



(12) 发明专利

(10) 授权公告号 CN 106507677 B

(45) 授权公告日 2021.03.12

(21) 申请号 201580027103.X

(22) 申请日 2015.04.15

(65) 同一申请的已公布的文献号
申请公布号 CN 106507677 A

(43) 申请公布日 2017.03.15

(30) 优先权数据

61/979,871 2014.04.15 US

62/062,006 2014.10.09 US

62/080,882 2014.11.17 US

(85) PCT国际申请进入国家阶段日
2016.11.22

(86) PCT国际申请的申请数据
PCT/US2015/025889 2015.04.15

(87) PCT国际申请的公布数据
W02015/160895 EN 2015.10.22

(73) 专利权人 伊鲁米那股份有限公司
地址 美国加利福尼亚州

(72) 发明人 克里斯蒂安·格勒克纳
阿米尔阿里·基亚 艾琳·博马蒂
莫利·何 海纓·李·格鲁内瓦德
斯科特·库尔斯腾

特瑞纳·法耶·奥索斯普拉罗普
达林·哈斯金斯 约书亚·伯吉斯
阿奴巴玛·康纳 丹尼尔·施凌曼
拉梅什·威德亚纳森

(74) 专利代理机构 北京安信方达知识产权代理
有限公司 11262

代理人 郑霞

(51) Int.Cl.

C12N 9/12 (2006.01)

C12N 15/54 (2006.01)

C12Q 1/6869 (2018.01)

(56) 对比文件

CN 1367840 A, 2002.09.04

US 2006040355 A1, 2006.02.23

CN 1319135 A, 2001.10.24

CN 102264914 A, 2011.11.30

Richard J. Gradman et al..A

bifunctional DNA binding region in Tn5
transposase.《Molecular Microbiology》
.2007, 讨论部分, 图1和表1.

审查员 赵鹏

权利要求书2页 说明书39页
序列表42页 附图37页

(54) 发明名称

用于改进插入序列偏倚和增加DNA输入耐受
性的修饰的转座酶

(57) 摘要

本文提供了用于改进核酸样品的片段化和
加标签的转座酶和反应条件, 特别地表现出改进
的插入序列偏倚的改变的转座酶和反应条件, 以
及使用以上描述的转座酶和反应条件的方法及
试剂盒。

1. 一种突变体Tn5转座酶,所述突变体Tn5转座酶具有转座酶活性,由自SEQ ID NO:1或SEQ ID NO:11突变的序列组成,所述序列与SEQ ID NO:1或SEQ ID NO:11相比的突变在位置Pro214和Ala338处,其中在位置Pro214处的突变是向精氨酸的突变并且在Ala338处的突变是向缬氨酸的突变。

2. 一种突变体Tn5转座酶,所述突变体Tn5转座酶具有转座酶活性,由自SEQ ID NO:1或SEQ ID NO:11突变的序列组成,所述序列与SEQ ID NO:1或SEQ ID NO:11相比的突变为在位置Pro214处向精氨酸的突变、在位置Ala338处向缬氨酸的突变和在位置Lys212处向精氨酸的突变;或者为在位置Pro214处向精氨酸的突变、在位置Ala338处向缬氨酸的突变、在Lys212处向精氨酸的突变和在Gly251处向精氨酸的置换突变。

3. 如权利要求2所述的突变体Tn5转座酶,其中所述序列与SEQ ID NO:1或SEQ ID NO:11相比的突变由以下突变组成:在位置Pro214处向精氨酸的突变、在位置Ala338处向缬氨酸的突变和在Lys212处向精氨酸的突变。

4. 如权利要求2所述的突变体Tn5转座酶,其中所述序列与SEQ ID NO:1或SEQ ID NO:11相比的突变由以下突变组成:在位置Pro214处向精氨酸的突变、在位置Ala338处向缬氨酸的突变、在Lys212处向精氨酸的突变和在Gly251处的突变是向精氨酸的突变。

5. 一种融合蛋白,所述融合蛋白包括与多肽融合结构域融合的如权利要求1至4中任一项所述的突变体Tn5转座酶。

6. 如权利要求5所述的融合蛋白,其中所述多肽融合结构域被融合至所述转座酶的N-末端或其中所述多肽融合结构域被融合至所述转座酶的C-末端。

7. 如权利要求6所述的融合蛋白,其中所述多肽融合结构域包含纯化标签。

8. 如权利要求6所述的融合蛋白,其中所述多肽融合结构域包含增加溶解度的标签。

9. 如权利要求8所述的融合蛋白,其中所述多肽融合结构域包含选自以下的结构域:麦芽糖结合蛋白(MBP)、延伸因子Ts(Tsf)、5-甲基胞嘧啶结合结构域及蛋白A。

10. 一种核酸分子,所述核酸分子编码如在权利要求1-4中任一项中定义的突变体Tn5转座酶。

11. 一种表达载体,所述表达载体包含权利要求10所述的核酸分子。

12. 一种宿主细胞,所述宿主细胞包含权利要求11所述的表达载体。

13. 一种用于体外转座的方法,所述方法包括:允许以下组分相互作用:(i)转座体复合物,所述转座体复合物包含根据权利要求1-4中任一项所述的突变体Tn5转座酶,和(ii)靶DNA。

14. 一种用于测序靶DNA的方法,所述方法包括:

(a) 使所述靶DNA与转座体复合物孵育,所述转座体复合物包含

(1) 根据权利要求1-4中任一项所述的突变体Tn5转座酶;以及

(2) 第一多核苷酸,所述第一多核苷酸包含:

(i) 3'部分,所述3'部分包含转座子末端序列,和

(ii) 第一标签,所述第一标签包含第一测序标签结构域,

所述孵育在所述靶DNA被片段化,且所述第一多核苷酸的3'转座子末端序列被转移至片段化的靶DNA的5'末端的条件下进行,

从而产生双链片段,在所述双链片段中所述5'末端经由所述第一标签加标签,且在所

述5'-加标签的链的3'末端处存在单链缺口；

(b) 将所述双链片段与修饰核酸的酶在第二标签被附连至所述5'-加标签的链的3'末端的条件下孵育，

(c) 通过提供聚合酶和对应于所述第一多核苷酸的一部分的扩增引物扩增所述(b)所获得的片段，从而产生具有在所述5'末端处的所述第一标签和在所述3'末端处的第二标签的加双标签片段的代表性文库；

(d) 提供第一测序引物，所述第一测序引物包含对应于所述第一测序标签结构域的部分；以及

(e) 延伸所述第一测序引物并且并行检测与所述加双标签片段的代表性文库的所述第一测序标签结构域相邻的核苷酸的身份。

15. 一种用于进行体外转座反应的试剂盒，所述试剂盒包含：转座体复合物，所述转座体复合物包含

(1) 根据权利要求1-4中任一项所述的突变体Tn5转座酶；和

(2) 多核苷酸，所述多核苷酸包括包含转座子末端序列的3'部分。

16. 一种用于体外转座的方法，所述方法包括：允许以下组分相互作用：(i) 转座体复合物，所述转座体复合物包含如权利要求1-4中任一项所述的突变体Tn5转座酶或如权利要求5-9中任一项所述的融合蛋白，(ii) 靶DNA；以及(iii) 反应缓冲液，所述反应缓冲液与在包含 Mg^{2+} 的缓冲液中的相同的酶相比，包含足以降低GC偏倚的有效浓度的 Co^{2+} 。

17. 如权利要求16所述的方法，其中所述 Co^{2+} 为在5mM至40mM范围内的浓度。

18. 如权利要求16所述的方法，其中所述 Co^{2+} 为在5mM至20mM范围内的浓度。

19. 如权利要求16所述的方法，其中所述 Co^{2+} 为在8mM至12mM范围内的浓度。

20. 一种融合蛋白，所述融合蛋白具有转座酶活性，由SEQ ID NO:26组成。

21. 一种突变体Tn5转座酶，所述突变体Tn5转座酶具有转座酶活性，由SEQ ID NO:19组成。

用于改进插入序列偏倚和增加DNA输入耐受性的修饰的转座酶

[0001] 相关申请

[0002] 本申请要求2014年4月15日提交的美国临时申请第61/979,871号;2014年10月9日提交的美国临时申请第62/062,006号;2014年11月17日提交的美国临时申请第62/080,882号的优先权,其在此通过引用以其整体并入。

技术领域

[0003] 本申请涉及但不限于用于改进插入序列偏倚和增加DNA输入耐受性的修饰的转座酶。

[0004] 背景

[0005] 转座酶在体外转座系统中是有用的。它们允许大规模片段化并将基因组DNA加标签,并且用于制备来自靶DNA的加标签的DNA片段的文库,以用于核酸分析方法诸如下一代测序和扩增方法。对这样的修饰的转座酶仍然存在需求,其具有改进的特性并且产生加标签的DNA片段,该加标签的DNA片段定性和定量地代表产生靶核酸的样品中的靶核酸。

[0006] 序列表

[0007] 本申请与电子格式的序列表一起提交。序列表作为命名为IP-1198-TW_Sequence_Listing.TXT的文件提供,2015年4月13日创建,大小为106Kb。该电子格式的序列表中的信息通过引用以其整体并入本文。

[0008] 概述

[0009] 本文提供了用于改进核酸样品的片段化和加标签的转座酶。本发明人已出人意料地鉴定了表现出改进的插入序列偏倚(bias)并具有许多其他相关优势的某些改变的转座酶。

[0010] 本文提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶。在一些实施方案中,突变体转座酶可包含在位置Asp248处的突变。在某些方面,在位置Asp248处的突变为置换突变。在某些方面,在位置Asp248处的置换突变可包括向选自以下组成的组的残基的突变:Tyr、Thr、Lys、Ser、Leu、Ala、Trp、Pro、Gln、Arg、Phe及His。

[0011] 在某些方面,在位置Asp248处的突变为在位置Asp248之后的插入突变。在某些方面,插入突变可包括在位置Asp248之后插入疏水性残基。在某些方面,插入突变可包括在位置Asp248之后插入缬氨酸残基。

[0012] 本文还提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Asp119处的突变。在某些方面,在位置Asp119处的突变为置换突变。在某些方面,在位置Asp119处的置换突变可包括向疏水性残基的突变。在某些方面,在位置Asp119处的置换突变可包括向亲水性残基的突变。在某些方面,在位置Asp119处的置换突变可包括向选自以下组成的组的残基的突变:Leu、Met、Ser、Ala及Val。

[0013] 本文还提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Trp125处的突变。在某些方面,在位置Trp125处的突变为置换突变。在某些方

面,在位置Trp125处的置换突变可包括向甲硫氨酸残基的突变。

[0014] 本文还提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Lys120处的突变。在某些方面,在位置Lys120处的突变为置换突变。在某些方面,在位置Lys120处的置换突变可包括向大体积芳香族残基的突变。在某些方面,在位置Lys120处的置换突变可包括向选自以下组成的组的残基的突变:Tyr、Phe、Trp及Glu。

[0015] 本文还提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Lys212和/或Pro214和/或Ala338处的突变。在某些方面,在位置Lys212和/或Pro214和/或Ala338处的一个或多个突变为置换突变。在某些方面,在位置Lys212处的置换突变包括向精氨酸的突变。在某些方面,在位置Pro214处的置换突变包括向精氨酸的突变。在某些方面,在位置Ala338处的置换突变包括向缬氨酸的突变。在一些实施方案中,转座酶还可包含在Gly251处的置换突变。在某些方面,在位置Gly251处的置换突变包括向精氨酸的突变。

[0016] 本文还提供了相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Glu146和/或Glu190和/或Gly251处的突变。在某些方面,在位置Glu146和/或Glu190和/或Gly251处的一个或多个突变为置换突变。在某些方面,在位置Glu146处的置换突变可包括向谷氨酰胺的突变。在某些方面,在位置Glu190处的置换突变可包括向甘氨酸的突变。在某些方面,在位置Gly251处的置换突变可包括向精氨酸的突变。

[0017] 还提供了改变的转座酶,所述改变的转座酶包含对包含SEQ ID NO:21的氨基酸序列的半保守结构域的置换突变,其中所述置换突变包括在位置2处向除了Trp、Asn、Val或Lys以外的任何残基的突变。在某些实施方案中,突变包括在位置2处向Met的置换。

[0018] 在任何以上描述的实施方案中,突变体Tn5转座酶还可包含在功能上等同于Tn5转座酶氨基酸序列中的Glu54和/或Met56和/或Leu372的位置处的置换突变。在某些实施方案中,转座酶包含与Tn5转座酶氨基酸序列中的Glu54Lys和/或Met56Ala和/或Leu372Pro同源的置换突变。

[0019] 本文还提供了突变体Tn5转座酶,所述突变体Tn5转座酶包含SEQ ID NO:2-10、12-20和25-26中任一个的氨基酸序列。

[0020] 本文还提供了融合蛋白,所述融合蛋白包含与另外的多肽融合的如在任何以上实施方案中定义的突变体Tn5转座酶。在一些实施方案中,与转座酶融合的多肽结构域可包含纯化标签、表达标签、溶解度标签或其组合。在一些实施方案中,与转座酶融合的多肽结构域可包含,例如,麦芽糖结合蛋白(MBP)。在一些实施方案中,与转座酶融合的多肽结构域可包含,例如,延伸因子Ts(Tsf)。

[0021] 本文还提供了核酸分子,所述核酸分子编码如在任何以上实施方案中定义的突变体Tn5转座酶。本文还提供了表达载体,所述表达载体包含以上描述的核酸分子。本文还提供了宿主细胞,所述宿主细胞包含以上描述的载体。

[0022] 本文还提供了用于体外转座的方法,所述方法包括:允许以下组分相互作用:(i)转座体(transposome)复合物,所述转座体复合物包含根据上文描述的任一实施方案的突变体Tn5转座酶,和(ii)靶DNA。

[0023] 本文还提供了用于使用上文描述的Tn5转座酶测序靶DNA的方法。在一些实施方案中,所述方法可包括(a)将靶DNA与转座体复合物孵育,所述转座体复合物包含(1)根据上文

描述的任一实施方案的突变体Tn5转座酶和(2)第一多核苷酸,所述第一多核苷酸包含(i) 3'部分,所述3'部分包含转座子末端序列和(ii)第一标签,所述第一标签包含第一测序标签结构域,所述孵育在藉此所述靶DNA被片段化,且所述第一多核苷酸的3'转座子末端序列被转移至该片段的5'末端的条件下进行,从而产生双链片段,在所述双链片段中5'末端经由第一标签加标签,且在5'-加标签链的3'末端处存在单链缺口;(b)将所述片段与修饰核酸的酶在藉此第二标签被附连至5'-加标签链的3'末端的条件下孵育;(c)任选地,通过提供聚合酶和对应于所述第一多核苷酸的一部分的扩增引物扩增所述片段,从而产生具有在5'末端处的第一标签和在3'末端处的第二标签的加双标签片段的代表性文库;(d)提供第一测序引物,所述第一测序引物包含对应于第一测序标签结构域的部分;以及(e)延伸所述第一测序引物并且并行检测与加双标签片段的代表性文库的第一测序标签结构域相邻的核苷酸的身份。

[0024] 本文还提供了用于进行体外转座反应的试剂盒。在一些实施方案中,试剂盒可包含转座体复合物,所述转座体复合物包含(1)根据上文描述的任一实施方案的突变体Tn5转座酶;和(2)包含含转座子末端序列的3'部分的多核苷酸。

[0025] 本文还提供了以下各项:

[0026] 项目1.一种修饰的转座酶,所述修饰的转座酶包括与多肽融合结构域融合的野生型或突变体转座酶。

[0027] 项目2.如项目1所述的修饰的转座酶,其中所述多肽融合结构域被融合至所述转座酶的N-末端。

[0028] 项目3.如项目1所述的修饰的转座酶,其中所述多肽融合结构域被融合至所述转座酶的C-末端。

[0029] 项目4.如项目1所述的修饰的转座酶,其中所述多肽融合结构域包含纯化标签。

[0030] 项目5.如项目1所述的修饰的转座酶,其中所述多肽融合结构域包含增加溶解度的标签。

[0031] 项目6.如项目1所述的修饰的转座酶,其中所述多肽融合结构域包含选自以下的结构域:麦芽糖结合蛋白(MBP)、延伸因子Ts(Tsf)、5-甲基胞嘧啶结合结构域及蛋白A。

[0032] 项目7.一种标签化反应缓冲液,所述标签化反应缓冲液与在包含 Mg^{2+} 的缓冲液中的相同的酶相比,包含转座酶和足以降低GC偏倚的有效浓度的 Co^{2+} 。

[0033] 项目8.如项目7所述的标签化反应缓冲液,其中所述 Co^{2+} 为在约5mM至约40mM范围内的浓度。

[0034] 项目9.如项目7所述的标签化反应缓冲液,其中所述 Co^{2+} 为在约5mM至约20mM范围内的浓度。

[0035] 项目10.如项目7所述的标签化反应缓冲液,其中所述 Co^{2+} 为在约8mM至约12mM范围内的浓度。

[0036] 项目11.如项目7所述的标签化反应缓冲液,其中所述 Co^{2+} 为在约10mM的浓度。

[0037] 项目12.一种用于体外转座的方法,所述方法包括:允许以下组分相互作用:(i)转座体复合物,所述转座体复合物包含转座酶,(ii)靶DNA;以及(iii)反应缓冲液,所述反应缓冲液与在包含 Mg^{2+} 的缓冲液中的相同的酶相比,包含足以降低GC偏倚的有效浓度的 Co^{2+} 。

- [0038] 项目13.如项目12所述的方法,其中所述 Co^{2+} 为在约5mM至约40mM范围内的浓度。
- [0039] 项目14.如项目12所述的方法,其中所述 Co^{2+} 为在约5mM至约20mM范围内的浓度。
- [0040] 项目15.如项目12所述的方法,其中所述 Co^{2+} 为在约8mM至约12mM范围内的浓度。
- [0041] 项目16.如项目12所述的方法,其中所述转座酶包含根据权利要求1-49中任一项所述的突变体Tn5转座酶。
- [0042] 项目17.如项目12所述的方法,其中所述转座酶包括Mos-1转座酶。
- [0043] 项目18.如项目12所述的方法,其中所述转座酶包括高活性Tn5转座酶。
- [0044] 项目19.一种突变体Tn5转座酶,其中所述突变体Tn5转座酶与野生型Tn5转座酶相比表现出增加的DNA输入耐受性。
- [0045] 项目20.如项目19所述的突变体Tn5转座酶,其中所述突变体转座酶包含SEQ ID NO:19的氨基酸序列。
- [0046] 项目21.如项目20所述的突变体Tn5转座酶,其中所述突变体Tn5转座酶在 ≥ 2.4 的突变体Tn5转座酶的nM数:输入DNA的ng数的比率表现出增加的DNA输入耐受性。
- [0047] 项目22.一种跨越一定范围的靶DNA输入量产生均一片段尺寸的方法,所述方法包括:
- [0048] (a) 提供靶DNA;其中所述靶DNA的量选自一定范围的靶DNA量,以及
- [0049] (b) 将所述靶DNA与转座酶孵育,其中所述转座酶的氨基酸序列包含来自SEQ ID NO:19的C-末端的至少400个氨基酸,且其中所述转座酶片段化所述靶DNA以产生均一片段尺寸。
- [0050] 项目23.如项目22所述的方法,其中所述靶DNA为基因组DNA。
- [0051] 项目24.如项目23所述的方法,其中所述基因组DNA为原核生物DNA。
- [0052] 项目25.如项目24所述的方法,其中所述原核生物DNA为细菌DNA。
- [0053] 项目26.如项目23所述的方法,其中所述基因组DNA为真核生物基因组DNA。
- [0054] 项目27.如项目26所述的方法,其中所述真核生物基因组DNA为人基因组DNA。
- [0055] 项目28.如项目22-27中任一项所述的方法,其中所述靶DNA的量在约1ng至200ng的范围内。
- [0056] 项目29.一种跨越一定范围的靶DNA输入量测序靶DNA的方法,所述方法包括:
- [0057] (a) 使所述靶DNA与转座体复合物孵育,所述转座体复合物包含
- [0058] (1) 突变体Tn5转座酶,所述突变体Tn5转座酶包含SEQ ID NO:19的氨基酸序列;和
- [0059] (2) 第一多核苷酸,所述第一多核苷酸包含:
- [0060] (i) 3'部分,所述3'部分包含转座子末端序列,和
- [0061] (ii) 第一标签,所述第一标签包含第一测序标签结构域,
- [0062] 其中所述靶DNA的量选自一定范围的靶DNA量,所述孵育在藉此所述靶DNA被片段化,且所述第一多核苷酸的3'转座子末端序列被转移至该片段的5'末端的条件下进行,
- [0063] 从而产生双链片段,在所述双链片段中所述5'末端经由所述第一标签加标签,且在所述5'-加标签链的3'末端处存在单链缺口;
- [0064] (b) 将所述片段与修饰核酸的酶在藉此第二标签被附连至所述5'-加标签链的3'末端的条件下孵育,
- [0065] (c) 任选地,通过提供聚合酶和对应于所述第一多核苷酸的一部分的扩增引物扩

增所述片段,从而产生具有在所述5'末端处的所述第一标签和在所述3'末端处的第二标签的加双标签片段的代表性文库;

[0066] (d) 提供第一测序引物,所述第一测序引物包含对应于所述第一测序标签结构域的部分;以及

[0067] (e) 延伸所述第一测序引物并且并行检测与所述加双标签片段的代表性文库的所述第一测序标签结构域相邻的核苷酸的身份。

[0068] 项目30.如项目29所述的方法,其中所述靶DNA的量在约1ng至200ng的范围内。

[0069] 项目31.如项目30所述的方法,其中突变体Tn5转座酶的nM数:输入DNA的ng数的比率 ≥ 2.4 。

[0070] 项目32.如项目22所述的方法,其中所述方法用于外显子组富集。

[0071] 项目33.如项目19所述的突变体Tn5转座酶,其中所述突变体转座酶具有SEQ ID NO:26的氨基酸序列。

[0072] 项目34.一种突变体Tn5转座酶,所述突变体Tn5转座酶包含SEQ ID NO:2-10、12-20和25-26中任一个的至少400个氨基酸序列的氨基酸序列。

[0073] 一个或更多个实施方案的细节在附图和以下的说明中描述。其他的特征、目标、和优势将由说明书和附图,以及由权利要求而明显。

[0074] 附图简述

[0075] 图1A为显示Tn5转座酶(1MUH)、Hermes转座酶(2BW3)、HIV整合酶(1ITG)、Mu转座酶(1BCM)及Mos1转座酶(3HOS)的催化核心结构域的结构比对的示意图。所示的编号代表Tn5转座酶中的氨基酸残基的编号。

[0076] 图1B为显示Tn5转座酶(1MUH,粉色)、Hermes转座酶(2BW3,黑色)、HIV整合酶(1ITG,棕黄色)、Mu转座酶(1BCM)及Mos1转座酶(3HOS,黄色)的结构比对催化核心结构域的示意图。Tn5转座酶W125位置以棒状表示来显示。

[0077] 图2为显示与Tn5对照相比,D248Y突变体Tn5转座酶的改变的序列插入偏倚的IVC图。

[0078] 图3为显示与Tn5对照相比,D119L突变体Tn5转座酶的改变的序列插入偏倚的IVC图。

[0079] 图4为显示与Tn5对照相比,W125M突变体Tn5转座酶的改变的序列插入偏倚的IVC图。

[0080] 图5为显示与Tn5对照相比,ia248V插入突变体Tn5转座酶的改变的序列插入偏倚的IVC图。

[0081] 图6为显示与Tn5对照相比,K120Y、K120F和K120W Tn5转座酶的改变的序列插入偏倚的IVC图。

[0082] 图7为显示与Tn5对照相比,三种突变体Tn5转座酶的改变的序列插入偏倚的IVC图。

[0083] 图8A为显示与Tn5对照相比由三种突变体Tn5转座酶创建的蜡状芽孢杆菌(*B.cereus*)文库中的AT/GC脱落(dropout)的图。图8B为显示与Tn5对照相比由三种突变体Tn5转座酶创建的蜡状芽孢杆菌文库的估计的文库尺寸的图。

[0084] 图9A为显示与Tn5对照相比,在由两种突变体Tn5转座酶创建的文库中的快速捕获

富集实验中的覆盖度均匀性的图。图9B为显示与Tn5对照相比,在由两种突变体Tn5转座酶创建的文库中的快速捕获富集实验中的10X和20X靶覆盖度和平均靶覆盖度的图。

[0085] 图10A为显示与Tn5对照相比,在由两种突变体Tn5转座酶创建的文库中的快速捕获富集实验中的独特读段和杂交选择文库尺寸的通过过滤器百分比的图。图10B为显示与Tn5对照相比,在由两种突变体Tn5转座酶创建的文库中的快速捕获富集实验中达到10X、20X和30X覆盖度的罚分的图。

[0086] 图11示出了使用不同的标签化 (tagmentation) 缓冲液制备的TS-Tn5059和TS-Tn5标签化 (tagmented) 的DNA文库中的独特分子数目的条形图。

[0087] 图12示出了使用不同的标签化缓冲液制备的TS-Tn5059和TS-Tn5标签化的DNA文库中的GC脱落百分比的条形图。

[0088] 图13示出了使用不同的标签化缓冲液制备的TS-Tn5059和TS-Tn5标签化的DNA文库中的AT脱落百分比的条形图。

[0089] 图14示出了使用标准缓冲液 (TD) 和钴缓冲液 (Co) 制剂制备的TS-Tn5059文库中的片段尺寸分布的生物分析仪迹线的图。

[0090] 图15示出了使用钴-DMSO (Co-DMSO)、NF2和HMW缓冲液制剂制备的TS-Tn5059文库中的片段尺寸分布的生物分析仪迹线的图。

[0091] 图16示出了使用标准缓冲液制剂 (TD) 和钴缓冲液 (Co) 制备的TS-Tn5文库中的片段尺寸分布的生物分析仪迹线的图。

[0092] 图17示出了使用钴-DMSO (Co-DMSO)、NF2和HMW缓冲液制剂制备的TS-Tn5文库中的片段尺寸分布的生物分析仪迹线的图。

[0093] 图18A、18B、18C和18D分别示出了TS-Tn5文库中的序列内容物的偏倚图、TS-TN5-Co文库中的序列内容物的偏倚图、TS-Tn5-Co-DMSO文库中的序列内容物的偏倚图及TS-Tn5-NF2文库中的序列内容物的偏倚图。

[0094] 图19A、19B、19C和19D分别示出了TS-Tn5059文库中的序列内容物的偏倚图、TS-TN5059-Co文库中的序列内容物的偏倚图、TS-Tn5059-Co-DMSO文库中的序列内容物的偏倚图及TS-Tn5059-NF2文库中的序列内容物的偏倚图。

[0095] 图20示出了使用不同的标签化缓冲液制备的MBP-Mos1标签化的文库中的平均读段总数和平均多样性的条形图。

[0096] 图21示出了MBP-Mos1标签化的文库中的GC和AT脱落的条形图。

[0097] 图22A、22B、22C和22D分别示出了Mos1-HEPES文库中的序列内容物的偏倚图、Mos1-HEPES-DMSO文库中的序列内容物的偏倚图、Mos1-HEPES-DMSO-Co文库中的序列内容物的偏倚图及Mos1-HEPES-DMSO-Mn文库中的序列内容物的偏倚图。

[0098] 图23示出了制备和富集用于外显子组测序的基因组DNA文库的方法的实例的流程图。

[0099] 图24A示出了使用TS-Tn5059转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库中的覆盖度的图。

[0100] 图24B示出了使用NexteraV2转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库中的覆盖度的图。

[0101] 图25A示出了使用TS-Tn5059转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库

中的缺口位置和缺口长度的图。

[0102] 图25B示出了使用NexteraV2转座体制备的标签化的蜡状芽孢杆菌基因组标签化的DNA文库中的缺口位置和缺口长度的图。

[0103] 图26示出了使用对25 ng人gDNA归一化为40nM(1X归一化的浓度)的TS-Tn5059的TDE1(Tn5版本-1)和TS-Tn5制备的标签化的基因组DNA文库中的片段尺寸分布的生物分析仪迹线的图。

[0104] 图27示出了采用25 ng人gDNA使用归一化为40nM(1X归一化的浓度)的TS-Tn5059的TDE1(Tn5版本-1)和TS-Tn5制备的标签化的基因组DNA文库中的尺寸分布的分析。

[0105] 图28示出了使用一定范围的DNA输入制备的标签化的基因组DNA文库中的片段尺寸分布的生物分析仪迹线的图。

[0106] 图29A示出了由第一使用者并使用Coriel人DNA制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0107] 图29B示出了由第二使用者并使用Coriel人DNA制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0108] 图30示出了由在6x“归一化”浓度的TDE1使用25 ng-100ng gDNA制备的Tn5版本1(TDE1)标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0109] 图31示出了由在6x“归一化”浓度的TS-Tn5使用25 ng-100ng gDNA制备的TS-Tn5标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0110] 图32示出了由在6x“归一化”浓度的TS-Tn5059使用10ng-100ng gDNA制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0111] 图33示出了由在6x“归一化”浓度的TS-Tn5059使用宽范围gDNA(5ng-500ng)制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图。

[0112] 详细描述

[0113] 在用于DNA测序的一些样品制备方法中,每个模板在插入物任一末端包含衔接子并且通常需要很多步骤以既修饰DNA或RNA又纯化期望的修饰反应产物。这些步骤通常在将适合的片段添加至流通池之前,在溶液中进行,在流通池中适合的片段通过引物延伸反应被偶联至表面,所述引物延伸反应将杂交的片段拷贝至共价地附连至表面的引物的末端上。然后这些‘接种(seeding)’模板通过数个扩增循环产生拷贝的模板的单克隆簇。然而,如其内容以其整体并入本文的U.S.2010/0120098中公开的,在溶液中将DNA转化为衔接子修饰的模板以准备用于簇形成和测序所需的步骤数目可通过使用转座酶介导的片段化和加标签(本文中称为标签化)来最小化。例如,标签化可用于片段化DNA,例如,如在用于Nextera™ DNA样品制备试剂盒(Illumina, Inc.)的工作流程中例示的,其中基因组DNA可通过工程化转座体来片段化,该工程化转座体同时片段化输入DNA并将输入DNA加标签,从而创建在片段末端包含独特连接物序列的片段化核酸分子群体。然而,对表现出改进的插入偏倚的转座酶存在需求。

[0114] 因此,本文提供了用于改进核酸样品的片段化和加标签的转座酶。本发明人已出人意料地鉴定了表现出改进的插入序列偏倚并具有许多其他相关优势的某些改变的转座酶。本文提供的改变的转座酶的一个实施方案为表现出改进的插入偏倚的转座酶。

[0115] 如本文使用的,术语“归一化转座体活性”指25 ng gDNA输入在50μl反应物中产生

以下的生物分析仪片段尺寸分布:曲线下总面积的的转座体的最小浓度:100-300bp=20%-30%;301-600bp=30%-40%;601-7,000bp=30-40%;100-7,000bp $\geq$ 90%。

[0116] 该最小浓度被称为1X。

[0117] 如在整个申请中使用的,转座体的浓度与归一化活性可互换使用。另外,如在整个申请中使用的,转座体的浓度与转座酶的浓度可互换使用。

[0118] 如本文使用,术语“插入偏倚”指转座酶对插入位点的序列偏好。例如,如果多核苷酸样品中的A/T/C/G的背景频率为平均分布的(25%A、25%T、25%C、25%G),则一种核苷酸相对于其他三种核苷酸在转座酶结合位点或切割位点处的任何过呈现(over-representation)反映在该位点处的插入偏倚。插入偏倚可使用本领域已知的多种方法中的任一种来测量。例如,可对插入位点进行测序,并且可将插入位点中的每个位置处的任何特定核苷酸的相对丰度进行比较,如下文实施例1中一般性陈述的。

[0119] “插入偏倚的改进”指在改变的转座酶的结合位点的一个或更多个位置处的特定碱基的频率被降低或增加以更接近该碱基在多核苷酸样品中的背景频率。改进可以是在该位置处的频率相对于在未改变的转座酶的该位置中的频率的增加。可选地,改进可以是在该位置处的频率相对于在未改变的转座酶的该位置中的频率的降低。因此,例如,如果多核苷酸样品中的T核苷酸的背景频率为0.25,且改变的转座酶将在转座酶结合位点中的特定位置处的T核苷酸的频率从高于0.25的频率降低至更接近0.25的频率,改变的转座酶具有插入偏倚的改进。类似地,例如,如果多核苷酸样品中的T核苷酸的背景频率为0.25,且改变的转座酶将在转座酶结合位点中的特定位置处的T核苷酸的频率从低于0.25的频率增加至更接近0.25的频率,改变的转座酶具有插入偏倚的改进。

[0120] 测量插入偏倚的一种方法是通过插入位点的大规模测序并测量在相对于插入位点的结合位点中的每个位置处的碱基的频率,如例如在Green等Mobile DNA (2012) 3:3中描述的,其通过引用以其整体并入本文。显示在每个位置处的丰度的典型工具为强度对循环分布图,例如,如在图2中示出的。如下文实施例1中描述的,通过转座子介导的加标签和片段化产生的片段末端可以以大规模测序,并且在插入位点的每个位置处的碱基分布频率可被测量以检测插入位点的一个或更多个位置处的偏倚。因此,例如,如图2中所示,位置(1)处对于‘G’核苷酸为0.55和对于‘A’核苷酸为0.16的频率的碱基分布反映了在该位置处对于G的强烈偏好和远离A的偏倚。作为另一个实例,且相比之下,如图3中示出的,在位置(20)处四种碱基中的每一种的碱基分布基本为0.25反映在该位置很少或没有序列偏倚。

[0121] 在本文提供的一些实施方案中,改变的转座酶提供在位于插入位点的上游或下游1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个或多于20个碱基的一个或更多个位点处的插入偏倚的降低。在一些实施方案中,改变的转座酶提供在位于插入位点的下游1个至15个碱基的一个或更多个位点处的插入偏倚的降低。在一些实施方案中,改变的转座酶提供在位于插入位点的上游1个至15个碱基的一个或更多个位点处的插入偏倚的降低。

[0122] 如下文更详细描述,本发明人已出人意料地发现,对在转座酶氨基酸序列的某些位置处的残基的一个或更多个突变导致在转座事件期间改进序列插入偏倚。这些改变的转座酶在高多样性和低多样性核酸样品的标签化中提供改进的性能,导致更大的覆盖度均匀性和被测序的多个区域的更少的脱落。

[0123] 如本文使用的,术语“DNA输入耐受性”指转座酶跨越一定范围的输入DNA量产生均一DNA片段尺寸的能力。

[0124] 如本文使用的,延伸因子的符号:TS与Tsf可互换使用。

[0125] 在一些实施方案中,输入DNA为基因组DNA。在一些实施方案中,输入DNA的范围可为0.001 μ g至1mg、1ng至1mg、1ng至900ng、1ng至500ng、1ng至300ng、1ng至250ng、1ng至100ng、5ng至250ng或5ng至100ng,且转座酶的浓度在5nM和500nM之间。在一些实施方案中,对于以上提及的输入DNA范围的转座酶的浓度为约25nM、30nM、35nM、40nM、50nM、60nM、65nM、70nM、75nM、80nM、90nM、95nM、100nM、125nM、130nM、140nM、150nM、175nM、180nM、190nM、200nM、210nM、225nM、230nM、240nM、250nM、260nM、275nM、280nM、290nM、300nM、325nM、350nM、360nM、375nM、380nM、390nM、400nM、425nM、450nM、475nM或500nM。在一些实施方案中,对于以上提及的输入DNA范围的转座酶的归一化浓度或归一化转座酶的浓度选自以下的范围:约0.1X至10X、1X至10X、3X至8X、4X至7X。在一些实施方案中,对于以上提及的输入DNA范围的转座酶的归一化浓度或归一化转座酶为约0.1X、0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1X、1.1X、1.2X、1.3X、1.4X、1.5X、1.6X、1.7X、1.8X、1.9X、2X、2.1X、2.2X、2.3X、2.4X、2.5X、2.6X、2.7X、2.8X、2.9X、3X、3.1X、3.2X、3.3X、3.4X、3.5X、3.6X、3.7X、3.8X、3.9X、4X、4.1X、4.2X、4.3X、4.4X、4.5X、4.6X、4.7X、4.8X、4.9X、5X、5.1X、5.2X、5.3X、5.4X、5.5X、5.6X、5.7X、5.8X、5.9X、6X、6.1X、6.2X、6.3X、6.4X、6.5X、6.6X、6.7X、6.8X、6.9X、7X或7.5X、8X、8.5X、9X、9.5X、10X。

[0126] 在一些实施方案中,输入DNA的量为1ng、2ng、3ng、4ng、5ng、6ng、7ng、8ng、9ng、10ng、11ng、12ng、13ng、15ng、20ng、25ng、30ng、35ng、40ng、45ng、50ng、55ng、60ng、65ng、70ng、75ng、80ng、85ng、90ng、95ng、100ng、110ng、115ng、120ng、125ng、130ng、135ng、140ng、150ng、155ng、160ng、165ng、170ng、180ng、185ng、190ng、195ng、200ng、210ng、220ng、225ng、230ng、235ng、240ng、245ng、250ng、260ng、270ng、280ng、290ng、300ng、325ng、350ng、375ng、400ng、425ng、450ng、475ng、500ng、525ng、550ng、600ng、650ng、700ng、750ng、800ng、850ng或900ng。在一些实施方案中,对于以上提及的输入DNA量的转座酶的浓度为约25nM、30nM、35nM、40nM、50nM、60nM、65nM、70nM、75nM、80nM、90nM、95nM、100nM、125nM、130nM、140nM、150nM、175nM、180nM、190nM、200nM、210nM、225nM、230nM、240nM、250nM、260nM、275nM、280nM、290nM、300nM、325nM、350nM、360nM、375nM、380nM、390nM、400nM、425nM、450nM、475nM或500nM。在一些实施方案中,对于以上提及的输入DNA量的转座酶的归一化浓度或归一化转座酶的浓度选自以下的范围:约0.1X至10X、1X至10X、3X至8X、4X至7X。在一些实施方案中,对于以上提及的输入DNA量的转座酶的归一化浓度或归一化转座酶为约0.1X、0.2X、0.3X、0.4X、0.5X、0.6X、0.7X、0.8X、0.9X、1X、1.1X、1.2X、1.3X、1.4X、1.5X、1.6X、1.7X、1.8X、1.9X、2X、2.1X、2.2X、2.3X、2.4X、2.5X、2.6X、2.7X、2.8X、2.9X、3X、3.1X、3.2X、3.3X、3.4X、3.5X、3.6X、3.7X、3.8X、3.9X、4X、4.1X、4.2X、4.3X、4.4X、4.5X、4.6X、4.7X、4.8X、4.9X、5X、5.1X、5.2X、5.3X、5.4X、5.5X、5.6X、5.7X、5.8X、5.9X、6X、6.1X、6.2X、6.3X、6.4X、6.5X、6.6X、6.7X、6.8X、6.9X、7X或7.5X、8X、8.5X、9X、9.5X、10X。

[0127] 在一些实施方案中,转座酶的nM浓度与输入DNA的ng量的比率为约0.5至5、1至5、2至5、2.1至3、或2.1至2.5。

[0128] 如本文使用的,术语“基因组DNA”指存在于包含编码多种细胞蛋白的一个或更多

个基因的细胞中的核酸。在一些实施方案中,基因组DNA来自原核生物体,例如,细菌和古细菌。在一些实施方案中,基因组DNA来真核生物体,例如,人、植物、真菌、变形虫。

[0129] 如本文使用的术语“突变体”或“修饰的”指当与野生型基因或基因产物相比时,显示序列和/或功能特性的改变(即,改变的特征)的基因或基因产物。“突变体”或“修饰的”还指当与野生型基因或基因产物相比时显示序列和/或功能特性中的改变(即,改变的特征)的在一个或多个特定核苷酸位置处的序列或在一个或多个特定密码子位置处的序列或在一个或多个特定氨基酸位置处的序列。

[0130] 如本文使用的“包括(Including)”具有与术语包含(comprising)相同的含义。

[0131] 如本文使用的“约”意指在定量方面加或减10%。

[0132] 如下文更详细描述,本發明人已出人意料地发现,对在转座酶氨基酸序列的某些位置处的残基的一个或多个突变导致增加的DNA输入耐受性,使得与野生型转座酶相比,突变体转座酶跨越一定范围的输入DNA量产生均一DNA片段尺寸。在一个实施方案中,与其他转座酶,例如,TS-Tn5和Tn5版本1(TDE1)相比,TS-Tn5059转座酶表现出增加的DNA输入耐受性。

[0133] 在一些实施方案中,与其他转座酶相比,TS-Tn5059表现出增加的DNA输入耐受性,其中输入DNA的范围在1ng至200ng基因组DNA之间,且TS-Tn5059的浓度在100-300nM之间。在其中与其他转座酶相比TS-Tn5059表现出增加的DNA输入耐受性的一些实施方案中,输入DNA的范围在5ng至200ng基因组DNA之间,且TS-Tn5059的浓度在100-250nM之间。在其中与其他转座酶相比TS-Tn5059表现出增加的DNA输入耐受性的一些实施方案中,输入DNA的范围在5ng至100ng基因组DNA之间,且TS-Tn5059的浓度在240nM和250nM之间。

[0134] 改进的插入偏倚连同增加的DNA输入耐受性提供比目前的Nextera®快速捕获方案(Illumina, Inc.)更快和更灵活的样品制备和外显子组富集方案。

[0135] 如本文使用的,术语“标签化(tagmentation)”指DNA通过包含与包含转座子末端序列的衔接子复合的转座酶的转座体复合物的修饰。标签化导致同时的DNA的片段化和衔接子与双链体片段的两条链的5'末端的连接。

[0136] 如本文使用的,“转座体复合物”或“转座体”包含至少一个转座酶和一个转座酶识别位点。在一些此类系统中,转座酶可与能够催化转座反应的转座子识别位点形成功能性复合物。转座酶可与转座酶识别位点结合,并在本文称为标签化的过程中将转座酶识别位点插入进靶核酸中。在一些此类插入事件中,转座酶识别位点的一条链可被转移进靶核酸中。

[0137] 本文提供的改变的转座酶可形成转座体复合物的一部分。示例性转座复合物包含,但不限于,高活性Tn5转座酶和Tn5型转座酶识别位点。高活性Tn5转座酶可包括在美国专利第5,925,545号、美国专利第5,965,443号、美国专利第7,083,980号和美国专利第7,608,434号以及Goryshin和Reznikoff, J. Biol. Chem., 273:7367 (1998) 中描述的那些,其每一个的内容通过引用以其整体并入本文。然而,应当理解,本文提供的改变的转座酶可用于能够以足以为其预期目的对靶核酸加标签的效率以随机或几乎随机的方式插入转座子末端的可用于本文提供的方法中的任何转座系统中。

[0138] 例如,提供的改变的转座酶可包括在功能上等同于Tn5氨基酸序列中的位点的一个或多个位置处的至少一个氨基酸置换突变。与Tn5同源的区域在本文中示出(如图1中

例示),并允许鉴定其他转座酶中的功能上等同的位点,其他转座酶例如,Hermes转座酶、HIV整合酶、Mu转座酶和Mos1转座酶。同样地,其他转座酶或整合酶中的功能上等同的位点对于本领域普通技术人员来说将通过例如进行Tn5氨基酸序列的序列比对并鉴定保守或半保守的残基或结构域而是明显的。因此,应当理解,可用于本文提供的某些实施方案的转座系统包括具有与Tn5的那些位点功能上等同的位点的任何已知的转座酶。例如,此类系统可包括MuA转座酶和包含R1和R2末端序列的Mu转座酶识别位点(Mizuuchi,K.,Cell,35:785,1983;Savilahti,H,等,EMBO J.,14:4893,1995)。

[0139] 被包括在本文提供的某些实施方案中的转座系统的更多实例包括金黄色葡萄球菌(*Staphylococcus aureus*)Tn552(Colegio等,J.Bacteriol.,183:2384-8,2001;Kirby C等,Mol.Microbiol.,43:173-86,2002)、Ty1(Devine&Boeke,Nucleic Acids Res.,22:3765-72,1994及国际公布W0 95/23875)、转座子Tn7(Craig,N L,Science.271:1512,1996;Craig,N L,于:Curr Top Microbiol Immunol.中的综述,204:27-48,1996)、Tn/0和IS10(Kleckner N,等,Curr Top Microbiol Immunol.,204:49-82,1996)、Mariner转座酶(Lampe D J,等,EMBO J.,15:5470-9,1996)、Tc1(Plasterk R H,Curr.Topics Microbiol.Immunol.,204:125-43,1996)、P元件(Gloor,G B,Methods Mol.Biol.,260:97-114,2004)、Tn3(Ichikawa&Ohtsubo,J Biol.Chem.265:18829-32,1990)、细菌插入序列(Ohtsubo&Sekine,Curr.Top.Microbiol.Immunol.204:1-26,1996)、逆转录病毒(Brown,等,Proc Natl Acad Sci USA,86:2525-9,1989)、以及酵母的逆转录转座子(Boeke&Corces,Annu Rev Microbiol.43:403-34,1989)。更多实例包括IS5、Tn10、Tn903、IS911以及转座酶家族酶的工程化版本(Zhang等,(2009)PLoS Genet.5:e1000689.Epub 2009Oct.16;Wilson C.等(2007)J.Microbiol.Methods 71:332-5)。以上引用的参考文献通过引用以其整体并入本文。

[0140] 简言之,“转座反应”为其中一个或更多个转座子在随机位点或几乎随机位点处被插入进靶核酸的反应。转座反应中的基本组分为转座酶和呈现转座子的核苷酸序列的DNA寡核苷酸以及形成功能性转座复合物或转座体复合物需要的其他组分,所述转座子的核苷酸序列包括转移的转座子序列及其互补物(即,非转移的转座子末端序列)。根据需要或期望,DNA寡核苷酸还可包含另外的序列(例如,衔接子或引物序列)。

[0141] 被添加至核酸的5'和/或3'末端的连接物可包括通用序列。通用序列为两个或更多个核酸分子共有,即共享的核苷酸序列的区域。任选地,两个或更多个核酸分子还具有序列差异的区域。因此,例如,5'连接物可包含相同或通用的核酸序列,且3'连接物可包含相同或通用的序列。在多种核酸分子的不同成员中可存在的通用序列可允许使用与通用序列互补的单个通用引物复制或扩增多种不同序列。

[0142] 转座酶突变体

[0143] 因此,本文提供了相对于野生型转座酶修饰的突变体转座酶。改变的转座酶可包含在功能上等同于下表1中列出的那些残基的一个或更多个位置处的至少一个氨基酸置换突变。表1列出了在已显示导致改进的插入偏倚的转座酶残基处的置换突变。如表1中列出的,本文提供的置换突变可处于任何功能性转座酶骨架中,诸如本文例示为SEQ ID NO:1的野生型Tn5转座酶、或具有对其他位点的另外突变的转座酶,该另外突变包括在称为高活性Tn5转座酶的转座酶序列中发现的那些突变,诸如,例如在美国专利第5,925,545号、美国专

利第5,965,443号、美国专利第7,083,980号及美国专利第7,608,434号的并入材料中示出的一种或更多种突变,以及如本文作为SEQ ID NO:11例示的。

[0144] 表1-导致改进的插入偏倚的突变的实例

[0145]

骨架	Tn5 突变体	SEQ ID NO:
Tn5 WT	D248Y D248T D248K D248S D248L D248A D248W D248P D248G D248R D248F D248H	2
Tn5 WT	D119L D119M D119S	3

[0146]

	D119A D119V	
Tn5 WT	W125M	4
Tn5 WT	iaD248	5
Tn5 WT	K120F K120Y K120E K120W	6
Tn5 WT	D248 至 Y、T、K、S、L、A、W、P、G、 R、F 或 H D119 至 L、M、S、A 或 V W125M K120F	7
Tn5 WT	K212R P214R A338V	8
Tn5 WT	K212R P214R G251R A338V	9
Tn5 WT	E146Q E190G G251R	10
Tn5 高活性	D248Y D248T D248K D248S D248L D248A D248W D248P D248G D248R D248F D248H	12
Tn5 高活性	D119L D119M D119S D119A D119V	13
Tn5 高活性	W125M	14
Tn5 高活性	iaD248	15
Tn5 高活性	K120F K120Y K120E K120W	16
Tn5 高活性	D248 至 Y、T、K、S、L、A、W、P、G、 R、F 或 H D119 至 L、M、S、A 或 V W125M K120F	17
Tn5 高活性	K212R P214R A338V	18
Tn5 高活性	K212R	19

[0147]		P214R G251R A338V	
	Tn5 高活性	E146Q E190G G251R	20

[0148] 如本领域理解的,以上表中列出的参考编号指野生型Tn5序列(SEQ ID NO:1)的氨基酸位置。本领域普通技术人员将理解编号可由于N-末端截短、插入或融合而改变。尽管位置编号可能已改变,以上列出的氨基酸的功能位置将保持相同。例如,在SEQ ID NO:25中列出的序列的前285个氨基酸残基包含大肠杆菌(E.coli)TS的N-末端融合,随后是SEQ ID NO:11的氨基酸残基2-476。如此,例如,SEQ ID NO:25的Pro 656功能上对应于SEQ ID NO:11的Pro 372。

[0149] 如此,在某些实施方案中,相对于野生型转座酶,本文提供的改变的转座酶包含在功能上等同于,例如,Tn5转座酶氨基酸序列中的Asp248、Asp119、Trp125、Lys120、Lys212、Pro214、Gly251、Ala338、Glu146和/或Glu190的一个或多个位置处的至少一个氨基酸置换突变。

[0150] 在一些实施方案中,突变体转座酶可包含在位置Asp248处的突变。在位置Asp248处的突变可以为,例如,置换突变或插入突变。在某些实施方案中,突变为对除了Asp以外的任何残基的置换突变。在某些实施方案中,在位置Asp248处的置换突变包括向选自以下组成的组的残基的突变:Tyr、Thr、Lys、Ser、Leu、Ala、Trp、Pro、Gln、Arg、Phe及His。

[0151] 在某些实施方案中,在位置Asp248处的突变为在位置Asp248之后的插入突变。在某些方面,插入突变可包括在Asp248之后插入任何残基。在某些方面,插入突变可包括在位置Asp248之后插入疏水性残基。疏水性残基是本领域技术人员已知的,并包括例如,Val、Leu、Ile、Phe、Trp、Met、Ala、Tyr及Cys。在某些方面,插入突变可包括在位置Asp248之后插入缬氨酸残基。

[0152] 本文提供的一些实施方案包括相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Asp119处的突变。在某些方面,在位置Asp119处的突变为置换突变。在某些方面,在位置Asp119处的置换突变可包括向疏水性残基的突变。疏水性残基是本领域技术人员已知的,并包括,例如,Val、Leu、Ile、Phe、Trp、Met、Ala、Tyr及Cys。在某些方面,在位置Asp119处的置换突变可包括向亲水性残基的突变。亲水性残基是本领域技术人员已知的,并包括,例如,Arg、Lys、Asn、His、Pro、Asp及Glu。在某些方面,在位置Asp119处的置换突变可包括向选自以下组成的组的残基的突变:Leu、Met、Ser、Ala及Val。

[0153] 本文提供的一些实施方案包括相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Trp125处的突变。在某些方面,在位置Trp125处的突变为置换突变。在某些方面,在位置Trp125处的置换突变可包括向甲硫氨酸残基的突变。

[0154] 本文提供的一些实施方案包括相对于野生型Tn5转座酶修饰的突变体Tn5转座酶,所述突变体转座酶包含在位置Lys120处的突变。在某些方面,在位置Lys120处的突变为置换突变。在某些方面,在位置Lys120处的置换突变可包括向大体积芳香族残基的突变。被表征为大体积芳香族残基的残基是本领域技术人员已知的,并包括,例如Phe、Tyr和Trp。在某些方面,在位置Lys120处的置换突变可包括向选自以下组成的组的残基的突变:Tyr、

Phe、Trp及Glu。

[0155] 本文提供的一些实施方案包括相对于野生型Tn5转座酶修饰的突变体Tn5转座酶，所述突变体转座酶包含在位置Lys212和/或Pro214和/或Ala338处的突变。在某些方面，在位置Lys212和/或Pro214和/或Ala338处的一个或更多个突变为置换突变。在某些方面，在位置Lys212处的置换突变包括向精氨酸的突变。在某些方面，在位置Pro214处的置换突变包括向精氨酸的突变。在某些方面，在位置Ala338处的置换突变包括向缬氨酸的突变。在一些实施方案中，转座酶还可包含在Gly251处的置换突变。在某些方面，在位置Gly251处的置换突变包括向精氨酸的突变。

[0156] 本文提供的一些实施方案包括相对于野生型Tn5转座酶修饰的突变体Tn5转座酶，所述突变体转座酶包含在位置Glu146和/或Glu190和/或Gly251处的突变。在某些方面，在位置Glu146和/或Glu190和/或Gly251处的一个或更多个突变为置换突变。在某些方面，在位置Glu146处的置换突变可包括向谷氨酰胺的突变。在某些方面，在位置Glu190处的置换突变可包括向甘氨酸的突变。在某些方面，在位置Gly251处的置换突变可包括向精氨酸的突变。

[0157] 在任何以上描述的实施方案中，突变体Tn5转座酶还可包含在功能上等同于Tn5转座酶氨基酸序列中的Glu54和/或Met56和/或Leu372的位置处的置换突变。在某些实施方案中，转座酶包含与Tn5转座酶氨基酸序列中的Glu54Lys和/或Met56Ala和/或Leu372Pro同源的置换突变。

[0158] 本文提供的一些实施方案包括包含SEQ ID NO:2-10和12-20中任一个的氨基酸序列的突变体Tn5转座酶。

[0159] 本文还提供了改变的转座酶，所述改变的转座酶包含对半保守结构域的置换突变。如本文使用的，术语“半保守结构域”指在多种转座酶间和/或在多种物种间完全保守或至少部分保守的转座酶的部分。半保守结构域包含位于转座酶的催化核心结构域中的氨基酸残基。已出人意料地发现，半保守结构域中的一个或更多个残基的突变影响转座酶活性，导致插入偏倚的改进。

[0160] 在一些实施方案中，半保守结构域包含具有SEQ ID NO:21列出的序列的氨基酸。SEQ ID NO:21对应于在本文作为SEQ ID NO:1列出的Tn5转座酶氨基酸序列的残基124-133。显示在半保守结构域中的多种转座酶间的保守性的结构比对示于图1中。图1中示出的转座酶序列包括Tn5转座酶(1MUH)、Hermes转座酶(2BW3)、HIV整合酶(1ITG)、Mu转座酶(1BCM)及Mos1转座酶(3HOS)的催化核心结构域。

[0161] 已出人意料地发现对半保守结构域中的一个或更多个残基的突变导致插入偏倚的改进。例如，在本文提供的改变的转座酶的一些实施方案中，置换突变包含在SEQ ID NO:21的位置2处向除了Trp以外的任何残基的突变。在某些实施方案中，改变的转座酶包含在SEQ ID NO:21的位置2处向Met的突变。

[0162] “功能上等同”意指在使用完全不同的转座酶的研究的情况下，对照转座酶将包含被认为出现于其他转座酶中在酶方面具有相同功能作用的氨基酸位置处的氨基酸置换。作为实例，在Mu转座酶中在位置288处从赖氨酸至甲硫氨酸的突变(K288M)将在功能上等同于在Tn5转座酶中在位置125处从色氨酸至甲硫氨酸的置换(W125M)。

[0163] 通常，在两个或更多个不同转座酶中的功能上等同的置换突变发生在转座酶的氨

氨基酸序列中的同源氨基酸位置处。因此,本文使用的术语“功能上等同”还包括与给定突变“位置上等同”或“同源”的突变,而不管突变的氨基酸的特定功能是否是已知的。基于序列比对和/或分子建模鉴定两种或更多种不同转座酶的氨基酸序列中的位置上等同的或同源的氨基酸残基是可能的。鉴定位置上等同和/或功能上等同的残基的序列比对和分子建模的实例示于图1中。因此,例如,如图1中示出的,半保守结构域中的残基被鉴定为Tn5转座酶氨基酸序列的位置124-133。Hermes转座酶、HIV整合酶、Mu转座酶和Mos1转座酶转座酶中的对应残基在图中被按垂直对齐标识,并且被认为位置上等同于以及功能上等同于Tn5转座酶氨基酸序列中的对应残基。

[0164] 上文描述的改变的转座酶可包含已知增强转座酶活性的一个或更多个方面的另外的置换突变。例如,在一些实施方案中,除了以上突变的任一种以外,改变的Tn5转座酶还可包含在功能上等同于Tn5转座酶氨基酸序列中的Glu54和/或Met56和/或Leu372的位置处的置换突变。可做出在功能上等同于Tn5转座酶氨基酸序列中的Glu54和/或Met56和/或Leu372的位置的一个或更多个位置处的导致改进的活性的多种置换突变的任一种,如本领域已知的和通过美国专利第5,925,545号、美国专利第5,965,443号、美国专利第7,083,980号和美国专利第7,608,434号的公开内容以及在Goryshin和Reznikoff, J. Biol. Chem., 273:7367 (1998) 的公开内容中例示的,其每一个的内容通过引用以其整体并入。在一些实施方案中,转座酶包含与Tn5转座酶氨基酸序列中的Glu54Lys和/或Met56Ala和/或Leu372Pro同源的置换突变。例如,置换突变可包括与Tn5转座酶氨基酸序列中的Glu54Lys和/或Met56Ala和/或Leu372Pro同源的置换突变。

[0165] 突变转座酶

[0166] 根据例如如以上讨论的转座酶模式和模式预测或者使用随机或半随机突变方法在本公开内容中任选地使用多种类型的诱变以,例如,修饰转座酶以产生变体。通常,任何可用的诱变程序可用于制备转座子突变体。此类诱变程序任选地包括针对感兴趣的一种或更多种活性(例如,改进的插入偏倚)选择突变核酸和多肽。可使用的程序包括,但不限于:定点诱变、随机点诱变、体外或体内同源重组(DNA改组和组合重叠PCR)、使用包含尿嘧啶模板的诱变、寡核苷酸定向诱变、硫代磷酸酯修饰的DNA诱变、使用缺口双链体DNA的诱变、点错配修复、使用修复缺陷型宿主菌株的诱变、限制性选择和限制性纯化、缺失诱变、通过全基因合成的诱变、简并PCR、双链断裂修复、以及本领域技术人员已知的许多其他程序。用于突变的起始转座酶可以为本文指出的任何转座酶,包括可用的转座酶突变体,诸如,例如在美国专利第5,925,545号、美国专利第5,965,443号、美国专利第7,083,980号和美国专利第7,608,434号以及Goryshin和Reznikoff, J. Biol. Chem., 273:7367 (1998) 的公开内容中鉴定的那些,其每一个的内容通过引用以其整体并入。

[0167] 任选地,诱变可通过来自天然存在的转座酶分子或已知的改变的或突变的转座酶(例如,使用如前述参考文献中提及的现有突变体转座酶)的已知信息,例如,序列、序列比较、物理特性、晶体结构和/或类似信息的来指导。然而,在另一类实施方案中,修饰可以是基本上随机的(例如,如在经典或“家族”DNA改组中,参见,例如,Cramer等(1998) “DNA shuffling of a family of genes from diverse species accelerates directed evolution” Nature 391:288-291)。

[0168] 有关突变形式的另外的信息见于:Sambrook等, Molecular Cloning--A

Laboratory Manual (第3版), 第1-3卷, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2000 ("Sambrook"); Current Protocols in Molecular Biology, F.M. Ausubel等, 编著, Current Protocols, a joint venture between Greene Publishing Associates, Inc. 和 John Wiley & Sons, Inc., (supplemented through 2011) ("Ausubel")) 以及 PCR Protocols A Guide to Methods and Applications (Innis等编著) Academic Press Inc. San Diego, Calif. (1990) ("Innis")。在引用的以下出版物和参考文献中提供了关于突变形式的另外细节: Arnold, Protein engineering for unusual environments, Current Opinion in Biotechnology 4:450-455 (1993); Bass等, Mutant Trp repressors with new DNA-binding specificities, Science 242:240-245 (1988); Bordo and Argos (1991) Suggestions for "Safe" Residue Substitutions in Site-directed Mutagenesis 217:721-729; Botstein & Shortle, Strategies and applications of in vitro mutagenesis, Science 229:1193-1201 (1985); Carter等, Improved oligonucleotide site-directed mutagenesis using M13 vectors, Nucl. Acids Res. 13:4431-4443 (1985); Carter, Site-directed mutagenesis, Biochem. J. 237:1-7 (1986); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, Methods in Enzymol. 154:382-403 (1987); Dale等, Oligonucleotide-directed random mutagenesis using the phosphorothioate method, Methods Mol. Biol. 57:369-374 (1996); Eghtedarzadeh & Henikoff, Use of oligonucleotides to generate large deletions, Nucl. Acids Res. 14:5115 (1986); Fritz等, Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, Nucl. Acids Res. 16:6987-6999 (1988); Grundstrom等, Oligonucleotide-directed mutagenesis by microscale "shot-gun" gene synthesis, Nucl. Acids Res. 13:3305-3316 (1985); Hayes (2002) Combining Computational and Experimental Screening for rapid Optimization of Protein Properties PNAS 99 (25) 15926-15931; Kunkel, The efficiency of oligonucleotide directed mutagenesis, in Nucleic Acids & Molecular Biology (Eckstein, F. 和 Lilley, D.M.J. 编著, Springer Verlag, Berlin)) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel等, Rapid and efficient site-specific mutagenesis without phenotypic selection, Methods in Enzymol. 154, 367-382 (1987); Kramer等, The gapped duplex DNA approach to oligonucleotide-directed mutation construction, Nucl. Acids Res. 12:9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, Methods in Enzymol. 154:350-367 (1987); Kramer等, Point Mismatch Repair, Cell 38:879-887 (1984); Kramer等, Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, Nucl. Acids Res. 16:7207 (1988); Ling等, Approaches to DNA mutagenesis: an overview, Anal Biochem. 254(2): 157-178 (1997); Lorimer 和 Pastan Nucl. Acids Res. 23, 3067-8 (1995); Mandecki, Oligonucleotide-directed double-strand break repair in plasmids of

Escherichia coli: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181 (1986); Nakamaye&Eckstein, Inhibition of restriction endonuclease *Nci* I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 14:9679-9698 (1986); Nambiar等, Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223:1299-1301 (1984); Sakamar和Khorana, Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14:6361-6372 (1988); Sayers等, Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers等, Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) *Nucl. Acids Res.* 16:803-814; Sieber, 等, *Nature Biotechnology*, 19:456-460 (2001); Smith, In vitro mutagenesis, *Ann. Rev. Genet.* 19:423-462 (1985); *Methods in Enzymol.* 100:468-500 (1983); *Methods in Enzymol.* 154:329-350 (1987); Stemmer, *Nature* 370, 389-91 (1994); Taylor等, The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13:8749-8764 (1985); Taylor等, The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13:8765-8787 (1985); Wells等, Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317:415-423 (1986); Wells等, Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Zoller&Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller&Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983); Zoller&Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, *Methods in Enzymol.* 154:329-350 (1987); Clackson等 (1991) "Making antibody fragments using phage display libraries" *Nature* 352:624-628; Gibbs等 (2001) "Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling" *Gene* 271:13-20; 以及 Hiraga和Arnold (2003) "General method for sequence-independent site-directed chimeragenesis" *J. Mol. Biol.* 330:287-296. 关于许多以上方法的另外细节可在 *Methods in Enzymology* 第154卷中发现, 其还描述了用多种诱变方法解决问题的有用控制。

[0169] 制备和分离重组转座酶

[0170] 通常, 编码如本文提供的转座酶的核酸可通过克隆、重组、体外合成、体外扩增和/或其他可用的方法制备。多种重组方法可用于表达编码如本文提供的转座酶的表达载体。

用于制备重组核酸、表达和分离表达产物的方法是本领域熟知和描述的。本文描述了许多示例性突变和突变组合,以及用于设计期望的突变的策略。用于制备和选择转座酶的催化结构域中的突变的方法见于本文并在美国专利第5,925,545号、美国专利第5,965,443号、美国专利第7,083,980号和美国专利第7,608,434号中例示,其通过引用以其整体并入。

[0171] 用于突变、重组和体外核酸操作方法(包括克隆、表达、PCR等)的另外有用的参考文献包括Berger和Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*第152卷Academic Press, Inc., San Diego, Calif. (Berger); Kaufman等(2003) *Handbook of Molecular and Cellular Methods in Biology and Medicine*第二版Ceske (编著)CRC Press (Kaufman); 以及The *Nucleic Acid Protocols Handbook* Ralph Rapley (编著) (2000) Cold Spring Harbor, Humana Press Inc (Rapley); Chen等(编著) *PCR Cloning Protocols*, 第二版(*Methods in Molecular Biology*, 第192卷)Humana Press; 以及在Viljoen等(2005) *Molecular Diagnostic PCR Handbook* Springer, ISBN 1402034032 中。

[0172] 另外,用于从细胞纯化质粒或其他相关核酸的大量试剂盒商购可得(参见,例如,两者都来自Pharmacia Biotech的EasyPrep™、FlexiPrep™;来自Stratagene的StrataClean™;以及来自Qiagen的QIAprep™)。任何分离的和/或纯化的核酸可被进一步操作以产生其他核酸,被用来转染细胞、被掺入进相关载体中以感染生物体用于表达等。典型的克隆载体包含转录和翻译终止子、转录和翻译起始序列、以及对调节特定靶核酸的表达有用的启动子。载体任选地包括包含至少一个独立终止子序列的通用表达盒、允许所述盒在真核生物或原核生物或两者中复制的序列(例如,穿梭载体)以及用于原核系统和真核系统两者的选择标志物。载体适合于在原核生物、真核生物或两者中复制和整合。

[0173] 在一些实施方案中,本文提供的转座酶被表达为融合蛋白。融合蛋白可增强如下特征:例如,转座酶的溶解度、表达和/或纯化。如本文使用的,术语“融合蛋白”指具有至少两个通常不存在于单个天然多肽中的多肽结构域的单条多肽链。因此,天然存在的蛋白及其点突变体不是如本文使用的“融合蛋白”。优选地,感兴趣的多肽经由肽键与至少一个多肽结构域融合,并且融合蛋白还可包括来源于单独蛋白的氨基酸部分之间的氨基酸的连接区域。与感兴趣的多肽融合的多肽结构域可增强感兴趣的多肽的溶解度和/或表达,并且还可提供纯化标签以允许从宿主细胞或培养物上清液或两者纯化重组融合蛋白。在表达、纯化和/或储存期间增加溶解度的多肽结构域是本领域熟知的,并包括,例如,麦芽糖结合蛋白(MBP)和延伸因子Ts (Tsf),如Fox, J.D.和Waugh D.S., *E.coli Gene Expression Protocols Methods in Molecular Biology*, (2003) 205:99-117以及Han等 *FEMS Microbiol. Lett.* (2007) 274:132-138例示的,其每一个通过引用以其整体并入本文。与感兴趣的多肽融合的多肽结构域可被融合在感兴趣的多肽的N-末端或C-末端处。术语“重组”指两个另外单独的序列区域段通过,例如,化学合成或通过由基因工程技术操作氨基酸或核酸的分离的区段的人工组合。

[0174] 在一个实施方案中,本发明提供了包含经修饰的Tn5转座酶和延伸因子Ts (Tsf)的转座酶融合蛋白。Tsf-Tn5融合蛋白可被组装成包含融合转座酶和游离转座子末端的功能性二聚体转座体复合物。与未融合的Tn5蛋白相比,Tsf-Tn5融合蛋白具有增加的溶解度和热稳定性。

[0175] 在另一个实施方案中,本发明提供了包含经修饰的Tn5转座酶和识别5-甲基胞嘧啶的蛋白结构域的转座酶融合蛋白。5-甲基胞嘧啶-Tn5融合蛋白可被组装成包含融合转座酶和游离转座子末端的功能性二聚体转座体复合物。5-甲基胞嘧啶结合蛋白结构域可,例如,用来将Tn5转座体复合物靶向基因组的甲基化区域。

[0176] 在又一个实施方案中,本发明提供了包含经修饰的Tn5转座酶和蛋白A抗体结合结构域的转座酶融合蛋白。蛋白A-Tn5融合蛋白可被组装成包含融合转座酶和游离转座子末端的功能性二聚体转座体复合物。蛋白A的抗体结合结构域可,例如,用来将Tn5转座体复合物靶向基因组的结合抗体的区域。

[0177] 本发明提供了包含经修饰的Tn5转座酶和延伸因子Ts (Tsf) 的全部或部分的转座酶融合蛋白。Tsf为可被用来增强在细菌表达系统,例如,大肠杆菌表达系统中表达的异源蛋白的溶解度的蛋白标签。Tsf增加异源蛋白的溶解度的能力可能是由于Tsf蛋白的固有的高折叠效率。在蛋白纯化实验中(数据未示出),Tsf作为与蛋白Tu的复合物被纯化。为了使Tu在复合物中结合,Tsf需要被正确折叠。Tsf-Tu复合物的纯化表明Tsf被正确折叠。大肠杆菌延伸因子TS的示例性核酸和对应的氨基酸序列分别显示为SEQ ID NO:22和23。

[0178] 在一个实例中,TS-Tn5融合蛋白通过将TS融合至高活性Tn5转座酶的N末端来构建。具有突变体Tn5转座酶蛋白的TS-融合的示例性氨基酸序列分别显示为SEQ ID NO:25和26。SEQ ID NO:24对应于编码SEQ ID NO:25的TS-突变体蛋白融合的核酸序列。

[0179] 尽管TS的全部或部分可被融合在N-末端或C-末端处,本领域技术人员应当理解,接头序列可被插入在TS序列和转座酶的N-末端或C-末端之间。在一些实施方案中,接头序列的长度可以为1个、2个、3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个、14个、15个、16个、17个、18个、19个、20个、21个、22个、23个、24个、25个、26个、27个、28个、29个、30个、31个、32个、33个、34个、35个、36个、37个、38个、39个、40个、41个、42个、43个、44个、45个、46个、47个、48个、49个、50个或更多个氨基酸。在一些实施方案中,融合蛋白的转座酶部分的一个或更多个氨基酸可缺失或用接头序列代替。在一些实施方案中,融合蛋白的转座酶部分的第一个甲硫氨酸可用两个氨基酸,例如,如SEQ ID NO:25和26中所示的Gly-Thr来代替。

[0180] TS-Tn5融合构建体在大肠杆菌中表达并被评价表达、溶解度和热稳定性。TS至Tn5的N-末端的融合增加了Tn5的溶解度。溶解度的增加可与转座体复合物的增加的稳健性和蛋白聚集的减少相关。与未融合的Tn5转座体的热稳定性相比,TS-Tn5转座体的热稳定性被实质上改进。在一个实例中,与未融合的Tn5对照相比,Tn5的热诱导的聚集在Tsf-Tn5融合构建体中被实质上降低。

[0181] 在一个应用中,TS-Tn5融合蛋白用于构建定向RNA-seq文库(例如,TotalScript™ RNA-Seq试剂盒,Illumina),以用于在下一代测序平台(例如,Illumina GA或HiSeq平台)上测序。

[0182] 在另一个应用中,TS-Tn5融合蛋白可用于归一化方法。在另一个应用中,TS溶解标签可用于表达和纯化其他修饰(突变体)Tn5转座酶。

[0183] 对TS融合标签特异性的抗体可用于拉下方法(pull-down process)中,以捕获转座体加标签的序列。在一个实例中,TS融合标签抗体为兔多克隆的。

[0184] 在另一个应用中,TS-Tn5转座体和抗-TS抗体可用于混合转座体方法中。例如,转

座体反应使用Tsf-Tn5转座体和Tn5转座体(即,未用TS加标签)来进行。抗-TS抗体用来特异性地拉下Tsf-Tn5转座体加标签的序列。

[0185] 本发明提供了包含经修饰的Tn5转座酶和识别5-甲基胞嘧啶的蛋白结构域的转座酶融合蛋白。5-甲基胞嘧啶结合蛋白结构域可,例如,用来将Tn5转座体复合物靶向基因组的甲基化区域。在一个应用中,5-甲基胞嘧啶结合结构域-Tn5转座体复合物可用来产生富含甲基的片段化和加标签(标签化)的文库,以用于甲基化分析。

[0186] 在一些实施方案中,与转座酶融合的多肽结构域包含蛋白A的抗体结合结构域。蛋白A为具有稳健折叠特征的相对小的紧密分子。蛋白A的抗体结合结构域可,例如,用来将Tn5转座体复合物靶向基因组的结合抗体的区域。例如,对5-甲基胞嘧啶特异的抗体可用来结合并鉴定基因组的甲基化区域。随后,基因组的结合抗体的区域可使用蛋白A-Tn5融合转座体复合物来靶向,以用于片段化和加标签(即,标签化)。

[0187] 在一些实施方案中,与转座酶融合的多肽结构域包含纯化标签。如本文使用的术语“纯化标签”指适合于纯化或鉴定多肽的任何肽序列。纯化标签特异性结合对纯化标签具有亲和力的另一部分。特异性结合纯化标签的此类部分通常被附连至基质或树脂,诸如琼脂糖珠。特异性结合纯化标签的部分包括抗体、其他蛋白(例如蛋白A或链霉亲和素)、镍离子或钴离子或树脂、生物素、直链淀粉、麦芽糖及环糊精。示例性纯化标签包括将结合金属离子诸如镍离子或钴离子的组氨酸(HIS)标签(诸如六组氨酸肽)。其他示例性纯化标签为myc标签(EQKLISEEDL)、Strep标签(WSHPQFEK)、Flag标签(DYKDDDDK)和V5标签(GKPIPNLLGLDST)。术语“纯化标签”还包括“表位标签”,即被抗体特异性识别的肽序列。示例性表位标签包括被单克隆抗-FLAG抗体特异性识别的FLAG标签。被抗-FLAG抗体识别的肽序列由序列DYKDDDDK或其实质上相同的变体组成。在一些实施方案中,与转座酶融合的多肽结构域包含两个或更多个标签,诸如如下文实施例1中例示的SUMO标签和STREP标签。术语“纯化标签”还包括纯化标签的实质上相同的变体。如本文使用的“实质上相同的变体”指与原始纯化标签相比被修饰(例如经由氨基酸置换、缺失或插入)但保留纯化标签与特异性识别纯化标签的部分特异性结合的特性的纯化标签的衍生物或片段。

[0188] 在一些实施方案中,与转座酶融合的多肽结构域包含表达标签。如本文使用的术语“表达标签”指可被附连至第二多肽并且被认为支持感兴趣的重组多肽的溶解度、稳定性和/或表达的任何肽或多肽。示例性表达标签包括Fc-标签和SUMO-标签。原则上,任何肽、多肽或蛋白可被用作表达标签。

[0189] 其他有用的参考文献,例如用于细胞分离和培养(例如,用于随后的核酸分离)的,包括,Freshney(1994) Culture of Animal Cells, a Manual of Basic Technique, 第三版, Wiley-Liss, New York以及其中引用的参考文献;Payne等(1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley&Sons, Inc. New York, N.Y.; Gamborg和Phillips(编著)(1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) 以及Atlas和Parks(编著) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, Fla。

[0190] 编码本文公开的重组转座酶的核酸也是本文提出的实施方案的特征。特定的氨基酸可由多个密码子编码,并且某些翻译系统(例如,原核细胞或真核细胞)通常表现出密码

子偏倚,例如,不同的生物体通常优先编码相同氨基酸的几个同义密码子中的一个。因此,本文提供的核酸任选地是“密码子优化的”,意指合成核酸以包括被用于表达转座酶的特定翻译系统优选的密码子。例如,当期望在细菌细胞(或甚至特定的细菌菌株)中表达转座酶时,可合成核酸以包括在该细菌细胞的基因组中最频繁发现的密码子,以用于有效表达转座酶。当期望在真核细胞中表达转座酶时,可采用类似的策略,例如,核酸可包括被该真核细胞优选的密码子。

[0191] 多种蛋白分离和检测方法是已知的,并且可用来,例如,从表达本文提供的重组转座酶的细胞的重组培养物分离转座酶。多种蛋白分离和检测方法是本领域熟知的,包括,例如,在以下中陈述的那些:R.Scopes,Protein Purification,Springer-Verlag,N.Y.(1982);Deutscher,Methods in Enzymology第182卷:Guide to Protein Purification,Academic Press,Inc.N.Y.(1990);Sandana(1997)Bioseparation of Proteins,Academic Press,Inc.;Bollag等(1996)Protein Methods,2.Sup.nd Edition Wiley-Liss,NY;Walker(1996)The Protein Protocols Handbook Humana Press,NJ,Harris and Angal(1990)Protein Purification Applications:A Practical Approach IRL Press at Oxford,Oxford,England;Harris and Angal Protein Purification Methods:A Practical Approach IRL Press at Oxford,Oxford,England;Scopes(1993)Protein Purification:Principles and Practice 3.Sup.rd Edition Springer Verlag,NY;Janson and Ryden(1998)Protein Purification:Principles,High Resolution Methods and Applications,第二版Wiley-VCH,NY;以及Walker(1998)Protein Protocols on CD-ROM Humana Press,NJ;以及其中引用的参考文献。关于蛋白纯化和检测方法的另外的细节可见于Satinder Ahuja编著,Handbook of Bioseparations,Academic Press(2000)。

[0192] 使用方法

[0193] 本文提供的改变的转座酶可用于测序程序,诸如体外转座技术。简言之,体外转座可通过使转座体复合物和靶DNA接触来引发。可容易地适于与本公开内容的转座酶一起使用的示例性转座程序和系统,例如,在WO 10/048605;US 2012/0301925;US 2013/0143774中被描述,其每一个通过引用以其整体并入本文。

[0194] 例如,在一些实施方案中,本文提供的转座酶可用于从包含任何感兴趣的dsDNA的靶DNA产生加标签的DNA片段文库(例如,用作下一代测序或扩增模板)的方法,所述方法包括:在体外转座反应中将靶DNA与至少一种转座酶和与所述转座酶形成转座复合物的转座子末端组合物在其中进入靶DNA的多种插入发生的条件下孵育并持续足够的时间,所述转座子末端组合物包含(i)转移链,所述转移链呈现转移的转座子末端序列和任选地,所转移的转座子末端序列的另外序列5'-,和(ii)非转移链,所述非转移链呈现与转移的转座子末端序列互补的序列,多种插入的每一个导致包含所述转移链或由所述转移链组成的第一标签连接至靶DNA中的核苷酸的5'-末端,从而片段化所述靶DNA并产生退火的5'-加标签的DNA片段的群体,所述退火的5'-加标签的DNA片段的每一个在所述5'-末端上具有所述第一标签;以及然后,将5'-加标签的DNA片段的3'-末端连接至所述第一标签或至第二标签,从而产生加标签的DNA片段的文库(例如,包含加标签的环状ssDNA片段或5'-和3'-加标签的DNA片段(或“加双标签的DNA片段”))。

[0195] 在一些实施方案中,用于体外转座反应的转座酶和转座子末端组合物或转座体组

合物的量在每50微升反应物每50纳克靶DNA约1皮摩尔和约25皮摩尔之间。在本发明的任何方法的一些优选的实施方案中,用于体外转座反应的转座酶和转座子末端组合物或转座体组合物的量在每50微升反应物每50纳克靶DNA约5皮摩尔和约50皮摩尔之间。在本发明的任何方法的一些优选的实施方案中,其中转座酶为高活性Tn5转座酶,且转座子末端组合物包含MEDS转座子末端组合物,或其中转座体组合物包含所述高活性Tn5转座酶和包含MEDS转座子末端的转座子末端组合物,用于体外转座反应的所述转座酶和转座子末端组合物或所述转座体组合物的量在每50微升反应物50纳克靶DNA约5皮摩尔和约25皮摩尔之间。在本发明的任何方法的一些优选的实施方案中,其中转座酶为高活性Tn5转座酶或MuA转座酶,用于体外转座反应的转座酶和转座子末端组合物或转座体组合物的终浓度为至少250nM;在一些其他实施方案中,高活性Tn5转座酶或MuA转座酶的终浓度及其各自的转座子末端组合物或转座体组合物的终浓度为至少500nM。

[0196] 在一些实施方案中,本发明提供了一种制备和富集用于外显子组测序的基因组DNA文库的方法。在多种实施方案中,本发明的方法使用改变的Tn5转座酶,例如,TS-Tn5059,用于制备基因组文库。在一个实施方案中,本发明的方法提供了具有降低的偏倚的基因组文库的制备,该降低的偏倚通过改变的转座酶(例如,TS-TN5059)的降低的插入序列偏倚驱动。基因组DNA使用TS-Tn5059的标签化提供了基因组跨越宽GC/AT范围的更完全的基因组覆盖度。

[0197] 在另一个实施方案中,本发明提供了一种使用具有增加的DNA输入耐受性的改变的转座酶制备基因组文库的方法。基因组DNA使用TS-Tn5059的标签化跨越一定范围的DNA输入量提供均匀的插入物尺寸。在一些实施方案中,本发明提供了一种外显子组测序的方法。

[0198] 标签化反应条件

[0199] 本文提供了用于标签化反应的反应条件和缓冲液。在一些实施方案中,二价阳离子被包括在标签化反应缓冲液中。在特定的实施方案中,二价阳离子可以为,例如, Co^{2+} 、 Mn^{2+} 、 Mg^{2+} 、 Cd^{2+} 或 Ca^{2+} 。二价阳离子可以任何适合的盐的形式被包括,诸如氯化物盐,例如, CoCl_2 、 MnCl_2 、 MgCl_2 、乙酸镁、 CdCl_2 或 CaCl_2 。在特定的实施方案中,标签化缓冲液包含 CoCl_2 ,如下文实施例中例示的。如实施例5中的实验证据证明的,在标签化缓冲液制剂中添加 CoCl_2 出人意料地改善了标签化期间的序列偏倚。

[0200] 在某些实施方案中,标签化缓冲液可具有,即,约或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的二价阳离子的浓度,或者为在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。在一些实施方案中,标签化缓冲液可具有,即,约,或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的 CoCl_2 的浓度,或者在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。在一些实施方案中,标签化缓冲液可具有,即,约,或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的 MnCl_2 的浓度,或者在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。在一些

实施方案中,标签化缓冲液可具有,即,约,或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的MgCl₂的浓度,或者在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。在一些实施方案中,标签化缓冲液可具有,即,约,或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的CdCl₂的浓度,或者在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。在一些实施方案中,标签化缓冲液可具有,即,约,或大于0.01mM、0.02mM、0.05mM、0.1mM、0.2mM、0.5mM、1mM、2mM、5mM、8mM、10mM、12mM、15mM、20mM、30mM、40mM、50mM、60mM、70mM、80mM、90mM、100mM的CaCl₂的浓度,或者在任何这些值之间的范围的二价阳离子的浓度,例如,0.01mM至0.05mM、0.02mM至0.5mM、8mM至12mM等等。

[0201] 在一些实施方案中,基因组DNA通过转座酶或标签化反应的片段化可在范围为25℃至70℃、37℃至65℃、50℃至65℃或50℃至60℃的温度下进行。在一些实施方案中,基因组DNA通过转座酶或标签化反应的片段化可在37℃、40℃、45℃、50℃、51℃、52℃、53℃、54℃、55℃、56℃、57℃、58℃、59℃、60℃、61℃、62℃、63℃、64℃或65℃下进行。

[0202] 编码改变的转座酶的核酸

[0203] 本文还提供了核酸分子,所述核酸分子编码本文提供的改变的转座酶。对于为转座酶的突变体版本的任何给定的改变的转座酶,该改变的转座酶的氨基酸序列且优选地还有编码转座酶的野生型核苷酸序列是已知的,根据分子生物学的基本原理获得编码突变体的核苷酸序列是可能的。例如,鉴于编码Tn5转座酶的野生型核苷酸序列是已知的,使用标准遗传密码推导出编码具有一个或多个氨基酸置换的任何给定突变体版本的Tn5的核苷酸序列是可能的。类似地,对于突变体版本的其他转座酶,可容易地导出核苷酸序列。然后,具有需要的核苷酸序列的核酸分子可使用本领域已知的标准分子生物学技术构建。

[0204] 根据本文提供的实施方案,定义的核酸不仅包括相同的核酸,而且包括任何小的碱基变异,包括,特别地,在导致归因于保守氨基酸置换中的简并密码子的同义密码子(指定相同氨基酸残基的不同密码子)的情况中的置换。术语“核酸序列”还包括与关于碱基变异给出的任何单链序列的互补序列。

[0205] 本文描述的核酸分子还可有利地被包括在适合的表达载体中,以在适合的宿主中表达由其编码的转座酶蛋白。将克隆的DNA掺入进适合的表达载体中用于随后转化所述细胞和随后选择所转化的细胞是本领域技术人员熟知的,如在Sambrook等(1989),Molecular cloning:A Laboratory Manual,Cold Spring Harbor Laboratory中提供的,其通过引用以其整体并入。

[0206] 此类表达载体包括具有被可操作地连接至能够影响所述DNA片段的表达的调节序列的根据本文提供的实施方案的核酸的载体,所述调节序列诸如启动子区域。术语“可操作地连接”指其中所描述的组件处于允许它们以它们以预期的方式起作用的关系的并置。此类载体可被转化进适合的宿主细胞中以提供根据本文提供的实施方案的蛋白的表达。

[0207] 核酸分子可编码成熟蛋白或具有前体序列(prosequence)的蛋白的核酸分子,包括编码前蛋白(preprotein)上然后被宿主细胞切割以形成成熟蛋白的前导序列的核酸分子。载体可以是,例如,装备有复制起点以及任选地用于表达所述核苷酸的启动子和任选地

启动子的调节子的质粒、病毒或噬菌体载体。载体可包含一种或更多种可选择标志物,诸如,例如,抗生素抗性基因。

[0208] 表达所需的调节元件包括结合RNA聚合酶并指导适当水平的转录起始的启动子序列以及用于核糖体结合的翻译起始序列。例如,细菌表达载体可包含启动子诸如lac启动子以及用于翻译起始的Shine-Dalgarno序列和起始密码子AUG。类似地,真核表达载体可包含RNA聚合酶II的异源或同源启动子、下游聚腺苷酸化信号、起始密码子AUG和用于脱离核糖体的终止密码子。此类载体可商购获得或由本领域熟知的方法描述的序列来组装。

[0209] 高等真核生物编码转座酶的DNA的转录可通过在载体中包含增强子序列来优化。增强子为作用于启动子以增加转录水平的DNA顺式作用元件。载体将通常还包含除了可选择标志物以外的复制起点。

[0210] 实施例1

[0211] 一般测定方法和条件

[0212] 以下段落描述了在下文呈现的实施例中使用的一般测定条件。

[0213] 使用Tn5用于人gDNA的WGS的标签化

[0214] 本部分描述了在下文实施例中使用的用于监测转座酶的插入偏倚的标签化测定。简言之,将50ng人基因组DNA与5 μ L TDE1在10mM Tris-乙酸盐,pH 7.6,25mM乙酸镁中在55 $^{\circ}$ C下孵育5min。然后,添加1/5反应物体积的125mM HEPES,pH 7.5,1M NaCl、50mM MgCl₂,随后添加Tn5转座酶至100nM,并在30 $^{\circ}$ C孵育60min。然后,反应物被清理(clean-up),并如在Illumina氏Nextera方案中描述扩增,并提交用于使用HiSeq 2000仪器测序。

[0215] Tn5转座体组装

[0216] 将Tn5与20 μ M退火的转座子在25mM HEPES,pH 7.6,125mM KCl、18.75mM NaCl、0.375mM EDTA、31.75%甘油中在室温下孵育30min。

[0217] 转座子组装

[0218] 通过将反应物加热至94 $^{\circ}$ C并将其缓慢冷却至室温,将转座子单独地在10mM Tris HCl,pH7.5,50mM NaCl、1mM EDTA中退火至40 μ M。对于Tn5-ME-A,Tn5马赛克末端序列A14 (Tn5 Mosaic End Sequence A14) (Tn5MEA) 被退火至Tn5非-转移的序列(NTS),且对于Tn5-ME-B,Tn5马赛克末端序列B15 (Tn5MEB) 被退火至Tn5非-转移的序列(NTS)。这些序列在下面示出:

[0219] Tn5MEA: 5' -TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

[0220] Tn5MEB: 5' -GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'

[0221] Tn5 NTS: 5' -CTGTCTCTTATACACATCT-3'

[0222] 2. 转座酶的克隆和表达

[0223] 本部分描述了用于克隆和表达在下文实施例中使用的多种转座酶突变体的方法。

[0224] 使用标准的定点诱变方法对编码转座酶的骨架基因序列的基因进行诱变。对于做出的每个突变,突变基因的固有序列(proper sequence)通过测序克隆的基因序列来确认。

[0225] 将Tn5转座酶基因克隆进修饰的pET11a质粒中。修饰的质粒包含来源于pASK5plus的纯化标签Strep Tag-II和来源于pET-SUMO的SUMO。使用BL21 (DE3) pLysY感受态细胞(New England Biolabs)进行表达。细胞在25 $^{\circ}$ C下生长至OD_{600nm}为0.5,并然后使用100 μ M IPTG诱导。在18 $^{\circ}$ C下进行表达19h。在蛋白酶抑制剂存在下在100mM TrisHCl,pH 8.0,1M NaCl、1mM

EDTA中使用微流化器裂解细胞沉淀。在裂解之后,将细胞裂解物与终浓度0.1%的脱氧胆酸盐孵育30分钟。添加聚乙烯亚胺至0.5%,然后以30,000xg离心20min。收集上清液并与等体积的饱和硫酸铵溶液混合,随后在冰上搅拌至少1h。然后将溶液以30,000xg离心20min,并将沉淀重悬于100mM Tris,pH8.0,1M NaCl、1mM EDTA和1mM DTT中。然后将重悬并过滤的溶液施加至使用AKTA纯化系统的Streptactin柱。使用100mM Tris,pH 8.0,1mM EDTA、1mM DTT、4M NaCl,随后是100mM Tris,pH 7.5,1mM EDTA、1mM DTT、100mM NaCl洗涤柱。使用100mM Tris,pH 7.5,1mM EDTA、1mM DTT、100mM NaCl和5mM脱硫生物素进行洗脱。使用100mM Tris pH 7.5,100mM NaCl、0.2mM EDTA和2mM DTT将洗脱物负载到肝素截留柱上。在使用相同缓冲液洗涤之后,使用100mM Tris pH 7.5,1M NaCl、0.2mM EDTA和2mM DTT用盐梯度洗脱Tn5变体。将级分收集、混合、浓缩,并添加甘油以得到50%的终浓度,之后在-20℃长期储存。

[0226] 3. 大肠杆菌基因组DNA中插入偏倚的IVC分析

[0227] 插入序列偏倚的分析使用IVC-图数据₁(强度对循环)进行。在对使用各自的DNA转座酶创建的DNA文库测序之后产生可用数据。简言之,如以上描述进行诱变和表达。转座酶与以上描述的转座子A和B孵育,并与大肠杆菌基因组DNA孵育以产生DNA文库。根据制造商的说明书,将每个文库在运行MiSeq快速化学的Illumina基因组分析仪系统(Illumina, Inc., San Diego, CA)上测序至少35个循环。使用Illumina RTA软件以在每个循环产生碱基调用(base call)和强度值。为了产生IVC图,将测序读段与大肠杆菌参考基因组比对,并将在每个循环的四个碱基的每一个的强度(发生)作为所有比对测序读段的全部强度值的分数作图。

[0228] 实施例2

[0229] 鉴定Tn5转座酶突变体的插入偏倚

[0230] 该实施例描述了使用IVC-图数据(强度对循环)分析插入序列偏倚。在对使用相应DNA转座酶创建的DNA文库测序之后可得该数据。该分析仅需要几个测序读段(20k-40k)以产生稳定的结果,并可在用来表达相应Tn5转座酶变体的大肠杆菌细胞裂解物中进行,且适合于HTS(高通量筛选)目的。

[0231] 对于多种单个氨基酸置换Tn5变体的代表性结果示于图2-7中。所示的变体指示,例如,通过在位置248处的置换的对称性损失(图2)、通过在位置119处的置换的变平的IVC图(图3)、通过在位置125处的置换或在位置248后的插入的IVC图的下半部中的减少的IVC(图4-5)、以及通过在位置K120处使用不同的芳香族氨基酸将重复从9bp增加至10bp(图6,表明对称性从1和第9bp变化至1和第10bp)。这些结果指示,与wt对照相比,指定的突变提供改进的插入偏倚。

[0232] 实施例3

[0233] 对细菌gDNA的全基因组测序

[0234] 进行这些实验,以比较多种转座酶突变体的a)估计的文库尺寸/多样性和b)AT/GC-脱落。这些实验用纯化的和活性归一化的Tn5转座酶变体进行。这些实验需要500k-1M测序读段/实验。

[0235] 结果通过使用指定的纯化的Tn5转座酶变体对蜡状芽孢杆菌gDNA进行标签化来获得。酶以活性来归一化并被设置为与Nextera™试剂盒一起销售的商业TDE1酶的活性匹配。

[0236] 指示为Tn5001的突变体具有与SEQ ID NO:11相同的氨基酸序列并用作“wt”对照。如以上表中示出的,指示为Tn5058、Tn5059和Tn5061的突变体分别具有与SEQ ID NO:18、19和20相同的氨基酸序列。实验以一式三份进行,数据显示所收集数据的平均值和标准偏差。“估计的文库尺寸”在不使用光学重复的情况下来计算,提供可重复的结果。

[0237] 如图8A中示出的,与Tn5001相比,Tn5058和Tn5059的AT-脱落存在显著减少,同时保持低的GC-脱落。类似地,如图8B中示出的,文库尺寸显著增加了1.7x (Tn5058) 和2.2x (Tn5059)。这些结果指示,与野生型转座酶相比,突变体Tn5058和Tn5059极大地改进了序列插入偏倚,导致下文实施例4中描述的进一步实验。

[0238] 实施例4

[0239] 对人gDNA的Nextera快速捕获富集实验

[0240] 以下实验用以上实施例3中描述的相同纯化的和活性归一化的Tn5转座酶突变体进行。这些实验通常需要40M-100M测序读段/实验,并且分析测序数据以比较a) 多样性、b) 富集、c) 覆盖度、d) 覆盖度均匀性、e) 罚分。

[0241] 捕获根据制造商的说明书使用Nextera快速捕获外显子组 (Illumina™) CEX池捕获探针以一式三份进行。如图9A中示出的,与wt对照Tn5001相比,所指示的突变体在覆盖度均匀性方面产生显著的改进。此外,如图9B中指示的,尽管平均靶覆盖度较低,所测试的突变体对10x和20x的靶标覆盖度产生统计学显著的改进。

[0242] 如图10A中示出的,所示的突变体产生了独特读段的数目和混合选择文库尺寸的增加。同样地,如图10B中示出的,与Tn5001相比,突变体产生较低的罚分,其为达到10x、20x或30x覆盖度深度所需的倍数更大的测序。这些结果指示,与对照相比,测试的突变体提供降低的插入偏倚和更均匀的覆盖度。

[0243] 实施例5

[0244] 标签化缓冲液组成对Tn5活性的影响

[0245] 进行以下实验以表征Tn5标签化缓冲液组成和反应条件对文库输出和测序度量的影响。

[0246] 为了评价标签化缓冲液组成和反应条件对文库输出和测序度量的影响,使用蜡状芽孢杆菌基因组DNA构建Tn5标签化的DNA文库。使用两种不同的Tn5转座酶构建标签化的文库,即突变体Tn5 (“TS-Tn5059”) 和对照高活性Tn5 (“TS-Tn5”)。TS是用于纯化Tn5和Tn5059蛋白的融合标签。相对于高活性Tn5氨基酸序列 (SEQ ID NO:11), Tn5059具有4个另外的突变K212R、P214R、G251R和A338V。TS-Tn5059在Tn5059的N-末端处包含TS标签。在一些实施方案中,TS-标签的C-末端可通过置换第一甲硫氨酸残基的接头与Tn5059的N-末端融合。在一些实施方案中,接头为Gly-Thr。

[0247] TS-Tn5059以10nM、40nM和80nM的终浓度来使用。TS-Tn5以4nM、15nM和30nM的终浓度来使用。将TS-Tn5059和TS-Tn5的酶浓度归一化 (使用标准缓冲液制剂), 以提供约相同水平的标签化活性,即,在10nM、40nM和80nM的TS-Tn5059与分别在4nM、15nM和30nM的Tn5具有约相同的活性。每个标签化的文库使用25ng输入的蜡状芽孢杆菌基因组DNA来制备。蜡状芽孢杆菌的基因组含量为约40%GC和约60%AT。

[0248] 将标签化缓冲液制备为2x制剂。2x制剂如下:标准缓冲液 (TD;20mM Tris乙酸盐, pH 7.6, 10mM乙酸镁和20%二甲基甲酰胺 (DMF); 钴缓冲液 (Co;20mM乙酸盐, pH 7.6, 和20mM

CoCl₂) ; 钴+DMSO缓冲液 (Co-DMSO; 20mM Tris乙酸盐, pH 7.6, 20mM CoCl₂和20% 二甲基亚砷 (DMSO)) ; 高分子量缓冲液 (HMW; 20mM Tris乙酸盐, pH 7.6, 和10mM乙酸镁) ; NF2缓冲液 (NF2; 20mM Tris乙酸盐, pH 7.6, 20mM CoCl₂和20% DMF) 。每天新鲜制备包含CoCl₂的标签化缓冲液。对于每个文库, 标签化反应通过在50μL总反应体积中将20μL蜡状芽孢杆菌基因组DNA (25ng) 、25μL 2x标签化缓冲液和5μL酶 (10x Ts-Tn5059或10x Ts-Tn5) 混合来进行。反应物在55℃孵育5分钟。在标签化反应之后, 根据标准Nextera™样品制备方案处理样品。在MiSeq装置上使用Illumina SBS (边合成边测序 (sequencing-by-synthesis)) 化学对文库测序。测序运行是使用V2 MiSeq试剂盒的2x71个循环。在生物分析仪上评价每个文库中的片段尺寸分布。

[0249] 图11示出了使用不同的标签化缓冲液制备的TS-Tn5059和TS-Tn5标签化的DNA文库中的独特分子数目的条形图100。文库中独特分子的数目为文库多样性 (复杂性) 的指示。图上的每个条代表一个标签化的文库。实验重复三次 (n=3) 。对照文库 (即, 使用标准的标签化缓冲液制备的文库) 被命名为“酶-酶浓度-DNA输入”。例如, 条形图100中的第一条被标记为“TS-Tn5059-10nM-25ng”并标示使用在标准缓冲液制剂中10nM的终浓度的TS-Tn5059和25ng的输入DNA制备的对照文库。使用修改的标签缓冲液制剂制备的文库被命名为“酶-酶浓度-缓冲液添加剂-DNA输入”。例如, 条形图100中的第四条被标记为“TS-Tn5059-10nM-Co-25ng”并标示使用在包含10mM CoCl₂的修改的标签化缓冲液中的10nM的终浓度的TS-Tn5059制备的文库。数据显示, 与在不添加CoCl₂的缓冲液 (即, 标准的缓冲液或HMW) 中制备的文库相比, 使用包含10mM CoCl₂的标签化缓冲液 (即, Co、Co-DMSO和NF2缓冲液) 制备的TS-Tn5059和TS-Tn5标签化的文库具有更高的平均多样性。

[0250] 图12示出了使用不同的标签化缓冲液制备的TS-Tn5059和TS-Tn5标签化的DNA文库中的GC脱落百分比的条形图200。对照文库和使用修改的标签化缓冲液制备的文库如图11中描述的来命名。GC脱落可被定义为基因组中的富含GC的区域从标签化的文库脱落 (缺失) 的百分比。数据显示, 对于使用标准的标签化缓冲液制备的对照TS-Tn5059和TS-Tn5文库, GC脱落的百分比相对低。数据还显示, 与在不添加CoCl₂的缓冲液 (即, 标准的缓冲液或HMW) 中制备的文库相比, 使用包含10mM CoCl₂的标签化缓冲液 (即, Co、Co-DMSO和NF2缓冲液) 制备的TS-Tn5059和TS-Tn5标签化的文库具有更高的GC脱落百分比 (即, 多达约6%) 。使用包含Co的缓冲液制备的文库中的GC脱落的增加被TS-Tn5059和TS-Tn5的浓度增加改善。例如, 与TS-Tn5059-10nm-25ng对照文库相比, TS-Tn5059-10nm-Co-25ng文库中的GC脱落百分比相对高。当TS-Tn5059的浓度增加至40nM (即, TS-Tn5059-40nm-Co-25ng) 和80nM (即, TS-Tn5059-80nm-Co-25ng) 时, GC脱落的百分比降低。

[0251] 图13示出了使用不同的标签化缓冲液制备的TS-Tn5059和TS-Tn5标签化的DNA文库中的AT脱落百分比的条形图300。对照文库和使用修改的标签化缓冲液制备的文库如图11中描述的来命名。AT脱落可被定义为基因组中的AT富含区域从标签化的文库脱落 (缺失) 的百分比。数据显示, 对于使用标准的标签化缓冲液制备的对照TS-Tn5059和TS-Tn5文库, 观察到一定量 (即, 分别为约1%至约3%和约7%至约3%) 的AT脱落。数据还显示, 与在不添加CoCl₂的缓冲液 (即, 标准的缓冲液或HMW) 中制备的文库相比, 使用低酶浓度 (即, 10nM) 和包含10mM CoCl₂的标签化缓冲液 (即, Co、Co-DMSO和NF2缓冲液) 制备的TS-Tn5059标签化的文库具有更低的AT脱落百分比。类似地, 与在不添加CoCl₂的缓冲液 (即, 标准的缓冲液或

HMW) 中制备的文库相比,使用包含10mM CoCl₂的标签化缓冲液(即,Co、Co-DMSO和NF2缓冲液)制备的TS-Tn5文库具有更低的AT脱落百分比。

[0252] 现在参考图12和13,在标签化缓冲液中添加CoCl₂ (10nM) (即,Co、Co-DMSO和NF2缓冲液)可在标签化的文库中“翻转(flip)”GC和AT脱落的百分比。例如,与TS-Tn5059-10nm-25ng对照文库相比,TS-Tn5059-10nm-Co-25ng文库中的GC脱落的百分比相对高(图12);而与TS-Tn5059-10nm-25ng对照文库相比,TS-Tn5059-10nm-Co-25ng文库中的AT脱落的百分比相对低(或无)(图13)。

[0253] 图14示出了使用标准缓冲液(TD)和钴缓冲液(Co)制剂制备的TS-Tn5059文库中的片段尺寸分布的生物分析仪迹线的图400。图400示出了,为Ts-Tn5059-10nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线410、为Ts-Tn5059-40nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线415、为Ts-Tn5059-80nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线420、为Ts-Tn5059-10nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线425、为Ts-Tn5059-10nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线430、以及为Ts-Tn5059-80nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线435。图400还示出了为以碱基对(bp)计的DNA片段尺寸的标准梯(ladder)的曲线440。梯中的片段尺寸(从左至右)示于表2中。

[0254]

表 2.尺寸梯	
梯峰	尺寸(bp)
1	35
2	50
3	100
4	150
5	200
6	300
7	400
8	500
9	600
10	700
11	1,000
12	2,000
13	3,000
14	7,000
15	10,380

[0255] 数据显示将在标签化反应中使用的TS-Tn5059的浓度从10nM增加至40nM和80nM将片段尺寸分布位移至更小的片段尺寸。在使用标准缓冲液(TD)制剂制备的文库中,片段尺寸分布的位移更加明显。例如,在使用钴缓冲(Co)制剂制备的文库中的片段尺寸分布在使用10nM TS-Tn5059制备的文库(曲线410)中为约3,000bp,且在使用40nM和80nM TS-Tn5059制备的文库(分别为曲线415和420)中为约1,000至约2,000bp。对于使用标准缓冲液(TD)制剂和80nM TS-Tn5059制备的文库(曲线435),片段尺寸分布为约200bp至约1,000bp。

[0256] 图15示出了使用钴-DMSO (Co-DMSO)、NF2和HMW缓冲液制剂制备的TS-Tn5059文库中的片段尺寸分布的生物分析仪迹线的图500。图500示出了,为Ts-Tn5059-10nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线510、为Ts-Tn5059-10nM-NF2-25 ng文库中的片

段尺寸分布的曲线的曲线515、为Ts-Tn5059-10nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线520、为Ts-Tn5059-40nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线525、为Ts-Tn5059-40nM-NF2-25 ng文库中的片段尺寸分布的曲线的曲线530、为Ts-Tn5059-40nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线535、为Ts-Tn5059-80nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线540、为Ts-Tn5059-80nM-NF2-25 ng文库中的片段尺寸分布的曲线的曲线545、以及为Ts-Tn5059-80nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线550。图500还示出了图14的图400的曲线440,其为以碱基对(bp)计的DNA片段尺寸的标准梯。

[0257] 数据显示,通常,将在标签化反应中使用的TS-Tn5059的浓度从10nM增加至40nM和80nM将片段尺寸分布位移至更小的片段尺寸。与使用Co-DMSO制备的文库(例如,曲线510和525)相比,在使用不包含CoCl₂的HMW缓冲液制备的文库(例如,曲线520和535)中,片段尺寸分布的变化更加明显。

[0258] 图16示出了使用标准缓冲液制剂(TD)和钴缓冲液(Co)制备的TS-Tn5文库中的片段尺寸分布的生物分析仪迹线的图600。图600示出了,为Ts-Tn5-4nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线610、为Ts-Tn5-15nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线615、为Ts-Tn5-30nM-Co-25 ng文库中的片段尺寸分布的曲线的曲线620、为Ts-Tn5-4nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线625、为Ts-Tn5-15nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线630、以及为Ts-Tn5-30nM-TD-25 ng文库中的片段尺寸分布的曲线的曲线635。图600还示出了图14的图400曲线440,其为以碱基对(bp)计的DNA片段尺寸的标准梯。

[0259] 数据显示将在标签化反应中使用的TS-Tn5的浓度从4nM增加至15nM和30nM将片段尺寸分布位移至更小的片段尺寸。在使用标准缓冲液(TD)制剂制备的文库中,片段尺寸分布的位移更加明显。该观察结果与图14的TS-Tn5059文库中的片段尺寸分布类似。

[0260] 图17示出了使用钴-DMSO(Co-DMSO)、NF2和HMW缓冲液制剂制备的TS-Tn5文库中的片段尺寸分布的生物分析仪迹线的图700。图700示出了,为Ts-Tn5-4nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线710、为Ts-Tn5-4nM-NF2-25 ng文库中的片段尺寸分布的曲线的曲线715、为Ts-Tn5-4nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线720、为Ts-Tn5-15nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线725、为Ts-Tn5-15nM-NF2-25 ng文库中的片段尺寸分布的曲线的曲线730、为Ts-Tn5-15nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线735、为Ts-Tn5-30nM-Co-DMSO-25 ng文库中的片段尺寸分布的曲线的曲线740、为Ts-Tn5-30nM-NF2-25 ng文库中的片段尺寸分布的曲线的曲线745、以及为Ts-Tn5-30nM-HMW-25 ng文库中的片段尺寸分布的曲线的曲线750。图700还示出了图14的图400的曲线440,其为以碱基对(bp)计的DNA片段尺寸的标准梯。

[0261] 数据显示将在标签化反应中使用的TS-Tn5的浓度从4nM增加至15nM和30nM将片段尺寸分布位移至更小的片段尺寸。在使用标准缓冲液(TD)制剂制备的文库中,片段尺寸分布的位移更加明显。该观察结果与图15的TS-Tn5059文库中的片段尺寸分布类似。

[0262] 通常,现在参考图14至17,在使用包含10nM CoCl₂的标签化缓冲液(例如,Co、Co-DMSO和NF2缓冲液)制备的TS-Tn5059和TS-Tn5文库中的片段尺寸大于在使用不含CoCl₂的标签化缓冲液(即,TD和HMW缓冲液)制备的TS-Tn5059和TS-Tn5文库中的片段尺寸。

[0263] 图18A、18B、18C和18D分别示出了TS-Tn5文库中的序列内容物的偏倚图800、TS-TN5-Co文库中的序列内容物的偏倚图830、TS-Tn5-Co-DMSO文库中的序列内容物的偏倚图840及TS-Tn5-NF2文库中的序列内容物的偏倚图850。偏倚图(或强度对循环数(IVC)图)将观察到的碱基(A、C、G或T)的比率绘制为SBS循环数的函数,并示出Tn5在标签化期间具有的优选的序列背景(sequence context)。

[0264] 偏倚图800、830、840和850各示出了为关于循环数的A含量的曲线的曲线810、为关于循环数的C含量的曲线的曲线815、为关于循环数的G含量的曲线的曲线820、以及为关于循环数的T含量的曲线的曲线825。例如,在图18A的TS-TN5文库中,代表碱基G的曲线820显示在第1个循环观察到的碱基的约38%为G;代表碱基T的曲线825显示在第1个循环观察到的碱基的约15%为T等。

[0265] 参考图18A,数据显示在使用标准的标签化缓冲液制剂制备的TS-Tn5文库中在约前15个SBS循环内观察到Tn5序列偏倚。在约15个循环之后,序列偏倚逐渐降低,且A、T、C和G含量反映期望的基因组组成。对于蜡状芽孢杆菌,基因组为约40%GC和约60%AT,这在从约第16个或第17个循环至第35个循环的偏倚图中表示,其中曲线810(即,A)和曲线825(即,T)汇聚在约30%(A+T~60%);且曲线815(即,C)和曲线820(即,G)汇聚在约20%(C+G~40%)。

[0266] 参考图18B、18C和18D,数据还显示在TS-Tn5-Co、TS-Tn5-Co-DMSO和Ts-Tn5-NF2文库中在约前15个SBS循环内观察到Tn5序列偏倚,TS-Tn5-Co、TS-Tn5-Co-DMSO和Ts-Tn5-NF2文库为使用包含CoCl₂的标签化缓冲液制备的文库。再次,在约15个循环之后,序列偏倚逐渐降低,且如参考图18A描述的,A、T、C和G含量反映期望的基因组组成。然而,在TS-Tn5-Co、TS-Tn5-Co-DMSO和Ts-Tn5-NF2文库中,曲线810(即,A)和曲线825(即,T)在约第10个循环至第15个循环开始朝向期望的基因组组成位移;且曲线815(即,C)和曲线820(即,G)在约第10个循环至第15个循环开始朝向期望的基因组组成位移。另外,当与图18A相比时,循环2-8之间的偏倚降低。数据显示在标签化缓冲液制剂中添加CoCl₂改善了标签化期间的Tn5序列偏倚。

[0267] 图19A、19B、19C和19D分别示出了TS-Tn5059文库中的序列内容物的偏倚图900、TS-TN5059-Co文库中的序列内容物的偏倚图930、TS-Tn5059-Co-DMSO文库中的序列内容物的偏倚图940及TS-Tn5059-NF2文库中的序列内容物的偏倚图950。偏倚图900、930、940和950各示出了为关于循环数的A含量的曲线的曲线910、为关于循环数的C含量的曲线的曲线915、为关于循环数的G含量的曲线的曲线920、以及为关于循环数的T含量的曲线的曲线925。

[0268] 参考图19A,数据显示在TS-Tn5059标签化的文库中,在约前15个SBS循环内观察到Tn5059序列偏倚。在约15个循环之后,序列偏倚降低,且如参考图18A描述的,A、T、C和G含量反映期望的基因组组成。然而,与图18A中示出的Tn5序列偏倚相比,突变体Tn5059显示出降低的序列偏倚。在TS-Tn5059文库中,曲线910(即,A)和曲线925(即,T)在约第10个循环至第15个循环开始朝向期望的基因组组成位移;且曲线915(即,C)和曲线920(即,G)在约第10个循环至第15个循环开始朝向期望的基因组组成位移。

[0269] 参考图19B、19C和19D,数据还显示在TS-Tn5059-Co、TS-Tn5059-Co-DMSO和Ts-Tn5059-NF2文库中在约前15个SBS循环内观察到Tn5059序列偏倚,TS-Tn5059-Co、TS-

Tn5059-Co-DMSO和Ts-Tn5059-NF2文库为使用包含CoCl₂的标签化缓冲液制备的文库。再次,在约15个循环之后,序列偏倚逐渐降低,且如参考图18A描述的,A、T、C和G含量反映期望的基因组组成。然而,在TS-Tn5059-Co文库中,曲线910(即,A)和曲线925(即,T)在约第5个循环开始朝向期望的基因组组成位移;且曲线915(即,C)和曲线920(即,G)在约第5个循环开始朝向期望的基因组组成位移。在TS-Tn5059-Co-DMSO和Ts-Tn5059-NF2文库中,曲线910(即,A)和曲线925(即,T)在约第5个循环之前开始朝向期望的基因组组成位移;且曲线915(即,C)和曲线920(即,G)在约第5个循环之前开始朝向期望的基因组组成位移。

[0270] 实施例6

[0271] 标签化缓冲液组成对Mos1活性的影响

[0272] 进行以下实验以表征Tn5标签化缓冲液组成和反应条件对文库输出和测序度量的影响。

[0273] Mos1标签化的DNA文库使用蜡状芽孢杆菌基因组DNA构建。用于构建标签化的文库的Mos1转座酶为MBP-Mos1融合蛋白。麦芽糖结合蛋白(MBP)为用于纯化Mos1蛋白的融合标签。MBP-Mos1以100μM的终浓度来使用。每个标签化的文库使用50ng输入的蜡状芽孢杆菌基因组DNA来制备。

[0274] 将标签化缓冲液制备为2x制剂。2x制剂如下:标准缓冲液(TD;20mM Tris乙酸盐,pH 7.6,10mM乙酸镁和20%二甲基甲酰胺(DMF);TD+NaCl(TD+NaCl;20mM Tris乙酸盐,pH 7.6,10mM乙酸镁、20%DMF和200mM NaCl);高分子量缓冲液(HMW;20mM Tris乙酸盐,pH 7.6和10mM乙酸镁);HEPES(50mM HEPES,pH 7.6,10mM乙酸镁、20%DMF);HEPES-DMSO(50mM HEPES pH 7.6,10mM乙酸镁和20%DMSO);HEPES-DMSO-Co(50mM HEPES,pH 7.6,20%DMSO和20mM CoCl₂)以及HEPES-DMSO-Mn(50mM HEPES,pH 7.6,20%DMSO和20mM锰(Mn))。每天新鲜制备包含CoCl₂的标签化缓冲液。

[0275] 对于每个文库,标签化反应通过在50μL总反应体积中将20μL蜡状芽孢杆菌基因组DNA(50ng)、25μL 2x标签化缓冲液和5μL酶(10x MBP-Mos1)混合来进行。反应物在30℃孵育60分钟。在标签化反应之后,根据标准Nextera™样品制备方案处理样品。在MiSeq装置上使用IlluminaSBS(边合成边测序(sequencing-by-synthesis))化学对文库测序。测序运行行为2x71个循环。

[0276] 图20示出了使用不同的标签化缓冲液制备的MBP-Mos1标签化的文库中的平均读段总数和平均多样性的条形图1000。读段总数为从流动池读段的总数。多样性为文库中独特分子的数目,并且用作文库复杂性的指示。图上的每对条代表一个标签化的文库。实验重复三次(n=3)。前两个图条,EZTn5-std-蜡状芽孢杆菌和NexteraV2-30C为分别在55℃和30℃使用Tn5和标准缓冲液制剂制备的比较文库。使用MBP-Mos1用于标签化反应制备的文库被命名为“酶-酶浓度-缓冲液”。例如,第三对图条被标记为“MBPMos1-100μM-TD”并标示使用在标准的标签化缓冲液(TD)中的在100μM的终浓度的MBP-Mos1制备的文库。数据显示。不同缓冲液对在相对相同数目或测序读段下由Mos1标签化制备的文库的多样性的影响。特别地,HEPES-DMSO-Mn缓冲液有助于增加文库的多样性。

[0277] 图21示出了MBP-Mos1标签化的文库中的GC和AT脱落的条形图1100。GC和AT脱落可被分别定义为基因组中从标签化的文库脱落(缺失)的富含GC的区域的百分比和AT富含区域的百分比。如图20中描述的命名文库。数据显示使用EZTn5和NexteraV2(即,Tn5转座酶)

制备的文库基本上没有GC脱落,但分别约7%和约5%的AT富含区域从标签化的文库中脱落。使用MBP-Mos1和标准的标签化缓冲液(MBPMos1-100 μ M-TD)制备的文库基本上没有AT脱落,但约2%或更少的富含GC的区域从标签化的文库中脱落。MBP-Mos1标签化的文库中的GC脱落百分比受到标签化缓冲液的组成的影响。GC脱落百分比在使用HMW、HEPES、HEPES-DMSO、HEPES-DMSO-Co和HEPES-DMSO-Mn缓冲液制备的MBP-Mos1标签化的文库中增加。

[0278] 图22A、22B、22C和22D分别示出了Mos1-HEPES文库中的序列含量的偏倚图1200、Mos1-HEPES-DMSO文库中的序列内容物的偏倚图1230、Mos1-HEPES-DMSO-Co文库中的序列内容物的偏倚图1240及Mos1-HEPES-DMSO-Mn文库中的序列内容物的偏倚图1250。偏倚图1200、1230、1240和1250各示出了为关于循环数的A含量的曲线的曲线1210、为关于循环数的C含量的曲线的曲线1215、为关于循环数计G含量的曲线的曲线1220、以及为关于循环数的T含量的曲线的曲线1225。

[0279] 参考图22A和22B,数据显示在Mos1-HEPES和Mos1-HEPES-DMSO标签化的文库中,在前几个SBS循环内观察到Mos1序列偏倚。在第1个SBS循环中,在整个流动池中,T的检测为约100%。在第2个SBS循环中,在整个流动池中,A的检测为约100%。在约4个循环之后,序列偏倚降低,且A、T、C和G含量反映期望的基因组组成。

[0280] 参考图22C和22D,数据还显示在Mos1-HEPES-DMSO-Co和Mos1-HEPES-DMSO-Mn标签化的文库中在前几个SBS循环中观察到Mos1序列偏倚,Mos1-HEPES-DMSO-Co和Mos1-HEPES-DMSO-Mn标签化的文库为使用分别用钴(Co)或锰(Mn)代替镁(Mg)的标签化缓冲液制备的文库。再次,在约4个循环之后,序列偏倚降低,且A、T、C和G含量反映期望的基因组组成。然而,在Mos1-HEPES-DMSO-Co和Mos1-HEPES-DMSO-Mn文库中,在第1个循环和第2个循环的曲线1210(即,A)和曲线1225(即,T),观察到朝向期望的基因组组成的位移。在Mos1-HEPES-DMSO-Mn文库中,朝向期望的基因组组成的位移更加明显。

[0281] 实施例7

[0282] TS-Tn5059文库制备和外显子组富集方案

[0283] 在一个实施方案中,本发明的方法提供了用于制备和富集基于Tn5转座体的外显子组文库的改进的工作流程。

[0284] 图23示出了制备和富集用于外显子组测序的基因组DNA文库的方法1260的实例的流程图。方法1260使用TS-Tn5059转座子和对目前Nextera®快速捕获方案的某些过程步骤的修改,以跨越一定范围的DNA输入量和测序度量提供改进的文库产量。例如,方法1260使用“双面(double-sided)”固相可逆固定(SPRI)方案(Agencourt AMPure XP beads; Beckman Coulter, Inc.)纯化标签化的DNA,并且在PCR扩增之前提供第一DNA片段尺寸选择步骤和第二DNA片段尺寸选择步骤。在另一个实例中,预浓缩过程被用来在外显子组富集之前浓缩片段化的DNA文库。方法1260包括,但不限于,以下步骤。

[0285] 在步骤1270,基因组DNA被转座体片标签化(加标签和片段化)。转座体同时片段化基因组DNA并将连接物序列添加至末端,允许随后通过PCR扩增。在一个实例中,转座体为TS-Tn5059。在标签化反应完成时,将标签化终止缓冲液添加至反应物中。标签化终止缓冲液可被修改以确保来自标签化的DNA的TS-Tn5059转座体复合物充分变性(例如,终止缓冲液中的SDS的浓度从0.1%增加至1.0%SDS结合高温加热)。

[0286] 在步骤1275,进行第一清理以从转座体纯化标签化的DNA并提供第一DNA片段尺寸

选择步骤。DNA片段尺寸可通过改变SPRI珠与DNA的体积与体积比率来选择(例如,1X SPRI = 1:1体积的SPRI:DNA)。例如,在第一尺寸选择中,选择SPRI珠与DNA的体积比率以结合大于一定尺寸的DNA片段(即,从样品中去除较大的DNA片段),而小于一定尺寸的DNA片段保留在上清液中。将其中具有经尺寸选择的DNA片段的上清液转移至干净的反应容器中,以用于后续处理。可弃去其上具有大DNA片段的SPRI珠。在一个实施方案中,SPRI珠的浓度可从0.8X至1.5X变化。在一个实施方案中,SPRI珠的浓度为0.8X。

[0287] 在步骤1280,进行第二次清理以进一步选择在一定尺寸范围内的DNA片段。例如,选择SPRI珠与DNA的体积比率以结合大于一定尺寸的DNA片段(即,在期望的尺寸范围内的DNA片段与SPRI珠结合)。较小的DNA片段保留在上清液中并弃去。然后将结合的DNA片段从SPRI珠洗脱,以用于随后处理。

[0288] 在任意的步骤1285,确定DNA片段尺寸分布。DNA片段尺寸分布,例如,使用生物分析仪确定。

[0289] 在步骤1290,经由有限循环PCR程序扩增纯化的标签化DNA。PCR步骤还添加标记(index) 1 (i7) 和标记2 (i5) 和测序、以及用于后续簇产生和测序所需的常见的连接物(P5和P7)。由于使用双面SPRI方法(即步骤1275和1280)选择期望的DNA片段尺寸范围,仅在期望的尺寸范围内的标签化的DNA片段可用于PCR扩增。因此,文库产量显著增加,并且随后的测序度量(例如,读段富集百分比)得到改进。

[0290] 在步骤1295,扩增的标签化的DNA文库使用基于珠的纯化方法纯化。

[0291] 在任意的步骤1300,确定PCR后的DNA片段尺寸分布。DNA片段尺寸分布,例如,使用生物分析仪确定。

[0292] 在步骤1310,将标签化的DNA文库预浓缩,然后是随后的杂交,以用于外显子组富集。例如,将标签化的DNA文库从约50 μ l预浓缩至约10 μ L。由于标签化的DNA文库被预浓缩,杂交动力学更快,且杂交时间降低。

[0293] 在步骤1320,进行用于外显子组富集的第一杂交。例如,将DNA文库与靶向感兴趣区域的生物素化捕获探针混合。DNA文库在约95 $^{\circ}$ C变性约10分钟,并在约58 $^{\circ}$ C与探针杂交约30分钟,总反应时间为约40分钟。

[0294] 在步骤1325,链霉亲和素珠用来捕获与感兴趣的靶向区域杂交的生物素化探针。使用两个加热的洗涤程序从珠中去除非特异性结合的DNA。然后将富集的文库从珠上洗脱并准备用于第二轮杂交。

[0295] 在步骤1330,用于外显子组富集的第二杂交使用与第一杂交相同的探针和阻断物来进行。例如,来自步骤155的洗脱的DNA文库在约95 $^{\circ}$ C变性约10分钟,并在约58 $^{\circ}$ C杂交约30分钟,总反应时间为约40分钟。第二杂交用来确保捕获区域的高特异性。

[0296] 在步骤1335,链霉亲和素珠用来捕获与感兴趣的靶向区域杂交的生物素化探针。使用两个加热的洗涤程序从珠中去除非特异性结合的DNA。然后将外显子组富集的文库从珠洗脱并通过10个循环的PCR扩增,为测序做准备。

[0297] 在步骤1340,外显子组富集的捕获样品(即,外显子组富集DNA文库)使用基于珠的纯化方案纯化。

[0298] 在步骤1345,PCR扩增外显子组富集DNA文库,以用于测序。

[0299] 在步骤1350,扩增的富集DNA文库任意地使用基于珠的纯化方法纯化。例如,使

用1X SPRI珠方案以去除可能干扰随后簇扩增和测序的不想要的产物(例如,过量的引物)。
[0300] 方法100在约11小时内提供文库制备和外显子组富集。如果省略了任选的步骤1285和1300,则方法1260在约9小时内提供文库制备和外显子组富集。

[0301] 实施例8

[0302] TS-Tn5059插入偏倚

[0303] 转座酶可在标签化反应中具有一定的插入位点(DNA序列)偏倚。DNA序列偏倚可导致基因组的某些区域(例如,富含GC或富含AT)从标签化的文库脱落。例如,Tn5转座酶对于基因组的富含GC的区域具有一定偏倚;因此,基因组的AT区域可在Tn5标签化的文库中脱落。为了提供更完全的基因组覆盖度,期望最小的序列偏倚。

[0304] 为了评价TS-Tn5059转座体对文库输出和测序度量的影响,使用标准的Nextera™ DNA文库制备试剂盒制备TS-Tn5059标签化的DNA文库,以用于全基因组测序和蜡状芽孢杆菌基因组DNA测序。TS-Tn5059以40nM的终浓度来使用。参考对照文库使用25nM NexteraV2转座体的标准的反应条件来制备。通过边合成边测序(SBS)评价文库。

[0305] 图24A示出了使用TS-Tn5059转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库中的覆盖度的图1400。随着GC含量增加,TS-Tn5059转座体变得对偏倚的水平增加耐受。图24B示出了使用NexteraV2转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库中的覆盖度的图1450。图24B显示了随着GC含量增加,富含GC的区域的Nextera V2覆盖度变得偏斜,具有增加的偏倚。数据显示,与使用NexteraV2制备的标签化的文库相比,使用TS-Tn5059制备的标签化的DNA文库具有跨越宽GC/AT范围的改进且更均匀的覆盖度与较低的插入偏倚。

[0306] 图25A示出了使用TS-Tn5059转座体制备的标签化的蜡状芽孢杆菌基因组DNA文库中的缺口位置和缺口长度的图1500。图25B示出了使用NexteraV2转座体制备的标签化的蜡状芽孢杆菌基因组标签化的DNA文库中的缺口位置和缺口长度的图。在TS-Tn5059标签化的文库中的缺口的数目为27。在NexteraV2标签化的文库中的缺口的数目为208。数据显示,与使用NexteraV2转座体制备的标签化的文库相比,使用TS-Tn5059转座体制备的标签化的DNA文库具有更均匀的覆盖度与较低的缺口。

[0307] 实施例9

[0308] TS-Tn5059 DNA输入耐受性

[0309] 标签化的DNA文库的制备使用酶促DNA片段化步骤(例如,转座体介导的标签化),并因此与,例如,机械片段化方法相比,可对DNA输入更敏感。在一个实例中,目前的Nextera®快速捕获富集方案已被优化为用于50ng的总基因组DNA输入。基因组DNA的较高量的输入可导致不完全标签化和较大的插入物尺寸,这可影响随后的富集性能。标签化反应中的较低量的基因组DNA输入或低质量基因组DNA可产生小于期望的插入物尺寸。较小的插入物可在随后的清理步骤期间丢失,并导致较低的文库多样性。

[0310] 为了评价不同DNA输入量对片段(插入物)尺寸分布的影响,TS-Tn5059标签化的DNA文库在多种酶浓度使用不同量的输入基因组DNA来制备,并将片段尺寸与其他转座酶获得的片段尺寸进行比较,其他转座酶的活性被归一化为40nM TS-Tn5059和25ng基因组DNA输入的活性。

[0311] 由40nM TS-Tn5059,归一化的TDE1(Tn5版本-1)和归一化的TS-Tn5并使用25ng人基因组DNA产生的片段的尺寸分布是相似的,如图26和27中示出的。

[0312] 然而,TS-Tn5059在较高的酶浓度和宽的输入DNA量范围内显示增加的DNA输入耐受性。图28示出了使用一定范围的DNA输入制备的标签化的基因组DNA文库中的片段尺寸分布的生物分析仪迹线的图1600。240nM的TS-Tn5059标签化的文库通过标签化来制备,经1.8X SPRI清理,随后生物分析仪迹线。参考对照文库使用目前的Nextera®快速捕获试剂盒(“Nextera”)和Agilent QXT试剂盒(“Agilent QXT”)来制备。标签化的文库使用25ng、50ng、75ng和100ng蜡状芽孢杆菌基因组DNA制备。数据显示,与使用Nextera或Agilent QXT转座体制备的文库相比,使用TS-Tn5059转座体制备的标签化的DNA文库跨越25ng至100ng DNA输入范围具有更一致的片段尺寸分布。当DNA输入的量从25ng增加到100ng时,TS-Tn5059标签化的文库中的标签化的DNA的产量增加,同时片段尺寸分布保持实质相同。相比之下,在75ng和100ng DNA输入时,Nextera和Agilent QXT片段文库显示DNA片段尺寸分布向较大片段尺寸的实质性位移。

[0313] 图29A示出了由第一使用者使用从5ng至100ng的不同人Coriel DNA输入制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图1700。图29B示出了由第二使用者制备的TS-Tn5059标签化的文库中的片段尺寸分布的生物分析仪迹线的图1750。标签化的文库使用5ng、10ng、25ng、50ng、75ng和100ng蜡状芽孢杆菌基因组DNA制备。图29A的图1700和图29B的图1750两者示出了,使用5ng DNA输入制备的标签化的文库中的片段尺寸分布的线1710、使用10ng DNA输入制备的标签化的文库中的片段尺寸分布的线1715、使用25ng DNA输入制备的标签化的文库中的片段尺寸分布的线1720、使用50ng DNA输入制备的标签化的文库中的片段尺寸分布的线1725、使用75ng DNA输入制备的标签化的文库中的片段尺寸分布的线1730、以及使用100ng DNA输入制备的标签化的文库中的片段尺寸分布的线1735。数据显示TS-Tn5059标签化的DNA文库中的片段尺寸分布与5ng至100ng的DNA输入范围是一致的。对于不同使用者,观察到片段尺寸分布的一致性。

[0314] 在另一个实例中,表3示出了在跨越从5ng至200ng的DNA输入范围的TS-Tn5059标签化的DNA文库中的中值文库插入物尺寸。

[0315]

表 3. 5 ng 至 200 ng DNA 输入的中值插入物尺寸									
输入 DNA (ng)		0	5	0	5	00	50	00	平均值 ± SD
中值插入物(bp)	64	69	58	44	71	78	79	75	167 ± 11

[0316] 在又一个实例中,表4示出了使用25ng、50ng、75ng和100ng DNA输入制备的TS-Tn5059标签化的DNA文库的文库插入物尺寸和外显子组富集测序度量。数据显示读段富集百分比(%)为约80%。使用目前的Nextera®快速捕获富集方案制备的标签化的文库的读段富集百分比为约60%(数据未显示)。数据还显示跨越从15ng至100ng的DNA输入范围的一致的插入物尺寸。

[0317]

表 4. 25 ng 至 100 ng DNA 输入的插入物尺寸和外显子组富集度量					
外显子组(Picard)度量	5 ng	0 ng	5 ng	00 ng	平均值±SD
%读段富集	8	9	0	7	±1 78
%重复	.6	.7	.9	.4	2. 9 ± 0.3
%零覆盖度	.6	.9	.2	.2	2. 2 ± 0.2
在 10x 的%外显子组覆盖度	1.6	2.4	2	2.4	82 .1 ± 0.4
插入物尺寸 150 ± 25 bp	67	69	68	76	17 0 ± 4
预富集文库量(ng/μL)	23	40	65	52	14 5 ± 15

[0318] 在另一个实例中,表5示出了在TS-Tn5059标签化的文库中跨越从25ng至100ng的DNA输入范围的预富集文库产量。

[0319]

表 5. 25 ng 至 100 ng DNA 输入的预富集文库产量				
实验编号	对于 15 μL 的产量 (25 ng 输入)	对于 15 μL 的产量 (50 ng 输入)	对于 15 μL 的产量 (75 ng 输入)	对于 15 μL 的产量 (100 ng 输入)
实验#1	1098	1468	2420	2240
实验#2	1332	2200	2480	1974
实验#3	1845	2100	2475	2295
实验#4	1860	2295	2895	2760
实验#5	1830	2895	1515	1665
实验#6	1785	2760	1725	1920

[0320] 在又一个实例中,表6示出了TS-Tn5059标签化的DNA文库的外显子组富集测序度量。以50ng的输入DNA开始,使用500ng、625ng和750ng的文库DNA输入制备文库,以用于外显子组富集。数据显示外显子组富集度量跨越一定范围的预富集文库输入量(即,500ng至750ng)是一致的。

[0321]

表 6. 500 ng 至 750 ng 预富集文库输入量的外显子组富集度量				
分配的 (marketing) 度量	00 ng	25 ng	50 ng	平均值±SD
%读段富集	2.2	1.7	2.3	82.1 ± 0.32
%重复	.1	.3	.9	4.1 ± 0.2
%零覆盖度	.7	.6	.7	1.67 ± 0.06
在 10x 的%外显子组覆盖度	4.3	5.4	5.6	85.5 ± 0.7
HS 文库尺寸	9M	0M	1M	50M + 1M
平均覆盖度	5.1	8.2	8.6	47.3 ± 1.92
HS 20x 罚分	.5	.3	.4	6.4 ± 0.1

[0322] 在又一个实例中,表7示出了使用目前的Nextera®快速捕获富集杂交方案(“NRC”)和图23的方法1260的富集步骤1310至1350制备的标签化的DNA文库的外显子组富集测序度量。数据显示,与使用目前的Nextera®快速捕获富集方案(“NRC”)中的杂交方案制备的文库相比,在使用图23的方法1260制备的TS-Tn5059标签化的文库中,改进和/或维

持外显子组富集度量。

[0323]

表 7.使用“NRC”和方法 100 杂交方案制备的标签化的 DNA 文库的外显子组富集测序度量		
关键外显子组度量	NRC 杂交	方法 100 杂交
读段富集(未填充(not padded))	54.3%	77.7%
读段富集(填充(padded))	64.3%	85%
平均覆盖度	50.3X	52.1X
零靶脱落	1.6%	1.9%
%重复(10 mil.读段)	3.1%	4%
在 10X 的覆盖度	83.3	84.3%
HS 文库尺寸	62.3 M	61 M
对靶选择的%	76.5%	85.2%
HS 20 x 罚分	7.4	6.8

[0324] 在单独的实验中,与归一化为相同浓度的Tn5版本-1和TS-Tn5转座酶相比,TS-Tn5059显示出在较高浓度(归一化为6X浓度)下增加的DNA输入耐受性。结果示于图30-33中。在6X“归一化”浓度下,Tn5版本1(图30)和TS-Tn5(图31)两者在25-100ng之间变化的gDNA输入显示片段尺寸分布位移。相比之下,在6X归一化浓度的TS-Tn5059未显示在10-100ng之间的DNA输入的显著的尺寸位移(图32)。当将DNA输入增加至200-500ng时,片段尺寸分布开始位移(图33)。TS-Tn5059的增加的DNA输入耐受性的结果总结在下表8中。

[0325] 表8:在最终50uL反应物中的TS-Tn5059 (nM) :gDNA (ng) 的比率

[0326]

TS-Tn5059 体积 (uL)	贮备溶液 体积(nM)	终 Rxn 体积(uL)	在 rxn 中的终浓度 (nM)	gDNA 输入(ng)	在 50 uL rxn 中的 TS-Tn5059 (nM) : gDNA (ng)的比率	注解
5	400	50	40	25	1.6	WGS 比率
5	400	50	40	50	0.8	
15	800	50	240 (6X)	500	0.48	
15	800	50	240 (6X)	400	0.6	
15	800	50	240 (6X)	300	0.8	
15	800	50	240 (6X)	200	1.2	
15	800	50	240 (6X)	100	2.4	观察到输入耐受性
15	800	50	240 (6X)	75	3.2	
15	800	50	240 (6X)	50	4.8	
15	800	50	240 (6X)	25	9.6	
15	800	50	240 (6X)	10	24	
15	800	50	240 (6X)	5	48	

[0327] 因此,对于TS-Tn5059,在比率 ≥ 2.4 (TS-Tn5059的nM数:输入DNA的ng数),不存在尺寸位移,表明DNA输入耐受性增加。

[0328] 贯穿该申请,提到多个出版物、专利和/或专利申请。通过引用以其全文将这些出

版物的公开内容在此并入该申请。

[0329] 在本文中术语包括意图是开放式的,不仅包括提及的要素,另外还包括任何另外的要素。

[0330] 已描述了很多实施方案。但是,将理解可进行多种改变。相应地,其他的实施方案在所附权利要求书的范围内。

[0001]	序列表																		
[0002]	<110> 伊鲁米那股份有限公司																		
[0003]	<120> 用于改进掺入序列偏倚的修饰的转座酶																		
[0004]	<130> IP-1198-PCT																		
[0005]	<160> 26																		
[0006]	<170> PatentIn版本3.5																		
[0007]	<210> 1																		
[0008]	<211> 476																		
[0009]	<212> PRT																		
[0010]	<213> 未知																		
[0011]	<220>																		
[0012]	<223> 转座酶Tn5																		
[0013]	<400> 1																		
[0014]	Met	Ile	Thr	Ser	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val			
[0015]	1				5						10				15				
[0016]	Phe	Ser	Ser	Ala	Ala	Leu	Gly	Asp	Pro	Arg	Arg	Thr	Ala	Arg	Leu	Val			
[0017]					20					25					30				
[0018]	Asn	Val	Ala	Ala	Gln	Leu	Ala	Lys	Tyr	Ser	Gly	Lys	Ser	Ile	Thr	Ile			
[0019]					35				40					45					
[0020]	Ser	Ser	Glu	Gly	Ser	Glu	Ala	Met	Gln	Glu	Gly	Ala	Tyr	Arg	Phe	Ile			
[0021]					50				55					60					
[0022]	Arg	Asn	Pro	Asn	Val	Ser	Ala	Glu	Ala	Ile	Arg	Lys	Ala	Gly	Ala	Met			
[0023]	65					70					75				80				
[0024]	Gln	Thr	Val	Lys	Leu	Ala	Gln	Glu	Phe	Pro	Glu	Leu	Leu	Ala	Ile	Glu			
[0025]					85					90					95				
[0026]	Asp	Thr	Thr	Ser	Leu	Ser	Tyr	Arg	His	Gln	Val	Ala	Glu	Glu	Leu	Gly			
[0027]					100					105					110				
[0028]	Lys	Leu	Gly	Ser	Ile	Gln	Asp	Lys	Ser	Arg	Gly	Trp	Trp	Val	His	Ser			
[0029]					115				120					125					
[0030]	Val	Leu	Leu	Leu	Glu	Ala	Thr	Thr	Phe	Arg	Thr	Val	Gly	Leu	Leu	His			
[0031]					130				135					140					
[0032]	Gln	Glu	Trp	Trp	Met	Arg	Pro	Asp	Asp	Pro	Ala	Asp	Ala	Asp	Glu	Lys			
[0033]	145					150					155				160				
[0034]	Glu	Ser	Gly	Lys	Trp	Leu	Ala	Ala	Ala	Ala	Thr	Ser	Arg	Leu	Arg	Met			
[0035]					165					170					175				
[0036]	Gly	Ser	Met	Met	Ser	Asn	Val	Ile	Ala	Val	Cys	Asp	Arg	Glu	Ala	Asp			
[0037]					180					185					190				
[0038]	Ile	His	Ala	Tyr	Leu	Gln	Asp	Lys	Leu	Ala	His	Asn	Glu	Arg	Phe	Val			
[0039]					195				200					205					
[0040]	Val	Arg	Ser	Lys	His	Pro	Arg	Lys	Asp	Val	Glu	Ser	Gly	Leu	Tyr	Leu			
[0041]					210				215					220					

[0042]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[0043]	225	230	235 240
[0044]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg		
[0045]	245	250	255
[0046]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[0047]	260	265	270
[0048]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		
[0049]	275	280	285
[0050]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		
[0051]	290	295	300
[0052]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[0053]	305	310	315 320
[0054]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		
[0055]	325	330	335
[0056]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[0057]	340	345	350
[0058]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[0059]	355	360	365
[0060]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[0061]	370	375	380
[0062]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[0063]	385	390	395 400
[0064]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[0065]	405	410	415
[0066]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[0067]	420	425	430
[0068]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[0069]	435	440	445
[0070]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[0071]	450	455	460
[0072]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[0073]	465	470	475
[0074]	<210> 2		
[0075]	<211> 476		
[0076]	<212> PRT		
[0077]	<213> 未知		
[0078]	<220>		
[0079]	<223> 转座酶Tn5		
[0080]	<220>		
[0081]	<221> 变体		
[0082]	<222> (248) .. (248)		
[0083]	<223> X = Y, T, K, S, L, A, W, P, G, R, F, H		

[0084]	<400> 2																			
[0085]	Met	Ile	Thr	Ser	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val				
[0086]	1				5					10					15					
[0087]	Phe	Ser	Ser	Ala	Ala	Leu	Gly	Asp	Pro	Arg	Arg	Thr	Ala	Arg	Leu	Val				
[0088]				20					25					30						
[0089]	Asn	Val	Ala	Ala	Gln	Leu	Ala	Lys	Tyr	Ser	Gly	Lys	Ser	Ile	Thr	Ile				
[0090]			35					40					45							
[0091]	Ser	Ser	Glu	Gly	Ser	Glu	Ala	Met	Gln	Glu	Gly	Ala	Tyr	Arg	Phe	Ile				
[0092]		50					55					60								
[0093]	Arg	Asn	Pro	Asn	Val	Ser	Ala	Glu	Ala	Ile	Arg	Lys	Ala	Gly	Ala	Met				
[0094]	65				70					75				80						
[0095]	Gln	Thr	Val	Lys	Leu	Ala	Gln	Glu	Phe	Pro	Glu	Leu	Leu	Ala	Ile	Glu				
[0096]				85				90					95							
[0097]	Asp	Thr	Thr	Ser	Leu	Ser	Tyr	Arg	His	Gln	Val	Ala	Glu	Glu	Leu	Gly				
[0098]			100					105					110							
[0099]	Lys	Leu	Gly	Ser	Ile	Gln	Asp	Lys	Ser	Arg	Gly	Trp	Trp	Val	His	Ser				
[0100]		115					120					125								
[0101]	Val	Leu	Leu	Leu	Glu	Ala	Thr	Thr	Phe	Arg	Thr	Val	Gly	Leu	Leu	His				
[0102]		130					135					140								
[0103]	Gln	Glu	Trp	Trp	Met	Arg	Pro	Asp	Asp	Pro	Ala	Asp	Ala	Asp	Glu	Lys				
[0104]	145				150					155				160						
[0105]	Glu	Ser	Gly	Lys	Trp	Leu	Ala	Ala	Ala	Ala	Thr	Ser	Arg	Leu	Arg	Met				
[0106]			165					170				175								
[0107]	Gly	Ser	Met	Met	Ser	Asn	Val	Ile	Ala	Val	Cys	Asp	Arg	Glu	Ala	Asp				
[0108]			180					185				190								
[0109]	Ile	His	Ala	Tyr	Leu	Gln	Asp	Lys	Leu	Ala	His	Asn	Glu	Arg	Phe	Val				
[0110]		195					200					205								
[0111]	Val	Arg	Ser	Lys	His	Pro	Arg	Lys	Asp	Val	Glu	Ser	Gly	Leu	Tyr	Leu				
[0112]		210					215				220									
[0113]	Tyr	Asp	His	Leu	Lys	Asn	Gln	Pro	Glu	Leu	Gly	Gly	Tyr	Gln	Ile	Ser				
[0114]	225				230					235				240						
[0115]	Ile	Pro	Gln	Lys	Gly	Val	Val	Xaa	Lys	Arg	Gly	Lys	Arg	Lys	Asn	Arg				
[0116]			245					250				255								
[0117]	Pro	Ala	Arg	Lys	Ala	Ser	Leu	Ser	Leu	Arg	Ser	Gly	Arg	Ile	Thr	Leu				
[0118]			260					265				270								
[0119]	Lys	Gln	Gly	Asn	Ile	Thr	Leu	Asn	Ala	Val	Leu	Ala	Glu	Glu	Ile	Asn				
[0120]		275					280					285								
[0121]	Pro	Pro	Lys	Gly	Glu	Thr	Pro	Leu	Lys	Trp	Leu	Leu	Leu	Thr	Ser	Glu				
[0122]		290					295				300									
[0123]	Pro	Val	Glu	Ser	Leu	Ala	Gln	Ala	Leu	Arg	Val	Ile	Asp	Ile	Tyr	Thr				
[0124]	305				310					315				320						
[0125]	His	Arg	Trp	Arg	Ile	Glu	Glu	Phe	His	Lys	Ala	Trp	Lys	Thr	Gly	Ala				

[0126]	325							330							335					
[0127]	Gly	Ala	Glu	Arg	Gln	Arg	Met	Glu	Glu	Pro	Asp	Asn	Leu	Glu	Arg	Met				
[0128]	340							345							350					
[0129]	Val	Ser	Ile	Leu	Ser	Phe	Val	Ala	Val	Arg	Leu	Leu	Gln	Leu	Arg	Glu				
[0130]	355							360							365					
[0131]	Ser	Phe	Thr	Leu	Pro	Gln	Ala	Leu	Arg	Ala	Gln	Gly	Leu	Leu	Lys	Glu				
[0132]	370							375							380					
[0133]	Ala	Glu	His	Val	Glu	Ser	Gln	Ser	Ala	Glu	Thr	Val	Leu	Thr	Pro	Asp				
[0134]	385					390					395					400				
[0135]	Glu	Cys	Gln	Leu	Leu	Gly	Tyr	Leu	Asp	Lys	Gly	Lys	Arg	Lys	Arg	Lys				
[0136]	405							410							415					
[0137]	Glu	Lys	Ala	Gly	Ser	Leu	Gln	Trp	Ala	Tyr	Met	Ala	Ile	Ala	Arg	Leu				
[0138]	420							425							430					
[0139]	Gly	Gly	Phe	Met	Asp	Ser	Lys	Arg	Thr	Gly	Ile	Ala	Ser	Trp	Gly	Ala				
[0140]	435							440							445					
[0141]	Leu	Trp	Glu	Gly	Trp	Glu	Ala	Leu	Gln	Ser	Lys	Leu	Asp	Gly	Phe	Leu				
[0142]	450							455							460					
[0143]	Ala	Ala	Lys	Asp	Leu	Met	Ala	Gln	Gly	Ile	Lys	Ile								
[0144]	465					470					475									
[0145]	<210> 3																			
[0146]	<211> 476																			
[0147]	<212> PRT																			
[0148]	<213> 未知																			
[0149]	<220>																			
[0150]	<223> 转座酶Tn5																			
[0151]	<220>																			
[0152]	<221> 变体																			
[0153]	<222> (119) .. (119)																			
[0154]	<223> X = L, M, S, A, V																			
[0155]	<400> 3																			
[0156]	Met	Ile	Thr	Ser	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val				
[0157]	1	5				10							15							
[0158]	Phe	Ser	Ser	Ala	Ala	Leu	Gly	Asp	Pro	Arg	Arg	Thr	Ala	Arg	Leu	Val				
[0159]	20							25							30					
[0160]	Asn	Val	Ala	Ala	Gln	Leu	Ala	Lys	Tyr	Ser	Gly	Lys	Ser	Ile	Thr	Ile				
[0161]	35							40							45					
[0162]	Ser	Ser	Glu	Gly	Ser	Glu	Ala	Met	Gln	Glu	Gly	Ala	Tyr	Arg	Phe	Ile				
[0163]	50							55							60					
[0164]	Arg	Asn	Pro	Asn	Val	Ser	Ala	Glu	Ala	Ile	Arg	Lys	Ala	Gly	Ala	Met				
[0165]	65					70					75					80				
[0166]	Gln	Thr	Val	Lys	Leu	Ala	Gln	Glu	Phe	Pro	Glu	Leu	Leu	Ala	Ile	Glu				
[0167]	85							90							95					

[0168]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[0169]		100	105
[0170]	Lys Leu Gly Ser Ile Gln Xaa Lys Ser Arg Gly Trp Trp Val His Ser		110
[0171]		115	120
[0172]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		125
[0173]		130	135
[0174]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		140
[0175]		145	150
[0176]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met		155
[0177]		165	170
[0178]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		175
[0179]		180	185
[0180]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		190
[0181]		195	200
[0182]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		205
[0183]		210	215
[0184]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		220
[0185]		225	230
[0186]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg		235
[0187]		245	250
[0188]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		255
[0189]		260	265
[0190]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		270
[0191]		275	280
[0192]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		285
[0193]		290	295
[0194]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		300
[0195]		305	310
[0196]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		315
[0197]		325	330
[0198]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		335
[0199]		340	345
[0200]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		350
[0201]		355	360
[0202]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		365
[0203]		370	375
[0204]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		380
[0205]		385	390
[0206]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		395
[0207]		405	410
[0208]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		415
[0209]		420	425
			430

[0210]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
[0211]	435 440 445
[0212]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
[0213]	450 455 460
[0214]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
[0215]	465 470 475
[0216]	<210> 4
[0217]	<211> 476
[0218]	<212> PRT
[0219]	<213> 未知
[0220]	<220>
[0221]	<223> 转座酶Tn5
[0222]	<400> 4
[0223]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[0224]	1 5 10 15
[0225]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[0226]	20 25 30
[0227]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[0228]	35 40 45
[0229]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
[0230]	50 55 60
[0231]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0232]	65 70 75 80
[0233]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0234]	85 90 95
[0235]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0236]	100 105 110
[0237]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Met Val His Ser
[0238]	115 120 125
[0239]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[0240]	130 135 140
[0241]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
[0242]	145 150 155 160
[0243]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
[0244]	165 170 175
[0245]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
[0246]	180 185 190
[0247]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
[0248]	195 200 205
[0249]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
[0250]	210 215 220
[0251]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser

[0252]	225	230	235	240
[0253]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg			
[0254]	245	250	255	
[0255]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu			
[0256]	260	265	270	
[0257]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn			
[0258]	275	280	285	
[0259]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu			
[0260]	290	295	300	
[0261]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr			
[0262]	305	310	315	320
[0263]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala			
[0264]	325	330	335	
[0265]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met			
[0266]	340	345	350	
[0267]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu			
[0268]	355	360	365	
[0269]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu			
[0270]	370	375	380	
[0271]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp			
[0272]	385	390	395	400
[0273]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys			
[0274]	405	410	415	
[0275]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu			
[0276]	420	425	430	
[0277]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala			
[0278]	435	440	445	
[0279]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu			
[0280]	450	455	460	
[0281]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile			
[0282]	465	470	475	
[0283]	<210> 5			
[0284]	<211> 477			
[0285]	<212> PRT			
[0286]	<213> 未知			
[0287]	<220>			
[0288]	<223> 转座酶Tn5			
[0289]	<400> 5			
[0290]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val			
[0291]	1	5	10	15
[0292]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val			
[0293]	20	25	30	

[0294]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile		
[0295]	35	40	45
[0296]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile		
[0297]	50	55	60
[0298]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met		
[0299]	65	70	75
[0300]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu		
[0301]	85	90	95
[0302]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[0303]	100	105	110
[0304]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser		
[0305]	115	120	125
[0306]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		
[0307]	130	135	140
[0308]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[0309]	145	150	155
[0310]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met		
[0311]	165	170	175
[0312]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[0313]	180	185	190
[0314]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[0315]	195	200	205
[0316]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[0317]	210	215	220
[0318]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[0319]	225	230	235
[0320]	Ile Pro Gln Lys Gly Val Val Asp Val Lys Arg Gly Lys Arg Lys Asn		
[0321]	245	250	255
[0322]	Arg Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr		
[0323]	260	265	270
[0324]	Leu Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile		
[0325]	275	280	285
[0326]	Asn Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser		
[0327]	290	295	300
[0328]	Glu Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr		
[0329]	305	310	315
[0330]	Thr His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly		
[0331]	325	330	335
[0332]	Ala Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg		
[0333]	340	345	350
[0334]	Met Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg		
[0335]	355	360	365

[0336]	Glu Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys
[0337]	370 375 380
[0338]	Glu Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro
[0339]	385 390 395 400
[0340]	Asp Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg
[0341]	405 410 415
[0342]	Lys Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg
[0343]	420 425 430
[0344]	Leu Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly
[0345]	435 440 445
[0346]	Ala Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe
[0347]	450 455 460
[0348]	Leu Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
[0349]	465 470 475
[0350]	<210> 6
[0351]	<211> 476
[0352]	<212> PRT
[0353]	<213> 未知
[0354]	<220>
[0355]	<223> 转座酶Tn5
[0356]	<220>
[0357]	<221> 变体
[0358]	<222> (120) .. (120)
[0359]	<223> X = Y, F, W, E
[0360]	<400> 6
[0361]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[0362]	1 5 10 15
[0363]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[0364]	20 25 30
[0365]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[0366]	35 40 45
[0367]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
[0368]	50 55 60
[0369]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0370]	65 70 75 80
[0371]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0372]	85 90 95
[0373]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0374]	100 105 110
[0375]	Lys Leu Gly Ser Ile Gln Asp Xaa Ser Arg Gly Trp Trp Val His Ser
[0376]	115 120 125
[0377]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His

[0378]	130	135	140
[0379]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[0380]	145	150	155
[0381]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met		
[0382]		165	170
[0383]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[0384]		180	185
[0385]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[0386]		195	200
[0387]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[0388]	210	215	220
[0389]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[0390]	225	230	235
[0391]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg		
[0392]		245	250
[0393]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[0394]		260	265
[0395]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		
[0396]		275	280
[0397]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		
[0398]	290	295	300
[0399]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[0400]	305	310	315
[0401]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		
[0402]		325	330
[0403]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[0404]		340	345
[0405]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[0406]		355	360
[0407]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[0408]	370	375	380
[0409]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[0410]	385	390	395
[0411]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[0412]		405	410
[0413]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[0414]		420	425
[0415]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[0416]		435	440
[0417]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[0418]	450	455	460
[0419]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		

[0420]	465	470	475
[0421]	<210> 7		
[0422]	<211> 476		
[0423]	<212> PRT		
[0424]	<213> 未知		
[0425]	<220>		
[0426]	<223> 转座酶Tn5		
[0427]	<220>		
[0428]	<221> 变体		
[0429]	<222> (119) .. (119)		
[0430]	<223> X = L, M, S, A, V		
[0431]	<220>		
[0432]	<221> 变体		
[0433]	<222> (120) .. (120)		
[0434]	<223> X = Y, F, W, E		
[0435]	<220>		
[0436]	<221> 变体		
[0437]	<222> (248) .. (248)		
[0438]	<223> X = Y, T, K, S, L, A, W, P, G, R, F, H		
[0439]	<400> 7		
[0440]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val		
[0441]	1 5 10 15		
[0442]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val		
[0443]	20 25 30		
[0444]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile		
[0445]	35 40 45		
[0446]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile		
[0447]	50 55 60		
[0448]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met		
[0449]	65 70 75 80		
[0450]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu		
[0451]	85 90 95		
[0452]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[0453]	100 105 110		
[0454]	Lys Leu Gly Ser Ile Gln Xaa Xaa Ser Arg Gly Trp Met Val His Ser		
[0455]	115 120 125		
[0456]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		
[0457]	130 135 140		
[0458]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[0459]	145 150 155 160		
[0460]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met		
[0461]	165 170 175		

[0462]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[0463]		180	185
[0464]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		190
[0465]		195	200
[0466]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		205
[0467]		210	215
[0468]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		220
[0469]		225	230
[0470]	Ile Pro Gln Lys Gly Val Val Xaa Lys Arg Gly Lys Arg Lys Asn Arg		235
[0471]		245	250
[0472]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		255
[0473]		260	265
[0474]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		270
[0475]		275	280
[0476]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		285
[0477]		290	295
[0478]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		300
[0479]		305	310
[0480]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		315
[0481]		325	330
[0482]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		335
[0483]		340	345
[0484]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		350
[0485]		355	360
[0486]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		365
[0487]		370	375
[0488]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		380
[0489]		385	390
[0490]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		395
[0491]		405	410
[0492]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		415
[0493]		420	425
[0494]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		430
[0495]		435	440
[0496]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		445
[0497]		450	455
[0498]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		460
[0499]		465	470
[0500]	<210> 8		475
[0501]	<211> 476		
[0502]	<212> PRT		
[0503]	<213> 未知		

[0504]	<220>
[0505]	<223> 转座酶Tn5
[0506]	<400> 8
[0507]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[0508]	1 5 10 15
[0509]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[0510]	20 25 30
[0511]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[0512]	35 40 45
[0513]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
[0514]	50 55 60
[0515]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0516]	65 70 75 80
[0517]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0518]	85 90 95
[0519]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0520]	100 105 110
[0521]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
[0522]	115 120 125
[0523]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[0524]	130 135 140
[0525]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
[0526]	145 150 155 160
[0527]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
[0528]	165 170 175
[0529]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
[0530]	180 185 190
[0531]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
[0532]	195 200 205
[0533]	Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
[0534]	210 215 220
[0535]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
[0536]	225 230 235 240
[0537]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
[0538]	245 250 255
[0539]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
[0540]	260 265 270
[0541]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
[0542]	275 280 285
[0543]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
[0544]	290 295 300
[0545]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr

[0546]	305	310	315	320
[0547]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala			
[0548]		325	330	335
[0549]	Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met			
[0550]		340	345	350
[0551]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu			
[0552]		355	360	365
[0553]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu			
[0554]		370	375	380
[0555]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp			
[0556]	385	390	395	400
[0557]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys			
[0558]		405	410	415
[0559]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu			
[0560]		420	425	430
[0561]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala			
[0562]		435	440	445
[0563]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu			
[0564]		450	455	460
[0565]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile			
[0566]	465	470	475	
[0567]	<210> 9			
[0568]	<211> 476			
[0569]	<212> PRT			
[0570]	<213> 未知			
[0571]	<220>			
[0572]	<223> 转座酶Tn5			
[0573]	<400> 9			
[0574]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val			
[0575]	1	5	10	15
[0576]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val			
[0577]		20	25	30
[0578]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile			
[0579]		35	40	45
[0580]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile			
[0581]		50	55	60
[0582]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met			
[0583]	65	70	75	80
[0584]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu			
[0585]		85	90	95
[0586]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly			
[0587]		100	105	110

[0588]	Lys	Leu	Gly	Ser	Ile	Gln	Asp	Lys	Ser	Arg	Gly	Trp	Trp	Val	His	Ser
[0589]		115						120					125			
[0590]	Val	Leu	Leu	Leu	Glu	Ala	Thr	Thr	Phe	Arg	Thr	Val	Gly	Leu	Leu	His
[0591]		130						135					140			
[0592]	Gln	Glu	Trp	Trp	Met	Arg	Pro	Asp	Asp	Pro	Ala	Asp	Ala	Asp	Glu	Lys
[0593]	145					150					155				160	
[0594]	Glu	Ser	Gly	Lys	Trp	Leu	Ala	Ala	Ala	Ala	Thr	Ser	Arg	Leu	Arg	Met
[0595]					165					170					175	
[0596]	Gly	Ser	Met	Met	Ser	Asn	Val	Ile	Ala	Val	Cys	Asp	Arg	Glu	Ala	Asp
[0597]					180					185					190	
[0598]	Ile	His	Ala	Tyr	Leu	Gln	Asp	Lys	Leu	Ala	His	Asn	Glu	Arg	Phe	Val
[0599]					195					200					205	
[0600]	Val	Arg	Ser	Arg	His	Arg	Arg	Lys	Asp	Val	Glu	Ser	Gly	Leu	Tyr	Leu
[0601]		210						215					220			
[0602]	Tyr	Asp	His	Leu	Lys	Asn	Gln	Pro	Glu	Leu	Gly	Gly	Tyr	Gln	Ile	Ser
[0603]	225					230					235				240	
[0604]	Ile	Pro	Gln	Lys	Gly	Val	Val	Asp	Lys	Arg	Arg	Lys	Arg	Lys	Asn	Arg
[0605]					245					250					255	
[0606]	Pro	Ala	Arg	Lys	Ala	Ser	Leu	Ser	Leu	Arg	Ser	Gly	Arg	Ile	Thr	Leu
[0607]					260					265					270	
[0608]	Lys	Gln	Gly	Asn	Ile	Thr	Leu	Asn	Ala	Val	Leu	Ala	Glu	Glu	Ile	Asn
[0609]					275					280					285	
[0610]	Pro	Pro	Lys	Gly	Glu	Thr	Pro	Leu	Lys	Trp	Leu	Leu	Leu	Thr	Ser	Glu
[0611]		290						295					300			
[0612]	Pro	Val	Glu	Ser	Leu	Ala	Gln	Ala	Leu	Arg	Val	Ile	Asp	Ile	Tyr	Thr
[0613]	305					310					315				320	
[0614]	His	Arg	Trp	Arg	Ile	Glu	Glu	Phe	His	Lys	Ala	Trp	Lys	Thr	Gly	Ala
[0615]					325					330					335	
[0616]	Gly	Val	Glu	Arg	Gln	Arg	Met	Glu	Glu	Pro	Asp	Asn	Leu	Glu	Arg	Met
[0617]					340					345					350	
[0618]	Val	Ser	Ile	Leu	Ser	Phe	Val	Ala	Val	Arg	Leu	Leu	Gln	Leu	Arg	Glu
[0619]					355					360					365	
[0620]	Ser	Phe	Thr	Leu	Pro	Gln	Ala	Leu	Arg	Ala	Gln	Gly	Leu	Leu	Lys	Glu
[0621]		370						375					380			
[0622]	Ala	Glu	His	Val	Glu	Ser	Gln	Ser	Ala	Glu	Thr	Val	Leu	Thr	Pro	Asp
[0623]	385					390					395				400	
[0624]	Glu	Cys	Gln	Leu	Leu	Gly	Tyr	Leu	Asp	Lys	Gly	Lys	Arg	Lys	Arg	Lys
[0625]					405					410					415	
[0626]	Glu	Lys	Ala	Gly	Ser	Leu	Gln	Trp	Ala	Tyr	Met	Ala	Ile	Ala	Arg	Leu
[0627]					420					425					430	
[0628]	Gly	Gly	Phe	Met	Asp	Ser	Lys	Arg	Thr	Gly	Ile	Ala	Ser	Trp	Gly	Ala
[0629]					435					440					445	

[0630]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
[0631]	450 455 460
[0632]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
[0633]	465 470 475
[0634]	<210> 10
[0635]	<211> 476
[0636]	<212> PRT
[0637]	<213> 未知
[0638]	<220>
[0639]	<223> 转座酶Tn5
[0640]	<400> 10
[0641]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[0642]	1 5 10 15
[0643]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[0644]	20 25 30
[0645]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[0646]	35 40 45
[0647]	Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
[0648]	50 55 60
[0649]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0650]	65 70 75 80
[0651]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0652]	85 90 95
[0653]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0654]	100 105 110
[0655]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
[0656]	115 120 125
[0657]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[0658]	130 135 140
[0659]	Gln Gln Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
[0660]	145 150 155 160
[0661]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
[0662]	165 170 175
[0663]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Gly Ala Asp
[0664]	180 185 190
[0665]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
[0666]	195 200 205
[0667]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
[0668]	210 215 220
[0669]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
[0670]	225 230 235 240
[0671]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg

[0672]	245	250	255
[0673]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[0674]	260	265	270
[0675]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		
[0676]	275	280	285
[0677]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		
[0678]	290	295	300
[0679]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[0680]	305	310	315
[0681]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		320
[0682]	325	330	335
[0683]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[0684]	340	345	350
[0685]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[0686]	355	360	365
[0687]	Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[0688]	370	375	380
[0689]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[0690]	385	390	395
[0691]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[0692]	405	410	415
[0693]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[0694]	420	425	430
[0695]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[0696]	435	440	445
[0697]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[0698]	450	455	460
[0699]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[0700]	465	470	475
[0701]	<210> 11		
[0702]	<211> 476		
[0703]	<212> PRT		
[0704]	<213> 未知		
[0705]	<220>		
[0706]	<223> 转座酶Tn5		
[0707]	<400> 11		
[0708]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val		
[0709]	1	5	10
[0710]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val		
[0711]	20	25	30
[0712]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile		
[0713]	35	40	45

[0714]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
[0715]	50 55 60
[0716]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0717]	65 70 75 80
[0718]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0719]	85 90 95
[0720]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0721]	100 105 110
[0722]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
[0723]	115 120 125
[0724]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[0725]	130 135 140
[0726]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
[0727]	145 150 155 160
[0728]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
[0729]	165 170 175
[0730]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
[0731]	180 185 190
[0732]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
[0733]	195 200 205
[0734]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
[0735]	210 215 220
[0736]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
[0737]	225 230 235 240
[0738]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
[0739]	245 250 255
[0740]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
[0741]	260 265 270
[0742]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
[0743]	275 280 285
[0744]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
[0745]	290 295 300
[0746]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
[0747]	305 310 315 320
[0748]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
[0749]	325 330 335
[0750]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
[0751]	340 345 350
[0752]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
[0753]	355 360 365
[0754]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
[0755]	370 375 380

[0756]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
[0757]	385 390 395 400
[0758]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
[0759]	405 410 415
[0760]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
[0761]	420 425 430
[0762]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
[0763]	435 440 445
[0764]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
[0765]	450 455 460
[0766]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
[0767]	465 470 475
[0768]	<210> 12
[0769]	<211> 476
[0770]	<212> PRT
[0771]	<213> 未知
[0772]	<220>
[0773]	<223> 转座酶Tn5
[0774]	<220>
[0775]	<221> 变体
[0776]	<222> (248) .. (248)
[0777]	<223> X = Y, T, K, S, L, A, W, P, G, R, F, H
[0778]	<400> 12
[0779]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[0780]	1 5 10 15
[0781]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[0782]	20 25 30
[0783]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[0784]	35 40 45
[0785]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
[0786]	50 55 60
[0787]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[0788]	65 70 75 80
[0789]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[0790]	85 90 95
[0791]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[0792]	100 105 110
[0793]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
[0794]	115 120 125
[0795]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[0796]	130 135 140
[0797]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys

[0798]	145	150	155	160
[0799]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met			
[0800]		165	170	175
[0801]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp			
[0802]		180	185	190
[0803]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val			
[0804]		195	200	205
[0805]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu			
[0806]		210	215	220
[0807]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser			
[0808]	225	230	235	240
[0809]	Ile Pro Gln Lys Gly Val Val Xaa Lys Arg Gly Lys Arg Lys Asn Arg			
[0810]		245	250	255
[0811]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu			
[0812]		260	265	270
[0813]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn			
[0814]		275	280	285
[0815]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu			
[0816]		290	295	300
[0817]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr			
[0818]	305	310	315	320
[0819]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala			
[0820]		325	330	335
[0821]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met			
[0822]		340	345	350
[0823]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu			
[0824]		355	360	365
[0825]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu			
[0826]		370	375	380
[0827]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp			
[0828]	385	390	395	400
[0829]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys			
[0830]		405	410	415
[0831]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu			
[0832]		420	425	430
[0833]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala			
[0834]		435	440	445
[0835]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu			
[0836]		450	455	460
[0837]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile			
[0838]	465	470	475	
[0839]	<210> 13			

[0840]	<211>	476
[0841]	<212>	PRT
[0842]	<213>	未知
[0843]	<220>	
[0844]	<223>	转座酶Tn5
[0845]	<220>	
[0846]	<221>	变体
[0847]	<222>	(119) .. (119)
[0848]	<223>	X = L, M, S, A, V
[0849]	<400>	13
[0850]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val	
[0851]	1	5 10 15
[0852]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val	
[0853]	20	25 30
[0854]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile	
[0855]	35	40 45
[0856]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile	
[0857]	50	55 60
[0858]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met	
[0859]	65	70 75 80
[0860]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu	
[0861]	85	90 95
[0862]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly	
[0863]	100	105 110
[0864]	Lys Leu Gly Ser Ile Gln Xaa Lys Ser Arg Gly Trp Trp Val His Ser	
[0865]	115	120 125
[0866]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His	
[0867]	130	135 140
[0868]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys	
[0869]	145	150 155 160
[0870]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met	
[0871]	165	170 175
[0872]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp	
[0873]	180	185 190
[0874]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val	
[0875]	195	200 205
[0876]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu	
[0877]	210	215 220
[0878]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser	
[0879]	225	230 235 240
[0880]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg	
[0881]	245	250 255

[0882]	Pro	Ala	Arg	Lys	Ala	Ser	Leu	Ser	Leu	Arg	Ser	Gly	Arg	Ile	Thr	Leu
[0883]				260					265					270		
[0884]	Lys	Gln	Gly	Asn	Ile	Thr	Leu	Asn	Ala	Val	Leu	Ala	Glu	Glu	Ile	Asn
[0885]				275					280					285		
[0886]	Pro	Pro	Lys	Gly	Glu	Thr	Pro	Leu	Lys	Trp	Leu	Leu	Leu	Thr	Ser	Glu
[0887]				290					295					300		
[0888]	Pro	Val	Glu	Ser	Leu	Ala	Gln	Ala	Leu	Arg	Val	Ile	Asp	Ile	Tyr	Thr
[0889]				305					310					315		320
[0890]	His	Arg	Trp	Arg	Ile	Glu	Glu	Phe	His	Lys	Ala	Trp	Lys	Thr	Gly	Ala
[0891]					325					330					335	
[0892]	Gly	Ala	Glu	Arg	Gln	Arg	Met	Glu	Glu	Pro	Asp	Asn	Leu	Glu	Arg	Met
[0893]					340					345					350	
[0894]	Val	Ser	Ile	Leu	Ser	Phe	Val	Ala	Val	Arg	Leu	Leu	Gln	Leu	Arg	Glu
[0895]					355					360					365	
[0896]	Ser	Phe	Thr	Pro	Pro	Gln	Ala	Leu	Arg	Ala	Gln	Gly	Leu	Leu	Lys	Glu
[0897]					370					375					380	
[0898]	Ala	Glu	His	Val	Glu	Ser	Gln	Ser	Ala	Glu	Thr	Val	Leu	Thr	Pro	Asp
[0899]					385					390					395	400
[0900]	Glu	Cys	Gln	Leu	Leu	Gly	Tyr	Leu	Asp	Lys	Gly	Lys	Arg	Lys	Arg	Lys
[0901]					405					410					415	
[0902]	Glu	Lys	Ala	Gly	Ser	Leu	Gln	Trp	Ala	Tyr	Met	Ala	Ile	Ala	Arg	Leu
[0903]					420					425					430	
[0904]	Gly	Gly	Phe	Met	Asp	Ser	Lys	Arg	Thr	Gly	Ile	Ala	Ser	Trp	Gly	Ala
[0905]					435					440					445	
[0906]	Leu	Trp	Glu	Gly	Trp	Glu	Ala	Leu	Gln	Ser	Lys	Leu	Asp	Gly	Phe	Leu
[0907]					450					455					460	
[0908]	Ala	Ala	Lys	Asp	Leu	Met	Ala	Gln	Gly	Ile	Lys	Ile				
[0909]					465					470					475	
[0910]	<210> 14															
[0911]	<211> 476															
[0912]	<212> PRT															
[0913]	<213> 未知															
[0914]	<220>															
[0915]	<223> 转座酶Tn5															
[0916]	<400> 14															
[0917]	Met	Ile	Thr	Ser	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val
[0918]					1		5				10				15	
[0919]	Phe	Ser	Ser	Ala	Ala	Leu	Gly	Asp	Pro	Arg	Arg	Thr	Ala	Arg	Leu	Val
[0920]					20					25					30	
[0921]	Asn	Val	Ala	Ala	Gln	Leu	Ala	Lys	Tyr	Ser	Gly	Lys	Ser	Ile	Thr	Ile
[0922]					35					40					45	
[0923]	Ser	Ser	Glu	Gly	Ser	Lys	Ala	Ala	Gln	Glu	Gly	Ala	Tyr	Arg	Phe	Ile

[0924]	50	55	60
[0925]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met		
[0926]	65	70	75
[0927]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu		
[0928]	85	90	95
[0929]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[0930]	100	105	110
[0931]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Met Val His Ser		
[0932]	115	120	125
[0933]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		
[0934]	130	135	140
[0935]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[0936]	145	150	155
[0937]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met		
[0938]	165	170	175
[0939]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[0940]	180	185	190
[0941]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[0942]	195	200	205
[0943]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[0944]	210	215	220
[0945]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[0946]	225	230	235
[0947]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg		
[0948]	245	250	255
[0949]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[0950]	260	265	270
[0951]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		
[0952]	275	280	285
[0953]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		
[0954]	290	295	300
[0955]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[0956]	305	310	315
[0957]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		
[0958]	325	330	335
[0959]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[0960]	340	345	350
[0961]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[0962]	355	360	365
[0963]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[0964]	370	375	380
[0965]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		

[0966]	385	390	395	400
[0967]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys			
[0968]	405	410	415	
[0969]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu			
[0970]	420	425	430	
[0971]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala			
[0972]	435	440	445	
[0973]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu			
[0974]	450	455	460	
[0975]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile			
[0976]	465	470	475	
[0977]	<210> 15			
[0978]	<211> 477			
[0979]	<212> PRT			
[0980]	<213> 未知			
[0981]	<220>			
[0982]	<223> 转座酶Tn5			
[0983]	<400> 15			
[0984]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val			
[0985]	1	5	10	15
[0986]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val			
[0987]	20	25	30	
[0988]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile			
[0989]	35	40	45	
[0990]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile			
[0991]	50	55	60	
[0992]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met			
[0993]	65	70	75	80
[0994]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu			
[0995]	85	90	95	
[0996]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly			
[0997]	100	105	110	
[0998]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser			
[0999]	115	120	125	
[1000]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His			
[1001]	130	135	140	
[1002]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys			
[1003]	145	150	155	160
[1004]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met			
[1005]	165	170	175	
[1006]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp			
[1007]	180	185	190	

[1008]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[1009]	195	200	205
[1010]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[1011]	210	215	220
[1012]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[1013]	225	230	235 240
[1014]	Ile Pro Gln Lys Gly Val Val Asp Val Lys Arg Gly Lys Arg Lys Asn		
[1015]	245	250	255
[1016]	Arg Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr		
[1017]	260	265	270
[1018]	Leu Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile		
[1019]	275	280	285
[1020]	Asn Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser		
[1021]	290	295	300
[1022]	Glu Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr		
[1023]	305	310	315 320
[1024]	Thr His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly		
[1025]	325	330	335
[1026]	Ala Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg		
[1027]	340	345	350
[1028]	Met Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg		
[1029]	355	360	365
[1030]	Glu Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys		
[1031]	370	375	380
[1032]	Glu Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro		
[1033]	385	390	395 400
[1034]	Asp Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg		
[1035]	405	410	415
[1036]	Lys Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg		
[1037]	420	425	430
[1038]	Leu Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly		
[1039]	435	440	445
[1040]	Ala Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe		
[1041]	450	455	460
[1042]	Leu Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[1043]	465	470	475
[1044]	<210> 16		
[1045]	<211> 476		
[1046]	<212> PRT		
[1047]	<213> 未知		
[1048]	<220>		
[1049]	<223> 转座酶Tn5		

[1050]	<220>
[1051]	<221> 变体
[1052]	<222> (120) .. (120)
[1053]	<223> X = Y, F, W, E
[1054]	<400> 16
[1055]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[1056]	1 5 10 15
[1057]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[1058]	20 25 30
[1059]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[1060]	35 40 45
[1061]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
[1062]	50 55 60
[1063]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
[1064]	65 70 75 80
[1065]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
[1066]	85 90 95
[1067]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
[1068]	100 105 110
[1069]	Lys Leu Gly Ser Ile Gln Asp Xaa Ser Arg Gly Trp Trp Val His Ser
[1070]	115 120 125
[1071]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
[1072]	130 135 140
[1073]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
[1074]	145 150 155 160
[1075]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
[1076]	165 170 175
[1077]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
[1078]	180 185 190
[1079]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
[1080]	195 200 205
[1081]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
[1082]	210 215 220
[1083]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
[1084]	225 230 235 240
[1085]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
[1086]	245 250 255
[1087]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
[1088]	260 265 270
[1089]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
[1090]	275 280 285
[1091]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu

[1092]	290	295	300
[1093]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[1094]	305	310	315 320
[1095]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		
[1096]	325	330	335
[1097]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[1098]	340	345	350
[1099]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[1100]	355	360	365
[1101]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[1102]	370	375	380
[1103]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[1104]	385	390	395 400
[1105]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[1106]	405	410	415
[1107]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[1108]	420	425	430
[1109]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[1110]	435	440	445
[1111]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[1112]	450	455	460
[1113]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[1114]	465	470	475
[1115]	<210> 17		
[1116]	<211> 476		
[1117]	<212> PRT		
[1118]	<213> 未知		
[1119]	<220>		
[1120]	<223> 转座酶Tn5		
[1121]	<220>		
[1122]	<221> 变体		
[1123]	<222> (119) .. (119)		
[1124]	<223> X = L, M, S, A, V		
[1125]	<220>		
[1126]	<221> 变体		
[1127]	<222> (120) .. (120)		
[1128]	<223> X = Y, F, W, E		
[1129]	<220>		
[1130]	<221> 变体		
[1131]	<222> (248) .. (248)		
[1132]	<223> X = Y, T, K, S, L, A, W, P, G, R, F, H		
[1133]	<400> 17		

[1134]	Met	Ile	Thr	Ser	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val
[1135]	1				5					10					15	
[1136]	Phe	Ser	Ser	Ala	Ala	Leu	Gly	Asp	Pro	Arg	Arg	Thr	Ala	Arg	Leu	Val
[1137]					20					25					30	
[1138]	Asn	Val	Ala	Ala	Gln	Leu	Ala	Lys	Tyr	Ser	Gly	Lys	Ser	Ile	Thr	Ile
[1139]					35					40					45	
[1140]	Ser	Ser	Glu	Gly	Ser	Lys	Ala	Ala	Gln	Glu	Gly	Ala	Tyr	Arg	Phe	Ile
[1141]					50					55					60	
[1142]	Arg	Asn	Pro	Asn	Val	Ser	Ala	Glu	Ala	Ile	Arg	Lys	Ala	Gly	Ala	Met
[1143]	65									70					75	
[1144]	Gln	Thr	Val	Lys	Leu	Ala	Gln	Glu	Phe	Pro	Glu	Leu	Leu	Ala	Ile	Glu
[1145]					85					90					95	
[1146]	Asp	Thr	Thr	Ser	Leu	Ser	Tyr	Arg	His	Gln	Val	Ala	Glu	Glu	Leu	Gly
[1147]					100					105					110	
[1148]	Lys	Leu	Gly	Ser	Ile	Gln	Xaa	Xaa	Ser	Arg	Gly	Trp	Met	Val	His	Ser
[1149]					115					120					125	
[1150]	Val	Leu	Leu	Leu	Glu	Ala	Thr	Thr	Phe	Arg	Thr	Val	Gly	Leu	Leu	His
[1151]					130					135					140	
[1152]	Gln	Glu	Trp	Trp	Met	Arg	Pro	Asp	Asp	Pro	Ala	Asp	Ala	Asp	Glu	Lys
[1153]	145									150					155	
[1154]	Glu	Ser	Gly	Lys	Trp	Leu	Ala	Ala	Ala	Ala	Thr	Ser	Arg	Leu	Arg	Met
[1155]					165					170					175	
[1156]	Gly	Ser	Met	Met	Ser	Asn	Val	Ile	Ala	Val	Cys	Asp	Arg	Glu	Ala	Asp
[1157]					180					185					190	
[1158]	Ile	His	Ala	Tyr	Leu	Gln	Asp	Lys	Leu	Ala	His	Asn	Glu	Arg	Phe	Val
[1159]					195					200					205	
[1160]	Val	Arg	Ser	Lys	His	Pro	Arg	Lys	Asp	Val	Glu	Ser	Gly	Leu	Tyr	Leu
[1161]					210					215					220	
[1162]	Tyr	Asp	His	Leu	Lys	Asn	Gln	Pro	Glu	Leu	Gly	Gly	Tyr	Gln	Ile	Ser
[1163]	225									230					235	
[1164]	Ile	Pro	Gln	Lys	Gly	Val	Val	Xaa	Lys	Arg	Gly	Lys	Arg	Lys	Asn	Arg
[1165]					245					250					255	
[1166]	Pro	Ala	Arg	Lys	Ala	Ser	Leu	Ser	Leu	Arg	Ser	Gly	Arg	Ile	Thr	Leu
[1167]					260					265					270	
[1168]	Lys	Gln	Gly	Asn	Ile	Thr	Leu	Asn	Ala	Val	Leu	Ala	Glu	Glu	Ile	Asn
[1169]					275					280					285	
[1170]	Pro	Pro	Lys	Gly	Glu	Thr	Pro	Leu	Lys	Trp	Leu	Leu	Leu	Thr	Ser	Glu
[1171]					290					295					300	
[1172]	Pro	Val	Glu	Ser	Leu	Ala	Gln	Ala	Leu	Arg	Val	Ile	Asp	Ile	Tyr	Thr
[1173]	305									310					315	
[1174]	His	Arg	Trp	Arg	Ile	Glu	Glu	Phe	His	Lys	Ala	Trp	Lys	Thr	Gly	Ala
[1175]					325					330					335	

[1176]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[1177]		340	350
[1178]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[1179]		355	365
[1180]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[1181]		370	380
[1182]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[1183]		385	400
[1184]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[1185]		405	415
[1186]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[1187]		420	430
[1188]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[1189]		435	445
[1190]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[1191]		450	460
[1192]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[1193]		465	475
[1194]	<210> 18		
[1195]	<211> 476		
[1196]	<212> PRT		
[1197]	<213> 未知		
[1198]	<220>		
[1199]	<223> 转座酶Tn5		
[1200]	<400> 18		
[1201]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val		
[1202]	1	5	15
[1203]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val		
[1204]		20	30
[1205]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile		
[1206]		35	45
[1207]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile		
[1208]		50	60
[1209]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met		
[1210]		65	80
[1211]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu		
[1212]		85	95
[1213]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[1214]		100	110
[1215]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser		
[1216]		115	125
[1217]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		

[1218]	130	135	140
[1219]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[1220]	145	150	155
[1221]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met		
[1222]	165	170	175
[1223]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[1224]	180	185	190
[1225]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[1226]	195	200	205
[1227]	Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[1228]	210	215	220
[1229]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[1230]	225	230	235
[1231]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg		
[1232]	245	250	255
[1233]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[1234]	260	265	270
[1235]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn		
[1236]	275	280	285
[1237]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu		
[1238]	290	295	300
[1239]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr		
[1240]	305	310	315
[1241]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala		
[1242]	325	330	335
[1243]	Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met		
[1244]	340	345	350
[1245]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu		
[1246]	355	360	365
[1247]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu		
[1248]	370	375	380
[1249]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp		
[1250]	385	390	395
[1251]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys		
[1252]	405	410	415
[1253]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[1254]	420	425	430
[1255]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[1256]	435	440	445
[1257]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[1258]	450	455	460
[1259]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		

[1260]	465	470	475
[1261]	<210> 19		
[1262]	<211> 476		
[1263]	<212> PRT		
[1264]	<213> 未知		
[1265]	<220>		
[1266]	<223> 转座酶Tn5		
[1267]	<400> 19		
[1268]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val		
[1269]	1 5 10 15		
[1270]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val		
[1271]	20 25 30		
[1272]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile		
[1273]	35 40 45		
[1274]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile		
[1275]	50 55 60		
[1276]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met		
[1277]	65 70 75 80		
[1278]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu		
[1279]	85 90 95		
[1280]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly		
[1281]	100 105 110		
[1282]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser		
[1283]	115 120 125		
[1284]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His		
[1285]	130 135 140		
[1286]	Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys		
[1287]	145 150 155 160		
[1288]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met		
[1289]	165 170 175		
[1290]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp		
[1291]	180 185 190		
[1292]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val		
[1293]	195 200 205		
[1294]	Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu		
[1295]	210 215 220		
[1296]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser		
[1297]	225 230 235 240		
[1298]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg		
[1299]	245 250 255		
[1300]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu		
[1301]	260 265 270		

[1302]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
[1303]	275 280 285
[1304]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
[1305]	290 295 300
[1306]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
[1307]	305 310 315 320
[1308]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
[1309]	325 330 335
[1310]	Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
[1311]	340 345 350
[1312]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
[1313]	355 360 365
[1314]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
[1315]	370 375 380
[1316]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
[1317]	385 390 395 400
[1318]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
[1319]	405 410 415
[1320]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
[1321]	420 425 430
[1322]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
[1323]	435 440 445
[1324]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
[1325]	450 455 460
[1326]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
[1327]	465 470 475
[1328]	<210> 20
[1329]	<211> 476
[1330]	<212> PRT
[1331]	<213> 未知
[1332]	<220>
[1333]	<223> 转座酶Tn5
[1334]	<400> 20
[1335]	Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
[1336]	1 5 10 15
[1337]	Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
[1338]	20 25 30
[1339]	Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
[1340]	35 40 45
[1341]	Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
[1342]	50 55 60
[1343]	Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met

[1344]	65	70	75	80
[1345]	Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu			
[1346]		85	90	95
[1347]	Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly			
[1348]		100	105	110
[1349]	Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser			
[1350]		115	120	125
[1351]	Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His			
[1352]		130	135	140
[1353]	Gln Gln Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys			
[1354]	145	150	155	160
[1355]	Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met			
[1356]		165	170	175
[1357]	Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Gly Ala Asp			
[1358]		180	185	190
[1359]	Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val			
[1360]		195	200	205
[1361]	Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu			
[1362]		210	215	220
[1363]	Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser			
[1364]	225	230	235	240
[1365]	Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg			
[1366]		245	250	255
[1367]	Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu			
[1368]		260	265	270
[1369]	Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn			
[1370]		275	280	285
[1371]	Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu			
[1372]		290	295	300
[1373]	Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr			
[1374]	305	310	315	320
[1375]	His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala			
[1376]		325	330	335
[1377]	Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met			
[1378]		340	345	350
[1379]	Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu			
[1380]		355	360	365
[1381]	Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu			
[1382]		370	375	380
[1383]	Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp			
[1384]	385	390	395	400
[1385]	Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys			

[1386]	405	410	415
[1387]	Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu		
[1388]	420	425	430
[1389]	Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala		
[1390]	435	440	445
[1391]	Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu		
[1392]	450	455	460
[1393]	Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile		
[1394]	465	470	475
[1395]	<210> 21		
[1396]	<211> 10		
[1397]	<212> PRT		
[1398]	<213> 未知		
[1399]	<220>		
[1400]	<223> 转座酶Tn5		
[1401]	<220>		
[1402]	<221> 变体		
[1403]	<222> (1) .. (1)		
[1404]	<223> x= W, R, K, P, T		
[1405]	<220>		
[1406]	<221> 变体		
[1407]	<222> (2) .. (2)		
[1408]	<223> x= W, N, V, K, M		
[1409]	<220>		
[1410]	<221> 变体		
[1411]	<222> (3) .. (3)		
[1412]	<223> x= V, F, I, T, L		
[1413]	<220>		
[1414]	<221> 变体		
[1415]	<222> (4) .. (4)		
[1416]	<223> x= H, L, W, C		
[1417]	<220>		
[1418]	<221> 变体		
[1419]	<222> (5) .. (5)		
[1420]	<223> x= S, G, V, F		
[1421]	<220>		
[1422]	<221> 变体		
[1423]	<222> (6) .. (6)		
[1424]	<223> x= V, A, W		
[1425]	<220>		
[1426]	<221> 变体		
[1427]	<222> (7) .. (7)		

- [1428] <223> x= L, T, V, Q, W
- [1429] <220>
- [1430] <221> 变体
- [1431] <222> (8) .. (8)
- [1432] <223> x= L, H, D
- [1433] <220>
- [1434] <221> 变体
- [1435] <222> (9) .. (9)
- [1436] <223> x= L, H, V
- [1437] <220>
- [1438] <221> 变体
- [1439] <222> (10) .. (10)
- [1440] <223> x= E, Y, A, K, Q
- [1441] <400> 21
- [1442] Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
- [1443] 1 5 10
- [1444] <210> 22
- [1445] <211> 852
- [1446] <212> DNA
- [1447] <213> 大肠杆菌(Escherichia coli)
- [1448] <400> 22
- [1449] atggctgaaa ttaccgcac cctggtaaaa gagctgcgtg agcgtactgg cgcaggcatg 60
- [1450] atggattgca aaaaagcact gactgaagct aacggcgaca tcgagctggc aatcgaaaac 120
- [1451] atgcgtaagt ccggtgctat taaagcagcg aaaaaagcag gcaacgttgc tgctgacggc 180
- [1452] gtgatcaaaa ccaaaatcga cggcaactac ggcatcattc tggaagttaa ctgccagact 240
- [1453] gacttcgttg caaaagacgc tggtttccag gcgttcgcag acaaagttct ggacgcagct 300
- [1454] gttgctggca aaatcactga cgttgaagtt ctgaaagcac agttcgaaga agaacgtgtt 360
- [1455] gcgctggtag cgaaaattgg tgaaaacatc aacattcgcc gcgttgctgc gctggaaggc 420
- [1456] gacgttctgg gttcttatca gcacgggtcg cgtatcggcg ttctggttgc tgctaaaggc 480
- [1457] gctgacgaag agctggttaa acacatcgct atgcacgttg ctgcaagcaa gccagaattc 540
- [1458] atcaaacggg aagacgtatc cgctgaagtg gtagaaaaag aataccaggt acagctggat 600
- [1459] atcgcgatgc agtctggtaa gccgaaagaa atcgcagaga aaatggttga aggccgcatg 660
- [1460] aagaaattca ccggcgaagt ttctctgacc ggtcagccgt tcgttatgga accaagcaaa 720
- [1461] actgttggtc agctgctgaa agagcataac gctgaagtga ctggcttcat ccgcttcgaa 780
- [1462] gtgggtgaag gcatcgagaa agttgagact gactttgcag cagaagttgc tgcgatgtcc 840
- [1463] aagcagtctt aa 852
- [1464] <210> 23
- [1465] <211> 283
- [1466] <212> PRT
- [1467] <213> 大肠杆菌
- [1468] <400> 23
- [1469] Met Ala Glu Ile Thr Ala Ser Leu Val Lys Glu Leu Arg Glu Arg Thr

[1470]	1	5					10					15				
[1471]	Gly	Ala	Gly	Met	Met	Asp	Cys	Lys	Lys	Ala	Leu	Thr	Glu	Ala	Asn	Gly
[1472]	20					25					30					
[1473]	Asp	Ile	Glu	Leu	Ala	Ile	Glu	Asn	Met	Arg	Lys	Ser	Gly	Ala	Ile	Lys
[1474]	35					40					45					
[1475]	Ala	Ala	Lys	Lys	Ala	Gly	Asn	Val	Ala	Ala	Asp	Gly	Val	Ile	Lys	Thr
[1476]	50					55					60					
[1477]	Lys	Ile	Asp	Gly	Asn	Tyr	Gly	Ile	Ile	Leu	Glu	Val	Asn	Cys	Gln	Thr
[1478]	65	70					75					80				
[1479]	Asp	Phe	Val	Ala	Lys	Asp	Ala	Gly	Phe	Gln	Ala	Phe	Ala	Asp	Lys	Val
[1480]	85					90					95					
[1481]	Leu	Asp	Ala	Ala	Val	Ala	Gly	Lys	Ile	Thr	Asp	Val	Glu	Val	Leu	Lys
[1482]	100					105					110					
[1483]	Ala	Gln	Phe	Glu	Glu	Glu	Arg	Val	Ala	Leu	Val	Ala	Lys	Ile	Gly	Glu
[1484]	115					120					125					
[1485]	Asn	Ile	Asn	Ile	Arg	Arg	Val	Ala	Ala	Leu	Glu	Gly	Asp	Val	Leu	Gly
[1486]	130					135					140					
[1487]	Ser	Tyr	Gln	His	Gly	Ala	Arg	Ile	Gly	Val	Leu	Val	Ala	Ala	Lys	Gly
[1488]	145	150					155					160				
[1489]	Ala	Asp	Glu	Glu	Leu	Val	Lys	His	Ile	Ala	Met	His	Val	Ala	Ala	Ser
[1490]	165					170					175					
[1491]	Lys	Pro	Glu	Phe	Ile	Lys	Pro	Glu	Asp	Val	Ser	Ala	Glu	Val	Val	Glu
[1492]	180					185					190					
[1493]	Lys	Glu	Tyr	Gln	Val	Gln	Leu	Asp	Ile	Ala	Met	Gln	Ser	Gly	Lys	Pro
[1494]	195					200					205					
[1495]	Lys	Glu	Ile	Ala	Glu	Lys	Met	Val	Glu	Gly	Arg	Met	Lys	Lys	Phe	Thr
[1496]	210					215					220					
[1497]	Gly	Glu	Val	Ser	Leu	Thr	Gly	Gln	Pro	Phe	Val	Met	Glu	Pro	Ser	Lys
[1498]	225	230					235					240				
[1499]	Thr	Val	Gly	Gln	Leu	Leu	Lys	Glu	His	Asn	Ala	Glu	Val	Thr	Gly	Phe
[1500]	245					250					255					
[1501]	Ile	Arg	Phe	Glu	Val	Gly	Glu	Gly	Ile	Glu	Lys	Val	Glu	Thr	Asp	Phe
[1502]	260					265					270					
[1503]	Ala	Ala	Glu	Val	Ala	Ala	Met	Ser	Lys	Gln	Ser					
[1504]	275					280										
[1505]	<210> 24															
[1506]	<211> 2283															
[1507]	<212> DNA															
[1508]	<213> 未知															
[1509]	<220>															
[1510]	<223> TS-Tn5核酸序列															
[1511]	<400> 24															

[1512]	atggctgaaa ttaccgcatc cctggtaaaa gagctgcgtg agcgtactgg cgcaggcatg	60
[1513]	atggattgca aaaaagcact gactgaagct aacggcgaca tcgagctggc aatcgaaaac	120
[1514]	atgcgtaagt ccggtgctat taaagcagcg aaaaaagcag gcaacgttgc tgctgacggc	180
[1515]	gtgatcaaaa ccaaaatcga cggcaactac ggcatcattc tggaagttaa ctgccagact	240
[1516]	gacttcgttg caaaagacgc tggtttccag gcgttcgcag acaaagttct ggacgcagct	300
[1517]	gttgctggca aaatcactga cgttgaagtt ctgaaagcac agttcgaaga agaacgtgtt	360
[1518]	gcgctggtag cgaaaattgg tgaaaacatc aacattcgcc gcgttgctgc gctggaaggc	420
[1519]	gacgttctgg gttcttatca gcacggtcgc cgtatcggcg ttcttggttg tgctaaaggc	480
[1520]	gctgacgaag agctggttaa acacatcgct atgcacgttg ctgcaagcaa gccagaattc	540
[1521]	atcaaacggg aagacgtatc cgctgaagtg gtagaaaaag aataccaggt acagctggat	600
[1522]	atcgcgatgc agtctggtaa gccgaaagaa atcgcagaga aaatggttga aggccgcatg	660
[1523]	aagaaattca ccggcgaagt ttctctgacc ggtcagccgt tcgttatgga accaagcaaa	720
[1524]	actgttggtc agctgctgaa agagcataac gctgaagtga ctggcttcat ccgcttcgaa	780
[1525]	gtgggtgaag gcatcgagaa agttgagact gactttgcag cagaagttgc tgcgatgtcc	840
[1526]	aagcagtctg gtaccataac ttctgctctt catcgtgcgg ccgactgggc taaatctgtg	900
[1527]	ttctcttcgg cggcgtggg tgatcctcgc cgtactgccc gcttggttaa cgtcgccgcc	960
[1528]	caattggcaa aatattctgg taaatcaata accatctcat cagagggtag taaagccgcc	1020
[1529]	caggaaggcg cttaccgatt tatccgcaat cccaacgttt ctgccgaggc gatcagaaag	1080
[1530]	gctggcgcca tgcaaacagt caagttggct caggagtttc ccgaactgct ggccattgag	1140
[1531]	gacaccacct ctttgagtta tcgccaccag gtcccggaag agcttggtgaa gctgggctct	1200
[1532]	attcaggata aatcccgcgg atggtgggtt cactccgttc tcttgctcga ggccaccaca	1260
[1533]	ttccgcaccg taggattact gcatcaggag tgggtgatgc gcccgatga ccctgccgat	1320
[1534]	gcggatgaaa aggagagtgg caaatggctg gcagcggccg caactagccg gttacgcatg	1380
[1535]	ggcagcatga tgagcaacgt gattgcggtc tgtgaccgcg aagccgatat tcatgcttat	1440
[1536]	ctgcaggaca aactggcgca taacgagcgc ttctgtgtgc gctccaagca cccacgcaag	1500
[1537]	gacgtagagt ctgggttgta tctgtacgac catctgaaga accaaccgga gttgggtggc	1560
[1538]	tatcagatca gcattccgca aaaggcgctg gtggataaac gcggtaaacg taaaaatcga	1620
[1539]	ccagcccgcga aggcgagctt gagcctgcgc agtgggcgca tcacgctaaa acaggggaat	1680
[1540]	atcacgctca acgcgtgct ggccgaggag attaacccgc ccaagggtga gaccccggtg	1740
[1541]	aaatggttgt tgctgaccag cgaaccggtc gagtcgctag cccaagcctt gcgcgtcatc	1800
[1542]	gacatttata cccatcgctg gcggatcgag gagttccata aggcattgaa aaccggagca	1860
[1543]	ggagccgaga ggcaacgcat ggaggagccg gataatctgg agcggtatgt ctcgatcctc	1920
[1544]	tcgtttgttg cggtcaggct gttacagctc agagaaagct tcacgccgcc gcaagcactc	1980
[1545]	agggcgcaag ggctgctaaa ggaagcgga cactagaaa gccagtccgc agaaacggtg	2040
[1546]	ctgaccccg atgaatgtca gctactgggc tatctggaca agggaaaacg caagcgcaaa	2100
[1547]	gagaaagcag gtagcttgca gtgggcttac atggcgatag ctagactggg cggttttatg	2160
[1548]	gacagcaagc gaaccggaat tgccagctgg ggccgctct gggaagggtg ggaagccctg	2220
[1549]	caaagtaaac tggatggctt tcttgccgcc aaggatctga tggcgcaggg gatcaagatc	2280
[1550]	tga	2283
[1551]	<210>	25
[1552]	<211>	760
[1553]	<212>	PRT

[1554]	<213> 未知																		
[1555]	<220>																		
[1556]	<223> TS-Tn5融合蛋白氨基酸序列																		
[1557]	<400> 25																		
[1558]	Met	Ala	Glu	Ile	Thr	Ala	Ser	Leu	Val	Lys	Glu	Leu	Arg	Glu	Arg	Thr			
[1559]	1				5					10					15				
[1560]	Gly	Ala	Gly	Met	Met	Asp	Cys	Lys	Lys	Ala	Leu	Thr	Glu	Ala	Asn	Gly			
[1561]					20					25					30				
[1562]	Asp	Ile	Glu	Leu	Ala	Ile	Glu	Asn	Met	Arg	Lys	Ser	Gly	Ala	Ile	Lys			
[1563]					35					40					45				
[1564]	Ala	Ala	Lys	Lys	Ala	Gly	Asn	Val	Ala	Ala	Asp	Gly	Val	Ile	Lys	Thr			
[1565]					50					55					60				
[1566]	Lys	Ile	Asp	Gly	Asn	Tyr	Gly	Ile	Ile	Leu	Glu	Val	Asn	Cys	Gln	Thr			
[1567]	65									70					75				80
[1568]	Asp	Phe	Val	Ala	Lys	Asp	Ala	Gly	Phe	Gln	Ala	Phe	Ala	Asp	Lys	Val			
[1569]										85					90				95
[1570]	Leu	Asp	Ala	Ala	Val	Ala	Gly	Lys	Ile	Thr	Asp	Val	Glu	Val	Leu	Lys			
[1571]										100					105				110
[1572]	Ala	Gln	Phe	Glu	Glu	Glu	Arg	Val	Ala	Leu	Val	Ala	Lys	Ile	Gly	Glu			
[1573]										115					120				125
[1574]	Asn	Ile	Asn	Ile	Arg	Arg	Val	Ala	Ala	Leu	Glu	Gly	Asp	Val	Leu	Gly			
[1575]										130					135				140
[1576]	Ser	Tyr	Gln	His	Gly	Ala	Arg	Ile	Gly	Val	Leu	Val	Ala	Ala	Lys	Gly			
[1577]	145									150					155				160
[1578]	Ala	Asp	Glu	Glu	Leu	Val	Lys	His	Ile	Ala	Met	His	Val	Ala	Ala	Ser			
[1579]										165					170				175
[1580]	Lys	Pro	Glu	Phe	Ile	Lys	Pro	Glu	Asp	Val	Ser	Ala	Glu	Val	Val	Glu			
[1581]										180					185				190
[1582]	Lys	Glu	Tyr	Gln	Val	Gln	Leu	Asp	Ile	Ala	Met	Gln	Ser	Gly	Lys	Pro			
[1583]										195					200				205
[1584]	Lys	Glu	Ile	Ala	Glu	Lys	Met	Val	Glu	Gly	Arg	Met	Lys	Lys	Phe	Thr			
[1585]										210					215				220
[1586]	Gly	Glu	Val	Ser	Leu	Thr	Gly	Gln	Pro	Phe	Val	Met	Glu	Pro	Ser	Lys			
[1587]	225									230					235				240
[1588]	Thr	Val	Gly	Gln	Leu	Leu	Lys	Glu	His	Asn	Ala	Glu	Val	Thr	Gly	Phe			
[1589]										245					250				255
[1590]	Ile	Arg	Phe	Glu	Val	Gly	Glu	Gly	Ile	Glu	Lys	Val	Glu	Thr	Asp	Phe			
[1591]										260					265				270
[1592]	Ala	Ala	Glu	Val	Ala	Ala	Met	Ser	Lys	Gln	Ser	Gly	Thr	Ile	Thr	Ser			
[1593]										275					280				285
[1594]	Ala	Leu	His	Arg	Ala	Ala	Asp	Trp	Ala	Lys	Ser	Val	Phe	Ser	Ser	Ala			
[1595]										290					295				300

[1596]	Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val Asn Val Ala Ala
[1597]	305 310 315 320
[1598]	Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile Ser Ser Glu Gly
[1599]	325 330 335
[1600]	Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile Arg Asn Pro Asn
[1601]	340 345 350
[1602]	Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met Gln Thr Val Lys
[1603]	355 360 365
[1604]	Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu Asp Thr Thr Ser
[1605]	370 375 380
[1606]	Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly Lys Leu Gly Ser
[1607]	385 390 395 400
[1608]	Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser Val Leu Leu Leu
[1609]	405 410 415
[1610]	Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His Gln Glu Trp Trp
[1611]	420 425 430
[1612]	Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys Glu Ser Gly Lys
[1613]	435 440 445
[1614]	Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met Gly Ser Met Met
[1615]	450 455 460
[1616]	Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp Ile His Ala Tyr
[1617]	465 470 475 480
[1618]	Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val Val Arg Ser Lys
[1619]	485 490 495
[1620]	His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu Tyr Asp His Leu
[1621]	500 505 510
[1622]	Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser Ile Pro Gln Lys
[1623]	515 520 525
[1624]	Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg Pro Ala Arg Lys
[1625]	530 535 540
[1626]	Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu Lys Gln Gly Asn
[1627]	545 550 555 560
[1628]	Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn Pro Pro Lys Gly
[1629]	565 570 575
[1630]	Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu Pro Val Glu Ser
[1631]	580 585 590
[1632]	Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr His Arg Trp Arg
[1633]	595 600 605
[1634]	Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala Gly Ala Glu Arg
[1635]	610 615 620
[1636]	Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met Val Ser Ile Leu
[1637]	625 630 635 640

[1638]	Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu Ser Phe Thr Pro
[1639]	645 650 655
[1640]	Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu Ala Glu His Val
[1641]	660 665 670
[1642]	Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp Glu Cys Gln Leu
[1643]	675 680 685
[1644]	Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys Glu Lys Ala Gly
[1645]	690 695 700
[1646]	Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu Gly Gly Phe Met
[1647]	705 710 715 720
[1648]	Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala Leu Trp Glu Gly
[1649]	725 730 735
[1650]	Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu Ala Ala Lys Asp
[1651]	740 745 750
[1652]	Leu Met Ala Gln Gly Ile Lys Ile
[1653]	755 760
[1654]	<210> 26
[1655]	<211> 760
[1656]	<212> PRT
[1657]	<213> 未知
[1658]	<220>
[1659]	<223> TS-Tn突变体蛋白
[1660]	<400> 26
[1661]	Met Ala Glu Ile Thr Ala Ser Leu Val Lys Glu Leu Arg Glu Arg Thr
[1662]	1 5 10 15
[1663]	Gly Ala Gly Met Met Asp Cys Lys Lys Ala Leu Thr Glu Ala Asn Gly
[1664]	20 25 30
[1665]	Asp Ile Glu Leu Ala Ile Glu Asn Met Arg Lys Ser Gly Ala Ile Lys
[1666]	35 40 45
[1667]	Ala Ala Lys Lys Ala Gly Asn Val Ala Ala Asp Gly Val Ile Lys Thr
[1668]	50 55 60
[1669]	Lys Ile Asp Gly Asn Tyr Gly Ile Ile Leu Glu Val Asn Cys Gln Thr
[1670]	65 70 75 80
[1671]	Asp Phe Val Ala Lys Asp Ala Gly Phe Gln Ala Phe Ala Asp Lys Val
[1672]	85 90 95
[1673]	Leu Asp Ala Ala Val Ala Gly Lys Ile Thr Asp Val Glu Val Leu Lys
[1674]	100 105 110
[1675]	Ala Gln Phe Glu Glu Glu Arg Val Ala Leu Val Ala Lys Ile Gly Glu
[1676]	115 120 125
[1677]	Asn Ile Asn Ile Arg Arg Val Ala Ala Leu Glu Gly Asp Val Leu Gly
[1678]	130 135 140
[1679]	Ser Tyr Gln His Gly Ala Arg Ile Gly Val Leu Val Ala Ala Lys Gly

[1680]	145	150	155	160
[1681]	Ala Asp Glu Glu Leu Val Lys His Ile Ala Met His Val Ala Ala Ser			
[1682]		165	170	175
[1683]	Lys Pro Glu Phe Ile Lys Pro Glu Asp Val Ser Ala Glu Val Val Glu			
[1684]		180	185	190
[1685]	Lys Glu Tyr Gln Val Gln Leu Asp Ile Ala Met Gln Ser Gly Lys Pro			
[1686]		195	200	205
[1687]	Lys Glu Ile Ala Glu Lys Met Val Glu Gly Arg Met Lys Lys Phe Thr			
[1688]	210	215	220	
[1689]	Gly Glu Val Ser Leu Thr Gly Gln Pro Phe Val Met Glu Pro Ser Lys			
[1690]	225	230	235	240
[1691]	Thr Val Gly Gln Leu Leu Lys Glu His Asn Ala Glu Val Thr Gly Phe			
[1692]		245	250	255
[1693]	Ile Arg Phe Glu Val Gly Glu Gly Ile Glu Lys Val Glu Thr Asp Phe			
[1694]		260	265	270
[1695]	Ala Ala Glu Val Ala Ala Met Ser Lys Gln Ser Gly Thr Ile Thr Ser			
[1696]		275	280	285
[1697]	Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val Phe Ser Ser Ala			
[1698]	290	295	300	
[1699]	Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val Asn Val Ala Ala			
[1700]	305	310	315	320
[1701]	Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile Ser Ser Glu Gly			
[1702]		325	330	335
[1703]	Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile Arg Asn Pro Asn			
[1704]		340	345	350
[1705]	Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met Gln Thr Val Lys			
[1706]		355	360	365
[1707]	Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu Asp Thr Thr Ser			
[1708]	370	375	380	
[1709]	Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly Lys Leu Gly Ser			
[1710]	385	390	395	400
[1711]	Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser Val Leu Leu Leu			
[1712]		405	410	415
[1713]	Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His Gln Glu Trp Trp			
[1714]		420	425	430
[1715]	Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys Glu Ser Gly Lys			
[1716]		435	440	445
[1717]	Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met Gly Ser Met Met			
[1718]	450	455	460	
[1719]	Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp Ile His Ala Tyr			
[1720]	465	470	475	480
[1721]	Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val Val Arg Ser Arg			

[1722]		485		490		495
[1723]	His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu Tyr Asp His Leu					
[1724]		500		505		510
[1725]	Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser Ile Pro Gln Lys					
[1726]		515		520		525
[1727]	Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg Pro Ala Arg Lys					
[1728]		530		535		540
[1729]	Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu Lys Gln Gly Asn					
[1730]	545		550		555	560
[1731]	Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn Pro Pro Lys Gly					
[1732]		565		570		575
[1733]	Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu Pro Val Glu Ser					
[1734]		580		585		590
[1735]	Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr His Arg Trp Arg					
[1736]		595		600		605
[1737]	Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala Gly Val Glu Arg					
[1738]		610		615		620
[1739]	Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met Val Ser Ile Leu					
[1740]	625		630		635	640
[1741]	Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu Ser Phe Thr Pro					
[1742]		645		650		655
[1743]	Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu Ala Glu His Val					
[1744]		660		665		670
[1745]	Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp Glu Cys Gln Leu					
[1746]		675		680		685
[1747]	Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys Glu Lys Ala Gly					
[1748]		690		695		700
[1749]	Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu Gly Gly Phe Met					
[1750]	705		710		715	720
[1751]	Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala Leu Trp Glu Gly					
[1752]		725		730		735
[1753]	Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu Ala Ala Lys Asp					
[1754]		740		745		750
[1755]	Leu Met Ala Gln Gly Ile Lys Ile					
[1756]		755		760		

71 Tn5核心 Hermes核心 整合酶核心 Mu核心 Mos1核心	AEAIRKAGAMQTVKLAQEFPELLAIEDTTSL DAKEKKALIGREIKSAVEKDGASATIDLWTD DCSPGIWQLDCTHL VEHLDAMQWINGDGYLH RIVTGDEKWI	124 WWVHSVLLLEAT..TFRTVGLLHQ.EWW..MRPDDPADADEKE...GKWL RNFLGVTLHYHENNELRD..L..LGLKSL..DFE...RS...TAENIY KVILLVAHVVA...SG.Y..IEA.EVI..PA...E...TGQETA PKTWFWDVK...TR.KI..LGW.RCD...VS...E...NIDSIR TMLCVWWD.Q...SG.V..IYY.ELKPGET...VNAARYQQQLI AAAT.SR..LRMGSMMSNV.IAVCDR...EAD...I...HAYLQDKL KKLKAIFSQFNVED..LSSI..KFVTDNR...G.AN..VV...KSLA.N.. YFLK.LA..GRW...P..VKTVHTDN...G.S..NFT...STTVK.AAC LSFMD.VV..TRYG.IPEDF.HITIDNTRGAA.NKWLTTGGAPNRYRFKVKEDDPKG.LFL NLNRA.LQ..RKRP.EYR.V.IFLHDNAP.SH.T..ARA...VRDTLE.TL. 223 AHNERFVVRSKHPRK...DVESGLY NIRINKLSKLALSLSIPA..SSASERTFSLAGN WAGIKQEFGMNKL...KKIIGQV L.MGAKMHWTSVVAGKGWGQAKPVERAFGVGGLEEVVDK NWEVLPAAAYSPDLAPSDY...HLFASMGHALAEQR	342 MEEPDLNLERMVVSILSFVAVRLQLRESFTP NR..IGQQTVDSSLFFLNSFYKNF... DQA.EHLKTAIQMAVFIHN... P..LFLKTMAEGVAMFN... DSY.ESVKKWADDEWFAAKDDEFYWRGIH..
--	--	---	--

图1A

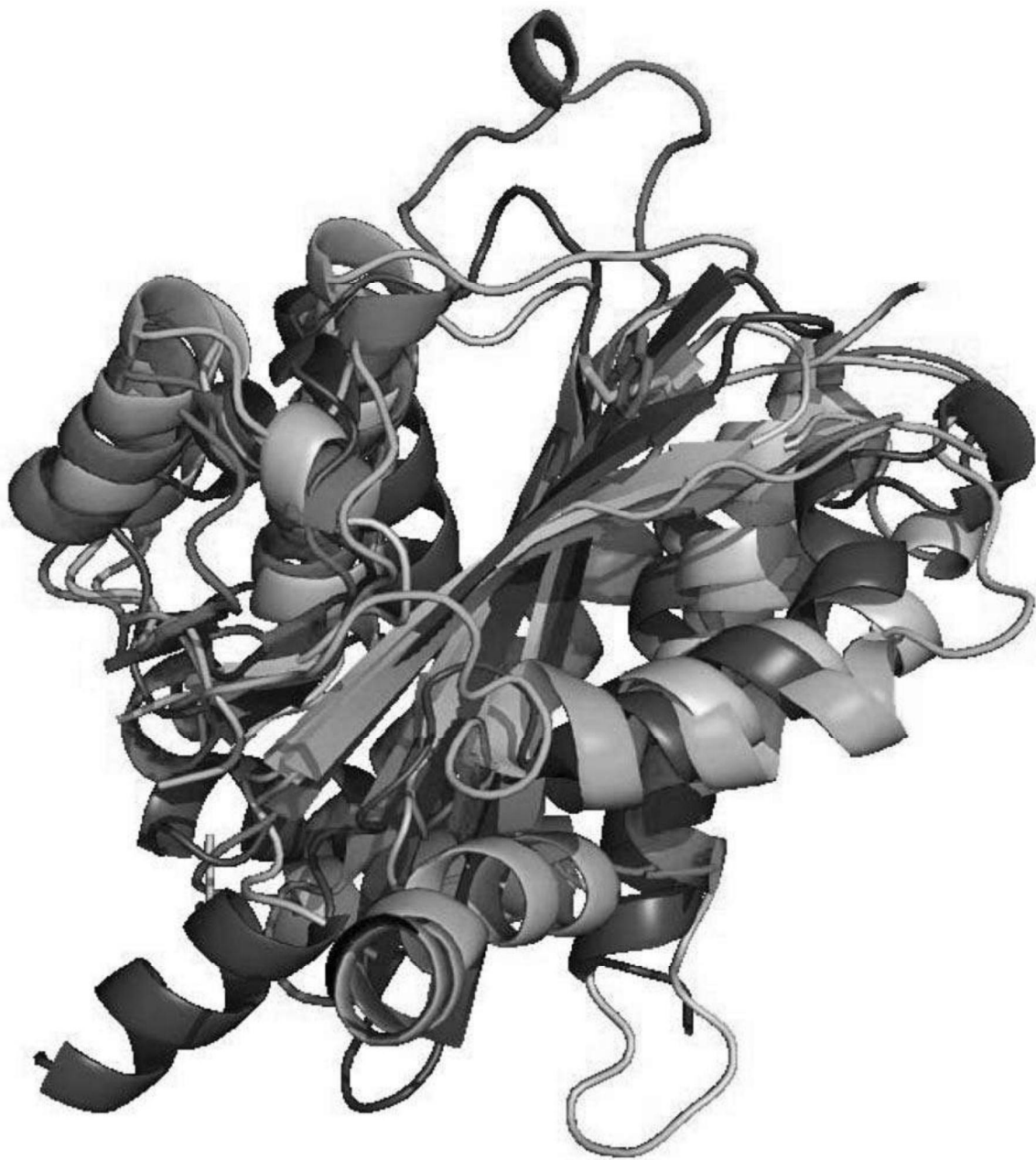


图1B

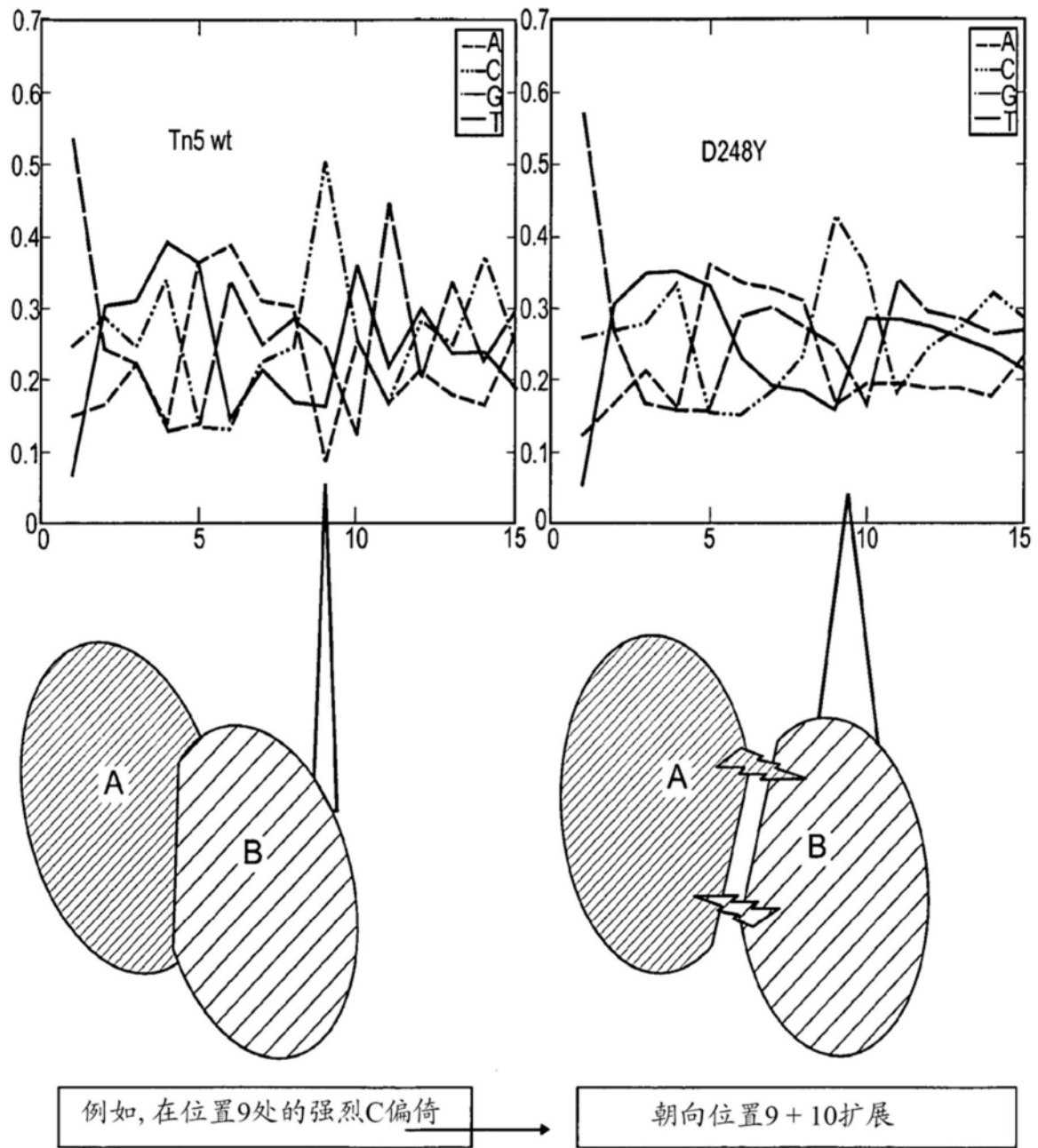
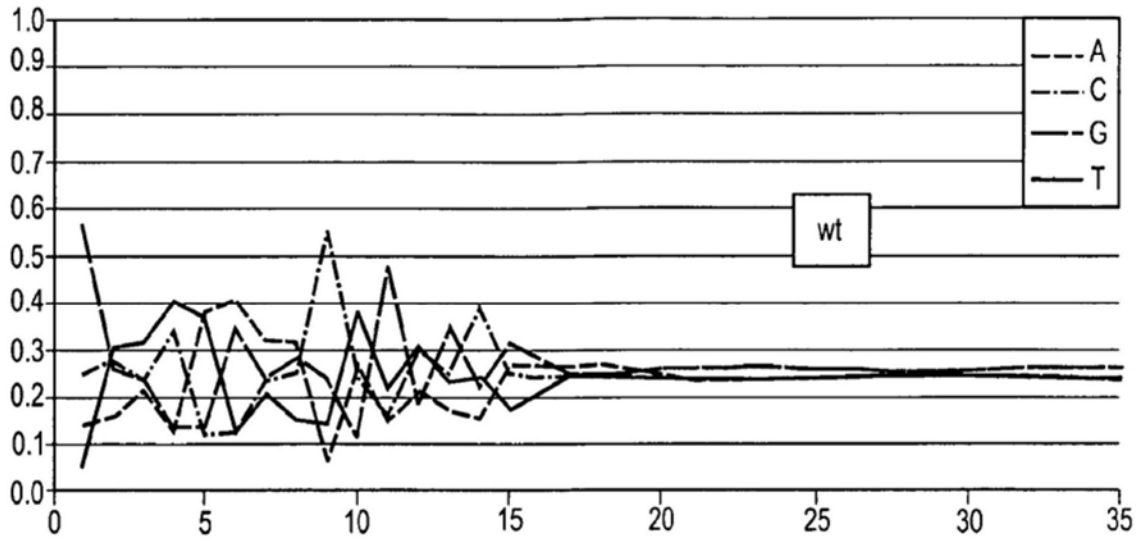


图2

对于384-10-A: A1的IVC图



对于96-4-SSM: B2的IVC图

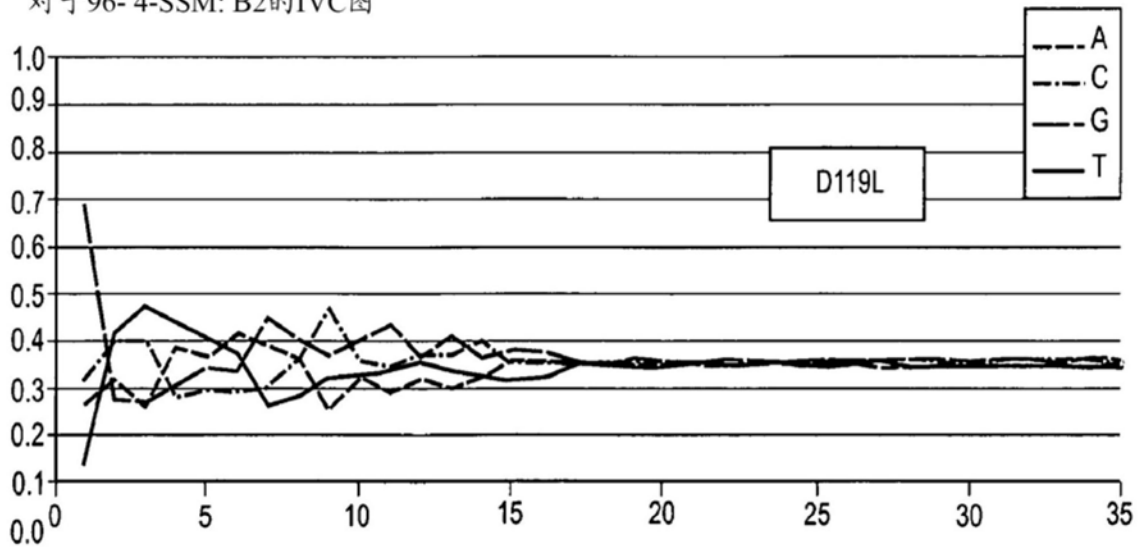


图3

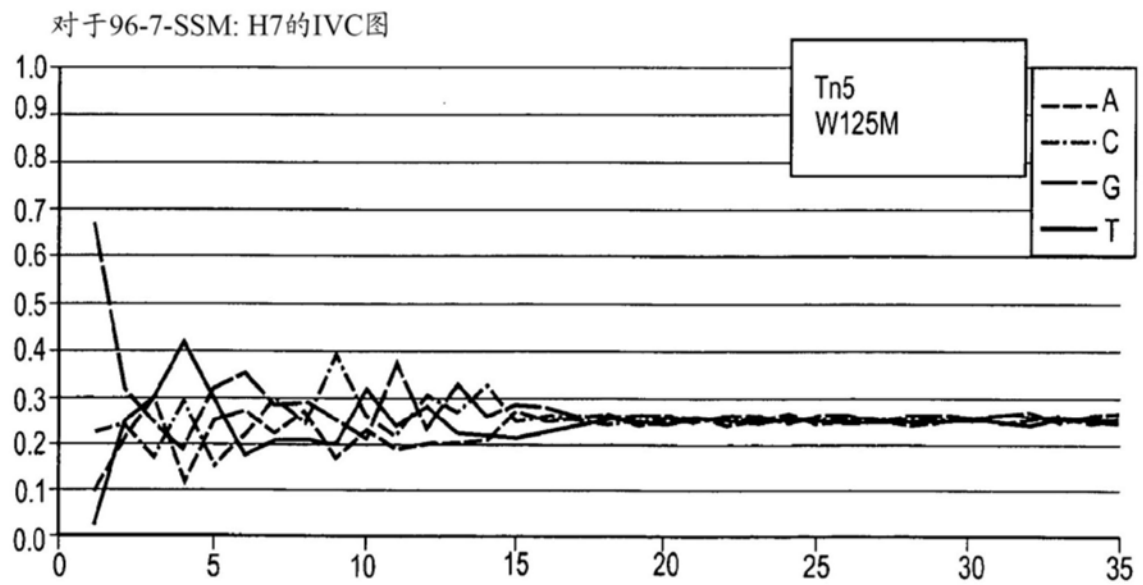
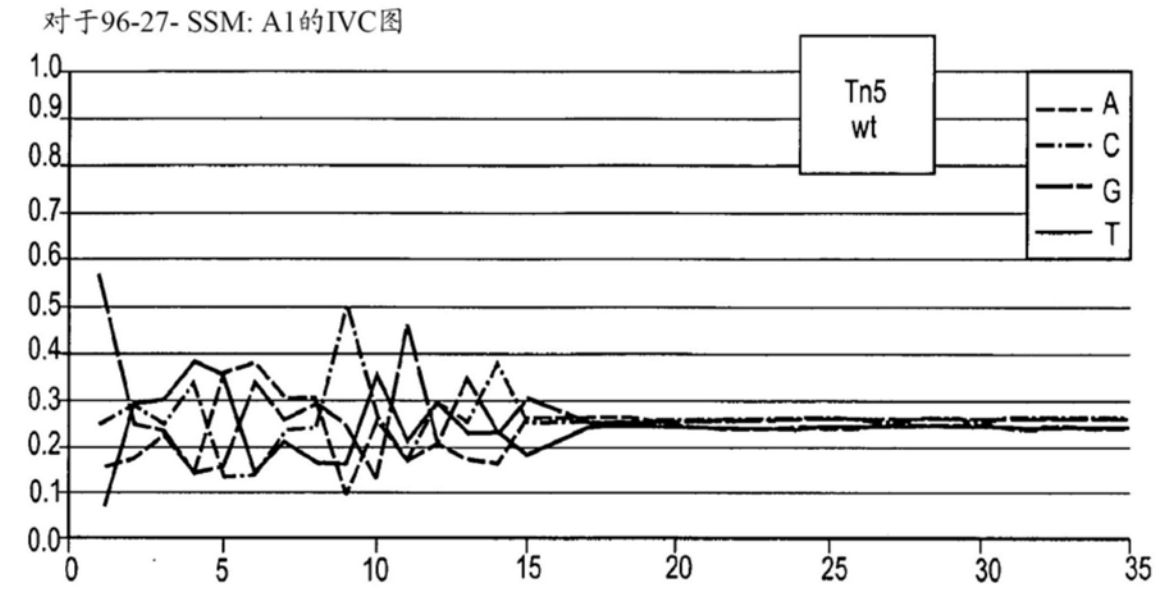


图4

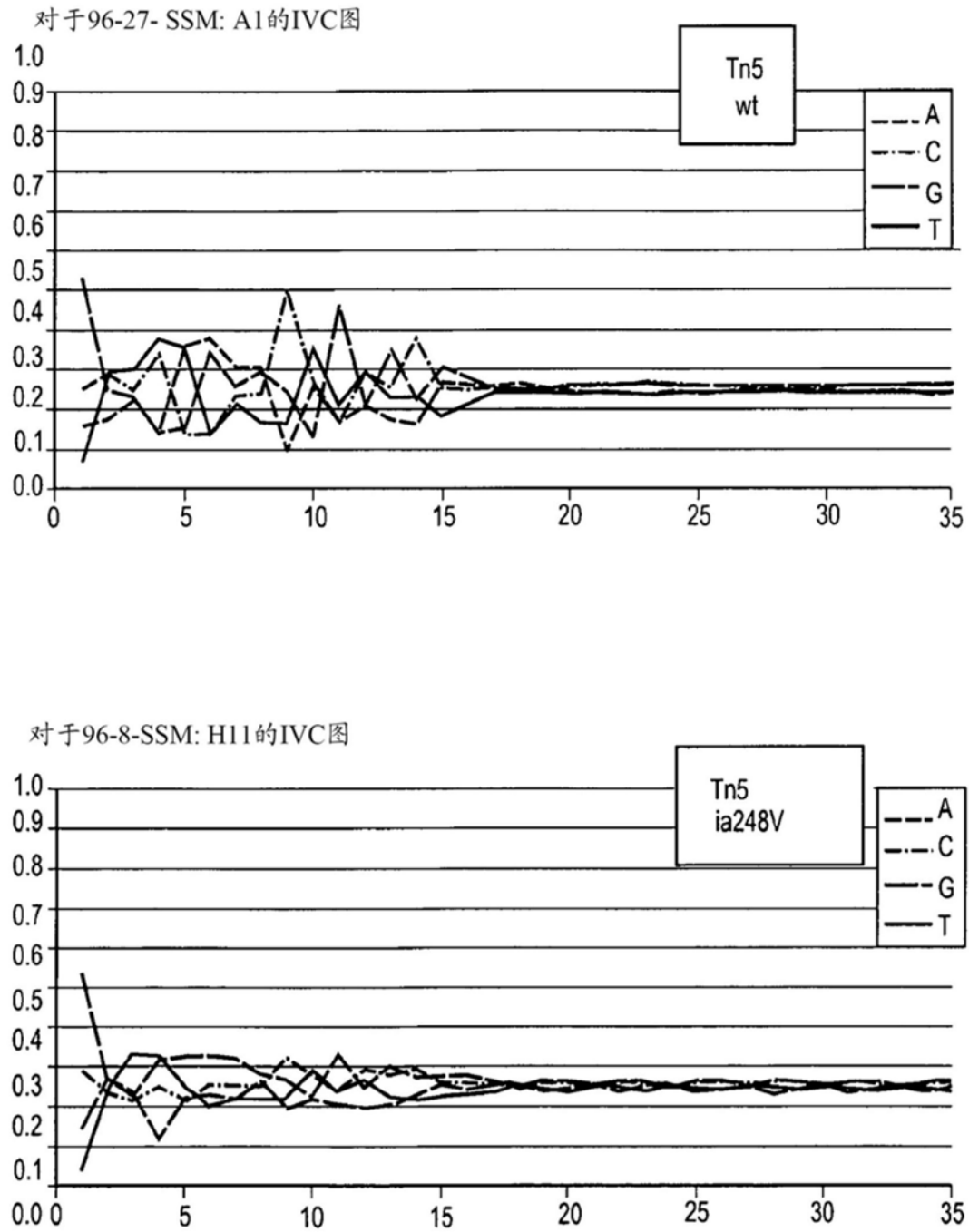
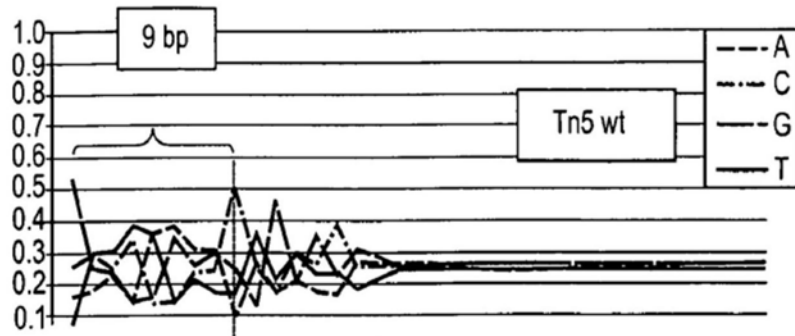
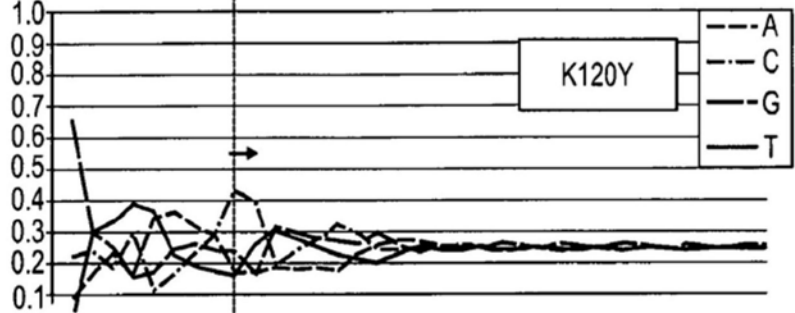


图5

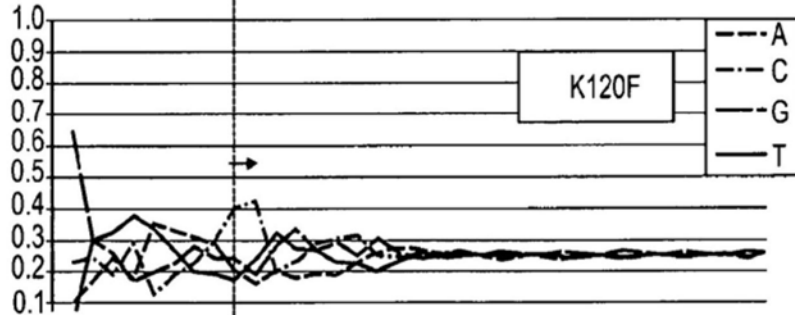
对于96-27-SSM: A1的IVC图



对于96-27-SSM: G11的IVC图



对于96-5-SSM: C5的IVC图



对于96-5-SSM: B3的IVC图

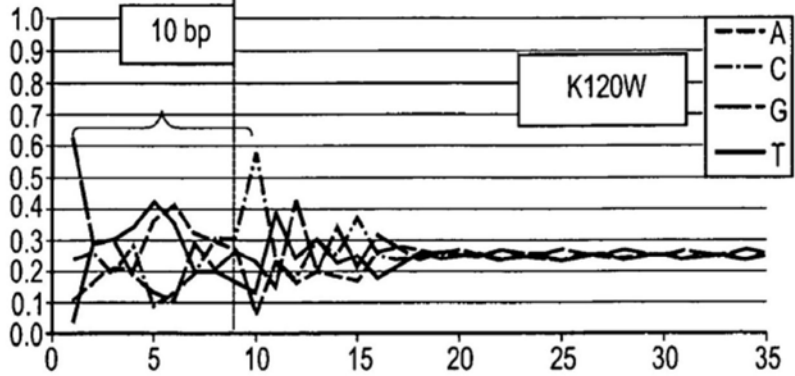


图6

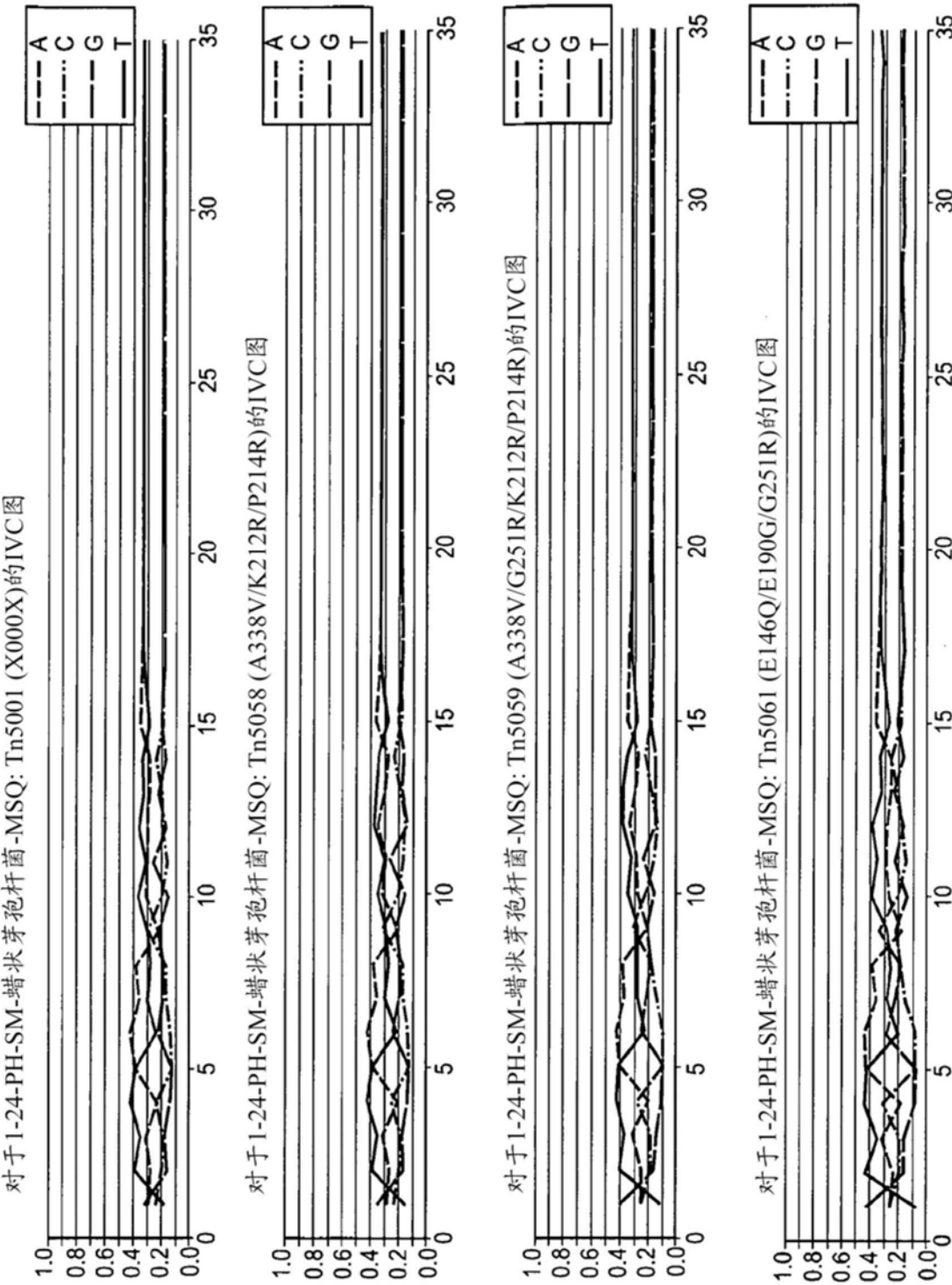


图7

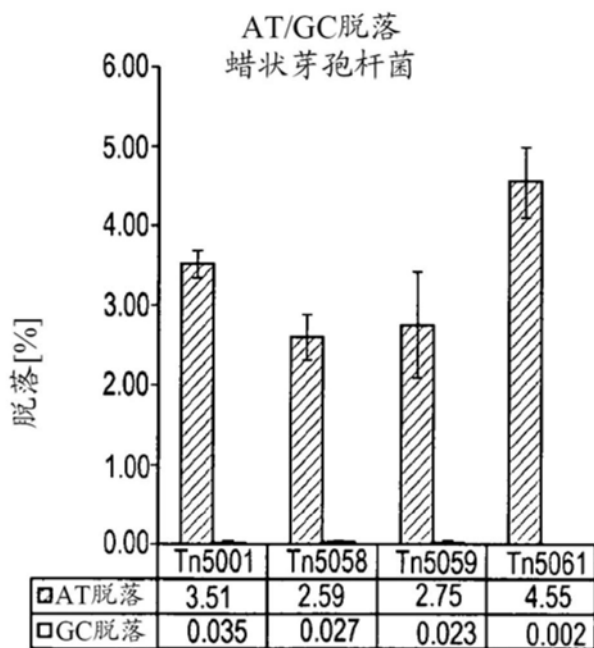


图8A

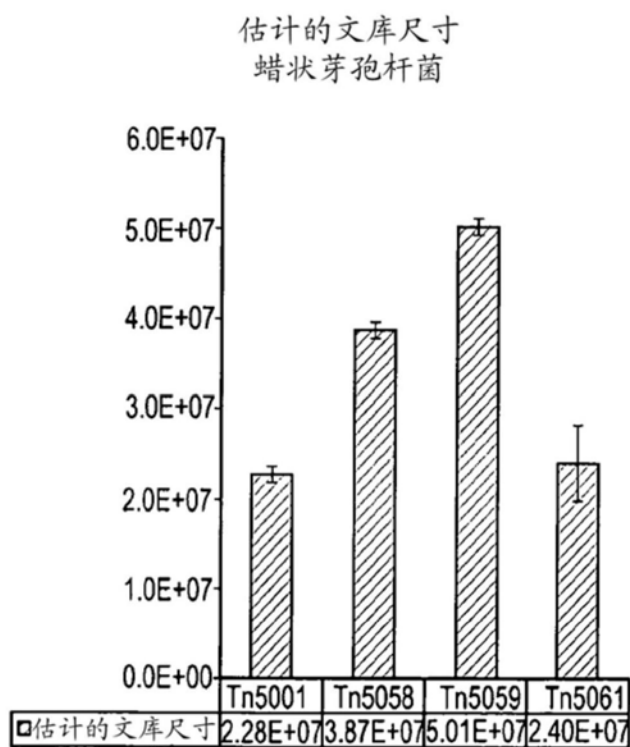


图8B

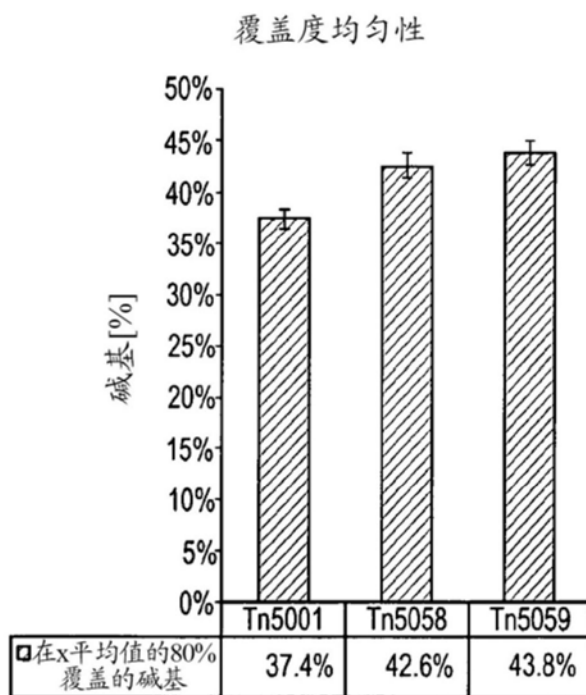


图9A

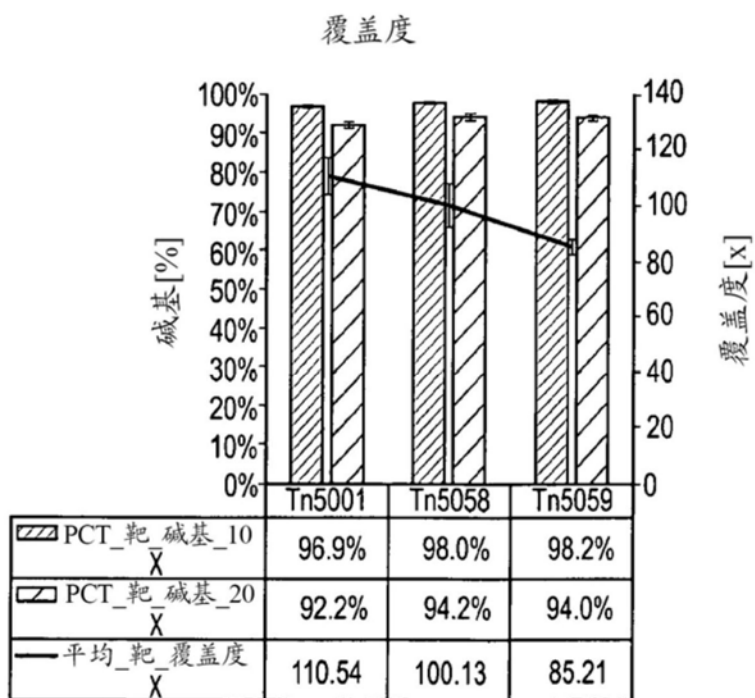


图9B

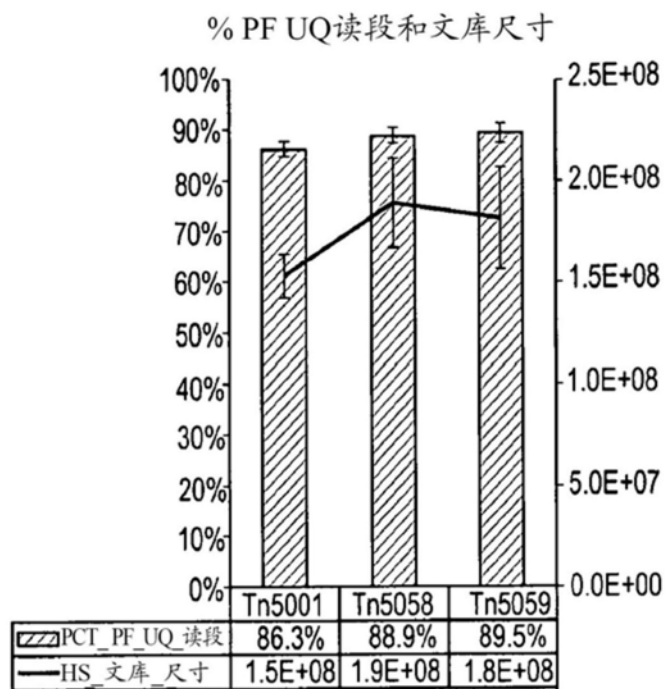


图10A

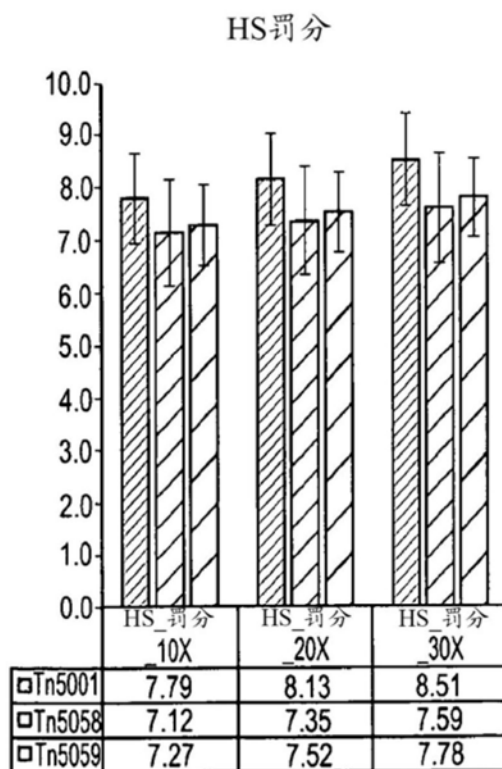


图10B

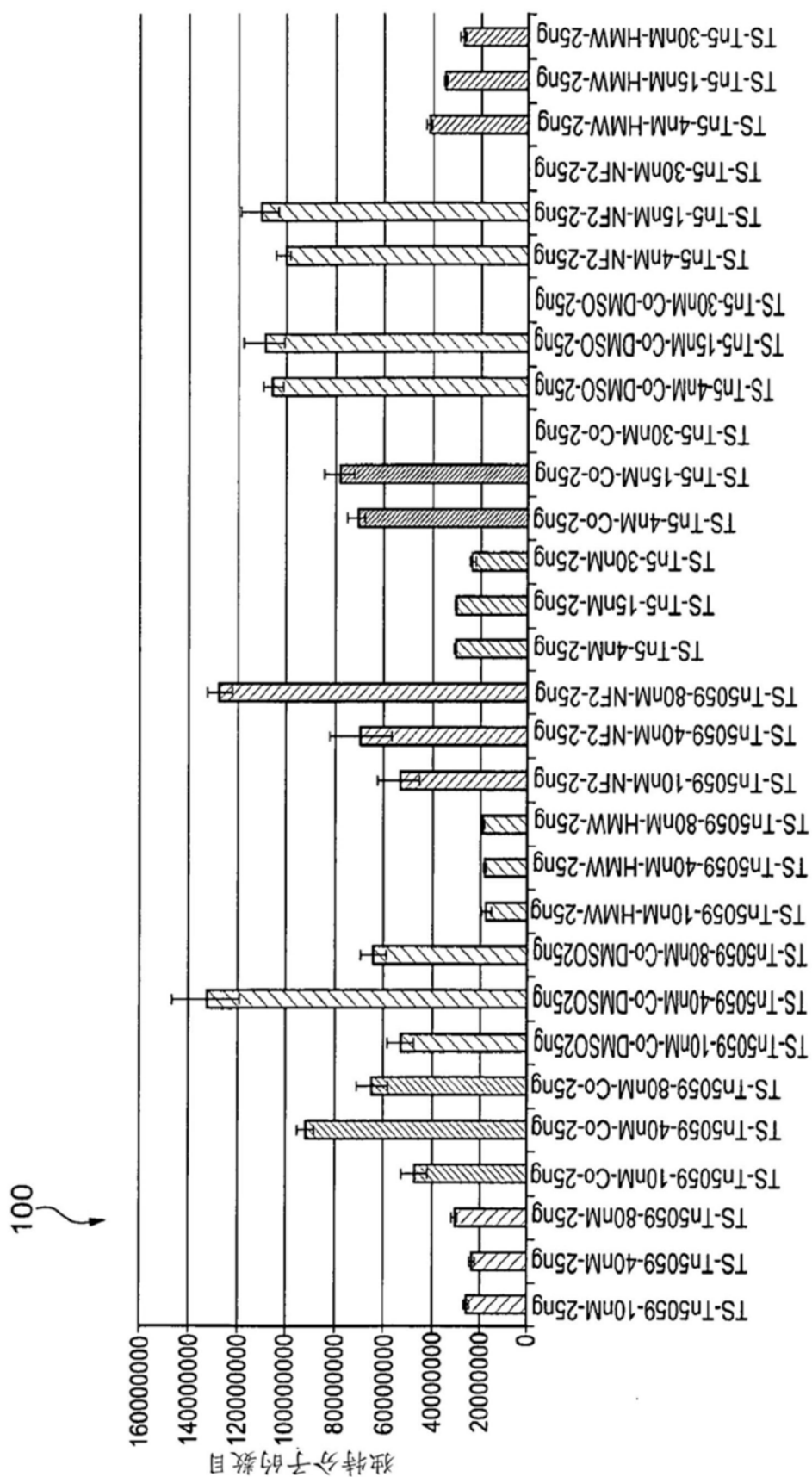


图11

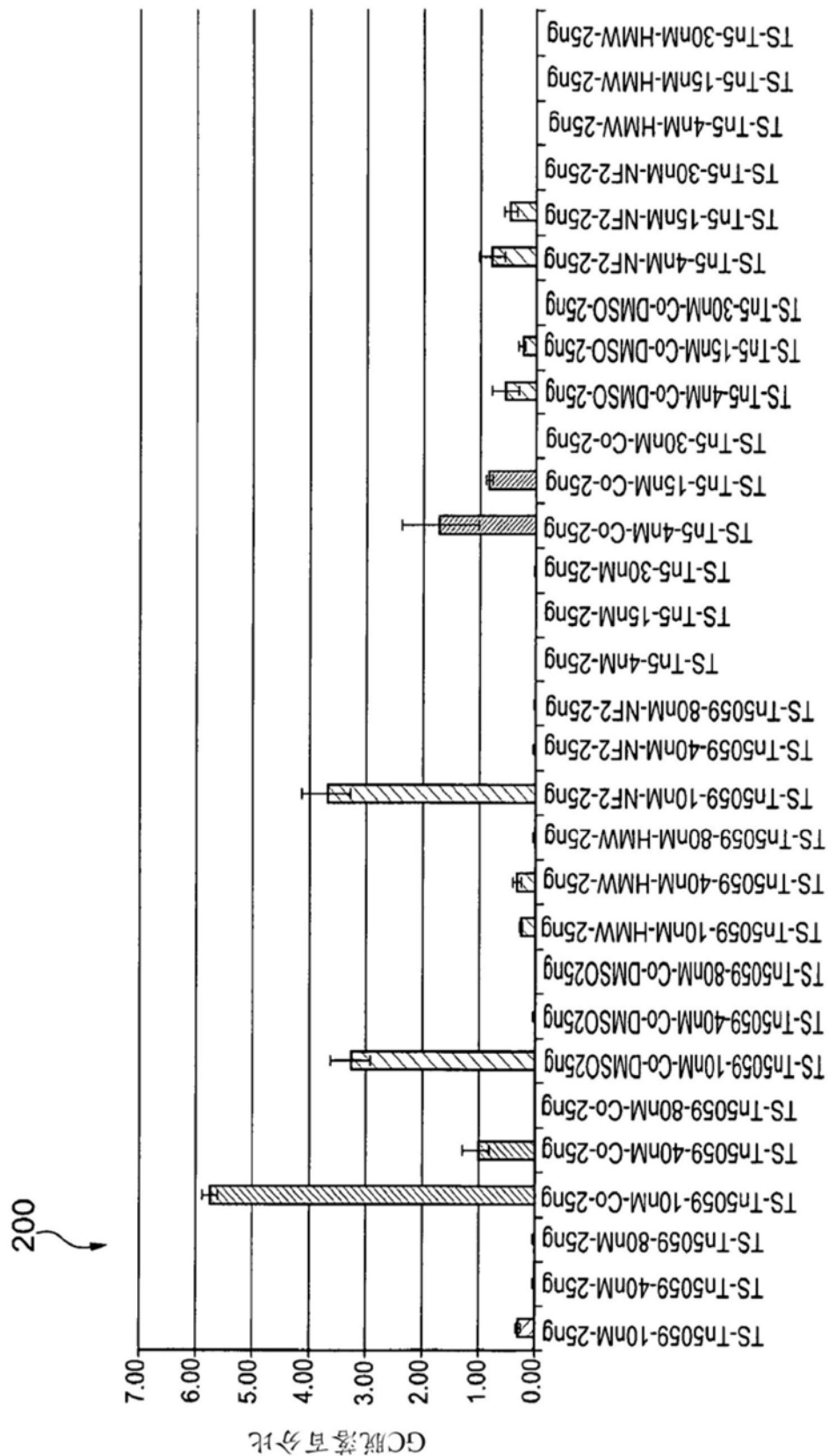


图12

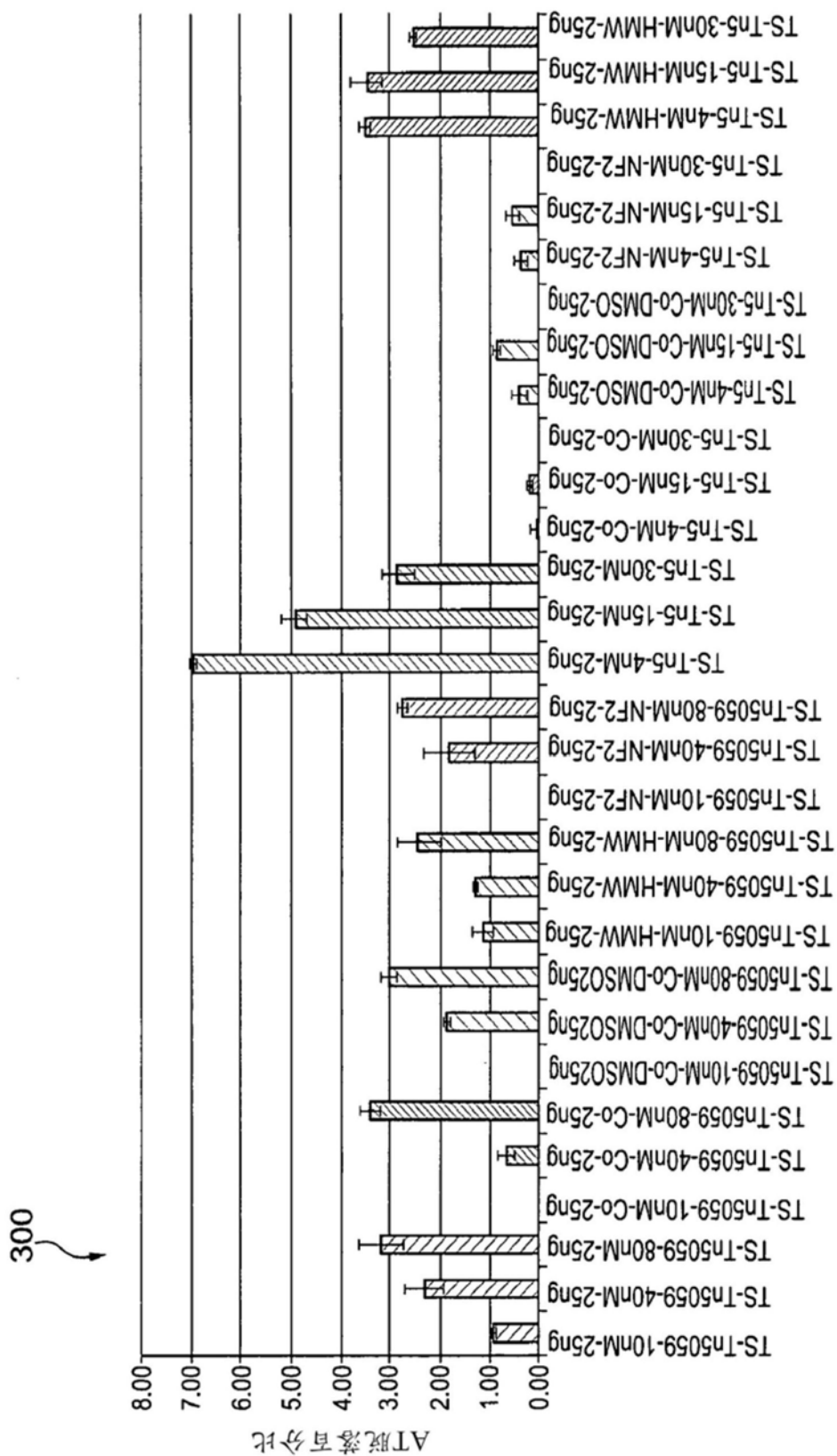


图13

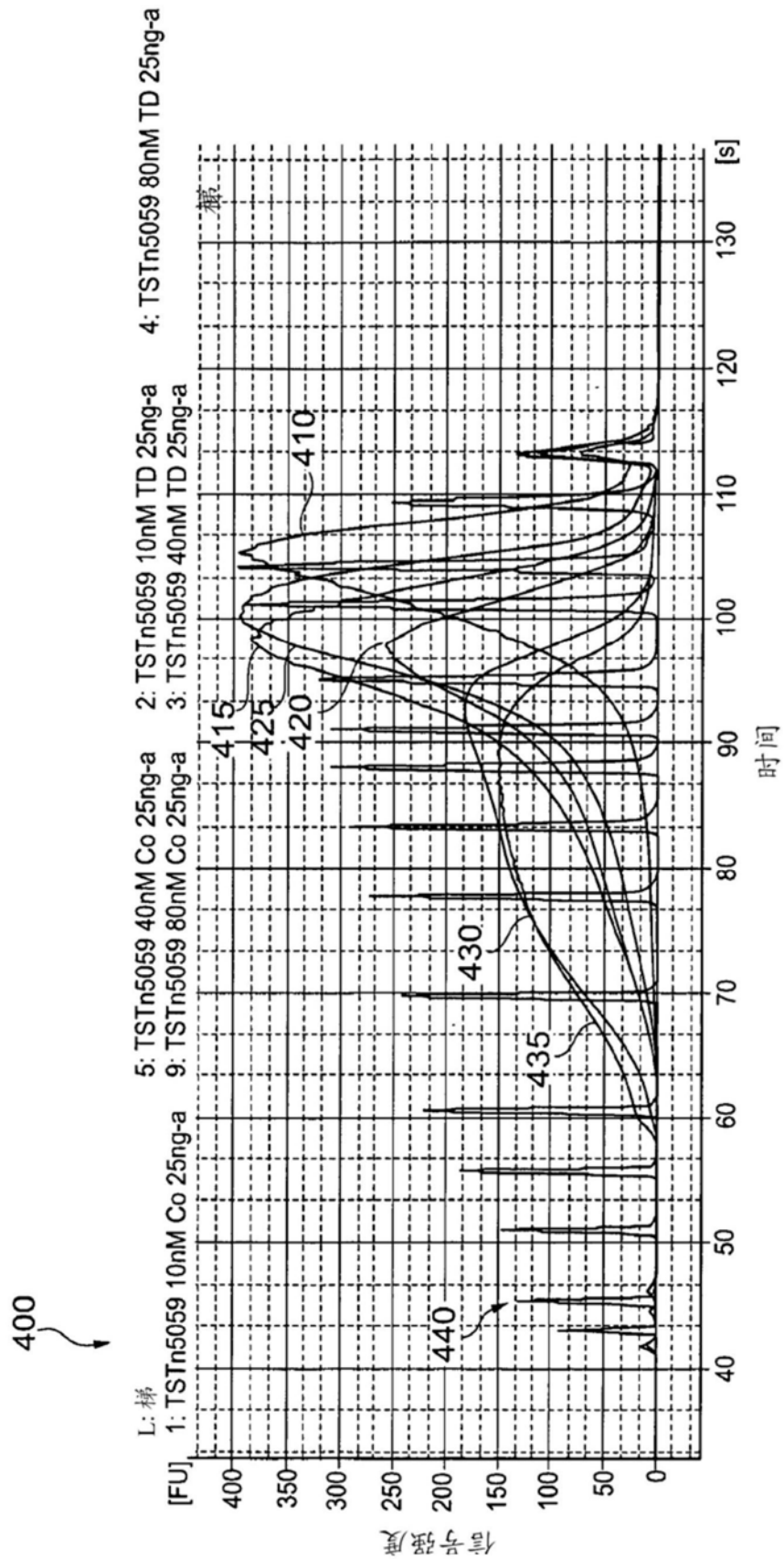


图14

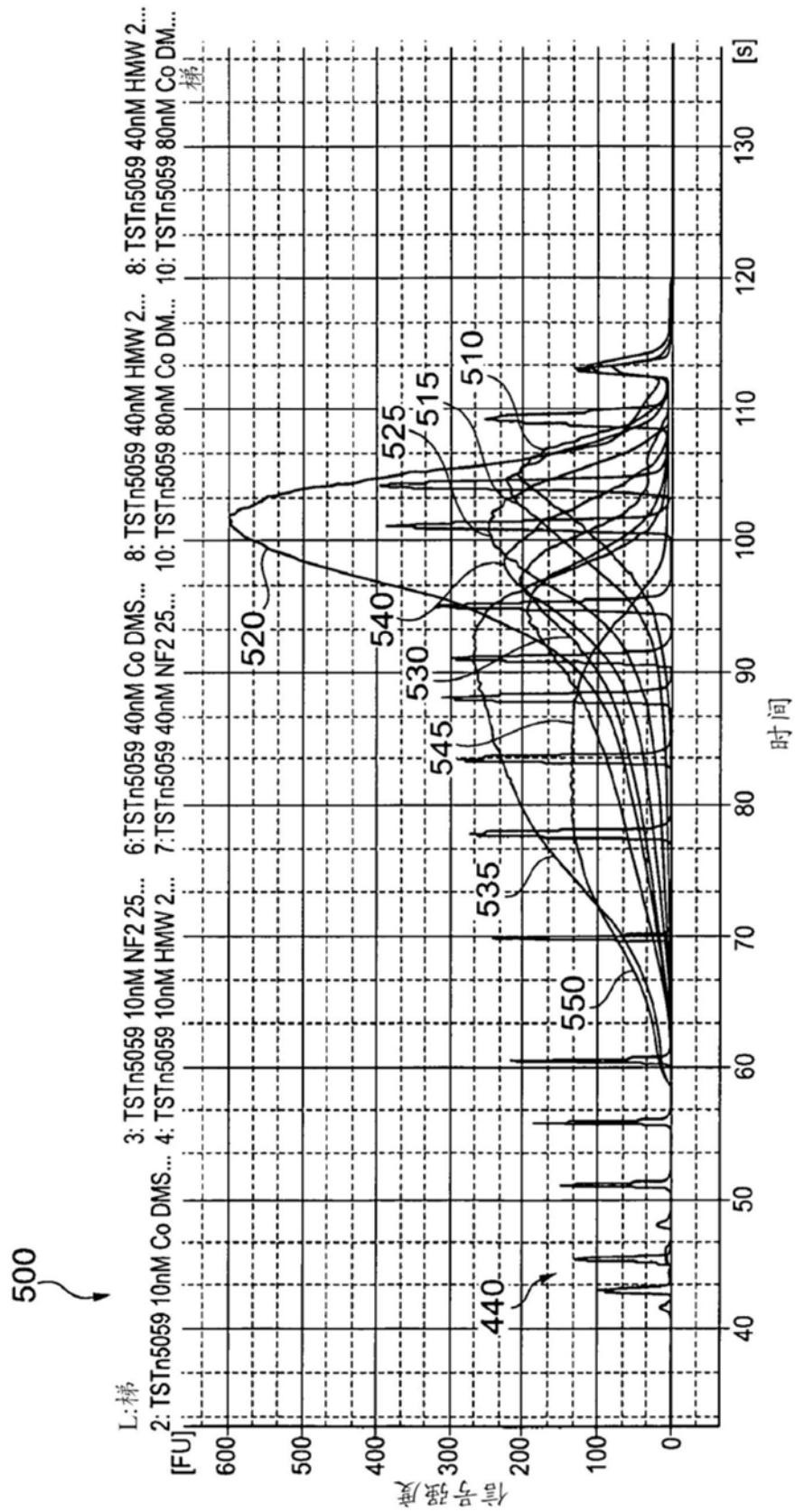


图15

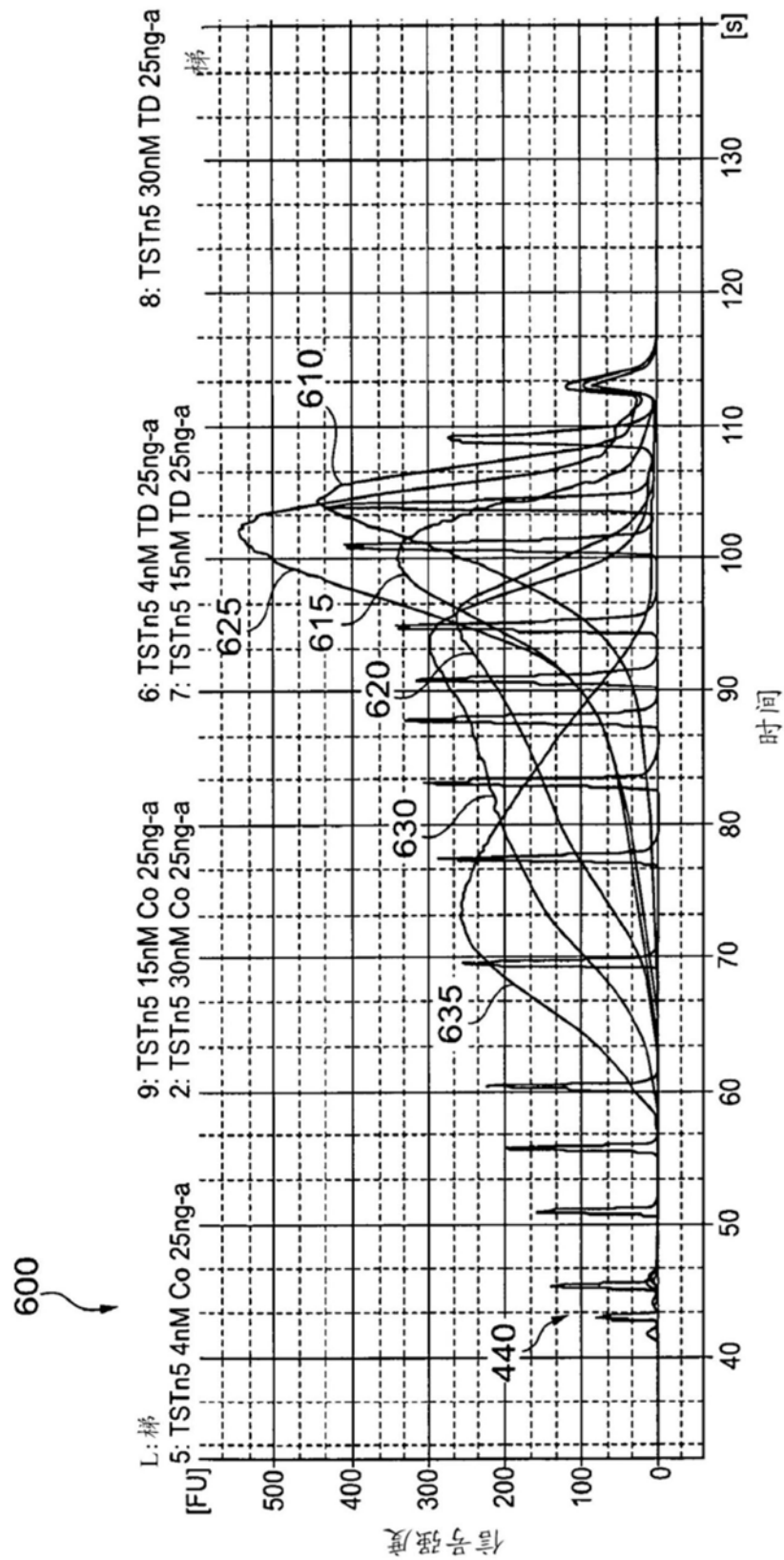


图16

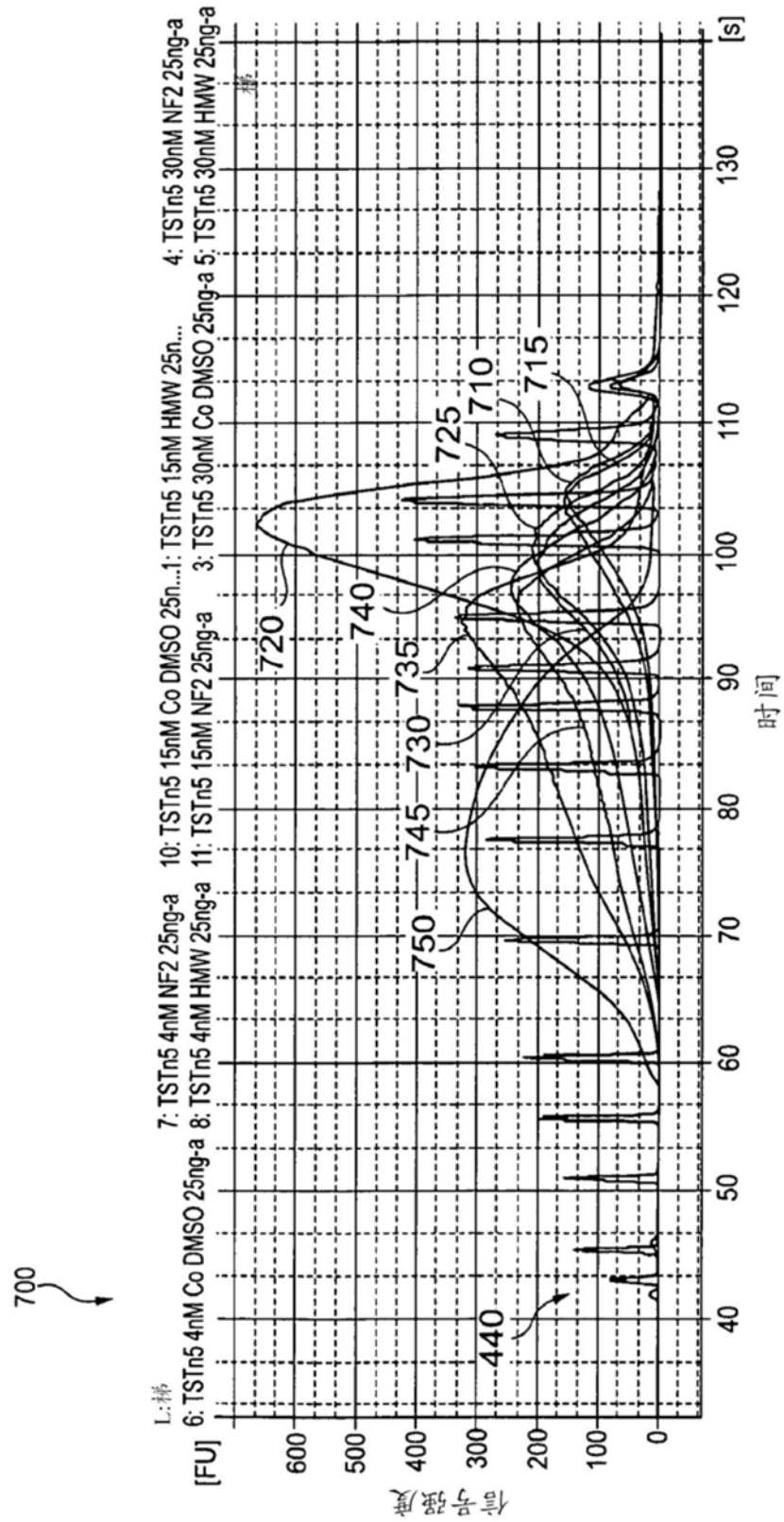


图17

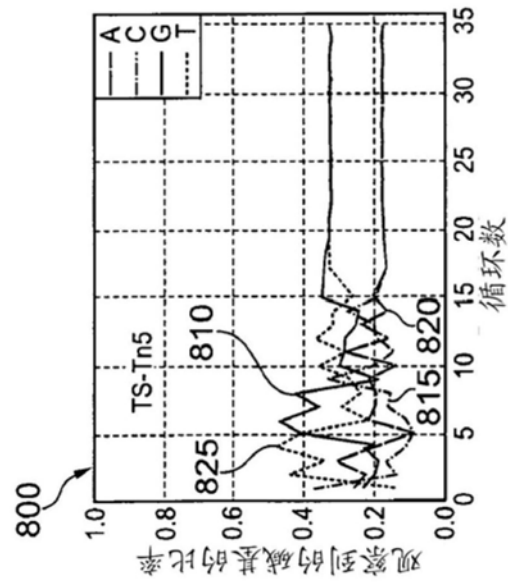


图18A

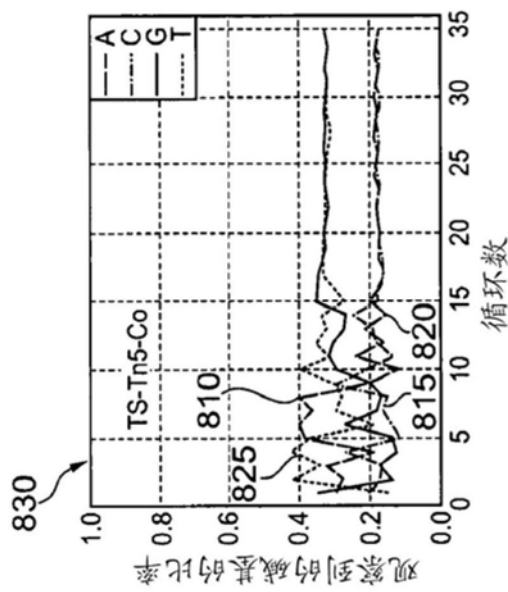


图18B

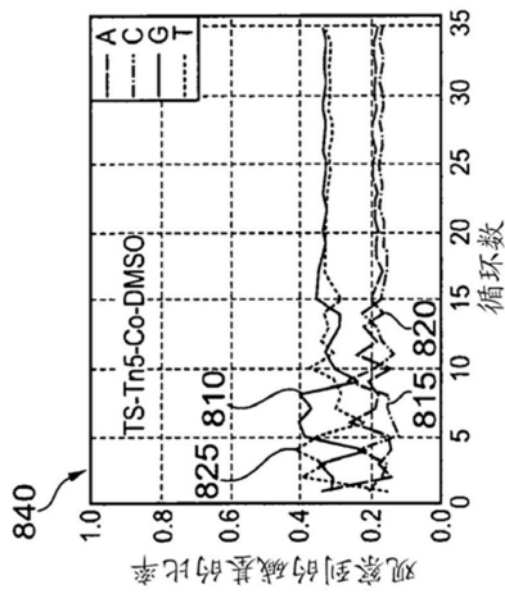


图18C

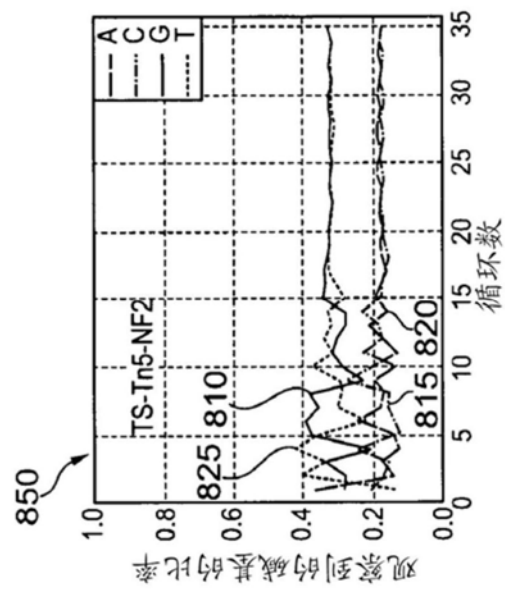


图18D

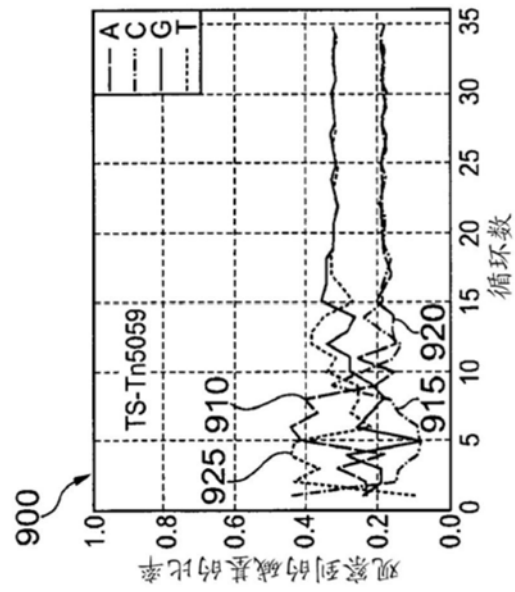


图19A

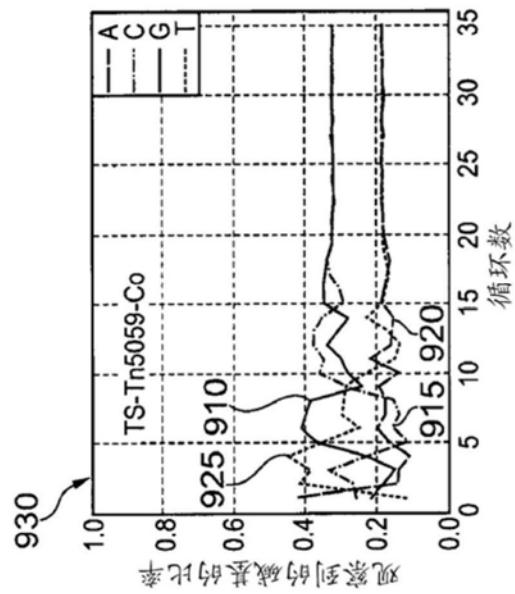


图19B

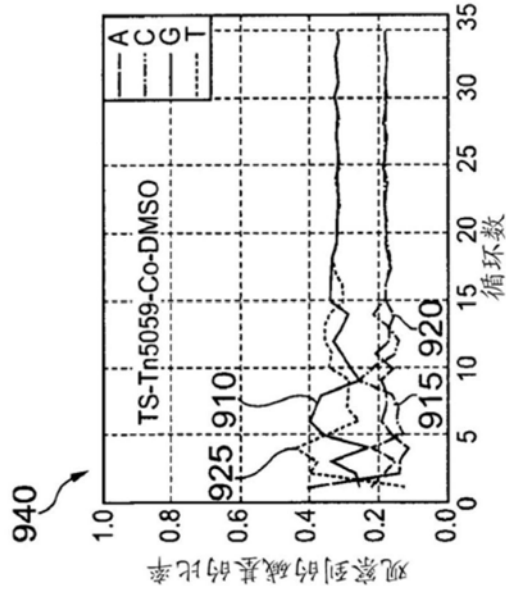


图19C

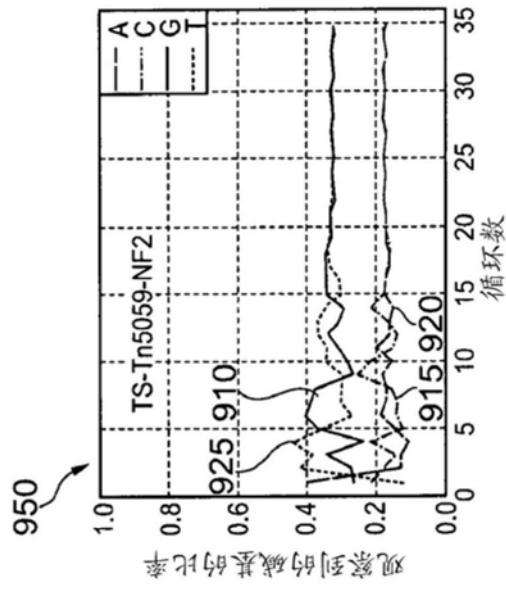


图19D

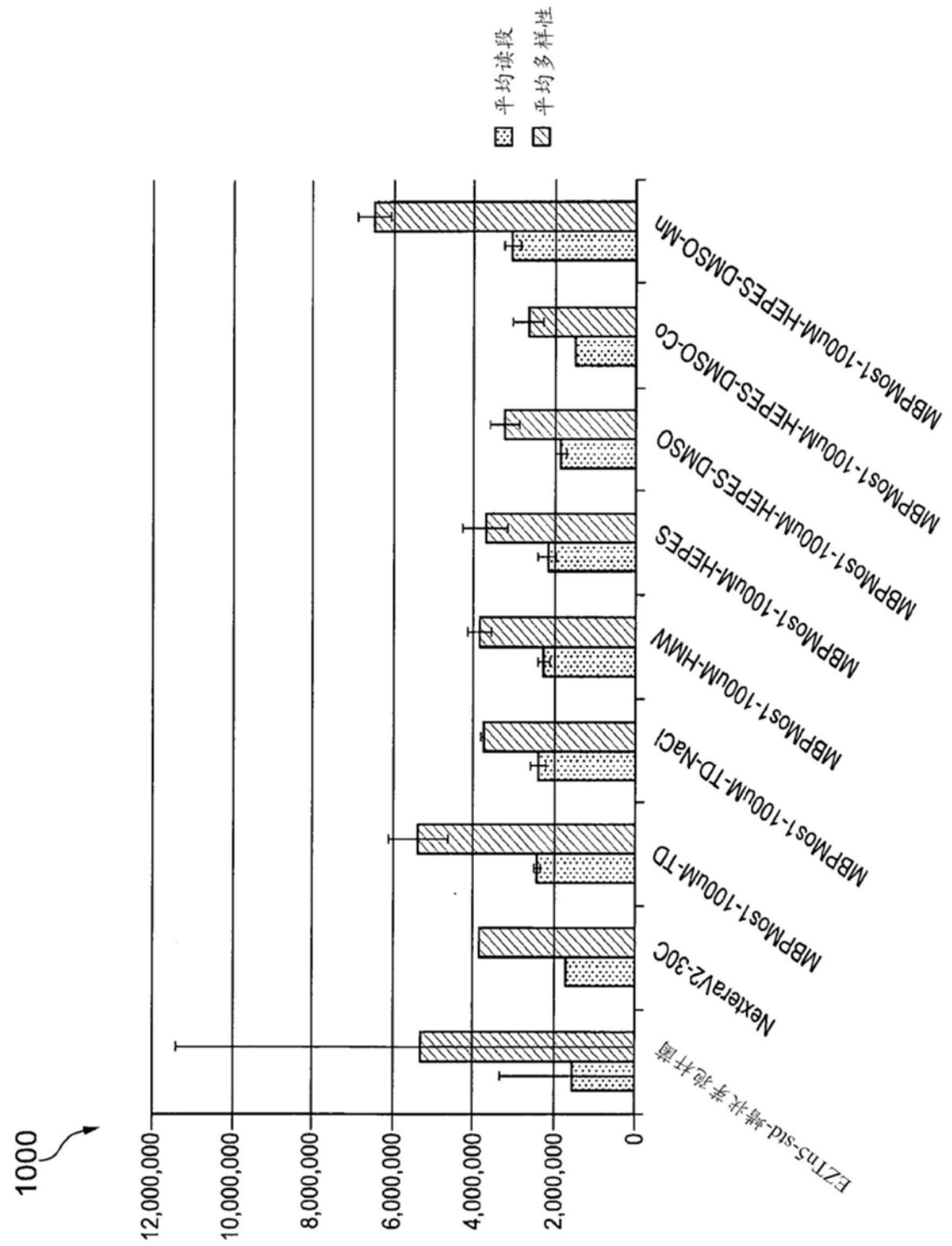


图20

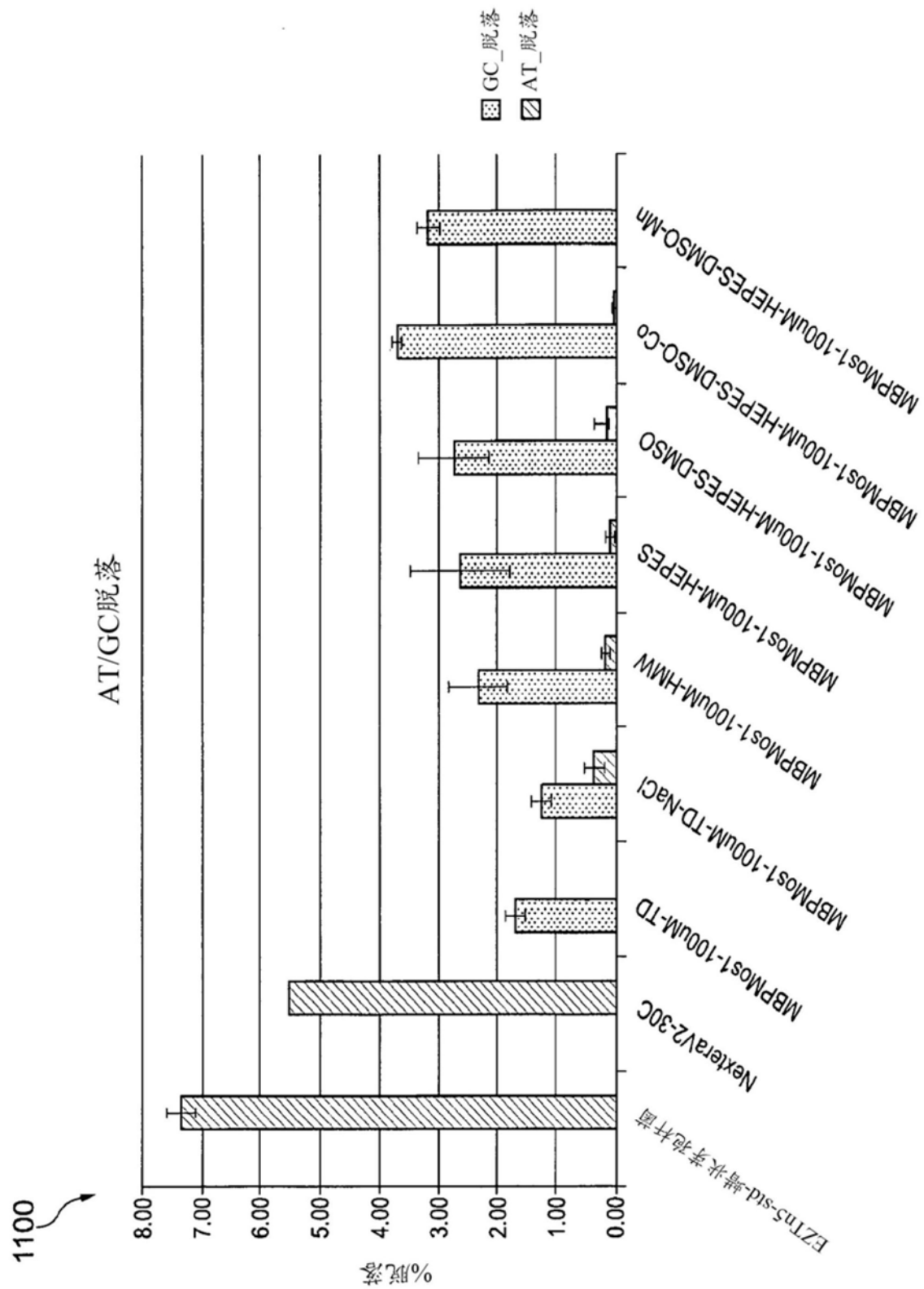


图21

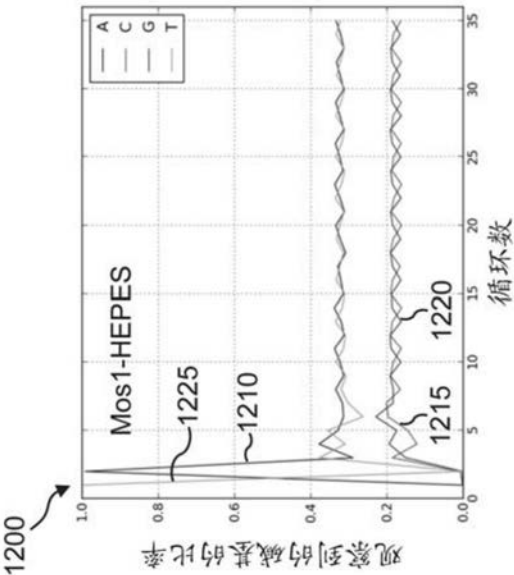


图22A

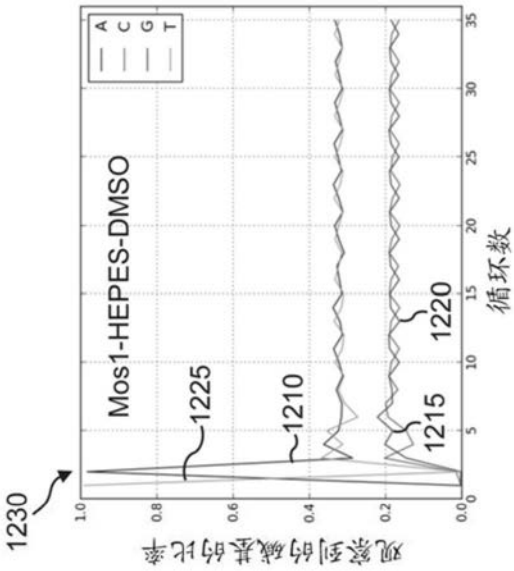


图22B

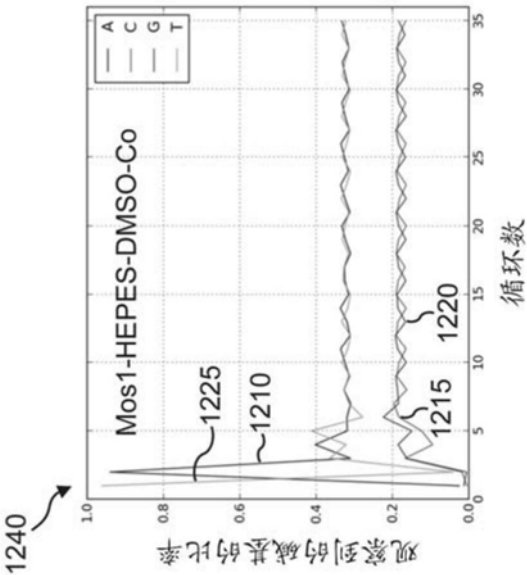


图22C

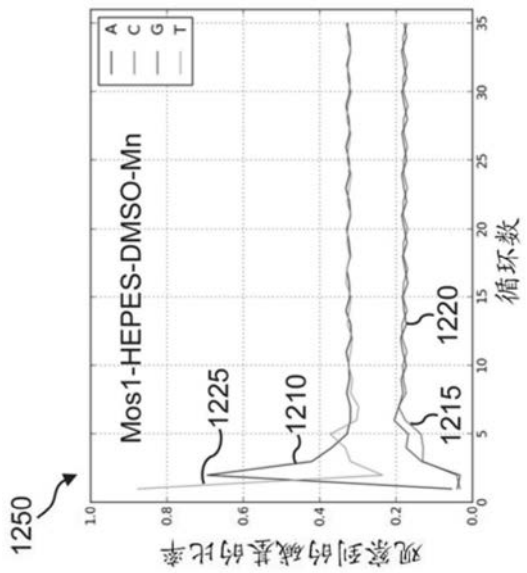


图22D

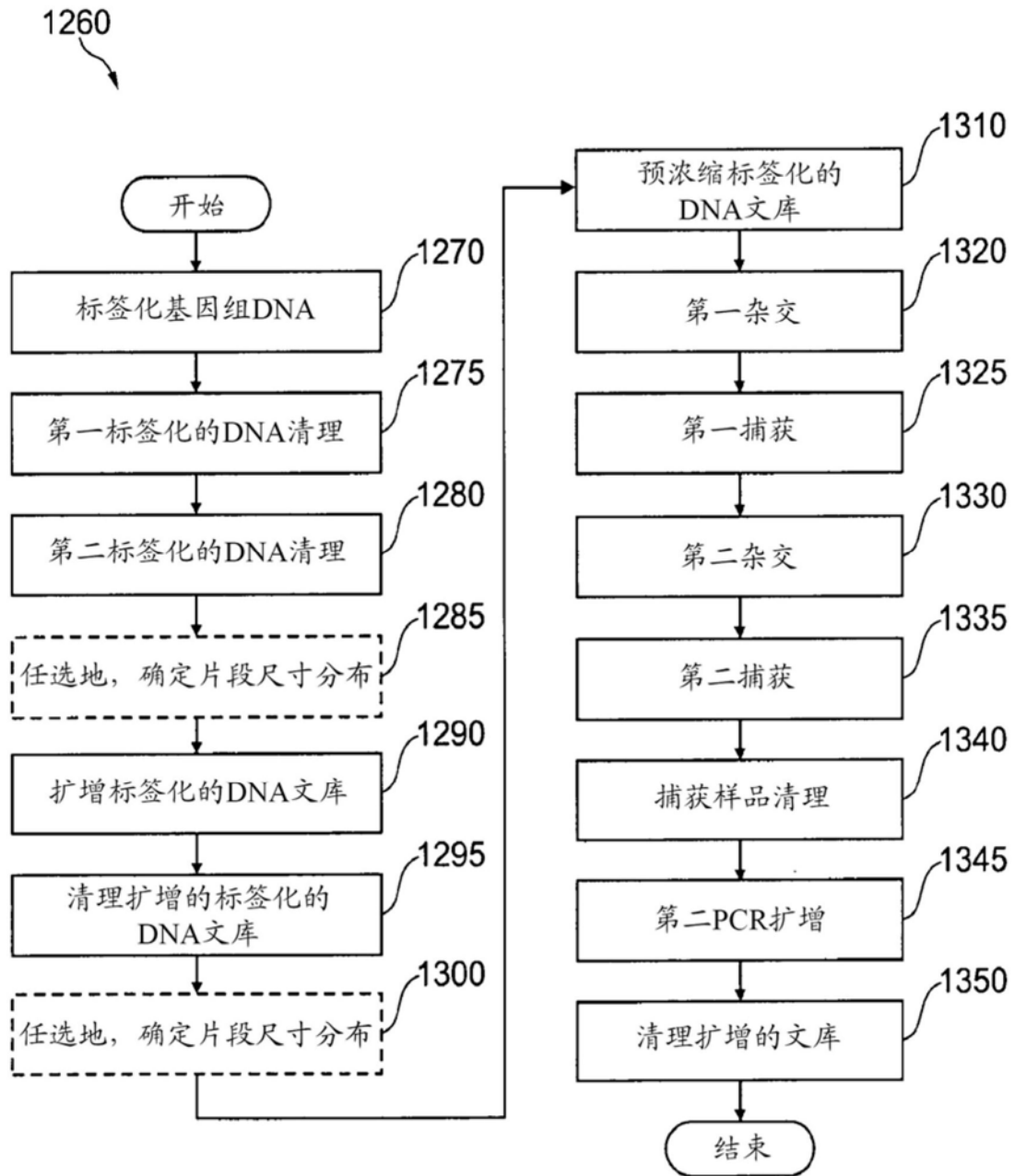


图23

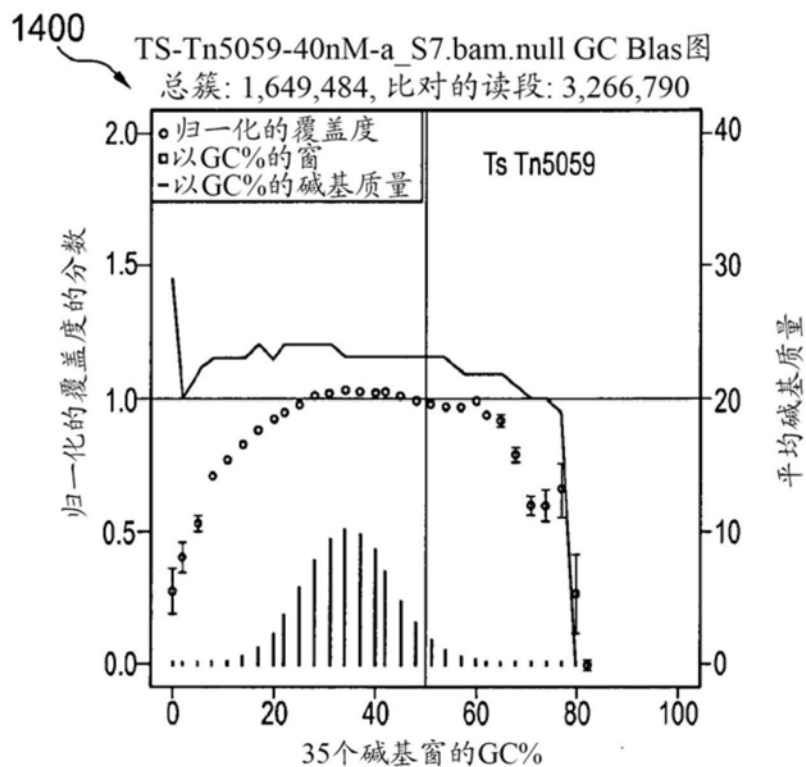


图24A

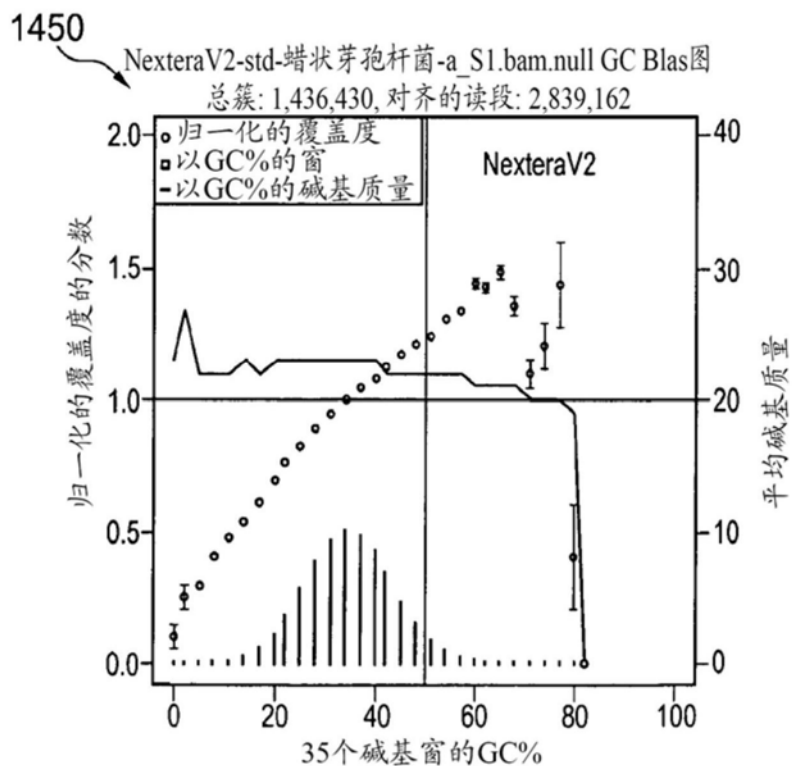


图24B

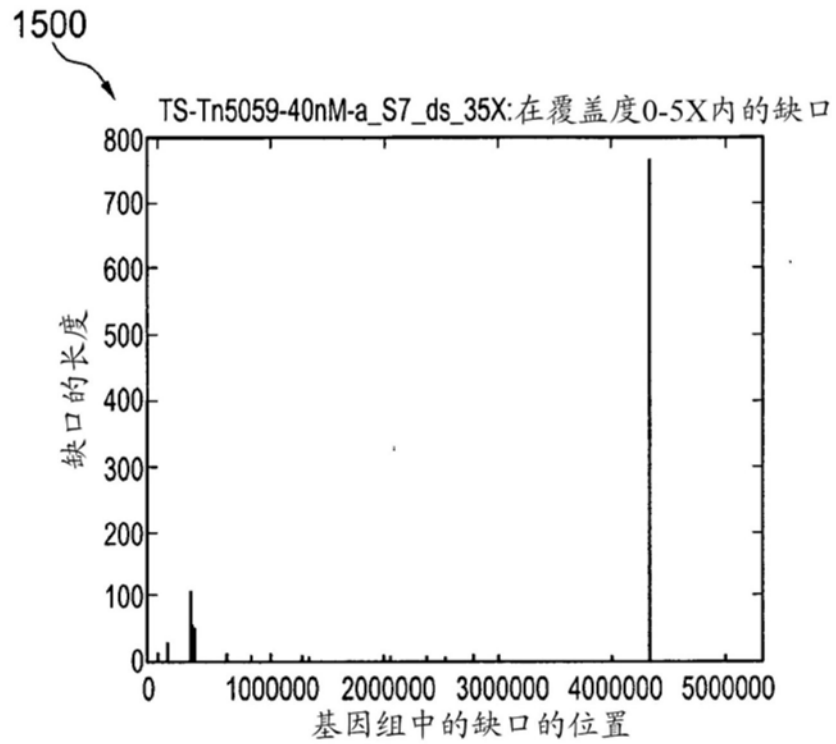


图25A

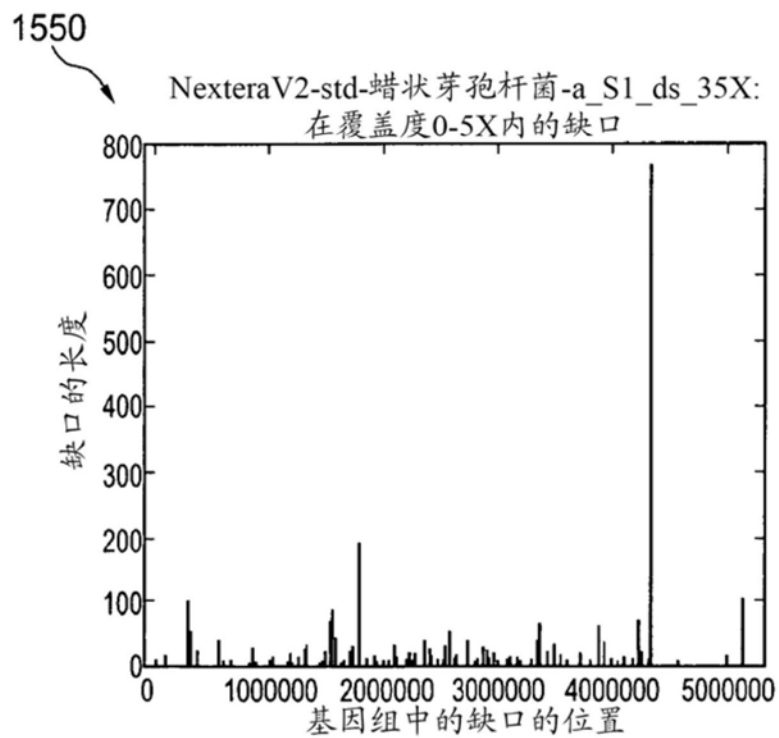


图25B

对25 ng人gDNA输入归一化为40nM的TS-Tn5059的
TDE1 (Tn5版本-1)和TS-Tn5的活性

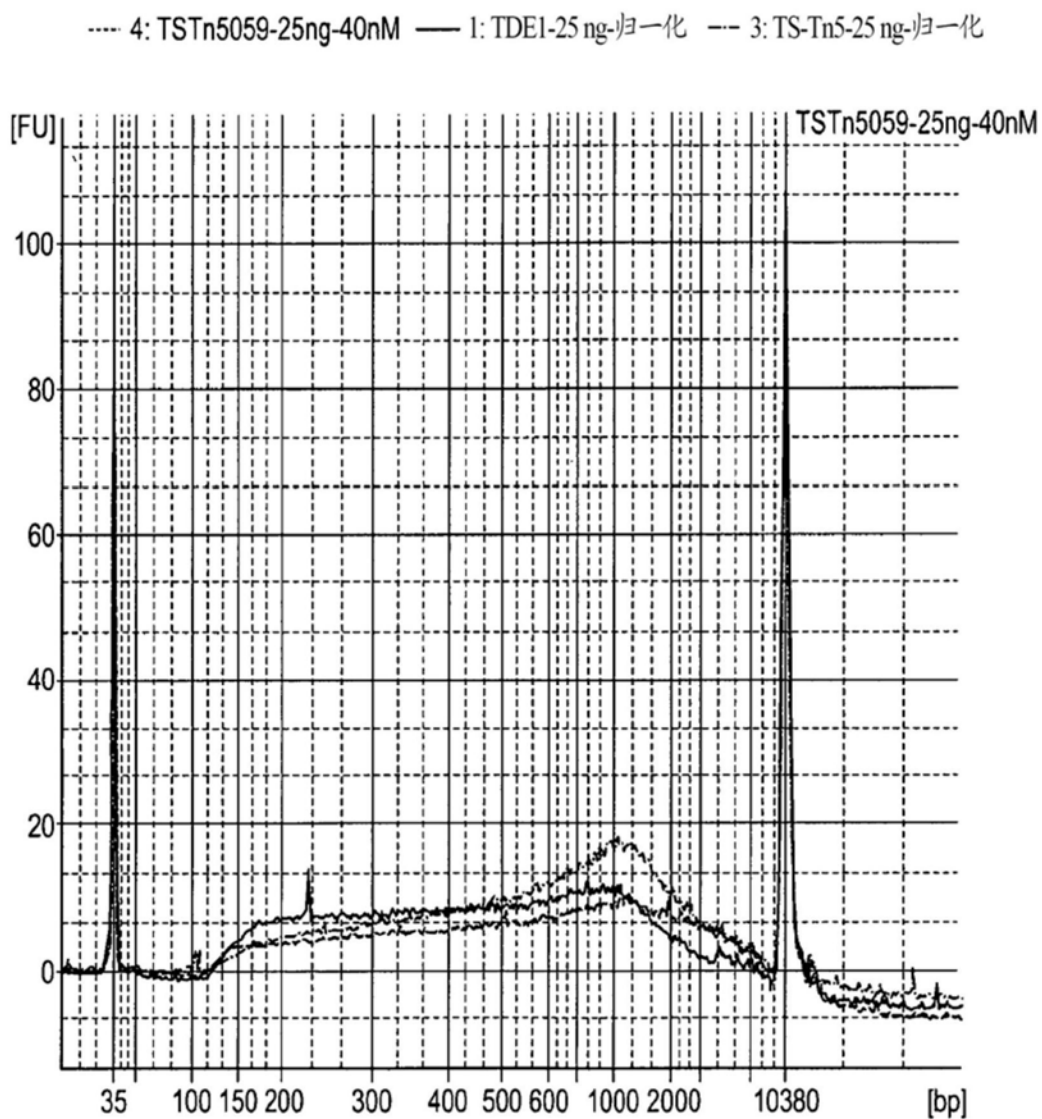


图26

片段尺寸分布的分析

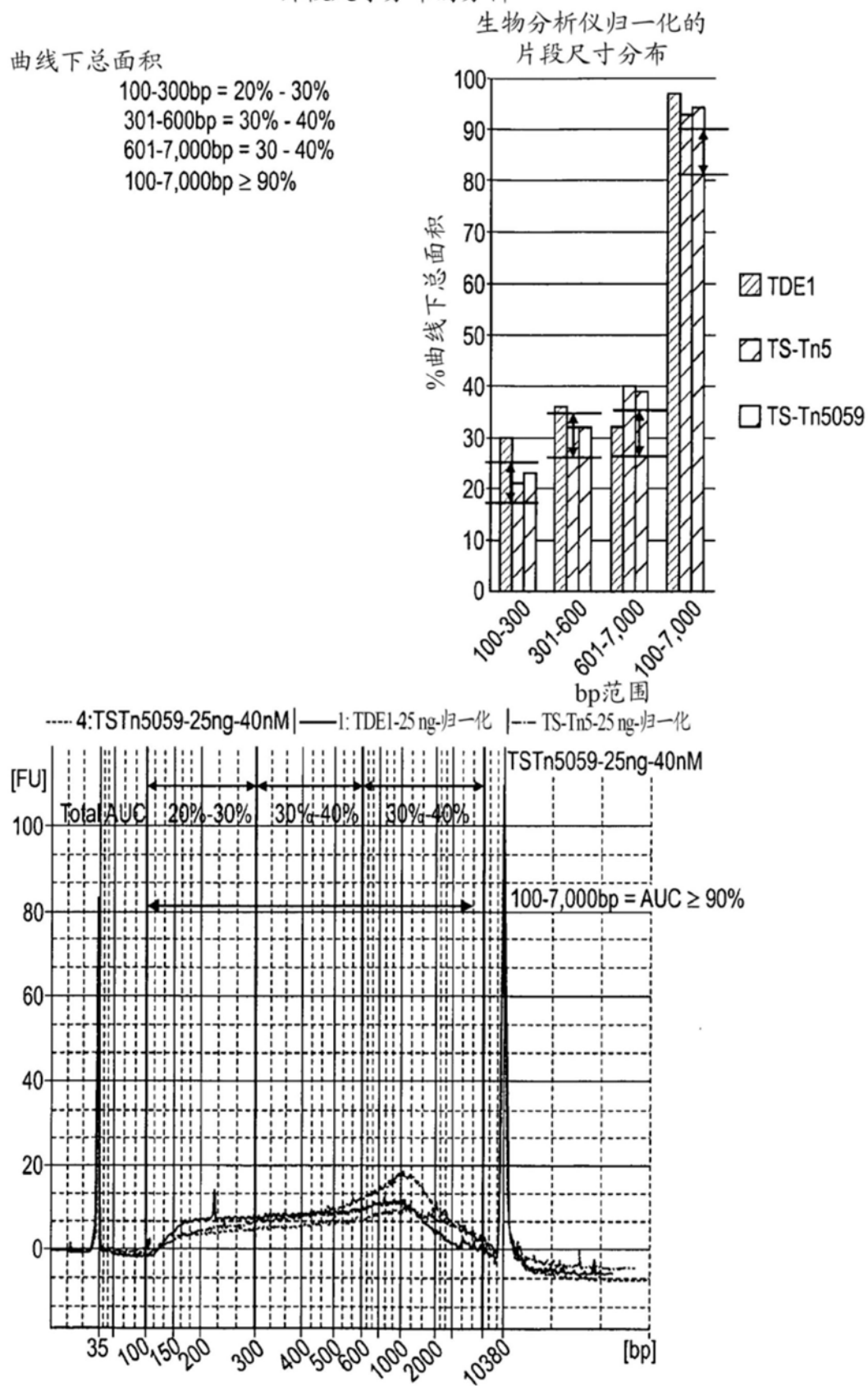


图27

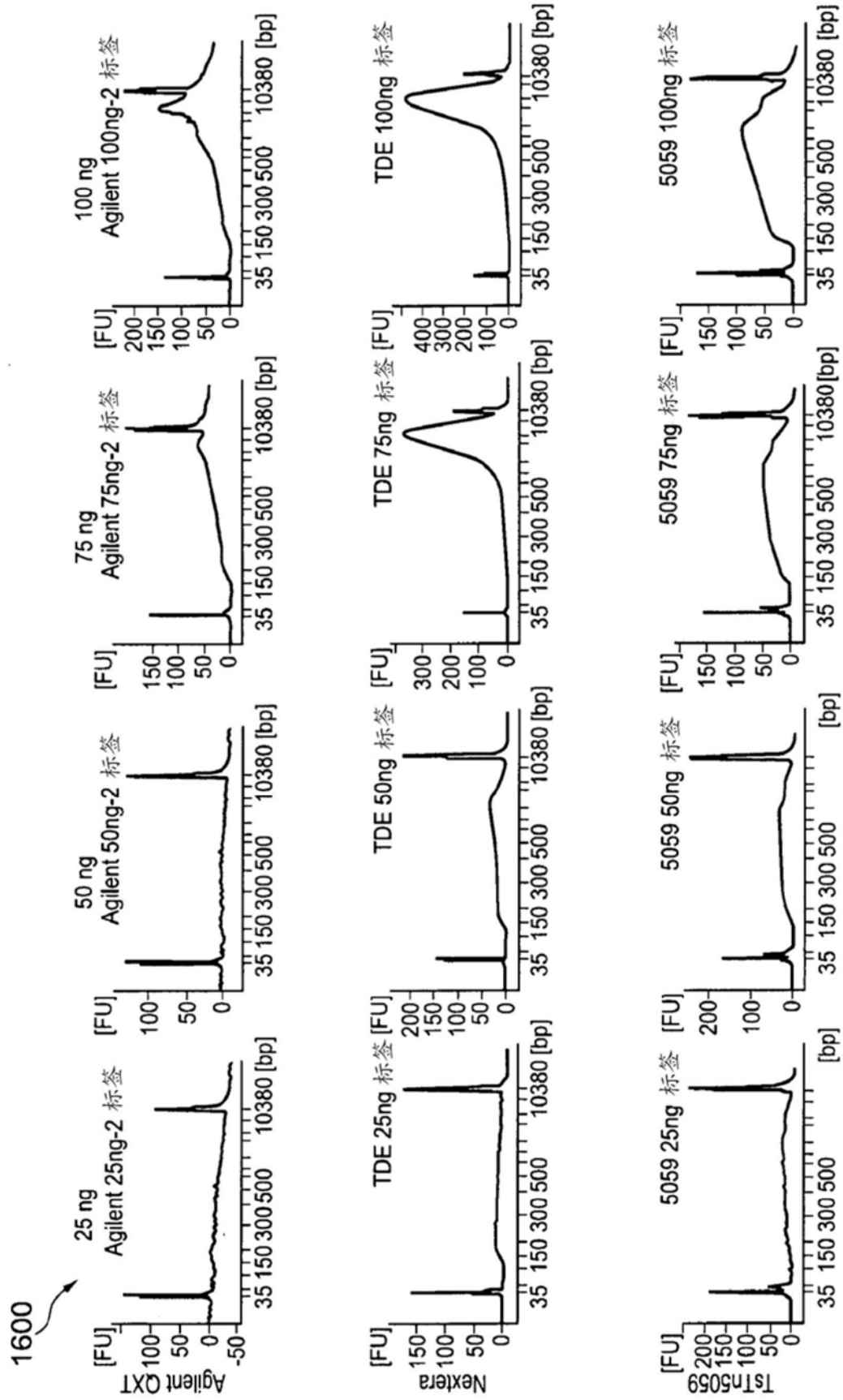


图28

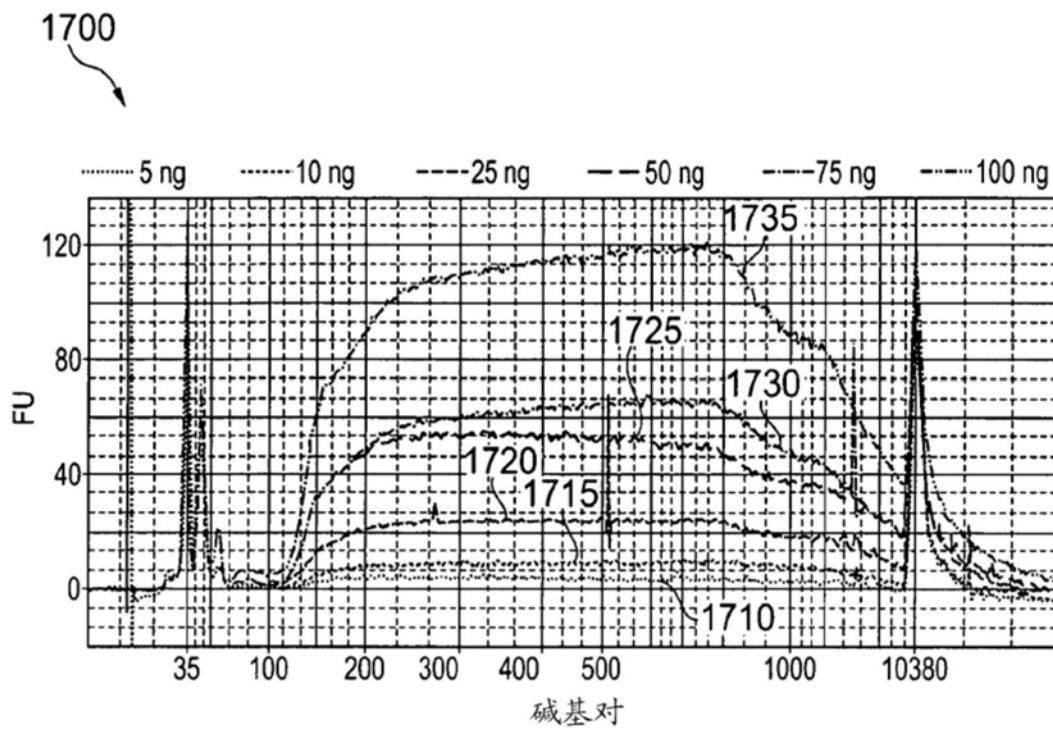


图29A

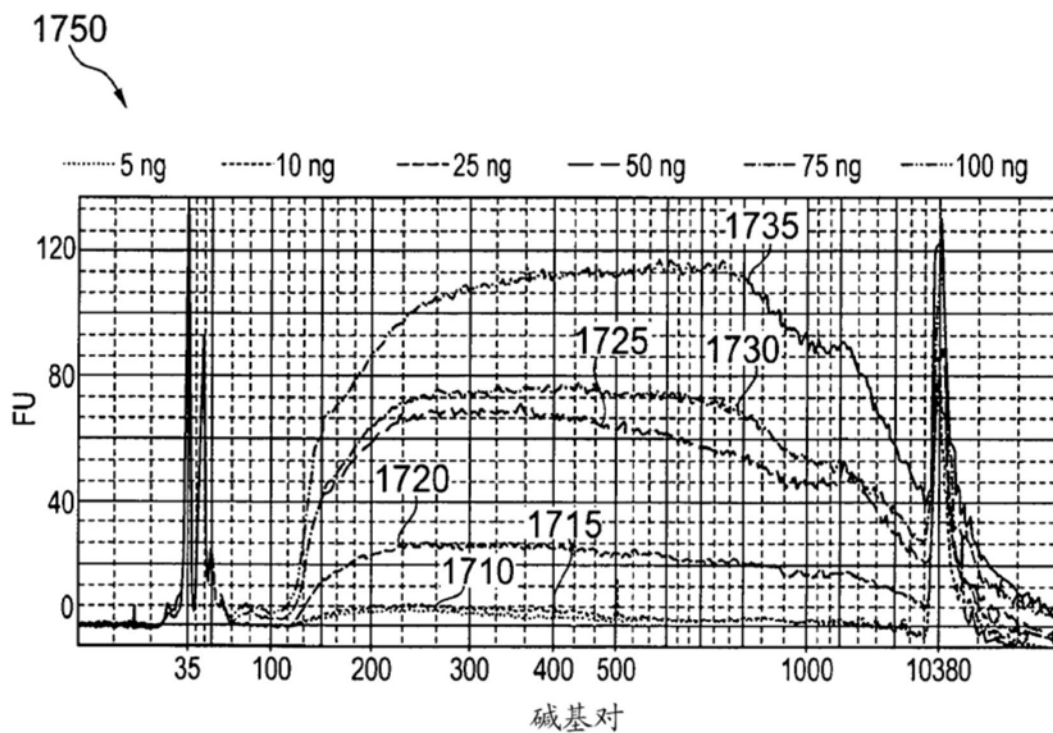


图29B

在6x“归一化”浓度的TDE1使用25 ng-100 ng gDNA

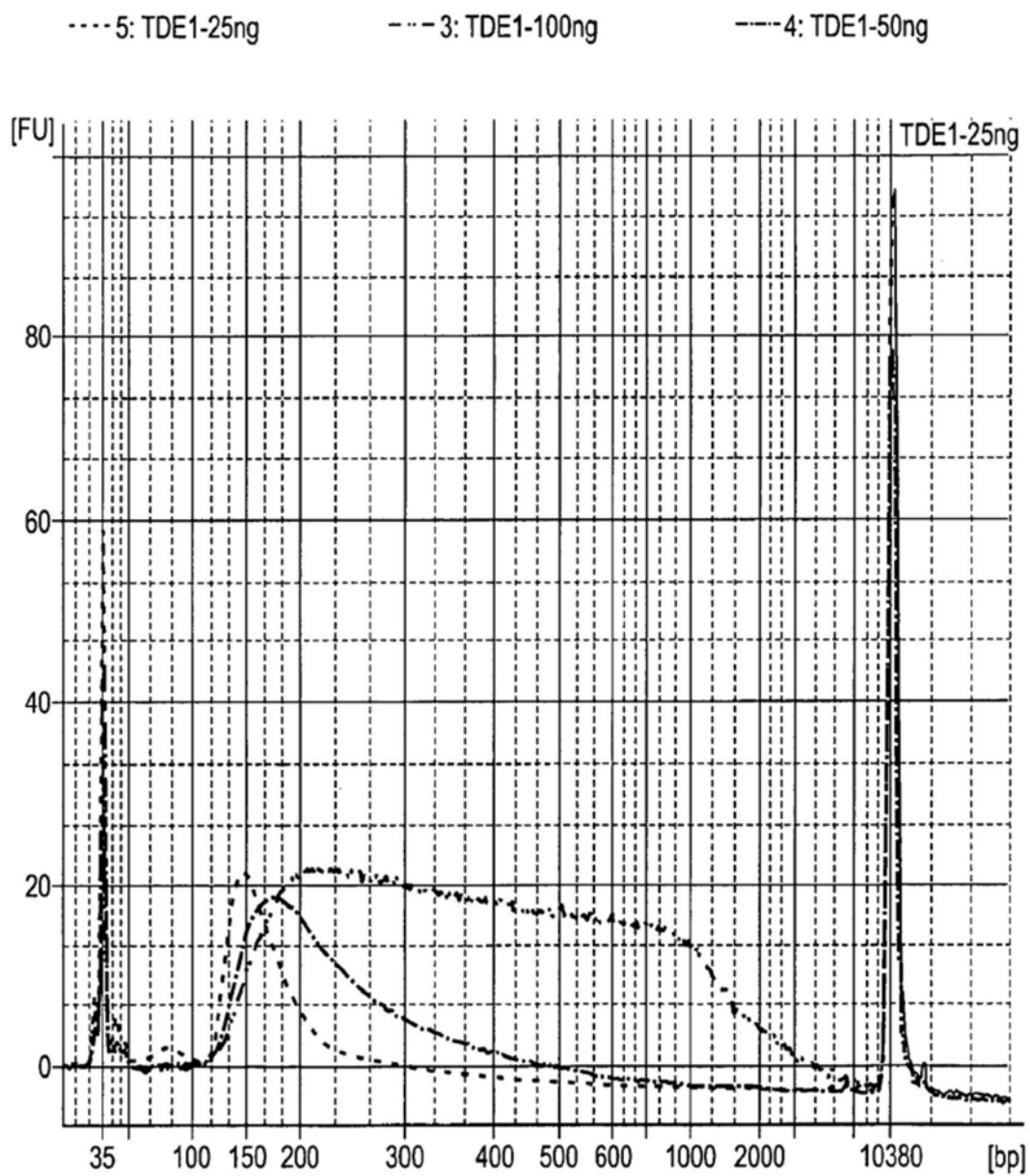


图30

在6x“归一化”浓度的TS-Tn5使用25 ng-100 ng gDNA

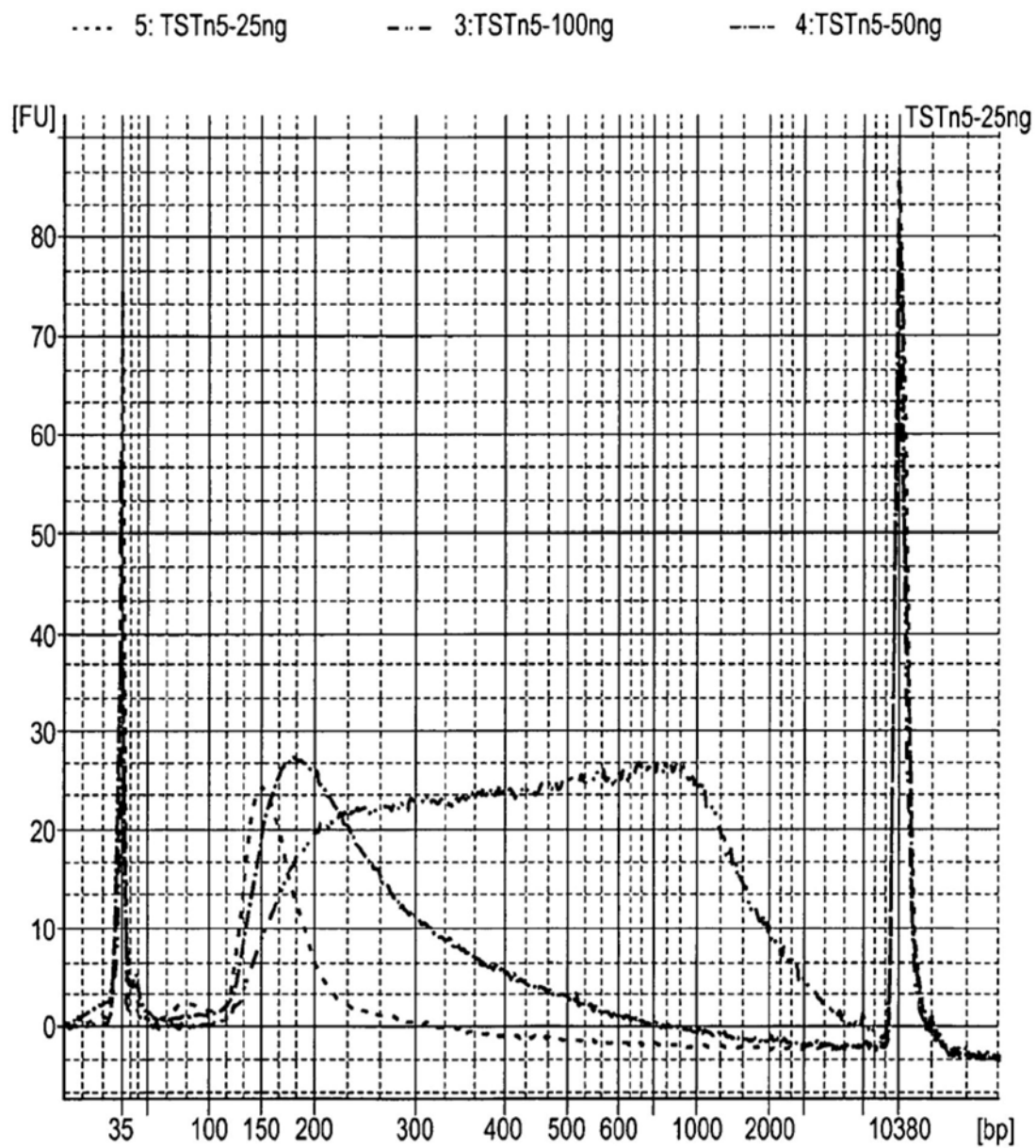


图31

在6x“归一化”浓度的TS-Tn5059使用10 ng-100 ng gDNA

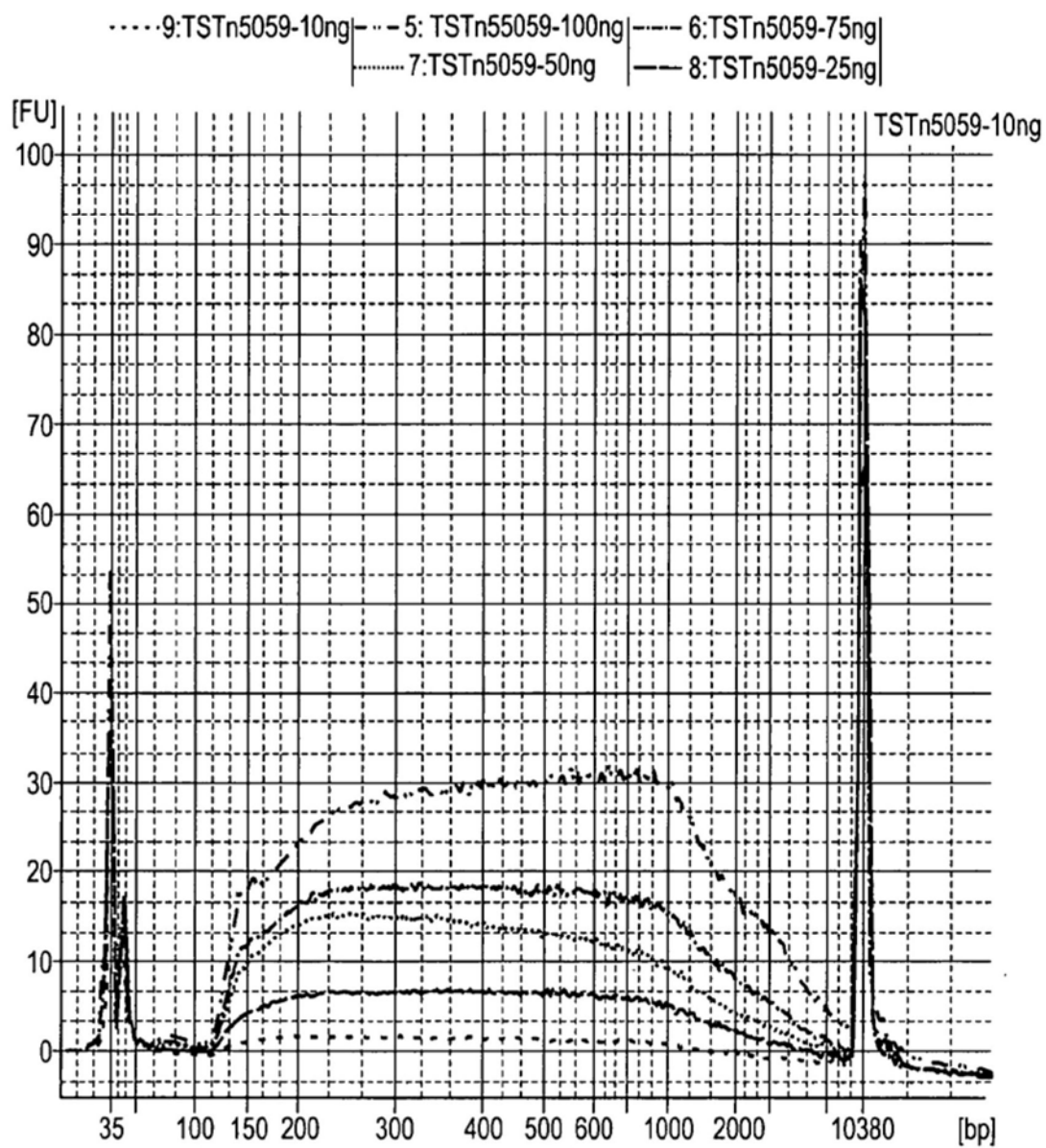


图32

在6x“归一化”浓度的TS-Tn5059使用5 ng-500 ng gDNA输入

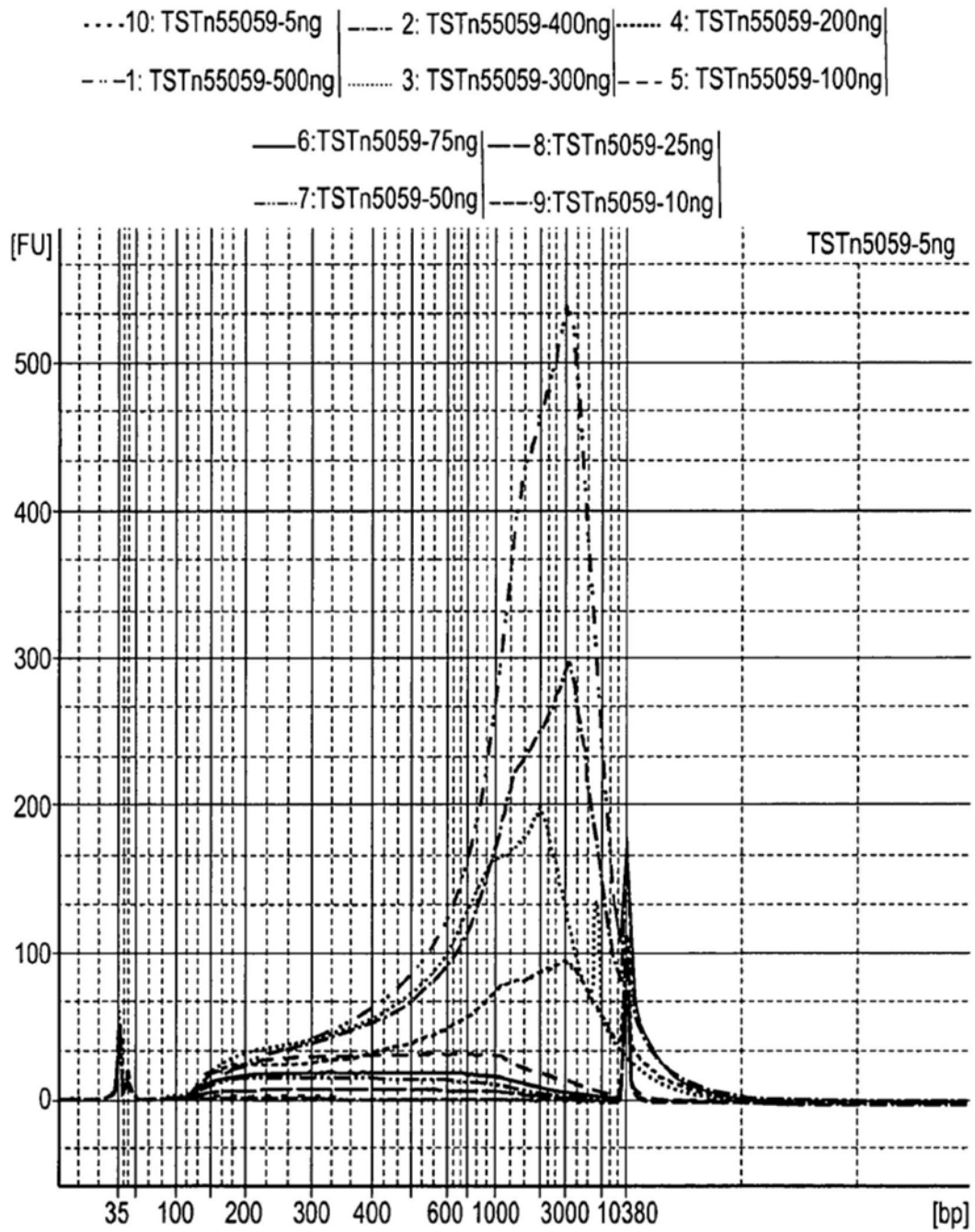


图33