



(86) Date de dépôt PCT/PCT Filing Date: 2002/01/22

(87) Date publication PCT/PCT Publication Date: 2002/07/25

(45) Date de délivrance/Issue Date: 2008/09/02

(85) Entrée phase nationale/National Entry: 2003/07/18

(86) N° demande PCT/PCT Application No.: EP 2002/000578

(87) N° publication PCT/PCT Publication No.: 2002/057459

(30) Priorité/Priority: 2001/01/22 (DE101 05 152.2)

(51) Cl.Int./Int.Cl. *C12N 15/55* (2006.01),
C12N 15/63 (2006.01), *C12N 15/79* (2006.01),
C12N 9/20 (2006.01)

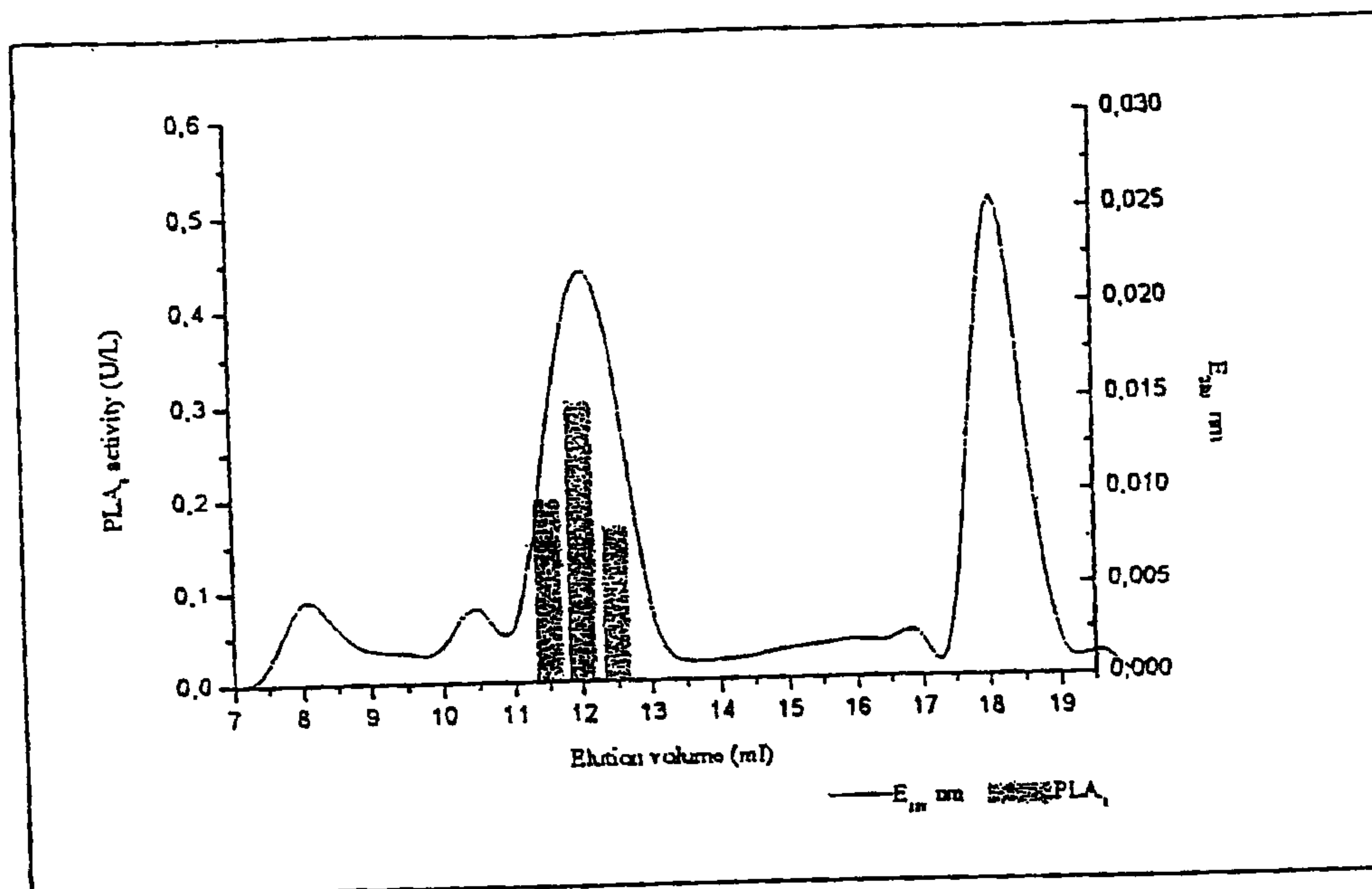
(72) Inventeurs/Inventors:
HARTMANN, MARCUS, DE;
GRENNINGLOH, MARCO, DE;
TIEDTKE, ARNO, DE

(73) Propriétaire/Owner:
CILIAN AG, DE

(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

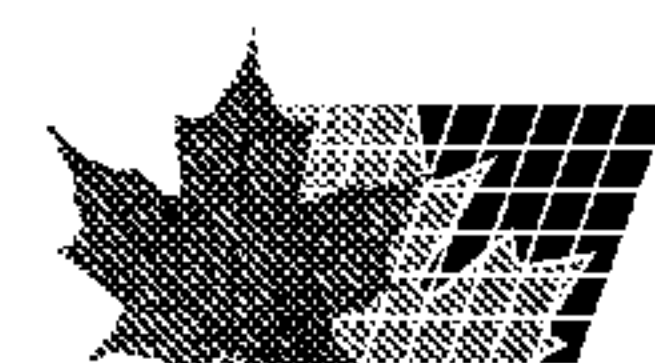
(54) Titre : SEQUENCE D'ADN DE L'ENZYME PHOSPHOLIPASE A1 DU CILIE TETRAHYMENA ET SON UTILISATION
(54) Title: DNA SEQUENCE OF THE ENZYME PHOSPHOLIPASE A1 OF CILIATE TETRAHYMENA, AND THE USE OF
THE SAME

SEC (size exclusion chromatography), elution
profile and enzyme activity of PLA₁



(57) Abrégé/Abstract:

A nucleic acid coding for the phospholipase A₁ from ciliates. In particular, the phospholipase A₁ has the amino acid sequence SEQ ID No. 7.



Abstract

A nucleic acid coding for the phospholipase A₁ from ciliates. In particular, the phospholipase A₁ has the amino acid sequence SEQ ID No. 7.

**DNA SEQUENCE OF THE ENZYME PHOPHOLIPASE A1 OF CILIATE
TETRAHYMENA, AND THE USE OF THE SAME**

The invention relates to a nucleic acid coding for the phospholipase A1 and the use thereof according to the preamble of claims 1 to 5.

5 Yeasts, bacteria and mammal cells are of great importance to the biotechnological preparation and production of recombinant active substances by the heterologous expression of foreign proteins. Bacterial expression systems based on *E. coli* or *B. subtilis* are used for the production of recombinant peptides or proteins, such as insulin, interleukin-2, tissue
10 plasminogen activator, proteases and lipases. In Gram-negative bacteria, the expression systems are mostly based on the use of genetic elements, such as the lac operon or the tryptophan operon. The proteins foreign to the host are produced either into "inclusion bodies" within the cell, or when expression systems based on β -lactamase genes are used, into the peri-
15 plasmic space. The production of recombinant proteins into the surrounding fermentation medium has not been established. In Gram-Positive bacteria, to date, almost exclusively cell-inherent proteins are introduced in expression systems and expressed.

Yeasts, such as *S. cerevisiae*, *Hansenula polymorpha*, *Kluyveromyces lactis*
20 or *Pichia pastoris*, are also employed for the heterologous expression of recombinant proteins, such as human factor XIIIa, bovine pro-chymosin, phytase or surface antigens. Here, the expression systems are based on shuttle vectors (vectors having both yeast and bacterial portions) which are based (depending on the yeast species) on the genetic elements of galactokinase-epimerase, methanol oxidase, acid phosphatase or alcohol-
25 dehydrogenase. As a rule, the recombinant protein is produced into the cytoplasm of the cell. When yeast-inherent signal sequences, such as the alpha factor, are used, the expressed proteins may also be secreted into the fermentation medium. The glycosylation of secreted proteins is effected
30 according to the "high mannose" type, and frequently there are hyperglyco-

- 2 -

sylations on the protein which may result in the formation of antibodies in the patient.

Mammal cells, such as various cell types from rodents (CHO cells, C127 cells) or simians (vero, CV-1 or COS cells) are also employed for the
5 heterologous expression of recombinant proteins. Here, the expression systems are based on recombinant viruses (BPV vector) or on shuttle vectors. To regulate the expression, viral SV40 enhancer/promoter systems or cellular enhancer elements are employed. The recombinant proteins, such as erythropoietin, are secreted into the fermentation medium because the
10 foreign genes usually bring their own signal sequences, which are understood by the expression system and used for targeting.

Further, for the biotechnological production of glycosylated extracellular enzymes, protozoans of the genus *Tetrahymena* are employed. *Tetrahymena* will grow on inexpensive fermentation media using standard fermentation
15 methods. For the transformation of such *Tetrahymena* cells, vectors are available which are based on the rDNA elements of *Tetrahymena*. For the heterologous expression of bacterial proteins in *Tetrahymena*, DNA constructs made from genes from *Tetrahymena* are employed. When suitable genetic elements for the regulation of the transcription, targeting
20 and glycosylation of foreign proteins are available, *Tetrahymena* is an ideal expression system for the inexpensive production of therapeutic recombinant proteins.

The Gram-negative bacterial expression systems used to date usually lead to the formation of "inclusion bodies" in the cell, accompanied by a denatur-
25 ing of the proteins. To recover the recombinant protein, the cells must be lysed, and the denatured inactive protein must be folded back to function. This causes additional cost-intensive process steps and reduces the yield of the desired protein. Glycosylation, which is important to eukaryotic proteins, is completely omitted. When Gram-positive bacterial expression systems are

- 3 -

used, degradation of the target protein due to high proteolytic activities in the fermentation broth is an additional problem.

When yeasts are used for heterologous expression, the desired target protein is also produced only into the cell, from where it must be removed
5 by cell lysis. As in bacterial expression systems, this causes additional time- and cost-intensive process steps. When yeast-inherent signal peptides are used, the foreign proteins are not correctly spliced and glycosylated for secretion.

In contrast, when mammal cell systems are employed for the production of
10 recombinant proteins, the desired proteins are found in the fermentation medium in an extracellular state, correctly spliced and glycosylated. However, what is disadvantageous here is, on the one hand, the low expression rate due to the defective processing and inefficient translation of genes which have been introduced into the genome of the production cell
15 line via viral vectors. On the other hand, the serum-containing fermentation media for mammal cells are extremely cost-intensive. In addition, the fermentation technology for the shear-sensitive cell lines is complicated and similarly expensive due to constructions for bubble-free aeration. Further problems arise from the high infection risk for the cell lines from mycoplas-
20 mas and viruses. All in all, the use of mammal cells for the biotechnological preparation of recombinant proteins results in very high costs, safety demands and low yields.

To the use of ciliates, such as *Tetrahymena*, the above mentioned draw-
backs in the production of recombinant proteins do not apply. Thus, for
25 example, some acid hydrolases which are involved in the digestion of food particles are exported from the cell in high quantities and with complex glycosylation.

In J. Euk. Microbiol. 43 (4), 1996, pages 295 to 303, Alam et al. describe the cloning of a gene which codes for the acid α -glucosidase of *Tetrahymena*

- 4 -

pyriformis. However, only a small portion of the protein is exported from the cell. Further, the International Patent Application WO 00/52176 describes the gene of a β -hexosaminidase from *Tetrahymena thermophila* which is known, however, to be exported from the cell to only about 80%.

5 However, to date, it has not been possible to cause glycosylated eukaryotic proteins to be expressed in *Tetrahymena* and also be exclusively secreted into the fermentation medium. This is because the DNA sequences of extracellular proteins inherent to *Tetrahymena* which are necessary for the construction of expression vectors and which exclusively export the foreign
10 protein into the surrounding fermentation medium have as yet been unknown. The DNA sequences of a protein which codes for the β -hexosaminidase of *Tetrahymena thermophila* are known. Such a sequence has been filed for a patent application under the International Patent application WO 00/52176. However, there is
15 a disadvantage of these sequences in that the pre/pro-peptides containing them will target a protein foreign to the host into the surrounding fermentation medium to only about 80%. This is due to the fact that the enzyme β -hexosaminidase is present to about 20% within the membrane under natural conditions, and only about 80% of the naturally produced enzyme is
20 exported from the cell. For this reason, pre/pro peptides of β -hexosaminidase, when positioned in front of a protein foreign to the host by genetic engineering methods, will target only about 20% of this protein foreign to the host into the cytoplasmic membrane on the surface of *Tetrahymena thermophila*. This is associated with a considerable process-
25 technological disadvantage for the production of recombinant active substances. On the one hand, the yield is decreased because part of the expressed protein remains in the cells bound to the membrane, and thus it is not possible to purify the entire expressed protein from the fermenter broth. On the other hand, the protein foreign to the host in the cell mem-
30 brane can exert toxic effects on the host cells and thus slow down the cell growth.

- 5 -

Further, no constitutive promoters of *Tetrahymena* which cause a consistent or continuous transcription of heterologous proteins have been known to date. To date, only promoters of histone and tubulin genes have been known (Bannon et al., 1984, Gaertig et al., 1993). However, a critical
5 disadvantage of these promoters is that their activation is dependent on the cell cycle. Genes of heterologous proteins which are linked to such cell-cycle-dependent promoters are caused to be expressed only in growing or dividing cells. This has considerable process-technological disadvantages since the desired protein is thus produced only in the logarithmic growth
10 phase. In the stationary growth phase in which the highest cell density and thus the highest performance of the expression organism (*Tetrahymena*) is reached in the production process, there is hardly any cell growth left and thus only a low expression of the heterologous protein takes place.

It is an object of the invention to provide a system which enables the
15 production of heterologous proteins in an expression system, after transformation into *Tetrahymena*, from the cells into the fermentation medium.

This object is achieved by a system in which a nucleic acid having the sequence SEQ ID No. 1 coding for a phospholipase A1 (SEQ ID No. 7) is employed. Advantageously, the expression product of this DNA is exported
20 from the cell in large amounts under culturing conditions. The expressed protein is exported into the surrounding culture medium to a high extent and is not contained in the membrane. The nucleic acid sequence according to the invention contains a promoter which causes a constitutive, i.e., cell-cycle-independent, transcription of the downstream genes of heterologous
25 proteins. Such constitutive transcription has the advantage that the proteins are continuously expressed by heterologous expression in the host organism without being affected by the cell cycle. Thus, the transcription of the foreign gene can be effected and the heterologous protein expressed also during the stationary growth phase with a low cell growth.

- 6 -

The DNA sequence of phospholipase A₁ according to the invention preferably includes an upstream region of PLA₁ (SEQ ID No. 2) which bears the promoter elements for the initiation of transcription, a signal peptide and a pro-peptide, further genetic elements for the targeting of proteins and, in particular, a downstream region of PLA₁ (SEQ ID No. 3) which contains genetic elements for the termination of transcription. The use of these nucleic acids in a vector enables the expression of heterologously expressed proteins independently of the cell cycle and to transport them selectively out of the cell and into the surrounding culture medium without expressed proteins becoming incorporated in the cytoplasmic membrane, whereby such proteins can be isolated from the fermentation broth without cell lysis.

Figure 1 shows a nucleic acid coding for the upstream region (SEQ ID No. 2), the coding region (SEQ ID No. 1) and the downstream region (SEQ ID No. 3) of phospholipase A₁ from ciliates.

Figure 2 shows a corresponding expression product of the nucleic acid according to SEQ ID No. 1. The invention also relates to the protein according to SEQ ID No. 7.

In particular, the invention also relates to the signal sequence (SEQ ID No. 6) of the protein according to the invention. Preferably, these are the amino acids 1 to 110 of the protein according to the invention (SEQ ID No. 5). The invention also relates to a nucleic acid coding for the N-terminal fragment (SEQ ID No. 3). This is preferably a fragment of the nucleic acids according to the invention (SEQ ID No. 4), especially having the nucleic acid sequence 1 to 155 according to Figure 1.

The nucleic acid sequence of the non-translated region (upstream region) (SEQ ID No. 2) upstream from the coding sequence region of the PLA₁ from *Tetrahymena* is positioned between position -275 and position -1 (represented in lowercase letters). The established non-translated region comprises 275 bases. As elements of a promoter, a TATA box is found on

- 7 -

positions -49 to -55 (printed in boldface), and a putative CAAT box is found between base -133 and base -136 (printed in boldface). The coding sequence range of the cDNA is represented in capital letters. The numbering of the sequence begins with the start codon ATG. Regions known from protein sequencing are boxed, and the stop codon is underlined. The mature protein is coded from base 331. The sequence listing from base 1 to base 330 represents the pre/pro sequence (SEQ ID No. 8) of PLA₁. The sequence listing from base 331 to base 963 is the sequence of the mature PLA₁ (SEQ ID No. 9). In position 961, there is the translation stop TGA, and in position 1039, there is the polyadenylation signal AAT AAA. The nucleic acid sequence from position 964 to position 1134, which is below the coding sequence of the PLA₁ of *Tetrahymena*, represents the downstream region of PLA₁ (SEQ ID No. 3) which is not translated (also represented in lowercase letters). In position 964 to position 1101, there is the region known from the sequencing of the cDNA, which was also confirmed by inverse PCR. After transcription, the poly-A tail is attached to the last codon of the mRNA (ttt, positions 1098-1101).

A further aspect of the invention is the use of a nucleic acid sequence of acid hydrolases according to the invention or parts thereof for the homologous or heterologous expression of recombinant proteins and peptides, and for homologous or heterologous recombination ("knock-out, "gene replacement").

The invention also relates to a method for the homologous or heterologous expression of proteins and peptides and for the homologous or heterologous recombination ("knock-out, "gene replacement") in which ciliates are transfected with a nucleic acid according to the invention.

The nucleic acids or parts thereof may be combined, in particular, with the enhancers, promoters, operators, origins, terminators, antibiotic resistances usual for the homologous or heterologous expression of proteins, or with

- 8 -

other nucleic acids or DNA fragments or all kinds of sequences from viroids, viruses, bacteria, archezoans, protozoans, fungi, plants, animals or humans.

In particular, the nucleic acid according to the invention is contained in a vector, a plasmid, a cosmid, a chromosome or minichromosome, a transposon, an IS element, an rDNA, or all kinds of circular or linear DNA or RNA.

The invention also relates to a method in which the nucleic acid or parts thereof according to the invention which code for phospholipase A₁ are combined with the usual, in homologous or heterologous expression, enhancers, such as the NF-1 region (a cytomegalovirus enhancer), promoters, such as the lac, trc, tic or tac promoters, the promoters of classes II and III of the T7 RNAP system, bacteriophage T7 and SP6 promoters, aprE, amylase or spac promoters for *Bacillus* expression systems, AOX1, AUG1 and 2 or GAPp promoters (*Pichia*) for yeast expression systems, RSV promoter (SV40 virus), CMV promoter (Cytomegalovirus), AFP promoter (adenoviruses) or metallothionine promoters for mammal expression systems, Sindbis virus promoters or Semlike forest virus promoters for insect cells, promoters for insect cell expression systems, such as hsp70, DS47, actin 5C or copia, plant-specific promoters, such as 35S promoter (cauliflower mosaic virus), amylase promoter or class I patatin promoter, operators, such as the tet operator, signal peptides, such as a-MF prepro signal sequences (*Saccharomyces*), origins, terminators, antibiotic and drug resistances, such as ampicillin, kanamycin, streptomycin, chloramphenicol, penicillin, amphotericin, cycloheximide, 6-methylpurine, paromomycin, hygromycin, α -amanatin, auxotrophy markers, such as the gene of dihydro-folate reductase, or other nucleic acids or DNA fragments, or all kinds of sequences from viroids, viruses, bacteria, archezoans, protozoans, fungi, plants, animals or humans.

In particular, the nucleic acid or parts thereof according to the invention are inserted into a vector, a plasmid, a cosmid, a chromosome or minichromo-

- 9 -

some, a transposon, an IS element, an rDNA, or all kinds of circular or linear DNA or RNA.

The skilled person will understand that nucleic acids having at least 40% homology with the nucleic acid according to SEQ ID No. 1 can also be employed according to the invention. The protein according to SEQ ID No. 2 can also be modified without losing its function. Thus, for example, so-called conservative exchanges of amino acids may be performed. Thus, for example, hydrophobic amino acids can be interchanged.

For the purification and isolation of phospholipase A₁ from *Tetrahymena* and for determining its sequence, the following methods can be used.

Recovery of PLA₁

PLA₁ was obtained from cell-free culture supernatants of *Tetrahymena thermophila*. Thus, the cells were fermented in a 2 l fermenter (Biostat MD, Braun Diessel Biotech, Melsungen, Germany) which was controlled over a digital controlling unit (DCU). The fermenter was first operated for 24 hours in a batch operation and then continuously. Harvesting of the cell-free culture supernatant was ensured through a perfusion module having a pore size of about 0.3 µm (S6/2, Enka, Wuppertal).

The fermentation was performed under the following parameters:

- the working volume was 2 liters;
- the perfusion rate was 2 liters/day;
- the revolutions per minute of the stirrer was limited to 800 rpm;
- the temperature was constantly at 30 °C;
- the pH value was kept constant at pH 7;
- the inoculation titer was at 50,000 cells/ml.

For the fermentation, the strain SB 1868 VII was used. This is a wild type strain of *Tetrahymena thermophila*.

- 10 -

The fermentation was performed over a period of 264 hours, and the harvests were tested for PLA₁ activity.

Purification of PLA₁

For the purification of PLA₁, 1 liter of cell-free culture supernatant from the fermentation was used. It was admixed with 140 g of ammonium sulfate and concentrated through an ultrafiltration unit (Pellikon XL, exclusion size 3 kDa, Millipore) to a volume of 50 ml. Subsequently, the sample was purified by hydrophobic interaction chromatography (20 x 1,6 Fractogel EMD Phenyl I 650, Merck, Darmstadt). The flow rate was 5 ml/min, and the eluate was collected in 5 ml fractions. The enzyme activity was measured by the deacylation of a radioactively labeled phospholipid (L-3-phosphatidylcholine, 1-palmitoyl-2-[1-¹⁴C]linoleoyl). Figure 3 shows the elution profile obtained, the sodium acetate gradient and the enzyme activities in the individual fractions.

The three fractions having the highest enzyme activities were combined and rebuffered into the starting buffer (Bis-Tris 20 mM, pH 6.5) for anion-exchange chromatography (AEC) by means of an ultrafiltration unit. Subsequently, the sample was charged onto the column (Q-Sepharose-Hiload-16/10, Pharmacia, Sweden), and the PLA₁ was eluted with a linear NaCl gradient (flow rate = 3 ml/min) from the column and collected in 5 ml fractions. Figure 4 shows the elution profile obtained, the NaCl gradient and the enzyme activities of the individual fractions.

From the fraction having the highest PLA₁ activity, 200 µl was removed and separated by size exclusion chromatography (SEC). For this purpose, a Superdex^{*} HR 75 30/10 column (Pharmacia, Sweden) was used. The flow rate in this chromatography was 0.6 ml/min, the eluate was collected in 200 µl fractions. Figure 5 shows the elution profile obtained and the enzyme activities of the individual fractions.

***Trade-mark**

- 11 -

The fractions obtained were examined for their purity using one-dimensional gel electrophoresis. Thus, two distinct bands were established at ~26 and ~28 kDa. Separation of these two bands by a two-dimensional gel electrophoresis resulted in a separation of the two bands into 2 and 3 spots, respectively, having different isoelectric points.

For the 26 kDa proteins, these were at pH 6.3 and 5.7, and for the 28 kDa proteins, they were at pH 6.3, 5.7 and 5.3. A final examination of these spots by mass fingerprint analysis showed, that these spots were isoforms of the same protein.

10 **Molecular-biological examination of PLA₁**

After the purity of the protein had been demonstrated, samples of the protein were blotted onto a PVDF membrane and subjected to initial sequencing from the N terminus. In addition, a further sample was tryptically digested and also subjected to initial sequencing. Using the protein sequences obtained thereby, oligonucleotide primers were prepared, which were then employed in reverse transcriptase PCR (3' RACE, rapid amplification of cDNA ends). Using this PCR technique, cDNA of phospholipase A₁ was successfully amplified and subsequently sequenced. The sequence obtained had a length of 633 bases and 729 bases, respectively, and the molecular weight of the mature protein derived therefrom is about 22.4 kDa. In the sequence derived, the oligopeptides of 22 amino acids (N-terminal) and 18 amino acids (within the protein) established from protein sequencing were found again to 100%. In addition to the sequence of the mature protein, the sequence of the pre/pro peptide could also be established by means of 5' RACE (rapid amplification of cDNA ends) (Figure 2). This is a peptide having a length of 110 amino acids which bears both the signal sequence and the pro peptide which inactivates the enzyme and is cleaved off only at the final place of activity of the enzyme. Sequence comparisons yielded no homologies with previously known phospholipases A₁, except for a consensus sequence of 5 amino acids (GxSxG), which is found in all lipases and

- 12 -

phospholipases and is discussed as a binding site for lipids or phospholipids. Further, the upstream and downstream sequences of PLA₁ were established by inverse PCR (Figure 1). Thus, genomic DNA was cut with restriction endonucleases, ligated with T4 ligase and finally amplified with inverse
5 primers. For the amplification of the upstream region of PLA₁ by inverse PCR, genomic DNA cut with the restriction endonuclease SspI was used. Thus, an upstream region of 275 bases could be established, and promoter elements identified. In position -136, there is a CAAT box which has a similar distance from the translation start as the CCAAT boxes of the histone
10 genes (-141 and -151) of *Tetrahymena* as found by Brunk and Sadler (1990). A TATA box, which fixes the exact starting point of transcription in eukaryotic genes, was identified on position -55. Its sequence corresponds to the consensus sequences found in eukaryotes. For the amplification of the downstream region, which contains the terminator for the transcription
15 of the PLA₁ from *Tetrahymena*, by inverse PCR, genomic DNA cut with BamHI was used. Thus, in addition to the downstream region known from 3'-RACE, another 222 bases could be established (Figure 3).

- 13 -

SEQUENCE LISTING

<110> Cilian AG

<120> A DNA sequence of the enzyme Phospholipase A1 from the ciliate Tetrahymena and use thereof

<130> 6603-230CA

<150> PCT/EP02/00578

<151> 2002-01-22

<150> 101 05 152.2

<151> 2001-01-22

<160> 9

<170> PatentIn Ver. 2.1

<210> 1

<211> 963

<212> DNA

<213> Tetrahymena thermophila

<400> 1

atgaacaaga ctctcatctt agctttagtt gttgttttgg ctttaactgc caccaccttg 60
gttgctttcc acaaccactc tcacaacatc agagttgact aagaccccg cactctcttc 120
aagcaattca agcaaactta caataagaag tatgctgatg ctactttcga aacctacaga 180
ttcgggtgtct tcacccaaaa cttagaaatc gtcaagactg actctacttt cgggtgtcacc 240
taattcatgg acttaactcc tgctgaattc gctcaacaat tcctcacttt acacgaaaag 300

- 14 -

gttaacagca ccgaagttta cagagcttaa ggtgaagcta ccgaagttga ctggactgcc 360
 aagggttaagg tcacccctgt taagaactaa gggtcttggtg gttcctgctg ggctttctcc 420
 accattgggtg ccgttgaatc tgctctttgg attgctgggc aagggtgaata aaacactctt 480
 aaccttgctg aataagaata agttgactgt gctaagtcct ccaagtacga ctctgaaggt 540
 tgcaacgggtg gttggatggt tgaagggttc aagtacatca tcgacaacaa gatctcctaa 600
 actgctaact atccctacac tgctaaggat ggtaagtgc aggacacctc ttccttcaag 660
 aagttctcta tttctaagta cgctgaaatc ccctaagggtg actgcaactc cctcaactct 720
 gccttagaac aaggtcctat ctccgttgct gttgatgcc ccaacttcta attctacact 780
 tctgggtgtct ttaaaaactg caaggccaac ctcaaccacg gtgtcctctt agttgccaac 840
 gttgactctt ctctcaagat caagaactcc tggggtcctt cttgggggtga aaagggtttc 900
 atcagattag ctgccggtta cacttgcggt gtctgcaatg ctgcctctta ccctattggt 960
 tga 963

<210> 2

<211> 275

<212> DNA

<213> Tetrahymena thermophila

<400> 2

aatatttatc aatgctactt ataattcttt tagtatgaga tatgatatgc tctttctctg 60
 ctagacttaa cttatgacat ttgaactttt aataaaagaa ttttttttat taaaaagcag 120
 agatttttaa tagaagaatc aatgactcat gaatttaata aagattttca agtgttttct 180
 aataccgact agctttataa attcacttat taatcaacga tataaaaatt atattaacaa 240
 atcaataaat aaaaaataa ataaaaacaa aacaa 275

<210> 3

<211> 171

- 15 -

<212> DNA

<213> *Tetrahymena thermophila*

<400> 3

aaaaacataa tccaaattaa aaaaaattac tcaaaactga taatataaaa aattaatttt 60
cataatttta atgtaaataa atacctttat atttgacgtt ttgtactcaa aataaattaa 120
agttaacaaa ccatatttat ttaattctac ttttcaattt ttaaaaatat a 171

<210> 4

<211> 1408

<212> DNA

<213> *Tetrahymena thermophila*

<400> 4

aatatttatc aatgctactt ataattcttt tagtatgaga tatgatatgc tctttctctg 60
ctagacttaa cttatgacat ttgaactttt aataaaagaa ttttttttat taaaaagcag 120
agatttttaa tagaagaatc aatgactcat gaatttaata aagattttca agtgttttct 180
aataccgact agctttataa attcacttat taatcaacga tataaaaatt atattaacaa 240
atcaataaat aaaaaaataa ataaaaacaa aacaaatgaa caagactctc atcttagctt 300
tagttgttgt tttggcttta actgccacca ccttggttgc tttccacaac cactctcaca 360
acatcagagt tgactaagac cccgccactc tcttcaagca attcaagcaa acttacaata 420
agaagtatgc tgatgctact ttcgaaacct acagattcgg tgtcttcacc caaaacttag 480
aaatcgtcaa gactgactct actttcgggtg tcaccttaatt catggactta actcctgctg 540
aattcgtca acaattcctc actttcacga aaaggttaac agcaccgaag tttacagagc 600
ttaagggtgaa gctaccgaag ttgactggac tgccaagggt aaggtcaccc ctgttaagaa 660
ctaaggttct tgtggttcct gctgggcttt ctccaccatt ggtgccgttg aatctgctct 720
ttggattgct ggtcaagggtg aataaaacac tcttaacctt gctgaataag aataagttga 780
ctgtgctaag tcccccaagt acgactctga aggttgcaac ggtgggttga tggttgaagg 840

- 16 -

tttcaagtac atcatcgaca acaagatctc ctaaactgct aactatccct acactgctaa 900
 ggatggtaag tgcaaggaca cctcttcctt caagaagttc tctatttcta agtacgctga 960
 aatcccctaa ggtgactgca actccctcaa ctctgcctta gaacaaggtc ctatctccgt 1020
 tgctgttgat gccaccaact tctaattcta cacttctggt gtcttttaaaa actgcaaggc 1080
 caacctcaac cacggtgtcc tcttagttgc caacgttgac tcttctctca agatcaagaa 1140
 ctcttgggggt ccttcttggg gtgaaaaggg tttcatcaga ttagctgccg gtaacacttg 1200
 cggtgtctgc aatgctgcct cttaccctat tgtttgaaaa aacataatcc aaattaaaaa 1260
 aaattactca aaactgataa tataaaaaat taattttcat aattttaatg taaataaata 1320
 cctttatatt tgacgttttg tactcaaaat aaattaaagt taacaaacca tatttattta 1380
 attctacttt tcaattttta aaaatata 1408

<210> 5

<211> 320

<212> PRT

<213> Tetrahymena thermophila

<400> 5

Met Asn Lys Thr Leu Ile Leu Ala Leu Val Gly Val Leu Ala Leu Thr

1

5

10

15

Ala Thr Thr Leu Val Ala Phe His Asn His Ser His Asn Ile Arg Val

20

25

30

Asp Gln Asp Pro Ala Thr Leu Phe Lys Gln Phe Lys Gln Thr Tyr Asn

35

40

45

Lys Lys Tyr Ala Asp Pro Thr Phe Glu Thr Tyr Arg Phe Gly Val Phe

50

55

60

- 17 -

Thr Gln Asn Leu Glu Ile Val Lys Thr Asp Ser Thr Phe Gly Val Thr
65 70 75 80

Gln Phe Met Asp Leu Thr Pro Ala Glu Phe Ala Gln Gln Phe Leu Thr
85 90 95

Leu His Glu Lys Val Asn Ser Thr Glu Val Tyr Arg Ala Gln Gly Glu
100 105 110

Ala Thr Glu Val Asp Trp Thr Ala Lys Gly Lys Val Thr Pro Val Lys
115 120 125

Asn Gln Gly Ser Cys Gly Ser Cys Trp Ala Phe Ser Thr Ile Gly Ala
130 135 140

Val Glu Ser Ala Leu Leu Ile Ala Gly Gln Gly Glu Gln Asn Thr Leu
145 150 155 160

Asn Leu Ala Glu Gln Glu Leu Val Asp Cys Ala Lys Ser Pro Lys Tyr
165 170 175

Asp Ser Glu Gly Cys Asn Gly Gly Trp Met Val Glu Gly Phe Lys Tyr
180 185 190

Ile Ile Asp Asn Lys Ile Ser Gln Thr Ala Asn Tyr Pro Tyr Thr Ala
195 200 205

Lys Asp Gly Lys Cys Lys Asp Thr Ser Ser Phe Lys Lys Phe Ser Ile

- 18 -

210	215	220
Ser Lys Tyr Ala Glu Ile Pro Gln Gly Asp Cys Asn Ser Leu Asn Ser		
225	230	235
Ala Leu Glu Gln Gly Pro Ile Ser Val Ala Val Asp Ala Thr Asn Phe		
245	250	255
Gln Phe Tyr Thr Ser Gly Val Phe Lys Asn Cys Lys Ala Asn Leu Asn		
260	265	270
His Gly Val Leu Leu Val Ala Asn Val Asp Ser Ser Leu Lys Ile Lys		
275	280	285
Asn Ser Trp Gly Pro Ser Trp Gly Glu Lys Gly Phe Ile Arg Leu Ala		
290	295	300
Ala Gly Asn Thr Cys Gly Val Cys Asn Ala Ala Ser Tyr Pro Ile Val		
305	310	315
320		

<210> 6

<211> 110

<212> PRT

<213> Tetrahymena thermophila

- 19 -

<400> 6

Met Asn Lys Thr Leu Ile Leu Ala Leu Val Gly Val Leu Ala Leu Thr

1 5 10 15

Ala Thr Thr Leu Val Ala Phe His Asn His Ser His Asn Ile Arg Val

20 25 30

Asp Gln Asp Pro Ala Thr Leu Phe Lys Gln Phe Lys Gln Thr Tyr Asn

35 40 45

Lys Lys Tyr Ala Asp Pro Thr Phe Glu Thr Tyr Arg Phe Gly Val Phe

50 55 60

Thr Gln Asn Leu Glu Ile Val Lys Thr Asp Ser Thr Phe Gly Val Thr

65 70 75 80

Gln Phe Met Asp Leu Thr Pro Ala Glu Phe Ala Gln Gln Phe Leu Thr

85 90 95

Leu His Glu Lys Val Asn Ser Thr Glu Val Tyr Arg Ala Gln

100 105 110

<210> 7

<211> 210

<212> PRT

<213> Tetrahymena thermophila

<400> 7

- 20 -

Gly Glu Ala Thr Glu Val Asp Trp Thr Ala Lys Gly Lys Val Thr Pro

1 5 10 15

Val Lys Asn Gln Gly Ser Cys Gly Ser Cys Trp Ala Phe Ser Thr Ile

20 25 30

Gly Ala Val Glu Ser Ala Leu Leu Ile Ala Gly Gln Gly Glu Gln Asn

35 40 45

Thr Leu Asn Leu Ala Glu Gln Glu Leu Val Asp Cys Ala Lys Ser Pro

50 55 60

Lys Tyr Asp Ser Glu Gly Cys Asn Gly Gly Trp Met Val Glu Gly Phe

65 70 75 80

Lys Tyr Ile Ile Asp Asn Lys Ile Ser Gln Thr Ala Asn Tyr Pro Tyr

85 90 95

Thr Ala Lys Asp Gly Lys Cys Lys Asp Thr Ser Ser Phe Lys Lys Phe

100 105 110

Ser Ile Ser Lys Tyr Ala Glu Ile Pro Gln Gly Asp Cys Asn Ser Leu

115 120 125

Asn Ser Ala Leu Glu Gln Gly Pro Ile Ser Val Ala Val Asp Ala Thr

130 135 140

Asn Phe Gln Phe Tyr Thr Ser Gly Val Phe Lys Asn Cys Lys Ala Asn

145 150 155 160

- 21 -

Leu Asn His Gly Val Leu Leu Val Ala Asn Val Asp Ser Ser Leu Lys

165

170

175

Ile Lys Asn Ser Trp Gly Pro Ser Trp Gly Glu Lys Gly Phe Ile Arg

180

185

190

Leu Ala Ala Gly Asn Thr Cys Gly Val Cys Asn Ala Ala Ser Tyr Pro

195

200

205

Ile Val

210

<210> 8

<211> 330

<212> DNA

<213> Tetrahymena thermophila

<400> 8

atgaacaaga ctctcatctt agctttagtt gttgttttgg ctttaactgc caccaccttg 60
gttgctttcc acaaccactc tcacaacatc agagttgact aagaccccg cactctcttc 120
aagcaattca agcaaactta caataagaag tatgctgatg ctactttcga aacctacaga 180
ttcgggtgtct tcacccaaaa cttagaaatc gtcaagactg actctacttt cgggtgtcacc 240
taattcatgg acttaactcc tgctgaattc gctcaacaat tcctcacttt acacgaaaag 300
gttaacagca ccgaagttaa cagagcttaa 330

<210> 9

- 22 -

<211> 633

<212> DNA

<213> *Tetrahymena thermophila*

<400> 9

ggtgaagcta ccgaagttga ctggactgcc aagggtgaagg tcaccctgt taagaactaa 60
ggttcttggtg gttcctgctg ggctttctcc accattggtg ccgttgaatc tgctctttgg 120
attgctggtc aagggtgaata aaacactctt aaccttgctg aataagaata agttgactgt 180
gctaagtccc ccaagtacga ctctgaagggt tgcaacgggtg gttggatggt tgaaggtttc 240
aagtacatca tcgacaacaa gatctcctaa actgctaact atccctacac tgctaaggat 300
ggtaagtgca aggacacctc ttccttcaag aagttctcta tttctaagta cgctgaaatc 360
ccctaagggtg actgcaactc cctcaactct gccttagaac aaggtcctat ctccgttgct 420
gttgatgcca ccaacttcta attctacact tctgggtgtct ttaaaaactg caaggccaac 480
ctcaaccacg gtgtcctctt agttgccaac gttgactctt ctctcaagat caagaactcc 540
tggggtcctt cttgggggtga aaagggtttc atcagattag ctgccggtaa cacttgcggt 600
gtctgcaatg ctgcctctta ccctattggt tga 633

1. A nucleic acid comprising coding and non-coding regions having the sequence of SEQ ID NO: 4, wherein said coding regions code for a phospholipase A₁ from *Tetrahymena thermophila*, and said non-coding regions represent the upstream and downstream regions of this gene.
2. The nucleic acid according to claim 1, wherein said non-coding upstream and downstream regions have the sequence of SEQ ID NO: 2 and SEQ ID NO: 3 respectively, and the coding regions have the sequence of SEQ ID NO: 1, SEQ ID NO: 8 or SEQ ID NO: 9.
3. An N-terminal signal sequence of the phospholipase A₁ according to claim 1, having the sequence of SEQ ID NO: 6.
4. A phospholipase A₁ protein having the amino acid sequence according to SEQ ID NO: 7.
5. A method for homologous or heterologous expression of a protein comprising transfecting a ciliate host cell with the nucleic acid as defined in claim 1 and culturing the host cell under conditions allowing expression of phospholipase A₁.
6. The method according to claim 5, wherein the transfected nucleic acid is contained in a vector, a plasmid or a cosmid.
7. A method for homologous or heterologous expression of a protein comprising transfecting a ciliate host cell with the nucleic acid as defined in claim 2 and culturing the host cell under conditions allowing expression of phospholipase A₁.
8. Vectors, plasmids, cosmids, chromosomes or minichromosomes, transposons, insertion sequence elements, rDNA or other kinds of circular or linear DNA or RNA containing the nucleic acid according to any one of claims 1 or 2.

Fig. 1

								aa	tat	tta	tca	atg	cta	ctt	ata	att	ctt	tta	gta	-241
tga	gat	atg	ata	igc	tct	ttc	tct	gct	aga	ctt	aac	tta	tga	cat	ttg	aac	ttt	taa	taa	-181
aag	aat	ttt	ttt	tat	taa	aaa	gca	gag	att	ttt	aat	aga	aga	atc	aat	gac	tca	tga	att	-121
taa	taa	aga	ttt	tca	agt	gtt	ttc	taa	tac	cga	cta	gct	tta	taa	att	cac	tta	tta	atc	-61
aac	gat	ata	aaa	att	ata	tta	aca	aat	caa	taa	ata	aaa	aaa	taa	ata	aaa	aca	aaa	caa	-1
ATG	AAC	AAG	ACT	CTC	ATC	TTA	GCT	TTA	GTT	GTT	GTT	TTG	GCT	TTA	ACT	GCC	ACC	ACC	TTG	60
GTT	GCT	TTC	CAC	AAC	CAC	TCT	CAC	AAC	ATC	AGA	GTT	GAC	TAA	GAC	CCC	GCC	ACT	CTC	TTC	120
AAG	CAA	TTC	AAG	CAA	ACT	TAC	AAT	AAG	AAG	TAT	GCT	GAT	GCT	ACT	TTC	GAA	ACC	TAC	AGA	180
TTC	GGT	GTC	TTC	ACC	CAA	AAC	TTA	GAA	ATC	GTC	AAG	ACT	GAC	TCT	ACT	TTC	GGT	GTC	ACC	240
TAA	TTC	ATG	GAC	TTA	ACT	CCT	GCT	GAA	TTG	GCT	CAA	CAA	TTC	CTC	ACT	TTA	CAC	GAA	AAG	300
GTT	AAC	AGC	ACC	GAA	GTT	TAC	AGA	GCT	TAA	GGT	GAA	GCT	ACC	GAA	GTT	GAC	TGG	ACT	GCC	360
AAG	GGT	AAG	GTC	ACC	CCT	GTT	AAG	AAC	TAA	GGT	TCT	TGT	GGT	TCC	TGC	TGG	GCT	TTC	TCC	420
ACC	ATT	GGT	GCC	GTT	GAA	TCT	GCT	CTT	TGG	ATT	GCT	GGT	CAA	GGT	GAA	TAA	AAC	ACT	CTT	480
AAC	CTT	GCT	GAA	TAA	GAA	TAA	GTT	GAC	TGT	GCT	AAG	TCC	CCC	AAG	TAC	GAC	TCT	GAA	GGT	540
TGC	AAC	GGT	GGT	TGG	ATG	GTT	GAA	GGT	TTC	AAG	TAC	ATC	ATC	GAC	AAC	AAG	ATC	TCC	TAA	600
ACT	GCT	AAC	TAT	CCC	TAC	ACT	GCT	AAG	GAT	GGT	AAG	TGC	AAG	GAC	ACC	TCT	TCC	TTC	AAG	660
AAG	TTC	TCT	ATT	TCT	AAG	TAC	GCT	GAA	ATC	CCC	TAA	GGT	GAC	TGC	AAC	TCC	CTC	AAC	TCT	720
GCC	TTA	GAA	CAA	GGT	CCT	ATC	TCC	GTT	GCT	GTT	GAT	GCC	ACC	AAC	TTC	TAA	TTC	TAC	ACT	780
TCT	GGT	GTC	TTT	AAA	AAC	TGC	AAG	GCC	AAC	CTC	AAC	CAC	GGT	GTC	CTC	TTA	GTT	GCC	AAC	840
GTT	GAC	TCT	TCT	CTC	AAG	ATC	AAG	AAC	TCC	TGG	GGT	CCT	TCT	TGG	GGT	GAA	AAG	GGT	TTC	900
ATC	AGA	TTA	GCT	GCC	GGT	AAC	ACT	TGC	GGT	GTC	TGC	AAT	GCT	GCC	TCT	TAC	CCT	ATT	GTT	960
TGA	aaa	aac	ata	atc	caa	att	aaa	aaa	aat	tac	tca	aaa	ctg	ata	ata	taa	aaa	att	aat	1020
ttt	cat	aat	ttt	aat	gta	aat	aaa	tac	ctt	tat	att	tga	cgt	ttt	gta	ctc	aaa	ata	aat	1080
taa	agt	taa	caa	acc	ata	ttt	att	taa	ttc	tac	ttt	tca	att	ttt	aaa	aat	ata			1140

5

Mature protein (from N-terminus to C-terminus, amino acid 111 to 342):

[illegible]

3/4

Figure 3: HIC (hydrophobic interaction chromatography), sodium acetate-gradient and enzyme activity of the eluted PLA₁

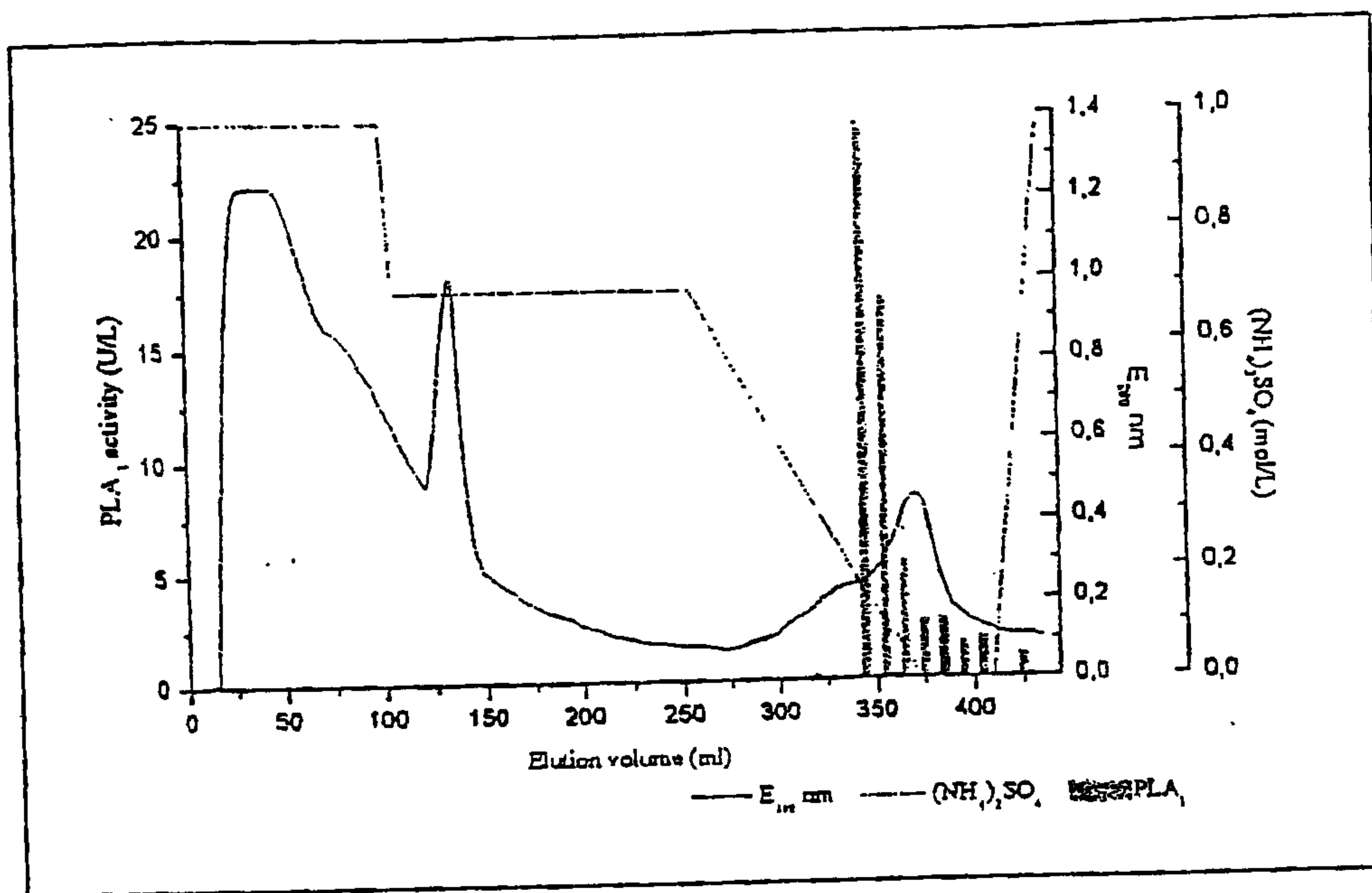
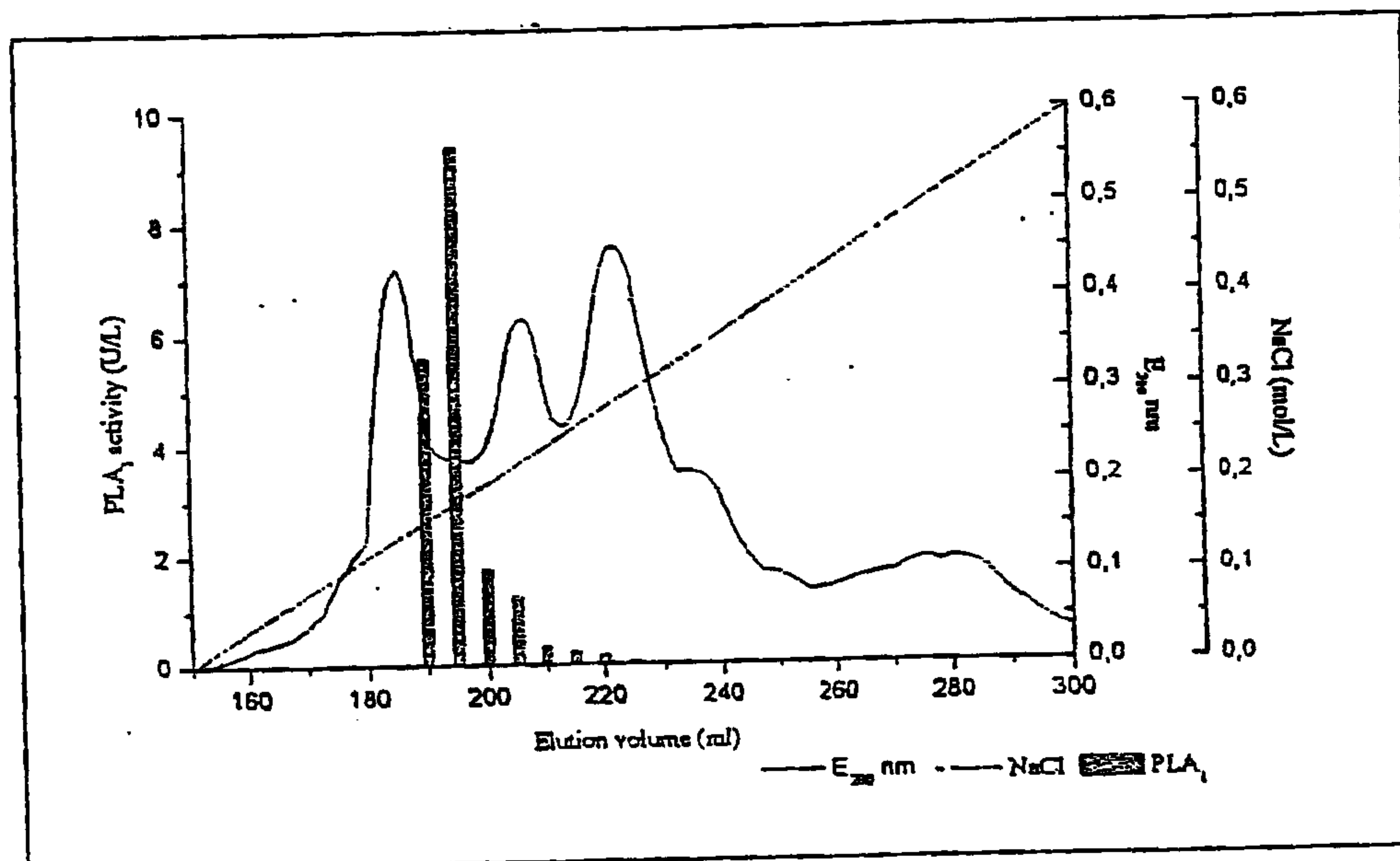
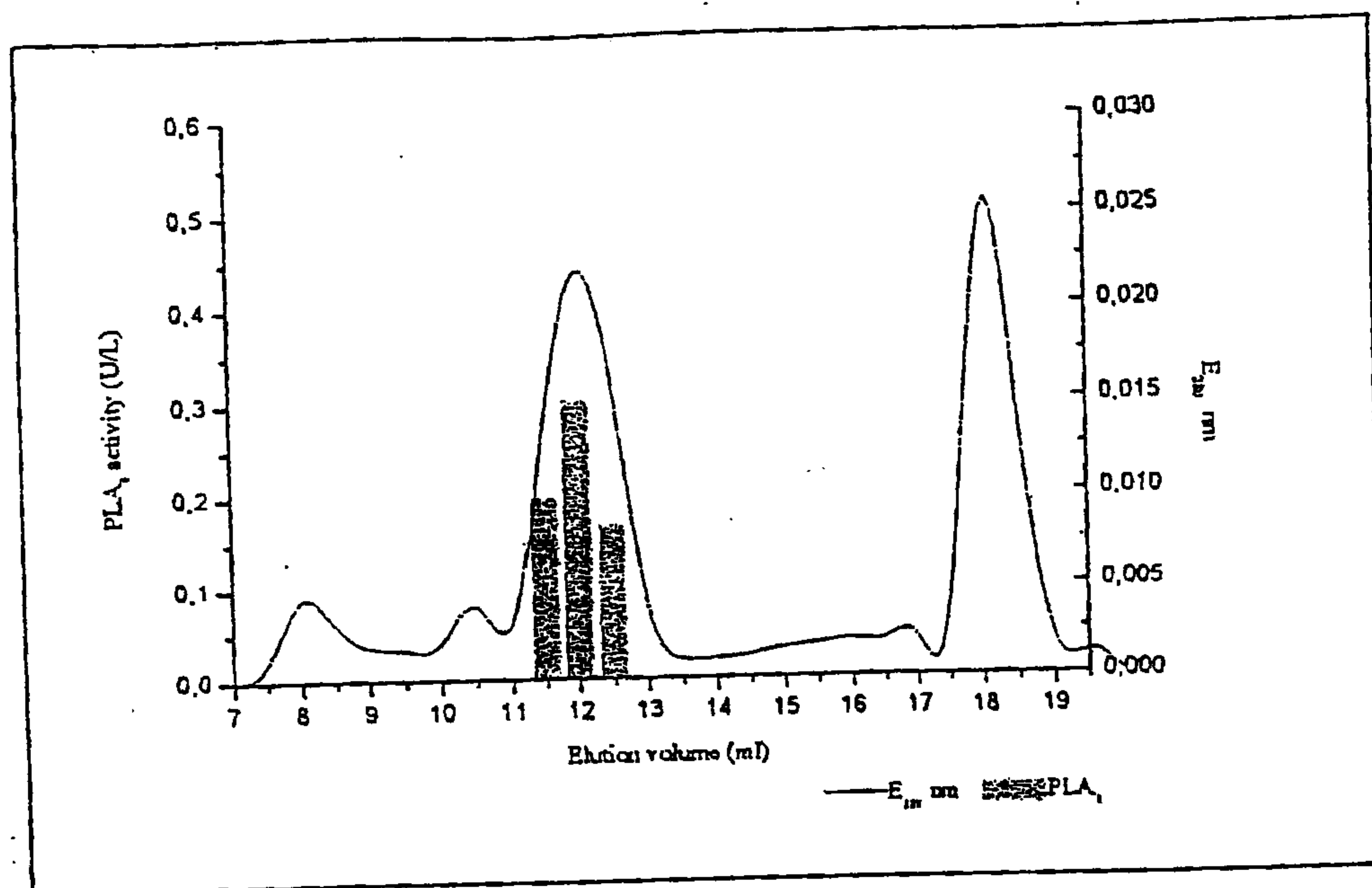


Figure 4: AEC (anion exchange chromatography), elution profile, sodium chloride gradient and enzyme activity of the eluted PLA₁



4/4

Figure 5: SEC (size exclusion chromatography), elution profile and enzyme activity of PLA₁



SEC (size exclusion chromatography), elution profile and enzyme activity of PLA₁

