(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 April 2012 (19.04.2012)

PCT

(10) International Publication Number
WO 2012/048856 A1

(51) International Patent Classification:
*C07K 14/62* (2006.01) *C12N 15/62* (2006.01)

(21) International Application Number:
PCT/EP2011/005090

(22) International Filing Date:
11 October 2011 (11.10.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10013554.0 12 October 2010 (12.10.2010) EP

(71) Applicant *(for all designated States except US)*: **GLUCOMETRIX PVS GMBH** [DE/DE]; Am Mühlenberg 11, 14476 Potsdam (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PONOMARENKO, Sophia** [DE/DE]; Göttweiger Str. 70, 94032 Passau (DE). **HERMANN, Stefan** [DE/DE]; Hegelallee 41, 14467 Potsdam (DE). **ECKERT, Kelvin** [DE/DE]; Corinthstr. 56, 10245 Berlin (DE).

(74) Agent: **BRITTINGER, Matthias**; TER MEER STEINMEISTER & PARTNER GBR, Mauerkircherstraße 45, 81679 München (DE).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

**Published**:

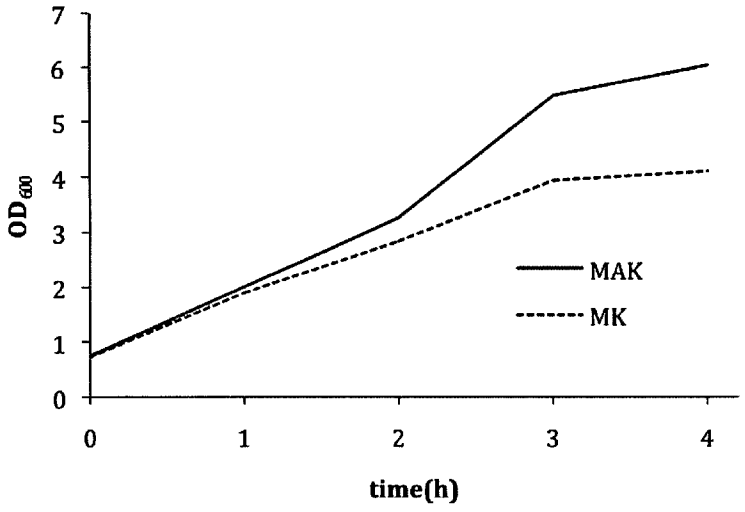
— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

*[Continued on next page]*

(54) **Title**: PROINSULIN WITH HELPER SEQUENCE

**Figure 1**




(57) **Abstract**: The invention provides with a polypeptide comprising a proinsulin precursor and a helper sequence. The helper sequence enhances the expression of the proinsulin precursor in a recombinant organism and increases the yields of insulin after purification and processing of the insulin precursor. The invention is thus further concerned with a method of producing insulin, in which proinsulin is expressed as recombinant polypeptide in an organism such as E. coli, isolated from the organism and then processed to yield active insulin.

WO 2012/048856 A1

— *with sequence listing part of description (Rule 5.2(a))*

# Proinsulin with helper sequence

## Field of the invention:

The invention is concerned with a polypeptide comprising a proinsulin precursor and a helper sequence. The helper sequence enhances the expression of the proinsulin precursor in a recombinant organism and increases the yields of insulin after purification and processing of the insulin precursor. The invention is thus further concerned with a method of producing insulin, in which proinsulin is expressed as recombinant polypeptide in an organism such as E. coli, isolated from the organism and then processed to yield active insulin.

## General remarks:

Insulin is a peptide hormone and regulates the blood glucose level. Therefore, insulin is administered to patients suffering from diabetes mellitus, a metabolic disorder characterized by an inadequate supply of insulin. According to the World Health Organization (WHO) some 285 million people will suffer from diabetes mellitus in 2010. Insulin therapy is essential to the survival of those with type 1 diabetes and is used to control the symptoms for those patients suffering from type 2 diabetes.

Human insulin consists of two separate chains, which are the A-chain of a length of 21 amino acids and the B-chain of a length of 30 amino acids. The A- and B-chains are joined together by a characteristic pattern of disulfide bridges. In the human pancreas, insulin is produced as a single pre-proinsulin chain in which the prospective A- and B-chains are linked together by the C-peptide and which further contains an N-terminal signal sequence. The formation of native insulin from proinsulin involves folding and the formation of correct disulfide bridges. Proinsulin is then processed proteolytically, which results in the cleavage of the C-peptide and release of the active hormone.

Traditionally, insulin was produced from animal sources such as bovine and porcine pancreatic preparations. Insulin produced from animal sources, however, differs from human

insulin and thus may elicit an adverse immune reaction. Due to the enormous demand for insulin as medicament, biosynthetic human insulin is manufactured for widespread clinical use using recombinant DNA technology.

Due to high rate of synthesis and rapid growth, the primary source for the manufacturing of biosynthetic recombinant insulin is its production in Escherichia coli (E. coli). There are two major approaches for the generation of recombinant human insulin from E. coli. In one approach, the A- and B-chains of human insulin are expressed separately, converted to their stable S-sulfonate derivatives, and subsequently combined to generate native insulin. The other approach is to produce proinsulin as a recombinant gene product, which is subsequently isolated from the host cell and processed in vitro by proteases to release the C-peptide and yield native insulin. The second approach is preferred, largely as a consequence of the requirement for only a single fermentation and subsequent purification protocol, thus it is more efficient than the two-chain combination approach.

However, manufacturing of recombinant proinsulin in E. coli is also accompanied with by several disadvantages. Upon expression in host cells high molecular weight aggregates are formed, often referred to as "inclusion bodies", which result from the inability of the expressed proteins to fold correctly in an unnatural cellular environment. The protein is present in the insoluble inclusion bodies in denatured form, thus requiring the use of detergents and denaturants to isolate and solubilize the protein. The isolated protein must subsequently be refolded in vitro. This includes the formation of the correct disulfide bridges and processing by proteases to convert the correctly folded proinsulin precursor into insulin.

Furthermore heterologous expression in host cells often leads to unspecific degradation of the heterologous protein, resulting in low protein yield and multiple degradation products.

The cumbersome and laborious procedures for the isolation of recombinant human insulin from E. coli involving proinsulin expression, isolation, refolding and processing often results in low yields of active product.

Expression yields often depend on the stability of the induced protein as it is affected by the protein sequence. A major factor influencing the half life of a protein is the N-terminus, as described by the N-end rule (Tobias et al. (1991), Science, 254(5036), 1374-7). The residues arginine, lysine, leucine, phenylalanine, tyrosine and tryptophan at the amino terminus (N-terminus) tend to decrease the half-life of the protein, i.e., the half-life can be in the order of two minutes, whereas other residues provide proteins having a half-life of more than ten hours for a protein differing only in the N-terminal amino acids.

Another approach to form more stable recombinant protein is to fuse the protein to another protein which is naturally present or which is easily expressed in the host organism. EP 0 871 474 B1 describes a method of insulin production in E. coli, in which the proinsulin polypeptide is fused to a superoxide dismutase (SOD) as leader sequence. After partial purification, the hybrid polypeptide is folded and subsequently processed with trypsin and carboxypeptidase to cleave off the leader peptide and the C-chain concomitantly.

Methods of isolating insoluble recombinant protein from inclusion bodies are well known in the prior art. These methods however, mostly result in low yield of active protein. Accordingly, a need exists to increase the yield of active protein such as human proinsulin after expression, isolation and processing of the recombinant protein from microorganisms such as E. coli.

**Object and summary of the invention:**

The underlying technical problem of the present invention thus is to enhance the expression of a proinsulin precursor in a heterologous organism and thereby increasing the yield of active recombinant insulin obtained from the proinsulin after expression. In particular, it is the object of the present invention to provide a recombinant polypeptide comprising the sequence of proinsulin, which is expressed in E. coli in high amounts and which can thus be easily isolated from inclusion bodies and which can further be easily processed into active insulin. Thus, it was another object of the present inventors to develop a process for obtaining recombinant active insulin, preferably active human insulin in high amounts.

The solution of the underlying technical problems is surprisingly that the yield of recombinant proinsulin can be increased by fusing to the amino terminus (N-terminus) of the amino acid chain of the proinsulin precursor a helper sequence which consists of the amino acid sequence methionine, alanine and lysine, which is written in the standard amino acid one-letter code as MAK.

Although it was known from the N-end rule in bacteria that some N-terminal amino acids may considerably increase the half-life time of proteins compared to other amino acids, it has surprisingly been found out by the inventors of the present invention that the amino acid sequence MAK located at the N-terminus of a polypeptide comprising an insulin precursor significantly increases the final cell density of a culture of a heterologous organism expressing the recombinant polypeptide. Additionally, it has been found out that the amino acid sequence MAK located at the N-terminus of a polypeptide comprising an insulin precursor also increases expression and stability of the recombinant polypeptide.

The present invention provides a polypeptide comprising,

i. a first peptidyl fragment consisting of the N-terminal amino acid sequence methionine, alanine and lysine (MAK), and
ii. a second peptidyl fragment, which is an insulin precursor.

In a preferred embodiment, the insulin precursor comprises insulin chains A and B, and most preferably is human proinsulin.

The invention further provides an isolated nucleic acid, which preferably is DNA, and comprising a nucleotide sequence encoding the polypeptide comprising the N-terminal MAK sequence and the insulin precursor.

The present invention further provides a recombinant cell such as a recombinant E. coli cell, which contains the nucleic acid encoding an amino aid sequence comprising the N-terminal MAK sequence and the insulin precursor.

The invention is further directed to a method of producing a polypeptide, the method comprising growing a recombinant cell such as an E. coli cell containing a nucleic acid such that the polypeptide encoded by the nucleic acid is expressed by the cell, and recovering the expressed polypeptide from the cell.

In a preferred embodiment, the method of producing insulin comprises the steps of

a. expressing the polypeptide comprising the N-terminal MAK sequence and the insulin precursor in Escherichia coli,
b. isolating the polypeptide from the recombinant cell,
c. subjecting the isolated polypeptide to a folding process which permits correct folding of the insulin precursor, and
d. subjecting the polypeptide to enzymatic processing to yield active insulin.

**Brief description of the drawings:**

Figure 1 shows growth curves of shaking cultures of E. coli cells transformed with a plasmid encoding DNA according to the subject invention compared to E. coli cells transformed with a plasmid according to a comparison example.

Figure 2 shows an SDS-PAGE analysis of protein samples obtained from E. coli shaking cultures after expression of a plasmid encoding DNA according to the subject invention and after expression with a plasmid according to a comparison example.

Figure 3 shows a semi quantitative analysis of the SDS-PA gels shown in Figure 2.

Figure 4 shows an SDS-PAGE analysis of protein samples obtained from E. coli cultures after growth in a fermenter and expression of a plasmid encoding DNA according to the subject invention compared to cultures after expression of a plasmid according to a comparison example.

Figure 5 shows a semi quantitative analysis of the SDS-PA gels shown in Figure 4.

**Detailed description of the invention:**

In the present invention the term "heterologous expression" means that the protein is experimentally put into a cell that does not normally make (i.e., express) that protein. Heterologous polypeptide or heterologous protein thus refers to the fact that the transferred DNA coding for a polypeptide or protein such as insulin or proinsulin was initially cloned from or derived from a different cell type or a different species from the recipient. For example, the gene encoding the proinsulin precursor can be made synthetically and then transferred into the host organism, which as native organism does not produce that polypeptide or protein. Therefore, the genetic material encoding for the polypeptide or protein is added to the recipient cell by recombinant cloning techniques known in the art. The genetic material that is transferred for the heterologous expression must be within a format that encourages the recipient cell to express the recombinant DNA as open reading frame (ORF) to synthesize a protein, i.e., it is put in an expression vector.

A "polypeptide" refers to a single linear chain of amino acids. A "protein" refers to a polypeptide, which has the ability to form into a specific conformation. Thus the terms polypeptide and protein can generally be used interchangeably for polypeptides of a specific length.

The term "recombinant DNA" refers to the form of artificial DNA such as a synthetical DNA or cDNA encoding a proinsulin precursor that is created through the introduction of the DNA into an organism such as E. coli for the purpose of expression of the polypeptide or protein encoded by the recombinant DNA. A recombinant protein thus is a protein that is derived from the recombinant DNA by expression of the recombinant DNA in the host cell. The recombinant DNA techniques required for transferring the recombinant DNA into a host organism and expression of the recombinant DNA to yield the recombinant protein in the organism are known to the one skilled in the art.

An "insulin precursor" refers to a molecule, which comprises, contains or is homologue to insulin chains A and B, comprising analogues, derivatives and fragments thereof. A human

insulin precursor refers to a polypeptide, which contains or is homologous to human insulin chains A and B, comprising analogues, derivatives and fragments thereof.

A "correctly folded" human insulin precursor refers to a molecule wherein the human insulin precursor has the conformation and disulfide bridges as found in a natural, biologically active human insulin, i.e., the disulfide bridges between a) A-6 and A-11, b) between A-7 and B-7 and c) between A-20 and B-19 are formed.

In a first aspect, the present invention provides an amino acid sequence, which comprises at the N-terminus the amino acid sequence consisting of methionine, alanine and lysine (MAK) and which further comprises an amino acid sequence of a proinsulin precursor. The N-terminal sequence MAK has the function of a helper sequence, which promotes the expression of the recombinant DNA encoding the polypeptide of the subject invention and which promotes stability of the polypeptide after recombinant expression in a host organism such as E. coli. Thus, the helper sequence MAK leads to an increased yield of protein after isolation, purification and processing of the recombinant polypeptide in order to obtain correctly folded insulin.

Preferably, the N-terminal helper sequence is located directly at the N-terminus of a proinsulin precursor. The insulin precursor preferably comprises the A- and B-chains of human proinsulin. Also described herein is a proinsulin precursor in which the A- and B-chains of human proinsulin are separated by a single amino acid or by an amino acid sequence consisting of 2 to 35 amino acids. Most preferably, the A- and B-chains are separated by the human C-peptide. In the most preferred embodiment, the human proinsulin precursor is human proinsulin and thus consists of the B-chain at its N-terminus followed by the C-peptide, which precedes the A-chain at its C-terminus. A concrete example of a polypeptide having the helper sequence MAK at its N-terminus and followed by the amino acid sequence of human proinsulin is given by SEQ ID NO 1:

MAKFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGP-GAGSLQPLALEGSLQKRGIVEQCCTSICSLYQLENYCN

The polypeptide of the present invention may further comprise a sequence, which promotes the purification of the recombinant polypeptide. A sequence, which is preferably used for purification of a recombinant polypeptide is a His-tag. The term His-tag or polyhistidine-tag is an amino acid motif in proteins that consists of at least four histidine (His) residues. The prevalently used His-tag consists of six histidine residues and is thus also known as hexa histidine-tag or 6xHis-tag. Such a His-tag can be used for affinity purification of the tagged recombinant protein, e.g. after expression in E. coli. Various purification kits for histidine-tagged proteins are available from Qiagen, Sigma, Thermo Scientific, GE Healthcare, Macherey-Nagel and others.

Different organisms often have different codon specificity to encode a single amino acid. As a consequence, it is preferred in the subject invention that the codon usage of the nucleotide sequence according to the present invention is adapted for the expression in the respective organism (E.L. Winnacker, Gene und Klone, Verlag Chemie, 1985, 224-241, Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nakamura, Y., Gojobori, T. and Ikemura, T. (2000) Nucl. Acids Res. 28, 292). For example, the DNA encoding the polypeptide comprising the helper sequence and the human proinsulin sequence of the present invention is preferably adapted for expression in the specific organism such as E. coli. Hence, in a preferred embodiment of the invention, the encoding DNA is adapted according to the codon usage of the host organism E. coli.

The subject invention thus relates to an isolated nucleic acid encoding a polypeptide having the helper sequence MAK at its N-terminus and further comprising an amino acid sequence of an insulin precursor. The isolated nucleic acid preferably is DNA. The invention further encompasses a nucleic acid sequence which is complementary to the nucleotide sequence encoding a polypeptide having the helper sequence MAK at its N-terminus and further comprising an amino acid sequence of an insulin precursor. The invention further encompasses an isolated nucleic acid which hybridizes under mild conditions to a nucleotide sequence, which is complementary to the nucleotide sequence encoding a polypeptide having the helper sequence MAK at its N-terminus and further comprising an amino acid sequence of an insulin precursor. A concrete example of such a nucleotide sequence, which preferably is DNA is given by SEQ ID NO 2:

```
                    20                  40                  60
atggcgaaat ttgtgaacca gcatctgtgc ggcagccatc tggtggaagc gctgtatctg

                    80                  100                 120
gtgtgcggcg aacgtggctt tttttatacc ccgaaaaccc gtcgtgaagc ggaagatctg

                   140                  160                 180
caagtgggcc aggtggaact gggtggtggt ccgggtgcgg gtagcctgca accgctggcg

                   200                  220                 240
ctggaaggca gcctgcaaaa acgtggcatt gtggaacagt gctgcaccag catttgcagc

                   260
ctgtatcagc tggaaaacta ttgcaactaa
```

In a further embodiment of the invention the amino acid sequence of the insulin precursor is encoded by the human proinsulin gene sequence or one or more fragments thereof. SEQ ID NO: 3 shows the naturally occurring coding sequence of human proinsulin, wherein the shown sequence is devoid of introns. The complete preproinsulin coding DNA sequence is according to GenBank Acc. No. J00265.1. Thus, the subject invention relates to an isolated nucleic acid encoding a polypeptide having the helper sequence MAK at its N-terminus and further comprising an amino acid sequence of an insulin precursor encoded by SEQ ID NO: 3 or one or more fragments thereof. SEQ ID NO 3 is:

```
                    20                  40                  60
atggccctgt ggatgcgcct cctgcccctg ctggcgctgc tggccctctg gggacctgac

                    80                  100                 120
ccagccgcag cctttgtgaa ccaacacctg tgcggctcac acctggtgga agctctctac

                   140                  160                 180
ctagtgtgcg gggaacgagg cttcttctac acacccaaga cccgccggga ggcagaggac
```

```
                  200                 220                 240
ctgcaggtgg ggcaggtgga gctgggcggg ggccctggtg caggcagcct gcagcccttg

                  260                 280                 300
gccctggagg ggtccctgca gaagcgtggc attgtggaac aatgctgtac cagcatctgc

                  320
tccctctacc agctggagaa ctactgcaac tag
```

The DNA of the invention may be obtained by standard procedures known in the art from cloned DNA, e.g., a DNA "library", by chemical synthesis, by cDNA cloning, or by the cloning of genomic DNA or fragments thereof, purified from the desired cell. Whatever the source, the gene should be molecularly cloned into a suitable vector for propagation of the gene.

For the expression of the polypeptide of the present invention, the DNA encoding the polypeptide is thus incorporated by standard cloning techniques into an expression vector. The expression vector provides all elements necessary for expression of the recombinant polypeptide in the heterlogous host. Suitable expression vectors are commercially available and include standard expression vectors for expression in E. coli such as pQET7 available from QIAGEN in which the gene encoding the recombinant protein is expressed under control of the T7 promoter. Transformation of the host cell by the expression vector can be achieved as described by Sambrook et al., Cold Spring Harbor Laboratory Press, 1998 and as known to the one skilled in the art. Suitable E. coli strains are commercially available and include various strains derived from E. coli BL21 such as E. coli C41, commercially available from Lucigen.

The host organism, preferably a prokaryotic cell, most preferably an E. coli cell is transformed by a DNA which is modified in a manner as described above and which encodes for the recombinant polypeptide of the present invention, by standard cloning techniques such as transformation of electro competent cells or chemically made competent cells.

Expression from certain promoters can be elevated in the presence of certain inducers; thus, expression of the genetically engineered polypeptide may be controlled. A preferred expression system is under control of the T7 promoter of E. coli and induced by the presence of IPTG (isopropyl β-D-1-thiogalactopyranoside).

The resulting heterologous host is then cultivated in a suitable medium. Suitable liquid media for growing the host organism include synthetical media, full or half media. Media for cultivation of E. coli include Luria Broth (LB), 2xYT or, in a particularly preferred embodiment of this invention, a fully synthetic medium based on a phosphate buffer, a nitrogen source like ammonium chloride, a carbon- and energy source like glucose or glycerol, trace elements, and an amino acid supplement to enhance growth (Korz, DJ *et al.* (1994), J. Biotech. 39, 59-65). Suitable conditions for cultivation are adapted to the organism according to standard procedures known to the skilled person. These include inoculation of the growth media with a starter culture and incubating the cells at a temperature between 25 °C and 42 °C. When grown in a flask, the concentration of dissolved oxygen is enhanced by rigorous shaking. In a fermenter, ample air supply is necessary. The pH of the culture should be kept between 5 and 8. At appropriate cell densities, expression of the recombinant protein is induced by addition of isopropyl-ß-D-galactopyranosid (IPTG) or lactose.

After sufficient growth of the heterologous organism, the cells are harvested e.g. by filtration or centrifugation and then disintegrated to further isolate the recombinant protein form the broken cells. Disintegration of cells can be achieved by high pressure homogenisation using a high pressure cell such as a french press cell. Other methods of disintegration of cells include enzymatic treatment with lysozyme and sonication.

In E. coli the recombinant polypeptide of the present invention is usually present in form of insoluble inclusion bodies. Thus, the inclusion bodies have to be isolated from the broken cells by washing the broken cells with mild detergents such as Tween 20 or Triton X 100 and with low concentrations of urea, preferably up to 2 M. The recombinant protein is then solubilised and renaturated under conditions, which allow formation of correct cystine bridges and refolding of the recombinant protein to its native conformation. Accordingly,

the process step of subjecting the isolated polypeptide to a folding process, which permits correct folding of the insulin precursor, preferably includes the formation of correct cystine bridges. Suitable conditions include the choice of an appropriate buffer and an appropriate pH for isolation and refolding of the recombinant protein as known in the art (Qiao, ZS *et al.* (2003), J. Biol. Chem. 278, 17800-17809; Winter, J *et al.* (2002), Anal. Biochem. 310, 148-155).

To obtain active insulin as product, the correctly folded insulin precursor has to be processed by enzymatic cleavage. Thus, the present invention further relates to a process for obtaining a correctly folded insulin-precursor, wherein the insulin precursor is processed by enzymatic cleavage in order to release the active insulin product. Enzymatic cleavage is achieved preferably by the proteases trypsin and carboxypeptidase B. Enzymatic cleavage can be achieved in separated process steps, in which in a first step the helper sequence MAK and most of the C-peptide, or any polypeptide located between the A and B chains, are released by incubation with trypsin, and the arginine at position B31 is trimmed off by carboxypeptidase B in a second step. In another embodiment of this invention, both enzymes are used simultaneously in a single process step. Thus, the helper sequence of the present invention has the additional advantage that it can easily be cleaved off by enzymatic processing due to the presence of the amino acid lysine at position 3 of the helper sequence.

Conditions for the processing of proinsulin to insulin are known to those skilled in the art (Jonasson, P. *et al.* (1996), Eur. J. Biochem. 236, 656-661; or as described in EP 0 367 161). Efficient enzymatic processing requires a slightly alkaline pH, which is preferably between 7,5 and 9, a temperature preferably between 4 °C and 37 °C, the presence of divalent cations such as calcium and/or magnesium and incubation times preferably between 15 min and 5 hours. The molar ratio of enzymes to proinsulin is preferably between 1:100 and 1:10000.

The processed and correctly folded Insulin is then purified in order to obtain the native product. Purification of the insulin product can be achieved by affinity chromatography, anion or cation exchange chromatography according to procedures known in the art. The

purified insulin can then be crystallized and/or lyophilized according to standard procedures.

The process steps as shortly described above are known from the prior art as listed above and can be adapted without inventive skill.

A method for the production of insulin from bacterial culture such as E. coli thus generally include the steps of:

a) cloning of the DNA sequence encoding the N-terminal MAK sequence and the proinsulin precursor into a suitable expression vector,
b) transforming a suitable organism for expression of the recombinant protein with the expression vector,
c) fermentation of the microorganism,
d) disintegration of cells,
e) isolation of inclusion bodies,
f) solubilization and renaturation of the proinsulin precursor,
g) incubation of the proinsulin precursor to allow formation of cystin bridges and folding of the proinsulin precursor,
h) concentration of the folded proinsulin precursor,
i) processing the proinsulin precursor with trypsin and carboxypeptidase to yield active insulin,
j) purification and crystallisation, and
k) lyophilisation of the active insulin.

It is within the knowledge of the skilled person that some of the steps may be modified or that further steps may be included such as further purification steps including chromatographic purification of the processed insulin.

The method steps as listed above will now be described in further detail by the following examples.

The examples describe preferred embodiments of the invention and are thus only intended for illustration purposes and are not intended to limit the subject invention in any way. As known to the skilled person, various modifications might be made to the described embodiments, without departing from the scope of the invention as defined by the scope of the subject claims. With regard to the methods as outlined below, i.e. the methods with regards to the expression of the polypeptide of the present invention in an heterologous organism, the methods for isolation of the recombinant protein from inclusion bodies, the methods of purifying the recombinant protein and subsequent refolding and processing, it will be understood that minor modifications in the protocols may be made without departing from the spirit or scope of the invention, i.e., with regard to the choice of solvents, buffers, detergents, denaturants, proteolytic enzymes, separation methods and chromatographic media. Such minor modifications are well within the knowledge of the skilled artisan and are intended to fall within the scope of the appended claims.

**Examples**

Example 1

Expression of pMAK-PI in E. coli grown in shaking cultures

The plasmid pMAK-PI is derived from an expression vector designed for the expression of a recombinant DNA in E. coli. The plasmid contains a gene according to the subject invention and given by SEQ ID NO: 2 encoding human proinsulin and containing a sequence encoding the N-terminal sequence methionine-alanine-lysine (MAK) located at the N-terminus of the proinsulin precursor.

The plasmid was transformed into E. coli C41 cells and grown on 2xYT solid medium containing kanamycin for selection of transformed cells. 300 cells were obtained on a petri dish. 5ml of 2xYT medium containing kanamycin were inoculated with a single colony and incubated overnight at 37°C under vigorous shaking. 400ml prewarmed 2xYT medium containing kanamycin were inoculated with the overnight culture and further incubated at 37°C under vigorous shaking. When the culture reached an optical density $OD_{600}$ of about

0.7, expression was induced by adding IPTG to a final concentration of 0.8 mM. Growth was checked by measuring the optical density and samples were taken in periodic intervals.

Comparative Example 1

Expression of pMK-PI in E. coli grown in shaking cultures

The plasmid pMK-PI differs from the plasmid pMAK-PI by the DNA sequence encoding for the N-terminal sequence preceding the proinsulin precursor. The encoded polypeptide has the N-terminal sequence Met-Lys-Proinsulin and thus differs from the encoded polypeptide of Example 1 in the amino acid alanine missing in the N-terminal helper sequence.

Growth and expression conditions of E. coli cells transformed with pMK-PI were as described above for Example 1.

As can be seen from Figure 1, cells containing the plasmid pMAK-PI reach a cell density after 4 h post induction, which is 50 % above the cell density of cells containing the plasmid pMK-PI grown under the same conditions and induced at the same time of cell density.

Figure 2 shows a comparison of the relative protein expression in an SDS-PAGE analysis. In this Figure, "M" depicts a molecular weight protein marker for evaluating the size of the expressed proteins. Columns designated with "A" show protein samples according to Example 1 and thus depict an example according to the invention, whereas columns designated with "K" show samples according to comparative Example 1. Shown are total protein samples of the overnight cultures (Ü/N) and after 0, 1 , 2 , 3 and 4 hours after IPTG induction.

Figure 3 shows a semi quantitative analysis achieved by a densitrometric scan of the protein samples of the SDS-PA gel shown in Figure 2 at 4 hours post induction. A numeric integration of the proinsulin peak located at the right-end side of the scans revealed that

expression of the recombinant DNA according to the invention yielded in a relative amount of proinsulin of about 10% based upon the total cellular protein, whereas the yield of the comparative construct was about 5 % based upon the total cellular protein and thus significantly lower than in the inventive Example.

As a result it can be seen by the above Experiments that the polypeptide according to the subject invention, i.e. a polypeptide having the helper sequence MAK at the N-terminus of a proinsulin precursor, firstly leads to an increased cell growth, which consequently leads to higher amounts of total cellular protein which can be used for the manufacture of insulin.

Secondly, the polypeptide of the subject invention leads to a higher relative amount of polypeptide comprising the proinsulin precursor when compared to the total cellular protein. The invention thus leads to an increased ratio of the polypeptide according to the subject invention to the total cellular protein. As a consequence, the higher starting amount of proinsulin precursor enables to yield higher amounts of active insulin after purification and processing of the proinsulin precursor.

Example 2

Expression of pMAK-PI in E. coli grown in a fermenter

The plasmid of Example 1 was used for expression in a fermenter with 1 l working volume. Cells were grown in a defined media (Korz, DJ *et al.* (1994), J. Biotech. 39, 59-65), supplemented with amino acids. Cells were grown at 37 °C. Dissolved oxygen was kept above 20 % saturation, pH was kept constant by the addition of aqueous ammonia whenever needed. When the optical density $OD_{600}$ reached a value of 5, protein expression was induced with IPTG. After 4 more hours, the optical density had reached 30, corresponding to a dry cell weight of 9 g/l.

Comparative Example 2

Expression of pMK-PI in E. coli grown in a fermenter

The plasmid of comparative Example 1 was used for expression in a fermenter under the same conditions as described for Example 2. The cell density at induction was the same as in Example 2.

Figure 4 shows a comparison of the relative protein expression in an SDS-PAGE analysis. In this Figure, line "MWM" depicts a molecular weight protein marker for evaluating the size of the expressed proteins. Columns designated with "MAK" show protein samples according to Example 2 and thus depict an example according to the invention, whereas columns designated with "MK" show samples according to comparative Example 2. Shown are total protein samples after 0, 2, 3, 4, 5 and 6 hours after IPTG induction (Example 2) and after 0, 4 and 6 hours (Comparative Example 2).

Figure 5 shows a semi quantitative analysis achieved by a densitrometric scan of the protein samples of the SDS-PA gel shown in Figure 4 at 6 hours post induction. A numeric integration of the proinsulin peak located at the right-end side of the scans revealed that expression of the recombinant DNA according to the invention yielded in an amount of proinsulin of about 8 to 10% based upon the total cellular protein, whereas the yield of the comparative construct was about 5 to 6 % based upon the total cellular protein and thus significantly lower than in the inventive Example.

It can be seen by a comparison of Example 2 with comparative Example 2 that the findings made above by the comparison of Example 1 with comparison Example 1 for the expression of a polypeptide according to the subject invention in shaking cultures equally apply for E. coli cultures grown in a fermenter.

As a result, it can be seen by the Examples according to the present invention that expression of a gene encoding a polypeptide containing a proinsulin precursor and the N-terminal helper sequence MAK leads to an increased growth rate of the heterologous host compared

to the comparative construct, which has the same N-terminal amino acid methionine but which differs in the overall helper sequence. It could be shown by the above experiments that the expression rate of genes encoding polypeptides comprising a proinsulin precursor and having methionine as starting amino acid, but which differ in the second amino acid, is significantly different. This finding is unexpected since for other systems, both alanine and lysine as second amino acid work equally well in enhancing protein expression (Bivona, L *et al.* (2010), Protein Expr. Purif. June 20 (epub ahead of print)).

By increasing the growth rate of the heterologous organism, the DNA according to the subject invention encoding for the helper sequence MAK and human proinsulin leads to an increased yield of recombinant protein. Accordingly, the subject invention enables to yield higher amounts of active insulin after purification and processing the proinsulin precursor due to higher starting amounts of recombinant protein.

As a consequence, it could be shown by the present invention that the growth rate of a heterologous organism expressing a gene encoding a recombinant polypeptide comprising a proinsulin precursor and gene expression said recombinant polypeptide not only depends on the N-terminal amino acid, but can be significantly increased by a helper sequence according to the subject invention.

## Claims

1. A polypeptide comprising,
   i. a first peptidyl fragment consisting of the N-terminal amino acid sequence methionine, alanine and lysine (MAK), and
   ii. a second peptidyl fragment, which is an insulin precursor.

2. The polypeptide of claim 1, wherein the insulin precursor comprises insulin chains A and B.

3. The polypeptide of claims 1 or 2, wherein the insulin precursor is human proinsulin.

4. The polypeptide of any of the previous claims, further comprising a sequence, which is capable of binding to a chromatographic matrix, preferably, an affinity tag.

5. The polypeptide of claim 4, wherein the affinity tag is a His-tag.

6. The polypeptide of claims 1, 2 or 3, consisting of the amino acid sequence of SEQ ID NO: 1.

7. An isolated nucleic acid comprising a nucleotide sequence encoding the polypeptide of any of claims 1 to 6.

8. The isolated nucleic acid of claim 7, wherein said nucleic acid is DNA.

9. The isolated nucleic acid of claims 7 or 8, which employs a plurality of alternative codons to those present in the naturally occurring wild-type human proinsulin coding sequence, said alternative codons causing no amino acid changes from wild-type human proinsulin, wherein at least a portion of said alternative codons are more preferred for usage in bacterial cells.

10. The isolated nucleic acid of claim 9, which consists of the nucleic acid sequence of SEQ ID NO: 2.

11. A recombinant cell containing the nucleic acid of any of claims 7 to 10.

12. The recombinant cell of claim 11, which is *Escherichia coli*.

13. A method of producing a polypeptide comprising growing a recombinant cell containing the nucleic acid of any of claims 7 to 10 such that the encoded polypeptide is expressed by the cell, and recovering the expressed polypeptide from the cell.

14. A method of producing insulin which comprises the steps of
    a. expressing the polypeptide according to any of claims 1 to 6 in *Escherichia coli*,
    b. isolating the polypeptide from the recombinant cell,
    c. subjecting the isolated polypeptide to a folding process which permits correct folding of the insulin precursor, and
    d. subjecting the polypeptide to enzymatic processing to yield active insulin.

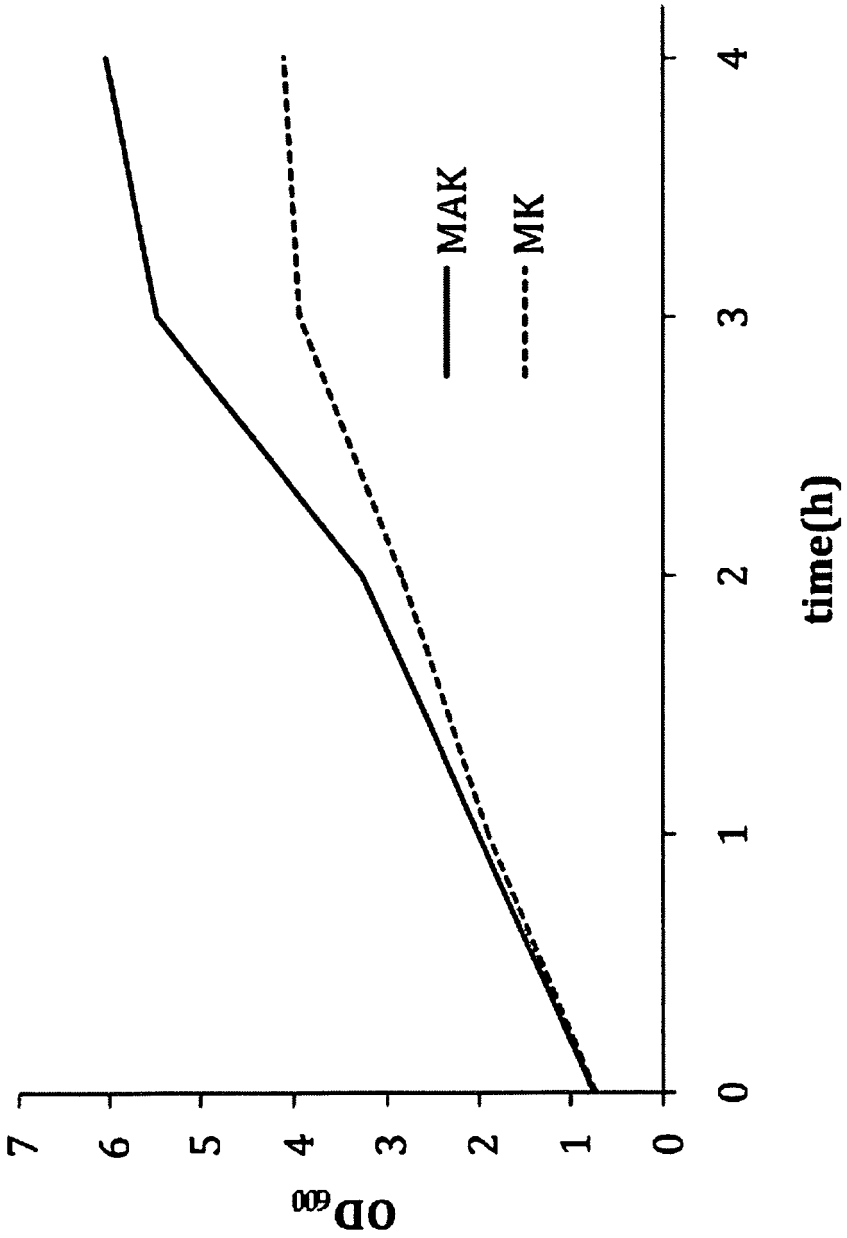



Figure 1

Figure 2

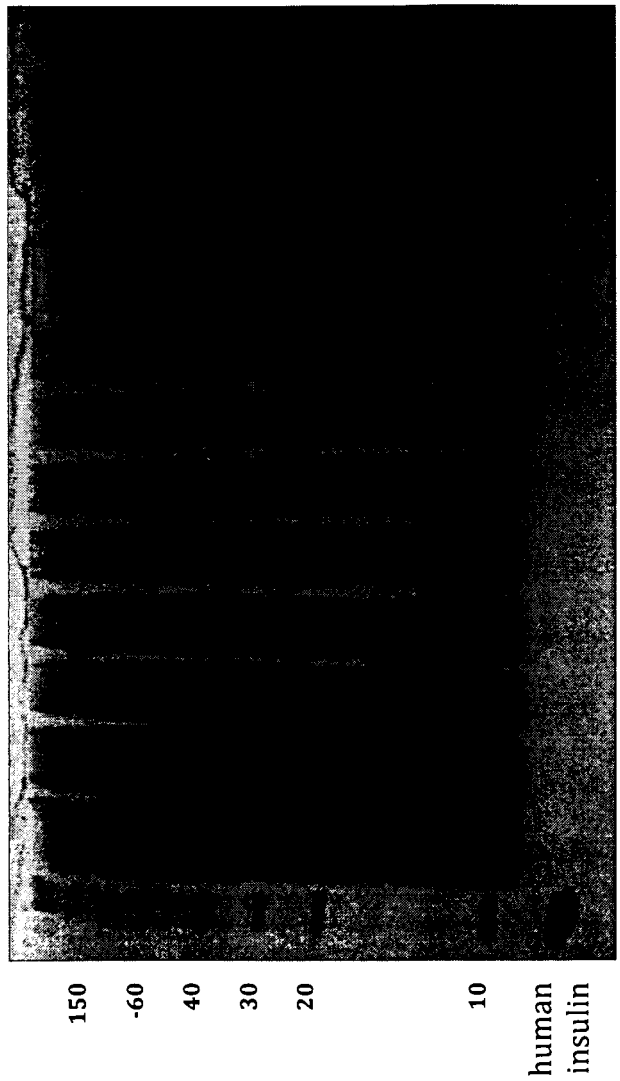


Figure 3

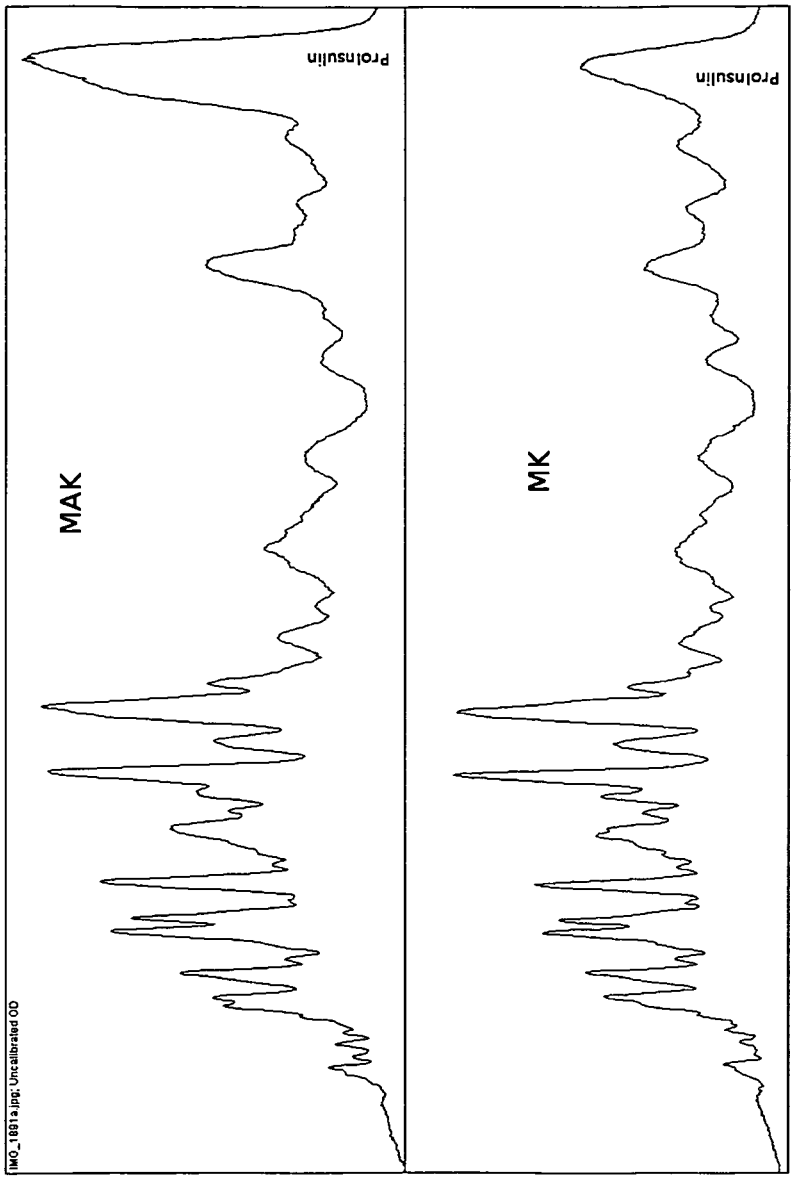

MAK
Prolinsulin
MK
Prolinsulin
IMO_1691a.jpg; Uncalibrated OD

Figure 4

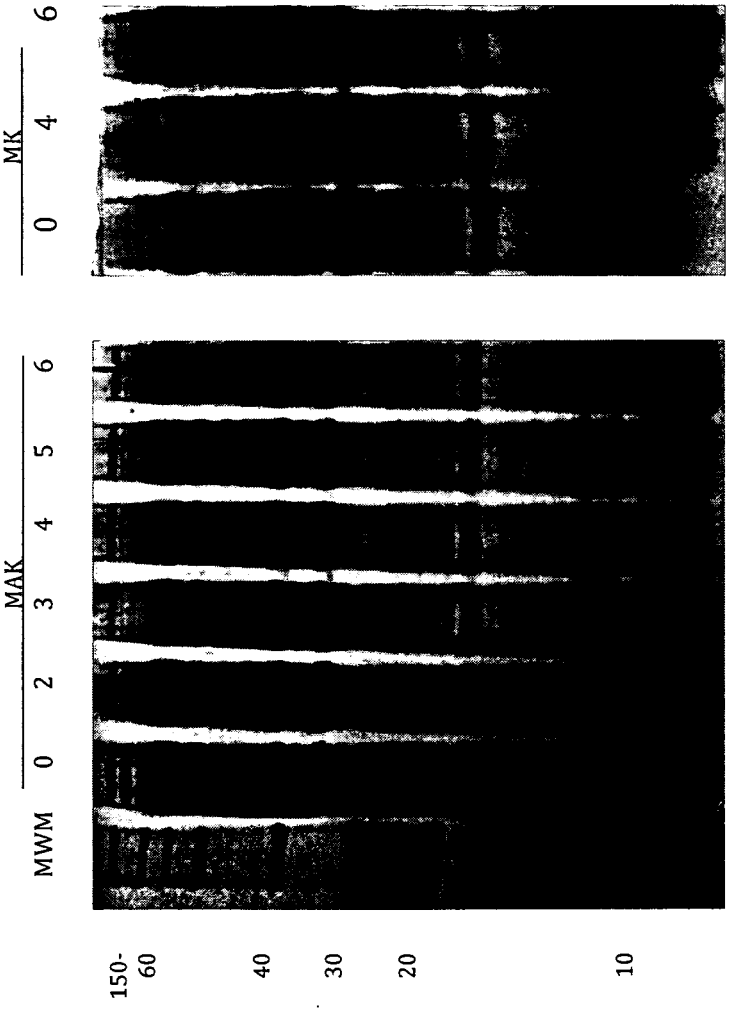
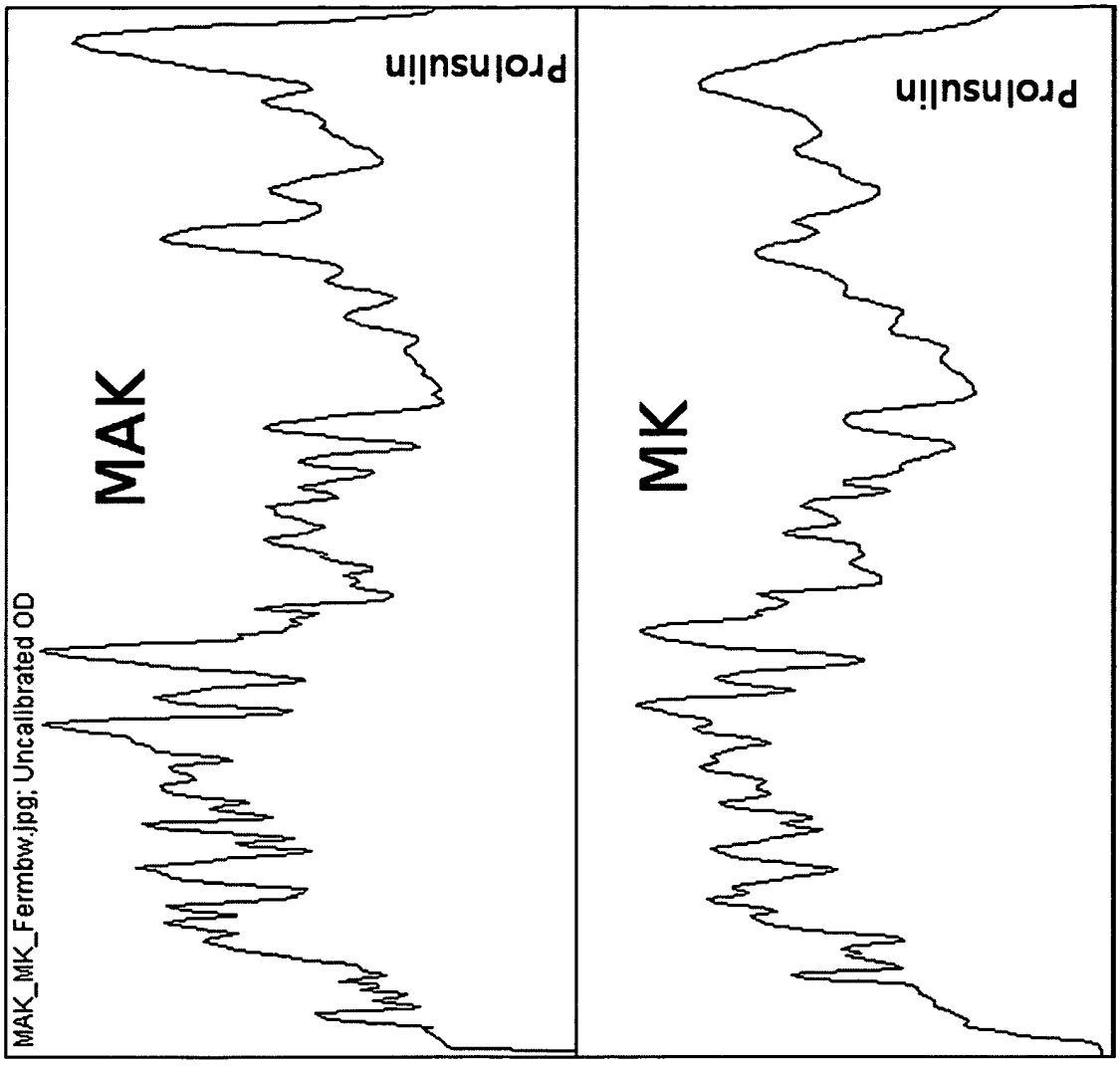


Figure 5


MAK_MK_Fermbw.jpg; Uncalibrated OD
MAK
Prolnsulin
MK
Prolnsulin

# INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2011/005090

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/62 C12N15/62
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)
C07K C12N

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 practical, search terms used)
EPO-Internal, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | EP 0 534 705 A2 (LILLY CO ELI [US])<br>31 March 1993 (1993-03-31)<br>claims; examples 29-32; table 2 | 1-14 |
| A | WO 01/18215 A1 (MAX PLANCK GES E V [DE];<br>JENNE DIETER [DE]; WILHARM ELKE [DE];<br>PARRY M) 15 March 2001 (2001-03-15)<br>claims | 1 |

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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 | Date of mailing of the international search report |
|---|---|
| 14 February 2012 | 22/02/2012 |

| Name and mailing address of the ISA/ | Authorized officer |
|---|---|
| European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016 | Huber, Angelika |

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | LILA R CASTELLANOS-SERRA ET AL: "Expression and folding of an interleukin-2-proinsulin fusion protein and its conversion into insulin by a single step enzymatic removal of the C-peptide and the N-terminal fused sequence", FEBS LETTERS, ELSEVIER, AMSTERDAM, NL, vol. 378, no. 2, 8 January 1996 (1996-01-08), pages 171-176, XP008141838, ISSN: 0014-5793, DOI: 10.1016/0014-5793(95)01437-3 [retrieved on 1999-11-30] abstract; figure 1 ----- | 1-14 |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2011/005090

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---|---|---|
| EP 0534705 A2 | 31-03-1993 | AU 655316 B2 | 15-12-1994 |
| | | AU 2528092 A | 25-03-1993 |
| | | CA 2079027 A1 | 25-03-1993 |
| | | EP 0534705 A2 | 31-03-1993 |
| | | JP 5304978 A | 19-11-1993 |
| | | US 5378613 A | 03-01-1995 |
| WO 0118215 A1 | 15-03-2001 | CA 2403330 A1 | 15-03-2001 |
| | | DE 19943177 A1 | 05-04-2001 |
| | | EP 1214428 A1 | 19-06-2002 |
| | | JP 2003529330 A | 07-10-2003 |
| | | US 2002192754 A1 | 19-12-2002 |
| | | WO 0118215 A1 | 15-03-2001 |