

Brevet N° **82757**
 du **10 septembre 80**
 Titre délivré : **L-2 FEV 1981**

GRAND-DUCHÉ DE LUXEMBOURG



Monsieur le Ministre
 de l'Economie Nationale et des Classes Moyennes
 Service de la Propriété Industrielle
 LUXEMBOURG

Demande de Brevet d'Invention

I. Requête

THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, 2200 University Avenue, à BERKELEY, Californie 94720, Etats-Unis d'Amérique, représentée par Monsieur Jacques de Muyser, agissant en qualité de mandataire

dépose ce dix septembre 1980 quatre-vingt à 15 heures, au Ministère de l'Economie Nationale et des Classes Moyennes, à Luxembourg :

1. la présente requête pour l'obtention d'un brevet d'invention concernant :
 "Gènes de la proinsuline humaine et le pré-proinsuline humaine, leur procédé de préparation et leurs applications".

déclare, en assumant la responsabilité de cette déclaration, que l'(es) inventeur(s) est (sont) :
 voir au verso

2. la délégation de pouvoir, datée de le
 3. la description en langue française de l'invention en deux exemplaires ;
 4. 2 planches de dessin, en deux exemplaires ;
 5. la quittance des taxes versées au Bureau de l'Enregistrement à Luxembourg,

le 10 septembre 1980
 revendique pour la susdite demande de brevet la priorité d'une (des) demande(s) de
 (6) brevet déposée(s) xx (7) aux Etats-Unis d'Amérique
 le 12 septembre 1979 (No. 75 192)

au nom de s inventeurs
 élit domicile pour lui (elle) et, si désigné, pour son mandataire, à Luxembourg
35, bld. Royal

sollicite la délivrance d'un brevet d'invention pour l'objet décrit et représenté dans les annexes susmentionnées, — avec ajournement de cette délivrance à /// mois.

Le mandataire

II. Procès-verbal de Dépôt

La susdite demande de brevet d'invention a été déposée au Ministère de l'Economie Nationale et des Classes Moyennes, Service de la Propriété Industrielle à Luxembourg, en date du :

10 septembre 1980

à 15 heures

Pr. le Ministre
 de l'Economie Nationale et des Classes Moyennes,
 p. d.

- 1.- Graeme BELL, 1611 Funston Avenue, à SAN FRANCISCO, Californie 94122, Etats-Unis d'Amérique
- 2.- Raymond Louis PICTET, 1524 Willard Street, à SAN FRANCISCO, Californie 94117, Etats-Unis d'Amérique
- 3.- Howard Michael GOODMAN, 3006 Clay Street, à SAN FRANCISCO, Californie 94115, Etats-Unis d'Amérique
- 4.- William J. RUTTER, 80 Everson Street, à SAN FRANCISCO, Californie 94131, Etats-Unis d'Amérique

REVENDEICATION DE LA PRIORITE

de la demande de brevet / du modèle d'utilité/

Et Aux ETATS-UNIS D'AMERIQUE

Du 12 SEPTEMBRE 1979

Mémoire Descriptif

déposé à l'appui d'une demande de

BREVET D'INVENTION

au

Luxembourg

au nom de : THE REGENTS OF THE UNIVERSITY OF CALIFORNIA

pour : "Gènes de la proinsuline humaine et de pré-proinsuline humaine, leur procédé de préparation et leurs applications".

L'invention fournit le gène humain isolé et purifié (cloné) codant pour la proinsuline et le gène humain codant pour la pré-proinsuline, des procédés d'isolement et de purification des gènes et un procédé de transfert des gènes à, et
 5 de reproduction des gènes dans, un microorganisme. Les gènes isolés et purifiés sont exprimés par un microorganisme hôte lorsqu'ils sont fusionnés avec un gène de procaryote hôte exprimable. Les deux gènes sont utilisables dans la production d'insuline humaine dans des buts thérapeutiques;

10 L'insuline est une hormone produite principalement par les cellules B du pancréas. Actuellement, l'utilisation de cette hormone dans le traitement du diabète est bien connue. Bien que les abattoirs fournissent des pancréas de boeuf et de porc comme sources d'insuline, une pénurie de cette hor-
 15 mone se fait sentir en raison du nombre croissant de diabétiques dans le monde. De plus, certains diabétiques sont allergiques à l'insuline de boeuf et de porc, avec des effets néfastes. Une possibilité de produire de l'insuline humaine en quantités suffisantes pour satisfaire les besoins mondiaux
 20 est donc hautement désirable. L'invention fournit des gènes qui sont insérables dans des microorganismes, qui sont utilisables dans la production d'insuline humaine.

L'insuline est constituée de deux chaînes de polypeptides connues sous le nom de chaîne A et B, reliées ensembles
 25 par des ponts disulfure. La chaîne A consiste en 21 acides aminés et la chaîne B en 30 acides aminés. Ces chaînes ne sont pas synthétisées indépendamment in vivo, mais dérivent d'un précurseur immédiat, appelé proinsuline. La proinsuline est une simple chaîne de polypeptide qui contient un peptide
 30 appelé le peptide C qui relie les chaînes A et B. Voir Steiner D.F. et al., Science 157, 697 (1967). Ce peptide C est excisé pendant l'emmagasinement de l'insuline dans les granules sécréteurs des cellules B du pancréas avant la sécrétion. Voir Tager H.S. et al., Ann. Rev. Biochem. 43, 509 (1974). On pen-

[illegible]

Dans les cellules B du pancréas, le produit de translation initial n'est pas la proinsuline elle-même, mais une pré-proinsuline qui contient plus de 20 acides aminés supplémentaires sur la terminaison amino de la proinsuline. Voir Cahn S.J. et al., Proc. Nat. Acad. Sci. USA 73, 1964 (1976) et Lomedico P.T. et al., Nucl. Acid Res. 3, 381 (1976). L'autre séquence d'acides aminés est appelée le peptide "signal".

30 Dans la pré-proinsuline humaine (voir la Fig.2 des dessins annexés) le peptide signal comporte 24 acides aminés et la séquence est la suivante : $\text{NH}_2\text{-Met}^{\text{-24}}\text{-Ala-Leu-Trp-Met}^{\text{-20}}\text{-Arg-Leu-}$
 $\text{Leu-Pro-Leu-Leu}^{\text{-10}}\text{-Ala-Leu-Leu-Ala-Leu-Trp-Gly-Pro-Asp-Pro-Ala-}$
 $\text{Ala-Ala}^{\text{-1}}$. On pense que la séquence de 24 acides aminés serait

un signal spécifique pour le transport vectorial du polypeptide synthétisé dans le réticulum endoplasmique de la cellule B, et est séparée de la proinsuline pendant cette phase. Voir Blobel G. et al., J. Cell. Biol. 67, 835 (1975).

5 On connaît plusieurs cas de peptides "signal" pour des protéines d'eucaryotes à transporter à travers des barrières de membrane. Une enzyme de scission spécifique a été observée dans un système sans cellules qui hydrolyse la liaison peptidique entre le peptide signal et la protéine active
10 en même temps qu'un passage à travers une membrane cellulaire. (Voir Blobel G. et al., Proc. Nat. Acad. Sci. USA 75, 361 (1978)).

De récents progrès en biochimie et dans la technologie de l'ADN recombinant ont rendu possible la synthèse de
15 protéines synthétiques dans des conditions déterminées indépendantes de l'organisme supérieur dont elles sont normalement isolées. Ces procédés de synthèse biochimiques utilisent des enzymes et des constituants sous-cellulaires du mécanisme de synthèse des protéines des cellules vivantes, soit in vi-
20 tro, dans des systèmes sans cellules, soit in vivo, dans des microorganismes. Dans l'un ou l'autre cas, l'élément clé est la fourniture d'un acide désoxyribonucléique (ADN) de séquence spécifique qui contient l'information nécessaire pour spécifier la séquence d'acides aminés désirée. Un tel ADN spéci-
25 fique est appelé ici un gène. La relation codante où une séquence de désoxynucléotides est utilisée pour spécifier la séquence d'acides aminés d'une protéine est décrite brièvement, ci-dessous, et fonctionne selon un ensemble fondamental de principes qui prévalent dans la totalité du domaine connu
30 des organismes vivants.

Un gène cloné peut être utilisé pour spécifier la séquence d'acides aminés des protéines synthétisées par des systèmes in vitro. Des systèmes de synthèse des protéines à partir de l'ADN sont bien connus dans la technique, voir, par
35 exemple, Zubay G., Ann. Rev. Genetics 7, 267 (1973). De plus,

un ADN à simple chaîne (spirale) peut être induit pour jouer le rôle d'ARN messager in vitro, fournissant une translation très fidèle de la séquence d'ADN (Salas, J. et al. J. Biol. Chem. 243, 1012 (1968). D'autres procédés bien connus dans la technique peuvent être utilisés en combinaison avec les procédés ci-dessus pour augmenter les rendements.

Des développements concernant la technologie de l'ADN recombinant ont permis l'isolement de gènes ou de portions de gènes spécifiques à partir d'organismes supérieurs, tels que l'homme et d'autres mammifères, et de transférer les gènes ou fragments à un microorganisme, tel qu'une bactérie ou une levure. Le gène transféré est reproduit et propagé lorsque le microorganisme transformé est reproduit. Il en résulte que le microorganisme transformé peut devenir capable de produire toute protéine pour laquelle code le gène ou le fragment, que ce soit une enzyme, une hormone, un antigène ou un anticorps ou une portion de ces substances. Le microorganisme transmet sa capacité à sa descendance, de sorte qu'en fait le transfert a pour résultat une nouvelle souche, possédant l'aptitude décrite. Voir par exemple Ullrich A. et al., Science 196, 1313 (1977) et Seeburg P.H. et al., Nature 270, 486 (1977). Un fait fondamental qui sert de base à l'application de cette technologie à des buts pratiques, est que l'ADN de tous les organismes vivants, des microbes à l'homme, est chimiquement le même, étant composé des quatre mêmes nucléotides. Les différences significatives résident dans les séquences de ces nucléotides dans la molécule d'ADN polymère. Les séquences de nucléotides sont utilisées pour spécifier les séquences d'acides aminés des protéines qui constituent l'organisme. Bien que la plupart des protéines de différents organismes diffèrent les unes des autres, la relation codante entre une séquence de nucléotides et une séquence d'acides aminés est fondamentalement la même pour tous les organismes. Par exemple, la même séquence de nucléotides qui code pour la séquence d'acides aminés de l'HGH (hormone gonadotrope

humaine) dans les cellules de l'hypophyse humaine, sera, une fois transférée à un microorganisme, reconnue comme codant pour la même séquence d'acides aminés.

Les abréviations utilisées ici sont expliquées dans le Tableau 2.

TABLEAU 2

	ADN	=	acide désoxyribonucléique
	ARN	=	acide ribonucléique
	cADN	=	ADN complémentaire
10		(synthétisé par voie enzymatique à partir d'une	séquence de mARN)
	mARN	=	ARN messenger
	dATP	=	désoxyadénosine triphosphate
	dGTP	=	désoxyguanosine triphosphate
15	dCTP	=	désoxycytidine triphosphate
	A	=	adénine
	T	=	thymine
	G	=	guanine
	C	=	cytosine
20	U	=	uracile
	ATP	=	adénosine triphosphate
	TTP	=	thymidine triphosphate
	EDTA	=	acide éthylène diamine tétraacétique

Les relations codantes entre une séquence de nucléotides dans l'ADN et une séquence d'acides aminés dans la protéine sont collectivement connues sous le nom de "code génétique", tel qu'indiqué dans le Tableau 3.

TABLEAU 3

30	<u>CODE GENETIQUE</u>			
	Phénylalanine(Phe)	TTK	Histidine(His)	CAK
	Leucine(Leu)	XTY	Glutamine(Gln)	CAJ
	Isoleucine(Ile)	ATM	Asparagine(Asn)	AAK
	Méthionine(Met)	ATG	Lysine(Lys)	AAJ

	Valine(Val)	GTL	Acide aspartique(Asp)	GAK
	Sérine(Ser)	QRS	Acide glutamique(Glu)	GAJ
	Proline(Pro)	CCL	Cystéine(Cys)	TGK
	Thréonine(Thr)	ACL	Tryptophane(Try)	TGG
5	Alanine(Ala)	GCL	Arginine(Arg)	WGZ
	Tyrosine(Tyr)	TAK	Glycine(Gly)	GGL
	Signal de terminaison	TAJ		
	Signal de terminaison	TGA		

10

Explication : chaque triplet de désoxynucléotide, de 3 lettres, correspond à un trinucéotide de mARN, ayant une extrémité 5' à gauche et une extrémité 3' à droite. Toutes les séquences d'ADN indiquées ici sont celles de la chaîne dont la séquence correspond à la séquence de mARN, l'uracile étant remplacé par la thymine. Les lettres représentent les bases purine ou pyrimidine formant la séquence de désoxynucléotides.

	A	=	adénine	J	=	A ou G
	G	=	guanine	K	=	T ou C
20	C	=	cytosine	L	=	A, T, C ou G
	T	=	thymine	M	=	A, C ou T
	X	=	T ou C si Y est A ou G			
	X	=	C si Y est C ou T			
	Y	=	A, G, C ou T si X est C			
25	Y	=	A ou G si X est T			
	W	=	C ou A si Z est A ou G			
	W	=	C si Z est C ou T			
	Z	=	A, G, C ou T si W est C			
	Z	=	A ou G si W est A			
30	QR	=	TC si S est A, G, C ou T			
	QR	=	AG si S est T ou C			
	S	=	A, G, C ou T si QR est TC			
	S	=	T ou C si QR est AG			

Une caractéristique importante du code, pour le but

de l'invention, réside dans le fait que chaque acide aminé est spécifié par une séquence de trinuécléotide, également connue comme triplet de nucléotide. Les liaisons phosphodiester réunissant des triplets adjacents ne peuvent pas être chimiquement distinguées de toutes les autres liaisons internuécléotides dans l'ADN. En conséquence la séquence de nucléotides ne peut pas être lue pour coder pour une séquence unique d'acides aminés sans une autre information pour déterminer le cadre de lecture qui est le terme utilisé pour indiquer le groupe de triplets utilisé par la cellule en décodant le message génétique.

De nombreuses techniques d'ADN recombinant utilisent deux classes de composés, les vecteurs de transfert et les enzymes de restriction, qui seront discutés successivement.

Un vecteur de transfert est une molécule d'ADN qui contient, entre autres, une information génétique qui assure sa propre reproduction lorsqu'elle est transférée à une souche de microorganisme hôte. Des exemples de vecteurs de transfert communément utilisés dans les génétiques bactériennes sont les plasmides et l'ADN de certains bactériophages. Bien que des plasmides aient été utilisés comme vecteurs de transfert pour le travail décrit ici, il est évident que d'autres types de vecteurs de transfert peuvent être également utilisés. "Plasmide" est le terme appliqué à toute unité d'ADN à reproduction autonome, qui peut être trouvée dans une cellule microbienne, autre que le génome de la cellule hôte elle-même.

L'ADN plasmique existe sous forme de structures cycliques à double chaîne, généralement de l'ordre de quelques millions de daltons de poids moléculaire, bien que certaines aient un poids moléculaire supérieur à 10^8 daltons. Elles ne représentent normalement qu'un faible pourcentage de l'ADN total de la cellule. L'ADN du vecteur de transfert est normalement séparable de l'ADN de la cellule hôte grâce à la grande différence de dimension entre eux. Les vecteurs de transfert portent une information génétique les rendant capables de se re-

produire dans la cellule hôte, dans la plupart des cas indépendamment de la vitesse de division de la cellule hôte.

Certains plasmides ont la propriété d'avoir une vitesse de reproduction réglable par le chercheur en modifiant les conditions de croissance. Au moyen de techniques appropriés, le
5 noyau de l'ADN plasmique peut être ouvert, un fragment d'ADN hétérologue inséré et le noyau refermé, formant une plus grosse molécule comprenant le segment d'ADN inséré. L'ADN bactériophage peut porter un segment d'ADN hétérologue insé-
10 ré au lieu de certains gènes provenant d'autres phages non-essentiels. Dans l'un ou l'autre cas, le vecteur de transfert sert de support ou de vecteur pour un fragment d'ADN hétérologue inséré.

Le transfert a lieu selon un procédé connu comme étant
15 tant une transformation. Pendant la transformation, les cellules bactériennes mélangées avec l'ADN plasmique incorporent des molécules entières de plasmide dans les cellules. Bien que le mécanisme du processus reste obscur, il est possible de maximaliser la proportion de cellules bactériennes capables de fixer l'ADN plasmique et par conséquent de se trans-
20 former par certains traitements déterminés par voie empirique. Une fois qu'une cellule a incorporé un plasmide, ce dernier est reproduit dans la cellule et les répliques du plasmide sont réparties dans les cellules filles lorsque la cellule
25 se divise. Toute information génétique contenue dans la séquence de nucléotides de l'ADN plasmique peut, en principe, être exprimée dans la cellule hôte. Typiquement, une cellule hôte transformée est reconnue par son acquisition de caractères portés sur le plasmide, tels que la résistance à certains anti-
30 antibiotiques. Des plasmides différents sont reconnaissables par les aptitudes différentes ou une combinaison différente d'aptitudes qu'ils confèrent à la cellule hôte les contenant. Tout plasmide donné peut être reproduit en quantité, en développant une culture pure de cellules contenant le plasmide
35 et en isolant l'ADN plasmique à partir de la culture.

Les endonucléases de restriction sont des enzymes hydrolytiques capables de catalyser une scission des molécules d'ADN en un site spécifique. Le lieu de l'action de l'endonucléase de restriction est déterminé par l'existence d'une

5 séquence de nucléotides spécifique. Une telle séquence est appelée le site de reconnaissance pour l'endonucléase de restriction. Des endonucléases de restriction de sources diverses ont été isolées et caractérisées en termes de la séquence de nucléotides de leurs sites de reconnaissance. Certaines

10 endonucléases de restriction hydrolysent les liaisons phosphodiester sur les deux chaînes au même point, produisant des extrémités tronquées. D'autres catalysent l'hydrolyse de liaisons séparées l'une de l'autre par quelques nucléotides, produisant des régions à simple chaîne libre à chaque extrémité

15 de la molécule scindée. Ces extrémités à simple chaîne sont auto-complémentaires, donc cohésives, et peuvent être utilisées pour réunir l'ADN hydrolysé. Etant donné que tout ADN capable d'être scindé par une telle enzyme doit contenir le même site de reconnaissance, les mêmes extrémités cohésives seront produites, de sorte qu'il est possible de réunir des séquences

20 hétérologues d'ADN qui ont été traitées avec une endonucléase de restriction à d'autres séquences ayant subi le même traitement. Voir Roberts R.J., Crit. Rev. Biochem., 4, 123 (1976). Les sites de restriction sont relativement rares, mais l'utilité générale des endonucléases de restriction a été considérablement élargie par la synthèse chimique d'oligonucléotides à double chaîne portant la séquence du site de restriction.

25 En conséquence, virtuellement tout segment d'ADN peut être accouplé à tout autre segment en fixant simplement l'oligonucléotide de restriction approprié aux extrémités de la molécule, et en soumettant le produit à l'action hydrolytique de l'endonucléase de restriction appropriée, d'où la production d'extrémités cohésives requises. Voir Heyneker H.L. et al. Nature 263, 748 (1976) et Scheller R.H. et al., Science 196,

30 177 (1977). Une caractéristique importante de la répartition

des sites de reconnaissance d'une endonucléase de restriction est le fait qu'ils sont répartis au hasard quant au cadre de lecture. En conséquence, la scission par une endonucléase de restriction peut se produire entre deux codons
 5 adjacents ou à l'intérieur d'un codon.

On dispose de procédés plus généraux de scission de l'ADN ou de modification de la séquence terminale. Un grand nombre d'endonucléases non-spécifiques peuvent être utilisées pour scinder statistiquement l'ADN, tel que discuté ci-
 10 dessous. Les séquences terminales peuvent être modifiées par la création de queues d'oligonucléotides de dA à une extrémité et de dT à l'autre, ou de dG et de dC, pour créer des sites de réunion sans avoir besoin de séquences de liaison spécifiques.

15 Le terme "expression" est utilisé pour souligner le fait qu'un organisme utilise rarement pour ne pas dire jamais, la totalité de ses aptitudes génétiques à un moment donné. Même dans des organismes relativement simples tels que des bactéries, de nombreuses protéines que la cellule est capable de synthétiser, ne sont pas synthétisées alors qu'elles
 20 pourraient l'être dans des conditions ambiantes appropriées. Lorsque le produit protéinique, codé par un gène donné, est synthétisé par l'organisme, le gène est dit "être exprimé". Si le produit protéinique n'est pas réalisé, le gène n'est
 25 pas exprimé. Normalement, l'expression des gènes dans E. coli est réglée tel que décrit généralement ci-après, de telle sorte que les protéines dont la fonction n'est pas utilisée dans un milieu donné, ne sont pas synthétisées et l'énergie métabolique est conservée.

30 Le mécanisme selon lequel l'expression du gène est réglée dans E. coli est bien connu, grâce aux études intensives de ces vingt dernières années. Voir, en général, Hayes W., The Genetics of Bacteria and Their Viruses, 2ème édition, John Wiley et Sons, New York (1968) et Watson J.D., The Molecu-
 35 lar Biology of the Gene, 3ème édition, Benjamin, Menlo Park,

Californie (1976). Ces études ont révélé que plusieurs gènes, normalement ceux codant pour des protéines exerçant des fonctions apparentées dans la cellule, sont trouvés groupés ensemble dans une séquence continue. Le groupe est appelé un

5 "opéron". Tous les gènes dans l'opéron sont transcrits dans la même direction, en commençant avec les codons codant pour l'acide aminé N-terminal de la première protéine dans la séquence et en continuant jusqu'à l'extrémité C-terminale de la dernière protéine dans l'opéron. Au commencement de l'opéron,

10 il existe une région de l'ADN, appelée la région de contrôle, qui comprend une variété d'éléments de contrôle dont l'opérateur, le promoteur et les séquences pour les sites de liaison ribosomiques. Le rôle de ces sites est de permettre que, sous leur contrôle, l'expression de ces gènes réponde aux

15 besoins de l'organisme. Par exemple, les gènes codant pour des enzymes requises exclusivement pour l'utilisation du lactose ne sont normalement pas beaucoup exprimés à moins que le lactose ou un de ses analogues ne soit réellement présent dans le milieu. Les fonctions de la région de contrôle nécessaires à l'expression sont l'initiation de la transcription

20 et l'initiation de la translation. L'expression du premier gène dans la séquence est initiée par l'initiation de la transcription et de la translation à la position codant pour l'acide aminé N-terminal de la première protéine de l'opéron. L'ex-

25 pression de chaque gène à partir de ce point est donc initiée à son tour, au moins jusqu'à ce qu'un signal de terminaison ou un autre opéron soit rencontré avec sa propre région de contrôle, accordée pour répondre à un jeu différent d'indications ambiantes. Bien que les détails correspondant à ce schéma général puissent varier énormément, le fait important est

30 que, pour être exprimé dans un procaryote tel que E. coli, un gène doit être convenablement situé quant à une région de contrôle ayant le rôle d'initiateur de transcription et d'initiateur de translation.

On a démontré que des gènes ne faisant pas normalement partie d'un opéron donné, peuvent être insérés dans l'opéron et contrôlés par lui. La démonstration classique en a été faite par Jacob F., et al., J. Mol. Biol., 13, 704 (1965).

5 Dans cette expérience, des gènes codant pour des enzymes impliquées dans un processus de biosynthèse de la purine, ont été transférés vers une région contrôlée par l'opéron du lactose. L'expression de l'enzyme biosynthétique de purine était visiblement réprimée en l'absence de lactose ou d'un analogue du lactose, et ne répondait pas aux indications ambiantes
10 réglant normalement son expression.

En plus de la région de l'opérateur réglant l'initiation de la transcription des gènes à partir d'elle, on sait qu'il existe des codons qui agissent comme des signaux d'arrêt, indiquant l'extrémité C-terminale d'une protéine donnée.
15 Voir Tableau 3. Ces codons sont connus comme des signaux de terminaison et également comme des codons extravagants car ils ne codent normalement pas pour un acide aminé quelconque. La suppression d'un signal de terminaison entre des gènes structuraux d'un opéron crée un gène fusionné qui pourrait
20 avoir pour résultat la synthèse d'une protéine chimérique constituée de deux séquences d'acides aminées codées par des gènes adjacents, réunies par une liaison peptidique. Cette synthèse de protéines chimériques lorsque des gènes sont fusionnés a été démontrée par Benzer S. et Champe S.P., Proc. Nat. Acad. Sci. USA 48, 114 (1962).
25

Dès qu'un gène a été isolé, purifié et inséré dans un vecteur de transfert, le résultat global de ce qu'on appelle "faire un clone" à partir du gène, sa disponibilité en quantité substantielle est assurée. Le gène ainsi cloné est transféré dans un microorganisme approprié où le gène se reproduit lorsque le microorganisme prolifère et à partir duquel le gène peut être de nouveau isolé par des moyens classiques. Cette technique fournit donc une source continuellement renouvelable
35 du gène pour d'autres manipulations, modifications et transferts

à d'autres vecteurs ou en d'autres lieux dans le même vecteur.

L'expression est obtenue en transférant le gène cloné, dans une orientation et un cadre de lecture appropriés, dans une région de contrôle telle qu'une lecture complète à partir du gène de procaryote se traduise par la synthèse d'une protéine chimérique comprenant la séquence d'acides aminés codés par le gène cloné. De nombreuses techniques de scission spécifique d'une protéine peuvent être utilisées pour scinder la protéine chimérique en un point désiré de façon à libérer la séquence d'acides aminés recherchés qui peut alors être purifiée par des moyens classiques. Des techniques de construction d'un vecteur de transfert d'expression où le gène cloné est convenablement juxtaposé avec une région de contrôle, sont décrites par Polisky B. et al., Proc. Nat. Acad. Sci. USA 73, 3900 (1976) ; Itakura K. et al., Science 198, 1056 (1977) ; Villa-Komaroff L. et al., Proc. Nat. Acad. Sci. USA 75, 3727 (1978) ; Mercereau-Puijalon O. et al., Nature 275, 505 (1978) ; Chang A.C.Y. et al., Nature 275, 617 (1978) et dans la demande de brevet US n° 933 035 de Rutter et al.

En résumé, le procédé qui permet de produire une protéine de mammifère, telle qu'une pré-proinsuline ou un proinsuline humaine, en appliquant la technologie de l'ADN recombinant, requiert d'abord l'isolement et la purification du gène de mammifère. Une fois cloné, le gène peut être produit en quantité, modifié par un moyen chimique ou enzymatique et transféré à un plasmide d'expression. Le gène cloné est également utilisable pour isoler des gènes voisins, ou si seul un fragment est cloné, pour isoler le gène entier, en utilisant le gène cloné comme modèle d'hybridation. De plus, le gène cloné est utilisable pour prouver par hybridation l'identité ou le caractère homologue des isolats indépendants de gènes identiques ou apparentés. En raison de la nature du code génétique, le gène cloné, lorsqu'il est transmis dans le cadre de lecture approprié, ne dirige que la production de la séquence d'acides aminés pour laquelle il code et aucune

autre.

L'isolement et la purification de la proinsuline du rat ont fait l'objet de divers travaux. Ullrich A. et al., voir ci-dessus, et Villa-Komaroff L. et al., voir ci-dessus, décrivent l'isolement et la purification du gène de la proinsuline du rat et un procédé de transfert et de reproduction de ce gène dans un microorganisme. Ullrich et al. ont recueilli plusieurs plasmides recombinants qui contenaient la séquence de codage pour la proinsuline, la région non transmise 3' et une partie du pré-peptide. L'expression de l'ADN de rat contenant la séquence codant pour l'insuline a été décrite dans la demande de brevet US n° 933 035. Villa-Komaroff et al. ont recueilli un plasmide recombinant qui contenait la séquence codant pour les acides aminés 4 à 86 de la proinsuline. Cette séquence de proinsuline a été séparée des acides aminés 24 à 182 de la pénicilline (β -lactamase) par la séquence codant pour six glycines. Cette séquence codant pour la proinsuline de pénicilline a été exprimée pour produire une protéine fusionnée. Ces articles décrivent une partie des procédés de base utilisés dans la technologie de l'ADN recombinant. Cependant, ils ne décrivent ni l'isolement, ni la purification du gène de la pré-proinsuline humaine ou du gène de la proinsuline humaine.

Crea R. et al., Proc. Nat. Acad. Sci USA 75, 5765 (1978), ont suivi une voie différente pour obtenir de l'insuline humaine. Cette voie consiste à synthétiser chimiquement des séquences codant pour (1) la chaîne A et (2) la chaîne B de l'insuline humaine, en utilisant des codons favorisés par E. coli. Ces deux séquences peuvent alors être insérées dans des plasmides qui peuvent être exprimés pour produire les chaînes A et B. L'insuline humaine pourrait alors être obtenue par formation des liaisons disulfure correctes entre les deux chaînes de protéine.

Le gène cloné pour la pré-proinsuline humaine est utilisable de diverses façons. La transposition à un vecteur de

transfert d'expression permet la synthèse de la pré-proinsuline par un microorganisme hôte transformé par le vecteur portant le gène cloné. La prolifération de l'hôte transformé se traduit par la synthèse de la pré-proinsuline comme partie d'une protéine chimérique. Si la portion de procaryote de la protéine fusionnée est la portion signal d'une protéine hôte excrétée ou sectionnée d'une autre façon en compartiments, l'excrétion ou la division en compartiments peuvent augmenter de façon importante la stabilité et la facilité de purification de la protéine fusionnée de pré-proinsuline. De plus, lorsque la portion de procaryote est courte, l'excrétion hors de l'hôte de procaryote peut être facilitée par le pré-peptide lui-même, si la pré-séquence fonctionne dans l'hôte de procaryote comme elle le fait dans la cellule d'eucaryote. Le gène de la pré-proinsuline peut également être utilisé pour obtenir le gène de la proinsuline en utilisant les techniques décrites ci-après.

Le gène cloné de la pré-proinsuline peut être utilisé dans diverses techniques pour produire la pré-proinsuline. La pré-proinsuline elle-même est utilisable en raison du fait qu'elle peut être convertie en proinsuline par des techniques enzymatiques et chimiques connues. Par exemple, le prépeptide peut être éliminé par une préparation enzymatique soluble, tel que décrit par Blobel G. et al., ci-dessus, spécifique pour l'élimination des peptides "signal". Le gène cloné de la proinsuline peut être utilisé dans diverses techniques pour la production de proinsuline. La proinsuline, produite à partir de l'un de ces gènes, est utilisable elle-même, car elle peut être convertie en insuline par des techniques enzymatiques et chimiques connues. Voir Kemmber W. et al., J. Biol. Chem. 242, 6786 (1971).

L'invention concerne un procédé de préparation d'un vecteur de transfert d'ADN, utilisable pour conserver et reproduire une séquence de désoxynucléotides codant pour la pré-proinsuline humaine, caractérisé en ce qu'on fait réagir une

séquence de désoxynucléotides codant pour la pré-proinsuline humaine avec une molécule d'ADN préparée par scission d'un vecteur de transfert par une enzyme de restriction.

5 Tel que décrit ici, un cADN ayant la séquence de base codant pour la pré-proinsuline humaine et un cADN ayant la séquence de base codant pour la proinsuline humaine sont isolés et purifiés (clonés). La structure des cADN clonés est vérifiée par analyse des séquences de nucléotides.

10 La source originale du matériau génétique est l'insulinoma humain (une tumeur produisant de l'insuline). Un ARN messager est isolé des cellules.

15 Un ADN complémentaire à l'ARN messager isolé (cADN) est synthétisé en utilisant la transcriptase inverse, dans deux cycles de réaction pour former le cADN à double chaîne. Le cADN à double chaîne, non-scindé, hétérogène, est traité pour fournir des séquences d'oligonucléotides de liaison, spécifiques, à chaque extrémité, afin de faciliter l'insertion dans un site de même spécificité de restriction sur un vecteur de transfert.

20 Pour l'isolement et la purification, on choisit un vecteur de transfert assurant une bonne sélection et des propriétés de reproduction stables. On insère le cADN traité dans un vecteur de transfert en un site prédéterminé sur le vecteur d'ADN pour former un vecteur de transfert recombinant, en utilisant des techniques courantes. Des cellules de microorganismes hôte sont transformées par le vecteur recombinant. Des agents de transformation ou transformants sont choisis selon les critères établis pour l'insertion en le site prédéterminé. Des colonies simples, dérivées chacune d'une seule cellule de microorganisme transformée, sont rassemblées et multipliées dans des cultures individuelles afin de permettre la reproduction des clones d'ADN du vecteur de transfert recombinant. On sélectionne l'ADN du vecteur de transfert pour les séquences d'insuline par un procédé d'hybridation de colonies modifié. Des colonies donnant un ADN de vecteur de

25
30
35

transfert recombinant pour les séquences d'insuline sont multipliées dans de plus grandes cultures individuelles pour isoler l'ADN du vecteur de transfert et le soumettre à une analyse. On choisit des clones appropriés pour une identification définitive de la séquence de nucléotides des gènes de pré-proinsuline ou de proinsuline clonés, et pour le transfert à des plasmides d'expression appropriés.

L'invention fournit en outre un procédé de préparation d'une séquence de désoxynucléotides codant pour la proinsuline humaine, utilisable dans la synthèse de l'insuline humaine, caractérisé en ce qu'on hydrolyse les extrémités 3' d'une séquence de désoxynucléotides codant pour la pré-proinsuline humaine, conservée et reproduite selon le procédé de l'invention, par un ou plusieurs cycles de digestion exonucléolytique progressive, catalysée par la T4 ADN polymérase agissant en présence d'un désoxynucléotide triphosphate, puis en hydrolysant les extrémités 5' de la séquence de désoxynucléotides par une digestion nucléolytique catalysée par une enzyme de nucléase spécifique pour un ADN à simple chaîne ayant une extrémité 5'.

L'invention fournit en outre un procédé de synthèse de la proinsuline humaine comprenant des peptides A et B séparés l'un de l'autre par un peptide C, caractérisé par l'incubation d'un microorganisme, transformé par un vecteur de transfert d'expression comprenant une séquence de désoxynucléotides codant pour la proinsuline humaine préparée tel que décrit ici, dans des conditions appropriées pour l'expression de la séquence codant pour la proinsuline humaine, et la purification de la proinsuline humaine à partir du lysat ou du milieu de culture du microorganisme.

L'invention fournit encore un procédé de synthèse de l'insuline humaine, caractérisé en ce qu'on oxyde des groupes sulfhydryle sur les peptides A et B de la proinsuline humaine, préparée tel que décrit ici, pour former des ponts disulfure entre les peptides A et B, puis on excise le peptide C par une

hydrolyse catalysée par une enzyme, spécifique pour les liaisons réunissant le peptide C aux peptides A et B.

L'invention fournit les éléments génétiques essentiels, nécessaires à la production d'insuline humaine par des techniques adaptables aux procédés industriels. On isole et purifie le gène structural naturel. Son expression dans un type de cellule hôte approprié donne un produit protéinique convertissable en insuline par des procédés connus. L'invention est fondamentalement basée sur la molécule de proinsuline et tire avantage du fait que la région du peptide C de la molécule de proinsuline permet un pliage spontané tel que les chaînes A et B sont convenablement juxtaposées. Dans cette configuration, l'appariement correct des groupes sulfhydryle est assuré et la formation des ponts disulfure, tels que trouvé dans la molécule d'insuline active, est facile à réaliser. L'excision du peptide C est effectuée par l'utilisation combinée de la trypsine, ou d'une enzyme ayant une spécificité similaire pour un substrat, et de la carboxypeptidase C ou de la cathepsine C pour éliminer l'arginine C-terminale sur la chaîne B, tel que décrit dans la technique antérieure.

La totalité de la partie structurale codante du gène de l'insuline humaine est clonée tel que décrit ici, comprenant une partie de la région non-transférée 5', toute la séquence du prépeptide, la totalité de la séquence codant pour les peptides B, C et A et la totalité de la région non-transférée 3'. Le choix de la cellule hôte décide des portions du gène cloné utilisées. La totalité de la séquence codant pour la pré-proinsuline sera utilisable pour transformer certaines espèces de cellules d'eucaryotes supérieurs, car la préséquence agit comme un simple peptide pour promouvoir l'excrétion du peptide hors de la cellule. De plus, ces lignées de cellules d'eucaryotes capables de répondre au peptide signal, catalysent l'élimination spécifique de la préséquence pendant le transport de la protéine à travers la membrane cellulaire.

En outre, ces cellules peuvent être capables de transformer ultérieurement le produit d'expression par une excision de la protéine C après la formation catalytique des ponts disulfure appropriés.

5 Dans les cellules hôtes de procaryotes, l'utilisation de la séquence codant pour la proinsuline est préférée selon l'invention. L'opération de conversion de la proinsuline en insuline est conduite en détail selon des procédés connus dans la technique. Deux types d'expression sont déjà connus,
10 impliquant tous les deux l'insertion de la séquence codant pour la proinsuline dans une région du vecteur de transfert d'ADN dont l'expression a été contrôlée par un promoteur de procaryote. Dans certains cas, une partie du gène structural de procaryote et du contrôle du promoteur est interposée de
15 telle sorte que le produit d'expression est une protéine fusionnée ayant sa portion N-terminale composée de la partie N-terminale de la protéine de procaryote et sa portion C-terminale est la séquence clonée, dans ce cas la proinsuline. L'expression de la proinsuline comme protéine humaine présente
20 l'avantage que la protéine fusionnée peut être plus stable dans la cellule hôte que la proinsuline elle-même. De plus, lorsque la protéine d'eucaryote est celle qui est normalement excrétée de la cellule hôte, la protéine fusionnée peut également être excrétée, rendant le produit d'expression plus facile à purifier. L'emploi de protéines fusionnées présente
25 l'inconvénient que ces protéines doivent être scindées de façon spécifique pour obtenir le produit désiré. Dans le cas de la proinsuline, il existe des techniques qui tirent partie de la séquence d'acides aminés de la protéine pour scinder
30 spécifiquement une protéine fusionnée, tel que décrit dans l'exemple 2.

La séquence de codage, clonée, peut également être exprimée directement dans la cellule de procaryote en insérant la séquence de façon qu'elle soit directement adjacente à un
35 promoteur. Cette conception est avantageuse en ce qu'une scis-

sion spécifique est alors inutile. Mais elle présente l'inconvénient de nécessiter une autre purification.

L'expression des gènes de mammifères insérés dans *E.coli* a été maintenant obtenue par insertion près des promoteurs lac, trp et β -lactamase. Le promoteur lac est utilisable du fait que ses propriétés génétiques sont bien caractérisées. Il existe deux sites d'insertion possibles, fournissant des conducteurs de procaryotes longs et courts pour la protéine fusionnée. On dispose d'une grande variété de variants génétiques, ayant des niveaux différents de répresseur endogène, qui peuvent être induits par la chaleur, etc. Le promoteur trp a l'avantage de fournir de hauts niveaux d'expression lorsqu'il est induit. Des plasmides d'expression ayant des sites d'insertion dans le promoteur trp sont disponibles pour les trois cadres de lecture. Le promoteur de bêta-lactamase fournit ce qui revient à une protéine signal de procaryote, car la lactamase est normalement excrétée. Le produit de fusion de β -lactamase est également excrété ou trouvé dans l'espace périplasmique. En conséquence, il faut moins d'opérations de purification pour obtenir la proinsuline pure.

L'invention tire avantage du rôle du peptide C de la proinsuline, en fait pour faciliter le pliage spontané de la proinsuline qui amène les chaînes A et B dans une configuration correcte de telle sorte que les groupes sulfhydryle sont convenablement appariés et les ponts disulfure corrects sont formés, tel qu'on l'observe dans une insuline provenant de la nature. Une comparaison des séquences du peptide C de différentes espèces montre que les séquences de peptide C ne sont pas beaucoup conservées pendant l'évolution. De nombreuses substitutions de la séquence d'acides aminés sont donc possibles dans le peptide C. En fait, le rôle du peptide C peut consister simplement à fournir une boucle d'acides aminés de longueur appropriée pour permettre aux chaînes A et B de se replier ensemble dans la configuration convenable. En consé-

quence, pratiquement toute séquence de codage pourrait être insérée à la place de la séquence du peptide C de la proinsuline humaine sans modifier substantiellement sa fonction primaire. Mais certains avantages plaident en faveur de l'utilisation du peptide C naturel ; par exemple la séparation du peptide des préparations d'insuline n'a pas besoin d'être complète, car le peptide C est un constituant naturel des préparations d'insuline. De plus, il peut être avantageux de conférer la configuration appropriée sur l'insuline ou d'effectuer la séparation par des enzymes.

L'invention démontre en outre l'universalité du code génétique. En particulier, des codons favorisés par les cellules de mammifères sont correctement transférés dans E.coli. Des codons de remplacement, codant pour le même acide aminé, pourraient être substitués dans la séquence sans affecter la séquence ou la fonction de la protéine exprimée. L'invention concerne donc toute variance de codage synthétique de la séquence de codage de base réellement clonée ici, dans la mesure où une telle variance code pour la pré-proinsuline humaine et la proinsuline humaine.

Les exemples non limitatifs suivants sont donnés à titre d'illustration de l'invention.

EXEMPLE 1

Gène cloné de la pré-proinsuline humaine

On clone le gène de l'insuline humaine à partir d'ARN extrait d'un insulinoma humain selon le procédé décrit par Ullrich et al., ci-dessus, pour isoler l'ARN de cellules d'îlots. On utilise l'ARN obtenu sous forme d'un culot après centrifugation dans CsCl 5,7 M contenant 100 mM d'EDTA, sans fractionnement, comme modèle pour la synthèse du cADN en employant la transcriptase inverse et la S1 nucléase, tel que décrit par Ullrich et al., ci-dessus. A partir de 130 µg d'ARN non-fractionné, on obtient environ 40 ng de cADN à double chaîne.

On traite le cADN non-fractionné avec la transférase terminale en présence de dCTP pour obtenir des queues oligo-C sur les terminaisons 3' des molécules de cADN. De même, on scinde le vecteur de transfert plasmique pBR322 au moyen de l'endonucléase de restriction Pst I et on traite par une transférase terminale en présence de dGTP pour produire des queues oligo-dG sur les terminaisons 3' de l'ADN plasmique linéaire. L'ADN plasmique ainsi traité ne peut pas former un ADN circulaire, car les extrémités des chaînes simples ne sont pas complémentaires. Toutefois, les extrémités oligo-dC du cADN traité sont complémentaires avec les extrémités oligo-dG du plasmide et peuvent donc être utilisées pour former des plasmides circulaires dans lesquels le cADN est inséré au site Pst I. De plus, l'insertion régénère les deux extrémités de la séquence de reconnaissance Pst I, aux deux extrémités de l'insert, permettant ainsi l'excision des séquences insérées (Voir Villa-Komaroff et al., ci-dessus).

On utilise le procédé de formation de queues ci-dessus pour produire des plasmides pBR-322 ayant une population hétérogène d'inserts de cADN au site Pst I. Ces plasmides seraient sensibles à l'ampicilline, car le site Pst I est dans le gène de la β -lactamase, mais ils restent résistants à la tétracycline. On transforme E.coli, souche X1776, par l'ADN plasmique contenant les inserts. On obtient 525 transformants résistants à la tétracycline. On procède à une reproduction sur plaque et on sélectionne les séquences d'insuline par une hybridation des colonies in situ, essentiellement tel que décrit par Gruenstein et Hogness, Proc. Nat. Acad. Sci., USA 72, 3961 (1975). On utilise le cADN de la pré-proinsuline de rat, préalablement isolé et purifié (Ullrich et al., ci-dessus) comme modèle d'hybridation, après avoir marqué le cADN de rat par une translation en utilisant l'ADN polymérase I. Etant donné que les séquences d'acides aminés de l'insuline de rat et de l'insuline humaine sont tout à fait semblables (insuline, 92 % d'homologie ; proinsuline, 83 % d'homologie),

on suppose que le cADN de rat isolé et purifié s'hybride de façon croisée avec les séquences de l'insuline humaine dans des conditions de rigueur réduite, connues dans la technique. Une autoradiographie révèle que 2 des 525 colonies s'hybrident avec le modèle de pré-proinsuline I de rat cloné. Ces deux colonies sont sensibles à l'ampicilline. Les plasmides isolés des colonies ont environ 250 paires de base et 500 paires de base de plus que le pBR322 lui-même. Le plasmide le plus long est désigné par l'abréviation pCHI-1.

On détermine la séquence de nucléotides du fragment de cADN inséré de pCHI-1 par le procédé de Maxam et Gilbert, Proc. Nat. Acad. Sci. USA 74, 560 (1977). Une représentation schématique de l'insert est illustrée sur la Fig. 1 des dessins annexés. Toute la région codant pour la pré-proinsuline humaine est contenue dans l'insert, avec une partie de la région 5' non-transférée, toute la région 3' non-transférée et une partie de la région 3'-polydA. Les sites de restriction utilisés pour scinder l'insert pour obtenir les séquences sont également montrés sur la Fig. 1., avec la direction de formation des séquences dans chaque fragment et les régions de chevauchement.

La séquence du mARN dérivée de la séquence du cADN, et la séquence d'acides aminés déduite de l'un des cadres de lecture sont représentées sur la Fig. 2 des dessins annexés. La structure primaire de la proinsuline humaine déterminée de cette façon concorde de façon précise avec celle obtenue par des expériences de formation de séquences d'acides aminés (Dayhoff M.D., Atlas of Protein Sequence and Structure, 5, Suppl. 2, pp. 127-130 (1976) et Suppl.3, pp. 150-151 (1978). Les acides aminés de la séquence codant pour la proinsuline sont numérotés de 1 à 86, ceux de la pré-séquence sont numérotés de -24 à -1. La Fig. 2 montre en outre l'emplacement de certains sites de restriction utilisables pour la construction du gène de la proinsuline à partir de l'insert pCHI-1. On comprendra que la séquence de cADN de la chaîne codante de

l'insert pCHI-1 est la même que celle représentée sur la Fig. 2 si ce n'est que la thymine (T) remplace l'uracile (U).

On prépare de grandes quantités de pCHI-1 et d'autres plasmides en transformant en outre E.coli HB-101. La souche
 5 HB-101 sert donc d'hôte approprié pour conserver et reproduire des plasmides ayant des inserts clonés, tel que décrit ici.

EXEMPLE 2

Construction du vecteur de transfert de la proinsuline

10 L'insert de cADN de pCHI-1 contient toute la séquence codant pour la pré-proinsuline humaine. Pour certaines applications, telles que le transfert à une autre espèce de cellule d'eucaryote, le processus normal et la séparation de la pré-séquence et du peptide C, et l'obtention d'une configura-
 15 tion convenablement repliée donnant une insuline active peuvent être nécessaires. Dans d'autres cas, un transfert à une cellule de procaryote telle qu'une bactérie, peut être nécessaire pour obtenir la protéine non-traitée. Dans cette dernière situation, la construction d'une séquence codant pour la
 20 proinsuline est avantageuse, car la proinsuline est facilement convertie en une insuline active in vitro, en utilisant des procédés connus dans la technique. Voir Kemmler W. et al, J. Biol. Chem. 242, 6786 (1971). On décrit ici trois procédés possibles pour construire une séquence codant pour la proin-
 25 suline. Un vecteur de transfert plasmique comprenant pBR322 avec une séquence codant pour la proinsuline, insérée, est désigné par pCHP-1.

A. Une séquence codant pour la chaîne B de l'insuline humaine, synthétisée par voie chimique, a été décrite par
 30 Crea R. et al. Proc. Nat. Acad. Sci. USA 75, 5765 (1978). La séquence de nucléotides, synthétique, diffère de la séquence naturelle décrite ici, car la séquence synthétique est conçue pour exploiter les attributions du codon plus favorisées par un hôte de procaryote, tel que E. coli. On incorpore donc un
 35 triplet codant pour la méthionine, juste avant l'acide aminé

1 (phe). Mais, par hasard, les deux séquences sont identiques dans la région des acides aminés 13-14, laquelle région ne contient que le site Alu I commun aux deux séquences. Les emplacements des sites Alu I sur la séquence naturelle sont
 5 indiqués sur la Fig.2.

On traite l'ADN de proinsuline synthétique par l'Alu I endonucléase pour obtenir deux fragments de 43 paires de base et environ 56 paires de base, respectivement. On scinde de même l'insert de cADN de pCHI-1, obtenu de préférence par
 10 scission en Hha I tel que décrit dans l'exemple 2B, par une hydrolyse partielle catalysée par l'Alu I endonucléase pour former des fragments de 75, 90, 110, 135, 165, 200, 245, 275, 375 et 455 paires de base, respectivement ; on fractionne ces fragments par électrophorèse sur gel pour obtenir le frag-
 15 ment de 375 paires de base, le résultat d'une scission en un seul site dans le codon pour l'acide aminé numéro 14.

On réunit par une ligature des extrémités tronquées les fragments de scission du gène synthétique et le fragment de scission à un seul site, de 375 paires de base, de l'insert de cADN. Une réunion correcte du fragment synthétique à
 20 43 paires de base avec le fragment de cADN à 375 paires de base est maximisée en assurant un excès molaire du cADN à 375 paires de base. Les possibilités d'une réunion incorrecte sont également réduites du fait que les fragments synthétiques
 25 ont des extrémités saillantes à simple chaîne qui ne sont pas complémentaires avec celles du fragment de cADN à 375 paires de base.

La molécule ainsi formée, un composite de la séquence synthétique codant pour la méthionine suivie des acides aminés 1 à 13, et de la séquence naturelle codant pour les acides aminés 14 à 86 de la proinsuline, constitue une séquence
 30 codant pour la proinsuline. La séquence codant pour la proinsuline peut être insérée dans n'importe quel plasmide d'expression choisi, par remplissage ou excision des extrémités
 35 à simple chaîne, puis fixation des oligonucléotides de liaison

appropriés.

L'expression donne une protéine fusionnée qui peut être scindée au résidu méthionine par traitement au bromure de cyanogène pour obtenir la proinsuline. La proinsuline est
 5 ensuite convertie en insuline par le procédé de Kemmler et al., ci-dessus.

B. L'insert de cADN de pCHI-1 a un site Hha I dans la séquence codant pour les acides aminés -14 à -13, tel que représenté sur la Fig. 2. De plus, le vecteur de transfert
 10 pBR322 a un site Hha I juste à 22 paires de base du site Pst I à l'extrémité 5' de l'insert. Il est donc possible d'isoler de nouveau une séquence comprenant la totalité de la séquence codant pour la proinsuline et une région de 22 paires de base de ADN pBR322, en traitant pCHI-1 par l'Hha I endonucléase.
 15 On préfère ce procédé à un nouvel isolement de l'insert avec la Pst I endonucléase, car l'insert contient deux sites Pst I internes et le rendement en ADN d'insert intact par un traitement à la Pst I endonucléase est faible. La séquence Hha I isolée est également parfaitement utilisable dans les procé-
 20 dés des exemples 2A et 2C décrits ici.

Le traitement de l'un ou l'autre des isolats par l'Hha I endonucléase se traduit par une scission de la chaîne plus entre les acides aminés -14 et -13 de la pré-séquence. (La chaîne plus est définie comme la chaîne dont la séquence
 25 de nucléotides correspond à la séquence du mARN. La chaîne moins est la chaîne dont la séquence est complémentaire de la séquence du mARN). La pré-séquence restante peut être spécifiquement séparée en exploitant l'activité 3' à 5' exonucléase de la T4 ADN polymérase qui agit sur la chaîne moins
 30 à l'extrémité de codage de la pré-séquence et sur la chaîne plus à l'extrémité opposée. La réaction exonucléolytique peut être arrêtée à un nucléotide défini en mettant le même nucléotide, sous la forme triphosphate, dans le mélange réactionnel. L'action exonucléolytique est alors contrecarrée
 35 par l'activité de polymérase de l'enzyme qui remplace conti-

nuellement le nucléotide spécifique, tel que décrit par Englund P.T. dans J. Biol. Chem. 246, 3269 (1971) et dans J. Mol. Biol. 66, 209 (1972).

La pré-séquence restante peut être digérée spécifiquement jusqu'au codon de phénylalanine N-terminal de l'acide aminé numéro 1 par trois cycles de digestion par la T4 polymérase. Le cycle 1 est effectué en présence de TTP qui termine la digestion à l'opposé du A du codon de glycine dans l'acide aminé en position -7. Le cycle 2 est effectué en présence de dCTP qui termine la digestion à l'opposé du G du codon en position -5 (Asp). Le cycle 3 est effectué en présence de dATP qui termine la digestion à l'opposé du T qui est le premier nucléotide de position 1 de la proinsuline.

Après chaque cycle de digestion par la T4 polymérase, le triphosphate du cycle juste complété doit être éliminé avant d'introduire le triphosphate nécessaire pour le cycle suivant. Pour séparer les triphosphates du mélange réactionnel, on utilise des minicolonnes de Sephadex G-75 (Marque déposée, Pharmacia Inc., Uppsala, Suède). Les colonnes peuvent être équilibrées avec un tampon approprié et des échantillons recueillis dans un tampon contenant le triphosphate du cycle suivant. De cette façon, l'enzyme migrant avec l'ADN ne digère pas l'ADN au-delà du prochain point d'arrêt choisi. Pour éviter une telle digestion dans la partie inférieure de la colonne après la résolution du triphosphate précédent, on effectue une chromatographie à froid (4 °C) et on active l'élution par centrifugation. On peut également inactiver l'enzyme par la chaleur, à 65 °C, avant l'opération de chromatographie. On amorce alors le cycle suivant par addition d'enzyme active, fraîche. Les trois cycles de digestion par la T4 polymérase fournissent une digestion complète et spécifique de la chaîne moins d'ADN jusqu'au premier codon de la séquence codant pour la proinsuline. La séquence de la chaîne plus à l'extrémité opposée de la molécule qui a également une extrémité 3', est telle que seuls quelques nucléotides sont

éliminés par les cycles de digestion ci-dessus.

La chaîne plus à l'extrémité de la pré-séquence est alors spécifiquement digérée avec la S1 exonucléase qui agit sur les extrémités à simple chaîne pour donner une molécule aux extrémités tronquées. Des précautions doivent être prises pour éviter une digestion partielle par la S1 nucléase au-delà du début du premier codon. Une telle digestion partielle peut avoir lieu en raison d'une "respiration" de l'hélice, un non-appariement partiel et provisoire des chaînes d'ADN. La "respiration" a lieu dans toute la molécule, mais le plus fréquemment dans les régions riches en A-T et aux extrémités des molécules d'ADN. Un non-appariement provisoire au codon numéro 1, riche en A-T, pourrait permettre une action S1 nucléolytique au-delà du point d'arrêt fixé. Cet inconvénient peut être évité en effectuant la digestion S1 dans des conditions d'intégrité maximale de l'hélice ; faible température (température ambiante ou plus basse) et haute teneur en sel normalement utilisé dans le tampon pour une réaction S1.

Des échantillons d'ADN obtenus aux stades successifs de la digestion par la S1 nucléase sont isolés et purifiés dans un vecteur de transfert approprié selon des procédés connus dans la technique. On procède à une analyse de la séquence du plus petit fragment Alu I de ces clones pour sélectionner les clones codant pour la proinsuline, dont la pré-séquence a été totalement séparée.

Un procédé attrayant pour isoler et purifier le cADN codant pour la proinsuline implique l'incorporation de queues oligo-A, en utilisant une transférase terminale, sur les extrémités 3' du cADN. Les queues oligo-T sont formées sur l'extrémité 3' en un site approprié sur un vecteur de transfert d'expression, tel que le site EcoRI dans le gène de β -galactosidase sur le plasmide pTS-1 (Ullrich A. et al.) (Voir également la demande de brevet US n° 933 035). L'insertion du cADN codant pour la proinsuline par le procédé ci-dessus fournit un insert correctement orienté, en phase quant au cadre de

lecture avec le gène de la β -galactidase du plasmide selon une probabilité de 1/6. L'expression des queues oligo-A se traduit par l'incorporation des résidus de lysine juste avant le début de la séquence de proinsuline. Une digestion modérée
 5 de la protéine fusionnée par la trypsine fournit la proinsuline qui est convertie en insuline active tel que décrit précédemment. Selon un autre procédé, la protéine fusionnée est traitée par une combinaison de trypsine et de carboxypeptidase B (ou de cathepsine B) pour donner une insuline active à
 10 partir de la protéine fusionnée en une seule réaction.

C. On construit une séquence codant pour la proinsuline par scission sélective à un site interne dans la région codant pour la proinsuline, puis on fixe une séquence synthétisée par voie chimique, codant pour cette partie de la ré-
 15 gion codant pour la proinsuline éliminée par la scission précédente. Comme source de région codant pour la proinsuline, on utilise le plasmide pCHI-1 qui est sélectivement excisé par traitement avec la Pst I endonucléase ou de préférence par traitement avec la Hha I endonucléase, tel que décrit dans
 20 l'exemple 2B.

Chaque fragment, après isolement, est traité par une phosphatase alcaline pour éliminer les groupes phosphate 5' terminaux, puis on scinde par traitement avec une endonucléase de restriction ayant un point de scission unique dans la sé-
 25 quence codant pour la proinsuline. Le site de restriction est de préférence proche d'une des extrémités de la séquence codant pour la proinsuline. Le site Alu I dans la région des acides aminés 13-14 fournit un point de scission approprié, (voir Fig.2). Le fragment Hha I de pCHI-1 est partiellement
 30 scindé par l'Alu I endonucléase pour fournir deux fragments d'environ 76 paires de base et 375 paires de base, respectivement. Les fragments Alu I sont fractionnés par électrophorèse sur gel, tel que décrit dans l'exemple 2A, et le fragment à 375 paires de base est recueilli.

35 On applique le procédé au phosphotriester (Itakura K.

et al., J. Biol. Chem. 250, 4592 (1975) et Itakura K. et al., J. Am. Chem. Soc. 97; 7327 (1975)) pour synthétiser une séquence de nucléotides codant pour les 13 premiers acides aminés de la proinsuline avec un G terminal 5' (sur la chaîne plus), pour compléter le codon pour l'alanine en position 14. La chaîne plus de l'ADN synthétique a la séquence 3'-TTTGTGAA-CCAACACCTGTGCGGCTCACACCTGGTGAAG-5' correspondant à la séquence naturelle. Cependant, il est possible de synthétiser d'autres séquences codant pour les mêmes acides aminés. De façon générale, la séquence est : 3'-TTKGTAAKCAJCAKXTYTGKGLQRSCAKXTYGTLCAJG-5'. La séquence obtenue est réunie par ses extrémités tronquées au fragment d'environ 375 paires de base du fragment Hha I de pCHI-1. Etant donné que ce dernier a un phosphate 5' uniquement à l'extrémité à réunir, les deux fragments sont réunis dans un ordre correct. Le fragment synthétique est correctement réuni au plus grand fragment dans environ 50 % des réactions.

L'ADN traité par une ligase est alors cloné et inséré dans un plasmide d'expression approprié, soit en formant une queue oligo-A, tel que décrit dans l'exemple 2B, soit en fixant des agents de liaison et en insérant dans des plasmides de cadres de lecture connus. Dans le cas d'inserts à queue oligo-A, on observe l'expression de la proinsuline dans environ 1/12 des clones. Dans le cas d'une insertion directe, lorsque le cadre de lecture est correct, la fréquence d'expression des clones est d'environ 50 %.

EXEMPLE 3

Expression de la pré-proinsuline et de la proinsuline humaines

Les inserts isolés et purifiés codant pour la pré-proinsuline (exemple 1) ou la proinsuline (exemple 2) sont insérés dans un vecteur de transfert d'expression. Lorsque l'insertion a lieu dans l'orientation correcte eu égard à l'initiation du transfert au site d'insertion, et que l'insert est en phase du cadre de lecture avec le promoteur et le site de fixation du ribosome, le produit protéinique du gène cloné

est synthétisé par les cellules hôtes de grande activité métabolique, transformées par le vecteur de transfert. Le produit protéinique est une protéine fusionnée si le vecteur de transfert d'expression contient une partie du gène de procar-

5 ryote entre le promoteur et le site d'insertion. Mais l'insertion peut être immédiatement adjacente à un site de promoteur. Dans de tels cas, la protéine codée par l'insert est synthétisée directement. Les deux techniques présentent des avantages et des inconvénients. Les protéines fusionnées ont l'a-

10 vantage d'avoir tendance à stabiliser la protéine étrangère codée par le gène inséré. Des propriétés fonctionnelles désirables, telles que l'excrétion de la cellule hôte, sont également conférées par la fusion avec certaines protéines hôtes telles que la β -lactamase. D'autre part, la séquence codée

15 par l'insert est difficile à purifier, étant donné le désir général d'éliminer spécifiquement la portion hôte de la protéine fusionnée. Cette élimination se fait par des techniques connues telles que décrites dans les exemples 2A et 2B. La synthèse directe de la protéine désirée supprime la nécessité

20 d'une scission spécifique, mais élimine la possibilité d'excrétion hors de la cellule hôte.

On a développé des plasmides d'expression dans lesquels l'expression est réglée par le promoteur lac (Itakura et al., Science 198, 1056 (1977), Ullrich A. et al., Excerpta Medica (1979)), par le promoteur trp (Martial et al., Science 205, 602 (1979)); et par le promoteur de β -lactamase, (demande de brevet US n° 44 647).

L'expression est décelée par la mesure d'un produit capable de s'associer par voie immunochimique avec un anti-

30 corps anti-insuline ou un anticorps anti-proinsuline. On a recours à un dosage radioimmunologique dans lequel l'anticorps est marqué de façon radioactive et des paires antigène-anticorps sont précipitées par une préparation de Staphylococcus aureus tués par la chaleur. (Voir Morgan et Lazarow, Diabetes, 12, 115 (1963) et Kessler S.W., J. Immunol. 115, 1617 (1975)).

Pour déceler l'expression dans des colonies de bactéries, on utilise le test de sélection radioimmun, tel que décrit par Erlich H.A. et al., Cell 10, 681 (1978) ou par Broome S. et Gilbert W., Proc. Nat. Acad. Sci. USA 75, 2746 (1978).

5 On décèle des protéines fusionnées indicatrices de l'expression en comparant les poids moléculaires de la protéine hôte contribuant à la partie N-terminale de la protéine fusionnée, dans les cellules hôtes transformées par des plasmides d'expression avec et sans insert. Une variante pré-
10 férée consiste à utiliser comme hôte E. coli, souche P678-54, produisant des minicellules. Des acides aminés marqués par un radioisotope sont incorporés dans des protéines à minicellules, en comparant les souches transformées avec des vecteurs de transfert d'expression avec et sans insertion de la séquen-
15 ce codant pour la proinsuline. Les protéines sont fractionnées par électrophorèse sur gel de SDS-acrylamide et les positions des protéines décelées par autoradiographie. L'expression de la proinsuline est indiquée par la présence d'une bande de protéine marquée, trouvée uniquement dans les mini-
20 cellules transformées par le plasmide d'expression de la proinsuline. La position de la bande d'électrophorèse fournit une mesure du poids moléculaire de la protéine exprimée et est conforme à la longueur connue du gène inséré et de la portion N-terminale du procaryote.

25 L'élimination de la portion de procaryote et la conversion de la proinsuline en insuline, in vitro, sont effectuées par des procédés connus, tel que décrit en détail ci-dessus.

REVEN DICATIONS

1 - Vecteur de transfert d'ADN comprenant une séquence de désoxynucléotides codant pour la pré-proinsuline humaine.

5 2 - Vecteur de transfert d'ADN comprenant une séquence de désoxynucléotides codant pour la proinsuline humaine.

3 - Microorganisme transformé par le vecteur de transfert suivant la revendication 1 ou 2.

10 4 - Vecteur de transfert d'ADN, caractérisé en ce qu'il comprend une séquence de désoxynucléotides pour la pré-proinsuline humaine comportant une chaîne plus ayant la séquence :

5'-TCCTTCTGCCATG₋₂₄GCL₋₂₃X₋₂₂TY₋₂₂TGG₋₂₁ATG₋₂₀W₋₁₉GZ₋₁₉X₋₁₈TY₋₁₈
 15 X₋₁₇TY₋₁₇CCL₋₁₆X₋₁₅TY₋₁₅X₋₁₄TY₋₁₄GCL₋₁₃X₋₁₂TY₋₁₂X₋₁₁TY₋₁₁
 GCL₋₁₀X₋₉TY₋₉TGG₋₈GGL₋₇CCL₋₆GAK₋₅CCL₋₄GCL₋₃GCL₋₂GCL₋₁TTK₁
 GTL₂AAK₃CAJ₄CAK₅X₆TY₆TGK₇GGL₈QR₉S₉CAK₁₀X₁₁TY₁₁GTL₁₂GAJ₁₃
 20 GCL₁₄X₁₅TY₁₅TAK₁₆X₁₇TY₁₇GTL₁₈TGK₁₉GCL₂₀GAJ₂₁W₂₂GZ₂₂GCL₂₃
 TTK₂₄TTK₂₅TAK₂₆ACL₂₇CCL₂₈AAJ₂₉ACL₃₀W₃₁GZ₃₁W₃₂GZ₃₂GAJ₃₃
 GCL₃₄GAJ₃₅GAK₃₆X₃₇TY₃₇CAJ₃₈GTL₃₉GGL₄₀CAJ₄₁GTL₄₂GAJ₄₃X₄₄
 TY₄₄GGL₄₅GGL₄₆GGL₄₇CCL₄₈GGL₄₉GCL₅₀GGL₅₁QR₅₂S₅₂X₅₃TY₅₃CAJ₅₄
 25 CCL₅₅X₅₆TY₅₆GCL₅₇X₅₈TY₅₈GAJ₅₉GGL₆₀QR₆₁S₆₁X₆₂TY₆₂CAJ₆₃AAJ₆₄
 W₆₅GZ₆₅GGL₆₆ATM₆₇GTL₆₈GAJ₆₉CAJ₇₀TGK₇₁TGK₇₂ACL₇₃QR₇₄S₇₄
 ATM₇₅TGK₇₆QR₇₇S₇₇X₇₈TY₇₈TAK₇₉CAJ₈₀X₈₁TY₈₁GAJ₈₂AAK₈₃TAK₈₄
 30 TGK₈₅AAK₈₆TAGACGCAGCCCGCAGGCAGCCCCCACC CGCCGCTCCTGCACCG
 AGAGAGATGGAATAAAGCCCTTGAACCAGC poly A-3'

où :

A est le radical désoxyadényle,

G est le radical désoxyguanyle,

- C est le radical désoxycytosyle,
 T est le radical thymidyle,
 J est A ou G ;
 K est T ou C ;
 5 L est A, T, C ou G ;
 M est A, C ou T ;
 X_n est T ou C si Y_n est A ou G , et C si Y_n est C ou T ;
 Y_n est A, G, C ou T si X_n est C, et A ou G si X_n est T ;
 W_n est C ou A si Z_n est G ou A, et C si Z_n est C ou T ;
 10 Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;
 QR_n est TC si S_n est A, G, C ou T et AG si S_n est T ou C ;
 S_n est A, G, C ou T si QR_n est TC et T ou C si QR_n est AG ;
 et les indices n indiquent la position dans la séquence d'acides aminés de la pré-proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotées à partir de l'extrémité amino.

- 5 - Vecteur de transfert d'ADN, caractérisé en ce qu'il comprend une séquence de désoxynucléotides codant pour
 20 la proinsuline humaine comportant une chaîne plus ayant la séquence :

5'-TTK₁GTL₂AAK₃CAJ₄CAK₅X₆TY₆TGK₇GGL₈QR₉S₉CAK₁₀X₁₁TY₁₁GTL₁₂GAJ₁₃
 GCL₁₄X₁₅TY₁₅TAK₁₆X₁₇TY₁₇GTL₁₈TGK₁₉GCL₂₀GAJ₂₁W₂₂GZ₂₂GCL₂₃
 25 TTK₂₄TTK₂₅TAK₂₆ACL₂₇CCL₂₈AAJ₂₉ACL₃₀W₃₁GZ₃₁W₃₂GZ₃₂GAJ₃₃
 GCL₃₄GAJ₃₅GAK₃₆X₃₇TY₃₇CAJ₃₈GTL₃₉GGL₄₀CAJ₄₁GTL₄₂GAJ₄₃X₄₄
 TY₄₄GGL₄₅GGL₄₆GGL₄₇CCL₄₈GGL₄₉GCL₅₀GGL₅₁QR₅₂S₅₂X₅₃TY₅₃CAJ₅₄
 CCL₅₅X₅₆TY₅₆GCL₅₇X₅₈TY₅₈GAJ₅₉GGL₆₀QR₆₁S₆₁X₆₂TY₆₂CAJ₆₃AAJ₆₄
 30 W₆₅GZ₆₅GGL₆₆ATM₆₇GTL₆₈GAJ₆₉CAJ₇₀TGK₇₁TGK₇₂ACL₇₃QR₇₄S₇₄
 ATM₇₅TGK₇₆QR₇₇S₇₇X₇₈TY₇₈TAK₇₉CAJ₈₀X₈₁TY₈₁GAJ₈₂AAK₈₃TAK₈₄
 TGK₈₅AAK₈₆TAG-3'

où :

- A est le radical désoxy adényle,
- G est le radical désoxyguanyle ,
- C est le radical désoxycytosyle ,
- 5 T est le radical thymidyle,
- J est A ou G ;
- K est T ou C ;
- L est A, T, C, ou G ;
- M est A, C ou T ;
- 10 X_n est T ou C si Y_n est A ou G ; et C si Y_n est C ou T ;
- Y_n est A, G, C ou T si X_n est C, et A ou G si X_n est T ;
- W_n est C ou A si Z_n est G ou A, et C si Z_n est C ou T ;
- Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;
- QR_n est TC si S_n est A, G, C ou T, et AG si S_n est T ou C ;
- 15 S_n est A, G, C ou T si QR_n est TC, et T ou C si QR_n est AG ;

et les indices n indiquent la position dans la séquence d'acides aminés de la proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotées à partir de l'extrémité amino.

6 - Microorganisme transformé par le vecteur de transfert selon la revendication 4 ou 5.

7 - Plasmide pCHI-1.

8 - Plasmide pCHP-1.

25 9 - Microorganisme transformé par le plasmide selon la revendication 7 ou 8.

10 - Microorganisme selon la revendication 9, caractérisé en ce que l'organisme est Escherichia coli.

30 11 - Microorganisme selon la revendication 10, caractérisé en ce que l'organisme est Escherichia coli HB-101.

12 - Protéine fusionnée, caractérisée en ce qu'elle comprend la séquence d'acides aminés de la pré-proinsuline humaine comme séquence C-terminale et une portion d'une protéine de procaryote comme séquence N-terminale.

13 - Protéine fusionnée, caractérisée en ce qu'elle comprend la séquence d'acides aminés de la proinsuline humaine comme séquence C-terminale et une portion d'une protéine de procaryote comme séquence N-terminale.

5 14 - Procédé de préparation d'une séquence de désoxynucléotides codant pour la proinsuline humaine, caractérisé en ce que

(1) on traite un fragment de cADN comprenant une séquence de nucléotides codant pour la pré-proinsuline humaine avec une endonucléase de restriction ayant un site de restriction dans la région codant pour le pré-peptide de la pré-proinsuline ;

(2) on traite la séquence de désoxynucléotides du stade (1) contenant la séquence codant pour la proinsuline avec la T4 ADN polymérase en présence d'un désoxynucléotide triphosphate, de telle sorte que la digestion exonucléolytique à partir des extrémités 3' ait lieu jusqu'à un point d'arrêt défini par le désoxynucléotide triphosphate ;

(3) on remplace le désoxynucléotide triphosphate par un désoxynucléotide triphosphate différent, de telle sorte que la digestion procède jusqu'à un second point d'arrêt défini ;

(4) on renouvelle l'opération du stade (3), si nécessaire, pour effectuer la digestion de l'extrémité 3' jusqu'au premier nucléotide de la séquence codant pour la proinsuline, la digestion finale étant conduite en présence de dATP ; et

(5) on traite le produit de désoxynucléotide obtenu au stade (4), par la S1 nucléase pour digérer l'extrémité 5' jusqu'au premier nucléotide de la séquence codant pour la proinsuline.

15 - Procédé selon la revendication 14, caractérisé en ce que l'endonucléase de restriction est Hha I.

16 - Procédé de préparation de la séquence de déso-

xynucléotides codant pour la proinsuline humaine, caractérisé en ce que :

(1) on scinde en deux fragments une séquence de désoxynucléotides codant pour la chaîne B de l'insuline humaine et ayant au moins un site de restriction en commun avec le cADN de la proinsuline humaine, la scission étant effectuée par une hydrolyse catalysée par une endonucléase de restriction reconnaissant le site de restriction commun ;

(2) on scinde un cADN codant pour la pré-proinsuline humaine en deux fragments, la scission étant effectuée par une hydrolyse catalysée par la même endonucléase de restriction que celle utilisée au stade (1) ; et

(3) on réunit par une liaison des extrémités tronquées les fragments obtenus aux stades (1) et (2), qui contiennent ensemble la séquence codant pour la proinsuline humaine.

17 - Procédé selon la revendication 16, caractérisé en ce que l'endonucléase de restriction est Alu I.

18 - Procédé de préparation de la séquence de désoxynucléotides codant pour la proinsuline humaine, caractérisé en ce que :

(1) on élimine une partie de la séquence de désoxynucléotides codant pour la proinsuline humaine, en traitant un cADN codant pour la pré-proinsuline humaine par une endonucléase de restriction ayant un site de reconnaissance dans la séquence codant pour la proinsuline ;

(2) on synthétise une séquence de désoxynucléotides capable de remplacer la partie éliminée et codant pour les mêmes acides aminés codés par la partie éliminée ; et

(3) on réunit les séquences de désoxynucléotides obtenues aux stades (1) et (2) en liant les extrémités tronquées.

19 - Procédé selon la revendication 18, caractérisé en ce que l'endonucléase de restriction utilisée au stade (1) est Alu I.

20 - Procédé de préparation d'insuline, caractérisé en ce que :

(1) on construit un microorganisme ayant une séquence codant pour un peptide comprenant : (a) une séquence de désoxynucléotides $(N_a N_b N_c)_j$, ayant une extrémité 5' et une extrémité 3', dans laquelle N_a , N_b et N_c peuvent être A, T, G ou C et j est un nombre entier de 0 à 100, $N_a N_b N_c$ n'étant pas TGA, TAA ou TAG, (b) une séquence de désoxynucléotides codant pour la chaîne B de l'insuline humaine fixée à l'extrémité 5' de $(N_a N_b N_c)_j$; et (c) une séquence de désoxynucléotides codant pour la chaîne A de l'insuline humaine fixée à l'extrémité 3' de $(N_a N_b N_c)_j$;

(2) on fait proliférer le microorganisme dans des conditions appropriées pour l'expression d'une séquence de désoxynucléotides codant pour le dit peptide, afin de produire le peptide ayant la chaîne B de l'insuline humaine fixée à sa terminaison N et la chaîne A de l'insuline humaine fixée à sa terminaison C ;

(3) on purifie le peptide avec les chaînes A et B fixées ;

(4) on traite le peptide purifié pour former des ponts disulfure, tel que trouvés dans l'insuline, entre les chaînes A et B ;

(5) on traite le peptide purifié réticulé avec une préparation d'enzyme capable d'exciser les chaînes A et B du peptide, les chaînes A et B excisées ayant les ponts disulfure, formant ainsi l'insuline ; et

(6) on purifie l'insuline produite au stade (5).

21 - Procédé selon la revendication 20, caractérisé en ce que la séquence de désoxynucléotides codant pour la chaîne B de l'insuline humaine comprend une chaîne plus ayant la séquence : 5'-TTK₁GTL₂AAK₃CAJ₄CAK₅X₆TY₆TGK₇GGL₈QR₉S₉

CAK₁₀X₁₁TY₁₁GTL₁₂GAJ₁₃GCL₁₄X₁₅TY₁₅TAK₁₆X₁₇TY₁₇GTL₁₈TGK₁₉

GCL₂₀GAJ₂₁W₂₂GZ₂₂GCL₂₃TTK₂₄TTK₂₅TAK₂₆ACL₂₇CCL₂₈AAJ₂₉ACL₃₀-3'

et la séquence de désoxynucléotides codant pour la chaîne
 A de l'insuline humaine comprend une chaîne "plus" ayant la
 séquence : 5'-GGL₆₆ATM₆₇GTL₆₈GAJ₆₉CAJ₇₀TGK₇₁TGK₇₂ACL₇₃QR₇₄
 S₇₄ATM₇₅TGK₇₆QR₇₇S₇₇X₇₈TY₇₈TAK₇₉CAJ₈₀X₈₁TY₈₁CAJ₈₂AAK₈₃
 5 TAK₈₄TGK₈₅AAK₈₆-3'

où :

A est le radical désoxyadényle,
 G est le radical désoxyguanyle,
 C est le radical désoxycytosyle,

10 T est le radical thymidyle,

J est A ou G ;

K est T ou C ;

L est A, T, C ou G ;

M est A, C ou T ;

15 X_n est T ou C si Y_n est A ou G , et C si Y_n est C ou T ;

Y_n est A, G, C ou T si X_n est C, et A ou G si X_n est T ;

W_n est C ou A si Z_n est G ou A, et C si Z_n est C ou T ;

Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;

QR_n est TC si S_n est A, G, C ou T, et AG si S_n est T ou C ;

20 S_n est A, G, C ou T si QR_n est TC, et T ou C si QR_n est AG ;

et les indices n indiquent la position dans la séquence d'acides aminés de la proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotés à partir de l'extrémité amino.
 25

22 - Procédé de préparation d'un vecteur de transfert d'ADN utilisable pour conserver et reproduire une séquence de désoxynucléotides codant pour la pré-proinsuline humaine caractérisé en ce qu'on fait réagir une séquence de désoxy-
 30 nucléotides codant pour la pré-proinsuline humaine avec une molécule d'ADN préparée en scindant un vecteur de transfert par une enzyme de restriction.

23 - Procédé selon la revendication 22, caractérisé en

ce que la séquence de désoxynucléotides codant pour la pré-proinsuline humaine comprend une chaîne plus ayant la séquence :

5 5'-ATG₋₂₄GCL₋₂₃X₋₂₂TY₋₂₂TGG₋₂₁ATG₋₂₀W₋₁₉GZ₋₁₉X₋₁₉TY₋₁₈
 X₋₁₇TY₋₁₇CCL₋₁₆X₋₁₅TY₋₁₅X₋₁₄TY₋₁₄GCL₋₁₃X₋₁₂TY₋₁₂X₋₁₁TY₋₁₁
 GCL₋₁₀X₋₉TY₋₉TGG₋₈GGL₋₇CCL₋₆GAK₋₅CCL₋₄GCL₋₃GCL₋₂GCL₋₁TTK₋₁
 GTL₂AAK₃CAJ₄CAK₅X₆TY₆TGK₇GGL₈QR₉S₉CAK₁₀X₁₁TY₁₁GTL₁₂GAJ₁₃
 10 GCL₁₄X₁₅TY₁₅TAK₁₆X₁₇TY₁₇GTL₁₈TGK₁₉GCL₂₀GAJ₂₁W₂₂GZ₂₂GCL₂₃
 TTK₂₄TTK₂₅TAK₂₆ACL₂₇CCL₂₈AAJ₂₉ACL₃₀W₃₁GZ₃₁W₃₂GZ₃₂GAJ₃₃
 GCL₃₄GAJ₃₅GAK₃₆X₃₇TY₃₇CAJ₃₈GTL₃₉GGL₄₀CAJ₄₁GTL₄₂GAJ₄₃X₄₄
 TY₄₄GGL₄₅GGL₄₆GGL₄₇CCL₄₈GGL₄₉GCL₅₀GGL₅₁QR₅₂S₅₂X₅₃TY₅₃CAJ₅₄
 15 CCL₅₅X₅₆TY₅₆GCL₅₇X₅₈TY₅₈GAJ₅₉GGL₆₀QR₆₁S₆₁X₆₂TY₆₂CAJ₆₃AAJ₆₄
 W₆₅GZ₆₅GGL₆₆ATM₆₇GTL₆₈GAJ₆₈CAJ₇₀TGK₇₁TGK₇₂ACL₇₃QR₇₄S₇₄
 ATM₇₅TGK₇₆QR₇₇S₇₇X₇₈TY₇₈TAK₇₉CAJ₈₀X₈₁TY₈₁GAJ₈₂AAK₈₃TAK₈₄
 TGK₈₅AAK₈₆TAG-3'

20

où

A est le radical désoxyadényle,

G est le radical désoxyguanyle,

C est le radical désoxycytosyle,

25 T est le radical thymidyle

J est A ou G ;

K est T ou C ;

L est A, T, C ou G ;

M est A, C ou T ;

30 X_n est T ou C si Y_n est A ou G , et C si Y_n est C ou T ;Y_n est A, G, C ou T si X_n est C, et A ou G si X_n est T ;W_n est C ou A si Z_n est G ou A, et C si X_n est C ou T ;Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;QR_n est TC si S_n est A, G, C ou T, et AG si S_n est T ou C ;35 S_n est A, G, C ou T si QR_n est TC, et T ou C si QR_n est AG ;

et les indices n indiquent la position dans la séquence d'acides aminés de la proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotées à partir de l'extrémité amino.

24 - Procédé de préparation d'une séquence de désoxynucléotides codant pour la proinsuline humaine, utilisable dans la synthèse de l'insuline humaine, caractérisé en ce qu'on hydrolyse les extrémités 3' d'une séquence de désoxynucléotides codant pour la pré-proinsuline humaine, conservée et reproduite selon le procédé de la revendication 22, par un ou plusieurs cycles d'une digestion exonucléolytique progressive, catalysée par la T4 ADN polymérase agissant en présence d'un désoxynucléotide triphosphate, puis on hydrolyse les extrémités 5' de la séquence de désoxynucléotides par une digestion nucléolytique catalysée par une enzyme du nucléase, spécifique pour l'ADN à une seule chaîne ayant une extrémité 5'.

25 - Procédé selon la revendication 24, caractérisé en ce que la séquence codant pour la préproinsuline humaine est hydrolysée en un site spécifique dans une réaction catalysée par la HhaI endonucléase avant la digestion par la T4 ADN polymérase.

26 - Procédé de préparation d'une séquence de désoxynucléotides codant pour la proinsuline humaine, utilisable dans la synthèse de l'insuline humaine, caractérisée en ce qu'on fait réagir un fragment de cADN codant pour une partie de la proinsuline humaine, préparée par scission d'un cADN codant pour la préproinsuline humaine, conservé et reproduit selon le procédé de la revendication 22, avec une endonucléase de restriction ayant un site de reconnaissance unique dans sa séquence de désoxynucléotides, avec un ADN synthétisé par voie chimique dont la séquence de désoxynucléotides code pour la partie de la proinsuline humaine non-codée par le fragment de cADN, dans une réaction de liaison des extrémités tronquées, catalysée par l'ADN ligase.

27 - Procédé selon la revendication 26, caractérisé en ce que l'endonucléase de restriction est Alu I.

28 - Procédé selon l'une quelconque des revendications 24 à 27, caractérisé en ce que la séquence de désoxy-nucléotides codant pour la proinsuline humaine comprend une chaîne "plus" ayant la séquence :

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

20 où :

A est le radical désoxyadényle,
G est le radical désoxyguanyle,
C est le radical désoxycytosyle,
T est le radical thymidyle,

25 J est A ou G ;

K est T ou C ;

L est A, T, C ou G ;

M est A, C ou T ;

X_n est T ou C si Y_n est A ou G , et C si Y_n est C ou T ;

30 Y_n est A, G, C ou T si X est C, et A ou G si X est T ;

W_n est C ou A si Z_n est G ou A, et C si Z_n est C ou T ;

Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;

QR_n est TC si S_n est A, G, C ou T, et AG si S_n est T ou C ;

S_n est A, G, C ou T si QR_n est TC, et T ou C si QR_n est AG ;

et les indices n indiquent la position dans la séquence d'acides aminés de la proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotées à partir de l'extrémité amino.

29 - Procédé de préparation du plasmide pCHI-1, selon l'une quelconque des revendications 22 ou 23, caractérisé en ce que la molécule d'ADN préparée par scission d'un vecteur de transfert avec une enzyme de restriction est pBR322 et l'enzyme de restriction est la Pst I endonucléase.

30 - Procédé de préparation d'un vecteur de transfert d'ADN, utilisable pour conserver et reproduire une séquence de désoxynucléotides codant pour la proinsuline humaine, caractérisé en ce qu'on fait réagir une séquence de désoxynucléotides codant pour la proinsuline humaine, préparée selon l'une quelconque des revendications 24 à 28, avec une molécule d'ADN préparée en scindant un vecteur de transfert par une enzyme de restriction.

31 - Procédé selon la revendication 30 pour la préparation du plasmide pCHP-1, caractérisé en ce que la molécule d'ADN préparée par scission d'un vecteur de transfert par une enzyme de restriction est pBR322 et l'enzyme de restriction est la Pst I endonucléase.

32 - Procédé de préparation d'un vecteur de transfert d'expression selon la revendication 22 ou 30, utilisable dans l'expression d'une séquence de désoxynucléotides codant pour la pré-proinsuline humaine ou la proinsuline humaine, caractérisé en ce que la molécule d'ADN préparée par scission d'un vecteur de transfert par une enzyme de restriction est scindée en un point situé dans la région de contrôle de l'expression.

33 - Procédé de préparation d'une protéine fusionnée, comprenant la séquence d'acides aminés de la pré-proinsuline humaine comme séquence C-terminale et une partie d'une protéine de procaryote comme séquence N-terminale, caractérisé en ce qu'on incube un microorganisme transformé par un vecteur de

transfert d'expression comprenant une séquence de désoxynuc-
cléotides pour la pré-proinsuline humaine, préparé par le
procédé selon la revendication 22 ou 23.

34 - Procédé de préparation d'une protéine fusionnée
5 comprenant la séquence d'acides aminés de la proinsuline hu-
maine comme séquence C-terminale et une partie d'une protéine
de procaryote comme séquence N-terminale, caractérisé en ce
qu'on incube un microorganisme transformé par un vecteur de
transfert d'expression comprenant une séquence de désoxynu-
10 cléotides codant pour la proinsuline humaine, préparé par le
procédé selon la revendication 30.

35 - Procédé de synthèse de la proinsuline humaine
comprenant des peptides A et B séparés l'un de l'autre par
un peptide C, caractérisé en ce qu'on incube un microorganis-
15 me, transformé par un vecteur de transfert d'expression com-
prenant une séquence de désoxynuccléotides codant pour la pro-
insuline humaine, préparé selon l'une quelconque des revendi-
cations 24 à 28, dans des conditions appropriées pour l'ex-
pression de la séquence codant pour la proinsuline humaine,
20 et on purifie la proinsuline humaine à partir du lysat ou du
milieu de culture du microorganisme.

36 - Procédé selon la revendication 35, caractérisé
en ce que la séquence de désoxynuccléotides codant pour la
chaîne B de l'insuline humaine comprend une chaîne plus ayant
25 la séquence : 5'-TTK₁GTL₂AAK₃CAJ₄CAK₅X₆TY₆TGK₇GGL₈QR₉S₉
CAK₁₀X₁₁TY₁₁GTL₁₂GAJ₁₃GCL₁₄X₁₅TY₁₅TAK₁₆X₁₇TY₁₇GTL₁₈TGK₁₉
GCL₂₀GAJ₂₁W₂₂GZ₂₂GCL₂₃TTK₂₄TTK₂₅TAK₂₆ACL₂₇CCL₂₈AAJ₂₉ACL₃₀-3'
et la séquence de désoxynuccléotides codant pour la chaîne A
30 comprend une chaîne plus ayant la séquence : 5'-GGL₆₆ATM₆₇
GTL₆₈GAJ₆₉CAJ₇₀TGK₇₁TGK₇₂ACL₇₃QR₇₄S₇₄ATM₇₅TGK₇₆QR₇₇S₇₇X₇₈TY₇₈
TAK₇₉CAJ₈₀X₈₁TY₈₁CAJ₈₂AAK₈₃TAK₈₄TGK₈₅AAK₈₆-3'
où :

A est le radical désoxyadényle,

G est le radical désoxyguanyle,

C est le radical désoxycytosyle,

T est le radical thymidyle,

J est A ou G ;

5 K est T ou C ;

L est A, T, C ou G ;

M est A, C ou T ;

X_n est T ou C si Y_n est A ou G , et C si Y_n est C ou T ;

Y_n est A, G, C ou T si X_n est C, et A ou G si X_n est T ;

10 W_n est C ou A si Z_n est G ou A, et C si Z_n est C ou T ;

Z_n est A, G, C ou T si W_n est C, et A ou G si W_n est A ;

QR_n est TC si S_n est A, G, C ou T, et AG si S_n est T ou C ;

S_n est A, G, C ou T si QR_n est TC, et T ou C si QR_n est AG ;

15 et les indices n indiquent la position dans la séquence d'acides aminés de la proinsuline humaine, à laquelle correspond chaque triplet dans la séquence de nucléotides, selon le code génétique, les positions des acides aminés étant numérotées à partir de l'extrémité amino.

37 - Procédé selon la revendication 35, caractérisé
20 en ce que la séquence de désoxynucléotides codant pour la proinsuline humaine comprend une chaîne plus ayant (a) la séquence de désoxynucléotides $(N_a N_b N_c)_j$ ayant une extrémité 5' et une extrémité 3', où N_a , N_b et N_c peuvent être A, T, G ou C et j est un nombre entier de 0 à 100, $N_a N_b N_c$ n'étant pas TGA,
25 TAA ou TAG ; (b) une séquence de désoxynucléotides codant pour la chaîne B de l'insuline humaine fixée à l'extrémité 5' de $(N_a N_b N_c)_j$; et (c) une séquence de désoxynucléotides codant pour la chaîne A de l'insuline humaine fixée à l'extrémité 3' de $(N_a N_b N_c)_j$.

30 38 - Procédé de synthèse de l'insuline humaine, caractérisé en ce qu'on oxyde des groupes sulfhydryle sur les peptides A et B de la proinsuline humaine, préparée selon l'une quelconque des revendications 35 à 37, pour former des

ponts disulfures entre les peptides A et B, puis on excise le peptide C par une hydrolyse catalysée par une enzyme, spécifique pour les liaisons reliant le peptide C aux peptides A et B.

FIG. 1

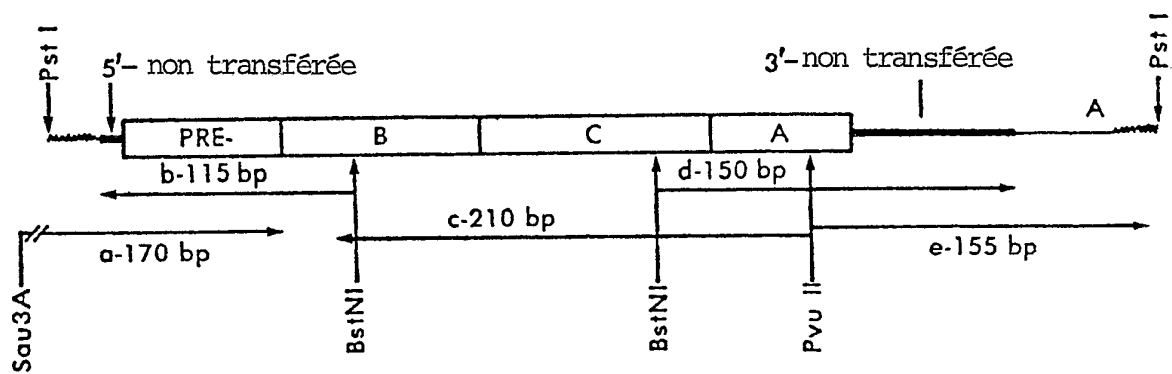


FIG. 2

-24 -20 -10
met ala leu trp met arg leu leu pro leu leu ala leu leu ala leu trp gly
UCCUUCUGCCAUG GCC CUG UGG AUG CGC CUC CUG CCC CUG CUG GCG CUG CUG GCC CUC UGG GGA

↑
Hha I

110

pro asp pro ala ala ala phe val asn gln his leu cys gly ser his leu val glu ala leu
CCU GAC CCA GCC GCA GCC UUU GUG AAC CAA CAC CUG UGC GGC UCA CAC CUG GUG GAA GCU CUC

Eco R II Alu I

Bst N 1

20
30
tyr leu val cys gly glu arg gly phe phe tyr thr pro lys thr arg arg glu ala glu asp
UAC CUA GUG UGC GGG GAA CGA GGC UUC UUC UAC ACA CCC AAG ACC CGC CGG GAG GCA GAG GAC

⁴⁰ leu gln val gly gln val glu leu gly gly gly pro gly ⁵⁰ ala gly ser leu gln pro leu ala
 CUG CAG GUG GGG CAG GUG GAG CUG GGC GGG GGC CCU GGU GCA GGC AGC CUG CAG CCC UUG GCC
 ↑
 Alu I

60
 leu glu gly ser leu gln lys arg gly ile val glu gln cys cys thr ser ile cys ser leu
 CUG GAG GGG UCC CUG CAG AAG CGU GGC AUU GUG GAA CAA UGC UGU ACC AGC AUC UGC UCC CUC

tyr gln leu glu asn tyr cys asn
 UAC CAG CUG GAG AAC UAC UGC AAC UAG ACGCAGCCCGCAGGCAGCCCCCACC CGCCGCCUCCUGCACC GAGAG

(Alu I
 Pvu II)

AGAUGGAAUAAAGCCCUUGAACCAGC