or not, depending on whether the user desires to enhance secretion of the protein of interest or if intracellular expression is preferred.

Further, in a particular embodiment of the invention the alpha-MF secretion signal sequence has been modified to enhance its secretory potential in *P. pastoris*. In a particular embodiment, the sequence for the alpha-MF secretion signal is SEQ ID NO: 50.

The MIE of the invention also contains the sequence encoding for a selection marker gene, which allows for rapid selection of clones containing the gene(s) of interest. Non-limiting selection marker genes appropriate for use in the MIE of the invention are lacZ, luciferase, β-GAL, CAT and fluorescent encoding protein genes such as GFP, BFP, YFP, CFP and their variants. In a particular embodiment, the selection marker gene is LacZ.
The MIE also contains a sequence encoding a detection and/or purification polypeptide. A “purification polypeptide” is, as generally known in the art, a peptide or protein useful for isolating or purifying the expressed protein. Said polypeptides often have affinity for a particular binding molecule, a property that is used for the purification/isolation. Likewise, a “detection polypeptide” is a peptide or protein useful for easily detecting the expressed protein. Sometimes the particular polypeptide serves both detection and purification purposes. Appropriate, non-limiting, purification and/or isolation polypeptides for the present MIE are: poly-histidine (6-20 histidine residues), FLAG-tag, StrepII-tag, HA epitope, c-Myc epitope, V5 epitope, calmodulin-binding peptide (CBP), enhanced green fluorescent protein (EGFP), enhanced yellow fluorescent protein (EYFP), Rudolph red fluorescent protein (RFP), monomeric Cherry fluorescent protein (mCherry), cyan fluorescent protein (CFP); these tags may be preceded by a protease cleavage sequence selected from the following: tobacco etch virus (TEV), 3C, enterokinase or carboxypeptidase A (CPA). In a particular embodiment the purification polypeptide is a poly-histidine sequence (KHs tag) comprising from 6 to 20 histidine residues following the Pmel restriction site. In particular, the KHs tag comprises from 6 to 16, or from 7 to 15, or from 8 to 12, for example 9, 10 or 11 histidine residues. Said KHs tag allows the GOI to be expressed with a C-terminal purification histidine tail.

The MIE of the invention may contain additional sequence elements. For instance, it may comprise further restriction sites and/or further detection and/or purification polypeptides. In one embodiment, the MIE additionally comprises a KpnI and/or MscI and/or XbaI restriction sites. In another embodiment, the MIE comprises a second KHs tag, wherein the second KHs tag comprises from 4 to 20 histidine residues. In particular, the second KHs tag comprises from 5 to 10, for example 6, 7, 8 or 9 histidine residues. In a particular embodiment, the MIE comprises the sequence elements as represented in Figure 2. Further particular embodiments define a MIE of SEQ ID NO: 31 and SEQ ID NO: 32.

In a particular embodiment the invention provides the MIE contains AOX1 promoter and terminator sequences devoid of Pmel and BstXI restriction sites, a BstBI restriction site, a secretion signal sequence, an Ncol restriction site, a selection marker gene, a Pmel restriction site, a sequence encoding for
a KHis tag comprising from 6 to 20 histidine residues and a BstXI restriction site. In another particular embodiment the MIE contains AOX1 promoter and terminator sequences devoid of Pmel and BstXI restriction sites, a BstBI restriction site, an alpha-MF secretion signal sequence, an Ncol restriction site, a selection marker gene, a Pmel restriction site, a sequence encoding for a KHis tag comprising from 6 to 20 histidine residues and a BstXI restriction site. In another particular embodiment the MIE contains AOX1 promoter and terminator sequences devoid of Pmel and BstXI restriction sites, a BstBI restriction site, an alpha-MF secretion signal sequence, an Ncol restriction site, a lacZ selection marker gene, a Pmel restriction site, a sequence encoding for a KHis tag comprising from 6 to 20 histidine residues and a BstXI restriction site. Further particular embodiments provide MIEs where the AOX1 sequence is SEQ ID NO: 1 and the alpha-MF sequence is SEQ ID NO: 50.

The invention also refers to an expression cassette obtained by inserting at least one GOI into the MIE. It is noted that the expression cassette of the invention may contain more than one, particularly from 1 to 10 GOI. The one or more GOI may be inserted into the MIE of the invention by methods known to the skilled person, e.g. by restriction enzyme digestion/ligation via compatible sites within the MIE or by recombination, preferably using ligase-independent methods, e.g. the SLIC method. In one embodiment, the GOI may be inserted between the Ncol and Pmel sites and the resulting expression cassette will then include the signal sequence, such that said GOI will be expressed as a secretion peptide(s). The GOI may be alternatively inserted between the BstBI and Pmel sites and will then be expressed as an intracellular peptide. The resulting expression cassettes usually contain a C-terminal KHis tag. Alternative embodiments provide for expression cassettes comprising a N-terminal KHis tag and/or further N- and C-terminal fusions. N-terminal fusions can be added to the expression cassette by cloning the GOI into Ncol/KpnI restricted MIE, and C-terminal fusions can be added similarly by cloning into Pmel/MscI restricted MIE. The XbaI site in the MIE can be used for the insertion of e.g., GFP and other C-terminal fusion tags.

According to one aspect, the invention provides a vector comprising a MIE or an expression cassette as defined above. In particular embodiments of the present invention, the vector of the invention comprising the above defined
MIE or expression cassette additionally comprises at least one site for integration of the nucleic acid into a vector or host cell.

Particularly preferred integration sites that may be incorporated into the present vector can be selected from site-specific recombinases recognition sites, in particular LoxP site or short flippase recognition target (FRT) site. In a particular embodiment, the vector of the invention comprises at least one LoxP site or FRT site for integration by Cre-loxP recombination or Flp-FRT recombination, respectively.

In particular embodiments of the present invention, the vector as described herein additionally contains sequences encoding for 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 zeocin, phleomycin, ampicillin, chloramphenicol, gentamycin, spectinomycin, blasticidin, hygromycin and kanamycin. In a particular embodiment, the antibiotic resistance marker is zeocin. In another particular embodiment, the sequence encoding for the zeocin resistance marker is optimized for improved zeocin expression in P. pastoris host. In another particular embodiment the sequence encoding for the zeocin resistance marker has SEQ ID NO: 33.

In addition to all the above, vectors of the invention usually comprise an origin of replication (ori). Examples of suitable oris are BR322, ColE1, as well as conditional origins of replication such as OriV and R6Kγ, 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.

One of the most imposing bottlenecks in the P. pastoris host is that genome integration is a prerequisite for expression, since very few autonomously replicating sequences are available. Genome integration requires large amounts of restriction enzyme-linearized plasmid per transformation (typically several micrograms pure plasmid per electroporation) and lengthy procedures for selection of high-copy integrands. The present invention provides for vectors comprising the PARS1 sequence that circumvent the need for
genome integration of the expression cassette thereby allowing expression tests to be run both from a self-replicating plasmid (like in conventional bacterial expression systems) and from a genome-integrated copy of the linearized plasmid (for stable protein production). Within the vector of the invention, PARS1 sequence is preferably inserted upstream the site for integration (for example upstream a loxP site).

In some embodiments, the vector comprises a MIE as defined above, at least one site for integration, a resistance marker and an origin of replication. In other embodiments, the vector additionally comprises a PARS1 sequence. In one embodiment, said origin of replication is BR322, said resistance marker is zeocin and said MIE comprises I-CeuI restriction site, for example a MIE of SEQ ID NO: 31. In another embodiment said origin of replication is BR322, said resistance marker is zeocin, said MIE comprises I-CeuI restriction site, for instance a MIE of SEQ ID NO: 31, and the vector additionally comprises a PARS1 sequence upstream from the site for integration. In another embodiment said origin of replication is a conditional origin of replication selected from OriV and R6Ky, said resistance marker is an antibiotic selected from the group consisting of ampicillin, chloramphenicol, gentamycin, spectinomycin, blasticidin, hygromycin and kanamycin, and said MIE comprises PI-SceI restriction site and AOX1 promoter lacking also SacI restriction site, for example a MIE of SEQ ID NO: 32.

Further particular embodiments of the present invention are vectors containing more than one of the sequence elements as defined above. For example, the vector may contain more than one recombination sequence for a site specific recombinase, e.g. 1 to 6, for example 2, 3, 4 or 5 of such recognition sequences, preferably 1 to 6, for example 2, 3, 4 or 5 loxP sites.

Particular embodiments of the invention provide a vector consisting of a sequence selected from the group consisting of SEQ ID NO: 34 to SEQ ID NO: 38.

In another aspect, the present invention is directed to a host cell containing the nucleic acid of the invention and/or the vector of the present invention. The host cells may be prokaryotic or eukaryotic. Prokaryotic hosts according to the present invention include bacteria, in particular E. coli such as
commercially available strains like TOP10, DH5α, HB101 etc., suitable for the propagation of the above vectors. Eukaryotic host cells are preferably *P. pastoris* strains for expression of the GOI. Virtually any *P. pastoris* strain can be used with the vectors according to the invention. The only requirement when zeocin-resistance cassette (Sh ble) is the resistance marker in the vector is that the Zeo resistance marker has not been used in the recipient strain, which covers most or all *P. pastoris* strains of industrial application. However, if the Zeo resistance marker had already been used in the recipient strain, the plasmid fusion mechanism allows to circumvent this limitation by adding any genetic locus suitable for integration into the multigene construct via the donor vectors (see below), therefore opening up the whole *P. pastoris* genome for integration; e.g., the gene for tubulin could be cloned into a Donor vector (e.g. pPDC, see below) and fused with the Acceptor vector carrying a GOI (e.g. pPACE-GOI), the resultant construct could be linearized at the ADE2 gene site therefore directing the plasmid toward the ADE2 locus exploited in the commercially available *P. pastoris* PichiaPink strains (Invitrogen). Otherwise, the present system does not impose any undue restraints on the yeast strain's genome and/or phenotype.

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 (Green M and Sambrook J (2012), *infra*).

The present invention also relates to a method for producing a vector containing multiple expression cassettes as defined above. 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. In certain embodiments, the vector of the invention comprises from 1 to 10 expression cassettes each containing a GOI, cloned in series by repeated use of the HE/BstXI multimerization cassette. In other embodiments, 1-10 GOI can be cloned into several different vectors of which one must be an acceptor vector, and then fused together by Cre-loxP recombination into a larger plasmid containing the complete set of GOIs cloned into the individual plasmids as explained below. In a further embodiment pertaining only to
proteins targeted to the secretory pathway of *P. pastoris*, the person skilled in
the art may choose to concatenate several GOI (typically, between 1-5) into a
polygene with a Kex2 or a furin cleavage sequence between the 3’ end of a
GOI and the 5’ of the following GOI. Post-translational proteolytic processing
5 of the polyprotein inside the endoplasmic reticulum should split the
polyprotein into the constituent proteins. Thus in other embodiments the
vector of the invention comprises from 1 to 10 expression cassettes,
preferably from 1 to 5 expression cassettes, wherein each cassette may
contain either 1 GOI or 1-5 concatenated GOI.

10 As already mentioned, the vectors of the invention may contain a site for
integration into another vector or host cell. In preferred embodiments, this site
for integration is a site-specific recombinase recognition sequence, such as a
loxp site or FRT site. This has the advantage that said vectors, which may
contain one or several expression cassettes obtained as described above,
can be assembled into a fusion vector comprising several vector entities
separated from each other by a 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.

15 A "vector entity" as used herein refers to a vector according to the invention
comprising at least one expression cassette, at least one site for integration,
a resistance marker, and which is suitable for propagation. Preferred
examples of the vector entities are those defined by SEQ ID NO: 34 to SEQ

20 ID NO: 38.

Thus, 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:

30 (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 so as to generate a mixture of fusions of the vector
entities comprising 2 to n of said vector entities;

35 (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;

5 (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 2 to 9, for example 3, 4, 5, 6, 7 or 8.

The term "culturing" means that the cells are incubated under the appropriate conditions for viability of the host cells.

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 is generally not specifically limited as long as a corresponding number of resistance markers are available. Examples of resistance markers useful in the context of this aspect of the present invention are as already defined above. Further details regarding the method for assembly of fusion vectors are disclosed in the examples below.

25 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; (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 as defined above.

According to a particular embodiment of the invention, the above-defined vector entities contain at least one multiple expression cassette. Each expression cassette present in the vector entity may contain one or more genes of interest. Preferred embodiments of the above fusion vector contain LoxP sites and Cre is the corresponding recombinase enzyme for vector assembly.

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 (antibiotic) 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 as defined above. In a particular embodiment, the selectable marker is R6Kγ 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 (1) of the above vector assembly method is transformed into bacterial cells lacking the pir gene (such E. coli strains TOP10, DH5α, HB101 or other commercially available pir+ cells). Donor vectors according to the present embodiment may contain AOX1
promoter lacking Sacl restriction site.

The vectors of the invention may be included in a kit for cloning and/or expression of proteins. One embodiment is directed to a kit for cloning and/or expression of proteins comprising at least one, for example from 1 to 10, vectors as disclosed herein. Preferably the kit comprises at least one vector, i.e. from 1 to 10 vectors, comprising a MIE or an expression cassette as defined above, an origin of replication such as BR322, ColE1, OriV or R6Kγ, a resistance marker, such as zeocin, ampicillin, chloramphenicol, gentamycin, spectinomycin, and kanamycin resistance markers, at least one site for integration such as a LoxP site or FRT site, and optionally a PARS1 sequence. In addition to the vectors as defined above, the kit of the invention may contain a host cell suitable for propagation of said vector(s). The kits may also contain *P. pastoris* host cells suitable for expression.

In one embodiment the kit comprises: (i) at least one (acceptor) vector comprising a MIE or expression cassette which contains l-CeuI as HE restriction site, a BR322 origin of replication, an antibiotic resistance marker, such as ampicillin, chloramphenicol, gentamycin, spectinomycin, blasticidin, hygromycin or kanamycin, a site for integration and, optionally, a PARS1 sequence; and/or (ii) at least one (donor) vector comprising a MIE which contains PI-SceI as HE restriction site, a conditional origin of replication, such as OriV or R6Kγ, an antibiotic resistance marker, such as ampicillin, chloramphenicol, gentamycin, spectinomycin or kanamycin and a site for integration. Preferably each vector included in the kit has a resistance marker different to the resistance marker of the other vectors. In a particular embodiment, the kit comprises: (i) at least one vector selected from the (acceptor) vectors with SEQ ID NO: 34 and SEQ ID NO: 35; and/or (ii) at least one (donor) vector selected from the vectors with SEQ ID NO: 36, SEQ ID NO: 37 and SEQ ID NO: 38. In particular embodiments the kit comprises both acceptor and donor vectors as defined above. In addition to the vectors as defined above, the kit of the invention may contain a host cell suitable for propagation of said vector(s). The kits may also contain *P. pastoris* host cells suitable for expression.

The kit according to the invention, as defined above, is particularly suitable for cloning and/or expression of multiprotein complexes in *P. pastoris*. In a
particular embodiment, the kit comprises acceptor and/or donor vectors as defined above, host cells suitable for propagation of the vectors and expression of the GOI and instructions for obtaining multiple-gene constructs and expression of the same to provide for the simultaneous expression of multiple proteins.

In another embodiment, the kit of the invention may comprise at least one vector comprising an expression cassette which contains one or more “helper genes”. In particular embodiments, said vector is of the kind of a donor vector as defined above, and said helper gene encodes for a protein serving a particular purpose related to the expression of the GOI, e.g., chaperons, activator, secretion enhancer, ribosome recycling factor, ribosome, wide-spectrum phosphatases, post-translational modifiers, etc. In the sense of the present invention, the vectors containing these helper genes are termed “helper vectors”. This helper vectors can be incorporated into the fusion vector following the vector assembly method described above and, when transformed into the appropriate P. pastoris host, the helper gene will be co-expressed with the GOI resulting in a positive impact in the yield or quality (or both) of the heterologous expression. Non-limiting examples of helper genes are listed in table 2.

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Helper gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. pastoris secretion</td>
<td>SchAC1p, PpHAC1p, PpPDI, PpBIP/Kar2</td>
</tr>
<tr>
<td>Secretion of mammalian proteins</td>
<td>Furin</td>
</tr>
<tr>
<td>General chaperones</td>
<td>Hsp40, Hsp90, Hsc70, HOP/StiP1</td>
</tr>
<tr>
<td>Genome targeting</td>
<td>Tubulin</td>
</tr>
</tbody>
</table>

Sc = S. cerevisiae
Pp = P. pastoris

Herein follows a complete description of preferred MIE, vectors, kits and methods of the present invention, all encompassed within a multi-protein expression system termed “MultiPichia”. General experimental procedures to generate multigene expression constructs are also provided.

A. MultiPichia System
A.1 MultiPichia vectors

The present invention provides as preferred exemplary embodiments small de novo designed vectors, which are called “Acceptor”, and “Donor” vectors (for plasmid maps see figure 3). Acceptor vectors for expression of proteins in *P. pastoris* [e.g. pPACE (SEQ ID NO: 34) and pBiPACE(SEQ ID NO: 35)] contain origins of replication derived from ColE1 (pMB1) and resistance marker (zeocin, Zeo). Donor vectors [e.g. pPDK (SEQ ID NO: 36), pPDC (SEQ ID NO: 37) and pPDS (SEQ ID NO: 38)] 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 (Kn), chloramphenicol (Cm), and spectinomycin (Sp). Preferably, one or two 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), which follows the basic layout outlined above. In particular, the MultiPichia MIE used in the present example contains, in addition to the HE and BstXI restriction sites, (1) the modified AOX1 promoter-terminator sequences, (2) alpha-MF secretion signal sequence flanked by restriction enzyme sites (5’ BstBI and 3’ NcoI); (3) a prokaryotic promoter upstream of the LacZα selection marker gene to support blue-white selection; (4) in the latter sequence, there are restriction sites designed to insert N-terminal (KpnI) and / or C-terminal fusion sequences (XbaI, Mscl) (e.g., StrepII, calmodulin binding peptide); and (5) an optional C-terminal lysine-decahistidine (Cter-KH10).

Obviously, due to the presence of the HE and BstXI sites, the expression cassettes can be easily interchanged between the vectors, either “empty” (LacZα) or with 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 which contain other promoter-terminator cassettes (e.g., GADH, TEF1) with any one of the resistance markers or others (e.g., blasticidin, hygromycin). The HE/BstXI site combinations can be used to multiply expression cassettes or to introduce alternative or multiple promoter sequences.
All Donor vectors contain a LoxP imperfect inverted repeat. This can be used for LoxP-mediated constructions (and deconstructions) of Acceptor / Donor multifusions as described above.

5

B.2 Multiple integration element (MIE)

The MIE 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 ligation-dependent as well as seamless cloning (e.g., BD InFusion, SLIC). In every case, the vector needs to be linearized, either by restriction digestion or by PCR with suitably chosen primers. Use of ultrahigh-fidelity polymerases such as Phusion (NEB) is preferred. The map of the present MIE is shown in Figure 1. DNA sequences for MIE are provided as SEQ ID NO. 31 (containing I-CeuI) and SEQ ID NO: 32 (containing PI-Scel). A map of the MIE used in the MultiPichia system is provided in Figure 2.

A.3 Tags, promoters and terminators

20 The AOX1 promoter is a strong inducible (with methanol) promoter that can be utilized in all strains of P. pastoris. Other alternative promoters are available (e.g., GADPH) or can be easily introduced should it be required.

In available vectors for P. pastoris a choice has to be made between cloning the gene insert in a vector for intracellular expression or for secretion. In the MultiPichia system, the same vectors can be used for both purposes because a α-MS sequence is always present in the vectors and the choice of localization is facilitated by the presence of a set of specific restriction sites. As exemplified in Figure 1, linearizing the vector with BstBI-Pmel allows intracellular expression by removing the αMF leader sequence, whereas linearizing with Ncol-Pmel allows secretion expression by ensuring the correct reading frame between the αMF and the gene of interest sequence. In the former case, the Kozak sequence that lies 5’ to the αMF has to be rebuilt after BstBI digestion since it is partially destroyed, which is conveniently done by the appropriate choice of PCR primers (infra).

The complete multiple integration element (MIE) in MultiPichia Donor and
Acceptor vectors, including promoters and terminators, excluding the HE and BstXI sites, can be exchanged with an expression cassette of choice by using restriction-ligation cloning with appropriate enzymes (e.g., EcoRI-Ncol, HE-Ncol). For example, the AOX1 promoter in pPACE can be replaced with the GAPDH promoter (pPACEGAP) for constitutive expression.

The MultiPichia Acceptor and Donor vectors provide optional N-terminal hexahistidine-3C tag (His6-3C) and / or C-terminal lysine-decahistidine (Cter-KH10) tags as well as a standardized mechanism for the introduction of further N- and C-terminal fusions. N-terminal fusions can be added to any one vector by cloning into Ncol-KpnI, and C-terminal fusions can be added similarly by cloning into Pmel-Mscl the desired cassette. The MultiPichia vectors also include an XbaI site between Pmel and Mscl, which can be used for the insertion of e.g., GFP and other C-terminal fusion tags.

A.4 Complex expression

For expression in *P. pastoris*, the MultiPichia multigene expression fusion vector is transformed into the appropriate expression host of choice. The present exemplary vectors (AOX1 promoter elements) can be used with all currently available *P. pastoris* strains. In the current implementation, the Acceptor vector carries a Zeocin resistance marker, hence if the multigene fusion is to be transformed into a *P. pastoris* strain harboring other Zeocin markers, alternative versions of the MultiPichia Acceptor vector with an alternative resistance that resolved the conflict should be prepared (e.g., blasticidin, hygromycin, G418). This can be easily achieved by PCR amplifying the vector less the resistance marker, and combine the resulting fragment with a PCR amplified resistance marker by SLIC or blunt-end ligation (using 5’-phosphorylated primers). All resistance markers present in the MultiPichia Donor vectors are intended for selection in *E. coli* and should not interfere with yeast expression.

Donor vectors of the present example depend for their propagation by *E. coli* of the pir gene product by virtue of the R6Ky conditional replication origin. These vectors cannot be used as such for expression in *P. pastoris* since they lack yeast-compatible resistance markers. However, the expression performance of the gene inserts present in Donor vectors can be
accomplished by fusion with an empty Acceptor vector, which would act as a vehicle into the expression host.

The present MultiPichia system for expression in *P. pastoris* contains two Acceptor vectors, pPACE and pBiPACE, which are identical except for the presence of a PARS1 (*Pichia* autonomous replication sequence 1) (Figure 3). pPACE requires genome integration and selection and leads to stable protein producing strains, whereas transformation of *P. pastoris* cells with pBiPACE leads to a transient phase whereby the vector stays as an episomal element. The advantage of pBiPACE over pPACE is that expression tests can be performed in a rapid and economic fashion while evaluating the relative performance of multiple constructs. Once a particular construct is selected for scaling up, linearization of pBiPACE or the use of pPACE will trigger genome integration of the expression cassette and a more robust expression strain.

These two approaches are outlined in Figure 4.

**B. Procedures**

**B.1 Cloning into MultiPichia vectors**

All Donor and Acceptors of the preferred embodiment for expression in *P. pastoris* contain an identical MIE with exception of the homing endonuclease site / BstXI tandem spanning the MIE (Figures 1 and 2). The MIE is designed for sequence and ligation-independent gene insertion methods. In addition, the MIE contains two sets of unique restriction sites (NcoI / KpnI and Pmel / MscI) that can be used as a classical multiple cloning site for conventional gene insertion by restriction / ligation. For automated applications, insertion of genes of interest is preferably done by recombination approaches (e.g., SLIC, BD InFusion).

**B.1.1 Single gene insertion into the MIE by SLIC**

There are several standard methods for insertion of single genes into vectors that have been published or commercialized (e.g., BD InFusion, SLIC). All these methods rely on the 3’ to 5’ exonuclease activity of DNA polymerases. In the absence of dNTPs, 5’ overhangs are created starting from blunt ends or overhangs by digestion from the 3’ end. If two DNA fragments contain the
same ca. 20-50 bp sequence at their ends, complementary overhangs can be produced that can anneal to produce a nicked double-stranded DNA joint fragment. This phenomenon can be exploited for ligation-independent combination of two or more DNA fragments containing terminal homologous sequences.

If T4 DNA polymerase or other commercially available DNA polymerases (e.g., BD InFusion) are used, seamless cloning can be carried out in a sequence and ligation-independent fashion and detailed protocols are available for the skilled person. This is advantageous for multiprotein expression since the longer constructs have a higher chance of containing restriction enzyme (either naturally present or introduced by unwanted mutagenesis) sites that would be incompatible with the conventional restriction / ligation cloning.

In the present invention, the SLIC strategy was tailored for inserting encoding DNAs amplified by Phusion polymerase into the MultiPichia Donor and Acceptor vectors. The streamlined, automatable protocol that is recommended for MultiPichia takes care of seamless integration of single genes into the expression cassettes and of concatamerization of expression cassettes into multigene constructs.

The following Protocol 1 represents an adaptation of the method described by Bieniossek et al. (2009, infra) and an improved process based on the method described by Li et al. (2007, infra). Protocol 1 is designed for manual operation although it is easily automatable to work in robotics platforms. Other cloning strategies can be used (e.g., BD InFusion) following the manufacturer’s recommendations.

**Protocol 1**: Single gene insertion by SLIC.

Reagents required:
- Phusion polymerase (2 U/ul)
- 5X HF / GC Buffer for Phusion polymerase
- dNTP mix (10 mM)
- T4 DNA polymerase (3 U/ul) (and 10X buffer)
- DpnI (20 U/ul)
E. coli competent cells
100 mM DTT, 2 M urea, 0.5 M EDTA
1 M IPTG, 100 mg/ml X-Gal or Bluo-Gal
Antibiotics

5

Step 1: Primer design

Primers for the SLIC procedure are designed such that their 5’ sequences provide homology regions complementary to the MultiPichia vectors. When treated with T4 DNA polymerase in the absence of dNTPs, long sticky ends would be generated.

Primers for the insert contain a DNA sequence corresponding to this homology region (“Adaptor sequence”), followed by a sequence that specifically anneals to the insert to be amplified (“Insert specific sequence”). Useful sequences used with the MultiPichia system are listed below (Table 3).

The “insert specific sequence” can be located upstream of a Kozak sequence if the gene of interest (GOI) is amplified from a vector that already contains expression elements and the MultiPichia vectors are linearized to eliminate the secretion peptide leader sequence. For secretion expression there is no need to include a Kozak sequence in the primers since it is already present in the vector. Only when intracellular expression is desired a Kozak sequence must be included in the forward primer so that a ribosome-binding site is provided in the final construct. A suitable Kozak sequence is contained in AAACGATG (underlined).

A choice can be made as to the inclusion of a stop codon (e.g., TAA) in the reverse primer between the “insert specific sequence” and the “vector adaptor sequence”. When a stop codon is included, a 3’ untagged protein product will be expressed; otherwise, the vector-encoding 3’ fusion will be appended to the expressed protein (e.g., in the default vector is a KHis10 tag).

Primers for PCR linearization of the vector backbone are designed to be complementary to the two adaptor sequences present in the primer pair chosen for insert amplification.
**Step 2:** PCR amplification of insert and vector  
**Step 3:** DpnI treatment of PCR products (optional)  
**Step 4:** Purification of PCR products

Steps 2-4 can be performed as per standard molecular biology techniques, e.g., Bieniossek et al. (2009, *supra*).

**Step 5:** T4 DNA polymerase exonuclease treatment  
**Step 6:** Mixing and annealing

The procedure for Steps 5 and 6 are available for a skilled person (e.g., Bieniossek et al. (2009, *supra*)). For the MultiPichia system the present invention incorporates a modification that increases the speed and ease of the method and which consists in carrying out both steps with premixed components (a “one-pot” reaction):

Reactions are prepared in 20 ul containing vector and insert in a 1:2 molar ratio (eluted in Step 4):

- 10X T4 DNA polymerase buffer: 2 ul
- 100 mM DTT: 1 ul
- 2 M urea: 2 ul
- DNA (vector and insert): 14 ul
- T4 DNA polymerase: 1 ul

Reactions are then carried out as follows; this sequence can be easily programmed in a thermocycler:

- **Incubation:** 23 °C for 20 min
- **Arrest:** Addition of 1 ul 0.5 M EDTA
- **Inactivation:** 75 °C for 20 min
- **Annealing:** 65 °C for 10 min, 37 °C for 30 min
- **Cooling:** 4 °C, on hold

**Step 7:** Transformation

Reaction mixtures (from Step 6) can be transformed into *E. coli* competent cells following standard transformation procedures.
Reactions for pPACE and pBiPACE plasmids are transformed as pACE and pACE2 of the ACEMBL system, except that 25 ug/ml Zeocin is used for selection. Since the MIE of the “empty” vectors encode the LacZα fragment, plating on Luria-Bertani medium with 1 mM IPTG and 0.1 mg/ml X-Gal or Bluo-Gal should result in white (recombinant) and / or blue (parental) colonies.

Reactions for Donor plasmids are transformed in pir+ E. coli cells as the Donor derivatives of the ACEMBL system, with the same antibiotics. For the Donor derivatives pPDC, pPDK, and pPDS, it is highly recommended that electrocompetent cells be used. Since the MIE of the “empty” vectors encode the LacZα fragment, plating on Luria-Bertani medium with 1 mM IPTG and 0.1 mg/ml X-Gal or Bluo-Gal should result in white (recombinant) and / or blue (parental) colonies.

Step 8: Plasmid analysis

This is done according to the ACEMBL protocols. The (optional) blue-white selection described in Step 7 should reduce the number of false positives and increase the suitability of MultiPichia for automated pipelines.

B.1.2 Sequential assembly in MIE by SLIC

The MIE according to the present invention can also be used to integrate genes of interest (GOIs) by using multi-fragment SLIC recombination. Multigene constructs in P. pastoris require their own promoter-terminator cassette if are to be expressed in different translation units (for an alternative approach vide infra).

A detailed protocol is outlined in the following Protocol 2:

Protocol 2. Sequential assembly by SLIC.

Reagents required:

- Phusion polymerase (2 U/ul)
- 5X HF / GC Buffer for Phusion polymerase
- dNTP mix (10 mM)
T4 DNA polymerase (3 U/ul) (and 10X buffer)
DpnI (20 U/ul)
E. coli competent cells
100 mM DTT, 2 M urea, 0.5 M EDTA
1 M IPTG, 100 mg/ml X-Gal or Bluo-Gal
Antibiotics

**Step 1:** Primer design

10 The MIE element according to the present embodiment is composed of tried-and-tested primer sequences. The latter sequences comprise the “Adaptor” sequences that can be used for inserting single genes or multigene constructs. Adaptor sequences form the 5’ ends of the primers used to amplify encoding DNAs to be inserted into the MIE. Insert specific sequences are appended to the 3’, and other DNA sequences (e.g., Kozak sequences, purification tags) can be added optionally as desired. Since each gene insert must be included into a promoter-terminator cassette, amplification of the terminator-promoter sequences are required to stitch together the GOIs in a multigene construct (except for the first one, which uses the vector-provided promoter, and the last one, which uses the vector-provided terminator).

Examples of successful adaptor sequences for a multigene construct are listed below (Table 3).

**Step 2:** PCR amplification of inserts and primers

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

**Step 4:** Purification of PCR products

All the above Steps are conducted as per Protocol 1.

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

**Step 6:** Mixing and annealing

As in Protocol 1, Steps 5 and 6 are best done together in the “one-pot” T4 DNA exonuclease treatment, mix and anneal procedure outlined above (Protocol 1). Even though several vector-to-insert molar ratios may be tried, a 1:2 ratio (for each insert) works best for the MultiPichia vectors.
Step 7: Transformation
Step 8: Plasmid analysis

Transformation can be performed as per Protocol 1. As above, blue-white selection can be used to select recombinant plasmids, specifically plasmids where the LacZα fragment has been replaced by the intended multigene construct cassette. Verification of the multigene construct by sequencing and (optionally) restriction digestion is highly recommended, especially as the number of fragments to combine becomes larger.

Table 3. Adaptor DNA sequences for single gene or multigene insertions into Multiphicha vectors by SLIC.

<table>
<thead>
<tr>
<th>Adaptor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>sAOX1InsFor (SEQ ID NO: 39)</td>
<td>Forward primer for insert amplification, for secretion using the α-MF sequence from S. cerevisiae. Needs to be followed by insert specific sequences (10-30 bp).</td>
</tr>
<tr>
<td>icAOX1InsFor (SEQ ID NO: 40)</td>
<td>Forward primer for insert amplification, for intracellular expression. The initiation ATG codon is included in the primer. Needs to be followed by insert specific sequences (10-30 bp).</td>
</tr>
<tr>
<td>AOX1InsRev (SEQ ID NO: 41)</td>
<td>Reverse primer for insert amplification (secretion or intracellular expression). Needs to be followed by insert specific sequences (10-30 bp). If a Cter-KHis10 tag is not desired, add a stop codon to the insert specific sequence (e.g., TAA).</td>
</tr>
<tr>
<td>AOX1VecFor (SEQ ID NO: 42)</td>
<td>Forward primer for vector amplification (secretion and intracellular expression), reverse complement of AOX1InsRev. No further extension required.</td>
</tr>
<tr>
<td>sAOX1VecRev (SEQ ID NO: 43)</td>
<td>Reverse primer for vector amplification for secretion expression, reverse complement of sAOX1InsFor. No further extension required.</td>
</tr>
<tr>
<td>icAOX1VecRev (SEQ ID NO: 44)</td>
<td>Reverse primer for vector amplification for intracellular expression, reverse complement of icAOX1InsFor.</td>
</tr>
<tr>
<td>Primer Name</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ttAOX1InsVecFor (SEQ ID NO: 45)</td>
<td>Forward primer for amplification of an insert 1 (GOI1) / vector AOX1 terminator (secretion or intracellular) expression. Needs to be preceded by insert 1 specific sequences (10-30 bp).</td>
</tr>
<tr>
<td>ttAOX1VecRev (SEQ ID NO: 46)</td>
<td>Forward primer for amplification of an insert 1 (GOI1) / vector AOX1 terminator (secretion or intracellular) expression. No further extension required.</td>
</tr>
<tr>
<td>pAOX1VecFor (SEQ ID NO: 47)</td>
<td>Forward primer for vector AOX1 promoter (secretion or intracellular expression) / insert 2 (GOI2) insert. No further extension is required.</td>
</tr>
<tr>
<td>pAOX1VecInsRev (SEQ ID NO: 48)</td>
<td>Reverse primer for vector AOX1 promoter (secretion or intracellular expression) / insert 2 (GOI2) insert. Needs to be preceded by insert 2 specific sequences (10-30 bp).</td>
</tr>
</tbody>
</table>

**B.1.3 Gene insertion by restriction / ligation**

The MIEs of the present invention provide a minimalistic set of restriction sites that can be used for restriction / ligation cloning if so desired, yielding functional expression cassettes. Specifically, the available restriction enzyme sites are NcoI and KpnI (5’, secretion), BstBI (5’, intracellular), and Pmel and MscI (3’). As with the seamless cloning strategies, in constructs targeted for secretion expression a choice is possible between Nter-His6-3C tags by using a 5’ KpnI site in the PCR fragment and a Cter-KHis10 tag using a 3’ Pmel site and no stop codon in the PCR fragment. Untagged N-terminal or C-terminal fragments are generated when NcoI is used for the 5’ end of the PCR fragment and MscI / Pmel plus a stop codon is used for the 3’ end of the PCR fragment. In constructs for intracellular expression the only available 5’ site is BstBI, which yields N-terminally untagged protein by default.
Genes of interest (GOI) can be subcloned into the MIEs by using standard cloning strategies of the MultiPichia vectors.

5 Protocol 3. Restriction / ligation cloning into the MIE.

Reagents required:
- Phusion polymerase (2 U/ul)
- 5X HF or GC Buffer for Phusion polymerase
- dNTP mix (10 mM)
- 10 mM BSA
- Ncol, KpnI, BstBI, Pmel, and/or Msci (and 10X Buffer)
- T4 DNA ligase (and 10X buffer)
- Antarctic Phosphatase (and 5X Buffer)
- E. coli competent cells
- 1 M IPTG, 100 mg/ml X-Gal or Bluo-Gal
- Antibiotics

Step 1: Primer design

For conventional cloning, PCR primers are designed such that they contain appropriate restriction sites, preceded by suitably chosen overhangs for efficient cleavage (e.g., NEB catalogue), and followed by ≥20 nucleotides overlapping with the gene of interest to be inserted.

All MultiPichia Donor and Acceptor vectors share the same restriction sites and MIEs; therefore they support classical cloning in a uniform manner for single-gene insertions. No additional sequences are required for secretion / intracellular expression constructs.

Step 2: Insert preparation
Step 3: Vector preparation
Step 4: Ligation

Steps 2-4 can be performed with the MultiPichia vectors as for the ACEMBL vectors, without modifications.
Step 5: Transformation
Step 6: Plasmid analysis

Steps 5 and 6 can be performed with the MultiPichia vectors as for the ACEMBL vectors. Blue-white selection is possible, as with the seamless cloning approach.

B.1.4 Multiplication by using the HE and BstXI sites

The MultiPichia Donor and Acceptor vectors contain MIE comprising a homing endonuclease (HE) site 5' to the promoter sequence and a especially chosen BstXI site 3' to the terminator sequence. I-Scel and a compatible BstXI site are in the Acceptor vectors and PI-Scel and a compatible BstXI site are in the Donor vectors. Each HE/BstXI site will generate compatible overhangs upon cleavage. When a HE/BstXI digested fragment is ligated into a HE site of a MultiPichia vector the asymmetry of the recognition sites ensures that the ligated extremes cannot be cut by any of the enzymes (the ligated HE and BstXI sites are “destroyed”). In this manner, successive rounds of multiplication can be performed iteratively.

Protocol 4. Multiplication by using HE / BstXI.

Reagents required:
- Homing endonucleases I-Scel, PI-Scel (and 10X Buffer)
- Restriction enzyme BstXI (and 10X Buffer)
- T4 DNA ligase (and 10X buffer)
- Antarctic Phosphatase (and 5X Buffer)
- *E. coli* competent cells
- Antibiotics

Step 1: Primer design

This Step is no longer necessary since all manipulations involve restriction digests and ligations.

Step 2: Insert preparation
Step 3: Vector preparation
4. **Step 4**: Ligation

5. **Step 5**: Transformation

6. **Step 6**: Plasmid analysis

5. **B.2 Cre-LoxP reaction of Acceptors and Donors**

All the MultiPichia Acceptor and Donor vectors contain the LoxP imperfect inverted repeat sequence (Seq ID NO: 49). Cre-loxP recombination to obtain fusion vectors may be performed as described in WO2010100278. In brief, recombination of two or more vectors bearing the LoxP sequence catalyzed by the Cre recombinase results in an equilibrium between all possible fusion vectors. Specific fusion vectors can be rescued by transforming the mixture in *E. coli* cells and plating them on appropriate antibiotics.

In contrast to WO2010100278, which discloses vectors for bacterial expression, the MultiPichia Acceptor vectors are the only ones that carry selection markers for yeast; therefore plasmid fusions must contain at least one Acceptor vector.

Fusion vectors in excess of 30 kb can be thus generated. The stability of these multigene constructs once inserted in *P. pastoris* genome can be tested by PCR amplification after serial passaging for more than 10 generations. Such stability tests show that the MultiPichia genome inserts are stable even when the selection antibiotic (e.g., Zeocin) is not present in the growth medium.

Since MultiPichia plasmids should be linearized for insertion into *P. pastoris* genome preferably by using ScaI, very large fusion plasmids should be constructed that do not have ScaI cleavage sites. This is most easily accomplished by custom gene synthesis and/or site-directed mutagenesis to remove any ScaI sites present in the genes of interest.

6. **C. MultiPichia kit for expression of proteins in *P. pastoris***

A kit according to the preferred embodiment for expression in *P. pastoris* contains:
- BW23473, BW23474 cells\(^1\) and/or Cre recombinase.
- pPACE, pPDK, pPDC, pPDS vectors
- pBiPACE vector

5\(^1\) *E. coli* strains expressing the pir gene for propagation of Donor derivatives (any pir+ *E. coli* strain could be used).

Optional components:
- Antibiotics: zeocin, chloramphenicol, kanamycin, spectinomycin.

10 - Enzymes:
  - 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)

15 Throughout the description and claims the word "comprise" and variations of the word, are not intended to exclude other technical features, additives, components, or steps. Additional objects, advantages and features of the invention will become apparent to those skilled in the art upon examination of the description or may be learned by practice of the invention. The following examples and drawings are provided by way of illustration, and they are not intended to be limiting of the present invention. Reference signs related to drawings and placed in parentheses in a claim, are solely for attempting to increase the intelligibility of the claim, and shall not be construed as limiting the scope of the claim. Furthermore, the present invention covers all possible combinations of particular and preferred embodiments described herein.

EXAMPLES

30 Examples of multiprotein expressions by using the above-described MultiPicha system are shown to illustrate the gene combination procedures outlined above. Reactions presented were carried out manually following the protocols provided in the above Section B.

35 Example 1: SLIC cloning into MultiPicha vectors: HPV16L1 and HPV18L1
Genes coding for full-length herpes virus-like particles (VLPs) 16L1 and 18L1 proteins (HPV16L1 and HPV18L1, respectively) were amplified from custom synthesized DNAs (Genscript) by using the following primer pairs for HPV16L1: icAOX1InsForHPV16L1 (SEQ IS NO: 51), which is a derivative of the icAOX1InsFor adaptor sequence (Table 3) elongated at its 3’ end with the 5’ end of the HPV16L1 gene, and AOX1InsRevHPV16L1 (SEQ ID NO: 52), a derivative of the AOX1InsRev adaptor sequence (Table 3) elongated at its 3’ end with the 3’ end of the HPV16L1 gene; and likewise for HPV18L1: icAOX1InsForHPV18L1 (SEQ ID NO: 53) and AOX1InsRevHPV18L1 (SEQ ID NO: 54), which are derivatives of the corresponding adaptors (Table 3). Both synthetic genes were amplified and prepared following the protocols described above. Linearized vector backbones were generated by digestion of pPACE with BstBI-Pmel. Above Protocol 1 (Section B.1.1) was followed, resulting in pPACE-HPV16L1his and pPACE-HPV18L1his (Figure 5; “his” means Cter-KHis10 tagged). The same constructs but with a stop codon were also prepared using above Protocol 1 (Section B.1.1), resulting in pPACE-HPV16L1 and pPACE-HPV18L1 (Figure 5). One clone per construct was linearized with Sacl and electroporated into P. pastoris SMD1168 cells. Yeast colonies were analyzed to find positive transformants. One transformant per construct was used for expression of tagged and untagged HPV16L1 and HPV18L1 proteins, which were purified by cation-exchange capture (5-ml HiTrap S XL, GE Healthcare) and S200 (Superdex 200, GE Healthcare) chromatography, resulting in pure proteins (Figure 5). All four proteins eluted from the S200 column with the expected molecular size (pentameric). For HPV16L1 and HPV18L1, the untagged versions were more stable than the corresponding oligohistidine tagged versions, which shows some proteolytic degradation.

The flexibility and convenience of the design of the MultiPichia Acceptor vector is illustrated by the facile choice of untagged or C-terminally tagged fusion constructs by including (or not) a stop codon at the cloning stage; for these targets, this minimal difference proved critical for the stability of the expressed product.

Example 2: SLIC cloning into pBiPACE and transient expression: Sch. pombe PSAT
The gene coding for full-length Schizosaccharomyces pombe phosphoserine aminotransferase (PSAT, SerC) (amino acids 1-389) was amplified from pOPINF-PS using icAOX1InsForPS (SEQ ID NO: 55) and AOX1InsRevPS (SEQ ID NO: 56), which are derivative of the adaptor sequences (Table 3) elongated at their 3’ ends with 5’ and 3’ insert specific sequences. The forward PCR primer is partially complementary to the N-terminal His6-3C tag encoded by pOPINF-PS, hence the resultant PCR amplicon contains an affinity oligohistidine tag; to avoid a second oligohistidine tag in the construct, the reverse primer insert a stop codon (TAA) at the 3’ end of the coding sequence. The PSAT gene (SEQ ID NO: 57) was amplified following the protocols above and inserted into the pBiPACE Acceptor vector previously linearized with BstBI-Pmel, following above Protocol 1 (Section B.1.1), resulting in pBiPACE-hisPS (Figure 6). The construct’s identity was verified by sequencing. The pBiPACE-hisPS plasmid was electroporated in P. pastoris SMD1168 cells, where it is maintained episomally. One colony from a zeocin-containing plate was used for expression of hisPS, which was purified by Ni²⁺ Immobilized metal ion affinity chromatography (IMAC) capture and S200 size exclusion chromatography, resulting in pure protein (Figure 6).

Example 3: Construction of a vector containing multiple expression cassettes for co-expression of anti-MUC1 IIB6 scFv and PpHAC1

Plasmids pPACE-αMUC1his (encoding a Cter-KHis8 tagged scFv derived from the αMUC1 IIB6 antibody, SEQ ID NO: 58) and pPACEDAαSacI-HAC1 (encoding the P. pastoris HAC1 transcription factor, SEQ ID NO: 59 cloned in a pPACE vector where the SacI site in the AOX1 promoter has been mutated) were prepared following the methods described above (Section B.1). The gene expression cassette encoding for PpHAC1 (comprising AOX1ΔSacI promoter, PpHAC1 gene, and the AOX1 terminator) was excised by digestion with I-CeuI and BstXI. The liberated fragment was inserted into I-CeuI digested and Antarctic-phosphatase treated pPACE-αMUC1his resulting in pPACE-αMUC1his-HAC1. The ligation mixture was then transformed into E. coli cells, the combined plasmid was recovered by plating on zeocin plates and verified by restriction digestion. The combined plasmid was linearized with SacI and electroporated into P. pastoris SMD1168 cells. Expression and purification by Ni²⁺ capture (IMAC) and S200 size exclusion chromatography resulted in enhanced secretion of anti-MUC1 IIB6 scFv (Figure 7).
Example 4: Construction of a vector containing multiple expression cassettes by Acceptor-donor fusion and co-expression of anti-MUC1 IIB6 scFv and PpHAC1

Plasmids pPACE-αMUC1his (encoding a Cter-KHis8 tagged scFv derived from the αMUC1 IIB6 antibody, SEQ ID NO: 58, targeted for secretion with the αMF leader sequence) and pPDCΔSacI-HAC1 (encoding the P. pastoris HAC1 transcription factor, SEQ ID NO: 59, cloned in a pPDC vector where the SacI site in the AOX1 promoter has been mutated), prepared following the methods described above (Section B.1), were assembled as illustrated in Figure 8. Following above methods (Section B.2), Cre recombination was used to generate the fusion plasmid from the two individual expression constructs, which was selected on LB agar plates with zeocin and chloramphenicol. Clones were verified by restriction analysis. The combined plasmid was linearized with SacI and electroporated into P. pastoris SMD1168 cells. Expression and Ni²⁺ IMAC capture and S200 size exclusion chromatography resulted in highly levels of secreted anti-MUC1 IIB6 scFv (Figure 8).

Example 5: Coexpression from Acceptor-Donor fusions: human fB / Furin

Plasmids pPACE-fB (encoding a Nter-His6 tagged human complement factor B, fB, SEQ ID NO: 60, targeted for secretion with the αMF leader sequence) and pPDCΔSacI-flagFurin (encoding Nter-FLAG Furin, SEQ ID NO: 61, cloned in a pPDC vector where the SacI site in the AOX1 promoter has been mutated), prepared following the methods described above (Section B.1), were assembled as illustrated in Figure 9. Following above methods (Section B.2), Cre recombination was used to generate the fusion plasmid from the two individual expression constructs, which was selected on LB agar plates with zeocin and chloramphenicol. Clones were verified by restriction analysis. The combined plasmid was linearized with SacI and electroporated into P. pastoris SMD1168 cells. Expression and Anti-FLAG affinity capture and S200 size exclusion chromatography resulted in highly levels of secreted human fB (Figure 9). Functional assays verified that the recombinant protein has authentic activity towards its cognate substrate (C3) (Figure 9).
Example 6: *Pichia pastoris* genome editing with CRISPR/Cas9

The modular design of the MultiPichia plasmids allow the generation of recombinant plasmids for applications other than recombinant protein production, such as genome editing and cell strain engineering. Plasmid pBiPACE-Cas9GFP-gRNA (Figure 10) was prepared following the methods above (Section B.1), where the Cas9GFP-gRNA insert, SEQ ID NO: 62, was synthesized as a single DNA molecule and inserted by restriction-ligation into the MIE between the pBiPACE vector’s BstBI and Pmel sites. The guide RNA (gRNA) encoded in the insert served to direct the Cas9GFP protein to a specific genomic locus where a mutation was introduced. The plasmid was electroporated into *P. pastoris* SMD1168 cells and induction of Cas9 was trigged by methanol induction. The successful introduction of the sequence substitution encoded in the gRNA DNA fragment was confirmed by sequencing, thereby demonstrating the possibility of performing genome engineering in *P. pastoris* with the CRISPR/Cas9 system.

REFERENCES CITED IN THE APPLICATION


WO2010100278


CLAIMS

1. An isolated multiple integration element (MIE) comprising, in the specified order:

5

a) a homing endonuclease site (HE);

b) a promoter sequence selected from the group consisting of AOX1, FLD1, FMD1, GAP and TEF1, wherein the promoter sequence is devoid of HE, BstBI, NcoI, Pmel and BstXI restriction sites if said restriction sites are present;

c) a BstBI restriction site;

d) a sequence encoding a secretion signal;

e) a NcoI restriction site;

f) a selection marker gene;

15

g) a Pmel restriction site;

h) a sequence encoding a detection and/or purification polypeptide;

i) a terminator sequence selected from the group consisting of AOX1, FLD1, FMD1, GAP and TEF1, wherein the terminator sequence is devoid of HE, BstBI, NcoI, Pmel and BstXI restriction sites if said restriction sites are present; and

j) a BstXI restriction site,

wherein the HE and BstXI sites are selected such that HE and BstXI result in compatible cohesive ends when cut by the homing endonuclease and the BstXI restriction enzyme, respectively, and the ligation product of HE and BstXI cohesive ends can neither be cleaved by the homing endonuclease nor the BstXI restriction enzyme.

2. The isolated MIE according to claim 1, wherein the promoter and terminator are AOX1 promoter and AOX1 terminator or, alternatively, wherein the secretion signal sequence is S. cerevisiae alpha-mating factor or, alternatively, wherein the promoter is AOX1 promoter, the terminator is AOX1 terminator and the secretion signal sequence is S. cerevisiae alpha-mating factor.

3. The isolated MIE according to any of the claims 1-2, wherein the selection marker gene is LacZ or, alternatively, wherein the detection or purification
polypeptide is selected from the group consisting of poly-histidine (6-20 histidine residues), FLAG-tag, StrepII-tag, HA epitope, c-Myc epitope, V5 epitope, calmodulin-binding peptide (CBP), enhanced green fluorescent protein (EGFP), enhanced yellow fluorescent protein (EYFP), Rudolph red fluorescent protein (RFP), monomeric Cherry fluorescent protein (mCherry), and cyan fluorescent protein (CFP) or, alternatively, wherein the selection marker gene is LacZ and the detection or purification polypeptide is selected from the group consisting of poly-histidine (6-20 histidine residues), FLAG-tag, StrepII-tag, HA epitope, c-Myc epitope, V5 epitope, calmodulin-binding peptide (CBP), enhanced green fluorescent protein (EGFP), enhanced yellow fluorescent protein (EYFP), Rudolph red fluorescent protein (RFP), monomeric Cherry fluorescent protein (mCherry), and cyan fluorescent protein (CFP).

4. The isolated MIE according to any of the claims 1-3, wherein the HE site is selected from the group consisting of I-CeuI site, a PI-ScI site, a I-ScI site, a I-Ppol site, a I-Crel site, and a I-Dmol site.

5. The isolated MIE according to any of the claims 1-4, wherein the HE/BstXI restriction sites are selected from the following HE/BstXI pairs: SEQ ID NO: 3/SEQ ID NO: 4, SEQ ID NO: 7/SEQ ID NO: 8, SEQ ID NO: 11/SEQ ID NO: 12, SEQ ID NO: 15/SEQ ID NO: 16, SEQ ID NO: 19/SEQ ID NO: 20, SEQ ID NO: 23/SEQ ID NO: 24, SEQ ID NO: 27/SEQ ID NO: 28, or their complementary sequences.

6. The isolated MIE according to any of the claims 1-5, which is one consisting of sequence SEQ ID NO: 31 or SEQ ID NO: 32.

7. An expression cassette derived from the MIE as defined in any of the claims 1-6 comprising one or more genes of interest.

8. A vector comprising the MIE according to any of the claims 1-6 or the expression cassette according to claim 7.

9. The vector according to claim 8, further comprising at least one of the sequence elements consisting of an antibiotic resistance marker, an origin of replication and from 1 to 6 sites for integration of the vector into another
vector or host cell.

10. A vector according to claim 9, wherein the site for integration is loxP.

11. The vector according to any of the claims 7-10, further comprising a PARS1 sequence.

12. The vector according to any of the claims 7-11 consisting of SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37 or SEQ ID NO: 38.

13. A host cell containing the MIE according to any of the claims 1-5, the expression cassette according to claim 6, or the vector according to any of the claims 7-12.

14. A kit for cloning and/or expressing proteins containing at least one vector according to any of the claims 9-12.

15. A method for producing multiple protein complexes comprising the steps of:

(i) producing a vector comprising from 1 to 10 expression cassettes as defined in claim 7
(ii) introducing the vector obtained in step (i) into a P. pastoris host cell; and
(iii) incubating the P. pastoris host cell under conditions allowing the simultaneous expression of the genes present in the vector.
FIG 3
FIG 5
FIG 6
FIG 7
FIG 9
FIG 10
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/00
ADD.

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

B. FIELDS SEARCHED

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

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

Electronic database consulted during the international search (name of database and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

[X] Further documents are listed in the continuation of Box C.  
[X] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"S" document member of the same patent family

Date of the actual completion of the international search
21 October 2016

Date of mailing of the international search report
09/11/2016

Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer
Young, Craig

Form PCT/ISA2/10 (second sheet) (April 2005)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>RAJ DEEPAK B THIMIRI GOVINDA ET AL: &quot;OminiBac: Universal Multigene Transfer Plasmids for Baculovirus Expression Vector Systems&quot;. STRUCTURAL GENOMICS: GENERAL APPLICATIONS HUMANA PRESS INC, 999 RIVERVIEW DR, STE 208, TOTOWA, NJ 07512-1165 USA SERIES: METHODS IN MOLECULAR BIOLOGY (ISSN 1064-3745(PRINT)), 2014, pages 123-130, XP009186414, the whole document</td>
<td>1-15</td>
</tr>
<tr>
<td>Y</td>
<td>CHRISTOPH BIENOSSEK ET AL: &quot;MultiBac: expanding the research toolbox for multiprotein complexes&quot;. TRENDS IN BIOCHEMICAL SCIENCES, vol. 37, no. 2, 1 February 2012 (2012-02-01), pages 49-57, XP05598236, ISSN: 0968-0004, DOI: 10.1016/J.TIBS.2011.10.005 the whole document</td>
<td>1-15</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>EP 2674438 A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP 2928290 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES 2554752 T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JP 2015519068 A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US 2015307899 A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WO 2014086981 A1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>