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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 July 2017 (27.07.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/126984 A1

(51) **International Patent Classification**:
C07K 14/62 (2006.01) *C12P 21/02* (2006.01)

(21) **International Application Number**:
PCT/PL2017/050003

(22) **International Filing Date**:
22 January 2017 (22.01.2017)

(25) **Filing Language**: Polish

(26) **Publication Language**: English

(30) **Priority Data**:
P.415888 22 January 2016 (22.01.2016) PL

(71) **Applicant**: **INSTYTUT BIOTECHNOLOGII I ANTYBIOTYKÓW** [PL/PL]; ul. Starościńska 5, 02-516 Warszawa (PL).

(72) **Inventors**: **BOROWICZ, Piotr**; ul. Starościńska 5, 02-516 Warszawa (PL). **PŁUCIENNICZAK, Andrzej**; ul. Starościńska 5, 02-516 Warszawa (PL). **PŁUCIENNICZAK, Grażyna**; ul. Starościńska 5, 02-516 Warszawa (PL). **SOKOŁOWSKA, Iwona**; ul. Starościńska 5, 02-516 Warszawa (PL). **ROMANIK-CHRUŚCIELEWSKA, Agnieszka**; ul. Starościńska 5, 02-516 Warszawa (PL). **ŁUKASIEWICZ, Natalia**; ul. Starościńska 5, 02-516 Warszawa (PL). **ZIELIŃSKI, Marcin**; ul. Starościńska 5, 02-516 Warszawa (PL). **ANTOSIK, Jarosław**; ul. Starościńska 5, 02-516 Warszawa (PL). **SOBOLEWSKA, Agnieszka**; ul. Starościńska 5, 02-516 Warszawa (PL). **MIKIEWICZ, Diana**; ul. Starościńska 5, 02-516 Warszawa (PL). **WÓJTOWICZ-KRAWIEC, Anna**; ul. Starościńska 5, 02-516 Warszawa (PL). **BARAN, Piotr**; ul. Starościńska 5, 02-516 Warszawa (PL). **BIERCZYŃSKA-KRZYSIK, Anna**; ul. Starościńska 5, 02-516 Warszawa (PL). **ODROWĄŻ-SYPNIEWSKI, Michał**; ul. Starościńska 5, 02-516 Warszawa (PL). **TREJCHMAN-MAŁECKA, Bożena**; ul. Starościńska 5, 02-516 Warszawa (PL). **STADNIK, Dorota**; ul. Starościńska 5, 02-516 Warszawa (PL). **STADNIK, Jacek**; ul. Starościńska 5, 02-516 Warszawa (PL). **SURMACZ-CHWEDORUK, Weronika**; ul. Starościńska 5, 02-516 Warszawa (PL). **ZIELIŃSKA, Joanna**; ul. Starościńska 5, 02-516 Warszawa (PL).

(74) **Agent**: **WITEK, Rafał**; WTS Patent Attorneys, Witek, Sniezko & Partners, ul. Rudolfa Weigla 12, 53-114 Wrocław (PL).

(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))*

— *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))*

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

WO 2017/126984 A1

(54) **Title**: A METHOD FOR PRODUCING INSULIN AND INSULIN DERIVATIVES, AND HYBRID PEPTIDE USED IN THIS METHOD

(57) **Abstract**: A method for producing insulin and derivatives thereof is disclosed, as well as a hybrid peptide used in this method.

A method for producing insulin and derivatives thereof, and hybrid peptide used in this method

The subject of the invention is a method for producing insulin and derivatives thereof, as well as a hybrid peptide used in this method.

Diabetes (diabetes mellitus), a metabolic disease (a group of metabolic diseases) of multifactorial etiology, is one of the most dangerous diseases of society and civilization, also sometimes called the epidemic of the XXI century. Due to serious complications, diabetes can lead to death of the patient if left untreated. In the case when chronic hyperglycemia is caused by insufficient secretion of insulin by the body - type 1 diabetes – the therapy requires continuous use of insulin and/or its analogues. The number of diabetic patients is 5-6% of the population, and the annual increase in the number of diabetics is approx. 4-6% depending on the region of the world. The incidence of this disease continues to rise; it is considered a disease of civilization - the current lifestyle greatly contributes to the development of diabetes - for this reason, so far efforts to stop the development of this disease have failed. Also, despite medical advances, insulin and its fast-acting and long-acting analogues form the basis for treatment of type 1 diabetes (10% of the total number of cases of the disease). Insulin and various derivatives thereof used in large quantities in the treatment of diabetes are produced in a large industrial scale, and for this reason production methods are constantly improved in order to increase process yield and lower production costs. The economics of the production process is also important for the growing number of biosimilar insulin manufacturers [e.g.: S. Gough, Practical Diabetes 2013, 30 (4), 146-147a, G. Dranitsaris, E. Amir, K. Dorwart, Drugs 2011, 71 (12) 1527-1536].

In the first years following the discovery of insulin, animal insulin secreted from the bovine or porcine pancreas was used as a human drug. Humanization of those insulins by replacement of the differing amino acids with the synthetic method was not applicable in industrial scale. It was not until the biosynthetic methods solved the problem of patient access to human insulin, and the first process was biotechnological generation of separate recombinant A and B chains of human insulin, and then assembling them into the same hormone as produced endogenously in the human pancreas (D. V. Goeddel *et al.*, Proc. Nat. Acad. Sci., 76: 106-110, 1979). Since the end of the last century, biosynthetic human insulin and analogues thereof, produced by the transformation of a single chain fusion protein, are the main source of this hormone as an exogenous drug. It is produced in recombinant bacterial systems, particularly in the cells of *Escherichia coli* (Polish patent No. PL 127 843), or yeast (L. Thim *et al.*, Proc. Nat. Acad. Sci. USA, 83: 6766-6770, 1986). The method of producing these drugs is generally described in numerous patent specifications and scientific publications and is based on the overexpression of a gene encoding preproinsulin, i.e. a hybrid polypeptide consisting of a leader protein and proinsulin, i.e. the insulin B-chain (or derivative thereof), a linker peptide and the insulin A-chain (or derivative thereof). Once the fusion protein is

isolated, the next manufacturing step is removal of the leader protein and the linker peptide, followed by isolation of the pure hormone. One example of such a process is a method of producing human insulin as described in Polish patent No. PL180818, and these methods are being continuously improved (e.g. Polish patents: PL167810, PL168114, PL177002, PL178466, PL180968, PL183284, PL191901, PL196626, PL198190, PL203195, PL203254, PL210437 and Polish patent applications P.309882, P.310007, P.320644, P.356005, P.374949, P.385586 and P.391975).

The biosynthetic human insulin and analogues thereof are the primary therapeutics in the treatment of type 1 diabetes and have significant value in the treatment of type 2 with different glucose control issues. Although the treatment with human insulin has brought satisfactory results in most cases for over 30 years, a number of analogues were created for improved metabolic control of diabetes due to their faster or prolonged action. For several years, the sales of insulin analogues have exceeded the half of the global insulin market, while patent protection of those drugs is beginning to expire. This situation allows entering the medical markets by new manufacturers of such insulins as biosimilar drugs, in accordance with the regulations of the European Union and the United States. For these reasons, efficient biosynthesis methods are becoming important, to provide high expression of the proteins of interest.

Polish patent No. PL180 818 describes an efficient method to produce recombinant human insulin using *E. coli* bacterial strain lacking the *cytR* repressor (described in European patent application No. EP.0303972) transformed with a plasmid with *deo* operon containing the coding region for the hybrid protein (precursor) including a leader peptide, 62 amino acids long derived from the N-terminus of modified human superoxide dismutase (CuZnSOD) preceded by the N-terminal amino acid Met and ending with the C-terminal Arg residue linking it to the insulin B-chain. The insulin B-chain is linked to the A-chain with a short linker peptide composed of the known linkers Lys-Arg (European patent specification No. EP195 691) or Arg (European patent specification No. EP347 781).

The aim of the invention was to provide an efficient process for producing insulin and analogues thereof that would exceed by several percent the yield of the process using the peptide C as a linker protein, as well as the hybrid proteins that are used in this method.

The subject of the invention is a polypeptide having the amino acid sequence of the formula:

X_n-B-Arg-Arg-A

where:

A is a polypeptide of the insulin A-chain or analogue thereof, preferably a sequence selected from SEQ. ID No.: 1-4.

B is a polypeptide of the insulin B-chain or analogue thereof, preferably a sequence selected from SEQ. ID No.: 5-7.

n is 0 or 1,

X is a leader protein polypeptide, preferably a sequence selected from SOD of SEQ. No.: 8 or UBI of SEQ. No.: 9.

Preferably, the polypeptide of the invention has an amino acid sequence selected from: SEQ. ID No.: 10-27.

The polypeptide of the invention is intended for the production of human insulin by expressing the polypeptide in a strain of *Escherichia coli* lacking *deo* repressors and/or *deo* (*cytR*) gene containing DNA encoding said polypeptide, wherein said DNA is present in a vector under the control of *deo*P_1P_2 promoter, while the hybrid polypeptide comprises a insulin B-chain covalently bound to the A-chain via Arg-Arg dipeptide.

Thus, a further subject of the invention is a method for producing human insulin comprising microbial expression of a recombinant protein containing insulin characterized in that it comprises:

(a) obtaining a hybrid polypeptide comprising proinsulin by expression of the hybrid polypeptide in a cell of an *Escherichia coli* strain lacking *deo* repressor and/or *deo* (*cytR*) gene containing DNA encoding the hybrid polypeptide, wherein the DNA is present in a vector under the control of *deo*P_1P_2 promoter, while the hybrid polypeptide comprises the insulin B-chain covalently bound to the A-chain via Arg-Arg dipeptide,

(b) recovering of the hybrid polypeptide,

(c) folding and formation of disulfide bonds within the hybrid polypeptide obtained in (b), avoiding prior subjection of the hybrid polypeptide to sulfitolysis,

(d) subjecting the hybrid polypeptide obtained in (c) folded, with disulfide bonds formed, to enzymatic digestion in order to produce insulin,

(e) purifying the insulin so produced.

The human insulin A-chain and the human insulin B-chain according to the invention is intended to mean both a polypeptide having the naturally occurring amino acid sequence and a sequence of its known analogue.

Preferably, the hybrid peptide is a peptide of the invention as defined above.

Preferably, at the step (a) the expression is performed in *E. coli* strain of the genotype F^- *ara Δ(pro lac) rpsL thi cytR recA* λ^- devoid of active *cytR* repressor protein.

Preferably, at the step (a) a plasmid *pIBA* of the sequence SEQ ID NO: 28 is used as the vector.

Preferably, step (a) comprises fermentation in the presence of glucose, glycerol or galactose.

Preferably, step (b) comprises:

(i) destruction of the cell wall of the bacterial cell or its fragments to form a lysate, preferably by the action of lysozyme and Triton X100 followed by sonication or disintegration;

(ii) isolation of intracellular inclusion body precipitate from the lysate by centrifugation; and

(iii) dissolving the inclusion body precipitate.

Preferably, step (c) comprises incubating the hybrid polypeptide, preferably in the presence of ascorbic acid, in particular at a concentration of about 2 moles per mole of SH groups present in the mixture, at a temperature of 4 to 37°C for about 1 to 30 hours, preferably about 5 hours, at a pH of 8.5 to 12, preferably at a pH of 11.0 to 11.25.

Preferably, the step (d) further comprises:

(i) adjusting the pH to about 8.8-9.0; and

(ii) digestion of the hybrid polypeptide with trypsin or optionally with deubiquitinating protease, e.g. UBP1ΔC or UBP1ΔC2, and subsequently optionally with carboxypeptidase B at a temperature of 16-37°C for 30 minutes to 16 hours.

Preferably, step (e) further comprises purifying the solution by low-pressure liquid chromatography on Sepharose Q prior to trypsin digestion.

Preferably, step (e) further comprises purifying the solution by low-pressure chromatography on DEAE-Sepharose prior to carboxypeptidase B digestion.

Preferably, step (e) further comprises purifying the resulting insulin by high-performance liquid chromatography, preferably to a purity of at least 98%, followed by crystallization, filtration and drying.

Preferably, the insulin purified at the step (e) is a protein selected from the group comprising: human insulin, human insulin $Lys^{B28}Pro^{B29}$ (lispro insulin), insulin $Gly^{A22}Arg^{B31}$ (GR insulin), human insulin $Ser^{A22}Arg^{B31}$ (SR insulin), insulin $Ser^{A22}Lys^{B3}Arg^{B31}$ (SK3R insulin), proinsulin B-ArgArg-A:dezAsn^{A21}Gly^{A21}.

Detailed description of the invention

As part of the search for high yield expression systems of insulins, studies have been planned on the effects of the type of linker peptide linking the A and B chains (proinsulin) on the yield of insulin and its analogues in the gene system described in the Polish patent No. PL 180 818.

The recombinant preproinsulin according to the invention is understood as a connection of proinsulin and an additional leader polypeptide, for example ubiquitin or superoxide dismutase (SOD), or fragments thereof. In the light of recognized terminology, the recombinant proinsulin is understood as a polypeptide chain, wherein the A and B chains of human insulin or an analogue thereof are connected to any (poly)peptide (linker), that allows them to fold by creating two disulphide bonds between the chains A and B and the third one - within the chain A.

The recombinant insulin according to the invention is meant human insulin or its recombinant analog of hypoglycaemic activity.

According to the research results, the particularly efficient method to produce insulin and analogues thereof according to the invention, comprising the microbial overexpression of a gene encoding the recombinant hybrid polypeptide, comprising proinsulin, in cells of *Escherichia coli* bacterial strains

lacking the *cytR* repressor transformed with an appropriate vector with relevant gene inserted under the control of a *deoP1P2* promoter, consists in that the proinsulin produced comprises the insulin B-chain covalently bound to the A-chain via Arg-Arg dipeptide as a linker peptide. In order to perform the studies, pIBA plasmid has been prepared on the basis of patent specification No. PL 180 818, containing a region encoding the *deoP1P2* promoter and a gene fragment of human superoxide dismutase (SOD, 62 aa) or ubiquitin (UBI) as described in U.S. patents 8,158,382 and 8,956,848, as well as by A. Wojtowicz *et al.*, Microbial Cell Fact. 2005, 4: 17 and Microb Cell Fact. 2014; 13 (1): 113. Further, a fragment encoding the relevant proinsulin is in the same reading frame – optionally modified B-chain, the linker peptide and optionally modified A-chain. The regulatory elements of the plasmid are derived from pBR322 vector (ATCC 31344). The *deoP1P2* promoter is derived from the bacterial *deo* operon composed of four closely related genes that regulate nucleotide and deoxynucleotide catabolism in *Escherichia coli* bacteria; its sequence was amplified from chromosomal DNA of *E. coli* K12 strain based on the nucleotide sequence from gene database [GenBank: AP009048]. Based on the strain of *E. coli* CSH50 (ATCC: 39111; Cheng *et al.*, Gene, 14 Suppl. 1-2: 121-130, 1981), the *E. coli* IBA strain was derived with the *F⁻ ara Δ(pro lac) rpsL thi cytR recA λ⁻* genotype, devoid of active *cytR* repressor protein by introducing a mutation into the gene using the method described in European patent application No. EP0303972. After transformation of *E. coli* IBA host strain with the pIBA plasmid, a system was obtained as described in Polish patent PL 180 818 to express the encoded fusion protein in a *pIBAINS* vector under the control of the *deoP1P2* promoter, which is negatively controlled by a chromosomal repressor protein, the product of *cytR* gene (K. Hammer-Jesperson *et al.*, Molec. Genet. 1975; 137: 327-335). The construction of plasmids used elements-DNA fragments encoding various leader proteins and various insulin derivatives, wherein the (pro)insulin A-chain was linked to the insulin B-chain via various linker peptides – examples of which are given below.

The production yield of insulin and its analogues in this expression system and with different linker peptides was compared with expression of these hormones with other strains of *E. coli* and a vector which is derived from the pIGAL plasmid (Gene Bank AY 424310), investigating in particular the relationship between the yield and the type of peptide linker.

The linker peptide sequences used were:

1. Lys-Arg
2. Arg
3. Arg-Arg
4. C-peptide

According to the invention the thoroughly investigated insulin and its derivatives-analogues were: recombinant human insulin (Polish patent specification No. PL128 599)

$Lys^{B28}Pro^{B29}$ human insulin (lispro insulin, Polish patent specification No. PL180 968)

$Gly^{A22}Arg^{B31}$ human insulin; (Polish patent specification No. PL219335 or US 08,618,048, GR insulin),

$Ser^{A22}Arg^{B31}$ human insulin (description of Polish patent application P.219335, SR insulin),

$Ser^{A22}Lys^{B3}Arg^{B31}$ human insulin (description of Polish patent application P.399287, SK3R insulin) and preproinsulin (precursor) G: leader protein – B-chain – linker – A-chain: dezAsnA21GlyA21.

The *E. coli* IBA was used as the *E. coli* host strain to study the yield; the essence of its properties is the inactive gene of *cytR* repressive protein, which is disclosed and described in the literature (e.g. Miller J H., Experiments in molecular genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1972) or in U.S. patent description 4,480,038 or Polish patent description PL 180 818 for *E. coli* strains deposited under accession numbers ATCC 69361 and 69363. Other strains widely used in research – *E. coli* DH5, *E. coli* DH5alpha. and *E. coli* HB101 – were used for comparative purposes.

The pIBA vector – *deoP1P2* promoter, tetracycline resistance, terA Trp transcription terminator – was used as the plasmid for protein expression in the yield studies, but for comparative purposes also other vectors were used: pIGAL1 derivative (Gene Bank AY 424310) – *pms* promoter, ampicillin resistance, ST1 transcription terminator; pIGAL1 derivative – *pms* promoter, tetracycline resistance, terA Trp transcription terminator and pIGAL1 derivative – *deo P1P2* promoter, tetracycline resistance, ST1 transcription terminator. A modified fragment of SOD, as in Polish patent specification No. PL 180 818 and a modified ubiquitin (UBI), as described in U.S. patent specification 8,158,382 and 8,956,848, were primarily used as the leader protein in the yield studies. In the present description, the terms SOD and UBI represent compounds described respectively in these patent specifications. The process for producing each of the insulin which is the subject of the present application is carried out with a classical method of genetic engineering and biotechnology. It consists in that a suitable strain is constructed, which in a process of biosynthesis (fermentation) produces the desired hybrid polypeptide (relevant preproinsulin), which is then transformed into the desired product, purified and isolated.

The system is universal to produce recombinant human insulin and its studied – slightly different – analogues. For this purpose modifications of the recombinant human preproinsulin gene were constructed using genetic techniques, such as site-specific mutagenesis reaction. Point mutagenesis reaction was carried out using a kit from Stratagene (Cat. No. 200518-5); plasmid DNA of pIBA and pIGAL plasmids was used as the template. Also any other DNA comprising the relevant sequence coding for recombined human proinsulin or preproinsulin can be used as the template.

For each insulin plasmid constructs – vectors – were prepared that encoded the studied (pre)proinsulin with a linker (poly) peptide, and elements-DNA fragments encoding various proinsulins were used for the plasmid construction – consisting of various linker peptides and of various insulin derivatives.

These vectors were used to transform competent cells of a suitable *Escherichia coli* strain, for example IBA, DH5α, DH5 or HB101, and it is possible to use cells of other *E. coli* strains or cells of other microorganisms, or other known cell lines suitable for recombinant protein expression. The plasmid containing the specific gene modification of recombinant human preproinsulin was isolated and sequenced to verify the correct transformation and nucleotide sequence. Single clones of the transformed strains were typically cultured in appropriate medium supplemented with selection antibiotic to prepare the bacterial material for the research bank, and the samples of bacterial cultures and 40% glycerol in the ratio 1:1 were deposited at -70°C. All research bank strains were subjected to extensive microbiological, biochemical and physicochemical testing, which confirmed their stability and compatibility of their properties with the requirements applicable to industrial banks. The course of subsequent biosynthesis was based on Polish patent specification No. PL 180 818. Following biosynthesis, the variants of the recombinant preproinsulin produced in *E. coli* strains were isolated after cell disruption as inclusion bodies, which were separated. Hybrid polypeptide with insulin or analogue thereof (corresponding preproinsulin) was isolated from the inclusion bodies, and then subjected to folding (renaturation). The resulting solution of folded hybrid protein was subjected to the controlled action of trypsin, similarly as in the case of a number of methods known and described previously (e.g. by Kemmler *et al.*, J. Biol. Chem., Vol. 246, pp. 6786-6791 (1971) or patent specifications US 6,686,177 or US 6,100,376), or deubiquitinating protease, as described in patent specifications US 8,158,382 and US 8,956,848, as well as by A. Wojtowicz *et al.*, Microbial Cell Fact. 2005, 4: 17 and Microb Cell Fact. 2014; 13 (1): 113, or both of these enzymes. In some cases (human insulin, lispro insulin) additional amino acids of the linker peptide were removed by carboxypeptidase B action; in other cases, this step was omitted. The resulting insulin and its analogues were subjected to a process of purification using known methods, mainly low-pressure chromatography, ultrafiltration, and HPLC. From the sufficiently purified – meeting pharmacopoeial requirements – solution of insulin or its analogue, the product was crystallized and dried.

In order to compare the expression yield for each host and each variant of the system, the biosynthesis, isolation and purification of the product was developed and optimized for each strain. The biosynthesis of *Escherichia coli* strains lacking the *cytR* repressor transformed with a suitable vector with relevant gene inserted under the control of appropriate *deoP1P2* promoter was carried out according to the description of Polish patent specification No. PL 180 818, in fermenters of 7.5 dm^3 in volume, using approx. 3-5% of the inoculation material and minimum, industrial medium for inoculum and production culture and controlling the induction of a desired product with glucose, according to example 2 of Polish patent specification No. PL 180 818. The biosynthesis of other *E. coli* strains was performed in accordance with the general methods of their culture. Inoculum culture was carried out for approx. 10 hours, until the end of the logarithmic phase of bacterial growth, and

production culture for approx. 16 hours. The subsequent process of product separation, transformation, and purification was also conducted as described in Polish patent No. PL 180 818 – optimized example 2 and 3. In particular, on completion of the production biosynthesis the resulting broth was cooled to approx. 10°C, *E. coli* cells were centrifuged, subjected to action of lysozyme and Triton X100, and then sonicated with ultrasound or disintegrated with pressure at 5-10°C. The inclusion bodies were centrifuged again, washed, and then subjected to renaturation, as generally described in example 3A of the Polish patent PL 180 818. Except for the proteins obtained according to Example 6, the further optimized process consisted in subjecting the solution to low-pressure liquid chromatography on DEAE Sepharose following citraconylation and digestion with trypsin or digestion with deubiquitinating protease and trypsin. After decitraconylation and precipitation and dissolution of the protein, the solution was subjected to another low-pressure liquid chromatography using Sepharose Q resin, followed by optional digestion of the linker peptide with carboxypeptidase B. After the final high-performance chromatography on Kromasil C8-18, the product purified at least to 98% purity was subjected to crystallization, filtration and drying. Detailed description of the entire process, completely comparable between specific insulins, is given in the examples. The process was performed in a pilot scale, using the methods and devices used in industry. After process optimization, three more batches of each of insulin were carried out to study the yield of the process, and the presented results are the average of each of them.

To compare the expression yield, at a fixed production process, analytical tests of five operations were selected from the unified process control:

(1) optical density (OD_{600}) of completed production culture;

(2) weight of the culture (isolated bacteria), expressed in g per 1 dm^3 of production culture (wet and/or dry);

(3) weight of the isolated inclusion bodies, expressed in g per 1 dm^3 of production culture (wet and/or dry or as dry weight of the slurry in a buffer);

(4) total protein (after precursor folding) as determined by a spectrophotometric method and expressed in AU/dm^3 or HPLC and expressed as % folded precursor in the total protein;

(5) weight of the purified final product expressed in mg per 1 dm^3 of production culture.

In the case of lispro insulin the analytical results of most steps are given – total protein determined by spectrophotometry and expressed in AU/dm^3 and the Bradford method and expressed in grams per 1 dm^3 of production culture.

A general block diagram of the production of insulin and its derivatives is shown in Scheme 1.

The invention discloses that the final yield of biosynthetic human insulin and its analogues, using the optimized technological process of biosynthesis, secretion, transformation, and purification, depending on the degree of overexpression of the gene encoding preproinsulin in the modified

Escherichia coli cells (with inactive *cytR* repressors that negatively regulate *deo* operon) transformed with a specific plasmid (with *deoP1P2* promoter) surprisingly depends primarily on the peptide linker used, which links the A and B chains of insulin.

According to the invention it was surprisingly found that the highest yield in this system is obtained when the peptide linker is Arg-Arg dipeptide.

Through the invention, specific conditions for preparation of human insulin and its analogues by genetic engineering are given, using the microbial overexpression of the relevant preproinsulin in the cells of Escherichia coli strains devoid of *deo* repressors and/or *deo* gene, containing plasmid DNA coding for an appropriate hybrid polypeptide in a vector under the control of *deoP1P2* promoter at the optimum performance – significantly increased when compared to the published data.

In the following embodiments of the invention the structure, yield and purity of all products were routinely determined by known methods of HPLC and UV spectrophotometry; to fully characterize the produced insulins, numerous modern methods and analytical techniques were used, including MS, CIEF, LC-MS/MS, GC, peptide mapping, NMR, crystallography and powder X-ray diffraction, CD, FTIR, ELISA or PCR and sequencing of amino acids and nucleotides.

In order to better explain the principle of the invention, the present description is supplemented with a detailed discussion of the exemplary embodiments, including the drawings, showing the general structure (Fig. 1) and the sequence (Fig. 2) of the *pIBA* plasmid, both ready for insertion of a hybrid protein gene and as an exemplary embodiment in the form of plasmid containing a gene encoding the insulin hybrid protein (Fig. 3 and Fig. 4).

No detailed descriptions of conventional bioengineering methods used in the construction of specific genes, plasmids, vectors with inserted genes encoding studied preproinsulins or methods of transformation of bacterial strains with these vectors, were given in the examples, as well as no methods for testing the stability of these strains. Such methods are well known to those skilled in the art and used in biotechnological laboratories, as well as described in numerous publications, of which the key ones are provided herein. The final products produced with method according to the invention are known insulins, most of the specific bioengineering methods for the construction of these recombinant bacterial strains are disclosed in the Polish patent PL180968 and Polish patent applications P.385586 and P.399287.

No detailed descriptions of standard analytical methods used for determination of content and concentrations of proteins and insulins at the specific manufacturing steps are given, since these methods are also well known to relevant skilled in the art and routinely used for protein determination.

The following examples, not limiting the scope of the invention, describe in detail the research methodology that allows in a fully controlled way to compare the obtained representative research results on the relationship between the process yield and the structure of the linker peptide.

Examples

I. General procedure of the process

Step 1. Research Bank – primary culture

Primary cultures of *E. coli* strain IBA, DH5 DH5alpha and HB101 containing the pIBA vector and pIGAL1 derivative as protein expression plasmids (Gene Bank AY 424310) were performed according to the requirements (ICH Topic Q 5 D: Quality of Biotechnological Products: Derivation and Characterization of Cell substrates Used for Production of Biotechnological/Biological Products and ICH Topic Q 5 B: Quality of Biotechnological Products: Analysis of the Expression Construct in Cell Lines Used for Production of r-DNA Derived Protein Products) in the basal medium:

The basal medium:

– casein hydrolyzate 20 g

– yeast extract 10 g

– NaCl 5.0 g

– water to a volume of 1 dm^3

– antibiotic (tetracycline or ampicillin) to a final concentration of 12.5 $\mu g/cm^3$ for TET or 100 $\mu g/cm^3$ for Amp

The flask with the basal medium and antibiotic was inoculated with a single bacterial colony obtained after transformation and shaken (220 rpm) at 37°C to an optical density OD_{600} of approx. 1.

The resulting culture was mixed with the freezing medium (1:1) and frozen at -70°C in 1 ml aliquots.

The freezing medium:

– K_2HPO_4 6.3 g

– KH_2PO_4 1.8 g

– sodium citrate x $2H_2O$ 0.57 g

– $MgSO_4$ x $7H_2O$ 0.09 g

– $(NH_4)_2SO_4$ 0.9 g

– glycerol 44.0 g

– water to a volume of 450 cm^3

For the resulting bacterial culture the following assays were performed:

• measurement of the optical density OD of the culture at 600 nm;

• checking the homogeneity of bacterial culture by preparation of live specimen.

Bacterial culture purity, morphological characteristics – appearance, shape of the cells – were evaluated. Observation of the specimen using Olympus CX 40 microscope;

• calculating the most probable number of colony forming units on the TSA medium and TSA with a specific antibiotic – cfu/ml;

• checking microbiological purity of the culture using Sabouraud media (to identify fungal contamination) and TSA (to identify contamination by other bacteria).

Step 2. Inoculum

The inoculum was prepared in flasks of 250-300 ml in volume filled with 50 ml of inoculum medium. To each flask the following sterile solutions were added:

• antibiotic (tetracycline or ampicillin, 12.5 mg/ml)	50 µl
• glucose 50%	500 µl

Additionally added to the cultures of *E. coli* IBA were:

• proline (100 mg/ml)	300 µl
• thiamine (50 mg/ml)	2-5 µl.

After inoculation of the flasks with primary culture (working bank), the inoculum was incubated at 30-37°C for 16-18 hours to OD_{600} approx. 0.7-1.0 on a shaker at 180-200 rpm.

The inoculum medium (components for 1 dm^3):

– K_2HPO_4	9 g
– KH_2PO_4	1 g
– NaCl	5 g
– NH_4Cl	1 g
– $MgSO_4 \cdot 7H_2O$	0.4 g
– iron-ammonium citrate	0.01 g
– yeast extract	0.2 g
– trace element solution	1 ml

• $MnSO_4 \times H_2O$	1 g/dm^3
• $ZnSO_4 \times 7\ H_2O$	2.78 g/dm^3
• $CoCl_2 \times 7\ H_2O$	2 g/dm^3
• $Na_2MoO_4 \times 2\ H_2O$	2 g/dm^3
• $CaCl_2 \times 2\ H_2O$	3 g/dm^3
• $CuSO_4 \times 5\ H_2O$	1.85 g/dm^3

- H_3BO_4 0.5 g/dm^3
- HCl (32%) 100 ml/dm^3

Step 3. Production culture (NewBrunswick BioFlo 310 fermenter, 7.5 dm^3 (starting volume 4 dm^3)

Initial culture parameters:

- Temperature: 37°C
- pH: 7.1 ± 0.01
- DO: 30-35%
- rotations: 250 rpm
- air flow: 5 dm^3/min.

To the sterile fermenter with 4 dm^3 of the production medium and antifoaming agent Antifoam 204 (0.5-1 ml, Sigma) sterile solutions were added:

- 6 g of $MgSO_4 \times 7H_2O$ in 80 cm^3 of water
- 0.2 g $CaCl_2 \times 2H_2O$ in 60 cm^3 of water
- glucose 50%, 4 cm^3

Additionally added to the *E. coli* IBA cultures:

- proline (100 mg/ml) 24 cm^3
- thiamine (50 mg/ml) 200-500 µl.

The production medium:

– K_2HPO4 8 g/dm^3

– KH_2PO4 2 g/dm^3

– NH_4Cl 3 g/dm^3

– iron-ammonium citrate 0.15 g/dm^3

– K_2SO_4 0.9 g/dm^3

– sodium citrate 3.0 g/dm^3

– trace element solution 3 cm^3

- $MnSO_4 \times H_2O$ 1 g/dm^3
- $ZnSO_4 \times 7\ H_2O$ 2.78 g/dm^3
- $CoCl_2 \times 7\ H_2O$ 2 g/dm^3
- $Na_2MoO_4 \times 2\ H_2O$ 2 g/dm^3
- $CaCl_2 \times 2\ H_2O$ 3 g/dm^3
- $CuSO_4 \times 5\ H_2O$ 1.85 g/dm^3
- H_3BO_4 0.5 g/dm^3
- HCl (32%) 100 ml/dm^3

Medium in the fermenter is inoculated with inoculum and cultured in appropriate conditions. 40% glucose solution, as a source of carbon and energy for biomass production, is continuously fed into the fermenter (supplemented with proline 10-30 mg/ml in the case of *E. coli* IBA) so that its concentration is 70-180 mg/dm^3.

pH (7.1 +/- 0.1) is controlled by cascade using 16% ammonia. The oxygen level in the medium is maintained at a level not lower than 30-35%. This parameter is controlled depending on the reading from DO electrode by the increase in impeller speed from 250 to 1000 rpm and in air flow from 5 to 10 dm^3/min. After reaching the maximum speed of the impeller and the maximum air flow (approx. 7-10 hours) the addition of glucose is stopped; OD_{600} value (in the case of *E. coli* IBA/pIBA) is approx. 25-35.

After a decline in glucose concentration, its addition is maintained so that the concentration is approx. 35-50 mg/dm^3 and the pH was kept at 7.1 +/- 0.1. This phase of the biosynthesis is carried out until the cells reach the stationary growth phase (increase in optical density of the culture is no longer observed) and lasts approx. 7-8 hours. At the end of the recombinant protein production phase OD_{600} is 50-60 in the case of *E. coli* IBA/pIBA strains. The culture is then cooled to a temperature below 20°C, preferably approx. 10°C.

Step 4. Centrifugation

The biomass is centrifuged using a fixed centrifuge at 6,000-8,000 rpm for 15 minutes and at a temperature of 4°C. The fermented broth may be centrifuged using a flow centrifuge.

Step 5. Isolation of inclusion bodies

The biomass is resuspended in 1 dm^3 of 50 mM Tris buffer, 0.5 M NaCl and 5 mM beta-mercaptoethanol, pH 7.5, and lysozyme is added to a concentration of 0.43 mg/cm^3 and 0.2 M EDTA to a concentration of 0.5%, then incubated for 30 minutes at room temperature and subjected to disintegration by sonication or pressure.

Step 6 - option 1: Sonication.

The material is cooled to approx. 10°C, divided into 4 parts, Triton is added to 1% and each portion is sonicated for 30 minutes with the amplitude of 33%. The sonicated sample is centrifuged at 8,000 rpm for 15 minutes at room temperature. The pellet is then resuspended in 1 dm^3 of 50 mM Tris buffer, 0.5 M NaCl and 5 mM beta-mercaptoethanol, pH 7.5 and sonicated again for 10 minutes. The inclusion bodies are centrifuged at 8,000 rpm for 15 minutes at room temperature.

Step 6 - option 2: Pressure disintegration.

The material is cooled to approx. 10°C, subjected to pressure disintegration (Panda+1000, GEA Niro Soavi) at a pressure of 800 bar at a rate that ensures total destruction of the cells (microscopic inspection), and cooling the suspension so that the temperature does not exceed approx. 10°C.

Step 7. Dissolution of inclusion bodies

The inclusion bodies were dissolved in adequate amount of 12 mM bicarbonate buffer and 0.2 mM EDTA to obtain absorbance values below 9 at a wavelength of 280 nm. The solution is adjusted to a pH value of 11.5 ± 0.5 with 2 M NaOH and stirred for 30 minutes at room temperature. Then the pH is adjusted to 10.8 ± 0.4 with 2 M HCl. The solution is clarified by centrifugation at 12,500 rpm for 15 minutes at 4°C.

Step 8. Renaturation

After dissolving the inclusion bodies, the protein is renatured by aeration – vigorous stirring for approx. 16-18 h at room temperature. The pH is then adjusted to pH 9.0 with 2 M HCl.

Step 9. Citraconylation

Citraconic anhydride is added portionwise to the protein solution in an amount calculated according to the formula: anhydride volume (cm^3) = 0.15 x solution volume [dm^3] x A_{280} and 2 M NaOH to maintain the pH in the range of 8.7-9.3. After adding the entire amount of the anhydride, the solution is stirred for 1 hour, then solution of 2 M ethanolamine in the volume of three times anhydride volume is added, followed by stirring for 30 minutes.

Step 10. Trypsination

After citraconylation the pH of the solution is adjusted to a value of 8.8 with 2 M HCl and diluted with water so that the conductivity reaches less than 5 mS. Then 1% trypsin solution is added in an amount calculated according to the formula:

volume of trypsin = A_{280} x volume of a solution of [dm^3] / 95)

and stirred for 16-18 hours at room temperature. The reaction is inhibited with aprotinin solution with a concentration of 1 mg/ml, added in an amount 21-times less than the volume of trypsin.

Step 11. Cutting off the leader protein - ubiquitin.

20 mM phosphoric acid is added to the protein solution at a temperature of 10°C to obtain a pH value of 7.5, and then at 37°C 19.2 cm^3 UBP1ΔC protease solution is added at a concentration of 1 mg/cm^3 and stirred for 1 hour. After cooling to 4°C the solution is clarified by centrifugation at 12,000 rpm for 15 minutes.

Step 12. The low-pressure chromatography on a bed of DEAE Sepharose

The solution at pH 8.6 ± 0.2 is loaded onto a column packed with DEAE resin (200 cm^3) and equilibrated with 0.5 M Tris, pH 8.6 ± 0.2 followed by 20 mM Tris, pH 8.6 ± 0.2. After loading, the column is washed with buffers at pH 8.6 ± 0.2: 20 mM Tris with NaCl until conductivity is 5 ± 1 mS, and then 20 mM Tris with NaCl until conductivity is 2 ± 1 mS and with 20 ± 5% isopropanol. The insulin protein is eluted with a buffer at pH 8.6 ± 0.2, consisting of 20 mM Tris with NaCl until conductivity is 5 ± 1 mS and supplemented with 25 ± 5% isopropanol.

Step 13. Decitraconylation

The main fraction eluted from the column is diluted 2 times, cooled to 4°C and acidified with 0.1 M HCl to a pH of 2.9 ± 0.2, then stirred 10 hours.

Step 14. Precipitation of insulin with zinc chloride.

A volume of 1 M zinc chloride, which corresponds to 1/50 of the volume of the main fraction, is added to the diluted main fraction, pH adjusted to 5.5 ± 0.5 and stirred for 1 hour. The suspension is then centrifuged at 9,000 rpm for 15 minutes at 4°C. The resulting pellet is suspended in water.

Step 15. Low-pressure chromatography on Sepharose Q resin.

The insulin suspension is supplemented with water to a volume calculated according to the formula: total volume = total amount of protein obtained after DEAE [AU A_{280}] / 7, Tris pH 8.6 is added to a concentration of 30 mM and 0.2 M EDTA solution is added in an amount calculated according to the formula: volume [cm^3] = total amount of protein obtained after DEAE [AU A_{280}] / 190, and then the pH is adjusted to 8.6. The solution of dissolved zinc salt of insulin at pH 8.6 is loaded onto a column packed with the Q resin (80 cm^3) and equilibrated with 0.5 M Tris, pH 8.6 and then 20 mM Tris, pH 8.6. After loading, the column is washed with 20 mM Tris buffer with 20 ± 5% isopropanol, pH 8.6, and then the insulin protein eluted with a buffer at pH 8.6 consisting of 20 mM TRIS with NaCl until conductivity is 3 ± 2 mS and with 25 ± 5% isopropanol.

Step 16. Reaction with carboxypeptidase B.

The main fraction eluted from the column is diluted two times and a solution of carboxypeptidase B at concentration of 2.5 mg/cm^3 is added in an amount calculated according to the formula: volume [μm^3] = total amount of protein obtained from Q [AU A_{280}] / 15. pH of the solution is adjusted to 8.8 and the solution is stirred for 16-18 hours at room temperature. The pH is then adjusted to 3 and the solution is purified by high-pressure chromatography.

Step 17. High-pressure chromatography RP-HPLC.

The separation is carried out on a Kromasil C8 or C18 column using an acetonitrile gradient:

Phase A	Phase B
95%	5%
75%	25%

Mobile phase A: 2,000 cm^3 of a 0.2 M sodium sulfate solution with 440 cm^3 of acetonitrile, pH 2.3 and a temperature of 25-30°C;

Mobile phase B: 1,250 cm^3 of a 0.2 M sodium sulfate solution with 1,250 cm^3 of acetonitrile, pH 2.3 and a temperature of 25-30°C;

The solution after reaction with carboxypeptidase B or a main fraction after purification on the Q Sepharose resin, following a double dilution and adjustment of the pH to 3, is purified in portions by

loading onto the column a single volume of solution having a total protein content of 200 AU, and then the main fraction is collected, so that product purity (HPLC) is at least 98%.

Step 18. Precipitation of insulin with zinc chloride.

The combined main fractions from the high-pressure chromatography are diluted 2 times and 1 M $ZnCl_2$ is added in a proportion of 3 cm^3 of zinc chloride per 150 cm^3 of the sample, and then the pH is adjusted to 6 ± 1, followed by stirring for 1 hour. The suspension is then centrifuged at 9,000 rpm for 15 minutes at 4°C and the pellet is resuspended in 10-15 cm^3 of water.

Step 19. Chromatography on a bed Sephadex G-25 and freeze-drying.

The slurry is adjusted to pH 3, filtered through a 0.22 µm filter and loaded onto a column filled with Sephadex G 25 resin (120 cm^3). Before loading the column is equilibrated with 5 mM ammonium acetate, pH 4, and the protein is eluted with 5 mM ammonium acetate, pH 4. The resulting material is subjected to freeze-drying.

II. Specific examples

Example 1. LysB28ProB29 human insulin (lispro insulin)

According to the general procedure of the process given above (section I), the production of biosynthetic lispro insulin was performed following steps 1-6, 6b, 7-10 and 12-19 and for the expression system 1.2. and 1.5. additionally step 11 (Table 1. Lispro insulin). The process was studied in 20 variants of 8 protein expression systems. The data obtained shows that the strains of *Escherichia coli* DH5α, DH5 and HB101, which perform well in laboratory conditions, with plasmids with *deoP1P2* and *pms* promoters are not suitable for production culture in the minimal medium (system 1.3., 1.4., 1.5., 1.6. and 1.7.). These strains are characterized by low growth on minimal technological medium, demonstrating a much lower optical density value in the production fermenters. A similarly low feasibility for production is shown by the IBA strain system with disrupted *cytR* repressor and pIGAL plasmid (system 1.8.), characterized by relatively better growth, but low content of inclusion bodies (microscopic observations).

The best expression yield is shown by the IBA/pIBA strain system, described in Polish patent specification No. PL 180 818, which was tested in 4 linker peptide variants and 2 leader protein variants (system 1.1. and 1.2.). All have good, comparable growth parameters in the minimal medium, but differ in weight of inclusion bodies and relative yield of the fusion peptide renaturation, and – consequently – the final product yield.

The highest yield of lispro insulin was obtained using the *E. coli* IBA/pIBA expression system producing A-chain linked to the B-chain of lispro insulin with ArgArg dipeptide (Table 1, Example No. 1.1.1. and 1.2.1.), exceeding the performance of other systems by at least 13-15%.

Example 2 GlyA22ArgB31 insulin (insulin GR).

According to the general procedure of the process given above (section I), the production of biosynthetic GR insulin was performed following steps 1-6, 6b, 7-10, 12-15 and 17-19 and for the expression system 2.2. additionally step 11 (Table 2. GR insulin). The process was studied in 12 variants of 5 protein expression systems (not using the variant with LysArg linker, as is does not lead to GR insulin). The data obtained confirms the findings of Example 1 related to the superiority of the expression system described in the Polish patent specification No. PL 180 818 over the other systems studies and shows similar differences in the yield of the process depending on type of the linker peptide that links the A and B chains of this insulin analogue. In particular, the highest yield of GR insulin was obtained using the *E. coli* IBA/pIBA expression system, producing A-chain linked to the B-chain of GR insulin with ArgArg dipeptide (Table 2, Example No. 2.1.1. and 2.2.1.) and exceeding the performance of other systems by at least 16-22%. The presented results show technological advantage of ArgArg linker regardless of the type of the leader protein.

Example 3. Human insulin.

According to the general procedure of the process given above (section I), the production of biosynthetic human insulin was performed in 8 variants of 2 protein expression systems. Steps 1-6, 6b, 7-10 and 12-19 and for the expression system 1.2. and 1.5. additionally step 11 were used in the studies. The data obtained shows (Table 3. Human insulin), that the IBA/pIBA strains system is characterized by the best expression yield, showing good, comparable growth parameters in the industrial minimal medium and differing in the weight of inclusion bodies produced and the relative yield of renaturation. Again the highest yield of human insulin was obtained using the *E. coli* IBA/pIBA expression system producing A-chain linked to the B-chain of human insulin with ArgArg dipeptide (Table 3 Example No. 3.1.1. and 3.2.1.). The analytical studies performed at different steps of the process indicate, that the higher yield in these expression system variants – by about 14-16% – is reached after the step of inclusion body preparation and renaturation, and it is kept along all other steps of the process.

Example 4. SerA22ArgB31 human insulin (SR insulin).

Studies of the yield of the SR insulin production process were performed according to the general procedure given above (section I), following steps 1-6, 6b, 7-10, 12-15 and 17-19 and for the

expression system 4.2. additionally step 11 (Table 2. GR insulin). The process was studied in 12 variants of 5 protein expression systems (not using the variant with LysArg linker, as is does not lead to SR insulin). The data obtained supports the previous results, constantly showing a higher yield of the *E. coli* IBA/pIBA expression system, producing A-chain linked to the B-chain of SR insulin with ArgArg dipeptide (Table 4 Example No. 4.1.1. and 4.2.1.) in comparison with other linker peptides of this system – by at least 15%.

Example 5. SerA22LizB3ArgB31 insulin (SK3R insulin)

According to the general procedure of the process given above, the production of biosynthetic SK3R insulin were performed in 6 variants of 2 protein expression systems. Steps 1-6, 6b, 7-10, 12-15 and 17-19 and for the expression system 5.2. additionally step 11 were used in the studies. The data obtained shows (Table 5. SK3R insulin) that the IBA/pIBA strains system is characterized by the best expression yield, showing good, comparable growth parameters in the industrial minimal medium and differing in the weight of inclusion bodies produced and the relative yield of renaturation. Again the highest yield of human insulin was obtained using the *E. coli* IBA/pIBA expression system producing A-chain linked to the B-chain of human insulin with ArgArg dipeptide (Table 5, Example No. 5.1.1. and 5.2.1.). The extensive analytical studies performed at different steps of the process indicate, that the higher yield in these expression system variants is kept along all other steps of the process after the step of inclusion body preparation and renaturation, the yield being higher than that obtained with other expression system variants by at least 14-16%.

Example 6. Preproinsulin (precursor) G: leader protein – B-chain – linker – A-chain: dezAsnA21GlyA21.

The versatility of the ArgArg linker advantage over others for preproinsulin expression was also demonstrated by studying the yield of a protein consisting of a leader protein preceding the B-chain, which was linked with the studied linkers to a modified A-chain of insulin (A: dezAsnA21GlyA21). The studies were performed using critical steps 1-8 of the general procedure of the process in 8 variants of 2 expression systems. The research results, presented in Table 6. Preproinsulin (precursor) G: leader protein – B-chain – linker – A-chain: dezAsnA21GlyA21 show that in the case of the ArgArg linker the yield of the process is 10% higher than in the other cases.

Based on the final product yield, confirmed by the protein content before purification, it was unexpectedly found that the best results in the production of recombinant human insulin and analogues thereof are obtained in an expression system consisting of an expression plasmid with *deoP1P2* promoter in bacterial strains of *Escherichia coli* lacking the *cytR* repressor gene (Polish patent No. PL 180 818) if in the plasmid encoding the fusion protein the A-chain of insulin is linked to

the B-chain with ArgArg dipeptide. Moreover, the difference in performance between the ArgArg dipeptide variant of the system and LyzArg is 13-22%, and is comparable to the improvement in system performance described in the Polish patent No. PL 180 818 for LizArg dipeptide and Arg in comparison with the C-peptide, which is 23% and 17%, respectively.

Genetic construction of pIBA/INSLys(31B)Arg(32B) pIBA/INS plasmid

The pIBA/INS plasmid of 4354 base pairs in size is composed of the following regulatory sequences and genes:

- from 7 bp to 942 bp a regulatory region with *deoP1P2* promoters is located,
- from 946 bp to 1137 bp there is a sequence encoding a fragment of human SOD gene,
- from 1138 bp to 1299 bp a sequence encoding the A and B chains, as well as the C chain (in the form of Lys-Arg linker) of recombinant human insulin INS is contained, the nucleotide sequence of insulin has been changed in accordance with the frequency of codon usage in *E. coli*.
- from 1312 bp to 1338 bp a region coding for Ter trpA transcription terminator is located,
- from 1508 bp to 2698 bp there is a tetracycline resistance gene Tet.

The structure of the pIBA plasmid is schematically shown in Figure 1, and its nucleotide sequence and amino acid sequence in Figure 2. The structure of the pIBA/INS plasmid comprising an exemplary gene encoding recombinant human insulin protein, wherein the A-chain is linked to the B-chain with LysArg dipeptide is schematically shown in Figure 3, and its nucleotide sequence and amino acid sequence in Figure 4.

Claims

1. A polypeptide having the amino acid sequence of the formula:

X_n-B-Arg-Arg-A

where:

A is a polypeptide of the insulin A-chain or analogue thereof, preferably a sequence selected from SEQ. ID No.: 1-4.

B is a polypeptide of the insulin B-chain or analogue thereof, preferably a sequence selected from SEQ. ID No.: 5-7.

n is 0 or 1,

X is a leader protein polypeptide, preferably of a sequence selected from SOD of SEQ. No.: 8 or UBI of SEQ. No.: 9.

2. The polypeptide according to claim 1, characterized in that it has an amino acid sequence selected from: SEQ. ID No.: 10-27.

3. A method for producing human insulin comprising microbial expression of a recombinant protein containing insulin, characterized in that it comprises:

(a) obtaining a hybrid polypeptide comprising proinsulin by expression of the hybrid polypeptide in a cell of an *Escherichia coli* strain lacking *deo* repressor and/or *deo* (*cytR*) gene containing DNA encoding the hybrid polypeptide, wherein the DNA is present in a vector under the control of *deo*P_1P_2 promoter, while the hybrid polypeptide comprises the insulin B-chain covalently bound to the A-chain via Arg-Arg dipeptide,

(b) recovering of the hybrid polypeptide,

(c) folding and formation of disulfide bonds within the hybrid polypeptide obtained in (b), avoiding prior subjection of the hybrid polypeptide to sulfitolysis,

(d) subjecting the hybrid polypeptide obtained in (c) folded, with disulfide bonds formed, to enzymatic digestion in order to produce insulin,

(e) purifying the insulin so produced.

4. The method of claim 3, characterized in that the hybrid peptide is a peptide defined in claims 1-2.

5. The method of claim. 3, characterized in that at the step (a) the expression is performed in *E. coli* strain of the genotype:

F⁻ ara Δ(pro lac) rpsL thi cytR recA λ⁻,
devoid of active *cytR* repressor protein.

6. The method of claim 3, characterized in that at the step (a) a plasmid *pIBA* of the sequence SEQ ID NO: 28 is used as the vector.

7. The method of claim 3, characterized in that the step (a) comprises fermentation in the presence of glucose, glycerol or galactose.

8. The method of claim 3, characterized in that the step (b) comprises:
(i) destruction of the cell wall of the bacterial cell or its fragments to form a lysate, preferably by the action of lysozyme and Triton X100 followed by sonication or disintegration;
(ii) isolation of intracellular inclusion body precipitate from the lysate by centrifugation; and
(iii) dissolving the inclusion body precipitate.

9. The method of claim 3, characterized in that the step (c) comprises incubating the hybrid polypeptide, preferably in the presence of ascorbic acid, in particular at a concentration of about 2 moles per mole of SH groups present in the mixture, at a temperature of 4 to 37°C for about 1 to 30 hours, preferably about 5 hours, at a pH of 8.5 to 12, preferably at a pH of 11.0 to 11.25.

10. The method of claim 3, characterized in that the step (d) further comprises:
(i) adjusting the pH to about 8.8-9.0; and
(ii) digestion of the hybrid polypeptide with trypsin and subsequently optionally with carboxypeptidase B at a temperature of 16-37°C for 30 minutes to 16 hours.

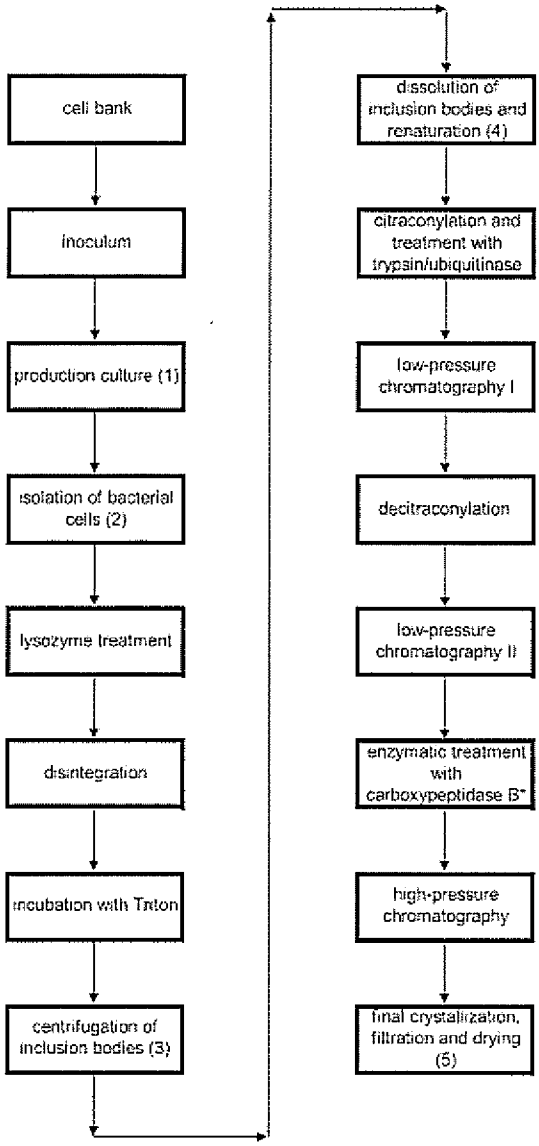
11. The method of claim 3, characterized in that the step (e) further comprises purifying the solution by low-pressure liquid chromatography on Sepharose Q prior to trypsin digestion.

12. The method of claim 3, characterized in that the step (e) further comprises purifying the solution by low-pressure chromatography on DEAE-Sepharose prior to carboxypeptidase B digestion.

13. The method of claim 3, characterized in that the step (e) further comprises purifying the resulting insulin by high-performance liquid chromatography, preferably to a purity of at least 98%, followed by crystallization, filtration and drying.

14. The method of claim 3, characterized in that the insulin purified at the step (e) is a protein selected from the group comprising: human insulin, human insulin $Lys^{B28}Pro^{B29}$ (lispro insulin), insulin $Gly^{A22}Arg^{B31}$ (GR insulin), human insulin $Ser^{A22}Arg^{B31}$ (SR insulin), insulin $Ser^{A22}Lys^{B3}Arg^{B31}$ (SK3R insulin), proinsulin B-ArgArg-A:dez$Asn^{A21}Gly^{A21}$.

Scheme 1



(1) Control of optical density OD_{600}
(2) Weight of the culture (bacteria) after centrifugation (or after centrifugation and drying)
(3) Weight of the pure inclusion bodies after centrifugation (or after centrifugation and drying)
(4) Total protein (after folding, spectrophotometric method or HPLC)
(5) Final product yield
* In the case of performing the operation

Table 1

1. LysB28ProB29 human insulin											
E. coli system/ vector characteristics	Linker	(1) OD	(2) Weight of culture		(3) Weight of inclusion bodies		(4) Relative content of folded precursor in total protein		(5) Final yield		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	% (HPLC)	% of max. value	mg/dm^3	% of max. value	
1.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	55	55.0	100%	13.8	100%	7.0	100%	39	100%	**1.1.1.**
	LysArg	54	55.0	100%	12.8	93%	5.7	81%	33	85%	**1.1.2.**
	Arg	51	54.4	99%	12.2	89%	5.7	82%	29	74%	**1.1.3.**
	C-peptide	53	55.0	100%	12.8	93%	5.9	85%	23	59%	**1.1.4.**
1.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	51	51.2	98%	12.9	100%	7.0	100%	35	100%	**1.2.1.**
	LysArg	51	51.2	98%	11.6	90%	5.7	81%	31	87%	**1.2.2.**
	Arg	48	51.1	98%	11.9	92%	5.7	82%	27	75%	**1.2.3.**
	C-peptide	48	52.0	100%	12.9	100%	6.6	94%	21	60%	**1.2.4.**
1.3. DH5α/pIGAL leader protein - SOD, resistance: ampicillin, transcription terminator - ST1	ArgArg	16	5.3		0.4		0.4		6**	-	**1.3.1.**
	LysArg	15	4.2		0.3		0.4		-*	-	**1.3.2.**
1.4. HB101/pIGAL leader protein - UBI, resistance: ampicillin, transcription terminator - ST1	ArgArg	12	3.1		0.2		~0.2		-*		**1.4.1.**
	LysArg	12	3.1		0.2		~0.2		-*		**1.4.2.**
1.5. DH5/pIGAL leader protein - UBI, resistance: tetracycline, transcription terminator - ST1	ArgArg	28	4.8		0.5		0.5		11**		**1.5.1**
	LysArg	29	5.0		0.5		0.5		9**		**1.5.2**
1.6. DH5/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - terA Trp	ArgArg	32	5.3		0.7		0.6		17**		**1.6.1.**
	LysArg	31	5.2		0.6		0.6		16**		**1.6.2.**

1.7. DH5/pIBA leader protein - SOD, resistance: tetracycline, transcription terminator - terA Trp	ArgArg	22	3.3		-*		-*		-*		**1.7.1.**
	LysArg	20	3.2		-*		-*		-*		**1.7.2.**
1.8. IBA/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - ST1	ArgArg	34	7.8		0.6		0.4		-*		**1.8.1.**
	LysArg	35	7.8		0.6		0.4		-*		**1.8.2.**

* - The process was interrupted at the purification step due to low content of the product.
** - Low and unrepresentative performance also results from inadequate scale of the purification process to low content of the product.

Table 2

2. GlyA22ArgB31 human insulin (GR insulin)											
E. coli system/ vector characteristics	Linker	(1) OD	(2) Dry weight of culture		(3) Dry weight of inclusion bodies		(4) Total protein (precursor) (A_{280})		(5) Final yield		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	AU/dm^3	% of max. value	mg/dm^3	% of max. value	
2.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	62	17.6	100%	8.8	100%	7860	100%	40	100%	**2.1.1.**
	Arg	59	17.4	99%	8	91%	6890	88%	31	78%	**2.1.2.**
	Peptyd C	62	17.5	99%	7.8	89%	6686	85%	24	61%	**2.1.3.**
2.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	68	17.9	99%	9.0	100%	7891	100%	40	100%	**2.2.1.**
	Arg	67	17.3	96%	8.1	90%	6714	85%	34	84%	**2.2.2.**
	Peptyd C	65	18.0	100%	8.3	92%	6890	88%	25	63%	**2.2.3.**
2.3. DH5/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - ST1	ArgArg	32	5.3		0.8				-*		**2.3.1**
	Arg	32	5.3		0.7				-*		**2.3.2**
2.4. DH5/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - terA Trp	ArgArg	26	4		-*				-*		**2.4.1**
	Arg	28	4		-*				-*		**2.4.2**
2.5. IBA/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - ST1	ArgArg	39	8.3		0.9				-*		**2.5.1**
	Arg	39	8.4		0.9				-*		**2.5.2**

Table 3

3. human insulin																								
E. coli system/ vector characte-ristics	Linker	(1) OD 600	(2) Weight of bacteria		(2) Dry weight of bacteria		(3) Weight of inclusion bodies		(3) Dry weight of inclusion bodies		(4) Total protein (precursor), A280		Total protein (after biotransformation), A280		Total protein (after DEAE), A280		Total protein (after Q), A280		Total protein (insulin after CPB), A280		Total protein (insulin after HPLC), Bratford method, A595		(5) Final yield		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	g/dm^3	% of max. value	g/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	mg/dm^3	% of max. value	mg/dm^3	% of max. value	
3.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	59	54.7	100%	26.9	100%	29.3	100%	14.3	100%	10070	100%	10133	100%	1411	100%	571	100%	471	100%	104	100%	82	100%	**3.1.1.**
	LysArg	61	55.0	100%	27.0	100%	27.5	94%	13.7	96%	8459	84%	8410	83%	1157	82%	468	82%	396	84%	87	85%	68	83%	**3.1.2.**
	Arg	59	53.6	98%	26.3	98%	24.8	90%	13.2	92%	8358	83%	8511	84%	1160	82%	465	82%	391	83%	85	83%	67	83%	**3.1.3.**
	C-peptide	53	54.7	100%	26.9	100%	22.9	92%	13.3	93%	7553	75%	7701	76%	1086	77%	457	80%	354	75%	76	74%	59	72%	**3.1.4.**
3.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	50	55.0	100%	27.0	100%	30.1	100%	14.3	100%	10012	100%	9039	100%	1128	100%	456	100%	368	100%	88	100%	72	100%	**3.2.1.**
	LysArg	48	55.0	100%	27.0	100%	28.3	94%	13.0	91%	8210	82%	7683	85%	936	83%	378	83%	316	86%	73	83%	61	85%	**3.2.2.**
	Arg	45	54.7	100%	26.9	100%	26.8	89%	13.3	93%	8203	82%	7593	84%	925	82%	374	82%	305	83%	76	85%	61	84%	**3.2.3.**
	C-peptide	44	52.5	96%	25.8	96%	27.1	90%	13.6	95%	6908	69%	6598	73%	790	70%	319	70%	265	72%	64	73%	54	74%	**3.2.4.**

Table 4

4. SerA22ArgB31 human insulin (SR insulin)											
E. coli system/ vector characteristics	Linker	(1) OD	(2) Dry weight of culture		(3) Dry weight of inclusion bodies		(4) Total protein (precursor) (A_{280})		(5) Final yield		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	AU/dm^3	% of max. value	mg/dm^3	% of max. value	
4.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	64	17.4	100%	8.8	100%	7860	100%	48	100%	**4.1.1.**
	Arg	68	17.5	99%	8.1	92%	6890	88%	41	85%	**4.1.2.**
	C-peptide	64	17.3	99%	8.2	93%	6686	85%	36	75%	**4.1.3.**
4.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	68	17.9	99%	9.0	100%	7891	100%	46	100%	**4.2.1.**
	Arg	67	17.3	99%	8.1	90%	6714	85%	39	85%	**4.2.2.**
	C-peptide	65	18.0	100%	8.3	92%	6890	88%	35	75%	**4.2.3.**
4.3. DH5/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - ST1	ArgArg	32	5.3						-*		**4.3.1**
	Arg	32	5.3						-*		**4.3.2**
4.4. DH5/pIBA leader protein - SOD, resistance: tetracycline, transcription terminator - terA Trp	ArgArg	26	4						-*		**4.4.1**
	Arg	28	4						-*		**4.4.2**
4.5. IBA/pIGAL leader protein - SOD, resistance: tetracycline, transcription terminator - ST1	ArgArg	39	8.3						-*		**4.5.1**
	Arg	39	8.4						-*		**4.5.2**

Table 5

5. SerA22LysB3ArgB31 human insulin (SK3R)																			
E. coli system/ vector characteristics	Linker	(1) OD 600	(2) Weight of bacteria		(3) Weight of inclusion bodies		(4) Total protein (precursor), A280		Total protein (after biotransformation), A280		Total protein (after DEAE), A280		Total protein (after Q), A280		Total protein (insulin after HPLC), Bradford method, A595		(5) Final yield		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	AU/dm^3	% of max. value	mg/dm^3	% of max. value	mg/dm^3	% of max. value	
5.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	57	52.3	100%	23.0	100%	7857	100%	7602	100%	1050	100%	326	100%	93	100%	78	100%	**5.1.1.**
	Arg	56	52.3	100%	21.2	92%	6757	86%	6386	84%	861	82%	274	84%	77	83%	66	84%	**5.1.2.**
	C-peptide	59	51.7	99%	21.6	94%	6600	84%	5778	76%	809	77%	261	80%	68	74%	56	72%	**5.1.3.**
5.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	55	51.8	97%	23.5	100%	7720	100%	7587	100%	1059	100%	332	100%	88	100%	72	100%	**5.2.1.**
	Arg	55	52.0	98%	22.1	94%	6330	82%	6373	84%	868	82%	272	82%	75	85%	62	86%	**5.2.2.**
	C-peptide	56	53.3	100%	22.3	95%	6639	86%	5918	78%	773	73%	249	75%	64	73%	55	76%	**5.2.3.**

Table 6

6. Preproinsulin (precursor) G: leader protein – B-chain – linker – A-chain: dezAsnA21GlyA21									
E. coli system/ vector characteristics	Linker	(1) OD	(2) Dry weight of bacteria		(3) Dry weight of inclusion bodies		(4) Total protein (precursor) [A280]		Example No.
			g/dm^3	% of max. value	g/dm^3	% of max. value	AU/dm^3	% of max. value	
6.1. IBA/pIBA leader protein - SOD, transcription terminator - terA Trp	ArgArg	58	17.4	99%	8.4	100%	6960	100%	**6.1.1.**
	LysArg	60	17.3	99%	7.9	94%	6264	90%	**6.1.2.**
	Arg	60	17.5	100%	7.8	93%	5916	85%	**6.1.3.**
	C-peptide	55	17.3	99%	8.1	96%	6334	91%	**6.1.4.**
6.2. IBA/pIBA leader protein - UBI, transcription terminator - terA Trp	ArgArg	55	17.0	100%	8.7	100%	7012	100%	**6.2.1.**
	LysArg	55	16.9	99%	8.1	93%	6381	91%	**6.2.2.**
	Arg	57	17.0	100%	8.0	92%	5890	84%	**6.2.3.**
	C-peptide	54	17.0	100%	8.4	97%	6451	92%	**6.2.4.**

map of the pIBA vector

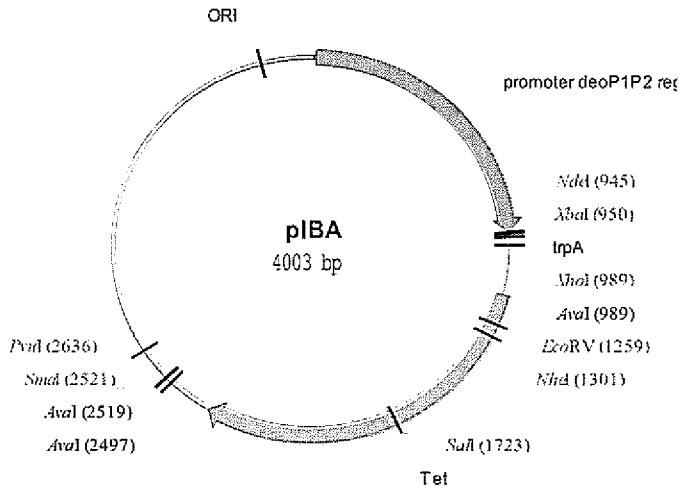


Fig. 1

Fig. 2

nucleotide sequence of the pIBA vector

```
   1 GAATTCGGAC CGTTGGCGAT GTGCGGTTTG CTACATTCAC AGATGTTCTT
     CTTAAGCCTG GCAACCGCTA CACGCCAAAC GATGTAAGTG TCTACAAGAA
  51 CGCCACTTCC AGCAGCAGGT CATCAGGGGT GATTTCAGGA TCGTAGATAA
     GCGGTGAAGG TCGTCGTCCA GTAGTCCCCA CTAAAGTCCT AGCATCTATT
 101 AGGTCAGGTT CGGTGAAACC TGCTTCAACT CTGCATCTGC ACGTAAGATC
     TCCAGTCCAA GCCACTTTGG ACGAAGTTGA GACGTAGACG TGCATTCTAG
 151 GCGCGGGTAA TGGGCGAATC AGACGGGCCG ATATTGGCGT GCATAAAGGC
     CGCGCCCATT ACCCGCTTAG TCTGCCCGGC TATAACCGCA CGTATTTCCG
 201 GTCTGGCAGG GTTCTGTCGA GGTAACGCCA GAAACGTTTT ATTCGAACAT
     CAGACCGTCC CAAGACAGCT CCATTGCGGT CTTTGCAAAA TAAGCTTGTA
 251 CGATCTCGTC TTGTGTTAGA ATTAATTCTA ACATACGGTT GCAACAACGC
     GCTAGAGCAG AACACAATCT TAATTAAGAT TGTATGCCAA CGTTGTTGCG
 301 ATCCAGTTGC CCCAGGTAGA CCGGCATCGA TGTGACCGAC GGTACGTGGT
     TAGGTCAACG GGGTCCATCT GGCCGTAGCT ACACTGGCTG CCATGCACCA
 351 GGTAAAGAAT GGTCAGCAGA GAGAGTGCGT CATCAAGATC TTTCGCGCCT
     CCATTTCTTA CCAGTCGTCT CTCTCACGCA GTAGTTCTAG AAAGCGCGGA
 401 TCCAGCTCCA GCCATTCGGA ACCGTTCGCC AGAAAACGGG CGTAATCGGG
     AGGTCGAGGT CGGTAAGCCT TGGCAAGCGG TCTTTTGCCC GCATTAGCCC
 451 TAAGACATAG CGCGGTTTGT ACGGCGCATG ACCTTCAAAC ATATCGCAGA
     ATTCTGTATC GCGCCAAACA TGCCGCGTAC TGGAAGTTTG TATAGCGTCT
 501 TTACACCTTC ATCCAGCGCG CGGCGGGCTT CGGCAGGAAG CTGTGGGTAA
     AATGTGGAAG TAGGTCGCGC GCCGCCCGAA GCCGTCCTTC GACACCCATT
 551 GGCAGATTGT TTTCTGCTTC CAGTGCCAGA AAATGGCGCT TCTGCTCCGG
     CCGTCTAACA AAAGACGAAG GTCACGGTCT TTTACCGCGA AGACGAGGCC
 601 GCTAAGCACT GGGCTGGTGA CAATTTGCTG GCAACGTTGT TGCAGTGCAT
     CGATTCGTGA CCCGACCACT GTTAAACGAC CGTTGCAACA ACGTCACGTA
 651 TTTCATGAGA AGTGGGCATC TTCTTTTCCT TTTATGCCGA AGGTGATGCG
     AAAGTACTCT TCACCCGTAG AAGAAAAGGA AAATACGGCT TCCACTACGC
 701 CCATTGTAAG AAGTTTCGTG ATGTTCACTT TGATCCTGAT GCGTTTGCCA
     GGTAACATTC TTCAAAGCAC TACAAGTGAA ACTAGGACTA CGCAAACGGT
 751 CCACTGACGC ATTCATTTGA AAGTGAATTA TTTGAACCAG ATCGCATTAC
     GGTGACTGCG TAAGTAAACT TTCACTTAAT AAACTTGGTC TAGCGTAATG
 801 AGTGATGCAA ACTTGTAAGT AGATTTCCTT AATTGTGATG TGTATCGAAG
     TCACTACGTT TGAACATTCA TCTAAAGGAA TTAACACTAC ACATAGCTTC
 851 TGTGTTGCGG AGTAGATGTT AGAATACTAA CAAACTCGCA AGGTGAATTT
     ACACAACGCC TCATCTACAA TCTTATGATT GTTTGAGCGT TCCACTTAAA
 901 TATTGGCGAC AAGCCTAGGT TTGTTTAACT TTAAGGAGAA ATCATATGTC
     ATAACCGCTG TTCGGATCCA AACAAATTGA AATTCCTCTT TAGTATACAG
 951 TAGAATCGAT GCCCGCTTAA TGAGCGGGCT TTTTTTTCTC GAGGACGTCT
     ATCTTAGCTA CGGGCGAATT ACTCGCCCGA AAAAAAAGAG CTCCTGCAGA
1001 AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG
     TTCTTTGGTA ATAATAGTAC TGTAATTGGA TATTTTTATC CGCATAGTGC
1051 AGGCCCTTTC GTCTTCAAGA CCTCTCATGT TTGACAGCTT ATCATCGATA
     TCCGGGAAAG CAGAAGTTCT GGAGAGTACA AACTGTCGAA TAGTAGCTAT
1101 AGCTTTAATG CGGTAGTTTA TCACAGTTAA ATTGCTAACG CAGTCAGGCA
     TCGAAATTAC GCCATCAAAT AGTGTCAATT TAACGATTGC GTCAGTCCGT
1151 CCGTGTATGA AATCTAACAA TGCGCTCATC GTCATCCTCG GCACCGTCAC
     GGCACATACT TTAGATTGTT ACGCGAGTAG CAGTAGGAGC CGTGGCAGTG
1201 CCTGGATGCT GTAGGCATAG GCTTGGTTAT GCCGGTACTG CCGGGCCTCT
     GGACCTACGA CATCCGTATC CGAACCAATA CGGCCATGAC GGCCCGGAGA
1251 TGCGGGATAT CGTCCATTCC GACAGCATCG CCAGTCACTA TGGCGTGCTG
     ACGCCCTATA GCAGGTAAGG CTGTCGTAGC GGTCAGTGAT ACCGCACGAC
1301 CTAGCGCTAT ATGCGTTGAT GCAATTTCTA TGCGCACCCG TTCTCGGAGC
     GATCGCGATA TACGCAACTA CGTTAAAGAT ACGCGTGGGC AAGAGCCTCG
1351 ACTGTCCGAC CGCTTTGGCC GCCGCCCAGT CCTGCTCGCT TCGCTACTTG
     TGACAGGCTG GCGAAACCGG CGGCGGGTCA GGACGAGCGA AGCGATGAAC
1401 GAGCCACTAT CGACTACGCG ATCATGGCGA CCACACCCGT CCTGTGGATC
     CTCGGTGATA GCTGATGCGC TAGTACCGCT GGTGTGGGCA GGACACCTAG
1451 CTCTACGCCG GACGCATCGT GGCCGGCATC ACCGGCGCCA CAGGTGCGGT
     GAGATGCGGC CTGCGTAGCA CCGGCCGTAG TGGCCGCGGT GTCCACGCCA
1501 TGCTGGCGCC TATATCGCCG ACATCACCGA TGGGGAAGAT CGGGCTCGCC
     ACGACCGCGG ATATAGCGGC TGTAGTGGCT ACCCCTTCTA GCCCGAGCGG
```

```
1551 ACTTCGGGCT CATGAGCGCT TGTTTCGGCG TGGGTATGGT GGCAGGCCCC
     TGAAGCCCGA GTACTCGCGA ACAAAGCCGC ACCCATACCA CCGTCCGGGG
1601 GTGGCCGGGG GACTGTTGGG CGCCATCTCC TTGCATGCAC CATTCCTTGC
     CACCGGCCCC CTGACAACCC GCGGTAGAGG AACGTACGTG GTAAGGAACG
1651 GGCGGCGGTG CTCAACGGCC TCAACCTACT ACTGGGCTGC TTCCTAATGC
     CCGCCGCCAC GAGTTGCCGG AGTTGGATGA TGACCCGACG AAGGATTACG
1701 AGGAGTCGCA TAAGGGAGAG CGTCGACCGA TGCCCTTGAG AGCCTTCAAC
     TCCTCAGCGT ATTCCCTCTC GCAGCTGGCT ACGGGAACTC TCGGAAGTTG
1751 CCAGTCAGCT CCTTCCGGTG GGCGCGGGGC ATGACTATCG TCGCCGCACT
     GGTCAGTCGA GGAAGGCCAC CCGCGCCCCG TACTGATAGC AGCGGCGTGA
1801 TATGACTGTC TTCTTTATCA TGCAACTCGT AGGACAGGTG CCGGCAGCGC
     ATACTGACAG AAGAAATAGT ACGTTGAGCA TCCTGTCCAC GGCCGTCGCG
1851 TCTGGGCCAT TTTCGGCGAG GACCGCTTTC GCTGGAGCGC GACGATGATC
     AGACCCGGTA AAAGCCGCTC CTGGCGAAAG CGACCTCGCG CTGCTACTAG
1901 GGCCTGTCGC TTGCGGTATT CGGAATCTTG CACGCCCTCG CTCAAGCCTT
     CCGGACAGCG AACGCCATAA GCCTTAGAAC GTGCGGGAGC GAGTTCGGAA
1951 CGTCACTGGT CCCGCCACCA AACGTTTCGG CGAGAAGCAG GCCATTATCG
     GCAGTGACCA GGGCGGTGGT TTGCAAAGCC GCTCTTCGTC CGGTAATAGC
2001 CCGGCATGGC GGCCGACGCG CTGGGCTACG TCTTGCTGGC GTTCGCGACG
     GGCCGTACCG CCGGCTGCGC GACCCGATGC AGAACGACCG CAAGCGCTGC
2051 CGAGGCTGGA TGGCCTTCCC CATTATGATT CTTCTCGCTT CCGGCGGCAT
     GCTCCGACCT ACCGGAAGGG GTAATACTAA GAAGAGCGAA GGCCGCCGTA
2101 CGGGATGCCC GCGTTGCAGG CCATGCTGTC CAGGCAGGTA GATGACGACC
     GCCCTACGGG CGCAACGTCC GGTACGACAG GTCCGTCCAT CTACTGCTGG
2151 ATCAGGGACA GCTTCAAGGA TCGCTCGCGG CTCTTACCAG CCTAACTTCG
     TAGTCCCTGT CGAAGTTCCT AGCGAGCGCC GAGAATGGTC GGATTGAAGC
2201 ATCATTGGAC CGCTGATCGT CACGGCGATT TATGCCGCCT CGGCGAGCAC
     TAGTAACCTG GCGACTAGCA GTGCCGCTAA ATACGGCGGA GCCGCTCGTG
2251 ATGGAACGGG TTGGCATGGA TTGTAGGCGC CGCCCTATAC CTTGTCTGCC
     TACCTTGCCC AACCGTACCT AACATCCGCG GCGGGATATG GAACAGACGG
2301 TCCCCGCGTT GCGTCGCGGT GCATGGAGCC GGGCCACCTC GACCTGAATG
     AGGGGCGCAA CGCAGCGCCA CGTACCTCGG CCCGGTGGAG CTGGACTTAC
2351 GAAGCCGGCG GCACCTCGCT AACGGATTCA CCACTCCAAG AATTGGAGCC
     CTTCGGCCGC CGTGGAGCGA TTGCCTAAGT GGTGAGGTTC TTAACCTCGG
2401 AATCAATTCT TGCGGAGAAC TGTGAATGCG CAAACCAACC CTTGGCAGAA
     TTAGTTAAGA ACGCCTCTTG ACACTTACGC GTTTGGTTGG GAACCGTCTT
2451 CATATCCATC GCGTCCGCCA TCTCCAGCAG CCGCACGCGG CGCATCTCGG
     GTATAGGTAG CGCAGGCGGT AGAGGTCGTC GGCGTGCGCC GCGTAGAGCC
2501 GCAGCGTTGG GTCCTGGCCC GGGACTCACC AGTCACAGAA AAGCATCTTA
     CGTCGCAACC CAGGACCGGG CCCTGAGTGG TCAGTGTCTT TTCGTAGAAT
2551 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT
     GCCTACCGTA CTGTCATTCT CTTAATACGT CACGACGGTA TTGGTACTCA
2601 GATAACACTG CGGCCAACTT ACTTCTGACA ACGATCGGAG GACCGAAGGA
     CTATTGTGAC GCCGGTTGAA TGAAGACTGT TGCTAGCCTC CTGGCTTCCT
2651 GCTAACCGCT TTTTTGCACA ACATGGGGGA TCATGTAACT CGCCTTGATC
     CGATTGGCGA AAAAACGTGT TGTACCCCCT AGTACATTGA GCGGAACTAG
2701 GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC
     CAACCCTTGG CCTCGACTTA CTTCGGTATG GTTTGCTGCT CGCACTGTGG
2751 ACGATGCCTG CAGCAATGGC AACAACGTTG CGCAAACTAT TAACTGGCGA
     TGCTACGGAC GTCGTTACCG TTGTTGCAAC GCGTTTGATA ATTGACCGCT
2801 ACTACTTACT CTAGCTTCCC GGCAACAATT AATAGACTGG ATGGAGGCGG
     TGATGAATGA GATCGAAGGG CCGTTGTTAA TTATCTGACC TACCTCCGCC
2851 ATAAAGTTGC AGGACCACTT CTGCGCTCGG CCCTTCCGGC TGGCTGGTTT
     TATTTCAACG TCCTGGTGAA GACGCGAGCC GGGAAGGCCG ACCGACCAAA
2901 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGCG GTATCATTGC
     TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCGC CATAGTAACG
2951 AGCACTGGGG CCAGATGGTA AGCCCTCCCG TATCGTAGTT ATCTACACGA
     TCGTGACCCC GGTCTACCAT TCGGGAGGGC ATAGCATCAA TAGATGTGCT
3001 CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT CGCTGAGATA
     GCCCCTCAGT CCGTTGATAC CTACTTGCTT TATCTGTCTA GCGACTCTAT
3051 GGTGCCTCAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA
     CCACGGAGTG ACTAATTCGT AACCATTGAC AGTCTGGTTC AAATGAGTAT
3101 TATACTTTAG ATTGATTTAA AACTTCATTT TTAATTTAAA AGGATCTAGG
     ATATGAAATC TAACTAAATT TTGAAGTAAA AATTAAATTT TCCTAGATCC
3151 TGAAGATCCT TTTTGATAAT CTCATGACCA AAATCCCTTA ACGTGAGTTT
     ACTTCTAGGA AAAACTATTA GAGTACTGGT TTTAGGGAAT TGCACTCAAA
3201 TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTTCTTG
     AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC
```

```
3251 AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAAACCAC
     TCTAGGAAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTTTTGGTG
3301 CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT
     GCGATGGTCG CCACCAAACA AACGGCCTAG TTCTCGATGG TTGAGAAAAA
3351 CCGAAGGTAA CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCCTTCT
     GGCTTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT GACAGGAAGA
3401 AGTGTAGCCG TAGTTAGGCC ACCACTTCAA GAACTCTGTA GCACCGCCTA
     TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT
3451 CATACCTCGC TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT
     GTATGGAGCG AGACGATTAG GACAATGGTC ACCGACGACG GTCACCGCTA
3501 AAGTCGTGTC TTACCGGGTT GGACTCAAGA CGATAGTTAC CGGATAAGGC
     TTCAGCACAG AATGGCCCAA CCTGAGTTCT GCTATCAATG GCCTATTCCG
3551 GCAGCGGTCG GGCTGAACGG GGGGTTCGTG CACACAGCCC AGCTTGGAGC
     CGTCGCCAGC CCGACTTGCC CCCCAAGCAC GTGTGTCGGG TCGAACCTCG
3601 GAACGACCTA CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC
     CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTCG
3651 GCCACGCTTC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG
     CGGTGCGAAG GGCTTCCCTC TTTCCGCCTG TCCATAGGCC ATTCGCCGTC
3701 GGTCGGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA AACGCCTGGT
     CCAGCCTTGT CCTCTCGCGT GCTCCCTCGA AGGTCCCCCT TTGCGGACCA
3751 ATCTTTATAG TCCTGTCGGG TTTCGCCACC TCTGACTTGA GCGTCGATTT
     TAGAAATATC AGGACAGCCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA
3801 TTGTGATGCT CGTCAGGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC
     AACACTACGA GCAGTCCCCC CGCCTCGGAT ACCTTTTTGC GGTCGTTGCG
3851 GGCCTTTTTA CGGTTCCTGG CCTTTTGCTG GCCTTTTGCT CACATGTTCT
     CCGGAAAAAT GCCAAGGACC GGAAAACGAC CGGAAAACGA GTGTACAAGA
3901 TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG
     AAGGACGCAA TAGGGGACTA AGACACCTAT TGGCATAATG GCGGAAACTC
3951 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT
     ACTCGACTAT GGCGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA
4001 GAG
     CTC
```

Fig. 2

Structure of pIBA/INSLys(31B)Arg(32B) plasmid encoding the recombinant protein lisarginsulin INS.

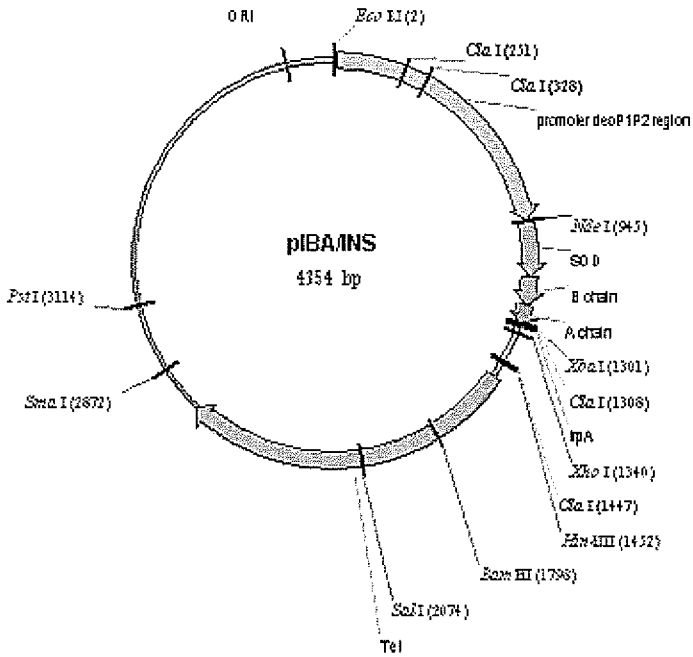


Fig. 3

Fig. 4

Nucleotide sequence and amino acid sequence of pIBA/INSLys(31B)Arg(32B) plasmid.

```
   1    GAATTCGGAC CGTTGGCGAT GTGCGGTTTG CTACATTCAC AGATGTTCTT
        CTTAAGCCTG GCAACCGCTA CACGCCAAAC GATGTAAGTG TCTACAAGAA
  51    CGCCACTTCC AGCAGCAGGT CATCAGGGGT GATTTCAGGA TCGTAGATAA
        GCGGTGAAGG TCGTCGTCCA GTAGTCCCCA CTAAAGTCCT AGCATCTATT
 101    AGGTCAGGTT CGGTGAAACC TGCTTCAACT CTGCATCTGC ACGTAAGATC
        TCCAGTCCAA GCCACTTTGG ACGAAGTTGA GACGTAGACG TGCATTCTAG
 151    GCGCGGGTAA TGGGCGAATC AGACGGGCCG ATATTGGCGT GCATAAAGGC
        CGCGCCCATT ACCCGCTTAG TCTGCCCGGC TATAACCGCA CGTATTTCCG
 201    GTCTGGCAGG GTTCTGTCGA GGTAACGCCA GAAACGTTTT ATTCGAACAT
        CAGACCGTCC CAAGACAGCT CCATTGCGGT CTTTGCAAAA TAAGCTTGTA
 251    CGATCTCGTC TTGTGTTAGA ATTAATTCTA ACATACGGTT GCAACAACGC
        GCTAGAGCAG AACACAATCT TAATTAAGAT TGTATGCCAA CGTTGTTGCG
 301    ATCCAGTTGC CCCAGGTAGA CCGGCATCGA TGTGACCGAC GGTACGTGGT
        TAGGTCAACG GGGTCCATCT GGCCGTAGCT ACACTGGCTG CCATGCACCA
 351    GGTAAAGAAT GGTCAGCAGA GAGAGTGCGT CATCAAGATC TTTCGCGCCT
        CCATTTCTTA CCAGTCGTCT CTCTCACGCA GTAGTTCTAG AAAGCGCGGA
 401    TCCAGCTCCA GCCATTCGGA ACCGTTCGCC AGAAAACGGG CGTAATCGGG
        AGGTCGAGGT CGGTAAGCCT TGGCAAGCGG TCTTTTGCCC GCATTAGCCC
 451    TAAGACATAG CGCGGTTTGT ACGGCGCATG ACCTTCAAAC ATATCGCAGA
        ATTCTGTATC GCGCCAAACA TGCCGCGTAC TGGAAGTTTG TATAGCGTCT
 501    TTACACCTTC ATCCAGCGCG CGGCGGGCTT CGGCAGGAAG CTGTGGGTAA
        AATGTGGAAG TAGGTCGCGC GCCGCCCGAA GCCGTCCTTC GACACCCATT
 551    GGCAGATTGT TTTCTGCTTC CAGTGCCAGA AAATGGCGCT TCTGCTCCGG
        CCGTCTAACA AAAGACGAAG GTCACGGTCT TTTACCGCGA AGACGAGGCC
 601    GCTAAGCACT GGGCTGGTGA CAATTTGCTG GCAACGTTGT TGCAGTGCAT
        CGATTCGTGA CCCGACCACT GTTAAACGAC CGTTGCAACA ACGTCACGTA
 651    TTTCATGAGA AGTGGGCATC TTCTTTTCCT TTTATGCCGA AGGTGATGCG
        AAAGTACTCT TCACCCGTAG AAGAAAAGGA AAATACGGCT TCCACTACGC
 701    CCATTGTAAG AAGTTTCGTG ATGTTCACTT TGATCCTGAT GCGTTTGCCA
        GGTAACATTC TTCAAAGCAC TACAAGTGAA ACTAGGACTA CGCAAACGGT
 751    CCACTGACGC ATTCATTTGA AAGTGAATTA TTTGAACCAG ATCGCATTAC
        GGTGACTGCG TAAGTAAACT TTCACTTAAT AAACTTGGTC TAGCGTAATG
 801    AGTGATGCAA ACTTGTAAGT AGATTTCCTT AATTGTGATG TGTATCGAAG
        TCACTACGTT TGAACATTCA TCTAAAGGAA TTAACACTAC ACATAGCTTC
 851    TGTGTTGCGG AGTAGATGTT AGAATACTAA CAAACTCGCA AGGTGAATTT
        ACACAACGCC TCATCTACAA TCTTATGATT GTTTGAGCGT TCCACTTAAA
                                                        MetAla·
 901    TATTGGCGAC AAGCCTAGGT TTGTTTAACT TTAAGGAGAA ATCATATGGC
        ATAACCGCTG TTCGGATCCA AACAAATTGA AATTCCTCTT TAGTATACCG
       ·ThrLysAla ValSerValLeu LysGlyAsp GlyProVal GlnGlyIleIle·
 951    TACTAAAGCT GTTTCTGTTT TAAAAGGTGA TGGTCCAGTT CAAGGAATTA
        ATGATTTCGA CAAAGACAAA ATTTTCCACT ACCAGGTCAA GTTCCTTAAT
       ·IAsnPheGlu GlnLysGlu SerAsnGlyPro ValLysVal TrpGlySer
1001    TTAATTTTGA ACAAAAAGAA AGTAATGGAC CAGTTAAAGT ATGGGGAAGT
        AATTAAAACT TGTTTTTCTT TCATTACCTG GTCAATTTCA TACCCCTTCA
      IleLysGlyLeu ThrGluGly LeuHisGly PheHisValHis GluPheGly·
1051    ATTAAAGGAC TTACTGAAGG CCTGCATGGA TTCCATGTTC ATGAGTTTGG
        TAATTTCCTG AATGACTTCC GGACGTACCT AAGGTACAAG TACTCAAACC
       ·AspAsnThr AlaGlySerThr SerAlaGly ProArgPhe ValAsnGlnHis
1101    AGATAATACA GCAGGCAGTA CTAGTGCAGG TCCTCGTTTT GTCAACCAGC
        TCTATTATGT CGTCCGTCAT GATCACGTCC AGGAGCAAAA CAGTTGGTCG
       ·HLeuCysGly SerHisLeu ValGluAlaLeu TyrLeuVal CysGlyGlu
1151    ACCTGTGTGG TTCTCACCTG GTTGAAGCAC TGTACCTGGT ATGTGGCGAA
        TGGACACACC AAGAGTGGAC CAACTTCGTG ACATGGACCA TACACCGCTT
      ArgGlyPhePhe TyrThrPro LysThrLys ArgGlyIleVal GluGlnCys
1201    CGTGGTTTCT TCTACACTCC TAAAACAAAG CGCGGCATCG TTGAACAGTG
        GCACCAAAGA AGATGTGAGG ATTTTGTTTC GCGCCGTAGC AACTTGTCAC
       ·CysThrSer IleCysSerLeu TyrGlnLeu GluAsnTyr CysAsn***
```

```
1251    CTGTACCTCT ATCTGTTCCC TGTACCAACT GGAGAACTAC TGCAATTAAT
        GACATGGAGA TAGACAAGGG ACATGGTTGA CCTCTTGATG ACGTTAATTA
1301    CTAGAATCGA TGCCCGCTTA ATGAGCGGGC TTTTTTTTCT CGAGGACGTC
        GATCTTAGCT ACGGGCGAAT TACTCGCCCG AAAAAAAAGA GCTCCTGCAG
1351    TAAGAAACCA TTATTATCAT GACATTAACC TATAAAAATA GGCGTATCAC
        ATTCTTTGGT AATAATAGTA CTGTAATTGG ATATTTTTAT CCGCATAGTG
1401    GAGGCCCTTT CGTCTTCAAG ACCTCTCATG TTTGACAGCT TATCATCGAT
        CTCCGGGAAA GCAGAAGTTC TGGAGAGTAC AAACTGTCGA ATAGTAGCTA
1451    AAGCTTTAAT GCGGTAGTTT ATCACAGTTA AATTGCTAAC GCAGTCAGGC
        TTCGAAATTA CGCCATCAAA TAGTGTCAAT TTAACGATTG CGTCAGTCCG
1501    ACCGTGTATG AAATCTAACA ATGCGCTCAT CGTCATCCTC GGCACCGTCA
        TGGCACATAC TTTAGATTGT TACGCGAGTA GCAGTAGGAG CCGTGGCAGT
1551    CCCTGGATGC TGTAGGCATA GGCTTGGTTA TGCCGGTACT GCCGGGCCTC
        GGGACCTACG ACATCCGTAT CCGAACCAAT ACGGCCATGA CGGCCCGGAG
1601    TTGCGGGATA TCGTCCATTC CGACAGCATC GCCAGTCACT ATGGCGTGCT
        AACGCCCTAT AGCAGGTAAG GCTGTCGTAG CGGTCAGTGA TACCGCACGA
1651    GCTAGCGCTA TATGCGTTGA TGCAATTTCT ATGCGCACCC GTTCTCGGAG
        CGATCGCGAT ATACGCAACT ACGTTAAAGA TACGCGTGGG CAAGAGCCTC
1701    CACTGTCCGA CCGCTTTGGC CGCCGCCCAG TCCTGCTCGC TTCGCTACTT
        GTGACAGGCT GGCGAAACCG GCGGCGGGTC AGGACGAGCG AAGCGATGAA
1751    GGAGCCACTA TCGACTACGC GATCATGGCG ACCACACCCG TCCTGTGGAT
        CCTCGGTGAT AGCTGATGCG CTAGTACCGC TGGTGTGGGC AGGACACCTA
1801    CCTCTACGCC GGACGCATCG TGGCCGGCAT CACCGGCGCC ACAGGTGCGG
        GGAGATGCGG CCTGCGTAGC ACCGGCCGTA GTGGCCGCGG TGTCCACGCC
1851    TTGCTGGCGC CTATATCGCC GACATCACCG ATGGGGAAGA TCGGGCTCGC
        AACGACCGCG GATATAGCGG CTGTAGTGGC TACCCCTTCT AGCCCGAGCG
1901    CACTTCGGGC TCATGAGCGC TTGTTTCGGC GTGGGTATGG TGGCAGGCCC
        GTGAAGCCCG AGTACTCGCG AACAAAGCCG CACCCATACC ACCGTCCGGG
1951    CGTGGCCGGG GGACTGTTGG GCGCCATCTC CTTGCATGCA CCATTCCTTG
        GCACCGGCCC CCTGACAACC CGCGGTAGAG GAACGTACGT GGTAAGGAAC
2001    CGGCGGCGGT GCTCAACGGC CTCAACCTAC TACTGGGCTG CTTCCTAATG
        GCCGCCGCCA CGAGTTGCCG GAGTTGGATG ATGACCCGAC GAAGGATTAC
2051    CAGGAGTCGC ATAAGGGAGA GCGTCGACCG ATGCCCTTGA GAGCCTTCAA
        GTCCTCAGCG TATTCCCTCT CGCAGCTGGC TACGGGAACT CTCGGAAGTT
2101    CCCAGTCAGC TCCTTCCGGT GGGCGCGGGG CATGACTATC GTCGCCGCAC
        GGGTCAGTCG AGGAAGGCCA CCCGCGCCCC GTACTGATAG CAGCGGCGTG
2151    TTATGACTGT CTTCTTTATC ATGCAACTCG TAGGACAGGT GCCGGCAGCG
        AATACTGACA GAAGAAATAG TACGTTGAGC ATCCTGTCCA CGGCCGTCGC
2201    CTCTGGGCCA TTTTCGGCGA GGACCGCTTT CGCTGGAGCG CGACGATGAT
        GAGACCCGGT AAAAGCCGCT CCTGGCGAAA GCGACCTCGC GCTGCTACTA
2251    CGGCCTGTCG CTTGCGGTAT TCGGAATCTT GCACGCCCTC GCTCAAGCCT
        GCCGGACAGC GAACGCCATA AGCCTTAGAA CGTGCGGGAG CGAGTTCGGA
2301    TCGTCACTGG TCCCGCCACC AAACGTTTCG GCGAGAAGCA GGCCATTATC
        AGCAGTGACC AGGGCGGTGG TTTGCAAAGC CGCTCTTCGT CCGGTAATAG
2351    GCCGGCATGG CGGCCGACGC GCTGGGCTAC GTCTTGCTGG CGTTCGCGAC
        CGGCCGTACC GCCGGCTGCG CGACCCGATG CAGAACGACC GCAAGCGCTG
2401    GCGAGGCTGG ATGGCCTTCC CCATTATGAT TCTTCTCGCT TCCGGCGGCA
        CGCTCCGACC TACCGGAAGG GGTAATACTA AGAAGAGCGA AGGCCGCCGT
2451    TCGGGATGCC CGCGTTGCAG GCCATGCTGT CCAGGCAGGT AGATGACGAC
        AGCCCTACGG GCGCAACGTC CGGTACGACA GGTCCGTCCA TCTACTGCTG
2501    CATCAGGGAC AGCTTCAAGG ATCGCTCGCG GCTCTTACCA GCCTAACTTC
        GTAGTCCCTG TCGAAGTTCC TAGCGAGCGC CGAGAATGGT CGGATTGAAG
2551    GATCATTGGA CCGCTGATCG TCACGGCGAT TTATGCCGCC TCGGCGAGCA
        CTAGTAACCT GGCGACTAGC AGTGCCGCTA AATACGGCGG AGCCGCTCGT
2601    CATGGAACGG GTTGGCATGG ATTGTAGGCG CCGCCCTATA CCTTGTCTGC
        GTACCTTGCC CAACCGTACC TAACATCCGC GGCGGGATAT GGAACAGACG
2651    CTCCCCGCGT TGCGTCGCGG TGCATGGAGC CGGGCCACCT CGACCTGAAT
        GAGGGGCGCA ACGCAGCGCC ACGTACCTCG GCCCGGTGGA GCTGGACTTA
2701    GGAAGCCGGC GGCACCTCGC TAACGGATTC ACCACTCCAA GAATTGGAGC
        CCTTCGGCCG CCGTGGAGCG ATTGCCTAAG TGGTGAGGTT CTTAACCTCG
2751    CAATCAATTC TTGCGGAGAA CTGTGAATGC GCAAACCAAC CCTTGGCAGA
        GTTAGTTAAG AACGCCTCTT GACACTTACG CGTTTGGTTG GGAACCGTCT
2801    ACATATCCAT CGCGTCCGCC ATCTCCAGCA GCCGCACGCG GCGCATCTCG
        TGTATAGGTA GCGCAGGCGG TAGAGGTCGT CGGCGTGCGC CGCGTAGAGC
2851    GGCAGCGTTG GGTCCTGGCC CGGGACTCAC CAGTCACAGA AAAGCATCTT
        CCGTCGCAAC CCAGGACCGG GCCCTGAGTG GTCAGTGTCT TTTCGTAGAA
2901    ACGGATGGCA TGACAGTAAG AGAATTATGC AGTGCTGCCA TAACCATGAG
        TGCCTACCGT ACTGTCATTC TCTTAATACG TCACGACGGT ATTGGTACTC
```

```
2951    TGATAACACT GCGGCCAACT TACTTCTGAC AACGATCGGA GGACCGAAGG
        ACTATTGTGA CGCCGGTTGA ATGAAGACTG TTGCTAGCCT CCTGGCTTCC
3001    AGCTAACCGC TTTTTTGCAC AACATGGGGG ATCATGTAAC TCGCCTTGAT
        TCGATTGGCG AAAAAACGTG TTGTACCCCC TAGTACATTG AGCGGAACTA
3051    CGTTGGGAAC CGGAGCTGAA TGAAGCCATA CCAAACGACG AGCGTGACAC
        GCAACCCTTG GCCTCGACTT ACTTCGGTAT GGTTTGCTGC TCGCACTGTG
3101    CACGATGCCT GCAGCAATGG CAACAACGTT GCGCAAACTA TTAACTGGCG
        GTGCTACGGA CGTCGTTACC GTTGTTGCAA CGCGTTTGAT AATTGACCGC
3151    AACTACTTAC TCTAGCTTCC CGGCAACAAT TAATAGACTG GATGGAGGCG
        TTGATGAATG AGATCGAAGG GCCGTTGTTA ATTATCTGAC CTACCTCCGC
3201    GATAAAGTTG CAGGACCACT TCTGCGCTCG GCCCTTCCGG CTGGCTGGTT
        CTATTTCAAC GTCCTGGTGA AGACGCGAGC CGGGAAGGCC GACCGACCAA
3251    TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCGC GGTATCATTG
        ATAACGACTA TTTAGACCTC GGCCACTCGC ACCCAGAGCG CCATAGTAAC
3301    CAGCACTGGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACACG
        GTCGTGACCC CGGTCTACCA TTCGGGAGGG CATAGCATCA ATAGATGTGC
3351    ACGGGGAGTC AGGCAACTAT GGATGAACGA AATAGACAGA TCGCTGAGAT
        TGCCCCTCAG TCCGTTGATA CCTACTTGCT TTATCTGTCT AGCGACTCTA
3401    AGGTGCCTCA CTGATTAAGC ATTGGTAACT GTCAGACCAA GTTTACTCAT
        TCCACGGAGT GACTAATTCG TAACCATTGA CAGTCTGGTT CAAATGAGTA
3451    ATATACTTTA GATTGATTTA AAACTTCATT TTTAATTTAA AAGGATCTAG
        TATATGAAAT CTAACTAAAT TTTGAAGTAA AAATTAAATT TTCCTAGATC
3501    GTGAAGATCC TTTTTGATAA TCTCATGACC AAAATCCCTT AACGTGAGTT
        CACTTCTAGG AAAAACTATT AGAGTACTGG TTTTAGGGAA TTGCACTCAA
3551    TTCGTTCCAC TGAGCGTCAG ACCCCGTAGA AAAGATCAAA GGATCTTCTT
        AAGCAAGGTG ACTCGCAGTC TGGGGCATCT TTTCTAGTTT CCTAGAAGAA
3601    GAGATCCTTT TTTTCTGCGC GTAATCTGCT GCTTGCAAAC AAAAAAACCA
        CTCTAGGAAA AAAAGACGCG CATTAGACGA CGAACGTTTG TTTTTTTGGT
3651    CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT
        GGCGATGGTC GCCACCAAAC AAACGGCCTA GTTCTCGATG GTTGAGAAAA
3701    TCCGAAGGTA ACTGGCTTCA GCAGAGCGCA GATACCAAAT ACTGTCCTTC
        AGGCTTCCAT TGACCGAAGT CGTCTCGCGT CTATGGTTTA TGACAGGAAG
3751    TAGTGTAGCC GTAGTTAGGC CACCACTTCA AGAACTCTGT AGCACCGCCT
        ATCACATCGG CATCAATCCG GTGGTGAAGT TCTTGAGACA TCGTGGCGGA
3801    ACATACCTCG CTCTGCTAAT CCTGTTACCA GTGGCTGCTG CCAGTGGCGA
        TGTATGGAGC GAGACGATTA GGACAATGGT CACCGACGAC GGTCACCGCT
3851    TAAGTCGTGT CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAGG
        ATTCAGCACA GAATGGCCCA ACCTGAGTTC TGCTATCAAT GGCCTATTCC
3901    CGCAGCGGTC GGGCTGAACG GGGGGTTCGT GCACACAGCC CAGCTTGGAG
        GCGTCGCCAG CCCGACTTGC CCCCCAAGCA CGTGTGTCGG GTCGAACCTC
3951    CGAACGACCT ACACCGAACT GAGATACCTA CAGCGTGAGC TATGAGAAAG
        GCTTGCTGGA TGTGGCTTGA CTCTATGGAT GTCGCACTCG ATACTCTTTC
4001    CGCCACGCTT CCCGAAGGGA GAAAGGCGGA CAGGTATCCG GTAAGCGGCA
        GCGGTGCGAA GGGCTTCCCT CTTTCCGCCT GTCCATAGGC CATTCGCCGT
4051    GGGTCGGAAC AGGAGAGCGC ACGAGGGAGC TTCCAGGGGG AAACGCCTGG
        CCCAGCCTTG TCCTCTCGCG TGCTCCCTCG AAGGTCCCCC TTTGCGGACC
4101    TATCTTTATA GTCCTGTCGG GTTTCGCCAC CTCTGACTTG AGCGTCGATT
        ATAGAAATAT CAGGACAGCC CAAAGCGGTG GAGACTGAAC TCGCAGCTAA
4151    TTTGTGATGC TCGTCAGGGG GGCGGAGCCT ATGGAAAAAC GCCAGCAACG
        AAACACTACG AGCAGTCCCC CCGCCTCGGA TACCTTTTTG CGGTCGTTGC
4201    CGGCCTTTTT ACGGTTCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTC
        GCCGGAAAAA TGCCAAGGAC CGGAAAACGA CCGGAAAACG AGTGTACAAG
4251    TTTCCTGCGT TATCCCCTGA TTCTGTGGAT AACCGTATTA CCGCCTTTGA
        AAAGGACGCA ATAGGGGACT AAGACACCTA TTGGCATAAT GGCGGAAACT
4301    GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC AGCGAGTCAG
        CACTCGACTA TGGCGAGCGG CGTCGGCTTG CTGGCTCGCG TCGCTCAGTC
4351    TGAG
        ACTC
```

Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/PL2017/050003

A. CLASSIFICATION OF SUBJECT MATTER
C07K14/62 (2006.01); C12P21/02 (2006.01)

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)
C07K14/62; C12P21/02; C12N15/17; A61K38/28

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)
EPODOC, WPI, STN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO2014122653 A1 (VALIN TECHNOLOGIES LTD) 14.08.2014 (paragraph [029], [032], [083]; fig.4; Seq Id No.1; Seq Id No.2)	1-2 3-14
X	EP0195691 A1 (NOVO INDUSTRI AS) 24.09.1986 (ex. 7, claim 5)	1-2
X	EP0622376 A1 (HOECHST AKTIENGESELLSCHAFT) 02.11.1994 (ex. 1, Seq Id No.3)	1-2
X	WO2009104199 A1 (BIOCON LTD) 27.08.2009 (p.46, Seq Id No.1)	1-2
Y	WO9620724 A1 (BIO-TECHNOLOGY GENERAL CORP) 11.07.1996 (figs. 3 and 4, p. 22 l. 29 – p. 23 l. 27, p. 24 l. 9 – p. 25 l. 5, p. 19 ll. 24-25, p. 29 ll. 10-25, p. 18 ll. 12-34)	3-14
Y	WO2010002283 A2 (INSTYTUT BIOTECHNOLOGII I ANTYBIOTYKÓW) 07.01.2010 (p. 5 l. 25 – p. 6 l. 8, p. 1 ll. 23-25)	3-14

☐ 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	Date of mailing of the international search report
18 May 2017	19 May 2017

Name and mailing address of the ISA/	Authorized officer
Visegrad Patent Institute / Branch Office PL Al. Niepodległości 188, 00-950 Warsaw, Poland Facsimile No. +48 22 579 00 01	Agnieszka Ucińska Telephone No. +48 22 579 06 25

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO2014122653 A1	14.08.2014	None	
EP0195691 A1	24.09.1986	AT69266 (T)	15.11.1991
		AU5500586 (A)	25.09.1986
		CA1340823 (C)	16.11.1999
		DK125786 (A)	17.11.1986
		DK129385 (A)	23.09.1986
		ES8802402 (A1)	16.08.1988
		ES8708251 (A1)	16.12.1987
		FI861198 (A)	23.09.1986
		GR860764 (B)	21.07.1986
		IE860737 (L)	22.09.1986
		JPH06133777 (A)	17.05.1994
		JPH0823985 (A)	30.01.1996
		JPS61221200 (A)	01.10.1986
		MX163837 (B)	25.06.1992
		NO861120 (A)	23.09.1986
		NZ215570 (A)	27.07.1989
		PT82241 (A)	01.04.1986
		ZA8601953 (B)	26.11.1986
EP0622376 A1	02.11.1994	AT203998 (T)	15.08.2001
		DK0622376 (T3)	12.11.2001
		ES2161726 (T3)	16.12.2001
		GR3036840 (T3)	31.01.2002
		JPH06321990 (A)	22.11.1994
		US5466666 (A)	14.11.1995
WO2009104199 A1	27.08.2009	BRPI0822775 (A2)	30.06.2015
		CN102015762 (A)	13.04.2011
		EP2242767 (A1)	27.10.2010
		IL207646 (A)	31.08.2016
		JP2011514151 (A)	06.05.2011
		JP2015083019 (A)	30.04.2015
		KR20100117612 (A)	03.11.2010
		MX2010009033 (A)	21.12.2010
		RU2010138656 (A)	27.03.2012
		US2011236925 (A1)	29.09.2011

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO2010002283 A2	07.01.2010	CA2729938 (A1)	07.01.2010
		CN102083855 (A)	01.06.2011
		EA201170149 (A1)	30.06.2011
		EP2321344 (A2)	18.05.2011
		EP2371853 (A2)	05.10.2011
		JP2011526886 (A)	20.10.2011
		PL385586 (A1)	18.01.2010
		US2011136736 (A1)	09.06.2011
		US2014121353 (A1)	01.05.2014
WO9620724 A1	11.07.1996	AT350053 (T)	15.01.2007
		AU1550095 (A)	24.07.1996
		BR9408638 (A)	21.09.2004
		CA2208095 (A1)	11.07.1996
		CZ9702031 (A3)	16.09.1998
		DE69434909 (T2)	09.08.2007
		DE95907193 (T1)	10.02.2005
		DK0871474 (T3)	07.05.2007
		EP0871474 (A1)	21.10.1998
		ES2279510 (T3)	16.08.2007
		HK1013594 (A1)	30.06.2006
		HUT78070 (A)	30.08.1999
		JPH10511849 (A)	17.11.1998
		NZ279002 (A)	25.02.1999
		PL321039 (A1)	24.11.1997
		PT871474 (E)	28.02.2007