



(12) 发明专利申请

(10) 申请公布号 CN 101892249 A

(43) 申请公布日 2010. 11. 24

(21) 申请号 201010222216. 4

C12Q 1/68 (2006. 01)

(22) 申请日 2002. 04. 16

C12Q 1/04 (2006. 01)

(30) 优先权数据

C12P 7/62 (2006. 01)

60/284, 066 2001. 04. 16 US

C12P 7/64 (2006. 01)

60/298, 796 2001. 06. 15 US

A01H 5/00 (2006. 01)

60/323, 269 2001. 09. 18 US

(62) 分案原申请数据

02811749. 2 2002. 04. 16

(71) 申请人 马泰克生物科学公司

地址 美国马里兰州

(72) 发明人 詹姆斯·G·梅茨 威廉·R·巴克利

詹姆斯·H·弗拉特 杰里·M·库纳

(74) 专利代理机构 北京市柳沈律师事务所

11105

代理人 闵丹

(51) Int. Cl.

C12N 15/52 (2006. 01)

C12N 15/11 (2006. 01)

C12N 5/10 (2006. 01)

C12N 1/15 (2006. 01)

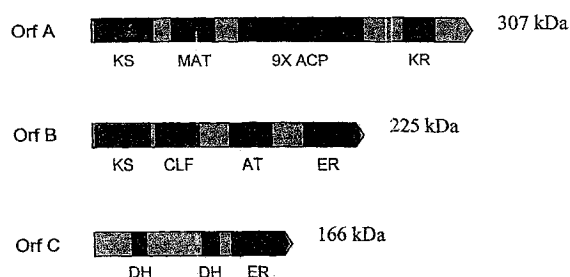
权利要求书 17 页 说明书 54 页 序列表 118 页
附图 3 页

(54) 发明名称

PUFA 聚酮化合物合酶系统及其用途

(57) 摘要

本发明一般涉及从非细菌生物分离或衍生的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统, 涉及其同源物, 涉及编码该 PUFA PKS 系统的生物学活性结构域的分离的核酸分子和重组核酸分子, 涉及制备和使用该系统用于生产所需生物活性分子的方法, 且涉及鉴定具有该 PUFA PKS 系统的新细菌和非细菌微生物的新方法。



1. 一种分离的核酸分子,它含有选自下组的核酸序列:
 - a. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;
 - b. 编码选自 SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ IDNO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;
 - c. 编码与(a)所述氨基酸序列中至少500个连续氨基酸有至少约60%相同的氨基酸序列的核酸序列,其中所述氨基酸序列具有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的生物学活性;
 - d. 编码与(b)所述氨基酸序列有至少约60%相同性的氨基酸序列的核酸序列,其中所述氨基酸序列具有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的生物学活性;和
 - e. 与(a), (b), (c), 或(d)的核酸序列完全互补的核酸序列。
2. 权利要求1的分离的核酸分子,其中所述核酸序列选自:
 - a. 编码与选自:SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6的氨基酸序列的至少500个连续氨基酸有至少约70%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFA PKS系统的至少一个结构域的生物学活性;和
 - b. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ IDNO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32的氨基酸序列有至少约70%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFAPKS系统的至少一个结构域的生物学活性。
3. 权利要求1的分离的核酸分子,其中所述核酸序列选自:
 - a. 编码与选自:SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6的氨基酸序列的至少500个连续氨基酸有至少约80%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFAPKS系统的至少一个结构域的生物学活性;和,
 - b. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ IDNO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32的氨基酸序列有至少约80%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFAPKS系统的至少一个结构域的生物学活性。
4. 权利要求1的分离的核酸分子,其中所述核酸序列选自:
 - a. 编码与选自:SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6的氨基酸序列的至少500个连续氨基酸有至少约90%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFAPKS系统的至少一个结构域的生物学活性;和,
 - b. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ IDNO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32的氨基酸序列有至少约90%相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有PUFAPKS系统的至少一个结构域的生物学活性。
5. 权利要求1的分离的核酸分子,其中所述核酸序列编码选自:SEQ IDNO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQID NO:13, SEQ ID NO:18, SEQ ID

NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 或其生物活性片段的氨基酸序列。

6. 权利要求 1 的分离的核酸分子,其中所述核酸分子包含选自:SEQ IDNO:1, SEQ ID NO:3,SEQ ID NO:5,SEQ ID NO:7,SEQ ID NO:9,SEQID NO:12,SEQ ID NO:17,SEQ ID NO:19,SEQ ID NO:21,SEQ ID NO:23,SEQ ID NO:25,SEQ ID NO:27,SEQ ID NO:29,或SEQ ID NO:31 的核酸序列。

7. 一种重组核酸分子,它包含与至少一个转录调控序列可操作地相连的权利要求 1 的核酸分子。

8. 用权利要求 7 的重组核酸分子转染的重组细胞。

9. 一种遗传修饰的微生物,其中所述微生物表达含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个生物学活性结构域的 PKS 系统,其中所述 PUFAPKS 系统的至少一个结构域由选自下组的核酸序列编码:

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:

(i) 选择产生至少一种 PUFA 的微生物;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

c. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;

d. 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ IDNO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;

e. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

f. 编码与选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,或SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性;

其中所述微生物被遗传修饰成影响所述 PKS 系统的活性。

10. 权利要求 9 的遗传修饰的微生物,其中所述微生物内源性表达含有所述 PUFA PKS 系统的至少一个结构域的 PKS 系统,且其中在编码所述 PUFA PKS 系统的所述至少一个结构域的核酸序列中存在所述遗传修饰。

11. 权利要求 10 的遗传修饰的微生物,其中在编码具有下列至少一种蛋白质的生物学活性的结构域的核酸序列中存在所述遗传修饰:丙二酰-CoA:ACP 酰基转移酶 (MAT), β -酮脂酰-ACP 合酶 (KS), 酮还原酶 (KR), 酰基转移酶 (AT), FabA-样 β -羟酰-ACP 脱水酶 (DH),

磷酸泛酰巯基乙胺转移酶,链长因子 (CLF),酰基载体蛋白 (ACP),烯酰 ACP-还原酶 (ER),催化反式 -2- 癸烯酰 -ACP 合成的酶,催化反式 -2- 癸烯酰 -ACP 向顺式 -3- 癸烯酰 -ACP 可逆异构化的酶,和催化顺式 -3- 癸烯酰 -ACP 延长成顺式 -11- 十八碳烯酸的酶。

12. 权利要求 10 的遗传修饰的微生物,其中在编码选自下组的氨基酸序列的核酸序列中存在所述遗传修饰:

a. 与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 70%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 70%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

13. 权利要求 10 的遗传修饰的微生物,其中在编码选自下组的氨基酸序列的核酸序列中存在所述遗传修饰:

a. 与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 80%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 80%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

14. 权利要求 10 的遗传修饰的微生物,其中在编码选自下组的氨基酸序列的核酸序列中存在所述遗传修饰:

a. 与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 90%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 90%相同性的氨基酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

15. 权利要求 10 的遗传修饰的微生物,其中在编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 或其生物活性片段的氨基酸序列的核酸序列中存在所述遗传修饰。

16. 权利要求 10 的遗传修饰的微生物,其中所述微生物是 Thraustochytrid。

17. 权利要求 16 的遗传修饰的微生物,其中所述 Thraustochytrid 来自于 Schizochytrium 属或破囊壶菌属。

18. 权利要求 10 的遗传修饰的微生物,其中所述微生物进一步被遗传修饰成重组表达编码来自细菌 PUFA PKS 系统的至少一个生物学活性结构域的至少一个核酸分子。

19. 权利要求 10 的遗传修饰的微生物,其中所述微生物进一步被遗传修饰成重组表达来自 I 型 PKS 系统的一个 PKS 系统的至少一个生物学活性结构域。

20. 权利要求 10 的遗传修饰的微生物,其中所述微生物进一步被遗传修饰成重组表达来自 II 型 PKS 系统的一个 PKS 系统的至少一个生物学活性结构域。

21. 权利要求 10 的遗传修饰的微生物,其中所述微生物进一步被遗传修饰成重组表达来自模块 PKS 系统的一个 PKS 系统的至少一个生物学活性结构域。

22. 权利要求 9 的遗传修饰的微生物,其中所述微生物内源性表达含有 PUFA PKS 系统的所述至少一个生物学活性结构域的 PUFA PKS 系统,且其中所述遗传修饰包含表达选自由编码来自第二种 PKS 系统的至少一个生物学活性结构域的重组核酸分子和编码影响所述 PUFA PKS 系统活性的蛋白质的重组核酸分子组成的组中的一个重组核酸分子。

23. 权利要求 22 的遗传修饰的微生物,其中所述重组核酸分子包含选自下组的核酸序列:

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:

(i) 选择产生至少一种 PUFA 的微生物;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

c. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;

d. 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;

e. 编码与 (a) 中所述氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性;和

f. 编码与 (b) 中所述氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性。

24. 权利要求 23 的遗传修饰的微生物,其中 (a) 中所述 Thraustochytrid 微生物来自于 Schizochytrium 属或破囊壶菌属。

25. 权利要求 23 的遗传修饰的微生物,其中所述重组核酸分子包含选自下组的核酸序列:

a. 编码与选自氨基酸序列 SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的至少 500 个连续氨基酸具有至少约 70% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 编码与选自 SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:

20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 70% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。

26. 权利要求 23 的遗传修饰的微生物, 其中所述重组核酸分子包含选自下组的核酸序列 :

a. 编码与选自 SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 80% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ; 和,

b. 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 80% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

27. 权利要求 23 的遗传修饰的微生物, 其中所述重组核酸分子包含选自下组的核酸序列 :

a. 编码与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 90% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性 ; 和,

b. 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 90% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

28. 权利要求 23 的遗传修饰的微生物, 其中所述重组核酸分子包含编码选自 :SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6, SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, SEQ ID NO :32 或其生物活性片段的氨基酸序列的核酸序列。

29. 权利要求 22 的遗传修饰的微生物, 其中所述重组核酸分子编码磷酸泛酰巯基乙胺转移酶。

30. 权利要求 22 的遗传修饰的微生物, 其中所述重组核酸分子包含编码来自细菌 PUFA PKS 系统的至少一个生物学活性结构域的核酸序列。

31. 权利要求 22 的遗传修饰的微生物, 其中所述重组核酸分子包含编码来自 I 型 PKS 系统的至少一个生物学活性结构域的核酸序列。

32. 权利要求 22 的遗传修饰的微生物, 其中所述重组核酸分子包含编码来自 II 型 PKS 系统的至少一个生物学活性结构域的核酸序列。

33. 权利要求 22 的遗传修饰的微生物, 其中所述重组核酸分子包含编码来自模块型 PKS 系统的至少一个生物学活性结构域的核酸序列。

34. 权利要求 9 的遗传修饰的微生物, 其中所述微生物通过用编码多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的所述至少一个结构域的重组核酸分子转染进行遗传修饰。

35. 权利要求 34 的遗传修饰的微生物, 其中所述重组核酸分子包含选自下组的核酸序

列：

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列；

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列：

(i) 选择产生至少一种 PUFA 的微生物；和，

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比，在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物；

c. 编码选自：SEQ ID NO：2，SEQ ID NO：4，SEQ ID NO：6，或其生物学活性片段的氨基酸序列的核酸序列；

d. 编码选自：SEQ ID NO：8，SEQ ID NO：10，SEQ ID NO：13，SEQ ID NO：18，SEQ ID NO：20，SEQ ID NO：22，SEQ ID NO：24，SEQ ID NO：26，SEQ ID NO：28，SEQ ID NO：30，SEQ ID NO：32，或其生物学活性片段的氨基酸序列的核酸序列；

e. 编码与 (a) 中所述氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性；和

f. 编码与 (b) 中所述氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性。

36. 权利要求 35 的遗传修饰的微生物，其中 (a) 中所述 Thraustochytrid 来自于 Schizochytrium 属或破囊壶菌属。

37. 权利要求 35 的遗传修饰的微生物，其中所述重组核酸分子包含选自下组的核酸序列：

a. 编码与选自：SEQ ID NO：2，SEQ ID NO：4，或 SEQ ID NO：6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 70% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性；和，

b. 编码与选自：SEQ ID NO：8，SEQ ID NO：10，SEQ ID NO：13，SEQ ID NO：18，SEQ ID NO：20，SEQ ID NO：22，SEQ ID NO：24，SEQ ID NO：26，SEQ ID NO：28，SEQ ID NO：30，或 SEQ ID NO：32 的氨基酸序列具有至少约 70% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

38. 权利要求 35 的遗传修饰的微生物，其中所述重组核酸分子包含选自下组的核酸序列：

a. 编码与选自：SEQ ID NO：2，SEQ ID NO：4，或 SEQ ID NO：6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 80% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性；和，

b. 编码与选自：SEQ ID NO：8，SEQ ID NO：10，SEQ ID NO：13，SEQ ID NO：18，SEQ ID NO：20，SEQ ID NO：22，SEQ ID NO：24，SEQ ID NO：26，SEQ ID NO：28，SEQ ID NO：30，或 SEQ ID NO：32 的氨基酸序列具有至少约 80% 相同性的氨基酸序列的核酸序列；其中所述氨基酸序

列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

39. 权利要求 35 的遗传修饰的微生物,其中所述重组核酸分子包含选自下组的核酸序列:

a. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 90% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 编码与选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 90% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

40. 权利要求 35 的遗传修饰的微生物,其中所述重组核酸分子包含编码选自:SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,SEQ ID NO:32 或其生物活性片段的氨基酸序列的核酸序列。

41. 权利要求 34 的遗传修饰的微生物,其中所述微生物被进一步遗传修饰成重组表达编码来自细菌 PUFA PKS 系统的至少一个生物学活性结构域的至少一种核酸分子。

42. 权利要求 34 的遗传修饰的微生物,其中所述微生物被进一步遗传修饰成重组表达来自 I 型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

43. 权利要求 34 的遗传修饰的微生物,其中所述微生物被进一步遗传修饰成重组表达来自 II 型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

44. 权利要求 34 的遗传修饰的微生物,其中所述微生物被进一步遗传修饰成重组表达来自模块型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

45. 一种遗传修饰的植物,其中所述植物被遗传修饰成重组表达包含多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个生物学活性结构域的 PKS 系统,其中所述结构域由选自下组的核酸序列编码:

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的核酸序列;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:

(i) 选择产生至少一种 PUFA 的微生物;和,

(ii) 从(i)中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

c. 编码选自:SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;

d. 编码选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;

e. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500

个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ; 和,

f. 编码与选自 : SEQ ID NO : 8, SEQ ID NO : 10, SEQ ID NO : 13, SEQ ID NO : 18, SEQ ID NO : 20, SEQ ID NO : 22, SEQ ID NO : 24, SEQ ID NO : 26, SEQ ID NO : 28, SEQ ID NO : 30, 或 SEQ ID NO : 32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

46. 权利要求 45 的遗传修饰的植物, 其中所述核酸序列选自 :

a. 编码与选自 : SEQ ID NO : 2, SEQ ID NO : 4, 或 SEQ ID NO : 6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 70% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ; 和,

b. 编码与选自 : SEQ ID NO : 8, SEQ ID NO : 10, SEQ ID NO : 13, SEQ ID NO : 18, SEQ ID NO : 20, SEQ ID NO : 22, SEQ ID NO : 24, SEQ ID NO : 26, SEQ ID NO : 28, SEQ ID NO : 30, 或 SEQ ID NO : 32 的氨基酸序列具有至少约 70% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。

47. 权利要求 45 的遗传修饰的植物, 其中所述核酸序列选自 :

a. 编码与选自 : SEQ ID NO : 2, SEQ ID NO : 4, 或 SEQ ID NO : 6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 80% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性 ; 和,

b. 编码与选自 : SEQ ID NO : 8, SEQ ID NO : 10, SEQ ID NO : 13, SEQ ID NO : 18, SEQ ID NO : 20, SEQ ID NO : 22, SEQ ID NO : 24, SEQ ID NO : 26, SEQ ID NO : 28, SEQ ID NO : 30, 或 SEQ ID NO : 32 的氨基酸序列具有至少约 80% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

48. 权利要求 45 的遗传修饰的植物, 其中所述核酸序列选自 :

a. 编码与选自 : SEQ ID NO : 2, SEQ ID NO : 4, 或 SEQ ID NO : 6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 90% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性 ; 和,

b. 编码与选自 : SEQ ID NO : 8, SEQ ID NO : 10, SEQ ID NO : 13, SEQ ID NO : 18, SEQ ID NO : 20, SEQ ID NO : 22, SEQ ID NO : 24, SEQ ID NO : 26, SEQ ID NO : 28, SEQ ID NO : 30, 或 SEQ ID NO : 32 的氨基酸序列具有至少约 90% 相同性的氨基酸序列的核酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

49. 权利要求 45 的遗传修饰的植物, 其中所述核酸序列编码选自 : SEQ ID NO : 2, SEQ ID NO : 4, SEQ ID NO : 6, SEQ ID NO : 8, SEQ ID NO : 10, SEQ ID NO : 13, SEQ ID NO : 18, SEQ ID NO : 20, SEQ ID NO : 22, SEQ ID NO : 24, SEQ ID NO : 26, SEQ ID NO : 28, SEQ ID NO : 30, SEQ ID NO : 32 或其生物活性片段的氨基酸序列。

50. 权利要求 45 的遗传修饰的植物, 其中 (a) 中所述 Thraustochytrid 来自于 Schizochytrium 属或破囊壶菌属。

51. 权利要求 45 的遗传修饰的植物, 其中所述植物被进一步遗传修饰成重组表达编码来自细菌 PUFAPKS 系统的至少一个生物学活性结构域的至少一个核酸分子。

52. 权利要求 45 的遗传修饰的植物, 其中所述植物被进一步遗传修饰成重组表达来自

I 型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

53. 权利要求 45 的遗传修饰的植物,其中所述植物被进一步遗传修饰成重组表达来自 II 型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

54. 权利要求 45 的遗传修饰的植物,其中所述植物被进一步遗传修饰成重组表达来自模块型 PKS 系统的一种 PKS 系统的至少一个生物学活性结构域。

55. 一种鉴定具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的微生物的方法,包括:

a. 选择产生至少一种 PUFA 的微生物;和,

b. 从 (a) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

其中将产生至少一种 PUFA 且在不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物鉴定为含有 PUFA PKS 系统的候选微生物。

56. 权利要求 55 的方法,其中步骤 (b) 包括从 (a) 中鉴定在不足约 2% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。

57. 权利要求 55 的方法,其中步骤 (b) 包括从 (a) 中鉴定在不足约 1% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。

58. 权利要求 55 的方法,其中步骤 (b) 包括从 (a) 中鉴定在约 0% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。

59. 权利要求 55 的方法,其中在 (a) 中所选微生物具有通过吞噬作用消费细菌的能力。

60. 权利要求 55 的方法,其中在 (a) 中所选微生物具有简单的脂肪酸分布特征。

61. 权利要求 55 的方法,其中在 (a) 中所选微生物是非细菌微生物。

62. 权利要求 55 的方法,其中在 (a) 中所选微生物是真核生物。

63. 权利要求 55 的方法,其中在 (a) 中所选微生物是 Thraustochytriales 目的一个成员。

64. 权利要求 55 的方法,其中在 (a) 中所选微生物在高于约 15°C 的温度下具有生产 PUFA 的能力。

65. 权利要求 55 的方法,其中在 (a) 中所选微生物在高于约 20°C 的温度下具有生产 PUFA 的能力。

66. 权利要求 55 的方法,其中在 (a) 中所选微生物在高于约 25°C 的温度下具有生产 PUFA 的能力。

67. 权利要求 55 的方法,其中在 (a) 中所选微生物在高于约 30°C 的温度中具有生产 PUFA 的能力。

68. 权利要求 55 的方法,其中在 (a) 中所选微生物具有产生生物体干重的约 5% 以上的所需生物活性化合物的能力。

69. 权利要求 55 的方法,其中在 (a) 中所选微生物具有产生生物体干重的约 10% 以上的所需生物活性化合物的能力。

70. 权利要求 55 的方法,其中在 (a) 中所选微生物的总脂肪酸约 30% 以上是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 3 个或更多个不饱和键的长链脂肪酸。

71. 权利要求 55 的方法,其中在 (a) 中所选微生物的总脂肪酸约 40% 以上是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 3 个或更多个不饱和键的长链脂肪酸。

72. 权利要求 55 的方法,其中在 (a) 中所选微生物的总脂肪酸约 30% 以上是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 4 个或更多个不饱和键的长链脂肪酸。

73. 权利要求 55 的方法,其中在 (a) 中所选微生物的总脂肪酸约 30% 以上是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 5 个或更多个不饱和键的长链脂肪酸。

74. 权利要求 55 的方法,还包含:

a. 检测所述生物体是否含有 PUFA PKS 系统。

75. 权利要求 74 的方法,其中所述检测步骤包括检测在所述微生物中与编码来自 Thraustochytrid PUFA PKS 系统的氨基酸序列的核酸序列在严格条件下杂交的核酸序列。

76. 权利要求 74 的方法,其中所述检测步骤包括检测在所述生物体中用来自 Thraustochytrid PUFA PKS 系统的核酸序列的寡核苷酸引物扩增的核酸序列。

77. 由权利要求 55 的方法鉴定的且被遗传修饰成调节所述 PUFA PKS 系统产生的分子的微生物。

78. 一种生产由聚酮化合物合酶系统产生的生物活性分子的方法,包括在有效产生所述生物活性分子的条件培养表达 PKS 系统的遗传修饰的生物体,所述 PKS 系统含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个生物学活性结构域,其中所述 PUFA PKS 系统的所述至少一个结构域由选自下组的核酸序列编码:

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:

(i) 选择产生至少一种 PUFA 的微生物;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

c. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;

d. 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;

e. 编码与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

f. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;

其中所述生物体被遗传修饰成影响所述 PKS 系统的活性。

79. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 来自于破囊壶菌属或 Schizochytrium 属。

80. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 是破囊壶菌 23B(ATCC 20892)。

81. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 是 Schizochytrium sp. SR21。

82. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 是 Schizochytrium 菌株 ATCC 20888。

83. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 是 Schizochytrium limaeium(IFO 32693)。

84. 权利要求 78 的方法, 其中 (a) 中所述 Thraustochytrid 是 Ulkenia(BP-5601)。

85. 权利要求 78 的方法, 其中所述核酸序列选自:

a. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 70% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,

b. 编码与选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 70% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

86. 权利要求 78 的方法, 其中所述核酸序列选自:

a. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 80% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性;和,

b. 编码与选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 80% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。

87. 权利要求 78 的方法, 其中所述核酸序列选自:

a. 编码与选自:SEQ ID NO:2,SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 90% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性;和,

b. 编码与选自:SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 90% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

88. 权利要求 78 的方法, 其中所述核酸序列编码选自:SEQ ID NO:2,SEQ ID NO:4,SEQ ID NO:6,SEQ ID NO:8,SEQ ID NO:10,SEQ ID NO:13,SEQ ID NO:18,SEQ ID NO:20,SEQ ID NO:22,SEQ ID NO:24,SEQ ID NO:26,SEQ ID NO:28,SEQ ID NO:30,SEQ ID NO:32 或其生物活性片段的氨基酸序列。

89. 权利要求 78 的方法,其中所述生物体内源性表达含有所述 PUFAPKS 系统的所述至少一个结构域的 PKS 系统,且其中所述遗传修饰处在编码所述 PUFAPKS 系统的所述至少一个结构域的核酸序列上。

90. 权利要求 89 的方法,其中所述遗传修饰与野生型生物体相比改变了由所述内源性 PKS 系统产生的至少一种产物。

91. 权利要求 78 的方法,其中所述生物体内源性表达含有所述 PUFAPKS 系统的所述至少一个生物活性结构域的 PKS 系统,且其中所述遗传修饰包括用选自:编码来自第二种 PKS 系统的至少一个生物学活性结构域的重组核酸分子和编码影响所述 PUFAPKS 系统活性的蛋白质的重组核酸分子组成的组中的一种重组核酸分子转染所述生物体。

92. 权利要求 91 的方法,其中所述遗传修饰使得与野生型生物体相比改变了由所述内源性 PKS 系统产生的至少一种产物。

93. 权利要求 78 的方法,其中所述生物体通过用编码所述多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的所述至少一个结构域的重组核酸分子转染而被遗传修饰。

94. 权利要求 78 的方法,其中所述生物体产生一种多不饱和脂肪酸 (PUFA) 的分布特征,该分布特征不同于没有遗传修饰的天然生物体。

95. 权利要求 78 的方法,其中所述生物体内源性表达非细菌 PUFAPKS 系统,且其中所述遗传修饰包括来自不同 PKS 系统的结构域取代编码所述非细菌 PUFA PKS 系统的至少一个结构域的核酸序列。

96. 权利要求 78 的方法,其中所述生物体内源性表达非细菌 PUFAPKS 系统,该系统通过用重组核酸分子转染所述生物体而被修饰,该重组核酸分子编码能对所述 PUFA PKS 系统产生的脂肪酸的链长进行调节的蛋白质。

97. 权利要求 96 的方法,其中编码调节脂肪酸链长的蛋白质的所述重组核酸分子取代了所述非细菌 PUFA PKS 系统中编码链长因子的核酸序列。

98. 权利要求 96 的方法,其中调节所述 PUFA PKS 系统产生的脂肪酸的链长的所述蛋白质是链长因子。

99. 权利要求 96 的方法,其中调节所述 PUFA PKS 系统产生的脂肪酸的链长的所述蛋白质是指导合成 C20 单位的链长因子。

100. 权利要求 96 的方法,其中所述生物体表达非细菌 PUFA PKS 系统,该在编码 FabA-样 β -羟酰 -ACP 脱水酶 (DH) 结构域的区域或编码 β -酮脂酰 -ACP 合酶 (KS) 的区域中含有遗传修饰,其中所述修饰与没有所述修饰相比改变了所述 PUFAPKS 系统产生的长链脂肪酸的比率。

101. 权利要求 100 的方法,其中所述修饰包含用不具有异构化活性的 DH 结构域取代所述非细菌 PUFA PKS 系统中的 FabA-样 β -羟酰 -ACP 脱水酶 (DH)。

102. 权利要求 100 的方法,其中所述修饰选自缺失所述区域的全部或部分,用来自不同生物的同源区域取代所述区域,或突变所述区域。

103. 权利要求 96 的方法,其中所述生物体表达非 -PUFAPKS 系统且其中所述遗传修饰包含用来自 PUFA PKS 系统的 FabA-样 β -羟酰 -ACP 脱水酶 (DH) 区域取代不具有异构化活性的 DH 区域。

104. 权利要求 78 的方法,其中所述生物体表达在烯酰 -ACP 还原酶 (ER) 区域中含有修

饰的非细菌 PUFAPKS 系统,其中所述修饰导致与没有所述修饰相比产生不同的化合物。

105. 权利要求 104 的方法,其中所述修饰选自缺失所述 ER 区域的全部或部分,用来自不同生物体的 ER 区域取代所述 ER 区域,或突变所述 ER 区域。

106. 权利要求 78 的方法,其中所述生物活性分子选自:抗炎症配制品,化疗剂,有效赋形剂,骨质疏松症药物,抗抑郁剂,抗惊厥剂,抗幽门螺杆菌 (*Helicobacter pylori*) 药物,治疗神经变性疾病药物,治疗变性肝脏疾病的药物,抗生素,或降胆固醇配制品。

107. 权利要求 78 的方法,其中所述生物活性分子是抗生素。

108. 权利要求 78 的方法,其中所述生物活性分子是多不饱和脂肪酸 (PUFA)。

109. 权利要求 78 的方法,其中所述生物活性分子是包含顺式构型的碳-碳双键的分子。

110. 权利要求 78 的方法,其中所述生物活性分子是在每第三个碳包含一个双键的分子。

111. 权利要求 78 的方法,其中所述生物体是微生物。

112. 权利要求 78 的方法,其中所述生物体是植物。

113. 产生具有不同于天然植物的多不饱和脂肪酸 (PUFA) 分布特征的植物的方法,包括对所述植物的细胞进行遗传修饰,以表达含有至少一种重组核酸分子的 PKS 系统,该重组核酸分子含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列,其中所述 PUFA PKS 系统的所述至少一个结构域由选自下组的核酸序列编码:

a. 编码来自 *Thraustochytrid* 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:

(i) 选择产生至少一种 PUFA 的微生物;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;

c. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;

d. 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;

e. 编码与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性;和,

f. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。

114. 一种修饰含有至少一种脂肪酸的终产物的方法,包括向所述终产物中添加一种由

重组宿主细胞产生的油类,该重组宿主细胞表达含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的至少一个重组核酸分子,其中所述 PUFA PKS 系统的至少一个结构域由选自下组的核酸序列编码:

- a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;
- b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列:
 - (i) 选择产生至少一种 PUFA 的微生物;和,
 - (ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物;
- c. 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;
- d. 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;
- e. 编码与选自:SEQ ID NO:2, SEQ ID NO:4,或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和,
- f. 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

115. 权利要求 114 的方法,其中所述终产物选自饮食添加剂,食品,药物配制品,人源化动物乳汁,或婴儿配制品。

116. 权利要求 115 的方法,其中所述药物配制品选自由抗炎症配制品,化疗剂,有效赋形剂,骨质疏松症药物,抗抑郁剂,抗惊厥剂,抗幽门螺杆菌药物,治疗神经变性疾病的药物,治疗变性肝脏疾病的药物,抗生素,和降胆固醇配制品组成的组中。

117. 权利要求 114 的方法,其中所述终产物用于治疗下述疾病:慢性炎症,急性炎症,胃肠疾病,癌症,恶病质,心脏再狭窄,神经变性疾病,肝脏变性疾病,血脂疾病,骨质疏松症,骨关节炎,自身免疫疾病,先兆子痫,早产,年老相关性黄斑病,肺病,或过氧化物酶体疾病。

118. 一种生产人源化动物乳汁的方法,包括用含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的至少一种重组核酸分子对产奶动物的产奶细胞进行遗传修饰,其中所述 PUFA PKS 系统的所述至少一个结构域由选自下组的核酸序列编码:

- a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;
- b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列;

(i) 选择产生至少一种 PUFA 的微生物 ;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物 ;

c. 编码选自 :SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6,或其生物学活性片段的氨基酸序列的核酸序列 ;

d. 编码选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, SEQ ID NO :32,或其生物学活性片段的氨基酸序列的核酸序列 ;

e. 编码与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ;和,

f. 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

119. 一种产生重组微生物的方法,包括对微生物细胞进行遗传修饰,以表达含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的至少一种重组核酸分子,其中所述 PUFA PKS 系统的所述至少一个结构域由选自下组的核酸序列编码 :

a. 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列 ;

b. 编码来自以下述方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列 :

(i) 选择产生至少一种 PUFA 的微生物 ;和,

(ii) 从 (i) 中鉴定与发酵培养基中在大于约 5% 饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物 ;

c. 编码选自 :SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6,或其生物学活性片段的氨基酸序列的核酸序列 ;

d. 编码选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, SEQ ID NO :32,或其生物学活性片段的氨基酸序列的核酸序列 ;

e. 编码与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ;和,

f. 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

120. 一种被修饰成表达多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的重组宿主细胞, 其中所述 PKS 催化重复的和非重复的酶反应, 且其中所述 PUFA PKS 系统包含:

- a. 至少两个烯酰 ACP- 还原酶 (ER) 结构域;
- b. 至少 6 个酰基载体蛋白 (ACP) 结构域;
- c. 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域;
- d. 至少一个酰基转移酶 (AT) 结构域;
- e. 至少一个酮还原酶 (KR) 结构域;
- f. 至少两个 FabA- 样 β -羟酰-ACP 脱水酶 (DH) 结构域;
- g. 至少一个链长因子 (CLF) 结构域; 和
- h. 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。

121. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统是真核 PUFA PKS 系统。

122. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统是藻类 PUFA PKS 系统。

123. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统是 Thraustochytriales PUFA PKS 系统。

124. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统是 Schizochytrium PUFA PKS 系统。

125. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统是破囊壶菌属 PUFAPKS 系统。

126. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统可在原核宿主细胞中表达。

127. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统可在真核宿主细胞中表达。

128. 权利要求 120 的重组宿主细胞, 其中所述宿主细胞是植物细胞。

129. 一种生产含有至少一种 PUFA 的产物的方法, 包括在有效产生该产物的条件下生长含有权利要求 128 的植物细胞的植物。

130. 权利要求 120 的重组宿主细胞, 其中所述宿主细胞是微生物细胞。

131. 一种生产含有至少一种 PUFA 的产物的方法, 包括在有效产生所述产物的条件下培养含有权利要求 130 的所述微生物细胞的培养物。

132. 权利要求 120 的重组宿主细胞, 其中所述 PUFA PKS 系统催化甘油三酯的直接产生。

133. 一种含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的遗传修饰的微生物, 其中所述 PKS 催化重复的和非重复的酶反应, 且其中所述 PUFAPKS 系统包含:

- a. 至少两个烯酰 ACP- 还原酶 (ER) 结构域;
- b. 至少 6 个酰基载体蛋白 (ACP) 结构域;
- c. 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域;
- d. 至少一个酰基转移酶 (AT) 结构域;
- e. 至少一个酮还原酶 (KR) 结构域;
- f. 至少两个 FabA- 样 β -羟酰-ACP 脱水酶 (DH) 结构域;
- g. 至少一个链长因子 (CLF) 结构域; 和

h. 至少一个丙二酰 -CoA:ACP 酰基转移酶 (MAT) 结构域 ;
其中该遗传修饰影响所述 PUFA PKS 系统的活性。

134. 权利要求 133 的遗传修饰的微生物,其中所述微生物是真核微生物。

135. 一种被修饰成表达非细菌多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的重组宿主细胞,其中所述非细菌 PUFA PKS 催化重复的和非重复的酶反应,且其中所述非细菌 PUFA PKS 系统包含:

- a. 至少一个烯酰 ACP- 还原酶 (ER) 结构域 ;
- b. 多个酰基载体蛋白 (ACP) 结构域 ;
- c. 至少两个 β - 酮脂酰 -ACP 合酶 (KS) 结构域 ;
- d. 至少一个酰基转移酶 (AT) 结构域 ;
- e. 至少一个酮还原酶 (KR) 结构域 ;
- f. 至少两个 FabA- 样 β - 羟酰 -ACP 脱水酶 (DH) 结构域 ;
- g. 至少一个链长因子 (CLF) 结构域 ;和
- h. 至少一个丙二酰 -CoA:ACP 酰基转移酶 (MAT) 结构域。

PUFA 聚酮化合物合酶系统及其用途

[0001] 本申请是基于申请号为 02811749.2, 申请日为 2002 年 4 月 16 日, 发明名称为“PUFA 聚酮化合物合酶系统及其用途”的发明申请的分案申请。

技术领域

[0002] 本发明涉及来自微生物, 包括诸如 *Thraustochytrid* 微生物的真核生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统。更具体地说, 本发明涉及编码非细菌 PUFA PKS 系统的核酸, 涉及非细菌 PUFA PKS 系统, 涉及含有非细菌 PUFA PKS 系统的遗传修饰的生物体, 且涉及制备和使用本文公开的非细菌 PUFA PKS 系统的方法。本发明还涉及鉴定含有 PUFA PKS 系统的细菌和非细菌微生物的方法。

背景技术

[0003] 聚酮化合物合酶 (PKS) 系统是本领域普遍已知的脂肪酸合酶 (FAS) 系统衍生的酶复合物, 但它通常被高度修饰以产生一般与脂肪酸有很小相似性的特异性产物。研究人员尝试了开发在文献中描述过的聚酮化合物合酶 (PKS) 系统, 它们属于一般称为 II 型, I 型和模块的三种基本类型之一。II 型系统的特征是可分离的蛋白质, 各蛋白质完成不同的酶促反应。酶协调作用以产生最终产物且该系统中的每个酶在最终产物的生产中一般参与几次。这类系统以类似于在植物和细菌中发现的脂肪酸合酶 (FAS) 系统的方式操作。I 型 PKS 系统与 II 型系统的相似之处在于酶以反复的方式用于生产最终产物。I 型与 II 型的区别在于酶活性发生在更大的蛋白质结构域中, 而不是与可分离的蛋白质相关。该系统类似于在动物和真菌中发现的 I 型 FAS 系统。

[0004] 与 I 型和 II 型系统相反, 在模块 PKS 系统中, 各酶结构域在最终产物的生产中只使用一次。该结构域在极大型蛋白质中发现且各反应的产物传递到 PKS 蛋白质的另一结构域上。另外, 在上述所有 PKS 系统中, 如果在最终产物中引入碳-碳双键, 通常为反式构型。

[0005] 在上述 I 型和 II 型 PKS 系统中, 每次循环进行同一组反应直到获得最终产物。在生物合成过程期间不允许导入特有反应。模块 PKS 系统需要的巨大蛋白质在反复的反应中不能节约利用 (即, 每次反应需要不同的结构域)。另外, 如上所述, 在所有以前所述 PKS 系统中以反式构型导入碳-碳双键。

[0006] 多不饱和脂肪酸 (PUFA) 是大多数真核生物中膜脂类的关键成份 (Lauritzen 等, *Prog. Lipid Res.* 40, 1(2001); McConnell 等, *Plant J.* 15, 521(1998)) 且是某些激素和信号分子的前体 (Heller 等, *Drugs* 55, 487(1998); Creelman 等, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48, 355(1997))。已知的 PUFA 合成途径包含通过延长和需氧去饱和反应加工脂肪酸合酶 (FAS) 产生的饱和 16:0 或 18:0 脂肪酸 (缩写 X:Y 表示含有 X 个碳原子和 Y 个顺式双键的酰基基团; PUFA 的双键位置相对于脂肪酸链的甲基碳 ($\omega 3$ 或 $\omega 6$) 用双键的系统亚甲基中断表示) (Sprecher, *Curr. Opin. Clin. Nutr. Metab. Care* 2, 135(1999); Parker-Barnes 等, *Proc. Natl. Acad. Sci. USA* 97, 8284(2000); Shanklin 等, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 611(1998))。从乙酰-CoA 开始, DHA 的合成需要约 30

个不同的酶活性和几乎 70 个反应,包括 4 个脂肪酸合成循环的重复步骤。聚酮化合物合酶 (PKSs) 完成一些与 FAS 相同的反应 (Hopwood 等, *Annu. Rev. Genet.* 24, 37 (1990) ;Bentley 等, *Annu. Rev. Microbiol.* 53, 411 (1999)) 且使用相同的小蛋白 (或结构域),即酰基载体蛋白 (ACP),作为生长碳链的共价连接位点。然而,在这些酶系统中,通常省去了 FAS 中所见的还原,脱水和还原的整个循环,从而产生高度衍生化的碳链,一般含有许多酮基和羟基以及反式构型的碳-碳双键。PKSs 的线型产物通常环化以形成复杂的生化试剂,包括抗生素和许多其它次级产物 (Hopwood 等, (1990) 出处同上 ;Bentley 等, (1999), 出处同上 ;Keating 等, *Curr. Opin. Chem. Biol.* 3, 598 (1999))。

[0007] 从包括希瓦氏菌属 (*Shewanella*) 种类海洋细菌的一些物种中已经报导了诸如二十二碳六烯酸 (DHA ;22:6 ω 3) 和二十碳五烯酸 (EPA ;20:5 ω 3) 的极长链 PUFA (Nichols 等, *Curr. Op. Biotechnol.* 10, 240 (1999) ;Yazawa, *Lipids* 31, S (1996) ;DeLong 等, *Appl. Environ. Microbiol.* 51, 730 (1986))。对来自希瓦氏菌属种类菌株 SCRC2738 的基因组片段 (克隆为质粒 pEPA) 的分析导致鉴定了 5 个可读框 (Orfs),总计 20Kb,它们是大肠杆菌中生产 EPA 的必要和充分条件 (Yazawa, (1996), 出处同上)。一些预测的蛋白结构域是 FAS 酶的同源物,而其它区域与已知功能的蛋白质无同源性。根据这些观察和生化研究,表明希瓦氏菌属中的 PUFA 合成包含 FAS 产生的 16- 或 18- 碳脂肪酸的延长和通过不明确的需氧去饱和酶插入双键 (Watanabe 等, *J. Biochem.* 122, 467 (1997))。对 5 个希瓦氏菌属 Orfs 编码的蛋白质序列的再检查得出了该假设并不正确的认识。5 个 Orfs 内的至少 11 个区域可鉴定为推定的酶结构域 (参见 Metz 等, *Science* 293 :290-293 (2001))。与基因数据库中的序列比较时,其中 7 个与 PKS 蛋白比与 FAS 蛋白相关性更强。该组中包含的结构域推定为编码丙二酰-CoA:ACP 酰基转移酶 (MAT), 3- 酮脂酰-ACP 合酶 (KS), 3- 酮脂酰-ACP 还原酶 (KR), 酰基转移酶 (AT), 磷酸泛酰巯基乙胺转移酶,链长 (或链起始) 因子 (CLF) 和非常罕见的 6 个 ACP 结构域簇 (即,存在两个以上聚集的 ACP 结构域在 PKS 或 FAS 序列中以前未见报导)。然而,三个区域与细菌 FAS 蛋白同源性更高。其中一个类似于最近描述的来自肺炎链球菌 (*Streptococcus pneumoniae*) 的三氯苯氧氯酚抗性烯酰还原酶 (ER) (Heath 等, *Nature* 406, 145 (2000) ;使用 LALIGN 程序 (模型, BLOSUM50 ;间隔缺口罚分, -10 ;延长罚分, -1) 比较 ORF8 肽与肺炎链球菌烯酰还原酶在 386aa 的重叠区内显示出 49% 的相似性)。两个区域是 fabA 编码的大肠杆菌 FAS 蛋白的同源物,它催化反式-2- 癸烯酰-ACP 的合成和该产物向顺式-3- 癸烯酰-ACP 的可逆异构化 (Heath 等, *J. Biol. Chem.* , 271, 27795 (1996))。因此,很可能希瓦氏菌属的 EPA 中的至少一些双键由 Orf7 中的 FabA- 样结构域催化的脱水酶-异构酶机制引入。

[0008] 厌氧生长的含有 pEPA 质粒的大肠杆菌细胞与需氧培养物积累相同水平的 EPA (Metz 等, 2001, 出处同上),表明在 EPA 合成中不涉及氧依赖型去饱和酶。当 pEPA 导入不能合成单不饱和脂肪酸且生长需要不饱和脂肪酸的大肠杆菌 fabB⁻ 突变体时,所得的细胞丧失其脂肪酸辅源营养。它们也比其它含有 pEPA 的菌株积累更高水平的 EPA,表明 EPA 与内源性产生的单不饱和脂肪酸竞争转移到甘油酯上。当含有 pEPA 的大肠杆菌细胞在 [¹³C]- 乙酸盐存在下生长时,对从细胞纯化的 EPA 进行 ¹³C-NMR 分析的数据证实了 EPA 的身份且提供了该脂肪酸从乙酰-CoA 和丙二酰-CoA 合成的证据 (参见 Metz 等, 2001, 出处同上)。来自含有 pEPA 的 fabB⁻ 细胞的无细胞匀浆从 [¹⁴C]- 丙二酰-CoA 合成 EPA 和饱和脂

肪酸。当匀浆分离成 200,000xg 高速沉淀和无膜的上清部分时,饱和脂肪酸合成局限于上清,与 II 型 FAS 酶的可溶性质一致 (Magnuson 等, Microbiol. Rev. 57, 522 (1993))。仅在高速沉淀部分中发现 EPA 的合成,表明 EPA 合成的发生可不依赖于大肠杆菌 FAS 酶或细胞质成份的可溶性中间物 (例如 16:0-ACP)。由于希瓦氏菌属 EPA 基因编码的蛋白质不特别疏水,因此 EPA 合成活性局限于该成分中反映了膜相联性酰基受体分子的要求。另外,与大肠杆菌 FAS 相反, EPA 合成是特异性的 NADPH- 依赖型且不需要 NADH。所有这些结果与编码多功能 PKS 的 pEPA 基因一致,该多功能 PKS 独立于 FAS, 延长酶, 和去饱和酶活性起作用以直接合成 EPA。很可能在希瓦氏菌属中鉴定的 PUFA 合成的 PKS 途径在海洋细菌中是普遍的。在 *Photobacterium profundum* (Allen 等, Appli. Environ. Microbiol. 65 :1710 (1999)) 和 *Moritella marina* (*Vibrio marinus*) (Tanaka 等, Biotechnol. Lett. 21 :939 (1999)) 中已经鉴定了与希瓦氏菌属基因簇高度同源的基因。

[0009] 对希瓦氏菌属进行的生化和分子遗传分析提供了聚酮化合物合酶能够从丙二酰-CoA 合成 PUFA 的确凿证据。由希瓦氏菌属 PKS 合成 EPA 的完整方案已经提出。与大肠杆菌 FabA 蛋白同源的蛋白质结构域的鉴定,和细菌 EPA 合成在厌氧条件下发生的观察结果提供了顺式双键的插入通过双功能脱水酶 /2- 反式, 3- 顺式异构酶 (DH/2, 3I) 的作用发生这一机制的证据。在大肠杆菌中, 3- 顺式酰基中间物与丙二酰-ACP 的缩合需要特定的酮脂酰-ACP 合酶且这支持在希瓦氏菌属基因簇中存在两个 KS (在 Orf5 和 Orf7 中) 的理论。然而, PKS 循环以两个碳的增量延长碳链,而在 EPA 产物中双键在每第三个碳处出现。如果通过 2- 反式, 2- 顺式异构化 (DH/2, 2I) 接着在延长的脂肪酸链中掺入顺式双键在 EPA 的 C-14 和 C-8 处产生双键就可解决这一差异。在例如, 视黄素类 (retinoid) 循环的 11- 顺式 - 视黄醛合成中已知会发生反式双键酶促转化成顺式构型而不发生键迁移 (Jang 等, J. Biol. Chem. 275, 28128 (2000))。尽管在希瓦氏菌属 PKS 中尚未鉴定这样的酶功能,但是它可能属于一个未鉴定的蛋白质结构域。

[0010] 希瓦氏菌属和另一海洋细菌 *Vibrio marinus* 中 PUFA 合成的 PKS 途径在美国专利号 6, 140, 486 (从 1998 年 6 月 4 日申请的, 发明名称为“通过在植物中表达聚酮化合物类合成基因生产多不饱和脂肪酸”的美国申请系列号 09/090, 793 出版, 在此引用以其整体作为参考) 中进行了详细描述。

[0011] 多不饱和脂肪酸 (PUFA) 据认为可用于营养, 制药, 工业, 和其它目的。来自天然来源和化学合成的 PUFA 的昂贵供应不足以满足商业需要。由于许多分离的去饱和酶和延长酶是从大多数植物物种中共有的亚油酸 (LA, 18:2 Δ 9, 12) 到更饱和且更长链 PUFA 的脂肪酸合成中所必需的, 因此改造植物宿主细胞以表达诸如 EPA 和 DHA 的 PUFA 可能需要表达 5 种或 6 种分离的酶活性以实现至少 EPA 和 DHA 的表达。另外, 为了生产可用量的该 PUFA, 可能需要其它的改造努力, 例如, 下调竞争底物的酶, 通过例如诱变改造成具有更高的酶活性或者将酶定向到质体细胞器上。因此有利的是从天然产生这些脂肪酸的物种获得包含 PUFA 生物合成的遗传材料并单独或与异源系统结合表达可改造成允许生产商品量的 PUFA 的该分离材料。

[0012] 在诸如希瓦氏菌属和 *Vibrio marinus* 的海洋细菌中发现的 PUFA PKS 系统 (参见美国专利号 6, 140, 486, 出处同上) 为商品 PUFA 生产的新方法提供了资源。然而, 这些海洋细菌具有的缺陷最终将限制其在商业水平上的利用。首先, 尽管美国专利号 6, 140, 486 公

开了海洋细菌 PUFA PKS 系统可用于遗传修饰植物,但是海洋细菌在寒冷的海洋环境中自然生活和生长且这些细菌的酶系统在 30℃ 以上不能很好地发挥功能。相反,许多农作物植物作为使用 PUFA PKS 系统进行遗传改造的有吸引力的目标在 30℃ 以上和变动到高于 40℃ 的温度下为正常的生长条件。因此,海洋细菌 PUFA PKS 系统预期不容易适应正常生长条件下的植物表达。而且海洋细菌 PUFA PKS 基因由于是细菌来源,可能与真核宿主细胞的基因组不相容,或者至少需要大量的修改以便在真核宿主中起作用。另外,已知的海洋细菌 PUFA PKS 系统不能直接产生甘油三酯,而直接产生甘油三酯是所期望的,因为甘油三酯是微生物中的脂类储存产物且因此可在微生物/植物细胞中以极高的水平(例如,高达 80-85% 的细胞重量)积累(与一般仅以低水平(例如最大值不超过细胞重量的 10-15%)积累的“结构”脂产物(例如磷脂)相反)。

[0013] 因此,本领域需要具有更大适应性的其它 PUFAPKS 系统用于商业用途。

[0014] 发明简述

[0015] 本发明的一个实施方案涉及含有选自如下的核酸序列的分离核酸分子:(a) 编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6,或其生物学活性片段的氨基酸序列的核酸序列;(b) 编码选自 SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物学活性片段的氨基酸序列的核酸序列;(c) 编码与(a)的氨基酸序列中至少 500 个连续氨基酸有至少约 60% 相同的氨基酸序列的核酸序列,其中该氨基酸序列具有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的生物学活性;(d) 编码与(b)的氨基酸序列至少约 60% 相同的氨基酸序列的核酸序列,其中该氨基酸序列具有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的生物学活性;和(e)与(a),(b),(c),或(d)的核酸序列完全互补的核酸序列。在另一方面,该核酸序列编码与(1)选自由 SEQ ID NO:2, SEQ ID NO:4,和 SEQ ID NO:6 组成的组中的氨基酸序列的至少 500 个连续氨基酸至少约 70% 相同,或至少约 80% 相同,或至少约 90% 相同,或相同的氨基酸序列;和/或(2)编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的氨基酸序列至少约 70% 相同的氨基酸序列的核酸序列。在一个优选的实施方案中,该核酸序列编码选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 和/或其生物学活性片段的氨基酸序列。在一个方面,该核酸序列选自:SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29,和 SEQ ID NO:31。

[0016] 本发明的另一实施方案涉及含有与至少一个转录调控序列可操作地相连的上述核酸分子的重组核酸分子。在另一实施方案中,本发明涉及直接用上述重组核酸分子转染的重组细胞。

[0017] 本发明的另一实施方案还涉及遗传修饰的微生物,其中该微生物表达含有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个生物学活性结构域的

PKS 系统。PUFAPKS 系统的至少一个结构域由选自如下的核酸序列编码：(a) 编码来自 *Thraustochytrid* 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列；(b) 编码来自本发明的筛选方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列；(c) 编码选自：SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, 和其生物学活性片段组成的组中的氨基酸序列的核酸序列；(d) 编码选自：SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, 或其生物学活性片段的氨基酸序列的核酸序列；(e) 编码与选自：SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列；其中该氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性；和, (f) 编码与选自：SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列；其中该氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。在该实施方案中, 遗传修饰该微生物以影响 PKS 系统的活性。在上文 (b) 中提到的本发明的筛选方法包括：(i) 选择产生至少一种 PUFA 的微生物；和, (ii) 从 (i) 中鉴定与发酵培养基中在大于 5% 饱和度, 且更优选 10% 的饱和度, 且更优选大于 15% 的饱和度, 且更优选大于 20% 的饱和度的溶氧条件下微生物产生的 PUFA 相比, 在发酵培养基中不足约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。

[0018] 在一个方面, 该微生物内源性表达含有 PUFA PKS 系统的至少一个结构域的 PKS 系统, 且其中在编码 PUFA PKS 系统的至少一个结构域的核酸序列中存在遗传修饰。例如, 在编码具有至少一种下列蛋白质的生物学活性的结构域的核酸序列中存在遗传修饰：丙二酰-CoA:ACP 酰基转移酶 (MAT), β -酮脂酰-ACP 合酶 (KS), 酮还原酶 (KR), 酰基转移酶 (AT), FabA-样 β -羟酰-ACP 脱水酶 (DH), 磷酸泛酰巯基乙胺转移酶, 链长因子 (CLF), 酰基载体蛋白 (ACP), 烯酰 ACP-还原酶 (ER), 催化反式-2-癸烯酰-ACP 合成的酶, 催化反式-2-癸烯酰-ACP 向顺式-3-癸烯酰-ACP 可逆异构化的酶, 和催化顺式-3-癸烯酰-ACP 延长成顺式-11-十八碳烯酸的酶。在一个方面, 在编码选自下组的氨基酸序列的核酸序列中存在遗传修饰：(a) 与选自：SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸具有至少约 70% 的相同性, 且优选至少约 80% 的相同性, 且更优选至少约 90% 的相同性且更优选相同的氨基酸序列；其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性；和, (b) 与选自：SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32 的氨基酸序列具有至少约 70% 相同性, 且优选至少约 80% 相同性, 且更优选至少约 90% 相同性且更优选相同的氨基酸序列；其中该氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性。

[0019] 在一个方面, 该遗传修饰的微生物是 *Thraustochytrid*, 它包括, 但不限于, 选自 *Schizochytrium* 和破囊壶菌属 (*Thraustochytrium*) 中的 *Thraustochytrid*。在另一方面, 该微生物进一步被遗传修饰成重组表达编码来自细菌 PUFAPKS 系统, I 型 PKS 系统, II 型 PKS 系统, 和 / 或模块 PKS 系统的至少一个生物学活性结构域的至少一个核酸分子。

[0020] 在该实施方案的另一面, 该微生物内源性表达含有 PUFA PKS 系统的至少一个生

物学活性结构域的 PUFA PKS 系统,且其中该遗传修饰包含表达选自自由编码来自第二种 PKS 系统的至少一个生物学活性结构域的重组核酸分子和编码影响 PUFA PKS 系统活性的蛋白质的重组核酸分子组成的组中的重组核酸分子。优选的是,该重组核酸分子包含任一上述核酸序列。

[0021] 在该实施方案的一个方面中,该重组核酸分子编码磷酸泛酰巯基乙胺转移酶。在另一方面,该重组核酸分子包含编码来自细菌 PUFA PKS 系统, I 型 PKS 系统, II 型 PKS 系统,和 / 或模块 PKS 系统的至少一个生物学活性结构域的核酸序列。

[0022] 在该实施方案的另一方面,通过用编码多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的重组核酸分子转染遗传修饰该微生物。该重组核酸分子可包括含有任一上述核酸序列的任意重组核酸分子。在一个方面,该微生物进一步被遗传修饰成重组表达编码来自细菌 PUFA PKS 系统, I 型 PKS 系统, II 型 PKS 系统,或模块 PKS 系统的至少一个生物学活性结构域的至少一个核酸分子。

[0023] 本发明的另一实施方案还涉及遗传修饰的植物,其中该植物被遗传修饰成重组表达含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个生物学活性结构域的 PKS 系统。该结构域可由任一上述核酸序列编码。在一个方面,该植物进一步被遗传修饰成重组表达编码来自细菌 PUFA PKS 系统, I 型 PKS 系统, II 型 PKS 系统,和 / 或模块 PKS 系统的至少一个生物学活性结构域的至少一个核酸分子。

[0024] 本发明的另一实施方案涉及鉴定具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的微生物的方法。该方法包括如下步骤:(a) 选择产生至少一种 PUFA 的微生物;和, (b) 从 (a) 中鉴定与发酵培养基中在大于 5% 饱和度的溶氧条件下,更优选 10% 的饱和度,更优选大于 15% 的饱和度,且更优选大于 20% 的饱和度时微生物产生的 PUFA 相比,在发酵培养基中不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。产生至少一种 PUFA 且在不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物鉴定为含有 PUFA PKS 系统的候选微生物。

[0025] 在该实施方案的一个方面,步骤 (b) 包含从 (a) 中鉴定在不超过约 2% 饱和度的溶氧条件下,且更优选在不超过约 1% 饱和度的溶氧条件下,甚至更优选在约 0% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。

[0026] 在该实施方案的另一方面,在 (a) 中选择的微生物具有通过吞噬作用消费细菌的能力。在另一方面,在 (a) 中选择的微生物具有简单的脂肪酸分布特征 (profile)。在另一方面,在 (a) 中选择的微生物是非细菌微生物。在另一方面,在 (a) 中选择的微生物是真核生物。在另一方面,在 (a) 中选择的微生物是 Thraustochytriales 目的一个成员。在另一方面,在 (a) 中选择的微生物在高于约 15°C,且优选高于约 20°C,且更优选高于约 25°C,且甚至更优选高于约 30°C 的温度下具有生产 PUFA 的能力。在另一方面,在 (a) 中选择的微生物具有以高于 5% 的生物体干重,且更优选高于 10% 的生物体干重生产所需生物活性化合物 (例如,脂类) 的能力。在还有另一方面,在 (a) 中选择的微生物其总脂肪酸高于 30% 是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 3 个或更多个不饱和键的长链脂肪酸,且优选的是,在 (a) 中选择的微生物其总脂肪酸高于 40% 是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 3 个或更多个不饱和键的长链脂肪酸。在另一方面,在 (a) 中选择的微生物其总脂肪酸高于 30% 是 C14:0, C16:0 和 C16:1,同时也产生至少一种具有 4 个或更多个

不饱和键的长链脂肪酸,且更优选同时也产生至少一种具有 5 个或更多个不饱和键的长链脂肪酸。

[0027] 在该实施方案的另一方面,该方法还包含检测生物体是否含有 PUFAPKS 系统的步骤 (c)。在该方面,该检测步骤可包括检测该微生物中与编码来自 *Thraustochytrid* PUFA PKS 系统的氨基酸序列的核酸序列在严格条件下杂交的核酸序列。另外,该检测步骤可包括检测生物体中用来自 *Thraustochytrid* PUFAPKS 系统的核酸序列的寡核苷酸引物扩增的核酸序列。

[0028] 本发明的另一实施方案涉及由上述筛选方法鉴定的微生物,其中该微生物被遗传修饰成通过 PUFA PKS 系统调节分子的生产。

[0029] 本发明还有另一实施方案涉及生产生物活性分子的方法,该生物活性分子由聚酮化合物合酶系统产生。该方法包括在有效生产生物活性分子的条件培养遗传修饰的生物体的步骤,该生物体表达含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个生物活性结构域的 PKS 系统。PUFA PKS 系统的结构域由任一上述核酸序列编码。

[0030] 在该实施方案的一个方面,该生物体内源性表达含有 PUFA PKS 系统的至少一个结构域的 PKS 系统,且该遗传修饰在编码 PUFA PKS 系统的至少一个结构域的核酸序列上。例如,该遗传修饰与野生型生物体相比可改变至少一种由内源性 PKS 系统产生的产物。

[0031] 在该实施方案的另一方面,该生物体内源性表达含有 PUFA PKS 系统的至少一个生物活性结构域的 PKS 系统,且该遗传修饰包括用选自:编码来自第二种 PKS 系统的至少一个生物学活性结构域的重组核酸分子和编码影响 PUFA PKS 系统活性的蛋白质的重组核酸分子组成的组中的重组核酸分子转染该生物体。例如,该遗传修饰与野生型生物体相比可改变由内源性 PKS 系统产生的至少一种产物。

[0032] 在该实施方案的还有另一方面中,通过用编码多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的重组核酸分子转染遗传修饰该生物体。在另一方面,该生物体产生的多不饱和脂肪酸 (PUFA) 分布特征不同于没有遗传修饰的天然生物体。在另一方面,该生物体内源性表达非细菌 PUFA PKS 系统,且其中该遗传修饰包括来自不同 PKS 系统的结构域取代编码非细菌 PUFA PKS 系统的至少一个结构域的核酸序列。

[0033] 在还有另一方面,该生物体内源性表达非细菌 PUFA PKS 系统,该系统通过用重组核酸分子转染该生物体而被修饰,该重组核酸分子编码的蛋白质调节 PUFA PKS 系统产生的脂肪酸的链长。例如,编码调节脂肪酸链长的蛋白质的该重组核酸分子可取代非细菌 PUFA PKS 系统中编码链长因子的核酸序列。在另一方面,调节 PUFA PKS 系统产生的脂肪酸链长的蛋白质是链长因子。在另一方面,调节 PUFA PKS 系统产生的脂肪酸链长的蛋白质是指导合成 C20 单位的链长因子。

[0034] 在一个方面,该生物体表达在选自编码 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域的区域和编码 β -酮脂酰-ACP 合酶 (KS) 的区域中含有遗传修饰的非细菌 PUFA PKS 系统,其中该修饰与没有该修饰相比改变了 PUFA PKS 系统产生的长链脂肪酸的比率。在一个方面,该修饰包含用不具有异构化活性的 DH 结构域取代非细菌 PUFA PKS 系统中的 FabA-样 β -羟酰-ACP 脱水酶 (DH)。在另一方面,该修饰选自缺失全部或部分该区域,用来自不同生物的同源区域取代该区域,和突变该区域组成的组中。

[0035] 在另一方面,该生物体表达 PKS 系统且该遗传修饰包含用来自 PUFAPKS 系统的

FabA-样 β -羟酰-ACP 脱水酶 (DH) 区域取代不具有异构化活性的 DH 区域。

[0036] 在另一方面,该生物体表达在烯酰-ACP 还原酶 (ER) 区域中含有修饰的非细菌 PUFA PKS 系统,其中该修饰导致与没有修饰相比产生不同的化合物。例如,该修饰可选自由缺失全部或部分 ER 区域,用来自不同生物体的 ER 区域取代该 ER 区域,和突变该 ER 区域组成的组中。

[0037] 在一个方面,由本方法产生的生物活性分子可包括,但不限于抗炎症配制品,化疗剂,有效赋形剂,骨质疏松症药物,抗抑郁剂,抗惊厥剂,抗幽门螺杆菌 (*Helicobacter pylori*) 药物,治疗神经变性疾病的药物,治疗变性肝脏疾病的药物,抗生素,和降胆固醇配制品。在一个方面,该生物活性分子是多不饱和脂肪酸 (PUFA)。在另一方面,该生物活性分子是包含顺式构型的碳-碳双键的分子。在另一方面,该生物活性分子是在每第三个碳包含一个双键的分子。

[0038] 在该实施方案的一个方面,该生物体是微生物,在另一方面,该生物体是植物。

[0039] 本发明的另一实施方案涉及产生一种植物的方法,该植物的多不饱和脂肪酸 (PUFA) 分布特征不同于天然植物,该方法包括遗传修饰该植物的细胞以表达含有至少一种重组核酸分子的 PKS 系统,该重组核酸分子含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列。该 PUFA PKS 系统的结构域由任一上述核酸序列编码。

[0040] 本发明还有另一实施方案涉及修饰含有至少一种脂肪酸的终产物的方法,包括向终产物中加入由重组宿主细胞产生的油脂,该重组宿主细胞表达至少一种重组核酸分子,该重组核酸分子含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列。PUFA PKS 系统的结构域由任一上述核酸序列编码。在一个方面,该终产物选自饮食添加剂,食品,药物配制品,人源化动物乳汁,和婴儿配制品组成的组中。药物配制品可包括,但不限于:抗炎症配制品,化疗剂,有效赋形剂,骨质疏松症药物,抗抑郁剂,抗惊厥剂,抗幽门螺杆菌药物,治疗神经变性疾病的药物,治疗变性肝脏疾病的药物,抗生素,和降胆固醇配制品。在一个方面,该终产物可用于治疗选自:慢性炎症,急性炎症,胃肠疾病,癌症,恶病质,心脏再狭窄,神经变性疾病,肝脏变性疾病,血脂疾病,骨质疏松症,骨关节炎,自身免疫疾病,先兆子痫,早产 (preterm birth),年老相关性黄斑病,肺病,和过氧化物酶体疾病组成的组中的一种症状。

[0041] 本发明还有另一实施方案涉及生产人源化动物乳汁的方法,包括用含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的至少一个重组核酸分子遗传修饰产奶动物的产奶细胞。该 PUFA PKS 系统的结构域由任一上述核酸序列编码。

[0042] 本发明还有另一实施方案涉及产生重组微生物的方法,包括遗传修饰微生物细胞以表达含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的至少一种重组核酸分子。该 PUFA PKS 系统的结构域由任一上述核酸序列编码。

[0043] 本发明还有另一实施方案涉及被修饰成表达多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的重组宿主细胞,其中该 PKS 催化重复的和非重复的酶反应。PUFA PKS 系统包含:(a) 至少两个烯酰 ACP-还原酶 (ER) 结构域;(b) 至少 6 个酰基载体蛋白 (ACP) 结构域;(c) 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域;(d) 至少一个酰基转移酶 (AT) 结构域;(e) 至少一个酮还原酶 (KR) 结构域;(f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域;(g) 至少一个链长因子 (CLF) 结构域;和 (h) 至少一个丙二酰-CoA:ACP 酰基转移

酶 (MAT) 结构域。在一个实施方案中, PUFA PKS 系统是真核 PUFA PKS 系统。在另一方面, PUFA PKS 系统是藻类 PUFA PKS 系统, 且优选 Thraustochytriales PUFAPKS 系统, 它包括, 但不限于 Schizochytrium PUFA PKS 系统或破囊壶菌属 PUFA PKS 系统。

[0044] 在该实施方案中, PUFA PKS 系统可在原核宿主细胞或在真核宿主细胞中表达。在一个方面, 该宿主细胞是植物细胞。因此, 本发明的一个实施方案是生产含有至少一种 PUFA 的产物的方法, 包括在有效产生该产物的条件下生长含有该植物细胞的植物。宿主细胞可以是微生物且在这种情况下, 本发明的一个实施方案是生产含有至少一种 PUFA 的产物的方法, 包括在有效产生该产物的条件下培养含有该微生物细胞的培养物。在一个方面, PKS 系统催化甘油三酯的直接生产。

[0045] 本发明还有另一实施方案涉及遗传修饰的微生物, 它含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统, 其中该 PKS 催化重复的和非重复的酶反应。PUFA PKS 系统包含: (a) 至少两个烯酰 ACP-还原酶 (ER) 结构域; (b) 至少 6 个酰基载体蛋白 (ACP) 结构域; (c) 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域; (d) 至少一个酰基转移酶 (AT) 结构域; (e) 至少一个酮还原酶 (KR) 结构域; (f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域; (g) 至少一个链长因子 (CLF) 结构域; 和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。该遗传修饰影响 PUFAPKS 系统的活性。在该实施方案的一个方面, 该微生物是真核微生物。

[0046] 本发明还有另一实施方案涉及被修饰成表达非细菌多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的重组宿主细胞, 其中该非细菌 PUFA PKS 催化重复的和非重复的酶反应。该非细菌 PUFA PKS 系统包含: (a) 至少一个烯酰 ACP-还原酶 (ER) 结构域; (b) 多个酰基载体蛋白 (ACP) 结构域; (c) 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域; (d) 至少一个酰基转移酶 (AT) 结构域; (e) 至少一个酮还原酶 (KR) 结构域; (f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域; (g) 至少一个链长因子 (CLF) 结构域; 和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。

附图说明

[0047] 图 1 是 Schizochytrium PUFA PKS 系统的结构域结构示意图。

[0048] 图 2 显示了来自 Schizochytrium 和希瓦氏菌属的 PKS 结构域的比较。

[0049] 图 3 显示了来自 Schizochytrium 的 PKS 结构域和来自念珠藻属 (Nostoc) 的其产物是不含任何双键的长链脂肪酸的相关 PKS 系统的比较。

[0050] 发明详述

[0051] 本发明一般涉及非细菌衍生的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统, 一般涉及含有非细菌 PUFA PKS 系统的遗传修饰的生物体, 涉及制备和使用该系统用于生产包括生物活性分子的目的产物的方法, 涉及鉴定具有该 PUFA PKS 系统的新真核微生物的新方法。本文所用的 PUFA PKS 系统一般具有下列鉴定特征: (1) 作为该系统的天然产物产生 PUFA; 和 (2) 它含有组装进复合物的几个多功能蛋白质, 该复合物同时进行脂肪酸链的反复加工和非反复加工, 包括在选定循环中的反式-顺式异构化和烯酰还原反应 (例如, 参见图 1)。

[0052] 更具体地说, 首先, 构成本发明基础的 PUFA PKS 系统产生多不饱和脂肪酸 (PUFA)

作为产物（即，内源性（天然）含有该 PKS 系统的生物体使用该系统制备 PUFA）。本文提及的 PUFA 优选是具有至少 16 碳的碳链长度的多不饱和脂肪酸，且更优选至少 18 碳，且更优选至少 20 碳，且更优选 22 或更多个碳，且具有至少 3 个或更多个双键，优选 4 个或更多个，且更优选 5 个或更多个，且甚至更优选 6 个或更多个双键，其中所有双键是顺式构型。本发明的一个目的是通过遗传操作或终产物的改造发现或创造产生具有所需链长和所需数目双键的多不饱和脂肪酸的 PKS 系统。PUFA 的例子包括，但不限于，DHA（二十二碳六烯酸（22:6， ω -3）），DPA（鲱油酸（C22:5， ω -6）），和 EPA（二十碳五烯酸（C20:5， ω -3））。

[0053] 其次，本文所述 PUFA PKS 系统同时参与反复和非反复的反应，使得该系统不同于以前所述 PKS 系统（例如，I 型，II 型或模块型系统）。更具体地说，本文所述 PUFA PKS 系统含有在每次循环期间显示功能的结构域以及仅在某些循环中显示功能的结构域。该系统的一个关键方面涉及与细菌 Fab A 酶表现出同源性的结构域。例如，大肠杆菌的 Fab A 酶表现出具有两种酶活性。它具有从含有羟基的碳链减去一个水分子（H₂O），在该碳链中留下一个反式双键的脱水活性。另外，它具有将反式双键转变成顺式构型的异构酶活性。该异构化与双键位置迁移到相邻碳结合完成。在 PKS（和 FAS）系统中，主碳链以 2 个碳的增量延长。因此可预知这些 PKS 系统产生 PUFA 产物所需的延长反应的次数。例如，产生 DHA（C22:6，均为顺式）需要 10 次延长反应。由于在终产物中仅有 6 个双键，这意味着在某些反应循环期间，双键被保留（作为顺式异构体），而在其它循环期间，双键在下一步延长前被还原。

[0054] 在发现海洋细菌的 PUFA PKS 系统（参见美国专利号 6,140,486）前，还不了解 PKS 系统具有这种反复性和选择性酶反应的组合，且认为它们不能以顺式构型产生碳-碳双键。然而，本发明所述 PUFA PKS 系统具有导入顺式双键的能力和改变循环中反应顺序的能力。

[0055] 因此，本发明人提出使用 PUFA PKS 系统的这些特征生产以前所述（II 型，I 型和模块型）PKS 系统不能产生的各种生物活性分子。这些生物活性分子包括，但不限于，多不饱和脂肪酸（PUFA），抗生素或其它生物活性化合物，其中许多分子将在下文讨论。例如，使用本文所述 PUFA PKS 基因结构的知识，可使用许多方法中的任一种改变 PUFA PKS 基因，或者将这些基因的部分与包括其它 PKS 系统的其它合成系统结合以便产生新产物。该系统的特定类型同时进行反复性和选择性反应的内在能力使得该系统能够产生对 PKS 系统的其它类型采用相似方法不能发现的产物。

[0056] 在一个方面，根据本发明的 PUFA PKS 系统包含至少下列生物活性结构域：(a) 至少两个烯酰 ACP-还原酶（ER）结构域；(b) 至少 6 个酰基载体蛋白（ACP）结构域；(c) 至少两个 β -酮脂酰-ACP 合酶（KS）结构域；(d) 至少一个酰基转移酶（AT）结构域；(e) 至少一个酮还原酶（KR）结构域；(f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶（DH）结构域；(g) 至少一个链长因子（CLF）结构域；和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶（MAT）结构域。这些结构域的功能分别是本领域公知的且在下文关于本发明的 PUFA PKS 系统中将进行详细描述。

[0057] 在另一实施方案中，PUFA PKS 系统包含至少下列生物活性结构域：(a) 至少一个烯酰 ACP-还原酶（ER）结构域；(b) 多个酰基载体蛋白（ACP）结构域（至少 4 个，且优选至少 5 个，且更优选至少 6 个，甚至更优选 7,8,9, 或 9 个以上）；(c) 至少两个 β -酮脂酰-ACP 合酶（KS）结构域；(d) 至少一个酰基转移酶（AT）结构域；(e) 至少一个酮还原酶（KR）结构

域 ; (f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域 ; (g) 至少一个链长因子 (CLF) 结构域 ; 和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。优选的是, 该 PUFA PKS 系统是非细菌 PUFA-PKS 系统。

[0058] 在一个实施方案中, 本发明的 PUFA PKS 系统是非细菌 PUFA PKS 系统。换句话说, 在一个实施方案中, 本发明的 PUFA PKS 系统从不是细菌的生物体中分离, 或者是来自不是细菌的诸如真核生物或古细菌的生物体的 PUFAPKS 系统的同源物或衍生物。真核生物根据细胞的分化程度与原核生物区分开来。具有更多分化的高等群体称为真核生物。具有较少分化的细胞的低等其它称为原核生物。一般来说, 原核生物不具有核膜, 细胞分裂期间不表现为有丝分裂, 仅有一个染色体, 其细胞质含有 70S 核糖体, 它们不具有任何线粒体, 内质网, 叶绿体, 溶酶体, 或高尔基体, 其鞭毛 (如果存在) 由单一原纤维组成。相反真核生物具有核膜, 它们在细胞分裂期间表现为有丝分裂, 它们具有许多染色体, 其细胞质含有 80S 核糖体, 它们具有线粒体, 内质网, 叶绿体 (在藻类中), 溶酶体和高尔基体, 且其鞭毛 (如果存在) 由许多原纤维组成。一般来说, 细菌是原核生物, 而藻类, 真菌, 原生生物, 原生动物和高等植物是真核生物。海洋细菌 (例如, 希瓦氏菌属和 *Vibrio marinus*) 的 PUFAPKS 系统不是本发明的基础, 尽管本发明确实包含使用来自这些细菌 PUFAPKS 系统的结构域与本发明的非细菌 PUFA PKS 系统的结构域相结合。例如, 根据本发明, 可产生遗传修饰的生物体, 它掺入了非细菌 PUFAPKS 功能域与细菌 PUFA PKS 功能域, 以及来自其它 PKS 系统 (I 型, II 型, 模块型) 或 FAS 系统的 PKS 功能域或蛋白质。

[0059] *Schizochytrium* 是以 DHA 和鲱油酸 (DPA ; 22:5 ω -6) 积累大量三酰甘油, 例如, 占干重 30% 的 DHA+DPA 的 *Thraustochytrid* 海洋微生物 (Barclay 等, *J. Appl. Phycol.* 6, 123 (1994))。在通过延长 / 去饱和途径合成 20- 和 22- 碳 PUFA 的真核生物中, 18-, 20- 和 22- 碳中间体的分子库相当大, 因此使用 [^{14}C]- 乙酸盐的体内标记实验揭示了预期中间体的清楚前体-产物动力学 (Gellerman 等, *Biochim. Biophys. Acta* 573 :23 (1979))。另外, 给该生物体提供的外源性放射标记的中间体转变成最终的 PUFA 产物。本发明人显示了 [$1-^{14}\text{C}$]- 乙酸盐迅速被 *Schizochytrium* 细胞吸收且掺入脂肪酸, 但在最短标记时间 (1 分钟) 时, 在脂肪酸中回收到含有 31% 标记的 DHA, 且该百分数在 10-15 分钟的 [^{14}C]- 乙酸盐掺入和随后 24 小时的培养物生长期间基本上保持不变 (参见实施例 3)。同样, DPA 通过实验表现出 10% 的标记。没有证据表明在 16- 或 18- 碳脂肪酸与 22- 碳多不饱和脂肪酸之间存在前体-产物关系。这些结果与从包含极小 (很可能与酶结合的) 中间体库的 [^{14}C]- 乙酸盐迅速合成 DHA 一致。来自 *Schizochytrium* 培养物的无细胞匀浆将 [$1-^{14}\text{C}$]- 丙二酰-CoA 掺入 DHA, DPA, 和饱和脂肪酸中。相同的生物合成活性在 100,000xg 的上清成分中保留但在膜沉淀中不存在。因此, *Schizochytrium* 中的 DHA 和 DPA 合成不涉及类似于其它真核生物所述膜结合的去饱和酶或脂肪酸延长酶 (Parker-Barnes 等, 2000, 出处同上 ; Shanklin 等, 1998, 出处同上)。这些分级分离数据与从希瓦氏菌属酶获得的数据相反 (参见 Metz 等, 2001, 出处同上) 且表明 *Schizochytrium* 酶使用不同的 (可溶性) 酰基受体分子, 例如 CoA。

[0060] 在共同未决的美国申请系列号 09/231,899 中, 从 *Schizochytrium* 构建了 cDNA 文库且测序了约 8,000 个随机克隆 (ESTs)。在这些资料组中, 仅鉴定了一个中度表达的基因 (全部序列的 0.3%) 是脂肪酸去饱和酶, 尽管一个单一克隆 (0.01%) 代表第二个推定的

去饱和酶。相反,图 2 中所示的与希瓦氏菌属 PKS 基因的 11 个结构域中的 8 个表现出同源性的序列均以 0.2-0.5% 的频率被鉴定。在美国申请系列号 09/231,899 中,测序了与希瓦氏菌属 PKS 基因表现出同源性的几个 cDNA 克隆,且各种克隆装配成代表两个部分可读框和一个完整可读框的核酸序列。含有美国申请系列号 09/231,899 所述第一个部分可读框的 cDNA 序列的核苷酸 390-4443(其中称为 SEQ ID NO:69)与本文称为 OrfA 的序列(SEQ ID NO:1)的核苷酸 4677-8730(加上终止密码子)匹配。含有美国申请系列号 09/231,899 所述第二个部分可读框的 cDNA 序列的核苷酸 1-4876(其中称为 SEQ ID NO:71)与本文称为 OrfB 的序列(SEQ ID NO:3)的核苷酸 1311-6177(加上终止密码子)匹配。含有美国申请系列号 09/231,899 所述完全可读框的 cDNA 序列的核苷酸 145-4653(其中称为 SEQ ID NO:76 且错误地称为部分可读框)与本文称为 OrfC 的序列(SEQ ID NO:5)的整个序列(加上终止密码子)匹配。

[0061] 本发明人对 cDNA 和基因组克隆的进一步测序得以鉴定 OrfA, OrfB 和 OrfC 各自的全长基因组序列并完全鉴定了与希瓦氏菌属中的那些结构域同源的结构域(参见图 2)。应注意在 *Schizochytrium* 中,基因组 DNA 与 cDNA 相同,因为根据本发明人的了解,在该生物体基因组中没有内含子。因此,提及来自 *Schizochytrium* 的核苷酸序列时可以是指基因组 DNA 或 cDNA。根据 *Schizochytrium* PKS 结构域与希瓦氏菌属的比较,很明显 *Schizochytrium* 基因组编码与希瓦氏菌属中能够催化 EPA 合成的蛋白质高度相似的蛋白质。*Schizochytrium* 中的该蛋白质构成催化 DHA 和 DPA 合成的 PUFAPKS 系统。正如本文进行的详细讨论,对希瓦氏菌属鉴定的反应程序的简单修饰将允许在 *Schizochytrium* 中合成 DHA。原核希瓦氏菌属和真核 *Schizochytrium* 基因之间的同源性表明 PUFA PKS 进行横向(lateral)基因转移。

[0062] 图 1 是来自 *Schizochytrium* PUFA PKS 系统的三个可读框的示意图,且包括该 PUFAPKS 系统的结构域结构。正如下面实施例 1 所述,各可读框的结构域结构如下:

[0063] 可读框 A(OrfA):

[0064] OrfA 的完整核苷酸序列在本文中表示为 SEQ ID NO:1。SEQ ID NO:1 的核苷酸 4677-8730 相应于美国申请系列号 09/231,899 中称为 SEQ ID NO:69 的序列的核苷酸 390-4443。因此,SEQ ID NO:1 的核苷酸 1-4676 代表在美国申请系列号 09/231,899 中未公开的另一序列。SEQ ID NO:1 的这一新区域编码 OrfA 中的下列结构域:(1)ORFA-KS 结构域;(2)ORFA-MAT 结构域;和(3)至少部分 ACP 结构域的区域(例如,至少 ACP 结构域 1-4)。应注意美国申请系列号 09/231,899 中的 SEQ ID NO:69 的核苷酸 1-389 与本文公开的 SEQ ID NO:1 中位置 4677 上游的 389 个核苷酸不匹配。因此,美国申请系列号 09/231,899 中 SEQ ID NO:69 的位置 1-389 似乎错误地放在该序列的核苷酸 390-4443 之后。这些前 389 个核苷酸中大多数(约位置 60-389)与本发明的 OrfA(SEQ ID NO:1)上游部分匹配,因此据信在美国申请系列号 09/231,899 中在制备连续的 cDNA 构建体的工作中出现了错误。在美国申请系列号 09/231,899 中出现排序错误的区域在高度重复序列的区域内(即,下面讨论的 ACP 区域),很可能在从各种 cDNA 克隆装配该序列时产生了某种混淆。

[0065] OrfA 是一个 8730 个核苷酸的序列(不包括终止密码子),它编码本文表示为 SEQ ID NO:2 的 2910 个氨基酸的序列。OrfA 内有 12 个结构域:(a) 一个 β -酮脂酰-ACP 合酶(KS)结构域;(b) 一个丙二酰-CoA:ACP 酰基转移酶(MAT)结构域;(c) 9 个酰基载体蛋白

(ACP) 结构域 ;和 (d) 一个酮还原酶 (KR) 结构域。

[0066] OrfA 的核苷酸序列以登录号 AF378327(氨基酸序列登录号 AAK728879) 保存在 GenBank 中。在标准 BLAST 检索(在所有 6 个可读框中用标准默认参数进行 BLAST 2.0 Basic BLAST 同源性检索,使用 blastp 进行氨基酸检索,blastn 用于核酸检索,且 blastX 用于核酸检索和翻译的氨基酸序列检索,其中,通过默认筛选查询序列的低复杂性区域(在 Altschul, S. F., Madden, T. L., Schaeffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997), “间隔的 BLAST 和 PSI-BLAST: 新产生的蛋白质数据库检索程序”。Nucleic Acids Res. 25 :3389-3402 中描述,本文引用以其整体作为参考)) 中比较了 OrfA 与已知的序列。在核酸水平上, OrfA 与任何已知的核苷酸序列没有明显的同源性。在氨基酸水平上,与 ORFA 同源性程度最高的序列是:念珠藻属种类 7120 异型囊胞糖脂合酶(登录号 NC_003272),它与 ORFA 在 1001 个氨基酸残基上有 42% 相同;和 *Moritella marinus* (*Vibrio marinus*) ORF8(登录号 AB025342),它与 ORFA 在 993 个氨基酸残基上有 40% 相同。

[0067] OrfA 的第一个结构域是 KS 结构域,本文也称为 ORFA-KS。该结构域包含在覆盖从 SEQ ID NO:1 (OrfA) 的约位置 1 和 40 之间的起始位点到 SEQ ID NO:1 的约位置 1428 和 1500 之间的终止位点的核苷酸序列内。含有编码 ORFA-KS 结构域的序列的核苷酸序列在本文中表示为 SEQ ID NO:7 (SEQ ID NO:1 的位置 1-1500)。含有 KS 结构域的氨基酸序列覆盖从 SEQ ID NO:2 (ORFA) 的约位置 1 和 14 之间的起始位点到 SEQ ID NO:2 的约位置 476 和 500 之间的终止位点。含有 ORFA-KS 结构域的氨基酸序列本文表示为 SEQ ID NO:8 (SEQ ID NO:2 的位置 1-500)。应注意 ORFA-KS 结构域含有一个活性位点基序:DXAC*(* 酰基结合位点 C₂₁₅)。

[0068] 根据本发明,具有 3- 酮脂酰 -ACP 合酶 (KS) 生物学活性(功能)的结构域或蛋白质鉴定为完成 FAS(和 PKS) 延长反应循环起始步骤的酶。用于延长的酰基基团通过硫酯键连接到酶活性位点的半胱氨酸残基上。在多步反应中,酰基酶进行与丙二酰 -ACP 的缩合形成酮脂酰 -ACP, CO₂ 和游离酶。KS 在延长循环中起关键作用且在许多系统中表现出比该反应循环中的其它酶具有更大的底物特异性。例如,大肠杆菌具有三个不同的 KS 酶,它们在该生物体的生理学中分别具有各自特定的作用 (Magnuson 等, Microbiol. Rev. 57, 522 (1993))。PUFA-PKS 系统的这两个 KS 结构域在 PUFA 生物合成反应程序中具有不同的作用。

[0069] 作为一类酶,KS 已被充分鉴定。许多证实的 KS 基因的序列是已知的,已经鉴定了活性位点基序且测定了一些的晶体结构。由于属于 KS 酶家族,通过与已知 KS 序列的同源性可容易地鉴定该蛋白质。

[0070] OrfA 的第二个结构域是 MAT 结构域,本文也称为 ORFA MAT。该结构域包含在覆盖从 SEQ ID NO:1 (OrfA) 的约位置 1723 和 1798 之间的起始位点到 SEQ ID NO:1 的约位置 2805 和 3000 之间的终止位点的核苷酸序列内。含有编码 ORFA-MAT 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:9 (SEQ ID NO:1 的位置 1723-3000)。含有 MAT 结构域的氨基酸序列覆盖从 SEQ ID NO:2 (ORFA) 的约位置 575 和 600 之间的起始位点到 SEQ ID NO:2 的约位置 935 和 1000 之间的终止位点。含有 ORFA-MAT 结构域的氨基酸序列本文表示为 SEQ ID NO:10 (SEQ ID NO:2 的位置 575-1000)。应注意该 ORFA-MAT 结构域含有一个活性位点基序:GHS*⁷⁰⁶XG(* 酰基结合位点 S₇₀₆),本文表示为 SEQ ID NO:11。

[0071] 根据本发明,具有丙二酰-CoA:ACP 酰基转移酶(MAT)生物学活性(功能)的结构域或蛋白质鉴定为从丙二酰-CoA 转移丙二酰半分子到 ACP 的结构域或蛋白质。除了该活性位点基序(GxSxG)外,这些酶具有一个延长的基序和关键位置的 Q 氨基酸,因此将它们鉴定为 MAT 酶(与 *Schizochytrium* Orf B 的 AT 结构域相反)。在一些 PKS 系统(但不是 PUFA PKS 结构域)中, MAT 结构域优先将甲基-或乙基-丙二酰装载到 ACP 基团上(从相应的 CoA 酯),从而在线型碳链中导入分支。MAT 结构域可通过其与已知 MAT 序列的同源性和其延长的基序结构被识别。

[0072] OrfA 的结构域 3-11 是 9 个串连的 ACP 结构域,本文也称为 ORFA-ACP(该序列的第一个结构域是 ORFA-ACP1,第二个结构域是 ORFA-ACP2,第三个结构域是 ORFA-ACP3,等)。第一个 ACP 结构域,即 ORFA-ACP1 包含在覆盖从 SEQ ID NO:1(OrfA)的约位置 3343 到约位置 3600 的核苷酸序列内。含有编码 ORFA-ACP1 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:12(SEQ ID NO:1 的位置 3343-3600)。含有第一个 ACP 结构域的氨基酸序列覆盖从 SEQ ID NO:2 的约位置 1115 到约位置 1200。含有 ORFA-ACP1 结构域的氨基酸序列在本文中表示为 SEQ ID NO:13(SEQ ID NO:2 的位置 1115-1200)。应注意该 ORFA-ACP1 结构域含有一个活性位点基序:LGIDS*(泛酰巯基乙胺结合基序 S_{1157}),本文以 SEQ ID NO:14 表示。

[0073] 所有 9 个 ACP 结构域的核苷酸和氨基酸序列高度保守且,因此各结构域的序列在本文中没有用单独的序列标识符表示。然而,根据本文公开的信息,本领域的技术人员可容易地测定含有其它 8 个 ACP 结构域中每一个的序列(参见下面的讨论)。

[0074] 所有 9 个 ACP 结构域一起覆盖从 SEQ ID NO:1 的约位置 3283 到约位置 6288 的 OrfA 区域,它相应于 SEQ ID NO:2 从约 1095 到约 2096 的氨基酸位置。含有所有 9 个结构域的全部 ACP 区域的核苷酸序列在本文中表示为 SEQ ID NO:16。以 SEQ ID NO:16 表示的区域包括各个 ACP 结构域之间的接头片段。9 个结构域的重复间隔为 SEQ ID NO:16 中约每隔 330 个核苷酸(相邻活性位点丝氨酸之间测定的实际氨基酸数目范围从 104 到 116 个氨基酸)。9 个 ACP 结构域中每个含有一个泛酰巯基乙胺结合基序 LGIDS*(本文中以 SEQ ID NO:14 表示),其中 S^* 是泛酰巯基乙胺结合位点丝氨酸(S)。泛酰巯基乙胺结合位点丝氨酸(S)位于各 ACP 结构域序列的中心。ACP 结构域区域的每个末端和各 ACP 结构域之间是高度富含脯氨酸(P)和丙氨酸(A)的区域,据信它是接头区域。例如, ACP 结构域 1 和 2 之间是序列:APAPVKAAAPAAPVASAPAPA,本文表示为 SEQ ID NO:15。9 个 ACP 结构域中每一个的活性位点丝氨酸残基(即,泛酰巯基乙胺结合位点)的位置,相对于 SEQ ID NO:2 的氨基酸序列如下:ACP1 = S_{1157} ;ACP2 = S_{1266} ;ACP3 = S_{1377} ;ACP4 = S_{1488} ;ACP5 = S_{1604} ;ACP6 = S_{1715} ;ACP7 = S_{1819} ;ACP8 = S_{1930} ;和 ACP9 = S_{2034} 。由于 ACP 结构域的平均大小是约 85 个氨基酸,不包括接头,且包括接头为约 110 个氨基酸,且活性位点丝氨酸约位于该结构域的中心,因此本领域的技术人员可容易地测定 OrfA 中 9 个 ACP 结构域中每一个的位置。

[0075] 根据本发明,具有酰基载体蛋白(ACP)生物学活性(功能)的结构域或蛋白质鉴定为小多肽(一般 80 到 100 个氨基酸长),它通过与该蛋白质的共价结合的辅因子的硫酯键连接用作延长脂肪酸链的载体。它们作为分离的单位或者作为更大蛋白质内的结构域存在。ACPs 通过将 CoA 的磷酸泛酰巯基乙胺基(phosphopantetheinyl)半分子转移到 ACP 高度保守的丝氨酸残基上从失活的无活性形式转变成有功能的有活性形式。酰基基团通过在磷酸泛酰巯基乙胺基半分子游离末端的硫酯键附着到 ACP 上。通过用放射性泛酰巯基乙胺

标记和通过与已知 ACPs 的序列同源性可鉴定 ACPs。存在上述基序 (LGIDS*) 的变异体也是 ACP 的一个标记。

[0076] OrfA 中的结构域 12 是 KR 结构域,本文也称为 ORFA-KR。该结构域包含在覆盖从 SEQ ID NO:1 的约位置 6598 的起始位点到 SEQ ID NO:1 的约位置 8730 的终止位点的核苷酸序列内。含有编码 ORFA-KR 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:17 (SEQ ID NO:1 的位置 6598-8730)。含有 KR 结构域的氨基酸序列覆盖从 SEQ ID NO:2 (ORFA) 的约位置 2200 的起始位点到 SEQ ID NO:2 的约位置 2910 的终止位点。含有 ORFA-KR 结构域的氨基酸序列本文表示为 SEQ ID NO:18 (SEQ ID NO:2 的位置 2200-2910)。KR 结构域内具有与短链乙醛脱氢酶 (KR 是该家族的一个成员) 同源的核心区。该核心区跨越从 SEQ ID NO:1 的约位置 7198 至约位置 7500,它相应于 SEQ ID NO:2 的氨基酸位置 2400-2500。

[0077] 根据本发明,具有酮还原酶活性,也称为 3-酮脂酰-ACP 还原酶 (KR) 生物学活性 (功能) 的结构域或蛋白质鉴定为催化 ACP 的 3-酮脂酰形式的吡啶-核苷-依赖型还原的结构域或蛋白质。它是脂肪酸从头生物合成延长循环中的第一个还原步骤和通常在聚酮化合物生物合成中进行的反应。观察到与烯酰 ACP 还原酶 (ER) 的一个家族, FAS 的另一还原酶 (但不是在 PUFA PKS 系统中存在的 ER 家族), 和短链乙醇脱氢酶家族有明显的序列相似性。对上述 PUFA PKS 区域的 Pfam 分析揭示了在核心区与短链乙醇脱氢酶家族的同源性。对相同区域的 Blast 分析揭示了核心区与已知的 KR 酶匹配且延长区与其它已鉴定的 PUFAPKS 系统的结构域同源。

[0078] 可读框 B (OrfB) :

[0079] OrfB 的完整核苷酸序列在本文中表示为 SEQ ID NO:3。SEQ ID NO:3 的核苷酸 1311-6177 相应于美国申请系列号 09/231,899 中称为 SEQ ID NO:71 的序列的核苷酸 1-4867 (在美国申请系列号 09/231,899 中该 cDNA 序列在终止密码子之外含有约 345 个另外的核苷酸,包括 polyA 尾)。因此,SEQ ID NO:1 的核苷酸 1-1310 代表在美国申请系列号 09/231,899 中未公开的另一序列。SEQ ID NO:3 的这一新区域含有由 OrfB 编码的 KS 结构域的大部分。

[0080] OrfB 是一个 6177 个核苷酸的序列 (不包括终止密码子),它编码一个 2059 个氨基酸的序列,本文表示为 SEQ ID NO:4。OrfB 内有 4 个结构域:(a) 一个 β -酮脂酰-ACP 合酶 (KS) 结构域;(b) 一个链长因子 (CLF) 结构域;(c) 一个酰基转移酶 (AT) 结构域;和,(d) 一个烯酰 ACP-还原酶 (ER) 结构域。

[0081] OrfB 的核苷酸序列以登录号 AF378328 (氨基酸序列登录号 AAK728880) 保存在 GenBank 中。按上述标准 BLAST 检索比较了 OrfB 与已知序列。在核酸水平上,OrfB 与任何已知的核苷酸序列没有明显的同源性。在氨基酸水平上,与 ORFB 同源性程度最大的序列是:与 ORFB 在 458 个氨基酸残基上有 53% 相同性的希瓦氏菌属种类假定蛋白 (登录号 U73935);与 ORFB 在 460 个氨基酸残基上有 53% 相同性的 *Moritella marinus* (*Vibrio marinus*) ORF11 (登录号 AB025342);与 ORFB 在 457 个氨基酸残基上有 52% 相同性的 *Photobacterium profundum* ω -3 多不饱和脂肪酸合酶 PfaD (登录号 AF409100);和与 ORFB 在 430 个氨基酸残基上有 53% 相同性的念珠藻属种类 7120 假定蛋白 (登录号 NC_003272)。

[0082] OrfB 中的第一个结构域是 KS 结构域,本文也称为 ORFB-KS。该结构域包含在覆盖从 SEQ ID NO:3 (OrfB) 的约位置 1 和 43 之间的起始位点到 SEQ ID NO:3 的约位置 1332 和

1350 之间的终止位点的核苷酸序列内。含有编码 ORFB-KS 结构域的序列的核苷酸序列本文表示为 SEQ ID NO :19 (SEQ ID NO :3 的位置 1-1350)。含有 KS 结构域的氨基酸序列覆盖从 SEQ ID NO :4 (ORFB) 的约位置 1 和 15 之间的起始位点到 SEQ ID NO :4 的约位置 444 和 450 之间的终止位点。含有 ORFB-KS 结构域的氨基酸序列本文表示为 SEQ ID NO :20 (SEQ ID NO :4 的位置 1-450)。应注意 ORFB-KS 结构域含有一个活性位点基序 :DXAC* (* 酰基结合位点 C₁₉₆)。KS 的生物学活性和鉴定具有该活性的蛋白质或结构域的方法在上文中描述。

[0083] OrfB 中的第二个结构域是 CLF 结构域,本文也称为 ORFB-CLF。该结构域包含在覆盖从 SEQ ID NO :3 (OrfB) 的约位置 1378 和 1402 之间的起始位点到 SEQ ID NO :3 的约位置 2682 和 2700 之间的终止位点的核苷酸序列内。含有编码 ORFB-CLF 结构域的序列的核苷酸序列本文表示为 SEQ ID NO :21 (SEQ ID NO :3 的位置 1378-2700)。含有 CLF 结构域的氨基酸序列覆盖从 SEQ ID NO :4 (ORFB) 的约位置 460 和 468 之间的起始位点到 SEQ ID NO :4 的约位置 894 和 900 之间的终止位点。含有 ORFB-CLF 结构域的氨基酸序列本文表示为 SEQ ID NO :22 (SEQ ID NO :4 的位置 460-900)。应注意 ORFB-CLF 结构域含有 KS 活性位点基序但没有结合酰基的半胱氨酸。

[0084] 根据本发明,一个结构域或蛋白质根据下面的理论称为链长因子 (CLF)。CLF 最初描述为具有 II 型 (分离的酶) PKS 系统的特征且假定在确定延长循环的次数,从而确定终产物的链长中起作用。CLF 的氨基酸序列与 KS 结构域具有同源性 (且认为与 KS 蛋白质形成异二聚体),但是它们没有活性位点半胱氨酸。CLF 在 PKS 系统中的作用目前尚有争议。新证据 (C. Bisang 等, Nature 401, 502 (1999)) 表明在 PKS 系统的引发中起作用 (提供需要延长的起始酰基基团)。在该作用中,据认为 CLF 结构域使丙二酸盐 (以丙二酰 -ACP) 去羧基,从而形成可转移到 KS 活性位点的乙酸基团。因此该乙酸盐充当可进行起始延长 (缩合) 反应的“引发”分子。已经鉴定了 II 型 CLF 的同源物在一些模块型 PKS 系统中是“装载”结构域。具有 CLF 序列特征的结构域在所有目前鉴定的 PUFAPKS 系统中均被发现且在各种情况下均发现它是多结构域蛋白的一部分。

[0085] 在 OrfB 中的第三个结构域是 AT 结构域,本文也称为 ORFB-AT。该结构域包含在覆盖从 SEQ ID NO :3 (OrfB) 的约位置 2701 和 3598 之间的起始位点到 SEQ ID NO :3 的约位置 3975 和 4200 之间的终止位点的核苷酸序列内。含有编码 ORFB-AT 结构域的序列的核苷酸序列本文表示为 SEQ ID NO :23 (SEQ ID NO :3 的位置 2701-4200)。含有 AT 结构域的氨基酸序列覆盖从 SEQ ID NO :4 (ORFB) 的约位置 901 和 1200 之间的起始位点到 SEQ ID NO :4 的约位置 1325 和 1400 之间的终止位点。含有 ORFB-AT 结构域的氨基酸序列本文表示为 SEQ ID NO :24 (SEQ ID NO :4 的位置 901-1400)。应注意该 ORFB-AT 结构域含有活性位点基序 GxS^{*}xG (* 酰基结合位 S₁₁₄₀),它是酰基转移酶 (AT) 蛋白的特征。

[0086] “酰基转移酶”或“AT”是指能进行许多不同酰基转移反应的一大类酶。Schizochytrium 结构域与目前检查的所有其它 PUFA PKS 系统中存在的结构域具有较好的同源性且与鉴定为具有特异性功能的一些酰基转移酶 (例如,与丙二酰 -CoA:ACP 酰基转移酶, MAT) 具有极弱的同源性。尽管与 MAT 的同源性弱,据信该 AT 结构域不起 MAT 的作用,因为它不具有该酶的延长基序结构特征 (参见上文的 MAT 结构域描述)。为了本说明书的目的,PUFAPKS 系统中的 AT 结构域的功能包括,但不限于:将脂肪酰基从 ORFA ACP 结构域转移到水上 (即,硫酯酶 - 以游离脂肪酸释放脂肪酰基),将脂肪酰基基团转移到诸如 CoA

的受体上,在各种 ACP 结构域之间转移酰基,或者将脂肪酰基转移到亲脂受体分子(例如,溶血磷脂酸)上。

[0087] OrfB 中的第四个结构域是 ER 结构域,本文也称为 ORFB-ER。该结构域包含在覆盖从 SEQ ID NO:3(OrfB) 的约位置 4648 的起始位点到 SEQ ID NO:3 的约位置 6177 的终止位点的核苷酸序列内。含有编码 ORFB-ER 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:25(SEQ ID NO:3 的位置 4648-6177)。含有 ER 结构域的氨基酸序列覆盖从 SEQ ID NO:4(ORFB) 的约位置 1550 的起始位点到 SEQ ID NO:4 的约位置 2059 的终止位点。含有 ORFB-ER 结构域的氨基酸序列本文表示为 SEQ ID NO:26(SEQ ID NO:4 的位置 1550-2059)。

[0088] 根据本发明,该结构域具有烯酰还原酶(ER)的生物学活性。ER 酶还原脂酰-ACP 中的反式双键(由 DH 活性导入),导致这些碳完全饱和。PUFA PKS 中的 ER 结构域与新鉴定的 ER 酶家族(Heath 等, Nature 406,145(2000))具有同源性。Heath 和 Rock 通过从肺炎链球菌克隆目的基因,纯化从该基因表达的蛋白质,并表明它在体外试验中具有 ER 活性鉴定了 ER 酶的这一新类型。OrfB 的 Schizochytrium ER 结构域的序列与肺炎链球菌 ER 蛋白质具有同源性。目前检查的所有 PUFA PKS 系统均含有至少一个具有与 Schizochytrium ER 结构域同源性极高的序列的结构域。Schizochytrium PUFA PKS 系统含有两个 ER 结构域(一个在 OrfB 上,一个在 OrfC 上)。

[0089] 可读框 C(OrfC) :

[0090] OrfC 的完整核苷酸序列在本文中表示为 SEQ ID NO:5。SEQ ID NO:5 的核苷酸 1-4509(即,完整可读框序列,不包括终止密码子)相应于美国申请系列号 09/231,899 中称为 SEQ ID NO:76 的序列的核苷酸 145-4653(美国申请系列号 09/231,899 中的 cDNA 序列在 OrfC 起始密码子的上游含有约 144 个核苷酸且在终止密码子之后含有约 110 个核苷酸,包括 polyA 尾)。OrfC 是一个 4509 个核苷酸的序列(不包括终止密码子),它编码 1503 个氨基酸的序列,本文表示为 SEQ ID NO:6。OrfC 内含有 3 个结构域:(a) 两个 FabA-样 β -羟酰-ACP 脱水酶(DH)结构域;和(b) 一个烯酰 ACP 还原酶(ER)结构域。

[0091] OrfC 的核苷酸序列以登录号 AF378329(氨基酸序列登录号 AAK728881)保存在 GenBank 中。按上述标准 BLAST 检索比较了 OrfC 与已知序列。在核酸水平上,OrfC 与任何已知的核苷酸序列没有明显的同源性。在氨基酸水平(Blastp)上,与 ORFC 同源性程度最大的序列是:与 ORFC 在 514 个氨基酸残基上有 45%相同性的 *Moritella marinus*(*Vibrio marinus*)ORF11(登录号 AB025342);与 ORFC 在 447 个氨基酸残基上有 49%相同性的希瓦氏菌属种类假定蛋白 8(登录号 U73935);与 ORFC 在 430 个氨基酸残基上有 49%相同性的念珠藻属种类假定蛋白(登录号 NC_003272);和与 ORFC 在 930 个氨基酸残基上有 37%相同性的希瓦氏菌属种类假定蛋白 7(登录号 U73935)。

[0092] OrfC 中的第一个结构域是 DH 结构域,本文也称为 ORFC-DH1。它是 OrfC 中的两个 DH 结构域之一,因此命名为 DH1。该结构域包含在覆盖从 SEQ ID NO:5(OrfC) 的约位置 1 和 778 之间的起始位点到 SEQ ID NO:5 的约位置 1233 和 1350 之间的终止位点的核苷酸序列内。含有编码 ORFC-DH1 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:27(SEQ ID NO:5 的位置 1-1350)。含有 DH1 结构域的氨基酸序列覆盖从 SEQ ID NO:6(ORFC) 的约位置 1 和 260 之间的起始位点到 SEQ ID NO:6 的约位置 411 和 450 之间的终止位点。含有 ORFC-DH1 结构域的氨基酸序列本文表示为 SEQ ID NO:28(SEQ ID NO:6 的位置 1-450)。

[0093] PUFA PKS 系统中两个 DH 结构域 (见下文 DH2) 的特征已经在前面部分中描述。这类酶从 β -酮脂酰-ACP 去掉 HOH 并在碳链中留下反式双键。PUFAPKS 系统的 DH 结构域与细菌的 FAS 系统相关性 DH 酶 (而不是其它 PKS 系统的 DH 结构域) 具有同源性。细菌 DH 的一个亚型, 即 FabA-样 DH, 具有顺反异构酶活性 (Heath 等, J. Biol. Chem., 271, 27795 (1996))。它与 FabA-样 DH 具有同源性表明 DH 结构域之一或者两个同时负责在 PUFA PKS 产物中插入顺式双键。

[0094] OrfC 中的第二个结构域是 DH 结构域, 本文也称为 ORFC-DH2。它是 OrfC 中两个 DH 结构域的第二个, 因此命名为 DH2。该结构域包含在覆盖从 SEQ ID NO:5 (OrfC) 的约位置 1351 和 2437 之间的起始位点到 SEQ ID NO:5 的约位置 2607 和 2850 之间的终止位点的核苷酸序列内。含有编码 ORFC-DH2 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:29 (SEQ ID NO:5 的位置 1351-2850)。含有 DH2 结构域的氨基酸序列覆盖从 SEQ ID NO:6 (ORFC) 的约位置 451 和 813 之间的起始位点到 SEQ ID NO:6 的约位置 869 和 950 之间的终止位点。含有 ORFC-DH2 结构域的氨基酸序列本文表示为 SEQ ID NO:30 (SEQ ID NO:6 的位置 451-950)。DH 的生物学活性已在上文中描述。

[0095] OrfC 中的第三个结构域是 ER 结构域, 本文也称为 ORFC-ER。该结构域包含在覆盖从 SEQ ID NO:5 (OrfC) 的约位置 2998 的起始位点到 SEQ ID NO:5 的约位置 4509 的终止位点的核苷酸序列内。含有编码 ORFC-ER 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:31 (SEQ ID NO:5 的位置 2998-4509)。含有 ER 结构域的氨基酸序列覆盖从 SEQ ID NO:6 (ORFC) 的约位置 1000 的起始位点到 SEQ ID NO:6 的约位置 1502 的终止位点。含有 ORFC-ER 结构域的氨基酸序列本文表示为 SEQ ID NO:32 (SEQ ID NO:6 的位置 1000-1502)。ER 的生物学活性已在上文中描述。

[0096] 本发明的一个实施方案涉及含有来自非细菌 PUFA PKS 系统的核酸序列, 其同源物, 其片段, 和 / 或与任一该核酸序列互补的核酸序列的分离的核酸分子。在一个方面, 本发明涉及含有选自下组的核酸序列的分离核酸分子: (a) 编码选自由 SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 及其生物学活性片段组成的组中的氨基酸序列的核酸序列; (b) 编码选自 SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:30, SEQ ID NO:32, 或其生物学活性片段的氨基酸序列的核酸序列; (c) 编码与 (a) 所述氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列, 其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性; (d) 编码与 (b) 所述氨基酸序列有至少约 60% 相同性的氨基酸序列的核酸序列, 其中所述氨基酸序列具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的生物学活性; 或 (e) 与 (a), (b), (c), 或 (d) 的核酸序列完全互补的核酸序列。在另一实施方案中, 本发明涉及包含编码几个 PUFA PKS 结构域的活性位点结构域或上述其它功能基序的序列的核酸序列。

[0097] 根据本发明, 具有 PUFAPKS 系统的至少一个结构域的生物学活性的氨基酸序列是具有以 Schizochytrium PUFA PKS 系统例证的, 本文详细描述 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列。Schizochytrium PUFA PKS 系统内的各结构域的生物学活性已在上文中详细描述。因此, 本发明的分离的核酸分子可编码任一 PUFA PKS 可读框

的翻译产物, PUFA PKS 结构域, 其生物学活性片段, 或天然 PUFA PKS 可读框或具有生物学活性的结构域的任一同源物。给定蛋白质或结构域的同源物是具有与天然对照氨基酸序列(即, 对照蛋白质或结构域)区别在于有一个或几个, 但不限于一个或几个的氨基酸缺失(例如, 蛋白质的截短形式, 如肽或片段), 插入, 颠倒, 取代和/或衍生化(例如, 通过糖基化, 磷酸化, 乙酰化, 豆蔻酰化, 异戊二烯化, 棕榈酸化, 酰胺化和/或添加糖基磷脂酰肌醇)的氨基酸序列的蛋白质或多肽。PUFA PKS 蛋白质或结构域的优选同源物在下文详细描述。应注意同源物可包括合成产生的同源物, 给定蛋白质或结构域的天然等位变异体, 或来自不同于产生对照序列的生物的生物体的同源序列。

[0098] 一般来说, 蛋白质或结构域的生物学活性或生物学作用是指体内(即, 在该蛋白质的天然生理学环境中)或体外(即, 在实验室条件下)测量或观察时由该蛋白质或结构域表现或完成的归因于该蛋白质或结构域的天然存在形式的任何功能。PUFA PKS 系统和组成 PUFA PKS 系统的单个蛋白质/结构域的生物学活性本文在别处有详细描述。蛋白质或结构域的修饰, 例如在同源物或模拟物(在下面讨论)中的修饰可产生与天然蛋白质或结构域具有相同生物学活性的蛋白质或结构域, 或者产生与天然蛋白质或结构域相比具有降低或增加的生物学活性的蛋白质或结构域。导致蛋白质或结构域表达降低或活性降低的修饰可称为蛋白质或结构域的失活(完全或部分), 下调, 或作用减小。同样, 导致蛋白质或结构域的表达增加或活性增加的修饰可称为蛋白质或结构域的放大, 超产, 活化, 增强, 上调或作用增加。PUFA PKS 系统的功能域是能够完成生物学功能(即, 具有生物学活性)的结构域(即, 结构域可以是蛋白质的一部分)。

[0099] 根据本发明, 分离的核酸分子是从其天然环境中取出(即, 对其进行了人工操作)的核酸分子, 其天然环境是天然状态下发现该核酸分子的基因组或染色体。因此, “分离的”不必反映该核酸分子被纯化的程度, 但是表明该核酸分子不包括天然状况下发现该核酸分子的完整基因组或完整染色体。分离的核酸分子可包括基因。包括基因的分离的核酸分子不是包括该基因的染色体的一个片段, 而是包括编码区和与该基因相关的调控区, 但是没有在同一染色体中天然发现的其它基因。分离的核酸分子也可包括一种特定的核酸序列, 其侧翼(即, 在该序列的 5' 和/或 3' 端)为天然状况下一般不是该特定核酸序列侧翼的其它核酸(即, 异源序列)。分离的核酸分子可包括 DNA, RNA(即, mRNA), 或 DNA 或 RNA 之一的衍生物(例如, cDNA)。尽管短语“核酸分子”主要是指物理学核酸分子且短语“核酸序列”主要是指核酸分子上的核苷酸序列, 但是两个短语可交换使用, 特别是对于能编码蛋白质或蛋白质的结构域的核酸分子或核酸序列。

[0100] 优选的是, 使用重组 DNA 技术(即, 聚合酶链式反应(PCR)扩增, 克隆)或化学合成产生本发明的分离的核酸分子。分离的核酸分子包括天然的核酸分子及其同源物, 包括, 但不限于, 天然等位基因变异体和修饰的核酸分子, 其中的核苷酸以这样的方式插入, 缺失, 取代, 和/或颠倒, 即该修饰对本文所述 PUFA PKS 系统的生物学活性产生了所需的影响。蛋白质同源物(例如, 由核酸同源物编码的蛋白质)在上文中已有详细讨论。

[0101] 可使用本领域的技术人员已知的许多方法来产生核酸分子同源物(参见, 例如, Sambrook 等, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989)。例如, 可使用各种技术修饰核酸分子, 该技术包括, 但不限于, 传统诱变技术和重组 DNA 技术, 例如定点诱变, 化学处理核酸分子以诱导突变, 限制性酶裂解核酸片段, 连接核酸

片段,核酸序列选定区域的 PCR 扩增和 / 或诱变,合成寡核苷酸混合物并连接混合物组以“构建”核酸分子的混合体及其组合。通过筛选核酸编码的蛋白质的功能和 / 或通过与野生型基因杂交可从修饰的核酸混合物中选择核酸分子同源物。

[0102] 本发明的核酸分子大小的最小值是足以形成探针或寡核苷酸引物的大小,该探针或寡核苷酸引物与本发明中有用的核酸分子的互补序列能够形成稳定的杂交体(例如,在中等,高度或极高严格条件下),或者其大小足以编码具有根据本发明的 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列。因此,编码这一蛋白质的核酸分子的大小取决于核酸组成和核酸分子与互补序列之间的同源性或相同性百分数以及杂交条件本身(例如,温度,盐浓度,和甲酰胺浓度)。用作寡核苷酸引物或探针的核酸分子大小的最小值一般是如果该核酸分子富含 GC 则为至少约 12 个到约 15 个核苷酸的长度且如果它们富含 AT 则为至少约 15 到约 18 个碱基的长度。对本发明的核酸分子大小的最大值没有限制,实际的限制是该核酸分子可包含足以编码 PUPA PKS 系统的结构域的生物活性片段,PUFA PKS 系统的完整结构域,PUFA PKS 系统的可读框(Orf)内的几个结构域,PUFA PKS 系统的一个完整的 Orf,或 PUFA PKS 系统的一个以上的 Orf。

[0103] 在本发明的一个实施方案中,分离的核酸分子含有选自:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32,或其生物活性片段的组中的核酸序列或基本上由其组成。在一个方面,该核酸序列选自:SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:12, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29,和 SEQ ID NO:31 的组中。在本发明的一个实施方案中,可产生在给定氨基酸序列的 C- 和 / 或 N- 末端各侧带有从至少一个,到多达约 20 个添加的异源氨基酸的任一上述 PUFA PKS 的氨基酸序列,以及该序列的同源物。所得的蛋白质或多肽可称为“基本上由给定的氨基酸序列组成”。根据本发明,异源性氨基酸是天然状况下不是在给定氨基酸序列侧翼发现(即,体内天然状况下未发现)的或在该基因中存在时如果使用产生该给定氨基酸序列的生物的标准密码子用法翻译天然序列中的该核苷酸时,不由编码给定氨基酸序列的天然核酸序列侧翼的核苷酸编码的氨基酸序列。同样,短语“基本上由其组成”在本文中当用于指核酸序列时,是指编码给定氨基酸序列的核酸序列在编码给定氨基酸序列的核酸序列 5' 和 / 或 3' 末端各侧带有从至少一个到多达约 60 个添加的异源核苷酸。异源性核苷酸是在天然基因中存在时不是在编码给定氨基酸序列的核酸序列侧翼天然发现(即,体内天然状况下未发现)的。

[0104] 本发明还包括分离的核酸分子,它含有编码具有 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列的核酸序列。在一个方面,该核酸序列编码包括:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或 SEQ ID NO:32 的任一 Schizochytrium PUFA PKS ORFs 或结构域的同源物,其中该同源物具有本文前面所述 PUFA PKS 系统的至少一个结构域的生物学活性。

[0105] 在本发明的一个方面,本发明包含的 Schizochytrium PUFA PKS 蛋白质或结构域的同源物含有与选自:SEQ ID NO:2, SEQ ID NO:4, 和 SEQ ID NO:6 的氨基酸序列的至少 500

个连续氨基酸有至少约 60% 相同性的氨基酸序列 ; 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。在另一方面, 该同源物的氨基酸序列与 SEQ ID NO :2, SEQ ID NO :4 和 SEQ ID NO :6 中任一个的至少约 600 个连续氨基酸, 且更优选至少约 700 个连续氨基酸, 且更优选至少约 800 个连续氨基酸, 且更优选至少约 900 个连续氨基酸, 且更优选至少约 1000 个连续氨基酸, 且更优选至少约 1100 个连续氨基酸, 且更优选至少约 1200 个连续氨基酸, 且更优选至少约 1300 个连续氨基酸, 且更优选至少约 1400 个连续氨基酸, 且更优选至少约 1500 个连续氨基酸, 或与 SEQ ID NO :6 的全长有至少约 60% 的相同性。在另一方面, 该同源物的氨基酸序列与 SEQ ID NO :2 或 SEQ ID NO :4 中任一个的至少约 1600 个连续氨基酸, 且更优选至少约 1700 个连续氨基酸, 且更优选至少约 1800 个连续氨基酸, 且更优选至少约 1900 个连续氨基酸, 且更优选至少约 2000 个连续氨基酸, 或与 SEQ ID NO :4 的全长有至少约 60% 的相同性。在另一方面, 该同源物的氨基酸序列与 SEQ ID NO :2 的至少约 2100 个连续氨基酸, 且更优选至少约 2200 个连续氨基酸, 且更优选至少约 2300 个连续氨基酸, 且更优选至少约 2400 个连续氨基酸, 且更优选至少约 2500 个连续氨基酸, 且更优选至少约 2600 个连续氨基酸, 且更优选至少约 2700 个连续氨基酸, 且更优选至少约 2800 个连续氨基酸, 且甚至更优选与 SEQ ID NO :2 的全长有至少约 60% 的相同性。

[0106] 在另一方面, 本发明包含的 Schizochytrium PUFA PKS 蛋白质或结构域的同源物含有与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列在上段所述任一连续氨基酸长度上有至少约 65% 的相同性, 且更优选至少约 70% 的相同性, 且更优选至少约 75% 的相同性, 且更优选至少约 80% 的相同性, 且更优选至少约 85% 的相同性, 且更优选至少约 90% 的相同性, 且更优选至少约 95% 的相同性, 且更优选至少约 96% 的相同性, 且更优选至少约 97% 的相同性, 且更优选至少约 98% 的相同性, 且更优选至少约 99% 的相同性的氨基酸序列, 其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

[0107] 在本发明的一个方面, 本发明包含的 Schizochytrium PUFA PKS 蛋白质或结构域的同源物含有与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列有至少约 60% 的相同性的氨基酸序列, 其中所述氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。在另一方面, 该同源物的氨基酸序列与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列有至少约 65% 的相同性, 且更优选至少约 70% 的相同性, 且更优选至少约 75% 的相同性, 且更优选至少约 80% 的相同性, 且更优选至少约 85% 的相同性, 且更优选至少约 90% 的相同性, 且更优选至少约 95% 的相同性, 且更优选至少约 96% 的相同性, 且更优选至少约 97% 的相同性, 且更优选至少约 98% 的相同性, 且更优选至少约 99% 的相同性, 其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

[0108] 根据本发明, 术语“邻近的”或“连续的”对于本文所述核酸或氨基酸序列是指以不中断的序列相连。例如, 对于第一个序列含有第二个序列的 30 个相邻 (或连续) 的氨基酸, 是指第一个序列包含与第二个序列中 30 个氨基酸残基的不中断的序列具有 100% 相同性的 30 个氨基酸残基的不中断的序列。同样, 对于第一个序列与第二个序列具有“100% 相同性”是指第一个序列与第二个序列完全匹配, 在核苷酸或氨基酸之间没有间隔。

[0109] 如本文所用,除非另有说明,对于相同性百分数(%)是指使用:(1)在所有6个可读框中用标准默认参数进行BLAST 2.0 Basic BLAST 同源性检索,使用blastp进行氨基酸检索,blastn用于核酸检索,且blastX用于核酸检索和翻译的氨基酸检索,其中,通过默认筛选查询序列的低复杂性区域(在Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997),“间隔的BLAST和PSI-BLAST:新产生的蛋白质数据库检索程序”。Nucleic Acids Res. 25:3389-3402中描述,本文引用以其整体作为参考);(2)BLAST 2序列对比(使用下述参数);(3)和/或使用标准默认参数的PSI-BLAST(位置特异性重复BLAST)进行的同源性评估。应注意由于BLAST 2.0 Basic BLAST和BLAST 2之间在标准参数上的一些区别,两个特定序列使用BLAST 2程序可能认为具有明显的同源性,而使用一个序列作为查询序列以BLAST 2.0 Basic BLAST进行检索时在最匹配的序列中可能没有鉴定出第二个序列。另外,PSI-BLAST提供了一个“提问档(profile)”检索的自动的,容易使用的版本,它是寻找序列同源性的一个敏感途径。该程序首先进行有间隔的BLAST数据库检索。PSI-BLAST程序使用来自任何有效序列对比的信息返回构建位置特异性记分矩阵,它代替查询序列用于下一轮数据库检索。因此,应明白可使用这些程序中的任一种测定相同性百分数。

[0110] 使用Tatusova和Madden,(1999),“Blast 2序列-用于比较蛋白质和核苷酸序列的一种新工具”,FEMS Microbiol Lett. 174:247-250所述BLAST 2序列可互相对比两个具体序列,本文引用该文献以其整体作为参考。使用BLAST 2.0算法以blastp或blastn进行BLAST 2序列对比,在两个序列之间进行有间隔的BLAST检索(BLAST 2.0)允许在所得的序列对比中引入间隔(,缺失和插入)。为了本文清楚的目的,使用如下标准默认参数进行BLAST 2序列对比。

[0111] 对于blastn,使用0 BLOSUM62矩阵:

[0112] 匹配得分=1

[0113] 错配罚分=-2

[0114] 空出间隔(5)和延伸间隔(2)罚分

[0115] 间隔_x下降(dropoff)(50)期望(expect)(10)序列大小(word size)(11)筛选(上)

[0116] 对于blastp,使用0 BLOSUM62矩阵:

[0117] 空出间隔(11)和延伸间隔(1)罚分

[0118] 间隔_x下降(50)期望(expect)(10)序列大小(3)筛选(上)。

[0119] 在本发明的另一实施方案中,具有本发明的PUFA PKS系统的至少一个结构域的生物活性的氨基酸序列包括与天然PUFA PKS蛋白质或多肽足够相似以致于编码该氨基酸序列的核酸序列在中等,高度,或极高严格条件(下文描述)下能够杂交到(即,与其杂交)编码天然PUFA PKS蛋白质或多肽的核酸分子(即,编码天然PUFA PKS蛋白质或多肽的核酸链的互补链)上的氨基酸序列。优选的是,具有本发明的PUFA PKS系统的至少一个结构域的生物活性的氨基酸序列由这样一种核酸序列编码,即该核酸序列在中等,高度或极高严格条件下杂交到编码含有由SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30,或SEQ ID NO:32中任一个代表的氨基酸

序列的蛋白质的核酸序列的互补链上。推定互补序列的方法对于本领域的技术人员是已知的。应注意由于氨基酸测序和核酸测序技术不是完全无误差的,因此本文提供的序列最多代表本发明的 PUFAPKS 结构域和蛋白质的表现序列。

[0120] 本文使用的杂交条件是指标准杂交条件,在该条件下使用核酸分子鉴定相似的核酸分子。该标准条件在例如,Sambrook 等,Molecular Cloning :ALaboratory Manual,Cold Spring Harbor Labs Press,1989 中公开。Sambrook 等,出处同上,在本文中引用以其整体作为参考(具体参见,第 9.31-9.62 页)。另外,计算合适的杂交和洗涤条件以完成允许各种程度的核苷酸错配的杂交的公式在例如,Meinkoth 等,1984,Anal. Biochem. 138,267-284 中公开;Meinkoth 等,出处同上,在本文中引用以其整体作为参考。

[0121] 更具体地说,本文提及的中等严格的杂交和洗涤条件是指在杂交反应中允许分离与用于探测的核酸分子具有至少约 70%的核酸序列相同性的核酸分子的条件(即,允许约 30%或更少的核苷酸错配的条件)。本文提及的高度严格的杂交和洗涤条件是指在杂交反应中允许分离与用于探测的核酸分子具有至少约 80%的核酸序列相同性的核酸分子的条件(即,允许约 20%或更少的核苷酸错配的条件)。本文提及的极高严格的杂交和洗涤条件是指在杂交反应中允许分离与用于探测的核酸分子具有至少约 90%的核酸序列相同性的核酸分子的条件(即,允许约 10%或更少的核苷酸错配的条件)。如上所述,本领域技术人员可使用 Meinkoth 等,出处同上中的公式计算合适的杂交和洗涤条件以达到这些特定的核苷酸错配水平。该条件可依赖于形成的是 DNA : RNA 还是 DNA : DNA 杂交体而变化。计算的 DNA : DNA 杂交体的解链温度比 DNA : RNA 杂交体低 10℃。在具体实施方案中,DNA : DNA 杂交体的严格杂交条件包括在 6X SSC(0.9M Na⁺) 的离子强度下在约 20℃和约 35℃之间(低度严格),更优选的是,在约 28℃和约 40℃之间(更严格),且甚至更优选的是,在约 35℃和约 45℃之间(甚至更严格)的温度下,在合适的洗涤条件下的杂交。在具体实施方案中,DNA : RNA 杂交体的严格杂交条件包括在 6X SSC(0.9M Na⁺) 的离子强度下在约 30℃和约 45℃之间,更优选的是,在约 38℃和约 50℃之间,且甚至更优选的是,在约 45℃和约 55℃之间的温度下,以同样严格的洗涤条件的杂交。这些值根据大于约 100 个核苷酸,0%甲酰胺和约 40%的 G+C 含量的分子的解链温度计算。另外,可按 Sambrook 等,出处同上,第 9.31 至 9.62 页提供的经验计算 T_m 值。一般来说,洗涤条件应当尽可能严格,且对于选定的杂交条件应是合适的。例如,杂交条件可包括盐和温度条件的组合,该温度比计算的特定杂交体的 T_m 低约 20-25℃,且洗涤条件一般包括盐和温度条件的组合,该温度比计算的特定杂交体的 T_m 低约 12-20℃。适用于 DNA : DNA 杂交体的杂交条件的一个例子包括在 6X SSC(50%甲酰胺)中在约 42℃下杂交 2-24 小时,接着进行洗涤步骤,包括在室温下在约 2X SSC 中的一次或多次洗涤,接着在更高温度和更低离子强度下进一步洗涤(例如,在约 37℃下在约 0.1X-0.5X SSC 中洗涤至少一次,接着在约 68℃下在约 0.1X-0.5X SSC 中洗涤至少一次)。

[0122] 本发明的另一实施方案包括含有重组载体和核酸分子的重组核酸分子,其中的核酸分子含有的核酸序列编码具有本文所述 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列。该核酸序列在上文中有详细描述。根据本发明,重组载体是一种改造的(即,人工产生的)核酸分子,它用作处理选定核酸序列并将该核酸序列导入宿主细胞的工具。因此重组载体适用于克隆,测序,和/或处理选定的核酸序列,例如通过表达和/或传

递选定的核酸序列到宿主细胞中以形成重组细胞。该载体一般含有异源核酸序列,异源核酸序列是天然状况下来发现与需要克隆或传递的核酸序列相连的核酸序列,尽管该载体也可含有天然状况下发现与本发明的核酸分子相连的或者用于表达本发明的核酸分子的调控核酸序列(例如,启动子,非翻译区)(下面详细讨论)。该载体可以是 RNA 或者是 DNA,是原核的或者真核的,且一般是一种质粒。该载体可作为染色体外因子(例如,质粒)维持或者可整合进重组生物(例如,微生物或植物)的染色体中。该完整载体可在宿主细胞内适当保留,或者在某些情况下,可删除质粒 DNA,留下本发明的核酸分子。整合的核酸分子可在染色体启动子控制下,在自身的或质粒的启动子控制下,或者在一些启动子的结合控制下。可整合单个或多个拷贝的核酸分子进入染色体中。本发明的重组载体可含有至少一个选择标记。

[0123] 在一个实施方案中,用于本发明的重组核酸分子的重组载体是一种表达载体。本文所用的短语“表达载体”用于指适合于产生编码产物(例如,目的蛋白)的载体。在该实施方案中,将编码需要产生的产物(例如,PUFA PKS 结构域)的核酸序列插入重组载体中以产生重组核酸分子。编码需要产生的蛋白质的核酸序列以这样的方式插入载体中,即将该核酸序列与载体中的调控序列可操作地连接使得能够在重组宿主细胞内转录和翻译该核酸序列。

[0124] 在另一实施方案中,用于本发明的重组核酸分子的重组载体是一种定向载体。本文所用的短语“定向载体”用于指这样一种载体,即它用于将特定核酸分子传递进重组宿主细胞,其中该核酸分子用于删除或灭活宿主细胞或微生物内的内源性基因(即,用于定向基因破坏或敲除技术)。该载体本领域也称为“敲除”载体。在该实施方案的一个方面,该载体的一部分,但更典型的是插入载体中的核酸分子(即,插入片段)具有与宿主细胞中靶基因(即,删除或失活针对的基因)的核酸序列同源的核酸序列。载体插入片段的核酸序列设计成与靶基因结合以便靶基因与插入片段进行同源重组,从而删除,灭活或减弱该内源性靶基因(即,通过突变或删除至少一部分内源性靶基因)。

[0125] 一般来说,重组核酸分子包括与一个或多个转录调控序列可操作地相连的至少一个本发明的核酸分子。本文所用的短语“重组分子”或“重组核酸分子”主要是指与转录调控序列可操作地相连的核酸分子或核酸序列,但是当核酸分子是本文讨论的重组分子时,上述短语可与短语“核酸分子”交换使用。根据本发明,短语“可操作地相连”是指核酸分子与转录调控序列以这样的方式相连,即该连接使得该分子转染(即,转化,转导,转染,接合或导入)进宿主细胞时能够表达。转录调控序列是控制转录开始,延长,或终止的序列。特别重要的转录调控序列是那些控制转录开始的序列,例如启动子,增强子,操纵子和阻抑序列。合适的转录调控序列包括在该重组核酸分子导入的宿主细胞或生物体中能起作用的任何转录调控序列。

[0126] 本发明的重组核酸分子也可含有其它调控序列,例如翻译调控序列,复制起点,和与该重组细胞相适应的其它调控序列。在一个实施方案中,包括整合进宿主细胞染色体中的那些分子的本发明的重组分子还含有分泌信号(即,信号片段核酸序列)以便使得表达的蛋白质从产生该蛋白质的细胞中分泌出来。合适的信号片段包括天然状况下与表达该蛋白质相关的信号片段或者能够指导根据本发明的蛋白质分泌的任何异源性信号片段。在另一实施方案中,本发明的重组分子包含使得表达蛋白能够传递到宿主细胞膜上并插入膜中

的前导序列。合适的前导序列包括天然状况下与该蛋白质相关的前导序列,或者能够指导该蛋白质传递并插入细胞膜中的任何异源性前导序列。

[0127] 本发明人发现 *Schizochytrium* PUFA PKS Orfs A 和 B 在基因组中紧密相连并测序了 Orfs 之间的区域。Orfs 以相反的方向排列且有 4244 个碱基对分开起始 (ATG) 密码子 (即,它们排列如下:3' OrfA5' -4244bp-5' OrfB3')。检查 4244bp 的基因间区域没有揭示出任何明显的 Orfs (在 BlastX 检索中没有发现明显的匹配)。Orfs A 和 B 两者在 *Schizochytrium* 中至少在产油期间都高度表达,意味着在该基因间区域包含有效启动子元件。据信这些遗传因子在转基因应用中作为双向启动子序列具有实用性。例如,在优选的实施方案中,可克隆该区域,在其各端放置任意目的基因并将该构建体导入 *Schizochytrium* (或者表明该启动子可发挥功能的一些其它宿主) 中。预期该调控元件在合适条件下可为两个导入的基因提供同等的高水平表达。含有 *Schizochytrium* PUFA PKS 调控元件 (例如,启动子) 的该调控区的完整核苷酸序列本文表示为 SEQ ID NO :36。

[0128] 同样, OrfC 在产油期间在 *Schizochytrium* 中高度表达且预期在其起始密码子的上游区域存在调控元件。已克隆并测序了 OrfC 的基因组 DNA 的上游区域并在本文中表示为 (SEQ ID NO :37)。该序列含有 3886nt,紧靠 OrfC 起始密码子的上游。检查该区域没有发现任何明显的 Orfs (即,在 BlastX 检索中没有发现明显的匹配)。据信在该区域包含的调控元件在合适的条件下可为放在其后的基因提供高水平的表达。另外,在合适条件下,该表达水平与在 A-B 基因间区域 (SEQ ID NO :36) 控制下的基因等同。

[0129] 因此,在一个实施方案中,按本文所述用于本发明的重组核酸分子可包括 SEQ ID NO :36 和 / 或 SEQ ID NO :37 内包含的 PUFA PKS 调控区。该调控区可包括至少具有基本 PUFA PKS 转录活性的 SEQ ID NO :36 和 / 或 SEQ ID NO :37 的任一部分 (片段)。

[0130] 可使用本发明的一种或多种重组分子产生本发明的编码产物 (例如,PUFA PKS 结构域,蛋白质,或系统)。在一个实施方案中,通过在有效产生蛋白质的条件下表达本文所述核酸分子来生产编码产物。产生编码蛋白的优选方法是通过用一种或多种重组分子转染宿主细胞以形成重组细胞。用于转染的合适宿主细胞包括,但不限于,可被转染的任何细菌,真菌 (例如,酵母),昆虫,植物或动物细胞。宿主细胞可以是未转染的细胞或者是已被至少一个其它重组核酸分子转染的细胞。

[0131] 根据本发明,术语“转染”用于指可将外源性核酸分子 (即,重组核酸分子) 插入细胞中的任一方法。术语“转化”当用于指将核酸分子导入微生物细胞,例如藻类,细菌和酵母时,可与术语“转染”交换使用。在微生物系统中,术语“转化”用于描述由于该微生物获得外源性核酸的遗传改变且与术语“转染”基本上同义。然而,在动物细胞中,转化被赋予了第二种含义,例如,它可指细胞变成癌细胞后在培养中生长特性的改变。因此,为了避免混淆,术语“转染”优选用于将外源核酸导入动物细胞的情况,且本文使用的术语“转染”一般包含动物细胞,植物细胞的转染和微生物细胞的转化,该术语适用的范围是将外源性核酸导入细胞。因此,转染技术包括,但不限于,转化,粒子轰击,电穿孔,显微注射,脂转染,吸附,感染和原生质体融合。

[0132] 本领域的技术人员将预料到使用重组 DNA 技术通过操纵例如,宿主细胞内核酸分子的拷贝数,转录这些核酸分子的效率,翻译所得转录子的效率,和翻译后修饰的效率可改进转染的核酸分子的表达调控。另外,可遗传改造启动子序列以便与天然启动子相

比表达水平提高。用于控制核酸分子表达的重组技术包括,但不限于,将该核酸分子整合进一个或多个宿主细胞染色体中,给质粒添加载体稳定性序列,取代或修饰转录控制信号(例如,启动子,操纵子,增强子),取代或修饰翻译控制信号(例如,核糖体结合位点,Shine-Dalgarno 序列),修饰核酸分子以符合该宿主细胞的密码子用法,和缺失使转录子不稳定的序列。

[0133] 上文对于重组核酸分子和宿主细胞的转染方面的一般性讨论试图应用于本文讨论的任一重组核酸分子,包括编码具有 PUFA PKS 的至少一个结构域的生物学活性的任一氨基酸序列的那些核酸分子,编码其它 PKS 系统的氨基酸序列的那些核酸分子,和编码其它蛋白质或结构域的那些核酸分子。

[0134] 本发明还涉及使用一种新方法鉴定具有在结构,结构域组成和/或功能上与 *Schizochytrium* PUFA PKS 系统同源的 PUFA PKS 系统的微生物。在一个实施方案中,该微生物是非细菌微生物,且优选的是,以该方法鉴定的微生物是真核微生物。另外,本发明涉及以该方法鉴定的微生物和这些微生物及来自这些微生物的 PUFA PKS 系统在用于根据本发明的 PUFA PKS 系统的各种应用(例如,遗传修饰的生物体和产生生物活性分子的方法)中的用途。本文所述并证实的独特筛选方法使得能够迅速鉴定含有与本发明的 *Schizochytrium* PUFA PKS 系统同源的 PUFA PKS 系统的新微生物菌株。申请人使用该方法发现并在本文中公开了破囊壶菌属微生物含有与 *Schizochytrium* 中的发现同源的 PUFA PKS 系统。该发现在下面实施例 2 中详细描述。

[0135] 通过单独使用下列方法或者以这些方法的任意组合可容易地鉴定/分离/筛选具有与 *Schizochytrium* 中的发现相似的 PUFA PKS 系统的微生物,例如本发明人发现并在实施例 2 中描述的破囊壶菌属微生物。

[0136] 一般来说,鉴定具有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的非细菌微生物的方法包括第一步 (a) 选择产生至少一种 PUFA 的微生物;和第二步 (b) 从 (a) 鉴定与发酵培养基中在大于 5% 饱和度,更优选 10% 的饱和度,更优选大于 15% 的饱和度,且更优选大于 20% 的饱和度的溶氧条件下所述微生物产生的 PUFA 相比,在发酵培养基中不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。产生至少一种 PUFA 且在不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物鉴定为含有 PUFA PKS 系统的候选微生物。鉴定含有 PUFA PKS 系统的优良候选微生物后,该方法可包括附加步骤 (c) 检测在步骤 (b) 中鉴定的生物是否含有 PUFA PKS 系统。

[0137] 在本发明的一个实施方案中,通过在低氧/缺氧条件下和有氧条件下培养筛选过程中选择的该微生物,并进一步测量该生物体中 PUFA 的含量进行步骤 (b),测定脂肪酸的分布特征,以及脂肪含量。通过比较在低氧/缺氧条件下的结果与在有氧条件下的结果,该方法提供了试验微生物是否含有本发明的 PUFA PKS 系统的有力征兆。这一优选的实施方案在下面详细描述。

[0138] 首先,在有氧条件下培养需要检查是否存在 PUFA PKS 系统的微生物菌株以诱导产生大量的细胞(微生物生物量)。作为该鉴定方法的一部分,随后将这些细胞放在低氧或缺氧培养条件下(例如不超过约 5% 饱和度的溶氧,更优选不超过约 2%,甚至更优选不超过约 1%,且最优选在培养基中约 0% 饱和度的溶氧)并使其再生长约 24-72 小时。在该方法中,微生物应在高于约 15°C,且更优选高于约 20°C,且甚至更优选高于约 25°C,且甚至更优

选高于约 30℃ 的温度下培养。在能在培养箱中（且因此在培养物中）诱导这类大气环境的培养箱中或者通过以在培养瓶 / 管本身中直接诱导该低氧环境的方式培养该细胞可容易地维持低氧或缺氧培养环境。

[0139] 在优选的培养方法中，微生物可在摇瓶中培养，它不同于一般含有少量培养基（不超过约 50% 的总容积且通常不超过约 25% 的总容积）以便在摇床上摇动时保持培养基有氧，该摇瓶反而超过其容积的约 50%，且更优选超过约 60%，且最优选超过其容积的约 75% 装满培养基。用培养基高装载该摇瓶防止它放在摇床上时在瓶中充分混合，从而防止氧扩散进培养基。因此随着微生物的生长，它们用尽培养基中已有的氧并自然在摇瓶中产生一个低氧或无氧的环境。

[0140] 培养期后，收获细胞并分析目的生物活性化合物（例如，脂类）的含量，但最具体的是含有两个或多个不饱和键，且更优选 3 个或更多双键，且甚至更优选 4 个或更多双键的化合物。对于脂类，将具有高于该微生物干重的约 5%，且更优选高于约 10%，且更优选高于约 15%，且甚至更优选高于约 20% 的该化合物的那些菌株鉴定为预期含有上述类型的新 PKS 系统。对于其它生物活性化合物，例如抗生素或以更少量合成的化合物，将具有高于该微生物干重的约 0.5%，且更优选高于约 0.1%，且更优选高于约 0.25%，且更优选高于约 0.5%，且更优选高于约 0.75%，且更优选高于约 1%，且更优选高于约 2.5%，且更优选高于约 5% 的该化合物的那些菌株鉴定为预期含有上述类型的新 PKS 系统。

[0141] 作为另一选择，或者与方法相结合，通过检查该菌株的脂肪酸分布特征（通过培养该生物体或者通过公开的或其它容易获得的来源获得）可鉴定预期含有本文所述新 PUFA PKS 系统的微生物菌株。如果该微生物含有高于约 30%，且更优选高于约 40%，且更优选高于约 45%，且甚至更优选高于约 50% 的其总脂肪酸是 C14:0, C16:0 和 / 或 C16:1，尽管也产生至少一种具有三个或更多个不饱和键，且更优选 4 个或更多个双键，且更优选 5 个或更多个双键，且甚至更优选 6 个或更多个双键的长链脂肪酸，那么将该微生物菌株鉴定为具有本发明所述类型的新 PUFA PKS 系统的可能候选菌株。在上述低氧条件下筛选该生物体，并证实产生了含有两个或更多个不饱和键的生物活性分子，这表明该生物体中存在新的 PUFA PKS 系统，通过分析该微生物的基因组可进一步证实这一点。

[0142] 通过筛选已知含有 C17:0 和 / 或 C17:1 脂肪酸（与上述高百分比的 C14:0, C16:0 和 C16:1 脂肪酸相结合）的真核菌株也可提高该方法的成功率，因为 C17:0 和 C17:1 脂肪酸是细菌（原核）为基础的或作用的脂肪酸生产系统的有效标记。鉴定含有新 PUFA PKS 系统的菌株的另一标记是该生物体产生简单的脂肪酸分布特征。根据本发明，“简单的脂肪酸分布特征”定义为该菌株以高于总脂肪酸 10% 的水平产生 8 种或更少的脂肪酸。

[0143] 使用这些方法或标记的任一种（单独或者优选相结合）使得本领域的技术人员能够容易地鉴定可信地预期含有本发明所述类型的新 PUFA PKS 系统的微生物菌株。

[0144] 在结合上述众多方法和标记的一个优选的实施方案，建立了一个新的生物学合理的筛选系统（使用摇瓶培养物）用于检测含有产生 PUFA 的 PKS 系统的微生物。该筛选系统按如下进行：

[0145] 将待测菌株 / 微生物的一部分培养物放在含有 50mL 培养基的 250mL 带挡板的摇瓶中（有氧处理），并将相同菌株的另一部分培养物放在含有 200mL 培养基的 250mL 无挡板的摇瓶中（缺氧 / 低氧处理）。根据被评估的微生物的类型和菌株采用不同的培养基。将

两种摇瓶放在 200rpm 的摇床上。培养 48-72 小时后,离心收获培养基并通过气相色谱法分析细胞的脂肪酸甲基酯含量以测定各培养物的下列数据:(1) 脂肪酸分布特征;(2) PUFA 含量;和 (3) 脂肪含量(约计为脂肪酸总量/细胞干重)。

[0146] 然后分析这些数据并回答下列 5 个问题(是/否):

[0147] 比较低氧/缺氧瓶的数据与有氧瓶的数据:

[0148] (1) 与有氧培养物相比低氧培养物中的 DHA(或其它 PUFA 含量)(以% FAME(脂肪酸甲酯))是否保持约相同或优选提高?

[0149] (2) 缺氧培养物中的 C14:0+C16:0+C16:1 是否超过约 40% 的 TFA?

[0150] (3) 在缺氧培养物中对于常规氧依赖型延长酶/去饱和酶途径是否有极少(FAME < 1%)或者没有前体(C18:3n-3+C18:2n-6+C18:3n-6)?

[0151] (4) 与有氧培养相比低氧培养中的脂肪含量(以脂肪酸总量/细胞干重表示)是否提高?

[0152] (5) 与有氧培养相比低氧培养中以占细胞干重百分数表示的 DHA(或其它 PUFA 含量)是否增加?

[0153] 如果前 3 个问题的答案为是,则它是该菌株含有形成长链 PUFA 的 PKS 遗传系统的良好征兆。答案为是的问题越多(优选前三个问题的答案必须为是),该菌株含有这种 PKS 遗传系统的征兆就越强。如果 5 个问题的答案都为是,那么就是该菌株含有形成长链 PUFA 的 PKS 遗传系统的极强征兆。缺少 18:3n-3/18:2n-6/18:3n-6 表明低氧条件关闭或者抑制了常规的 PUFA 合成途径。高 14:0/16:0/16:1 脂肪是细菌作用的脂肪酸合成分布特征(存在 C17:0 和 17:1 也是它的指标)和简单的脂肪酸分布特征的初步指标。在低氧条件下 PUFA 合成和含有 PUFA 的脂肪合成增加是 PUFA PKS 系统的直接征兆,因为该系统不需要氧形成高度不饱和脂肪酸。

[0154] 最后,在本发明的鉴定方法中,一旦鉴定了优良的候选菌株,优选筛选该微生物以检测该微生物是否含有 PUFA PKS 系统。例如,可筛选该微生物的基因组以检测是否存在编码本文所述 PUFA PKS 系统的结构域的一种或多种核酸序列。优选的是,该检测步骤包括合适的核酸检测方法,例如,对目的微生物中的一种或多种核酸序列的杂交,扩增和测序。用于该检测方法的探针和/或引物可来源于任何已知的 PUFA PKS 系统,包括美国专利号 6,140,486 所述海洋细菌 PUFA PKS 系统,或美国申请系列号 09/231,899 和本文所述 Thraustochytrid PUFA PKS 系统。一旦鉴定了新的 PUFA PKS 系统,来自这些系统的遗传物质也可用于检测其它新的 PUFA PKS 系统。以鉴定和检测序列为目的的核酸杂交,扩增和测序方法是本领域熟知的。使用这些检测方法,可评估序列同源性和结构域的结构(例如,各种 PUFA PKS 功能域的存在,数目和/或排列)并与本文所述已知 PUFA PKS 系统进行比较。

[0155] 在一些实施方案中,可使用生物学试验鉴定 PUFA PKS 系统。例如,在美国申请系列号 09/231,899 的实施例 7 中,描述了使用一些类型的脂肪酸合成系统的熟知抑制剂,即,硫乳霉素的关键实验的结果。本发明人表明在 Schizochytrium 的所有细胞中可特异性抑制 PUFA 的合成而不抑制短链饱和脂肪酸的合成。该结果具有如下意义:本发明人从 Schizochytrium 的 cDNA 序列分析中了解到在 Schizochytrium 中存在 I 型脂肪酸合酶系统。已知硫乳霉素不能抑制 I 型 FAS 系统,这与本发明人的数据一致,即,用硫乳霉素处理没有抑制饱和脂肪酸(在 Schizochytrium 中主要是 C14:0 和 C16:0)的生产。在文献和本

发明人自己的数据中没有硫乳霉素对 C14:0 或 C16:0 脂肪酸的延长或其去饱和化（即，短链饱和脂肪酸通过传统途径转变成 PUFA）具有任何抑制作用的征兆。因此，硫乳霉素强烈抑制 *Schizochytrium* 中的 PUFA 生产的事实表明传统 PUFA 合成途径在 *Schizochytrium* 中不能产生 PUFA，而是涉及一个不同的合成途径。另外，以前已经确定硫乳霉素抑制希瓦氏菌属 PUFA PKS 系统（注意本发明的 PUFA PKS 系统同时具有 I 型和 II 型系统的因子），且已知硫乳霉素是 II 型 FAS 系统（例如在大肠杆菌中发现的系统）的抑制剂。因此，该实验表明 *Schizochytrium* 以不涉及 I 型 FAS 的途径产生 PUFA。使用相似的理论和检测步骤可检测使用本文公开的新筛选方法鉴定的微生物中的 PUFA PKS 系统。

[0156] 另外，实施例 3 显示了其它的生化数据，该数据提供了 *Schizochytrium* 中的 PUFA 不通过传统途径产生的证据（即，在所有细胞中没有观察到 C16:0 与 DHA 之间的前体产物动力学，且体外 PUFA 合成可从膜成份分离——除了插入一系列双键中的第一个双键的 $\delta 9$ 去饱和酶外，传统 PUFA 合成途径的所有脂肪酸去饱和酶都与细胞膜相联）。这类生化数据可用于在以上述新筛选方法中鉴定的微生物中检测 PUFA PKS 活性。

[0157] 需要使用本发明的筛选 / 鉴定方法筛选的优选微生物菌株选自：细菌，藻类，真菌，原生动物或原生生物组成的组中，但最优选从由藻类，真菌，原生动物和原生生物组成的真核微生物中选择。这些微生物优选在高于约 15°C，更优选高于约 20°C，甚至更优选高于约 25°C 且最优选高于约 30°C 的温度下能够生长并产生含有两个或多个不饱和键的生物活性化合物。

[0158] 在本发明该方法的一些实施方案中，可在超过约 20°C，优选超过约 25°C 且甚至更优选超过约 30°C 的温度下产生 PUFA 的细菌中鉴定新的细菌 PUFAPKS 系统。如本文前面所述，美国专利 6,140,486 中所述海洋细菌，希瓦氏菌属和 *Vibrio marinus* 在较高温度下不产生 PUFA，这限制了来自这些细菌的 PUFA PKS 系统的有用性，特别是在野外条件下的植物应用中的有用性。因此，在一个实施方案中，本发明的筛选方法可用于鉴定具有 PUFA PKS 系统且能在较高温度（例如，超过约 20, 25, 或 30°C）下生长并产生 PUFA 的细菌。在该实施方案中，可将诸如制霉菌素（抗真菌剂）或放线菌酮（真核蛋白合成抑制剂）的真核生物生长抑制剂加入琼脂板中用于从下文所述生境 / 生态位类型收集的水样品 / 土壤样品培养 / 选择起始菌株。该方法可帮助选择没有（或有最小的）真核菌株污染的细菌菌株富集物。该选择方法与在高温（例如，30°C）下培养平板，然后选择产生至少一种 PUFA 的菌株相结合可初步鉴定具有在高温下有效的 PUFA PKS 系统的候选细菌菌株（与现有技术中仅在不超过约 20°C 且更优选低于约 5°C 的温度下表现出 PUFA 生产的那些细菌菌株相反）。

[0159] 收集优选的微生物类型用于筛选根据本发明的 PUFA PKS 系统的地点包括下列任一处：低氧环境（或靠近这些类型的低氧环境的地点，包括在动物的肠道中，该动物包括消耗微生物或含有微生物的食物的无脊椎动物（包括滤食性生物类型）），含有低氧或无氧的水中生境（包括淡水，盐湖和海洋），且特别是处于或者靠近海洋中的低氧环境（区域）。该微生物菌株优选不是专性厌氧菌，而是可适应在有氧及低氧或无氧环境中均可存活。同时含有有氧及低氧或无氧环境的土壤环境也是发现这些生物体的优良环境，特别是在含水生境或暂时含水生境的这些土壤类型中。

[0160] 特别优选的微生物菌株可以是在其部分生活周期期间它能够通过诸如吞噬作用，吞噬营养或内吞能力的机制消耗整个细菌细胞（食细菌生物）和 / 或在其生活周期中具有

以变形虫状阶段或裸露原生质体存在的一个时期的一种菌株（选自藻类，真菌（包括酵母），原生动物或原生生物组成的组中）。该营养方法如果发生错误且该细菌细胞（或其DNA）不能被消化反而有效掺入真核细胞中将极大地增加将细菌 PKS 系统转移进真核细胞的可能性。

[0161] 能够食细菌（特别是通过吞噬作用或内吞）的微生物菌株（不是 Thraustochytrids 的成员）可在下列微生物纲（包括但不限于举例的属）中找到：

[0162] 在藻类和藻类样微生物（包括 stramenopiles）中：裸藻纲（例如眼虫属（Euglena），和 Peranerna），金藻纲（例如赭单胞菌属（Ochromonas）），Dinobryaceae 纲（例如 Dirzobryon, Platychrysis, 和 Chrysochromulina 属），甲藻纲（包括 Crypthecodinium, Gymnodinium, Peridinium, Ceratium, Gyrodinium, 和 Oxyrrhis 属），隐藻纲（例如隐藻属（Cryptomonas），和 Rhodomonas），黄藻纲（例如 Olisthodiscus 属）（且包括存在变形虫状阶段的藻类形式，如鞭毛虫 Rhizochloridaceae, 和 Aphanochaete pascheri, Bumilleria stigeoclonium 和 Vaucheria geminata 的游动孢子 / 配子），Eustigmatophyceae 纲, 和 Prymnesiophyceae 纲（包括 Prymnesium 和 Diacronema 属）。

[0163] 在 Stramenopiles 中包括：Proteromonads, Orpalines, Developayella, Diplophorys, Larbrinthulids, Thraustochytrids, Bicosocids, 卵菌纲, Hypochytridiomycetes, Commation, Reticulosphaera, Pelagomonas, Pelapococcus, Ollicola, Aureococcus, Parmales, Raphidiophytes, Synurids, Rhizochromulinaales, Pedinellales, Dictyochales, Chrysomeridales, Sarcinochrysidales, Hydrurales, Hibberdiales, 和 Chromulinales。

[0164] 在真菌中有：粘菌纲（形成粘变形体）- 粘液霉菌, 聚粘菌亚纲, 包括 Acrasiceae 目（例如 Sappinia 属），Guttulinaceae 纲（例如 Guttulinopsis, 和 Guttulina 属），Dictysteliaceae 纲（例如宿曲滴菌属（Acrasis），盘基网柄菌属（Dictyostelium），Polysphondylium, 和 Coenonia），和 Phycomyces 纲, 包括壶菌目, Ancylistales, Blastocladales, Monoblepharidales, 水霉目, 霜霉目, 毛霉菌目, 和虫霉目。

[0165] 在原生动物中有：具有能够食细菌（包括通过吞噬作用）的生活阶段的原生动物株可选自分类为纤毛虫, 鞭毛虫或变形虫的类型。原生动物纤毛虫包括的种类：漏斗虫, 肾形虫, 管口虫, Haptorids, 核残迹虫, 寡膜虫, Polyhymenophora（旋毛虫）, Prostomes 和吸管纲。原生动物鞭毛虫包括 Biosocids, Bodonids, 单鞭滴虫, Chrysophytes（例如 Anthophysa 属, Chrysamoeba, 金球藻虫属, 树滴虫属, 锥囊藻虫属, 鱼鳞藻虫属, 赭滴虫属, Paraphysomonas, Poterioochromonas, Spumella, Syncrypta, 黄群藻虫属, 和 Uroglena），领鞭毛虫, Cryptophytes（例如唇滴虫属, 隐滴虫属, Cyanomonas, 和 Goniomonas）, Dinoflagellates, Diplomonads, Euglenoids, Heterolobosea, Pedinellids, Pelobionts, 胶领鞭虫, Pseudodendromonads, Spongomonads 和 Volvocales（和其它鞭毛虫, 包括未分类的鞭毛虫 Artodiscus 属, Clautriavia, 曳鞭毛虫属, Kathablepharis 和 Multicilia）。变形虫状原生动物包括的种类有：太阳虫, Centrohelids, Desmothoracids, Diplophryids, Eumamoebae, Heterolobosea, Leptomyxids, Nucleariids filose 变形虫, Peleobionts, Testate 变形虫和 Vampyrellids（且包括未分类的变形虫属 Gymnophrys, Biomyxa, Microcometes, Reticulomyxa, Belonocystis, Elaeorhaxis, Allelogromia,

Gromia 或 Lieberkuhnia)。原生动物包括如下的目:Percolomonadeae, Heterolobosea, Lyromonadea, Pseudociliata, Trichomonadea, Hypermastigae, Heteromiteae, Telonemea, Cyathobodonea, Ebridea, Ppytomyxea, Opalineae, Kinetomonadea, Hemimastigae, Protostelea, Myxagastrea, Dictyostelea, Choanomonadea, Apicomonadea, Eogregarinea, Neogregarinea, Coelotrolphea, Eucoccidea, 血孢子虫亚目, Piroplasmea, Spirotrichea, Prostomatea, Litostomatea, Phyllopharyngea, Nassophorea, Oligohymenophorea, Colpodea, Karyorelicta, Nucleohelea, Centrohelea, Acantharea, Sticholonchea, Polycystinea, Phaeodarea, Lobosea, Filosea, Athalamea, Monothalamea, Polythalamea, Xenophyophorea, Schizocladea, Holosea, Entamoebae, 粘盘孢目, 纺线孢子目, Halosporea, Paramyxea, Rhombozoa 和 Orthonectea。

[0166] 本发明的一个优选的实施方案包括从上述一种优选的生境中收集的上述微生物的菌株。

[0167] 本发明的一个实施方案涉及使用上述新 PUFAPKS 筛选方法鉴定的任何微生物, 涉及 PUFA PKS 基因及其编码的蛋白质, 且涉及该微生物和 / 或 PUFAPKS 基因和蛋白质 (包括其同源物和片段) 在本文所述任一方法中的用途。具体地说, 本发明包含以本发明的筛选方法鉴定的生物体, 该生物体随后被遗传修饰成通过所述 PUFA PKS 系统调节生物活性分子的生产。

[0168] 本发明的另一实施方案还涉及一种分离的核酸分子, 该分子含有编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一种生物学活性结构域或其生物学活性片段的核酸序列。如上所述, 本发明人成功使用该方法鉴定了一种具有 PUFA PKS 系统的非细菌微生物, 该系统用于鉴定含有 PUFA PKS 系统的 Thraustochytriales 目的其它成员。实施例 2 描述了三种该微生物的鉴定。具体地说, 本发明人使用本发明的筛选方法鉴定了认为极有可能含有 PUFA PKS 系统的破囊壶菌 23B(ATCC20892), 接着检测了破囊壶菌属种类 23B 基因组中与本文公开的 Schizochytrium PUFA PKS 基因杂交的序列。还将 Schizochytrium lirinacium(IFO 32693) 和 Ulkenia(BP5601) 鉴定为含有 PUFA PKS 系统的良好候选菌株。根据这些数据和 Thraustochytriales 目的成员之间的相似性, 据信现在使用本发明提供的方法和工具可容易地鉴定许多其它 Thraustochytriales 的 PUFAPKS 系统。因此, 本发明包括 Thraustochytriales 的 PUFAPKS 系统和其部分和 / 或其同源物 (例如, 蛋白质, 结构域及其片段), 含有该系统和其部分和 / 或其同源物的遗传修饰的生物体, 使用该微生物和 PUFA PKS 系统的方法。

[0169] 开发导致对 Thraustochytrids 的分类学进行了修正。分类学家将 Thraustochytrids 放在藻类或藻类状原生生物中。然而, 由于分类不确定, 对于本发明最好认为本发明所述该菌株是 Thraustochytrids (目: Thraustochytriales; 科: Thraustochytriaceae; 属: 破囊壶菌属, Schizochytrium, Labyrinthuloides, 或 Japonochytrium)。对于本发明, labyrinthulids 的许多成员被认为包含在 Thraustochytrids 中。分类学改变在下文中概括。本文公开的某些单细胞微生物菌株是 Thraustochytriales 目的成员。Thraustochytrids 是分类史进化的海洋真核生物。Moss(1986), Bahnweb 和 Jackle(1986), 以及 Chamberlain 和 Moss(1988) 对

Thraustochytrids 的分类地位问题进行了综述。根据本发明,短语“Thraustochytrid”,“Thraustochytriales 微生物”和“Thraustochytriales 目的微生物”可互换使用。

[0170] 为了方便,分类学家首先将 Thraustochytrids 与藻菌目(藻类状真菌)的其它无色游动孢子真核生物放在一起。然而,藻菌目这一名称最后从分类学地位中去掉,并将 Thraustochytrids 保留在卵菌纲(双鞭毛游动孢子真菌)中。最初假定卵菌纲与异鞭毛藻类有亲缘关系,且由 Barr(Barr,1981, Biosystems 14:359-370)总结的,广泛的超微结构和生化研究最终支持了这一假定。事实上卵菌纲被 Leedale(Leedale,1974, Taxon 23:261-270)和其他藻类学家作为异鞭毛藻类的一部分接受。然而,事实上由于其异氧特性的方便性,卵菌纲和 Thraustochytrids 被真菌学家(研究真菌的科学家)而不是藻类学家(研究藻类的科学家)大量研究。

[0171] 从另一分类学角度来看,进化生物学家对于真核生物如何进化形成了两个系统学派。一个学说认为膜结合细胞器通过一系列的内共生的外生起源(Margulis,1970, Origin of Eukaryotic Cells, Yale University Press, New Haven);例如线粒体起源于细菌的细胞内共生生物,叶绿体起源于蓝藻类,鞭毛起源于螺旋体。另一学说提出膜结合的细胞器从原核祖先的无膜结合的系统通过自生加工逐渐进化(Cavalier-Smith,1975, Nature(Lond.)256:462-468)。然而两组进化生物学家都将卵菌纲和 Thraustochytrids 从真菌中取出并与 chromophyte 藻类一起放在 Chromophyta 界中(Cavalier-Smith,1981, BioSystems 14:461-481)(该界最近扩展到包括其它原生生物且该界的成员现在称为 Stramenopiles)或与所有藻类一起放在 Protoctista 界中(Margulis 和 Sagen,1985, Biosystems 18:141-147)。

[0172] 随着电子显微镜的发展,对 Thraustochytrids 的两个属,即,破囊壶菌属和 Schizochytrium 的游动孢子超微结构的研究(Perkins,1976,第 279-312 页,见“Recent Advances in Aquatic Mycology”(ed.E.B.G Jones), John Wiley & Sons, New York; Kazama,1980, Can. J. Bot. 58:2434-2446; Barr,1981, Biosystems 14:359-370)提供了 Thraustochytriaceae 与卵菌纲仅具有较远的亲缘关系的良好证据。另外,对 5S 核糖体 RNA 序列的对应分析(多元统计形式)得到的遗传数据表明 Thraustochytriales 明显是一个独特的真核类型,完全独立于真菌,且与红藻和褐藻,以及与卵菌纲的成员有最近的亲缘关系(Mannella,等,1987, Mol. Evol. 24:228-235)。大多数分类学家同意将 Thraustochytrids 从卵菌纲中取出(Bartnicki-Garcia,1987,第 389-403 页,见“Evolutionary Biology of the Fungi”(eds. Rayner, A. D. M., Brasier, C. M. & Moore, D.), Cambridge University Press, Cambridge)。

[0173] 总之,采用 Cavalier-Smith 的分类系统(Cavalier-Smith,1981, BioSystems 14:461-481,1983; Cavalier-Smith,1993, Microbiol Rev. 57:953-994),将 Thraustochytrids 与 chromophyte 藻类一起分类进 Chromophyta(Stramenopiles) 界中。最近 Cavalier Smith 等使用 Heterokonta 的 18s rRNA 标记证实 Thraustochytrids 是 chromists 而不是真菌再次确认了该分类地位(Cavalier-Smith 等,1994, Phil. Tran. Roy. Soc. London Series BioSciences 346:387-397)。该分类将它们放在与真菌完全不同的一个界中,而真菌都放在 Eufungi 界中。因此 Thraustochytrids 的分类地位总结如下:

[0174] 界:Chromophyta(Stramenopiles)

[0175] 门:Heterokonta

[0176] 目:Thraustochytriales

[0177] 科:Thraustochytriaceae

[0178] 属:破囊壶菌属, Schizochytrium, Labyrinthuloides, 或 Japonochytrium

[0179] 一些早期的分类学家将破囊壶菌属的一些原始的成员(具有变形虫状生活阶段的成员)分入一个称为 Ulkenia 的独立属中。然而,现在知道大多数(如果不是全部的话)Thraustochytrids(包括破囊壶菌属和 Schizochytrium)表现出变形虫状阶段且有人认为 Ulkenia 本身不是一个正确的属。本文使用的破囊壶菌属将包括 Ulkenia。

[0180] 尽管门和界的高级分类中的分类学地位不确定,但是 Thraustochytrids 保持特有的和特征性的分类,其成员可分类在 Thraustochytriales 目内。

[0181] 多不饱和脂肪酸(PUFA)是高等真核生物中的基本膜成份和许多脂类衍生的信号分子的前体。本发明的 PUFA PKS 系统使用不需要饱和脂肪酸去饱和并延长的 PUFA 合成途径。该途径由结构和机制上均不同于以前认识的 PKSs 的 PUFA PKSs 催化。顺式双键的产生表明包含位置特异性异构酶,据信这些酶用于产生新家族的抗生素。

[0182] 为了使用本发明的 PUFA PKS 系统产生明显高产的各种生物活性分子,可对一种生物体,优选微生物或植物,进行遗传修饰以影响 PUFA PKS 系统的活性。在一个方面,该生物体内源性含有并表达 PUFA PKS 系统,且该遗传修饰可以是对内源性 PUFA PKS 系统的一种或多种功能域的遗传修饰,因此该修饰对 PUFA PKS 系统的活性具有某种影响。在另一方面,该生物体可内源性含有并表达 PUFA PKS 系统,且该遗传修饰可以是导入至少一种外源性核酸序列(例如,重组核酸分子),其中该外源性核酸序列编码第二种 PKS 系统的至少一个生物学活性结构域或蛋白质和/或影响所述 PUFA PKS 系统的活性的蛋白质(例如,磷酸泛酰巯基乙胺转移酶(PPTase),下文讨论)。在还有另一实施方案中,该生物体不必内源性(天然)含有 PUFA PKS 系统,而是经遗传修饰导入编码具有 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列的至少一个重组核酸分子。在该方面,通过在生物体中导入或提高 PUFA PKS 活性影响 PUFA PKS 的活性。与这些方面中任一种相关的各种实施方案在下面更详细地讨论。

[0183] 因此,根据本发明,一个实施方案涉及遗传修饰的微生物,其中该微生物表达含有多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个生物学活性结构域的 PKS 系统。PUFA PKS 系统的至少一个结构域由选自如下的核酸序列编码:(a) 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的核酸序列;(b) 编码来自以本发明的筛选方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列;(c) 编码与选自:SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和, (d) 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 和 SEQ ID NO:32 组成的组中的氨基酸序列有至少约 60% 相同性的氨基酸序列的核酸序列;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。该遗传修饰影响生物体中 PKS 系统的活性。(b) 部分中提及的筛选方法已在上文中详细描述且包括如下步骤:(a) 选择产生

至少一种 PUFA 的微生物 ;和, (b) 从 (a) 中鉴定与发酵培养基中在大于约 5% 饱和度, 且优选约 10%, 且更优选约 15%, 且更优选约 20% 的饱和度的溶氧条件下微生物产生的 PUFA 相比, 在发酵培养基中不超过约 5% 饱和度的溶氧条件下具有产生增加的 PUFA 的能力的微生物。该遗传修饰的微生物可包括任意一种或多种上文鉴定的核酸序列, 和 / 或任一上文详细描述 Schizochytrium PUFA PKS ORFs 或结构域的任一其它同源物。

[0184] 本文所用的遗传修饰的微生物包括遗传修饰的细菌, 原生生物, 显微藻类, 真菌, 或其它微生物, 且特别是本文所述 Thraustochytriales 目 (例如, Thraustochytrid) 的任一属 (例如, Schizochytrium, 破囊壶菌属, Japonochytrium, Labyrinthuloides)。该遗传修饰的微生物具有从其正常 (即野生型或天然) 形式进行了修饰 (即, 突变或改变) 以便达到所需结果 (即, 增加或修饰的 PUFA PKS 活性和 / 或使用 PKS 系统产生所需产物) 的基因组。使用传统的菌株开发和 / 或分子遗传技术可完成对微生物的遗传修饰。该技术是本领域已知的且一般公开用于微生物, 例如, 参见 Sambrook 等, 1989, Molecular Cloning: A Laboratory Manual, 冷泉港实验室出版。该参考文献 Sambrook 等, 出处同上在本文中引用以其整体作为参考。遗传修饰的微生物包括这样的微生物, 其中的核酸分子以该修饰提供微生物内的所需效果的方式被插入, 缺失或修饰 (即, 突变 ;例如, 通过插入, 缺失, 取代, 和 / 或核苷酸倒置)。

[0185] 根据本发明修饰的优选微生物宿主细胞包括, 但不限于, 任一细菌, 原生生物, 显微藻类, 真菌, 或原生动物。在一个方面, 遗传修饰的优选微生物包括, 但不限于, Thraustochytriales 目的任一微生物。用于本发明的特别优选的宿主细胞可包括来自如下属的微生物, 该属包括, 但不限于 :破囊壶菌属, Labyrinthuloides, Japonochytrium, 和 Schizochytrium。这些属内优选的种类包括, 但不限于 :任一 Schizochytrium 种类, 包括 Schizochytrium aggregatum, Schizochytrium limacinum, Schizochytrium minutum ;任一破囊壶菌属种类 (包括以前的 Ulkenia 种类, 例如 U. Visurgensis, U. Amoeboida, U. Sarkariana, U. Profunda, U. radiata, U. Minuta 和 Ulkenia sp. BP-5601), 且包括 Thraustochytrium stratum, Thraustochytrium aureum, Thraustochytrium roseum ;和任一 Japonochytrium 种类。特别优选的 Thraustochytriales 菌株包括, 但不限于 :Schizochytrium sp. (S31) (ATCC 20888) ;Schizochytrium sp. (S8) (ATCC 20889) ;Schizochytrium sp. (LC-RM) (ATCC 18915) ;Schizochytrium sp. (SR21) ;Schizochytrium aggregatum (Goldstein et Belsky) (ATCC 28209) ;Schizochytrium limacinum (Honda et Yokochi) (IF032693) ;破囊壶菌属种类 (23B) (ATCC 20891) ;Thraustochytrium striatum (Schneider) (ATCC 24473) ;Thraustochytrium aureum (Goldstein) (ATCC 34304) ;Thraustochytrium roseum (Goldstein) (ATCC 28210) ;和 Japonochytrium sp. (L1) (ATCC 28207)。用于遗传修饰的合适宿主微生物的其它例子包括, 但不限于, 酵母, 包括啤酒糖酵母 (Saccharomyces cerevisiae), 卡尔斯伯糖酵母 (Saccharomyces carlsbergensis), 或诸如念珠菌属 (Candida), 克鲁维氏酵母属 (Kluyveromyces) 的其它酵母, 或其它真菌, 例如, 诸如曲霉属 (Aspergillus), 链孢霉属 (Neurospora), 青霉属 (Penicillium) 的丝状真菌, 等。细菌细胞也可用作宿主。它包括可用于发酵方法的大肠杆菌。另外, 诸如乳酸杆菌属 (Lactobacillus) 种类或杆菌属 (Bacillus) 种类的宿主也可用作宿主。

[0186] 本发明的另一实施方案涉及遗传修饰的植物,其中该植物被遗传修饰成重组表达含有多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个生物活性结构域的 PKS 系统。该结构域由选自如下的一种核酸序列编码:(a) 编码来自 *Thraustochytrid* 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;(b) 编码来自本文所述筛选和选择方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列(参见上文在遗传修饰的微生物的讨论中的方法的简要小结);(c) 编码选自由:SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, 和其生物学活性片段组成的组中的氨基酸序列的核酸序列;(d) 编码选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, 或其生物学活性片段的氨基酸序列的核酸序列;(e) 编码与选自:SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性;和/或(f) 编码与选自:SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。该遗传修饰的植物可包括任意一种或多种上文鉴定的核酸序列,和/或任一上文详细描述 *Schizochytrium* PUFA PKS ORFs 或结构域的任一其它同源物。

[0187] 本文所用的遗传修饰的植物可包括任一遗传修饰的植物,包括高等植物且特别是任何可消费的植物或用于产生本发明的所需生物活性分子的植物。该遗传修饰的植物具有从其正常(即野生型或天然)形式进行了修饰(即,突变或改变)以便实现所需结果(即,增加或修饰的 PUFA PKS 活性和/或使用该 PKS 系统产生所需产物)的基因组。使用传统的品系开发和/或分子遗传技术可完成对植物的遗传修饰。产生将编码所需氨基酸序列的重组核酸分子插入植物基因组的转基因植物的方法是本领域已知的。根据本发明用于遗传修饰的优选植物优选是适合于被包括人类的动物消费的植物。

[0188] 根据本发明被遗传修饰的优选植物(即,植物宿主细胞)包括,但不限于任何高等植物,且特别是可消费的植物,包括农作物植物且特别是使用其油类的植物。该植物可包括,例如:canola,大豆,油菜,亚麻,玉米,红花,向日葵和烟草。其它优选的植物包括已知产生用作药物试剂,调味剂,neutraceutical 试剂,功能食品成份或美容活性剂的那些植物或遗传修饰成产生这些化合物/试剂的植物。

[0189] 根据本发明,遗传修饰的微生物或植物包括使用重组技术修饰的微生物或植物。本文使用的导致基因表达,基因功能,或基因产物(即,由该基因编码的蛋白)功能降低的遗传修饰可指基因的失活(完全或部分),缺失,中断,阻断或下调。例如,导致该基因编码的蛋白质的功能下降的基因遗传修饰可由该基因的完全缺失(即,该基因不存在,且因此该蛋白质不存在),导致蛋白质不完全翻译或不翻译(例如,蛋白质不表达)的基因突变,或降低或消除该蛋白的天然功能(即表达降低或无酶活性或作用的蛋白质)的基因突变引起。导致基因表达或功能增强的遗传修饰可指基因的扩增,过度表达,激活,增强,增加,或上调。

[0190] 根据本发明的微生物或植物的遗传修饰优选影响植物表达的 PKS 系统的活性,无

论该 PKS 系统是内源性且进行了遗传修饰的系统,即向生物体中导入了重组核酸分子的内源性系统,还是完全通过重组技术提供的系统。根据本发明,“影响 PKS 系统的活性”包括与不存在该遗传修饰相比引起由该生物体表达的 PKS 系统中任何可检测的或可测量的改变或修饰的任一遗传修饰。PKS 系统中可检测的改变或修饰可包括,但不限于:将 PKS 系统活性导入生物体中使得该生物体目前具有可测量的 / 可检测的 PKS 系统活性(即,遗传修饰前该生物体不含有 PKS 系统),将来自与该生物体内源性表达的 PKS 系统不同的 PKS 系统的功能域导入该生物体中使得 PKS 系统的活性得到修饰(例如,将细菌 PUFA PKS 结构域或 I 型 PKS 结构域导入内源性表达非细菌 PUFA PKS 系统的生物体中),PKS 系统产生的生物活性分子的量改变(例如,与不存在遗传修饰相比该系统产生更多(量增大)或更少(量减少)的给定产物),PKS 系统产生的生物活性分子的类型改变(例如,该系统产生新的或不同的产物,或由该系统天然产物的变异体),和 / 或 PKS 系统产生的多个生物活性分子的比率改变(例如,与不存在该遗传修饰相比该系统产生不同比率的一种 PUFA 与另一 PUFA 的比率,产生完全不同的脂类分布特性,或者与天然构型相比将各种 PUFA 放在三酰甘油的不同位置)。该遗传修饰包括任一类型的遗传修饰且特别包括通过重组技术和通过传统诱变作出的修饰。

[0191] 应注意提及 PUFA PKS 系统中功能域或蛋白质的活性增加是指在含有该结构域或蛋白质(或导入了该结构域或蛋白质)的生物体中导致该结构域或蛋白质系统的功能增强的任何遗传修饰且可包括结构域或蛋白质的活性更高(例如,比活或体内酶活性),结构域或蛋白质系统的抑制或降解减少,和该结构域或蛋白质的过度表达。例如,基因拷贝数增加,通过使用与天然启动子相比产生更高表达水平的启动子来增加表达水平,或通过遗传改造或传统诱变改变基因以增加该基因编码的结构域或蛋白质的活性。

[0192] 同样,提及 PUFA PKS 系统中功能域或蛋白质的活性降低是指在含有该结构域或蛋白质(或导入了该结构域或蛋白质)的生物体中导致该结构域或蛋白质功能降低的任何遗传修饰且包括该结构域或蛋白质的活性降低,该结构域或蛋白质的抑制或降解增加,和该结构域或蛋白质的表达减少或消除。例如,通过抑制或减少该结构域或蛋白质的生产,“敲除”编码该结构域或蛋白质的基因或其部分,降低结构域或蛋白质活性,或抑制该结构域或蛋白质的活性可减小本发明的结构域或蛋白质的作用。抑制或减少结构域或蛋白质的生产可包括将编码该结构域或蛋白质的基因放在需要在生长培养基中存在诱导化合物的启动子的控制下。通过形成耗尽培养基中的该诱导剂的条件,编码该结构域或蛋白质的基因表达(且因此蛋白质的合成)可被关闭。抑制或减小结构域或蛋白质的活性也可包括使用类似于在美国专利号 4,743,546 中所述切除技术方案,本文引用以供参考。使用该方法,可将编码目的蛋白的基因克隆到特定遗传序列之间,允许从基因组中特异性,受控制地切除该基因。通过,例如,按美国专利号 4,743,546 转换培养物的培养温度,或者通过一些其它的物理或营养信号可促进切除。

[0193] 在本发明的一个实施方案中,遗传修饰包括修饰编码具有本文所述非细菌 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列的核酸序列。该修饰可针对内源性(天然)表达的非细菌 PUFA PKS 系统内的氨基酸序列,因此通过,例如,传统诱变和选择技术和 / 或分子遗传技术,包括遗传工程技术可对天然含有该系统的微生物进行遗传修饰。遗传工程技术可包括,例如,使用定向重组载体缺失内源性基因的一部分,或者用异源性序

列取代内源性基因的一部分。可导入宿主基因组的异源性序列的例子包括编码来自诸如不同的非细菌 PUFA PKS 系统,细菌 PUFA PKS 系统,I 型 PKS 系统,II 型 PKS 系统,或模块 PKS 系统的另一 PKS 系统的至少一个功能域的序列。导入宿主基因组的其它异源性序列包括编码不是 PKS 系统的结构域的蛋白质或功能域,但是将影响内源性 PKS 系统活性的序列。例如,可将编码磷酸泛酰巯基乙胺转移酶的核酸分子导入宿主基因组中(下文讨论)。可对内源性 PUFA PKS 系统作出的特异性修饰在下文中详细讨论。

[0194] 在本发明的该实施方案的另一方面,该遗传修饰可包括:向宿主中导入:(1) 编码具有非细菌 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列的重组核酸分子;和/或(2) 编码影响 PUFA PKS 系统的活性的蛋白质或功能域的重组核酸分子。该宿主可包括:(1) 不表达任一 PKS 系统的宿主细胞,其中将 PKS 系统的所有功能域导入该宿主细胞中,且其中至少一种功能域来自于非细菌 PUFA PKS 系统;(2) 表达具有非细菌 PUFA PKS 系统的至少一个功能域的 PKS 系统(内源性或重组)的宿主细胞,其中导入的重组核酸分子可编码至少一种其它非细菌 PUFA PKS 结构域功能或者影响宿主 PKS 系统的活性的另一蛋白质或结构域;和(3) 表达不必包括来自非细菌 PUFAPKS 的结构域功能的 PKS 系统(内源性或重组)的宿主细胞,且其中导入的重组核酸分子包括编码非细菌 PUFA PKS 系统的至少一个功能域的核酸序列。换句话说,本发明试图包括任何遗传修饰的生物体(例如,微生物或植物),其中该生物体包含至少一种非细菌 PUFA PKS 结构域功能(内源性或者通过重组修饰),且其中当该生物体包含功能性 PKS 系统时该遗传修饰对非细菌 PUFA PKS 结构域功能或 PKS 系统具有可测量的影响。

[0195] 因此,使用本发明的非细菌 PUFA PKS 系统,例如,利用来自 *Thraustochytrid* PUFA PKS 系统的基因,可使用基因混合将 PUFA 产物的范围扩展到包括 EPA,DHA,ARA,GLA,SDA 及其它产物,以及产生广泛的生物活性分子,包括抗生素,其它药用化合物,和其它所需产物。获得这些生物活性分子的方法不仅包括混合来自各种生物的基因,而且还包括遗传修饰本文公开的非细菌 PUFA PKS 基因的方法。对本发明的非细菌 PUFAPKS 系统的遗传基础和结构域结构的了解提供了设计产生各种生物活性分子的遗传修饰的新生物体的基础。尽管本发明人预期可混合和修饰任一 PKS 结构域和有关基因,但是以举例的方式,下面讨论了在遗传修饰和生物活性分子的生产中对 PUFA-PKS 系统的各种可能的操作。

[0196] 例如,在一个实施方案中,通过修饰 CLF(链长因子)结构域改变非细菌 PUFA-PKS 系统的产物,例如由 *Thraustochytrids* 产生的那些产物。该结构域是 II 型(分离的酶) PKS 系统的特征。其氨基酸序列与 KS(酮合酶对)结构域具有同源性,但它缺乏活性位点半胱氨酸。CLF 可发挥确定延长循环次数,且因此确定最终产物的链长的功能。在本发明的该实施方案中,使用目前对 FAS 和 PKS 合成的了解情况,提供了通过定点修饰非细菌 PUFA-PKS 系统产生 ARA 的合理策略。在文献中关于 CLF 在 PKS 系统中的功能存在争论(C. Bisang 等, *Nature* 401,502(1999)) 且认识到其它结构域可能与最终产物的链长确定相关。然而,有意义的是 *Schizochytrium* 同时产生 DHA(C22:6, ω -3) 和 DPA(C22:5, ω -6)。在 PUFA-PKS 系统中,在碳链生长的合成期间导入顺式双键。由于在该分子合成的早期出现 ω -3 和 ω -6 双键的放置,因此预期它们不会影响随后的最终产物链长确定。因此,不受理论的约束,本发明人相信将指导 C20 单元(代替 C22 单元)合成的因子(例如,CLF) 导入 *Schizochytrium* PUFA-PKS 系统中将导致产生 EPA(C20:5, ω -3) 和 ARA(C20:4, ω -6)。例如,在异源性系统

中,通过将来自产生 EPA 的系统的 CLF(例如,来自发光菌属 (*Photobacterium*) 的 CLF) 直接取代进 *Schizochytrium* 基因组可利用该 CLF。然后分析所得的转化子的脂肪酸分布特征的变化以鉴定产生 EPA 和 / 或 ARA 的转化子。

[0197] 除了依赖于异源性系统(重组系统,例如可导入植物中的系统)的开发外,CLF 的概念可在 *Schizochytrium* 中利用(即,通过修饰 *Schizochytrium* 基因组)。转化和异源性重组已经在 *Schizochytrium* 中得到证实。通过构建带有 OrfB 的 CLF 被来自 C20 PUFA-PKS 系统的 CLF 取代的克隆可利用它。编码区的下游可插入标记基因。然后可转化野生型细胞,选择标记表型,然后筛选插入了该新的 CLF 的细胞。接着可分析它们对脂肪酸分布特征的任何影响以鉴定产生 EPA 和 / 或 ARA 的转化子。如果发现不同于与 CLF 相关的一些因子影响终产物的链长,则可采用相似的策略改变这些因子。

[0198] 涉及改变 PUFA-PKS 产物的另一优选的实施方案包括修饰或取代 β -羟酰-ACP 脱水酶 / 酮合酶对。在大肠杆菌的顺式-11-十八碳烯酸 (C18:1, Δ 11) 合成期间,据信顺式双键的产生依赖于特异性 DH 酶, β -羟酰-ACP 脱水酶,即 FabA 基因的产物。该酶从 β -羟酰-ACP 去掉 HOH 并在碳链中留下一个反式双键。DH 的一个亚型,即 FabA-样 DH 具有顺反异构酶活性 (Heath 等,1996,出处同上)。细菌和非细菌 PUFA-PKS 系统的一个新方面是存在两个 FabA-样 DH 结构域。不受理论的约束,本发明人相信这些 DH 结构域中的一个或两者都具有顺反异构酶活性 (DH 结构域的操作在下面进行更详细地讨论)。

[0199] 大肠杆菌中不饱和脂肪酸合成的另一方面是需要特定的 KS 酶, β -酮脂酰-ACP 合酶, FabB 基因的产物。它是完成对连接到活性位点的半胱氨酸残基(通过硫酯键)上的脂肪酸与丙二酰-ACP 进行缩合的酶。在逐步反应中,释放 CO₂ 且以两个碳延长线型链。据信只有该 KS 延长含有双键的碳链。只有当双键为顺式构型时发生该延长,如果为反式构型,在延长前双键被烯酰-ACP 还原酶 (ER) 还原 (Heath 等,1996,出处同上)。至今鉴定的所有 PUFA-PKS 系统都具有两个 KS 结构域,一个与大肠杆菌的 FabB-样 KS 比与其它 KS 具有更高的同源性。同样,不受理论的约束,本发明人相信在 PUFA-PKS 系统中, DH (FabA-样) 和 KS (FabB-样) 酶结构域的特异性和相互作用确定了终产物中顺式双键的数目和位置。由于 2-碳延长反应的次数比 PUFA-PKS 终产物中存在的双键数目更大,因此可确定在一些延长循环中发生了完全还原。因此 DH 和 KS 结构域可用作改变 DHA/DPA 比率或其它长链脂肪酸比率的靶。通过导入来自其它系统的同源性结构域或者通过对这些基因片段的诱变可修饰和 / 或评估它们。

[0200] 在另一实施方案中,可修饰或取代 ER(烯酰-ACP 还原酶,一种还原脂肪酸-ACP 中的反式双键产生完全饱和碳的酶) 结构域以改变 PKS 系统制备的产物类型。例如,本发明人已知 *Schizochytrium* PUFA-PKS 系统与以前描述的细菌系统的区别在于它具有两个(而不是一个) ER 结构域。不受理论的约束,本发明人相信这些 ER 结构域可有效影响所得的 PKS 生产产物。通过分别敲除单个结构域或者通过修饰其核苷酸序列或者通过用来自其它生物的 ER 结构域的取代可改变所得的 PKS 产物。

[0201] 在另一实施方案中,可将编码不是 PKS 系统的一部分,但影响 PKS 系统的蛋白质或结构域的核酸分子导入生物体中。例如,上文所述所有 PUFAPKS 系统含有多个,串连的, ACP 结构域。ACP(作为独立的蛋白质或者作为更大蛋白质的一个结构域) 需要附着到磷酸泛酰巯基乙胺辅因子上以产生有活性的,完整 (holo)-ACP。磷酸泛酰巯基乙胺与 apo-ACP 的附

着通过酶超家族成员,即磷酸泛酰巯基乙胺基转移酶 (PPTase) 完成 (Lambalot R. H., 等, Chemistry and Biology, 3, 923 (1996))。

[0202] 与其它 PKS 和 FAS 系统相比,本发明人推定通过特异性,内源性,PPTase 可完成对 Schizochytrium ORFA 蛋白质中存在的多个 ACP 结构域的激活。在 Schizochytrium 中尚未鉴定编码该推定的 PPTase 的基因。如果在 Schizochytrium 中存在该基因,可预料到用于尝试鉴定和克隆它的一些方法。这些方法包括 (但不限于):从活跃生长的 Schizochytrium 细胞制备的 cDNA 文库的产生和部分测序 (注意,在目前获得的 Schizochytrium cDNA 文库组中鉴定了与 PPTase 的序列具有同源性的一个序列;然而,它似乎是多结构域的 FAS 蛋白质的一部分,且本身不编码所需的 OrfA 特异性 PPTase);使用在许多 PPTase 中存在的氨基酸基序设计的简并寡核苷酸引物用于 PCR 反应 (以获得用于筛选基因组或 cDNA 文库的核酸探针分子);基于蛋白质-蛋白质相互作用的遗传方法 (例如,酵母双杂交系统),其中 ORFA-ACP 结构域可用作“钓饵”以发现“靶” (即, PPTase);以及纯化和部分测序酶本身作为产生用于筛选基因组或 cDNA 文库的核酸探针的工具。

[0203] 异源性 PPTase 也许能够激活 Schizochytrium ORFA ACP 结构域也是可预料到的。已经表明一些 PPTases,例如枯草杆菌 (*Bacillus subtilis*) 的 sfp 酶 (Lambalot 等, 出处同上) 和 *Streptomyces verticillus* 的 svp 酶 (Sanchez 等, 2001, Chemistry & Biology 8: 725-738), 具有广泛的底物耐受性。可试验这些酶以观察它们是否激活 Schizochytrium ACP 结构域。另外,最近的一篇文献描述了真菌 PKS 蛋白质在烟草中的表达 (Yalpani 等, 2001, The Plant Cell 13: 1401-1409)。在该转基因植物中检测到导入的 PKS 系统 (由展青霉 (*Penicillium patulum*) 的 6-甲基水杨酸 (6-methylsalicylic acid) 合酶基因编码) 的产物,尽管在这些植物中不存在相应的真菌 PPTase。这表明内源性植物 PPTase(s) 识别并激活真菌 PKS ACP 结构域。与该观察结果相关联,本发明人在拟南芥整体基因组数据库中鉴定了很可能编码 PPTases 的两个序列 (基因)。这些序列 (GenBank 登录号: AAG51443 和 AAC05345) 目前列为编码“未知蛋白”。根据在翻译的蛋白质序列中存在包括: G(I/V)D 和 WxxKE(A/S)xxK (SEQ ID NO: 33) 的一些标记基序 (在 Lambalot 等, 1996 中作为所有 PPTases 的特征列出) 将它们鉴定为推定的 PPTases。另外,这两个推定的蛋白质含有一般在 PPTases 中发现的另外两个基序,这两个基序一般与 PKS 和非核糖体肽合成系统相关;即 FN(I/L/V)SHS (SEQ ID NO: 34) 和 (I/V/L)G(I/L/V)D(I/L/V) (SEQ ID NO: 35)。而且,这些基序在该蛋白质序列的预期相关位置出现。很可能在诸如烟草的其它植物中存在该拟南芥基因的同源物。同样,可克隆并表达这些基因以观察它们编码的酶是否可激活 Schizochytrium ORFA ACP 结构域,或者作为选择,可在该转基因植物中直接表达 OrfA (靶向质体或细胞质)。

[0204] 可识别 ORFA ACP 结构域作为底物的另一异源性 PPTase 是念珠藻属 (*Nostoc*) 种类的 PCC 7120 (以前称为鱼腥藻属 (*Anabaena*) 种类的 PCC 7120) 的 Het I 蛋白质。正如美国专利号 6, 140, 486 中所述,希瓦氏菌属的一些 PUFA-PKS 基因与在念珠藻属中发现的 PKS 基因簇中存在的蛋白质结构域具有高度同源性 (该专利的图 2)。该念珠藻属 PKS 系统与长链 (C26 或 C28) 羟基脂肪酸的合成相关,该脂肪酸与糖类半分子酯化形成异形囊胞细胞壁的一部分。这些念珠藻属 PKS 结构域也与 Schizochytrium PKS 蛋白质的 Orfs B 和 C 中发现的结构域 (即,相应于在希瓦氏菌属 PKS 蛋白质中发现的那些结构域的相同结

构域)具有高度同源性。直到最近,在 GenBank 数据库中没有一个念珠藻属 PKS 结构域与 *Schizochytrium* OrfA 的任一结构域(或同源性希瓦氏菌属 Orf 5 蛋白质)具有高度同源性。然而,最近已经测序了念珠藻属的全部基因组,因此,现在可得到紧靠该 PKS 基因簇上游区域的序列。在该区域中有三个 Orfs 与 OrfA 的该结构域(KS,MAT,ACP 和 KR)具有同源性(见图 3)。在这组中包含两个 ACP 结构域,都与 ORFAACP 结构域表现出高度同源性。在念珠藻属 PKS 基因簇的末端是编码 Het I PPTase 的基因。以前,不清楚 Het I 酶的底物是什么,然而在该基因簇中新鉴定的 Orf(Hg1 E)中存在串连的 ACP 结构域强烈地启发了本发明人该底物是那些 ACPs。*Schizochytrium* 与念珠藻属的 ACP 结构域的同源性,以及在两种蛋白质中该结构域的串连排列使得 Het I 很可能是 *Schizochytrium* ORFA ACPs 的异源性激活的候选物。本发明人相信首次认识到并推测了念珠藻属 Het I PPTase 的该用途。

[0205] 正如在 Metz 等,2001,出处同上中所述,PUFA PKS 系统的一个新特征是存在两个脱水酶结构域,两者都与大肠杆菌的 FabA 蛋白质具有同源性。利用上述新的念珠藻属 PKS 基因序列,现在可比较两个系统及其产物。念珠藻属基因簇(从 Hg1E 到 Het I)中结构域的顺序本发明人将它们确定为(参见图 3):

[0206] KS-MAT-2xACP, KR, KS, CLF-AT, ER(HetM, HetN)HetI

[0207] 在 *Schizochytrium* PUFA-PKS 的 Orfs A, B 和 C 中,该顺序(OrfA-B-C)是:

[0208] KS-MAT-9xACP-KR KS-CLF-AT-ER DH-DH-ER

[0209] 可见结构域顺序相对应(也具有较高的氨基酸序列同源性)。念珠藻属 PKS 系统的产物是不含双键(顺式或反式)的长链羟基脂肪酸(具有一个或两个羟基的 C26 或 C28)。*Schizochytrium* PKS 系统的产物是长链多不饱和脂肪酸(C22,具有 5 个或 6 个双键,均为顺式)。两个结构域组之间的明显差异是在 *Schizochytrium* 蛋白质中存在两个 DH 结构域,正好该结构域涉及 DHA 和 DPA 中顺式双键的形成(推测念珠藻属系统中的 HetM 和 HetN 涉及包含羟基且也含有 DH 结构域,其起源不同于在 PUFA 中发现的 DH)。另外,在 *Schizochytrium* Orfs B 和 C 中两份 ER 结构域的作用尚不清楚(在其它鉴定的 PUFA PKS 系统中不存在第二个 ER 结构域)。两组结构域之间的氨基酸序列同源性隐含着一种进化关系。可认为该 PUFA PKS 基因组起源于(在进化意义上)一种插入了 DH(FabA 样)结构域的祖先念珠藻属样 PKS 基因组。DH 结构域的加入导致在新的 PKS 终产物结构中导入了顺式双键。

[0210] *Schizochytrium* 和念珠藻属 PKS 结构域结构的比较以及 *Schizochytrium* 与希瓦氏菌属 PUFA-PKS 蛋白质之间的结构域组织的比较证实了改变结构域顺序以及插入新的结构域产生新终产物的天然能力。另外,现在可在实验室改造该基因以产生新产物。这些观察结果的含义是以定向或随机的方式继续改造该系统可影响终产物。例如,在优选的实施方案中,可设想用 PUFA-PKS 系统的一个 DH(FabA-样)结构域取代不具有异构化活性的 DH 结构域,可能产生具有顺式和反式双键混合体的分子。*Schizochytrium* PUFAPKS 系统的普遍产物是 DHA 和 DPA(C22:5 ω 6)。如果改造产生 C20 脂肪酸的系统,可预期该产物是 EPA 和 ARA(C20:4 ω 6)。这可为 ARA 提供一种新来源。也可使用产生不同的 DHA 与 DPA 比率的有关 PUFA-PKS 系统的结构域取代,例如,使用来自破囊壶菌属 23B 的基因(本文首次鉴定了其 PUFAPKS 系统)。

[0211] 另外,可设想特异性改变 *Schizochytrium* PUFA PKS 系统(至今描述的其它 PUFA PKS 系统没有两个 ER 结构域)中的一个 ER 结构域(例如,去掉,或失活)以测定其对终产

物分布特征的影响。可使用复杂性更高或更低的方案对 PUFA-PKS 蛋白质的各个不同的结构域以定向的方式尝试相似的策略。当然可不限于改造单个结构域。最后,通过混合来自 PUFA-PKS 系统和其它 PKS 或 FAS 系统(例如,I 型,II 型,模块型)的结构域可扩展该研究以产生全部范围的新终产物。例如,可将 PUFA-PKS DH 结构域导入正常情况下在其终产物中不掺入顺式双键的系统中。

[0212] 因此,本发明包含通过遗传修饰生物体中编码具有根据本发明的非细菌 PUFA PKS 系统的至少一个功能域的生物学活性的氨基酸序列的至少核酸序列,和/或表达含有编码该氨基酸序列的核酸序列的至少一个重组核酸分子来遗传修饰微生物或植物细胞的方法。上文已经详细描述了该序列,用于遗传修饰生物体的方法,和具体修饰的各种实施方案。一般来说,该方法用于产生可生产一种或多种特定的生物活性分子的特定遗传修饰的生物体。

[0213] 本发明的一个实施方案涉及被修饰成表达多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的重组宿主细胞,其中该 PKS 催化重复的和非重复的酶反应,且其中该 PUFA PKS 系统包括:(a)至少两个烯酰 ACP-还原酶(ER)结构域;(b)至少6个酰基载体蛋白(ACP)结构域;(c)至少两个 β -酮脂酰-ACP合酶(KS)结构域;(d)至少一个酰基转移酶(AT)结构域;(e)至少一个酮还原酶(KR)结构域;(f)至少两个 FabA-样 β -羟酰-ACP脱水酶(DH)结构域;(g)至少一个链长因子(CLF)结构域;和(h)至少一个丙二酰-CoA:ACP酰基转移酶(MAT)结构域。在一个实施方案中,PUFA PKS 系统是真核 PUFA PKS 系统。在一个优选的实施方案中,PUFA PKS 系统是藻类 PUFA PKS 系统。在一个更优选的实施方案中,该 PUFA PKS 系统是 Thraustochytriales PUFA PKS 系统。该 PUFA PKS 系统可包括,但不限于 Schizochytrium PUFA PKS 系统,和破囊壶菌属 PUFA PKS 系统。在一个实施方案中,该 PUFA PKS 系统可在原核宿主细胞中表达。在另一实施方案中,该 PUFA PKS 系统可在真核宿主细胞中表达。

[0214] 本发明的另一实施方案涉及被修饰成表达非细菌 PUFA PKS 系统的重组宿主细胞,其中该 PKS 系统催化重复的和非重复的酶反应,且其中该非细菌 PUFA PKS 系统包括至少下列生物学活性结构域:(a)至少一个烯酰 ACP-还原酶(ER)结构域;(b)多个酰基载体蛋白(ACP)结构域(至少4个);(c)至少两个 β -酮脂酰-ACP合酶(KS)结构域;(d)至少一个酰基转移酶(AT)结构域;(e)至少一个酮还原酶(KR)结构域;(f)至少两个 FabA-样 β -羟酰-ACP脱水酶(DH)结构域;(g)至少一个链长因子(CLF)结构域;和(h)至少一个丙二酰-CoA:ACP酰基转移酶(MAT)结构域。

[0215] 本发明的该实施方案的一个方面涉及产生含有至少一种 PUFA 的产物的方法,包括在有效产生该产物的条件下生长含有上述任一重组宿主细胞的植物,其中该重组宿主细胞是植物细胞。本发明的该实施方案的另一方面涉及生产含有至少一种 PUFA 的产物的方法,包括在有效产生该产物的条件下培养含有任一上述重组宿主细胞的培养物,其中该宿主细胞是微生物细胞。在一个优选的实施方案中,该宿主细胞中的 PKS 系统催化三酰甘油的定向生产。

[0216] 本发明的另一实施方案涉及含有非细菌多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的微生物,其中该 PKS 催化重复的和非重复的酶反应,且其中该 PUFA PKS 系统包括:(a)至少两个烯酰 ACP-还原酶(ER)结构域;(b)至少6个酰基载体蛋白(ACP)结构

域 ;(c) 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域 ;(d) 至少一个酰基转移酶 (AT) 结构域 ;(e) 至少一个酮还原酶 (KR) 结构域 ;(f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域 ;(g) 至少一个链长因子 (CLF) 结构域 ;和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。优选的是,该微生物是非细菌微生物且更优选的是真核微生物。

[0217] 本发明还有另一实施方案涉及含有非细菌多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的微生物,其中该 PKS 催化重复的和非重复的酶反应,且其中该 PUFA PKS 系统包括 : (a) 至少一个烯酰 ACP-还原酶 (ER) 结构域 ;(b) 多个酰基载体蛋白 (ACP) 结构域 (至少 4 个) ;(c) 至少两个 β -酮脂酰-ACP 合酶 (KS) 结构域 ;(d) 至少一个酰基转移酶 (AT) 结构域 ;(e) 至少一个酮还原酶 (KR) 结构域 ;(f) 至少两个 FabA-样 β -羟酰-ACP 脱水酶 (DH) 结构域 ;(g) 至少一个链长因子 (CLF) 结构域 ;和 (h) 至少一个丙二酰-CoA:ACP 酰基转移酶 (MAT) 结构域。

[0218] 在本发明的一个实施方案中,预期可结合诱变程序与选择性筛选程序以获得目的生物活性分子。它可包括寻找各种生物活性化合物的方法。该寻找不限于产生具有顺式双键的那些分子。该诱变方法可包括,但不限于 : 化学诱变,基因改组,编码特定酶结构域的基因区域的交换,或局限于这些基因的特定区域的诱变,以及其它方法。

[0219] 例如,可使用高通量诱变方法影响或优化目的生物活性分子的生产。一旦形成有效的模型系统,可以高通量方式修饰这些基因。可设想在两种水平上利用这些技术。首先,如果可设计足够的选择性筛选方法用于产生目的产物 (例如,ARA),则可用于尝试改变该系统以产生该产物 (例如,代替诸如上文讨论的其它策略,或者与其协同作用)。另外,如果上文列出的该策略导致一组基因不产生该目的产物,那么可使用该高通量技术优化该系统。例如,如果导入的结构域仅在相当低的温度下起作用,则可设计允许去掉该限制的选择方法。在本发明的一个实施方案中,筛选方法可用于鉴定具有与本文所述 *Schizochytrium* 的 PUFA PKS 系统相似的新 PKS 系统的其它非细菌生物体 (参见上文)。在该生物体中鉴定的同源性 PKS 系统可用于与本文所述用于 *Schizochytrium*,以及用于其它遗传物质来源的方法相似的方法中,其中从该来源产生,进一步修饰和 / 或突变 PKS 系统用于在该微生物中,在另一微生物中,或在高等植物中表达以产生各种化合物。

[0220] 应认识到可导入天然 (内源性,原始的) PKS 系统中的随机的或者定向的许多遗传改变会导致酶功能的失活。本发明优选的实施方案包括的系统仅选择不抑制 PKS 系统产生一种产物的能力的那些修饰。例如,大肠杆菌的 FabB- 菌株不能合成不饱和脂肪酸且为了生长需要向培养基中补充可代替其正常不饱和脂肪酸的脂肪酸 (参见 Metz 等, 2001, 出处同上)。然而,当用有功能的 PUFA-PKS 系统 (即,在大肠杆菌宿主中产生 PUFA 产物的系统,参见 (Metz 等, 2001, 出处同上, 图 2A) 转化该菌株时可取消该需要 (对培养基的补充需要)。转化的 FabB- 菌株现在需要有功能的 PUFA-PKS 系统 (以产生该不饱和脂肪酸) 用于生长而不需要补充该脂肪酸。该例子中的关键因素在于广泛的不饱和脂肪酸的生产可以满足 (即使不饱和脂肪酸代替例如支链脂肪酸)。因此,在本发明的另一优选的实施方案中,可在本文公开的一种或多种 PUFAPKS 基因中产生大量突变,然后转化适当修饰的 FabB- 菌株 (例如在含有 ER 结构域的表达构建体中产生突变并转化具有其它必需结构域在独立的质粒中或者整合进染色体中的 FabB 菌株) 并仅选择不需补充培养基就能生长 (即,仍然具有产生能补充 FabB- 缺陷的分子) 的那些转化子。可开发其它筛选工具用于寻找在有活性

的 PKS 系统的该选择性亚型中产生的特定的化合物（例如，对于脂肪酸使用 GC）。可设想许多相似的选择性筛选工具用于目的生物活性分子。

[0221] 如上所述，在本发明的一个实施方案中，遗传修饰的微生物或植物包括合成所需生物活性分子（产物）的能力增强或者新导入了合成特定产物（例如，合成特定抗生素）的能力的微生物或植物。根据本发明，“增强合成产物的能力”是指在与该产物合成相关的途径中使得与在相同条件下培养或生长的野生型微生物或植物相比该微生物或植物产生的产物量（包括任何以前不存在的产物的生产）增加的任何增强或上调。产生该遗传修饰的生物体的方法在上文中已有详细描述。

[0222] 本发明的一个实施方案是通过生长或培养本发明的遗传修饰的微生物或植物（在上文中详细描述）产生所需生物活性分子（也称为产物或化合物）的方法。该方法包括分别在发酵培养基中培养或在合适的环境，例如土壤中生长具有按本文前面所述且根据本发明遗传修饰的微生物或植物的步骤。在一个优选的实施方案中，产生本发明的生物活性分子的方法包括在有效产生该生物活性分子的条件培养表达包含多不饱和脂肪酸（PUFA）聚酮化合物合酶（PKS）系统的至少一个生物活性结构域的 PKS 系统的遗传修饰的生物体的步骤。在该优选的方面，PUFA PKS 系统的至少一个结构域由选自下组的核酸序列编码：(a) 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸（PUFA）聚酮化合物合酶（PKS）系统的至少一个结构域的核酸序列；(b) 编码来自以本发明的新筛选方法鉴定的微生物（在上文中详细描述）的 PUFA PKS 系统的至少一个结构域的核酸序列；(c) 编码选自：SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, 和其生物学活性片段组成的组中的氨基酸序列的核酸序列；(d) 编码选自：SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, 或其生物学活性片段的氨基酸序列的核酸序列；(e) 编码与选自：SEQ ID NO:2, SEQ ID NO:4, 或 SEQ ID NO:6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列；其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性；和，(f) 编码与选自：SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:13, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列；其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。在该方法的这一优选方面中，该生物体被遗传修饰成影响 PKS 系统的活性（在上文中详细描述）。用于与本发明的 PUFA PKS 系统相关的遗传修饰的优选宿主细胞在上文中描述。

[0223] 在生产本发明的所需生物活性化合物的方法中，在合适的培养基中在有效产生该生物活性化合物的条件下培养或生长遗传修饰的微生物。合适的或有效的培养基是指本发明的遗传修饰的微生物在其中培养时，能够产生所需产物的任何培养基。该培养基一般是含有可同化的碳，氮和磷酸盐来源的含水培养基。该培养基也可包含合适的盐，无机物，金属和其它营养成分。本发明的微生物可在常规发酵生物反应器中培养。可通过包括，但不限于，分批，补料分批，细胞再循环，和连续发酵的任何发酵方法培养该微生物。用于根据本发明的潜在宿主微生物的优选生长条件是本领域熟知的。该遗传修饰的微生物产生的所需生物活性分子可使用常规分离和纯化技术从发酵培养基中回收。例如，可过滤或离心该发酵培养基以去掉微生物，细胞碎片和其它特定物质，并通过例如，离子交换，色谱法，萃取，

溶剂萃取,膜分离,电透析,逆向渗透,蒸馏,化学衍生化和结晶的常规方法从无细胞上清中回收该产物。作为选择,可使用产生所需化合物的微生物,或其提取物和各种分级分离成份而不必从该产物中去除微生物成份。

[0224] 在生产本发明的所需生物活性化合物的方法中,在发酵培养基中培养或在诸如土壤的合适培养基中生长遗传修饰的植物。合适的,或有效的发酵培养基在上文中详细讨论。用于高等植物的合适生长培养基包括用于植物的任何生长培养基,包括,但不限于,土壤,沙土,支持根生长的任何其它颗粒培养基(例如,蛭石,perlite,等)或水耕法培养基,以及优化该高等植物生长的合适的光照,水和营养补充物。本发明的遗传修饰的植物可通过按照本发明遗传修饰的 PKS 系统的活性改造成产生大量所需产物。通过从该植物中提取该化合物的纯化方法可回收该化合物。在一个优选的实施方案中,通过收获该植物回收该化合物。在该实施方案中,可以其天然状态消费该植物或者进一步加工成可消费的产物。

[0225] 如上所述,在一个方面,用于本发明的遗传修饰的微生物可内源性含有并表达 PUFA PKS 系统,且该遗传修饰可以是对内源性 PUFA PKS 系统的一个或多个功能域的遗传修饰,因此该修饰对 PUFA PKS 系统的活性具有一些影响。在另一方面,该生物体可内源性含有并表达 PUFA PKS 系统,且该遗传修饰可以是导入至少一个外源性核酸序列(例如,重组核酸分子),其中该外源性核酸分子编码第二个 PKS 系统的至少一个生物活性结构域或蛋白质和/或影响所述 PUFA PKS 系统的活性的蛋白质(例如,磷酸泛酰巯基乙胺基转移酶(PPTase),下文讨论)。在还有另一方面,该生物体不必内源性(天然)含有 PUFA PKS 系统,但是被遗传修饰成导入至少一个编码具有 PUFA PKS 系统的至少一个结构域的生物学活性的氨基酸序列的重组核酸分子。在该方面,通过在该生物体中导入或增加 PUFA PKS 活性来影响 PUFA PKS 的活性。分别与这些方面相关的各种实施方案已在上文中详细讨论。

[0226] 在产生生物活性化合物的方法的一个实施方案中,与野生型生物体相比,该遗传修饰改变了由内源性 PKS 系统产生的至少一种产物。

[0227] 在另一实施方案中,该生物体内源性表达含有该 PUFA PKS 系统的至少一种生物学活性结构域的 PKS 系统,且该遗传修饰包含用选自:编码来自第二个 PKS 系统的至少一个生物学活性结构域的重组核酸分子和编码影响 PUFA PKS 系统的活性的蛋白质的重组核酸分子组成的组中的重组核酸分子转染该生物体。在该实施方案中,与野生型生物体相比,该遗传修饰优选改变由内源性 PKS 系统产生的至少一种产物。第二个 PKS 系统可包括另一 PUFA PKS 系统(细菌或非细菌),I 型 PKS 系统,II 型 PKS 系统,和/或模块型 PKS 系统。影响 PKS 系统的活性的蛋白质的例子已在上文中描述(例如,PPTase)。

[0228] 在另一实施方案中,该生物体通过用编码多不饱和脂肪酸(PUFA)聚酮化合物合酶(PKS)系统的至少一个结构域的重组核酸分子转染来进行遗传修饰。该重组核酸分子在本文前面进行了详细描述。

[0229] 在另一实施方案中,该生物体内源性表达非细菌 PUFA PKS 系统,且该遗传修饰包括用来自不同 PKS 系统的结构域取代编码该非细菌 PUFA PKS 系统的至少一个结构域的核酸序列。在另一实施方案中,该生物体内源性表达非细菌 PUFA PKS 系统,且通过用编码调节 PUFA PKS 系统产生的脂肪酸的链长的蛋白质的重组核酸分子转染该生物体进行了修饰。在一个方面,编码调节脂肪酸链长的蛋白质的重组核酸分子取代编码该非细菌 PUFA PKS 系统中的链长因子的核酸序列。在另一方面,调节 PUFA PKS 系统产生的脂肪酸的链长的蛋白

质是链长因子。在另一方面,调节 PUFA PKS 系统产生的脂肪酸的链长的蛋白质是指导 C20 单位合成的链长因子。

[0230] 在另一实施方案中,该生物体表达在选自自由编码 β -羟酰-ACP 脱水酶 (DH) 的结构域和编码 β -酮脂酰-ACP 合酶 (KS) 的结构域组成的组中的一个结构域中包含遗传修饰的非细菌 PUFA PKS 系统,其中与没有该修饰相比,该修饰改变了由 PUFA PKS 系统产生的长链脂肪酸的比率。在该实施方案的一个方面,该修饰选自自由该结构域的全部或部分缺失,用来自不同生物体的同源性结构域取代该结构域,和该结构域的突变组成的组中。

[0231] 在另一实施方案中,该生物体表达在烯酰-ACP 还原酶 (ER) 结构域中含有修饰的非细菌 PUFA PKS 系统,其中与没有该修饰相比,该修饰导致产生不同化合物。在该实施方案的一个方面,该修饰选自自由该 ER 结构域的全部或部分缺失,用来自不同生物体的 ER 结构域取代该 ER 结构域,和该 ER 结构域的突变组成的组中。

[0232] 在产生生物活性分子的方法的一个实施方案中,该生物体产生的多不饱和脂肪酸 (PUFA) 分布特征不同于没有遗传修饰的天然生物体。

[0233] 用于产生生物活性分子的许多其它遗传修饰对于本领域的技术人员而言在本说明书的教导下是显而易见的,且各种其它修饰已在本文前面进行了讨论。本发明包含导致产生所需生物活性分子的与本文所述 PUFAPKS 系统相关的任何遗传修饰。

[0234] 根据本发明的生物活性分子包括具有生物学活性且可由包含具有本文所述非细菌 PUFA PKS 系统的至少一个功能域的生物学活性的至少一种氨基酸序列的 PKS 系统产生的任何分子(化合物,产物,等)。该生物活性分子可包括,但不限于:多不饱和脂肪酸 (PUFA),抗炎症配制品,化疗剂,有效赋形剂,骨质疏松症药物,抗抑郁剂,抗惊厥剂,抗幽门螺杆菌 (*Helicobacter pylori*) 药物,治疗神经变性疾病的药物,治疗变性肝脏疾病的药物,抗生素,和降胆固醇配制品。本发明的非细菌 PUFA PKS 系统的一个优势是该系统具有诱导顺式构型的碳-碳双键的能力,且该分子在每第三个碳包含一个双键。该能力可用于产生各种化合物。

[0235] 优选的是,目的生物活性化合物由该遗传修饰的微生物以高于约 0.05%,且优选高于约 0.1%,且更优选高于约 0.25%,且更优选高于约 0.5%,且更优选高于约 0.75%,且更优选高于约 1%,且更优选高于约 2.5%,且更优选高于约 5%,且更优选高于约 10%,且更优选高于约 15%,且更优选高于约 20%的微生物干重的量产生。对于脂类化合物,优选的是,该化合物以高于约 5%的微生物干重的量产生。对于其它生物活性化合物,例如抗生素或以更少量合成的化合物,将在微生物干重中具有该化合物的那些菌株鉴定为预期含有上述类型的新 PKS 系统。在一些实施方案中,特定的生物活性分子(化合物)由该微生物分泌,而不是累积。因此,该生物活性分子一般从培养基中回收且产生的分子浓度随该微生物和培养规模而变化。

[0236] 本发明的一个实施方案涉及修饰含有至少一种脂肪酸的终产物的方法,包括向所述终产物中添加一种由重组宿主细胞产生的油类,该重组宿主细胞表达至少一个含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的重组核酸分子。该 PUFA PKS 系统是任一非细菌 PUFA PKS 系统,且优选选自:(a) 编码来自 *Thraustochytrid* 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列;(b) 编码来自以本文公开的新筛选方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列;

(c) 编码选自由 :SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6, 和其生物学活性片段组成的组中的氨基酸序列的核酸序列 ;(d) 编码选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, SEQ ID NO :32, 或其生物学活性片段的氨基酸序列的核酸序列 ;(e) 编码与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中该氨基酸序列具有 PUFAPKS 系统的至少一个结构域的生物学活性 ;和 (f) 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, 或 SEQ ID NO :32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。上文中已经详细描述了这些核酸序列的变化。

[0237] 优选的是, 该终产物选自由食品, 饮食添加剂, 药物配制品, 人源化动物乳汁, 和婴儿配制品组成的组中。合适的药物配制品包括, 但不限于 : 抗炎症配制品, 化疗剂, 有效赋形剂, 骨质疏松症药物, 抗抑郁剂, 抗惊厥剂, 抗幽门螺杆菌药物, 治疗神经变性疾病的药物, 治疗变性肝脏疾病的药物, 抗生素, 和降胆固醇配制品。在一个实施方案中, 该终产物用于治疗选自由 : 慢性炎症, 急性炎症, 胃肠疾病, 癌症, 恶病质, 心脏再狭窄, 神经变性疾病, 肝脏变性疾病, 血脂疾病, 骨质疏松症, 骨关节炎, 自身免疫疾病, 先兆子痫, 早产, 年老相关性黄斑病, 肺病, 和过氧化物酶体疾病组成的组中的一种症状。

[0238] 合适的食品包括, 但不限于, 精细面包商品, 面包卷 (bread and rolls), 早餐谷类食品, 加工的和未加工的乳酪, 调味品 (番茄酱, 蛋黄酱, 等), 乳制品 (奶, 酸乳酪), 布丁和胶质甜点, 碳酸饮料, 茶, 粉末状饮料混合物, 加工的鱼产物, 水果类饮料, 口香糖, 硬甜食, 冷冻乳制品, 加工的肉类产品, 坚果和坚果饼 (nut-based spreads), 意大利面食, 加工的家禽产物, 肉汁和酱, 马铃薯片和其它片状食品或酥脆食品, 巧克力和其它甜食, 汤和汤混合物, 大豆类产物 (奶, 饮料, 乳酪, whiteners), 以植物油为基础的饼 (spreads), 和蔬菜类饮料。

[0239] 本发明的另一实施方案涉及生产人源化的动物乳汁的方法。该方法包括用至少一种含有编码 PUFA PKS 系统的至少一个生物活性结构域的核酸序列的重组核酸分子遗传修饰产奶动物的产奶细胞的步骤。该 PUFA PKS 系统是一种非细菌 PUFA PKS 系统, 且优选的是, 该 PUFA PKS 系统的至少一个结构域由选自下组的核酸序列编码 : (a) 编码来自 Thraustochytrid 微生物的多不饱和脂肪酸 (PUFA) 聚酮化合物合酶 (PKS) 系统的至少一个结构域的核酸序列 ;(b) 编码来自以本文前面描述的新筛选方法鉴定的微生物的 PUFA PKS 系统的至少一个结构域的核酸序列 ;(c) 编码选自由 :SEQ ID NO :2, SEQ ID NO :4, SEQ ID NO :6, 和其生物学活性片段组成的组中的氨基酸序列的核酸序列 ;(d) 编码选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22, SEQ ID NO :24, SEQ ID NO :26, SEQ ID NO :28, SEQ ID NO :30, SEQ ID NO :32, 或其生物学活性片段的氨基酸序列的核酸序列 ;(e) 编码与选自 :SEQ ID NO :2, SEQ ID NO :4, 或 SEQ ID NO :6 的氨基酸序列的至少 500 个连续氨基酸有至少约 60% 相同性的氨基酸序列的核酸序列 ;其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性 ;和 / 或 (f) 编码与选自 :SEQ ID NO :8, SEQ ID NO :10, SEQ ID NO :13, SEQ ID NO :18, SEQ ID NO :20, SEQ ID NO :22,

SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, 或 SEQ ID NO:32 的氨基酸序列具有至少约 60% 相同性的氨基酸序列的核酸序列; 其中该氨基酸序列具有 PUFA PKS 系统的至少一个结构域的生物学活性。

[0240] 遗传修饰宿主细胞并产生遗传修饰的非人产奶动物的方法是本领域已知的。可修饰的宿主动物的例子包括牛, 绵羊, 猪, 山羊, 牦牛, 等, 它们易于进行遗传操作和克隆用于迅速扩大表达转基因的群体。对于动物, 通过修饰基因调控区可使 PKS- 样转基因适应于在靶器官, 组织和体液中表达。特别所需是在宿主动物乳腺乳汁中生产 PUFA。

[0241] 提供下面的实施例用于说明的目的而不打算用于限制本发明的范围。

实施例

[0242] 实施例 1

[0243] 下面的实施例描述了对来自 *Schizochytrium* 的 PKS 相关性序列的进一步分析。

[0244] 本发明人使用 PCT 公开号 WO 0042195 的实施例 8 和 9 和美国申请系列号 09/231,899 中列出的一般方法测序了包含 *Schizochytrium* PUFA PKS 系统中所有三个可读框 (Orfs) 全长的基因组 DNA。Schizochytrium PKS 蛋白质中的生物学活性结构域在图 1 中以图示进行了描述。Schizochytrium PUFA PKS 系统的结构域结构更具体地描述如下。

[0245] 可读框 A (OrfA) :

[0246] OrfA 的完整核苷酸序列本文表示为 SEQ ID NO:1。OrfA 是一个 8730 个核苷酸的序列 (不包括终止密码子), 它编码一个 2910 个氨基酸的序列, 本文表示为 SEQ ID NO:2。OrfA 内有 12 个结构域:

[0247] (a) 一个 β -酮脂酰 -ACP 合酶 (KS) 结构域;

[0248] (b) 一个丙二酰 -CoA:ACP 酰基转移酶 (MAT) 结构域;

[0249] (c) 9 个酰基载体蛋白 (ACP) 结构域;

[0250] (d) 一个酮还原酶 (KR) 结构域。

[0251] OrfA 内包含的结构域根据如下测定:

[0252] (1) Pfam 程序分析的结果 (Pfam 是蛋白质结构域或保守蛋白区域的多个序列对比的数据库)。该序列对比表示了可暗示该蛋白质功能的一些进化保守结构。从 Pfam 序列对比建立的概形隐藏马尔可夫模型 (概形 HMMs) 对于自动识别属于已有蛋白质家族的新蛋白质非常有用, 即使该同源性较弱。不同于标准的逐对序列对比方法 (例如 BLAST, FASTA), Pfam HMMs 可明智地处理多结构域蛋白质。涉及所用的 Pfam 版本的参考文献是: Bateman A, Birney E, Cerruti L, Durbin R, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) *Nucleic Acids Research* 30(1):276-280; 和 / 或

[0253] (2) 使用 BLAST 2.0 Basic BLAST 同源性检索与细菌 PUFA-PKS 系统 (例如, 希瓦氏菌属) 的同源性比较, 使用 blastp 用标准默认参数进行氨基酸检索, 其中, 通过默认筛选查询序列的低复杂性区域 (在 Altschul, S. F., Madden, T. L., Schaaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997), “间隔的 BLAST 和 PSI-BLAST: 新产生的蛋白质数据库检索程序”。*Nucleic Acids Res.* 25:3389-3402 中描述, 本文引用以其整体作为参考)。

[0254] 据信提供的单个结构域的序列含有编码功能域的全长序列, 且在 Orf 内可含有其

它侧翼序列。

[0255] ORFA-KS

[0256] OrfA 中的第一个结构域是 KS 结构域,本文也称为 ORFA-KS。该结构域包含在覆盖从 SEQ ID NO:1(OrfA) 的约位置 1 和 40 之间的起始位点到 SEQ ID NO:1 的约位置 1428 和 1500 之间的终止位点的核苷酸序列内。含有编码 ORFA-KS 结构域的序列的核苷酸序列在本文中表示为 SEQ ID NO:7(SEQ ID NO:1 的位置 1-1500)。含有 KS 结构域的氨基酸序列覆盖从 SEQ ID NO:2(ORFA) 的约位置 1 和 14 之间的起始位点到 SEQ ID NO:2 的约位置 476 和 500 之间的终止位点。含有 ORFA-KS 结构域的氨基酸序列本文表示为 SEQ ID NO:8(SEQ ID NO:2 的位置 1-500)。应注意 ORFA-KS 结构域含有一个活性位点基序:DXAC*(* 酰基结合位点 C₂₁₅)。

[0257] ORFA-MAT

[0258] OrfA 的第二个结构域是 MAT 结构域,本文也称为 ORFA MAT。该结构域包含在覆盖从 SEQ ID NO:1(OrfA) 的约位置 1723 和 1798 之间的起始位点到 SEQ ID NO:1 的约位置 2805 和 3000 之间的终止位点的核苷酸序列内。含有编码 ORFA-MAT 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:9(SEQ ID NO:1 的位置 1723-3000)。含有 MAT 结构域的氨基酸序列覆盖从 SEQ ID NO:2(ORFA) 的约位置 575 和 600 之间的起始位点到 SEQ ID NO:2 的约位置 935 和 1000 之间的终止位点。含有 ORFA-MAT 结构域的氨基酸序列本文表示为 SEQ ID NO:10(SEQ ID NO:2 的位置 575-1000)。应注意该 ORFA-MAT 结构域含有一个活性位点基序:GHS*⁷⁰⁶XG(* 酰基结合位点 S₇₀₆),本文表示为 SEQ ID NO:11。

[0259] ORFA-ACP#1-9

[0260] OrfA 的结构域 3-11 是 9 个串连的 ACP 结构域,本文也称为 ORFA-ACP(该序列的第一个结构域是 ORFA-ACP1,第二个结构域是 ORFA-ACP2,第三个结构域是 ORFA-ACP3,等)。第一个 ACP 结构域,即 ORFA-ACP1 包含在覆盖从 SEQ ID NO:1(OrfA) 的约位置 3343 到约位置 3600 的核苷酸序列内。含有编码 ORFA-ACP1 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:12(SEQ ID NO:1 的位置 3343-3600)。含有第一个 ACP 结构域的氨基酸序列覆盖从 SEQ ID NO:2 的约位置 1115 到约位置 1200。含有 ORFA-ACP 1 结构域的氨基酸序列在本文中表示为 SEQ ID NO:13(SEQ ID NO:2 的位置 1115-1200)。应注意该 ORFA-ACP1 结构域含有一个活性位点基序:LGIDS*(* 泛酰巯基乙胺结合基序 S₁₁₅₇),本文以 SEQ ID NO:14 表示。所有 9 个 ACP 结构域的核苷酸和氨基酸序列高度保守且,因此各结构域的序列在本文中没有用单独的序列标识符表示。然而,根据该信息,本领域的技术人员可容易地测定其它 8 个 ACP 结构域各自的序列。9 个结构域的重复间隔是 SEQ ID NO:1 的接近约 110 到约 330 核苷酸。

[0261] 所有 9 个 ACP 结构域一起覆盖从 SEQ ID NO:1 的约位置 3283 到约位置 6288 的 OrfA 区域,它相应于 SEQ ID NO:2 从约 1095 到约 2096 的氨基酸位置。该区域包括各个 ACP 结构域之间的接头片段。9 个 ACP 结构域中每个含有一个泛酰巯基乙胺结合基序 LGIDS*(本文中以 SEQ ID NO:14 表示),其中 * 是泛酰巯基乙胺结合位点 S。ACP 结构域区域的每个末端和各 ACP 结构域之间是高度富含脯氨酸(P)和丙氨酸(A)的区域,据信它是接头区域。例如,ACP 结构域 1 和 2 之间是序列:APAPVKAAAPAAPVASAPAPA,本文表示为 SEQ ID NO:15。

[0262] ORFA-KR

[0263] OrfA 中的结构域 12 是 KR 结构域,本文也称为 ORFA-KR。该结构域包含在覆盖从 SEQ ID NO:1 的约位置 6598 的起始位点到 SEQ ID NO:1 的约位置 8730 的终止位点的核苷酸序列内。含有编码 ORFA-KR 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:17 (SEQ ID NO:1 的位置 6598-8730)。含有 KR 结构域的氨基酸序列覆盖从 SEQ ID NO:2 (ORFA) 的约位置 2200 的起始位点到 SEQ ID NO:2 的约位置 2910 的终止位点。含有 ORFA-KR 结构域的氨基酸序列本文表示为 SEQ ID NO:18 (SEQ ID NO:2 的位置 2200-2910)。KR 结构域内具有与短链乙醛脱氢酶 (KR 是该家族的一个成员) 同源的核心区。该核心区跨越从 SEQ ID NO:1 的约位置 7198 至约位置 7500,它相应于 SEQ ID NO:2 的氨基酸位置 2400-2500。

[0264] 可读框 B (OrfB) :

[0265] OrfB 的完整核苷酸序列在本文中表示为 SEQ ID NO:3。OrfB 是一个 6177 个核苷酸的序列 (不包括终止密码子),它编码一个 2059 个氨基酸的序列,本文表示为 SEQ ID NO:4。OrfB 内有 4 个结构域:

[0266] (a) β -酮脂酰-ACP 合酶 (KS) 结构域;

[0267] (b) 一个链长因子 (CLF) 结构域;

[0268] (c) 一个酰基转移酶 (AT) 结构域;

[0269] (d) 一个烯酰 ACP-还原酶 (ER) 结构域。

[0270] ORFB 内包含的结构域根据如下结果确定:(1) 上述用 Pfam 程序分析的结果;和/或(2) 也在上文描述的使用 BLAST 2.0 Basic BLAST 同源性检索与细菌 PUFA-PKS 系统 (例如,希瓦氏菌属) 的同源性比较。据信提供的单个结构域的序列含有编码功能域的全长序列,且在 Orf 内可含有其它侧翼序列。

[0271] ORFB-KS

[0272] OrfB 中的第一个结构域是 KS 结构域,本文也称为 ORFB-KS。该结构域包含在覆盖从 SEQ ID NO:3 (OrfB) 的约位置 1 和 43 之间的起始位点到 SEQ ID NO:3 的约位置 1332 和 1350 之间的终止位点的核苷酸序列内。含有编码 ORFB-KS 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:19 (SEQ ID NO:3 的位置 1-1350)。含有 KS 结构域的氨基酸序列覆盖从 SEQ ID NO:4 (ORFB) 的约位置 1 和 15 之间的起始位点到 SEQ ID NO:4 的约位置 444 和 450 之间的终止位点。含有 ORFB-KS 结构域的氨基酸序列本文表示为 SEQ ID NO:20 (SEQ ID NO:4 的位置 1-450)。应注意 ORFB-KS 结构域含有一个活性位点基序:DXAC* (*酰基结合位点 C₁₉₆)。

[0273] ORFB-CLF

[0274] OrfB 中的第二个结构域是 CLF 结构域,本文也称为 ORFB-CLF。该结构域包含在覆盖从 SEQ ID NO:3 (OrfB) 的约位置 1378 和 1402 之间的起始位点到 SEQ ID NO:3 的约位置 2682 和 2700 之间的终止位点的核苷酸序列内。含有编码 ORFB-CLF 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:21 (SEQ ID NO:3 的位置 1378-2700)。含有 CLF 结构域的氨基酸序列覆盖从 SEQ ID NO:4 (ORFB) 的约位置 460 和 468 之间的起始位点到 SEQ ID NO:4 的约位置 894 和 900 之间的终止位点。含有 ORFB-CLF 结构域的氨基酸序列本文表示为 SEQ ID NO:22 (SEQ ID NO:4 的位置 460-900)。应注意 ORFB-CLF 结构域含有 KS 活性位点基序但没有结合酰基的半胱氨酸。

[0275] ORFB-AT

[0276] 在 OrfB 中的第三个结构域是 AT 结构域,本文也称为 ORFB-AT。该结构域包含在覆盖从 SEQ ID NO:3(OrfB) 的约位置 2701 和 3598 之间的起始位点到 SEQ ID NO:3 的约位置 3975 和 4200 之间的终止位点的核苷酸序列内。含有编码 ORFB-AT 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:23(SEQ ID NO:3 的位置 2701-4200)。含有 AT 结构域的氨基酸序列覆盖从 SEQ ID NO:4(ORFB) 的约位置 901 和 1200 之间的起始位点到 SEQ ID NO:4 的约位置 1325 和 1400 之间的终止位点。含有 ORFB-AT 结构域的氨基酸序列本文表示为 SEQ ID NO:24(SEQ ID NO:4 的位置 901-1400)。应注意该 ORFB-AT 结构域含有 AT 活性位点基序 GxS^{*}xG(* 酰基结合位 S₁₁₄₀)。

[0277] ORFB-ER

[0278] OrfB 中的第四个结构域是 ER 结构域,本文也称为 ORFB-ER。该结构域包含在覆盖从 SEQ ID NO:3(OrfB) 的约位置 4648 的起始位点到 SEQ ID NO:3 的约位置 6177 的终止位点的核苷酸序列内。含有编码 ORFB-ER 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:25(SEQ ID NO:3 的位置 4648-6177)。含有 ER 结构域的氨基酸序列覆盖从 SEQ ID NO:4(ORFB) 的约位置 1550 的起始位点到 SEQ ID NO:4 的约位置 2059 的终止位点。含有 ORFB-ER 结构域的氨基酸序列本文表示为 SEQ ID NO:26(SEQ ID NO:4 的位置 1550-2059)。

[0279] 可读框 C(OrfC) :

[0280] OrfC 的完整核苷酸序列在本文中表示为 SEQ ID NO:5。OrfC 是一个 4509 个核苷酸的序列(不包括终止密码子),它编码 1503 个氨基酸的序列,本文表示为 SEQ ID NO:6。OrfC 内含有 3 个结构域:

[0281] (a) 两个 FabA-样 β -羟酰-ACP 脱水酶(DH) 结构域;

[0282] (b) 一个烯酰 ACP 还原酶(ER) 结构域。

[0283] ORFC 内包含的结构域根据如下结果确定:(1) 上述用 Pfam 程序分析的结果;和/或(2) 也在上文描述的使用 BLAST 2.0 Basic BLAST 同源性检索与细菌 PUFA-PKS 系统(例如,希瓦氏菌属)的同源性比较。据信提供的单个结构域的序列含有编码功能域的全长序列,且在 Orf 内可含有其它侧翼序列。

[0284] ORFC-DH1

[0285] OrfC 中的第一个结构域是 DH 结构域,本文也称为 ORFC-DH1。它是 OrfC 中的两个 DH 结构域之一,因此命名为 DH1。该结构域包含在覆盖从 SEQ ID NO:5(OrfC) 的约位置 1 和 778 之间的起始位点到 SEQ ID NO:5 的约位置 1233 和 1350 之间的终止位点的核苷酸序列内。含有编码 ORFC-DH1 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:27(SEQ ID NO:5 的位置 1-1350)。含有 DH1 结构域的氨基酸序列覆盖从 SEQ ID NO:6(ORFC) 的约位置 1 和 260 之间的起始位点到 SEQ ID NO:6 的约位置 411 和 450 之间的终止位点。含有 ORFC-DH1 结构域的氨基酸序列本文表示为 SEQ ID NO:28(SEQ ID NO:6 的位置 1-450)。

[0286] ORFC-DH2

[0287] OrfC 中的第二个结构域是 DH 结构域,本文也称为 ORFC-DH2。它是 OrfC 中两个 DH 结构域的第二,因此命名为 DH2。该结构域包含在覆盖从 SEQ ID NO:5(OrfC) 的约位置 1351 和 2437 之间的起始位点到 SEQ ID NO:5 的约位置 2607 和 2850 之间的终止位点的核苷酸序列内。含有编码 ORFC-DH2 结构域的序列的核苷酸序列本文表示为 SEQ ID NO:29(SEQ ID NO:5 的位置 1351-2850)。含有 DH2 结构域的氨基酸序列覆盖从 SEQ ID NO:

6 (ORFC) 的约位置 451 和 813 之间的起始位点到 SEQ ID NO :6 的约位置 869 和 950 之间的终止位点。含有 ORFC-DH2 结构域的氨基酸序列本文表示为 SEQ ID NO :30 (SEQ ID NO :6 的位置 451-950)。

[0288] ORFC-ER

[0289] OrfC 中的第三个结构域是 ER 结构域, 本文也称为 ORFC-ER。该结构域包含在覆盖从 SEQ ID NO :5 (OrfC) 的约位置 2998 的起始位点到 SEQ ID NO :5 的约位置 4509 的终止位点的核苷酸序列内。含有编码 ORFC-ER 结构域的序列的核苷酸序列本文表示为 SEQ ID NO :31 (SEQ ID NO :5 的位置 2998-4509)。含有 ER 结构域的氨基酸序列覆盖从 SEQ ID NO :6 (ORFC) 的约位置 1000 的起始位点到 SEQ ID NO :6 的约位置 1502 的终止位点。含有 ORFC-ER 结构域的氨基酸序列本文表示为 SEQ ID NO :32 (SEQ ID NO :6 的位置 1000-1502)。

[0290] 实施例 2

[0291] 下面的实施例描述了使用本发明的筛选方法鉴定含有根据本发明的 PUFAPKS 系统的三个其它的非细菌生物。

[0292] 按照美国临时申请系列号 60/298, 796 中所述和本文详细描述筛选方法培养破囊壶菌 23B (ATCC 20892)。

[0293] 用于检测含有产生 PUFA 的 PKS 系统的微生物而形成的生物学合理的筛选系统 (使用摇瓶培养物) 按如下进行:

[0294] 将待测菌株 / 微生物的 2mL 培养物放在含有 50mL 培养基的 250mL 带挡板的摇瓶中 (有氧处理), 并将相同菌株的另一 2mL 培养物放在含有 200mL 培养基的 250mL 无挡板的摇瓶中 (缺氧处理)。将两种摇瓶放在 200rpm 的摇床上。培养 48-72 小时后, 离心收获培养物并通过气相色谱法分析细胞的脂肪酸甲基酯以测定各培养物的下列数据: (1) 脂肪酸分布特征; (2) PUFA 含量; (3) 脂肪含量 (评估为脂肪酸总量 (TFA))。

[0295] 然后分析这些数据并回答下列 5 个问题:

[0296] 选择标准: 低氧 / 缺氧瓶与有氧瓶 (是 / 否):

[0297] (1) 与有氧培养物相比低氧培养物中的 DHA (或其它 PUFA 含量) (以 % FAME) 是否保持约相同或优选提高?

[0298] (2) 缺氧培养物中的 C14:0+C16:0+C16:1 是否超过约 40% 的 TFA?

[0299] (3) 在缺氧培养物中对于常规氧依赖型延长酶 / 去饱和酶途径是否有极少 (FAME > 1%) 或者没有前体 (C18:3n-3+C18:2n-6+C18:3n-6)?

[0300] (4) 与有氧培养相比低氧培养中的脂肪含量 (以脂肪酸总量 / 细胞干重表示) 是否提高?

[0301] (5) 与有氧培养相比低氧培养中以占细胞干重百分数表示的 DHA (或其它 PUFA 含量) 是否增加?

[0302] 如果前 3 个问题的答案为是, 则它是该菌株含有形成长链 PUFA 的 PKS 遗传系统的良好征兆。答案为是的问题越多 (优选前三个问题的答案必须为是), 该菌株含有这种 PKS 遗传系统的征兆就越强。如果 5 个问题的答案都为是, 那么就是该菌株含有形成长链 PUFA 的 PKS 遗传系统的极强征兆。

[0303] 在上述方法之后, 使用破囊壶菌属种类 23B (ATCC 20892) 的冷冻小瓶接种含有 50mL RCA 培养基的 250mL 摇瓶。25°C 下在摇床上摇动 (200rpm) 培养物 72 小时。RCA 培养

基含有如下成份：

[0304] RCA 培养基

[0305] 去离子水 1000mL

[0306] Reef Crystals[®]海盐 40g/L

[0307] 葡萄糖 20g/L

[0308] 谷氨酸一钠 (MSG) 20g/L

[0309] 酵母提取物 1g/L

[0310] PII 金属 * 5mL/L

[0311] 维生素混合物 * 1mL/L

[0312] pH 7.0

[0313] *PII 金属混合物和维生素混合物与美国专利号 5, 130, 742 所述相同, 本文引用以其整体作为参考。

[0314] 然后使用 25mL 72 小时的旧培养物接种含有 50mL 低氮 RCA 培养基 (10g/L MSG 代替 20g/L) 的另一 250mL 摇瓶, 且使用另一 25mL 的培养物接种含有 175mL 低氮 RCA 培养基的 250mL 摇瓶。然后将两个摇瓶放在 25℃ 下的摇床 (200rpm) 上 72 小时。然后通过离心收获细胞并冷冻干燥。使用标准气相色谱步骤 (例如在 US 5, 130, 742 中所述) 分析干燥细胞的脂肪含量和脂肪酸分布特征和含量。

[0315] 破囊壶菌属 23B 的筛选结果如下：

[0316] 以 % FAME 表示的 DHA 是否增加？ 是 (38 → 44%)

[0317] C14:0+C16:0+C16:1 是否高于约 40% TFA？ 是 (44%)

[0318] 是否无 C18:3(n-3) 或 C18:3(n-6)？ 是 (0%)

[0319] 脂肪含量是否增加？ 是 (增加 2 倍)

[0320] DHA (或其它 HUFA 含量) 是否增加？ 是 (增加 2.3- 倍)

[0321] 该结果, 特别是在低氧条件下 DHA 含量 (表示为 % FAME) 明显增加有力地表明在该破囊壶菌属菌株中存在产生 PUFA 的 PKS 系统。

[0322] 为了提供证实存在 PUFA PKS 系统的其它数据, 使用来自 Schizochytrium 菌株 20888 的 PKS 探针对破囊壶菌 23B 进行 southern 印迹, 其中菌株 20888 已经确定含有产生 PUFA 的 PKS 系统 (即, 上述 SEQ ID Nos :1-32)。使用 Southern 印迹技术检测与来自 PKS PUFA 合成基因的杂交探针同源的破囊壶菌 23B 基因组 DNA 的片段。用 ClaI 或 KpnI 限制性核酸内切酶消化破囊壶菌 23B 基因组 DNA, 通过琼脂糖凝胶电泳分离 (0.7% 琼脂糖, 在标准 Tris- 乙酸盐 -EDTA 缓冲液中), 并通过毛细管转移印迹到 Schleicher & SchuellNytran Supercharge 膜上。使用两个地高辛配基标记的杂交探针, 一个对 Schizochytrium PKS OrfB 的烯酰还原酶 (ER) 区域 (Orf B 的核苷酸 5012-5511 ;SEQ ID NO :3) 具有特异性, 另一个对在 Schizochytrium PKS OrfC 的开始处的保守区 (OrfC 的核苷酸 76-549 ;SEQ ID NO :5) 具有特异性。

[0323] OrfB-ER 探针在破囊壶菌 23B 基因组 DNA 中检测到一个约 13kb 的 ClaI 片段和一个约 3.6kb 的 KpnI 片段。OrfC 探针在破囊壶菌 23B 基因组 DNA 中检测到一个约 7.5kb 的 ClaI 片段和一个约 4.6kb 的 KpnI 片段。

[0324] 最后, 使用地高辛配基标记的相应于 Schizochytrium 20888PUFA-PKS 基因下列

片段:OrfA的核苷酸7385-7879(SEQ ID NO:1),OrfB的核苷酸5012-5511(SEQ ID NO:3),和OrfC的核苷酸76-549(SEQ ID NO:5)的探针筛选由来自破囊壶菌23B基因组DNA的DNA片段插入载体 λ FIX II(Stratagene)组成的重组基因组文库。这些探针分别从破囊壶菌23B文库检测到阳性噬斑,表明Schizochytrium PUFA-PKS基因与破囊壶菌23B基因之间具有广泛同源性。

[0325] 总之,这些结果证实破囊壶菌23B基因组DNA含有与来自Schizochytrium 20888的PKS基因同源的序列。

[0326] 本文包括的该Thraustochytrid微生物作为用于上述实施方案的这些基因的另一来源。

[0327] 破囊壶菌23B(ATCC 20892)在其脂肪酸分布特征方面明显不同于Schizochytrium sp.(ATCC 20888)。破囊壶菌23B具有高达14:1的DHA:DPA(n-6)比率,相比之下Schizochytrium(ATCC 20888)中仅为2-3:1。破囊壶菌23B也具有高水平的C20:5(n-3)。与已知的Schizochytrium PUFA PKS系统相比对破囊壶菌23B的PUFA PKS系统中的结构域的分析给我们提供了关于如何修饰这些结构域以影响使用这些系统产生的PUFA的比率和类型的关键信息。

[0328] 使用上述筛选方法可鉴定含有PUFA PKS系统的其它潜在候选菌株。本发明人鉴定的具有PUFA PKS系统的两个其它菌株是Schizochytrium limacium(SR21) Honda & Yokochi(IF032693)和Ulkenia(BP-5601)。除了在N2培养基(葡萄糖:60g/L;KH₂PO₄:4.0g/L;酵母提取物:1.0g/L;玉米浆:1mL/L;NH₄NO₃:1.0g/L;人工海盐(Reef Crystals):20g/L;所有上述浓度在去离子水中混合)中外按上述筛选两个菌株。对于Schizochytrium和Ulkenia这两个菌株,用于破囊壶菌23B的上述前三个筛选问题的答案均为是,即(Schizochytrium-有氧对缺氧的DHA%FAME为32→41%,58%的14:0/16:0/16:1,0%的前体)和(Ulkenia-有氧对缺氧的DHA%FAME为28→44%,63%的14:0/16:0/16:1,0%的前体),表明这些菌株是含有PUFA PKS系统的良好候选菌株。各菌株的最后两个问题得到了否定的答案:在S. limacium中脂肪从61%干重下降到22%干重,且DHA从21%下降到9%干重,Ulkenia中脂肪从59%下降到21%干重且DHA从16%下降到9%干重。这些Thraustochytrid微生物本文也声明作为用于上述实施方案的基因的另外来源。

[0329] 实施例3

[0330] 下面的实施例证实了Schizochytrium中的DHA和DPA合成不涉及膜结合的去饱和酶或脂肪酸延长酶,与对其它真核生物所述情况一样(Parker-Barnes等,2000,出处同上;Shanklin等,1998,出处同上)。

[0331] Schizochytrium积累大量富含DHA和鲱油酸(DPA;22:5 ω 6)的三酰甘油;例如,占干重30%的DHA+DPA。在通过延长/去饱和化途径合成20-和22-碳的PUFA的真核生物中,18-,20-和22-碳中间体的分子库相当大,因此使用[¹⁴C]-乙酸盐的体内标记实验揭示了预测的中间体的清楚前体-产物动力学。另外,给该生物体提供的外源性放射标记的中间体转变成最终的PUFA产物。

[0332] 在开始时以单次脉冲给2日龄的培养物补充[1-¹⁴C]乙酸盐。然后离心收获细胞样品并提取脂类。另外,通过测量离心之前和之后样品的放射活性估计细胞的[1-¹⁴C]乙酸盐吸收。以AgNO₃-TLC(溶剂,己烷:二乙基醚:乙酸,体积比为70:30:2)分离从总细

胞脂类衍生的脂肪酸甲酯。以气相色谱法证实脂肪酸带的身份,通过液闪计数测量它们的放射活性。结果表明 $[1-^{14}\text{C}]$ -乙酸盐被 *Schizochytrium* 细胞迅速吸收并掺入脂肪酸中,但是在最短标记时间(1分钟)时,DHA 含有在脂肪酸中回收的标记物的 31%且在 $[^{14}\text{C}]$ -乙酸盐掺入的 10-15 分钟期间和随后 24 小时的培养物生长中该百分数基本上保持不变(数据未显示)。同样,在整个实验中 DPA 代表了 10%的标记物。没有关于 16- 或 18- 碳脂肪酸与 22- 碳多不饱和脂肪酸之间存在前体-产物关系的证据。这些结果与从包含极小(很可能是酶结合的)中间体分子库的 $[^{14}\text{C}]$ -乙酸盐迅速合成 DHA 一致。

[0333] 接着,在含有 2mM DTT, 2mM EDTA, 和 10%甘油的 100mM 磷酸盐缓冲液 (pH 7.2) 中通过用玻璃珠搅拌破碎细胞。以 100,000g 离心无细胞匀浆 1 小时。在 25°C 下在补充了 20 μM 乙酰-CoA, 100 μM $[1-^{14}\text{C}]$ 丙二酰-CoA (0.9Gbpq/mol), 2mM NADH, 和 2mM NADPH 的匀浆缓冲液中培养总匀浆, 沉淀 (H-S 沉淀), 和上清 (H-S 上清) 分级分离成份的相当等分试样 60 分钟。提取试验物, 制备脂肪酸甲酯并在用 Instantimager (Packard Instruments, Meriden, CT) 检测放射性前按上文所述分离。结果表明来自 *Schizochytrium* 培养物的无细胞匀浆将 $[1-^{14}\text{C}]$ -丙二酰-CoA 掺入 DHA, DPA, 和饱和脂肪酸中(数据未显示)。100,000xg 的上清成份保留相同的生物合成活性但在膜沉淀中不存在该活性。将这些数据与细菌酶试验期间获得的数据(参见 Metz 等, 2001, 出处同上)相对照且可表明使用了不同的(可溶的)酰基受体分子。因此, *Schizochytrium* 中的 DHA 和 DPA 合成不涉及膜结合的去饱和酶或脂肪酸延长酶, 与对其它真核生物所述一样。

[0334] 尽管已经详细描述了本发明的各种实施方案, 但是显而易见的是本领域的技术人员可想到对这些实施方案进行改良和改编。然而应清楚地明白, 这些改良和改编在下面权利要求书所述本发明的范围内。

序列表

<110> 马泰克生物科学公司 (MARTEK BIOSCIENCES CORPORATION)

<120>PUFA 聚酮化合物合酶系统及其用途

<130>2997-29-PCT

<150>60/284, 066

<151>2001-04-16

<150>60/298, 796

<151>2001-06-15

<150>60/323, 269

<151>2001-09-18

<160>37

<170>PatentIn version 3.1

<210>1

<211>8730

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (8730)

<223>

<400>1

atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc	48
Met Ala Ala Arg Leu Gln Gln Lys Gly Gly Glu Met Asp Thr Arg	
1 5 10 15	
att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg	96
Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val	
20 25 30	

cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat	144
Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp	
35 40 45	
ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag	192
Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys	
50 55 60	
acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag	240
Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu	
65 70 75 80	
tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag	288
Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu	
85 90 95	
gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc	336
Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala	
100 105 110	
ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc	384
Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile	
115 120 125	
ggc tgc gtg ctc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc	432
Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe	
130 135 140	
tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg	480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met	
145 150 155 160	
ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc	528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala	
165 170 175	
aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac	576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn	
180 185 190	
gtc acc gcc ggt cgc tgc acc aac acc ttc aac ctc gac ggc atg aac	624
Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn	
195 200 205	
tgc gtt gtc gac gcc gca tgc gcc tcg tcc ctc atc gcc gtc aag gtc	672
Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val	
210 215 220	
gcc atc gac gag ctg ctc tac ggt gac tgc gac atg atg gtc acc ggt	720
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly	
225 230 235 240	

gcc acc tgc acg gat aac tcc atc ggc atg tac atg gcc ttc tcc aag	768
Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys	
245 250 255	
acc ccc gtg ttc tcc acg gac ccc agc gtg cgc gcc tac gac gaa aag	816
Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys	
260 265 270	
aca aag ggc atg ctc atc ggc gag ggc tcc gcc atg ctc gtc ctc aag	864
Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys	
275 280 285	
cgc tac gcc gac gcc gtc cgc gac ggc gat gag atc cac gct gtt att	912
Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile	
290 295 300	
cgc ggc tgc gcc tcc tcc agt gat ggc aag gcc gcc ggc atc tac acg	960
Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr	
305 310 315 320	
ccc acc att tcg ggc cag gag gag gcc ctc cgc cgc gcc tac aac cgc	1008
Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg	
325 330 335	
gcc tgt gtc gac ccg gcc acc gtc act ctc gtc gag ggt cac ggc acc	1056
Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr	
340 345 350	
ggt act ccc gtt ggc gac cgc atc gag ctc acc gcc ttg cgc aac ctc	1104
Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu	
355 360 365	
ttt gac aag gcc tac ggc gag ggc aac acc gaa aag gtc gct gtg ggc	1152
Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly	
370 375 380	
agc atc aag tcc agc atc ggc cat ctc aag gcc gtc gcc ggt ctc gcc	1200
Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala	
385 390 395 400	
ggt atg atc aag gtc atc atg gcg ctc aag cac aag act ctc ccg ggc	1248
Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly	
405 410 415	
acc atc aac gtc gac aac cca ccc aac ctc tac gac aac acg ccc atc	1296
Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile	
420 425 430	
aac gag tcc tcg ctc tac att aac acc atg aac cgc ccc tgg ttc ccg	1344
Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro	
435 440 445	

ccc cct ggt gtg ccc cgc cgc gcc ggc att tcg agc ttt ggc ttt ggt	1392
Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly	
450 455 460	
ggc gcc aac tac cac gcc gtc ctc gag gag gcc gag ccc gag cac acg	1440
Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr	
465 470 475 480	
acc gcg tac cgc ctc aac aag cgc ccg cag ccc gtg ctc atg atg gcc	1488
Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala	
485 490 495	
gcc acg ccc gcg gcc ctc cag tcg ctc tgc gag gcc cag ctc aag gag	1536
Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu	
500 505 510	
ttc gag gcc gcc atc aag gag aac gag acc gtc aag aac acc gcc tac	1584
Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr	
515 520 525	
atc aag tgc gtc aag ttc ggc gag cag ttc aaa ttc cct ggc tcc atc	1632
Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile	
530 535 540	
ccg gcc aca aac gcg cgc ctc ggc ttc ctc gtc aag gat gct gag gat	1680
Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp	
545 550 555 560	
gcc tgc tcc acc ctc cgt gcc atc tgc gcc caa ttc gcc aag gat gtc	1728
Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val	
565 570 575	
acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc cgc gcc	1776
Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala	
580 585 590	
aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc ggc cag	1824
Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln	
595 600 605	
ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac tgg ccc	1872
Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro	
610 615 620	
cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc aag gtc	1920
Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val	
625 630 635 640	
gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc tac ccg	1968
Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro	
645 650 655	

cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag atc tcc Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser 660 665 670	2016
ctc acc gcc tac tcgcag ccc tcg acc ctg gcc tgc gct ctc ggt gcc Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala 675 680 685	2064
ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc gcc ggc Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly 690 695 700	2112
cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc gtc gac His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp 705 710 715 720	2160
cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc atg ggc Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly 725 730 735	2208
ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc gtc att Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile 740 745 750	2256
ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc tgg ctc Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu 755 760 765	2304
ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc gtc gaa Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu 770 775 780	2352
ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc cgc gtc Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val 785 790 795 800	2400
gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg gag aac Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn 805 810 815	2448
gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc cgc acc Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr 820 825 830	2496
ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag acc tac Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr 835 840 845	2544
ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc agc gtc Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val 850 855 860	2592

aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg cgc atc	2640
Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile	
865 870 875 880	
ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc tcc gag	2688
Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu	
885 890 895	
acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac ccg gcc	2736
Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala	
900 905 910	
tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc cag ctc	2784
Ser Gly Thr Asp Ser Asp Tle Gln Leu Arg Asp Ala Ala Val Gln Leu	
915 920 925	
gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac gcc ccc	2832
Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro	
930 935 940	
gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc ctc cgc	2880
Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg	
945 950 955 960	
ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc cgc gac	2928
Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp	
965 970 975	
gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc gcc gca	2976
Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala	
980 985 990	
ccg ctc atc aag gcc ccg gag ccc gtt gtc gac gag gcc gcc aag cgc	3024
Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg	
995 1000 1005	
gag gcc gag cgt ctc cag aag gag ctt cag gat gcc cag cgc cag	3069
Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln	
1010 1015 1020	
ctc gac gac gcc aag cgc gcc gcc gcc gag gcc aac tcc aag ctc	3114
Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu	
1025 1030 1035	
gcc gct gcc aag gag gag gcc aag acc gcc gct gct tcg gcc aag	3159
Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys	
1040 1045 1050	
ccc gca gtt gac act gct gtt gtc gaa aag cat cgt gcc atc ctc	3204
Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu	
1055 1060 1065	

aag tcc atg ctc gcg gag ctc gat ggc tac gga tcg gtc gac gct	3249
Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala	
1070	1075
tct tcc ctc cag cag cag cag cag cag acg gcc ccc gcc ccg	3294
Ser Ser Leu Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro	
1085	1090
gtc aag gct gct gcg cct gcc gcc ccc gtt gcc tcg gcc cct gcc	3339
Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala	
1100	1105
ccg gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc	3384
Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val	
1115	1120
atg gag gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc	3429
Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile	
1130	1135
gag gct gac atg gag ctc gag acc gag ctc ggc att gac tcc atc	3474
Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile	
1145	1150
aag cgt gtc gag atc ctc tcc gag gtc cag gcc atg ctc aat gtc	3519
Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val	
1160	1165
gag gcc aag gat gtc gat gcc ctc agc cgc act cgc act gtt ggt	3564
Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly	
1175	1180
gag gtt gtc aac gcc atg aag gcc gag atc gct ggc agc tct gcc	3609
Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala	
1190	1195
ccg gcg cct gct gcc gct gct ccg gct ccg gcc aag gct gcc cct	3654
Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro	
1205	1210
gcc gcc gct gcg cct gct gtc tcg aac gag ctt ctc gag aag gcc	3699
Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala	
1220	1225
gag acc gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag	3744
Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu	
1235	1240
act gac atg atc gag tcc gac atg gag ctc gag act gag ctc ggc	3789
Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly	
1250	1255
	1260

att gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc	3834
Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala	
1265	1270 1275
atg ctc aac gtc gag gcc aag gac gtc gac gct ctc agc cgc act	3879
Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr	
1280	1285 1290
cgc act gtg ggt gag gtc gtc aac gcc atg aag gct gag atc gct	3924
Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala	
1295	1300 1305
ggt ggc tct gcc ccg gcg cct gcc gcc gct gcc cca ggt ccg gct	3969
Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Pro Gly Pro Ala	
1310	1315 1320
gct gcc gcc cct gcg cct gcc gcc gcc gcc cct gct gtc tcg aac	4014
Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn	
1325	1330 1335
gag ctt ctt gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc	4059
Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala	
1340	1345 1350
gcc aag act ggc tac gag act gac atg atc gag tcc gac atg gag	4104
Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu	
1355	1360 1365
ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag att	4149
Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile	
1370	1375 1380
ctc tcc gag gtc cag gcc atg ctc aac gtc gag gcc aag gac gtc	4194
Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val	
1385	1390 1395
gac gct ctc agc cgc acc cgc act gtt ggc gag gtc gtc gat gcc	4239
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala	
1400	1405 1410
atg aag gcc gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc	4284
Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala	
1415	1420 1425
gct gct cct gct ccg gct gct gcc gcc cct gcg cct gcc gcc cct	4329
Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Pro	
1430	1435 1440
gcg cct gct gtc tcg agc gag ctt ctc gag aag gcc gag act gtc	4374
Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val	
1445	1450 1455

gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act gac atg	4419
Val Met Glu yal Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met	
1460	1465
atc gag tcc gac atg gag ctc gag acc gag ctc ggc att gac tcc	4464
Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser	
1475	1480
atc aag cgt gtc gag att ctc tcc gag gtc cag gcc atg ctc aac	4509
Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn	
1490	1495
gtc gag gcc aag gac gtc gac gct ctc agc cgc acc cgc act gtt	4554
Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val	
1505	1510
ggc gag gtc gtc gat gcc atg aag gcc gag atc gct ggt ggc tct	4599
Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser	
1520	1525
gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc gcc	4644
Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala	
1535	1540
cct gcg cct gcc gcc cct gcg cct gcc gcc cct gcg cct gct gtc	4689
Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val	
1550	1555
tcg agc gag ctt ctc gag aag gcc gag act gtc gtc atg gag gtc	4734
Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val	
1565	1570
ctc gcc gcc aag act ggc tac gag act gac atg att gag tcc gac	4779
Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp	
1580	1585
atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc	4824
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val	
1595	1600
gag att ctc tcc gag gtt cag gcc atg ctc aac gtc gag gcc aag	4869
Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys	
1610	1615
gac gtc gac gct ctc agc cgc act cgc act gtt ggt gag gtc gtc	4914
Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val	
1625	1630
gat gcc atg aag gct gag atc gct ggc agc tcc gcc tcg gcg cct	4959
Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro	
1640	1645
	1650

gcc gcc gct gct cct gct ccg gct gct gcc gct cct gcg ccc gct Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala 1655 1660 1665	5004
gcc gcc gcc cct gct gtc tcg aac gag ctt ctc gag aaa gcc gag Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu 1670 1675 1680	5049
act gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag act Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr 1685 1690 1695	5094
gac atg atc gag tcc gac atg gag ctc gag act gag ctc ggc att Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile 1700 1705 1710	5139
gac tcc atc aag cgt gtc gag atc ctc tcc gag gtt cag gcc atg Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met 1715 1720 1725	5184
ctc aac gtc gag gcc aag gac gtc gat gcc ctc agc cgc acc cgc Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg 1730 1735 1740	5229
act gtt ggc gag gtt gtc gat gcc atg aag gcc gag atc gct ggt Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly 1745 1750 1755	5274
ggc tct gcc ccg gcg cct gcc gcc gct gcc cct gct ccg gct gcc Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala 1760 1765 1770	5319
gcc gcc cct gct gtc tcg aac gag ctt ctc gag aag gcc gag act Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr 1775 1780 1785	5364
gtc gtc atg gag gtc ctc gcc gcc aag act ggc tac gag acc gac Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp 1790 1795 1800	5409
atg atc gag tcc gac atg gag ctc gag acc gag ctc ggc att gac Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp 1805 1810 1815	5454
tcc atc aag cgt gtc gag att ctc tcc gag gtt cag gcc atg ctc Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu 1820 1825 1830	5499
aac gtc gag gcc aag gac gtc gat gct ctc agc cgc act cgc act Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr 1835 1840 1845	5544

gtt ggc gag gtc gtc gat gcc atg aag gct gag atc gcc ggc agc Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser 1850 1855 1860	5589
tcc gcc ccg gcg cct gcc gcc gct gct cct gct ccg gct gct gcc Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala 1865 1870 1875	5634
gct cct gcg ccc gct gcc gct gcc cct gct gtc tcg agc gag ctt Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu 1880 1885 1890	5679
ctc gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys 1895 1900 1905	5724
act ggc tac gagact gac atg att gag tcc gac atg gag ctc gag Thr Gly Tyr GluThr Asp Met Ile Glu Ser Asp Met Glu Leu Glu 1910 1915 1920	5769
act gag ctc ggc att gac tcc atc aag cgt gtc gag atc ctc tcc Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser 1925 1930 1935	5814
gag gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gat gcc Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala 1940 1945 1950	5859
ctc agc cgc acc cgc act gtt ggc gag gtt gtc gat gcc atg aag Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys 1955 1960 1965	5904
gcc gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc gct gcc Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala 1970 1975 1980	5949
cct gct ccg gct gcc gcc gcc cct gct gtc tcg aac gag ctt ctt pro Ala Pro Ala Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu 1985 1990 1995	5994
gag aag gcc gag acc gtc gtc atg gag gtc ctc gcc gcc aag act Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr 2000 2005 2010	6039
ggc tac gag acc gac atg atc gag tcc gac atg gag ctc gag acc Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr 2015 2020 2025	6084
gag ctc ggc att gac tcc atc aag cgt gtc gag att ctc tcc gag Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu 2030 2035 2040	6129

gtt cag gcc atg ctc aac gtc gag gcc aag gac gtc gac gct ctc Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu 2045 2050 2055	6174
agc cgc act cgc act gtt ggc gag gtc gtc gat gcc atg aag gct Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala 2060 2065 2070	6219
gag atc gct ggt ggc tct gcc ccg gcg cct gcc gcc gct gct cct Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro 2075 2080 2085	6264
gcc tcg gct ggc gcc gcg cct gcg gtc aag att gac tcg gtc cac Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His 2090 2095 2100	6309
ggc gct gac tgt gat gat ctt tcc ctg atg cac gcc aag gtg gtt Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val 2105 2110 2115	6354
gac atc cgc cgc ccg gac gag ctc atc ctg gag cgc ccc gag aac Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn 2120 2125 2130	6399
cgc ccc gtt ctc gtt gtc gat gac ggc agc gag ctc acc ctc gcc Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala 2135 2140 2145	6444
ctg gtc cgc gtc ctc ggc gcc tgc gcc gtt gtc ctg acc ttt gag Leu Val Arg Val Leu Gly Ala Cys Val Val Val Leu Thr Phe Glu 2150 2155 2160	6489
ggt ctc cag ctc gct cag cgc gct ggt gcc gct gcc atc cgc cac Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His 2165 2170 2175	6534
gtg ctc gcc aag gat ctt tcc gcg gag agc gcc gag aag gcc atc Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile 2180 2185 2190	6579
aag gag gcc gag cag cgc ttt ggc gct ctc ggc gcc ttc atc tcg Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser 2195 2200 2205	6624
cag cag gcg gag cgc ttc gag ccc gcc gaa atc ctc gcc ttc acg Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr 2210 2215 2220	6669
ctc atg tgc gcc aag ttc gcc aag gct tcc ctc tgc acg gct gtg Leu Met Cys Ala Lys Phe Ala Lys Ala Ser Leu Cys Thr Ala Val 2225 2230 2235	6714

gct ggc ggc cgc ccg gcc ttt atc ggt gtg gcg cgc ctt gac ggc Ala Gly Gly Arg Pro Ala Phe Ile Gly Val Ala Arg Leu Asp Gly 2240 2245 2250	6759
cgc ctc gga ttc act tcg cag ggc act tct gac gcg ctc aag cgt Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser Asp Ala Leu Lys Arg 2255 2260 2265	6804
gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag acc atc ggc ctc Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys Thr Ile Gly Leu 2270 2275 2280	6849
gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg gac att gct Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val Asp Ile Ala 2285 2290 2295	6894
cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg cgc gag Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val Arg Glu 2300 2305 2310	6939
atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc gca Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly Ala 2315 2320 2325	6984
aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly 2330 2335 2340	7029
aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser 2345 2350 2355	7074
ggc ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr 2360 2365 2370	7119
cgc cag atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys 2375 2380 2385	7164
gtc tct gcg agc gaa ccg gca tgg tgc gct ggc atc act gac gag Val Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu 2390 2395 2400	7209
aag gct gtg caa aag gct gct acc cag gag ctc aag cgc gcc ttt Lys Ala Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe 2405 2410 2415	7254
agc gct ggc gag ggc ccc aag ccc acg ccc cgc gct gtc act aag Ser Ala Gly Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys 2420 2425 2430	7299

ctt gtg ggc tct gtt ctt ggc gct cgc gag gtg cgc agc tct att	7344
Leu Val Gly Ser Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile	
2435	2440 2445
gct gcg att gaa gcg ctc ggc ggc aag gcc atc tac tcg tcg tgc	7389
Ala Ala Ile Glu Ala Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys	
2450	2455 2460
gac gtg aac tct gcc gcc gac gtg gcc aag gcc gtg cgc gat gcc	7434
Asp Val Asn Ser Ala Ala Asp Val Ala Lys Ala Val Arg Asp Ala	
2465	2470 2475
gag tcc cag ctc ggt gcc cgc gtc tcg ggc atc gtt cat gcc tcg	7479
Glu Ser Gln Leu Gly Ala Arg Val Ser Gly Ile Val His Ala Ser	
2480	2485 2490
ggc gtg ctc cgc gac cgt ctc atc gag aag aag ctc ccc gac gag	7524
Gly Val Leu Arg Asp Arg Leu Ile Glu Lys Lys Leu Pro Asp Glu	
2495	2500 2505
ttc gac gcc gtc ttt ggc acc aag gtc acc ggt ctc gag aac ctc	7569
Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu	
2510	2515 2520
ctc gcc gcc gtc gac cgc gcc aac ctc aag cac atg gtc ctc ttc	7614
Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe	
2525	2530 2535
agc tcg ctc gcc ggc ttc cac ggc aac gtc ggc cag tct gac tac	7659
Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr	
2540	2545 2550
gcc atg gcc aac gag gcc ctt aac aag atg ggc ctc gag ctc gcc	7704
Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala	
2555	2560 2565
aag gac gtc tcg gtc aag tcg atc tgc ttc ggt ccc tgg gac ggt	7749
Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly	
2570	2575 2580
ggc atg gtg acg ccg cag ctc aag aag cag ttc cag gag atg ggc	7794
Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly	
2585	2590 2595
gtg cag atc atc ccc cgc gag ggc ggc gct gat acc gtg gcg cgc	7839
Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg	
2600	2605 2610
atc gtg ctc ggc tcc tcg ccg gct gag atc ctt gtc ggc aac tgg	7884
Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp	
2615	2620 2625

cgc acc ccg tcc aag aag gtc ggc tcg gac acc atc acc ctg cac	7929
Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His	
2630	2635
2640	
cgc aag att tcc gcc aag tcc aac ccc ttc ctc gag gac cac gtc	7974
Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val	
2645	2650
2655	
atc cag ggc cgc cgc gtg ctg ccc atg acg ctg gcc att ggc tcg	8019
Tle Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser	
2660	2665
2670	
ctc gcg gag acc tgc ctc ggc ctc ttc ccc ggc tac tcg ctc tgg	8064
Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp	
2675	2680
2685	
gcc att gac gac gcc cag ctc ttc aag ggt gtc act gtc gac ggc	8109
Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly	
2690	2695
2700	
gac gtc aac tgc gag gtg acc ctc acc ccg tcg acg gcg ccc tcg	8154
Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser	
2705	2710
2715	
ggc cgc gtc aac gtc cag gcc acg ctc aag acc ttt tcc agc ggc	8199
Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly	
2720	2725
2730	
aag ctg gtc ccg gcc tac cgc gcc gtc atc gtg ctc tcc aac cag	8244
Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln	
2735	2740
2745	
ggc gcg ccc ccg gcc aac gcc acc atg cag ccg ccc tcg ctc gat	8289
Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp	
2750	2755
2760	
gcc gat ccg gcg ctc cag ggc tcc gtc tac gac ggc aag acc ctc	8334
Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu	
2765	2770
2775	
ttc cac ggc ccg gcc ttc cgc ggc atc gat gac gtg ctc tcg tgc	8379
Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys	
2780	2785
2790	
acc aag agc cag ctt gtg gcc aag tgc agc gct gtc ccc ggc tcc	8424
Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser	
2795	2800
2805	
gac gcc gct cgc ggc gag ttt gcc acg gac act gac gcc cat gac	8469
Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp	
2810	2815
2820	

```

ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg      8514
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val
      2825                      2830                      2835
cgc cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc      8559
Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg
      2840                      2845                      2850
atc gtc cag cac cgc ccg gtc  ccg cag gac aag cccttc tac att      8604
Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile
      2855                      2860                      2865
acc ctc cgc tcc aac cag tcg ggc ggt cac tcc cag cac aag cac      8649
Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His
      2870                      2875                      2880
gcc ctt cag ttc cac aac gag cag ggc gat ctc ttc att gat gtc      8694
Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val
      2885                      2890                      2895
cag gct tcg gtc atc gcc acg gac agc ctt gcc ttc                      8730
Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala Phe
      2900                      2905                      2910

```

<210>2

<211>2910

<212>PRT

<213>Schizochytrium sp.

<400>2

```

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

```

Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile		
115	120	125
Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe		
130	135	140
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met		
145	150	155
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala		
165	170	175
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn		
180	185	190
Val Thr Ala Gly Arg Cys Thr Asn Thr Phe Asn Leu Asp Gly Met Asn		
195	200	205
Cys Val Val Asp Ala Ala Cys Ala Ser Ser Leu Ile Ala Val Lys Val		
210	215	220
Ala Ile Asp Glu Leu Leu Tyr Gly Asp Cys Asp Met Met Val Thr Gly		
225	230	235
Ala Thr Cys Thr Asp Asn Ser Ile Gly Met Tyr Met Ala Phe Ser Lys		
245	250	255
Thr Pro Val Phe Ser Thr Asp Pro Ser Val Arg Ala Tyr Asp Glu Lys		
260	265	270
Thr Lys Gly Met Leu Ile Gly Glu Gly Ser Ala Met Leu Val Leu Lys		
275	280	285
Arg Tyr Ala Asp Ala Val Arg Asp Gly Asp Glu Ile His Ala Val Ile		
290	295	300
Arg Gly Cys Ala Ser Ser Ser Asp Gly Lys Ala Ala Gly Ile Tyr Thr		
305	310	315
Pro Thr Ile Ser Gly Gln Glu Glu Ala Leu Arg Arg Ala Tyr Asn Arg		
325	330	335
Ala Cys Val Asp Pro Ala Thr Val Thr Leu Val Glu Gly His Gly Thr		
340	345	350
Gly Thr Pro Val Gly Asp Arg Ile Glu Leu Thr Ala Leu Arg Asn Leu		
355	360	365
Phe Asp Lys Ala Tyr Gly Glu Gly Asn Thr Glu Lys Val Ala Val Gly		
370	375	380
Ser Ile Lys Ser Ser Ile Gly His Leu Lys Ala Val Ala Gly Leu Ala		
385	390	395
Gly Met Ile Lys Val Ile Met Ala Leu Lys His Lys Thr Leu Pro Gly		
405	410	415
Thr Ile Asn Val Asp Asn Pro Pro Asn Leu Tyr Asp Asn Thr Pro Ile		

420	425	430
Asn Glu Ser Ser Leu Tyr Ile Asn Thr Met Asn Arg Pro Trp Phe Pro		
435	440	445
Pro Pro Gly Val Pro Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly		
450	455	460
Gly Ala Asn Tyr His Ala Val Leu Glu Glu Ala Glu Pro Glu His Thr		
465	470	475
Thr Ala Tyr Arg Leu Asn Lys Arg Pro Gln Pro Val Leu Met Met Ala		480
485	490	495
Ala Thr Pro Ala Ala Leu Gln Ser Leu Cys Glu Ala Gln Leu Lys Glu		
500	505	510
Phe Glu Ala Ala Ile Lys Glu Asn Glu Thr Val Lys Asn Thr Ala Tyr		
515	520	525
Ile Lys Cys Val Lys Phe Gly Glu Gln Phe Lys Phe Pro Gly Ser Ile		
530	535	540
Pro Ala Thr Asn Ala Arg Leu Gly Phe Leu Val Lys Asp Ala Glu Asp		
545	550	555
Ala Cys Ser Thr Leu Arg Ala Ile Cys Ala Gln Phe Ala Lys Asp Val		
565	570	575
Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe Arg Ala		
580	585	590
Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser Gly Gln		
595	600	605
Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn Trp Pro		
610	615	620
Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser Lys Val		
625	630	635
Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu Tyr Pro		
645	650	655
Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys Ile Ser		
660	665	670
Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu Gly Ala		
675	680	685
Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala Ala Gly		
690	695	700
His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys Val Asp		
705	710	715
Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile Met Gly		
725	730	735

Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala Val Ile			
740	745	750	
Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val Trp Leu			
755	760	765	
Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser Val Glu			
770	775	780	
Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe Arg Val			
785	790	795	800
Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met Glu Asn			
805	810	815	
Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe Arg Thr			
820	825	830	
Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu Thr Tyr			
835	840	845	
Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser Ser Val			
850	855	860	
Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala Arg Ile			
865	870	875	880
Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val Ser Glu			
885	890	895	
Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn Pro Ala			
900	905	910	
Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val Gln Leu			
915	920	925	
Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp Ala Pro			
930	935	940	
Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr Leu Arg			
945	950	955	960
Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val Arg Asp			
965	970	975	
Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly Ala Ala			
980	985	990	
Pro Leu Ile Lys Ala Pro Glu Pro Val Val Asp Glu Ala Ala Lys Arg			
995	1000	1005	
Glu Ala Glu Arg Leu Gln Lys Glu Leu Gln Asp Ala Gln Arg Gln			
1010	1015	1020	
Leu Asp Asp Ala Lys Arg Ala Ala Ala Glu Ala Asn Ser Lys Leu			
1025	1030	1035	
Ala Ala Ala Lys Glu Glu Ala Lys Thr Ala Ala Ala Ser Ala Lys			

1040	1045	1050
Pro Ala Val Asp Thr Ala Val Val Glu Lys His Arg Ala Ile Leu		
1055	1060	1065
Lys Ser Met Leu Ala Glu Leu Asp Gly Tyr Gly Ser Val Asp Ala		
1070	1075	1080
Ser Ser Leu Gln Gln Gln Gln Gln Gln Thr Ala Pro Ala Pro		
1085	1090	1095
Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser Ala Pro Ala		
1100	1105	1110
Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val		
1115	1120	1125
Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile		
1130	1135	1140
Glu Ala Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile		
1145	1150	1155
Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val		
1160	1165	1170
Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly		
1175	1180	1185
Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala		
1190	1195	1200
Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Lys Ala Ala Pro		
1205	1210	1215
Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala		
1220	1225	1230
Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu		
1235	1240	1245
Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly		
1250	1255	1260
Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala		
1265	1270	1275
Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr		
1280	1285	1290
Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala		
1295	1300	1305
Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala		
1310	1315	1320
Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn		
1325	1330	1335

Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala		
1340	1345	1350
Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu		
1355	1360	1365
Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile		
1370	1375	1380
Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val		
1385	1390	1395
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala		
1400	1405	1410
Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala		
1415	1420	1425
Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Pro		
1430	1435	1440
Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val		
1445	1450	1455
Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met		
1460	1465	1470
Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser		
1475	1480	1485
Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn		
1490	1495	1500
Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val		
1505	1510	1515
Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser		
1520	1525	1530
Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala		
1535	1540	1545
Pro Ala Pro Ala Ala Pro Ala Pro Ala Ala Pro Ala Pro Ala Val		
1550	1555	1560
Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val		
1565	1570	1575
Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp		
1580	1585	1590
Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val		
1595	1600	1605
Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys		
1610	1615	1620
Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val		

1625	1630	1635
Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro		
1640	1645	1650
Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala		
1655	1660	1665
Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu		
1670	1675	1680
Thr Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr		
1685	1690	1695
Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile		
1700	1705	1710
Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met		
1715	1720	1725
Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg		
1730	1735	1740
Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly		
1745	1750	1755
Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala		
1760	1765	1770
Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr		
1775	1780	1785
Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp		
1790	1795	1800
Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp		
1805	1810	1815
Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu		
1820	1825	1830
Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr		
1835	1840	1845
Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser		
1850	1855	1860
Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala		
1865	1870	1875
Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu		
1880	1885	1890
Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys		
1895	1900	1905
Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu		
1910	1915	1920

Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser		
1925	1930	1935
Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala		
1940	1945	1950
Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys		
1955	1960	1965
Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala		
1970	1975	1980
Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu Leu Leu		
1985	1990	1995
Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr		
2000	2005	2010
Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr		
2015	2020	2025
Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu		
2030	2035	2040
Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu		
2045	2050	2055
Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala		
2060	2065	2070
Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro		
2075	2080	2085
Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser Val His		
2090	2095	2100
Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val Val		
2105	2110	2115
Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn		
2120	2125	2130
Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala		
2135	2140	2145
Leu Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu		
2150	2155	2160
Gly Leu Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His		
2165	2170	2175
Val Leu Ala Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile		
2180	2185	2190
Lys Glu Ala Glu Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser		
2195	2200	2205
Gln Gln Ala Glu Arg Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr		

2210	2215	2220
Leu Met Cys Ala Lys Phe	Ala Lys Ala Ser Leu Cys Thr Ala Val	
2225	2230	2235
Ala Gly Gly Arg Pro Ala Phe	Ile Gly Val Ala Arg Leu Asp Gly	
2240	2245	2250
Arg Leu Gly Phe Thr Ser Gln	Gly Thr Ser Asp Ala Leu Lys Arg	
2255	2260	2265
Ala Gln Arg Gly Ala Ile Phe	Gly Leu Cys Lys Thr Ile Gly Leu	
2270	2275	2280
Glu Trp Ser Glu Ser Asp Val	Phe Ser Arg Gly Val Asp Ile Ala	
2285	2290	2295
Gln Gly Met His Pro Glu Asp	Ala Ala Val Ala Ile Val Arg Glu	
2300	2305	2310
Met Ala Cys Ala Asp Ile Arg	Ile Arg Glu Val Gly Ile Gly Ala	
2315	2320	2325
Asn Gln Gln Arg Cys Thr Ile	Arg Ala Ala Lys Leu Glu Thr Gly	
2330	2335	2340
Asn Pro Gln Arg Gln Ile Ala	Lys Asp Asp Val Leu Leu Val Ser	
2345	2350	2355
Gly Gly Ala Arg Gly Ile Thr	Pro Leu Cys Ile Arg Glu Ile Thr	
2360	2365	2370
Arg Gln Ile Ala Gly Gly Lys	Tyr Ile Leu Leu Gly Arg Ser Lys	
2375	2380	2385
Val Ser Ala Ser Glu Pro Ala	Trp Cys Ala Gly Ile Thr Asp Glu	
2390	2395	2400
Lys Ala Val Gln Lys Ala Ala	Thr Gln Glu Leu Lys Arg Ala Phe	
2405	2410	2415
Ser Ala Gly Glu Gly Pro Lys	Pro Thr Pro Arg Ala Val Thr Lys	
2420	2425	2430
Leu Val Gly Ser Val Leu Gly	Ala Arg Glu Val Arg Ser Ser Ile	
2435	2440	2445
Ala Ala Ile Glu Ala Leu Gly	Gly Lys Ala Ile Tyr Ser Ser Cys	
2450	2455	2460
Asp Val Asn Ser Ala Ala Asp	Val Ala Lys Ala Val Arg Asp Ala	
2465	2470	2475
Glu Ser Gln Leu Gly Ala Arg	Val Ser Gly Ile Val His Ala Ser	
2480	2485	2490
Gly Val Leu Arg Asp Arg Leu	Ile Glu Lys Lys Leu Pro Asp Glu	
2495	2500	2505

Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly Leu Glu Asn Leu		
2510	2515	2520
Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met Val Leu Phe		
2525	2530	2535
Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser Asp Tyr		
2540	2545	2550
Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu Ala		
2555	2560	2565
Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly		
2570	2575	2580
Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly		
2585	2590	2595
Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg		
2600	2605	2610
Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp		
2615	2620	2625
Arg Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His		
2630	2635	2640
Arg Lys Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val		
2645	2650	2655
Ile Gln Gly Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser		
2660	2665	2670
Leu Ala Glu Thr Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp		
2675	2680	2685
Ala Ile Asp Asp Ala Gln Leu Phe Lys Gly Val Thr Val Asp Gly		
2690	2695	2700
Asp Val Asn Cys Glu Val Thr Leu Thr Pro Ser Thr Ala Pro Ser		
2705	2710	2715
Gly Arg Val Asn Val Gln Ala Thr Leu Lys Thr Phe Ser Ser Gly		
2720	2725	2730
Lys Leu Val Pro Ala Tyr Arg Ala Val Ile Val Leu Ser Asn Gln		
2735	2740	2745
Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro Pro Ser Leu Asp		
2750	2755	2760
Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly Lys Thr Leu		
2765	2770	2775
Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu Ser Cys		
2780	2785	2790
Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly Ser		

2795	2800	2805
Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr	Asp Ala His Asp	
2810	2815	2820
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met	Leu Val Trp Val	
2825	2830	2835
Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn	Ser Ile Gln Arg	
2840	2845	2850
Ile Val Gln His Arg Pro Val Pro Gln Asp Lys	Pro Phe Tyr Ile	
2855	2860	2865
Thr Leu Arg Ser Asn Gln Ser Gly Gly His Ser	Gln His Lys His	
2870	2875	2880
Ala Leu Gln Phe His Asn Glu Gln Gly Asp Leu	Phe Ile Asp Val	
2885	2890	2895
Gln Ala Ser Val Ile Ala Thr Asp Ser Leu Ala	Phe	
2900	2905	2910

<210>3

<211>6177

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (6177)

<223>

<400>3

atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag	48
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys	
1 5 10 15	
cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc	96
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr	
20 25 30	
aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag	144
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys	
35 40 45	
gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac	192
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr	
50 55 60	

aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac	240
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr	
65 70 75 80	
ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc	288
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu	
85 90 95	
aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca	336
Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr	
100 105 110	
cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc	384
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	
115 120 125	
cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc	432
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu	
130 135 140	
ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag	480
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln	
145 150 155 160	
tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc	528
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala	
165 170 175	
tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc	576
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val	
180 185 190	
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat	624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp	
195 20 0205	
cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc	672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys	
210 215 220	
ctg ccg gag ccc ttt ttc atc ctt tcg ggc ttt tcc acc ttc cag gcc	720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala	
225 230 235 240	
atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac	768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp	
245 250 255	
agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag	816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys	
260 265 270	

cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt	864
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu	
275 280 285	
ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc	912
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro	
290 295 300	
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att	960
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile	
305 310 315 320	
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc	1008
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly	
325 330 335	
acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt	1056
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe	
340 345 350	
gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac	1104
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His	
355 360 365	
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc	1152
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser	
370 375 380	
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc	1200
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr	
385 390 395 400	
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag	1248
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu	
405 410 415	
acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt	1296
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly	
420 425 430	
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc	1344
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala	
435 440 445	
gcc tgc acg ggc cac gac tcc att tct gcg ctc tcg gcc cgc tgc ggc	1392
Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly	
450 455 460	
ggt gaa agc aac atg cgc atc gcc atc act ggt atg gac gcc acc ttt	1440
Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe	
465 470 475 480	

ggc gct ctc aag gga ctc gac gcc ttc gag cgc gcc att tac acc ggc	1488
Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly	
485 490 495	
gct cac ggt gcc atc cca ctc cca gaa aag cgc tgg cgc ttt ctc ggc	1536
Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly	
500 505 510	
aag gac aag gac ttt ctt gac ctc tgc ggc gtc aag gcc acc ccg cac	1584
Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His	
515 520 525	
ggc tgc tac att gaa gat gtt gag gtc gac ttc cag cgc ctc cgc acg	1632
Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr	
530 535 540	
ccc atg acc cct gaa gac atg ctc ctc cct cag cag ctt ctg gcc gtc	1680
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val	
545 550 555 560	
acc acc att gac cgc gcc atc ctc gac tcg gga atg aaa aag ggt ggc	1728
Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly	
565 570 575	
aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac ctc gag ctc tac cgt	1776
Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg	
580 585 590	
cac cgt gct cgc gtc gct ctc aag gag cgc gtc cgc cct gaa gcc tcc	1824
His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser	
595 600 605	
aag aag ctc aat gac atg atg cag tac att aac gac tgc ggc aca tcc	1872
Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser	
610 615 620	
aca tcg tac acc tcg tac att ggc aac ctc gtc gcc acg cgc gtc tcg	1920
Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser	
625 630 635 640	
tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg atc acc gag ggc aac	1968
Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn	
645 650 655	
aac tcc gtc tac cgc tgc gcc gag ctc ggc aag tac ctc ctc gag acc	2016
Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr	
660 665 670	
ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc gat ctc tgc ggc agt	2064
Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser	
675 680 685	

gcc gaa aac ctt tac gtc aag tct cgc cgc ttc aag gtg tcc acc tcc	2112
Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser	
690 695 700	
gat acc ccg cgc gcc agc ttt gac gcc gcc gcc gat ggc tac ttt gtc	2160
Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val	
705 710 715 720	
ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt gag act agc tgc acc	2208
Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr	
725 730 735	
aag gac gac cgt atc tac gct tgc atg gat gcc atc gtc cct ggc aac	2256
Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn	
740 745 750	
gtc cct agc gcc tgc ttg cgc gag gcc ctc gac cag gcg cgc gtc aag	2304
Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys	
755 760 765	
ccg ggc gat atc gag atg ctc gag ctc agc gcc gac tcc gcc cgc cac	2352
Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His	
770 775 780	
ctc aag gac ccg tcc gtc ctg ccc aag gag ctc act gcc gag gag gaa	2400
Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu	
785 790 795 800	
atc ggc ggc ctt cag acg atc ctt cgt gac gat gac aag ctc ccg cgc	2448
Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg	
805 810 815	
aac gtc gca acg ggc agt gtc aag gcc acc gtc ggt gac acc ggt tat	2496
Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr	
820 825 830	
gcc tct ggt gct gcc agc ctc atc aag gct gcg ctt tgc atc tac aac	2544
Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn	
835 840 845	
cgc tac ctg ccc agc aac ggc gac gac tgg gat gaa ccc gcc cct gag	2592
Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu	
850 855 860	
gcg ccc tgg gac agc acc ctc ttt gcg tgc cag acc tcg cgc gct tgg	2640
Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp	
865 870 875 880	
ctc aag aac cct ggc gag cgt cgc tat gcg gcc gtc tcg ggc gtc tcc	2688
Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser	
885 890 895	

gag acg cgc tgc tgc tat tcc gtg ctc ctc tcc gaa gcc gag ggc cac	2736
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His	
900 905 910	
tac gag cgc gag aac cgc atc tgc ctc gac gag gag gcg ccc aag ctc	2784
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu	
915 920 925	
att gtg ctt cgc gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac	2832
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp	
930 935 940	
aag atc cgc gag cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag	2880
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu	
945 950 955 960	
tcc gag ctc aag gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc gcc	2928
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly	
965 970 975	
gag acc ctt gcc cag gat gcc gct tct tca ggc tcg caa aag ccc ctc	2976
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu	
980 985 990	
gct ctc agc ctc gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag	3024
Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu	
995 1000 1005	
ctc gcg gcc aag ggt atc ccg cgc tgc ctc aag atg cgc cgc gat	3069
Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp	
1010 1015 1020	
tgg agc tcc cct gct ggc agc cgc tac gcg cct gag ccg ctc gcc	3114
Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala	
1025 1030 1035	
agc gac cgc gtc gcc ttc atg tac ggc gaa ggt cgc agc cct tac	3159
Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr	
1040 1045 1050	
tac ggc atc acc caa gac att cac cgc att tgg ccc gaa ctc cac	3204
Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His	
1055 1060 1065	
gag gtc atc aac gaa aag acg aac cgt ctc tgg gcc gaa ggc gac	3249
Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp	
1070 1075 1080	
cgc tgg gtc atg ccg cgc gcc agc ttc aag tcg gag ctc gag agc	3294
Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser	
1085 1090 1095	

cag cag caa gag ttt gat cgc aac atg att gaa atg ttc cgt ctt	3339
Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu	
1100 1105 1110	
gga atc ctc acc tca att gcc ttc acc aat ctg gcg cgc gac gtt	3384
Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val	
1115 1120 1125	
ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt ctt ggc gag	3429
Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu	
1130 1135 1140	
att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc atc tcc	3474
Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser	
1145 1150 1155	
gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac aag	3519
Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys	
1160 1165 1170	
gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att	3564
Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile	
1175 1180 1185	
cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg	3609
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val	
1190 1195 1200	
cgc ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc	3654
Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser	
1205 1210 1215	
aag tac gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc	3699
Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu	
1220 1225 1230	
att agc ggc aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc	3744
Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu	
1235 1240 1245	
ggt ggc aac att cct gcg ctt ccc gtg acc cag ggc atg tgc ggc	3789
Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly	
1250 1255 1260	
cac tgc ccc gag gtg gga cct tat acc aag gat atc gcc aag atc	3834
His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile	
1265 1270 1275	
cat gcc aac ctt gag ttc ccc gtt gtc gac ggc ctt gac ctc tgg	3879
His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp	
1280 1285 1290	

acc aca atc aac cag aag cgc ctc gtg cca cgc gcc acg ggc gcc	3924
Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala	
1295 1300 1305	
aag gac gaa tgg gcc cct tct tcc ttt ggc gag tac gcc ggc cag	3969
Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln	
1310 1315 1320	
ctc tac gag aag cag gct aac ttc ccc caa atc gtc gag acc att	4014
Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile	
1325 1330 1335	
tac aag caa aac tac gac gtc ttt gtc gag gtt ggg ccc aac aac	4059
Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn	
1340 1345 1350	
cac cgt agc acc gca gtg cgc acc acg ctt ggt ccc cag cgc aac	4104
His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn	
1355 1360 1365	
cac ctt gct ggc gcc atc gac aag cag aac gag gat gct tgg acg	4149
His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr	
1370 1375 1380	
acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt gtt cct	4194
Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro	
1385 1390 1395	
ggc gtc acg atc tcg ccg ctg tac cac tcc aag ctt gtg gcg gag	4239
Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu	
1400 1405 1410	
gct cag gct tgc tac gct gcg ctc tgc aag ggt gaa aag ccc aag	4284
Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys	
1415 1420 1425	
aag aac aag ttt gtg cgc aag att cag ctc aac ggt cgc ttc aac	4329
Lys Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn	
1430 1435 1440	
agc aag gcg gac ccc atc tcc tcg gcc gat ctt gcc agc ttt ccg	4374
Ser Lys Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro	
1445 1450 1455	
cct gcg gac cct gcc att gaa gcc gcc atc tcg agc cgc atc atg	4419
Pro Ala Asp Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met	
1460 1465 1470	
aag cct gtc gct ccc aag ttc tac gcg cgt ctc aac att gac gag	4464
Lys Pro Val Ala Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu	
1475 1480 1485	

cag gac gag acc cga gat ccg atc ctc aac aag gac aac gcg ccg	4509
Gln Asp Glu Thr Arg Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro	
1490	1495 1500
tct tct tct tct tct tct tct tct tct tct tct tct tct tct tct	4554
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser	
1505	1510 1515
ccg tcg cct gct cct tcg gcc ccc gtg caa aag aag gct gct ccc	4599
Pro Ser Pro Ala Pro Ser Ala Pro Val Gln Lys Lys Ala Ala Pro	
1520	1525 1530
gcc gcg gag acc aag gct gtt gct tcg gct gac gca ctt cgc agt	4644
Ala Ala Glu Thr Lys Ala Val Ala Ser Ala Asp Ala Leu Arg Ser	
1535	1540 1545
gcc ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt	4689
Ala Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser	
1550	1555 1560
gcc tcc ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc	4734
Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val	
1565	1570 1575
att gtg ccg ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc	4779
Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe	
1580	1585 1590
atg aaa acg tac ggt gtt tcg gcg cct ctg tac acg ggc gcc atg	4824
Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met	
1595	1600 1605
gcc aag ggc att gcc tct gcg gac ctc gtc att gcc gcc ggc cgc	4869
Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg	
1610	1615 1620
cag ggc atc ctt gcg tcc ttt ggc gcc ggc gga ctt ccc atg cag	4914
Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln	
1625	1630 1635
gtt gtg cgt gag tcc atc gaa aag att cag gcc gcc ctg ccc aat	4959
Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn	
1640	1645 1650
ggc ccg tac gct gtc aac ctt atc cat tct ccc ttt gac agc aac	5004
Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn	
1655	1660 1665
ctc gaa aag ggc aat gtc gat ctc ttc ctc gag aag ggt gtc acc	5049
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr	
1670	1675 1680

ttt gtc gag gcc tcg gcc ttt atg acg ctc acc ccg cag gtc gtg	5094
Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val	
1685 1690 1695	
cgg tac cgc gcg gct ggc ctc acg cgc aac gcc gac ggc tcg gtc	5139
Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val	
1700 1705 1710	
aac atc cgc aac cgt atc att ggc aag gtc tcg cgc acc gag ctc	5184
Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu	
1715 1720 1725	
gcc gag atg ttc atg cgt cct gcg ccc gag cac ctt ctt cag aag	5229
Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu Gln Lys	
1730 1735 1740	
ctc att gct tcc ggc gag atc aac cag gag cag gcc gag ctc gcc	5274
Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu Ala	
1745 1750 1755	
cgc cgt gtt ccc gtc gct gac gac atc gcg gtc gaa gct gac tcg	5319
Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser	
1760 1765 1770	
ggt ggc cac acc gac aac cgc ccc atc cac gtc att ctg ccc ctc	5364
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu	
1775 1780 1785	
atc atc aac ctt cgc gac cgc ctt cac cgc gag tgc ggc tac ccg	5409
Ile Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro	
1790 1795 1800	
gcc aac ctt cgc gtc cgt gtg ggc gcc ggc ggt ggc att ggg tgc	5454
Ala Asn Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys	
1805 1810 1815	
ccc cag gcg gcg ctg gcc acc ttc aac atg ggt gcc tcc ttt att	5499
Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile	
1820 1825 1830	
gtc acc ggc acc gtg aac cag gtc gcc aag cag tcg ggc acg tgc	5544
Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys	
1835 1840 1845	
gac aat gtg cgc aag cag ctc gcg aag gcc act tac tcg gac gta	5589
Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val	
1850 1855 1860	
tgc atg gcc ccg gct gcc gac atg ttc gag gaa ggc gtc aag ctt	5634
Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu	
1865 1870 1875	

cag gtc ctc aag aag gga acc atg ttt ccc tcg cgc gcc aac aag	5679
Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys	
1880	1885
ctc tac gag ctc ttt tgc aag tac gac tcg ttc gag tcc atg ccc	5724
Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro	
1895	1900
ccc gca gag ctt gcg cgc gtc gag aag cgc atc ttc agc cgc gcg	5769
Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala	
1910	1915
ctc gaa gag gtc tgg gac gag acc aaa aac ttt tac att aac cgt	5814
Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg	
1925	1930
ctt cac aac ccg gag aag atc cag cgc gcc gag cgc gac ccc aag	5859
Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys	
1940	1945
ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc ctg gcg agc	5904
Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser	
1955	1960
cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac tac cag	5949
Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln	
1970	1975
gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc aag	5994
Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys	
1985	1990
gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc	6039
Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val	
2000	2005
gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc	6084
Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg	
2015	2020
cgt ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc	6129
Arg Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala	
2030	2035
gct ctt gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag	6174
Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys	
2045	2050
ctg	6177
Leu	

<210>4

<211>2059

<212>PRT

<213>Schizochytrium sp.

<220>

<221>misc_feature

<222>(370).. (370)

<223> 在 370 位的' Xaa' 代表 Leu

<220>

<221>misc_feature

<222>(371).. (371)

<223> 在 371 位的' Xaa' 代表 Ala 或 Val.

<400>4

```

Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys
1           5           10           15
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr
          20           25           30
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys
          35           40           45
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr
          50           55           60
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr
65           70           75           80
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu
          85           90           95
Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr
          100          105          110
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu
          115          120          125
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu
          130          135          140
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln
145          150          155          160
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala
          165          170          175
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val

```

180	185	190
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp		
195	200	205
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys		
210	215	220
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala		
225	230	235
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp		
245	250	255
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys		
260	265	270
Arg Leu Asp Asp Ala Ile Arg Asp Gly Asp His Ile Tyr Gly Thr Leu		
275	280	285
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro		
290	295	300
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile		
305	310	315
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly		
325	330	335
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe		
340	345	350
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His		
355	360	365
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser		
370	375	380
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr		
385	390	395
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu		
405	410	415
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly		
420	425	430
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala		
435	440	445
Ala Cys Thr Gly His Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly		
450	455	460
Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe		
465	470	475
Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly		
485	490	495

Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly		
500	505	510
Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His		
515	520	525
Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr		
530	535	540
Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val		
545	550	555
Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly		
565	570	575
Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg		
580	585	590
His Arg Ala Arg Val Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser		
595	600	605
Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser		
610	615	620
Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser		
625	630	635
Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn		
645	650	655
Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr		
660	665	670
Gly Glu Val Asp Gly Val Val Val Ala Gly Val Asp Leu Cys Gly Ser		
675	680	685
Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser		
690	695	700
Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val		
705	710	715
Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr		
725	730	735
Lys Asp Asp Arg Ile Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn		
740	745	750
Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys		
755	760	765
Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His		
770	775	780
Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu		
785	790	795
Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg		800

	805	810	815
Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr			
	820	825	830
Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn			
	835	840	845
Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu			
	850	855	860
Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp			
865	870	875	880
Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser			
	885	890	895
Glu Thr Arg Ser Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His			
	900	905	910
Tyr Glu Arg Glu Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu			
	915	920	925
Ile Val Leu Arg Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp			
	930	935	940
Lys Ile Arg Glu Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu			
945	950	955	960
Ser Glu Leu Lys Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly			
	965	970	975
Glu Thr Leu Ala Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu			
	980	985	990
Ala Leu Ser Leu Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu			
	995	1000	1005
Leu Ala Ala Lys Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp			
	1010	1015	1020
Trp Ser Ser Pro Ala Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala			
	1025	1030	1035
Ser Asp Arg Val Ala Phe Met Tyr Gly Glu Gly Arg Ser Pro Tyr			
	1040	1045	1050
Tyr Gly Ile Thr Gln Asp Ile His Arg Ile Trp Pro Glu Leu His			
	1055	1060	1065
Glu Val Ile Asn Glu Lys Thr Asn Arg Leu Trp Ala Glu Gly Asp			
	1070	1075	1080
Arg Trp Val Met Pro Arg Ala Ser Phe Lys Ser Glu Leu Glu Ser			
	1085	1090	1095
Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu Met Phe Arg Leu			
	1100	1105	1110

Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala Arg Asp Val		
1115	1120	1125
Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu Gly Glu		
1130	1135	1140
Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile Ser		
1145	1150	1155
Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys		
1160	1165	1170
Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile		
1175	1180	1185
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val		
1190	1195	1200
Arg Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser		
1205	1210	1215
Lys Tyr Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu		
1220	1225	1230
Ile Ser Gly Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu		
1235	1240	1245
Gly Gly Asn Ile Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly		
1250	1255	1260
His Cys Pro Glu Val Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile		
1265	1270	1275
His Ala Asn Leu Glu Phe Pro Val Val Asp Gly Leu Asp Leu Trp		
1280	1285	1290
Thr Thr Ile Asn Gln Lys Arg Leu Val Pro Arg Ala Thr Gly Ala		
1295	1300	1305
Lys Asp Glu Trp Ala Pro Ser Ser Phe Gly Glu Tyr Ala Gly Gln		
1310	1315	1320
Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln Ile Val Glu Thr Ile		
1325	1330	1335
Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val Gly Pro Asn Asn		
1340	1345	1350
His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro Gln Arg Asn		
1355	1360	1365
His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala Trp Thr		
1370	1375	1380
Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val Pro		
1385	1390	1395
Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu		

1400	1405	1410
Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys		
1415	1420	1425
Lys Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn		
1430	1435	1440
Ser Lys Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro		
1445	1450	1455
Pro Ala Asp Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met		
1460	1465	1470
Lys Pro Val Ala Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu		
1475	1480	1485
Gln Asp Glu Thr Arg Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro		
1490	1495	1500
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser		
1505	1510	1515
Pro Ser Pro Ala Pro Ser Ala Pro Val Gln Lys Lys Ala Ala Pro		
1520	1525	1530
Ala Ala Glu Thr Lys Ala Val Ala Ser Ala Asp Ala Leu Arg Ser		
1535	1540	1545
Ala Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser		
1550	1555	1560
Ala Ser Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val		
1565	1570	1575
Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe		
1580	1585	1590
Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met		
1595	1600	1605
Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Arg		
1610	1615	1620
Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met Gln		
1625	1630	1635
Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn		
1640	1645	1650
Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn		
1655	1660	1665
Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr		
1670	1675	1680
Phe Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val		
1685	1690	1695

Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val		
1700	1705	1710
Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu		
1715	1720	1725
Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu Gln Lys		
1730	1735	1740
Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu Ala		
1745	1750	1755
Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser		
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu		
1775	1780	1785
Ile Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro		
1790	1795	1800
Ala Asn Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys		
1805	1810	1815
Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile		
1820	1825	1830
Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys		
1835	1840	1845
Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser Asp Val		
1850	1855	1860
Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu		
1865	1870	1875
Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys		
1880	1885	1890
Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro		
1895	1900	1905
Pro Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala		
1910	1915	1920
Leu Glu Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg		
1925	1930	1935
Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys		
1940	1945	1950
Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser		
1955	1960	1965
Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp Tyr Gln		
1970	1975	1980
Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile Lys		
1985	1990	1995

Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val
 2000 2005 2010
 Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg
 2015 2020 2025
 Arg Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala
 2030 2035 2040
 Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys
 2045 2050 2055
 Leu

<210>5

<211>4509

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (4509)

<223>

<400>5

atg gcg ctc cgt gtc aag acg aac aag aag cca tgc tgg gag atg acc	48
Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr	
1 5 10 15	
aag gag gag ctg acc agc ggc aag acc gag gtg ttc aac tat gag gaa	96
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu	
20 25 30	
ctc ctc gag ttc gca gag ggc gac atc gcc aag gtc ttc gga ccc gag	144
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu	
35 40 45	
ttc gcc gtc atc gac aag tac ccg cgc cgc gtg cgc ctg ccc gcc cgc	192
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg	
50 55 60	
gag tac ctg ctc gtg acc cgc gtc acc ctc atg gac gcc gag gtc aac	240
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn	
65 70 75 80	
aac tac cgc gtc ggc gcc cgc atg gtc acc gag tac gat ctc ccc gtc	288
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val	
85 90 95	

aac gga gag ctc tcc gag ggc gga gac tgc ccc tgg gcc gtc ctg gtc	336
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val	
100 105 110	
gag agt ggc cag tgc gat ctc atg ctc atc tcc tac atg ggc att gac	384
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp	
115 120 125	
ttc cag aac cag ggc gac cgc gtc tac cgc ctg ctc aac acc acg ctc	432
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu	
130 135 140	
acc ttt tac ggc gtg gcc cac gag ggc gag acc ctc gag tac gac att	480
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile	
145 150 155 160	
cgc gtc acc ggc ttc gcc aag cgt ctc gac ggc ggc atc tcc atg ttc	528
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe	
165 170 175	
ttc ttc gag tac gac tgc tac gtc aac ggc cgc ctc ctc atc gag atg	576
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met	
180 185 190	
cgc gat ggc tgc gcc ggc rtc ttc acc aac gag gag crc gac gcc ggc	624
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly	
195 200 205	
aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc	672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile	
210 215 220	
cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag	720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys	
225 230 235 240	
acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg	768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp	
245 250 255	
gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa	816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys	
260 265 270	
ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac	864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp	
275 280 285	
cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc	912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile	
290 295 300	

ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag	960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln	
305 310 315 320	
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag	1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys	
325 330 335	
atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac	1056
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp	
340 345 350	
ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa	1104
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln	
355 360 365	
atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag	1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu	
370 375 380	
atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac	1200
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn	
385 390 395 400	
atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc	1248
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg	
405 410 415	
atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac	1296
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp	
420 425 430	
ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac	1344
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn	
435 440 445	
ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc	1392
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly	
450 455 460	
ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc	1440
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala	
465 470 475 480	
gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc	1488
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro	
485 490 495	
gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag	1536
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys	
500 505 510	

gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc	1584
Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro	
515 520 525	
tct ttt gcg ccc teg gcc tac aag ccg cgc aac atc gcc ttt acg ccc	1632
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro	
530 535 540	
ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg	1680
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro	
545 550 555 560	
ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg	1728
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met	
565 570 575	
tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc	1776
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg	
580 585 590	
agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct	1824
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser	
595 600 605	
gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag	1872
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys	
610 615 620	
ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac	1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr	
625 630 635 640	
aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag	1968
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu	
645 650 655	
atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc	2016
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro	
660 665 670	
ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac	2064
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn	
675 680 685	
gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc	2112
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg	
690 695 700	
aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc	2160
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val	
705 710 715 720	

cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag	2208
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys	
725 730 735	
ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag	2256
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln	
740 745 750	
gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac	2304
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn	
755 760 765	
aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc	2352
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly	
770 775 780	
agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc	2400
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu	
785 790 795 800	
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc	2448
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val	
805 810 815	
tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg	2496
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr	
820 825 830	
gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg	2544
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser	
835 840 845	
gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag	2592
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu	
850 855 860	
gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa	2640
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln	
865 870 875 880	
ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc	2688
Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg	
885 890 895	
ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc	2736
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile	
900 905 910	
gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc	2784
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly	
915 920 925	

ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc	2832
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg	
930 935 940	
gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc	2880
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala	
945 950 955 960	
tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc	2928
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro	
965 970 975	
gct gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag	2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys	
980 985 990	
acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg	3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro	
995 1000 1005	
acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag	3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln	
1010 1015 1020	
gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc	3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg	
1025 1030 1035	
tcc ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc	3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly	
1040 1045 1050	
gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc	3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala	
1055 1060 1065	
ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc	3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro	
1070 1075 1080	
atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg	3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu	
1085 1090 1095	
cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac	3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp	
1100 1105 1110	
agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc	3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly	
1115 1120 1125	

gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag	3429
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln	
1130 1135 1140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt	3474
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly	
1145 1150 1155	
tcg gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc	3519
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr	
1160 1165 1170	
gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc	3564
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu	
1175 1180 1185	
gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag	3609
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu	
1190 1195 1200	
ctc gcg cgc cgc gtt ccc gtc gcc gac gat atc gct gtc gag gct	3654
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala	
1205 1210 1215	
gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc	3699
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu	
1220 1225 1230	
ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc	3744
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly	
1235 1240 1245	
tac ccc gcg cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc	3789
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val	
1250 1255 1260	
ggc tgc ccg cag gcc gcc gcc gcc gcg ctc acc atg ggc gcc gcc	3834
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala	
1265 1270 1275	
ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc	3879
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly	
1280 1285 1290	
acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg	3924
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser	
1295 1300 1305	
gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc	3969
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val	
1310 1315 1320	

aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc	4014
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala	
1325	1330 1335
aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc	4059
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser	
1340	1345 1350
atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag	4104
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys	
1355	1360 1365
cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att	4149
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile	
1370	1375 1380
aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac	4194
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp	
1385	1390 1395
ccc aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt	4239
Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu	
1400	1405 1410
gcc agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac	4284
Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp	
1415	1420 1425
tac cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc	4329
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe	
1430	1435 1440
atc aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc	4374
Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro	
1445	1450 1455
tgt gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac	4419
Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr	
1460	1465 1470
ctg cgc cgt ctc aac gcc ctg cgc agc gac ccg cgc att gac ctc	4464
Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu	
1475	1480 1485
gag acc gag gat gct gcc ttt gtc tac gag ccc acc aac gcg ctc	4509
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu	
1490	1495 1500

<210>6

<211>1503

<212>PRT

<213>Schizochytrium sp.

<400>6

```

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr
1           5           10           15
Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu
          20           25           30
Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu
          35           40           45
Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg
          50           55           60
Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn
65           70           75           80
Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val
          85           90           95
Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val
          100          105          110
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp
          115          120          125
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu
          130          135          140
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile
145          150          155          160
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe
          165          170          175
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met
          180          185          190
ArgAsp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly
          195          200          205
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile
          210          215          220
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys
225          230          235          240
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp
          245          250          255
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys
          260          265          270
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp

```

275	280	285
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile		
290	295	300
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln		
305	310	315
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys		
325	330	335
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp		
340	345	350
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln		
355	360	365
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu		
370	375	380
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn		
385	390	395
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg		
405	410	415
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp		
420	425	430
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn		
435	440	445
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly		
450	455	460
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala		
465	470	475
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro		
485	490	495
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys		
500	505	510
Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro		
515	520	525
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro		
530	535	540
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro		
545	550	555
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met		
565	570	575
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg		
580	585	590

Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser		
595	600	605
Asp Leu Lys His Val Ash Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys		
610	615	620
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr		
625	630	635
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu		
	645	650
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro		
	660	665
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn		
	675	680
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg		
	690	695
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val		
705	710	715
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys		
	725	730
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln		
	740	745
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn		
	755	760
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly		
	770	775
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu		
785	790	795
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val		
	805	810
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr		
	820	825
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser		
	835	840
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu		
	850	855
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln		
865	870	875
Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg		
	885	890
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile		

900	905	910
Val Ser Val Asp Ala His Asp Gly	Val Val Asp Leu Val Ala Asp Gly	
915	920	925
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg		
930	935	940
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala		
945	950	955
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro		
965	970	975
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys		
980	985	990
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro		
995	1000	1005
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln		
1010	1015	1020
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg		
1025	1030	1035
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly		
1040	1045	1050
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala		
1055	1060	1065
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro		
1070	1075	1080
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu		
1085	1090	1095
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp		
1100	1105	1110
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly		
1115	1120	1125
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln		
1130	1135	1140
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly		
1145	1150	1155
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr		
1160	1165	1170
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu		
1175	1180	1185
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu		
1190	1195	1200

Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala		
1205	1210	1215
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu		
1220	1225	1230
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly		
1235	1240	1245
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val		
1250	1255	1260
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala		
1265	1270	1275
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly		
1280	1285	1290
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser		
1295	1300	1305
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val		
1310	1315	1320
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala		
1325	1330	1335
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser		
1340	1345	1350
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys		
1355	1360	1365
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile		
1370	1375	1380
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp		
1385	1390	1395
Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu		
1400	1405	1410
Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp		
1415	1420	1425
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe		
1430	1435	1440
Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro		
1445	1450	1455
Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr		
1460	1465	1470
Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu		
1475	1480	1485
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu		

1490	1495	1500	
<210>7			
<211>600			
<212>DNA			
<213>Schizochytrium sp.			
<220>			
<221>CDS			
<222>(1).. (600)			
<223>			
<400>7			
atg gcg gcc cgt ctg cag gag caa aag gga ggc gag atg gat acc cgc			48
Met Ala Ala Arg Leu Gln Glu Gln Lys Gly Gly Glu Met Asp Thr Arg			
1 5 10 15			
att gcc atc atc ggc atg tcg gcc atc ctc ccc tgc ggc acg acc gtg			96
Ile Ala Ile Ile Gly Met Ser Ala Ile Leu Pro Cys Gly Thr Thr Val			
20 25 30			
cgc gag tcg tgg gag acc atc cgc gcc ggc atc gac tgc ctg tcg gat			144
Arg Glu Ser Trp Glu Thr Ile Arg Ala Gly Ile Asp Cys Leu Ser Asp			
35 40 45			
ctc ccc gag gac cgc gtc gac gtg acg gcg tac ttt gac ccc gtc aag			192
Leu Pro Glu Asp Arg Val Asp Val Thr Ala Tyr Phe Asp Pro Val Lys			
50 55 60			
acc acc aag gac aag atc tac tgc aag cgc ggt ggc ttc att ccc gag			240
Thr Thr Lys Asp Lys Ile Tyr Cys Lys Arg Gly Gly Phe Ile Pro Glu			
65 70 75 80			
tac gac ttt gac gcc cgc gag ttc gga ctc aac atg ttc cag atg gag			288
Tyr Asp Phe Asp Ala Arg Glu Phe Gly Leu Asn Met Phe Gln Met Glu			
85 90 95			
gac tcg gac gca aac cag acc atc tcg ctt ctc aag gtc aag gag gcc			336
Asp Ser Asp Ala Asn Gln Thr Ile Ser Leu Leu Lys Val Lys Glu Ala			
100 105 110			
ctc cag gac gcc ggc atc gac gcc ctc ggc aag gaa aag aag aac atc			384
Leu Gln Asp Ala Gly Ile Asp Ala Leu Gly Lys Glu Lys Lys Asn Ile			
115 120 125			
ggc tgc gtg crc ggc att ggc ggc ggc caa aag tcc agc cac gag ttc			432
Gly Cys Val Leu Gly Ile Gly Gly Gly Gln Lys Ser Ser His Glu Phe			

130	135	140	
tac tcg cgc ctt aat tat gtt gtc gtg gag aag gtc ctc cgc aag atg			480
Tyr Ser Arg Leu Asn Tyr Val Val Val Glu Lys Val Leu Arg Lys Met			
145	150	155	160
ggc atg ccc gag gag gac gtc aag gtc gcc gtc gaa aag tac aag gcc			528
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala			
165	170	175	
aac ttc ccc gag tgg cgc ctc gac tcc ttc cct ggc ttc ctc ggc aac			576
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn			
180	185	190	
gtc acc gcc ggt cgc tgc acc aac			600
Val Thr Ala Gly Arg Cys Thr Asn			
195	200		

<210>8

<211>200

<212>PRT

<213>Schizochytrium sp.

<400>8

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

145	150	155	160
Gly Met Pro Glu Glu Asp Val Lys Val Ala Val Glu Lys Tyr Lys Ala			
	165	170	175
Asn Phe Pro Glu Trp Arg Leu Asp Ser Phe Pro Gly Phe Leu Gly Asn			
	180	185	190
Val Thr Ala Gly Arg Cys Thr Asn			
	195	200	

<210>9

<211>1278

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (1278)

<223>

<400>9

gat gtc acc aag gag gcc tgg cgc ctc ccc cgc gag ggc gtc agc ttc	48
Asp Val Thr Lys Glu Ala Trp Arg Leu Pro Arg Glu Gly Val Ser Phe	
1 5 10 15	
cgc gcc aag ggc atc gcc acc aac ggc gct gtc gcc gcg ctc ttc tcc	96
Arg Ala Lys Gly Ile Ala Thr Asn Gly Ala Val Ala Ala Leu Phe Ser	
20 25 30	
ggc cag ggc gcg cag tac acg cac atg ttt agc gag gtg gcc atg aac	144
Gly Gln Gly Ala Gln Tyr Thr His Met Phe Ser Glu Val Ala Met Asn	
35 40 45	
tgg ccc cag ttc cgc cag agc att gcc gcc atg gac gcc gcc cag tcc	192
Trp Pro Gln Phe Arg Gln Ser Ile Ala Ala Met Asp Ala Ala Gln Ser	
50 55 60	
aag gtc gct gga agc gac aag gac ttt gag cgc gtc tcc cag gtc ctc	240
Lys Val Ala Gly Ser Asp Lys Asp Phe Glu Arg Val Ser Gln Val Leu	
65 70 75 80	
tac ccg cgc aag ccg tac gag cgt gag ccc gag cag aac ccc aag aag	288
Tyr Pro Arg Lys Pro Tyr Glu Arg Glu Pro Glu Gln Asn Pro Lys Lys	
85 90 95	
atc tcc ctc acc gcc tac tcg cag ccc tcg acc ctg gcc tgc gct ctc	336
Ile Ser Leu Thr Ala Tyr Ser Gln Pro Ser Thr Leu Ala Cys Ala Leu	

100	105	110	
ggt gcc ttt gag atc ttc aag gag gcc ggc ttc acc ccg gac ttt gcc			384
Gly Ala Phe Glu Ile Phe Lys Glu Ala Gly Phe Thr Pro Asp Phe Ala			
115	120	125	
gcc ggc cat tcg ctc ggt gag ttc gcc gcc ctc tac gcc gcg ggc tgc			432
Ala Gly His Ser Leu Gly Glu Phe Ala Ala Leu Tyr Ala Ala Gly Cys			
130	135	140	
gtc gac cgc gac gag ctc ttt gag ctt gtc tgc cgc cgc gcc cgc atc			480
Val Asp Arg Asp Glu Leu Phe Glu Leu Val Cys Arg Arg Ala Arg Ile			
145	150	155	160
atg ggc ggc aag gac gca ccg gcc acc ccc aag gga tgc atg gcc gcc			528
Met Gly Gly Lys Asp Ala Pro Ala Thr Pro Lys Gly Cys Met Ala Ala			
165	170	175	
gtc att ggc ccc aac gcc gag aac atc aag gtc cag gcc gcc aac gtc			576
Val Ile Gly Pro Asn Ala Glu Asn Ile Lys Val Gln Ala Ala Asn Val			
180	185	190	
tgg ctc ggc aac tcc aac tcg cct tcg cag acc gtc atc acc ggc tcc			624
Trp Leu Gly Asn Ser Asn Ser Pro Ser Gln Thr Val Ile Thr Gly Ser			
195	200	205	
gtc gaa ggt atc cag gcc gag agc gcc cgc ctc cag aag gag ggc ttc			672
Val Glu Gly Ile Gln Ala Glu Ser Ala Arg Leu Gln Lys Glu Gly Phe			
210	215	220	
cgc gtc gtg cct ctt gcc tgc gag agc gcc ttc cac tcg ccc cag atg			720
Arg Val Val Pro Leu Ala Cys Glu Ser Ala Phe His Ser Pro Gln Met			
225	230	235	240
gag aac gcc tcg tcg gcc ttc aag gac gtc atc tcc aag gtc tcc ttc			768
Glu Asn Ala Ser Ser Ala Phe Lys Asp Val Ile Ser Lys Val Ser Phe			
245	250	255	
cgc acc ccc aag gcc gag acc aag ctc ttc agc aac gtc tct ggc gag			816
Arg Thr Pro Lys Ala Glu Thr Lys Leu Phe Ser Asn Val Ser Gly Glu			
260	265	270	
acc tac ccc acg gac gcc cgc gag atg ctt acg cag cac atg acc agc			864
Thr Tyr Pro Thr Asp Ala Arg Glu Met Leu Thr Gln His Met Thr Ser			
275	280	285	
agc gtc aag ttc ctc acc cag gtc cgc aac atg cac cag gcc ggt gcg			912
Ser Val Lys Phe Leu Thr Gln Val Arg Asn Met His Gln Ala Gly Ala			
290	295	300	
cgc atc ttt gtc gag ttc gga ccc aag cag gtg ctc tcc aag ctt gtc			960
Arg Ile Phe Val Glu Phe Gly Pro Lys Gln Val Leu Ser Lys Leu Val			

305	310	315	320	
tcc gag acc ctc aag gat gac ccc tcg gtt gtc acc gtc tct gtc aac				1008
Ser Glu Thr Leu Lys Asp Asp Pro Ser Val Val Thr Val Ser Val Asn				
	325	330	335	
ccg gcc tcg ggc acg gat tcg gac atc cag ctc cgc gac gcg gcc gtc				1056
Pro Ala Ser Gly Thr Asp Ser Asp Ile Gln Leu Arg Asp Ala Ala Val				
	340	345	350	
cag ctc gtt gtc gct ggc gtc aac ctt cag ggc ttt gac aag tgg gac				1104
Gln Leu Val Val Ala Gly Val Asn Leu Gln Gly Phe Asp Lys Trp Asp				
	355	360	365	
gcc ccc gat gcc acc cgc atg cag gcc atc aag aag aag cgc act acc				1152
Ala Pro Asp Ala Thr Arg Met Gln Ala Ile Lys Lys Lys Arg Thr Thr				
	370	375	380	
ctc cgc ctt tcg gcc gcc acc tac gtc tcg gac aag acc aag aag gtc				1200
Leu Arg Leu Ser Ala Ala Thr Tyr Val Ser Asp Lys Thr Lys Lys Val				
385	390	395	400	
cgc gac gcc gcc atg aac gat ggc cgc tgc gtc acc tac ctc aag ggc				1248
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly				
	405	410	415	
gcc gca ccg ctc atc aag gcc ccg gag ccc				1278
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro				
	420	425		

<210>10

<211>426

<212>PRT

<213>Schizochytrium sp.

<400>10

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

Tyr	Pro	Arg	Lys	Pro	Tyr	Glu	Arg	Glu	Pro	Glu	Gln	Asn	Pro	Lys	Lys
				85					90					95	
Ile	Ser	Leu	Thr	Ala	Tyr	Ser	Gln	Pro	Ser	Thr	Leu	Ala	Cys	Ala	Leu
				100					105					110	
Gly	Ala	Phe	Glu	Ile	Phe	Lys	Glu	Ala	Gly	Phe	Thr	Pro	Asp	Phe	Ala
				115					120					125	
Ala	Gly	His	Ser	Leu	Gly	Glu	Phe	Ala	Ala	Leu	Tyr	Ala	Ala	Gly	Cys
				130					135					140	
Val	Asp	Arg	Asp	Glu	Leu	Phe	Glu	Leu	Val	Cys	Arg	Arg	Ala	Arg	Ile
				145					150					155	
Met	Gly	Gly	Lys	Asp	Ala	Pro	Ala	Thr	Pro	Lys	Gly	Cys	Met	Ala	Ala
				165					170					175	
Val	Ile	Gly	Pro	Asn	Ala	Glu	Asn	Ile	Lys	Val	Gln	Ala	Ala	Asn	Val
				180					185					190	
Trp	Leu	Gly	Asn	Ser	Asn	Ser	Pro	Ser	Gln	Thr	Val	Ile	Thr	Gly	Ser
				195					200					205	
Val	Glu	Gly	Ile	Gln	Ala	Glu	Ser	Ala	Arg	Leu	Gln	Lys	Glu	Gly	Phe
				210					215					220	
Arg	Val	Val	Pro	Leu	Ala	Cys	Glu	Ser	Ala	Phe	His	Ser	Pro	Gln	Met
				225					230					235	
Glu	Asn	Ala	Ser	Ser	Ala	Phe	Lys	Asp	Val	Ile	Ser	Lys	Val	Ser	Phe
				245					250					255	
Arg	Thr	Pro	Lys	Ala	Glu	Thr	Lys	Leu	Phe	Ser	Asn	Val	Ser	Gly	Glu
				260					265					270	
Thr	Tyr	Pro	Thr	Asp	Ala	Arg	Glu	Met	Leu	Thr	Gln	His	Met	Thr	Ser
				275					280					285	
Ser	Val	Lys	Phe	Leu	Thr	Gln	Val	Arg	Asn	Met	His	Gln	Ala	Gly	Ala
				290					295					300	
Arg	Ile	Phe	Val	Glu	Phe	Gly	Pro	Lys	Gln	Val	Leu	Ser	Lys	Leu	Val
				305					310					315	
Ser	Glu	Thr	Leu	Lys	Asp	Asp	Pro	Ser	Val	Val	Thr	Val	Ser	Val	Asn
				325					330					335	
Pro	Ala	Ser	Gly	Thr	Asp	Ser	Asp	Ile	Gln	Leu	Arg	Asp	Ala	Ala	Val
				340					345					350	
Gln	Leu	Val	Val	Ala	Gly	Val	Asn	Leu	Gln	Gly	Phe	Asp	Lys	Trp	Asp
				355					360					365	
Ala	Pro	Asp	Ala	Thr	Arg	Met	Gln	Ala	Ile	Lys	Lys	Lys	Arg	Thr	Thr
				370					375					380	
Leu	Arg	Leu	Ser	Ala	Ala	Thr	Tyr	Val	Ser	Asp	Lys	Thr	Lys	Lys	Val

385 390 395 400
Arg Asp Ala Ala Met Asn Asp Gly Arg Cys Val Thr Tyr Leu Lys Gly
405 410 415
Ala Ala Pro Leu Ile Lys Ala Pro Glu Pro
420 425

<210>11

<211>5

<212>PRT

<213>Schizochytrium sp.

<220>

<221>MISC_FEATURE

<222>(4)..(4)

<223>x =任意氨基酸

<400>11

Gly His Ser Xaa Gly

1 5

<210>12

<211>258

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1)..(258)

<223>

<400>12

gct gtc tcg aac gag ctt ctt gag aag gcc gag act gtc gtc atg gag 48

Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu

1 5 10 15

gtc ctc gcc gcc aag acc ggc tac gag acc gac atg atc gag gct gac 96

Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu Ala Asp

20 25 30

atg gag ctc gag acc gag ctc ggc att gac tcc atc aag cgt gtc gag 144

Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu

35	40	45	
atc ctc tcc gag gtc cag gcc atg ctc aat gtc gag gcc aag gat gtc			192
Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val			
50	55	60	
gat gcc ctc agc cgc act cgc act gtt ggt gag gtt gtc aac gcc atg			240
Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asn Ala Met			
65	70	75	80
aag gcc gag atc gct ggc			258
Lys Ala Glu Ile Ala Gly			
85			

<210>13

<211>86

<212>PRT

<213>Schizochytrium sp.

<400>13

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

<210>14

<211>5

<212>PRT

<213>Schizochytrium sp.

<400>14

Leu Gly Ile Asp Ser	
1	5

<210>15

<211>21

<212>PRT

<213>Schizochytrium sp.

<400>15

Ala Pro Ala Pro Val Lys Ala Ala Ala Pro Ala Ala Pro Val Ala Ser

1

5

10

15

Ala Pro Ala Pro Ala

20

<210>16

<211>3006

<212>DNA

<213>Schizochytrium sp.

<400>16

```

gccccgccc cggtaaggc tgctgcgcct gccgcccccg ttgcctcggc ccctgccccg      60
gctgtctcga acgagcttct tgagaaggcc gagactgtcg tcatggaggt cctcgccgcc      120
aagaccggct acgagaccga catgatcgag gctgacatgg agctcgagac cgagctcggc      180
attgactcca tcaagcgtgt cgagatcctc tccgagggtcc aggccatgct caatgtcgag      240
gccaaaggatg tcgatgccct cagccgcact cgcactgttg gtgaggttgt caacgccatg      300
aaggccgaga tcgttggcag ctctgccccg gcgcctgtcg ccgctgctcc ggctccggcc      360
aaggctgccc ctgccgcgc tgccctgtct gtctcgaacg agcttctcga gaaggccgag      420
accgtcgtca tggaggctct cgccgccaag actggctacg agactgacat gatcgagtcc      480
gacatggagc tcgagactga gctcggcatt gactccatca agcgtgtcga gatcctctcc      540
gaggttcagg ccatgtctaa cgtcgaggcc aaggacgtcg acgctctcag ccgcactcgc      600
actgtgggtg aggtcgtcaa cgccatgaag gctgagatcg ctgggtggctc tgccccggcg      660
cctgccgccc ctgccccagg tccggctgct gccgccccctg cgcctgccgc cgccgcccct      720
gctgtctcga acgagcttct tgagaaggcc gagaccgtcg tcatggaggt cctcgccgcc      780
aagactggct acgagactga catgatcgag tccgacatgg agctcgagac cgagctcggc      840
attgactcca tcaagcgtgt cgagattctc tccgagggtcc aggccatgct caacgtcgag      900
gccaaaggacg tcgacgtctc cagccgcacc cgcactgttg gcgaggctgt cgatgccatg      960
aaggccgaga tcgttgggtg ctctgccccg gcgcctgccg ccgctgctcc tgctccggct     1020
gctgccgccc ctgcgcctgc cgccccctgcg cctgctgtct cgagcgagct tctcgagaag     1080
gccgagactg tcgtcatgga ggctcctgcc gccaaagactg gctacgagac tgacatgate     1140
gagtccgaca tggagctcga gaccgagctc ggcatgact ccatcaagcg tgtcgagatt     1200
ctctccgagg tccaggccat gctcaacgtc gaggccaagg acgtcgacgc tctcagccgc     1260
accgcactg ttggcgaggt cgtcgatgcc atgaaggccg agatcgctgg tggtctctgcc     1320

```

ccggcgccctg	ccgccgctgc	tcctgtctccg	gctgtctgccg	cccctgcgcc	tgccgcccct	1380
gcgcctgccg	cccctgcgcc	tgtgtgtctcg	agcgagcttc	tcgagaaggc	cgagactgtc	1440
gtcatggagg	tcctcgccgc	caagactggc	tacgagactg	acatgattga	gtccgacatg	1500
gagctcgaga	ccgagctcgg	cattgactcc	atcaagcgtg	tcgagattct	ctccgaggtt	1560
caggccatgc	tcaacgtcga	ggccaaggac	gtcgacgctc	tcagccgcac	tcgcaactgtt	1620
ggtgaggtcg	tcgatgccat	gaaggctgag	atcgctggca	gctccgcctc	ggcgccctgcc	1680
gccgtgtctc	ctgtctccggc	tgtgtccgct	cctgcgcccg	ctgccgccgc	ccctgtctgtc	1740
tcgaacgagc	ttctcgagaa	agccgagact	gtcgtcatgg	aggtcctcgc	cgccaagact	1800
ggctacgaga	ctgacatgat	cgagtccgac	atggagctcg	agactgagct	cggcattgac	1860
tccatcaagc	gtgtcgagat	cctctccgag	gttcaggcca	tgtctaacgt	cgaggccaag	1920
gacgtcgatg	ccctcagccg	cacccgcact	gttggcgagg	ttgtcgatgc	catgaaggcc	1980
gagatcgctg	gtggctctgc	cccggcgcc	gccgccgctg	cccctgtctc	ggctgccgcc	2040
gcccctgtctg	tctcgaacga	gcttctcgag	aaggccgaga	ctgtcgtcat	ggaggtcctc	2100
gccccaaga	ctggctacga	gaccgacatg	atcgagtccg	acatggagct	cgagaccgag	2160
ctcggcattg	actccatcaa	gcgtgtcgag	attctctccg	aggttcaggc	catgctcaac	2220
gtcgaggcca	aggacgtcga	tgtctcagc	cgcactcgca	ctgttggcga	ggtcgtcgat	2280
gccatgaagg	ctgagatcgc	cggcagctcc	gccccggcgc	ctgccgccgc	tgtcctgtct	2340
ccggctgtctg	ccgtctctgc	gcccgtctgc	gctgcccctg	ctgtctcgag	cgagcttctc	2400
gagaaggccg	agaccgtcgt	catggaggtc	ctcgccgcca	agactggcta	cgagactgac	2460
atgattgagt	ccgacatgga	gtctcgagact	gagctcggca	ttgactccat	caagcgtgtc	2520
gagatcctct	ccgaggttca	ggccatgtct	aacgtcgagg	ccaaggacgt	cgatgccctc	2580
agccgcaccc	gcactgtttg	cgaggttgtc	gatgccatga	aggccgagat	cgctggtggc	2640
tctgccccgg	cgcctgcgc	cgtgtcccct	gtcccggtg	ccgccgcccc	tgtgtctctg	2700
aacgagcttc	ttgagaaggc	cgagaccgtc	gtcatggagg	tcctcgccgc	caagactggc	2760
tacgagaccg	acatgatcga	gtccgacatg	gagctcgaga	ccgagctcgg	cattgactcc	2820
atcaagcgtg	tcgagattct	ctccgagggt	caggccatgc	tcaacgtcga	ggccaaggac	2880
gtcgacgctc	tcagccgcac	tcgcaactgt	ggcgaggctg	tcgatgccat	gaaggctgag	2940
atcgctgggtg	gctctgcccc	ggcgccctgcc	gccgctgtct	ctgcctcggc	tggcgccgcg	3000
cctgcg						3006

<210>17

<211>2133

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (2133)

<223>

<400>17

ttt ggc gct ctc ggc ggc ttc atc tcg cag cag gcg gag cgc ttc gag	48
Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg Phe Glu	
1 5 10 15	
ccc gcc gaa atc ctc ggc ttc acg ctc atg tgc gcc aag ttc gcc aag	96
Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe Ala Lys	
20 25 30	
gct tcc ctc tgc acg gct gtg gct ggc ggc cgc ccg gcc ttt atc ggt	144
Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe Ile Gly	
35 40 45	
gtg gcg cgc ctt gac ggc cgc ctc gga ttc act tcg cag ggc act tct	192
Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly Thr Ser	
50 55 60	
gac gcg ctc aag cgt gcc cag cgt ggt gcc atc ttt ggc ctc tgc aag	240
Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu Cys Lys	
65 70 75 80	
acc atc ggc ctc gag tgg tcc gag tct gac gtc ttt tcc cgc ggc gtg	288
Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg Gly Val	
85 90 95	
gac att gct cag ggc atg cac ccc gag gat gcc gcc gtg gcg att gtg	336
Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val	
100 105 110	
cgc gag atg gcg tgc gct gac att cgc att cgc gag gtc ggc att ggc	384
Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly	
115 120 125	
gca aac cag cag cgc tgc acg atc cgt gcc gcc aag ctc gag acc ggc	432
Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly	
130 135 140	
aac ccg cag cgc cag atc gcc aag gac gac gtg ctg ctc gtt tct ggc	480
Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly	
145 150 155 160	
ggc gct cgc ggc atc acg cct ctt tgc atc cgg gag atc acg cgc cag	528
Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln	
165 170 175	
atc gcg ggc ggc aag tac att ctg ctt ggc cgc agc aag gtc tct gcg	576
Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala	
180 185 190	
agc gaa ccg gca tgg tgc gct ggc atc act gac gag aag gct gtg caa	624

Ser	Glu	Pro	Ala	Trp	Cys	Ala	Gly	Ile	Thr	Asp	Glu	Lys	Ala	Val	Gln	
195				200				205								
aag	gct	gct	acc	cag	gag	ctc	aag	cgc	gcc	ttt	agc	gct	ggc	gag	ggc	672
Lys	Ala	Ala	Thr	Gln	Glu	Leu	Lys	Arg	Ala	Phe	Ser	Ala	Gly	Glu	Gly	
210				215				220								
ccc	aag	ccc	acg	ccc	cgc	gct	gtc	act	aag	ctt	gtg	ggc	tct	gtt	ctt	720
Pro	Lys	Pro	Thr	Pro	Arg	Ala	Val	Thr	Lys	Leu	Val	Gly	Ser	Val	Leu	
225				230				235				240				
ggc	gct	cgc	gag	gtg	cgc	agc	tct	att	gct	gcg	att	gaa	gcg	ctc	ggc	768
Gly	Ala	Arg	Glu	Val	Arg	Ser	Ser	Ile	Ala	Ala	Ile	Glu	Ala	Leu	Gly	
245				250				255								
ggc	aag	gcc	atc	tac	tcg	tcg	tgc	gac	gtg	aac	tct	gcc	gcc	gac	gtg	816
Gly	Lys	Ala	Ile	Tyr	Ser	Ser	Cys	Asp	Val	Asn	Ser	Ala	Ala	Asp	Val	
260				265				270								
gcc	aag	gcc	gtg	cgc	gat	gcc	gag	tcc	cag	ctc	ggt	gcc	cgc	gtc	tcg	864
Ala	Lys	Ala	Val	Arg	Asp	Ala	Glu	Ser	Gln	Leu	Gly	Ala	Arg	Val	Ser	
275				280				285								
ggc	atc	gtt	cat	gcc	tcg	ggc	gtg	ctc	cgc	gac	cgt	ctc	atc	gag	aag	912
Gly	Ile	Val	His	Ala	Ser	Gly	Val	Leu	Arg	Asp	Arg	Leu	Ile	Glu	Lys	
290				295				300								
aag	ctc	ccc	gac	gag	ttc	gac	gcc	gtc	ttt	ggc	acc	aag	gtc	acc	ggt	960
Lys	Leu	Pro	Asp	Glu	Phe	Asp	Ala	Val	Phe	Gly	Thr	Lys	Val	Thr	Gly	
305				310				315				320				
ctc	gag	aac	ctc	ctc	gcc	gcc	gtc	gac	cgc	gcc	aac	ctc	aag	cac	atg	1008
Leu	Glu	Asn	Leu	Leu	Ala	Ala	Val	Asp	Arg	Ala	Asn	Leu	Lys	His	Met	
325				330				335								
gtc	ctc	ttc	agc	tcg	ctc	gcc	ggc	ttc	cac	ggc	aac	gtc	ggc	cag	tct	1056
Val	Leu	Phe	Ser	Ser	Leu	Ala	Gly	Phe	His	Gly	Asn	Val	Gly	Gln	Ser	
340				345				350								
gac	tac	gcc	atg	gcc	aac	gag	gcc	ctt	aac	aag	atg	ggc	ctc	gag	ctc	1104
Asp	Tyr	Ala	Met	Ala	Asn	Glu	Ala	Leu	Asn	Lys	Met	Gly	Leu	Glu	Leu	
355				360				365								
gcc	aag	gac	gtc	tcg	gtc	aag	tcg	atc	tgc	ttc	ggt	ccc	tgg	gac	ggt	1152
Ala	Lys	Asp	Val	Ser	Val	Lys	Ser	Ile	Cys	Phe	Gly	Pro	Trp	Asp	Gly	
370				375				380								
ggc	atg	gtg	acg	ccg	cag	ctc	aag	aag	cag	ttc	cag	gag	atg	ggc	gtg	1200
Gly	Met	Val	Thr	Pro	Gln	Leu	Lys	Lys	Gln	Phe	Gln	Glu	Met	Gly	Val	
385				390				395				400				
cag	atc	atc	ccc	cgc	gag	ggc	ggc	gct	gat	acc	gtg	gcg	cgc	atc	gtg	1248

Gln	Ile	Ile	Pro	Arg	Glu	Gly	Gly	Ala	Asp	Thr	Val	Ala	Arg	Ile	Val	
405				410				415								
ctc	ggc	tcc	tcg	ccg	gct	gag	atc	ctt	gtc	ggc	aac	tgg	cgc	acc	ccg	1296
Leu	Gly	Ser	Ser	Pro	Ala	Glu	Ile	Leu	Val	Gly	Asn	Trp	Arg	Thr	Pro	
420				425				430								
tcc	aag	aag	gtc	ggc	tcg	gac	acc	atc	acc	ctg	cac	cgc	aag	att	tcc	1344
Ser	Lys	Lys	Val	Gly	Ser	Asp	Thr	Ile	Thr	Leu	His	Arg	Lys	Ile	Ser	
435				440				445								
gcc	aag	tcc	aac	ccc	ttc	ctc	gag	gac	cac	gtc	atc	cag	ggc	cgc	cgc	1392
Ala	Lys	Ser	Asn	Pro	Phe	Leu	Glu	Asp	His	Val	Ile	Gln	Gly	Arg	Arg	
450				455				460								
gtg	ctg	ccc	atg	acg	ctg	gcc	att	ggc	tcg	ctc	gcg	gag	acc	tgc	ctc	1440
Val	Leu	Pro	Met	Thr	Leu	Ala	Ile	Gly	Ser	Leu	Ala	Glu	Thr	Cys	Leu	
465				470				475				480				
ggc	ctc	ttc	ccc	ggc	tac	tcg	ctc	tgg	gcc	att	gac	gac	gcc	cag	ctc	1488
Gly	Leu	Phe	Pro	Gly	Tyr	Ser	Leu	Trp	Ala	Ile	Asp	Asp	Ala	Gln	Leu	
485				490				495								
ttc	aag	ggt	gtc	act	gtc	gac	ggc	gac	gtc	aac	tgc	gag	gtg	acc	ctc	1536
Phe	Lys	Gly	Val	Thr	Val	Asp	Gly	Asp	Val	Asn	Cys	Glu	Val	Thr	Leu	
500				505				510								
acc	ccg	tcg	acg	gcg	ccc	tcg	ggc	cgc	gtc	aac	gtc	cag	gcc	acg	ctc	1584
Thr	Pro	Ser	Thr	Ala	Pro	Ser	Gly	Arg	Val	Asn	Val	Gln	Ala	Thr	Leu	
515				520				525								
aag	acc	ttt	tcc	agc	ggc	aag	ctg	gtc	ccg	gcc	tac	cgc	gcc	gtc	atc	1632
Lys	Thr	Phe	Ser	Ser	Gly	Lys	Leu	Val	Pro	Ala	Tyr	Arg	Ala	Val	Ile	
530				535				540								
gtg	ctc	tcc	aac	cag	ggc	gcg	ccc	ccg	gcc	aac	gcc	acc	atg	cag	ccg	1680
Val	Leu	Ser	Asn	Gln	Gly	Ala	Pro	Pro	Ala	Asn	Ala	Thr	Met	Gln	Pro	
545				550				555				560				
ccc	tcg	ctc	gat	gcc	gat	ccg	gcg	ctc	cag	ggc	tcc	gtc	tac	gac	ggc	1728
Pro	Ser	Leu	Asp	Ala	Asp	Pro	Ala	Leu	Gln	Gly	Ser	Val	Tyr	Asp	Gly	
565				570				575								
aag	acc	ctc	ttc	cac	ggc	ccg	gcc	ttc	cgc	ggc	atc	gat	gac	gtg	ctc	1776
Lys	Thr	Leu	Phe	His	Gly	Pro	Ala	Phe	Arg	Gly	Ile	Asp	Asp	Val	Leu	
580				585				590								
tcg	tgc	acc	aag	agc	cag	ctt	gtg	gcc	aag	tgc	agc	gct	gtc	ccc	ggc	1824
Ser	Cys	Thr	Lys	Ser	Gln	Leu	Val	Ala	Lys	Cys	Ser	Ala	Val	Pro	Gly	
595				600				605								
tcc	gac	gcc	gct	cgc	ggc	gag	ttt	gcc	acg	gac	act	gac	gcc	cat	gac	1872

Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp	
610	615 620
ccc ttc gtg aac gac ctg gcc ttt cag gcc atg ctc gtc tgg gtg cgc	1920
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg	
625	630 635 640
cgc acg ctc ggc cag gct gcg ctc ccc aac tcg atc cag cgc atc gtc	1968
Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val	
	645 650 655
cag cac cgc ccg gtc ccg cag gac aag ccc ttc tac att acc ctc cgc	2016
Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg	
	660 665 670
tcc aac cag tcg ggc ggt cac tcc cag cac aag cac gcc ctt cag ttc	2064
Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe	
	675 680 685
cac aac gag cag ggc gat ctc ttc att gat gtc cag gct tcg gtc atc	2112
His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile	
	690 695 700
gcc acg gac agc ctt gcc ttc	2133
Ala Thr Asp Ser Leu Ala Phe	
705	710

<210>18

<211>711

<212>PRT

<213>Schizochytrium sp.

<400>18

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

Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala Ile Val		
100	105	110
Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly Ile Gly		
115	120	125
Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu Thr Gly		
130	135	140
Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val Ser Gly		
145	150	155
Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr Arg Gln		
165	170	175
Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val Ser Ala		
180	185	190
Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala Val Gln		
195	200	205
Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly Glu Gly		
210	215	220
Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser Val Leu		
225	230	235
Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala Leu Gly		
245	250	255
Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala Asp Val		
260	265	270
Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg Val Ser		
275	280	285
Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile Glu Lys		
290	295	300
Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val Thr Gly		
305	310	315
Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys His Met		
325	330	335
Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly Gln Ser		
340	345	350
Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu Glu Leu		
355	360	365
Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp Asp Gly		
370	375	380
Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met Gly Val		
385	390	395
Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg Ile Val		

	405	410	415
Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg Thr Pro			
	420	425	430
Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys Ile Ser			
	435	440	445
Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly Arg Arg			
	450	455	460
Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr Cys Leu			
465	470	475	480
Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala Gln Leu			
	485	490	495
Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val Thr Leu			
	500	505	510
Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala Thr Leu			
	515	520	525
Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala Val Ile			
	530	535	540
Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met Gln Pro			
545	550	555	560
Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr Asp Gly			
	565	570	575
Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp Val Leu			
	580	585	590
Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val Pro Gly			
	595	600	605
Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala His Asp			
	610	615	620
Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp Val Arg			
625	630	635	640
Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg Ile Val			
	645	650	655
Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr Leu Arg			
	660	665	670
Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu Gln Phe			
	675	680	685
His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser Val Ile			
	690	695	700
Ala Thr Asp Ser Leu Ala Phe			
705	710		

<210>19

<211>1350

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (1350)

<223>

<400>19

atg gcc gct cgg aat gtg agc gcc gcg cat gag atg cac gat gaa aag	48
Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys	
1 5 10 15	
cgc atc gcc gtc gtc ggc atg gcc gtc cag tac gcc gga tgc aaa acc	96
Arg Ile Ala Val Val Gly Met Ala Val Gln Tyr Ala Gly Cys Lys Thr	
20 25 30	
aag gac gag ttc tgg gag gtg ctc atg aac ggc aag gtc gag tcc aag	144
Lys Asp Glu Phe Trp Glu Val Leu Met Asn Gly Lys Val Glu Ser Lys	
35 40 45	
gtg atc agc gac aaa cga ctc ggc tcc aac tac cgc gcc gag cac tac	192
Val Ile Ser Asp Lys Arg Leu Gly Ser Asn Tyr Arg Ala Glu His Tyr	
50 55 60	
aaa gca gag cgc agc aag tat gcc gac acc ttt tgc aac gaa acg tac	240
Lys Ala Glu Arg Ser Lys Tyr Ala Asp Thr Phe Cys Asn Glu Thr Tyr	
65 70 75 80	
ggc acc ctt gac gag aac gag atc gac aac gag cac gaa ctc ctc ctc	288
Gly Thr Leu Asp Glu Asn Glu Ile Asp Asn Glu His Glu Leu Leu Leu	
85 90 95	
aac ctc gcc aag cag gca ctc gca gag aca tcc gtc aaa gac tcg aca	336
Asn Leu Ala Lys Gln Ala Leu Ala Glu Thr Ser Val Lys Asp Ser Thr	
100 105 110	
cgc tgc ggc atc gtc agc ggc tgc ctc tcg ttc ccc atg gac aac ctc	384
Arg Cys Gly Ile Val Ser Gly Cys Leu Ser Phe Pro Met Asp Asn Leu	
115 120 125	
cag ggt gaa ctc ctc aac gtg tac caa aac cat gtc gag aaa aag ctc	432
Gln Gly Glu Leu Leu Asn Val Tyr Gln Asn His Val Glu Lys Lys Leu	
130 135 140	

ggg gcc cgc gtc ttc aag gac gcc tcc cat tgg tcc gaa cgc gag cag	480
Gly Ala Arg Val Phe Lys Asp Ala Ser His Trp Ser Glu Arg Glu Gln	
145 150 155 160	
tcc aac aaa ccc gag gcc ggt gac cgc cgc atc ttc atg gac ccg gcc	528
Ser Asn Lys Pro Glu Ala Gly Asp Arg Arg Ile Phe Met Asp Pro Ala	
165 170 175	
tcc ttc gtc gcc gaa gaa ctc aac ctc ggc gcc ctt cac tac tcc gtc	576
Ser Phe Val Ala Glu Glu Leu Asn Leu Gly Ala Leu His Tyr Ser Val	
180 185 190	
gac gca gca tgc gcc acg gcg ctc tac gtg ctc cgc ctc gcg cag gat	624
Asp Ala Ala Cys Ala Thr Ala Leu Tyr Val Leu Arg Leu Ala Gln Asp	
195 200 205	
cat ctc gtc tcc ggc gcc gcc gac gtc atg ctc tgc ggt gcc acc tgc	672
His Leu Val Ser Gly Ala Ala Asp Val Met Leu Cys Gly Ala Thr Cys	
210 215 220	
ctg ccg gag ccc ttt ttc atc ctt tgc ggc ttt tcc acc ttc cag gcc	720
Leu Pro Glu Pro Phe Phe Ile Leu Ser Gly Phe Ser Thr Phe Gln Ala	
225 230 235 240	
atg ccc gtc ggc acg ggc cag aac gtg tcc atg ccg ctg cac aag gac	768
Met Pro Val Gly Thr Gly Gln Asn Val Ser Met Pro Leu His Lys Asp	
245 250 255	
agc cag ggc ctc acc ccg ggt gag ggc ggc tcc atc atg gtc ctc aag	816
Ser Gln Gly Leu Thr Pro Gly Glu Gly Gly Ser Ile Met Val Leu Lys	
260 265 270	
cgt ctc gat gat gcc atc cgc gac ggc gac cac att tac ggc acc ctt	864
Arg Leu Asp Asp Ala Ile Arg Asn Gly Asp His Ile Tyr Gly Thr Leu	
275 280 285	
ctc ggc gcc aat gtc agc aac tcc ggc aca ggt ctg ccc ctc aag ccc	912
Leu Gly Ala Asn Val Ser Asn Ser Gly Thr Gly Leu Pro Leu Lys Pro	
290 295 300	
ctt ctc ccc agc gag aaa aag tgc ctc atg gac acc tac acg cgc att	960
Leu Leu Pro Ser Glu Lys Lys Cys Leu Met Asp Thr Tyr Thr Arg Ile	
305 310 315 320	
aac gtg cac ccg cac aag att cag tac gtc gag tgc cac gcc acc ggc	1008
Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly	
325 330 335	
acg ccc cag ggt gat cgt gtg gaa atc gac gcc gtc aag gcc tgc ttt	1056
Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe	
340 345 350	

gaa ggc aag gtc ccc cgt ttc ggt acc aca aag ggc aac ttt gga cac	1104
Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His	
355 360 365	
acc cts gyc gca gcc ggc ttt gcc ggt atg tgc aag gtc ctc ctc tcc	1152
Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser	
370 375 380	
atg aag cat ggc atc atc ccg ccc acc ccg ggt atc gat gac gag acc	1200
Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr	
385 390 395 400	
aag atg gac cct ctc gtc gtc tcc ggt gag gcc atc cca tgg cca gag	1248
Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu	
405 410 415	
acc aac ggc gag ccc aag cgc gcc ggt ctc tcg gcc ttt ggc ttt ggt	1296
Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly	
420 425 430	
ggc acc aac gcc cat gcc gtc ttt gag gag cat gac ccc tcc aac gcc	1344
Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala	
435 440 445	
gcc tgc	1350
Ala Cys	
450	

<210>20

<211>450

<212>PRT

<213>Schizochytrium sp.

<220>

<221>misc_feature

<222>(370).. (370)

<223>在 370 位的' Xaa' 代表 eu.

<220>

<221>misc_feature

<222>(371).. (371)

<223>在 371 位的' Xaa' 代表 Ala 或 Val.

<400>20

Met Ala Ala Arg Asn Val Ser Ala Ala His Glu Met His Asp Glu Lys

1	5	10	15
Arg Ile Ala Val	Val Gly Met Ala Val	Gln Tyr Ala Gly	Cys Lys Thr
20	25	30	
Lys Asp Glu Phe	Trp Glu Val Leu Met	Asn Gly Lys Val	Glu Ser Lys
35	40	45	
Val Ile Ser Asp	Lys Arg Leu Gly Ser	Asn Tyr Arg Ala	Glu His Tyr
50	55	60	
Lys Ala Glu Arg	Ser Lys Tyr Ala Asp	Thr Phe Cys Asn	Glu Thr Tyr
65	70	75	80
Gly Thr Leu Asp	Glu Asn Glu Ile Asp	Asn Glu His Glu	Leu Leu Leu
85	90	95	
Asn Leu Ala Lys	Gln Ala Leu Ala Glu	Thr Ser Val Lys	Asp Ser Thr
100	105	110	
Arg Cys Gly Ile	Val Ser Gly Cys Leu	Ser Phe Pro Met	Asp Asn Leu
115	120	125	
Gln Gly Glu Leu	Leu Asn Val Tyr Gln	Asn His Val Glu	Lys Lys Leu
130	135	140	
Gly Ala Arg Val	Phe Lys Asp Ala Ser	His Trp Ser Glu	Arg Glu Gln
145	150	155	160
Ser Asn Lys Pro	Glu Ala Gly Asp Arg	Arg Ile Phe Met	Asp Pro Ala
165	170	175	
Ser Phe Val Ala	Glu Glu Leu Asn Leu	Gly Ala Leu His	Tyr Ser Val
180	185	190	
Asp Ala Ala Cys	Ala Thr Ala Leu Tyr	Val Leu Arg Leu	Ala Gln Asp
195	200	205	
His Leu Val Ser	Gly Ala Ala Asp Val	Met Leu Cys Gly	Ala Thr Cys
210	215	220	
Leu Pro Glu Pro	Phe Phe Ile Leu Ser	Gly Phe Ser Thr	Phe Gln Ala
225	230	235	240
Met Pro Val Gly	Thr Gly Gln Asn Val	Ser Met Pro Leu	His Lys Asp
245	250	255	
Ser Gln Gly Leu	Thr Pro Gly Glu Gly	Gly Ser Ile Met	Val Leu Lys
260	265	270	
Arg Leu Asp Asp	Ala Ile Arg Asp Gly	Asp His Ile Tyr	Gly Thr Leu
275	280	285	
Leu Gly Ala Asn	Val Ser Asn Ser Gly	Thr Gly Leu Pro	Leu Lys Pro
290	295	300	
Leu Leu Pro Ser	Glu Lys Lys Cys Leu	Met Asp Thr Tyr	Thr Arg Ile
305	310	315	320

Asn Val His Pro His Lys Ile Gln Tyr Val Glu Cys His Ala Thr Gly
 325 330 335
 Thr Pro Gln Gly Asp Arg Val Glu Ile Asp Ala Val Lys Ala Cys Phe
 340 345 350
 Glu Gly Lys Val Pro Arg Phe Gly Thr Thr Lys Gly Asn Phe Gly His
 355 360 365
 Thr Xaa Xaa Ala Ala Gly Phe Ala Gly Met Cys Lys Val Leu Leu Ser
 370 375 380
 Met Lys His Gly Ile Ile Pro Pro Thr Pro Gly Ile Asp Asp Glu Thr
 385 390 395 400
 Lys Met Asp Pro Leu Val Val Ser Gly Glu Ala Ile Pro Trp Pro Glu
 405 410 415
 Thr Asn Gly Glu Pro Lys Arg Ala Gly Leu Ser Ala Phe Gly Phe Gly
 420 425 430
 Gly Thr Asn Ala His Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala
 435 440 445
 Ala Cys
 450

<210>21

<211>1323

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (1323)

<223>

<400>21

tcg gcc cgc tgc ggc ggt gaa agc aac atg cgc atc gcc atc act ggt	48
Ser Ala Arg Cys Gly Gly Glu Ser Asn Met Arg Ile Ala Ile Thr Gly	
1 5 10 15	
atg gac gcc acc ttt ggc gct ctc aag gga ctc gac gcc ttc gag cgc	96
Met Asp Ala Thr Phe Gly Ala Leu Lys Gly Leu Asp Ala Phe Glu Arg	
20 25 30	
gcc att tac acc ggc gct cac ggt gcc atc cca ctc cca gaa aag cgc	144
Ala Ile Tyr Thr Gly Ala His Gly Ala Ile Pro Leu Pro Glu Lys Arg	
35 40 45	

tgg cgc ttt ctc ggc aag gac aag gac ttt ctt gac ctc tgc ggc gtc	192
Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe Leu Asp Leu Cys Gly Val	
50 55 60	
aag gcc acc ccg cac ggc tgc tac att gaa gat gtt gag gtc gac ttc	240
Lys Ala Thr Pro His Gly Cys Tyr Ile Glu Asp Val Glu Val Asp Phe	
65 70 75 80	
cag cgc ctc cgc acg ccc atg acc cct gaa gac atg ctc ctc cct cag	288
Gln Arg Leu Arg Thr Pro Met Thr Pro Glu Asp Met Leu Leu Pro Gln	
85 90 95	
cag ctt ctg gcc gtc acc acc att gac cgc gcc atc ctc gac tcg gga	336
Gln Leu Leu Ala Val Thr Thr Ile Asp Arg Ala Ile Leu Asp Ser Gly	
100 105 110	
atg aaa aag ggt ggc aat gtc gcc gtc ttt gtc ggc ctc ggc acc gac	384
Met Lys Lys Gly Gly Asn Val Ala Val Phe Val Gly Leu Gly Thr Asp	
115 120 125	
ctc gag ctc tac cgt cac cgt gct cgc gtc gct ctc aag gag cgc gtc	432
Leu Glu Leu Tyr Arg His Arg Ala Arg Val Ala Leu Lys Glu Arg Val	
130 135 140	
cgc cct gaa gcc tcc aag aag ctc aat gac atg atg cag tac att aac	480
Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp Met Met Gln Tyr Ile Asn	
145 150 155 160	
gac tgc ggc aca tcc aca tcg tac acc tcg tac att ggc aac ctc gtc	528
Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser Tyr Ile Gly Asn Leu Val	
165 170 175	
gcc acg cgc gtc tcg tcg cag tgg ggc ttc acg ggc ccc tcc ttt acg	576
Ala Thr Arg Val Ser Ser Gln Trp Gly Phe Thr Gly Pro Ser Phe Thr	
180 185 190	
atc acc gag ggc aac aac tcc gtc tac cgc tgc gcc gag ctc ggc aag	624
Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg Cys Ala Glu Leu Gly Lys	
195 200 205	
tac ctc ctc gag acc ggc gag gtc gat ggc gtc gtc gtt gcg ggt gtc	672
Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly Val Val Val Ala Gly Val	
210 215 220	
gat ctc tgc ggc agt gcc gaa aac ctt tac gtc aag tct cgc cgc ttc	720
Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr Val Lys Ser Arg Arg Phe	
225 230 235 240	
aag gtg tcc acc tcc gat acc ccg cgc gcc agc ttt gac gcc gcc gcc	768
Lys Val Ser Thr Ser Asp Thr Pro Arg Ala Ser Phe Asp Ala Ala Ala	
245 250 255	

gat ggc tac ttt gtc ggc gag ggc tgc ggt gcc ttt gtg ctc aag cgt	816
Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly Ala Phe Val Leu Lys Arg	
260 265 270	
gag act agc tgc acc aag gac gac cgt atc tac gct tgc atg gat gcc	864
Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile Tyr Ala Gys Met Asp Ala	
275 280 285	
atc gtc cct ggc aac gtc cct agc gcc tgc ttg cgc gag gcc ctc gac	912
Ile Val Pro Gly Asn Val Pro Ser Ala Cys Leu Arg Glu Ala Leu Asp	
290 295 300	
cag gcg cgc gtc aag ccg ggc gat atc gag atg ctc gag ctc agc gcc	960
Gln Ala Arg Val Lys Pro Gly Asp Ile Glu Met Leu Glu Leu Ser Ala	
305 310 315 320	
gac tcc gcc cgc cac ctc aag gac ccg tcc gtc ctg ccc aag gag ctc	1008
Asp Ser Ala Arg His Leu Lys Asp Pro Ser Val Leu Pro Lys Glu Leu	
325 330 335	
act gcc gag gag gaa atc ggc ggc ctt cag acg atc ctt cgt gac gat	1056
Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln Thr Ile Leu Arg Asp Asp	
340 345 350	
gac aag ctc ccg cgc aac gtc gca acg ggc agt gtc aag gcc acc gtc	1104
Asp Lys Leu Pro Arg Asn Val Ala Thr Gly Ser Val Lys Ala Thr Val	
355 360 365	
ggt gac acc ggt tat gcc tct ggt gct gcc agc ctc atc aag gct gcg	1152
Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala Ser Leu Ile Lys Ala Ala	
370 375 380	
ctt tgc atc tac aac cgc tac ctg ccc agc aac ggc gac gac tgg gat	1200
Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser Asn Gly Asp Asp Trp Asp	
385 390 395 400	
gaa ccc gcc cct gag gcg ccc tgg gac agc acc ctc ttt gcg tgc cag	1248
Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser Thr Leu Phe Ala Cys Gln	
405 410 415	
acc tcg cgc gct tgg ctc aag aac cct ggc gag cgt cgc tat gcg gcc	1296
Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly Glu Arg Arg Tyr Ala Ala	
420 425 430	
gtc tcg ggc gtc tcc gag acg cgc tcg	1323
Val Ser Gly Val Ser Glu Thr Arg Ser	
435 440	

<210>22

<211>441

<212>PRT

<213>Schizochytrium sp.

<400>22

Ser	Ala	Arg	Cys	Gly	Gly	Glu	Ser	Asn	Met	Arg	Ile	Ala	Ile	Thr	Gly
1				5					10					15	
Met	Asp	Ala	Thr	Phe	Gly	Ala	Leu	Lys	Gly	Leu	Asp	Ala	Phe	Glu	Arg
			20					25					30		
Ala	Ile	Tyr	Thr	Gly	Ala	His	Gly	Ala	Ile	Pro	Leu	Pro	Glu	Lys	Arg
		35					40					45			
Trp	Arg	Phe	Leu	Gly	Lys	Asp	Lys	Asp	Phe	Leu	Asp	Leu	Cys	Gly	Val
	50					55				60					
Lys	Ala	Thr	Pro	His	Gly	Cys	Tyr	Ile	Glu	Asp	Val	Glu	Val	Asp	Phe
65				70					75					80	
Gln	Arg	Leu	Arg	Thr	Pro	Met	Thr	Pro	Glu	Asp	Met	Leu	Leu	Pro	Gln
			85					90					95		
Gln	Leu	Leu	Ala	Val	Thr	Thr	Ile	Asp	Arg	Ala	Ile	Leu	Asp	Ser	Gly
			100					105					110		
Met	Lys	Lys	Gly	Gly	Asn	Val	Ala	Val	Phe	Val	Gly	Leu	Gly	Thr	Asp
		115				120					125				
Leu	Glu	Leu	Tyr	Arg	His	Arg	Ala	Arg	Val	Ala	Leu	Lys	Glu	Arg	Val
	130					135					140				
Arg	Pro	Glu	Ala	Ser	Lys	Lys	Leu	Asn	Asp	Met	Met	Gln	Tyr	Ile	Asn
145					150					155				160	
Asp	Cys	Gly	Thr	Ser	Thr	Ser	Tyr	Thr	Ser	Tyr	Ile	Gly	Asn	Leu	Val
			165					170					175		
Ala	Thr	Arg	Val	Ser	Ser	Gln	Trp	Gly	Phe	Thr	Gly	Pro	Ser	Phe	Thr
		180						185					190		
Ile	Thr	Glu	Gly	Asn	Asn	Ser	Val	Tyr	Arg	Cys	Ala	Glu	Leu	Gly	Lys
		195				200						205			
Tyr	Leu	Leu	Glu	Thr	Gly	Glu	Val	Asp	Gly	Val	Val	Val	Ala	Gly	Val
	210					215						220			
Asp	Leu	Cys	Gly	Ser	Ala	Glu	Asn	Leu	Tyr	Val	Lys	Ser	Arg	Arg	Phe
225					230					235				240	
Lys	Val	Ser	Thr	Ser	Asp	Thr	Pro	Arg	Ala	Ser	Phe	Asp	Ala	Ala	Ala
			245					250					255		
Asp	Gly	Tyr	Phe	Val	Gly	Glu	Gly	Cys	Gly	Ala	Phe	Val	Leu	Lys	Are
		260						265					270		
Glu	Thr	Ser	Cys	Thr	Lys	Asp	Asp	Arg	Ile	Tyr	Ala	Cys	Met	Asp	Ala

275					280					285						
Ile	Val	Pro	Gly	Asn	Val	Pro	Ser	Ala	Cys	Leu	Arg	Glu	Ala	Leu	Asp	
290					295					300						
Gln	Ala	Arg	Val	Lys	Pro	Gly	Asp	Ile	Glu	Met	Leu	Glu	Leu	Ser	Ala	
305					310					315					320	
Asp	Ser	Ala	Arg	His	Leu	Lys	Asp	Pro	Ser	Val	Leu	Pro	Lys	Glu	Leu	
325					330					335						
Thr	Ala	Glu	Glu	Glu	Ile	Gly	Gly	Leu	Gln	Thr	Ile	Leu	Arg	Asp	Asp	
340					345					350						
Asp	Lys	Leu	Pro	Arg	Asn	Val	Ala	Thr	Gly	Ser	Val	Lys	Ala	Thr	Val	
355					360					365						
Gly	Asp	Thr	Gly	Tyr	Ala	Ser	Gly	Ala	Ala	Ser	Leu	Ile	Lys	Ala	Ala	
370					375					380						
Leu	Cys	Ile	Tyr	Asn	Arg	Tyr	Leu	Pro	Ser	Asn	Gly	Asp	Asp	Trp	Asp	
385					390					395					400	
Glu	Pro	Ala	Pro	Glu	Ala	Pro	Trp	Asp	Ser	Thr	Leu	Phe	Ala	Cys	Gln	
405					410					415						
Thr	Ser	Arg	Ala	Trp	Leu	Lys	Asn	Pro	Gly	Glu	Arg	Arg	Tyr	Ala	Ala	
420					425					430						
Val	Ser	Gly	Val	Ser	Glu	Thr	Arg	Ser								
435					440											

 $\langle 210 \rangle_{23}$ $\langle 211 \rangle 1500$

$\langle 212 \rangle$ DNA

<213>Schizochytrium sp.

 $\langle 220 \rangle$ $\langle 221 \rangle$ CDS $\langle 222 \rangle (1) \dots (1500)$

⟨223⟩

 $\langle 400 \rangle_{23}$

tgc	tat	tcc	gtg	ctc	ctc	tcc	gaa	gcc	gag	ggc	cac	tac	gag	cgc	gag	48
Cys	Tyr	Ser	Val	Leu	Leu	Ser	Glu	Ala	Glu	Gly	His	Tyr	Glu	Arg	Glu	
1				5					10					15		
aac	cgc	atc	tgc	ctc	gac	gag	gag	gcg	ccc	aag	ctc	att	gtg	ctt	cgc	96
Asn	Arg	Ile	Ser	Leu	Asp	Glu	Glu	Ala	Pro	Lys	Leu	Ile	Val	Leu	Arg	
				20				25					30			

gcc gac tcc cac gag gag atc ctt ggt cgc ctc gac aag atc cgc gag	144
Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu	
35 40 45	
cgc ttc ttg cag ccc acg ggc gcc gcc ccg cgc gag tcc gag ctc aag	192
Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys	
50 55 60	
gcg cag gcc cgc cgc atc ttc ctc gag ctc ctc ggc gag acc ctt gcc	240
Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala	
65 70 75 80	
cag gat gcc gct tct tca ggc tcg caa aag ccc ctc gct ctc agc ctc	288
Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu	
85 90 95	
gtc tcc acg ccc tcc aag ctc cag cgc gag gtc gag ctc gcg gcc aag	336
Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys	
100 105 110	
ggt atc ccg cgc tgc ctc aag atg cgc cgc gat tgg agc tcc cct gct	384
Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala	
115 120 125	
ggc agc cgc tac gcg cct gag ccg ctc gcc agc gac cgc gtc gcc ttc	432
Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe	
130 135 140	
atg tac ggc gaa ggt cgc agc cct tac tac ggc atc acc caa gac att	480
Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile	
145 150 155 160	
cac cgc att tgg ccc gaa ctc cac gag gtc atc aac gaa aag acg aac	528
His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn	
165 170 175	
cgt ctc tgg gcc gaa ggc gac cgc tgg gtc atg ccg cgc gcc agc ttc	576
Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe	
180 185 190	
aag tcg gag ctc gag agc cag cag caa gag ttt gat cgc aac atg att	624
Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile	
195 200 205	
gaa atg ttc cgt ctt gga atc ctc acc tca att gcc ttc acc aat ctg	672
Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu	
210 215 220	
gcg cgc gac gtt ctc aac atc acg ccc aag gcc gcc ttt ggc ctc agt	720
Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser	
225 230 235 240	

ctt ggc gag att tcc atg att ttt gcc ttt tcc aag aag aac ggt ctc	768
Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu	
245 250 255	
atc tcc gac cag ctc acc aag gat ctt cgc gag tcc gac gtg tgg aac	816
Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn	
260 265 270	
aag gct ctg gcc gtt gaa ttt aat gcg ctg cgc gag gcc tgg ggc att	864
Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile	
275 280 285	
cca cag agt gtc ccc aag gac gag ttc tgg caa ggc tac att gtg cgc	912
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg	
290 295 300	
ggc acc aag cag gat atc gag gcg gcc atc gcc ccg gac agc aag tac	960
Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr	
305 310 315 320	
gtg cgc ctc acc atc atc aat gat gcc aac acc gcc ctc att agc ggc	1008
Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly	
325 330 335	
aag ccc gac gcc tgc aag gct gcg atc gcg cgt ctc ggt ggc aac att	1056
Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile	
340 345 350	
cct gcg ctt ccc gtg acc cag ggc atg tgc ggc cac tgc ccc gag gtg	1104
Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val	
355 360 365	
gga cct tat acc aag gat atc gcc aag atc cat gcc aac ctt gag ttc	1152
Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe	
370 375 380	
ccc gtt gtc gac ggc ctt gac ctc tgg acc aca atc aac cag aag cgc	1200
Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg	
385 390 395 400	
ctc gtg cca cgc gcc acg ggc gcc aag gac gaa tgg gcc cct tct tcc	1248
Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser	
405 410 415	
ttt ggc gag tac gcc ggc cag ctc tac gag aag cag gct aac ttc ccc	1296
Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro	
420 425 430	
caa atc gtc gag acc att tac aag caa aac tac gac gtc ttt gtc gag	1344
Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu	
435 440 445	

```

gtt ggg ccc aac aac cac cgt agc acc gca gtg cgc acc acg ctt ggt      1392
Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly
      450                      455                      460
ccc cag cgc aac cac ctt gct ggc gcc atc gac aag cag aac gag gat      1440
Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp
465                      470                      475                      480
gct tgg acg acc atc gtc aag ctt gtg gct tcg ctc aag gcc cac ctt      1488
Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu
      485                      490                      495
gtt cct ggc gtc      1500
Val Pro Gly Val
      500

```

<210>24

<211>500

<212>PRT

<213>Schizochytrium sp.

<400>24

```

Cys Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu
1              5              10              15
Asn Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg
      20              25              30
Ala Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu
      35              40              45
Arg Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys
      50              55              60
Ala Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala
65              70              75              80
Gln Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu
      85              90              95
Val Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys
      100             105             110
Gly Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala
      115             120             125
Gly Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe
      130             135             140
Met Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile
145             150             155             160

```

His Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn		
165	170	175
Arg Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe		
180	185	190
Lys Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile		
195	200	205
Glu Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu		
210	215	220
Ala Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser		
225	230	235
Leu Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu		
245	250	255
Ile Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn		
260	265	270
Lys Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile		
275	280	285
Pro Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg		
290	295	300
Gly Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr		
305	310	315
Val Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly		
325	330	335
Lys Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile		
340	345	350
Pro Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val		
353	360	365
Gly Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe		
370	375	380
Pro Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg		
385	390	395
Leu Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser		
405	410	415
Phe Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro		
420	425	430
Gln Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu		
435	440	445
Val Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly		
450	455	460
Pro Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp		

465	470	475	480
Ala Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu			
	485	490	495
Val Pro Gly Val			
500			

<210>25

<211>1530

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (1530)

<223>

<400>25

ctg ctc gat ctc gac agt atg ctt gcg ctg agc tct gcc agt gcc tcc	48
Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser	
1 5 10 15	
ggc aac ctt gtt gag act gcg cct agc gac gcc tcg gtc att gtg ccg	96
Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro	
20 25 30	
ccc tgc aac att gcg gat ctc ggc agc cgc gcc ttc atg aaa acg tac	144
Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr	
35 40 45	
ggt gtt tcg gcg cct ctg tac acg ggc gcc atg gcc aag ggc att gcc	192
Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala	
50 55 60	
tct gcg gac ctc gtc att gcc gcc ggc cgc cag ggc atc ctt gcg tcc	240
Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser	
65 70 75 80	
ttt ggc gcc ggc gga ctt ccc atg cag gtt gtg cgt gag tcc atc gaa	288
Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu	
85 90 95	
aag att cag gcc gcc ctg ccc aat ggc ccg tac gct gtc aac ctt atc	336
Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile	
100 105 110	
cat tct ccc ttt gac agc aac ctc gaa aag ggc aat gtc gat ctc ttc	384

His	Ser	Pro	Phe	Asp	Ser	Asn	Leu	Glu	Lys	Gly	Asn	Val	Asp	Leu	Phe	
115				120				125								
ctc	gag	aag	ggt	gtc	acc	ttt	gtc	gag	gcc	tcg	gcc	ttt	atg	acg	ctc	432
Leu	Glu	Lys	Gly	Val	Thr	Phe	Val	Glu	Ala	Ser	Ala	Phe	Met	Thr	Leu	
130				135				140								
acc	ccg	cag	gtc	gtg	cgg	tac	cgc	gcg	gct	ggc	ctc	acg	cgc	aac	gcc	480
Thr	Pro	Gln	Val	Val	Arg	Tyr	Arg	Ala	Ala	Gly	Leu	Thr	Arg	Asn	Ala	
145				150				155				160				
gac	ggc	tcg	gtc	aac	atc	cgc	aac	cgt	atc	att	ggc	aag	gtc	tcg	cgc	528
Asp	Gly	Ser	Val	Asn	Ile	Arg	Asn	Arg	Ile	Ile	Gly	Lys	Val	Ser	Arg	
165				170				175								
acc	gag	ctc	gcc	gag	atg	ttc	atg	cgt	cct	gcg	ccc	gag	cac	ctt	ctt	576
Thr	Glu	Leu	Ala	Glu	Met	Phe	Met	Arg	Pro	Ala	Pro	Glu	His	Leu	Leu	
180				185				190								
cag	aag	ctc	att	gct	tcc	ggc	gag	atc	aac	cag	gag	cag	gcc	gag	ctc	624
Gln	Lys	Leu	Ile	Ala	Ser	Gly	Glu	Ile	Asn	Gln	Glu	Gln	Ala	Glu	Leu	
195				200				205								
gcc	cgc	cgt	gtt	ccc	gtc	gct	gac	gac	atc	gcg	gtc	gaa	gct	gac	tcg	672
Ala	Arg	Arg	Val	Pro	Val	Ala	Asp	Asp	Ile	Ala	Val	Glu	Ala	Asp	Ser	
210				215				220								
ggt	ggc	cac	acc	gac	aac	cgc	ccc	atc	cac	gtc	att	ctg	ccc	ctc	atc	720
Gly	Gly	His	Thr	Asp	Asn	Arg	Pro	Ile	His	Val	Ile	Leu	Pro	Leu	Ile	
225				230				235				240				
atc	aac	ctt	cgc	gac	cgc	ctt	cac	cgc	gag	tgc	ggc	tac	ccg	gcc	aac	768
Ile	Asn	Leu	Arg	Asp	Arg	Leu	His	Arg	Glu	Cys	Gly	Tyr	Pro	Ala	Asn	
245				250				255								
ctt	cgc	gtc	cgt	gtg	ggc	gcc	ggc	ggt	ggc	att	ggg	tgc	ccc	cag	gcg	816
Leu	Arg	Val	Arg	Val	Gly	Ala	Gly	Gly	Gly	Ile	Gly	Cys	Pro	Gln	Ala	
260				265				270								
gcg	ctg	gcc	acc	ttc	aac	atg	ggt	gcc	tcc	ttt	att	gtc	acc	ggc	acc	864
Ala	Leu	Ala	Thr	Phe	Asn	Met	Gly	Ala	Ser	Phe	Ile	Val	Thr	Gly	Thr	
275				280				285								
gtg	aac	cag	gtc	gcc	aag	cag	tcg	ggc	acg	tgc	gac	aat	gtg	cgc	aag	912
Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly	Thr	Cys	Asp	Asn	Val	Arg	Lys	
290				295				300								
cag	ctc	gcg	aag	gcc	act	tac	tcg	gac	gta	tgc	atg	gcc	ccg	gct	gcc	960
Gln	Leu	Ala	Lys	Ala	Thr	Tyr	Ser	Asp	Val	Cys	Met	Ala	Pro	Ala	Ala	
305				310				315				320				
gac	atg	ttc	gag	gaa	ggc	gtc	aag	ctt	cag	gtc	ctc	aag	aag	gga	acc	1008

Asp Met Phe GlG Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr	
325	330
atg ttt ccc tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac	1056
Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr	
340	345
gac tcg ttc gag tcc atg ccc ccc gca gag ctt gcg cgc gtc gag aag	1104
Asp Ser Phe Glu Ser Met Pro Pro Ala Glu Leu Ala Arg Val Glu Lys	
355	360
cgc atc ttc agc cgc gcg ctc gaa gag gtc tgg gac gag acc aaa aac	1152
Arg Ile Phe Ser Arg Ala Leu Glu Glu Val Trp Asp Glu Thr Lys Asn	
370	375
ttt tac att aac cgt ctt cac aac ccg gag aag atc cag cgc gcc gag	1200
Phe Tyr Ile Asn Arg Leu His Asn Pro Glu Lys Ile Gln Arg Ala Glu	
385	390
cgc gac ccc aag ctc aag atg tcg ctg tgc ttt cgc tgg tac ctg agc	1248
Arg Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Ser	
405	410
ctg gcg agc cgc tgg gcc aac act gga gct tcc gat cgc gtc atg gac	1296
Leu Ala Ser Arg Trp Ala Asn Thr Gly Ala Ser Asp Arg Val Met Asp	
420	425
tac cag gtc tgg tgc ggt cct gcc att ggt tcc ttc aac gat ttc atc	1344
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ser Phe Asn Asp Phe Ile	
435	440
aag gga act tac ctt gat ccg gcc gtc gca aac gag tac ccg tgc gtc	1392
Lys Gly Thr Tyr Leu Asp Pro Ala Val Ala Asn Glu Tyr Pro Cys Val	
450	455
gtt cag att aac aag cag atc ctt cgt gga gcg tgc ttc ttg cgc cgt	1440
Val Gln Ile Asn Lys Gln Ile Leu Arg Gly Ala Cys Phe Leu Arg Arg	
465	470
ctc gaa att ctg cgc aac gca cgc ctt tcc gat ggc gct gcc gct ctt	1488
Leu Glu Ile Leu Arg Asn Ala Arg Leu Ser Asp Gly Ala Ala Ala Leu	
485	490
gtg gcc agc atc gat gac aca tac gtc ccg gcc gag aag ctg	1530
Val Ala Ser Ile Asp Asp Thr Tyr Val Pro Ala Glu Lys Leu	
500	505

<210>26

<211>510

<212>PRT

<213>Schizochytrium sp.

<400>26

```

Leu Leu Asp Leu Asp Ser Met Leu Ala Leu Ser Ser Ala Ser Ala Ser
1           5           10           15
Gly Asn Leu Val Glu Thr Ala Pro Ser Asp Ala Ser Val Ile Val Pro
          20           25           30
Pro Cys Asn Ile Ala Asp Leu Gly Ser Arg Ala Phe Met Lys Thr Tyr
          35           40           45
Gly Val Ser Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala
          50           55           60
Ser Ala Asp Leu Val Ile Ala Ala Gly Arg Gln Gly Ile Leu Ala Ser
65           70           75           80
Phe Gly Ala Gly Gly Leu Pro Met Gln Val Val Arg Glu Ser Ile Glu
          85           90           95
Lys Ile Gln Ala Ala Leu Pro Asn Gly Pro Tyr Ala Val Asn Leu Ile
          100          105          110
His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe
          115          120          125
Leu Glu Lys Gly Val Thr Phe Val Glu Ala Ser Ala Phe Met Thr Leu
          130          135          140
Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu Thr Arg Asn Ala
145          150          155          160
Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg
          165          170          175
Thr Glu Leu Ala Glu Met Phe Met Arg Pro Ala Pro Glu His Leu Leu
          180          185          190
Gln Lys Leu Ile Ala Ser Gly Glu Ile Asn Gln Glu Gln Ala Glu Leu
          195          200          205
Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser
          210          215          220
Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile
225          230          235          240
Ile Asn Leu Arg Asp Arg Leu His Arg Glu Cys Gly Tyr Pro Ala Asn
          245          250          255
Leu Arg Val Arg Val Gly Ala Gly Gly Gly Ile Gly Cys Pro Gln Ala
          260          265          270
Ala Leu Ala Thr Phe Asn Met Gly Ala Ser Phe Ile Val Thr Gly Thr
          275          280          285

```

Val	Asn	Gln	Val	Ala	Lys	Gln	Ser	Gly	Thr	Cys	Asp	Asn	Val	Arg	Lys
290						295					300				
Gln	Leu	Ala	Lys	Ala	Thr	Tyr	Ser	Asp	Val	Cys	Met	Ala	Pro	Ala	Ala
305					310					315				320	
Asp	Met	Phe	Glu	Glu	Gly	Val	Lys	Leu	Gln	Val	Leu	Lys	Lys	Gly	Thr
			325						330					335	
Met	Phe	Pro	Ser	Arg	Ala	Asn	Lys	Leu	Tyr	Glu	Leu	Phe	Cys	Lys	Tyr
			340						345				350		
Asp	Ser	Phe	Glu	Ser	Met	Pro	Pro	Ala	Glu	Leu	Ala	Arg	Val	Glu	Lys
		355						360				365			
Arg	Ile	Phe	Ser	Arg	Ala	Leu	Glu	Glu	Val	Trp	Asp	Glu	Thr	Lys	Asn
	370					375					380				
Phe	Tyr	Ile	Asn	Arg	Leu	His	Asn	Pro	Glu	Lys	Ile	Gln	Arg	Ala	Glu
385				390						395				400	
Arg	Asp	Pro	Lys	Leu	Lys	Met	Ser	Leu	Cys	Phe	Arg	Trp	Tyr	Leu	Ser
			405						410					415	
Leu	Ala	Ser	Arg	Trp	Ala	Asn	Thr	Gly	Ala	Ser	Asp	Arg	Val	Met	Asp
			420					425					430		
Tyr	Gln	Val	Trp	Cys	Gly	Pro	Ala	Ile	Gly	Ser	Phe	Asn	Asp	Phe	Ile
	435						440					445			
Lys	Gly	Thr	Tyr	Leu	Asp	Pro	Ala	Val	Ala	Asn	Glu	Tyr	Pro	Cys	Val
	450					455					460				
Val	Gln	Ile	Asn	Lys	Gln	Ile	Leu	Arg	Gly	Ala	Cys	Phe	Leu	Arg	Arg
465					470					475				480	
Leu	Glu	Ile	Leu	Arg	Asn	Ala	Arg	Leu	Ser	Asp	Gly	Ala	Ala	Ala	Leu
			485						490					495	
Val	Ala	Ser	Ile	Asp	Asp	Thr	Tyr	Val	Pro	Ala	Glu	Lys	Leu		
			500						505				510		

<210>27

<211>4512

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (4512)

<223>

163

195	200	205	
aag ggc gtc gtc ttc acc cgc ggc gac ctc gcc gcc cgc gcc aag atc			672
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile			
210	215	220	
cca aag cag gac gtc tcc ccc tac gcc gtc gcc ccc tgc ctc cac aag			720
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys			
225	230	235	240
acc aag ctc aac gaa aag gag atg cag acc ctc gtc gac aag gac tgg			768
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp			
245	250	255	
gca tcc gtc ttt ggc tcc aag aac ggc atg ccg gaa atc aac tac aaa			816
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys			
260	265	270	
ctc tgc gcg cgt aag atg ctc atg att gac cgc gtc acc agc att gac			864
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp			
275	280	285	
cac aag ggc ggt gtc tac ggc ctc ggt cag ctc gtc ggt gaa aag atc			912
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile			
290	295	300	
ctc gag cgc gac cac tgg tac ttt ccc tgc cac ttt gtc aag gat cag			960
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln			
305	310	315	320
gtc atg gcc gga tcc ctc gtc tcc gac ggc tgc agc cag atg ctc aag			1008
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys			
325	330	335	
atg tac atg atc tgg ctc ggc ctc cac ctc acc acc gga ccc ttt gac			1056
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp			
340	345	350	
ttc cgc ccg gtc aac ggc cac ccc aac aag gtc cgc tgc cgc ggc caa			1104
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln			
355	360	365	
atc tcc ccg cac aag ggc aag ctc gtc tac gtc atg gag atc aag gag			1152
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu			
370	375	380	
atg ggc ttc gac gag gac aac gac ccg tac gcc att gcc gac gtc aac			1200
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn			
385	390	395	400
atc att gat gtc gac ttc gaa aag ggc cag gac ttt agc ctc gac cgc			1248
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg			

405	410	415	
atc agc gac tac ggc aag ggc gac ctc aac aag aag atc gtc gtc gac			1296
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp			
420	425	430	
ttt aag ggc atc gct ctc aag atg cag aag cgc tcc acc aac aag aac			1344
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn			
435	440	445	
ccc tcc aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc			1392
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly			
450	455	460	
ccc gag gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc			1440
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala			
465	470	475	480
gcc ccg gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc			1488
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro			
485	490	495	
gtt gcc ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag			1536
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys			
500	505	510	
gag atg tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc			1584
Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro			
515	520	525	
tct ttt gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc			1632
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro			
530	535	540	
ttc ccc ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg			1680
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro			
545	550	555	560
ctc acc tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg			1728
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met			
565	570	575	
tgc ctc ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc			1776
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg			
580	585	590	
agc ccc gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct			1824
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser			
595	600	605	
gac ctc aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag			1872
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys			

610	615	620	
ggt acc atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac			1920
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr			
625	630	635	640
aag ggc gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag			1968
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu			
645	650	655	
atc gcc ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc			2016
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro			
660	665	670	
ctg acc atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac			2064
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn			
675	680	685	
gcc gag ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc			2112
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg			
690	695	700	
aac gtc acc aag tgc act ggc tac agc atg ctc ggc gag atg ggc gtc			2160
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val			
705	710	715	720
cac cgc ttc acc ttt gag ctc tac gtc gat gat gtg ctc ttt tac aag			2208
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys			
725	730	735	
ggc tcg acc tcg ttc ggc tgg ttc gtg ccc gag gtc ttt gcc gcc cag			2256
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln			
740	745	750	
gcc ggc ctc gac aac ggc cgc aag tcg gag ccc tgg ttc att gag aac			2304
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn			
755	760	765	
aag gtt ccg gcc tcg cag gtc tcc tcc ttt gac gtg cgc ccc aac ggc			2352
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly			
770	775	780	
agc ggc cgc acc gcc atc ttc gcc aac gcc ccc agc ggc gcc cag ctc			2400
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu			
785	790	795	800
aac cgc cgc acg gac cag ggc cag tac ctc gac gcc gtc gac att gtc			2448
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val			
805	810	815	
tcc ggc agc ggc aag aag agc ctc ggc tac gcc cac ggt tcc aag acg			2496
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr			

820	825	830	
gtc aac ccg aac gac tgg ttc ttc tcg tgc cac ttt tgg ttt gac tcg			2544
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser			
835	840	845	
gtc atg ccc gga agt ctc ggt gtc gag tcc atg ttc cag ctc gtc gag			2592
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu			
850	855	860	
gcc atc gcc gcc cac gag gat ctc gct ggc aaa gca cgg cat tgc caa			2640
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln			
865	870	875	880
ccc cac ctt tgt gca cgc ccc cgg gca aga tca agc tgg aag tac cgc			2688
pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg			
885	890	895	
ggc cag ctc acg ccc aag agc aag aag atg gac tcg gag gtc cac atc			2736
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile			
900	905	910	
gtg tcc gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc			2784
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly			
915	920	925	
ttc ctc tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc			2832
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg			
930	935	940	
gtg cgc atc gcc tcc ggt gag gcc cct gcc gcc gcc tcc tcc gcc gcc			2880
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala			
945	950	955	960
tct gtg ggc tcc tcg gct tcg tcc gtc gag cgc acg cgc tcg agc ccc			2928
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro			
965	970	975	
gct gtc gcc tcc ggc ccg gcc cag acc atc gac ctc aag cag ctc aag			2976
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys			
980	985	990	
acc gag ctc ctc gag ctc gat gcc ccg ctc tac ctc tcg cag gac ccg			3024
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro			
995	1000	1005	
acc agc ggc cag ctc aag aag cac acc gac gtg gcc tcc ggc cag			3069
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln			
1010	1015	1020	
gcc acc atc gtg cag ccc tgc acg ctc ggc gac ctc ggt gac cgc			3114
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg			

1025	1030	1035	
tcc ttc atg gag acc tac ggc gtc gtc gcc ccg ctg tac acg ggc			3159
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly			
1040	1045	1050	
gcc atg gcc aag ggc att gcc tcg gcg gac ctc gtc atc gcc gcc			3204
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala			
1055	1060	1065	
ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc ggc ggc ctc ccc			3249
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro			
1070	1075	1080	
atg cac cac gtg cgc gcc gcc ctc gag aag atc cag gcc gcc ctg			3294
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu			
1085	1090	1095	
cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct ttt gac			3339
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp			
1100	1105	1110	
agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag ggc			3384
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly			
1115	1120	1125	
gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag			3429
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln			
1130	1135	1140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt			3474
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly			
1145	1150	1155	
tcg gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc			3519
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr			
1160	1165	1170	
gag ctc gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc			3564
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu			
1175	1180	1185	
gag aag ctc atc gcc tcg ggc gag atc acc cag gag cag gcc gag			3609
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu			
1190	1195	1200	
ctc gcg cgc cgc gtt ccc gtc gcc gac gat atc gct gtc gag gct			3654
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala			
1205	1210	1215	
gac tcg ggc ggc cac acc gac aac cgc ccc atc cac gtc atc ctc			3699
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu			

1220	1225	1230	
ccg ctc atc atc aac ctc cgc aac cgc ctg cac cgc gag tgc ggc			3744
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly			
1235	1240	1245	
tac ccc gcg cac ctc cgc gtc cgc gtt ggc gcc ggc ggt ggc gtc			3789
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val			
1250	1255	1260	
ggc tgc ccg cag gcc gcc gcc gcc gcg ctc acc atg ggc gcc gcc			3834
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala			
1265	1270	1275	
ttc atc gtc acc ggc act gtc aac cag gtc gcc aag cag tcc ggc			3879
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly			
1280	1285	1290	
acc tgc gac aac gtg cgc aag cag ctc tcg cag gcc acc tac tcg			3924
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser			
1295	1300	1305	
gat atc tgc atg gcc ccg gcc gcc gac atg ttc gag gag ggc gtc			3969
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val			
1310	1315	1320	
aag ctc cag gtc ctc aag aag gga acc atg ttc ccc tcg cgc gcc			4014
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala			
1325	1330	1335	
aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc gac tcc			4059
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser			
1340	1345	1350	
atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc aag			4104
Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys			
1355	1360	1365	
cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att			4149
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			
1370	1375	1380	
aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac			4194
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp			
1385	1390	1395	
ccc aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt			4239
Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu			
1400	1405	1410	
gcc agc cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac			4284
Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp			

1415	1420	1425	
tac cag gtc tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc			4329
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe			
1430	1435	1440	
atc aag ggc acc tac ctc gac ccc gct gtc tcc aac gag tac ccc			4374
Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro			
1445	1450	1455	
tgt gtc gtc cag atc aac ctg caa atc ctc cgt ggt gcc tgc tac			4419
Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr			
1460	1465	1470	
ctg cgc cgt ctc aac gcc ctg cgc aac gac ccg cgc att gac ctc			4464
Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu			
1475	1480	1485	
gag acc gag gat gct gcc ttt gtc taG gag ccc acc aac gcg ctc			4509
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu			
1490	1495	1500	
taa			4512

<210>28

<211>1503

<212>PRT

<213>Schizochytrium sp.

<400>28

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr

1 5 10 15

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu

20 25 30

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu

35 40 45

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg

50 55 60

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn

65 70 75 80

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val

85 90 95

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val

100 105 110

Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp

115 120 125

Phe	Gln	Asn	Gln	Gly	Asp	Arg	Val	Tyr	Arg	Leu	Leu	Asn	Thr	Thr	Leu
130						135				140					
Thr	Phe	Tyr	Gly	Val	Ala	His	Glu	Gly	Glu	Thr	Leu	Glu	Tyr	Asp	Ile
145					150					155					160
Arg	Val	Thr	Gly	Phe	Ala	Lys	Arg	Leu	Asp	Gly	Gly	Ile	Ser	Met	Phe
				165					170					175	
Phe	Phe	Glu	Tyr	Asp	Cys	Tyr	Val	Asn	Gly	Arg	Leu	Leu	Ile	Glu	Met
				180					185					190	
Arg	Asp	Gly	Cys	Ala	Gly	Phe	Phe	Thr	Asn	Glu	Glu	Leu	Asp	Ala	Gly
		195					200					205			
Lys	Gly	Val	Val	Phe	Thr	Arg	Gly	Asp	Leu	Ala	Ala	Arg	Ala	Lys	Ile
		210					215					220			
Pro	Lys	Gln	Asp	Val	Ser	Pro	Tyr	Ala	Val	Ala	Pro	Cys	Leu	His	Lys
225					230					235					240
Thr	Lys	Leu	Asn	Glu	Lys	Glu	Met	Gln	Thr	Leu	Val	Asp	Lys	Asp	Trp
				245					250					255	
Ala	Ser	Val	Phe	Gly	Ser	Lys	Asn	Gly	Met	Pro	Glu	Ile	Asn	Tyr	Lys
				260					265					270	
Leu	Cys	Ala	Arg	Lys	Met	Leu	Met	Ile	Asp	Arg	Val	Thr	Ser	Ile	Asp
				275				280						285	
His	Lys	Gly	Gly	Val	Tyr	Gly	Leu	Gly	Gln	Leu	Val	Gly	Glu	Lys	Ile
				290				295						300	
Leu	Glu	Arg	Asp	His	Trp	Tyr	Phe	Pro	Cys	His	Phe	Val	Lys	Asp	Gln
305					310					315					320
Val	Met	Ala	Gly	Ser	Leu	Val	Ser	Asp	Gly	Cys	Ser	Gln	Met	Leu	Lys
				325					330					335	
Met	Tyr	Met	Ile	Trp	Leu	Gly	Leu	His	Leu	Thr	Thr	Gly	Pro	Phe	Asp
				340					345					350	
Phe	Arg	Pro	Val	Asn	Gly	His	Pro	Asn	Lys	Val	Arg	Cys	Arg	Gly	Gln
				355					360					365	
Ile	Ser	Pro	His	Lys	Gly	Lys	Leu	Val	Tyr	Val	Met	Glu	Ile	Lys	Glu
				370					375					380	
Met	Gly	Phe	Asp	Glu	Asp	Asn	Asp	Pro	Tyr	Ala	Ile	Ala	Asp	Val	Asn
385					390					395					400
Ile	Ile	Asp	Val	Asp	Phe	Glu	Lys	Gly	Gln	Asp	Phe	Ser	Leu	Asp	Arg
				405					410					415	
Ile	Ser	Asp	Tyr	Gly	Lys	Gly	Asp	Leu	Asn	Lys	Lys	Ile	Val	Val	Asp
				420					425					430	
Phe	Lys	Gly	Ile	Ala	Leu	Lys	Met	Gln	Lys	Arg	Ser	Thr	Asn	Lys	Asn

435					440					445						
Pro	Ser	Lys	Val	Gln	Pro	Val	Phe	Ala	Asn	Gly	Ala	Ala	Thr	Val	Gly	
450					455					460						
Pro	Glu	Ala	Ser	Lys	Ala	Ser	Ser	Gly	Ala	Ser	Ala	Ser	Ala	Ser	Ala	
465					470					475					480	
Ala	Pro	Ala	Lys	Pro	Ala	Phe	Ser	Ala	Asp	Val	Leu	Ala	Pro	Lys	Pro	
485					490					495						
Val	Ala	Leu	Pro	Glu	His	Ile	Leu	Lys	Gly	Asp	Ala	Leu	Ala	Pro	Lys	
300					505					510						
Glu	Met	Ser	Trp	His	Pro	Met	Ala	Arg	Ile	Pro	Gly	Asn	Pro	Thr	Pro	
515					520					525						
Ser	Phe	Ala	Pro	Ser	Ala	Tyr	Lys	Pro	Arg	Asn	Ile	Ala	Phe	Thr	Pro	
530					535					540						
Phe	Pro	Gly	Asn	Pro	Asn	Asp	Asn	Asp	His	Thr	Pro	Gly	Lys	Met	Pro	
545					550					555					560	
Leu	Thr	Trp	Phe	Asn	Met	Ala	Glu	Phe	Met	Ala	Gly	Lys	Val	Ser	Met	
565					570					575						
Cys	Leu	Gly	Pro	Glu	Phe	Ala	Lys	Phe	Asp	Asp	Ser	Asn	Thr	Ser	Arg	
580					585					590						
Ser	Pro	Ala	Trp	Asp	Leu	Ala	Leu	Val	Thr	Arg	Ala	Val	Ser	Val	Ser	
595					600					605						
Asp	Leu	Lys	His	Val	Asn	Tyr	Arg	Asn	Ile	Asp	Leu	Asp	Pro	Ser	Lys	
610					615					620						
Gly	Thr	Met	Val	Gly	Glu	Phe	Asp	Cys	Pro	Ala	As	pAla	Trp	Phe	Tyr	
625					630					635					640	
Lys	Gly	Ala	Cys	Asn	Asp	Ala	His	Met	Pro	Tyr	Ser	Ile	Leu	Met	Glu	
645					650					655						
Ile	Ala	Leu	Gln	Thr	Ser	Gly	Val	Leu	Thr	Ser	Val	Leu	Lys	Ala	Pro	
660					665					670						
Leu	Thr	Met	Glu	Lys	Asp	Asp	Ile	Leu	Phe	Arg	Asn	Leu	Asp	Ala	Asn	
675					680					685						
Ala	Glu	Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg	
690					695					700						
Asn	Val	Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val	
705					710					715					720	
His	Arg	Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys	
725					730					735						
Gly	Ser	Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln	
740					745					750						

Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn		
755	760	765
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly		
770	775	780
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu		
785	790	795
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val		
805	810	815
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr		
820	825	830
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser		
835	840	845
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu		
850	855	860
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln		
865	870	875
Pro His Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg		
885	890	895
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile		
900	905	910
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly		
915	920	925
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg		
930	935	940
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala		
945	950	955
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro		
965	970	975
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys		
980	985	990
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro		
995	1000	1005
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln		
1010	1015	1020
Ala Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg		
1025	1030	1035
Ser Phe Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly		
1040	1045	1050
Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala		

1055	1060	1065
Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro		
1070	1075	1080
Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu		
1085	1090	1095
Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp		
1100	1105	1110
Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly		
1115	1120	1125
Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln		
1130	1135	1140
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly		
1145	1150	1155
Ser Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr		
1160	1165	1170
Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu		
1175	1180	1185
Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu		
1190	1195	1200
Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala		
1205	1210	1215
Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile Leu		
1220	1225	1230
Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly		
1235	1240	1245
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val		
1250	1255	1260
Gly Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala		
1265	1270	1275
Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly		
1280	1285	1290
Thr Cys Asp Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser		
1295	1300	1305
Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val		
1310	1315	1320
Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala		
1325	1330	1335
Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Asp Ser		
1340	1345	1350

Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe Lys		
1355	1360	1365
Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile		
1370	1375	1380
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp		
1385	1390	1395
Pro Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu		
1400	1405	1410
Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp		
1415	1420	1425
Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe		
1430	1435	1440
Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro		
1445	1450	1455
Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr		
1460	1465	1470
Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu		
1475	1480	1485
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu		
1490	1495	1500

<210>29

<211>1500

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1).. (1500)

<223>

<400>29

aag gtt cag ccc gtc ttt gcc aac ggc gcc gcc act gtc ggc ccc gag	48
Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu	
1 5 10 15	
gcc tcc aag gct tcc tcc ggc gcc agc gcc agc gcc agc gcc gcc ccg	96
Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro	
20 25 30	
gcc aag cct gcc ttc agc gcc gat gtt ctt gcg ccc aag ccc gtt gcc	144

Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala	
35 40 45	
ctt ccc gag cac atc ctc aag ggc gac gcc ctc gcc ccc aag gag atg	192
Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met	
50 55 60	
tcc tgg cac ccc atg gcc cgc atc ccg ggc aac ccg acg ccc tct ttt	240
Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe	
65 70 75 80	
gcg ccc tcg gcc tac aag ccg cgc aac atc gcc ttt acg ccc ttc ccc	288
Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro	
85 90 95	
ggc aac ccc aac gat aac gac cac acc ccg ggc aag atg ccg ctc acc	336
Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr	
100 105 110	
tgg ttc aac atg gcc gag ttc atg gcc ggc aag gtc agc atg tgc ctc	384
Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu	
115 120 125	
ggc ccc gag ttc gcc aag ttc gac gac tcg aac acc agc cgc agc ccc	432
Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro	
130 135 140	
gct tgg gac ctc gct ctc gtc acc cgc gcc gtg tct gtg tct gac ctc	480
Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu	
145 150 155 160	
aag cac gtc aac tac cgc aac atc gac ctc gac ccc tcc aag ggt acc	528
Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr	
165 170 175	
atg gtc ggc gag ttc gac tgc ccc gcg gac gcc tgg ttc tac aag ggc	576
Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly	
180 185 190	
gcc tgc aac gat gcc cac atg ccg tac tcg atc ctc atg gag atc gcc	624
Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala	
195 200 205	
ctc cag acc tcg ggt gtg ctc acc tcg gtg ctc aag gcg ccc ctg acc	672
Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr	
210 215 220	
atg gag aag gac gac atc ctc ttc cgc aac ctc gac gcc aac gcc gag	720
Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu	
225 230 235 240	
ttc gtg cgc gcc gac ctc gac tac cgc ggc aag act atc cgc aac gtc	768

Phe	Val	Arg	Ala	Asp	Leu	Asp	Tyr	Arg	Gly	Lys	Thr	Ile	Arg	Asn	Val		
				245				250				255					
acc	aag	tgc	act	ggc	tac	agc	atg	ctc	ggc	gag	atg	ggc	gtc	cac	cgc	816	
Thr	Lys	Cys	Thr	Gly	Tyr	Ser	Met	Leu	Gly	Glu	Met	Gly	Val	His	Arg		
				260				265				270					
ttc	acc	ttt	gag	ctc	tac	gtc	gat	gat	gtg	ctc	ttt	tac	aag	ggc	tcg	864	
Phe	Thr	Phe	Glu	Leu	Tyr	Val	Asp	Asp	Val	Leu	Phe	Tyr	Lys	Gly	Ser		
				275				280				285					
acc	tcg	ttc	ggc	tgg	ttc	gtg	ccc	gag	gtc	ttt	gcc	gcc	cag	gcc	ggc	912	
Thr	Ser	Phe	Gly	Trp	Phe	Val	Pro	Glu	Val	Phe	Ala	Ala	Gln	Ala	Gly		
				290				295				300					
ctc	gac	aac	ggc	cgc	aag	tcg	gag	ccc	tgg	ttc	att	gag	aac	aag	gtt	960	
Leu	Asp	Asn	Gly	Arg	Lys	Ser	Glu	Pro	Trp	Phe	Ile	Glu	Asn	Lys	Val		
				305				310				315				320	
ccg	gcc	tcg	cag	gtc	tcc	tcc	ttt	gac	gtg	cgc	ccc	aac	ggc	agc	ggc	1008	
Pro	Ala	Ser	Gln	Val	Ser	Ser	Phe	Asp	Val	Arg	Pro	Asn	Gly	Ser	Gly		
				325				330				335					
cgc	acc	gcc	atc	ttc	gcc	aac	gcc	ccc	agc	ggc	gcc	cag	ctc	aac	cgc	1056	
Arg	Thr	Ala	Ile	Phe	Ala	Asn	Ala	Pro	Ser	Gly	Ala	Gln	Leu	Asn	Arg		
				340				345				350					
cgc	acg	gac	cag	ggc	cag	tac	ctc	gac	gcc	gtc	gac	att	gtc	tcc	ggc	1104	
Arg	Thr	Asp	Gln	Gly	Gln	Tyr	Leu	Asp	Ala	Val	Asp	Ile	Val	Ser	Gly		
				355				360				365					
agc	ggc	aag	aag	agc	ctc	ggc	tac	gcc	cac	ggt	tcc	aag	acg	gtc	aac	1152	
Ser	Gly	Lys	Lys	Ser	Leu	Gly	Tyr	Ala	His	Gly	Ser	Lys	Thr	Val	Asn		
				370				375				380					
ccg	aac	gac	tgg	ttc	ttc	tcg	tgc	cac	ttt	tgg	ttt	gac	tcg	gtc	atg	1200	
Pro	Asn	Asp	Trp	Phe	Phe	Ser	Cys	His	Phe	Trp	Phe	Asp	Ser	Val	Met		
				385				390				395				400	
ccc	gga	agt	ctc	ggt	gtc	gag	tcc	atg	ttc	cag	ctc	gtc	gag	gcc	atc	1248	
Pro	Gly	Ser	Leu	Gly	Val	Glu	Ser	Met	Phc	Gln	Leu	Val	Glu	Ala	Ile		
				405				410				415					
gcc	gcc	cac	gag	gat	ctc	gct	ggc	aaa	gca	cgg	cat	tgc	caa	ccc	cac	1296	
Ala	Ala	His	Glu	Asp	Leu	Ala	Gly	Lys	Ala	Arg	His	Cys	Gln	Pro	His		
				420				425				430					
ctt	tgt	gca	cgc	ccc	cgg	gca	aga	tca	agc	tgg	aag	tac	cgc	ggc	cag	1344	
Leu	Cys	Ala	Arg	Pro	Arg	Ala	Arg	Ser	Ser	Trp	Lys	Tyr	Arg	Gly	Gln		
				435				440				445					
ctc	acg	ccc	aag	agc	aag	aag	atg	gac	tcg	gag	gtc	cac	atc	gtg	tcc	1392	

Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser
 450 455 460
 gtg gac gcc cac gac ggc gtt gtc gac ctc gtc gcc gac ggc ttc ctc 1440
 Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu
 465 470 475 480
 tgg gcc gac agc ctc cgc gtc tac tcg gtg agc aac att cgc gtg cgc 1488
 Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg
 485 490 495
 atc gcc tcc ggt 1500
 Ile Ala Ser Gly
 500

<210>30

<211>500

<212>PRT

<213>Schizochytrium sp.

<400>30

Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly Pro Glu
 1 5 10 15
 Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala Ala Pro
 20 25 30
 Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro Val Ala
 35 40 45
 Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys Glu Met
 50 55 60
 Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro Ser Phe
 65 70 75 80
 Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro Phe Pro
 85 90 95
 Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro Leu Thr
 100 105 110
 Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met Cys Leu
 115 120 125
 Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg Ser Pro
 130 135 140
 Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser Asp Leu
 145 150 155 160
 Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys Gly Thr

	165		170		175
Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr Lys Gly					
	180		185		190
Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu Ile Ala					
	195		200		205
Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro Leu Thr					
	210		215		220
Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn Ala Glu					
225		230		235	240
Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg Asn Val					
	245		250		255
Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val His Arg					
	260		265		270
Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys Gly Ser					
	275		280		285
Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln Ala Gly					
	290		295		300
Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn Lys Val					
305		310		315	320
Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly Ser Gly					
	325		330		335
Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu Asn Arg					
	340		345		350
Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val Ser Gly					
	355		360		365
Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr Val Asn					
	370		375		380
Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser Val Met					
385		390		395	400
Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu Ala Ile					
	405		410		415
Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln Pro His					
	420		425		430
Leu Cys Ala Arg Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg Gly Gln					
	435		440		445
Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile Val Ser					
	450		455		460
Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly Phe Leu					
465		470		475	480

Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg Val Arg
 485 490 495
 Ile Ala Ser Gly
 500

<210>31

<211>1512

<212>DNA

<213>Schizochytrium sp.

<220>

<221>CDS

<222>(1)..(1512)

<223>

<400>31

gcc ccg ctc tac ctc tcg cag gac ccg acc agc ggc cag ctc aag aag	48
Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys	
1 5 10 15	
cac acc gac gtg gcc tcc ggc cag gcc acc atc gtg cag ccc tgc acg	96
His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr	
20 25 30	
ctc ggc gac ctc ggt gac cgc tcc ttc atg gag acc tac ggc gtc gtc	144
Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val	
35 40 45	
gcc ccg ctg tac acg ggc gcc atg gcc aag ggc att gcc tcg gcg gac	192
Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp	
50 55 60	
ctc gtc atc gcc gcc ggc aag cgc aag atc ctc ggc tcc ttt ggc gcc	240
Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala	
65 70 75 80	
ggc ggc ctc ccc atg cac cac gtg cgc gcc gcc ctc gag aag atc cag	288
Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln	
85 90 95	
gcc gcc ctg cct cag ggc ccc tac gcc gtc aac ctc atc cac tcg cct	336
Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro	
100 105 110	
ttt gac agc aac ctc gag aag ggc aac gtc gat ctc ttc ctc gag aag	384
Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys	

115	120	125	
ggc gtc act gtg gtg gag gcc tcg gca ttc atg acc ctc acc ccg cag			432
Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln			
130	135	140	
gtc gtg cgc tac cgc gcc gcc ggc ctc tcg cgc aac gcc gac ggt tcg			480
Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser			
145	150	155	160
gtc aac atc cgc aac cgc atc atc ggc aag gtc tcg cgc acc gag ctc			528
Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu			
165	170	175	
gcc gag atg ttc atc cgc ccg gcc ccg gag cac ctc ctc gag aag ctc			576
Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu			
180	185	190	
atc gcc tcg ggc gag atc acc cag gag cag gcc gag ctc gcg cgc cgc			624
Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg			
195	200	205	
gtt ccc gtc gcc gac gat atc gct gtc gag gct gac tcg ggc ggc cac			672
Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His			
210	215	220	
acc gac aac cgc ccc atc cac gtc atc ctc ccg ctc atc atc aac ctc			720
Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu			
225	230	235	240
cgc aac cgc ctg cac cgc gag tgc ggc tac ccc gcg cac ctc cgc gtc			768
Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val			
245	250	255	
cgc gtt ggc gcc ggc ggt ggc gtc ggc tgc ccg cag gcc gcc gcc gcc			816
Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala			
260	265	270	
gcg ctc acc atg ggc gcc gcc ttc atc gtc acc ggc act gtc aac cag			864
Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln			
275	280	285	
gtc gcc aag cag tcc ggc acc tgc gac aac gtg cgc aag cag ctc tcg			912
Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser			
290	295	300	
cag gcc acc tac tcg gat atc tgc atg gcc ccg gcc gcc gac atg ttc			960
Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe			
305	310	315	320
gag gag ggc gtc aag ctc cag gtc ctc aag aag gga acc atg ttc ccc			1008
Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro			

325	330	335	
tcg cgc gcc aac aag ctc tac gag ctc ttt tgc aag tac gac tcc ttc			1056
Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe			
340	345	350	
gac tcc atg cct cct gcc gag ctc gag cgc atc gag aag cgt atc ttc			1104
Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe			
355	360	365	
aag cgc gca ctc cag gag gtc tgg gag gag acc aag gac ttt tac att			1152
Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			
370	375	380	
aac ggt ctc aag aac ccg gag aag atc cag cgc gcc gag cac gac ccc			1200
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro			
385	390	395	400
aag ctc aag atg tcg ctc tgc ttc cgc tgg tac ctt ggt ctt gcc agc			1248
Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser			
405	410	415	
cgc tgg gcc aac atg ggc gcc ccg gac cgc gtc atg gac tac cag gtc			1296
Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val			
420	425	430	
tgg tgt ggc ccg gcc att ggc gcc ttc aac gac ttc atc aag ggc acc			1344
Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr			
435	440	445	
tac ctc gac ccc gct gtc tcc aac gag tac ccc tgt gtc gtc cag atc			1392
Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile			
450	455	460	
aac ctg caa atc ctc cgt ggt gcc tgc tac ctg cgc cgt ctc aac gcc			1440
Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala			
465	470	475	480
ctg cgc aac gac ccg cgc att gac ctc gag acc gag gat gct gcc ttt			1488
Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe			
485	490	495	
gtc tac gag ccc acc aac gcg ctc			1512
Val Tyr Glu Pro Thr Asn Ala Leu			
500			

<210>32

<211>504

<212>PRT

<213>Schizochytrium sp.

<400>32

Ala Pro Leu Tyr Leu Ser Gln Asp Pro Thr Ser Gly Gln Leu Lys Lys
 1 5 10 15
 His Thr Asp Val Ala Ser Gly Gln Ala Thr Ile Val Gln Pro Cys Thr
 20 25 30
 Leu Gly Asp Leu Gly Asp Arg Ser Phe Met Glu Thr Tyr Gly Val Val
 35 40 45
 Ala Pro Leu Tyr Thr Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp
 50 55 60
 Leu Val Ile Ala Ala Gly Lys Arg Lys Ile Leu Gly Ser Phe Gly Ala
 65 70 75 80
 Gly Gly Leu Pro Met His His Val Arg Ala Ala Leu Glu Lys Ile Gln
 85 90 95
 Ala Ala Leu Pro Gln Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro
 100 105 110
 Phe Asp Ser Asn Leu Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys
 115 120 125
 Gly Val Thr Val Val Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln
 130 135 140
 Val Val Arg Tyr Arg Ala Ala Gly Leu Ser Arg Asn Ala Asp Gly Ser
 145 150 155 160
 Val Asn Ile Arg Asn Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu
 165 170 175
 Ala Glu Met Phe Ile Arg Pro Ala Pro Glu His Leu Leu Glu Lys Leu
 180 185 190
 Ile Ala Ser Gly Glu Ile Thr Gln Glu Gln Ala Glu Leu Ala Arg Arg
 195 200 205
 Val Pro Val Ala Asp Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His
 210 215 220
 Thr Asp Asn Arg Pro Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu
 225 230 235 240
 Arg Asn Arg Leu His Arg Glu Cys Gly Tyr Pro Ala His Leu Arg Val
 245 250 255
 Arg Val Gly Ala Gly Gly Gly Val Gly Cys Pro Gln Ala Ala Ala Ala
 260 265 270
 Ala Leu Thr Met Gly Ala Ala Phe Ile Val Thr Gly Thr Val Asn Gln
 275 280 285
 Val Ala Lys Gln Ser Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ser

290	295	300	
Gln Ala Thr Tyr Ser Asp Ile Cys Met Ala Pro Ala Ala Asp Met Phe			
305	310	315	320
Glu Glu Gly Val Lys Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro			
	325	330	335
Ser Arg Ala Asn Lys Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe			
	340	345	350
Asp Ser Met Pro Pro Ala Glu Leu Glu Arg Ile Glu Lys Arg Ile Phe			
	355	360	365
Lys Arg Ala Leu Gln Glu Val Trp Glu Glu Thr Lys Asp Phe Tyr Ile			
	370	375	380
Asn Gly Leu Lys Asn Pro Glu Lys Ile Gln Arg Ala Glu His Asp Pro			
385	390	395	400
Lys Leu Lys Met Ser Leu Cys Phe Arg Trp Tyr Leu Gly Leu Ala Ser			
	405	410	415
Arg Trp Ala Asn Met Gly Ala Pro Asp Arg Val Met Asp Tyr Gln Val			
	420	425	430
Trp Cys Gly Pro Ala Ile Gly Ala Phe Asn Asp Phe Ile Lys Gly Thr			
	435	440	445
Tyr Leu Asp Pro Ala Val Ser Asn Glu Tyr Pro Cys Val Val Gln Ile			
	450	455	460
Asn Leu Gln Ile Leu Arg Gly Ala Cys Tyr Leu Arg Arg Leu Asn Ala			
465	470	475	480
Leu Arg Asn Asp Pro Arg Ile Asp Leu Glu Thr Glu Asp Ala Ala Phe			
	485	490	495
Val Tyr Glu Pro Thr Asn Ala Leu			
	500		

<210>33

<211>9

<212>PRT

<213> 人工序列

<220>

<223>motif

<220>

<221>MISC_FEATURE

<222>(2).. (3)

<223>x = 任意氨基酸

<220>

<221>MISC_FEATURE

<222>(6).. (6)

<223>x = A 或 S

<220>

<221>MISC_FEATURE

<222>(7).. (8)

<223>x = 任意氨基酸

<400>33

Trp Xaa Xaa Lys Glu Xaa Xaa Xaa Lys

1 5

<210>34

<211>6

<212>PRT

<213> 人工序列

<220>

<223>motif

<220>

<221>MISC_FEATURE

<222>(3).. (3)

<223>x = I 或 L 或 V

<400>34

Phe Asn Xaa Ser His Ser

1 5

<210>35

<211>5

<212>PRT

<213> 人工序列

<220>

<223>motif

<220>

<221>MISC_FEATURE

<222>(1).. (5)

<223>x = I 或 L 或 V

<400>35

Xaa Gly Xaa Asp Xaa

1

5

<210>36

<211>4244

<212>DNA

<213>Schizochytrium sp.

<400>36

tttctctctc	tcgagctggt	gctgctgctg	ctgctgctgc	tgcttccttg	ctggttctca	60
cgtccgttcg	atcaagecgt	cgctcgctcg	accgatcggt	gcgtgcgtgc	gtgcgtgagt	120
cttgttgcc	ggcagccgca	ggctgtctgt	ctgtttgtgt	agttttaccc	tcgggggttcg	180
gggtctgcct	gcctcccgct	cccgcgccgc	gccgcgccgt	tccaccccgc	tcgcctccgc	240
ccatcgggcc	tcgcctctc	gcgcgcacg	catcgcgcg	atcgcatgca	tcatgctgcc	300
acgcacgggg	ggacgcgcgc	cccgcgtccc	ccgcgcgcgc	cgctcgctgc	tggcgatgcc	360
gtcgccgccc	tccttccttc	cctcgccctc	tcttcctccc	gagccccctt	gtcttccttc	420
gccccgcag	cggcgcgag	gaagcgagga	gagcggggag	gagagaagaa	aagaaaagaa	480
aagaaaagaa	aataacagcg	ccgtctcgcg	cagacgcgcg	cggccgcgtg	cgaggcggcg	540
tgatggggct	tctcgtggcg	cggctgcggc	ctggcccggc	ctcgcccttg	aggtgcaggc	600
tttgggagag	aagagtggga	cgcggagaag	ataagatggt	gccatggcgc	aggacggaga	660
ggttgctgaa	acttcttcga	gcggcacagg	cgatggcgag	agaccgacag	ctgccggcgc	720
ggaggggatg	gatacctccc	gaggctggca	tggacgagct	ggccgcgcgg	atctggctgg	780
ccgcgcggcg	gtgggtccgg	aggcgcgagg	tgggttttct	tcatacctga	taccatacgg	840
tattcattct	tcctctccag	gaaggaagca	agtcacatag	agtatcacta	gcctaattgat	900
ggactctatg	ttttagggca	cgtcggagca	gaaggcgcga	gcgattcgaa	tgcgagcgat	960
agatacagca	cagagacctt	gccggcgacg	cggatgcagg	cgagcacgca	cgcaccgcac	1020
gcacggcagc	ggtgcacgcg	ctcctcgga	gatgcacggt	tctgcgccgc	gcctttacat	1080
tttttgattt	taggtgggtg	gcctgccact	ttgaacatca	tccacaagtc	aacgcagcat	1140
caagaggcaa	gcaagtacat	acatccattc	gaattcaagt	tcaagagacg	cagcaacagc	1200
cgccgctccg	ctcaagctgc	agctagctgg	ctgacagggc	tcgctggctg	tagtggaaaa	1260
ttccattcac	ttttctgcat	ccgcggccag	caggcccgtg	cgcacgttct	ctcgtttggt	1320

tgttcgttcg	tgcgtgcgtg	cgtgcgtccc	agctgcctgt	ctaactctgcc	gcgcgatcca	1380
acgaccctcg	gtcgtgcceg	caagcgaaac	ccgacgccga	cctggccaat	gccgcaagaa	1440
tgctaagcgc	gcagcaatgc	tgagagtaat	cttcagccca	ccaagtcatt	atcgtgccc	1500
aagtctccat	cgcagccaca	ttcaggtttt	ctctctctct	ccctccctct	ctttctgccg	1560
ggagagaagg	aaagacccgc	cgcgcgcgc	tctgcgcctg	tgacgggctg	tccgttgtaa	1620
gccctcttag	acagtcccta	ggtgccgggc	gccgcgcgc	ctccgtcgca	ggcacacgta	1680
ggcggccacg	ggttcccccc	gcacctcca	cacctcttc	ccccgcagcc	ggaccgcgcg	1740
ccgtctgctt	acgcacttcg	cgcggcgcgc	gcccgcgaac	ccgagcgcgt	gctgtgggcg	1800
ccgtcttccg	gccgcgtcgg	aggtcgtccc	cgcgcgcgc	tactccgggt	cctgtgcggt	1860
acgtacttaa	tattaacagt	gggacctcgc	acaggacctg	acggcagcac	agacgtcgcc	1920
gcctcgcata	gctggggacg	caggcgaggc	atcccggcgc	ggccccgcac	cggggaggct	1980
gcggggcggc	ctcttcceg	cggcgccgc	atcaggcgga	tgacgcaaga	gccctcgcag	2040
tcgctcgctc	gcgggagcgc	agcgcggcgc	cagcgtggcc	aagctcccgc	cccttctggc	2100
tggttgcata	cctgcctgcc	tgcctgcctg	cgtgcgtgcg	tgctgcgtg	ccttcgtgcg	2160
tgcctgcctt	cgtgcgtgcg	tgctgagtg	cggcggaaga	gggatcatgc	gaggatcaat	2220
cacccgcgc	acctcgactt	ttgaagaagc	cgcgatgcga	tgcgatgcga	tgcgatgcga	2280
cgcgataccg	tcgagggcta	cgaagcgagt	ctggccggcc	gtcatacaac	gcacgttttc	2340
gagaaggagg	gctggcggag	gcgtgcatgc	cggcgacctat	tgcaaacgcg	gcgtctcgtg	2400
gctggcgaag	gtgcctggag	gatctaacga	tcgtgctat	gatgctatag	ctgtgctgat	2460
ccccggtcca	ttccaccacg	tctgtgcctg	ccgcctgacc	tgcgcttggc	tttccttcaa	2520
gttctcctcc	gccgggcctt	caggaccgag	acgagacctg	cagctgcagc	tagactcgcg	2580
ctcgctcgcg	gaggattcgc	cggccgcggg	gccggacggg	actcgcgagg	tcacacggcc	2640
gccggcgatc	gcgatggctg	tgctgacgta	ctcgctcgctg	gcagccgtac	gtcagcgacg	2700
ccgcctccgt	attgtggatt	cgttagttgg	ttgttgggtg	atttgttgat	taattttttt	2760
gttcgtaggc	ttggttatag	ctaatagttt	agtttatact	ggtgctcttc	ggtgctgatt	2820
tagctcgact	tgggtccaca	ccactgcccc	tctactgtga	atggatcaat	ggacgcacga	2880
cgggcccagc	aaagtgcgcg	agtgaggtaa	cctaagcaac	ggcggctctc	agaggggacg	2940
cacgccctcc	gtcgcagtca	gtccagacag	gcagaaaagc	gtcttaggga	ccacgcacgc	3000
acgcacgcac	gcacgcacgc	ccgcacgcac	gctccctccc	tcgcgtgcct	atttttttag	3060
gcttccttcc	gcacgggcct	acctctcgct	ccctcgctc	gccgcaccag	gcggcagcag	3120
cgataacctg	cgggtccgcc	tcgctacgc	gctcagccgc	agctcagccc	agccgcgagc	3180
tagggtttgt	tcgtcctgaa	ttgtttgatt	tgatttgatt	tgatttgatc	cgatccgatc	3240
cgatctgate	tgatttgctt	tgctttgctt	tgctctccctc	ccggcgcgga	ccaagcgtcc	3300
gtctgcgcgc	cgcagcttcc	cttctctctc	cagccctcct	tctgctccccg	cctctcgcgc	3360
aagcacgcag	cttcgcgcgc	gcataccggtc	ggtcggtcgg	tcgatcgacc	cgcctgcgcg	3420
tgctgctgtg	gccgggcttt	tctccatcgg	cgactctttc	ttctccatac	gtcctaactac	3480
gtacatacat	actgcgggct	tctctctctt	ccagcgcggc	gacggcgga	ggctgcgacg	3540
tcgtcgccgc	cgcgggcgcgc	gcgcgcgcgc	ccgccgcgcgc	ccgcgtcgca	ggcctcgtc	3600
gccgcgcgcg	ctccgctccg	ctccgaggcc	gcgagagggc	cgcggcggcg	cgatggatgg	3660

```

atggatggat ggatggatgg atggattttg ttgatcgatg gcggcgcatg ggcggagatg 3720
agcgaggacg agcgcgcgag cgcggcagcc ggattcgcag ggcctcgctc gcctcgcgcc 3780
cgctgccgcg cccgccttgc gagcctgcgc cgcgagcgag cgagcgagcg agcggggctt 3840
tctttgtctc gcgcgcgcgt tggcctcgctg tgtcttgcgc ttgcgtagcg ggcgccgcgg 3900
tggaagatgg ctcatccaat cgacccattc acgcacgcac tccggcgcg c agagaaggcc 3960
gaggaggagc agcaagcaaa ccaaaagctc tcgcgctcgc ggtctcgggc tcgagcggtc 4020
tcggagagag agtcttgcgg cgaccaccgg cagcagcagc agcagcagca gcgctgtcga 4080
gcacgagcac gagcagcagc acgagcacga gcattcgagc aagaggacag acacggttgt 4140
cagcgcctag ctgcctcgat acagaaagag gcgggttggg cgtaaaaaaa aaggagcacg 4200
caagccgcca gccagccagc tagctagcca gcctgcctgc caaa 4244

```

<210>37

<211>3886

<212>DNA

<213>Schizochytrium sp.

<220>

<221>misc_feature

<222>(2115).. (2115)

<223>n = a, c, g, 或 t

<400>37

```

gatcttgatt gccaaagctct ggattgtcga ttccgatgaa tcgagctctt tgttgtcgag 60
ctctggcttg ccgagctttc agaaatagac aaaattgccg agttcctgat tcgggggctc 120
tcgattgcca aggtctggtg gattctcgaa ctctcgattg tcaaaatctt ggtcgtctcg 180
tcggattctt tcctgatttg ttttgtcaag accttgagat tgtgcaaaac cttgatcggt 240
gacaaaccct tgatcgacag cagcctttca tcacgctcag ctcttgtcat tgattatatt 300
ccccctgaca gccaacacct tgatgcaggg tctcaacctt gatTTTTTgga ggccatcatc 360
agcatcacgc cccggcactc accctcaaca ttcgacagcc aacgcttttt tttcttcgac 420
taggatctga gaataaaagc aggtcaccac gaccgtaggc caacgcgaca accatggaaa 480
taaagtgaca acgaacgact tgcaagttta aatgtaaaga gcagcaattg cccgcccaca 540
gacaaatgaa agcaggcgcc gagtcttatt tgaggagggtg ggcctgtggc aatgggcgaa 600
agaaaatcaa ggacaaggag agcaggttac gtaccggtat actggtatac gtacatggat 660
ggttcttggc aagttgacgg gatgtgtgcg agtgaccgtg gtagttaacg aaagagccgc 720
aagggaagg aaagcaagag aatgcagact ttccacagg atggatgggt ccgcagcttg 780
ccgcatgatg aaacgctgta ttccacctgg cacgtgggtg cgcacgcgcc cacatatgat 840
cgcgcgggcg ggtgtattat acattttccc cctcaggctt actgccatcc ctccatgcgt 900
cgctcgtgcg aacgacgcaa gcctttcgca tcgtgcagcc tctttctggg aaggcaagag 960
ctaaacccaa acctaaacga aagaacattt ttacctctct ctctctccca ttggtcgcgt 1020

```

gcgctccgcc	gctcgctcct	cctcctgcc	gtgtcgcgcc	ctaacttccc	ccctccctcc	1080
ctccctccct	ccctccctct	ctcctgccac	cgccccctctc	tccgcgctgc	gtgcggtgct	1140
gccctggacc	aatggcatgc	tgttgcacgc	tcggcgcatg	acgcaagccg	cttcgcaatt	1200
tccgatcag	atctcggcgg	ggcgtgcgcc	gcgggggtcac	tgcggacctg	ccgcggcccc	1260
tgtttctttc	acatccatca	tgtcctccaa	acctccgcct	cctccacgca	cgtagcgacg	1320
cccgtcgca	cgcgcgcact	gccgtgcga	aagcaagcgc	ccgcccgcgc	cccggcgacg	1380
ggaaggcggc	cgcggtctcc	ctccgcggtt	gcctcgctcc	cgcgcggggc	tgggcgggca	1440
gcagaaggcg	ggtggcgggc	gcggttccg	tcttcgtcag	cggcctacgt	cggcgggcgc	1500
gcgcgagact	acgcatgccc	ttgcgtcatg	cgctcgacag	tagccgccgc	gggcctagcg	1560
tttccgctgg	cgccgcgcct	aagcccccg	cgcgcacggt	attgccgcga	taccgtacgg	1620
ccaagaccgc	cgcagacgtc	ggccctctcg	cggccagcca	gccagcagcg	cagcggagga	1680
agagcgcgca	ggcgcggcgg	gagggcgggc	gcggagcagc	gcagagcggg	gcggagcagc	1740
gcggagcaga	acgggcagac	tcggagcggg	cagggcgggc	agagctttgg	ggtttaagga	1800
ccgggttacc	ggcgaagtga	gcggttgcgg	ggagcggctg	tgggaggggt	gagtacgcaa	1860
gcacgatgcg	agcgagagag	agacgtgcc	gcgaatcaag	aaggtaggcg	cgctgcgagg	1920
cgcggcgggc	gagcggagcg	agggagaggg	agagggagag	agagggaggg	agacgtcgcc	1980
gcggcggggc	ctggcctggc	ctggtttggc	ttggtcagcg	cggccttgct	cgagcgtgca	2040
gctggagtgg	ggtggattca	tttggatttt	cttttgtttt	tgtttttctc	tctttcccgg	2100
aaagtgttgg	ccggnccgtg	ttctttgttt	tgattttctc	aaaagtgttg	gtggttggtt	2160
ctctctcttg	gctctctgtc	aggcggctcc	gtccacgccc	cggcctctcc	tctcctctcc	2220
tctcctctcc	tctcgtgcg	tatacgtacg	tacgtttgta	tacgtacata	catcccgcgc	2280
gccgtgccgg	cgagggtttg	ctcagcctgg	agcaatgcga	tgcgatgcga	tgcgatgcga	2340
cgcgacgcga	cgcgagtcac	tggttcgcgc	tgtggctgtg	gcttgcttgc	ttacttgctt	2400
tcgagctctc	ccgctttctt	ctttccttct	cacgccacca	ccaacgaaag	aagatcggcc	2460
ccggcacgcc	gctgagaagg	gctggcgggc	atgacggcac	gcgcgcccgc	tgccacgttg	2520
gcgctcgctg	ctgctgctgc	tgtgtgctgt	gctgtgctgt	ctgctgctgc	tgtgtcttct	2580
gcgcgcaggc	tttggccacga	ggccggcggtg	ctggccgctg	ccgcttccag	tccgcgtgga	2640
gagatcgaat	gagagataaa	ctggatggat	tcatcgaggg	atgaatgaac	gatggttggga	2700
tgcctttttc	ctttttcagg	tccacagcgg	gaagcaggag	cgcgtgaatc	tgccgccatc	2760
cgcatacgtc	tgcategcat	cgcategcat	gcacgcacgc	ctcgccggga	gccacagacg	2820
ggcgacaggg	cggccagcca	gccaggcagc	cagccaggca	ggcaccagag	ggccagagag	2880
cgcgcctcac	gcacgcgcgc	cagtgcgcgc	atcgctcgca	gtgcagacct	tgattccccg	2940
cgcggatctc	cgcgagcccg	aaacgaagag	cgcggtacgg	gcccattccta	gcgtcgcttc	3000
gcaccgcate	gcategcatc	gcgttcccta	gagagtagta	ctcgacgaag	gcaccatttc	3060
cgcgtctctc	ttcggcgcgga	tcgaggcccc	cggcgccgcg	acgatcgcgg	cggccgcggc	3120
gctggcgggc	gccctggcgc	tcgcgctggc	ggccgcccgc	ggcgtctggc	cctggcgcgc	3180
gcgggcgccc	caggaggagc	ggcagcggct	gctcgccgcc	agagaagagc	gcgccggggc	3240
cggggaggga	cggggaggag	aaggagaagg	cgcgcaaggc	ggccccgaaa	gagaagaccc	3300
tggacttgaa	cgcgaagaag	aagaagaagg	agaagaagtt	gaagaagaag	aagaagaagg	3360

agaggaagtt gaagaagacg aggagcaggc gcgttccaag gcgcgttctc ttccggaggc	3420
gcgttccagc tgcggcgggcg gggcgggctg cggggcgggc gcgggcgcgg gtgcgggcag	3480
aggggacgcg cgcgcggagg cggagggggc cgagcgggag cccctgctgc tgcggggcgc	3540
ccgggccgca ggtgtggcgc gcgcgacgac ggaggcgacg acgccagcgg ccgcgacgac	3600
aaggccggcg gcgtcggcgg gcggaaggcc ccgcgcggag caggggcggg agcaggacaa	3660
ggcgcaggag caggagcagg gccgggagcg ggagcgggag cgggcggcgg agcccaggc	3720
agaaccaat cgagatccag agcgagcaga ggccggccgc gagcccgagc ccgcgccgca	3780
gactactagt accgctgcgg aatcacagca gcagcagcag cagcagcagc agcagcagca	3840
gcagcagcag ccacgagagg gagataaaga aaaagcggca gagacg	3886

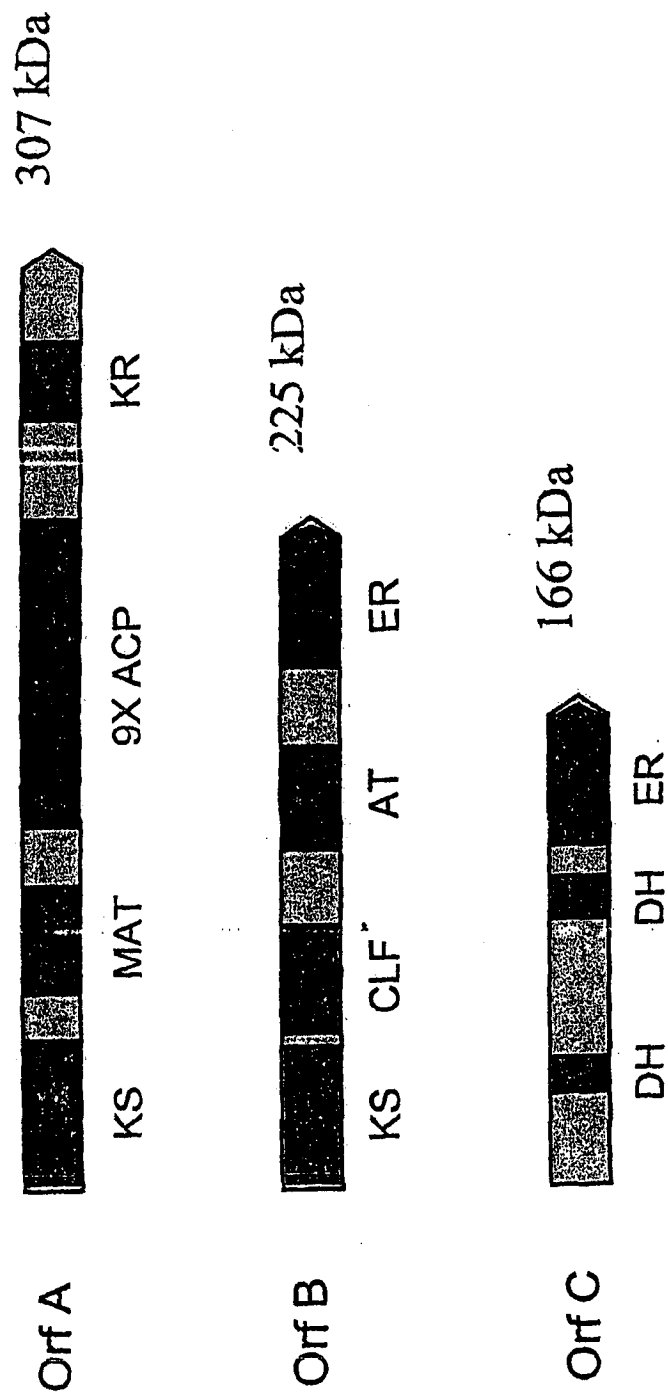
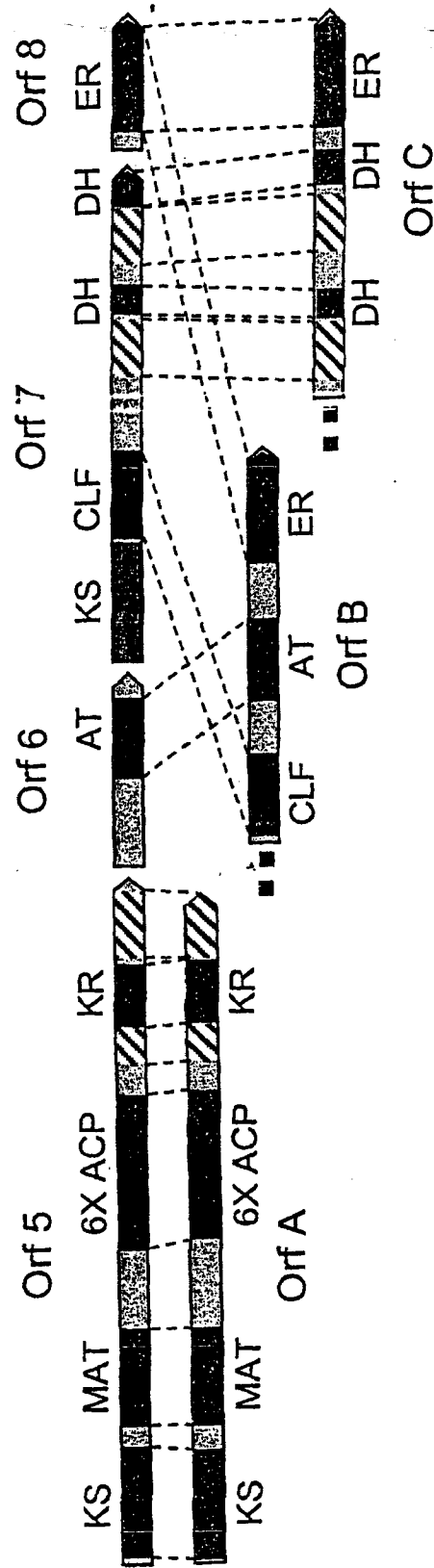


图 1

PKS Orfs/结构域的比较

Schizochytrium 与 *Shewanella*



氨基酸序列同源性



图 2

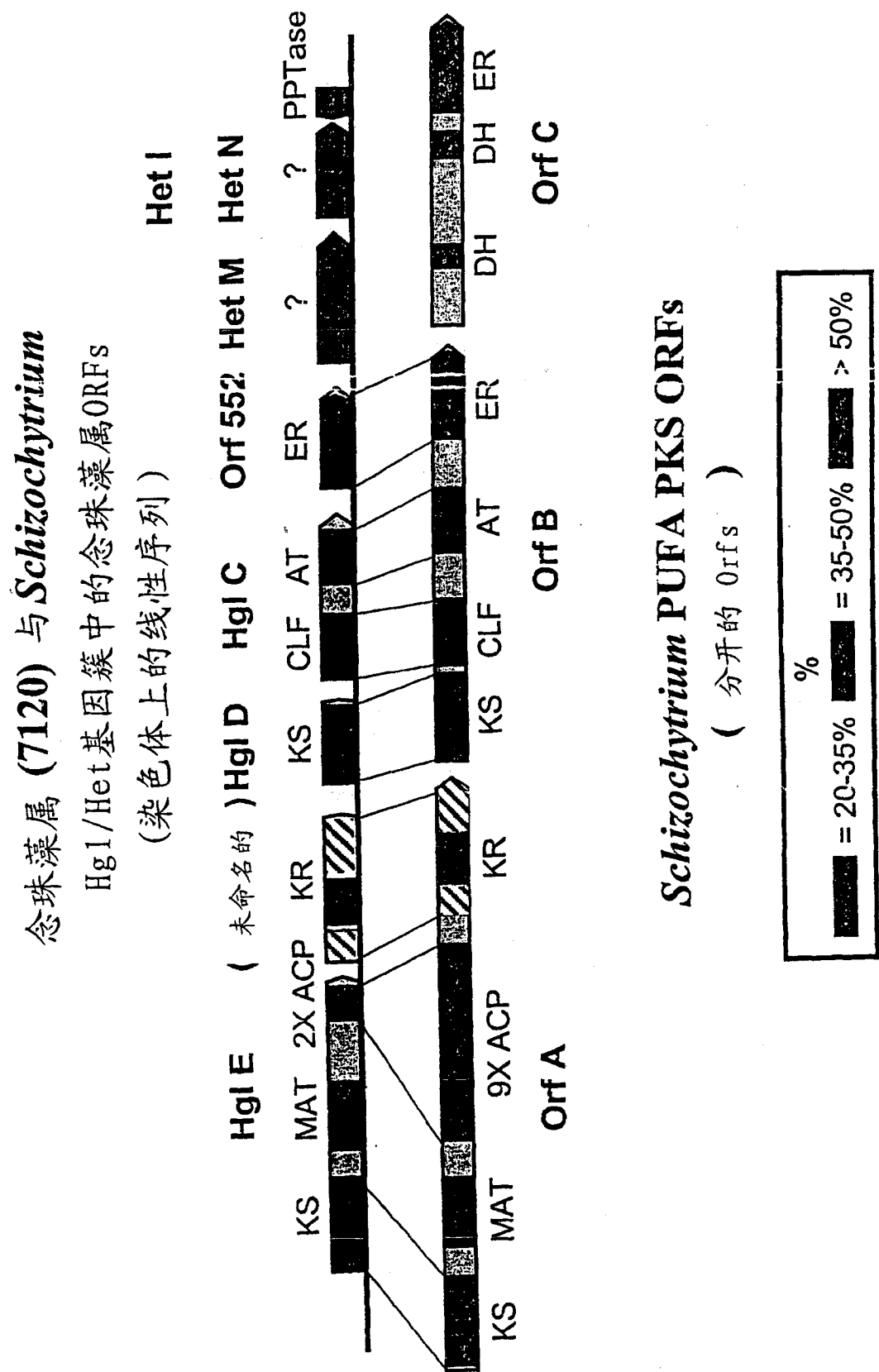


图 3

193