



(22) Date de dépôt/Filing Date: 2004/12/02

(41) Mise à la disp. pub./Open to Public Insp.: 2005/06/05

(45) Date de délivrance/Issue Date: 2012/03/06

(30) Priorité/Priority: 2003/12/05 (EP03 028 101.8)

(51) Cl.Int./Int.Cl. *C12N 9/48* (2006.01)

(72) Inventeurs/Inventors:

GLASER, STEPHAN, DE;
GEIPEL, FRANK, DE;
KIRSCHBAUM, THOMAS, DE;
REXER, BERNHARD, DE;
THALHOFER, JOHANN-PETER, DE;
MUELLER, RAINER, DE;
GIESSEL, CLAUDIA, DE;
...

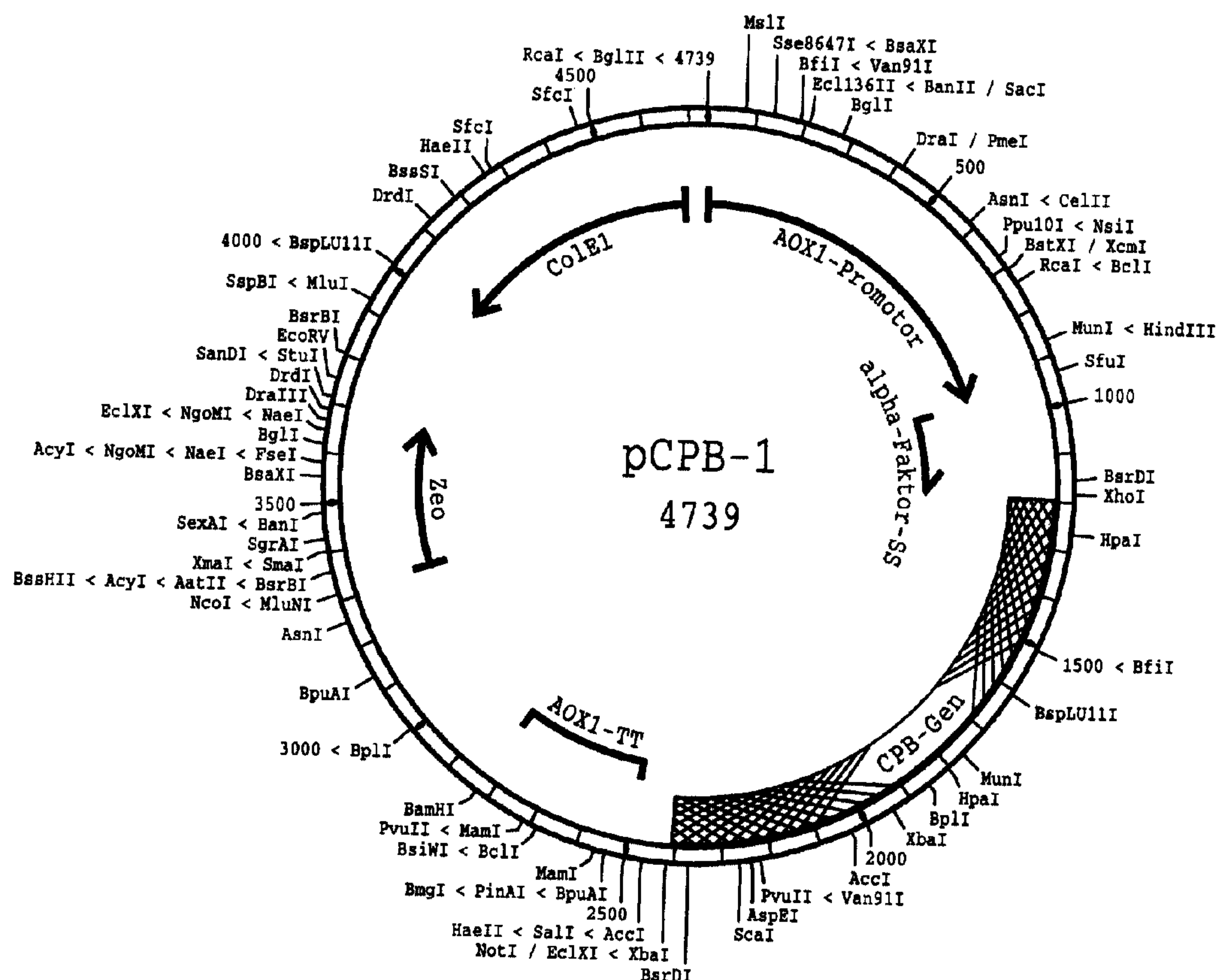
(73) Propriétaire/Owner:

F. HOFFMANN-LA ROCHE AG, CH

(74) Agent: BORDEN LADNER GERVAIS LLP

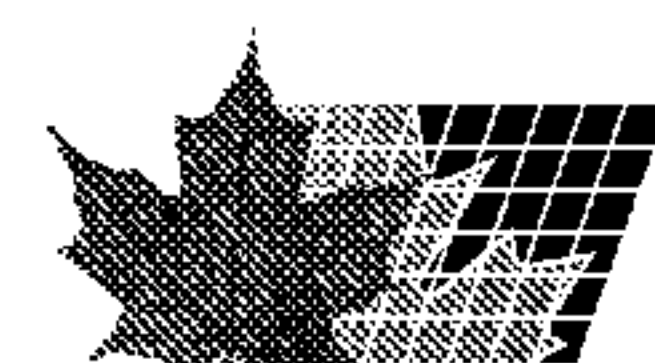
(54) Titre : CARBOXYPEPTIDASE B EXPRIMEE PAR ECOMBINAISON ET SA PURIFICATION

(54) Title: RECOMBINANTLY EXPRESSED CARBOXYPEPTIDASE B AND PURIFICATION THEREOF



(57) Abrégé/Abstract:

The invention provides a method to produce a protein with carboxypeptidase B activity from a pro-carboxypeptidase B zymogen, derived from a non-animal host organism. Carboxypeptidase B is activated from the zymogen using non-denaturing conditions. Particularly, the activation is performed under conditions that avoid unwanted non-covalent binding of the propeptide to the activated carboxypeptidase B enzyme.



(72) Inventeurs(suite)/Inventors(continued): ECKSTEIN, HELLMUT, DE; WOLF, ELVIRA, DE

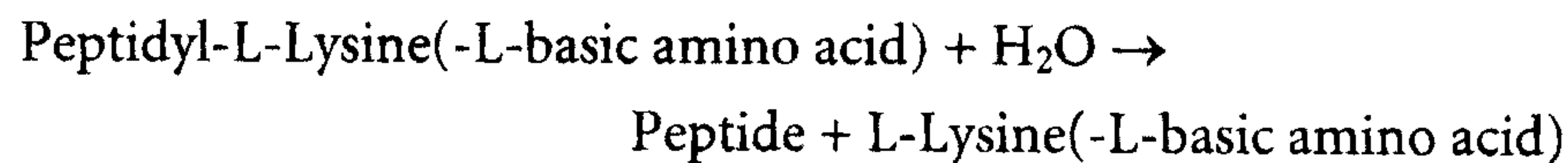
Abstract

The invention provides a method to produce a protein with carboxypeptidase B activity from a pro-carboxypeptidase B zymogen, derived from a non-animal host organism. Carboxypeptidase B is activated from the zymogen using non-denaturing conditions. Particularly, the activation is performed under conditions that avoid unwanted non-covalent binding of the propeptide to the activated carboxypeptidase B enzyme.

Recombinantly expressed carboxypeptidase B and purification thereof

The present invention pertains to the field of biotechnology. More specifically, the invention pertains to the production and further processing of an enzymatically inactive precursor of a protein with carboxypeptidase B activity. An aspect of the invention pertains to the activation and concomitant purification of the protein with carboxypeptidase B activity.

Carboxypeptidases are enzymes that hydrolyze C-terminal peptide bonds. The carboxypeptidase family includes metallo-, serine, and cysteine carboxypeptidases. According to their substrate specificity, these enzymes are referred to as carboxypeptidase A (cleaving aliphatic residues) or carboxypeptidase B (cleaving basic amino residues). Carboxypeptidase B (EC 3.4.17.2) is an enzyme that catalyzes hydrolysis of the basic amino acids, Lysine, Arginine, and Ornithine from the C-terminal position in polypeptides:



The precursor of carboxypeptidase B is a zymogen or proenzyme in the pancreas of most vertebrates (reviewed by Folk J. Carboxypeptidase B, in: The Enzymes 3, P. Boyer, Academic Press, NY, 57, 1971). Carboxypeptidase B may function in the further degradation of products of tryptic digestion, for which it is specific, very much as carboxypeptidase A is specific for products of chymotrypsin. Carboxypeptidase B also has an esterase activity which is related to the metal content of the enzyme (Folk, J., and Gladner, J.: Biochim Biophys Acta (1961) 48, 139-47; Zisapel, N. and Sokolovsky, M, Eur J Biochem (1975) 54, 541-7).

A "zymogen" or a "proenzyme" is the inactive precursor of an enzyme. In other words, a zymogen or a proenzyme is the inactive precursor of a protein with enzymatic activity such as a protein with carboxypeptidase B activity. Many proteins with enzymatic activity are synthesized as such inactive precursors and are subsequently activated by cleavage of one or a few specific peptides bonds. Activation of enzymes and other proteins by specific proteolysis occurs frequently in biological systems. For example: (a) The fibrous protein collagen which is

- 2 -

present in skin and bone is derived from procollagen which is a soluble precursor. (b) Some protein hormones are synthesized as inactive precursors. For example, insulin is derived from pro-insulin by proteolytic removal of a peptide. (c) The digestive enzymes that hydrolyze proteins are synthesized and secreted as zymogens in the stomach and pancreas. Carboxypeptidase B belongs to this group. (d) Blood clotting is mediated by a cascade of proteolytic activations that assures a rapid and amplified response to trauma.

In vivo the zymogen of carboxypeptidase B, that is to say "pro-carboxypeptidase B", is translated in pancreatic cells as a pre-protein termed "pre-pro-carboxypeptidase B".

Pre-pro-carboxypeptidase B comprises at its N-terminus a signal peptide. The amino acid sequence of pre-pro-carboxypeptidase B describes the primary translational product. The signal peptide directs pre-pro-carboxypeptidase B polypeptide to the secretory pathway of the pancreatic cell. During the secretion process the leader peptide is cleaved off proteolytically, thereby converting pre-pro-carboxypeptidase B to pro-carboxypeptidase B.

Pro-carboxypeptidase B is devoid of enzymatic activity. In vivo activation usually takes place in the small intestine. Trypsin cleaves pro-carboxypeptidase B, thereby generating an enzymatically inactive propeptide and the proteolytically active carboxypeptidase B enzyme.

As an example for pre-pro-carboxypeptidase, SEQ ID NO:1 depicts the amino acid sequence of rat pre-pro-carboxypeptidase B as published by Clauser E. et al., J. Biol. Chem. 1988, Vol. 263, 17837-45. Accordingly, the first 13 amino acids of the pre-pro-carboxypeptidase B polypeptide represent the signal peptide and the following 95 amino acids constitute the propeptide. The activation product, i.e. the active rat carboxypeptidase B enzyme contains 307 amino acids with a calculated molecular weight of 35,228 Da.

Using trypsin pro-carboxypeptidase B can also be activated in vitro to form carboxypeptidase B. The co-formation of the propeptide and the enzyme moiety has been shown by analytical methods like SDS gel electrophoreses as well as reversed phase HPLC by Ventura et al. J. Biol. Chem. (1999) 274(28), 19925-33.

- 3 -

Activation experiments in vitro under various conditions have shown that while a proteolytically active carboxypeptidase B enzyme is readily generated, under non-denaturing conditions the propeptide remains non-covalently attached to a large number of carboxypeptidase B molecules. This could be shown when denaturing
5 conditions were applied to the complex of the propeptide and the enzyme. Thus, non-denaturing purification methods and particularly anion exchange chromatography so far do not provide sufficient means to separate the non-covalently attached propeptide from the carboxypeptidase B enzyme.

Because of its high specificity for C-terminal basic amino acids, carboxypeptidase B
10 has found wide use, e.g. in end-group analysis for sequence determination. However, of major economical importance are industrial applications which involve proteolytic processing. Particularly, carboxypeptidase B is used in production processes that lead to insulin for medical use and that involve proteolytic cleavage of pro-insulin.

Regarding insulin as a medical product a particular problem arises by virtue of the
15 carboxypeptidase B propeptide that is non-covalently attached to the carboxypeptidase B enzyme. In said insulin production processes, subsequent to proteolytic cleavage of pro-insulin, the insulin fragment is purified under denaturing conditions. As a consequence, it is possible that the carboxypeptidase B
20 propeptide co-purifies with insulin. In such a case separation of the carboxypeptidase B propeptide from the insulin molecule can be a laborious process which at the same time is likely to decrease the final yield of purified insulin.

Commercially available carboxypeptidase B purified from porcine pancreas may
25 not be totally free of other proteolytic activities. Moreover, as it is the case for most animal-derived products, porcine carboxypeptidase B may contain infectious agents such as viruses, prions, or other biologically active components with deleterious potential for human health. It is therefore desired to obtain carboxypeptidase B from an alternative source. In addition, it is desired that
30 carboxypeptidase B from the alternative source is substantially free of the propeptide.

It is known to the art that carboxypeptidase B can be produced recombinantly in transformed host organisms (e.g. Ventura et al. J. Biol. Chem. (1999) 274(28), 19925-33) and purified therefrom.

5 WO 01/51624 discloses recombinant expression of porcine pro-carboxypeptidase B in the methylotrophic yeast *Pichia pastoris*. Following activation of the zymogen by way of tryptic cleavage, a first purification step includes hydrophobic chromatography. Following the addition of soybean trypsin inhibitor, the activated enzyme is further purified using Q-sepharose^{*} chromatography. The purification method of WO 01/51624 involves multiple steps and there is concomitant loss of
10 carboxypeptidase B product. In addition, there is the need to separate the soybean trypsin inhibitor as well as the carboxypeptidase B propeptide from the product.

DE 19 915 938 discloses recombinant expression of a Histidine-tagged human pro-carboxypeptidase B in the methylotrophic yeast *Pichia pastoris*. The product is purified using affinity chromatography and by virtue of the Histidine-tag.
15 Subsequently, pro-carboxypeptidase B is activated using trypsin and the activation reaction is terminated by the addition of soybean trypsin inhibitor. By way of filtering through a membrane which excludes particles above a molecular weight of about 30,000 Da, carboxypeptidase B is separated from trypsin, the propeptide, the inhibitor as well as other fragments that were generated during the course of the
20 purification procedure. However, as the molecular weight of carboxypeptidase B enzyme is about 37,000 Da, losses can be expected due to the filtration step, depending on the size distribution of the pores in the filter used. Moreover, as the activation step is performed on the dissolved zymogen there is the need to further separate the propeptide from the carboxypeptidase B enzyme.

25 In WO 96/23064 recombinant expression of rat pro-carboxypeptidase B in *E.coli* is described. Inclusion bodies of insoluble pro-carboxypeptidase B are produced. The zymogen is renatured and subsequently purified further. The activation step is performed on the dissolved zymogen and there is the need to further separate the propeptide from the carboxypeptidase B enzyme. Owing to the
30 denaturation/renaturation steps the yield of carboxypeptidase B enzyme obtainable using this method is comparably low.

*Trade-mark

- 5 -

To the knowledge of the inventors, no specific method to separate activated carboxypeptidase B from the non-activated form is known.

5 The problem to be solved by the present invention is therefore to provide an alternative method to produce a protein with carboxypeptidase B activity from a zymogen, whereby non-covalent attachment of the propeptide to the protein with carboxypeptidase B activity is avoided. Another problem to be solved by the present invention is to provide an alternative method to separate the protein with carboxypeptidase B activity from the propeptide and from residual zymogen following the activation step. Another problem to be solved by the present
10 invention is to provide an alternative method to separate under non-denaturing conditions the protein with carboxypeptidase B activity from the propeptide and from residual zymogen, following the activation step.

According to the invention a method is provided to produce a protein with carboxypeptidase B activity, comprising the steps of a) providing a vector
15 comprising a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag and a signal peptide, whereby optionally between the Histidine-tag and the signal peptide or between the Histidine-tag and rat carboxypeptidase B a spacer sequence is inserted; (b) transforming a microbial host organism with the vector; (c) cultivating the
20 microbial host organism in a growth medium containing nutrients and a carbon source, whereby the microbial host organism expresses the pre-protein and secretes the Histidine-tagged pro-carboxypeptidase B into the growth medium; (d) immobilizing the secreted Histidine-tagged pro-carboxypeptidase B in the growth medium of step (c) on a particulate metal chelate affinity matrix capable of binding
25 the Histidine-tag, and washing the particulate metal chelate affinity matrix, whereby the Histidine-tagged pro-carboxypeptidase B is immobilized; (e) incubating the particulate metal chelate affinity matrix with the immobilized Histidine-tagged pro-carboxypeptidase B of step (d) in a buffer containing trypsin, thereby cleaving proteolytically the pro-carboxypeptidase B moiety and releasing
30 the protein with carboxypeptidase B activity into the liquid phase, whereby the Histidine-tagged propeptide moiety is immobilized; (f) separating the liquid phase containing the protein with carboxypeptidase B activity from the particulate metal chelate affinity matrix, whereby the Histidine-tagged propeptide moiety is

- 6 -

immobilized on the particulate metal chelate affinity matrix ; and (g) purifying the protein with carboxypeptidase B activity from the liquid phase of step (f).

5 Certain terms are used with particular meaning, or are defined for the first time, in this description of the present invention. For the purposes of the present invention, the following terms are defined by their art-accepted definitions, when such exist, except that when those definitions conflict or partially conflict with the definitions set forth below. In the event of a conflict in definition, the meaning of the terms are first defined by the definitions set forth below.

10 Amino acid identification uses the the three-letter abbreviations as well as the single-letter alphabet of amino acids, i.e., Asp D Aspartic acid, Ile I Isoleucine, Thr T Threonine, Leu L Leucine, Ser S Serine, Tyr Y Tyrosine, Glu E Glutamic acid, Phe F PhenylAlanine, Pro P Proline, His H Histidine, Gly G Glycine, Lys K Lysine, Ala A Alanine, Arg R Arginine, Cys C Cysteine, Trp W Tryptophan, Val V Valine, Gln Q Glutamine, Met M Methionine, Asn N Asparagine. An amino acid at a particular
15 position in an amino acid sequence is given by its three-letter abbreviation and a number. As an example, referring to the amino acid sequence of native rat pre-pro carboxypeptidase B of SEQ ID NO: 1, "His14" denotes the Histidine residue at amino acid position 14.

20 The term "comprising" is used in the description of the invention and in the claims to mean "including, but not necessarily limited to".

The term "pre-protein" refers to a primary translation product comprising at its N-terminus a signal peptide. In the context of the present invention the "pre-protein" is the precursor of zymogen or pro-enzyme, i.e. pro-carboxypeptidase B. The pro-enzyme results from post-translational processing of the pre-protein.

25 A "signal peptide" is a cleavable signal sequence of amino acids present in the pre-protein form of a secretable protein. Proteins transported across the cell membrane, i.e. "secreted", typically have an N- terminal sequence rich in hydrophobic amino acids about 15 to 30 amino acids long. Sometime during the process of passing through the membrane, the signal sequence is cleaved by a signal peptidase
30 (Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (eds), Molecular Biology of the Cell, fourth edition, 2002, Garland Science Publishing). Many

- 7 -

sources of signal peptides are well known to those skilled in the art and can include, for example, the amino acid sequence of the α -factor signal peptide from *Saccharomyces cerevisiae* and the like. In general, the pre-protein N-terminus of essentially any secreted protein is a potential source of a signal peptide suitable for use in the present invention. A signal peptide can also be bipartite comprising two signal peptides directing the pre-protein to a first and a second cellular compartment. Bipartite signal peptides are cleaved off stepwise during the course of the secretory pathway. A preferred example therefor is the prepro peptide of the α -factor from *Saccharomyces cerevisiae* (Waters et al., J. Biol. Chem. 263 (1988) 6209-14). An even more preferred example is the amino acid sequence encoded by the nucleotide sequence from position 1 to position 255 in SEQ ID NO: 3.

Pre-proteins with a secretory N-terminal signal peptide are directed to enter the "secretory pathway". The secretory pathway comprises the processes of post-translational processing and finally results in secretion of pro-carboxypeptidase B. In the present document it is understood that proteins secreted by methylotrophic yeast strains have passed through the secretory pathway.

The term "post-translational processing" denotes the modification steps a pre-protein is subjected to, in order result in a protein in a cellular or extracellular compartment. In the context of the present invention post-translational processing of pre-pro-carboxypeptidase B results in secreted pro-carboxypeptidase B.

Enzymatic proteolytic cleavage of pro-carboxypeptidase B, e.g. by trypsin, is referred to as "activation". It is well known to the skilled artisan that the molecular events leading to activation depend on proteolytic processing. Trypsin cleaves pro-carboxypeptidase B, thereby generating an enzymatically inactive "propeptide" or "propeptide moiety" and a proteolytically active "enzyme moiety", that is to say carboxypeptidase B. By way of example, enzyme moiety of rat pro-carboxypeptidase B, that is to say the "pro-carboxypeptidase B moiety" of the rat procarboxypeptidase B polypeptide is the polypeptide given by the amino acid sequence from position 191 to position 497 in SEQ ID NO: 4.

The propeptide of a pro-carboxypeptidase B usually comprises about 95 amino acids. It is also known that proteolytic cleavage of pro-carboxypeptidase B by

trypsin not only activates carboxypeptidase B but also may form tryptic fragments of carboxypeptidase B.

5 A “methylophilic yeast” is defined as a yeast that is capable of utilising methanol as its carbon source. The term also comprises laboratory strains thereof. In case a methylophilic yeast strain is auxotrophic and because of this needs to be supplemented with an auxillary carbon-containing substance such as, e.g. Histidine in the case of a methylophilic yeast strain unable to synthesise this amino acid in sufficient amounts, this auxillary substance is regarded as a nutrient but not as a carbon source.

10 A “vector” is defined as DNA which can comprise, i.e. carry and maintain the DNA fragment of the invention, including, for example, phages and plasmids. These terms are understood by those of skill in the art of genetic engineering. The term “expression cassette” denotes a nucleotide sequence encoding a pre-protein, operably linked to a promoter and a terminator. As for vectors containing an
15 expression cassette, the terms “vector” and “expression vector” are synonyms.

“Transformation” means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration.

20 The term “episome” denotes a unit of genetic material composed of a series of genes that sometimes has an independent existence in a host cell and at other times is integrated into a chromosome of the cell, replicating itself along with the chromosome. An example for an episome is the vector of Figure 1.

The term “expression” and the verb “to express” denote transcription of DNA sequences and/or the translation of the transcribed mRNA in a host organism resulting in a pre-protein, i.e. not including post-translational processes.

25 A nucleotide sequence “encodes” a peptide or protein when at least a portion of the nucleotide sequence, or its complement, can be directly translated to provide the amino acid sequence of the peptide or protein, or when the isolated nucleotide sequence can be used, alone or as part of an expression vector, to express the peptide or protein in vitro, in a prokaryotic host cell, or in a eukaryotic host cell.

All nucleotide sequences are written in the direction from the 5' (stands for prime) end to the 3' end also referred to as 5' to 3'.

5 A "promoter" is a regulatory nucleotide sequence that stimulates transcription. These terms are understood by those of skill in the art of genetic engineering. Like a promoter, a "promoter element" stimulates transcription but constitutes a sub-fragment of a larger promoter sequence.

10 The term "operably linked" refers to the association of two or more nucleic acid fragments on a single vector so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence, i.e. a nucleotide sequence encoding a protein or a pre-protein, when it is capable of affecting the expression of that coding sequence, i.e., that the coding sequence is under the transcriptional control of the promoter.

15 A "peptide bond" is a covalent bond between two amino acids in which the α -amino group of one amino acid is bonded to the alpha-carboxyl group of the other amino acid.

Detailed description of the invention

20 A first embodiment of the invention is a method to produce a protein with carboxypeptidase B activity, comprising the steps of a) providing a vector comprising a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag and a signal peptide, whereby optionally between the Histidine-tag and the signal peptide or between the Histidine-tag and rat carboxypeptidase B a spacer sequence is inserted; (b) transforming a microbial host organism with the vector; (c) cultivating the
25 microbial host organism in a growth medium containing nutrients and a carbon source, whereby the microbial host organism expresses the pre-protein and secretes the Histidine-tagged pro-carboxypeptidase B into the growth medium; (d) immobilizing the secreted Histidine-tagged pro-carboxypeptidase B in the growth medium of step (c) on a particulate metal chelate affinity matrix capable of binding
30 the Histidine-tag, and washing the particulate metal chelate affinity matrix, whereby the Histidine-tagged pro-carboxypeptidase B is immobilized; (e) incubating the particulate metal chelate affinity matrix with the immobilized Histidine-tagged pro-

- 10 -

carboxypeptidase B of step (d) in a buffer containing trypsin, thereby cleaving proteolytically the pro-carboxypeptidase B moiety and releasing the protein with carboxypeptidase B activity into the liquid phase, whereby the Histidine-tagged propeptide moiety is immobilized; (f) separating the liquid phase containing the protein with carboxypeptidase B activity from the particulate metal chelate affinity matrix, whereby the Histidine-tagged propeptide moiety is immobilized on the particulate metal chelate affinity matrix ; and (g) purifying the protein with carboxypeptidase B activity from the liquid phase of step (f).

When constructing a vector to be used for transformation of a microbial host strain, e.g. a methylotrophic yeast strain, some molecular cloning techniques may require the addition of a linker or spacer sequence to a nucleotide sequence that encodes a functional element of the pre-protein. A spacer sequence may also be generated, e.g., when a DNA fragment comprising the coding sequence of a functional element of the pre-protein is excised from a cloning vector using a restriction endonuclease generating a fragment that is longer than the coding sequence. Other reasons why linker or spacer nucleotide sequences may occur are possible. Nucleotide sequences encoding functional elements of the pre-protein in this regard encode (a) the signal peptide, (b) the Histidine-tag and (c) the pro-carboxypeptidase B moiety. Thus, owing to the presence of such a linker nucleotide sequence, additional amino acids, i.e. a "spacer sequence", may be inserted at the junctions of any two functional elements of the pre-protein. An example therefor is the spacer sequence consisting of a Serine residue and an Alanine residue at amino acid positions 86 and 87 in SEQ ID NO: 3 and SEQ ID NO: 4. Preferred in the pre-protein of the invention are spacer sequences consisting of not more than four amino acid residues. More preferred are spacer sequences consisting of two amino acid residues. The insertion of a linker nucleotide sequence encoding a spacer sequence is "optional" in the sense that said insertion may facilitate the joining of the separate nucleotide sequences encoding two functional elements, in order to form a nucleotide sequence encoding the pre-protein or a precursor thereof.

A Histidine-tag is an amino acid sequence containing preferably 6 consecutive Histidines. As the Histidines represent the essential portion, there are few additional amino acids comprised in the Histidine-tag moiety. Importantly, the Histidine-tag used in the present invention does not add another relevant trypsin cleavage site to the secreted protein. Thus, the trypsin cleavage site of the zymogen

remains the preferred cleavage site under the conditions applied during on-column activation of immobilized Histidine-tagged pro-carboxypeptidase B. Upon on-column activation of the immobilized Histidine-tagged pre-carboxypeptidase B the propeptide remains bound to the particulate metal chelating affinity matrix via the Histidine-tag.

Purification of secreted Histidine-tagged pro-carboxypeptidase B is facilitated by immobilized metal affinity chromatography. This method is a widely employed method to purify recombinant proteins containing a short affinity-tag consisting of histidine residues (Histidine-tag). Immobilized metal-affinity chromatography (described by Porath, J. et al., *Nature* 258 (1975) 598–599) is based on the interaction between a transition metal ion (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) immobilized on a particulate metal chelating affinity matrix and specific amino acid side chains. Histidine is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices, as electron donor groups on the histidine imidazole ring readily form coordination bonds with the immobilized transition metal. Peptides containing sequences of consecutive histidine residues are efficiently retained on particulate metal chelating affinity. Following washing of the matrix material, peptides containing histidine sequences such as a Histidine-tag can be easily eluted by either adjusting a low (acidic) pH of the column buffer or by adding free imidazole.

The method to purify proteins with histidine residues was first described in by Hochuli, E. et al., *J. Chromatogr.* 411 (1987) 177–184. The document describes a nitrilotriacetic acid (NTA) adsorbent for metal-chelate affinity chromatography. The NTA resin forms a quadridentate chelate and is especially suitable for metal ions with coordination numbers of six, since two valencies remain for the reversible binding of biopolymers. Dihydrofolate reductase with a Histidine-tag was successfully purified with Ni^{2+} -NTA matrices as described by Hochuli, E. et al., *Bio/Technology* 6 (1988) 1321–1325. The purification efficiency of this system was dependent on the length of the Histidine-tag and the solvent system. While the system worked efficiently with His6-tagged proteins under denaturing conditions, His3-tagged proteins were efficiently purified under physiological conditions. However, His6-tagged proteins can be bound to Ni^{2+} -NTA matrices under non-denaturing conditions in low- or high-salt buffers. After binding, the target protein can be eluted by an imidazole gradient from 0.8 to 250 mM. Washing with a low

- 12 -

concentration of imidazole (e.g. 0.8 mM) can be used to reduce unspecific binding of host proteins with histidines.

In the present invention, the immobilization step is preferably performed in the presence of imidazole. Under these conditions, unspecific binding of other Histidine-containing proteins has been found to be reduced. Preferred concentrations of imidazole in this regard are in the range between 0.01 mM and 1 mM.

Another material that has been developed to purify Histidine-tagged proteins is TALON.* It consists of a Co^{2+} - carboxylmethyldaspartate (Co^{2+} -CMA), which is coupled to a solid-support resin. TALON has been reported to exhibit less unspecific protein binding than the Ni^{2+} -NTA resin, resulting in higher elution product purity (Chaga, G. et al., *Biotechnol. Appl. Biochem.* 29 (1999) 19–24; Chaga, G. et al., *J. Chromatogr. A* 864 (1999) 257–256).

Histidine-tags are commonly placed on either the N- or the C-terminus of recombinant proteins. Optimal placement of the tag is protein-specific. Purification using Histidine tags has been carried out successfully using a number of expression systems including bacteria (Chen, B.P. and Hai, T., *Gene* 139(1994) 73–75; Rank, K.B. et al., *Protein Expr. Purif.* 22 (2001) 258–266), yeast (Borsing, L. et al. *Biochem. Biophys. Res. Commun.* 240 (1997) 586–590; Kaslow, D.C. and Shiloach, J., *Bio/Technology* 12 (1994) 494–499), mammalian cells (Janknecht, R. et al. *Proc. Natl. Acad. Sci. USA* 88 (1991) 8972–8976; Janknecht, R. and Nordheim, A., *Gene* 121 (1992) 321–324), and baculovirus-infected insect cells (Kuusinen, A. et al., *Eur. J. Biochem.* 233 (1995) 720–726; Schmidt, M. et al. *Protein Expr. Purif.* 12 (1998) 323–330). More than 100 structures of Histidine-tagged proteins have been deposited in the Protein Data Bank. Proteins with a Histidine-tag may vary slightly as far as their mosaicity and diffraction compared to the native protein (Hakansson, K. et al. *Acta Crystallogr. D Biol. Crystallogr.* 56 (2000) 924–926). In principle, it cannot be excluded that the Histidine-tag may interfere with protein activity (Wu, J. and Filutowicz, M., *Acta Biochim. Pol.* 46 (1999) 591–599), although the relatively small size and charge of the Histidine-tag ensure that protein activity is rarely affected. Moving the Histidine-tag to the opposite terminus (Halliwell, C.M. et al., *Anal Biochem.* 295 (2001) 257–261) or carrying out the purification under denaturing conditions often solves this problem.

*Trade-mark

- 13 -

The secreted Histidine-tagged pro-carboxypeptidase B in the growth medium is immobilized on a particulate metal chelate affinity matrix capable of binding a Histidine tag. For instance, the Histidine-tagged pro-carboxypeptidase B can be adsorbed to metal ions immobilized on a metal-chelating resin. As described above, such resins are well known to the skilled artisan. A preferred particulate metal chelate affinity matrix is chromatography material that is coated with nickel-nitrilotriacetic acid (Ni-NTA). In this regard, the person skilled in the art is aware of EP 0 253 303, EP 0 282 042, and EP 1 069 131. Qiagen commercializes the QIAexpress^{*} purification system which can be used for the immobilization step. The same company provides an expression vector (pQE) that can be used to produce Histidine-tagged fusion polypeptides. In this regard, the person skilled in the art is also aware of US 5,284,933 and US 5,310,663.

Purification of protein with a metal center may require especially adapted methods for purification because the metal can be absorbed by the NTA. Purification under anaerobic conditions may also require especially adapted methods for purification because Ni^{2+} -NTA is reduced. Nevertheless, purification of pre-proteins with a Histidine-tag is one of the the most commonly used method.

Carboxypeptidase B contains Zn^{2+} as a cofactor. For this reason, instead of using Ni^{2+} -NTA as a matrix for chromatographic purification, a metal chelating affinity matrix loaded with Zn^{2+} is preferred. Thereby any accidental and unwanted replacement of the Zn^{2+} cofactor by Ni^{2+} is avoided.

An even more preferred particulate metal chelate affinity matrix is therefore Zn^{2+} -loaded (i.e. Zn^{2+} ions immobilized on) StreamlineTM Chelating adsorbent (Amersham Biosciences). StreamlineTM adsorbents are based on agarose. The macroporous structure of the highly cross-linked agarose matrices combines good binding capacities for large molecules, such as proteins, with high chemical and mechanical stability. High mechanical stability is an important property of a matrix to be used in expanded bed chromatography to reduce the effects of attrition when particles are moving freely in the expanded bed. Particles made only of organic material have limited density and would need to have very large diameters for the high sedimentation velocity required. Such large particle diameters result in long diffusional path lengths, which cause considerable mass transfer resistance, counteracting productivity. StreamlineTM adsorbents are therefore based on a

^{*}Trade-mark

composite particle containing an inert core material that is denser than organic materials. Such particles can be designed so that their sedimentation velocity is high also at a reasonable particle size.

- 5 Expanded bed adsorption is a single pass operation in which desired proteins are purified from crude, particulate containing feed-stock without the need for separate clarification, concentration and initial purification. The expansion of the adsorbent bed creates a distance between the adsorbent particles, i.e. increased voidage (void volume fraction) in the bed, which allows for unhindered passage of cells, cell debris and other particulates during application of crude feed to the column.
- 10 The particulate metal chelate affinity matrix (such as the Streamline™) adsorbent is expanded and equilibrated by applying an upward liquid flow to the column. A stable fluidized bed is formed when the adsorbent particles are suspended in equilibrium due to the balance between particle sedimentation velocity and upward liquid flow velocity. The column adaptor is positioned in the upper part of the
- 15 column during this phase. Crude, unclarified feed, i.e. culture medium containing secreted Histidine-tagged pro-carboxypeptidase B and the microbial host organism, is applied to the expanded bed with the same upward flow as used during expansion and equilibration. Histidine-tagged pro-carboxypeptidase B is immobilized by way of binding with the Histidine-tag to the adsorbent while cell
- 20 debris, cells, particulates and contaminants pass through unhindered. Weakly bound material, such as residual cells, cell debris and other type of particulate material, is washed out from the expanded bed using upward liquid flow. When the weakly retained material has been washed out from the bed the captured Histidine-tagged pro-carboxypeptidase B is activated by on-column cleavage with trypsin,
- 25 whereby the upward flow mode is maintained. Carboxypeptidase B is released and can be washed from the column. The flow-through contains the target protein, increased in concentration, clarified, partly purified, and ready for further purification by packed bed chromatography.

- 30 How to use a preferred particulate metal chelate affinity matrix, i.e. Streamline™ Chelating adsorbent (Amersham Biosciences) and how to set up, optimize and perform expanded bed chromatography is further detailed in in the Pharmacia Biotech manual "Expanded bed adsorption, principles and methods" (ISBN 91-

- 15 -

630-5519-8). General instructions how to regenerate the column are also described therein.

5 A large number of microbial host organisms can be used in the present invention. The main requirement is that the microbial host organism is engineered and cultured such that it secretes Histidine-tagged pro-carboxypeptidase B into the growth medium. In a preferred embodiment the microbial host organism is a prokaryotic organism. In this regard, the skilled artisan is well aware of commercially available bacterial systems for recombinant expression and secretion which are based on bacteria such as *E. coli*, *Bacillus* sp., *Staphylococcus* sp.. In a
10 more preferred embodiment the microbial host organism is a microbial eukaryotic organism. Very much preferred is a yeast species. In an even more preferred embodiment, the microbial organism is a methylotrophic yeast strain.

Thus, in another embodiment of the invention, Histidine-tagged rat pro-carboxypeptidase B is produced by recombinant means using methylotrophic yeast
15 as a non-animal host organism. Methylotrophic yeasts have the biochemical pathways necessary for methanol utilization and are classified into four genera, based upon cell morphology and growth characteristics: *Hansenula*, *Pichia*, *Candida*, and *Torulopsis*. The most highly developed methylotrophic host systems utilize *Pichia pastoris* (*Komagataella pastoris*) and *Hansenula polymorpha* (*Pichia angusta*).
20

Expression of heterologous proteins in yeast is described in US 5,618,676, US 5,854,018, US 5,856,123, and US 5,919,651.

Yeast organisms produce a number of proteins that are synthesized intracellularly but have a function outside the cell. These extracellular proteins are referred to as
25 secreted proteins. Initially the secreted proteins are expressed inside the cell in the form of a precursor or a pre-protein containing an N-terminal signal peptide ensuring effective direction of the expressed product into the secretory pathway of the cell, across the membrane of the endoplasmic reticulum. The signal peptide is generally cleaved off from the desired product during translocation. Cleavage is
30 effected proteolytically by a signal peptidase. A particular sub-sequence of amino acids of the signal peptide is recognised and cleaved by the signal peptidase. This sub-sequence is referred to as signal peptidase cleavage site. Once having entered

- 16 -

the secretory pathway, the protein is transported to the Golgi apparatus. From the Golgi apparatus the proteins are distributed to the plasma membrane, lysosomes and secretory vesicles.

5 Secreted proteins are confronted with different environmental conditions as opposed to intracellular proteins. Part of the processes of the secretory pathway is to stabilise the maturing extracellular proteins. Therefore, pre-proteins that are passed through the secretory pathway of yeast undergo specific posttranslational processing. For example, processing can comprise the generation of disulfide bonds to form intramolecular cross-links. Moreover, certain amino acids of the protein
10 can be glycosylated.

Several approaches have been suggested for the expression and secretion in yeast of proteins heterologous to yeast. EP 0 116 201 describes a process by which proteins heterologous to yeast are transformed by an expression vector harboring DNA encoding the desired protein, a signal peptide and a peptide acting as a signal
15 peptidase cleavage site. A culture of the transformed organism is prepared and grown, and the protein is recovered from culture media. For use in yeast cells a suitable signal peptide has been found to be the α -factor signal peptide from *Saccharomyces cerevisiae* (US 4,870,008).

During secretion, the yeast enzyme KEX-2 is the signal peptidase which recognizes a
20 Lysine-Arginine sequence as its cleavage site in the pre-protein. KEX-2 cleaves at the junction to the sequence of the desired protein. As a result, the desired gene product is released and free of the leader portions, i.e. the signal peptide of the pre-protein. KEX-2 endoprotease was originally characterised in *Saccharomyces* yeast where it specifically processes the precursor of mating type α -factor and a killer
25 factor (Julius, D., et al., Cell 37 (1984) 1075-1089). Methylophilic yeast species such as *Pichia pastoris* share the KEX-2-type protease (similar role and function) with *Saccharomyces cerevisiae* (Werten, M.W., et al., Yeast 15 (1999) 1087-1096).

A well-established methylophilic yeast species exemplarily described as host for high-level recombinant protein expression is *Pichia pastoris* (US 4,683,293,
30 US 4,808,537, US 4,812,405, US 4,818,700, US 4,837,148, US 4,855,231,
US 4,857,467, US 4,879,231, US 4,882,279, US 4,885,242, US 4,895,800,
US 4,929,555, US 5,002,876, US 5,004,688, US 5,032,516, US 5,122,465,

- 17 -

US 5,135,868, US 5,166,329, WO 00/56903). In the absence of glucose, *Pichia pastoris* uses methanol as a carbon source which at the same time is a hallmark of a methylotrophic organism. The alcohol oxidase (AOX1) promoter given in SEQ ID NO: 5 controls expression of alcohol oxidase, which catalyses the first step in methanol metabolism. Typically, 30% of the total soluble protein in methanol-induced cells is alcohol oxidase. Several *Pichia* expression vectors carry the AOX1 promoter and use methanol to induce high-level expression of desired heterologous proteins. Expression constructs also integrate into the *Pichia pastoris* genome, creating a transformed and genetically stable host.

Using an expression vector encoding a heterologous pre-protein comprising a signal peptide or a signal peptide with a signal peptidase cleavage site, and a desired protein, methylotrophic yeast strains such as *Pichia pastoris* strains can be manipulated in order to secrete the desired product into the growth medium from where the secreted protein can be purified. It may be advantageous to produce nucleotide sequences encoding the pre-protein possessing a substantially different codon usage.

Regarding the encoded pro-carboxypeptidase B polypeptide, the nucleotide sequence comprised in SEQ ID NO: 3 is different from previously published nucleotide sequences such as SEQ ID NO: 1 because of the degeneracy of the genetic code. The nucleotide sequence of SEQ ID NO: 3 from position 286 to position 1497 encodes the same polypeptide as the nucleotide sequence of SEQ ID NO: 1 from position 40 to position 1248. However, SEQ ID NO: 3 incorporates two amino acid exchanges, i.e. Lys201Asn and Arg329Asp. These amino acid exchanges are also documented in WO 96/23064. The term "degenerate code" indicates that in the genetic code a particular amino acid can be coded by two or more different codons. Degeneracy occurs because of the fact that of the 64 possible base triplets, 3 are used to code the stop signals, and the other 61 are left to code for only 20 different amino acids.

Thus, codons may be selected to increase the rate at which expression of the pre-protein occurs in a particular yeast expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding the pre-protein, without altering the encoded amino acid sequence, include the production of RNA

- 18 -

transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence. The nucleotide sequence of SEQ ID NO: 3 is an example for a coding sequence that has been optimized in this fashion.

5 Using a vector comprising the nucleotide sequence encoding the pre-protein that is competent for expression, i.e. operably linked to a promoter or promoter element and to a terminator or terminator element, as well as to sequences required for efficient translation, the host organism is transformed with the vector and transformants are selected. Transformants are then analyzed with respect to the
10 yield of recombinant protein secreted into the growth medium. Transformants secreting the highest quantities of recombinant protein are selected. Thus, transformants secreting the highest amounts of Histidine-tagged pro-carboxypeptidase B are selected.

On the one hand, expression yield is dependent on proper targeting of the desired
15 product, that is to say targeting the pre-protein to the secretory pathway of yeast by means of the signal peptide. An example for a signal peptide is the α -factor signal peptide from *Saccharomyces cerevisiae* encoded in SEQ ID NO: 3 from position 1 to position 255. On the other hand, expression yield can be increased by increasing the dosage of the gene encoding the desired product, i.e. the copy number of the
20 expression construct in the host organism is amplified. One way to accomplish this is by multiple transformation of an expression vector encoding the desired product. Another way is to introduce the gene encoding the desired product into the host organism using a first and a second expression vector, whereby the second expression vector is based on a selectable marker which differs from the selectable
25 marker used in the first expression vector. The second expression vector encoding the same desired product can even be introduced when the host organism already carries multiple copies of a first expression vector (US 5,324,639; Thill, G.P., et al., Positive and Negative Effects of Multi-Copy Integrated Expression in *Pichia pastoris*, International Symposium on the Genetics of Microorganisms 2 (1990),
30 pp. 477-490; Vedvick, T., et al., J. Ind. Microbiol. 7 (1991) 197-201; Werten, M.W., et al., Yeast 15 (1999) 1087-1096).

Secretion of pro-carboxypeptidase B directs the recombinant protein to the extracytoplasmic space from where it diffuses into the growth medium. Thus, in a

- 19 -

preferred embodiment of the invention the methylotrophic yeast is grown in liquid culture and secretes pro-carboxypeptidase B into the liquid growth medium, i.e. the liquid culture medium. This allows a very efficient separation of yeast biomass from the recombinant protein using, e.g. expanded bed chromatography techniques. As a result, pro-carboxypeptidase B purified from the yeast source organism can be freed efficiently from other not desired enzyme activities.

Regarding the vector with which, in a preferred embodiment, the methylotrophic yeast strain is transformed, the person skilled in the art is well aware of expression vectors that allow the construction of fusion polypeptides containing a so called "Histidine tag", "(poly)Histidine tag" or "Histidine-tag". In the preferred pre-protein of present invention, a Histidine-tag is fused to the N-terminus of the pro-carboxypeptidase B polypeptide. The Histidine-tag in total comprises six consecutive Histidine residues. The skilled artisan will note in this regard that the N-terminal amino acid of the pro-carboxypeptidase B polypeptide is a Histidine, too. Thus, in the annotated nucleotide sequence encoding the pre-protein there is an overlap of the fused coding sequences in that the codon for the last Histidine of the Histidine-tag also represents the first Histidine of the pro-carboxypeptidase B moiety.

After activation, i.e. the tryptic cleavage of the immobilized Histidine-tagged pro-carboxypeptidase B, the Histidine-tagged propeptide moiety remains bound to the particulate metal chelate affinity matrix. The activated carboxypeptidase B is released and separated in step (f) from the particulate metal chelate affinity matrix. Thus, separating particulate metal chelate affinity matrix from the liquid phase at the same time separates the propeptide from the carboxypeptidase B enzyme. Using this method non-covalent attachment of the propeptide to the carboxypeptidase B molecule is avoided.

It is preferred that the vector is comprising a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag (Histidine-tag) and a signal peptide, whereby optionally between the Histidine-tag and the signal peptide a spacer sequence may be inserted. It is more preferred that the amino acid sequence of the rat pro-carboxypeptidase B is the amino acid sequence from position 14 to position 415 in SEQ ID NO: 3. It is also preferred that the signal peptide contains a signal peptidase cleavage site which

- 20 -

is located adjacent to the Histidine-tag or adjacent to the spacer sequence. It is very much preferred that the amino acid sequence of the expressed pre-protein is the amino acid sequence in SEQ ID NO: 4. It is even more preferred that the nucleotide sequence encoding rat pro-carboxypeptidase B is the nucleotide sequence from position 286 to position 1,497 in SEQ ID NO: 3. It is even more preferred that the nucleotide sequence encoding the pre-protein is the nucleotide sequence in SEQ ID NO: 3. It is even more preferred that the nucleotide sequence encoding the pre-protein is operably linked to a promoter or promoter element.

To enable transcription of the nucleotide sequence encoding the pre-protein it is preferred that the nucleotide sequence encoding the pre-protein is operably linked to a promoter or promoter element. Very much preferred is a promoter or promoter element from *Pichia pastoris*, even more preferred is the *Pichia pastoris* AOX1 promoter given in SEQ ID NO: 5. It is also preferred that in addition the nucleotide sequence that encodes the pre-protein is operably linked with a terminator sequence that directs termination of transcription in the methylotrophic yeast strain. Very preferred is a terminator from *Pichia pastoris*, even more preferred is the *Pichia pastoris* AOX1 terminator.

It is further preferred that the vector is a plasmid capable of being replicated as an episome in the methylotrophic yeast strain. Thus, the preferred plasmid is a circular nucleic acid molecule that comprises an origin of replication directing replication of the episome in the methylotrophic yeast strain. Moreover, the plasmid comprises a selectable marker that is expressed in the methylotrophic yeast strain, whereby the selectable marker allows to select for the presence of the plasmid in the methylotrophic yeast strain. A very much preferred selectable marker is a ZeocinTM resistance gene, that is the native form or a genetically engineered variant of the *Sh ble* gene from *Streptoalloteichus hindustanus* (Drocourt, D., et al., *Nucleic Acids Res.* 18 (1990) 4009; Carmels, T., et al., *Curr. Genet.* 20 (1991) 309-314). Another very much preferred selectable marker confers resistance against aminoglycoside antibiotics such as Hygromycin and G418 (Southern, P.J., and Berg, P., *J. Mol. Appl. Genet.* 1 (1982) 327-341). An example for such a selectable marker is an aminoglycoside phosphotransferase gene.

It is further preferred that an artificial chromosome capable of being replicated in the methylotrophic yeast strain contains the vector. Thus, the preferred artificial

chromosome is a linear nucleic acid molecule that comprises at least one origin of replication, a centromere and terminal telomeres, thereby controlling replication, integrity and mitotic/meiotic distribution of the artificial chromosome in the methylotrophic yeast strain. Moreover, the vector that is contained in the artificial chromosome comprises a selectable marker that is expressed in the methylotrophic yeast strain and that allows to select for the presence of the vector in the artificial chromosome that is replicated in the methylotrophic yeast strain. A very much preferred selectable marker is a Zeocin[™] resistance gene, that is the native form or an artificial variant of the Sh ble gene from *Streptoalloteichus hindustanus*. Another very much preferred selectable marker confers resistance against aminoglycoside antibiotics such as Hygromycin and G418. An example for such a selectable marker is an aminoglycoside phosphotransferase gene.

It is even more preferred that a chromosome of the methylotrophic yeast strain contains the vector. It is very much preferred that the vector has a nucleotide sequence identical to a chromosomal sequence, thus allowing integration of the vector into the host chromosome by site-specific recombination. To this end, the *Pichia pastoris* AOX1 locus is even more preferred as a locus for integration the host chromosome by site-specific recombination. It is also very much preferred that, the vector comprises a selectable marker that is expressed in the methylotrophic yeast strain and that allows to select for the presence of the vector in the methylotrophic yeast strain. A very much preferred selectable marker is a Zeocin[™] resistance gene, that is the native form or an artificial variant of the Sh ble gene from *Streptoalloteichus hindustanus*. Another very much preferred selectable marker confers resistance against aminoglycoside antibiotics such as Hygromycin and G418. An example for such a selectable marker is an aminoglycoside phosphotransferase gene.

The person skilled in the art is aware of the fact that the yield of secreted Histidine-tagged pro-carboxypeptidase B obtainable from growth medium, such as liquid growth medium, can be increased when the number of copies of the nucleotide sequence encoding the pre-protein is increased. Thus, the yield of secreted Histidine-tagged pro-carboxypeptidase B molecule obtainable from growth medium can be increased when number of copies of the vector in the genome of the methylotrophic yeast strain is increased. For example, the copy number of the vector can be increased by subjecting the methylotrophic yeast strain to repeated

- 22 -

transformations of the vector and repeated selection rounds using increasing concentrations of the selective agent against which the selective marker comprised in the vector confers resistance (US 5,324,639; Thill, G.P., et al., Positive and Negative Effects of Multi-Copy Integrated Expression in *Pichia pastoris*, International Symposium on the Genetics of Microorganisms 2 (1990), pp. 477-490; Vedvick, T., et al., J. Ind. Microbiol. 7 (1991) 197-201).

The person skilled in the art is also aware of the fact that repeated transformations can be carried out using more than one vector. For example, repeated transformations can be carried out using a first and a second vector, whereby the first and the second vector encode the same pre-protein, whereby in the first and in the second vector the nucleotide sequence encoding the pre-protein is operably linked to a promoter or promoter element, whereby the same pre-protein is expressed and Histidine-tagged pro-carboxypeptidase B is secreted, and whereby the first and the second vector confer resistance to a first and a second selection marker.

An example for a first selective marker is the *Sh ble* gene, that is the ZeocinTM resistance gene (Drocourt, D., et al., Nucleic Acids Res. 18 (1990) 4009; Carmels, T., et al., Curr. Genet. 20 (1991) 309-314). The protein encoded by the *Sh ble* gene binds ZeocinTM stoichiometrically and with a strong affinity. The binding of ZeocinTM inhibits its toxic activity thereby selecting for transformants containing the *Sh ble* gene. It is known to a person skilled in the art that increasing the concentration of ZeocinTM as the selective agent in the medium selects for an increase in the number of copies of the vector expressing the *Sh ble* gene. It is therefore advantageous to use a vector with the *Sh ble* gene as a selectable marker to generate by repeated transformation multiple transformants of the methylotrophic yeast strain containing multiple copies of the vector. It is furthermore advantageous that transformations are repeated and selection for even more resistant transformants is repeated until for the transformed methylotrophic yeast strain no further increase of the level of resistance to ZeocinTM is obtained anymore or no further increase of the ZeocinTM concentration in the selection medium is possible anymore.

In case a first and a second vector are used, an example for a second selection marker is resistance against aminoglycoside antibiotics (Southern, P.J., and Berg, P.,

- 23 -

J. Mol. Appl. Genet. 1 (1982) 327-341) such as G418. Thus, an exemplarily second vector expresses a resistance gene that confers resistance against G418. For example, there are several aminoglycoside phosphotransferases known to the art that confer resistance to aminoglycoside antibiotics (van Treeck, U., et al., Antimicrob Agents Chemother. 19 (1981) 371-380; Beck, E., et al., Gene 19 (1982) 327-336). The aminoglycoside phosphotransferase I (APH-I) enzyme has the ability to inactivate the antibiotic G418 and is an established selectable marker in yeast (Chen, X.J., and Fukuhara, H., Gene, Vol. 69 (1988) 181-192).

Thus, for the purpose of further increasing the dosage of the nucleotide sequence encoding the pre-protein, the second vector is advantageously used for further rounds of transformation and selection, whereby in this case a preferred selective agent is G418 and whereby for transformation the methylotrophic yeast strain transformed with the first vector is used.

It is further preferred the methylotrophic yeast strain is a *Hansenula*, *Pichia*, *Candida* or *Torulopsis* species. In a very much preferred embodiment of the invention, the methylotrophic yeast strain is selected from the group consisting of *Pichia pastoris*, *Hansenula polymorpha*, *Candida boidinii* and *Torulopsis glabrata*.

Even more preferred *Pichia pastoris* strains are deposited at the American Type Culture Collection (ATCC) with the accession numbers 201178, 201949, 204162, 204163, 204164, 204165, 204414, 204415, 204416, 204417, 20864, 28485, 34614, 60372, 66390, 66391, 66392, 66393, 66394, 66395, 76273, 76274, and 90925.

Yet, an even more preferred methylotrophic yeast strain is the *Pichia pastoris* strain with the American Type Culture Collection accession number 76273 or a derivative thereof.

Even more preferred *Hansenula polymorpha* strains are deposited at the American Type Culture Collection with the accession numbers 14754, 200499, 200500, 200501, 200502, 200503, 200504, 200505, 200506, 200507, 200508, 200509, 200510, 200511, 200512, 200513, 200838, 200839, 201322, 204205, 22023, 26012, 34438, 36669, 38626, 44954, 44955, 46059, 48180, 58401, 62809, 64209, 66057, 76722, 76723, 76760, 90438, 96694, 96695, MYA-335, MYA-336, MYA-337, MYA-338, MYA-339, and MYA-340.

- 24 -

Even more preferred *Candida boidinii* strains are deposited at the American Type Culture Collection with the accession numbers 18810, 201209, 20432, 26175, 32195, 32929, 36351, 38256, 38257, 44637, 46498, 48180, 56294, 56507, 56897, 60364, 62807, 90439, 90441, 96315, and 96926.

5 Even more preferred *Torulopsis glabrata* strains are deposited at the American Type Culture Collection with the accession numbers 15126, 15545, 2001, 22019, 26512, 28226, 28290, 32312, 32554, 32936, 34147, 34449, 36909, 38326, 4135, 46433, 48435, 58561, 66032, 750, and 90030.

10 Another embodiment of the invention is a transformed *Pichia pastoris* strain with a chromosome that contains a vector comprising a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag and a signal peptide, operably linked with the *Pichia pastoris* AOX1 promoter according to SEQ ID NO: 5 or a promoter element thereof, whereby the nucleotide sequence that encodes the pre-protein is the
15 nucleotide sequence of SEQ ID NO: 3. It is also preferred that the vector comprises a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag (Histidine-tag) and a signal peptide, whereby optionally between the Histidine-tag and the signal peptide a spacer sequence may be inserted.

20 Yet another embodiment of the invention is a protein with carboxypeptidase B activity which is substantially free of carboxypeptidase B propeptide, obtainable by the method according to the invention. The term "substantially free" denotes that the propeptide is below the detection level using mass spectroscopic detection means. An exemplary analysis is described in Example 8.

25 As indicated in Example 8, the purified protein with carboxypeptidase B activity predominantly lacks the C-terminal Tyrosine (Tyr497 in SEQ ID NO:4). Only a very small fraction of the purified protein still contains this amino acid. Thus, according to the invention the protein with carboxypeptidase B activity which is substantially free of carboxypeptidase B propeptide has the amino acid sequence of
30 SEQ ID NO:4 from position 191 to position 496.

Example 9 illustrates that after the first purification step according to Example 7 the collected pools contain product with almost equal amounts of product with and without C-terminal Tyrosine. During the process the terminal Tyrosine residues are progressively removed. After the second chromatographic purification step the predominant protein species lacks the C-terminal Tyrosine almost completely. A possible but not tested explanation therefor is that the host organism used for expressing and secreting the product produces a protein with carboxypeptidase Y activity. Carboxypeptidase Y is known from *Saccharomyces cerevisiae* to be capable of cleaving off C-terminal amino acids with a broad specificity. Although carboxypeptidase Y is known to be a vacuolar enzyme (Kato, M. et al. Eur. J. Biochem 270 (2003) 4587-4593) lysed yeast cells may give rise to noticeable carboxypeptidase Y activity in the fermentation broth. However, since removal of the C-terminal Tyrosine leaves the overall enzymatic properties of the protein with carboxypeptidase B activity unchanged and since no further amino acids are removed from the C-terminus, inactivation or additional separation of the hypothetical carboxypeptidase Y activity is not deemed to be necessary.

Yet another embodiment of the invention is the use of a protein with carboxypeptidase B activity which is substantially free of carboxypeptidase B propeptide according to the invention for proteolytic cleavage of a peptide bond. It is preferred that cleavage of a peptide bond is effected by catalyzing hydrolysis of the basic amino acids, Lysine, Arginine, and Ornithine from the C-terminal position in a polypeptide. It is further preferred that the amino acid sequence of the protein with carboxypeptidase B activity is the amino acid sequence from position 191 to position 496 of SEQ ID NO: 4.

Very much preferred is the use of a protein with carboxypeptidase B activity which is substantially free of carboxypeptidase B propeptide according to the invention characterized in that a peptide bond of a precursor of insulin is cleaved. Processes therefor have been described, e.g., in EP 0 264 250 and EP 0 195 691. Accordingly, a precursor of insulin is proteolytically cleaved by trypsin and a protein with carboxypeptidase B activity. It is even more preferred that the protein with carboxypeptidase B activity catalyzes the hydrolytic cleavage of the basic amino acids, Lysine, Arginine, and Ornithine, from a C-terminal position of the tryptic proteolysis product of the insulin precursor.

- 26 -

Yet another embodiment of the invention is a reagent solution containing a protein with carboxypeptidase B activity which is substantially free of carboxypeptidase B propeptide, according to the invention. It is preferred that the protein with carboxypeptidase B activity in the reagent solution is capable of catalyzing the hydrolytic cleavage of the basic amino acids, Lysine, Arginine, and Ornithine, from a C-terminal position of a peptide or a polypeptide. It is even more preferred that the amino acid sequence of the protein with carboxypeptidase B activity in the reagent solution is the amino acid sequence from position 191 to position 496 of SEQ ID NO: 4. A reagent solution containing a protein with carboxypeptidase B activity usually is an aqueous solution which in addition contains buffer salts. However, further ingredients are possible and are well known to the skilled artisan. An example for a further ingredient is trypsin.

The following examples, references, sequence listings and figures are provided to aid the understanding of the present invention, the true scope of which is set forth in the appended claims. It is understood that modifications can be made in the procedures set forth without departing from the spirit of the invention.

Description of the Figures

Figure 1 Graphic representation of the vector pCPB-1. The vector contains a nucleotide sequence encoding a pre-protein, that is a fusion protein comprising the α -factor signal peptide from *Saccharomyces cerevisiae* (denoted "alpha-Faktor-SS") and the Histidine-tagged pro-carboxypeptidase B. The nucleotide sequence encoding a pre-protein is under transcriptional control of the AOX1 promoter (denoted "AOX1-Promotor") and the AOX1 terminator (denoted "AOX1-TT"). The vector confers resistance against Zeocin®.

Figure 2a Diagram depicting the result of an exemplary mass spectrometry analysis of a carboxypeptidase B preparation according to Example 8. The x-axis indicates m/z values (m = ion mass; z = ion charge). The asterisks denote peaks which correspond to the respective carboxypeptidase B dimers.

- Figure 2b** Diagram depicting the result of an exemplary mass spectrometry analysis of a carboxypeptidase B preparation according to Example 8. The x-axis indicates ion molecular weight in [Da].
- 5 **Figure 3** Section (a) shows the hypothetical m/z values and peaks which correspond to the carboxypeptidase B propeptide, also referred to as the 12,132 Da peptide. The section denoted (b) shows the corresponding clipping of the spectrum according to Figure 1a. The x-axes of sections (a) and (b) are aligned and have identical scaling.
- 10 **Figure 4** Section (a) shows the hypothetical m/z values and peaks which correspond to the carboxypeptidase B propeptide, also referred to as the 10,746 Da peptide. The section denoted (b) shows the corresponding clipping of the spectrum according to Figure 1a. The x-axes of sections (a) and (b) are aligned and have identical scaling.
- 15
- Figure 5** Clipping of a spectrum obtained with the enzyme preparation according to Example 8.
- Figure 6** Diagram depicting the result of mass spectrometry analysis of a carboxypeptidase B preparation. Analysis was performed on a sample taken after the expanded bed chromatography, as explained in Example 9. The x-axis of section (a) indicates m/z values (m = ion mass; z = ion charge). The x-axis of section (b) indicates ion molecular weight in [Da].
- 20
- Figure 7** Diagram depicting the result of mass spectrometry analysis of a carboxypeptidase B preparation. Analysis was performed on a sample taken after the Q-Sepharose ff^{TM} , as explained in Example 9. The x-axis of section (a) indicates m/z values (m = ion mass; z = ion charge). The x-axis of section (b) indicates ion molecular weight in [Da].
- 25

Figure 8 Diagram depicting the result of mass spectrometry analysis of a carboxypeptidase B preparation. Analysis was performed on the final product of the purification procedure, as explained in Example 9. The x-axis of section (a) indicates m/z values (m = ion mass; z = ion charge). The x-axis of section (b) indicates ion molecular weight in [Da].

Example 1

Synthesis of the gene encoding the pre-protein

DNA techniques were performed according to standard procedures (Sambrook, Fritsch & Maniatis, Molecular Cloning, A Laboratory Manual, 3rd edition, CSHL Press, 2001). Reagents for molecular biological work were used according to the recommendations of the suppliers.

A nucleotide sequence encoding an artificial pre-pro-carboxypeptidase B gene according to SEQ ID NO: 3 was synthesized de-novo. Portions of the nucleotide sequence were synthesized as 24 single-stranded DNA oligonucleotides with a length of between 54 and 90 nucleotides. The single-stranded oligonucleotides represented an alternating fashion overlapping portions of the leading strand and the lagging strand of SEQ ID NO: 3. Each oligonucleotide was designed such that the 5' and 3' ends overlapped with the neighbouring oligonucleotide. As an exception, the oligonucleotides representing the 5' and the 3' terminus of SEQ ID NO: 3 only overlapped with the 3' and 5' ends of the neighbouring fragments, respectively. The sequences of the overlaps were chosen such that during an annealing reaction unspecific annealing was avoided. The oligonucleotides representing the 5' and the 3' terminus of SEQ ID NO: 3 additionally contained linker sequences with recognition sites for restriction endonucleases, in order to facilitate further steps of molecular cloning, such as the insertion of the artificial coding sequence into expression vectors. A preferred restriction site for the oligonucleotides representing the 5' terminus of SEQ ID NO: 3 was XhoI. A preferred restriction site for the oligonucleotides representing the 3' terminus of SEQ ID NO: 3 was NotI.

A nucleic acid molecule comprising the nucleotide sequence according SEQ ID NO: 3 was synthesized stepwise by way of PCR (polymerase chain reaction). In principle,

- 29 -

the first two oligonucleotides representing adjacent and partially overlapping portions of the leading and the lagging strand were added to a PCR reaction mixture. Following several cycles of PCR, a contiguous double-stranded fragment representing both portions of the leading and the lagging strand was obtained. The double-stranded fragment was optionally purified, mixed with the consecutive adjacent and partially overlapping single-stranded oligonucleotide and was subjected to a further round of PCR.

Using the procedure described above, three larger fragments of the nucleotide sequence according SEQ ID NO: 3 were synthesized independently. Each fragment could be present in a mixture of by-products. Therefore, the fragments were electrophoresed in an agarose gel and identified by their size. Agarose blocks containing the desired fragments were excised and the DNA fragments were isolated using the QiaQuick Gel Extraction Kit (Qiagen)*. Other extraction methods are also possible.

Of the three fragments, the second fragment at its 5' end had a sequence overlap with the 3' end of the first fragment, and whereby the second fragment at its 3' end had a sequence overlap with the 5' end of the third fragment. The full-length sequence according to SEQ ID NO: 3 was again synthesized in a stepwise fashion. The three fragments were united in a PCR reaction mixture. The annealing temperatures were chosen taking into account the overlapping sequence with the lowest melting point. Five cycles of PCR were performed, followed by the addition of an additional pair of primers complementary to the 5' end and the 3' end of the desired full-length product. Using an annealing temperature chosen with respect to the annealing temperature of the newly added primer with the lower melting temperature, further 25 cycles of PCR were performed and the full-length fragment comprising the nucleotide sequence of SEQ ID NO: 3 was obtained. The full-length fragment (i.e. the full-length pre-protein-encoding DNA including linker sequences with restriction endonuclease cleavage sites) was inserted into a vector and propagated in a transformed host organism. A preferred host organism for propagation purposes was E. coli.

The nucleotide sequence of the full-length DNA comprising the sequence encoding the pre-protein of SEQ ID NO: 4 was verified by sequencing. The final full-length product was amplified using PCR.

*Trade-mark

Example 2**Construction of vectors, transformation, expression**

For further steps regarding the construction of expression vectors, transformation, expression of the pre-protein and secretion of Histidine-tagged pro-carboxypeptidase B into growth media, the methods suggested and described in the
 5 Invitrogen manuals "Pichia Expression Kit" Version M 011102 25-0043, "pPICZ A, B, and C" Version D 110801 25-0148, "pPICZ α A, B, and C" Version E 010302 25-0150, and "pPIC9K" Version E 030402 25-0106 were applied. Reference is also made to further vectors, yeast strains and media mentioned therein. Basic methods
 10 of molecular biology were applied as described in Sambrook, Fritsch & Maniatis, Molecular Cloning, A Laboratory Manual, 3rd edition, CSHL Press, 2001.

As a result, several vectors were obtained whereby each comprised an expression cassette containing capable of expressing the pre-protein according to SEQ ID NO: 4 in methylotrophic yeast strains. A preferred yeast strain was a Pichia pastoris
 15 strain. Each vector in addition comprised, among other genetic elements, an origin of replication capable of replicating the vector autonomously in methylotrophic yeast cells. In addition the vectors were designed such as to enable integration into the genome of the host cell. Another genetic element on each vector was a further expression cassette expressing a selectable marker.

20 Repeated transformations of the Pichia pastoris host strain were performed in order to increase the yield of secreted Histidine-tagged pro-carboxypeptidase B.

Example 3**Protein quantification**

Quantification of purified rat carboxypeptidase B was performed by way of
 25 measuring the extinction at 280 nm using cuvettes having a width of 1 cm. The concentration was determined according to the following formula:

$$\text{Protein [mg/ml]} = \frac{10 \text{ [mg/ml]} * \Delta E_{\text{Probe}} * \text{dilution factor}}{21.4}$$

Example 4**Protein activity**

5 The activity of carboxypeptidase B was determined using 1.5 M Hippuryl-L-Arginine (Bachem G-2265) in 100 mM Tris pH 7.8 at 25°C and cuvettes having a width of 0.5 cm.

Example 5**HPLC analytics**

10 Activation of carboxypeptidase B from Histidine-tagged pro-carboxypeptidase B was monitored using HPLC (high pressure liquid chromatography) and gel filtration separation means (Superdex 75 HR 10/30, Amersham Biosciences). The chromatography buffer was 0.1 M Tris pH 7.5, 0.3 M NaCl. The flux was 0.5 ml/min.

Retention times of carboxypeptidase B and Histidine-tagged pro-carboxypeptidase B differed unambiguously (1.2 ml) facilitating their recognition on elution profiles.

15 **Example 6**

Fermentation

20 General methods were followed according to descriptions in the literature (Higgins, D.R., Cregg, J.M. (1998) Pichia Protocols. In: Methods in Molecular Biology Vol. 103, the whole document but in particular pp 107-120; Stratton, J., Chiruvolu, V., Meagher, M. (1998) High cell-density fermentation. In: Pichia Fermentation Guidelines, Invitrogen).

25 Pre-cultures of the transformed Pichia pastoris strain capable of expressing the pre-protein of SEQ ID NO: 4 were made at 30°C in yeast nitrogen base medium without amino acids (DIFCO). The pre-culture was used to inoculate a fermentation medium. The fermentation medium contained minerals, that is to say H₃PO₄, CaSO₄ x 2 H₂O, K₂SO₄, MgSO₄ x 7 H₂O, KOH, NaOH, the trace elements CuSO₄ x 5 H₂O, KJ, MnSO₄ x H₂O, Na₂MoO₄ x 2 H₂O, H₃BO₄, ZnCl₂, CoCl₂ x 6 H₂O, and FeSO₄ x 7 H₂O. The fermentation medium also contained biotin and glycerol. The

- 32 -

pH was adjusted to pH 5.0 and maintained at that value using NH_3 which at the same time served as a source for Nitrogen.

5 When glycerol being the first carbon source was exhausted, methanol was added, thereby inducing expression of the pre-protein that was driven by the AOX1 promoter.

Example 7

Purification

10 The fraction of the liquid mass of the fermentation broth was adjusted to 5-13% by adding a buffer containing 20 mM Tris pH 7.5, 1M NaCl and 1 mM imidazole. The diluted fermentation broth was passed through a column with a particulate metal chelate affinity matrix . The preferred particulate metal chelate affinity matrix was Zn^{2+} -loaded Streamline™ Chelating (Amersham Biosciences). Loading the particulate metal chelate affinity matrix in the column with Histidine-tagged pro-carboxypeptidase B was done using the expanded bed chromatography technique.

15 Under the conditions as described above, the Histidine-tagged pro-carboxypeptidase B in the diluted fermentation broth binds, i.e. adsorbs to the particulate metal chelate affinity matrix .

Following the binding step, the particulate metal chelate affinity matrix in the (still expanded) column was washed with a washing buffer containing 20 mM Tris pH 7.5, 1 M NaCl, and 1 mM Imidazole. Several (at least two) washing steps were performed under conditions allowing Histidine-tagged pro-carboxypeptidase B to remain bound to the particulate metal chelate affinity matrix . The first washing step was performed using the washing buffer additionally containing 11% [volume by volume] of glycerol. Any additional washing step was performed using the

20

25 washing buffer without glycerol.

Activation of carboxypeptidase B was effected by way of on-column cleavage with trypsin. Following the washing step, a buffer (20 mM Tris pH 7.5, 1M NaCl and 1 mM imidazole) containing trypsin (13-20 U/ml) was pumped through the column, whereby the buffer that had left the column was re-cycled into the column.

30 After 150-300 min the buffer now additionally containing activated carboxypeptidase B was removed. The column was washed (two to three washing

- 33 -

steps) in upward flow mode using washing buffer, whereby the flow-through was collected. The buffer now additionally containing activated carboxypeptidase B and the flow-through from the washing steps were pooled. Tryptic activity was stopped by adding Benzamidinium hydrochloride to a final concentration of 1 mM.

5 The buffer containing activated carboxypeptidase B and trypsin was filtered to remove last traces of residual biomass. The filtered buffer was subjected to diafiltration by tangential flow filtration. A dialysis buffer was used which contained Tris pH 8.3 at a concentration of 60-100 mM, and further containing 1 mM Benzamidinium hydrochloride, 0.1 mM ZnCl₂. Diafiltration resulted in an
10 exchange of the buffer in which the activated carboxypeptidase was present.

The subsequent step was Q-Sepharose ffTM chromatography. The buffer in which the activated carboxypeptidase was present was loaded on a column containing Q-Sepharose ffTM. A washing step was applied using 60-100 mM Tris pH 8.3, 1 mM Benzamidinium hydrochloride, 0.1 mM ZnCl₂. Following the washing step, the
15 column was eluted by applying a gradient extending to 125 mM NaCl. Fractions were collected. Fractions containing carboxypeptidase B were pooled.

Example 8

Mass spectrometry of purified carboxypeptidase B

Carboxypeptidase B was purified according to Example 7. Prior to mass
20 spectrometry the enzyme preparation was desalted using a micro HPLC device (ABI-120A) equipped with an autosampler (Gilson 234). The column used was a Vydac Protein C4 cartridge (Vydac catalogue no. 214GD51) together with the appropriate guard holder. An aliquot of the enzyme preparation containing 50 µg - 100 µg of protein was loaded on the column and eluted using a gradient of elution
25 buffer A (3.5% "pro analysi" grade HCOOH [formic acid] in HPLC grade water [Baker]) and buffer B (80% HPLC grade acetonitrile, 5% HPLC grade HCOOH in HPLC grade water [Baker]). The gradient was applied by eluting for the first 5 min with 5% buffer B, 95% Buffer A and then for 3 more min with 100% buffer B. During HPLC the temperature of the column was kept at 35°C. The enzyme was
30 detected at a wave length of 280 nm. It was found that the enzyme eluted between t_{5 min} and t_{8 min}. The protein peak was collected (from 30 mAU – 40 mAU; peak maximum >800 mAU).

Mass spectrometry was performed on a Q-ToF 2™ (Waters Micromass, Manchester, UK) equipped with a nanospray interface. The device was capable of selecting peptide or protein ions with an m/z of 200 – 2000 and protein ions with an m/z of 800 – 2000. Spray capillaries were from Proxeon (“Medium”, catalogue no. ES 387).
5 Measurements were taken using variable parameters with respect to capillary voltage, cone voltage, and MS profiles. Parameters were adjusted for each sample to be analyzed regarding the measurement range. For analysis the software MassLynx 4.0™ with the MaxEnt 1™ package (Waters) was used. Using Maximum Entropy processing applied to mass spectrometry data MaxEnt 1™ aids the
10 enhancement of complex spectra by allowing the deconvolution of overlapping multiply-charged spectra produced by the analysis of protein mixtures. Indicated masses below 10,000 Da are monoisotopic, masses indicated as equal to or above 10,000 Da are averaged masses.

By way of mass spectrometry it was found that the enzyme preparation, i.e. the
15 purified carboxypeptidase B gives rise to essentially three different mass peaks: (1) the main peak [designated A in Figure 2b] corresponding to a (averaged) mass of 35,015 Da indicating carboxypeptidase B lacking the C-terminal Tyrosine; (2) a minor peak corresponding to a mass of about 70,026 Da was also found (not shown) and indicates the dimer of carboxypeptidase B lacking the C-terminal
20 Tyrosine; (3) a minor peak corresponding to a mass of about 35,176 Da indicates the carboxypeptidase B including the C-terminal Tyrosine.

Theoretically, peaks corresponding to the Histidine-tagged propeptide should reflect the 12,132 Da peptide of 105 amino acids generated from the amino acid sequence of of SEQ ID NO: 4 by (a) the KEX-2 signal peptidase cleaving between
25 Arg85 and Ser86 and (b) trypsin cleaving between Arg190 and Ala191. Also, fragments of the Histidine-tagged propeptide can be postulated. These include the 10,746 Da peptide comprising the sequence from Ser86 to Arg178 and the 1,404 Da peptide comprising the sequence from Asn179 to Arg190. The “(a)” sections of Figures 3 and 4 indicate the hypothetical peaks that one would expect if the
30 12,132 Da and 10,746 Da peptides were present in the enzyme preparation. The “(b)” sections represent enlarged clippings of spectra like the one shown in Figure 2a. It can be seen that at the positions of the hypothetical peaks nothing is detected in the spectra of the enzyme preparation. In addition, Figure 5 shows a clipping of a spectrum obtained with the enzyme preparation. The hypothetical

- 35 -

1,404 Da peptide would be expected to give rise to peaks at $m/z = 702.8$ ($[M+2H]^{2+}$) and $m/z = 468.9$ ($[M+3H]^{3+}$). However, no conspicuous peaks indicating the presence of the propeptide are present.

5 Furthermore, the N-terminal amino acid sequence of the polypeptide corresponding to the main peak was determined. The 15 N-terminal amino acids were found to be "ASGHSYTKYNNWETI". These are identical with the amino acids 191 to 205 of SEQ ID NO: 4, i.e. they represent the N-terminus of the activated carboxypeptidase B enzyme.

Example 9

10 Removal of C-terminal Tyrosine

Carboxypeptidase B was purified according to Example 7. A sample was taken from the pooled carboxypeptidase B containing fractions and washing solutions. Mass spectrometry was performed according to Example 8. Results are shown in Figure 6. Figure 7 illustrates the results of a sample after the Q-sepharose ff^{TM} chromatography and Figure 8 illustrates the endproduct. It becomes clear that in 15 the beginning about 50% or more of carboxypeptidase B contains the C-terminal Tyrosine. After the second chromatography the predominant protein species lacks the C-terminal Tyrosine already. However, C-terminal degradation was restricted to the C-terminal Tyrosine.

20

List of References

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (eds), Molecular Biology of the Cell, fourth edition, 2002, Garland Science Publishing
- Beck, E., et al., *Gene* 19 (1982) 327-336
- 5 Borsing, L. et al. *Biochem. Biophys. Res. Commun.* 240 (1997) 586-590
- Carmels, T., et al., *Curr. Genet.* 20 (1991) 309-314
- Chaga, G. et al., *Biotechnol. Appl. Biochem.* 29 (1999) 19-24
- Chaga, G. et al., *J. Chromatogr. A* 864 (1999) 257-256
- Chen, B.P. and Hai, T., *Gene* 139(1994) 73-75
- 10 Chen, X.J., and Fukuhara, H., *Gene*, Vol. 69 (1988) 181-192
- Clauser E. et al., *J. Biol. Chem.* 1988, Vol. 263, 17837 - 45
- DE 19 915 938
- Drocourt, D., et al., *Nucleic Acids Res.* 18 (1990) 4009
- EP 0 116 201
- 15 EP 0 195 691
- EP 0 253 303
- EP 0 264 250
- EP 0 282 042
- EP 1 069 131
- 20 Folk J. Carboxypeptidase B, in: *The Enzymes* 3, P. Boyer, Academic Press, NY, 57, 1971
- Folk, J., and Gladner, J.: *Biochim Biophys Acta* (1961) 48, 139-47
- Hakansson, K. et al. *Acta Crystallogr. D Biol. Crystallogr.* 56 (2000) 924-926
- Halliwell, C.M. et al., *Anal Biochem.* 295 (2001) 257-261
- 25 Higgins, D.R., Cregg, J.M. (1998) *Pichia Protocols*. In: *Methods in Molecular Biology* Vol. 103, the whole document but in particular pp 107-120
- Hochuli, E. et al., *Bio/Technology* 6 (1988) 1321-1325
- Hochuli, E. et al., *J. Chromatogr.* 411 (1987) 177-184
- Janknecht, R. and Nordheim, A., *Gene* 121 (1992) 321-324
- 30 Janknecht, R. et al. *Proc. Natl. Acad. Sci. USA* 88 (1991) 8972-8976
- Julius, D., et al., *Cell* 37 (1984) 1075-1089
- Kaslow, D.C. and Shiloach, J., *Bio/Technology* 12 (1994) 494-499
- Kato, M. et al. *Eur. J. Biochem* 270 (2003) 4587-4593
- Kuusinen, A. et al., *Eur. J. Biochem.* 233 (1995) 720-726

- 37 -

Pharmacia Biotech manual "Expanded bed adsorption, principles and methods".
ISBN 91-630-5519-8

Porath, J. et al., *Nature* 258 (1975) 598–599

Rank, K.B. et al., *Protein Expr. Purif.* 22 (2001) 258–266

5 Sambrook, Fritsch & Maniatis, *Molecular Cloning, A Laboratory Manual*, 3rd
edition, CSHL Press, 2001

Schmidt, M. et al. *Protein Expr. Purif.* 12 (1998) 323–330

Southern, P.J., and Berg, P., *J. Mol. Appl. Genet.* 1 (1982) 327-341

Stratton, J., Chiruvolu, V., Meagher, M. (1998) High cell-density fermentation. In:
10 *Pichia Fermentation Guidelines*, Invitrogen

Thill, G.P., et al., Positive and Negative Effects of Multi-Copy Integrated Expression
in *Pichia pastoris*, *International Symposium on the Genetics of*
Microorganisms 2 (1990), pp. 477-490

US 4,683,293

15 US 4,808,537

US 4,812,405

US 4,818,700

US 4,837,148

US 4,855,231

20 US 4,857,467

US 4,870,008

US 4,879,231

US 4,882,279

US 4,885,242

25 US 4,895,800

US 4,929,555

US 5,002,876

US 5,004,688

US 5,032,516

30 US 5,122,465

US 5,135,868

US 5,166,329

US 5,284,933

US 5,310,663

35 US 5,324,639

US 5,618,676

- 38 -

- US 5,854,018
US 5,856,123
US 5,919,651
van Treeck, U., et al., Antimicrob Agents Chemother. 19 (1981) 371-380
5 Vedvick, T., et al., J. Ind. Microbiol. 7 (1991) 197-201
Ventura et al. J. Biol. Chem. (1999) 274(28), 19925-33
Waters et al., J. Biol. Chem. 263 (1988) 6209-14
Werten, M.W., et al., Yeast 15 (1999) 1087-1096
WO 00/56903
10 WO 01/51624
WO 96/23064
Wu, J. and Filutowicz, M., Acta Biochim. Pol. 46 (1999) 591-599
Zisapel, N. and Sokolovsky, M, Eur J Biochem (1975) 54, 541-7

- 39 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: F. Hoffmann-La Roche AG
(B) STREET: Grenzacherstrasse 124
(C) CITY: CH-4070 Basel
(E) COUNTRY: SWITZERLAND

(ii) TITLE OF INVENTION: Recombinantly expressed carboxypeptidase
B and purification thereof

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Borden Ladner Gervais LLP
(B) STREET: 100 Queen Street
(C) CITY: Ottawa
(D) PROVINCE: Ontario
(E) COUNTRY: Canada
(F) POSTAL CODE: K1P 1J9

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy Disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn version 3.2

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: Unknown
(B) FILING DATE: 2004-12-02

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: EP 03028101.8
(B) FILING DATE: 2003-12-05

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: David Conn
(B) REGISTRATION NUMBER: 3960
(C) REFERENCE/DOCKET NUMBER: PAT 58097-1

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (613) 237-5160
(B) TELEFAX: (613) 787-3558

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1248 bases
(B) TYPE: nucleic acid

ATG	TTG	CTG	CTA	CTG	GCC	CTG	GTG	AGT	GTG	GCC	TTG	GCT	CAT	GCT	TCC	48
Met	Leu	Leu	Leu	Leu	Ala	Leu	Val	Ser	Val	Ala	Leu	Ala	His	Ala	Ser	
1				5					10					15		
GAG	GAG	CAC	TTT	GAT	GGC	AAC	CGG	GTG	TAC	CGT	GTC	AGT	GTA	CAT	GGT	96
Glu	Glu	His	Phe	Asp	Gly	Asn	Arg	Val	Tyr	Arg	Val	Ser	Val	His	Gly	
			20					25					30			
GAA	GAT	CAC	GTC	AAC	TTA	ATT	CAG	GAG	CTA	GCC	AAC	ACC	AAA	GAG	ATT	144
Glu	Asp	His	Val	Asn	Leu	Ile	Gln	Glu	Leu	Ala	Asn	Thr	Lys	Glu	Ile	
		35					40					45				
GAT	TTC	TGG	AAA	CCA	GAT	TCT	GCT	ACA	CAA	GTG	AAG	CCT	CTC	ACT	ACA	192
Asp	Phe	Trp	Lys	Pro	Asp	Ser	Ala	Thr	Gln	Val	Lys	Pro	Leu	Thr	Thr	
	50					55					60					

- 41 -

GTT GAC TTT CAT GTT AAA GCA GAA GAT GTT GCT GAT GTG GAG AAC TTT	240
Val Asp Phe His Val Lys Ala Glu Asp Val Ala Asp Val Glu Asn Phe	
65 70 75 80	
CTG GAG GAG AAT GAA GTT CAC TAT GAG GTA CTG ATA AGC AAC GTG AGA	288
Leu Glu Glu Asn Glu Val His Tyr Glu Val Leu Ile Ser Asn Val Arg	
85 90 95	
AAT GCT CTG GAA TCC CAG TTT GAT AGC CAC ACC CGT GCA AGT GGA CAC	336
Asn Ala Leu Glu Ser Gln Phe Asp Ser His Thr Arg Ala Ser Gly His	
100 105 110	
AGC TAC ACC AAG TAC AAC AAG TGG GAA ACG ATT GAG GCG TGG ATT CAA	384
Ser Tyr Thr Lys Tyr Asn Lys Trp Glu Thr Ile Glu Ala Trp Ile Gln	
115 120 125	
CAA GTT GCC ACT GAT AAT CCA GAC CTT GTC ACT CAG AGC GTC ATT GGA	432
Gln Val Ala Thr Asp Asn Pro Asp Leu Val Thr Gln Ser Val Ile Gly	
130 135 140	
ACC ACA TTT GAA GGA CGT AAC ATG TAT GTC CTC AAG ATT GGC AAA ACT	480
Thr Thr Phe Glu Gly Arg Asn Met Tyr Val Leu Lys Ile Gly Lys Thr	
145 150 155 160	
AGA CCG AAT AAG CCT GCC ATC TTC ATC GAT TGT GGT TTC CAT GCA AGA	528
Arg Pro Asn Lys Pro Ala Ile Phe Ile Asp Cys Gly Phe His Ala Arg	
165 170 175	
GAG TGG ATT TCT CCT GCA TTC TGT CAG TGG TTT GTG AGA GAG GCT GTC	576
Glu Trp Ile Ser Pro Ala Phe Cys Gln Trp Phe Val Arg Glu Ala Val	
180 185 190	
CGT ACC TAT AAT CAA GAG ATC CAC ATG AAA CAG CTT CTA GAT GAA CTG	624
Arg Thr Tyr Asn Gln Glu Ile His Met Lys Gln Leu Leu Asp Glu Leu	
195 200 205	
GAT TTC TAT GTT CTG CCT GTG GTC AAC ATT GAT GGC TAT GTC TAC ACC	672
Asp Phe Tyr Val Leu Pro Val Val Asn Ile Asp Gly Tyr Val Tyr Thr	
210 215 220	
TGG ACT AAG GAC AGA ATG TGG AGA AAA ACC CGC TCT ACT ATG GCT GGA	720
Trp Thr Lys Asp Arg Met Trp Arg Lys Thr Arg Ser Thr Met Ala Gly	
225 230 235 240	
AGT TCC TGC TTG GGT GTA AGA CCC AAC AGG AAT TTT AAT GCT GGC TGG	768
Ser Ser Cys Leu Gly Val Arg Pro Asn Arg Asn Phe Asn Ala Gly Trp	
245 250 255	
TGT GAA GTG GGA GCT TCT CGG AGT CCC TGC TCT GAA ACT TAC TGT GGA	816
Cys Glu Val Gly Ala Ser Arg Ser Pro Cys Ser Glu Thr Tyr Cys Gly	
260 265 270	
CCA GCC CCA GAG TCT GAA AAA GAG ACA AAG GCC CTG GCA GAT TTC ATC	864
Pro Ala Pro Glu Ser Glu Lys Glu Thr Lys Ala Leu Ala Asp Phe Ile	
275 280 285	

- 42 -

CGC AAC AAC CTC TCC ACC ATC AAG GCC TAC CTG ACC ATC CAC TCA TAC	912
Arg Asn Asn Leu Ser Thr Ile Lys Ala Tyr Leu Thr Ile His Ser Tyr	
290 295 300	
TCA CAG ATG ATG CTC TAC CCT TAC TCC TAT GAC TAC AAA CTG CCT GAG	960
Ser Gln Met Met Leu Tyr Pro Tyr Ser Tyr Asp Tyr Lys Leu Pro Glu	
305 310 315 320	
AAC TAT GAG GAA TTG AAT GCC CTG GTG AAA GGT GCG GCA AAG GAG CTT	1008
Asn Tyr Glu Glu Leu Asn Ala Leu Val Lys Gly Ala Ala Lys Glu Leu	
325 330 335	
GCC ACT CTG CAT GGC ACC AAG TAC ACA TAT GGC CCA GGA GCT ACA ACA	1056
Ala Thr Leu His Gly Thr Lys Tyr Thr Tyr Gly Pro Gly Ala Thr Thr	
340 345 350	
ATC TAT CCT GCT GCT GGG GGA TCT GAC GAC TGG TCT TAT GAT CAG GGA	1104
Ile Tyr Pro Ala Ala Gly Gly Ser Asp Asp Trp Ser Tyr Asp Gln Gly	
355 360 365	
ATC AAA TAT TCA TTT ACC TTT GAA CTC CGG GAT ACA GGC TTC TTT GGC	1152
Ile Lys Tyr Ser Phe Thr Phe Glu Leu Arg Asp Thr Gly Phe Phe Gly	
370 375 380	
TTT CTC CTT CCT GAG TCT CAG ATC CGC CAG ACC TGT GAG GAG ACA ATG	1200
Phe Leu Leu Pro Glu Ser Gln Ile Arg Gln Thr Cys Glu Glu Thr Met	
385 390 395 400	
CTT GCA GTC AAG TAC ATT GCC AAT TAT GTC CGA GAA CAT CTA TAT TAG	1248
Leu Ala Val Lys Tyr Ile Ala Asn Tyr Val Arg Glu His Leu Tyr	
405 410 415	

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Rattus norvegicus

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

Met Leu Leu Leu Leu Ala Leu Val Ser Val Ala Leu Ala His Ala Ser
1 5 10 15
Glu Glu His Phe Asp Gly Asn Arg Val Tyr Arg Val Ser Val His Gly
20 25 30

- 43 -

Glu	Asp	His	Val	Asn	Leu	Ile	Gln	Glu	Leu	Ala	Asn	Thr	Lys	Glu	Ile	35	40	45	
Asp	Phe	Trp	Lys	Pro	Asp	Ser	Ala	Thr	Gln	Val	Lys	Pro	Leu	Thr	Thr	50	55	60	
Val	Asp	Phe	His	Val	Lys	Ala	Glu	Asp	Val	Ala	Asp	Val	Glu	Asn	Phe	65	70	75	80
Leu	Glu	Glu	Asn	Glu	Val	His	Tyr	Glu	Val	Leu	Ile	Ser	Asn	Val	Arg	85	90	95	
Asn	Ala	Leu	Glu	Ser	Gln	Phe	Asp	Ser	His	Thr	Arg	Ala	Ser	Gly	His	100	105	110	
Ser	Tyr	Thr	Lys	Tyr	Asn	Lys	Trp	Glu	Thr	Ile	Glu	Ala	Trp	Ile	Gln	115	120	125	
Gln	Val	Ala	Thr	Asp	Asn	Pro	Asp	Leu	Val	Thr	Gln	Ser	Val	Ile	Gly	130	135	140	
Thr	Thr	Phe	Glu	Gly	Arg	Asn	Met	Tyr	Val	Leu	Lys	Ile	Gly	Lys	Thr	145	150	155	160
Arg	Pro	Asn	Lys	Pro	Ala	Ile	Phe	Ile	Asp	Cys	Gly	Phe	His	Ala	Arg	165	170	175	
Glu	Trp	Ile	Ser	Pro	Ala	Phe	Cys	Gln	Trp	Phe	Val	Arg	Glu	Ala	Val	180	185	190	
Arg	Thr	Tyr	Asn	Gln	Glu	Ile	His	Met	Lys	Gln	Leu	Leu	Asp	Glu	Leu	195	200	205	
Asp	Phe	Tyr	Val	Leu	Pro	Val	Val	Asn	Ile	Asp	Gly	Tyr	Val	Tyr	Thr	210	215	220	
Trp	Thr	Lys	Asp	Arg	Met	Trp	Arg	Lys	Thr	Arg	Ser	Thr	Met	Ala	Gly	225	230	235	240
Ser	Ser	Cys	Leu	Gly	Val	Arg	Pro	Asn	Arg	Asn	Phe	Asn	Ala	Gly	Trp	245	250	255	
Cys	Glu	Val	Gly	Ala	Ser	Arg	Ser	Pro	Cys	Ser	Glu	Thr	Tyr	Cys	Gly	260	265	270	
Pro	Ala	Pro	Glu	Ser	Glu	Lys	Glu	Thr	Lys	Ala	Leu	Ala	Asp	Phe	Ile	275	280	285	
Arg	Asn	Asn	Leu	Ser	Thr	Ile	Lys	Ala	Tyr	Leu	Thr	Ile	His	Ser	Tyr	290	295	300	
Ser	Gln	Met	Met	Leu	Tyr	Pro	Tyr	Ser	Tyr	Asp	Tyr	Lys	Leu	Pro	Glu	305	310	315	320

- 44 -

```

Asn Tyr Glu Glu Leu Asn Ala Leu Val Lys Gly Ala Ala Lys Glu Leu
      325                      330                      335

Ala Thr Leu His Gly Thr Lys Tyr Thr Tyr Gly Pro Gly Ala Thr Thr
      340                      345                      350

Ile Tyr Pro Ala Ala Gly Gly Ser Asp Asp Trp Ser Tyr Asp Gln Gly
      355                      360                      365

Ile Lys Tyr Ser Phe Thr Phe Glu Leu Arg Asp Thr Gly Phe Phe Gly
      370                      375                      380

Phe Leu Leu Pro Glu Ser Gln Ile Arg Gln Thr Cys Glu Glu Thr Met
      385                      390                      395                      400

Leu Ala Val Lys Tyr Ile Ala Asn Tyr Val Arg Glu His Leu Tyr
      405                      410                      415

```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1497 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Artificial

(ix) FEATURE:

- (D) OTHER INFORMATION: Artificial sequence encoding pre-pro-carboxypeptidase B polypeptide, optimized for expression in methylotrophic yeast.

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: (1)..(1497)
- (D) OTHER INFORMATION: Coding sequence including stop codons.

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: (1)..(255)
- (D) OTHER INFORMATION: Sequence encoding a Saccharomyces cerevisiae alpha-factor signal peptide sequence which contains a signal peptidase cleavage site.

(ix) FEATURE:

- (A) NAME/KEY: misc_feature
- (B) LOCATION: (256)..(261)

- 45 -

(D) OTHER INFORMATION: Sequence encoding a spacer.

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: (262)..(288)

(D) OTHER INFORMATION: Sequence encoding the Histidine-tag, RGSHHHHHH.

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: (286)..(570)

(D) OTHER INFORMATION: Sequence encoding the rat carboxypeptidase B propeptide; the first codon is also part of the Histidine-tag. The sequence incorporates the amino acid exchanges Lys14Asn and Arg142Asp documented in WO 96/23064.

(ix) FEATURE:

(A) NAME/KEY: misc_feature

(B) LOCATION: (571)..(1497)

(D) OTHER INFORMATION: Sequence encoding the enzyme moiety of rat carboxypeptidase B

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

ATG AGA TTT CCT TCA ATT TTT ACT GCT GTT TTA TTC GCA GCA TCC TCC	48
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser	
1 5 10 15	
GCA TTA GCT GCT CCA GTC AAC ACT ACA ACA GAA GAT GAA ACG GCA CAA	96
Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln	
20 25 30	
ATT CCG GCT GAA GCT GTC ATC GGT TAC TCA GAT TTA GAA GGG GAT TTC	144
Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe	
35 40 45	
GAT GTT GCT GTT TTG CCA TTT TCC AAC AGC ACA AAT AAC GGG TTA TTG	192
Asp Val Ala Val Leu Pro Phe Ser Asn Ser Thr Asn Asn Gly Leu Leu	
50 55 60	
TTT ATA AAT ACT ACT ATT GCC AGC ATT GCT GCT AAA GAA GAA GGG GTA	240
Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val	
65 70 75 80	
TCT CTC GAG AAG AGA TCC GCT AGA GGT TCT CAC CAC CAT CAC CAT CAC	288
Ser Leu Glu Lys Arg Ser Ala Arg Gly Ser His His His His His His	
85 90 95	
GCT TCT GAG GAG CAC TTC GAC GGT AAC AGA GTT TAC AGA GTT TCT GTT	336
Ala Ser Glu Glu His Phe Asp Gly Asn Arg Val Tyr Arg Val Ser Val	
100 105 110	

- 46 -

CAC	GGT	GAG	GAC	CAC	GTT	AAC	TTG	ATT	CAA	GAG	TTG	GCT	AAC	ACT	AAG	384
His	Gly	Glu	Asp	His	Val	Asn	Leu	Ile	Gln	Glu	Leu	Ala	Asn	Thr	Lys	
		115					120					125				
GAG	ATT	GAC	TTC	TGG	AAG	CCA	GAC	TCT	GCT	ACT	CAA	GTT	AAG	CCA	TTG	432
Glu	Ile	Asp	Phe	Trp	Lys	Pro	Asp	Ser	Ala	Thr	Gln	Val	Lys	Pro	Leu	
	130					135					140					
ACT	ACT	GTT	GAC	TTC	CAC	GTT	AAG	GCT	GAG	GAC	GTT	GCC	GAT	GTT	GAA	480
Thr	Thr	Val	Asp	Phe	His	Val	Lys	Ala	Glu	Asp	Val	Ala	Asp	Val	Glu	
145					150					155					160	
AAC	TTC	TTG	GAG	GAG	AAC	GAG	GTT	CAC	TAC	GAA	GTT	TTG	ATC	TCT	AAC	528
Asn	Phe	Leu	Glu	Glu	Asn	Glu	Val	His	Tyr	Glu	Val	Leu	Ile	Ser	Asn	
				165				170						175		
GTT	CGT	AAC	GCT	TTG	GAA	TCC	CAA	TTC	GAC	TCT	CAC	ACT	AGA	GCT	TCT	576
Val	Arg	Asn	Ala	Leu	Glu	Ser	Gln	Phe	Asp	Ser	His	Thr	Arg	Ala	Ser	
			180					185					190			
GGT	CAC	TCT	TAC	ACT	AAG	TAC	AAC	AAC	TGG	GAG	ACT	ATT	GAG	GCT	TGG	624
Gly	His	Ser	Tyr	Thr	Lys	Tyr	Asn	Asn	Trp	Glu	Thr	Ile	Glu	Ala	Trp	
		195					200					205				
ATT	CAA	CAA	GTT	GCT	ACT	GAC	AAC	CCA	GAC	TTG	GTT	ACT	CAA	TCT	GTT	672
Ile	Gln	Gln	Val	Ala	Thr	Asp	Asn	Pro	Asp	Leu	Val	Thr	Gln	Ser	Val	
	210					215					220					
ATT	GGT	ACT	ACT	TTC	GAG	GGT	AGA	AAC	ATG	TAC	GTT	TTG	AAG	ATT	GGT	720
Ile	Gly	Thr	Thr	Phe	Glu	Gly	Arg	Asn	Met	Tyr	Val	Leu	Lys	Ile	Gly	
225					230					235					240	
AAG	ACT	AGA	CCA	AAC	AAG	CCA	GCT	ATT	TTC	ATT	GAC	TGT	GGT	TTC	CAC	768
Lys	Thr	Arg	Pro	Asn	Lys	Pro	Ala	Ile	Phe	Ile	Asp	Cys	Gly	Phe	His	
				245				250						255		
GCT	AGA	GAA	TGG	ATT	TCC	CCA	GCT	TTC	TGT	CAA	TGG	TTC	GTT	AGA	GAG	816
Ala	Arg	Glu	Trp	Ile	Ser	Pro	Ala	Phe	Cys	Gln	Trp	Phe	Val	Arg	Glu	
			260					265					270			
GCT	GTT	AGA	ACT	TAC	AAC	CAA	GAG	ATT	CAC	ATG	AAG	CAA	TTG	TTG	GAC	864
Ala	Val	Arg	Thr	Tyr	Asn	Gln	Glu	Ile	His	Met	Lys	Gln	Leu	Leu	Asp	
		275					280					285				
GAG	TTG	GAC	TTC	TAC	GTT	TTG	CCA	GTT	GTT	AAC	ATT	GAC	GGT	TAC	GTT	912
Glu	Leu	Asp	Phe	Tyr	Val	Leu	Pro	Val	Val	Asn	Ile	Asp	Gly	Tyr	Val	
	290					295					300					
TAC	ACT	TGG	ACT	AAG	GAC	AGA	ATG	TGG	AGA	AAG	ACT	CGT	TCC	ACT	ATG	960
Tyr	Thr	Trp	Thr	Lys	Asp	Arg	Met	Trp	Arg	Lys	Thr	Arg	Ser	Thr	Met	
305					310					315					320	
GCT	GGT	TCT	TCT	TGC	CTT	GGT	GTC	GAT	CCA	AAT	AGA	AAC	TTT	AAC	GCT	1008
Ala	Gly	Ser	Ser	Cys	Leu	Gly	Val	Asp	Pro	Asn	Arg	Asn	Phe	Asn	Ala	
				325				330					335			

- 47 -

GGT TGG TGT GAG GTC GGT GCT TCT AGA TCC CCA TGC TCT GAA ACT TAC	1056
Gly Trp Cys Glu Val Gly Ala Ser Arg Ser Pro Cys Ser Glu Thr Tyr	
340 345 350	
TGT GGT CCT GCT CCT GAA TCT GAA AAG GAG ACT AAG GCT TTG GCT GAC	1104
Cys Gly Pro Ala Pro Glu Ser Glu Lys Glu Thr Lys Ala Leu Ala Asp	
355 360 365	
TTC ATT AGA AAC AAC TTG TCT ACT ATT AAG GCT TAC TTG ACT ATT CAC	1152
Phe Ile Arg Asn Asn Leu Ser Thr Ile Lys Ala Tyr Leu Thr Ile His	
370 375 380	
TCT TAC TCT CAA ATG ATG TTG TAC CCA TAC TCT TAC GAC TAC AAG TTG	1200
Ser Tyr Ser Gln Met Met Leu Tyr Pro Tyr Ser Tyr Asp Tyr Lys Leu	
385 390 395 400	
CCA GAA AAC TAC GAG GAG TTG AAC GCT TTG GTT AAG GGT GCT GCT AAA	1248
Pro Glu Asn Tyr Glu Glu Leu Asn Ala Leu Val Lys Gly Ala Ala Lys	
405 410 415	
GAA TTG GCT ACT TTG CAC GGT ACT AAA TAC ACT TAC GGT CCA GGT GCT	1296
Glu Leu Ala Thr Leu His Gly Thr Lys Tyr Thr Tyr Gly Pro Gly Ala	
420 425 430	
ACT ACT ATT TAC CCA GCT GCT GGT GGT TCT GAC GAC TGG TCT TAC GAC	1344
Thr Thr Ile Tyr Pro Ala Ala Gly Gly Ser Asp Asp Trp Ser Tyr Asp	
435 440 445	
CAA GGT ATT AAG TAC TCT TTC ACT TTC GAG TTG AGA GAT ACT GGT TTC	1392
Gln Gly Ile Lys Tyr Ser Phe Thr Phe Glu Leu Arg Asp Thr Gly Phe	
450 455 460	
TTC GGT TTC TTG TTG CCT GAG TCC CAA ATT AGA CAA ACT TGT GAG GAA	1440
Phe Gly Phe Leu Leu Pro Glu Ser Gln Ile Arg Gln Thr Cys Glu Glu	
465 470 475 480	
ACC ATG TTG GCT GTT AAG TAC ATT GCT AAC TAC GTT AGA GAG CAC TTG	1488
Thr Met Leu Ala Val Lys Tyr Ile Ala Asn Tyr Val Arg Glu His Leu	
485 490 495	
TAC TAA TAA	1497
Tyr	

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 497 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- 48 -

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Artificial

(ix) FEATURE:

(D) OTHER INFORMATION: Artificial sequence encoding
pre-pro-carboxypeptidase B
polypeptide, optimized for expression
in methylotrophic yeast.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met	Arg	Phe	Pro	Ser	Ile	Phe	Thr	Ala	Val	Leu	Phe	Ala	Ala	Ser	Ser	1	5	10	15
Ala	Leu	Ala	Ala	Pro	Val	Asn	Thr	Thr	Thr	Glu	Asp	Glu	Thr	Ala	Gln	20	25	30	
Ile	Pro	Ala	Glu	Ala	Val	Ile	Gly	Tyr	Ser	Asp	Leu	Glu	Gly	Asp	Phe	35	40	45	
Asp	Val	Ala	Val	Leu	Pro	Phe	Ser	Asn	Ser	Thr	Asn	Asn	Gly	Leu	Leu	50	55	60	
Phe	Ile	Asn	Thr	Thr	Ile	Ala	Ser	Ile	Ala	Ala	Lys	Glu	Glu	Gly	Val	65	70	75	80
Ser	Leu	Glu	Lys	Arg	Ser	Ala	Arg	Gly	Ser	His	His	His	His	His	His	85	90	95	
Ala	Ser	Glu	Glu	His	Phe	Asp	Gly	Asn	Arg	Val	Tyr	Arg	Val	Ser	Val	100	105	110	
His	Gly	Glu	Asp	His	Val	Asn	Leu	Ile	Gln	Glu	Leu	Ala	Asn	Thr	Lys	115	120	125	
Glu	Ile	Asp	Phe	Trp	Lys	Pro	Asp	Ser	Ala	Thr	Gln	Val	Lys	Pro	Leu	130	135	140	
Thr	Thr	Val	Asp	Phe	His	Val	Lys	Ala	Glu	Asp	Val	Ala	Asp	Val	Glu	145	150	155	160
Asn	Phe	Leu	Glu	Glu	Asn	Glu	Val	His	Tyr	Glu	Val	Leu	Ile	Ser	Asn	165	170	175	
Val	Arg	Asn	Ala	Leu	Glu	Ser	Gln	Phe	Asp	Ser	His	Thr	Arg	Ala	Ser	180	185	190	
Gly	His	Ser	Tyr	Thr	Lys	Tyr	Asn	Asn	Trp	Glu	Thr	Ile	Glu	Ala	Trp	195	200	205	
Ile	Gln	Gln	Val	Ala	Thr	Asp	Asn	Pro	Asp	Leu	Val	Thr	Gln	Ser	Val	210	215	220	
Ile	Gly	Thr	Thr	Phe	Glu	Gly	Arg	Asn	Met	Tyr	Val	Leu	Lys	Ile	Gly	225	230	235	240

Thr Met Leu Ala Val Lys Tyr Ile Ala Asn Tyr Val Arg Glu His Leu
485 490 495

Tyr

- 50 -

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 938 bases
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Pichia pastoris*

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: (1)..(938)
- (D) OTHER INFORMATION: *Pichia pastoris* AOX promoter

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

AGATCTAACA	TCCAAAGACG	AAAGGTTGAA	TGAAACCTTT	TTGCCATCCG	ACATCCACAG	60
GTCCATTCTC	ACACATAAGT	GCCAAACGCA	ACAGGAGGGG	ATACACTAGC	AGCAGACCGT	120
TGCAAACGCA	GGACCTCCAC	TCCTCTTCTC	CTCAACACCC	ACTTTTGCCA	TCGAAAAACC	180
AGCCCAAGTTA	TTGGGCTTGA	TTGGAGCTCG	CTCATTCCAA	TTCCTTCTAT	TAGGCTACTA	240
ACACCATGAC	TTTATTAGCC	TGTCTATCCT	GGCCCCCTG	GCGAGGTTCA	TGTTTGTTTA	300
TTTCCGAATG	CAACAAGCTC	CGCATTACAC	CCGAACATCA	CTCCAGATGA	GGGCTTTCTG	360
AGTGTGGGGT	CAAATAGTTT	CATGTTCCCC	AAATGGCCCA	AAACTGACAG	TTTAAACGCT	420
GTCTTGGAAC	CTAATATGAC	AAAAGCGTGA	TCTCATCCAA	GATGAACTAA	GTTTGGTTCG	480
TTGAAATGCT	AACGGCCAGT	TGGTCAAAAA	GAAACTTCCA	AAAGTCGGCA	TACCGTTTGT	540
CTTGTTTGGT	ATTGATTGAC	GAATGCTCAA	AAATAATCTC	ATTAATGCTT	AGCGCAGTCT	600
CTCTATCGCT	TCTGAACCCC	GGTGCACCTG	TGCCGAAACG	CAAATGGGGA	AACACCCGCT	660
TTTTGGATGA	TTATGCATTG	TCTCCACATT	GTATGCTTCC	AAGATTCTGG	TGGGAATACT	720
GCTGATAGCC	TAACGTTCAT	GATCAAAATT	TAACTGTTCT	AACCCCTACT	TGACAGCAAT	780
ATATAAACAG	AAGGAAGCTG	CCCTGTCTTA	AACCTTTTTT	TTTATCATCA	TTATTAGCTT	840
ACTTTCATAA	TTGCGACTGG	TTCCAATTGA	CAAGCTTTTG	ATTTTAACGA	CTTTTAACGA	900
CAACTTGAGA	AGATCAAAAA	ACAATAATT	ATTCGAAA			938

Patent Claims

1. A method to produce a protein with carboxypeptidase B activity, comprising the steps of
 - 5 a) providing a vector comprising a nucleotide sequence which encodes a pre-protein consisting of the rat pro-carboxypeptidase B that is N-terminally fused to a Histidine-tag and a signal peptide,

whereby optionally between the Histidine-tag and the signal peptide or between the Histidine-tag and rat carboxypeptidase B a spacer sequence is inserted;
 - 10 (b) transforming a microbial host organism with the vector;
 - (c) cultivating the microbial host organism in a growth medium containing nutrients and a carbon source, whereby the microbial host organism expresses the pre-protein and secretes the Histidine-tagged pro-carboxypeptidase B into the growth medium;
 - 15 (d) immobilizing the secreted Histidine-tagged pro-carboxypeptidase B in the growth medium of step (c) on a particulate metal chelate affinity matrix capable of binding the Histidine-tag, and washing the particulate metal chelate affinity matrix, whereby the Histidine-tagged pro-carboxypeptidase B is immobilized;
 - 20 (e) incubating the particulate metal chelate affinity matrix with the immobilized Histidine-tagged pro-carboxypeptidase B of step (d) in a buffer containing trypsin, thereby cleaving proteolytically the pro-carboxypeptidase B moiety and releasing the protein with carboxypeptidase B activity into the liquid phase,
 - 25 whereby the Histidine-tagged propeptide moiety is immobilized;

- 52 -

- (f) separating the liquid phase containing the protein with carboxypeptidase B activity from the particulate metal chelate affinity matrix,
whereby the Histidine-tagged propeptide moiety is immobilized; and
 - (g) purifying the protein with carboxypeptidase B activity from the liquid phase of step (f).
2. The method according to claim 1, wherein the microbial host strain is a methylotrophic yeast strain.
3. The method according to claim 1 or 2, wherein the amino acid sequence of the rat pro-carboxypeptidase B is the amino acid sequence from position 14 to position 415 in SEQ ID NO: 3.
4. The method according to any one of claims 1 to 3, wherein the signal peptide contains a signal peptidase cleavage site which is located adjacent to the Histidine-tag or adjacent to the spacer sequence.
5. The method according to any one of claims 1 to 4, wherein the amino acid sequence of the expressed pre-protein is the amino acid sequence in SEQ ID NO: 4.
6. The method according to any one of claims 1, 2, 4 and 5, wherein the nucleotide sequence encoding rat pro-carboxypeptidase B is the nucleotide sequence from position 286 to position 1,497 in SEQ ID NO: 3.
7. The method according to any one of claims 1 to 4, wherein the nucleotide sequence encoding the pre-protein is the nucleotide sequence in SEQ ID NO: 3.

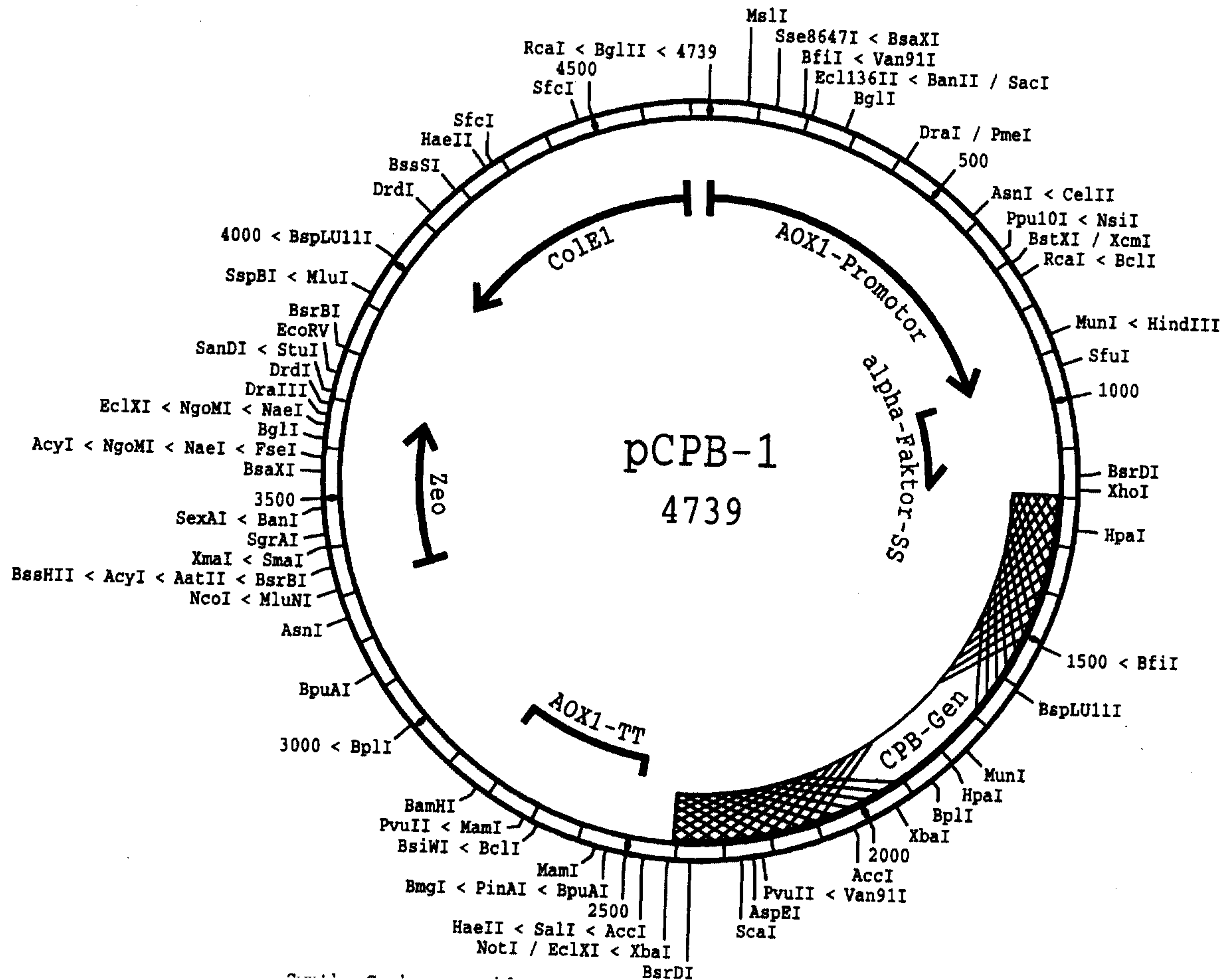
Figure 1

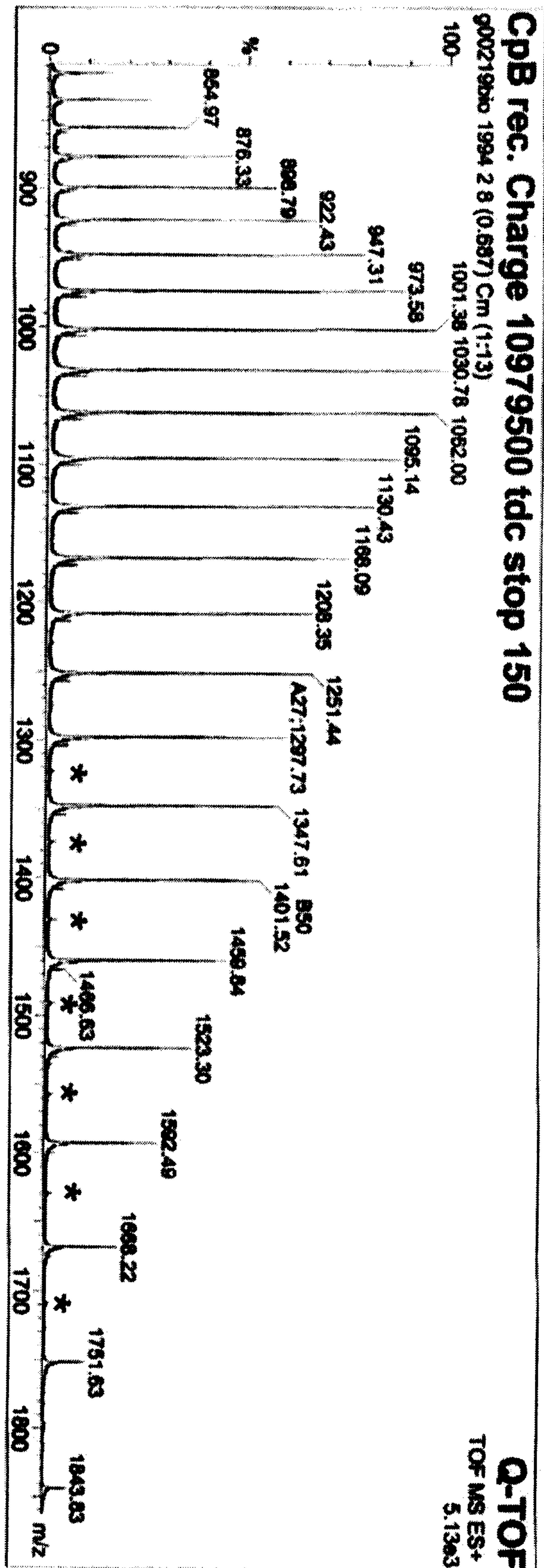
Figure 2a

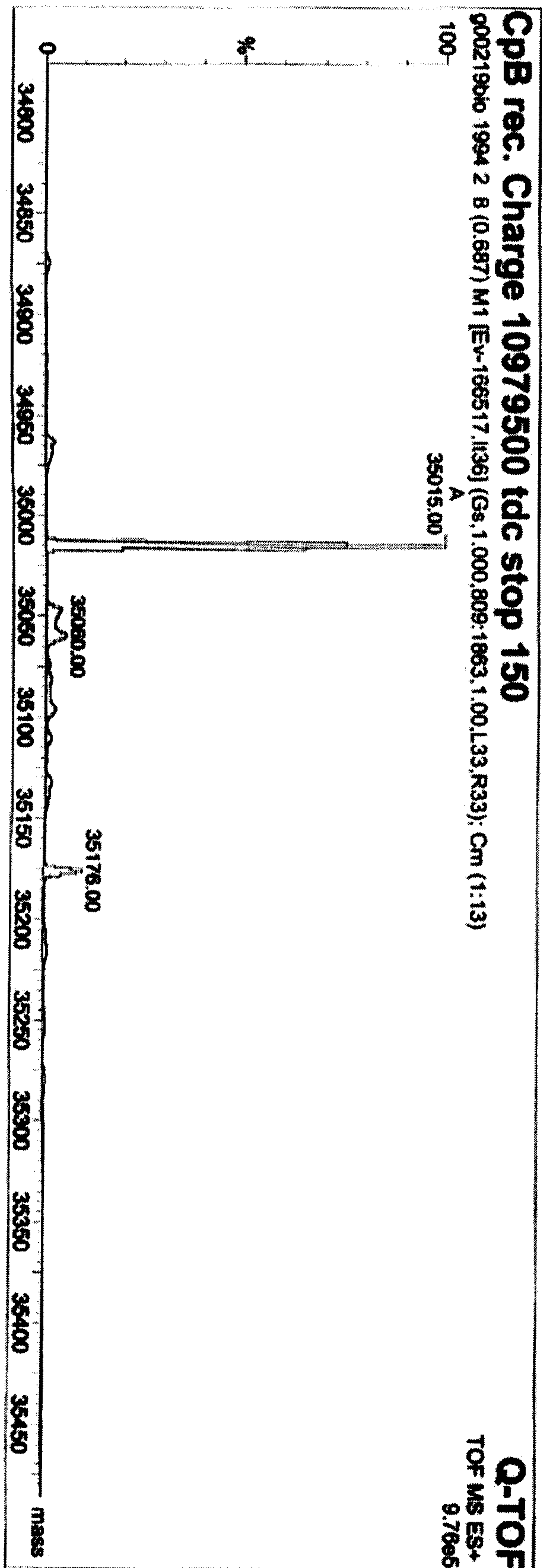
Figure 2b

Figure 3

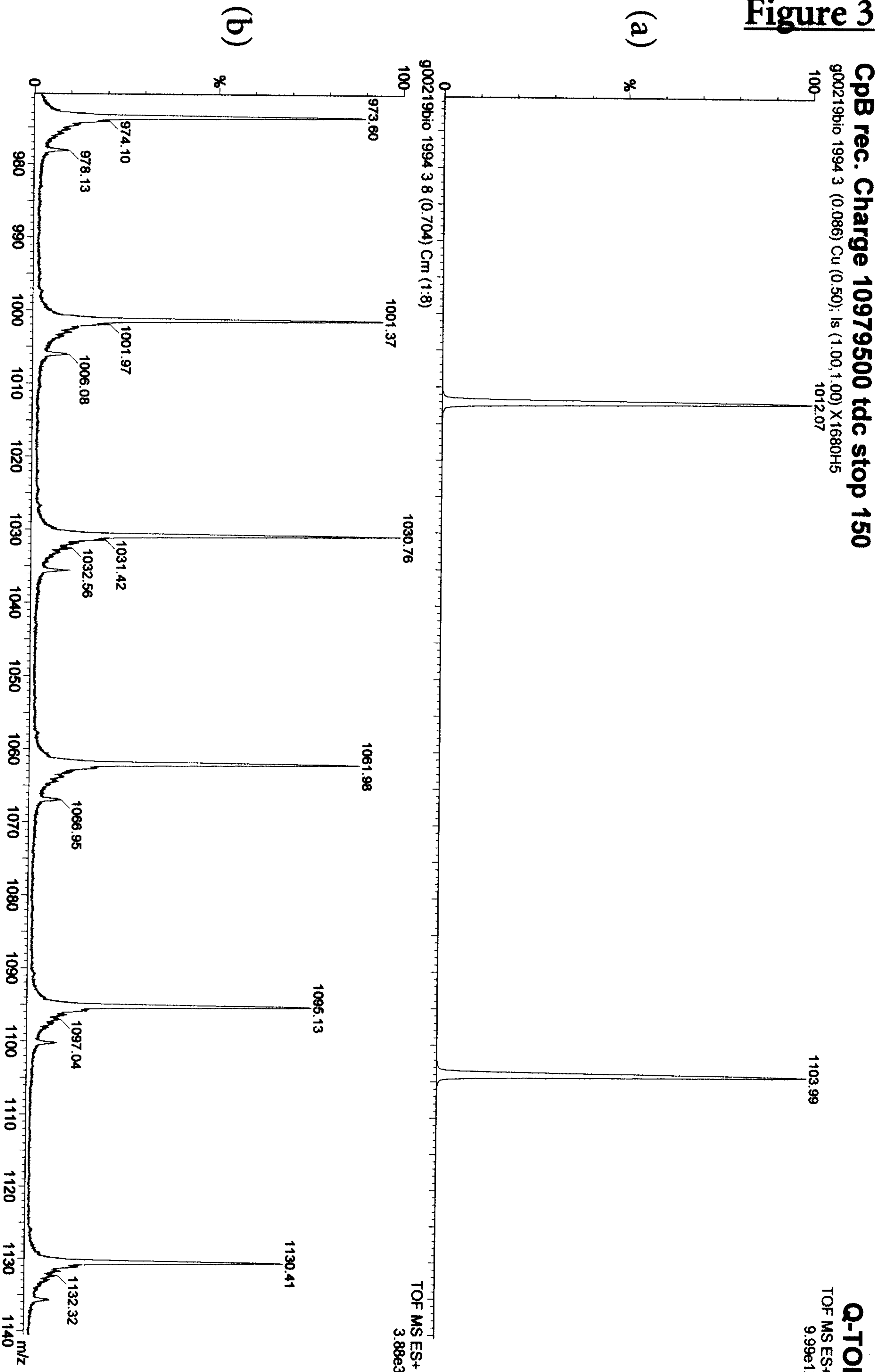


Figure 4

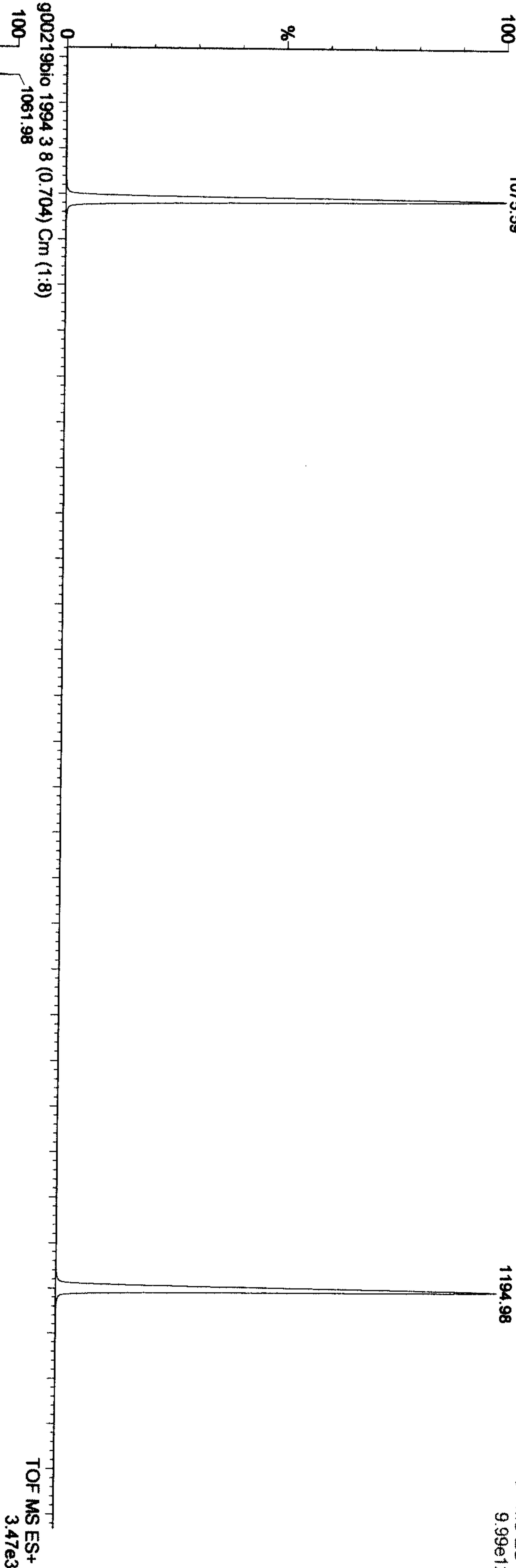
CpB rec. Charge 10979500 tdc stop 150

g00219bio 1994 3 (0.086) Cu (0.50); Is (1.00,1.00) X1488H4

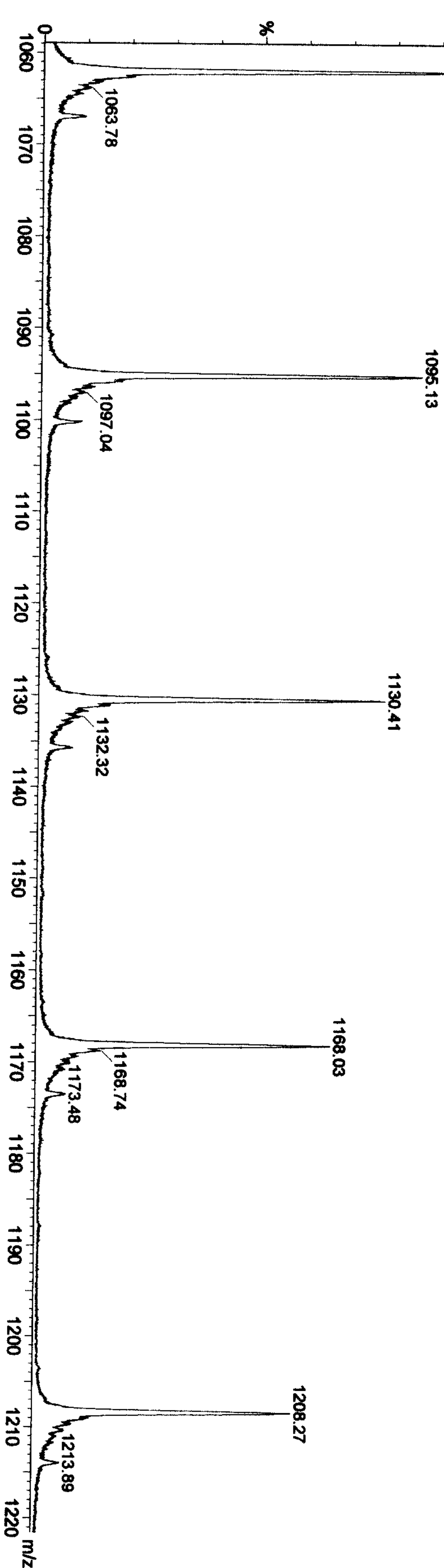
1075.59

Q-TOF
TOF MS ES+
9.99e12

(a)



(b)



6/9

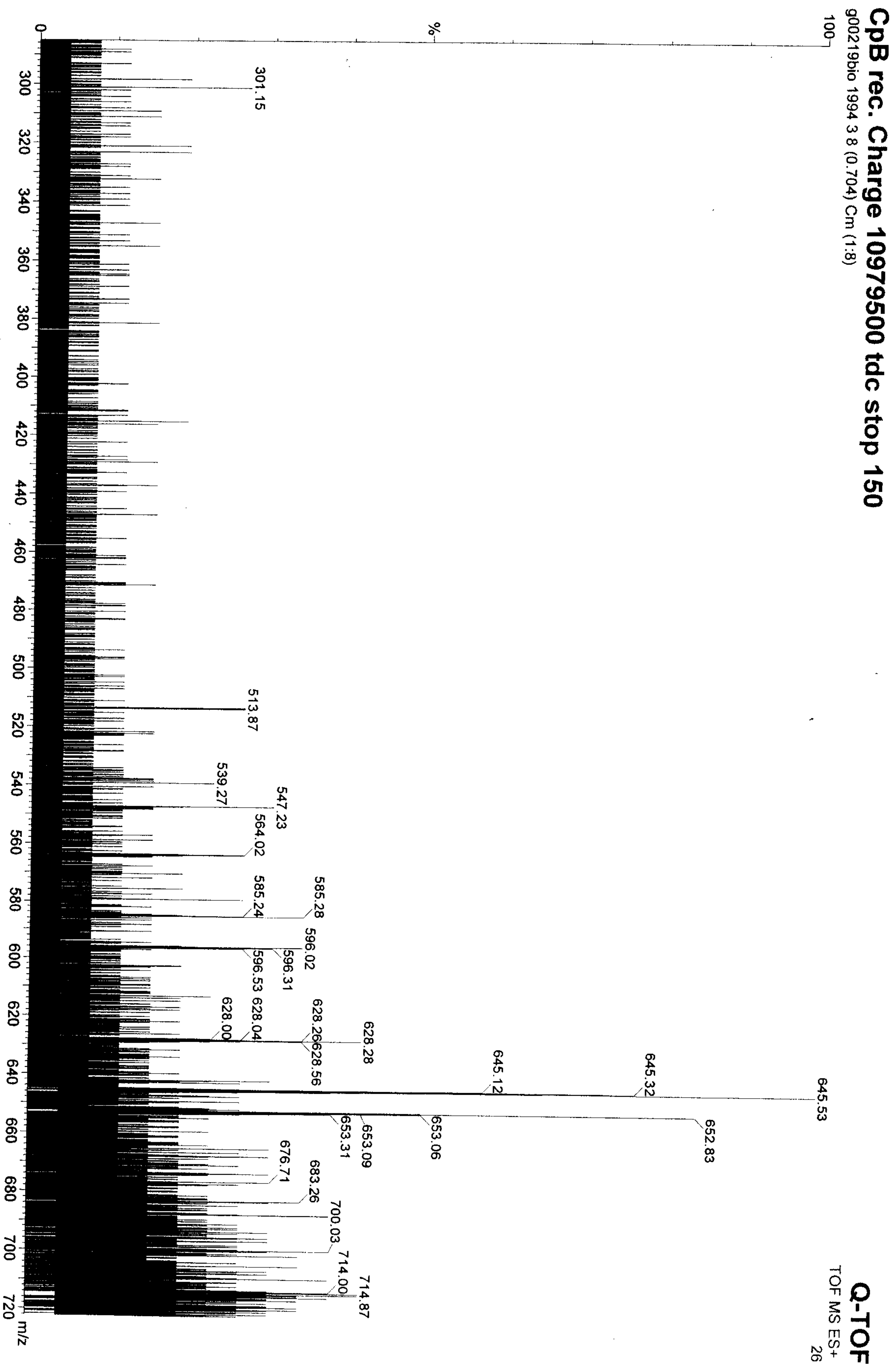
Figure 5

Figure 6

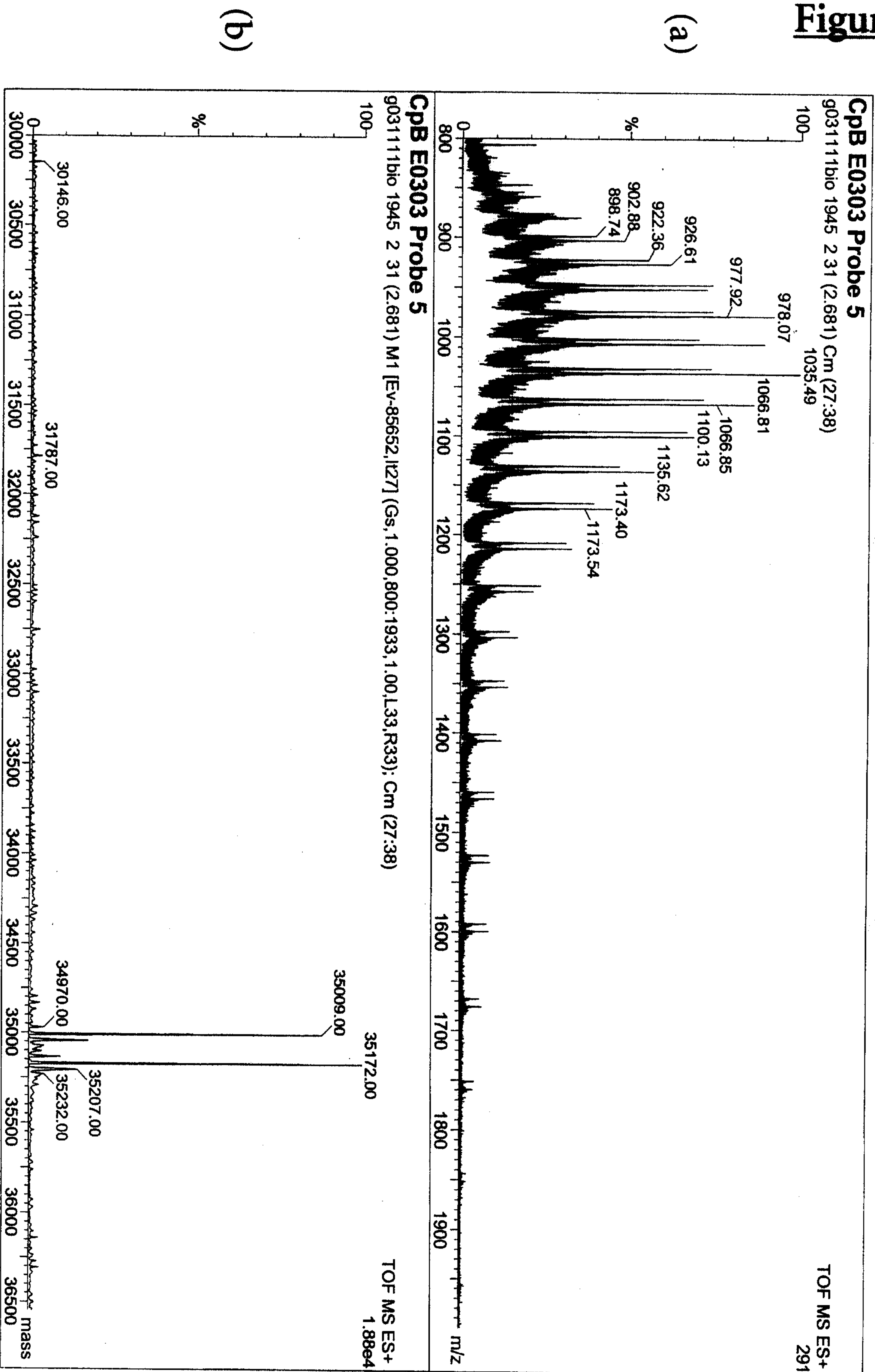


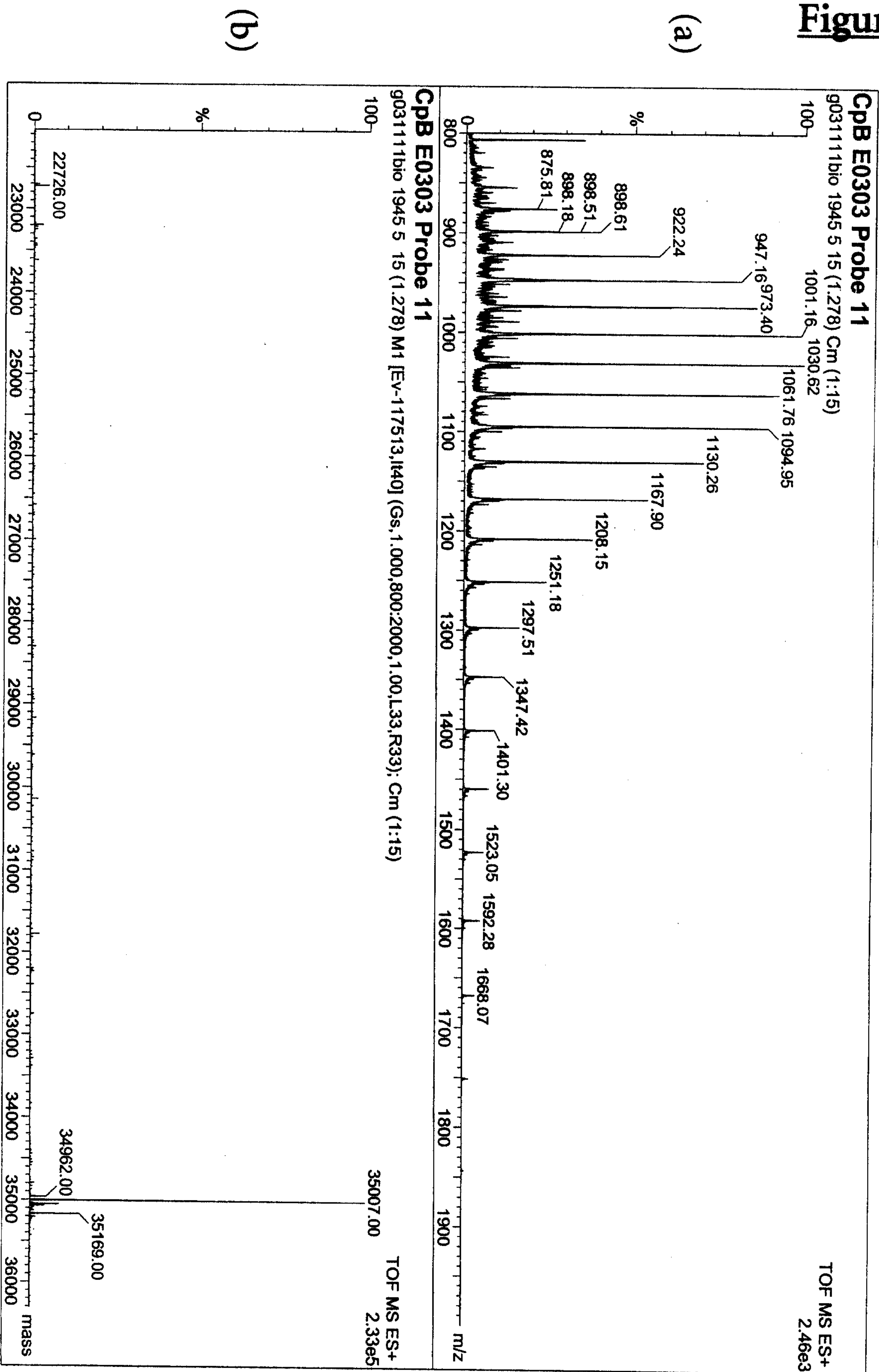
Figure 7

Figure 8

