Recombinant expression of multiprotein complexes using polygenes

The present invention relates to a recombinant polynucleotide encoding, each within a single open reading frame (ORF), at least two polygenes each coding for at least three biologically active polypeptides wherein at least two of the polypeptides encoded by the genes constituting the polygenes are of non-viral origin, at least two of the polypeptides encoded by the genes constituting the polygenes are each capable of at least transiently interacting with at least one other polypeptide encoded by a gene of said polygenes, and the genes constituting each polygene are connected to one another by a sequence coding for at least one protease cleavage site and/or by a sequence coding for at least one self-cleaving peptide. Further embodiments of the present invention are a vector containing the recombinant polynucleotide, a host cell containing the recombinant polynucleotide and/or the vector and a non-human transgenic animal transformed with the recombinant polynucleotide and/or the vector. The present invention also relates to methods for the production of the polynucleotide and for the manufacture of multiprotein complexes. The embodiments of the present invention are particularly useful in gene therapy, drug candidate screening, vaccine production and crystallisation of multiprotein complexes for structural investigation.

An intense focus of biological research efforts in the post-genomic era is the elucidation of protein interaction networks (interactome). Since many of the identified multiprotein complexes are not present in sufficient quantities in their native cells for detailed molecular biological analysis, their study is dependent on recombinant technologies for large-scale heterologous protein production. Recombinant expression methods require a disproportionate investment in both labor and

CONFIRMATION COPY

materials prior to multiprotein expression, and subsequent to expression do not provide flexibility for rapidly altering the multiprotein components for revised expression studies.

There are several recombinant technologies that are currently used to obtain multisubunit complexes. Proteins can for example be expressed in isolation in *E. coli* either in soluble form or as inclusion bodies, purified and then reconstituted with similarly produced proteins *in vitro* into multiprotein complexes. Eukaryotic cells (e.g. mammalian or yeast cells) can also be used as hosts in transient expression experiments. This methodology is entirely dependent on the existence of an efficient *in vitro* reconstitution protocol. While this strategy may yield acceptable results for more simple systems with small subunit sizes, it is generally not applicable for more complicated multiprotein complexes containing many, and also large, subunits (e.g. close to all higher eukaryotic – in particular human – regulatory complexes).

Co-expression has been recognised as a superior alternative to the strategy of *in vitro* reconstitution as outlined above. Several co-expression systems have been developed in the past both for prokaryotic and eukaryotic expression. In prokaryotic systems, co-expression can be achieved by generating a single plasmid containing all genes of choice or by co-transforming several plasmids containing one or two genes and different resistance markers and replicons.

Co-expression in eukaryotic cells has been realised by using the baculovirus system, initially with limited success by co-infection with several viruses, and later and more successfully by expressing all proteins from a single virus, offering many advantages and eliminating several limitations present in prokaryotic systems (such as comparatively small subunit sizes, lack of authentic processing, difficult expression of eukaryotic (especially human) proteins etc.). For the baculovirus system, expression from a single virus has been shown to increase yields dramatically (Berger et al. (2004) *Nature Biotech.* 22, 1583-1587; see also Comment (2004), *Nature Biotech.* 22, vii, New & Views (2004), *Nature Biotech.* 22, 152, Research Highlights, *Nature Methods* 2, 7 (2005); Bertolotti-Ciarlet et al. (2003) *Vaccine* 21, 3885-3900), while decisively reducing the logistic demands especially for large scale production.

A major improvement of multiprotein expression was the provision of the modular system for the generation of multigene expression cassettes provided by the present inventors, which is disclosed in WO 2005/085456 A1 (PCT/EP2004/013381; see also Berger et al. (2004) *ibid.*). The MultiBac technology described in WO 2005/085456 A1 (PCT/EP2004/013381) enables the simple generation of multigene expression cassettes as well as modification and revision of expression experiments (Berger et al. (2004) *ibid.*).

However, a hindrance for successful expression and *in vivo* assembly of multisubunit complexes, in particular with many (6, 7, 8 or more), subunits (which constitute the majority of eukaryotic, e.g. human, gene regulatory complexes) is found in the fact that the relative expression levels of these subunits typically vary significantly based on in many cases not fully understood mechanisms (e.g. transcription and translation efficacy, protein stability, mRNA stability and secondary structures etc.). As a consequence, the subunit which is expressed in the least amount in an intrinsically unbalanced system will dictate the overall success of the multisubunit complex production experiment by limiting total complex yield. Accordingly, the transcription/translation machinery will produce excess amounts of other components which are not incorporated in the process thus "wasting" cellular transcription/translation resources. Individual expression levels typically vary several fold (e.g. up to 10 fold or more) with respect to each other, entailing losses which are refractory to a successful production of the desired multisubunit complexes, in particular in the case of complexes with more than 4, such as 5, 6, 8, 10 or more subunits (e.g. in the case of many eukaryotic gene regulatory complexes).

Viruses of the picornavirus super-group have a genome consisting of a single-stranded RNA molecule in sense orientation containing a single or two ORFs that code(s) for a polyprotein comprising the viral proteins which are connected to one another by cleavage sites of a viral protease or by self-cleaving peptides (reviewed, e.g., in Ryan et al. (1997) J. Gener. Virol. 78, 699-723).

The general concept of expressing a polyprotein through a recombinant virus for the production of protein complexes has been applied in the reconstitution of a TCR (T cell receptor):CD3 complex (Szymczak et al. (2004) Nature Biotech. 5, 589-594). The

authors used two recombinant retroviral vectors wherein one vector contained the sequences encoding the two TCR subunits whereas the other vector encoded a polyprotein comprising the four CD3 subunits. The subunits were connected by self-cleaving 2A peptide sequences derived from aphthoviruses. One disadvantage of this approach is that, in order to reconstitute the complete complex, two separate vectors must be prepared and two transfections are necessary.

The viral polyprotein approach has been applied in baculovirus expression systems for small constructs such as heterodimeric IL-12 (Kokuho et al. (1999) Vet. Immunol. Immunopathol. 72, 289-302) and fusion proteins comprising a nuclear targeting signal derived from baculoviral polyhedrin and a protein of interest (US 5,179,007).

Therefore, the technical problem underlying the present invention is to provide a new system for improved expression of multiple proteins.

The solution of the above technical problem is provided by the embodiments defined in the claims.

In particular, the present invention provides a polynucleotide encoding, each within a single open reading frame (ORF), at least two polygenes each coding for at least three biologically active polypeptides wherein at least two of the polypeptides encoded by the genes constituting the polygenes are of non-viral origin, at least two of the polypeptides encoded by the genes constituting the polygenes are each capable of at least transiently interacting with at least one other polypeptide encoded by a gene of said polygenes, and the genes constituting each polygene are connected to one another by a sequence coding for at least one protease cleavage site and/or by a sequence coding for at least one self-cleaving peptide.

The polynucleotide according to the present invention may be a DNA, RNA or a polynucleotide comprising one ore more synthetic nucleotide analogues. The polynucleotide may be present in single or double stranded form. DNA, in particular double-stranded DNA, forms are especially preferred. The polynucleotide of the present invention may be produced by chemical synthesis. Preferred polynucleotide

constructs of the present invention are made by recombinant gene technology (see, e.g., Sambrook et al. "Molecular Cloning", Cold Spring Harbor Laboratory, 1989).

A "polygene" as used herein is a nucleic acid sequence that encodes at least three biologically active polypeptides in a single ORF. Thus, each "gene" constituting the polygene is a nucleic acid sequence coding for a polypeptide, in particular a protein or fragment, variant, mutant or analogue thereof, having a specific, in particular structural, regulatory or enzymatic, function. Preferably the "gene" encoding the polypeptide comprises the coding region of a cDNA encoding the structural, regulatory or enzymatic protein or fragment, variant, mutant or analogue thereof.

A "fragment" of the polypeptide encoded by a gene contained in the polygenes means a part or region of the original polypeptide, preferably a fragment retaining at least one of the functions of the complete protein. A "variant" of the polypeptide encoded by a gene contained in the polygene means a polypeptide that is a functional or non-functional equivalent of the original polypeptide derived from another species or a functional or non-functional derivative of the original polypeptide that arises from alternative splicing or post-translational processing. A "mutant" of the polypeptide encoded by a gene contained in the polygene means a polypeptide that is derived from a naturally occurring protein by insertion, substitution, addition and/or deletion of one or more amino acid residues. An "analogue" of the polypeptide encoded by a gene contained in the polygene means functional equivalent of the original polypeptide that may even have a non-related amino acid sequence but exerts the same function as the polypeptide it is analogous to.

Correspondingly, on the nucleic acid level, a gene "fragment" is a part or region of the original gene the "fragment" is derived from. The gene "variant" has a sequence that is found in a different species compared to the original gene, or it may encode a splicing variant or post-translationally processed version of the polypeptide in question. The "mutant" is derived from the parent gene by insertion, substitution, addition and/or deletion of one or more nucleotides. The "analogue" of a gene encodes a functional equivalent of the polypeptide encoded by the parent gene.

At least two of the genes in the polygenes according to the present invention are of non-viral origin. "Non-viral" means that the nucleic acid sequence encoding the polypeptide (representing a functional protein or a fragment, variant, analogue or mutant thereof) is originally not found in or not derived from the genome of a virus. In particular, nucleotide sequences comprised in the polygene of the polynucleotide according to the present invention stem from eukaryotes and/or prokaryotes.

Thus, according to the present invention, the genes encoding the subunits (such as a multiprotein complex or members of a metabolic pathway or any other proteins that at least potentially interact at least transiently with one another) of a multisubunit assembly are present in at least two open reading frames (ORFs). The sequences encoding the subunits (polypeptides) of the assembly are present in at least two polygenes wherein the genes constituting each polygene are connected to one another by a sequence (there may be more than one) coding for a protease cleavage site (i.e. a sequence comprising the recognition site of a protease) or at least one self-cleaving peptide.

According to a preferred embodiment of the present invention the protease(s) capable of cleaving the cleavage sites encoded by the sequence(s) connecting the genes constituting the polygenes is/are encoded by the polynucleotide of the present invention. More preferably, the gene(s) encoding the protease(s) is/are part of at least one of the polygenes.

Suitable protease cleavages sites and self-cleaving peptides are known to the skilled person (see, e.g., in Ryan et al. (1997) J. Gener. Virol. 78, 699-722; Scymczak et al. (2004) Nature Biotech. 5, 589-594). Preferred examples of protease cleavage sites are the cleavage sites of potyvirus NIa proteases (e.g. tobacco etch virus protease), potyvirus HC proteases, potyvirus P1 (P35) proteases, byovirus NIa proteases, byovirus RNA-2-encoded proteases, aphthovirus L proteases, enterovirus 2A proteases, rhinovirus 2A proteases, picorna 3C proteases, comovirus 24K proteases, nepovirus 24K proteases, RTSV (rice tungro spherical virus) 3C-like protease, PYVF (parsnip yellow fleck virus) 3C-like protease, thrombin, factor Xa and enterokinase. Due to its high cleavage stringency, TEV (tobacco etch virus) protease cleavage sites are particularly preferred. Thus, the genes of the polygenes according to the

present invention are preferably connected by a stretch of nucleotides comprising a nucleotide sequence encoding an amino acid sequence of the general form EXXYXQ(G/S) wherein X represents any amino acid (cleavage by TEV occurs between Q and G or Q and S). Most preferred are linker nucleotide sequences coding for ENLYFQG and ENLYFQS, respectively.

Preferred self-cleaving peptides (also called "cis-acting hydrolytic elements", CHYSEL; see deFelipe (2002) Curr. Gene Ther. 2, 355-378) are derived from potyvirus and cardiovirus 2A peptides. Especially preferred self-cleaving peptides are selected from 2A peptides derived from FMDV (foot-and-mouth disease virus), equine rhinitis A virus, *Thosea asigna* virus and porcine teschovirus.

At least two of the polypeptides encoded by the polygenes of the present invention are capable of at least transiently interacting with one other polypeptide encoded by the polygenes, or they are at least suspected to be capable of at least transiently interacting with another polypeptide encoded by a gene contained in the polygenes. Typical "interactions" formed between the polypeptides include covalent binding, hydrogen bonds, electrostatic interactions and Van-der-Waals interactions. "Transient" interactions are common to biomolecules, in particular proteins, and are typically represented by interactions between enzymes and their substrates, receptors and their (agonistic or antagonistic) ligands, interactions between members of metabolic pathways and interactions between proteins of regulatory (e.g. gene regulatory) complexes.

The polypeptides encoded by the nucleotide sequences constituting the polygenes of the present invention may be the same or different. Thus, each polygene present in the constructs of the invention may contain one or more copy of each nucleotide sequence encoding a protein of interest. In this manner it is, e.g., possible to provide constructs that serve for optimal expression of the desired proteins, in particular in case proteins are normally expressed at different levels and/or are present in a macromolecular assembly in different stoichiometries. Therefore, in case a polypeptide is poorly expressed in commonly used systems, two or more copies of the corresponding coding sequence may be integrated into one or more polygene(s) of an inventive construct. The same approach may be used, in case a polypeptide is

present as a dimer, trimer or multimer in a desired complex. In this manner, the constructs of the present invention may be assembled individually according to the requirements (expression levels, stoichiometry etc.) of any complex or other macromolecular assembly a person skilled in the art desires to express and/or to purify.

It is further preferred that the genes constituting the polygenes are selected from the group consisting of genes encoding members of multiprotein complexes and genes encoding members of metabolic pathways. Preferred multiprotein complexes are gene regulatory protein complexes such as transcription factor complexes, transport complexes such as complexes involved in nuclear and/or cellular transport, protein folding complexes, receptor/ligand complexes, cell-cell recognition complexes, complexes involved in apoptosis, complexes involved in cell cycle regulation etc. Members of metabolic pathways are, e.g. members of carbohydrate metabolism (such as glycolysis, gluconeogenesis, citric acid cycle, glycogen biosynthesis, galactose pathway, calvin cycle etc.), lipid metabolism (such as triacylglycerol metabolism, activation of fatty acids, β-oxidation of fatty acids (even chain/odd chain), α-oxidation pathway, fatty acid biosynthesis, cholesterol biosynthesis etc.), amino acid metabolism such as glutamate reactions, Krebs-Henseleit urea cycle, shikimate pathway, Phe and Tyr biosynthesis, Trp biosynthesis etc.), energy metabolism (such as oxidative phosphorylation, ATP synthesis, photosynthesis, methane metabolism etc.) nucleic acid metabolism (purin and pyrimidine biosynthesis and degradation, DNA replication etc.). Members of multiprotein complexes and members of metabolic pathway may be taken from, e.g. http://www.biocarta.com/genes/index.asp and G. Michal (ed.) Biochemical Pathways, 1. edition, John Wiley & Sons, Hoboken, NJ, USA, 1999, the disclosure content of which is hereby incorporated by reference.

Each polygene according to the present invention contains at least 3 genes, i.e. sequences encoding a biologically active polypeptide. More preferred are polygenes encoding 4, 5 6 or more or even more proteins. As mentioned above, it is preferred that the protease(s) capable of cleaving the protease cleavage sites connecting the polypeptides is/are encoded by at least one of the polygenes.

According to a preferred embodiment, the polynucleotide of the present invention contains at least two promoter sequences which are each operatively linked to one of the polygenes, thus capable of controlling the expression of the polygenes. Suitable promoters in the constructs of the present invention may be selected from the group consisting of polh, p10 and pXIV very late baculoviral promoters, vp39 baculoviral late promoter, vp39polh baculoviral late/very late promoter, $P_{cap/polh}$, *pcna*, *etl*, *p35*, *da26* baculoviral early promoters; CMV, SV40, UbC, EF-1α, RSVLTR, MT, P_{DS47}, Ac5, P_{GAL} and P_{ADH}. The promoter sequences may be the same for all polygenes, or different promoters may be selected for the different polygenes.

Preferably, the each ORF containing a polygene of the present invention is flanked by a terminator sequence such as SV40, HSVtk or BGH (bovine growth hormone).

The polynucleotide according the present invention may contain further regulatory sequences such as enhancers or suppressor sequences.

It is further preferred that the polynucleotide according to the present invention contains at least one site for its integration into a vector or host cell. Such an integration site will allow for the convenient genomic or transient incorporation of the polynucleotide into vectors (such as virus) and host cells (e.g. eukaryotic host cells), respectively. Sites for genomic integration are more preferred.

Especially preferred integration sites are those which are compatible for the polynucleotide's integration into a virus. More preferably, the integration site is compatible for the polynucleotide's integration into a virus selected from the group consisting of adenovirus, adeno-associated virus (AAV), autonomous parvovirus, herpes simple virus (HSV), retrovirus, rhadinovirus, Epstein-Barr virus, lentivirus, semliki forest virus and baculovirus.

In a further preferred embodiment, the integration site is compatible for the polynucleotide's integration into a eukaryotic host cell which may preferably be selected from the group consisting of mammalian (such as human cells, e.g. HeLa, Huh7, HEK293, HepG2, KATO-III, IMR32, MT-2, pancreatic β cells, keratinocytes, bone-marrow fibroblasts, CHP212, primary neural cells, W12, SK-N-MC, Saos-2,

WI38, primary hepatocytes, FLC3, 143TK-, DLD-1, umbilical vein cells, embryonic lung fibroblasts, primary foreskin fibroblasts, osteosarcoma cells, MRC5, MG63 cells etc.), porcine (such as CPL, FS-13, PK-15) cells, bovine (such as MDB, BT) cells, ovine (such as FLL-YFT) cells, *C. elegans* cells, yeast (such as *S. cerevisiae, S. pombe, C. albicans, P. pastoris*) cells, and insect cells (such as *S. frugiperda*, e.g. Sf9, Sf21, Express Sf+, High Five H5 cells, *D. melanogaster*, e.g. S2 Schneider cells).

Particularly preferred integration sites are selected from the transposon elements of Tn7, λ integrase-specific attachment sites and SSRs (site specific recombinases), preferably the cre-lox specific (LoxP) site or the FLP recombinase specific recombination (FRT) site.

In a preferred embodiment of the present invention the polynucleotide additionally comprises one or more resistance markers for selecting host cells with desired properties based on a resistance to otherwise toxic substances. Examples of suitable resistance markers are those providing resistance against ampicillin, chloramphenicol, gentamycin, spectinomycin and/or kanamycin.

For its incorporation into a prokaryotic host cell, the polynucleotide of the present invention preferably comprises a conditional R6Kγ origin of replication for making propagation dependent on the *pir* gene in a prokaryotic host.

Especially preferred embodiments of the polynucleotide of the present invention result by the insertion of the polygenes as expression cassettes into constructs disclosed in WO 2005/085456 A1 (PCT/EP2004/013381).

Therefore, it is preferred that the polynucleotide of the present invention comprises a functional arrangement according to the following Formula I

$$X - \underline{T1 - MCS1 - P1} - [A - B] - \underline{P2 - MCS2 - T2} - Y \qquad (I)$$

comprising

(a) at least two expression cassettes $\underline{T1 - MCS1 - P1}$ and $\underline{P2 - MCS1 - T2}$ in a head-to-head, head-to-tail or tail-to-tail arrangement, each comprising a multiple cloning site MCS1 or MCS2, flanked by a promoter P1 and a

terminator sequence T1 for MCS1 and flanked by a promoter P2 and a terminator sequence T2 for MCS2

(b) at least one multiplication module M in between the promoters P1 and P2 comprising at least two restriction sites A and B

(c) at least two restriction sites X and Y each flanking one of the expression cassettes,

wherein

(i) restriction sites A and X as well as B and Y are compatible, but

(ii) the ligation products of AY and BX are not enzymatically cleavable by restriction enzymes a, b, x or y specific for restriction sites A, B, X and Y, and

(iii) restriction sites A and B as well as restriction sites X and Y are incompatible,

wherein each polygene is inserted into one of the expression cassettes.

With respect to further preferred embodiments of the constructs having the arrangement of Formula (I) it is expressly referred to WO 2005/085456 A1 (PCT/EP2004/013381). In particular, restriction sites A and B in the multiplication module M are selected from the group consisting of restriction sites BstZ17I, SpeI, ClaI and NruI or restriction sites cleaved by isoschizomers thereof. Isoschizomers are restriction enzymes that have identical cleavage sites. Furthermore, preferred examples of the restriction sites X and Y are restriction sites selected from the group consisting of PmeI and AvrII or restriction sites cleaved by isoschizomers thereof.

Particularly preferred polynucleotides having each polygene inserted into one of the expression cassettes contained in the above formula (I) comprise the following features:

(a) promoters P1 and P2 are selected from the group consisting of polh and p10;

(b) terminator sequences are selected from the group consisting of SV40 and HSVtk;

(c) restriction sites A and B in the multiplication module M are selected from group consisting of restriction sites BstZ171I, SpeI, ClaI and NruI;

(d) restriction sites X and Y are selected from the group consisting of restriction sites PmeI and AvrII; and

(e) sites for virus integration are selected from the group consisting of cre-lox and Tn7.

With respect to the production of polynucleotides having the above arrangement according to Formula (I) it is expressly referred to WO 2005/085456 A1 (PCT/EP2004/013381).

The provision of polynucleotides of the present invention encoding several biologically active polypeptides within two or more ORFs each containing a polygene provides a major improvement with respect to the cloning and expression of genes coding for members of multisubunit protein complexes: On the one hand, assembly of all subunit genes into a single ORF is often impossible or highly difficult because of the huge size and numbers of coding sequences to be coupled. On the other hand, efficient assembly of several or all members of a multisubunit complex each present in separate expression cassettes has often turned out to be highly inefficient, since the overall complex yield is determined by the least expressed subunit. According to the present invention the subunits of a multiprotein assembly are encoded by at least two polygenes (each representing a single ORF) each coding for at least three polypeptides (preferably of non-viral origin) which results in an optimal compromise between manageability of the construct and its constituents (in particular assembly of the polygenes) and expression efficiency (in particular in case the polynucleotide of the present invention is present in a suitable vector).

Therefore, a further embodiment of the present invention is a vector containing the above-described polynucleotide. The vector may be selected from the group consisting of plasmids, expression vectors and transfer vectors. More preferably, the vector of the present invention is useful for eukaryotic gene transfer, transient or viral vector-mediated gene transfer.

Especially preferred vectors are eukaryotic expression vectors such as viruses selected from adenovirus, adeno-associated virus (AAV), autonomous parvovirus, herpes simple virus (HSV), retrovirus, rhadinovirus, Epstein-Barr virus, lentivirus, semliki forest virus and baculovirus. Most preferred vectors of the present invention are baculovirus expression vectors. Preferred baculovirus of the present invention are embodiments wherein the genes *v-cath* and *chiA* are functionally disrupted, since this leads to improved maintenance of cellular compartments during infection and

protein expression. The *v-cath* gene encodes the viral protease V-CATH which is activated by upon cell death by a process dependent on a juxtaposed gene on the viral DNA, *chiA*, which codes for a chitinase. Both genes are preferably disrupted to eliminate V-CATH activity and to gain the option of utilising chitin affinity chromatography without interference form the *chiA* gene product. The quality of the expression products generated by a baculovirus system lacking functionally active *v-cath* and *chiA* genes is significantly improved because of the reduction of viral-dependent proteolytic activity and cell lysis.

Preferably, vectors according to the present invention comprise a site for SSRs, preferably LoxP for cre-lox site specific recombination. More preferably, the cre-lox site is located in one or both of the baculoviral gene v-cath and chiA so as to disrupt their function.

The vector of the present invention preferably contains one ore more marker genes for selection of hosts successfully transfected with the correctly assembled vector. Examples of suitable marker genes are luciferase, β-Gal, CAT, genes encoding fluorescent proteins such as GFP, BFP, YFP, CFP and variants thereof, and the lacZα gene. The marker gene(s) may be functionally equivalent variants, mutants, fragments or analogues of the mentioned examples or other suitable markers known to the skilled person. Variants, mutants or analogues preferably show a homology of at least 75%, more preferably 85%, especially preferred 90 %, in particular at least 95% on the amino acid level in comparison to the marker said variant, mutant or analogue is derived from.

In another preferred embodiment the vector of the present invention comprises a transposon element, preferably the Tn7 attachment site. More preferably, such a transposon element, e.g. the Tn7 attachment site, is located within a marker gene such that a successful integration by transposition can be assessed by testing the phenotype provided by the functional marker gene.

Preferred transfer vectors of the present invention are based on pFBDM or pUCDM as disclosed in WO 2005/085456 A1 (PCT/EP2004/13381; see SEQ ID NO: 1 and 2 as well as Fig. 1 and 2, respectively disclosed therein). Further preferred transfer

vectors of the present invention are based on derivatives of the above pFBDM and pUCDM, respectively, vectors.

Examples of particularly preferred derivatives of pFBDM and pUCDM are transfer vectors pSPL (Fig. 3), pFL (Fig. 4), pKL (Fig. 5) and pKDM (Fig. 6). Like pUCDM, pSPL contains a conditional origin of replication (R6Kγ). pFL (like pFBDM) contains a high copy-number replication origin (ColE1). pKDM and pKL have low-copy replication origins derived from pBR322. In analogy to pFBDM, pFL, pKL and pKDM contain transposon elements (Tn7R, Tn7L). Vectors pSPL, pFL and pKL have a LoxP imperfect inverted repeat flanking the dual expression cassette (as does pUCDM). All vectors contain the above-described multiplication module (M) for generating multigene cassettes1. pFL and pKL (and derivatives) are acceptor vectors, pUCDM and pSPL (and derivatives) are donor vectors in Cre-mediated plasmid fusions.

Important features of the above preferred examples of transfer vectors for generating the constructs of the present invention are summarised in the following Table 1.

Tab. 1: Features of preferred transfer vectors

Vector	Antibiotic resistance marker	Replicon (source)	Host strain	Recombination and multiplication elements	Usage
pFBDM	Ampicillin Gentamycin	ColE1	TOP10*	Tn7L, Tn7R, multiplication module M	integration in MultiBac ** Tn7 site
pFL	Ampicillin Gentamycin	ColE1	TOP10	Tn7L, Tn7R, LoxP, multiplication module M	acceptor for plasmid fusions; integration in MultiBac ** Tn7 site
pKDM	Kanamycin Gentamycin	pBR322	TOP10	Tn7L, Tn7R, multiplication module M	integration in MultiBac ** Tn7 site
pKL	Kanamycin Gentamycin	pBR322	TOP10	Tn7L, Tn7R, LoxP, multiplication module M	acceptor for plasmid fusions; integration in MultiBac ** Tn7 site
pUCDM	Chloramphenicol	R6kγ	BW23473	LoxP, multiplication module M	donor for plasmid fusions; integration in MultiBac ** LoxP site
pSPL	Spectinomycin	R6Kγ	BW23473	LoxP, multiplication module M	donor for plasmid fusions; integration in MultiBac ** LoxP site

* or any other general laboratory cloning strain (*recA⁻ endA⁻ pir⁻*)

** see WO 2005/085456 A1

Therefore, the polygenes of the present invention are inserted into a vector such as pFBDM, pUCDM, pSPL, pFL, pKL or pKDM at the multiple cloning sites (MCS1 and MCS2), either by restriction enzyme cleavage and ligation or via recombination (e.g. using the BD In-Fusion enzyme). The baculovirus transfer vectors pFBDM, pUCDM, pSPL, pFL, pKL and pKDM comprise modified recipient baculovirus DNA engineered for improved protein production and allow for a simple and rapid method to integrate genes via two access sites (attTn7 and LoxP) into this baculoviral DNA in *E. coli* cells tailored for this purpose.

According to a further embodiment the present invention provides a host cell containing the polynucleotide and/or the vector of the invention.

Examples of preferred host cells are mammalian cells, such as human, rodent, porcine cells such as CPL, FS-13 and PK-15, bovine cells such as MDB and BT, ovine cells such as FLL-YFT, *C. elegans* cells, yeast cells such as *S. cerevisiae, S. pombe, P. pastoris* and *C. albicans*, insect cells such as cells from *S. frugiperda*, preferably Sf9, Sf21, Express Sf+ or High Five h5 cells, cells from *D. melanogaster* such as S2 Schneider cells, and bacteria such as *E. coli*, preferably strains Top10, Dh5α, DH10α, HB101, TG1, BW23473 and BW23474.

Preferred human cells are selected from HeLa, Huh7, HEK293, HepG2, KATO-III, IMR32, MT-2, pancreatic β cells, keratinocytes, bone-marrow fibroblasts, CHP212, primary neural cells, W12, SK-N-MC, Saos-2, WI38, primary hepatocytes, FLC3, 143TK-, DLD-1, umbilical vein cells, embryonic lung fibroblasts, primary foreskin fibroblasts, osteosarcoma cells, MRC5 and MG63 cells.

Host cells comprising a polynucleotide and/or vector according to the invention may be isolated cells or they may be present in tissues or organs.

A further embodiment of the present invention relates to a non-human transgenic animal being transformed with at least one polynucleotide sequence and/or vector of the invention. Preferred transgenic animals are rodent, porcine, bovine and *C. elegans* species.

The transgenic animal of the present invention is particularly useful for the elucidation of the role of multiprotein complexes or for screening of compounds for their biological activities *in vivo*.

A further embodiment of the present invention is a method for the production of the polynucleotide as defined above comprising the steps of:

(a) providing, preferably amplifying, the coding regions of the genes constituting the at least two polygenes;

(b) providing said coding regions with the sequences coding for the at least one protease cleavage site and/or the at least one self-cleaving peptide; and

(c) assembling the fragments resulting from steps (a) and (b) such that a single ORF results in each polygene; and

(d) combining the at least two polygenes into a single polynucleotide.

Another aspect of the invention is a method for the production of the vector according to the present invention comprising the steps of

(a) generating at least two polygenes each comprising at least three genes within a single ORF as defined above (preferably by the above method for the production of the polynucleotide of the present invention); and

(b) cloning the polygenes into a plasmid or viral vector,

wherein at least one of the genes of each polygene is of non-viral origin and at least two of the polypeptide encoded by the genes are capable of at least transiently interacting with one other polypeptide encoded by the genes.

Preferably, one of the genes assembled into the polygenes is a gene encoding a protease capable of cleaving the protease cleavage sites connecting the polypeptides encoded by the polygenes. Preferred proteases are as defined above.

The construction of the polygenes as well as the vectors of the present invention can be carried out through various molecular biological techniques which are generally known to a person skilled in the art (see, e.g., Ausubel et al. (eds.) Current Protocols in Molecular Biology, John Wiley & Sons, Hoboken, NJ, USA, 2003). The production of a polygene may be carried out, e.g. by PCR amplification of the nucleotide sequences coding for the particular polypeptides (e.g. using corresponding cDNA

templates), preferably by usage of a primer (either 5' or 3') providing the sequence(s) coding for the at least one protease cleavage site and/or the at least one self-cleaving peptide. Preferably, the primers further contain a recognition sequence of a suitable restriction enzyme. Preferably, each primer contains a restriction site that is different from the restriction site of the other primer such that a directional ligation of the resulting amplification product with another amplification product and/or a linearised vector containing the same restriction sites is possible. According to another preferred embodiment for the production of constructs by directional ligation, the primers may contain a recognition sequence of a restriction enzyme that produces an overhang which is not self-ligatable such as RsrII or BstEII. In case the primers themselves do not contain a restriction site, it is also possible to provide the amplification products with adapters comprising the desired restriction site(s). Of course, it is also possible to provide the coding regions for the desired polypeptides using any source (besides amplification). For example, the required sequences may be already present as such or in corresponding vectors from which the sequences may be cut out by appropriate restriction enzymes. Constructs that do not contain appropriate restriction sites etc. may be provided with the appropriate sequences (restriction site(s), protease cleavage site(s)/self-cleavable peptide(s), linkers etc.) by ligation of suitable adapters containing the required elements. The amplification products or any other appropriate construct (containing the restriction sites) are then cut with the appropriate restriction enzymes. Of course, the sequences of any primers used are preferably selected such that, after final assembly of the polygenes, preferably into a suitable vector, a single ORF results for each polygene.

According to a preferred embodiment of the invention, the amplification products or other appropriate sequences may then be ligated sequentially or simultaneously into a suitable vector, e.g. at MCS1 or MCS2 of the MultiBac vector system referred to above. For example, one polygene may be introduced into pFDBM and another polygene may be ligated into pUCDM (as described in WO 2005/085456 A1 (PCT/EP2004/013381)). The resulting constructs are then used for the production of corresponding bacmids by *cre-lox* site-specific recombination (pUCDM derivative) and Tn7 specific transposition (pFBDM derivative) yielding a baculoviral expression vector ready for infection of corresponding insect cells for expression and purification of the multiprotein complex of interest.

Besides producing the polygenes and introducing these into a suitable vector by restriction/ligation it is also possible to assemble such constructs by homologous recombination using appropriate recombinases. Suitable examples of recombinase-based cloning techniques are the In-Fusion® system available from BD Biosciences Clontech, Heidelberg, Germany (see *Clontechniques*, October 2002, p. 10) and the Red®/ET® recombination system available from Gene Bridges GmbH, Dresden, Germany (see WO-A-99/29837; http://www.genebridges.com).

The In-Fusion® system requires 15 bp homologous regions in a DNA molecule to be fused into a linear construct (such as an appropriate vector) having the corresponding homologous regions. Accordingly, a vector containing a polygene of the present invention may assembled sequentially or simultaneously by providing the constituting coding regions (together with the linker sequences coding for a protease cleavage site and/or a self-cleaving peptide) with appropriate homologous sequences of 15 bp. If desired, the 15 bp homologous sequences may be selected such that the constituting coding regions are assembled in a desired order. Of course, appropriate homology regions may be introduced by PCR amplification of the corresponding fragments using primers containing the desired sequences.

The Red®/ET® recombination system is different from the In-Fusion® system in that homology sequences of 40 to 60 base pairs are required, but the construct in which a fragment is to be inserted needs not to be linear. The recombination is carried out *in vivo* in a host, preferably an *E. coli* strain, expressing the dual recombinase system "ET" (RecE/RecT or Redα/Redβ). Thus, the fragments constituting the polygenes may be directly transformed together with an appropriate vector into the appropriate host, preferably *E. coli* cells. In this manner, each polygene may be assembled in the suitable vector either sequentially fragment by fragment (preferably comprising the coding region for each member polypeptide of the multiprotein complex + at least one protease cleavage site sequence/self-cleaving peptide sequence), or by simultaneous transformation of all fragments.

Especially preferred polynucleotide constructs of the present invention contain polygenes of at least similar length, since the inventors have found that the

expression of corresponding polyproteins from such constructs results in comparable expression levels. According to the present invention "polygenes of similar length" means that the lengths of the nucleotide sequences of the polygenes differ from one another by not more than 50 %, more preferably not more than 30 %, in particular not more than 20 % or even less.

Furthermore, the present invention provides a method for the production of multiprotein complexes *in vitro* comprising the steps of

(a) cultivating the host cell according to the present invention in a suitable medium under conditions allowing the expression of the polygenes; and

(b) recovering the expression products encoded by the polygenes from the medium and/or the host cells.

The present invention also relates to a method for the production of multiprotein complexes *in vivo* comprising the steps of

(a) generating at least two polygenes each comprising at least three genes within a single ORF as defined above; and

(b) transforming the polygenes into an animal such that the polygenes are expressed in said animal.

Preferably the transformation of the animal with the polygenes according to step (b) is effected by means of a vector, in particular a viral vector, more preferably a baculovirus vector. Baculoviruses are especially useful vehicles for delivery of polygenes into mammalian species. The above *in vivo* method is preferably carried out in mammals, *C. elegans* or insects. Particularly preferred examples of suitable animal species are defined herein above.

The embodiments of the present invention are also useful for the preparation of vaccines directed against multisubunit assemblies of proteins. Complexes of multiple subunits often display different epitopes compared to the individual proteins constituting the complexes. Therefore, the multiprotein complexes produced according to the present invention display the naturally occurring relevant epitopes in a more appropriate fashion, thus providing better antigen targets for antibody production.

Recently, virion-like particles (VLPs) consisting of four proteins from the sever acute respiratory syndrome (SARS) coronavirus were made using a recombinant baculovirus expression vector (cf. Mortola et al. (2004), *FEBS Lett.* 576, 174-178). The effective expression of such infectious particles for the preparation of vaccines will be greatly facilitated using the polygene expression system according to the present invention. In particular, the high-yield expression of multisubunit assemblies that contain substantially more polypeptide than the example of SARS-VLPs is made available by the expression tools of the present invention.

Therefore, the present invention further relates to a method for the production of a vaccine comprising the steps of

(a) administering at least one polynucleotide and/or vector of the present invention to a mammal, whereby the polygene of the invention is expressed within the mammal;

(b) optionally administering an adjuvans to the mammal; and

(c) optionally isolating the antibodies and/or spleen cells producing antibodies specific for at least one of the polypeptide encoded by the polygenes.

The present invention provides a convenient and simple approach for the recombinant production of multiprotein assemblies. These multisubunit assemblies may be tested for protein complex interactions or modifications of the proteins constituting the multisubunit assembly. The multisubunit assemblies produced according to the present invention may also be assayed for their interaction with candidate compounds (small organic molecules, nucleic acids, peptides, polypeptides etc.) that may exert a biologically significant activity being of medical value.

Therefore, the present invention is also directed to a method for assaying protein complex interactions or protein modifications.

According to a preferred embodiment, the present invention provides a method for the screening of protein complex interactions or modifications of multiprotein complexes *in vitro* comprising the steps of

(a) providing a host cell according to the present invention containing at least two polygenes;

(b) maintaining the host cell under conditions that allow expression of the polygenes; and

(c) detecting interactions between or modifications of the polypeptides encoded by the polygenes.

Another preferred embodiment of the present invention is a method for *in vitro* screening of candidate compounds capable of (i) interacting with a multiprotein complex and/or (ii) modification of proteins within a multiprotein complex and/or (iii) inhibiting interactions within or between multiprotein complexes and/or inhibiting modifications of proteins within a multiprotein complex, comprising the steps of

(a) providing a host cell according to the present invention containing at least two polygenes;

(b) maintaining the host cells under conditions that allow expression of the polygenes;

(c) contacting a candidate compound with the host cell; and

(d) detecting interactions of the expression products with the candidate compound and/or interactions between the expression products and/or modifications of the expression products and/or inhibition of interactions between the expression products.

The polynucleotides and/or vectors of the present invention are also suitable for the screening of protein – protein, protein – (multi)protein complex or multiprotein complex – multiprotein complex interactions or modifications (phosphorylation, glycosylation etc.) of multiprotein complexes *in vivo*.

Thus, a further preferred embodiment of the present invention is a method for *in vivo* screening of candidate compounds capable of (i) interacting with a multiprotein complex and/or (ii) modification of proteins within a multiprotein complex and/or (iii) inhibiting interactions within or between multiprotein complexes and/or inhibiting modifications of (a) proteins within a multiprotein complex, comprising the steps of

(a) providing an animal comprising at least one polynucleotide and/or vector of the invention containing at least two polygenes as defined above, whereby the polygenes are expressed in the animal;

(b) administering a candidate compound to the animal; and

(c) detecting interactions of the expression products with the candidate compound and/or interactions between the expression products and/or modifications of the expression products and/or inhibition of interactions between the expression products.

The multiprotein expression tools of the present invention are also of medical use. In particular, bioactive multiprotein complexes as well as medically advantageous combinations of proteins, e.g. antibody mixtures, optionally in combination with interleukins and/or adjuvans can be administered to an animal or human by means of the polynucleotides and/or the gene delivery vectors of the present invention.

Accordingly, the present invention further relates to the use of the polynucleotide and/or the vector and/or the host cell described above for the preparation of a medicament comprising a polygene transfer vehicle for gene therapy.

Tremendous efforts are being made to develop gene delivery systems for therapeutic applications. Gene therapy has been the focus of intense enthusiasm but also criticism in the past. To date, major progress has been made in evaluating gene therapy in clinical trials on the way to achieving safe and applicable clinical *in vivo* and *ex vivo* strategies for human diseases (see Worgall S. (2004) *Peadiatr. Nephrol.*). Overall, gene therapy now stands as a very promising avenue for the correction of genetic as well as acquired disorders entailing permanent or transient expression of a therapeutic gene product (Worgall S., *ibid.*). Recombinant vectors based on virus, in particular those that are not replication competent in mammalian hosts (e.g. baculoviral vectors) have emerged recently as a powerful tool for mammalian cell gene delivery and have been successfully applied to a whole range of mammalian cell lines including human, primate, rodent, bovine, procine and ovine cells (reviewed in Kost and Condreay (2002) *Trends Biotech.* 20, 173-180). To obtain complex gene transfer/therapy effects, both *ex vivo* and *in vivo*, an increasing demand has arisen for polycistronic viral vectors to accomplish more powerful results

rather by combined gene therapy than by single gene therapy (de Felipe (2002), *Curr. Gene Ther.* 2, 355-378; Planelles (2003) *Meth. Mol. Biol.* 229, 273-284). The requirement for the incorporation of accessory genes into a carrier virus that is to be administered *in vivo*, e.g. to block inactivation by the complement system, has also been demonstrated by using a pseudotyped baculovirus with baculoviral gp64 envelope proteins that carried a human decay-accelerating factor protein domain fusion (Hueser et al. (2001) *Nat. Biotech.* 19, 451-455), exemplifying the necessity to provide recombinant modifications on the virus production level in addition to the multiple genes to be transferred for therapeutic purposes.

Accordingly, recombinant baculovirus of the present invention are preferred for preparing gene therapeutic medicaments. More preferably, the vector used for the medicament of the present invention is a baculovirus comprising at least two polygenes as defined above encoding

(i) one or more therapeutic polypeptide(s) and

(ii) one or more baculoviral proteins

In a preferred embodiment, the protein(s) according to (ii) are humanised baculoviral proteins expressed from pseudotyped baculovirus, preferably a humanised baculovirus envelope protein gp64, e.g. gp64 fused with a human protein such as for example decay accelerating factor.

Furthermore, the present invention relates to an *in vivo* gene therapeutic method comprising the steps of

(a) providing a polygene transfer vehicle comprising a polynucleotide according to the invention; and

(b) administering the polygene transfer vehicle to a patient suffering from a genetic disorder.

The present invention further provides an *ex vivo* gene therapeutic method comprising the steps of

(a) collecting cells of a patient suffering from a genetic disorder;

(b) transforming the collected cells with a polygene transfer vehicle comprising a polynucleotide according to the present invention; and

(c) administering the transformed cells to the patient.

The multiprotein complexes produced according to the present invention may advantageously be used in biophysical studies, in particular structural studies using crystallographical, electron-microscopical and/or NMR techniques, protein chemical studies, in particular for protein-protein interactions, and for drug development.

Thus, the present invention is directed to the use of the polynucleotide and/or the vector and/or the host cell of the present invention for the crystallisation of multiprotein complexes.

A further embodiment of the present invention is a kit for the preparation of multiple-protein complexes comprising

(a) primers for PCR amplification of the coding sequences constituting the polygenes;

(b) a plasmid or viral vector; and

(c) optionally host cells suitable for the propagation of the plasmid or vector

The primers are conveniently designed to match the needs for producing a single ORF for each polygene and may contain restriction sites for ligation (sequentially or simultaneously) into the plasmid or viral vector and/or the primers may contain sequences for assembling the polygenes and/or insertion into the plasmid or viral vector by homologous recombination (e.g. using the In-Fusion® system or Red®/ET® system as described above).

The Figures show:

Fig. 1 shows the nucleotide sequence (SEQ ID NO: 1) and deduced amino acid sequence (SEQ ID NO: 2) of a PCR product coding for human TATA-Box-Binding Protein (hTPB) core (hTBPc, c-terminal fragment of the full-length protein truncated at position 159). Positions of RsrII restriction sites (present in the primer sequences) are indicated.

Fig. 2 shows photographs of agarose gel electrophoretic analyses of *in vitro* ligation of hTBPc gene segments and subcloning of the mixture. The PCR-amplified hTPBc gene was digested by RsrII and purified (lane 1). Incubation with ligase yields a ladder of concatamers containing 1,2,3 and more genes linked in one ORF each (lane 2, lane 3 is MBI DNA Marker 1kb ladder). Subcloning of the mixture of the thus-yielded expression constructs containing one polygene each with differing numbers of linked hTBPc genes that can be liberated by restriction digest using RsrII (lanes 4-7). Digestion outside of the inserted polygene evidences 1 (lane 8), 2 (lane 9), 3 (lane 10) and 5 (lane 11) hTBPc genes that yielded a single ORF in each case.

Fig. 3 shows a schematic representation of the basic transfer vector pSPL underlying preferred transfer vector constructs of the present invention.

Fig. 4 shows a schematic representation of the basic transfer vector pFL underlying preferred transfer vector constructs of the present invention.

Fig. 5 shows a schematic representation of the basic transfer vector pKL underlying preferred transfer vector constructs of the present invention.

Fig. 6 shows a schematic representation of the basic transfer vector pKDM underlying preferred transfer vector constructs of the present invention.

Fig. 7 shows a schematic representation of the transfer vector construct pFBDO[hTBPc]3.

Fig. 8 shows the nucleotide sequence of pFBDO[hTBPc]3 (SEQ ID NO: 3).

Fig. 9 shows a schematic representation of the transfer vector construct pUCDMCSTAF1TBPcTAF2.

Fig. 10 shows the nucleotide sequence of pUCDMCSTAF1TBPcTAF2 (SEQ ID NO: 4).

Fig. 11 shows a schematic representation of the transfer vector construct pFBDO[HisTEVTAF6TAF9]his.

Fig. 12 shows the nucleotide sequence of pFBDO[HisTEVTAF6TAF9]his (SEQ ID NO: 5).

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

EXAMPLES

Example 1: Production of polygenes and ligation into expression vectors

The principle of generating polygenes is shown here by using human TATA-Box-Binding protein (hTBP) core (hTBPc, c-terminal fragment of the full-length protein truncated at position 159). The gene encoding hTBPc was amplified by polymerase chain reaction (PCR) using a sense primer annealing to the 5' end of the gene containing an overhang possessing a RsrII restriction site and further encoding an amino acid spacer and a Tobacco-Etch-Virus (TEV) cleavage site. The antisense primer annealed to the 3' terminus of the gene and contained an RsrII restriction site. RsrII is a restriction enzyme that produces an asymmetric overhang of 3 nucleotides which do not self ligate, therefore, the restriction product is asymmetric and ligation yields a directional product. The PCR product was digested with RsrII and purified. The DNA (SEQ ID NO: 1) and deduced amino acid sequence (SEQ ID NO: 2) of the PCR product are shown in Fig. 1.

Ligation yielded concatamers of hTBPc as shown in Fig. 2. Subcloning of the *in vitro* ligation reaction mixture into an appropriate vector yielded expression constructs containing polygenes encoding 1,2,3, and 5 hTBP proteins in a single polyprotein separated by TEV protease cleavage sites. A schematic representation and the nucleotide sequence (SEQ ID NO: 3) of one of the resulting expression vectors (pFBDO[hTBPc]3) are shown in Figs. 3 and 4, respectively.

Example 2: Generation of baculoviral transfer vectors containing polygenes encoding subunits of a human general transcription factor

A polygene was generated encoding a polyprotein comprising human TBP associated factors hTAF1 and hTAF2 in addition to hTBPc inserted into a transfer vector pUCDM (see WO 2005/085456 A1 (PCT/EP2004/013381)) for baculovirus expression, with the genes separated by sequences encoding an amino acid spacer and a TEV protease site. A schematic representation of the resulting construct pUCDMCSTAF1TBPcTAF2 is shown in Fig. 9. The nucleotide sequence of the construct is shown in Fig. 10 (SEQ ID NO: 4). A further construct was generated containing a polygene encoding a polyprotein comprising TEV protease and human TBP associated factors hTAF6 and hTAF9 inserted into the transfer vector pFBDM (see WO 2005/085456 A1 (PCT/EP2004/013381)) for baculovirus expression, with the genes separated by sequences encoding an amino acid spacer and a TEV protease site. A schematic representation of the resulting construct pFDDO[HisTEVTAF6TAF9]his is shown in Fig. 11. The nucleotide sequence of this construct is shown in Fig. 12. (SEQ ID NO: 5)

Example 3: Preparation of bacmid constructs, infection of insect cells and protein expression

For the construction of bacmids constructs comprising the above two polygenes, the constructs pUCDMCSTAF1TBPcTAF2 (pUCDM derivative) and pFDDO[HisTEVTAF6TAF9]his (pFBDM derivative) were each introduced into DH10MultiBac^{Cre} cells as described in Examples 5 (for pUCDMCSTAF1TBPcTAF2; *Cre-lox* site-specific recombination) and 6 (for pFDDO[HisTEVTAF6TAF9]his; Tn7 transposition) of WO 2005/085456 A1 (PCT/EP2004/013381). If desired, one-step transposition / cre-lox site-specific recombination can be carried DH10MultiBac^{Cre} cells as described in WO 2005/085456 A1 (PCT/EP2004/013381) as well. Bacmid preparation, infection of insect cells and protein expression was carried out according to established protocols (see, e.g., O'Reilly et al. (1994) "Baculovirus expression vectors. A laboratory manual" Oxford University Press, New York – Oxford; "Bac-to-Bac^{TM} Baculovirus Expression Systems Manual" Invitrogen, Life Technologies, Inc., 2000).

The following Sequence Listing is part of the present description, wherein the sequences are as follows:

SEQ ID NO: 1 is the nucleotide sequence of the PCR product coding for human TATA-Box-Binding Protein (hTPB) core (hTBPc, c-terminal fragment of the full-length protein truncated at position 159) shown in Fig. 1.

SEQ ID NO: 2 is the amino acid sequence of the human TATA-Box-Binding Protein core (hTBPc) shown in Fig. 1.

SEQ ID NO: 3 is the nucleotide sequence of pFBDO[hTBPc]3 shown in Fig. 8.

SEQ ID NO: 4 is the nucleotide sequence of pUCDMCSTAF1TBPcTAF2 shown in Fig. 10.

SEQ ID NO: 5 is the nucleotide sequence of pFBDO[HisTEVTAF6TAF9]his shown in Fig. 12.

Claims

1. A polynucleotide encoding, each within a single open reading frame (ORF), at least two polygenes each coding for at least three biologically active polypeptides wherein at least two of the polypeptides encoded by the genes constituting the polygenes are of non-viral origin, at least two of the polypeptides encoded by the genes constituting the polygenes are each capable of at least transiently interacting with at least one other polypeptide encoded by a gene of said polygenes, and the genes constituting each polygene are connected to one another by a sequence coding for at least one protease cleavage site and/or by a sequence coding for at least one self-cleaving peptide.

2. The polynucleotide of claim 1 wherein each polygene is operatively linked to a promoter sequence capable of controlling the expression of the polygene.

3. The polynucleotide of claim 1 or 2 wherein each ORF is flanked by a terminator sequence.

4. The polynucleotide according to any one of the preceding claims wherein the protease cleavage site is selected from the group consisting of the cleavage sites of potyvirus NIa proteases, potyvirus HC proteases, potyvirus P1 (P35) proteases, byovirus NIa proteases, byovirus RNA-2-encoded proteases, aphthovirus L proteases, enterovirus 2A proteases, rhinovirus 2A proteases, picorna 3C proteases, comovirus 24K proteases, nepovirus 24K proteases, RTSV (rice tungro spherical virus) 3C-like protease, PYVF (parsnip yellow fleck virus) 3C-like protease, thrombin, factor Xa and enterokinase.

5. The polynucleotide of claim 4 wherein the protease cleavage site is a TEV (tobacco etch virus) protease cleavage site.

6. The polynucleotide according to any one of claims 1 to 4 wherein the self-cleaving peptide is selected from the group consisting of aphthovirus 2A peptides and cardiovirus 2A peptides.

7. The polynucleotide of claim 6 wherein the self-cleaving peptide is the FMDV (foot-and-mouth disease virus) 2A peptide.

8. The polynucleotide according to any one of the preceding claims further containing the gene(s) for the protease(s) capable of cleaving the protease cleavage site(s).

9. The polynucleotide according to claim 5 wherein the protease gene(s) is/are part of at least one of the polygenes.

10. The polynucleotide according to any one of the preceding claims wherein at least one polygene comprises at least 4 genes.

11. The polynucleotide according to claim 10 wherein each polygene comprises at least 4 genes.

12. The polynucleotide according to any one of claims 1 to 11 wherein the lengths of the nucleotide sequences of the polygenes differ from one another by less than 50 %, preferably less than 30 %, more preferably 20 %.

13. The polynucleotide according to any one of the preceding claims further comprising at least one site for its integration into a vector or host cell.

14. The polynucleotide according to any one of the preceding claims wherein the genes constituting the polygenes are selected from the group consisting of genes encoding members of multiprotein complexes and genes encoding members of metabolic pathways.

15. The polynucleotide according to any one of the preceding claims comprising a functional arrangement according to Formula I

$$X - \underline{T1 - MCS1 - P1} - [A - B] - \underline{P2 - MCS2 - T2} - Y \qquad (I)$$

comprising

(a) at least two expression cassettes T1 – MCS1 – P1 and P2 – MCS1 – T2 in a head-to-head, head-to-tail or tail-to-tail arrangement, each comprising a multiple cloning site MCS1 or MCS2, flanked by a promoter P1 and a terminator sequence T1 for MCS1 and flanked by a promoter P2 and a terminator sequence T2 for MCS2

(b) at least one multiplication module M in between the promoters P1 and P2 comprising at least 2 restriction sites A and B

(c) at least two sites X and Y each flanking one of the expression cassettes,

wherein

(d) restriction sites A and X as well as B and Y are compatible, but

(e) the ligation products of AY and BX are not enzymatically cleavable by restriction enzymes a, b, x or y specific for restriction sites A, B, X and Y, respectively, and

(f) restriction sites A and B as well as X and Y are incompatible, and

wherein each polygene is inserted into one of the expression cassettes.

16. The polynucleotide of claim 15 wherein the restriction sites A and B in the multiplication module M are selected from the group consisting of restriction sites BstZ17I, SpeI, ClaI and NruI or restriction sites cleaved by isoschizomers thereof.

17. The polynucleotide of claim 15 to 16 wherein the restriction sites X and Y are selected from the group consisting of restriction sites PmeI and AvrII or restriction sites cleaved by isoschizomers thereof.

18. The polynucleotide according to any one of claims 15 to 17 wherein

(a) promoters P1 and P2 are selected from the group consisting of polh and p10;

(b) terminator sequences are selected from the group consisting of SV40 and HSVtk;

(c) restriction sites A and B in the multiplication module M are selected from group consisting of restriction sites BstZ171I, SpeI, ClaI and NruI;

(d) restriction sites X and Y are selected from the group consisting of restriction sites PmeI and AvrII; and

(e) sites for virus integration are selected from the group consisting of cre-lox and Tn7.

19. A vector comprising the polynucleotide according to any one of the preceding claims.

20. The vector according to claim 19 being a eukaryotic expression vector or transfer vector.

21. The vector according to claim 19 or 20 being a virus.

22. The vector according to claim 21 being a baculovirus expression vector.

23. The vector according to claim 22 wherein the baculovirus genes v-cath and chiA are functionally disrupted.

24. The vector according to any one of claims 19 to 23 comprising a site for site specific recombinases (SSRs).

25. The vector according to claim 24 comprising a LoxP site for cre-lox specific recombination.

26. The vector according to any one of claims 19 to 25 comprising a transposon element.

27. The vector according to claim 26 wherein the transposon element is the Tn7 attachment site.

28. A host cell containing the polynucleotide according to any one of claims 1 to 18 and/or the vector according to any one of claims 19 to 27.

29. The host cell according to claim 28, which is selected from the group consisting of bacteria, yeast cells, insect cells, nematode cells and mammalian cells.

30. The host cell according to claim 29, which is selected from the group consisting of *E. coli*, *S. cerevisiae*, *P. pastoris*, *S. pombe*, *C. albicans*, Sf9 cells, *C. elegans* cells, human cells, rodent cells, porcine cells, bovine cells and ovine cells.

31. A transgenic non-human animal transformed with the polynucleotide according to any one of claims 1 to 28 and/or the vector according to any one of claims 19 to 27.

32. The animal according to claim 31, which is selected from rodent, porcine, bovine and *C. elegans*.

33. A method for the production of the polynucleotide according to any one of claims 1 to 18 comprising the steps of

(a) amplifying the coding regions of the genes constituting the polygenes;

(b) providing said coding regions with the sequences coding for the at least one protease cleavage site and/or the at least one self-cleaving peptide;

(c) assembling the fragments resulting from steps (a) and (b) such that a single ORF results in each polygene;

(d) combining the at least two polygenes into a single polynucleotide.

34. A method for the production of multiprotein complexes *in vitro* comprising the steps of

(a) cultivating the host cell according to any one of claims 28 to 30 in a suitable medium under conditions allowing the expression of the polygenes; and

(b) recovering the expression products encoded by the polygenes from the medium and/or the host cells.

35. Use of the polynucleotide according to any one of claims 1 to 18 and/or the vector according to any one of claims 19 to 27 and/or the host cell according to any one of claims 28 to 30 for the preparation of a medicament comprising a polygene transfer vehicle for gene therapy.

36. Use of the polynucleotide according to any one of claims 1 to 18 and/or the vector according to any one of claims 19 to 27 and/or the host cell according to any one of claims 28 to 30 for the crystallisation of multiprotein complexes.