



(19)中華民國智慧財產局

(12)發明說明書公告本

(11)證書號數：TW I682997 B

(45)公告日：中華民國 109 (2020) 年 01 月 21 日

(21)申請案號：104112180

(22)申請日：中華民國 104 (2015) 年 04 月 15 日

(51)Int. Cl. : C12N9/12 (2006.01) C12Q1/68 (2018.01)

(30)優先權：2014/04/15 美國 61/979,871

2014/10/09 美國 62/062,006

2014/11/17 美國 62/080,882

(71)申請人：美商伊路米納有限公司 (美國) ILLUMINA, INC. (US)

美國

(72)發明人：克勒克納 克利斯丁 GLOECKNER, CHRISTIAN (US)；奇拉 阿米拉利 KIA, AMIRALI (IR)；波瑪堤 厄尼 BOMATI, ERIN (US)；赫 莫里 HE, MOLLY (US)；格呂嫩瓦爾德 海音 里 GRUNENWALD, HAIYING LI (US)；克爾史丹 史考特 KUERSTEN, SCOTT (US)；歐薩斯普拉羅普 崔妮 法耶 OSOTHPRAROP, TRINA FAYE (US)；哈斯金 達林 HASKINS, DARIN (US)；柏各司 喬許華 BURGESS, JOSHUA (US)；肯納 安努帕瑪 KHANNA, ANUPAMA (US)；施林曼 丹尼爾 SCHLINGMAN, DANIEL (US)；瓦伊德亞納坦 拉梅什 VAIDYANATHAN, RAMESH (US)

(74)代理人：閻啟泰；林景郁

(56)參考文獻：

Gradman RJ et al., "A bifunctional DNA binding region in Tn5 transposase", Molecular Microbiology 67(3), 528-540, 2007/12/16
 Ason B and Reznikoff WS, "DNA sequence bias during Tn5 transposition", J. Mol. Biol., 335(5):1213-1225, 2004/01/30

審查人員：施雅儀

申請專利範圍項數：19 項 圖式數：33 共 164 頁

(54)名稱

用於改善插入序列傾向及增加DNA輸入耐受度之經修飾的轉位酶

(57)摘要

本文呈現用於改善核酸樣品之片段化及標記之轉位酶及反應條件，特定言之展現改善之插入序列傾向之經改變轉位酶及反應條件，以及使用其之方法及套組。

Presented herein are transposase enzymes and reaction conditions for improved fragmentation and tagging of nucleic acid samples, in particular altered transposases and reaction conditions which exhibit improved insertion sequence bias, as well as methods and kits using the same.

指定代表圖：



圖1A

處之突變為取代突變。在某些態樣中，位置 Trp125 處之取代突變可包含變成甲硫胺酸殘基之突變。

【0010】 本文亦呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Lys120 處之突變。在某些態樣中，位置 Lys120 處之突變為取代突變。在某些態樣中，位置 Lys120 處之取代突變可包含變成大型芳族殘基之突變。在某些態樣中，位置 Lys120 處之取代突變可包含變成選自由以下組成之群的殘基之突變：Tyr、Phe、Trp 及 Glu。

【0011】 本文亦呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Lys212 及/或 Pro214 及/或 Ala338 處之突變。在某些態樣中，位置 Lys212 及/或 Pro214 及/或 Ala338 處之一或多個突變為取代突變。在某些態樣中，位置 Lys212 處之取代突變包含變成精胺酸之突變。在某些態樣中，位置 Pro214 處之取代突變包含變成精胺酸之突變。在某些態樣中，位置 Ala338 處之取代突變包含變成纈胺酸之突變。在一些具體實例中，轉位酶可另外包含 Gly251 處之取代突變。在某些態樣中，位置 Gly251 處之取代突變包含變成精胺酸之突變。

【0012】 本文亦呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Glu146 及/或 Glu190 及/或 Gly251 處之突變。在某些態樣中，位置 Glu146 及/或 Glu190 及/或 Gly251 處之一或多個突變為取代突變。在某些態樣中，位置 Glu146 處之取代突變可包含變成麩醯胺酸之突變。在某些態樣中，位置 Glu190 處之取代突變可包含變成甘胺酸之突變。在某些態樣中，位置 Gly251 處之取代突變可包含變成精胺酸之突變。

【0013】 亦提供包含對含有 SEQ ID NO: 21 之胺基酸序列之半保守域

之取代突變的經改變轉位酶，其中該取代突變包含在位置 2 處變成除 Trp、Asn、Val 或 Lys 以外的任何殘基之突變。在某些具體實例中，突變包含在位置 2 處變成 Met 之取代。

【0014】 在上述具體實例中之任一者中，突變 Tn5 轉位酶可另外包含功能上等效於 Tn5 轉位酶胺基酸序列中之 Glu54 及/或 Met56 及/或 Leu372 之位置處之取代突變。在某些具體實例中，轉位酶包含與 Tn5 轉位酶胺基酸序列中之 Glu54Lys 及/或 Met56Ala 及/或 Leu372Pro 同源之取代突變。

【0015】 本文亦呈現包含 SEQ ID NO: 2-10 及 12-20 中之任一者之胺基酸序列之突變 Tn5 轉位酶。

【0016】 本文亦呈現包含融合至額外多肽的如任何以上具體實例中所定義之突變 Tn5 轉位酶之融合蛋白。在一些具體實例中，融合至轉位酶之多肽域可包含純化標籤、表現標籤、溶解度標籤或其組合。在一些具體實例中，融合至轉位酶之多肽域可包含例如麥芽糖結合蛋白（MBP）。在一些具體實例中，融合至轉位酶之多肽域可包含例如延伸因子 Ts（Tsf）。

【0017】 本文亦呈現編碼如任何以上具體實例中所定義之突變 Tn5 轉位酶之核酸分子。本文亦呈現包含上文所述之核酸分子之表現載體。本文亦呈現包含上文所述之載體之宿主細胞。

【0018】 本文亦呈現試管內轉位之方法，其包含：允許以下組分相互作用：(i) 包含如上文描述之任一具體實例之突變 Tn5 轉位酶之轉位體（transposome）複合物，及 (ii) 目標 DNA。

【0019】 本文亦呈現使用上文描述之 Tn5 轉位酶定序目標 DNA 之方法。在一些具體實例中，該方法可包含 (a) 在藉以使得目標 DNA 經片段

化，且第一聚核苷酸之 3'轉位子末端序列轉移至片段之 5'末端之條件下藉由包含（1）如上文描述之具體實例中之任一者之突變 Tn5 轉位酶；及（2）包含（i）包含轉位子末端序列之 3'部分，及（ii）包含第一定序標籤域之第一標籤之轉位體複合物培育目標 DNA，進而產生雙股片段，其中 5'末端經第一標籤標記，且在 5'標記股之 3'末端處存在單股間隙；（b）藉由核酸修飾酶在藉以使得第二標籤附接至 5'標記股之 3'末端之條件下培育片段，（c）視情況藉由提供聚合酶及對應於第一聚核苷酸之一部分的擴增引子來擴增片段，進而產生在 5'末端處具有第一標籤且在 3'末端處具有第二標籤之代表性二標記片段庫；（d）提供包含對應於第一定序標籤域之部分的第一定序引子；及（e）並行地擴展第一定序引子及偵測鄰接於二標記片段之代表性庫之第一定序標籤域之核苷酸之身分。

【0020】 本文中亦呈現用於進行試管內轉位反應之套組。在一些具體實例中，套組可包含轉位體複合物，其包含（1）如上文描述之具體實例中之任一者之突變 Tn5 轉位酶；及（2）包含有含轉位子末端序列之 3'部分的聚核苷酸。

【0021】 一或多個具體實例之細節闡述於以下隨附圖式及實施方式中。其他特徵、目標及優勢將自實施方式及附圖以及申請專利範圍顯而易見。

【圖式簡單說明】

【0022】

圖 1A 為顯示 Tn5 轉位酶（1MUH）、Hermes 轉位酶（2BW3）、HIV 整合酶（1ITG）、Mu 轉位酶（1BCM）及 Mos1 轉位酶（3HOS）之催化核心域之

結構比對的示意圖。顯示之編號表示 Tn5 轉位酶中之胺基酸殘基之編號。

圖 1B 為顯示 Tn5 轉位酶 (1MUH, 粉紅色)、Hermes 轉位酶 (2BW3, 黑色)、HIV 整合酶 (1ITG, 茶色)、Mu 轉位酶 (1BCM) 及 Mos1 轉位酶 (3HOS, 黃色) 之催化核心域之結構比對的示意圖。Tn5 轉位酶 W125 位置以棒狀圖示顯示。

圖 2 為顯示 D248Y 突變 Tn5 轉位酶相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 3 為顯示 D119L 突變 Tn5 轉位酶相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 4 為顯示 W125M 突變 Tn5 轉位酶相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 5 為顯示 ia248V 插入突變 Tn5 轉位酶相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 6 為顯示 K120Y、K120F 及 K120W Tn5 轉位酶插入突變體相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 7 為顯示三種突變 Tn5 轉位酶相比於 Tn5 對照組之經改變序列插入傾向之 IVC 曲線圖。

圖 8A 為顯示相比於 Tn5 對照組的藉由三種突變 Tn5 轉位酶產生之蠟狀芽孢桿菌庫中之 AT/GC 丟失之圖。圖 8B 為顯示相比於 Tn5 對照的藉由三種突變 Tn5 轉位酶測試之蠟狀芽孢桿菌庫之估計之庫大小之圖。

圖 9A 為顯示相比於 Tn5 對照組的藉由兩種突變 Tn5 轉位酶測試之庫中之快速捕獲富集實驗中之覆蓋度均一性之圖。圖 9B 為顯示相比於 Tn5 對照

組的藉由兩種突變 Tn5 轉位酶產生之庫中之快速捕獲富集實驗中之 10x 及 20x 目標覆蓋度及平均目標覆蓋度之圖。

圖 10A 為顯示相比於 Tn5 對照組的藉由兩種突變 Tn5 轉位酶產生之庫中之快速捕獲富集實驗中之通過獨特讀數之過濾的 % 及混合選擇庫尺寸之圖。圖 10B 為顯示相比於 Tn5 對照組的藉由兩個突變 Tn5 轉位酶產生之庫中之快速捕獲富集實驗中達到 10x、20x 及 30x 覆蓋度之罰分之圖。

圖 11 顯示使用不同標籤化 (tagmentation) 緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之獨特分子數之柱狀圖。

圖 12 顯示使用不同標籤化緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之 GC 丟失 % 之柱狀圖。

圖 13 顯示使用不同標籤化緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之 AT 丟失 % 之柱狀圖。

圖 14 顯示使用標準緩衝劑 (TD) 及鈷緩衝劑 (Co) 調配物製備之 TS-Tn5059 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 15 顯示使用鈷-DMSO (Co-DMSO) NF2 及 HMW 緩衝劑調配物製備之 TS-Tn5059 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 16 顯示使用標準緩衝劑調配物 (TD) 及鈷緩衝劑 (Co) 製備之 TS-Tn5 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 17 顯示使用鈷-DMSO (Co-DMSO)、NF2 及 HMW 緩衝劑調配物製備之 TS-Tn5 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 18A、18B、18C 及 18D 分別顯示 TS-Tn5 庫中之序列含量之傾向圖、TS-TN5-Co 庫中之序列含量之傾向圖、TS-Tn5-Co-DMSO 庫中之序列含量之

傾向圖及 TS-Tn5-NF2 庫中之序列含量之傾向圖。

圖 19A、19B、19C 及 19D 分別顯示 TS-Tn5059 庫中之序列含量之傾向圖、TS-TN5059-Co 庫中之序列含量之傾向圖、TS-Tn5059-Co-DMSO 庫中之序列含量之傾向圖及 TS-Tn5059-NF2 庫中之序列含量之傾向圖。

圖 20 顯示在使用不同標籤化緩衝劑製備之 MBP-Mos1 標籤化庫中之平均讀數總數及平均多樣性之柱狀圖。

圖 21 顯示 MBP-Mos1 標籤化庫中之 GC 及 AT 丟失之柱狀圖。

圖 22A、22B、22C 及 22D 分別顯示 Mos1-HEPES 庫中之序列含量之傾向圖、Mos1-HEPES-DMSO 庫中之序列含量之傾向圖、Mos1-HEPES-DMSO-Co 庫中之序列含量之傾向圖及 Mos1-HEPES-DMSO-Mn 庫中之序列含量之傾向圖。

圖 23 說明製備及富集用於外顯子組定序之基因體 DNA 庫之方法之實施例之流程圖。

圖 24A 顯示使用 TS-Tn5059 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之覆蓋度之曲線圖。

圖 24B 顯示使用 NexteraV2 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之覆蓋度之曲線圖。

圖 25A 顯示使用 TS-Tn5059 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之間隙位置及間隙長度之曲線圖。

圖 25B 顯示使用 NexteraV2 轉位體製備之標籤化蠟狀芽孢桿菌基因體標籤化 DNA 庫中之間隙位置及間隙長度之曲線圖。

圖 26 顯示對於 25ng 人類 gDNA 使用針對在 40nM (1x標準化濃度) 下

【0026】 如在整個本申請案中所使用，轉位體之濃度與標準化活性可互換使用。另外，如在整個本申請案中所使用，轉位體之濃度與轉位酶之濃度可互換使用。

【0027】 如本文中所使用，術語「插入傾向」係指轉位酶對插入位點之序列偏好。舉例而言，若聚核苷酸樣品中之 A/T/C/G 之背景頻率同等分佈（25% A、25% T、25% C、25% G），則在轉位酶結合位點或裂解位點對一種核苷酸超過其他三種核苷酸之任何過度表現反映彼位點處之插入傾向。插入傾向可使用此項技術中已知之多種方法中之任一者量測。舉例而言，如下文實施例 1 中一般闡述，可定序插入位點且可比較任何特定核苷酸在插入位點中之各位置處之相對豐度。

【0028】 「插入傾向之改善」指示特定鹼基在經改變轉位酶之結合位點之一或多個位置處之頻率減少或增加以較接近聚核苷酸樣品中之該鹼基之背景頻率。改善可為相對於未改變轉位酶中之彼位置之頻率的彼位置處之頻率之增加。或者，改善可為相對於未改變轉位酶中之彼位置之頻率的彼位置處之頻率之減少。因此，舉例而言，若聚核苷酸樣品中之 T 核苷酸之背景頻率為 0.25，且經改變轉位酶將轉位酶結合位點中之指定位置處之 T 核苷酸之頻率自高於 0.25 之頻率減少至較接近 0.25 之頻率，則經改變轉位酶具有插入傾向之改善。類似地，舉例而言，若聚核苷酸樣品中之 T 核苷酸之背景頻率為 0.25，且經改變轉位酶將轉位酶結合位點中之指定位置處之 T 核苷酸之頻率自低於 0.25 之頻率增加至較接近 0.25 之頻率，則經改變轉位酶具有插入傾向之改善。

【0029】 一種量測插入傾向之方法為藉由插入位點之大規模定序及

量測相對於插入位點的結合位點中之各位置處之鹼基之頻率，如例如在 Green 等人 *Mobile DNA* (2012) 3:3 中所述，該文獻以全文引用之方式併入本文中。顯示各位置處之豐度之典型工具為強度相對於循環分佈曲線圖，例如如圖 2 中所示。如在下文實施例 1 中所述，藉由轉位子介導標記及片段化產生之片段末端可以大規模定序，且可量測在插入位點之各位置處之鹼基之分佈頻率以偵測在插入位點之一或多個位置處之傾向。因此，舉例而言，如在圖 2 中指示，就『G』核苷酸而言之 0.55 及就『A』核苷酸而言之 0.16 之位置（1）處之鹼基分佈頻率反映在彼位置處對 G 之強烈偏好及遠離 A 之傾向。作為另一實例，且相比之下，如圖 3 中所示，位置（20）處之鹼基分佈對於四種鹼基中之每一者基本上為 0.25，反映在彼位置處之極小或無序列傾向。

【0030】 在本文呈現之一些具體實例中，經改變轉位酶在位於插入位點上游或下游 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20 或 20 個以上鹼基處之一或多個位點處提供插入傾向之減小。在一些具體實例中，經改變轉位酶在位於插入位點下游 1 至 15 個鹼基處之一或多個位點處提供插入傾向之減小。在一些具體實例中，經改變轉位酶在位於插入位點上游 1 至 15 個鹼基處之一或多個位點處提供插入傾向之減小。

【0031】 如在下文中更詳細地描述，本發明人已出人意料地發現在轉位酶胺基酸序列之某些位置對殘基之一或多個突變導致在轉位事件期間的經改善序列插入傾向。此等經改變轉位酶在高及低相異性核酸樣品之標籤化中給出經改善效能，導致定序之各區域之較大覆蓋度均一性及較少丟失。

【0032】 如本文中所使用，術語「DNA 輸入耐受度」係指轉位酶產生針對一定範圍之輸入 DNA 量之均一 DNA 片段尺寸之能力。

【0033】 如本文所使用，延伸因子之記法 TS 可與 Tsf 互換使用。

【0034】 在一些具體實例中，輸入 DNA 為基因體 DNA。在一些具體實例中，輸入 DNA 之範圍可為 0.001 μ g 至 1 mg、1 ng 至 1 mg、1 ng 至 900 ng、1 ng 至 500 ng、1 ng 至 300 ng、1 ng 至 250 ng、1 ng 至 100 ng、5 ng 至 250 ng 或 5 ng 至 100 ng 且轉位酶之濃度在 5 nM 與 500 nM 之間。在一些具體實例中，就輸入 DNA 之上述範圍而言之轉位酶之濃度為約 25 nM、30 nM、35 nM、40 nM、50 nM、60 nM、65 nM、70 nM、75 nM、80 nM、90 nM、95 nM、100 nM、125 nM、130 nM、140 nM、150 nM、175 nM、180 nM、190 nM、200 nM、210 nM、225 nM、230 nM、240 nM、250 nM、260 nM、275 nM、280 nM、290 nM、300 nM、325 nM、350 nM、360 nM、375 nM、380 nM、390 nM、400 nM、425 nM、450 nM、475 nM 或 500 nM。在一些具體實例中，就輸入 DNA 之上述範圍而言之標準化濃度之轉位酶或標準化轉位體之濃度係選自約 0.1x 至 10x、1x 至 10x、3x 至 8x、4x 至 7x 之範圍。在一些具體實例中，就輸入 DNA 之上述範圍而言之轉位酶或標準化轉位體之標準化濃度為約 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。

【0035】 在一些具體實例中，輸入 DNA 之量為 1 ng、2 ng、3 ng、4 ng、5 ng、6 ng、7 ng、8 ng、9 ng、10 ng、11 ng、12 ng、13 ng、15 ng、20 ng、25 ng、30 ng、35 ng、40 ng、45 ng、50 ng、55 ng、60 ng、65 ng、70 ng、75 ng、80 ng、85 ng、90 ng、95 ng、100 ng、110 ng、115 ng、120 ng、125 ng、130 ng、135 ng、140 ng、150 ng、155 ng、160 ng、165 ng、170 ng、180 ng、185 ng、190 ng、195 ng、200 ng、210 ng、220 ng、225 ng、230 ng、235 ng、240 ng、245 ng、250 ng、260 ng、270 ng、280 ng、290 ng、300 ng、325 ng、350 ng、375 ng、400 ng、425 ng、450 ng、475 ng、500 ng、525 ng、550 ng、600 ng、650 ng、700 ng、750 ng、800 ng、850 ng 或 900 ng。在一些具體實例中，就輸入 DNA 之上述量而言之轉位酶之濃度為約 25 nM、30 nM、35 nM、40 nM、50 nM、60 nM、65 nM、70 nM、75 nM、80 nM、90 nM、95 nM、100 nM、125 nM、130 nM、140 nM、150 nM、175 nM、180 nM、190 nM、200 nM、210 nM、225 nM、230 nM、240 nM、250 nM、260 nM、275 nM、280 nM、290 nM、300 nM、325 nM、350 nM、360 nM、375 nM、380 nM、390 nM、400 nM、425 nM、450 nM、475 nM 或 500 nM。在一些具體實例中，就輸入 DNA 之上述量而言之標準化濃度之轉位酶或標準化轉位體之濃度係選自約 0.1x 至 10x、1x 至 10x、3x 至 8x、4x 至 7x 之範圍。在一些具體實例中，就輸入 DNA 之上述量而言之轉位酶或標準化轉位體之標準化濃度為約 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。

【0036】 在一些具體實例中，轉位酶之 nM 濃度與輸入 DNA 之 ng 量之比為約 0.5 至 5、1 至 5、2 至 5、2.1 至 3 或 2.1 至 2.5。

【0037】 如本文中所使用，術語「基因體 DNA」係指存在於包含一或多個編碼細胞之各種蛋白之基因之細胞中的核酸。在一些具體實例中，基因體 DNA 來自原核生物體，例如細菌及古菌。在一些具體實例中，基因體 DNA 來自真核生物體，例如人類、植物、真菌、阿米巴原蟲。

【0038】 如本文所用之術語「突變」或「經修飾」係指在相比於野生型基因或基因產物時，在序列及或功能特性中顯示改變（亦即經改變特徵）之基因或基因產物。「突變」或「經修飾」亦指位於一或多個特定核苷酸位置之序列，或位於一或多個特定密碼子位置之序列，或位於一或多個特定胺基酸位置之序列，其在相比於野生型基因或基因產物時，在序列及或功能特性中顯示改變（亦即經改變特徵）。

【0039】 如本文所用之「包括」與術語包含具有相同含義。

【0040】 如本文所用之「約」在數量上意謂加上或減去 10%。

【0041】 如在下文中更詳細地敘述，本發明人已出人意料地發現在轉位酶胺基酸序列之某些位置處對殘基之一或多個突變導致 DNA 輸入耐受度增加，使得突變轉位酶相比於野生型轉位酶，針對一定範圍之輸入 DNA 量產生均一 DNA 片段尺寸。在一個具體實例中，TS-Tn5059 轉位酶展現相比於其他轉位酶，例如 TS-Tn5 及 Tn5 型式 1（TDE1）增加之 DNA 輸入耐受度。

【0042】 在一些具體實例中，TS-Tn5059 展現相比於其他轉位酶增加之 DNA 輸入耐受度，其中輸入 DNA 之範圍在 1 ng 至 200 ng 基因體 DNA 之間且 TS-Tn5059 之濃度在 100-300 nM 之間。在 TS-Tn5059 展現相比於其他轉位酶增加之 DNA 輸入耐受度之一些具體實例中，輸入 DNA 之範圍在 5 ng 至 200 ng 基因體 DNA 之間且 TS-Tn5059 之濃度在 100-250 nM 之間。在 TS-Tn5059 展現相比於其他轉位酶增加之 DNA 輸入耐受度之一些具體實例中，輸入 DNA 之範圍在 5 ng 至 100 ng 基因體 DNA 之間且 TS-Tn5059 之濃度在 240 nM 與 250 nM 之間。

【0043】 改善之插入傾向連同增加之 DNA 輸入耐受度提供比當前 Nextera®快速捕獲方案（Illumina 公司）更快及更靈活之樣品製備及外顯子組富集方案。

【0044】 如本文中所使用，術語「標籤化」係指藉由與包含轉位子末端序列之接附子複合之包含轉位酶之轉位體複合物的 DNA 修飾。標籤化導致同時的 DNA 片段化及接附子接合至雙螺旋片段之兩條股之 5'末端。

【0045】 如本文所使用，「轉位體複合物」或「轉位體」包含至少轉位酶及轉位酶識別位點。在一些此類系統中，轉位酶可與轉位子識別位點形成能夠催化轉位反應之功能複合物。轉位酶可在本文中稱為標籤化之方法中結合至轉位酶識別位點且將轉位酶識別位點插入至目標核酸中。在一些此類插入事件中，轉位酶識別位點之一條股可轉移至目標核酸中。

【0046】 本文中呈現之經改變轉位酶可形成轉位體複合物的一部分。例示性轉位複合物包括（但不限於）超高活性 Tn5 轉位酶及 Tn5 型轉位酶識別位點。超高活性 Tn5 轉位酶可包括美國專利第 5,925,545 號、美國

專利第 5,965,443 號、美國專利第 7,083,980 號及美國專利第 7,608,434 號以及 Goryshin 及 Reznikoff, *J. Biol. Chem.*, 273:7367 (1998)之揭示內容中所述之彼等轉位酶，該文獻中之每一者之內容以全文引用之方式併入本文中。然而，應瞭解，本文中呈現之經改變轉位酶可用於任何能夠將轉位子末端在足夠效率之情況下以隨機或幾乎隨機方式插入至標籤目標核酸之轉位系統，因為其預期目的可用於提供之方法中。

【0047】 舉例而言，呈現之經改變轉位酶可包括包含至少一個在功能上等效於 Tn5 胺基酸序列中之位點之一或多個位置處之胺基酸取代突變。於 Tn5 具有同源性之區域闡述於本文中，如圖 1 中所例示且允許鑑別其他轉位酶，例如 Hermes 轉位酶、HIV 整合酶、Mu 轉位酶及 Mos1 轉位酶中之功能等效位點。同樣，其他轉位酶或整合酶中之功能等效位點將為一般熟習此項技術者顯而易見，例如藉由進行 Tn5 胺基酸序列之序列比對及鑑別保守或半保守殘基或域。因此，應瞭解，可在本文提供之某些具體實例之情況下使用之轉位系統包括任何具有與 Tn5 之位點功能等效之位點的已知轉位酶。舉例而言，此類系統可包括 MuA 轉位酶及 Mu 轉位酶識別位點，其包含 R1 及 R2 末端序列 (Mizuuchi, K., *Cell*, 35:785, 1983; Savilahti, H 等人, *EMBO J.*, 14:4893, 1995)。

【0048】 包含於本文提供之某些具體實例中之轉位系統之更多實例包括金黃色葡萄球菌 Tn552 (Colegio 等人, *J. Bacteriol.*, 183: 2384-8, 2001; Kirby C 等人, *Mol. Microbiol.*, 43: 173-86, 2002)、Ty1 (Devine 及 Boeke, *Nucleic Acids Res.*, 22: 3765-72, 1994 及國際公開案 WO 95/23875)、轉位子 Tn7 (Craig, N L, *Science*. 271: 1512, 1996; Craig, N L, Review in: *Curr Top Microbiol Immunol.*,

204:27-48, 1996) 、Tn/O 及 IS10 (Kleckner N 等人, Curr Top Microbiol Immunol., 204:49-82, 1996) 、Mariner 轉位酶 (Lampe D J 等人, EMBO J., 15: 5470-9, 1996) 、Tc1 (Plasterk R H, Curr. Topics Microbiol. Immunol., 204: 125-43, 1996) 、P Element (Gloor, G B, Methods Mol. Biol., 260: 97-114, 2004) 、Tn3 (Ichikawa 及 Ohtsubo, J Biol. Chem. 265:18829-32, 1990) 、細菌插入序列 (Ohtsubo 及 Sekine, Curr. Top. Microbiol. Immunol. 204: 1-26, 1996) 、反轉錄病毒 (Brown 等人, Proc Natl Acad Sci USA, 86:2525-9, 1989) 及酵母菌之反轉錄轉位子 (Boeke 及 Corces, Annu Rev Microbiol. 43:403-34, 1989) 。更多實例包括 IS5 、Tn10 、Tn903 、IS911 及轉位酶家族酶之工程改造形式 (Zhang 等人, (2009) PLoS Genet. 5:e1000689. 電子版 2009 年 10 月 16 日 ; Wilson C. 等人 (2007) J. Microbiol. Methods 71:332-5) 。上文所引用之參考文獻以全文引用的方式併入本文中。

【0049】 簡言之，「轉位反應」為其中一或多個轉位子插入至隨機位點或幾乎隨機位點處之目標核酸中的反應。轉位反應中之必需組分為轉位酶及展現轉位子之核苷酸序列之 DNA 寡核苷酸，包括經轉移轉位子序列及其補體（亦即非轉移轉位子末端序列）以及形成功能轉位或轉位體複合物所需的其他組分。DNA 寡核苷酸可按需要或所需另外包含其他序列（例如接附子或引子序列）。

【0050】 添加至核酸之 5'及/或 3'末端之接附子可包含通用序列。通用序列為兩個或兩個以上核酸分子共同，亦即共用之核苷酸序列區域。視情況，兩個或兩個以上核酸分子亦具有序列差異之區域。因此，舉例而言，5'接附子可包含一致或通用核酸序列且 3'接附子可包含一致或通用序列。可存在於複數個核酸分子之不同成員中之通用序列可允許使用與通用序列互

補之單一通用引子複製或擴增多個不同序列。

轉位酶突變體

【0051】 因此，本文呈現相對於野生型轉位酶經修飾之突變轉位酶。經改變轉位酶可包含至少一個在功能上等效於下表 1 中所闡述之彼等殘基之一或多個位置處之胺基酸取代突變。表 1 闡述已顯示導致改善之插入傾向的轉位酶殘基處之取代突變。如表 1 中所闡述，本文中呈現之取代突變可在任何功能轉位酶主鏈中，諸如在本文中例示為 SEQ ID NO: 1 之野生型 Tn5 轉位酶，或具有對於其他位點之其他突變之轉位酶，包括發現與稱為超高活性 Tn5 轉位酶之轉位酶序列中之彼等突變，諸如美國專利第 5,925,545 號、美國專利第 5,965,443 號、美國專利第 7,083,980 號及美國專利第 7,608,434 號之併入材料中所闡述之一或多個突變，及如在本文中例示為 SEQ ID NO: 11。

表 1-導致改善之插入傾向之突變之實例

主鏈	Tn5 突變體	SEQ ID NO:
Tn5 WT	D248Y D248T D248K D248S D248L D248A D248W D248P D248G D248R D248F D248H	2
Tn5 WT	D119L D119M D119S D119A D119V	3
Tn5 WT	W125M	4
Tn5 WT	iaD248	5
Tn5 WT	K120F K120Y K120E K120W	6
Tn5 WT	Y、T、K、S、L、A、W、P、G、R、F 或 H 之 D248 L、M、S、A 或 V 之 D119 W125M K120F	7
Tn5 WT	K212R P214R A338V	8
Tn5 WT	K212R P214R G251R A338V	9
Tn5 WT	E146Q E190G G251R	10
Tn5 超高活性	D248Y D248T D248K D248S D248L D248A	12

	D248W D248P D248G D248R D248F D248H	
Tn5 超高活性	D119L D119M D119S D119A D119V	13
Tn5 超高活性	W125M	14
Tn5 超高活性	iaD248	15
Tn5 超高活性	K120F K120Y K120E K120W	16
Tn5 超高活性	Y、T、K、S、L、A、W、P、G、R、F 或 H 之 D248 L、M、S、A 或 V 之 D119 W125M K120F	17
Tn5 超高活性	K212R P214R A338V	18
Tn5 超高活性	K212R P214R G251R A338V	19
Tn5 超高活性	E146Q E190G G251R	20

【0052】 如在此項技術中所理解，上表中所列之參考編號係指野生型 Tn5 序列（SEQ ID NO: 1）之胺基酸位置。一般熟習此項技術者應理解，編號可由於 N 末端截斷、插入或融合而變化。以上列出之胺基酸之功能位置將保持相同，儘管位置編號可能已改變。舉例而言，SEQ ID NO: 25 中闡述之序列之第一組 285 個胺基酸殘基包含大腸桿菌 TS 之 N 末端融合，接著為 SEQ ID NO: 11 之胺基酸殘基 2-476。因此，舉例而言，SEQ ID NO: 25 之 Pro 656 功能上對應於 SEQ ID NO: 11 之 Pro 372。

【0053】 因此，在某些具體實例中，本文中呈現之經改變轉位酶包含功能上等效於例如 Tn5 轉位酶胺基酸序列中之 Asp248、Asp119、Trp125、Lys120、Lys212、Pro214、Gly251、Ala338、Glu146 及/或 Glu190 之一或多個位置處相對於野生型轉位酶之至少一個胺基酸取代突變。

【0054】 在一些具體實例中，突變轉位酶可包含位置 Asp248 處之突變。位置 Asp248 處之突變可例如為取代突變或插入突變。在某些具體實例中，突變為變成除 Asp 以外之任何殘基之取代突變。在某些具體實例中，位置 Asp248 處之取代突變包括變成選自由以下組成之群的殘基之突變：Tyr、Thr、Lys、Ser、Leu、Ala、Trp、Pro、Gln、Arg、Phe 及 His。

【0055】 在某些具體實例中，位置 Asp248 處之突變為位置 Asp248 之後的插入突變。在某些態樣中，插入突變可包含在 Asp248 之後插入任何殘基。在某些態樣中，插入突變可包含在位置 Asp248 之後插入疏水性殘基。疏水性殘基為熟習此項技術者已知且包括例如 Val、Leu、Ile、Phe、Trp、Met、Ala、及 Cys。在某些態樣中，插入突變可包含在位置 Asp248 之後插入羧胺酸殘基。

【0056】 本文中呈現之一些具體實例包括相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Asp119 處之突變。在某些態樣中，位置 Asp119 處之突變為取代突變。在某些態樣中，位置 Asp119 處之取代突變可包含變成疏水性殘基之突變。疏水性殘基為熟習此項技術者已知且包括例如 Val、Leu、Ile、Phe、Trp、Met、Ala、Tyr 及 Cys。在某些態樣中，Asp119 位置處之取代突變可包含變成親水性殘基之突變。親水性殘基為熟習此項技術者已知且包括例如 Arg、Lys、Asn、His、Pro、Asp

及 Glu。在某些態樣中，位置 Asp119 處之取代突變可包含變成選自由以下組成之群的殘基之突變：Leu、Met、Ser、Ala 及 Val。

【0057】 本文中呈現之一些具體實例包括相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Trp125 處之突變。在某些態樣中，位置 Trp125 處之突變為取代突變。在某些態樣中，位置 Trp125 處之取代突變可包含變成甲硫胺酸殘基之突變。

【0058】 本文中呈現之一些具體實例包括相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Lys120 處之突變。在某些態樣中，位置 Lys120 處之突變為取代突變。在某些態樣中，位置 Lys120 處之取代突變可包含變成大型芳族殘基之突變。特性化為大型芳族殘基之殘基為熟習此項技術者已知且包括例如 Phe、Tyr 及 Trp。在某些態樣中，位置 Lys120 處之取代突變可包含變成選自由以下組成之群的殘基之突變：Tyr、Phe、Trp 及 Glu。

【0059】 本文中呈現之一些具體實例包括相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Lys212 及/或 Pro214 及/或 Ala338 處之突變。在某些態樣中，位置 Lys212 及/或 Pro214 及/或 Ala338 處之一或多個突變為取代突變。在某些態樣中，位置 Lys212 處之取代突變包含變成精胺酸之突變。在某些態樣中，位置 Pro214 處之取代突變包含變成精胺酸之突變。在某些態樣中，位置 Ala338 處之取代突變包含變成纈胺酸之突變。在一些具體實例中，轉位酶可另外包含 Gly251 處之取代突變。在某些態樣中，位置 Gly251 處之取代突變包含變成精胺酸之突變。

【0060】 本文中呈現之一些具體實例包括相對於野生型 Tn5 轉位酶

經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Glu146 及/或 Glu190 及/或 Gly251 處之突變。在某些態樣中，位置 Glu146 及/或 Glu190 及/或 Gly251 處之一或多個突變為取代突變。在某些態樣中，位置 Glu146 處之取代突變可包含變成麩醯胺酸之突變。在某些態樣中，位置 Glu190 處之取代突變可包含變成甘胺酸之突變。在某些態樣中，位置 Gly251 處之取代突變可包含變成精胺酸之突變。

【0061】 在上述具體實例中之任一者中，突變 Tn5 轉位酶可另外包含功能上等效於 Tn5 轉位酶胺基酸序列中之 Glu54 及/或 Met56 及/或 Leu372 之位置處之取代突變。在某些具體實例中，轉位酶包含與 Tn5 轉位酶胺基酸序列中之 Glu54Lys 及/或 Met56Ala 及/或 Leu372Pro 同源之取代突變。

【0062】 本文中呈現之一些具體實例包括突變 Tn5 轉位酶，其包含 SEQ ID NO: 2-10 及 12-20 中之任一者之胺基酸序列。

【0063】 本文中亦呈現包含對半保守域之取代突變之經改變轉位酶。如本文中所使用，術語「半保守域」係指在各種轉位酶中及/或在各種物質中完全保守，或至少部分保守之轉位酶的一部分。半保守域包含滯留於轉位酶之催化核心域中之胺基酸殘基。已出人意料地發現半保守域中之一或多個殘基之突變影響轉位酶活性，導致插入傾向之改善。

【0064】 在一些具體實例中，半保守域包含具有 SEQ ID NO: 21 中闡述之序列之胺基酸。SEQ ID NO: 21 對應於在本文中闡述為 SEQ ID NO: 1 之 Tn5 轉位酶胺基酸序列之殘基 124-133。顯示半保守域中之各種轉位酶中之保守之結構比對闡述於圖 1 中。圖 1 中顯示之轉位酶序列包括 Tn5 轉位酶 (1MUH)、Hermes 轉位酶 (2BW3)、HIV 整合酶 (1ITG)、Mu 轉位酶 (1BCM)

及 Mos1 轉位酶 (3HOS) 之催化核心域。

【0065】 已出人意料地發現半保守域中之一或多個殘基之突變導致插入傾向之改善。舉例而言，在本文中呈現之經改變轉位酶之一些具體實例中，取代突變包含在 SEQ ID NO: 21 之位置 2 處變成除 Trp 以外之任何殘基的突變。在某些具體實例中，經改變轉位酶包含在 SEQ ID NO: 21 之位置 2 處變成 Met 之突變。

【0066】 「功能上等效」意謂對照轉位酶在使用完全不同轉位酶之研究的情況下將含有認為出現於在酶中具有相同功能作用之其他轉位酶中之胺基酸位置處之胺基酸取代。舉例而言，Mu 轉位酶中之位置 288 處離胺酸至甲硫胺酸之突變 (K288M) 將功能上等效於 Tn5 轉位酶中之位置 125 處色胺酸至甲硫胺酸之取代 (W125M)。

【0067】 一般而言，兩種或兩種以上不同轉位酶中之功能上等效取代突變出現於轉位酶之胺基酸序列中之同源胺基酸位置處。因此，本文中使用之術語「功能上等效」亦涵蓋與給定突變「位置上等效」或「同源」之突變，無論突變胺基酸之特定功能是否已知。有可能基於序列比對及/或分子模擬鑑別兩種或兩種以上不同轉位酶之胺基酸序列中之位置上等效或同源胺基酸殘基。鑑別位置上等效及/或功能上等效殘基之序列比對及分子模擬之實例闡述於圖 1 中。因此，舉例而言，如圖 1 中所示，半保守域中之殘基鑑別為 Tn5 轉位酶胺基酸序列之位置 124-133。Hermes 轉位酶、HIV 整合酶、Mu 轉位酶及 Mos1 轉位酶中之對應殘基在圖中鑑別為垂直排列型的且是為位置上等效於以及功能上等效於 Tn5 轉位酶胺基酸序列中之對應殘基。

【0068】 上文描述之經改變轉位酶可包含已知增強轉位酶活性之一或多個態樣之其他取代突變。舉例而言，在一些具體實例中，除以上突變中之任一者以外，經改變 Tn5 轉位酶可另外包含功能上等效於 Tn5 轉位酶胺基酸序列中之 Glu54 及/或 Met56 及/或 Leu372 之位置處之取代突變。功能上等效於導致活性改善之 Tn5 轉位酶胺基酸序列中之 Glu54 及/或 Met56 及/或 Leu372 之位置處之一或多個位置處之多種取代突變中之任一者可如此項技術中已知且藉由美國專利第 5,925,545 號、美國專利第 5,965,443 號、美國專利第 7,083,980 號及美國專利第 7,608,434 號之揭示內容以及 Goryshin 及 Reznikoff, *J. Biol. Chem.*, 273:7367 (1998)之揭示內容例示進行，該文獻中之每一者以全文引用之方式併入本文中。具體實例，轉位酶包含與 Tn5 轉位酶胺基酸序列中之 Glu54Lys 及/或 Met56Ala 及/或 Leu372Pro 同源之取代突變。舉例而言，取代突變可包含與 Tn5 轉位酶胺基酸序列中之 Glu54Lys 及/或 Met56Ala 及/或 Leu372Pro 同源之取代突變。

突變轉位酶

【0069】 各種類型之突變誘發視情況用於本發明中，例如用以修飾轉位酶以產生變異體，例如根據如上文所論述之轉位酶模型及模型預測，或使用隨機或半隨機突變方法。一般而言，任何可用的突變誘發程序可用於製造轉位酶突變體。此類突變誘發程序視情況包括關於一或多個相關活性（例如改善之插入傾向）選擇突變核酸及多肽。可使用之程序包括（但不限於）：定點突變誘發、隨機點突變誘發、試管內或活體內同源重組（DNA 改組及組合重疊 PCR）、使用含有模板之尿嘧啶之突變誘發、寡核苷酸定向突變誘發、硫代磷酸酯修飾 DNA 突變誘發、使用有間隙的雙螺旋 DNA 之

突變誘發、點失配修復、使用修復缺陷宿主菌株、限制選擇及限制純化之突變誘發、缺失突變誘發、藉由總基因合成之突變誘發、簡併 PCR、雙股斷裂修復及熟習此項技術者已知之許多其他程序。用於突變之起始轉位酶可為本文中所指出之彼等轉位酶中之任一者，包括可用的轉位酶突變體，諸如美國專利第 5,925,545 號、美國專利第 5,965,443 號、美國專利第 7,083,980 號及美國專利第 7,608,434 號以及 Goryshin 及 Reznikoff, J. Biol. Chem., 273:7367 (1998)之揭示內容中鑑別之彼等者，該文獻中之每一者以全文引用的方式併入本文中。

【0070】 視需要，突變誘發可藉由來自天然存在的轉位酶分子或已知經改變或突變轉位酶（例如使用如前述參考文獻所指出之現有突變轉位酶）之已知資訊，例如如上文所論述的序列、序列比較、物理特性、晶體結構及/或其類似者引導。然而，在另一類具體實例中，修飾可基本上為隨機的（例如在經典或「家族」DNA 改組中，參見例如 Cramer 等人 (1998) 「DNA shuffling of a family of genes from diverse species accelerates directed evolution」Nature 391:288-291）。

【0071】 突變形式之額外資訊發現於 Sambrook 等人，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 等人編, Current Protocols, Greene Publishing Associates 公司與 John Wiley & Sons 公司之合資企業 (2011 年補充) (「Ausubel」) 及 PCR Protocols A Guide to Methods and Applications (Innis 等人編) Academic Press 公司 San Diego, Calif. (1990) (「Innis」) 中。本文中所引用之以下出版物及參考

文獻提供突變形式之其他細節：Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993)；Bass 等人, Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988)；Bordo 及 Argos (1991) Suggestions for "Safe" Residue Substitutions in Site-directed Mutagenesis 217:721-729；Botstein 及 Shortle, Strategies and applications of in vitro mutagenesis, *Science* 229:1193-1201 (1985)；Carter 等人 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 等人, Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996)；Eghtedarzadeh 及 Henikoff, Use of oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986)；Fritz 等人, Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, *Nucl. Acids Res.* 16: 6987-6999 (1988)；Grundstrom 等人, 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, *Nucleic Acids & Molecular Biology* (Eckstein, F.及 Lilley, D. M. J.編, Springer Verlag, Berlin)) (1987)；Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. USA* 82:488-492

(1985) ; Kunkel 等人, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987) ; Kramer 等人, The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984) ; Kramer 及 Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987) ; Kramer 等人, Point Mismatch Repair, *Cell* 38:879-887 (1984) ; Kramer 等人, Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988) ; Ling 等人, Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997) ; Lorimer 及 Pastan *Nucleic 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 及 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 等人, Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223: 1299-1301(1984) ; Sakamar 及 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 等人, Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988) ; Sayers 等人, 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 等人, *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 等人, The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13: 8749-8764 (1985) ; Taylor 等人, The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13: 8765-8787 (1985) ; Wells 等人, Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986) ; Wells 等人, Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985) ; Zoller 及 Smith, Oligonucleotide-directed mutagenesis using M 13-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 及 Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983) ; Zoller 及 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 等人 (1991) 「Making antibody fragments using phage display libraries」 *Nature* 352:624-628 ; Gibbs 等人 (2001) 「Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling」 *Gene* 271:13-20 ;

及 Hiraga 及 Arnold (2003) 「General method for sequence-independent site-directed chimeragenesis: J. Mol. Biol. 330:287-296。許多以上方法之其他細節可發現於 Methods in Enzymology 第 154 卷中，其亦描述藉由各種突變誘發方法有用地控制解決難題的問題。

製造及分離重組轉位酶

【0072】 一般而言，如本文中呈現之編碼轉位酶之核酸可藉由選殖、重組、試管內合成、試管內擴增及/或其他可用方法製得。多種重組方法可用於表現編碼如本文中呈現之轉位酶之表現載體。製造重組核酸、表現及分離所表現產物之方法已為所熟知且描述於此項技術中。多種例示性突變及突變組合，以及設計所需突變之策略描述於本文中。製造及選擇轉位酶之催化域中之突變之方法發現於本文中且例示於美國專利第 5,925,545 號、美國專利第 5,965,443 號、美國專利第 7,083,980 號及美國專利第 7,608,434 號中，該專利以全文引用的方式併入。

【0073】 關於突變、重組及試管內操縱方法（包括選殖、表現、PCR 及其類似者）之其他有用參考文獻包括 Berger 及 Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology 第 152 卷 Academic Press 公司, San Diego, Calif. (Berger)；Kaufman 等人 (2003) Handbook of Molecular and Cellular Methods in Biology and Medicine 第二版 Ceske (編) CRC Press (Kaufman)；及 The Nucleic Acid Protocols Handbook Ralph Rapley (編) (2000) Cold Spring Harbor, Humana Press 公司 (Rapley)；Chen 等人 (編) PCR Cloning Protocols, 第二版 (Methods in Molecular Biology, 第 192 卷) Humana Pres；及 Viljoen 等人 (2005) Molecular Diagnostic PCR Handbook Springer, ISBN 1402034032。

【0074】 另外，用於自細胞純化質粒或其他相關核酸之過多套組為市售的（參見例如 EasyPrep™、FlexiPrep™，二者均購自 Pharmacia Biotech；StrataClean™，購自 Stratagene；及 QIAprep™，購自 Qiagen）。任何分離及/或純化之核酸可另外經操縱以生產其他核酸、用於轉染細胞、併入相關載體中以感染有機體以用於表現及/或其類似者。典型的選殖載體含有轉錄及轉譯終止子、轉錄及轉譯起始序列及適用於調節特定目標核酸之表現之啟動子。載體視情況包含通用表現卡匣，該表現卡匣含有至少一個獨立的終止子序列、允許在真核細胞或原核生物或二者中複製卡匣之序列（例如，穿梭載體）及選擇標記，其用於原核及真核系統二者。載體適合於在原核細胞、真核細胞或二者中之複製及整合。

【0075】 在一些具體實例中，本文中呈現之轉位酶表現為融合蛋白。融合蛋白可增強諸如轉位酶之溶解度、表現及/或純化的特徵。如本文所使用，術語「融合蛋白」係指具有至少兩個通常不存在於單一天然多肽中之多肽域之單一多肽鏈。因此，如本文所用，天然存在的蛋白及其點突變體不為「融合蛋白」。較佳地，相關多肽與至少一個多肽域經由肽鍵融合且融合蛋白亦可包括來源於獨立蛋白之胺基酸部分之間的胺基酸之連接區。融合至相關多肽之多肽域可增強相關多肽之溶解度及/或表現且亦可提供純化標籤以允許純化來自宿主細胞或培養上清液或二者之重組融合蛋白。增加表現、純化及/或儲存期間之溶解度之多肽域已在此項技術中為所熟知且包括例如麥芽糖結合蛋白（MBP）及延伸因子 Ts（Tsf），其如 Fox, J.D.及 Waugh D.S., *E. coli* Gene Expression Protocols *Methods in Molecular Biology*, (2003) 205:99-117 及 Han 等人 *FEMS Microbiol. Lett.* (2007) 274:132-138 所例示，該

文獻中之每一者以全文引用的方式併入本文中。融合至相關多肽之多肽域可融合於相關多肽之 N 端或 C 端。術語「重組」係指人工組合兩個否則分離之序列片段，例如藉由化學合成或藉由基因工程改造技術操縱胺基酸或核酸之分離片段。

【0076】 在一個具體實例中，本發明提供包含經修飾之 Tn5 轉位酶及延伸因子 Ts (Tsf) 之轉位酶融合蛋白。Tsf-Tn5 融合蛋白可組裝成包含融合轉位酶及自由轉位子末端之功能二聚轉位體複合物。Tsf-Tn5 融合蛋白具有相比於未融合 Tn5 蛋白增加之溶解度及熱穩定性。

【0077】 在另一具體實例中，本發明提供包含經修飾之 Tn5 轉位酶及識別 5-甲基胞嘧啶之蛋白域之轉位酶融合蛋白。5-甲基胞嘧啶-Tn5 融合蛋白可組裝成包含融合轉位酶及自由轉位子末端之功能二聚轉位體複合物。5-甲基胞嘧啶結合蛋白域可例如用於將 Tn5 轉位體複合物靶向基因體之甲基化區。

【0078】 在另一具體實例中，本發明提供包含經修飾之 Tn5 轉位酶及蛋白 A 抗體結合域之轉位酶融合蛋白。蛋白 A-Tn5 融合蛋白可組裝成包含融合轉位酶及自由轉位子末端之功能二聚轉位體複合物。蛋白 A 之抗體結合域可例如用於將 Tn5 轉位體複合物靶向基因體之抗體結合區。

【0079】 本發明提供包含經修飾之 Tn5 轉位酶及所有或部分延伸因子 TS (Tsf) 之轉位酶融合蛋白。Tsf 為可用於增強表現於細菌表現系統，例如大腸桿菌表現系統中之異源蛋白之溶解度的蛋白標籤。Tsf 增加異源蛋白之溶解度之能力可歸因於 Tsf 蛋白之內在高摺疊效率。在蛋白純化實驗（資料未示出）中，Tsf 以與蛋白 Tu 之複合物形式純化。為了使 Tu 結合於

複合物中，Tsf 需要恰當摺疊。Tsf-Tu 複合物之純化表明 Tsf 經恰當摺疊。大腸桿菌延伸因子 TS 之例示性核酸及對應胺基酸序列分別顯示為 SEQ ID NO: 22 及 23。

【0080】 在一實例中，藉由將 TS 融合至超高活性 Tn5 轉位酶之 N 端構建 TS-Tn5 融合蛋白。與突變 Tn5 轉位酶蛋白之 TS 融合之例示性胺基酸序列分別顯示為 SEQ ID NO: 25 及 26。SEQ ID NO: 24 對應於編碼 SEQ ID NO: 25 之 TS 突變蛋白融合物之核酸序列。

【0081】 儘管所有或部分 TS 可融合於 N 端或 C 端，熟習此項技術者應瞭解，可在 TS 序列與轉位酶之 N 端或 C 端之間插入連接序列。在一些具體實例中，連接序列之長度可為 1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50 或 50 個以上的胺基酸。在一些具體實例中，融合蛋白之轉位酶部分之一或多個胺基酸可缺失或經連接序列置換。在一些具體實例中，融合蛋白之轉位酶部分之第一甲硫胺酸可經兩個胺基酸，例如如 SEQ ID NO: 25 及 26 中所指示之 Gly-Thr 置換。

【0082】 TS-Tn5 融合構築體表現於大腸桿菌中且關於表現、溶解度及熱穩定性進行評估。TS 融合至 Tn5 之 N 端增加 Tn5 之溶解度。溶解度之增加可與轉位體複合物之增加之穩固性及蛋白聚集之減少相關。TS-Tn5 轉位體之熱穩定性相比於未融合 Tn5 轉位體之熱穩定性實質上經改善。在一實例中，相比於未融合 Tn5 對照物，Tn5 之誘導性聚集在 Tsf-Tn5 融合構築體中實質上減少。

【0083】 在一個應用中，TS-Tn5 融合蛋白用於構築定向 RNA-seq 庫（例如 TotalScript™ RNA-Seq 套組, Illumina）以在下一代定序平台（例如 Illumina GA 或 HiSeq 平台）上定序。

【0084】 在另一應用中，TS-Tn5 融合蛋白可用於標準化方法中。在另一應用中，TS 溶解標籤可用於其他經修飾（突變）Tn5 轉位酶之表現及純化。

【0085】 對 TS 融合標籤具有特異性之抗體可用於下拉方法中以捕獲轉位體標記之序列。在一實例中，TS 融合標籤抗體為家兔多株抗體。

【0086】 在另一應用中，TS-Tn5 轉位體及抗-TS 抗體可用於混合轉位體方法中。舉例而言，使用 Tsf-Tn5 轉位體及 Tn5 轉位體（亦即未經 TS 標記）進行轉位體反應。抗-TS 抗體尤其用於下拉 Tsf-Tn5 轉位體標記之序列。

【0087】 本發明提供一種包含經修飾 Tn5 轉位酶及識別 5-甲基胞嘧啶之蛋白域之轉位酶融合蛋白。5-甲基胞嘧啶結合蛋白域可例如用於將 Tn5 轉位體複合物靶向基因體之甲基化區。在一個應用中，5-甲基胞嘧啶結合域-Tn5 轉位體複合物可用於產生用於甲基化分析之甲基富集片段化及標記（標籤化）庫。

【0088】 在一些具體實例中，融合至轉位酶之多肽域包含蛋白 A 之抗體結合域。蛋白 A 為具有穩固摺疊特徵之相對較小、緻密分子。蛋白 A 之抗體結合域可例如用於將 Tn5 轉位體複合物靶向基因體之抗體結合區。舉例而言，對 5-甲基胞嘧啶具有特異性之抗體可用於結合至基因體之甲基化區且對其進行鑑別。基因體之抗體結合區可隨後使用蛋白 A-Tn5 融合轉位體複合物經靶向以用於片段化及標記（亦即標籤化）。

【0089】 在一些具體實例中，融合至轉位酶之多肽域包含純化標籤。如本文所用之術語「純化標籤」係指任何適合於純化或鑑別多肽之肽序列。純化標籤特異性結合至對於純化標籤具有親和力之另一部分。特異性結合至純化標籤之此類部分通常附接至基質或樹脂，諸如瓊脂糖珠粒。特異性結合至純化標籤之部分包括抗體、其他蛋白（例如蛋白 A 或抗生蛋白鏈菌素）、鎳或鈷離子或樹脂、生物素、直鏈澱粉、麥芽糖及環糊精。例示性純化標籤包括組胺酸（HIS）標籤（諸如六組胺酸肽），其將結合至金屬離子，諸如鎳或鈷離子。其他例示性純化標籤為 myc 標籤（EQKLISEEDL）、Strep 標籤（WSHPQFEK）、Flag 標籤（DYKDDDDK）及 V5 標籤（）。術語「純化標籤」亦包括「抗原決定基標籤」，亦即藉由抗體特定識別之肽序列。例示性抗原決定基標籤包括 FLAG 標籤，其藉由單株抗-FLAG 抗體特定識別。藉由抗-FLAG 抗體識別之肽序列由序列 DYKDDDDK 或其實質上一致的變異體組成。在一些具體實例中，融合至轉位酶之多肽域包含兩種或兩種以上標籤，諸如 SUMO 標籤及 STREP 標籤，如下文實施例 1 中所例示。術語「純化標籤」亦包括純化標籤之實質上一致的變異體。如本文所用之「實質上一致的變異體」係指純化標籤之衍生物或片段，其相比於初始純化標籤經修飾（例如經由胺基酸取代、缺失或插入），但保留特異性結合至特定識別純化標籤之部分的純化標籤特性。

【0090】 在一些具體實例中，融合至轉位酶之多肽域包含表現標籤。如本文所用之術語「表現標籤」係指可附接至第二多肽且應支持相關重組多肽之溶解度、穩定性及/或表現之任何肽或多肽。例示性表現標籤包括 Fc 標籤及 SUMO 標籤。大體上，任何肽、多肽或蛋白可用作表現標籤。

【0091】 其他有用參考文獻（例如關於細胞分離及培養（例如用於後續核酸分離））包括 Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, 第三版, Wiley-Liss, New York 及其中所引用之參考文獻；Payne 等人 (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons 公司 New York, N.Y.；Gamborg and Phillips (編) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York)及 Atlas 及 Parks (編) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.。

【0092】 編碼本文揭示之重組轉位酶之核酸亦為本文中呈現之具體實例之特徵。特定胺基酸可藉由多個密碼子編碼，且某些轉譯系統（例如原核或真核細胞）通常展現密碼子傾向，例如不同有機體通常偏好編碼相同胺基酸之若干同義密碼子中之一者。因此，本文中呈現之核酸視情況經「密碼子最佳化」，意謂核酸經合成以包括對於用於表現轉位酶之特定轉譯系統較佳之密碼子。舉例而言，當需要表現細菌細胞（或甚至特定細菌菌株）中之轉位酶時，核酸可經合成以包括最頻繁地發現於彼細菌細胞之基因體中之密碼子以有效表現轉位酶。當需要表現真核細胞中之轉位酶時，可採用類似策略，例如核酸可包括對於彼真核細胞較佳之密碼子。

【0093】 多種蛋白分離及偵測方法為已知的且可用於分離轉位酶，例如自表現本文中呈現之重組轉位酶之細胞之重組培養物。多種蛋白分離及偵測方法已在此項技術中為所熟知，包括例如以下文獻中所闡述之彼等方法：R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982)；Deutscher, *Methods in Enzymology* 第 182 卷: *Guide to Protein Purification*, Academic Press

公司 N.Y. (1990) ; Sandana (1997) Bioseparation of Proteins, Academic Press 公司 ; Bollag 等人 (1996) Protein Methods, 第 2 增補本 Wiley-Liss, NY ; Walker (1996) The Protein Protocols Handbook Humana Press, NJ , Harris 及 Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England ; Harris 及 Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England ; Scopes (1993) Protein Purification: Principles and Practice 第 3 增補本 Springer Verlag, NY ; Janson 及 Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, 第二版 Wiley-VCH, NY ; 及 Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ ; 及其中所引用之參考文獻。關於蛋白純化及偵測方法之其他細節可發現於 Satinder Ahuja 編, Handbook of Bioseparations, Academic Press (2000)中。

使用方法

【0094】 本文中呈現之經改變轉位酶可用於定序程序,諸如試管內轉位技術。簡言之,試管內轉位可藉由使轉位體複合物與目標 DNA 接觸起始。可易於適用於本發明之轉位酶之例示性轉位程序及系統描述於例如 WO 10/048605 ; US 2012/0301925 ; US 2013/0143774 中,該專利中之每一者以全文引用之方式併入本文中。

【0095】 舉例而言,在一些具體實例中,本文中呈現之轉位酶可用於自包含任何相關 dsDNA 之目標 DNA 產生標記 DNA 片段之庫(例如用作下一代定序或擴增模板)之方法中,該方法包含:藉由至少一個轉位酶及轉位子末端組成物在試管內轉位反應中培育目標 DNA,轉位酶藉由該轉位子末端組成物形成轉位複合物,該轉位子末端組成物包含(i)展現經轉移轉

位子末端序列及視情況存在之經轉移轉位子末端序列之另一序列 5'之轉移股，及 (ii) 在其中出現多個插入至目標 DNA 中之插入之條件下且持續究其而言足夠之時間展現與經轉移轉位子末端序列互補之序列之轉移股，該插入中之每一者導致包含轉移股或由其組成之第一標籤接合至目標 DNA 中之核苷酸之 5'末端，進而將目標 DNA 片段化且產生經退火 5'標記之 DNA 片段之群，其中之每一者在 5'末端上具有第一標籤；且隨後將 5'標記之 DNA 片段之 3'末端接合至第一標籤或第二標籤，進而產生標記 DNA 片段之庫(例如包含標記圓形 ssDNA 片段或 5'及 3'標記之 DNA 片段(或「二標記 DNA 片段」))。

【0096】 在一些具體實例中，轉位酶及轉位子末端組成物或用於試管內轉位反應中之轉位體組成物之量在每 50 微升反應物每 50 奈克目標 DNA 約 1 皮莫耳與約 25 皮莫耳之間。在任一本發明之方法之一些較佳具體實例中，轉位酶及轉位子末端組成物或用於試管內轉位反應中之轉位體組成物之量在每 50 微升反應物每 50 奈克目標 DNA 約 5 皮莫耳與約 50 皮莫耳之間。在任一本發明之方法之一些較佳具體實例中，其中轉位酶為超高活性 Tn5 轉位酶且轉位子末端組成物包含 MEDS 轉位子末端組成物或其中轉位體組成物包含該超高活性 Tn5 轉位酶及包含 MEDS 轉位子末端之轉位子末端組成物，該轉位酶及轉位子末端組成物或用於試管內轉位反應中之該轉位體組成物之量在每 50 微升反應物每 50 奈克目標 DNA 約 5 皮莫耳與約 25 皮莫耳之間。在任一本發明之方法之一些較佳具體實例中，其中轉位酶為超高活性 Tn5 轉位酶或 MuA 轉位酶，轉位酶及轉位子末端組成物或用於試管內轉位反應中之轉位體組成物之最終濃度為至少 250 nM；在一些其他具

體實例中，超高活性 Tn5 轉位酶或 MuA 轉位酶及其各別轉位子末端組成物或轉位體組成物之最終濃度為至少 500 nM。

【0097】 在一些具體實例中，本發明提供一種製備及富集基因體 DNA 庫以用於外顯子組定序之方法。在各種具體實例中，本發明方法使用經改變 Tn5 轉位酶，例如 TS-Tn5059 以用於製備基因體庫。在一個具體實例中，本發明方法提供製備具有減少的藉由經改變轉位酶（例如 TS-TN5059）之減少之插入型序列傾向驅動之傾向之基因體庫。使用 TS-Tn5059 之基因體 DNA 之標籤化提供跨越寬 GC/AT 範圍之基因體之更完整覆蓋度。

【0098】 在另一具體實例中，本發明提供一種使用具有增加之 DNA 輸入耐受度之經改變轉位酶製備基因體庫之方法。使用 TS-Tn5059 之基因體 DNA 之標籤化提供針對一定範圍之 DNA 輸入量的均一插入尺寸。在一些具體實例中，本發明提供一種外顯子組定序之方法。

標籤化反應條件

【0099】 此處呈現用於標籤化反應之反應條件及緩衝劑。在一些具體實例中，二價陽離子包含於標籤化反應緩衝劑中。在特定具體實例中，二價陽離子可例如為 Co^{2+} 、 Mn^{2+} 、 Mg^{2+} 、 Cd^{2+} 或 Ca^{2+} 。二價陽離子可以任何適合之鹽，諸如氯化物鹽，例如 CoCl_2 、 MnCl_2 、 MgCl_2 、乙酸鎂、 CdCl_2 或 CaCl_2 之形式包含。在特定具體實例中，標籤化緩衝劑包含 CoCl_2 ，如在下文實施例中所例示。如藉由實施例 5 中之實驗證據所表明，在標籤化緩衝劑調配物中添加 CoCl_2 出人意料地改善標籤化期間的序列傾向。

【0100】 在某些具體實例中，標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、

10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之二價陽離子濃度，或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。在一些具體實例中，標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之 CoCl_2 濃度或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。在一些具體實例中，標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之 MnCl_2 濃度或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。在一些具體實例中。標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之 MgCl_2 濃度或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。在一些具體實例中。標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之 CdCl_2

濃度或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。在一些具體實例中，標籤化緩衝劑可具有約 0.01 mM、0.02 mM、0.05 mM、0.1 mM、0.2 mM、0.5 mM、1 mM、2 mM、5 mM、8 mM、10 mM、12 mM、15 mM、20 mM、30 mM、40 mM、50 mM、60 mM、70 mM、80 mM、90 mM、100 mM 或大於其之 CaCl_2 濃度或在此等值中之任一者之間的範圍內，例如 0.01 mM 至 0.05 mM、0.02 mM 至 0.5 mM、8 mM 至 12 mM 等之二價陽離子濃度。

【0101】 在一些具體實例中，基因體 DNA 藉由轉位酶之片段化或標籤化反應可在 25°C 至 70°C、37°C 至 65°C、50°C 至 65°C 或 50°C 至 60°C 之溫度的範圍內進行。在一些具體實例中，基因體 DNA 藉由轉位酶之片段化或標籤化反應可在 37°C、40°C、45°C、50°C、51°C、52°C、53°C、54°C、55°C、56°C、57°C、58°C、59°C、60°C、61°C、62°C、63°C、64°C 或 65°C 下進行。

編碼經改變轉位酶之核酸

【0102】 本文進一步呈現編碼本文中呈現之經改變轉位酶之核酸分子。對於作為胺基酸序列已知且編碼轉位酶之野生型核苷酸序列較佳亦已知之轉位酶之突變型式之任何給定經改變轉位酶，可根據分子生物學之基本原理獲得編碼該突變體之核苷酸序列。舉例而言，倘若編碼 Tn5 轉位酶之野生型核苷酸序列已知，可使用標準基因密碼推斷編碼具有一或多個胺基酸取代之 Tn5 之任何給定突變型式之核苷酸序列。類似地，對於其他轉位酶之突變形式，可容易推導出核苷酸序列。具有所需核苷酸序列之核酸分子可接著使用此項技術中已知之標準分子生物學技術構築。

【0103】 根據本文中呈現之具體實例，指定之核酸不僅包括一致核酸，且亦包括任何較小鹼基變異，尤其包括在由於保守胺基酸取代中之簡併密碼產生同義密碼子（指定相同胺基酸殘基之不同密碼子）之情況下之取代。術語「核酸序列」亦包括關於鹼基變異給定之任何單股序列之互補序列。

【0104】 本文所述之核酸分子亦可有利地包括於適合之表現載體中以在適合之宿主中表現自其編碼之轉位酶蛋白。將選殖 DNA 併入至適合之表現載體中以用於該細胞之後續轉化及經轉化細胞之後續選擇為熟習此項技術者熟知，如 Sambrook 等人 (1989), *Molecular cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory 中所提供，該文獻以全文引用的方式併入本文中。

【0105】 該種表現載體包括具有根據本文中呈現之具體實例之核酸可操作地連接於能夠影響該 DNA 片段之表現之調節序列（諸如啟動子區）的載體。術語「可操作地連接」係指所描述之組分呈准許其以其預期方式起作用的關係的並接。此類載體可轉化至適合之宿主細胞中以提供根據本文中呈現之具體實例之蛋白之表現。

【0106】 核酸分子可編碼成熟蛋白或具有前序列之蛋白，包括編碼前蛋白上之前導序列，該前蛋白隨後由宿主細胞裂解以形成成熟蛋白。載體可例如為質體、病毒或噬菌體載體，其配備有複製起點，及視情況存在之用於表現該核苷酸之啟動子及視情況存在之啟動子之調節子。載體可含有一或多個選擇性標記物，諸如抗生素抗性基因。

【0107】 表現所需的調節元件包括啟動子序列以結合 RNA 聚合酶及

導引適當水準之轉錄起始以及轉譯起始序列以用於核糖體結合。舉例而言，細菌表現載體可包括啟動子（諸如 lac 啟動子）以及用於轉譯起始的 Shine-Dalgarno 序列及起始密碼子 AUG。類似地，真核表現載體可包括用於 RNA 聚合酶 II 之異源或同源啟動子、下游聚腺苷酸化信號、起始密碼子 AUG 及用於分離核糖體之終止密碼子。此類載體可商業上獲得或自所述序列藉由此項技術中熟知之方法組裝。

【0108】 藉由高級真核生物轉錄編碼轉位酶之 DNA 可藉由在載體中包括增強子序列而最佳化。增強子為作用於啟動子以增加轉錄水準之 DNA 之順式作用元件。除包括選擇性標記物以外，載體亦將一般包括複製起點。

實施例 1

一般分析方法及條件

【0109】 以下段落描述用於下文中呈現之實施例中之一般分析條件。

使用 TN5 進行 WGS 對人類 gDNA 進行標籤化

【0110】 此部分描述用於下文實施例中之標籤化分析，用於監測轉位酶之插入傾向。簡言之，在 55°C 下將 50 ng 人類基因體 DNA 與 5 µL TDE1 一起在 10 mM Tris-乙酸鹽（pH 7.6）、25 mM 乙酸鎂中培育 5 分鐘。隨後添加 1/5 反應體積之 125 mM HEPES（pH 7.5）、1 M NaCl、50 mM MgCl₂，接著添加 Tn5 轉位體至 100 nM 且在 30°C 下培育 60 分鐘。隨後清除反應物且如 Illumina's Nextera 方案中所述擴增且遞交用於使用 HiSeq 2000 儀器定序。

Tn5 轉位體組裝

【0111】 將 Tn5 與經黏接至 20 µM 之轉位子在室溫下在 25 mM HEPES（pH 7.6）、125 mM KCl、18.75 mM NaCl、0.375 mM EDTA、31.75%

甘油中培育 30 分鐘。

轉位子組裝

【0112】 藉由將反應物加熱至 94°C 且將其緩慢冷卻至室溫而在 10 mM TrisHCl (pH 7.5)、50 mM NaCl、1 mM EDTA 中將轉位子分開黏接至 40 μM。對於 Tn5-ME-A, Tn5 嵌合末端序列 (Mosaic End Sequence) A14 (Tn5MEA) 經黏接至 Tn5 非轉移序列 (NTS), 且對於 Tn5-ME-B, Tn5 嵌合末端序列 B15 (Tn5MEB) 經黏接至 Tn5 非轉移序列 (NTS)。此等序列如下所指示：

Tn5MEA : 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-3'

Tn5MEB : 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'

Tn5 NTS : 5'-CTGTCTCTTATACACATCT-3'

2. 轉位酶之選殖及表現

【0113】 此部分描述用於選殖及表現用於以下實施例中之各種轉位酶突變體之方法。

【0114】 使用標準定點突變誘發方法對編碼轉位酶之主鏈基因序列之基因進行突變誘發。對於進行之各突變，藉由定序選殖基因序列確認突變基因之恰當序列。

【0115】 將 Tn5 轉位酶基因選殖至經修飾 pET11a 質體中。經修飾質體含有來源於 pASK5plus 之純化標籤 StrepTag-II 及來源於 pET-SUMO 之 SUMO。使用 BL21 (DE3) pLysY 感受態細胞 (New England Biolabs) 進行表現。細胞在 25°C 下生長至 0.5 之 OD_{600nm}，接著使用 100 μM IPTG 誘導。在 18°C 下進行表現持續 19 h。使用微流化劑將細胞集結粒在蛋白酶抑制劑存在下溶解於 100 mM TrisHCl (pH 8.0)、1 M NaCl、1 mM EDTA 中。在溶解之

後，將細胞溶解物與去氧膽酸鹽一起培育 30 分鐘至 0.1% 之最終濃度。添加聚乙二亞胺至 0.5%，隨後在 30,000 $\times g$ 下離心 20 分鐘。收集上清液且與相等體積之飽和硫酸銨溶液混合，接著在冰上攪拌至少 1 h。隨後將溶液在 30,000 $\times g$ 下離心 20 分鐘且將集結粒再懸浮於 100 mM Tris (pH 8.0)、1 M NaCl、1 mM EDTA 及 1 mM DTT 中。隨後使用 AKTA 純化系統將再懸浮且過濾之溶液施加至 Streptactin 管柱。使用 100 mM Tris (pH 8.0)、1 mM EDTA、1 mM DTT、4 M NaCl，接著用 100 mM Tris (pH 7.5)、1 mM EDTA、1 mM DTT、100 mM NaCl 洗滌管柱。使用 100 mM Tris (pH 7.5)、1 mM EDTA、1 mM DTT、100 mM NaCl 及 5 mM 去硫生物素進行洗提。使用 100 mM Tris (pH 7.5)、100 mM NaCl、0.2 mM EDTA 及 2 mM DTT 將洗提液加載至肝素捕獲管柱上。在使用相同緩衝劑洗滌之後，使用 100 mM Tris (pH 7.5)、1 M NaCl、0.2 mM EDTA 及 2 mM DTT 以鹽梯度洗提 Tn5 變異體。將洗提份收集、合併、濃縮且添加甘油以產生 50% 最終濃度，隨後長期儲存於 -20°C 下。

3. 大腸桿菌基因體 DNA 中之插入傾向之 IVC 分析

【0116】 使用 IVC 曲線資料（強度相對於循環）進行插入序列傾向之分析。在定序使用各別 DNA 轉位酶創建之 DNA 庫之後產生此可用資料。簡言之，如上文所述進行突變誘發及表現。將轉位酶與上文所述之轉位子 A 及 B 一起培育，且與大腸桿菌基因體 DNA 一起培育以產生 DNA 庫。在運作 MiSeq Fast 化學方法之 Illumina 基因體分析儀系統 (Illumina 公司, San Diego, CA) 上根據製造商說明書對各庫定序至少 35 個循環。Illumina RTA 軟體用於在各循環產生鹼基響應及強度值。為了產生 IVC 曲線，將定序讀數與大腸桿菌參考基因體比對且將各循環之四種鹼基中之每一者之強度（出現率）

標繪為所有比對定序讀數之所有強度值之分數。

實施例 2

鑑別 Tn5 轉位酶突變體之插入傾向

【0117】 此實施例描述使用 IVC 曲線資料（強度相對於循環）分析插入序列傾向。此資料在定序使用各別 DNA 轉位酶創建之 DNA 庫之後可用。此分析僅需要一些定序讀數（20k-40k）以得到穩定結果且可在用於表現各別 Tn5 轉位酶變異體之大腸桿菌細胞溶解物中進行，且適合於 HTS（高通量篩選）目的。

【0118】 各種單一胺基酸取代 Tn5 變異體之代表性結果闡述於圖 2-7 中。舉例而言，顯示之變異體表明藉由位置 248 處之取代之對稱性損失（圖 2）、藉由位置 119 處之取代之扁平 IVC 曲線（圖 3）、藉由位置 125 處之取代或位置 248 之後的插入之 IVC 曲線之後半部中之 IVC 減少（圖 4-5）及藉由在位置 K120 處使用不同芳族胺基酸的 9 bp 至 10 bp 之複製增加（圖 6，顯示自 1 及第 9 至 1 及第 10 bp 之對稱性變化）。此等結果表明指定突變相比於 wt 對照組提供改善之插入傾向。

實施例 3

細菌性 gDNA 之全基因體定序

【0119】 進行此等實驗以關於 a) 估計之庫大小/多樣性及 b) AT/GC 丟失比較各種轉位酶突變體。藉由經純化及活性標準化 Tn5 轉位酶變異體進行此等實驗。此等實驗需要 500k-1M 定序讀數/實驗。

【0120】 藉由使用指定純化 Tn5 轉位酶變異體對蠟狀芽孢桿菌（*B. cereus*）gDNA 進行標籤化獲得結果。酶根據活性標準化且設定成匹配與

Nextera™ 套組一起銷售之商業 TDE1 酶之活性。

【0121】 指定為 Tn5001 之突變體與 SEQ ID NO: 11 具有相同胺基酸序列且充當「wt」對照組。如上表中所示，指定為 Tn5058、Tn5059 及 Tn5061 之突變體分別與 SEQ ID NO: 18、19 及 20 具有相同胺基酸序列。一式三份地進行實驗，資料顯示收集之資料之平均值及標準差。「估計之庫大小」在不使用光學重複的情況下計算，提供可再現結果。

【0122】 如圖 8A 中所示，相比於 Tn5001，Tn5058 及 Tn5059 存在 AT 丟失之顯著減少，同時保持 GC 丟失較低。類似地，如圖 8B 中所示，存在庫尺寸的 1.7x (Tn5058) 及 2.2x (Tn5059) 之顯著增加。此等結果表明相比於野生型轉位酶，突變體 Tn5058 及 Tn5059 極大地改善序列插入傾向，導致在下文實施例 4 中所述之其他實驗。

實施例 4

人類 gDNA 之 Nextera 快速捕獲富集實驗

【0123】 在與上文實施例 3 中所述相同之純化及活性標準化 Tn5 轉位酶變異體之情況下進行以下實驗。此等實驗通常需要 40M-100M 定序讀數/實驗，且分析定序資料以比較 a) 多樣性、b) 富集、c) 覆蓋度、d) 覆蓋度均一性、e) 罰分。

【0124】 根據製造商說明書使用 Nextera 快速捕獲外顯子組 (Illumina™) CEX 集區捕獲探針一式三份地進行捕獲。如圖 9A 中所示，相比於 wt 對照 Tn5001，指定突變體產生覆蓋度均一性之顯著改善。另外，如圖 9B 中指示，藉由目標覆蓋度之 10x 及 20x 之測試突變體產生統計顯著改善，儘管平均目標覆蓋度較低。

【0125】 如圖 10A 中所示，指定突變體產生獨特讀數數目及混合選擇庫大小之增加。同樣，如圖 10B 中所示，突變體產生相比於 Tn5001 較低之罰分，其為達到 10x、20x 或 30x 深度之覆蓋度所需的倍數較大定序。此等結果表明測試突變體相比於對照組提供更大插入傾向及更均一覆蓋度。

實施例 5

標籤化緩衝劑組成對 Tn5 活性之影響

【0126】 進行以下實驗以特性化 Tn5 標籤化緩衝劑組成及反應條件對庫輸出及定序度量值之影響。

【0127】 為了評估標籤化緩衝劑組成及反應條件對庫輸出及定序度量值之影響，使用蠟狀芽孢桿菌基因體 DNA 構築 Tn5 標籤化 DNA 庫。兩種不同 Tn5 轉位酶用於構築標籤化庫，亦即突變 Tn5 (「TS-Tn5059」) 及對照超高活性 Tn5 (「TS-Tn5」)。TS 為用於純化 Tn5 及 Tn5059 蛋白之融合標籤。Tn5059 具有 4 個就超高活性 Tn5 胺基酸序列 (SEQ ID NO: 11) 而言之額外突變體 K212R、P214R、G251R 及 A338V。TS-Tn5059 在 Tn5059 之 N 端包含 TS 標籤。在一些具體實例中，TS-標籤之 C 端可藉由連接子融合至 Tn5059 之 N 端，其替代第一甲硫胺酸殘基。在一些具體實例中，連接子為 Gly-Thr。

【0128】 TS-Tn5059 以 10、40 及 80 nM 之最終濃度使用。TS-Tn5 以 4、15 及 30 nM 之最終濃度使用。TS-Tn5059 及 TS-Tn5 之酶濃度經標準化 (使用標準緩衝劑調配物) 以提供大致相同之標籤化活性位準，亦即 10、40 及 80 nM 處之 TS-Tn5059 分別與 4、15 及 30 nM 處之 Tn5 具有大致相同活性位準。使用 25 ng 輸入量之蠟狀芽孢桿菌基因體 DNA 製備各標籤化庫。蠟狀

芽孢桿菌之基因體含量為約 40% GC 及約 60% AT。

【0129】 標籤化緩衝劑製備為 2×調配物。2×調配物為如下：標準緩衝劑（TD；20 mM Tris 乙酸鹽，pH 7.6、10 mM 乙酸鎂及 20%二甲基甲醯胺（DMF）；鈷緩衝劑（Co；20 mM Tris 乙酸鹽，pH 7.6 及 20 mM CoCl_2 ）；鈷+DMSO 緩衝劑（Co-DMSO；20 mM Tris 乙酸鹽，pH 7.6、20 mM CoCl_2 及 20%二甲亞砜（DMSO））；高分子量緩衝劑（HMW；20 mM Tris 乙酸鹽，pH 7.6 及 10 mM 乙酸鎂）；NF2 緩衝劑（NF2；20 mM Tris 乙酸鹽，pH 7.6、20 mM CoCl_2 及 20% DMF）。每日製備新的包括 CoCl_2 之標籤化緩衝劑。對於各庫，藉由以 50 μL 之總反應體積混合 20 μL 蠟狀芽孢桿菌基因體 DNA（25 ng）、25 μL 2×標籤化緩衝劑及 5 μL 酶（10× Ts-Tn5059 或 10× Ts-Tn5）進行標籤化反應。在 55°C 下培育反應物 5 分鐘。在標籤化反應之後，根據標準 Nextera™ 樣品製備方案加工樣品。使用 Illumina's SBS(合成定序法)化學方法在 MiSeq 裝置上對庫定序。定序操作為使用 V2 MiSeq 套組之 2×71 循環。在生物分析儀上評估各庫中之片段尺寸分佈。

【0130】 圖 11 顯示使用不同標籤化緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之獨特分子數之柱狀圖 100。庫中之獨特分子數為庫之多樣性（複雜度）之示度。圖上之各條表示標籤化庫。重複實驗三次（ $n=3$ ）。藉由「酶-酶濃度-DNA 輸入」指示對照庫（亦即使用標準標籤化緩衝劑製備之庫）。舉例而言，柱狀圖 100 中之第一條標記為「TS-Tn5059-10nM-25ng」且表示使用標準緩衝劑調配物中 10 nM 之最終濃度之 TS-Tn5059 及 25 ng 之輸入 DNA 製備之對照庫。使用經修飾標籤化緩衝劑調配物製備之庫由「酶-酶濃度-緩衝劑添加劑-DNA 輸入」指示。舉例

而言，柱狀圖 100 中之第四條標記為「TS-Tn5059-10nM-Co-25ng」且表示使用包括 10 mM CoCl_2 之經修飾標籤化緩衝劑中 10 nM 之最終濃度之 TS-Tn5059 製備之庫。資料顯示使用包括 10 mM CoCl_2 之標籤化緩衝劑（亦即 Co、Co-DMSO 及 NF2 緩衝劑）製備之 TS-Tn5059 及 TS-Tn5 標籤化庫具有相比於不添加 CoCl_2 之緩衝劑（亦即標準緩衝劑或 HMW）中製備之庫較高之平均多樣性。

【0131】 圖 12 顯示使用不同標籤化緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之 GC 丟失%之柱狀圖 200。如圖 11 中所述來表示使用經修飾標籤化緩衝劑製備之對照庫及庫。GC 丟失可定義為基因體中自標籤化庫除去（不存在）之 GC 富集區之百分比。資料顯示對於使用標準標籤化緩衝劑製備之對照 TS-Tn5059 及 TS-Tn5 庫，GC 丟失之百分比相對較低。資料亦顯示使用包括 10 mM CoCl_2 之標籤化緩衝劑（亦即 Co、Co-DMSO 及 NF2 緩衝劑）製備之 TS-Tn5059 及 TS-Tn5 標籤化庫具有相比於在不添加 CoCl_2 之緩衝劑（亦即標準緩衝劑或 HMW）中製備之庫較高之 GC 丟失%（亦即達至約 6%）。使用含 Co 緩衝劑製備之庫中之 GC 丟失之增加藉由增加 TS-Tn5059 及 TS-Tn5 之濃度改善。舉例而言，TS-Tn5059-10nm-Co-25ng 庫中之 GC 丟失%相比於 TS-Tn5059-10nm-25ng 對照庫相對較高。隨著 TS-Tn5059 之濃度增加至 40 nM（亦即 TS-Tn5059-40nm-Co-25ng）及 80 nM（亦即 TS-Tn5059-80nm-Co-25ng），GC 丟失%減小。

【0132】 圖 13 顯示使用不同標籤化緩衝劑製備之 TS-Tn5059 及 TS-Tn5 標籤化 DNA 庫中之 AT 丟失%之柱狀圖 300。如圖 11 中所述來表示使用經修飾標籤化緩衝劑製備之對照庫及庫。AT 丟失可定義為基因體中自

標籤化庫除去（不存在）之 AT 富集區之百分比。資料顯示對於使用標準標籤化緩衝劑製備之對照 TS-Tn5059 及 TS-Tn5 庫，觀測到一定量（亦即分別地，約 1%至約 3%及約 7%至約 3%）之 AT 丟失。資料亦顯示使用低酶濃度（亦即 10 nM）及包括 10 mM CoCl₂之標籤化緩衝劑（亦即 Co、Co-DMSO 及 NF2 緩衝劑）製備之 TS-Tn5059 標籤化庫相比於在不添加 CoCl₂之緩衝劑（亦即標準緩衝劑或 HMW）中製備之庫具有較低 AT 丟失%。類似地，使用包括 10 mM CoCl₂之標籤化緩衝劑（亦即 Co、Co-DMSO 及 NF2 緩衝劑）製備之 TS-Tn5 庫具有相比於在不添加 CoCl₂之緩衝劑（亦即標準緩衝劑或 HMW）中製備之庫較低之 AT 丟失%。

【0133】 現參看圖 12 及 13，在標籤化緩衝劑（亦即 Co、Co-DMSO 及 NF2 緩衝劑）中添加 CoCl₂（10 nM）可「翻轉」標籤化庫中之 GC 及 AT 丟失%。舉例而言，TS-Tn5059-10nm-Co-25ng 庫中之 GC 丟失%（圖 12）相比於 TS-Tn5059-10nm-25ng 對照庫相對較高；而 TS-Tn5059-10nm-Co-25ng 庫中之 AT 丟失%（圖 13）相比於 TS-Tn5059-10nm-25ng 對照庫相對較低（或無）。

【0134】 圖 14 顯示使用標準緩衝劑（TD）及鈷緩衝劑（Co）調配物製備之 TS-Tn5059 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖 400。曲線圖 400 顯示為 Ts-Tn5059-10nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 410、為 Ts-Tn5059-40nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 415、為 Ts-Tn5059-80nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 420、為 Ts-Tn5059-10nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 425、為 Ts-Tn5059-10nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 430 及為

Ts-Tn5059-80nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 435。曲線圖 400 亦顯示為鹼基對 (bp) 中之 DNA 片段尺寸之標準梯的曲線 440。梯中之片段尺寸 (從左向右) 顯示於表 2 中。

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

【0135】 資料顯示將用於標籤化反應中之 TS-Tn5059 之濃度自 10 nM 增加至 40 nM 及 80 nM 將片段尺寸分佈移位至較小片段尺寸。片段尺寸分佈之移位在使用標準緩衝劑 (TD) 調配物製備之庫中更明顯。舉例而言，在使用鈷緩衝劑 (Co) 調配物製備之庫中之片段尺寸分佈在使用 10 nM TS-Tn5059 製備之庫中為約 3,000 bp (曲線 410) 且在使用 40 及 80 nM TS-Tn5059 製備之庫中為約 1,000 至約 2,000 bp (分別地，曲線 415 及 420)。對於使用標準緩衝劑 (TD) 調配物及 80 nM TS-Tn5059 (曲線 435) 製備之庫，片段尺寸分佈為約 200 bp 至約 1,000 bp。

【0136】 圖 15 顯示使用鈷-DMSO (Co-DMSO)、NF2 及 HMW 緩衝劑調配物製備之 TS-Tn5059 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖

500。曲線圖 500 顯示為 Ts-Tn5059-10nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 510、為 Ts-Tn5059-10nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 515、為 Ts-Tn5059-10nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 520、為 Ts-Tn5059-40nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 525、為 Ts-Tn5059-40nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 530、為 Ts-Tn5059-40nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 535、為 Ts-Tn5059-80nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 540、為 Ts-Tn5059-80nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 545 及為 Ts-Tn5059-80nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 550。曲線圖 500 亦顯示圖 14 之曲線圖 400 之曲線 440，其為鹼基對 (bp) 中之 DNA 片段尺寸之標準梯。

【0137】 資料顯示一般而言，將用於標籤化反應中之 TS-Tn5059 之濃度自 10 nM 增加至 40 nM 及 80 nM 將片段尺寸分佈移位至較小片段尺寸。相比於使用 Co-DMSO 製備之庫 (例如曲線 510 及 525)，片段尺寸分佈之移位在使用不包括 CoCl_2 之 HMW 緩衝劑製備之庫 (例如曲線 520 及 535) 中更明顯。

【0138】 圖 16 顯示使用標準緩衝劑調配物 (TD) 及鈷緩衝劑 (Co) 製備之 TS-Tn5 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖 600。曲線圖 600 顯示為 Ts-Tn5-4nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 610、為 Ts-Tn5-15nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 615、為 Ts-Tn5-30nM-Co-25 ng 庫中之片段尺寸分佈之曲線的曲線 620、為 Ts-Tn5-4nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 625、為

Ts-Tn5-15nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 630 及為 Ts-Tn5-30nM-TD-25 ng 庫中之片段尺寸分佈之曲線的曲線 635。曲線圖 600 亦顯示圖 14 之曲線圖 400 之曲線 440，其為鹼基對 (bp) 中之 DNA 片段尺寸之標準梯。

【0139】 資料顯示將用於標籤化反應中之 TS-Tn5 之濃度自 4 nM 增加至 15 nM 及 30 nM 將片段尺寸分佈移位至較小片段尺寸。片段尺寸分佈之移位在使用標準緩衝劑 (TD) 調配物製備之庫中更明顯。此觀測結果與圖 14 之 TS-Tn5059 庫中之片段尺寸分佈類似。

【0140】 圖 17 顯示使用鈷-DMSO (Co-DMSO)、NF2 及 HMW 緩衝劑調配物製備之 TS-Tn5 庫中之片段尺寸分佈之生物分析儀跡線之曲線圖 700。曲線圖 700 顯示為 Ts-Tn5-4nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 710、為 Ts-Tn5-4nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 715、為 Ts-Tn5-4nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 720、為 Ts-Tn5-15nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 725、為 Ts-Tn5-15nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 730、為 Ts-Tn5-15nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 735、為 Ts-Tn5-30nM-Co-DMSO-25 ng 庫中之片段尺寸分佈之曲線的曲線 740、為 Ts-Tn5-30nM-NF2-25 ng 庫中之片段尺寸分佈之曲線的曲線 745 及為 Ts-Tn5-30nM-HMW-25 ng 庫中之片段尺寸分佈之曲線的曲線 750。曲線圖 700 亦顯示圖 14 之曲線圖 400 之曲線 440，其為鹼基對 (bp) 中之 DNA 片段尺寸之標準梯。

【0141】 資料顯示將用於標籤化反應中之 TS-Tn5 之濃度自 4 nM 增加

至 15 nM 及 30 nM 將片段尺寸分佈移位至較小片段尺寸。片段尺寸分佈之移位在使用標準緩衝劑 (TD) 調配物製備之庫中更明顯。此觀測結果與圖 15 之 TS-Tn5059 庫中之片段尺寸分佈類似。

【0142】 一般而言，現參看圖 14 至 17，使用包括 10 nM CoCl_2 之標籤化緩衝劑 (例如 Co、Co-DMSO 及 NF2 緩衝劑) 製備之 TS-Tn5059 及 TS-Tn5 庫中之片段尺寸大於使用無 CoCl_2 之標籤化緩衝劑 (亦即 TD 及 HMW 緩衝劑) 製備之 TS-Tn5059 及 TS-Tn5 庫中之片段尺寸。

【0143】 圖 18A、18B、18C 及 18D 分別顯示 TS-Tn5 庫中之序列含量之傾向圖 800、TS-TN5-Co 庫中之序列含量之傾向圖 830、TS-Tn5-Co-DMSO 庫中之序列含量之傾向圖 840 及 TS-Tn5-NF2 庫中之序列含量之傾向圖 850。傾向圖 (或強度相對於循環數 (IVC) 圖) 描繪觀測之鹼基 (A、C、G 或 T) 隨 SBS 循環數變化之比且顯示在標籤化期間 Tn5 具有之較佳序列環境。

【0144】 傾向圖 800、830、840 及 850 各顯示為依據循環數之 A 含量之曲線的曲線 810、依據循環數之 C 含量之曲線的曲線 815、為依據循環數之 G 含量之曲線的曲線 820 及為依據循環數之 T 含量之曲線的曲線 825。舉例而言，在圖 18A 之 TS-TN5 庫中，表示鹼基 G 之曲線 820 顯示約 38% 在第 1 循環觀測之鹼基為 G；表示鹼基 T 之曲線 825 顯示約 15% 在第 1 循環觀測之鹼基為 T 等。

【0145】 參看圖 18A，資料顯示經 TS-Tn5 庫中之 SBS 之約首先 15 個循環觀測到 Tn5 序列傾向，該庫使用標準標籤化緩衝劑調配物製備。在約 15 個循環之後，序列傾向逐漸減小且 A、T、C 及 G 含量反映預期基因體組成。對於蠟狀芽孢桿菌，基因體為約 40% GC 及約 60% AT，其表示於

傾向圖中之約第 16 或 17 循環至第 35 循環，其中曲線 810（亦即 A）及曲線 825（亦即 T）會合於約 30%處（A+T 約 60%）；且曲線 815（亦即 C）曲線 820（亦即 G）會合於約 20%處（C+G 約 40%）。

【0146】 參看圖 18B、18C 及 18D，資料亦顯示經 TS-Tn5-Co、TS-Tn5-Co-DMSO 及 Ts-Tn5-NF2 庫中之 SBS 之約首先 15 個循環觀測到 Tn5 序列傾向，該庫為使用包括 CoCl_2 之標籤化緩衝劑製備之庫。同樣，在約 15 循環之後，序列傾向逐漸減小且 A、T、C 及 G 含量反映如參看圖 18A 所描述之預期基因體組成。然而，在 TS-Tn5-Co、TS-Tn5-Co-DMSO 及 Ts-Tn5-NF2 庫中，曲線 810（亦即 A）及曲線 825（亦即 T）在約第 10 循環至第 15 循環開始朝向預期基因體組成移位；且曲線 815（亦即 C）及曲線 820（亦即 G）在約第 10 循環至第 15 循環開始朝向預期基因體組成移位。另外，當相比於圖 18A 時，循環 2-8 之間的傾向減小。資料顯示在標籤化緩衝劑調配物中添加 CoCl_2 改善標籤化期間的 Tn5 序列傾向。

【0147】 圖 19A、19B、19C 及 19D 分別顯示 TS-Tn5059 庫中之序列含量之傾向圖 900、TS-TN5059-Co 庫中之序列含量之傾向圖 930、TS-Tn5059-Co-DMSO 庫中之序列含量之傾向圖 940 及 TS-Tn5059-NF2 庫中之序列含量之傾向圖 950。傾向圖 900、930、940 及 950 各顯示為依據循環數之 A 含量之曲線的曲線 910、依據循環數之 C 含量之曲線的曲線 915、為依據循環數之 G 含量之曲線的曲線 920 及為依據循環數之 T 含量之曲線的曲線 925。

【0148】 參看圖 19A，資料顯示經 TS-Tn5059 標籤化庫中之 SBS 之約首先 15 個循環觀測到 Tn5059 序列傾向。在約 15 循環之後，序列傾向減小

且 A、T、C 及 G 含量反映如參看圖 18A 所描述之預期基因體組成。然而，突變體 Tn5059 顯示相比於圖 18A 中顯示之 Tn5 序列傾向減小之序列傾向。在 TS-Tn5059 庫中，曲線 910（亦即 A）及曲線 925（亦即 T）在約第 10 循環至第 15 循環開始朝向預期基因體組成移位；且曲線 915（亦即 C）及曲線 920（亦即 G）在約第 10 循環至第 15 循環開始朝向預期基因體組成移位。

【0149】 參看圖 19B、19C 及 19D，資料亦顯示經 TS-Tn5059-Co、TS-Tn5059-Co-DMSO 及 Ts-Tn5059-NF2 庫中之 SBS 之約首先 15 個循環觀測到 Tn5059 序列傾向，該庫為使用包括 CoCl_2 之標籤化緩衝劑製備之庫。同樣，在約 15 循環之後，序列傾向逐漸減小且 A、T、C 及 G 含量反映如參看圖 18A 描述之預期基因體組成。然而，在 TS-Tn5059-Co 庫中，曲線 910（亦即 A）及曲線 925（亦即 T）在約第 5 循環開始朝向預期基因體組成移位；且曲線 915（亦即 C）及曲線 920（亦即 G）在約第 5 循環開始朝向預期基因體組成移位。在 TS-Tn5059-Co-DMSO 及 Ts-Tn5059-NF2 庫中，曲線 910（亦即 A）及曲線 925（亦即 T）在第 5 循環之前開始朝向預期基因體組成移位；且曲線 915（亦即 C）及曲線 920（亦即 G）在第 5 循環之前開始朝向預期基因體組成移位。

實施例 6

標籤化緩衝劑組成物對 Mos1 活性之影響

【0150】 進行以下實驗以特性化 Tn5 標籤化緩衝劑組成及反應條件對庫輸出及定序度量值之影響。

【0151】 使用蠟狀芽孢桿菌基因體 DNA 構築 Mos1 標籤化 DNA 庫。用於構築標籤化庫之 Mos1 轉位酶為 MBP-Mos1 融合蛋白。麥芽糖結合蛋白

(MBP) 為用於純化 Mos1 蛋白之融合標籤。MBP-Mos1 以 100 μ M 之最終濃度使用。使用 50 ng 輸入量之蠟狀芽孢桿菌基因體 DNA 製備各標籤化庫。

【0152】 標籤化緩衝劑製備為 2 $\times$ 調配物。2 $\times$ 調配物為如下：標準緩衝劑 (TD；20 mM Tris 乙酸鹽，pH 7.6、10 mM 乙酸鎂及 20% 二甲基甲醯胺 (DMF)；TD+NaCl (TD-NaCl；20 mM Tris 乙酸鹽，pH 7.6、10 mM 乙酸鎂、20% DMF 及 200 mM NaCl)；高分子量緩衝劑 (HMW；20 mM Tris 乙酸鹽，pH 7.6 及 10mM 乙酸鎂)；HEPES (50 mM HEPES，pH 7.6、10 mM 乙酸鎂、20% DMF)；HEPES-DMSO (50 mM HEPES pH 7.6、10 mM 乙酸鎂及 20% DMSO)；HEPES-DMSO-Co (50 mM HEPES，pH 7.6、20% DMSO 及 20 mM CoCl_2) 及 HEPES-DMSO-Mn (50 mM HEPES，pH 7.6、20% DMSO 及 20 mM 錳 (Mn))。每日製備新的包括 CoCl_2 之標籤化緩衝劑。

【0153】 對於各庫，藉由以 50 μ L 之總反應體積混合 20 μ L 蠟狀芽孢桿菌基因體 DNA (50 ng)、25 μ L 2 $\times$ 標籤化緩衝劑及 5 μ L 酶 (10 $\times$ MBP-Mos1) 進行標籤化反應。在 30 $^{\circ}$ C 下培育反應物 60 分鐘。在標籤化反應之後，根據標準 Nextera™ 樣品製備方案加工樣品。使用 Illumina's SBS (合成定序法) 化學方法在 MiSeq 裝置上對庫定序。定序操作為 2 $\times$ 71 循環。

【0154】 圖 20 顯示在使用不同標籤化緩衝劑製備之 MBP-Mos1 標籤化庫中之平均讀數總數及平均多樣性之柱狀圖 1000。讀數總數為來自流動細胞之讀數的總數。多樣性為庫中之獨特分子數且用作庫複雜度之示度。圖上之每一對條表示標籤化庫。重複實驗三次 (n=3)。首先兩個圖條 EZTn5-std-bcereus 及 NexteraV2-30C 分別為使用 Tn5 及標準緩衝劑調配物在 55 $^{\circ}$ C 及 30 $^{\circ}$ C 下製備之比較庫。用於標籤化反應的使用 MBP-Mos1 製備之庫

藉由「酶-酶濃度-緩衝劑」指示。舉例而言，第三對圖條標記為「MBPMos1-100 μ M-TD」且指示使用標準標籤化緩衝劑（TD）中 100 μ M 之最終濃度下之 MBP-Mos1 製備之庫。資料顯示不同緩衝劑對藉由 Mos1 標籤化在相對相同數目或定序讀數下製備之庫之多樣性的影響。特定言之，HEPES-DMSO-Mn 緩衝劑幫助增加庫之多樣性。

【0155】 圖 21 顯示 MBP-Mos1 標籤化庫中之 GC 及 AT 丟失之柱狀圖 1100。GC 及 AT 丟失可分別定義為基因體中自標籤化庫除去（不存在）之 GC 富集區及 AT 富集區之百分比。庫如圖 20 中所述來表示。資料顯示使用 EZTn5 及 NexteraV2（亦即 Tn5 轉位酶）製備之庫基本上無 GC 丟失，但分別自標籤化庫除去約 7%及約 5% AT 富集區。使用 MBP-Mos1 及標準標籤化緩衝劑（MBPMos1-100 μ M-TD）製備之庫基本上無 AT 丟失，但自標籤化庫除去約 2%或 2%以下之 GC 富集區。MBP-Mos1 標籤化庫中之 GC 丟失%由標籤化緩衝劑之組成實現。GC 丟失%在使用 HMW、HEPES、HEPES-DMSO、HEPES-DMSO-Co 及 HEPES-DMSO-Mn 緩衝劑製備之 MBP-Mos1 標籤化庫中增加。

【0156】 圖 22A、22B、22C 及 22D 分別顯示 Mos1-HEPES 庫中之序列含量之傾向圖 1200、Mos1-HEPES-DMSO 庫中之序列含量之傾向圖 1230、Mos1-HEPES-DMSO-Co 庫中之序列含量之傾向圖 1240 及 Mos1-HEPES-DMSO-Mn 庫中之序列含量之傾向圖 1250。傾向圖 1200、1230、1240 及 1250 各顯示為依據循環數之 A 含量之曲線的曲線 1210、依據循環數之 C 含量之曲線的曲線 1215、為依據循環數之 G 含量之曲線的曲線 1220 及為依據循環數之 T 含量之曲線的曲線 1225。

【0157】 參看圖 22A 及 22B，資料顯示經 Mos1-HEPES 及 Mos1-HEPES-DMSO 標籤化庫中之 SBS 之首先少數循環觀測到 Mos1 序列傾向。在第一 SBS 循環中，偵測之 T 在整個流動細胞中為約 100%。在第二 SBS 循環中，偵測之 A 在整個流動細胞中為約 100%。在約 4 個循環之後，序列傾向減小且 A、T、C 及 G 含量反映預期基因體組成。

【0158】 參看圖 22C 及 22D，資料亦顯示經 Mos1-HEPES-DMSO-Co 及 Mos1-HEPES-DMSO-Mn 標籤化庫中之 SBS 之首先少數循環觀測到 Mos1 序列傾向，該庫為使用分別藉由鈷（Co）或錳（Mn）置換鎂（Mg）之標籤化緩衝劑製備之庫。同樣，在約 4 個循環之後，序列傾向減小且 A、T、C 及 G 含量反映預期基因體組成。然而，在 Mos1-HEPES-DMSO-Co 及 Mos1-HEPES-DMSO-Mn 庫中，曲線 1210（亦即 A）及曲線 1225（亦即 T）在第 1 循環及第 2 循環觀測到朝向預期基因體組成之移位。朝向預期基因體組成之移位 Mos1-HEPES-DMSO-Mn 庫中更明顯。

實施例 7

TS-Tn5059 庫製備及外顯子組富集方案

【0159】 在一個具體實例中，本發明方法提供一種製備及富集基於 Tn5 轉位體之外顯子組庫之流線型工作流。

【0160】 圖 23 說明製備及富集用於外顯子組定序之基因體 DNA 庫之方法 1260 之實施例之流程圖。方法 1260 使用 TS-Tn5059 轉位體及對於當前 Nextera®快速捕獲方案之某些方法步驟之修改以提供針對一定範圍之 DNA 輸入量及定序度量值的經改善庫產率。舉例而言，方法 1260 使用「雙面」固相可逆固定（SPRI）方案（Agencourt AMPure XP 珠粒；Beckman Coulter

公司) 以純化標籤化 DNA 且在 PCR 擴增之前提供第一 DNA 片段尺寸選擇步驟及第二 DNA 片段尺寸選擇步驟。在另一實施例中，預濃縮方法用於在外顯子組富集之前濃縮標籤化 DNA 庫。方法 1260 包括 (但不限於) 以下步驟。

【0161】 在步驟 1270 中，基因體 DNA 藉由轉位體標籤化 (標記及片段化)。轉位體同時將基因體 DNA 片段化且將接附子序列添加至末端，允許藉由 PCR 之後續擴增。在一個實施例中，轉位體為 TS-Tn5059。在完成標籤化反應時，將標籤化終止緩衝劑添加至反應中。標籤化終止緩衝劑可經修飾以確保 TS-Tn5059 轉位體複合物自標籤化 DNA 之足夠變性 (例如終止緩衝劑中之 SDS 之濃度自 0.1% 增加至 1.0% SDS 以及高熱量)。

【0162】 在步驟 1275 中，進行第一清除以純化來自轉位體之標籤化 DNA 且提供第一 DNA 片段尺寸選擇步驟。DNA 片段尺寸可藉由改變 SPRI 珠粒與 DNA 之體積比 (例如 $1 \times \text{SPRI} = 1:1$ 體積 $\text{SPRI}:\text{DNA}$) 選擇。舉例而言，在第一尺寸選擇中，選擇 SPRI 珠粒與 DNA 之體積比以結合大於某一尺寸之 DNA 片段 (亦即自樣品移除較大 DNA 片段)，同時將小於某一尺寸之 DNA 片段保持於上清液中。將其中具有尺寸選擇 DNA 片段之上清液轉移至清潔反應容器中以用於後續處理。可丟棄在上面具有較大 DNA 片段之 SPRI 珠粒。在一個具體實例中，SPRI 珠粒之濃度可在 $0.8 \times$ 至 $1.5 \times$ 之範圍內變化。在一個具體實例中，SPRI 珠粒之濃度為 $0.8 \times$ 。

【0163】 在步驟 1280 中，進行第二清除以進一步選擇某一尺寸範圍內之 DNA 片段。舉例而言，選擇 SPRI 珠粒與 DNA 之體積比以結合大於某一尺寸之 DNA 片段 (亦即所需尺寸範圍內之 DNA 片段結合至 SPRI 珠粒)。

較小 DNA 片段保持於上清液中且丟棄。隨後自 SPRI 珠粒洗提結合 DNA 片段以用於後續處理。

【0164】 在視情況存在之步驟 1285 中，測定 DNA 片段尺寸分佈。DNA 片段尺寸分佈例如使用生物分析儀測定。

【0165】 在步驟 1290 中，純化標籤化 DNA 經由有限循環 PCR 程式擴增。PCR 步驟亦添加指數 1 (i7) 及指數 2 (i5) 及定序，以及後續叢集產生及定序所需的常用接附子 (P5 及 P7)。由於雙面 SPRI 方法 (亦即步驟 1275 及 1280) 用於選擇所需 DNA 片段尺寸範圍，僅所需尺寸範圍內之標籤化 DNA 片段可用於 PCR 擴增。因此，庫產率顯著增加且後續定序度量值 (例如讀數富集%) 經提高。

【0166】 在步驟 1295 中，使用基於珠粒之純化方法純化擴增標籤化 DNA 庫。

【0167】 在視情況存在之步驟 1300 中，測定 PCR 後的 DNA 片段尺寸分佈。DNA 片段尺寸分佈例如使用生物分析儀測定。

【0168】 在步驟 1310 中，標籤化 DNA 庫在用於外顯子組富集之後續雜交之前預濃縮。舉例而言，標籤化 DNA 庫自約 50 μ l 預濃縮至約 10 μ L。由於標籤化 DNA 庫經預濃縮，雜交動力學較快且雜交時間減少。

【0169】 在步驟 1320 中，進行用於外顯子組富集之第一雜交。舉例而言，DNA 庫與靶向關注區之生物素標記捕獲探針混合。DNA 庫在約 95 $^{\circ}$ C 下變性約 10 分鐘且在約 58 $^{\circ}$ C 下持續約 30 分鐘雜交至探針，總反應時間為約 40 分鐘。

【0170】 在步驟 1325 中，抗生蛋白鏈菌素珠粒用於捕獲雜交至靶向

關注區之生物素標記探針。兩個加熱洗滌程序用於自珠粒移除非特異性結合 DNA。隨後自珠粒洗提富集庫且準備用於第二輪雜交。

【0171】 在步驟 1330 中，使用與第一雜交相同之探針及阻斷劑進行用於外顯子組富集之第二雜交。舉例而言，來自步驟 155 之經洗提 DNA 庫在約 95°C 下變性約 10 分鐘且在約 58°C 下持續約 30 分鐘雜交，總反應時間為約 40 分鐘。第二雜交用於確保捕獲區之高特異性。

【0172】 在步驟 1335 中，抗生蛋白鏈菌素珠粒用於捕獲雜交至靶向關注區之生物素標記探針。兩個加熱洗滌程序用於自珠粒移除非特異性結合 DNA。隨後自珠粒洗提外顯子組富集庫且藉由為定序做準備之十個循環的 PCR 擴增。

【0173】 在步驟 1340 中，使用基於珠粒之純化方案純化外顯子組富集捕獲樣品（亦即外顯子組富集 DNA 庫）。

【0174】 在步驟 1345 中，外顯子組富集 DNA 庫經 PCR 擴增以用於定序。

【0175】 在步驟 1350 中，使用基於珠粒之純化方案視情況純化擴增富集 DNA 庫。舉例而言，1× SPRI 珠粒方案用於移除可能干擾後續叢集擴增及定序之非所需產物（例如過量引子）。

【0176】 方法 100 提供在約 11 小時內之庫製備及外顯子組富集。若省去視情況存在之步驟 1285 及 1300，則方法 1260 提供在約 9 小時內之庫製備及外顯子組富集。

實施例 8

TS-Tn5059 插入傾向

【0177】 轉位酶可在標籤化反應中具有某一插入位點（DNA 序列）傾向。DNA 序列傾向可造成基因體之某些區（例如 GC rich 或 AT 富集區）自標籤化庫丟失。舉例而言，Tn5 轉位酶具有對於基因體之 GC 富集區之某一傾向；因此，基因體之 AT 區可在 Tn5 標籤化庫中丟失。為提供基因體之更完整覆蓋度，需要最小序列傾向。

【0178】 為評估 TS-Tn5059 轉位體對庫輸出及定序度量值之影響，使用用於全基因體定序及蠟狀芽孢桿菌基因體 DNA 之標準 Nextera™ DNA 庫製備套組製備 TS-Tn5059 標籤化 DNA 庫。以 40 nM 之最終濃度使用 TS-Tn5059。使用 25 nM NexteraV2 轉位體之標準反應條件製備參考對照庫。藉由合成定序法（SBS）評估庫。

【0179】 圖 24A 顯示使用 TS-Tn5059 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之覆蓋度之曲線圖 1400。隨著 GC 含量增加，TS-Tn5059 轉位體變得對增加之傾向水準具抗性。圖 24B 顯示使用 NexteraV2 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之覆蓋度之曲線圖 1450。圖 24B 表明隨著 GC 含量增加，GC 富集區之 Nextera V2 覆蓋度變得偏斜，具有增加傾向。資料顯示相比於使用 NexteraV2 製備之標籤化庫，使用 TS-Tn5059 製備之標籤化 DNA 庫具有跨越寬 GC/AT 範圍之經改善及更均勻覆蓋度以及更低插入傾向。

【0180】 圖 25A 顯示使用 TS-Tn5059 轉位體製備之標籤化蠟狀芽孢桿菌基因體 DNA 庫中之間隙位置及間隙長度之曲線圖 1500。圖 25B 顯示使用 NexteraV2 轉位體製備之標籤化蠟狀芽孢桿菌基因體標籤化 DNA 庫中之間隙位置及間隙長度之曲線圖。TS-Tn5059 標籤化庫中之間隙數目為 27。

NexteraV2 標籤化庫中之間隙數目為 208。資料顯示相比於使用 NexteraV2 轉位體製備之標籤化庫，使用 TS-Tn5059 轉位體製備之標籤化 DNA 庫具有更均勻覆蓋度以及更少間隙。

實施例 8

TS-Tn5059 DNA 輸入耐受度

【0181】 製備標籤化 DNA 庫使用酶促 DNA 片段化步驟（例如轉位體介導標籤化）且因此，相比於例如機械片段化方法，可對 DNA 輸入較敏感。在一個實例中，當前 Nextera®快速捕獲富集方案已關於輸入 50 ng 之總基因體 DNA 最佳化。基因體 DNA 之較高質量輸入可導致不完全標籤化及較大插入尺寸，其可影響後續富集效能。標籤化反應中之基因體 DNA 之較低質量輸入或低品質基因體 DNA 可產生小於預期之插入尺寸。較小插入物可在後續清除步驟期間損失且導致較低庫多樣性。

【0182】 為評估不同 DNA 輸入量對片段（插入）尺寸分佈之影響，使用各種酶濃度下之各種量之輸入基因體 DNA 製備 TS-Tn5059 標籤化 DNA 庫且將片段尺寸相比於關於其他轉位酶獲得之片段尺寸，該轉位酶之活性標準化為 40 nM TS-Tn5059 及 25 ng 基因體 DNA 輸入之活性。

【0183】 如圖 26 及 27 中所示，藉由 40nM TS-Tn5059、標準化 TDE1（Tn5 型式-1）及標準化 TS-Tn5 及使用 25 ng 人類基因體 DNA 產生之片段之尺寸分佈類似。

【0184】 然而，TS-Tn5059 顯示 DNA 輸入耐受度在較高酶濃度及經廣泛範圍之輸入 DNA 量增加。圖 28 顯示使用一系列 DNA 輸入量製備之標籤化基因體 DNA 庫中之片段尺寸分佈之生物分析儀跡線之圖 1600。藉由標

籤化、1.8× SPRI 清除繼之以生物分析儀追蹤製備 240 nM 之 TS-Tn5059 標籤化庫。使用當前 Nextera®快速捕獲套組（「Nextera」）及 Agilent QXT 套組（「Agilent QXT」）製備參考對照庫。使用 25、50、75 及 100 ng 之蠟狀芽孢桿菌基因體 DNA 製備標籤化庫。資料顯示相比於使用 Nextera 或 Agilent QXT 轉位體製備之庫，使用 TS-Tn5059 轉位體製備之標籤化 DNA 庫具有跨越 25 至 100 ng DNA 輸入範圍更一致之片段尺寸分佈。隨著 DNA 輸入量自 25 ng 增加至 100 ng，TS-Tn5059 標籤化庫中之標籤化 DNA 之產率增加，同時片段尺寸分佈保持實質上相同。相比之下，在 75 ng 及 100 ng DNA 輸入量下，Nextera 及 Agilent QXT 標籤化庫顯示 DNA 片段尺寸分佈實質上移位至較大片段尺寸。

【0185】 圖 29A 顯示藉由第一使用者使用 5 ng 至 100 ng 之改變輸入量之人類 Coriel DNA 製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖 1700。圖 29B 顯示藉由第二使用者製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖 1750。使用 5、10、25、50、75 及 100 ng 之蠟狀芽孢桿菌基因體 DNA 製備標籤化庫。圖 29A 之曲線圖 1700 及圖 29B 之曲線圖 1750 均顯示使用 5 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1710、使用 10 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1715、使用 25 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1720、使用 50 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1725、使用 75 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1730 及使用 100 ng 之 DNA 輸入量製備之標籤化庫中之片段尺寸分佈之線 1735。資料顯示 TS-Tn5059 標籤化 DNA 庫中之片段

尺寸分佈在 5 至 100 ng 的範圍內之 DNA 輸入量中一致。對於不同使用者觀測到片段尺寸分佈之一致性。

【0186】 在另一實施例中，表 3 顯示跨越 5 至 200 ng 之範圍之 DNA 輸入量的 TS-Tn5059 標籤化 DNA 庫中之中值庫插入尺寸。

表 3. 5 ng 至 200 ng DNA 輸入量之情況下之中值插入尺寸									
輸入 DNA (ng)	5	10	25	50	75	100	150	200	平均值±標準差
中值插入 (bp)	164	169	158	144	171	178	179	175	167 ± 11

【0187】 在另一實施例中，表 4 顯示使用 25、50、75 及 100 ng DNA 輸入量製備之 TS-Tn5059 標籤化 DNA 庫之庫插入尺寸及外顯子組富集定序度量值。資料顯示讀數富集百分比 (%) 為約 80%。使用當前 Nextera®快速捕獲富集方案製備之標籤化庫之讀數富集%為約 60% (資料未示出)。資料亦顯示跨越 15 ng 至 100 ng 之 DNA 輸入範圍之一致插入尺寸。

表 4. 25 至 100 ng DNA 輸入量之情況下之插入尺寸及外顯子組富集度量值					
外顯子組 (皮卡德 (Picard)) 度量值	25 ng	50 ng	75 ng	100 ng	平均值±標準差
讀數富集%	78	79	80	77	78 ± 1
重複%	3.6	2.7	2.9	2.4	2.9 ± 0.3
零覆蓋度%	2.6	1.9	2.2	2.2	2.2 ± 0.2
10x下之外顯子組覆蓋度%	81.6	82.4	82	82.4	82.1 ± 0.4
插入尺寸 150±25 bp	167	169	168	176	170 ± 4
預富集庫定量 (ng/μL)	123	140	165	152	145 ± 15

【0188】 在另一實施例中，表 5 顯示 TS-Tn5059 標籤化庫中跨越 25 ng 至 100 ng 之 DNA 輸入範圍之預富集庫產率。

表 5. 25 至 100 ng DNA 輸入量之情況下之預富集庫產率				
實驗編號	15 μL 之產率 (25 ng 輸入量)	15 μL 之產率 (50 ng 輸入量)	15 μL 之產率 (75 ng 輸入量)	15 μL 之產率 (100 ng 輸入量)
實驗編號 1	1098	1468	2420	2240
實驗編號 2	1332	2200	2480	1974
實驗編號 3	1845	2100	2475	2295
實驗編號 4	1860	2295	2895	2760
實驗編號 5	1830	2895	1515	1665
實驗編號 6	1785	2760	1725	1920

【0189】 在另一實施例中，表 6 顯示 TS-Tn5059 標籤化 DNA 庫之外



顯子組富集定序度量值。開始於 50 ng 之輸入 DNA，使用 500 ng、625 ng 及 750 ng 輸入量之用於外顯子組富集之庫 DNA 製備庫。資料顯示外顯子組富集度量值跨越預富集庫輸入量之範圍（亦即 500 ng 至 750 ng）一致。

表 6. 500 至 750 ng 預富集庫輸入量之外顯子組富集度量值				
營銷度量值	500 ng	625 ng	750 ng	平均值±標準差
讀數富集%	82.2	81.7	82.3	82.1 ± 0.32
重複%	4.1	4.3	3.9	4.1 ± 0.2
零覆蓋度%	1.7	1.6	1.7	1.67 ± 0.06
10x下之外顯子組覆蓋度%	84.3	85.4	85.6	85.5 ± 0.7
HS 庫尺寸	49M	50M	51M	50M ± 1M
平均覆蓋度	45.1	48.2	48.6	47.3 ± 1.92
HS20x罰分	6.5	6.3	6.4	6.4 ± 0.1

【0190】 在另一實施例中，表 7 顯示使用當前 Nextera®快速捕獲富集雜交方案（「NRC」）及圖 23 之方法 1260 之富集步驟 1310 至 1350 製備之標籤化 DNA 庫之外顯子組富集定序度量值。資料顯示相比於使用當前 Nextera®快速捕獲富集方案（「NRC」）中之雜交方案製備之庫，外顯子組富集度量值在使用圖 23 之方法 1260 製備之 TS-Tn5059 標籤化庫中經提高及/或得以維持。

表 7. 使用「NRC」及方法 100 雜交方案製備之標籤化 DNA 庫之外顯子組富集定序度量值		
關鍵外顯子組度量值	NRC 雜交	方法 100 雜交
讀數富集（未填充）	54.3%	77.7%
讀數富集（填充）	64.3%	85%
平均覆蓋度	50.3x	52.1x
零目標丟失	1.6%	1.9%
重複%（1 千萬讀數）	3.1%	4%
10x下之覆蓋度	83.3	84.3%
HS 庫尺寸	62.3 M	61 M
目標之選擇%	76.5%	85.2%
HS 20x罰分	7.4	6.8

【0191】 在獨立實驗中，TS-Tn5059 表明相比於標準化至相同濃度之 Tn5 型式-1 及 TS-Tn5 轉位酶在較高濃度（標準化至 6x濃度）下增加之 DNA 輸入耐受度。結果顯示於圖 30-33 中。6x「標準化」濃度下之 Tn5 型式 1（圖

30) 及 TS-Tn5 (圖 31) 均在 gDNA 輸入量在 25-100ng 之間改變之情況下顯示片段尺寸分佈移位。相比之下，6x標準化濃度下之 TS-Tn5059 在 10-100 ng 之間的 DNA 輸入量下未顯示顯著尺寸移位 (圖 32)。片段尺寸分佈在將 gDNA 輸入增加至 200-500 ng 時開始移位(圖 33)。TS-Tn5059 之增加之 DNA 輸入耐受度之結果概述於下表 8 中。

表 8：最終 50 μ L 反應中之 TS-Tn5059 (nM) :gDNA (ng) 之比

TS-Tn5059 體積 (μ L)	儲備體積 (nM)	最終 Rxn 體積 (μ L)	rxn 中之最 終濃度 (nM)	gDNA 輸入 量 (ng)	50 μ L rxn 中之 TS-Tn5059 (nM) :gDNA (ng) 之比	註釋
5	400	50	40	25	1.6	WGS 比
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	可見之 輸入耐 受度
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	

【0192】 因此，對於 ≥ 2.4 之比率 (nM TS-Tn5059:ng 輸入 DNA) 下之 TS-Tn5059，不存在表明 DNA 輸入耐受度增加之尺寸移位。

【0193】 在整個本申請案中，已參考各種公開案、專利及/或專利申請案。此等公開案之揭示內容特此以全文引用之方式併入本申請案中。

【0194】 術語包含在本文中意欲為開放的，不僅包括所述要素，且另外包涵任何其他要素。

【0195】 已描述多個具體實例。儘管如此，應瞭解，可進行各種修改。因此，其他具體實例在以下申請專利範圍之範疇內。

序列表

<110> 伊路米納有限公司
<120> 用於改善插入序列傾向之經修飾的轉位酶
<130> IP-1198-TW
<160> 26
<170> PatentIn 3.5版
<210> 1
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 1
Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
1 5 10 15

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 2
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (248)..(248)
<223> X = Y, T, K, S, L, A, W, P, G, R, F, H

<400> 2

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val

195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Xaa Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475



<210> 3
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (119)..(119)
<223> X = L, M, S, A, V

<400> 3

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Xaa Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 4
 <211> 476
 <212> PRT
 <213> 未知

<220>
 <223> 轉位酶Tn5

<400> 4

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
 20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
 35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
 50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
 65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
 85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
 100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Met Val His Ser
 115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
 130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
 145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met
 165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
 180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
 195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
 210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
 225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 5
<211> 477
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5



<400> 5

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Val Lys Arg Gly Lys Arg Lys Asn
245 250 255

Arg Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr
260 265 270

Leu Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile
275 280 285

Asn Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser
290 295 300

Glu Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr
305 310 315 320

Thr His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly
325 330 335

Ala Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg
340 345 350

Met Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg
355 360 365

Glu Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys
370 375 380

Glu Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro
385 390 395 400

Asp Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg
405 410 415

Lys Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg
420 425 430

Leu Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly
435 440 445

Ala Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe
450 455 460

Leu Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

- <210> 6
- <211> 476
- <212> PRT
- <213> 未知
- <220>
- <223> 轉位酶Tn5
- <220>
- <221> 變異體
- <222> (120)..(120)
- <223> X = Y, F, W, E
- <400> 6



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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Xaa Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn

275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

- <210> 7
- <211> 476
- <212> PRT
- <213> 未知
- <220>
- <223> 轉位酶Tn5
- <220>
- <221> 變異體
- <222> (119)..(119)
- <223> X = L, M, S, A, V
- <220>
- <221> 變異體
- <222> (120)..(120)
- <223> X = Y, F, W, E



<220>
<221> 變異體
<222> (248)..(248)
<223> X = Y, T, K, S, L, A, W, P, G, R, F, H

<400> 7
Met Ile Thr Ser Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val
1 5 10 15

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Xaa Xaa Ser Arg Gly Trp Met Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Xaa Lys Arg Gly Lys Arg Lys Asn Arg

—

—

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn

275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 9
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 9

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30



Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Glu Ala Met Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr

第 18 頁

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Gln Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Gly Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met

340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Leu Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 11
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 11

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95



Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu

370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 12
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (248)..(248)
<223> X = Y, T, K, S, L, A, W, P, G, R, F, H

<400> 12

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Xaa Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 13
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (119)..(119)
<223> X = L, M, S, A, V

<400> 13

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110



Lys Leu Gly Ser Ile Gln Xaa Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp

第 26 頁

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu

420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 15
<211> 477
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 15

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175



Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Val Lys Arg Gly Lys Arg Lys Asn
245 250 255

Arg Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr
260 265 270

Leu Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile
275 280 285

Asn Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser
290 295 300

Glu Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr
305 310 315 320

Thr His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly
325 330 335

Ala Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg
340 345 350

Met Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg
355 360 365

Glu Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys
370 375 380

Glu Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro
385 390 395 400

Asp Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg
405 410 415

Lys Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg
420 425 430

Leu Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly
435 440 445

Ala Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe

450

455

460

Leu Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 16
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (120)..(120)
<223> X = Y, F, W, E

<400> 16

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Xaa Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190



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

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 17
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<220>
<221> 變異體
<222> (119)..(119)
<223> X = L, M, S, A, V

<220>
<221> 變異體
<222> (120)..(120)
<223> X = Y, F, W, E

<220>
<221> 變異體
<222> (248)..(248)
<223> X = Y, T, K, S, L, A, W, P, G, R, F, H

<400> 17

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Xaa Xaa Ser Arg Gly Trp Met Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160



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

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 18
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 18

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190



Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Gly Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 19
<211> 476
<212> PRT
<213> 未知

<220>
<223> 轉位酶Tn5

<400> 19

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Glu Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Arg His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220



Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Val Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

<210> 20
<211> 476
<212> PRT

<213> 未知

<220>

<223> 轉位酶Tn5

<400> 20

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

Phe Ser Ser Ala Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val
20 25 30

Asn Val Ala Ala Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile
35 40 45

Ser Ser Glu Gly Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile
50 55 60

Arg Asn Pro Asn Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met
65 70 75 80

Gln Thr Val Lys Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu
85 90 95

Asp Thr Thr Ser Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly
100 105 110

Lys Leu Gly Ser Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser
115 120 125

Val Leu Leu Leu Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His
130 135 140

Gln Gln Trp Trp Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys
145 150 155 160

Glu Ser Gly Lys Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met
165 170 175

Gly Ser Met Met Ser Asn Val Ile Ala Val Cys Asp Arg Gly Ala Asp
180 185 190

Ile His Ala Tyr Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val
195 200 205

Val Arg Ser Lys His Pro Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu
210 215 220

Tyr Asp His Leu Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser
225 230 235 240

Ile Pro Gln Lys Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg
245 250 255

Pro Ala Arg Lys Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu
260 265 270

Lys Gln Gly Asn Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn
275 280 285

Pro Pro Lys Gly Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu
290 295 300

Pro Val Glu Ser Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr
305 310 315 320

His Arg Trp Arg Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala
325 330 335

Gly Ala Glu Arg Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met
340 345 350

Val Ser Ile Leu Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu
355 360 365

Ser Phe Thr Pro Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu
370 375 380

Ala Glu His Val Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp
385 390 395 400

Glu Cys Gln Leu Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys
405 410 415

Glu Lys Ala Gly Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu
420 425 430

Gly Gly Phe Met Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala
435 440 445

Leu Trp Glu Gly Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu
450 455 460

Ala Ala Lys Asp Leu Met Ala Gln Gly Ile Lys Ile
465 470 475

- <210> 21
- <211> 10
- <212> PRT
- <213> 未知
- <220>
- <223> 轉位酶Tn5
- <220>
- <221> 變異體

<222> (1)..(1)
<223> x= W, R, K, P, T

<220>
<221> 變異體
<222> (2)..(2)
<223> x= W, N, V, K, M

<220>
<221> 變異體
<222> (3)..(3)
<223> x= V, F, I, T, L

<220>
<221> 變異體
<222> (4)..(4)
<223> x= H, L, W, C

<220>
<221> 變異體
<222> (5)..(5)
<223> x= S, G, V, F

<220>
<221> 變異體
<222> (6)..(6)
<223> x= V, A, W

<220>
<221> 變異體
<222> (7)..(7)
<223> x= L, T, V, Q, W

<220>
<221> 變異體
<222> (8)..(8)
<223> x= L, H, D

<220>
<221> 變異體
<222> (9)..(9)
<223> x= L, H, V

<220>
<221> 變異體
<222> (10)..(10)
<223> x= E, Y, A, K, Q

<400> 21

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
1 5 10

<210> 22
<211> 852
<212> DNA
<213> 大腸桿菌

<400> 22
atggctgaaa ttaccgcatc cctggtaaaa gagctgcgtg agcgtactgg cgcaggcatg 60
atggattgca aaaaagcact gactgaagct aacggcgaca tcgagctggc aatcgaaaac 120
atgcgtaagt ccggtgctat taaagcagcg aaaaaagcag gcaacgttgc tgctgacggc 180
gtgatcaaaa ccaaaatcga cggcaactac ggcatcattc tggaagttaa ctgccagact 240
gacttcgttg caaaagacgc tggtttccag gcgttcgcag acaaagtict ggacgcagct 300



gttgctggca aaatcactga cgttgaagtt ctgaaagcac agttcgaaga agaacgtgtt 360
gcgctggtag cgaaaattgg tgaaaacatc aacattcgcc gcgttgctgc gctggaaggc 420
gacgttctgg gttcttatca gcacggtgcg cgtatcggcg ttctggttgc tgctaaaggc 480
gctgacgaag agctgggtta acacatcgct atgcacgttg ctgcaagcaa gccagaattc 540
atcaaaccgg aagacgtatc cgctgaagtg gtagaaaaag aataccaggt acagctggat 600
atcgcgatgc agtctggtta gccgaaagaa atcgcagaga aaatggttga aggccgcatg 660
aagaaattca ccggcggaagt ttctctgacc ggtcagccgt tcgttatgga accaagcaaa 720
actgttggtc agctgctgaa agagcataac gctgaagtga ctggcttcac ccgcttcgaa 780
gtgggtgaag gcatcgagaa agttgagact gactttgcag cagaagtgc tgcgatgtcc 840
aagcagtcct aa 852

<210> 23
<211> 283
<212> PRT
<213> 大腸桿菌

<400> 23

Met Ala Glu Ile Thr Ala Ser Leu Val Lys Glu Leu Arg Glu Arg Thr
1 5 10 15

Gly Ala Gly Met Met Asp Cys Lys Lys Ala Leu Thr Glu Ala Asn Gly
20 25 30

Asp Ile Glu Leu Ala Ile Glu Asn Met Arg Lys Ser Gly Ala Ile Lys
35 40 45

Ala Ala Lys Lys Ala Gly Asn Val Ala Ala Asp Gly Val Ile Lys Thr
50 55 60

Lys Ile Asp Gly Asn Tyr Gly Ile Ile Leu Glu Val Asn Cys Gln Thr
65 70 75 80

Asp Phe Val Ala Lys Asp Ala Gly Phe Gln Ala Phe Ala Asp Lys Val
85 90 95

Leu Asp Ala Ala Val Ala Gly Lys Ile Thr Asp Val Glu Val Leu Lys
100 105 110

Ala Gln Phe Glu Glu Glu Arg Val Ala Leu Val Ala Lys Ile Gly Glu
115 120 125

Asn Ile Asn Ile Arg Arg Val Ala Ala Leu Glu Gly Asp Val Leu Gly
130 135 140

Ser Tyr Gln His Gly Ala Arg Ile Gly Val Leu Val Ala Ala Lys Gly
145 150 155 160

Ala Asp Glu Glu Leu Val Lys His Ile Ala Met His Val Ala Ala Ser
165 170 175

Lys Pro Glu Phe Ile Lys Pro Glu Asp Val Ser Ala Glu Val Val Glu
180 185 190

Lys Glu Tyr Gln Val Gln Leu Asp Ile Ala Met Gln Ser Gly Lys Pro
195 200 205

Lys Glu Ile Ala Glu Lys Met Val Glu Gly Arg Met Lys Lys Phe Thr
210 215 220

Gly Glu Val Ser Leu Thr Gly Gln Pro Phe Val Met Glu Pro Ser Lys
225 230 235 240

Thr Val Gly Gln Leu Leu Lys Glu His Asn Ala Glu Val Thr Gly Phe
245 250 255

Ile Arg Phe Glu Val Gly Glu Gly Ile Glu Lys Val Glu Thr Asp Phe
260 265 270

Ala Ala Glu Val Ala Ala Met Ser Lys Gln Ser
275 280

<210> 24
<211> 2283
<212> DNA
<213> 未知

<220>
<223> TS-Tn5 核酸序列

<400> 24
atggctgaaa ttaccgcata cctggtaaaa gagctgcgtg agcgtactgg cgcaggcatg 60
atggattgca aaaaagcact gactgaagct aacggcgaca tcgagctggc aatcgaaaac 120
atgcgtaagt ccggtgctat taaagcagcg aaaaaagcag gcaacgttgc tgctgacggc 180
gtgatcaaaa ccaaaatcga cggcaactac ggcatcattc tggaagttaa ctgccagact 240
gacttcgttg caaaagacgc tggtttccag gcgttcgcag acaaagtctt ggacgcagct 300
gttgctggca aaatcactga cgttgaagtt ctgaaagcac agttcgaaga agaacgtgtt 360
gcgctggtag cgaaaattgg tgaaaacatc aacattcgcc gcgttgctgc gctggaaggc 420
gacgttctgg gttcttatca gcacggtgcg cgtatcgcg ttctggttgc tgctaaaggc 480
gctgacgaag agctgggtta acacatcgct atgcacgttg ctgcaagcaa gccagaattc 540
atcaaaccgg aagacglatc cgctgaagtg gtagaaaaag aataccaggt acagctggat 600
atcgcgatgc agtctggtta gccgaaagaa atcgagaga aaatggttga aggccgcatg 660
aagaaattca ccggcgaagt ttctctgacc ggtcagccgt tcgttatgga accaagcaaa 720
actgttggtc agctgctgaa agagcataac gctgaagtga ctggcttcat ccgcttcgaa 780
gtgggtgaag gcatcgagaa agttgagact gactttgcag cagaagtigc tgcgatgtcc 840



aagcagictg gtaccataac ttctgtcttt catcgtgcgg ccgactgggc taaatctgtg 900
ttctcttcgg cggcgctggg tgatcctcgc cgtactgccc gcttgggttaa cgtcgccgcc 960
caattggcaa aatatctctg taaatcaata accatctcat cagagggtag taaagccgcc 1020
caggaaggcg cttaccgatt tatccgcaat cccaacgttt ctgccgaggc gatcagaaag 1080
gctggcgcca tgcaaacagt caagtggct caggagtttc ccgaactgct ggccattgag 1140
gacaccacct ctttgagtta tggccaccag gtgcgccgaag agcttggcaa gctgggctct 1200
attcaggata aatcccgcgg atgggtgggtt cactccgttc tcttgctcga ggccaccaca 1260
ttccgcaccg taggattact gcatcaggag tgggtggatgc gcccggatga ccctgccgat 1320
gcggtatgaaa aggagagtgg caaatggctg gcagcggccg caactagccg gttacgcatg 1380
ggcagcatga tgagcaacgt gattgcggtc tgtgaccgag aagccgatat tcatgcttat 1440
ctgcaggaca aactggcgca taacgagcgc ttcgtgggtgc gctccaagca cccacgcaag 1500
gacgtagagt ctgggttgta tctgtacgac catctgaaga accaaccgga gttgggtggc 1560
tatcagatca gcattccgca aaagggcgtg gtggataaac gcggtaaacg taaaaatcga 1620
ccagcccga aggcgagctt gagcctgcgc agtgggcgca tcacgctaaa acaggggaat 1680
atcacgtca acgcggtgct ggccgaggag attaacccgc ccaagggtga gaccccgttg 1740
aaatggttgt tgcagaccag cgaaccggtc gaggcgttag cccaagcctt gcgcgtcatc 1800
gacatttata cccatcgtg gcggatcgag gaggttccata aggcattgaa aaccggagca 1860
ggagccgaga ggcaacgcat ggaggagccg gataatctgg agcggtatgt ctcgatcctc 1920
tcgtttgttg cggtcaggct gttacagctc agagaaagct tcacgccgcc gcaagcactc 1980
agggcgcaag ggcgtctaaa ggaagcggaa cacgtagaaa gccagtccgc agaaacggtg 2040
ctgaccccg atgaatgtca gctactgggc tatctggaca agggaaaaacg caagcgcaaa 2100
gagaaagcag giagcttgca gtgggcttac atggcgatag ctagactggg cggttttatg 2160
gacagcaagc gaaccggaat tgccagctgg ggcgccctct gggaagggtg ggaagccctg 2220
caaagtaaac tggatggctt tcttgccgcc aaggatctga tggcgagggg gatcaagatc 2280
tga 2283

<210> 25
<211> 760
<212> PRT
<213> 未知

<220>
<223> TS-Tn5 融合蛋白氨基酸序列

<400> 25

Met Ala Glu Ile Thr Ala Ser Leu Val Lys Glu Leu Arg Glu Arg Thr
1 5 10 15

Gly Ala Gly Met Met Asp Cys Lys Lys Ala Leu Thr Glu Ala Asn Gly
20 25 30

Asp Ile Glu Leu Ala Ile Glu Asn Met Arg Lys Ser Gly Ala Ile Lys
35 40 45

Ala Ala Lys Lys Ala Gly Asn Val Ala Ala Asp Gly Val Ile Lys Thr
50 55 60

Lys Ile Asp Gly Asn Tyr Gly Ile Ile Leu Glu Val Asn Cys Gln Thr
65 70 75 80

Asp Phe Val Ala Lys Asp Ala Gly Phe Gln Ala Phe Ala Asp Lys Val
85 90 95

Leu Asp Ala Ala Val Ala Gly Lys Ile Thr Asp Val Glu Val Leu Lys
100 105 110

Ala Gln Phe Glu Glu Glu Arg Val Ala Leu Val Ala Lys Ile Gly Glu
115 120 125

Asn Ile Asn Ile Arg Arg Val Ala Ala Leu Glu Gly Asp Val Leu Gly
130 135 140

Ser Tyr Gln His Gly Ala Arg Ile Gly Val Leu Val Ala Ala Lys Gly
145 150 155 160

Ala Asp Glu Glu Leu Val Lys His Ile Ala Met His Val Ala Ala Ser
165 170 175

Lys Pro Glu Phe Ile Lys Pro Glu Asp Val Ser Ala Glu Val Val Glu
180 185 190

Lys Glu Tyr Gln Val Gln Leu Asp Ile Ala Met Gln Ser Gly Lys Pro
195 200 205

Lys Glu Ile Ala Glu Lys Met Val Glu Gly Arg Met Lys Lys Phe Thr
210 215 220

Gly Glu Val Ser Leu Thr Gly Gln Pro Phe Val Met Glu Pro Ser Lys
225 230 235 240

Thr Val Gly Gln Leu Leu Lys Glu His Asn Ala Glu Val Thr Gly Phe
245 250 255

Ile Arg Phe Glu Val Gly Glu Gly Ile Glu Lys Val Glu Thr Asp Phe
260 265 270

Ala Ala Glu Val Ala Ala Met Ser Lys Gln Ser Gly Thr Ile Thr Ser
275 280 285

Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val Phe Ser Ser Ala
290 295 300

Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val Asn Val Ala Ala



第 45 頁

Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr His Arg Trp Arg
595 600 605

Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala Gly Ala Glu Arg
610 615 620

Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met Val Ser Ile Leu
625 630 635 640

Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu Ser Phe Thr Pro
645 650 655

Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu Ala Glu His Val
660 665 670

Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp Glu Cys Gln Leu
675 680 685

Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys Glu Lys Ala Gly
690 695 700

Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu Gly Gly Phe Met
705 710 715 720

Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala Leu Trp Glu Gly
725 730 735

Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu Ala Ala Lys Asp
740 745 750

Leu Met Ala Gln Gly Ile Lys Ile
755 760

<210> 26
<211> 760
<212> PRT
<213> 未知

<220>
<223> TS-Tn 突變蛋白

<400> 26

Met Ala Glu Ile Thr Ala Ser Leu Val Lys Glu Leu Arg Glu Arg Thr
1 5 10 15

Gly Ala Gly Met Met Asp Cys Lys Lys Ala Leu Thr Glu Ala Asn Gly
20 25 30

Asp Ile Glu Leu Ala Ile Glu Asn Met Arg Lys Ser Gly Ala Ile Lys
35 40 45

Ala Ala Lys Lys Ala Gly Asn Val Ala Ala Asp Gly Val Ile Lys Thr



50	55	60
Lys Ile Asp Gly Asn Tyr Gly Ile Ile Leu Glu Val Asn Cys Gln Thr		
65	70	75 80
Asp Phe Val Ala Lys Asp Ala Gly Phe Gln Ala Phe Ala Asp Lys Val		
	85	90 95
Leu Asp Ala Ala Val Ala Gly Lys Ile Thr Asp Val Glu Val Leu Lys		
	100	105 110
Ala Gln Phe Glu Glu Glu Arg Val Ala Leu Val Ala Lys Ile Gly Glu		
	115	120 125
Asn Ile Asn Ile Arg Arg Val Ala Ala Leu Glu Gly Asp Val Leu Gly		
	130	135 140
Ser Tyr Gln His Gly Ala Arg Ile Gly Val Leu Val Ala Ala Lys Gly		
	145	150 155 160
Ala Asp Glu Glu Leu Val Lys His Ile Ala Met His Val Ala Ala Ser		
	165	170 175
Lys Pro Glu Phe Ile Lys Pro Glu Asp Val Ser Ala Glu Val Val Glu		
	180	185 190
Lys Glu Tyr Gln Val Gln Leu Asp Ile Ala Met Gln Ser Gly Lys Pro		
	195	200 205
Lys Glu Ile Ala Glu Lys Met Val Glu Gly Arg Met Lys Lys Phe Thr		
	210	215 220
Gly Glu Val Ser Leu Thr Gly Gln Pro Phe Val Met Glu Pro Ser Lys		
	225	230 235 240
Thr Val Gly Gln Leu Leu Lys Glu His Asn Ala Glu Val Thr Gly Phe		
	245	250 255
Ile Arg Phe Glu Val Gly Glu Gly Ile Glu Lys Val Glu Thr Asp Phe		
	260	265 270
Ala Ala Glu Val Ala Ala Met Ser Lys Gln Ser Gly Thr Ile Thr Ser		
	275	280 285
Ala Leu His Arg Ala Ala Asp Trp Ala Lys Ser Val Phe Ser Ser Ala		
	290	295 300
Ala Leu Gly Asp Pro Arg Arg Thr Ala Arg Leu Val Asn Val Ala Ala		
	305	310 315 320
Gln Leu Ala Lys Tyr Ser Gly Lys Ser Ile Thr Ile Ser Ser Glu Gly		
	325	330 335

Ser Lys Ala Ala Gln Glu Gly Ala Tyr Arg Phe Ile Arg Asn Pro Asn
 340 345 350
 Val Ser Ala Glu Ala Ile Arg Lys Ala Gly Ala Met Gln Thr Val Lys
 355 360 365
 Leu Ala Gln Glu Phe Pro Glu Leu Leu Ala Ile Glu Asp Thr Thr Ser
 370 375 380
 Leu Ser Tyr Arg His Gln Val Ala Glu Glu Leu Gly Lys Leu Gly Ser
 385 390 395 400
 Ile Gln Asp Lys Ser Arg Gly Trp Trp Val His Ser Val Leu Leu Leu
 405 410 415
 Glu Ala Thr Thr Phe Arg Thr Val Gly Leu Leu His Gln Glu Trp Trp
 420 425 430
 Met Arg Pro Asp Asp Pro Ala Asp Ala Asp Glu Lys Glu Ser Gly Lys
 435 440 445
 Trp Leu Ala Ala Ala Ala Thr Ser Arg Leu Arg Met Gly Ser Met Met
 450 455 460
 Ser Asn Val Ile Ala Val Cys Asp Arg Glu Ala Asp Ile His Ala Tyr
 465 470 475 480
 Leu Gln Asp Lys Leu Ala His Asn Glu Arg Phe Val Val Arg Ser Arg
 485 490 495
 His Arg Arg Lys Asp Val Glu Ser Gly Leu Tyr Leu Tyr Asp His Leu
 500 505 510
 Lys Asn Gln Pro Glu Leu Gly Gly Tyr Gln Ile Ser Ile Pro Gln Lys
 515 520 525
 Gly Val Val Asp Lys Arg Arg Lys Arg Lys Asn Arg Pro Ala Arg Lys
 530 535 540
 Ala Ser Leu Ser Leu Arg Ser Gly Arg Ile Thr Leu Lys Gln Gly Asn
 545 550 555 560
 Ile Thr Leu Asn Ala Val Leu Ala Glu Glu Ile Asn Pro Pro Lys Gly
 565 570 575
 Glu Thr Pro Leu Lys Trp Leu Leu Leu Thr Ser Glu Pro Val Glu Ser
 580 585 590
 Leu Ala Gln Ala Leu Arg Val Ile Asp Ile Tyr Thr His Arg Trp Arg
 595 600 605

Ile Glu Glu Phe His Lys Ala Trp Lys Thr Gly Ala Gly Val Glu Arg
610 615 620

Gln Arg Met Glu Glu Pro Asp Asn Leu Glu Arg Met Val Ser Ile Leu
625 630 635 640

Ser Phe Val Ala Val Arg Leu Leu Gln Leu Arg Glu Ser Phe Thr Pro
645 650 655

Pro Gln Ala Leu Arg Ala Gln Gly Leu Leu Lys Glu Ala Glu His Val
660 665 670

Glu Ser Gln Ser Ala Glu Thr Val Leu Thr Pro Asp Glu Cys Gln Leu
675 680 685

Leu Gly Tyr Leu Asp Lys Gly Lys Arg Lys Arg Lys Glu Lys Ala Gly
690 695 700

Ser Leu Gln Trp Ala Tyr Met Ala Ile Ala Arg Leu Gly Gly Phe Met
705 710 715 720

Asp Ser Lys Arg Thr Gly Ile Ala Ser Trp Gly Ala Leu Trp Glu Gly
725 730 735

Trp Glu Ala Leu Gln Ser Lys Leu Asp Gly Phe Leu Ala Ala Lys Asp
740 745 750

Leu Met Ala Gln Gly Ile Lys Ile
755 760

發明專利說明書

(本說明書格式、順序，請勿任意更動)

【發明名稱】(中文/英文)

用於改善插入序列傾向及增加 DNA 輸入耐受度之經修飾的轉位酶

MODIFIED TRANSPOSASES FOR IMPROVED INSERTION SEQUENCE

BIAS AND INCREASED DNA INPUT TOLERANCE

【相關申請案】

【0001】 本申請案主張 2014 年 4 月 15 日申請之美國臨時申請案第 61/979,871 號；2014 年 10 月 9 日申請之第 62/062,006 號；2014 年 11 月 17 日申請之第 62/080,882 號之優先權，該申請案以全文引用的方式併入本文中。

【技術領域】

【0002】 本案係關於轉位酶酵素和反應條件以及使用其之方法與套組。

【先前技術】

【0003】 轉位酶適用於試管內轉位系統。其允許基因體 DNA 之大規模片段化及標記且適用於自用於核酸分析方法，諸如下一代定序及擴增方法之目標 DNA 製造標記 DNA 片段之庫 (library)。仍需要具有改善特性且產生定性及定量代表產生其之樣品中之目標核酸之標記 DNA 片段之經修飾的轉位酶。

【序列表】

【0004】 本申請案與電子格式之序列表一起提出申請。序列表以 2015 年 4 月 13 日創建的名稱為 IP-1198-TW_Sequencelisting.txt 之文件形式提供，

其大小為 106 Kb。電子格式之序列表中之資訊以全文引用之方式併入本文中。

【發明內容】

【0005】 本文呈現用於改善核酸樣品之片段化及標記之轉位酶。本發明人已出人意料地鑑別某些展現改善的插入序列傾向且具有多種其他相關優勢之經改變轉位酶。

【0006】 本文呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶。在一些具體實例中，突變轉位酶可包含位置 Asp248 處之突變。在某些態樣中，位置 Asp248 處之突變為取代突變。在某些態樣中，位置 Asp248 處之取代突變可包含變成選自由以下組成之群的殘基之突變：Tyr、Thr、Lys、Ser、Leu、Ala、Trp、Pro、Gln、Arg、Phe 及 His。

【0007】 在某些態樣中，位置 Asp248 處之突變為位置 Asp248 之後的插入突變。在某些態樣中，插入突變可包含在位置 Asp248 之後插入疏水性殘基。在某些態樣中，插入突變可包含在位置 Asp248 之後插入纈胺酸殘基。

【0008】 本文亦呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Asp119 處之突變。在某些態樣中，位置 Asp119 處之突變為取代突變。在某些態樣中，位置 Asp119 處之取代突變可包含變成疏水性殘基之突變。在某些態樣中，Asp119 位置處之取代突變可包含變成親水性殘基之突變。在某些態樣中，位置 Asp119 處之取代突變可包含變成選自由以下組成之群的殘基之突變：Leu、Met、Ser、Ala 及 Val。

【0009】 本文亦呈現相對於野生型 Tn5 轉位酶經修飾之突變 Tn5 轉位酶，該突變轉位酶包含位置 Trp125 處之突變。在某些態樣中，位置 Trp125

之 TS-Tn5059 標準化之 TDE1 (Tn5 型式-1) 及 TS-Tn5 製備之標籤化基因體 DNA 庫中之片段尺寸分佈之生物分析儀跡線之圖。

圖 27 顯示使用 25ng 人類 gDNA 使用針對在 40nM (1x標準化濃度) 下之 TS-Tn5059 標準化之 TDE1 (Tn5 型式-1) 及 TS-Tn5 製備之標籤化基因體 DNA 庫中之尺寸分佈之分析。

圖 28 顯示使用一系列 DNA 輸入量製備之標籤化基因體 DNA 庫中之片段尺寸分佈之生物分析儀跡線之圖。

圖 29A 顯示藉由第一使用者;及使用 Coriel 人類 DNA 製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 29B 顯示藉由第二使用者及使用 Coriel 人類 DNA 製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 30 顯示藉由使用 25ng-100ng gDNA 之 6x「標準化」濃度下之 TDE1 製備之 Tn5 型式 1 (TDE1) 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 31 顯示藉由使用 25ng-100ng gDNA 之 6x「標準化」濃度下之 TS-Tn5 製備之 TS-Tn5 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 32 顯示藉由使用 10ng-100ng gDNA 之 6x「標準化」濃度下之 TS-Tn5059 製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀跡線之曲線圖。

圖 33 顯示藉由使用更寬範圍之 gDNA (5 ng-500 ng) 之 6x「標準化」濃度下之 TS-Tn5059 製備之 TS-Tn5059 標籤化庫中之片段尺寸分佈之生物分析儀痕量之曲線圖。

【實施方式】

【0023】 在用於 DNA 定序之一些樣品製備方法中，各模板在插入物之任一末端處含有接附子且通常需要多個步驟以修飾 DNA 或 RNA 及純化修飾反應之所需產物。此等步驟典型地在將經調適片段添加至流動細胞之前在溶液中進行，該片段在流動細胞中藉由將雜交片段複製至共價連接於表面之引子之末端上之引子延長反應偶合至表面。此等『接種』模板隨後經由若干擴增循環產生複製模板之單株叢集。然而，如美國 2010/0120098（其含量全文併入本文中）中揭示，將 DNA 轉化為準備用於叢集形成及定序之溶液中之接附子經修飾模板所需的步驟數可藉由使用在本文中稱為標籤化之轉位酶介導片段化及標記最小化。舉例而言，標籤化可用於對 DNA 片段化，例如如用於 Nextera™ DNA 樣品製備套組（Illumina 公司）之工作流中例示，其中基因體 DNA 可藉由同時將輸入 DNA 片段化及標記之工程改造轉位體片段化，進而產生一群在片段末端包含獨特接附子序列之片段化核酸分子。然而，需要展現改善之插入傾向之轉位酶。

【0024】 因此，本文呈現用於改善核酸樣品之片段化及標記之轉位酶。本發明人已出人意料地鑑別某些展現改善的插入序列傾向且具有多種其他相關優勢之經改變轉位酶。本文中呈現之經改變轉位酶之一個具體實例為展現改善之插入傾向之轉位酶。

【0025】 如本文中所使用，術語「標準化轉位體活性」係指在 50 μ l 反應物中產生以下生物分析儀片段尺寸分佈的關於 25ng gDNA 輸入量之轉位體之最小濃度：總曲線下面積：100-300 bp=20%-30%；301-600 bp=30%-40%；601 -7,000 bp=30%-40%；100 -7,000 bp $\geq$ 90%。此最小濃度稱為 1x。

【符號說明】**【0196】**

100：柱狀圖

200：柱狀圖

300：柱狀圖

400：曲線圖

410：曲線

415：曲線

420：曲線

425：曲線

430：曲線

435：曲線

440：曲線

500：曲線圖

510：曲線

515：曲線

520：曲線

525：曲線

530：曲線

535：曲線

540：曲線

545：曲線

550：曲線

600：曲線圖

610：曲線

615：曲線

620：曲線

625：曲線

630：曲線

635：曲線

700：曲線圖

710：曲線

715：曲線

720：曲線

725：曲線

730：曲線

735：曲線

740：曲線

745：曲線

750：曲線

800：傾向圖

810：曲線

815：曲線

820：曲線

825：曲線

830：傾向圖

840：傾向圖

850：傾向圖

900：傾向圖

910：曲線

915：曲線

920：曲線

925：曲線

930：傾向圖

940：傾向圖

950：傾向圖

1000：柱狀圖

1100：柱狀圖

1200：傾向圖

1210：曲線

1215：曲線

1220：曲線

1225：曲線

1230：傾向圖

1240：傾向圖

1250：傾向圖

1260：方法

1270：步驟

1275：步驟

1280：步驟

1285：步驟

1290：步驟

1295：步驟

1300：步驟

1310：步驟

1320：步驟

1325：步驟

1330：步驟

1335：步驟

1340：步驟

1345：步驟

1350：步驟

1400：曲線圖

1450：曲線圖

1500：曲線圖

1550：曲線圖

1600：圖

1700：曲線圖

1710：線

1715：線

1720：線

1725：線

1730：線

1735：線

1750：曲線圖

I682997

發明摘要

※ 申請案號：104112180

※ 申請日：104年4月15日

※IPC 分類：C12N 9/12 (2006.01)
C12Q 1/68 (2018.01)

【發明名稱】(中文/英文)

用於改善插入序列傾向及增加 DNA 輸入耐受度之經修飾的轉位酶

MODIFIED TRANSPOSASES FOR IMPROVED INSERTION SEQUENCE

BIAS AND INCREASED DNA INPUT TOLERANCE

【中文】

本文呈現用於改善核酸樣品之片段化及標記之轉位酶及反應條件，特定言之展現改善之插入序列傾向之經改變轉位酶及反應條件，以及使用其之方法及套組。

【英文】

Presented herein are transposase enzymes and reaction conditions for improved fragmentation and tagging of nucleic acid samples, in particular altered transposases and reaction conditions which exhibit improved insertion sequence bias, as well as methods and kits using the same.

104 年 7 月 9 日替換頁

【代表圖】**【本案指定代表圖】：**第（ 1A ）圖。**【本代表圖之符號簡單說明】：**

無

【本案若有化學式時，請揭示最能顯示發明特徵的化學式】：

無



108 年 8 月 28 日替換頁

申請專利範圍

1. 一種突變 Tn5 轉位酶，其具有轉位酶活性，其包含 SEQ ID NO: 1 或 SEQ ID NO: 11，其包含對應於 SEQ ID NO: 1 或 SEQ ID NO: 11 中之位置 Pro214 處的取代突變，且包含對應於 SEQ ID NO: 1 或 SEQ ID NO: 11 中之位置 Ala338 處的取代突變，以及視需要地包含對應於 SEQ ID NO: 1 或 SEQ ID NO: 11 中之 Glu54 及/或 Met56 及/或 Lys212 及/或 Leu372 的位置處的取代突變。
2. 如申請專利範圍第 1 項之突變 Tn5 轉位酶，其中該突變 Tn5 轉位酶包含 SEQ ID NO: 1 或 SEQ ID NO: 11 中之位置 Pro214、Ala338 及 Lys212 處的取代突變，且視需要包含對應於 SEQ ID NO: 1 或 SEQ ID NO: 11 中之 Glu54 及/或 Met56 及/或 Leu372 的位置處的取代突變。
3. 如申請專利範圍第 1 項或第 2 項之突變 Tn5 轉位酶，其中該位置 Lys212 處之突變包含變成精胺酸之取代突變。
4. 如申請專利範圍第 1 項或第 2 項之突變 Tn5 轉位酶，其中該位置 Pro214 處之突變包含變成精胺酸之取代突變。
5. 如申請專利範圍第 1 項或第 2 項之突變 Tn5 轉位酶，其中該位置 Ala338 處之突變包含變成纈胺酸之取代突變。
6. 如申請專利範圍第 1 項或第 2 項之突變 Tn5 轉位酶，其中該轉位酶進一步包含對應於 SEQ ID NO: 1 或 SEQ ID NO: 11 中之 Gly251 的位置處之取代突變。
7. 如申請專利範圍第 6 項之突變 Tn5 轉位酶，其中該位置 Gly251 處之取代突變包含變成精胺酸之突變。

108年8月28日替換頁

8. 如申請專利範圍第 1 項之突變 Tn5 轉位酶，其包含 SEQ ID NO:19。
9. 一種融合蛋白，其包含融合至多肽融合域之如申請專利範圍第 1 項至第 7 項中任一項之突變 Tn5 轉位酶。
10. 如申請專利範圍第 9 項之融合蛋白，其中該多肽融合域融合至該突變 Tn5 轉位酶之 N 端；或
其中該多肽融合域融合至該突變 Tn5 轉位酶之 C 端。
11. 如申請專利範圍第 10 項之融合蛋白，其中該多肽融合域包含增加溶解度之標籤，或
其中該多肽融合域包含選自以下者之域：麥芽糖結合蛋白（MBP）、延伸因子 Ts（Tsf）、5-甲基胞嘧啶結合域及蛋白 A。
12. 如申請專利範圍第 9 項之融合蛋白，其包含 SEQ ID NO:26。
13. 一種核酸分子，其編碼如申請專利範圍第 1 項至第 12 項中任一項之突變或融合蛋白 Tn5 轉位酶。
14. 一種表現載體，其包含如申請專利範圍第 13 項之核酸分子。
15. 一種宿主細胞，其包含如申請專利範圍第 14 項之載體。
16. 一種用於試管內轉位之方法，其包含允許以下組分相互作用：(i) 包含如申請專利範圍第 1 項至第 12 項中任一項之突變或融合蛋白 Tn5 轉位酶之轉位體（transposome）複合物，及 (ii) 目標 DNA。
17. 一種定序目標 DNA 之方法，其包含：
 - (a) 使該目標 DNA 與包含以下者之轉位體複合物一起培育：
 - (1) 如申請專利範圍第 1 項至第 12 項中任一項之突變或融合蛋白 Tn5 轉位酶；及

108 年 8 月 28 日 替換頁

(2) 包含以下之第一聚核苷酸：

(i) 包含轉位子末端序列之 3'部分，及

(ii) 包含第一定序標籤域之第一標籤，

該培育係在藉以使得該目標 DNA 被片段化，且該第一聚核苷酸之 3'轉位子末端序列被轉移至該片段之 5'末端之條件下，

從而產生雙股片段，其中 5'末端經該第一標籤標記，且在該 5'標記股之 3'末端處存在單股間隙；

(b) 使該片段與核酸修飾酶一起在藉以使得第二標籤附接至該 5'標記股之 3'末端之條件下培育，

(c) 視情況藉由提供聚合酶及對應於該第一聚核苷酸之一部分的擴增引子來擴增該片段，從而產生在 5'末端處具有該第一標籤且在 3'末端處具有第二標籤之代表性二標記片段庫 (library)；

(d) 提供包含對應於該第一定序標籤域之部分的第一定序引子；及

(e) 並行地擴展該第一定序引子及偵測與該代表性二標記片段庫之第一定序標籤域鄰接之核苷酸之身分 (identity)。

18. 一種用於進行試管內轉位反應之套組，其包含：

包含以下者之轉位體複合物：

(1) 如申請專利範圍第 1 項至第 12 項中任一項之突變或融合蛋白 Tn5 轉位酶；及

(2) 包含含有轉位子末端序列之 3'部分之聚核苷酸。

19. 一種用於試管內轉位之方法，其包含允許以下組分相互作用：(i) 包含

轉位酶之轉位體複合物，(ii) 目標 DNA；及 (iii) 包含 Co^{2+} 的反應緩

108 年 8 月 28 日 替換頁

衝劑， Co^{2+} 之濃度為相比於在包含 Mg^{2+} 之緩衝劑中之相同酶足以減少 GC 傾向之有效濃度，其中該轉位酶酵素包含如申請專利範圍第 1 項至第 12 項中任一項之突變或融合蛋白 Tn5 轉位酶。