



(51) International Patent Classification:
C12N 9/64 (2006.01)

(21) International Application Number:
PCT/EP2017/050303

(22) International Filing Date:
9 January 2017 (09.01.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN2016/070391
7 January 2016 (07.01.2016) CN
16158052.7 1 March 2016 (01.03.2016) EP

(71) Applicant: **NOVO NORDISK A/S** [DK/DK]; Novo Allé,
2880 Bagsværd (DK).

(72) Inventors: **LIU, Yun**; c/o Novo Nordisk R/D Center
China, Building 2, 20 Life Science Park Rd, Changping
District, Beijing, 102206 (CN). **GAO, Xiang**; c/o Novo
Nordisk R/D Center China, Building 2, 20 Life Science
Park Rd, Changping District, Beijing, 102206 (CN).
THØGERSEN, Henning; C/O Novo Nordisk A/S, Novo
Allé, 2880 Bagsværd (DK).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA,
MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG,
NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS,
RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY,
TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN,
ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: MODIFIED ENTEROKINASE LIGHT CHAIN AND ITS PREPARATION METHOD

(57) Abstract: The present invention is related to novel mammalian enterokinase analogues such as mammalian enterokinase light chain analogues and methods of making such. Also described herein is a method for cleaving proteins having an enterokinase cleavage site.



WO 2017/118752 A1

MODIFIED ENTEROKINASE LIGHT CHAIN AND ITS PREPARATION METHOD

TECHNICAL FIELD

The present invention is related to novel mammalian enterokinase analogues,
5 methods of making such and the use of said mammalian enterokinase analogues for
cleaving proteins having an enterokinase cleavage site.

BACKGROUND

In *E.coli* many exogenous proteins are expressed as fusion proteins, which have to
be cleaved to release the mature, active protein. For that purpose a processing enzyme is
10 needed, preferably one which cleaves directly at the junction leaving no extra amino acids on
the product. Serine protease enterokinase is such an enzyme.

The serine protease enterokinase (in short enterokinase or EK), also known as
enteropeptidase, is a heterodimeric glycoprotein, a mammalian enzyme catalyzing the
conversion of trypsinogen into active trypsin. Enterokinase has exhibits high specificity for
15 the substrate sequence Asp-Asp-Asp-Asp-Lys ((Asp)₄-Lys, DDDDK, SEQ ID NO: 16), where
it selectively cleaves after lysine. The high specificity for the recognition site makes
enterokinase (EK) a useful tool for in vitro cleavage of fusion proteins. Enterokinase isolated
from bovine duodenal mucosa exhibits a molecular weight (MW) of 150,000 and a
carbohydrate content of 35 percent. The enzyme is comprised of a heavy chain
20 (MW~115,000) and a disulfide-linked light chain (MW~35,000) (Liepnieks et al., J. Biol.
Chem., 254(5): 1677-1683 (1979)). The function of the heavy chain is to anchor the enzyme
to the mucosal membrane. The light chain acts as the catalytic subunit.

The wild type bovine enterokinase light chain generally exhibits good activity in the
presence of various detergents and denaturants over a wide pH range (4.5–9.5) and
25 temperature range (4–45 °C). Therefore, the enterokinase light chain as a powerful tool has
been used in biotechnology for the in vitro cleavage of fusion proteins.

However, the complicated production processes and low production yield extracted
from animals, such as porcine and bovine, has set a limitation to EK application in
biotechnology. Recently, much effort has been made to establish a recombinant process to
30 obtain enterokinase or enterokinase analogues in *E.coli*. Recombinant enterokinase light
chain in *E.coli* has been obtained by secretion of active enterokinase light chain or by
intracellular accumulation of inclusion bodies of inactive enterokinase light chain, refolding
and activation.

The secretion process in *E.coli* aiming at a soluble EK product leads to a very low yield and expensive purification procedure.

In order to get a uniform product, the EK is preferred to be produced as insoluble material in inclusion bodies. However the low yield of EK light chain is bottle neck for its industrial application. A main reason for the low production yield is the low refolding yield of EK light chain. EK light chain is very likely to precipitate during refolding. Thus, when EK light chain is expressed in inclusion bodies, the concentration of soluble protein in refolding buffer shall be kept at very low level. It means that a very large volume of container shall be used for refolding. Therefore, from production perspective, it is not efficient or even practical at all. Another bottle neck for producing EK light chain in inclusion bodies is that it is difficult to obtain correctly refolded protein, because there are 4 disulfide bridges in the EK light chain.

An object of the invention is to provide a process to increase the refolding yield of a mammalian enterokinase analogue.

Another object of the invention is to provide a mammalian enterokinase analogue with improved refolding yield.

SUMMARY

In one aspect, the invention provides an improved production process for obtaining enterokinase light chain analogues.

In a further aspect, the invention relates to a method for improving production yield of an enterokinase light chain analogue produced in inclusion body. Also or alternatively, the invention relates to a method for improving the refolding yield of an enterokinase light chain analogue produced in inclusion body.

Also or alternatively, the invention provides a fusion protein comprising an enterokinase light chain analogue and a fusion tag selected from Dsb family members with signal sequence deleted; or from Dsb family members not comprising signal sequence.

In another aspect, the invention also relates to an enterokinase light chain analogue that comprises mutants that facilitate refolding of the enterokinase light chain analogue.

In another aspect, the invention provides a method for recombinantly producing a peptide or protein in a bacterial or yeast host cell. Also or alternatively, the method comprises using the enterokinase light chain analogue of the present invention or produced by the method of the present invention.

In one aspect of the invention, a bovine enterokinase light chain analogue is obtained which comprises at least one substitution in position 209 from hydrophobic to a

hydrophilic charged amino acid(s). In one aspect, the bovine enterokinase light chain analogue according to the invention further comprises a substitution in position 112. In one aspect of the invention, a bovine enterokinase light chain analogue is obtained which comprises at least one substitution in position 134 and/or 135 from hydrophobic to a hydrophilic charged amino acid(s).

BRIEF DESCRIPTION OF DRAWINGS

Figure 1: Flowchart for EK refolding process.

Figure 2: Comparison of the activities of EK_L(C112A, L134K, I135K) and EK_L(C112A, L134K, I135K, Y209R) with EK_L at concentration of 0.1 µg/ml). The activity of EK_L (column A) was defined as 100%. The relative activities of EK_L(C112A, L134K and I135K) (column B) and EK_L(C112A, L134K, I135K and Y209R) (column C) were 107% and 97%, respectively.

Figure 3: Comparison of refolding yield of EK_{LM} with different fusion tags. **A** refers to Thrx; **B** refers to DsbA; **C** refers to ΔssDsbA; **D** refers to ΔssDsbC; **E** refers to ΔssDsbG; **F** refers to GST; **G** refers to NRDH-Redoxin; **H** refers to ERp19; **I** refers to hAdrenodoxin; **J** refers to hGlutaredoxin; **K** refers to hPDI; and **L** refers to streptococcol-glutaredoxin.

Figure 4: Screening for mutants to improve refolding efficiency. Many substitutions of EK_L were systemically screened. Y209R was identified that can further increase the refolding yield. "Control" is an EK_L with a substitute of C112A. The other tested EK_L analogues were EK_L with the substitute of C112A and the substitute indicated under each bar.

Figure 5: Confirming mutants that can improve refolding efficiency and the synergistic effect of mutants with fusion tags. **A** refers to ΔssDsbA-EK_L(C112A, L134K, I135K, and L213K); **B** refers to ΔssDsbA-EK_L(C112A, L134K, and I135K); **C** refers to ΔssDsbA-EK_L(C112A, L134K, I135K, and Y209R); **D** refers to ΔssDsbC-EK_L(C112A, L134K, I135K, and L213K); **E** refers to ΔssDsbC-EK_L(C112A, L134K, and I135K); **F** refers to ΔssDsbC-EK_L(C112A, L134K, I135K, and Y209R); **G** refers to ΔssDsbG-EK_L(C112A, L134K, I135K, and L213K); **H** refers to ΔssDsbG-EK_L(C112A, L134K, and I135K); **I** refers to ΔssDsbG-EK_L(C112A, L134K, I135K, and Y209R); and **J** refers to ΔssDsbA-EK_L(C112A and Y209R).

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NO: 1 gives the amino acid sequence of full length bovine enterokinase light chain.

5 SEQ ID NO: 2 gives the amino acid sequence of bovine enterokinase light chain analogue EK_L(C112A, L134 K, I135K).

SEQ ID NO: 3 gives the amino acid sequence of bovine enterokinase light chain analogue EK_L(C112A, L134 K, I135K, Y209R).

10 SEQ ID NO: 4 gives the amino acid sequence of the Linker that is located between the fusion tag and the enterokinase light chain analogues in the fusion proteins prepared in the Examples.

SEQ ID NO: 5 gives the amino acid sequence of the fusion protein consisting of ΔssDsbA, Linker and EK_L(C112A, L134 K, I135K, Y209R).

SEQ ID NO: 6 gives the amino acid sequence of DsbA.

SEQ ID NO: 7 gives the amino acid sequence of DsbC.

15 SEQ ID NO: 8 gives the amino acid sequence of DsbG.

SEQ ID NO: 9 gives the amino acid sequence of Trx.

SEQ ID NO: 10 gives the amino acid sequence of hNRDH redoxin.

SEQ ID NO: 11 gives the amino acid sequence of hPDI.

SEQ ID NO: 12 gives the amino acid sequence of GST.

20 SEQ ID NO: 13 gives the amino acid sequence of Erp19.

SEQ ID NO: 14 gives the amino acid sequence of hGlutaredoxin.

SEQ ID NO: 15 gives the amino acid sequence of Streptococcol-glutaredoxin.

SEQ ID NO: 16 gives the amino acid sequence of hAdrenodoxin.

SEQ ID NO: 17 gives the amino acid sequence of a cleavage site.

25 **DESCRIPTION**

The present invention is related to a production process for obtaining enterokinase light chain analogues.

In a further aspect, the invention relates to a method for improving the production yield of an enterokinase light chain analogue produced in inclusion body.

30 Also or alternatively, the invention relates to a method for improving refolding an enterokinase light chain analogue of the invention.

Also or alternatively, the invention relates to a method for increasing the amount of correctly refolded enterokinase light chain analogue produced in inclusion body.

In a further aspect, the invention provides a method for producing an enterokinase light chain analogue by using a fusion tag to express enterokinase light chain analogue in a fusion protein, in which the fusion tag is selected from Dsb family members with signal sequence deleted. In a further aspect of the invention, the fusion tag is selected from DsbA, DsbC or DsbG, with signal sequence deleted. In another or alternative aspect of the invention, the fusion tag is selected from Dsb family members that do not comprise a signal sequence. In a further aspect of the invention, the fusion tag is selected from NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.

The term "fusion protein" as used herein is meant to refer to a protein created through genetic engineering from two or more proteins or peptides. As used herein, a fusion protein can comprise a fusion tag and an enterokinase light chain analogue. In a further aspect of the invention, a fusion protein as used herein can refer to a protein in which a Asp-Asp-Asp-Lys (D4K) sequence has been intentionally introduced for specific cleavage.

Dsb (disulphide bond) system/family has been known and studied. Components of the system catalyse the formation of disulfide bridges, a process that is crucial for protein structure stabilization and activity. Dsb family in E. coli comprises several enzymes that form two distinct pathways in the periplasm of the bacteria, which all comprise signal sequences. One of the two pathways is 'oxidative pathway', which introduces disulphide bonds into substrate proteins. E. coli-secreted proteins show a strong bias for even numbers of cysteine residues and most contain a single pair of cysteines. For proteins with more than two cysteines, incorrect disulphide bonds could be formed. The other pathway, 'isomerization pathway', reshuffles these non-native disulphide bonds, thereby recuing misfolded substrates. The DSB 'oxidative pathway' in E. coli comprises DsbA and DsbB. The Dsb 'isomerization pathway' in E. coli comprises DsbC and DsbG. Besides the Dsb family enzymes in periplasm of E. coli, there are many other enzymes that catalyse the formation of disulfide bridges in mammalian cell, other bacteria, or the cytoplasm of E. coli, such as NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, E coli Glutaredoxin like NRDH redoxin.

DsbA is responsible for the introduction of disulphide into the newly synthesized proteins that are translocated to the periplasm. It is known that DsbA introduces disulphides into the substrate protein predominantly between consecutive cysteine residues. Thus, in proteins with multiple cysteines and non-consecutive disulphide bonds, this can lead to the formation of incorrect disulphides. See Begoña Heras et al. (2009). Nature Reviews

Microbiology 7, 215-225 (March 2009). Since DSB system works in the periplasm of *E. coli*, DsbA has been used for expressing proteins by secretion process.

According to another aspect, the invention also relates to an enterokinase light chain analogue that comprises mutants (e.g. the one or more substitutions) to facilitate refolding of
5 the enterokinase light chain analogue.

Also or alternatively, the mutants (e.g. the one or more substitutions) in bovine enterokinase light chain analogues of the invention can have synergistic effect with fusion tags to facilitate refolding of the enterokinase light chain analogue.

Mammalian enterokinases are carbohydrate containing heterodimers with a heavy
10 chain of 650-800 amino acids and a catalytic light chain of around 235 amino acids and an overall homology of 75-80% (Liepniecks et al., J. Biol. Chem. 254, 1677 (1979), Matsushima et al., J. Biol. Chem. 269 (31), 19976 (1994), Kitamoto et al., Biochemistry 34, 4562 (1995) for bovine, porcine and human enterokinase, respectively). The term "bovine enterokinase" as used herein means the bovine enterokinase enzyme whose structure and properties are
15 well-known. Further studies of the catalytic light chains are reported in LaVallie et al., J. Biol. Chem. 268 (31), 23311-17 (1993) on the bovine EK and in Matsushima et al., J. Biochem. 125, 947, (1999) on the porcine EK.

"An enterokinase light chain" according to the invention is herein to be understood as bovine enterokinase light chain or an enterokinase light chain from another species such
20 as porcine or human enterokinase light chain.

The term "bovine enterokinase light chain" as used herein means the light chain of bovine enterokinase having 4 disulphide bridges. The bovine enterokinase light chain is e.g. described in LaVallie et al, above.

The term "wild type enterokinase light chain" as used herein (e.g. as "wild type
25 bovine enterokinase light chain") is intended to mean an enterokinase light chain before any substitutions according to the invention have been applied thereto. Similarly, "wild type bovine enterokinase light chain" as used herein is intended to mean bovine enterokinase light chain, as described e.g. in LaVallie et al, above.

The term "enterokinase light chain analogue" or "bovine enterokinase light chain
30 analogue" as used herein means a modified bovine enterokinase light chain with enterokinase activity, wherein one or more amino acid residues of the enterokinase light chain have been substituted by other amino acid residues and/or wherein one or more amino acid residues have been deleted from the enterokinase light chain and/or wherein one or more amino acid residues have been added and/or inserted to the enterokinase light chain.

In one embodiment an enterokinase light chain analogue comprises less than 10 amino acid modifications (substitutions, deletions, additions (including insertions) and any combination thereof) relative to bovine enterokinase light chain, alternatively less than 9, 8, 7, 6, 5, 4, 3 or 2 modifications relative to bovine enterokinase light chain. In one aspect an enterokinase light chain analogue comprises 5 amino acid modifications, in one aspect 4 amino acid modifications, in one aspect 3 amino acid modifications, in one aspect 2 amino acid modifications and in one aspect 1 amino acid modification relative to bovine enterokinase light chain.

Modifications in the enterokinase molecule light chain are denoted stating the position and the one or three letter code for the amino acid residue substituting the native amino acid residue. Using the one letter codes for amino acids, terms like 112A, 134K, 135K and 209R designates that the amino acid in position 112, 134, 135 and 209, respectively, is A, K or R. Using the three letter codes for amino acids, the corresponding expressions are 112Ala, 134Lys, 135Lys and 209Arg, respectively. Thus, e.g., 112Ala,134Lys,135Lys, 209Arg bovine enterokinase light chain is an analogue of bovine enterokinase light chain where the amino acid in position 112 is substituted with alanine, the amino acid in position 134 is substituted with lysine, the amino acid in position 135 is substituted with lysine, and the amino acid in position 209 is substituted with Arginine.

Herein, the term "amino acid residue" is an amino acid from which, formally, a hydroxy group has been removed from a carboxy group and/or from which, formally, a hydrogen atom has been removed from an amino group.

Examples of bovine enterokinase light chain analogues are such wherein Cys in position 112 may be substituted with a number of amino acids including Ala and Ser; Leu in position 134 is substituted with Lys or another charged amino acid; at position 135 where Ile is substituted with Lys or another charged amino acid; or at position 209 where Tyr is substituted with Arg or another charged amino acid.

Further examples of bovine enterokinase light chain analogues according to the invention include, without limitation: 112Ala,134Lys bovine enterokinase light chain; 112Ala,135Lys bovine enterokinase light chain; 112Ala, 209Arg bovine enterokinase light chain; 112Ala,134Lys,135Lys bovine enterokinase light chain; 112Ala,134Lys,135Lys, 209Arg bovine enterokinase light chain; and any such combinations including substitutions with other charged amino acids.

The bovine enterokinase light chain analogue described by the present invention, maintains enterokinase wild type catalytic activity for use as a restriction proteases to specifically cleave fusion proteins.

In one aspect, a bovine enterokinase light chain analogue of the invention has full enterokinase activity compared to wild type bovine enterokinase. In one aspect, a bovine enterokinase light chain analogue of the invention has a substantially equivalent functional or biological activity as wild type bovine enterokinase. For example, a bovine enterokinase light chain analogue has substantially equivalent functional or biological activities (i.e., is a functional equivalent) of the polypeptide having the amino acid sequence set forth as SEQ ID NO: 1 (e.g., has a substantially equivalent enteropeptidase activities).

Also or alternatively, the mutants introduced to EK_L were designed not near or inside the catalytic domain of EK_L and the mutants are not expected to alter the protein conformation. Thus, the catalytic activities of these analogues should be comparable to EK_L .

The term "protease" is intended to include any polypeptide/s, alone or in combination with other polypeptides, that break peptide bonds between amino acids of proteins.

The term "catalytic activity" or "proteolytic activity" is meant to refer to the cleavage activity of a substrate by an enzyme. In particular embodiments, the term refers to the enzymatic cleavage by enteropeptidases, or the enterokinase activity.

As used herein, enterokinase activity means the capability of cleaving peptide or protein substrates at a specific site; for protein substrates, this is generally following the sequence (Asp)₄-Lys, or a similar sequence such as those described in Light et al., Anal. Biochem. 106: 199(1980); (a cluster of negatively charged amino acids followed by a positively charged amino acid). Typically, such activity is measured by activation of trypsinogen by cleaving the N-terminal propeptide (containing (Asp)₄-Lys) with the enterokinase or enterokinase analogue and subsequently assaying the amount of active trypsin generated using tosyl-arginine-methylester (TAME). Alternatively, enterokinase activity can be measured directly by incubating the enzyme with the peptide substrate Gly (Asp)₄-Lys-ss-naphthylamide and measuring the increase in fluorescence (excitation at 337 nm, emission at 420 nm) generated by cleavage and release of the ss-NA (ss-naphthylamide) moiety. See, e.g., Grant et al., Biochem. Biophys. Acta. 567:207(1979). Bovine enterokinase is also active on some trypsin substrates like TAME and BAEE (benzyl-arginine-ethyl-ester).

In one aspect, less aggregation during the refolding process of a bovine enterokinase light chain analogue according to the invention is obtained relative to the aggregation obtained during the refolding process of wild type bovine enterokinase light chain. In one aspect, a bovine enterokinase light chain analogue according to the invention is

more soluble or has less precipitation during the refolding process relative to wild type bovine enterokinase light chain.

Also or alternatively, the enterokinase light chain analogue of the invention incorporated as one of the fusion protein partners to yet another protein, such as a fusion tag, was released by the addition of a minimal amount of exogenous enterokinase. The released enterokinase light chain analogues in turn can catalyze many more proteolytic cleavages of fusion proteins. In this way, large amounts of enterokinase activity can be produced from a fusion protein.

In one embodiment, the protein comprising is an enterokinase light chain analogue of the invention is a fusion protein. In another embodiment, the fusion protein is a recombinant fusion protein. In a further embodiment, the protein is bacterially produced. In a more particular embodiment, the protein is a synthetic protein.

The method for refolding a bovine enterokinase light chain analogue according to the invention may be carried out by denaturation in urea, followed by oxidative refolding in glutathione or another re-dox environment.

Also or alternatively, the invention relates to a method for increasing the concentration of solubilized fusion proteins comprising enterokinase light chain analogue in refolding buffer to obtain correctly refolded enterokinase light chain analogue of the invention.

In one aspect of the invention, the fusion tag is removed during refolding, i.e. during dilution and incubation under refolding conditions. It has thus been found that refolding and activation may be obtained without addition of an activation enzyme (this may also be referred to herein as auto-activation). In one aspect of the invention, the linker connecting the fusion tag and the bovine enterokinase light chain analogue of the invention is removed by autocleavage (may also be referred to herein as auto-activation).

Also or alternatively, the invention provides a fusion protein comprising a fusion tag, an enterokinase light chain analogue, and a linker. In a further aspect of the invention, the linker comprises the sequence (Asp)₄-Lys, or a similar sequence such as those described in Light et al., Anal. Biochem. 106: 199(1980); (a cluster of negatively charged amino acids followed by a positively charged amino acid). In a further aspect of the invention, the linker has the amino acid sequence of SEQ ID NO: 4.

Nucleic acid forms encoding enterokinase light chain analogues of the present invention are also within the scope of the invention. Nucleic acids according to the invention include genomic DNA (gDNA), complementary DNA (cDNA), synthetic DNA prepared by chemical synthesis as well as DNA with deletions or substitutions, allelic variants and

sequences that hybridize thereto under stringent conditions as long as they encode enterokinase light chain analogues of the present invention.

In one embodiment a nucleic acid is provided wherein said nucleic acid comprises a polynucleotide sequence, and wherein said nucleic acid encodes a mammalian enterokinase light chain analogue such as a bovine enterokinase light chain analogue according to the invention. In one embodiment, the nucleic acid is operably linked to an inducible promoter. In one embodiment, a recombinant vector is provided which comprises the nucleic acid operably linked to the inducible promoter. In one embodiment, the inducible promoter is selected from a group consisting of AraB, T7, trp, lac, and tac.

A further embodiment of the invention provides a host cell comprising the recombinant vector comprising the polynucleotide sequence coding for the amino acid sequence of a mammalian enterokinase light chain analogue such as a bovine enterokinase light chain analogue according to the invention.

A further aspect of the invention provides the host cell comprising the recombinant vector comprising the polynucleotide sequence coding for the amino acid sequence encoding a mammalian enterokinase light chain analogue such as a bovine enterokinase light chain analogue according to the invention. In one embodiment, the host cell is selected from a group consisting of *E.coli*, *B.subtilis*, *S.saccharomyces* and *A.oryzae*.

The production of polypeptides, e.g., enterokinase light chain, is well known in the art. The bovine enterokinase light chain analogue may for instance be produced by classical peptide synthesis, e.g., solid phase peptide synthesis using t-Boc or Fmoc chemistry or other well established techniques, see, e.g., Greene and Wuts, "Protective Groups in Organic Synthesis", John Wiley & Sons, 1999. The bovine enterokinase light chain analogue may also be produced by a method which comprises culturing a host cell containing a DNA sequence encoding the analogue and capable of expressing the bovine enterokinase light chain analogue in a suitable nutrient medium under conditions permitting the expression of the bovine enterokinase light chain analogue. Several recombinant methods may be used in the production of bovine enterokinase light chain and bovine enterokinase light chain analogues. Examples of methods which may be used in the production of enterokinase in microorganisms such as, e.g., *Escherichia coli* and *Saccharomyces cerevisiae* are, e.g., disclosed in WO 94/16083.

Typically, the bovine enterokinase light chain analogue is produced by expressing a DNA sequence encoding the bovine enterokinase light chain analogue or a precursor thereof in a suitable host cell by well known technique as disclosed in e.g. WO 94/16083.

The bovine enterokinase light chain analogues of the invention may be recovered from the cell culture medium or from the cells. The bovine enterokinase light chain analogues of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

In one aspect, the bovine enterokinase light chain analogues of the present invention are purified using anion exchange chromatography. In a further aspect, the anion exchange chromatography is followed by hydrophobic interaction chromatography. In one aspect, the bovine enterokinase light chain analogues of the present invention are purified using Q HP anion exchange chromatography. In a further aspect, the Q HP anion exchange chromatography is followed by Phenyl FF hydrophobic interaction chromatography.

In one aspect of the present invention an improved process for production of a mammalian enterokinase light chain analogue such as a bovine enterokinase light chain analogue is provided, wherein said method comprises the steps:

a) culturing the host cells in a growth medium comprising inducer, wherein the host cells comprise a polynucleotide sequence encoding the amino acid sequence of the enterokinase light chain analogue and a fusion tag, wherein the fusion tag is selected from Dsb family members with signal sequence deleted or from Dsb family members that do not comprise signal sequence;

b) recovering the cells with enterokinase light chain analogue in inclusion bodies

c) solubilizing and refolding the enterokinase light chain analogue; and

d) purifying the enterokinase light chain analogue.

The invention provides a new recombinant process for production of mammalian enterokinase light chain analogue such as a bovine enterokinase light chain analogue in *E. coli* in a very efficient and economic way.

The expression of a bovine enterokinase light chain analogue according to the invention may e.g. be localized in the inclusion bodies of *E. coli* or in the secreted material of yeast. In one embodiment expression of enterokinase is localized in the inclusion bodies of *E. coli*.

Various strains of *E. coli* are useful as host cells for the production of non-glycosylated, homogeneous enterokinase activity are also well-known in the art. A non-exclusive list of such strains includes *E. coli* B BL21 DE3, *E. coli* K12 W3110, MC1061, DH1,

K803, HB101, JM101 and other K12 like strains. Alternatively, other bacterial species may be used, including *B. subtilis*, various strains of *Pseudomonas*, other bacilli and the like.

Many strains of yeast cells, known to those skilled in the art, are also available as host cells for expression of the enterokinase activity of the present invention. Yeast cells are especially useful as a host for pre/pro fusion to mature enterokinase. When expressed using
5 a suitable yeast vector, the fusion is secreted by virtue of a signal peptide. In a further aspect, the fusion tag is selected from Dsb family with signal sequence.

When the bovine enterokinase light chain analogue of this invention is expressed in bacterial cells, it may be expressed intracellularly usually as inclusion bodies, or it may be
10 secreted from bacterial cells in active form if a secretory signal is included. Where necessary or desired, the enterokinase activity may be obtained by conventional methods such as solubilization of protein in urea or guanidine HCl, followed by dilution to reduce the concentration of these reagents and treatment with oxidizing agents such as dithiothreitol or ss-mercapto ethanol to enhance refolding.

15 In one aspect of the invention, a process for preparing a bovine enterokinase light chain analogue in *E. coli* cells is obtained, wherein the *E. coli* cells are transformed with a plasmid carrying the bovine enterokinase light chain analogue gene and an inducible promoter by fermentation involving batch and fed batch stages and isolation and purification of the expressed protein from the cultures.

20 In one aspect of the invention, a refolding process for a bovine enterokinase light chain analogue according to the invention is obtained, wherein the expression of the enterokinase light chain analogue is in the form of inclusion bodies in recombinant *E. coli*. In one embodiment denaturation followed by refolding in a redox system is used.

The enterokinase light chain analogues of the invention may be used in a method
25 for cleaving proteins having an enterokinase cleavage site, and especially fusion proteins having such a cleavage site engineered into their sequence. The amounts needed are readily determined empirically by one skilled in the art.

Another particular aspect of the invention teaches a method for cleavage of a protein containing an Asp-Asp-Asp-Asp-Lys cleavage site using any of the bovine
30 enterokinase light chain analogues of the invention described herein, the method comprising contacting the protein with any of the bovine enterokinase light chain analogues of the invention, and wherein the contacting of the protein with the bovine enterokinase light chain analogue results in specific cleavage.

In a further aspect, the invention teaches a method for the preparation of
35 recombinant protein using any of the bovine enterokinase light chain analogues according to

the invention as described herein, the method comprising providing a recombinant fusion protein containing a Asp-Asp-Asp-Asp-Lys cleavage site, and contacting the fusion protein with any of the bovine enterokinase light chain analogues of the invention, wherein contacting the recombinant fusion protein with the bovine enterokinase light chain analogue results in Asp-Asp-Asp-Asp-Lys specific cleavage and preparation of recombinant protein.

Also, the fusion protein of the invention may comprise a fusion tag selected from Dsb family members with signal sequence deleted or from Dsb family members not comprising signal sequence, an enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain.

The enterokinase light chain of the fusion protein of the invention may be derived from a mammalian species, or an analogue thereof. The mammalian species may be bovine, human or porcine.

Also, the Dsb family fusion tag may be DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.

Also, the enterokinase light chain may be a bovine enterokinase light chain (SEQ ID NO:1). Alternatively, the enterokinase light chain may be a bovine enterokinase light chain analogue comprises a substitution in position 112 of SEQ ID NO:1; wherein optionally the substitution in position 112 is C112A.

Also or alternatively, the enterokinase light chain may be a bovine enterokinase light chain analogue comprises a substitution in position 209 of SEQ ID NO:1; wherein optionally the substitution is Y209R.

Also, the bovine enterokinase light chain analogue may further comprise substitutions in positions 134 and 135 of SEQ ID NO:1, wherein optionally the substitutions in positions 134 and 135 are L134K and I135K.

Also, the linker of the fusion protein may comprise a cleavage site, wherein the amino acid sequence of the cleavage site is Asp-Asp-Asp-Asp-Lys.

Also or alternatively, the fusion protein of the invention may comprise a DsbA tag with signal sequence deleted, a linker and a bovine enterokinase light chain analogue with substitutions of C112A, L134K, I135K, and Y209R.

Also or alternatively, the fusion protein of the invention may comprise the amino acid sequence shown in SEQ ID NO:5.

In one aspect the invention provides a bovine enterokinase light chain analogue comprising a substitution in position 209 of SEQ ID NO:1, wherein optionally the substitution in position 209 is Y209R. In a further aspect the bovine enterokinase light chain analogue

may further comprising a substitution in position 112, wherein optionally the substitution in position 112 is C112A. In a further aspect the bovine enterokinase light chain analogue may further comprise substitutions in positions 134 and 135, wherein optionally the substitutions in positions 134 and 135 are L134K and I135K.

5 In one aspect the invention provides a bovine enterokinase light chain analogue, wherein the analogue comprising substitutions of C112A, L134K, I135K, and Y209R, wherein the amino acid sequence of the analogue may be as shown in SEQ ID NO: 3.

In another aspect the invention provides a method for production of an enterokinase light chain, comprising the steps: a) culturing host cells comprising a polynucleotide
10 sequence encoding a fusion protein, wherein the fusion protein comprising a fusion tag selected from Dsb family members with signal sequence deleted or from Dsb family members not comprising signal sequence, the enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain; b) inducing expression of the fusion protein in inclusion bodies; c) solubilizing, refolding the fusion protein; and d) purifying the
15 enterokinase light chain or the analogue thereof.

In yet another aspect the invention provides a method for recombinantly producing a protein in a bacterial or yeast host cell, comprising: a) expressing in yeast or bacteria a fusion protein comprising the protein to be produced; b) cleaving the fusion protein with a bovine enterokinase light chain analogue as defined herein; and c) purifying the produced
20 protein; wherein optionally the fusion protein expressed in step a) further comprises an Asp-Asp-Asp-Lys cleavage site.

In a further aspect of the methods, the Dsb family fusion tag may be DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.
25 In a further aspect of the methods, the activating is achieved through auto-activation. In a further aspect of the methods, the linker comprises a cleavage site of Asp-Asp-Asp-Lys. In a further aspect of the method, the host cell is E. coli. In a further aspect of the methods, the protein to be produced is a GLP-1 peptide, a GLP-1 analogue, an insulin, or an insulin analogue. In a further aspect, the methods of the invention may encompass the fusion
30 protein as defined herein.

The terms "aspect" and "embodiment" may be used interchangeably herein.

The following is a non-limiting list of aspects according to the invention:

Aspect 1. A fusion protein, comprising a fusion tag selected from Dsb family
35 members with signal sequence deleted or from Dsb family members not comprising signal

sequence, an enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain.

Aspect 2. The fusion protein of aspect 1, wherein the enterokinase light chain is derived from a mammalian species, or an analogue thereof.

5 Aspect 3. The fusion protein of aspect 2, wherein the mammalian species is bovine, human or porcine.

Aspect 4. The fusion protein of aspect 1, wherein the Dsb family fusion tag is DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.

10 Aspect 5. The fusion protein of aspect 1, wherein the enterokinase light chain is a bovine enterokinase light chain (SEQ ID NO:1).

Aspect 6. The fusion protein of aspect 1, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 112 of SEQ ID NO:1.

15 Aspect 7. The fusion protein of aspect 6, wherein the substitution in position 112 is C112A.

Aspect 8. The fusion protein of aspect 1, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 209 of SEQ ID NO:1.

20 Aspect 9. The fusion protein of aspect 8, wherein the substitution is Y209R.

Aspect 10. The fusion protein of any one of aspects 6 to 9, wherein the bovine enterokinase light chain analogue further comprising substitutions in positions 134 and 135 of SEQ ID NO:1.

25 Aspect 11. The fusion protein of aspect 10, wherein the substitutions in positions 134 and 135 are L134K and I135K.

Aspect 12. The fusion protein of aspect 1, wherein the linker comprises a cleavage site, wherein the amino acid sequence of the cleavage site is Asp-Asp-Asp-Asp-Lys.

30 Aspect 13. A fusion protein, comprising a DsbA tag with signal sequence deleted, a linker and a bovine enterokinase light chain analogue with substitutions of C112A, L134K, I135K, and Y209R.

Aspect 14. A fusion protein, wherein the amino acid sequence of the fusion protein is shown in SEQ ID NO:5.

Aspect 15. A bovine enterokinase light chain analogue, comprising a substitution in position 209 of SEQ ID NO:1.

Aspect 16. The bovine enterokinase light chain analogue of aspect 15, wherein the substitution in position 209 is Y209R.

Aspect 17. The bovine enterokinase light chain analogue of aspect 15, wherein the analogue further comprising a substitution in position 112.

5 Aspect 18. The bovine enterokinase light chain analogue of aspect 17, wherein the substitution in position 112 is C112A.

Aspect 19. The bovine enterokinase light chain analogue of aspect 17, wherein the analogue further comprising substitutions in positions 134 and 135.

10 Aspect 20. The bovine enterokinase light chain analogue of aspect 19, wherein the substitutions in positions 134 and 135 are L134K and I135K.

Aspect 21. A bovine enterokinase light chain analogue, wherein the analogue comprising substitutions of C112A, L134K, I135K, and Y209R.

Aspect 22. The bovine enterokinase light chain analogue of aspect 21, wherein the amino acid sequence of the analogue is show SEQ ID NO: 3.

15 Aspect 23. A method for production of an enterokinase light chain, comprising the steps:

- a) culturing host cells comprising a polynucleotide sequence encoding a fusion protein, wherein the fusion protein comprising a fusion tag selected from Dsb family members with signal sequence deleted or from Dsb family members not comprising
- 20 signal sequence, the enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain;
- b) inducing expression of the fusion protein in inclusion bodies;
- c) solubilizing, refolding the fusion protein; and
- d) purifying the enterokinase light chain or the analogue thereof.

25 Aspect 24. The method of aspect 23, wherein the Dsb family fusion tag is DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.

Aspect 25. The method of aspect 23, wherein the fusion protein is according to any one of aspects 1-12.

30 Aspect 26. The method of aspect 23, wherein the activating is achieved through auto-activation.

Aspect 27. The method of aspect 23, wherein the linker comprises a cleavage site of Asp-Asp-Asp-Asp-Lys.

Aspect 28. The method of aspect 23, wherein host cell is E. coli.

Aspect 29. A method for recombinantly producing a protein in a bacterial or yeast host cell, comprising:

- a) expressing in yeast or bacteria a fusion protein comprising the protein to be produced;
- 5 b) cleaving the fusion protein with a bovine enterokinase light chain analogue according to any one of aspects 15-22; and
- c) purifying the produced protein.

Aspect 30. The method for recombinantly producing a protein according to aspect 29, wherein the fusion protein expressed in step a) further comprises an Asp-Asp-Asp-Asp-
10 Lys cleavage site.

Aspect 31. The method for recombinantly producing a protein according to aspect 29 or aspect 30, wherein the host cell is *E. coli*.

Aspect 32. The method for recombinantly producing a protein according to any one of aspects 29-31, wherein the protein to be produced is a GLP-1 peptide, a GLP-1 analogue,
15 an insulin, or an insulin analogue.

The following is a non-limiting list of further aspects according to the invention:

Aspect 33. A fusion protein, comprising a fusion tag selected from Dsb family members with signal sequence deleted or from Dsb family members not comprising signal
20 sequence, an enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain.

Aspect 34. The fusion protein of aspect 33, wherein the enterokinase light chain is derived from a mammalian species, or an analogue thereof.

Aspect 35. The fusion protein of aspect 34, wherein the mammalian species is
25 bovine, human or porcine.

Aspect 36. The fusion protein of any one of aspects 33-35, wherein the Dsb family fusion tag is DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or *E. coli* Glutaredoxin like NRDH redoxin.

Aspect 37. The fusion protein of any one of aspects 33-36, wherein the enterokinase
30 light chain is a bovine enterokinase light chain (SEQ ID NO:1).

Aspect 38. The fusion protein of any one of aspects 33-37, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 112 of SEQ ID NO:1.

Aspect 39. The fusion protein of aspect 38, wherein the substitution in position 112 is C112A.

Aspect 40. The fusion protein of any one of aspects 33-39, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 209 of SEQ ID NO:1.

Aspect 41. The fusion protein of aspect 40, wherein the substitution is Y209R.

Aspect 42. The fusion protein of any one of aspects 33-41, wherein the bovine enterokinase light chain analogue further comprising substitutions in positions 134 and 135 of SEQ ID NO:1.

Aspect 43. The fusion protein of aspect 42, wherein the substitutions in positions 134 and 135 are L134K and I135K.

Aspect 44. The fusion protein of any one of aspects 33-43, wherein the linker comprises a cleavage site, wherein the amino acid sequence of the cleavage site is Asp-Asp-Asp-Lys.

Aspect 45. A fusion protein, comprising a DsbA tag with signal sequence deleted, a linker and a bovine enterokinase light chain analogue with substitutions of C112A, L134K, I135K, and Y209R.

Aspect 46. A fusion protein, wherein the amino acid sequence of the fusion protein is shown in SEQ ID NO:5.

Aspect 47. A bovine enterokinase light chain analogue, comprising a substitution in position 209 of SEQ ID NO:1.

Aspect 48. The bovine enterokinase light chain analogue of aspect 47, wherein the substitution in position 209 is Y209R.

Aspect 49. The bovine enterokinase light chain analogue of aspect 47 or 48, wherein the analogue further comprising a substitution in position 112.

Aspect 50. The bovine enterokinase light chain analogue of aspect 49, wherein the substitution in position 112 is C112A.

Aspect 51. The bovine enterokinase light chain analogue of any one of aspects 47-50, wherein the analogue further comprises substitutions in positions 134 and 135.

Aspect 52. The bovine enterokinase light chain analogue of aspect 51, wherein the substitutions in positions 134 and 135 are L134K and I135K.

Aspect 53. A bovine enterokinase light chain analogue, wherein the analogue comprising substitutions of C112A, L134K, I135K, and Y209R.

Aspect 54. The bovine enterokinase light chain analogue of aspect 53, wherein the amino acid sequence of the analogue is SEQ ID NO: 3.

Aspect 55. A method for production of an enterokinase light chain, comprising the steps: a) culturing host cells comprising a polynucleotide sequence encoding a fusion protein, wherein the fusion protein comprising a fusion tag selected from Dsb family members with signal sequence deleted or from Dsb family members not comprising signal sequence, the enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain; b) inducing expression of the fusion protein in inclusion bodies; c) solubilizing, refolding the fusion protein; and d) purifying the enterokinase light chain or the analogue thereof.

Aspect 56. The method of aspect 55, wherein the Dsb family fusion tag is DsbA, DsbC, or DsbG, with signal sequence deleted; or is NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, or E. coli Glutaredoxin like NRDH redoxin.

Aspect 57. The method of aspect 55 or 56, wherein the fusion protein is as defined in any one of aspects 33-46.

Aspect 58. The method of any one of aspects 55-57, wherein the activating is achieved through auto-activation.

Aspect 59. The method of any one of aspects 55-58, wherein the linker comprises a cleavage site of Asp-Asp-Asp-Asp-Lys.

Aspect 60. The method of any one of aspects 55-59, wherein host cell is E. coli.

Aspect 61. A method for recombinantly producing a protein in a bacterial or yeast host cell, comprising: a) expressing in yeast or bacteria a fusion protein comprising the protein to be produced; b) cleaving the fusion protein with a bovine enterokinase light chain analogue as defined in any one of aspects 47-54; and c) purifying the produced protein.

Aspect 62. The method for recombinantly producing a protein according to aspect 61, wherein the fusion protein expressed in step a) further comprises an Asp-Asp-Asp-Asp-Lys cleavage site.

Aspect 63. The method for recombinantly producing a protein according to aspect 61 or 62, wherein the host cell is E. coli.

Aspect 64. The method for recombinantly producing a protein according to any one of aspects 61-63, wherein the protein to be produced is a GLP-1 peptide, a GLP-1 analogue, an insulin, or an insulin analogue.

EXAMPLES

Abbreviations:

EK: enterokinase

EK_L: wild type bovine enterokinase light chain

EK_L analogue: EK_L with mutation(s), e.g., substitutions. For example, the EK_L analogue could be EK_L(C112A), EK_L(C112A, Y209R), EK_L(C112A, L134 K, I135K), EK_L(C112A, L134 K, I135K, Y209R), etc.

5 ΔssDsbA: DsbA with signal sequence deleted.

ΔssDsbC: DsbC with signal sequence deleted.

ΔssDsbG: DsbG with signal sequence deleted.

Trx: Thioredoxin

GST: glutathione S-transferase

10 ERp19: endoplasmic reticulum protein 19

hPDI: human protein disulfide isomerase

Linker: a linker comprised in the fusion protein between the fusion tag and the EK_L or EK_L analogue.

15 Tag: a fusion tag fused with the N-terminal of EK_L or EK_L analogue. For example, the fusion tag could be one of ΔssDsbA, DsbC, ΔssDsbC, DsbG, ΔssDsbG, GST, NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin, *E coli* Glutaredoxin like NRDH redoxin, etc.

Tag-Linker-EK_L /Tag-Linker-EK_L analogue: EK_L or EK_L analogue fused with an N-terminal fusion tag and the Linker.

20 IPTG: Isopropyl β-D-1-thiogalactopyranoside

Tris: Tris(hydroxymethyl) aminomethane

DTT: Dithiothreitol

GSSG: Glutathione disulfide

GSH: Glutathione

25 FDM: Fermentation defined medium

LC-MS: Liquid chromatography–mass spectrometry

SDS-PAGE: Sodium dodecyl sulfate polyacrylamide gel electrophoresis

PCR reaction: Polymerase chain reaction

30 PEG1000: Polyethylene Glycol 1000, a polyethylene glycol with approximate molecular weight 1000.

Example 1: Plasmid construction for expressing fusion proteins of the invention

The DNA sequence encoding DsbA-Linker-EK_L was amplified and introduced into vector pET39b (Novagen) by homologous recombination to obtain construct pET39b-DsbA-Linker-EK_L.

35

For the constructs with other tags (e.g., DsbC, DsbG, Trx, GST, hNRDH-Redoxin, ERp19, hAdrenodoxin, hGlutaredoxin, hPDI, streptococcol-glutaredoxin), the DNA sequences of the tags were synthesized and replaced the DNA sequence of DsbA in pET39b-DsbA-Linker-EK_L by homologous recombination. To obtain ΔssDsbA, ΔssDsbC and ΔssDsbG, the signal sequences were deleted from DsbA, DsbC and DsbG by using-
 5 QuikChange® XL Site-Directed Mutagenesis Kit (Stratagene).

Mutants in EK_L, such as M48K, L65K, L74K, C112A, I125K, L134K, I135K, V143K, Y209R, Q210R, Q210K etc., were introduced into the constructs by using QuikChange® XL Site-Directed Mutagenesis Kit (Stratagene).

10

Amino acid sequence of EK_L:

IVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGL
 HMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPICLPEENQVFPP
 15 GRICSIAGWGALIYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAGGVDSC
 QGDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLH
 (SEQ ID NO: 1)

Amino acid sequence of EK_L(C112A, L134 K, I135K):

20 IVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGL
 HMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPIALPEENQVFPP
 GRICSIAGWGAKKYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAGGVDS
 CQDSGGPLMCQENNRWLLAGVTSFGYQCALPNRPGVYARVPRFTEWIQSFLH
 (SEQ ID NO: 2)

25 Underlined: mutants at positions 112, 134 and 135

Amino acid sequence of EK_L(C112A, L134 K, I135K, Y209R):

IVGGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGL
 HMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPIALPEENQVFPP
 30 GRICSIAGWGAKKYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAGGVDS
 CQDSGGPLMCQENNRWLLAGVTSFGRQCALPNRPGVYARVPRFTEWIQSFLH
 (SEQ ID NO: 3)

Underlined: mutants at positions 112, 134, 135 and 209

35 **Amino acid sequence of the Linker:**

GS GSGHMH HHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDK
(SEQ ID NO: 4)

Amino acid sequence of Δ ssDsbA-Linker- EK_L(C112A, L134 K, I135K, Y209R):

5 MAQYEDGKQYTTLEKPVAGAPQVLEFFSFFCPHCYQFEEVLHISDNVKKKLPEGVKMTKYH
VNFMGGDLGKDLTQAWAVAMALGVEDKVTVPFLFEGVQKTQTIRSASDIRDVFINAGIKGEEY
DAAWNSFVVKSLVAQQEKAADVQLRGVPAMFVNGKYQLNPQGMDTSNMDV FVQQYADT
VKYLSEKKGSGSGHMH HHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKIV
GGSDSREGAWPWVVALYFDDQQVCGASLVSRDWLVSAAHCVYGRNMEPSKWKAVLGL
10 **HMASNLTSPQIETRLIDQIVINPHYNKRRKNNDIAMMHLEMKVNYTDYIQPIALPEENQVFPP**
GRICSIAGWGA**K****KYQGSTADVLQEADVPLLSNEKCQQQMPEYNITENMVCAGYEAGGVD**
SCQGDSSGGLMCQENNRWLLAGVTSFGRQCALPNRPGVYARVPRFTEWISFLH**
 (SEQ ID NO: 5)**

Underlined: Δ ssDsbA;

15 Regular: the Linker;

Bold: EK_L(C112A, L134 K, I135K, Y209R) (note: underlined and regular: substitutes of C112A, L134K, and I135K)

Amino acid sequence of tags:

20

DsbA

MKKIWLALAGLVLA**FSASAA**QYEDGKQYTTLEKPVAGAPQVLEFFSFFCPHCYQF
 EEVLHISDNVKKKLPEGVKMTKYHVNFMGGDLGKDLTQAWAVAMALGVEDKVTV
 PLFEGVQKTQTIRSASDIRDVFINAGIKGEEYDAAWNSFVVKSLVAQQEKAADVQ
 25 LRGVPAMFVNGKYQLNPQGMDTSNMDV FVQQYADTVKYLSEKK
 (SEQ ID NO: 6)

Bold: signal sequence of DsbA

DsbC

30 **MKKGFMLFTLLAA****FSGFAQ**ADDAAIQQTAKMGIKSSDIQPAPVAGMKTVLTNSG
 VLYITDDGKHIIQGPMYDVSGTAPVNVNKMMLKQLNALEKEMIVYKAPQEKHVITV
 FTDITCGYCHKLHEQMADYNALGITVRYLAFFPRQGLSDAEKEMKAIWCAKDKNK
 AFDDVMAGKSVAPASCDVDIADHYALGVQLGVSGTPAVVLSNGTLVPGYQPPKEM
 KEFLDEHQKMTSGK
 35 (SEQ ID NO: 7)

Bold: signal sequence of DsbC

DsbG

5 **MLKKILLALLPAIAFAEELPAPVKAIEKQGITIIKTFDAPGGMKGYLGKYQDMGVTI**
YLTPDGKHAISGYMYNEKGENLSNTLIEKEIYAPAGREMWQRMEQSHWLLDGKKD
APVIVYVFADPFCPYCKQFWQQARPWVDSGKVQLRTLLVGVIKPESPATAAAILAS
KDKPAKTWQQYEASGGKLLNVPANVSTEQMKVLSKDNEKLMDDLGANVTPAIYYM
SKENTLQQAVGLPDQKTLNIIMGNK

(SEQ ID NO: 8)

10 Bold: signal sequence of DsbG

Trx

MSDKIIHLTDDSFDTDLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVA
KLNIDQNPGTAPKYGIRGIPTLLLKFNGEVAATKVGALSKGQLKEFLDANLA

15 (SEQ ID NO: 9)

hNRDH redoxin

GDHMFNRPNRNDVDDGVQDIQNDVNQLADSLESVLKSWGSDAKGEAEAARSKA
QALLKETRARMHGRTRVQQAARDAVAARILLFVKDPGVAWVQQLR

20 (SEQ ID NO: 10)

hPDI

MDAPEEEDHVLVLRKSNFAEALAAHKYLLVEFYAPWCGHCKALAPEYAKAAGKLLK
AEGSEIRLAKVDATEESDLAQYGVRYPTIKFFRNGDTASPKEYTAGREADDIVN
25 WLKKRTGPAATTLPDGAAAESLVESSEVAVIGFFKDVEDSAKQFLQAAEAIDDIPF
GITSNSDVFVSKYQLDKDGVVLFKKFDEGRNNEFEVTKENLLDFIKHNQLPLVIEFT
EQTAPKIFGGEIKTHILLFLPKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDHTDNQR
ILEFFGLKKEECPAVRLITLLEEMTKYKPESEELTAERITEFCHRFLEGKIKPHLMSQ
ELPEDWDKQPVKVLVGKNFEDVAFDEKKNVFVEFYAPWCGHCKQLAPIWDKLGE
30 TYKDHENIVIAKMDSTANEVEAVKVHSFPTLKFFPASADRTVIDYNGERTLDGFKKF
LESGGQDGAGDDDDLEDLEEAEPPDMEEDDDQKAVKDEL

(SEQ ID NO: 11)

GST

35 MSPILGYWKIKGLVQPTRLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLP

YYIDGDVKLTQSMAlRYIADKHNMLGGCPKERAElSMLEGAVLDIRYGVSRlAYSKD
 FETLKVDfLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMC
 LDAFPKLVCfKKRIEAIPQIDKYLKSSKYIAWPLQGwQATFGGGDHPPKSD
 (SEQ ID NO: 12)

5

Erp19

MHNGLGKGFGDHIHWRTLEDGKKEAAASGLPLMVIIHKSWCGACKALKPKFAEST
 EISELSHNfVMVNLEDEEEPKDEDFSPDGGYIPRILFLDPSGKVHPEIINENGnPSY
 KYFYVSAEQVVQGMKEAQERLTGDAFRKKHLEDEL

10

(SEQ ID NO: 13)

hGlutaredoxin

MAQEFVNCKIQPGKVvVfIKPTCPYCRRAQEILSQLPIKQGLLEFVDITATNHTNEIQ
 DYlQQLTGARTVPRVfIGKDCIGGCSDLVSLQQSGELLTRLKQIGALQ

15

(SEQ ID NO: 14)

Streptococcol-glutaredoxin

MVTVYSKNNCVQCKMTRFLDSNNVAYREINLDEQPEYIDQVKELGFSAAPIIQTPT
 EVfSGfQPGKLKQLA

20

(SEQ ID NO: 15)

hAdrenodoxin

MSSSEDKITVHFInRDGETLTtKGKVGDsLLDVVENNLdIDGfGACEGTLACSTC
 HLIFEDHIYEKLDAITDEENDMLDLAYGLTDRSRLGCQICLTkSMdNMtVRVPETVA
 DARQSIDVGKTS

25

(SEQ ID NO: 16)

Example 2: Expression of fusion proteins as inclusion bodies

E.coli strain BL21 DE3 were transformed with the constructs prepared in Example 1.

30 Cells were grown in shake flask with LB medium at 37°C for 6 hours. Then cells were induced with 1.0mM IPTG at an OD600 of 1.0, and then grew for 6 hours at 37°C before harvesting by centrifugation. The expression bands from different constructs were confirmed by LC-MS. The fusion proteins were expressed as inclusion bodies at similar level, except for pET39b-hPDI-Linker-EK_L(C112A, L134K, I135K,. Moreover, for all the tested constructs,
 35 the expression yield of the fusion proteins was no less than 150mg/L in shake flask.

The expression of Trx-Linker-EK_L(C112A, L134K, I135K) and ΔssDsbA-Linker-EK_L(C112A, L134 K, I135K) was up scaled and tested in fermenters. Cells from a glycerol stock were inoculated on an EC1 plate and grew overnight at 37°C, and then washed with 0.9% sodium chloride to suspend the cells. The culture grew in a fermenter containing fermentation defined medium (FDM) at 37°C for 16 hours. Then cells were induced with 1.0mM IPTG at an OD600 of 150, and then grew for 6 hours at 37°C before harvesting by centrifugation. Based on UPLC analysis, expression yields of 4g/L were achieved for both ΔssDsbA-Linker-EK_L(C112A, L134K, I135K) and Trx-Linker-EK_L(C112A, L134K, I135K).

10 Example 3: Refolding, auto-catalytic activation and purification

Cells from fermentation (constructs being expressed as inclusion bodies) obtained in Example 2 were re-suspended in lysis buffer (1:10, v/v) containing 20mM Tris, pH 8.0, and lysed by French press. Inclusion bodies were precipitated at 20,000g for 1 hour at 4°C, and then washed once by lysis buffer. The inclusion bodies were solubilized at 5 mg/ml in buffer containing 20mM Tris, 8M urea, pH8.0, 20mM DTT and incubated at 4°C for 3 hours. After centrifuged at 20,000g for 30min, the solubilized fusion proteins was diluted 20 fold into refolding buffer containing 20mM Tris, pH 9.0, 1.2 M Urea, 0.5% PEG 1000, 1.5% β-cyclodextrin and 1mM GSSG, 3 mM GSH, 5mM cysteine, 0.1mM cysteine, and then incubated at 4 °C at least for 12 hours for refolding.

During dilution and incubation for refolding, auto-catalytic cleavage occurred, and fully active EK_L or EK_L analogues without fusion tags and Linker was liberated from the fusion proteins by the escaped active enzyme, which specifically cleaved fusion tag off at DDDDK recognition site just before the N terminal of enzyme sequence. Finally, the active enzymes were purified and concentrated by one step anion exchange chromatographic purification (Q HP column).

The refolding process scheme and conditions are shown in Figure 1.

With respect to the various fusion tags, the optimal concentration for refolding was also explored.

Based on tests at different concentrations of fusion proteins, it was found that the amount of purified (i.e., correctly refolded) EK_L(C112A, L134K, I135K) from a fixed volume was dependent upon the concentration of soluble fusion protein in the refolding buffer, and reached a maximum when the concentration was 250 μg/ml. All the fusion proteins with ΔssDsbA tag were refolded at the concentration of 250μg/ml.

In contrast, for all the other fusion tags tested, the maximum amount of purified enzyme from a fixed volume was obtained at the concentration much lower than 250μg/ml.

Example 4: Enzyme assays

The enzymatic activity was measured directly using a fluorogenic substrate, GDDDDK-Beta-naphthylamide. The reaction was started with addition of 1ul sample into each well of Fluorescent 96 well plate containing 100ul of reaction buffer. After incubating for 10 seconds, the fluorescence was measured with Fluostar OPTIMA (excitation at 340nm and emission at 420nm). The enzyme activity was defined by arbitrary unit (EU), which derived from slope*60/30,000, where the slope represented linear range.

The mutants introduced to EK_L were designed not near or inside the catalytic domain of EK_L and the mutants are not expected to alter the protein conformation. Thus, the catalytic activities of these analogues should be comparable to EK_L. Activities of EK_L(C112A, L134K, I135K) and EK_L(C112A, L134K, I135K, Y209R) were tested and compared with the activity of EK_L at the same concentration of purified refolded protein (0.1 µg/ml). The activity of EK_L (column A in Fig. 2) was defined as 100% here. The relative activities of EK_L(C112A, L134K and I135K) (column B in Fig. 2) and EK_L(C112A, L134K, I135K and Y209R) (column C in Fig. 2) were 107% and 97%, respectively. Please see Figure 2.

Since the catalytic activities of the EK_L analogues are comparable, the activity of protein derived from the same amount of inclusion body correlates with the refolding efficiency. The catalytic activity of EK_L is 5000 EU for 1mg/ml EK_L. Thus, the final concentration of correctly refolded EK_L analogue (mg/L) can be calculated by dividing the tested catalytic activity (EU/ml) by 5.

The EK_L analogues prepared were all stable when stored in buffer containing 20mM Tris, 200mM NaCl at -80°C, 4°C, or even at room temperature. No apparent degradation or decrease of activity was observed after 3 months storage.

Example 5: Effect of various mutants on the refolding of EK_L analogues

Mutants were introduced to EK_L by using QuikChange® XL Site-Directed Mutagenesis Kit (Stratagene) to obtain the constructs DsbA-Linker-EK_L analogues from DsbA-Linker-EK_L.

DsbA (Disulphide oxidoreductase) is an oxidase responsible for generation of disulfide bonds in proteins of E. coli. DsbA is localized in periplasm of E. coli, and is responsible for the introduction of disulfide bonds into newly synthesized proteins that are translocated to the periplasm.

This experiment was designed as a high throughput assay to screen mutants that improve refolding of EK_L analogues. Fusion proteins of DsbA-Linker-EK_L analogues were

synthesized as secreted proteins from *E. coli*, so that disulfide bonds were introduced to EK_L analogues in periplasm space. This mimics the in vitro refolding process.

When fusion proteins were secreted into the supernatant, fully active EK_L analogues were liberated from the fusion proteins by the escaped active enzymes, which specifically
5 cleaved fusion tag off at DDDDK recognition site just before the enzyme sequence.

Activities of secreted EK_L analogues were tested to reflect the yield of EK_L analogues with correct conformation. The EK_L analogues with mutants that have the highest activity were then synthesized again through inclusion body process described in Example 2 and refolded as described in Example 3. The activity of purified EK_L analogues prepared by
10 inclusion body and refolding process were tested by the activity assay described in Example 4, to confirm the mutants' effect on improving refolding.

Procedure:

15 Cells transformed with pET39b-DsbA-Linker- EK_L analogues were stored in glycerol stock. The EK_L analogues with mutants to be tested were EK_L(C112A, M48K), EK_L(C112A, L65K), EK_L(C112A, L74K), EK_L(C112A, I125K), EK_L(C112A, V143K), EK_L(C112A, Y209R), EK_L(C112A, Q210R), EK_L(C112A, Q210K). The positive control was cells transformed with pET39b-DsbA-Linker-EK_L(C112A), and the negative control was cells transformed with
20 pET39b.

Aliquots of 40µl stock solution comprising the cells were grown at 37°C for 4 hrs for inoculation. Then aliquots of 10 µl inoculated culture were added to 170 µl/well seed medium (LB+1%glucose+100µg/ml Kanamycin) respectively in 96-well plate. Cells then grew in Kuner shaker under the condition of 220rpm, 30°C, 80% humidity for overnight.

25 Aliquots of 495 µl/well culture and expression medium (EC1+100µg/ml Kanamycin) were added to 96-well plate. Aliquots of 5 µl from seed culture were added to 495 µl/well culture and expression medium (inoculation of 1:100 v/v). The culture was started as OD=0.05, then grew at 37°C for 3 hrs. When OD reached 4.5, cells were induced by adding 5 µl of 10mM IPTG to each well. Cells continued to grow at 25°C for overnight.

30 Cells in each well of 96-well plate was diluted 10 fold with culture and expression medium (EC1+100µg/ml Kanamycin). The 96-well plate was centrifuged at 4000rpm for 10 mins to separate cells and supernatant. Aliquots of 200 µl supernatant/well were applied to STI column for purification. The purified EK_L analogues were also tested for enzymatic activity test as described in Example 4.

Example 6: ResultsEffect of fusion tags on refolding efficiency and purification

To compare the refolding efficiency of fusion tags on EK_L or EK_L analogues, we tested twelve tags: Trx, DsbA, ΔssDsbA, ΔssDsbC, ΔssDsbG, GST, NRDH-Redoxin, ERp1, hAdrenodoxin, hGlutaredoxin, hPDI, and streptococcal-glutaredoxin on refolding of EK_L(C112A, L134K, I135K). All the tested fusion tags have Trx-like domain.

Cells were transformed with pET39b-DsbA-Linker-EK_L(C112A, L134K, I135K), pET39b-ΔssDsbA-Linker-EK_L(C112A, L134K, I135K), pET39b-Trx-Linker-EK_L(C112A, L134K, I135K), pET39b-hNRDH-Redoxin-Linker-EK_L(C112A, L134K, I135K), pET39b-ERp19-Linker-EK_L(C112A, L134K, I135K), pET39b-hAdrenodoxin-Linker-EK_L(C112A, L134K, I135K), pET39b-hGlutaredoxin-Linker-EK_L(C112A, L134K, I135K), pET39b-hPDI-Linker-EK_L(C112A, L134K, I135K), pET39b-streptococcal-glutaredoxin-Linker-EK_L(C112A, L134K, I135K), etc.). The transformed cells were cultured and expressed the fusion proteins according to the process described in Example 3.

No apparent leaky expression was observed by SDS-PAGE before IPTG induction.

The catalytic activities were tested according to Example 4. The results are shown in Table 1 and Figure 3. Please note the activities were tested on EK_L analogues obtained from the same amount of inclusion bodies. Thus, the activities were correlated with the refolding yield of the EK_L analogues.

Table 1: Effect of fusion tags on refolding efficiency

Column	Tags	Catalytic Activity (EU/ml)
A	Trx	225
B	DsbA	0
C	ΔssDsbA	325
D	ΔssDsbC	208
E	ΔssDsbG	9
F	GST	110
G	NRDH-Redoxin	0
H	ERp19	74
I	hAdrenodoxin	100
J	hGlutaredoxin	27
K	hPDI	15
L	streptococcal-glutaredoxin	26

Trx tag has been commonly used for facilitate the refolding of EK_L or EK_L analogues. Compared with Trx tag, ΔssDsbA tag significantly improved the total refolding

yield of EK_L(C112A, L134K, I135K) up to 40%. Due to the large difference on surface charge states between ΔssDsbA tag and EK_L(C112A, L134K, I135K), the auto-cleaved ΔssDsbA were easily removed from EK_L(C112A, L134K, I135K) by Q HP column. ΔssDsbC tag also facilitated refolding of EK_L(C112A, L134K, I135K) at a level comparable to Trx tag.

5 Based on tests at different concentrations of ΔssDsbA-Linker-EK_L(C112A, L134K, I135K), it was found that the amount of correctly refolded EK_L(C112A, L134K, I135K) from a fixed volume was dependent upon the concentration of soluble fusion protein in the refolding buffer, and reached a maximum when the concentration was 250 μg/ml. All the fusion proteins with ΔssDsbA tag were refolded at the concentration of 250μg/ml.

10 In contrast, for all the other fusion tags tested, the maximum amount of purified enzyme from a fixed volume was obtained at the concentration of soluble protein much lower than 250 μg/ml. For Trx-Linker-EK_L(C112A, L134K, I135K), the amount of purified enzyme from a fixed volume reached the maximum when the fusion protein concentration was 120 μg/ml, which was the highest concentration among all the other tested fusion tags. This means that using ΔssDsbA as the fusion tag can make the production process more efficient. For example, using a holding tank with the same volume, the amount of enzyme purified from ΔssDsbA fusion protein would be more than two times that of the enzyme purified from Trx fusion protein.

20 During the anion exchange chromatographic purification (Q HP column), it was found that DsbA and ΔssDsbA were easier to be separated from EK_L or other EK_L analogues, such as EK_L(C112A, L134K, I135K). This is due to the different surface charge property between DsbA/ssDsbA and EK_L(C112A, L134K, I135K).

Screening for mutants to improve refolding efficiently by secretion process

25 Constructs of pET39b-DsbA-Linker-EK_L analogues comprising M48K, L65K, L74K, C112A, I125K, L134K, I135K, V143K, Y209R, Q210R, and Q210K respectively in EK_L analogue were prepared according to Example 1. Effect of these mutants on refolding of EK_L analogues were tested according to Example 5. Results can be found in Table 2 and Figure 4. The EK_L analogues were prepared as secreted proteins in supernatant. Activities were tested to see if such mutants improve di-sulfide bonds forming in EK_L analogues, which mimics in vitro refolding. Y209R was identified as the mutant improving the refolding yield the most, comparing with other mutants.

Table 2: Effect of mutants on di-sulfide bonds forming in EK_L analogues

Mutants	Catalytic Activity
---------	--------------------

	(EU/ml)
M48K	3.4
L65K	5.1
L74K	4.5
I125K	3.6
V143K	4.3
Y209R	12.75
Q210R	7.1
Q210K	0
Control(C112A)	10

Confirming mutants that can improve refolding efficiency and the synergistic effect with fusion tags

Cells were transformed with pET39b- Δ ssDsbA-Linker-EK_L(C112A, L134K, I135K),
 5 pET39b- Δ ssDsbA-Linker-EK_L(C112A, L134K, I135K, L213K), pET39b- Δ ssDsbA-Linker-
 EK_L(C112A, L134K, I135K, Y209R), pET39b- Δ ssDsbA-Linker-EK_L(C112A, Y209R); pET39b-
 Δ ssDsbC-Linker-EK_L(C112A, L134K, I135K), pET39b- Δ ssDsbC-Linker-EK_L(C112A, L134K,
 I135K, L213K), pET39b- Δ ssDsbC-Linker-EK_L(C112A, L134K, I135K, Y209R); pET39b-
 Δ ssDsbG-Linker-EK_L(C112A, L134K, I135K), pET39b- Δ ssDsbG-Linker-EK_L(C112A, L134K,
 10 I135K, L213K), pET39b- Δ ssDsbG-Linker-EK_L(C112A, L134K, I135K, Y209R). The
 transformed cells were cultured and expressed the fusion proteins according to the process
 described in Example 3.

The catalytic activities of EK_L analogues were tested according to Example 4. The
 results are shown in Table 3 and Figure 5. Please note the activities were tested on EK_L
 15 analogues obtained from the same amount of inclusion bodies. Thus, the activities were
 correlated with the refolding yield of the EK_L analogues.

Table 3

Column	Tags and EK _L analogues	Catalytic Activity (EU/ml)
A	Δ ssDsbA, EK _L (C112A, L134K, I135K)	138
B	Δ ssDsbA EK _L (C112A, L134K, I135K, L213K)	325
C	Δ ssDsbA EK _L (C112A, L134K, I135K, Y209R)	371
D	Δ ssDsbC EK _L (C112A, L134K, I135K)	79
E	Δ ssDsbC EK _L (C112A, L134K, I135K, L213K)	231
F	Δ ssDsbC EK _L (C112A, L134K, I135K, Y209R)	236

G	Δ ssDsbG EK _L (C112A, L134K, I135K)	30
H	Δ ssDsbG EK _L (C112A, L134K, I135K, L213K)	31
I	Δ ssDsbG EK _L (C112A, L134K, I135K, Y209R)	191
J	Δ ssDsbA EK _L (C112A, Y209R)	195

Based on the results above, Y209R was proved to have the best synergistic effect with other mutation sites and fusion tags to facilitate the refolding of EK_L analogues.

5 While certain features of the invention have been illustrated and described herein, many modifications, substitutions, changes, and equivalents will now occur to those of ordinary skill in the art. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the invention.

10 All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference in their entirety and to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein (to the maximum extent permitted by law).

15 All headings and sub-headings are used herein for convenience only and should not be construed as limiting the invention in any way.

20 The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

 The citation and incorporation of patent documents herein is done for convenience only and does not reflect any view of the validity, patentability, and/or enforceability of such patent documents.

25 This invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law.

CLAIMS

1. A fusion protein, comprising a fusion tag selected from Dsb family members with signal sequence deleted, an enterokinase light chain, and a linker between the fusion tag and the enterokinase light chain.

5

2. The fusion protein of claim 1, wherein the enterokinase light chain is derived from a mammalian species, or an analogue thereof.

3. The fusion protein of claim 2, wherein the mammalian species is bovine, human or porcine.

10

4. The fusion protein of any one of the preceding claims, wherein the Dsb family fusion tag is DsbA, DsbC, or DsbG, with signal sequence deleted.

15

5. The fusion protein of any one of the preceding claims, wherein the enterokinase light chain is a bovine enterokinase light chain (SEQ ID NO:1).

6. The fusion protein of any one of the preceding claims, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 112 of SEQ ID NO:1.

20

7. The fusion protein of claim 6, wherein the substitution in position 112 is C112A.

8. The fusion protein of any one of the preceding claims, wherein the enterokinase light chain is a bovine enterokinase light chain analogue comprises a substitution in position 209 of SEQ ID NO:1.

25

9. The fusion protein of claim 8, wherein the substitution is Y209R.

10. The fusion protein of any one of claims 6 to 9, wherein the bovine enterokinase light chain analogue further comprising substitutions in positions 134 and 135 of SEQ ID NO:1.

30

11. The fusion protein of claim 10, wherein the substitutions in positions 134 and 135 are L134K and I135K.

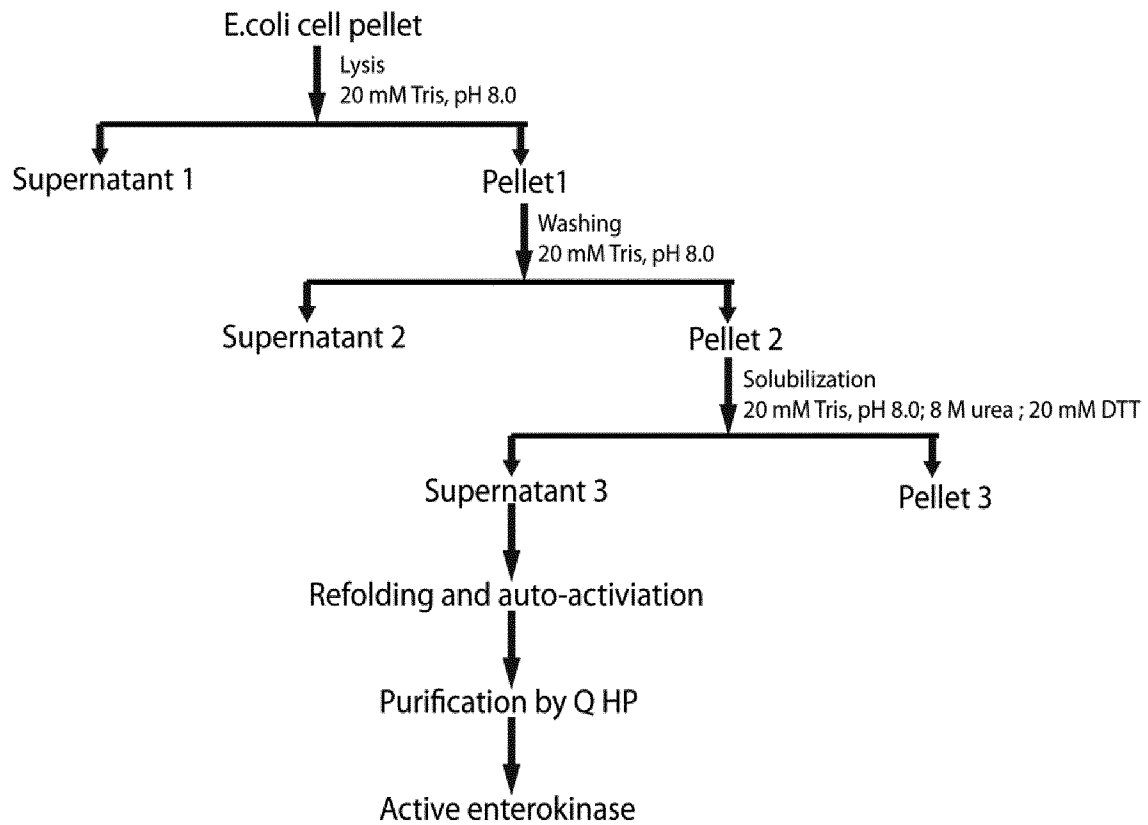
35

12. The fusion protein of any one of the preceding claims, wherein the linker comprises a cleavage site, wherein the amino acid sequence of the cleavage site is Asp-Asp-Asp-Lys.

5 13. A fusion protein, comprising a DsbA tag with signal sequence deleted, a linker and a bovine enterokinase light chain analogue with substitutions of C112A, L134K, I135K, and Y209R.

10 | ~~1445~~. A Bovine enterokinase light chain analogue comprising substitutions of C112A, L134K, I135K, and Y209R.

1/5

**Fig. 1/5**

2/5

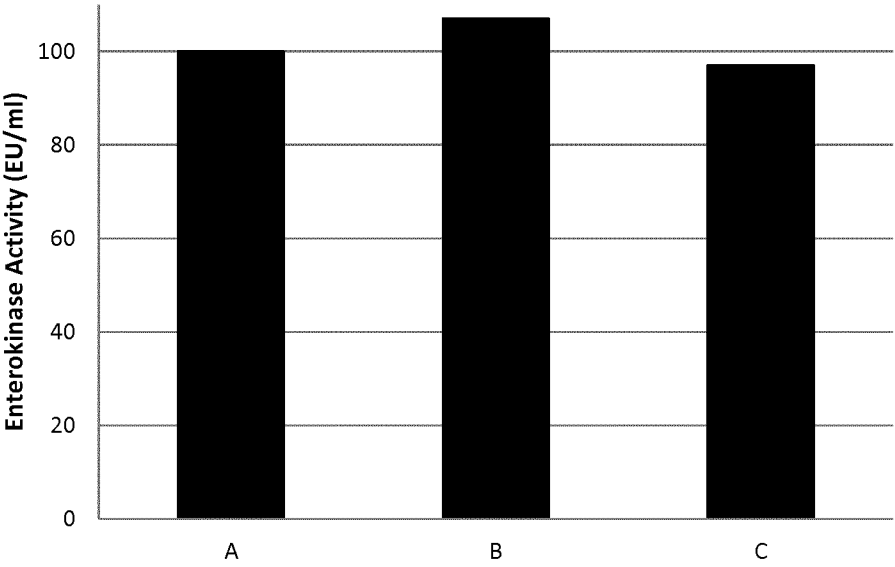


Fig. 2/5

3/5

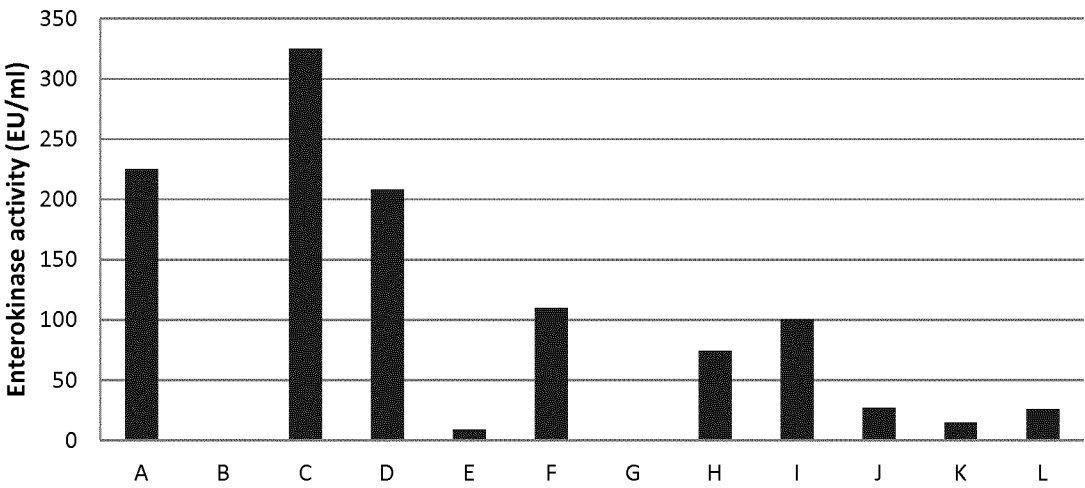


Fig.3/5

4/5

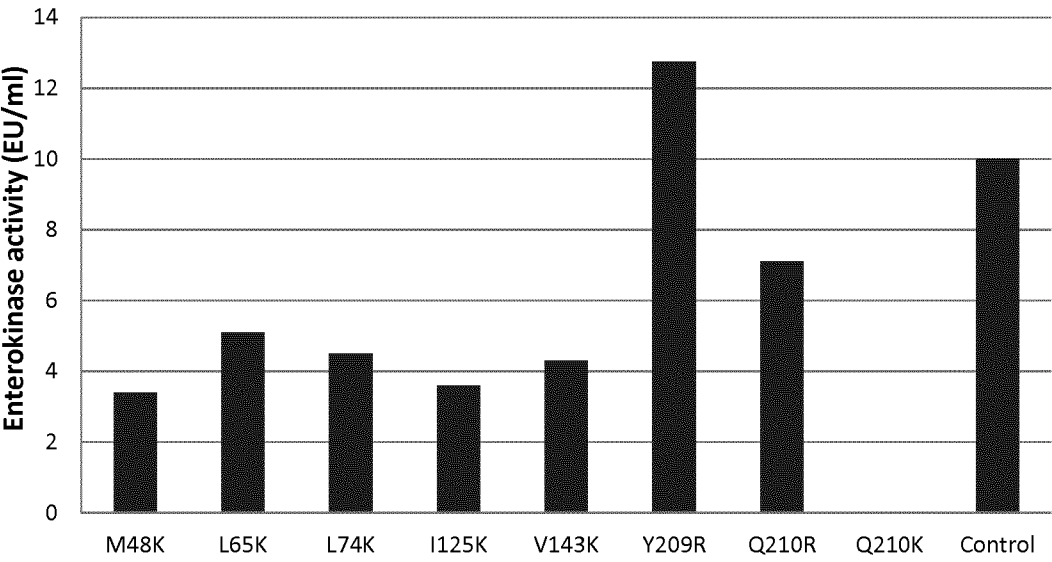


Fig. 4/5

5/5

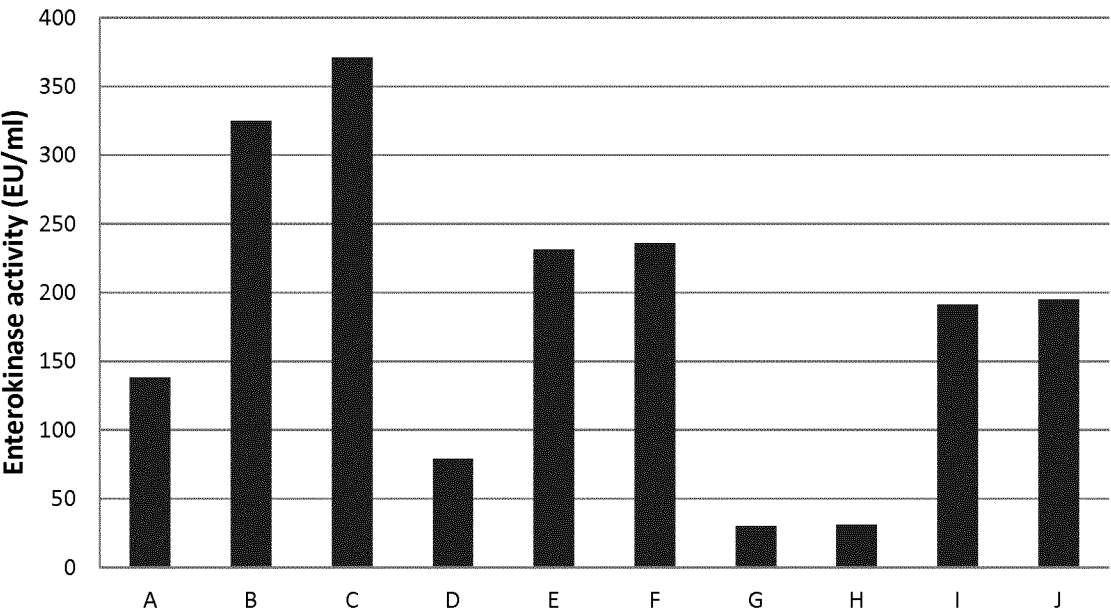


Fig. 5/5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/050303

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/64
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBASE, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COLLINS-RACIE LISA A ET AL: "Production of recombinant bovine enterokinase catalytic subunit in Escherichia coli using the novel secretory fusion partner Dsba", BIOTECHNOLOGY. THE INTERNATIONAL MONTHLY FOR INDUSTRIAL BIOLOGY, NATURE PUBLISHING GROUP, US, vol. 13, no. 9, 1 January 1995 (1995-01-01), pages 982-987, XP002156881, ISSN: 0733-222X, DOI: 10.1038/NBT0995-982 abstract	1-14
A	----- -/--	15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

22 February 2017

Date of mailing of the international search report

07/03/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Griesinger, Irina

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2017/050303

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEI HUANG ET AL: "Functional Expression and Purification of Bovine Enterokinase Light Chain in Recombinant Escherichia coli", PREPARATIVE BIOCHEMISTRY AND BIOTECHNOLOGY, vol. 37, no. 3, 1 May 2007 (2007-05-01), pages 205-217, XP055289903, US ISSN: 1082-6068, DOI: 10.1080/10826060701386695	1-14
A	abstract	15
X	----- CN 101 624 588 A (SHANGHAI ZHANGJIANG BIOLOGY TE) 13 January 2010 (2010-01-13)	1-14
Y	figure 1	15
Y	----- WO 2013/092855 A1 (NOVO NORDISK AS [DK]) 27 June 2013 (2013-06-27)	15
A	claims 1,2	1-14
X	----- PETER SIMEONOV ET AL: "Surface supercharged human enteropeptidase light chain shows improved solubility and refolding yield", PROTEIN ENGINEERING, DESIGN AND SELECTION, OXFORD JOURNAL, LONDON, GB , vol. 24, no. 3 1 March 2011 (2011-03-01), pages 261-268, XP002679971, ISSN: 1741-0126, DOI: 10.1093/PROTEIN/GZQ104 Retrieved from the Internet: URL:http://peds.oxfordjournals.org/content/24/3/261 [retrieved on 2010-11-16]	15
A	abstract	1-14
A	----- WO 95/04076 A1 (CELTRIX PHARMA [US]) 9 February 1995 (1995-02-09) abstract; claims 1-7	1-15
A	----- Novagen: "pET-39b(+) Vector", , 1 December 1998 (1998-12-01), XP055290097, Retrieved from the Internet: URL:https://www.embl.de/pepcore/pepcore_services/strains_vectors/vectors/pdf/pET-39b_map.pdf [retrieved on 2016-07-20] the whole document	1-14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/050303

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 101624588 A	13-01-2010	CN 101386847 A	18-03-2009
		CN 101613687 A	30-12-2009
		CN 101624586 A	13-01-2010
		CN 101624587 A	13-01-2010
		CN 101624588 A	13-01-2010

WO 2013092855 A1	27-06-2013	EP 2794875 A1	29-10-2014
		JP 2015502167 A	22-01-2015
		US 2015064744 A1	05-03-2015
		WO 2013092855 A1	27-06-2013

WO 9504076 A1	09-02-1995	AT 280234 T	15-11-2004
		AU 688363 B2	12-03-1998
		CA 2168429 A1	09-02-1995
		DE 69434083 D1	25-11-2004
		DE 69434083 T2	09-02-2006
		EP 0722461 A1	24-07-1996
		JP 3370094 B2	27-01-2003
		JP H09504166 A	28-04-1997
		US 5563046 A	08-10-1996
		US 5629172 A	13-05-1997
		US 5830706 A	03-11-1998
		WO 9504076 A1	09-02-1995
