This Sheet Has Been Replaced



960825|-8
|P03

H1-241

HOECHST AKTIENGESELLSCHAFT HOE 94/F 042 Dr. TH/St

Process for obtaining insulin having correctly linked cystine bridges

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

A 6 - A 11
A 7 - B 7
A 20 - B 19

The letters A and B stand for the particular insulin amino acid chain, and the number for the position of the amino acid residue, which is counted from the amino to the carboxyl end of the particular amino acid chain. Disulfide bridges can also be formed between two human insulin molecules so that infinitely many different disulfide bridges can easily be formed.

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 linked with one another via a C peptide having 35 amino acid residues. The formation of the disulfide bridges occurring in human insulin takes place via an interme-diate, 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

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

obtaining proinsulin having correctly linked cysteine bridges is also 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 linked cystine bridges" is understood as meaning the disulfide bridges which occur in biologically active insulin from mammals.

Genetic engineering processes enable 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 Escherichia coli cells modified by genetic engineering do not have any correctly linked 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 bromide cleavage of the fusion protein - isolation of the cleavage product containing the proinsulin sequence - protection of the cysteine residues of proinsulin by S-sulfonate groups - formation of the correctly linked cystine bridges - sulfitolysis - desalting of the proinsulin - chromatographic purification of the proinsulin containing correctly linked 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.

The 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 multi-stage process path, serious losses have to be expected. From the stage of the isolated fusion protein, an up to 40 % loss of proinsulin must be expected through cyanogen bromide cleavage, sulfitolysis and purification of the proinsulin (EP 0 055 945). Similarly high losses

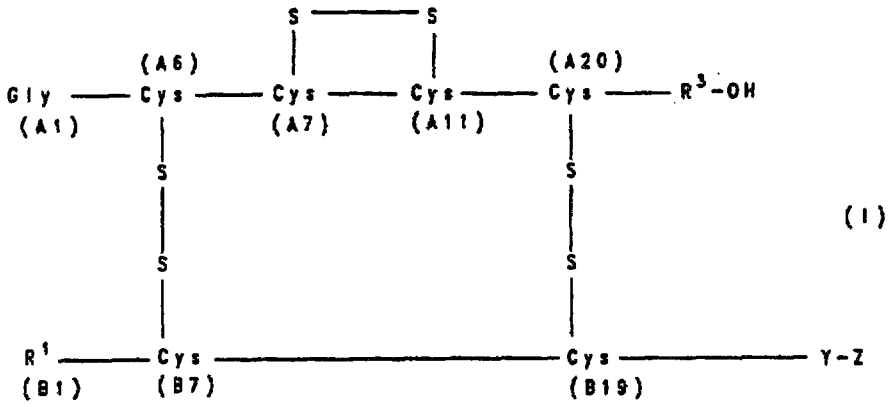
THIS SHEET HAS BEE

can occur in the course of the following purification steps up to the final product.

Yield increases in the preparation of human insulin or insulin derivatives by genetic engineering can be achieved if the number of process steps necessary can be substantially decreased.

The object of the present invention was to develop a process for obtaining insulin having correctly linked cystine bridges in the insulin amino acid chain, in which fewer process steps are necessary and, all in all, lower purification losses occur.

A process for obtaining insulin of the formula I



(I)

has now been found,

This Sheet Has Been Replaced

which comprises

A) reacting a protein of the formula II

R^2-R^1-B2-B29-Y-X-Gly-A2-A20-R^3 (II)

with an amount of a mercaptan which yields 2 to 10 -SH radicals of the mercaptan per cysteine residue of the protein of the formula II, in the presence of at least one chaotropic auxiliary in an aqueous medium at a pH of 10 to 11 and a concentration of the protein of the formula II of 0.05 to 0.3 g per liter of aqueous medium and

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

B) reacting the resulting proinsulin having correctly linked cystine bridges with trypsin, a trypsin-like enzyme and optionally additionally with carboxypeptidase B or a mixture of the enzymes mentioned to give the insulin of the formula I having correctly linked cystine bridges,

C) treating the reaction product thus obtained with 3 to 50 g of a hydrophobic adsorber resin per liter of aqueous medium at a pH of 4 to 7,

D) isolating the adsorber resin containing adsorbed insulin of the formula I and

E) desorbing the insulin of the formula I from the adsorber resin;

in this case, in formulae I and II

X is
a) an amino acid residue from the group consisting of Lys and Arg or
b) a peptide having 2 to 35 amino acid residues, containing the amino acid residue Arg or Lys at the N-terminal and carboxyl end of the peptide,

Y is a genetically encodable amino acid residue,

Z is
a) an amino acid residue from the group consisting of Arg and Lys,
b) a peptide having 2 or 3 amino acid residues, containing the amino acid residue Arg or Lys at the carboxyl end of the peptide or
c) OH,

R^1 is a phenylalanine residue or a covalent bond,

R^2 is
a) a hydrogen atom,
b) an amino acid residue from the group consisting of Lys and Arg or
c) a peptide having 2 to 45 amino acid residues, containing the amino acid residue Lys or Arg at the carboxyl end of the peptide,

R^3 is a genetically encodable amino acid residue and

the residues A2 - A20 correspond to the amino acid sequence of the A chain of human insulin, animal insulin or an insulin derivative and

the residues B2 - B29 correspond to the amino acid

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

sequence of the B chain of human insulin, animal insulin or an insulin derivative.

A protein of the formula II is preferably employed; in this case

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

Y is an amino acid residue from the group consisting of Ala, Thr and Ser,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 25 amino acid residues, containing Arg at the carboxyl end of the peptide,

R^3 is an amino acid residue from the group consisting of Asn, Ser and Gly,

and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the B and A chains of human insulin and

an insulin of the formula I having correctly linked cystine bridges results, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg; the peptide residue Arg-Arg or OH.

A protein of the formula II is particularly preferably employed; in this case

X is the amino acid residue Arg or a peptide having 2 to 35 amino acid residues, two basic amino acid residues, in particular Arg and/or Lys, being at the start and end of the peptide,

Y is the amino acid residue Thr,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 15 amino acid residues, at whose carboxyl end is the amino acid residue Arg,

R^3 is the amino acid residue Gly or Asn,

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin and

an insulin of the formula I having correctly linked cystine bridges is formed, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or Lys-Lys or OH.

Surprisingly, it has been found that the protein of the formula II does not have to be completely reduced in a reaction step customarily preceding the "refolding" (R. Jaenicke, R. Rudolph, 1989, Folding Proteins, in Protein Structure, a practical Approach; Ed. Creighton, T.E., pages 191-224, JRL Press Oxford; EP 0 219 874) and that, in the procedure described, in spite of a high proportion of foreign protein (50 - 90 %) folding yields are achieved which are at a level comparable with the folding of appropriately purified proinsulin containing -SH protective groups.

It has further surprisingly been found that the enzymatic reaction of correctly folded proinsulin to give the insulin of the formula I having correctly linked cystine bridges is also possible in high yield in the presence of 50 to 90 % of foreign protein and the insulin of the formula I binds with good selectivity to a hydrophobic adsorber resin.

Compared with the previously known methods, the process according to the invention enables substantially more rapid preparation with higher yields. The reason for this is the avoidance of the sulfitolysis and optionally cyanogen bromide cleavage process steps and the previously unknown possibility of converting proinsulin directly from its denatured state in the presence of limited amounts of chaotropic salts in high yields into a proinsulin having correctly bonded cystine bridges and then of converting it enzymatically into the insulin of

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

the formula I without further purification. It is thereby possible to separate the resulting insulin of the formula I from the folding solution by adsorption. The insulins of the formula I can thus be obtained from the adsorbent without intermediate isolation or purification after desorption. The process according to the invention leads to a substantially shortened preparation, which because of shorter standing times and the avoidance of losses makes possible yield increases of 25-100 % compared with the known processes.

The amino acid sequence of peptides and proteins is designated from the N-terminal end of the amino acid chain. The information given in parentheses in formula I, e.g. A6, A20, B1, B7 or B19, corresponds to the position of amino acid residues in the A or B chains of the insulin.

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 stand for the term "genetically encodable amino acid radical".

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 hens. The term radicals A2 - A20 and B2 - B29 of insulin derivatives stands for the corresponding amino acid sequences of human insulin which are formed by the replacement of amino acids by other genetically encodable amino acids.

The residue Z is always part of the amino acid sequence of X and is formed by the activity of the proteases such as trypsin, trypsin-like enzymes or carboxypeptidase B. The residue R^3 is the amino acid residue which is in position A21 of the A chain of insulin. The residue Y is the amino acid residue which is in position B30 of the B chain of insulin.

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

Trypsin or trypsin-like enzymes are proteases which cleave amino acid chains at the arginine or lysine residue.

Carboxypeptidase B is an exoprotease which splits off 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).

The A chain of human insulin 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.

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.

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

The term hydrophobic adsorber resin stands for nonionic, hydrophobic, crosslinked polymers and/or copolymers, for example polyacrylates or polystyrene or copolymers of styrene and divinylbenzene, in particular polymeric adsorber resins having a large surface area and many large pores, e.g. commercial preparations of the companies Rohm and Haas or Mitsubishi Chemical Industries Ltd. such as XAD16, XAD1600 or HP20.

The term mercaptan is understood as meaning compounds which contain at least one -SH group. Water-soluble mercaptans are preferred. Examples of these compounds are

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

dithiothreitol, dithioerythritol, 2-mercaptoethanol, cysteine, glutathione, methyl thioglycolate, 3-mercapto-1,2-propanediol and 3-mercaptopropionic acid.

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

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

The procedure in process step A) is as follows:
The proteins are dissolved in a chaotropic auxiliary or in mixtures of various chaotropic auxiliaries. Preferably, urea or guanidine hydrochloride is used in a concentration of 6 to 9 M, preferably 8 M, based on water as a solvent. The pH of the reaction mixture is 8.5 to 10.8. Examples of buffers which can be used are phosphate, tri(hydroxymethyl)aminomethane (tris), borate or glycine buffer. The concentration of the buffer substances in the aqueous medium is up to 0.5 M, preferably from 0.005 M to 0.5 M, particularly preferably from 0.05 to 0.1 M. The protein mixture is then mixed with an aqueous mercaptan solution. By this means, the following concentrations are established in the reaction solution (based on water):

The concentration of the protein of the formula II is 0.05 to 0.6 g/l, preferably 0.1 to 0.3 g/l. The amount of

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

the mercaptan is 2 to 10 -SH residues of the mercaptan per cysteine residue of the protein of the formula II, preferably 6 to 9. The pH of the solution is 10 to 11, preferably 10.8. The abovementioned buffer components are used. The exact pH is established by addition of sodium hydroxide solution. The concentration of the buffer components is 0.005 to 0.5 M, preferably 0.05 to 0.1 M.

The concentration of the chaotropic auxiliary in the reaction mixture containing the mercaptan is less than 1 M, preferably 0.1 to 0.8 M, in particular 0.3 to 0.6 M. The mercaptan employed is preferably cysteine or 2-mercaptoethanol, alone or as a mixture.

The temperature during the folding in process step A) is 0°C to 50°C, preferably 2°C to 30°C, in particular 4°C to 10°C. The reaction time is 3 to 12 hours, preferably 4 to 8 hours, in particular 5 hours.

The result of process step A) is a proinsulin which contains correctly linked cystine bridges.

In process step B), the reaction solution from process step A) is adjusted to a pH of 6.5 to 9, preferably 8 to 9. Trypsin or a trypsin-like enzyme is added thereto in an amount from 1 to 3 mg based on 3 g of proinsulin having correctly linked cystine bridges. Optionally, carboxypeptidase B is added thereto in an amount of 3 to 8 units based on 1 g of insulin of the formula I. A mixture of trypsin and carboxypeptidase B can also be added. The solution obtained is then stirred at a temperature of 4°C to 15°C for from 4 to 16 hours. The result of process step B) is an insulin of the formula I which contains correctly linked cystine bridges.

In process step C), the reaction solution from process step B) is adjusted to a pH of 2.5 to 4.5 with an acid such as hydrochloric acid or sulfuric acid. If turbidity of the solution should occur, this is optionally removed

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

by filtration or centrifugation. The remaining solution is treated with a hydrophobic adsorber resin. Resins of the XAD or HP20 type have proven suitable. The amount of the resin is 3 to 50 g per liter of reaction solution, preferably 20 to 40 g/l.

The suspension obtained can be treated with a salt such as NaCl up to a concentration of 0.1 M, and is then stirred for 3 to 4 hours. The adsorption of the insulin of the formula I on the resin is checked by sampling and high-pressure liquid chromatographic analysis (HPLC analysis). As soon as insulin is no longer detectable in the solution, the adsorber resin is separated off from the aqueous reaction solution (process step D). This is carried out, for example, by filtration or centrifugation according to known methods. The resin containing adsorbed insulin of the formula I is washed with a purely aqueous or buffer-containing aqueous solution. In particular, it is washed until the conductivity in the wash solution is less than 0.4 mS/cm.

The desorption (process step E) of the insulin of the formula I is carried out by known methods, which depend on the hydrophobic adsorber resin used. The desorption is carried out in the case of XAD or HP20 adsorber resins, for example, by means of an aqueous solution which contains 20 to 65 % of a water-miscible, nonionic organic solvent, e.g. a (C_1-C_6)-alkanol. Preferably, 35 % of isopropanol or n-propanol is used in the solvent water. The desorption of the insulin of the formula I is carried out, for example, by washing with the isopropanol solution. The washing can be carried out on a stirred pressure filter by mixing with the isopropanol solution or by chromatography in a column. The volume ratio of resin to isopropanol solution is 1:1 to 1:10, preferably 1:1.1 to 1:2. The washing steps can be repeated one to five times, preferably twice. The combined isopropanol fractions can be used directly or after dilution with water for further chromatographic purifications.

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

In the following examples the process according to the invention is described in detail. Percentage data relate to the weight, if not stated otherwise.

Example 1

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

Proinsulin sequence 1 (Seq. Id No. 4):

Met 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 1 corresponds to the formula II, in this case

X is a C peptide from human insulin,
Y is Thr (B30),
R^1 is Phe (B1),
R^2 is a peptide having 11 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 1 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are removed by centrifugation and disrupted by customary high-pressure homogenization. The fusion protein inclusion bodies released are isolated by centrifugation.

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

2600 g of the isolated fusion protein inclusion bodies (based on dry matter after freeze drying) having the proinsulin sequence 1 are dissolved in 200 l of an 8 M guanidine hydrochloride solution at pH 10.8. Optionally, after centrifugation of small amounts of suspended matter, the clear solution is stirred into 1800 l of an aqueous cysteine solution (250 g of cysteine hydrochloride hydrate) at a pH of 10.8 and a temperature of 4°C. The proportion of the insulin-containing fusion protein is determined with the aid of an SDS electrophoresis scan (U.K. Lämmli, Nature, Volume 227, pages 680-685, 1970). It is 45 %. After conclusion of the folding reaction, the content of proinsulin sequence I having correctly linked cystine bridges in the reaction mixture is determined as 760 g with the aid of analytical HPLC.

The 2000 l of solution is adjusted to a pH of 8.5 using 6N HCl. 500 mg of trypsin is then added. 305 g of insulin 2 are formed according to HPLC measurement.

Insulin 2 corresponds to the formula I, in this case

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).

Insulin 2 consists of the Seq Id No. 6 and 9, which are linked to one another via correctly bonded cystine bridges.

The solution is adjusted to a pH of 3.5 using 6N HCl and the conductivity value is adjusted to about 10 mS/cm with the aid of saturated NaCl solution. A slight turbidity is removed by centrifugation. 75 kg of HP20 (Mitsubishi Chemical Industries Ltd., Düsseldorf, Germany) are added

THIS SHEET HAS BEEN REPLACED

This Sheet Has Been Replaced

to the clear solution. The suspension is slowly stirred at 4°C for about 16 hours until insulin 2 is no longer to be detected in the supernatant. The resin is removed by filtration through a suction filter and washed with water until the conductivity in the wash solution is less than 0.4 mS/cm. The product is desorbed by suspension of the resin in 75 l of 50 % strength aqueous isopropanol solution. Filtration and incubation with the isopropanol solution is repeated twice. After 60 minutes, the resin is filtered and then incubated several times in a 35 % strength isopropanol solution and filtered. The yield of insulin 2 is 288 g.

The eluates which contain insulin 2 can be further purified immediately on a chromatography column after dilution with water and pH adjustment.

HPLC analysis

0.5 g of protein are dissolved in 40 ml of a solution of 6 M guanidine hydrochloride, 50 mM tris, pH 8.5, 5 mM ethylenediaminetetraacetate (EDTA), 1 % 2-mercaptoethanol and 10 mM dithiothreitol at 95°C for 2 min 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 running time:	40 min.

The gradient is characterized by the following amounts of buffer B after the corresponding running times:
10 min 25 %, 12 min 60 %, 13 min 90 %, 15 min 100 %.

Flow rate:	1 ml/min
Detection:	215 nm

THIS SHEET HAS BEEN REPLACED

Retention time
of insulin: about 19 min

This Sheet Has Been Replaced

Example 2

A fusion protein having the following amino acid sequence is prepared by fermentation of a genetically modified E. coli (EP 0 347 781):
Fusion protein 3 (Seq Id No. 5)

Met Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr Gln Leu
Gln Leu Glu His Leu Leu Leu Asp Leu Gln Met Ile Leu
Asn Gly Ile Asn Asn Tyr Lys Asn Pro Lys Leu Thr Arg
Met Ile Glu Gly 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 Gly Ile Val
Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
Glu Asn Tyr Cys Asn

This fusion protein contains the amino acid sequence of proinsulin 3. The expressed fusion protein 3 collects in the E. coli cells and forms inclusion bodies. The disruption of the cells is carried out as in Example 1. The fusion proteins 3 thus obtained are subjected to cyanogen bromide cleavage, and proinsulin 3 is formed (Seq Id No. 8).
Proinsulin 3 corresponds to the formula II, in this case
X is Arg
Y is Thr (B30),
R^1 is Phe (B1),
R^2 is a peptide having 4 amino acid residues,
R^3 is Asn (A21) and
A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin.

After cyanogen bromide cleavage, the content of proinsulin 3 is determined quantitatively with the aid of SDS electrophoresis in a freeze-dried sample containing 9.2 % insulin-containing protein.

THIS SHEET HAS BEEN REPLACED

2000 g of proinsulin 3 are incubated with cysteine hydrochloride hydrate as in Example 1. A content of 1260 g of proinsulin 3 having correctly linked cystine bridges is then determined in the total reaction mixture by means of analytical HPLC. The product is treated with 500 mg of trypsin as in Example 1.

Insulin 4 is formed. Insulin 4 corresponds to the formula I, in this case

Y is Thr (B30),

Z is Arg,

R^1 is Phe (B1)

R^3 is Asn (A21) and

A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin.

Insulin 4 consists of the Seq Id No. 1 and 7, which are linked to one another by correctly bonded cystine bridges.

The resulting solution (2 m^3) contains 800 g of the insulin derivative and is treated with 75 kg of HP20 as in Example 1. After adsorption and desorption, the combined eluate contains 703 g of insulin 4 (88 % yield).

This Sheet Has Been Replaced



THIS SHEET HAS BEEN REPLACED

(1) GENERAL INFORMATION:

(i) APPLICANT

(A) NAME: Hoechst Aktiengesellschaft

(B) STREET: -

(C) CITY: Frankfurt am Main

(D) COUNTRY: Federal Republic of Germany

(E) POSTAL CODE: 65926

(F) TELEPHONE: 069-305-6047

(G) TELEFAX: 069-35-7175

(H) TELEX: 041234-700 hod

(ii) TITLE OF INVENTION: Process for obtaining insulin having correctly linked cystine bridges

(iii) NUMBER OF SEQUENCES: 9

(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 (EPA)

(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

This Sheet Has Been Replaced



(ii) MOLECULE TYPE: Protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Escherichia coli

(ix) FEATURES:

(A) NAME/KEY: Protein

(B) LOCATION: 1..21

THIS SHEET HAS BEEN REPLACED

(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

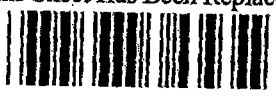
(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

This Sheet Has Been Replaced



(ii) MOLECULE TYPE: Protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Escherichia coli

THIS SHEET HAS BEEN REPLACED

(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: 97 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..97

This Sheet Has Been Replaced



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

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

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

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

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

THIS SHEET HAS BEEN REPLACED

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

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

Gly
```

(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:

```
Met Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr Gln Leu Gln Leu Glu
1               5                   10                  15

His Leu Leu Leu Asp Leu Gln Met Ile Leu Asn Gly Ile Asn Asn Tyr
            20                  25                  30

Lys Asn Pro Lys Leu Thr Arg Met Ile Glu Gly Arg Phe Val Asn Gln
        35                  40                  45

His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly
    50                  55                  60

Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr 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. 6:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

This Sheet Has Been Replaced



THIS SHEET HAS BEEN REPLACED

(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: 31 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..31

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

```
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
            20                  25                  30
```

This Sheet Has Been Replaced



THIS SHEET HAS BEEN REPLACED

(2) INFORMATION FOR SEQ ID NO. 8:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 56 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..56

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 8:

```
Ile Glu Gly Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val
1               5                  10                  15

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

Lys Thr Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu
        35                  40                  45

Tyr Gln Leu Glu Asn Tyr Cys Asn
    50                  55
```

(2) INFORMATION FOR SEQ ID NO. 9:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

This Sheet Has Been Replaced



(ii) MOLECULE TYPE: Protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Escherichia coli

(ix) FEATURES:

(A) NAME/KEY: Protein

(B) LOCATION: 1..21

THIS SHEET HAS BEEN REPLACED

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 9:

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

This Sheet Has Been Replaced

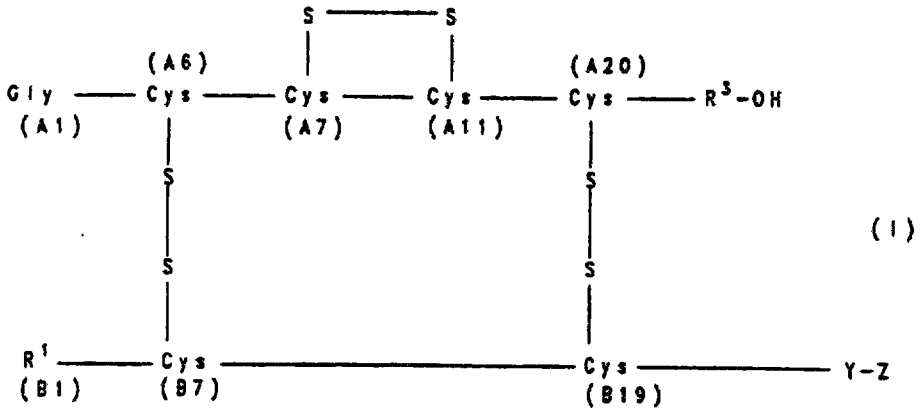


THIS SHEET HAS BEEN REPLACED

Patent claims

HOE 94/F 042

1. A process for obtaining insulin of the formula I



This Sheet Has Been Replaced

which comprises

A) reacting a protein of the formula II

R^2-R^1-B2-B29-Y-X-Gly-A2-A20-R^3 (II)

with an amount of a mercaptan which yields 2 to 10 -SH radicals of the mercaptan per cysteine residue of the protein of the formula II, in the presence of at least one chaotropic auxiliary in an aqueous medium at a pH of 10 to 11 and a concentration of the protein of the formula II of 0.05 to 0.3 g per liter of aqueous medium and

B) reacting the resulting proinsulin having correctly linked cystine bridges with trypsin, a trypsin-like enzyme and optionally additionally with carboxypeptidase B or a mixture of the enzymes mentioned to give the insulin of the formula I having correctly linked cystine bridges,

C) treating the reaction product thus obtained with 3 to 50 g of a hydrophobic adsorber resin per liter of aqueous

THIS SHEET HAS BEEN REPLACED

medium at a pH of 4 to 7,

D) isolating the adsorber resin containing adsorbed insulin of the formula I and

E) desorbing the insulin of the formula I from the adsorber resin;

in this case, in formulae I and II

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

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

Y is a genetically encodable amino acid residue,

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

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

c) OH,

R^1 is a phenylalanine residue or a covalent bond,

R^2 is a) a hydrogen atom,

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

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

R^3 is a genetically encodable amino acid residue and

the residues A2 - A20 correspond to the amino acid sequence of the A chain of human insulin, animal insulin or an insulin derivative and

the residues B2 - B29 correspond to the amino acid sequence of the B chain of human insulin, animal insulin or an insulin derivative.

This Sheet Has Been Replaced

THIS SHEET HAS BEEN REPLACED

* C L M S *

R^3 is the amino acid residue Gly or Asn, and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin and an insulin of the formula I having correctly linked cystine bridges is formed, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or Lys-Lys or OH.

4. The process as claimed in one or more of claims 1 to 3, wherein cysteine or cysteine hydrochloride hydrate is employed as the mercaptan.

5. The process as claimed in one or more of claims 1 to 4, wherein a crosslinked polystyrene, polyacrylate or copolymer of polystyrene and divinylbenzene is employed as the hydrophobic adsorber resin.

6. The process as claimed in one or more of claims 1 to 5, wherein the insulin of the formula I is desorbed from the adsorber resin using isopropanol or n-propanol.

7. The process as claimed in one or more of claims 1 to 6, wherein 0.01 to 0.1 M salt is added in the process step C).

8. An insulin of the formula I, obtained by a process as claimed in one or more of claims 1 to 7.

This Sheet Has Been Replaced

THIS SHEET HAS BEEN REPLACED

Process for obtaining insulin having correctly linked cystine bridges

A process for obtaining insulin having correctly linked cystine bridges from a corresponding proinsulin amino acid chain in the presence of mercaptan, chaotropic auxiliaries and at a pH of 10 to 11 is described, as well as the direct conversion of the proinsulin to insulin with the aid of enzymes and direct isolation of the insulin from the reaction mixture with the aid of hydrophobic adsorber resins.

This Sheet Has Been Replaced



THIS SHEET HAS BEEN REPLACED

REPLACEMENT SHEET
25 FEB 1997

Process for obtaining insulin having correctly linked cystine bridges

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

A 6 - A 11

A 7 - B 7

A 20 - B 19

The letters A and B stand for the particular insulin amino acid chain, and the number for the position of the amino acid residue, which is counted from the amino to the carboxyl end of the particular amino acid chain. Disulfide bridges can also be formed between two human insulin molecules so that infinitely many different disulfide bridges can easily be formed.

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 linked with one another via a C peptide having 35 amino acid residues. The formation of the disulfide bridges occurring in human insulin takes place via an interme-diate, 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

REPLACEMENT SHEET
25 FEB 1997

obtaining proinsulin having correctly linked cysteine bridges is also 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 linked cystine bridges" is understood as meaning the disulfide bridges which occur in biologically active insulin from mammals.

Genetic engineering processes enable 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 Escherichia coli cells modified by genetic engineering do not have any correctly linked 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 bromide cleavage of the fusion protein - isolation of the cleavage product containing the proinsulin sequence - protection of the cysteine residues of proinsulin by S-sulfonate groups - formation of the correctly linked cystine bridges - sulfitolysis - desalting of the proinsulin - chromatographic purification of the proinsulin containing correctly linked 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.

The 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 multi-stage process path, serious losses have to be expected. From the stage of the isolated fusion protein, an up to 40 % loss of proinsulin must be expected through cyanogen bromide cleavage, sulfitolysis and purification of the proinsulin (EP 0 055 945). Similarly high losses

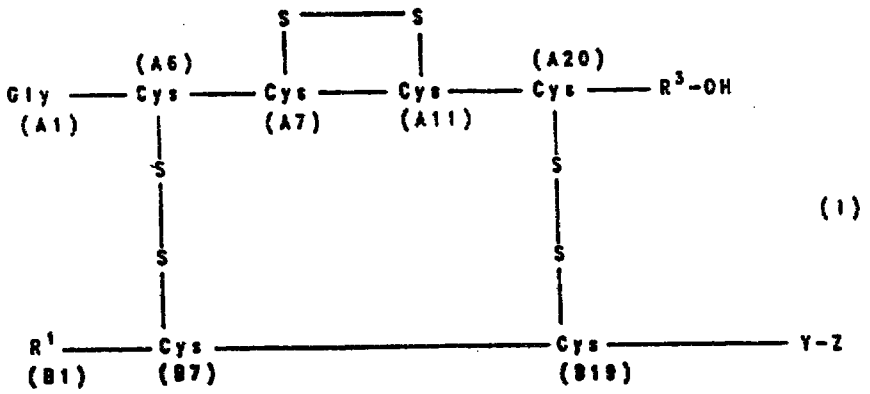
REPLACEMENT SHEET
25 FEB 1997

can occur in the course of the following purification steps up to the final product.

Yield increases in the preparation of human insulin or insulin derivatives by genetic engineering can be achieved if the number of process steps necessary can be substantially decreased.

The object of the present invention was to develop a process for obtaining insulin having correctly linked cystine bridges in the insulin amino acid chain, in which fewer process steps are necessary and, all in all, lower purification losses occur.

A process for obtaining insulin of the formula I



(I)

has now been found,

which comprises

A) reacting a protein of the formula II

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

with an amount of a mercaptan which yields 2 to 10 -SH radicals of the mercaptan per cysteine residue of the protein of the formula II, in the presence of at least one chaotropic auxiliary in an aqueous medium at a pH of 10 to 11 and a concentration of the protein of the formula II of 0.05 to 0.3 g per liter of aqueous medium and

REPLACEMENT SHEET
25 FEB 1997

B) reacting the resulting proinsulin having correctly linked cystine bridges with trypsin, a trypsin-like enzyme and optionally additionally with carboxypeptidase B or a mixture of the enzymes mentioned to give the insulin of the formula I having correctly linked cystine bridges,

C) treating the reaction product thus obtained with 3 to 50 g of a hydrophobic adsorber resin per liter of aqueous medium at a pH of 4 to 7,

D) isolating the adsorber resin containing adsorbed insulin of the formula I and

E) desorbing the insulin of the formula I from the adsorber resin;

in this case, in formulae I and II

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

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

Y is a genetically encodable amino acid residue,

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

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

c) OH,

R^1 is a phenylalanine residue or a covalent bond,

R^2 is a) a hydrogen atom,

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

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

R^3 is a genetically encodable amino acid residue and

the residues A2 - A20 correspond to the amino acid sequence of the A chain of human insulin, animal insulin or an insulin derivative and

the residues B2 - B29 correspond to the amino acid

REPLACEMENT SHEET
25 FEB 1997

sequence of the B chain of human insulin, animal insulin or an insulin derivative.

A protein of the formula II is preferably employed; in this case

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

Y is an amino acid residue from the group consisting of Ala, Thr and Ser,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 25 amino acid residues, containing Arg at the carboxyl end of the peptide,

R^3 is an amino acid residue from the group consisting of Asn, Ser and Gly,

and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the B and A chains of human insulin and

an insulin of the formula I having correctly linked cystine bridges results, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or OH.

A protein of the formula II is particularly preferably employed; in this case

X is the amino acid residue Arg or a peptide having 2 to 35 amino acid residues, two basic amino acid residues, in particular Arg and/or Lys, being at the start and end of the peptide,

Y is the amino acid residue Thr,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 15 amino acid residues, at whose carboxyl end is the amino acid residue Arg,

R^3 is the amino acid residue Gly or Asn,

REPLACEMENT SHEET
25 FEB 1997

and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin and

an insulin of the formula I having correctly linked cystine bridges is formed, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or Lys-Lys or OH.

Surprisingly, it has been found that the protein of the formula II does not have to be completely reduced in a reaction step customarily preceding the "refolding" (R. Jaenicke, R. Rudolph, 1989, Folding Proteins, in Protein Structure, a practical Approach; Ed. Creighton, T.E., pages 191-224, JRL Press Oxford; EP 0 219 874) and that, in the procedure described, in spite of a high proportion of foreign protein (50 - 90 %) folding yields are achieved which are at a level comparable with the folding of appropriately purified proinsulin containing -SH protective groups.

It has further surprisingly been found that the enzymatic reaction of correctly folded proinsulin to give the insulin of the formula I having correctly linked cystine bridges is also possible in high yield in the presence of 50 to 90 % of foreign protein and the insulin of the formula I binds with good selectivity to a hydrophobic adsorber resin.

Compared with the previously known methods, the process according to the invention enables substantially more rapid preparation with higher yields. The reason for this is the avoidance of the sulfitolysis and optionally cyanogen bromide cleavage process steps and the previously unknown possibility of converting proinsulin directly from its denatured state in the presence of limited amounts of chaotropic salts in high yields into a proinsulin having correctly bonded cystine bridges and then of converting it enzymatically into the insulin of

REPLACEMENT SHEET
25 FEB 1997

the formula I without further purification. It is thereby possible to separate the resulting insulin of the formula I from the folding solution by adsorption. The insulins of the formula I can thus be obtained from the adsorbent without intermediate isolation or purification after desorption. The process according to the invention leads to a substantially shortened preparation, which because of shorter standing times and the avoidance of losses makes possible yield increases of 25-100 % compared with the known processes.

The amino acid sequence of peptides and proteins is designated from the N-terminal end of the amino acid chain. The information given in parentheses in formula I, e.g. A6, A20, B1, B7 or B19, corresponds to the position of amino acid residues in the A or B chains of the insulin.

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 stand for the term "genetically encodable amino acid radical".

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 hens. The term radicals A2 - A20 and B2 - B29 of insulin derivatives stands for the corresponding amino acid sequences of human insulin which are formed by the replacement of amino acids by other genetically encodable amino acids.

The residue Z is always part of the amino acid sequence of X and is formed by the activity of the proteases such as trypsin, trypsin-like enzymes or carboxypeptidase B. The residue R^3 is the amino acid residue which is in position A21 of the A chain of insulin. The residue Y is the amino acid residue which is in position B30 of the B chain of insulin.

REPLACEMENT SHEET
25 FEB 1997

Trypsin or trypsin-like enzymes are proteases which cleave amino acid chains at the arginine or lysine residue.

Carboxypeptidase B is an exoprotease which splits off 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).

The A chain of human insulin 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.

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.

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

The term hydrophobic adsorber resin stands for nonionic, hydrophobic, crosslinked polymers and/or copolymers, for example polyacrylates or polystyrene or copolymers of styrene and divinylbenzene, in particular polymeric adsorber resins having a large surface area and many large pores, e.g. commercial preparations of the companies Rohm and Haas or Mitsubishi Chemical Industries Ltd. such as XAD16, XAD1600 or HP20.

The term mercaptan is understood as meaning compounds which contain at least one -SH group. Water-soluble mercaptans are preferred. Examples of these compounds are

REPLACEMENT SHEET
25 FEB 1997

dithiothreitol, dithioerythritol, 2-mercaptoethanol, cysteine, glutathione, methyl thioglycolate, 3-mercapto-1,2-propanediol and 3-mercaptopropionic acid.

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

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

The procedure in process step A) is as follows:
The proteins are dissolved in a chaotropic auxiliary or in mixtures of various chaotropic auxiliaries. Preferably, urea or guanidine hydrochloride is used in a concentration of 6 to 9 M, preferably 8 M, based on water as a solvent. The pH of the reaction mixture is 8.5 to 10.8. Examples of buffers which can be used are phosphate, tri(hydroxymethyl)aminomethane (tris), borate or glycine buffer. The concentration of the buffer substances in the aqueous medium is up to 0.5 M, preferably from 0.005 M to 0.5 M, particularly preferably from 0.05 to 0.1 M. The protein mixture is then mixed with an aqueous mercaptan solution. By this means, the following concentrations are established in the reaction solution (based on water):

The concentration of the protein of the formula II is 0.05 to 0.6 g/l, preferably 0.1 to 0.3 g/l. The amount of

REPLACEMENT SHEET
25 FEB 1997

the mercaptan is 2 to 10 -SH residues of the mercaptan per cysteine residue of the protein of the formula II, preferably 6 to 9. The pH of the solution is 10 to 11, preferably 10.8. The abovementioned buffer components are used. The exact pH is established by addition of sodium hydroxide solution. The concentration of the buffer components is 0.005 to 0.5 M, preferably 0.05 to 0.1 M.

The concentration of the chaotropic auxiliary in the reaction mixture containing the mercaptan is less than 1 M, preferably 0.1 to 0.8 M, in particular 0.3 to 0.6 M. The mercaptan employed is preferably cysteine or 2-mercaptoethanol, alone or as a mixture.

The temperature during the folding in process step A) is 0°C to 50°C, preferably 2°C to 30°C, in particular 4°C to 10°C. The reaction time is 3 to 12 hours, preferably 4 to 8 hours, in particular 5 hours.

The result of process step A) is a proinsulin which contains correctly linked cystine bridges.

In process step B), the reaction solution from process step A) is adjusted to a pH of 6.5 to 9, preferably 8 to 9. Trypsin or a trypsin-like enzyme is added thereto in an amount from 1 to 3 mg based on 3 g of proinsulin having correctly linked cystine bridges. Optionally, carboxypeptidase B is added thereto in an amount of 3 to 8 units based on 1 g of insulin of the formula I. A mixture of trypsin and carboxypeptidase B can also be added. The solution obtained is then stirred at a temperature of 4°C to 15°C for from 4 to 16 hours. The result of process step B) is an insulin of the formula I which contains correctly linked cystine bridges.

In process step C), the reaction solution from process step B) is adjusted to a pH of 2.5 to 4.5 with an acid such as hydrochloric acid or sulfuric acid. If turbidity of the solution should occur, this is optionally removed

REPLACEMENT SHEET
25 FEB 1997

by filtration or centrifugation. The remaining solution is treated with a hydrophobic adsorber resin. Resins of the XAD or HP20 type have proven suitable. The amount of the resin is 3 to 50 g per liter of reaction solution, preferably 20 to 40 g/l.

The suspension obtained can be treated with a salt such as NaCl up to a concentration of 0.1 M, and is then stirred for 3 to 4 hours. The adsorption of the insulin of the formula I on the resin is checked by sampling and high-pressure liquid chromatographic analysis (HPLC analysis). As soon as insulin is no longer detectable in the solution, the adsorber resin is separated off from the aqueous reaction solution (process step D). This is carried out, for example, by filtration or centrifugation according to known methods. The resin containing adsorbed insulin of the formula I is washed with a purely aqueous or buffer-containing aqueous solution. In particular, it is washed until the conductivity in the wash solution is less than 0.4 mS/cm.

The desorption (process step E) of the insulin of the formula I is carried out by known methods, which depend on the hydrophobic adsorber resin used. The desorption is carried out in the case of XAD or HP20 adsorber resins, for example, by means of an aqueous solution which contains 20 to 65 % of a water-miscible, nonionic organic solvent, e.g. a (C_1-C_6)-alkanol. Preferably, 35 % of isopropanol or n-propanol is used in the solvent water. The desorption of the insulin of the formula I is carried out, for example, by washing with the isopropanol solution. The washing can be carried out on a stirred pressure filter by mixing with the isopropanol solution or by chromatography in a column. The volume ratio of resin to isopropanol solution is 1:1 to 1:10, preferably 1:1.1 to 1:2. The washing steps can be repeated one to five times, preferably twice. The combined isopropanol fractions can be used directly or after dilution with water for further chromatographic purifications.

REPLACEMENT SHEET
25 FEB 1997

In the following examples the process according to the invention is described in detail. Percentage data relate to the weight, if not stated otherwise.

Example 1

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

Proinsulin sequence 1 (Seq. Id No. 4):

Met 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 1 corresponds to the formula II, in this case

X is a C peptide from human insulin,
Y is Thr (B30),
R^1 is Phe (B1),
R^2 is a peptide having 11 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 1 collects in the E. coli cells and forms inclusion bodies. After completion of the fermentation, the cells are removed by centrifugation and disrupted by customary high-pressure homogenization. The fusion protein inclusion bodies released are isolated by centrifugation.

REPLACEMENT SHEET
25 FEB 1997

2600 g of the isolated fusion protein inclusion bodies (based on dry matter after freeze drying) having the proinsulin sequence 1 are dissolved in 200 l of an 8 M guanidine hydrochloride solution at pH 10.8. Optionally, after centrifugation of small amounts of suspended matter, the clear solution is stirred into 1800 l of an aqueous cysteine solution (250 g of cysteine hydrochloride hydrate) at a pH of 10.8 and a temperature of 4°C. The proportion of the insulin-containing fusion protein is determined with the aid of an SDS electrophoresis scan (U.K. Lämmli, Nature, Volume 227, pages 680-685, 1970). It is 45 %. After conclusion of the folding reaction, the content of proinsulin sequence I having correctly linked cystine bridges in the reaction mixture is determined as 760 g with the aid of analytical HPLC.

The 2000 l of solution is adjusted to a pH of 8.5 using 6N HCl. 500 mg of trypsin is then added. 305 g of insulin 2 are formed according to HPLC measurement.

Insulin 2 corresponds to the formula I, in this case

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).

Insulin 2 consists of the Seq Id No. 6 and 9, which are linked to one another via correctly bonded cystine bridges.

The solution is adjusted to a pH of 3.5 using 6N HCl and the conductivity value is adjusted to about 10 mS/cm with the aid of saturated NaCl solution. A slight turbidity is removed by centrifugation. 75 kg of HP20 (Mitsubishi Chemical Industries Ltd., Düsseldorf, Germany) are added

REPLACEMENT SHEET
25 FEB 1997

to the clear solution. The suspension is slowly stirred at 4°C for about 16 hours until insulin 2 is no longer to be detected in the supernatant. The resin is removed by filtration through a suction filter and washed with water until the conductivity in the wash solution is less than 0.4 mS/cm. The product is desorbed by suspension of the resin in 75 l of 50 % strength aqueous isopropanol solution. Filtration and incubation with the isopropanol solution is repeated twice. After 60 minutes, the resin is filtered and then incubated several times in a 35 % strength isopropanol solution and filtered. The yield of insulin 2 is 288 g.

The eluates which contain insulin 2 can be further purified immediately on a chromatography column after dilution with water and pH adjustment.

HPLC analysis

0.5 g of protein are dissolved in 40 ml of a solution of 6 M guanidine hydrochloride, 50 mM tris, pH 8.5, 5 mM ethylenediaminetetraacetate (EDTA), 1 % 2-mercaptoethanol and 10 mM dithiothreitol at 95°C for 2 min 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 running time:	40 min.

The gradient is characterized by the following amounts of buffer B after the corresponding running times:
10 min 25 %, 12 min 60 %, 13 min 90 %, 15 min 100 %.

Flow rate:	1 ml/min
Detection:	215 nm

REPLACEMENT SHEET
25 FEB 1997

Retention time
of insulin: about 19 min

Example 2

A fusion protein having the following amino acid sequence is prepared by fermentation of a genetically modified E. coli (EP 0 347 781):
Fusion protein 3 (Seq Id No. 5)

Met Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr Gln Leu
Gln Leu Glu His Leu Leu Leu Asp Leu Gln Met Ile Leu
Asn Gly Ile Asn Asn Tyr Lys Asn Pro Lys Leu Thr Arg
Met Ile Glu Gly 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 Gly Ile Val
Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
Glu Asn Tyr Cys Asn

This fusion protein contains the amino acid sequence of proinsulin 3. The expressed fusion protein 3 collects in the E. coli cells and forms inclusion bodies. The disruption of the cells is carried out as in Example 1. The fusion proteins 3 thus obtained are subjected to cyanogen bromide cleavage, and proinsulin 3 is formed (Seq Id No. 8).
Proinsulin 3 corresponds to the formula II, in this case
X is Arg
Y is Thr (B30),
R^1 is Phe (B1),
R^2 is a peptide having 4 amino acid residues,
R^3 is Asn (A21) and
A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin.

After cyanogen bromide cleavage, the content of proinsulin 3 is determined quantitatively with the aid of SDS electrophoresis in a freeze-dried sample containing 9.2 % insulin-containing protein.

REPLACEMENT SHEET
25 FEB 1997

2000 g of proinsulin 3 are incubated with cysteine hydrochloride hydrate as in Example 1. A content of 1260 g of proinsulin 3 having correctly linked cystine bridges is then determined in the total reaction mixture by means of analytical HPLC. The product is treated with 500 mg of trypsin as in Example 1.

Insulin 4 is formed. Insulin 4 corresponds to the formula I, in this case

Y is Thr (B30),

Z is Arg,

R^1 is Phe (B1)

R^3 is Asn (A21) and

A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin.

Insulin 4 consists of the Seq Id No. 1 and 7, which are linked to one another by correctly bonded cystine bridges.

The resulting solution (2 m^3) contains 800 g of the insulin derivative and is treated with 75 kg of HP20 as in Example 1. After adsorption and desorption, the combined eluate contains 703 g of insulin 4 (88 % yield).

REPLACEMENT SHEET
25 FEB 1997

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT

(A) NAME: Hoechst Aktiengesellschaft

(B) STREET: -

(C) CITY: Frankfurt am Main

(D) COUNTRY: Federal Republic of Germany

(E) POSTAL CODE: 65926

(F) TELEPHONE: 069-305-6047

(G) TELEFAX: 069-35-7175

(H) TELEX: 041234-700 hod

(ii) TITLE OF INVENTION: Process for obtaining insulin having correctly linked cystine bridges

(iii) NUMBER OF SEQUENCES: 9

(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 (EPA)

(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

REPLACEMENT SHEET
25 FEB 1997

(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

REPLACEMENT SHEET
25 FEB 1997

(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: 97 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..97

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

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

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

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

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

REPLACEMENT SHEET
25 FEB 1997

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

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

Gly

(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:

Met Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr Gln Leu Gln Leu Glu
1 5 10 15

His Leu Leu Leu Asp Leu Gln Met Ile Leu Asn Gly Ile Asn Asn Tyr
20 25 30

Lys Asn Pro Lys Leu Thr Arg Met Ile Glu Gly Arg Phe Val Asn Gln
35 40 45

His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly
50 55 60

Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr 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. 6:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

REPLACEMENT SHEET
25 FEB 1997

(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: 31 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..31

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

```
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
            20                  25                  30
```

REPLACEMENT SHEET
25 FEB 1997

(2) INFORMATION FOR SEQ ID NO. 8:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 56 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..56

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 8:

```
Ile Glu Gly Arg Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val
1               5                  10                  15

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

Lys Thr Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu
        35                  40                  45

Tyr Gln Leu Glu Asn Tyr Cys Asn
    50                  55
```

(2) INFORMATION FOR SEQ ID NO. 9:

(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

REPLACEMENT SHEET
25 FEB 1997

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 9:

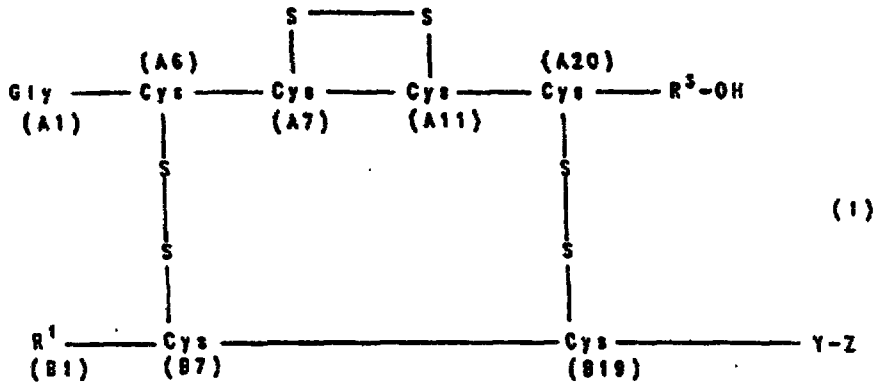
Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu
1 5 10 15

Glu Asn Tyr Cys Gly
20

REPLACEMENT SHEET
25 FEB 1997

Patent claims HOE 94/F 042

1. A process for obtaining insulin of the formula I



* C L M S *

which comprises

A) reacting a protein of the formula II

$R^2-R^1-B2-B29-Y-X-Gly-A2-A20-R^3$ (II)

with an amount of a mercaptan which yields 2 to 10 -SH radicals of the mercaptan per cysteine residue of the protein of the formula II, in the presence of at least one chaotropic auxiliary in an aqueous medium at a pH of 10 to 11 and a concentration of the protein of the formula II of 0.05 to 0.3 g per liter of aqueous medium and

B) reacting the resulting proinsulin having correctly linked cystine bridges with trypsin, a trypsin-like enzyme and optionally additionally with carboxypeptidase B or a mixture of the enzymes mentioned to give the insulin of the formula I having correctly linked cystine bridges,

C) treating the reaction product thus obtained with 3 to 50 g of a hydrophobic adsorber resin per liter of aqueous

REPLACEMENT SHEET
25 FEB 1997

* C L M S *

medium at a pH of 4 to 7,

D) isolating the adsorber resin containing adsorbed insulin of the formula I and

E) desorbing the insulin of the formula I from the adsorber resin;

in this case, in formulae I and II

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

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

Y is a genetically encodable amino acid residue,

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

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

c) OH,

R^1 is a phenylalanine residue or a covalent bond,

R^2 is a) a hydrogen atom,

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

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

R^3 is a genetically encodable amino acid residue and

the residues A2 - A20 correspond to the amino acid sequence of the A chain of human insulin, animal insulin or an insulin derivative and

the residues B2 - B29 correspond to the amino acid sequence of the B chain of human insulin, animal insulin or an insulin derivative.

REPLACEMENT SHEET
25 FEB 1997

2. The process as claimed in claim 1, wherein a protein of the formula II is employed, in this case

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

Y is an amino acid residue from the group consisting of Ala, Thr and Ser,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 25 amino acid residues, containing Arg at the carboxyl end of the peptide,

R^3 is an amino acid residue from the group consisting of Asn, Ser and Gly,

and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin and

an insulin of the formula I having correctly linked cystine bridges results, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or OH.

3. The process as claimed in claim 1, wherein a protein of the formula II is employed, in this case

X is the amino acid residue Arg or a peptide having 2 to 35 amino acid residues, two basic amino acid residues, in particular Arg and/or Lys, being at the start and end of the peptide,

Y is the amino acid residue Thr,

R^1 is the amino acid residue Phe,

R^2 is a) a hydrogen atom or

b) a peptide having 2 to 15 amino acid residues, at whose carboxyl end is the amino acid residue Arg,

REPLACEMENT SHEET
25 FEB 1997

* C L M S *

R^3 is the amino acid residue Gly or Asn, and the residues A2 - A20 and B2 - B29 correspond to the amino acid sequence of the A and B chains of human insulin and an insulin of the formula I having correctly linked cystine bridges is formed, in this case Y, R^1, R^2, R^3, A2-A20 and B2-B29 have the abovementioned definition and

Z is the amino acid residue Arg, the peptide residue Arg-Arg or Lys-Lys or OH.

4. The process as claimed in one or more of claims 1 to 3, wherein cysteine or cysteine hydrochloride hydrate is employed as the mercaptan.

5. The process as claimed in one or more of claims 1 to 4, wherein a crosslinked polystyrene, polyacrylate or copolymer of polystyrene and divinylbenzene is employed as the hydrophobic adsorber resin.

6. The process as claimed in one or more of claims 1 to 5, wherein the insulin of the formula I is desorbed from the adsorber resin using isopropanol or n-propanol.

7. The process as claimed in one or more of claims 1 to 6, wherein 0.01 to 0.1 M salt is added in the process step C).

8. An insulin of the formula I, obtained by a process as claimed in one or more of claims 1 to 7.

REPLACEMENT SHEET
25 FEB 1997

Abstract

Process for obtaining insulin having correctly linked cystine bridges

A process for obtaining insulin having correctly linked cystine bridges from a corresponding proinsulin amino acid chain in the presence of mercaptan, chaotropic auxiliaries and at a pH of 10 to 11 is described, as well as the direct conversion of the proinsulin to insulin with the aid of enzymes and direct isolation of the insulin from the reaction mixture with the aid of hydrophobic adsorber resins.