

[19] 中华人民共和国国家知识产权局



[12] 发明专利说明书

专利号 ZL 02806850.5

[51] Int. Cl.

C12N 15/62 (2006.01)

C12N 15/70 (2006.01)

C07K 14/62 (2006.01)

C07K 14/815 (2006.01)

C12N 1/21 (2006.01)

C12P 21/02 (2006.01)

[45] 授权公告日 2009 年 6 月 17 日

[11] 授权公告号 CN 100500853C

[51] Int. Cl. (续)

C12P 21/06 (2006.01)

[22] 申请日 2002.2.8 [21] 申请号 02806850.5

[30] 优先权

[32] 2001.2.20 [33] DE [31] 10108212.6

[86] 国际申请 PCT/EP2002/001307 2002.2.8

[87] 国际公布 WO2002/068660 英 2002.9.6

[85] 进入国家阶段日期 2003.9.19

[73] 专利权人 塞诺菲-安万特德国有限公司

地址 德国法兰克福

共同专利权人 瓦克尔化学股份公司

[72] 发明人 P·哈伯曼 J·埃特尔

[56] 参考文献

EP -0158564 -A1 1985.10.16

EP -0511393 -A1 1992.11.4

WO -9109125 -A1 1991.6.27

审查员 宋智刚

[74] 专利代理机构 北京市中咨律师事务所

代理人 黄革生 林柏楠

权利要求书 3 页 说明书 18 页

[54] 发明名称

用于向细菌培养物上清液中分泌目的蛋白质的融合蛋白质

[57] 摘要

本发明涉及包含融合部分和目的蛋白质的融合蛋白质，两种蛋白质的联合导致融合蛋白质分泌至细菌宿主的上清液中，而且目的蛋白质以其正确的三维结构存在。

1. 编码融合蛋白质的 DNA，其具有-F-As_m-R_n-Y-形式，
其中，

F 为编码 lepirudin、Ser - 蛭素或 Ala - 蛭素的 DNA 序列，所述 DNA 序列编码的氨基酸序列使得 Y 所编码的胰岛素原、胰岛素或其衍生物分泌至发酵培养基中，

As 为编码由遗传密码编码的氨基酸的三核苷酸序列，

m 为 0 - 10 的整数，

R 为精氨酸密码子，

n 为 0 或 1，并且

Y 为编码目的蛋白质的 DNA 序列，在发酵培养基中所述目的蛋白质正确折叠并且为融合蛋白质的一部分，其中所述目的蛋白质为胰岛素原、胰岛素或其衍生物。

2. 根据权利要求 1 所述的 DNA，其中所述 DNA 为 P-S-F-As_m-R_n-Y-T 的形式，其中

P 为启动子

S 为编码信号序列的 DNA 序列，所述信号序列导致最优产量，其为来源于荧光假单胞菌 (*Pseudomonas fluorescens*) 的 oprF 基因；编码鼠伤寒沙门氏菌 (*S. typhimurium*) 外膜蛋白质-fim D-的信号序列的 DNA；编码大肠杆菌 (*E. coli*) 碱性磷酸酶前体蛋白质的信号序列的 DNA 序列；编码信号序列 smompa 的 DNA 序列，所述信号序列 smompa 来源于粘质沙雷氏菌 (*Serratia marcescens*) 编码主要外膜蛋白质的 ompA 基因；编码信号序列 ecoompc 的 DNA 序列，所述信号序列 ecoompc 来源于大肠杆菌编码主要外膜蛋白质的 ompC 基因；编码信号序列 af009352 的 DNA 序列，所述信号序列 af009352 来源于枯草芽孢杆菌 (*Bacillus subtilis*) osmoprotectant 结合蛋白前体-opuCC-；编码信号序列 aeoxyna 的 DNA 序列，所述信号序列 aeoxyna 来源于豚鼠气单胞菌 (*Aeromonas caviae*) 编

码木聚糖酶 I 前体的 xynA 基因;或者编码信号序列 stomps1 的 DNA 序列, 所述信号序列 stomps1 来源于鼠伤寒沙门氏菌编码外膜蛋白质 S1 的基因,

T 为使表达增强的非翻译 DNA 序列。

3. 蛋白质, 其由权利要求 1 或 2 的 DNA 所编码。

4. 质粒, 其含有权利要求 1 或 2 所述的 DNA。

5. 宿主细胞, 其含有权利要求 4 所述的质粒。

6. 宿主细胞, 其含有权利要求 1 或 2 所述的 DNA。

7. 根据权利要求 5 或 6 所述的宿主细胞, 其中细胞选自大肠杆菌、枯草芽孢杆菌和链霉菌 (*Streptomyces*)。

8. 根据权利要求 7 所述的宿主细胞, 其中权利要求 4 所述的质粒或者权利要求 1 或 2 所述的 DNA 整合至宿主细胞的基因组中。

9. 发酵生产融合蛋白质的方法, 在该方法中

(a) 权利要求 1 或 2 所述的 DNA 分子在权利要求 5 或 6 所述的宿主细胞中表达; 并且

(b) 将表达的融合蛋白质分离。

10. 根据权利要求 9 所述的方法, 其中将上清液与宿主细胞分开以分离表达的蛋白质, 然后将表达的蛋白质从上清液中分离出来。

11. 根据权利要求 9 或 10 所述的方法, 其中用于分离表达蛋白质的步骤选自微量过滤、疏水相互作用层析和离子交换层析。

12. 根据权利要求 9 或 10 所述的方法, 其中表达蛋白质的分离包括步骤: 沉淀培养基或上清液中的组分而使表达蛋白质保持在溶液中。

13. 根据权利要求 9 或 10 所述的方法, 其中发酵后, 于 pH 6-9 向发酵上清液中加入硫醇或盐酸半胱氨酸, 产生浓度为 0.05 到 2.5 mM 的游离 SH 基团。

14. 根据权利要求 9 或 10 所述的方法, 其中在从宿主细胞分离出发酵上清液后, 将宿主细胞在新鲜培养基中重复培养, 并从培养中获得的每一上清液中将释放的融合蛋白质分离出来。

15. 生产胰岛素原、胰岛素或胰岛素衍生物的方法, 其中

-
- a) 根据权利要求 9 或 10 所述的方法获得表达的目的蛋白质，
 - b) 利用酶促或化学切割的方法释放目的蛋白质，并且
 - c) 将其分离。

16. 权利要求 15 的方法，其中所述目的蛋白质为胰岛素或胰岛素衍生物。

用于向细菌培养物上清液中分泌目的蛋白质的融合蛋白质

发明描述

本发明涉及融合蛋白质，其包含目的蛋白质和融合部分，所述两种蛋白质的联合导致融合蛋白质分泌至细菌宿主上清液中，并且目的蛋白质以正确的三维结构存在。融合蛋白质的基因序列为可在细菌宿主中表达的表达盒的一部分。本发明涉及对利用此表达盒的该融合蛋白质进行发酵、表达以及后处理的方法；涉及含有此表达盒的质粒；涉及细菌宿主细胞，所述细菌宿主细胞含有整合至染色体中和/或作为复制子（例如作为质粒）的此表达盒；涉及以蛭素或其衍生物作为融合部分的所述融合蛋白质；涉及生产胰岛素或胰岛素衍生物的方法；以及涉及所述表达盒在从蛭素或其衍生物制备融合蛋白质以及生产胰岛素或胰岛素衍生物的方法中的使用。

对基于重组蛋白质生产药物的优化方法的发展应在尽可能情况下满足以下两点，第一，方法应尽量节约成本，以及第二，产物应达到最高纯度。

对此而言，表达系统的选择决定了特定生产方法的过程，并且如熟练技术人员所显而易见的，新的蛋白质-化学技术的发展、多种生物化学可能性以及已知技术新的联合总能使现有方法的改进成为可能。

对用于合成的宿主细胞系统的选择由期望蛋白质的特性决定。细菌（诸如大肠杆菌）代表了利用其可以在便宜培养基中快速生产粗产量为几克蛋白质的系统。该系统尤其对于不需修饰并且可在体外复性至其生物学活性形式的蛋白质是有用的。对于大量需要的蛋白质（诸如胰岛素），旨在得到导致蛋白质以包含体形式在细胞内聚积的表达速率。细胞裂解后，将蛋白质溶解并使其在之后进一步的方法步骤中折叠。然而，折叠过程并不定量。这可能是由于包含体形成过程中的不可逆破坏、细胞裂解时的相应破坏以及折叠过程中的错误所造成。随后，“错误”折叠或修饰的分子不得

不在进一步的分离步骤中除去。这不利于节约生产成本。另外，在终产物中也会再次出现痕量所述分子。由于药物要满足高纯度标准，因此需要适当细致和高成本的纯化。由于有利的成本/粗产量比，由大肠杆菌向培养基中分泌正确折叠形式的目的蛋白质的方法将是值得期望的。然而，至今这只是在一些特例中获得成功。

国际专利申请 PCT/EP00/08537 描述了这种特例。当利用特异分泌信号序列时，利用大肠杆菌成功地合成和分泌了克级的 lepirudin(药物 Refludan[®]的活性成分)。德国专利申请 No. 100 33 195.2 (未发表) 描述了一种双功能蛋白质，其包括蛭素和蛭素衍生物以及来源于壁虱的 Xa 因子抑制剂及其衍生物。所述蛋白质同样可以利用大肠杆菌大量合成和分泌。除这一发现之外，随后又令人惊讶地发现蛭素不仅仅可以以 TAP 融合蛋白质的形式大量分泌，而且也可作为与多肽(诸如胰岛素原衍生物)的融合蛋白质的一部分而大量分泌；还发现其具有生物学活性并且令人吃惊地，融合部分诸如胰岛素原以正确的三维结构存在。由于可以省却经细胞内表达后再进行体外折叠的步骤(所述步骤可造成不可忽略的产量损失)，以及由此导致的更简单的蛋白质纯化方法，这一出乎意料的结果使得利用细菌宿主/载体系统以更节约成本的方法生产诸如胰岛素等成为可能。另一优势是不需加入离液辅助剂来溶解融合蛋白质，而在传统方法中利用大肠杆菌生产胰岛素时这是必须的。生态学上，这通过避免相应废物从而导致更少的环境污染。

蛭类(Hirudo)中的水蛭已经形成了例如多种凝血酶抑制剂蛭素同工型。蛭素通过分子的人为变异(例如 N-末端氨基酸的改变)已进行了优化以适应药学需要(见 EP-A 0 324 712)。

本发明包括在形成融合蛋白质(例如与猿猴胰岛素原或其衍生物)时蛭素和蛭素变体的使用。本发明特定实施方案使用一种天然蛭素同工型(所有天然蛭素同工型总称为“蛭素”)。天然同工型有例如 Val-Val-蛭素或 Ile-Thr-蛭素。本发明其它实施方案使用天然蛭素同工型的变体。变体来源于天然蛭素同工型，但其与天然同工型相比含有例如氨基酸添加和/或氨基

酸缺失和/或氨基酸改变。蛭素变体可以含有天然蛭素同工型的交替肽片断和新的氨基酸。蛭素变体为人所公知并被描述在例如 DE 3 430 556 中。蛭素变体可以以蛋白质的形式通过商业途径得到 (Calbiochem® Biochemicals, 目录号 377-853, -950-960)。

胰岛素为具有 51 个氨基酸的多肽, 所述 51 个氨基酸分布于两条氨基酸链上: 21 个氨基酸的 A 链和 30 个氨基酸的 B 链。两条链彼此通过 2 个二硫桥连接。多年来胰岛素组合物被用于糖尿病治疗。这不仅包括使用天然胰岛素, 还包括使用胰岛素衍生物和类似物。

胰岛素衍生物为天然胰岛素 (即人胰岛素或动物胰岛素) 的衍生物, 其通过至少一个天然氨基酸残基的置换和/或至少一个氨基酸残基和/或有机残基的添加而与相应天然胰岛素不同, 否则, 所述胰岛素衍生物与天然胰岛素相同。总的来说, 胰岛素衍生物与人胰岛素相比具有稍有不同的作用。

起效加速的胰岛素衍生物在 EP 0 214 826、EP 0 375 437 和 EP 0 678 522 中描述。EP 0 214 826 等涉及 B27 和 B28 的置换。EP 0 678 522 描述了在 B29 位具有不同氨基酸 (优选脯氨酸, 但不为谷氨酸) 的胰岛素衍生物。EP 0 375 437 涉及在 B28 位为赖氨酸或精氨酸的胰岛素衍生物, 如需要, 其还可以在 B3 和/或 A21 进行额外修饰。

EP 0 419 504 公开了胰岛素衍生物, 其通过 B3 位的天冬酰胺和 A5、A15、A18 或 A21 位的至少一个其它氨基酸的修饰对化学修饰产生耐受性。

WO 92/00321 描述了在 B1-B6 位至少一个氨基酸被赖氨酸或精氨酸替代的胰岛素衍生物。根据 WO 92/00321, 这种胰岛素展现出延长的作用。

当利用基因工程生产胰岛素和胰岛素衍生物时, 经常表达胰岛素前体, 即“胰岛素原”, 其包含 B、C 和 A 链。所述胰岛素原经适当和正确折叠以及形成二硫桥后可以通过酶促或化学去除 C 链而转变为胰岛素或胰岛素衍生物。胰岛素原经常以融合蛋白质形式表达。“不想要的”融合部分同样需要化学或酶促去除。

对本领域熟练技术人员显而易见的是, 重组细胞的培养、增殖和发酵

方法取决于所选择的重组宿主/载体系统。这同样是本发明的主题。

此融合蛋白质在酸性培养基中显示出令人惊讶地良好溶解性，这导致在此蛋白质的化学后处理方面的独特优势。首先，上清液中许多不需要组分可在所述条件下沉淀，而其次肽酶或蛋白酶在所述条件下失活。因此，在操作的最后，酸化发酵肉汤将使以下成为可能，即从融合蛋白质中直接分离不需要的上清液蛋白质连同宿主细胞，以及在进一步的步骤中浓缩所述融合蛋白质。这同样是本发明的主题。

在发酵的最后，折叠过程可能不会 100% 完全。添加硫醇或例如盐酸半胱氨酸可以使此过程完全。这同样是本发明的主题。

如果这两个蛋白质通过内切蛋白酶特异识别的氨基酸接头融合，那么目的蛋白质就可以直接以活性形式切下，其中所述内切蛋白酶仅在氨基酸接头处而不在其它位置有效切割融合蛋白质。在生产胰岛素时，蛭素和胰岛素原之间的接头优选在其羧基末端含有精氨酸。这样就使得可在同时的处理中利用胰蛋白酶的转换作用来切除融合部分以及将胰岛素原转变为单或双-Arg-胰岛素。所述接头必须相对于胰岛素加工过程进行优化以使蛭素部分的切除不慢于连接胰岛素 B 和 A 链的 C 肽序列或其衍生物的切除。这同样是本发明的主题。可用的表达系统的例子有载体 pJF 118，其描述在欧洲专利 0 468 539 的图 1 中。

含有编码胰岛素原或胰岛素原衍生物的 DNA 序列的质粒在例如专利 EP-A 0 489 780 和 PCT/EP 00/08537 中描述。

含有蛭素序列的质粒 pK 152 (根据 EP-A 0 324 712) 被用作蛭素 DNA 序列的来源。

与跨过细菌内膜有关的目的蛋白质的输出相容性对分泌而言非常重要。在这一点，选择对不同蛋白质而言或多或少为最佳的信号序列非常重要，专利申请 PCT/EP00/08537 描述了基于 PCR 的信号序列筛选系统。既然蛭素活性令人惊讶地保持完整并且由此可以通过凝血酶抑制试验在上清液中容易地检测到，这一系统也可以应用于以蛭素作为 N 端融合部分的融合蛋白质上。

于是，本发明涉及编码融合蛋白质的 DNA（替代术语：表达盒），其具有 $-F-As_m-R_n-Y-$ 的形式，

其中

F 为编码氨基酸序列的 DNA 序列，所述氨基酸序列允许蛋白质 Y 分泌至发酵培养基，

As 为化学键或编码可由遗传密码编码的氨基酸的 DNA 序列，

m 为 0-10 的整数

R 为化学键或精氨酸密码子，

n 为 0 或 1，并且

Y 为编码目的蛋白质的 DNA 序列，在发酵培养基中所述目的蛋白质正确折叠并且为融合蛋白质的一部分，

其中尤其选择编码蛭素或其衍生物（F）和胰岛素原或其衍生物（Y）的 DNA 序列。

本发明还涉及 $P-S-F-As_m-R_n-Y-T$ 形式的表达盒（替代术语：DNA 分子），

其中

P 为启动子，

S 为编码信号序列的 DNA 序列，所述信号序列可导致最优产量，

T 为可增强表达的非翻译 DNA 序列。

本发明还涉及含有上述表达盒的质粒，并且还涉及含有所述质粒的宿主细胞或优选含有整合至宿主基因组中的表达盒的宿主细胞，所述宿主细胞选自大肠杆菌、枯草芽孢杆菌（*B.subtilis*）和链霉菌（*Streptomyces*）。

本发明也涉及发酵生产上述融合蛋白质的方法，在所述方法中

(a)如上所述的 DNA 分子在如上所述的宿主细胞中表达，并且

(b)分离表达的融合蛋白质；

特别地，在所述方法中，将上清液与宿主细胞分开以分离表达蛋白质，然后将表达蛋白质从上清液中分离出来；在所述方法中，用于沉淀后浓缩上清液中的表达蛋白质的步骤选自微量过滤、疏水相互作用层析

和离子交换层析；在所述方法的一个特定实施方案中，分离表达蛋白质包括将培养基或上清液中的组分沉淀而使表达蛋白质保持在溶液中的步骤；在所述方法的本发明另一优选实施方案中，发酵后，向 pH 6-9 的发酵上清液中加入硫醇或盐酸半胱氨酸，产生 0.05 到 2.5 mM 的游离 SH 基浓度。

本发明的一个特定实施方案包括从宿主细胞中分离发酵上清液，进一步在新鲜培养基中培养宿主细胞并从上清液中分离释放的融合蛋白质。换言之，本发明另一实施方案为如上所述的方法，在所述方法中，从宿主细胞分离发酵上清液后，将宿主细胞在新鲜培养基中重复培养，并且从培养过程中获得的每一上清液中分离释放的融合蛋白质。

本发明还涉及生产胰岛素或胰岛素衍生物的方法，在该方法中，

- (a) 从通过如上所述方法获得的表达蛋白质中
- (b) 利用酶促或化学切割释放目的蛋白质，尤其是胰岛素或胰岛素衍生物，并且
- (C) 将目的蛋白质分离。

下述实施例用于更详尽描述本发明，而无意限制本发明。

实施例 1: 附加于来自荧光假单胞菌 (*Pseudomonas fluorescens*) oprF 基因产物的信号序列之后的 lepirudin-GNSAR-猿猴胰岛素原融合蛋白质的构建

专利申请 PCT/EP00/08537 实施例 2 描述了一种用于大肠杆菌的表达载体，其通过荧光假单胞菌 oprF 基因产物的信号序列使 Refludan 表达和分泌至培养基中 (De, E.等人, FEMS 微生物学通讯 (FEMS Microbiol Lett.), 127, 263-272, 1995)。这一载体用于构建 Refludan-GNSAR-猿猴胰岛素原融合蛋白质 (GNSAR = SEQ ID NO.:1)，并且将其定义为 pBpfu_hir。

另外的起始材料为 pJF118(EP 0468 539)和 pK152(PCT/EP00/08537)

质粒 DNA。下列寡核苷酸为必需：

引物 pfuf1

5'GGTTCTCTTA TTGCCGCTAC TTCTTTCGGC GTTCTGGCAc ttacgtatac tgactgca
3'

(SEQ ID NO.: 2)

引物 insu11HindIII

5' - TTTTTAAGCTTCATGTTTGA CAGCTTATCA T -3' (SEQ ID NO.: 3)

引物 Hir_insf1

5' ATCCCTGAGG AATACCTTCA GGGAAATTCG GCACGATTG TG - 3'(SEQ ID
NO.: 4)

引物 Hir_insrev1

5' - CACAAATCGT GCCGAATTC CCTGAAGGTA TTCCTCAGGG AT -3'(SEQ ID
NO.: 5)

引物 pfuf1 与表达载体上编码信号序列与 lepirudin 的接点的 DNA 区域杂交。

引物 Hir_insrev1 中显示为粗体的部分与质粒 pINT90d 上编码前胰岛素原和猿猴胰岛素原序列的接点的 DNA 区域杂交，并且与质粒 pK152 上蛭素序列的 3'末端序列杂交。引物 Hir_insrev1 与引物 Hir_insf1 100% 互补。

引物 insu11HindIII标明了克隆在 pINT90d 中并编码猿猴胰岛素原序列的 DNA 区域的 3'末端，并额外带有限制酶 HindIII识别的六核苷酸序列。

利用 Hir_insf1/ Insu11hindIII引物对和质粒 pINT90d 为模板以及利用 pfuf1/Hir_insrev 引物对和质粒 pBpfu_hir 为模板进行两个标准聚合酶链式

反应。将两反应产物混和，并取出一小份利用引物 pfuf1/ Insu11HindIII进行第三个聚合酶链式反应。结果为含有信号（部分）-lepirudin-GNSAR-猿猴胰岛素原序列的 DNA 产物。将 DNA 片断利用限制酶 BamH I 和 HindIII消化，其中 BamH I 在 lepirudin 序列中而 HindIII在胰岛素原编码序列的 3'末端切割。

在平行的反应中，载体 pBpfu 用所述两个酶消化并分离大的载体片断。两反应的分离产物在 T4 连接酶反应中连接。利用连接混合物转化大肠杆菌菌株 K12 Mc1061 感受态细胞（Sambrook 等人，分子克隆（Molecular Cloning）（冷泉港实验室出版社 Cold Spring Harbor Laboratory Press）1989）并涂布至含有 25 µg/ml 氨苄青霉素的 NA 平板上。从转化体中分离质粒 DNA 进行鉴定。同时，将经质粒分析鉴定的转化体制平板以保种。DNA 通过限制性分析和 DNA 序列分析进行鉴定。鉴定为正确的质粒命名为 pBpfuHir_Ins。

实施例 2: 附加至鼠伤寒沙门氏菌(*S. typhimurium*)外膜蛋白质(fimD)信号序列后的 Ser-蛭素-GNSAR-猿猴胰岛素原融合蛋白质的构建
构建根据实施例 1 所述方案进行。

专利申请 PCT/EP00/08537 实施例 10 描述了通过鼠伤寒沙门氏菌外膜蛋白质信号序列（Rioux, C.R., Friedrich, M.J. 和 Kadner, R.J.；细菌学杂志（J. Bacteriol.）172（11），6217-6222（1990））用于输出 lepirudin 的载体的构建。所产生的质粒命名为 pBstyfim_hir，用于实验目的。质粒 pK152 和 pINT90d 的 DNA 在各自场合作为模板。

构建需要 4 个引物。

引物 insu11HindIII、Hir_insf1 和 Hir_insrev1 如实施例 1 所述。

引物 styfimf1 为新合成的，并且其序列如下：

5' CGGCGCTGAG TCTCGCCTTA TTTTCTCACC TATCTTTTGC CTCTacgtat
actgactgcaCTG 3' (SEQ ID NO.: 6)

显示为粗体的 DNA 三联体表示丝氨酸密码子。其结果产生在氨基酸序列的 1 位由丝氨酸代替亮氨酸的蛭素。

与实施例 1 相类似, 用 Hir_insfl/ Insu11HindIII 引物对及质粒 pINT90d 为模板以及利用 styfimflser/Hir_insrev 引物对及 pK152 DNA 为模板进行两个标准聚合酶链式反应。将两反应产物混和, 并取出一小份利用引物 styfimflser/Insu11HindIII 进行第三个聚合酶链式反应。结果为含有信号 (部分) - Ser - 蛭素 - GNSAR - 猿猴胰岛素原序列的 DNA 产物。将 DNA 片断利用限制酶 BamH I 和 HindIII 消化。

在平行的反应中, 载体 pBstyfim_Hir 用所述两个酶消化并将大的载体片断分离出来。两反应的分离产物在 T4 连接酶反应中连接。利用连接混合物转化大肠杆菌菌株 K12 Mc1061 感受态细胞。从转化体中分离质粒 DNA 进行鉴定。同时, 将经质粒分析鉴定的转化体制平板以保种。DNA 通过限制性分析和 DNA 序列分析进行鉴定。鉴定为正确的质粒命名为 pBstyfim_SerHir_Ins。

实施例 3: 附加至大肠杆菌碱性磷酸酶前体蛋白质信号序列之后的 Ala-蛭素-R-猿猴胰岛素原融合蛋白质的构建

大肠杆菌碱性磷酸酶前体蛋白质具有信号序列:

MKQSTIALAL LPLLFTPVT A (SEQ ID NO.: 7)

(Shuttleworth H., Taylor J., Minton N.; Nucleic Acids Res. 14:8689, (1986)).

此肽序列通过 GCG 程序 Backtranslate (Wisconsin Package 10.1 版, Genetics Computer Group (GCG), Madison, Wisc.) 使用大肠杆菌高密码子使用标准翻译成 DNA。

这产生下述序列:

5'ATGAAACAGTCGACCATCGCGCTGGCGCTGCTGCCGCTGCTGTTACCCCCGGT
TACCAAAGCG 3' (SEQ ID NO.: 8)

为将此序列克隆并加至编码蛭素（其特征在于1位为丙氨酸（EP-A 0 448 093））的DNA序列上，所述序列由以下显示为粗体的序列延长：

5' TTTTTTGAATTCATGAAACAGTCGACCATCGCGCTGGCGCTGCTGCCGCTGCTGTTTCAC
CCCGGTTACCAAAG -CG GCTacgtat actgactgcaCTG (SEQ ID NO.: 9)

如下部分重叠的两寡核苷酸序列即来源于此。

引物 phoaf1 具有以下序列：

5' CTGCTGCCGCTGCTGTTCAACCCGGTTACCAAAGCG GCTACG
TATACTGACTGCACTG -3' (SEQ ID NO.: 10)

引物 phoaf2 具有以下序列：

5' TTTTTTGAATTCATGAAACAGTCGACCATCGCGCTGGCGCTGCTGCCGCTGCTG -3'
(SEQ ID NO.: 11)

表达载体的构建需要引物 insu11HindIII、Hir_insf2 和 Hir_insrev2 以及质粒 pK152、pINT90d 和 pJF118 的 DNA。

引物 Hir_insf2 具有以下序列：

5' - ATCCCTGAGGAATACCTTCAGcgaTTTGTGAACCAGCAC C -3' (SEQ ID NO.
12)

引物 Hir_insrev2 具有以下序列：

5' - GGTGCTGGTTCACAAAtcqCTGAAGGTA TTCCTCAGGG AT-3' (SEQ ID NO.13)

粗体中的大写字母表示与胰岛素原杂交的序列，而普通字体中的大写字母表示与蛭素序列 3' 末端重叠的序列。带下划线的粗体小写字母表示接头精氨酸的密码子。

与实施例 1 相类似，利用 Hir_insf1/ Insu11HindIII 引物对及质粒 pINT90d 的 DNA 为模板以及利用 phoaf1/Hir_insrev 引物对及 pK152 DNA 为模板进行两个标准聚合酶链式反应。将两反应的产物混和，并取出一小

份利用引物 phoaf1/ Insu11HindIII进行第三个聚合酶链式反应。结果为含有信号 - Ala - 蛭素 - GNSAR - 猿猴胰岛素原序列的 DNA 产物。将 DNA 片断利用限制酶 BamH I 和 HindIII消化。在平行的反应中,载体 pJF118 用所述两个酶消化并分离大的载体片断。两反应的分离产物在 T4 连接酶反应中连接。利用连接混合物转化大肠杆菌菌株 K12 Mc1061 感受态细胞。并且从转化体中分离质粒 DNA 进行鉴定。同时,将经质粒分析鉴定的转化体制平板以保种。DNA 通过限制性分析和 DNA 序列分析进行鉴定。鉴定为正确的质粒命名为 pNS22。

实施例 4: 凝血酶抑制试验

蛭素浓度根据 Griebbach 等人 (血栓形成研究 (Thrombosis Research) 37, 347 - 350 页, 1985) 的方法确定。为此,测定中包括特定量的 Refludan 标准以建立校正曲线,从所述曲线可直接确定产物的量 (mg/l)。生物学活性也是对融合蛋白质中胰岛素原组分正确折叠的直接测定。或者,可以利用金黄色葡萄球菌 (*S. aureus*) 蛋白质水解消化和随后 RP-HPLC 系统的分析来确定正确 S-S 桥形成。

实施例 5: 融合蛋白质的表达

重组细胞在含有 100 µg/ml 氨苄青霉素的 2YT 培养基 (每升: 16 克胰蛋白胨、10 克酵母提取物、5 克 NaCl) 中培养过夜。将过夜培养物用新鲜培养基 1:50 稀释并将细胞培养至密度为约 0.8 OD₆₀₀。

然后通过添加 IPTG 至 0.05 - 2 mM 来诱导表达。以这种方式诱导的细胞进一步孵育 3 - 26 小时。

3 小时后,可清楚地测量到上清液中蛭素的抗凝血酶作用。由于经考马斯亮蓝染色后的 SDS PAGE 显示只有在诱导细胞中有可在 Western 印迹分析中与多克隆抗胰岛素抗体反应的新条带,因此所述作用归因于目的融合蛋白质的分泌。在发酵实验中,诱导只在培养至显著高的光密度时开始。此处优选基于基本培养基的合成培养基。

细胞生产力可以通过利用抽取细菌“奶”（milking）的原理增加，即通过在最佳诱导时间后，小心从上清液中移去细胞并进一步将细胞在新鲜培养基中孵育，然后可以向该培养基中再次加入诱导剂。之后平行地从收获的上清液中制备胰岛素。

实施例 6：融合蛋白质的纯化

诱导完毕后，将细胞上清液调节至 pH 2.5 - 3，并通过离心或过滤将细胞和上清液组分除去。将沉淀后的上清液加至阳离子交换柱（S-Hyper DF, Source 30S）上并在 30 % 2-丙醇存在下利用 150 到 450 mM NaCl 线性梯度于 pH 3.5 进行分级分离。各级分利用 RP-HPLC 方法分析。胰岛素原-蛭素融合蛋白质在约 300 mM 的 NaCl 浓度处洗脱下来。将足够纯的级分合并，用 0.1 % TFA 稀释并利用泵将其加至 RP 柱（PLRP-S 7.5×50 mm）上。洗脱利用 25 - 50 % 乙腈梯度进行。合并两组级分。去除溶剂后将材料冻干。通过 SDS 聚丙烯酰胺电泳检查材料的纯度。纯化的融合蛋白质利用质谱（ESI）分析。通过实验确定的融合蛋白质分子量与去除信号肽后其理论预期分子量相符。

实施例 7：二硫桥连接的确证

用胰蛋白酶消化融合蛋白质，并将形成的片断通过 RP-HPLC 方法以及随后利用质谱方法分析。标识为 de-(B30)胰岛素的片断基于其 5706 Da 的质量被成功鉴定。这一产物用金黄色葡萄球菌（*S. aureus*）V8 蛋白酶消化。RP-HPLC 分析显示出预期的肽谱。

胰蛋白酶切割按照以下所述进行：

将冻干的融合蛋白质溶于 50 mM Tris-HCl pH 8（1 mg/ml），并加入胰蛋白酶（每 mg 融合蛋白质 1 μ g）。在反应结束后，pH 调至 3 以失活胰蛋白酶。

金黄色葡萄球菌（*S. aureus*）消化按照以下所述进行：

分离的 de-(B30)胰岛素于 pH 8 溶于水，加入金黄色葡萄球菌（*S.*

aureus) 蛋白酶(胰岛素量的 1/50)。并将混合物于 37℃ 孵育 5 小时, 而后室温过夜。

实施例 8: 胰岛素的纯化

不同于上清液中大多数其它由于宿主细胞自发裂解或是由于分泌而存在的多肽, 融合蛋白质令人惊讶地在 pH2.5 - 3.5 条件下不沉淀。因此培养基就可以适当酸化, 并且随后在沉淀作用完成后通过离心或微量过滤除去沉淀和细胞, 并进行浓缩。

随后, 将培养基调节至 pH 6.8, 并且平行地通过分析性 HPLC 测定法确定融合蛋白质含量。测定后向上清液中加入胰蛋白酶使其约为每 1 - 1.5 mg 融合蛋白质 1 μ g。在室温下孵育约 4 小时后, 纯化利用阳离子交换树脂在 2 - 丙醇存在下、pH 3.5 进行。洗脱通过在缓冲液中应用 0.15 至 0.45 M 的梯度进行。

双 - Arg - 胰岛素在约 0.3 M 处洗脱。经 1:1 稀释后, 双 - Arg - 胰岛素通过添加 10% 强度 ZnCl_2 溶液于 pH 6.8 从含胰岛素的级分中沉淀出来。将胰岛素滤出并随后将其溶于 0.05 M Tris-HCl (pH 8.5) 而产生 2 mg/ml 的溶液。

随后向每 100 ml 溶液中加入约 1 单位羧肽酶 B, 并在反应进行时轻微搅动。然后用柠檬酸调节 pH 至 5.5, 并将胰岛素在 ZnCl_2 存在下结晶。将结晶移出、溶解, 经 RP-HPLC 纯化后, 胰岛素再经结晶纯化。

实施例 9: 在培养基中直接处理融合蛋白质

在表达期的最后, 将培养基调节至 pH 6.8 并随后边搅动边加入胰蛋白酶至每升 4 - 8 mg。孵育约 4 小时后, 将以此方式处理的发酵肉汤调节至 pH 2.5 - 3。沉淀 1 - 6 小时后, 将 pH 升至 3.5, 并通过阳离子交换层析在 30% 2 - 丙醇存在下纯化形成的双 - Arg - 胰岛素。洗脱通过 0.05 - 0.5 M 的 NaCl 梯度进行。含有产物的级分用水 1:1 稀释并随后加入 ZnCl_2 , 使之形成 0.1% 强度的 ZnCl_2 溶液。双 - Arg - 胰岛素在 pH 6.8 时沉淀然后可例

如根据实施例 8 转变为胰岛素。

实施例 10: 用于融合蛋白质分泌的其它信号序列

利用专利申请 PCT/EP00/08537 描述的技术可以检测到能导致蛭素 - 胰岛素原融合蛋白质分泌的其它信号序列:

信号序列 smompa, 其来源于粘质沙雷氏菌 (*Serratia marcescens*) 的主要外膜蛋白质的 ompA 基因(GenEMBL 数据库定位: SMOMPA, 1364bp DNA BCT 30-3-1995)

信号序列 ecoompc, 其来源于大肠杆菌编码主要外膜蛋白质的 ompC 基因(GenEMBL 数据库定位: SMOMPA, 1364bp, DNA BCT 30-3-1995)

信号序列 af009352, 其来源于枯草芽孢杆菌 (*Bacillus subtilis*) osmoprotectant 结合蛋白前体 (opuCC) (GenEMBL 数据库定位: AF009532, 4500bp, DNA BCT 23-7-1997)

信号序列 aeoxyna, 其来源于豚鼠气单胞菌 (*Aeromonas caviae*) 编码木聚糖酶 I 前体的 xynA 基因(GenEMBL 数据库定位: AEOXYNA, 1139bp, DNA BCT 07-2-1999)

信号序列 stomps1, 其来源于鼠伤寒沙门氏菌 (*S. typhi*) 编码外膜蛋白质 S1 的基因 (GenEMBL 数据库定位: STOMPS1, 1938bp, DNA BCT 24-8-1995)

序列表

<110> 安万特医药德国有限公司 (Aventis Pharma Deutschland GmbH)

<120> 用于向细菌培养物上清液中分泌目的蛋白质的融合蛋白质

<130> DEAV2001/0009

<140> 10108212.6

<141> 2001-02-20

<160> 13

<170> PatentIn Ver. 2.1

<210> 1

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 人工序列的描述:pBpfu_hir

<400> 1

Gly Asn Ser Ala Arg

1 5

<210> 2

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:pfuf1

<400> 2

ggttctctta ttgccgtac ttctttcggc gttctggcac ttacgtatac tgactgca 58

<210> 3

<211> 31

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:insu11hind111

<400> 3

tttttaagct tcatgtttga cagcttatca t

31

<210> 4

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:Hir_insfl

<400> 4

atccctgagg aataccttca gggaaattcg gcacgatttg tg

42

<210> 5

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:Hir_insrev1

<400> 5

cacaaatcgt gccgaatttc cctgaaggta ttcctcaggg at

42

<210> 6

<211> 63

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:styfimf1

<400> 6

cggcgctgag tctcgctta tttctcacc tatcttttgc ctctacgtat actgactgca
ctg

60
63

<210> 7
<211> 21
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述: 碱性磷酸酶(信号序列)

<400> 7
Met Lys Gln Ser Thr Ile Ala Leu Ala Leu Leu Pro Leu Leu Phe Thr
1 5 10 15

Pro Val Thr Lys Ala
20

<210> 8
<211> 63
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述:碱性磷酸酶 (信号序列)

<400> 8
atgaaacagt cgaccatcgc gctggcgctg ctgccgctgc tgttcacccc ggttaccaa 60
gcg 63

<210> 9
<211> 97
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述:克隆片段

<400> 9
ttttttgaat tcatgaaaca gtcgaccatc gcgctggcgc tgctgccgct gctgttcacc 60
ccggttacca aagcggctac gtatactgac tgcactg 97

<210> 10
<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:phoaf1

<400> 10

ctgctgccgc tgctgttcac cccggttacc aaagcggcta cgtatactga ctgcactg 58

<210> 11

<211> 54

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:phoaf2

<400> 11

ttttttgaat tcatgaaaca gtcgaccatc gcgctggcgc tgctgccgct gctg 54

<210> 12

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:Hir_insf2

<400> 12

atccctgagg aataccttca gcgatttgtg aaccagcacc 40

<210> 13

<211> 40

<212> DNA

<213> 人工序列

<220>

<223> 人工序列的描述:Hir_insrev2

<400> 13

ggtgctggtt cacaaatcgc tgaaggtatt cctcagggat 40