



(12)发明专利申请

(10)申请公布号 CN 107109380 A

(43)申请公布日 2017.08.29

(21)申请号 201580066254.6

(22)申请日 2015.10.06

(30)优先权数据

1417712.5 2014.10.07 GB

PCT/GB2015/051291 2015.05.01 GB

(85)PCT国际申请进入国家阶段日
2017.06.05

(86)PCT国际申请的申请数据

PCT/GB2015/052916 2015.10.06

(87)PCT国际申请的公布数据

W02016/055777 EN 2016.04.14

(71)申请人 牛津纳米孔技术公司
地址 英国牛津郡

(72)发明人 安德鲁·海伦
瑞贝卡·维多利亚·鲍恩

马克·布鲁斯 拉科马·贾亚辛格
约瑟夫·哈格里夫·劳埃德
绍博尔奇·苏罗尔斯
伊丽莎白·杰恩·华莱士
克里斯多夫·彼得·约德

(74)专利代理机构 北京同立钧成知识产权代理
有限公司 11205
代理人 彭雪瑞 臧建明

(51)Int.Cl.

C12N 9/14(2006.01)

C12N 15/55(2006.01)

C12N 15/63(2006.01)

C12Q 1/68(2006.01)

权利要求书7页 说明书92页
序列表73页 附图9页

(54)发明名称

经修饰的酶

(57)摘要

本发明涉及修饰的Dda解旋酶,其可用于控制多核苷酸的移动,并且特别有助于对多核苷酸进行测序。

1. 一种DNA依赖性ATPase (Dda) 解旋酶, 其中:
 - (a) 至少一个与单链DNA (ssDNA) 中一个或多个核苷酸相互作用的氨基酸被取代; 和
 - (b) 与跨膜孔相互作用的解旋酶的一部分包括一个或多个修饰,其中, 所述解旋酶具有控制多核苷酸移动的能力。
2. 根据权利要求1所述的解旋酶, 其中, 在 (a) 中, 用包含较大侧链 (R基团) 的氨基酸取代至少一个与单链DNA (ssDNA) 中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸。
3. 根据权利要求2所述的解旋酶, 其中, 所述解旋酶包括:
 - (a) SEQ ID NO:8的变体, 其中, 所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396和Y415中的至少一个; 或者
 - (b) SEQ ID NO:9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22或23的变体, 其中, 所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是与SEQ ID NO:8中的H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396和Y415相对应的氨基酸中的至少一个。
4. 根据权利要求2或3所述的解旋酶, 其中, 所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是至少一个插入在ssDNA中的核苷酸之间的氨基酸。
5. 根据权利要求4所述的解旋酶, 其中, 所述解旋酶包括:
 - (a) SEQ ID NO:8的变体, 其中, 所述至少一个插入在ssDNA中的核苷酸之间的氨基酸是P89, F98和V150中的至少一个; 或者
 - (b) SEQ ID NO:9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22或23的变体, 其中, 所述至少一个插入在ssDNA中的核苷酸之间的氨基酸是与SEQ ID NO:8中的P89, F98和V150相对应的氨基酸中的至少一个。
6. 根据权利要求2至5中任一项所述的解旋酶, 其中, 所述较大侧链 (R基团) 包括增加数目的碳原子, 具有增加的长度, 增加的分子体积和/或具有增加的范德华体积。
7. 根据权利要求2至6中任一项所述的解旋酶, 其中, 所述较大侧链 (R基团) 增加了所述至少一个氨基酸与所述ssDNA中一个或多个核苷酸之间的 (i) 静电相互作用; (ii) 氢键和/或 (iii) 阳离子- π (阳离子- π) 相互作用。
8. 根据权利要求2至7中任一项所述的解旋酶, 其中, 包含较大侧链 (R基团) 的氨基酸不是丙氨酸 (A)、半胱氨酸 (C)、甘氨酸 (G)、硒代半胱氨酸 (U)、甲硫氨酸 (M)、天冬氨酸 (D) 或谷氨酸 (E)。
9. 根据前述权利要求中任一项所述的解旋酶, 其中
 - 组氨酸 (H) 被 (i) 精氨酸 (R) 或赖氨酸 (K); (ii) 谷氨酰胺 (Q) 或天冬酰胺 (N) 或 (iii) 苯丙氨酸 (F)、酪氨酸 (Y) 或色氨酸 (W) 取代;
 - 天冬酰胺 (N) 被 (i) 精氨酸 (R) 或赖氨酸 (K); (ii) 谷氨酰胺 (Q) 或组氨酸 (H) 或 (iii) 苯丙氨酸 (F)、酪氨酸 (Y) 或色氨酸 (W) 取代;
 - 脯氨酸 (P) 被 (i) 精氨酸 (R) 或赖氨酸 (K); (ii) 谷氨酰胺 (Q)、天冬酰胺 (N)、苏氨酸 (T) 或组氨酸 (H); (iii) 酪氨酸 (Y)、苯丙氨酸 (F) 或色氨酸 (W) 或 (iv) 亮氨酸 (L)、缬氨酸 (V) 或异亮氨酸 (I) 取代;
 - 苯丙氨酸 (F) 被 (i) 精氨酸 (R) 或赖氨酸 (K); (ii) 组氨酸 (H) 或 (iii) 酪氨酸 (Y) 或色氨酸 (W) 取代;

酸(W)取代;

天冬氨酸(D)被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代;

缬氨酸(V)被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代;

丝氨酸(S)被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代;和/或

酪氨酸(Y)被(i)精氨酸(R)或赖氨酸(K)或(iii)色氨酸(W)取代。

10. 根据前述权利要求中任一项所述的解旋酶,其中,所述解旋酶是SEQ ID NO:8的变体,并且包含:

- H82N;
- H82Q; - F276W; - F98W/V150L;
- H82W; - F276R; - F98W/V150N;
- N88R; - F276K; - F98W/V150W;
- N88H; - F276H; - F98W/V150H;
- N88W; - S287K; - F98W/P152W;
- N88Y; - S287R; - F98W/P152F;
- P89L; - S287W; - F98W/P152Y;
- P89V; - S287F; - F98W/P152H;
- P89I; - H396Y; - F98W/P152I;
- P89E; - H396F; - F98W/P152L;
- P89T; - H396Q; - F98W/P152V;
- P89F; - H396K; - F98W/F240W;
- D121H; - Y415W; - F98W/F240Y;
- D121Y; - Y415R; - F98W/F240H;
- D121K; - F98W/H82N; - F98W/F276W;
- V150I; - F98W/H82Q; - F98W/F276R;
- V150L; - F98W/H82W; - F98W/F276K;
- V150N; - F98W/N88R; - F98W/F276H;
- V150W; - F98W/N88H; - F98W/S287K;
- V150H; - F98W/N88W; - F98W/S287R;
- P152W; - F98W/N88Y; - F98W/S287W;
- P152F; - F98W/P89L; - F98W/S287F;

- P152Y; - F98W/P89V; - F98W/H396Y;
- P152H; - F98W/P89I; - F98W/H396F;
- P152I; - F98W/P89T; - F98W/H396Q;
- P152L; - F98W/P89F; - F98W/Y415W;
- P152V; - F98W/D121H; 或者
- F240W; - F98W/D121Y; - F98W/Y415R。
- F240Y; - F98W/D121K;
- F240H; - F98W/V150I;

11. 根据权利要求2至7中任一项所述的解旋酶,其中,具有较大侧链(R基团)的氨基酸是非天然氨基酸。

12. 根据前述权利要求中任一项所述的解旋酶,其中,在(a)中,至少一个与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸被取代。

13. 根据权利要求12所述的解旋酶,其中,所述取代增加了所述至少一个氨基酸和所述ssDNA中一个或多个磷酸基团之间的(i)静电相互作用;(ii)氢键和/或(iii)阳离子- π (阳离子- π)相互作用。

14. 根据权利要求13所述的解旋酶,其中,所述取代增加了位点的净正电荷。

15. 根据权利要求12至14中任一项所述的解旋酶,其中,所述解旋酶包括:

(a) SEQ ID NO:8的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸是H64,T80,S83,N242,K243,N293,T394和K397中的至少一个;或者

(b) SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸是与SEQ ID NO:8中的H64,T80,S83,N242,K243,N293,T394和K397相对应的至少一个氨基酸。

16. 根据权利要求15所述的解旋酶,其中

组氨酸(H)被(i)精氨酸(R)或赖氨酸(K);(ii)天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)或苏氨酸(T);(iii)苯丙氨酸(F)、色氨酸(W)或酪氨酸(Y)取代;

丝氨酸(S)被(i)精氨酸(R)、组氨酸(H)或赖氨酸(K);(ii)天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)或组氨酸(H)或(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代;

天冬酰胺(N)被(i)精氨酸(R)、组氨酸(H)或赖氨酸(K);(ii)丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代;

赖氨酸(K)被(i)精氨酸(R)或组氨酸(H);(ii)丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代;和/或

酪氨酸(Y)被(i)精氨酸(R)或组氨酸(H);(ii)丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代。

17. 根据权利要求12至16中任一项所述的解旋酶,其中,所述氨基酸被非天然氨基酸取代。

18. 根据权利要求12至17中任一项所述的解旋酶,其中,所述解旋酶是SEQ ID NO:8的变体,其包含一个或多个的(a)H64N,H64Q,H64K或H64F;(b)T80K,T80Q或T80N;(c)S83H,

S83N, S83K, S83T, S83R或S83Q (d) N242H或N242Q; (e) K243Q或K243H; (f) N293Q, N293K或N293H; (g) T394K, T394H或T394N和 (h) K397R, K397H或K397Y。

19. 根据前述权利要求中任一项所述的解旋酶, 其中, 所述解旋酶是SEQ ID NO:8的变体, 其包含以下位点上的取代:

- F98/H64, 例如F98W/H64N, F98W/H64Q, F98W/H64K或F98W/H64F;
- F98/T80, 例如F98W/T80K, F98W/T80Q, F98W/T80N;
- F98/H82, 例如F98W/H82N, F98W/H82Q或F98W/H82W;
- F98/S83, 例如F98W/S83H, F98W/S83N, F98W/S83K, F98W/S83T, F98W/S83R或F98W/S83Q;
- F98/N242, 例如F98W/N242H, F98W/N242Q, F98W/K243Q或F98W/K243H;
- F98/N293, 例如F98W/N293Q, F98W/N293K, F98W/N293H, F98W/T394K, F98W/T394H, F98W/T394N, F98W/H396Y, F98W/H396F, F98W/H396Q或F98W/H396K; 或者
- F98/K397, 例如F98W/K397R, F98W/K397H或F98W/K397Y。

20. 根据前述权利要求中任一项所述的解旋酶, 其中, 所述解旋酶包含

(a) SEQ ID NO:8的变体, 并且与跨膜孔相互作用的解旋酶的一部分包括SEQ ID NO:8中的位点1, 2, 3, 4, 5, 6, 51, 176, 177, 178, 179, 180, 181, 185, 189, 191, 193, 194, 195, 197, 198, 199, 200, 201, 202, 203, 204, 207, 208, 209, 210, 211, 212, 213, 216, 219, 220, 221, 223, 224, 226, 227, 228, 229, 247, 254, 255, 256, 257, 258, 259, 260, 261, 298, 300, 304, 308, 318, 319, 321, 337, 347, 350, 351, 405, 415, 422, 434, 437, 438;

(b) SEQ ID NO:8的变体, 并且与跨膜孔相互作用的解旋酶的一部分包括SEQ ID NO:8中的位点1, 2, 4, 51, 177, 178, 179, 180, 185, 193, 195, 197, 198, 199, 200, 202, 203, 204, 207, 208, 209, 210, 211, 212, 216, 221, 223, 224, 226, 227, 228, 229, 254, 255, 256, 257, 258, 260, 304, 318, 321, 347, 350, 351, 405, 415, 422, 434, 437和438;

(c) SEQ ID NO:8的变体, 并且与跨膜孔相互作用的解旋酶的一部分包括SEQ ID NO:8中的位点1, 2, 178, 179, 180, 185, 195, 197, 198, 199, 200, 202, 203, 207, 209, 210, 212, 216, 221, 223, 226, 227, 255, 258, 260, 304, 350和438; 或者

(d) SEQ ID NO:9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22或23的变体, 并且与跨膜孔相互作用的解旋酶的一部分包含与(a), (b)或(c)中限定的SEQ ID NO:8中的位点对应的位点。

21. 根据权利要求20所述的解旋酶, 其中, 所述解旋酶包含SEQ ID NO:8的变体, 其包含在(a) K194 (b) W195 (c) D198 (d) K199和(e) E258上的一个或多个修饰。

22. 根据前述权利要求中任一项所述的解旋酶, 其中, 所述解旋酶是SEQ ID NO:8的变体, 其包含在以下位点上的取代:

- F98/K194/H64, 例如, F98W/K194L/H64N, F98W/K194L/H64Q, F98W/K194L/H64K或者F98W/K194L/H64F;
- F98/K194/T80, 例如, F98W/K194L/T80K, F98W/K194L/T80Q或者F98W/K194L/T80N;
- F98/K194/H82, 例如, F98W/K194L/H82N, F98W/K194L/H82Q或者F98W/K194L/H82W
- F98/S83/K194, 例如, F98W/S83H/K194L, F98W/S83T/K194L, F98W/S83R/K194L, F98W/S83Q/K194L, F98W/S83N/K194L, F98W/S83K/K194L, F98W/N88R/K194L, F98W/N88H/K194L,

F98W/N88W/K194L或者F98W/N88Y/K194L;

-F98/S83/K194/F276, 例如, F98W/S83H/K194L/F276K;

-F98/P89/K194, 例如, F98W/P89L/K194L, F98W/P89V/K194L, F98W/P89I/K194L或者F98W/P89T/K194L;

-F98/D121/K194, 例如, F98W/D121H/K194L, F98W/D121Y/K194L或者F98W/D121K/K194L;

-F98/V150/K194, 例如, F98W/V150I/K194L, F98W/V150L/K194L, F98W/V150N/K194L, F98W/V150W/K194L或者F98W/V150H/K194L;

-F98/P152/K194, 例如, F98W/P152W/K194L, F98W/P152F/K194L, F98W/P152Y/K194L, F98W/P152H/K194L, F98W/P152I/K194L, F98W/P152L/K194L或者F98W/P152V/K194L;

-F98/F240/K194, 例如, F98W/F240W/K194L, F98W/F240Y/K194L或者F98W/F240H/K194L;

-F98/N242/K194, 例如, F98W/N242H/K194L或者F98W/N242Q/K194L;

-F98/K194/F276, 例如, F98W/K194L/F276K, F98W/K194L/F276H, F98W/K194L/F276W或者F98W/K194L/F276R;

-F98/K194/S287, 例如, F98W/K194L/S287K, F98W/K194L/S287R, F98W/K194L/S287W或者F98W/K194L/S287F;

-F98/N293/K194, 例如, F98W/N293Q/K194L, F98W/N293K/K194L或者F98W/N293H/K194L;

-F98/T394/K194, 例如, F98W/T394K/K194L, F98W/T394H/K194L或者F98W/T394N/K194L;

-F98/H396/K194, 例如, F98W/H396Y/K194L, F98W/H396F/K194L, F98W/H396Q/K194L或者F98W/H396K/K194L;

-F98/K397/K194, 例如, F98W/K397R/K194L, F98W/K397H/K194L或者F98W/K397Y/K194L;或者

-F98/Y415/K194, 例如, F98W/Y415W/K194L或者F98W/Y415R/K194L。

23. 根据前述权利要求中任一项所述的解旋酶, 其中

(a) 至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入到(i) 塔域和/或(ii) 销域和/或(iii) 1A (RecA式马达) 域;

(b) 至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入到钩域和/或2A (RecA式马达) 域;或者

(c) 对解旋酶进行修饰以减少其表面负电荷。

24. 根据前述权利要求中任一项所述的解旋酶, 其中, 所述解旋酶是SEQ ID NO:8的变体, 并且包含E94C, F98W, C109A, C136A, K194L和A360C。

25. 一种构建体, 其包含根据前述权利要求中任一项所述的解旋酶和附加的多核苷酸结合部分, 其中, 所述解旋酶与所述多核苷酸结合部分连接, 并且所述构建体具有控制多核苷酸移动的能力。

26. 根据权利要求25所述的构建体, 其中, 所述构建体包含两个或多个根据权利要求1至24中任一项所述的解旋酶。

27. 一种多核苷酸,其包含对根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体进行编码的序列。

28. 一种载体,其包含与启动子可操作地连接的根据权利要求27所述的多核苷酸。

29. 一种宿主细胞,其包含根据权利要求25所述的载体。

30. 一种制备根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体的方法,其包括对根据权利要求27所述的多核苷酸进行表达,用根据权利要求28所述的载体对细胞进行转染或对根据权利要求29所述的宿主细胞进行培养。

31. 一种控制多核苷酸的移动的方法,包括使所述多核苷酸与根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体接触,从而控制所述多核苷酸的移动。

32. 根据权利要求31所述的方法,其中,所述方法是用于控制多核苷酸穿过跨膜孔的移动。

33. 一种表征目标多核苷酸的方法,包括:

(a) 使所述目标多核苷酸与跨膜孔和根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体接触,使得所述解旋酶或构建体控制所述目标多核苷酸穿过所述孔的移动;以及

(b) 随着所述多核苷酸相对于所述孔移动,获取一个或多个测量值,其中所述测量值代表所述目标多核苷酸的一个或多个特征,并由此表征所述目标多核苷酸。

34. 根据权利要求33所述的方法,其中,所述一个或多个特征选自 (i) 所述目标多核苷酸的长度; (ii) 所述目标多核苷酸的同一性; (iii) 所述目标多核苷酸的序列; (iv) 所述目标多核苷酸的二级结构和 (v) 所述目标多核苷酸是否是经修饰的。

35. 根据权利要求34所述的方法,其中,所述目标多核苷酸通过甲基化、氧化、损伤,用一个或多个蛋白质或一个或多个标记物、标签或间隔物进行了修饰。

36. 根据权利要求33至35中任一项所述的方法,其中,通过电测量和/或光测量来测量所述目标多核苷酸的一个或多个特征。

37. 根据权利要求36所述的方法,其中,所述电测量是电流测量、阻抗测量、隧道测量或场效应晶体管 (FET) 测量。

38. 根据权利要求37所述的方法,其中,所述方法包括:

(a) 使所述目标多核苷酸与跨膜孔和根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体接触,使得所述解旋酶或构建体控制目标多核苷酸穿过所述孔的移动;以及

(b) 随着所述多核苷酸相对于所述孔移动,测量流经所述孔的电流,其中所述电流代表所述目标多核苷酸的一个或多个特征,并由此表征所述目标多核苷酸。

39. 根据权利要求33至38中任一项所述的方法,其中,所述方法还包括这一步骤,即跨所述孔施加电压以在所述孔和所述解旋酶或构建体之间形成复合物。

40. 根据权利要求33至38中任一项所述的方法,其中,所述多核苷酸的至少一部分是双链的。

41. 根据权利要求33至41中任一项所述的方法,其中,所述孔是跨膜蛋白孔或固态孔。

42. 根据权利要求41所述的方法,其中,所述跨膜蛋白孔衍生自溶血素、白细胞介素、耻

垢分枝杆菌孔蛋白A (MspA)、MspB、MspC、MspD、胞溶素、外膜孔蛋白F (OmpF)、外膜孔蛋白G (OmpG)、外膜磷脂酶A、奈瑟氏菌自转运脂蛋白 (NaIP) 和WZA。

43. 根据权利要求42所述的方法, 其中, 所述跨膜蛋白由SEQ ID NO:2所示的8个相同的亚基形成, 或 (b) 所述亚基的变体形成, 其中, 基于氨基酸同一性, 所述8个亚基中的一个或多个相比于SEQ ID NO:2在整个序列上具有至少50%的同源性并具有孔活性。

44. 一种用于形成表征目标多核苷酸的传感器的方法, 包括: 在 (a) 孔和 (b) 根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体之间形成复合物, 从而形成用于表征目标多核苷酸的传感器。

45. 根据权利要求44所述的方法, 其中, 所述复合物通过以下步骤形成: (a) 在目标多核苷酸的存在下使所述孔和所述解旋酶或构建体接触, 以及 (a) 跨所述孔施加电势。

46. 根据权利要求45所述的方法, 其中, 所述电势是电压电势或化学电势。

47. 根据权利要求46所述的方法, 其中, 通过将所述孔与所述解旋酶或构建体共价连接而形成所述复合物。

48. 一种用于表征目标多核苷酸的传感器, 其包含在 (a) 孔和 (b) 根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体之间的复合物。

49. 根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体的用途, 为控制目标多核苷酸穿过孔的移动。

50. 一种用于表征目标多核苷酸的试剂盒, 其包含

(a) 孔和根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体; 或者

(b) 根据权利要求1至24中任一项所述的解旋酶或根据权利要求25或26所述的构建体和一个或多个负载部分。

51. 一种用于表征样品中的目标多核苷酸的装置, 包括 (a) 多个孔和 (b) 多个根据权利要求1至24中任一项所述的解旋酶或多个根据权利要求25或26所述的构建体。

52. 一种用于制备根据权利要求1至24中任一项所述的解旋酶的方法, 包括:

(a) 提供解旋酶; 以及

(b) 对所述解旋酶进行修饰以制备根据权利要求1至24中任一项所述的解旋酶。

53. 根据权利要求52所述的方法, 其中, 所述方法还包括 (c) 确定是否所得到的解旋酶能够控制多核苷酸的移动。

54. 一种制备根据权利要求25或26所述的构建体的方法, 包括将根据权利要求1至24中任一项所述的解旋酶与附加的多核苷酸结合部分连接, 从而制备出所述构建体。

55. 根据权利要求54所述的方法, 其中, 所述方法还包括确定是否所得到的构建体能够控制多核苷酸的移动。

56. 一系列与多核苷酸连接的两个或多个解旋酶, 其中, 两个或多个解旋酶中的至少一个是根据权利要求1至24中任一项所述的解旋酶。

经修饰的酶

技术领域

[0001] 本发明涉及一种修饰的Dda解旋酶,其可用于控制多核苷酸的移动,并且特别有助于测序多核苷酸。

背景技术

[0002] 目前需要一种具有广泛的应用范围的快速且廉价的多核苷酸(如DNA或RNA)测序和鉴定技术。现有的技术是缓慢的且昂贵的,这主要由于它们依赖于扩增技术来产生大量的多核苷酸,且需要大量特定的用于信号检测的荧光化学物质。

[0003] 跨膜孔(纳米孔)作为用于聚合物和各种小分子的直接的、电生物传感器,具有很大的潜力。特别是,目前作为一种有潜力的DNA测序技术的纳米孔得到了许多关注。

[0004] 当跨纳米孔施加电势时,在,当分析物如核苷酸在桶(barrel)中短暂停留一段时间时,会产生电流的变化。所述核苷酸的纳米孔检测能产生已知特征和持续时间的电流变化。在链测序方法中,单个多核苷酸链穿过所述孔并实现对核苷酸的鉴定。链测序可包括使用多核苷酸结合蛋白来控制所述多核苷酸穿过所述孔的移动。

发明内容

[0005] 令人出乎意料的是,发明人已经鉴定出特异性Dda突变体,该突变体具有提升的能力以控制多核苷酸穿过孔的移动。本发明的突变体显示减少的向前滑移(slipping)。这是DNA通过至少4个连续核苷酸并且通常通过超过10个连续核苷酸相对孔向前移动的现象。向前滑移可能涉及100个或更多个连续的核苷酸的向前移动,这对于每个多核苷酸可能发生多次。对于多核苷酸测序,向前滑移可能是有问题的。由发明人鉴定的突变体通常包含突变的组合,即(1)与单链DNA(ssDNA)中的核苷酸相互作用的氨基酸的一个或多个取代,和(2)与跨膜孔相互作用的突变体的一部分上的一个或多个修饰。

[0006] 因此,本发明提供了一种DNA依赖性ATPase(Dda)解旋酶,在其中(a)与单链DNA(ssDNA)中的一个或多个核苷酸相互作用的至少一个氨基酸被取代,和(b)与跨膜孔相互作用的解旋酶的一部分包括一个或多个修饰,其中,所述解旋酶具有控制多核苷酸的移动的能力。

[0007] 本发明还提供:

[0008] 一种构建体,其包含本发明所述的解旋酶和附加的多核苷酸结合部分,其中,所述解旋酶与所述多核苷酸结合部分连接,并且所述构建体具有控制多核苷酸的移动的能力;

[0009] 一种多核苷酸,其包含对本发明所述的解旋酶或本发明所述的构建体进行编码的序列;

[0010] 一种载体,其包含与启动子可操作地连接的本发明所述的多核苷酸;

[0011] 一种宿主细胞,其包含本发明所述的载体;

[0012] 一种制备本发明所述的解旋酶或本发明所述的构建体的方法,其包括对本发明所述的多核苷酸进行表达,用本发明所述的载体对细胞进行转染或对本发明所述的宿主细胞

进行培养；

[0013] 一种控制多核苷酸的移动的方法，包括使所述多核苷酸与本发明所述的解旋酶或本发明所述的构建体接触，从而控制多核苷酸的移动；

[0014] 一种表征目标多核苷酸的方法，包括：(a) 使所述目标多核苷酸与跨膜孔和本发明所述的解旋酶或本发明所述的构建体接触，使得所述解旋酶控制目标多核苷酸穿过所述孔的移动；以及 (b) 随着所述多核苷酸相对于所述孔移动，获取一个或多个测量值，其中所述测量值代表所述目标多核苷酸的一个或多个特征，并由此表征所述目标多核苷酸；

[0015] 一种用于形成表征目标多核苷酸的传感器的方法，包括：在 (a) 孔和 (b) 本发明所述的解旋酶或本发明所述的构建体之间形成复合物，从而形成用于表征目标多核苷酸的传感器；

[0016] 一种用于表征目标多核苷酸的传感器，其包含在 (a) 孔和 (b) 本发明所述的解旋酶或本发明所述的构建体之间的复合物；

[0017] 本发明所述的解旋酶或本发明所述的构建体的用途，为控制目标多核苷酸穿过孔的移动。

[0018] 一种用于表征目标多核苷酸的试剂盒，其包含 (a) 孔和 (b) 本发明所述的解旋酶或本发明所述的构建体；

[0019] 一种用于表征样品中目标多核苷酸的装置，包括 (a) 多个孔和 (b) 多个本发明所述的解旋酶或多个本发明所述的构建体；和

[0020] 一系列与多核苷酸连接的两个或多个解旋酶，其中，两个或多个解旋酶中的至少一个是本发明所述的解旋酶。

附图说明

[0021] 图1示出了T4 Dda-E94C/A360C/C109A/C136A (具有E94C/A360C/C109A/C136A突变的SEQ ID NO:8) 相对于MspA- (G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K) 8 (具有G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K突变的SEQ ID NO:2=MspA突变体1) 或MspA- (缺失-L74/G75/D118/L119) D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K) 8 (具有D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K突变并缺失L74/G75/D118/L119氨基酸的SEQ ID NO:2=MspA突变体2) 的三种不同的初始模拟取向。运行1和运行2之间的区别在于尽管孔和酶处于相同位点，酶和孔均具有不同的侧链构象。在运行3中，酶相对于纳米孔略微倾斜。

[0022] 图2示出了纳米孔MspA突变体1与酶突变体1的相互作用点的图表 (y轴标签=孔/酶接触的数量，x轴标签=孔氨基酸残基数量)。图表的每一行示出了不同酶/纳米孔取向例如，运行1-3的相互作用点。

[0023] 图3示出了酶突变体1与MspA突变体1的相互作用点的图表 (y轴标签=孔/酶接触的数量，x轴标签=酶氨基酸残基数量)。图表的每一行示出了不同酶/纳米孔取向例如，运行1-3的相互作用点。

[0024] 图4示出了纳米孔MspA突变体2与酶突变体1的相互作用点的图表 (y轴标签=孔/酶接触的数量，x轴标签=孔氨基酸残基数量)。图表的每一行示出了不同酶/纳米孔取向例如，运行1-3的相互作用点。

[0025] 图5示出了酶突变体1与MspA突变体2的相互作用点的图表(y轴标签=孔/酶接触点数量,x轴标签=酶氨基酸残基数量)。图表的每一行示出了不同酶/纳米孔取向例如,运行1-3的相互作用点。

[0026] 图6(A)示出了图表(y轴标签=孔氨基酸残基数量,x轴标签=酶氨基酸残基数量)的两个区域,其示出了来自运行1的孔(MspA突变体2)中的哪些氨基酸与酶(酶突变体1)中的特定氨基酸相互作用。图6(B)示出了图表(y轴标签(a1)=孔氨基酸残基数量,y轴标签(a2)=孔/酶接触的数量,x轴标签=酶氨基酸残基数量)的一个区域,其示出了来自运行3的孔(MspA突变体2)中的哪些氨基酸与酶(酶突变体1)中的特定氨基酸相互作用。图表中的灰色带条表示氨基酸之间的相互作用。灰色带条的暗度对应酶/孔之间的相互作用的数量,其中深灰色=相互作用较多且浅灰色=相互作用较少。每个盒子中的第一氨基酸对应MspA突变体2中的相互作用的氨基酸,第二氨基酸对应酶突变体1中的相互作用的氨基酸。

[0027] 图7示出了实施例3中使用的DNA构建体X。A段对应30个iSpC3间隔物。B段对应SEQ ID NO:60。标记C对应实验中使用的酶突变体。D段对应4个iSp18间隔物。E段对应SEQ ID NO:61。F端对应4个iSpC3间隔物。G段对应SEQ ID NO:62。H段对应SEQ ID NO:63。I段对应SEQ ID NO:64。J段对应3'胆固醇。

[0028] 图8示出了当解旋酶(T4 Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8,然后(Δ M1)G1))控制DNA构建体X穿过CsgG-Eco纳米孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有Y51T/F56Q突变的SEQ ID NO:66,其中StrepII(C)是SEQ ID NO:67并连接在C末端)的移位时的示例性电流迹线(y轴标签=电流(pA),x轴标签=所有三条迹线的时间(s))。B和C段示出了电流迹线A的放大区域。

[0029] 图9示出了当解旋酶(T4 Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8,然后(Δ M1)G1))控制DNA/RNA构建体Y穿过MspA纳米孔的移位时的示例性电流迹线(y轴标签=电流(pA),x轴标签=时间(s))。标记为1的区域对应RNA区域,标记为2的区域对应DNA区域。

[0030] 序列表说明

[0031] SEQ ID NO:1表示编码了野生型MspA单体的密码子优化的多核苷酸序列。该突变体缺乏信号序列。

[0032] SEQ ID NO:2表示成熟形式的野生型MspA单体的氨基酸序列。该突变体缺乏信号序列。

[0033] SEQ ID NO:3表示编码 α -溶血素-E111N/K147N(α -HL-NN;Stoddart等,PNAS,2009;106(19):7702-7707)的一种单体的多核苷酸序列。

[0034] SEQ ID NO:4表示 α -HL-NN的一种单体的氨基酸序列。

[0035] SEQ ID NO:5至7表示MspB,C和D的氨基酸序列。

[0036] SEQ ID NO:8至23表示表1和表2中所示的Dda解旋酶的氨基酸序列。

[0037] SEQ ID NO:24表示优选的HhH域的氨基酸序列。

[0038] SEQ ID NO:25表示来自噬菌体(bacteriophage)RB69的ssb的氨基酸序列,其由gp32基因编码。

[0039] SEQ ID NO:26表示来自噬菌体T7的ssb的氨基酸序列,其由gp2.5基因编码。

- [0040] SEQ ID NO:27表示来自疱疹病毒1 (Herpes virus 1) 的UL42持续合成性因子的氨基酸序列。
- [0041] SEQ ID NO:28表示PCNA的亚基1的氨基酸序列。
- [0042] SEQ ID NO:29表示PCNA的亚基2的氨基酸序列。
- [0043] SEQ ID NO:30表示PCNA的亚基3的氨基酸序列。
- [0044] SEQ ID NO:31表示Phi29DNA聚合酶的氨基酸序列。
- [0045] SEQ ID NO:32表示来自疱疹病毒1的UL42持续合成性因子的氨基酸序列 (1至319)。
- [0046] SEQ ID NO:33表示来自噬菌体RB69的ssb的氨基酸序列,即SEQ ID NO:25,其C末端缺失 (gp32RB69CD)。
- [0047] SEQ ID NO:34表示来自噬菌体T7 (gp2.5T7-R21IDe1) 的ssb的氨基酸序列 (1至210)。全长蛋白在SEQ ID NO:96中进行表示。
- [0048] SEQ ID NO:35表示Hel308 H1a的第5个域的氨基酸序列。
- [0049] SEQ ID NO:36表示Hel308 Hvo的第5个域的氨基酸序列。
- [0050] SEQ ID NO:37表示 (HhH) 2域的氨基酸序列。
- [0051] SEQ ID NO:38表示 (HhH) 2- (HhH) 2域的氨基酸序列。
- [0052] SEQ ID NO:39表示人线粒体SSB (HsmtSSB) 的氨基酸序列。
- [0053] SEQ ID NO:40表示来自Phi29 DNA聚合酶的p5蛋白的氨基酸序列。
- [0054] SEQ ID NO:41表示来自大肠杆菌 (E.coli) 的野生型SSB的氨基酸序列。
- [0055] SEQ ID NO:42表示来自噬菌体T4的ssb的氨基酸序列,其由gp32基因编码。
- [0056] SEQ ID NO:43表示EcoSSB-CterAla的氨基酸序列。
- [0057] SEQ ID NO:44表示EcoSSB-CterNGGN的氨基酸序列。
- [0058] SEQ ID NO:45表示EcoSSB-Q152de1的氨基酸序列。
- [0059] SEQ ID NO:46表示EcoSSB-G117de1的氨基酸序列。
- [0060] SEQ ID NO:47表示拓扑异构酶V Mka (如甲烷嗜热菌) (Methanopyrus Kandleri) 的氨基酸序列。
- [0061] SEQ ID NO:48表示拓扑异构酶V Mka (如甲烷嗜热菌) 的域H-L的氨基酸序列。
- [0062] SEQ ID NO:49表示突变体S (大肠杆菌) (Escherichia coli) 的氨基酸序列。
- [0063] SEQ ID NO:50表示Sso7d (硫磺矿硫化叶菌) (Sulfolobus solfataricus) 的氨基酸序列。
- [0064] SEQ ID NO:51表示Sso10b1 (硫磺矿硫化叶菌P2) (Sulfolobus solfataricus P2) 的氨基酸序列。
- [0065] SEQ ID NO:52表示Sso10b2 (硫磺矿硫化叶菌P2) 的氨基酸序列。
- [0066] SEQ ID NO:53表示色氨酸阻遏物 (大肠杆菌) 的氨基酸序列。
- [0067] SEQ ID NO:54表示λ阻遏物 (肠杆菌噬菌体λ) (Enterobacteria phage lambda) 的氨基酸序列。
- [0068] SEQ ID NO:55表示Cren7 (组蛋白中的泉古菌Cren7Sso) (Histone crenarchaea Cren7Sso) 的氨基酸序列。
- [0069] SEQ ID NO:56表示人组蛋白 (智人) (Homo sapiens) 的氨基酸序列。

[0070] SEQ ID NO:57表示dsbA(肠杆菌噬菌体T4)(Enterobacteria phage T4)的氨基酸序列。

[0071] SEQ ID NO:58表示Rad51(智人)的氨基酸序列。

[0072] SEQ ID NO:59表示PCNA滑移夹钳(海洋细菌JL354)(Citromicrobium bathyomarinum JL354)的氨基酸序列。

[0073] SEQ ID NO:60至64表示实施例3中使用的多核苷酸序列。

[0074] SEQ ID NO:65表示编码来自大肠杆菌K-12菌株MC4100亚株(Escherchia coli Str.K-12substr.MC4100)的野生型CsgG单体的密码子优化的多核苷酸序列。该单体缺乏信号序列。

[0075] SEQ ID NO:66表示来自大肠杆菌K-12菌株MC4100亚株的成熟形式的野生型CsgG单体的氨基酸序列。该单体缺乏信号序列。用于此CsgG的缩写=CsgG-Eco。

[0076] SEQ ID NO:67表示StrepII(C)的氨基酸序列。

[0077] SEQ ID NO:68至73表示实施例5中使用的多核苷酸序列。

具体实施方式

[0078] 应理解,公开的产品和方法的不同应用可以根据本领域的具体需求而调整。可以理解本文中使用的术语仅是为了描述本发明的具体实施方式的目的,而不意为对本发明的限制。

[0079] 此外,如本说明书和所附权利要求书中所使用的,除非另有说明,单数形式“一”、“一个”和“所述”包括复数指代。因此,例如,涉及“解旋酶”时包括“多个解旋酶”,涉及“修饰”时包括两个或多个所述修饰,涉及“跨膜蛋白孔”时包括两个或多个所述孔,等。

[0080] 本文所引用的所有公开物、专利和专利申请,无论在前文或后文,均以引用的方式全文引入。

[0081] 修饰Dda解旋酶

[0082] 本发明提供一种修饰的Dda解旋酶。以下将对一个或多个具体的修饰进行更详细的讨论。本发明所述的修饰包括如下所述的一个或多个取代。

[0083] 采用跨膜孔和解旋酶对多核苷酸进行的表征,例如,测序,通常涉及分析由k个核苷酸组成的聚合物单元,其中,k是正整数(即,“k-聚体”)。这在申请号为PCT/GB2012/052343(公开为W02013/041878)的国际申请中进行了讨论。当目标多核苷酸相对于孔移动或移动穿过孔时,通常通过测量流过孔的电流来分析多核苷酸内的不同k-聚体。多核苷酸相对于孔移动,例如移动穿过孔可以被看作是从一个k-聚体移动到另一个或从k-聚体移动到k-聚体。

[0084] 本发明的修饰的解旋酶提供目标多核苷酸相对于跨膜孔,例如,通过跨膜孔的更一致移动。随着目标多核苷酸相对于孔移动,例如,移动穿过孔时,解旋酶优选提供从一个k-聚体到另一个k-聚体或从k-聚体到k-聚体的更一致移动。解旋酶使得目标多核苷酸能够相对于跨膜孔更平稳地移动,例如,更平稳地移动通过跨膜孔。解旋酶优选提供目标多核苷酸相对于跨膜孔,例如,通过跨膜孔的更规则或较少的不规则的移动。

[0085] 对一个或多个与ssDNA中一个或多个核苷酸相互作用的氨基酸进行的修饰,特别是取代,使得经修饰的解旋酶显示减少的向前滑移。这是DNA通过至少4个连续核苷酸并且

通常通过超过10个连续核苷酸相对孔向前移动的现象。向前滑移可能涉及100个或更多个连续的核苷酸的向前移动,这对于每个多核苷酸可能发生多次。对于多核苷酸测序,向前滑移可能是有问题的。

[0086] 修饰通常使解旋酶显示的向前滑移的频率降低至少10%、至少20%、至少30%、至少40%、至少50%、至少60%、至少70%、至少80%或至少90%。修饰通常会消除向前滑移,即,使解旋酶显示的向前滑移的频率100%降低。修饰通常使解旋酶显示的向前滑移的长度减少到10个核苷酸或更少,例如,9个核苷酸或更少、8个核苷酸或更少、7个核苷酸或更少、6个核苷酸或更少、5个核苷酸或更少、4个核苷酸或更少、3个核苷酸或更少、2个核苷酸或更少或者1个核苷酸。修饰优选地降低解旋酶显示的向前滑移的频率和长度。

[0087] 可以使用本领域已知的任何方法来测量向前滑移。通常,在纳米孔系统中,例如,下述的纳米孔系统中测定解旋酶控制多核苷酸的移动的能力和向前滑移的发生率。如实施例中所述,可以确定解旋酶控制多核苷酸的移动的能力和向前滑移的发生率。

[0088] 修饰,特别是与跨膜孔相互作用的解旋酶的一部分中的修饰,通常降低与目标多核苷酸相对于跨膜孔移动,例如,穿过跨膜孔,相关联的噪声。当正在分析k-聚体时,在任何维度上的目标多核苷酸的不必要移动通常导致对于k-聚体的在电流特征或电平中的噪声。本发明的解旋酶可以通过减少与目标多核苷酸中一个或多个k-聚体(例如,每个k-聚体)相关联的不必要移动来降低这种噪声。解旋酶可以降低与目标多核苷酸中一个或多个k-聚体(例如,每个k-聚体)的电流电平或特征相关联的噪声。

[0089] 在优选实施方案中,目标多核苷酸是双链的,并且解旋酶降低与互补链的移动相关联的噪声比其降低与模板链的移动相关联的噪声程度上要大和/或,解旋酶增加互补链的移动一致性比其增加模板链的移动一致性程度上要大。这对双链目标多核苷酸的链测序是有利的。双链多核苷酸的两个位点优选通过桥连部分,例如,发夹环或发夹环适配器连接。这将在以下进行更详细的讨论。换言之,本发明的修饰的解旋酶更擅长于控制多核苷酸的移动。解旋酶可以控制多核苷酸的移动的程度通常通过以下更详细讨论的修饰发生改变。

[0090] 对本发明的解旋酶进行修饰。与相应的野生型解旋酶或天然解旋酶相比,修饰的解旋酶通常是经过修饰的。本发明的解旋酶是人造的或非天然的。

[0091] 本发明的修饰的解旋酶对于链测序期间控制多核苷酸的移动是有用工具。解旋酶可以以至少两种主动操作模式(当解旋酶提供所有必需组分以促进移动时,例如,ATP和 Mg^{2+})和一种非主动操作模式(当解旋酶未提供必需组分以促进移动时)控制DNA的移动。当提供所有必需组分以促进移动时,解旋酶沿着DNA在5' -3' 方向上移动,但DNA在纳米孔中的取向(取决于DNA的哪个末端被捕获)意味着酶可以用于逆着施加的场将DNA从纳米孔中移出,或顺着施加的场将DNA移进到纳米孔中。当DNA的3' 末端被捕获时,解旋酶逆着电压施加的场方向运作,将螺旋DNA从纳米孔中牵引出来并进入顺式腔。然而,当在纳米孔中DNA的5' 末端被捕获时,解旋酶顺着电压施加的场方向运作,将螺旋DNA推入纳米孔并进入反式腔。当解旋酶未提供必需组分以促进移动时,其可以与DNA结合并且当通过施加的场将其牵引进入孔中时,其作为减慢DNA移动的制动器。在非主动模式下,DNA是否在3' 或5' 末端被捕获是无关紧要的,而是施加的场将DNA牵引进入纳米孔朝向反式侧,其中,酶作为制动器。当处于非主动模式时,解旋酶对DNA的移动控制可以以多种方式进行描述,包括棘轮

(ratcheting)、滑移和制动。

[0092] 在对多核苷酸,特别是500个或更多个核苷酸构成的多核苷酸进行测序时出现的问题是,控制多核苷酸移动的分子马达可能与多核苷酸脱离。这使得多核苷酸能够被快速地并且以不受控制的方式沿所施加的场的方向牵引穿过孔。本发明的修饰的解旋酶不太可能从正被测序的多核苷酸中解链或脱离。当修饰的解旋酶控制多核苷酸移动穿过纳米孔时,其可以增加多核苷酸的读取长度。在本发明的修饰的解旋酶控制下,移动整个多核苷酸穿过纳米孔的能力使其能够与已知方法相比以改进的精度和速度来估计多核苷酸的特征,例如,其序列。随着链长度增加并且需要持续改进的分子马达时,这变得更加重要。本发明的修饰的解旋酶在控制由500个或更多个核苷酸例如,1000个、5000个、10000个、20000个、50000个、100000个或更多个核苷酸构成的目标多核苷酸的移动时特别有效。

[0093] 本发明的修饰的解旋酶也是等温聚合酶链式反应(PCR)的有用工具。在这种方法中,双链DNA的链通常先通过本发明的解旋酶分离,并通过单链DNA(ssDNA)结合蛋白进行包覆。在第二步中,两个序列特异性引物通常与DNA模板的每个边界杂交。然后可以使用DNA聚合酶将退火的引物延伸到模板来制备双链DNA,本发明的解旋酶然后将两种新合成的DNA产物用作底物,进入下一轮反应。因此,发展同步的链反应,从而导致所选目标序列的指数扩增。

[0094] 修饰的解旋酶具有控制多核苷酸移动的能力。可以采用本领域已知的任何方法测定解旋酶控制多核苷酸移动的能力。例如,解旋酶可以与多核苷酸接触,并且可以使用标准方法确定多核苷酸的位点。修饰的解旋酶控制多核苷酸移动的能力通常在纳米孔系统例如,下文所述的,特别是如实施例中所所述的纳米孔系统中进行测定。

[0095] 本发明的修饰的解旋酶可以是孤立的、基本上孤立的、纯化的或基本上纯化的。如果解旋酶完全没有任何其它组分,如脂质、多核苷酸、孔单体或其它蛋白质,则其是孤立的或纯化的。如果将解旋酶与不会干扰其预期用途的载体或稀释剂混合,则其是基本上孤立的。例如,如果解旋酶以以下形式存在,即,包含小于10%、小于5%、小于2%或小于1%的其它组分(例如脂质、多核苷酸、孔单体或其它蛋白质),则其是基本上孤立的或基本上纯化的。

[0096] 根据本发明,可以对任何Dda解旋酶进行修饰。以下对优选的Dda解旋酶进行讨论。

[0097] Dda解旋酶通常包括以下五个域:1A(RecA式(RecA-like)马达)域、2A(RecA式马达)域、塔域(tower domain)、销域(pin domain)和钩域(hook domain)(Xiaoping He等人,Structure;20:1189-1200)。可以使用蛋白质建模来鉴定这些域,在结晶状态下对蛋白质进行X射线衍射测量(Rupp B(2009).Biomolecular Crystallography:Principles,Practice and Application to Structural Biology.New York:Garland Science.),在溶液中对蛋白质进行核磁共振(NMR)光谱(Mark Ranee;Cavanagh,John;Wayne J.Fairbrother;Arthur W.Hunt III;Skelton,NNicholas J.(2007).Protein NMR spectroscopy:principles and practice(2nd ed.).Boston:Academic Press.)或在冷冻水合状态下对蛋白质进行低温电子显微镜检查(van Heel M,Gowen B,Matadeen R,Orlova EV,Finn R,Pape T,Cohen D,Stark H,Schmidt R,Schatz M,Patwardhan A(2000).“Single-particle electron cryo-microscopy:towards atomic resolution.”.Q Rev Biophys.33:307-69)。通过上述方法确定的蛋白质的结构信息可从蛋白质数据库(PDB)中公开获得。

[0098] 蛋白质建模利用了蛋白质结构比同源物中蛋白质序列更保守这一事实。因此,蛋白质原子分辨率模型的产生取决于对一个或多个蛋白质结构进行鉴定,所述蛋白质结构可能与查询序列的结构相类似。为了评估是否存在合适的蛋白质结构以便用作构建蛋白质模型的“模板”,对蛋白质数据数据库(PDB)进行搜索。如果蛋白质结构与查询序列的序列同一性共享合理的水平,则认为其是合适的模板。如果存在这样的模板,则将模板序列与查询序列进行“对齐”,即,将查询序列中的残基映射到模板残基上。然后,使用序列比对和模板结构来产生查询序列的结构模型。因此,蛋白质模型的质量取决于序列对齐和模板结构的质量。

[0099] 可以进行模拟以评估哪些氨基酸与酶结合位点内的ssDNA中的核苷酸接触。可以利用GROMACS软件包版本4.0.5,采用AMBER-99SB力场和TIP3P水模型进行模拟。优选方法在实施例中进行了公开。

[0100] 本发明的修饰

[0101] 本发明的解旋酶是这样一种解旋酶,其中,至少一个与单链DNA(ssDNA)中一个或多个核苷酸相互作用的氨基酸被取代。可以取代任何数量的氨基酸,例如,1个或更多个、2个或更多个、3个或更多个、4个或更多个、5个或更多个或者6个或更多个氨基酸。当解旋酶沿着ssDNA移动或者当ssDNA移动通过解旋酶时,氨基酸可以顺序地与不同核苷酸相互作用。每个被取代的氨基酸每次可以与任何数量的核苷酸相互作用,例如,每次一个、两个、三个或更多个核苷酸。与单链DNA中一个或多个核苷酸相互作用的氨基酸可以采用上述蛋白质建模来进行鉴定。

[0102] 碱基和/或糖相互作用

[0103] 本发明的解旋酶优选是这样一种解旋酶,其中,用包含较大侧链(R基团)的氨基酸取代至少一个与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸。可以取代任何数量的氨基酸,例如,1个或更多个、2个或更多个、3个或更多个、4个或更多个、5个或更多个或者6个或更多个氨基酸。每个氨基酸可以与碱基、糖或碱基和糖相互作用。可以采用上述蛋白质建模来鉴定与单链DNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸。

[0104] 以下表1总结了根据本发明可进行修饰的优选Dda解旋酶。

[0105]

Dda 同源物 (SEQ ID NO:)		生息场所	蛋白质数 据库	长度	至 1993 的序列 同一性/ %	D/E 数 量相对 K/R 氨 基酸	# C
Rma-DSM (SEQ ID NO: 9)	海洋红嗜热 盐菌 (Rhodothermu s marinus)	温和型嗜 盐生物，中 等嗜热菌 >65°C	D0MKQ2	678	21	-84/+85	2
Csp (SEQ ID NO: 10)	蓝杆藻 (菌株 ATCC 51142) (Cyanothece sp. (strain	海洋细菌	B1X365	496	24	-76/+76	5

[0106]

	ATCC 51142))						
Sru (SEQ ID NO: 11)	Salinibacter ruber	极端嗜盐, 35-45°C	Q2S429	421	26	-78/+54	3
Sgo (SEQ ID NO: 12)	Sulfurimonas gotlandica GD1	生息场所: 深海热液 喷口、海岸 沉积物	B6BJ43	500	27	-72/+64	2
Vph12B8 (SEQ ID NO: 13)	海丽特弧菌噬 菌体 12B8(Vibrio phage henriette 12B8)	宿主发现 在盐水、胃 虫中	M4MBC3	450	27	-62/+47	6
Vph (SEQ ID NO: 14)	弧菌噬菌体 phi-pp2(Vibri o phage phi-pp2)	宿主发现 在盐水、胃 虫中	I6XGX8	421	39	-55/+45	5
Aph65 (SEQ ID NO: 15)	气单胞菌噬菌 体 65(Aeromonas phage 65)	宿主发现 在淡水/淡 盐水、胃虫 中	E5DRP6	434	40	-57/+48	4
AphCC2 (SEQ ID NO: 16)	气单胞菌噬菌 体 CC2(Aeromon as phage CC2)	宿主发现 在淡水/淡 盐水、胃虫 中	I6XH64	420	41	-53/+44	4
Cph (SEQ ID NO: 17)	阪崎肠杆菌噬 菌体 vB CsaM GAP161(Cron obacter phage vB CsaM GAP161)	肠杆菌科 的宿主成 员	K4FBD0	443	42	-59/+57	4
Kph (SEQ ID NO: 18)	克雷伯氏菌噬 菌体 KP15(Klebsiel la phage KP15)	肠杆菌科 的宿主成 员	D5JF67	442	44	-59/+58	5

[0107]

SphIME13 (SEQ ID NO: 19)	寡养单胞菌噬 菌体 IME13(<i>Stenot rophomonas phage IME13</i>)	宿主发现 在土壤中	J7HXT5	438	51	-58/+59	7
AphAc42 (SEQ ID NO: 20)	不动杆菌噬菌 体 Ac42(<i>Acineto bacter phage Ac42</i>)	宿主发现 在土壤中	E5EYE6	442	59	-53/+49	9
SphSP18 (SEQ ID NO: 21)	志贺氏杆菌噬 菌体 SP18(<i>Shigella phage SP18</i>)	肠杆菌科 的宿主成 员	E3SFA5	442	59	-55/+55	9
Yph (SEQ ID NO: 22)	耶尔森氏菌噬 菌体 phiR1-RT(<i>Yer sinia phage phiR1-RT</i>)	肠杆菌科 的宿主成 员	I7J3V8	439	64	-52/+52	7
SphS16 (SEQ ID NO: 23)	沙门氏菌噬菌 体(<i>Salmonella phage S16</i>)	肠杆菌科 的宿主成 员	M1EA88	441	72	-56/+55	5
1993 (SEQ ID NO: 8)	肠道菌噬菌体 T4(<i>Enterobate ria phage T4</i>)	肠杆菌科 的宿主成 员	P32270	439	100	-57/+58	5

[0108] 表1

[0109] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是H82,N88,P89,F98,D121,V150,P152,F240,F276,S287,H396和Y415中的至少一个。这些数字对应SEQ ID NO:8中的相关位点,并且相比SEQ ID NO:8可能需要在在一个或多个氨基酸已经在变体中插入或删除的情况下进行改变。如上所述,本领域技术人员可以确定变体中的相应位点。本发明的解旋酶优选包含SEQ ID NO:8的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是F98和一个或多个H82,N88,P89,D121,V150,P152,F240,F276,S287,H396和Y415,例如,F98/H82,F98/N88,F98/P89,F98/D121,F98/V150,F98/P152,F98/F240,F98/F276,F98/S287或F98/H396。

[0110] 本发明的解旋酶优选为SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是至少一个与SEQ ID NO:8中的H82,N88,P89,F98,D121,V150,P152,F240,F276,

S287, H396和Y415相对应的氨基酸。本发明的解旋酶优选包含SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸是与SEQ ID NO:8中F98对应的氨基酸和一个或多个与SEQ ID NO:8中H82, N88, P89, D121, V150, P152, F240, F276, S287, H396和Y415对应的氨基酸,例如,与F98/H82, F98/N88, F98/P89, F98/D121, F98/V150, F98/P152, F98/F240, F98/F276, F98/S287或F98/H396对应的氨基酸。

[0111] 表2 示出了与SEQ ID NO:8中的H82, N88, P89, F98, D121, V150, P152, F240, F276, S287, H396和Y415对应的SEQ ID NO:9至23中的氨基酸。

[0112]

SEQ ID NO:												
8	H82	N88	P89	F98	D121	V150	P152	F240	F276	S287	H396	Y415
9	H81	F87	D88	S105	S131	V181	Q183	R274	H313	G314	H428	H447
10	A144	Q150	P151	G158	N187	V217	K218	W307	F344	K355	H455	H473
11	H87	R93	L94	G100	G126	V154	N155	Y240	Y277	G280	H377	R395
12	H112	K118	P119	E128	G154	I185	N187	Y273	F309	K310	H414	H433
13	H93	V99	M100	D106	N132	I159	Q160	Y267	M302	G303	H400	K419
14	H74	H80	P81	F90	D114	V143	H145	Y230	M266	P273	H378	Y397
15	I178	I184	P85	F94	S117	E147	A149	Y235	M271	I279	I1387	Y406
16	H65	H71	P72	F81	S104	V133	H135	F222	M258	I266	H373	Y392
17	H84	S90	P91	F100	D126	V155	A157	Y243	V279	T290	H399	A418
18	H84	S90	P91	F100	D126	V155	T157	Y243	V279	V290	H398	A417
19	Q82	N88	P89	F98	T121	V150	E152	Y237	M274	K285	H393	Q412
20	H88	N94	P95	F104	D127	V156	P158	F246	I282	S293	H399	K418
21	H84	N90	P91	F100	D123	V152	P154	Y242	M278	S289	H399	M418
22	H83	N89	P90	F99	D122	V151	P153	Y241	M277	H288	H396	M415
23	H83	N89	P90	F99	D122	V151	P153	F241	M277	H288	H397	M416

[0113] 表2

[0114] 至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸优选是至少一个插入在ssDNA中的核苷酸之间的氨基酸。可以对插入在ssDNA中核苷酸之间的氨基酸进行如上所述的建模。所述至少一个插入在ssDNA中的核苷酸之间的氨基酸优选为SEQ ID NO:8中的P89, F98和V150中的至少一个, 例如P89, F98, V150, P89/F98, P89/V150, F98/V150或P89/F98/V150。

[0115] 所述至少一个插入在SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的ssDNA中的核苷酸之间的氨基酸优选为至少一个与SEQ ID NO:8中的P89, F98和V150对应的氨基酸, 例如, P89, F98, V150, P89/F98, P89/V150, F98/V150或P89/F98/V150。相应的氨基酸如上表2所示。

[0116] 较大R基团

[0117] 较大侧链(R基团) 优选地 (a) 包含增加数量的碳原子 (b) 具有增加的长度 (c) 具有

增加的分子体积和/或(d)具有增加的范德华体积。较大侧链(R基团)优选(a);(b);(c);(d);(a)和(b);(a)和(c);(a)和(d);(b)和(c);(b)和(d);(c)和(d);(a),(b)和(c);(a),(b)和(d);(a),(c)和(d);(b),(c)和(d);或(a),(b),(c)和(d)。可以采用本领域中的标准方法对(a)至(d)中的每一个进行测量。

[0118] 较大侧链(R基团)优选地增加所述至少一个氨基酸与所述ssDNA中一个或多个核苷酸之间的(i)静电相互作用(ii)氢键和/或(iii)阳离子- π (阳离子- π)相互作用,比如增加(i),(ii),(iii);(i)和(ii);(i)和(iii);(ii)和(iii);以及(i),(ii)和(iii)。技术人员可以判断是否R基团增加了这些相互作用中的任何一个。例如,在(i)中,诸如精氨酸(R)、组氨酸(H)和赖氨酸(K)等带正电荷的氨基酸具有增加静电相互作用的R基团。例如,在(ii)中,诸如天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)和组氨酸(H)等氨基酸具有增加氢键的R基团。例如,在(iii)中,诸如苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)等芳族氨基酸具有增加阳离子- π (阳离子- π)相互作用的R基团。以下的具体取代被标记为(i)至(iii)以反映这些变化。其它可能的取代被标记为(iv)。这些(iv)取代通常增加侧链(R基团)的长度。

[0119] 包含较大侧链(R)的氨基酸可以是非天然氨基酸。所述非天然氨基酸可以是以下讨论的任何一种。

[0120] 包含较大侧链(R基团)的氨基酸优选不是丙氨酸(A)、半胱氨酸(C)、甘氨酸(G)、硒代半胱氨酸(U)、甲硫氨酸(M)、天冬氨酸(D)或谷氨酸(E)。

[0121] 组氨酸(H)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)或天冬酰胺(N)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。组氨酸(H)更优选被(a)N、Q或W或(b)Y、F、Q或K取代。

[0122] 天冬酰胺(N)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)或组氨酸(H)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。天冬酰胺(N)更优选被R、H、W或Y取代。

[0123] 脯氨酸(P)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)、苏氨酸(T)或组氨酸(H);(iii)酪氨酸(Y)、苯丙氨酸(F)或色氨酸(W)或(iv)亮氨酸(L)、缬氨酸(V)或异亮氨酸(I)取代。脯氨酸(P)更优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)、苏氨酸(T)或组氨酸(H);(iii)苯丙氨酸(F)或色氨酸(W)或(iv)亮氨酸(L)、缬氨酸(V)或异亮氨酸(I)取代。脯氨酸(P)更优选被(a)F,(b)L、V、I、T或F或(c)W、F、Y、H、I、L或V取代。

[0124] 缬氨酸(V)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代。缬氨酸(V)更优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代。缬氨酸(V)更优选被I或H或I、L、N、W或H取代。

[0125] 苯丙氨酸(F)优选被(i)精氨酸(R)或赖氨酸(K);(ii)组氨酸(H)或(iii)酪氨酸(Y)或色氨酸(W)取代。苯丙氨酸(F)更优选被(a)W,(b)W、Y或H,(c)W、R或K或(d)K、H、W或R取代。

[0126] 谷氨酰胺(Q)优选被(i)精氨酸(R)或赖氨酸(K)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。

[0127] 丙氨酸(A)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代。

[0128] 丝氨酸(S)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)异亮氨酸(I)或亮氨酸(L)取代。丝氨酸(S)优选被K、R、W或F取代。

[0129] 赖氨酸(K)优选被(i)精氨酸(R)或(iii)酪氨酸(Y)或色氨酸(W)取代。

[0130] 精氨酸(R)优选被(iii)酪氨酸(Y)或色氨酸(W)取代。

[0131] 甲硫氨酸(M)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。

[0132] 亮氨酸(L)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)或天冬酰胺(N)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。

[0133] 天冬氨酸(D)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。天冬氨酸(D)更优选被H、Y或K取代。

[0134] 谷氨酸(E)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H)或(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)取代。

[0135] 异亮氨酸(I)优选被(i)精氨酸(R)或赖氨酸(K);(ii)谷氨酰胺(Q)、天冬酰胺(N)或组氨酸(H);(iii)苯丙氨酸(F)、酪氨酸(Y)或色氨酸(W)或(iv)亮氨酸(L)取代。

[0136] 酪氨酸(Y)优选被(i)精氨酸(R)或赖氨酸(K)或(iii)色氨酸(W)取代。酪氨酸(Y)更优选被W或R取代。

[0137] 解旋酶更优选包含SEQ ID NO:8的变体,并且包含(a)P89F;(b)F98W;(c)V150I;(d)V150H;(e)P89F和F98W;(f)P89F和V150I;(g)P89F和V150H;(h)F98W和V150I;(i)F98W和V150H;(j)P89F、F98W和V150I或(k)P89F、F98W和V150H。

[0138] 解旋酶更优选包含SEQ ID NO:8的变体,其包含:

- H82N;	35	- N88R;	- N88Y;
- H82Q;		- N88H;	- P89L;
[0139] - H82W;		- N88W;	40 - P89V;
- P89I;		- S287R;	55 - F98W/V150W;
- P89E;		- S287W;	- F98W/V150H;

	- P89T;	30	- S287F;		- F98W/P152W;
	- P89F;		- H396Y;		- F98W/P152F;
	- D121H;		- H396F;		- F98W/P152Y;
	- D121Y;		- H396Q;	60	- F98W/P152H;
	- D121K;		- H396K;		- F98W/P152I;
	- V150I;	35	- Y415W;		- F98W/P152L;
	- V150L;		- Y415R;		- F98W/P152V;
	- V150N;		- F98W/H82N;		- F98W/F240W;
	- V150W;		- F98W/H82Q;	65	- F98W/F240Y;
	- V150H;		- F98W/H82W;		- F98W/F240H;
	- P152W;	40	- F98W/N88R;		- F98W/F276W;
	- P152F;		- F98W/N88H;		- F98W/F276R;
[0140]	- P152Y;		- F98W/N88W;		- F98W/F276K;
	- P152H;		- F98W/N88Y;	70	- F98W/F276H;
	- P152I;		- F98W/P89L;		- F98W/S287K;
	- P152L;	45	- F98W/P89V;		- F98W/S287R;
	- P152V;		- F98W/P89I;		- F98W/S287W;
	- F240W;		- F98W/P89T;		- F98W/S287F;
	- F240Y;		- F98W/P89F;	75	- F98W/H396Y;
	- F240H;		- F98W/D121H;		- F98W/H396F;
	- F276W;	50	- F98W/D121Y;		- F98W/H396Q;
	- F276R;		- F98W/D121K;		- F98W/Y415W;
	- F276K;		- F98W/V150I;		或者
	- F276H;		- F98W/V150L;	80	- F98W/Y415R。
	- S287K;		- F98W/V150N;		

[0141] 磷酸基团相互作用

[0142] 本发明的解旋酶优选是这样一种解旋酶,即,至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸被取代。可以取代任何数量的氨基酸,例如,1个或更多个、2个或更多个、3个或更多个、4个或更多个、5个或更多个或者6个或更多个氨基酸。ssDNA中的核苷酸各自包含三个磷酸基团。被取代的每个氨基每次可以与任何数量的磷酸基团相互作用,例如,每次一个、两个或三个磷酸基团。与一个或多个磷酸基团相互作用的氨基酸可以采用上述蛋白质建模来进行鉴定。

[0143] 取代优选增加所述至少一个氨基酸和所述ssDNA中一个或多个磷酸基团之间的(i)静电相互作用;(ii)氢键和/或(iii)阳离子- π (阳离子- π)相互作用。以下使用标记(i)、(ii)和(iii)对增加(i)、(ii)和(iii)的优选取代进行讨论。

[0144] 取代优选增加位点的净正电荷。可以采用本领域已知的方法来测量任何位点处的净电荷。例如,可采用等电点来限定氨基酸的净电荷。通常在约7.5处测量净电荷。取代优选是用带正电荷、不带电荷、非极性或芳族的氨基酸取代带负电荷的氨基酸。带负电荷的氨基酸是带有净负电荷的氨基酸。带负电荷的氨基酸包括但不限于天冬氨酸(D)和谷氨酸(E)。带正电荷的氨基酸是带有净正电荷的氨基酸。带正电荷的氨基酸可以是天然存在的或非天

然存在的。带正电的氨基酸可以是合成的或修饰的。例如，带有净正电荷的修饰氨基酸可以专门设计用于本发明。对氨基酸进行的多种不同类型的修饰是本领域公知的。优选的天然存在的带正电荷的氨基酸包括但不限于组氨酸(H)、赖氨酸(K)和精氨酸(R)。

[0145] 不带电荷的氨基酸、非极性氨基酸或者芳族氨基酸可以是天然存在的或非天然存在的。其可以是合成的或修饰的。不带电荷的氨基酸没有净电荷。合适的带电氨基酸包括但不限于半胱氨酸(C)、丝氨酸(S)、苏氨酸(T)、甲硫氨酸(M)、天冬酰胺(N)和谷氨酰胺(Q)。非极性氨基酸具有非极性侧链。合适的非极性氨基酸包括但不限于甘氨酸(G)、丙氨酸(A)、脯氨酸(P)、异亮氨酸(I)、亮氨酸(L)和缬氨酸(V)。芳族氨基酸具有芳族侧链。合适的芳族氨基酸包括但不限于组氨酸(H)、苯丙氨酸(F)、色氨酸(W)和酪氨酸(Y)。

[0146] 解旋酶优选包含SEQ ID NO:8的变体,其中,至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸是H64,T80,S83,N242,K243,N293,T394和K397中的至少一个。这些数字对应SEQ ID NO:8中的相关位点,并且相比SEQ ID NO:8可能需要在在一个或多个氨基酸已经在变体中插入或删除的情况下进行改变。本领域技术人员可以确定如上所述的变体中的相应位点。

[0147] 解旋酶优选包含SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其中,所述至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸是与SEQ ID NO:8中的H64,T80,S83,N242,K243,N293,T394和K397相对应的至少一个氨基酸。

[0148] 表3示出了与SEQ ID NO:8中的H64,T80,S83,N242,K243,N293,T394和K397对应的SEQ ID NO:9至23中的氨基酸。

[0149]

SEQ ID NO:								
8	H64	T80	S83	N242	K243	N293	T394	K397
9	G63	T79	S82	N276	A277	N317	T426	K429
10	N121	T142	K145	N309	K310	N361	T453	K456
11	H66	T85	S88	N242	K243	N286	T375	R378
12	H89	T110	S113	N275	K276	V315	T412	K415
13	H68	T91	S94	N269	A270	N308	T398	K401
14	H56	T72	S75	N232	D233	N279	T376	K379
15	H60	T76	S79	N237	K238	N285	T385	K388
16	H47	T63	S66	N224	K225	N272	T371	K374
17	H66	T82	S85	N245	D246	N296	T397	K400

[0150]

18	H66	T82	S85	N245	A246	N296	T396	K399
19	H64	T80	S83	N239	N240	N291	T391	K394
20	H70	T86	S89	N248	K249	N299	T397	K400
21	H66	T82	S85	N244	K245	N295	T397	K400
22	H65	T81	S84	N243	K244	N294	T394	K397
23	H65	T81	K84	N243	K244	N294	T395	K398

[0151] 表3

[0152] 组氨酸(H)优选被(i)精氨酸(R)或赖氨酸(K);(ii)天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)或苏氨酸(T);(iii)苯丙氨酸(F)、色氨酸(W)或酪氨酸(Y)取代。组氨酸(H)优选被(a)N、Q、K或F或者(b)N、Q或W取代。

[0153] 苏氨酸(T)优选被(i)精氨酸(R)、组氨酸(H)或赖氨酸(K);(ii)天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)或组氨酸(H)或者(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代。苏氨酸(T)更优选被(a)K、Q或N或者(b)K、H或N取代。

[0154] 丝氨酸(S)优选被(i)精氨酸(R)、组氨酸(H)或赖氨酸(K);(ii)天冬酰胺(N)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或者(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代。丝氨酸(S)更优选被H、N、K、T、R或Q取代。

[0155] 天冬酰胺(N)优选被(i)精氨酸(R)、组氨酸(H)或赖氨酸(K);(ii)丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或者(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代。天冬酰胺(N)更优选被(a)H或Q或者(b)Q、K或H取代。

[0156] 赖氨酸(K)优选被(i)精氨酸(R)或组氨酸(H);(ii)天冬酰胺(N)、丝氨酸(S)、谷氨酰胺(Q)、苏氨酸(T)或组氨酸(H)或者(iii)苯丙氨酸(F)、色氨酸(W)、酪氨酸(Y)或组氨酸(H)取代。赖氨酸(K)更优选被(a)Q或H或者(b)R、H或Y取代。

[0157] 解旋酶更优选包含SEQ ID NO:8的变体,并且包含例如(a)H64N,H64Q,H64K或H64F;(b)T80K,T80Q或T80N;(c)S83H,S83N,S83K,S83T,S83R或S83Q;(d)N242H或N242Q;(e)K243Q或K243H;(f)N293Q,N293K或N293H;(g)T394K,T394H或T394N或(h)K397R,K397H或K397Y中的所有的一个或多个。

[0158] 组合

[0159] 解旋酶可以是这样一种解旋酶,其中,(a)至少一个与ssDNA中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸被包含较大侧链(R基团)的氨基酸取代;和(b)至少一个与ssDNA中一个或多个核苷酸的一个或多个磷酸基团相互作用的氨基酸被取代。解旋酶优选包含:

[0160] (a)SEQ ID NO:8的变体,其包含如上限定的F98上的取代,以及在H64,T80,S83,N242,K243,N293,T394和K397中一个或多个上的取代;或者

[0161] (b)SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其包含在与F98对应的氨基酸上的取代和在一个或多个与SEQ ID NO:8中H64,T80,S83,N242,K243,N293,T394和K397对应的氨基酸上的取代。

[0162] 解旋酶优选是SEQ ID NO:8的变体,其包含以下位点上的取代:

- [0163] -F98/H64, 例如, F98W/H64N, F98W/H64Q, F98W/H64K或者F98W/H64F;
- [0164] -F98/T80, 例如, F98W/T80K, F98W/T80Q, F98W/T80N;
- [0165] -F98/H82, 例如, F98W/H82N, F98W/H82Q或者F98W/H82W;
- [0166] -F98/S83, 例如, F98W/S83H, F98W/S83N, F98W/S83K, F98W/S83T, F98W/S83R或者F98W/S83Q;
- [0167] -F98/N242, 例如, F98W/N242H, F98W/N242Q, F98W/K243Q或者F98W/K243H;
- [0168] -F98/N293, 例如, F98W/N293Q, F98W/N293K, F98W/N293H, F98W/T394K,
- [0169] F98W/T394H, F98W/T394N, F98W/H396Y, F98W/H396F, F98W/H396Q或者F98W/H396K;
或者
- [0170] -F98/K397, 例如, F98W/K397R, F98W/K397H或者F98W/K397Y。
- [0171] SEQ ID NO:9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22或23中的优选组合包括与上面列出的SEQ ID NO:8中的组合对应的氨基酸的组合。
- [0172] 孔相互作用
- [0173] 本发明的解旋酶还是这样一种解旋酶, 其中, 与跨膜孔相互作用的解旋酶的一部分包含一个或多个修饰, 优选一个或多个取代。所述与跨膜孔相互作用的解旋酶的一部分通常是当解旋酶用于控制多核苷酸穿过孔的移动时与跨膜孔相互作用的解旋酶的一部分, 例如, 以下进行更详细的讨论。当解旋酶用于控制多核苷酸穿过孔的移动时, 所述一部分通常包含与孔相互作用或接触的氨基酸, 例如, 以下进行更详细的讨论。在施加电势下, 当解旋酶与正移动穿过孔的多核苷酸结合或连接时, 所述一部分通常包含与孔相互作用或接触的氨基酸。
- [0174] 在SEQ ID NO:8中, 与跨膜孔相互作用的一部分通常包含以下位点上的氨基酸, 即, 1, 2, 3, 4, 5, 6, 51, 176, 177, 178, 179, 180, 181, 185, 189, 191, 193, 194, 195, 197, 198, 199, 200, 201, 202, 203, 204, 207, 208, 209, 210, 211, 212, 213, 216, 219, 220, 221, 223, 224, 226, 227, 228, 229, 247, 254, 255, 256, 257, 258, 259, 260, 261, 298, 300, 304, 308, 318, 319, 321, 337, 347, 350, 351, 405, 415, 422, 434, 437, 438。这些数字对应SEQ ID NO:8中的相关位点, 并且相比SEQ ID NO:8可能需要在的一个或多个氨基酸已经在变体中插入或删除的情况下进行改变。如上所述, 本领域技术人员可以确定变体中的相应位点。与跨膜孔相互作用的一部分优选包含以下位点上的氨基酸:
- [0175] (a) SEQ ID NO:8中的位点: 1, 2, 4, 51, 177, 178, 179, 180, 185, 193, 195, 197, 198, 199, 200, 202, 203, 204, 207, 208, 209, 210, 211, 212, 216, 221, 223, 224, 226, 227, 228, 229, 254, 255, 256, 257, 258, 260, 304, 318, 321, 347, 350, 351, 405, 415, 422, 434, 437和438; 或者
(b) SEQ ID NO:8中的位点: 1, 2, 178, 179, 180, 185, 195, 197, 198, 199, 200, 202, 203, 207, 209, 210, 212, 216, 221, 223, 226, 227, 255, 258, 260, 304, 350和438。
- [0176] 与跨膜孔相互作用的一部分优选包含SEQ ID NO:8中的位点K194, W195, K198, K199和E258上的一个或多个氨基酸, 例如, 2个、3个、4个或5个。SEQ ID NO:8的变体优选包含在(a) K194; (b) W195; (c) D198; (d) K199和(e) E258中一个或多个上的修饰。SEQ ID NO:8的变体优选包含: (a) K194, 例如, K194L; (b) W195, 例如, W195A; (c) D198, 例如, D198V; (d) K199, 例如, K199L和(e) E258, 例如, E258L; 中一个或多个上的取代。变体可以包括{a}; {b}; {c}; {d}; {e}; {a, b}; {a, c}; {a, d}; {a, e}; {b, c}; {b, d}; {b, e}; {c, d}; {c, e}; {d, e}; {a, b,

c} ; {a,b,d} ; {a,b,e} ; {a,c,d} ; {a,c,e} ; {a,d,e} ; {b,c,d} ; {b,c,e} ; {b,d,e} ; {c,d,e} ; {a,b,c,d} ; {a,b,c,e} ; {a,b,d,e} ; {a,c,d,e} ; {b,c,d,e} ; 或者 {a,b,c,d,e}。当修饰的多核苷酸结合蛋白与衍生自MspA的孔,特别是以下讨论的任何修饰孔相互作用时,本段中阐述的修饰或取代是优选的。

[0177] 与跨膜孔相互作用的多核苷酸结合蛋白的一部分优选包含SEQ ID NO:8中位点194或199上的氨基酸。变体优选包含K194A,K194V,K194F,K194D,K194S,K194W或者K194L和/或K199A,K199V,K199F,K199D,K199S,K199W或者K199L。在SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23中,与跨膜孔相互作用的一部分通常包含与上述SEQ ID NO:8中的那些位点对应的位点上的氨基酸。可以采用以下比对(alignment)方式鉴定与SEQ ID NO:9中的那些位点对应的SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22和23中的氨基酸。

[0178]

SEQ ID NO:8	K194	W195	D198	K199	E258
SEQ ID NO:9	L230	E231	H234	Y235	R293
SEQ ID NO:10	W259	N260	T263	Y264	E326
SEQ ID NO:11	A192	D193	F196	G197	A259
SEQ ID NO:12	I224	K225	D228	F229	Q292
SEQ ID NO:13	Q213	D214	Y217	A218	A286
SEQ ID NO:14	Q185	W186	T189	N190	N248
SEQ ID NO:15	G190	W191	P194	N195	K253
SEQ ID NO:16	G177	W178	Q181	N182	K240
SEQ ID NO:17	K200	M201	P204	M205	K261
SEQ ID NO:18	K200	P201	P204	L205	K261
SEQ ID NO:19	K193	W194	E197	K198	A256
SEQ ID NO:20	N200	W201	E204	N205	N264
SEQ ID NO:21	G196	W197	D200	C201	E260
SEQ ID NO:22	G195	W196	E199	N200	E259
SEQ ID NO:23	S195	W196	E199	K200	Q259

[0179] 优选组合

[0180] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含F98上的,例如,F98R,F98K,F98Q,F98N,F98H,F98Y,F98F或F98W上的取代;以及K194上的,例如,K194A,K194V,K194F,K194D,K194S,K194W或K194L上的取代;和/或K199上的,例如,K199A,K199V,K199F,K199D,K199S,K199W或K199L上的取代。本发明的解旋酶优选包含SEQ ID NO:9,10,11,12,13,14,15,16,17,18,19,20,21,22或23的变体,其包含在与SEQ ID NO:8中的F98相对应的位点上的取代以及在与SEQ ID NO:8中的K194和/或K199相对应的位点上的取代。这些相应的位点可以被SEQ ID NO:8中F98,K194和K199的上述任何氨基酸取代。

[0181] 解旋酶优选是SEQ ID NO:8的变体,其包含以下位点上的取代:

[0182] -F98/K194/H64,例如,F98W/K194L/H64N,F98W/K194L/H64Q,F98W/K194L/H64K或者F98W/K194L/H64F;

- [0183] -F98/K194/T80, 例如, F98W/K194L/T80K, F98W/K194L/T80Q或者F98W/K194L/T80N;
- [0184] -F98/K194/H82, 例如, F98W/K194L/H82N, F98W/K194L/H82Q或者F98W/K194L/H82W
- [0185] -F98/S83/K194, 例如, F98W/S83H/K194L, F98W/S83T/K194L, F98W/S83R/K194L, F98W/S83Q/K194L, F98W/S83N/K194L, F98W/S83K/K194L, F98W/N88R/K194L, F98W/N88H/K194L, F98W/N88W/K194L或者F98W/N88Y/K194L;
- [0186] -F98/S83/K194/F276, 例如, F98W/S83H/K194L/F276K;
- [0187] -F98/P89/K194, 例如, F98W/P89L/K194L, F98W/P89V/K194L, F98W/P89I/K194L或者F98W/P89T/K194L;
- [0188] -F98/D121/K194, 例如, F98W/D121H/K194L, F98W/D121Y/K194L或者F98W/D121K/K194L;
- [0189] -F98/V150/K194, 例如, F98W/V150I/K194L, F98W/V150L/K194L, F98W/V150N/K194L, F98W/V150W/K194L或者F98W/V150H/K194L;
- [0190] -F98/P152/K194, 例如, F98W/P152W/K194L, F98W/P152F/K194L, F98W/P152Y/K194L, F98W/P152H/K194L, F98W/P152I/K194L, F98W/P152L/K194L或者F98W/P152V/K194L;
- [0191] -F98/F240/K194, 例如, F98W/F240W/K194L, F98W/F240Y/K194L或者F98W/F240H/K194L;
- [0192] -F98/N242/K194, 例如, F98W/N242H/K194L或者F98W/N242Q/K194L;
- [0193] -F98/K194/F276, 例如, F98W/K194L/F276K, F98W/K194L/F276H, F98W/K194L/F276W或者F98W/K194L/F276R;
- [0194] -F98/K194/S287, 例如, F98W/K194L/S287K, F98W/K194L/S287R, F98W/K194L/S287W或者F98W/K194L/S287F;
- [0195] -F98/N293/K194, 例如, F98W/N293Q/K194L, F98W/N293K/K194L或者F98W/N293H/K194L;
- [0196] -F98/T394/K194, 例如, F98W/T394K/K194L, F98W/T394H/K194L或者F98W/T394N/K194L;
- [0197] -F98/H396/K194, 例如, F98W/H396Y/K194L, F98W/H396F/K194L, F98W/H396Q/K194L或者F98W/H396K/K194L;
- [0198] -F98/K397/K194, 例如, F98W/K397R/K194L, F98W/K397H/K194L或者F98W/K397Y/K194L;或者
- [0199] -F98/Y415/K194, 例如, F98W/Y415W/K194L或者F98W/Y415R/K194L。
- [0200] 在任何以上组合中, 可以用W195, D198, K199和E258中的任一个取代K194。
- [0201] 解旋酶优选是SEQ ID NO:9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22或23的变体, 其包含在与上述SEQ ID NO:8中的组合相对应的氨基酸上的取代。
- [0202] 塔域和/或销域和/或1A域中的修饰
- [0203] 本发明的解旋酶优选是这样一种解旋酶, 其中, 至少一个半胱氨酸残基(即, 一个或多个半胱氨酸残基)和/或至少一个非天然氨基酸(即, 一个或多个非天然氨基酸)被引入i) 塔域和/或(ii) 销域和/或(iii) 1A (RecA式马达) 域, 其中, 解旋酶具有控制多核苷酸的移动的能力。这些类型的修饰在申请号为PCT/GB2014/052736 (W02015/055981) 的国际申请中

公开。可以将至少一个半胱氨酸残基和/或至少一个非天然氨基酸引入塔域；销域；1A域；塔域和销域；塔域和1A域或者塔域、销域和1A域。

[0204] 本发明的解旋酶优选是这样一种解旋酶，其中，至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入 (i) 塔域和 (ii) 销域和/或1A (RecA式马达) 域中的每个，即，引入塔域和销域；塔域和1A域或者塔域、销域和1A域。

[0205] 可以将任何数量的半胱氨酸残基和/或非天然氨基酸引入每个域。例如，可以引入1,2,3,4,5,6,7,8,9,10或更多个半胱氨酸残基，和/或可以引入1,2,3,4,5,6,7,8,9,10或更多个非天然氨基酸。可以仅引入一个或多个半胱氨酸残基。可以仅引入一个或多个非天然氨基酸。可以引入一个或多个半胱氨酸残基和一个或多个非天然氨基酸的组合。

[0206] 优选通过取代引入所述至少一个半胱氨酸残基和/或至少一个非天然氨基酸。这样做的方法是本领域已知的。

[0207] 这些修饰并不妨碍解旋酶与多核苷酸结合。这些修饰降低了多核苷酸从解旋酶解链或脱离的能力。换言之，所述一个或多个修饰通过防止从多核苷酸链解离而增加了解旋酶的持续合成性 (processivity) 能力。所述一个或多个修饰通常还增加了酶的热稳定性，所述修饰给其提供了改善的结构稳定性，从而有利于链测序。

[0208] 非天然氨基酸是在解旋酶中不天然存在的氨基。非天然氨基酸优选不是组氨酸、丙氨酸、异亮氨酸、精氨酸、亮氨酸、天冬酰胺、赖氨酸、天冬氨酸、甲硫氨酸、半胱氨酸、苯丙氨酸、谷氨酸、苏氨酸、谷氨酰胺、色氨酸、甘氨酸、缬氨酸、脯氨酸、丝氨酸或酪氨酸。非天然氨基酸更优选地不是前句话中的二十种氨基酸中的任一个或硒代半胱氨酸。

[0209] 本发明中使用的优选非天然氨基酸包括但不限于4-叠氮基-L-苯丙氨酸 (Faz)、4-乙酰基-L-苯丙氨酸、3-乙酰基-L-苯丙氨酸、4-乙酰乙酰基-L-苯丙氨酸、0-烯丙基-L-酪氨酸、3-(苯硒基)-L-丙氨酸、0-2-丙炔-1-基-L-酪氨酸、4-(二羟基硼基)-L-苯丙氨酸、4-[(乙基硫烷基)羰基]-L-苯丙氨酸、(2S)-2-氨基-3-4-[(丙烷-2-基硫烷基)羰基]苯基；丙氨酸、(2S)-2-氨基-3-4-[(2-氨基-3-硫烷基丙酰基)氨基]苯基；丙氨酸、0-甲基-L-酪氨酸、4-氨基-L-苯丙氨酸、4-氰基-L-苯丙氨酸、3-氰基-L-苯丙氨酸、4-氟-L-苯丙氨酸、4-碘-L-苯丙氨酸、4-溴-L-苯丙氨酸、0-(三氟甲基)酪氨酸、4-硝基-L-苯丙氨酸、3-羟基-L-酪氨酸、3-氨基-L-酪氨酸、3-碘代-L-酪氨酸、4-异丙基-L-苯丙氨酸、3-(2-萘基)-L-丙氨酸、4-苯基-L-苯丙氨酸、(2S)-2-氨基-3-(萘-2-基氨基)丙氨酸、6-(甲基硫烷基)正亮氨酸、6-氧代-L-赖氨酸、D-酪氨酸、(2R)-2-羟基-3-(4-羟基苯基)丙氨酸、(2R)-2-氨基辛酸酯3-(2,2'-双吡啶-5-基)-D-丙氨酸、2-氨基-3-(8-羟基-3-喹啉基)丙氨酸、4-苯甲酰基-L-苯丙氨酸、S-(2-硝基苄基)半胱氨酸、(2R)-2-氨基-3-[(2-硝基苄基)硫烷基]丙氨酸、(2S)-2-氨基-3-[(2-硝基苄基)氧基]丙氨酸、0-(4,5-二甲氧基-2-硝基苄基)-L-丝氨酸、(2S)-2-氨基-6-[(2-硝基苄基)氧基]羰基；氨基)己酸、0-(2-硝基苄基)-L-酪氨酸、2-硝基苯丙氨酸、4-[(E)-苯基二氮烯基]-L-苯丙氨酸、4-[3-(三氟甲基)-3H-双吡丙啶-3-基]-D-苯丙氨酸 (4-[3-(Trifluoromethyl)-3H-diaziren-3-yl]-D-phenylalanine)、2-氨基-3-[[5-(二甲基氨基)-1-萘基]磺酰基氨基]丙氨酸、(2S)-2-氨基-4-(7-羟基-2-氧代-2H-色烯-4-基)丁氨酸、(2S)-3-[(6-乙酰萘-2-基)氨基]-2-氨基丙氨酸、4-(羧甲基)苯丙氨酸、3-硝基-L-酪氨酸、0-磺基-L-酪氨酸、(2R)-6-乙酰胺基-2-氨基己酸酯、1-甲基组氨酸、2-氨基壬酸、2-氨基癸酸、L-高半胱氨酸、5-硫烷基正缬氨酸、6-硫烷基-L-正亮氨酸5-(甲基硫烷基)-L-正缬氨

酸、N⁶-[(2R,3R)-3-甲基-3,4-二氢-2H-吡咯-2-基]羰基;-L-赖氨酸、N⁶-[(苄氧基)羰基]赖氨酸、(2S)-2-氨基-6-[(环戊基羰基)氨基]己酸、N⁶-[(环戊氧基)羰基]-L-赖氨酸、(2S)-2-氨基-6-[(2R)-四氢呋喃-2-基羰基]氨基;己酸、(2S)-2-氨基-8-[(2R,3S)-3-乙炔基四氢呋喃-2-基]-8-氧代辛酸、N⁶-(叔丁氧基羰基)-L-赖氨酸、(2S)-2-羟基-6-[(2-甲基-2-丙酰基)氧基]羰基;氨基)己酸、N⁶-[(烯丙氧基)羰基]赖氨酸、(2S)-2-氨基-6-[(2-叠氮基苄基)氧基]羰基;氨基)己酸、N⁶-L-脯氨酰基-L-赖氨酸、(2S)-2-氨基-6-[(丙-2-炔-1-基氧基)羰基]氨基;己酸和N⁶-[(2-叠氮基乙氧基)羰基]-L-赖氨酸。最优选的非天然氨基酸是4-叠氮基-L-苯丙氨酸(Faz)。

[0210] 以下表4(分为两部分)对每个Dda同源物(SEQ ID NO:8至23)中构成每个域的残基进行了鉴定。

[0211]

同源物	SEQ ID NO	1A	2A
Dda-Rma-DSM	9	M1-I84+R113-Y211	R212-E294+G422-S678
Dda-Csp	10	M1-L147+S166-V240	R241-N327+A449-G496
Dda-Sru	11	M1-L90+E108-H173	R174-D260+A371-V421
Dda-Sgo	12	M1-L115+N136-V205	R206-K293+I408-L500
Dda-Vph12B8	13	M1-L96+F114-V194	R195-D287+V394-Q450
Dda-Vph	14	M1-L77+V96-V166	R167-T249+L372-N421
Dda-Aph65	15	M1-M81+L99-M171	R172-T254+L381-K434
Dda-AphCC2	16	M1-M68+M86-M158	R159-T241+L367-K420
Dda-Cph	17	M1-L87+A108-M181	R182-T262+L393-V443
Dda-Kph	18	M1-L87+A108-M181	R182-T262+L392-V442
Dda-SphIME13	19	M1-L85+T103-K176	R177-N257+L387-V438
Dda-AphAc42	20	M1-L91+V109-M183	R184-T265+L393-I442
Dda-SphSP18	21	M1-L87+M105-M179	R180-T261+L393-V442
Dda-Yph	22	M1-L86+V104-K178	R179-T260+L390-I439
Dda-SphS16	23	M1-L86+V104-M178	R179-T260+L391-V441
Dda-1993	8	M1-L85+V103-K177	R178-T259+L390-V439

[0212]

同源物	SEQ ID	塔	销	钩
Dda-Rma-DSM	9	G295-N309+F316-Y421	Y85-L112	A310-L315
Dda-Csp	10	V328-P342+N360-Y448	K148-N165	V343-L359
Dda-Sru	11	A261-T275+T285-Y370	G91-E107	W276-L284
Dda-Sgo	12	G294-I307+T314-Y407	G116-T135	R308-Y313
Dda-Vph12B8	13	V288-E301+N307-N393	G97-P113	M302-W306
Dda-Vph	14	S250-P264+E278-S371	K78-E95	V265-I277
Dda-Aph65	15	K255-P269+T284-S380	K82-K98	V270-F283
Dda-AphCC2	16	D242-P256+T271-S366	K69-K85	V257-F270
Dda-Cph	17	T263-P277+N295-P392	K88-K107	L278-Y294

Dda-Kph	18	D263-P277+N295-A391	K88-K107	L278-Y294
Dda-SphIME13	19	A258-P272+N290-P386	K86-G102	L273-F289
Dda-AphAc42	20	L266-P280+N298-A392	K92-D108	L281-F297
Dda-SphSP18	21	D262-P276+N294-A392	K88-E104	H277-F293
Dda-Yph	22	D261-P275+N293-A389	K87-E103	L276-F292
Dda-SphS16	23	E261-P275+T293-A390	K87-E103	L276-F292
Dda-1993	8	D260-P274+N292-A389	K86-E102	L275-F291

[0213] 表4

[0214] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基D260-P274和N292-A389)和/或(ii)销域(残基K86-E102)和/或(iii)1A域(残基M1-L85和V103-K177)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N292-A389中。

[0215] 本发明的解旋酶优选包含SEQ ID NO:9的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基G295-N309和F316-421)和/或(ii)销域(残基Y85-L112)和/或(iii)1A域(残基M1-I84和R113-Y211)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基F316-Y421中。

[0216] 本发明的解旋酶优选包含SEQ ID NO:10的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基V328-P342和N360-Y448)和/或(ii)销域(残基K148-N165)和/或(iii)1A域(残基M1-L147和S166-V240)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N360-Y448中。

[0217] 本发明的解旋酶优选包含SEQ ID NO:11的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基A261-T275和T285-Y370)和/或(ii)销域(残基G91-E107)和/或(iii)1A域(残基M1-L90和E108-H173)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基T285-Y370中。

[0218] 本发明的解旋酶优选包含SEQ ID NO:12的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基G294-I307和T314-Y407)和/或(ii)销域(残基G116-T135)和/或(iii)1A域(残基M1-L115和N136-V205)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基T314-Y407中。

[0219] 本发明的解旋酶优选包含SEQ ID NO:13的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基V288-E301和N307-N393)和/或(ii)销域(残基G97-P113)和/或(iii)1A域(残基M1-L96和F114-V194)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N307-N393中。

[0220] 本发明的解旋酶优选包含SEQ ID NO:14的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基S250-P264和E278-S371)和/或(ii)销域(残基K78-E95)和/或(iii)A域(残基M1-L77和V96-V166)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基E278-S371中。

[0221] 本发明的解旋酶优选包含SEQ ID NO:15的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入到(i)塔域(残基K255-P269和T284-S380)和/或(ii)销域(残基K82-K98)和/或(iii)1A域(残基M1-M81和L99-M171)。至少一个半胱氨酸残基和/或至

少一个非天然氨基酸优选被引入到塔域的残基T284-S380中。

[0222] 本发明的解旋酶优选包含SEQ ID NO:16的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基D242-P256和T271-S366)和/或(ii)销域(残基K69-K85)和/或(iii)A域(残基M1-M68和M86-M158)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基T271-S366中。

[0223] 本发明的解旋酶优选包含SEQ ID NO:17的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基T263-P277和N295-P392)和/或(ii)销域(残基K88-K107)和/或(iii)1A域(残基M1-L87和A108-M181)。优选将所述至少一个半胱氨酸残基和/或至少一个非天然氨基酸引入塔域的残基N295-P392中。

[0224] 本发明的解旋酶优选包含SEQ ID NO:18的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基D263-P277和N295-A391)和/或(ii)销域(残基K88-K107)和/或(iii)1A域(残基M1-L87和A108-M181)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N295-A391中。

[0225] 本发明的解旋酶优选包含SEQ ID NO:19的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基A258-P272和N290-P386)和/或(ii)销域(残基K86-G102)和/或(iii)1A域(残基M1-L85和T103-K176)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N290-P386中。

[0226] 本发明的解旋酶优选包含SEQ ID NO:20的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基L266-P280和N298-A392)和/或(ii)销域(残基K92-D108)和/或(iii)1A域(残基M1-L91和V109-M183)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N298-A392中。

[0227] 本发明的解旋酶优选包含SEQ ID NO:21的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基D262-P276和N294-A392)和/或(ii)销域(残基K88-E104)和/或(iii)1A域(残基M1-L87和M105-M179)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N294-A392中。

[0228] 本发明的解旋酶优选包含SEQ ID NO:22的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基D261-P275和N293-A389)和/或(ii)销域(残基K87-E103)和/或(iii)1A域(残基M1-L86和V104-K178)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基N293-A389中。

[0229] 本发明的解旋酶优选包含SEQ ID NO:23的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(残基E261-P275和T293-A390)和/或(ii)销域(残基K87-E103)和/或(iii)1A域(残基M1-L86和V104-M178)。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到塔域的残基T293-A390中。

[0230] 本发明的解旋酶优选包含SEQ ID NO:8至23中任一项的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域和(ii)销域和/或1A域。本发明的解旋酶更优选包含SEQ ID NO:8至23中任一项的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入(i)塔域(ii)销域和(iii)1A域。如上所述,可以引入任何数量和组合的半胱氨酸残基和非天然氨基酸。

[0231] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含(i)E94C和/或A360C;(ii)

E93C和/或K358C; (iii) E93C和/或A360C; (iv) E93C和/或E361C; (v) E93C和/或K364C; (vi) E94C和/或L354C; (vii) E94C和/或K358C; (viii) E93C和/或L354C; (ix) E94C和/或E361C; (x) E94C和/或K364C; (xi) L97C和/或L354C; (xii) L97C和/或K358C; (xiii) L97C和/或A360C; (xiv) L97C和/或E361C; (xv) L97C和/或K364C; (xvi) K123C和/或L354C; (xvii) K123C和/或K358C; (xviii) K123C和/或A360C; (xix) K123C和/或E361C; (xx) K123C和/或K364C; (xxi) N155C和/或L354C; (xxii) N155C和/或K358C; (xxiii) N155C和/或A360C; (xxiv) N155C和/或E361C; (xxv) N155C和/或K364C; (xxvi) (i) 至 (xxv) 和G357C中的任一个; (xxvii) (i) 至 (xxv) 和Q100C中的任一个; (xxviii) (i) 至 (xxv) 和I127C中的任一个; (xxix) (i) 至 (xxv) 和Q100C和I127C中的任一个; (xxx) E94C和/或F377C; (xxxi) N95C; (xxxii) T91C; (xxxiii) Y92L, E94Y, Y350N, A360C和Y363N; (xxxiv) E94Y和A360C; (xxxv) A360C; (xxxvi) Y92L, E94C, Y350N, A360Y和Y363N; (xxxvii) Y92L, E94C和A360Y; (xxxviii) E94C和/或A360C和F276A; (xxxix) E94C和/或L356C; (xl) E93C和/或E356C; (xli) E93C和/或G357C; (xlii) E93C和/或A360C; (xliii) N95C和/或W378C; (xliv) T91C和/或S382C; (xlv) T91C和/或W378C; (xlvi) E93C和/或N353C; (xlvii) E93C和/或S382C; (xlviii) E93C和/或K381C; (xlix) E93C和/或D379C; (l) E93C和/或S375C; (li) E93C和/或W378C; (lii) E93C和/或W374C; (liii) E94C和/或N353C; (liv) E94C和/或S382C; (lv) E94C和/或K381C; (lvi) E94C和/或D379C; (lvii) E94C和/或S375C; (lviii) E94C和/或W378C; (lix) E94C和/或W374C; (lx) E94C和A360Y; (lxi) E94C, G357C和A360C或 (lxii) T2C, E94C和A360C。在 (i) 至 (lxii) 中的任一个中, 和/或优选为和。

[0232] 本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体, 其包含与 (i) 至 (lxii) 中任一项所定义的SEQ ID NO:8中的那些位点相对应的位点上的半胱氨酸残基。可以采用以下SEQ ID NO:8至23的比对来鉴定与SEQ ID NO:8中的那些位点相对应的SEQ ID NO:9至23中任一个的位点。本发明的解旋酶优选包含SEQ ID NO:11的变体, 其包含 (a) D99C和/或L341C; (b) Q98C和/或L341C或 (d) Q98C和/或A340C。本发明的解旋酶优选包含SEQ ID NO:15的变体, 其包含D90C和/或A349C。本发明的解旋酶优选包含SEQ ID NO:21的变体, 其包含D96C和/或A362C。

[0233] 本发明的解旋酶优选包含与 (i) 至 (lxii) 中任一项所定义的SEQ ID NO:8至23中任一个的变体, 其中, Faz代替半胱氨酸被引入到一个或多个特定位点上。Faz代替半胱氨酸可以被引入到每个特定位点上。本发明的解旋酶优选包含SEQ ID NO:8的变体, 其包含 (i) E94Faz和/或A360C; (ii) E94C和/或A360Faz; (iii) E94Faz和/或A360Faz; (iv) Y92L, E94Y, Y350N, A360Faz和Y363N; (v) A360Faz; (vi) E94Y和A360Faz; (vii) Y92L, E94Faz, Y350N, A360Y和Y363N; (viii) Y92L, E94Faz和A360Y; (ix) E94Faz和A360Y; 和 (x) E94C, G357Faz和A360C。

[0234] 本发明的解旋酶优选还包含一个或多个来自销域的单一氨基酸缺失。可以制备任何数量的单一氨基酸缺失, 例如1, 2, 3, 4, 5, 6, 7, 8, 9, 10或更多个。解旋酶更优选包含SEQ ID NO:8的变体, 其包含E93的缺失、E95的缺失或E93和E95的缺失。解旋酶更优选包含SEQ ID NO:8的变体, 其包含 (a) E94C, N95的缺失和A360C; (b) E93的缺失, E94的缺失, N95的缺失, 和A360C; (c) E93的缺失, E94C, N95的缺失和A360C或 (d) E93C, N95的缺失和A360C。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体, 其包含与SEQ ID NO:8中的E93对

应的位点的缺失、与SEQ ID NO:8中的E95对应的位点的缺失或与SEQ ID NO:8中的E93和E95对应的位点的缺失。

[0235] 本发明的解旋酶优选还包含一个或多个来自钩域的单一氨基酸缺失。可以制备任何数量的单一氨基酸缺失,例如1,2,3,4,5,6,7,8,9,10或更多个。解旋酶更优选包含SEQ ID NO:8的变体,其包含任何数量的位点T278至S287的缺失。解旋酶更优选包含SEQ ID NO:8的变体,其包含(a) E94C, Y279至K284的缺失和A360C; (b) E94C, T278的缺失, Y279, V286和S287以及A360C; (c) E94C, I281的缺失和K284和用单个G的取代和A360C; (d) E94C, K280的缺失和P2845和用单个G的取代和A360C; 或(e) Y279至K284的缺失、E94C、F276A和A230C。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,其包含与SEQ ID NO:8中的278至287对应位点的任何数量的缺失。

[0236] 本发明的解旋酶优选还包含一个或多个来自销域的单一氨基酸缺失和一个或多个来自钩域的单一氨基酸缺失。

[0237] 本发明的解旋酶优选是这样一种解旋酶,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域和/或2A (RecA式) 域中。如上所述,任何数量和组合的半胱氨酸残基和非天然氨基酸可以被引入用于塔、销和1A域。

[0238] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L275-F291) 和/或2A (RecA式) 域(残基R178-T259和L390-V439) 中。

[0239] 本发明的解旋酶优选包含SEQ ID NO:9的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基A310-L315) 和/或2A (RecA式) 域(残基R212-E294和G422-S678) 中。

[0240] 本发明的解旋酶优选包含SEQ ID NO:10的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基V343-L359) 和/或2A (RecA式) 域(残基R241-N327和A449-G496) 中。

[0241] 本发明的解旋酶优选包含SEQ ID NO:11的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基W276-L284) 和/或2A (RecA式) 域(残基R174-D260和A371-V421) 中。

[0242] 本发明的解旋酶优选包含SEQ ID NO:12的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基R308-Y313) 和/或2A (RecA式) 域(残基R206-K293和I408-L500) 中。

[0243] 本发明的解旋酶优选包含SEQ ID NO:13的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基M302-W306) 和/或2A (RecA式) 域(残基R195-D287和V394-Q450) 中。

[0244] 本发明的解旋酶优选包含SEQ ID NO:14的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基V265-I277) 和/或2A (RecA式) 域(残基R167-T249和L372-N421) 中。

[0245] 本发明的解旋酶优选包含SEQ ID NO:15的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基V270-F283) 和/或2A (RecA式) 域(残基R172-T254和L381-K434) 中。

[0246] 本发明的解旋酶优选包含SEQ ID NO:16的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基V257-F270)和/或2A (RecA式)域(残基R159-T241和L367-K420)中。

[0247] 本发明的解旋酶优选包含SEQ ID NO:17的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L278-Y294)和/或2A (RecA式)域(残基R182-T262和L393-V443)中。

[0248] 本发明的解旋酶优选包含SEQ ID NO:18的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L278-Y294)和/或2A (RecA式)域(残基R182-T262和L392-V442)中。

[0249] 本发明的解旋酶优选包含SEQ ID NO:19的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L273-F289)和/或2A (RecA式)域(残基R177-N257和L387-V438)中。

[0250] 本发明的解旋酶优选包含SEQ ID NO:20的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L281-F297)和/或2A (RecA式)域(残基R184-T265和L393-I442)中。

[0251] 本发明的解旋酶优选包含SEQ ID NO:21的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基H277-F293)和/或2A (RecA式)域(残基R180-T261和L393-V442)中。

[0252] 本发明的解旋酶优选包含SEQ ID NO:22的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L276-F292)和/或2A (RecA式)域(残基R179-T260和L390-I439)中。

[0253] 本发明的解旋酶优选包含SEQ ID NO:23的变体,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被进一步引入到钩域(残基L276-F292)和/或2A (RecA式)域(残基R179-T260和L391-V441)中。

[0254] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含(i) I181C; (ii) Y279C; (iii) I281C; 和(iv) E288C中的一个或多个。解旋酶可以包含(i)至(iv)的任何组合,例如(i); (ii); (iii); (iv); (i)和(ii); (i)和(iii); (i)和(iv); (ii)和(iii); (ii)和(iv); (iii)和(iv); 或(i), (ii), (iii)和(iv)。解旋酶更优选包含SEQ ID NO:8的变体,其包含(a) E94C, I281C和A360C或(b) E94C, I281C, G357C和A360C。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,其包含与(i)至(iv)、(a)和(b)中所定义的SEQ ID NO:8中的那些位点相对应的一个或多个位点上的半胱氨酸残基。解旋酶可以包含这些变体中的任一个,其中,Faz代替半胱氨酸被引入到一个或多个特定位点(或每个特定位点)上。

[0255] 本发明的解旋酶被进一步修饰以减少其表面负电荷。可以以与上述公开的Dda域相同的方式对表面残基进行鉴定。表面负电荷通常是表面带负电荷的氨基酸,例如,天冬氨酸(D)和谷氨酸(E)。

[0256] 解旋酶优选地被修饰,以通过用一个或多个带正电荷的氨基酸、不带电荷的氨基酸、非极性氨基酸和/或芳族氨基酸取代一个或多个带负电荷的氨基酸,或者通过引入一个或多个带正电荷的氨基酸,优选与一个或多个带负电荷的氨基酸相邻,来中和一个或多个表面负电荷。合适的带正电荷的氨基酸包括但不限于组氨酸(H)、赖氨酸(K)和精氨酸(R)。

不带电荷的氨基酸没有净电荷。合适的带电氨基酸包括但不限于半胱氨酸(C)、丝氨酸(S)、苏氨酸(T)、甲硫氨酸(M)、天冬酰胺(N)和谷氨酰胺(Q)。非极性氨基酸具有非极性侧链。合适的非极性氨基酸包括但不限于甘氨酸(G)、丙氨酸(A)、脯氨酸(P)、异亮氨酸(I)、亮氨酸(L)和缬氨酸(V)。芳香族氨基酸具有芳香族侧链。合适的芳香族氨基酸包括但不限于组氨酸(H)、苯丙氨酸(F)、色氨酸(W)和酪氨酸(Y)。

[0257] 优选的取代包括但不限于用R取代E、用K取代E、用N取代E、用K取代D以及用R取代D。

[0258] 本发明的解旋酶优选包含SEQ ID NO:8的变体,并且所述一个或多个带负电荷的氨基酸是D5,E8,E23,E47,D167,E172,D202,D212和E273中的一个或多个。可以中和任何数量的这些氨基酸,例如,它们中的1,2,3,4,5,6,7或8个。可以中和任何组合。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,并且所述一个或多个带负电荷的氨基酸对应SEQ ID NO:8中的D5,E8,E23,E47,D167,E172,D202,D212和E273的一个或多个。可以采用以下比对方式确定与SEQ ID NO:8中的D5,E8,E23,E47,D167,E172,D202,D212和E273相对应的SEQ ID NO:9至23中的氨基酸。本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含(a)E94C,E273G和A360C或(b)E94C,E273G,N292G和A360C。

[0259] 优选通过除去一个或多个天然半胱氨酸残基来对本发明的解旋酶进行进一步修饰。可以除去任何数量的天然半胱氨酸残基。SEQ ID NO:9至23中的每一个中的半胱氨酸残基数量在表1(如#C)中示出。优选通过取代来除去所述一个或多个半胱氨酸残基。优选用丙氨酸(A)、丝氨酸(S)或缬氨酸(V)取代所述一个或多个半胱氨酸残基。本发明的解旋酶优选包含SEQ ID NO:8的变体,并且所述一个或多个天然半胱氨酸残基是C109,C114,C136,C171和C412中的一个或多个。可以除去任何数量和组合的这些半胱氨酸残基。例如,SEQ ID NO:8的变体可以包含C109;C114;C136;C171;C412;C109和C114;C109和C136;C109和C171;C109和C412;C114和C136;C114和C171;C114和C412;C136和C171;C136和C412;C171和C412;C109,C114和C136;C109,C114和C171;C109,C114和C412;C109,C136和C171;C109,C136和C412;C109,C171和C412;C114,C136和C171;C114,C136和C412;C114,C171和C412;C136,C171和C412;C109,C114,C136和C171;C109,C114,C136和C412;C109,C114,C171和C412;C109,C136,C171和C412;C114,C136,C171和C412;或C109,C114,C136,C171和C412。

[0260] 本发明的解旋酶优选是这样一种解旋酶,其中,至少一个半胱氨酸残基(即,一个或多个半胱氨酸残基)和/或至少一个非天然氨基酸(即,一个或多个非天然氨基酸)仅被引入到塔域中。以上对合适的修饰进行了讨论。

[0261] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含以下突变:

[0262] -E93C和K364C;

[0263] -E94C和K364C;

[0264] -E94C和A360C;

[0265] -L97C和E361C;

[0266] -L97C和E361C和C412A;

[0267] -K123C和E361C;

[0268] -K123C,E361C和C412A;

[0269] -N155C和K358C;

- [0270] -N155C, K358C和C412A;
- [0271] -N155C和L354C;
- [0272] -N155C, L354C和C412A;
- [0273] - δ E93, E94C, δ N95和A360C;
- [0274] -E94C, δ N95和A360C;
- [0275] -E94C, Q100C, I127C和A360C;
- [0276] -L354C;
- [0277] -G357C;
- [0278] -E94C, G357C和A360C;
- [0279] -E94C, Y279C和A360C;
- [0280] -E94C, I281C和A360C;
- [0281] -E94C, Y279Faz和A360C;
- [0282] -Y279C和G357C;
- [0283] -I281C和G357C;
- [0284] -E94C, Y279C, G357C和A360C;
- [0285] -E94C, I281C, G357C和A360C;
- [0286] -E8R, E47K, E94C, D202K和A360C;
- [0287] -D5K, E23N, E94C, D167K, E172R, D212R和A360C;
- [0288] -D5K, E8R, E23N, E47K, E94C, D167K, E172R, D202K, D212R和A360C;
- [0289] -E94C, C114A, C171A, A360C和C412D;
- [0290] -E94C, C114A, C171A, A360C和C412S;
- [0291] -E94C, C109A, C136A和A360C;
- [0292] -E94C, C109A, C114A, C136A, C171A, A360C和C412S;
- [0293] -E94C, C109V, C114V, C171A, A360C和C412S;
- [0294] -C109A, C114A, C136A, G153C, C171A, E361C和C412A;
- [0295] -C109A, C114A, C136A, G153C, C171A, E361C和C412D;
- [0296] -C109A, C114A, C136A, G153C, C171A, E361C和C412S;
- [0297] -C109A, C114A, C136A, G153C, C171A, K358C和C412A;
- [0298] -C109A, C114A, C136A, G153C, C171A, K358C和C412D;
- [0299] -C109A, C114A, C136A, G153C, C171A, K358C和C412S;
- [0300] -C109A, C114A, C136A, N155C, C171A, K358C和C412A;
- [0301] -C109A, C114A, C136A, N155C, C171A, K358C和C412D;
- [0302] -C109A, C114A, C136A, N155C, C171A, K358C和C412S;
- [0303] -C109A, C114A, C136A, N155C, C171A, L354C和C412A;
- [0304] -C109A, C114A, C136A, N155C, C171A, L354C和C412D;
- [0305] -C109A, C114A, C136A, N155C, C171A, L354C和C412S;
- [0306] -C109A, C114A, K123C, C136A, C171A, E361C和C412A;
- [0307] -C109A, C114A, K123C, C136A, C171A, E361C和C412D;
- [0308] -C109A, C114A, K123C, C136A, C171A, E361C和C412S;

- [0309] -C109A, C114A, K123C, C136A, C171A, K358C和C412A;
- [0310] -C109A, C114A, K123C, C136A, C171A, K358C和C412D;
- [0311] -C109A, C114A, K123C, C136A, C171A, K358C和C412S;
- [0312] -C109A, C114A, C136A, G153C, C171A, E361C和C412A;
- [0313] -E94C, C109A, C114A, C136A, C171A, A360C和C412D;
- [0314] -E94C, C109A, C114V, C136A, C171A, A360C和C412D;
- [0315] -E94C, C109V, C114A, C136A, C171A, A360C和C412D;
- [0316] -L97C, C109A, C114A, C136A, C171A, E361C和C412A;
- [0317] -L97C, C109A, C114A, C136A, C171A, E361C和C412D;或者
- [0318] -L97C, C109A, C114A, C136A, C171A, E361C和C412S。

[0319] 钩域和/或2A域中的修饰

[0320] 在一个实施方案中,本发明的解旋酶是这样一种解旋酶,其中,至少一个半胱氨酸残基和/或至少一个非天然氨基酸被引入到钩域和/或2A (RecA式马达) 域中,其中,所述解旋酶具有控制多核苷酸的移动的能力。至少一个半胱氨酸残基和/或至少一个非天然氨基酸优选被引入到钩域和2A (RecA式马达) 域中。

[0321] 可以将任何数量的半胱氨酸残基和/或非天然氨基酸引入到每个域中。例如,1,2,3,4,5,6,7,8,9,10或更多个半胱氨酸残基可以被引入,和/或1,2,3,4,5,6,7,8,9,10或更多个非天然氨基酸可以被引入。可以仅引入一个或多个半胱氨酸残基。可以仅引入一个或多个非天然氨基酸。可以引入一个或多个半胱氨酸残基和一个或多个非天然氨基酸的组合。

[0322] 优选通过取代引入所述至少一个半胱氨酸残基和/或至少一个非天然氨基酸。这样做的方法是本领域已知的。以上对钩域和/或2A (RecA式马达) 域的合适修饰进行了讨论。

[0323] 本发明的解旋酶优选为SEQ ID NO:8的变体,其包含(a) Y279C, I181C, E288C, Y279C和I181C; (b) Y279C和E288C; (c) I181C和E288C或(d) Y279C, I181C和E288C。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,其包含与(a)至(d)中所定义的SEQ ID NO:8中的那些位点相对应的一个或多个位点上的突变。

[0324] 表面修饰

[0325] 在一个实施方案中,对解旋酶进行修饰以减少其表面负电荷,其中,解旋酶具有控制多核苷酸的移动的能力。以上对合适的修饰进行了讨论。可以中和任何数量的表面负电荷。本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含以下突变:

- [0326] -E273G;
- [0327] -E8R, E47K和D202K;
- [0328] -D5K, E23N, D167K, E172R和D212R;或者
- [0329] -D5K, E8R, E23N, E47K, D167K, E172R, D202K和D212R。

[0330] 其它修饰的解旋酶

[0331] 在一个实施方案中,本发明的解旋酶包含SEQ ID NO:8的变体,其包含:

- [0332] -A360K;
- [0333] -Y92L和/或A360Y;
- [0334] -Y92L, Y350N和Y363N;

- [0335] -Y92L和/或Y363N;或者
- [0336] -Y92L。
- [0337] 其它修饰
- [0338] 除了以上公开的特定突变之外,SEQ ID NO:8的变体可以包含以下突变中的一个或多个:
- [0339] -K38A;
- [0340] -T91F;
- [0341] -T91N;
- [0342] -T91Q;
- [0343] -T91W;
- [0344] -V96E;
- [0345] -V96F;
- [0346] -V96L
- [0347] -V96Q;
- [0348] -V96R;
- [0349] -V96W;
- [0350] -V96Y;
- [0351] -P274G;
- [0352] -V286F;
- [0353] -V286W;
- [0354] -V286Y;
- [0355] -F291G;
- [0356] -N292F;
- [0357] -N292G;
- [0358] -N292P;
- [0359] -N292Y;
- [0360] -G294Y;
- [0361] -G294F;
- [0362] -K364A;和
- [0363] -W378A。
- [0364] 除了以上公开的特定突变外,SEQ ID NO:8的变体可以包含:

[0365]

- K38A, E94C 和 A360C;
- H64K; E94C 和 A360C;
- H64N; E94C 和 A360C;
- H64Q; E94C 和 A360C;
- H64S; E94C 和 A360C;
- H64W, E94C 和 A360C;
- T80K, E94C 和 A360C;
- T80K, S83K, E94C, N242K, N293K 和 A360C;
- T80K, S83K, E94C, N242K, N293K, A360C 和 T394K;
- T80K, S83K, E94C, N293K 和 A360C;
- T80K, S83K, E94C, A360C 和 T394K;
- T80K, S83K, E94C, A360C 和 T394N;
- T80K, E94C, N293K 和 A360C;
- T80N, E94C 和 A360C;
- H82A, E94C 和 A360C;
- H82A, P89A, E94C, F98A 和 A360C;
- H82F, E94C 和 A360C;
- H82Q, E94C, A360C;
- H82R, E94C 和 A360C;
- H82W, E94C 和 A360C;
- H82W, P89W, E94C, F98W 和 A360C;
- H82Y, E94C 和 A360C;
- S83K, E94C 和 A360C;

[0366]

- T80K, E94C, N242K 和 A360C;
- T80K, E94C, N242K, N293K 和 A360C;
- N88H, E94C 和 A360C;
- N88Q, E94C 和 A360C;
- P89A, E94C 和 A360C;
- P89A, F98W, E94C 和 A360C;
- P89A, E94C, F98Y 和 A360C;
- P89A, E94C, F98A 和 A360C;
- P89F, E94C 和 A360C;
- P89S, E94C 和 A360C;
- P89T, E94C 和 A360C;
- P89W, E94C, F98W 和 A360C;
- P89Y, E94C 和 A360C;
- T91F, E94C 和 A360C;
- T91N, E94C 和 A360C;
- T91Q, E94C 和 A360C;
- T91W, E94C 和 A360C;
- E94C, V96E 和 A360C;
- E94C, V96F 和 A360C;
- E94C, V96L 和 A360C;
- E94C, V96Q 和 A360C;
- E94C, V96R 和 A360C;
- E94C, V96W 和 A360C;
- E94C, V96Y 和 A360C;
- E94C, F98A 和 A360C;
- E94C, F98L 和 A360C;
- E94C, F98V 和 A360C;
- E94C, F98Y 和 A360C;
- E94C; F98W 和 A360C;
- E94C, V150A 和 A360C;
- E94C, V150F 和 A360C;
- E94C, VI 501 和 A360C;
- E94C, V150K 和 A360C;
- E94C, V150L 和 A360C;
- E94C, V150S 和 A360C;
- E94C, V150T 和 A360C;
- E94C, A360C 和 T394K;
- S83K, T80K, E94C, A360C 和 T394K;
- S83N, E94C 和 A360C;
- S83T, E94C 和 A360C;
- E94C, V150W 和 A360C;
- E94C, V150Y 和 A360C;
- E94C, F240Y 和 A360C;
- E94C, F240W 和 A360C;
- E94C, N242K 和 A360C;
- E94C, N242K, N293K 和 A360C;
- E94C, P274G 和 A360C;
- E94C, L275G 和 A360C;
- E94C, F276A 和 A360C;
- E94C, F276I 和 A360C;
- E94C, F276M 和 A360C;
- E94C, F276V 和 A360C;
- E94C, F276W 和 A360C;
- E94C, F276Y 和 A360C;
- E94C, V286F 和 A360C;
- E94C, V286W 和 A360C;
- E94C, V286Y 和 A360C;
- E94C, S287F 和 A360C;
- E94C, S287W 和 A360C;
- E94C, S287Y 和 A360C;
- E94C, F291G 和 A360C;
- E94C, N292F 和 A360C;
- E94C, N292G 和 A360C;
- E94C, N292P 和 A360C;
- E94C, N292Y 和 A360C;
- E94C, N293F 和 A360C;
- E94C, N293K 和 A360C;
- E94C, N293Q 和 A360C;
- E94C, N293Y 和 A360C;
- E94C, G294F 和 A360C;
- E94C, G294Y 和 A360C;
- E94C, A36C 和 K364A;
- E94C, A360C, W378A;
- E94C, A360C 和 Y415F;

[0367]

- E94C, A360C 和 H396Q;
- E94C, A360C 和 H396S;
- E94C, A360C 和 H396W;
- E94C, A360C 和 Y415K;
- E94C, A360C 和 Y415M; 或者
- E94C, A360C 和 Y415W。

[0368] 本发明的解旋酶优选包含SEQ ID NO:8的变体,其包含(a) E94C/A360C/W378A或者(b) E94C/A360C/C109A/C136A/W378A或者(d) E94C/A360C/C109A/C136A/W378A和然后(Δ M1)G1G2(即,缺失M1,然后加入G1和G2)。

[0369] SEQ ID NO:8至23中任一个的优选变体具有(除了本发明的修饰之外),一个甘氨酸残基(G)取代的N-末端甲硫氨酸(M)。在实施例中,这显示为(Δ M1)G1。也可以将其称为M1G。上述任一变体还可以包括M1G。

[0370] 本发明最优选的解旋酶包含SEQ ID NO:8的变体,其包含:

[0371] (a) E94C/F98W/A360C/C109A/C136A/K194L;

[0372] (b) M1G/E94C/F98W/A360C/C109A/C136A/K194L;

[0373] (c) E94C/F98W/A360C/C109A/C136A/K199L;或

[0374] (d) M1G/E94C/F98W/A360C/C109A/C136A/K199L。

[0375] 变体

[0376] 解旋酶的变体是一种酶,其具有从野生型解旋酶的氨基酸序列变化而来的氨基酸序列,并具有多核苷酸结合活性。具体地,SEQ ID NO:8至23中任一个的变体是一种酶,其具有由SEQ ID NO:8至23中任一个的氨基酸序列变化而来并具有多核苷酸结合活性的氨基酸序列。可以采用本领域已知的方法来确定多核苷酸结合活性。合适的方法包括但不限于荧光各向异性、色氨酸荧光和电泳迁移率变动分析(EMSA)。例如,可以如实施例所述确定变体与单链多核苷酸结合的能力。

[0377] 变体具有解旋酶活性。这可以以各种方式进行测量。例如,可以采用电生理学、荧光测定或ATP水解来测量变体沿着多核苷酸移位的能力。

[0378] 变体可以包括促进多核苷酸对解旋酶进行编码的修饰和/或促进其在高盐浓度和/或室温下活性的修饰。

[0379] 对于整个长度的SEQ ID NO:8至23中任一个的氨基酸序列,基于氨基酸相似性或同一性,变体将优选与该序列至少20%同源。更优选地,基于氨基酸同一性,变体多肽与整个序列的SEQ ID NO:8至23中任一个的氨基酸序列可以至少30%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%且更优选至少95%、97%或99%同源。在100或更多,例如150,200,300,400或500或更多的连续氨基酸长度上,可以具有至少70%,例如至少80%、至少85%、至少90%或至少95%的氨基酸同一性(“严格同源性”)。如下所述,对同源性进行确定。相对于SEQ ID NO:2和4,变体可以以下述任一方式不同于野生型序列。具体地,除了上述特定修饰之外,SEQ ID NO:8至23中任一个的变体可以包括下述一个或多个取代、一个或多个缺失和/或一个或多个添加。

[0380] SEQ ID NO:8至23中任一个的优选变体在氨基-(N-)末端和/或羧基(C-)末端上具有非天然氨基酸,例如,Faz。SEQ ID NO:8至23中任一个的优选变体在氨基-(N-)末端和/或羧基(C-)末端具有半胱氨酸残基。SEQ ID NO:8至23中任一个的优选变体在氨基-(N-)末端

具有半胱氨酸残基,在羧基(C-)末端具有非天然氨基酸,例如,Faz,反之亦然。

[0381] SEQ ID NO:8的优选变体包含以下修饰E54G,D151E,I196N和G357A的一个或多个,例如,全部。

[0382] 无连接

[0383] 在一个优选实施方案中,本发明的修饰的解旋酶中引入的半胱氨酸和/或非天然氨基酸彼此之间不连接。

[0384] 连接引入的两个或多个半胱氨酸和/或非天然氨基酸

[0385] 在另一个优选实施方案中,本发明的修饰的解旋酶中引入的两个或多个半胱氨酸和/或非天然氨基酸彼此连接。这通常降低本发明的解旋酶从多核苷酸中解链的能力。

[0386] 任何数量和组合的两个或更多个所述引入的半胱氨酸和/或非天然氨基酸可以彼此连接。例如,3,4,5,6,7,8或多个半胱氨酸和/或非天然氨基酸可以彼此连接。一个或多个半胱氨酸可以与一个或多个半胱氨酸连接。一个或多个半胱氨酸可以与一个或多个非天然氨基酸(例如,Faz)连接。一个或多个非天然氨基酸(例如,Faz)可以与一个或多个非天然氨基酸(例如,Faz)连接。

[0387] 两个或多个半胱氨酸和/或非天然氨基酸可以以任何方式连接。连接可以是瞬时的,例如,非共价连接。甚至瞬时连接将减少多核苷酸从解旋酶中解链。

[0388] 两个或多个半胱氨酸和/或非天然氨基酸优选通过亲和分子连接。合适的亲和分子是本领域已知的。亲和分子优选(a)互补多核苷酸(国际申请号PCT/GB10/000132(公开为W02010/086602);(b)抗体或其片段以及互补表位(Biochemistry 6thEd,W.H.Freeman and co(2007) pp953-954);(c)肽拉链(O'Shea et al.,Science 254(5031):539-544);(d)能够通过 β -片体添加(β -sheets augmentation)的相互作用(Remaut and Waksman Trends Biochem.Sci.(2006) 31 436-444);(e)能够氢键、 π 堆叠(π -stacking)或形成盐桥;(f)轮烷(Xiang Ma and He Tian Chem.Soc.Rev.,2010,39,70-80);(g)适体和互补蛋白(James, W.in Encyclopedia of Analytical Chemistry,R.A.Meyers(Ed.) pp.4848-4871 John Wiley&Sons Ltd,Chichester,2000)或(h)半螯合剂(Hammerstein et al.J Biol Chem.2011April 22;286(16):14324-14334)。对于(e),氢键发生在与电负性原子结合的质子和另一个电负性原子之间。 π 堆叠需要两个可以堆叠在一起的芳族环,其中,两个芳族环的平面是平行的。盐桥位于可以使其电子在几个原子上离域(delocalize)的基团之间,例如,天冬氨酸与精氨酸之间。

[0389] 两个或多个部分可以通过六-组氨酸标签(hexa-his tag)或Ni-NTA瞬时连接。

[0390] 两个或多个半胱氨酸和/或非天然氨基酸优选永久连接。在本发明的上下文中,在使用解旋酶时如果连接未断裂或者在使用者的一部分上不进行干预的情况下连接不会断裂,则连接是永久性的,干预的情况例如,使用还原反应来打开-S-S-键。

[0391] 两个或多个半胱氨酸和/或非天然氨基酸优选共价连接。可以采用本领域已知的任何方法使两个或更多个半胱氨酸和/或非天然氨基酸共价连接。

[0392] 两个或多个半胱氨酸和/或非天然氨基酸可以通过其天然存在的氨基酸例如,半胱氨酸、苏氨酸、丝氨酸、天冬氨酸、天冬酰胺、谷氨酸和谷氨酰胺共价连接。可以对天然存在的氨基酸进行修饰以便于连接。例如,可以通过酰化、磷酸化、糖基化或法尼基化对天然存在的氨基酸进行修饰。其它合适的修饰是本领域已知的。对天然存在的氨基酸进行的修

饰可以是转译后修饰。两个或更多个半胱氨酸和/或非天然氨基酸可以通过被引入到其序列的氨基酸进行连接。这些氨基酸优选通过取代被引入。引入的氨基酸可以是促进连接的半胱氨酸或非天然氨基酸。合适的非天然氨基酸包括但不限于4-叠氮基-L-苯丙氨酸(Faz)、Liu C.C. and Schultz P.G., *Annu. Rev. Biochem.*, 2010, 79, 413-444中的图1中包含的任一个编号为1-71的氨基酸或下列氨基酸中的任一个。引入的氨基酸可以如上所述进行修饰。

[0393] 在优选实施方案中,使用连接器来连接两个或更多个半胱氨酸和/或非天然氨基酸。以下对连接器分子进行更详细的讨论。一种合适的连接方法是半胱氨酸连接。这将在以下进行更详细的讨论。两个或更多个半胱氨酸和/或非天然氨基酸优选采用一个或多个例如,两个或三个连接器连接。所述一个或多个连接器可以设计成减小如上所述的开口尺寸或闭合开口。如果正在使用一个或多个连接器来闭合如上所述的开口,则所述一个或多个连接器的至少一部分优选是取向的,从而当其通过解旋酶结合时不与多核苷酸平行。更优选地,所有连接器均以这种方式取向。如果正在使用一个或多个连接器来闭合如上所述的开口,则所述一个或多个连接器的至少一部分优选地在方向,即当其通过解旋酶结合时以不与多核苷酸平行的方向穿过开口。更优选地,所有连接器均以这种方式穿过开口。在这些实施方案中,所述一个或多个连接器的至少一部分可能垂直于多核苷酸。这种取向有效地闭合了开口,使得多核苷酸不能通过开口从解旋酶中解开。

[0394] 每个连接器可以具有两个或多个功能性末端,例如,两个、三个或四个功能性末端。连接器中的末端的合适结构在本领域中是公知的。

[0395] 一个或多个连接器的一个或多个末端优选与解旋酶共价连接。如果一端为共价连接,则一个或多个连接器可以瞬时连接上述两个或多个半胱氨酸和/或非天然氨基酸。如果两个或所有末端为共价连接,则一个或多个连接器永久连接两个或多个半胱氨酸和/或非天然氨基酸。

[0396] 一个或多个连接器优选为氨基酸序列和/或化学交联剂。

[0397] 合适的氨基酸连接器例如,肽连接器是本领域已知的。通常设计氨基酸或肽连接器的长度、柔性和亲水性,使得其减小开口的尺寸,但不会扰乱解旋酶的功能。优选的柔性肽连接器是2至20个,例如4,6,8,10或16个,丝氨酸和/或甘氨酸氨基酸的链。更优选的柔性连接器包括(SG)₁, (SG)₂, (SG)₃, (SG)₄, (SG)₅, (SG)₈, (SG)₁₀, (SG)₁₅或(SG)₂₀,其中,S是丝氨酸,G是甘氨酸。优选的刚性连接器是2至30个,例如,4,6,8,16或24个,脯氨酸氨基酸的链。更优选的刚性连接器包括(P)₁₂,其中,P是脯氨酸。连接器的氨基酸序列优选包含多核苷酸结合部分。以下对这种部分和与其使用相关的优点进行讨论。

[0398] 合适的化学交联剂是本领域公知的。合适的化学交联剂包括但不限于包括以下官能团的那些化学交联剂:马来酰亚胺、活性酯、琥珀酰亚胺、叠氮化物、炔烃(例如,二苯并环辛炔(DIBO或DBCO)、二氟环炔烃和线性炔烃)、膦(例如,无痕和非无痕施陶丁格结合中使用的那些)、卤代乙酰基(例如,碘乙酰胺)、光气型试剂、磺酰氯化物试剂、异硫氰酸酯、酰基卤化物、肼、二硫化物、乙烯基砜、氮杂环丙烷和光敏试剂(例如,芳基叠氮化物、二氮杂环丙烷)。

[0399] 氨基酸和官能团之间的反应可以是自发的,例如,半胱氨酸/马来酰亚胺,或者可能需要外部试剂,例如,用于连接叠氮化物和线性炔烃的Cu(I)。

[0400] 连接器可以包括跨越所需距离的任何分子。连接器在长度上可以从一个碳(光气型连接器)到许多埃(Angstrom)进行变化。线性分子的示例包括但不限于聚乙二醇(PEG)、多肽、多糖、脱氧核糖核酸(DNA)、肽核酸(PNA)、苏糖核酸(TNA)、甘油核酸(GNA)、饱和和不饱和的烃、聚酰胺。这些连接器可以是惰性的或反应性的,特别是其在限定位点上可以是化学断裂的(cleavable),或者可以用荧光团或配体进行自身修饰。连接器优选是耐二硫苏糖醇(DTT)的。

[0401] 优选的交联剂包括2,5-二氧代吡咯烷-1-基3-(吡啶-2-基二硫烷基)丙酸酯、2,5-二氧代吡咯烷-1-基4-(吡啶-2-基二硫烷基)丁酸乙酯和2,5-二氧代吡咯烷-1-基8-(吡啶-2-基二硫烷基)辛酸酯、二马来酰亚胺PEG 1k、二马来酰亚胺PEG 3.4k、二马来酰亚胺PEG5k、二马来酰亚胺PEG 10k、双(马来酰亚胺基)乙烷(BMOE)、双马来酰亚胺己烷(BMH)、1,4-双马来酰亚胺丁烷(BMB)、1,4-双马来酰亚胺基-2,3-二羟基丁烷(BMDB)、BM[PEO]2(1,8-双马来酰亚胺二乙二醇)、BM[PEO]3(1,11-双马来酰亚胺三乙二醇)、三[2-马来酰亚胺氨基乙基]胺(TMEA)、DTME二硫代二马来酰亚胺乙烷、双马来酰亚胺PEG3、双马来酰亚胺PEG11、DBCO-马来酰亚胺、DBCO-PEG4-马来酰亚胺、DBCO-PEG4-NH₂、DBCO-PEG4-NHS、DBCO-NHS、DBCO-PEG-DBCO 2.8kDa、DBCO-PEG-DBCO 4.0kDa、DBCO-15原子-DBCO、DBCO-26原子-DBCO、DBCO-35原子-DBCO、DBCO-PEG4-S-S-PEG3-生物素、DBCO-S-S-PEG3-生物素、DBCO-S-S-PEG11-生物素、(琥珀酰亚胺基3-(2-吡啶基二硫代)丙酸酯(SPDP)和马来酰亚胺-PEG(2kDa)-马来酰亚胺(α , ω -双马来酰亚胺基聚乙二醇)。最优选的交联剂是马来酰亚胺-丙基-SRDFWRS-(1,2-乙二胺)-丙基-马来酰亚胺。

[0402] 一个或多个连接器可以是可断裂的。这将在以下进行更详细的讨论。

[0403] 可以使用彼此特异的两个不同连接器来连接两个或更多个半胱氨酸和/或非天然氨基酸。其中一个连接器与一个部分连接,另一个与另一个部分连接。连接器应当反应以形成本发明的修饰的解旋酶。可以使用申请号为PCT/GB10/000132(公布为W02010/086602)的国际申请中描述的杂交连接器来连接两个或更多个半胱氨酸和/或非天然氨基酸。具体地,可以使用两个或更多个连接器来连接两个或更多个半胱氨酸和/或非天然氨基酸,其中,所述两个或更多个连接器各自包含可杂交的区域和能够形成共价键的基团。连接器中的可杂交区域将两个或更多个半胱氨酸和/或非天然氨基酸杂交并连接。连接的半胱氨酸和/或非天然氨基酸然后通过基团之间形成的共价键耦合。根据本发明,可以使用申请号为PCT/GB10/000132(公开为W02010/086602)的国际申请中公开的任何特异性连接器。

[0404] 可以对两个或更多个半胱氨酸和/或非天然氨基酸进行修饰,然后使用对这两个修饰特异的化学交联剂来连接。可以使用上述任一交联剂。

[0405] 可以对连接器进行标记。合适的标记物包括但不限于荧光分子(例如Cy3或AlexaFluor®555)、放射性同位素、例如,¹²⁵I, ³⁵S,酶、抗体、抗原、多核苷酸和诸如生物素等配体。这种标记物允许对连接器的数量进行量化。标记物还可以是可断裂的纯化标签,例如,生物素,或在鉴定方法中出现的特异性序列,例如,本身不存在于蛋白质中但通过胰蛋白酶消化而释放的肽。

[0406] 连接两个或更多个半胱氨酸的优选方法是通过半胱氨酸连接。这可以通过双功能化学交联剂或通过末端出现半胱氨酸残基的氨基酸连接器来介导。

[0407] 可以对任何双官能连接器的长度、反应性、特异性、刚性和溶解性进行设计以确保

开口尺寸充分减小并且解旋酶的功能被保留。合适的连接器包括双马来酰亚胺交联剂,例如,1,4-双(马来酰亚胺基)丁烷(BMB)或双(马来酰亚胺基)己烷。双功能连接器的一个缺点是如果优选在特定位点进行连接,则要求解旋酶不进一步包含表面可接近(accessible)的半胱氨酸残基,因为双功能连接器与表面可接近的半胱氨酸残基的结合可能难以控制,并且可能影响底物结合或活性。如果解旋酶确实包含几个可接近的半胱氨酸残基,则可能需要对解旋酶进行修饰以去除它们,同时确保修饰不影响解旋酶的折叠或活性。这在申请号为PCT/GB10/000133(公开为W02010/086603)的国际申请中进行了讨论。可以通过对例如,肽连接器上的相邻残基进行修饰来增强半胱氨酸残基的反应性。例如,侧链精氨酸、组氨酸或赖氨酸残基的碱基会将半胱氨酸硫醇基团的pKa改变为更具反应性的S⁻基团的pKa。半胱氨酸残基的反应性可以被硫醇保护基团,例如5,5'-二硫代双-(2-对硝基苯甲酸)(dTNB)保护。这些保护基团可以在连接器连接之前与解旋酶的一个或多个半胱氨酸残基反应。可以采用固定在小珠(beads)上的还原试剂(例如,固定化的三(2-羧乙基)膦,TCEP)来使表面可接近的半胱氨酸选择性去保护。以下对半胱氨酸连接进行更详细的讨论。

[0408] 另一种优选的连接方法是通过Faz连接。这可以通过双功能化学连接器或通过末端出现Faz残基的多核苷酸连接器来介导。

[0409] 其它修饰

[0410] 也可以对本发明的解旋酶进行修饰以增加(i)塔域与(ii)销域和/或1A域之间的吸引力。根据本发明,可以进行任何已知的化学修饰。这些类型的修饰在申请号为PCT/GB2014/052736(W02015/055981)的国际申请中进行了公开。

[0411] 具体地,本发明提供本发明的解旋酶,其中,至少一个带电荷的氨基酸被引入到(i)塔域和/或(ii)销域和/或(iii)1A(RecA式马达)域中,其中,解旋酶具有控制多核苷酸移动的能力。如上所述,可以对解旋酶控制多核苷酸移动的能力进行测量。本发明优选提供本发明的解旋酶,其中,至少一个带电荷的氨基酸被引入到(i)塔域和(ii)销域和/或1A域中。

[0412] 所述至少一个带电荷的氨基酸可以带负电荷的或带正电荷的。所述至少一个带电荷的氨基酸优选地与其在解旋酶中相互作用的任一氨基酸所带的电荷相反。例如,至少一个带正电荷的氨基酸可以被引入到塔域中的与销域中的带负电荷的氨基酸发生相互作用的位点上。所述至少一个带电荷的氨基酸通常被引入到位点,即在野生型(即,未修饰的)解旋酶中的不带电荷的位点上。可以使用所述至少一个带电荷的氨基酸来取代解旋酶中至少一个带相反电荷的氨基酸。例如,带正电荷的氨基酸可以被用来取代带负电荷的氨基酸。

[0413] 合适的带电荷的氨基酸如上讨论。所述至少一个带电荷的氨基酸可以是天然的,例如,精氨酸(R)、组氨酸(H)、赖氨酸(K)、天冬氨酸(D)或谷氨酸(E)。可选地,所述至少一个带电荷的氨基酸可以是人造的或非天然的。可以将任何数量的带电荷的氨基酸引入每个域中。例如,可以将1,2,3,4,5,6,7,8,9,10或更多个带电荷的氨基酸引入每个域中。

[0414] 解旋酶优选包含SEQ ID NO:8的变体,其包含在以下一个或多个位点上带正电荷的氨基酸:(i) 93;(ii) 354;(iii) 360;(iv) 361;(v) 94;(vi) 97;(vii) 155;(viii) 357;(ix) 100和(x) 127。解旋酶优选包含SEQ ID NO:8的变体,其包含在以下一个或多个位点上带负电荷的氨基酸:(i) 354;(ii) 358;(iii) 360;(iv) 364;(v) 97;(vi) 123;(vii) 155;(viii) 357;(ix) 100和(x) 127。解旋酶优选包含SEQ ID NO:9至23中任一个的变体,所述变体在与

(i)至(x)中任一项所定义的SEQ ID NO:8中的那些位点相对应的位点上包含带正电荷的氨基酸或带负电荷的氨基酸。可以采用以下SEQ ID NO:8至23的比对来鉴定与SEQ ID NO:8中的那些位点相对应的SEQ ID NO:9至23中任一个的位点。

[0415] 解旋酶优选包含SEQ ID NO:8的变体,其通过至少一个带电荷的氨基酸的引入来修饰,使得其在以下位点上包含带相反电荷的氨基酸:(i) 93和354;(ii) 93和358;(iii) 93和360;(iv) 93和361;(v) 93和364;(vi) 94和354;(vii) 94和358;(viii) 94和360;(ix) 94和361;(x) 94和364;(xi) 97和354;(xii) 97和358;(xiii) 97和360;(xiv) 97和361;(xv) 97和364;(xvi) 123和354;(xvii) 123和358;(xviii) 123和360;(xix) 123和361;(xx) 123和364;(xxi) 155和354;(xxii) 155和358;(xxiii) 155和360;(xxiv) 155和361;(xxv) 155和364。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,其在与(i)至(xxv)中任一项所定义的SEQ ID NO:8中的那些位点相对应的位点上包含带相反电荷的氨基酸。

[0416] 本发明还提供一种解旋酶,其中,(i)至少一个带电荷的氨基酸被引入到塔域中;以及(ii)至少一个带相反电荷的氨基酸被引入到销域和/或1A(RecA式马达)域中,其中,所述解旋酶具有控制多核苷酸移动的能力。所述至少一个带电荷的氨基酸可以带负电荷,所述至少一个带相反电荷的氨基酸可以带正电荷,反之亦然。以上对合适的带电荷的氨基酸进行了讨论。可以引入任何数量的带电荷的氨基酸和任何数量的带相反电荷的氨基酸。例如,可以引入1,2,3,4,5,6,7,8,9,10或更多个带电荷的氨基酸和/或可以引入1,2,3,4,5,6,7,8,9,10或更多个带相反电荷的氨基酸。

[0417] 通常在野生型解旋酶中不带电荷的位点上引入带电荷的氨基酸。可采用带电荷的氨基酸中的一种或两种来取代解旋酶中带电荷的氨基酸。例如,带正电荷的氨基酸可以被用来取代带负电荷的氨基酸。可以在上述(i)塔域和(ii)销域和/或1A域中的任一位点上引入带电荷的氨基酸。通常引入带相反电荷的氨基酸使得其在所得的解旋酶中相互作用。解旋酶优选包含SEQ ID NO:8的变体,其中,在以下位点上引入带相反电荷的氨基酸:(i) 97和354;(ii) 97和360;(iii) 155和354;或(iv) 155和360。本发明的解旋酶优选包含SEQ ID NO:9至23中任一个的变体,其在与(i)至(iv)中任一项所定义的SEQ ID NO:8中的那些位点相对应的位点上包含带相反电荷的氨基酸。

[0418] 构建体

[0419] 本发明还提供一种构建体,其包含本发明的修饰的解旋酶和附加的多核苷酸结合部分,其中,所述解旋酶与所述多核苷酸结合部分连接,所述构建体具有控制多核苷酸移动的能力。所述构建体是人造的或非天然的。

[0420] 本发明的构建体是在链测序期间控制多核苷酸移动的有用工具。本发明的构建体相比本发明的修饰的解旋酶更不可能从正进行测序的多核苷酸中脱离开来。当构建体控制多核苷酸穿过纳米孔的移位时,所述构建体可甚至提供更大的多核苷酸的读取长度。

[0421] 还可以设计与特异性多核苷酸序列结合的目标构建体。如下进行更详细的讨论,多核苷酸结合部分可以与特异性多核苷酸序列结合,从而将构建体的解旋酶部分靶向特异性序列。

[0422] 构建体具有控制多核苷酸移动的能力。这可以如上所述进行确定。

[0423] 本发明的构建体可以是分离的、基本上分离的、纯化的或基本上纯化的。如果构建体完全没有任何其它组分,如脂质、多核苷酸或孔单体,则其是分离的或纯化的。如果将构

建体与不会干扰其预期用途的载体或稀释剂混合,则其是基本上分离的。例如,如果构建体以下形式存在,即,包含小于10%、小于5%、小于2%或小于1%的其它组分(例如脂质、多核苷酸或孔单体),则其是基本上分离的或基本上纯化的。

[0424] 解旋酶可以是上述本发明解旋酶中的任一种。

[0425] 解旋酶优选与附加的多核苷酸结合部分共价连接。解旋酶可以在不止一个位点,例如两个或三个上与所述部分连接。

[0426] 可以采用本领域已知的任何方法将解旋酶与所述部分共价连接。关于连接两个或多个部分的合适方法如上讨论。

[0427] 解旋酶和部分可以单独制备,然后连接在一起。两个组分可以以任何构型进行连接。例如,它们可以通过其末端(即,氨基或羧基末端)氨基酸进行连接。合适的构型包括但不限于所述部分的氨基末端与解旋酶的羧基末端连接,反之亦然。可选地,两个组分可以通过其序列内的氨基酸进行连接。例如,所述部分可以与解旋酶的环区域中的一个或多个氨基酸连接。在优选实施方案中,所述部分的末端氨基酸与解旋酶的环区域中的一个或多个氨基酸连接。

[0428] 在优选实施方案中,解旋酶与所述部分进行化学连接,例如,通过上述一个或多个连接器分子。在另一优选实施方案中,解旋酶与所述部分进行遗传融合。如果整个构建体从单个多核苷酸序列表达,则解旋酶与部分进行遗传融合。可以以任何方式组合解旋酶和部分的编码序列以形成编码构建体的单个多核苷酸序列。孔与核酸结合蛋白的遗传融合在申请号为PCT/GB09/001679(公开为W02010/004265)的国际申请中进行了讨论。

[0429] 解旋酶和部分可以以任何构型进行遗传融合。解旋酶和部分可以通过其末端氨基酸进行融合。例如,所述部分的氨基末端可以与解旋酶的羧基末端进行融合,反之亦然。所述部分的氨基酸序列优选以框架(frame)的形式添加到解旋酶的氨基酸序列中。换言之,所述部分优选插入解旋酶的序列中。在这些实施方案中,解旋酶和部分通常在两个位点上,即,通过所述部分的氨基和羧基末端氨基酸连接。如果所述部分插入解旋酶的序列中,则所述部分的氨基和羧基末端氨基酸优选非常接近,并且各自与解旋酶或其变体的序列中的相邻氨基酸连接。在优选实施方案中,所述部分插入到解旋酶的环区域中。

[0430] 解旋酶可以与所述部分直接连接。解旋酶优选采用上述一个或多个,例如两个或三个连接器与所述部分连接。所述一个或多个连接器可以设计成限制所述部分的移动性。可以对解旋酶和/或部分进行修饰以促进上述一个或多个连接器连接。

[0431] 可断裂的连接器可用作辅助,以将构建体与非连接的组分分离,并可用于进一步控制合成反应。例如,异双官能连接器可以与解旋酶反应,但不与所述部分反应。如果连接器的自由端可用于将解旋酶蛋白质与表面结合,则可以从混合物中除去第一反应中未反应的解旋酶。随后,连接器可以进行断裂以暴露出与所述部分反应的基团。此外,通过遵循此连接反应的序列,可以首先优化解旋酶的反应条件,然后在连接器断裂后优化所述部分的反应条件。因为连接器被限制在与其连接的区域内,所以第二反应也将更多地针对与所述部分反应的正确位点。

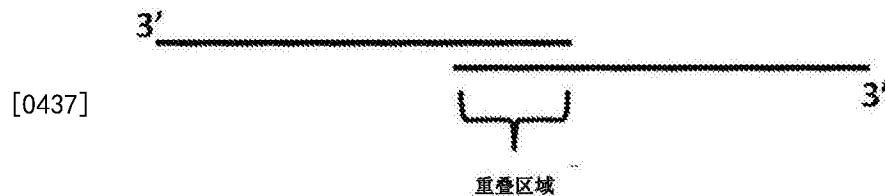
[0432] 在解旋酶/交联剂复合物与所述部分共价连接之前,解旋酶可以与双功能交联剂共价连接。可选地,在双功能交联剂/部分复合物与解旋酶连接之前,所述部分可以与双功能交联剂共价连接。解旋酶和部分可以同时与化学交联剂共价连接。

[0433] 将解旋酶与所述部分连接的优选方法是如上所述的半胱氨酸连接和Faz连接。在优选实施方案中,反应性半胱氨酸存在于肽连接器上,所述肽连接器遗传地连接到所述部分。这意味着不一定需要额外的修饰来从所述部分中去除其它可接近的半胱氨酸残基。

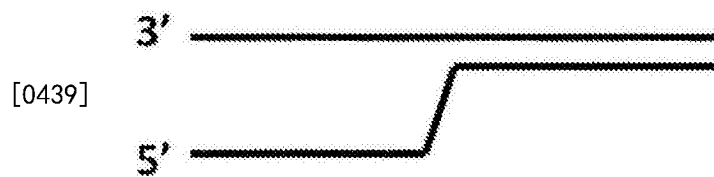
[0434] 可以通过将连接器的浓度保持在大量过剩的解旋酶和/或部分中来防止解旋酶或部分自身发生交联。可选地,可以使用“钥匙和锁”结构,其中使用了两个连接器。每个连接器的仅一端可以一起反应以形成更长的连接器,而连接器的另一端各自与构建体的不同部分(即,解旋酶或部分)反应。这将在以下进行更详细的讨论。

[0435] 对连接位点进行挑选使得当构建体与多核苷酸接触时,解旋酶和部分均可以与多核苷酸结合并控制其移动。

[0436] 可以使用解旋酶和部分的多核苷酸结合活性来促进连接。例如,可以使用互补多核苷酸来使解旋酶和部分在它们杂交时结合在一起。解旋酶可以与一个多核苷酸结合,部分可以与互补多核苷酸结合。然后可以使两个多核苷酸彼此杂交。这将使解旋酶与部分紧密接触,使得连接反应更有效率。这特别有助于在正确取向上连接两个或更多个解旋酶以便控制目标多核苷酸的移动。以下示出了可能使用的互补多核苷酸的示例。



[0438] 对于解旋酶-Phi29构建体,可以采用以下DNA。



[0440] 可以将标签添加到构建体中,使得构建体的纯化变得更加容易。然后,如果需要去除这些标签,则可以对这些标签进行化学或酶断裂。也可以包括荧光团或发色团,这些也可以是可断裂的。

[0441] 构建体的简单纯化方法是在每个蛋白质(即,解旋酶和部分)上包括不同的纯化标签,例如,六-组氨酸标签和Strep-标签®。如果两种蛋白质彼此不同,则这个方法特别有用。采用两种标签使得仅具有两种标签的物种能够很容易被纯化。

[0442] 如果两种蛋白质不具有两种不同的标签,则可以使用其它方法。例如,可以除去具有游离表面半胱氨酸的蛋白质或连接有连接器的蛋白质,这些蛋白质未反应形成构建体,例如,对马来酰亚胺连接器使用碘乙酰胺树脂。

[0443] 基于不同的DNA持续合成性(processivity)性质,本发明的构建体也可以从未反应的蛋白质中进行纯化。具体地,当本发明的构建体控制多核苷酸穿过纳米孔的移位时,基于对多核苷酸的亲和性增加、一旦结合从多核苷酸脱离的可能性降低和/或多核苷酸的读取长度增加,所述构建体从未反应的蛋白质中进行纯化。

[0444] 还可以设计与特异性多核苷酸序列结合的目标构建体。如下进行更详细的讨论,多核苷酸结合部分可以与特异性多核苷酸序列结合,从而将构建体的解旋酶部分靶向特异性序列。

[0445] 多核苷酸结合部分 (Polynucleotide binding moiety)

[0446] 本发明的构建体包含多核苷酸结合部分。多核苷酸结合部分是能够与多核苷酸结合的多肽。所述部分优选能够与限定的多核苷酸序列进行特异性结合。换言之,所述部分优选与特异性多核苷酸序列结合,但是显示比与不同序列的结合强度小至少10倍,或更优选地显示比与不同序列的结合强度小至少100倍,或最优选地显示比与不同序列的结合强度小至少1000倍。所述不同序列可以是随机序列。在一些实施方案中,所述部分与特异性多核苷酸序列结合,但不能测量与不同序列的结合。可以使用与特异性序列结合的部分来设计被定向到这种序列的构建体。

[0447] 所述部分通常与多核苷酸相互作用,并修饰其至少一个性质。所述部分可以通过对多核苷酸进行断裂以形成单个核苷酸或核苷酸的较短链,例如二核苷酸或三核苷酸,来对其进行修饰。所述部分可以通过将多核苷酸取向或将其移动到特定位点,即控制其移动,来对其进行修饰。

[0448] 多核苷酸(例如,核酸)是包含两个或更多个核苷酸的大分子。多核苷酸或核酸可以包含任何核苷酸的任何组合。核苷酸可以是天然存在的或人造的。可以对目标多核苷酸中的一个或多个核苷酸进行氧化或甲基化。目标多核苷酸中的一个或多个核苷酸可以被损坏。例如,多核苷酸可以包含嘧啶二聚体。此类二聚体通常与紫外线导致的损坏相关,且是皮肤黑素瘤的首要原因。可以对目标多核苷酸中的一个或多个核苷酸进行修饰,例如,用标记物或标签。合适的标记物如上所述。目标多核苷酸可以包含一个或多个间隔物。

[0449] 核苷酸通常包含核碱基、糖和至少一个磷酸基团。核碱基通常是杂环的。核碱基包括但不限于嘌呤和嘧啶,更具体地为腺嘌呤、鸟嘌呤、胸腺嘧啶、尿嘧啶和胞嘧啶。糖通常是戊糖。核苷酸糖包括但不限于核糖和脱氧核糖。核苷酸通常是核糖核苷酸或脱氧核糖核苷酸。核苷酸通常包含单磷酸、二磷酸或三磷酸。磷酸可以连接在核苷酸的5'或3'侧上。

[0450] 核苷酸包括但不限于单磷酸腺苷(AMP)、单磷酸鸟苷(GMP)、单磷酸胸苷(TMP)、单磷酸尿苷(UMP)、单磷酸胞苷(CMP)、单磷酸5-甲基胞苷、二磷酸5-甲基胞苷、三磷酸5-甲基胞苷、单磷酸5-羟甲基胞苷、二磷酸5-羟甲基胞苷、三磷酸5-羟甲基胞苷、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、脱氧单磷酸腺苷(dAMP)、脱氧单磷酸鸟苷(dGMP)、脱氧单磷酸胸苷(dTMP)、脱氧单磷酸尿苷(dUMP)和脱氧单磷酸胞苷(dCMP)。核苷酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP、dCMP和dUMP。

[0451] 核苷酸可以是脱碱基的(即缺少碱基)。核苷酸也可以缺少碱基和糖(即为C3间隔物)。

[0452] 多核苷酸中的核苷酸可以以任何方式彼此连接。如在核酸中一样,核苷酸通常通过其糖和磷酸基团连接。如嘧啶二聚体中一样,所述核苷酸可以通过其碱基连接。

[0453] 多核苷酸可以是单链或双链的。多核苷酸的至少一部分优选是双链的。

[0454] 多核苷酸可以是核酸,例如,脱氧核糖核酸(DNA)或核糖核酸(RNA)。目标多核苷酸可以包含与一条DNA链杂交的一条RNA链。多核苷酸可以是本领域已知的任何合成核酸,例如,肽核酸(PNA)、甘油核酸(GNA)、核糖核酸(TNA)、锁定核酸(LNA)或其它具有核苷酸侧链的合成聚合物。

[0455] 优选所述部分的三级结构是已知的。对所述部分三维结构的了解使得能够对所述部分进行修饰以促进其在本发明构建体中的功能。

[0456] 所述部分可以是任何尺寸并具有任何结构。例如,所述部分可以是寡聚物,例如,二聚体或三聚体。所述部分优选是由一种单体形成的小的球形多肽。这种部分易于处理,并且不太可能干扰解旋酶控制多核苷酸移动的能力,特别是如果融合到或插入到解旋酶序列中的话。

[0457] 所述部分的氨基和羧基末端优选非常接近。所述部分的氨基和羧基末端更优选存在于所述部分的同一面上。这种实例便于将所述部分插入解旋酶的序列中。例如,如果所述部分的氨基和羧基末端非常接近,则每个末端可以通过基因融合与解旋酶序列中的相邻氨基酸连接。

[0458] 还优选已知所述部分的活性位点的位置和功能。这防止对消除了所述部分的活性的活性位点进行修饰。还使得所述部分能够与解旋酶连接,使得所述部分与多核苷酸结合并控制其移动。对部分可以结合多核苷酸并定向多核苷酸的方式的了解也使得有效的构建体能够被设计。

[0459] 本发明的构建体在链测序时是有用的。所述部分优选在缓冲液背景中与多核苷酸结合,所述缓冲液背景与链测序和核苷酸的鉴别相容。所述部分优选在远高于正常生理水平的盐浓度,例如100mM至2M的盐浓度中至少具有残留活性。更优选地,对所述部分进行修饰以增加其在高盐浓度下的活性。也可以对所述部分进行修饰以改善其持续合成性、稳定性和保质期。

[0460] 可以通过表征来自诸如嗜盐、中度嗜盐细菌、嗜热和中度嗜热生物等嗜极生物的多核苷酸结合部分来确定合适的修饰,并通过定向进化方法来改变嗜温或嗜热核酸外切酶的耐盐性、稳定性和温度依赖性。

[0461] 多核苷酸结合部分优选包含一个或多个域,其独立地选自螺旋-发夹-螺旋(HhH)域、真核单链结合蛋白(SSB)、细菌SSB、古细菌SSB、病毒SSB、双链结合蛋白、滑移夹具、持续合成性因子、DNA结合环、复制起始蛋白、端粒结合蛋白、阻遏物、锌指(zinc fingers)和增殖细胞核抗原(PCNA)。

[0462] 螺旋-发夹-螺旋(HhH)域是多肽基序,其以序列非特异性方式结合DNA。其在与聚合酶融合时呈现盐稳定性和持续合成性,并增加其热稳定性。合适的域包括来自如甲烷嗜热菌(*Methanopyrus kandleri*) (SEQ ID NO:47)的拓扑异构酶V中的域H(残基696-751)和域HI(残基696-802)。如下所述,多核苷酸结合部分可以是如SEQ ID NO:48中所示的SEQ ID NO:47的域H-L。来自如甲烷嗜热菌的拓扑异构酶V是如下讨论的双链结合蛋白的示例。

[0463] HhH域优选包含SEQ ID NO:24或37或38中所示的序列或其变体。当该域用于本发明的构建体中时,该域增加了解旋酶的持续合成性和耐盐性。SEQ ID NO:24或37或38的变体是具有由SEQ ID NO:24或37或38的氨基酸序列变化而来并且具有多核苷酸结合活性的氨基酸序列的蛋白质。可以对此进行如上所述的测量。基于氨基酸的同一性,变体通常与整个序列的SEQ ID NO:24或37或38具有至少50%的同源性(或上述与解旋酶相关的%同源性中的任一个),并且具有多核苷酸结合活性。变体可以以上述与解旋酶相关的或下述与孔相关的任何方式不同于SEQ ID NO:24或37或38。如上所述,变体优选包含一个或多个取代的半胱氨酸残基和/或一个或多个取代的Faz残基以便于与解旋酶连接。

[0464] SSB以序列非特异性方式用高亲和力结合单链DNA。SSB以各种形式存在于生命的所有域中,并结合DNA作为单体或多聚体。采用氨基酸序列比对和对数(如隐马尔可夫模

型),可以根据序列同源性对SSB进行分类。Pfam家族(PF00436)包括所有与已知SSB序列相似的蛋白质。然后,可以根据蛋白质的结构分类(Structural Classification of Proteins,SCOP)对这群SSB进行进一步分类。SSB属于以下系属:分类(Class);所有 β 蛋白,折叠;OB-折叠,总科:核酸结合蛋白,家族:单链DNA结合域,SSB。在这个家族中,可以根据亚科对SSB进行分类,其中,在每个亚科中通常表征多个类型物种。

[0465] SSB可以来自真核生物,比如来自人类、小鼠、大鼠、真菌、原生动物或植物;来自原核生物,比如细菌和古细菌;或来自病毒。

[0466] 真核SSB作为复制蛋白A(RPA)已知。在大多数情况下,真核SSB是由不同尺寸单位形成的异源三聚体。一些较大单位(例如,酿酒酵母的RPA70)是稳定的并且以单体形式结合ssDNA。

[0467] 细菌性SSB结合DNA作为稳定的同源四聚体(例如,大肠杆菌、耻垢分枝杆菌和幽门螺杆菌)或同聚二聚体(例如,耐辐射球菌和海栖热袍菌)。来自古细菌基因组的SSB被认为与真核RPA有关。它们中只有几个,例如由泉古菌硫磺矿硫化叶菌(*crenarchaeote Sulfolobus solfataricus*)编码的SSB,是同聚四聚体。大多数其它物种中的SSB与真核生物中的复制蛋白更接近,被称为RPA。在这些物种的一部分中,其被证明是单体的(詹氏甲烷球菌和嗜热自养甲烷杆菌)。但是,其它古细菌物种,包括火焰色古球状菌(*Archaeoglobus fulgidus*)和*Methanococcoides burtonii*,似乎各自包含两个开放阅读框,其序列与RPA序列相似。在蛋白质水平没有证据,并且没有关于DNA结合能力或寡聚状态的公开数据。然而,这些基因的每一个中存在的两种寡核苷酸/寡糖(OB)折叠(其中一个*M.burtonii* ORF存在下的三种OB折叠)表明它们也结合单链DNA。

[0468] 病毒性SSB结合DNA作为单体。这一点以及病毒性SSB相对较小的尺寸使得其适合于与其它蛋白质进行遗传融合,例如,通过柔性肽连接器。可选地,SSB可以分开进行表达并通过化学方法(例如,半胱氨酸、非天然氨基酸)与其它蛋白质连接。这将在以下进行更详细的讨论。

[0469] SSB优选为(i)包含羧基末端(C-末端)区域的SSB,所述区域不具有净负电荷;或(ii)在其C-末端区域包含一个或多个修饰的修饰SSB,其减少C末端区域的净负电荷。这种SSB不阻塞跨膜孔,因此能够表征目标多核苷酸。

[0470] SSB,包含不具有净负电荷的C-末端区域,的示例包括但不限于人线粒体SSB(HsmtSSB;SEQ ID NO:39,人复制蛋白A 70kDa亚基,人复制蛋白A 14kDa亚基,来自尖毛虫的端粒末端结合蛋白 α 亚基,来自尖毛虫的端粒末端结合蛋白 β 亚基的核心域,来自粟酒裂殖酵母的端粒保护蛋白1(Pot1),人Pot1,来自小鼠或大鼠的BRCA2的OB-折叠域,来自phi29(SEQ ID NO:40)的p5蛋白或这些蛋白中任何一个的变体。变体是具有由野生型蛋白质的氨基酸序列变化而来并且具有单链多核苷酸结合活性的氨基酸序列的蛋白质。可以采用本领域已知(和如上所述)的方法来测定多核苷酸结合活性。例如,变体结合单链多核苷酸的能力可以如实施例所述进行确定。

[0471] 基于SEQ ID NO:39或40在整个序列上的氨基酸同一性,SEQ ID NO:39或40的变体通常与SEQ ID NO:39或40具有至少50%的同源性(或上述与解旋酶相关的%同源性中的任一个),并且具有单链多核苷酸结合活性。变体可以以与解旋酶相关的上述任一方式不同于SEQ ID NO:39或40。具体地,变体可以具有如表7和表8中所示的一个或多个保守取代。

[0472] SSB,需要在其C-末端区域进行一个或多个修饰以减少净负电荷,的示例包括但不限于大肠杆菌的SSB (EcoSSB;SEQ ID NO:41,结核分枝杆菌的SSB,耐辐射球菌的SSB,嗜热栖热菌的SSB,来自硫磺矿硫化叶菌的SSB,人复制蛋白A 32kDa亚基 (RPA32) 片段,来自酿酒酵母的CDC13 SSB,来自大肠杆菌的Primosomal复制蛋白N (PriB),来自拟南芥的PriB,假定蛋白At4g28440,来自T4的SSB (gp32;SEQ ID NO:42),来自RB69的SSB (gp32;SEQ ID NO:25),来自T7的SSB (gp2.5;SEQ ID NO:26)或这些蛋白质中任一个的变体。因此,本发明方法中使用的SSB可以衍生自这些蛋白质中的任一个。

[0473] 除了C-末端区域中的一个或多个修饰之外,本方法中使用的SSB还可以包括额外的修饰,其在C-末端区域的外部或者不减少C-末端区域净负电荷。换言之,本发明方法中使用的SSB衍生自野生型蛋白质的变体。变体是具有由野生型蛋白质的氨基酸序列变化而来并且具有单链多核苷酸结合活性的氨基酸序列的蛋白质。如上所述,可以确定多核苷酸结合活性。

[0474] 本发明中使用的SSB可以衍生自SEQ ID NO:25,26,41或42的变体。换言之,可以将SEQ ID NO:25,26,41或42的变体用作本发明中使用的SSB的起点,但是实际使用的SSB还包括在其C-末端区域中的一个或多个修饰,这减少了C-末端区域的净负电荷。基于SEQ ID NO:25,26,41或42在整个序列上的氨基酸同一性,SEQ ID NO:25,26,41或42的变体通常与SEQ ID NO:25,26,41或42具有至少50%的同源性(或上述与解旋酶相关的%同源性中的任一个),并具有单链多核苷酸结合活性。变体可以以上述与解旋酶相关的任一方式不同于SEQ ID NO:25,26,41或42。具体地,变体可以具有如表7和表8中所示的一个或多个保守取代。

[0475] 根据正常的蛋白N到C命名法,可以直接鉴定SSB的C-末端区域。SSB的C-末端区优选为大约SSB的C-末端端部的最后三分之一,例如,SSB的C-末端端部的最后三分之一。SSB的C-末端区域更优选为SSB的C-末端端部的大约最后四分之一、五分之一或八分之一,例如,SSB的C-末端端部的最后四分之一、五分之一或八分之一。可以根据SSB蛋白质一级结构的氨基酸的数量或根据SSB蛋白质一级结构的实际长度来测量SSB的最后三分之一、四分之一、五分之一或八分之一。在N至C方向上的各种氨基酸的长度是本领域已知的。

[0476] C-末端区域优选地是从SSB的C末端端部的大约最后10个到大约最后60个氨基酸。C-末端区域更优选地是SSB的C-末端端部的大约最后15个、大约最后20个、大约最后25个、大约最后30个、大约最后35个、大约最后40个、大约最后45个、大约最后50个或大约最后55个氨基酸。

[0477] C-末端区域通常包含甘氨酸和/或脯氨酸富集区域。该脯氨酸/甘氨酸富集区域给C-末端区域提供了灵活性并且可用于鉴定C-末端区域。

[0478] 申请号为PCT/GB2013/051924(公开为W02014/013259)的国际申请中公开了用于减少净负电荷的适当修饰。SSB可以是该国际申请中公开的任一SSB。

[0479] 修饰的SSB最优选包含选自SEQ ID NO:33,34,43至46中所示序列的序列。

[0480] 双链结合蛋白用高亲和力结合双链DNA。合适的双链结合蛋白包括但不限于突变体S (MutS;NCBI参考序列:NP_417213.1;SEQ ID NO:49)、Sso7d(硫磺矿硫化叶菌P2;NCBI参考序列:NP_343889.1;SEQ ID NO:50;《核酸研究》(Nucleic Acids Research),2004,第32卷第3号,1197-1207)、Sso10b1 (NCBI参考序列:NP_342446.1;SEQ ID NO:51)、Sso10b2

(NCBI参考序列:NP_342448.1;SEQ ID NO:52)、色氨酸阻遏物(Trp阻遏物;NCBI参考序列:NP_291006.1;SEQ ID NO:53)、 λ 抑制子(NCBI参考序列:NP_040628.1;SEQ ID NO:54)、Cren7(NCBI参考序列:NP_342459.1;SEQ ID NO:55)、主要组蛋白类别H1/H5,H2A,H2B,H3和H4(NCBI参考序列:NP_066403.2,SEQ ID NO:56)、dsbA(NCBI参考序列:NP_049858.1;SEQ ID NO:57)、Rad51(NCBI参考序列:NP_002866.2;SEQ ID NO:58)、滑移夹具和拓扑异构酶V Mka(SEQ ID NO:47)或这些蛋白中任一个的变体。基于SEQ ID NO:47,49,50,51,52,53,54,55,56,57或58在整个序列上的氨基酸同一性,SEQ ID NO:47,49,50,51,52,53,54,55,56,57或58的变体通常与SEQ ID NO:47,49,50,51,52,53,54,55,56,57或58具有至少50%的同源性(或上述与解旋酶相关的%同源性中的任一个),并且具有单链多核苷酸结合活性。变体可以以上述与解旋酶相关的任一方式不同于SEQ ID NO:47,49,50,51,52,53,54,55,56,57或58。具体地,变体可以具有表7和表8中所示的一个或多个保守取代。大多数聚合酶通过与滑移夹具相互作用来实现持续合成性。通常,这些滑移夹具是包围dsDNA的多聚体蛋白(同源二聚体或同源三聚体)。这些滑移夹具需要辅助蛋白(夹具装载器)来使其在ATP依赖过程中集合在DNA螺旋周围。它们也不作为拓扑系链直接与DNA接触。由于滑移夹具通过聚合酶域以特定方式与其同源聚合酶相互作用,所以该片段可以与解旋酶融合,以便刺激解旋酶补充到滑移夹具上。这种相互作用可以通过共价键的生成(引入半胱氨酸或非天然氨基酸)进一步稳定。

[0481] 与DNA滑移夹具相关,持续合成性因子是病毒性蛋白,该病毒性蛋白将其同源聚合酶锚定到DNA,从而导致所生成片段的长度显著增加。该病毒性蛋白可以是单体的(正如来自单纯疱疹病毒1的UL42一样)或多聚体的(来自巨细胞病毒的UL44是二聚体),其在DNA链周围不形成闭合环,并且其直接与DNA接触。UL42在不降低其相应聚合酶的速率的情况下,增加了持续合成性,这表明其以不同于SSB的模式与DNA相互作用。UL42优选包含SEQ ID NO:27或SEQ ID NO:32其变体中所示的序列。SEQ ID NO:27或32的变体是具有由SEQ ID NO:27或32的氨基酸序列变化而来并且具有多核苷酸结合活性的氨基酸序列的蛋白质。可以对此进行如上所述的测量。基于SEQ ID NO:27或32在整个序列上的氨基酸同一性,变体通常与SEQ ID NO:27或32具有至少50%的同源性(或上述与解旋酶相关的%同源性中的任一个),并且具有多核苷酸结合活性。变体可以以上述与解旋酶相关或下述与孔相关的任一方式不同于SEQ ID NO:27或SEQ ID NO:32。如上所述,变体优选包含一个或多个取代的半胱氨酸残基和/或一个或多个取代的Faz残基,以便于与解旋酶连接。

[0482] 可以通过遗传融合或化学连接(半胱氨酸、非天然氨基酸)完成UL42与解旋酶的连接。由于结合UL42的聚合酶多肽在晶体结构中是可见的,所以这35个氨基酸(残基1200-1235)可以与解旋酶的C-末端融合,该多肽和持续合成性因子之间的天然亲和力用于形成复合物。可以通过引入共价相互作用(半胱氨酸或非天然氨基酸)使相互作用稳定。一个选择是利用天然UL42半胱氨酸(C300),其位于多肽相互作用位点附近,并将位点突变引入到聚合酶多肽(例如,L1234C)中。

[0483] 增加聚合酶持续合成性的已知方法是通过利用噬菌体T7DNA聚合酶(残基258-333)的大肠杆菌硫氧还蛋白(Trx)和硫氧还蛋白结合域(TBD)之间的相互作用。Trx与TBD结合导致多肽的构象变成结合DNA的构象。TBD被认为对DNA链进行钳制(clamp down)并限制聚合酶的脱离率,从而增加了持续合成性。通过将TBD转移到非持续合成性聚合酶上来制备

嵌合聚合酶,从而导致聚合片段的长度增加1000倍。并没有尝试将TBD与任何其它类别的蛋白质连接,但却设计了TBD和Trx之间的共价连接,并且可以用于稳定相互作用。

[0484] 一些解旋酶在体内使用辅助蛋白来实现持续合成性(例如,对于大肠杆菌Rep解旋酶的来自噬菌体 Φ x174的cisA和来自噬菌体M13的基因II蛋白)。这些蛋白质中的一些已显示与多于一种解旋酶相互作用(例如,MutL作用于UvrD和Rep,但程度不尽相同)。这些蛋白质具有固有的DNA结合能力,其中一些鉴定特异性DNA序列。这些辅助蛋白中的一些本身与特异性DNA序列共价连接的能力也可用于为解旋酶活性创建设定起始点。

[0485] 对染色体末端进行保护的蛋白质以高度特异性方式与端粒ssDNA序列结合。这种能力可以按原样利用,也可以通过使用位点突变进行利用以便消除序列特异性。

[0486] 小DNA (Small DNA) 结合基序(如螺旋-转角-螺旋) 鉴定特异性DNA序列。至于噬菌体434阻遏物,设计62残基片段并显示出保留DNA结合能力和特异性。

[0487] 真核蛋白中的基序丰富,锌指由大约30个与DNA以特定方式结合的氨基酸组成。通常,每个锌指只鉴定三个DNA碱基,但可以对多个锌指进行连接以获得较长序列的鉴定。

[0488] 增殖细胞核抗原 (PCNA) 形成非常紧的夹具(环状物),其沿dsDNA或ssDNA上下滑移。来自泉古菌的PCNA在其为异源三聚体时是独特的,因此可以使一个亚基功能化(functionalise)并保留活性。它的亚基在SEQ ID NO:28,29和30中示出。PCNA优选为包含SEQ ID NO:28,29和30中所示序列或其变体的三聚体。PCNA滑移夹具(NCBI参考序列:ZP_06863050.1;SEQ ID NO:59)形成二聚体。PCNA优选为包含SEQ ID NO:59或其变体的二聚体。变体是具有由SEQ ID NO:28,29,30或59的氨基酸序列变化而来并具有多核苷酸结合活性的氨基酸序列的蛋白质。可以对此进行如上所述测量。变体通常是三聚体或二聚体并且具有多核苷酸结合活性,其中,所述三聚体基于SEQ ID NO:28,29和30在整个序列上的氨基酸同一性,包含与SEQ ID NO:28,29和30具有至少50%同源性的序列(或上述与解旋酶相关的%同源性中的任一个);所述二聚体基于SEQ ID NO:59在整个序列上的氨基酸同一性,包含与SEQ ID NO:59具有至少50%同源性的序列(或上述与解旋酶相关的%同源性中的任一个)。变体可以以上述与解旋酶相关或下述与孔相关的任一方式包含不同于SEQ ID NO:28,29,30或59的序列。如上所述,变体优选包含一个或多个取代的半胱氨酸残基和/或一个或多个取代的Faz残基以便于与解旋酶连接。在优选实施方案中,对来自泉古菌的PCNA的亚基1和2(即,SEQ ID NO:28和29或其变体)进行连接,例如,遗传融合,并将所得蛋白质与解旋酶连接以形成本发明的构建体。在使用构建体期间,一旦构建体结合了多核苷酸,可以加入亚基3(即,SEQ ID NO:30或其变体)以使PCNA夹具(或环状物)变完整。在优选实施方案中,将PCNA滑移夹具的一个单体(即,SEQ ID NO:59或其变体)与解旋酶连接,例如,遗传融合,以形成本发明的构建体。在使用构建体期间,一旦构建体结合了多核苷酸,可以加入第二个单体(即,SEQ ID NO:59或其变体)以使PCNA夹具(或环状物)变完整。

[0489] 多核苷酸结合基序可以选自以下表5所示基序中的任一种。

[0490] 表5 合适的多核苷酸结合基序

[0491]

编号	名称	分类	生物体	组织	序列	功能形式	MW (Da)	备注
1	SSBEco	ssb	大肠杆菌	1QVC, 1EYG	P0AGE0	同源四聚体	18975	
2	SSBBhe	ssb	韩瑟勒巴通氏菌	3LGJ, 3PGZ	<u>Q6G302</u>	同源四聚体	16737	仅结构
3	SSBCbu	ssb	贝氏柯克斯体	3TQY	<u>Q83EP4</u>	同源四	17437	仅结构

[0492]

						聚体		
4	SSBTma	ssb	海栖热袍菌	1Z9F	<u>Q9WZ73</u>	同源二聚体	16298	小、热稳定、非盐依赖性DNA结合
5	SSBHpy	ssb	幽门螺杆菌	2VW9	<u>O25841</u>	同源四聚体	20143	
6	SSBDra	ssb	耐辐射球菌	1SE8	Q9RY51	同源二聚体	32722	
7	SSBTaq	ssb	水生栖热菌	2FXQ	Q9KH06	同源二聚体	30026	
8	SSBMsm	ssb	耻垢分枝杆菌	3A5U, 1X3E	Q9AF15	同源四聚体	17401	四聚体比大肠杆菌更稳定, 少盐依赖性结合
9	SSBSso	ssb/RPA	硫磺矿硫化叶菌	1O7I	Q97W73	同源四聚体	16138	与RPA的相似性
10	SSBMHsmt	ssb	智人	3ULL	Q04837	同源四聚体	17260	
11	SSBMle	ssb	麻风	<u>3AFP</u>	<u>P46390</u>	同	17701	

[0493]

			分枝杆菌			源四聚体		
12	gp32T4	ssb	噬菌体 T4	1GPC	P03695	单体	33506	不存在 DNA 时为同源二聚体，结合 DNA 时为单体
13	gp32RB69	ssb	噬菌体 RB69	2A1K	Q7Y265	单体	33118	
14	gp2.5T7	ssb	噬菌体 T7	1JE5	<u>P03696</u>	单体	25694	
15	UL42	持续合成性因子	疱疹病毒 1	1DML	<u>P10226</u>	单体	51159	结合 ssDNA、dsDNA，结构显示与聚合酶连接
16	UL44	持续合成性因子	疱疹病毒 5(巨细胞病毒)	1YYP	P16790	同源二聚体	46233	在 DNA 上形成 C 形状夹具
17	pf8	持续合成性因子	KSHV	3I2M	<u>Q77ZG5</u>	同源二聚体	42378	
18	RPAMja	RPA	詹氏甲烷球菌	3DM3	Q58559	单体	73842	包含 4 个 OB 折叠。片段结构
19	RPAMm	RPA	甲烷	3E0E,	Q6LYF9	单	71388	核心域结

[0494]

	a		球菌	2K5V		体		构
20	RPAMth	RPA	嗜热自养甲烷杆菌			单体	120000	显示直接与 Hel308 相互作用。序列来自纸。
21	RPA70Sce	RPA	酿酒酵母	1YNX	P22336	异源三聚体	70348	单元具有两种 OB 折叠并结合 DNA
22	RPAMbu1	RPA	伯顿甲烷球菌		Q12V72	?	41227	鉴定出三种 OB 折叠
23	RPAMbu2	RPA	伯顿甲烷球菌		Q12W96	?	47082	鉴定出两种 OB 折叠
24	RPA70Hsa	RPA	智人	1JMC	<u>P27694</u>	异源三聚体	68138	
25	RPA14Hsa	RPA	智人	3KDF	<u>P35244</u>	异源三聚体	13569	与 RPA32 的复合物
26	gp45T4	滑移夹具	噬菌体 T4	1CZD	P04525	同源三聚体	24858	环状线型 DNA
27	BetaEco	滑移夹具	大肠杆菌	3BEP	<u>P0A988</u>	同源二聚体	40587	环状线型 DNA，可结合剩余 ssDNA
28	PCNASc	滑移夹具	酿酒	1PLQ,	<u>P15873</u>	同	28916	环状线型

[0495]

	e		酵母	3K4X		源二聚体		DNA
29	PCNATko	滑移夹具	鹿儿岛栖热球菌	3LX1	<u>Q5JF32</u>	同源二聚体	28239	
30	PCNAHvo	滑移夹具	沃氏嗜盐富饶菌	<u>3IFV</u>	<u>D0VWY8</u>	同源二聚体	26672	
31	PCNAPfu	滑移夹具	强烈炽热球菌	1GE8	<u>O73947</u>	同源二聚体	28005	
32	PCNAMbu	滑移夹具	伯顿甲烷球菌		Q12U18	同源二聚体	27121	推断同源
33	BetaMtu	滑移夹具	结核分支杆菌	3P16	<u>Q50790</u>	同源二聚体	42113	
34	BetaTma	滑移夹具	海栖热袍菌	1VPK	<u>Q9WYA0</u>	同源二聚体	40948	
35	BetaSpy	滑移夹具	酿脓链球菌	2AVT	Q9EVR1	同源二聚	41867	

[0496]

						体		
36	gp45RB69	滑移夹具	噬菌体 RB69	1B77	<u>O80164</u>	同源二聚体	25111	结构显示与来自聚合酶的多肽相互作用
37	p55Hsa	DNA 结合蛋白	智人 (线粒体)	2G4C, 3IKL, 3IKM	<u>Q9UHN</u>	单体	54911	与特定聚合酶域相互作用
38	p55Dme	DNA 结合蛋白	黑腹果蝇 (Drosophila melanogaster)		Q9VJV8	单体	41027	与表现耐盐性、持续合成性且活性增加的聚合酶 γ 相关联
39	p55Xla	DNA 结合蛋白	非洲爪蟾		Q9W6G7	单体	52283	
40	RepDsau	复制起始蛋白	金黄色葡萄球菌		P08115	同源二聚体	37874	增加 PcrA 的持续合成性, 共价且特异性连接 DNA
41	G2P	复制起始蛋白质	肠杆菌噬菌体		P69546	单体	46168	增加 Rep 的持续合成性, 共价且特异性连接 DNA
42	MutLEco	错配修复蛋白	大肠杆菌	1BKN, 1B62, 1B63	P23367	同源二聚体	67924	增加 UvrD (和 Rep) 的持续合成性
43	KuMtu	DNA 修复	结核		O05866	同	30904	增加

[0497]

		蛋白	分支杆菌			源二聚体		UvrD1 的持续合成性。结构可用于人 Ku
44	OnTEBP	端粒结合蛋白	尖毛虫- α		<u>P16458</u>		41446	
45	EcrTEBP	端粒结合蛋白	游仆虫		Q06183	单体	53360	与 OnTEBP 同源，基因组中没有 β 亚基
46	TteTEBP	端粒结合蛋白	嗜热四膜虫 (<i>Tetrahymena thermophila</i>)		Q23FB9	异源二聚体	53073	与 OnTEBP- α 同源
47	pot1Spo	端粒结合蛋白	粟酒裂殖酵母		O13988	单体	64111	与 TEBP 相关
48	Cdc13pSce	端粒结合蛋白	酿酒酵母		C7GSV7	单体	104936	与端粒 DNA 特异性结合
49	C1	阻遏物	噬菌体 434		P16117	同源二聚体	10426	特异性结合 DNA 作为同二聚体
50	LexA	阻遏物	大肠杆菌	1LEB	<u>P0A7C2</u>	同源二聚体	22358	特异性结合 DNA 作为同二聚体

[0498] 多核苷酸结合部分优选衍生自多核苷酸结合酶。多核苷酸结合酶是一种多肽，其能够与多核苷酸结合并相互作用和修饰多核苷酸的至少一个性质。所述酶可以通过对多核苷酸进行断裂以形成单个核苷酸或核苷酸的较短链，例如二核苷酸或三核苷酸，来对其进行修饰。所述酶可以通过将多核苷酸取向或移动到特定定位点来对其进行修饰。只要多核苷

酸结合部分能够结合多核苷酸并控制其移动,其就不需要显示酶活性。例如,所述部分可以衍生自己已经进行修饰以便消除其酶活性的酶,或者可以在防止其作为酶的情况下进行使用。

[0499] 多核苷酸结合部分优选衍生自溶核酶。所述酶更优选衍生自酶分类(EC)组3.1.11,3.1.13,3.1.14,3.1.15,3.1.16,3.1.21,3.1.22,3.1.25,3.1.26,3.1.27,3.1.30和3.1.31中的任一个。所述酶可以是申请号为PCT/GB10/000133(公开为W02010/086603)的国际申请中公开的任一个。

[0500] 优选的酶是核酸外切酶、聚合酶、解旋酶和拓扑异构酶,例如,旋转酶。合适的核酸外切酶包括但不限于来自大肠杆菌的核酸外切酶I、来自大肠杆菌的核酸外切酶III、嗜热链球菌(*T. thermophilus*)中的RecJ和噬菌体λ核酸外切酶、TatD核酸外切酶及其变体。

[0501] 聚合酶优选为部分分类(Moiety Classification)(MC)组2.7.7.6,2.7.7.7,2.7.7.19,2.7.7.48和2.7.7.49中的任一个。聚合酶优选是DNA依赖性DNA聚合酶、RNA依赖性DNA聚合酶、DNA依赖性RNA聚合酶或RNA依赖性RNA聚合酶。聚合酶可以是PyroPhage® 3173 DNA聚合酶(其可以从Lucigen®公司商业购得)、SD聚合酶(可以从Bioron®商业购得)或其变体。多核苷酸结合部分优选衍生自Phi29DNA聚合酶(SEQ ID NO:31)。所述部分可以包含SEQ ID NO:101中所示的序列或其变体。SEQ ID NO:31的变体是具有由SEQ ID NO:31的氨基酸序列变化而来并具有多核苷酸结合活性的氨基酸序列的酶。可以对此进行如上所述测量。所述变体可以包括促进多核苷酸结合和/或促进其在高盐浓度和/或室温下活性的修饰。

[0502] 基于氨基酸同一性,对于SEQ ID NO:31的氨基酸序列的整个长度,变体优选与该序列至少50%同源。更优选地,基于氨基酸同一性,对于SEQ ID NO:31的氨基酸序列的整个序列,变体多肽与该序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,更优选地,至少95%、97%或99%同源。在200或更多,例如230、250、270或280或更多的连续氨基酸长度上时,可以具有至少80%,例如至少85%、90%或95%的氨基酸同一性(“严格同源性”)。同源性如上所述进行确定。变体可以按下述针对SEQ ID NO:2和4所描述的多种方式中的任意一种不同于野生型序列。

[0503] 解旋酶可以是上述任一种,包括SEQ ID NO:8至23中的任一个。以下对解旋酶二聚体和多聚体进行详细讨论。多核苷酸结合部分可以是衍生自解旋酶的多核苷酸结合域。例如,多核苷酸结合部分优选包含SEQ ID NO:35或36中所示的序列或其变体。SEQ ID NO:35或36的变体是具有由SEQ ID NO:35或36的氨基酸序列变化而来并且具有多核苷酸结合活性的氨基酸序列的蛋白质。可以对此进行如上所述测量。所述变体可以包括促进多核苷酸结合和/或促进其在高盐浓度和/或室温下活性的修饰。

[0504] 基于氨基酸同一性,对于SEQ ID NO:35或36的氨基酸序列的整个长度,变体优选与该序列至少50%同源。更优选地,基于氨基酸同一性,对于SEQ ID NO:35或36的氨基酸序列的整个序列,变体多肽与该序列至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%,更优选地,至少95%、97%或99%同源。在40或更多,例如50、60、70或80或更多的连续氨基酸长度上时,可以具有至少80%,例如,至少85%、90%或95%的氨基酸同一性(“严格同源性”)。同源性如上所述进行确定。变体可以按下述与SEQ ID NO:2和4所描述的多种方式中的任意一种不同于野生型序列。

[0505] 拓扑异构酶优选为部分分类(MC)组5.99.1.2和5.99.1.3中的任一个。

[0506] 多核苷酸结合部分可以是上述酶中的任一种。

[0507] 所述部分可以用显示标记物进行标记。所述标记物可以是上述任一种。

[0508] 所述部分可以与任何产生部分的生物体(moiety-producing organism),诸如大肠杆菌、嗜热链球菌或噬菌体分离开来,或通过合成或重组方式进行制备。例如,所述部分可以通过如下所述的体外转译和转录来合成。所述部分可以在如下所述纯化后大规模生产。

[0509] 解旋酶寡聚物

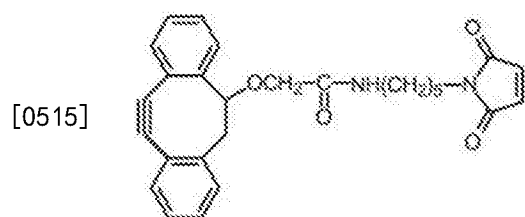
[0510] 通过以上讨论很明显,多核苷酸结合部分优选衍生自解旋酶。例如,其可以是来自解旋酶的多核苷酸域。所述部分更优选包含一个或多个解旋酶。解旋酶可以是上述与本发明的构建物相关的任一种,包括本发明的解旋酶和根据本发明未进行修饰的解旋酶。在这些实施方案中,本发明的构建体当然包括两个或更多个连接在一起的解旋酶。根据本发明,对解旋酶中的至少一个进行修饰。构建体可以包含两个、三个、四个、五个或更多个解旋酶。换言之,本发明的构建体可以包含解旋酶二聚体、解旋酶三聚体、解旋酶四聚体、解旋酶五聚体等。

[0511] 两个或更多个解旋酶可以以任何方向连接在一起。可以通过每个解旋酶中相同的氨基酸位点或空间上邻近的氨基酸位点连接相同或相似的解旋酶。这被称为“头对头”形式。可选地,可以通过每个解旋酶的相对或不同侧上的位点连接相同或相似的解旋酶。这被称为“头对尾”形式。由三个相同或相似的解旋酶组成的解旋酶三聚体可以包括头对头和头对尾形式。

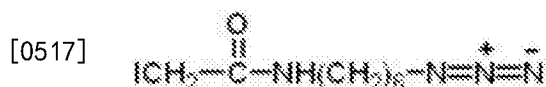
[0512] 两个或更多个解旋酶可以彼此不同(即,构建体是异源二聚体、三聚体、四聚体或五聚体等)。例如,本发明的构建体可以包括(a)本发明的一个或多个解旋酶和一个或多个根据本发明未进行修饰的解旋酶或(b)本发明的两个或更多个不同的解旋酶。构建体可以包含相同的Dda解旋酶的不同变体。例如,构建体可以包含上述解旋酶中任一个的两种变体,其中,在每个变体的不同位点上引入一个或多个半胱氨酸残基或Faz残基。在这种情况下,解旋酶可以是头对尾的形式。

[0513] 异源二聚体可以以两种可能的方式形成。第一种涉及使用如上所述的同源双功能连接器。可以用大量过量的连接器以这样一种方式,即一个连接器与蛋白质的一个分子连接,对解旋酶变体中的一个进行修饰。然后将这种经连接器修饰的变体从未修饰的蛋白质、可能的同源二聚体和未反应的连接器中纯化出来以便与另一个解旋酶变体反应。然后将所得到的二聚体从其它物质中纯化出来。

[0514] 第二种涉及使用异源双功能连接器。例如,可以用第一PEG连接器对解旋酶变体中的一个进行修饰,其中所述第一PEG连接器在其一端包含马来酰亚胺或碘乙酰胺官能团和另一端包含环辛炔官能团(DIBO)。如下示出了这种情况的示例:



[0516] 可以用第二PEG连接器对第二个解旋酶变体进行修饰,所述第二PEG连接器在其一端包含马来酰亚胺或碘乙酰胺官能团和另一端包含叠氮官能团。如下示出了这种情况的示例:

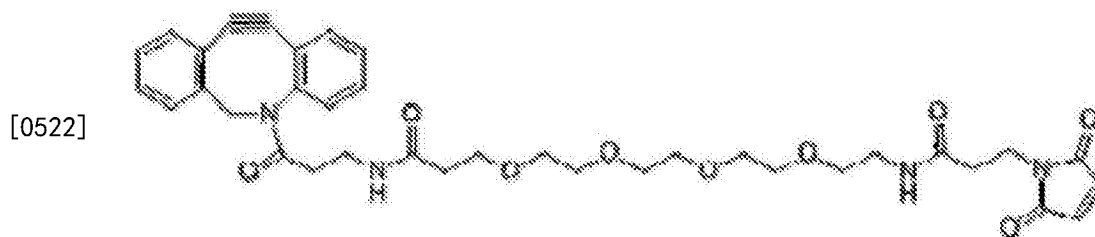


[0518] 然后可以对具有两个不同连接器的两个解旋酶变体进行纯化并点击在一起(使用无铜点击化学copper free click chemistry)以制备二聚体。无铜点击化学由于其期望性质而被用于这些应用中。例如,它对蛋白质是快速、干净且无毒的。然而,其它合适的生物正交化学包括但不限于施陶丁格化学、肼或酰肼/醛或酮试剂(HyNic+4FB化学,包括所有的Solulink™试剂)、Diels-Alder试剂对和硼酸/水杨酸异羟肟酸(salicyhydroxamate)试剂。

[0519] 这两种连接同一解旋酶的两个不同变体的方式对于上述任一构建体也是有效的,其中,解旋酶和部分彼此不同,例如,两个不同解旋酶的二聚体和解旋酶-聚合酶二聚体。

[0520] 类似的方法也可用于连接不同的Faz变体。可以用大量过量的连接器以这样一种方式,即一个连接器与蛋白质的一个分子连接,对一个Faz变体进行修饰。然后可以将这种经连接器修饰的Faz变体从未修饰的蛋白质、可能的同源二聚体和未反应的连接器中纯化出来以便与第二个Faz变体反应。然后可以将所得到的二聚体从其它物质中纯化出来。

[0521] 也可以通过连接同一解旋酶或不同解旋酶的半胱氨酸变体和Faz变体来制备异源二聚体。可以将一端包含马来酰亚胺或碘乙酰胺官能团和另一端包含DBCO官能团的异源双功能PEG连接器用于突变体的这种组合中。如下示出了这种连接器的示例(DBCO-PEG4-马来酰亚胺):

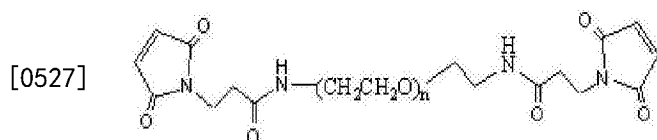


[0523] 通过改变两个官能团之间的PEG单元的数量可以改变连接器的长度。

[0524] 解旋酶异源三聚体可以包括三种不同类型的解旋酶。对于包含多于三种解旋酶的寡聚物也是如此。构建体内的两个或更多个解旋酶可以是同一种解旋酶的不同变体,例如,SEQ ID NO:8至23中任一个的不同变体。可以在不同位点上对不同变体进行修饰,以便于通过所述不同位点进行连接。因此,异源三聚体可以是头对尾和头对头形式。

[0525] 本发明构建体中的两个或更多个解旋酶可以彼此相同(即,构建体是同源二聚体、三聚体、四聚体或五聚体等)。在这些实施方案中,优选使用每个解旋酶中的相同位点对解旋酶进行连接。因此,解旋酶是头对头连接。可以使用在相同位点上被取代到解旋酶中的半胱氨酸残基或Faz残基对解旋酶进行连接。

[0526] 可以采用包含诸如马来酰亚胺或碘乙酰胺等硫醇反应性基团的同源双功能连接器来连接相同解旋酶变体中的半胱氨酸残基。这些官能团可以位于聚乙二醇(PEG)链的末端,如下示例:



[0528] 连接器的长度可以变化以适应所需应用。例如， n 可以是2, 3, 4, 8, 11, 12, 16或更多。PEG连接器由于其具有良好的性质，如水溶性，而是合适的。其它非PEG连接器也可用于半胱氨酸连接。

[0529] 通过使用类似的方法，也可以将相同的Faz变体制成同源二聚体。可以使用具有DIBO官能团的同源双功能连接器来连接相同Faz变体的两个分子，从而利用不含 Cu^{2+} 的点击化学制备同源二聚体。以下给出了连接器的示例：



[0531] PEG连接器的长度可以变化，包括2, 4, 8, 12, 16或更多个PEG单元。也可以制备这种连接器以包含荧光标签以便于量化。也可以将这种荧光标签并入马来酰亚胺连接器。

[0532] 如果本发明的解旋酶中引入了两个或更多个半胱氨酸残基或非天然氨基酸并且不同解旋酶单体中的不同半胱氨酸或非天然氨基酸连接在一起，则也可以以头对尾形式制备同源二聚体或更长的同型寡聚物。例如，可以从包含Y279C和G357C的SEQ ID NO:8的变体中制备同源寡聚物，一个单体中的279处的C可以与另一单体中的357处的C连接。类似地，可以从包含I281C和G357C的SEQ ID NO:8的变体中制备同源寡聚物，一个单体中的281处的C可以与另一单体中的357处的C连接。Faz代替C被引入到这些位点时，也是如此。这种C和Faz突变体使得能够产生一系列或一连串的解旋酶。

[0533] 多核苷酸序列

[0534] 本发明提供一种包含对本发明解旋酶或本发明构建体进行编码的序列的多核苷酸。所述多核苷酸可以由这样的序列组成。多核苷酸可以是上述任一种。

[0535] 可以使用本领域已知的方法来表达本文中所述蛋白质中的任一种。可以使用本领域的标准方法来分离和复制多核苷酸序列。可以从产生生物体（例如，*Methanococcoides burtonii*）的解旋酶和/或产生生物体（例如，大肠杆菌）的SSB中提取染色体DNA。可以采用涉及特异性引物的PCR来对编码了感兴趣序列的基因进行扩增。然后可以将扩增的序列并入重组可复制载体中，比如克隆载体。所述载体可用于在相容的宿主细胞中复制多核苷酸。因此，可以通过将编码了感兴趣序列的多核苷酸引入可复制载体中，将载体引入相容的宿主细胞中，并在引起载体复制的条件下培养宿主细胞来制备多核苷酸序列。载体可以从宿主细胞中回收出来。本领域已知用于克隆多核苷酸的合适宿主细胞，以下对其进行更详细的描述。

[0536] 多核苷酸序列可以克隆到合适的表达载体中。在表达载体中，多核苷酸序列通常与控制序列可操作地连接，该控制序列能够通过宿主细胞提供编码序列的表达。这种表达载体可用于表达构建体。

[0537] 术语“可操作地连接”是指并列连接，其中，所描述的组分处于允许其以预期方式发挥作用的的关系。控制序列可操作地连接编码序列是以这一方式，即编码序列的表达在与

控制序列相容的条件下实现,进行连接。可以将相同或不同多核苷酸的多个副本引入载体中。

[0538] 然后将表达载体引入合适的宿主细胞中。因此,可以通过将编码了构建体的多核苷酸序列插入表达载体中,将载体引入相容的细菌宿主细胞中,并在引起多核苷酸表达的条件下培养宿主细胞来制备构建体。

[0539] 载体可以是,例如,提供有复制起点的质粒、病毒或噬菌体载体,可选地,表达所述多核苷酸序列的启动子以及,可选地,启动子的调节子。载体可以包含一个或多个选择性标记基因,例如,氨苄青霉素抗性基因。可以选择启动子和其它表达调节信号以便与宿主细胞相容,对此设计表达载体。通常使用T7、trc、lac、ara或 λ 启动子。

[0540] 宿主细胞通常以高水平表达构建体。选择用多核苷酸序列转化的宿主细胞与用于转化细胞的表达载体相容。宿主细胞通常是细菌,优选大肠杆菌。任一具有 λ DE3溶原性细菌的细胞,例如,Rosetta2 (DE3) pLys、C41 (DE3)、BL21 (DE3)、JM109 (DE3)、B834 (DE3)、TUNER、Origami和Origami B,均可以表达包含T7启动子的载体。

[0541] 系列

[0542] 本发明还提供一系列与多核苷酸连接(或结合)的两个或更多个解旋酶,其中,所述两个或更多个解旋酶中的至少一个是本发明的解旋酶。所述系列可以包括任何数量的解旋酶,例如,2,3,4,5,6,7,8,9,10或更多个解旋酶。任何数量的解旋酶可以是本发明的解旋酶。所有所述两个或更多个解旋酶优选是本发明的解旋酶。本发明的一个或多个解旋酶可以是上述任一种。

[0543] 两个或更多个解旋酶可以是相同的解旋酶或者可以是不同的解旋酶。例如,如果所述系列包含本发明的两个或更多个解旋酶,则其可以相同或可以不同。

[0544] 所述系列可以包含任何数量和任何组合的本发明解旋酶。一系列两个或更多个解旋酶优选包含至少两个本发明的解旋酶。所述系列可以包含两个或更多个解旋酶,每个解旋酶包含SEQ ID NO:8的变体,所述变体包含(或仅包含)(a) P89F; (b) V150I; (c) V150H; (d) P89F和F98W; (e) P89F和V150I; (f) P89F和V150H; (g) F98W和V150I; (h) F98W和V150H; (i) P89F, F98W和V150I或(j) P89F, F98W和V150H。

[0545] 所述系列可以包含两个或更多个解旋酶,每个解旋酶包含SEQ ID NO:8的变体,所述变体包含(i) E94C/A360C; (ii) E94C/A360C,然后(Δ M1) G1G2(即,缺失M1,然后加入G1和G2); (iii) E94C/A360C/C109A/C136A; (iv) E94C/A360C/C109A/C136A,然后(Δ M1) G1G2(即,缺失M1,然后加入G1和G2); (v) E94C/A360C/W378A; (vi) E94C/A360C/W378A,然后(Δ M1) G1G2(即,缺失M1,然后加入G1和G2); (vii) E94C/A360C/C109A/C136A/W378A或(viii) E94C/A360C/C109A/C136A/W378A,然后(Δ M1) G1G2(即,缺失M1,然后加入G1和G2)。所述系列中的本发明一种解旋酶优选包含包含了(i)至(iv)中一个的SEQ ID NO:8的变体,所述系列中的本发明另一种解旋酶优选包含含有(v)至(viii)中一个的SEQ ID NO:8的变体。

[0546] 除了一个或多个本发明的解旋酶之外,所述系列可以包括不是本发明一部分的一个或多个解旋酶。一个或多个解旋酶可以是或衍生自Hel308解旋酶、RecD解旋酶(例如, TraI解旋酶或TrwC解旋酶)、XPD解旋酶或Dda解旋酶(例如,SEQ ID NO:8至23中的任一个)。一个或多个解旋酶可以是申请号为PCT/GB2012/052579(公开为W02013/057495); PCT/GB2012/053274(公开为W02013/098562); PCT/GB2012/053273(公开为W02013/098561);

PCT/GB2013/051925 (公开为W02014/013260); PCT/GB2013/051924 (公开为W0 2014/013259) 和PCT/GB2013/051928 (公开为W02014/013262); 以及PCT/GB2014/052736 (W02015/055981) 的国际申请中公开的解旋酶、修饰的解旋酶或解旋酶构建体中的任一个。具体地, 优选对一个或多个解旋酶进行修饰以减小多核苷酸结合域中的开口大小, 多核苷酸通过该开口在至少一个构象状态下可以从解旋酶中解链。这在W02014/013260中进行了公开。

[0547] 所述系列中的两个或更多个解旋酶可以彼此分开。所述系列中的两个或更多个解旋酶可以在多核苷酸移动穿过孔时通过跨膜孔集合在一起。所述系列中的两个或更多个解旋酶可以彼此接触。

[0548] 两个或更多个解旋酶优选彼此不连接, 除过经由多核苷酸连接在一起。两个或更多个解旋酶优选彼此不共价连接。

[0549] 两个或更多个解旋酶可以彼此连接或共价连接。解旋酶可以以任何顺序并使用任何方法进行连接。一系列连接的解旋酶可以称为一连串。

[0550] 下文将对本发明所述系列可以连接/结合的多核苷酸进行更详细的讨论。

[0551] 本发明的方法

[0552] 本发明提供一种控制目标多核苷酸移动的方法。所述方法包括使目标多核苷酸与本发明的修饰的解旋酶或本发明的构建体接触, 从而控制多核苷酸移动。所述方法优选跨孔施加电势的情况下进行。如下进行更详细的讨论, 施加的电势通常导致孔和解旋酶或构建体之间形成复合物。所施加的电势可以是电压电势。可选地, 施加的电势可以是化学电势。这方面的示例是在两性分子层上使用盐梯度。盐梯度在Holden等人2007年7月11日《美国化学会杂志》(J. Am. Chem. Soc.) 129 (27): 8650-5中进行了公开。

[0553] 本发明还提供一种表征目标多核苷酸的方法。所述方法包括 (a) 使目标多核苷酸与本发明的跨膜孔和修饰的解旋酶或本发明的构建体接触, 使得解旋酶或构建体控制目标多核苷酸穿过孔的移动。所述方法还包括 (b) 随着所述多核苷酸相对于所述孔移动, 获取一个或多个测量值, 其中所述测量值代表所述目标多核苷酸的一个或多个特征, 并由此表征所述目标多核苷酸。

[0554] 在本发明的所有方法中, 解旋酶或构建体可以是上述任一种。

[0555] 在这些方法中可以使用任何数量的本发明解旋酶。例如, 可以使用1, 2, 3, 4, 5, 6, 7, 8, 9, 10或更多个解旋酶。如果使用了两个或更多个本发明的解旋酶, 则其可以相同或不同。对于本发明所述系列, 以上讨论了合适的数量和组合。这些同样适用于本发明的方法。

[0556] 如果使用了两个或更多个解旋酶, 则其可以彼此连接。两个或更多个解旋酶可以彼此共价连接。解旋酶可以以任何顺序并使用任何方法进行连接。本发明中使用的优选解旋酶构建体在申请号为PCT/GB2013/051925 (公开为W02014/013260); PCT/GB2013/051924 (公开为W02014/013259) 和PCT/GB2013/051928 (公开为W02014/013262); 以及PCT/GB2014/052736 (W02015/055981) 的国际申请中进行了描述。

[0557] 如果使用了两个或更多个解旋酶, 其优选彼此不连接, 除了经由多核苷酸连接之外。两个或更多个解旋酶更优选地彼此不共价连接。

[0558] 步骤 (a) 和 (b) 优选在如上所述跨孔施加电势的情况下进行。在一些实例中, 随着多核苷酸相对于孔移动, 流经所述孔的电流被用于确定目标多核苷酸的序列。这就是链测序。

[0559] 本发明的方法用于表征目标多核苷酸。以上对多核苷酸进行了限定。

[0560] 可以使用该方法表征全部或仅一部分的目标多核苷酸。目标多核苷酸可以是任何长度。例如，多核苷酸长度可以是至少10个、至少50个、至少100个、至少150个、至少200个、至少250个、至少300个、至少400个或至少500个核苷酸对。多核苷酸长度可以是1000个或更多个核苷酸对、5000个或更多个核苷酸对或100000个或更多个核苷酸对。

[0561] 目标多核苷酸存在于任何合适的样品中。本发明通常在已知包含或怀疑包含目标多核苷酸的样品上进行。可选地，可对样品实施本发明，以确认在样品中的存在是已知的或期望的一个或多个目标多核苷酸的同源性。

[0562] 样品可以是生物样品。本发明可以针对从任何生物体或微生物中获得或提取的样品在体外实施。样品可以是非生物样品。非生物样品优选为流体样品。非生物样品的示例包括手术液、水如饮用水、海水或河水，以及实验室试验用的试剂。

[0563] 跨膜孔是在一定程度上穿过膜的结构。它允许通过施加的电势驱动水合离子使其跨膜流动或在膜内流动。跨膜孔通常穿过整个膜，使得水合离子可以从膜的一侧流到膜的另一侧。然而，跨膜孔不是必须跨越膜。其可以在一端封闭。例如，孔可以是膜中的凹陷，水合离子可以沿着膜中的凹陷流动或者流入其中。

[0564] 本发明中可使用任一跨膜孔。孔可以是生物的或人造的。合适的孔包括但不限于蛋白质孔、多核苷酸孔和固态孔。

[0565] 根据本发明，可以使用任何膜。合适的膜是本领域公知的。膜优选为两性分子层。两性分子层是由两亲分子，诸如磷脂形成的层，其具有至少一个亲水部分和至少一个亲脂或疏水部分。两亲分子可以是合成的或天然存在的。非天然存在的两亲物和形成单层的两亲物是本领域已知的，包括：例如，嵌段共聚物 (Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450)。嵌段共聚物是聚合物材料，其中，两个或更多个单体亚单元聚合在一起形成单一聚合物链。嵌段共聚物具有的性质通常由每个单体亚单元贡献。然而，嵌段共聚物可以具有从各个亚单元形成的聚合物所不具有的独特性质。可以对嵌段共聚物进行改造，使得其中一个单体亚单元是疏水性的（即亲脂性的），而其它亚单元在水介质中是亲水性的。在这种情况下，嵌段共聚物可以具有两亲性质，并且可以形成模仿生物膜的结构。嵌段共聚物可以是二嵌段（由两个单体亚单元组成），但也可以由两个以上的单体亚单元构成，以形成相当于两亲物质的更复杂的结构。共聚物可以是三嵌段、四嵌段或五嵌段共聚物。

[0566] 两性分子层可以是单层或双层。两性分子层通常是平面脂质双分子层或支撑双层。

[0567] 两性分子层通常是脂质双分子层。脂质双分子层是细胞膜的模型，并作为一系列实验研究的优秀平台。例如，脂质双分子层可通过单通道记录器用于膜蛋白的体外研究。可选地，脂质双分子层可用作生物传感器用以检测一系列物质的存在。脂质双分子层可以是任何脂质双分子层。合适的脂质双分子层包括但不限于平面脂质双分子层、支撑双层或脂质体。脂质双分子层优选为平面脂质双分子层。合适的脂质双分子层在申请号为PCT/GB08/000563（公开为W02008/102121）的国际申请、申请号为PCT/GB08/004127（公开为W02009/077734）和申请号为PCT/GB2006/001057（公开为W02006/100484）的国际申请中进行了公开。

[0568] 用于形成脂质双分子层的方法是本领域已知的。在实施例中公开了合适的方法。

脂质双分子层通常由Montal和Mueller的方法形成(Proc.Natl.Acad.Sci.USA.,1972;69:3561-3566),其中的脂质单分子层在穿过孔的任一侧负载在水溶液/空气界面上,所述孔垂直于所述界面。

[0569] Montal和Mueller的方法受到欢迎,因为其是形成适用于蛋白孔插入的优质脂质双分子层的划算且较直接的方法。形成双分子层的其它常见方法包括尖端浸渍、涂覆双层和脂质体双分子层的膜片钳(patch-clamping)。

[0570] 在优选实施方案中,形成脂质双分子层如国际申请No.PCT/GB08/004127(公开为WO 2009/077734)中所描述。

[0571] 在另一个优选实施方案中,膜是固态层。固态层不属于生物来源。换言之,固态层不是由生物环境获得或分离出的,所述生物环境例如生物体或细胞,或生物学上可获得结构的合成制造形式。固态层可以由有机材料和无机材料形成,包括但不限于,微电子材料,绝缘材料,例如Si₃N₄,Al₂O₃和SiO₂,有机和无机聚合物如聚酰胺,塑料如特氟隆(Teflon®)或柔性体如双组分加成固化硅橡胶,以及玻璃。固态层可由单原子层例如石墨烯,或只有几个原子厚度的层形成。合适的石墨层在国际申请No.PCT/US2008/010637(公开为WO 2009/035647)中公开。

[0572] 所述方法通常使用(i)包含孔的人造两性分子层;(ii)分离的,天然存在的包含孔的脂质双分子层,或(iii)含有孔嵌入其中的细胞。所述方法通常使用人造两性分子层,例如人造脂质双分子层实施。除了孔,该分子层可包含其他跨膜和/或膜内蛋白以及其他分子。合适的设备和条件在下面讨论。本发明的方法通常在体外实施。

[0573] 多核苷酸可以与膜连接(coupled to)。优选使用一个或多个锚定物将多核苷酸与膜连接。可以使用任何已知的方法将多核苷酸与膜连接。

[0574] 每个锚定物包含与多核苷酸连接(或结合)的基团和与膜连接(或结合)的基团。每个锚定物可以与多核苷酸和/或膜共价连接(或结合)。

[0575] 可以使用任何数量的锚定物,例如,2,3,4或更多个锚定物,将每个多核苷酸与膜连接。例如,可以使用两个锚定物将一个多核苷酸与膜连接,每个锚定物分别与多核苷酸和膜连接(或结合)。

[0576] 一个或多个锚定物可以包括一个或多个解旋酶和/或一个或多个分子制动器。

[0577] 如果膜是两性分子层,例如,共聚物膜或脂质双分子层,则一个或多个锚定物优选包含存在于膜中的多肽锚定物和/或存在于膜中的疏水锚定物。疏水锚定物优选为脂质、脂肪酸、固醇、碳纳米管、多肽、蛋白质或氨基酸,例如,胆固醇、棕榈酸酯或生育酚。在优选实施方案中,一个或多个锚定物不是检测器。

[0578] 可以对膜的组分,例如两亲分子、共聚物或脂质,进行化学修饰或官能化以形成一个或多个锚定物。膜组分的合适的化学修饰和合适的官能化方式的示例将在以下进行更详细的讨论。可以对任何比例的膜组分进行官能化,例如,至少0.01%、至少0.1%、至少1%、至少10%、至少25%、至少50%或100%。

[0579] 多核苷酸可以直接与膜连接。用于将多核苷酸与膜连接的一个或多个锚定物优选包含连接器。一个或多个锚定物可以包括一个或多个,例如2,3,4或更多的,连接器。可以使用一个连接器将多于一个,例如,2,3,4或更多个,多核苷酸与膜连接。

[0580] 优选的连接器包括但不限于聚合物,例如,多核苷酸,聚乙二醇(PEG),多糖和多

肽。这些连接器可以是直链、支链或环状的。例如,连接器可以是环状多核苷酸。多核苷酸可以与环状多核苷酸连接器上的互补序列杂交。

[0581] 一个或多个锚定物或一个或多个连接器可以包括能够被切割以断裂(broken down)的组分,例如,限制位点或光敏基团。

[0582] 官能化连接器及其可以连接分子的方式在本领域中是已知的。例如,经马来酰亚胺基官能化的连接器会与蛋白质中的半胱氨酸残基反应并与半胱氨酸残基连接。在本发明的上下文中,蛋白质可以存在于膜中或可以用于与多核苷酸连接(或结合)。这将在以下进行更详细的讨论。

[0583] 可以使用“钥匙和锁”结构来避免多核苷酸交联。每个连接器的仅一端可以一起反应以形成更长的连接器,连接器的另一端各自分别与多核苷酸或膜反应。这种连接器在申请号为PCT/GB10/000132(公开为W02010/086602)的国际申请中进行了描述。

[0584] 在下述测序实施方案中优选使用连接器。如果多核苷酸直接与膜永久性连接,这意味着当其与孔相互作用时不会脱离连接,由于膜和检测器之间的距离,测序运行不能继续到多核苷酸末端,所以一些序列数据将丢失。如果使用连接器,则可以完全处理多核苷酸。

[0585] 连接可以是永久的或稳定的。换言之,连接可以是这样,即,当多核苷酸与孔相互作用时保持与膜连接。

[0586] 连接可以是瞬时的。换言之,连接可以是这样,即,当多核苷酸与孔相互作用时可以与膜分离开来。

[0587] 对于某些应用,例如适体检测,优选连接的瞬态性质。如果永久性或稳定性连接器直接与多核苷酸的5'或3'末端连接,并且连接器比膜和跨膜孔通道之间的距离短,则由于测序运行不能继续到多核苷酸的末端,一些序列数据将丢失。如果连接是瞬时的,则当连接端随机脱离膜时,可以完全处理多核苷酸。以下对形成永久性/稳定性或瞬时连接的化学基团进行更详细的讨论。可以使用胆固醇或脂肪酰基链,将多核苷酸与两性分子层或三嵌段共聚物膜瞬时连接。可以使用长度为6至30个碳原子的任一脂肪酰基链,例如,十六烷酸。

[0588] 在优选实施方案中,多核苷酸,例如核酸,与两性分子层,例如三嵌段共聚物膜或脂质双分子层,连接。使用各种不同的系链策略将核酸与合成的脂质双分子层连接。这些在以下表6中进行了总结。

[0589] 表6

[0590]

锚定物 包含	连接 类型	参考文献
硫醇	稳定	Yoshina-Ishii, C. and S. G. Boxer (2003) . "Arrays of mobile tethered vesicles on supported lipid bilayers." <u>J Am Chem Soc</u> 125 (13) :3696-7
生物素	稳定	Nikolov, V., R. Lipowsky, et al. (2007) . "Behavior of giant vesicles with anchored DNA molecules." <u>Biophys J</u> 92 (12) :4356-68
胆固醇	瞬时	Pfeiffer, I. and F. Hook (2004) . "Bivalent cholesterol-based coupling of oligonucleotides to lipid membrane assemblies." <u>J Am Chem Soc</u> 126 (33) :10224-5
表面活性剂 (例如, 脂 质、棕榈酸 酯等)	稳定	van Lengerich, B., R. J. Rawle, et al. "Covalent attachment of lipid vesicles to a fluid-supported bilayer allows observation of DNA-mediated vesicle interactions." <u>Langmuir</u> 26 (11) :8666-72

[0591] 可以使用合成反应中经修饰的亚磷酰胺对合成的多核苷酸和/或连接器进行官能化,这与直接加入合适的锚定基团,如胆固醇、生育酚、棕榈酸酯、硫醇、脂质和生物素基团,容易相容。这些不同的连接化学物质提供了一系列与多核苷酸连接的选择。每个不同的修饰基团以稍微不同的方式与多核苷酸连接,并且连接并不总是永久性的,所以为多核苷酸与膜连接提供了不同的停留时间(dwelling times)。瞬时连接的优点如上所述。

[0592] 如果可以向多核苷酸添加互补的反应基团或锚定基团,则也可以通过许多其它方式来实现多核苷酸与连接器或官能化膜的连接。向多核苷酸的任一末端添加反应基团已被报道。可以使用T4多核苷酸激酶和ATP γ S向ssDNA或dsDNA的5'添加硫醇基团(Grant, G.P. and P.Z. Qin(2007). "A facile method for attaching nitroxide spin labels at the 5' terminus of nucleic acids." Nucleic Acids Res 35(10):e77)。可以使用T4多核苷酸激酶和 γ -[2-叠氮基乙基]-ATP或 γ -[6-叠氮基己基]-ATP向ssDNA或dsDNA的5'-磷酸基团添加叠氮基团。使用硫醇或点击化学,可以将系链(tether)与多核苷酸共价连接,其中,所述系链包含硫醇、碘乙酰胺OPSS或马来酰亚胺基团(对硫醇具有反应性)或DIBO(二苯并环辛炔dibenzocyclooctyne)或炔基(对叠氮化物具有反应性)。可以采用末端转移酶使化学基团(例如,生物素、硫醇和荧光团)的选择更多样化,从而将修饰的寡核苷酸引入ssDNA的3' (Kumar, A., P. Tchen, et al. (1988). "Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyl transferase." Anal Biochem 169(2):376-82)。链霉亲和素/生物素和/或链霉亲和素/脱硫生物素连接可用于任何其它多核苷酸。以下实施例描述了如何使用链霉亲和素/生物素和链霉亲和素/脱硫生物素将多核苷酸与膜连接。可以使用核苷酸经过适当修饰的末端转移酶(例如,胆固醇或棕榈酸酯)来将锚定物直接添加到多核苷酸中,这也是有可能的。

[0593] 一个或多个锚定物优选通过杂交将多核苷酸与膜连接。一个或多个锚定物中的杂交使得连接能够以所述瞬时方式进行。杂交可以存在于一个或多个锚定物的任何部分中,

例如,在一个或多个锚定物与多核苷酸之间、一个或多个锚定物内或者一个或多个锚定物与膜之间。例如,连接器可以包含两个或更多个杂交在一起的多核苷酸,例如,3,4或5个多核苷酸。一个或多个锚定物可以与多核苷酸杂交。一个或多个锚定物可以直接与多核苷酸杂交,或直接与连接到多核苷酸的Y适配器和/或前导序列杂交,或直接与连接到多核苷酸的桥连部分适配器,例如发夹环适配器,杂交(如下所述)。可选地,一个或多个锚定物可以与一个或多个,例如,2或3个,杂交到多核苷酸的中间多核苷酸(或“夹板(splint)”)杂交、与连接到多核苷酸的Y适配器和/或前导序列杂交,或与连接到多核苷酸的桥连部分适配器杂交(如下所述)。

[0594] 一个或多个锚定物可以包括单链或双链多核苷酸。锚定物的一部分可以与单链或双链多核苷酸连接。采用T4RNA连接酶I对ssDNA的短片段的连接已被报道(Troutt, A.B., M.G. McHeyzer-Williams等人(1992); “Ligation-anchored PCR: a simple amplification technique with single-sided specificity.” *Proc Natl Acad Sci U S A* 89(20): 9823-5)。可选地,单链或双链多核苷酸可以与双链多核苷酸连接,然后通过热或化学变性分离这两条链。对于双链多核苷酸,可以将一段单链多核苷酸添加到所述双链体的一端或两端,或将双链多核苷酸添加到所述双链体的一端或两端。对于将单链多核苷酸添加到双链多核苷酸,这可以同连接到单链多核苷酸其它区域一样,使用T4RNA连接酶I来实现。对于将双链多核苷酸添加到双链多核苷酸,此时连接可以用分别在多核苷酸和添加的多核苷酸上的互补的3'dA/dT尾部(这是常规针对许多样品制备应用进行的,以防止串多联体(concatemer)或者二聚体形成)或使用由多核苷酸的限制消化以及相容适配器的连接而产生的“粘性末端”,进行“平端终止(blunt-ended)”。然后,当双链体熔化时,如果使用单链多核苷酸进行连接,则每个单链将具有5'或3'修饰,或者如果使用双链多核苷酸进行连接,则在5'末端、3'末端或这两个末端上发生修饰。

[0595] 如果多核苷酸是合成链,则可以在化学合成多核苷酸期间引入一个或多个锚定物。例如,可以采用连接有反应基团的引物对多核苷酸进行合成。

[0596] 腺苷酸化多核苷酸是连接反应中的中间体,其中,单磷酸腺苷与多核苷酸的5'-磷酸基团连接。可以使用各种试剂盒来生成该中间体,例如,来自NEB的5' DNA腺苷酸化试剂盒。通过在反应中用ATP取代修饰的核苷酸三磷酸,然后就可以向多核苷酸的5'添加反应基团(如硫醇、胺、生物素、叠氮化物等)。使用具有适当修饰的核苷酸的5' DNA腺苷酸化试剂盒(例如,胆固醇或棕榈酸酯)也可以将锚定物直接添加到多核苷酸中。

[0597] 对基因组DNA片段进行扩增的常见技术是使用聚合酶链反应(PCR)。此处,采用两个合成的寡核苷酸引物,就可以生成同一DNA片段的多个副本,其中,对于每个副本,双链体中每个链的5'将是合成的多核苷酸。可以通过使用聚合酶向单链或双链DNA的3'末端添加单个或多个核苷酸。可以使用的聚合酶的示例包括但不限于末端转移酶、克列诺酶(Klenow)和大肠杆菌多Poly(A)聚合酶。通过在反应中用ATP取代修饰的核苷酸三磷酸,然后就可以将诸如胆固醇、硫醇、胺、叠氮化物、生物素或脂质等锚定物引入双链多核苷酸。因此,扩增的多核苷酸的每个副本将包含锚定物。

[0598] 理想情况下,多核苷酸与膜连接而不必使多核苷酸官能化。这可以通过将一个或多个锚定物(例如,多核苷酸结合蛋白或化学基团)与膜连接并使一个或多个锚定物与多核苷酸相互作用或通过使膜官能化来得以实现。一个或多个锚定物可以通过本文所述的任何

方法与膜连接。具体地,一个或多个锚定物可以包含一个或多个连接器,例如,马来酰亚胺官能化连接器。

[0599] 在本实施方案中,多核苷酸通常是RNA、DNA、PNA、TNA或LNA,并且可以是双链或单链。本实施方案特别适用于基因组DNA多核苷酸。

[0600] 一个或多个锚定物可以包含与单链或双链多核苷酸、多核苷酸内的特异性核苷酸序列或多核苷酸内的修饰核苷酸的模式或存在于多核苷酸上的任何其它配体连接、结合或相互作用的任何基团。

[0601] 适用于锚定物的结合蛋白包括但不限于:大肠杆菌单链结合蛋白、P5单链结合蛋白、T4 gp32单链结合蛋白、TOPO V dsDNA结合区域、人组蛋白、大肠杆菌HU DNA结合蛋白和其它古细菌、原核或真核单链或双链多核苷酸(或核酸)结合蛋白,包括以下所列那些。

[0602] 特异性核苷酸序列可以是转录因子、核糖体、内切核酸酶、拓扑异构酶或复制起始因子所鉴定的序列。修饰核苷酸的模式可以是甲基化或损伤模式。

[0603] 一个或多个锚定物可以包含与多核苷酸连接、结合、嵌入或相互作用的任何基团。所述基团可以通过静电、氢键或范德华相互作用嵌入多核苷酸或与多核苷酸相互作用。这种基团包括赖氨酸单体、聚赖氨酸(其将与ssDNA或dsDNA相互作用)、溴化乙锭(其将嵌入dsDNA)、通用碱基或通用核苷酸(其可以与任何多核苷酸杂交)和钨复合物(其可以与甲基化碱基反应)。因此,可以使用一个或多个连接到膜的通用核苷酸将多核苷酸与膜连接。每个通用核苷酸可以采用一个或多个连接器与膜连接。通用核苷酸优选包含以下核碱基之一:次黄嘌呤、4-硝基吡啶、5-硝基吡啶、6-硝基吡啶、甲酰基吡啶、3-硝基吡咯、硝基咪唑、4-硝基吡唑、4-硝基苯并咪唑、5-硝基吡唑、4-氨基苯并咪唑或苯基(C6-芳族环)。通用核苷酸更优选包含以下核苷之一:2'-脱氧肌苷、肌苷、7-脱氮-2'-脱氧肌苷、7-脱氮肌苷、2-氮杂-脱氧肌苷、2-氮杂-肌苷、2-O'-甲基肌苷、4-硝基吡啶2'-脱氧核糖核苷、4-硝基吡啶核糖核苷、5-硝基吡啶2'-脱氧核糖核苷、5-硝基吡啶核糖核苷、6-硝基吡啶2'-脱氧核糖核苷、6-硝基吡啶核糖核苷、3-硝基吡咯2'-脱氧核糖核苷、3-硝基吡咯核糖核苷、次黄嘌呤的无环糖类似物、硝基咪唑2'-脱氧核糖核苷、硝基咪唑核糖核苷、4-硝基吡唑2'-脱氧核糖核苷、4-硝基吡唑核糖核苷、4-硝基苯并咪唑2'-脱氧核糖核苷、4-硝基苯并咪唑核糖核苷、5-硝基吡唑2'-脱氧核糖核苷、5-硝基吡唑核糖核苷、4-氨基苯并咪唑2'-脱氧核糖核苷、4-氨基苯并咪唑核糖核苷、苯基C-核糖核苷、苯基C-2'-脱氧核糖核苷、2'-脱氧神经氨酸、2'-脱氧肌氨酸、K-2'-脱氧核糖、P-2'-脱氧核糖和吡咯烷。通用核苷酸更优选包含2'-脱氧肌苷。通用核苷酸更优选为IMP或dIMP。通用核苷酸最优选为dPMP(2'-脱氧-P-核苷单磷酸)或dKMP(N6-甲氧基-2,6-二氨基嘌呤单磷酸)。

[0604] 一个或多个锚定物可以通过胡斯坦(Hoogsteen)氢键(其中,两个核碱基通过氢键保持在一起)或反向胡斯坦氢键(其中,一个核碱基相对于另一个核碱基旋转180°)与多核苷酸连接(或结合)。例如,一个或多个锚定物可以包含一个或多个核苷酸、一个或多个寡核苷酸或一个或多个与多核苷酸形成胡斯坦氢键或反向胡斯坦氢键的多核苷酸。这些类型的氢键允许第三个多核苷酸链缠绕在双链螺旋周围并形成三链体。一个或多个锚定物可以通过与双链双链体形成三链体而与双链多核苷酸连接(或结合)。

[0605] 在本实施方案中,至少1%、至少10%、至少25%、至少50%或100%的膜组分可以进行官能化。

[0606] 当一个或多个锚定物包含蛋白时,它们能够无需进行进一步官能化而直接锚定到膜中,例如,如果它们已经具有与膜相容的外部疏水区域。这种蛋白的示例包括但不限于跨膜蛋白、膜内蛋白和膜蛋白。可选地,蛋白可以用与膜相容的基因融合疏水区域表达。这种疏水性蛋白质区域是本领域已知的。

[0607] 一个或多个锚定物优选在与膜接触之前与多核苷酸混合,但是一个或多个锚定物可以与膜接触并随后与多核苷酸接触。

[0608] 另一方面,可以使用上述方法对多核苷酸进行官能化,使得其可以被特异性结合基团识别。具体地,可以采用诸如生物素(用于与链霉亲和素结合)、直链淀粉(用于与麦芽糖结合蛋白或融合蛋白结合)、Ni-NTA(用于与多组氨酸或多组氨酸标记蛋白结合)或肽(例如,抗原)等配体对多核苷酸进行官能化。

[0609] 根据优选实施方案,当多核苷酸与旋入孔的前导序列连接时,可使用一个或多个锚定物来将多核苷酸与膜连接。以下对前导序列进行更详细的讨论。优选地,将多核苷酸与优先旋入孔的前导序列连接(例如,联接)。这种前导序列可以包含均聚多核苷酸或无碱区域。通常将前导序列设计成直接或经由一个或多个中间多核苷酸(或夹板)与一个或多个锚定物杂交。在这种情况下,一个或多个锚定物通常包含与前导序列中的序列或一个或多个中间多核苷酸(或夹板)中的序列互补的多核苷酸序列。在这种情况下,一个或多个夹板通常包含与前导序列中的序列互补的多核苷酸序列。

[0610] 化学连接中所用的分子的示例是EDC(1-乙基-3-[3-二甲基氨基丙基]碳二亚胺盐酸盐)。也可以采用市售的试剂盒(Thermo Pierce,料号:22980)向多核苷酸的5'添加反应基团。合适的方法包括但不限于使用组氨酸残基和Ni-NTA的瞬时亲和连接以及通过反应性半胱氨酸、赖氨酸或非天然氨基酸进行更强健稳定的共价连接。

[0611] 跨膜孔优选为跨膜蛋白孔。跨膜蛋白孔是允许水合离子(例如,分析物)从膜的一侧流向膜的另一侧的多肽或多肽集合。在本发明中,跨膜蛋白孔能够形成这样的孔,其允许由施加电势所驱动的水合离子从膜的一侧流向另一侧。跨膜蛋白孔优选允许诸如核苷酸等分析物从膜的一侧(例如,脂质双分子层)流向另一侧。跨膜蛋白孔允许诸如DNA或RNA等多核苷酸移动穿过孔。

[0612] 跨膜蛋白孔可以是单体或寡聚物。所述孔优选由多个重复的亚基,例如至少6个、至少7个、至少8个或至少9个亚基组成。所述孔优选由6,7,8或9个亚基组成。所述孔优选为六聚、六聚体,七聚体,八聚体或九聚体的孔。所述孔可以是同源寡聚物(homo-oligomer)或异源寡聚物(hetero-oligomer)。

[0613] 跨膜蛋白孔通常包含离子可以流过的桶或通道。所述孔的亚基通常围绕中心轴并且向跨膜桶或通道或跨膜 α -螺旋束或通道提供链(strands)。

[0614] 跨膜蛋白孔的桶或通道通常包含有助于与分析物,例如核苷酸、多核苷酸或核酸,相互作用的氨基酸。这些氨基酸优选位于桶或通道的收缩(constriction)附近。跨膜蛋白孔通常包含一个或多个带正电荷的氨基酸,例如,精氨酸、赖氨酸或组氨酸或芳族氨基酸如酪氨酸或色氨酸。这些氨基酸通常有助于孔和核苷酸、多核苷酸或核酸之间的相互作用。

[0615] 本发明中所使用的跨膜蛋白孔可以衍生自 β -桶孔或 α -螺旋束孔。 β -桶孔包含由 β -链形成的桶或通道。合适的 β -桶孔包括但不限于 β -毒素,例如 α -溶血素、炭疽毒素和白细胞介素,以及细菌的外膜蛋白/孔蛋白,例如耻垢分枝杆菌孔蛋白(Msp),比如:MspA、MspB、

MspC或MspD、CsgG、外膜孔蛋白F (OmpF)、外膜孔蛋白G (OmpG)、外膜磷脂酶A和奈瑟氏菌自转运脂蛋白 (NalP), 和其它孔如lysenin。 α -螺旋束孔包括由 α -螺旋形成的桶或通道。合适的 α -螺旋束孔包括但不限于内膜蛋白和 α 外膜蛋白, 例如, WZA和ClyA毒素。跨膜孔可以衍生自lysenin。申请号为PCT/GB2013/050667 (公开为W02013/153359) 的国际申请中对衍生自lysenin的合适孔进行了公开。申请号为PCT/EP2015/069965的国际申请中对衍生自CsgG的合适孔进行了公开。跨膜孔可以衍生自Msp或 α -溶血素 (α -HL)。

[0616] 跨膜蛋白孔优选衍生自Msp, 优选衍生自MspA。这种孔是寡聚的, 并且通常包含衍生自Msp的7, 8, 9或10个单体。孔可以是衍生自含相同单体的Msp的同源寡聚孔。可选地, 所述孔可以是衍生自Msp的异聚寡聚孔, 所述Msp包含至少一个不同于其它单体的单体。优选地, 孔衍生自MspA或其同源物或并系同源物 (paralog)。

[0617] 衍生自Msp的单体通常包含SEQ ID NO:2所示的序列或其变体。SEQ ID NO:2是野生型MspA单体。SEQ ID NO:2的变体是具有由SEQ ID NO:2的氨基酸序列变化而来并且其具有形成孔的能力的氨基酸序列的多肽。可以使用本领域已知的任何方法来对变体形成孔的能力进行测定。例如, 可以将变体与其它合适的亚基一起插入两性分子层, 并且可以对其寡聚化形成孔的能力进行确定。本领域已知用于将亚基插入诸如两性分子层等膜中的方法。例如, 亚基可以以纯化形式悬浮在包含脂质双分子层的溶液中, 使得其扩散到所述脂质双分子层, 并通过结合到脂质双分子层而插入, 并组装成功能状态。可选地, 可以采用“拾取和放置”方法将亚基直接插入膜中, 所述方法在M.A.Holden和H.Bayley《美国化学会杂志》2005, 127, 6502-6503以及申请号为PCT/GB2006/001057 (公开为W02006/100484) 的国际申请中进行了描述。

[0618] 对于整个长度的SEQ ID NO:2的氨基酸序列, 基于氨基酸同一性, 变体优选与该序列至少50%同源。更优选地, 基于氨基酸同一性, 变体与整个序列的SEQ ID NO:2的氨基酸序列可以至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%且更优选至少95%、97%或99%同源。在100或更多, 例如125, 150, 175或200或更多个的连续氨基酸, 可以具有至少80%, 例如, 至少85%、90%或95%的氨基酸同一性 (严格同源性)。

[0619] 可使用本领域的标准方法来确定同源性。例如, UWGCG软件包提供了可用于计算同源性的BESTFIT程序, 例如在其默认设置使用 (Devereux et al (1984) *Nucleic Acids Research* 12, p387-395)。可使用PILEUP、BLAST和PSIBLAST算法来计算同源性或比对序列 (例如, 鉴定等效残基或相应序列 (通常在它们的默认设置)), 例如, 如Altschul S.F. (1993) *J Mol Evol* 36:290-300; Altschul, S.F et al (1990) *J Mol Biol* 215:403-10中所描述的。进行BLAST分析的软件可通过美国国家生物技术信息中心 (<http://www.ncbi.nlm.nih.gov/>) 公开获得。

[0620] 变体可以包含以下突变: D90N, D91N, D93N, D118R, D134R和E139K。变体可以是申请号为PCT/GB2012/050301 (W02012/107778) 的国际申请中公开的变体中的任一种。

[0621] 变体优选 (a) 不包含位点90上的天冬氨酸 (D); (b) 不包含位点91上的天冬氨酸 (D); (c) 包含位点93上的天冬氨酸 (D) 或谷氨酸 (E); 和 (d) 包含一个或多个修饰, 其降低了在单体的帽形成区域和/或桶形成区域中的向内的氨基酸的净负电荷。(d) 中优选的突变包括但不限于D118R, Q126R, D134R和E139K。变体优选包含位点93上的D90N, D91N, D或E;

D118R, D134R和E139K。变体可以是申请号为PCT/GB2015/051290的国际申请中所公开变体中的任一种。

[0622] SEQ ID NO:2是MspA单体的MS-(B1)8突变体。相比于MspA,变体可以包含MspB,C或D单体中的任何突变。MspB,C和D的成熟形式在SEQ ID NO:5至7中进行示出。具体地,变体可以包含存在于MspB中的以下取代:A138P。变体可以包含存在于MspC中的以下取代中的一个或多个:A96G,N102E和A138P。变体可以包含存在于MspD中的以下突变中的一个或多个:G1缺失,L2V,E5Q,L8V,D13G,W21A,D22E,K47T,I49H,I68V,D91G,A96Q,N102D,S103T,V104I,S136K和G141A。变体可以包含MspB,C和D的突变体和取代物中的一个或多个的组合。变体优选包含突变L88N。SEQ ID NO:2的变体除了MS-(B1)8的全部突变之外还具有突变L88N,并被称为MS-(B2)8。本发明中使用的孔优选为MS-(B2)8。进一步的优选变体包含突变G75S/G77S/L88N/Q126R。SEQ ID NO:2的变体除了MS-(B1)8的全部突变之外还具有突变G75S/G77S/L88N/Q126R,并被称为MS-(B2C)8。本发明中使用的孔优选为MS-(B2)8或MS-(B2C)8。

[0623] 除了上述取代之外,还可以对SEQ ID NO:2的氨基酸序列进行氨基酸取代,例如,多达1,2,3,4,5,10,20或30个取代。保守取代用具有类似化学结构、类似化学性质或类似侧链体积的其它氨基酸取代氨基酸。引入的氨基酸相比其所取代的氨基酸可能具有相似的极性、亲水性、疏水性、碱度、酸度、中性或带电性。可选地,保守取代可以引入另一种芳族或脂肪族氨基酸来取代预先存在的芳族或脂肪族氨基酸。保守的氨基酸变化是本领域公知的,并且可以根据以下表7中所定义的20个主要氨基酸的性质来进行选择。当氨基酸具有相似的极性时,也可以结合表8中氨基酸侧链的亲水性水平来对此进行确定。

[0624] 表7氨基酸的化学性质

[0625]

Ala	脂肪族、疏水、中性	Met	疏水、中性
Cys	极性、疏水、中性	Asn	极性、亲水、中性
Asp	极性、亲水、带电荷 (-)	Pro	疏水、中性
Glu	极性、亲水、带电荷 (-)	Gln	极性、亲水、中性
Phe	芳族、疏水、中性	Arg	极性、亲水、带电荷 (+)
Gly	脂肪族、中性	Ser	极性、亲水、中性
His	芳族、极性、亲水、带电荷 (+)	Thr	极性、亲水、中性
Ile	脂肪族、疏水、中性	Val	脂肪族、疏水、中性

[0626]

Lys	极性、亲水、带电荷 (+)	Trp	芳族、疏水、中性
Leu	脂肪族、疏水、中性	Tyr	芳族、极性、疏水

[0627] 表8亲水性水平

	侧链	亲水性
[0628]	Ile	4.5
	Val	4.2
	Leu	3.8
	Phe	2.8
	Cys	2.5
	Met	1.9
	Ala	1.8
	Gly	-0.4
	Thr	-0.7
	Ser	-0.8
	Trp	-0.9
	Tyr	-1.3
	Pro	-1.6
	His	-3.2
	Glu	-3.5
	Gln	-3.5
	Asp	-3.5
	Asn	-3.5
	Lys	-3.9
	Arg	-4.5

[0629] SEQ ID NO:2的氨基酸序列的一个或多个氨基酸残基可另外从上述多肽中缺失。多达1,2,3,4,5,10,20或30个以上残基可以缺失。

[0630] 变体可以包括SEQ ID NO:2的片段。这种片段保留形成孔的活性。所述片段长度可以至少为50,100,150或200个氨基酸。这种片段可用于制备孔。片段优选包含SEQ ID NO:2的孔形成域。片段必须包含SEQ ID NO:2的残基88,90,91,105,118和134中的一个。通常,片段包含SEQ ID NO:2的残基88,90,91,105,118和134中的全部。

[0631] 可以将一个或多个氨基酸可选地或额外地加入到上述多肽中。可以在SEQ ID NO:2的氨基酸序列的氨基末端或羧基末端或多肽变体或其片段上提供延伸。所述延伸可能相当短,例如,长度为1至10个氨基酸。可选地,所述延伸可以更长,例如,多达50或100个氨基酸。根据本发明,可以将载体蛋白与氨基酸序列融合。以下将对其它融合蛋白进行更详细的讨论。

[0632] 如上所述,变体是具有由SEQ ID NO:2的氨基酸序列变化而来且保留其形成孔能力的氨基酸序列的多肽。变体通常包含SEQ ID NO:2中负责孔形成的区域。包含 β -桶的Msp的孔形成能力由每个亚基中的 β -片体(β -sheets)提供。SEQ ID NO:2的变体通常包含SEQ ID NO:2中形成 β -片体的区域。只要所得变体具有成孔能力,就可以对形成 β -片体的SEQ ID NO:2的区域进行一个或多个修饰。SEQ ID NO:2的变体优选包含其 α -螺旋和/或环状区域内的一个或多个修饰,例如,取代、增加或缺失。

[0633] 可以对衍生自Msp的单体进行修饰以帮助对它们的鉴定或纯化,例如,通过添加组氨酸残基(hist标签)、天冬氨酸残基(asp标签)、链霉亲和素标签或旗标,或通过添加用以促进其从细胞中分泌的信号序列,多肽在所述细胞中不天然含有这种序列。引入基因标签的替代方法是将标签化学反应到孔上的天然或改造位点上。这种情况的示例是将凝胶转移试剂与在孔外部设计的半胱氨酸反应。已经证明这是一种分离溶血素异源寡聚物的方法

(Chem Biol.1997Jul;4(7):497-505)。

[0634] 可以用显示标记物 (revealing label) 对衍生自Msp的单体进行标记。显示标记物可以是允许对孔进行检测的任何合适的标记物。以上对合适的标记物进行了描述。

[0635] 也可以使用D-氨基酸来制备衍生自Msp的单体。例如,衍生自Msp的单体可以包含L-氨基酸和D-氨基酸的混合物。这对制备这种蛋白质或肽而言是本领域常见的。

[0636] 衍生自Msp的单体包含一个或多个有助于鉴定核苷酸特异性的修饰。衍生自Msp的单体也可以包含其它非特异性修饰,只要这些修饰不影响孔的形成即可。本领域已知许多非特异性侧链修饰,并且可以针对衍生自CsgG的单体的侧链进行。这种修饰包括:例如,通过下述进行的氨基酸还原性烷基化:通过与醛反应,然后用NaBH₄还原,用亚氨酰乙酸甲酯脎化或用乙酸酐酰化。

[0637] 可以使用本领域已知的标准方法来制备衍生自Msp的单体。可以通过合成或重组方式制备衍生自Msp的单体。例如,可以通过体外转译和转录 (IVTT) 对孔进行合成。申请号为PCT/GB09/001690 (公开为W02010/004273)、PCT/GB09/001679 (公开为W02010/004265) 或PCT/GB10/000133 (公开为W02010/086603) 的国际申请中对合适的孔制备方法进行了讨论。将孔插入膜的方法如下所述。

[0638] 跨膜蛋白孔也优选衍生自 α -溶血素 (α -HL)。

[0639] 在一些实施方案中,对跨膜蛋白孔进行化学修饰。可以以任何方式并在任何位点上对孔进行化学修饰。优选通过将分子与一个或多个半胱氨酸 (半胱氨酸连接) 连接,将分子与一个或多个赖氨酸连接,将分子与一个或多个非天然氨基酸连接,对表位进行酶修饰或对末端进行修饰来对跨膜蛋白孔进行化学修饰。进行这些修饰的合适方法是本领域公知的。可以通过连接任何分子来对跨膜蛋白孔进行化学修饰。例如,可以通过连接染料或荧光团来对孔进行化学修饰。

[0640] 可以对孔中任何数量的单体进行化学修饰。优选对一个或多个单体,例如2,3,4,5,6,7,8,9或10个单体进行如上所述的化学修饰。

[0641] 可以通过对相邻残基进行修饰来增强半胱氨酸残基的反应性。例如,侧链精氨酸、组氨酸或赖氨酸残基的碱基会将半胱氨酸硫醇基团的pKa改变为更具反应性的S⁻基团的pKa。半胱氨酸残基的反应性可以被诸如dTNB等硫醇保护基团保护。这些保护基团可以在连接器连接之前与孔的一个或多个半胱氨酸残基反应。

[0642] 可以直接将分子 (其中,孔进行化学修饰) 与孔连接,或者通过连接器进行连接,正如申请号为PCT/GB09/001690 (公布为W02010/004273)、PCT/GB09/001679 (公开为W02010/004265) 或PCT/GB10/000133 (公开为W02010/086603) 的国际申请中所公开的。

[0643] 解旋酶或构建体可以与孔共价地连接。解旋酶或构建体优选不与孔共价地连接。向孔和解旋酶或构建体施加电压通常导致形成能够对目标多核苷酸进行测序的传感器。这将在以下进行更详细的讨论。

[0644] 可以对本文所述的任何蛋白质 (即,解旋酶、跨膜蛋白孔或构建体) 进行修饰以有助于对它们的鉴定或纯化,例如,通过添加组氨酸残基 (his标签)、天冬氨酸残基 (asp标签)、链霉亲和素标签、旗号标签 (flag tag)、SUMO标签、GST标签或MBP标签,或者通过添加促进它们从细胞分泌的信号序列,其中多肽不天然含有这类序列。引入遗传标签的替代方法是将标签化学反应到解旋酶、孔或构建体上的天然或改造位点上。这种情况的示例是将

凝胶转移试剂与在孔外部设计的半胱氨酸反应。已经证明这是一种分离溶血素异源寡聚物的方法 (Chem Biol.1997Jul;4 (7) 497-505)。

[0645] 可以用显示标记物标记解旋酶、孔或构建体。显示标记物可以是允许对孔进行检测的任何合适标记。合适的标记物包括但不限于荧光分子、放射性同位素,例如,¹²⁵I、³⁵S、酶、抗体、抗原、多核苷酸和诸如生物素等配体。

[0646] 可以通过合成或重组方式制备蛋白质。例如,可以通过体外转译和转录 (IVTT) 对解旋酶、孔或构建体进行合成。可以对解旋酶、孔或构建体的氨基酸序列进行修饰用以包括非天然存在的氨基酸或用以增加蛋白质的稳定性。当通过合成方式制备蛋白质时,可以在制备过程中引入这些氨基酸。也可以在合成或重组制备之后改变解旋酶、孔或构建体。

[0647] 也可以使用D-氨基酸来制备解旋酶、孔或构建体。例如,孔或构建体可以包含L-氨基酸和D-氨基酸的混合物。这在制备这种蛋白质或肽时是本领域常见的。

[0648] 解旋酶、孔或构建体也可以包含其它非特异性修饰,只要这些修饰不影响孔的形成或者解旋酶或构建体的功能即可。本领域已知许多非特异性侧链修饰,并可以对所述蛋白的侧链进行。这种修饰包括:例如通过下述进行的氨基酸还原性烷基化:与醛反应然后用NaBH₄还原,用亚氨酰乙酸甲酯进行脎基化或用乙酸酐进行酰化。

[0649] 可以使用本领域已知的标准方法来制备解旋酶、孔和构建体。编码了解旋酶、孔或构建体的多核苷酸序列可以采用本领域的标准方法进行衍生和复制。编码解旋酶、孔或构建体的多核苷酸序列可以使用本领域的标准技术来在细菌宿主细胞中表达。可以通过对来自重组表达载体的多肽进行原位表达来在细胞中制备解旋酶、孔和/或构建体。表达载体可选地携带诱导型启动子以控制多肽的表达。这些方法在Sambrook,J.and Russell,D.(2001).Molecular Cloning:A Laboratory Manual,3rd Edition.Cold Spring Harbor Laboratory Press,Cold Spring Harbor,NY中进行了描述。

[0650] 从产生蛋白质的生物体中或重组表达后经任意蛋白质液相色谱系统纯化后,大规模地制备解旋酶、孔和/或构建体。典型的蛋白质液相色谱系统包括FPLC、AKTA系统、Bio-Cad系统、Bio-Rad BioLogic系统和Gilson HPLC系统。

[0651] 本发明的方法涉及测量目标多核苷酸的一个或多个特征。所述方法可能涉及测量目标多核苷酸的两个、三个、四个或五个或更多个特征。所述一个或多个特征优选选自(i)目标多核苷酸的长度;(ii)目标多核苷酸的同源性;(iii)目标多核苷酸的序列;(iv)目标多核苷酸的二级结构和(v)目标多核苷酸是否经过修饰。根据本发明,可以对(i)至(v)的任一组合进行测量。

[0652] 对于(i),例如,可以通过确定目标多核苷酸和孔之间的相互作用次数或目标多核苷酸和孔之间的相互作用持续时间来测量多核苷酸的长度。

[0653] 对于(ii),可以以多种方式测量多核苷酸的同源性。可以结合目标多核苷酸的序列的测定,或不联合目标多核苷酸的序列的测定来测定多核苷酸的同源性。前者是直接的;对多核苷酸进行测序,并由此进行鉴定。后者可以通过几种方式完成。例如,可以测定多核苷酸中特定基序的存在(而无需测定多核苷酸的剩余序列)。可选地,所述方法中测定的特定的电和/或光信号的测量值可鉴定来自特定来源的目标多核苷酸。

[0654] 对于(iii),可以如先前所述对多核苷酸的序列进行确定。合适的测序方法,特别是那些使用电测量值的测序方法,描述于Stoddart D et al.,Proc Natl Acad Sci,12;

106 (19) :7702-7, Lieberman KR et al, J Am Chem Soc. 2010; 132 (50) :17961-72, 和国际申请WO 2000/28312中。

[0655] 对于(iv), 所述二级结构可以以多种方式测量。例如, 如果所述方法包含电测量, 二级结构可以利用穿过孔的停留时间的改变或电流变化进行测量。这使得能够区别出单链和双链多核苷酸的区域。

[0656] 对于(v), 可以测定任何存在或不存在修饰。该方法优选包括确定所述目标多核苷酸是否用一个或多个蛋白质或一个或多个标记物、标签或间隔物通过甲基化、氧化、损伤进行了修饰。特异性修饰将导致与孔的特异性相互作用, 其可以使用下述方法进行测定。例如, 可以基于孔与每个核苷酸的相互作用过程中穿过孔的电流, 区别胞嘧啶与甲基胞嘧啶。

[0657] 可以进行各种不同类型的测量。这包括但不限于: 电测量和光测量。可能的电测量包括: 电流测量、阻抗测量、隧道测量(Ivanov AP等人, Nano Lett. 2011 Jan 12; 11 (1) :279-85) 和FET测量(公开号为W02005/124888的国际申请)。光测量可以与电测量相结合(Soni GV等人, Rev Sci Instrum. 2010 Jan; 81 (1) :014301)。测量可以是跨膜电流测量, 例如, 流过孔的离子电流的测量。

[0658] 可以使用标准单通道记录设备进行电测量, Stoddart D等人, Proc Natl Acad Sci, 12; 106 (19) :7702-7, Lieberman KR等人, J Am Chem Soc. 2010; 132 (50) :17961-72, 和国际申请WO-2000/28312所描述的。可选地, 可以采用多通道系统进行电测量, 例如, 如公开号为W02009/077734和公开号为W02011/067559的国际申请中所描述的

[0659] 在优选实施方案中, 所述方法包括:

[0660] (a) 使目标多核苷酸与本发明的跨膜孔和解旋酶或本发明的构建体接触, 使得目标多核苷酸移动穿过孔, 并且解旋酶或构建体控制目标多核苷酸穿过孔的移动; 以及

[0661] (b) 随着多核苷酸相对于孔移动, 测量流经所述孔的电流, 其中, 所述电流代表目标多核苷酸的一个或多个特征, 并由此所述表征目标多核苷酸。

[0662] 可以采用适用于研究膜/孔系统的任何设备来实施所述方法, 在所述系统中孔存在于膜中。可以采用适用于感测跨膜孔的任何设备来实施所述方法。例如, 所述设备包括一个室, 所述室包括水溶液和将该室分割为两部分的屏障(barrier)。所述障碍物通常具有裂缝, 其中在缝隙中形成包括孔的膜。或者该屏障形成其中存在孔的膜。

[0663] 可以采用申请号为PCT/GB08/000562 (W02008/102120) 的国际申请中所描述的设备来执行所述方法。

[0664] 该方法可以包括随着多核苷酸相对于所述孔移动, 测量通过所述孔的电流。因此, 所述设备还可以包括能够跨膜和孔施加电势并测量电流信号的电路。可以使用膜片钳或电压钳来执行所述方法。所述方法优选包含使用电压钳。

[0665] 本发明的方法可包括随着所述多核苷酸相对于所述孔移动来测量流过所述孔的电流。用于测量通过跨膜蛋白孔的离子电流的合适的条件是本领域已知的, 并且在实施例中公开。该方法通过跨膜和孔施加的电压进行实施。所使用的电压通常为+2V至-2V, 通常为-400mV至+400mV。所使用的电压优选在具有下限和上限的范围内, 其中, 所述下限选自-400mV, -300mV, -200mV, -150mV, -100mV, -50mV, -20mV和0mV, 所述上限独立地选自+10mV, +20mV, +50mV, +100mV, +150mV, +200mV, +300mV和+400mV。所使用的电压范围更优选在100mV至240mV, 最优选在120mV至220mV。可通过对孔施加提高的电势来提高对不同核苷酸的分辨

力。

[0666] 该方法通常在任何载荷子,如金属盐,例如碱金属盐,卤盐,例如氯盐,如碱金属氯盐的存在下实施。载荷子可包括离子型液体或有机盐,例如四甲基氯化铵,三甲基氯化铵,氯化苯甲基铵,或1-乙基-3-甲基咪唑鎓氯化物。在上述示例性装置中,盐存在于所述室中的水溶液中。通常使用氯化钾(KCl)、氯化钠(NaCl)、氯化铯(CsCl)或亚铁氰化钾和铁氰化钾的混合物。优选KCl、NaCl和亚铁氰化钾和铁氰化钾的混合物。盐浓度可为饱和的。盐浓度可以是3M或更低,通常为0.1至2.5M、0.3至1.9M、0.5至1.8M、0.7至1.7M、0.9至1.6M或1M至1.4M。盐浓度优选为150mM至1M。令人惊奇地是,Hel308、XPD、RecD和TraI解旋酶在高盐浓度下起作用。优选使用至少0.3M,例如至少0.4M、至少0.5M、至少0.6M、至少0.8M、至少1.0M、至少1.5M、至少2.0M、至少2.5M或至少3.0M的盐浓度来执行所述方法。高盐浓度提供高信噪比,并使得在正常电流波动的背景下,代表核苷酸存在的电流能被识别。

[0667] 通常在存在缓冲液的情况下执行该方法。在上述示例性设备中,缓冲液存在于所述室中以水溶液存在。本发明的方法中可使用任何缓冲液。通常,缓冲液是磷酸盐缓冲液。其它合适的缓冲液包括但不限于HEPES和Tris-HCl缓冲液。通常在pH值为4.0至12.0、4.5至10.0、5.0至9.0、5.5至8.8、6.0至8.7或7.0至8.8或7.5至8.5的情况下执行所述方法。所用的pH值优选为约7.5。

[0668] 可以在温度为0℃至100℃、15℃至95℃、16℃至90℃、17℃至85℃、18℃至80℃、19℃至70℃或20℃至60℃的情况下执行所述方法。通常在室温下执行所述方法。可选地,该方法可选地在支持酶功能的温度下,例如约37℃实施。

[0669] 可以在存在自由核苷酸或自由核苷酸类似物和/或促进解旋酶或构建体起作用的酶辅助因子的情况下执行所述方法。还可以在不存在自由核苷酸或自由核苷酸类似物并且不存在酶辅助因子的情况下执行所述方法。自由核苷酸可以是上述任一个或多个单个核苷酸。游离核苷酸包括但不限于单磷酸腺苷(AMP)、二磷酸腺苷(ADP)、三磷酸腺苷(ATP)、单磷酸鸟苷(GMP)、二磷酸鸟苷(GDP)、三磷酸鸟苷(GTP)、单磷酸胸苷(TMP)、二磷酸胸苷(TDP)、三磷酸胸苷(TTP)、单磷酸尿苷(UMP)、二磷酸尿苷(UDP)、三磷酸尿苷(UTP)、单磷酸胞苷(CMP)、二磷酸胞苷(CDP)、三磷酸胞苷(CTP)、环单磷酸腺苷(cAMP)、环单磷酸鸟苷(cGMP)、脱氧单磷酸腺苷(dAMP)、脱氧二磷酸腺苷(dADP)、脱氧三磷酸腺苷(dATP)、脱氧单磷酸鸟苷(dGMP)、脱氧二磷酸鸟苷(dGDP)、脱氧三磷酸鸟苷(dGTP)、脱氧单磷酸胸苷(dTMP)、脱氧二磷酸胸苷(dTDP)、脱氧三磷酸胸苷(dTTP)、脱氧单磷酸尿苷(dUMP)、脱氧二磷酸尿苷(dUDP)、脱氧三磷酸尿苷(dUTP)、脱氧单磷酸胞苷(dCMP)、脱氧二磷酸胞苷(dCDP)和脱氧三磷酸胞苷(dCTP)。自由核苷酸优选选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP或dCMP。自由核苷酸优选为三磷酸腺苷(ATP)。酶辅助因子是使解旋酶或构建体起作用的因子。酶辅助因子优选为二价金属阳离子。二价金属阳离子优选为 Mg^{2+} 、 Mn^{2+} 、 Ca^{2+} 或 Co^{2+} 。酶辅助因子最优选是 Mg^{2+} 。

[0670] 目标多核苷酸可以以任何顺序接触解旋酶或构建体和孔。优选地,当目标多核苷酸与解旋酶或构建体和孔接触时,目标多核苷酸首先与解旋酶或构建体形成复合物。当跨孔施加电压时,目标多核苷酸/解旋酶或构建体复合物然后与孔形成复合物并控制多核苷酸穿过孔的移动。

[0671] 其它方法

[0672] 本发明还提供了形成传感器以用于表征目标多核苷酸的方法。所述方法包括：在孔和本发明的解旋酶或本发明的构建体之间形成复合物。解旋酶可以是上述任一种。有关本发明的系列和方法，可以使用以上讨论的本发明的解旋酶的任何数量和组合。

[0673] 可以通过在存在目标多核苷酸的情况下使孔与解旋酶或构建体接触，然后在孔上施加电势来形成复合物。施加的电势可以是如上所述的化学电势或电压电势。可选地，可以通过将孔与解旋酶或构建体共价地连接来形成复合物。用于共价连接的方法是本领域已知的并在以下国际申请中进行了公开，例如，公开在申请号为PCT/GB09/001679（公开为W02010/004265）和申请号为PCT/GB10/000133（公开为W02010/086603）的国际申请。复合物是一种用于表征目标多核苷酸的传感器。所述方法优选包括在衍生自Msp的孔和本发明的解旋酶或本发明的构建体之间形成复合物。结合本发明方法所讨论的上述实施方案中的任一个同样适用于所述方法。本发明还提供采用本发明方法制备的传感器。

[0674] 试剂盒

[0675] 本发明还提供一种用于表征目标多核苷酸的试剂盒。

[0676] 在一个实施方案中，试剂盒包含 (a) 本发明的孔或 (b) 本发明的解旋酶或本发明的构建体。孔可以是上述任一种。

[0677] 在另一个实施方案中，试剂盒包含 (a) 本发明的解旋酶或本发明的构建体和 (b) 一个或多个负载部分。每个负载部分可以是能够与目标多核苷酸连接的任何部分。只要解旋酶或构建体可以结合并且其可以与目标多核苷酸连接，每个负载部分可以是任何长度。

[0678] 一个或多个负载部分优选是合成的或人造的。一个或多个负载部分优选是非天然的。

[0679] 合适的负载部分包括但不限于聚合物连接器、化学连接器、多核苷酸或多肽。一个或多个负载部分优选包含多核苷酸或载体多核苷酸。在这种实施方案中，解旋酶或构建体优选与多核苷酸结合（或连接）。可以使用上述任一种多核苷酸。优选地，一个或多个负载部分包含DNA，RNA，修饰的DNA（例如，脱碱基DNA）、RNA、PNA、LNA、BNA或PEG。一个或多个负载部分更优选包含单链或双链DNA或RNA。

[0680] 一个或多个负载部分优选包含与一个或多个多核苷酸结合蛋白结合（或连接）的单链多核苷酸。

[0681] 一个或多个负载部分中的至少一个优选为Y适配器。Y适配器通常包含 (a) 双链区域和 (b) 单链区域或另一端不互补的区域。如果Y适配器包括单链区域，则可以将其描述为具有突出端（overhang）。在Y适配器中存在非互补区域使得适配器具有Y形状，因为与双链部分不同，这两条链通常彼此不杂合。Y适配器优选地包括一个或多个能够将Y适配器与膜连接的锚定物。以上对锚定物进行了更详细的讨论。优选的锚定物是胆固醇。

[0682] Y适配器优选地包含优先旋入孔的前导序列。前导序列有利于本发明的方法。前导序列被设计为优先旋入孔，从而促进目标多核苷酸相对于孔，例如，穿过孔移动。前导序列也可以用于将多核苷酸与如上所述的一个或多个锚定物连接。

[0683] 前导序列通常包含聚合物。聚合物优选是带负电荷的。聚合物优选为多核苷酸，例如，DNA或RNA，修饰的多核苷酸（例如，脱碱基DNA）、PNA、LNA、BNA、聚乙二醇（PEG）或多肽。前导序列优选包含多核苷酸，更优选包含单链多核苷酸。前导序列可以包含任何上述多核苷酸。单链前导序列最优选包含单链DNA，例如，poly dT片段。前导序列优选地包括如下所述

的间隔物。

[0684] 前导序列可以是任何长度,但是长度通常为10至150个核苷酸,例如,长度为20至150个核苷酸。前导系列的长度通常取决于方法中使用的跨膜孔。

[0685] 一个或多个负载部分中的至少一个优选是桥连部分。桥连部分最优选是发夹环或发夹环适配器。可以使用本领域已知的方法来设计合适的发夹环适配器。发夹环可以是任何长度。如果用作负载部分,发夹环通常为400个或更少的核苷酸,例如,350个或更少的核苷酸、300个或更少的核苷酸、250个或更少的核苷酸、200个或更少的核苷酸、150个或更少的核苷酸、100个或更少的核苷酸、90个或更少的核苷酸、80个或更少的核苷酸、70个或更少的核苷酸、60个或更少的核苷酸、50个或更少的核苷酸、40个或更少的核苷酸、30个或更少的核苷酸、20个或更少的核苷酸或者10个或更少的核苷酸。发夹环的长度优选为约1至400、2至300、5至200、6至100个核苷酸。当多核苷酸的两个互补部分杂交形成双链序列(称为茎(stem))时,形成发夹环。如果用作负载部分,则发夹环的茎优选为200个或更少的核苷酸对,例如,150个或更少的核苷酸对、100个或更少的核苷酸对、90个或更少的核苷酸对、80个或更少的核苷酸对、70个或更少的核苷酸对、60个或更少的核苷酸对、50个或更少的核苷酸对、40个或更少的核苷酸对、30个或更少的核苷酸对、20个或更少的核苷酸对或者10个或更少的核苷酸对。一个或多个多核苷酸结合蛋白通常与发夹的环结合,即,不是茎。

[0686] 如果目标多核苷酸是双链的,则一个或多个负载部分优选包含Y适配器且可选地包含桥连部分,例如,发夹环适配器。如果至少一个或多个负载部分是Y适配器,则其可以联合未结合或连接有任何多核苷酸结合蛋白的桥连适配器进行使用。

[0687] 解旋酶或构建体可以在一个或多个负载部分上的一个或多个间隔物处停滞。间隔物在申请号为PCT/GB2014/050175(WO2014/135838)的国际申请中进行了限定。优选的间隔物包括但不限于硝基咪唑、5-硝基咪唑、肌苷、吡啶、2-氨基嘌呤、2-6-二氨基嘌呤、5-溴脱氧尿苷、反式胸苷(反式dT)、反式二脱氧胸苷(ddT)、二脱氧胞苷(ddC)、5-甲基胞苷、5-羟甲基胞苷、2'-O-甲基RNA碱基、异脱氧胞苷(Iso-dC)、异脱氧鸟苷(Iso-dG)、iSpC3基团(即,缺少糖和碱的核苷酸)、可光裂解(PC)基团、己二醇基团、间隔物9(iSp9)基团、间隔物18(iSp18)基团、聚合物或硫醇连接。间隔物可以包括这些基团的任一组合。许多这些基团可以从IDT®(Integrated DNA Technologies®)商购获得。

[0688] 可以使用任何数量的一个或多个负载部分。所述方法可以包括连接两个或更多个负载部分,每个负载部分具有与其结合(连接)的解旋酶或构建体。例如,负载部分可以与目标多核苷酸的每个末端连接。在这种实施方案中,一个负载部分优选是Y适配器,另一个负载部分可以是桥连部分,例如,发夹环适配器。

[0689] 一个或多个负载部分可以以任何方式与目标多核苷酸连接。一个或多个负载部分优选与目标多核苷酸共价地连接。

[0690] 一个或多个负载部分最优选与目标多核苷酸连接。一个或多个负载部分可以与多核苷酸的任一端连接,即,5'或3'末端。负载部分可以与目标多核苷酸的两端连接。可以使用本领域已知的任何方法将一个或多个负载部分与多核苷酸连接。可以在不存在ATP或使用 γ -S-ATP(ATP γ S)代替ATP的情况下,将一个或多个负载部分与多核苷酸连接。

[0691] 可以采用连接酶,例如T4 DNA连接酶、大肠杆菌DNA连接酶、Taq DNA连接酶、Tma DNA连接酶和9°N DNA连接酶将一个或多个负载部分连接。所述连接酶优选在实施例3中所

述的条件下使用。

[0692] 一旦负载部分已经与目标多核苷酸连接,解旋酶或构建体优选保持与负载部分结合(连接)。根据本发明,在其连接之后,解旋酶或构建体可以从一个或多个负载部分中解开连接。

[0693] 对于本发明方法所讨论的任何实施方案同样适用于试剂盒。解旋酶可以是上述任一种。针对本发明系列和方法,试剂盒可以包括上述本发明的解旋酶的任何数量和组合。

[0694] 试剂盒还可以包含膜的组分,例如,形成两性分子层,例如脂质双分子层所需的磷脂。

[0695] 本发明的试剂盒可以额外地包含一个或多个能够使上述任何实施方案得以进行的其它试剂或仪器。这种试剂或仪器包括以下一个或多个:合适的缓冲液(水溶液)、从受体(例如,血管或包含针的仪器)获得样品的装置、用于扩增和/或表达多核苷酸的装置、如上所限定的膜或电压或膜片钳装置。试剂可以以干燥状态存在于试剂盒中,使得流体样品使试剂重新悬浮。可选地,试剂盒还可以包括能够使试剂盒用于本发明方法中的说明书或有关所述方法可用于哪种生物体的细节。可选地,试剂盒可以包含核苷酸。

[0696] 设备

[0697] 本发明还提供一种用于表征目标多核苷酸的设备。所述设备包括多个孔和多个本发明的解旋酶或多个本发明的构建体。所述设备优选地还包括用于实施本发明方法的指令。所述设备可以是用于多核苷酸分析的任何常规设备,例如,阵列或芯片。结合本发明方法所讨论的上述任一实施方式同样适用于本发明设备。解旋酶可以是结合本发明构建体所讨论的上述任一种,包括本发明的解旋酶和根据本发明未经过修饰的解旋酶。所述设备可以包括任何数量和组合的本发明解旋酶。

[0698] 所述设备优选地设置用以执行本发明的方法。

[0699] 所述设备优选地包括:

[0700] 传感器装置,其能够支撑多个孔和膜并且可操作地采用孔和膜进行多核苷酸表征;和

[0701] 至少一个端口,用于传送进行表征的材料。

[0702] 可选地,所述设备优选地包括:

[0703] 传感器装置,其能够支撑多个孔和膜并且可操作地采用孔和膜进行多核苷酸表征;和

[0704] 至少一个储存器,用于容纳进行表征的材料。

[0705] 所述设备更优选包括:

[0706] 传感器装置,其能够支撑多个孔与膜,并且可操作地采用孔和膜进行多聚核苷酸表征;

[0707] 至少一个储存器,用于容纳进行表征的材料;

[0708] 流体系统,用以可控地将材料从所述至少一个储存器向所述传感器装置供给;和

[0709] 一个或多个容器,用于接收相应的样本,所述流体系统用于选择性地使样品从所述一个或多个容器向所述传感器装置供给。所述设备可以是申请号为PCT/GB08/004127(公开为W02009/077734)、PCT/GB10/000789(公开为W02010/122293)、PCT/GB 10/002206(公开为W02011/067559)或PCT/US99/25679(公开为W000/28312)的国际申请中描述的任一种。

[0710] 本发明解旋酶的制备方法

[0711] 本发明还提供制备本发明的修饰的解旋酶的方法。所述方法包括提供Dda解旋酶并对所述解旋酶进行修饰以形成本发明的修饰的解旋酶。

[0712] 所述方法优选还包括确定解旋酶是否能够控制多核苷酸的移动。以上对这种情况的测定方法进行了描述。如果可以控制多核苷酸的移动,则解旋酶已被正确修饰,并且制备出本发明的解旋酶。如果不能控制多核苷酸的移动,则尚未制备出本发明的解旋酶。

[0713] 本发明构建体的制备方法

[0714] 本发明还提供一种制备本发明构建体的方法。所述方法包括将本发明的解旋酶与额外的多核苷酸结合部分连接,优选共价地连接。上述任何解旋酶和部分可用于所述方法中。如上所述,对共价连接的位点和方法进行选择。

[0715] 所述方法优选还包括确定构建体是否能够控制多核苷酸的移动。对以上这种情况的测定方法进行描述。如果可以控制多核苷酸的移动,则解旋酶和部分已被正确连接并且制备出本发明的构建体。如果不能控制多核苷酸的移动,则尚未制备出本发明的构建体。

[0716] 以下实施例对本发明进行说明。

[0717] 实施例1

[0718] 本实施例描述了用以研究MspA-(G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K) 8 (具有G75S/G77S/L88N/D90N/D91N/D118R/Q126R/D134R/E139K突变的SEQ ID NO:2=MspA突变体1) 或MspA-(De1-L74/G75/D118/L119) D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K) 8 (具有D56N/E59R/L88N/D90N/D91N/Q126R/D134R/E139K突变以及缺失L74/G75/D118/L119氨基酸的SEQ ID NO:2=MspA突变体2) 与T4 Dda-E94C/A360C/C109A/C136A (具有E94C/A360C/C114A/C171A/C421D突变的SEQ ID NO:8=酶突变体1a) 之间的相互作用所进行的模拟。

[0719] 利用GROMACS软件包版本4.0.5,采用GROMOS 53a6力场和SPC水模型进行模拟。

[0720] MspA突变体1和MspA突变体2模型基于蛋白质数据库中发现的MspA的晶体结构,记录码为1UUN。使用PyMOL进行相关突变,在MspA突变体2的情况下,将残基L74/G75/D118/L119从桶中删除。然后,使用最陡下降算法对所得到的孔模型进行能量最小化。酶突变体1a模型基于蛋白质数据库中发现的Dda1993结构,记录码为3UPU。同样,使用PyMOL进行相关突变,并且使用最陡下降算法对模型进行能量最小化。

[0721] 然后,将酶突变体1a模型置于MspA突变体1和MspA突变体2之上。对酶突变体1a/MspA突变体1和酶突变体1a/MspA突变体2系统进行三次模拟,其中酶突变体1a的取向在每次模拟中有所不同(三次不同模拟取向的动画示意图,参见图1)。将孔放入包含DPPC分子的脂质膜中,并对模拟箱进行溶剂化处理。在整个模拟过程中,向孔的主干(backbone)施加限制。然而,酶不受限制。使用300K的Berendsen恒温器和Berendsen恒压器,将所述系统在NPT系统中模拟40ns。

[0722] 使用GROMACS分析软件和本地写入代码分析酶和孔之间的接触。图2至图5示出了在MspA突变体1(图2和3)和MspA突变体2(图4和5)与酶突变体1a相互作用的氨基酸残基。下表示出了孔和酶中的氨基酸位点,其中,发现所述孔和酶相互作用(表9示出了在MspA突变体1和酶突变体1a之间测量相互作用时所观察到的MspA突变体1氨基酸接触点,表10示出了在MspA突变体1和酶突变体1a之间测量相互作用时所观察到的酶突变体1a氨基酸接触点,

表11示出了在MspA突变体2和酶突变体1a之间测量相互作用时所观察到的MspA突变体2氨基酸接触点,图12示出了在MspA突变体2和酶突变体1a之间测量相互作用时所观察到的酶突变体1a氨基酸接触点)。图6示出了孔(MspA突变体2)中哪些氨基酸与酶(酶突变体1a)中的特定氨基酸相互作用。模拟数据可以用于鉴定酶突变体1a的一部分,其是被修饰的以改善酶和纳米孔之间的相互作用,从而提供目标多核苷酸相对于跨膜孔,例如穿过跨膜孔的更一致的移动。

	运行图 1	运行图 2	运行图 3
	孔氨基酸残基	孔氨基酸残基	孔氨基酸残基
[0723]	57	56	57
	59	57	136
	136	136	59
	134	139	134
	56	52	56
	54	134	12
	12	138	139
	169	55	58
	14	59	137
	58		14
	55		48
	52		169
	138		
[0724]	139		
	137		

[0725] 表9

[0726]

运行图1	运行图2	运行图3
酶氨基酸残基	酶氨基酸残基	酶氨基酸残基
2	180	255
180	199	216
179	202	221
178	1	227
227	4	351
4	51	321
177	434	254
212	179	258
1	178	224
194	177	257

204	197	256
176	5	223
213	201	212
3	181	308
216	200	207
211	6	350
202		228
224		210
223		319
191		304
199		209
201		347
434		261
405		260
255		247

[0727] 表10

[0728]	运行图 1	运行图 2	运行图 3
	孔氨基酸残基	孔氨基酸残基	孔氨基酸残基
	59	59	56
[0729]	57	57	59
	134	169	57
	136	134	136
	169	136	12
	56	56	14
	137	54	134
	58	14	54
	14	12	169
	135		53
	60		
	170		

[0730] 表11

	运行图 1	运行图 2	运行图 3
	酶氨基酸残基	酶氨基酸残基	酶氨基酸残基
[0731]	350	202	199
	258	180	197
	223	179	185
	195	212	198
	198	258	207
	438	211	202
	260	198	223
	207	265	180
	226	260	209
	304	259	210
	200	255	203
	227	1	204
	347	200	437
	321	300	200
	422	203	211
	318	261	405
	415	216	227
	210	177	258
	229	213	212
	255	207	256
	224	337	216
	228	204	189
[0732]	208	434	228
	193	298	220
	256		219

[0733] 表12

[0734] 实施例2

[0735] 本实施例描述了用以研究两种不同酶(野生型Dda 1993(SEQ ID NO:8))和T4Dda-E94C/A360C(具有E94C/A360C突变的SEQ ID NO:8)=酶突变体18)和多核苷酸之间的相互作用所进行的模拟。

[0736] 进行模拟是为了评估哪些残基与酶结合位点内的DNA接触。利用GROMACS软件包版本4.0.5,采用AMBER-99SB力场和TIP3P水模型进行模拟。

[0737] 对两种酶进行了模拟,即,野生型Dda1993和酶突变体18。酶突变体18以闭合复合物形式进行模拟,使得在E94C和A360C之间存在二硫键。野生型Dda 1993的初始结构是基于蛋白质数据库中可用的结构,记录码为3UPU。该PDB文件中的结构是Dda 1993-K38A。因此,

在野生型Dda 1993模拟中,使用PyMOL将残基38回复突变为赖氨酸。酶突变体18模型也基于3UPU中的结构,其中,在PyMOL中进行相关突变。在两种酶模拟中模拟的DNA是存在于3UPU的晶体结构中的DNA(DNA序列是poly(dT))。然后,使用最陡下降算法对得到的酶/DNA模型进行能量最小化。然后,对模拟箱进行溶剂化处理,并进行另一轮的能量最小化。在整个模拟过程中,酶和DNA不受限制。使用300K的Berendsen恒温器和Berendsen恒压器,将所述系统在NPT系统中模拟20ns。

[0738] 使用GROMACS分析软件和本地写入代码分析酶和DNA之间的接触。下表示出了在两种酶中与存在于晶体结构3UPU中的DNA相互作用的氨基酸(表13和表14示出了在野生型Dda 1993和DNA之间测量相互作用时所观察到的野生型Dda 1993氨基酸接触点,表15和表16示出了在酶突变体18和DNA之间测量相互作用时所观察到的酶突变体18氨基酸接触点)。模拟数据可用于鉴定Dda 1993酶和酶突变体18中的位点,其可以突变以改善酶和DNA之间的相互作用,从而提供目标多核苷酸相对于跨膜孔,例如穿过跨膜孔的更一致的移动。

[0739]

氨基酸残基 WT Dda 1993
N242
K397
H396
N293
H82
H64
F98
V150
Y415
T63

氨基酸残基 WT Dda 1993
I87
D417
P62
S287
H414
K243
M119
A416
L420
K86

F240
T241
T394
T80
F276
P89
N88
S83
I289
P152
P274

表 13

N292
V96
D151
N155
W378
E288
T278
V286
K123
R148
L97

表 14

[0740]

氨基酸残基酶突变体 18
H82
K397
N242
H64
V150
H396
T241
N293
T63
N88
F98
F240
T80
S83
P89
T394
S287
F276
Y415
I87
D417

表 15

氨基酸残基酶突变体 18
V286
I289
P62
N292
R148
K101
K243
A416
K86
L420
N155
E288
P274
D151
P152
P285
M119
T278
D121
K284
Q272

表 16

[0741] 实施例3

[0742] 本实施例使用具有多种不同解旋酶的T4 Dda-E94C/C109A/C136A/A360C (具有E94C/C109A/C136A/A360C突变并且然后(Δ M1)G1的SEQ ID NO:8) (酶突变体1) 比较控制DNA构建体X (参见图7) 穿过纳米孔的移动。所有测试的解旋酶控制DNA穿过纳米孔的移动, 并在随着DNA移位穿过纳米孔时观察电流变化。测试的解旋酶具有a) 至少一个与单链DNA (ssDNA) 中的一个或多个核苷酸相互作用的氨基酸取代或b) 与跨膜孔相互作用的在解旋酶的一部分中的一个或多个修饰或者a) 和b) 两种变化。本实施例研究补体向前滑移 (complement slips forward) 数量/3.6kb、补体向前滑移数量/kb、由于向前滑移而在构建体X中缺失的%碱基、补体中向前滑移的总长度以及向前滑移的平均长度。本实施例中研究的解旋酶沿着多核苷酸的5' 至3' 方向上移动。当孔捕获到多核苷酸的5' 末端 (解旋酶移动远离的一端) 时, 解旋酶在施加电势而产生的场的方向上工作, 并使螺旋多核苷酸移动进入孔中并进入反式腔室。在本实施例中, 向前滑移涉及DNA相对于孔向前移动 (即, 朝着其3' 末端并远离其5' 末端) 至少4个连续的核苷酸。

[0743] 材料和方法

[0744] 在建立实验之前, 用适当的酶将DNA构建体X (最终浓度为0.1nM) 在室温下预培养5分钟 (参见以下提供的酶列表 (添加到纳米孔系统的最终浓度为10nM, 在缓冲液 (253mM KCl、50mM磷酸钾、pH值为8.0的2mM EDTA) 中提供) 。5分钟后, 将TMAD (添加到纳米孔系统的最终浓度为100 μ M) 添加到预混合物中, 并将混合物再培养5分钟。最后, 将MgCl₂ (添加到纳米孔系统的最终浓度为2mM) 、ATP (添加到纳米孔系统的最终浓度为2mM) 和KCl (添加到纳米孔系统的最终浓度为500mM终浓度) 添加到预混合物中。

[0745] 在缓冲液 (25mM磷酸钾缓冲液、150mM亚铁氰化钾 (II) 、150mM铁氰化钾 (III) , pH值为8.0) 中, 从插入嵌段共聚物中的单个MspA纳米孔 (MspA-(De1-L74/G75/D118/L119) D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K) 8 (具有D56F/E59R/L88N/D90N/D91N/Q126R/D134R/E139K突变并缺失L74/G75/D118/L119氨基酸的SEQ ID NO:2) (MspA突变体3)) 中获得电测量值。在单孔插入嵌段共聚物得以实现之后, 然后将2mL缓冲液 (25mM磷酸钾缓冲液、150mM亚铁氰化钾 (II) 、150mM铁氰化钾 (III) , pH值为8.0) 流过系统以除去任何过量的MspA纳米孔。然后将150 μ L pH值为8.0的500mM KCl、25mM磷酸钾流过系统。10分钟后, 再将150 μ L pH值为8.0的500mM KCl、25mM磷酸钾流过系统, 然后将酶 (参见以下列表, 最终浓度为10nM) 、DNA构建体X (最终浓度为0.1nM) 、燃料 (fuel) (最终浓度为2mM的MgCl₂, 最终浓度为2mM的ATP) 预混合物 (总计150 μ L) 流入单个纳米孔实验系统。在-140mV下运行实验, 并监测经解旋酶控制的DNA移动。

[0746] 结果

[0747] 对许多不同的解旋酶进行研究, 以为了确定至少一个或多个取代对被认为与DNA构建体X相互作用的解旋酶区域或者被认为与纳米孔相互作用的一个或多个修饰的影响。为了鉴定经解旋酶控制的DNA移位得以改善的解旋酶, 对五个不同的参数进行了研究: 1) 补体向前滑移数量/3.6kb; 2) 补体向前滑移数量/kb; 3) 由于向前滑移而在构建体X中缺失的碱基%; 4) 补体中向前滑移的总长度; 5) 向前滑移的平均长度。

[0748] 使用以下步骤计算每千碱基或每3.6kb向前滑移测量值: 1) 使用HMM算法将经解旋酶控制的DNA移动映射到模型; 2) 然后, 将经解旋酶控制的DNA移动进行过滤; 3) 对映射的经

解旋酶控制的DNA移动进行检查以确保映射准确;4) 根据千碱基或3.6kb随着至少4个连续核苷酸的向前滑移移动,确定被分类的转变。由于向前滑移而在构建体X中缺失的%碱基是由于沿着DNA构建体X向前滑移而在构建体X中缺失的碱基数量表示为百分比的测量。补体滑移的总长度是链的补体片段中所有滑移的总和。平均滑移长度是链的补体片段中所有滑移的总和除以补体中的总滑移数量。

[0749] 以下表17示出了相比酶突变体1所测试的不同酶。在所测试的酶中,突变体2至13和15至17在与酶突变体1相比时呈现参数1至5中至少一个的改进。

[0750] 突变体5至13具有至少一个核苷酸,其与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用,被包含较大侧链(R基团)并且在与跨膜孔相互作用的解旋酶的一部分中没有一个或多个修饰的氨基酸取代。突变体5至13在与酶突变体1相比时均呈现参数1至5中至少一个的改进,这导致包含较大侧链(R基团)并且与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸被取代。显然,在与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的位点上用较大侧链基团进行至少一个取代导致移动控制得以改善。

[0751] 突变体14具有至少一个氨基酸,其与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用,被包含较小侧链(R基团)并且在与跨膜孔相互作用的解旋酶的一部分中没有一个或多个修饰的氨基酸取代。突变体14在与酶突变体1相比时呈现参数1至5中至少一个的改进,这导致包含较小侧链(R基团)并且与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸被取代。显然,在与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的位点上用较小侧链基团进行至少一个取代导致移动控制变差。

[0752] 突变体4具有至少一个氨基酸取代,其与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个磷酸基相互作用,至少一个与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸被包含较大侧链(R基团)并且在与跨膜孔相互作用的解旋酶的一部分中没有一个或多个修饰的氨基酸取代。突变体4在与酶突变体1相比时呈现参数1至5中至少一个的改进,这导致氨基酸组合被取代,例如,第一种是与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个磷酸基相互作用;第二种是与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用。此外,突变体4在与酶突变体9相比时呈现参数1至5中至少一个的改进,这导致与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个磷酸基相互作用的氨基酸被取代。显然,在与一个或多个核苷酸的糖和/或碱基相互作用的位点上用较大侧链基团进行取代以及进行与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个磷酸基相互作用的取代导致移动控制得以改善。

[0753] 突变体2具有至少一个氨基酸,其与单链DNA(ssDNA)中一个或多个核苷酸的一个或多个糖和/或碱基相互作用,被包含较大侧链(R基团)并且在与跨膜孔相互作用的解旋酶的一部分中没有一个或多个修饰的氨基酸取代。突变体2在与酶突变体1相比时呈现参数1至5中至少一个的改进,这导致组合发生变化,例如,第一种是与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代;第二种是与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰。突变体2在与酶突变体9相比时也呈现参数1至5中至少一个的改进,这导致第二种情况,即,在与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰。此外,突变体2在与酶突变体16相比时呈现参数1至5中至少一个的改进,这导致第一

种情况,即,与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代。显然,进行组合变化(第一种是与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代;第二种是与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰)导致酶的移动控制呈现改进。

[0754] 突变体3具有至少一个氨基酸,其与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用,被包含较大侧链(R基团)并且在与跨膜孔相互作用的解旋酶的一部分中有一个或多个修饰的氨基酸取代。突变体3在与酶突变体1相比时呈现参数1至5中至少一个的改善,这导致组合发生变化,例如,第一种是在与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代;第二种是在与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰。突变体3在与酶突变体9相比时也呈现参数1至5中至少一个的改进,这导致第二种情况,即,与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰。此外,突变体3在与酶突变体17相比时也呈现参数1至5中至少一个的改进,这导致第一种情况,即,与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代。显然,进行组合取代(第一种是与单链DNA(ssDNA)中一个或多个核苷酸的糖和/或碱基相互作用的至少一个取代;第二种是与跨膜孔相互作用的解旋酶的一部分中的一个或多个修饰)导致酶的移动控制呈现改进。

[0755] 酶ID

[0756] 酶突变体1=T4Dda-E94C/C109A/C136A/A360C(具有E94C/C109A/C136A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0757] 酶突变体2=T4 Dda-E94C/F98W/C109A/C136A/K199L/A360C(具有E94C/F98W/C109A/C136A/K199L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0758] 酶突变体3=T4Dda-F98W/E94C/C109A/C136A/K194L/A360C(具有F98W/E94C/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0759] 酶突变体4=T4Dda-S83H/E94C/F98W/C109A/C136A/A360C(具有S83H/E94C/F98W/C109A/C136A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0760] 酶突变体5=T4Dda-E94C/F98W/C109A/C136A/F276K/A360C(具有E94C/F98W/C109A/C136A/F276K/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0761] 酶突变体6=T4Dda-E94C/F98W/C109A/C136A/S287R/A360C(具有E94C/F98W/C109A/C136A/S287R/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0762] 酶突变体7=T4Dda-E94C/F98W/C109A/C136A/S287W/A360C(具有E94C/F98W/C109A/C136A/S287W/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0763] 酶突变体8=T4Dda-E94C/F98W/C109A/C136A/S287F/A360C(具有E94C/F98W/C109A/C136A/S287F/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0764] 酶突变体9=T4 Dda-E94C/F98W/C109A/C136A/A360C(具有E94C/F98W/C109A/C136A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0765] 酶突变体10=T4Dda-P89F/E94C/C109A/C136A/A360C(具有P89F/E94C/C109A/C136A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0766] 酶突变体11=T4Dda-E94C/C109A/C136A/V150H/A360C(具有/C109A/C136A/V150H/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0767] 酶突变体12=T4Dda-E94C/C109A/C136A/V150I/A360C(具有E94C/C109A/C136A/V150I/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0768] 酶突变体13=T4Dda-E94C/C109A/C136A/P152F/A360C(具有E94C/C109A/C136A/P152F/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0769] 酶突变体14=T4Dda-E94C/F98A/C109A/C136A/A360C(具有E94C/F98A/C109A/C136A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0770] 酶突变体15=T4Dda-E94C/C109A/C136A/K199L/A360C(具有E94C/C109A/C136A/K199L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0771] 酶突变体16=T4Dda-E94C/C109A/C136A/K194L/A360C(具有E94C/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0772] 酶突变体17=T4Dda-E94C/C109A/C136A/W195A/A360C(具有E94C/C109A/C136A/W195A/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)

[0773]

酶	补体滑移数量 /3.6kb	补体滑移数量 /kb	通过滑移缺失 的%补体碱基	补体滑移 的总长度	平均滑移长度
1	16	4.4	10.5	377	23.6
2	6	1.7	3.6	129	21.5
3	3	0.8	1.9	70	23.3
4	6	1.7	3.4	121	24
5	7.7	2.1	3.9	140	22
6	8.1	2.3	4.4	158	27
7	5.8	1.6	3.1	112	27
8	6.7	1.9	3.7	134	28
9	11	3.1	6.5	233	21.2
10	17	4.7	10.3	372	21.9
11	6	1.7	3.7	134	22.3
12	12	3.3	8.7	314	26.2
13	12	3.3	9.9	355	29.6
14	17	4.7	11.7	420	24.7
15	8	2.2	5.3	191	23.9
16	4	1.1	2.8	100	25.0
17	15	4.2	10.3	371	24.7

[0774] 表17

[0775] 实施例4

[0776] 本实施例示出了解旋酶T4Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)如何控制DNA构建体X(参见图7)通过CsgG纳米孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C)9(具有Y51T/F56Q突变的SEQ ID NO:66,其中,StrepII(C)是SEQ ID NO:67并且在C末端上进行连接)的移动。

[0777] 材料和方法

[0778] 如实施例3的材料和方法部分中所述,制备DNA构建体X解旋酶(T4Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO: 8并且然后(Δ M1)G1)预混合物。

[0779] 除了纳米孔是CsgG而不是MspA之外,以实施例3中所述的类似方法,从插入嵌段共聚物的单个CsgG纳米孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有Y51T/F56Q突变的SEQ ID NO:66,其中,StrepII(C)是SEQ ID NO:67并在C末端上进行连接)中获得电测量值。

[0780] 结果

[0781] 观察到解旋酶控制DNA移动,随着T4Dda-E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)控制DNA构建体X(参见图7))穿过CsgG纳米孔(CsgG-Eco-(Y51T/F56Q)-StrepII(C))9(具有Y51T/F56Q突变的SEQ ID NO:66,其中,StrepII(C)是SEQ ID NO:67并且在C末端上进行连接)移动。图8A中示出了解旋酶控制DNA移动的电迹线的示例,图8B和图8C示出了同一迹线的放大视图。

[0782] 实施例5

[0783] 本实施例示出了如何将发夹与RNA链的3'末端连接,并对RNA链进行逆转录以产生RNA/DNA杂交体。随后,将非RNA多核苷酸与RNA/DNA杂交体中RNA链的5'末端连接以便负载DNA解旋酶,T4Dda(E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)。观察到经解旋酶控制的RNA/DNA构建体穿过纳米孔的移动。

[0784] 材料和方法

[0785] 1.发夹连接

[0786] 将以下表18中列出的试剂混合并置于以下表19中程序的热循环仪装置上。然后使用Agencourt Ampure SPRI珠以每 μ L样品1.8 μ L SPRI珠的比例对混合物进行纯化。纯化后,使用表20中的Life Technologies Super Script II:试剂进行逆转录,根据厂商协议进行混合,并放置在表21中程序的热循环仪装置上。然后使用Agencourt Ampure SPRI珠以每 μ L样品1.8 μ L SPRI珠的比例对混合物进行纯化。该样品被称为逆转录样品1。

[0787]

试剂	体积	物料浓度	最终浓度
RNA 链(3'末端用开放阅读框 SEQ ID NO: 68 多聚腺苷酸化)	0.72 μ l	556ng/ μ l	0.2 μ M
polyT 发夹 (SEQ ID NO: 72 的 5'末端与磷酸基连接,SEQ ID NO: 72 的 3'末端与 4 个 iSpC3 间隔物连接, 4 个 iSpC3 间隔物的另一末端与 SEQ ID NO: 73 的 5'末端连接)	0.4 μ l	50 μ M	1 μ M
T4 DNA 连接酶缓冲液	4 μ l	5 $\times$	1 $\times$
T4 DNA 连接酶	1 μ l	2000U/ μ l	2000 U
NF H ₂ O	13.88 μ l		
总计	20 μ l		

[0788] 表18

[0789]

循环数量	步骤	温度(°C)	时间
1	连接	16	2:00:00

[0790] 表19

[0791]

试剂	体积	物料浓度	最终浓度
在 polyT 发夹连接后的 RNA	7 ul	35.5ng/ul	248.5ng/反应
dNTPs	1 ul	每个 10uM	0.5 uM
NF H ₂ O	5 ul		
第一链缓冲液	4 ul	5×	1×
0.1M DTT	2 ul	0.1M	0.01M
Super Script II	1 ul	200U	200U
总计	20 ul		

[0792] 表20

[0793]

循环数量	步骤	温度(°C)	时间
1	逆转录	42	0:50:00
2	变性	70	0:15:00

[0794] 表21

[0795] 随后,通过混合表22中列出的试剂并将混合物置于表23中程序的热循环仪装置上,将“非RNA多核苷酸”(30个SpC3间隔物连接到SEQ ID NO:69的5'末端,SEQ ID NO:69的3'末端连接到4个iSp18间隔物的5'末端,4个iSp18间隔物的3'末端连接到SEQ ID NO:70的5'末端,SEQ ID NO:70的3'末端连接到4个5-硝基吡啶的5'末端,4个5-硝基吡啶的3'末端连接到RNA序列CAAGGG)与RNA多核苷酸(其在以上步骤中进行逆转录)连接。然后,采用Agencourt Ampure SPRI珠以每μL样品1.8μL SPRI珠的比例对混合物进行纯化。该样品称为连接样品1。

[0796]

试剂	体积	物料浓度	最终浓度
转录样品 1	1.5 ul	166ng/ul	250ng/反应
T4 RNA 连接酶 1 反应缓冲液	2 ul	10×	1×
“非 RNA 多核苷酸 (见以上描述)”	2.4 ul	50 uM	8.33 uM
ATP	0.4 ul	50mM	1 mM
NF H ₂ O	0.8 ul		
T4 RNA 连接酶 1	2.9 ul	10U/ul	29U
PEG 8k	10 ul	50%	25%
总计	20 ul		

[0797] 表22

[0798]

循环数量	步骤	温度(°C)	时间
1	连接	16	4:00:00

[0799] 表23

[0800] 对以下表24中列出的试剂进行混合并在65℃下进行培养,然后以0.1℃/秒的速率进行冷却。该样品称为DNA/RNA构建体Y。

[0801]

试剂	体积	物料浓度	最终浓度
连接样品 1	9 μL	~1 μM	942 μM
锚定物 (SEQ ID NO: 71 的 3'末端与 6 个 iSp18 间隔物的 5'末端连接, 2 个胸腺嘧啶和 3'胆固醇 TEG)	0.36 μL	100 μM	3.77 μM
10 mM TRIS pH 7.5 50 mM NaCl	0.19 μL	50×	1×

[0802]

总计	9.55 μL		

[0803] 表24

[0804] 电生理学

[0805] 将DNA/RNA构建体Y与2μL 17.4μM的T4Dda (E94C/F98W/C109A/C136A/K194L/A360C (具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(ΔM1)G1)培养20分钟。然后向培养过的混合物中添加2.1μL的800μM TMAD,并在室温下保持10分钟。然后将该样品稀释到缓冲液(276μL的500mM KCl、25mM的磷酸钾;pH值为8.0) MgCl₂ (4μL, 150mM) 和ATP (4μL, 150mM) 中,总体积为300μL。

[0806] 在缓冲液(25mM磷酸钾缓冲液、150mM亚铁氰化钾、150mM铁氰化钾;pH值为8.0)中,从插入嵌段共聚物的单个MspA纳米孔中获得电测量值。在实现单个孔插入嵌段共聚物中之后,然后将缓冲液(2mL,25mM磷酸钾缓冲液、150mM亚铁氰化钾、150mM铁氰化钾;pH值为8.0)流过系统以除去任何过量的MspA纳米孔。

[0807] 在添加DNA/RNA构建体Y和解旋酶之前,将过量的缓冲液(500mM KCl、25mM磷酸钾;pH值为8.0)流过系统。最后,向纳米孔系统添加与DNA/RNA构建体Y结合的T4Dda (E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变并且然后(Δ M1)G1)的SEQ ID NO:8)),在-140mV下进行实验,并且检测经解旋酶控制的DNA移动。

[0808] 结果:

[0809] 本实施例示出了如何将非RNA多核苷酸与RNA(其已经进行过转录)连接以便负载DNA解旋酶,T4Dda (E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1)),随后观察到解旋酶控制构建体移动。图9中示出了经T4 Dda (E94C/F98W/C109A/C136A/K194L/A360C(具有E94C/F98W/C109A/C136A/K194L/A360C突变的SEQ ID NO:8并且然后(Δ M1)G1))解旋酶控制的移动的示例。

[0810] 本发明的优选Dda解旋酶的比对(SEQ ID NO:8至23)

[0811]

[illegible]

序列表

<110> 牛津纳米孔技术公司

<120> 经修饰的酶

<130> N406654W0

<150> GB1417712.5

<151> 2014-10-07

<150> PCT/GB2015/051291

<151> 2015-05-01

<160> 73

<170> PatentIn version 3.5

<210> 1

<211> 552

<212> DNA

<213> 耻垢分枝杆菌 (*Mycobacterium smegmatis*)

<400> 1

```

ggcctggata acgaacttag cctggtggac ggccaagatc gcacgctgac ggtgcaacaa 60
tgggatacct tcctgaatgg tgtgtttccg ctggatcgta accgcctgac ccgtgaatgg 120
tttcattccg gtcgcgcaaa atatatcgtc gcaggcccggt gtgctgacga attcgaaggc 180
acgctggaac tgggttatca gattggcttt ccgtggtcac tgggcgttgg tatcaacttc 240
tcgtacacca cgccgaatat tctgatcgat gacggtgata ttaccgcacc gccgtttggc 300
ctgaacagcg tgattacgcc gaacctgttt ccgggtgtta gcattctctgc cgatctgggc 360
aacggtccgg gcattcaaga agtggcaacc tttagtgtgg acgtttccgg cgctgaaggc 420
gggtgtcgcg tgtctaagtc ccacggtacc gttacgggcg cggccggcgg tgtcctgctg 480
cgtccgttcg cgcgcctgat tgcgagcacc ggcgactctg ttacgaccta tggcgaaccg 540
tggaatatga ac                                     552

```

<210> 2

<211> 184

<212> PRT

<213> 耻垢分枝杆菌

<400> 2

```

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

```

<210> 4

<211> 293

<212> PRT

<213> 人工序列

<220>

<223> α溶血素突变体 (E111N/K147N)

<400> 4

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

260	265	270
Lys Trp Thr Asp Arg Ser Ser Glu Arg Tyr Lys Ile Asp Trp Glu Lys		
275	280	285
Glu Glu Met Thr Asn		
290		
<210> 5		
<211> 184		
<212> PRT		
<213> 耻垢分枝杆菌		
<400> 5		
Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu		
1	5	10
Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp		
20	25	30
Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr		
35	40	45
Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu		
50	55	60
Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe		
65	70	75
Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Ala		
85	90	95
Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly		
100	105	110
Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val		
115	120	125
Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val		
130	135	140
Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu		
145	150	155
Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr		
165	170	175
Tyr Gly Glu Pro Trp Asn Met Asn		
180		
<210> 6		
<211> 184		
<212> PRT		
<213> 耻垢分枝杆菌		
<400> 6		

Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
 1 5 10 15
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
 20 25 30
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 35 40 45
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
 50 55 60
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
 65 70 75 80
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asp Asp Gly Asp Ile Thr Gly
 85 90 95
 Pro Pro Phe Gly Leu Glu Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
 100 105 110
 Val Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
 115 120 125
 Ala Thr Phe Ser Val Asp Val Ser Gly Pro Ala Gly Gly Val Ala Val
 130 135 140
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
 145 150 155 160
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
 165 170 175
 Tyr Gly Glu Pro Trp Asn Met Asn
 180

<210> 7

<211> 183

<212> PRT

<213> 耻垢分枝杆菌

<400> 7

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

Tyr Thr Thr Pro Asn Ile Leu Ile Asp Gly Gly Asp Ile Thr Gln Pro
 85 90 95
 Pro Phe Gly Leu Asp Thr Ile Ile Thr Pro Asn Leu Phe Pro Gly Val
 100 105 110
 Ser Ile Ser Ala Asp Leu Gly Asn Gly Pro Gly Ile Gln Glu Val Ala
 115 120 125
 Thr Phe Ser Val Asp Val Lys Gly Ala Lys Gly Ala Val Ala Val Ser
 130 135 140
 Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu Arg
 145 150 155 160
 Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr Tyr
 165 170 175
 Gly Glu Pro Trp Asn Met Asn
 180

<210> 8

<211> 439

<212> PRT

<213> 肠杆菌噬菌体T4 (Enterobacteria phage T4)

<400> 8

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

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

<210> 9

<211> 678

<212> PRT

<213> 海洋红嗜热盐菌 (Rhodothermus marinus)

<400> 9

```

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

```


290	295	300
Leu Val Asn Arg Asn Ala Pro Leu His Gly Leu Phe Asn Gly Asp Leu		
305	310	315
Val Leu Val Glu Thr Val Gly Pro Leu Glu His Arg Arg Val Gly Arg		320
	325	330
Arg Gly Arg Pro Pro Val Asp Leu Tyr Phe Arg Asp Val Glu Leu Leu		335
	340	345
Tyr Pro His Glu Lys Pro Arg Asn Arg Ile Arg Cys Lys Leu Leu Glu		350
	355	360
Asn Leu Leu Glu Ser Pro Asp Gly Gln Leu Ser Pro Asp Ile Ile Gln		365
	370	380
Ala Leu Leu Ile Asp Phe Tyr Arg Arg His Pro Ser Leu Lys His Gly		
385	390	395
Ser Ser Glu Phe Arg Leu Met Leu Ala Asn Asp Ala Tyr Phe Asn Ala		400
	405	410
Leu His Val Arg Tyr Gly Tyr Ala Met Thr Val His Lys Ala Gln Gly		415
	420	425
Gly Glu Trp Lys Arg Ala Thr Val Val Phe Asn Asp Trp Arg His Phe		430
	435	440
Arg His Ala Glu Phe Phe Arg Trp Ala Tyr Thr Ala Ile Thr Arg Ala		445
	450	460
Arg Glu Glu Leu Leu Thr Ile Gly Ala Pro Ser Phe Glu Ala Leu Ser		
465	470	475
Asp Met Arg Trp Gln Pro Ala Pro Ser Val Pro Ala Pro Glu Gln Ala		480
	485	490
Ala Glu Asn Ala Thr Arg Phe Pro Leu Lys Ala Leu Glu Thr Tyr His		495
	500	505
Gln Arg Leu Ser Glu Ala Leu Thr Ala Ala Gly Ile Glu Thr Thr Gly		510
	515	520
Val Glu Leu Leu Gln Tyr Ala Val Arg Tyr His Leu Ala Arg Ala Asp		525
	530	540
Arg Thr Thr Arg Ile Gln Tyr Tyr Tyr Arg Gly Asp Gly Gln Ile Ser		
545	550	555
Arg Ile Val Thr Leu Gly Gly Ala Asp Asp Pro Glu Leu Thr Gln Gln		560
	565	570
Ala Tyr Ala Leu Phe Glu Arg Ile Leu Ser Glu Pro Pro Ala Asp Ser		575
	580	585
Gly Glu Leu Pro Glu Asn Pro Leu Leu Arg Glu Phe Leu Glu Arg Ala		590
	595	600
		605

His Leu Arg Leu Glu Gly Ser Gly Ile Arg Ile Val His Trp Lys Glu
 610 615 620
 Met Pro Tyr Ala Leu Arg Leu Tyr Phe Ser Ala Asp Gly Glu Asn Val
 625 630 635 640
 Thr Ile Asp Phe Tyr Tyr Asn Arg Arg Gly Val Trp Thr His Ala Gln
 645 650 655
 Glu Val Gly Arg Ser Ser Ser Gly Ala Leu Phe Ala Arg Ile Gln Ser
 660 665 670
 Leu Leu Gln Ala Asp Ser
 675
 <210> 10
 <211> 496
 <212> PRT
 <213> 蓝丝菌属 (Cyanothecae) ATCC51142
 <400> 10
 Met Ser Gln Ser Val Val Val Pro Asp Glu Leu Gly Glu Ile Ile Thr
 1 5 10 15
 Ala Val Ile Glu Phe Tyr Gln Asp Ala Val Asp Lys Ile Glu Pro Lys
 20 25 30
 Ile Val Phe Leu Glu Leu Arg Lys Asn Val Val Asp Trp Val Ser Arg
 35 40 45
 Thr Gln Leu Lys Ile Glu Glu Lys Glu Ile Gln Ala Thr Gly Leu Thr
 50 55 60
 Arg Gln Gln Gln Thr Ala Tyr Lys Glu Met Ile Asn Phe Ile Glu Asn
 65 70 75 80
 Ser Ser Glu Gln Tyr Phe Arg Leu Ser Gly Tyr Ala Gly Thr Gly Lys
 85 90 95
 Ser Phe Leu Met Ala Lys Val Ile Glu Trp Leu Lys Gln Glu Asp Tyr
 100 105 110
 Lys Tyr Ser Val Ala Ala Pro Thr Asn Lys Ala Ala Lys Asn Leu Thr
 115 120 125
 Gln Ile Ala Arg Ser Gln Gly Ile Lys Ile Glu Ala Thr Thr Val Ala
 130 135 140
 Lys Leu Leu Lys Leu Gln Pro Thr Ile Asp Val Asp Thr Gly Gln Gln
 145 150 155 160
 Ser Phe Glu Phe Asn Ser Glu Lys Glu Leu Glu Leu Lys Asp Tyr Asp
 165 170 175
 Val Ile Ile Ile Asp Glu Tyr Ser Met Leu Asn Lys Asp Asn Phe Arg
 180 185 190

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

<210> 11

<211> 421

<212> PRT

<213> *Salinibacter ruber*

<400> 11

```

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

```

275	280	285
Glu Val Arg Val Lys Lys Ala Gln Val Glu Thr Phe Glu Ala Asp Asp		
290	295	300
Gln Ser Glu Trp Thr Val Trp Glu Leu Lys Ile Arg Thr Pro Gly Arg		
305	310	315
Gly Leu Thr Arg Thr Ile His Val Leu His Glu Glu Glu Arg Glu Arg		
325	330	335
Tyr Glu Asn Ala Leu Glu Arg Arg Arg Gly Lys Ala Glu Asp Asp Pro		
340	345	350
Ser Lys Trp Asp Arg Phe Phe Glu Leu Arg Glu Arg Phe Ala Arg Val		
355	360	365
Asp Tyr Ala Tyr Ala Thr Thr Val His Arg Ala Gln Gly Ser Thr Tyr		
370	375	380
Asp Thr Val Phe Val Asp His Arg Asp Leu Arg Val Cys Arg Gly Glu		
385	390	395
Glu Arg Gly Ala Leu Leu Tyr Val Ala Val Thr Arg Pro Ser Arg Arg		
405	410	415
Leu Ala Leu Leu Val		
420		
<210> 12		
<211> 500		
<212> PRT		
<213> Sullfurimonas gotlandica GD1		
<400> 12		
Met Lys Ile Leu Asn Lys Glu Thr Tyr Lys Leu Ser Leu His Gln Glu		
1	5	10
Glu Val Phe Thr Gln Ile Val Ser Gln Leu Asp Thr Lys Val Ser Ser		
20	25	30
Ile Leu Lys Ser Thr Asn Ile Glu Asp Tyr Leu Leu Ser Leu Thr Gly		
35	40	45
Pro Ala Gly Thr Gly Lys Thr Phe Leu Thr Thr Gln Ile Ala Lys Tyr		
50	55	60
Leu Val Glu Lys Arg Lys Glu Ser Glu Tyr Pro Met Ser Ser Asp Phe		
65	70	75
Asp Phe Thr Ile Thr Ala Pro Thr His Lys Ala Val Gly Val Leu Ser		
85	90	95
Lys Leu Leu Arg Glu Asn Asn Ile Gln Ser Ser Cys Lys Thr Ile His		
100	105	110
Ser Phe Leu Gly Ile Lys Pro Phe Ile Asp Tyr Thr Thr Gly Glu Glu		

115	120	125
Lys Phe Val Val Asp Lys Thr Asn Lys Arg Lys Asp Arg Thr Ser Ile		
130	135	140
Leu Ile Val Asp Glu Ser Ser Met Ile Gly Asn Thr Leu Tyr Glu Tyr		
145	150	155
Ile Leu Glu Ala Ile Glu Asp Lys Arg Val Asn Val Val Leu Phe Ile		
165	170	175
Gly Asp Pro Tyr Gln Leu Leu Pro Ile Glu Asn Ser Lys Asn Glu Ile		
180	185	190
Tyr Asp Leu Pro Asn Arg Phe Phe Leu Ser Glu Val Val Arg Gln Ala		
195	200	205
Glu Asn Ser Tyr Ile Ile Arg Val Ala Thr Lys Leu Arg Glu Arg Ile		
210	215	220
Lys Asn Gln Asp Phe Ile Ser Leu Gln Gln Phe Phe Gln Glu Asn Met		
225	230	235
Glu Asp Glu Ile Thr Phe Phe His Asn Lys Glu Ala Phe Leu Glu Asp		
245	250	255
Phe Tyr Lys Glu Glu Glu Trp Tyr Lys Glu Asn Lys Ile Leu Ala Thr		
260	265	270
Tyr Lys Asn Lys Asp Val Asp Ala Phe Asn Lys Ile Ile Arg Asn Lys		
275	280	285
Phe Trp Glu Gln Lys Gly Asn Thr Thr Pro Ser Thr Leu Leu Ala Gly		
290	295	300
Asp Met Ile Arg Phe Lys Asp Ala Tyr Thr Val Gly Asp Ile Thr Ile		
305	310	315
Tyr His Asn Gly Gln Glu Leu Gln Leu Gly Ser Thr Glu Val Lys Tyr		
325	330	335
His Asp Ser Leu His Ile Glu Tyr Trp Glu Cys Lys Ser Ile Tyr Ala		
340	345	350
Leu Glu Gln Gln Val Phe Arg Val Val Asn Pro Asp Ser Glu Ala Val		
355	360	365
Phe Asn Gln Lys Leu Gln Ser Leu Ala Thr Lys Ala Lys Gln Ala Lys		
370	375	380
Phe Pro Asp Asn Lys Lys Leu Trp Lys Leu Tyr Tyr Glu Thr Arg Asn		
385	390	395
Met Phe Ala Asn Val Gln Tyr Ile His Ala Ser Thr Ile His Lys Leu		
405	410	415
Gln Gly Ser Thr Tyr Asp Val Ser Tyr Ile Asp Ile Phe Ser Leu Val		
420	425	430

His Asn His Tyr Met Ser Asp Glu Glu Lys Tyr Arg Leu Leu Tyr Val
 435 440 445
 Ala Ile Thr Arg Ala Ser Lys Asp Ile Lys Ile Phe Met Ser Ala Phe
 450 455 460
 Asp Arg Thr Ser Asp Glu Lys Val Ile Ile Asn Asn Gln Asn Ser Glu
 465 470 475 480
 Thr Met Asn Thr Leu Lys Gln Leu His Asp Ile Asp Ile Ile Leu Lys
 485 490 495
 Asp Leu Asp Leu
 500
 <210> 13
 <211> 450
 <212> PRT
 <213> 弧菌噬菌体亨瑞特(Vibrio phage henriette)12B8
 <400> 13
 Met Ala Asp Phe Glu Leu Thr Leu Gly Gln Lys Thr Val Leu Gly Glu
 1 5 10 15
 Val Ile Ser Thr Ile Leu Lys Pro Val Asn Leu Asn Asp Thr Ser Arg
 20 25 30
 Phe His Thr Met His Gly Pro Ala Gly Ser Gly Lys Thr Thr Val Leu
 35 40 45
 Gln Arg Ile Ile Ser Gln Ile Pro Ala Tyr Lys Thr Ile Gly Phe Cys
 50 55 60
 Ser Pro Thr His Lys Ser Val Lys Val Ile Arg Arg Met Ala Arg Glu
 65 70 75 80
 Ala Gly Ile Ser His Arg Val Asp Ile Arg Thr Ile His Ser Ala Leu
 85 90 95
 Gly Leu Val Met Lys Pro Val Arg Gly Asp Glu Val Leu Val Lys Glu
 100 105 110
 Pro Phe Ala Glu Glu Arg Ile Tyr Asp Val Leu Ile Ile Asp Glu Ala
 115 120 125
 Gly Met Leu Asn Asp Glu Leu Ile Met Tyr Ile Leu Glu Ser Gln Ser
 130 135 140
 Ser Lys Val Ile Phe Val Gly Asp Met Cys Gln Ile Gly Pro Ile Gln
 145 150 155 160
 Ser Asn Leu Pro Glu Glu Asp Gly Tyr Thr Pro Thr Ser Thr Asp Asp
 165 170 175
 Val Ser Lys Val Phe Thr Glu Val Glu Met Met Ser Ala Leu Thr Glu
 180 185 190

Val Val Arg Gln Ala Glu Gly Ser Pro Ile Ile Gln Leu Ala Thr Glu		
195	200	205
Phe Arg Leu Ala Gln Asp Asp Ile Tyr Ala Asp Leu Pro Arg Ile Val		
210	215	220
Thr Asn Thr Thr Pro Asp Gly Asn Gly Ile Ile Thr Met Pro Asn Gly		
225	230	235 240
Asn Trp Val Asp Ser Ala Val Ala Arg Phe Gln Ser Asp Gln Phe Lys		
	245	250 255
Glu Asp Pro Asp His Cys Arg Ile Val Cys Tyr Thr Asn Ala Met Val		
	260	265 270
Asp Leu Cys Asn Asp Leu Val Arg Lys Arg Leu Phe Gly Ala Asp Val		
	275	280 285
Pro Glu Trp Leu Glu Asp Glu Ile Leu Val Ala Gln Glu Met Gly Ser		
	290	295 300
Thr Trp Asn Asn Ala Asp Glu Leu Arg Ile Val Ser Ile Asp Asp His		
305	310	315 320
Phe Asp Gln Gln Tyr Glu Val Pro Cys Trp Arg Met Gln Leu Glu Ser		
	325	330 335
Val Glu Asp His Lys Leu His Asn Ala Leu Val Val Lys Gly Asp Tyr		
	340	345 350
Ile Glu Asp Phe Lys Phe Arg Leu Asn Ala Ile Ala Glu Arg Ala Asn		
	355	360 365
Thr Asp Lys Asn Met Ser Gly Met His Trp Lys Glu Phe Trp Gly Met		
	370	375 380
Arg Lys Lys Phe Asn Thr Phe Lys Asn Val Tyr Ala Gly Thr Ala His		
385	390	395 400
Lys Ser Gln Gly Ser Thr Phe Asp Tyr Thr Tyr Val Phe Thr Pro Asp		
	405	410 415
Phe Tyr Lys Phe Gly Ala Thr Met Thr Ile Lys Arg Leu Leu Tyr Thr		
	420	425 430
Ala Ile Thr Arg Ser Arg Tyr Thr Thr Tyr Phe Ala Met Asn Thr Gly		
	435	440 445
Ala Gln		
450		
<210> 14		
<211> 421		
<212> PRT		
<213> 弧菌噬菌体 (Vibrio phage) phi-pp2		
<400> 14		

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

305	310	315	320
Glu Phe Thr Val Leu Trp Asp Asp Leu Gln Lys Glu Arg Phe Ala Arg			
	325	330	335
Tyr Leu Ser Val Ala Ala Gly Thr Tyr Lys Ser Met Arg Gly Asn Thr			
	340	345	350
Lys Arg Tyr Trp Arg Ala Phe Trp Gly Leu Lys Glu Gln Met Ile Glu			
	355	360	365
Thr Lys Ser Leu Gly Ala Ser Thr Val His Lys Ser Gln Gly Thr Thr			
	370	375	380
Val Lys Gly Val Cys Leu Tyr Thr Gln Asp Met Gly Tyr Ala Glu Pro			
385	390	395	400
Glu Ile Leu Gln Gln Leu Val Tyr Val Gly Leu Thr Arg Pro Thr Asp			
	405	410	415
Trp Ala Leu Tyr Asn			
	420		
<210> 15			
<211> 434			
<212> PRT			
<213> 气单胞菌属噬菌体(Aeromonas phage)65			
<400> 15			
Met Ser Glu Ser Glu Ile Thr Leu Thr Pro Ser Gln Asn Met Ala Val			
1	5	10	15
Asn Glu Val Lys Asn Gly Thr Gly His Ile Thr Ile Ser Gly Pro Pro			
	20	25	30
Gly Ser Gly Lys Thr Phe Leu Val Lys Tyr Leu Ile Lys Met Leu Gly			
	35	40	45
Asp Glu Leu Gly Thr Val Leu Ala Ala Pro Thr His Gln Ala Lys Ile			
	50	55	60
Val Leu Thr Glu Met Ser Gly Ile Glu Ala Cys Thr Ile His Ser Leu			
65	70	75	80
Met Lys Ile His Pro Glu Thr Leu Glu Asp Ile Gln Ile Phe Asp Gln			
	85	90	95
Ser Lys Leu Pro Asp Leu Ser Asn Ile Arg Tyr Leu Ile Val Glu Glu			
	100	105	110
Ala Ser Met His Ser Lys Thr Leu Phe Lys Ile Thr Met Lys Ser Ile			
	115	120	125
Pro Pro Thr Cys Arg Ile Ile Ala Ile Gly Asp Lys Asp Gln Ile Gln			
	130	135	140
Pro Glu Glu His Ala Gln Gly Glu Leu Ser Pro Tyr Phe Thr Asp Pro			

145	150	155	160
Arg Phe Ser Gln Ile Arg Leu Thr Asp Ile Met Arg Gln Ser Leu Asp			
	165	170	175
Asn Pro Ile Ile Gln Val Ala Thr Lys Ile Arg Glu Gly Gly Trp Ile			
	180	185	190
Glu Pro Asn Trp Asn Arg Asp Thr Lys Thr Gly Val Tyr Lys Val Ser			
	195	200	205
Gly Ile Thr Asp Leu Val Asn Ser Tyr Leu Arg Ala Val Lys Thr Pro			
210	215	220	
Glu Asp Leu Thr Lys Tyr Arg Phe Leu Ala Tyr Thr Asn Lys Val Val			
225	230	235	240
Asn Lys Val Asn Ser Ile Val Arg Glu His Val Tyr Lys Thr Lys Leu			
	245	250	255
Pro Phe Ile Glu Gly Glu Lys Ile Val Leu Gln Glu Pro Val Met Val			
	260	265	270
Glu His Glu Asp Asp Thr Ile Glu Thr Ile Phe Thr Asn Gly Glu Val			
	275	280	285
Val Thr Ile Asn Glu Ile Glu Val Phe Asp Arg Thr Ile Arg Ile Asp			
290	295	300	
Gly Ser Pro Glu Phe Lys Val Asn Ala Ala Lys Leu Ser Val Ser Ser			
305	310	315	320
Asp Tyr Ser Gly Ile Glu His Asp Phe Cys Val Leu Tyr Gly Ser Glu			
	325	330	335
Ser Arg Leu Glu Phe Glu Tyr Gln Leu Ser Glu Ser Ala Gly Asn Ile			
	340	345	350
Lys Gln Met Gly Lys Gly Gly Asn Gln Arg Ser Ala Trp Lys Ser Phe			
	355	360	365
Trp Ala Ala Lys Lys Met Phe Ile Glu Thr Lys Ser Leu Gly Ala Ser			
	370	375	380
Thr Ile His Lys Ser Gln Gly Ser Thr Val Lys Gly Val Trp Leu Ala			
385	390	395	400
Leu His Asp Ile His Tyr Ala Asp Glu Glu Leu Lys Gln Gln Leu Val			
	405	410	415
Tyr Val Gly Val Thr Arg Pro Thr Asp Phe Cys Leu Tyr Phe Asp Gly			
	420	425	430
Thr Lys			
<210> 16			
<211> 420			
<212> PRT			

<213> 气单胞菌属噬菌体CC2

<400> 16

```

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

```

290	295	300
Val Thr Ser Asp Phe Ser Gly Val Thr His Asp Ile Met Ser Val Tyr		
305	310	315
Gly Glu Asp Ser Lys Ala Glu Phe Asn Tyr Gln Leu Ser Glu Ala Ala		320
	325	330
Ala Val Ile Lys Gln Met Gln Arg Gly Gln Thr Lys Ala Ala Trp Ala		335
	340	345
Ser Phe Trp Asp Ala Lys Lys Thr Phe Thr Glu Thr Lys Ser Leu Gly		350
	355	360
Ala Cys Thr Ile His Lys Ser Gln Gly Ser Thr Val Lys Gly Val Trp		365
370	375	380
Leu Gly Leu His Asp Ile Ser Tyr Ala Asp Thr Asp Leu Gln Gln Gln		
385	390	395
Leu Val Tyr Val Gly Val Thr Arg Pro Thr Asp Phe Cys Leu Tyr Phe		400
	405	410
Asp Gly Ser Lys		415
420		
<210> 17		
<211> 443		
<212> PRT		
<213> 阪崎肠杆菌噬菌体(Cronobacter phage) vB CsaM GAP161		
<400> 17		
Met Ser Glu Leu Thr Phe Asp Asp Leu Ser Asp Asp Gln Lys Ser Ala		
1	5	10
His Asp Arg Val Ile His Asn Ile Gln Asn Ala Ile His Thr Thr Ile		15
	20	25
Thr Gly Gly Pro Gly Val Gly Lys Thr Thr Leu Val Lys Phe Val Phe		30
	35	40
Asn Thr Leu Lys Gly Leu Gly Ile Ser Gly Ile Trp Leu Thr Ala Pro		45
50	55	60
Thr His Gln Ala Lys Asn Val Leu Ala Ala Ala Thr Gly Met Asp Ala		
65	70	75
Thr Thr Ile His Ser Ala Leu Lys Ile Ser Pro Val Thr Asn Glu Glu		80
	85	90
Leu Arg Val Phe Glu Gln Gln Lys Gly Lys Lys Ala Pro Asp Leu Ser		95
	100	105
Thr Cys Arg Val Phe Val Val Glu Glu Val Ser Met Val Asp Met Asp		110
	115	120
Leu Phe Arg Ile Ile Arg Arg Ser Ile Pro Ser Asn Ala Val Ile Leu		125

130	135	140
Gly Leu Gly Asp Lys Asp Gln Ile Arg Pro Val Asn Ala Asp Gly Arg		
145	150	155
Val Glu Leu Ser Pro Phe Phe Asp Glu Glu Ile Phe Asp Val Ile Arg		
165	170	175
Met Asp Lys Ile Met Arg Gln Ala Glu Gly Asn Pro Ile Ile Gln Val		
180	185	190
Ser Arg Ala Val Arg Asp Gly Lys Met Leu Lys Pro Met Ser Val Gly		
195	200	205
Asp Leu Gly Val Phe Gln His Ala Asn Ala Val Asp Phe Leu Arg Gln		
210	215	220
Tyr Phe Arg Arg Val Lys Thr Pro Asp Asp Leu Ile Glu Asn Arg Met		
225	230	235
Phe Ala Tyr Thr Asn Asp Asn Val Asp Lys Leu Asn Ala Thr Ile Arg		
245	250	255
Lys His Leu Tyr Lys Thr Thr Glu Pro Phe Ile Leu Asp Glu Val Ile		
260	265	270
Val Met Gln Glu Pro Leu Val Gln Glu Met Arg Leu Asn Gly Gln Ile		
275	280	285
Phe Thr Glu Ile Val Tyr Asn Asn Asn Glu Lys Ile Arg Val Leu Glu		
290	295	300
Ile Ile Pro Arg Arg Glu Val Ile Lys Ala Glu Lys Cys Asp Glu Lys		
305	310	315
Ile Glu Ile Glu Phe Tyr Leu Leu Lys Thr Val Ser Leu Glu Glu Glu		
325	330	335
Thr Glu Ala Gln Ile Gln Val Val Val Asp Pro Val Met Lys Asp Arg		
340	345	350
Leu Gly Asn Tyr Leu Ala Tyr Val Ala Ser Thr Tyr Lys Arg Ile Lys		
355	360	365
Gln Gln Thr Gly Tyr Lys Ala Pro Trp His Ser Phe Trp Ala Ile Lys		
370	375	380
Asn Lys Phe Gln Asp Val Lys Pro Leu Pro Val Cys Thr Tyr His Lys		
385	390	395
Ser Gln Gly Ser Thr Tyr Asp His Ala Tyr Met Tyr Thr Arg Asp Ala		
405	410	415
Tyr Ala Phe Ala Asp Tyr Asp Leu Cys Lys Gln Leu Ile Tyr Val Gly		
420	425	430
Val Thr Arg Ala Arg Tyr Thr Val Asp Tyr Val		
435	440	

<210> 18

<211> 442

<212> PRT

<213> 克雷伯氏菌噬菌体(Klebsiella phage)KP15

<400> 18

```

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

```

Val Met Gln Glu Pro Leu Val Thr Glu Gly Arg Val Asn Gly Val Ser		
275	280	285
Phe Val Glu Val Ile Tyr Asn Asn Asn Glu Gln Ile Lys Ile Leu Glu		
290	295	300
Ile Ile Pro Arg Ser Asp Thr Ile Lys Ala Asp Arg Cys Asp Pro Val		
305	310	315
Gln Ile Asp Tyr Phe Leu Met Lys Thr Glu Ser Met Phe Glu Asp Thr		
	325	330
Lys Ala Asp Ile Gln Val Ile Ala Asp Pro Val Met Gln Glu Arg Leu		
	340	345
Gly Asp Tyr Leu Asn Tyr Val Ala Phe Gln Tyr Lys Lys Met Lys Gln		
	355	360
Glu Thr Gly Tyr Lys Ala Pro Trp Tyr Ser Phe Trp Gln Ile Lys Asn		
	370	375
Lys Phe Gln Thr Val Lys Ala Leu Pro Val Cys Thr Tyr His Lys Gly		
385	390	395
Gln Gly Ser Thr Tyr Asp His Ser Tyr Met Tyr Thr Arg Asp Ala Tyr		
	405	410
Ala Tyr Ala Asp Tyr Glu Leu Cys Lys Gln Leu Leu Tyr Val Gly Thr		
	420	425
Thr Arg Ala Arg Phe Thr Val Asp Tyr Val		
	435	440
<210> 19		
<211> 438		
<212> PRT		
<213> 寡养单胞菌噬菌体 (Stenotrophomonas phage) IME13		
<400> 19		
Met Val Thr Tyr Asp Asp Leu Thr Val Gly Gln Lys Asp Ala Ile Glu		
1	5	10
Lys Ala Leu Gln Ala Met Arg Thr Lys Arg His Ile Thr Ile Arg Gly		
	20	25
Pro Ala Gly Ser Gly Lys Thr Thr Met Thr Arg Phe Leu Leu Glu Arg		
	35	40
Leu Phe Gln Thr Gly Gln Gln Gly Ile Val Leu Thr Ala Pro Thr His		
	50	55
Gln Ala Lys Lys Glu Leu Ser Lys His Ala Leu Arg Lys Ser Tyr Thr		
65	70	75
Ile Gln Ser Val Leu Lys Ile Asn Pro Ser Thr Leu Glu Glu Asn Gln		
	85	90
		95

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

405	410	415
Pro Asp Val Ala Gln Glu Leu Ile Tyr Val Gly Asn Thr Arg Ala Arg		
420	425	430
Lys Asn Val Cys Phe Val		
435		
<210> 20		
<211> 442		
<212> PRT		
<213> 不动杆菌噬菌体 (Acinetobacter phage) Ac42		
<400> 20		
Met Asn Phe Glu Asp Leu Thr Glu Gly Gln Lys Asn Ala Tyr Thr Ala		
1 5 10 15		
Ala Ile Lys Ala Ile Glu Thr Val Pro Ser Ser Ser Ala Glu Lys Arg		
20 25 30		
His Leu Thr Ile Asn Gly Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr		
35 40 45		
Lys Phe Leu Ile Ala Glu Leu Ile Arg Arg Gly Glu Arg Gly Val Tyr		
50 55 60		
Leu Ala Ala Pro Thr His Gln Ala Lys Lys Val Leu Ser Gln His Ala		
65 70 75 80		
Gly Met Glu Ala Ser Thr Ile His Ser Leu Leu Lys Ile Asn Pro Thr		
85 90 95		
Thr Tyr Glu Asp Ser Thr Thr Phe Glu Gln Lys Asp Val Pro Asp Met		
100 105 110		
Ser Glu Cys Arg Val Leu Ile Cys Asp Glu Ala Ser Met Tyr Asp Leu		
115 120 125		
Lys Leu Phe Gln Ile Leu Met Ser Ser Ile Pro Leu Cys Cys Thr Val		
130 135 140		
Ile Ala Leu Gly Asp Ile Ala Gln Ile Arg Pro Val Glu Pro Gly Ala		
145 150 155 160		
Phe Glu Gly Gln Val Ser Pro Phe Phe Thr Tyr Glu Lys Phe Glu Gln		
165 170 175		
Val Ser Leu Thr Glu Val Met Arg Ser Asn Ala Pro Ile Ile Asp Val		
180 185 190		
Ala Thr Ser Ile Arg Thr Gly Asn Trp Ile Tyr Glu Asn Val Ile Asp		
195 200 205		
Gly Ala Gly Val His Asn Leu Thr Ser Glu Arg Ser Val Lys Ser Phe		
210 215 220		
Met Glu Lys Tyr Phe Ser Ile Val Lys Thr Pro Glu Asp Leu Phe Glu		

225	230	235	240
Asn Arg Leu Leu Ala Phe Thr Asn Lys Ser Val Asp Asp Leu Asn Lys			
	245	250	255
Ile Val Arg Lys Lys Ile Tyr Asn Thr Leu Glu Pro Phe Ile Asp Gly			
	260	265	270
Glu Val Leu Val Met Gln Glu Pro Leu Ile Lys Ser Tyr Thr Tyr Glu			
	275	280	285
Gly Lys Lys Val Ser Glu Ile Val Phe Asn Asn Gly Glu Met Val Lys			
	290	295	300
Val Leu Cys Cys Ser Gln Thr Ser Asp Glu Ile Ser Val Arg Gly Cys			
305	310	315	320
Ser Thr Lys Tyr Met Val Arg Tyr Trp Gln Leu Asp Leu Gln Ser Leu			
	325	330	335
Asp Asp Pro Asp Leu Thr Gly Ser Ile Asn Val Ile Val Asp Glu Ala			
	340	345	350
Glu Ile Asn Lys Leu Asn Leu Val Leu Gly Lys Ser Ala Glu Gln Phe			
	355	360	365
Lys Ser Gly Ala Val Lys Ala Ala Trp Ala Asp Trp Trp Lys Leu Lys			
	370	375	380
Arg Asn Phe His Lys Val Lys Ala Leu Pro Cys Ser Thr Ile His Lys			
385	390	395	400
Ser Gln Gly Thr Ser Val Asp Asn Val Phe Leu Tyr Thr Pro Cys Ile			
	405	410	415
His Lys Ala Asp Ser Gln Leu Ala Gln Gln Leu Leu Tyr Val Gly Ala			
	420	425	430
Thr Arg Ala Arg His Asn Val Tyr Tyr Ile			
	435	440	
<210> 21			
<211> 442			
<212> PRT			
<213> 志贺氏杆菌噬菌体 (Shigella phage) SP18			
<400> 21			
Met Ile Lys Phe Glu Asp Leu Asn Thr Gly Gln Lys Glu Ala Phe Asp			
1	5	10	15
Tyr Ile Thr Glu Ala Ile Gln Arg Arg Ser Gly Glu Cys Ile Thr Leu			
	20	25	30
Asn Gly Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Val Ile			
	35	40	45
Asp His Leu Val Arg Asn Gly Val Met Gly Ile Val Leu Ala Ala Pro			

50	55	60
Thr His Gln Ala Lys	Lys Val Leu Ser Lys	Leu Ser Gly Gln Thr Ala
65	70	75
Asn Thr Ile His Ser	Ile Leu Lys Ile Asn	Pro Thr Thr Tyr Glu Asp
85	90	95
Gln Asn Ile Phe Glu	Gln Arg Glu Met Pro	Asp Met Ser Lys Cys Asn
100	105	110
Val Leu Val Cys Asp	Glu Ala Ser Met Tyr	Asp Gly Ser Leu Phe Lys
115	120	125
Ile Ile Cys Asn Ser	Val Pro Glu Trp Cys	Thr Ile Leu Gly Ile Gly
130	135	140
Asp Met His Gln Leu	Gln Pro Val Asp Pro	Gly Ser Thr Gln Gln Lys
145	150	155
Ile Ser Pro Phe Phe	Thr His Pro Lys Phe	Lys Gln Ile His Leu Thr
165	170	175
Glu Val Met Arg Ser	Asn Ala Pro Ile Ile	Glu Val Ala Thr Glu Ile
180	185	190
Arg Asn Gly Gly Trp	Phe Arg Asp Cys Met	Tyr Asp Gly His Gly Val
195	200	205
Gln Gly Phe Thr Ser	Gln Thr Ala Leu Lys	Asp Phe Met Val Asn Tyr
210	215	220
Phe Gly Ile Val Lys	Asp Ala Asp Met Leu	Met Glu Asn Arg Met Tyr
225	230	235
Ala Tyr Thr Asn Lys	Ser Val Glu Lys Leu	Asn Asn Ile Ile Arg Arg
245	250	255
Lys Leu Tyr Glu Thr	Asp Lys Ala Phe Leu	Pro Tyr Glu Val Leu Val
260	265	270
Met Gln Glu Pro His	Met Lys Glu Leu Glu	Phe Glu Gly Lys Lys Phe
275	280	285
Ser Glu Thr Ile Phe	Asn Asn Gly Gln Leu	Val Arg Ile Lys Asp Cys
290	295	300
Lys Tyr Thr Ser Thr	Ile Leu Arg Cys Lys	Gly Glu Ser His Gln Leu
305	310	315
Val Ile Asn Tyr Trp	Asp Leu Glu Val Glu	Ser Ile Asp Glu Asp Glu
325	330	335
Glu Tyr Gln Val Asp	Arg Ile Lys Val Leu	Pro Glu Asp Gln Gln Pro
340	345	350
Lys Phe Gln Ala Tyr	Leu Ala Lys Val Ala	Asp Thr Tyr Lys Gln Met
355	360	365

Lys Ala Ala Gly Lys Arg Pro Glu Trp Lys Asp Phe Trp Lys Ala Arg
 370 375 380
 Arg Thr Phe Leu Lys Val Arg Ala Leu Pro Val Ser Thr Ile His Lys
 385 390 395 400
 Ala Gln Gly Val Ser Val Asp Lys Ala Phe Ile Tyr Thr Pro Cys Ile
 405 410 415
 His Met Ala Glu Ala Ser Leu Ala Ser Gln Leu Ala Tyr Val Gly Ile
 420 425 430
 Thr Arg Ala Arg Tyr Asp Ala Tyr Tyr Val
 435 440
 <210> 22
 <211> 439
 <212> PRT
 <213> 耶尔森氏菌噬菌体(Yersinia phage)phiR1-RT
 <400> 22
 Met Ile Thr Tyr Asp Asp Leu Thr Asp Gly Gln Lys Ser Ala Phe Asp
 1 5 10 15
 Asn Thr Met Glu Ala Ile Lys Asn Lys Lys Gly His Ile Thr Ile Asn
 20 25 30
 Gly Pro Ala Gly Thr Gly Lys Thr Thr Leu Thr Lys Phe Ile Ile Asp
 35 40 45
 His Leu Ile Lys Thr Gly Glu Ala Gly Ile Ile Leu Cys Ala Pro Thr
 50 55 60
 His Gln Ala Lys Lys Val Leu Ser Lys Leu Ser Gly Met Asp Ala Ser
 65 70 75 80
 Thr Ile His Ser Val Leu Lys Ile Asn Pro Thr Thr Tyr Glu Glu Asn
 85 90 95
 Gln Ile Phe Glu Gln Arg Glu Val Pro Asp Leu Ala Ala Cys Arg Val
 100 105 110
 Leu Ile Cys Asp Glu Ala Ser Phe Tyr Asp Arg Lys Leu Phe Gly Ile
 115 120 125
 Ile Leu Ala Thr Val Pro Ser Trp Cys Thr Val Ile Ala Leu Gly Asp
 130 135 140
 Lys Asp Gln Leu Arg Pro Val Thr Pro Gly Glu Ser Glu Gln Gln Leu
 145 150 155 160
 Ser Pro Phe Phe Ser His Ala Lys Phe Lys Gln Val His Leu Thr Glu
 165 170 175
 Ile Lys Arg Ser Asn Gly Pro Ile Ile Gln Val Ala Thr Asp Ile Arg
 180 185 190

Asn Gly Gly Trp Leu Ser Glu Asn Ile Val Asp Gly Glu Gly Val His		
195	200	205
Ala Phe Asn Ser Asn Thr Ala Leu Lys Asp Phe Met Ile Arg Tyr Phe		
210	215	220
Asp Val Val Lys Thr Ala Asp Asp Leu Ile Glu Ser Arg Met Leu Ala		
225	230	235 240
Tyr Thr Asn Lys Ser Val Asp Lys Leu Asn Gly Ile Ile Arg Arg Lys		
	245	250 255
Leu Tyr Glu Thr Asp Lys Pro Phe Ile Asn Gly Glu Val Leu Val Met		
	260	265 270
Gln Glu Pro Leu Met Lys Glu Leu Glu Phe Asp Gly Lys Lys Phe His		
	275	280 285
Glu Ile Val Phe Asn Asn Gly Gln Leu Val Lys Ile Leu Tyr Ala Ser		
	290	295 300
Glu Thr Ser Thr Phe Ile Ser Ala Arg Asn Val Pro Gly Glu Tyr Met		
305	310	315 320
Ile Arg Tyr Trp Asn Leu Glu Val Glu Thr Ala Asp Ser Asp Asp Asp		
	325	330 335
Tyr Ala Thr Ser Gln Ile Gln Val Ile Cys Asp Pro Ala Glu Met Thr		
	340	345 350
Lys Phe Gln Met Phe Leu Ala Lys Thr Ala Asp Thr Tyr Lys Asn Ser		
	355	360 365
Gly Val Lys Ala Tyr Trp Lys Asp Phe Trp Ser Val Lys Asn Lys Phe		
	370	375 380
Lys Lys Val Lys Ala Leu Pro Val Ser Thr Ile His Lys Ser Gln Gly		
385	390	395 400
Cys Thr Val Asn Asn Thr Phe Leu Tyr Thr Pro Cys Ile His Met Ala		
	405	410 415
Asp Ala Gln Leu Ala Lys Gln Leu Leu Tyr Val Gly Ala Thr Arg Ala		
	420	425 430
Arg Thr Asn Leu Tyr Tyr Ile		
	435	
<210> 23		
<211> 441		
<212> PRT		
<213> 沙门氏菌噬菌体 (Salmonella phage) S16		
<400> 23		
Met Ile Thr Phe Glu Gln Leu Thr Ser Gly Gln Lys Leu Ala Phe Asp		
1	5	10 15

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

Leu Asn Gly Asn Lys Gly Phe Ser Ser Glu Asp Lys Gly Glu Trp Lys		
20	25	30
Leu Lys Leu Asp Ala Ser Gly Asn Gly Gln Ala Val Ile Arg Phe Leu		
35	40	45
Pro Ala Lys Thr Asp Asp Ala Leu Pro Phe Ala Ile Leu Val Asn His		
50	55	60
Gly Phe Lys Lys Asn Gly Lys Trp Tyr Ile Glu Thr Cys Ser Ser Thr		
65	70	75
His Gly Asp Tyr Asp Ser Cys Pro Val Cys Gln Tyr Ile Ser Lys Asn		
85	90	95
Asp Leu Tyr Asn Thr Asn Lys Thr Glu Tyr Ser Gln Leu Lys Arg Lys		
100	105	110
Thr Ser Tyr Trp Ala Asn Ile Leu Val Val Lys Asp Pro Gln Ala Pro		
115	120	125
Asp Asn Glu Gly Lys Val Phe Lys Tyr Arg Phe Gly Lys Lys Ile Trp		
130	135	140
Asp Lys Ile Asn Ala Met Ile Ala Val Asp Thr Glu Met Gly Glu Thr		
145	150	155
Pro Val Asp Val Thr Cys Pro Trp Glu Gly Ala Asn Phe Val Leu Lys		
165	170	175
Val Lys Gln Val Ser Gly Phe Ser Asn Tyr Asp Glu Ser Lys Phe Leu		
180	185	190
Asn Gln Ser Ala Ile Pro Asn Ile Asp Asp Glu Ser Phe Gln Lys Glu		
195	200	205
Leu Phe Glu Gln Met Val Asp Leu Ser Glu Met Thr Ser Lys Asp Lys		
210	215	220
Phe Lys Ser Phe Glu Glu Leu Asn Thr Lys Phe Asn Gln Val Leu Gly		
225	230	235
Thr Ala Ala Leu Gly Gly Ala Ala Ala Ala Ala Ser Val Ala Asp		
245	250	255
Lys Val Ala Ser Asp Leu Asp Asp Phe Asp Lys Asp Met Glu Ala Phe		
260	265	270
Ser Ser Ala Lys Thr Glu Asp Asp Phe Met Ser Ser Ser Ser Ser Asp		
275	280	285
Asp Gly Asp Leu Asp Asp Leu Leu Ala Gly Leu		
290	295	
<210> 26		
<211> 232		
<212> PRT		

<213> 细菌噬菌体 (Bacteriophage) T7

<400> 26

```

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

```

<210> 27

<211> 324

<212> PRT

<213> 疱疹病毒 (Herpes virus) 1

<400> 27

```

Met Asp Ser Pro Gly Gly Val Ala Pro Ala Ser Pro Val Glu Asp Ala
1           5           10           15

```

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

<210> 28

<211> 251

<212> PRT

<213> 人工序列

<220>

<223> PCNA的亚基1.

<400> 28

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

<210> 29

<211> 245

<212> PRT

<213> 人工序列

<220>

<223> PCNA的亚基2.

<400> 29

```

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

```

<210> 30

<211> 246

<212> PRT

<213> 人工序列

<220>

<223> PCNA的亚基3.

<400> 30

```

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

```

<210> 31

<211> 608

<212> PRT

<213> 枯草芽孢杆菌噬菌体(Bacillus subtilis phage)phi29

<400> 31

```

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

```

Gly Lys Tyr Val Trp Asp Glu Asp Tyr Pro Leu His Ile Gln His Ile		
275	280	285
Arg Cys Glu Phe Glu Leu Lys Glu Gly Tyr Ile Pro Thr Ile Gln Ile		
290	295	300
Lys Arg Ser Arg Phe Tyr Lys Gly Asn Glu Tyr Leu Lys Ser Ser Gly		
305	310	315
Gly Glu Ile Ala Asp Leu Trp Leu Ser Asn Val Asp Leu Glu Leu Met		
	325	330
Lys Glu His Tyr Asp Leu Tyr Asn Val Glu Tyr Ile Ser Gly Leu Lys		
	340	345
Phe Lys Ala Thr Thr Gly Leu Phe Lys Asp Phe Ile Asp Lys Trp Thr		
	355	360
Tyr Ile Lys Thr Thr Ser Glu Gly Ala Ile Lys Gln Leu Ala Lys Leu		
	370	375
Met Leu Asn Ser Leu Tyr Gly Lys Phe Ala Ser Asn Pro Asp Val Thr		
	385	390
Gly Lys Val Pro Tyr Leu Lys Glu Asn Gly Ala Leu Gly Phe Arg Leu		
	405	410
Gly Glu Glu Glu Thr Lys Asp Pro Val Tyr Thr Pro Met Gly Val Phe		
	420	425
Ile Thr Ala Trp Ala Arg Tyr Thr Thr Ile Thr Ala Ala Gln Ala Cys		
	435	440
Tyr Asp Arg Ile Ile Tyr Cys Asp Thr Asp Ser Ile His Leu Thr Gly		
	450	455
Thr Glu Ile Pro Asp Val Ile Lys Asp Ile Val Asp Pro Lys Lys Leu		
	465	470
Gly Tyr Trp Ala His Glu Ser Thr Phe Lys Arg Ala Lys Tyr Leu Arg		
	485	490
Gln Lys Thr Tyr Ile Gln Asp Ile Tyr Met Lys Glu Val Asp Gly Lys		
	500	505
Leu Val Glu Gly Ser Pro Asp Asp Tyr Thr Asp Ile Lys Phe Ser Val		
	515	520
Lys Cys Ala Gly Met Thr Asp Lys Ile Lys Lys Glu Val Thr Phe Glu		
	530	535
Asn Phe Lys Val Gly Phe Ser Arg Lys Met Lys Pro Lys Pro Val Gln		
	545	550
Val Pro Gly Gly Val Val Leu Val Asp Asp Thr Phe Thr Ile Lys Ser		
	565	570
Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys Gly Gly Gly Ser		

	580		585		590
Gly Gly Gly Ser Gly Gly Ser Ala Trp Ser His Pro Gln Phe Glu Lys					
	595		600		605
<210> 32					
<211> 318					
<212> PRT					
<213> 疱疹病毒 1					
<400> 32					
Thr Asp Ser Pro Gly Gly Val Ala Pro Ala Ser Pro Val Glu Asp Ala					
1	5		10		15
Ser Asp Ala Ser Leu Gly Gln Pro Glu Glu Gly Ala Pro Cys Gln Val					
	20		25		30
Val Leu Gln Gly Ala Glu Leu Asn Gly Ile Leu Gln Ala Phe Ala Pro					
	35		40		45
Leu Arg Thr Ser Leu Leu Asp Ser Leu Leu Val Met Gly Asp Arg Gly					
	50		55		60
Ile Leu Ile His Asn Thr Ile Phe Gly Glu Gln Val Phe Leu Pro Leu					
65	70		75		80
Glu His Ser Gln Phe Ser Arg Tyr Arg Trp Arg Gly Pro Thr Ala Ala					
	85		90		95
Phe Leu Ser Leu Val Asp Gln Lys Arg Ser Leu Leu Ser Val Phe Arg					
	100		105		110
Ala Asn Gln Tyr Pro Asp Leu Arg Arg Val Glu Leu Ala Ile Thr Gly					
	115		120		125
Gln Ala Pro Phe Arg Thr Leu Val Gln Arg Ile Trp Thr Thr Thr Ser					
	130		135		140
Asp Gly Glu Ala Val Glu Leu Ala Ser Glu Thr Leu Met Lys Arg Glu					
145	150		155		160
Leu Thr Ser Phe Val Val Leu Val Pro Gln Gly Thr Pro Asp Val Gln					
	165		170		175
Leu Arg Leu Thr Arg Pro Gln Leu Thr Lys Val Leu Asn Ala Thr Gly					
	180		185		190
Ala Asp Ser Ala Thr Pro Thr Thr Phe Glu Leu Gly Val Asn Gly Lys					
	195		200		205
Phe Ser Val Phe Thr Thr Ser Thr Cys Val Thr Phe Ala Ala Arg Glu					
	210		215		220
Glu Gly Val Ser Ser Ser Thr Ser Thr Gln Val Gln Ile Leu Ser Asn					
225	230		235		240
Ala Leu Thr Lys Ala Gly Gln Ala Ala Ala Asn Ala Lys Thr Val Tyr					

				245				250				255			
Gly	Glu	Asn	Thr	His	Arg	Thr	Phe	Ser	Val	Val	Val	Asp	Asp	Cys	Ser
				260				265				270			
Met	Arg	Ala	Val	Leu	Arg	Arg	Leu	Gln	Val	Gly	Gly	Gly	Thr	Leu	Lys
				275				280				285			
Phe	Phe	Leu	Thr	Thr	Pro	Val	Pro	Ser	Leu	Cys	Val	Thr	Ala	Thr	Gly
				290				295				300			
Pro	Asn	Ala	Val	Ser	Ala	Val	Phe	Leu	Leu	Lys	Pro	Gln	Lys		
305				310				315							
<210> 33															
<211> 233															
<212> PRT															
<213> 细菌噬菌体 RB69															
<400> 33															
Lys	Gly	Phe	Ser	Ser	Glu	Asp	Lys	Gly	Glu	Trp	Lys	Leu	Lys	Leu	Asp
1				5				10				15			
Ala	Ser	Gly	Asn	Gly	Gln	Ala	Val	Ile	Arg	Phe	Leu	Pro	Ala	Lys	Thr
				20				25				30			
Asp	Asp	Ala	Leu	Pro	Phe	Ala	Ile	Leu	Val	Asn	His	Gly	Phe	Lys	Lys
				35				40				45			
Asn	Gly	Lys	Trp	Tyr	Ile	Glu	Thr	Cys	Ser	Ser	Thr	His	Gly	Asp	Tyr
50				55				60							
Asp	Ser	Cys	Pro	Val	Cys	Gln	Tyr	Ile	Ser	Lys	Asn	Asp	Leu	Tyr	Asn
65				70				75				80			
Thr	Asn	Lys	Thr	Glu	Tyr	Ser	Gln	Leu	Lys	Arg	Lys	Thr	Ser	Tyr	Trp
				85				90				95			
Ala	Asn	Ile	Leu	Val	Val	Lys	Asp	Pro	Gln	Ala	Pro	Asp	Asn	Glu	Gly
				100				105				110			
Lys	Val	Phe	Lys	Tyr	Arg	Phe	Gly	Lys	Lys	Ile	Trp	Asp	Lys	Ile	Asn
				115				120				125			
Ala	Met	Ile	Ala	Val	Asp	Thr	Glu	Met	Gly	Glu	Thr	Pro	Val	Asp	Val
130				135				140							
Thr	Cys	Pro	Trp	Glu	Gly	Ala	Asn	Phe	Val	Leu	Lys	Val	Lys	Gln	Val
145				150				155				160			
Ser	Gly	Phe	Ser	Asn	Tyr	Asp	Glu	Ser	Lys	Phe	Leu	Asn	Gln	Ser	Ala
				165				170				175			
Ile	Pro	Asn	Ile	Asp	Asp	Glu	Ser	Phe	Gln	Lys	Glu	Leu	Phe	Glu	Gln
				180				185				190			
Met	Val	Asp	Leu	Ser	Glu	Met	Thr	Ser	Lys	Asp	Lys	Phe	Lys	Ser	Phe

195	200	205
Glu Glu Leu Asn Thr Lys Phe Asn Gln Val Leu Gly Thr Ala Ala Leu		
210	215	220
Gly Gly Ala Ala Ala Ala Ala Ser		
225	230	
<210> 34		
<211> 210		
<212> PRT		
<213> 细菌噬菌体 T7		
<400> 34		
Ala Lys Lys Ile Phe Thr Ser Ala Leu Gly Thr Ala Glu Pro Tyr Ala		
1	5	10
Tyr Ile Ala Lys Pro Asp Tyr Gly Asn Glu Glu Arg Gly Phe Gly Asn		
20	25	30
Pro Arg Gly Val Tyr Lys Val Asp Leu Thr Ile Pro Asn Lys Asp Pro		
35	40	45
Arg Cys Gln Arg Met Val Asp Glu Ile Val Lys Cys His Glu Glu Ala		
50	55	60
Tyr Ala Ala Ala Val Glu Glu Tyr Glu Ala Asn Pro Pro Ala Val Ala		
65	70	75
Arg Gly Lys Lys Pro Leu Lys Pro Tyr Glu Gly Asp Met Pro Phe Phe		
85	90	95
Asp Asn Gly Asp Gly Thr Thr Thr Phe Lys Phe Lys Cys Tyr Ala Ser		
100	105	110
Phe Gln Asp Lys Lys Thr Lys Glu Thr Lys His Ile Asn Leu Val Val		
115	120	125
Val Asp Ser Lys Gly Lys Lys Met Glu Asp Val Pro Ile Ile Gly Gly		
130	135	140
Gly Ser Lys Leu Lys Val Lys Tyr Ser Leu Val Pro Tyr Lys Trp Asn		
145	150	155
Thr Ala Val Gly Ala Ser Val Lys Leu Gln Leu Glu Ser Val Met Leu		
165	170	175
Val Glu Leu Ala Thr Phe Gly Gly Gly Glu Asp Asp Trp Ala Asp Glu		
180	185	190
Val Glu Glu Asn Gly Tyr Val Ala Ser Gly Ser Ala Lys Ala Ser Lys		
195	200	205
Pro Arg		
210		
<210> 35		

<211> 99

<212> PRT

<213> 嗜冷嗜盐菌(*Halorubrum lacusprofundi*)

<400> 35

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

Asp Phe Arg

<210> 36

<211> 102

<212> PRT

<213> 富盐菌(*Haloferax volcanii*)

<400> 36

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

Ser Leu Gly Asp Phe Glu

100

<210> 37

<211> 55

<212> PRT

<213> 人工序列

<220>

<223> (HhH) 2 域.

<400> 37

```

Trp Lys Glu Trp Leu Glu Arg Lys Val Gly Glu Gly Arg Ala Arg Arg
1           5           10           15
Leu Ile Glu Tyr Phe Gly Ser Ala Gly Glu Val Gly Lys Leu Val Glu
           20           25           30
Asn Ala Glu Val Ser Lys Leu Leu Glu Val Pro Gly Ile Gly Asp Glu
           35           40           45
Ala Val Ala Arg Leu Val Pro
           50           55

```

<210> 38

<211> 107

<212> PRT

<213> 人工序列

<220>

<223> (HhH) 2-(HhH) 2 域.

<400> 38

```

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

```

<210> 39

<211> 132

<212> PRT

<213> 智人(Homo sapiens)

<400> 39

```

Glu Ser Glu Thr Thr Thr Ser Leu Val Leu Glu Arg Ser Leu Asn Arg

```

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

<210> 40

<211> 123

<212> PRT

<213> 枯草芽孢杆菌噬菌体(Bacillus subtilis phage)phi29

<400> 40

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

<210> 41

<211> 177

<212> PRT

<213> 大肠杆菌(Escherichia coli)

<400> 41

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

Phe

<210> 42

<211> 301

<212> PRT

<213> 肠杆菌噬菌体T4

<400> 42

Met	Phe	Lys	Arg	Lys	Ser	Thr	Ala	Glu	Leu	Ala	Ala	Gln	Met	Ala	Lys
1				5					10					15	
Leu	Asn	Gly	Asn	Lys	Gly	Phe	Ser	Ser	Glu	Asp	Lys	Gly	Glu	Trp	Lys
			20					25					30		
Leu	Lys	Leu	Asp	Asn	Ala	Gly	Asn	Gly	Gln	Ala	Val	Ile	Arg	Phe	Leu
		35				40					45				
Pro	Ser	Lys	Asn	Asp	Glu	Gln	Ala	Pro	Phe	Ala	Ile	Leu	Val	Asn	His

50	55	60
Gly Phe Lys Lys Asn Gly Lys Trp Tyr Ile Glu Thr Cys Ser Ser Thr		
65	70	75
His Gly Asp Tyr Asp Ser Cys Pro Val Cys Gln Tyr Ile Ser Lys Asn		80
85	90	95
Asp Leu Tyr Asn Thr Asp Asn Lys Glu Tyr Ser Leu Val Lys Arg Lys		
100	105	110
Thr Ser Tyr Trp Ala Asn Ile Leu Val Val Lys Asp Pro Ala Ala Pro		
115	120	125
Glu Asn Glu Gly Lys Val Phe Lys Tyr Arg Phe Gly Lys Lys Ile Trp		
130	135	140
Asp Lys Ile Asn Ala Met Ile Ala Val Asp Val Glu Met Gly Glu Thr		
145	150	155
Pro Val Asp Val Thr Cys Pro Trp Glu Gly Ala Asn Phe Val Leu Lys		
165	170	175
Val Lys Gln Val Ser Gly Phe Ser Asn Tyr Asp Glu Ser Lys Phe Leu		
180	185	190
Asn Gln Ser Ala Ile Pro Asn Ile Asp Asp Glu Ser Phe Gln Lys Glu		
195	200	205
Leu Phe Glu Gln Met Val Asp Leu Ser Glu Met Thr Ser Lys Asp Lys		
210	215	220
Phe Lys Ser Phe Glu Glu Leu Asn Thr Lys Phe Gly Gln Val Met Gly		
225	230	235
Thr Ala Val Met Gly Gly Ala Ala Ala Thr Ala Ala Lys Lys Ala Asp		
245	250	255
Lys Val Ala Asp Asp Leu Asp Ala Phe Asn Val Asp Asp Phe Asn Thr		
260	265	270
Lys Thr Glu Asp Asp Phe Met Ser Ser Ser Ser Gly Ser Ser Ser Ser		
275	280	285
Ala Asp Asp Thr Asp Leu Asp Asp Leu Leu Asn Asp Leu		
290	295	300
<210> 43		
<211> 177		
<212> PRT		
<213> 人工序列		
<220>		
<223> EcoSSB-CterAla		
<400> 43		
Ala Ser Arg Gly Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly Gln		

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

Phe

<210> 44

<211> 177

<212> PRT

<213> 人工序列

<220>

<223> EcoSSB-CterNGGN

<400> 44

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

Gly	Gln	Leu	Arg	Thr	Arg	Lys	Trp	Thr	Asp	Gln	Ser	Gly	Gln	Asp	Arg		
85								90				95					
Tyr	Thr	Thr	Glu	Val	Val	Val	Asn	Val	Gly	Gly	Thr	Met	Gln	Met	Leu		
100								105				110					
Gly	Gly	Arg	Gln	Gly	Gly	Gly	Ala	Pro	Ala	Gly	Gly	Asn	Ile	Gly	Gly		
115								120				125					
Gly	Gln	Pro	Gln	Gly	Gly	Trp	Gly	Gln	Pro	Gln	Gln	Pro	Gln	Gly	Gly		
130								135				140					
Asn	Gln	Phe	Ser	Gly	Gly	Ala	Gln	Ser	Arg	Pro	Gln	Gln	Ser	Ala	Pro		
145								150				155				160	
Ala	Ala	Pro	Ser	Asn	Glu	Pro	Pro	Met	Asn	Phe	Gly	Gly	Asn	Ile	Pro		
165								170				175					

Phe

<210> 45

<211> 152

<212> PRT

<213> 人工序列

<220>

<223> EcoSSB-Q152del

<400> 45

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

145	150
<210> 46	
<211> 117	
<212> PRT	
<213> 人工序列	
<220>	
<223> EcoSSB-G117del	
<400> 46	
Ala Ser Arg Gly Val Asn Lys Val Ile Leu Val Gly Asn Leu Gly Gln	
1 5 10 15	
Asp Pro Glu Val Arg Tyr Met Pro Asn Gly Gly Ala Val Ala Asn Ile	
20 25 30	
Thr Leu Ala Thr Ser Glu Ser Trp Arg Asp Lys Ala Thr Gly Glu Met	
35 40 45	
Lys Glu Gln Thr Glu Trp His Arg Val Val Leu Phe Gly Lys Leu Ala	
50 55 60	
Glu Val Ala Ser Glu Tyr Leu Arg Lys Gly Ser Gln Val Tyr Ile Glu	
65 70 75 80	
Gly Gln Leu Arg Thr Arg Lys Trp Thr Asp Gln Ser Gly Gln Asp Arg	
85 90 95	
Tyr Thr Thr Glu Val Val Val Asn Val Gly Gly Thr Met Gln Met Leu	
100 105 110	
Gly Gly Arg Gln Gly	
115	
<210> 47	
<211> 984	
<212> PRT	
<213> 如甲烷嗜热菌 (Methanopyrus kandleri)	
<400> 47	
Met Ala Leu Val Tyr Asp Ala Glu Phe Val Gly Ser Glu Arg Glu Phe	
1 5 10 15	
Glu Glu Glu Arg Glu Thr Phe Leu Lys Gly Val Lys Ala Tyr Asp Gly	
20 25 30	
Val Leu Ala Thr Arg Tyr Leu Met Glu Arg Ser Ser Ser Ala Lys Asn	
35 40 45	
Asp Glu Glu Leu Leu Glu Leu His Gln Asn Phe Ile Leu Leu Thr Gly	
50 55 60	
Ser Tyr Ala Cys Ser Ile Asp Pro Thr Glu Asp Arg Tyr Gln Asn Val	
65 70 75 80	

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

385	390	395	400
Ala Tyr Arg Ala Ala Val Glu Ile Gln Leu Ala Glu Leu Thr Lys Lys			
	405	410	415
Glu Gly Val Gly Arg Lys Thr Ala Glu Arg Leu Leu Arg Ala Phe Gly			
	420	425	430
Asn Pro Glu Arg Val Lys Gln Leu Ala Arg Glu Phe Glu Ile Glu Lys			
	435	440	445
Leu Ala Ser Val Glu Gly Val Gly Glu Arg Val Leu Arg Ser Leu Val			
	450	455	460
Pro Gly Tyr Ala Ser Leu Ile Ser Ile Arg Gly Ile Asp Arg Glu Arg			
465	470	475	480
Ala Glu Arg Leu Leu Lys Lys Tyr Gly Gly Tyr Ser Lys Val Arg Glu			
	485	490	495
Ala Gly Val Glu Glu Leu Arg Glu Asp Gly Leu Thr Asp Ala Gln Ile			
	500	505	510
Arg Glu Leu Lys Gly Leu Lys Thr Leu Glu Ser Ile Val Gly Asp Leu			
	515	520	525
Glu Lys Ala Asp Glu Leu Lys Arg Lys Tyr Gly Ser Ala Ser Ala Val			
	530	535	540
Arg Arg Leu Pro Val Glu Glu Leu Arg Glu Leu Gly Phe Ser Asp Asp			
545	550	555	560
Glu Ile Ala Glu Ile Lys Gly Ile Pro Lys Lys Leu Arg Glu Ala Phe			
	565	570	575
Asp Leu Glu Thr Ala Ala Glu Leu Tyr Glu Arg Tyr Gly Ser Leu Lys			
	580	585	590
Glu Ile Gly Arg Arg Leu Ser Tyr Asp Asp Leu Leu Glu Leu Gly Ala			
	595	600	605
Thr Pro Lys Ala Ala Ala Glu Ile Lys Gly Pro Glu Phe Lys Phe Leu			
	610	615	620
Leu Asn Ile Glu Gly Val Gly Pro Lys Leu Ala Glu Arg Ile Leu Glu			
625	630	635	640
Ala Val Asp Tyr Asp Leu Glu Arg Leu Ala Ser Leu Asn Pro Glu Glu			
	645	650	655
Leu Ala Glu Lys Val Glu Gly Leu Gly Glu Glu Leu Ala Glu Arg Val			
	660	665	670
Val Tyr Ala Ala Arg Glu Arg Val Glu Ser Arg Arg Lys Ser Gly Arg			
	675	680	685
Gln Glu Arg Ser Glu Glu Glu Trp Lys Glu Trp Leu Glu Arg Lys Val			
	690	695	700

Gly	Glu	Gly	Arg	Ala	Arg	Arg	Leu	Ile	Glu	Tyr	Phe	Gly	Ser	Ala	Gly
705							710				715				720
Glu	Val	Gly	Lys	Leu	Val	Glu	Asn	Ala	Glu	Val	Ser	Lys	Leu	Leu	Glu
			725						730						735
Val	Pro	Gly	Ile	Gly	Asp	Glu	Ala	Val	Ala	Arg	Leu	Val	Pro	Gly	Tyr
			740						745						750
Lys	Thr	Leu	Arg	Asp	Ala	Gly	Leu	Thr	Pro	Ala	Glu	Ala	Glu	Arg	Val
			755						760						765
Leu	Lys	Arg	Tyr	Gly	Ser	Val	Ser	Lys	Val	Gln	Glu	Gly	Ala	Thr	Pro
			770						775						780
Asp	Glu	Leu	Arg	Glu	Leu	Gly	Leu	Gly	Asp	Ala	Lys	Ile	Ala	Arg	Ile
785							790				795				800
Leu	Gly	Leu	Arg	Ser	Leu	Val	Asn	Lys	Arg	Leu	Asp	Val	Asp	Thr	Ala
							805				810				815
Tyr	Glu	Leu	Lys	Arg	Arg	Tyr	Gly	Ser	Val	Ser	Ala	Val	Arg	Lys	Ala
							820				825				830
Pro	Val	Lys	Glu	Leu	Arg	Glu	Leu	Gly	Leu	Ser	Asp	Arg	Lys	Ile	Ala
			835						840						845
Arg	Ile	Lys	Gly	Ile	Pro	Glu	Thr	Met	Leu	Gln	Val	Arg	Gly	Met	Ser
			850						855						860
Val	Glu	Lys	Ala	Glu	Arg	Leu	Leu	Glu	Arg	Phe	Asp	Thr	Trp	Thr	Lys
865							870				875				880
Val	Lys	Glu	Ala	Pro	Val	Ser	Glu	Leu	Val	Arg	Val	Pro	Gly	Val	Gly
							885				890				895
Leu	Ser	Leu	Val	Lys	Glu	Ile	Lys	Ala	Gln	Val	Asp	Pro	Ala	Trp	Lys
							900				905				910
Ala	Leu	Leu	Asp	Val	Lys	Gly	Val	Ser	Pro	Glu	Leu	Ala	Asp	Arg	Leu
							915				920				925
Val	Glu	Glu	Leu	Gly	Ser	Pro	Tyr	Arg	Val	Leu	Thr	Ala	Lys	Lys	Ser
							930				935				940
Asp	Leu	Met	Arg	Val	Glu	Arg	Val	Gly	Pro	Lys	Leu	Ala	Glu	Arg	Ile
945							950				955				960
Arg	Ala	Ala	Gly	Lys	Arg	Tyr	Val	Glu	Glu	Arg	Arg	Ser	Arg	Arg	Glu
							965				970				975
Arg	Ile	Arg	Arg	Lys	Leu	Arg	Gly								
							980								

<210> 48

<211> 299

<212> PRT

<213> 如甲烷嗜热菌

<400> 48

```

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

```

290	295
<210> 49	
<211> 853	
<212> PRT	
<213> 大肠杆菌	
<400> 49	
Met Ser Ala Ile Glu Asn Phe Asp Ala His Thr Pro Met Met Gln Gln	
1 5 10 15	
Tyr Leu Arg Leu Lys Ala Gln His Pro Glu Ile Leu Leu Phe Tyr Arg	
20 25 30	
Met Gly Asp Phe Tyr Glu Leu Phe Tyr Asp Asp Ala Lys Arg Ala Ser	
35 40 45	
Gln Leu Leu Asp Ile Ser Leu Thr Lys Arg Gly Ala Ser Ala Gly Glu	
50 55 60	
Pro Ile Pro Met Ala Gly Ile Pro Tyr His Ala Val Glu Asn Tyr Leu	
65 70 75 80	
Ala Lys Leu Val Asn Gln Gly Glu Ser Val Ala Ile Cys Glu Gln Ile	
85 90 95	
Gly Asp Pro Ala Thr Ser Lys Gly Pro Val Glu Arg Lys Val Val Arg	
100 105 110	
Ile Val Thr Pro Gly Thr Ile Ser Asp Glu Ala Leu Leu Gln Glu Arg	
115 120 125	
Gln Asp Asn Leu Leu Ala Ala Ile Trp Gln Asp Ser Lys Gly Phe Gly	
130 135 140	
Tyr Ala Thr Leu Asp Ile Ser Ser Gly Arg Phe Arg Leu Ser Glu Pro	
145 150 155 160	
Ala Asp Arg Glu Thr Met Ala Ala Glu Leu Gln Arg Thr Asn Pro Ala	
165 170 175	
Glu Leu Leu Tyr Ala Glu Asp Phe Ala Glu Met Ser Leu Ile Glu Gly	
180 185 190	
Arg Arg Gly Leu Arg Arg Arg Pro Leu Trp Glu Phe Glu Ile Asp Thr	
195 200 205	
Ala Arg Gln Gln Leu Asn Leu Gln Phe Gly Thr Arg Asp Leu Val Gly	
210 215 220	
Phe Gly Val Glu Asn Ala Pro Arg Gly Leu Cys Ala Ala Gly Cys Leu	
225 230 235 240	
Leu Gln Tyr Ala Lys Asp Thr Gln Arg Thr Thr Leu Pro His Ile Arg	
245 250 255	
Ser Ile Thr Met Glu Arg Glu Gln Asp Ser Ile Ile Met Asp Ala Ala	

260	265	270
Thr Arg Arg Asn Leu Glu Ile Thr	Gln Asn Leu Ala Gly Gly Ala Glu	
275	280	285
Asn Thr Leu Ala Ser Val Leu Asp	Cys Thr Val Thr Pro Met Gly Ser	
290	295	300
Arg Met Leu Lys Arg Trp Leu His	Met Pro Val Arg Asp Thr Arg Val	
305	310	315
Leu Leu Glu Arg Gln Gln Thr Ile	Gly Ala Leu Gln Asp Phe Thr Ala	
325	330	335
Gly Leu Gln Pro Val Leu Arg Gln	Val Gly Asp Leu Glu Arg Ile Leu	
340	345	350
Ala Arg Leu Ala Leu Arg Thr Ala	Arg Pro Arg Asp Leu Ala Arg Met	
355	360	365
Arg His Ala Phe Gln Gln Leu Pro	Glu Leu Arg Ala Gln Leu Glu Thr	
370	375	380
Val Asp Ser Ala Pro Val Gln Ala	Leu Arg Glu Lys Met Gly Glu Phe	
385	390	395
Ala Glu Leu Arg Asp Leu Leu Glu	Arg Ala Ile Ile Asp Thr Pro Pro	
405	410	415
Val Leu Val Arg Asp Gly Gly Val	Ile Ala Ser Gly Tyr Asn Glu Glu	
420	425	430
Leu Asp Glu Trp Arg Ala Leu Ala	Asp Gly Ala Thr Asp Tyr Leu Glu	
435	440	445
Arg Leu Glu Val Arg Glu Arg Glu	Arg Thr Gly Leu Asp Thr Leu Lys	
450	455	460
Val Gly Phe Asn Ala Val His Gly	Tyr Tyr Ile Gln Ile Ser Arg Gly	
465	470	475
Gln Ser His Leu Ala Pro Ile Asn	Tyr Met Arg Arg Gln Thr Leu Lys	
485	490	495
Asn Ala Glu Arg Tyr Ile Ile Pro	Glu Leu Lys Glu Tyr Glu Asp Lys	
500	505	510
Val Leu Thr Ser Lys Gly Lys Ala	Leu Ala Leu Glu Lys Gln Leu Tyr	
515	520	525
Glu Glu Leu Phe Asp Leu Leu Leu	Pro His Leu Glu Ala Leu Gln Gln	
530	535	540
Ser Ala Ser Ala Leu Ala Glu Leu	Asp Val Leu Val Asn Leu Ala Glu	
545	550	555
Arg Ala Tyr Thr Leu Asn Tyr Thr	Cys Pro Thr Phe Ile Asp Lys Pro	
565	570	575

Gly Ile Arg Ile Thr Glu Gly Arg His Pro Val Val Glu Gln Val Leu		
580	585	590
Asn Glu Pro Phe Ile Ala Asn Pro Leu Asn Leu Ser Pro Gln Arg Arg		
595	600	605
Met Leu Ile Ile Thr Gly Pro Asn Met Gly Gly Lys Ser Thr Tyr Met		
610	615	620
Arg Gln Thr Ala Leu Ile Ala Leu Met Ala Tyr Ile Gly Ser Tyr Val		
625	630	635
Pro Ala Gln Lys Val Glu Ile Gly Pro Ile Asp Arg Ile Phe Thr Arg		
645	650	655
Val Gly Ala Ala Asp Asp Leu Ala Ser Gly Arg Ser Thr Phe Met Val		
660	665	670
Glu Met Thr Glu Thr Ala Asn Ile Leu His Asn Ala Thr Glu Tyr Ser		
675	680	685
Leu Val Leu Met Asp Glu Ile Gly Arg Gly Thr Ser Thr Tyr Asp Gly		
690	695	700
Leu Ser Leu Ala Trp Ala Cys Ala Glu Asn Leu Ala Asn Lys Ile Lys		
705	710	715
Ala Leu Thr Leu Phe Ala Thr His Tyr Phe Glu Leu Thr Gln Leu Pro		
725	730	735
Glu Lys Met Glu Gly Val Ala Asn Val His Leu Asp Ala Leu Glu His		
740	745	750
Gly Asp Thr Ile Ala Phe Met His Ser Val Gln Asp Gly Ala Ala Ser		
755	760	765
Lys Ser Tyr Gly Leu Ala Val Ala Ala Leu Ala Gly Val Pro Lys Glu		
770	775	780
Val Ile Lys Arg Ala Arg Gln Lys Leu Arg Glu Leu Glu Ser Ile Ser		
785	790	795
Pro Asn Ala Ala Ala Thr Gln Val Asp Gly Thr Gln Met Ser Leu Leu		
805	810	815
Ser Val Pro Glu Glu Thr Ser Pro Ala Val Glu Ala Leu Glu Asn Leu		
820	825	830
Asp Pro Asp Ser Leu Thr Pro Arg Gln Ala Leu Glu Trp Ile Tyr Arg		
835	840	845
Leu Lys Ser Leu Val		
850		
<210> 50		
<211> 64		
<212> PRT		

<213> 硫磺矿硫化叶菌 (Sulfolobus solfataricus)

<400> 50

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

<210> 51

<211> 99

<212> PRT

<213> 硫磺矿硫化叶菌P2 (Sulfolobus solfataricus P2)

<400> 51

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

Arg Lys Lys

<210> 52

<211> 88

<212> PRT

<213> 硫磺矿硫化叶菌P2

<400> 52

```
Thr Glu Lys Leu Asn Glu Ile Val Val Arg Lys Thr Lys Asn Val Glu
1           5           10           15
Asp His Val Leu Asp Val Ile Val Leu Phe Asn Gln Gly Ile Asp Glu
           20           25           30
Val Ile Leu Lys Gly Thr Gly Arg Glu Ile Ser Lys Ala Val Asp Val
           35           40           45
```

Tyr Asn Ser Leu Lys Asp Arg Leu Gly Asp Gly Val Gln Leu Val Asn
 50 55 60
 Val Gln Thr Gly Ser Glu Val Arg Asp Arg Arg Arg Ile Ser Tyr Ile
 65 70 75 80
 Leu Leu Arg Leu Lys Arg Val Tyr
 85

<210> 53

<211> 107

<212> PRT

<213> 大肠杆菌

<400> 53

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

<210> 54

<211> 237

<212> PRT

<213> 肠道菌噬菌体λ(Enterobacteria phage lambda)

<400> 54

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

65	70	75	80
Ile Ala Arg Glu Ile Tyr Glu Met Tyr Glu Ala Val Ser Met Gln Pro			
	85	90	95
Ser Leu Arg Ser Glu Tyr Glu Tyr Pro Val Phe Ser His Val Gln Ala			
	100	105	110
Gly Met Phe Ser Pro Glu Leu Arg Thr Phe Thr Lys Gly Asp Ala Glu			
	115	120	125
Arg Trp Val Ser Thr Thr Lys Lys Ala Ser Asp Ser Ala Phe Trp Leu			
	130	135	140
Glu Val Glu Gly Asn Ser Met Thr Ala Pro Thr Gly Ser Lys Pro Ser			
	145	150	155
Phe Pro Asp Gly Met Leu Ile Leu Val Asp Pro Glu Gln Ala Val Glu			
	165	170	175
Pro Gly Asp Phe Cys Ile Ala Arg Leu Gly Gly Asp Glu Phe Thr Phe			
	180	185	190
Lys Lys Leu Ile Arg Asp Ser Gly Gln Val Phe Leu Gln Pro Leu Asn			
	195	200	205
Pro Gln Tyr Pro Met Ile Pro Cys Asn Glu Ser Cys Ser Val Val Gly			
	210	215	220
Lys Val Ile Ala Ser Gln Trp Pro Glu Glu Thr Phe Gly			
	225	230	235

<210> 55

<211> 60

<212> PRT

<213> 中的泉古菌(Crenarchaea)

<400> 55

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

<210> 56

<211> 136

<212> PRT

<213> 智人

<400> 56

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

<210> 57

<211> 89

<212> PRT

<213> 肠杆菌噬菌体T4

<400> 57

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

<210> 58

<211> 339

<212> PRT

<213> 智人

<400> 58

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

164

225	230	235	240
Leu Thr Ser Lys Leu Ile Asp Gly Thr Phe Pro Asp Tyr Ser Arg Val			
	245	250	255
Ile Pro Thr Gly Asn Asp Lys Leu Leu Arg Leu Asp Pro Lys Ala Phe			
	260	265	270
Phe Gln Gly Val Asp Arg Val Ala Thr Ile Ala Thr Glu Lys Thr Arg			
	275	280	285
Ala Val Lys Met Gly Leu Asp Glu Asp Lys Val Thr Leu Ser Val Thr			
	290	295	300
Ser Pro Asp Asn Gly Thr Ala Ala Glu Glu Ile Ala Ala Glu Tyr Lys			
305	310	315	320
Ala Glu Gly Phe Glu Ile Gly Phe Asn Ala Asn Tyr Leu Lys Asp Ile			
	325	330	335
Leu Gly Gln Ile Asp Ser Asp Thr Val Glu Leu His Leu Ala Asp Ala			
	340	345	350
Gly Ala Pro Thr Leu Ile Arg Arg Asp Glu Asn Ser Pro Ala Leu Tyr			
	355	360	365
Val Leu Met Pro Met Arg Val			
	370	375	

<210> 60

<211> 12

<212> DNA

<213> 人工序列

<220>

<223> 实施例3中使用的序列.

<400> 60

tttttttttt tt 12

<210> 61

<211> 3667

<212> DNA

<213> 人工序列

<220>

<223> 实施例3中使用的序列.

<400> 61

```

ggttgtttct gttggtgctg atattgcggc gtctgcttgg gtgtttaacc tgccatcaga 60
tttggtttgt tagtcgcttt ttttttttgg aatttttttt ttggaatttt ttttttgcgc 120
taacaacctc ctgccgtttt gcccgtgcat atcggtcacg aacaaatctg attactaaac 180
acagtagcct ggatttggtc tatcagtaat cgaccttatt cctaattaaa tagagcaaat 240
ccccttattg ggggtaagac atgaagatgc cagaaaaaca tgacctgttg gccgccattc 300

```

```

tcgcggcaaa ggaacaaggc atcggggcaa tccttgcggt tgcaatggcg taccttcgcg 360
gcagatataa tggcgggtgcg tttaaaaaa cagtaatcga cgcaacgatg tgcgccatta 420
tcgcctagtt cattcgtgac cttctcgact tcgccggact aagtagcaat ctgcgttata 480
taacgagcgt gtttatcggc tacatcggta ctgactcgat tggttcgctt atcaaacgct 540
tcgctgctaa aaaagccgga gtagaagatg gtagaaatca ataatcaacg taaggcgctt 600
ctcgatatgc tggcgtggtc ggagggaact gataacggac gtcagaaaac cagaaatcat 660
ggttatgacg tcattgtagg cggagagcta ttactgatt actccgatca ccctcgcaa 720
cttgtcacgc taaacccaaa actcaaatca acaggcgccg gacgctacca gcttctttcc 780
cggttggtggg atgcctaccg caagcagctt ggctgaaag acttctctcc gaaaagtcag 840
gacgctgtgg cattgcagca gattaaggag cgtggcgctt tacctatgat tgatcgtggt 900
gatatccgtc aggcaatcga ccgttgacgc aatatctggg ctactgcc gggcgctggt 960
tatggtcagt tcgagcataa ggctgacagc ctgattgcaa aattcaaaga agcgggaggga 1020
acggtcagag agattgatgt atgagcagag tcaccgcgat tatctccgt ctggttatct 1080
gcatcatcgt ctgcctgtca tgggtgttta atcattaccg tgataacgcc attacctaca 1140
aagcccagcg cgacaaaaat gccagagaac tgaagctggc gaacgcggca attactgaca 1200
tgcagatgcg tcagcgtgat gttgtgcgc tcgatgcaa atacacgaag gagttagctg 1260
atgctaaagc tgaaaatgat gctctgcgtg atgatgttg cgctggctgt cgtcggttgc 1320
acatcaaagc agtctgtcag tcagtgcgtg aagccaccac cgcctccggc gtggataatg 1380
cagcctccc ccgactggca gacaccgtg aacgggatta ttaccctc agagagaggc 1440
tgatcactat gcaaaaacaa ctggaaggaa ccagaagta tattaatgag cagtgcagat 1500
agagttgccc atatcgatgg gcaactcatg caattattgt gagcaataca cacgcgttc 1560
cagcggagta taaatgccta aagtaataaa accgagcaat ccatttacga atgtttgctg 1620
ggtttctgtt ttaacaacat tttctgcgc gccacaaatt ttggctgcat cgacagtttt 1680
cttctgccc attccagaaa cgaagaaatg atgggtgatg gtttccttg gtgctactgc 1740
tgccggttg ttttgaacag taaacgtctg ttgagcat cctgtaataa gcagggccag 1800
cgcagtagcg agtagcattt tttcatggt gttattccc atgctttttg aagttcgcag 1860
aatcgatatg gtagaaaatt aaacaaccc taaacaatga gttgaaatt catattgtta 1920
atatttatta atgtatgtca ggtgcgatga atcgtcattg tattcccga ttaactatgt 1980
ccacagccct gacggggaac ttctctgcgg gagtgtccgg gaataattaa aacgatgcac 2040
acagggttta gcgcgtacac gtattgcatt atgccaacgc cccggtgctg acacggaaga 2100
aaccggacgt tatgatttag cgtggaaaga tttgttagt gttctgaatg ctctcagtaa 2160
atagtaatga attatcaaag gtatagtaat atcttttatg ttcatggata tttgtaacct 2220
atcggaaaac tcctgctta gcaagatttt ccctgtattg ctgaaatgtg atttctcttg 2280
atttcaacct atcataggac gtttctataa gatgcgtgtt tcttgagaat ttaacattta 2340
caacctttt aagtcctttt attaacacgg tgttatcgtt ttctaacacg atgtgaatat 2400
tatctgtggc tagatagtaa atataatgtg agacgttgtg acgttttagt tcagaataaa 2460
acaattcaca gtctaaatct tttcgcaatt gatcgaatat ttctttaaaa atggcaacct 2520
gagccattgg taaaacctc catgtgatac gagggcgct agtttgcatt atcgttttta 2580
tcgtttcaat ctggtctgac ctcttgtgt tttgttgatg atttatgtca aatattagga 2640

```

```

atgttttcac ttaatagtat tggttgcgta acaaagtgcg gtcctgctgg cattctggag 2700
ggaaatacaa cgcacagatg tatgtaaggc caacgtgctc aaatcttcat acagaaagat 2760
ttgaagtaat attttaaccg ctagatgaag agcaagcgca tggagcgaca aaatgaataa 2820
agaacaatct gctgatgata cctccgtgga tctgattcgt gtaaaaaata tgcttaatag 2880
caccatttct atgagttacc ctgatgttgt aattgcatgt atagaacata aggtgtctct 2940
ggaagcattc agagcaattg aggcagcggt ggtgaagcac gataataata tgaaggatta 3000
ttccctgggtg gttgactgat caccataact gctaatacatt caaactatct agtctgtgac 3060
agagccaaca cgcagtcctgt cactgtcagg aaagtggtaa aactgcaact caattactgc 3120
aatgccctcg taattaagtg aatttacaat atcgtcctgt tcggaggga gaacgcggga 3180
tggtcattct tcatcacttt taattgatgt atatgctctc tttctgacg ttagtctccg 3240
acggcagggt tcaatgaccc aggtgagaa attcccggac cttttttgct caagagcgat 3300
gttaatttgt tcaatcattt ggtaggaaa gcggatgttg cgggttggtt ttctgcgggt 3360
tctgttcttc gttgacatga ggttgccccg tattcagtgt cgtgatttg tattgtctga 3420
agttgttttt acgttaagtt gatgcagatc aattaatacg atacctgcgt cataattgat 3480
tatttgacgt ggttgatgg cctccacgca cgttgtgata ttagatgat aatcattatc 3540
actttacggg tcctttccgg tgaaaaaaaa ggtacaaaaa aaaacatcgt cgtgagtagt 3600
gaaccgtaag catgtaggac gtcctgtcgc tgtgtctcgg aactgattg acacggttta 3660
gtagagc

```

<210> 62

<211> 3617

<212> DNA

<213> 人工序列

<220>

<223> 实施例3中使用的序列.

<400> 62

```

tttttttttt tttttttttt ttttttttcg agacacagcg acaggacgtc ctacatgctt 60
acggttcaact actcacgacg atgttttttt tggtagcttt tttttcaccg gaaaggaccc 120
gtaaagtgat aatgattatc atctacatat cacaacgtgc gtggaggcca tcaaacacg 180
tcaaataatc aattatgacg caggtatcgt attaattgat ctgcatcaac ttaacgtaaa 240
aacaacttca gacaatacaa atcagcgaca ctgaatacgg ggaacctca tgtcaacgaa 300
gaacagaacc cgcagaacaa caaccgcaa catecgcttt cctaaccaaa tgattgaaca 360
aattaacatc gctcttgagc aaaaagggtc cgggaatttc tcagcctggg tcattgaagc 420
ctgccgtcgg agactaacgt cagaaaagag agcatataca tcaattaaaa gtgatgaaga 480
atgaacatcc cgcgttcttc cctccgaaca ggacgatatt gtaaattcac ttaattacga 540
gggcattgca gtaattgagt tgcagtttta ccaatttctt gacagtgaca gactgcgtgt 600
tggtctctgtc acagactaaa tagtttgaat gattagcagt tatggtgatc agtcaaccac 660
caggaataaa tccttcatat tattatcgtg cttcaccaac gctgcctcaa ttgctctgaa 720
tgcttcacga gacaccttat gttctataca tgcaattaca acatcagggt aactcataga 780
aatggtgcta ttaagcatat tttttacacg aatcagatcc acggagggat catcagcaga 840

```

```

ttgttcttta ttcatTTTgt cgtccatgc gcttgcTctt catctagcgg ttaaaatatt 900
acttcaaate tttctgtatg aagatttgag cacgttggcc ttacatacat ctgtcggTtg 960
tatttccctc cagaatgcc a gcaggaccgc actttgttac gcaaccaata ctattaagtg 1020
aaaacattcc taatatttga cataaatcat caacaaaaca caaggaggTc agaccagatt 1080
gaaacgataa aaacgataat gcaaactacg cgccctcgta tcacatggaa ggTTTTacca 1140
atggctcagg ttgccatttt taaagaaata ttcgatcaag tgcgaaaaga tttagactgt 1200
gaattgtttt attctgaact aaaacgtcac aacgtctcac attatattta ctatctagcc 1260
acagataata ttcacatcgt gttagaaaac gataacaccg tgTTaataaa aggacttaaa 1320
aaggttgtaa atgtTaaatt ctcaagaaac acgcattctta tagaaacgtc ctatgatagg 1380
ttgaaatcaa gagaaatcac atttcagcaa tacaggga aa atcttgctaa agcaggagtt 1440
ttccgatggg ttacaaatat ccatgaacat aaaagatatt actatacctt tgataattca 1500
ttactattta ctgagagcat tcagaacact acacaaatct tTccacgcta aatcataacg 1560
tccggtttct tccgtgtcag caccggggcg ttggcataat gcaatacgtg tacgcgctaa 1620
accctgtgtg catcgtttta attattcccg gacactcccg cagagaagtt ccccgTcagg 1680
gctgtggaca tagTTaatcc gggaatacaa tgacgattca tcgcacctga catacattaa 1740
taaataattaa caatatgaaa tttcaactca ttgtttaggg tttgtTTaat tttctacaca 1800
tacgattctg cgaacttcaa aaagcatcgg gaataacacc atgaaaaaaa tgctactcg 1860
tactgcgctg gccctgctta ttacaggatg tgctcaacag acgtttactg ttcaaaacaa 1920
accggcagca gtagcaccaa aggaaaccat caccatcat ttcttcgttt ctggaattgg 1980
gcagaagaaa actgtcgatg cagccaaaat ttgtggcggc gcagaaaatg ttgtTaaac 2040
agaaaccag caaacattcg taaatggatt gctcggTTTT attactttag gcatttatac 2100
tccgctggaa gcgcgtgtgt attgctcaca ataattgcat gagttgcca tcgatatggg 2160
caactctate tgcactgctc attaatatac ttctgggttc cttccagttg tttttgcata 2220
gtgatcagcc tctctctgag ggtgaaataa tcccgTtcag cggTgtctgc cagtcggggg 2280
gaggctgcat tatccacgcc ggaggcggtg gtggcttcac gcactgactg acagactgct 2340
ttgatgtgca accgacgacg accagcggca acatcatcac gcagagcatc attttcagct 2400
ttagcatcag ctaactcctt cgtgtatttt gcatcgagcg cagcaacatc acgctgacgc 2460
atctgcatgt cagtaattgc cgcgttcgcc agcttcagtt ctctggcatt tttgtcgcg 2520
tgggctttgt aggtaattgc gttatcacgg taatgattaa cagcccatga caggcagacg 2580
atgatgcaga taaccagagc ggagataatc gcggtgactc tgctcataca tcaatctctc 2640
tgaccgttcc gcccgttctt ttgaattttg caatcagget gtcagcetta tgctcgaact 2700
gaccataacc agcgcccggc agtgaagccc agatattgct gcaacggTcg attgcctgac 2760
ggatatcacc acgatcaatc ataggTaaag cgccacgctc cttaatctgc tgcaatgcca 2820
cagcgtcctg acttttcgga gagaagtctt tcaggccaag ctgcttgcgg taggcatccc 2880
accaacggga aagaagctgg tagcgtccgg cgctgttga tttgagtttt gggtttagcg 2940
tgacaagttt gcgagggtga tcggagtaat cagTaaatag ctctccgctt acaatgacgt 3000
cataaccatg atttctggtt ttctgacgtc cgttatcagt tccctccgac cagccagca 3060
tatcgaggaa cgcttacgt tgattattga tttctacat cttctactcc ggctttttta 3120
gcagcgaagc gtttgataag cgaaccaatc gagtcagTac cgatgtagcc gataaacacg 3180

```

ctcgttatat aagcgagatt gctacttagt ccggcggaagt cgagaaggte acgaatgaac 3240
 taggcgataa tggcgcacat cgttgcgtcg attactgttt ttgtaaacgc accgccatta 3300
 tatctgccgc gaaggtagcg cattgcaaac gcaaggattg ccccgatgcc ttgttccttt 3360
 gccgcgagaa tggcggccaa caggatcatgt ttttctggca tcttcatgtc ttacccccaa 3420
 taaggggatt tgctctatct aattaggaat aaggctgatt actgatagaa caaatccagg 3480
 ctactgtgtt tagtaatcag atttgttcgt gaccgatatg cacgggcaaa acggcaggag 3540
 gttgttagcg caaaaaaaaa attccaaaaa aaaaattcca aaaaaaaaaa gcgactaaca 3600
 aacacaatct gatggca 3617

<210> 63

<211> 68

<212> DNA

<213> 人工序列

<220>

<223> 实施例3中使用的序列.

<400> 63

ggttaaacac ccaagcagac gccgcaatat cagcaccaac agaaacaacc tttagagcga 60
 gcggtcaa 68

<210> 64

<211> 15

<212> DNA

<213> 人工序列

<220>

<223> 实施例3中使用的序列.

<400> 64

ttgaccgctc gcctc 15

<210> 65

<211> 786

<212> DNA

<213> 大肠杆菌

<400> 65

tgtctgaccg caccgccgaa agaagcggca cgtecgacce tgatgccgcg tgcacagtct 60
 tataaagatc tgacccatct gccggctccg acgggcaaaa tttttgttag cgtctataac 120
 atccaggacg aaaccggtca atttaaaccg taccggcgca gtaatttctc cacggccgtt 180
 ccgcagagtg caaccgctat gctggtcacg gcactgaaag attcccgttg gttcattccg 240
 ctggaacgcc agggcctgca aaacctgctg aatgaacgta aaattatccg cgcagctcag 300
 gaaaacggta ccgtggccat taacaatcgt attccgctgc aaagcctgac cgccgcaaac 360
 atcatgggtg aaggctctat catcggttac gaatcaaag tcaaatacgg cggtgtgggc 420
 gcacgttatt ttggcattgg tgctgatacc cagtaccaac tggaccagat cgcagttaac 480
 ctgcgcgtgg ttaatgtcag caccggcgaa attctgagct ctgtgaatac cagcaaaacg 540

atcctgtctt acgaagtgca ggctggtgtt tttcgtttca ttgattatca acgcctgctg 600
 gaaggcgaag tcggttacac ctcaaacgaa ccggtgatgc tgtgtctgat gtcggcgatt 660
 gaaacgggtg ttatttttct gatcaatgat ggcacgacc gtggtctgtg ggatctgcag 720
 aacaaagccg aacgtcaaaa tgacattctg gtgaaatacc gccacatgag tgttccgccg 780
 gaatcc 786

<210> 66

<211> 262

<212> PRT

<213> 大肠杆菌

<400> 66

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

225	230	235	240
Asn Lys Ala Glu Arg Gln Asn Asp Ile Leu Val Lys Tyr Arg His Met			
	245	250	255
Ser Val Pro Pro Glu Ser			
260			

<210> 67

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> StepII (C) 的氨基酸序列

<400> 67

Ser Ala Trp Ser His Pro Gln Phe Glu Lys

1 5 10

<210> 68

<211> 1653

<212> RNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 68

```

auggaagacg ccaaaaacau aaagaaaggc ccggcgccau ucuaucgcu ggaagaugga 60
accgcuggag agcaacugca uaaggcuau aagagauacg ccugguucc uggaacaauu 120
gcuuuuacag augcacauau cgagguggac aucacuuacg cugaguacuu cgaaaugucc 180
guucggguugg cagaagcuau gaaacgauau gggcgauau caaauacag aaucgucgua 240
ugcagugaaa acucucuua auucuuuau ccgguguuugg gcgcguuuu uaucggaguu 300
gcaguugcgc ccgcgaacga cauuuauau gaacgugaau ugcucaacag uaugggcuu 360
ucgcagccua ccgugguguu cguuuccaaa aagggguugc aaaaauuuu gaacgugcaa 420
aaaaagcucc caaucaucca aaaaauuuu aucauggauu cuaaaacgga uuaccagga 480
uuucagucga uguacacguu cgucacauu caucuaccuc ccgguuuuua ugaauacgau 540
uuugugccag aguccuucga uagggacaag acaauugcac ugaucaugaa cuccucugga 600
ucuacugguc ugccuaaagg ugucgcucug ccuauagaa cugccugcgu gagauucug 660
caugccagag auccuauuuu uggcaaucaa aucauuccgg auacugcgau uuuuaguguu 720
guuccauucc aucacgguuu uggaauuuu acuacacucg gauuuuugau auguggauuu 780
cgagucgucu uaauguauag auuugaagaa gagcuguuuc ugaggagccu ucaggauuac 840
aagauucaa gugcgcugcu ggugccaacc cuauucuccu ucuucgcaa aagcacucug 900
auugacaaau acgauuuuac uauuuuacac gaaauugcuu cugguggcgc ucccucucu 960
aaggaagucg gggaagcggu ugccaagagg uuccaucugc cagguaucag gcaaggauau 1020
gggcucacug agacuacauc agcuauucug auuacacccg agggggauga uaaaccgggc 1080

```

gcggucggua aaguuguucc auuuuuugaa gcgaagguug uggaucugga uaccgggaaa 1140
 acgcugggcg uuaaucaaag aggcgacug ugugugagag guccuaugau uauguccggu 1200
 uauguaaaca auccggaagc gaccaacgcc uugauugaca aggauuggaug gcuacauucu 1260
 ggagacauag cuuacuggga cgaagacgaa cacuucuua ucguugaccg ccugaagucu 1320
 cugauuaagu acaaaggcua ucagguggcu cccgcugaau uggaauccau cuugcuccaa 1380
 caccccaaca ucuucgacgc aggugucgca ggucuucccg acgaugacgc cggugaacuu 1440
 cccgccgccg uuguuguuuu ggagcacgga aagacgauga cggaaaaaga gaucguggau 1500
 uacgucgcca gucaaguaac aaccgcgaaa aaguugcgcg gaggaguugu guuuguggac 1560
 gaaguaccga aaggucuuac cggaaaacuc gacgcaagaa aaucagaga gauccucaua 1620
 aaggccaaga agggcggaaa gaucgccgug uaa 1653

<210> 69

<211> 10

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 69

cccccccca 10

<210> 70

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 70

ggttgtttct gttggtgctg atattgc 27

<210> 71

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 71

gcaatatcag caccaacaga aacaacct 28

<210> 72

<211> 48

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 72

cgttctgttt atgtttcttg gacactgatt gacacggttt agtagaac 48

<210> 73

<211> 51

<212> DNA

<213> 人工序列

<220>

<223> 实施例5中使用的序列.

<400> 73

ccccccccca cccccccccca cccccccccca agaaacataa acagaacgtt t 51

运行3

运行2

运行1

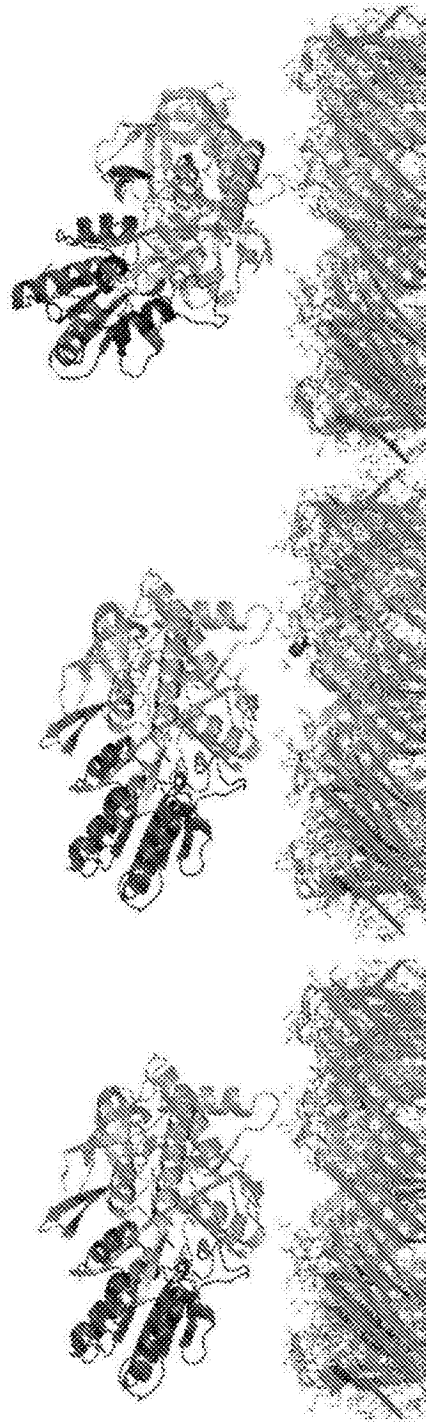


图1

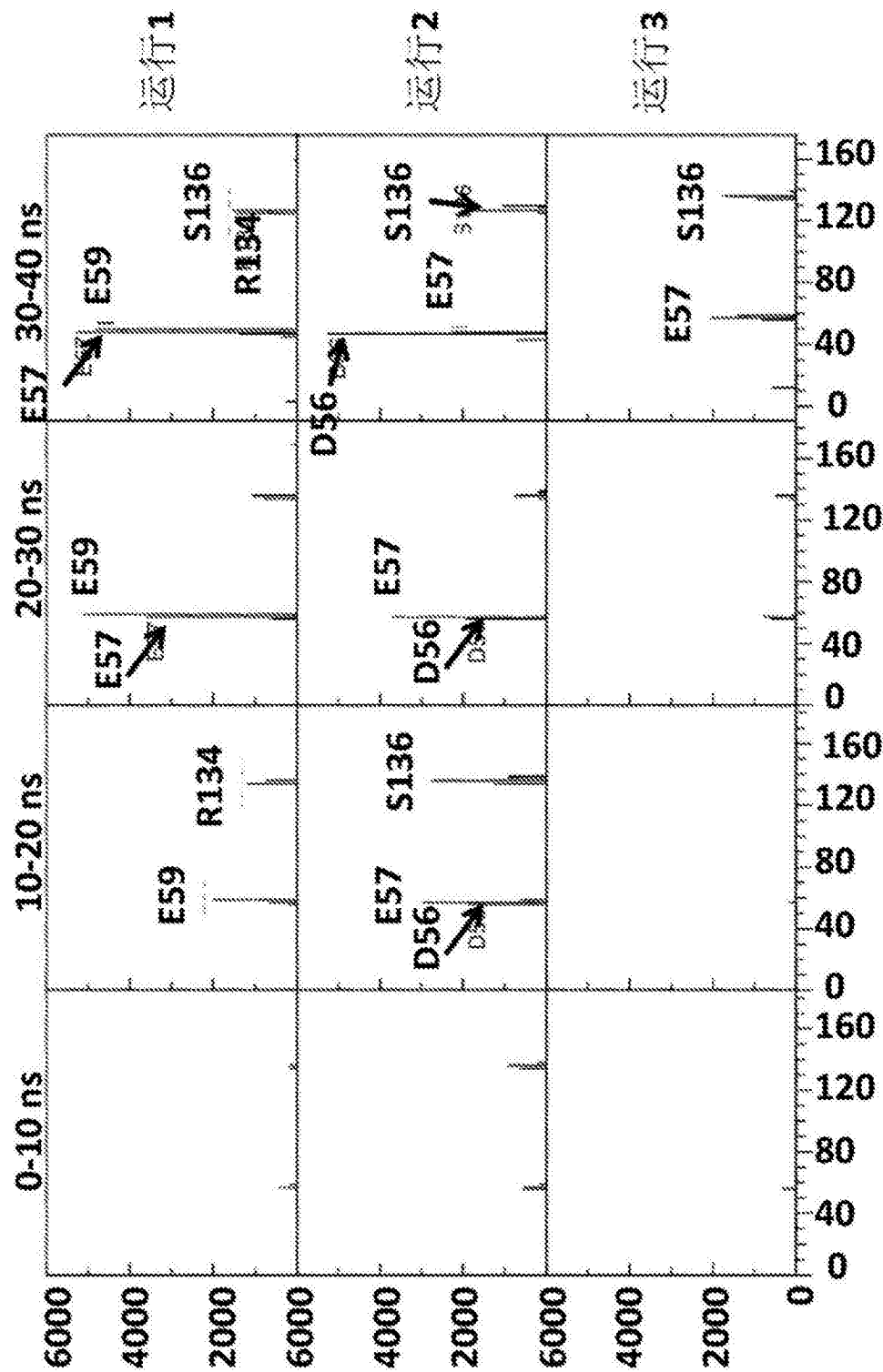


图2

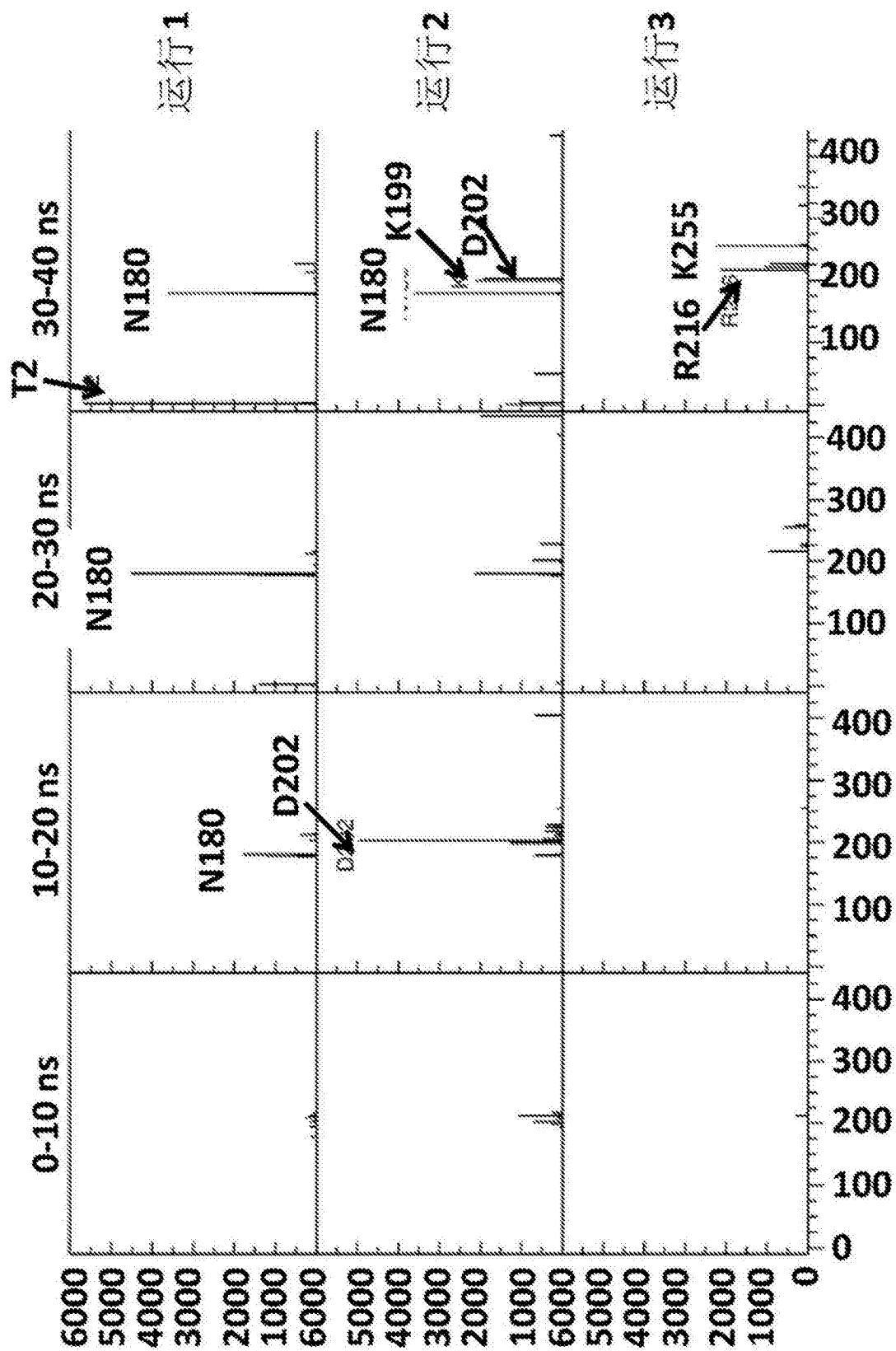


图3

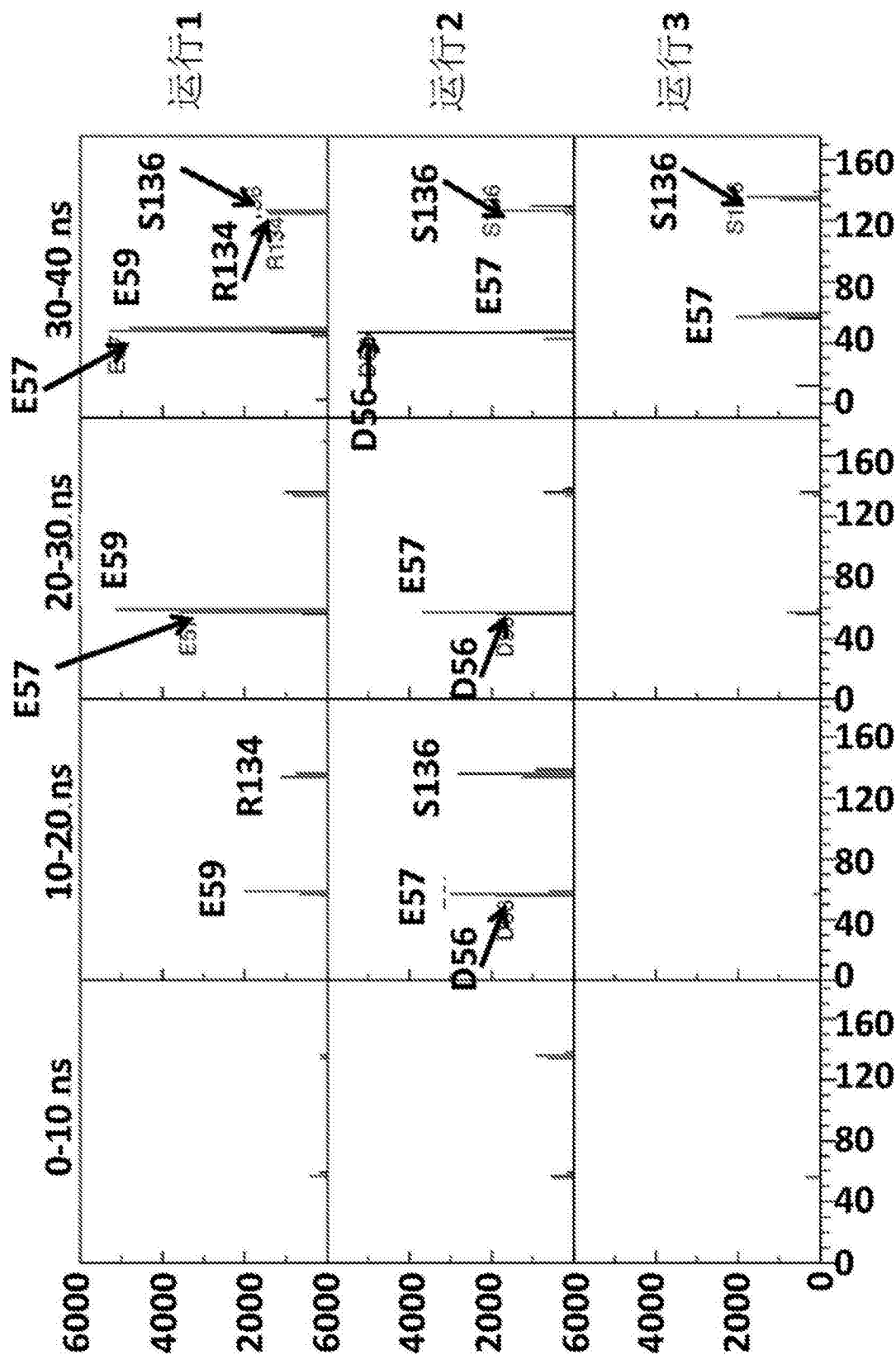


图4

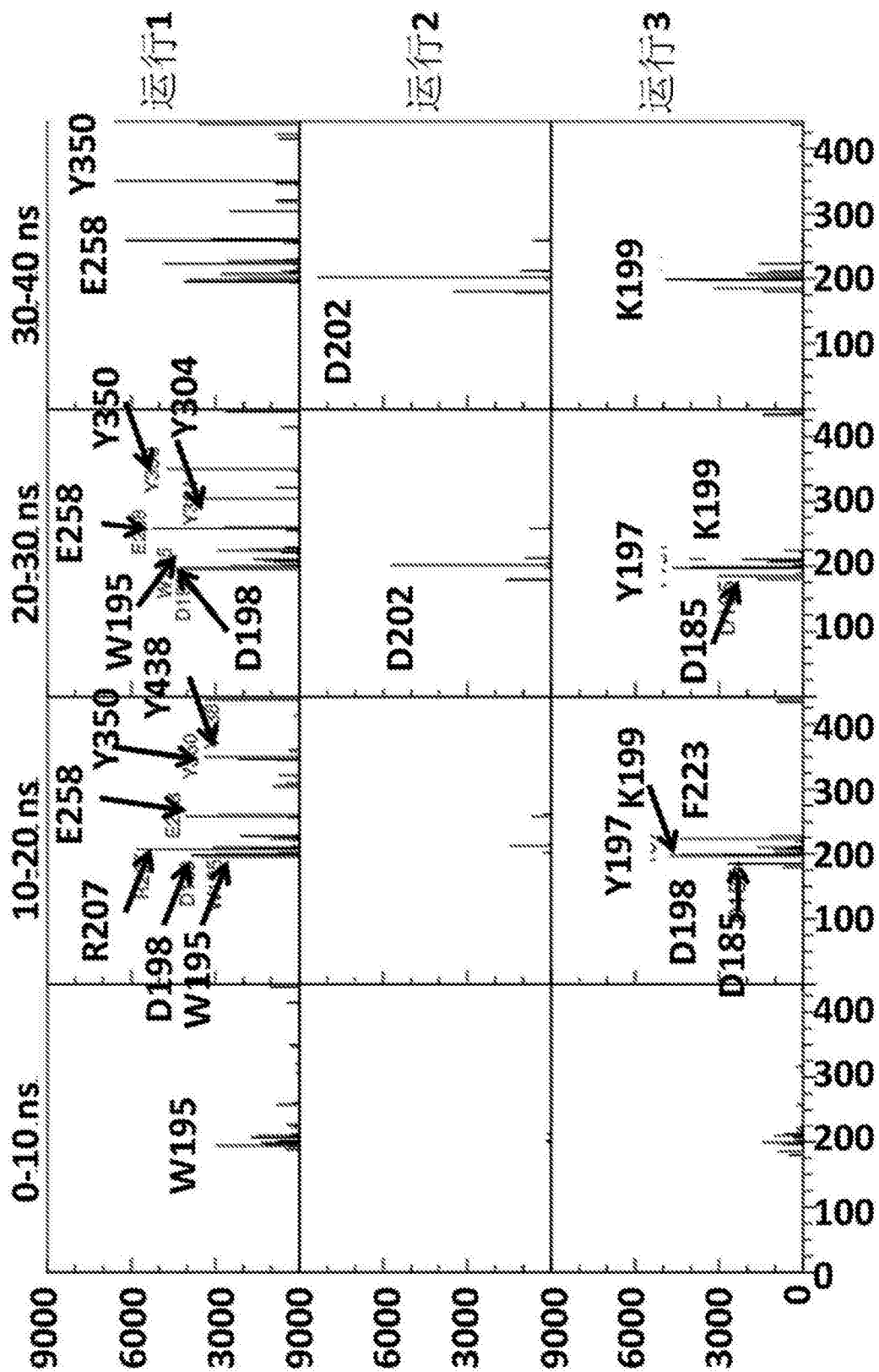


图5

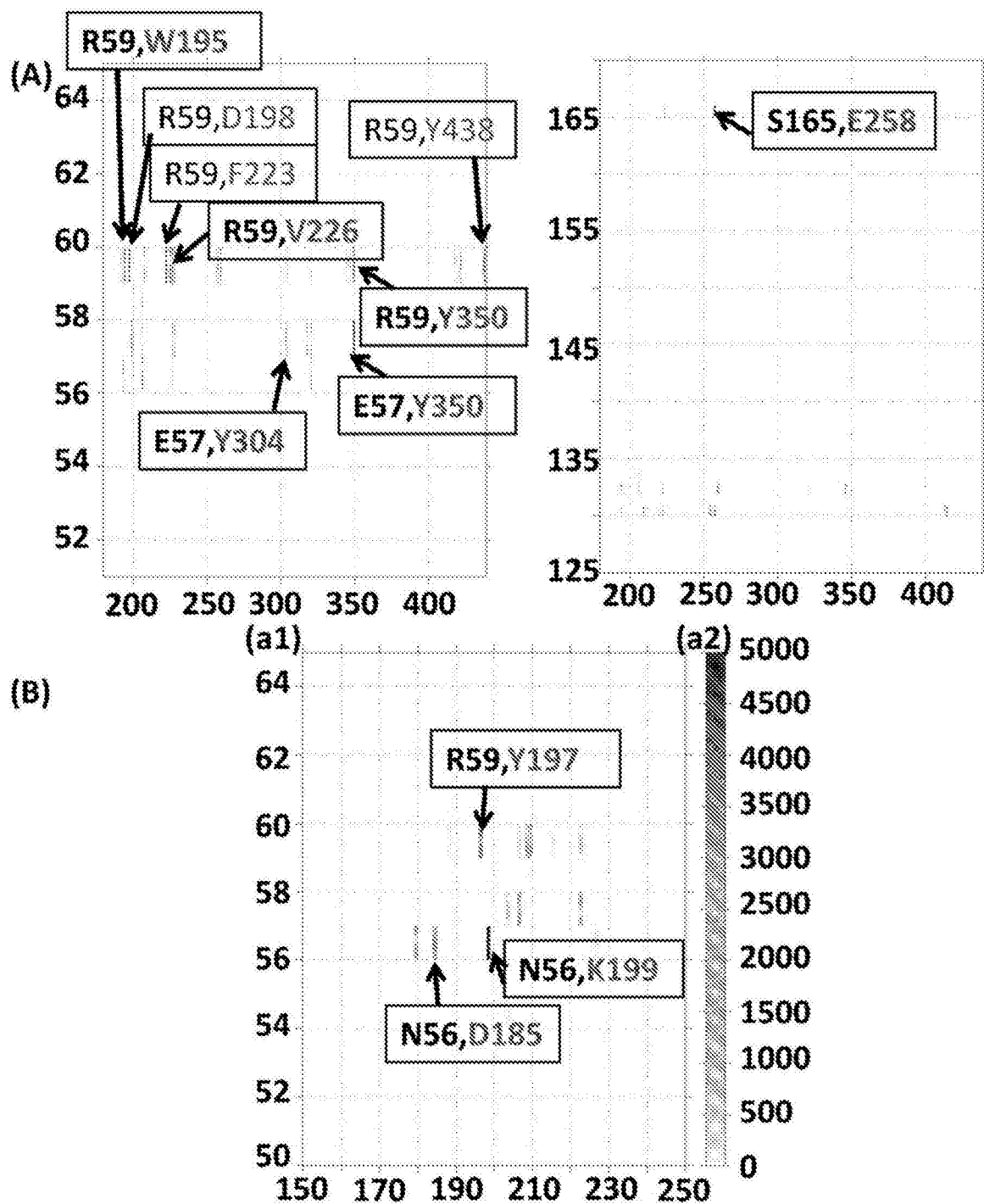


图6

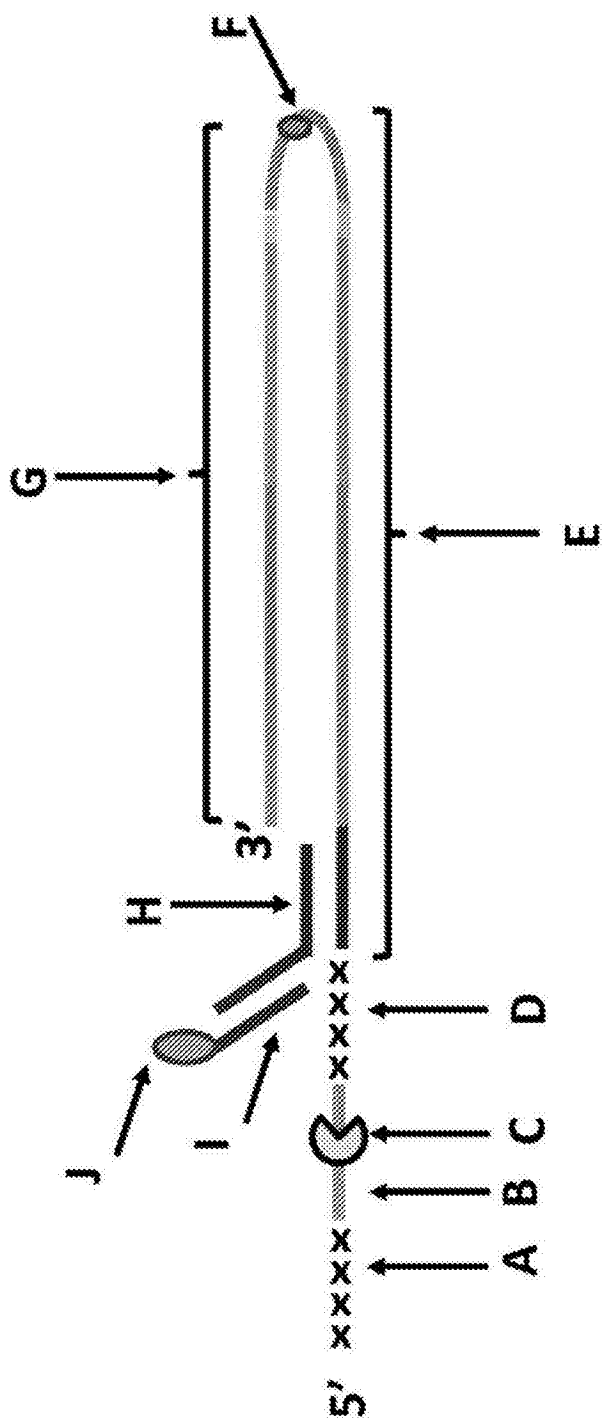


图7

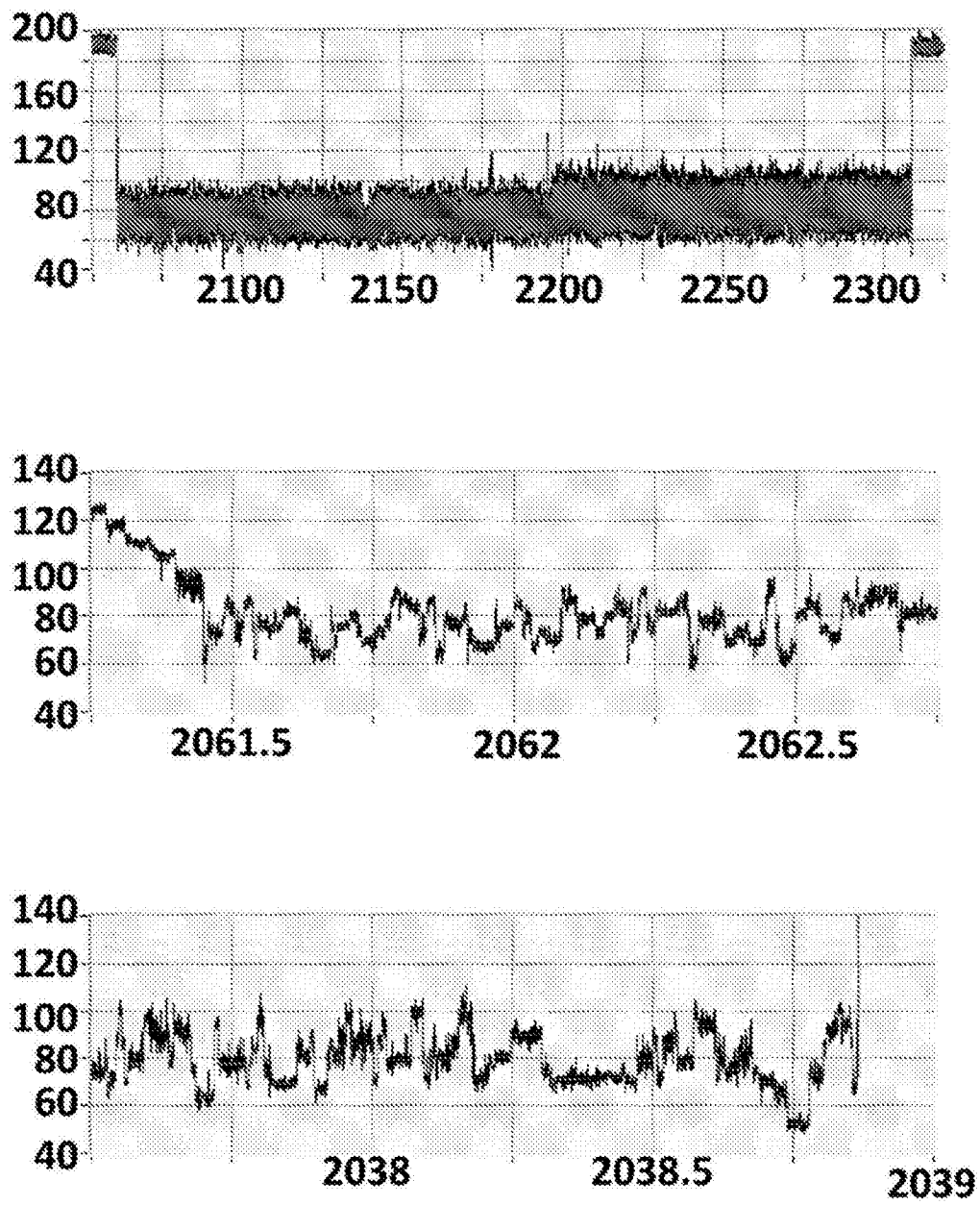


图8

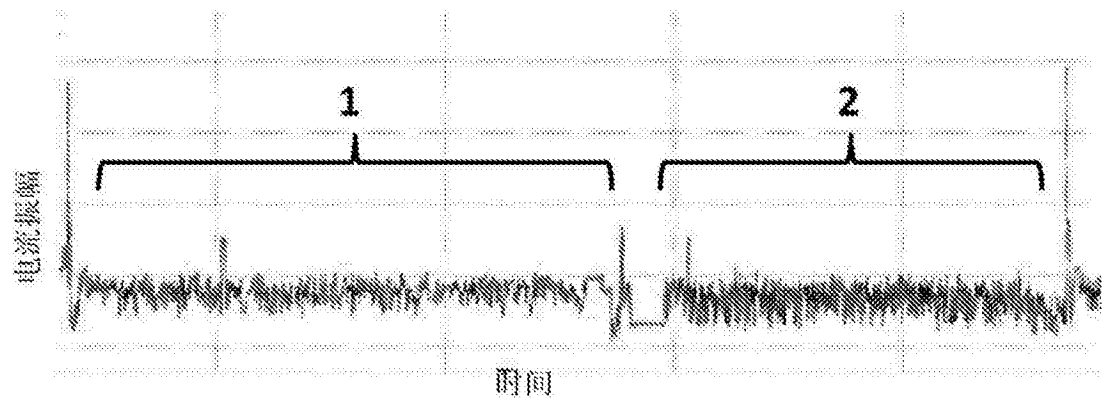


图9