OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,298,069**
(86) 1998/07/23
(87) 1999/02/04

(72) REETZ, MANFRED T., DE
(72) ZONTA, ALBIN, DE
(72) SCHIMOSSEK, KLAUS, DE
(72) LIEBETON, KLAUS, DE
(72) JÄGER, KARL-ERICH, DE
(71) STUDIENGESELLSCHAFT KOHLE MBH, DE

(51) Int.Cl.[7] C12N 15/55, C12Q 1/44, C12N 1/21, C12N 9/20, C12N 15/10
(30) 1997/07/25 (197 31 990.4) DE

(54) **PROCEDE DE FABRICATION ET D'IDENTIFICATION DE NOUVELLES HYDROLASES A PROPRIETES AMELIOREES**

(54) **METHOD FOR PRODUCING AND IDENTIFYING NEW HYDROLASES HAVING IMPROVED PROPERTIES**

(57) L'invention concerne un procédé de fabrication et d'identification de mutants hydrolases à propriétés améliorées eu égard à la stéréosélectivité ou à la régiosélectivité, à l'activité catalytique ou à la stabilité au cours de réactions chimiques.

(57) The invention relates to a method for producing and identifying hydrolase mutants having improved properties regarding stereoselectivity and site selectivity, catalytic activity or stability during chemical reactions.

**PCT**

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 :
C12N 15/55, 15/10, 9/20, 1/21, C12Q 1/44

A2

(11) Internationale Veröffentlichungsnummer: **WO 99/05288**

(43) Internationales Veröffentlichungsdatum: 4. Februar 1999 (04.02.99)

(21) Internationales Aktenzeichen: PCT/EP98/04612

(22) Internationales Anmeldedatum: 23. Juli 1998 (23.07.98)

(30) Prioritätsdaten:
197 31 990.4 25. Juli 1997 (25.07.97) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: STUDIENGESELLSCHAFT KOHLE MBH [DE/DE]; Kaiser–Wilhelm–Platz 1, D–45470 Mülheim an der Ruhr (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: REETZ, Manfred, T. [DE/DE]; Kaiser–Wilhelm–Platz 1, D–45470 Mülheim an der Ruhr (DE). ZONTA, Albin [DE/DE]; Kaiser–Wilhelm–Platz 1, D–45470 Mülheim an der Ruhr (DE). SCHIMOSSEK, Klaus [DE/DE]; Kaiser–Wilhelm–Platz 1, D–45470 Mülheim an der Ruhr (DE). LIEBETON, Klaus [DE/DE]; Kaiser–Wilhelm–Platz 1, D–45470 Mülheim an der Ruhr (DE). JÄGER, Karl–Erich [DE/DE]; Ruhr Universität Bochum, Fakultät f. Biologie, Universitätsstrasse 150, D–44790 Bochum (DE).

(74) Anwälte: VON KREISLER, Alek usw.; Deichmannhaus am Hauptbahnhof, Bahnhofsvorplatz, D–50667 Köln (DE).

(81) Bestimmungsstaaten: CA, JP, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht
*Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.*
*Mit Angaben über hinterlegtes biologisches Material, eingereicht gemäss Regel 13bis getrennt von der Beschreibung.*

(54) Title: METHOD FOR PRODUCING AND IDENTIFYING NEW HYDROLASES HAVING IMPROVED PROPERTIES

(54) Bezeichnung: VERFAHREN ZUR HERSTELLUNG UND IDENTIFIZIERUNG VON NEUEN HYDROLASEN MIT VERBESSERTEN EIGENSCHAFTEN

(57) Abstract

The invention relates to a method for producing and identifying hydrolase mutants having improved properties regarding stereoselectivity and site selectivity, catalytic activity or stability during chemical reactions.

(57) Zusammenfassung

Die vorliegende Erfindung betrifft ein Verfahren zur Herstellung und Identifizierung von Hydrolase–Mutanten mit verbesserten Eigenschaften hinsichtlich der Stereo– oder Regioselektivität, katalytischen Aktivität oder Stabilität bei chemischen Umsetzungen.

SMB

# Method for Producing and Identifying New Hydrolases Having Improved Properties

The present invention relates to a process for the preparation and identification of hydrolase mutants having improved properties with respect to stereo- or regioselectivity, catalytic activity or stability in chemical reactions.

## Prior art:

Hydrolases are among the most wide-spread enzymes in organic synthesis. As a subgroup of the hydrolases, esterases and lipases, in particular, catalyze a wide variety of reactions, such as the hydrolysis of carboxylic acid esters, or the synthesis of esters or transesterifications in organic solvents. Due to their high stereoselectivity, stability and their being readily available, they are interesting for numerous industrial processes. Thus, for example, lipases have been industrially employed for the optical resolution of chiral alcohols, acids or amines, for the preparation of optically pure medicaments, natural substances, plant protective agents or high-grade fats and oils (K. Faber, *Biotransformations in Organic Chemistry*, Springer-Verlag, Berlin, 2nd Ed. 1995). Nevertheless, the enantioselectivity of a lipase or esterase with respect to a given substrate cannot be predicted with certainty, and in many cases, the reactions proceed with only moderate optical yields. Therefore, there is a need for a process for the preparation of hydrolases which enables a well-aimed optimization of enantioselectivity

with respect to a desired product and the special process conditions, such as temperature and solvent. Although effects on the enantioselectivity of lipases could be studied using the molecular-biological method of *in vitro* mutagenesis, which is customary today (K. Hult, M. Holmquist, M. Martinelle, *European Symposium on Biocatalysis*, Graz, 1993, Abstracts, L-4), an optimization with respect to a particular substrate which would have led to an enzyme useful in organic synthesis could not be achieved.

The most important possible applications of genetic engineering include protein desig, wherein mutations are introduced base-specifically into the gene sequence of the corresponding protein based on known structural data using *in vitro* mutagenesis. By selectively substituting amino acids, enzymes having improved catalytical activity or stability could already be prepared in this way (A. Shaw, R. Bott, *Current Opinion in Structural Biology*, 1996, 6, 546). This technique, the so-called oligonucleotide-directed or site-directed mutagenesis, is based on the substitution of a short sequence segment of the gene coding for the naturally occurring enzyme (wild type) by a synthetically mutagenized oligonucleotide. Subsequent expression of the gene results in an enzyme mutant which may have advantageous properties. In a method derived therefrom, the so-called cassette mutagenesis, oligonucleotides with partially randomized sequences are used. This provides a library of mutants of a limited size, which can then be tested with respect to its properties.

Despite of the advantages of these established methods, they are hardly suitable for the stepwise optimization of an enzyme or for the generation of enzymes having novel properties. The fact that our understanding of the laws governing protein folding and the structure-function relationship of proteins is still incomplete is the main reason for the failing of many projects in the field of the so-called rational

protein design. In addition, a stepwise optimization process according to the classical method is relatively labor-consuming and does not ensure a significant improvement of the enzyme properties per se.

More recently, novel molecular-biological methods of mutagenesis have been described (D.W. Leung, E. Chen, D.V. Goeddel, *Technique*, 1989, 1, 11, and W.P.C. Stemmer, A. Crameri, PCT WO 95/22625) which are based on the polymerase chain reaction known from the literature (R.K. Saiki, S.J. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, N. Arnheim, *Science*, 1985, 230, 1350). Instead of site-directed mutagenesis, these methods employ combinatorial methods for the generation of extensive mutant libraries which are subsequently screened for mutants having positive properties using suitable screening methods. This mimics the naturally occurring evolutive processes of replication and recombination, mutation and selection on a molecular level. This method, described as *in vitro* evolution (or *directed evolution*), has already proven useful in some cases as a suitable method for obtaining new biocatalysts (W.P.C. Stemmer, *Nature*, 1994, 370, 389, and F.H. Arnold, *Chemical Engineering Science*, 1996, 51, 5091).

In spite of the progress made in this field, this method cannot yet be generally transferred to all classes of enzymes, since suitable test methods for identifying mutants with positive properties are lacking in most cases. Such methods are a sine qua non, however, in view of the large number of mutated enzyme variants to be expected in the production of combinatorial mutant libraries. Especially in the case of the lipases which are interesting for industrial processes, the production of mutants with improved stereoselectivity by the methods of *in vitro* evolution has not been successful to date, because an efficient screening method for enantioselectivity testing still does not exist. The classical method for determining the enantioselectivity of a lipase- or esterase-catalyzed reaction is based on the separation of the reaction

products and educts by liquid or gas chromatography using chirally modified stationary phases. However, due to the enormous number of samples to be processed in the screening of extensive mutant libraries, this method is unsuitable since chromatographical separations with chirally modified columns are time-consuming, being only capable of sequential processing. Another as yet unsolved problem is the difficulty, frequently to observe, of expressing functional lipases or esterases in host organisms with a sufficiently high activity yield. However, this is indispensable to a high-performance screening system since too low enzyme activities are difficult to detect in the determination of enantioselectivity due to the limited sensitivity of a test system.

Object of the invention:

Therefore, it has been the object of the present invention to provide a simple process for the preparation of mutated hydrolases, especially lipases or esterases, having improved stereo- or regioselectivity, catalytic activity and stability towards particular substrates (e.g., carboxylic acids, alcohols, amines, or their derivatives), which process additionally enables a rapid identification of positive mutants from extensive mutant libraries, and the use of the enzymes thus prepared in the optical resolution of chiral alcohols, acids and amines, and their derivatives.

Description of the invention:

The present invention relates to

(1) a process for the preparation and identification of hydrolase mutants having improved properties with respect to stereo- or regioselectivity, catalytic activity or stability, characterized in that

a) a starting hydrolase gene is mutagenized by a modified polymerase chain

reaction (PCR), wherein the mutation rate and total number of mutations in the amplified DNA is adjusted by adjusting the concentrations of $Mg^{2+}$, $Mn^{2+}$ and of the deoxynucleotides and by adjusting the number of cycles; and/or

b) one or more starting hydrolase genes, one or more hydrolase genes mutated according to step a), or mixtures of one or more starting hydrolase genes and one or more hydrolase genes mutated according to step a) are mutagenized by enzymatically fragmenting said genes, followed by enzymatic reassembly of the fragments produced to give complete recombinant hydrolase genes;

c) the mutated hydrolase genes obtained according to step a) or b) are transformed into a host organism; and

d) hydrolase mutants having improved properties, expressed by transformants obtained in step c), are identified by a test method;

(2) a hydrolase mutant obtainable by the process defined under (1);

(3) a DNA sequence coding for the hydrolase mutant defined under (2);

(4) a vector comprising the DNA sequence defined under (3);

(5) a transformant comprising a DNA sequence as defined under (3) and/or a vector as defined under (4);

(6) a process for the preparation of hydrolase mutants having improved properties, comprising culturing a transformant according to (5); and

(7) a method for testing catalysts for stereo- or regioselectivity, wherein equal amounts of the catalyst are added to a test substrate and to the pure stereo- or regioisomers of the test substrate, provided with a chromophorous group which causes a spectrometrically determinable change of absorption or emission upon cleavage by the catalyst, in separate test vessels, and the stereo- or regioselectivity is determined from the ratio of the linear initial reaction rates obtained.

As a rule, the preparation of the new biocatalysts starts with the isolation of a lipase or esterase gene from the organism of origin. This may be any microbial, plant and animal organism which is the carrier of a lipase or esterase gene. The isolation of the gene can be effected according to the methods known from the literature (J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989, New York). Usually, the genomic DNA is fragmented using restriction endonucleases, and the gene fragments obtained are cloned in a host organism (e.g., *E. coli*). Then, using oligonucleotides with sequence homology to a segment of the lipase or esterase gene, the gene is identified within the gene library in hybridization experiments, followed by isolation thereof.

Surprisingly, it has been found according to the invention that naturally occurring hydrolase genes can be mutagenized by a modified polymerase chain reaction (PCR), changing certain reaction parameters, to obtain an extensive mutant library which can be screened for mutants having improved enantioselectivity using a novel test method.

The novelty of the process resides in that an extensive randomized mutant library can be established, starting with a naturally occurring lipase or esterase gene (the so-called wild type gene), using a modified PCR (hereinafter referred to as *mutagenizing PCR*). It has been found that the mutation rate during the PCR can be adjusted in a well-aimed manner by changing the components of the PCR. The number of mutations in the lipase gene in question (the mutation rate) can be controlled by varying the concentrations of $Mg^{2+}$ and/or of the deoxyoligonucleotides and/or the addition of $Mn^{2+}$ ions. Preferably, the following concentrations are used depending on the DNA polymerase employed:

$Mg^{2+}$: 1.5 mM - 8.0 mM
dNTP : 0.05 mM - 1.0 mM
$Mn^{2+}$: 0.0 mM - 3.0 mM

In addition, it has been found that the number of cycles in the PCR correlates with the number of mutations: the higher the selected number of cycles, the higher is the total number of mutations. By

means of this parameter, the diversity of the mutant library can be adjusted.

For determining the mutation rate, the purified PCR products are sequenced. The mutation rate can be determined by comparing the sequences obtained with the sequence of the wild type gene.

Table 1 shows the mutation rate as a function of the concentration of the above mentioned components of the PCR in the amplification of the lipase gene from *P. aeruginosa (lipA)*.

Table 1

| Exp. | $Mg^{2+}$ (mM) | $Mn^{2+}$ (mM) | dATP/ dGTP (mM) | dTTP/ dCTP (mM) | Mutation rate (mutations/ 1000 bp[1)]) |
|---|---|---|---|---|---|
| 1 | 6.1 | - | 0.2 | 0.2 | 1-2 |
| 2 | 7.0 | 0.5 | 0.2 | 1.0 | 15-20 |

[1)] bp = base pairs

From the sequencing results, it can further be seen that the transition and transversion types of mutation occur in about the same statistical frequency. In contrast, deletions and insertions are rarely observed. In addition, the mutations are uniformly distributed over the entire lipase gene. Thus, a mutant library with statistically uniformly distributed mutations can be produced by the method described. A mutation rate of 1-2 mutations/hydrolase gene has proven advantageous. Thereby, it is prevented that a negative mutation will mask a mutation with a positive effect, as would be the case if several mutations occurred per one hydrolase gene. In order to obtain a complete mutant library, each with one amino acid substitution per enzyme molecule, 5415 mutants must theoretically be generated in a lipase consisting of 285 amino

acids (here: lipase from *P. aeruginosa*). This value results from the following formula:

$$N = 19 \times M \times 285! / [(285 - M)! \times M!]$$

with N = number of mutants, and M = number of amino acid substitutions per one lipase molecule. According to the invention, it could be surprisingly shown that positive mutants are found in even substantially smaller sized libraries, a mutation rate of 1-2 having been employed.

The mutated lipase or esterase genes obtained by the process described are ligated into a suitable expression vector and then transformed into a host organism, e.g., *E. coli*. Then, the transformed cells are plated on agar plates and cultured. If the expression rate is sufficiently high, the colonies obtained can be transferred to microtitration plates provided with a liquid medium and, after growth has started, can be directly employed in a screening test. In the case where only little enzyme is formed in the expression of the lipase gene or the gene product is not correctly folded in the host organism used (inclusion bodies) or incompletely secreted into the culture medium, it will be advantageous to reclone the mutated genes in another host organism, preferably the original organism.

In order to obtain sufficiently high enzyme activities, the individual bacterial clones which contain a mutated lipase or esterase gene are transferred from the agar plates into the wells of commercially available microtitration plates and cultured in liquid medium. Preferably, microtitration plates having 96 wells per plate are employed. The growth of the bacteria can be monitored by measuring the cell density ($OD_{600}$ value). It is advantageous to inoculate a second microtitration plate in parallel in this way in order to have a reference for the later

identification of positive clones. After the growth of the bacteria glycerol is conveniently added to the reference plate, which is then stored at -80 °C until used for identification. If the bacteria are secreting the enzyme into the extracellular space (as with the lipase from *P. aeruginosa*), the cells in the microtitration plates are centrifuged off, and the supernatant with the lipase or esterase activity is used for the screening test. In the case where the bacteria (e.g., *E. coli*) accumulate the enzyme in the periplasm, a cell wall lysis must be preliminarily done, wherein methods known from the literature, such as lysozyme treatment, can be used.

By culturing the corresponding clones from the reference plate, sufficient plasmid DNA can be isolated which can be used for the characterization of the mutated lipase or esterase gene. The mutations are localized within the gene by sequencing. One advantage of the invention is the fact that the mutated gene in a positive clone can be further optimized with respect to its properties in further mutation cycles by the process described, even without knowing the exact position of the mutations. Thus, the isolated lipase or esterase gene is again used in a PCR modified according to the above stated conditions (*mutagenizing PCR*). This procedure may be repeated until the properties of the lipase or esterase mutant meet the requirements of the stereoselective reaction.

For a further optimization of the identified positive mutants, the process described can be extensive in that the DNA of several positive mutants is first fragmented and then can be reassembled into functional lipase or esterase genes in a combinatorial process according to W.P.C. Stemmer (*Nature*, 1994, 370, 389). The thus obtained *in vitro* recombinant library is subsequently expressed, and the recombinant gene products are examined for improved enantioselectivity using the test methods according to the invention. The advantage of this method

is that the positive properties of different lipase or esterase mutants may be added in one new recombinant gene due to the recombination, which eventually may result in a further improvement of the lipase or esterase. The course of the method described is as follows:

Using the enzyme DNase I (e.g., from bovine pancreas), the lipase or esterase genes are first cleaved into fragments having a preferably length of between 25 bp and 100 bp. The size of the fragments can be checked by separating them by means of agarose electrophoresis and comparing with corresponding DNA length markers. The DNA fragments thus obtained are purified to free them from adhering DNase. The *in vitro* recombination is performed under the conditions of a conventional PCR, but without adding any PCR primers. In analogy with conventional PCR, one cycle is comprised of three steps: a) denaturing, b) annealing and c) elongation. During annealing, hybridization occurs of sequence-homologous fragments which may be derived from different mutated lipase or esterase genes. In the subsequent elongation step, the strands are completed by the DNA polymerase so that new recombinant lipase genes are eventually obtained. The optimum number of cycles is determined in a preliminary experiment. Thus, after every 5 cycles, a small sample of the reaction mixture is separated by agarose gel electrophoresis to determine the cycle in which the maximum of the size distribution of the recombinants in the range of the size of the enzyme gene. A number of cycles of between 30 and 45 is preferably selected. The band obtained in the agarose gel which corresponds in size to the lipase or esterase gene is purified and amplified by a conventional PCR. The PCR product is purified and, following ligation into a suitable vector (plasmid), transformed into *E. coli*. As already discussed in the paragraph dealing with *mutagenizing PCR*, it may be required to reclone in another host organism if the lipase activity should be too low after expression in *E. coli*. The recombinants ob-

tained are grown in microtitration plates for the test for enantioselectivity.

In a variant of the invention, the described methods of *mutagenizing PCR* and *in vitro* recombination for the production of mutant or recombinant libraries can be performed successively or repeated in any order and frequency desired in order to optimize the enantioselectivity of the lipase or esterase. Preferably, at least one mutation cycle is performed in the beginning using *mutagenizing PCR*. This may then be followed by an *in vitro* recombination cycle, wherein the best positive mutant clones are respectively employed. By monitoring the enantioselectivity of the enzyme mutants obtained, the optimization process can be followed.

In another variant of the invention, positive lipase or esterase mutants identified by the screening of mutant or recombinant libraries can be further optimized using classical directed mutagenesis or cassette mutagenesis. Thus, the mutation in the lipase or esterase gene is first localized by sequencing. This gene is subsequently again mutagenized by means of "wobbled" primers at the codons coding for positive mutants. The thus obtained mutant library of a limited size can then be expressed and screened for improved enantioselectivity.

Positive lipase or esterase mutants identified by the screening of mutant or recombinant libraries can be further optimized using site-directed saturation mutagenesis. Thus, the positive mutation in the lipase or esterase gene is first localized by sequencing. Then, using any method of site-directed mutagenesis which allows for the exchange of multiple bases, this gene is changed in such a way that all possible codons are formed at the site of the gene which codes for the position to be optimized. This provides a library of mutants of a limited size in which mutants the amino acid originally present in the amino acid position to be optimized has been replaced by the remaining 19

amino acids. The thus obtained mutant library of a limited size can then be expressed and screened for improved enantioselectivity.

In a variant of the method described, the lipase or esterase gene of the wild type enzyme is employed for *in vitro* recombination together with the positive mutants found. This can result in backcrossings in which mutations having neutral or negative properties can be eliminated. Following expression, the recombinant library obtained can be examined for improved enantioselectivity.

In another variant of the method described, hydrolase genes from different organisms are employed for *in vitro* recombination, provided they possess sufficient sequence homology with the originally employed hydrolase gene.

In a variant of the method, the *in vitro* recombination is performed under the conditions of the modified PCR described. Thus, the concentrations of the $Mg^{2+}$ or $Mn^{2+}$ ions and of the deoxynucleotides (dNTPs) are changed to adjust the mutation rate during *in vitro* recombination in a well-aimed manner.

The invention further relates to test methods which allow for the identification of enzyme mutants having improved stereoselectivity or regioselectivity from extensive mutant libraries. Thus, after centrifuging off the bacterial cells, two aliquots of the enzyme-containing supernatant are transferred to adjacent wells of a new microtitration plate. After addition of the two enantiomeric pure substrates in the two wells, respectively, the activity of the lipase or esterase is determined by spectrophotometry. The measurements are performed in a commercially available spectral photometer for microtitration plates. This allows for a high sample throughput. The selection of the substrate depends on the type of chiral compound for which optimization of the lipase or

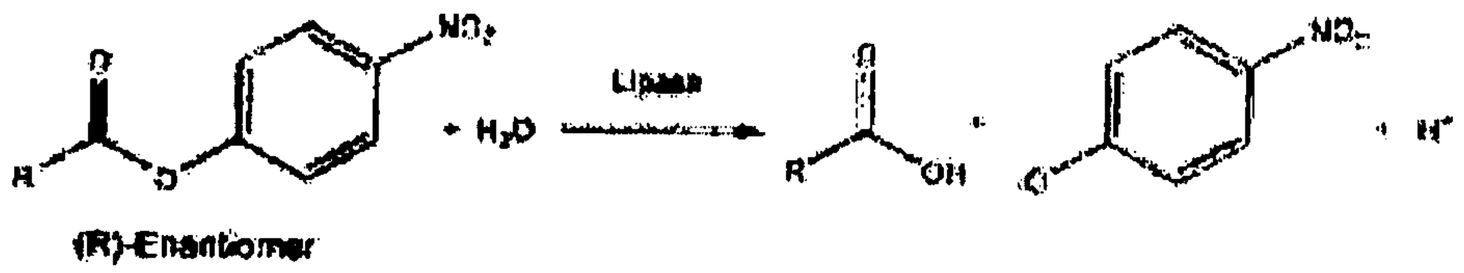
esterase is to be effected. The method is particularly suitable for chiral carboxylic acids, alcohols and amines.

In the case of chiral carboxylic acids or chiral COOH-functional compounds, the two corresponding p-nitrophenyl esters of the *(R)*- and *(S)*-acids are employed as test substrates. Formula 1 shows the principle of the test method wherein R represents any organic residue having at least one asymmetric center.

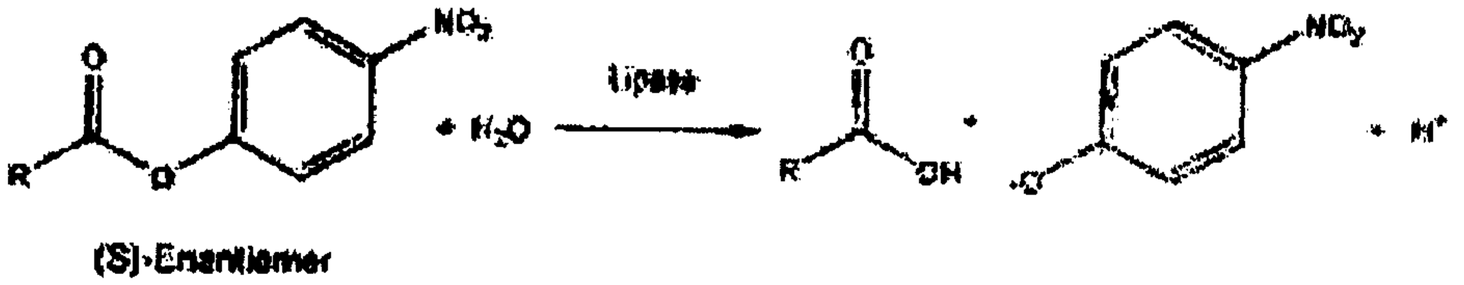
Formula 1

Scheme of the test method for stereoselectivity for chiral carboxylic acids or COOH-functional compounds

Reaction 1:




Reaction 2:



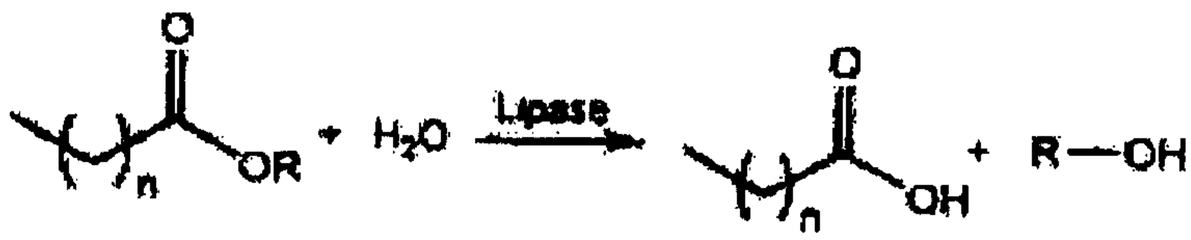

Due to the high absorbance of the p-nitrophenolate anion released in the hydrolase-catalyzed ester hydrolysis ($\lambda_{max}$ = 405 nm, $E_{max}$ = 14,000), a highly sensitive test method results by which an activity determination can be performed even for low substrate concentrations. The enantioselectivity of the hydrolase mutants can be determined with sufficient accuracy from the quotient of the hydrolysis rates $V_{app(R)}$ and $V_{app(S)}$ for the *(R)*- and the *(S)*-ester, respectively. Since both test reactions contain only one enantiomer (either the R- or the S-ester), the absence of a competing reaction with the other enantiomer must be taken into account when the enantioselectivity is determined. Although this kinetic effect may lead to the calculation of inaccurate enantioselectivities, it has been found that the apparent enantioselectivities obtained by the presented method ($E_{app}$) are

sufficiently telling with respect to the enantioselectivity of the mutated lipases. $E_{app}$ is obtained as $V_{app(R)}/V_{app(S)}$. Another advantage is the simple performance and good reproducibility of the test, which is also suitable for screening with a high sample throughput.

In the case of chiral alcohols or chiral OH-functional compounds, fatty acid esters of the two optically pure alcohols are employed in the test for stereoselectivity. The chain length of the fatty acids is within a range of from $C_2$ to $C_{18}$. As the alcohol component, primary, secondary and tertiary alcohols and their derivatives having at least one asymmetric center can be used. Solutions of the esters of the *(R)*- and *(S)*-alcohols are hydrolysed with culture supernatants of the hydrolase mutants in adjacent wells of a microtitration plate. The hydrolysis rates $V_{app(R)}$ and $V_{app(S)}$ for the *(R)*- and the *(S)*-ester, respectively, are a measure of the enantioselectivity of the enzyme mutant examined. Detection is effected through a coupled enzyme reaction (H.U. Bergmeyer, *Grundlagen der enzymatischen Analyse,* Verlag Chemie, Weinheim, 1977) in which the continuous release of the fatty acid is monitored. The dye produced is assayed by colorimetry at 546 nm ($\varepsilon = 19.3\ l \cdot mmol^{-1} \cdot cm^{-1}$). The concentrations of the enzymes, cofactors and coenzymes of auxiliary reactions 2 and 3 (see Formula 2) and of the indicator reaction 4 must be selected in such a way that the lipase- or esterase-catalyzed reaction to be determined is rate-determining. The quotient of the hydrolysis rates for the *(R)*- and the *(S)*-ester, respectively, corresponds to the apparent enantioselectivity ($E_{app}$). In one variant, the fatty acid amides of chiral amines or $NH_2$- or NHR-functional compounds are employed instead of the optically pure esters. Formula 2 shows the scheme of the test system.

Formula 2

Scheme of the test method for stereoselectivity for chiral alcohols; R represents any organic residue having at least one asymmetric center; abbreviations: CoA (coenzyme A), ATP (adenosine-5'-triphosphate), AMP (adenosine-5'-monophosphate)

1.) 


free fatty acid

2.) free fatty acid + CoA + ATP $\xrightarrow{\text{Acyl-CoA Synthetase}}$ acyl-CoA + AMP + pyrophosphate

3.) acyl-CoA + $O_2$ $\xrightarrow{\text{Acyl-CoA Oxidase}}$ enoyl-CoA + $H_2O_2$

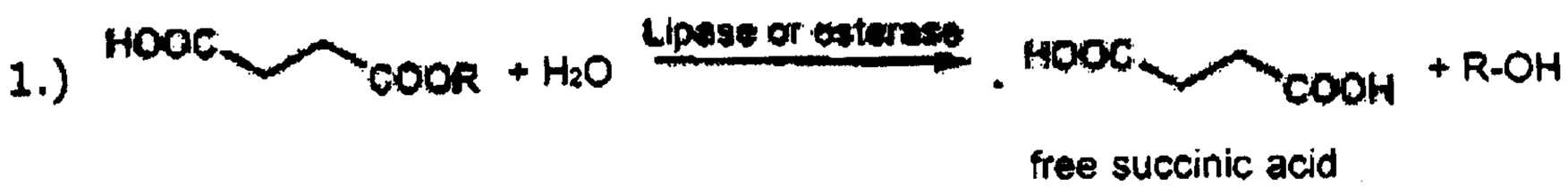
4.) $H_2O_2$ + 4-aminoantipyrine + 2,4,6-tribromo-3-hydroxybenzoic acid $\xrightarrow{\text{Peroxidase}}$ red dye + 2 $H_2O$ + HBr

In a variant of the method, the corresponding esters and amides of succinic acid can be employed instead of the fatty acid esters or amides. The latter have the advantage, over the fatty acids, of being more soluble in aqueous solutions or aqueous-organic solvents. The measurement is performed by UV spectrometry at 340 nm ($\varepsilon$ = 6.3 l·mmol$^{-1}$·cm$^{-1}$). In this test method too, it has to be taken care that the hydrolase-catalyzed reaction 1 be rate-determining. The quotient of the hydrolysis rates $V_{app(R)}$ and $V_{app(S)}$ for the *(R)*- and the *(S)*-ester, respectively, corresponds to the apparent enantioselectivity ($E_{app}$). In one variant, the fatty acid amides of chiral amines are employed instead of the optically pure esters. Both primary and secondary

amines may be employed as the amine component. The scheme of the test system is represented in Formula 3.

Formula 3

Scheme of the test method for stereoselectivity for chiral alcohols; R represents any organic residue having at least one asymmetric center; abbreviations: CoA (coenzyme A), ITP (inosine-5'-triphosphate), IDP (inosine-5'-diphosphate), NADH/$NAD^+$ (reduced/oxidized nicotinamide adenine dinucleotide)




The test for the identification of hydrolase mutants having improved stereoselectivity may further be performed in such a way that both stereoisomers are contained in the test reaction. Thus, the separated measurements of the *(R)*- and *(S)*-enantiomers can be dispensed with.

The test principle starts with binding a racemic mixture of the chiral substrate to a solid phase. Through an ester or amide linkage to this chiral compound, a radioactively labeled organic residue is bound. Two cases can be distinguished:

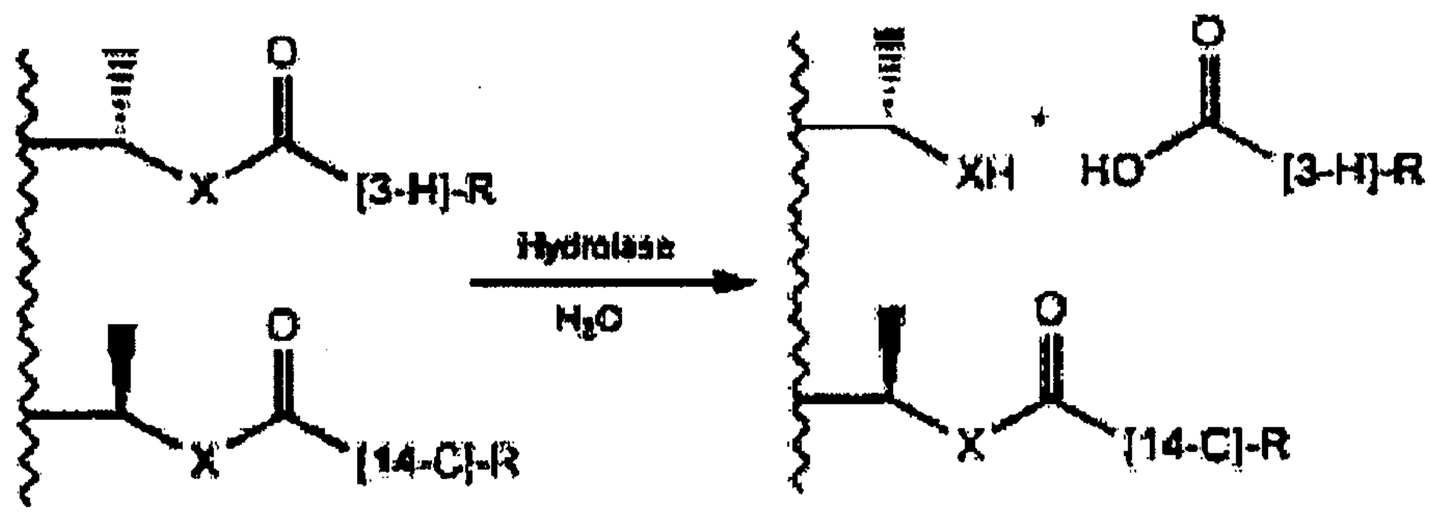
a) Solid-phase bound chiral carboxylic acid: the carboxy function is esterified with a radioactively labeled alcohol.

b) Solid-phase bound chiral alcohol or chiral amine, or OH- or $NH_2$-functional (or NHR-functional) compounds: the hydroxy or amine function is labeled with a radioactively labeled carboxylic acid.

It is critical that the two enantiomers of the racemic mixture bound to the solid phase be labeled with different isotopes. Preferably, $^{3}H$- and $^{14}C$-labeled compounds are used. As the solid phase, all usual organic functionalized polymers as well as inorganic functionalized supports can be employed. Preferably, solid phases based on polystyrene and silica gel supports are employed. The chiral radioactively labeled compounds are then bound to the solid phase wherein the coupling to the solid phase must be adapted to the chemical nature of the chiral substrate. Formula 4 shows the scheme of the modified solid phase and the principle of the test method.

Formula 4

Scheme of the solid-phase screening test for stereoselectivity with a dual radioactively labeled substrate; X = O, NH; R is a radioactively labeled organic residue




Approximately equal amounts of the thus modified support can be dispensed to small reaction vessels (e.g., the wells of microtitration plates) and then admixed with the culture supernatants of the hydrolase mutants. In the subsequent reaction, the radioactively labeled components (carboxylic acid or alcohol) are hydrolysed from the solid phase and released into the liquid medium. An aliquote of the medium is then removed and examined for the amount of radioactivity in a scintillation counter. From the ratio between the two different isotopes, the enantiomeric excess and the conversion of the reaction and thus the enantioselectivity of the mutated esterase or lipase can be calculated. By using regioisomeric test compounds, the tests described can also be used for the identification of hydrolase mutants having improved regioselectivity. Instead of hydrolase mutants, other catalysts may also be employed for determining the stereo- or regioselectivity.

The test for enantioselectivity of the hydrolase mutants prepared by the process described may also be performed by a capillary-electrophoretical separation using chirally modified capillaries which allow for a direct separation of the enantiomeric substrates or products

of the hydrolase-catalyzed test reaction. Here, the test substrates can be employed as a racemate. The separation may be effected both in capillaries and by the use of prepared microchips which allow for electrophoretical separation and parallel running of the analyses for a high sample throughput. In both cases, it is a precondition that the enantiomers can be separated by capillary electrophoresis.

The invention will now be further illustrated by the following Examples and Figures.

Figure 1 shows the experimentally obtained measured curves for the determination of the apparent enantioselectivity ($E_{app}$) in the hydrolysis of *(R)*- and *(S)*-2-methyldecanoic acid p-nitrophenyl ester with culture supernatants of the lipase mutants P1B 01-E4, P2B 08-H3, P3B 13-D10, P4B 04-H3, P5B 14-C11, P4BSF 03-G10, and the wild type lipase from *P. aeruginosa* (the slopes have the unit [mOD/min]).

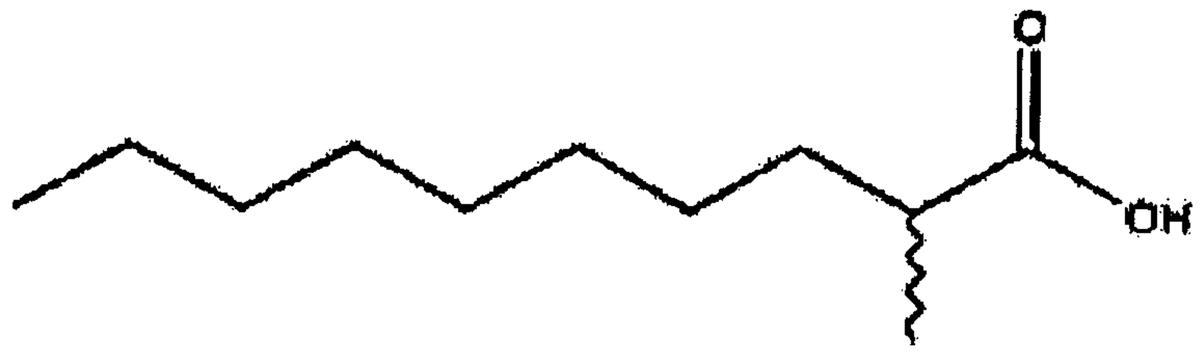
Figure 2: Comparison of the DNA sequences of the lipase mutants P1B 01-H1, P1B 01-E4, P2B 08-H3, P3B 13-D10, P4B 04-H3, P5B 14-C11 and P4BSF 03-G10 S155F with the sequence of the wild type of lipase from *P. aeruginosa* (the mutated bases with respect to the wild type are boxed, the origin of the mature lipase mutants is at base 163 or at base 162 in the wild type).

Example 1

In the following Example, the gene of the lipase from *P. aeruginosa* (isolation according to K.-E. Jäger, Ruhr-Universität Bochum) has been used for an optimization. The substrate for which the enantioselectivity of the lipase was to be improved was *(R,S)*-2-methyldecanoic acid. A

lipase mutant with a preference for the *(S)*-enantiomer was to be developed. The screening test was performed with *(R)*- and *(S)*-2-methyl-decanoic acid p-nitrophenyl ester.

Formula 5




*(R,S)*-2-methyldecanoic acid

Bacterial strains

*E. coli* JM109:

e14-(McrA), *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17($r_K$-$m_K$+), *sup*E44, *rel*A1, $\Delta$(*lac-pro*AB), [F' *tra*$\Delta$36 *pro*AB *lac*$I^q$ $Z\Delta M15$] (Stratagene)

*P. aeruginosa* PABST7.1:

lacUV5/lac$I^q$ controlled T7-polymerase gene stably integrated in the chromosome of strain *P. aeruginosa* PABS, which bears a deletion in the structural gene of lipase *lipA* (K.-E. Jaeger *et al., J. Mol. Cat. Part B*, 1997, in press)

Plasmids

pMut5: *BamHI/ApaI* fragment (1046 bp) of the *P. aeruginosa* lipase gene *lipA* in the vector pBluescript KSII (Stratagene)

pUCPL6A: *BamHI/HindIII* fragment (2.8 kbp) comprising the *P. aeruginosa* lipase operon in the vector pUCPKS (Watson *et al.*, *Gene* 1996, 172, 163) under the control of the T7 promoter

Culturing of bacteria

*E. coli* JM109 is grown over night (16 h) at 37 °C in 5 ml of LB medium on a test tube roller. For *P. aeruginosa* PABST7.1, 1 mM IPTG is added to the medium. For the screening test, *P. aeruginosa* PABST7.1 is grown in microtitration plates on a rotary shaker, the culture volume being 200 µl and the incubation being prolonged to 36-48 h. Antibiotics are added in the following concentrations:

*E. coli* JM109: ampicillin 100 µg/ml; *P. aeruginosa* PABST7.1: carbenicillin 200 µg/ml, tetracyclin 50 µg/ml

Mutagenizing PCR

The lipase gene *lipA* is amplified using the plasmid pMut5 linearized with endonuclease *Xmn* I as a template and the following PCR primers:

A: 5'-GCGCAATTAACCCTCACTAAAGGGAACAAA-3';

B: 5'-GCGTAATACGACTCACTATAGGGCGAA-3'

After purification of the PCR product using a Qiagen Qiaquick Column®, it serves as a template in a mutagenic PCR. The reaction conditions are as follows: a 100 µl reaction volume contains 16.6 mM $(NH_4)_2SO_4$; 67 mM Tris-HCl (pH 8.8); 6.1 mM $MgCl_2$; 6.7 µM EDTA (pH 8.0); 0.2 mM dNTPs; 10 mM mercaptoethanol; 10 µl of DMSO; 10 pmol each of the primers; 0.1 ng of template DNA; and 1 U of Taq polymerase (Goldstar, Eurogentec). The reaction volume is covered with a layer of 100 µl of paraffin. Ten parallel reactions were performed which were combined after completion of the reaction. The cycling protocol is as follows: A 2 min denaturation at 98 °C is followed by 25 cycles with 1 min at 94 °C, 2 min at 64 °C, 1 min at 72 °C on a Robocycler 40

(Stratagene), followed by incubation for 7 min at 72 °C. The Taq polymerase is added after the denaturation of the 1st cycle. The sequencing of the PCR products yields an error rate of about 1-2 base substitutions per 1000 bp.

Cloning of the PCR products

The PCR products are precipitated with ethanol and resuspended in distilled water. After restriction with *ApaI* and *BamHI*, the 1046 bp fragment formed is purified using a Qiagen Qiaquick Column® and ligated into the correspondingly prepared vector pUCPL6A using T4 DNA ligase (MBI Fermentas) for 2 h at room temperature. The reaction volume is diluted 1:5 and transformed into 200 µl of competent cells of *E. coli* JM109 prepared by the method of *Hanahan* (*J. Mol. Biol.* 1983, 166, 557). For this purpose, the DNA and cells are stored on ice for 1 h and incubated with shaking at 42 °C for 2 min and, after the addition of 700 µl of LB medium, at 37 °C for 45 min. The cell suspension is subsequently plated onto LB (ampicillin 100 µg/ml) plates. Sixty nanograms of the PCR product employed in the ligation reaction will yield about 1500 colonies. All colonies are resuspended in sterile LB medium, the plasmid DNA is purified and transformed into *P. aeruginosa* PABST7.1 by electroporation according to the method of *Farinha* and *Kropinski* (*FEMS Microbiol. Lett.* 1990, 70, 221). The 96 wells of the microtitration plates are inoculated with one colony each and treated as described in Culturing of bacteria. To obtain the culture supernatant, which is to be employed subsequently in the test for stereoselectivity, the microtitration plates are centrifuged at 4000 rpm for 30 min.

Test for stereoselectivity

The lipase-containing culture supernatants obtained by centrifugation are pipetted in two aliquots into adjacent wells of a microtitration plate. The test volume is 100 µl and is composed of the following components (Table 2):

Table 2

Composition of the reaction mixture in the test for improved enantioselectivity of lipase mutants

| (R) reaction | (S) reaction |
|---|---|
| 50 µl of culture supernatant | 50 µl of culture supernatant |
| 40 µl of 10 mM Tris/HCl buffer, pH 7.5 | 40 µl of 10 mM Tris/HCl buffer, pH 7.5 |
| 10 µl of substrate solution [10 mg/ml (R)-2-methyldecanoic acid p-nitrophenyl ester in DMF] | 10 µl of substrate solution [10 mg/ml (S)-2-methyldecanoic acid p-nitrophenyl ester in DMF] |

After the addition of the Tris/HCl buffer to the supernatants, the microtitration plate is incubated at 30 °C for about 5 min. After addition of the substrate solution, the reaction is continuously monitored for 10 min by spectrophotometry at 410 nm at 30 °C. From the linear rise of the absorption curve, which is a measure of the constant initial rate of the hydrolysis, the apparent enantioselectivity ($E_{app}$) is determined. Thus, the slopes measured in the linear region of the initial rates of the reactions for the pair of enantiomers are divided by one another to obtain the value of the apparent enantioselectivity of the corresponding lipase mutant.

Determination of stereoselectivity by gas chromatography

Selected positive clones are grown in 5 ml liquid cultures (LB medium), and after centrifugation and removal of the bacterial pellet, the lipase-containing supernatant is employed for the reaction. As the substrate,

100 µl of a solution of racemic *(R,S)*-2-methyldecanoic acid p-nitrophenyl ester (10 mg/ml in dimethylformamide) is used. This solution is admixed with 700 µl of 10 mM Tris/HCl buffer, pH 7.5. The reaction is started by adding 100 µl of culture supernatant and performed at 30 °C and 1000 rpm in Eppendorf reaction vessels. After 2.5 h, samples of 200 µl each are removed and transferred to an Eppendorf vessel filled with 200 µl of dichloromethane. After the addition of 25 µl of 20% aqueous hydrochloric acid, the products and educts are extracted (vortex shaker, 1 min). Finally, the organic phase is used for gas-chromatographic analysis (GC). Separation of the enantiomers of the free 2-methyldecanoic acid is achieved thereby.

Separation conditions of GC:

| | |
|---|---|
| Instrument: | Hewlett Packard 5890 |
| Column: | 25 m 2.6 DM 3 Pent β-CD/80% SE 54 |
| Detector: | FID |
| Temperature: | 230 °C inlet; 80-190 °C with 2 °C/min |
| Gas: | 0.6 bar $H_2$ |
| Sample quantity: | 0.1 ml |

Results (1st cycle)

Of the about 1000 clones examined which had been obtained by *mutagenizing PCR* from the starting DNA (wild type gene of lipase from *P. aeruginosa*), 12 were identified to have an improved enantioselectivity over that of the corresponding wild type enzyme. Finally, 3 clones were selected and their enantioselectivity determined by GC analysis.

Table 3

Selected lipase mutants with improved enantioselectivity (1st cycle)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calculated from GC) |
|---|---|---|---|---|---|
| Wild type | 21.8 | 14.9 | 1.5 | 2.4 / 15.3 | 1.1 |
| P1B 01-E4 | 128.4 | 43.2 | 3.0 | 36.1 / 23.2 | 2.4 |
| P1B 01-F12 | 78.8 | 35.7 | 2.2 | 14.1 / 30.5 | 1.4 |
| P1B 01-H1 | 158.7 | 56.2 | 2.8 | 37.6 / 4.5 | 2.2 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) $E = \ln[1-c(1+ee_P)]/\ln[1-c(1-ee_P)]$ with c = conversion, $ee_P$ = ee value of the product

The DNA of the clone P1B 01-E4 served as the starting point for a new cycle of PCR mutagenization. Thus, the plasmid pUCPL6A was isolated from the clone and transformed into *E. coli* JM109 as described above. After the preparation of the plasmid DNA, the 1046 bp fragment was obtained by restriction with *ApaI* and *BamHI* and subsequent purification and ligated into the correspondingly prepared plasmid pMut5. After transformation and plasmid isolation, this plasmid served as template DNA in a *mutagenizing PCR* under the conditions as described above. The DNA obtained from the *mutagenizing PCR* served to prepare a new mutant library (2nd generation).

Results (2nd cycle)

From the mutant library of the 2nd generation, about 2200 clones were used for the screening test. Ten mutants with an improved enantioselectivity over that of mutant P1B 01-E4 were identified. Two mutants (P2B 04-G11 and P2B 08-H3) were examined more closely by GC analysis.

Table 4

Selected lipase mutants with improved enantioselectivity (2nd cycle)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calculated from GC) |
|---|---|---|---|---|---|
| P2B 04-G11 | 224.9 | 52.3 | 4.3 | 47.8 / 30.0 | 3.4 |
| P2B 08-H3 | 310.8 | 67.4 | 4.6 | 56.6 / 19.3 | 4.1 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) $E = \ln[1-c(1+ee_p)]/\ln[1-c(1-ee_p)]$ with c = conversion, $ee_p$ = ee value of the product

Clone P2B 08-H3 was used for the next mutation cycle (3rd generation).

Results (3rd cycle)

From the mutant library of the 3rd generation, about 2400 clones were used for the screening test. One mutant (P3B 13-D10) with an improved enantioselectivity over that of mutant P2B 08-H3 was identified. It was examined further by GC analysis.

Table 5

Selected lipase mutants with improved enantioselectivity (3rd cycle)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calculated from GC) |
|---|---|---|---|---|---|
| P3B 13-D10 | 240.0 | 35.2 | 6.9 | 74.8 / 34.6 | 10.2 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) E = ln[1-c(1+ee$_P$)]/ln[1-c(1-ee$_P$)] with c = conversion, ee$_P$ = ee value of the product

Results (4th cycle)

From the mutant library of the 4th generation, about 2000 clones were used for the screening test. Four mutants with an improved enantioselectivity over that of mutant P3B 13-D10 were identified. They were examined further by GC analysis.

Table 6

Selected lipase mutants with improved enantioselectivity (4th cycle)

| Mutant | $V_{app}$(S) [mOD/min] | $V_{app}$(R) [mOD/min] | $E_{app}$ [1] | % ee (by GC)/ % conversion | E value [2] (calculated from GC) |
|---|---|---|---|---|---|
| P4B 04-H3 | 355.6 | 26.5 | 13.4 | 81.0 / 20.0 | 11.2 |
| P4B 01-F2 | 162.4 | 13.8 | 11.7 | 82.1 / 5.0 | 10.6 |
| P4B 15-G1 | 315.4 | 28.1 | 11.2 | 80.0 / 18.0 | 10.7 |
| P4B 15-H7 | 288.0 | 25.1 | 11.5 | 78.4 / 22.0 | 10.2 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) E = ln[1-c(1+ee$_P$)]/ln[1-c(1-ee$_P$)] with c = conversion, ee$_P$ = ee value of the product

The clone P4B04-H3 was inserted in the next mutation cycle (5$^{th}$ generation).

Results (5th cycle)

From the mutant library of the 5th generation, about 5200 clones were used for the screening test. Two mutants with an improved enantioselectivity over that of mutant P4B 04-H3 were identified. They were examined further by GC analysis.

Table 7

Selected lipase mutants with improved enantioselectivity (5th cycle)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calculated from GC) |
|---|---|---|---|---|---|
| P5B 14-C11 | 275.9 | 17.3 | 15.9 | 77.0 / 43.0 | 13.7 |
| P5B 08-F2 | 124.0 | 8.7 | 14.3 | 79.7 / 40.3 | 15.1 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) $E = \ln[1-c(1+ee_p)]/\ln[1-c(1-ee_p)]$ with c = conversion, $ee_p$ = ee value of the product

Sequencing of the positive mutants

By sequencing the positive mutants, the mutations could be localized within the lipase genes (see Figure 2). After assigning the base triplets to the corresponding amino acids, the following amino acid substitutions result with respect to the wild type lipase from *P. aeruginosa*:

P1B 01-H1: T103I (Thr103 → Ile103), S149G (Ser149 → Gly149)

P1B 01-E4: S149G (Ser149 → Gly149)

P2B 08-H3: S149G (Ser149 → Gly149), S155L (Ser155 → Leu155)

P3B 13-D10: S149G (Ser149 → Gly149), S155L (Ser155 → Leu155), V47G (Val47 → Gly47)

P4B 04-H3: S149G (Ser149 → Gly149), S155L (Ser155 → Leu155), V47G (Val47 → Gly47), S33N (Ser33 → Asn33), F259L (Phe259 → Leu259)

P5B 14-C11: S149G (Ser149 → Gly149), S155L (Ser155 → Leu155), V47G (Val47 → Gly47), S33N (Ser33 → Asn33), F259L (Phe259 → Leu259), K110R (Lys110 → Arg110)

Mutants P1B 01-E4, P2B 08-H3 and P3B 13-D10 were deposited on July 16, 1997, with the DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, D-38124 Braunschweig, Mascheroder Weg 1b, under the designations of DSM 11 658, DSM 11 659 and DSM 11 659, respectively.

Mutants P5B 14-C11 and P4B 04-H3 were deposited on July 20, 1998, with the DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, D-38124 Braunschweig, Mascheroder Weg 1b, under the designations of DSM 12 320 and DSM 12 322, respectively.

Example 2

The protocols for the culturing of the bacteria, the *mutagenizing PCR* and the test method for enantioselectivity are analogous to those of Example 1. However, in this Example, the preparation of extensive mutant libraries is effected by *in vitro* recombination.

The DNA used for the *in vitro* recombination is either generated by *mutagenizing PCR* or obtained by combining the DNA from any number of clones from one or more clone generations formed by repeated *mutagenizing PCR*. If the PCR products of a *mutagenizing PCR* are the starting point for obtaining DNA for the *in vitro* recombination, the procedure is as follows: The PCR products of the *mutagenizing PCR* (see Example 1) are purified, cleaved with the restriction endonucleases *Apa* I and *BamH* I, ligated into the correspondingly cleaved vector pMUTS and then transformed into *E. coli* JM 109. The plasmid DNA from all transformation clones is isolated. If some number of selected clones from one or more generations of mutant clones are the starting point for obtaining DNA for the *in vitro* recombination, then the plasmid DNA of the vector pMUT5 is isolated and combined with the respective variants of the lipase gene of P. *aeruginosa*. In both cases, the further

procedure is as follows: Restriction with the endonuclease *Pvu* II yields a 1430 bp fragment which comprises the binding sites of primers A and B already used in the *mutagenizing PCR*, in addition to the structural gene for the lipase from *P. aeruginosa*. This fragment is purified and cleaved into randomly generated fragments by incubation with deoxyribonuclease I (DNase I from bovine pancreas). The size of the fragments and the error rate of the subsequent reassembling can be influenced by selecting the incubation conditions.

DNase I treatment

In a total volume of 100 µl, 3 µg of *Pvu* II fragments in 50 mM Tris/HCl, pH 7.5, 10 mM $MgCl_2$ or 10 mM $MnCl_2$, respectively, and 50 µg/ml BSA is incubated at 23 °C with 0.075 U DNase I for 10-25 min or 1-10 min, respectively. The reaction is terminated by incubation at 93 °C for 10 min. Depending on the reaction time, fragments of smaller than 500 bp to smaller than 10 bp are obtained. In the case where only a particular range of sizes is used, these fragments can be obtained from agarose gels by selective electro-blotting on DEAE membrane (according to F.M. Ausubel *et al.*, eds., *Current Protocols in Molecular Biology*, John Wiley and Sons, 1989). After purification of the fragments by the Qiagen Nucleotide Removal Kit® (Qiagen), the following reassembling reaction is performed.

Reassembling reaction

10-30 ng of the fragments derived from the DNase I restriction are subjected to the following PCR cycles in 75 mM Tris/HCl, pH 9.0, 20 mM $(NH_4)_2SO_4$, 0.01% (w/v) Tween® 20, 1.5 mM $MgCl_2$, 0.2 mM dNTPs with 2 U Goldstar Taq polymerase (Eurogentec) in a total volume of 50 µl: 2 min at 94 °C, 40 cycles of 1 min at 94 °C, 2 min at 52 °C and 1 min

at 72 °C, finally 7 min at 72 °C. The Taq polymerase is added after the 1 minute denaturing step of the 1st cycle.

PCR

1 µl from the reassembling reaction is employed in a subsequent PCR reaction, which is composed as described for the reassembling reaction , with the following differences: instead of the DNase I generated fragments, 1 µl of the reassembling reaction is employed as the template DNA. In addition, primers A and B in a concentration of 0.2 mM and 10% dimethylsulfoxide are added. The cycling protocol is as follows:

2 min at 98 °C, 30 cycles of 1 min at 94°C, 2 min at 64 °C, 1 min at 72 °C and finally 7 min at 72 °C; parallel runs are performed. The PCR products formed in these reactions are purified, restricted with the Restriction endonucleases Apa I and Bam HI and cloned as described in the paragraph Mutagenizing PCR of Example 1.

Results (*in vitro* recombination):

Twelve clones of the 1st generation of the mutant library obtained by *mutagenizing PCR* (see Example 1) were used for the *in vitro* recombination. The following clones which had shown improved enantioselectivity in the screening test were used:

P1B 01-A2, P1B 01-A6, P1B 01-D2, P1B 01-D5, P1B 01-E1, P1B 01-E4, P1B 01-F3, P1B 01-F11, P1B 01-H1, P1B 01-H3, P1B 01-F12.

The DNA of these clones recombined according to the procedure described above is cloned as stated in the paragraph Mutagenizing PCR, and the culture supernatants are employed in the test for enanti-

oselectivity. About 1000 recombinant clones were tested. The two identified recombinants S2A 01-E11 and S2A 02-G3 exhibit a significant improvement of enantioselectivity over the best mutant of the 1st generation (P1B 01-E4) from Example 1.

Table 8

Selected lipase mutants with improved enantioselectivity (*in vitro* recombination)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calc.from GC) |
|---|---|---|---|---|---|
| S2A 01-E11 | 145.6 | 41.6 | 3.5 | 41.0 / 27.0 | 2.8 |
| S2A 02-G3 | 210.8 | 62.0 | 3.4 | 38.0 / 23.0 | 2.5 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) $E = \ln[1-c(1+ee_p)]/\ln[1-c(1-ee_p)]$ with c = conversion, $ee_p$ = ee value of the product

Example 3

Side-directed saturation mutagenesis in the amino acid position 155 of lipase mutant P3B 13-D10:

Plasmids:

pMut5 13D10: BamHI/ApaI fragment (1046 bp) of the gene of mutant P3B 13D10 for the lipase from *P. aeruginosa* in pBluescript KS II

pMut5$_{\Delta}$AK 13D10: Deletion of the AflIII/KpnI fragment in pMut5 13D10

A fragment of the gene for the lipase from mutant P3B 13D10 is amplified using plasmid pMut5 13D10, linearized by endonuclease XmnI, and the following PCR primers:

A: 5'-GCGCAATTAACCCTCACTAAAGGGAACAAA-3'

M: 5'-GGTACGCAGAATNNNCTGGGCTCGC-3'

where N represents A or C or G or T.

The reaction conditions are as follows: A 50 µl reaction volume contains 75 mM Tris/HCl, pH 9.0 (at 25 °C); 20 mM $(NH_4)_2SO_4$; 1.5 mM $MgCl_2$; 0.01% (w/v) Tween® 20; 10% (v/v) DMSO; 10 pmol of each of the primers; 0.1 ng of the template DNA; and 2 U of Taq polymerase (Goldstar, Eurogentec). The cycling protocol is as follows: A 2 min denaturation at 98 °C is followed by 30 cycles with 1 min at 94 °C, 2 min at 64 °C, 1 min at 72 °C on a Robocycler 40 (Stratagene), followed by incubation for 7 min at 72 °C. The Taq polymerase is added after the denaturation of the 1st cycle. After purification of the PCR products by agarose gel electrophoresis and elution of the DNA from the agarose gel using the Nucleospin Extract Kit (Macherey & Nagel), it was used as a primer (socalled megaprimer) in a subsequent PCR. Thus, the lipase gene is amplified on the plasmid pMut5$_{\Delta}$AK 13D10, linearized by endonuclease XmnI, using the following PCR primers and the above described reaction conditions:

A: 5'-GCGCAATTAACCCTCACTAAAGGGAACAAA-3'

B (megaprimer): 5'-GCGTAATACGACTCACTATAGGGCGAA-3'

The reaction conditions and the cycling protocol are as described above, except that 1-10 ng of the megaprimer is added to the reaction mixture. The cloning of the PCR products is effected as described under Cloning of the PCR Products.

Results (saturation mutagenesis, 3rd generation, P3B13-D10)

From the mutant library of the saturation mutagenesis (3rd generation, P3B 13-D10), about 900 clones were used for the screening test. One mutant (P4BSF 03-G10) with an improved enantioselectivity over that of mutant P3B 13-D10 was identified. It was examined further by GC analysis.

Table 9

Selected lipase mutant with improved enantioselectivity (3rd generation, P3B 13-D10)

| Mutant | $V_{app}(S)$ [mOD/min] | $V_{app}(R)$ [mOD/min] | $E_{app}$ [1]) | % ee (by GC)/ % conversion | E value [2]) (calculated from GC) |
|---|---|---|---|---|---|
| P4BSF 03-G10 | 384.7 | 25.3 | 15.2 | 87.3 / 19.0 | 20.4 |

1) $E_{app} = V_{app}(S)/V_{app}(R)$

2) $E = \ln[1-c(1+ee_P)]/\ln[1-c(1-ee_P)]$ with c = conversion, $ee_P$ = ee value of the product

Sequencing of the positive mutants

By sequencing the positive mutants, the mutations could be localized within the lipase gene (see Figure 2). After assigning the base triplets to the corresponding amino acids, the following amino acid substitution resulted with respect to mutant P3B 13-D10:

P4BSF 03-G10 : L155F (Leu155 → Phe155)

Mutant P4BSF 03-G10 was deposited on July 20, 1998, with the DSMZ - Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, D-38124 Braunschweig, Mascheroder Weg 1b, under the designation of DSM 12 321.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Studiengesellschaft Kohle mbH
(B) STREET: Kaiser-Wilhelm-Platz 1
(C) CITY: Muelheim an der Ruhr
(E) COUNTRY: Germany
(F) POSTAL CODE (ZIP): 45470

(ii) TITLE OF INVENTION: A Process for the Preperation and Identification of Novel Hydrolases Having Improved Properties

(iii) NUMBER OF SEQUENCES: 21

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCGCAATTAA CCCTCACTAA AGGGAACAAA 30

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCGTAATACG ACTCACTATA GGGCGAA 27

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC    60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TAT CTG CTC CCC CTC       111
                           Met Lys Lys Lys Tyr Leu Leu Pro Leu
                           -26 -25                 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG      159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
       -15                 -10                 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC      207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
  1               5                  10                  15

ATG CTC GGC TTC GAC AAC ATC CTC GGG GTC GAC TAC TGG TTC GGC ATT      255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
             20                  25                  30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GTC      303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Val
         35                  40                  45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG      351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
     50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC      399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
 65                  70                  75

GCC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT      447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
 80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ATC AGC GTC GGC GCC CCG CAC AAG CGT      495
Pro Asp Leu Ile Ala Ser Ala Ile Ser Val Gly Ala Pro His Lys Gly
                100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCC GGT TCG GCC GGC      543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
            115                 120                 125

GAG GCA GTC CTC TCC GCC CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC      591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
        130                 135                 140

TTC CTT TCC AGC GGC GGC ACC GGT ACG CAG AAT TCA CTG GGC TCG CTG      639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Ser Leu Gly Ser Leu
    145                 150                 155

GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG      687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175
```

```
CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC      735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC      783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
            195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG      831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG      879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
    225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC TTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                 1049
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
Met Lys Lys Lys Tyr Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                  -5                   1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
                10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
            25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Val Ser Gln Leu Asp Thr Ser Glu
        40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
    55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
                75                  80                  85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
            90                  95                  100
```

```
Ile Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
        105                 110                 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
    120                 125                 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135                 140                 145                 150

Gly Thr Gln Asn Ser Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
                155                 160                 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
            170                 175                 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
        185                 190                 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
    200                 205                 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215                 220                 225                 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
                235                 240                 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
            250                 255                 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
        265                 270                 275

Arg Leu Lys Asn Ala Ser Leu
    280                 285
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC   60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC      111
                           Met Lys Lys Lys Ser Leu Leu Pro Leu
                           -26 -25                 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG     159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
        -15                 -10                 -5
```

```
GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC     207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
  1               5                  10                  15

ATG CTC GGC TTC GAC AAC ATC CTC GGG GTC GAC TAC TGG TTC GGC ATT     255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
              20                  25                  30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GTC     303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Val
          35                  40                  45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG     351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
      50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC     399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
  65                  70                  75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT     447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT     495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
              100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC     543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
          115                 120                 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC     591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
      130                 135                 140

TTC CTT TCC AGC GGC GGC ACC GGT ACG CAG AAT TCA CTG GGC TCG CTG     639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Ser Leu Gly Ser Leu
  145                 150                 155

GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG     687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC     735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
              180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC     783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
          195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG     831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
      210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG     879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
  225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG     927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255
```

```
AAC CAG GTC TTC GCC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                  1049
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                  -5                  1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
            10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
        25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Val Ser Gln Leu Asp Thr Ser Glu
    40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
                75                  80                  85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
            90                  95                  100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
        105                 110                 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
    120                 125                 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135                 140                 145                 150

Gly Thr Gln Asn Ser Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
                155                 160                 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
            170                 175                 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
        185                 190                 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
    200                 205                 210
```

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215 220 225 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
235 240 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
250 255 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
265 270 275

Arg Leu Lys Asn Ala Ser Leu
280 285

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC 60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC 111
Met Lys Lys Lys Ser Leu Leu Pro Leu
-26 -25 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG 159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
-15 -10 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC 207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
1 5 10 15

ATG CTC GGC TTC GAC AAC ATC CTT GGG GTC GAC TAC TGG TTC GGC ATT 255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
20 25 30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GTC 303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Val
35 40 45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG 351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
50 55 60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC 399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
65 70 75

```
GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT      447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
 80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT      495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
                100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC      543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
            115                 120                 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC      591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
        130                 135                 140

TTC CTT TCC AGC GGC GGC ACC GGT ACG CAG AAT TTA CTG GGC TCG CTG      639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Leu Leu Gly Ser Leu
    145                 150                 155

GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG      687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC      735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC      783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
            195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG      831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG      879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
    225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC TTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                 1049
```

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                 -5                  1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
                10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
            25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Val Ser Gln Leu Asp Thr Ser Glu
        40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
                75                  80                  85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
            90                  95                  100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
        105                 110                 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
    120                 125                 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135                 140                 145                 150

Gly Thr Gln Asn Leu Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
                155                 160                 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
            170                 175                 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
        185                 190                 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
    200                 205                 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215                 220                 225                 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
                235                 240                 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
            250                 255                 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
        265                 270                 275

Arg Leu Lys Asn Ala Ser Leu
    280                 285
```

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1047 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:84..1016

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:162..1016

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

```
GGATCCCCGG TTCTCCCGGA AGGATTCGGG CGATGGCTGG CAGGACGCGC CCCTCGGCCC     60

CATCAACCTG AGATGAGAAC AAC ATG AAG AAG AAG TCT CTG CTC CCC CTC          110
                          Met Lys Lys Lys Ser Leu Leu Pro Leu
                          -26 -25                 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG        158
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
      -15                 -10                 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC        206
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
  1               5                  10                  15

ATG CTC GGC TTC GAC AAC ATC CTC GGG GTC GAC TAC TGG TTC GGC ATT        254
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
              20                  25                  30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GTC        302
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Val
          35                  40                  45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG        350
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
      50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC        398
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
  65                  70                  75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT        446
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
 80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT        494
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
             100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC        542
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
         115                 120                 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC ACC        590
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
     130                 135                 140

TTC CTT TCC AGC GGC AGC ACC GGT ACG CAG AAT TCA CTG GGC TCG CTG        638
Phe Leu Ser Ser Gly Ser Thr Gly Thr Gln Asn Ser Leu Gly Ser Leu
 145                 150                 155
```

```
GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG      686
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC      734
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC      782
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
            195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG      830
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG      878
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
225                 230                 235


GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      926
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC TTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      974
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1016
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC C                                  1047
```

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                 -5                  1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
            10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
        25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Val Ser Gln Leu Asp Thr Ser Glu
    40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
```

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
75 80 85 90 95 100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
105 110 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
120 125 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Ser Thr
135 140 145 150

Gly Thr Gln Asn Ser Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
155 160 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
170 175 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
185 190 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
200 205 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215 220 225 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
235 240 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
250 255 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
265 270 275

Arg Leu Lys Asn Ala Ser Leu
280 285

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC 60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC 111

```
                                  Met Lys Lys Lys Ser Leu Leu Pro Leu
                                  -26 -25                 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG     159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
        -15                 -10                 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC     207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
      1               5                  10                  15

ATG CTC GGC TTC GAC AAC ATC CTT GGG GTC GAC TAC TGG TTC GGC ATT     255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
                  20                  25                  30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GGC     303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Gly
              35                  40                  45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG     351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
          50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC     399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
      65                  70                  75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT     447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
  80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT     495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
                 100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC     543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
             115                 120                 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC     591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
         130                 135                 140

TTC CTT TCC AGC GGC CGC ACC GGT ACG CAG AAT TTA CTG GGC TCG CTG     639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Leu Leu Gly Ser Leu
     145                 150                 155

GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG     687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
 160                 165                 170                 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC     735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                 180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC     783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
             195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG     831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
         210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG     879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
     225                 230                 235
```

```
GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC TTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG               1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
        275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                  1049
```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                 -5                  1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
                10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
            25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Gly Ser Gln Leu Asp Thr Ser Glu
        40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
    55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
                75                  80                  85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
            90                  95                  100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
        105                 110                 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
    120                 125                 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135                 140                 145                 150

Gly Thr Gln Asn Leu Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
                155                 160                 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
            170                 175                 180
```

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
185 190 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
200 205 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215 220 225 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
235 240 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
250 255 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
265 270 275

Arg Leu Lys Asn Ala Ser Leu
280 285

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1050 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC 60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC 111
Met Lys Lys Lys Ser Leu Leu Pro Leu
-26 -25 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG 159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
-15 -10 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC 207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
1 5 10 15

ATG CTC CCC TTC GAC AAC ATC CTT GGG GTC GAC TAC TGG TTC GGC ATT 255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
20 25 30

CCC AAC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GGC 303
Pro Asn Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Gly
35 40 45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG 351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln

```
          50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC     399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
 65                  70                  75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT     447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
 80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT     495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
                100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC     543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
            115                 120                 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC     591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
        130                 135                 140

TTC CTT TCC AGC GGC GGC ACC GGT ACG CAG AAT TTA CTG GGC TCG CTG     639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Leu Leu Gly Ser Leu
    145                 150                 155

GAG TCG CTG AAC AGC GAG GGT CCC GCG CGC TTC AAC GCC AAG TAC CCG     687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175

CAG CGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCT TAC AAG GTC AAC     735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC     783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
            195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG     831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG     879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
    225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG     927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC CTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC     975
Asn Gln Val Leu Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG             1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CCG                               1050
```

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25 -20 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10 -5 1 5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
10 15 20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Asn Ala Leu Arg Arg Asp
25 30 35

Gly Ala Gln Val Tyr Val Thr Glu Gly Ser Gln Leu Asp Thr Ser Glu
40 45 50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
55 60 65 70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
75 80 85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
90 95 100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
105 110 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
120 125 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135 140 145 150

Gly Thr Gln Asn Leu Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
155 160 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
170 175 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
185 190 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
200 205 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215 220 225 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
235 240 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Leu Gly Leu Thr
250 255 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
265 270 275

Arg Leu Lys Asn Ala Ser Leu
280 285

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

```
GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC    60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC      111
                           Met Lys Lys Lys Ser Leu Leu Pro Leu
                           -26 -25                 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG     159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
        -15                 -10                 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC     207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
  1               5                  10                  15

ATG CTC GGC TTC GAC AAC ATC CTT GGG GTC GAC TAC TGG TTC GGC ATT     255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
            20                  25                  30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GGC     303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Gly
        35                  40                  45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG     351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
    50                  55                  60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC     399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
65                  70                  75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT     447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
 80                  85                  90                  95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AGG GGT     495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Arg Gly
            100                 105                 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC     543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
```

```
                    115                 120                 125
GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTC ATC AGC      591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
        130                 135                 140

TTC CTT TCC AGC GGC GGC ACC GGT ACG CAG AAT TTA CTG GGC TCG CTG      639
Phe Leu Ser Ser Gly Gly Thr Gly Thr Gln Asn Leu Leu Gly Ser Leu
    145                 150                 155

GAG TCG CTG AAC AGT GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG      687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160                 165                 170                 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCT TAC AAG GTC AAC      735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
                180                 185                 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCG CCG CTG ACC AAC TTC      783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe
            195                 200                 205

CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG      831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG      879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
    225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC CTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Leu Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                 1049
```

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                  -5                  1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
                10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
            25                  30                  35
```

Gly Ala Gln Val Tyr Val Thr Glu Gly Ser Gln Leu Asp Thr Ser Glu
40 45 50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
55 60 65 70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
75 80 85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
90 95 100

Thr Ser Val Gly Ala Pro His Arg Gly Ser Asp Thr Ala Asp Phe Leu
105 110 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
120 125 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Thr
135 140 145 150

Gly Thr Gln Asn Leu Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
155 160 165

Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
170 175 180

Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
185 190 195

Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
200 205 210

Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215 220 225 230

Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
235 240 245

Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Leu Gly Leu Thr
250 255 260

Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
265 270 275

Arg Leu Lys Asn Ala Ser Leu
280 285

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1049 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..1017

(ix) FEATURE:
(A) NAME/KEY: mat_peptide

(B) LOCATION:163..1017

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGATCCCCCG GTTCTCCCGG AAGGATTCGG GCGATGGCTG GCAGGACGCG CCCCTCGGCC 60

CCATCAACCT GAGATGAGAA CAAC ATG AAG AAG AAG TCT CTG CTC CCC CTC 111
Met Lys Lys Lys Ser Leu Leu Pro Leu
-26 -25 -20

GGC CTG GCC ATC GGT CTC GCC TCT CTC GCT GCC AGC CCT CTG ATC CAG 159
Gly Leu Ala Ile Gly Leu Ala Ser Leu Ala Ala Ser Pro Leu Ile Gln
-15 -10 -5

GCC AGC ACC TAC ACC CAG ACC AAA TAC CCC ATC GTG CTG GCC CAC GGC 207
Ala Ser Thr Tyr Thr Gln Thr Lys Tyr Pro Ile Val Leu Ala His Gly
1 5 10 15

ATG CTC GGC TTC GAC AAC ATC CTT GGG GTC GAC TAC TGG TTC GGC ATT 255
Met Leu Gly Phe Asp Asn Ile Leu Gly Val Asp Tyr Trp Phe Gly Ile
20 25 30

CCC AGC GCC TTG CGC CGT GAC GGT GCC CAG GTC TAC GTC ACC GAA GGC 303
Pro Ser Ala Leu Arg Arg Asp Gly Ala Gln Val Tyr Val Thr Glu Gly
35 40 45

AGC CAG TTG GAC ACC TCG GAA GTC CGC GGC GAG CAG TTG CTG CAA CAG 351
Ser Gln Leu Asp Thr Ser Glu Val Arg Gly Glu Gln Leu Leu Gln Gln
50 55 60

GTG GAG GAA ATC GTC GCC CTC AGC GGC CAG CCC AAG GTC AAC CTG ATC 399
Val Glu Glu Ile Val Ala Leu Ser Gly Gln Pro Lys Val Asn Leu Ile
65 70 75

GGC CAC AGC CAC GGC GGG CCG ACC ATC CGC TAC GTC GCC GCC GTA CGT 447
Gly His Ser His Gly Gly Pro Thr Ile Arg Tyr Val Ala Ala Val Arg
80 85 90 95

CCC GAC CTG ATC GCT TCC GCC ACC AGC GTC GGC GCC CCG CAC AAG GGT 495
Pro Asp Leu Ile Ala Ser Ala Thr Ser Val Gly Ala Pro His Lys Gly
100 105 110

TCG GAC ACC GCC GAC TTC CTG CGC CAG ATC CCA CCG GGT TCG GCC GGC 543
Ser Asp Thr Ala Asp Phe Leu Arg Gln Ile Pro Pro Gly Ser Ala Gly
115 120 125

GAG GCA GTC CTC TCC GGG CTG GTC AAC AGC CTC GGC GCG CTG ATC AGC 591
Glu Ala Val Leu Ser Gly Leu Val Asn Ser Leu Gly Ala Leu Ile Ser
130 135 140

TTC CTT TCC AGC GGC GGC ATC GGT ACG CAG AAT TTT CTG GGC TCG CTG 639
Phe Leu Ser Ser Gly Gly Ile Gly Thr Gln Asn Phe Leu Gly Ser Leu
145 150 155

GAG TCG CTG AAC AGC GAG GGT GCC GCG CGC TTC AAC GCC AAG TAC CCG 687
Glu Ser Leu Asn Ser Glu Gly Ala Ala Arg Phe Asn Ala Lys Tyr Pro
160 165 170 175

CAG GGC ATC CCC ACC TCG GCC TGC GGC GAA GGC GCC TAC AAG GTC AAC 735
Gln Gly Ile Pro Thr Ser Ala Cys Gly Glu Gly Ala Tyr Lys Val Asn
180 185 190

GGC GTG AGC TAT TAC TCC TGG AGC GGT TCC TCC CCG CTG ACC AAC TTC 783
Gly Val Ser Tyr Tyr Ser Trp Ser Gly Ser Ser Pro Leu Thr Asn Phe

```
                195                 200                 205
CTC GAT CCG AGC GAC GCC TTC CTC GGC GCC TCG TCG CTG ACC TTC AAG      831
Leu Asp Pro Ser Asp Ala Phe Leu Gly Ala Ser Ser Leu Thr Phe Lys
        210                 215                 220

AAC GGC ACC GCC AAC GAC GGC CTG GTC GGC ACC TGC AGT TCG CAC CTG      879
Asn Gly Thr Ala Asn Asp Gly Leu Val Gly Thr Cys Ser Ser His Leu
    225                 230                 235

GGC ATG GTG ATC CGC GAC AAC TAC CGG ATG AAC CAC CTG GAC GAG GTG      927
Gly Met Val Ile Arg Asp Asn Tyr Arg Met Asn His Leu Asp Glu Val
240                 245                 250                 255

AAC CAG GTC TTC GGC CTC ACC AGC CTG TTC GAG ACC AGC CCG GTC AGC      975
Asn Gln Val Phe Gly Leu Thr Ser Leu Phe Glu Thr Ser Pro Val Ser
                260                 265                 270

GTC TAC CGC CAG CAC GCC AAC CGC CTG AAG AAC GCC AGC CTG              1017
Val Tyr Arg Gln His Ala Asn Arg Leu Lys Asn Ala Ser Leu
            275                 280                 285

TAGGACCCCG GCCGGGGCCT CGGCCCGGGC CC                                 1049
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
Met Lys Lys Lys Ser Leu Leu Pro Leu Gly Leu Ala Ile Gly Leu Ala
-26 -25                 -20                 -15

Ser Leu Ala Ala Ser Pro Leu Ile Gln Ala Ser Thr Tyr Thr Gln Thr
-10                  -5                   1                   5

Lys Tyr Pro Ile Val Leu Ala His Gly Met Leu Gly Phe Asp Asn Ile
                10                  15                  20

Leu Gly Val Asp Tyr Trp Phe Gly Ile Pro Ser Ala Leu Arg Arg Asp
            25                  30                  35

Gly Ala Gln Val Tyr Val Thr Glu Val Ser Gln Leu Asp Thr Ser Glu
        40                  45                  50

Val Arg Gly Glu Gln Leu Leu Gln Gln Val Glu Glu Ile Val Ala Leu
    55                  60                  65                  70

Ser Gly Gln Pro Lys Val Asn Leu Ile Gly His Ser His Gly Gly Pro
                75                  80                  85

Thr Ile Arg Tyr Val Ala Ala Val Arg Pro Asp Leu Ile Ala Ser Ala
            90                  95                  100

Thr Ser Val Gly Ala Pro His Lys Gly Ser Asp Thr Ala Asp Phe Leu
        105                 110                 115

Arg Gln Ile Pro Pro Gly Ser Ala Gly Glu Ala Val Leu Ser Gly Leu
    120                 125                 130

Val Asn Ser Leu Gly Ala Leu Ile Ser Phe Leu Ser Ser Gly Gly Ile
```

```
135                 140                 145                 150
Gly Thr Gln Asn Phe Leu Gly Ser Leu Glu Ser Leu Asn Ser Glu Gly
                155                 160                 165
Ala Ala Arg Phe Asn Ala Lys Tyr Pro Gln Gly Ile Pro Thr Ser Ala
            170                 175                 180
Cys Gly Glu Gly Ala Tyr Lys Val Asn Gly Val Ser Tyr Tyr Ser Trp
        185                 190                 195
Ser Gly Ser Ser Pro Leu Thr Asn Phe Leu Asp Pro Ser Asp Ala Phe
    200                 205                 210
Leu Gly Ala Ser Ser Leu Thr Phe Lys Asn Gly Thr Ala Asn Asp Gly
215                 220                 225                 230
Leu Val Gly Thr Cys Ser Ser His Leu Gly Met Val Ile Arg Asp Asn
                235                 240                 245
Tyr Arg Met Asn His Leu Asp Glu Val Asn Gln Val Phe Gly Leu Thr
            250                 255                 260
Ser Leu Phe Glu Thr Ser Pro Val Ser Val Tyr Arg Gln His Ala Asn
        265                 270                 275
Arg Leu Lys Asn Ala Ser Leu
    280                 285
```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GCGCAATTAA CCCTCACTAA AGGGAACAAA 30

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GGTACGCAGA ATNNNCTGGG CTCGC 25

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "synthetic DNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GCGTAATACG ACTCACTATA GGGCGAA

27

# CLAIMS:

1. A process for the preparation and identification of hydrolase mutants having improved properties with respect to stereo- or regioselectivity, characterized in that
a) a starting hydrolase gene is mutagenized by a modified polymerase chain reaction (PCR), wherein the mutation rate and total number of mutations in the amplified DNA is adjusted by adjusting the concentrations of $Mg^{2+}$, $Mn^{2+}$ and of the deoxynucleotides and by adjusting the number of cycles;
b) optionally one or more hydrolase genes mutated according to step a), or mixtures of one or more starting hydrolase genes and one or more hydrolase genes mutated according to step a) are mutagenized by enzymatically fragmenting said genes, followed by enzymatic reassembly of the fragments produced to give complete recombinant hydrolase genes;
c) the mutated hydrolase genes obtained according to step a) or b) are transformed into a host organism; and
d) hydrolase mutants having improved properties, expressed by transformants obtained in step c), are identified by a test method.
2. The process according to claim 1, wherein an average mutation rate of 1-2 base substitutions, per one hydrolase gene to be mutagenized, is adjusted in the PCR in step a) by adjusting the concentrations of $Mg^{2+}$, $Mn^{2+}$ and of the deoxynucleotides.
3. The process according to claim 1, wherein a hydrolase gene mutagenized in a PCR previously performed according to claim 1 is used as the starting hydrolase gene in step a).

4. The process according to claim 1, wherein the enzymatic fragmentation of the hydrolase genes in step b) is performed using a deoxyribonuclease.

5. The process according to claim 1, wherein the reassembly of the fragments in step b) is effected enzymatically by means of a thermostable DNA polymerase using temperature cycles in which the parameters of temperature and duration of cycles are adjusted.

6. The process according to claim 1, wherein the mutation rate is adjusted during the enzymatic reassembly in step b) by adjusting the concentrations of $Mg^{2+}$, $Mn^{2+}$ and of the deoxynucleotides.

7. The process according to claim 1, wherein the completely recombined hydrolase genes are amplified by a polymerase chain reaction in step b) after completion of the reassembly reaction.

8. The process according to claim 1, wherein either modified hydrolase genes obtained from step a) according to claim 1 or 2 or several hydrolase genes mutagenized according to claim 3 are subjected to fragmentation and reassembly in step b).

9. The process according to claim 1, wherein synthetically prepared gene fragments are additionally used for the reassembly in step b).

10. The process according to claim 1, wherein hydrolase gene fragments from different organisms sharing a sequence homology of at least 60% can be used for the reassembly in step b).

11. The process according to claim 2 or 6, wherein the hydrolase mutants are lipase or esterase mutants, and the concentration of the magnesium ions is from 1.5 to 8.0 mM, preferably from 5.8 to 6.4 mM, and the concentration of the manganese ions is from 0.0 to 3.0 mM, preferably < 0.3 mM.

12. The process according to claim 2 or 6, wherein the hydrolase mutants are lipase or esterase mutants, and the concentration of the deoxynucleotide triphosphates is from 0.05 to 1.0 mM, preferably 0.2 mM.

13. The process according to claim 1, wherein for the test for stereo- or regioselectivity of the hydrolase mutants in step d), a test substrate is provided with a chromophorous group which causes a spectrometrically determined change of absorption or emission upon cleavage by the catalyst, and equal amounts of the hydrolase mutants are added to the pure stereo- or regioisomers of the test substrate in separate test vessels, and the stereo- or regioselectivity can be determined from the ratio of the linear initial reaction rates obtained.

14. The process according to claim 13, wherein the stereo- or regioisomers of a compound with a UV/VIS-active or fluorescence-active molecular group bound through a carboxylic acid ester or carboxylic acid amide linkage are used as the test substrate.

15. The process according to claim 14, wherein said UV/VIS-active molecular group is a p-nitrophenyl residue.

16. The process according to claim 1, wherein the test for stereo- or regioselectivity in step d) is effected through determination of the change of concentration with time of free fatty acids or succinic acid, wherein the corresponding stereo- or regioisomeric carboxylic acid esters or amides are hydrolyzed in separate vessels by means of the hydrolase mutants to give free fatty acids or succinic acid.

17. The process according to claim 1, wherein the test for stereo- or regioselectivity in step d) is effected through measuring the radioactivity, wherein the hydrolase mutants are reacted with stereo- or regioisomers having different radioactive labels in one functional group, and wherein the mixture of the stereo- or regioisomers is fixed on a support.

18. The process according to claim 17, wherein one of the stereo- or regioisomers of the support-bound mixture of isomeric compounds is labeled with the radioisotope $^{3}H$, and the other stereo- or regioisomer is labeled with the radioisotope $^{14}C$.

19. The process according to claim 1, wherein the test for stereoselectivity in step d) is effected through the capillary-electrophoretic determination of the reaction products and educts of a test reaction, the separation of the stereoisomeric reaction products and educts being performed in chirally modified capillaries.

20. The process according to claims 13 to 19, wherein several reactions are performed in parallel in microtitration plates.

21. The process according to claim 1, wherein the position of the codon coding for the changed amino acid is localized by sequencing in the mutants having improved properties identified in step d), followed

by generating a set of hydrolase genes with all possible codons for this position by means of site-directed saturation mutagenesis, and the mutated hydrolase genes thus obtained are further treated in analogy with steps c) and d) of claim 1.

22. The process according to claim 21, wherein the localization of the position of the codon coding for the changed amino acid is effected through DNA sequencing.

23. A hydrolase mutant obtainable by a process according to one or more of claims 1 to 22.

24. The hydrolase mutant according to claim 23 which is a lipase mutant.

25. The hydrolase mutant according to claim 23 which is an esterase mutant.

26. The hydrolase mutant according to claim 24 which is a lipase mutant of the starting lipase from the strain *P. aeruginosa*.

27. The hydrolase mutant according to claim 26 which is obtainable by expression from the transformants P1B 01-E4 (DSM 11 658), P2B 08-H3 (DSM 11 659), P3B 13-D10 (DSM 11 660), P4B 04-H3 (DSM 12 322), P5B 14-C11 (DSM 12 320) or P4BSF 03-G10 (DSM 12 321).

28. The hydrolase mutant according to claim 24 which has the amino acid sequence of the mature proteins shown in SEQ ID NOS. 4, 6, 8, 12, 14, 16 or 18.

29. A DNA sequence coding for a hydrolase mutant according to one or more of claims 23 to 28.

30. The DNA sequence according to claim 29 which comprises a DNA sequence shown in SEQ ID NOS. 3, 5, 7, 11, 13, 15 or 17.

31. A vector comprising a DNA sequence according to claim 29 or 30.

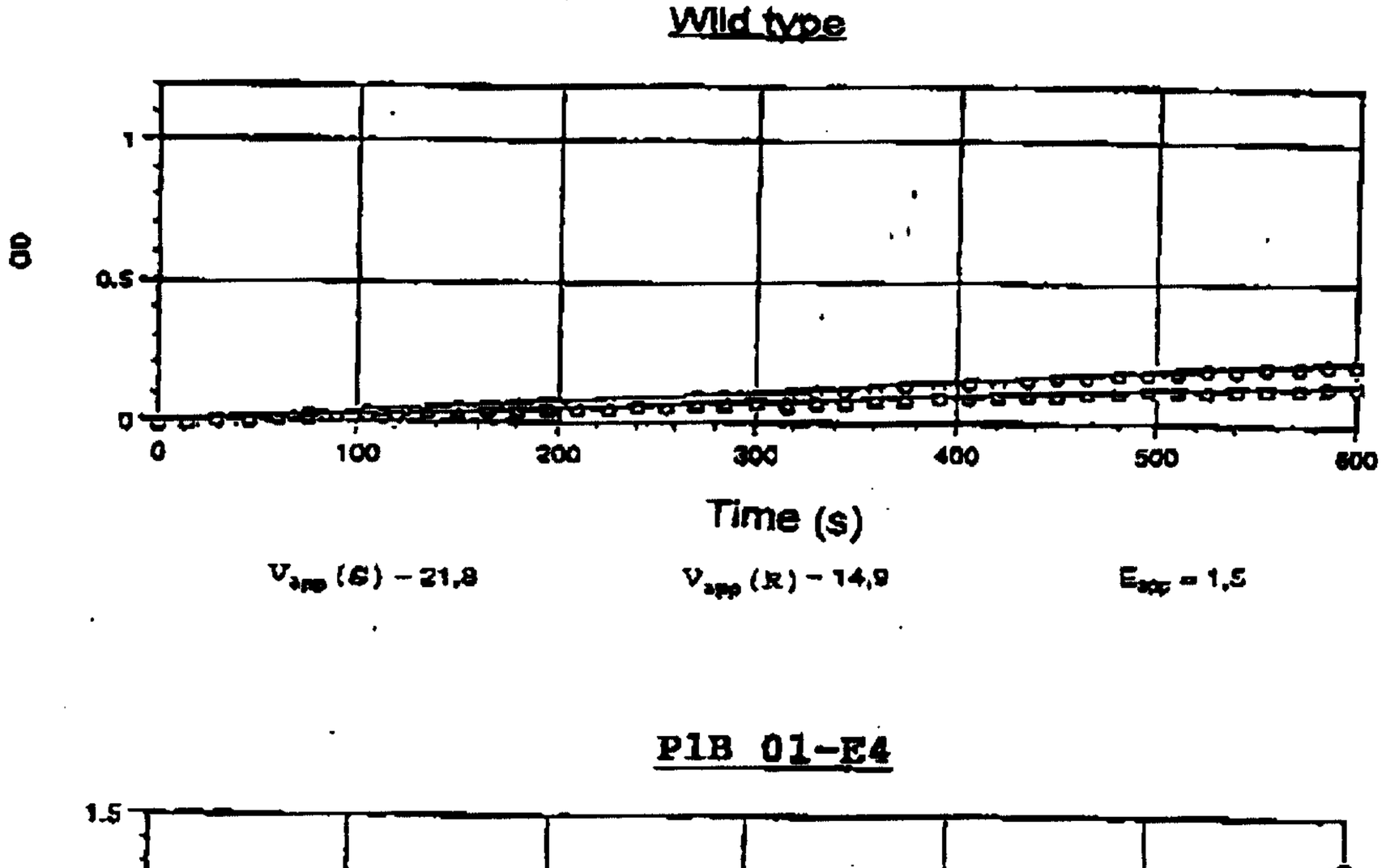
32. A transformant comprising a DNA sequence according to claim 29 or 30 and/or a vector according to claim 31.

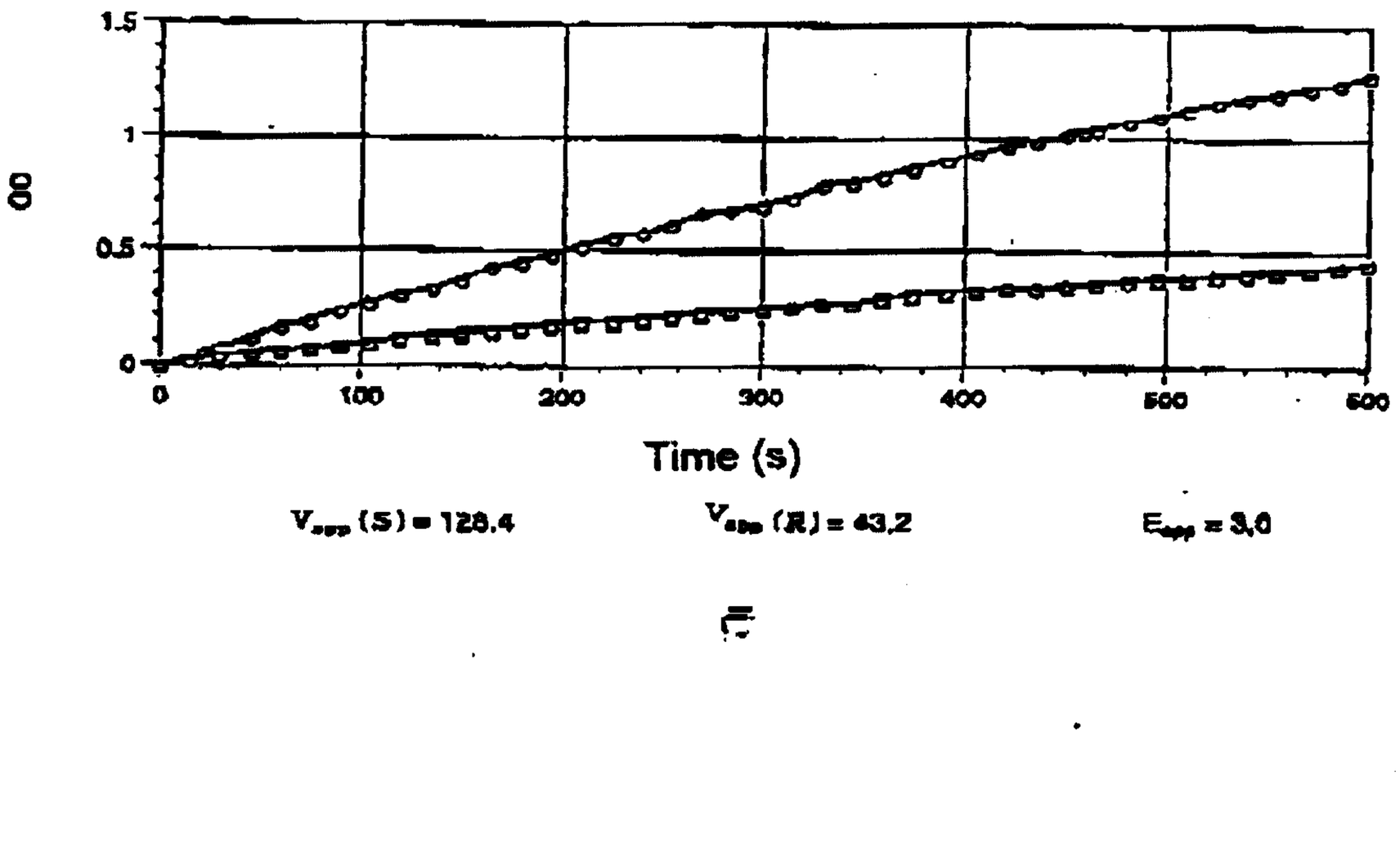
33. The transformant according to claim 32 which is transformant P1B 01-E4 (DSM 11 658), P2B 08-H3 (DSM 11 659), P3B 13-D10 (DSM 11 660), P4B 04-H3 (DSM 12 322), P5B 14-C11 (DSM 12 320) or P4BSF 03-G10 (DSM 12 321).

34. A process for the preparation of hydrolase mutants having improved properties, comprising culturing a transformant according to claim 32 or 33.

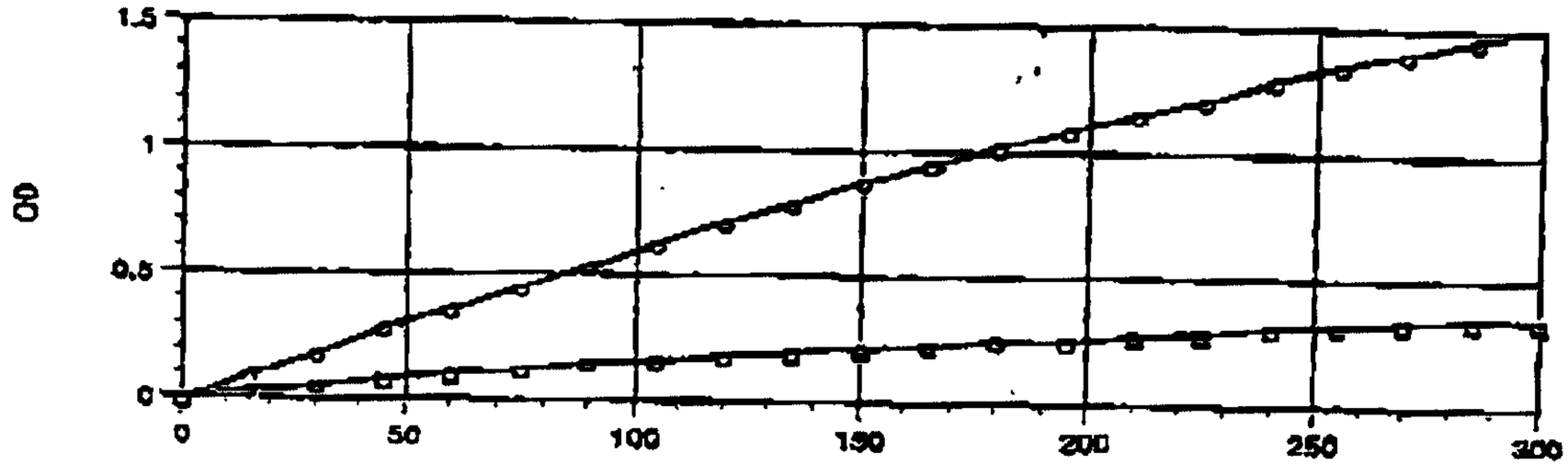
35. A method for testing catalysts for stereo- or regioselectivity, wherein equal amounts of the catalyst are added to a test substrate and to the pure stereo- or regioisomers of the test substrate, provided with a chromophorous group which causes a spectrometrically determinable change of absorption or emission upon cleavage by the catalyst, in separate test vessels, and the stereo- or regioselectivity is determined from the ratio of the linear initial reaction rates obtained.

Fig. 1


Wild type
OD
1
0.5
0
0
100
200
300
400
500
600
Time (s)
$V_{app}$ (S) = 21,8
$V_{app}$ (R) = 14,9
$E_{app}$ = 1,5

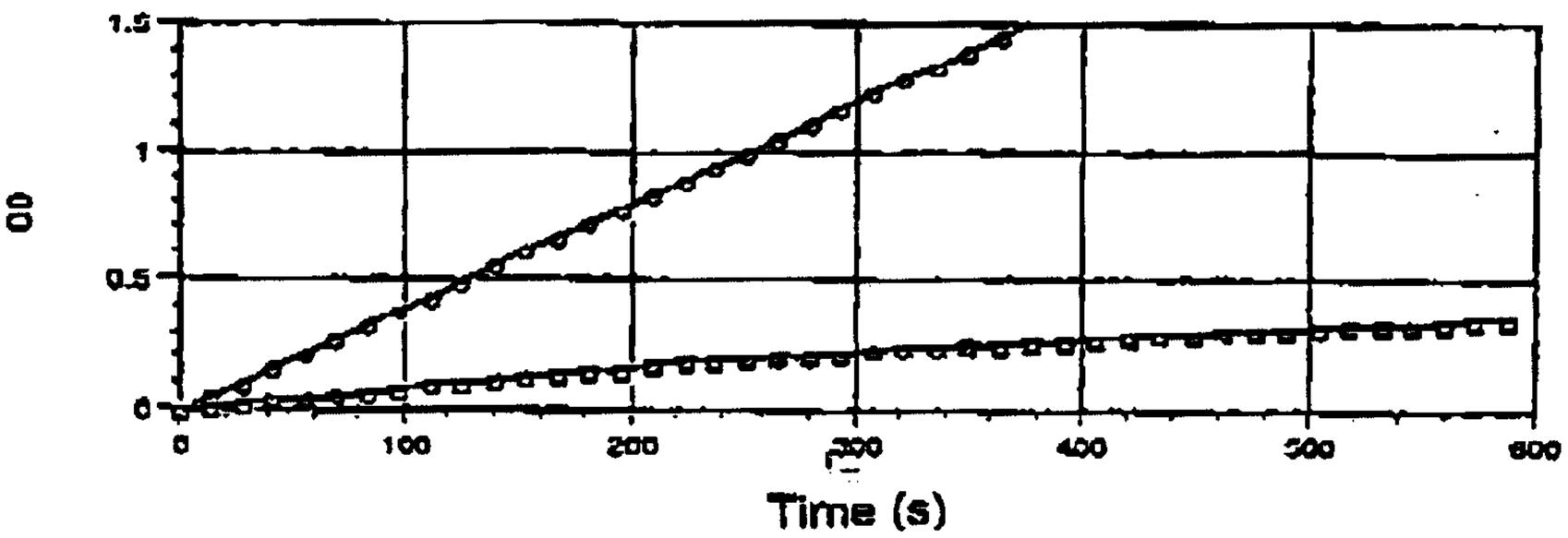

P1B 01-E4
OD
1.5
1
0.5
0
0
100
200
300
400
500
600
Time (s)
$V_{app}$ (S) = 128.4
$V_{app}$ (R) = 43,2
$E_{app}$ = 3,0

**P2B 08-H3**



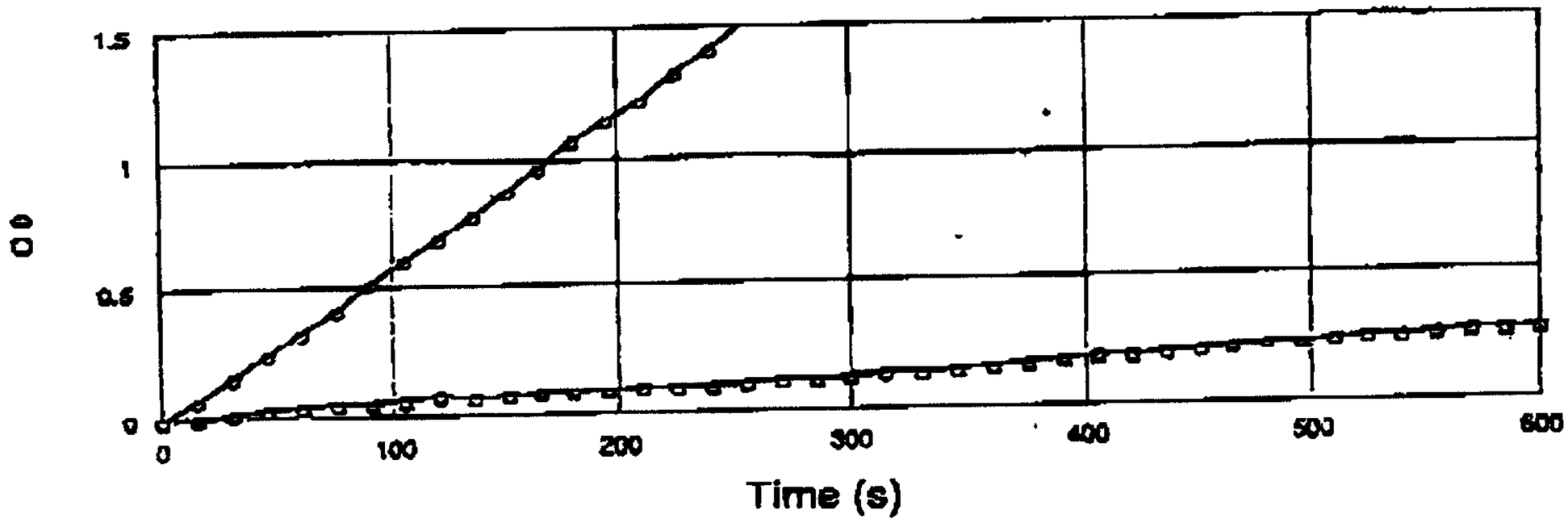

$V_{app}$ (S) = 310,8 $V_{app}$ (R) = 67,4 $E_{corr}$ = 4,6

**P3B 13-D10**



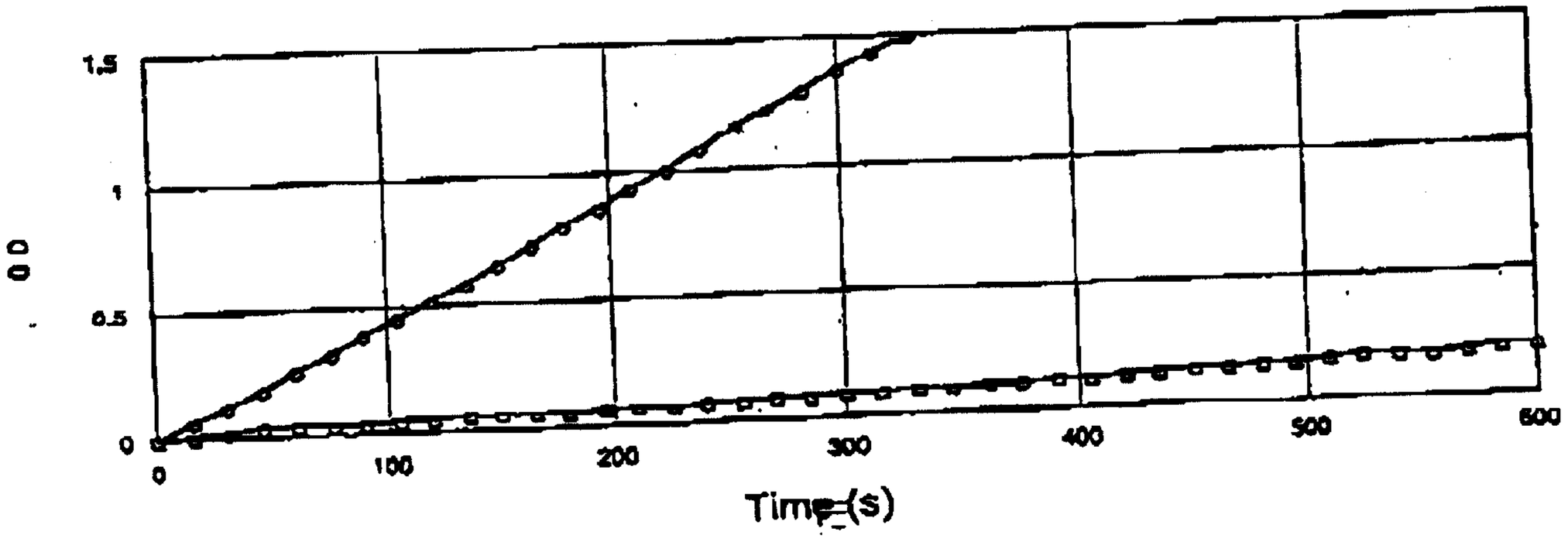

$V_{app}$ (S) = 241,1 $V_{app}$ (R) = 25,2 $E_{app}$ = 6,9

**P4B 04-H3**



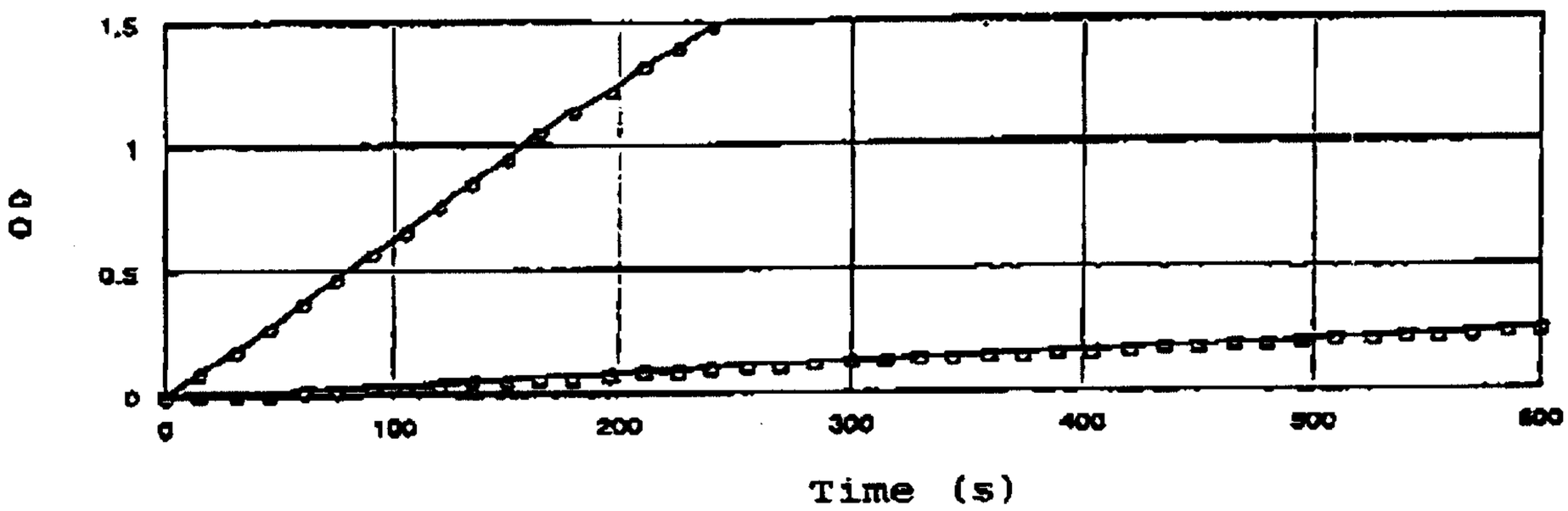

$V_{app}$ (S) = 355,6 $V_{app}$ (R) = 26,5 $E_{app}$ = 13,4

**P5B 14-C11**



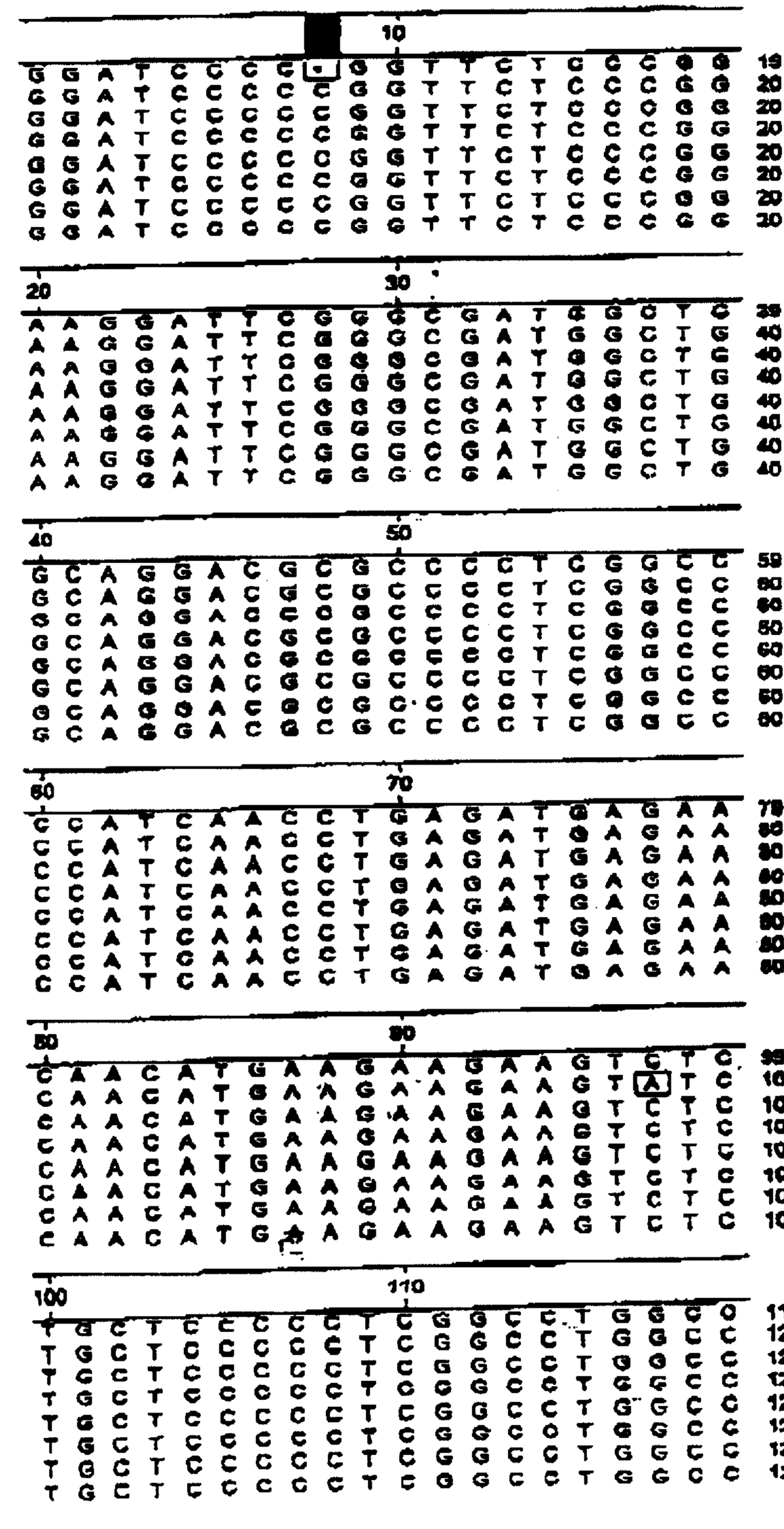

$V_{app}$ (S) = 275,9 $V_{app}$ (R) = 17.3 $E_{app}$ = 15,9

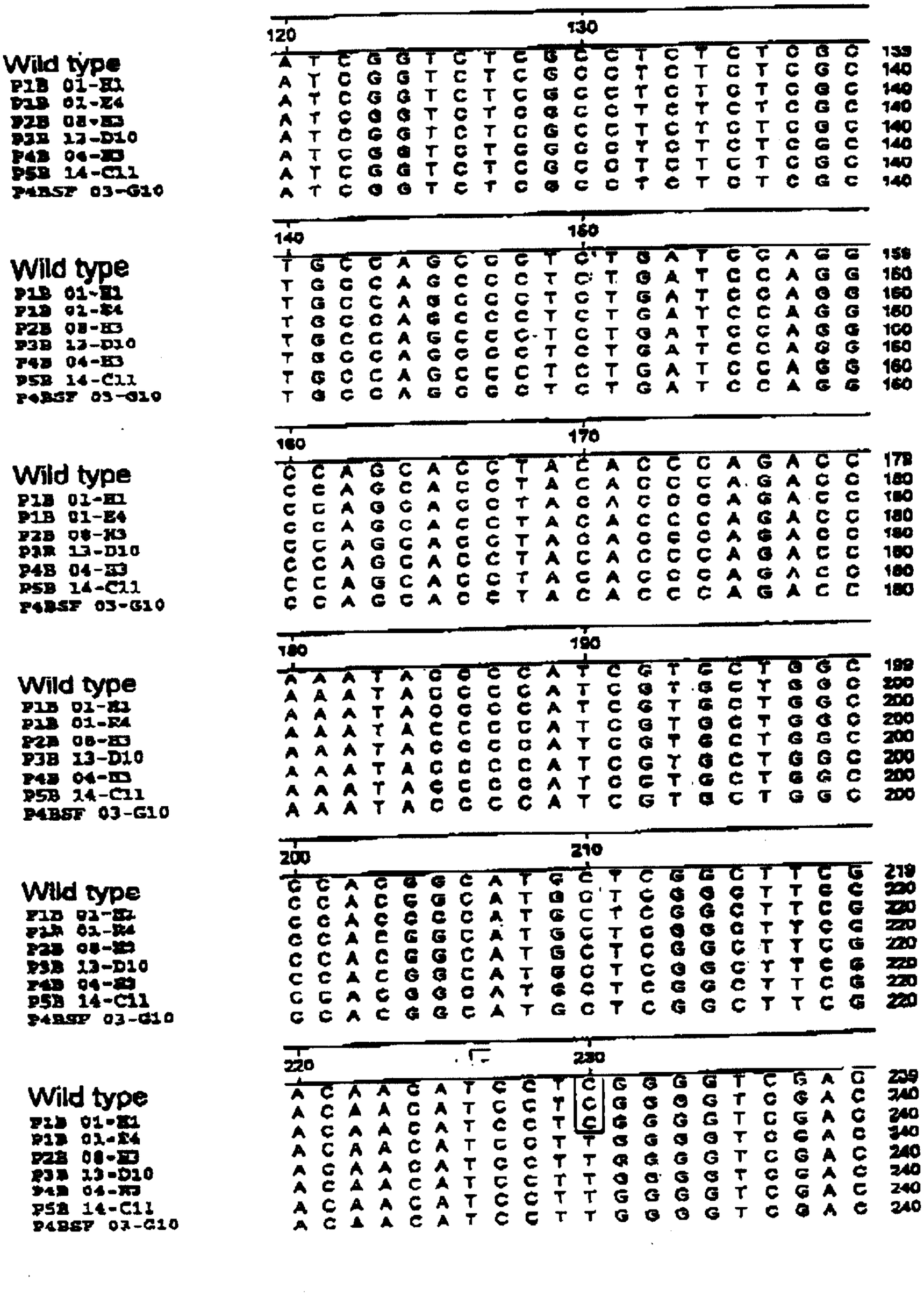
**P4BSF 03-G10**

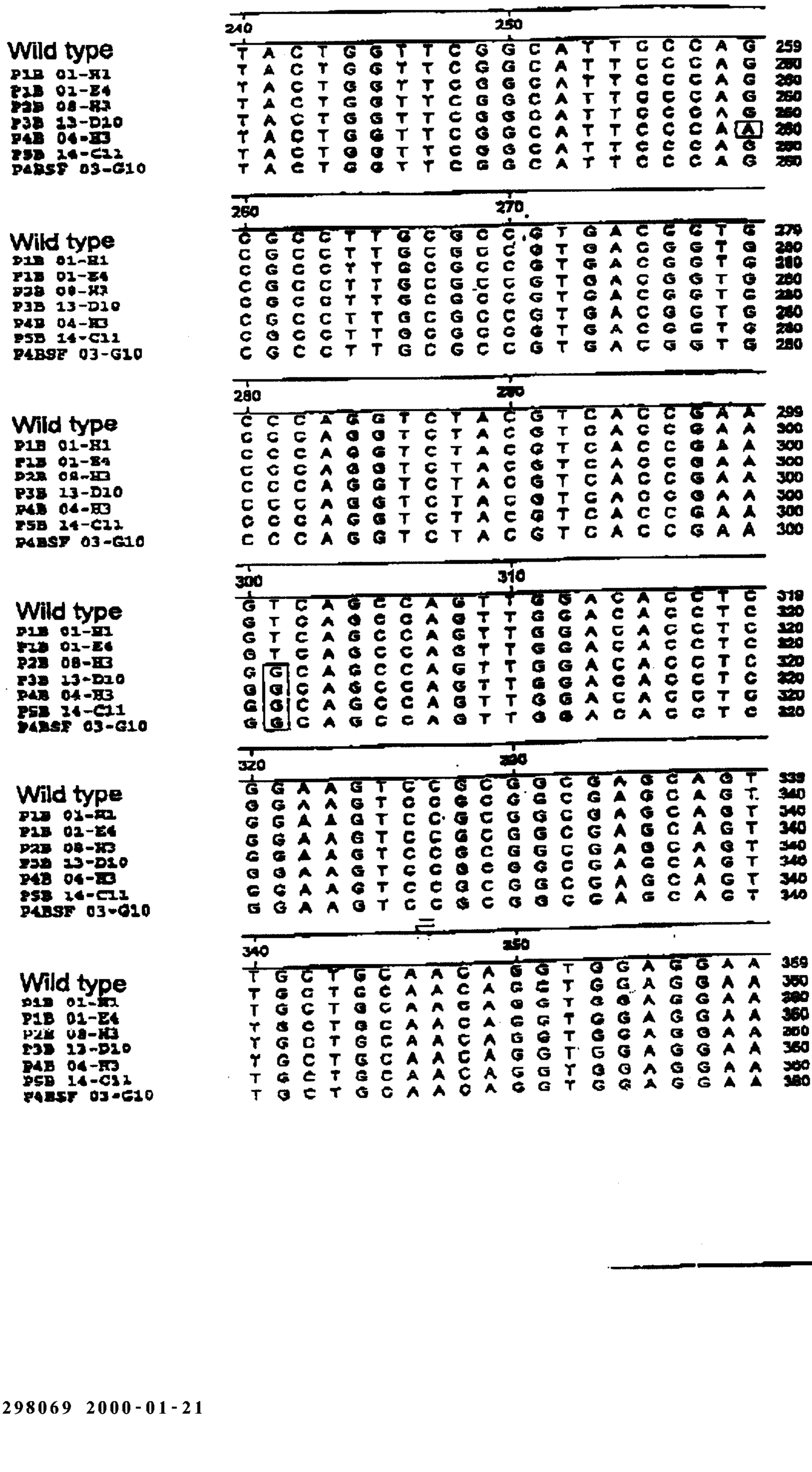


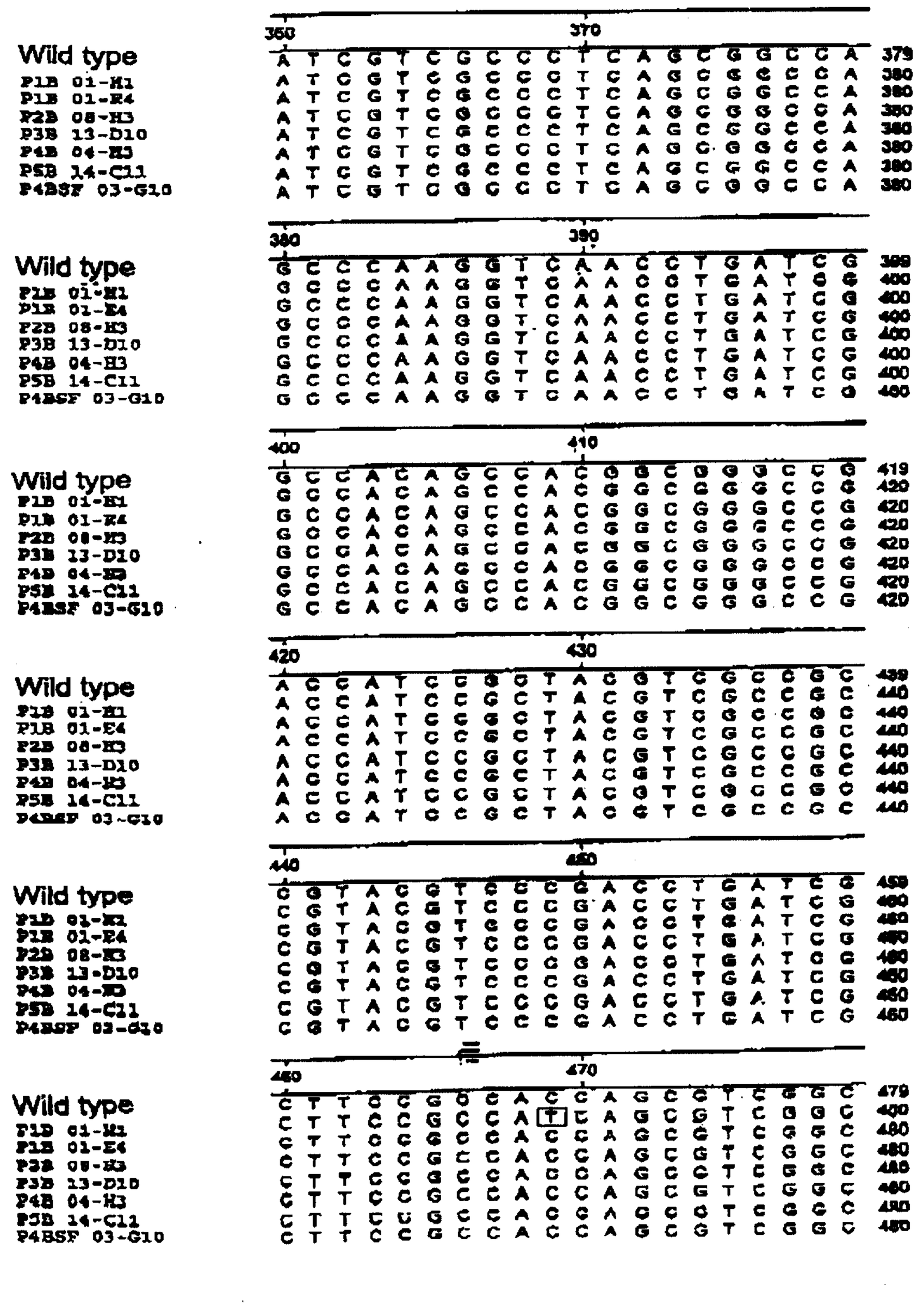

$V_{app}$ (S) = 384.7 $V_{app}$ (R) = 25.3 $E_{app}$ = 15.2

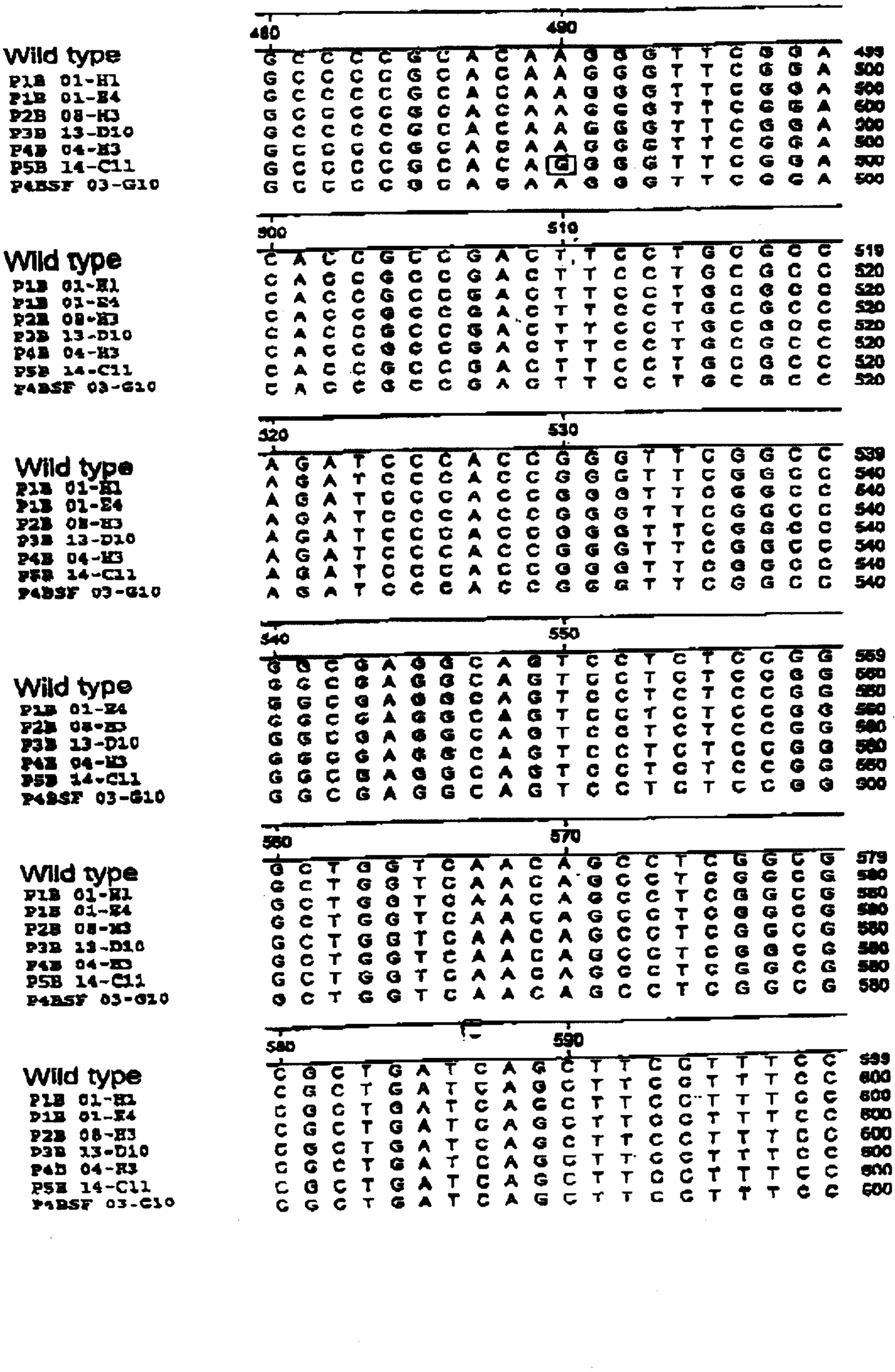
Fig. 2

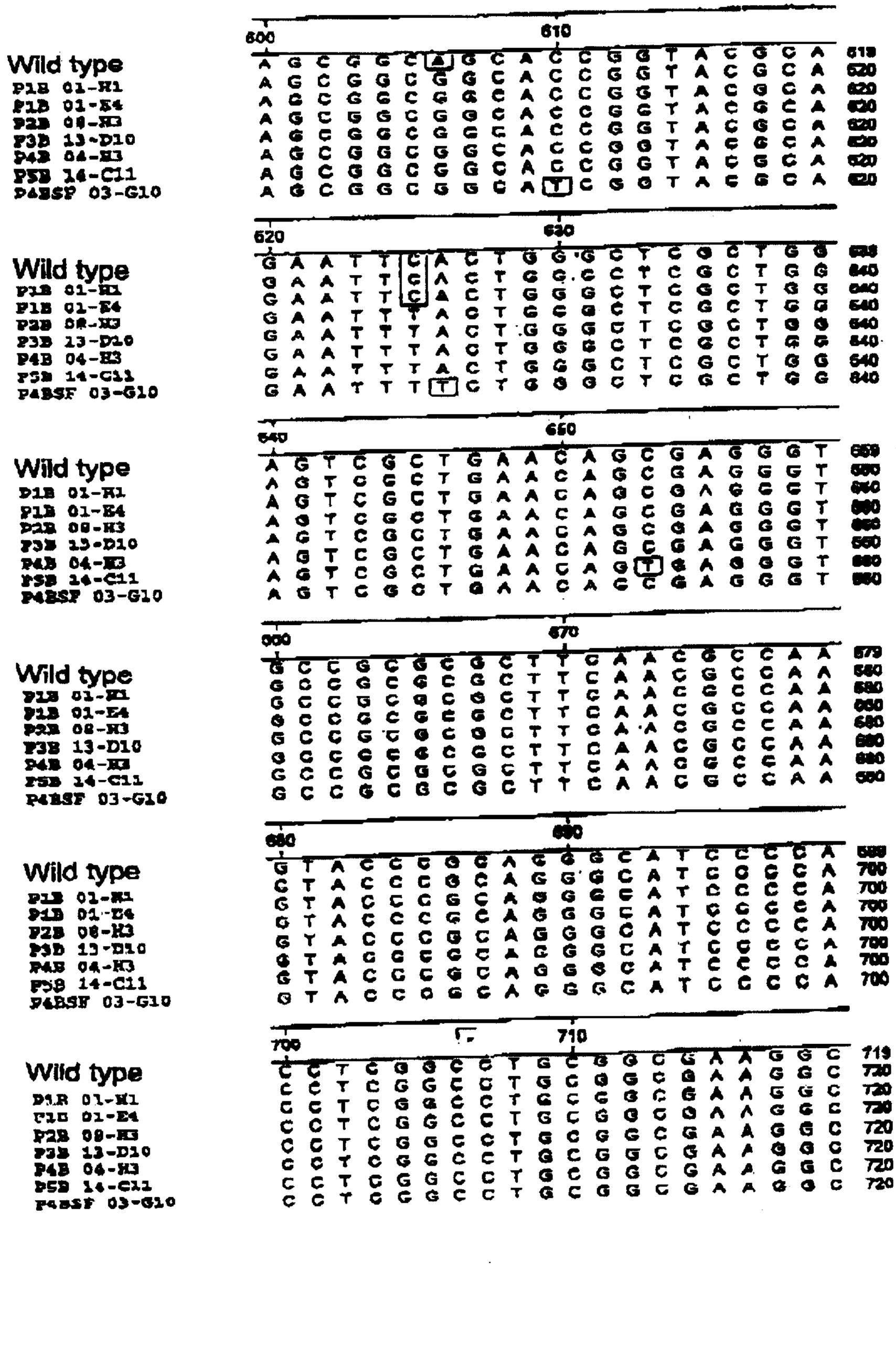

Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4B6F 03-G10
10
20
30
40
50
60
70
80
90
100
110


120
130
140
150
160
170
180
190
200
210
220
230
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10

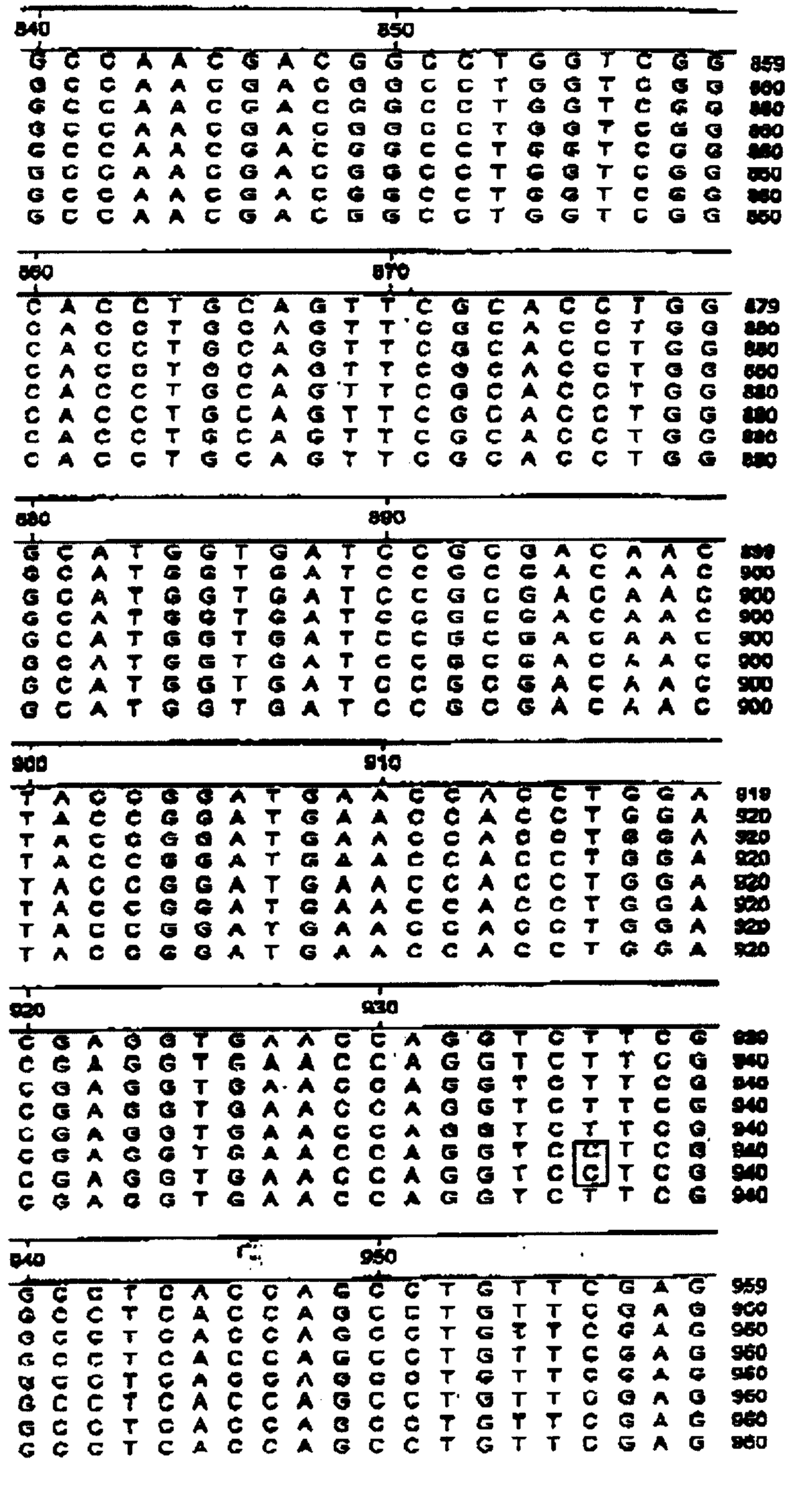

240
250
260
270
280
290
300
310
320
330
340
350
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10


360
370
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
379
380
390
400
410
419
420
430
439
440
450
459
460
470
479


480
490
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
500
510
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
520
530
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
540
550
Wild type
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
560
570
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
580
590
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10


600
610
Wild type
P1B 01-H1
P1B 01-E4
P2B 08-H3
P3B 13-D10
P4B 04-H3
P5B 14-C11
P4BSF 03-G10
620
630
640
650
660
670
680
690
700
710

```
                720                                730
Wild type       G C C T A C A A G G T C A A C G G C G T  739
P1B 01-H1       G C C T A C A A G G T C A A C G G C G T  740
P1B 01-E4       G C C T A C A A G G T C A A C G G C G T  740
P2B 08-H3       G C C T A C A A G G T C A A C G G C G T  740
P3B 13-D10      G C C T A C A A G G T C A A C G G C G T  740
P4B 04-H3       G C T T A C A A G G T C A A C G G C G T  740
P5B 14-C11      G C T T A C A A G G T C A A C G G C G T  740
P4B3F 03-G10    G C C T A C A A G G T C A A C G G C G T  740

                740                                750
Wild type       G A G C T A T T A C T C C T G G A G C G  759
P1B 01-H1       G A G C T A T T A C T C C T G G A G C G  760
P1B 01-E4       G A G C T A T T A C T C C T G G A G C G  760
P2B 08-H3       G A G C T A T T A C T C C T G G A G C G  760
P3B 13-D10      G A G C T A T T A C T C C T G G A G C G  760
P4B 04-H3       G A G C T A T T A C T C C T G G A G C G  760
P5B 14-C11      G A G C T A T T A C T C C T G G A G C G  760
P4B3F 03-G10    G A G C T A T T A C T C C T G G A G C G  760

                760                                770
Wild type       G T T C C T C G C C G C T G A C C A A C  779
P1B 01-H1       G T T C C T C G C C G C T G A C C A A C  780
P1B 01-E4       G T T C C T C G C C G C T G A C C A A C  780
P2B 08-H3       G T T C C T C G C C G C T G A C C A A C  780
P3B 13-D10      G T T C C T C G C C G C T G A C C A A C  780
P4B 04-H3       G T T C C T C G C C G C T G A C C A A C  780
P5B 14-C11      G T T C G T C G C C G C T G A C C A A C  780
P4B3F 03-G10    G T T C C T C G C C G C T G A C C A A C  780

                780                                790
Wild type       T T C C T C G A T C C G A G C G A C G C  799
P1B 01-H1       T T C C T C G A T C C G A G C G A C G C  800
P1B 01-E4       T T C C T C G A T C C G A G C G A C G C  800
P2B 08-H3       T T C C T C G A T C C G A G C G A C G C  800
P3B 13-D10      T T C C T C G A T C C G A G C G A C G C  800
P4B 04-H3       T T C C T C G A T C C G A G C G A C G C  800
P5B 14-C11      T T C C T C G A T C C G A G C G A C G C  800
P4B3F 03-G10    T T C C T C G A T C C G A G C G A C G C  800

                800                                810
Wild type       C T T C C T C G G C G C T C C T C G T C  819
P1B 01-H1       C T T C C T C G G C G C T C C T C G T C  820
P1B 01-E4       C T T C C T C G G C G C T C C T C G T C  820
P2B 08-H3       C T T C C T C G G C G C T C C T C G T C  820
P3B 13-D10      C T T C C T C G G C G C T C C T C G T C  820
P4B 04-H3       C T T C C T C G G C G C T C C T C G T C  820
P5B 14-C11      C T T C C T C G G C G C T C C T C G T C  820
P4B3F 03-G10    C T T C C T C G G C G C T C C T C G T C  820

                820                                830
Wild type       T G A C C T T C A A G A A C G G C A C C  839
P1B 01-H1       T G A C C T T C A A G A A C G G C A C C  840
P1B 01-E4       T G A C C T T C A A G A A C G G C A C C  840
P2B 08-H3       T G A C C T T C A A G A A C G G C A C C  840
P3B 13-D10      T G A C C T T C A A G A A C G G C A C C  840
P4B 04-H3       T G A C C T T C A A G A A C G G C A C C  840
P5B 14-C11      T G A C C T T C A A G A A C G G C A C C  840
P4B3F 03-C10    T G A C C T T C A A G A A C G G C A C C  840
```


840 850
Wild type G C C A A C G A C G G C C T G G T C G G 859
P1B 01-H1 G C C A A C G A C G G C C T G G T C G G 860
P1B 01-E4 G C C A A C G A C G G C C T G G T C G G 860
P2B 08-H3 G C C A A C G A C G G C C T G G T C G G 860
P3B 13-D10 G C C A A C G A C G G C C T G G T C G G 860
P4B 04-H3 G C C A A C G A C G G C C T G G T C G G 860
P5B 14-C11 G C C A A C G A C G G C C T G G T C G G 860
P4BSF 03-G10 G C C A A C G A C G G C C T G G T C G G 860
860 870
Wild type C A C C T G C A G T T C G C A C C T G G 879
P1B 01-H1 C A C C T G C A G T T C G C A C C T G G 880
P1B 01-E4 C A C C T G C A G T T C G C A C C T G G 880
P2B 08-H3 C A C C T G C A G T T C G C A C C T G G 880
P3B 13-D10 C A C C T G C A G T T C G C A C C T G G 880
P4B 04-H3 C A C C T G C A G T T C G C A C C T G G 880
P5B 14-C11 C A C C T G C A G T T C G C A C C T G G 880
P4BSF 03-G10 C A C C T G C A G T T C G C A C C T G G 880
890 900
Wild type G C A T G G T G A T C C G C G A C A A C 899
P1B 01-H1 G C A T G G T G A T C C G C G A C A A C 900
P1B 01-E4 G C A T G G T G A T C C G C G A C A A C 900
P2B 08-H3 G C A T G G T G A T C C G C G A C A A C 900
P3B 13-D10 G C A T G G T G A T C C G C G A C A A C 900
P4B 04-H3 G C A T G G T G A T C C G C G A C A A C 900
P5B 14-C11 G C A T G G T G A T C C G C G A C A A C 900
P4BSF 03-G10 G C A T G G T G A T C C G C G A C A A C 900
910 919
Wild type T A C C G G A T G A A C C C A C C T G G A
P1B 01-H1 T A C C G G A T G A A C C C A C C T T G G A 920
P1B 01-E4 T A C C G G A T G A A C C C A C C T T G G A 920
P2B 08-H3 T A C C G G A T G A A C C C A C C T T G G A 920
P3B 13-D10 T A C C G G A T G A A C C C A C C T T G G A 920
P4B 04-H3 T A C C G G A T G A A C C C A C C T T G G A 920
P5B 14-C11 T A C C G G A T G A A C C C A C C T T G G A 920
P4BSF 03-G10 T A C C G G A T G A A C C C A C C T T G G A 920
930 939
Wild type C G A G G T G A A C C A G G T C T T C G
P1B 01-H1 C G A G G T G A A C C A G G T C T T C G 940
P1B 01-E4 C G A G G T G A A C C A G G T C T T C G 940
P2B 08-H3 C G A G G T G A A C C A G G T C T T C G 940
P3B 13-D10 C G A G G T G A A C C A G G T C T T C G 940
P4B 04-H3 C G A G G T G A A C C A G G T C T T C G 940
P5B 14-C11 C G A G G T G A A C C A G G T C T T C G 940
P4BSF 03-G10 C G A G G T G A A C C A G G T C T T C G 940
940 950
Wild type G C C T C A C C A G C C T G T T C G A G 959
P1B 01-H1 G C C T C A C C A G C C T G T T C G A G 960
P1B 01-E4 G C C T C A C C A G C C T G T T C G A G 960
P2B 08-H3 G C C T C A C C A G C C T G T T C G A G 960
P3B 13-D10 G C C T C A C C A G C C T G T T C G A G 960
P4B 04-H3 G C C T C A C C A G C C T G T T C G A G 960
P5B 14-C11 G C C T C A C C A G C C T G T T C G A G 960
P4BSF 03-G10 G C C T C A C C A G C C T G T T C G A G 960

960 970

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wild type | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 979 |
| P1B 01-H1 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P1B 01-E4 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P2B 08-H3 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P3B 13-D10 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P4B 04-H3 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P5B 14-C11 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |
| P4BSF 03-G10 | A | C | C | A | G | C | C | C | G | G | T | C | A | G | C | G | T | C | T | A | 980 |

990 1000

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wild type | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 999 |
| P1B 01-H1 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P1B 01-E4 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P2B 08-H3 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P3B 13-D10 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P4B 04-H3 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P5B 14-C11 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |
| P4BSF 03-G10 | C | C | G | C | C | A | G | C | A | C | G | C | C | A | A | C | C | G | C | C | 1000 |

1000 1010

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wild type | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1019 |
| P1B 01-H1 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P1B 01-E4 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P2B 08-H3 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P3B 13-D10 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P4B 04-H3 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P5B 14-C11 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |
| P4BSF 03-G10 | T | G | A | A | G | A | A | C | G | C | C | A | G | C | C | T | G | T | A | G | 1020 |

1020 1030

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wild type | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1039 |
| P1B 01-H1 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P1B 01-E4 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P2B 08-H3 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P3B 13-D10 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P4B 04-H3 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P5B 14-C11 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |
| P4BSF 03-G10 | G | A | C | C | C | C | G | G | C | C | G | G | G | G | C | C | T | C | G | G | 1040 |

1040

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Wild type | C | C | C | G | G | G | C | C | | | | | | | | | | | | | 1047 |
| P1B 01-H1 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1049 |
| P1B 01-E4 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1049 |
| P2B 08-H3 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1049 |
| P3B 13-D10 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1049 |
| P4B 04-H3 | C | C | C | G | G | G | C | C | C | G | | | | | | | | | | | 1050 |
| P5B 14-C11 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1048 |
| P4BSF 03-G10 | C | C | C | G | G | G | C | C | C | | | | | | | | | | | | 1049 |