

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2019/0024058 A1 MOLINA ESPEJA et al.

Jan. 24, 2019 (43) Pub. Date:

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

(71) Applicant: CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS (CSIC), Madrid (ES)

(72) Inventors: Patricia MOLINA ESPEJA, Madrid (ES); Francisco José PLOU GASCA, Madrid (ES); Miguel ALCALDE GALEOTE, Madrid (ES); Patricia GÓMEZ DE SANTOS, Madrid (ES)

(73) Assignee: CONSEJO SUPERIOR INVESTIGACIONES CIERNTIFICAS (CSIC), Madrid (ES)

15/775,369 Appl. No.:

PCT Filed: Nov. 14, 2016

(86) PCT No.: PCT/ES2016/070809

§ 371 (c)(1),

(2) Date: Sep. 11, 2018

(30)Foreign Application Priority Data

(ES) P201531641 Nov. 13, 2015

Publication Classification

(51) **Int. Cl.** C12N 9/08 (2006.01)C12N 15/80 (2006.01)

(52) U.S. Cl. CPC C12N 9/0065 (2013.01); C12R 1/685 (2013.01); C12N 15/80 (2013.01)

(57)ABSTRACT

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.

Specification includes a Sequence Listing.

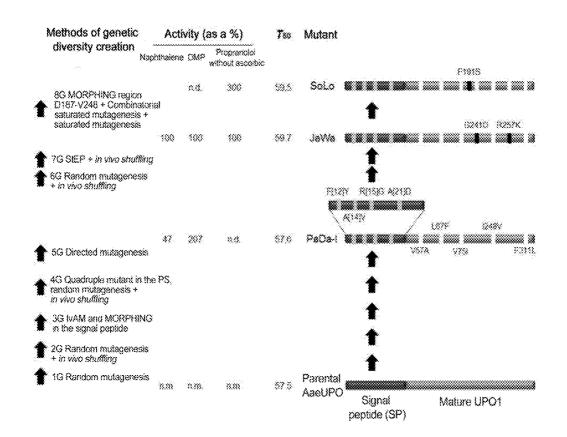
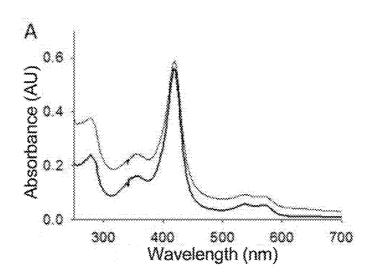
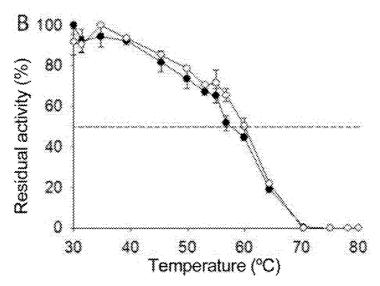


FIG. 1 33.55 * **#**287× 880 Mature UPO1 -• 02440 83883 ** Rispo A210 peptide (SP) Signal PaCad . *** #2## #2## AaeUPO 000 Parental 38778 Mutant 2 8 88 8 Naphthalene CMP Propramo 8 8 8 Activity (as a %) C) 8 ## ## 8 : :: 8 83.88 Č. 4G Quadruple mutant in the PS, random mutagenesis + in vivo shuffling 8G MORPHING region D187-V248 + Combinatorial saturated mutagenesis + saturated mutagenesis 7G StEP + in vivo shuffing 3G IVAM and MORPHING in the signal peptide 2G Random mutagenesis + in vivo shuffing 1G Random mutagenesis 6G Random mutagenesis + in vivo shuffling 5G Directed mutagenesis Methods of genetic diversity creation

FIG. 2



Jan. 24, 2019 Sheet 2 of 13



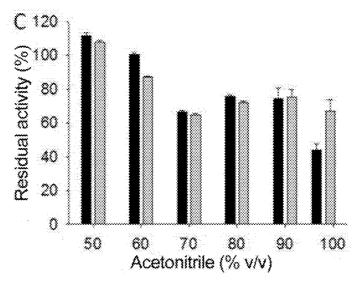
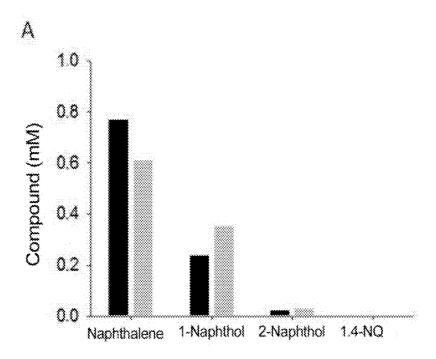


FIG. 3



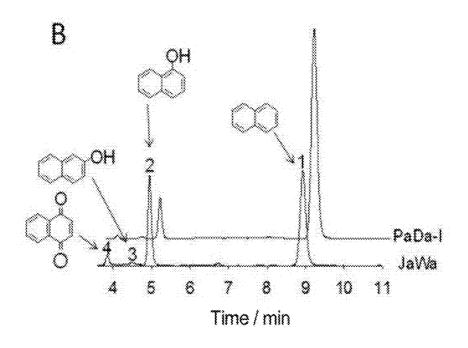
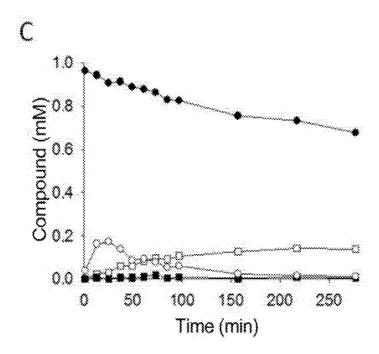


FIG. 3 (cont.)



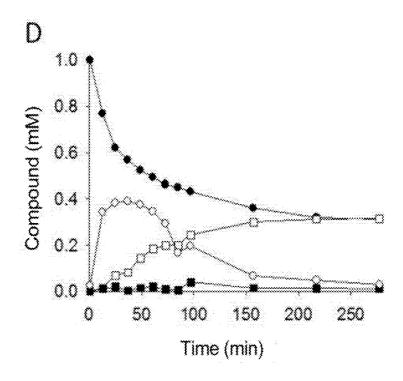


FIG. 4

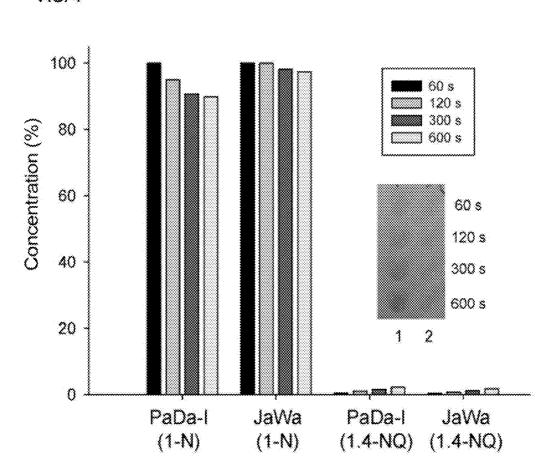
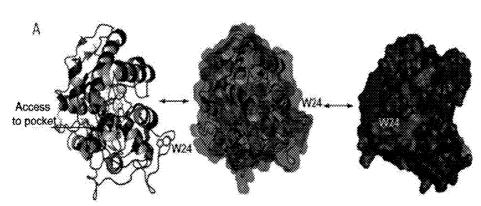


FIG. 5



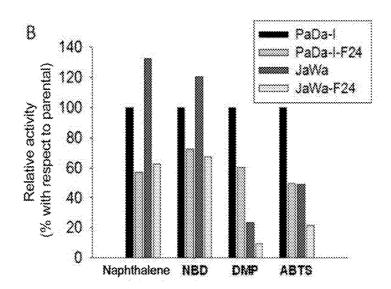
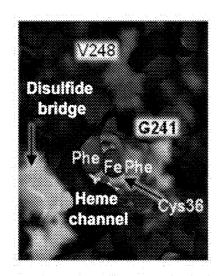
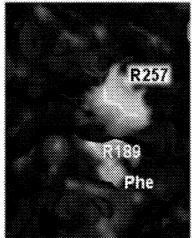


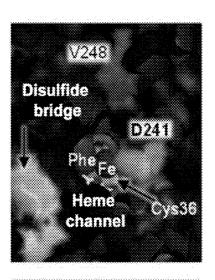
FIG. 6

A





8



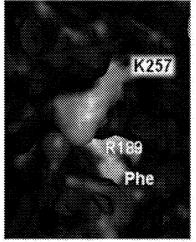
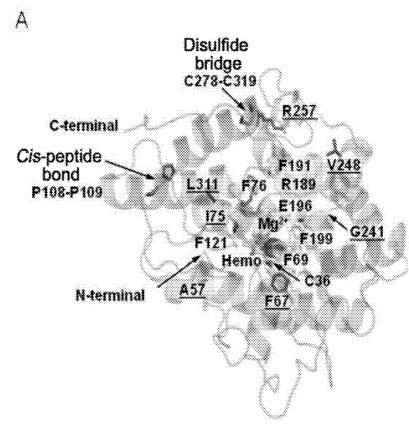


FIG. 7



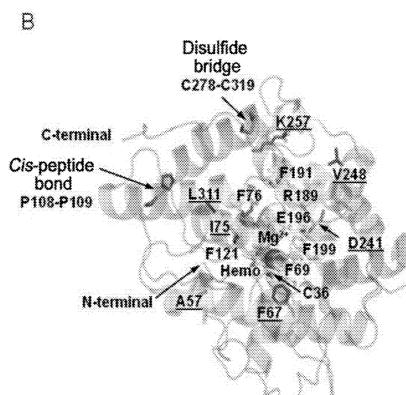


FIG. 8

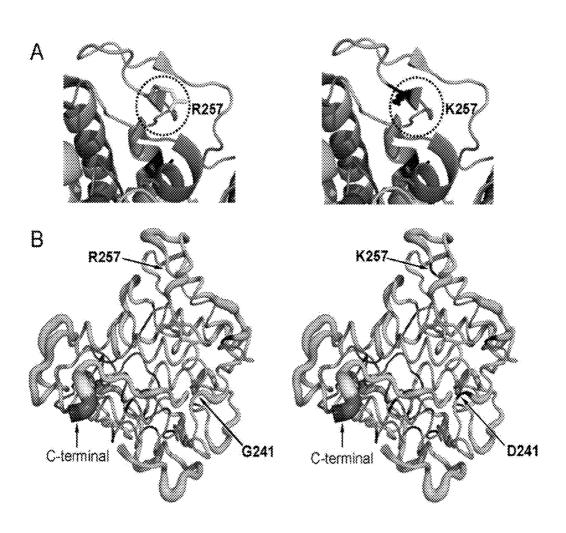


FIG. 9

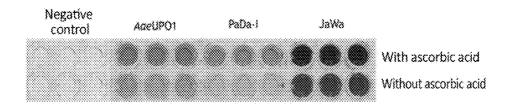


FIG. 10

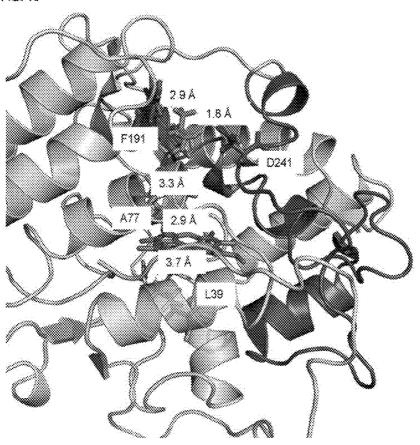


FIG. 11

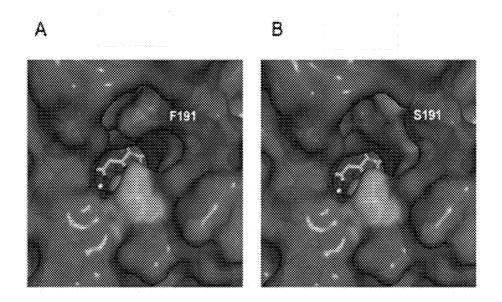
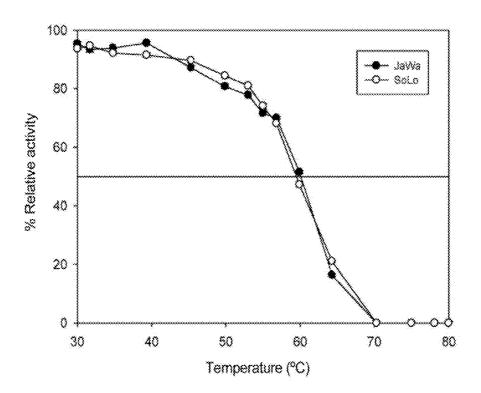


FIG. 12



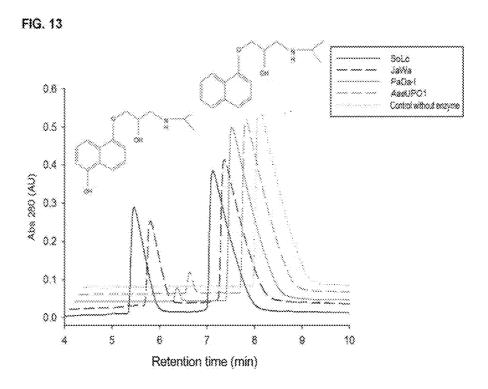


FIG. 14

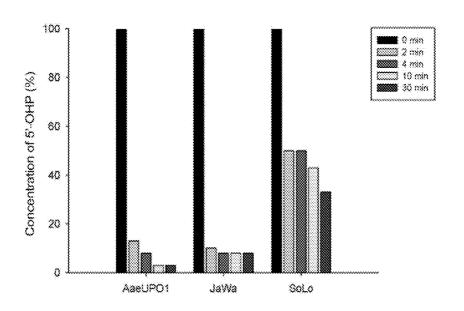
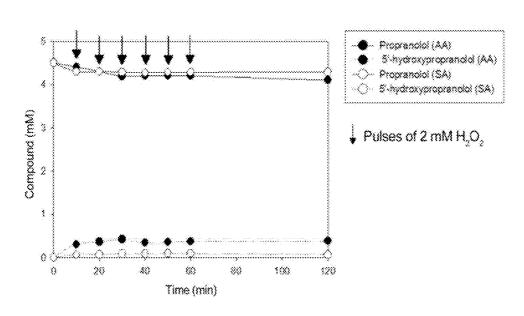
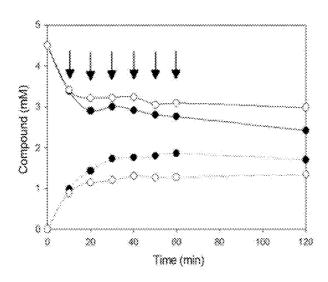


FIG. 15

AaeUPO1



SoLo



MUTANTS OF UNSPECIFIC PEROXYGENASE WITH HIGH MONOOXYGENASE ACTIVITY AND USES THEREOF

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 carbonhydrogen 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

[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)Hdependent 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 (NA-DPH) 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 betablocker 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 (AaeUPO) and Coprinellus radians (CraUPO). 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 inter-

[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 selfsufficient. With minimal requirements, just catalytic concentrations of H₂O₂ (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 selfsufficient, 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 (AaeUPO 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* (AaeUPO 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 AaeUPO1 of SEQ ID NO: 2, encoded by the sequence SEQ ID NO: 1. Similarly, the nucleotide sequence that encodes the native signal peptide of AaeUPO1 (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 AaeUPO1 (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 AaeUPO1 (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 T₅₀=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_{ca}/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:

[0014] i) they show a high production rate,

[0015] ii) they show high activity,

[0016] iii) they show high stability,

[0017] 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'-hydroxy-propranolol,

[0018] 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,

[0019] vi) shows enhanced catalytic efficiency for naphthalene up to values of $6.2 \times 10^5 \text{ s}^{-1} \text{ M}^{-1}$; and for 5'-hydroxypropranolol of $3.1 \times 10^6 \text{ s}^{-1} \text{ M}^{-1}$, two orders of magnitude higher than those of any enzyme described,

[0020] 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,

[0021] viii) They have a regioselectivity against 1-naphthol of approximately 97% and of approximately 99% against 5'-hydroxypropranolol.

[0022] 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.

[0023] 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:

[0024] SEQ ID NO: 1—Nucleotide sequence of the gene that encodes AaeUPO1 without signal peptide.

[0025] SEQ ID NO: 2—Polypeptide sequence of AaeUPO1 without signal peptide.

[0026] SEQ ID NO: 3—Nucleotide sequence of the gene that encodes AaeUPO1 with wild-type signal peptide.

[0027] SEQ ID NO: 4—Polypeptide sequence of AaeUPO1 with wild-type signal peptide.

[0028] SEQ ID NO: 5—Nucleotide sequence of the gene that encodes AaeUPO1 with modified signal peptide.

[0029] SEQ ID NO: 6—Polypeptide sequence AaeUPO1 with modified signal peptide.

[0030] SEQ ID NO: 7—Nucleotide sequence that encodes the wt-JaWa variant without signal peptide.

[0031] SEQ ID NO: 8—Polypeptide sequence of the wt-JaWa variant without signal peptide.

[0032] SEQ ID NO: 9—Nucleotide sequence that encodes the wt-JaWa variant with wild-type signal peptide.

[0033] SEQ ID NO: 10—Polypeptide sequence of the wt-JaWa variant with wild-type signal peptide.

[0034] SEQ ID NO: 11—Nucleotide sequence that encodes the wt-JaWa variant with modified signal peptide.

[0035] SEQ ID NO: 12—Polypeptide sequence of the wt-JaWa variant with modified signal peptide.

[0036] SEQ ID NO: 13—Nucleotide sequence that encodes the PaDa-I variant without signal peptide.

[0037] SEQ ID NO: 14—Polypeptide sequence of the PaDa-I variant without signal peptide.

[0038] SEQ ID NO: 15—Nucleotide sequence that encodes the PaDa-I variant with wild-type signal peptide.

[0039] SEQ ID NO: 16—Polypeptide sequence of the PaDa-I variant with wild-type signal peptide.

[0040] SEQ ID NO: 17—Nucleotide sequence that encodes the PaDa-I variant with modified signal peptide

[0041] SEQ ID NO: 18—Polypeptide sequence of the PaDa-I variant with modified signal peptide.

[0042] SEQ ID NO: 19—Nucleotide sequence that encodes the JaWa variant without signal peptide.

[0043] SEQ ID NO: 20—Polypeptide sequence of the JaWa variant without signal peptide.

[0044] SEQ ID NO: 21—Nucleotide sequence that encodes the JaWa variant with wild-type signal peptide.

[0045] SEQ ID NO: 22—Polypeptide sequence of the JaWa variant with wild-type signal peptide.

[0046] SEQ ID NO: 23—Nucleotide sequence that encodes the JaWa variant with modified peptide.

[0047] SEQ ID NO: 24—Polypeptide sequence of the JaWa variant with modified signal peptide.

[0048] SEQ ID NO: 25—Nucleotide sequence that encodes the native signal peptide of AaeUPO1.

[0049] SEQ ID NO: 26—Polypeptide sequence of the native signal peptide of AaeU P01

- [0050] 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 AaeUPO1 of SEQ ID NO: 26.
- [0051] 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.
- [0052] SEQ ID NO: 29—Nucleotide sequence that encodes the W24F variant obtained from the PaDa-I mutant of SEQ ID NO: 17.
- [0053] SEQ ID NO: 30—Polypeptide sequence that encodes the W24F variant obtained from the PaDa-I mutant of SEQ ID NO: 18.
- [0054] SEQ ID NO: 31—Nucleotide sequence that encodes the W24F variant obtained from the JaWa mutant of SEQ ID NO: 23.
- [0055] SEQ ID NO: 32—Polypeptide sequence that encodes the W24F variant obtained from the JaWa mutant of SEQ ID NO: 24.
- [0056] SEQ ID NO: 37—Nucleotide sequence that encodes the SoLo variant without signal peptide.
- [0057] SEQ ID NO: 38—Polypeptide sequence of the SoLo without signal peptide.
- [0058] SEQ ID NO: 39—Nucleotide sequence that encodes the SoLo variant with wild-type signal peptide.
- [0059] SEQ ID NO: 40—Polypeptide sequence of the SoLo variant with wild-type signal peptide.
- [0060] SEQ ID NO: 41—Nucleotide sequence that encodes the SoLo variant with modified signal peptide.
- [0061] SEQ ID NO: 42—Polypeptide sequence of the SoLo variant with modified signal peptide.
- [0062] SEQ ID NO: 61—Nucleotide sequence that encodes the wt-SoLo variant without signal peptide.
- [0063] SEQ ID NO: 62—Polypeptide sequence of the wt-SoLo variant without signal peptide.
- [0064] SEQ ID NO: 63—Nucleotide sequence that encodes the wt-SoLo variant with wild-type signal peptide.
- [0065] SEQ ID NO: 64—Polypeptide sequence of the wt-SoLo variant with wild-type signal peptide.
- [0066] SEQ ID NO: 65—Nucleotide sequence that encodes the wt-SoLo variant with modified signal peptide.
- [0067] SEQ ID NO: 66—Polypeptide sequence of the wt-SoLo variant with modified signal peptide.
- [0068] 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.
- [0069] 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 naphtha-

- lene of k_{ca}/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:
 - [0070] i) Functional heterologous expression in yeast (0.2 g/L),
 - [0071] ii) increase in catalytic constants and efficiencies.
 - [0072] iii) increased stability against various factors (temperature, co-solvents).
- [0073] 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:
 - [0074] i) it shows a high production rate,
 - [0075] ii) it shows high activity,
 - [0076] iii) it shows high stability,
 - [0077] 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),
 - [0078] 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'-hydroxy-propranolol of up to 3.6 fold,
 - [0079] vi) it shows enhanced catalytic efficiency for naphthalene up to values of $6.2 \times 10^5 \text{ s}^{-1} \text{M}^{-1}$; and for 5'-hydroxypropranolol of $3'1 \times 10^6 \text{ s}^{-1} \text{ M}^{-1}$, two orders of magnitude higher than those of any enzyme described.
 - [0080] 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,
 - [0081] viii) it shows regioselectivity against 1-naphthol of approximately 97% and of approximately 99% against 5'-hydroxypropranolol.
- [0082] 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 H₂O₂ 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.
- [0083] 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.
- **[0084]** 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.

[0085] 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. [0086] 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.

[0087] 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 (AaeUPO1), 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).

[0088] 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 (AaeUPO1), 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).

[0089] 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 (AaeUPO1), 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).

[0090] 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.

tion is the coding sequence of an AaeUPO1 peroxygenase variant with the described enhanced activity, whose nucleotide sequence corresponds to:

[0091] a) nucleic acid molecules of the isolated polynucleotide sequence or in its complementary strand,

[0092] b) nucleic acid molecules whose complementary strand is capable of hybridising in astringent conditions with a polynucleotide sequence of (a), or

[0093] c) nucleic acid molecules, whose sequence differs from (a) and/or (b) due to the degeneration of the genetic code.

[0094] 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. [0095] The polynucleotides that encode the polypeptides of amino acid sequences described in the invention correspond to variants obtained by means of directed evolution of AaeUPO1 peroxygenase (E.C. 1.11.2.1). Said protein, AaeUPO1, 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. [0096] The term "polynucleotide", as used in the description, relates to polymeric forms of nucleotides of any length, both ribonucleotides and deoxyribonucleotides.

[0097] 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, Wis.), 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.

[0098] 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 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 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.

[0099] 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.

[0100] 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

[0101] 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.

[0102] 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.

[0103] 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.

[0104] 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.

[0105] 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.

[0106] 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).

[0107] 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.

[0108] 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:

[0109] 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),

[0110] 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 (1248V),

[0111] 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),

[0112] 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

[0113] 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).

[0114] 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.

[0115] 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:

[0116] 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),

[0117] 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),

[0118] 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

[0119] 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).

[0120] 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'-hydroxypropanol 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-1 variant.

[0121] 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).

[0122] 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).

[0123] 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).

[0124] 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).

[0125] 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.

[0126] 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). [0127] 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 AaeUPO1, 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 SEO 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).

[0128] 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.

[0129] 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).

[0130] 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).

[0131] 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.

[0132] 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:

- [0133] 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),
- [0134] 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),
- [0135] 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
- [0136] 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).

[0137] 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.

[0138] 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).

[0139] 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).

[0140] 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).

[0141] 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.

[0142] 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 (AaeUPO1, 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).

[0143] 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).

[0144] The term "peptide", "polypeptide" or "protein", as used in the description, relates to a polymeric form of amino acids of any length.

[0145] 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.

[0146] 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.

[0147] 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 $6.2 \times 10^5 \, \mathrm{s}^{-1} \, \mathrm{M}^{-1}$, a decrease of approximately 1.5 fold in the ratio 1.4naphthoquinone: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 3.1×10⁶ 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

[0148] 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.

[0149] 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.

[0150] 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).

[0151] 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:

- [0152] 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),
- [0153] 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),
- [0154] 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).
- [0155] 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
- [0156] 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).

[0157] 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.

[0158] 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:

- [0159] 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)
- [0160] 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),
- [0161] 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
- [0162] 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).

[0163] 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.

[0164] 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).

[0165] 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).

[0166] 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).

[0167] 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).

[0168] 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).

[0169] 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).

[0170] 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).

[0171] 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.

[0172] 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).

[0173] 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).

[0174] 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 G241 D and R257K mutations, respectively, obtaining the JaWa variant of SEQ ID NO: 20.

[0175] 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.

[0176] 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.

[0177] 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:

[0178] 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),

[0179] 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),

[0180] 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

[0181] 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).

[0182] 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.

[0183] 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).

[0184] 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).

[0185] 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).

[0186] 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.

[0187] 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.

[0188] 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.

[0189] 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, Wis.), pRSET or PrEP plasmids (Invitrogen, San Diego, Calif.), pMAL plasmids (New England Biolabs, Beverly, Mass.), 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.

[0190] 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.

[0191] 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.

[0192] 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.

[0193] 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.

[0194] 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.

[0195] 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.

[0196] 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).

[0197] Preferably, the signal peptide is that of the AaeUPO1 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:

- [0198] 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).
- [0199] 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).
- [0200] 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
- [0201] 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).

[0202] 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.

[0203] 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.

[0204] 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.

[0205] 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.

[0206] A host cell of the present invention includes prokaryotic and eukaryotic cells. Prokaryotes include gramnegative organisms (for example, Escherichia coli) or grampositive 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.

[0207] 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*.

[0208] 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*.

[0209] Another aspect described in the present invention relates to the method for obtaining the polypeptide of the invention, which comprises the following steps:

- [0210] a) Introducing the vector of the invention, as described earlier, in an appropriate host cell (host cell of the invention),
- [0211] b) culturing the host cell of the invention in an appropriate medium, and
- [0212] 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.
- [0213] 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.

[0214] Another aspect of the invention relates to a host cell culture of the invention.

[0215] 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.

[0216] 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.

[0217] 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

[0218] 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 AaeUPO1 peroxygenase expressed in the yeast, or with respect to UPO variants such as, for example, the PaDa-I variant.

[0219] 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.

[0220] 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).

[0221] 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.

[0222] 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.

[0223] 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.

[0224] 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

[0225] FIG. 1 Directed evolution of AaeUPO1. 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.

[0226] FIG. 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 $\rm H_2O_2$ and 0.3 mM ABTS). The error bars indicate standard deviations.

[0227] FIG. 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 H₂O₂ (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: 1.2naphthalene 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

[0228] FIG. 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 1 mM $\rm H_2O_2$ (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.

[0229] FIG. 5 W24F variants obtained by means of directed mutagenesis. A) Model built on the crystal structure of the AaeUPO1 enzyme (PDB access number: 2YOR), comprising the mutations of the JaWa variant as well as the W24F modification with respect to wild AaeUPO1. 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 H₂O₂ 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.

[0230] FIG. 6 Mutations in the UPO variants described in the invention. Model built on the structure of the AaeUPO1 crystal (PDB access number: 2YOR). A) PaDa-I; B) JaWa. The V248 mutant stems from the previous evolution path-

way. 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 $\mathrm{Fe^{3+}}$ is represented as a sphere.

[0231] FIG. 7 Protein model of A) PaDa-I and B) JaWa. The protein model for PaDa-I (A) was built on the structure of the AaeUPO1 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.

[0232] FIG. 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.

[0233] FIG. 9 Assay of 4-AAP (4-aminoantipyrine) with different pure UPO variants (AaeUPO1, PaDa-1 and JaWa). The reactions were performed at room temperature and their composition was as follows: 0.2 μ M of each pure UPO variant, 50 mM pH 7.0 potassium phosphate buffer, 5 mM propranolol, 2 mM H₂O₂ (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.

[0234] FIG. 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.

[0235] FIG. 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.

[0236] FIG. 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.

[0237] FIG. 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). [0238] FIG. 14 Turnover rates of AaeUPO, 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.

[0239] FIG. 15 Calculation of the total turnover number (TTN) of AaeUPO and SoLo. The assay was carried out

using 0.03 μ M 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

[0240] 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

[0241] 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, Oreg., USA), while the naphthalene is from Acros Organics (Geel, Belgium).

[0242] 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).

[0243] The competent *Escherichia coli* XL2-Blue cells and the Genemorph II Random Mutagenesis (Mutazyme II) kit were obtained from Agilent Technologies (Santa Clara, Calif., USA) and the iProof high-fidelity DNA polymerase was acquired from Bio-Rad (Hercules, Calif., USA). The BamHI and XhoI restriction enzymes were obtained from New England Biolabs (Ipswich, Mass., 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, Calif., 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

[0244] 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)

[0245] 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)

[0246] 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'-gggaggggtgaatgtaagc-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 AaeUPO1 or PaDa-I.

W24F Variants

[0247] 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'-ctcaccca

<u>tttaagccgcttcgacctggcgatattcgtggac-3'</u>) and the F24REV primer of sequence SEQ ID NO: 36 (5'-gtccacgaatatcgcca-ggtcgaagcggcttaaatgggtgag-3'). The changes made to said primer to perform the mutagenesis appear underlined in the nucleotide sequence thereof.

[0248] 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 μ L, 3% DMSO, 0.5 μ M 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

[0249] 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 MgSO₄, 31.6 mL of absolute ethanol, 1 mL of 25 g/L chloramphenicol and ddH₂O 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)

[0250] The mother plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 minutes at 3,500 RPM and 4° C. 20 μ L 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 μ L of reaction mixture were added with 2,6-dimethoxyphenol (DMP) or naphthalene to the daughter plates using a pipetting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, USA).

[0251] The DMP reaction mixture was composed of 100 mM pH 7.0 potassium phosphate buffer, 3 mM DMP and 1 mM $\rm H_2O_2$. 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 H₂O₂. 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 (T₅₀) (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

[0252] 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

[0253] 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 ddH₂O up to 1,000 mL) at 30° $\stackrel{\circ}{\text{C}}$. 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 ddH₂O 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

[0254] 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 ddH₂O 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 (OD₆₀₀=0.1). This culture of each clone was incubated for 72 hours at 25° C. and 220 RPM (at peak UPO activity; OD₆₀₀=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 H₂O₂). 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 T₅₀ 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 [0255] 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, OD₆₀₀ 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; OD₆₀₀=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 AaeUPO1 Variants

[0256] The purification of the recombinant AaeUPO variants described in the present invention was carried out by means of ion-exchange chromatography (ÄKTA 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 subjected 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

[0257] 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 H₂O₂; and for the rest of the substrates, in 100 mM pH 7.0 potassium phosphate buffer, 2 mM H₂O₂ (DMP) or 1 mM H₂O₂ (NBD and naphthalene, in 20% of acetonitrile—final concentration). For H₂O₂, 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_{cav} 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

[0258] 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×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% ddH₂O (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 H₂O₂ in 100 mM pH 7.0 potassium phosphate buffer (1 mL of final volume). The reaction started with the addition of H₂O₂ 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).

[0259] 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 $\rm H_2O_2$ in 100 mM pH 7.0 potassium phosphate buffer (0.2 mL of final volume). [0260] The standard deviations were less than 5% in all cases

Analysis Using MALDI-TOF-MS and Determination of the Isoelectric Point

[0261] 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 μ g 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)

[0262] 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

[0263] 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.

[0264] 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 (FIG. 1).

[0265] 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 |
| CT2 (nm) | 540 | 537 | 537 |

¹Estimated by SDS-PAGE;

[0266] As can be observed in Table 1 and in FIG. 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 T₅₀-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) (FIG. 2). [0267] 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 AaeUPO 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 (FIG. 3A).

[0268] 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 (FIG. 3B, C, D).

[0269] 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 $\rm H_2O_2$ 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.

[0270] 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

²estimated using MALDI-TOF;

³estimated according to the amino acid composition.

⁴Estimated in culture supernatants. n.d. not determined.

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.

[0271] 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 $\rm H_2O_2$, while 100 mM pH 7.0 potassium phosphate and 2 mM $\rm H_2O_2$ (DMP) or 1 mM (naphthalene or NBD, in 20% acetonitrile—final concentration) was used for the other buffers. For the $\rm H_2O_2$, 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) | $8.2 \times 10^6 \pm 6 \times 10^5$ | $6.9 \times 10^5 \pm 6.3 \times 10^4$ | |
| DMP | K _m (μM) | 126 ± 14 | 866 ± 108 | |
| | $k_{cat}(s^{-1})$ | 68 ± 2 | 142 ± 8 | |
| | k_{cat}/K_m $(s^{-1} M^{-1})$ | $5.4 \times 10^5 \pm 4.8 \times 10^4$ | $1.6 \times 10^5 \pm 1.2 \times 10^4$ | |
| Naphthalene | $K_m (\mu M)$ | 578 ± 106 | 127 ± 27 | |
| | $k_{cat}(s^{-1})$ | 229 ± 17 | 78 ± 3 | |
| | k_{cat}/K_m $(s^{-1} M^{-1})$ | $4 \times 10^5 \pm 4 \times 10^4$ | $6.2 \times 10^5 \pm 1.1 \times 10^5$ | |
| NBD | K _m (μM) | 483 ± 95 | 769 ± 80 | |
| | k_{cat} (s ⁻¹) | 338 ± 22 | 154 ± 8 | |
| | k_{cat}/K_m $(s^{-1} M^{-1})$ | | $2.0 \times 0^5 \pm 1.2 \times 10^4$ | |
| H_2O_2 | $K_m (\mu M)$ | 486 ± 55 | $1,250 \pm 300$ | |
| 2 2 | $k_{cat}(s^{-1})$ | | 447 ± 40 | |
| | $\begin{array}{c} \mathbf{k}_{cat}/\mathbf{K}_{m} \\ (\mathbf{s}^{-1}\ \mathbf{M}^{-1}) \end{array}$ | $5.0 \times 10^5 \pm 4.2 \times 10^4$ | $3.6 \times 10^5 \pm 5.9 \times 10^4$ | |

[0272] As can be observed in Table 2, the k_{ca}/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_{ca}/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.

[0273] 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 1,4naphthoquinone:1-naphthol (FIG. 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 (FIG. 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 AaeUPO1 crystal was closely examined and a variant was built by mean of directed mutagenesis in Trp24 (FIG. 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.

[0274] 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 (FIG. 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

[0275] The mutations of the JaWa variant were mapped (SEQ ID NO: 24) onto the structure of the wild AaeUPO1 (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 (FIG. 6 and FIG. 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 hemethiolate domain (in which there is a Glu196-Arg189 acidbase 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 catalysisrelevant 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 (FIG. 8)). B factor makes reference to the rigidity/flexibility of a protein or amino acid region present in a protein or peptide.

[0276] 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)

[0277] 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 $\rm H_2O_2$ and without requiring the addition of antioxidants such as ascorbic acid to the reaction.

[0278] 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

[0279] 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, Oreg., USA), while the naphthalene, propranolol and potassium persulfate are from Acros Organics (Geel, Belgium). 5-hydroxypropranolol was acquired from Santa Cruz Biotechnology (Santa Cruz, Calif., USA).

[0280] The competent *Escherichia coli* XL2-Blue cells and Pfu ultra DNA polymerase were obtained from Agilent Technologies (Santa Clara, Calif., USA) and iProof high-fidelity DNA polymerase was acquired from Bio-Rad (Hercules, Calif., USA). The BamHI and XhoI restriction enzymes were obtained from New England Biolabs (Ipswich, Mass., 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, Calif., USA). The NucleoSpin Plasmid kit is from Macherey-Nagel (Düren, Germany) and the oligonucleotides used were synthesised by Metabion (Bay-

ern, Germany). All the chemical compounds are of the highest purity available in the market.

Directed Evolution

[0281] 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)

[0282] 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'-gcgcattcaagactccattg-3'), 90 nM MJaWa Rev (SEQ ID NO: 44; 5'-gatcttgccgacattttttcc-3'), 0.3 mM deoxynucleotide triphosphates (dNTPs, 0.075 mM of each), 0.1 mM or $0.2 \text{ mM MnCl}_2, 1.5 \text{ mM MgCl}_2, 0.05 \text{ U/}\mu\text{L Taq DNA}$ 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 nonmutagenic 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'-caggeteatectatgeagece-3') and 0.5 µM RMLC (SEQ ID NO: 34; 5'-gggagggggtgaatgtaage-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 highfidelity 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)

[0283] 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.

[0284] To this end, three PCRs were performed in a final volume of 50 μ L. All contained 3% of DMSO, 0.3 mM dNTPs (0.075 mM each), 0.05 U/ μ L PfuUltra DNA polymerase and 2 ng/ μ L of template, but each with different primers. PCR 1 with 0.25 μ M of RMLN (SEQ ID NO: 33), 0.25 μ M of F76 VHG R

```
(SEQ ID NO: 47; 5'-gcaaqtccqtaatqaqattqccqtccacaaqqtqqqccqcatatgtg qccdbqattqcqqc-3),
```

$0.25~\mu M$ of F76 NDT R

[0285]

```
(SEQ ID No: 48; 5'-<u>gcaaqtccqtaatqaqattqccqtccacaaqqtqqqccqc</u>atatgtggcahngattgcggc-3'
```

and 0.25 μM of F76 TGG R

[0286]

```
(SEQ ID NO: 49; 5'-gcaagtccgtaatgagattgccgtccacaaggtgggccgcatatgtggccagattgcggc-3').
```

PCR 2 con 0.25 µM of HF F

[0287]

```
(SEQ ID NO: 50; 5'-geggeceacettgtggaeggeaateteattaeggaettge-3'
```

0.25 µM of S191 VHG R

[0288]

```
(SEQ ID NO: 51; 5'-<u>cccatccacaaaaagattcqcqqqqaaqqtqqtctcqccq</u>taagca gtcdbgaacctaaag-3'
```

0.25 µM of S191 NDT R

[0289]

```
(SEQ ID NO: 52; 5'-cccatccacaaaaaqattcqcqqqqaaqqtqqtctcqccqtaagcaqtahnqaacctaaaq-3')
```

y $0.25~\mu M$ of S191 TGG R

[0290]

```
(SEQ ID NO: 53; 5'-cccatccacaaaaagattcqcqqqqaaqqtqqtctcqccqtaaqcaqtccaqaacctaaaq-3').
```

PCR 3 con 0.25 µM de HF F-RMLC

[0291]

```
(SEQ ID NO: 54; 5'-cggcgaciaccaccttccccgcgaatctttttgtggatggg-3')
```

and 0.25 μ M 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)

[0292] There is a phenylalanine triad in the catalytic pocket of AaeUPO, 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.

[0293] Library F69: two PCRs were performed in a final volume of 50 μ L. The first contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μ M RMLN (SEQ ID NO: 33), 0.5 μ M F69 R (SEQ ID NO: 55; 5'-gaagattgcggcttgattgtcmnnattgaatc-3'), 0.02 U/ μ L iProof DNA polymerase and 2 ng/ μ L 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 μ M RMLC (SEQ ID NO: 34), 0.5 μ M F69 F (SEQ ID NO: 56; 5'-cgcggttcaggaaggattcaatnn-

kgacaatc-3'), 0.02 U/ μ L iProof DNA polymerase and 2 ng/ μ L of template (SoLo comprising SEQ ID NO: 41).

[0294] 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'-catactggegtegect-temnnggtgecatge-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'-ggactcaatgageatggeaccnn-kgaaggeg-3'), 0.02 U/μL iProof DNA polymerase and 2 ng/μL of template (SoLo comprising SEQ ID NO: 41).

[0295] F199 library: two PCRs were performed in a final volume of 50 μ L. The first contained 3% of DMSO, 0.2 mM dNTPs (0.05 mM of each), 0.5 μ M RMLN (SEQ ID NO: 33), 0.5 μ M F199 R (SEQ ID NO: 59; 5'-ccacaaaaagattcgcgggmnnggtggtctcg-3'), 0.02 U/ μ L iProof DNA polymerase and 2 ng/ μ L 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 μ M RMLC (SEQ ID NO: 34), 0.5 μ M F199 F (SEQ ID NO: 60; 5'-ctactgcttacggcgagaccaccnnkcccgcg-3'), 0.02 U/ μ L iProof DNA polymerase and 2 ng/ μ L of template (SoLo comprising SEQ ID NO: 41).

[0296] 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 AaeUPO1, to analyse the synthesis of HDMs, taking 5'-hydroxypropranolol with each by way of example.

Preparation of the Mutant Libraries

[0297] 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 MgSO₄, 31.6 mL of absolute ethanol, 1 mL of 25 g/L chloramphenicol and ddH₂O 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)

[0298] The mother plates were centrifuged (Eppendorf 5810R centrifuge, Germany) for 10 minutes at 3,500 RPM and 4° C. 20 μ L 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 μ L of reaction mixture with propranolol were added to

the daughter plates using a pipetting robot (Multidrop Combi Reagent Dispenser, Thermo Scientific, USA).

[0299] The reaction mixture with propranolol was composed of 50 mM pH 7.0 potassium phosphate buffer, 5 mM propranolol and 2 mM H₂O₂ 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, rescreenings were carried out, in addition to a third rescreening wherein kinetic stability was determined (T_{50}) (P. Molina-Espeja, et al. Appl. Environ. Microbiol. 2014. 80, 3496-3507).

Second Re-Screening

[0300] 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 ddH2O 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 LBamp 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 ddH2O 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

[0301] 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 ddH₂O 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 (OD₆₀₀=0.1). This culture of each clone was incubated for 72 hours at 25° C. and 220 RPM (at peak UPO activity;

OD₆₀₀=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 H₂O₂). 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 T₅₀ was determined as as the temperature value at which the protein loses 50% of its initial activity after incubating for 30

Production of UPO Recombinant Variants in S. cerevisiae [0302] 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 (OD₆₀₀=0.1). This culture of each clone was incubated for 72 hours at 25° C. and at 220 RPM (at peak UPO activity; OD₆₀₀=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 AaeUPO1 Variants

[0303] 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, Calif., USA) through a membrane with a pore size of 10 kDa (Millipore) using a peristaltic pump (Masterflex Easy Load; Cole-Parmer, Vernon Hills, Ill.). 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

[0304] The kinetic constants of the variants of the invention, AaeUPO, PaDa-I, JaWa and SoLo, for ABTS were estimated in 100 mM pH 4.0 sodium phosphate/citrate buffer and 2 mM H₂O₂; and for the other substrates, in 100 mM pH 7.0 potassium phosphate buffer and 2 mM H₂O₂ (propranolol). For H₂O₂, 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

[0305] 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×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% ddH₂O (in both cases with 0.1% of acetic acid) up to 90% methanol and 10% ddH₂O 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 H₂O₂ in 50 mM pH 7.0 potassium phosphate buffer (0.5 mL of final volume). The reaction was started with the addition of H₂O₂ 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 H₂O₂ 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 H₂O₂ 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 H₂O₂ was added every 10 minutes, monitoring the reaction in each addition point taking different aliquots. The standard deviations were less than 5% in all

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

[0306] 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

[0307] 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 (FIG. 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 (FIG. 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

[0308] 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.

[0309] 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.

[0310] As can be observed in FIG. 12, the SoLo variant of SEQ ID NO: 42 showed very similar kinetic thermostability to that of the JaWa mutant (SEQ ID NO: 24).

[0311] The propranolol transformation reaction performed by the wild AaeUPO 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 FIG. 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.

[0312] The kinetic value of AaeUPO, JaWa and SoLo for propranolol, and for ABTS and $\rm H_2O_2$ (Table 3) were determined.

TABLE 3

| | Kine | tic parameters for the var | riants of the invention ar | ıd for wild AaeUPO. | |
|-------------|--------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| Substrate | Kinetic constants | AaeUPO1 | 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 |
| | $K_{cat}/K_m (s^{-1}M^{-1})$ | $8.8 \times 10^6 \pm 6.9 \times 10^5$ | $8.2 \times 10^6 \pm 6.0 \times 10^5$ | $6.9 \times 10^5 \pm 6.3 \times 10^4$ | $6.4 \times 10^5 \pm 6.7 \times 10^4$ |
| Propranolol | Km (µM) | $2,239 \pm 333$ 150 ± 12 | $2,268 \pm 220$ | 244 ± 92 | 391 ± 97 |
| | K_{cat} (s ⁻¹) | 150 ± 12 | 212 ± 11 | 765 ± 76 | 497 ± 35 |
| | $K_{car}/K_m (s^{-1}M^{-1})$ | $6.7 \times 10^4 \pm 4.8 \times 10^3$ | $9.3 \times 10^4 \pm 4.3 \times 10^3$ | $3.1 \times 10^6 \pm 0.9 \times 10^5$ | $1.3 \times 10^6 \pm 0.2 \times 10^5$ |
| Naphthalene | Km (µM) | 156 ± 20 | 578 ± 106 | 127 ± 27 | 789 ± 96 |
| | K_{cat} (s ⁻¹) | 92 ± 3 | 229 ± 17 | 78 ± 3 | 337 ± 20 |
| | $K_{cat}/K_{m} (s^{-1}M^{-1})$ | $5.9 \times 10^5 \pm 5.9 \times 10^4$ | $4.0 \times 10^5 \pm 4.0 \times 10^4$ | $6.2 \times 10^5 \pm 1.1 \times 10^4$ | $4.3 \times 10^5 \pm 2.8 \times 10^4$ |
| H_2O_2 | Km (µM) | $1,370 \pm 162$ | 486 ± 55 | $1,250 \pm 153$ | $1,430 \pm 153$ |
| | $K_{cat} (s^{-1})$ | 290 ± 15 | 238 ± 8 | 446 ± 23 | 446.23. |
| | $K_{cat}/K_m (s^{-1}M^{-1})$ | $2.1 \times 10^5 \pm 1.5 \times 10^4$ | $5.0 \times 10^5 \pm 4.2 \times 10^4$ | $3.1 \times 10^5 \pm 1.8 \times 10^4$ | $3.1 \times 10^5 \pm 1.8 \times 10^4$ |

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 MnCl₂) 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 (FIG. 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.

[0313] 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 AaeUPO with the naphthalene as substrate, differentiating itself in the total turnover values, which are higher for JaWa.

[0314] 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 FIG. 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.

[0315] When the reaction was monitored for long reaction times with the addition of 2 mM $\rm H_2O_2$, 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. (FIG. 15).

TABLE 4

Determination of the total turnover numbers (TTNs) for the variants of the invention and for wild AaeUPO.

| | T | TNs |
|--------|--------------------|-----------------------|
| | With ascorbic acid | Without ascorbic acid |
| AaeUPO | 14,000 | 3,000 |
| JaWa | 48,000 | 15,000 |
| SoLo | 62,000 | 45,000 |

SEQUENCE LISTING

```
<160> NUMBER OF SEQ ID NOS: 66
<210> SEQ ID NO 1
<211> LENGTH: 987
<212> TYPE: DNA
<213> ORGANISM: Agrocybe aegerita
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(987)
<400> SEOUENCE: 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
                                    10
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
           2.0
                               2.5
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
                           40
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
                       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
                    70
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
                                   90
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
ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
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
                        135
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
                    150
                                       155
145
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
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
```

| | | | | | | | | | | - | con | tin | ued | | |
|--|--------------|--------------|-----------|------------|------------|------------|------------|-----------|------------|------------|------------|------------|-----------|------------|-----|
| | | 180 | | | | | 185 | | | | | 190 | | | |
| gct tac Ala Tyr | | | | | | | | | | | | - | | _ | 624 |
| agg gac Arg Asp 210 | _ | | _ | | _ | _ | _ | _ | _ | | _ | | | | 672 |
| ttc agc Phe Ser 225 | | | | | | | | | | | | | | | 720 |
| ggc aca Gly Thr | | _ | | _ | _ | | _ | _ | | | _ | _ | | | 768 |
| aga aat Arg Asn | _ | | _ | | | _ | | | _ | - | | | | | 816 |
| gac ttt Asp Phe | | | | _ | _ | _ | | | | | _ | | | _ | 864 |
| gtc aag Val Lys 290 | | | | _ | | _ | _ | | _ | | _ | | _ | | 912 |
| aat acg Asn Thr 305 | | | _ | | | | _ | | _ | _ | _ | | _ | | 960 |
| cag gtc Gln Val | | | | | _ | - | tga | | | | | | | | 987 |
| <210> SI <211> LI <212> TY <213> OF | ENGTH PE: | H: 32 PRT | 28 | ocybe | e aeç | gerit | ta | | | | | | | | |
| <400> SI | EQUEN | ICE : | 2 | | | | | | | | | | | | |
| Glu Pro 1 | Gly | Leu | Pro 5 | Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | |
| Val Asn | Asp | Glu 20 | | His | | _ | _ | | | _ | | _ | _ | Ile | |
| Arg Gly | Pro 35 | CÀa | Pro | Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | |
| Pro Arg 50 | Asn | Gly | Val | Ala | Thr 55 | Pro | Val | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | |
| Glu Gly 65 | Leu | Asn | Phe | Asp 70 | Asn | Gln | Ala | Ala | Val 75 | Phe | Ala | Thr | Tyr | Ala 80 | |
| Ala His | Leu | Val | Asp 85 | Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | |
| Arg Lys | Thr | Arg 100 | Leu | Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | |
| Gly Gly | Leu 115 | Asn | Glu | His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | |
| Arg Gly 130 | Asp | Ala | Phe | Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | |
| Phe Glu 145 | Gln | Leu | Val | Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | |
| | | | | | | | | | | | | | | | |

| Asn | Leu | Thr | Val | Ala 165 | Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | |
|------------|------------|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| Ile | Ala | Thr | Asn 180 | Pro | Asn | Phe | Ser | Phe | Val | Asp | Phe | Arg | Phe | Phe | Thr | |
| Ala | Tyr | Gly 195 | Glu | Thr | Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | |
| Arg | Asp 210 | Asp | Gly | Gln | Leu | Asp 215 | | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | |
| Phe 225 | | Arg | Met | Pro | Asp 230 | | | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | |
| | Thr | Gly | Val | Glu 245 | | Val | Ile | Gln | Ala 250 | | Pro | Met | Gln | Pro | | |
| Arg | Asn | Val | Gly 260 | | Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | | Ser | |
| Asp | Phe | Ser 275 | | Pro | Cha | Leu | Met 280 | | Glu | Lys | Phe | | | Ile | Thr | |
| Val | - | | Leu | Tyr | Pro | | | Thr | Val | Gln | | 285 Arg | Lys | Ala | Leu | |
| | 290 Thr | Asn | Leu | Asp | | 295 Phe | Phe | Gln | Gly | | 300 Ala | Ala | Gly | Cys | | |
| 305 Gln | Val | Phe | Pro | Tyr | 310 Gly | Arg | Asp | | | 315 | | | | | 320 | |
| | | EQ II | | | | | | | | | | | | | | |
| | | ENGTI PE : | | | | | | | | | | | | | | |
| <213 | > OF | RGANI | ISM: | | ocyb | e ae | geri | ta | | | | | | | | |
| | | EATUF AME/F | | CDS | | | | | | | | | | | | |
| | | CATI EATUR | | (1) | (1 | 113) | | | | | | | | | | |
| <221 | . > NA | AME/F | KEY: | | _pept | | - 1 | | | | | | | | | |
| | | | | | 0) | (111) | b) | | | | | | | | | |
| <400 | > SI | EQUE | ICE : | 3 | | | | | | | | | | | | |
| - | | | | | _ | | | | ttg Leu | _ | | - | | | - | 48 |
| | • | • | -40 | | | | | -35 | _ | | 2 | | -30 | 1 | | |
| - | _ | | | _ | | _ | | _ | gcc | | | _ | _ | _ | _ | 96 |
| val | uta | -25 | LIO | дар | тÀт | WIG | -20 | пеп | Ala | ату | пеп | -15 | GIII | GIII | GIU | |
| _ | - | - | | | | | | | gcc | _ | | | | | | 144 |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg -1 | | Pro | Gly | Leu | Pro 5 | |
| | | | | | | _ | | _ | aag | _ | | | _ | | _ | 192 |
| Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lуз 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | |
| | | | | | | | | | gat | | | | | | | 240 |
| His | Pro | Trp | Lуs 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | Cha | Pro | |
| | | | | | | | | | tac | | | | | | | 288 |
| σιγ | ьeu | Asn 40 | ınr | ьeu | Ala | ser | H1s | стх | Tyr | ьeu | Pro | Arg 50 | asn | стХ | val | |
| - | | _ | | | | | | | gtt Val | _ | - | | | | | 336 |
| ліа | 55 | FIU | vaı | GIII | 116 | 60 | LOII | та | vaı | GIII | 65 | сту | шeu | Poll | F116 | |
| | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | con | CIII | ueu | | |
|------------|-------------------------------------|----------------|--------------|-----|-------------------|------|------|-----|-----|-----------|-----|-----|------------|-----|-----------|------|
| Asp 70 | Asn | Gln | Ala | Ala | Val 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
| | | | | _ | gac Asp | _ | _ | _ | | | _ | _ | _ | | | 432 |
| | | | - | | cca Pro | | | _ | | _ | | | | | | 480 |
| | | | | _ | ggc Gly | _ | _ | - | _ | | _ | | - | _ | | 528 |
| | | | | | gat Asp | | | | _ | | | _ | _ | _ | _ | 576 |
| - | | - | | - | ttt Phe 155 | | | | | | | | | - | | 624 |
| | - | | _ | | aag Lys | _ | | | _ | | | | | | | 672 |
| | | | | - | gac | | | | | | - | | | | | 720 |
| | | | - | | ctt Leu | | | _ | 000 | _ | | _ | _ | | _ | 768 |
| | _ | _ | - | - | gca Ala | | _ | | | | | _ | _ | _ | | 816 |
| _ | - | | | - | gca Ala 235 | | _ | _ | _ | - | | | | - | | 864 |
| ~ | _ | | _ | ~ | cat His | | _ | _ | | | _ | | ~ | | _ | 912 |
| | | _ | | | gtc Val | _ | | | | | _ | | | | | 960 |
| | | | | | aaa Lys | | | | | | | | | | | 1008 |
| _ | | _ | _ | | cag Gln | | _ | | _ | | | _ | | | _ | 1056 |
| | | | | | gtc Val 315 | | | | | | | | | | | 1104 |
| | cga Arg | - | tga | | | | | | | | | | | | | 1116 |
| <21 <21 | 0 > SI 1 > LI 2 > T 3 > OI | ENGTI YPE : | H: 3' PRT | 71 | ocyb | e ae | geri | ī a | | | | | | | | |
| < 40 | 0> S1 | EQUEI | ICE : | 4 | | | | | | | | | | | | |
| Met | Lys | Tyr | Phe -40 | Pro | Leu | Phe | Pro | Thr | Leu | Val | Tyr | Ala | Val -30 | Gly | Val | |

<220> FEATURE: <221> NAME/KEY: CDS

-continued

Val Ala Phe Pro Asp Tyr Ala Ser Leu Ala Gly Leu Ser Gln Glu Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln Glu Gly Leu Asn Phe Asp Asn Gln Ala Ala Val Phe Ala Thr Tyr Ala Ala His Leu Val Asp 80 Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu 95 Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu 110 His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe 125 Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val 140 Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala 155 Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr 190 Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln 205 Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro 220 Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp 300 Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr Gly Arg Asp <210> SEQ ID NO 5 <211> LENGTH: 1116 <212> TYPE: DNA <213> ORGANISM: Agrocybe aegerita

| | | | | | | | | | | | | 0011 | CIII | | | | | |
|------------|--------------------------------------|----------------|-------------|------|-------|------|-----|-----|-----|-----|-----|------|------|-----|-----|-----|--|--|
| <22 <22 | 2 > L0 0 > F1 1 > N2 2 > L0 | EATUI AME/I | RE: KEY: | mat. | _pept | tide | 6) | | | | | | | | | | | |
| < 40 | 0> S1 | EQUEI | NCE: | 5 | | | | | | | | | | | | | | |
| _ | aaa Lys | | | | _ | | | | _ | _ | | _ | | | _ | 48 | | |
| - | gct Ala | | | - | | - | | _ | - | | | _ | | _ | - | 96 | | |
| _ | gac Asp -10 | - | | | | | | | - | - | Glu | | | | | 144 | | |
| | ggt Gly | | | | | _ | | - | | _ | | | - | | - | 192 | | |
| | cca Pro | | _ | _ | | - | | | _ | | _ | | | - | | 240 | | |
| | ctc Leu | | | _ | _ | | | | | | _ | _ | | | _ | 288 | | |
| - | acc Thr 55 | _ | | | | | | | - | _ | _ | | | | | 336 | | |
| - | aat Asn | | - | - | _ | | _ | | | | _ | | | | _ | 384 | | |
| | aat Asn | | | _ | - | _ | _ | - | | | - | _ | _ | | | 432 | | |
| | gly ggg | | | | | | | | | | | | | | | 480 | | |
| | ggc Gly | | | - | | - | _ | - | _ | | - | | - | - | | 528 | | |
| | ggc Gly 135 | | | | | | | | | | | | | | | 576 | | |
| | tac Tyr | | | | | | | | | | | | | | | 624 | | |
| | gag Glu | | | | | | | | | | | | | | | 672 | | |
| | ttc Phe | | | - | _ | | | | | | _ | | | | | 720 | | |
| | ttc Phe | | | | | | | | | | | | | | | 768 | | |
| | gat Asp 215 | _ | - | - | _ | | _ | | | | | _ | _ | _ | | 816 | | |
| gac | gat | ttc | ttc | cgc | gca | ccc | agc | ccg | aga | agt | ggc | aca | gga | gtc | gag | 864 | | |

| | | | | | | | | | | | - | con | tin | ued | | |
|------------|-------------------------------------|----------------|--------------|------------|------------|-------|------------|------------|------------|------------|-----|------------|------------|------------|------------|------|
| Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Gly | Thr | Gly | Val | Glu 245 | |
| - | gtt Val | | _ | _ | | | _ | _ | | | _ | | _ | | _ | 912 |
| | aac Asn | _ | | | _ | _ | | | | | - | | | | | 960 |
| _ | ttg Leu | _ | | | | | _ | | | _ | - | _ | | | | 1008 |
| _ | aat Asn 295 | _ | _ | | _ | | _ | | - | | | _ | | | _ | 1056 |
| | ttc Phe | | _ | | _ | _ | _ | | _ | | _ | _ | | | | 1104 |
| | cga Arg | | tga | | | | | | | | | | | | | 1116 |
| <21 <21 | 0 > SI 1 > LI 2 > T 3 > OI | ENGTI YPE : | H: 3' PRT | 71 | ocybe | e aeg | gerit | :a | | | | | | | | |
| < 40 | 0> SI | EQUEI | ICE : | 6 | | | | | | | | | | | | |
| Met | Lys | Tyr | Phe | Pro | Leu | Phe | Pro | Thr | Leu | Val | Tyr | Ala | Val -30 | Gly | Val | |
| Val | Ala | Phe -25 | Pro | Asp | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| | Asp -10 | | | | | -5 | | | | -1 | 1 | | | | 5 | |
| | Gly | | | 10 | | | | | 15 | | | | - | 20 | | |
| | Pro | _ | 25 | | | | | 30 | _ | | | | 35 | | | |
| | Thr | 40 | | | | | 45 | | - | | | 50 | | | | |
| Asp | 55 Asn | | | | Val | 60 | | | | Ala | 65 | | | | Asp | |
| 70 Gly | Asn | Leu | Ile | | 75 Asp | Leu | Leu | Ser | | 80 Gly | Arg | Lys | Thr | _ | 85 Leu | |
| Thr | Gly | Pro | Asp 105 | 90 Pro | Pro | Pro | Pro | Ala 110 | 95 Ser | Val | Gly | Gly | Leu 115 | 100 Asn | Glu | |
| His | Gly | Thr 120 | | Glu | Gly | Asp | Ala 125 | | Met | Thr | Arg | Gly 130 | | Ala | Phe | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | Phe | Glu | Gln | Leu | Val | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| | | | | | | | | | | | | | | | | |

| Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
|--------------------------------------|---|--|----------------------------------|----------------------|------------|------------|------------|-------------------|------------|------------|------------|------------|-------------------|------------|------------|------------|
| Thr | Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln 200 205 210 Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro 215 220 225 Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu | | | | | | | | | | | | | | | |
| Leu | 215 220 225 sp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val | | | | | | | | | | | | | Pro | | |
| Asp 230 | _ | Phe | Phe | Arg | | Pro | Ser | Pro | Arg | | Gly | Thr | Gly | Val | Glu 245 | |
| | | Ile | Gln | Ala 250 | | Pro | Met | Gln | Pro 255 | | Arg | Asn | Val | Gly 260 | | |
| Ile | Asn | Ser | Tyr 265 | | Val | Asp | Pro | Thr 270 | | Ser | Asp | Phe | Ser 275 | | Pro | |
| CÀa | Leu | Met | | Glu | Lys | Phe | | | Ile | Thr | Val | _ | | Leu | Tyr | |
| Pro | | 280 Pro | Thr | Val | Gln | | 285 Arg | Lys | Ala | Leu | | 290 Thr | Asn | Leu | Asp | |
| | 295 Phe | Phe | Gln | Gly | | 300 Ala | Ala | Gly | Cys | | 305 Gln | Val | Phe | Pro | _ | |
| 310 Gly | Arg | Asp | | | 315 | | | | | 320 | | | | | 325 | |
| <21: <22: <22: <22: <22: | 3 > 01 0 > F1 3 > 0' 0 > F1 1 > NA | YPE: RGAN: EATUR THER EATUR AME/I | SM: RE: INF RE: KEY: | Art: ORMA' CDS | rion | : wt | _ | | riant | = | | | | | | |
| < 40 | 0 > S1 | EQUEI | ICE : | 7 | | | | | | | | | | | | |
| | | gga Gly | | | | | | | | | _ | | _ | _ | _ | 48 |
| | | gac Asp | | - | | | | _ | _ | | _ | | | _ | | 96 |
| | | cct Pro 35 | | | | | | | | | | | | | | 144 |
| | | aat Asn | | | | | | | | | | | | | | 192 |
| - | | ctc Leu | | | - | | | - | - | - | | _ | | | | 240 |
| | | ctt Leu | | | | | | | | | | | | | | 288 |
| 000 | | | | | | | | | | | | | | | | |
| | | acg Thr | | | | | | | | | | | | | | 336 |
| Arg ggt | Lys gga | | Arg 100 aat | Leu gag | Thr cat | Gly ggc | Pro | Asp 105 ttc | Pro gaa | Pro ggc | Pro gac | Pro gcc | Ala 110 agt | Ser atg | Val | 336 384 |

| | | | -continued |
|-------------------|---|-------------------------------|---|
| 130 | 135 | | 140 |
| _ | | | gga gga gga aaa tac 480 Gly Gly Lys Tyr 160 |
| | | | cgc att caa gac tcc 528 Arg Ile Gln Asp Ser 175 |
| | | | ttt agg ttc ttt act 576 Phe Arg Phe Phe Thr 190 |
| | y Glu Thr Thr Phe | | ttt gtg gat ggg cgc 624 Phe Val Asp Gly Arg 205 |
| | | Met Asp Ala Ala | cgg agt ttt ttc caa 672 Arg Ser Phe Phe Gln 220 |
| | | | ccc agc ccg aga agt 720 Pro Ser Pro Arg Ser 240 |
| | | | cct atg cag ccc gga 768 Pro Met Gln Pro Gly 255 |
| _ | | | gac cca aca tcc tct 816 Asp Pro Thr Ser Ser 270 |
| - | r Thr Pro Cys Leu | | ttc gtc aac ata acg 864 Phe Val Asn Ile Thr 285 |
| | _ | Pro Thr Val Gln | ctt cgc aaa gcc ctt 912 Leu Arg Lys Ala Leu 300 |
| - | - | | gcc gct gga tgt acc 960 Ala Ala Gly Cys Thr 320 |
| | c cca tac ggg cga e Pro Tyr Gly Arg 325 | | 987 |
| <220> FEAT | TH: 328 : PRT NISM: Artificial URE: | Sequence nthetic Construct | = |
| <400> SEQU | ENCE: 8 | | |
| Glu Pro Gl | y Leu Pro Pro Gly 5 | Pro Leu Glu Asn 10 | Ser Ser Ala Lys Leu 15 |
| Val Asn As | p Glu Ala His Pro 20 | Trp Lys Pro Leu 25 | Arg Pro Gly Asp Ile 30 |
| Arg Gly Pro | | Asn Thr Leu Ala 40 | Ser His Gly Tyr Leu 45 |
| Pro Arg As: 50 | n Gly Val Ala Thr 55 | Pro Val Gln Ile | Ile Asn Ala Val Gln |
| Glu Gly Le | u Asn Phe Asp Asn 70 | Gln Ala Ala Val | Phe Ala Thr Tyr Ala 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 Va | al |
| Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Th | hr |
| Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Le | eu |
| Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Ty 145 150 155 16 | yr 60 |
| Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Se 165 170 175 | er |
| Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Th | hr |
| Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Ar 195 200 205 | rg |
| Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gl 210 215 220 | ln |
| Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Se 225 230 235 24 | er 40 |
| Asp Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gl 245 250 255 | ly |
| Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Se 260 265 270 | er |
| Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Th | hr |
| Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Le 290 295 300 | eu |
| Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Th 305 310 315 32 | hr 20 |
| Gln Val Phe Pro Tyr Gly Arg Asp 325 | |
| <pre><210> SEQ ID NO 9 <211> LENGTH: 1116 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Wt-JaWa variant with wild signal pept <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (1)(1113) <220> FEATURE: <221> NAME/KEY: mat_peptide <222> LOCATION: (130)(1116)</pre> | tide |
| <400> SEQUENCE: 9 | |
| atg aaa tat ttt ccc ctg ttc cca acc ttg gtc ttc gca gcg agg gt Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Va -40 -35 -30 | |
| gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag ga Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Gln Gl -25 -20 -15 | |
| ttg gac gct ata atc cca aca ctc gag gcc cga gag cca gga tta cc Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pr -10 -5 -1 1 | ro |
| cct ggt cct ctc gag aat agc tct gca aag ttg gtg aac gac gag gc | ct 192 |

| Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|--|
| | | | _ | _ | | _ | | | gat Asp | | _ | | | _ | | 240 | |
| | | | | _ | _ | | | | tac Tyr | | _ | _ | | | _ | 288 | |
| - | | _ | | | | | | | gtt Val | _ | - | | | | | 336 | |
| Asp 70 | Asn | Gln | Ala | Āla | Val 75 | Phe | Ala | Thr | tat Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | 384 | |
| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | atc Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | 432 | |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | tcc Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | 480 | |
| His | Gly | Thr 120 | Phe | Ğlu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Āla | Phe | 528 | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | acg Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | 576 | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | aaa Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | 624 | |
| Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | gac Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | 720 | |
| Asn | Phe | Ser | Phe 185 | Val | Aap | Phe | Arg | Phe 190 | Phe | Thr | Āla | Tyr | Gly 195 | Glu | Thr | 720 768 | |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | ggg Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | 816 | |
| Leu | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe aga | Gln | Phe 225 | Ser | Arg | Met | Pro | 864 | |
| Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | aga Arg | Ser 240 | Asp | Thr | Gly | Val | Glu 245 | 912 | |
| Val | Val | Ile | Gln | Āla 250 | His | Pro | Met | Gln | Pro 255 | Gly | Lys | Asn | Val | Gly 260 | Lys | | |
| Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | tcc Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | 960 | |
| | | | | | | | | | ata Ile | | | | | | | 1008 | |
| _ | | _ | _ | | _ | | _ | | gcc Ala | | | _ | | | _ | 1056 | |
| ttc | ttc | ttc | cag | gga | gtc | gcc | gct | gga | tgt | acc | cag | gtc | ttc | cca | tac | 1104 | |

| Phe 310 | Phe | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Cys | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | |
|------------------------------|------------|----------------------------------|-----------------------------|-----------|------------|------------|------------|------------|------------|------------|-----------|------------|------------|------------|------------|------|
| | cga Arg | | tga | | | | | | | | | | | | | 1116 |
| <21: <21: <21: <22: |)> FI | ENGTI (PE : RGAN: EATUI | H: 3' PRT ISM: RE: | 71 Art | ific: | | | | Const | ruct | Ξ | | | | | |
| < 40 | O> SI | EQUEI | NCE: | 10 | | | | | | | | | | | | |
| Met | ГÀЗ | Tyr | Phe | Pro | Leu | Phe | Pro | Thr | Leu | Val | Phe | Ala | Ala -30 | Arg | Val | |
| Val | Ala | Phe -25 | Pro | Ala | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr -5 | Leu | Glu | Ala | Arg -1 | Glu 1 | Pro | Gly | Leu | Pro 5 | |
| Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | |
| His | Pro | Trp | Lys 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | Сув | Pro | |
| Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | Pro | Arg 50 | Asn | Gly | Val | |
| Ala | Thr 55 | Pro | Val | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Leu | Asn | Phe | |
| Asp | Asn | Gln | Ala | Ala | Val 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
| Gly | Asn | Leu | Ile | Thr | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | |
| His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | Phe | Glu | Gln | Leu | Val | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly | Glu | Leu | Arg | Phe | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn | Phe | Ser | Phe | Val | Asp | Phe | Arg | Phe | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe | Ser | Arg | Met | Pro | |
| Asp 230 | | Phe | Phe | Arg | Ala 235 | | Ser | Pro | Arg | Ser 240 | | Thr | Gly | Val | Glu 245 | |
| | Val. | Tle | Gln | د [۵ | His | Pro | Met | Gln | Pro | | Lare | Δen | Val | Glv | | |
| | | | | 250 | | | | | 255 | _ | | | | 260 | - | |
| Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | |

| Cys | Leu | Met 280 | Tyr | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | |
|---|---|---|--|-------------------------|--------------|------------|-------------------|-----|------|------|------------|------------|-------|-------|-------|--------|
| Pro | | | Thr | Val | Gln | | Arg | Lys | Ala | Leu | | | Asn | Leu | Asp | |
| Phe | 295 Phe | Phe | Gln | Gly | Val | 300 Ala | Ala | Gly | Сув | Thr | 305 Gln | Val | Phe | Pro | Tyr | |
| 310 | | | | - | 315 | | | - | - | 320 | | | | | 325 | |
| GIY | Arg | Asp | | | | | | | | | | | | | | |
| <21: <21: <21: <22: <22: <22: <22: <22: | 1 > L1 2 > T 3 > O1 0 > F1 3 > O' 1 > N 1 | EATU THER EATU AME/: OCAT: EATU AME/: | H: 1 DNA ISM: RE: INF RE: KEY: ION: RE: KEY: | Art. ORMA' CDS (1) mat. | TION(1 _pept | : Wt | | | rian | t wi | th m | odif | ied : | signa | al pe | eptide |
| < 40 | 0 > SI | EQUE | NCE : | 11 | | | | | | | | | | | | |
| _ | | | | | - | | cca Pro | | _ | _ | | _ | | | - | 48 |
| - | _ | | | _ | | - | tca Ser -20 | _ | - | | | _ | _ | _ | - | 96 |
| _ | _ | _ | | | | | ctc Leu | | _ | Arg | | | | | | 144 |
| | | | | | | _ | tct Ser | - | _ | _ | | | - | | - | 192 |
| | | | _ | _ | | _ | cct Pro | | _ | | _ | | | _ | | 240 |
| | | | | _ | - | | cac His 45 | | | | _ | _ | | | - | 288 |
| _ | | _ | | | | | aac Asn | | _ | _ | _ | | | | | 336 |
| _ | | | _ | _ | _ | | gcc Ala | | | | _ | | | | _ | 384 |
| | | | | _ | _ | _ | ctg Leu | _ | | | _ | _ | _ | | | 432 |
| | | | - | | | | ccc Pro | - | | - | | | | | | 480 |
| | | | | - | | - | gcc Ala 125 | - | _ | | _ | | - | - | | 528 |
| | | | | | - | | aat Asn | | _ | | | - | _ | _ | - | 576 |
| gac | tac | agc | aac | cga | ttt | gga | gga | gga | aaa | tac | aat | ctt | acc | gtc | gcg | 624 |

| | | | | | | | | | | | - | con | tin | ued | | |
|----------------------|---|---|-------------------------|---------------------------|------------|-----------|-----------|-----------|-----------|------------|-----------|---------------|-------------|-------------------|------------|------|
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| 999 Gly | | | _ | | _ | _ | | | _ | | | | | | | 672 |
| aat Asn | | | | - | _ | | | | | | - | | | | | 720 |
| acc Thr | | | | | | | | _ | | _ | | _ | - | ggc Gly | _ | 768 |
| cta Leu | | | | | | | | | | | | | | | | 816 |
| gac Asp 230 | - | | | - | - | | _ | _ | - | - | - | | | - | | 864 |
| _ | - | | _ | - | | | _ | _ | | | | | _ | ggc Gly 260 | _ | 912 |
| atc Ile | | _ | | | _ | _ | | | | | - | | | acc Thr | | 960 |
| - | _ | _ | | | | | _ | | | _ | _ | _ | | ctc Leu | | 1008 |
| ccg Pro | | _ | _ | | _ | | _ | | _ | | | _ | | | _ | 1056 |
| ttc Phe 310 | | | _ | | _ | - | _ | | _ | | _ | _ | | | | 1104 |
| ggg Gly | - | - | tga | | | | | | | | | | | | | 1116 |
| <220 <223 <400 | .> LE :> T\ :> OF !> FE :> O\ | ENGTH (PE: RGAN) EATUH THER | H: 3' PRT ISM: RE: INFO | 71 Art: DRMA' 12 | rion | - | nthe | ic (| | | | 7 .7 - | 77-7 | C1- | 17-7 | |
| Met | - | - | -40 | | | | | -35 | | | - | | -30 | - | | |
| Val | | -25 | | _ | | | -20 | | | | | -15 | | | | |
| | -10 | | | | | -5 | | | | -1 | 1 | | | | 5 | |
| Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | |
| His | Pro | Trp | Lув 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | Cys | Pro | |
| Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | Pro | Arg 50 | Asn | Gly | Val | |
| Ala | Thr 55 | Pro | Val | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Leu | Asn | Phe | |
| | | | | | | | | | | | | | | | | |

| Asp 70 | Asn | Gln | Ala | Ala | Val 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
|----------------------|----------------------------------|-----------------------|---------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|----|
| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | ГÀа | Thr | Arg 100 | Leu | |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | |
| His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly | Glu | Leu | Arg | Phe 170 | ГÀа | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe 225 | Ser | Arg | Met | Pro | |
| Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Asp | Thr | Gly | Val | Glu 245 | |
| Val | Val | Ile | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Lys | Asn | Val | Gly 260 | Lys | |
| Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | |
| CAa | Leu | Met 280 | Tyr | Glu | ГÀа | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | |
| Pro | Asn 295 | Pro | Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp | |
| Phe 310 | Phe | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Cha | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | |
| Gly | Arg | Asp | | | | | | | | | | | | | | |
| <211 <212 <213 |)> SE L> LE 2> TY 3> OR | ENGTH PE: RGANI | H: 98 DNA SM: | 34 | fici. | al S | Seque | ence | | | | | | | | |
| <223 |)> FE 3> OI | HER | INFO | RMAI | : NOI | PaI | a-I | vari | ant | with | nout | sigr | nal p | epti | de | |
| <221 |)> FE L> NA 2> LC | ME/F | ŒY: | | . (98 | 34) | | | | | | | | | | |
| < 400 |)> SE | EQUEN | ICE : | 13 | | | | | | | | | | | | |
| | cca Pro | | | | | | | | | | | | | | | 18 |
| | aac Asn | _ | | - | | | | _ | _ | | _ | | | - | | 6 |
| - | gga Gly | | _ | | | | | | _ | - | | | | | | 4 |

| cog age ant gge gtt og acc cog oge can at at at ac gg gtt eag Pro Ary Arn Gly Val Ala The Thr Pro Ala Gln Ile Ile Ann Ala Val Gln S5 San gga tte cast to gge cast cas goe goa atto tte goe act at gog Glu Gly Phe Aen Phe Aen Ann Ann Gln Ala Ala Ile Ihe Ala Thr Tyr Ala S5 Coc cac ctt gtg gac ggs act cat ggg gct gat cca gas get tte dtg age att gag Ala His Leu Val Aen Gly Awn Leu Ile Thr Amp Leu Leu Ser Ile Gly S6 cgc aag acg cgg ctc act ggg gct gat cca cca ccc ccc gct tcc gtt Aen His Leu Val Aen Glu His Gly Pro Aen Pro Pro Pro Pro Pro Pro Ala Ser Val Ilo S6 Cgc aag acg cgg ctc act ggg gc ct gat cca cca ccc ccc gct tcc gtt Aen His Leu Val Aen Glu His Gly Pro Aen Pro | | | | | | | | | | | | | | | | | | | |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
| ôu diy Phe Amp Phe Amp Amn Gin Âla Âla 11e Phe Âla Thr Tyx Âla 75 80 288 gcc cac ctt gtg gac ggc sat ctc att acg gac ttg ctg agc atc gga 288 Ala His Leu Val Amp Giy Amn Leu Ile Thr Amp Leu Leu Ser 11e Giy 35 egc aag acg cgg ctc act ggg gat ctg ca cac acc acc acc acc ggt Lec gtt 314 Arg Lys Thr Amp Leu Thr Giy Pro Amp Pro Pro Pro Pro Pro Ala Ser Val 316 ggt ggat ctc act gga gat tig gac act gac act gac act gac gac gac gac gac gac gat gat gac gac fill girl gac ham Giu Giy Amp Ala Ser Wet Thr 384 115 110 1120 cga ggt gac gac ttc ttt ggc aac aac cac gat ttc aat gag acg ctc gat ttc att gar gap gac gac gat gat gac gac gac gat ttc att gar gap ga gas aat acc gat gat gac gac gat gat gat gac gac gat gat gat gac gac gat gat gat gat gac gac gat gat gat gat gat gac gac gat | _ | Arg | | | - | _ | Thr | _ | | | | Ile | | | _ | _ | 192 | | |
| Ala His Leu Val App Gly Aen Leu 11e Thr App Leu Leu Ser I1e Gly 95 cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt Arg Lyg Thr Arg Leu Thr Gly Pro App Pro Pro Pro Pro Pro Pata Ser Val 110 ggt ggg ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Aen Glu His Gly Thr Phe Glu Gly App Ala Ser Wet Thr 110 ggt ggg gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Aen Glu His Gly Thr Phe Glu Gly App Ala Ser Wet Thr 115 cga ggt gac gga ttc ttt ggc acc aca cac gat ttc aat gag acg ctc Arg Gly App Ala Phe Phe Gly Aen Ann His App Phe Am Glu Thr Leu 130 130 ttc gaa cag ttg gtt gac tac agc aca cgat ttc aat gag agg aca acc acc acc gat tcc att acc gtc gcg ggg ggg gga acc ttc gat acc agc gat gac gcd gad gac acc ttc acc gtc gcg ggg ggg gga acc gtc gcd ggg ggg ggg gga acc gcc gcd gac ggg ggg ggg gga acc gcc gcd gac ggg ggg ggg gga acc gcc gcd gac gac gcc gcd ggg ggg ggg gga acc gcc gcd gac gac gcc gcd gac gac gac gcc gcd tcc ggt tcc aag ggc atc cac gac acc acc acc acc acc acc acc ac | Glu | | | | | Asp | | | _ | _ | Ile | | _ | | | Ala | 240 | | |
| Arg Lye Thr Arg Leu Thr GUY Pro App Pro | | | | | Asp | | | | | Thr | | | | | Ile | | 288 | | |
| Gly Sly Leu Ann Glù His Gly Thr Phe Slu Sly Aep Ala Ser Met Thr 115 Cga ggt gac goa ttc ttt ggc acc acc cac gat ttc act gag acg ctc Arg Sly Aep Ala Phe Phe Gly Aen Aen Hie Aep Phe Aen Glu Thr Leu 130 130 Let cgaa cag ttg gtt gac tac acc acc gat ttc gag gga gga aca tac Phe Glu Gln Leu Val Aep Tyr Ser Aen Arg Phe Gly Gly Gly Lyr Tyr 145 146 act cta acc gtc ggc ggg gag ctc cgt ttc acc gac acc acc act acc acc acc acc acc acc a | | | | Arg | | | | | Asp | | | | | Ala | | | 336 | | |
| Arg Gity Asp Ala Phe Phe Gity Asp Ala His Asp Phe Asm Giu Thr Leu 130 ttc gaa cag ttg gtt gat tac agc aac cga ttt gga gga gga ga aa tac Phe Giu Gin Leu Val Amp Tyr Ser Asm Arg Phe Giy Giy Giy Lys Tyr 145 at ctt acc gtc gcg ggg gg gc ctc cgt ttc aag cga ct tac agc att cac gat cac Asm Leu Thr Val Ala Giy Giu Leu Arg Phe Lys Arg Iie Gin Amp Ser 170 180 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tt act 180 180 att ggg acc aac cac ccc aat ttc tcc ttt gtt gac ttt agg ttc tt act 180 180 att ggg aga acc aac ccc aat ttc tcc gg gad gag gg gg gcc Ala Tyr Giy Giu Thr Thr Phe Pro Ala Asm Leu Phe Val Amp Phe Arg Giy Arg 195 agg acg acg cag cac acc acc ccc ga at atg gat gct gat gg gg gg agg gac gac ggc cag cta gaa tat ga gat gct gac cgg agt ttt ttc caa Ana Tyr Giy Giu Thr Thr Phe Pro Ala Asm Leu Phe Val Amp Giy Arg 195 agg acg acg cac ggc cag cta gaa tat ga gat gct gac cgg agt ttt ttc caa Ang Amp Amp Giy Gin Leu Amp Met Amp Ala Ala Arg Ser Phe Phe Gin 210 210 220 agg aca agg cg ca gat gtc gaa gat ttc ttc cgc gca ccc agc ccg aga agt ttc aag cgt atg ctg gaa gtc gag gt gt gac gac gac acc agc ccg aga agt ttc aag cgt atg far gta cac gat ctc cat gc aag acc 235 aga aat gtc gag gta gtt gta cag gct cat cct atg cag occ gga agg aca gag gtc gag gta gtt gta cag gct cat cct atg cag occ gga agg aca gag gtc gag gta gtt gta cag gct cat cct atg cag occ gga agg aca gag gtc gag gta gt gta cac acc gtc gac cca aca tcc ttc arg Ann Val Giy Lys Iie Ann Ser Tyr Thr Val Amp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg acc acc gtc gac cca aca tcc tct Arg Ann Val Giy Lys Iie Ann Ser Tyr Thr Val Amp Pro Thr Ser Ser 270 gac agg tcc gac gat tcc ccc tgc acc acc gcg gt gac ccc acc acc tcc tcc ccc acc acc tcc tcc acc acc gcg gg ga gac gac gac acc acc | | | Leu | | | | | Thr | | _ | | - | Āla | _ | _ | | 384 | | |
| Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 145 146 146 150 150 150 160 160 160 160 170 170 170 17 | _ | Gly | _ | - | | | Gly | | | | _ | Phe | | | _ | | 432 | | |
| Ash Leu Thr Val Ala Gly Glu Leu Arg Phe Lya Arg Ile Gln Asg Ser 116 Ala Thr Yan Phe Coc at the Ser Phe Val Asg Phe Arg Phe Phe Thr 180 | Phe | _ | _ | _ | - | Asp | | _ | | _ | Phe | | | | | Tyr | 480 | | |
| The Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr 190 get tac gag gag acc acc tect cocc gag aat cett ttt gtg gad gag gag gag gag acc acc Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 205 agg gac gac gac gag cta gat atg gat gat gat gat gat gat gat g | | | | - | Ala | | | | - | Phe | _ | _ | | | Asp | | 528 | | |
| Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 205 agg gac gac gac gac agt at gat atg gat gct gac cag agt ttt ttc caa Arg Asp Gly Gln Leu Asp Met Asp Ala Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt ttc ttc agc gca ccc agc ccg aga agt phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 240 ggc aca gga gtc gag gta gtt gta cag gct cat cct atg cag ccc gga act gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly 255 aga at gtc gga aga at aca ac agc tac acc gtc gac cca aca tcc tct Arg Asp No Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gga aca att gct gac cca aca tcc tct acc acc acc acc acc a | | | | Asn | | | | | Phe | _ | _ | | | Phe | | | 576 | | |
| Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 225 ggc aca gga gtc gag gta gtt gta cag ggt cat cct atg cag ccg aga agt Gly Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly 250 aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct Arg Asp Val Gly Lys Ile Asp Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gcc aca gga att tcc ccc tgc ttg atg tac acc gtc gac cca aca tcc tct Arg Asp Nan Val Gly Lys Ile Asp Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gas aca tcc tct Arg Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asp Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag gtc gcg gcg gcd gcd ctc 280 aat acg aat ccc gat ttc tta ttc cag gga gtc gc gc gcd gcd gda acc 300 aat acg aat ccc gat ttc tta ttc cag gga gtc gcc gcd gcd gda acc 310 Asp Thr Asp Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr 305 aca gtc ttc cca tac ggg cga gat Gln Val Phe Pro Tyr Gly Arg Asp | _ | | Gly | | | | | Pro | | | | | Val | - | | - | 624 | | |
| Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 240 ggc aca gga gtc gag gta gtt gta gtt gta cag gct cat cct atg cag gcc gga gly and glu val val glu val val glu ala gta gcc gtg gad gag at gtg gag gta gtg val gta | | Asp | _ | | _ | | Asp | _ | _ | _ | _ | Arg | _ | | | | 672 | | |
| Gly Thr Gly Val Glu Val Val Val Gln Ala His Pro Met Gln Pro Gly 255 aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc gac cca aca aca tcc tct Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aca aca acc ctc tct Asp Pro Thr Ser Ser 270 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa acc ctc tct Asp Pro Thr Val Asn Ile Thr 285 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa acg cct tys Ala Leu 290 aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc 260 aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc 260 aat acg at ctc cat ac ggg cga gat Gly Val Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat Gly Arg Asp | Phe | - | _ | _ | | Asp | _ | | | _ | Āla | | - | _ | - | Ser | 720 | | |
| Arg Asn Val GIV Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac aca acg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Asp Pro Thr Val Glu Lys Phe Val Asn Ile Thr 280 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Asp Pro Thr Val Glu Lys Asp Asp Pro Thr Val Glu Lys Asp Asp Pro Thr Val Glu Cys Thr 310 aat acg aat ctc gat ttc tta ttc cag gga gtc gar gtc gcc gtc gga tgt acc Asp Thr Asn Leu Asp Phe Leu Phe Glu Gly Val Ala Gly Cys Thr 320 cag gtc ttc cca tac ggg cga gat Gly Arg Asp Pro Tyr Gly Arg Asp Pro Thr Val Asp Pro Tyr Gly Arg Asp | | | | - | Glu | _ | _ | _ | _ | Āla | | | _ | _ | Pro | | 768 | | |
| Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc tta ttc cag gga gtc gcc gct gga tgt acc Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat Gln Val Phe Pro Tyr Gly Arg Asp | | | | Gly | | | | | Tyr | | | | | Thr | | | 816 | | |
| 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 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 Gln Val Phe Pro Tyr Gly Arg Asp | - | | Ser | | | _ | _ | Met | | | | | Val | | | _ | 864 | | |
| 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 | | Lys | | | | | Asn | | | | | Leu | | | | | 912 | | |
| Gln Val Phe Pro Tyr Gly Arg Asp | Asn | _ | | | _ | Phe | | | _ | | Val | - | _ | | _ | Thr | 960 | | |
| | _ | - | | | Tyr | | _ | - | | | | | | | | | 984 | | |

<210> SEQ ID NO 14 <211> LENGTH: 328 <212> TYPE: PRT

```
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<400> SEQUENCE: 14
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu
Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln
Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala
Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
                              105
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
                 120
Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu
                     135
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr
                   150
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                                   170
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr
                              185
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
                           200
 \hbox{Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln} \\
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
                                       235
Gly Thr Gly Val Glu Val Val Gln Ala His Pro Met Gln Pro Gly
Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu
Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr
                   310
                                       315
Gln Val Phe Pro Tyr Gly Arg Asp
              325
<210> SEQ ID NO 15
<211> LENGTH: 1113
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PaDa-I variant with wild signal peptide
```

| | | | | | | | | | | | _ | con | C TIII | ueu | | |
|------------------------------|----------------------------------|----------------|-----------------------------|-----|--------------------|-------------------|----|---|---|---|-----|-----|--------|-----|---|-----|
| <221 <222 <220 <221 | L> NA 2> LO 0> FI L> NA | EATUF AME/F | KEY: ION: RE: KEY: | (1) | (1: _pept 0) | | 3) | | | | | | | | | |
| < 400 |)> SI | EQUE | ICE : | 15 | | | | | | | | | | | | |
| | | | | | | ttc Phe | | | | | | | | | | 48 |
| | | | | | | gcc Ala | | | | | | | | | | 96 |
| _ | _ | _ | | | | aca Thr -5 | | | _ | _ | Glu | | | | | 144 |
| | | | | | | agc Ser | | | | | | | | | | 192 |
| | | | | | | cga Arg | | | | | | | | | | 240 |
| | | | | _ | _ | tct Ser | | | | | _ | _ | | | _ | 288 |
| | | | | | | ata Ile 60 | | | | | | | | | | 336 |
| - | | | - | - | | ttc Phe | _ | | | | - | | | | - | 384 |
| | | | | _ | _ | ttg Leu | _ | _ | | | _ | _ | _ | | | 432 |
| | | | - | | | ccc Pro | | _ | | _ | | | | | | 480 |
| | | | | - | | gac Asp | _ | _ | _ | | _ | | - | _ | | 528 |
| | | | | | - | ttc Phe 140 | | | _ | | | - | _ | _ | - | 576 |
| | | | | | | gga Gly | | | | | | | | | | 624 |
| | | | - | | _ | cgc Arg | | | - | | | | | | | 672 |
| | | | | - | _ | ttt Phe | | | | | - | | | | | 720 |
| | | | | | | ttt Phe | | - | | - | | _ | - | | _ | 768 |
| | - | _ | - | - | - | cgg Arg 220 | - | | | | | - | _ | - | | 816 |

| _ | | | | | | | | | | | | _ | con | tin | ued | | |
|---|----------------------|--------------|------------|--------------------|-----------|-----------|-------------------|------------|------------|-----------|------------|-----------|------------|------------|------------|------------|------|
| Ì | - | _ | | | _ | _ | ccc Pro | _ | _ | _ | _ | | | | _ | | 864 |
| | | | | | | | cct Pro | | | | | | | | | | 912 |
| | | | _ | | | _ | gac Asp | | | | | _ | | | | | 960 |
| | - | _ | _ | | | | ttc Phe | - | | | _ | - | _ | | | | 1008 |
| | _ | | _ | _ | | _ | ctt Leu 300 | _ | | _ | | | _ | | | _ | 1056 |
| I | | | | | | | gcc Ala | | | | | | | | | | 1104 |
| | | cga Arg | - | | | | | | | | | | | | | | 1113 |
| • | <212 <213 <220 | > T: > OF | EATUI | PRT ISM: RE: | Art | | ial : : Syr | _ | | Const | truc | t | | | | | |
| • | < 400 |)> SI | EQUEI | ICE : | 16 | | | | | | | | | | | | |
| ľ | /let | ГХа | Tyr | Phe | Pro | Leu | Phe | Pro | Thr -35 | Leu | Val | Phe | Ala | Ala -30 | Arg | Val | |
| 7 | /al | Ala | Phe -25 | Pro | Ala | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| I | Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg -1 | | Pro | Gly | Leu | Pro 5 | |
| I | Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | |
| Ι | lis | Pro | Trp | Lуs 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | Cys | Pro | |
| (| Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | Pro | Arg 50 | Asn | Gly | Val | |
| 1 | Ala | Thr 55 | Pro | Ala | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Phe | Asn | Phe | |
| | Asp 70 | Asn | Gln | Ala | Ala | Ile 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
| (| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | |
| | ľhr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | |
| Ι | His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| I | ?he | Gly 135 | Asn | Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | Phe | Glu | Gln | Leu | Val | |
| | Asp 150 | Tyr | Ser | Asn | Arg | Phe | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| | | | | | | | | | | | | | | | | | |

| Gly Glu Le | | | | | | | | | | | | | | |
|--|---|---|---|--|---|---|--|--|---|---|---|--|---|------------------|
| | eu Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn Phe Se | er Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr Phe Pr | | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu Asp Me 215 | et Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe 225 | Ser | Arg | Met | Pro | |
| Asp Asp Ph 230 | ne Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Gly | Thr | Gly | Val | Glu 245 | |
| Val Val Va | al Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Arg | Asn | Val | Gly 260 | Lys | |
| Ile Asn Se | er Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | |
| Cys Leu Me | _ | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | |
| Pro Asn Pr 295 | o Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp | |
| Phe Leu Ph 310 | ne Gln | Gly | Val 315 | Ala | Ala | Gly | Cys | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | |
| Gly Arg As | gp | | | | | | | | | | | | | |
| <212> TYPE <213> ORGA <220> FEAT | NISM: | Arti | ifici | ial S | Seque | ence | | | | | | | | |
| <pre><223> OTHE <220> FEAT <221> NAME <222> LOCF <220> FEAT <221> NAME <222> LOCF</pre> | TURE: E/KEY: ATION: TURE: E/KEY: | CDS (1). | (11 _pept | l13) | | vari | lant | with | n moo | difi€ | ed si | ignal | l peptide | |
| <220> FEAT <221> NAME <222> LOCF <220> FEAT <221> NAME | TURE: E/KEY: ATION: TURE: E/KEY: ATION: | CDS (1). mat_ (130 | (11 _pept | l13) | | vari | iant | with | n moo | difi€ | ed si | igna] | l peptide | |
| <pre><220> FEAT <221> NAME <222> LOCF <220> FEAT <221> NAME <222> LOCF</pre> | TURE: E/KEY: ATION: TURE: E/KEY: ATION: URE: URE: E/KEY: ATION: | CDS (1). mat_ (130 | (11 pept)) | 113) :ide (1113 | 3) cca | acc | ttg | gtc | tac | gca | gtg | a aa | gtc | 48 |
| <pre><220> FEAT <221> NAME <222> LOCE <220> FEAT <221> NAME <222> LOCE <400> SEQU atg aaa ta</pre> | TURE: Z/KEY: ATION: TURE: Z/KEY: ATION: UENCE: ATION: tttt Y Phe -40 :t cct ne Pro | CDS (1). mat_ (130) 17 ccc Pro | pept()) | ide (1113 ttc Phe | cca Pro | acc Thr -35 ttg | ttg Leu gcc | gtc Val ggc | tac Tyr ctc | gca Ala agc | gtg Val -30 cag | ggg Gly cag | gtc Val | 48 |
| <pre><220> FEAT <221> NAME <222> LOCY <220> FEAT <221> NAME <222> LOCY <400> SEQU atg aaa ta Met Lys Ty gtt get tt Val Ala Ph</pre> | CURE: E/KEY: ATION: CURE: E/KEY: ATION: UENCE: ATION: UENCE: at ttt vr Phe -40 ct cct ne Pro | CDS (1). mat_(130) 17 ccc Pro gac Asp | ctg Leu | ttc Phe gcc Ala | cca Pro tca Ser -20 | acc Thr -35 ttg Leu | ttg Leu gcc Ala | gtc Val ggc Gly cga | tac Tyr ctc Leu | gca Ala agc Ser -15 | gtg Val -30 cag Gln | ggg Gly cag Gln | gtc Val gaa Glu cct | |
| <pre><220> FEAT <221> NAME <222> LOCE <220> FEAT <221> NAME <222> LOCE <400> SEQU atg aaa ta Met Lys Ty gtt gct tt Val Ala Ph</pre> | CURE: E/KEY: ATION: E/KEY: ATION: UENCE: ATION: UENCE: At ttt AT | CDS (1). mat_(130 17 ccc Pro gac Asp atc Ile | ctg Leu tac Tyr cca Pro | ttc Phe gcc Ala aca Thr -5 agc | cca Pro tca Ser -20 ctc Leu | acc Thr -35 ttg Leu gag Glu | ttg Leu gcc Ala gcc Ala | gtc Val ggc Gly cga Arg -1 ttg | tac Tyr ctc Leu gag Glu 1 | gca Ala agc Ser -15 cca Pro | gtg Val -30 cag Gln gga Gly | ggg Gly cag Gln tta Leu | gtc Val gaa Glu cct Pro 5 | 96 |
| <pre><220> FEAT <221> NAME <222> LOCY <220> FEAT <221> NAME <222> LOCY <400> SEQU atg aaa ta Met Lys Ty gtt get tt Val Ala Ph -1 ttg gac gc Leu Asp Al -10 cct ggt cc</pre> | TURE: E/KEY: ATION: TURE: E/KEY: ATION: JENCE: ATION: JENCE: At ttt AT Phe -40 At cct at at a Ile At at a Ile At ctc Co Leu Agg aag | CDS (1). mat_(130 ccc Pro dcc Asp atc Ile gag Glu 10 ccg | ctg Leu tac Tyr cca Pro | ttc Phe gcc Ala aca Thr -5 agc Ser | cca Pro tca Ser -20 ctc Leu tct Ser | acc Thr -35 ttg Leu gag Glu gca Ala | ttg Leu gcc Ala gcc Ala aag Lys 15 | gtc Val ggc Gly cga Arg -1 ttg Leu | tac Tyr ctc Leu gag Glu 1 gtg Val | gca Ala agc Ser -15 cca Pro aac Asn | gtg Val -30 cag Gln gga Gly gac Asp | ggg Gly cag Gln tta Leu gag Glu 20 tgc | gtc Val gaa Glu cct Pro 5 gct Ala | 96 144 |
| <pre><220> FEAT <221> NAME <222> LOCE <220> FEAT <221> NAME <222> LOCE <400> SEQU atg aaa ta Met Lys Ty gtt gct tt Val Ala Ph</pre> | TURE: E/KEY: ATION: TURE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: ATION: URE: E/KEY: E/KEY: ATION: URE: E/KEY: | CDS (1). mat_(130 ccc Pro) gac Asp atc Ile gag Glu 10 ccg Pro ctg | ctg Leu tac Tyr cca Pro aat Asn ctt Leu gca | ttc Phe gcc Ala aca Thr -5 agc Ser | cca Pro tca Ser -20 ctc Leu tct Ser cct Pro | acc Thr -35 ttg Leu gag Glu gca Ala ggc Gly 30 | ttg Leu gcc Ala gcc Ala aag Lys 15 gat Asp | gtc Val ggc Gly cga Arg -1 ttg Leu att Ile | tac Tyr ctc Leu gag Glu 1 gtg Val cgt Arg | gca Ala agc Ser -15 cca Pro aac Asn gga Gly | gtg Val -30 cag Gln gga Gly gac Asp cet Pro 35 | ggg Gly cag Gln tta Leu gag Glu 20 tgc Cys | gtc Val gaa Glu cct Pro 5 gct Ala cct Pro | 96 144 192 |

| _ | | | _ | _ | | ttc Phe | _ | | | | _ | | | | _ | 384 |
|---|------------|---|---|---|---|-------------------|---|---|---|---|---|---|---|---|---|------|
| | | | | _ | _ | ttg Leu | _ | - | | | _ | _ | _ | | | 432 |
| | | | _ | | | ccc Pro | | _ | | _ | | | | | | 480 |
| | | | | - | | gac Asp | - | _ | _ | | _ | | - | _ | | 528 |
| | | | | | _ | ttc Phe 140 | | | _ | | | _ | _ | _ | - | 576 |
| - | | _ | | - | | gga Gly | | | | | | | | _ | | 624 |
| | | | _ | | _ | cgc Arg | | | _ | | | | | | | 672 |
| | | | | - | _ | ttt Phe | | | | | - | | | | | 720 |
| | | | | | | ttt Phe | | | | | | | | | | 768 |
| | _ | _ | - | - | _ | cgg Arg 220 | _ | | | | | _ | _ | _ | | 816 |
| _ | - | | | _ | - | ccc Pro | _ | _ | _ | - | | | | - | | 864 |
| - | _ | _ | _ | - | | cct Pro | _ | _ | | | _ | | _ | | _ | 912 |
| | | _ | | | _ | gac Asp | | | | | _ | | | | | 960 |
| | | | | | | ttc Phe | | | | | | | | | | 1008 |
| | | | | | | ctt Leu 300 | | | | | | | | | | 1056 |
| | | | _ | | _ | gcc Ala | _ | | _ | | _ | _ | | | | 1104 |
| | cga Arg | _ | | | | | | | | | | | | | | 1113 |

<210> SEQ ID NO 18 <211> LENGTH: 371

<212> TYPE: PRT <213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic Construct

| < 400 |)> SI | EQUEI | ICE : | 18 | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Met | Lys | Tyr | Phe | Pro | Leu | Phe | Pro | Thr | Leu | Val | Tyr | Ala | Val -30 | Gly | Val |
| Val | Ala | Phe -25 | Pro | Asp | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr -5 | Leu | Glu | Ala | Arg -1 | Glu 1 | Pro | Gly | Leu | Pro 5 |
| Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala |
| His | Pro | Trp | Lys 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | Сув | Pro |
| Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | Pro | Arg 50 | Asn | Gly | Val |
| Ala | Thr 55 | Pro | Ala | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Phe | Asn | Phe |
| Asp 70 | Asn | Gln | Ala | Ala | Ile 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 |
| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu |
| His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | ГЛа | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 |
| Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro |
| Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln |
| Leu | Asp 215 | Met | Asp | | Ala | _ | | Phe | Phe | | Phe 225 | Ser | Arg | Met | Pro |
| Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Gly | Thr | Gly | Val | Glu 245 |
| Val | Val | Val | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Arg | Asn | Val | Gly 260 | Lys |
| Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro |
| Cys | Leu | Met 280 | Tyr | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr |
| Pro | Asn 295 | Pro | Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp |
| Phe 310 | Leu | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Cys | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 |
| Gly | Arg | Asp | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | |

<210> SEQ ID NO 19 <211> LENGTH: 984

| | | | | | | | | | | | | | tin | ueu — | | | | |
|--------------------------------------|--|-------|-----------------------------------|-------|------|-------|-------------------|---|-------|-------|-------|------|-------|----------|---|-----|--|--|
| <213 <220 <223 <220 <221 | 3 > OF 0 > FF 3 > OT 0 > FF L > NF | EATUR | SM: RE: INFO RE: KEY: | ORMA' | rion | : Jal | Seque Wa va | | nt w: | ithou | ıt s: | igna | l pej | ptide | € | | | |
| | | | | | (9 | 54) | | | | | | | | | | | | |
| | | EQUE | | | | | | | | | | | | | | | | |
| | | | | | | | cct Pro | | | | _ | | - | _ | _ | 48 | | |
| | | _ | | _ | | | tgg Trp | _ | _ | | _ | | | _ | | 96 | | |
| - | | | _ | | | | aat Asn 40 | | _ | _ | | | | | | 144 | | |
| _ | _ | | | - | - | | ccg Pro | | | | | | | - | - | 192 | | |
| - | | | | | _ | | caa Gln | _ | _ | | | _ | | | | 240 | | |
| - | | | | - | | | ctc Leu | | _ | - | _ | _ | - | | | 288 | | |
| - | _ | _ | | | | | cct Pro | _ | | | | | _ | | _ | 336 | | |
| | | | | | | | acc Thr 120 | | - | | - | _ | _ | _ | | 384 | | |
| - | | _ | - | | | | aac Asn | | | _ | | | | _ | | 432 | | |
| | _ | _ | _ | - | - | | agc Ser | | - | | | | | | | 480 | | |
| | | | - | | | | ctc Leu | - | | _ | - | | | _ | | 528 | | |
| | 5 5 | | | | | | tcc Ser | | _ | _ | | | | | | 576 | | |
| - | | | | | | | ccc Pro 200 | | | | | | - | | - | 624 | | |
| | - | - | | _ | | - | atg Met | - | - | _ | | - | | | | 672 | | |
| | _ | _ | _ | | - | - | ttc Phe | | - | _ | | _ | _ | - | _ | 720 | | |
| - | | | - | | - | - | gta Val | _ | - | | | _ | _ | | | 768 | | |
| | | _ | | _ | | | agc Ser | | | - | - | | | | | 816 | | |

| | | | | | | | | | | | - | con | tin | ued | | |
|--------------------------|---|----------------------------------|-----------------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| | | | 260 | | | | | 265 | | | | | 270 | | | |
| | ttt Phe | | | | | | | | | | | | | | | 864 |
| _ | aag Lys 290 | | | | _ | | _ | _ | | _ | | _ | | _ | | 912 |
| | acg Thr | | | | | | | | | | | | | | | 960 |
| _ | gtc Val | | | | | _ | _ | | | | | | | | | 984 |
| <21 <21 <21 <22 | 0 > Si 1 > Li 2 > T 3 > Oi 0 > Fi 3 > O' | ENGTI YPE : RGAN: EATUI | H: 3: PRT ISM: RE: | 28 Art | | | - | | Const | truc | t | | | | | |
| | 0 > S | | | | | | | | | | | | | | | |
| Glu 1 | Pro | Gly | Leu | Pro 5 | Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lуs 15 | Leu | |
| Val | Asn | Asp | Glu 20 | Ala | His | Pro | Trp | Lys 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | |
| Arg | Gly | Pro 35 | Cys | Pro | Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | |
| Pro | Arg 50 | Asn | Gly | Val | Ala | Thr 55 | Pro | Ala | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | |
| Glu 65 | Gly | Phe | Asn | Phe | Asp 70 | Asn | Gln | Ala | Ala | Ile 75 | Phe | Ala | Thr | Tyr | Ala 80 | |
| Ala | His | Leu | Val | Asp 85 | Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | |
| Arg | Lys | Thr | Arg 100 | Leu | Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | |
| Gly | Gly | Leu 115 | | Glu | | Gly | | | | _ | _ | | | Met | Thr | |
| Arg | Gly 130 | | Ala | Phe | Phe | Gly 135 | Asn | Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | |
| Phe | Glu | Gln | Leu | Val | Asp 150 | - | Ser | Asn | Arg | Phe | Gly | Gly | Gly | Lys | Tyr 160 | |
| Asn | Leu | Thr | Val | Ala 165 | _ | Glu | Leu | Arg | Phe | Lys | Arg | Ile | Gln | Asp 175 | Ser | |
| Ile | Ala | Thr | Asn 180 | Pro | Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe | Phe | Thr | |
| Ala | Tyr | Gly 195 | Glu | Thr | Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | |
| Arg | Asp 210 | Asp | Gly | Gln | Leu | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | |
| Phe | Ser | Arg | Met | Pro | Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | |
| Asp | Thr | Gly | Val | Glu 245 | Val | Val | Val | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | |
| | | | | | | | | | | | | | | | | |

| _ | | | | | | | | | | | | | | | | |
|---|--|--|---|-----------------------|-------------------------------------|-----------------------|------------|------------|-------|------------|------------|------------|------------|----------------|------------|-----|
| Lys | Asn | Val | Gly 260 | ГÀз | Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | |
| Asp | Phe | Ser 275 | Thr | Pro | Cys | Leu | Met 280 | Tyr | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | |
| Val | Lys 290 | Ser | Leu | Tyr | Pro | Asn 295 | Pro | Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | |
| Asn 305 | Thr | Asn | Leu | Asp | Phe | Leu | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Сла | Thr 320 | |
| Gln | Val | Phe | Pro | Tyr 325 | Gly | Arg | Asp | | | | | | | | | |
| <211 <212 <213 <220 <223 <220 <221 <222 <222 <222 | .> LE :> TY :> OF :> OF :> NE :> LC :> NE :> NE :> NE :> NE | EATUR THER EATUR AME/R CATUR EATUR AME/R | H: 1: DNA SM: SE: INF SE: CEY: ON: SE: CEY: | Art. ORMA CDS (1) mat | ific: TION (1: _pept 0) | : Jai 113) Eide | Wa va | | nt wi | ith v | wild | sig | nal p | p e pt: | ide | |
| < 400 |)> SE | EQUE | ICE : | 21 | | | | | | | | | | | | |
| _ | | | | | ctg Leu | | | | _ | - | | - | | | - | 48 |
| _ | _ | | | _ | tac Tyr | _ | | _ | _ | | | _ | _ | _ | _ | 96 |
| _ | - | - | | | cca Pro | | | | - | _ | Glu | | | | | 144 |
| | | | | | aat Asn | _ | | _ | _ | _ | | | _ | | _ | 192 |
| | | | | | ctt Leu | | | | | | | | | | | 240 |
| | | | | | gca Ala | | | | | | | | | | | 288 |
| | | | | | ata Ile | | | | | | | | | | | 336 |
| | | | | | atc Ile 75 | | | | | | | | | | | 384 |
| | | | | | gac | | | | | | | | | | | 432 |
| | | | | | cca Pro | | | | | | | | | | | 480 |
| | | | | - | ggc Gly | _ | - | _ | _ | | - | | - | - | | 528 |
| ttt | ggc | aac | aac | cac | gat | ttc | aat | gag | acg | ctc | ttc | gaa | cag | ttg | gtt | 576 |

| -continued | |
|---|------|
| Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val | |
| gac tac agc aac cga ttt gga gga gga aaa tac aat ctt acc gtc gcg Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala 150 155 160 165 | 624 |
| ggg gag ctc cgt ttc aag cgc att caa gac tcc att gcg acc aac ccc Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro 170 175 180 | 672 |
| aat ttc tcc ttt gtt gac ttt agg ttc ttt act gct tac ggc gag acc Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr 185 190 195 | 720 |
| acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc agg gac gac ggc cag Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln 200 205 210 | 768 |
| cta gat atg gat gct gca cgg agt ttt ttc caa ttc agc cgt atg cct Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro 215 220 225 | 816 |
| gac gat ttc ttc cgc gca ccc agc ccg aga agt gac aca gga gtc gag Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu 230 235 240 245 | 864 |
| gta gtt gta cag gct cat cct atg cag ccc gga aaa aat gtc ggc aag Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys 250 255 260 | 912 |
| atc aac agc tac acc gtc gac cca aca tcc tct gac ttt tcc acc ccc Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro 265 270 275 | 960 |
| tgc ttg atg tac gag aaa ttc gtc aac ata acg gtc aag tca ctc tac Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr 280 285 290 | 1008 |
| ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt aat acg aat ctc gat Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp 295 300 305 | 1056 |
| ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr 310 315 320 325 | 1104 |
| ggg cga gat Gly Arg Asp | 1113 |
| <210> SEQ ID NO 22 <211> LENGTH: 371 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic Construct | |
| <400> SEQUENCE: 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 55 | Pro | Ala | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Phe | Asn | Phe | | |
| Asp 70 | Asn | Gln | Ala | Ala | Ile 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | | |
| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | | |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | | |
| His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | | |
| Gly | Glu | Leu | Arg | Phe 170 | ГЛа | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | | |
| Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | | |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | | |
| Leu | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe 225 | Ser | Arg | Met | Pro | | |
| Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Asp | Thr | Gly | Val | Glu 245 | | |
| Val | Val | Val | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Lys | Asn | Val | Gly 260 | Lys | | |
| Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | | |
| CÀa | Leu | Met 280 | Tyr | Glu | ГÀа | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | | |
| Pro | Asn 295 | Pro | Thr | Val | Gln | Leu 300 | Arg | ГЛа | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp | | |
| Phe 310 | Leu | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | CÀa | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | | |
| Gly | Arg | Asp | | | | | | | | | | | | | | | |
| <211 <212 <213 <220 <223 <220 <221 <222 <220 <221 <222 | D> SI L> LE L> TY 3> OF FI 3> OT L> NA L> NA C2> LO C2> LO C2> LO C3 C4 C4 C4 C5 C5 C6 C6 C7 C | ENGTH (PE: RGAN: REATUR PHER EATUR AME/I DCAT: AME/I AME/I DCAT: | H: 1: DNA SM: SE: INFO SE: CEY: ON: SE: CON: | Art: ORMA: CDS (1) mat_ (13) | rion: (11 | : JaN | - ∛ava | | it wi | ith v | wild | mod: | ified | d siģ | gnal | peption peption in the period of the period | de |
| atg | aaa | tat | ttt | ccc | _ | | | | _ | - | | - | | | _ | | 48 |
| Met | Lys | Tyr | Phe -40 | Pro | Leu | Phe | Pro | Thr | Leu | Val | Tyr | Ala | Val -30 | Gly | Val | | |
| gtt | gct | ttt | cct | gac | tac | gcc | tca | ttg | gcc | ggc | ctc | agc | cag | cag | gaa | | 96 |

| Val | Ala | Phe -25 | Pro | Asp | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
|-----|-------------------|------------|-----|-----|-----|-----|------------|-----|-----|-----|-----|------------|-----|-----|-----|------|
| _ | gac Asp -10 | _ | | | | | | | _ | _ | Glu | | | | | 144 |
| | ggt Gly | | | | | - | | - | _ | _ | | | - | | - | 192 |
| | cca Pro | | _ | _ | | _ | | | - | | - | | | - | | 240 |
| | ctc Leu | | | _ | _ | | | | | | _ | _ | | | - | 288 |
| - | acc Thr 55 | _ | | | | | | | - | _ | - | | | | | 336 |
| _ | aat Asn | | - | - | | | _ | | | | _ | | | | - | 384 |
| | aat Asn | | | | | | | | | | | | | | | 432 |
| | ggg ggg | | - | | | | | _ | | _ | | | | | | 480 |
| | ggc Gly | | | _ | | _ | _ | _ | _ | | _ | | _ | _ | | 528 |
| | ggc Gly 135 | | | | _ | | | | _ | | | _ | _ | _ | - | 576 |
| _ | tac Tyr | _ | | _ | | | | | | | | | | _ | | 624 |
| | gag Glu | | - | | _ | - | | | - | | | | | | | 672 |
| | ttc Phe | | | | | | | | | | | | | | | 720 |
| | ttc Phe | | | | | | | | | | | | | | | 768 |
| | gat Asp 215 | | | | | | | | | | | | | | | 816 |
| | gat Asp | | | | | | | | | | | | | | | 864 |
| | gtt Val | | | | | | | | | | | | | | | 912 |
| | aac Asn | - | | | - | - | | | | | - | | | | | 960 |
| tac | tta | atg | tac | gag | aaa | ttc | gtc | aac | ata | acg | gtc | aag | tca | ctc | tac | 1008 |

| -continued | |
|---|------|
| 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 Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp 295 300 305 | 1056 |
| ttc tta ttc cag gga gtc gcc gct gga tgt acc cag gtc ttc cca tac Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr 310 315 320 325 | 1104 |
| ggg cga gat Gly Arg Asp | 1113 |
| <210> SEQ ID NO 24 <211> LENGTH: 371 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: Synthetic Construct | |
| <400> SEQUENCE: 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 Arq Phe Lys Arq 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
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
          250 255
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
                            270
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
Gly Arg Asp
<210> SEQ ID NO 25
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Agrocybe aegerita
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(129)
<400> SEOUENCE: 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
                                 10
gtt gct ttt cct gcc tac gcc tca ttg gcc ggc ctc agc cag cag gaa
                                                                      96
\label{thm:conditional} \mbox{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
<210> SEQ ID NO 26
<211> LENGTH: 43
<212> TYPE: PRT
<213> ORGANISM: Agrocybe aegerita
<400> SEQUENCE: 26
Met Lys Tyr Phe Pro Leu Phe Pro Thr Leu Val Phe Ala Ala Arg Val
Val Ala Phe Pro Ala Tyr Ala Ser Leu Ala Gly Leu Ser Gln Glu
Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg
<210> SEQ ID NO 27
<211> LENGTH: 129
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Signal peptide modified with respect to the
     wild signal peptide from A. aegerita
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (129)
<400> SEQUENCE: 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
            5
                                10
```

| | | c ggc ctc agc cag cag gaa a Gly Leu Ser Gln Gln Glu 30 | 96 |
|---|--|--|----|
| | c cca aca ctc gag gcc e Pro Thr Leu Glu Ala 40 | 9 | 29 |
| <pre><210> SEQ ID NO 28 <211> LENGTH: 43 <212> TYPE: PRT <213> ORGANISM: Ar <220> FEATURE: <223> OTHER INFORM</pre> | tificial Sequence ATION: Synthetic Cons | struct | |
| <400> SEQUENCE: 28 | | | |
| Met Lys Tyr Phe Pro | o Leu Phe Pro Thr Leu 10 | ı Val Tyr Ala Val Gly Val 15 | |
| Val Ala Phe Pro As 20 | p Tyr Ala Ser Leu Ala 25 | a Gly Leu Ser Gln Gln Glu 30 | |
| Leu Asp Ala Ile Il 35 | e Pro Thr Leu Glu Ala 40 | a Arg | |
| <pre><210> SEQ ID NO 29 <211> LENGTH: 1113 <212> TYPE: DNA <213> ORGANISM: Ar <220> FEATURE: <223> OTHER INFORM <220> FEATURE: <221> NAME/KEY: CD. <222> LOCATION: (1 <220> FEATURE: <221> NAME/KEY: ma <222> LOCATION: (1)</pre> | ATION: W24F variant o S)(1113) t_peptide | obtained from the PaDa-I variant | |
| <400> SEQUENCE: 29 | | | |
| - | | g gtc tac gca gtg ggg gtc 1 Val Tyr Ala Val Gly Val -30 | 48 |
| | | c ggc ctc agc cag cag gaa a Gly Leu Ser Gln Gln Glu -15 | 96 |
| | | c cga gag cca gga tta cct 1 a Arg Glu Pro Gly Leu Pro -1 1 5 | 44 |
| | | g ttg gtg aac gac gag gct 1 s Leu Val Asn Asp Glu Ala 20 | 92 |
| - | | t att cgt gga cct tgc cct 2 p Ile Arg Gly Pro Cys Pro 35 | 40 |
| | | c ctc ccg aga aat ggc gtt 2 r Leu Pro Arg Asn Gly Val 50 | 88 |
| | | t cag gaa gga ttc aat ttc 3 1 Gln Glu Gly Phe Asn Phe 65 | 36 |
| | _ | t gcg gcc cac ctt gtg gac 3 r Ala Ala His Leu Val Asp 80 85 | 84 |

| | | | | _ | gac | _ | _ | - | | | _ | _ | _ | | | 432 |
|------------------------------|----------------|----------------------------------|-----------------------------|-----------|-------------------|-------|-----|-------|--------|--------|---|---|---|---|---|------|
| | | | _ | | cca Pro | | | _ | | _ | | | | | | 480 |
| | | | | - | ggc Gly | _ | - | - | _ | | _ | | _ | _ | | 528 |
| | | | | | gat Asp | | | | _ | | | _ | _ | _ | _ | 576 |
| - | | _ | | _ | ttt Phe 155 | | | | | | | | | _ | | 624 |
| | | | | | aag Lys | | | | | | | | | | | 672 |
| | | | | - | gac | | | | | | - | | | | | 720 |
| | | | | | ctt Leu | | | | | | | | | | | 768 |
| | _ | _ | - | - | gca Ala | | _ | | | | | _ | _ | _ | | 816 |
| - | _ | | | - | gca Ala 235 | | _ | _ | _ | _ | | | | - | | 864 |
| - | _ | _ | _ | - | cat His | | _ | _ | | | _ | | _ | | _ | 912 |
| | | _ | | | gtc Val | _ | | | | | _ | | | | | 960 |
| ~ | | _ | | | aaa Lys | | ~ | | | _ | ~ | _ | | | | 1008 |
| | | | | | cag Gln | | | | | | | | | | | 1056 |
| | | | | | gtc Val 315 | | | | | | | | | | | 1104 |
| | cga Arg | - | | | | | | | | | | | | | | 1113 |
| <213 <213 <213 <220 |)> FI | ENGTI YPE : RGAN: EATUI | H: 3' PRT ISM: RE: | 71 Art | ific: | | _ | | Jon of | - 7011 | - | | | | | |
| | 3> 0. 0> SI | | | | TION | . ayı | 161 | -1C (| COIIS | LIUC | | | | | | |
| | | ~ | | | | | | | | | | | | | | |

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 Glu -20 Leu Asp Ala Ile Ile Pro Thr Leu Glu Ala Arg Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala His Pro Phe Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe 120 125 Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val 140 Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro 170 175 Asn Phe Ser Phe Val Asp Phe Arg Phe Phe Thr Ala Tyr Gly Glu Thr 190 Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Asp Asp Gly Gln 205 Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Gly Thr Gly Val Glu 235 240 Val Val Val Gln Ala His Pro Met Gln Pro Gly Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr 315 320 Gly Arg Asp <210> SEQ ID NO 31 <211> LENGTH: 1113 <212> TYPE: DNA <213> ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: W24F variant obtained from the Jawa variant <220> FEATURE:

| | | | | | | | | | | _ | COII | tını | ueu | | | |
|---|----------------------|----------------------|----------------|-------------|-----|----|---|---|---|-----|------|------|-----|---|-----|--|
| <221> I <222> I <220> I <221> I <222> I | LOCA FEAT NAME | TION URE: /KEY | : (1) : mat | (1 _pept | ide | 3) | | | | | | | | | | |
| <400> \$ | SEQU | ENCE | : 31 | | | | | | | | | | | | | |
| atg aaa Met Lys | | | e Pro | _ | | | | _ | _ | | _ | | | _ | 48 | |
| gtt gct Val Ala | | e Pro | _ | | _ | | _ | _ | | | _ | _ | _ | _ | 96 | |
| ttg gad Leu Asp -10 | p Āl | | | | | | | - | - | Glu | | | | | 144 | |
| cct ggt Pro Gly | | | | | _ | | - | _ | _ | | | - | | - | 192 | |
| cac cca His Pro | | | | | | | | | | | | | | | 240 | |
| ggt cto Gly Leo | | | | | | | | | | | | | | | 288 | |
| gca aco Ala Thi 55 | r Pr | | | | | | | - | _ | - | | | | | 336 | |
| gac aat Asp Asi 70 | | - | - | | | - | | | | - | | | | - | 384 | |
| ggc aat Gly Ası | | | _ | _ | _ | _ | _ | | | _ | _ | _ | | | 432 | |
| act ggg | _ | _ | Pro | | | | _ | | _ | - | - | | | - | 480 | |
| cat ggo His Gly | | r Phe | | | | | | | | | | | | | 528 | |
| ttt ggd Phe Gly 135 | y As | | | _ | | | | _ | | | _ | _ | _ | _ | 576 | |
| gac tac Asp Ty: 150 | _ | | _ | | | | | | | | | | _ | | 624 | |
| ggg gaq Gly Glı | - | _ | | _ | - | | | - | | | | | | | 672 | |
| aat tto Asn Pho | | | e Val | - | | | | | | - | | | | | 720 | |
| acc tto Thr Phe | | o Ala | | | | | _ | | _ | | - | - | | _ | 768 | |
| cta gat Leu Asp 21! | p Me | | _ | - | | _ | | | | | _ | _ | _ | | 816 | |

| | | | | | | | | | - | con | tin | ued | | |
|---|--------------------------------------|------------|------------|-----------|------------|------------|-----------|------------|-----------|------------|------------|-----|------------|------|
| gac gat tt Asp Asp Ph 230 | | - | _ | | _ | _ | _ | _ | - | | | _ | | 864 |
| gta gtt gt Val Val Va | _ | _ | | | _ | _ | | | | | _ | | _ | 912 |
| atc aac ag Ile Asn Se | | | | | | | | | | | | | | 960 |
| tgc ttg at Cys Leu Me 28 | et Tyr | | | | | | | | | | | | | 1008 |
| ccg aat cc Pro Asn Pr 295 | | | | | | | | | | | | | | 1056 |
| ttc tta tt Phe Leu Ph 310 | _ | | _ | - | _ | | _ | | _ | _ | | | | 1104 |
| ggg cga ga Gly Arg As | | | | | | | | | | | | | | 1113 |
| <210> SEQ <211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHE | STH: 3' E: PRT ANISM: TURE: | 71 Art: | | | _ | | Const | truci | t | | | | | |
| <400> SEQU | JENCE : | 32 | | | | | | | | | | | | |
| Met Lys Ty | -40 | | | | | -35 | | | - | | -30 | - | | |
| Val Ala Ph | 25 | - | - | | -20 | | | - | | -15 | | | | |
| Leu Asp Al -10 Pro Gly Pr | | | | -5 | | | | -1 | 1 | | | | 5 | |
| His Pro Ph | | 10 | | | | | 15 | | | | - | 20 | | |
| Gly Leu As | 25 | | | J | | 30 | - | | J | • | 35 | • | | |
| 40 Ala Thr Pr | | Gln | Ile | Ile | 45 Asn | Ala | Val | Gln | Glu | 50 Gly | Phe | Asn | Phe | |
| 55 Asp Asn Gl | ln Ala | Ala | | 60 Phe | Ala | Thr | Tyr | | 65 Ala | His | Leu | Val | _ | |
| 70 Gly Asn Le | eu Ile | Thr | 75 Asp | Leu | Leu | Ser | Ile 95 | 80 Gly | Arg | Lys | Thr | Arg | 85 Leu | |
| Thr Gly Pr | o Asp 105 | | Pro | Pro | Pro | Ala 110 | | Val | Gly | Gly | Leu 115 | | Glu | |
| His Gly Th | | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| Phe Gly As | sn Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | Phe | Glu | Gln | Leu | Val | |
| Asp Tyr Se 150 | er Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| | | | | | | | | | | | | | | |

| Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
|--|--------------------------------------|--|-----------------------------|------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|----|
| Asn : | Phe | Ser | Phe | Val | Asp | Phe | Arg | Phe | Phe | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr | Phe | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu . | Asp 215 | | Asp | Ala | Ala | Arg 220 | | Phe | Phe | Gln | Phe | | Arg | Met | Pro | |
| Asp . | | Phe | Phe | Arg | Ala 235 | | Ser | Pro | Arg | Ser 240 | | Thr | Gly | Val | Glu 245 | |
| Val ' | Val | Val | Gln | Ala 250 | | Pro | Met | Gln | Pro 255 | | FÀa | Asn | Val | Gly 260 | | |
| Ile . | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | | Ser | Asp | Phe | Ser 275 | | Pro | |
| Cys : | Leu | Met 280 | | | Lys | Phe | Val 285 | | Ile | Thr | Val | Lys 290 | | Leu | Tyr | |
| Pro . | Asn 295 | | Thr | Val | Gln | Leu 300 | | Lys | Ala | Leu | Asn 305 | | Asn | Leu | Asp | |
| Phe : | | Phe | Gln | Gly | Val 315 | | Ala | Gly | Cys | Thr | | Val | Phe | Pro | Tyr 325 | |
| Gly . | Arg | Asp | | | 313 | | | | | 320 | | | | | 323 | |
| <210 <211 <212 <213 <220 <223 | > LE > TY > OF > FE > OT | ENGTI PE: RGANI EATUI PHER | H: 2. DNA ISM: RE: INF | 2 Art ORMA | | | | | r | | | | | | | |
| cctc | tata | ict t | taa | cgtc | aa g | 3 | | | | | | | | | | 22 |
| <210 <211 <212 <213 <220 <223 | > LE > TY > OF > FE | ENGTI PE: RGAN: EATUI | H: 2 DNA ISM: RE: | 0 Art | | | | | r | | | | | | | |
| <400 | > SE | EQUEI | ICE: | 34 | | | | | | | | | | | | |
| ggga | gggd | gt (| gaat | gtaa | gc | | | | | | | | | | | 20 |
| <210 <211 <212 <213 <220 <223 | > LE > TY > OF > FE | ENGTI PE: RGANI EATUI | H: 4. DNA ISM: RE: | 2 Art | | | _ | | mer | | | | | | | |
| <400 | > SE | EQUEI | ICE : | 35 | | | | | | | | | | | | |
| ctca | ccca | ıtt t | aag | ccgc | tt c | gacci | tggc | g ata | atte | gtgg | ac | | | | | 42 |
| <210 <211 <212 <213 <220 <223 | > LE > TY > OF > FE | ENGTI PE: RGANI EATUI | H: 4. DNA ISM: RE: | 2 Art | | | _ | | mer | | | | | | | |

| SECONDECE: 36 |
|--|
| 2110. SBQ ID NO 37 2111. LENOTH: 984 2112. TYPE: DBA 2123. ORGANISK: Artificial Sequence 2203. OFREATURE: 2213. ORGANISK: Artificial Sequence 2216. PEATURE: 2223. OFREATURE: 2221. NEMB_KEY: CBS 2222. LOCATION: (11(984) 2400. SEQUENCE: 37 gag coa gga tta cat cot ggt cot ctc gag aat agc tct gca aag ttg Giu Pro Gly Leu Pro Pro Gly Pro Leu Glu Ann Ser Ser Ala Lyr Leu 1 1 1 25 10 25 26 27 28 29 29 20 21 29 29 20 21 29 29 20 21 29 29 20 21 29 20 21 29 20 21 29 20 21 29 20 20 21 29 20 20 20 20 20 20 20 20 20 |
| c2112 TMPS; DNA c2123 ORGANISM: Artificial Sequence c220- FRATURE: c2215 OFREXIDE: c2215 NAME/KEY: CDS c221- NAME/KEY: CDS c222- LOCATION: (1)(984) c400- SEQUENCE: 37 gag coa gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Amn Ser Ser Ala Lyz Leu 1 |
| gag cca gga tta cct cct ggt cct ctc gag aat agc tct gca aag ttg Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Aan 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 Val Aan Aap Glu Ala His Pro Trp Lys Pro Leu Arg Pro Cly Aap Ile 20 25 20 cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac gag tac ctc Arg Gly Pro Cys Pro Gly Leu Aan Thr Leu Ala Ser His Gly Tyr Leu 35 40 45 ccc gag aat ggc gtt gca acc ccg gcg caa att att act agc ggt tcag Pro Arg Aan Gly Val Ala Thr Pro Ala Gln Ile Ile Aan Ala Val Gln 50 55 60 gaa gga ttc aat ttc gac aat caa gcc gca atc ttc gcc aca tat gcg Glu Gly Phe Aan Phe Aap Aan 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 gag Ala His Leu Val Aap Gly Aan Leu Ile Thr Aap Leu Leu Ser Ile Gly 85 90 cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt Arg Lys Thr Arg Leu Thr Gly Pro Aap Pro Pro Pro Pro Pro Ala Ser Val 100 105 110 ggt gga ctc aat gag cat gca cc ttc gaa ggc gac gca gt atc Gly Gly Leu Aan Glu His Gly Thr Phe Glu Gly Aap Ala Ser Met Thr 115 120 125 cga agg tgac gca tct ttt tgc aac aac cac gat ttc aat gag acc gc Gly Gly Leu Aan Glu His Gly Pro Aap Pro |
| Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Ann Ser Ser Ala Lys Leu 1 5 10 15 gtg aac gac gag get cac cac tgg aag ceg ctt ega cet gag at att Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Cly Asp Ile 20 25 30 cgt gga cet tgc cet ggt ctc aat act ctg gea tct cac ggg tac ctc Arg Gly Pro Cys Pro Gly Leu Asm Thr Leu Ala Ser His Gly Tyr Leu 35 40 ccg aga aat ggc gtt gca acc ceg gcg caa ata ata ata aca geg gtt cag Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln 50 55 60 gaa gga ttc aat tct gac act cac gcg gca atc tct gcc aca tat gcg Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala 65 70 80 gac cac ctt gtg gac gga act ctc att acg gac ttg ctg agc atc gga Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly 85 90 cgc aag acg cgg ctc act ggg cct gat cca cac ccc ccc gct tcc gtt Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Pro Ala Ser Wal 100 105 110 ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr 115 120 cga ggt gac gca ttc ttt ggc aac acc acc acc gat tc at gag acc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 135 120 cga ggt gac gca ttc ttt ggc aac acc acc gat ttc aat gag acc ct Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 131 125 cqa ggt gac gca ttc ttt ggc aac acc cac gca ttc aat gac acc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 131 120 cat tac gc gc ggg gag ctc cgt ttc aag gcg att caa gac tcc Arg Leu Thr Val Ala Cly Glu Leu Arg Phe Lyc Arg Ile Gln Asp Ser 165 170 gct tac gg gag acc acc att cc ccc gcg aat ctt ttt ggg gag gga aca tc 165 170 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt ggg gat ggc 281 act cta ggc gag acc acc ttc ccc gcg aat ctt ttt ggg gat ggc 282 act gg gg gag acc acc ttc ccc gcg gat ctt ttt ggg gat ggc 284 arg gag gac gac gac gac acc ttc ccc gcg gat ctt ttt ggg gat ggc 285 acc gg gac gac gac gac acc ttc ccc gcg gat ctt ttt ggg gat ggc 286 arg gag gac gac gac gat at gat ttc ttc g |
| Vail Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile 20 25 26 21 25 20 2 |
| Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu 35 |
| Pro Arg Asn Gly Val Åla Thr Pro Åla Gln Ile Ile Asn Åla Val Gln 55 60 60 60 55 60 60 60 60 60 60 60 60 60 60 60 60 60 |
| Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala 65 70 80 288 gcc cac ctt gtg gac ggc at ctc att acg gac ttg ctg agc atc gga Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly 95 cgc aag acg cgg ctc act ggg cct gat cca cca ccc ccc gct tcc gtt Arg Lys Thr Arg Leu Thr Gly Pro App Pro Pro Pro Pro Pro Ala Ser Val 100 110 ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr 115 120 120 125 cga ggt gac gca ttc ttt ggc aac acc cac gat ttc aat gag acg ctc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 140 ttc gaa cag ttg gtt gac tac agc acc gat ttc gag gga gga aaa tac Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 150 160 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag gcg att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 att gcg acc acc ca att ct cct ttg ttg tgac ttt agg ttg tct ct act 175 att gcg acc acc ca att ct ccc gcg at ctt ttt gtg gat ggg cgc Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr 180 gct tac ggc gag acc acc ttc ccc gcg at ctt ttt gtg gat ggc gc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 acg gac gac gac gca cag cta gat atg gat gcc acg agt ttt ttc caa Arg Asp Asp Alp Gly Gln Leu Asp Masp Ala Ala Arg Ser Phe Phe Gln 210 ctc agc cgt atg cct gac gat ttc ccc gcc acc acc gat ttt ttc caa Arg Asp Asp Gly Gln Leu Asp Masp Ala Ala Arg Ser Phe Phe Gln 210 ctc agc cgt atg cct gac gat ttc ttc ccg gca ccc acc acc acc acc acc acc ac |
| Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly 85 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 ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr 115 cga ggt gac gca ttc ttt ggc aac acc cac gat ttc aat gag acg ctc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga gaa aaa tac Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 176 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 186 at ctg gac acc acc caat ttc tcc ttt gtt gac ttt agg ttc tct act 187 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 180 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg gc gc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 acg gac gac gac gcc act gat atg gat gct gac cgg agt ttt ttc caa Arg Asp Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 acg gac gac gac cac cac at gat atg gat gcc acc acc gac acc acc acc acc acc acc |
| Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val 1100 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 cga ggt gac gca ttc ttt ggc aac acc cac gat ttc aat gag acg ctc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttt cat att gg ser Info 185 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 175 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt gg gtt gg ttc act gg gg gag acc acc ttc ccc gcg aat ctt ttg gat gg gg gg gg cgc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 agg gac gac gac gcc atg cta gat atg gat gct gca cgg agt ttt ttc caa Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asp Ala Ala Arg Ser Phe Phe Gln 210 210 220 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt tt ttc caa Grac Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 2210 Etc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt Typ Arg Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Pro Arg Ser |
| Cay ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg ctc Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 180 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 180 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggc cgc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 agg gac gac gac gac cta gat atg gat gct gca cgg agt ttt ttc caa Asp Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc ccc agc acc acc cagc ccg aga agt ttc tcc aag 720 ttc agc cgt atg cct gac gat ttc ttc ccc agc acc acc agc ccc agc aga agt 720 ttc agc cgt atg cct gac gat ttc ttc ccc agc acc acc agc ccc agc aga agt 720 Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu 130 ttc gaa cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 att gcg acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 170 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttg gt gat ggc gcg Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 agg gac gac gac gc cag cta gat atg gat gct gca cgg agt ttt ttc caa Asp Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt ttc tc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccc agc ccg aga agt ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccc agc ccg aga agt ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccc agc ccg aga agt ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccc agc ccg aga agt ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccc agc ccg aga agt Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| Phe Glu Glu Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Lys Tyr 145 |
| Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 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 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 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 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 |
| Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr 180 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggg cgc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 200 agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt Tyr Gly Gly Gln Leu Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| 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 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 Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| 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 Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser |
| |

| _ | | | | | | | | | | | | con | tin | uea | | | |
|------------------------------|-------------------------|------------|-----------------------------|------------|------------|------------|-------------------|------------|-----------|-----------|-----------|------------|------------|------------|------------|-----|--|
| _ | | | _ | | _ | _ | gta Val | _ | _ | | | _ | _ | | | 768 | |
| | | - | | _ | | | agc Ser | | | _ | - | | | | | 816 | |
| | | | | | | | atg Met 280 | | | | | | | | | 864 | |
| | | | | | | | ccg Pro | | | | | | | | | 912 | |
| | _ | | | _ | | | ttc Phe | _ | | _ | _ | _ | | _ | | 960 | |
| _ | - | | | | Gly 999 | _ | - | | | | | | | | | 984 | |
| <211 <212 <213 <220 | .> LE !> T\ !> OF | EATUR | H: 32 PRT ISM: RE: | 28 Art | | | Seque nthet | | Const | truct | t | | | | | | |
| < 400 |)> SI | EQUEI | ICE : | 38 | | | | | | | | | | | | | |
| Glu 1 | Pro | Gly | Leu | Pro 5 | Pro | Gly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | | |
| Val | Asn | Asp | Glu 20 | Ala | His | Pro | Trp | Lys 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | | |
| Arg | Gly | Pro 35 | Сув | Pro | Gly | Leu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | | |
| Pro | Arg 50 | Asn | Gly | Val | Ala | Thr 55 | Pro | Ala | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | | |
| Glu 65 | Gly | Phe | Asn | Phe | Asp | Asn | Gln | Ala | Ala | Ile 75 | Phe | Ala | Thr | Tyr | Ala 80 | | |
| Ala | His | Leu | Val | Asp 85 | Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | | |
| Arg | Lys | Thr | Arg 100 | | Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | | |
| Gly | Gly | Leu 115 | Asn | Glu | His | Gly | Thr | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | | |
| Arg | Gly 130 | Asp | Ala | Phe | Phe | Gly 135 | Asn | Asn | His | Asp | Phe | Asn | Glu | Thr | Leu | | |
| Phe 145 | Glu | Gln | Leu | Val | Asp | Tyr | Ser | Asn | Arg | Phe | Gly | Gly | Gly | Lys | Tyr 160 | | |
| Asn | Leu | Thr | Val | Ala 165 | Gly | Glu | Leu | Arg | Phe | Lys | Arg | Ile | Gln | Asp 175 | Ser | | |
| Ile | Ala | Thr | Asn 180 | Pro | Asn | Phe | Ser | Phe | Val | Asp | Phe | Arg | Phe | Ser | Thr | | |
| Ala | Tyr | Gly 195 | | Thr | Thr | Phe | Pro | | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | | |
| Arg | Asp | | Gly | Gln | Leu | Asp | Met | Asp | Ala | Ala | Arg | | Phe | Phe | Gln | | |

| Phe 225 | Ser | Arg | Met | Pro | Asp 230 | Asp | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | |
|---|--|--|--|---------------------|--------------|------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-----|
| Asp | Thr | Gly | Val | Glu 245 | Val | Val | Val | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | |
| Lys | Asn | Val | Gly 260 | Lys | Ile | Asn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | |
| Asp | Phe | Ser 275 | Thr | Pro | CAa | Leu | Met 280 | Tyr | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | |
| Val | Lys 290 | Ser | Leu | Tyr | Pro | Asn 295 | Pro | Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | |
| Asn 305 | Thr | Asn | Leu | Asp | Phe 310 | Leu | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Cha | Thr 320 | |
| Gln | Val | Phe | Pro | Tyr 325 | Gly | Arg | Asp | | | | | | | | | |
| <211 <212 <213 <220 <223 <220 <221 <221 <221 <221 |) > FE 3 > OT 0 > FE 1 > NA 2 > LO 0 > FE 1 > NA | ENGTH PE: CGANI EATUF CHER ME/F CATUF AME/F | H: 11 DNA SM: SE: INFO SE: CEY: ON: SE: CEY: | Arti DRMAT CDS (1). | ION: .(11 | | uO Vâ | | nt wi | th v | vild | sigr | nal p | oepti | de | |
| < 400 |)> SE | EQUEN | ICE : | 39 | | | | | | | | | | | | |
| | | | | | | ttc Phe | | | | | | | | | | 48 |
| - | _ | | | - | | gcc Ala | | _ | - | | | _ | _ | _ | _ | 96 |
| | | | | | | aca Thr -5 | | | | | | | | | | 144 |
| | | | | | | agc Ser | | | | | | | | | | 192 |
| | | | _ | _ | | cga Arg | | | _ | | _ | | | _ | | 240 |
| | | | | _ | _ | tct Ser | | | | | _ | _ | | | - | 288 |
| | | | | | | ata Ile 60 | | | | | | | | | | 336 |
| | | | | | | ttc Phe | | | | | | | | | | 384 |
| | | | | | | ttg Leu | | | | | | | | | | 432 |
| | | | | | | ccc Pro | | | | | | | | | | 480 |

| | | | | | | | | | | | | con | CTU | uea | | |
|------------------------------|--|----------------------------------|-----------------------------|-----------|-----|-----|------------|-----|-----|-----------|-----|------------|------------|-----|----------|------|
| | | | 105 | | | | | 110 | | | | | 115 | | | |
| | ggc Gly | | | - | | - | - | - | _ | | - | | - | - | | 528 |
| | ggc Gly 135 | | | | _ | | | | _ | | | _ | _ | _ | _ | 576 |
| - | tac Tyr | _ | | - | | | | | | | | | | _ | | 624 |
| | gag Glu | | _ | | _ | _ | | | _ | | | | | | | 672 |
| | ttc Phe | | | - | _ | | | | | | - | | | | | 720 |
| | ttc Phe | | | | | | | _ | | _ | | _ | _ | | _ | 768 |
| | gat Asp 215 | _ | - | - | _ | | _ | | | | | - | _ | _ | | 816 |
| - | gat Asp | | | _ | - | | _ | _ | _ | _ | - | | | - | | 864 |
| _ | gtt Val | _ | _ | - | | | _ | _ | | | | | _ | | _ | 912 |
| | aac Asn | - | | | - | - | | | | | - | | | | | 960 |
| _ | ttg Leu | _ | | | | | - | | | _ | - | _ | | | | 1008 |
| _ | aat Asn 295 | _ | _ | | _ | | _ | | _ | | | _ | | | _ | 1056 |
| | tta Leu | | | | | | | | | | | | | | | 1104 |
| | cga Arg | - | | | | | | | | | | | | | | 1113 |
| <21: <21: <21: <22: | 0 > SI 1 > LI 2 > T 3 > OI 0 > FI 3 > O | ENGTI (PE : RGAN: EATUI | H: 3' PRT [SM: RE: | 71 Art | | | - | | | truc | t | | | | | |
| < 40 | 0> SI | EQUEI | ICE : | 40 | | | | | | | | | | | | |
| Met | Lys | Tyr | Phe | Pro | Leu | Phe | Pro | Thr | | Val | Phe | Ala | Ala -30 | Arg | Val | |
| Val | Ala | Phe -25 | Pro | Ala | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg -1 | | Pro | Gly | Leu | Pro 5 | |

His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu Pro Arg Asn Gly Val Ala Thr Pro Ala Gln Ile Ile Asn Ala Val Gln Glu Gly Phe Asn Phe Asp Asn Gln Ala Ala Ile Phe Ala Thr Tyr Ala Ala His Leu Val Asp Gly Asn Leu Ile Thr Asp Leu Leu Ser Ile Gly Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val Gly Gly Leu Asn Glu 110 His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr Arg Gly Asp Ala Phe 120 125 130Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu Phe Glu Gln Leu Val 140 Asp Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr Asn Leu Thr Val Ala 155 160 Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser Ile Ala Thr Asn Pro 170 Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln 205 Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro 220 $\hbox{Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu} \\$ 235 Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro 270 Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr Gly Arg Asp <210> SEQ ID NO 41 <211> LENGTH: 1113 <212> TYPE: DNA <213 > ORGANISM: Artificial Sequence <220> FEATURE: <223> OTHER INFORMATION: SoLo variant with modified signal peptide <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (1)..(1113) <220> FEATURE: <221> NAME/KEY: mat_peptide <222> LOCATION: (130) .. (1113)

Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu Val Asn Asp Glu Ala

| | | | | | | | | | | | - | con | tin | ued | | | | |
|------|-------------------|------|-------|----|---|---|---|---|---|---|-----|-----|-----|-----|---|-----|--|--|
| < 40 | 0> SI | EQUE | ICE : | 41 | | | | | | | | | | | | | | |
| _ | aaa Lys | | | | _ | | | | _ | _ | | _ | | | - | 48 | | |
| _ | gct Ala | | | _ | | _ | | _ | _ | | | _ | _ | _ | _ | 96 | | |
| _ | gac Asp -10 | - | | | | | | | - | - | Glu | | | | | 144 | | |
| | ggt Gly | | | | | _ | | _ | _ | _ | | | _ | | _ | 192 | | |
| | cca Pro | | _ | _ | | _ | | | _ | | _ | | | _ | | 240 | | |
| | ctc Leu | | | _ | _ | | | | | | _ | _ | | | _ | 288 | | |
| - | acc Thr 55 | _ | | | | | | | - | _ | - | | | | | 336 | | |
| _ | aat Asn | | _ | _ | | | _ | | | | _ | | | | _ | 384 | | |
| | aat Asn | | | _ | _ | _ | _ | - | | | _ | _ | _ | | | 432 | | |
| | ggg Gly | | - | | | | | - | | - | | | | | | 480 | | |
| | ggc Gly | | | - | | _ | - | - | _ | | _ | | - | _ | | 528 | | |
| | ggc Gly 135 | | | | _ | | | | _ | | | _ | _ | _ | - | 576 | | |
| _ | tac Tyr | - | | _ | | | | | | | | | | _ | | 624 | | |
| | gag Glu | | _ | | _ | _ | | | _ | | | | | | | 672 | | |
| | ttc Phe | | | _ | _ | | | | | | _ | | - | | | 720 | | |
| | ttc Phe | | | | | | | | | | | | | | | 768 | | |
| | gat Asp 215 | _ | _ | - | - | | _ | | | | | _ | _ | _ | | 816 | | |
| - | gat Asp | | | - | - | | _ | _ | - | - | - | | | _ | | 864 | | |
| - | gtt Val | _ | _ | - | | | _ | _ | | | | | _ | | _ | 912 | | |
| | | | | | | | | | | - | • | | | • | - | | | |

| | | | | | | | | | | | - | con | tin | ued | | |
|---------------------------------|--|---|-------------------------------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------|
| | | | | 250 | | | | | 255 | | | | | 260 | | |
| | aac Asn | | | | | | | | | | | | | | | |
| _ | ttg Leu | _ | | | | | _ | | | _ | _ | _ | | | | |
| _ | aat Asn 295 | _ | _ | | _ | | _ | | _ | | | _ | | | _ | |
| | tta Leu | | _ | | _ | _ | _ | | _ | | _ | _ | | | | |
| | cga Arg | _ | | | | | | | | | | | | | | 1113 |
| <21 <21 <21 <22 <22 | 0 > SI 1 > LI 2 > TI 3 > OI 0 > FI 3 > OI | ENGTH YPE: RGANI EATUR THER | H: 3' PRT ISM: RE: INFO | 71 Art ORMA | | | _ | | Const | truc | t | | | | | |
| | 0> SI | | | | Lou | Dho | Dro | Thr | Lou | 7707 | Tr rac | 71. | Wal. | Cl | 7707 | |
| мес | Lys | ıyr | -40 | PIO | ьеи | Pne | PIO | -35 | ьеи | val | Tyr | Ala | -30 | GIY | Val | |
| Val | Ala | Phe -25 | Pro | Asp | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg -1 | | Pro | Gly | Leu | Pro 5 | |
| | Gly | | | 10 | | | | | 15 | | | | | 20 | | |
| | Pro | _ | 25 | | | | | 30 | | | _ | | 35 | | | |
| Ī | Leu | 40 | | | | | 45 | | | | | 50 | | | | |
| | Thr 55 | | | | | 60 | | | | | 65 | | | | | |
| Asp 70 | Asn | GIn | Ala | Ala | 11e 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
| Gly | Asn | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | |
| Thr | Gly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | |
| His | Gly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | |
| Asp 150 | Tyr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly | Glu | Leu | Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn | Phe | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Ser | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| | | | | | | | | | | | | | | | | |

<210> SEQ ID NO 47

```
Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg Arg Asp Asp Gly Gln
                           205
Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln Phe Ser Arg Met Pro
                       220
Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser Asp Thr Gly Val Glu
                  235
Val Val Val Gln Ala His Pro Met Gln Pro Gly Lys Asn Val Gly Lys
Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser Asp Phe Ser Thr Pro
Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr Val Lys Ser Leu Tyr
Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu Asn Thr Asn Leu Asp
Phe Leu Phe Gln Gly Val Ala Ala Gly Cys Thr Gln Val Phe Pro Tyr
                   315
Gly Arg Asp
<210> SEO ID NO 43
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MJaWa Fw primer
<400> SEQUENCE: 43
gcgcattcaa gactccattg
                                                                      20
<210> SEQ ID NO 44
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: MJaWa Rev primer
<400> SEQUENCE: 44
gatcttgccg acatttttc c
                                                                      21
<210> SEQ ID NO 45
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HFJaWa Fw primer
<400> SEQUENCE: 45
caggeteate etatgeagee e
<210> SEQ ID NO 46
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HFJaWa Rev primer
<400> SEQUENCE: 46
                                                                      2.4
caaaggagaa attggggttg gtcg
```

```
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: VHG R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (51)..(51)
<223> OTHER INFORMATION: d is a or g or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (52)..(52)
<223> OTHER INFORMATION: b is g or c or t
<400> SEQUENCE: 47
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggcc dbgattgcgg
<210> SEQ ID NO 48
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: NDT R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (51)..(51)
<223> OTHER INFORMATION: h is a or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (52)..(52)
<223> OTHER INFORMATION: n is a or g or c or t
<400> SEQUENCE: 48
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggca hngattgcgg
                                                                       60
                                                                       61
<210> SEQ ID NO 49
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: TGG R primer
<400> SEQUENCE: 49
gcaagtccgt aatgagattg ccgtccacaa ggtgggccgc atatgtggcc cagattgcgg
<210> SEQ ID NO 50
<211> LENGTH: 40
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HF F primer
<400> SEQUENCE: 50
                                                                       40
gcggcccacc ttgtggacgg caatctcatt acggacttgc
<210> SEQ ID NO 51
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: S191 VHG R primer
<220> FEATURE:
```

```
<221> NAME/KEY: misc_feature
<222> LOCATION: (50)..(50)
<223> OTHER INFORMATION: d is a or g or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (51)..(51)
<223> OTHER INFORMATION: b is g or c or t
<400> SEQUENCE: 51
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtcd bgaacctaaa
<210> SEQ ID NO 52
<211> LENGTH: 61
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: NDT R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (50)..(50)
<223> OTHER INFORMATION: h is a or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (51)..(51)
<223> OTHER INFORMATION: h is a or g or c or t
<400> SEQUENCE: 52
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtah ngaacctaaa
                                                                       60
                                                                       61
<210> SEQ ID NO 53
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: S191 TGG R primer
<400> SEQUENCE: 53
cccatccaca aaaagattcg cggggaaggt ggtctcgccg taagcagtcc agaacctaaa
                                                                       60
                                                                       61
<210> SEQ ID NO 54
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: HF F-RMLC primer
<400> SEQUENCE: 54
cggcgagacc accttccccg cgaatctttt tgtggatggg
<210> SEQ ID NO 55
<211> LENGTH: 32
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F69 R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (22)..(22)
<223> OTHER INFORMATION: m is a or c
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (23)..(23)
```

```
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
<223> OTHER INFORMATION: n is a or g or c or t
<400> SEQUENCE: 55
gaagattgcg gcttgattgt cmnnattgaa tc
                                                                       32
<210> SEQ ID NO 56
<211> LENGTH: 32
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F69 F primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25)..(25)
<223> OTHER INFORMATION: k is g or t
<400> SEQUENCE: 56
cgcggttcag gaaggattca atnnkgacaa tc
                                                                       32
<210> SEQ ID NO 57
<211> LENGTH: 31
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F121 R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (19) .. (19)
<223> OTHER INFORMATION: m is a or c
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (20)..(20)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(21)
<223> OTHER INFORMATION: n is a or g or c or t
<400> SEQUENCE: 57
catactggcg tcgccttcmn nggtgccatg c
                                                                       31
<210> SEQ ID NO 58
<211> LENGTH: 31
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F121 F primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (22)..(22)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
```

```
<223> OTHER INFORMATION: k is g or t
<400> SEQUENCE: 58
ggactcaatg agcatggcac cnnkgaaggc g
                                                                       31
<210> SEQ ID NO 59
<211> LENGTH: 32
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F199 R primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (20)..(20)
<223> OTHER INFORMATION: m is a or c
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(21)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (22)..(22)
<223> OTHER INFORMATION: n is a or g or c or t
<400> SEQUENCE: 59
                                                                       32
ccacaaaaag attcgcgggm nnggtggtct cg
<210> SEO ID NO 60
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: F199 F primer
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (24)..(24)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25)..(25)
<223> OTHER INFORMATION: n is a or g or c or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (26)..(26)
<223> OTHER INFORMATION: k is g or t
<400> SEQUENCE: 60
                                                                       32
ctactgctta cggcgagacc accnnkcccg cg
<210> SEQ ID NO 61
<211> LENGTH: 987
<212> TYPE: DNA
<213 > ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Wt-SoLo variant without signal peptide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) .. (987)
<400> SEQUENCE: 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
               5
                                    10
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
           2.0
                                25
cgt gga cct tgc cct ggt ctc aat act ctg gca tct cac ggg tac ctc
```

| Arg Cly Pro Cys Pro Cly Leu Aan Thr Leu Ala Ser His Cly Tyr Leu 45 Cog aga act ggs gtt gca acc cog gtg cas ata ata aca gog gtt cag Pro Arg Aan Cly Val Ala Thr Pro Val Gin Tie Ite Aan Ala Val Gin So So Gas gga cite act tic gee act can goe gos git tie goe act tat gog Qiu Cly Leu Aan Pho Amp Aen Gin Ala Ala Val Pho Ala Thr Tyr Ala 65 Go Go Cac cit gtg gae ggc cat cat cat acg gae tit ge goe act gag acg acg gae gee atg Go Cac cit gtg gae ggc cat cat cat acg gae tit ge goe acg Go Cag aag acg og cot cat ggg cot get gee acc aca coc coc got to gtt Arg Lys Thr Ang Leu Thr Gly Pro Anp Pro Pro Pro Pro Ala Ser Val 105 105 Gg Gg Gg Gg Gg Gg Gg G |
|---|
| pro Aig Aan Gly Val Ala Thr Pro Val Gln He He Aen Ala Val Gln 50 gaa gga ctc aat ttc gac aat caa gec gca gtc ttc gcc aca tat geg Gls Gly Leu Aen Phe Amp Aan Gln Ala Ala Val Phe Ala Thr Tyr Ala 75 gec cac ctt gtg gac gga at ctc att acg gac ttc gtg agc atc gga Ala His Leu Val Asp Gly Amn Leu He Thr Amp Leu Leu Ser He Gly 80 ggc aag acg cgg ctc act ggg cct gat cca cac ccc ccc gct tcc gtt 80 ggt aga ctc aat gag cat ggc cat ctc gan ggc gac gcc act gas gac gac gac atc gar Ala His Leu Val Asp Gly Amn Leu He Thr Amp Leu Leu Ser He Gly 80 ggt gac ctc aat gag cat ggc act ctc gan ggc gac gcc act atc acc gly 81 ggt gac ctc aat gag cat ggc act tcc gtd 100 ggt gac ctc aat gag cat ggc act tcc gtd 110 ggt gac ctc aat gag cat ggc acc ttc gan ggc gac gcc act atc acc Gly Gly Leu Aen Glu His Gly Thr Phe Glu Gly Amp Ala Ser Met Thr 110 ggt gac ctc aat gag cat ggc acc cac cac gat ttc aat gag acg ctc Arg Gly Amp Ala Phe Phe Gly Amn Amn His Ram Phe Amn Glu Thr Leu 130 131 ttc gan cag ttg gt gac tac agc cac ttg sgs gga gga aat acc Phe Glu Gln Leu Val Amp Tyr Ser Amn Arg Phe Gly Gly Gly Lye Tyr 145 ttc gan cag ttg gt gac tac agc ctc 155 aat ctt acc gtc gag gag gdc cc gt ttc aag gcc att cac agc atc cc Amn Leu Thr Ala Gly Glu Leu Arg Phe Lye Arg He Gln Amp Ser 175 att gcg acc acc ccc act ttc ccc gcg aat ctt ttt gtc gac tgc gcc Ala Tyr Gly Glu Thr Thr Phe For Ala Amn Leu Phe Val Amp Gly Arg 185 ggt tac ggc gac acc acc ttc ccc gcg aat ctt ttt gtc gat ggc gc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Amn Leu Phe Val Amp Gly Arg 210 ggc aca gga gto gag gta gta gta tac ac ga gt gt gtc gac agg agt ttt tc cac Arg Ang Amp Amp Gly Gln Leu Arg Phe Pro Pro Pro Pro Amp Pro For Pro Arg Ser 225 ggc aca gga gto gag gta gta gta acc acc gtc gcc agc act acc tct ctc Arg Amn Thr Amn Pro Amp Pro |
| Glu Gily Leu Aen Phe App Aen Gln Ala Ala Val Phe Ala Thr Tyr Ala 65 75 65 75 65 76 65 76 65 77 65 66 76 67 76 67 77 68 80 80 80 80 80 80 80 80 80 80 80 80 80 |
| Ala His Leu Val Amp Oly Am Leu Ile Thr Amp Leu Leu Ser Ile Oly 85 cgo aag acg cgd ctc act ggg cct gat cca cca ccc ccc gct tcc gtt Arg Lyo Thr Arg Leu Thr Gly Pro Amp Pro Pro Pro Pro Pro Pro Ala Ser Val 100 ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Amn Glu His Gly Thr Phe Glu Gly Amp Ala Ser Met Thr 115 cga ggt gac gca ttc ttt ggc aac aac cac gat ttc aat gag acg cct Arg Gly Amp Ala Phe Phe Gly Amn Amn His Amp Phe Amn Glu Thr Leu 130 ttc gaa cag ttg gtt gac tac agc aac cga ttt gag gga gga gac aac tac Phe Glu Gln Leu Val Amp Tyr Ser Amn Arg Phe Gly Gly Gly Lyo Tyr 145 aat ctt acc gtc gcg ggg gga ctc cgt ttc aag gac gat at ag acc cg acc gtt gag gga gga cac acc cac gat ttt gag gga gga aaa tac Phe Glu Gln Leu Val Amp Tyr Ser Amn Arg Phe Gly Gly Gly Lyo Tyr 165 aat ctt acc gtc gcg ggg gga ctc cgt ttc aag gcc att caa gac tcc Amn Leu Thr Val Ala Gly Glu Lau Arg Phe Lyo Ang Ile Gln Amp Ser 175 att gga acc aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act Ile Ala Thr Amn Phe Amn Phe Ser Phe Val Amp Phe Arg Phe Ser Thr 180 ggct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat tgg ttc tct act 195 agg gag gac gac gag cac acc ttc ccc gcg aat ctt ttt gtg gat ttt tcc aa Arg Amp Amp Amp Gly Gln Leu Amp Met Amp Ala Ala Amp Ser Phe Phe Gln 210 200 agg gac gac ggc gag cta gat atg atg atg cac gac gac ga agt ttt ttc caa Arg Amp |
| Arg Lye Thr Arg Leu Thr Gly Pro App Pro Pro Pro Pro Pro Ala Ser Val 100 ggt gga ctc aat gag cat ggc acc ttc gaa ggc gac gcc agt atg acc Gly Gly Leu Aen Glu His Gly Thr Phe Glu Gly Aep Ala Ser Met Thr 115 cga ggt gac gca ttc ttt ggc aac acc cac gat ttc aat gag acg ctc Arg Gly Aep Ala Phe Phe Gly Aan Aen His Aep Phe Aen Glu Thr Leu 130 135 135 136 ttc gga cag ttg gtt gac tac agc aac cga ttt gga gga gga aaa tac Phe Glu Gln Leu Val Aep Tyr Ser Aen Arg Phe Gly Gly Gly Lye Tyr 145 165 175 175 aat cac gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Aen Leu Thr Val Ala Gly Glu Leu Arg Phe Lye Arg Ile Gln Aep Ser 165 165 176 att gga cac aac ccc aat ttc tcc ttt gtt gac ttt agg ttc tct act 110 Ala Thr Aen Pro Aen Phe Ser Phe Val Aep Phe Arg Phe Ser Thr 180 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttg gga tgg agg agg gac gac gac gac gac acc ctc acc gcg aat ctt ttg gga tgg agg agg gac gac gac gac gac acc ttc ccc gcg aat ctt ttg gga tgg agg gac gac gac gac gac acc ttc ccc gcg aat ctt ttg gga tgg agg gac gac gac gac gac acc acc ttc ccc gcg aat ctt ttg gga tgg agg gac gac gac gac gac acc acc ttc ccc gcg aat ctt ttg gga tgg agg gac gac gac gac gac acc acc ttc ccc gcg aat ctt ttg gga tgg agg acc gac gac gac gac acc acc ttc ccc gcg aat ctt ttg gga tag gg gac gac gac gac gac acc acc ttc ccc gcg aat ctt ttc gtg gat gag gac gac gac gac gac acc acc ttc ccc gcg acc gag agt ttt tc caa Arg Aep Aep Gly Gln Leu Aep Met Aep Ala Ala Arg Ser Phe Phe Gln 210 215 216 227 228 228 229 229 235 235 240 245 245 246 247 248 249 249 240 240 240 240 240 240 240 240 240 240 |
| Gly Gly Leu Asm Glu Hie Gly Thr Phe Glu Gly Asp Ala Ser Met Thr 115 cga ggt gas gas the tett ggs aac aac cac gat the aat gag acg ctc Arg Gly Asp Ala Phe Phe Gly Asm Asm His Asp Phe Asm Glu Thr Leu 130 tte gaa cag ttg gtt gas tac ags aac cgs ttt gag ggs ggs aas tac Phe Glu Gln Leu Val Asp Tyr Ser Asm Arg Phe Gly Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att cac gac tcc Asm Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 166 att gog acc aac ccc aat ttc tec ttt gtt gac ttt agg ttc tc act 170 att gog acc aac ccc aat tec tec ttt gtt gac ttt agg ttc tc tc 186 att ggg acg aga acc acc ttc ccc gcg aat ctt ttg tgg ag ggg cgc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asm Leu Phe Val Asp Gly Arg 195 200 agg gac gac gac gac gat at atg gat gct gca cgg agt ttt tc caa Arg Asp Asp Asp Gly Glu Leu Arg May Ala Ala Arg Ser Phe Phe Gln 210 220 ttc ag gc gad acc acc ttc ccc gcg acc acc aga ttt ttc caa Arg Asp Asp Asp Gly Glu Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 221 ttc agc gt atg ctt gac gat ttt ttc tcc cgc gca ccc agc ccq aga act 122 ttc agc gt atg cct gac gat ttt ttc tcc cgc gca ccc agc ccq aga 172 ttc agc gt atg ctt gac gat ttt ttc tcc cgc gca ccc agc ccq aga 235 326 327 328 329 320 320 322 321 322 325 326 326 327 328 328 329 320 320 321 322 325 326 326 327 328 328 329 320 320 321 320 320 321 320 320 |
| Arg Gly Aep Ala Phe Phe Cly Aen Aen Hie Aep Phe Aen Glu Thr Leu 130 130 130 135 135 136 137 138 139 130 130 130 130 130 130 130 |
| Phe Glu Glu Cheu Val App Tyr Ser Asn Arg Phe Gly Gly Gly Lys Tyr 145 aat ctt acc gtc gcg ggg gag ctc cgt ttc aag cgc att caa gac tcc Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln App Ser 165 170 att gcg acc acc caa ttc tcc ttt gtt gac ttt agt tcc tct act Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr 180 gct tac ggc gag acc acc ttc ccc gcg aat ctt ttt gtg gat ggc gcc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 200 agg gac gac gac gac ta gat atg gat gcc gcg agt ttt ttc caa Arg Anp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc cc aga agt Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 225 aga aca gag gtc gag gta gtt ata cag gct cat cct atg cag ccc aga agt Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc aac agc tac acc gtc gac caa acc tct ct Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 265 gc cag ttt cc acc ccc tgc ttg atg tac gas aat tc gtc aac aac acc ct atc Arg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gc aag tcc ct tac ccg acc gag gtc ag ttc ctc cag gag gtc cag cct gaa gcc ctt tcc acc ccc tgc ttg atg tac gag aac acc acc ctc gcs ga gtc ag tcc acc acc acc ctc Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gc aag tcc acc tcc tac ccg acc gag gtc ga gtc gcc gcc gcg ga ttg acc Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Ala Cly Cys Thr 305 aat acc gat ttc cca tac ggg cg ag tt tg Gln Val Phe Pro Tyr Gly Arg Asp |
| Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser 165 165 170 175 175 165 175 175 175 175 175 175 175 175 175 17 |
| Ile Åla Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr 180 get tac ggc gag acc acc ttc coc gcg aat ctt ttt gtg gat ggg cgc Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 195 agg gac gac ggc cag cta gat atg gat ggt gca cgg agt ttt ttc caa Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc cg aga agt ttc ttc agc cgt atg ctt gac ggt gtt cac agc ccg aga agt pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 225 ggc aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga acc gga gtt gal Val Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct Arg Asp Nan Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc gac aca aca tcc tct Arg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aac ata acc acc acc acc ttc cac acc ac |
| Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg 200 agg gac gac ggc cag cta gat atg gat gct gca cgg agt ttt ttc caa Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 225 ggc aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc ac ac agc tca cac gtc gac cac act cct tc Arg Asp Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 260 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc gac aca act acc acc acc Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc ttc ttc cag gga gtc gc gcc gct gga tgt acc Asn Chan Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat tga Gln Val Phe Pro Tyr Gly Arg Asp |
| Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln 210 ttc agc cgt atg cct gac gat ttc ttc cgc gca ccc agc ccg aga agt Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 235 ggc aca gga gtc gag gta gtt at ata cag gct cat cct atg cag ccc gga Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc aca acc gtc gac cca aca tcc tct Arg Asp Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 265 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt 285 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt 291 gat ttt Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc Asn Thr Asn Leu Asp Phe |
| Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser 235 ggc aca gga gtc gag gta gtt ata cag gct cat cct atg cag ccc gga Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 270 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc Asn Thr Asn Leu Asp Phe Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat tga Gln Val Phe Pro Tyr Gly Arg Asp |
| Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly 245 aga aat gtc ggc aag atc aac agc tac acc gtc gac cca aca tcc tct Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 260 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc gac aac ata acg 864 Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 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 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 cag gtc ttc cca tac ggg cga gat tga Glu Val Phe Pro Tyr Gly Arg Asp |
| Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser 265 gac ttt tcc acc ccc tgc ttg atg tac gag aaa ttc gtc aac ata acg Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc ttc tc cag gga gtc gcc gct gga tgt acc Asn Thr Asn Leu Asp Phe Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat tga Gln Val Phe Pro Tyr Gly Arg Asp |
| Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr 275 gtc aag tca ctc tac ccg aat ccg acg gtg cag ctt cgc aaa gcc ctt Val Lys Ser Leu Tyr Pro Asn Pro Thr Val Gln Leu Arg Lys Ala Leu 290 aat acg aat ctc gat ttc ttc ttc cag gga gtc gcc gct gga tgt acc Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr 305 cag gtc ttc cca tac ggg cga gat tga Gln Val Phe Pro Tyr Gly Arg Asp 987 |
| 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 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 |
| 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 |
| Gln Val Phe Pro Tyr Gly Arg Asp |
| |

```
<210> SEQ ID NO 62
<211> LENGTH: 328
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic Construct
<400> SEQUENCE: 62
Glu Pro Gly Leu Pro Pro Gly Pro Leu Glu Asn Ser Ser Ala Lys Leu
Val Asn Asp Glu Ala His Pro Trp Lys Pro Leu Arg Pro Gly Asp Ile
\hbox{Arg Gly Pro Cys Pro Gly Leu Asn Thr Leu Ala Ser His Gly Tyr Leu}\\
Pro Arg Asn Gly Val Ala Thr Pro Val Gln Ile Ile Asn Ala Val Gln
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
Arg Lys Thr Arg Leu Thr Gly Pro Asp Pro Pro Pro Pro Ala Ser Val
                   105
Gly Gly Leu Asn Glu His Gly Thr Phe Glu Gly Asp Ala Ser Met Thr
                          120
\hbox{Arg Gly Asp Ala Phe Phe Gly Asn Asn His Asp Phe Asn Glu Thr Leu}\\
                      135
Phe Glu Gln Leu Val Asp Tyr Ser Asn Arg Phe Gly Gly Lys Tyr
                150
                                       155
Asn Leu Thr Val Ala Gly Glu Leu Arg Phe Lys Arg Ile Gln Asp Ser
                                  170
Ile Ala Thr Asn Pro Asn Phe Ser Phe Val Asp Phe Arg Phe Ser Thr
                               185
Ala Tyr Gly Glu Thr Thr Phe Pro Ala Asn Leu Phe Val Asp Gly Arg
Arg Asp Asp Gly Gln Leu Asp Met Asp Ala Ala Arg Ser Phe Phe Gln
Phe Ser Arg Met Pro Asp Asp Phe Phe Arg Ala Pro Ser Pro Arg Ser
Gly Thr Gly Val Glu Val Val Ile Gln Ala His Pro Met Gln Pro Gly
Arg Asn Val Gly Lys Ile Asn Ser Tyr Thr Val Asp Pro Thr Ser Ser
Asp Phe Ser Thr Pro Cys Leu Met Tyr Glu Lys Phe Val Asn Ile Thr
                 280
{\tt Val\ Lys\ Ser\ Leu\ Tyr\ Pro\ Asn\ Pro\ Thr\ Val\ Gln\ Leu\ Arg\ Lys\ Ala\ Leu}
                       295
Asn Thr Asn Leu Asp Phe Phe Phe Gln Gly Val Ala Ala Gly Cys Thr
                 310
                                       315
Gln Val Phe Pro Tyr Gly Arg Asp
             325
```

<210> SEQ ID NO 63 <211> LENGTH: 1116 <212> TYPE: DNA

| _ | | | | | | | | | | | | | | | | | |
|-------|------------------|-------|-----------|------------|-------|------------|----------|-------|------------|------|------------|-------|------------|------------|----------|------|--|
| | 3 > OF 0 > FF | | | Art: | ific: | ial : | Seque | ence | | | | | | | | | |
| < 22 | 3 > 0 | THER | INF | ORMA' | TION | : Wt | -SoLo | o va: | riant | c wi | th w | ild : | signa | al pe | eptid | le | |
| | 0> FI 1> NA | | | CDS | | | | | | | | | | | | | |
| <22 | 2 > LO | CAT: | ON: | | (1 | 113) | | | | | | | | | | | |
| | 0> FI 1> NA | | | mat_ | _pept | ide | | | | | | | | | | | |
| <22 | 2 > L(| CAT: | ION: | (13) | 0) | (111 | 6) | | | | | | | | | | |
| < 40 | 0> SI | EQUEI | ICE: | 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 -30 | Gly | Val | | |
| | | | | | | | | | | | | | | | | | |
| - | gct Ala | | | - | | _ | | _ | - | | | _ | _ | _ | _ | 96 | |
| | | -25 | | | | | -20 | | | | | -15 | | | | | |
| ttg | gac | gct | ata | atc | cca | aca | ctc | gag | gcc | cga | gag | cca | gga | tta | cct | 144 | |
| Leu | Asp -10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg | | Pro | Gly | Leu | Pro 5 | | |
| | | a | a+- | ~ | a | | . | ~ · · | | | | | ~- · | ~ | | 100 | |
| | ggt Gly | | | | | _ | | _ | _ | _ | | | _ | | _ | 192 | |
| | | | | 10 | | | | | 15 | | | | | 20 | | | |
| | cca | | _ | _ | | _ | | | _ | | _ | | | _ | | 240 | |
| His | Pro | Trp | Lуs 25 | Pro | Leu | Arg | Pro | GIY | Asp | Ile | Arg | GIY | Pro 35 | Cys | Pro | | |
| aat | ata | 225 | aat | ata | aa2 | tat | a2a | aaa | tag | ata | aaa | 202 | 225 | aaa | at t | 288 | |
| | ctc Leu | | | _ | _ | | | | | | _ | _ | | | _ | 200 | |
| | | 40 | | | | | 45 | | | | | 50 | | | | | |
| - | acc | _ | | | | | | | _ | _ | _ | | | | | 336 | |
| АІА | Thr 55 | Pro | vai | GIN | 11e | 60 | Asn | Ala | vai | GIN | 65 | GIY | Leu | Asn | Pne | | |
| gac | aat | caa | acc | aca | atc | ttc | acc | aca | tat | aca | acc | cac | ctt | ata | gac | 384 | |
| Asp | Asn | | - | - | Val | | - | | | Āla | _ | | | | Asp | | |
| 70 | | | | | 75 | | | | | 80 | | | | | 85 | | |
| - | aat Asn | | | _ | _ | _ | _ | _ | | | _ | _ | _ | | | 432 | |
| Cly | non | БСС | 110 | 90 | пър | пси | Leu | DCI | 95 | GLY | mg | цур | **** | 100 | шеш | | |
| act | ggg | cct | gat | cca | cca | ccc | ccc | gct | tcc | gtt | ggt | gga | ctc | aat | gag | 480 | |
| | Gly | | Asp | | | | | Āla | | - | | | Leu | | | | |
| | | | 105 | | | | | 110 | | | | | 115 | | | | |
| | ggc Gly | | | | | | | | | | | | | | | 528 | |
| | 4 | 120 | | | 4 | r | 125 | · | | · | , | 130 | - | | | | |
| ttt | ggc | aac | aac | cac | gat | ttc | aat | gag | acg | ctc | ttc | gaa | cag | ttg | gtt | 576 | |
| Phe | Gly 135 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | | |
| | | | | _ | | | | | | _ | | | | | | 66.4 | |
| - | tac Tyr | - | | - | | | | | | | | | | - | | 624 | |
| 150 | | | | | 155 | | | | | 160 | | | | | 165 | | |
| 999 | gag | ctc | cgt | ttc | aag | cgc | att | caa | gac | tcc | att | gcg | acc | aac | ccc | 672 | |
| Gly | Glu | Leu | Arg | Phe 170 | ГЛа | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | | |
| | | | | 1/0 | | | | | 1/3 | | | | | 100 | | | |
| | ttc Phe | | | - | - | | | | | | - | | | | | 720 | |
| -1211 | | | 185 | | | 10 | 9 | 190 | | | | -1- | 195 | J_ u | | | |
| acc | ttc | ccc | gca | aat | ctt | ttt | gta | gat | gqa | cqc | aqq | gac | gac | gqc | caq | 768 | |
| | Phe | Pro | | | | | Val | _ | | _ | | Āsp | - | | _ | | |
| | | 200 | | | | | 205 | | | | | 210 | | | | | |

| _ | | | | | | | | | | | | con | tin' | ued _ | | | |
|--|---|--|--|---|---|---------------------------------------|--|--|--|---|--------------------------------------|---|--|---------------------------------------|--------------------------------------|------|--|
| cta g Leu A 2 | | _ | - | - | _ | | _ | | | | | _ | _ | _ | | 816 | |
| gac g Asp A 230 | | | | _ | _ | | _ | _ | _ | _ | | | | _ | | 864 | |
| gta g Val V | | | _ | - | | | _ | _ | | | _ | | - | | _ | 912 | |
| atc a Ile A | | _ | | | _ | _ | | | | | _ | | | | | 960 | |
| tgc t Cys L | eu | | | | | | | | | | | | | | | 1008 | |
| ccg a Pro A 2 | | _ | _ | | _ | | _ | | _ | | | _ | | | _ | 1056 | |
| ttc t Phe P 310 | | | _ | | _ | _ | - | | _ | | _ | _ | | | | 1104 | |
| ggg c Gly A | _ | - | tga | | | | | | | | | | | | | 1116 | |
| <210><211><212><213> | · TY | PE: | PRT | | ifia: | ial (| Secure | ence | | | | | | | | | |
| <211> | OR FE | PE: GANI ATUR HER | PRT SM: E: INFO | Art: ORMA: | | | _ | | Const | ruci | t | | | | | | |
| <211><212><213><223> | TY OR FE OT | PE: GANI ATUR HER | PRT SM: E: INFO | Art: DRMA: | rion | : Syr | nthet | ic (| | | | Ala | Val | Gly | Val | | |
| <211><212><213><220><223> 400 | TY OR FE OT SE | PE: GANI ATUF HER QUEN | PRT SM: E: INFO ICE: Phe -40 | Art: DRMA: 64 Pro | FION Leu | : Syr Phe | nthet Pro | Thr | Leu | Val | Tyr | | -30 | | | | |
| <pre><211> <212> <213> <220> <223> <400> Met L Val A Leu A</pre> | TY OR FE OT SE | PE: GANI ATUF HER QUEN Tyr Phe -25 | PRT ISM: RE: INFO ICE: Phe -40 Pro | Art: DRMA: 64 Pro Asp | rion Leu Tyr | : Syr Phe Ala | Pro Ser | Thr -35 Leu | Leu Ala | Val Gly | Tyr Leu Glu | Ser -15 | -30 Gln | Gln | Glu | | |
| <pre><211> <212> <213> <220> <223> <400> Met L Val A Leu A</pre> | TYY OR OR FE OT SE | PE: GANI ATUR HER QUEN Tyr Phe -25 | PRT SM: E: INFO ICE: Phe -40 Pro | Art: DRMAT 64 Pro Asp | Leu Tyr Pro | Phe Ala Thr | Pro Ser -20 Leu | Thr -35 Leu Glu | Leu Ala Ala | Val Gly Arg | Tyr Leu Glu 1 | Ser -15 Pro | -30 Gln Gly | Gln | Glu Pro 5 | | |
| <211><212><213><213><220><223> 400 Met L Val A | TYYS OR FE OT SE VYS Asp 10 | PE: GANI ATUR HER QUEN Tyr Phe -25 Ala | PRT (SM: CE: INFO INFO INFO INFO INFO INFO INFO INFO | Art: 64 Pro Asp Ile Glu 10 | Leu Tyr Pro Asn | Phe Ala Thr -5 Ser | Pro Ser -20 Leu Ser | Thr -35 Leu Glu | Leu Ala Ala Lys 15 | Val Gly Arg -1 Leu | Tyr Leu Glu 1 Val | Ser -15 Pro Asn | -30 Gln Gly Asp | Gln Leu Glu 20 | Glu Pro 5 Ala | | |
| <211><212><213><213><220><223> <400> Met L Val A Leu A Pro G | TYY OR OR OT SE | PE: GANI ATUF HER QUEN Tyr Phe -25 Ala Pro | PRT SM: RE: INFC ICE: Phe -40 Pro Ile Leu Lys 25 | Art: DRMA: 64 Pro Asp Ile Glu 10 Pro | Leu Tyr Pro Asn | : Syn Phe Ala Thr -5 Ser Arg | Pro Ser -20 Leu Ser | Thr -35 Leu Glu Ala Gly 30 | Leu Ala Ala Lys 15 | Val Gly Arg -1 Leu | Tyr Leu Glu 1 Val | Ser -15 Pro Asn Gly | -30 Gln Gly Asp Pro 35 | Gln Leu Glu 20 Cys | Glu Pro 5 Ala Pro | | |
| <pre><211> <212> <213> <220> <223> <400> Met L Val A Leu A Pro G His P Gly L Ala T</pre> | TYYS OR SELYS Ala Asp 10 Gly Pro | PE: GANI ATUF HER PQUEN Tyr Phe -25 Ala Pro Trp | PRT SM: SE: INFO Pro Ile Leu Lys 25 Thr | Art: DRMA: 64 Pro Asp Ile Glu 10 Pro Leu | Leu Tyr Pro Asn Leu | Phe Ala Thr -5 Ser Arg | Pro Ser -20 Leu Ser Pro His 45 | Thr -35 Leu Glu Ala Gly 30 | Leu Ala Ala Lys 15 Asp | Val Gly Arg -1 Leu Ile | Tyr Leu Glu 1 Val Arg | Ser -15 Pro Asn Gly Arg 50 | -30 Gln Gly Asp Pro 35 Asn | Gln Leu Glu 20 Cys | Glu Pro 5 Ala Pro Val | | |
| <pre><211> <212> <213> <220> <223> <400> Met L Val A Leu A Pro G His P Gly L Ala T</pre> | TYY OR OR OT | PE: GANI ATUR HER QUEN Tyr Phe -25 Ala Pro Trp Asn 40 Pro | PRT SM: CE: INFO INFO INFO INFO INFO INFO INFO INFO | Art: 64 Pro Asp Ile Glu 10 Pro Leu Gln | Tyr Pro Asn Leu Ala | Phe Ala Thr -5 Ser Arg Ser Ile 60 | Pro Ser -20 Leu Ser Pro His 45 Asn | Thr -35 Leu Glu Ala Gly 30 Gly | Leu Ala Ala Lys 15 Asp Tyr | Val Gly Arg -1 Leu Ile Leu Gln | Tyr Leu Glu 1 Val Arg Pro Glu 65 | Ser -15 Pro Asn Gly Arg 50 | -30 Gln Gly Asp Pro 35 Asn | Glu Glu 20 Cys Gly Asn | Glu Pro 5 Ala Pro Val | | |
| <pre><211><212><213><220><223></pre> <pre><400> Met L Val A Leu A Pro G His P Gly L Ala T 5 Asp A</pre> | TY OR | PE: GANIATUR CHER COULEN TYP Phe -25 Ala Pro Asn 40 Pro Gln | PRT SM: LE: INFO ICE: Phe -40 Ile Leu Lys 25 Thr Val | Art: ORMAN 64 Pro Asp Ile Glu 10 Pro Leu Gln Ala | TION Leu Tyr Pro Asn Leu Ala Ile Val 75 | Phe Ala Thr -5 Ser Arg Ser Ile 60 Phe | Pro Ser -20 Leu Ser Pro His 45 Asn | Thr -35 Leu Glu Ala Gly 30 Gly | Leu Ala Ala Lys 15 Asp Tyr Val | Val Gly Arg -1 Leu Ile Leu Gln Ala 80 | Tyr Leu Glu 1 Val Arg Pro Glu 65 Ala | Ser -15 Pro Asn Gly Arg 50 Gly | -30 Gln Gly Asp Pro 35 Asn Leu Leu | Gln Leu Glu 20 Cys Gly Asn | Glu Pro 5 Ala Pro Val Phe Asp 85 | | |
| <pre><211><212><213><220><223></pre> <pre><400> Met L Val A Leu A Pro G His P Gly L Ala T 5 Asp A 70</pre> | TYYOUR TYPE | PE: GANIATURE COLUMN TYP Phe -25 Ala Pro Trp Asn 40 Pro Gln Leu | PRT SM: SM: VE: INFC ICE: Phe -40 Pro Ile Leu Lys 25 Thr Val Ala Ile | Art: ORMA: 64 Pro Asp Ile Glu 10 Pro Leu Gln Ala Thr 90 | TION Leu Tyr Pro Asn Leu Ala Ile Val 75 Asp | Phe Ala Thr -5 Ser Arg Ser Ile 60 Phe | Pro Ser -20 Leu Ser Pro His 45 Asn Ala | Thr -35 Leu Glu Ala Gly 30 Gly Ala Thr | Leu Ala Ala Lys 15 Asp Tyr Val Tyr Ile 95 | Val Gly Arg -1 Leu Ile Leu Gln Ala 80 Gly | Tyr Leu Glu 1 Val Arg Pro Glu 65 Ala | Ser -15 Pro Asn Gly Arg 50 Gly His | -30 Gln Gly Asp Pro 35 Asn Leu Leu | Gln Leu Glu 20 Cys Gly Asn Val | Glu Pro 5 Ala Pro Val Phe Asp 85 Leu | | |

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

| Asp Tyr Se | | | | | | | | | | | | | | |
|---|---|---|--|---|---|---|--|--|---|---|---|--|---|-----------------|
| | r Asn | | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly Glu Le | ı Arg | Phe 170 | Lys | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn Phe Se | r Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Ser | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr Phe Pro | | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu Asp Me 215 | a Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe 225 | Ser | Arg | Met | Pro | |
| Asp Asp Ph | e Phe | _ | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Gly | Thr | Gly | Val | Glu 245 | |
| Val Val Il | ∋ Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Arg | Asn | Val | Gly 260 | Lys | |
| Ile Asn Se | r Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | |
| Cys Leu Me 28 | _ | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | |
| Pro Asn Pro 295 | o Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp | |
| Phe Phe Phe 310 | ∋ Gln | - | Val 315 | Ala | Ala | Gly | Cys | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | |
| Gly Arg As | Þ | | | | | | | | | | | | | |
| <210> SEQ | ID NO | 65 | | | | | | | | | | | | |
| <pre><211> LENG <212> TYPE <213> ORGA <220> FEAT <223> OTHE <221> NAME <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA</pre> | : DNA NISM: URE: R INFO URE: /KEY: FION: URE: /KEY: | Arti DRMAT CDS (1). | 'ION: .(11 pept | : Wt- 113) | -SoLo | | riant | : wit | ch mo | odif: | ied : | signa | al peptió | le |
| <pre><212> TYPE <213> ORGA <220> FEAT <222> OTHE <220> FAT <221> NAME <222> LOCA <220> FEAT <221> NAME</pre> | : DNA NISM: JRE: R INFO JRE: /KEY: FION: JRE: /KEY: FION: | Arti ORMAT CDS (1). mat_ | 'ION: .(11 pept | : Wt- 113) | -SoLo | | riant | : wit | ch mo | odif: | ied : | signa | al peptic | le |
| <pre><212> TYPE <213> ORGA <220> FEAT <223> OTHE <220> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME</pre> | : DNA NISM: JRE: R INFO JRE: /KEY: FION: JRE: /KEY: FION: FION: ENCE: | Arti DRMAT CDS (1). mat_ (130 65 ccc | ·!ION: .(11 pept .)(| : Wt- 113) ide (1116 | SoLo | vai | ttg | gtc | tac | gca | gtg | 999 | gtc | le 48 |
| <pre><212> TYPE <213> ORGA' <220> FEAT' <222> OTHE <220> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA <2400> SEQU </pre> | : DNA NISM: JRE: ZR INFO JRE: /KEY: FION: FION: ENCE: ttt The -40 cct Pro | Arti DRMAT CDS (1). mat_; (130 65 ccc Pro | . (11 pept) (ctg Leu | : Wt- l13) ide (1116 ttc Phe | SoLo CCa Pro | acc Thr -35 | ttg Leu gcc | gtc Val ggc | tac Tyr | gca Ala agc | gtg Val -30 cag | ggg Gly cag | gtc Val | |
| <pre><212> TYPE <213> ORGA <220> FEAT <222> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA <220> FEAT <400> SEQU atg aaa ta Met Lys Ty gtt gct tt Val Ala Ph</pre> | E DNA NISM: URE: R INFO URE: /KEY: FION: URE: /KEY: FION: ENCE: ttt r Phe -40 c cct | Arti RMAT CDS (1). mat: (130 65 ccc Pro gac Asp | ctg Leu | ttc Phe | cca Pro tca Ser -20 | acc Thr -35 ttg Leu | ttg Leu gcc Ala | gtc Val ggc Gly cga | tac Tyr ctc Leu gag | gca Ala agc Ser -15 cca | gtg Val -30 cag Gln | ggg Gly cag Gln | gtc Val gaa Glu cct | 48 |
| <pre><212> TYPE <213> ORGA <220> FEAT <223> OTHE <220> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA <200> FEAT <221> TOCA <200 FEAT <210 FEAT <221> TOCA <200 FEAT <221 FEAT <222 FEAT <221 FEAT <222 FEAT <222 FEAT <222 FEAT <222 FEAT <223 FEAT <23 FEAT <23 FEAT <24 FEAT <24 FEAT <24 FEAT <25 FEAT <26 FEAT <27 FEAT <27 FEAT <27 FEAT <28 FEAT <28 FEAT <28 FEAT <29 FEAT <29 FEAT <29 FEAT <29 FEAT <29 FEAT <29 FEAT <20 FEA</pre> | EDNA NISM: JRE: RINFO JRE: /KEY: FION: JRE: /KEY: FION: ENCE: -40 -40 | Arti RMAT CDS (1). mat (130 65 ccc Pro gac Asp atc Ile | ctg Leu tac Tyr | ttc Phe gcc Ala aca Thr -5 | cca Pro tca Ser -20 ctc Leu | acc Thr -35 ttg Leu gag Glu | ttg Leu gcc Ala gcc Ala | gtc Val ggc Gly cga Arg -1 ttg | tac Tyr ctc Leu gag Glu 1 gtg | gca Ala agc Ser -15 cca Pro | gtg Val -30 cag Gln gga Gly | ggg Gly cag Gln tta Leu | gtc Val gaa Glu cct Pro 5 | 4 8 |
| <pre><212> TYPE <213> ORGA <220> FEAT <222> FEAT <222> FEAT <221> NAME <222> LOCA <220> FEAT <221> NAME <222> LOCA <20> FEAT <21> NAME <221> TYPE </pre> <pre><400> SEQU </pre> <pre>atg aaa ta Met Lys Ty </pre> <pre>gtt gct tt Val Ala Ph </pre> <pre>-2 ttg gac gc Leu Asp Al <10</pre> <pre>cct ggt cc</pre> | E DNA NISM: JRE: JRE: /KEY: FION: JRE: /KEY: FION: ENCE: C ttt Phe -40 C ctt Pro C ata A lle C tcu C Leu G aag | Arti RMAT CDS (1). mat (130 65 ccc Pro gac Asp atc Ile gag Glu 10 ccg | . (11 pept ctg Leu tac Tyr cca Pro aat Asn | ttc Phe gcc Ala aca Thr -5 agc | SoLocal Solocal Solocal Ser -20 ctc Leu tct Ser cct | acc Thr -35 ttg Leu gag Glu gca Ala | ttg Leu gcc Ala gcc Ala aag Lys 15 | gtc Val ggc Gly cga Arg -1 ttg Leu | tac Tyr ctc Leu gag Glu 1 gtg Val | gca Ala agc Ser -15 cca Pro aac Asn | gtg Val -30 cag Gln gga Gly gac Asp | ggg Gly cag Gln tta Leu gag Glu 20 tgc | gtc Val gaa Glu cct Pro 5 gct Ala | 48 96 144 |

| _ | | | | | | | | | | | | | | | | |
|-----|-------------------|-----|-----|-----|------------|-----|-----|-----|-----|------------|-----|-----|-----|-----|------------|------|
| _ | acc Thr 55 | _ | | | | | | | - | _ | _ | | | | | 336 |
| | aat Asn | | | | | | | | | | | | | | | 384 |
| | aat Asn | | | _ | _ | _ | _ | - | | | _ | _ | _ | | | 432 |
| | ggg Gly | | _ | | | | | _ | | _ | | | | | | 480 |
| | ggc Gly | | | - | | _ | - | - | _ | | _ | | _ | _ | | 528 |
| | ggc Gly 135 | | | | _ | | | | _ | | | _ | _ | _ | - | 576 |
| _ | tac Tyr | _ | | _ | | | | | | | | | | _ | | 624 |
| | g gag Glu | | - | | _ | _ | | | - | | | | | | | 672 |
| | ttc n Phe | | | - | _ | | | | | | - | | | | | 720 |
| | ttc Phe | | | | | | | _ | | _ | | _ | - | | - | 768 |
| | gat Asp 215 | _ | - | - | _ | | _ | | | | | _ | _ | _ | | 816 |
| _ | gat Asp | | | - | - | | _ | _ | _ | - | | | | - | | 864 |
| - | gtt Val | | _ | _ | | | _ | _ | | | _ | | _ | | _ | 912 |
| | aac Asn | | | | | | | | | | | | | | | 960 |
| | ttg Leu | | | | | | | | | | | | | | | 1008 |
| | aat Asn 295 | Pro | _ | | _ | | _ | | - | | | _ | | | - | 1056 |
| | . ++- | ttc | _ | gga | gtc | gcc | gct | gga | - | | | - | | | | 1104 |
| 310 | Phe | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Cys | Thr 320 | GIn | Val | Phe | Pro | Tyr 325 | |

<210> SEQ ID NO 66 <211> LENGTH: 371 <212> TYPE: PRT <213> ORGANISM: Artificial Sequence

| <220> <223> | | | | ORMA' | TION | : Syı | nthet | cic (| Const | truc | t | | | | | |
|----------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--|
| <400> | SE | EQUE | ICE : | 66 | | | | | | | | | | | | |
| Met L | ıλa | Tyr | Phe -40 | Pro | Leu | Phe | Pro | Thr -35 | Leu | Val | Tyr | Ala | Val -30 | Gly | Val | |
| Val A | Ala | Phe -25 | Pro | Asp | Tyr | Ala | Ser -20 | Leu | Ala | Gly | Leu | Ser -15 | Gln | Gln | Glu | |
| Leu A | Asp 10 | Ala | Ile | Ile | Pro | Thr | Leu | Glu | Ala | Arg -1 | | Pro | Gly | Leu | Pro 5 | |
| Pro G | Sly | Pro | Leu | Glu 10 | Asn | Ser | Ser | Ala | Lys 15 | Leu | Val | Asn | Asp | Glu 20 | Ala | |
| His P | ro | Trp | Lув 25 | Pro | Leu | Arg | Pro | Gly 30 | Asp | Ile | Arg | Gly | Pro 35 | СЛа | Pro | |
| Gly L | eu | Asn 40 | Thr | Leu | Ala | Ser | His 45 | Gly | Tyr | Leu | Pro | Arg 50 | Asn | Gly | Val | |
| Ala T | hr 55 | Pro | Val | Gln | Ile | Ile 60 | Asn | Ala | Val | Gln | Glu 65 | Gly | Leu | Asn | Phe | |
| Asp A | sn | Gln | Ala | Ala | Val 75 | Phe | Ala | Thr | Tyr | Ala 80 | Ala | His | Leu | Val | Asp 85 | |
| Gly A | an | Leu | Ile | Thr 90 | Asp | Leu | Leu | Ser | Ile 95 | Gly | Arg | Lys | Thr | Arg 100 | Leu | |
| Thr G | Sly | Pro | Asp 105 | Pro | Pro | Pro | Pro | Ala 110 | Ser | Val | Gly | Gly | Leu 115 | Asn | Glu | |
| His G | Sly | Thr 120 | Phe | Glu | Gly | Asp | Ala 125 | Ser | Met | Thr | Arg | Gly 130 | Asp | Ala | Phe | |
| Phe G | 31y .35 | Asn | Asn | His | Asp | Phe 140 | Asn | Glu | Thr | Leu | Phe 145 | Glu | Gln | Leu | Val | |
| Asp T 150 | yr | Ser | Asn | Arg | Phe 155 | Gly | Gly | Gly | Lys | Tyr 160 | Asn | Leu | Thr | Val | Ala 165 | |
| Gly G | lu | Leu | Arg | Phe 170 | ГÀа | Arg | Ile | Gln | Asp 175 | Ser | Ile | Ala | Thr | Asn 180 | Pro | |
| Asn P | he | Ser | Phe 185 | Val | Asp | Phe | Arg | Phe 190 | Ser | Thr | Ala | Tyr | Gly 195 | Glu | Thr | |
| Thr P | he | Pro 200 | Ala | Asn | Leu | Phe | Val 205 | Asp | Gly | Arg | Arg | Asp 210 | Asp | Gly | Gln | |
| Leu A | Asp 215 | Met | Asp | Ala | Ala | Arg 220 | Ser | Phe | Phe | Gln | Phe 225 | Ser | Arg | Met | Pro | |
| Asp A | ap | Phe | Phe | Arg | Ala 235 | Pro | Ser | Pro | Arg | Ser 240 | Gly | Thr | Gly | Val | Glu 245 | |
| Val V | al | Ile | Gln | Ala 250 | His | Pro | Met | Gln | Pro 255 | Gly | Arg | Asn | Val | Gly 260 | Lys | |
| Ile A | sn | Ser | Tyr 265 | Thr | Val | Asp | Pro | Thr 270 | Ser | Ser | Asp | Phe | Ser 275 | Thr | Pro | |
| Cys L | eu | Met 280 | Tyr | Glu | Lys | Phe | Val 285 | Asn | Ile | Thr | Val | Lys 290 | Ser | Leu | Tyr | |
| Pro A | sn 195 | Pro | Thr | Val | Gln | Leu 300 | Arg | Lys | Ala | Leu | Asn 305 | Thr | Asn | Leu | Asp | |
| Phe P | he | Phe | Gln | Gly | Val 315 | Ala | Ala | Gly | Сув | Thr 320 | Gln | Val | Phe | Pro | Tyr 325 | |
| Gly A | arg | Asp | | | | | | | | | | | | | | |

- 1. A polynucleotide that encodes a polypeptide with peroxygenase activity, wherein the polypeptide amino acid sequence encoding having an identity of at least 70% with SEQ ID NO: 2 (AaeUPO1), 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 of claim 1 further comprising 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 of claim 1 wherein the 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 of claim **1** further comprising the nucleotide sequence encoding the signal peptide of SEQ ID NO: 26.
- 5. The polynucleotide of claim 4 wherein 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, of claim **1** wherein the sequence is selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 9 and SEQ ID NO: 7.
- 7. The polynucleotide, of claim 1 wherein the sequence is selected from the group consisting of SEQ ID NO: 23, SEQ ID NO: 21 and SEQ ID NO: 19.

- **8**. The polynucleotide, of claim **1** wherein the sequence is selected from the group consisting of SEQ ID NO: 41, SEQ ID NO: 39 and SEQ ID NO: 37.
- **9**. The polypeptide is encoded by any of the nucleotide sequences according to claim **1**.
- 10. The polypeptide of claim 9 wherein the sequence is selected from the group 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 of claim 10 wherein the sequence is selected from the group consisting of: SEQ ID NO: 24, SEQ ID NO: 22 and SEQ ID NO: 20.
- 12. The polypeptide of claim 10 wherein sequence is selected from the group consisting of: SEQ ID NO: 38, SEQ ID NO: 40 and SEQ ID NO: 42
- 13. A method for obtaining the polypeptide, according to of claim 9 comprising the following operations:
 - i. Introducing a vector with the polynucleotide, according to claim 1, in a suitable host cell.
 - ii. Culturing the host cell in a suitable medium, and,
 - iii. Purifying the synthesised polypeptide.
- 14. A host cell comprising the polynucleotide according to claim 1 and which is capable of producing the polypeptide according to claim 9.
- 15. The host cell, according to claim 14, wherein the host cell is a yeast or fungus.
- **16**. The host cell, according to claim **15**, wherein the yeast belongs to the genus *Saccharomyces* sp or *Pichia* sp.
- 17. The host cell, according to claim 15, wherein the fungus belongs to the genus *Aspergillus* sp.
- 18. A kit comprising at least one polypeptide according to claim 9.
- 19. An electronic device comprising at least one polypeptide according to claim 9.
 - 20. (canceled)
 - **21**. (canceled)
 - 22. (canceled)
 - 23. (canceled)
 - 24. (canceled)
 - 25. (canceled)
 - 26. (canceled)
 - **27**. (canceled)
- 28. A method for hydroxylation of at least one hydrocarbon comprising the use of the polypeptide according to claim 9, of the host cell according to claim 14, of the kit according to claim 18 or of 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.
- **30**. A biosensor comprising the polypeptide according to claim **9**, the host cell according to claim **14**, the kit according to claim **18** or the electronic device according to claim **19**.
- 31. A composition comprising the polypeptide according to claim 9
- **32**. The composition according to claim **31**, wherein the composition is a pharmaceutical, nutritional or cosmetic composition.
- 33. The device according to claim 19, wherein further containing immobilized enzymes.

* * * * *