(12) PATENT **(11)** Application No. **AU 199880763 B2**
(19) AUSTRALIAN PATENT OFFICE **(10)** Patent No. **744824**

(54) Title
Improved process for obtaining insulin precursors having correctly bonded cystine bridges

(51)[7] International Patent Classification(s)
C07K 014/62

(21) Application No: **199880763** (22) Application Date: **1998.08.17**

(30) Priority Data

(31) Number	(32) Date	(33) Country
19735711.3	**1997.08.18**	**DE**

(43) Publication Date : **1999.02.25**
(43) Publication Journal Date : **1999.02.25**
(44) Accepted Journal Date : **2002.03.07**

(71) Applicant(s)
Aventis Pharma Deutschland GmbH

(72) Inventor(s)
Franz-Josef Rubroder; Reinhold Keller

(74) Agent/Attorney
WATERMARK PATENT and TRADEMARK ATTORNEYS,Locked Bag 5,HAWTHORN VIC 3122

Abstract

Improved process for obtaining insulin precursors having correctly bonded cystine bridges

The present invention relates to an improved process for obtaining a precursor of insulins or insulin derivatives having correctly bonded cystine bridges in the presence of cysteine or cysteine hydrochloride and of a chaotropic auxiliary.

AUSTRALIA

Patents Act 1990

ORIGINAL
COMPLETE SPECIFICATION
STANDARD PATENT

Application Number:

Lodged:

Invention Title: IMPROVED PROCESS FOR OBTAINING INSULIN PRECURSORS HAVING CORRECTLY BONDED CYSTINE BRIDGES

The following statement is a full description of this invention, including the best method of performing it known to us :-

Description

Improved process for obtaining insulin precursors having correctly bonded cystine bridges

The present invention relates to an improved process for obtaining a precursor of insulin or insulin derivatives having correctly bonded cystine bridges in the presence of cysteine or cysteine hydrochloride and of a chaotropic auxiliary.

Human insulin is a protein with two amino acid chains together having 51 amino acid residues. 6 cysteine residues are found in the two amino acid chains, each two cysteine residues being bonded to one another via a disulfide bridge. In biologically active human insulin, the A and B chains are bonded to one another via two cystine bridges, and a further cystine bridge is found in the A chain. Within a human insulin molecule, looked at statistically, there are 15 possibilities for the formation of disulfide bridges. In biologically active human insulin, only one of the 15 possibilities is found. The following cysteine residues are linked to one another in human insulin:

A 6 - A 11

A 7 - B 7

A 20 - B 19

The letters A and B represent the respective insulin amino acid chain and the number represents the position of the amino acid residue, which is counted from the amino to the carboxyl end of the respective amino acid chain. Disulfide bridges can also be formed between two human insulin molecules such that incalculably many different disulfide bridges can easily result.

A known process for the preparation of human insulin is based on the use of human proinsulin. Human proinsulin is a protein having a linear amino acid chain of 86 amino acid residues, the B and A chains of the human insulin being bonded to one another via a C peptide having 35 amino acid residues. The formation of the disulfide bridges found in human insulin takes place via an intermediate, the cysteine residues of the human insulin being provided with a sulfur protective group, e.g. an S-sulfonate ($-S-SO_3^-$) group (EP 0 037 255). A process for obtaining proinsulin having correctly bonded cystine bridges is additionally known (Biochemistry, 60, (1968), pages 622 to 629), which starts from proinsulin obtained from porcine pancreas, in which the cysteine residues are present as thiol residues (-SH). The term "correctly bonded cystine bridges" is understood as meaning the disulfide bridges which are found in biologically active insulin from mammals.

Recombinant DNA processes allow precursors of insulin or insulin derivatives, in particular human proinsulin or proinsulin which has an amino acid sequence and/or amino acid chain length differing from human insulin to be prepared in microorganisms. The proinsulins prepared from genetically modified Escherichia coli cells do not have any correctly bonded cystine bridges. A process for obtaining human insulin using E. coli (EP 0 055 945) is based on the following process steps:
Fermentation of the microorganisms - cell disruption - isolation of the fusion protein - cyanogen halide cleavage of the fusion protein - isolation of the cleavage product having the proinsulin sequence - protection of the cystine residues of proinsulin by S-sulfonate groups - chromatographic purification of the S-sulfonate - formation of the correctly bonded cystine bridges - desalting of the proinsulin - chromatographic purification of the proinsulin having correctly bonded cystine bridges - concentration of the proinsulin solution - chromatographic purification of the concentrated proinsulin solution - enzymatic cleavage of the proinsulin to obtain human insulin - chromatographic purification of the resulting human insulin.

Disadvantages of this process are the number of process steps and the losses in the purification steps, which lead to a low yield of insulin. Because of the multistage process route, considerable losses have to be accepted. From the stage of the isolated fusion protein via cyanogen halide cleavage, sulfitolysis and purification of the proinsulin, an up to 40% loss of proinsulin has to be expected (EP 0 055 945). Similarly high losses can occur in the course of the subsequent purification steps as far as the final product.

Yield increases in the preparation of human insulin or insulin derivatives by recombinant DNA means can be achieved if the number of process steps necessary can be significantly reduced.

EP 0 600 372 A1 (or US 5,473,049) and EP 0 668 292 A2 disclose an appropriately improved process for obtaining insulins or insulin derivatives, in which the insulin precursor or precursor of the insulin derivative whose cystine bridges are not present in correctly linked form is reacted in the presence of a mercaptan, for example cysteine, and of at least one chaotropic auxiliary, for example urea or guanidine hydrochloride, to give an insulin precursor or precursor of the insulin derivative having correctly bonded cystine bridges. In the known process, these proteins are first dissolved in a very low concentration in aqueous solutions of a chaotropic auxiliary or of mixtures of various chaotropic auxiliaries. The protein mixture is then mixed with an aqueous mercaptan solution.

Surprisingly, it has now been found that the yields of correctly folded precursors of insulins or insulin derivatives can be increased and the reaction times for the folding process can be reduced by not bringing the precursor into solution in a first step by means of the chaotropic auxiliary, but by first introducing the mercaptan, namely cysteine or cysteine hydrochloride, into the aqueous suspension of the precursor and only

bringing about the dissolution of the precursor in a subsequent step by introduction into an aqueous solution of the chaotropic auxiliary and finally bringing about the correct folding of the precursor by dilution of the mixture to a preferred cysteine or cysteine hydrochloride concentration with introduction of the mixture into an appropriate amount of water.

Accordingly, the present invention relates to a process for obtaining a precursor of insulins or insulin derivatives having correctly bonded cystine bridges in the presence of cysteine or cysteine hydrochloride and of a chaotropic auxiliary, which comprises successively carrying out the following steps:

(a) mixing an aqueous suspension of the precursor of insulins or insulin derivatives with an amount of cysteine or cysteine hydrochloride which results in 1 to 15 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor,

(b) introducing the cysteine- or cysteine hydrochloride-containing suspension of the precursor into a 4 to 9 molar solution of the chaotropic auxiliary at a pH of approximately 8 to approximately 11.5 and a temperature of approximately 15 to approximately 55°C, keeping the mixture obtained at this temperature for approximately 10 to 60 minutes and

(c) introducing the mixture at a pH of approximately 8 to approximately 11.5 and a temperature of approximately 5 to approximately 30°C into an amount of water which results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 1 to 5 mM and of the chaotropic auxiliary to 0.2 to 1.0 M.

Preferably, the process is one wherein

in step (a) the amount of cysteine or cysteine hydrochloride corresponds to an amount which results in 1 to 6 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor,

in step (b) the cysteine- or cysteine hydrochloride-containing suspension of the precursor is introduced into a 4 to 9 molar solution of the chaotropic auxiliary at a pH of 8 to 11 and a temperature of 30 to 45°C and the mixture obtained is kept for 20 to 40 minutes at this temperature and

in step (c) the mixture is introduced at a pH of 8 to 11 and at a temperature of 15 to 20°C into an amount of water which results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 1 to 5 mM and a concentration of the chaotropic auxiliary of 0.2 to 1.0 M.

Chaotropic auxiliaries are compounds which break hydrogen bridges in aqueous solution, for example ammonium sulfate, guanidine hydrochloride, ethylene carbonate, thiocyanate, dimethyl sulfoxide and urea.

In the process according to the present invention, the chaotropic auxiliary employed is preferably guanidine, guanidine hydrochloride or particularly preferably urea.

The concentration of the chaotropic auxiliary in step (b) of the process according to the invention is preferably 7.0 to 9M, the temperature in step (b) is preferably 40°C and the pH in step (b) is preferably 10 to 11.

In the process according to the invention, the pH in step (c) is preferably 10 to 11. In step (c) of the process according to the present invention, the amount of water into which the mixture is introduced is preferably selected such that this results in a dilution of the cysteine or cysteine hydrochloride

concentration in the mixture to 2.5 to 3 mM and a concentration of the chaotropic auxiliary of 0.5 M.

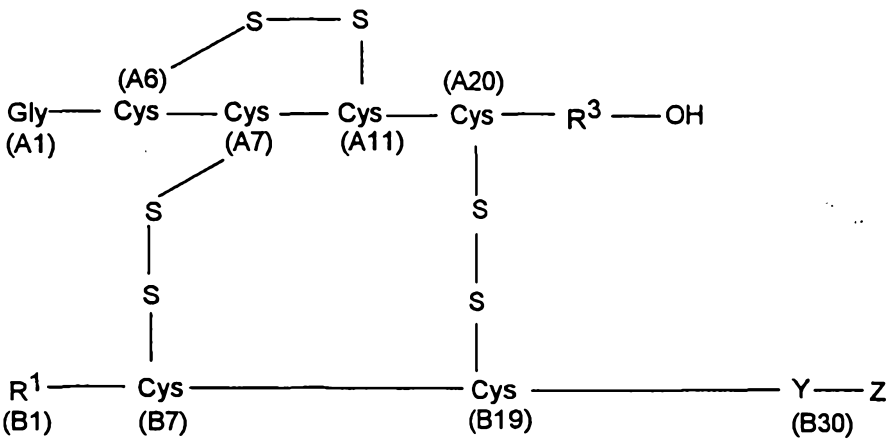
Particularly preferably, the process according to the invention is one wherein the concentration of the chaotropic auxiliary in step (b) is approximately 8 M, the temperature in step (b) is approximately 40°C, the pH in step (b) is approximately 10.6, the pH in step (c) is approximately 10.6 and in step (c) the amount of water results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 2.5 to 3 mM and a concentration of the chaotropic auxiliary of 0.5 M.

The result of the process according to the present invention is a precursor of insulins or insulin derivatives, in particular a proinsulin, whose cystine bridges are correctly bonded.

Insulin derivatives are derivatives of naturally occurring insulins, namely human insulin (see SEQ ID NO 1 = A chain of human insulin; see SEQ ID NO 2 = B chain of human insulin, sequence listing) or animal insulins, which differ by substitution of at least one naturally occurring amino acid residue and/or addition of at least one amino acid residue and/or organic residue of the corresponding, otherwise identical naturally occurring insulin.

From the precursor of the insulins or insulin derivatives having correctly bonded cystine bridges obtained with the aid of the process according to the present invention, it is finally possible according to the process described in EP 0 600 372 A1 (or US 5,473,049) or in EP 0 668 292 A2 to prepare an insulin or an insulin derivative having correctly bonded cystine bridges by enzymatic cleavage by means of trypsin or a tripsin-like enzyme and, if appropriate, additionally by means of carboxypeptidase B and subsequent purification on an adsorber resin.

The insulin or insulin derivative which can be prepared from the precursor can preferably be described by formula I



(I),

in which

Y is a genetically encodable amino acid residue,

Z is a) an amino acid residue from the group consisting of His, Arg and Lys,

b) a peptide having 2 or 3 amino acid residues, comprising the amino acid residue Arg or Lys at the carboxyl end of the peptide,

c) a peptide having 2 - 35 genetically encodable amino acids, comprising 1 to 5 histidine residues, or

d) OH,

R^1 is a phenylalanine residue (Phe) or a covalent bond,

R^3 is a genetically encodable amino acid residue,

where the radicals A2 - A20 of the amino acid sequence of the A chain of human insulin not shown for the simplification of the formula I correspond to animal insulin or an insulin derivative and the radicals B2 - B29 of the amino acid sequence of the B chain of human insulin not shown for the simplification of the formula I correspond to animal insulin or an insulin derivative.

The amino acid sequence of peptides and proteins is indicated from the N-terminal end of the amino acid chain onward. The details in formula I in brackets, e.g. A6, A20, B1, B7 or B19, correspond to the position of amino acid residues in the A or B chains of the insulin.

The term "genetically encodable amino acid residue" represents the amino acids Gly, Ala, Ser, Thr, Val, Leu, Ile, Asp, Asn, Glu, Gln, Cys, Met, Arg, Lys, His, Tyr, Phe, Trp, Pro and selenocysteine.

The terms "residues A2 - A20" and "residues B2 - B29" of animal insulin are understood as meaning, for example, the amino acid sequences of insulin from cattle, pigs or chickens. The term "residues A2 - A20" and "B2 - B29" of insulin derivatives represents the corresponding amino acid sequences of human insulin which are formed by the replacement of amino acids by other genetically encodable amino acids.

The A chain of human insulin, for example, has the following sequence (SEQ ID NO.: 1):

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu

Glu Asn Tyr Cys Asn.

The B chain of human insulin has the following sequence (SEQ ID NO.: 2):

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr.

In this case, in formula I R^3 is asparagine (Asn), R^1 is phenylalanine (Phe), Y is threonine (Thr) and Z is OH.

The process according to the present invention is accordingly particularly suitable for obtaining a precursor of insulins or insulin derivatives having

the formula II, whose cystine bridges (not shown in formula II) are correctly folded,

$$R^2\text{-}R^1\text{-(B2-B29)-Y-X-Gly-(A2-A20)-}R^3 \qquad \text{(II)},$$

in which

R^2 is a) a hydrogen atom,

b) an amino acid residue from the group consisting of lysine (Lys) and arginine (Arg) or

c) a peptide having 2 to 45 amino acid residues, comprising the amino acid residue lysine (Lys) or arginine (Arg) at the carboxyl end of the peptide

R^1 is a phenylalanine residue (Phe) or a covalent bond,

(B2-B29) are the amino acid residues in the positions B2 to B29 of the B chain of human insulin, animal insulin or an insulin derivative which is optionally varied in one or more of these positions,

Y is a genetically encodable amino acid residue,

X is a) an amino acid residue from the group consisting of lysine (Lys) and arginine (Arg),

b) a peptide having 2 to 35 amino acid residues, comprising the amino acid residue lysine (Lys) or arginine (Arg) at the N-terminal end and at the carboxyl end of the peptide, or

c) a peptide having 2 to 35 genetically encodable amino acids, comprising 1 to 5 histidine residues,

(A2-A20) are the amino acid residues in the positions A2 to A20 of the B chain of human insulin, animal insulin or an insulin derivative which is optionally varied in one or more of these positions and

R^3 is a genetically encodable amino acid residue.

1. Preferably, in formula II:

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 25 amino acid residues, comprising the amino acid residue arginine (Arg) at the carboxyl end of the peptide,

R^1 is a phenylalanine residue (Phe),

(B2-B29) are the amino acid residues in the positions B2 to B29 of the B chain of human insulin,

Y is an amino acid residue from the group consisting of alanine (Ala), threonine (Thr) and serine (Ser),

X is the amino acid residue arginine (Arg) or a peptide having the amino acid sequence of the C chain of human insulin,

(A2-A20) are the amino acid residues in the positions A2 to A20 of the B chain of human insulin and

R^3 is an amino acid residue from the group consisting of asparagine (Asn), serine (Ser) and glycine (Gly).

The C chain of human insulin has the following sequence (SEQ ID NO.: 3):

Arg Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg.

2. Preferably, in formula II:

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 15 amino acid residues, at whose carboxyl end is found an arginine residue (Arg),

R^1	is a phenylalanine residue (Phe),
(B2-B29)	are the amino acid residues in the positions B2 to B29 of the B chain of human insulin,
Y	is a threonine residue (Thr),
X	is the amino acid residue arginine (Arg) or a peptide having 2 to 35 amino acid residues, where at the beginning and at the end of the peptide there are two basic amino acid residues, in particular arginine (Arg) and/or lysine (Lys),
(A2-A20)	are the amino acid residues in the positions A2 to A20 of the B chain of human insulin and
R^3	is the amino acid residue asparagine (Asn) or glycine(Gly).

The residue Z of the insulin or of the insulin derivative of the formula I is, as a rule, part of the amino acid sequence of X of the precursor of the formula II and results due to the activity of the proteases such as trypsin, trypsin-like enzyme or carboxypeptidase B. The radical R^3 is the amino acid residue which is in position A21 of the A chain of insulin. The radical Y is the amino acid residue which is in position B30 of the B chain of insulin.

Trypsin or trypsin-like enzymes are proteases which cleave amino acid chains at the arginine or lysine residue.
Carboxypeptidase B is an exoprotease which removes basic amino acid residues such as Arg or Lys which are at the carboxy-terminal end of amino acid chains. (Kemmler et al., J. Biol. Chem. 246, pages 6786-6791).

From the precursor mentioned under 1, it is possible, for example, to obtain an insulin or insulin derivative of the formula I having correctly linked cystine bridges, where Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the meaning mentioned under 1 and Z is an argine residue (Arg), a peptide residue Arg-Arg or -OH.

From the precursor mentioned under 2, it is possible, for example, to obtain an insulin or insulin derivative of the formula I having correctly linked cystine bridges, where Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the meaning mentioned under 2 and Z is an arginine residue (Arg), a peptide residue Arg-Arg or Lys-Lys or -OH.

The precursor of the formula II can be formed in microorganisms with the aid of a large number of genetic constructs (EP 0 489 780, EP 0 347 781, EP 0 453 969). The genetic constructs are expressed in microorganisms such as Escherichia coli or Streptomycetes during fermentation. The proteins formed are deposited in the interior of the microorganisms (EP 0 489 780) or secreted into the fermentation solution.

For the process according to the invention, precursors of insulins or of insulin derivatives of the formula II can be employed which, directly after the cell disruption, are still contaminated with a large number of proteins which originate from the fermentation solution and from the microorganisms. The precursors of the formula II, however, can also be employed in prepurified form, for example after precipitation or chromatographic purification.

Example 1 (Comparison Example, Prior Art)

By fermentation of genetically modified Escherichia coli cells (EP 0 489 780), a fusion protein having the following amino acid sequence is prepared.

Proinsulin sequence 1 (SEQ ID NO.: 4):

Ala Thr Thr Ser Thr Gly Asn Ser Ala Arg Phe Val Asn Gln His Leu

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln

Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn

Proinsulin sequence 1 corresponds to the formula II, in this formula

X is C-peptide from human insulin (SEQ ID NO.: 3)

Y is Thr (B30),

R^1 is Phe (B1),

R^2 is a peptide having 10 amino acid residues,

R^3 is Asn (A21) and

A2 - A20 is the amino acid sequence of the A chain of human insulin (amino acid residues 2 to 20) and B2 - B29 is the amino acid sequence of the B chain of human insulin (amino acid residues 2 to 29).

The expressed fusion protein having the proinsulin sequence 1 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are separated off by centrifugation and disrupted by customary high-pressure homogenization. The fusion protein inclusion bodies released are isolated by centrifugation.

20 kg of the isolated fusion protein inclusion bodies (based on dry matter after freeze drying; the proportion of the insulin-containing fusion protein is determined with the aid of HPLC; it is 50%) having the proinsulin sequence 1 are dissolved in 550 l of an 8 M urea solution at pH 10.6. If appropriate, after centrifugation of small amounts of substances causing turbidity, the clear solution is stirred into 9000 l of an aqueous cysteine solution (5 kg of cysteine hydrochloride hydrate) at a pH of 10.6 and a temperature of 4°C.

After completion of the folding reaction after about 24 h, the content of proinsulin sequence I having correctly bonded cystine bridges in the reaction batch is determined with the aid of analytical HPLC as 3.0 kg, corresponding to a conversion of 30%.

The 9500 l of solution is adjusted to a pH of 5.0 using 1N HCl and separated. A pH of 9 is then set by addition of 1N sodium hydroxide solution. 3 g of trypsin are introduced into the solution. 1.25 kg of an insulin having 2 carboxy-terminal arginine residues results according to HPLC measurement.

After cleavage using carboxypeptidase B, human insulin results, which is additionally purified with the aid of chromatographic methods.

Human insulin corresponds to the formula I, in this formula

Y is Thr (B30),

Z is OH,

R^1 is Phe (B1),

R^3 is Asn (A21) and

A2 - A20 is the amino acid sequence of the A chain of human insulin (amino acid residues 2 to 20) and B2-B29 is the amino acid sequence of the B chain of human insulin (amino acid residues 2 to 29)

Human insulin 2 consists of the SEQ ID NO.: 1 and 2, which are connected to one another via correctly bonded cystine bridges.

As described in EP 0 668 292, the solution is concentrated and purified by means of adsorber resin. The eluate, which contains insulin 2, can immediately be further purified on a chromatographic column after dilution with water and pH adjustment.

HPLC analysis

0.5 g of protein is dissolved for 2 min in 40 ml of a solution of 6 M guanidine hydrochloride, 50 mM tris, pH 8.5, 5 mM ethylenediamine

tetraacetate (EDTA), 1% 2-mercaptoethanol, 10 mM dithiothreitol at 95°C and then centrifuged at 14000 g for 20 min. 0.02 ml of the clear supernatant is applied to a high-pressure liquid chromatography column.

Column:	®Nucleogel RP 300-5/46 (Macherey & Nagel, Aachen, Germany)
Gradient:	Buffer A: 0.1% trifluoroacetic acid (TFA)
	Buffer B: 0.09% TFA in acetonitrile
Temperature:	55°C
Total run time:	40 min

The gradient is distinguished by the following amounts of buffer B after the corresponding run times:

10 min 25%, 12 min 60 %, 13 min 90 %, 15 min 100 %.

Flow rate:	1 ml/min
Detection:	215 nm
Retention time of insulin:	approximately 19 min

Example 2 (Process according to the present invention)

By fermentation of genetically *modified Escherichia coli cells* (EP 0 489 780), a fusion protein having the amino acid sequence shown in Example 1 is prepared (proinsulin sequence 1. SEQ ID NO.: 4).

The expressed fusion protein having the proinsulin sequence 1 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are separated off by centrifugation and disrupted by customary high-pressure homogenization. The released fusion protein inclusion bodies are isolated by centrifugation.

5 kg of cysteine hydrochloride hydrate are added to the aqueous fusion protein suspension, which contains 40 kg of fusion protein (determined by freeze drying of an aliquot).

The suspension (the proportion of the insulin-containing fusion protein is determined with the aid of HPLC. It is 50%.) having the proinsulin sequence 1 is dissolved at 40°C in 550 l of an 8 M urea solution at pH 10.2. The clear solution is stirred into 9000 l of water at a pH of 10.6 and a temperature of 15°C.

After completion of the folding reaction after about 5 hours, the content of proinsulin sequence 1 having correctly bonded cystine bridges in the reaction batch is determined with the aid of analytical HPLC as 10.0 kg, corresponding to a conversion of 50%.

The 9500 l of solution is adjusted to a pH of 5.0 using 1N HCl and separated. A pH of 9 is then set by addition of 1N sodium hydroxide solution. 10 g of trypsin are introduced into the solution. 4 kg of an insulin having 2 carboxy-terminal arginine residues result. After cleavage using carboxypeptidase B, human insulin (SEQ ID NO.: 1 and 2 having correctly bonded cystine bridges) results.

The solution is concentrated and purified by means of adsorber resin.

The eluate which contains human insulin can immediately be purified further on a chromatography column after dilution with water and pH adjustment.

Example 3 (Comparison Example, Prior Art)

By fermentation of genetically modified Escherichia coli cells (EP 0 489 780), a fusion protein having the following amino acid sequence is prepared.

Proinsulin sequence 2 (SEQ ID NO.: 5):

Ala Thr Thr Ser Thr Gly Asn Ser Ala Arg Phe Val Asn Gln His Leu

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln

Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Gly

Proinsulin sequence 2 corresponds to the formula II, in this formula

X is C-peptide of human insulin (SEQ ID NO.: 3),

Y is Thr (B30),

R^1 is Phe (B1),

R^2 is a peptide having 10 amino acid residues,

R^3 is Gly (A21) and

A2 - A20 is the amino acid sequence of the A chain of human insulin (amino acid residues 2 to 20) and B2 - B29 is the amino acid sequence of the B chain of human insulin (amino acid residues 2 to 29).

The expressed fusion protein having the proinsulin sequence 2 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are separated off by centrifugation and disrupted by customary high-pressure homogenization. The released fusion protein inclusion bodies are isolated by centrifugation.

20 kg of the isolated fusion protein inclusion bodies (based on dry matter after freeze drying; the proportion of the insulin-containing fusion protein is determined with the aid of HPLC. It is 50%.) having the proinsulin sequence 2 are dissolved at 20°C in 550 l of an 8 M urea solution at pH

10.6. The clear solution is stirred into 9000 l of an aqueous cysteine solution (5 kg of cysteine hydrochloride hydrate) at a pH of 10.6 and a temperature of 4°C.

After completion of the folding reaction after about 24 hours, the content of proinsulin sequence 2 having correctly bonded cystine bridges in the reaction batch is determined with the aid of analytical HPLC as 3.0 kg, corresponding to a conversion of 30%.

The 9500 l of solution is adjusted to a pH of 5.0 using 1N HCl and separated. A pH of 9 is then set by addition of 1N sodium hydroxide solution. 3 g of trypsin are introduced into the solution. 0.98 kg of an insulin derivative having 2 carboxy-terminal arginine residues results according to HPLC measurement. This insulin derivative corresponds to the formula I, where

Y is Thr (B30),

Z is Arg-Arg,

R^1 is Phe (B1),

R^3 is Gly (A21) and

A2 - A20 is the amino acid sequence of the A chain of human insulin (amino acid residues 2 to 20) and B2-B29 is the amino acid sequence of the B chain of human insulin (amino acid residues 2 to 29), and consists of the SEQ ID NO.: 6 and 7, which are connected to one another via correctly bonded cystine bridges.

The solution is concentrated and purified by means of adsorber resin.

The eluate which contains the insulin derivative can immediately be purified further on a chromatography column after dilution with water and pH adjustment.

Example 4 (Process according to the present invention)

By fermentation of genetically modified Escherichia coli cells (EP 0 489 780), the fusion protein having the proinsulin sequence 2 (SEQ ID NO.: 5) is prepared according to Example 3.

The expressed fusion protein having the proinsulin sequence 2 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are separated off by centrifugation and disrupted by customary high-pressure homogenization. The released fusion protein inclusion bodies are isolated by centrifugation.

5 kg of cysteine hydrochloride hydrate are added to the aqueous fusion protein suspension, which contains 40 kg of fusion protein (determined by freeze drying of an aliquot).

The suspension (the proportion of the insulin-containing fusion protein is determined with the aid of HPLC. It is 50%.) having the proinsulin sequence 2 is dissolved at 40°C in 550 l of an 8 M urea solution at pH 10.2. The clear solution is stirred into 9000 l of water at a pH of 10.6 and a temperature of 15°C.
After completion of the folding reaction after about 5 hours, the content of proinsulin sequence I having correctly bonded cystine bridges in the reaction batch is determined with the aid of analytical HPLC as 10.0 kg, corresponding to a conversion of 50%.

The 9500 l of solution is adjusted to a pH of 5.0 using 1 N HCl and separated. A pH of 9 is then set by addition of 1N sodium hydroxide solution. 10 g of trypsin are introduced into the solution. 2.8 kg of the insulin derivative result (HPLC measurement), which consists of the sequences SEQ ID NO.: 6 and 7 which are linked to one another via correctly bonded cystine bridges.

The solution is concentrated by means of adsorber resin and purified.

The eluate which contains the insulin derivative can immediately be purified further on a chromatography column after dilution with water and pH adjustment.

"comprises/comprising" when used in this specification is taken to specify the presence of stated features, integers, steps or components but does not preclude the presence or addition of one or more other features, integers, steps, components or groups thereof.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Hoechst Marion Roussel Deutschland GmbH
(B) STREET: -
(C) CITY: Frankfurt am Main
(E) COUNTRY: Federal Republic of Germany
(F) POSTAL CODE: 65926
(G) TELEPHONE: 069-305-5307
(H) TELEFAX: 069-35-7175
(I) TELEX: 041234-700 hod

(ii) TITLE OF INVENTION: Improved process for obtaining insulin precursors having correctly bonded cystine bridges

(iii) NUMBER OF SEQUENCES: 7

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Asn
20

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..30

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

```
 Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu
Tyr
 1               5                  10                  15

 Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
             20                  25                  30
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..35

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

```
 Arg Arg Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly
Gly
 1               5                  10                  15

 Gly Pro Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser
Leu
             20                  25                  30

 Gln Lys Arg
         35
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..96

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

Ala Thr Thr Ser Thr Gly Asn Ser Ala Arg Phe Val Asn Gln His Leu
1 5 10 15

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
20 25 30

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln
35 40 45

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln
50 55 60

Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln
65 70 75 80

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
85 90 95

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..96

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

Ala Thr Thr Ser Thr Gly Asn Ser Ala Arg Phe Val Asn Gln His Leu
1 5 10 15

Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg
20 25 30

Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg Glu Ala Glu Asp Leu Gln
35 40 45

Val Gly Gln Val Glu Leu Gly Gly Gly Pro Gly Ala Gly Ser Leu Gln
50 55 60

Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys Arg Gly Ile Val Glu Gln
65 70 75 80

Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Gly
85 90 95

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1 5 10 15

Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg
20 25 30

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Escherichia coli

(ix) FEATURES:
(A) NAME/KEY: Protein
(B) LOCATION: 1..21

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Gly
20

Patent claims xx

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A process for obtaining a precursor of insulins or insulin derivatives having correctly bonded cystine bridges in the presence of cysteine or cysteine hydrochloride and of a chaotropic auxiliary, which comprises successively carrying out the following steps:

(a) mixing an aqueous suspension of the precursor of insulins or insulin derivatives with an amount of cysteine or cysteine hydrochloride which results in 1 to 15 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor,

(b) introducing the cysteine- or cysteine hydrochloride-containing suspension of the precursor into a 4 to 9 molar solution of the chaotropic auxiliary at a pH of approximately 8 to approximately 11.5 and a temperature of approximately 15 to approximately 55°C, keeping the mixture obtained at this temperature for approximately 10 to 60 minutes and

(c) introducing the mixture at a pH of approximately 8 to approximately 11.5 and a temperature of approximately 5 to approximately 30°C into an amount of water which results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 1 to 5 mM and of the chaotropic auxiliary to 0.2 to 1.0 M.

2. The process as claimed in claim 1, wherein

in step (a) the amount of cysteine or cysteine hydrochloride corresponds to an amount which results in 1 to 6 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor,

in step (b) the cysteine- or cysteine hydrochloride-containing suspension of the precursor is introduced into a 4 to 9 molar solution of the chaotropic

auxiliary at a pH of 8 to 11 and a temperature of 30 to 45°C and the mixture obtained is kept for 20 to 40 minutes at this temperature and

in step (c) the mixture is introduced at a pH of 8 to 11 and at a temperature of 15 to 20°C into an amount of water which results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 1 to 5 mM and a concentration of the chaotropic auxiliary of 0.2 to 1.0 M.

3. The process as claimed in claim 1 or 2, wherein the chaotropic auxiliary is guanidine or guanidine hydrochloride.

4. The process as claimed in claim 1 or 2, wherein the chaotropic auxiliary is urea.

5. The process as claimed in one or more of claims 1 to 4, wherein the concentration of the chaotropic auxiliary in step (b) is 7.0 to 9 M.

6. The process as claimed in one or more of claims 1 to 5, wherein the temperature in step (b) is 40°C.

7. The process as claimed in one or more of claims 1 to 6, wherein the pH in step (b) is 10 to 11.

8. The process as claimed in one or more of claims 1 to 7, wherein the pH in step (c) is 10 to 11.

9. The process as claimed in one or more of claims 1 to 8, wherein in step (c) the amount of water results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to 2.5 to 3 mM and a concentration of the chaotropic auxiliary to 0.5 M.

10. The process as claimed in one or more of claims 2 to 9, wherein the concentration of the chaotropic auxiliary in step (b) is approximately 8 M, the temperature in step (b) is approximately 40°C, the pH in step (b) is approximately 10.6, the pH in step (c) is approximately 10.6 and in step (c) the amount of water results in a dilution of the concentration of the cysteine or of the cysteine hydrochloride in the mixture to approximately 2.5 to 3 mM and a concentration of the chaotropic auxiliary of 0.5 M.

11. The process as claimed in one or more of claims 1 to 10, wherein the precursor of the insulins or of the insulin derivatives has the sequence according to the formula II

$$R^2\text{-}R^1\text{-(B2-B29)-Y-X-Gly-(A2-A20)-}R^3 \qquad \text{(II)},$$

in which

R^2 is a) a hydrogen atom,

b) an amino acid residue from the group consisting of lysine (Lys) and arginine (Arg) or

c) a peptide having 2 to 45 amino acid residues, comprising the amino acid residue lysine (Lys) or arginine (Arg) at the carboxyl end of the peptide

R^1 is a phenylalanine residue (Phe) or a covalent bond,

(B2-B29) are the amino acid residues in the positions B2 to B29 of the B chain of human insulin, animal insulin or an insulin derivative which is optionally varied in one or more of these positions,

Y is a genetically encodable amino acid residue,

X is a) an amino acid residue from the group consisting of lysine (Lys) and arginine (Arg) or

b) a peptide having 2 to 35 amino acid residues,

comprising the amino acid residue lysine (Lys) or arginine (Arg) at the N-terminal end and at the carboxyl end of the peptide, or

c) a peptide having 2 to 35 genetically encodable amino acids, comprising 1 to 5 histidine residues,

(A2-A20) are the amino acid residues in the positions A2 to A20 of the B chain of human insulin, animal insulin or an insulin derivative which is optionally varied in one or more of these positions and

R^3 is a genetically encodable amino acid residue.

12. The process as claimed in claim 11, wherein in formula II

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 25 amino acid residues, comprising the amino acid residue arginine (Arg) at the carboxyl end of the peptide,

R^1 is a phenylalanine residue (Phe),

(B2-B29) are the amino acid residues in the positions B2 to B29 of the B chain of human insulin,

Y is an amino acid residue from the group consisting of alanine (Ala), threonine (Thr) and serine (Ser),

X is the amino acid residue arginine (Arg) or a peptide having the amino acid sequence of the C chain of human insulin,

(A2-A20) are the amino acid residues in the positions A2 to A20 of the B chain of human insulin and

R^3 is an amino acid residue from the group consisting of asparagine (Asn), serine (Ser) and glycine (Gly).

13. The process as claimed in claim 11, wherein in formula II

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 15 amino acid residues, at whose carboxyl end is found an arginine residue (Arg),

R^1 is a phenylalanine residue (Phe),

(B2-B29) are the amino acid residues in the positions B2 to B29 of the B chain of human insulin,

Y is a threonine residue (Thr),

X is the amino acid residue arginine (Arg) or a peptide having 2 to 35 amino acid residues, where at the beginning and at the end of the peptide there are two basic amino acid residues, in particular arginine (Arg) and/or lysine (Lys),

(A2-A20) are the amino acid residues in the positions A2 to A20 of the B chain of human insulin and

R^3 is the amino acid residue asparagine (Asn) or glycine (Gly).

14. A process for obtaining an insulin or an insulin derivative thereof having correctly bonded cysteine bridges in the presence of cysteine or cysteine hydrochloride and of a chaotropic auxiliary, which comprises the following steps:

(a) mixing an aqueous suspension of a precursor of insulins or insulin derivative with an amount of cysteine or cysteine hydrochloride which results in approximately 1 to 15 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor;

(b) introducing the cysteine or cysteine hydrochloride-containing suspension of the precursor into a 4 to 9 molar solution of the chaotropic auxiliary at a pH of approximately 8 to approximately 11.5 and a temperature of approximately 15°C to approximately 55°C, keeping the mixture obtained at this temperature for approximately 10 to 60 minutes;

AUSTRALIAN PATENT OFFICE

(c) diluting the cysteine or the cysteine hydrochloride concentration to approximately 1 to 5 mM and diluting the chaotropic auxiliary concentration to approximately 0.2 to 1.0 M in the mixture at a pH of approximately 8 to 11.5 and a temperature of approximately 5°C to 30°C;

(d) cleaving with trypsin or a trypsin-like enzyme the precursor of the insulin or insulin derivative obtained by this process; and

(e) purifying the insulin or insulin derivative having correctly bonded cysteine bridges, wherein the precursor of insulin or insulin derivative of step (a) is obtained by expressing a fusion protein in a microorganism.

15. The process according to claim 14, wherein

in step (a) the amount of cysteine or cysteine hydrochloride corresponds to an amount which results in 1 to 6 SH residues of the cysteine or cysteine hydrochloride per cysteine residue of the precursor,

in step (b) the cysteine or cysteine hydrochloride-containing suspension of the precursor is added into an approximately 4 to 9 molar solution of the chaotropic auxiliary at a pH of approximately 8 to 11 and a temperature of approximately 30°C to 45°C and the mixture obtained is kept for approximately 20 to 40 minutes at this temperature; and

wherein the pH in step (c) is approximately 8 to 11 and the temperature is approximately 15°C to 20°C.

AUSTRALIAN PATENT OFFICE

16. The process according to claim 14, further comprising cleavage of the precursor of insulin or insulin derivative obtained in step (d) with carboxypeptidase B or a carboxypeptidase-like enzyme.

DATED this 10th day of January 2002

HOECHST MARION ROUSSEL DEUTSCHLAND GMBH

WATERMARK PATENT & TRADEMARK ATTORNEYS
290 BURWOOD ROAD
HAWTHORN VICTORIA 3122
AUSTRALIA

P14769AU00 KJS:AMT:SLB

AUSTRALIAN PATENT OFFICE