(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 3 375 869 A1

(12) EUROPEAN PATENT APPLICATION
published in accordance with Art. 153(4) EPC

(43) Date of publication:
19.09.2018 Bulletin 2018/38

(21) Application number: **16863726.2**

(22) Date of filing: **14.11.2016**

(51) Int Cl.:
C12N 9/08 *(2006.01)*

(86) International application number:
PCT/ES2016/070809

(87) International publication number:
WO 2017/081355 (18.05.2017 Gazette 2017/20)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
MA MD

(30) Priority: **13.11.2015 ES 201531641**

(71) Applicant: **Consejo Superior De Investigaciones Científicas (CSIC)**
28006 Madrid (ES)

(72) Inventors:
- **MOLINA ESPEJA, Patricia**
  28049 Madird (ES)
- **PLOUGASCA, Francisco José**
  28049 Madrid (ES)
- **ALCALDE GALEOTE, Miguel**
  28049 Madrid (ES)
- **GÓMEZ DE SANTOS, Patricia**
  28049 Madrid (ES)

(74) Representative: **Pons**
Glorieta Ruben Dario 4
28010 Madrid (ES)

(54) **MUTANTS OF UNSPECIFIC PEROXYGENASE WITH HIGH MONOOXYGENASE ACTIVITY AND USES THEREOF**

(57) The invention relates to an unspecific peroxygenase of the *Agrocybe aegerita* fungus, obtained by means of directed molecular evolution to facilitate the functional expression thereof in an active, soluble and stable form. The peroxygenase described in the invention shows a significant improvement in the functional expression thereof, improved monooxygenase activity and reduced peroxidase activity, in relation to the monooxygenase and peroxidase activities showed by the unspecific wild-type peroxygenase of *A. aegerita*. The peroxygenase of the invention is useful in chemical processes, including industrial transformations such as the selective oxyfunctionalisation of carbon-hydrogen bonds of various organic compounds.

EP 3 375 869 A1

Description

TECHNICAL FIELD OF THE ART

[0001] The present invention belongs to the field of molecular biology, recombinant DNA technology and biotechnology. Specifically, it relates to a peroxygenase enzyme with enhanced functional expression in an active, soluble and stable form, showing improved peroxygenase activity and reduced peroxidase activity with respect to the native enzyme or wild-type, and which has been obtained through a process of directed molecular evolution. Said enzyme may be used in chemical processes, including industrial transformations such as the selective oxyfunctionalisation of carbon-hydrogen bonds of various organic compounds, preferably those hydroxylation processes that transform naphthalene into 1-naphthol and/or propranolol into 5'-hydroxypropranolol.

STATE OF THE ART

[0002] The methods of organic synthesis, preferably processes aimed at selective oxyfunctionalisation of carbon-hydrogen bonds of various organic compounds and, more specifically, those compounds that by hydroxylation processes give rise to other products with characteristics more suitable for different uses such as, for example, synthesis of agrochemical products, herbicides, insecticides, pharmaceuticals, cosmetics and dye precursors, are currently carried out using chemical catalysts, such as sulfonic acid and platinum compounds, which are highly polluting products, with low turnover numbers and reduced regioselectivity, in addition to high energy consumption (high temperatures and pressures), high production costs and large release of waste.

[0003] In the search for a more environmentally friendly alternative and, thus, prevent the aforementioned drawbacks of the use of chemical catalysts in this type of reactions, microorganisms such as *Cunninghamella, Bacillus cereus* ATCC14579, the green algae *Chlorella* and various fungi and enzymes which transform, by means of hydroxylation procedures, for example naphthalene into 1- naphthol, have been used.

[0004] In this regard, enzymes with monooxygenase activity which conduct selective oxyfunctionalisation of aromatic rings may offer a more ecological alternative to conventional chemical processes.

[0005] For example, in the case of the aromatic hydrocarbon 1-naphthol, naphthalene-based synthesis is carried out with enzymes that show monooxygenase activity. Specifically, P450 monooxygenases are enzymes that show such activity and which have been subjected to engineering for different purposes over the years, from the selective hydroxylation of alkanes -including terminal hydroxylation- to the unnatural cyclopropanation of olefins by means of carbon transfer. Said P450 monooxygenase enzymes transform naphthalene into 1-naphthol either by means of the peroxide shunt pathway or by means of its NAD(P)H-dependent natural activity (H. J. Zhanglin, F.H. Arnold, Nature 1999. 399, 670-673; P. C. Cirino, F. H. Arnold, Angew. Chem. Int. Ed. 2003. 42, 3299-3301; P. Meinhold, et al. Adv. Synth. Catal. 2006. 348, 763-772; P. S. Coelho, et al. Science 2013. 339, 307-310). More recently, the evolution of the toluene ortho-monooxygenase enzyme (TOM) and its involvement in the process of a cell biocatalytic system has also been described (K. A. Canada, et al. J. Bacteriol. 2002. 184, 344-349; L. Rui, et al. Appl. Environ. Microbiol. 2004. 70, 3246-3252; J. Garikipati, et al. Appl. Environ. Microbiol. 2009. 75, 6545-6552). In all these cases, the low enzyme stability of the aforementioned enzymes, along with the high requirements in terms of high-cost redox cofactors (NADPH) and associated reducing domains (flavins), have prevented the industrial use thereof in the synthesis of the aromatic hydrocarbon 1-naphthol from naphthalene.

[0006] Furthermore, Human Drug Metabolites (HDMs) are the result of the metabolism of pharmaceutical compounds, mainly by hepatic P450 monooxygenase enzymes. For the pharmaceutical industry, the toxicity evaluation, effectiveness and activity of these metabolites is key, but to date the chemical synthesis thereof produces very low yields besides being very complicated. The most important HDMs include, namely, those derived from hydrocarbon propranolol, such as 5'- hydroxypropranolol. Propranolol is a beta-blocker drug commonly used for the treatment of hypertension, migraine prophylaxis in children and attenuation of physical manifestations of anxiety. Heretofore, known enzymatic alternatives for obtaining propranolol derivatives are P450 monooxygenase enzymes or unspecific fungal peroxygenases such as *Agrocybe aegerita* (*Aae*UPO) and *Coprinellus radians* (*Cra*UPO). Specifically, P450 monooxygenases require cellular environments and/or expensive redox cofactors (NADPH), in addition to associated reducing domains (flavins), and show low operational stabilities and low regioselectivity. Furthermore, the specific fungal peroxygenases described require antioxidants such as ascorbic acid to prevent the subsequent oxidation of the product of interest.

[0007] One of the enzymes studied for the synthesis of the aforementioned compounds, 1-naphthol and 5'-hydroxypropranolol, was the enzyme UPO (*Unspecific PerOxygenase,* E.C. 1.11.2.1), secreted by the basidiomycete fungus *Agrocybe aegerita,* and known as the first "true" natural aromatic peroxygenase. The enzyme AaeUPO has properties resembling those of P450 monooxygenase enzymes as regards the selective oxyfunctionalisation of carbon-hydrogen bonds of various organic compounds. AaeUPO is an extracellular, highly active and stable enzyme, besides not requiring cofactors or auxiliary redox flavoproteins, *i.e.* it is self-sufficient. With minimal requirements, just catalytic concentrations

of H_2O_2 (acting as an enzyme co-oxidant - primary electron acceptor- and oxygen source), AaeUPO is capable of carrying out a wide variety of highly complex transformations in organic synthesis, such as for example the hydroxylation of aromatic and aliphatic compounds, olefin epoxidation, N- and S-oxidation of heterocyclic compounds or breakage of ether linkages, among many others. Furthermore, it has natural mono(per)oxygenase activity, such as P450 monooxygenase enzyme, and peroxidase on phenolic substrates (M. Kluge, et al. Appl. Microbiol. Biotechnol. 2009. 81, 1071-1076). The coexistence of both activities, peroxygenase and peroxidase, in the same enzyme is a problem when the objective is to use this enzyme in an industrial process, since the products of hydroxylation of AaeUPO always appear with different amounts of oxidation products derived from the former. This is especially true in the case of aromatic hydroxylations wherein the product(s) released by the peroxygenase activity may in turn again be substrates for the peroxidase activity of the UPO, promoting the formation of quinones involving non-enzymatic polymerisation which affects the overall efficiency of the process.

[0008] Therefore, in the state of the art there is a need for enzymes showing improved monooxygenase activity, to the detriment of its peroxidase activity, together with high enzyme stability, high regioselectivity and which are self-sufficient, *i.e.* they do not require the presence of cofactors to carry out their monooxygenase activity. It is also important to note that said enzymes require robust expression systems that provide high levels of active enzyme. Therefore, these enzymes, due to the aforementioned characteristics, are suitable for use in methods of organic synthesis, preferably in processes of oxyfunctionalisation, oxidation or selective hydroxylation of hydrocarbons in general, both aromatic and aliphatic linear, branched and cyclic, preferably the method of hydroxylation of cyclic aromatic compounds, both single cyclic or condensed compounds, more preferably the method of hydroxylation for the synthesis of 1-naphthol and/or synthesis of 5'-hydroxypropranolol, where said processes are carried out in a single step under mild conditions, such as ambient temperature, atmospheric pressure and in an aqueous solution, with low organic co-solvent content, to reduce energy consumption, as well as the harmful effects of chemical synthesis.

DESCRIPTION OF THE INVENTION

[0009] The present invention describes the directed evolution of the unspecific peroxygenase UPO (E.C. 1.11.2.1) of *A. aegerita* (*Aae*UPO of SEQ ID NO: 1), to obtain variants or mutants showing a functional expression in a soluble, active and highly stable form in a eukaryote heterologous host, preferably *Saccharomyces cerevisiae or Pichia pastoris,* besides showing an improved peroxygenase activity and reduced peroxidase activity relative to the wild-type UPO enzyme of *A. aegerita* (SEQ ID NO: 2) expressed in *S. cerevisiae.* Said variants or mutants, due to the aforementioned characteristics, are suitable for use in methods of organic synthesis, preferably in processes of oxyfunctionalisation, oxidation or selective hydroxylation of hydrocarbons in general, both aromatic and aliphatic linear, branched and cyclic, preferably the method of hydroxylation of cyclic aromatic compounds, both single cyclic or condensed compounds, more preferably the method of hydroxylation for the synthesis of 1-naphthol and/or synthesis of 5'-hydroxypropranolol wherein these processes are carried out in a single step, without requiring the presence of cofactors, under mild conditions such as ambient temperature, atmospheric pressure and in an aqueous solution, with low organic co-solvent content, to reduce energy consumption, as well as the adverse consequences of the chemical synthesis.

[0010] The peroxygenase UPO1 of *A. aegerita* (*Aae*UPO of SEQ ID NO: 1) was subjected to several cycles of laboratory-directed evolution combined with semi-rational approaches (*i.e.* rational semi-rational and random design methods were used) for the different variants described herein. On the one hand, the peroxygenase UPO1 of *A. aegerita* (AaeUPO of SEQ ID NO: 1) was subjected to five cycles of directed evolution, giving rise to the mutant, hereinafter and throughout the present invention PaDa-I, SEQ ID NO: 14 and which is encoded by the nucleotide sequence SEQ ID NO: 13. Said PaDa-I mutant comprises the L67F, I248V, F311L, V75I and V57A mutations with respect to wild AaeUP01 of SEQ ID NO: 2, encoded by the sequence SEQ ID NO: 1. Similarly, the nucleotide sequence that encodes the native signal peptide of *Aae*UPO1 (SEQ ID NO: 25) was also subjected to directed evolution cycles and gave rise to a modified or evolved signal peptide of SEQ ID NO: 27, as described in P. Molina-Espeja et al. Appl. Environ. Microbiol. 2014. 80, 3496-3507. In this manner, the PaDa-I mutant that comprised the evolved signal peptide (SEQ ID NO: 27) was obtained, whose nucleotide sequence is SEQ ID NO: 17, which encodes the PaDa-I peptide of SEQ ID NO: 18. Said PaDa-I mutant, as demonstrated by the inventors (P. Molina-Espeja, et al. Appl. Environ. Microbiol. 2014. 80, 3496-507) has high functional expression, enhanced catalytic constants, high thermostability and greater resistance to the presence of organic co-solvents with respect to the wild-type UPO expressed in S. *cerevisiae.* Enzyme substrate promiscuity was preserved performing a dual assay in High-Throughput Screening (HTS) format to explore both oxidative activities and those relating to oxygen transfer from mutant libraries, besides incorporating an assay to avoid the loss of kinetic thermostability.

[0011] Two new cycles of laboratory-directed evolution were carried out based on the previously described PaDa-I mutant, which gave rise to the JaWa variant of SEQ ID NO: 23, with two added mutations in the protein sequence SEQ ID NO: 24: G241D y R257K, regarding the sequence of the PaDa-I mutant. In this manner, the JaWa mutant of the nucleotide sequence SEQ ID NO: 23 or SEQ ID NO: 19 is obtained, which encode the peptides of SEQ ID NO: 24 or

SEQ ID NO: 20, depending on whether or not they have the evolved or modified signal peptide of SEQ ID NO: 28 encoded for the nucleotide sequence of SEQ ID NO: 27. On the other, these two new mutations, G241D and R257K, were also incorporated to the sequence of the native peroxygenase *Aae*UPO1 (SEQ ID NO: 1) by means of directed mutagenesis, giving rise to a variant we will call wt-JaWa of SEQ ID NO: 8 or SEQ ID NO: 12, respectively encoded by the nucleotide sequences SEQ ID NO: 7 or SEQ ID NO: 11, depending on whether or not the evolved signal peptide of SEQ ID NO: 28 encoded for the nucleotide sequence of SEQ ID NO: 27.

[0012] Based on the JaWa mutant SEQ ID NO: 23 encoded for the nucleotide sequence SEQ ID NO: 24, previously described, another three new laboratory-directed evolution cycles were carried out which gave rise to the SoLo variant of SEQ ID NO: 41, with an added mutation in the protein sequence SEQ ID NO: 42: F191S, with respect to the sequence of the JaWa mutant. In this manner, the SoLo mutant of the nucleotide sequence SEQ ID NO: 41 or SEQ ID NO: 37 is obtained, which encode the peptides of SEQ ID NO: 42 or SEQ ID NO: 38, depending on whether or not they have the evolved or modified signal peptide of SEQ ID NO: 28, encoded by the nucleotide sequence of SEQ ID NO: 27. Furthermore, this new mutation, F191S, was also incorporated to the sequence of the native peroxygenase *Aae*UPO1 (SEQ ID NO: 1) by means of directed mutagenesis, giving rise to a variant we will call wt-SoLo of SEQ ID NO: 62 or SEQ ID NO: 66, respectively encoded by the nucleotide sequences SEQ ID NO: 61 or SEQ ID NO: 65, depending on whether or not they have the evolved signal peptide of SEQ ID NO: 28 encoded for the nucleotide sequence of SEQ ID NO: 27.

[0013] Thus, the variants described herein, preferably the variants JaWa and SoLo, have all the characteristics and advantages previously mentioned for the PaDa-I mutant, but also show a greater increase in thermostability (values of T_{50}= 59.7°C, an increase in thermostability of 2°C, with respect to the variant PaDa-I), greater stability against the presence of co-solvents and kinetic values against naphthalene of k_{cat}/K_m of around 1.56 fold higher than those described for the PaDa-I variant when said mutants are expressed in a heterologous organism, preferably in yeasts, for the case of the variant JaWa and around 1.47 fold higher in k_{cat} for the case of the variant SoLo. Therefore, the main advantages of the variants with improved peroxygenase activity and reduced peroxidase activity, with respect to wild AaeUPO, or to other variants of the state of the art, such as for example the variant PaDa-I, are as follows:

i) they show a high production rate,
ii) they show high activity,
iii) they show high stability,
iv) they show an increase in TTN of 2.5 fold (TTN of approximately 50,000) in the case of the synthesis of 1-naphthol and of three fold in the absence of antioxidants (45,000 for the SoLo mutant against 15,000 of the JaWa mutant) or of 15 fold (3,000 in the case of wild AaeUPO) for the synthesis of 5'-hydroxypropranolol,
v) shows an increase in k_{cat} for 1-naphthol of up to 1.5 fold and an increase in k_{cat} for 5'-hydroxypropranolol of up to 3.6 fold,
vi) shows enhanced catalytic efficiency for naphthalene up to values of $6.2 \times 10^5\ s^{-1}M^{-1}$; and for 5'-hydroxypropranolol of $3.1 \times 10^6\ s^{-1}\ M^{-1}$, two orders of magnitude higher than those of any enzyme described,
vii) show a reduction of approximately 1.5 fold in the ratio 1.4-naphthoquinone:1-naphthol, and up to 50% less oxidation with respect to 5'-hydroxypropranolol,
viii) They have a regioselectivity against 1-naphthol of approximately 97% and of approximately 99% against 5'-hydroxypropranolol.

[0014] Therefore, the present invention provides new peroxygenases showing all the aforementioned advantages over native or wild-type peroxygenase, such as the functional expression in a heterologous organism, preferably, *S. cerevisiae* or *P. pastoris,* as well as with respect to other variants or mutants currently known in the state of the art, such as the PaDa-I variant. Additionally, the variants described herein have greater selectivity and the highest total turnover numbers (TTN) for methods of organic synthesis, preferably in processes of oxyfunctionalisation or selective oxidation of hydrocarbons in general, both aromatic and aliphatic linear, branched and cyclic, preferably the method of hydroxylation of cyclic aromatic compounds, both single cyclic or condensed compounds, more preferably the method of hydroxylation for the synthesis of 1-naphthol and/or synthesis of 5'-hydroxypropranolol, known to date for this enzyme superfamily. Heterologously secreted in an active, soluble and very stable form, these variants carry out selective aromatic oxygenations in the absence of cofactors NAD(P)H and reductase domains. Its self-sufficient mono(per)oxygenase activity, together with its reduced peroxidase activity, make these UPO variants a valuable biocatalyst for the future of applications in the field of organic synthesis.

[0015] Thus, the present invention relates to the amino acid sequences of said peroxygenase variants, and the nucleotide sequences that encode said peroxygenase variants. Below is a list of the polynucleotides and polypeptides described herein:

- SEQ ID NO: 1 - Nucleotide sequence of the gene that encodes *Aae*UPO1 without signal peptide.
- SEQ ID NO: 2 - Polypeptide sequence of *Aae*UPO1 without signal peptide.

- SEQ ID NO: 3 - Nucleotide sequence of the gene that encodes *Aae*UPO1 with wild-type signal peptide.
- SEQ ID NO: 4 - Polypeptide sequence of *Aae*UPO1 with wild-type signal peptide.
- SEQ ID NO: 5 - Nucleotide sequence of the gene that encodes *Aae*UPO1 with modified signal peptide.
- SEQ ID NO: 6 - Polypeptide sequence *Aae*UPO1 with modified signal peptide.
- SEQ ID NO: 7 - Nucleotide sequence that encodes the wt-JaWa variant without signal peptide.
- SEQ ID NO: 8 - Polypeptide sequence of the wt-JaWa variant without signal peptide.
- SEQ ID NO: 9 - Nucleotide sequence that encodes the wt-JaWa variant with wild-type signal peptide.
- SEQ ID NO: 10 - Polypeptide sequence of the wt-JaWa variant with wild-type signal peptide.
- SEQ ID NO: 11 - Nucleotide sequence that encodes the wt-JaWa variant with modified signal peptide.
- SEQ ID NO: 12 - Polypeptide sequence of the wt-JaWa variant with modified signal peptide.
- SEQ ID NO: 13 - Nucleotide sequence that encodes the PaDa-I variant without signal peptide.
- SEQ ID NO: 14 - Polypeptide sequence of the PaDa-I variant without signal peptide.
- SEQ ID NO: 15 - Nucleotide sequence that encodes the PaDa-I variant with wild-type signal peptide.
- SEQ ID NO: 16 - Polypeptide sequence of the PaDa-I variant with wild-type signal peptide.
- SEQ ID NO: 17 - Nucleotide sequence that encodes the PaDa-I variant with modified signal peptide.
- SEQ ID NO: 18 - Polypeptide sequence of the PaDa-I variant with modified signal peptide.
- SEQ ID NO: 19 - Nucleotide sequence that encodes the JaWa variant without signal peptide.
- SEQ ID NO: 20 - Polypeptide sequence of the JaWa variant without signal peptide.
- SEQ ID NO: 21 - Nucleotide sequence that encodes the JaWa variant with wild-type signal peptide.
- SEQ ID NO: 22 - Polypeptide sequence of the JaWa variant with wild-type signal peptide.
- SEQ ID NO: 23 - Nucleotide sequence that encodes the JaWa variant with modified peptide.
- SEQ ID NO: 24 - Polypeptide sequence of the JaWa variant with modified signal peptide.
- SEQ ID NO: 25 - Nucleotide sequence that encodes the native signal peptide of *Aae*UPO1.
- SEQ ID NO: 26 - Polypeptide sequence of the native signal peptide of *Aae*UPO1
- SEQ ID NO: 27 - Nucleotide sequence that encodes the modified signal peptide comprising mutations F[12]Y, A[14]V, R[15]G and A[21]D with respect to the nucleotide sequence that encodes the native signal peptide of *Aae*UPO1 of SEQ ID NO: 26.
- SEQ ID NO: 28 - Polypeptide sequence of the modified signal peptide comprising the mutations F[12]Y, A[14]V, R[15]G and A[21]D with respect to the polypeptide sequence of SEQ ID NO: 26.
- SEQ ID NO: 29 - Nucleotide sequence that encodes the W24F variant obtained from the PaDa-I mutant of SEQ ID NO: 17.
- SEQ ID NO: 30 - Polypeptide sequence that encodes the W24F variant obtained from the PaDa-I mutant of SEQ ID NO: 18.
- SEQ ID NO: 31 - Nucleotide sequence that encodes the W24F variant obtained from the JaWa mutant of SEQ ID NO: 23.
- SEQ ID NO: 32 - Polypeptide sequence that encodes the W24F variant obtained from the JaWa mutant of SEQ ID NO: 24.
- SEQ ID NO: 37 - Nucleotide sequence that encodes the SoLo variant without signal peptide.
- SEQ ID NO: 38 - Polypeptide sequence of the SoLo without signal peptide.
- SEQ ID NO: 39 - Nucleotide sequence that encodes the SoLo variant with wild-type signal peptide.
- SEQ ID NO: 40 - Polypeptide sequence of the SoLo variant with wild-type signal peptide.
- SEQ ID NO: 41 - Nucleotide sequence that encodes the SoLo variant with modified signal peptide.
- SEQ ID NO: 42 - Polypeptide sequence of the SoLo variant with modified signal peptide.
- SEQ ID NO: 61 - Nucleotide sequence that encodes the wt-SoLo variant without signal peptide.
- SEQ ID NO: 62 - Polypeptide sequence of the wt-SoLo variant without signal peptide.
- SEQ ID NO: 63 - Nucleotide sequence that encodes the wt-SoLo variant with wild-type signal peptide.
- SEQ ID NO: 64 - Polypeptide sequence of the wt-SoLo variant with wild-type signal peptide.
- SEQ ID NO: 65 - Nucleotide sequence that encodes the wt-SoLo variant with modified signal peptide.
- SEQ ID NO: 66 - Polypeptide sequence of the wt-SoLo variant with modified signal peptide.

[0016] The authors of the present invention have used a methodological combination based on directed evolution and mutagenesis and have obtained peroxygenase variants or mutants that resolve the need for a biocatalyst with high activity and thermostability, a high functional production rate, in addition to showing enhanced peroxygenase activity and reduced peroxidase activity, with respect to the wild-type UPO enzyme or even with respect to other UPO variants such as the PaDa-I variant.

[0017] The peroxygenases of the present invention, preferably the so-called JaWa and SoLo variants, are highly stable against temperature (values of T_{50}= 59.7°C/59.5 °C, an increase in thermostability of 2°C with respect to the PaDa-I variant, being T_{50} the temperature at which the enzyme maintains 50% of its initial activity after 10 minutes of incubation)

and against the presence of co-solvents. Said peroxygenases have kinetic values with respect to naphthalene of k_{cat}/K_m of around 1.56 fold higher than those described for the PaDa-I variant and around 46 fold higher than that described for wild AaeUPO with respect to propranolol, expressed in a heterologous organism, preferably yeasts, due to which its evolutionary design has given rise to:

i) Functional heterologous expression in yeast (0.2 g/L),
ii) increase in catalytic constants and efficiencies,
iii) increased stability against various factors (temperature, co-solvents).

[0018] Therefore, the main advantages of the variants with enhanced peroxygenase activity and reduced peroxidase activity, with respect to the wild UPO, or to other variants of the state of the art, such as for example the PaDa-I variant, as mentioned earlier, are as follows:

i) it shows a high production rate,
ii) it shows high activity,
iii) it shows high stability,
iv) it shows an increase in TTN of up to 2.5 fold (TTN of approximately 50,000), for the case of synthesis of 1-naphthol, and for the synthesis of 5'-hydroxypropranolol of three fold in the absence of antioxidants (45,000 for the SoLo mutant against 15,000 of the JaWa mutant) or 15 fold (3,000 in the case of wild AaeUPO),
v) it shows an increase in k_{cat} for 1-naphthol of up to 1.5 times and an increase in k_{cat} for 5'-hydroxypropranolol of up to 3.6 fold,
vi) it shows enhanced catalytic efficiency for naphthalene up to values of 6.2 x 10^5 $s^{-1}M^{-1}$; and for 5'-hydroxypropranolol of 3'1 x 10^6 s^{-1} M^{-1}, two orders of magnitude higher than those of any enzyme described,
vii) it shows a reduction of approximately 1.5 fold in the ratio 1.4-naphthoquinone:1-naphthol and up to 50% less oxidation on 5'-hydroxypropranolol,
viii) it shows regioselectivity against 1-naphthol of approximately 97% and of approximately 99% against 5'-hydroxypropranolol.

[0019] For the purposes of the present invention, the term "peroxygenase" relates to the unspecific peroxygenase enzyme in accordance with EC 1.11.2.1, which catalyses the insertion of an oxygen atom from H_2O_2 or other peroxide which acts as a source of oxygen, in a wide variety of substrates. For the purposes of the present invention, peroxygenase is preferably unspecific peroxygenase (UPO) secreted by the basidiomycete fungus *A. aegerita,* whose nucleotide sequence is SEQ ID NO: 3 or SEQ ID NO: 1 that encodes a protein whose amino acid sequence is SEQ ID NO: 4 or SEQ ID NO: 2, depending on whether or not it comprises a signal peptide, respectively.

[0020] The terms "oxygen donors", "oxidising agent" and "oxidant" relate to a substance, molecule or compound that donates oxygen to a substrate in an oxidation reaction.

[0021] Typically, the oxygen donor is reduced (it accepts electrons). By way of example, non-limiting oxygen donors include molecular oxygen or dioxygen (O_2) and peroxides, including alkyl peroxides such as t-butyl, cumene hydroperoxide, paracetic acid and, more preferably, hydrogen peroxide (H_2O_2). A "peroxide" is any compound other than molecular oxygen (O_2) which has two oxygen atoms bonded to each other.

[0022] For the purposes of the present invention, the term "mutant" or "variant", used indistinctly throughout the present invention and relating to the UPO peroxygenases of the invention obtained by means of the methods described herein and which have at least two mutations, preferably at least three mutations, more preferably at least four mutations, more preferably at least five mutations, more preferably at least six mutations, more preferably at least seven mutations, more preferably at least eight mutations, more preferably at least nine mutations, more preferably at least ten mutations, more preferably at least eleven mutations and more preferably at least twelve mutations, resulting from greater peroxygenase activity and lower peroxidase activity, in addition to all the aforementioned advantages, than that showed by the corresponding native or wild-type UPO enzyme or any other UPO variant, preferably the PaDa-I variant, expressed in a heterologous host, preferably in yeasts of the genus *Saccharomyces* sp. and *Pichia* sp. and more preferably in the *S. cerevisiae* and *P. pastoris* species.

[0023] For the purposes of the present invention, the term "cofactor" relates to any substance that is necessary or beneficial to the activity of an enzyme. "Coenzyme" means a cofactor that interacts directly with and serves to promote a reaction catalysed by an enzyme. Many coenzymes also serve as carriers. For example, NAD^+ and $NADP^+$ carry hydrogen atoms from one enzyme to another (in the form of NADH and NADPH, respectively). An "auxiliary protein" means any protein substance necessary or beneficial to the activity of an enzyme.

[0024] In a first aspect, the present invention relates to a polynucleotide that encodes a polypeptide with peroxygenase activity, hereinafter polynucleotide of the invention, characterised in that the amino acid sequence of the polypeptide encoding show an identity of at least 70% with SEQ ID NO: 2 (*Aae*UPO1), and comprising at least two amino acid

alterations in the positions homologous to positions 241 and 257 of the sequence, which replace the amino acids: original glycine (G) by ascorbic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K).

[0025] In a preferred embodiment of the nucleotide of the invention, it is characterised in that the amino acid sequence of the polypeptide encoding showing an identity of at least 70% with SEQ ID NO: 2 (*Aae*UPO1), and further comprises an amino acid alteration in the homologous position to position 191 of the sequence SEQ ID NO: 2, which replaces the original amino acid phenylalanine (F) by serine (S) (F191S).

[0026] In another preferred embodiment of the polynucleotide of the invention, it is characterised in that the amino acid sequence of the polypeptide encoding showing an identity of at least 70% with SEQ ID NO: 2 (*Aae*UPO1), and comprises the amino acid alterations in the homologous positions 241, 257 and 191 of said sequence, which replace the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D), original arginine (R) by lysine (K) in position 257 (R257K) and original phenylalanine (F) by serine (S) (F191S).

[0027] With the information supplied in the present invention, a person skilled in the art is capable of identifying nucleotide sequences homologous to those described in the present invention and that encode peroxygenase with identical characteristics to those described for the peroxygenase of the invention. Therefore, the polynucleotide of the invention is the coding sequence of an *Aae*UPO1 peroxygenase variant with the described enhanced activity, whose nucleotide sequence corresponds to:

a) nucleic acid molecules of the isolated polynucleotide sequence or in its complementary strand,
b) nucleic acid molecules whose complementary strand is capable of hybridising in astringent conditions with a polynucleotide sequence of (a), or
c) nucleic acid molecules, whose sequence differs from (a) and/or (b) due to the degeneration of the genetic code.

[0028] The term "astringent conditions" or "astringent hybridisation conditions" makes reference to conditions in which a hybridisation probe with its target sequence has a higher level than that of the other sequences (*i.e.* at least two fold higher than the base). The astringent conditions depend on the nature of the sequence and may vary according to the circumstances. Fully homologous target sequences can be identified by controlling astringency and washing conditions. Alternatively, astringency conditions may be adjusted to allow certain non-homologous pairings which may be detected at lower homology levels. A probe generally has less than 1,000 nucleotides in length and optionally less than 500 nucleotides. An average person skilled in the art has a deep understanding of nucleic acid hybridisation techniques.

[0029] The polynucleotides that encode the polypeptides of amino acid sequences described in the invention correspond to variants obtained by means of directed evolution of *Aae*UPO1 peroxygenase (E.C. 1.11.2.1). Said protein, *Aae*UPO1, corresponds to the nucleotide or polynucleotide sequence of SEQ ID NO: 3 or SEQ ID NO: 1, that are the coding sequences of the polypeptide with the amino acid sequence SEQ ID NO: 4 or SEQ ID NO: 2, depending on whether or not it comprises the signal peptide, respectively.

[0030] The term "polynucleotide", as used in the description, relates to polymeric forms of nucleotides of any length, both ribonucleotides and deoxyribonucleotides.

[0031] The term "identity" or "percentage of identity" between two sequences (nucleic acids or proteins) is understood to be the designation of a percentage of nucleotides or identical amino acid residues between the two compared sequences, obtained after the best alignment, being said percentage purely statistic and wherein the differences between the two sequences are distributed randomly and along the entire length. The term "best alignment" or "optimum alignment" is understood to be the designation of the alignment whereby the percentage of identity determined as described below is the highest. Comparisons between two nucleotide or amino acid sequences are traditionally performed: comparing these sequences once optimally aligned, performing said comparison by segment or by "comparison window" to identify and compare local regions of similarity regions. The optimum alignment of these sequences for comparison can be performed, in particular, with the help of one of the following algorithm: the local homology algorithm, Smith and Waterman (1981); the local homology algorithm, Neddleman and Wunsch (1970); the similarity search method, Pearson and Lipman (1988); the computer programs that use these algorithms (GAP, BESTFIT, BLASTP, BLASTN, BLASTX, TBLASTX, FASTA and TFASTA in the Wisconsin Genetics software package (*Genetics Computer Group,* 575 *Science* Dr., Madison, WI), or the Internet servers in particular of the National Centre for Biotechnology (NCBI) (http ://www.ncbi.nlm.nih.gov), EMBL (http ://www.embl.org) and the Ensembl project (http ://www.ensembl.org)). In order to obtain optimum alignment, the BLAST program is preferably used, with the BLOSUM 62 matrix. The PAM or PAM250 matrices may also be used, in addition to an identity matrix for the nucleotide sequences.

[0032] In a preferred aspect of the invention, the polynucleotide and polypeptide sequences described herein comprise at least approximately 60%, at least approximately 65%, at least approximately 70%, at least approximately 75%, at least approximately 80%, at least approximately 85%, at least approximately 88% of identity, at least approximately 89%, at least approximately 90%, at least approximately 91%, at least approximately 92%, at least approximately 93%, at least approximately 94%, at least approximately 95%, at least approximately 96%, at least approximately 97%, at least approximately 98%, at least approximately 99% or 100% of identity against a reference sequence, when compared

and aligned for a maximum correspondence against a comparison window or designated region as measured using the aforementioned algorithms.

[0033] The term "homology" or "percentage of homology" (percentage of homology, identity+similarity) is determined using homology comparison software, such as BLASP, TBLASTN or tBLASTX, of the National Centre of Biotechnology Information (NCBI), using the specific parameters. For the purposes of the present invention, the term "homology" relates to the identity of two or more nucleic acid sequences or to the identity of two or more amino acid sequences. Homologous sequences include "paralogous" and "orthologous". The term "paralogous" relates to gene duplications within the genome of a species, giving rise to paralogous genes. The term "orthologous" relates to homologous genes in different organisms due to the ancestral relationship.

[0034] In a preferred aspect, the polynucleotides that encode the polypeptide of the present invention show an enhancement of at least 20%, for example, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or at least 100% in the peroxygenase activity of the polynucleotide that encodes the mature polypeptide of SEQ ID NO: 4, SEQ ID NO: 2, SEQ ID NO: 14 or SEQ ID NO: 18.

[0035] In a preferred aspect, the polynucleotides that encode the polypeptide of the present invention show a reduction of at least 20%, for example, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or at least 100% in the peroxydase activity of the polynucleotide that encodes the mature polypeptide of SEQ ID NO: 4, SEQ ID NO: 2, SEQ ID NO: 14 or SEQ ID NO: 18.

[0036] The term "allelic variation" means any of two or more alternative forms of a gene that occupies the same chromosome locus. Allelic variation occurs naturally through mutation and can lead to polymorphism within populations. Gene mutations may be silent (without changes in the encoded polypeptide) or may encode polypeptides with altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

[0037] The term "encodes", as used in the description, makes reference to the correlation existing between the nucleotide triplets or codons in a DNA sequence and the amino acids that form the peptides, the amino acid sequences or the proteins. Where it states that a nucleotide sequence encodes a peptide, it means that when said nucleotide sequence is transcribed to messenger RNA (mRNA) and this mRNA is translated, said peptide will be generated.

[0038] For the purposes of the present invention, the term "encoding sequence" or sequence "that encodes" a polypeptide, protein or enzyme is a nucleotide sequence which, when expressed, gives rise to the production of this polypeptide, protein o enzyme, *i.e.* the nucleotide sequence encodes an amino acid sequence for that polypeptide, protein or enzyme. A coding sequence is "under the control" of sequences that control cell transcription and translation when the RNA polymerase transcribes the mRNA-coding sequence, which is subsequently transcribed and translated into the protein encoded by the coding sequence. Preferably, the coding sequence is a double-stranded DNA sequence which is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulating sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic, DNA, cDNA and synthetic DNA sequences or a recombinant nucleotide sequence. If the coding sequence is intended for expression in a eukaryotic cell, a transcription termination sequence and polyadenylation signal will be generally located 3' to the coding sequence.

[0039] The term "cDNA" is defined herein as a DNA molecule that can be prepared for reverse transcription using a mature, full-length mRNA molecule obtained from a eukaryotic cell. cDNA lacks sequences of introns that are normally present in the corresponding genomic DNA. The transcription of primary (initial) RNA is a mRNA precursor which is processed through a series of steps before appearing as mature, full-length mRNA. These steps include the elimination of intronic sequences through a process called linking. Therefore, cDNA derived from mRNA lacks any intronic sequence.

[0040] The term "gene" relates to a DNA sequence that encodes or corresponds to a particular amino acid sequence comprising all or part of one or more proteins or enzymes, and may or may not include regulatory DNA sequences, such as promoter sequences, which determine for example the conditions in which the gene is expressed. Some genes, which are not structured genes, may be transcribed from DNA to RNA, but are not translated into an amino acid sequence. Other genes can function as structural gene regulators or as DNA transcription regulators. A gene that encodes a protein of the invention for use in an expression system, if the DNA is genomic or cDNA, can be isolated from any source, particularly using fungal cDNA or a genomic library. Methods for obtaining genes are well known in the art, for example, Sambrook et al. (supra).

[0041] Thus, in a preferred object of the invention, the polynucleotide that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described in the present invention, said encoded polypeptide comprises the amino acid replacements: glycine (G) in position 241 and arginine (R) in position 257, of SEQ ID NO: 2, by the amino acids: aspartic acid (D) and lysine (K), respectively, giving rise to the G241D and R257K mutations in said sequence. In a preferred embodiment of the invention, the polynucleotide described herein further comprises, in addition to the G241D and R257K mutations, an additional amino acid alteration in the homologous position to position 191 of

said SEQ ID NO: 2 which replaces the original amino acid phenylalanine (F) by serine (S), giving rise to the mutation F191S.

[0042] In another particular embodiment of the nucleotide of the invention, it can further comprise the two aforementioned mutations, common to all the UPO mutants obtained in the present invention, or alternatively the three previously described mutations, at least one of the following mutations, whether isolated or in combinations thereof:

a) replacement of the original amino acid leucine (L) by the amino acid phenylalanine (F) in the homologous position to position 67 of SEQ ID NO: 2 (L67F),
b) replacement of the original amino acid isoleucine (I) by the amino acid valine (V) in the homologous position to position 248 of SEQ ID NO: 2 (I248V),
c) replacement of the original amino acid phenylalanine (F) by the amino acid leucine (L) in the homologous position to position 311 of SEQ ID NO: 2 (F311L),
d) replacement of the original amino acid valine (V) by the amino acid isoleucine (I) in the homologous position to position 75 of SEQ ID NO: 2 (V75I), and
e) replacement of the original amino acid valine (V) by the amino acid alanine (A) in the homologous position to 57 of SEQ ID NO: 2 (V57A).

[0043] In another particular embodiment of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, with respect to a wild-type UPO enzyme of SEQ ID NO: 2, or with respect to a variant with UPO activity such as, for example, the PaDa-I variant of SEQ ID NO: 14, as described herein, said encoded polypeptide is characterised in that it can further comprise the nucleotide sequence that encodes the signal peptide of SEQ ID NO: 26.

[0044] In another particular embodiment of the nucleotide of the invention, which encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said polypeptide is characterised in that the polynucleotide sequence encoding the signal peptide of SEQ ID NO: 26, has further at least one of the following additional mutation or any of its combinations:

a) replacement of the amino acid phenylalanine (F) by the amino acid tyrosine (Y) in the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y),
b) replacement of the amino acid alanine (A) by the amino acid valine (V) in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),
c) replacement of the amino acid arginine (R) by the amino acid glycine (G) in the homologous position to position 15 of SEQ ID NO: 26 (R[15]G), and
d) replacement of the amino acid alanine (A) by the amino acid aspartic (D) in the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

[0045] All these mutations and combinations thereof give rise to peroxygenase mutants or variants having a wide spectrum of biotechnological applications, specifically with high functional expression, high monooxygenase activity to the detriment of the peroxidase activity, high thermostability and greater resistance to the presence of organic co-solvents, maintenance of regioselectivity against 1-naphthol, reduction in the ratio 1.4-naphthoquinone:1-naphthol, enhanced catalytic efficiency for naphthalene; additionally, it improves regioselectivity against 5'-hydroxypropranolol up to 99%, reduces the oxidation of 5'-hydroxypropranol up to 50% and enhances catalytic efficiency for propranolol by two orders of magnitude for different applications, with respect to the wild-type UPO or respect to other UPO variants, such as the PaDa-I variant.

[0046] In a preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acidic alterations G241D and R257K, with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 9 that encodes the variant of SEQ ID NO: 10, or with SEQ ID NO: 7 that encodes the variant of SEQ ID NO: 8 (UPO wt-JaWa UPO variants, with and without signal peptide, respectively).

[0047] In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide has the amino acidic alterations G241D, R257K and additionally F191S, with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 63 that encodes the variant of SEQ ID NO: 64, or with SEQ ID NO: 61 that encodes the variant of SEQ ID NO: 62 (UPO wt-SoLo variants, with and without signal peptide, respectively).

[0048] In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide

shows the amino acid alterations G241D and R257K, with respect to SEQ ID NO: 2, or the amino acid alterations G241D, R257K and F191S, with respect to SEQ ID NO: 2, and further comprise the amino acid alterations F[12]Y, A[14]V, R[15] G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polynucleotide of the invention is selected from the list consisting of: SEQ ID NO: 11 that encodes the variant of SEQ ID NO: 12 (UPO mutant wt-JaWa variant with modified signal peptide) and SEQ ID NO: 65 that encodes the variant of SEQ ID NO: 66 (UPO mutant wt-SoLo with modified signal peptide).

[0049] Thus, in another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D, R257K, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2. In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D, R257K, F191S, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 21 that encodes the variant of SEQ ID NO: 22, or with SEQ ID NO: 19 that encodes the variant of SEQ ID NO: 20 (JaWa variants, with and without signal peptide, respectively). In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 39 that encodes the variant of SEQ ID NO: 40, or with SEQ ID NO: 37 that encodes the variant of SEQ ID NO: 38 (SoLo variants, with and without signal peptide, respectively).

[0050] In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D, R257K, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D in the signal peptide of SEQ ID NO: 26. In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with peroxygenase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D, R257K, F191S, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 23 that encodes the variant of SEQ ID NO: 24 (JaWa variant with modified signal peptide). In another particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 41 that encodes the variant of SEQ ID NO: 42 (SoLo variant with modified signal peptide).

[0051] Since the peroxygenases secreted by ligninolytic basidiomycetes fungi may be considered to be related in terms of their evolution, it is to be expected that the global identity of the genes will be 50% or higher and, more specifically, at the level of the amino acid sequence corresponding to SEQ ID NO: 4 or SEQ ID NO: 2 (peroxygenase *Aae*UPO1, with and without signal peptide, respectively), or of the amino acid sequence corresponding to SEQ ID NO: 18 or SEQ ID NO: 14 (peroxygenase PaDa-I, with and without modified signal peptide, respectively), is 70% or higher. The correspondence between the amino acid sequence of the artificial peroxygenase(s) that are the objects of the invention and the sequence of other peroxygenases can be determined by means of method known in the art. For example, they can be determined by direct comparison of the amino acid sequence information of the putative peroxygenase and the amino acid sequence corresponding to SEQ ID NO: 24 or SEQ ID NO: 20 of this specification (JaWa peroxygenase variant, with and without modified signal peptide, respectively) or to SEQ ID NO: 42 or SEQ ID NO: 38 (SoLo peroxygenase variant, with and without modified signal peptide, respectively).

[0052] With the information provided in the present invention, a person skilled in the art is capable of combining the previously described mutation to generate new peroxygenase variants with improved peroxygenase activity and reduced peroxidase activity, in addition to the other functional characteristics mentioned herein.

[0053] Another of the objects described herein relates to a polynucleotide sequence that encodes a polypeptide with peroxygenase activity, characterised in that the amino acid sequence of the polypeptide it encodes shows an identity of at least of 70% with SEQ ID NO: 14 (PaDa-I), and in that it comprises at least two amino acid alterations in the homologous positions to positions 241 and 257 of said sequence, replacing the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K). In a preferred embodiment, the polynucleotide sequence that encodes a polypeptide as described herein further comprises an additional amino acid alteration in the homologous position to position 191 of said sequence SEQ ID NO: 14, replacing the original amino acid phenylalanine (F) by serine (S) in position 191 (F191S).

[0054] Alternatively, another of the objects described in the present invention relate to a polynucleotide sequence that encodes a polypeptide with peroxygenase activity, characterised in that the amino acid sequence of the polypeptide shows an identity of at least 70% with SEQ ID NO: 14 (PaDa-I), and which comprises the amino acids alanine (A), phenylalanine (F), isoleucine (I), valine (V) and leucine (L) in positions 57, 67, 75, 248 and 31, respectively, with respect to SEQ ID NO: 14, characterised in that it further comprises two amino acid alterations in the homologous positions to positions 241 and 257 of said sequence, replacing the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K) and optionally, it may further comprise an

additional amino acid alteration in position 191 of said sequence SEQ ID NO: 14, which replace the original amino acid phenylalanine (F) by serine (S) (F191S).

[0055] In a particular embodiment of the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, with respect to a variant with UPO activity of SEQ ID NO: 14, as described herein, said encoded polypeptide is characterised in that it can further comprise the nucleotide sequence that encode the signal peptide of SEQ ID NO: 26.

[0056] In another particular embodiment of the nucleotide of the invention, which encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said polypeptide is characterised in that the polynucleotide sequence that encodes the signal peptide of SEQ ID NO: 26, has further at least one of the following additional mutation or any of its combinations:

a) replacement of the amino acid phenylalanine (F) by the amino acid tyrosine (Y) in the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y),
b) replacement of the amino acid alanine (A) by the amino acid valine (V) in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),
c) replacement of the amino acid arginine (R) by the amino acid glycine (G) in the homologous position to position 15 of SEQ ID NO: 26 (R[15]G), and
d) replacement of the amino acid alanine (A) by the amino acid aspartic (D) in the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

[0057] All these mutations give rise to mutants or variants of the peroxygenases with a wide spectrum of biotechnological applications, specifically with high functional expression, high monooxygenase activity and low peroxidase activity, high thermostability, greater resistance to the presence of organic co-solvents, maintenance of regioselectivity against 1-naphthol, decrease in the ratio 1.4-naphthoquinone:1-naphthol, enhanced catalytic efficiency for naphthalene, decreasing oxidation by up to 50% on 5'-hydroxypropranolol, enhancement of catalytic efficiency by two orders of magnitude, for different applications, with respect to the PaDa-I variant of SEQ ID NO: 18.

[0058] In a preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D and R257K, with respect to SEQ ID NO: 14. In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with the characteristics and advantages mentioned herein, said encoded polypeptide shows the amino acid alterations G241D, R257K and F191S, with respect to SEQ ID NO: 14. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 21 that encodes the variant of SEQ ID NO: 22, or with SEQ ID NO: 19 that encodes the variant of SEQ ID NO: 20 (UPO JaWa variant, with and without signal peptide, respectively). In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 39 that encodes the variant of SEQ ID NO: 40, or with SEQ ID NO: 37 that encodes the variant of SEQ ID NO: 38 (SoLo UPO variant, with and without signal peptide, respectively).

[0059] In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D and R257K, with respect to SEQ ID NO: 14, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 23 that encodes the variant of SEQ ID NO: 24 (UPO JaWa variant with modified signal peptide).

[0060] In another preferred embodiment of the invention, the polynucleotide of the invention that encodes a polypeptide with improved peroxygenase activity and reduced peroxidase activity, as described herein, said encoded polypeptide shows the amino acid alterations G241D, R257K and F191S, with respect to SEQ ID NO: 14, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polynucleotide of the invention corresponds to SEQ ID NO: 41 that encodes the variant of SEQ ID NO: 42 (UPO SoLo variant with modified signal peptide).

[0061] As mentioned earlier, with the information supplied in the present invention, a person skilled in the art is capable of combining the previously described mutations to generate new peroxygenase variants with improved peroxygenase activity and reduced peroxidase activity, in addition to the other functional characteristics mentioned herein.

[0062] Another object described in the present invention relates to the amino acid sequence encoded by the polynucleotide of the invention, hereinafter polypeptide of the invention, characterised in that it shows a sequence identity of at least 70% with SEQ ID NO: 2 (*Aae*UPO1, without signal peptide) and because it comprises at least two amino acid alterations, preferably replacements, in the homologous positions to positions 241 and 257 of said sequence, which replace the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K).

[0063] In a preferred embodiment, the polypeptide of the invention further comprises an additional amino acid alteration,

preferably a replacement, in the homologous position to position 191 of SEQ ID NO: 2, which replace the original amino acid phenylalanine (F) by serine (S) in position 191 (F191S).

[0064] The term "peptide", "polypeptide" or "protein", as used in the description, relates to a polymeric form of amino acids of any length.

[0065] Thus, in a preferred aspect of the invention, the replacements of the amino acids: glycine (G) in position 241 and arginine (R) in position 257 of SEQ ID NO: 2, by the amino acids aspartic acid (D) and lysine (K), respectively, gives rise to the G241D and R257K mutations, respectively, obtaining the wt-JaWa variant of SEQ ID NO: 8.

[0066] In another preferred aspect of the invention, the replacement of the amino acid phenylalanine (F) in position 191 of SEQ ID NO: 2, by the amino acid serine (S), gives rise to the F191S mutation, obtaining the wt-SoLo variant of SEQ ID NO: 62.

[0067] The polypeptide of the invention can also show additional mutations to those mentioned earlier that improve its activity and stability, both thermal and in the presence of different co-solvents and their functional expression in heterologous organisms. Additionally, the variants with improved peroxygenase activity and reduced peroxidase activity, show an increase in TTN of approximately 2.5 fold, an increase in k_{cat} for 1-naphthol of up to 1.5 fold, enhanced catalytic efficiency for naphthalene of up to $6.2 \times 10^5\ s^{-1}\ M^{-1}$, a decrease of approximately 1.5 fold in the ratio 1.4-naphthoquinone:1-naphthol, and regioselectivity against 1-naphthol of approximately 97%, in addition to an increase in TTN for the synthesis of 5'-hydroxypropranolol of 3 fold in the absence of antioxidants (45,000 for the SoLo mutant against 15,000 of the JaWa mutant) or of 15 fold (3,000 in the case of wild AaeUPO), an increase in k_{cat} for 5'-hydroxypropranolol of up to 3.6 fold, enhanced catalytic efficiency for 5'-hydroxypropranolol of up to $3.1 \times 10^6\ s^{-1}\ M^{-1}$, two orders of magnitude higher than those of any enzyme described and show 50% less oxidation on 5'-hydroxypropranolol. These mutations described earlier in this invention can show various combinations jointly with the mutation described earlier, as is known to a person skilled in the art.

[0068] In a preferred aspect, the polypeptides of the present invention show an improvement of at least 20%, for example, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or at least 100% in the peroxygenase activity with respect to the peroxygenase activity of the mature polypeptide of SEQ ID NO: 4, SEQ ID NO: 2, SEQ ID NO: 14 or SEQ ID NO: 18.

[0069] In a preferred aspect, the polypeptides of the present invention show a reduction of at least 20%, for example, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or at least 100% in the peroxydase activity with respect to the peroxygenase activity of the mature polypeptide of SEQ ID NO: 4, SEQ ID NO: 2, SEQ ID NO: 14 or SEQ ID NO: 18.

[0070] In a particular embodiment of the polypeptide of the invention, it may comprise, in addition to the two aforementioned mutations, common to all the UPO mutants obtained and described in the present invention, an additional mutation comprising the replacement of the original amino acid phenylalanine (F) by the amino acid serine (S) in the homologous position to position 191 of SEQ ID NO: 2 (F191S).

[0071] In a particular embodiment of the polypeptide of the invention, it can comprise, in addition to the aforementioned mutations, whether isolated or in combinations thereof:

a) replacement of the original amino acid leucine (L) by the amino acid phenylalanine (F) in the homologous position to position 2 of SEQ ID NO: 2 (L67F),
b) replacement of the original amino acid isoleucine (I) by the amino acid valine (V) in the homologous position to position 248 of SEQ ID NO: 2 (I248V),
c) replacement of the original amino acid phenylalanine (F) by the amino acid leucine (L) in the homologous position to position 311 of SEQ ID NO: 2 (F311L),
d) replacement of the original amino acid valine (V) by the amino acid isoleucine (I) in the homologous position to position 75 of SEQ ID NO: 2 (V75I), and
e) replacement of the original amino acid valine (V) by the amino acid alanine (A) in the homologous position to 57 of SEQ ID NO: 2 (V57A).

[0072] In another preferred embodiment of the polypeptide of the invention, it is characterised in that can further comprise the sequence that encodes the signal peptide of SEQ ID NO: 26.

[0073] In another preferred embodiment of the polypeptide of the invention, it is characterised in that it also has at least one of the following additional mutations or any of its combinations in the nucleotide sequence that encodes the signal peptide of SEQ ID NO: 26:

a) replacement of the amino acid phenylalanine (F) by the amino acid tyrosine (Y) in the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y)

b) replacement of the amino acid alanine (A) by the amino acid valine (V) in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),

c) replacement of the amino acid arginine (R) by the amino acid glycine (G) in the homologous position to position 15 of SEQ ID NO: 26 (R[15]G), and

d) replacement of the amino acid alanine (A) by the amino acid aspartic (D) in the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

[0074] All these mutations give rise to mutants or variants of the peroxygenases with a wide spectrum of biotechnological applications, specifically with high functional expression, high monooxygenase activity and low peroxidase activity, high thermostability, greater resistance to the presence of organic co-solvents, greater regioselectivity and an increase in TTN, for different applications, with respect to the wild-type UPO, or with respect to other UPO mutants such as, for example, the PaDa-I mutant.

[0075] Thus, in a preferred embodiment of the invention, the polypeptide has amino acid alterations G241D and R257K with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 10 or of SEQ ID NO: 8 (wt-JaWa variant, with and without signal peptide, respectively).

[0076] Thus, in a preferred embodiment of the invention, the polypeptide shows the amino acid alteration F191S with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 64 or of SEQ ID NO: 62 (wt-SoLo variant, with and without signal peptide, respectively).

[0077] In another preferred embodiment of the invention, the polypeptide show the amino acids alterations G241D and R257K with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 12 (wt-JaWa variant, with modified signal peptide). In another preferred embodiment of the invention, the polypeptide shows the amino acid alteration F191S with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 66 (wt-SoLo variant, with modified signal peptide).

[0078] Thus, in another preferred embodiment of the invention, the polypeptide of the invention has the amino alterations G241D, R257K, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 22 or of SEQ ID NO: 20 (JaWa variant, with and without signal peptide, respectively).

[0079] Thus, in another preferred embodiment of the invention, the polypeptide of the invention has the amino acid alterations G241D, R257K, F191S, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 40 or of SEQ ID NO: 38 (SoLo variant, with and without signal peptide, respectively).

[0080] In another preferred embodiment of the invention, the polypeptide shows the amino acid alterations G241D, R257K, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 24 (JaWa variant, with modified signal peptide).

[0081] In another preferred embodiment of the invention, the polypeptide shows the amino acid alterations G241D, R257K, F191S, L67F, I248V, F311L, V57A and V75I, with respect to SEQ ID NO: 2, and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 42 (SoLo variant, with modified signal peptide).

[0082] With the information provided in the present invention, a person skilled in the art is capable of combining the previously described mutations to generate new peroxygenase variants with improved peroxygenase activity and reduced peroxidase activity and greater stability, in addition to comprising the functional characteristics mentioned throughout this specification.

[0083] Another object described in the present invention relates to the amino acid sequence that encodes the polynucleotide of the invention, characterised in that its sequence show an identity of at least 70% with SEQ ID NO: 14 (PaDa-I, without signal peptide), and in that it comprises at least two amino acid alterations, preferably replacements, in the homologous positions to positions 241 and 257 of said sequence, replacing the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K). In a preferred embodiment, the amino acid sequence encoded by the polynucleotide of the invention further comprises an additional amino acid alteration, preferably a replacement, in the homologous position to position 191 of said sequence SEQ ID NO: 14, replacing the original amino acid phenylalanine (F) by serine (S) in position 191 (F191S).

[0084] Alternatively, the present invention also relates to the amino acid sequence coded by the polynucleotide of the invention, characterised in that it shows a sequence identity of at least 70% with SEQ ID NO: 14 (PaDa-I), and which

comprises the amino acids alanine (A), phenylalanine (F), isoleucine (I), valine (V) and leucine (L) in positions 57, 67, 75, 248 and 311, respectively, with respect to SEQ ID NO: 14, characterised in that it further comprises two amino acid alterations in homologous positions to positions 241 and 257 of said sequence, replacing the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D) and original arginine (R) by lysine (K) in position 257 (R257K). Also alternatively, the amino acid sequence coded by the polynucleotide of the invention, characterised in that it shows a sequence identity of at least 70% with SEQ ID NO: 14 (PaDa-I), and which comprises the amino acids alanine (A), phenylalanine (F), isoleucine (I), valine (V) and leucine (L) in positions 57, 67, 75, 248 and 311, respectively, with respect to SEQ ID NO: 14, characterised in that it further comprises at least three amino acid alterations in homologous positions in positions 241, 257 and 191 of said sequence, replacing the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D), original arginine (R) by lysine (K) in position 257 (R257K) and original phenylalanine (F) by serine (S) (F191S).

[0085] Thus, in a preferred aspect of the invention, the replacements of the amino acids: glycine (G) in position 241 and arginine (R) in position 257 of SEQ ID NO: 14, by the amino acids aspartic acid (D) and lysine (K), respectively, gives rise to the G241D and R257K mutations, respectively, obtaining the JaWa variant of SEQ ID NO: 20.

[0086] In another preferred aspect of the invention, the replacements of the amino acids: glycine (G) in position 241, arginine (R) in position 257 and phenylalanine (F) in position 191 of SEQ ID NO: 14, by the amino acids aspartic acids (D), lysine (K) and serine (S), respectively, gives rise to the G241D, R257K and F191S mutations, respectively, obtaining the SoLo variant of SEQ ID NO: 38.

[0087] In another preferred embodiment of the polypeptide of the invention, it is characterised in that it can further comprises the sequence that encodes the signal peptide of SEQ ID NO: 26.

[0088] In another preferred embodiment of the polypeptide of the invention, it is characterised in that has further at least one of the following additional mutations or any of its combinations in the nucleotide sequence that encodes the signal peptide of SEQ ID NO: 26:

a) replacement of the amino acid phenylalanine (F) by the amino acid tyrosine (Y) in the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y),
b) replacement of the amino acid alanine (A) by the amino acid valine (V) in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),
c) replacement of the amino acid arginine (R) by the amino acid glycine (G) in the homologous position to position 15 of SEQ ID NO: 26 (R[15]G), and
d) replacement of the amino acid alanine (A) by the amino acid aspartic (D) in the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

[0089] As mentioned earlier, all these mutations give rise to peroxygenase mutants or variants with a wide spectrum of biotechnological applications, specifically with high functional expression, high monooxygenase activity and low peroxidase activity, high thermostability, greater resistance to the presence of organic co-solvents, greater regioselectivity and increase in TTN, for different applications, with respect to the PaDa-I variant.

[0090] Thus, in a preferred embodiment of the invention, the polypeptide shows the amino acid alterations G241D and R257K with respect to SEQ ID NO: 14. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 22 or SEQ ID NO: 20 (JaWa variant, with and without signal peptide, respectively).

[0091] In another preferred embodiment of the invention, the polypeptide shows the amino acid alterations G241D, R257K and F191S with respect to SEQ ID NO: 14. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 40 or SEQ ID NO: 38 (SoLo variant, with and without signal peptide, respectively).

[0092] In another preferred embodiment of the invention, the polypeptide show the amino acids alterations G241D and R257K with respect to SEQ ID NO: 14, also alternatively shows the alteration F191S and further comprises the amino acid alterations F[12]Y, A[14]V, R[15]G and A[21]D, in the signal peptide of SEQ ID NO: 26. In a particular embodiment of the invention, the polypeptide of the invention corresponds to the peptide of SEQ ID NO: 24 or with the peptide of SEQ ID NO: 42 (JaWa or SoLo variants, with modified signal peptide, respectively).

[0093] Another object described in the present invention relates to the use of the polypeptide of the invention in methods of organic synthesis, preferably in processes of oxyfunctionalisation or selective oxidation of hydrocarbon in general, both aromatic and linear aliphatic, branched and cyclic (alkanes such as propane, 2,3-dimethylbutane or cyclohexane, fatty acids such as lauric acid), linear, branched and cyclic unsaturated hydrocarbonated chains (olefins such as propene, 2-methyl-2-butene or limonene), more preferably in the production of 1-naphthol for applications in the textile industry (dyes), agrochemicals (herbicides, pesticides) or in bioremediation, more preferably in the production of HDMs and even more preferably in the production of 5'-hydroxypropranolol. Also for cosmetic and/or food applications, synthesis of metabolites for drugs or pharmaceutical compositions, other bioremediation processes, preferably, transformation of

recalcitrant PAHs (polycyclic aromatic hydrocarbons) into less-polluting derivatives, biosensor design, preferably, immunoassays for detection by means of chemoluminescence and in the manufacture of bioelectronic devices containing immobilised enzymes. Additionally, the polypeptides described in the present invention can transform any compound that is a substrate of AaeUPO, such a for example: O- and N- can dealkylate compounds such as tetrahydrofurane or lidocaine, respectively; heterocyclic compounds showing sulphur or nitrogen atoms in their structure, wherein said compounds may be S- or N-oxygenated, as in the case of dibenzothiophene or pyridine, respectively.

[0094] The polynucleotide of the invention can be found isolated as such or forming part of gene constructions or vectors which allow the propagation of said polynucleotides in suitable host cells. Such gene expression vectors include control sequences such as, for example, translation (such as start and stop codes) and transcription (for example, promoter-operator regions, binding sites) control elements. The vectors according to the invention may include bacterial plasmids and viral vectors, and other vectors in accordance with the well-known and documented methods in the state of the art, and can be expressed in a variety of different expression systems, also well known and documented. A variety of techniques that can be used to introduce such vectors in prokaryotic or eukaryotic cells (host cells) for expression thereof are also known. Suitable transformation or transfection techniques are well known to the person skilled in the art and are described in the state of the art. Therefore, in another aspect, the invention relates to a vector, hereinafter vector of the invention, that comprises the polynucleotide of the invention as described earlier.

[0095] The term "nucleic acid construction" as used herein relates to a nucleic acid molecule - single or double-stranded- which is isolated from a naturally occurring gene or which is modified to contain nucleic acid segments in such a manner that it would not do otherwise should it occur naturally or that is synthetic. The term "nucleic acid construction" is synonymous of the term "expression cassette" when the nucleic acid construct contains the control sequence required for the expression of an encoding sequence of the present invention.

[0096] The terms "vector" or "expression vector" relate to the vehicle whereby a DNA or RNA sequence (for example, a heterologous gene) can be introduced in a host cell, for the purpose of transforming the host and promoting the expression (for example, transcription and translation) of the sequence introduced. The vectors typically comprise the DNA of a transmissible agent, wherein the foreign DNA encodes a protein inserted using restriction enzyme technology. A common type of vector is a "plasmid", which is generally a double-stranded DNA molecule, which can easily accept additional DNA (foreign) and that can be easily introduced in a suitable host cell. A large number of vectors, including plasmidic and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. Non-limiting examples include pKK plasmids (Clonetech), pUC plasmids, pET plasmids (Novagen, Inc., Madison, WI), pRSET or PrEP plasmids (Invitrogen, San Diego, CA), pMAL plasmids (New England Biolabs, Beverly, MA), pGAPZaA, pcWori+, pET-26b (+), pXTD14, pYEX-S1, pMAL and pET22-b (+), or the plasmid used in the present invention, pJRoC30, donated by Dr. Frances Arnold, of the Californian Institute of Technology (CALTECH, USA). Recombinant clonation vectors often include one or more replication systems for cloning or expression, one or more markers for selection in the host, for example, resistance to antibiotics, and one or more expression cassettes. Suitable vectors for insertion of said polynucleotide are vectors derived from expression vectors in prokaryotes such as, by way of example, pUC18, pUC19, Bluescript and its derivatives, mp18, mp19, pBR322, pMB9, Co1E1, pCR1, RP4, phages and "launch" vectors, such as pSA3 and pAT28; expression vectors in yeasts such as the 2 micron plasmid of *S. cerevisiae,* integration plasmids, YEP vectors, centromere and similar plasmids; expression vectors in insect cells such as pAC series vectors and pVL series expression vectors; expression vectors in plant cells such as piBi, pEarleyGate, PAVA, pCAMBIA, PGSA, PGWB, PMDC, PMY, pore and similar series, and other expression vectors in eukaryotic cells, including baculovirus suitable for transfection of insect cells using any commercially available baculovirus system. Other vectors can be used as desired by a person skilled in the art. Routine experimentation in biotechnology can be used to determine the most suitable vectors for use with the invention, if different to that described in the Examples. In general, the choice of the vector depends on the size of the polynucleotide and of the host cell to be used in the methods of this invention.

[0097] The term "control sequences" is defined herein to include all the necessary components for the expression of the polypeptide coding sequences of the present invention. Each control sequence may be native or foreign to the nucleotide sequence that encodes the native or foreign polypeptide therebetween. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, pro-peptide sequence, promoter, signal peptide sequence and transcription terminator. The control sequences include, at least, a promoter and translation and transcription stop signals.

[0098] The control sequences may have links in order to introduce specific restriction sites that facilitate the linkage of the control sequences with the coding region of the nucleotide sequence that encodes a polypeptide.

[0099] A "promoter sequence" is a DNA regulatory region capable of binding to the RNA polymerase in a cell and initiating the transcription of a gene (direction 3') downstream from the coding sequence. For the purpose of defining this invention, the promoter sequence is limited at its 3' terminus by the transcription start site and extends upstream (5' direction) to include the minimum number of necessary bases or elements to begin the transcription at detectable levels above the base.

[0100] The expression "operationally linked" relates to a juxtaposition wherein the components thus described have

a relationship that allows them to function intentionally. A control sequence "operationally linked" to a coding sequence is linked in such a manner that the expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0101] In a preferred embodiment, the genetic construction of the invention further comprises a polynucleotide that encodes a signal peptide enhanced by directed evolution which favours the functional expression of the polypeptide of the invention.

[0102] The term "signal peptide", as used in the description, relates to a peptide which is located at the amino end of a polypeptide or protein, and whose function is to direct the localisation of the protein at different compartments of the cell (nucleus, mitochondria, chloroplast, endoplasmic reticulum (ER), Golgi apparatus (GA), etc.) or to the extracellular space, in the case that the protein is secreted.

[0103] The signal peptide of the factor α is a polypeptide with 83 amino acids. The first 19 amino acids constitute the pre-leader that directs the polypeptide being created towards the ER. After entering the ER, the pre-leader is cleaved by a peptidase, giving rise to a pro-protein. At this point, the N-glycosylations of three asparagine residues facilitate the transit of the pro-protein of the ER to the GA. In the GA, the pro-leader can act as a chaperone until it is processed by the proteases KEX1, KEX2 and STE13 (M.A. Romanos, et al., 1992. Yeast 8, 423-488; J. R. Shuster, 1991. Curr. Opin. Biotechnol. 2, 685-690). Additionally, the pro-leader seems to be involved in an indicated vacuolar process, which is detrimental to heterologous secretion (J. A. Rakestraw, et al. Biotechnol. Bioeng. 2009. 103, 1192-1201).

[0104] Preferably, the signal peptide is that of the *Aae*UPO1 of the nucleotide sequence SEQ ID NO: 25 which encodes the amino acid sequence SEQ ID NO: 26. In a more preferred embodiment, the signal peptide comprises at least one of the following mutations or any combination thereof:

a) the replacement of the original phenylalanine (F) amino acid by the tyrosine (Y) amino acid in the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y),
b) the replacement of the original alanine (A) amino acid by the valine (V) amino acid in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),
c) the replacement of the original arginine (R) amino acid by the glycine (G) amino acid in the homologous position to position 15 of SEQ ID NO: 26 (R[15]G), and
d) the replacement of the original alanine (A) amino acid by the aspartic acid (D) amino acid in the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

[0105] In another even more preferred embodiment, the signal peptide of the invention corresponds to the peptide sequence SEQ ID NO: 28, encoded by the nucleotide sequence SEQ ID NO: 27. Said signal peptide favours the functional expression of the polypeptide of the invention.

[0106] Another object described in the present invention relates to a host cell characterised in that it comprises the nucleotide of the invention and is capable of producing the polypeptide of the invention as described throughout the present document.

[0107] As used in the present specification, a "host cell" includes any culturable cell that can be modified through the introduction of DNA not contained naturally in the cell, hereinafter host cell of the invention. Preferably, a host cell is that in which the nucleotide of the invention can be expressed, giving rise to a stable, post-translationally modified polypeptide located in the appropriate subcellular compartment. The choice of an appropriate host cell can also be influenced by the choice of the detection signal.

[0108] For example, the use of constructions with reporter genes (for example, lacZ, luciferase, thymidine kinase or GFP) can provide a selectable signal by activating or inhibiting the transcription of the gene of interest in response to a transcription-regulating protein. The phenotype of the host cell must be considered in order to achieve an optimal selection or screening.

[0109] A host cell of the present invention includes prokaryotic and eukaryotic cells. Prokaryotes include gram-negative organisms (for example, *Escherichia coli*) or gram-positive organisms (for example, bacteria of the genus *Bacillus sp.*). Prokaryotic cells are used, preferably, to propagate the transcription-control sequence of the vector that contains the polynucleotide(s) of the invention, which will make it possible to obtain a larger number of copies of the vector containing the polynucleotide(s) that is/are the object of the invention. The appropriate prokaryotic host cells for transforming this vector include, for example, *E. coli, Bacillus subtilis, Salmonella typhimurium* and other species within the genera *Pseudomonas, Streptomyces and Staphylococcus.* Eukaryotic cells include, *inter alia,* yeast cells, plant cells, fungus cells, insect cells, mammal cells and parasite organism cells (for example, *Trypanosomas*). As used herein, the term yeast does not include only yeast in the strictly taxonomic sense, *i.e.* unicellular organisms, but also multicellular fungi similar to yeasts or filamentous fungi. Examples of species include *Kluyveromyces lactis, Schizosaccharomyces pombe* and *Ustilago maydis,* with *S. cerevisiae and P. pastoris* as preferred organisms. Other yeasts that can be used in the production of the polyamino acid sequence(s) of the present invention are *Neurospora crassa, Aspergillus niger, A. nidulans, A. sojae, A. oryzae, Candida tropicalis and Hansenula polymorpha.* Mammal host cell culture systems include established

cell lines such as COS cells, L cells, 3T3 cells, Chinese hamster ovarian cells (CHO), embryonic stem cells, with BHK, HeK or HeLa cells such as preferred cells. Eukaryotic cells are, preferably, used for the expression of the recombinant gene through the application of the transcription regulation sequence or the expression vector of the present invention.

[0110] Brewer's yeast S. *cerevisiae* is a unicellular fungus that belongs to the Superkingdom Eukarya (Metazoa/Fungi group), Kingdom *Fungi,* Subkingdom *Dikarya,* Phylum *Ascomycota,* Subphylum *Saccharomycotina,* Class *Saccharomycetes,* Order *Saccharomycetales,* Family *Saccharomycetaceae* and Genus *Saccharomyces.*

[0111] The methylotrophic yeast *P. pastoris* belongs to the Superkingdom *Eukarya,* (Metazoa/Fungi group), Kingdom *Fungi,* Subkingdom *Dikarya,* Phylum *Ascomycota,* Subphylum *Saccharomycotina,* Class *Saccharomycetes,* Order *Saccharomycetales,* Family *Saccharomycetaceae* and Genus *Komagataella.*

[0112] Another aspect described in the present invention relates to the method for obtaining the polypeptide of the invention, which comprises the following steps:

a) Introducing the vector of the invention, as described earlier, in an appropriate host cell (host cell of the invention),
b) culturing the host cell of the invention in an appropriate medium, and
c) purifying the polypeptide of the invention with improved peroxygenase activity and reduced peroxidase activity, with respect to the same activities of a wild-type AaeUPO enzyme or of a variant with UPO activity such as, for example, the PaDa-I variant.

[0113] The terms "purify", "isolate", "isolation" or "purification" of the polypeptides or enzymes described in the present invention relate to the separation of the peptides of the invention and, alternatively, to their concentration, as of the culture medium of the cell of the invention. The methods for separating and purifying polypeptides are well known in the art, without limitation, differential solubility, chromatography, electrophoresis or isoelectrofocus techniques. For some purposes, it is preferable to produce the polypeptide in a recombinant system wherein the protein contains an additional sequence ticket that facilitates the purification, such as, but not limited to, polyhistidine. Chromatography techniques can be based on the molecular weight, load or affinity of the protein and can be performed in a column, on paper or in a plate. Protein separation can be performed, for example, using Fast Protein Liquid Chromatography (FPLC), in an automated system that significantly reduces purification time and enhances purification performance.

[0114] Another aspect of the invention relates to a host cell culture of the invention.

[0115] A host cell culture relates to the process of maintaining and growing the host cells. Cell cultures require controlled conditions: temperature, pH, gas percentages (oxygen and carbon dioxide), in addition to the presence of appropriate nutrients to allow cellular viability and division. Cell cultures can be developed in solid substrates such as agar, or in liquid medium, which makes it possible to culture large amounts of cells in suspension.

[0116] Another object of the invention relates to the use of the host cell of the invention, or of the host cell culture of the invention, to obtain the polypeptide of the invention. Preferably, the host cell of the invention is a yeast, more preferably of the genera *Saccharomyces* sp. or *Pichia* sp and, even more preferably, the species are *Saccharomyces cerevisiae* or *Pichia pastoris.*

[0117] Peroxygenases, as in the case of the polypeptides of the invention, are known for their large number of applications such as, for example, their use in organic synthesis, preferably in processes of oxyfunctionalisation, oxidation or selective hydroxylation of hydrocarbons in general, both aromatic and aliphatic, linear, branched and cyclic, preferably the method of hydroxylation of cyclic aromatic compounds, both simple or condensed cyclic compounds, more preferably a method of hydroxylation for the synthesis of 1-naphthol and/or synthesis of 5'-hydroxypropranolol, limonene derivatives for cosmetic and/or nutritional applications, synthesis of drug metabolites or pharmaceutical compositions, synthesis of 1-naphthol for dyes, herbicides or pesticides, bioremediation (transformation of recalcitrant PAHs) and biosensor design (chemoluminescence detection immunoassays). Thus, the polypeptide of the invention and the host cell of the invention may have any of the currently known uses for these enzymes in the state of the art.

[0118] Another aspect of the invention relates to the use of the polynucleotide of the invention, or of the vectors, or genetic constructions of the invention, or of the host cell of the invention, for obtaining enzymes with improved peroxygenase activity and reduced peroxidase activity, which show a high production rate, high regioselectivity, preferably against 1-naphthol and/or against propranolol, and high thermostability with respect to the wild-type or native *Aae*UPO1 peroxygenase expressed in the yeast, or with respect to UPO variants such as, for example, the PaDa-I variant.

[0119] Thus, another object of the invention relates to the use of the polypeptide of the invention in the manufacture of diagnosis/prognosis kits for biomedical purposes for detecting metabolites and measuring their concentration in, for example, blood, saliva, tear and/or urine samples.

[0120] Another particular object of the invention relates to the use of the polypeptide of the invention in the manufacture of electronic devices containing immobilised enzymes for, for example, biomedical diagnosis by detecting metabolites and measuring their concentration *in vivo* through, by way of example, wireless nanodevices that work on different physiological fluids (blood, saliva, tears and/or urine).

[0121] Diagnosis kits for biomedical purposes and electronic devices containing immobilised enzymes, specifically

the polypeptides described in the present invention, also form part of the invention.

[0122] Thus, another object described in the present invention relates to a kit or to an electronic device comprising at least one polypeptide as described in the present invention.

[0123] Another object described in the present invention relates to methods of organic synthesis, preferably in processes of oxyfunctionalisation, oxidation or selective hydroxylation of hydrocarbons in general, both aromatic and aliphatic, linear, branched and cyclic, preferably the method of hydroxylation of cyclic aromatic compounds, of both simple or condensed cyclic compounds, more preferably the method of hydroxylation for the synthesis of 1-naphthol and/or synthesis of 5'-hydroxypropranolol, through the use of variants, of the host cell, of the kit, or of the device of the invention.

[0124] Throughout the description and the claims, the word "comprises" and its variants are not intended to exclude other technical characteristics, additives, components or steps. For the persons skilled in the art, other objects, advantages and characteristics of the invention will be inferred partly from the description and partly from the practice of the invention. The following examples and drawings are provided by way of example of the invention and are not intended to limit the present invention.

BRIEF DESCRIPTION OF THE FIGURES

[0125]

Figure 1 Directed evolution of *Aae*UPO1. From cycles 1 to 5, the enzyme was improved in terms of functional expression and activity (the accumulated mutations are detailed as light grey rectangles). Starting from the parental AaeUPO, it was subjected to five directed evolution cycles until obtaining the PaDa-I mutant, which was subjected to two more cycles of directed evolution, in this case to improve the production capacity of 1-naphthol (the new mutations appear as black rectangles), and three further cycle grouped together in a single generation to improve the production of 5'-hydroxypropranolol. The activities (as a %) stem from measurements using microcultures of *S. cerevisiae* in 96-well microplates of the second re-screening. Thermostability (T_{50}) was determined using flask culture supernatants: n.m. not measurable, n.d. not determined.

Figure 2 Biochemical characteristics of the variants of the invention. A) Spectroscopic characteristics of the PaDa-I (thin line) and JaWa (thick line) mutants at rest. AU, arbitrary units. **B)** Thermostability analysis (T_{50}) of the PaDa-I (black circles) and JaWa (white circles) mutants. The experiments were carried out using culture supernatants and each point represents the average value and standard deviation of three individual experiments. **C)** Stability of the PaDa-I (black bars) and JaWa (grey bars) mutants at high acetonitrile concentrations. The stabilities were determined after 5 hours of incubation of the enzyme in increasing concentrations of the co-solvent (from 50% to 100%) at 20°C in 10 mM pH 7.0 potassium phosphate buffer. After that time, aliquots were taken and analysed using ABTS substrate (100 mM pH 4.0 sodium phosphate/citrate buffer, 2 mM H_2O_2 and 0.3 mM ABTS). The error bars indicate standard deviations.

Figure 3 Transformation of naphthalene by means of the variants described in the invention. A) Products formed after 15 minutes of reaction stopped with 20 μL of HCl 37% (PaDa-I, black bars; JaWa, grey bars). The reactions were carried out at room temperature using 6.6 nM of pure enzyme, 100 mM pH 7.0 of potassium phosphate buffer, 1 mM naphthalene, 20% acetonitrile and 1 mM H_2O_2 (1 mL of final volume). As can be observed in the figure, the products obtained were mainly naphthalene, 1-naphthol and 2-naphthol. **B)** Chromatograms of the naphthalene transformation reaction after 270 minutes (1: naphthalene; 2: 1-naphthol; 3: 2-naphthol and 4: 1.4-naphthoquinone (1.4-NQ)). **C)** and **D)** Monitoring of the reaction for 270 minutes (without adding HCl) for the PaDa-I (C) and JaWa (D) mutants. Black circles: naphthalene; white circles: 1.2-naphthalene oxide; white squares: 1-naphthol and black squares: 2-naphthol. Total turnover numbers (TTN, expressed as μmoles of product/μmoles of enzyme) were calculated using the production value of 1-naphthol after 270 minutes.

Figure 4 Conversion of naphthalene at 1-naphthol by means of the PaDa-I and JaWa variants. The reactions were performed at room temperature and their composition was as follows: 40 nM of pure enzyme, 100 mM pH 7.0 potassium phosphate buffer, 1 mM naphthalene, 20% acetonitrile and 1mM H_2O_2 (1 mL of final volume). 1-N: 1-naphthol; 1,4-NQ: 1-4-naphthoquinone. Each reaction was performed in triplicate and were stopped with HCl (pH<1) at different times (between 60 and 600 s). *Inset:* polymeric colorimetric products derived from 1.4-naphthoquinone, 1: PaDa-I and 2: JaWa.

Figure 5 W24F variants obtained by means of directed mutagenesis. A) Model built on the crystal structure of the *Aae*UPO1 enzyme (PDB access number: 2YOR), comprising the mutations of the JaWa variant as well as the W24F modification with respect to wild *Aae*UPO1. The model is shown without a surface, with a transparent surface and with an opaque surface, showing position W24. **B)** Activity of the W24F variants using different substrates with respect to their respective parentals, relativised to the PaDa-I activity. The experiments were carried out using 100 mL flask culture supernatants. The buffer used was 100 mM pH 7.0 potassium phosphate buffer, except for the ABTS, in which case 100 mM pH 4.0 sodium phosphate/citrate was used. The components of the mixture were: 0.5

mM naphthalene, 1 mM NBD, 3 mM DMP and 0.3 mM ABTS. In all cases, 1 mM H_2O_2 and 15% acetonitrile were added to the mixtures. For the activity with naphthalene, the Fast Red method was applied (after 10 minutes of reaction, Fast Red was added -final concentration 0.5 mM- and when the red colour appeared and became stabilised, final absorbance was measured). The molar extinction coefficients are: naphthalene+Fast Red, ε_{510}=4,700 M^{-1} cm^{-1}; NBD, ε_{425}=9,700 M^{-1} cm^{-1}; DMP, ε_{469}=27,500 M^{-1} cm^{-1} and ABTS, ε_{418}=36,000 M^{-1} cm^{-1}.

Figure 6 Mutations in the UPO variants described in the invention. Model built on the structure of the *Aae*UPO1 crystal (PDB access number: 2YOR). **A)** PaDa-I; **B)** JaWa. The V248 mutant stems from the previous evolution pathway. The phenylalanine (Phe) residues are responsible for the accommodation of the substrates in the catalytic pocket, the Cys36 residue is the axial heme ligand; R189 is a component of the acid-base pair involved in the catalysis, and heme Fe^{3+} is represented as a sphere.

Figure 7 Protein model of A) PaDa-I and B) JaWa. The protein model for PaDa-I (A) was built on the structure of the *Aae*UPO1 crystal (PDB access number: 2YOR) and the software PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC. The new mutations of the PaDa-I mutant with respect to the native UPO are shown underlined, while the residues with a zig-zag underline are those which have been changed in JaWa (B). The image shows the five Phe that participate in the accommodation of the substrate: Phe 69, Phe 76, Phe 121, Phe 191 and Phe 199; the two catalytic residues are R189 and E196.

Figure 8 B factors for the evolved UPOs of the present invention. Representation of the B factors (obtained using PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC.) of the PaDa-I variant (left) and the JaWa variant (right). Said B factors make reference to the rigidity/flexibility of a protein region or of an amino acid. **A)** Detail of the mutation in position 257, located on the surface: darker shades indicated greater rigidity. **B)** Representation in "putty" mode of the complete structure of the PaDa-I and JaWa variants. The greater the thickness of the lines, the greater the flexibility.

Figure 9 Assay of 4-AAP (4-aminoantipyrine) with different pure UPO variants (*Aae*UPO1, PaDa-I and JaWa). The reactions were performed at room temperature and their composition was as follows: 0.2 μM of each pure UPO variant, 50 mM pH 7.0 potassium phosphate buffer, 5 mM propranolol, 2 mM H_2O_2 (0.05 mL of final volume) and, in the case of reactions with ascorbic acid, it was added to a concentration of 4 mM. Each reaction was performed in triplicate.

Figure 10 Molecular docking with JaWa and propranolol. Amino acids that interact with propranolol are indicated, with the distances therefrom. The zone selected for MORPHING experiments due to its proximity to the protein-substrate contact points is indicated in dark grey.

Figure 11 Mutations in SoLo variants with respect to the JaWa variant described in the invention. Model built on the structure of the PaDa-I crystal. **A)** JaWa; **B)** SoLo.

Figure 12 Thermostability analysis (T_{50}) of the JaWa (black circles) and SoLo (white circles) mutants. The experiments were carried out using culture supernatants and each point represents the average value and standard deviation of three individual experiments.

Figure 13 Chromatogram showing the enzyme reactions. The reactions were performed at room temperature and their composition was as follows: 0.03 μM of each pure UPO variant, 50 mM pH 7.0 of potassium phosphate buffer, 4 mM propranolol, 2 mM H_2O_2 (0.5 mL of final volume).

Figure 14 Turnover rates of *Aae*UPO, JaWa and SoLo. The reaction mixture contained 0.03 μM of each pure UPO variant, 0.4 mM 5'-hydroxypropranolol, and 2 mM H_2O_2 in 50 mM pH 7.0 potassium phosphate buffer (0.3 mL of final volume). The disappearance of the product 5'-hydroxypropranolol can be observed due to the formation of its corresponding quinone by means of the peroxidase activity of the enzyme.

Figure 15 Calculation of the total turnover number (TTN) of AaeUPO and SoLo. The assay was carried out using 0.03 μM of each pure enzyme, 4 mM propranolol and 2 mM H_2O_2 in 50 mM pH 7.0 potassium phosphate buffer and in the same manner, but also with 4 mM ascorbic acid. In both cases, 2 mM H_2O_2 was added every 10 minutes, monitoring the reaction in each addition point taking different aliquots.

EXAMPLES

[0126] Following are examples of the invention by means of assays carried out by the inventors, which evidence the effectiveness of the product of the invention. The following examples serve to illustrate the invention and must not be considered to limit the scope thereof.

Example 1. Obtainment and characterisation of the variants of the present invention.

MATERIALS AND METHODS

Reagents and enzymes

[0127] ABTS (2,2'-azino-bis(3-ethylbenzothiazolin-6-sulfonic) acid), DMP (2,6-dimetoxiphenol), benzyl alcohol, 1-naphthol, 2-naphthol, 1,4-naphthoquinone, Fast Red (*Fast Red TR Salt hemi(zinc chloride) salt*), Taq DNA polymerase and the *Saccharomyces cerevisiae* transformation kit were obtained from Sigma-Aldrich (Saint Louis, MO, USO). NBD (5-nitro-1,3-benzodioxole) was acquired from TCI America (Portland, OR, USA), while the naphthalene is from Acros Organics (Geel, Belgium).

[0128] The cDNA of *upo1* (C1A-2 clone) of *A. aegerita* was provided by Dr. Martin Hofrichter (M. J. Pecyna, et al. Appl. Microbiol. Biotechnol. 2009, 84, 885-897).

[0129] The competent *Escherichia coli* XL2-Blue cells and the Genemorph II Random Mutagenesis (Mutazyme II) kit were obtained from Agilent Technologies (Santa Clara, CA, USA) and the iProof high-fidelity DNA polymerase was acquired from Bio-Rad (Hercules, CA, USA). The BamHI and XhoI restriction enzymes were obtained from New England Biolabs (Ipswich, MA, USA) and the protease-deficient strain of *S. cerevisiae* BJ5465 from LGCPromochem (Barcelona, Spain). The Zymoprep Yeast Plasmid Miniprep and Zymoclean Gel DNA Recovery kits are marketed by Zymo Research (Orange, CA, USA). The NucleoSpin Plasmid kit is from Macherey-Nagel (Düren, Germany) and the oligonucleotides used were synthesised by Isogen Life Science (Barcelona, Spain). All the chemical compounds are of the highest purity available in the market.

Directed evolution

[0130] The PaDa-I mutant (SEQ ID NO: 18) comprising the mutated signal peptide of SEQ ID NO: 28, was obtained as described in P. Molina-Espeja, et al. Appl. Environ. Microbiol. 2014. 80, 3496.-3507. After each evolution cycle, the PCR products were loaded in a semi-preparatory agarose gel and were purified using the Zymoclean Gel DNA Recovery kit. The DNA fragments recovered were cloned in the pJRoC30 plasmid under the control of the GAL1 promoter linearised with BamHI and XhoI (wherewith the parental or predecessor gene is also eliminated). The linearised plasmid was loaded in a low-melting-point preparatory agarose gel and was purified using the Zymoclean Gel DNA Recovery kit.

First generation (1G)

[0131] In order to obtain the variants described in the present invention, an error-prone PCR was performed in a final volume of 50 μL. This reaction contained 3% dimethyl sulfoxide (DMSO), 0.37 μM of RMLN (SEQ ID NO: 33 5'-cctctatactttaacgtcaagg-3'), 0.37 μM of RMLC (SEQ ID NO: 34 5'-gggagggcgtgaatgtaagc-3'), 0.8 mM deoxynucleotide triphosphate (dNTPs, 0.2 mM each), 0.05 U/μL of Mutazyme II (Genemorph II kit, Stratagene) and 2.822 ng of template (pJRoC30 plasmid (from the California Institute of Technology (CALTECH, USA), which comprises the nucleotide sequence of the PaDa-I mutant of SEQ ID NO:17, 300 ng of the target DNA). This mutagenic PCR was performed in a gradient thermocyclator (Mycycler, Bio-Rad, USA), determining the following parameters: 95°C 2 min (1 cycle); 94°C 45 s, 53°C 45 s and 74°C 3 min (28 cycles); and 74°C 10 min (1 cycle). 200 ng of the PCR product were mixed with 100 g of the linearised plasmid and competent *S. cerevisiae* cells were transformed so as to produce *in vivo* DNA shuffling and cloning (using the yeast transformation kit for such purpose). The volume resulting from the transformation was plated in (solid) minimal plates (for SC drop-out plates, said (solid) minimum consists of 100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L uracil-free amino acid supplement, 100 mL of 20% glucose, 20 g of bacto agar, 700 mL of distilled water and 1 mL of 25 g/L chloramphenicol) were incubated for three days at 30°C. The individual colonies that were formed were selected and subjected to a dual colorimetric High-Throughput Screening (HTS) assay, to efficiently explore mutant libraries without altering enzyme stability thereof, in addition to various re-screenings, as described below.

Second generation (2G)

[0132] Mutagenic StEP (Staggered Extension Process) was performed using the best mutants obtained in the first generation (H. Zhao, et al. Nat Biotechnol. 1998. 16, 258-261; E. Garcia-Ruiz, et al. Biochem. J. 2012. 441, 487-498) combined with *in vivo* shuffling. The conditions of the StEP PCR were: 3% DMSO, 90 nM RMLN (SEQ ID NO: 33 5'-cctctatactttaacgtcaagg-3'), 90 nM RMLC (SEQ ID NO: 34 5'-gggagggcgtgaatgtaagc-3'), 0.3 mM dNTPs (0.075 mM each), 0.05 U/μL Taq DNA polymerase and 16 ng of the templates (pJRoC30 with the four best mutants of the first generation). The PCRs were performed in a gradient thermocyclator using the following parameters: 95°C 5 min (1 cycle); 94°C 30 s, 55°C 20 s (90 cycles). 200 ng of the PCR products were mixed with 100 ng of the linearised plasmid

and transformed into competent *S. cerevisiae* cells). The rest of the procedure was followed as explained previously to obtain the first generation. In this evolution cycle a new variant, JaWa, was obtained, wherein the two new mutations took place: G241D and R257K, with respect to any of the enzymes *Aae*UPO1 or PaDa-I.

W24F variants

[0133] Two individual high-fidelity PCRs were performed for each PaDa-I variant (PaDa-I of SEQ ID NO: 18, encoded by SEQ ID NO: 17) and JaWa (SEQ ID NO: 24, encoded by SEQ ID NO: 23), using the nucleotide sequences that encode both as a template and thereby introducing the change required in their sequence. Starting the numbering of the *upo1* gene of SEQ ID NO: 1 from the start of the mature protein of SEQ ID NO: 2, the two nucleotide changes made were G71T and G72T (change in codon: TGG -W- to TTT -F). Two primers were designed for these PCRs, wherein the aforementioned changes were included. Said primers were the F24FOR primer of sequence SEQ ID NO: 35 (F24FOR: 5'-ctcacccatttaagccgcttcgacctggcgatattcgtggac-3') and the F24REV primer of sequence SEQ ID NO: 36 (5'-gtccacgaatatcgccaggtcgaagcggcttaaatgggtgag-3'). The changes made to said primer to perform the mutagenesis appear underlined in the nucleotide sequence thereof.

[0134] The conditions of these PCRs were: (i) in a final volume of 50 μL, 3% DMSO, 0.5 μM RMLN (SEQ ID NO: 33), 0.5 μM F24REV of SEQ ID NO: 36, 1 mM dNTPs (0.25 mM each), 0.02 U/μL of iProof high-fidelity DNA polymerase and 10 ng of the templates; or (ii) in a final volume of 50 μL, 3% DMSO, 0.5 μM F24FOR of SEQ ID NO: 35, 0.5 μM RMLC of SEQ ID NO: 34, 1 mM dNTPs (0.25 of each), 0.02 U/μL of iProof high-fidelity DNA polymerase and 10 ng of the templates. The following parameters were used: (i) 98°C 30 s (1 cycle), 98°C 10 s, 47°C 25 s, 72°C 15 s (28 cycles) and 72°C 10 min (1 cycle); or (ii) 98°C 30 s (1 cycle), 98°C 10 s, 58°C 25 s, 72°C 45 s (35 cycles) and 72°C 10 min (1 cycle). 200 ng of the two PCR products corresponding to their respective template were mixed with 100 g of the linearised plasmid and were transformed into S. *cerevisiae* in order to perform the *in vivo* assembly of the genes and cloning using the *In Vivo* Overlap Extension (IVOE) technique (M. Alcalde. Methods Mol. Biol. 2010. 634, 3, -14).

Preparation of the mutant libraries

[0135] Individual colonies corresponding to clones were selected and inoculated in 96 sterile wells (Greiner Bio-One GmbH, Germany), hereinafter mother plates, with 200 μL/minimal medium for expression per well (100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L, 67 mL of 1M pH 6.0 potassium phosphate buffer, 111 mL of 20% galactose, 22 mL of 0.1 M $MgSO_4$, 31.6 mL of absolute ethanol, 1 mL of 25 g/L chloramphenicol and *dd*H_2O up to 1,000 mL). Column 6 of each column was inoculated with the corresponding parental and well H1 with untransformed *S. cerevisiae*. The plates were sealed to avoid evaporation and were incubated at 30°C, 220 RPM and 80% of relative humidity (in a Minitron, INFORS, Switzerland) for five days.

Dual colorimetric High-Throughput Screening (HTS)

[0136] The mother plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 minutes at 3,500 RPM and 4°C. 20 μL of supernatant were transferred from these mother plates to two replica daughter plates with the help of a Freedom EVO liquid handling robot (Tecan, Switzerland). 180 μL of reaction mixture were added with 2,6-dimethoxy-phenol (DMP) or naphthalene to the daughter plates using a pipetting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, USA).

[0137] The DMP reaction mixture was composed of 100 mM pH 7.0 potassium phosphate buffer, 3 mM DMP and 1 mM H_2O_2. Simultaneously, this same screening assay was carried out but adding 10% acetonitrile to the reaction mixture in order to determine changes in the activity caused by the appearance of resistance to this organic co-solvent (present in the naphthalene screening reaction mixture, necessary so it remains dissolved). The reaction mixture with naphthalene contained 100 mM pH 7.0 potassium phosphate buffer, 0.5 mM naphthalene, 10% acetonitrile and 1 mM H_2O_2. The plates were briefly agitated and initial absorbance was measured at 469 nm and 510 nm, respectively, using a plate reader for such purpose (SPECTRAMax Plus 384, Molecular Devices, USA). After a reaction time of 10 minutes, 20 μL of Fast Red (Fast Red TR Salt hemi(zinc chloride) salt) were added to each naphthalene screening well (so that its final concentration in each well was 0.5 mM). The plates were kept at room temperature until they turned orange (DMP) or red (naphthol-Fast Red), at which time the absorbance was newly measured. The values were normalised against the parental of each plate. In order to rule out false positives, two re-screenings were carried out, in addition to a third re-screening wherein kinetic stability was determined (T_{50}) (P. Molina-Espeja, et al. Appl. Environ. Microbiol. 2014. 80, 3496-3507). The Fast Red compound was specifically coupled to the 1-naphthol to form an azo-type red dye that can be measured at 510 nm (ε_{510} = 4,700 M^{-1} cm^{-1}), wavelength at which the interference in the measurement produced by the culture medium is minimal.

First re-screening

[0138] The best screening clones were selected (~50 clones), of which 5 μL aliquots were taken and transferred to sterile plates containing of 200 μL minimal medium for expression per well. Columns 1 and 12 plus rows A and H were not inoculated, for the purpose of avoiding evaporation and, thus, the appearance of false positives. They were incubated for 5 days at 30°C and 220 RPM. The parental was treated in the same manner (row D, wells 7-11). The plates were treated following the same protocol as the previously described screening.

Second re-screening

[0139] An aliquot with the ~10 best clones of the first re-screening was inoculated in 3 mL of YPD culture medium (10 g of yeast extract, 20 g of peptone, 100 mL of 20% glucose, 1 mL of 25 g/L chloramphenicol and *dd*H_2O up to 1,000 mL) at 30°C and 220 RPM for 16 hours. The plasmids of those cultures were extracted using the Zymoprep Yeast Plasmid Miniprep kit. Due to the impurity and low concentration of the DNA extracted, the plasmids were transformed into supercompetent *E. coli* XL2-Blue cells and plated in LB-amp plates (Luria-Bertani medium is composed of 5 g of yeast extract, 10 g of peptone, 10 g of NaCl, 100 mg of ampicillin and *dd*H_2O up to 1,000 mL). An individual colony was selected from each clone, inoculated in 5 mL of LB and grown for 16 hours at 37°C and at 250 RPM. The plasmids were extracted using the NucleoSpin Plasmid kit and transformed into competent *S. cerevisiae* cells (as well as with the parental). Five individual colonies of each clone were selected and inoculated to undergo the same previously described screening protocol.

Third re-screening. Thermostability assay

[0140] An individual *S. cerevisiae* colony was selected with the corresponding clone (grown in a SC drop-out minimal medium plate: 100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L uracil-free amino acid supplement, 100 mL of 20% glucose, 1 mL of 25 g/L chloramphenicol and *dd*H_2O up to 1,000 mL) was inoculated in 2 mL of selective minimal medium (as in the SC plate medium, but with 20 g of bacto agar and rafinose instead of galactose) and was incubated for 48 hour at 30°C and 220 RPM. An aliquot of this culture was taken such that, upon inoculating it in 5 mL of new minimal medium, optical density at 600 nm would have a value of 0.25 (optical density, OD_{600}=0.25). This starter was incubated until completing two full growth cycles (between 6 and 8 hours), at which time 1 mL of cells were taken to inoculate 9 mL of expression medium in a 100 mL flask (OD_{600}=0.1). This culture of each clone was incubated for 72 hours at 25°C and 220 RPM (at peak UPO activity; OD_{600}=25-30), the cells were separated by centrifugation (10 minutes at 4,500 RPM and 4°C) and supernatant was filtered (using a glass and nitrocellulose filter with a pore size of 0.45 μm). Appropriate dilutions of the supernatants were prepared so that aliquots of 20 μL would give rise to a linear response in kinetic mode. 50 μL of supernatant were used for each point in a temperature gradient created by means of thermocyclator, from 30 to 80°C. After incubating for 10 minutes, the aliquots were cooled in ice for 10 minute and tempered at room temperature for 5 minutes. Lastly, these supernatants were subjected to the colorimetric assay using ABTS (100 mM pH 4.0 sodium phosphate/citrate buffer, 0.3 mM ABTS and 2 mM H_2O_2). The thermostability values were calculated in accordance with the ratio between the residual activities incubated at different temperatures and the value of initial activity at room temperature. The value of T_{50} was determined as the value of the temperature at which the protein loses 50% of it initial activity after incubating for 10 minutes.

Production of UPO recombinant variants in *S. cerevisiae*

[0141] An independent *S. cerevisiae* colony that comprised the corresponding variant of the invention was selected from a SC drop-out minimal medium plate and inoculated in 20 mL of liquid SC minimal medium, cultures which were incubated at 48 h at 30°C and 220 RPM. An aliquot of this culture was taken so that, upon inoculating it in 100 mL of new minimal medium, OD_{600} would have a value of 0.25. This starter was incubated until completing two full growth cycles (between 6 and 8 hours), at which time 100 mL of cells were taken to inoculate 900 mL of minimal medium for expression in a 2,000 mL flask (OD_{600}=0.1). This culture of each clone was incubated for 72 hours at 25°C at at 220 RPM (at peak UPO activity; OD_{600}=25-30), the cells were separated by centrifugation (10 minutes at 4,500 RPM and 4°C) and the supernatant was filtered (with glass and nitrocellulose filter with a pore size of 0.45 μm).

Purification of recombinant *Aae*UPO1 variants

[0142] The purification of the recombinant AaeUPO variants described in the present invention was carried out by means of ion-exchange chromatography (ÄKT A purifier, GE Healthcare). The raw extract was firstly treated by fractional precipitation with ammonium sulphate (55%, first cut) and, after eliminating the pellet, the supernatant was newly sub-

jected to precipitation with ammonium sulphate (85%, second cut). The final pellet was re-suspended in the 10 mM pH 4.3 sodium phosphate/citrate buffer (buffer A) and the sample was filtered and loaded on a strong cation-exchange column (HiTrap SP FF, GE Healthcare), pre-balanced with buffer A. The proteins were eluded by means of a linear gradient of 0 to 25% of buffer A with 1 M of NaCl in 55 mL and of 25 to 100% of buffer A with 1 M NaCl in 5 mL, at a flow rate of 1 mL/min. The fractions with UPO activity were recovered, concentrated and dialysed in 10 mM pH 6.5 Bis Tris buffer (buffer B) and loaded on a high-resolution anion-exchange column (Biosuite Q, Waters), pre-balanced with buffer B. The proteins were eluded by means of a linear gradient of 0 to 15% of buffer B with 1 M of NaCl in 40 mL y de 15 a 100% de buffer B with 1 M NaCl in 5 mL, at a flow rate of 1 mL/min. The fractions with UPO activity were recovered, concentrated and dialysed in 50 mM pH 7.0 potassium phosphate buffer and stored at 4°C. Reinheitszahl [Rz] [A_{418}/A_{280}] values of ~2 were obtained. The fractions of the different purification steps were analysed in a 12% SDS/PAGE acrylamide gel, dyed with Coomassie blue. The concentrations of the raw extracts of these steps were determined by means of Bradford reagent and BSA as standard.

Kinetic constants values

[0143] The kinetic constants of the variants of the invention for ABTS were estimated in 100 mM pH 4.0 sodium phosphate/citrate buffer and 2 mM H_2O_2; and for the rest of the substrates, in 100 mM pH 7.0 potassium phosphate buffer, 2 mM H_2O_2 (DMP) or 1 mM H_2O_2 (NBD and naphthalene, in 20% of acetonitrile - final concentration). For H_2O_2, benzyl alcohol was used as substrate at the corresponding saturation conditions. The reactions were performed in triplicate and the oxidations of the substrates were followed by spectrophotometric changes (ABTS: ε_{418}=36,000 M^{-1} cm^{-1}; DMP: ε_{469}=27,500 M^{-1} cm^{-1}; NBD: ε_{425}=9,700 M^{-1} cm^{-1}, naphthalene: ε_{303}=2,010 M^{-1} cm^{-1}, and benzyl alcohol: ε_{280}=1,400 M^{-1} cm^{-1}). The kinetics for naphthalene were performed following the protocol described in M. G. Kluge, et al. Appl. Microbiol. Biotechnol. 2007. 75, 1473-1478. In order to calculate the values of K_m and k_{cat}, values of V_{max} were represented at substrate concentrations and the hyperbole function was adjusted (using SigmaPlot 10.0, wherein the parameter a is equal to k_{cat} and the parameter b, to K_m).

HPLC analysis

[0144] The reactions were analysed by means of chromatography in reverse phase (HPLC). The equipment is composed by a tertiary pump (Varian-Agilent Technologies, USA) coupled to an autosampler (Merck Millipore, MA, USA); an ACE C18 PFP column was used for separation (pentafluorophenyl, 15 cm x 4.6 cm) at 45°C and detection was performed using a photodiode detector (PDA) (Varian-Agilent Technologies, USA). The mobile phase selected was 70% methanol and 30% *dd*H_2O (in both cases with 0.1% of acetic acid) at a flow rate of 0.8 mL/min. The reaction was quantified at 268 nm (based on standard HPLCs). For the 15 minute reaction, the mixture contained 6.6 nM of pure enzyme, 1 mM naphthalene, 20% acetonitrile and 1 mM H_2O_2 in 100 mM pH 7.0 potassium phosphate buffer (1 mL of final volume). The reaction started with the addition of H_2O_2 and stopped with 20 μL of 37% HCl. For long reaction times, the conditions used were those described earlier but without stopping the reaction with HCl. A sample of 10 μL was injected and analysed at different reaction times (from 1 to 270 minutes).

[0145] For the kinetic values of the 1-naphthol, the reaction was performed using 40 nM of pure enzyme, 1 mM 1-naphthol, 20% acetonitrile and 1 mM H_2O_2 in 100 mM pH 7.0 potassium phosphate buffer (0.2 mL of final volume).

[0146] The standard deviations were less than 5% in all cases.

Analysis using MALDI-TOF-MS and determination of the isoelectric point

[0147] The analyses were performed using an Autoflex III MALDI-TOF-TOF unit with smartbeam laser (Bruker Daltonics). The samples were evaluated in positive mode. The method was calibrated using BSA with standard, thereby covering a range of 15,000 to 70,000 Da. In order to determine the isoelectric point of the UPO variants, 8 μg of pure enzyme were subjected to two-dimensional electrophoresis. These experiments were carried out at the Proteomic and Genomic Service of the Biological Research Centre (CIB-CSIC, Spain).

Analysis by Liquid Chromatography/Mass Spectrometry (LC/MS)

[0148] These analyses were performed using a mass spectrometer with a Q-TOF hybrid analyser (QSTAR, ABSciex, MA, USA). Electrospray (ESI) was used as an ionisation source and, as ionising phase, methanol. In this case, the entrance system was direct injection in a HPLC 1100 (Agilent Technologies, USA). The resolution of the assay corresponds to 9,000 FWHM (*Full Width at Half Maximum*), accuracy, 5-10 ppm and was performed in negative mode.

RESULTS

[0149] Taking the PaDa-I mutant enzyme of SEQ ID NO: 18 encoded by SEQ ID NO: 17 as parental to carry out the directed evolution experiments, UPO mutant libraries were built by means of random mutagenesis and recombination by StEP and *in vivo* DNA shuffling with the objective of obtaining a mutant enzyme or variant that shows less peroxidase activity on the 1-naphthol, while boosting peroxygenase activity on the naphthalene, also taking into account that said variant must be expressed robustly in heterologous organisms and secreted in an active, soluble and very stable form. To this end, each variant obtained in the mutant libraries was subjected to *ad hoc* double screening for the purpose of obtaining the variants with the aforementioned capabilities, greater peroxygenase activity against naphthalene and less peroxidase activity against 1-naphthol.

[0150] After subjecting the PaDa-I mutant (SEQ ID NO: 17) to two cycles of directed evolution (~ 4,000 clones analysed), a double mutant was identified which was called JaWa and which comprises the nucleotide sequence SEQ ID NO: 23, that encodes the variant of SEQ ID NO: 24. Said JaWa mutant (SEQ ID NO: 24) comprises the G241D and R257K mutations with respect to the PaDa-I mutant of SEQ ID NO: 18, with a peroxygenase activity on microplate that doubled that of its parental and a peroxidase activity that was reduced to half (**Figure 1**).

[0151] Both variants, PaDa-I and JaWa, were produced, purified at homogeneity (Reinheitszahl [Rz] [A_{418}/A_{280}] value ~2) and biochemically characterised. No changes were detected with regard to general spectral characteristics, processing of the N-terminus, molecular mass or degree of glycosylation **(Table 1).**

Table 1. Biochemical characteristics of wild-type AaeUPO (SEQ ID NO: 4) and of the PaDa-I (SEQ ID NO: 18) y JaWa (SEQ ID NO: 24) variants.

Spectroscopic and biochemical characteristics	Wild-type UPO	PaDa-I	JaWa
Pm (Da)[1]	46,000	52,000	52,000
Pm (Da)[2]	n.d.	51,100	51,100
Pm (Da)[3]	35,942	35,914	35,944
Degree of glycosylation (%)	22	30	30
Thermal stability, T_{50} (°C)[4]	n.d.	57.6	59.7
pI	4.9-5.7	5.5	5.3
Optimum pH for ABTS	4.0	4.0	4.0
Optimum pH for DMP	7.0	6.0	6.0
Optimum pH for naphthalene	6.5	6.0	6.0
Rz, (A_{418}/A_{280})	2.4	1.8	2.3
Soret region (nm)	420	418	418
CT1 (nm)	572	570	570
Spectroscopic and biochemical characteristics	Wild-type UPO	PaDa-I	JaWa
CT2 (nm)	540	537	537

[1]Estimated by SDS-PAGE; [2]estimated using MALDI-TOF; [3]estimated according to the amino acid composition. [4]Estimated in culture supernatants. n.d. not determined.

[0152] As can be observed in **Table 1** and in **Figure 2,** the JaWa mutant enzyme of SEQ ID NO: 24 showed greater kinetic thermostability than the PaDa-I variant of SEQ ID NO: 18 (2°C higher T_{50} -temperature at which the enzyme retains 50% of its activity after 10 minutes of incubation-), in addition to higher stability in the presence of acetonitrile, necessary for the bioavailability of the naphthalene (the solubility of the naphthalene in water is 31.7 mg/L) (**Figure 2**).

[0153] The naphthalene transformation reaction performed by the JaWa (SEQ ID NO: 24) and PaDa-I (SEQ ID NO: 18) mutants and that was analysed by means of HPLC-PDA has evidenced that the oxygenation of the naphthalene by *Aae*UPO occurs through an unstable intermediary compound, 1,2-naphthalene oxide (epoxide). It undergoes quick hydrolysis to naphthol (1- and 2-naphthol) when the pH is acid (M. Kluge, et al. Appl. Microbiol. Biotechnol. 2009. 81, 1071-1076). Therefore, the distribution of the resulting products after 15 minutes of reaction was firstly measured (stopped with HCl). Both the PaDa-I (SEQ ID NO: 18) and JaWa (SEQ ID NO: 24) variants demonstrated similar regioselectivity (92% 1-naphthol, 8% 2-naphthol), but the JaWa variant showed a significant increase in the production of 1-naphthol (156% more than PaDa-I) without detectable traces of 1,4-naphthoquinone, its oxidation product (**Figure 3A**).

[0154] When the long reaction times were monitored (270 minutes at pH 7.0 without stopping the reaction), a similar

behaviour was observed, which indicates that the transformation of the 1,2-naphthalene oxide to naphtholes also occurs at neutral pH, although it is true that, at lower speed, traces of 1,4-naphthoquinone were also detected (**Figure 3B, C, D**).

[0155] While with both variants, PaDa-I and JaWa, the formation of the epoxide intermediary reached its maximum at ~40 minutes (due to the oxidative damage caused by the H_2O_2 in all the peroxidases), regioselectivity increased to 97% of 1-naphthol. This result corresponds to the loss of selectivity observed in acid conditions given by a greater reactivity of the epoxide.

[0156] The composition of the resulting products did not vary for any of the PaDa-I (SEQ ID NO: 18) and JaWa (SEQ ID NO: 24) variants, as observed in the mass spectrometry analysis performed, but the differences between the two mutants in terms of production performance were very significant, reaching values of 0.14 and 0.32 mM of 1-naphthol for PaDa-I and JaWa, respectively. The JaWa variant obtained total turnover numbers (TTN) of nearly 50,000 against the 20,000 of PaDa-I.

[0157] Additionally, the kinetic values of the two variants were determined using substrates of both peroxygenase and peroxidase activity (**Table 2**), as described in the section on materials and methods. Briefly, the kinetic constants for the ABTS were measured in 100 mM pH 4.0 sodium phosphate/citrate buffer and 2 mM H_2O_2, while 100 mM pH 7.0 potassium phosphate and 2 mM H_2O_2 (DMP) or 1 mM (naphthalene or NBD, in 20% acetonitrile -final concentration) was used for the other buffers. For the H_2O_2, benzyl alcohol was used as substrate to the corresponding saturation conditions.

Table 2. Kinetic parameters for PaDa-I (SEQ ID NO: 18) and JaWa (SEQ ID NO: 24) variants.

Substrate	Kinetic constants	PaDa-I	JaWa
ABTS	K_m (μM)	48.0 ± 4.5	181 ± 22
	k_{cat} (s^{-1})	395 ± 13	125 ± 5
	k_{cat}/K_m ($s^{-1}M^{-1}$)	$8.2x10^6$ ± $6x10^5$	$6.9x10^5$ ± $6.3x10^4$
DMP	K_m (μM)	126 ± 14	866 ± 108
	k_{cat} (s^{-1})	68 ± 2	142 ± 8
	k_{cat}/K_m ($s^{-1}M^{-1}$)	$5.4x10^5$ ± $4.8x10^4$	$1.6x10^5$ ± $1.2x10^4$
Naphthalene	K_m (μM)	578 ± 106	127 ± 27
	k_{cat} (s^{-1})	229 ± 17	78 ± 3
	k_{cat}/K_m ($s^{-1}M^{-1}$)	$4x10^5$ ± $4x10^4$	$6.2x10^5$ ± $1.1x10^5$
NBD	K_m (μM)	483 ± 95	769 ± 80
	k_{cat} (s^{-1})	338 ± 22	154 ± 8
	k_{cat}/K_m ($s^{-1}M^{-1}$)	$7x10^5$ ± $9.9x10^4$	$2.0x0^5$ ± $1.2x10^4$
H_2O_2	K_m (μM)	486 ± 55	1,250 ± 300
	k_{cat} (s^{-1})	238 ± 8	447 ± 40
	k_{cat}/K_m ($s^{-1}M^{-1}$)	$5.0x10^5$ ± $4.2x10^4$	$3.6x10^5$ ± $5.9x10^4$

[0158] As can be observed in **Table 2,** the k_{cat}/K_m value (catalytic efficiency) for naphthalene was 1.5 times higher for the JaWa variant (SEQ ID NO: 24) with respect to the PaDa-I variant (SEQ ID NO: 18). Also, the peroxidase activity of the JaWa variant (SEQ ID NO: 24) was reduced (with a significant decrease in catalytic efficiencies of 3 to 11 times for the substrates of peroxidase activity DMP and ABTS, respectively). The k_{cat}/K_m value for H_2O_2 with benzyl alcohol as substrate was also affected. In the results obtained with NBD, another oxygen transfer substrate such as naphthalene, the trend is similar, *i.e.* k_{cat} decreases in the JaWa variant while the affinity to the K_m substrate improves, despite the fact that this entails higher k_{cat}/K_m for the PaDa-I variant. The fact that the catalytic efficiency of the JaWa variant for NBD has not improved is significant, since it is not a substrate used in the screenings of this part of the evolution. However, the fact that the tendency of the catalytic constant and affinity to the substrate is similar in two monooxygenase substrates indicates that there is an enzyme action mechanism acting in some way to favour the formation of 1-naphtol while reducing peroxidase activity.

[0159] To confirm the decrease in peroxidase activity with respect to the hydroxylation of the naphthalene, the values of the catalytic constant were measured by using HPLC (μmol product μmol enzyme^{-1} min^{-1}) for the conversion of 1-naphthol into 1,4-naphthoquinone. Although the catalytic constant of the PaDa-I variant (SEQ ID NO: 18) for 1-naphthol was already low (200 min^{-1}), with the JaWa variant (SEQ ID NO: 24) this value decreased to 92 min^{-1}, in addition to a reduction of ~1.5 times in the ratio 1,4-naphthoquinone:1-naphthol (**Figure 4**). This effect can also be observed at first glance, since the polymeric products produced in the reaction with the PaDa-I variants (SEQ ID NO: 18) (due to non-enzymatic quinone regrouping processes) are coloured (**Figure 4**). There are hypotheses in literature on the possibility

that UPO is similar to CPO in the existence of different sites with peroxidase activity in its structure. To suppress these alternative peroxidation pathways, the structure of the *A*aeUPO1 crystal was closely examined and a variant was built by mean of directed mutagenesis in Trp24 (**Figure 5A**), a highly oxidable residue, found on the protein surface, using the PaDa-I (SEQ ID NO: 18) and JaWa (SEQ ID NO: 24) variants as templates, as described in the section on materials and methods.

[0160] Next, the activities of the PaDa-I-W24F (SEQ ID NO: 30) and JaWa-W24F (SEQ ID NO: 32) variants were determined. The W24F mutation reduced 60% of the peroxidase activity in both variants and with all the tested substrates, but caused a decrease in the peroxygenase activity, with a reduction of 50% in the activity on the naphthalene and NBD (**Figure 5B**). This indicates that the Trp24 residue probably also affects the peroxygenase activity of the UPO.

Example 2. Mutational analysis of the variants of the invention

[0161] The mutations of the JaWa variant were mapped (SEQ ID NO: 24) onto the structure of the wild *Aae*UPO1 (SEQ ID NO: 4), which shows a very characteristic catalytic pocket wherein linkage with the substrate takes place, dominated by a Phe triad (Phe69-Phe121-Phe199) involved in the correct orientation of the aromatic compounds (**Figure 6 and Figure 7**). The G241D mutation is at the entrance to the heme channel. The dramatic change of a Gly, apolar and small, for an Asp, loaded and larger, seems to narrow the cavity, which can affect the accommodation of the naphthalene in the catalytic pocket. This theory is not consistent with the fact that the affinity to naphthalene was improved in the JaWa variant, with a decrease in its K_m of 3 times (**Table 2**). On the contrary, the introduction of a negative charge in the heme-thiolate domain (in which there is a Glu196-Arg189 acid-base pair involved in the formation of the Compound I -porphyrin with a radical cation and oxo-Fe IV=O-) may negatively affect the k_{cat} value, depending on the chemical nature of the bound substrate. The R257K mutation is located on the surface of the protein, far from catalysis-relevant regions, but is at the start of a "pathway" towards the catalytic R189 residue. It is a known fact that some peroxidases show various surface-exposed entrances for electron-mediated substrate oxidation through a long-range electron transfer pathway towards the heme domain, as also described in the present work for W24F variants. In this regard, the R257K replacement may be affecting any of these circuits with a possible beneficial lateral effect on thermostability through localised remodelling in the secondary structure (the two mutations, G241D and R257K, vary the estimation of factor B (**Figure 8**)). B factor makes reference to the rigidity/flexibility of a protein or amino acid region present in a protein or peptide.

[0162] These results evidence that the UPO variants described herein show greater selectivity and the highest TTN known for the production of 1-naphthol for this enzyme superfamily to date. Additionally, as demonstrated, said variants are heterologously secreted in an active, soluble and very stable form, being capable of carrying out selective aromatic oxygenations in the absence of NAD(P)H cofactors and reductase domains. Their self-sufficient mono(per)oxygenase activity make this UPO variant a valuable biocatalyst for application in the field of organic synthesis.

Example 3. Obtainment and characterisation of variants of the invention for the synthesis of human drug metabolites (HDMs).

[0163] The most important HDMs include, namely, derivatives of propranolol, a beta-blocker drug commonly used for the treatment of hypertension, migraine prophylaxis in children and attenuation of physical manifestations of anxiety. This example shows how the UPO variants of the invention are capable of forming 5'-hydroxypropranolol from propranolol oxygenation, without inorganic pollutants, at room temperature, atmospheric pressure and in the absence of organic solvents, in a single step, with catalytic concentrations of H_2O_2 and without requiring the addition of antioxidants such as ascorbic acid to the reaction.

[0164] In addition to the variants described in Example 1, a new variant was built based on the JaWa variant, which even showed an improvement in the production of 5'-hydroxypropranolol with respect to said JaWa mutant. Following is a description of the obtainment of a new variant called SoLo comprising SEQ ID NO: 42 and which is encoded by the nucleotide sequence SEQ ID NO: 41.

MATERIALS AND METHODS

Reagents and enzymes

[0165] ABTS (2,2'-azino-bis(3-ethylbenzothiazolin-6-sulfonic acid)), L-ascorbic acid, 4-aminoantipyrine, benzyl alcohol, Taq DNA polymerase and the *Saccharomyces cerevisiae* transformation kit were obtained from Sigma-Aldrich (Saint Louis, MO, USA). NBD (5-nitro-1,3-benzodioxole) was acquired from TCI America (Portland, OR, USA), while the naphthalene, propranolol and potassium persulfate are from Acros Organics (Geel, Belgium). 5-hydroxypropranolol was acquired from Santa Cruz Biotechnology (Santa Cruz, CA, USA).

[0166] The competent *Escherichia coli* XL2-Blue cells and Pfu ultra DNA polymerase were obtained from Agilent Technologies (Santa Clara, CA, USA) and iProof high-fidelity DNA polymerase was acquired from Bio-Rad (Hercules, CA, USA). The BamHI and XhoI restriction enzymes were obtained from New England Biolabs (Ipswich, MA, USA) and the protease-deficient strain of *S. cerevisiae* BJ5465 from LGCPromochem (Barcelona, Spain). The Zymoprep Yeast Plasmid Miniprep and Zymoclean Gel DNA Recovery kits are marketed by Zymo Research (Orange, CA, USA). The NucleoSpin Plasmid kit is from Macherey-Nagel (Düren, Germany) and the oligonucleotides used were synthesised by Metabion (Bayern, Germany). All the chemical compounds are of the highest purity available in the market.

Directed evolution

[0167] Based on the JaWa mutant comprising SEQ ID NO: 24, which is encoded by the nucleotide sequence SEQ ID NO: 23, after each evolution cycle, the PCR products were loaded on a semi-preparatory agarose gel and purified using the Zymoclean Gel DNA Recovery kit. The recovered DNA fragments were cloned in the pJRoC30 plasmid under the control of the GAL1 promoter linearised with BamHI and XhoI (also eliminating the parental gel or predecessor). The linearised plasmid was loaded in a low-melting-point preparatory agarose gel and was purified using the Zymoclean Gel DNA Recovery kit.

First generation (1G)

[0168] To obtain the SoLo mutant (SEQ ID NO: 42, encoded by SEQ ID NO: 41), docking studies were performed on the JaWa mutant (SEQ ID NO: 24, encoded by SEQ ID NO: 23) using the *Molecular Operating Environment* program (MOE, *Chemical Computing Group Inc.* http://www.chemcomp/com) and propranolol as a substrate. Based on these, a region of the protein was selected to be subjected to random mutagenesis using the MORPHING technique (*Mutagenic Organized Recombination Process by Homologous in vivo Grouping*) (D. González-Perez et al., PLoS ONE 2014. 9:e90919). To obtain the different variants additional to those described earlier, two error-prone PCRs were performed in a specific zone of the nucleotide sequence (SEQ ID NO: 23) that encodes that JaWa mutant (SEQ ID NO: 24), specifically in the coding zone from the D187-V248 region of the JaWa mutant of SEQ ID NO: 24 in a final volume of 50 μL. These reactions contained 3% of dimethyl sulfoxide (DMSO), 90 nM MJaWa Fw (SEQ ID NO: 43; 5'-gcgcattcaa-gactccattg-3'), 90 nM MJaWa Rev (SEQ ID NO: 44; 5'-gatcttgccgacatttttttcc-3'), 0.3 mM deoxynucleotide triphosphates (dNTPs, 0.075 mM of each), 0.1 mM or 0.2 mM $MnCl_2$, 1.5 mM $MgCl_2$, 0.05 U/μL Taq DNA polymerase and 1 ng/μl of the template (pJRoC30 plasmid from the California Institute of Technology (CALTECH, USA), comprising the nucleotide sequence of the JaWa mutant of SEQ ID NO: 23). This mutagenic PCR was performed in a gradient thermocyclator (Mycycler, Bio-Rad, EEUU), determining the following parameters: 94°C 2 min (1 cycle); 94°C 45 s, 48°C 30 s and 72°C 90 s (28 cycles); and 72°C 10 min (1 cycle). Furthermore high-fidelity PCRs were performed in the fragments that must remain non-mutagenic in a final volume of 50 μL. These reactions contained 3% of dimethyl sulfoxide (DMSO), 0.5 μM HFJaWa Fw (SEQ ID NO: 45; 5'- caggctcatcctatgcagccc-3') and 0.5 μM RMLC (SEQ ID NO: 34; 5'-gggagggcgtgaatg-taagc-3') or 0.5 μM HFJaWa Rev (SEQ ID NO: 46; 5'-caaaggagaaattggggttggtcg-3') and 0.5 μM RMLN (SEQ ID NO: 33; 5'-cctctatactttaacgtcaagg-3') for the other high-fidelity fragment, 1 mM dNTPs (0.25 mM of each), 0.05 U/μL PfuUltra DNA polymerase and 2 ng/μL of template. These reactions were performed in the same gradient thermocyclator, determining the following parameters: 95°C 2 min (1 cycle); 95°C 45 s, 48°C 30 s and 72°C 90 s (28 cycles); and 72°C 10 min (1 cycle). 200 ng of PCR products were mixed with 100 ng of the linearised plasmid and competent *S. cerevisiae* cells were transformed such as to produce *in vivo* shuffling of the DNA and cloning (using the yeast transformation kit for such purpose). The volume resulting from the transformation was plated in minimal solid medium plates (for SC drop-out plates, said minimal solid medium consists of 100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L uracil-free amino acid supplement, 100 mL of 20% glucose, 20 g bacto agar, 700 mL of distilled water and 1 mL of 25 g/L chloramphenicol) and were incubated for 3 days at 30°C. The individual colonies that were formed were selected and subjected to the dual colorimetric High-Throughput Screening (HTS) assay to efficiently explore mutant libraries without altering the enzyme stability thereof, in addition to various re-screenings, as described below. In this evolution cycle, a new variant was obtained called SoLo, which comprises the nucleotide sequence SEQ ID NO: 41, that encodes the variant of SEQ ID NO: 42, wherein a new mutation took place: F191S, with respect to the JaWa variant (SEQ ID NO: 24).

Second Generation (2G)

[0169] Since the mutation that appeared in the SoLo variant (SEQ ID NO: 42) is found in one of the two phenylalanines that delimit the entrance to the heme channel, combinatorial saturation mutagenesis (CSM) was performed using the 22c-trick method, as described in S. Kille, et al. ACS Synth. Biol. 2013. 2,83-92, in positions S191 and F76.

[0170] To this end, three PCRs were performed in a final volume of 50 μL. All contained 3% of DMSO, 0.3 mM dNTPs (0.075 mM each), 0.05 U/μL PfuUltra DNA polymerase and 2 ng/μL of template, but each with different primers. PCR

1 with 0.25 μM of RMLN (SEQ ID NO: 33), 0.25 μM of F76 VHG R (SEQ ID NO: 47; 5'-gcaagtccgtaatgagattgccgtccacaagtgggccgcatatgtggc*ccdb*gattgcggc-3'), 0.25 μM of F76 NDT R (SEQ ID NO: 48; 5'-gcaagtccgtaatgagattgccgtccacaaggtgggccgcatatgtggc*ahn*gattgcggc-3') and 0.25 μM of F76 TGG R (SEQ ID NO: 49; 5'-gcaagtccgtaatgagattgccgtccacaaggtgggccgcatatgtggc*ccca*gattgcggc-3'). PCR 2 con 0.25 μM of HF F (SEQ ID NO: 50; 5'-gcggcccaccttgtggacggcaatctcattacggacttgc-3'), 0.25 μM of S191 VHG R (SEQ ID NO: 51; 5'-cccatccacaaaaagattcgcggggaaggtggtctcgccgtaagcagt*ccdb*gaacctaaag-3'), 0.25 μM of S191 NDT R (SEQ ID NO: 52; 5'-cccatccacaaaaagattcgcggggaaggtggtctcgccgtaagcagt*ahn*gaacctaaag-3') y 0.25 μM of S191 TGG R (SEQ ID NO: 53; 5'-cccatccacaaaaagattcgcggggaaggtggtctcgccgtaagcagt*ccca*gaacctaaag-3'). PCR 3 con 0,25 μM de HF F-RMLC (SEQ ID NO: 54; 5'-cggcgagaccaccttccccgcgaatctttttgtggatggg-3') and 0.25 μM of RMLC (SEQ ID NO: 34). The underlined regions are those in which *in vivo* DNA assembly occurs and the region in italics is the changed codon (where N = A/T/C/G; D = no C; V = no T, H = no G; and B = no A). These reactions were performed in the gradient thermocyclator, determining the following parameters: 95°C 2 min (1 cycle); 95°C 45 s, 48°C 45 s and 72°C 60 s (28 cycles); and 72°C 10 min (1 cycle). 200 ng of each of the PCR products were mixed with 100 ng of the linearised plasmid and transformed into competent *S. cerevisiae* cells. The rest of the procedure was followed as explained previously to obtain the first generation. No improved variant was obtained with respect to the SoLo mutant.

Third generation (3G)

[0171] There is a phenylalanine triad in the catalytic pocket of *Aae*UPO, PaDa-I and JaWa (F69-F121-F199). Due to the complex catalytic pocket and to the fact that these phenylalanines are in charge correctly orienting the aromatic substrates, it was decided to carry out mutagenesis on these residues with NNK degenerated codons (N = A/T/C/G; D; K = T/G, M = A/C) independently, *i.e.* creating three different libraries.

[0172] Library F69: two PCRs were performed in a final volume of 50 μL. The first contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLN (SEQ ID NO: 33), 0.5 μM F69 R (SEQ ID NO: 55; 5'-gaagattgcggcttgattgtc-*mnn*attgaatc-3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41). And the second contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLC (SEQ ID NO: 34), 0.5 μM F69 F (SEQ ID NO: 56; 5'-cgcggttcaggaaggattcaat*nnk*gacaatc-3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41).

[0173] F121 library: two PCRs were performed in a final volume of 50 μL. The first contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLN (SEQ ID NO: 33), 0.5 μM F121 R (SEQ ID NO: 57; 5'- catactggcgtcgccttc-*mnn*ggtgccatgc -3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41). And the second contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLC (SEQ ID NO: 34), 0.5 μM F121 F (SEQ ID NO: 58; 5'-ggactcaatgagcatggcacc*nnk*gaaggcg-3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41).

[0174] F199 library: two PCRs were performed in a final volume of 50 μL. The first contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLN (SEQ ID NO: 33), 0.5 μM F199 R (SEQ ID NO: 59; 5'-ccacaaaaagattcgcgg*mnn*ggtggtctcg-3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41). And the second contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μM RMLC (SEQ ID NO: 34), 0.5 μM F199 F (SEQ ID NO: 60; 5'-ctactgcttacggcgagaccacc*nnk*cccgcg-3'), 0.02 U/μL *iProof* DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41).

[0175] These reactions were performed in the gradient thermocyclator, determining the following parameters: 98°C 30 s (1 cycle); 98°C 10 s, 48°C 30 s and 72°C 30 s (28 cycles); and 72°C 10 min (1 cycle). 200 ng of each of the PCR products were mixed with 100 ng of the linearised plasmid (each library separately) and transformed into competent *S. cerevisiae* cells. The rest of the method was followed as explained earlier to obtain the first and second generation. Neither was any variant better than SoLo found (SEQ ID NO: 42), due to which this mutant was selected, together with the JaWa mutant (SEQ ID NO: 24) and the parental *Aae*UPO1, to analyse the synthesis of HDMs, taking 5'-hydroxypropranolol with each by way of example.

Preparation of the mutant libraries

[0176] Individual colonies corresponding to clones were selected and inoculated in sterile 96-well plates (Greiner Bio-One GmbH, Germany), hereinafter mother plates, with 200 μL/minimal medium for expression per well (100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L uracil-free amino acid supplement, 67 mL of 1 M pH 6.0 potassium phosphate buffer, 111 mL of 20% galactose, 22 mL of 0.1 M $MgSO_4$, 31.6 mL of absolute ethanol, 1 mL of 25 g/L chloramphenicol and *dd*H_2O up to 1,000 mL). Column 6 of each column was inoculated with the corresponding parental and well H1 with *S. cerevisiae* transformed with the pJRoC30-MtL plasmid (laccase without functional expression). The plates were sealed to avoid evaporation and were incubated at 30°C, 220 RPM and 80% of relative humidity (in a Minitron, INFORS, Switzerland) for five days.

Dual colorimetric High-Throughput Screening (HTS)

[0177] The mother plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 minutes at 3,500 RPM and 4°C. 20 μL of supernatant of these mother plates were transferred to two replica daughter plates with the help of a Freedom EVO liquid-handling robot (Tecan, Switzerland). 50 μL of reaction mixture with propranolol were added to the daughter plates using a pipetting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, USA).

[0178] The reaction mixture with propranolol was composed of 50 mM pH 7.0 potassium phosphate buffer, 5 mM propranolol and 2 mM H_2O_2 to detect the peroxygenase activity of the enzyme on the substrate and its subsequent peroxidase activity on the product. This same screening assay was simultaneously carried out but adding ascorbic acid (4 mM) to the reaction mixture in order to exclusively detect the peroxygenase activity of the enzyme on propranolol and avoid the subsequent peroxidase activity. Without ascorbic, the plates were incubated for 30 minutes and with ascorbic for 60 minutes. Subsequently, by means of the 4 aminoantipyrine (4-AAP, C. R. Otey and J. M. Joern, Methods Mol. Biol. 2003. 230, 141-8) the amount of product formed per well was revealed. The plates were briefly agitated and absorbance measured at 530 nm, using a plate reader for such purpose (SPECTRAMax Plus 384, Molecular Devices, USA). The values were normalised against the parental of each plate. To rule out false positives, re-screenings were carried out, in addition to a third re-screening wherein kinetic stability was determined (T_{50}) (P. Molina-Espeja, et al. Appl. Environ. Microbiol. 2014. 80, 3496-3507).

Second re-screening

[0179] An aliquot with the ~10 best screening clones was inoculated in 3 mL of YPD culture medium (10 g of yeast extract, 20 g of peptone, 100 mL of 20% glucose, 1 mL of 25 g/L chloramphenicol and *dd*H2O up to 1,000 mL) at 30°C and 220 RPM for 24 hours. The plasmids of those cultures were extracted using the Zymoprep Yeast Plasmid Miniprep kit. Due to the impurity and low concentration of the DNA extracted, the plasmids were transformed into supercompetent *E. coli* XL2-Blue cells and plated in LB-amp plates (Luria-Bertani medium is composed of 5 g of yeast extract, 10 g of peptone, 10 g of NaCl, 100 mg of ampicillin and *dd*H2O up to 1,000 mL). An individual colony was selected from each clone, inoculated in 5 mL of LB and grown for 16 hours at 37°C and at 250 RPM. The plasmids were extracted using the NucleoSpin Plasmid kit and transformed into competent *S. cerevisiae* cells (as in the parental, which in the first generation is JaWa and in the second and third is SoLo). Five individual colonies of each clone were selected and inoculated to undergo the same previously described screening protocol.

Third re-screening. Thermostability assay

[0180] An individual *S. cerevisiae* colony was selected with the corresponding clone (grown on a SC drop-out minimal medium plate: 100 mL of 6.7% yeast nitrogen base, 100 mL of 19.2 g/L uracil-free amino acid supplement, 100 mL of 20% glucose, 1 mL of 25 g/L chloramphenicol and *dd*H_2O up to 1,000 mL), was inoculated in 3 mL of selective minimal medium (like the SC plate medium, but with 20 g of bacto agar and rafinose instead of galactose) and incubated for 48 hours at 30°C and 220 RPM. An aliquot of this culture was taken such that, upon inoculating it in 5 mL of new minimal medium, optical density at 600 nm would have a value of 0.25 (optical density, OD_{600}=0.25). This starter was incubated until completing two full growth cycles (between 6 and 8 hours), at which time 1 mL of cells were taken to inoculate 9 mL of expression medium in a 100 mL flask (OD_{600}=0.1). This culture of each clone was incubated for 72 hours at 25°C and 220 RPM (at peak UPO activity; OD_{600}=25-30), the cells were separated by centrifugation (10 minutes at 4,500 RPM and 4°C) and the supernatant was filtered (using a glass and nitrocellulose filter with a pore size of 0.45 μm). Appropriate supernatant dilutions were prepared so that aliquots of 20 μL would give rise to a linear response in kinetic mode. 50 μL of supernatant were used for each point at a temperature gradient created using a thermocyclator, from 30 to 80°C. After incubating for 10 minutes, the aliquots were cooled in ice for 10 minutes and tempered at room temperature for 5 minutes. Lastly, these supernatants were subjected to the colorimetric assay using ABTS (100 mM pH 4.0 sodium phosphate/citrate buffer, 0.3 mM ABTS and 2 mM H_2O_2). The thermostability values were calculated in accordance with the ratio between the residual activities incubated at different temperatures and the value of initial activity at room temperature. The value of T_{50} was determined as as the temperature value at which the protein loses 50% of its initial activity after incubating for 30 minutes.

Production of UPO recombinant variants in *S. cerevisiae*

[0181] An independent *S. cerevisiae* colony that comprised the corresponding variant of the invention, on the one hand JaWa and on the other SoLo, was selected from a SC drop-out minimal medium plate and inoculated in 20 mL of liquid SC minimal medium, cultures that were incubated for 48 hours at 30°C and 220 RPM. An aliquot of this culture was taken so that, upon inoculating it in 100 mL of new minimal medium, OD_{600} would have a value of 0.25. This starter

was incubated until completing two full growth cycles (between 6 and 8 hours), at which time 100 mL of cells were taken to inoculate 900 mL of minimal medium for expression in a 2,000 mL flask (OD_{600}=0.1). This culture of each clone was incubated for 72 hours at 25°C and at 220 RPM (at peak UPO activity; OD_{600}=25-30), the cells were separated by centrifugation (10 minutes at 4,500 RPM and 4°C) and the supernatant was filtered (using a glass and nitrocellulose filter with a pore size of 0.45 μm).

Purification of recombinant *Aae*UPO1 variants

[0182] The purification of the variants described in the present invention, JaWa and SoLo, was carried out using cation-exchange chromatography followed by anion-exchange chromatography (ÄKTA purifier, GE Healthcare). The raw extract was concentrated and dialysed in 20 mM pH 3.3 sodium phosphate/citrate buffer (buffer A) by means of tangential ultrafiltration (Pellicon; Millipore, Temecula, CA, USA) through a membrane with a pore size of 10 kDa (Millipore) using a peristaltic pump (Masterflex Easy Load; Cole-Parmer, Vernon Hills, IL). The sample was filtered and loaded on a strong cation-exchange column (HiTrap SP FF, GE Healthcare), pre-balanced with buffer A. The proteins were eluded by means of a linear gradient of 0 to 40% of buffer A with 1M NaCl in 60 mL and from 40 to 100% of buffer A with 1 M NaCl in 5 mL, at a flow rate of 1 mL/min. The fractions with UPO activity were recovered, concentrated and dialysed in 20 mM pH 7.8 Tris-HCl buffer (buffer B) and loaded on a high-resolution anion-exchange column (Biosuite Q, Waters), pre-balanced with buffer B. The proteins were eluded by means of a linear gradient of 0 to 20% of buffer B with 1 M NaCl in 40 mL and from 20 to 100% of buffer B with 1 M NaCl in 5 mL, at a flow rate of 1 mL/min. The fractions with UPO activity were recovered, concentrated and dialysed in 10 mM pH 7.0 potassium phosphate buffer and stored at 4°C. Reinheitszahl [Rz] [A_{418}/A_{280}] values of ~2 were obtained. The fractions of the different purification steps were analysed in a 12% SDS/PAGE acrylamide gel, dyed with Coomassie blue. The concentrations of the raw extracts of these steps were determined by means of Bradford reagent and BSA as standard.

Kinetic constants values

[0183] The kinetic constants of the variants of the invention, *Aae*UPO, PaDa-I, JaWa and SoLo, for ABTS were estimated in 100 mM pH 4.0 sodium phosphate/citrate buffer and 2 mM H_2O_2; and for the other substrates, in 100 mM pH 7.0 potassium phosphate buffer and 2 mM H_2O_2 (propranolol). For H_2O_2, benzyl alcohol was used as substrate at the corresponding saturation conditions. The reactions were performed in triplicate and the oxidations of the substrates were followed by spectrophotometric changes (ABTS: ε_{418}=36,000 M^{-1} cm^{-1}; Propranolol: ε_{325}: 1,996 M^{-1} cm^{-1}; and benzyl alcohol: ε_{280}=1,400 M^{-1} cm^{-1}). The kinetics for propranolol were performed calculating ε_{325} experimentally at pH 7.0. In order to calculate the values of K_m and k_{cat}, values of V_{max} were represented at substrate concentrations and the hyperbole function was adjusted (using SigmaPlot 10.0, wherein the parameter *a* is equal to k_{cat} and the parameter *b*, to K_m).

HPLC analysis

[0184] The reactions were analysed by means of chromatography in reverse phase (HPLC). The equipment was composed of a tertiary pump (Varian-Agilent Technologies, USA) coupled to an autosampler (Merck Millipore, MA, USA); for the separation, a Zorbax Eclipse plus C18 column (15 cm x 4.6 cm) at 40°C was used and the detection was performed using a photodiode detector (PDA) (Varian, Agilent Technologies, USA). The mobile phase selected was a gradient from 10% methanol and 90% *dd*H_2O (in both cases with 0.1% of acetic acid) up to 90% methanol and 10% *dd*H_2O at a flow rate of 0.8 mL/min. The reaction was quantified at 280 nm (based on HPLC standards). For the 15 minute reaction, the mixture contained 0.03 μM of pure enzyme, 4 mM propranolol and 2 mM H_2O_2 in 50 mM pH 7.0 potassium phosphate buffer (0.5 mL of final volume). The reaction was started with the addition of H_2O_2 and was stopped with 20 μL of 37% HCl. In order to determine the turnover rates of the variants with 5'-hydroxypropranolol (product of interest), the mixture contained 0.03 μM of pure enzyme, 0.4 mM 5'-hydroxypropranolol and 2 mM H_2O_2 in 50 mM pH 7.0 potassium phosphate buffer (0.3 mL of final volume). In order to calculate the total turnover number (TTN) of the assayed variants, the assay was carried out using 0.03 μM of pure enzyme, 5 mM propranolol and 2 mM H_2O_2 in 50 mM pH 7.0 potassium phosphate buffer and in the same manner, but also adding 4 mM ascorbic acid. In both cases, 2 mM H_2O_2 was added every 10 minutes, monitoring the reaction in each addition point taking different aliquots. The standard deviations were less than 5% in all cases.

Analysis by Liquid Chromatography/Mass Spectrometry (LC/MS)

[0185] These analyses were performed using a mass spectrometer with a Q-TOF hybrid analyser (QSTAR, ABSciex, MA, USA). Electrospray (ESI) was used as an ionisation source and, as ionising phase, methanol. In this case, the

entrance system was direct injection in a HPLC 1100 (Agilent Technologies, USA). The resolution of the assay corresponds to 9,000 FWHM (*Full Width at Half Maximum*), accuracy at 5-10 ppm and it was performed in positive mode.

RESULTS

[0186] The activity of the different UPO variants was evaluated by means of the 4-AAP assay to determine the most appropriate starting point for determining the capacity of said variants for HDM synthesis (**Figure 9**). As can be observed in the figure, the variant with the greatest activity against propranolol and best ratio among its activity with and without ascorbic was JaWa (SEQ ID NO: 24, encoded by SEQ ID NO: 23), due to which it was the mutant selected for the docking assays (**Figure 10**). Based on these results, wherein it was observed that the substrate interacted with a series of residues of the catalytic pocket and of the heme access channel, a region of the JaWa mutant that was in direct contact with the substrate was selected (residues D187-V248 of SEQ ID NO: 24). The objective is to obtain a mutant enzyme or variant that shows less peroxidase activity on 5'-hydroxypropranolol (which is the product of the reaction with propranolol) while improving peroxygenase activity on propranolol, also taking into account that said variant must be expressed robustly in heterologous organisms and secreted in an active, soluble and very stable form. To this end, each variant obtained in the mutant libraries was subjected to double screening designed *ad hoc* for the purpose of obtaining the variants with the aforementioned capabilities, greater peroxygenase activity on propranolol (measured in the presence of ascorbic acid) and less peroxidase activity against 5'-hydroxypropranolol (in the absence of ascorbic acid). Two libraries with different mutagenic rates (concentration of $MnCl_2$) were analysed, identifying a single mutant in both libraries and repeatedly to that called SoLo and which comprises the nucleotide sequence SEQ ID NO: 41 that encodes the variant of SEQ ID NO: 42. Said SoLo mutant (SEQ ID NO: 42) has the F191S mutation (**Figure 11**) with respect to the JaWa mutant of SEQ ID NO: 24, with a peroxygenase activity on microplate 30% higher than its parental (JaWa) and decrease in peroxidase activity of more than two fold.

[0187] Two further cycles of evolution (2G and 3G) were performed using the SoLo variant (SEQ ID NO: 41) as parental, wherein no enhanced variant was detected.

Both variants, JaWa (SEQ ID NO: 24) and SoLo (SEQ ID NO: 42), were produced, purified at homogeneity (Reinheitszahl [Rz] [A_{418}/A_{280}] value ~2) and biochemically characterised.

[0188] As can be observed in **Figure 12,** the SoLo variant of SEQ ID NO: 42 showed very similar kinetic thermostability to that of the JaWa mutant (SEQ ID NO: 24).

[0189] The propranolol transformation reaction performed by the wild *Aae*UPO enzyme (SEQ ID NO: 2), and the PaDa-I (SEQ ID NO: 18), JaWa (SEQ ID NO: 24) and SoLo (SEQ ID NO: 42) variants in the absence of ascorbic acid and was analysed using HPLC-PDA is included in **Figure 13.** It can be observed that both JaWa and SoLo are those that produce the largest amount of 5'-hydroxypropranolol, in addition to having 99% of regioselectivity, since traces of neither 4'-hydroxypropranolol nor N-desisopropyl propranolol (DYP) were detected.

[0190] The kinetic value of *Aae*UPO, JaWa and SoLo for propranolol, and for ABTS and H_2O_2 (**Table 3**) were determined.

Table 3. Kinetic parameters for the variants of the invention and for wild *Aae*UPO.

Substrate	Kinetic constants	*Aae*UPO1	PaDa-I	JaWa	SoLo
ABTS	Km (μM)	25-0±2.5	48.8±4.5	181±22	568±91
	K_{cat} (s^{-1})	221±6	395±13	125±5	365±23
	K_{cat}/K_m ($s^{-1}M^{-1}$)	$8.8x10^6±6.9x10^5$	$8.2x10^6±6.0x10^5$	$6.9x10^5±6.3x10^4$	$6.4x10^5±6.7x10^4$
Propranolol	Km (μM)	2,239±333	2,268±220	244±92	391±97
	K_{cat} (s^{-1})	150±12	212±11	765±76	497±35
	K_{cat}/K_m ($s^{-1}M^{-1}$)	$6.7x10^4±4.8x10^3$	$9.3x10^4±4.3x10^3$	$3.1x10^6±0.9x10^5$	$1.3x10^6±0.2x10^5$
Naphthalene	K_m (μM)	156±20	578±106	127±27	789±96
	K_{cat} (s^{-1})	92±3	229±17	78±3	337±20
	K_{cat}/K_m ($s^{-1}M^{-1}$)	$5.9x10^5±5.9x10^4$	$4.0x10^5±4.0x10^4$	$6.2x10^5±1.1x10^4$	$4.3x10^5±2.8x10^4$
H_2O_2	Km (μM)	1,370±162	486±55	1,250±153	1,430±153
	K_{cat} (s^{-1})	290±15	238±8	446±23	446.23.
	K_{cat}/K_m ($s^{-1}M^{-1}$)	$2.1x10^5±1.5x10^4$	$5.0x10^5±4.2x10^4$	$3.1x10^5±1.8x10^4$	$3.1x10^5±1.8x10^4$

[0191] As can be observed in **Table 3,** both the JaWa (SEQ ID NO: 24) and SoLo (SEQ ID NO: 42) variants increased the k_{cat}/K_m (catalytic efficiency) values for propranolol by two orders of magnitude. It can also be observed that the JaWa

(SEQ ID NO: 24) and SoLo (SEQ ID NO: 42) variants show a reduction in peroxidase activity, measured with ABTS, of one order of magnitude in catalytic efficiency, being the affinity to the substrate, in the case of the SoLo variant, three fold worse with respect to its parental. The values for H_2O_2 with benzyl alcohol were not affected. As in the case of the propranolol between JaWa and SoLo, JaWa has kinetic constants similar to *Aae*UPO with the naphthalene as substrate, differentiating itself in the total turnover values, which are higher for JaWa.

[0192] Since the kinetics with propranolol of the JaWa and SoLo variants are very similar, the turnover rates were calculated with 5'-hydroxypropranolol as a substrate in the absence of ascorbic acid, in order to evaluate the peroxidase activity of each variant against its propranolol reaction product. In **Figure 14** it can be observed that JaWa and AaeUPO oxidise practically the entire product, but SoLo is capable of maintaining approximately 50% thereof without oxidising. It follows that the SoLo variant (SEQ ID NO: 42), has significantly reduced its peroxidase activity on its own product, allowing higher performances in the production of this propranolol metabolite.

[0193] When the reaction was monitored for long reaction times with the addition of 2 mM H_2O_2, the total turnover numbers (TTNs) were determined, obtaining a value of 45,000 for SoLo, 15,000 for JaWa and 3,000 for AaeUPO in the absence of ascorbic acid; and in the presence of ascorbic acid, 62,000 for SoLo, 48,000 for JaWa and 14,000 for AaeUPO (**Table 4**). This implies that, even by adding ascorbic acid to the reaction, the independent use of this antioxidant in the reaction medium is possible, simplifying the process. (**Figure 15**).

Table 4. Determination of the total turnover numbers (TTNs) for the variants of the invention and for wild *Aae*UPO.

	TTNs	
	With ascorbic acid	**Without ascorbic acid**
***Aae*UPO**	14,000	3,000
JaWa	48,000	15,000
SoLo	62,000	45,000

SEQUENCE LISTING

<110> Centro Superior de Investigaciones Científicas (CSIC)

<120> MUTANTS OF UNSPECIFIC PEROXYGENASE WITH HIGH MONOOXYGENASE

<130> 1641.1156

<150> P201531641
<151> 2015-11-13

<160> 66

<170> PatentIn version 3.5

<210> 1
<211> 987
<212> DNA
<213> Agrocybe aegerita

<220>
<221> CDS
<222> (1)..(987)

<400> 1

```
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg        48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                  10                  15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att        96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc       144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

ccg aga aat ggc gtt gca acc ccg gtg caa ata ata aac gcg gtt cag       192
Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

gaa gga ctc aat ttc gac aat caa gcc gca gtc ttc gcc aca tat gcg       240
Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65                  70                  75                  80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga       288
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt       336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc       384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc       432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac       480
```

```
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc  528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc ttt act  576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc  624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa  672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt  720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

ggc aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga  768
Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct  816
Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg  864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt  912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc  960
Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

cag gtc ttc cca tac ggg cga gat tga                              987
Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 2
<211> 328
<212> PRT
<213> Agrocybe aegerita

<400> 2

```
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30
```

```
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65                  70                  75                  80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285
```

```
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 3
<211> 1116
<212> DNA
<213> Agrocybe aegerita

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1116)

<400> 3

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc         336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100
```

```
act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt ata cag gct cat cct atg cag ccc gga aga aat gtc ggc aag        912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 4

<211> 371
<212> PRT
<213> Agrocybe aegerita

<400> 4

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1                   5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
```

185 190 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
200 205 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
215 220 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230 235 240 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
250 255 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
265 270 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
280 285 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
295 300 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310 315 320 325

Gly Arg Asp

<210> 5
<211> 1116
<212> DNA
<213> Agrocybe aegerita

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1116)

<400> 5
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc 48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
-40 -35 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa 96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
-25 -20 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct 144

```
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1                5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct  192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct  240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt  288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc  336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac  384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc  432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag  480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc  528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt  576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg  624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc  672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc  720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag  768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct  816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag  864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245
```

```
gta gtt ata cag gct cat cct atg cag ccc gga aga aat gtc ggc aag        912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 6
<211> 371
<212> PRT
<213> Agrocybe aegerita

<400> 6

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85
```

```
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 7
<211> 987
<212> DNA
<213> Artificial Sequence

<220>
<223> wt-JaWa Variant

<220>
<221> CDS
<222> (1)..(987)

<400> 7

```
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg          48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att          96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc         144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

ccg aga aat ggc gtt gca acc ccg gtg caa ata ata aac gcg gtt cag         192
Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

gaa gga ctc aat ttc gac aat caa gcc gca gtc ttc gcc aca tat gcg         240
Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65                  70                  75                  80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga         288
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt         336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc         384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc         432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac         480
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc         528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc ttt act         576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190
```

```
gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc        624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa        672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt        720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

gac aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga        768
Asp Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aaa aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct        816
Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg        864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt        912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc        960
Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

cag gtc ttc cca tac ggg cga gat tga                                    987
Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 8
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 8

```
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
    50                  55                  60
```

```
Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65                  70                  75                  80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Asp Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320
```

```
Gln Val Phe Pro Tyr Gly Arg Asp
                325


<210>  9
<211>  1116
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Wt-JaWa variant with wild signal peptide


<220>
<221>  CDS
<222>  (1)..(1113)

<220>
<221>  mat_peptide
<222>  (130)..(1116)

<400>  9
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc         336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag         480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115
```

```
cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

gta gtt ata cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 10
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 10

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195
```

```
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 11
<211> 1116
<212> DNA
<213> Artificial Sequence

<220>
<223> Wt-JaWa variant with modified signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1116)

<400> 11

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc        48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa        96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15
```

```
ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct        144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1                5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct        192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct        240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt        288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc        336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac        384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc        432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
```

```
230                 235                 240                 245

gta gtt ata cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 12
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 12

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65
```

```
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
```

310 315 320 325

Gly Arg Asp

<210> 13
<211> 984
<212> DNA
<213> Artificial Sequence

<220>
<223> PaDa-I variant without signal peptide

<220>
<221> CDS
<222> (1)..(984)

<400> 13

```
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg          48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                  10                  15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att          96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc         144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

ccg aga aat ggc gtt gca acc ccg gcg caa ata ata aac gcg gtt cag         192
Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

gaa gga ttc aat ttc gac aat caa gcc gca atc ttc gcc aca tat gcg         240
Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65                  70                  75                  80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga         288
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt         336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc         384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc         432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac         480
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc         528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
```

```
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc ttt act        576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc        624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa        672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt        720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

ggc aca gga gtc gag gta gtt gta cag gct cat cct atg cag ccc gga        768
Gly Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct        816
Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg        864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt        912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc        960
Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

cag gtc ttc cca tac ggg cga gat                                      984
Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 14
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 14

```
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45
```

```
Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65                  70                  75                  80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Gly Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
```

```
    290                 295                 300

Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

Gln Val Phe Pro Tyr Gly Arg Asp
                325

<210>  15
<211>  1113
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  PaDa-I with wild signal peptide

<220>
<221>  CDS
<222>  (1)..(1113)

<220>
<221>  mat_peptide
<222>  (130)..(1113)

<400>  15
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gtc        48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa        96
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct       144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct       192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct       240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt       288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc       336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac       384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc       432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
```

```
                    90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aga aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                          1113
Gly Arg Asp
```

<210> 16
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 16

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
```

170 175 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
185 190 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
200 205 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
215 220 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230 235 240 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
250 255 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
265 270 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
280 285 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
295 300 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310 315 320 325

Gly Arg Asp

<210> 17
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> PaDa-I with modified signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 17
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc 48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val

```
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa        96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct       144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct       192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct       240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt       288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc       336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac       384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc       432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag       480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc       528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt       576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg       624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc       672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc       720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag       768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct       816
```

```
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag      864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aga aat gtc ggc aag      912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc      960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac     1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat     1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac     1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                        1113
Gly Arg Asp
```

<210> 18
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 18

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
```

```
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290
```

```
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 19
<211> 984
<212> DNA
<213> Artificial Sequence

<220>
<223> JaWa without signal peptide

<220>
<221> CDS
<222> (1)..(984)

<400> 19

```
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg          48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att          96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
            20                  25                  30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc         144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

ccg aga aat ggc gtt gca acc ccg gcg caa ata ata aac gcg gtt cag         192
Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

gaa gga ttc aat ttc gac aat caa gcc gca atc ttc gcc aca tat gcg         240
Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65                  70                  75                  80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga         288
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt         336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc         384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc         432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac         480
```

```
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc  528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc ttt act  576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc  624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa  672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt  720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

gac aca gga gtc gag gta gtt gta cag gct cat cct atg cag ccc gga  768
Asp Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aaa aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct  816
Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg  864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt  912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc  960
Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

cag gtc ttc cca tac ggg cga gat                                 984
Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 20
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 20

```
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1               5                   10                  15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
```

```
                20                  25                  30

Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
        35                  40                  45

Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
    50                  55                  60

Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65                  70                  75                  80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
            180                 185                 190

Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Asp Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270
```

```
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 21
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> JaWa with wild signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 21

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc         336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
```

```
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc        432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325
```

ggg cga gat 1113
Gly Arg Asp

<210> 22
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 22

Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
-40 -35 -30

Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
-25 -20 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
-10 -5 -1 1 5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
10 15 20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
25 30 35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
40 45 50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
55 60 65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70 75 80 85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
90 95 100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
105 110 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
120 125 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
135 140 145

```
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 23
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> JaWa with wild modified signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide

<222> (130)..(1113)

<400> 23

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc         336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag         480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc         528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt         576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg         624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc         672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc         720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195
```

```
acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                          1113
Gly Arg Asp
```

<210> 24
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 24

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20
```

```
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275
```

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
280 285 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
295 300 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310 315 320 325

Gly Arg Asp

<210> 25
<211> 129
<212> DNA
<213> Agrocybe aegerita

<220>
<221> CDS
<222> (1)..(129)

<400> 25
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gtc 48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
1 5 10 15

gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa 96
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
20 25 30

ttg gac gct ata atc cca aca ctc gag gcc cga 129
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg
35 40

<210> 26
<211> 43
<212> PRT
<213> Agrocybe aegerita

<400> 26

Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
1 5 10 15

Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
20 25 30

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg
35 40

<210> 27
<211> 129
<212> DNA

<213> Artificial Sequence

<220>
<223> Signal peptide modified with respect to the wild signal peptide from A. aegerita

<220>
<221> CDS
<222> (1)..(129)

<400> 27
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc 48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
1 5 10 15

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa 96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
20 25 30

ttg gac gct ata atc cca aca ctc gag gcc cga 129
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg
35 40

<210> 28
<211> 43
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 28

Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
1 5 10 15

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
20 25 30

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg
35 40

<210> 29
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> W24F variant obtained from the PaDa-I variant

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 29

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1                5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca ttt aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Phe Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc         336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag         480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc         528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt         576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg         624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc         672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc         720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag         768
```

```
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct      816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag      864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aga aat gtc ggc aag      912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc      960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac     1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat     1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac     1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                        1113
Gly Arg Asp
```

<210> 30
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 30

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Phe Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
```

```
                25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275
```

```
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 31
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> W24F variant obtained from the Jawa variant

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 31

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1                5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca ttt aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Phe Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc         336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
```

```
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc        432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325
```

ggg cga gat 1113
Gly Arg Asp

<210> 32
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 32

Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
-40 -35 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
-25 -20 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
-10 -5 -1 1 5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
10 15 20

His Pro Phe Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
25 30 35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
40 45 50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
55 60 65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70 75 80 85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
90 95 100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
105 110 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
120 125 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
135 140 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150 155 160 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
170 175 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr
185 190 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
200 205 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
215 220 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230 235 240 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
250 255 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
265 270 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
280 285 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
295 300 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310 315 320 325

Gly Arg Asp

<210> 33
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> RMLN primer

<400> 33
cctctatact ttaacgtcaa gg 22

<210> 34
<211> 20
<212> DNA

<213> Artificial Sequence

<220>
<223> Cebador RMLC

<400> 34
gggagggcgt gaatgtaagc 20

<210> 35
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> F24FOR primer

<400> 35
ctcacccatt taagccgctt cgacctggcg atattcgtgg ac 42

<210> 36
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> F24REV primer

<400> 36
gtccacgaat atcgccaggt cgaagcggct taaatgggtg ag 42

<210> 37
<211> 984
<212> DNA
<213> Artificial Sequence

<220>
<223> SoLo variant without signal peptide

<220>
<221> CDS
<222> (1)..(984)

<400> 37
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg 48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1 5 10 15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att 96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
20 25 30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc 144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
35 40 45

ccg aga aat ggc gtt gca acc ccg gcg caa ata ata aac gcg gtt cag 192
Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
50 55 60

```
gaa gga ttc aat ttc gac aat caa gcc gca atc ttc gcc aca tat gcg        240
Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65                  70                  75                  80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga        288
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt        336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc        384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc        432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac        480
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc        528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act        576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr
            180                 185                 190

gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc        624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa        672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt        720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

gac aca gga gtc gag gta gtt gta cag gct cat cct atg cag ccc gga        768
Asp Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aaa aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct        816
Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg        864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt        912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc        960
Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320
```

cag gtc ttc cca tac ggg cga gat 984
Gln Val Phe Pro Tyr Gly Arg Asp
325

<210> 38
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 38

Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1 5 10 15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
20 25 30

Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
35 40 45

Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
50 55 60

Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
65 70 75 80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
85 90 95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
100 105 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
115 120 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
130 135 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145 150 155 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
165 170 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr
180 185 190

```
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Asp Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 39
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> SoLo variant with wild signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 39

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gtc        48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa        96
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15
```

```
ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct        144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1                5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct        192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct        240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt        288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc        336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac        384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc        432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag        480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc tct act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245
```

```
gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                          1113
Gly Arg Asp
```

<210> 40
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 40

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
            -40                 -35                 -30

Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65
```

```
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325
```

Gly Arg Asp

<210> 41
<211> 1113
<212> DNA
<213> Artificial Sequence

<220>
<223> SoLo variant with modified signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1113)

<400> 41

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gcg caa ata ata aac gcg gtt cag gaa gga ttc aat ttc         336
Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

gac aat caa gcc gca atc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag         480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115
```

```
cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc        528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc tct act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230                 235                 240                 245

gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag        912
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat                                                          1113
Gly Arg Asp
```

<210> 42
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 42

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195
```

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
200 205 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
215 220 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
230 235 240 245

Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
250 255 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
265 270 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
280 285 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
295 300 305

Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310 315 320 325

Gly Arg Asp

<210> 43
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> MJaWa Fw primer

<400> 43
gcgcattcaa gactccattg 20

<210> 44
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> MJaWa Rev primer

<400> 44
gatcttgccg acatttttttc c 21

<210> 45

<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> HFJaWa Fw primer

<400> 45
caggctcatc ctatgcagcc c 21

<210> 46
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> HFJaWa Rev primer

<400> 46
caaaggagaa attggggttg gtcg 24

<210> 47
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> VHG R primer

<220>
<221> misc_feature
<222> (51)..(51)
<223> d is a or g or t

<220>
<221> misc_feature
<222> (52)..(52)
<223> b is g or c or t

<400> 47
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggcc dbgattgcgg 60

c 61

<210> 48
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> NDT R primer

<220>
<221> misc_feature
<222> (51)..(51)
<223> h is a or c or t

<220>

<221> misc_feature
<222> (52)..(52)
<223> n is a or g or c or t

<400> 48
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggca hngattgcgg 60

c 61

<210> 49
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> TGG R primer

<400> 49
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggcc cagattgcgg 60

c 61

<210> 50
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> HF F primer

<400> 50
gcggcccacc ttgtggacgg caatctcatt acggacttgc 40

<210> 51
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> S191 VHG R primer

<220>
<221> misc_feature
<222> (50)..(50)
<223> d is a or g or t

<220>
<221> misc_feature
<222> (51)..(51)
<223> b is g or c or t

<400> 51
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtcd bgaacctaaa 60

g 61

<210> 52
<211> 61

<212> DNA
<213> Artificial Sequence

<220>
<223> NDT R primer

<220>
<221> misc_feature
<222> (50)..(50)
<223> h is a or c or t

<220>
<221> misc_feature
<222> (51)..(51)
<223> h is a or g or c or t

<400> 52
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtah ngaacctaaa 60

g 61

<210> 53
<211> 61
<212> DNA
<213> Artificial Sequence

<220>
<223> S191 TGG R primer

<400> 53
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtcc agaacctaaa 60

g 61

<210> 54
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> HF F-RMLC primer

<400> 54
cggcgagacc accttccccg cgaatctttt tgtggatggg 40

<210> 55
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> F69 R primer

<220>
<221> misc_feature
<222> (22)..(22)
<223> m is a or c

<220>
<221> misc_feature
<222> (23)..(23)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a or g or c or t

<400> 55
gaagattgcg gcttgattgt cmnnattgaa tc 32

<210> 56
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> F69 F primer

<220>
<221> misc_feature
<222> (23)..(23)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (25)..(25)
<223> k is g or t

<400> 56
cgcggttcag gaaggattca atnnkgacaa tc 32

<210> 57
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> F121 R primer

<220>
<221> misc_feature
<222> (19)..(19)
<223> m is a or c

<220>
<221> misc_feature
<222> (20)..(20)
<223> n is a or g or c or t

<220>
<221> misc_feature

<222> (21)..(21)
<223> n is a or g or c or t

<400> 57
catactggcg tcgccttcmn nggtgccatg c 31

<210> 58
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> F121 F primer

<220>
<221> misc_feature
<222> (22)..(22)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (23)..(23)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (24)..(24)
<223> k is g or t

<400> 58
ggactcaatg agcatggcac cnnkgaaggc g 31

<210> 59
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> F199 R primer

<220>
<221> misc_feature
<222> (20)..(20)
<223> m is a or c

<220>
<221> misc_feature
<222> (21)..(21)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (22)..(22)
<223> n is a or g or c or t

<400> 59
ccacaaaaag attcgcgggm nnggtggtct cg 32

<210> 60
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> F199 F primer

<220>
<221> misc_feature
<222> (24)..(24)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (25)..(25)
<223> n is a or g or c or t

<220>
<221> misc_feature
<222> (26)..(26)
<223> k is g or t

<400> 60
ctactgctta cggcgagacc accnnkcccg cg 32

<210> 61
<211> 987
<212> DNA
<213> Artificial Sequence

<220>
<223> Wt-SoLo variant without signal peptide

<220>
<221> CDS
<222> (1)..(987)

<400> 61
gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg 48
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1 5 10 15

gtg aac gac gag gct cac cca tgg aag ccg ctt cga cct ggc gat att 96
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
20 25 30

cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc 144
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
35 40 45

ccg aga aat ggc gtt gca acc ccg gtg caa ata ata aac gcg gtt cag 192
Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
50 55 60

gaa gga ctc aat ttc gac aat caa gcc gca gtc ttc gcc aca tat gcg 240
Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65 70 75 80

gcc cac ctt gtg gac ggc aat ctc att acg gac ttg ctg agc atc gga 288

```
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
                85                  90                  95

cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt        336
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
            100                 105                 110

ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc        384
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
        115                 120                 125

cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc        432
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
    130                 135                 140

ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac        480
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145                 150                 155                 160

aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc        528
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                165                 170                 175

att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act        576
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr
            180                 185                 190

gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc        624
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
        195                 200                 205

agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa        672
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt        720
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

ggc aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga        768
Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct        816
Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg        864
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt        912
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc        960
Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

cag gtc ttc cca tac ggg cga gat tga                                    987
Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 62
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 62

Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
1 5 10 15

Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
20 25 30

Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
35 40 45

Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
50 55 60

Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala
65 70 75 80

Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
85 90 95

Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
100 105 110

Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
115 120 125

Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
130 135 140

Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
145 150 155 160

Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
165 170 175

Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr
180 185 190

Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
195 200 205

```
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
    210                 215                 220

Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
225                 230                 235                 240

Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
                245                 250                 255

Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
            260                 265                 270

Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
        275                 280                 285

Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
    290                 295                 300

Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
305                 310                 315                 320

Gln Val Phe Pro Tyr Gly Arg Asp
                325
```

<210> 63
<211> 1116
<212> DNA
<213> Artificial Sequence

<220>
<223> Wt-SoLo variant with wild signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1116)

<400> 63

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc        48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa        96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct       144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct       192
```

```
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct  240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt  288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc  336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac  384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc  432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag  480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc  528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt  576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg  624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc  672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc tct act gct tac ggc gag acc  720
Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag  768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct  816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag  864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt ata cag gct cat cct atg cag ccc gga aga aat gtc ggc aag  912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260
```

```
atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 64
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 64

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85
```

```
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

<210> 65
<211> 1116
<212> DNA
<213> Artificial Sequence

<220>
<223> Wt-SoLo variant with modified signal peptide

<220>
<221> CDS
<222> (1)..(1113)

<220>
<221> mat_peptide
<222> (130)..(1116)

<400> 65

```
atg aaa tat ttt ccc ctg ttc cca acc ttg gtc tac gca gtg ggg gtc          48
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

gtt gct ttt cct gac tac gcc tca ttg gcc ggc ctc agc cag cag gaa          96
Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cct         144
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5              -1  1               5

cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gct         192
Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

cac cca tgg aag ccg ctt cga cct ggc gat att cgt gga cct tgc cct         240
His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

ggt ctc aat act ctg gca tct cac ggg tac ctc ccg aga aat ggc gtt         288
Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

gca acc ccg gtg caa ata ata aac gcg gtt cag gaa gga ctc aat ttc         336
Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

gac aat caa gcc gca gtc ttc gcc aca tat gcg gcc cac ctt gtg gac         384
Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

ggc aat ctc att acg gac ttg ctg agc atc gga cgc aag acg cgg ctc         432
Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

act ggg cct gat cca cca ccc ccc gct tcc gtt ggt gga ctc aat gag         480
Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

cat ggc acc ttc gaa ggc gac gcc agt atg acc cga ggt gac gca ttc         528
His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130
```

```
ttt ggc aac aac cac gat ttc aat gag acg ctc ttc gaa cag ttg gtt        576
Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg        624
Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc        672
Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

aat ttc tcc ttt gtt gac ttt agg ttc tct act gct tac ggc gag acc        720
Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag        768
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210

cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct        816
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

gac gat ttc ttc cgc gca ccc agc ccg aga agt ggc aca gga gtc gag        864
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

gta gtt ata cag gct cat cct atg cag ccc gga aga aat gtc ggc aag        912
Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc        960
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac       1008
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat       1056
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

ttc ttc ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac       1104
Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

ggg cga gat tga                                                      1116
Gly Arg Asp
```

<210> 66
<211> 371
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 66

```
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Tyr Ala Val Gly Val
            -40                 -35                 -30

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Glu
        -25                 -20                 -15

Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro
    -10                 -5                  -1  1               5

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala
                10                  15                  20

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro
            25                  30                  35

Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val
        40                  45                  50

Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe
    55                  60                  65

Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp
70                  75                  80                  85

Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu
                90                  95                  100

Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu
            105                 110                 115

His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe
        120                 125                 130

Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val
    135                 140                 145

Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala
150                 155                 160                 165

Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro
                170                 175                 180

Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr
            185                 190                 195

Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
        200                 205                 210
```

```
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
    215                 220                 225

Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu
230                 235                 240                 245

Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys
                250                 255                 260

Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
            265                 270                 275

Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
        280                 285                 290

Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
    295                 300                 305

Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
310                 315                 320                 325

Gly Arg Asp
```

Claims

1. A polynucleotide that encodes a polypeptide with peroxygenase activity, **characterised in that** the polypeptide amino acid sequence encoding having an identity of at least 70% with SEQ ID NO: 2 (*Aae*UPO1), and comprising at least two amino acid alterations in the homologous positions to positions 241 and 257 of the sequence, which replace the amino acids: original glycine (G) by aspartic acid (D) in position 241 (G241D mutation) and original arginine (R) by lysine (K) in position 257 (R257K mutation).

2. The polynucleotide, according to claim 1, **characterised in that** further comprises an amino acid alteration at the homologous position to position 191 of the sequence SEQ ID NO: 2, which replaces the original amino acid phenylalanine (F) by serine (S) (F191S).

3. The polynucleotide, according to any of claims 1 to 2, **characterised in that** its amino acid sequence further comprises at least one of the following additional mutations or any combination thereof:

   a) The replacement of the original amino acid leucine (L) by the amino acid phenylalanine (F) at the homologous position to position 67 of SEQ ID NO: 2 (L67F mutation),
   b) The replacement of the original amino acid isoleucine (I) by the amino acid valine (V) at the homologous position to position 248 of SEQ ID NO: 2 (I248V mutation),
   c) The replacement of the original amino acid phenylalanine (F) by the amino acid leucine (L) at the position homologous to position 311 of SEQ ID NO: 2 (F311L mutation),
   d) The replacement of the original amino acid valine (V) by the amino acid isoleucine (I) at the homologous position to position 75 of SEQ ID NO: 2 (V75I mutation) and
   e) The replacement of the original amino acid valine (V) by the amino acid alanine (A) at the homologous position to position 57 of SEQ ID NO: 2 (V57A mutation).

4. The polynucleotide, according to any of claims 1 to 3, **characterised in that** further having the nucleotide sequence that encodes the signal peptide of SEQ ID NO: 26.

5. The polynucleotide, according to claim 4, **characterised in that** the nucleotide sequence that encodes the signal peptide of SEQ ID NO: 26 comprises at least one of the following mutations or any combination thereof:

a) The replacement of the original amino acid phenylalanine (F) by the amino acid tyrosine (Y) at the homologous position to position 12 of SEQ ID NO: 26 (F[12]Y),
b) The replacement of the original amino acid alanine (A) by the amino acid valine (V) in the homologous position to position 14 of SEQ ID NO: 26 (A[14]V),
c) The replacement of the original amino acid arginine (R) by the original acid glycine (G) at the homologous position to position 15 of SEQ ID NO: 26 (R[15]G) and
d) The replacement of the original amino acid alanine (A) by the amino acid aspartic (D) at the homologous position to position 21 of SEQ ID NO: 26 (A[21]D).

6. The polynucleotide, according to claims 1 to 5, **characterised in that** its sequence is selected from the list consisting of SEQ ID NO: 11, SEQ ID NO: 9 and SEQ ID NO: 7.

7. The polynucleotide, according to claims 1 to 5, **characterised in that** its sequence is selected from the list consisting of SEQ ID NO: 23, SEQ ID NO: 21 and SEQ ID NO: 19.

8. The polynucleotide, according to claims 1 to 5, **characterised in that** its sequence is selected from the list consisting of SEQ ID NO: 41, SEQ ID NO: 39 and SEQ ID NO: 37.

9. The polypeptide, **characterised in that** it is encoded by any of the nucleotide sequences according to claims 1 to 8.

10. The polypeptide, according to claim 9, **characterised in that** its sequence is selected from the list consisting of: SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 24, SEQ ID NO: 38, SEQ ID NO: 40 and SEQ ID NO: 42.

11. The polypeptide, according to claim 10, **characterised in that** its sequence is selected from the list consisting of: SEQ ID NO: 24, SEQ ID NO: 22 and SEQ ID NO: 20.

12. The polypeptide, according to claim 10, **characterised in that** its sequence is selected from the list consisting of: SEQ ID NO: 38, SEQ ID NO: 40 and SEQ ID NO: 42

13. A method for obtaining the polypeptide, according to any of claims 9 to 12, **characterised in that** it comprises the following steps:

i. Introducing a vector with the polynucleotide, according to any of claims 1 to 8, in a suitable host cell,
ii. culturing the host cell in a suitable medium, and,
iii. purifying the synthesised polypeptide.

14. A host cell **characterised in that** it comprises the polynucleotide according to claims 1 to 8 and is capable of producing the polypeptide according to claims 9 to 12.

15. The host cell, according to claim 14, **characterised in that** it is a yeast or fungus.

16. The host cell, according to claim 15, **characterised in that** the yeast belongs to the genus *Saccharomyces* sp or *Pichia* sp, particularly the *Saccharomyces cerevisiae* or *Pichia pastoris* species.

17. The host cell, according to claim 15, **characterised in that** the fungus belongs to the genus *Aspergillus* sp, particularly the *Aspergillus oryzae, Aspergillus niger or Aspergillus sojae species.*

18. A kit comprising at least one polypeptide according to claims 9 to 12.

19. An electronic device comprising at least one polypeptide according to claims 9 to 12.

20. Use of the polypeptide according to claims 9 to 12, the host cell according to any of claims 14 to 17, the kit according to claim 18 or the electronic device according to claim 19, in organic synthesis processes.

21. The use, according to claim 20, wherein the organic synthesis process is a hydrocarbon oxyfunctionalisation procedure, wherein the hydrocarbons are selected from any of the following list consisting of: aromatic, aliphatic, linear, branched and cyclic hydrocarbons.

22. The use according to any of claims 20 or 21, wherein the organic synthesis process is selected from the list consisting of: hydroxylation processes for the synthesis of 1-naphthol and/or 5'-hydroxypropranolol.

23. The use of the polypeptide according to any of claims 9 to 12, the host cell according to any of claims 14 to 17, the kit according to claim 18 or the electronic device according to claim 19, in bioremediation processes.

24. The use of the polypeptide according to claims 9 to 12, the host cell according to claims 14 to 17, the kit according to claim 18 or the electronic device according to claim 19, in the manufacture of biosensors.

25. The use of the polypeptide, according to any of claims 9 to 12, in the manufacture of pharmaceutical, nutritional and/or cosmetic compositions.

26. The use of the polypeptide, according to any of claims 9 to 12, in the manufacture of electronic devices containing immobilised enzymes.

27. The use of the electronic device, according to claim 19, to manufacture a pharmaceutical composition of biomedical diagnostics.

28. A method for hydroxylation of at least one hydrocarbon comprising the use of the polypeptide according to any of claims 9 to 12, the host cell according to any of claims 14 to 17, the kit according to claim 18 or the electronic device according to claim 19.

29. The method, according to claim 28, wherein the compound obtained from the hydroxylation is selected from the list consisting of: 1-naphthol and 5'-hydroxypropranolol.

FIG. 1

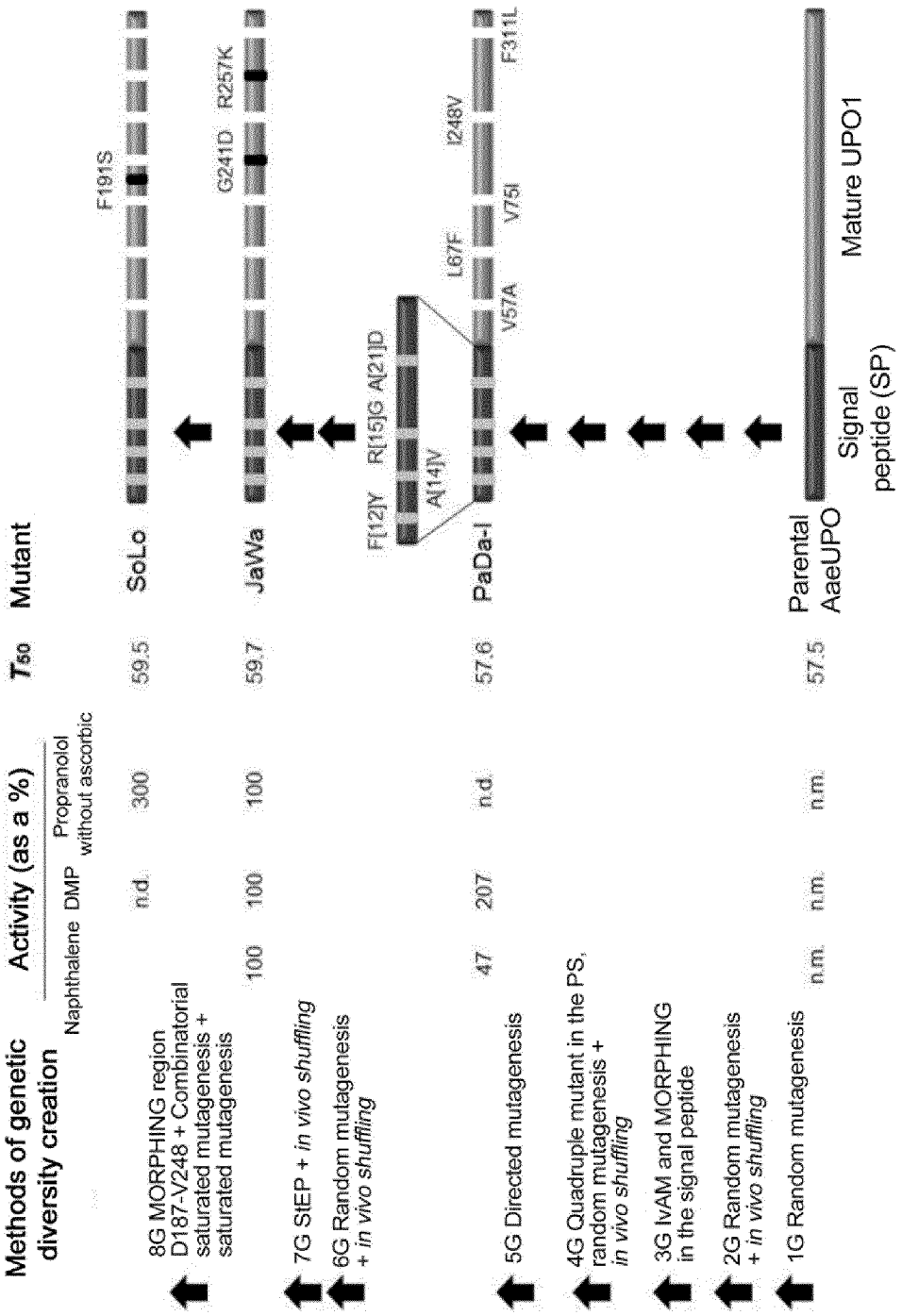

Methods of genetic diversity creation
Activity (as a %)
Naphthalene
DMP
Propranolol without ascorbic
T50
Mutant
8G MORPHING region D187-V248 + Combinatorial saturated mutagenesis + saturated mutagenesis
n.d.
300
59.5
SoLo
F191S
100
100
100
59.7
JaWa
G241D
R257K
7G StEP + in vivo shuffling
6G Random mutagenesis + in vivo shuffling
F[12]Y
R[15]G
A[21]D
A[14]V
L67F
I248V
47
207
n.d.
57.6
PaDa-I
V57A
V75I
F311L
5G Directed mutagenesis
4G Quadruple mutant in the PS, random mutagenesis + in vivo shuffling
3G IvAM and MORPHING in the signal peptide
2G Random mutagenesis + in vivo shuffling
1G Random mutagenesis
n.m.
n.m.
n.m.
57.5
Parental AaeUPO
Signal peptide (SP)
Mature UPO1

FIG. 2

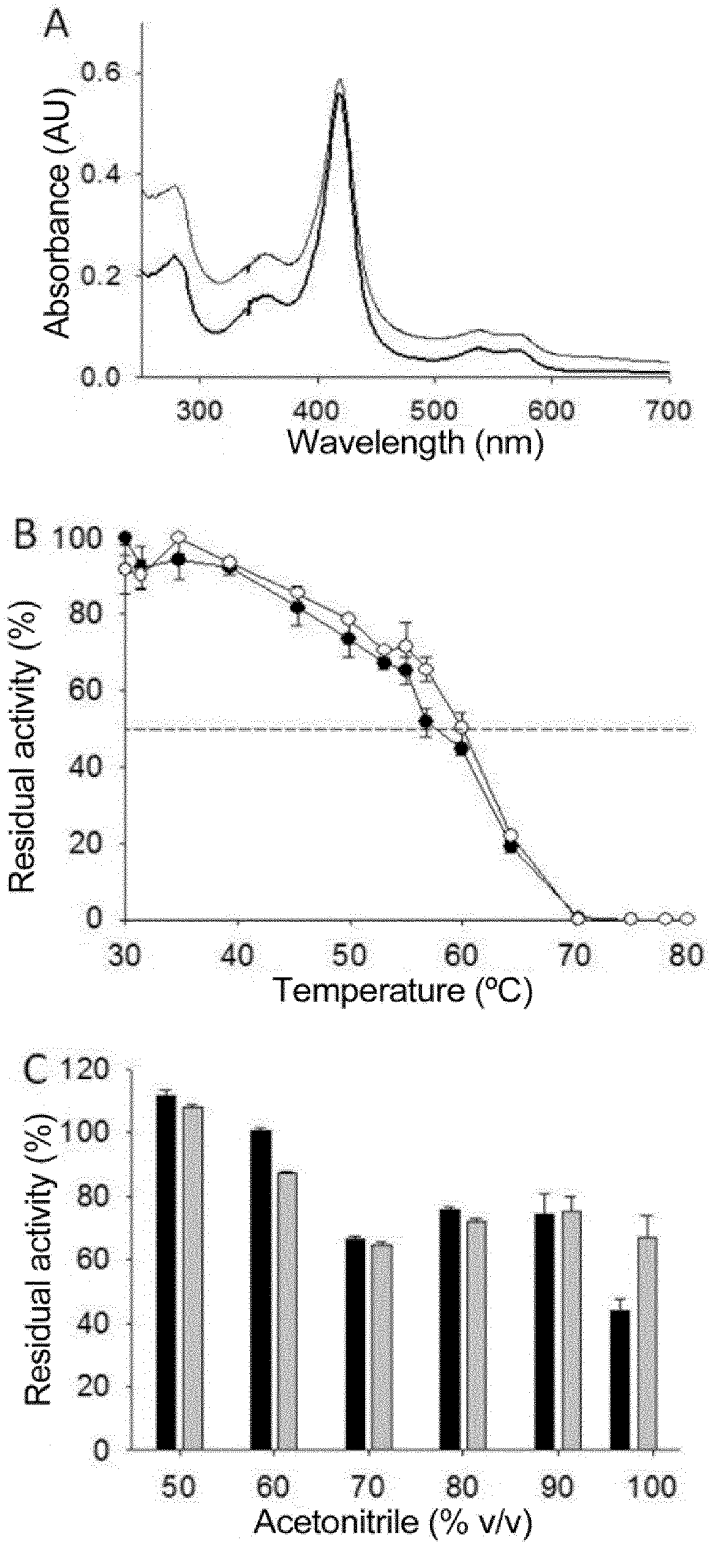

A
Absorbance (AU)
0.0
0.2
0.4
0.6
300
400
500
600
700
Wavelength (nm)
B
Residual activity (%)
0
20
40
60
80
100
30
40
50
60
70
80
Temperature (°C)
C
Residual activity (%)
0
20
40
60
80
100
120
50
60
70
80
90
100
Acetonitrile (% v/v)

FIG. 3

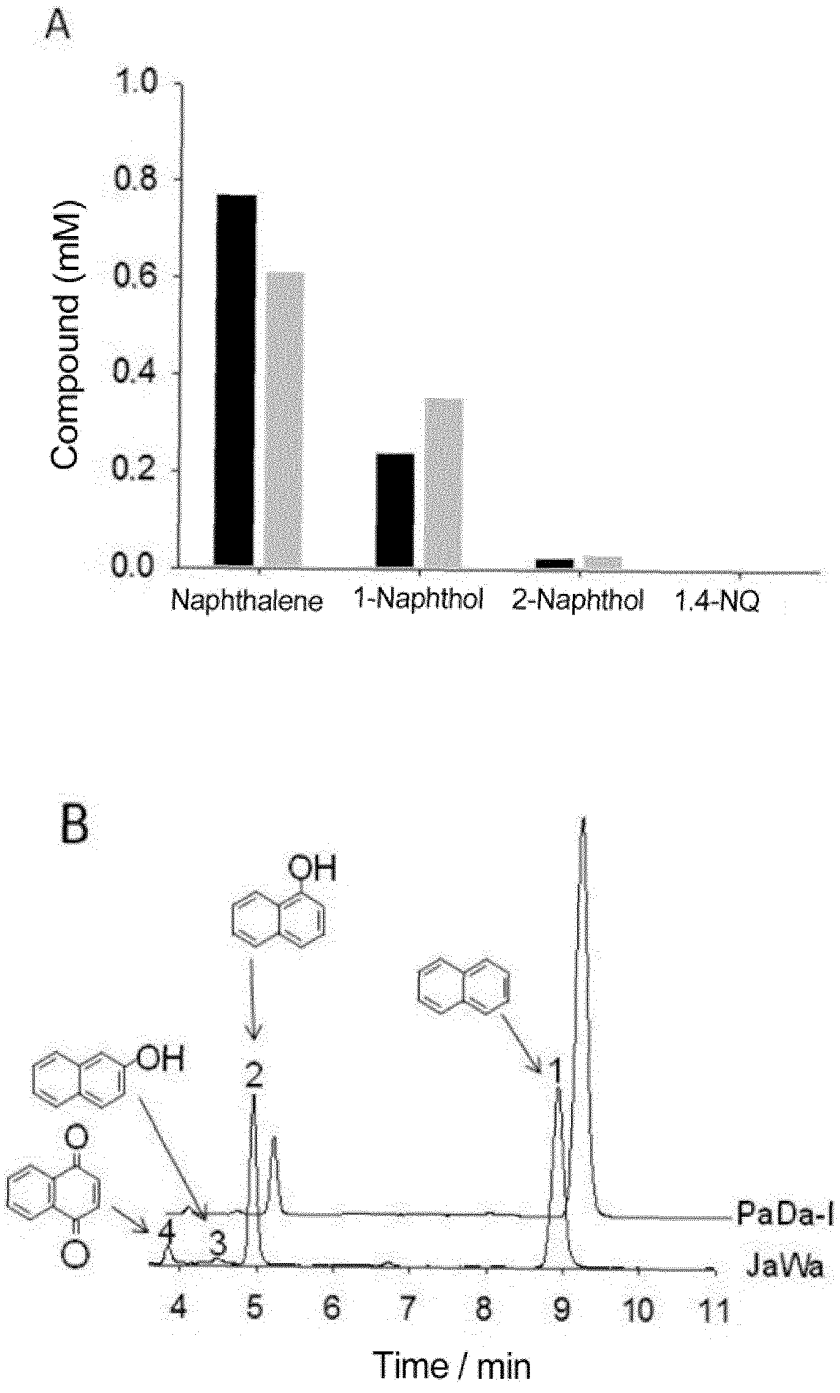

A
1.0
0.8
0.6
0.4
0.2
0.0
Compound (mM)
Naphthalene
1-Naphthol
2-Naphthol
1.4-NQ
B
OH
OH
O
O
2
1
4
3
PaDa-I
JaWa
4
5
6
7
8
9
10
11
Time / min

FIG. 3 (cont.)

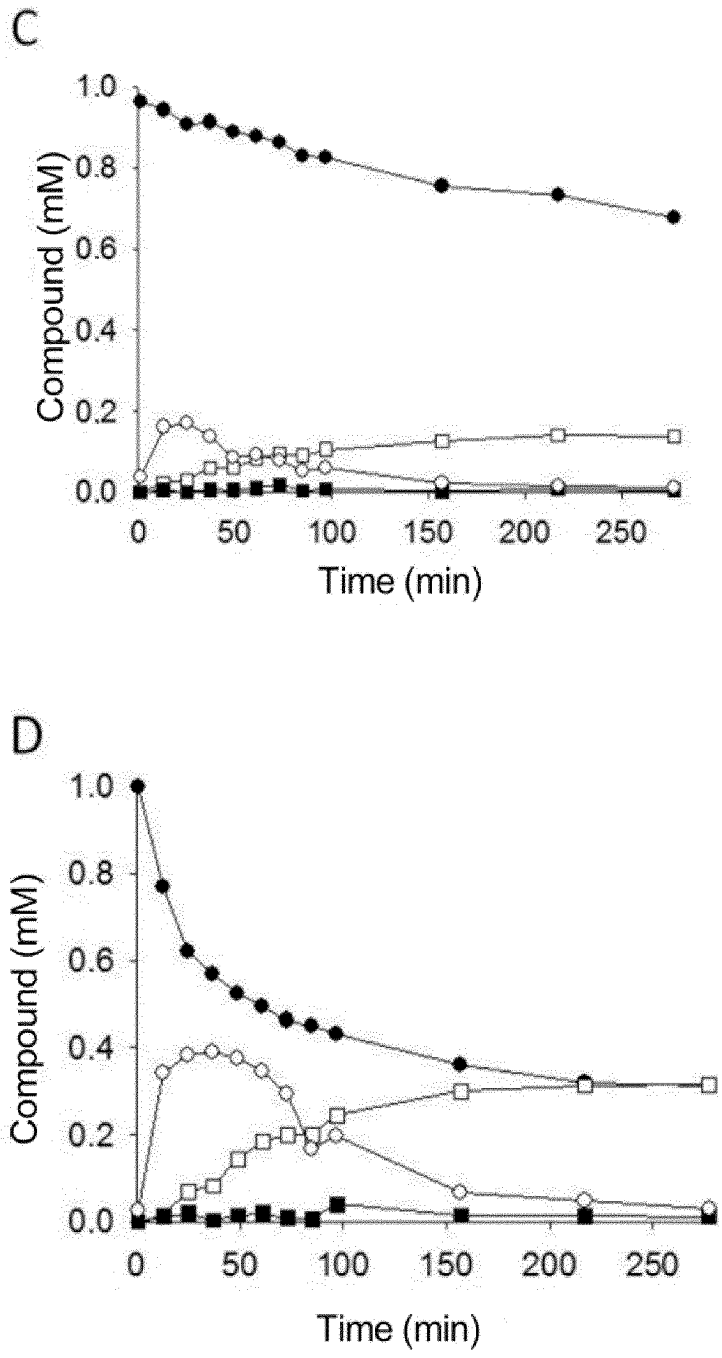

C
1.0
0.8
0.6
0.4
0.2
0.0
Compound (mM)
0
50
100
150
200
250
Time (min)
D
1.0
0.8
0.6
0.4
0.2
0.0
Compound (mM)
0
50
100
150
200
250
Time (min)

FIG. 4

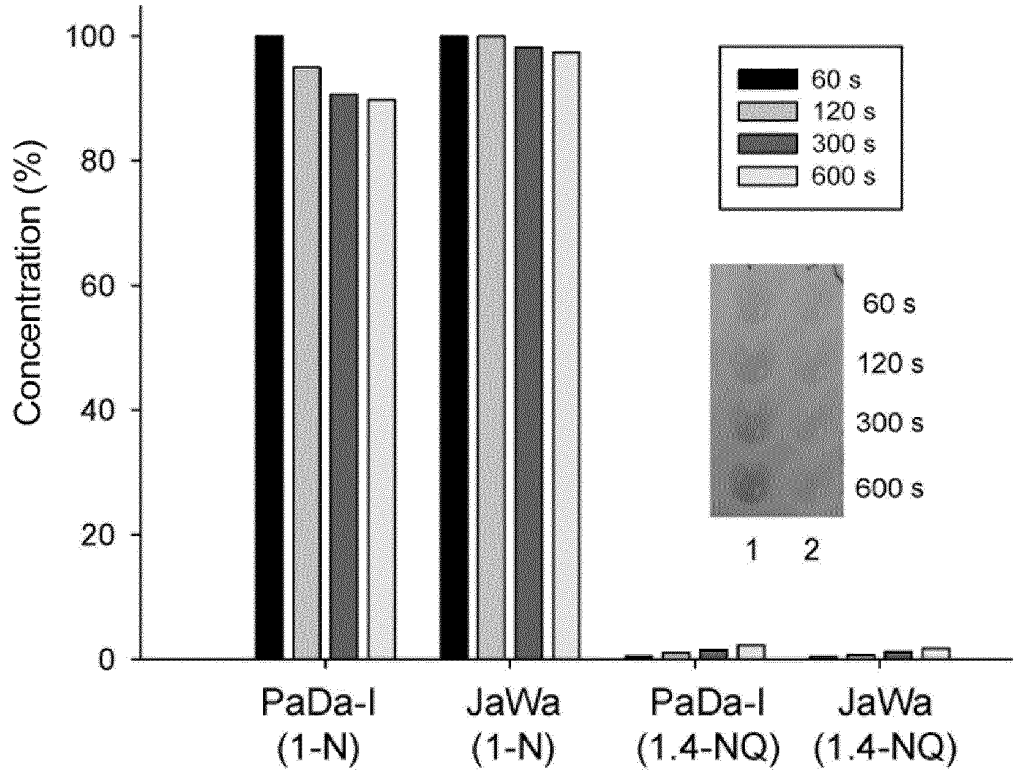

100
80
60
40
20
0
Concentration (%)
60 s
120 s
300 s
600 s
60 s
120 s
300 s
600 s
1
2
PaDa-I (1-N)
JaWa (1-N)
PaDa-I (1.4-NQ)
JaWa (1.4-NQ)

FIG. 5

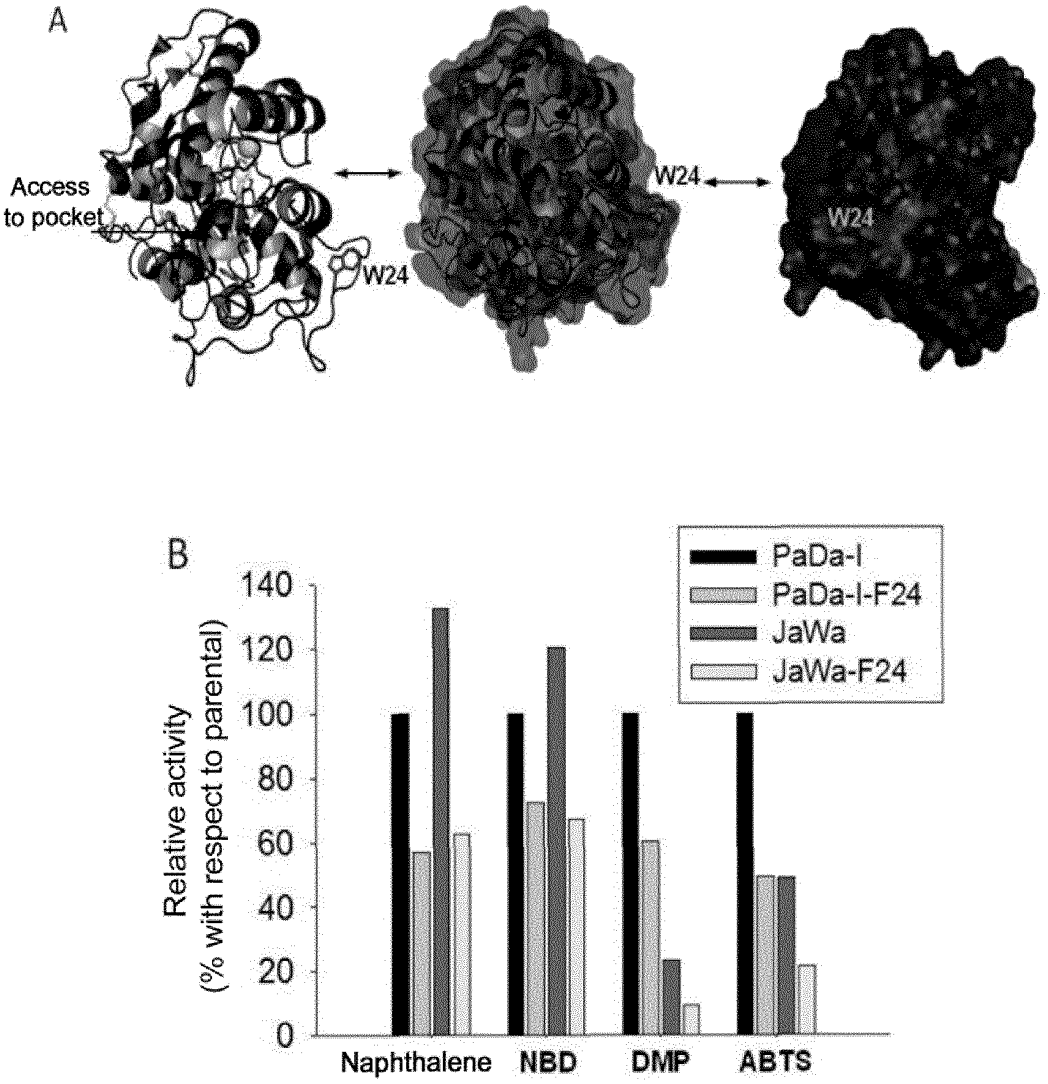

A
Access
to pocket
W24
W24
W24
B
Relative activity
(% with respect to parental)
140
120
100
80
60
40
20
0
PaDa-I
PaDa-I-F24
JaWa
JaWa-F24
Naphthalene
NBD
DMP
ABTS

FIG. 6

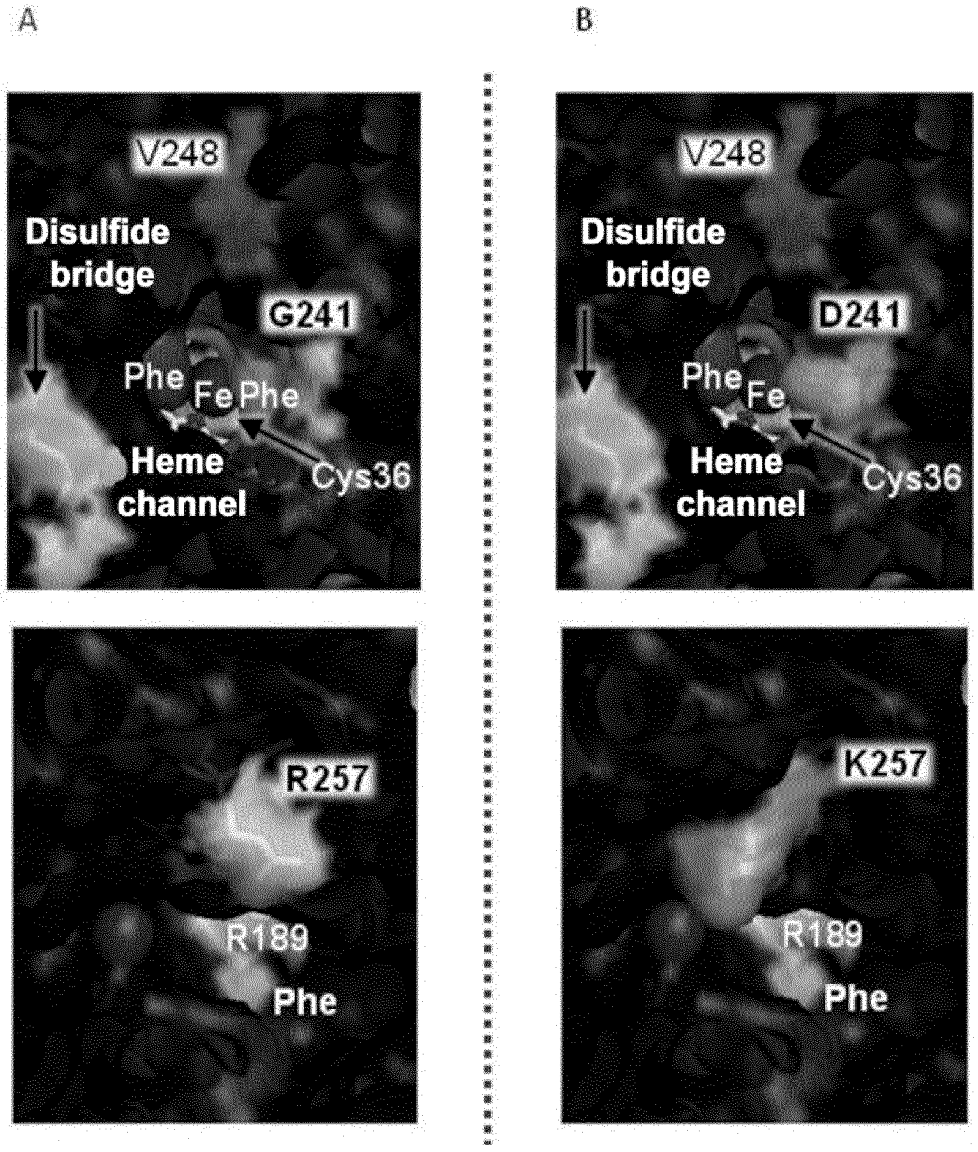

A
B
V248
Disulfide bridge
G241
Phe
Fe
Phe
Heme channel
Cys36
R257
R189
Phe
V248
Disulfide bridge
D241
Phe
Fe
Heme channel
Cys36
K257
R189
Phe

FIG. 7

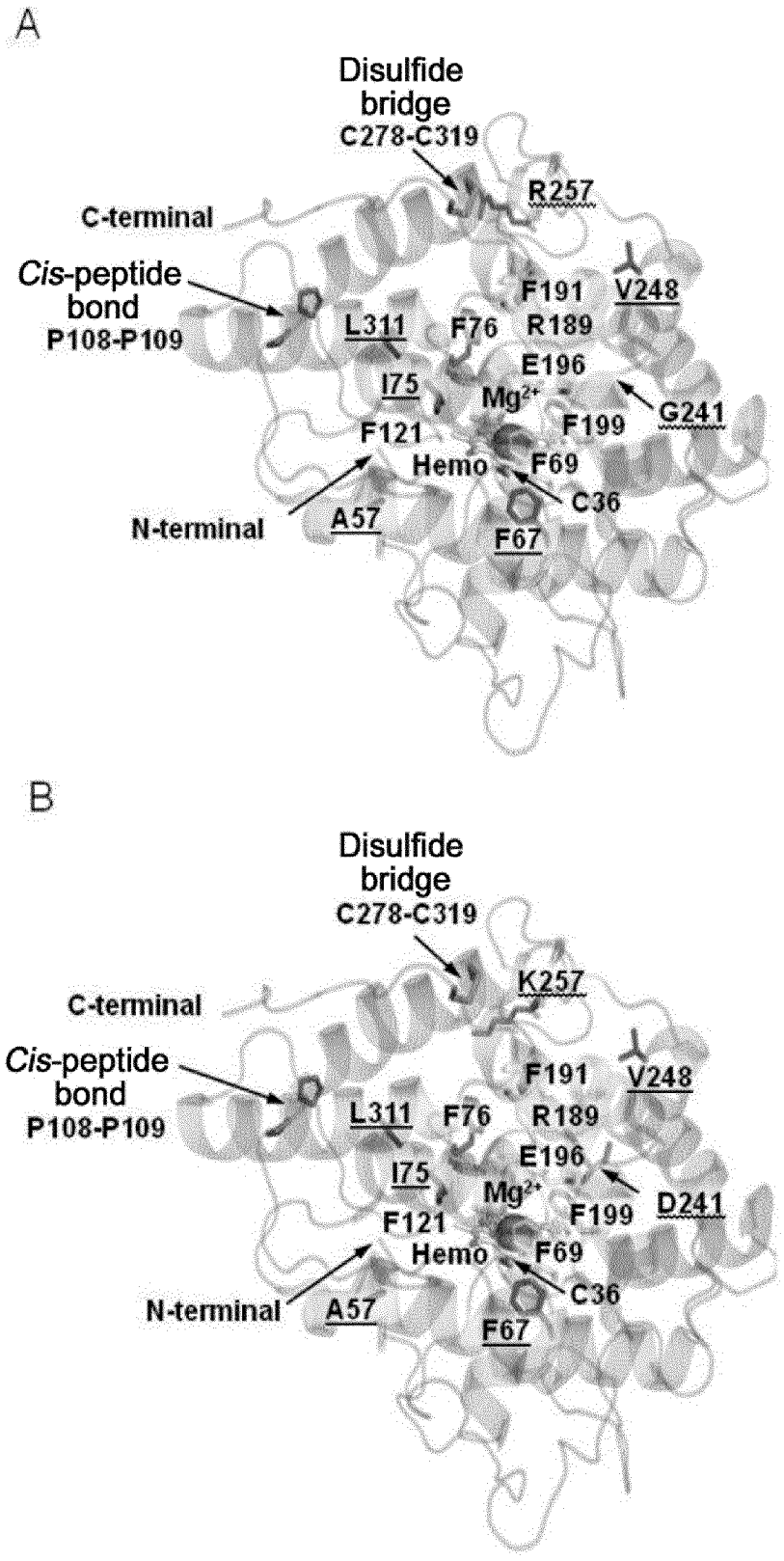

A
Disulfide bridge C278-C319
R257
C-terminal
Cis-peptide bond P108-P109
F191
V248
L311
F76
R189
E196
I75
Mg2+
G241
F121
F199
Hemo
F69
C36
N-terminal
A57
F67
B
Disulfide bridge C278-C319
K257
C-terminal
Cis-peptide bond P108-P109
F191
V248
L311
F76
R189
E196
I75
Mg2+
D241
F121
F199
Hemo
F69
C36
N-terminal
A57
F67

FIG. 8

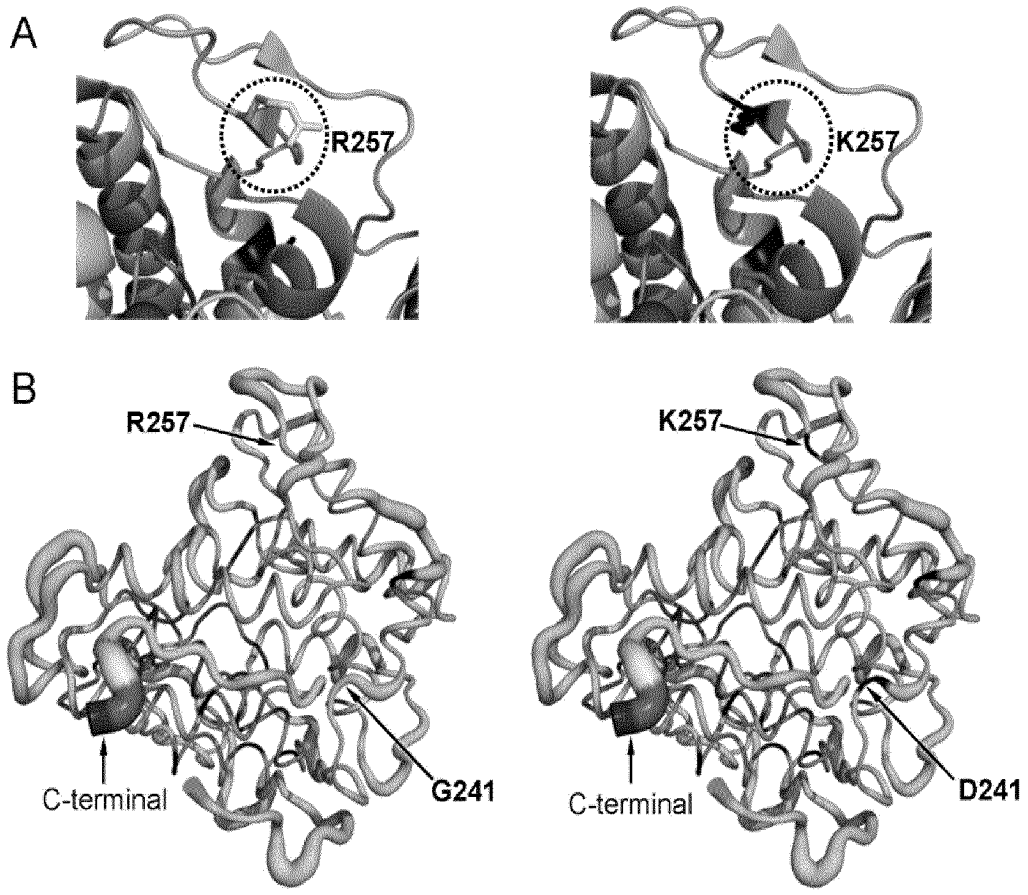


FIG. 9

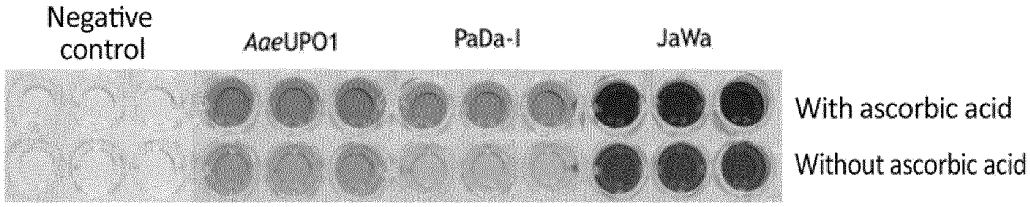

Negative control
AaeUPO1
PaDa-I
JaWa
With ascorbic acid
Without ascorbic acid

FIG. 10

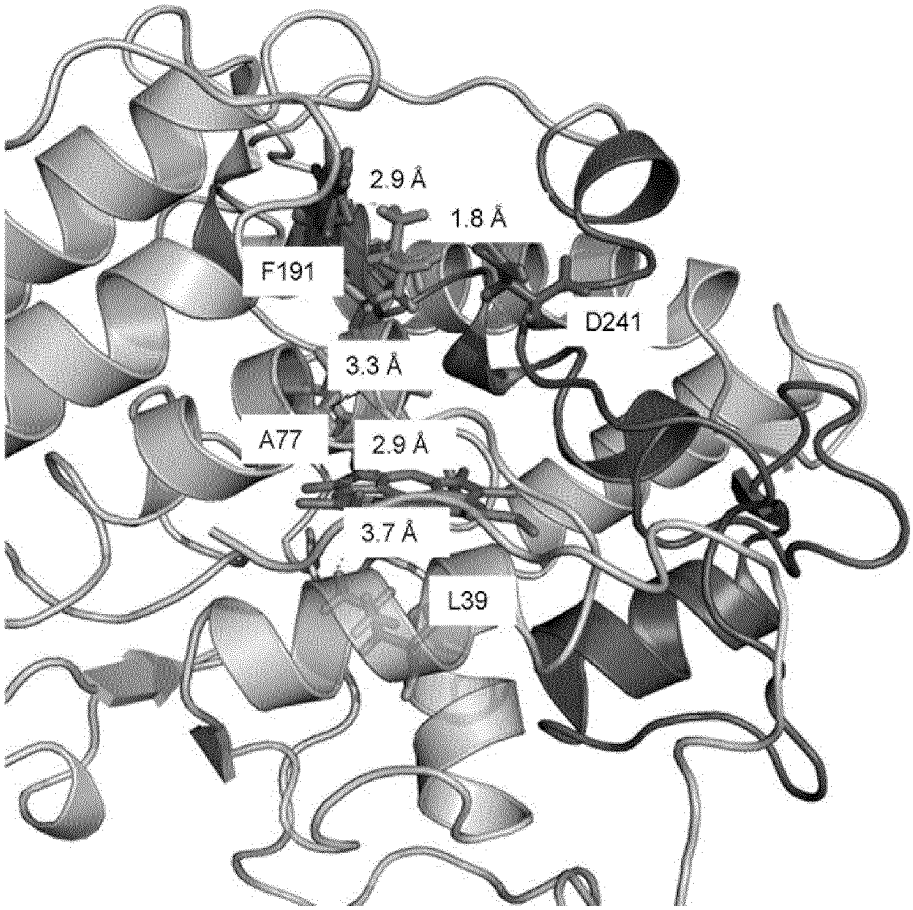

2.9 Å
1.8 Å
F191
D241
3.3 Å
A77
2.9 Å
3.7 Å
L39

FIG. 11

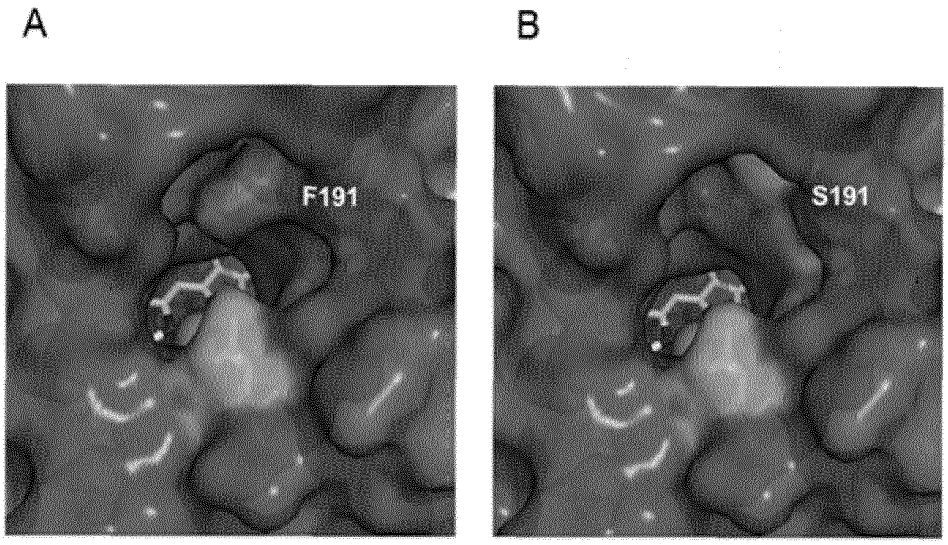

A
B
F191
S191

FIG. 12

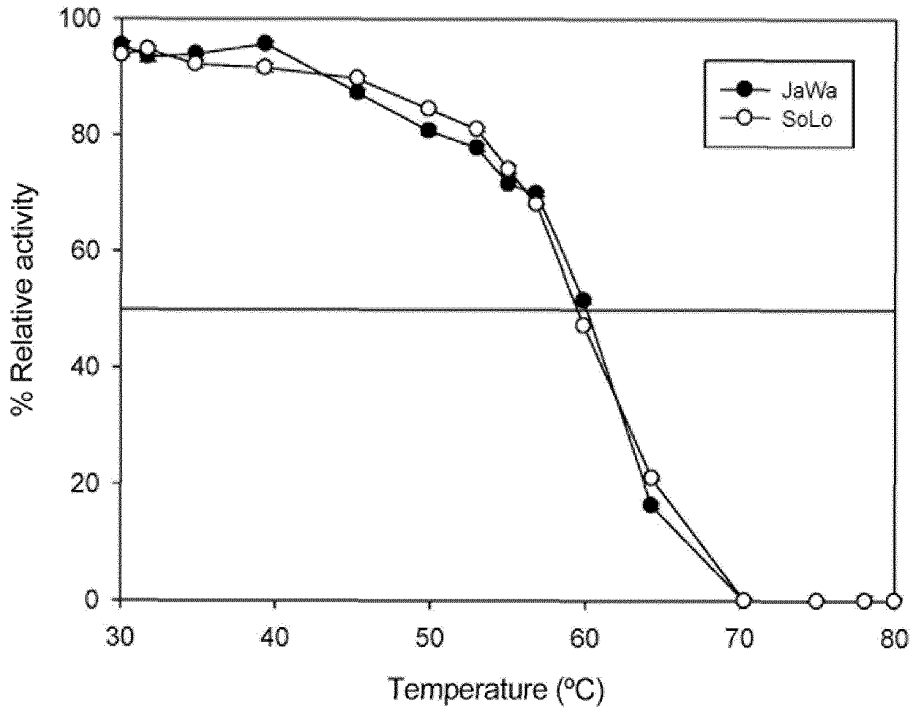

100
80
60
40
20
0
% Relative activity
30
40
50
60
70
80
Temperature (°C)
JaWa
SoLo

FIG. 13

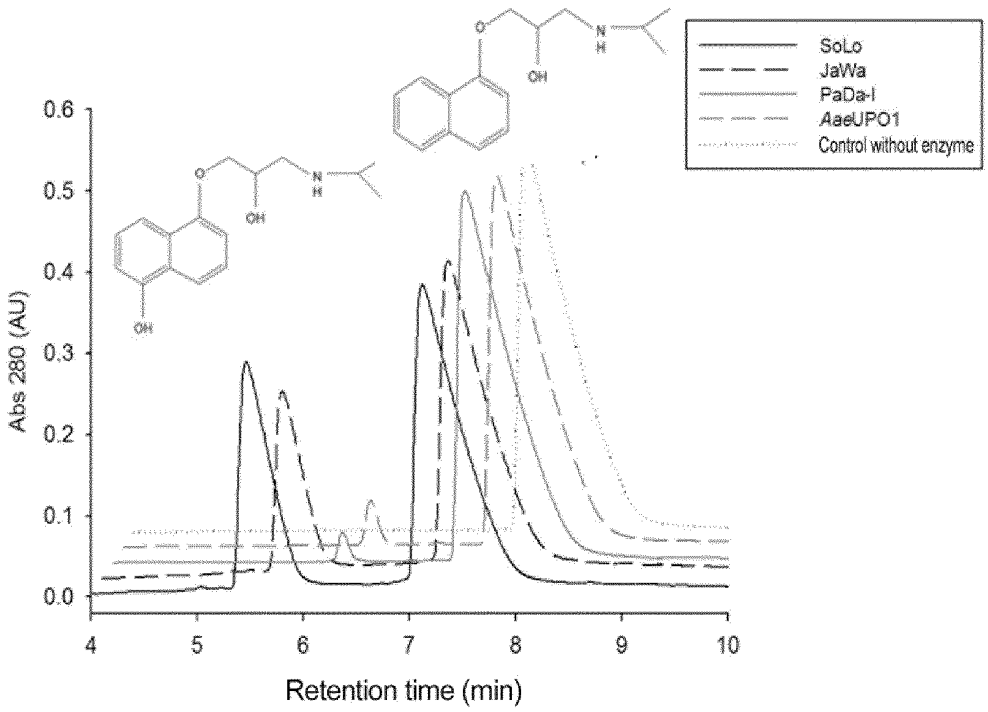

SoLo
JaWa
PaDa-I
AaeUPO1
Control without enzyme
Abs 280 (AU)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
4
5
6
7
8
9
10
Retention time (min)

FIG. 14

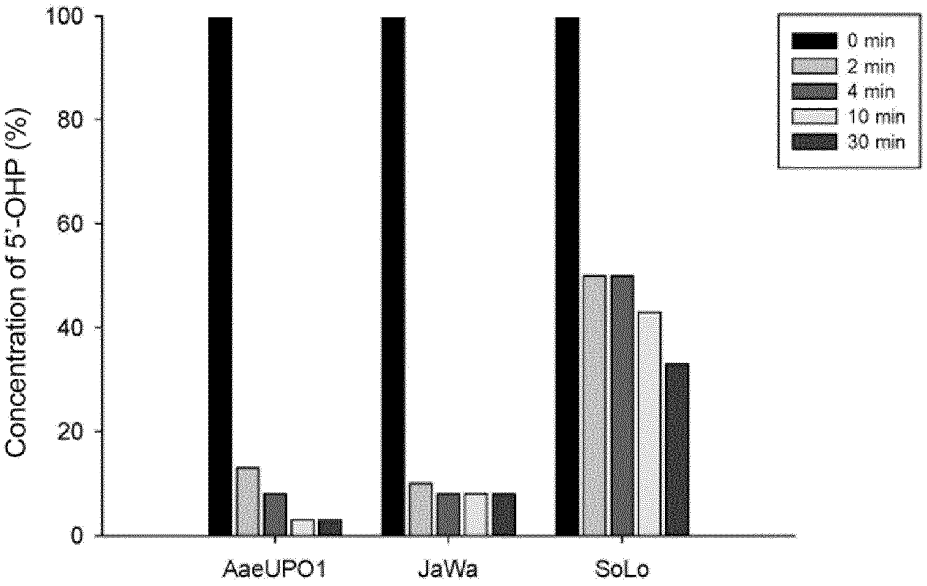
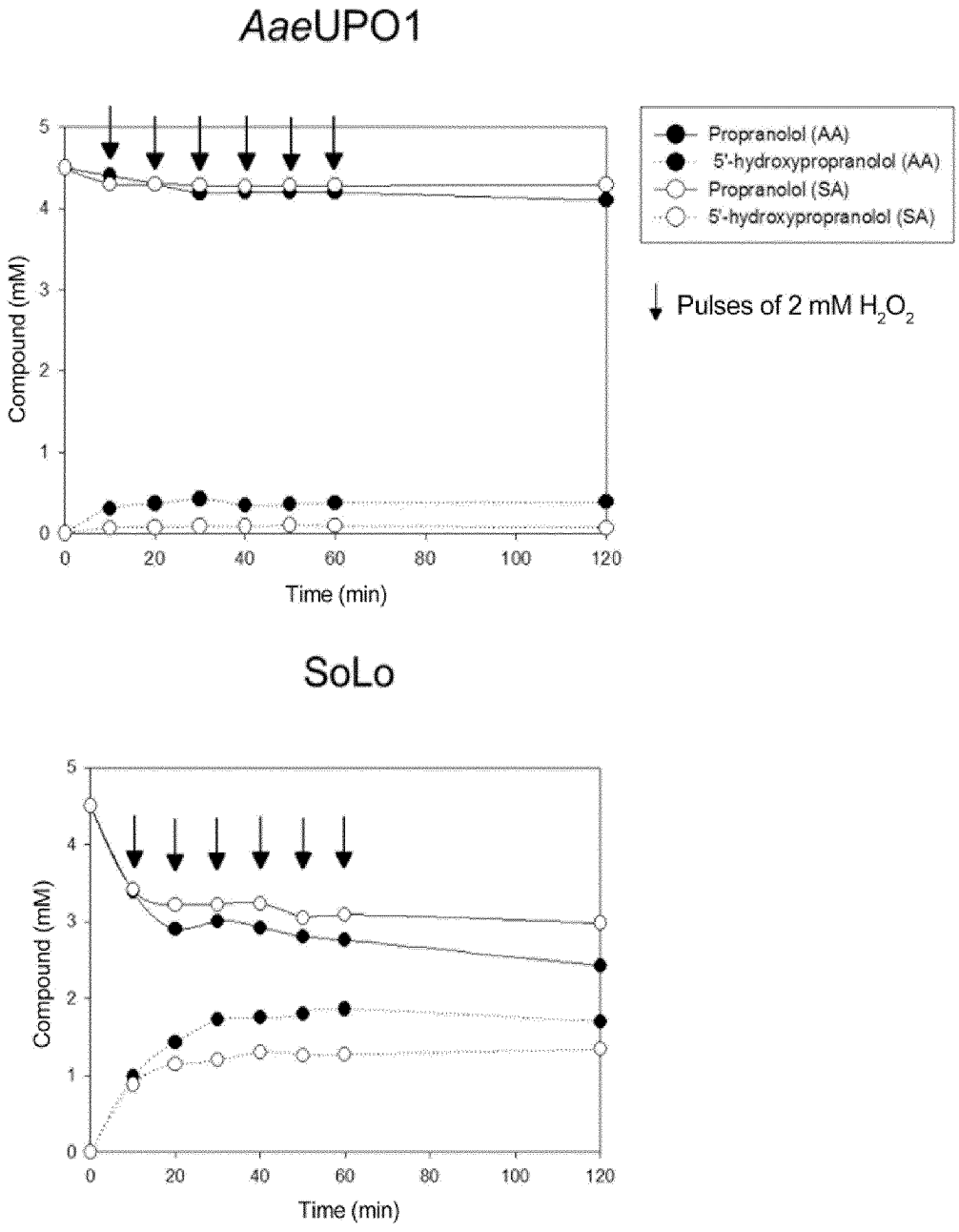

0 min
2 min
4 min
10 min
30 min
Concentration of 5'-OHP (%)
0
20
40
60
80
100
AaeUPO1
JaWa
SoLo

FIG. 15


AaeUPO1
Compound (mM)
Time (min)
Propranolol (AA)
5'-hydroxypropranolol (AA)
Propranolol (SA)
5'-hydroxypropranolol (SA)
Pulses of 2 mM H2O2
SoLo
Compound (mM)
Time (min)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/ES2016/070809

A. CLASSIFICATION OF SUBJECT MATTER

C12N9/08 (2006.01)

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 data base consulted during the international search (name of data base and, where practicable, search terms used)

EPODOC, INVENES, REGISTRY, BIOSIS, INTERNET

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MOLINA-ESPEJA, P. et al. "Directed evolution of unspecific peroxygenase from *Agrocybe aegerita*". APPLIED AND ENVIRONMENTAL MICROBIOLOGY. June 2014, Vol. 80, Nº 11, pages 3496-3507, doi:10.1128/AEM.00490-14, the whole document.	1-29
Y	WO 2015079064 A2 (NOVOZYMES A/S) 04.06.2015, claims.	1-29
Y	WO 0172999 A1 (CALIFORNIA INSTITUTE OF TECHNOLOGY) 04.10.2001, examples and claims.	1-29
A	PECYNA, M.J. et al. "Molecular characterization of aromatic peroxygenase from *Agrocybe aegerita*". APPL. MICROBIOL. BIOTECHNOL. 12.05.2009. Vol. 84, Nº 5, pages 885-897, doi:10.1007/s00253-009-2000-1, the whole document.	1-29

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance.
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure use, exhibition, or other means.
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other documents , such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07/03/2017	**(09/03/2017)**
Name and mailing address of the ISA/ OFICINA ESPAÑOLA DE PATENTES Y MARCAS Paseo de la Castellana, 75 - 28071 Madrid (España) Facsimile No.: 91 349 53 04	Authorized officer M. Novoa Sanjurjo Telephone No. 91 3498466

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/ES2016/070809

C (continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of documents, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	MOLINA-ESPEJA, P. et al. "Synthesis of 1-Naphtol by a natural peroxygenase engineered by direct evolution". CHEM. BIO. CHEM. 21.01.2016. Vol. 17, N° 4, pages 341-349, doi:10.1002/cbic.201500493, the whole document.	1-29

Form PCT/ISA/210 (continuation of second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/ES2016/070809

Patent document cited in the search report	Publication date	Patent family member(s)	Publication date
WO2015079064 A2	04.06.2015	US2016376566 A1	29.12.2016
		CN105793418 A	20.07.2016
		EP3074509 A2	05.10.2016
WO0172999 A1	04.10.2001	US2003153042 A1	14.08.2003
		AU7576300 A	08.10.2001

Form PCT/ISA/210 (patent family annex) (January 2015)

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **H. J. ZHANGLIN ; F.H. ARNOLD.** *Nature,* 1999, vol. 399, 670-673 **[0005]**
- **P. C. CIRINO ; F. H. ARNOLD.** *Angew. Chem. Int. Ed.,* 2003, vol. 42, 3299-3301 **[0005]**
- **P. MEINHOLD et al.** *Adv. Synth. Catal.,* 2006, vol. 348, 763-772 **[0005]**
- **P. S. COELHO et al.** *Science,* 2013, vol. 339, 307-310 **[0005]**
- **K. A. CANADA et al.** *J. Bacteriol.,* 2002, vol. 184, 344-349 **[0005]**
- **L. RUI et al.** *Appl. Environ. Microbiol.,* 2004, vol. 70, 3246-3252 **[0005]**
- **J. GARIKIPATI et al.** *Appl. Environ. Microbiol.,* 2009, vol. 75, 6545-6552 **[0005]**
- **M. KLUGE et al.** *Appl. Microbiol. Biotechnol.,* 2009, vol. 81, 1071-1076 **[0007] [0153]**
- **P. MOLINA-ESPEJA et al.** *Appl. Environ. Microbiol.,* 2014, vol. 80, 3496-3507 **[0010] [0137] [0178]**
- **P. MOLINA-ESPEJA et al.** *Appl. Environ. Microbiol.,* 2014, vol. 80, 3496-507 **[0010]**
- **M.A. ROMANOS et al.** *Yeast,* 1992, vol. 8, 423-488 **[0103]**
- **J. R. SHUSTER.** *Curr. Opin. Biotechnol.,* 1991, vol. 2, 685-690 **[0103]**
- **J. A. RAKESTRAW et al.** *Biotechnol. Bioeng.,* 2009, vol. 103, 1192-1201 **[0103]**
- **M. J. PECYNA et al.** *Appl. Microbiol. Biotechnol.,* 2009, vol. 84, 885-897 **[0128]**
- **P. MOLINA-ESPEJA et al.** *Appl. Environ. Microbiol.,* 2014, vol. 80, 3496, , 3507 **[0130]**
- **H. ZHAO et al.** *Nat Biotechnol.,* 1998, vol. 16, 258-261 **[0132]**
- **E. GARCIA-RUIZ et al.** *Biochem. J.,* 2012, vol. 441, 487-498 **[0132]**
- **M. ALCALDE.** *Methods Mol. Biol.,* 2010, vol. 634 (3), 14 **[0134]**
- **M. G. KLUGE et al.** *Appl. Microbiol. Biotechnol.,* 2007, vol. 75, 1473-1478 **[0143]**
- **D. GONZÁLEZ-PEREZ et al.** *PLoS ONE,* 2014, vol. 9, 90919 **[0168]**
- **S. KILLE et al.** *ACS Synth. Biol.,* 2013, vol. 2, 83-92 **[0169]**
- **C. R. OTEY ; J. M. JOERN.** *Methods Mol. Biol.,* 2003, vol. 230, 141-8 **[0178]**