

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

C07K 19/00

[12] 发明专利申请公开说明书

C07K 14/62 C07K 14/61

C12N 15/62 C12N 1/21

//(C12N1/21 C12R1: 19)

[21] 申请号 98813941.3

[43] 公开日 2001 年 4 月 11 日

[11] 公开号 CN 1291199A

[22] 申请日 1998.3.31 [21] 申请号 98813941.3

[86] 国际申请 PCT/CN98/00052 1998.3.31

[87] 国际公布 WO99/50302 英 1999.10.7

[85] 进入国家阶段日期 2000.9.29

[71] 申请人 通化安泰克生物工程公司

地址 中国吉林省通化市东宝新村

[72] 发明人 甘忠如

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 卢新华 杨九昌

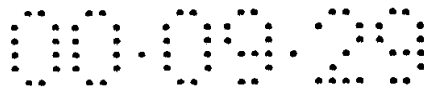
权利要求书 6 页 说明书 32 页 附图页数 2 页

[54] 发明名称 含有分子内伴侣样序列的嵌合蛋白及其在胰岛素生产中的应用

[57] 摘要

本发明涉及含有与靶蛋白,优选胰岛素前体相连的分子内伴侣(IMC)样序列的嵌合蛋白。本发明还涉及获得含胰岛素前体的正确折叠的嵌合蛋白的方法,该方法包括,特别是,将含有与胰岛素前体相连的 IMC 样序列的非正确折叠的嵌合蛋白与至少一种离液辅助剂接触。本发明还涉及筛选氨基酸的测定方法,该氨基酸序列能够通过含有与胰岛素前体相连的 IMC 样序列的嵌合蛋白促进胰岛素前体的折叠。

I S S N 1 0 0 8 - 4 2 7 4



权 利 要 求 书

1. 一种嵌合蛋白，该蛋白自 N 端至 C 端包含：

5 a) 一种第一肽段，该肽段由一段与含有人生长激素 (hGH) 蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 40% 的氨基酸序列组成，其中同一性百分率由大小与 hGH 的区域相同的一段氨基酸序列决定；

b) 一个精氨酸残基，或一个赖氨酸残基，或一种至少由 2 个氨基酸组成的第二肽段，其中该肽段的最 C 端氨基酸残基为 Arg 或 Lys 残基；以及

10 c) 一种第三肽段，该肽段包含一段由 2 个以上半胱氨酸残基的氨基酸序列组成，并且该肽段不是 hGH 蛋白的一部分。

2. 权利要求 1 的嵌合蛋白，其中第一肽段由一段与 hGH 蛋白的区域的同一性至少为 60% 的氨基酸序列组成。

15 3. 权利要求 1 的嵌合蛋白，其中第一肽段能够被抗 hGH 抗体结合。

4. 权利要求 1 的嵌合蛋白，其中第一肽段由 SEQ ID NO:1 的氨基酸序列组成。

5. 权利要求 1 的嵌合蛋白，其中第一肽段由 SEQ ID NO:2 的氨基酸序列组成。

20 6. 权利要求 1 的嵌合蛋白，其中第二肽段由 SEQ ID NO:3 的氨基酸序列组成。

7. 权利要求 1 的嵌合蛋白，其中第三肽段为胰岛素前体。

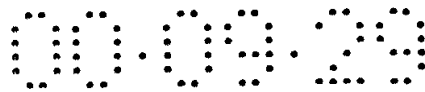
8. 权利要求 7 的嵌合蛋白，其中胰岛素前体来源于人。

25 9. 权利要求 8 的嵌合蛋白，其中人胰岛素前体能够被抗人胰岛素抗体结合。

10. 权利要求 8 的嵌合蛋白，其中人胰岛素前体由 SEQ ID NO:4 的氨基酸序列组成。

30 11. 权利要求 8 的嵌合蛋白，其中在胰岛素前体中的人胰岛素前体 B 链和 A 链由一个氨基酸残基或一段由 2-34 个氨基酸残基组成的肽段隔开。

12. 权利要求 8 的嵌合蛋白，其中人胰岛素前体由 SEQ ID NO:5



的氨基酸序列组成。

13. 一种由 SEQ ID NO:6 的氨基酸序列组成的嵌合蛋白。

14. 一种由 SEQ ID NO:7 的氨基酸序列组成的嵌合蛋白。

5 15. 一种分离的核酸，该核酸含有编码权利要求 1 的嵌合蛋白的核苷酸序列。

16. 一种分离的核酸，该核酸含有编码权利要求 13 的嵌合蛋白的核苷酸序列。

17. 一种分离的核酸，该核酸含有编码权利要求 14 的嵌合蛋白的核苷酸序列。

10 18. 权利要求 15 的核酸，其中所述核酸为 DNA。

19. 一种分离的核酸，该核酸含有与权利要求 15 的核苷酸序列互补的核苷酸序列。

20. 一种分离的核酸，该核酸能够与权利要求 18 的 DNA 中编码第一、第二和第三肽段的核苷酸序列杂交。

15 21. 一种含有权利要求 15 的核酸的重组细胞。

22. 一种含有权利要求 16 的核酸的重组细胞。

23. 一种含有权利要求 17 的核酸的重组细胞。

20 24. 一种产生嵌合蛋白的方法，该方法包括将含有权利要求 15 的核酸的重组细胞进行培养以使细胞表达编码的嵌合蛋白，并回收表达的嵌合蛋白。

25. 一种产生嵌合蛋白的方法，该方法包括将含有权利要求 16 的核酸的重组细胞进行培养以使细胞表达编码的嵌合蛋白，并回收表达的嵌合蛋白。

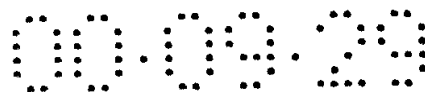
25 26. 一种产生嵌合蛋白的方法，该方法包括将含有权利要求 17 的核酸的重组细胞进行培养以使细胞表达编码的嵌合蛋白，并回收表达的嵌合蛋白。

27. 权利要求 24 的方法的产物。

28. 权利要求 25 的方法的产物。

29. 权利要求 26 的方法的产物。

30 30. 一种获得含有胰岛素前体的正确折叠第一嵌合蛋白的方法，该方法包括在一种水性介质中将含有胰岛素前体的非正确折叠第二



嵌合蛋白与至少一种离液辅助剂相接触，所述含有胰岛素前体的第二嵌合蛋白由一种经一个或多个可切除氨基酸残基与胰岛素前体隔离的分子内伴侣（IMC）样肽段组成；其中所述 IMC 样肽段：

a) 含有约 20-200 个氨基酸残基；

5 b) 不是胰岛素前体或其一部分；并且

c) 可促进胰岛素前体的折叠，以至于将含有胰岛素前体的非正确折叠第二嵌合蛋白与离液辅助剂接触时，含有胰岛素前体的正确折叠第一嵌合蛋白的产量要高于将不含有所述 IMC 样肽段的非正确折叠胰岛素前体与同样的离液辅助剂接触时的正确折叠胰岛素前体的产量。

10

31. 权利要求 30 的方法，其中胰岛素前体来源于人。

32. 权利要求 31 的方法，其中人胰岛素前体能够被抗人胰岛素抗体结合。

15 33. 权利要求 31 的方法，其中人胰岛素前体由 SEQ ID NO:4 的氨基酸序列组成。

34. 权利要求 31 的方法，其中在胰岛素前体中的人胰岛素前体 B 链和 A 链由一个氨基酸残基或一种由 2-34 个氨基酸残基组成的肽段隔开。

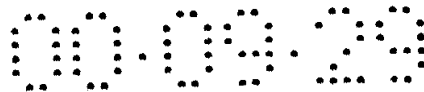
20 35. 权利要求 31 的方法，其中人胰岛素前体由 SEQ ID NO:5 的氨基酸序列组成。

36. 权利要求 30 的方法，其中 IMC 样肽段与胰岛素前体相比含有更高比例的带电氨基酸残基。

25 37. 权利要求 30 的方法，其中在 IMC 样肽段中，N 端一半包含的带正电的氨基酸残基多于带负电的氨基酸残基，而 C 端一半包含的带负电的氨基酸残基则多于带正电的氨基酸残基。

38. 权利要求 30 的方法，其中 IMC 样肽段由一段与含有人生长激素（hGH）蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 40% 的氨基酸序列组成，其中的同一性百分率由大小与 hGH 的区域相同的一段氨基酸序列决定。

30 39. 权利要求 38 的方法，其中 IMC 样肽段由一段与含有人生长激素（hGH）蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 60%



的氨基酸序列组成。

40. 权利要求 30 的方法，其中 IMC 样肽段能够被抗 hGH 抗体结合。

41. 权利要求 38 的方法，其中 IMC 样肽段由 SEQ ID NO:1 的氨基酸序列组成。

42. 权利要求 38 的方法，其中 IMC 样肽段由 SEQ ID NO:2 的氨基酸序列组成。

43. 权利要求 30 的方法，其中可切除氨基酸残基为 Arg 残基或 Lys 残基。

44. 权利要求 30 的方法，其中可切除氨基酸残基由 SEQ ID NO:3 的氨基酸序列组成。

45. 权利要求 30 的方法，其中在含有胰岛素前体的非正确折叠第二嵌合蛋白中，IMC 样肽段位于该嵌合蛋白的 N 端。

46. 权利要求 30 的方法，其中在含有胰岛素前体的非正确折叠第二嵌合蛋白中，IMC 样肽段位于该嵌合蛋白的 C 端。

47. 权利要求 30 的方法，其中在含有胰岛素前体的非正确折叠第二嵌合蛋白中，IMC 样肽段位于胰岛素前体的 B 链和 A 链之间。

48. 权利要求 30 的方法，其中 IMC 样肽段含有一个或多个可切除的氨基酸残基，这些残基与含有胰岛素前体的第二嵌合蛋白中隔离 IMC 样肽段和胰岛素前体的一个或多个可切除的氨基酸残基相同。

49. 权利要求 48 的方法，其中可切除氨基酸残基为一个 Arg 残基或一个 Lys 残基。

50. 权利要求 30 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白由 SEQ ID NO:6 的氨基酸序列组成。

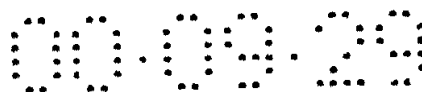
51. 权利要求 30 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白由 SEQ ID NO:7 的氨基酸序列组成。

52. 权利要求 30 的方法，其中离液辅助剂选自盐酸胍、碳酸亚乙酯、硫氰酸、二甲亚砜和脲。

53. 权利要求 52 的方法，其中的离液辅助剂为脲。

54. 权利要求 53 的方法，其中脲的浓度约为 2.0-8.0 M。

55. 权利要求 54 的方法，其中脲的浓度约为 3.0-6.0 M。



56. 权利要求 30 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白在 pH 约为 8.0-10.5 的水性介质中与至少一种离液辅助剂相接触，而含有胰岛素前体的非正确折叠第二嵌合蛋白的浓度约为 0.05-15g/L。

5 57. 权利要求 56 的方法，其中的 pH 值约维持在 9.0-10.0。

58. 权利要求 56 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白约为 0.5-5.0g/L。

59. 权利要求 58 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白约为 2.0-3.0g/L。

10 60. 权利要求 30 的方法，该方法进一步包括将含有胰岛素前体的非正确折叠第二嵌合蛋白与一定量的硫醇相接触，其中硫醇的量可为含有胰岛素前体的非正确折叠第二嵌合蛋白的每个半胱氨酸残基提供少于 5 个的硫醇基。

15 61. 权利要求 60 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白同时与硫醇和离液辅助剂相接触。

62. 权利要求 60 的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白依次与硫醇和离液辅助剂相接触。

20 63. 权利要求 60 的方法，其中硫醇用量为含有胰岛素前体的非正确折叠第二嵌合蛋白的每个半胱氨酸残基提供约 0.07-1.0 个硫醇基。

64. 权利要求 60 的方法，其中硫醇的选择集合包括二硫苏糖醇、二硫赤藓糖醇、2-巯基乙醇、半胱氨酸、巯基乙醇酸甲酯、3-巯基-1,2-丙二醇和 3-巯基丙酸。

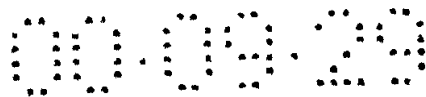
65. 权利要求 64 的方法，其中的硫醇为 2-巯基乙醇。

25 66. 权利要求 30 的方法，该方法进一步包括将含有胰岛素前体的正确折叠第一嵌合蛋白从含有胰岛素前体的非正确折叠第二嵌合蛋白中分离。

67. 权利要求 67 的方法，其中将含有胰岛素前体的第一嵌合蛋白从含有胰岛素前体的第二嵌合蛋白中分离的方法采用超滤法。

30 68. 权利要求 67 的方法，其中实施超滤的 pH 值约为 8.0-11.0。

69. 权利要求 68 的方法，其中实施超滤的 pH 值约为 9.0-10.0。



70. 权利要求 30 的方法的产物。

71. 一种对氨基酸序列促进胰岛素前体折叠的能力进行筛选的测定方法，该方法包括：

5 (a) 改变权利要求 1 的嵌合蛋白中第一肽段的氨基酸序列，获得具有所述改变的所述嵌合蛋白，在一定条件下于水性介质中将具有所述改变的所述嵌合蛋白与至少一种离液辅助剂接触足够长的时间，以使所述嵌合蛋白正确折叠，并测量具有所述改变的所述嵌合蛋白的折叠产量；

10 (b) 获得用于步骤 (a) 的但没有任何步骤 (a) 所述氨基酸序列改变的嵌合蛋白，在与步骤 (a) 相同的条件下于水性介质中将没有任何步骤 (a) 所述氨基酸序列改变的嵌合蛋白与相同于步骤 (a) 所用的离液辅助剂接触与步骤 (a) 相同的时间，并测量嵌合蛋白的折叠产量；以及

15 (c) 将分别在步骤 (a) 和步骤 (b) 中测量的嵌合蛋白折叠产量进行比较，

其中若步骤 (a) 测得的产量基本上等于或大于步骤 (b) 测得的产量，则表明该氨基酸序列可促进胰岛素前体的折叠。

72. 权利要求 71 的测定方法，其中的嵌合蛋白由 SEQ ID NO:6 的氨基酸序列组成。

20 73. 权利要求 71 的测定方法，其中的嵌合蛋白由 SEQ ID NO:7 的氨基酸序列组成。

74. 权利要求 71 的测定方法，其中的离液辅助剂为脲。

25 75. 权利要求 71 的测定方法，该方法进一步包括分别将步骤 a) 和步骤 b) 的嵌合蛋白与一定量的硫醇相接触，其中硫醇的量可为该嵌合蛋白的每个半胱氨酸残基提供少于 5 个的硫醇基。

76. 权利要求 75 的测定方法，其中的硫醇为 2-巯基乙醇。

77. 权利要求 71 的测定方法的产物。



说明书

含有分子内伴侣样序列的嵌合蛋白及其在胰岛素生产中的应用

1. 发明背景

5

1. 1. 技术背景

10

本发明涉及一种含有与目标蛋白相连接的分子内伴侣样 (IMC) 序列的嵌合蛋白。详细而言, 本发明涉及一种含有与一种胰岛素前体相连接的 IMC 样序列的嵌合蛋白。本发明还涉及一种用于获得含有胰岛素前体的正确折叠嵌合蛋白的方法, 该方法尤其包括将含有一种与胰岛素前体结合的 IMC 样序列的非正确折叠嵌合蛋白与至少一种离液辅助剂相接触。此外, 本发明还涉及一种测定法, 用于对一种氨基酸序列通过含有与胰岛素前体结合的 IMC 样序列的嵌合蛋白来促进胰岛素前体折叠的能力进行筛选。

15

1. 2. 背景知识

1. 2. 1. 分子内伴侣与蛋白折叠

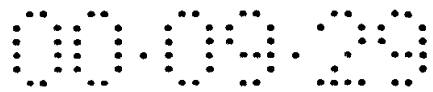
20

25

分子伴侣的定义是指能够协助其它多肽正确折叠但又不作为功能性组装结构的成分的一类蛋白 (Shinde and Inouye, TIBS, 1993, 18:442-446)。分子内伴侣 (IMCs) 是将要折叠的目标蛋白前体的一部分, 在缺少 IMCs 的情况下, 目标蛋白分子将不具有足够的信息来完成正确的自折叠 (Inouye, Enzyme, 1991, 45:314-321)。IMCs 的独特性质包括: a) IMC 与目标蛋白利用一个肽键形成单一多肽而相互连接; b) IMC 对目标蛋白的活性构象形成而言是绝对必需的, 但它不是目标蛋白的功能所必需的; c) 一旦蛋白折叠完成, IMC 即自动去除或由其它肽链内切酶切除; d) IMC 不具有催化剂作用, 即一分子 IMC 只能使一分子目标蛋白再折叠; 以及 e) IMC 是一种只对目标蛋白才发挥作用的具有高度特异性的“特制”多肽 (Inouye, Enzyme, 1991, 45:314-321)。

30

最近的研究表明, IMC 或前肽能够以分子间方式帮助目标蛋白折叠, 即 IMC 或前肽不是通过肽键与目标蛋白相结合, 而是作为独立的肽加入折叠反应 (美国专利号 5, 719, 021)。但值得注意的是, 美国



专利号 5,719,021 中使用的前肽是目标蛋白的天然前肽,或是与目标蛋白的功能相同且所含氨基酸序列与目标蛋白相似的多肽的前肽。此外,美国专利号 5,719,021 中描述的分子间反应必须是在有利于疏水相互作用的水性离子缓冲介质中进行。

5 IMCs 的实例包括枯草杆菌蛋白酶前肽、 α -细胞裂解蛋白酶前肽、羧肽酶 Y 前肽和泛素前肽 (Shinde and Inouye, TIBS, 1993, 18:442-446)。枯草杆菌蛋白酶 IMC 序列的特征有: a) 与目标蛋白相比, IMC 含有更高比例的带电氨基酸残基; b) 这些带电残基在 IMC 内的分布极不均匀,即 N 端部分包含的带正电的残基多于带负电的残基,而 C 端部分包含的带负电的残基则多于带正电的残基; c) IMC 10 内的丝氨酸和苏氨酸残基分布也不均匀; d) IMC 具有高反应性含量的芳香族残基; e) IMC 含有一段 9 个残基的疏水序列 (Inouye, Enzyme, 1991, 45:314-321)。在 α -细胞裂解蛋白酶和羧肽酶 Y 中也发现了类似的对带电残基的倾向性 (Inouye, Enzyme, 1991, 45:314-321)。

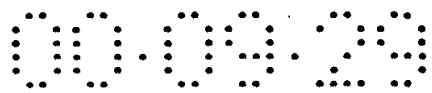
15

1. 2. 2. 成熟的人生长激素的氨基酸序列

成熟的人生长激素 (hGH) 的氨基酸序列公开于 Ikehara et al., Proc. Natl. Acad. Sci. USA, 1984, 81:5956-5960。目前该领域还没有提出成熟的 hGH 或其任何部分可具有 IMC 或前肽的作用。事实上,成熟的 hGH 或其任何部分根本不能被认为是一种前肽,因为根据 20 定义,任何前序列、原序列或前原序列都要从成熟序列中去除。

1. 2. 3. 人胰岛素的结构

胰岛素是一种已知氨基酸序列和结构特征的明确定义的肽 25 (Watson et al., Recombinant DNA--A Short Course; Scientific American Books, W.H. Freeman Co., New York, 1983, pp.231-235; Norman and Litwack, In Hormones, Academic Press, New York, 1987, pp.264-317)。该激素含有 A 链 (21 个氨基酸) 和 B 链 (30 个氨基酸) 两种独立的肽链,如图 1B 所示,两条肽链通过二硫键相结合。胰岛素原是胰岛素的生物前体,它是由 C 肽将 A 链和 B 链连接 30 起来形成的单一链肽 (图 1A)。



1. 2. 4. 由重组方法生产人胰岛素

人胰岛素是第一种在细菌内制备的与人胰岛素肽的序列相同的动物蛋白 (Watson et al., Recombinant DNA--A Short Course; Scientific American Books, W.H. Freeman Co., New York, 1983, pp. 231-235)。第一次在实验室中成功地表达出人胰岛素是在 1978 年, 批准将人胰岛素作为治疗药物则是在 1982 年 (Johnson, Science, 1983, 219:632-637)。

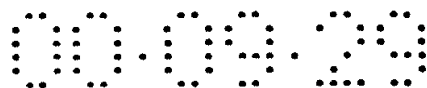
10

1. 2. 4. 1. 双链法

该方法是利用由包含人胰岛素 A 链或 B 链 DNA 编码序列的质粒分别转化的大肠杆菌经独立发酵而将每条胰岛素链作为 β -半乳糖苷酶 (β -gal) 融合蛋白来加以制备。产物处在细胞内并且主要出现在胞质包涵体中 (Williams et al., Science, 1982, 215(5):687-689)。大肠杆菌中产生的重组蛋白通常占总蛋白的 10-40% (Burgess, Protein Engineering; Oxender, D.L., Fox, C.F., Eds.; Alan R. Liss, Inc.; New York, 1987; pp. 71-82)。

从包涵体中提出融合蛋白, 用 CNBr 对 β -半乳糖苷酶与 A 链或 B 链间的 Met 残基进行化学裂解, 纯化以获得分离的 A 链和 B 链。然后于存在有限量硫醇的条件下将肽以 A-B 链为 2:1 的比例混合并诱导折叠 (S-磺酸化形式), 以获得有活性的激素 (Chance et al., In Peptides: Synthesis-Structure-Function, Rich D.M. Gross, E., Eds., Pierce Chemical Co., Rockford, II 1981, pp. 721-728; Frank and Chance, In Quo Vadis? Therapeutic Agents Produced by Genetic Engineering, Joyesuk et al., Eds., Sanoff Group, Toulouse-Labege, France, 1985, pp137-148)。24 小时后的产率根据所用 B 链的量可约为 60% (Chance et al., In Insulins, Growth Hormone and Recombinant DNA Technology, Raven Press, New York, 1981, pp. 71-85; Johnson, Fluid Phase Equilib., 1986, 29:109-123)。Goeddel et al., Proc. Natl. Acad. Sci. U.S.A., 1979, 76(1):106-110 中获得了类似的结果, 其表达的 A 链或 B 链融

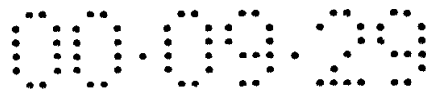
25
30



合蛋白占细胞总蛋白的 20%。在随后的 S-磺酰化链的折叠过程中可产生 50-80% 的正确折叠。

β -gal 融合蛋白的大尺寸可限制其产量，这是因为 β -gal (-1000 个氨基酸) 与胰岛素 A 链或 B 链 (分别为 21 或 30 个氨基酸) 形成的融合蛋白可从细胞核糖体解离 (链在翻译过程中过早终止)，从而产生不完整的胰岛素肽 (Burnett, *Experimental Manipulation of Gene Expression*, Inouye, Ed., Academic Press, New York, 1983, pp. 259-277; Hall, *Invisible Frontiers--The Race to Synthesize a Human Gene*, Atlantic Monthly Press, New York, 1987)。该方法的关键改进在于它用色氨酸 (Trp) 操纵子替代乳糖操纵子 (β -gal 系统) 以获得较小的融合蛋白。Trp 操纵子包含一系列的 5 个细菌基因，这些基因依次合成色氨酸合成代谢所需的酶。与 β -gal 的 1000 个氨基酸相比，这些酶之一的 TrpE 只含有 190 个氨基酸。由于 Trp 启动子在大肠杆菌中属于强启动子，因此在胰岛素 A 链或 B 链基因前接 TrpE 基因的优势是能够使总蛋白中融合蛋白的产量从 5-10% 增加到 20-30% (Hall, *Invisible Frontiers--The Race to Synthesize a Human Gene*, Atlantic Monthly Press, New York, 1987)。这将使多肽的表达量与 lac (即 β -gal) 系统相比至少增加 10 倍 (Burnett, *Experimental Manipulation of Gene Expression*, Inouye, Ed., Academic Press, New York, 1983, pp. 259-277)。当大肠杆菌发酵缺乏色氨酸时，Trp 操纵子即启动 (Hall, *Invisible Frontiers--The Race to Synthesize a Human Gene*, Atlantic Monthly Press, New York, 1987; Etienne-Decent, In *Genetic Biochemistry: From Gene to Protein*, Ellis Horwood Limited, Chichester, U. K., 1988, pp. 125-127)。由于细胞团块能够首先达到最大，因此这种特性在发酵其间十分有利。然后可在适当时机使发酵培养基变为色氨酸缺乏，以此来启动细胞的胰岛素生产系统。

发酵结束后，回收细胞并将其裂解。然后将细胞碎片与包涵体分离，并将包涵体溶于一种溶剂，尽管对其细节还不是很清楚 (Wheelwright, *Protein Purification*, Oxford University Press; New York, 1991, pp. 217)。有时是将包涵体溶于 6M 盐酸胍和 0.1M



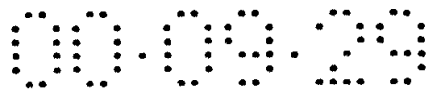
二硫苏糖醇中 (Burgess, Protein Engineering, Oxender and Fox, Eds., Alan R. Liss, Inc., New York, 1987, pp. 71-82)。接着将 Trp-LE-Met-A 链和 Trp-LE-Met-B 链经 CNBr 裂解产生胰岛素 A 链和 B 链。对 A 链和 B 链的进一步修饰包括：氧化亚硫酸盐解
5 [sulfitolysis]、纯化并结合，产生粗制胰岛素。粗制胰岛素经离子交换、大小排阻层析及反相高效液相层析 (RP HPLC) 产生纯化的重组人胰岛素 (Frank and Chance, Munch Med. Wschr, 1983, 125(Suppl. 1):514-520)。

10 1. 2. 4. 2. 胰岛素原法 (胞内型)

人胰岛素也可利用能够产生完整胰岛素原的重组微生物来产生，以替代分别产生 A 链和 B 链的方法 (Kroeff et al., J. Chromatogr, 1989, 481:45-61)。首先将 mRNA 复制成 cDNA，再利用化学方法合成甲硫氨酸密码子并将其连接在胰岛素原 cDNA 的 5'
15 端。将 cDNA 插入能够被引入并在大肠杆菌中生长的质粒载体的细菌基因中。通过破坏甲硫氨酸连接可使胰岛素原从细菌酶 (β -gal) 片段 (或作为选择从 Trp-LE/Met 胰岛素原 (Trp 胰岛素原)) 中释放。胰岛素原链经过折叠可形成正确的分子内二硫键，然后用酶将 C 肽切除，从而产生人胰岛素 (Frank and Chance, Munch Med. Wschr, 1983, 125(Suppl. 1):514-520)。比较而言，前述的双链法要更加复杂。
20

Dorschug 等人构建了编码含有小胰岛素原 (B-Arg-A) 的融合蛋白的重组质粒，并在大肠杆菌 (包涵体形式) 和酵母 (分泌形式) 中表达了融合蛋白，然后经 CNBr 裂解和氧化亚硫酸盐解制备出正确折叠的小胰岛素原，继而用胰蛋白酶和羧肽酶 B 处理，使正确折叠的小胰岛素原转化为人胰岛素 (EP 0, 347, 781 B1; IL 9, 562, 511 B 和 AU
25 611, 303 B2)。

Tottrup and Carlson, Biotechnol. Bioeng, 1990, 35:339-348 中则采用以一种优化的分批培养发酵方式培养的酵母系统，据报导，超氧化物歧化酶-人胰岛素原融合蛋白 (SOD-PI) 的产量为
30 1500mg/L。SOD-PI 将作为重组人胰岛素市场的原材料；但终产物的产量未见报导。



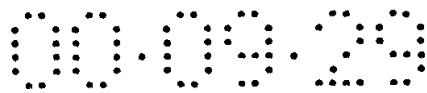
最近, Castellanos-Serra et al., FEBS Letters, 1996, 378:171-176 在大肠杆菌中表达了一种胰岛素原融合蛋白, 该融合蛋白带有一段经赖氨酸残基与胰岛素原 N 端相连的修饰的白介素-2 N 端肽 (1-22 位氨基酸残基)。从包涵体中分离融合胰岛素原, 并通过氧化亚硫酸盐解进行再折叠, 然后用胰蛋白酶和羧肽酶 B 经长时间的反应使其转化为正确的融合蛋白胰岛素。IL2-胰岛素原融合蛋白能够在无需首先去除 IL2 片段、甚至不需要去除 CNBr 以及相关纯化步骤的情况下正确地折叠。但氧化亚硫酸盐解及相关纯化步骤在使用 IL2-胰岛素原融合蛋白的方法中不能取消。

10

1. 2. 4. 3. 胰岛素原方法 (分泌型)

Villa-Komaroff et al., Proc. Natl. Acad. Sci. U.S., 1978, 75(8):3727-3731 中首次描述了大肠杆菌内用于人胰岛素原的分泌系统。Thim 等人构建了编码含有一种修饰的酵母配对因子 $\alpha 1$ 前导序列和一种胰岛素前体的融合蛋白的重组质粒 (Thim et al., Proc. Natl. Acad. Sci. USA, 1986, 83:6766-6770)。该前导序列能够将融合蛋白引入酵母细胞的分泌途径并使融合蛋白受到 Lys-Arg 加工酶系统的作用。在胰岛素原和其它含有短间隔肽 (6 个或更多氨基酸残基) 以替代 C 肽的胰岛素前体内, A 链和 B 链之间的一个或两个二元序列上还出现部分加工。但在胰岛素前体分子内缺乏间隔肽的情况下, 如 B-Arg-Arg-A (其中 A 和 B 分别为人胰岛素原的 A 链和 B 链), 则未观测到加工过程。这种类型的单链胰岛素前体能够经胰蛋白酶和羧肽酶 B 处理, 在酶的作用下转化为胰岛素。在 Diers et al., Drug Biotechnology Regulations (Scientific Basis and Practices), Chiu and Gueriguian, Eds., Marcel Dekker, Inc., New York, 1991, pp. 167-177 中对一种未折叠肽进行了描述, 即一段前导肽或原肽段, 接着是一段 Lys-Arg 序列、B 链 (1-29 位氨基酸)、一种短肽桥, 然后接 A 链 (1-21 位氨基酸)。在此前体中, 胰岛素 B 链的 29 位氨基酸通过一种含有一个与 A 链靠近的碱性氨基酸的短连接肽与 A 链的 1 位氨基酸连接。人胰岛素的产生是利用转肽作用然后将形成的酯键水解。进一步纯化则还需一些层析步骤。

30



1. 2. 5. 胰岛素前体的折叠

人胰岛素是一种具有两条氨基酸链的共 51 个氨基酸残基的蛋白。两条氨基酸链上共有 6 个半胱氨酸残基，每条链上各有两个半胱氨酸残基通过二硫键相互连接。从统计学角度来看，一个人胰岛素分子内形成的二硫键有 15 种可能性。但这 15 种可能性中只有一种存在于具有生物活性的人胰岛素中，其二硫键如下：1) A6-A11；2) A7-B7；3) A20-B19。

人胰岛素中的二硫键形成是利用一种中间方式来实现，即为人胰岛素的半胱氨酸提供一种硫保护性基团，如 S-磺酸根 (-S-SO₃) 基团 (EP 0, 037, 255)。此外，也有使用其半胱氨酸残基以巯基 (-SH) 形式存在的猪胰岛素原来获得具有正确连接的半胱氨酸桥的胰岛素原 (Biochemistry, 1968, 60:622-629)。Obermeier 等人描述了一种在存在硫醇、离液辅助剂和疏水性吸附树脂的情况下由浓度为 0.05-0.3g/L 的相应胰岛素原氨基酸链获得半胱氨酸桥正确连接的胰岛素原的方法 (美国专利号 5, 473, 049)。在美国专利号 5, 473, 049 描述的方法中删去了氧化亚硫酸盐解步骤。但是胰岛素蛋白只能在低浓度下折叠，这在很大程度上降低了该方法的商业价值。此外，大量硫醇和疏水性吸附树脂的使用增加了处理的复杂性以及下游纯化过程的成本。从美国专利号 5, 473, 049 的公开内容中无法确定删去氧化亚硫酸盐解步骤所带来的利益是否会超出其增加的下游纯化成本。

上文对参考文献的引用并非是承认这些参考文献在技术上优先于本发明。

2. 发明概述

本发明涉及一种嵌合蛋白，该蛋白自 N 端至 C 端包含：a) 一种第一肽段，该肽段包含一段与含有人生长激素 (hGH) 蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 40% 的氨基酸序列，其中同一性百分率由大小与 hGH 的区域相同的一段氨基酸序列决定；b) 一个精氨酸，或一个赖氨酸，或一种至少含有 2 个氨基酸的第二肽段，其中该肽段的最 C 端氨基酸残基为 Arg 或 Lys；c) 一种第三肽段，该肽段



包含一段含有 2 个以上半胱氨酸的氨基酸序列，并且该肽段不是 hGH 蛋白的一部分。更详细而言，本发明涉及一种其第三肽段为一种胰岛素前体的嵌合蛋白。

5 本发明还涉及一种获得含有胰岛素前体的正确折叠第一嵌合蛋白的方法，该方法包括在一种水性介质中将含有胰岛素前体的非正确折叠第二嵌合蛋白与至少一种离液辅助剂相接触，其中所述含有胰岛素前体的第二嵌合蛋白包含一种经一个或多个可切除氨基酸残基与胰岛素前体隔离的分子内伴侣（IMC）样肽段；其中所述 IMC 样肽段：
10 a) 含有约 20-200 个氨基酸残基； b) 不是胰岛素前体或其一部分；
并且 c) 可促进胰岛素前体的折叠，以至于将含有胰岛素前体的非正确折叠第二嵌合蛋白与离液辅助剂接触时，含有胰岛素前体的正确折叠第一嵌合蛋白的产量要高于将不含有所述 IMC 样肽段的非正确折叠胰岛素前体与同样的离液辅助剂接触时的正确折叠胰岛素前体的产量。

15 本发明另涉及一种对氨基酸序列促进胰岛素前体折叠的能力进行筛选的测定方法，该方法包括：(a) 改变在 4.2 节中公开的含有胰岛素前体的嵌合蛋白的第一肽段的氨基酸序列，获得具有所述改变的所述嵌合蛋白，在一定条件下于水性介质中将具有所述改变的所述嵌合蛋白与至少一种离液辅助剂接触足够长的时间，以使所述嵌合蛋白
20 正确折叠，并测量具有所述改变的所述嵌合蛋白的折叠产量；(b) 获得用于步骤 (a) 的但没有任何步骤 (a) 所述氨基酸序列改变的嵌合蛋白，在与步骤 (a) 相同的条件下于水性介质中将没有任何步骤 (a) 所述氨基酸序列改变的嵌合蛋白与相同于步骤 (a) 所用的离液辅助剂接触与步骤 (a) 相同的时间，并测量嵌合蛋白的折叠产量；以及 (c) 将分
25 别在步骤 (a) 和步骤 (b) 中测量的嵌合蛋白折叠产量进行比较，其中若步骤 (a) 测得的产量基本上等于或大于步骤 (b) 测得的产量，则表明该氨基酸序列可促进胰岛素前体的折叠。

3. 附图简述

30 图 1A 和 1B. 胰岛素原以及正确形成二硫键的成熟胰岛素的结构。1A 描绘胰岛素原结构。1B 描绘正确形成二硫键的成熟胰岛素的结构。

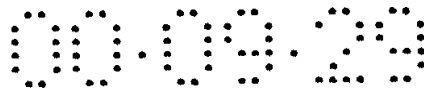


图 2. hGH-小胰岛素原(mini-proinsulin) (SEQ ID NO:6) 表达载体 (pZRhi-1) 图谱。

4. 发明详述

5 重组方法使我们有可能在微生物中生产人胰岛素原、或氨基酸序列和/或氨基酸链的长度与天然人胰岛素不同的胰岛素原。在大肠杆菌等微生物中生产人胰岛素原或其衍生物的一个主要问题是胞内降解 (Ladisich and Kohlmann, *Biotechnol. Prog.*, 1992, 2:469-478)。此外, 利用重组方法在微生物中生产的人胰岛素原或其衍生物不具有正确连接的半胱氨酸桥 (美国专利号 5, 473, 049)。

10 在本发明之前已了解的一种由重组方法获得人胰岛素的方法是基于以下步骤: 1) 用表达含有人胰岛素原或其衍生物的融合蛋白的载体转化微生物并将该微生物发酵; 2) 细胞裂解; 3) 分离融合蛋白; 4) 用溴化氰裂解融合蛋白; 5) 分离具有胰岛素原序列的裂解产物; 15 6) 氧化亚硫酸盐解; 7) 形成正确连接的半胱氨酸桥; 8) 胰岛素原脱盐; 9) 将具有正确连接的半胱氨酸桥的胰岛素原层析纯化; 9) 胰岛素原溶液浓缩; 10) 将浓缩的胰岛素原溶液层析纯化; 11) 将胰岛素原酶解以获得人胰岛素; 以及 12) 将所得人胰岛素层析纯化 (EP 0, 055, 945)。该方法的缺点是步骤繁琐, 并且纯化过程的损失大, 20 导致胰岛素产量低。在分离的融合蛋白经溴化氰裂解、氧化亚硫酸盐解以及胰岛素原纯化的过程中, 胰岛素原的损失可能会高达 40% (EP 0, 055, 945)。反之, 如果能够显著减少必要步骤的数量, 重组生产的胰岛素或其衍生物的产量将显著提高。

本发明的目标是开发一种重组方法, 用于获得具有正确连接的半胱氨酸桥的人胰岛素, 该方法所需步骤较少, 从而可获得较高产量的人胰岛素。本发明的另一目标是开发一种可在上述方法中使用的含有胰岛素前体的嵌合蛋白。本发明的又一目标则是开发一种测定方法, 用于筛选经肽键与胰岛素前体连接时可促进胰岛素前体折叠的氨基酸序列。

30 本申请人已找到一些肽序列, 这些序列不但能够使胰岛素序列免受微生物宿主的胞内降解, 而且与当时已有的人胰岛素表达系统相比



具有以下优点：当通过肽键与胰岛素前体连接时，1) 可促进融合胰岛素前体的折叠；2) 可提高融合蛋白的溶解性并降低融合蛋白的分子间相互作用，从而使融合胰岛素前体能够在具有商业意义的高浓度下折叠；3) 无需溴化氰裂解、氧化亚硫酸盐解以及相关纯化步骤；

5 以及 4) 无需使用高浓度的硫醇或疏水性吸附树脂。

本申请人惊奇地发现，本发明的目标可通过将一种 IMC 样序列经一个或多个可切除氨基酸残基与胰岛素前体相连接来加以实现。IMC 样序列具有某些 IMC 序列的特性，如协助目标蛋白折叠、与其目标蛋白相比含有更高比率的带电氨基酸残基、带电氨基酸残基的两极化分布、以及具有一段似乎是为目标蛋白“定制”的序列。但本发明使用的 IMC 样序列在某些重要方面与 IMC 序列不同。首先，IMC 样序列与目标蛋白属于不同种，即不是目标蛋白的前肽。举例来说，如果胰岛素前体是将要折叠的目标蛋白，那么 IMC 样序列就不是胰岛素前体或其一部分。其次，IMC 样序列的大小约为 20-200 个氨基酸残基。

10

此外，与早先的技术学说 (Castellanos-serra et al., FEBS Letters, 1996, 378:171-176) 不同的是，本申请人惊奇地发现，若 IMC 样序列中的一个或多个可切除氨基酸残基与隔离 IMC 样序列和胰岛素前体的一个或多个可切除氨基酸残基相同，则能够在折叠后将 IMC 样序列片段去除，从而简化下游的纯化步骤。

15

为清楚地说明本公开内容，而并非作为其限制，本发明详述将分为以下几个小节。

20

4. 1. 编码 4. 2. 节公开的嵌合蛋白的核酸

本发明提供一种分离的核酸，该核酸含有编码 4. 2 节公开的嵌合蛋白的核酸序列。

25

在一项具体实施方案中，本发明提供一种分离的核酸，该核酸含有编码具有 SEQ ID NO:6 的氨基酸序列的嵌合蛋白的核酸序列。

在另一项特定实施方案中，本发明提供一种分离的核酸，该核酸含有编码具有 SEQ ID NO:7 的氨基酸序列的嵌合蛋白的核酸序列。

在一项优选实施方案中，本发明提供一种分离的 DNA 分子，该 DNA 分子含有编码 4. 2 节公开的嵌合蛋白的核苷酸序列。

30



在另一项优选实施方案中，本发明提供一种分离的核酸，该核酸含有与编码 4.2 节公开的嵌合蛋白的核苷酸序列互补的核苷酸序列。

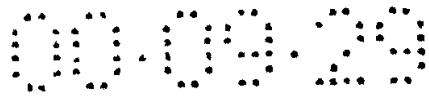
在又一项优选实施方案中，本发明提供一种分离的核酸，该核酸能够与编码 4.2 节公开的嵌合蛋白的 DNA 中的第一、第二和第三肽段的编码核苷酸序列杂交。

含有编码 4.2 节公开的嵌合蛋白、或其任何片段、类似物或衍生物的核苷酸序列的核酸可利用任何该领域已知的方法获得。核酸也可完全利用化学方法合成。作为选择，编码嵌合蛋白各个片段，即第一、第二或第三肽段，的核酸可利用分子克隆方法获得或从预定细胞中纯化获得。然后可通过化学或酶学方法将编码嵌合蛋白各个片段的核酸连接在一起，形成含有编码 4.2 节公开的嵌合蛋白、或其任何片段、类似物或衍生物的核苷酸序列的核酸。

任何人类细胞均可作为 hGH 核酸分离的核酸来源。任何哺乳动物细胞均可作为胰岛素核酸分离的核酸来源。胰岛素的核酸编码序列的分离可来源于哺乳动物、人类、猪、牛、猫、鸟类、马、犬类以及其它啮齿类或灵长类动物等。

该 DNA 可利用该领域已知的标准方法从 DNA 克隆(如 DNA“文库”)中获得，或通过化学合成、cDNA 克隆获得、或通过由预定细胞纯化的基因组 DNA 或其片段的克隆获得(例如见, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D. M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U. K. Vol. I, II.)。来源于基因组 DNA 的克隆除编码区外可能还包含 DNA 调控区和内含子区；来源于 cDNA 的克隆则只包含外显子序列。无论哪种来源的基因均需要将该分子克隆到适当载体中进行基因扩增。

在来源于 cDNA 的基因的分子克隆过程中，cDNA 可通过该领域所熟知的方法由细胞总 RNA 或 mRNA 产生。该基因也可由基因组 DNA 获得，其产生的 DNA 片段(如利用限制性内切酶或通过机械剪切)中将有一些可编码预定基因。然后可根据大小利用标准技术将线性 DNA 片段分离，这些技术包括但不局限于琼脂糖凝胶电泳和聚丙烯酰胺凝胶



电泳以及柱层析。

一旦获得含有编码 4.2 节公开的嵌合蛋白、或其任何片段、类似物或衍生物的核苷酸序列的核酸，即可通过核酸测序（可利用该领域所熟知的任何方法）并与已知序列对比来证实其同一性。DNA 序列分析的实施可采用该领域已知的任何技术，这些技术包括但不限于 Maxam and Gilbert 法（Maxam and Gilbert, 1980, Meth. Enzymol. 65:499-560）、Sanger 双脱氧法（Sanger et al., 1977, Proc. Natl. Acad. Sci. U.S.A., 74:54-63）、使用 T7 DNA 聚合酶（Tabor and Richadson, 美国专利号 4,795,699）、使用 DNA 自动测序仪（如 Applied Biosystems, Foster City, CA）或 PCT Publication WO 97/15690 描述的方法。

能够与含有编码 4.2 节公开的嵌合蛋白、或其任何片段、类似物或衍生物的核苷酸序列的核酸杂交的核酸可在低、高、或中等严格条件下通过核酸杂交来分离（也可参考 Shilo and Weinberg, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:6789-6792）。

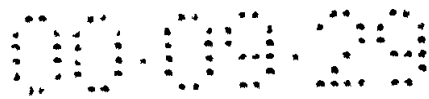
4. 2. 嵌合蛋白

本发明提供一种嵌合蛋白，该蛋白从 N 端至 C 端包含：a) 一种第一肽段，该肽段由一段与含有人生长激素（hGH）蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 40% 的氨基酸序列组成，其中同一性百分率由大小与 hGH 的区域相同的一段氨基酸序列决定；b) 一个 Arg 残基，或一个 Lys 残基，或一种至少含有 2 个氨基酸的第二肽段，其中该肽段的最 C 端氨基酸残基为 Arg 或 Lys；c) 一种第三肽段，该肽段由一段含有 2 个以上半胱氨酸残基的氨基酸序列组成，并且该肽段不是 hGH 蛋白的一部分。

在一优选实施方案中，本发明提供一种上述的嵌合蛋白，其中第一肽段由一段与 hGH 蛋白的区域的同一性至少为 60% 的氨基酸序列组成。

在另一优选实施方案中，本发明提供一种上述的嵌合蛋白，其中第一肽段能够被抗 hGH 抗体结合。

在一更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其



中第一肽段由 SEQ ID NO:1 的氨基酸序列组成。

在另一项更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其中第一肽段由 SEQ ID NO:2 的氨基酸序列组成。

5 在一优选实施方案中，本发明提供一种上述的嵌合蛋白，其中第二肽段由 SEQ ID NO:3 的氨基酸序列组成。

在一具体实施方案中，本发明提供一种上述的嵌合蛋白，其中第三肽段为胰岛素前体。

在一优选实施方案中，本发明提供一种上述的嵌合蛋白，其中胰岛素前体来源于人。

10 在一更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其中人胰岛素前体能够被抗人胰岛素抗体结合。

在另一更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其中人胰岛素前体含有 SEQ ID NO:4 的氨基酸序列。

15 在又一更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其中在胰岛素前体中的人胰岛素前体 B 链和 A 链由一个氨基酸残基或一种由 2-34 个氨基酸残基组成的肽段隔开。

在又一更优选的实施方案中，本发明提供一种上述的嵌合蛋白，其中人胰岛素前体由 SEQ ID NO:5 的氨基酸序列组成。

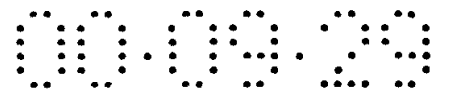
20 在一最优选的实施方案中，本发明提供一种由 SEQ ID NO:6 的氨基酸序列组成的嵌合蛋白。

在另一最优选的实施方案中，本发明提供一种由 SEQ ID NO:7 的氨基酸序列组成的嵌合蛋白。

4. 3. 获得 4.2 节公开的嵌合蛋白

25 4.2 节公开的嵌合蛋白及其衍生物、类似物和片段可通过该领域已知的方法获得，这些方法包括但不局限于重组表达法、从天然资源中纯化，以及化学合成。

30 举例来说，4.2 节公开的嵌合蛋白可通过重组蛋白表达技术获得。重组表达时，将编码 4.2 节公开的嵌合蛋白的基因或其部分插入到一种适合于在特定宿主细胞中表达的克隆载体中。有大量的该领域已知的载体-宿主系统可以使用。可能的载体包括，但不局限于，质



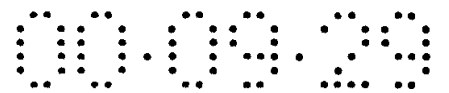
粒或已修饰的病毒，但这些载体系统必须与所用的宿主细胞相容。这些载体包括，但不局限于， λ 噬菌体衍生物等噬菌体、或 pBR332 或 pUC 质粒衍生物等质粒，或 Bluescript 载体 (Stratagene)。插入克隆载体可通过，例如，将 DNA 片段与带有互补粘性末端的克隆载体相连接的方法来实现。但如果克隆载体上不存在与切碎该 DNA 的限制位点相互补的位点，则可通过酶学方法对 DNA 分子的末端进行修饰。作为选择，也可通过在 DNA 末端连接核苷酸序列 (衔接物) 的方法来产生所需的任何位点；这些连接的衔接物可包含经化学方法合成的编码限制性内切酶识别序列的特定寡核苷酸。通过转化、转染、感染、电穿孔等方法将重组分子引入宿主细胞，以产生该基因的多个拷贝。

在一种可选的方法中，可以在“鸟枪法”将预定基因插入适当克隆载体后对其进行鉴定和分离。在插入克隆载体之前，预定基因可通过，例如大小分部分离法，进行富集。

在特定实施方案中，用结合了编码 4.2 节公开的嵌合蛋白的分离基因的重组 DNA 分子、cDNA、或合成的 DNA 序列转化宿主细胞，可产生该基因的多个拷贝。因此，通过转化体培养、从转化体中分离重组 DNA 分子，以及在必要时从分离的重组 DNA 中回收插入基因的方法可获得大量该基因。

编码 4.2 节公开的嵌合蛋白及其衍生物、类似物和片段的核苷酸序列或其具有功能活性的类似物或片段或其它衍生物可被插入到适当的表达载体，即含有转录并翻译插入的蛋白编码序列所必需的元件的载体，中。表达蛋白编码序列可采用各种载体-宿主系统。这些系统包括，但不局限于，用病毒 (如痘苗病毒、腺病毒等) 感染的哺乳动物细胞系统；用病毒 (如杆状病毒) 感染的昆虫细胞系统；诸如含有酵母载体的酵母等微生物，或用噬菌体 DNA、质粒 DNA、或粘粒 DNA 转化的细菌。载体的表达元件在长度和特异性上各不相同。根据所用的宿主-载体系统可采用多种适当的转录和翻译元件中的任何一种。

任何前述的用于将 DNA 片段插入载体的方法都可用来构建含有嵌合基因并包含适当转录/翻译调控信号和蛋白编码序列的表达载体。这些方法包括体外重组体 DNA 和合成技术以及体内重组体 (基因重组)。编码 4.2 节公开的嵌合蛋白及其衍生物、类似物和片段的核酸



序列的表达可由第二种核酸序列调控，以便在重组 DNA 分子转化的宿主细胞中表达 4.2 节公开的嵌合蛋白。举例来说，4.2 节公开的嵌合蛋白的表达可利用该领域已知的任何启动子/增强子元件来加以调控。可用来调控 4.2 节公开的嵌合蛋白的表达的启动子包括，但不局
5 限于，SV40 早期启动子区 (Bernoist and Chambon, 1981, *Nature* 290:304-310)、Rous 肉瘤病毒 3' 长末端重复中包含的启动子 (Yamamoto, et al., 1980, *Cell* 22:787-797)、疱疹胸苷激酶启动子 (Wagner et al., 1981, *Proc. Natl. Acad. Sci. U.S.A.*, 78:1441-1445)、金属硫蛋白基因调控序列 (Brinster et al., 1982, 10 *Nature* 296:39-42); β -内酰胺酶启动子等原核表达载体 (Villakamaroff, et al., 1978, *Proc. Natl. Acad. Sci. U.S.A.*, 75:3727-3731)、tac 启动子 (DeBoer, et al., 1983, *Proc. Natl. Acad. Sci. U.S.A.*, 80:21-25); 或 Trp E 启动子 (Hall, *Invisible Frontiers--The Race to Synthesize a Human Gene*, Atlantic 15 Monthly Press, New York, 1987) 也可参考 *Science American*, 1980, 242:74-94 中的“重组细菌中的有益蛋白”; 来自酵母或其它真菌的启动子元件, 如 Gal4 启动子、ADC (乙醇脱氢酶) 启动子、PKG (磷酸甘油激酶) 启动子、碱性磷酸酶启动子, 以及具有组织特异性并用于转基因动物的下列动物转录调控区: 在胰腺腺泡细胞中具有活性的 20 弹性蛋白酶 I 基因调控区 (Swift et al., 1984, *Cell* 38:639-646; Ornitz et al., 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409; MacDonald, 1987, *Hepatology* 7:425-515); 在胰腺 β 细胞中具有活性的胰岛素基因调控区 (Hanahan, 1985, *Nature* 315:115-122)、在淋巴细胞中具有活性的免疫球蛋白基因调控区 25 (Grosschedl et al., 1984, *Cell* 38:647-658; Adams et al., 1985, *Nature* 318:533-538; Alexander et al., 1984, *Mol. Cell. Biol.* 7:1436-1444)、在睾丸、乳房、淋巴和肥大细胞中具有活性的小鼠乳癌病毒调控区 (Leder et al., 1986, *Cell* 45:485-495)、在肝脏中具有活性的白蛋白基因调控区 (Pinkert et al., 1987, *Gene and 30 Devel.* 1:268-276)、在肝脏中具有活性的 α 胎蛋白基因调控区 (Krumlauf et al., 1985, *Mol. Cell. Biol.* 5:1639-1648; Hammer

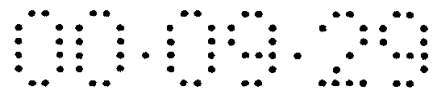


et al., 1987, *Science* 235:53-58)、在肝脏中具有活性的 α 1-抗胰蛋白酶基因调控区 (Kelsey et al., 1987, *Gene and Devel.* 1:161-171)、在骨髓细胞中具有活性的 β -球蛋白基因调控区 (Mogram et al., 1985, *Nature* 315:338-340; Kollias et al., 1986, *Cell* 46:89-94)、在脑的少突胶质细胞中具有活性的髓鞘碱性蛋白基因调控区 (Readhead et al., 1987, *Cell* 48:703-712)、在骨骼肌中具有活性的肌球蛋白轻链-2 基因调控区 (Sani, 1985, *Nature* 314:283-286)、以及在下丘脑中具有活性的促性腺释放激素基因调控区 (Mason et al., 1986, *Science* 234:1372-1378)。

10 举例来说, 可用载体可包含一种经操作与编码 4.2 节公开的嵌合蛋白的核酸相连接的启动子、一个或多个复制区、以及任选的一个或多个可选标记 (如抗生素抗性基因)。

含有编码 4.2 节中公开的嵌合蛋白或其衍生物、类似物或片段的基因插入片段的表达载体可通过三种常用方法鉴定: a) 核酸杂交, 15 b) “标记”基因的功能的出现或缺失, 以及 c) 插入序列的表达。在第一种方法中, 利用含有与编码 4.2 节公开的嵌合蛋白的插入基因同源的序列的探针经核酸杂交可检测出编码 4.2 节公开的嵌合蛋白的基因的存在。在第二种方法中, 根据在载体中插入编码 4.2 节公开的嵌合蛋白的基因而导致特定“标记”基因的功能 (如胸苷激酶活性、抗 20 生素抗性、转化表型、杆状病毒中形成包涵体等) 的出现或损失可鉴定并筛选重组载体/宿主系统。例如, 若编码 4.2 节公开的嵌合蛋白的基因被插入到载体的标记基因序列中, 即可利用该标记基因功能的缺失来鉴定含有编码 4.2 节公开的嵌合蛋白的插入序列的重组体。在第三种方法中, 可通过测定重组体表达的嵌合蛋白产物来鉴定重组表 25 达载体。这些测定方法可基于, 例如, 4.2 节公开的嵌合蛋白在体外测定体系中的物理或功能性质, 如与抗 hGH 或抗胰岛素抗体结合。

一旦鉴定并分离出特定的重组 DNA 分子, 即可利用该领域已知的几种方法将其扩增。一旦建立起适当的宿主系统和培养条件, 即可将重组表达载体扩增并大量制备。如前所述, 可用的表达载体包括, 但 30 不局限于, 以下载体或其衍生物: 痘苗病毒或腺病毒等人病毒或动物病毒; 杆状病毒等昆虫病毒; 酵母载体; 噬菌体载体 (如 λ 噬菌体载



体)，以及质粒和粘粒 DNA 载体。

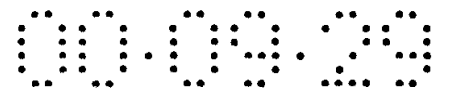
此外，还可选择能够调节插入基因的表达或能够按特定方式对基因产物进行修饰和加工的宿主细胞菌株。某些诱导因子的存在可提高特定启动子的表达；从而提高 4.2 节公开的嵌合蛋白的基因工程表达量。此外，不同的宿主细胞具有特异的蛋白翻译和翻译后加工以及修饰（如糖基化、磷酸化）特点和机制。可选择适当的细胞系或宿主系统以确保对所表达的外源蛋白进行所需的修饰和加工。例如，用细菌系统表达可生产非糖基化的核心蛋白产物。用哺乳动物细胞表达可确保异种蛋白的“天然”糖基化。此外，不同的载体/宿主表达系统可不同程度地影响加工反应。

cDNA 和基因组序列均能够被克隆并表达。

4.2 节公开的嵌合蛋白或其片段、类似物或衍生物也可利用标准方法加以分离和纯化，这些方法包括层析（如离子交换层析、亲和层析和分子筛）、离心、差示溶解度法或其它用于蛋白存活的标准技术。功能性质的评定可采用任何适当的测定方法。

编码 4.2 节公开的嵌合蛋白或其片段、类似物或衍生物的核酸序列可进行体外或体内突变，以产生并/或破坏翻译序列、起始序列、和/或终止序列，或在编码区产生变异。任何该领域已知的诱变技术均可使用，这些技术包括，但不局限于，体外定向诱变（Hutchinson et al., 1978, *J. Biol. Chem.* 253:6551）、使用 TAB 接头（Pharmacia）、含有突变的 PCR 引物等。

诱变的实验方法主要包括定向突变以及后面的变化基因产物的表型检测。根据需要，一些更常用的定向突变规程可使用提供单链 DNA 的载体，也可使用提供双链 DNA 的载体。通常，使用这些载体的诱变规程如下。合成一种诱变引物，即一种与要改变的序列互补，但包含一个或少量已改变、增加或缺失的碱基。利用 DNA 聚合酶在体外将引物延伸，并经过其它一些处理后，将双链 DNA 转染到细菌细胞中。接着利用多种方法鉴定具有所需突变的 DNA，并由含有突变序列的克隆纯化出所需蛋白。对于更长的序列则通常需要一些额外的克隆步骤，因为长插入片段（长于 2000 个碱基）在这些载体中不稳定。其规程为该领域的技术人员所了解，而用于定向诱变的试剂盒可从众多的生



物技术供应公司购得，如 Amersham Life Science, Inc. (Arlington Heights, IL) 和 Stratagene Cloning Systems (La Jolla, CA)。

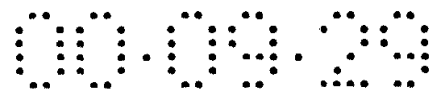
此外，可利用化学方法合成 4.2 节公开的嵌合蛋白或其片段、类似物或衍生物（见 Clark-Lewis et al., 1991, Biochem. 5 30:3128-3135; Merrifield, 1963, J. Amer. Chem. Soc. 85:2149-2156）。例如，可采用固相技术合成 4.2 节公开的嵌合蛋白或其片段、类似物或衍生物，再从树脂上裂解，并通过制备型高效液相色谱纯化（例如见 Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., N.Y., pp. 50-60）。
10 4.2 节公开的嵌合蛋白或其片段、衍生物和类似物也可利用肽合成仪来合成，合成肽的组成可利用氨基酸分析或测序方法来进行验证（如 Edman 降解法；见 Creighton, 1983, Proteins, Structures and Molecular Principles, W.H. Freeman and Co., N.Y., pp. 34-49）。

4.2 节公开的嵌合蛋白或其片段、衍生物和类似物可完全通过氨基酸残基的顺序添加来合成，作为选择，也可利用该领域所熟知的技术由亚组份片段连接而成，这些技术包括，如片段缩聚（Shin et al., 15 1992, Biosci. Biotech. Biochem. 56:404-408; Nyfeler et al., 1992, Peptide, Proc. 12th Amer. Pep. Soc., Smith and River (eds), Leiden, pp661-663; Nokihiro et al., 1990, Protein Rearsch Foundation, Yanaihara (ed), Osaka, pp315-320）。

在次优选实施方案中，4.2 节公开的嵌合蛋白或其片段、衍生物和类似物可通过该蛋白的蛋白酶解，然后由上述的标准方法纯化而获得。

25 4. 4. 获得正确折叠的胰岛素前体

本发明提供一种获得含有胰岛素前体的正确折叠第一嵌合蛋白的方法，该方法包括：在一种水性介质中将含有胰岛素前体的非正确折叠第二嵌合蛋白与至少一种离液辅助剂相接触，其中所述含有胰岛素前体的第二嵌合蛋白包含一种经一个或多个可切除氨基酸残基与胰岛素前体隔离的内分子伴侣（IMC）样肽段；其中所述 IMC 样肽段：
30 a) 含有约 20-200 个氨基酸残基； b) 不是胰岛素前体或其一部分；



并且 c) 可促进胰岛素前体的折叠，以至于将含有胰岛素前体的非正确折叠第二嵌合蛋白与离液辅助剂接触时，含有胰岛素前体的正确折叠第一嵌合蛋白的产量要高于将不含有所述 IMC 样肽段的非正确折叠胰岛素前体与同样的离液辅助剂接触时的正确折叠胰岛素前体的产量。

5

文中使用的术语“胰岛素前体”是指这样一种分子，该分子 1) 含有人胰岛素 A 链和 B 链，或其类似物、衍生物和片段，2) 含有 6 个半胱氨酸残基，3) 具有可切除的连接胰岛素 A 链和 B 链的连接部分，并且 4) 能够被抗人胰岛素抗体结合。人胰岛素前体的实例包括，但不局限于，Ladisich and Kohlmann, *Biotechnol. Prog.*, 1992, 8:469-478; Thim et al., *Proc. Natl. Acad. Sci. USA*, 1986, 83:6766-6770; U.S Patent No. 5,473,049; U.S Patent No. 5,457,066; and EP 0,347,781 B1 所公开的那些。

10

术语“正确折叠的”人胰岛素前体或含有胰岛素前体的嵌合蛋白是指一种分子，该分子中的人胰岛素前体具有与天然生物活性人胰岛素相同的构象和二硫键，即形成 a) A-6 与 A-11, b) A-7 与 B-7, c) A-20 与 B-19 的二硫键。术语“非正确折叠的”人胰岛素前体或含有胰岛素前体的嵌合蛋白是指一种分子，该分子中的人胰岛素前体不具有与天然生物活性人胰岛素相同的构象或二硫键，或二者都不具备。

15

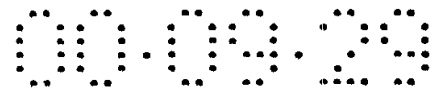
在一优选实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中的胰岛素前体来源于人。同样优选的是，人胰岛素前体能够被抗人胰岛素抗体结合。仍然优选的是，人胰岛素前体含有 SEQ ID NO:4 的氨基酸序列。亦为优选的还有，在人胰岛素前体中，人胰岛素前体的 B 链与 A 链由一个氨基酸残基或由 2-34 个氨基酸残基组成的肽段隔开。更优选的是，人胰岛素前体由 SEQ ID NO:5 的氨基酸序列组成。

20

25

在一项优选实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中 IMC 样肽段与胰岛素前体相比含有更高比例的带电氨基酸残基。同样优选的是，在 IMC 样肽段中，N 端部分包含的带正电的氨基酸残基多于带负电的氨基酸残基，而 C 端部分包含的带负电的氨基酸残基则多于带正电的氨基酸残基。仍然优

30



选的是，IMC 样肽段包含一段与含有人生长激素（hGH）蛋白 N 端至少前 20 个氨基酸的区域的同一性至少为 40% 的氨基酸序列，其中同一性百分率由大小与 hGH 的区域相同的一段氨基酸序列决定。亦为优选的还有，IMC 样肽段由一段与含有人生长激素（hGH）蛋白 N 端至少前 5 20 个氨基酸的区域的同一性至少为 60% 的氨基酸序列组成。同样优选的还有，IMC 样肽段能够被抗人生长激素抗体结合。更优选的为，IMC 样肽段含有 SEQ ID NO:1 的氨基酸序列。同样更优选的为，IMC 样肽段由 SEQ ID NO:2 的氨基酸序列组成。

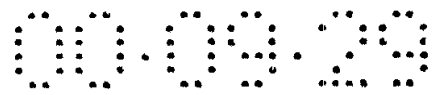
10 在一项优选实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中可切除的氨基酸残基为 Arg 或 Lys。同样优选的是，可切除氨基酸残基由 SEQ ID NO:3 的氨基酸序列组成。

15 在一项特定实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中在含有胰岛素前体的非正确折叠第二嵌合蛋白中，IMC 样肽段位于该嵌合蛋白的 N 端。在另一项类似的特定实施方案中，IMC 样肽段位于该嵌合蛋白的 C 端。在又一项类似的特定实施方案中，IMC 样肽段位于胰岛素前体的 B 链和 A 链之间。

20 在一项优选实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中 IMC 样肽段含有一个或多个可切除的氨基酸残基，这些残基与含有胰岛素前体的第二嵌合蛋白中隔离 IMC 样肽段和胰岛素前体的一个或多个可切除的氨基酸残基相同。更优选的为，可切除的氨基酸残基为一个 Arg 残基或一个 Lys 残基。

25 在一项最优选的实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白由 SEQ ID NO:6 的氨基酸序列组成。同样最优选的是，这种含有胰岛素前体的非正确折叠第二嵌合蛋白由 SEQ ID NO:7 的氨基酸序列组成。

30 离液辅助剂为能够在水溶液中打断氢键的复合物。在一项特定实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中离液辅助剂的选自盐酸胍、碳酸亚乙酯、硫氰酸、二甲亚砜和脲。优选的离液辅助剂为脲。更优选的是，脲的浓度约为



2.0 -8.0 M。最优选的是，脲的浓度约为 3.0-6.0 M。

5 在另一项特定实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，其中含有胰岛素前体的非正确折叠第二嵌合蛋白在 pH 约为 8.0-10.5 的水性介质中与至少一种离液辅助剂相接触，而含有胰岛素前体的非正确折叠第二嵌合蛋白的浓度约为 0.05-15g/L。优选的 pH 值约维持在 9.0-10.0。同样优选的，含有胰岛素前体的非正确折叠第二嵌合蛋白约为 0.5-5.0g/L。更优选的是，含有胰岛素前体的非正确折叠第二嵌合蛋白约为 2.0-3.0g/L。

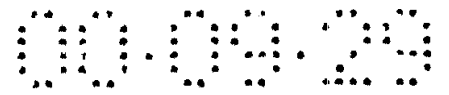
10 硫醇为至少含有一个-SH 基团的可溶于水的复合物。在另一项特定实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，该方法进一步包括将含有胰岛素前体的非正确折叠第二嵌合蛋白与一定量的硫醇相接触，其中硫醇的量可为含有胰岛素前体的非正确折叠第二嵌合蛋白的每个半胱氨酸残基提供少于 5 个的硫醇基。在另一项特定实施方案中，含有胰岛素前体的非正确折叠第二嵌合蛋白同时与硫醇和离液辅助剂相接触。在另一项特定实施方案中，含有胰岛素前体的非正确折叠第二嵌合蛋白依次与硫醇和离液辅助剂相接触。优选的硫醇用量可为含有胰岛素前体的非正确折叠第二嵌合蛋白的每个半胱氨酸残基提供约 0.07-1.0 个硫醇基。同样优选的是，硫醇选自二硫苏糖醇、二硫赤藓糖醇、2-巯基乙醇、半胱氨酸、巯基乙醇酸甲酯、3-巯基-1,2-丙二醇和 3-巯基丙酸。更优选的硫醇为 2-巯基乙醇。

25 在另一特定实施方案中，本发明提供一种获得上述含有胰岛素前体的正确折叠嵌合蛋白的方法，该方法进一步包含将含有胰岛素前体的正确折叠第一嵌合蛋白从含有胰岛素前体的非正确折叠第二嵌合蛋白中分离。优选的方法是通过超滤将含有胰岛素前体的第一嵌合蛋白从含有胰岛素前体的第二嵌合蛋白中分离。更优选的实施超滤的 pH 值约 8.0-11.0。最优选的实施超滤的 pH 值约为 9.0-10.0。

在另一项特定实施方案中，本发明通过上述的方法提供一种含有胰岛素前体的正确折叠嵌合蛋白。

30

4. 5. 筛选一种可促进胰岛素前体折叠的氨基酸序列



本发明提供一种对氨基酸序列促进胰岛素前体折叠的能力进行筛选的测定方法，该方法包括：(a) 改变 4.2 节公开的含有胰岛素前体的嵌合蛋白的第一肽段的氨基酸序列，获得具有所述改变的所述嵌合蛋白，在一定条件下于水性介质中将具有所述改变的所述嵌合蛋白与至少一种离液辅助剂接触足够长的时间，以使所述嵌合蛋白正确折叠，并测量具有所述改变的所述嵌合蛋白的折叠产量；(b) 获得用于步骤(a)的但没有任何步骤(a)所述氨基酸序列改变的嵌合蛋白，在与步骤(a)相同的条件下于水性介质中将没有任何步骤(a)所述氨基酸序列改变的嵌合蛋白与相同于步骤(a)所用的离液辅助剂接触与步骤(a)相同的时间，并测量嵌合蛋白的折叠产量；以及(c) 将分别在步骤(a)和步骤(b)中测量的嵌合蛋白折叠产量进行比较，其中若步骤(a)测得的产量基本上等于或大于步骤(b)测得的产量，则表明该氨基酸序列可促进胰岛素前体的折叠。

4.2 节公开的嵌合蛋白中第一肽段的氨基酸序列可通过任何该领域已知的诱变技术来加以改变。优选的方法是利用 4.3 节所述的诱变技术。

在上述测定法的一项优选实施方案中，嵌合蛋白由 SEQ ID NO:6 的氨基酸序列组成。

在上述测定法的另一项优选实施方案中，嵌合蛋白由 SEQ ID NO:7 的氨基酸序列组成。

在上述测定法的又一项优选实施方案中，离液辅助剂为脲。

在上述测定法的另一项优选实施方案中，该测定法进一步包括分别将步骤 a) 和步骤 b) 的嵌合蛋白与一定量的硫醇相接触，其中硫醇的量可为该嵌合蛋白的每个半胱氨酸残基提供少于 5 个的硫醇巯基。更优选的硫醇为 2-巯基乙醇。

在一项特定实施方案中提供上述测定法的产物。

5. 实施例

利用化学方法合成一种 DNA 片段，该 DNA 片段编码由 SEQ ID NO:6 的氨基酸序列组成的 hGH-小鼠胰岛素原。将该 DNA 片段克隆到 Trp 启动子调控的细菌表达载体中。将含有 hGH-小鼠胰岛素原的表达载体转化到

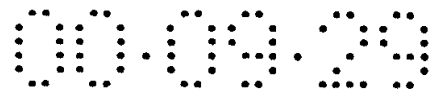
大肠杆菌 PR1 菌株内，并将重组细胞培养于含有微量元素的 M9-CA 培养基中。从包涵体内回收 hGH-小胰岛素原融合蛋白，并在一定条件下折叠，使折叠的 hGH-小胰岛素原融合蛋白中形成的二硫键与正确折叠的人胰岛素原中的二硫键相同，即形成的二硫键为 A6-A11、A7-B7 和 A20-B19。通过 100K 超滤将折叠的 hGH-小胰岛素原融合蛋白从未折叠的融合蛋白中分离出来。用胰蛋白酶消化分离的正确折叠 hGH-小胰岛素原融合蛋白，以形成正确折叠的 Arg(B31)-人胰岛素。通过阳离子交换层析将 Arg(B31)-人胰岛素纯化为至少 90% 的纯度。然后用羧肽酶 B 消化已纯化的 Arg(B31)-人胰岛素，以形成正确折叠的人胰岛素，然后再通过反相 HPLC 纯化。利用 N 端序列分析、分子量测定以及肽图法确定如此产生的纯人胰岛素的性质。

5. 1. hGH-小胰岛素原表达载体的构建

根据 Gan et al., Gene, 1989, 79:159-166 中公开的方法化学合成编码含有 SEQ ID NO:6 的氨基酸序列的 hGH-小胰岛素原的 DNA 片段。合成的 DNA 片段中包含一个 5'端 Cla I 位点和一个 3'端 Hind III 位点。简单地说，即利用化学方法合成一段由 5'端 Cla I 到 3'端 Kpn I 的片段和一段由 5'端 Kpn I 到 3'端 Hind III 的片段并分别克隆到 pUC18 载体中，其中 Kpn I 可切割编码 SEQ ID NO:6 的 51 位和 52 为氨基酸残基的核酸序列。然后将编码 SEQ ID NO:6 的完整氨基酸序列的 DNA 片段亚克隆到修饰的 pATH2 载体中，以使 hGH-小胰岛素原的表达受到 Trp 启动子和 SD 序列的调控。所得载体 pZRhi-1 (图 2) 则用于 hGH-小胰岛素原融合蛋白的表达。

5. 2. hGH-小胰岛素原融合蛋白的表达

将 hGH-小胰岛素原融合蛋白的表达载体转化到大肠杆菌菌株 RR1 或 K12 W3110 中。将转化的大肠杆菌细胞培养于含有微量元素的 M9-CA 培养基中。对 hGH-小胰岛素原融合蛋白在摇瓶和发酵罐中的表达水平均进行检测。在这两种情况下均可观察到相当于大肠杆菌总蛋白的 20-25% 的高水平表达。



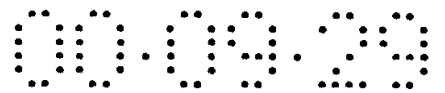
5. 3. hGH-小鼠胰岛素原融合蛋白的重折叠

hGH-小鼠胰岛素原融合蛋白是以不可溶的“包涵体”形式表达。为释放包含体，在 800 巴的高压匀浆器中将大肠杆菌细胞破碎。通过 10000g 离心去除细胞碎片和可溶性大肠杆菌蛋白。用水将含有 hGH-小鼠胰岛素原融合蛋白的包涵体清洗三次。所得包涵体沉淀用作折叠的原材料，其中的 hGH-小鼠胰岛素原融合蛋白约为 90% 的纯度。将包涵体溶于 pH 10.4 并含有 2-6 mM 巯基乙醇的 8M 脲中，使 hGH-小鼠胰岛素原融合蛋白的浓度为 20-30 mg/ml。离心去除不溶性物质。用低浓度的脲将上清稀释 10 倍，使脲的终浓度为 3-6M，pH 为 9-10，巯基乙醇的终浓度则为 0.2-0.6 mM。在 4℃ 和 3.2 M 脲浓度、pH 9.3 的条件下进行常规折叠。利用溶于 0.1% 磷酸缓冲液的 30-47% 乙腈梯度经 C4 反相 HPLC 监控该折叠过程。在这样的条件下，正确折叠的 hGH-小鼠胰岛素原的滞留时间约为 23 分钟。折叠可在 24 小时内完成。再折叠率约为 70%。

15 通过 100K 超滤将折叠混合物分部分离。利用 10K 超滤系统将滤出液组份中正确折叠的 hGH-小鼠胰岛素原融合蛋白浓缩。在 pH 3.5 下用水去除脲。超滤步骤的产率超过 85%。

5. 4. 正确折叠的 hGH-小鼠胰岛素原的胰蛋白酶裂解

20 胰蛋白酶裂解时，正确折叠的 hGH-小鼠胰岛素原融合蛋白的浓度约为 10-12 mg/ml，优选的为 10 mg/ml。胰蛋白酶与 hGH-小鼠胰岛素原融合蛋白的比例范围约为 1:60-1:250，优选的为 1:100。pH 值约维持在 10-11，优选的为 10.8。于 4℃ 下将裂解进行约 1-5 小时，优选的为 3.5 小时。用磷酸缓冲液将 pH 值调为 3.5 以终止裂解反应。反相 HPLC 分析结果表明，该裂解步骤的产率高于 95%。在高于 10 的 pH 下，胰蛋白酶可作用于以下 Arg 残基：人小鼠胰岛素原 B 链和 A 链之间的 Arg 残基、hGH 片段与小鼠胰岛素原片段之间的 Arg 残基、hGH 内的 Arg 残基。hGH 片段经胰蛋白酶消化产生几个小片段和一个 B 链的 C 25 端带有一个额外 Arg 的人胰岛素，该胰岛素被命名为 Arg(B31)-人胰岛素。人胰岛素的 Arg(B22) 在上述条件下不能被切除，推测可能是由于三维结构的阻碍造成的。



用溶于 10mM 柠檬酸缓冲液的 NaCl 作为洗脱液，由阳离子交换层析将 Arg(B31)-人胰岛素纯化。纯化的 Arg(B31)-人胰岛素具有高于 90% 的纯度。

5 5. 5. Arg(B31)-人胰岛素转化为人胰岛素

用羧肽酶 B 去除 Arg(B31)-人胰岛素的 C 端 Arg。正确折叠的 Arg(B31)-人胰岛素的所用浓度约为 10 mg/ml。羧肽酶 B 与 Arg(B31)-人胰岛素的比例约维持在 1:1000。37℃ 下于 50mM Tris-HCl 缓冲液 pH8.0 中将裂解进行约 1 小时。用磷酸缓冲液将 pH 值调为 3.5 以终止反应。人胰岛素的产率高于 99%。

5. 6. 人胰岛素的纯化

15 将羧肽酶 B 消化产生的人胰岛素上样到已用 pH3.0 的 0.1% 磷酸缓冲液平衡的 C8 反相 HPLC 柱上。利用溶于 pH3.0 的 0.1% 磷酸缓冲液的 17-35% 的乙腈梯度洗脱人胰岛素。收集胰岛素组份，并加入浓度为 0.125-0.2 M 的乙酸，pH 值为 6.0。4℃ 下将如此产生的胰岛素结晶。人胰岛素的纯度高于 99%。

5. 7. 纯化人胰岛素的特性

20 利用标准 Edman 降解法测定纯化人胰岛素和 WHO 标准人胰岛素的 N 端前 15 个氨基酸。纯化人胰岛素 A 链和 B 链的 N 端序列均与 WHO 标准人胰岛素的相同。

 利用 VG Platform 质谱分析法确定纯化人胰岛素与 WHO 标准人胰岛素的分子量。两者的 M/Z 均为 5807.7。

25 用 V8 蛋白酶消化纯化人胰岛素和 WHO 标准人胰岛素。用 C18 反相 HPLC 分析消化片段。两者的肽图模式完全相同（见 Thim et al., Genetics and Molecular Biology of Industrial Microorganisms, Hershberger et al., Ed. American Society for Microbiology, 1989, p322-328）。

30 本发明并不局限于文中叙述的微生物或文中所述的特定实施方案的范围。事实上，由于前文的描述及附图，本发明除文中所述之外

的各种变化对该领域的专家而言均是显而易见的。这些变化将仍处于附加权利要求的范围之内。

文中引用了许多参考文献，其公开内容将被全面引入作为参考。

6. 序列表

(1) 一般信息

(i) 申请人: Ganz, Z. R.

(ii) 发明题目: 含分子内伴侣样序列的嵌和蛋白及其在胰岛

5 素生产上的应用

(iii) 序列数: 7

(iv) 通信地址:

(A) 收件人

(B) 街

10 (C) 市

(D) 省/州

(E) 国家

(F) 邮编

(v) 计算机可读形式

15 (A) 介质类型: 3.5 寸盘

(B) 计算机: IBM PC

(C) 操作系统: DOS

(D) 软件: WordPerfect5.1

(vi) 当前申请数据

20 (A) 申请号: 待发

(B) 提交日期: 同时提交

(C) 分类

(vii) 优先申请数据

(A) 申请号

25 (B) 申请日期

(viii) 律师/代理人信息

(A) 名字

(B) 登记号

(C) 参考/文档号

30 (ix) 电信信息

(A) 电话

(B) 电传

(C) 电报

(2) SEQ ID NO:1 的信息

(i) 序列特征

5 (A) 长度:49 个氨基酸

(B) 类型:氨基酸

(C) 拓扑结构:线性

(ii) 分子类型:蛋白

(xi) 序列描述: SEQ ID NO:1

10

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 Pro
 50

(3) SEQ ID NO:2 的信息

(i) 序列特征

(A) 长度:92 个氨基酸

15 (B) 类型:氨基酸

(C) 拓扑结构:线性

(xi) 序列描述: SEQ ID NO:2

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 Pro Gln Thr Ser Leu Ser Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60
 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75
 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln
 85 90 95

(4) SEQ ID NO:3 的信息

(i) 序列特征

- 5 (A) 长度:6 个氨基酸
 (B) 类型:氨基酸
 (C) 拓扑结构:线性

(xi) 序列描述: SEQ ID NO:3

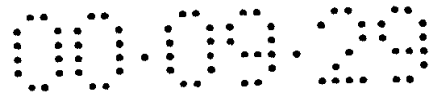
Leu Gly Thr Gly Pro Arg

1 5

(5) SEQ ID NO:4 的信息

- 10 (i) 序列特征
 (A) 长度:86 个氨基酸
 (B) 类型:氨基酸
 (C) 拓扑结构:线性

(xi) 序列描述: SEQ ID NO:4



Phe Val Asn Gln His leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr 15
1 5 10
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Arg 30
 20 25
Glu Ala Glu Asp Leu Gln Val Gly Gln Val Glu Leu Gly Gly Gly Pro 45
 35 40 45
Gly Ala Gly Ser Leu Gln Pro Leu Ala Leu Glu Gly Ser Leu Gln Lys 60
 50 55 60
Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln 80
65 70 75
Leu Glu Asn Tyr Cys Asn 90
 85 90

(6) SEQ ID NO:5 的信息

(i) 序列特征

- 5 (A) 长度:52 个氨基酸
 (B) 类型:氨基酸
 (C) 拓扑结构:线性

(xi) 序列描述: SEQ ID NO:5

Phe Val Asn Gln His leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr 15
1 5 10
Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr Arg Gly 30
 20 25 30
Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln Leu Glu 45
 35 40 45
Asn Tyr Cys Asn 55
 50 55

10

(7) SEQ ID NO:6 的信息

(i) 序列特征

- 15 (A) 长度:107 个氨基酸
 (B) 类型:氨基酸
 (C) 拓扑结构:线性

(xi) 序列描述: SEQ ID NO:6

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 Pro Leu Gly Thr Gly Pro Arg Phe Val Asn Gln His leu Cys Gly Ser
 50 55 60
 His Leu Val Glu Ala Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe
 65 70 75 80
 Tyr Thr Pro Lys Thr Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile
 85 90 95
 Cys Ser Leu Tyr Gln Leu Glu Asn Tyr Cys Asn
 100 105 110

(8) SEQ ID NO:7 的信息

- 5 (i) 序列特征
- (A) 长度:150 个氨基酸
 - (B) 类型:氨基酸
 - (C) 拓扑结构:线性
- (xi) 序列描述: SEQ ID NO:7

Met Phe Pro Thr Ile Pro Leu Ser Arg Leu Phe Asp Asn Ala Met Leu
 1 5 10 15
 Arg Ala His Arg Leu His Gln Leu Ala Phe Asp Thr Tyr Gln Glu Phe
 20 25 30
 Glu Glu Ala Tyr Ile Pro Lys Glu Gln Lys Tyr Ser Phe Leu Gln Asn
 35 40 45
 Pro Gln Thr Ser Leu Ser Phe Ser Glu Ser Ile Pro Thr Pro Ser Asn
 50 55 60
 Arg Glu Glu Thr Gln Gln Lys Ser Asn Leu Glu Leu Leu Arg Ile Ser
 65 70 75 80
 Leu Leu Leu Ile Gln Ser Trp Leu Glu Pro Val Gln Leu Gly Thr Gly
 85 90 95
 Pro Arg Phe Val Asn Gln His leu Cys Gly Ser His Leu Val Glu Ala
 100 105 110
 Leu Tyr Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr
 115 120 125
 Arg Gly Ile Val Glu Gln Cys Cys Thr Ser Ile Cys Ser Leu Tyr Gln
 130 135 140
 Leu Glu Asn Tyr Cys Asn
 145 150

说明书附图



图 1A

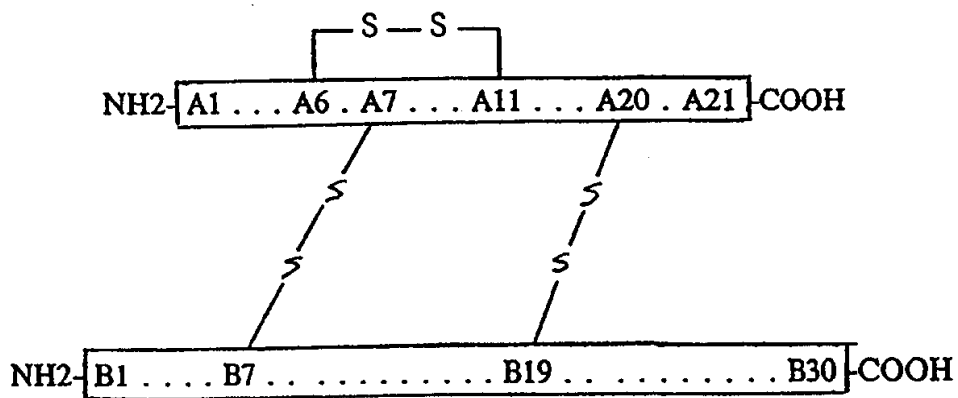


图 1B

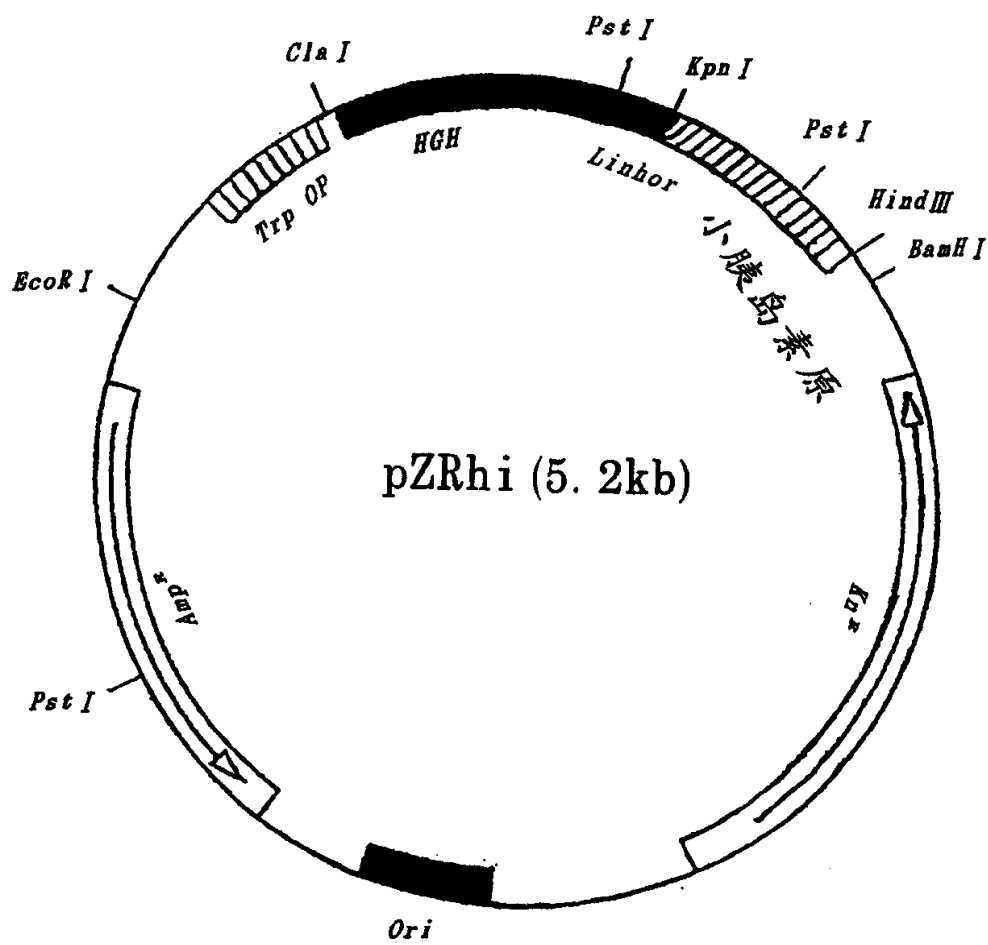


图 2