[19]中华人民共和国专利局

[51]Int.Cl[6]
C12N 5/04
C12N 15/29 C12N 15/31
C12N 15/82 C07K 14/415
A01H 4/00

[12] 发明专利申请公开说明书

[21] 申请号 95193569.0

[43]公开日 1997 年 6 月 4 日

[11] 公开号 CN 1151183A

[22]申请日 95.4.13

[30]优先权

[32]94.4.13 [33]US[31]08/227,360

[86]国际申请 PCT/US95/04570 95.4.13

[87]国际公布 WO95/28478 英 95.10.26

[85]进入国家阶段日期 96.12.12

[71]申请人 综合医院公司

地址 美国马萨诸塞州

共同申请人 加州大学评议会

[72]发明人 F·M·奥斯贝 B·J·施塔斯卡维兹
A·F·宾特 D·达尔贝克 片桐文章
B·N·孔凯尔 M·N·明德因诺斯
G·-L·余

[74]专利代理机构 中国专利代理(香港)有限公司

代理人 吴玉和 姜建成

权利要求书 3 页 说明书 90 页 附图页数 11 页

[54]发明名称 RPS2基因及其应用

[57]摘要

本发明涉及基本上纯的编码拟南芥 Rps2 多肽的 DNA，基本上纯的 Rps2 多肽，以及利用该 DNA 在转基因植物细胞和整株植物中表达 Rps2 多肽以使转基因植物获得对病原体的抗性。

权　利　要　求　书

1. 编码 Rps 多肽的基本上纯的 DNA。

2. 权利要求 1 中的 DNA，其中所述 DNA 包括 *RPS2* 基因［SEQ. ID.NO:1］。

3. 权利要求 1 中的 DNA，其中所述 DNA 是基因组 DNA。

4. 权利要求 1 中的 DNA，其中所述 DNA 是 cDNA。

5. 权利要求 1 中的 DNA，其中所述 DNA 是拟南芥属植物的 DNA。

6. 具有图 2 所示的序列［SEQ. ID.NO:1］的基本上纯的 DNA 或其简并变体，所述 DNA 编码图 2 所示的开放阅读框架“a”的氨基酸序列［SEQ. ID.NOS:2-5］。

7. 与图 2 所示的 DNA 序列［SEQ. ID.NO:1］有大约 50%或更高序列一致性的基本上纯的 DNA。

8. 权利要求 1 或 2 中的 DNA，其中所述 DNA 是与所述多肽的表达调控序列可操纵式连锁的；

其中的调控序列包括一个启动子。

9. 权利要求 8 中的 DNA，其中所述启动子是一个组成型启动子。

10. 权利要求 8 中的 DNA，其中所述启动子可被一种或多种外部试剂诱导。

11. 权利要求 8 中的 DNA，其中所述启动子是细胞型特异性的。

12. 含有权利要求 1 中的 DNA 的细胞。

13. 权利要求 12 中的细胞，所述细胞是植物细胞。

14. 权利要求 13 中的植物细胞，所述植物细胞是对由植物病原体引起的疾病有抗性的，所述植物病原体携带有一个能产生被 Rps 多肽识别的信号的无毒基因。

15. 权利要求 14 中的植物细胞，所述植物病原体携带 *avrRpt2* 基因。

16. 权利要求 14 中的植物细胞，所述植物细胞来自拟南芥、番茄、大豆、菜豆、玉米、小麦和水稻等植物。

17. 权利要求 14 中的植物细胞，所述植物病原体是丁香假单胞菌。

18. 权利要求 13 中的植物细胞，其中所述植物细胞还包含与调控序列可操纵式连锁的 *avrRpt2* 基因；其中所述调控序列包括一个启动子。

19. 权利要求 18 中的植物细胞，其中所述启动子是组成型启动子。

20. 权利要求 18 中的植物细胞，其中所述启动子可被一种或多种外部试剂诱导。

21. 权利要求 18 中的植物细胞，其中所述启动子是细胞型特异性的。

22. 含有整合到所述植物基因组中的权利要求 1 中的 DNA 的转基因植物，其中所述 DNA 在该转基因植物中表达。

23. 含有整合到所述植物基因组中的权利要求 8 中的 DNA 的转基因植物，其中所述 DNA 在该转基因植物中表达。

24. 由权利要求 18 中的植物细胞长成的转基因植物，其中所述 DNA 和 *avrRpt2* 基因在该转基因植物中表达。

25. 权利要求 22 中的转基因植物的种子。

26. 权利要求 23 中的转基因植物的种子。

27. 权利要求 24 中的转基因植物的种子。

28. 权利要求 22 中的转基因植物的细胞。

29. 权利要求 23 中的转基因植物的细胞。

30. 一种使植物获得对植物病原体的抗性的方法，其包括：制备含有权利要求 1 中的 DNA 的转基因植物细胞，该 DNA 整合到所述转基因植物细胞基因组中并被置于能在所述植物细胞中表达的位置；由该植物细胞生长出转基因植物，其中所述 DNA 在该转基因植物中表达。

31. 一种检测植物细胞中抗性基因的方法，其包括：在杂交条件下使权利要求 1 中的 DNA 或其长度大于大约 18 个核苷酸的部分与从所述植物细胞中得到的基因组 DNA 制备物接触，上述条件可以用来检测与图 2 所示的序列 [SEQ. ID.NO:1] 有大约 50 % 或更高序列一致性的 DNA 序列。

32. 一种制备 Rps2 多肽的方法，其包括：

制备用编码 Rps2 多肽，并被置于能在所述细胞中表达的位置的 DNA 转化的细胞；

将该转化细胞在适宜条件下培养以使所述 DNA 表达;

分离出所述 Rps2 多肽。

33. 一种使转基因植物获得对植物病原体的抗性的方法，其包括: 制备含有权利要求 8 中的 DNA 的转基因植物细胞，该 DNA 整合到该转基因植物细胞的基因组中并被置于能在所述植物细胞中表达的位置;

由该植物细胞生长出转基因植物，其中所述 DNA 在该转基因植物中表达。

34. 一种使转基因植物获得对植物病原体的抗性的方法，其包括: 由权利要求 18 中的植物细胞生长出转基因植物，其中所述 DNA 和 *avrRpt2* 基因在该转基因植物中表达。

35. 一种从植物中分离与 *RPS2* [SEQ. ID.NO:1]有序列一致性的抗病基因或其部分的方法，其包括:

用寡聚核苷酸引物通过 PCR 扩增该抗病基因或其部分，该引物

(a)每一个的长度均大于 13 个核苷酸;

(b)每一个都有与图 2 核苷酸序列 [SEQ. ID.NO:1] 中某个区域的相对的 DNA 链互补的区域;

(c)可任选地包含在扩增产物中产生限制性酶切位点的序列; 以及

分离出该抗病基因或其部分。

36. 一种基本上纯的 Rps2 多肽

37. 权利要求 32 中的多肽，其含有与图 2 所示的氨基酸序列 [SEQ. ID.NOS:2-5]基本一致的氨基酸序列。

38. 一种含有权利要求 1 中的 DNA 的载体，该载体能在含有该载体的细胞中指导所述 DNA 编码的多肽的表达。

39. 一种含有与调控序列可操纵式连锁的 *avrRpt2* [SEQ. ID.NO:105] 基因的 DNA 的载体，其中调控序列包括一个启动子。

40. 一种含有与调控序列可操纵式连锁的权利要求 1 中的 DNA 和 *avrRpt2* 基因 [SEQ. ID.NO:105] 的 DNA 的载体，其中调控序列包括一个启动子。

说　明　书

RPS2 基因及其应用

发明背景

本发明涉及重组植物核酸和多肽以及它们在使转基因植物具有对病原体的抗病性方面的应用。

植物有多种抵抗病原体的策略。其中一种防御反应是过敏反应(HR)，它包括迅速将受感染组织的坏死限定在局部。在某些宿主－病原体相互作用中，遗传学分析揭示了不致病病原体的特定无毒(*avr*)基因中基因对基因的对应性(gene-for-gene correspondence)，所述不致病病原体可以在具有特定抗性基因的宿主中引发过敏反应。

发明概述

简而言之，本发明涉及基本上纯的编码 Rps 多肽（下面将要描述）的 DNA（例如基因组 DNA、cDNA 或合成 DNA）。在相关方面，本发明也涉及含有上述基本上纯的编码 Rps 多肽的 DNA 的载体，细胞(如植物细胞）和转基因植物或其种子。

在优选实施方案中，*RPS* 基因［SEQ.ID NO:5］是指拟南芥属（Arabidopsis）植物的 *RPS2* 基因。在不同的优选实施方案中，细胞是指从转基因植物细胞得到的转化植物细胞。在相关方面，本发明涉及含有编码 RPS 多肽的转基因的转基因植物，所述 RPS 多肽在易被表达 *avrRpt2* 无毒基因［SEQ.ID NO:105］的病原体或表达能被 RPS 多肽以类似方式识别的无毒信号的病原体感染的植物组织中表达。

另一方面，本发明涉及基本上纯的含一个启动子的 DNA，该启动子能在易被表达 *avrRpt 2* 无毒基因［SEQ.ID NO:105］的细菌病原体感染的植物组织中表达 *RPS2* 基因［SEQ.ID NO:1］。

在优选的实施方案中，启动子是指 *RPS* 基因的天然启动子。此外，转录和翻译调节区域也优选地为 *RPS* 基因的天然调节区域。

本发明中的转基因植物优选地为易被表达无毒基因（优选地是 *avrRpt2* 无毒基因［SEQ.ID NO:105］）的病原体所感染的植物。在优选实施方案中，转基因植物选自但不限于拟南芥、番茄、大豆、菜豆、玉米、小麦和水稻。

在另一方面，本发明涉及一种使植物具有对病原体的抗性的方法，包括：(a)制备含有编码 Rps2 多肽的转基因的转基因植物细胞，其中转基因被整合到转基因植物基因组中并被放在表达位点；(b)由转基因植物细胞培养出表达 *RPS2* 转基因的转基因植物。

另一方面，本发明涉及一种检测植物细胞中抗病基因的方法，包括：(a)在杂交条件下使 *RPS2* 基因［SEQ.ID NO:1］或者该基因中长度大于 18 个核苷酸的片段与待测植物细胞的基因组 DNA 接触，该方法可用于检测与图 2 中编码 Rps2 多肽［SEQ.ID NOS:2-5］的 DNA 序列有约 50 % 以上一致性的 DNA 序列。

另一方面，本发明涉及一种生产 Rps2 多肽的方法，包括：(a)制备用编码 Rps2 多肽的 DNA 转化并将该 DNA 放在表达位点的细胞；(b)将转化细胞在适宜 DNA 表达条件下培养；和(c)分离出 Rps2 多肽。

另一方面，本发明涉及基本上纯的 Rps2 多肽，该多肽优选地具有一段含 50 个以上氨基酸的序列，该序列与图 2 中所示开放阅读框架“a”的一段含 50 个以上氨基酸的序列基本一致，该多肽最优选地是拟南芥（Arabidopsis thaliana）的 Rps2 多肽。

另一方面，本发明涉及一种使转基因植物获得对不携带 *avrRpt2* 无毒基因的病原体的抗性的方法，包括：(a)制备含有编码 Rps2 多肽的转基因和编码 *avrRpt2* 基因产物［SEQ.ID No:106］的转基因的转基因植物细胞，其中，转基因整合到转基因植物基因组中并被放在表达位点；*avrRpt2* 转基因（如果需要的话，也可为 *RPS2* 基因）受适于控制基因表达的调控序列的控制；(b)由转基因植物细胞培养出转基因植物，其中 *RPS2* 和 *avrRpt2* 转基因在转基因植物中得到表达。

另一方面，本发明涉及一种使转基因植物获得对不表达无毒基因的病原体的抗性的方法，包括：(a)制备带有整合到基因组中的转基因的转基因植物细胞，所述转基因中含有一个受控于特定启动子的 *RPS2* 基因，该启动子能使 *RPS2* 基因组成型表达，(b)由转基因植物细胞培养出组成型表达 *RPS2* 转基因的转基因植物。

另一方面，本发明涉及一种使转基因植物获得可调控的对不表达无毒基因的病原体的抗性的方法，包括：(a)制备带有整合到基因组中的转基因的转基因植物细胞，所述转基因中含有一个受控于特定启动子的 *RPS2* 基因，该启动子能使 *RPS2* 基因可调控地表达；(b)由转基因植物

细胞培养出可调控地表达 *RPS2* 转基因的转基因植物。在优选实施方案中， *RPS2* 基因的表达是利用一个组织特异性或细胞型特异性的启动子，或利用一个能被外源信号或试剂（如化学信号或化学试剂）所激活的启动子。

"抗病基因"是指编码一种能引发植物细胞或组织防御反应的多肽的基因。 RPS 基因是一种与图 2 中的 *RPS2* 序列[SEQ ID NO:1]或其一部分有大约 50 % 以上序列一致性的抗病基因。 *RPS2* 基因是拟南芥中编码 Rps2 抗病多肽 [SEQ .ID NOS:2-5] 的抗病基因。

"多肽"是指任意氨基酸链，不考虑其长度和翻译后修饰（例如糖基化或磷酸化）。

"基本一致"是指该多肽或核酸与参考多肽或核酸在序列上有50%，优选地为 85%，更优选地为 90%，最优选地为 95%的同源性。对于多肽，所比较的序列长度一般为至少 16 个氨基酸，优选地为至少 20 个氨基酸，更优选地为至少 25 个氨基酸，最优选地为至少 35 个氨基酸。对于核酸，所比较的序列长度一般至少为 50 个核苷酸，优选地为至少 60 个核苷酸，更优选地为至少 75 个核苷酸，最优选地为至少 110 个核苷酸。

序列一致性分析一般采用序列分析软件（如 Wisconsin 大学生物技术中心的遗传学计算机小组的序列分析软件包，大学路 1710 号，Madison, WI53705 ）。该软件能对不同的置换、缺失、置换以及其他修饰赋予同源度值来进行相似序列的比较。典型的保守性置换包括以下置换：甘氨酸，丙氨酸；缬氨酸，异亮氨酸，亮氨酸；天冬氨酸，谷氨酸，天冬酰胺，谷氨酰胺；丝氨酸，苏氨酸；赖氨酸，精氨酸；苯丙氨酸，酪氨酸。

"基本上纯的多肽"是指已经分离除去了其天然共存成分的 Rps2 多肽。一般地，当一种多肽在除去了与它天然共存的蛋白质和有机分子后其重量占总重量的 60 % 就可称之为基本上纯的。优选的 Rps2 多肽占总重量的 75 % 以上，更优选的为 90 % 以上，最优选的为 99 % 以上。基本上纯的 Rps2 多肽可以从天然来源（如植物细胞）中提取；可以表达编码 Rps2 多肽的重组核酸；也可以化学合成该蛋白。纯度的测量可以用任何适当的方法，如柱层析、聚丙烯酰胺凝胶电泳、高效液相色谱分析等。

当一种蛋白在天然状态下从与它共存的杂质中分离出来后，该蛋白就是与其天然共存成分高度分离的了。因此，如果一种蛋白质是由化学合成的或是在与天然产生该蛋白的细胞所不同的胞内系统中生产的，那它就是与其天然共存成分高度分离的。因此，基本上纯的多肽包括那些来源于真核生物却在大肠杆菌或其他原核生物中合成的多肽。

“基本上纯的 DNA”是指这样的 DNA，它不含有在该 DNA 来源生物的基因组中天然存在于该基因两侧的基因。因此，它包括例如插入到载体中的重组 DNA；插入到自复制质粒或病毒中的重组 DNA；插入到原核或真核基因组中的重组 DNA；或不含其他片段而以独立分子形式存在的 DNA（如 cDNA，用 PCR 或限制性内切酶消化所产生的基因组片段或 cDNA 的片段）。它还包括杂种基因中编码附加多肽序列的那一部分重组 DNA。

“转化细胞”是指这样的细胞，在它本身（或它的祖先细胞）中利用重组 DNA 技术导入了编码 Rps2 多肽（如其中所用的）的 DNA 分子。

“置于能够表达的位置”指该 DNA 分子被放在一段指导该序列的转录和翻译（即促进例如 Rps2 多肽或重组蛋白或 RNA 分子的产生）的 DNA 序列的相邻位置。

“报导基因”是指其表达能被检测的基因；这样的基因包括但不限于β－葡糖苷酸酶（GUS）、荧光素酶、氯霉素转乙酰酶（CAT），以及β－半乳糖苷酶。

“启动子”是指足以指导转录的极小的序列。在本发明中，启动子也包括一些足以引起启动子依赖性的能被细胞型特异性和组织特异性外部信号或试剂所调控的或能被外部信号或试剂所诱导的基因表达的启动子元件，这种元件可能位于天然基因的 5'或 3'端。

“可操纵式连锁”是指一个基因和一段调节序列以这样的方式连接在一起，当恰当的分子（如转录激活蛋白）结合到调节序列上时该基因能表达。

“植物细胞”是指能自我繁殖的含有一个质体并被半透膜包裹的细胞。如仍需要进一步繁殖，还要有细胞壁。其中所用的植物细胞包括藻类、氰细菌、悬浮培养的种子、胚、分生组织、愈伤组织、叶、根、芽、配子体、孢子体、花粉和小孢子等。

“转基因”是指被人工插入细胞内并成为由该细胞长成的生物体的

基因组的一部分的 DNA 片段，还包括与转基因生物部分或全部异源（即不同）的基因，或与某生物的内源基因同源的基因。

“转基因的”是指含有人工插入细胞内并成为由该细胞长成的生物体的基因组的一部分的 DNA 片段的细胞。其中所用的转基因生物通常指转基因植物，DNA（转基因）已被人为地插入到该植物的核基因组或质体基因组中。

“病原体”是指侵入活的植物组织细胞能诱导植物组织发生疾病反应的生物。

本发明的其他特征和优点将在下面的优选实施方案和权利要求中阐述。

发明详述

首先说明图表

附图

图 1A - 1F 是为确定 *RPS2* 基因座(locus)的克隆位点而作的物理和 RFLP 分析图谱。

图 1A 是拟南芥染色体 IV 的相关片段的遗传学和 RFLP 图谱的排列。由 1993 年的植物学杂志（4:745-750）上 Lister 和 Dean 所发表的图谱修改而成。RFLP 的标记 L11F11 存在于 YUP11F11 YAC 克隆的左臂。

图 1B 是 *RPS2* 基因座周围的相关 YACs 的排列图。YAC 构建体 YUP16G5，YUP18G9 和 YUP11F11 由 Pennsylvania 大学的 J. Ecker 提供。YAC 构建体 EW3H7，EW11D4，EW11E4 和 EW9C3 由 Ciba-Geigy 公司的 E. Ward 提供。

图 1C 是 *RPS2* 位点周围的科斯质粒（cosmid）克隆的排列图谱。科斯质粒克隆 H 是从 EW3H7 YAC 克隆中得到的，科斯质粒克隆 E 是从 EW11E4 YAC 克隆得到的，垂直箭头表示 RFLP 标记在生态型 La-er 和 *RPS2-101N* 植物之间的相对位置。RFLP 标记是用 Southern 印迹法筛选鉴定的，用了 50 种以上的限制性酶消化，并以相应的科斯质粒克隆的全部或部分作为探针。图 1C 中的科斯质粒克隆由法国的 J. Giraudat, C. N.R. S., Gif-Sur-Yvette 提供。

图 1D 和 1E 分别是科斯质粒 E4 - 4 和 E4 - 6 的 EcoRI 限制性内切酶的酶切位点图。*RPS2* 基因座周围的重组切断点存在于 EcoRI 限制

性酶切所得的 4.5 和 7.5kb 片段中。

图 1F 是编码 RNA 转录本的基因的大致位置图，该 RNA 转录本是用 PolyA$^+$ RNA 印迹分析法鉴定的。转录本的大小在每个转录本下面以千碱基对为单位标出。

图 2 是含有 *RPS2*[SEQ. ID. NO:1]基因座的 cDNA - 4 的全部核苷酸序列图。三个阅读框架在核苷酸下面标出。阅读框架“ a ”的推测的氨基酸序列 [SEQ.ID. NOS:2-5] 也已标出，共有 909 个氨基酸残基。图 2 中的开放阅读框架“ a ”中由起始密码子 ATG 编码的甲硫氨酸已被圈出。图 2 中起始密码子 ATG 中的 A 是第 31 个核苷酸。

图 3 是 *avrRpt2* 基因 [SEQ. ID NO:105] 核苷酸序列及其推测的氨基酸序列 [SEQ.ID. NO:106] 。下面加了横线的是一个可能的核糖体结合位点。开放阅读框架 3'末端用水平箭头标出的是反向重复序列。开放阅读框架中推测的氨基酸序列在相应核苷酸序列下标出。

图 4 是为了确定 P4104 和 P4115 所携带的 DNA（编码 cDNA - 4 ）的功能而进行的互补作用分析的简图，该 DNA 可使原本没有 *RPS2* 抗病活性的拟南芥具有 *RPS2* 抗病活性。“ Genome”行上的垂直短线表示限制性酶 EcoRI 的识别位点，其上的数字表示酶切所得的 DNA 片段（亦见图 1E ）的大小，单位是千碱基对(kb)。相对的“ cDNAs ”是编码 RNA 转录本（亦见图 1F ）的序列的大概位置；箭头表示 cDNAs 4， 5 和 6 的转录方向。在功能互补实验中，用前面提到的拟南芥基因组 DNA 序列转化 *rps2-201C* / *rps2-201C* 植物，这些序列被携带在已标出名称的质粒（双科斯质粒载体 pSLJ4541 的衍生物）上，并用土壤杆菌（ Agrobacterium)介导转化法转至植物中。所得的转化子与表达 *avrRpt2* 的丁香假单胞菌（ P.syringae)共同培养后的抗病性表现型用“ Sus.”（感病，无抗病反应）或“ Res.”（抗病）表示。

对病原体抗性的遗传学基础

其中概要介绍植物宿主与微生物病原体之间的相互作用。植物受可能的病原体侵袭后将产生以下各种不同的结果：或者病原体在宿主中成功地扩散，引起相关病症；或者病原体的生长被宿主的防御反应所停止。在某些植物－病原体的相互作用中，活性防御反应的可见标志被称为过敏反应或“ HR ”。过敏反应包括受感染部位附近细胞迅速坏死，可能还包括可见的干枯斑的产生。能诱导特定宿主发生过敏反应的病原

体是对该宿主不致病的，宿主是抗病的，这种植物－病原体间的相互作用是不相容的。能在特定宿主上增殖并引发疾病的菌株是致病的；这种情况下的宿主是感病的，植物－病原体间的相互作用是相容的。

"经典"遗传学分析已被成功地用于阐明一系列特定宿主种类的栽培品种（或不同的野生品种）中致病或不致病的真菌或细菌病原体的一系列菌株（小种）与特定的宿主植物之间相互识别的遗传学基础。在许多这种情况下，对宿主和病原体的遗传学分析表明许多不致病的真菌和细菌品系与致病的品系之间的差别在于前者有一个或多个无毒(avr)基因，该基因能与宿主中的"抗病"基因相互对应。这种无毒基因－抗病基因的对应关系被称为"基因对基因"（"gene-for-gene"）模型（Crute等，(1985) pp 197-309，植物抗病机制，R. S. S. Fraser 编；Ellingboe，（1981）植物病理学年鉴 19:125-143；Flor，（1971）植物病理学年鉴 9:275-296；Keen 和 Staskawicz，（1988）同上；Keen 等，生物技术在控制植物病原体中的应用. I. Chet，编，John Wiley & Sons，1993，pp.65 - 88)根据该模型的一个简单的描述，植物抗病基因编码无毒基因所产生的信号分子的特异性受体。然后，信号传导通路将该信号传送到一系列引起宿主过敏反应和其他防御反应的靶基因上（Gabriel 和 Rolfe，(1990)植物病理学年鉴 28:365-391）。这只是一个简单的推测模型，无毒－抗病基因相互作用的分子基础目前仍不清楚。

通过许多细菌无毒基因的克隆，"基因对基因假说"中一个基本的预言已在分子水平上被令人信服地证实，（Innes 等，（1993）细菌学杂志 175:4859-4869；Dong 等，（1991）植物细胞 3:61-72; Whelan 等，(1991) 植物细胞 3:49-59; Staskawicz 等，（1987）细菌学杂志 69:5789-5794; Gabriel 等，（1986）美国国家科学院院刊 83:6415-6419; Keen 和 Staskawicz，（1988）微生物学年鉴 42:421-440; Kobayashi 等，（1990）分子植物－微生物相互作用 3:94-102 和（1990）3:103-111)。这些被克隆的无毒基因中的许多与相关宿主植物中的单个抗病基因是相互对应的，而且当它们被转移到其它致病菌株中时引起不致病的表现型。Innes 等人从感染番茄的假丁香单胞菌（*Pseudomonas Syringae*）中分离出了 *avrRpt2* 基因座（Innes 等，（1993）细菌学杂志 173:4859-4869）。图 3 是 *avrRpt2* 基因的核苷酸序列[SEQ. ID NO:105]和推测的氨基酸序列[SEQ. ID NO:6]。

当被病原体侵袭时，能引发植物作出相应反应的已知信号有：欧文氏杆菌属(Erwinia) (Wei 等，(1992) 科学 257:85-88) 和假单胞菌属(Psendomonas)(He 等，(1993)细胞 73:1255-1266)中的 harpins; 枝孢属(Cladosporium)中的 *avr4* (Joosten 等，(1994)自然 367:384-386)和 avr9 肽(van den Ackerveken 等，(1992)植物杂志 2:359-366); 假单胞菌属 (Arlat 等，(1994)EMBO J. 13:543-553)中的 *PopA1*； avrD 产生的脂多糖(Midland 等，(1993)有机化学杂志(J. Org. Chem) 58:2940-2945); 嘴孢属(Rhynchosporium)中的 *NIP1*(Hahn 等，(1993) 分子植物－微生物相互作用 6:745-754).

与无毒基因相比，人们对与无毒基因产生的特异性信号相对应的植物抗性基因的了解要少得多。植物抗性基因 *RPS2*(rps 代表抗丁香假单胞菌 resistance to Pseudomonas syringae)，是从前未发现的植物抗病基因族中的第一个新基因，它与一个特定的无毒基因 (*avrRpt2*)相对应，有关 *RPS2* 克隆的工作见 1993 年分子植物－微生物相互作用杂志 6:434-443 上 Yu 等人的文章，以及 1993 年植物细胞杂志 5:865-875 上 Kunkel 等人的文章。

一个明显与此无关的与特定植物抗病基因 Pto 相对应的无毒基因已被从番茄(*Lycopersicon esculentum*)中分离出来(Martin 等，(1993)科学 262:1432-1436)。表达该 *Pto* 基因的番茄能抵抗表达 *avrPto* 无毒基因的番茄丁香假单胞菌(*Psendomonas syringae pv. tomato*)的感染。从 *Pto* 基因 DNA 序列推测出的氨基酸序列与丝氨酸－苏氨酸蛋白激酶高度同源，表明 *Pto* 与信号传导有关。但 *RPS2* 与 *Pto* 或其它任何已知的蛋白激酶都没有相似性，这表明 *RPS2* 是一个新型植物抗病基因族的代表。

已有人报导从玉米(*Zea mays*)中分离出了一种叫 *Hm1* 的种特异性抗性基因 (Johal 和 Briggs, (1992)科学 258:985-987) *Hm1* 通过有控制地降解一种真菌毒素来以引发对特定种类的真菌病原体旋孢霉(*Cochliobolus carbonum*)的抗性，这一策略的机制与无毒基因特异性的 *RPS2-avrRpt2* 抗病机制不同。

应用本发明中所克隆的 *RPS2* 基因可更容易地构建对特定病原体具有抗性的植物，并能克服经典的育种方法不能在种间转移抗病基因的缺陷(Keen 等，(1993)同上) 。下面将要描述的是拟南芥 *RPS* 基因座，*RPS2* 基因组 DNA 以及 *RPS2* cDNA 的克隆和表征。对于 *avrRpt2* 基因，

*RPS2*基因以及突变体*rps2-101C*, *rps2-102C*和*rps2-201C*(也称*rps2-201*)的描述见（ Dong 等人(1991)植物细胞 3:61-72; Yu 等人(1993)同上; Kunkel 等人(1993)同上; Whalen 等人，(1991)同上; Innes 等人，(1993)同上。一个名为*rps2-101N*的突变体也被分离出来。*RPS2*基因的鉴定和克隆将在下面描述。

RPS2 克服对携带 *avrRpt2* 基因的病原体的敏感性

为了证实病原体中的无毒基因与宿主中的抗病基因之间的遗传学关系，有必要首先分离出一个无毒基因。通过对已知的与拟南芥属农作物病原体假单孢菌属菌株的筛选，得到了高毒性的菌株 maculicola 丁香假单胞菌(*Psm*)ES4326, 番茄丁香假单胞菌(*Pst*)DC300, 和一个无毒菌种 *Pst* MM1065,并分析了它们在野生型拟南芥植株中的相对生长能力(Dong 等，(1991)植物细胞, 3:61-72; Walen 等，(1991)植物细胞 3:49-59; MM1065 在 Walen 等的文章中被称为 JL1065 ）。*Psm* ES4326 或 *Pst* DC3000 两天后能在拟南芥叶片中扩增 10^4 倍并导致水渍样病斑，*Pst* MM1065 48 小时后在拟南芥叶片中最多只能扩增 10 倍，并导致轻度枯黄病斑。因此，抗病性就与严格抑制病原体的生长有关。

利用下面文献中的标准技术克隆出了 *Pst* MM1065 菌株中的无毒（ avr ）基因; Dong 等（ 1991 ），植物细胞 3:61-72 ; Whalen 等，（ 1991 ）同上; Innes 等，（ 1993 ）同上。从该菌株中分离出的无毒基因称为 *avrRpt2* 。正常情况下，如前所述，毒性菌株 *Psm* E4326 或 *Pst* DC3000 在 48 小时后引起病症的出现。相比之下，*Psm* ES4326/ *avrRpt2* 或 *Pst* DC3000/ *avrRpt2* 在 16 小时内诱导产生可见的坏死性过敏反应（ HR ），且其繁殖倍数比野生型拟南芥叶片中的 *Psm* ES4326 或 *Pst* DC3000 低 50 倍（ Dong 等，（ 1991 ），同上; Whalen 等，（ 1991 ）同上）。因此，野生型拟南芥的抗病性至少要求有一个存在于病原体中的无毒基因或无毒基因产生的一个信号。

利用用克隆的 *avrRpt2* 基因寻找能引起对携带 *avrRpt2* 基因的病原体的抗病性的突变基因的方法分离出了四个拟南芥抗病突变体（ Yu 等，(1993), 同上; Kankel 等(1993), 同上）。这四个拟南芥突变体在受到 *Psm* ES4326/*avrRpt2* 或 *Pst* DC3000/ *avrRpt2* 感染时不能产生过敏反应，就像失去了抗病能力的植株一样。在其中一个突变体中，约

3000 株 5 至 6 周龄的经过乙基甲磺酸（ EMS ）诱变得到的 M_2 生态型哥伦比亚(Col-0 植株）植株被人工接种 *Psm* ES4326/ *avrRpt2*，从中得到了一个单突变体 *rps2-101C*, (resistance to Psendomonas syringae (抗丁香假单胞菌）)，（ Yu 等，（ 1993 ），同上）。

对第二个突变体，用一种能特异性扩增不能产生过敏反应的突变体的方法进行了分离（ Yu 等，(1993),同上）。对生长于培养皿中的 10 天龄的拟南芥幼苗进行感染 Phaseolicola 丁香假单胞菌（ *Psp* ） NPS3121 和感染 *Psp* NPS3121/*avrRpt2* 的对照实验，发现约 90 % 的被 *Psp* NPS3121 感染的植株存活，但约 90 % - 95 % 的被 *Psp* NPS3121/*avrRpt2* 感染的植株死亡。显然， *Psp* NPS3121/*avrRpt2* 感染完整的拟南芥幼苗诱导产生了系统过敏反应，后者常能使幼苗致死。在对照实验中，感染了 Psp NPS3121 的幼苗能够存活是因为 *Psp* NPS3121 对拟南芥是弱病原体。用 *Psp* NPS3121/*avrRpt2* 感染 4000 株 EMS-诱变的哥伦比亚 M_2 幼苗分离到了第二个抗病突变体。得到了 200 个存活株。这些存活株被移植到土壤中，当植株进入成熟期时利用人工接种的方法重筛选。在这 200 株中，有一株在被人工接种 *Psm* ES4326/ *avrRpt2* 时不能产生过敏反应，该突变体被称为 *rps2- 102C* (Yu 等，(1993),同上)。

第三个突变体 *rps2-201C*, 是在约 7500 株用双环氧丁酮诱变的拟南芥生态型 Col-0 的种子生长而成的 M_2 植株中筛选得到的（ Kunkel 等，（ 1993 ），同上）。将这种植株的完整叶片莲座浸泡到含有 Pst DC3000/ *avrRpt2* 细菌和表面活性剂 Silwet L-77 的溶液中以使植株被接种（ Whalen 等，(1991), 同上）。让植株在人工控制的环境中生长 3 至 4 天，此时可见病症的发展。通过筛选得到了四个突变系（含有等位基因 *rps2-201C*, *rps2-202C*, *rps2-203C* 和 *rps2-204C* ），其中 *rps2-201C* 纯合体植株是进一步研究的首选植株（ Kunkel 等，(1993). 同上和直接应用）。

第四个 *rps2* 的突变体 *rps2-101N* 的分离尚未见报导，这第四个被分离出来的或是一个突变体，或是一个敏感的拟南芥生态型。拟南芥 Nossen 生态型的种子经过γ - 射线照射后被密集地播种到平板上，让其穿过一张尼龙网发芽和生长。当生长到 5 至 6 周时，翻转平板，使植株部分浸入在含有 *Psm* ES4326/*avrRpt2* 培养物的盘子中，然后让植株在真

空干燥器中被感染。用该法接种的植株能在 24 小时内产生过敏反应。利用这种筛选方法筛选了约 40000 株植株，发现了一个敏感植株。随后的 RFLP 分析表明该植株可能不是一个 Nossen 突变体，而是另一个对 *Psm* ES4326/ *avrRpt2* 敏感的拟南芥生态型。该植株被称为 *rps2-101N*。被分离出来的突变体 *rps2-101C*, *rps2-102C*, *rps2-201C* 和 *rps2-101N* 被统称为“ *rps2* 突变体”。

rps2 突变体不能对克隆的无毒基因 *avrRpt2* 产生特异性反应

RPS2 基因产物是对携带有无毒基因 *avrRpt2* 的病原体产生抗性所特别必需的。Rps2 多肽上一个导致其功能丧失或降低的突变表现为不能对病原体感染产生过敏反应。当被 *Psm* ES4326/ *avrRpt2*， *Pst* DC3000/ *avrRpt2* 或 *Pst* NPS3121 / *avrRpt2* 感染时， *rps2* 突变体表现出病症或无效反应。特别是没有诱导产生过敏反应，这表明该植株是敏感的并且已丧失对病原体的抗性，即使病原体含有 *avrRpt2* 基因。

不管 *avrRpt2* 基因存在与否，病原体在 *rps2* 突变植株叶片中的生长情况是相同的。对比 *Psm* ES4326 和 *Psm* ES4326 / *avrRpt2* 在 *RPS2* 突变体中的生长情况发现二者在 *RPS2* 突变体中繁殖得同样好，且其速率与 *Psm* ES4326 在野生型拟南芥叶片中的一样。 *Pst* DC3000 和 *Pst* DC3000/ *avrRpt2* 在 *rps2* 突变体中的生长情况与之相似。

当 *rps2* 突变体受到携带有其他无毒基因如 *Psm* ES4326/ *avrB*， *Pst* DC3000/ *avrB*， *Psm* ES4326/ *avrRpml*, *Pst* DC3000/ *avrRpml* 的假单胞菌病原体感染时能产生过敏反应。能对 *avrRpt2* 之外的无毒基因产生过敏反应表明用 *avrRpt2* 筛选分离得到的 *rps2* 突变体对 *avrRpt2* 具有特异性。

RPS2 基因的作图和克隆

对 *rps2* 突变体 *rps2-101C*， *rps2-102C*， *rps2-201C* 和 *rps-101N* 的遗传学分析表明它们都象单孟德尔基因座一样分离，而且它们四个很象等位基因。采用含有基于聚合酶链式反应(PCR)的标记的标准 RFLP 作图法发现这四个 *RPS2* 突变体位于染色体 IV 的底部(Yu 等, (1993), 同上; Kunkel 等, (1993), 同上; 和 Mindrionos, M. , 未发表）。分离分析表明 *RPS2-101C* 和 *RPS2-102C* 与 PCR 标记 PG11 紧密连锁，同时

用 RFLP 标记 M600 来确定 *RPS2-201C* 突变体在染色体上的位置（图 1A）（Yu 等，(1993)，同上；Kunkel 等，(1993)，同上）。随后 *RPS2* 被定位在 PG11 的靠近着丝粒的一端。

RPS2/rps2 杂合体植株表现出的防御反应介于野生型和 *rps2/rps2* 纯合体突变植株之间（Yu 等，(1993)，同上；和 Kunkel 等，(1993)，同上；）杂合体植株在 *Psm* ES4326/ *avrRpt2* 或 *Pst* DC3000/ *avrRpt2* 感染时能产生过敏反应，但过敏反应的出现比野生型植株慢且需要更高的最低种菌浓度（Yu 等，(1993)，同上；和 Kunkel 等，(1993)，同上）。

RPS2 基因的高分辨率图谱和 *RPS2* cDNA 的分离

为了在图谱基础上克隆 *RPS2* 基因，将 *rps2-101N* 与 *rps2-101N* 与 Landsberg *erecta RPS2/RPS2* 杂交。F_1 代植物自花传粉所得的 165 个 F_2 代植株又自花传粉产生 F_3 代。标准的 RFLP 绘图方法表明：*rps2-101N* 图谱与 RFLP 标记 PG11 相邻且在 PG11 的靠近着丝粒的一端。为了得到更为详细的图谱位置，将 *rps2-101N/rps-101*N 与含有隐性突变 *cer2* 和 *ap2* 的双标记 Landsberg erecta 品系杂交。*cer2* 和 *ap2* 的遗传图距约为 15cM，*rps2* 基因座就在这一区间内，收集表现 *CER2 ap2* 或 *cer2 AP2* 基因型 F_2 代植株，让其自交，对每一 F_2 代植株的至少 20 个 F_3 代植株接种 *Psm* ES4326/*avrRpt2*，以记录 *RPS2*。在从包含每一个 F_2 系的约 20 个 F_3 代植株的库中制备 DNA。利用 *CER2 ap2* 和 *cer2 AP2* 重组体进行图 1 所示的染色体步查。

如图 1 所示，*RPS2* 是图谱中一个被科斯质粒克隆 E4-4 和 E4-6 跨越的 28-35 kb 区域。该区域至少含有 6 个能产生可检测的转录本的基因。RNA 印迹分析表明它们转录本的大小或它们在 *rps2* 突变体中的表达量没有大的差别。分离得到了每一个转录本的 cDNA 克隆，且对其中 5 个进行了测序，正如下面将讲到的，其中的一个转录本 cDNA－4 与 *RPS2* 的基因座位一致。通过这一研究，从哥伦比亚生态型野生型植株中获得了三个与 *RPS2* 相对应的的相互独立的 cDNA 克隆（cDNA-4-4, cDNA-4-5, 和 cDNA-4-11）。RNA 印迹分析发现 *RPS2* 转录本的表观大小为 3.8 和 3.1 kb.。

第四个独立的 cDNA-4 克隆（cDNA-4-2453）是在一项独立研究中利用在图谱基础上分离 *RPS2* 的方法获得的。含有连续的、重叠的拟南

芥生态型Col-0基因组DNA插入片段的酵母人工染色体(YAC)克隆被鉴定出来，该插入片段是从*RPS2*的M600区开始的，跨度约900kb。拟南芥YAC文库见J. Ecker和E. Ward,同上；以及E. Grill(Grill和Somerville(1991)分子遗传学 226-484-490)。科斯质粒“H”和“E”是从YAC插入片段中获得的，并被用于*RPS2*的分离（图1）。

*RPS2*的遗传学和物理位置已用物理作图的RFLP、RAPD（随机扩增多态性DNA）和CAPS（断裂式扩增多态性序列）标记更精确地确定。利用基因型*RPS2/RPS2*(No-0野生型)和*rps2-201/rps2-201*(Col-0背景）的植株杂交后的性状分离数进行遗传学绘图。用17B7LE，PG11，M600和其他标记对*RPS2*座位绘图。为进行高分辨率遗传学作图，利用YAC和科斯质粒克隆的插入端片段产生了一系列紧密连锁的RFLP标记（图1）（Kunkel等，（1993），同上；Konieczny和Ausubel(1993)植物学杂志4:403-410；和Chang等，(1988)美国国家科学院院刊85:6856-6860)。然后利用科斯质粒克隆E4-4和E4-6来鉴定表达的转录本（图1F中的cDNA-4,-5,-6,-7,-8），包括cDNA－4－2453克隆。

*RPS2*的DNA序列分析

对来自野生型Col-0植株突变体*rps2-101C*, *rps2-102C*, *rps2-201C*和*rps2-101N*的cDNA-4的DNA序列分析表明cDNA-4与*RPS2*一致。对*rps2-101C*, *rps2-102C*和*rps2-201C*的DNA序列分析显示了其与野生型序列之间的不同（见表1）。表1中的数字是从编码第一个甲硫氨酸的密码子ATG开始的，其中以A为第一个核苷酸。对与突变体*rps2-102C*相对应的cDNA－4的DNA序列分析表明它与野生型在第476个氨基酸残基上有差别。对与*rps2-101N*上的cDNA－4相对应的cDNA的DNA序列分析表明它在富含亮氨酸重复区域（LRR）内的第581个氨基酸残基处有一个长10个碱基对的插入片段。这导致*RPS2*阅读框架的移动。突变型*rps2-101C*含有一个导致链终止密码子形成的突变。突变体等位基因*rps2-201C*的DNA序列表明其与螺旋－环－螺旋(Helix-loop-Helix)位元(motif)有相似性的LRR区内的一个氨基酸发生了突变，这进一步支持了认为该位置就是*RPS2*基因的看法。DNA和氨基酸序列见图2[分别为SEQ.ID.NO:1和SEQ.ID NOS:2-5]。

表1

突变体	野生型	位置	突变
rps2-101C	703 TGA 705	704	TAA 终止密码子
rps2-101N	1741 GTG 1743	1741	GTGGAGTTGTATG 插入片段
rps2-102C 476	1426 AGA 1428 精氨酸	1427	AAA 氨基酸 赖氨酸
rps2-201C	2002 ACC 2004 苏氨酸	2002	CCC 氨基酸 脯氨酸

对与野生型 Col-0 植株的 *RPS2* 相对应的 cDNA - 4 的 DNA 序列分析发现了一个跨度为 2,751bp 的开放阅读框架（在两个终止密码子之间）。该阅读框架中，从第一个甲硫氨酸密码子到 3' - 终止密码子共 2727bp, 编码一段 909 个氨基酸的多肽（见图 2 中的开放阅读框架“a”）。该氨基酸序列的分子量为 104,460, PI 为 6.51。

RPS2 属于一类新型抗病基因， Rps2 多肽的结构不象原来克隆并发表的唯一的一个无毒基因特异性的植物抗病基因 *Pto* 的产物的蛋白质结构，后者有一个可能的蛋白激酶结构域。通过上述对推测的氨基酸序列的分析， *RPS2* 含有几个在其它真核和原核蛋白中保守不同的蛋白结构域。这些结构域包括但不限于富含亮氨酸重复区（ LRR ）（ Kobe 和 Deisenhofer, (1994)自然 366:751-756); P-环（ P-loop) Saraste 等，（ 1990)生物学的趋势（ TIBS ） 15:430-434 ; 螺旋 - 环 - 螺旋， Murre 等，（ 1989)细胞 56:777-783; 亮氨酸拉链(leucine Zipper)Rodrigues 和 Park(1993)分子细胞生物学 13:6711-6722)。 *RPS2* 的氨基酸序列中含有一个 LRR 位元（ LRR 位元从第 505 个氨基酸残基到第 867 个氨基酸残基）。该位元存在于许多已知蛋白质中且被认为与蛋白 - 蛋白相互作用有关，也许因此它能与其他和植物抗病性有关的蛋白质相互作用。 Rps2 多肽 LRR 的 N-端部分，如啤酒酵母（ *Saccharomyces cerevisiae*)LRR 的 N - 端部分是腺苷酸环化酶 CYR1。一个推测为跨膜结构域的区域（ Klein 等，（ 1985 ） Biochim., Biophys, Acta 815:468-476) 位于 LRR N - 端的第 350 个氨基酸残基到第 365 个氨基酸残基之间。一个 ATP/GTP 结合位点位元（ P - 环）被推测为位于第 177 个氨基酸残基和第 194 个氨基酸残基之间。

通过以上对推测的氨基酸序列的分析，Rps2 多肽可能有一个膜受体结构。该结构由一个 N－端胞外区和一个 C－端胞质区组成。另一方面，Rps2 的拓扑学结构可能正相反：含一个 N－端胞质区和一个 C－端胞外区。LRR 位元通常在胞外区，Rps2 的 LRR 含有 5 个潜在的 N－糖基化位点，

用功能互补法鉴定 *RPS2*

rps2-201 纯合体和拟南芥相应的基因组 DNA 的互补实验从功能上证实了编码 cDNA－4 的基因组区域带有 *RPS2* 活性。从 YACs EW11D4，EW9C3 和 YUP11F1（见图 1 和图 4）上的 *RPS2* 区构建了含有野生型拟南芥 DNA 连续重叠序列的科斯质粒。科斯质粒载体是由 pSLJ4541（得自 J. Jones, Sainsbury 研究院，Norwich, 英格兰）构建而成（称为“双科斯质粒”），其中 pSLJ4541 含有允许在土壤杆菌介导的转化过程中将插入片段整合到植物基因组中的序列。科斯质粒“H”和“E”（图 1）被用于鉴定含有拟南芥基因组 *RPS2* 区域的克隆。

利用土壤杆菌介导的转化，40 个以上的含有插入的 *RPS2* 区 DNA 的双科斯质粒被用于转化 *rps2-201* 纯合突变体（Chang 等，(1990) 第四届国际拟南芥研讨会会议摘要，第 28 页，维也纳，奥地利）。重组子仍为敏感型（用观察在感染携带 *avrRpt2* 的 Psp 3121 和不含 *avrRpt2* 的 *Psp* 3121 后均不能产生过敏反应等方法确定），这表明插入的 DNA 不含功能性 *RPS2*。这些科斯质粒能产生图 4 所示的“Sus.”或敏感表现型。获得 *avrRpt2*-特异性抗病性的重组子（用观察在感染带有 *avrRpt2* 的 *Psp* 3121 后能产生强过敏反应，而感染不含 *avrRpt2* 的 *Psp* 3121 后不产生过敏反应等方法来确定）表明插入的 DNA 含有能产生图 4 所示的“Res.”或抗病表现型的功能性 *RPS2* 基因，用 pD4 双科斯质粒得到的转化子表现出上述强过敏反应。经典遗传学分析（用共转化选择标记物赋予的抗病表现型和抗卡那霉素表现型的紧密连锁）和 Southern 杂交分析证实转化子中含有插入 DNA。这些结果表明 *RPS2* 被科斯质粒 pD4 (见图 4）上的一个 18kb 的拟南芥基因组片段所编码。

为了进一步确定 *RPS2* 基因座的位置和证实它能使 *rps2-201* 纯合突变体产生抗病毒表现型，测验了六个含有部分重叠基因组 DNA 插入片段的双科斯质粒。根据与 *RPS2* 区五个 cDNA 克隆相应的转录位置选出了重叠插入子 pD2, pD4, pD14, pD15, pD27 和 pD47（图 4）。在这些

转化实验中，采用真空感染法（ Bechtolcd 等，(1993) C. R. Acad. Sci. 巴黎，316:1194-1199 ）进行土壤杆菌介导的转化，采用根转化/再生方法进行了科斯质粒 pD2, pD14, pD15, pD39 和 pD46 的土壤杆菌介导的转化（ ValveeKens 等，（ 1988 ），PNAS 85:5536-5540 ）。分析 *RPS2* 活性的病原体接种实验的结果见图 4 。

利用含有完整编码区和 cDNA-4 或 cDNA-6 基因组片段上游 1kb 以上区段的双科斯质粒进行了附加的转化实验。用真空感染转化法获得了三个独立的转化子，它们含有 *rps2-201C* 纯合背景的野生型 cDNA-6 基因组片段（ 图 4 中的 pAD431 ）。这些植株都不表现 *avrRpt2* 依赖性的抗病性。用野生型基因组 cDNA-4 (p4104 和 p4115，每一个都含有与所有 cDNA-4 的开放阅读框架一致的 Col-0 基因组片段，再加上约 1.7kb 的 5'上游片段和终止子 3'下游约 0.3kb 的片段转化 *RPS2*-201C 纯合的突变体。这些 p4104 和 p4115 转化子表现出与野生型 *RPS2* 纯合体（ *rps2* 即来自该纯合体）相似的抗病表现型。当用 cDNA - 4 基因组片段转化时，突变体(*rps2-101N* 和 *rps2-101C* 纯合体）也表现出 *avrRpt2* 依赖性的抗病性。

用 *RPS2* 序列检测其它抗性基因

利用 *RPS2* cDNA 作为探针对拟南芥基因组 DNA 进行 DNA 印迹法分析，发现拟南芥含有几个能与 *RPS2* 或 *RPS2* 的一部分杂交的 DNA 片段，这说明拟南芥基因组中有几个相关基因。

从以上描述及图 2 所示的核苷酸序列 [SEQ ID NO:1] 可知，分离出的与 *RPS2* 基因有 50 % 或更高序列一致性的其它植物抗病基因是可能的。可以用含有 *RPS2* 基因或其中长于 18 个核苷酸的片段的寡聚核苷酸探针来进行这种检测和分离。编码 RPS2 多肽 [SEQ. ID. NOS:2-5] 的特定结构特征的序列的探针优选地是能分离具有相似结构域的各种抗病基因的探针。利用标准的技术进行杂交，例如 Ausubel 等，《当代分子生物学方法》, John Wiley & Sons, (1989)。

例如，检测 *RPS2* 基因的高度严格的条件包括：杂交：约 42 ℃，约 50 % 甲酰胺；第一次洗：约 65 ℃，约 2x SSC, 1% SDS；第二次洗：约 65 ℃，约 0.1% SSC，检测与 *RPS2* 基因有 5 % 序列一致性的 *RPS* 基因的低严格条件如杂交：约 42 ℃，无甲酰胺；第一次洗：约 42 ℃，约 6x SSC，约 1% SDS；第二次洗：约 50 ℃，约 6X SSC，约 1 % SDS。

上述例子中所用的探针是一个约350核苷酸的编码Rps2的LRR区域中间部分的DNA片段。在低严格条件下，用BamHI消化的拟南芥基因组DNA中至少有5条DNA带与编码RPS2 的LRR位元中间部分的DNA有足够的序列一致性使之可以杂交。用编码RPS2的LRR位元以外的Rps2的N－端*RPS2*基因中一段300个核苷酸的部分作为探针得到了相似的结果。

其他抗病基因的分离采用了分子生物学领域技术人员熟知的PCR扩增技术，利用设计的寡聚核苷酸引物只扩增与*RPS2*有序列一致性的基因中位于两个寡聚核苷酸之间的那一段序列。引物被优选地设计为便于扩增的产物能克隆到合适的载体中去。

*RPS2*在转基因植物细胞和植物中的表达

用将含有能表达RPS2多肽的*RPS2*基因的一个DNA片段导入植物的方法已成功地使*RPS2*基因在对携带*avrRpt2*的病原体敏感的植物中表达。已经有了大量适合植物细胞稳定转染或适合构建转基因植物的载体; 有关这些载体的描述见Pouwels等,《克隆载体实验手册》, 1985, Supp, 1987; Weissbach和Weissbach,《植物分子生物学方法》, 学术出版社, 1989; 以及Gelvin等, 《植物分子生物学手册》, Kluwer 学术出版社,1990。典型的植物表达载体包括(1)在5'和3'调节序列的转录控制之下有一个或多个克隆的植物基因, 和(2)一个明显的选择标记。如果需要, 这些植物表达载体可能包含一个启动子调节区域(例如一个控制可诱导的或组成型的、环境调节的或发育调节的、细胞特异性或组织特异性表达的调节区域), 一个转录起始位点, 一个核糖体结合位点, 一个RNA加工信号, 一个转录终止位点, 和/或一个多聚腺苷酸信号。

根据本发明, 能用于植物抗病基因表达的有用的植物启动子的例子是一个花叶病毒的启动子, 例如花椰菜花叶病毒(CaMV)的35S启动子。这些启动子能引起在大多数植物组织中的高水平表达, 而且这些启动子的活性不依赖于病毒编码的蛋白质。 CaMV是35S和19S启动子的来源。在大多数转基因植物组织中, CaMV 35S启动子是一个强启动子(例见Odel等,自然 313:810,(1985)) MV启动子在单子叶植物中也具高活性(见Dekeyser等, 植物细胞2:591, (1990); Terada和Shimamoto, 分子遗传学和遗传工程 220:389, (1990))。

其他有用的植物启动子包括但不限于 nopaline 合成酶启动子（ An 等，植物生物学 88:547,（1988）)和章鱼碱合成酶启动子（ Fromm 等，植物细胞 1:977,（1989）)。

在某些应用中，可能需要在一种合适的组织中，以一个合适的水平，或在一个合适的发育阶段生产 *RPS2* 基因或 *avrRpt2* 基因产物。因此，存在着各种基因启动子，每一个启动子的调节序列中都有其独有的特性，根据环境，激素和/或发育情况进行相应的调节。这些启动子控制的有(1)热调节的基因表达（例见 Callis 等，植物生理学 88:965,（1988））,(2)光调节的基因表达（如豌豆 rbcS-3A 启动子，见 Kuhlemeier 等，植物细胞（ Plant Cell)1:471,（1989））;玉米 rbcS-启动子，见 Schaffner 和 Sheen，植物细胞 3:997,（1991）；或在豌豆中发现的叶绿素 a / b - 结合蛋白基因，见 Simpson 等， EMBO J. 4:2723,（1985）),(3)激素调节的基因表达（例如小麦 Em 基因中的抗坏血酸敏感序列，见 Marcotte 等，植物细胞 1:969,（1989））,(4)创伤诱导的基因表达（例如 *Wun I*，见 Siebertz 等，植物细胞 1:961,(1989)），或(5)器官特异性的基因表达（例如块茎特异性的贮藏蛋白基因，见 Roshal 等， EMBO J. 6:1155,(1987)；玉米中的 23-kDa 玉米醇蛋白基因，见 Schernthaner 等，EMBO J. 7:1249,（1988）；或法国菜豆β - 菜豆蛋白基因，见 Bustos 等，植物细胞 1:839,（1989））。

植物表达载体可以含有 RNA 加工信号，例如内含子，它对于高效的 RNA 合成和积累是非常重要的（ Callis 等，基因和发育 1:1183,（1987））。 RNA 拼接序列的位置能影响植物中转基因的表达水平。由此看来，要调节基因表达水平，转基因中编码 RPS2 多肽的序列的上游或下游应当有一个内含子。

除了上述 5’ 调控序列，表达载体还应当包括调控区，后者广泛存在于植物基因的 3’ 区（ Thornburg 等，美国国家科学院院刊 84:744,（1987）; An 等，植物细胞 1:115,（1989））。例如，表达载体中包含 3’ 终止子区可增加 mRNA 的稳定性。从土豆的 PI - II 终止子区可以获得一个这样的终止子区。此外，其他常用终止子区来自章鱼碱或 no-paline 合成酶信号。

典型的植物表达载体也包括一个明显的选择标记基因，用于鉴定转化的细胞。用于植物系统的选择标记基因包括编码抗生素抗性的基因，

例如编码对潮霉素，卡那霉素，博莱霉素， G418，链霉素或壮观霉素的抗性的基因，光合作用必需基因可作为光合作用缺失品系的选择标记。编码对除草剂抗性的基因也可作为选择标记；有用的除草剂抗性基因有编码 phosphinothricin 乙酰转移酶的 bar 基因，它能引起对广谱除草剂 Basta® (Hoechst AG, 法兰克福，德国）的抗性。

对植物细胞对某一特定的选择性试剂的敏感性的测定，以及对能有效地杀死即使不是全部，至少也是大多数转化细胞的上述选择性试剂的浓度的测定，使选择性标记的有效利用变得更加容易。用于烟草转化的抗生素浓度如下： 75-100 g /ml（卡那霉素）， 20-50μg /ml（潮霉素），或 5-10μg /ml（博莱霉素）。选择抗除草剂的转化子的步骤见例如 Vasil I. K.,《植物细胞培养和体细胞遗传学》，第 I， II， III 卷，实验方法及其应用学术出版社， 纽约， 1984 。

精通植物分子生物学的人都清楚，基因表达的水平显然不仅与启动子， RNA 加工信号和终止子元件的重组有关，还与这些元件如何被用于提高基因表达水平有关。

植物的转化

由于植物表达载体的构建，有几种标准的方法可将重组遗传物质导入宿主植物中以产生转基因植物。这些方法包括（ 1 ）土壤杆菌介导的转化（根瘤土壤杆菌或发根病土壤杆菌）（例见 Lichtenstein 和 Fuller，遗传工程，第 6 卷， PWJ Rigby 编，伦敦，学术出版社， 1985 ），（ 2 ）颗粒递呈系统（例见 Gordon -Kamm 等，植物细胞 2 : 603 ，（ 1990 ）；或 BioRad 技术公告 1687, 同上），（ 3 ）微注射方法（例见 Green 等，植物组织和细胞培养）,学术出版社，纽约， 1987 ），（ 4 ）聚乙二醇法（例见 Draper 等，植物细胞生理学） 23 : 451 ，（ 1982 ）；或 Zhang 和 Wu, 遗传学理论和应用 76:835 ，（ 1988 ））。（ 5 ）脂质体介导的 DNA 吸收法（例见 Freeman 等，植物细胞生理学 25:1353 ，（ 1984 ）），（ 6 ）电融合法（例见 Gelvin 等，同上； Dekeyser 等，同上；或 Fromm 等，自然 319:791,（ 1986 ）），和（ 7 ）涡旋法（例见 Kindle, K.，美国国家科学院院刊 87:1228,（ 1990 ））。

下面是一个土壤杆菌介导的植物转化的实施例。对将转移到植物细胞基因组中的基因的处理一般分为两个阶段。首先，在大肠杆菌

（*E.coli*）中进行所有的克隆和DNA修饰步骤，用接合法使含有所构建的目的基因的质粒转移到土壤杆菌中。然后，用所得到的土壤杆菌转化植物细胞。对于普通的植物表达载体来说，质粒含有一个能在土壤杆菌中复制的复制原点和在大肠杆菌中高拷贝数复制的复制原点。因此，在转移到土壤杆菌和导入到植物中以前就能很容易地在大肠杆菌中生产和检测转基因。载体上能携带抗性基因，一个用于细菌的筛选，例如抗链霉素基因，其余的将在植物中表达，例如编码对卡那霉素抗性的基因或抗除草剂基因。此外质粒还具有限制性内切酶位点以便能加入一个或多个与合适的调节序列可操纵性连锁的转基因和定向T－DNA边界序列，后者在被土壤杆菌的转移功能识别时，能界定将转移到植物中去的区域。

在另一个实施例中，可用将沉积有DNA的钨微弹射入植物细胞中的方法转化植物细胞。在射击用的Biolistic仪器（Bio-Rad, Hercules, 加拿大）中，火药冲力（管径为22的活塞式冲力）或空气驱动力推动一个大塑料发射粒穿过枪管。大塑料发射粒的前部装有含有DNA沉淀的钨粒悬浊液。发射粒在点火后冲向丙烯纤维档板，其上有一个大发射粒不能穿过的小孔。结果，大塑料发射粒撞击在档板上，而钨微弹继续穿过板上的小孔射向靶。本发明中的靶可以是任何植物细胞、组织、种子或胚。被导入细胞中的微发射粒上的DNA嵌入到核中或叶绿体中。

现在，植物细胞中转基因的转移和表达对于本领域的技术人员来说已是常规操作。它已成为研究基因表达、增加农作物多样性和提高商业利益的主要方法。

转基因植物的再生

用植物表达载体转化的植物细胞可以用标准的植物组织培养技术，从如单个细胞、愈伤组织或叶盘再生。本领域的技术人员熟知，来自几乎任何植物的各种细胞、组织和器官都能成功地通过培养再生成一个完整的植株。这样的技术例见Vasil, 同上; Green等, 同上; Weissbach和Weissbach 同上; 和Gelvin等,同上。

一个可能的实施例是，将一个携带有一个选择标记基因（如卡那霉素抗性）和一个克隆的 *RPS2* 基因的载体转化到土壤杆菌中，其中的 *RPS2* 基因处于其自身的启动子和终止子控制下，如果需要的话，也可处于外源调控序列比如35S CaMV启动子和nopaline合成酶终止子的控

制之下。用含有载体的土壤杆菌转化叶组织，参见 Horsch 等（科学 227:1229,（1985））。几周（如 3-5 周）后在含有卡那霉素（如 100μg /ml）的植物组织培养基上筛选可能的转化子，然后将具有卡那霉素抗性的茎放在不含激素的植物组织培养基上使其生根，然后选出具有卡那霉素抗性的植株在温室中生长。如果需要，可将自花授粉的转基因植物的种子播种到无土培养基上并在温室中生长， 用在无激素的卡那霉素培养基上播种表面灭菌的种子的方法来筛选具有卡那霉素抗性的后代，可用标准技术分析转基因的插入（见 Ausubel 等，同上； Gelvin 等，同上；）

表达选择标记的转基因植物可用标准的免疫印迹法以及 DNA 和 RNA 检测技术来筛选转基因 DNA 的遗传。每个阳性转基因植物和它的转基因后代在与其他用相同转基因得到的转基因植物相比较时是不同的。在大多数情况下，转基因 DNA 是随机地整合到植物基因组 DNA 中去的，整合位点严重影响转基因表达的水平、组织以及发展模式。因此，对于每一个转基因通常要筛选大量转基因品系以鉴定和选出具有最合适的表达方式的植物。

用转基因表达的水平来评价转基因品系。首先用 RNA 水平上的表达对阳性表达植物进行鉴定和定量。采用了标准的 RNA 分析技术，包括用所设计的寡聚核苷酸引物通过 PCR 扩增专一性地扩增转基因 RNA 模板，以及用转基因特异性的探针做溶液杂交分析（例见 Ausubel 等，同上）。然后，用 RPS2 多肽特异性抗体进行 Western 免疫印迹分析以分析 RNA 阳性植物中的蛋白质的表达情况（例见 Ausubel 等，同上），此外，用转基因特异性的核苷酸探针和抗体做标准的原位杂交和免疫细胞化学分析，可以确定转基因组织中的表达位点。

一旦 RPS2 多肽在任何细胞或转基因植物中表达（例如以上的描述），就能用任何标准技术如亲和层析法将它分离出来。举例来说，一种抗 RPS2 抗体（如用 Ausubel 等，同上所描述的方法或任何标准技术制得）可吸附到柱子上并被用于多肽的分离。根据标准方法，在亲和层析之前要对产 *RPS2* 细胞进行裂解和分级分离（例见 Ausubel 等，同上）。如果需要，分离后的重组多肽可进一步被纯化，如用高效液相色谱法（ HPLC ）（例见 Fisher,生化和分子生物学实验技术， Work 和 Burdon, 编辑， Elsevier， 1980 ）。

这些多肽表达和纯化的普通技术也可用于生产和分离有用的 *RPS2*

片段或其类似物。

用途

将 *RPS2* 导入转化的植物细胞中可使植物获得对携带 *avrRpt2* 无毒基因的细菌病原体的抗性。例如用本发明中的表达 *RPS2* 的转基因植物可以改变原来对携带无毒基因 *avrRpt2* 的植物病原体敏感的植物的抗病性，该法简单且费用低廉。

本发明还可模拟宿主抗性的天然机制使植物获得对广谱病原体的抗性。首先，在没有病原体信号的条件下使 *RPS2* 转基因在植物细胞中高水平表达以产生组成型的植物防御反应。与植物防御反应的产生相联系的表达水平可通过防御反应基因的表达水平来测定，见 Dong 等，同上。然后，利用可控制的启动子如组织特异性启动子、细胞型特异性启动子或能被外源信号或试剂诱导的启动子使 *RPS2* 转基因表达，以限制防御反应中暂时的组织表达。最后，使 *RPS2* 基因产物和 *avrRpt2* 基因产物共表达。使 *RPS2* 基因表达的有其天然启动子，组成型表达启动子比如 CaMV 35S 启动子，组织特异性或细胞型特异性启动子或能被外源信号或试剂激活的启动子。 *RPS2* 和 *avrRpt2* 的共表达类似于与植物防御反应起始相关的基因产物的生成，同时在宿主植物和病原体中不存在特异性的抗性基因—无毒基因对的情况下能使植物获得对病原体的抗性。

本发明还提供用于在植物细胞中表达的具有图 2 所示序列 [SEQ.ID.NO:1]的一段核酸，或编码图 2 中开放阅读框架“ a ”的氨基酸序列 [SEQ. ID.NOS:2-5] 的核酸的简并变种。

本发明还可以利用图 2 所示的 *RPS2* 序列 [SEQ. ID.NO:1] 或其中长度大于 18 个核苷酸的一段作为探针，来分离与 *RPS2* 有 50 % 或更高一致性的核酸序列。适当降低的杂交条件被用于分离与图 2 所示的 *RPS2* 序列 [SEQ. ID.NO:1] 有 50 % 或更高一致性的 DNA 序列。

本发明可使植物获得抗病性，主要是农作物，特别是重要的农作物如番茄、胡椒、玉米、小麦、水稻和豆科植物如大豆和菜豆，或是任何对携带无毒基因如无毒基因 *avrRpt2* 的病原体敏感的植物。这种病原体包括但不限于丁香假单胞菌菌株。

本发明还包括 Rps2 多肽的任何生物活性片段或其类似物。其中的“生物活性”是指在体内具有任何图 2 所示的 Rps2 多肽 [SEQ.

ID.NOS:2-5] 特征的活性。有用的 Rps2 片段或 Rps2 类似物是指在任何针对抗病基因产物活性的生物学试验中表现出生物学活性的那些片段或类似物，这些试验参见 Dong 等（ 1991 ），同上； Yu 等（ 1993 ）同上； Kunkel 等（ 1993 ）同上； Whalen 等（ 1991 ），同上。尤其是有生物活性的 Rps2 多肽片段或类似物能够使植物获得对携带无毒基因 *avrRpt2* 的病原体的基本抗性。基本抗性是指至少能部分降低植物对携带无毒基因 *avrRpt2* 的病原体的敏感性。

优选的类似物包括 Rps2 多肽（或其生物活性片段），其序列与野生型的序列的差别仅在于保守氨基酸的取代，例如一个氨基酸取代另一个特征相似的氨基酸（如缬氨酸取代甘氨酸，精氨酸取代赖氨酸等）或在于一个或多个非保守氨基酸的取代，缺失或插入，但不影响多肽的生物学活性。

类似物或与天然存在的 Rps2 多肽在氨基酸序列上有所不同，或能通过不涉及序列的途径得到修饰，或二者皆有。本发明中的类似物与天然存在的 Rps2 多肽序列 [SEQ. ID.NOS:2-5] 中一段长度为 20 个氨基酸残基，优选地为 40 个氨基酸残基，更优选地为全长序列有至少为 70 %，优选地为 80 %，更优选地为 90 %，最优选地为 95 %，甚至是 99% 的同源性。

一级结构的改变包括天然或诱导的遗传学变异，还包括含有除天然存在的 L - 氨基酸之外的残基（如 D - 氨基酸）或非天然存在或合成的氨基酸（如β 或γ 氨基酸）的类似物。本发明中还包括被体内的多肽化学衍生作用包括乙酰化，甲基化，磷酸化，羧化或糖基化所修饰的 Rps2 多肽。

除了基本全长多肽，本发明还包括多肽的具有生物活性的片段。在其中，术语“片段”，用于多肽，一般指其长度为至少 20 个残基，更典型的是指至少 40 个残基，优选地是指至少 60 个残基。 Rps2 多肽的片段可用本领域的技术人员所熟悉的方法来产生。可用其中所描述的方法使某个候选片段具有 *RPS2* 的生物学活性。本发明还包括含有生物活性非必需残基（如通过 mRNA 拼接的变化或蛋白质加工过程的变化而增加的残基）的 *RPS2* 多肽。

其他实施方案将在后面的权利要求中叙述。

序列表

(1)一般资料:

(i) 申请人: Ausubel, Frederick M.
Staskawicz, Brian J.
Brent, Andrew F.
Dahlbeck, Douglas
Katagiri, Fumiaki
Kunkel, Barbara N.
Mindrinos, Michael N.
Yu, Guo-Liang

(ii)发明名称: RPS2 基因及其应用

(iii)序列数: 106

(iv)地址:

(A) ADDRESSEE: Fish & Richardson
(B)街道: 225 Franklin Street Suite 3100
(C)城市: Boston
(D)州: MA
(E)国家: USA
(F) ZIP: 02110-2904

(v)计算机可读形式:

(A)介质类型:软盘
(B)计算机: IBM PC 兼容型
(C)操作系统: PC-DOS/MS-DOS
(D)软件: PatentIn Release #1.0, Version #1.30B

(vi)当前申请数据:

(A)申请号: US 08/227,360

(B)提交日期: 13-APR-1994

(C)分类:

(viii) 律师/代理人资料:

(A) 姓名: Clark, Paul T.

(B)注册号: 30,162

(C) 参考/登记号: 00786/230001

(ix)通讯资料:

(A)电话: (617) 542-5070

(B)电传: (617) 542-8906

(C) TELEX: 100254

(2) SEQ ID NO:1 的资料:

(i) 序列特征:

(A) 长度: 2903 个碱基对

(B) 类型: 核酸

(C) 链型: 单链

(D) 拓扑类型: 线性

(ii) 分子类型: DNA (基因组)

(xi) 序列描述: SEQ ID NO:1:

AAGTAAAAGA AAGAGCGAGA AATCATCGAA ATGGATTTCA TCTCATCTCT TATCGTTGGC 60

TGTGCTCAGG TGTTGTGTGA ATCTATGAAT ATGGCGGAGA GAAGAGGACA TAAGACTGAT 120

CTTAGACAAG CCATCACTGA TCTTGAAACA GCCATCGGTG ACTTGAAGGC CATACGTGAT 180

GACCTGACTT TACGGATCCA ACAAGACGGT CTAGAGGGAC GAAGCTGCTC AAATCGTGCC 240

AGAGAGTGGC TTAGTGCGGT GCAAGTAACG GAGACTAAAA CAGCCCTACT TTTAGTGAGG 300

TTTAGGCGTC GGGAACAGAG GACGCGAATG AGGAGGAGAT ACCTCAGTTG TTTCGGTTGT 360

GCCGACTACA AACTGTGCAA GAAGGTTTCT GCCATATTGA AGAGCATTGG TGAGCTGAGA 420

GAACGCTCTG AAGCTATCAA AACAGATGGC GGGTCAATTC AAGTAACTTG TAGAGAGATA 480

CCCATCAAGT CCGTTGTCGG AAATACCACG ATGATGGAAC AGGTTTTGGA ATTTCTCAGT 540

GAAGAAGAAG AAAGAGGAAT CATTGGTGTT TATGGACCTG GTGGGGTTGG GAAGACAACG 600

TTAATGCAGA GCATTAACAA CGAGCTGATC ACAAAAGGAC ATCAGTATGA TGTACTGATT 660

TGGGTTCAAA TGTCCAGAGA ATTCGGCGAG TGTACAATTC AGCAAGCCGT TGGAGCACGG 720

TTGGGTTTAT CTTGGGACGA GAAGGAGACC GGCGAAAACA GAGCTTTGAA GATATACAGA 780

GCTTTGAGAC AGAAACGTTT CTTGTTGTTG CTAGATGATG TCTGGGAAGA GATAGACTTG 840

GAGAAAACTG GAGTTCCTCG ACCTGACAGG GAAAACAAAT GCAAGGTGAT GTTCACGACA 900

CGGTCTATAG CATTATGCAA CAATATGGGT GCGGAATACA AGTTGAGAGT GGAGTTTCTG 960

GAGAAGAAAC ACGCGTGGGA GCTGTTCTGT AGTAAGGTAT GGAGAAAAGA TCTTTTAGAG 1020

TCATCATCAA TTCGCCGGCT CGCGGAGATT ATAGTGAGTA AATGTGGAGG ATTGCCACTA 1080

GCGTTGATCA CTTTAGGAGG AGCCATGGCT CATAGAGAGA CAGAAGAAGA GTGGATCCAT 1140

GCTAGTGAAG TTCTGACTAG ATTTCCAGCA GAGATGAAGG GTATGAACTA TGTATTTGCC 1200

CTTTTGAAAT TCAGCTACGA CAACCTCGAG AGTGATCTGC TTCGGTCTTG TTTCTTGTAC 1260

TGCGCTTTAT TCCCAGAAGA ACATTCTATA GAGATCGAGC AGCTTGTTGA GTACTGGGTC 1320

GGCGAAGGGT TTCTCACCAG CTCCCATGGC GTTAACACCA TTTACAAGGG ATATTTTCTC 1380

ATTGGGGATC TGAAAGCGGC ATGTTTGTTG GAAACCGGAG ATGAGAAAAC ACAGGTGAAG 1440

ATGCATAATG TGGTCAGAAG CTTTGCATTG TGGATGGCAT CTGAACAGGG GACTTATAAG 1500

GAGCTGATCC TAGTTGAGCC TAGCATGGGA CATACTGAAG CTCCTAAAGC AGAAAACTGG 1560

CGACAAGCGT TGGTGATCTC ATTGTTAGAT AACAGAATCC AGACCTTGCC TGAAAAACTC 1620

ATATGCCCGA AACTGACAAC ACTGATGCTC CAACAGAACA GCTCTTTGAA GAAGATTCCA 1680

ACAGGGTTTT TCATGCATAT GCCTGTTCTC AGAGTCTTGG ACTTGTCGTT CACAAGTATC 1740

ACTGAGATTC CGTTGTCTAT CAAGTATTTG GTGGAGTTGT ATCATCTGTC TATGTCAGGA 1800

ACAAAGATAA GTGTATTGCC ACAGGAGCTT GGGAATCTTA GAAAACTGAA GCATCTGGAC 1860

CTACAAAGAA CTCAGTTTCT TCAGACGATC CCACGAGATG CCATATGTTG GCTGAGCAAG 1920

CTCGAGGTTC TGAACTTGTA CTACAGTTAC GCCGGTTGGG AACTGCAGAG CTTTGGAGAA 1980

GATGAAGCAG AAGAACTCGG ATTCGCTGAC TTGGAATACT TGGAAAACCT AACCACACTC 2040

GGTATCACTG TTCTCTCATT GGAGACCCTA AAAACTCTCT TCGAGTTCGG TGCTTTGCAT 2100

AAACATATAC AGCATCTCCA CGTTGAAGAG TGCAATGAAC TCCTCTACTT CAATCTCCCA 2160

TCACTCACTA ACCATGGCAG GAACCTGAGA AGACTTAGCA TTAAAAGTTG CCATGACTTG 2220

GAGTACCTGG TCACACCCGC AGATTTTGAA AATGATTGGC TTCCGAGTCT AGAGGTTCTG 2280

ACGTTACACA GCCTTCACAA CTTAACCAGA GTGTGGGGAA ATTCTGTAAG CCAAGATTGT 2340

CTGCGGAATA TCCGTTGCAT AAACATTTCA CACTGCAACA AGCTGAAGAA TGTCTCATGG 2400

GTTCAGAAAC TCCCAAAGCT AGAGGTGATT GAACTGTTCG ACTGCAGAGA GATAGAGGAA 2460

TTGATAAGCG AACACGAGAG TCCATCCGTC GAAGATCCAA CATTGTTCCC AAGCCTGAAG 2520

ACCTTGAGAA CTAGGGATCT GCCAGAACTA AACAGCATCC TCCCATCTCG ATTTTCATTC 2580

CAAAAAGTTG AAACATTAGT CATCACAAAT TGCCCCAGAG TTAAGAAACT GCCGTTTCAG 2640

GAGAGGAGGA CCCAGATGAA CTTGCCAACA GTTTATTGTG AGGAGAAATG GTGGAAAGCA 2700

CTGGAAAAAG ATCAACCAAA CGAAGAGCTT TGTTATTTAC CGCGCTTTGT TCCAAATTGA 2760

TATAAGAGCT AAGAGCACTC TGTACAAATA TGTCCATTCA TAAGTAGCAG GAAGCCAGGA 2820

AGGTTGTTCC AGTGAAGTCA TCAACTTTCC ACATAGCCAC AAAACTAGAG ATTATGTAAT 2880

CATAAAAACC AAACTATCCG CGA 2903

(2) SEQ ID NO:2 的资料:

(i) 序列特征:

(A) 长度: 885 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:2:

Lys Lys Glu Arg Glu Ile Ile Glu Met Asp Phe Ile Ser Ser Leu Ile
1 5 10 15

Val Gly Cys Ala Gln Val Leu Cys Glu Ser Met Asn Met Ala Glu Arg
20 25 30

Arg Gly His Lys Thr Asp Leu Arg Gln Ala Ile Thr Asp Leu Arg Ile
35 40 45

Gln Gln Asp Gly Leu Glu Gly Arg Ser Cys Ser Asn Arg Ala Arg Glu
50 55 60

Trp Leu Ser Ala Val Gln Val Thr Glu Thr Lys Thr Ala Leu Leu Leu
65 70 75 80

Val Arg Phe Arg Arg Arg Glu Gln Arg Thr Arg Met Arg Arg Arg Tyr
85 90 95

Leu Ser Cys Phe Gly Cys Ala Asp Tyr Lys Leu Cys Lys Lys Val Ser
100 105 110

Ala Ile Leu Lys Ser Ile Gly Glu Leu Arg Glu Arg Ser Glu Ala Ile
115 120 125

Lys Thr Asp Gly Gly Ser Ile Gln Val Thr Cys Arg Glu Ile Pro Ile
130 135 140

Lys Ser Val Val Gly Asn Thr Thr Met Met Glu Gln Val Leu Glu Phe
145 150 155 160

Leu Ser Glu Glu Glu Glu Arg Gly Ile Ile Gly Val Tyr Gly Pro Gly
165 170 175

Gly Val Gly Lys Thr Thr Leu Met Gln Ser Ile Asn Asn Glu Leu Ile
180 185 190

Thr Lys Gly His Gln Tyr Asp Val Leu Ile Trp Val Gln Met Ser Arg
195 200 205

Glu Phe Gly Glu Cys Thr Ile Gln Gln Ala Val Gly Ala Arg Leu Gly
210 215 220

Leu Ser Trp Asp Glu Lys Glu Thr Gly Glu Asn Arg Ala Leu Lys Ile
225 230 235 240

Tyr Arg Ala Leu Arg Gln Lys Arg Phe Leu Leu Leu Leu Asp Asp Val
245 250 255

Trp Glu Glu Ile Asp Leu Glu Lys Thr Gly Val Pro Arg Pro Asp Arg
260 265 270

Glu Asn Lys Cys Lys Val Met Phe Thr Thr Arg Ser Ile Ala Leu Cys
275 280 285

Asn Asn Met Gly Ala Glu Tyr Lys Leu Arg Val Glu Phe Leu Glu Lys
290 295 300

Lys His Ala Trp Glu Leu Phe Cys Ser Lys Val Trp Arg Lys Asp Leu
305 310 315 320

Leu Glu Ser Ser Ser Ile Arg Arg Leu Ala Glu Ile Ile Val Ser Lys
325 330 335

Cys Gly Gly Leu Pro Leu Ala Leu Ile Thr Leu Gly Gly Ala Met Ala
340 345 350

His Arg Glu Thr Glu Glu Glu Trp Ile His Ala Ser Glu Val Leu Thr
355 360 365

Arg Phe Pro Ala Glu Met Lys Gly Met Asn Tyr Val Phe Ala Leu Leu

370 375 380

Lys Phe Ser Tyr Asp Asn Leu Glu Ser Asp Leu Leu Arg Ser Cys Phe
385 390 395 400

Leu Tyr Cys Ala Leu Phe Pro Glu Glu His Ser Ile Glu Ile Glu Gln
405 410 415

Leu Val Glu Tyr Trp Val Gly Glu Gly Phe Leu Thr Ser Ser His Gly
420 425 430

Val Asn Thr Ile Tyr Lys Gly Tyr Phe Leu Ile Gly Asp Leu Lys Ala
435 440 445

Ala Cys Leu Leu Glu Thr Gly Asp Glu Lys Thr Gln Val Lys Met His
450 455 460

Asn Val Val Arg Ser Phe Ala Leu Trp Met Ala Ser Glu Gln Gly Thr
465 470 475 480

Tyr Lys Glu Leu Ile Leu Val Glu Pro Ser Met Gly His Thr Glu Ala
485 490 495

Pro Lys Ala Glu Asn Trp Arg Gln Ala Leu Val Ile Ser Leu Leu Asp
500 505 510

Asn Arg Ile Gln Thr Leu Pro Glu Lys Leu Ile Cys Pro Lys Leu Thr
515 520 525

Thr Leu Met Leu Gln Gln Asn Ser Ser Leu Lys Lys Ile Pro Thr Gly
530 535 540

Phe Phe Met His Met Pro Val Leu Arg Val Leu Asp Leu Ser Phe Thr
545 550 555 560

Ser Ile Thr Glu Ile Pro Leu Ser Ile Lys Tyr Leu Val Glu Leu Tyr
565 570 575

His Leu Ser Met Ser Gly Thr Lys Ile Ser Val Leu Pro Gln Glu Leu
580 585 590

Gly Asn Leu Arg Lys Leu Lys His Leu Asp Leu Gln Arg Thr Gln Phe
595 600 605

Leu Gln Thr Ile Pro Arg Asp Ala Ile Cys Trp Leu Ser Lys Leu Glu
610 615 620

Val Leu Asn Leu Tyr Tyr Ser Tyr Ala Gly Trp Glu Leu Gln Ser Phe
625 630 635 640

Gly Glu Asp Glu Ala Glu Glu Leu Gly Phe Ala Asp Leu Glu Tyr Leu
645 650 655

Glu Asn Leu Thr Thr Leu Gly Ile Thr Val Leu Ser Leu Glu Thr Leu
660 665 670

Lys Thr Leu Phe Glu Phe Gly Ala Leu His Lys His Ile Gln His Leu
675 680 685

His Val Glu Glu Cys Asn Glu Leu Leu Tyr Phe Asn Leu Pro Ser Leu
690 695 700

Thr Asn His Gly Arg Asn Leu Arg Arg Leu Ser Ile Lys Ser Cys His
705 710 715 720

Asp Leu Glu Tyr Leu Val Thr Pro Ala Asp Phe Glu Asn Asp Trp Leu
725 730 735

Pro Ser Leu Glu Val Leu Thr Leu His Ser Leu His Asn Leu Arg Cys
740 745 750

Ile Asn Ile Ser His Cys Asn Lys Leu Lys Asn Val Ser Trp Val Gln
755 760 765

Lys Leu Pro Lys Leu Glu Val Ile Glu Leu Phe Asp Cys Arg Glu Ile
770 775 780

Glu Glu Leu Ile Ser Glu His Glu Ser Pro Ser Val Glu Asp Pro Thr
785 790 795 800

Leu Phe Pro Ser Leu Lys Thr Leu Arg Thr Arg Asp Leu Pro Glu Leu
805 810 815

Asn Ser Ile Leu Pro Ser Arg Phe Ser Phe Gln Lys Val Glu Thr Leu
820 825 830

Val Ile Thr Asn Cys Pro Arg Val Lys Lys Leu Pro Phe Gln Glu Arg
835 840 845

Arg Thr Gln Met Asn Leu Pro Thr Val Tyr Cys Glu Glu Lys Trp Trp
850 855 860

Lys Ala Leu Glu Lys Asp Gln Pro Asn Glu Glu Leu Cys Tyr Leu Pro
865 870 875 880

Arg Phe Val Pro Asn

(2) SEQ ID NO:3 的资料:

(i) 序列特征:

(A) 长度: 20 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:3:

Glu His Ser Val Gln Ile Cys Pro Phe Ile Ser Ser Arg Lys Pro Gly
1 5 10 15

Arg Leu Phe Gln
20

(2) SEQ ID NO:4 的资料:

(i) 序列特征:

(A) 长度: 6 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:4:

Ser His Gln Leu Ser Thr
1 5

(2) SEQ ID NO:5 的资料:

(i) 序列特征:

(A) 长度: 11 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:5:

Arg Leu Cys Asn His Lys Asn Gln Thr Ile Arg
1 5 10

(2) SEQ ID NO:6 的资料:

(i) 序列特征:

(A) 长度: 28 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:6:

Ser Lys Arg Lys Ser Glu Lys Ser Ser Lys Trp Ile Ser Ser His Leu

1 5 10 15

Leu Ser Leu Ala Val Leu Arg Cys Cys Val Asn Leu

20 25

(2) SEQ ID NO:7 的资料:

(i) 序列特征:

(A) 长度: 25 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:7:

Ile Trp Arg Arg Glu Glu Asp Ile Arg Leu Ile Leu Asp Lys Pro Ser

1 5 10 15

Leu Ile Leu Lys Gln Pro Ser Val Thr

20 25

(2) SEQ ID NO:8 的资料:

(i) 序列特征:

(A) 长度: 6 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:8:

Arg Pro Tyr Val Met Thr
1 5

(2) SEQ ID NO:9 的资料:

(i) 序列特征:
(A) 长度: 8 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:9:

Leu Tyr Gly Ser Asn Lys Thr Val
1 5

(2) SEQ ID NO:10 的资料:

(i) 序列特征:
(A) 长度: 17 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:10:

Arg Asp Glu Ala Ala Gln Ile Val Pro Glu Ser Gly Leu Val Arg Cys
1 5 10 15

Lys

(2) SEQ ID NO:11 的资料:

(i) 序列特征:

(A) 长度: 8 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:11:

Arg Arg Leu Lys Gln Pro Tyr Phe
1 5

(2) SEQ ID NO:12 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:12:

Gly Leu Gly Val Gly Asn Arg Gly Arg Glu
1 5 10

(2) SEQ ID NO:13 的资料:

(i) 序列特征:
(A) 长度: 22 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:13:

Gly Gly Asp Thr Ser Val Val Ser Val Val Pro Thr Thr Asn Cys Ala
1 5 10 15

Arg Arg Phe Leu Pro Tyr
20

(2) SEQ ID NO:14 的资料:

(i) 序列特征:
(A) 长度: 5 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:14:

Arg Ala Leu Val Ser
1 5

(2) SEQ ID NO:15 的资料:

(i) 序列特征:
(A) 长度: 15 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:15:

Glu Asn Ala Leu Lys Leu Ser Lys Gln Met Ala Gly Gln Phe Lys
1 5 10 15

(2) SEQ ID NO:16 的资料:

(i) 序列特征:
(A) 长度: 15 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:16:

Leu Val Glu Arg Tyr Pro Ser Ser Pro Leu Ser Glu Ile Pro Arg
1 5 10 15

(2) SEQ ID NO:17 的资料:

(i) 序列特征:

(A) 长度: 29 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:17:

Trp Asn Arg Phe Trp Asn Phe Ser Val Lys Lys Lys Lys Glu Glu Ser
1 5 10 15

Leu Val Phe Met Asp Leu Val Gly Leu Gly Arg Gln Arg
20 25

(2) SEQ ID NO:18 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:18:

Cys Arg Ala Leu Thr Thr Ser
1 5

(2) SEQ ID NO:19 的资料:

(i) 序列特征:
(A) 长度: 9 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:19:

Ser Gln Lys Asp Ile Ser Met Met Tyr
1 5

(2) SEQ ID NO:20 的资料:

(i) 序列特征:
(A) 长度: 36 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:20:

Phe Gly Phe Lys Cys Pro Glu Asn Ser Ala Ser Val Gln Phe Ser Lys
1 5 10 15

Pro Leu Glu His Gly Trp Val Tyr Leu Gly Thr Arg Arg Arg Pro Ala
20 25 30

Lys Thr Glu Leu
35

(2) SEQ ID NO:21 的资料:

(i) 序列特征:
(A) 长度: 5 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:21:

Arg Tyr Thr Glu Leu
1 5

(2) SEQ ID NO:22 的资料:

(i) 序列特征:
(A) 长度: 8 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:22:

Asp Arg Asn Val Ser Cys Cys Cys
1 5

(2) SEQ ID NO:23 的资料:

(i) 序列特征:

(A) 长度: 6 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:23:

Met Met Ser Gly Lys Arg
1 5

(2) SEQ ID NO:24 的资料:

(i) 序列特征:

(A) 长度: 17 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:24:

Thr Trp Arg Lys Leu Glu Phe Leu Asp Leu Thr Gly Lys Thr Asn Ala
1 5 10 15

Arg

(2) SEQ ID NO:25 的资料:

(i) 序列特征:

(A) 长度: 6 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:25:

Cys Ser Arg His Gly Leu
1 5

(2) SEQ ID NO:26 的资料:

(i) 序列特征:

(A) 长度: 11 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:26:

His Tyr Ala Thr Ile Trp Val Arg Asn Thr Ser
1 5 10

(2) SEQ ID NO:27 的资料:

(i) 序列特征:

(A) 长度: 23 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:27:

Glu Trp Ser Phe Trp Arg Arg Asn Thr Arg Gly Ser Cys Ser Val Val
1 5 10 15

Arg Tyr Gly Glu Lys Ile Phe
20

(2) SEQ ID NO:28 的资料:

(i) 序列特征:

(A) 长度: 11 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:28:

Ser His His Gln Phe Ala Gly Ser Arg Arg Leu
1 5 10

(2) SEQ ID NO:29 的资料:

(i) 序列特征:
(A) 长度: 7 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:29:

Val Asn Val Glu Asp Cys His
1 5

(2) SEQ ID NO:30 的资料:

(i) 序列特征:
(A) 长度: 19 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:30:

Glu Glu Pro Trp Leu Ile Glu Arg Gln Lys Lys Ser Gly Ser Met Leu
1 5 10 15

Val Lys Phe

(2) SEQ ID NO:31 的资料:

(i) 序列特征:
(A) 长度: 6 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:31:

Leu Asp Phe Gln Gln Arg
1 5

(2) SEQ ID NO:32 的资料:

(i) 序列特征:
(A) 长度: 6 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:32:

Thr Met Tyr Leu Pro Phe
1 5

(2) SEQ ID NO:33 的资料:

(i) 序列特征:
(A) 长度: 27 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:33:

Asn Ser Ala Thr Thr Thr Ser Arg Val Ile Cys Phe Gly Leu Val Ser
1 5 10 15

Cys Thr Ala Leu Tyr Ser Gln Lys Asn Ile Leu
20 25

(2) SEQ ID NO:34 的资料:

(i) 序列特征:
(A) 长度: 14 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:34:

Lys Arg His Val Cys Trp Lys Pro Glu Met Arg Lys His Arg
1 5 10

(2) SEQ ID NO:35 的资料:

(i) 序列特征:
(A) 长度: 33 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:35:

Arg Ser Ser Ser Leu Leu Ser Thr Gly Ser Ala Lys Gly Phe Ser Pro
1 5 10 15
Ala Pro Met Ala Leu Thr Pro Phe Thr Arg Asp Ile Phe Ser Leu Gly
20 25 30

Ile

(2) SEQ ID NO:36 的资料:

(i) 序列特征:
(A) 长度: 22 个氨基酸
(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:36:

Arg Cys Ile Met Trp Ser Glu Ala Leu His Cys Gly Trp His Leu Asn
1 5 10 15

Arg Gly Leu Ile Arg Ser
20

(2) SEQ ID NO:37 的资料:

(i) 序列特征:

(A) 长度: 20 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:37:

Leu Ser Leu Ala Trp Asp Ile Leu Lys Leu Leu Lys Gln Lys Thr Gly
1 5 10 15

Asp Lys Arg Trp
20

(2) SEQ ID NO:38 的资料:

(i) 序列特征:

(A) 长度: 15 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:38:

Ile Thr Glu Ser Arg Pro Cys Leu Lys Asn Ser Tyr Ala Arg Asn
1 5 10 15

(2) SEQ ID NO:39 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:39:

Cys Ser Asn Arg Thr Ala Leu
1 5

(2) SEQ ID NO:40 的资料:

(i) 序列特征:

(A) 长度: 46 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:40:

Arg Arg Phe Gln Gln Gly Phe Ser Cys Ile Cys Leu Phe Ser Glu Ser
1 5 10 15

Trp Thr Cys Arg Ser Gln Val Ser Leu Arg Phe Arg Cys Leu Ser Ser
20 25 30

Ile Trp Trp Ser Cys Ile Ile Cys Leu Cys Gln Glu Gln Arg
35 40 45

(2) SEQ ID NO:41 的资料:

(i) 序列特征:

(A) 长度: 12 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:41:

Val Tyr Cys His Arg Ser Leu Gly Ile Leu Glu Asn

1 5 10

(2) SEQ ID NO:42 的资料:

(i) 序列特征:

(A) 长度: 21 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:42:

Ser Ile Trp Thr Tyr Lys Glu Leu Ser Phe Phe Arg Arg Ser His Glu
1 5 10 15

Met Pro Tyr Val Gly
20

(2) SEQ ID NO:43 的资料:

(i) 序列特征:

(A) 长度: 5 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:43:

Ala Ser Ser Arg Phe
1 5

(2) SEQ ID NO:44 的资料:

(i) 序列特征:

(A) 长度: 32 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:44:

Thr Cys Thr Thr Val Thr Pro Val Gly Asn Cys Arg Ala Leu Glu Lys
1 5 10 15

Met Lys Gln Lys Asn Ser Asp Ser Leu Thr Trp Asn Thr Trp Lys Thr
20 25 30

(2) SEQ ID NO:45 的资料:

(i) 序列特征:

(A) 长度: 12 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:45:

Pro His Ser Val Ser Leu Phe Ser His Trp Arg Pro
1 5 10

(2) SEQ ID NO:46 的资料:

(i) 序列特征:
(A) 长度: 38 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:46:

Lys Leu Ser Ser Ser Ser Val Leu Cys Ile Asn Ile Tyr Ser Ile Ser
1 5 10 15

Thr Leu Lys Ser Ala Met Asn Ser Ser Thr Ser Ile Ser His His Ser
20 25 30

Leu Thr Met Ala Gly Thr
35

(2) SEQ ID NO:47 的资料:

(i) 序列特征:
(A) 长度: 27 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:47:

Glu Asp Leu Ala Leu Lys Val Ala Met Thr Trp Ser Thr Trp Ser His
1 5 10 15

Pro Gln Ile Leu Lys Met Ile Gly Phe Arg Val
20 25

(2) SEQ ID NO:48 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:48:

Arg Tyr Thr Ala Phe Thr Thr
1 5

(2) SEQ ID NO:49 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:49:

Pro Glu Cys Gly Glu Ile Leu
1 5

(2) SEQ ID NO:50 的资料:

(i) 序列特征:
(A) 长度: 6 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:50:

Phe Arg Asn Ser Gln Ser
1 5

(2) SEQ ID NO:51 的资料:

(i) 序列特征:
(A) 长度: 10 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:51:

Ala Lys Ile Val Cys Gly Ile Ser Val Ala
1 5 10

(2) SEQ ID NO:52 的资料:

(i) 序列特征:
(A) 长度: 7 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:52:

Thr Phe His Thr Ala Thr Ser
1 5

(2) SEQ ID NO:53 的资料:

(i) 序列特征:
(A) 长度: 8 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:53:

Leu Asn Cys Ser Thr Ala Glu Arg
1 5

(2) SEQ ID NO:54 的资料:

(i) 序列特征:
(A) 长度: 16 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:54:

Ala Asn Thr Arg Val His Pro Ser Lys Ile Gln His Cys Ser Gln Ala
1 5 10 15

(2) SEQ ID NO:55 的资料:

(i) 序列特征:
(A) 长度: 7 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:55:

Glu Leu Gly Ile Cys Gln Asn
1 5

(2) SEQ ID NO:56 的资料:

(i) 序列特征:
(A) 长度: 15 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:56:

Thr Ala Ser Ser His Leu Asp Phe His Ser Lys Lys Leu Lys His
1 5 10 15

(2) SEQ ID NO:57 的资料:

(i) 序列特征:
(A) 长度: 19 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:57:

Ser Ser Gln Ile Ala Pro Glu Leu Arg Asn Cys Arg Phe Arg Arg Gly
1 5 10 15

Gly Pro Arg

(2) SEQ ID NO:58 的资料:

(i) 序列特征:

(A) 长度: 47 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:58:

Thr Cys Gln Gln Phe Ile Val Arg Arg Asn Gly Gly Lys His Trp Lys
1 5 10 15

Lys Ile Asn Gln Thr Lys Ser Phe Val Ile Tyr Arg Ala Leu Phe Gln
20 25 30

Ile Asp Ile Arg Ala Lys Ser Thr Leu Tyr Lys Tyr Val His Ser
35 40 45

(2) SEQ ID NO:59 的资料:

(i) 序列特征:

(A) 长度: 33 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:59:

Val Ala Gly Ser Gln Glu Gly Cys Ser Ser Glu Val Ile Asn Phe Pro
1 5 10 15

His Ser His Lys Thr Arg Asp Tyr Val Ile Ile Lys Thr Lys Leu Ser
20 25 30

Ala

(2) SEQ ID NO:60 的资料:

(i) 序列特征:
(A) 长度: 25 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:60:

Val Lys Glu Arg Ala Arg Asn His Arg Asn Gly Phe His Leu Ile Ser
1 5 10 15

Tyr Arg Trp Leu Cys Ser Gly Val Val

(2) SEQ ID NO:61 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:61:

Ile Tyr Glu Tyr Gly Gly Glu Lys Arg Thr
1 5 10

(2) SEQ ID NO:62 的资料:

(i) 序列特征:

(A) 长度: 5 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:62:

Leu Glu Gly His Thr
1 5

(2) SEQ ID NO:63 的资料:

(i) 序列特征:

(A) 长度: 23 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:63:

Pro Asp Phe Thr Asp Pro Thr Arg Arg Ser Arg Gly Thr Lys Leu Leu
1 5 10 15

Lys Ser Cys Gln Arg Val Ala
20

(2) SEQ ID NO:64 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:64:

Cys Gly Ala Ser Asn Gly Asp
1 5

(2) SEQ ID NO:65 的资料:

(i) 序列特征:

(A) 长度: 8 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:65:

Asn Ser Pro Thr Phe Ser Glu Val
1 5

(2) SEQ ID NO:66 的资料:

(i) 序列特征:

(A) 长度: 35 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:66:

Ala Ser Gly Thr Glu Asp Ala Asn Glu Glu Glu Ile Pro Gln Leu Phe
1 5 10 15

Arg Leu Cys Arg Leu Gln Thr Val Gln Glu Gly Phe Cys His Ile Glu

Glu His Trp
35

(2) SEQ ID NO:67 的资料:

(i) 序列特征:
(A) 长度: 5 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:67:

Ala Glu Arg Thr Leu
1 5

(2) SEQ ID NO:68 的资料:

(i) 序列特征:
(A) 长度: 13 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:68:

Ser Tyr Gln Asn Arg Trp Arg Val Asn Ser Ser Asn Leu
1 5 10

(2) SEQ ID NO:69 的资料:

(i) 序列特征:

(A) 长度: 22 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:69:

Arg Asp Thr His Gln Val Arg Cys Arg Lys Tyr His Asp Asp Gly Thr
1 5 10 15

Gly Phe Gly Ile Ser Gln
20

(2) SEQ ID NO:70 的资料:

(i) 序列特征:

(A) 长度: 24 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:70:

Arg Arg Arg Lys Arg Asn His Trp Cys Leu Trp Thr Trp Trp Gly Trp
1 5 10 15

Glu Asp Asn Val Asn Ala Glu His
20

(2) SEQ ID NO:71 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:71:

Gln Arg Ala Asp His Lys Arg Thr Ser Val
1 5 10

(2) SEQ ID NO:72 的资料:

(i) 序列特征:

(A) 长度: 55 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:72:

Cys Thr Asp Leu Gly Ser Asn Val Gln Arg Ile Arg Arg Val Tyr Asn
1 5 10 15

Ser Ala Ser Arg Trp Ser Thr Val Gly Phe Ile Leu Gly Arg Glu Gly
20 25 30

Asp Arg Arg Lys Gln Ser Phe Glu Asp Ile Gln Ser Phe Glu Thr Glu
35 40 45

Thr Phe Leu Val Val Ala Arg
50 55

(2) SEQ ID NO:73 的资料:

(i) 序列特征:

(A) 长度: 15 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:73:

Cys Leu Gly Arg Asp Arg Leu Gly Glu Asn Trp Ser Ser Ser Thr
1 5 10 15

(2) SEQ ID NO:74 的资料:

(i) 序列特征:

(A) 长度: 9 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:74:

Arg Asp Arg Arg Arg Val Asp Pro Cys
1 5

(2) SEQ ID NO:75 的资料:

(i) 序列特征:
(A) 长度: 41 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:75:

Gln Gly Lys Gln Met Gln Gly Asp Val His Asp Thr Val Tyr Ser Ile
1 5 10 15

Met Gln Gln Tyr Gly Cys Gly Ile Gln Val Glu Ser Gly Val Ser Gly
20 25 30

Glu Glu Thr Arg Val Gly Ala Val Leu
35 40

(2) SEQ ID NO:76 的资料:

(i) 序列特征:

(A) 长度: 21 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:76:

Gly Met Glu Lys Arg Ser Phe Arg Val Ile Ile Asn Ser Pro Ala Arg
1 5 10 15

Gly Asp Tyr Ser Glu
20

(2) SEQ ID NO:77 的资料:

(i) 序列特征:

(A) 长度: 17 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:77:

Met Trp Arg Ile Ala Thr Ser Val Asp His Phe Arg Arg Ser His Gly

1 5 10 15

Ser

(2) SEQ ID NO:78 的资料:

(i) 序列特征:

(A) 长度: 24 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:78:

Ile Ser Ser Arg Asp Glu Gly Tyr Glu Leu Cys Ile Cys Pro Phe Glu
1 5 10 15

Ile Gln Leu Arg Gln Pro Arg Glu
20

(2) SEQ ID NO:79 的资料:

(i) 序列特征:

(A) 长度: 24 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:79:

Ser Ala Ser Val Leu Phe Leu Val Leu Arg Phe Ile Pro Arg Arg Thr
1 5 10 15

Phe Tyr Arg Asp Arg Ala Ala Cys
20

(2) SEQ ID NO:80 的资料:

(i) 序列特征:
(A) 长度: 14 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:80:

Val Leu Gly Arg Arg Arg Val Ser His Gln Leu Pro Trp Arg
1 5 10

(2) SEQ ID NO:81 的资料:

(i) 序列特征:
(A) 长度: 22 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:81:

His His Leu Gln Gly Ile Phe Ser His Trp Gly Ser Glu Ser Gly Met
1 5 10 15

Phe Val Gly Asn Arg Arg
20

(2) SEQ ID NO:82 的资料:

(i) 序列特征:
(A) 长度: 7 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:82:

Glu Asn Thr Gly Glu Asp Ala
1 5

(2) SEQ ID NO:83 的资料:

(i) 序列特征:
(A) 长度: 43 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:83:

Lys Thr His Met Pro Glu Thr Asp Asn Thr Asp Ala Pro Thr Glu Gly
1 5 10 15

Leu Phe Glu Glu Asp Ser Asn Arg Val Phe His Ala Tyr Ala Cys Ser
20 25 30

Gln Ser Leu Gly Leu Val Val His Lys Tyr His
35 40

(2) SEQ ID NO:84 的资料:

(i) 序列特征:

(A) 长度: 11 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:84:

Cys Gly Gln Lys Leu Cys Ile Val Asp Gly Ile
1 5 10

(2) SEQ ID NO:85 的资料:

(i) 序列特征:

(A) 长度: 5 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:85:

Gly Ala Asp Pro Ser
1 5

(2) SEQ ID NO:86 的资料:

(i) 序列特征:
(A) 长度: 14 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:86:

Ser Arg Lys Leu Ala Thr Ser Val Gly Asp Leu Ile Val Arg
1 5 10

(2) SEQ ID NO:87 的资料:

(i) 序列特征:
(A) 长度: 6 个氨基酸
(B) 类型: 氨基酸

(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:87:

Gln Asn Pro Asp Leu Ala
1 5

(2) SEQ ID NO:88 的资料:

(i) 序列特征:
(A) 长度: 31 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:88:

Asp Ser Val Val Tyr Gln Val Phe Gly Gly Val Val Ser Ser Val Tyr
1 5 10 15

Val Arg Asn Lys Asp Lys Cys Ile Ala Thr Gly Ala Trp Glu Ser
20 25 30

(2) SEQ ID NO:89 的资料:

(i) 序列特征:
(A) 长度: 47 个氨基酸

(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:89:

Lys Thr Glu Ala Ser Gly Pro Thr Lys Asn Ser Val Ser Ser Asp Asp
1 5 10 15

Pro Thr Arg Cys His Met Leu Ala Glu Gln Ala Arg Gly Ser Glu Leu
20 25 30

Val Leu Gln Leu Arg Arg Leu Gly Thr Ala Glu Leu Trp Arg Arg
35 40 45

(2) SEQ ID NO:90 的资料:

(i) 序列特征:
(A) 长度: 7 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:90:

Ser Arg Arg Thr Arg Ile Arg
1 5

(2) SEQ ID NO:91 的资料:

(i) 序列特征:

(A) 长度: 30 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:91:

Leu Gly Ile Leu Gly Lys Pro Asn His Thr Arg Tyr His Cys Ser Leu
1 5 10 15

Ile Gly Asp Pro Lys Asn Ser Leu Arg Val Arg Cys Phe Ala
20 25 30

(2) SEQ ID NO:92 的资料:

(i) 序列特征:

(A) 长度: 7 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:92:

Thr Tyr Thr Ala Ser Pro Arg
1 5

(2) SEQ ID NO:93 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:93:

Thr Pro Leu Leu Gln Ser Pro Ile Thr His
1 5 10

(2) SEQ ID NO:94 的资料:

(i) 序列特征:

(A) 长度: 8 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:94:

Pro Trp Gln Glu Pro Glu Lys Thr
1 5

(2) SEQ ID NO:95 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:95:

Leu Gly Val Pro Gly His Thr Arg Arg Phe
1 5 10

(2) SEQ ID NO:96 的资料:

(i) 序列特征:

(A) 长度: 58 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:96:

Leu Ala Ser Glu Ser Arg Gly Ser Asp Val Thr Gln Pro Ser Gln Leu
1 5 10 15

Asn Gln Ser Val Gly Lys Phe Cys Lys Pro Arg Leu Ser Ala Glu Tyr
20 25 30

Pro Leu His Lys His Phe Thr Leu Gln Gln Ala Glu Glu Cys Leu Met
35 40 45

Gly Ser Glu Thr Pro Lys Ala Arg Gly Asp
50 55

(2) SEQ ID NO:97 的资料:

(i) 序列特征:

(A) 长度: 33 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:97:

Thr Val Arg Leu Gln Arg Asp Arg Gly Ile Asp Lys Arg Thr Arg Glu
1 5 10 15

Ser Ile Arg Arg Arg Ser Asn Ile Val Pro Lys Pro Glu Asp Leu Glu
20 25 30

Asn

(2) SEQ ID NO:98 的资料:

(i) 序列特征:

(A) 长度: 18 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:98:

Gly Ser Ala Arg Thr Lys Gln His Pro Pro Ile Ser Ile Phe Ile Pro
1 5 10 15

Lys Ser

(2) SEQ ID NO:99 的资料:

(i) 序列特征:

(A) 长度: 10 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:99:

Asn Ile Ser His His Lys Leu Pro Gln Ser
1 5 10

(2) SEQ ID NO:100 的资料:

(i) 序列特征:

(A) 长度: 18 个氨基酸

(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:100:

Glu Thr Ala Val Ser Gly Glu Glu Asp Pro Asp Glu Leu Ala Asn Ser
1 5 10 15

Leu Leu

(2) SEQ ID NO:101 的资料:

(i) 序列特征:
(A) 长度: 20 个碱基对
(B) 类型: 核酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:101:

GGTAGTGAGT AGAGAATAAC 20

(2) SEQ ID NO:102 的资料:

(i) 序列特征:
(A) 长度: 13 个氨基酸

(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:102:

Glu Leu Arg Ala Leu Cys Thr Asn Met Ser Ile His Lys
1 5 10

(2) SEQ ID NO:103 的资料:

(i) 序列特征:
(A) 长度: 23 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:103:

Gln Glu Ala Arg Lys Val Val Pro Val Lys Ser Ser Thr Phe His Ile
1 5 10 15

Ala Thr Lys Leu Glu Ile Met
20

(2) SEQ ID NO:104 的资料:

(i) 序列特征:

(A) 长度: 6 个氨基酸
(B) 类型: 氨基酸
(C) 链型: 不相关
(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:104:

Lys Pro Asn Tyr Pro Arg
1 5

(2) SEQ ID NO:105 的资料:

(i) 序列特征:
(A) 长度: 1491 个碱基对
(B) 类型: 核酸
(C) 链型: 单链
(D) 拓扑类型: 线性

(ii) 分子类型: DNA (基因组)

(xi) 序列描述: SEQ ID NO:105:

ATCGATTGAT CTCTGGCTCA GTGCGAGTAG TCCATTTGAG AGCAGTCGTA GCCCCGCGTG 60

GCGCATCATG GAGCTATTTG GAATTTTCGC AGGGTTATCG ATTCGTAGTG GGAACCCATT 120

CATTGTTTGG AACCACCAAC GGACGACTTA ACAAGCTCCC CGAGGTGCAT GATGAAAATT 180

GCTCCAGTTG CCATAAATCA CAGCCCGCTC AGCAGGGAGG TCCCGTCACA CGCGGCACCC 240

ACTCAGGCAA AGCAAACCAA CCTTCAATCT GAAGCTGGCG ATTTAGATGC AAGAAAAAGT 300

AGCGCTTCAA GCCCGGAAAC CCGCGCATTA CTCGCTACTA AGACAGTACT CGGGAGACAC 360

AAGATAGAGG TTCCGGCCTT TGGAGGGTGG TTCAAAAAGA AATCATCTAA GCACGAGACG 420

GGCGGTTCAA GTGCCAACGC AGATAGTTCG AGCGTGGCTT CCGATTCCAC CGAAAAACCT 480

TTGTTCCGTC TCACGCACGT TCCTTACGTA TCCCAAGGTA ATGAGCGAAT GGGATGTTGG 540

TATGCCTGCG CAAGAATGGT TGGCCATTCT GTCGAAGCTG GGCCTCGCCT AGGGCTGCCG 600

GAGCTCTATG AGGGAAGGGA GGCGCCAGCT GGGCTACAAG ATTTTTCAGA TGTAGAAAGG 660

TTTATTCACA ATGAAGGATT AACTCGGGTA GACCTTCCAG ACAATGAGAG ATTTACACAC 720

GAAGAGTTGG GTGCACTGTT GTATAAGCAC GGGCCGATTA TATTTGGGTG GAAAACTCCG 780

AATGACAGCT GGCACATGTC GGTCCTCACT GGTGTCGATA AAGAGACGTC GTCCATTACT 840

TTTCACGATC CCCGACAGGG GCCGGACCTA GCAATGCCGC TCGATTACTT TAATCAGCGA 900

TTGGCATGGC AGGTTCCACA CGCAATGCTC TACCGCTAAG TAGCAGGGTA TCTTCACGTG 960

GCGGCATCAT GACAAGCCCA TGATGCCGCC AGCAGCTACC TGAATGCCGT CTGGCTTTTT 1020

GGTCCCTATT GTCGTATCCG GAAGATGACG TCAAAGAATC TCGGCAAGAG CTTTCTTGCT 1080

CGACTCCTCA GCTTCCGGAT CGATCAGGTC GCTTGCCAGA GCGCGCTTGT CCATGAGCAT 1140

CTGCCACAGC TGCTGGTCGA TGGTGTCCTC AGCTAAAGGG ATTTTGACGA CAACCATGCG 1200

CAACTGCCCG TTGCGATACG CTCGATCCTG AAGCCCCGGT GTCCATGGCA GCCCCAAGAA 1260

AAAGACATAG TTCGCCGCTG TGAGGTTGTA GCCTGTGCCG GCGGCCGACC TGGTCCCGAT 1320

AAACACCCTG CAGTCCGGAT CCTGCTGGAA AGCATCAATC GCCTTCTGCC GCTTCTTGGG 1380

CGAGTCACTG CCCACCAACG TCACGCACCC GACGCCAAGC TTGAGGCAGT GCTCCCGCAA 1440

CGTGGCCACG GATTCCTGAT ACTCGCAGAA GAGGATCACC TTGTCGTCGA C 1491

(2) SEQ ID NO:106 的资料:

(i) 序列特征:

(A) 长度: 255 个氨基酸

(B) 类型: 氨基酸

(C) 链型: 不相关

(D) 拓扑类型: 线性

(ii) 分子类型: 蛋白质

(xi) 序列描述: SEQ ID NO:106:

Met Lys Ile Ala Pro Val Ala Ile Asn His Ser Pro Leu Ser Arg Glu
1 5 10 15

Val Pro Ser His Ala Ala Pro Thr Gln Ala Lys Gln Thr Asn Leu Gln
20 25 30

Ser Glu Ala Gly Asp Leu Asp Ala Arg Lys Ser Ser Ala Ser Ser Pro
35 40 45

Glu Thr Arg Ala Leu Leu Ala Thr Lys Thr Val Leu Gly Arg His Lys

50 55 60

Ile Glu Val Pro Ala Phe Gly Gly Trp Phe Lys Lys Lys Ser Ser Lys
65 70 75 80

His Glu Thr Gly Gly Ser Ser Ala Asn Ala Asp Ser Ser Ser Val Ala
85 90 95

Ser Asp Ser Thr Glu Lys Pro Leu Phe Arg Leu Thr His Val Pro Tyr
100 105 110

Val Ser Gln Gly Asn Glu Arg Met Gly Cys Trp Tyr Ala Cys Ala Arg
115 120 125

Met Val Gly His Ser Val Glu Ala Gly Pro Arg Leu Gly Leu Pro Glu
130 135 140

Leu Tyr Glu Gly Arg Glu Ala Pro Ala Gly Leu Gln Asp Phe Ser Asp
145 150 155 160

Val Glu Arg Phe Ile His Asn Glu Gly Leu Thr Arg Val Asp Leu Pro
165 170 175

Asp Asn Glu Arg Phe Thr His Glu Glu Leu Gly Ala Leu Leu Tyr Lys
180 185 190

His Gly Pro Ile Ile Phe Gly Trp Lys Thr Pro Asn Asp Ser Trp His
195 200 205

Met Ser Val Leu Thr Gly Val Asp Lys Glu Thr Ser Ser Ile Thr Phe
210 215 220

His Asp Pro Arg Gln Gly Pro Asp Leu Ala Met Pro Leu Asp Tyr Phe
225 230 235 240

Asn Gln Arg Leu Ala Trp Gln Val Pro His Ala Met Leu Tyr Arg
245 250 255

说 明 书 附 图

图 1A

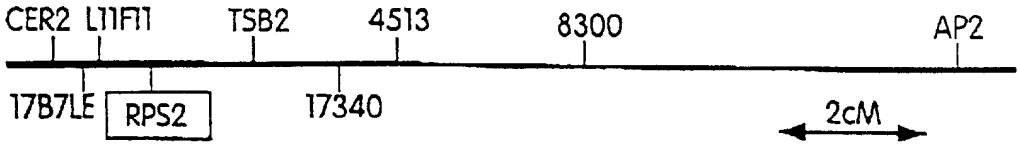


图 1B

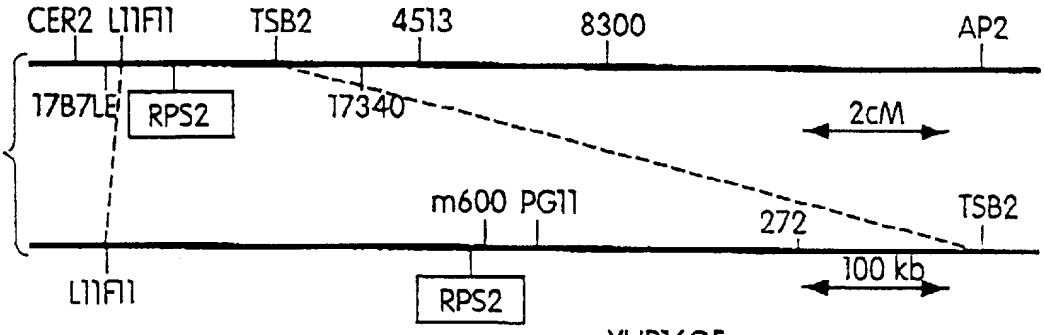


图 1C

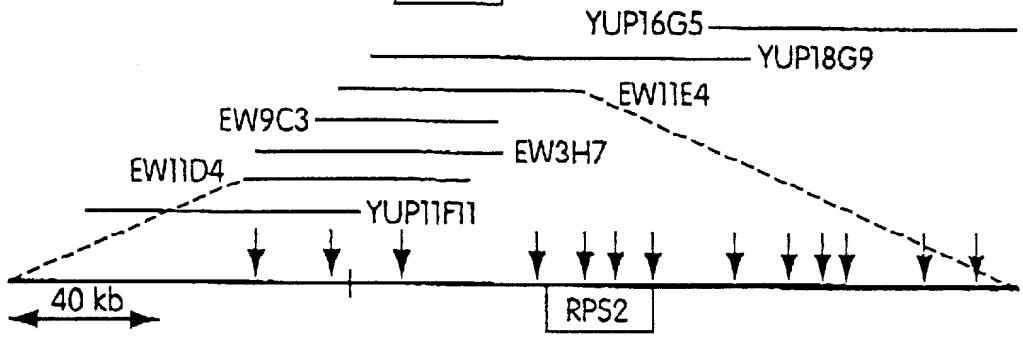


图 1D

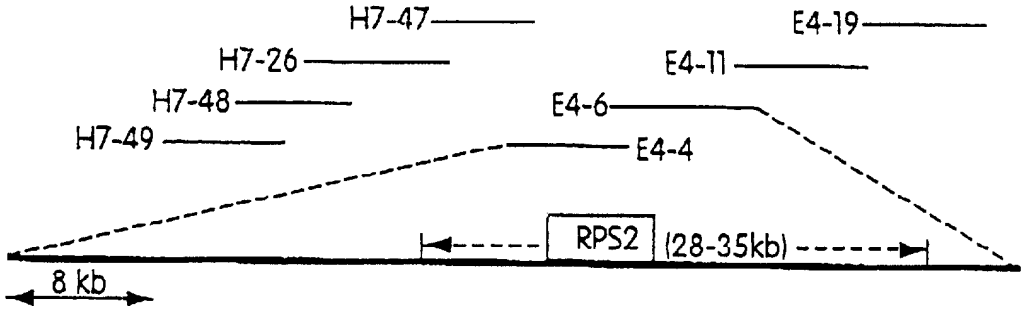


图 1E

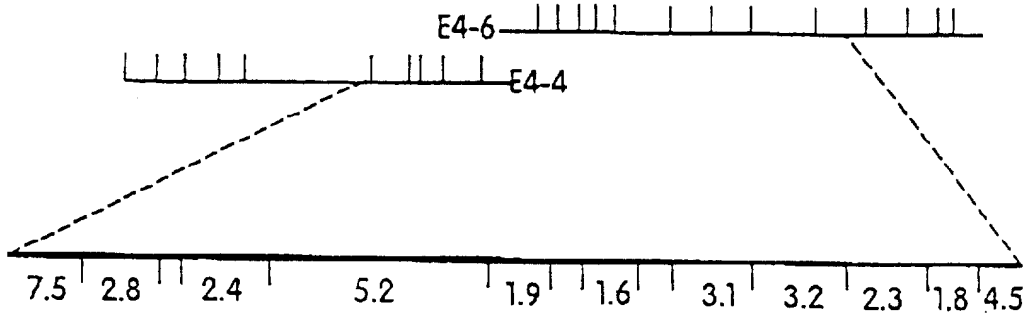
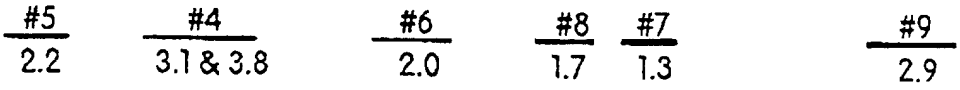


图 1F



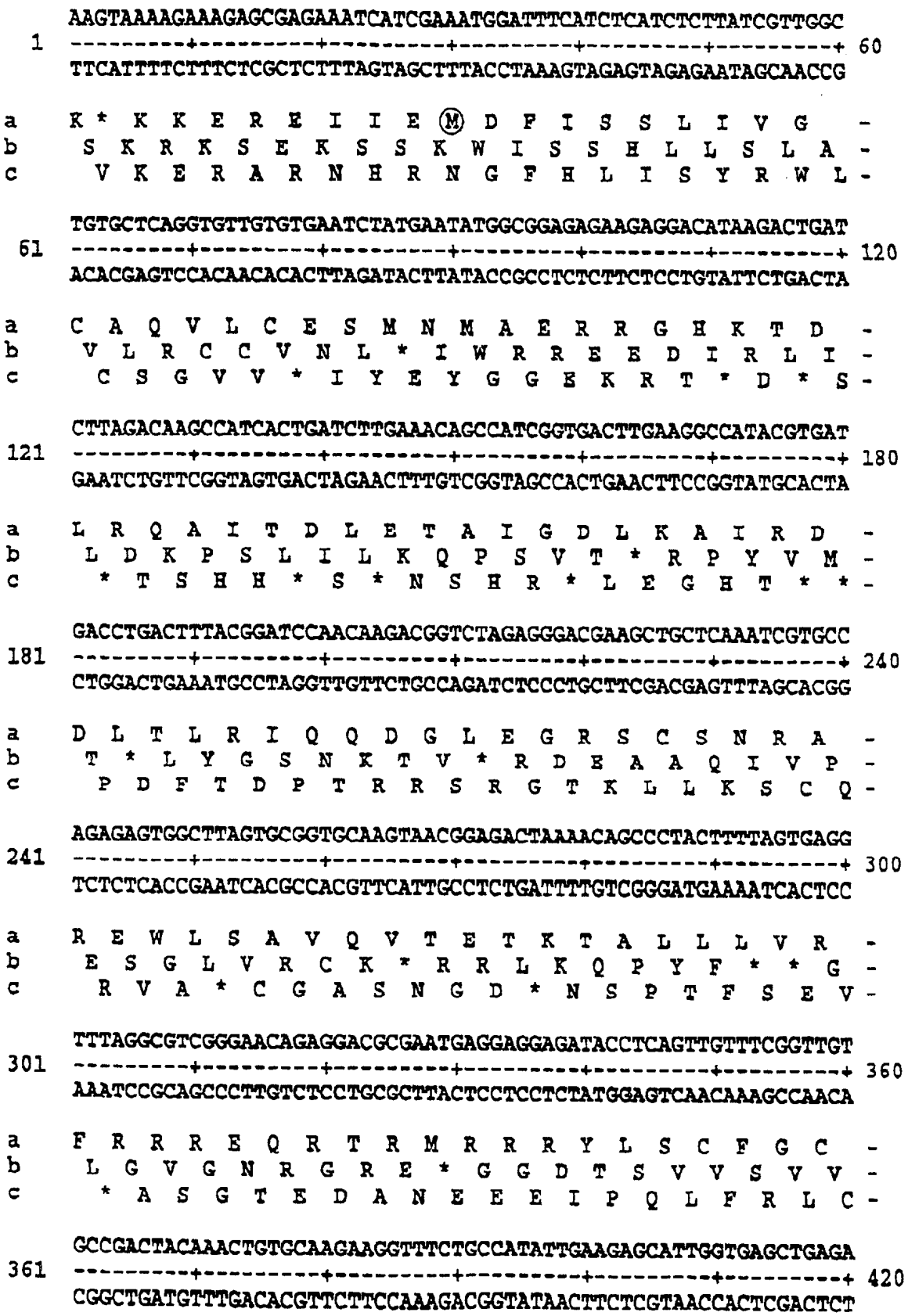

AAGTAAAAGAAAGAGCGAGAAATCATCGAAATGGATTTCATCTCATCTCTTATCGTTGGC
TTCATTTTCTTTCTCGCTCTTTAGTAGCTTTACCTAAAGTAGAGTAGAGAATAGCAACCG
a K * K K E R E I I E M D F I S S L I V G -
b S K R K S E K S S K W I S S H L L S L A -
c V K E R A R N H R N G F H L I S Y R W L -
TGTGCTCAGGTGTTGTGTGAATCTATGAATATGGCGGAGAGAAGAGGACATAAGACTGAT
ACACGAGTCCACAACACACTTAGATACTTATACCGCCTCTCTTCTCCTGTATTCTGACTA
a C A Q V L C E S M N M A E R R G H K T D -
b V L R C C V N L * I W R R E E D I R L I -
c C S G V V * I Y E Y G G E K R T * D * S -
CTTAGACAAGCCATCACTGATCTTGAAACAGCCATCGGTGACTTGAAGGCCATACGTGAT
GAATCTGTTCGGTAGTGACTAGAACTTTGTCGGTAGCCACTGAACTTCCGGTATGCACTA
a L R Q A I T D L E T A I G D L K A I R D -
b L D K P S L I L K Q P S V T * R P Y V M -
c * T S H H * S * N S H R * L E G H T * * -
GACCTGACTTTACGGATCCAACAAGACGGTCTAGAGGGACGAAGCTGCTCAAATCGTGCC
CTGGACTGAAATGCCTAGGTTGTTCTGCCAGATCTCCCTGCTTCGACGAGTTTAGCACGG
a D L T L R I Q Q D G L E G R S C S N R A -
b T * L Y G S N K T V * R D E A A Q I V P -
c P D F T D P T R R S R G T K L L K S C Q -
AGAGAGTGGCTTAGTGCGGTGCAAGTAACGGAGACTAAAACAGCCCTACTTTTAGTGAGG
TCTCTCACCGAATCACGCCACGTTCATTGCCTCTGATTTTGTCGGGATGAAAATCACTCC
a R E W L S A V Q V T E T K T A L L L V R -
b E S G L V R C K * R R L K Q P Y F * * G -
c R V A * C G A S N G D * N S P T F S E V -
TTTAGGCGTCGGGAACAGAGGACGCGAATGAGGAGGAGATACCTCAGTTGTTTCGGTTGT
AAATCCGCAGCCCTTGTCTCCTGCGCTTACTCCTCCTCTATGGAGTCAACAAAGCCAACA
a F R R R E Q R T R M R R R Y L S C F G C -
b L G V G N R G R E * G G D T S V V S V V -
c * A S G T E D A N E E E I P Q L F R L C -
GCCGACTACAAACTGTGCAAGAAGGTTTCTGCCATATTGAAGAGCATTGGTGAGCTGAGA
CGGCTGATGTTTGACACGTTCTTCCAAAGACGGTATAACTTCTCGTAACCACTCGACTCT
1
61
121
181
241
301
361
60
120
180
240
300
360
420

图 2A

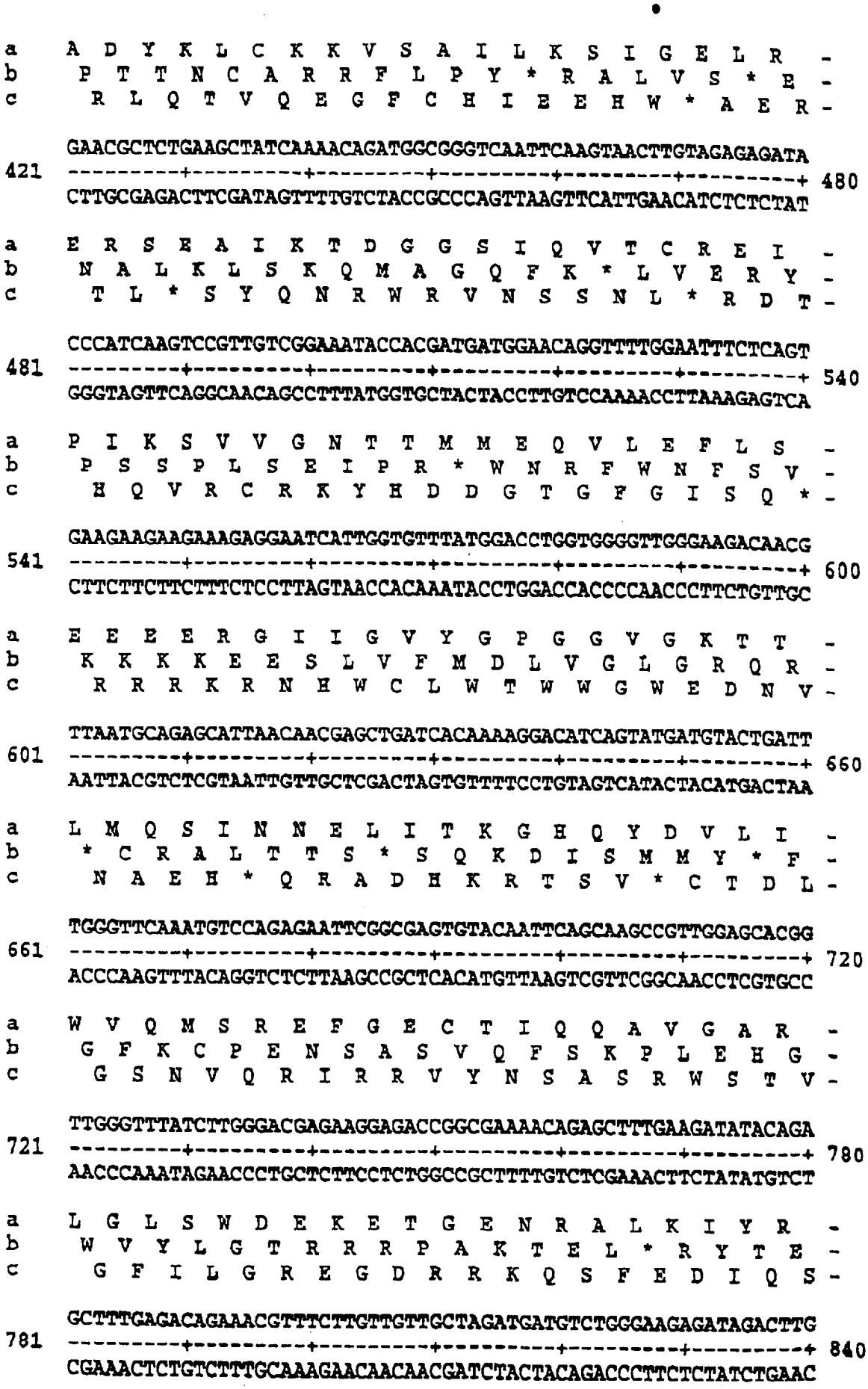

a A D Y K L C K K V S A I L K S I G E L R -
b P T T N C A R R F L P Y * R A L V S * E -
c R L Q T V Q E G F C H I E E H W * A E R -
GAACGCTCTGAAGCTATCAAAACAGATGGCGGGTCAATTCAAGTAACTTGTAGAGAGATA
421
CTTGCGAGACTTCGATAGTTTTGTCTACCGCCCAGTTAAGTTCATTGAACATCTCTCTAT
480
a E R S E A I K T D G G S I Q V T C R E I -
b N A L K L S K Q M A G Q F K * L V E R Y -
c T L * S Y Q N R W R V N S S N L * R D T -
CCCATCAAGTCCGTTGTCGGAAATACCACGATGATGGAACAGGTTTTGGAATTTCTCAGT
481
GGGTAGTTCAGGCAACAGCCTTTATGGTGCTACTACCTTGTCCAAAACCTTAAAGAGTCA
540
a P I K S V V G N T T M M E Q V L E F L S -
b P S S P L S E I P R * W N R F W N F S V -
c H Q V R C R K Y H D D G T G F G I S Q * -
GAAGAAGAAGAAAGAGGAATCATTGGTGTTTATGGACCTGGTGGGGTTGGGAAGACAACG
541
CTTCTTCTTCTTTCTCCTTAGTAACCACAAATACCTGGACCACCCCAACCCTTCTGTTGC
600
a E E E E R G I I G V Y G P G G V G K T T -
b K K K K E E S L V F M D L V G L G R Q R -
c R R R K R N H W C L W T W W G W E D N V -
TTAATGCAGAGCATTAACAACGAGCTGATCACAAAAGGACATCAGTATGATGTACTGATT
601
AATTACGTCTCGTAATTGTTGCTCGACTAGTGTTTTCCTGTAGTCATACTACATGACTAA
660
a L M Q S I N N E L I T K G H Q Y D V L I -
b * C R A L T T S * S Q K D I S M M Y * F -
c N A E H * Q R A D H K R T S V * C T D L -
TGGGTTCAAATGTCCAGAGAATTCGGCGAGTGTACAATTCAGCAAGCCGTTGGAGCACGG
661
ACCCAAGTTTACAGGTCTCTTAAGCCGCTCACATGTTAAGTCGTTCGGCAACCTCGTGCC
720
a W V Q M S R E F G E C T I Q Q A V G A R -
b G F K C P E N S A S V Q F S K P L E H G -
c G S N V Q R I R R V Y N S A S R W S T V -
TTGGGTTTATCTTGGGACGAGAAGGAGACCGGCGAAAACAGAGCTTTGAAGATATACAGA
721
AACCCAAATAGAACCCTGCTCTTCCTCTGGCCGCTTTTGTCTCGAAACTTCTATATGTCT
780
a L G L S W D E K E T G E N R A L K I Y R -
b W V Y L G T R R R P A K T E L * R Y T E -
c G F I L G R E G D R R K Q S F E D I Q S -
GCTTTGAGACAGAAACGTTTCTTGTTGTTGCTAGATGATGTCTGGGAAGAGATAGACTTG
781
CGAAACTCTGTCTTTGCAAAGAACAACAACGATCTACTACAGACCCTTCTCTATCTGAAC
840

图 2B

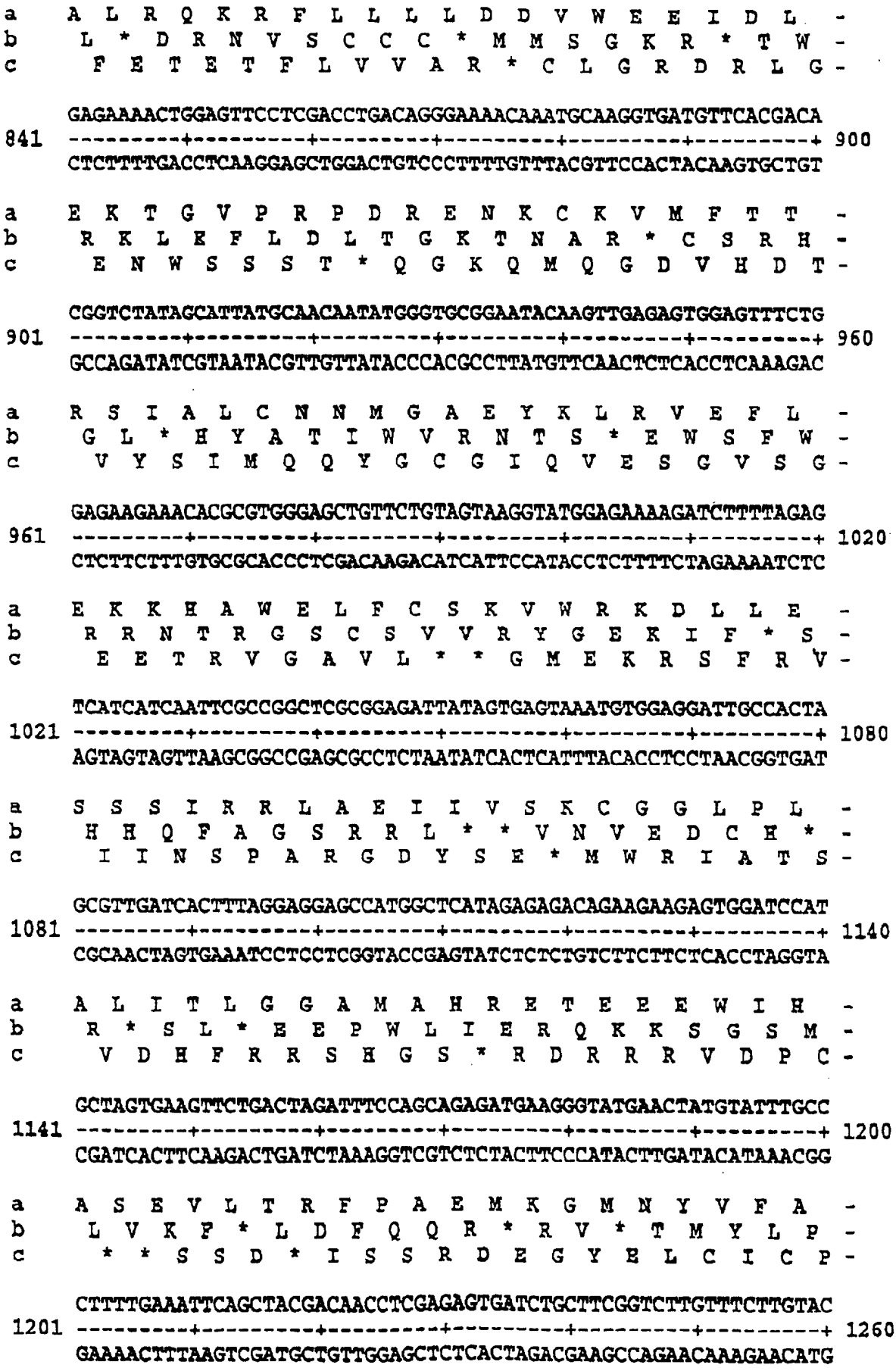

a A L R Q K R F L L L L D D V W E E I D L -
b L * D R N V S C C C * M M S G K R * T W -
c F E T E T F L V V A R * C L G R D R L G -
GAGAAAACTGGAGTTCCTCGACCTGACAGGGAAAACAAATGCAAGGTGATGTTCACGACA
841 900
CTCTTTTGACCTCAAGGAGCTGGACTGTCCCTTTTGTTTACGTTCCACTACAAGTGCTGT
a E K T G V P R P D R E N K C K V M F T T -
b R K L E F L D L T G K T N A R * C S R H -
c E N W S S S T * Q G K Q M Q G D V H D T -
CGGTCTATAGCATTATGCAACAATATGGGTGCGGAATACAAGTTGAGAGTGGAGTTTCTG
901 960
GCCAGATATCGTAATACGTTGTTATACCCACGCCTTATGTTCAACTCTCACCTCAAAGAC
a R S I A L C N N M G A E Y K L R V E F L -
b G L * H Y A T I W V R N T S * E W S F W -
c V Y S I M Q Q Y G C G I Q V E S G V S G -
GAGAAGAAACACGCGTGGGAGCTGTTCTGTAGTAAGGTATGGAGAAAAGATCTTTTAGAG
961 1020
CTCTTCTTTGTGCGCACCCTCGACAAGACATCATTCCATACCTCTTTTCTAGAAAATCTC
a E K K H A W E L F C S K V W R K D L L E -
b R R N T R G S C S V V R Y G E K I F * S -
c E E T R V G A V L * * G M E K R S F R V -
TCATCATCAATTCGCCGGCTCGCGGAGATTATAGTGAGTAAATGTGGAGGATTGCCACTA
1021 1080
AGTAGTAGTTAAGCGGCCGAGCGCCTCTAATATCACTCATTTACACCTCCTAACGGTGAT
a S S S I R R L A E I I V S K C G G L P L -
b H H Q F A G S R R L * * V N V E D C H * -
c I I N S P A R G D Y S E * M W R I A T S -
GCGTTGATCACTTTAGGAGGAGCCATGGCTCATAGAGAGACAGAAGAAGAGTGGATCCAT
1081 1140
CGCAACTAGTGAAATCCTCCTCGGTACCGAGTATCTCTCTGTCTTCTTCTCACCTAGGTA
a A L I T L G G A M A H R E T E E E W I H -
b R * S L * E E P W L I E R Q K K S G S M -
c V D H F R R S H G S * R D R R R V D P C -
GCTAGTGAAGTTCTGACTAGATTTCCAGCAGAGATGAAGGGTATGAACTATGTATTTGCC
1141 1200
CGATCACTTCAAGACTGATCTAAAGGTCGTCTCTACTTCCCATACTTGATACATAAACGG
a A S E V L T R F P A E M K G M N Y V F A -
b L V K F * L D F Q Q R * R V * T M Y L P -
c * * S S D * I S S R D E G Y E L C I C P -
CTTTTGAAATTCAGCTACGACAACCTCGAGAGTGATCTGCTTCGGTCTTGTTTCTTGTAC
1201 1260
GAAAACTTTAAGTCGATGCTGTTGGAGCTCTCACTAGACGAAGCCAGAACAAAGAACATG

图 2C

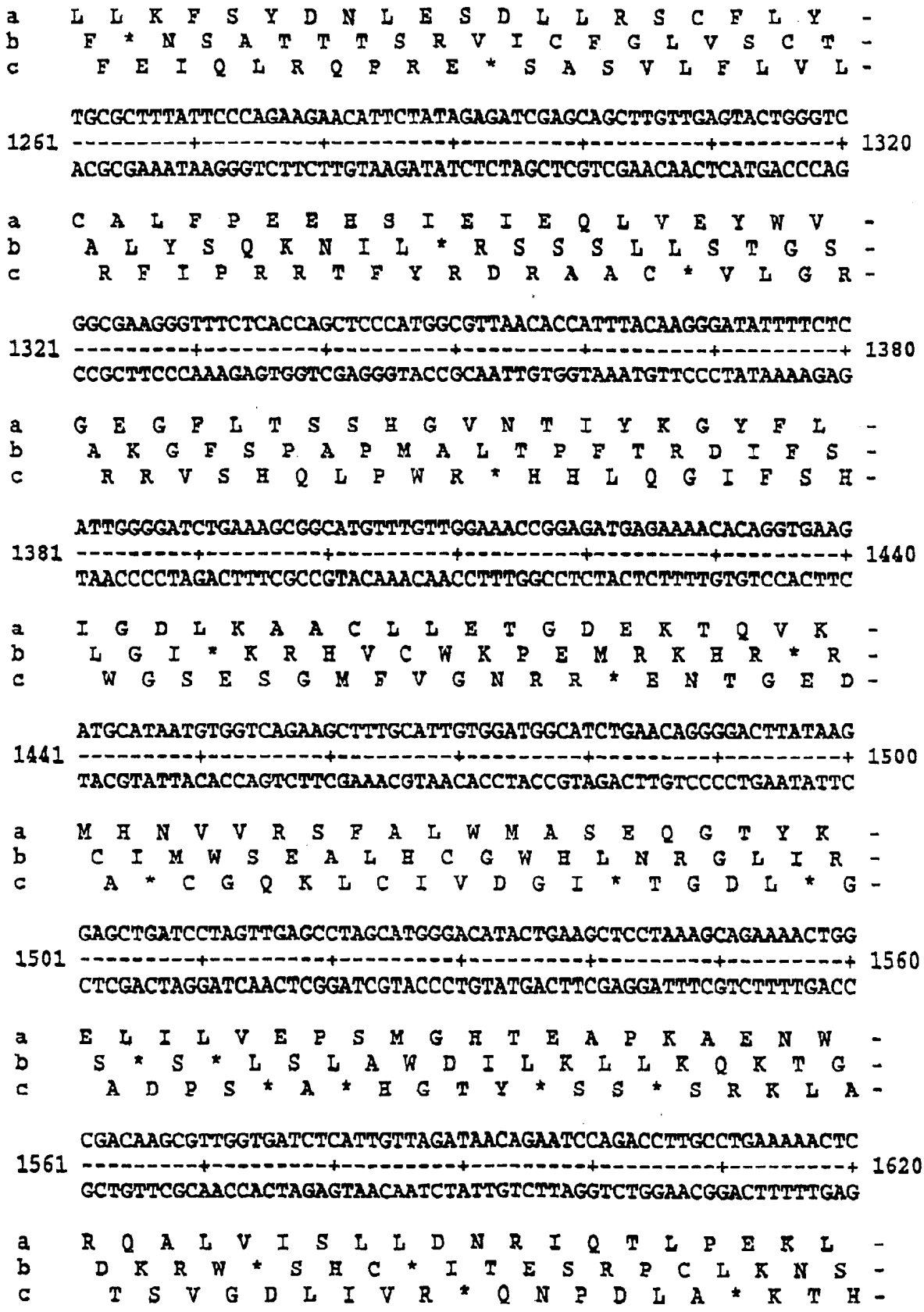

a L L K F S Y D N L E S D L L R S C F L Y -
b F * N S A T T T S R V I C F G L V S C T -
c F E I Q L R Q P R E * S A S V L F L V L -
TGCGCTTTATTCCCAGAAGAACATTCTATAGAGATCGAGCAGCTTGTTGAGTACTGGGTC
1261 ---------+---------+---------+---------+---------+---------+ 1320
ACGCGAAATAAGGGTCTTCTTGTAAGATATCTCTAGCTCGTCGAACAACTCATGACCCAG
a C A L F P E E H S I E I E Q L V E Y W V -
b A L Y S Q K N I L * R S S S L L S T G S -
c R F I P R R T F Y R D R A A C * V L G R -
GGCGAAGGGTTTCTCACCAGCTCCCATGGCGTTAACACCATTTACAAGGGATATTTTCTC
1321 ---------+---------+---------+---------+---------+---------+ 1380
CCGCTTCCCAAAGAGTGGTCGAGGGTACCGCAATTGTGGTAAATGTTCCCTATAAAAGAG
a G E G F L T S S H G V N T I Y K G Y F L -
b A K G F S P A P M A L T P F T R D I F S -
c R R V S H Q L P W R * H H L Q G I F S H -
ATTGGGGATCTGAAAGCGGCATGTTTGTTGGAAACCGGAGATGAGAAAACACAGGTGAAG
1381 ---------+---------+---------+---------+---------+---------+ 1440
TAACCCCTAGACTTTCGCCGTACAAACAACCTTTGGCCTCTACTCTTTTGTGTCCACTTC
a I G D L K A A C L L E T G D E K T Q V K -
b L G I * K R H V C W K P E M R K H R * R -
c W G S E S G M F V G N R R * E N T G E D -
ATGCATAATGTGGTCAGAAGCTTTGCATTGTGGATGGCATCTGAACAGGGGACTTATAAG
1441 ---------+---------+---------+---------+---------+---------+ 1500
TACGTATTACACCAGTCTTCGAAACGTAACACCTACCGTAGACTTGTCCCCTGAATATTC
a M H N V V R S F A L W M A S E Q G T Y K -
b C I M W S E A L H C G W H L N R G L I R -
c A * C G Q K L C I V D G I * T G D L * G -
GAGCTGATCCTAGTTGAGCCTAGCATGGGACATACTGAAGCTCCTAAAGCAGAAAACTGG
1501 ---------+---------+---------+---------+---------+---------+ 1560
CTCGACTAGGATCAACTCGGATCGTACCCTGTATGACTTCGAGGATTTCGTCTTTTGACC
a E L I L V E P S M G H T E A P K A E N W -
b S * S * L S L A W D I L K L L K Q K T G -
c A D P S * A * H G T Y * S S * S R K L A -
CGACAAGCGTTGGTGATCTCATTGTTAGATAACAGAATCCAGACCTTGCCTGAAAAACTC
1561 ---------+---------+---------+---------+---------+---------+ 1620
GCTGTTCGCAACCACTAGAGTAACAATCTATTGTCTTAGGTCTGGAACGGACTTTTTGAG
a R Q A L V I S L L D N R I Q T L P E K L -
b D K R W * S H C * I T E S R P C L K N S -
c T S V G D L I V R * Q N P D L A * K T H -

图 2D

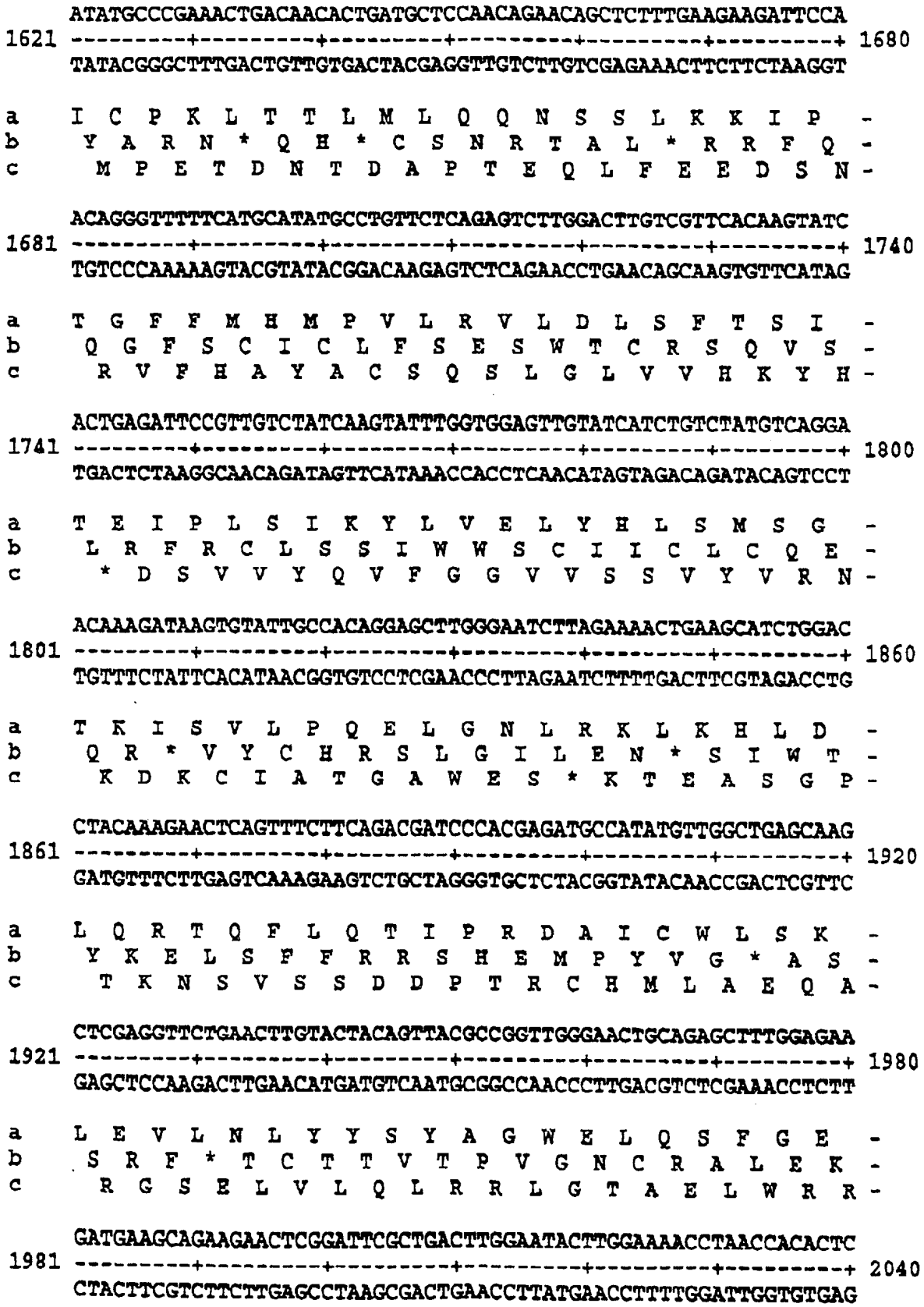

ATATGCCCGAAACTGACAACACTGATGCTCCAACAGAACAGCTCTTTGAAGAAGATTCCA
1621 ---------+---------+---------+---------+---------+---------+ 1680
TATACGGGCTTTGACTGTTGTGACTACGAGGTTGTCTTGTCGAGAAACTTCTTCTAAGGT
a I C P K L T T L M L Q Q N S S L K K I P -
b Y A R N * Q H * C S N R T A L * R R F Q -
c M P E T D N T D A P T E Q L F E E D S N -
ACAGGGTTTTTCATGCATATGCCTGTTCTCAGAGTCTTGGACTTGTCGTTCACAAGTATC
1681 ---------+---------+---------+---------+---------+---------+ 1740
TGTCCCAAAAAGTACGTATACGGACAAGAGTCTCAGAACCTGAACAGCAAGTGTTCATAG
a T G F F M H M P V L R V L D L S F T S I -
b Q G F S C I C L F S E S W T C R S Q V S -
c R V F H A Y A C S Q S L G L V V H K Y H -
ACTGAGATTCCGTTGTCTATCAAGTATTTGGTGGAGTTGTATCATCTGTCTATGTCAGGA
1741 ---------+---------+---------+---------+---------+---------+ 1800
TGACTCTAAGGCAACAGATAGTTCATAAACCACCTCAACATAGTAGACAGATACAGTCCT
a T E I P L S I K Y L V E L Y H L S M S G -
b L R F R C L S S I W W S C I I C L C Q E -
c * D S V V Y Q V F G G V V S S V Y V R N -
ACAAAGATAAGTGTATTGCCACAGGAGCTTGGGAATCTTAGAAAACTGAAGCATCTGGAC
1801 ---------+---------+---------+---------+---------+---------+ 1860
TGTTTCTATTCACATAACGGTGTCCTCGAACCCTTAGAATCTTTTGACTTCGTAGACCTG
a T K I S V L P Q E L G N L R K L K H L D -
b Q R * V Y C H R S L G I L E N * S I W T -
c K D K C I A T G A W E S * K T E A S G P -
CTACAAAGAACTCAGTTTCTTCAGACGATCCCACGAGATGCCATATGTTGGCTGAGCAAG
1861 ---------+---------+---------+---------+---------+---------+ 1920
GATGTTTCTTGAGTCAAAGAAGTCTGCTAGGGTGCTCTACGGTATACAACCGACTCGTTC
a L Q R T Q F L Q T I P R D A I C W L S K -
b Y K E L S F F R R S H E M P Y V G * A S -
c T K N S V S S D D P T R C H M L A E Q A -
CTCGAGGTTCTGAACTTGTACTACAGTTACGCCGGTTGGGAACTGCAGAGCTTTGGAGAA
1921 ---------+---------+---------+---------+---------+---------+ 1980
GAGCTCCAAGACTTGAACATGATGTCAATGCGGCCAACCCTTGACGTCTCGAAACCTCTT
a L E V L N L Y Y S Y A G W E L Q S F G E -
b S R F * T C T T V T P V G N C R A L E K -
c R G S E L V L Q L R R L G T A E L W R R -
GATGAAGCAGAAGAACTCGGATTCGCTGACTTGGAATACTTGGAAAACCTAACCACACTC
1981 ---------+---------+---------+---------+---------+---------+ 2040
CTACTTCGTCTTCTTGAGCCTAAGCGACTGAACCTTATGAACCTTTTGGATTGGTGTGAG

图 2E

```
a   D E A E E L G F A D L E Y L E N L T T L -
b    M K Q K N S D S L T W N T W K T * P H S -
c     * S R R T R I R * L G I L G K P N H T R-

    GGTATCACTGTTCTCTCATTGGAGACCCTAAAAACTCTCTTCGAGTTCGGTGCTTTGCAT
2041 ---------+---------+---------+---------+---------+---------+ 2100
    CCATAGTGACAAGAGAGTAACCTCTGGGATTTTTGAGAGAAGCTCAAGCCACGAAACGTA

a   G I T V L S L E T L K T L F E F G A L H -
b    V S L F S H W R P * K L S S S S V L C I -
c     Y H C S L I G D P K N S L R V R C F A * -

    AAACATATACAGCATCTCCACGTTGAAGAGTGCAATGAACTCCTCTACTTCAATCTCCCA
2101 ---------+---------+---------+---------+---------+---------+ 2160
    TTTGTATATGTCGTAGAGGTGCAACTTCTCACGTTACTTGAGGAGATGAAGTTAGAGGGT

a   K H I Q H L H V E E C N E L L Y F N L P -
b    N I Y S I S T L K S A M N S S T S I S H -
c     T Y T A S P R * R V Q * T P L L Q S P I-

    TCACTCACTAACCATGGCAGGAACCTGAGAAGACTTAGCATTAAAAGTTGCCATGACTTG
2161 ---------+---------+---------+---------+---------+---------+ 2220
    AGTGAGTGATTGGTACCGTCCTTGGACTCTTCTGAATCGTAATTTTCAACGGTACTGAAC

a   S L T N H G R N L R R L S I K S C H D L -
b    H S L T M A G T * E D L A L K V A M T W -
c     T H * P W Q E P E K T * H * K L P * L G-

    GAGTACCTGGTCACACCCGCAGATTTTGAAAATGATTGGCTTCCGAGTCTAGAGGTTCTG
2221 ---------+---------+---------+---------+---------+---------+ 2280
    CTCATGGACCAGTGTGGGCGTCTAAAACTTTTACTAACCGAAGGCTCAGATCTCCAAGAC

a   E Y L V T P A D F E N D W L P S L E V L -
b    S T W S H P Q I L K M I G F R V * R F * -
c     V P G H T R R F * K * L A S E S R G S D-

    ACGTTACACAGCCTTCACAACTTAACCAGAGTGTGGGGAAATTCTGTAAGCCAAGATTGT
2281 ---------+---------+---------+---------+---------+---------+ 2340
    TGCAATGTGTCGGAAGTGTTGAATTGGTCTCACACCCCTTTAAGACATTCGGTTCTAACA

a   T L H S L H N L T R V W G N S V S Q D C -
b    R Y T A F T T * P E C G E I L * A K I V -
c     V T Q P S Q L N Q S V G K F C K P R L S-

    CTGCGGAATATCCGTTGCATAAACATTTCACACTGCAACAAGCTGAAGAATGTCTCATGG
2341 ---------+---------+---------+---------+---------+---------+ 2400
    GACGCCTTATAGGCAACGTATTTGTAAAGTGTGACGTTGTTCGACTTCTTACAGAGTACC

a   L R N I R C I N I S H C N K L K N V S W -
b    C G I S V A * T F H T A T S * R M S H G -
c     A E Y P L H K H F T L Q Q A E E C L M G-

    GTTCAGAAACTCCCAAAGCTAGAGGTGATTGAACTGTTCGACTGCAGAGAGATAGAGGAA
2401 ---------+---------+---------+---------+---------+---------+ 2460
    CAAGTCTTTGAGGGTTTCGATCTCCACTAACTTGACAAGCTGACGTCTCTCTATCTCCTT
```

图 2F

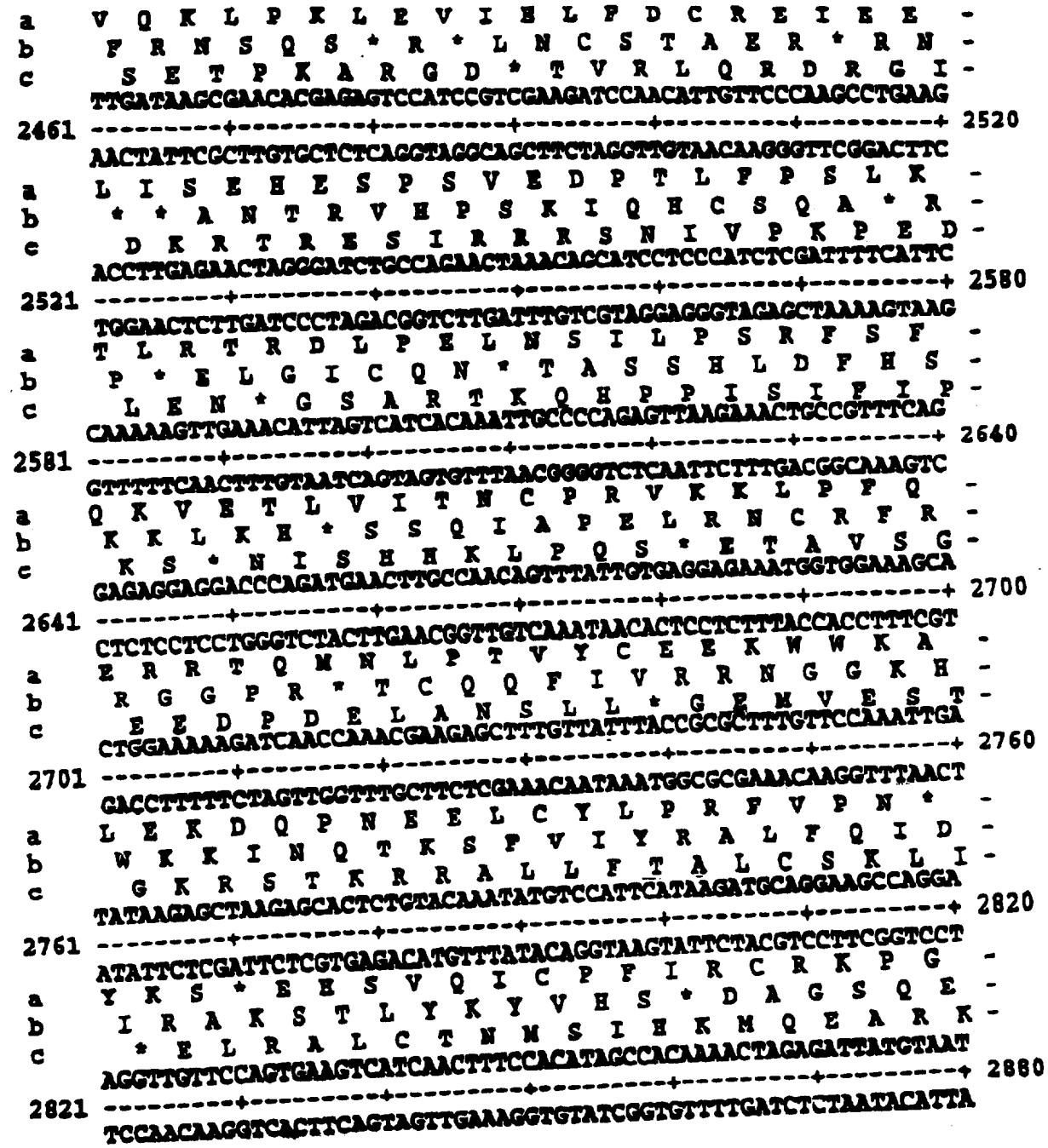


图 2G

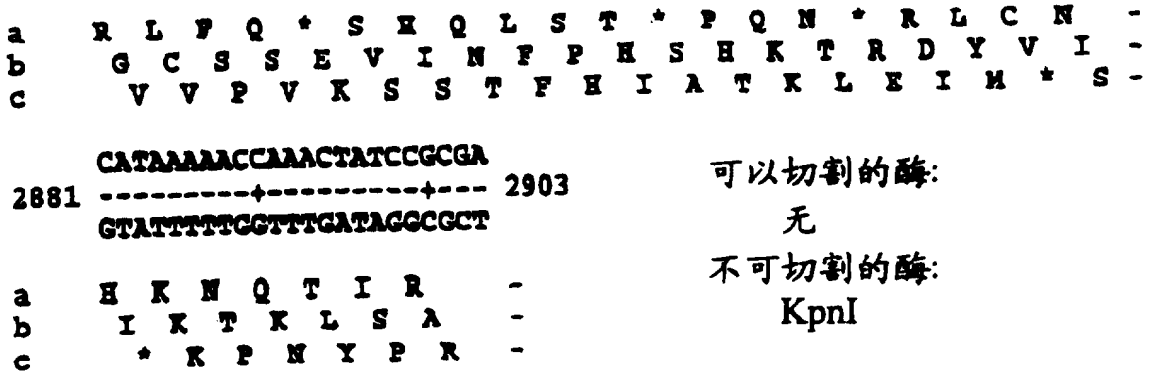


图 2H

```
-146
ATCGATTGATCTCTGGCTCAGTGCGAGTAGTCCATTTGAGAGCAGTCGTAGCCCCGCGTG  -86

GCGCATCATGGAGCTATTTGGAATTTTCGCAGGGTTATCGATTCGTAGTGGGAACCCATT  -26

                1
CATTGTTTGGAACCACCAACGGACGACTTAACAAGCTCCCCGAGGTGCATGATGAAAATT   35
                                                 MetLysIle

GCTCCAGTTGCCATAAATCACAGCCCGCTCAGCAGGGAGGTCCCGTCACACGCGGCACCC   95
AlaProValAlaIleAsnHisSerProLeuSerArgGluValProSerHisAlaAlaPro

ACTCAGGCAAAGCAAACCAACCTTCAATCTGAAGCTGGCGATTTAGATGCAAGAAAAAGT  155
ThrGlnAlaLysGlnThrAsnLeuGlnSerGluAlaGlyAspLeuAspAlaArgLysSer

AGCGCTTCAAGCCCGGAAACCCGCGCATTACTCGCTACTAAGACAGTACTCGGGAGACAC  215
SerAlaSerSerProGluThrArgAlaLeuLeuAlaThrLysThrValLeuGlyArgHis

AAGATAGAGGTTCCGGCCTTTGGAGGGTGGTTCAAAAAGAAATCATCTAAGCACGAGACG  275
LysIleGluValProAlaPheGlyGlyTrpPheLysLysLysSerSerLysHisGluThr

GGCGGTTCAAGTGCCAACGCAGATAGTTCGAGCGTGGCTTCCGATTCCACCGAAAAACCT  335
GlyGlySerSerAlaAsnAlaAspSerSerSerValAlaSerAspSerThrGluLysPro

TTGTTCCGTCTCACGCACGTTCCTTACGTATCCCAAGGTAATGAGCGAATGGGATGTTGG  395
LeuPheArgLeuThrHisValProTyrValSerGlnGlyAsnGluArgMetGlyCysTrp

TATGCCTGCGCAAGAATGGTTGGCCATTCTGTCGAAGCTGGGCCTCGCCTAGGGCTGCCG  455
TyrAlaCysAlaArgMetValGlyHisSerValGluAlaGlyProArgLeuGlyLeuPro

GAGCTCTATGAGGGAAGGGAGGCGCCAGCTGGGCTACAAGATTTTTCAGATGTAGAAAGG  515
GluLeuTyrGluGlyArgGluAlaProAlaGlyLeuGlnAspPheSerAspValGluArg

TTTATTCACAATGAAGGATTAACTCGGGTAGACCTTCCAGACAATGAGAGATTTACACAC  575
PheIleHisAsnGluGlyLeuThrArgValAspLeuProAspAsnGluArgPheThrHis
```

图 3A

```
GAAGAGTTGGGTGCACTGTTGTATAAGCACGGGCCGATTATATTTGGGTGGAAAACTCCG  635
GluGluLeuGlyAlaLeuLeuTyrLysHisGlyProIleIlePheGlyTrpLysThrPro

AATGACAGCTGGCACATGTCGGTCCTCACTGGTGTCGATAAAGAGACGTCGTCCATTACT  695
AsnAspSerTrpHisMetSerValLeuThrGlyValAspLysGluThrSerSerIleThr

TTTCACGATCCCCGACAGGGGCCGGACCTAGCAATGCCGCTCGATTACTTTAATCAGCGA  755
PheHisAspProArgGlnGlyProAspLeuAlaMetProLeuAspTyrPheAsnGlnArg

TTGGCATGGCAGGTTCCACACGCAATGCTCTACCGCTAAGTAGCAGGGTATCTTCACGTG  815
LeuAlaTrpGlnValProHisAlaMetLeuTyrArgEnd

GCGGCATCATGACAAGCCCATGATGCCGCCAGCAGCTACCTGAATGCCGTCTGGCTTTTT  875

GGTCCCTATTGTCGTATCCGGAAGATGACGTCAAAGAATCTCGGCAAGAGCTTTCTTGCT  935

CGACTCCTCAGCTTCCGGATCGATCAGGTCGCTTGCCAGAGCGCGCTTGTCCATGAGCAT  995

CTGCCACAGCTGCTGGTCGATGGTGTCCTCAGCTAAAGGGATTTTGACGACAACCATGCG 1055

CAACTGCCCGTTGCGATACGCTCGATCCTGAAGCCCCGGTGTCCATGGCAGCCCCAAGAA 1115

AAAGACATAGTTCGCCGCTGTGAGGTTGTAGCCTGTGCCGGCGGCCGACCTGGTCCCGAT 1175

AAACACCCTGCAGTCCGGATCCTGCTGGAAAGCATCAATCGCCTTCTGCCGCTTCTTGGG 1235

CGAGTCACTGCCCACCAACGTCACGCACCCGACGCCAAGCTTGAGGCAGTGCTCCCGCAA 1295

CGTGGCCACGGATTCCTGATACTCGCAGAAGAGGATCACCTTGTCGTCGAC  1346
```

图 3B

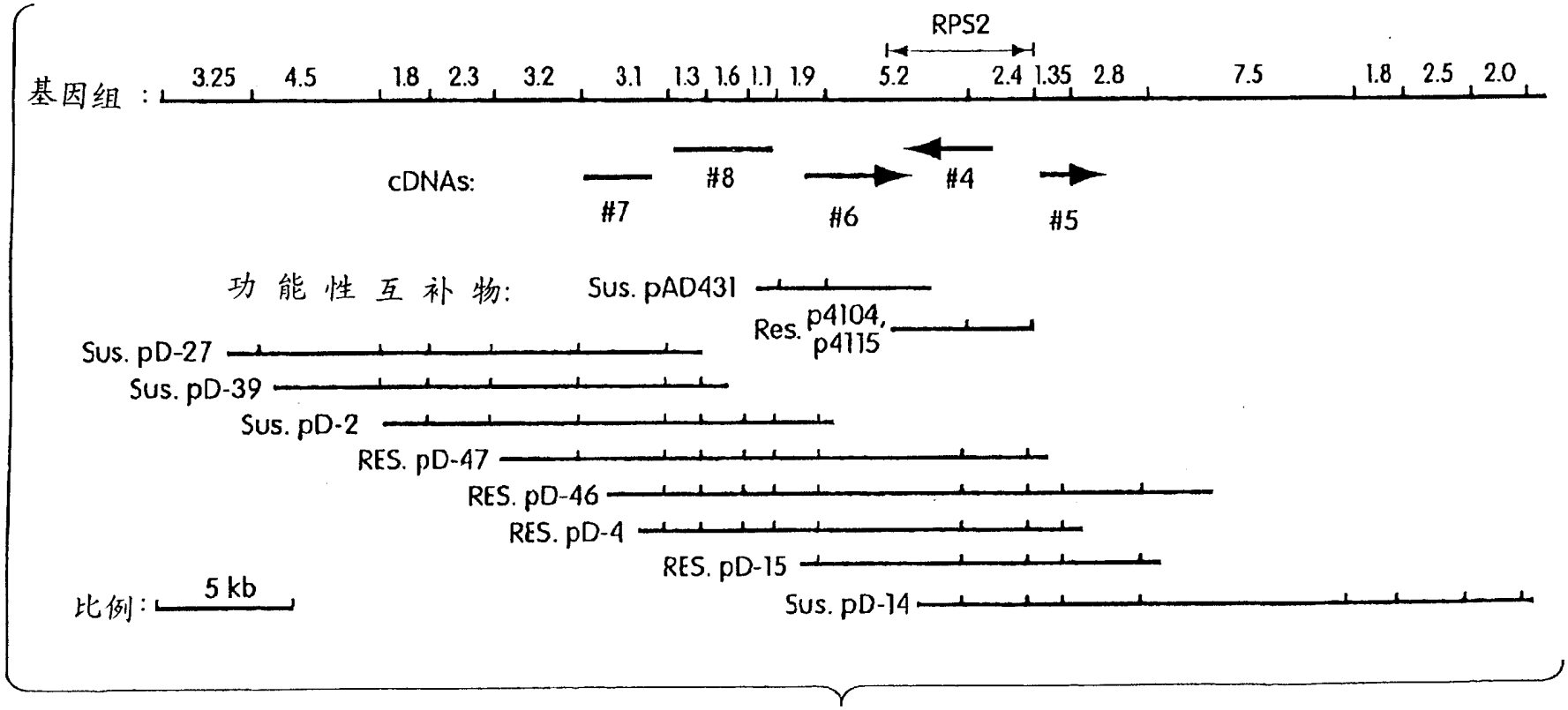

RPS2
基因组：
3.25 4.5 1.8 2.3 3.2 3.1 1.3 1.6 1.1 1.9 5.2 2.4 1.35 2.8 7.5 1.8 2.5 2.0
cDNAs:
#7
#8
#6
#4
#5
功能性互补物:
Sus. pAD431
Res. p4104, p4115
Sus. pD-27
Sus. pD-39
Sus. pD-2
RES. pD-47
RES. pD-46
RES. pD-4
RES. pD-15
Sus. pD-14
比例：
5 kb

图 4