




# 中华人民共和国国家知识产权局

100037 北京市阜成门外大街2号万通新世界广场8层 中国国际贸易促进委员会专利商标事务所 程泳 <i>0260564</i>	发文日期  2007年04月13日
申请号: 2006100715457	
申请人: 基因库公司	
发明创造名称: 用特定序列组合体检测核酸的方法	

## 发明专利申请公布及进入实质审查程序通知书

上述专利申请,经初步审查符合专利法及其实施细则的有关规定,根据专利法第三十四条规定,该申请已在第23卷,第13期发明专利公报上予以公布。

根据申请人提出的实质审查请求,经审查符合专利法第三十五条,实施细则第九十条规定,该专利申请已进入实质审查程序。

注:附公布说明书一份。

提示:

1.自本通知书发文日起,该申请即进入实质审查程序,自此专利申请人向专利局递交各种专利文件时,应在文件上注明实质审查程序。

2.根据专利法实施细则第五十一条第一款的规定,发明专利申请人在收到国务院专利行政部门发出的发明专利进入实质审查阶段通知书之日起三个月内,可以对发明专利申请主动提出修改。



审查员: 冷鲁

审查部门: 初审及流程管理部

2007年4月3日

21229  
2002.8



回函清寄: 100088 北京市海淀区蓟门桥西土城路6号 国家知识产权局专利局受理处收  
(注: 凡寄给审查员个人的信函不具有法律效力)

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
C12Q 1/68 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200610071545.7

[43] 公开日 2007年3月28日

[11] 公开号 CN 1936020A

[22] 申请日 1995.12.7

[21] 申请号 200610071545.7

分案原申请号 95197558.7

[30] 优先权

[32] 1994.12.9 [33] US [31] 80/353,476

[71] 申请人 基因库公司

地址 美国华盛顿

[72] 发明人 苏珊·威恩英格

阿瑟·M·威恩英格

[74] 专利代理机构 中国国际贸易促进委员会专利商  
标事务所

代理人 程泳

权利要求书 10 页 说明书 155 页 附图 29 页

## [54] 发明名称

用特定序列组合体检测核酸的方法

## [57] 摘要

本发明是一种具有高度灵敏性和特异性的检测和定位样品中特定核酸序列的新方法。该方法和该方法中使用的新组合体涉及使用探针核酸、产生核酸结合区和使用靶核酸结合装配体来检测和定位靶核酸。即使存在具有类似序列的核酸，该方法也可以完成所说靶核酸的检测和定位。该方法使得各个特定结合事件所产生的信号高度扩增。具体地说，所说方法和组合体提供来用于检测样品中 HIV 和 HPV 核酸。这些方法和组合体可以用于疾病诊断、遗传监测、法医侦探以及核酸混合物分析。用于所说检测方法中的某些新组合体在预防或治疗病原性疾病上有用。



1. 一种探针核酸(PNA), 该核酸包含:

(A) 单链序列 1/2 TBR, 该序列在杂交条件下可以和存在于靶核酸(TNA)中的 1/2 TBR 形成杂交体 TBR;

(B) 单链序列 1/2 BBR, 该序列在杂交条件下可以和辅助核酸(BNA)中的约 0-10 个 1/2 BBR 形成杂交体 BBR; 和

(C) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

其中所说的 TBR 可以以高亲和力和 TBA 结合, 所说的 TBA 是一种能够区分配对的 TBR 和具有未配对的核苷酸的 TBR 的物质, 并且, 所说的 BBR 可以以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分配对的 BBR 和具有未配对的核苷酸的 BBR 的物质。

2. 一种辅助核酸(BNA), 该核酸包含:

(A) 1/2 BBR, 其具有和 PNA 或另外的 BNA 中 1/2 BBR 序列互补的序列, 并且其在杂交条件下可以和 PNA 形成杂交体 BBR;

(B) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上; 和

(C) 附加的杂交位点 1/2 BBR, 用来和附加的 BNA 杂交;

其中所说的 BBR 可以以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分配对的 BBR 和具有未配对的核苷酸的 BBR 的物质。

3. 一种发夹核酸(HNA), 该核酸包含单链序列 1/2 BBR, 该序列在杂交条件下能够形成发夹结构, 并且同时和 BNA 结合形成能够结合 BBA 的 BBR, 其中所说的 BBR 可以以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分完美的 BBR 和具有未配对的核苷酸的 BBR 的物质。

4. 权利要求 1 的 PNA, 其中所说的 TBR 由一个或多个核酸结合蛋白、DNA 结合蛋白、DNA-RNA 杂交结合蛋白或 RNA 结合蛋白的识别位点组成。

5. 权利要求4的PNA, 其中所说的TBR是存在于病原体基因组中的核酸结合蛋白识别位点, 或者是和脊椎动物基因组病理状态相关的结合位点, 或者是存在于污染发酵过程的生物体基因组中的核酸结合蛋白识别位点。

6. 权利要求4的PNA, 其中所说的TBR是HIV-LTR或其一部分。

7. 一种用于检测和定位特定TNA序列的方法, 该方法包括以下步骤:

(A) 使所说的TNA与权利要求1的PNA杂交;

(B) 使所说的TNA与含有1/2 BBR的BNA杂交, 所说1/2 BBR的序列和PNA中的1/2 BBR序列互补;

(C) 将含有TBR和BBR的步骤(A)和(B)的产物加入到含有TBA的表面、液体或其它介质中;

(D) 将BBA加入到步骤C的混合物中, 其中所说的BBA包含:

(I) 能够选择性地结合BBR的分子或分子的一部分;

(II) 可检测的指示剂; 和

(E) 检测连接到BBA上的指示剂所产生的信号。

8. 权利要求7的方法, 其中所说的指示剂是蛋白质, 包括可以催化反应导致产生有色反应产物的酶, 放射性核素, 有色小球。

9. 一种检测样品中特定靶核酸TNA存在的方法, 该方法包括以下步骤:

(A) 使所说的样品和探针核酸PNA接触, 如果样品中存在所说的TNA, 则PNA与TNA杂交形成靶结合区TBR, 此TBR可以和靶结合装配体TBA结合; 和

(B) 使已经和所说的PNA接触的样品和TBA接触, 此TBA可以结合任何由所说的PNA与样品中所说的TNA杂交形成的TBR。

10. 一种具有高灵敏性和特异性的检测和定位特定核酸序列的方法, 该方法包括:

(A) 将含有1/2 BBR和1/2 TBR的PNA加入到含有或怀疑含有包含1/2 TBR序列的TNA样品中, 以形成具有靶结合区TBR的复合体, 此TBR由分别存在于PNA和TNA中的互补1/2 TBR的杂交形成;

(B) 将步骤A中形成的TBR结合到固定化的TBA上, 以形成TBA-TNA-PNA复合体;

(C) 将包含辅助结合区 1/2 BBR 的辅助核酸 BNA 加入到步骤 B 中形成的复合体中, 以便 BNA 中的 1/2 BBR 与存在于 PNA 中的 1/2 BBR 序列杂交, 或者加入到存在于已经和 PNA 结合的 BNA 中的 1/2 BBR 上形成 BBR, 以便形成 TBA-TNA-PNA-(BNA)<sub>n</sub> 复合体;

(D) 将含有 1/2 BBR 序列的发夹核酸 HNA 加入到步骤 C 中形成的复合体中, 以便 HNA 中的 1/2 BBR 和任何存在于步骤 C 的复合体的 BNA 中的可用的 1/2 BBR 序列杂交, 由此在步骤 C 的 TBA-TNA-PNA-(BNA)<sub>n</sub> 复合体上加帽使 BNA 延伸, 形成 TBA-TNA-PNA-(BNA)<sub>n</sub>-HNA 复合体;

(E) 将和指示剂部分连接的辅助结合装配体加入到步骤 D 中所形成的 TBA-TNA-PNA-(BNA)<sub>n</sub>-HNA 复合体中, 以便形成 TBA-TNA-PNA (BNA-BBA)<sub>n</sub>-HNA 复合体; 和

(F) 检测由连接到步骤 (E) 中 TBA-TNA-PNA-(BNA-BBA)<sub>n</sub>-HNA 复合体中的 TBA, PNA, BNA, BBA 或 HNA 上的指示剂所产生的信号;  
其中所说的 TNA 包含:

(i) 一个或多个特定 1/2 TBR 核酸序列, 此特定序列在特定样品中的存在与否有待证实;

所说的 PNA 包含:

(i) 单链序列 1/2 TBR, 该序列在杂交条件下可以和存在于靶核酸 (TNA) 中的 1/2 TBR 形成杂交体 TBR;

(ii) 单链序列 1/2 BBR, 该序列在杂交条件下可以和辅助核酸 (BNA) 中的 0-10 个 1/2 BBR 形成杂交体 BBR; 和

(iii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

所说的 BNA 包含:

(i) 1/2 BBR, 其具有和 PNA 或另外的 BNA 中 1/2 BBR 序列互补的序列, 并且其在杂交条件下可以和 PNA 形成杂交体 BBR;

(ii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上; 和

(iii) 附加的其它 BNA 的杂交位点 1/2 BBR;

(iv) 序列 1/2 BBR, 其可以与已经和 PNA 杂交的 BNA 杂交;

所说的 BBA 包含:

(i) 能够选择性地结合 BBR 的分子或分子的一部分; 和

(ii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

并且所说的 TBA 包含

(i) 能够选择性地结合 TBR 的分子或分子的一部分; 和

(ii) 非连接的支持物和/或指示剂, 或者连接的支持物或其它定位方式和/或指示剂, 所说的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

11. 用于在固相中检测靶多核苷酸存在的杂交方法, 该方法涉及将靶多核苷酸(如果存在于样品中)直接或经媒介俘获结构固定化在固相的俘获位点上; 在所说的固定化过程之前、之中或之后将一种可检测的标记连接到所说的靶多核苷酸(如果存在)上, 并且在俘获位点上检测所说的标记(如果具有任何所说的标记存在), 改进包括:

(A) 使用靶结合装配体 TBA 作为达到固定化所说的靶多核苷酸的手段, 其中所说的 TBA 只结合到由特定的探针核酸 PNA 和所说的靶核酸形成的独特的杂交体上, 以便形成可以被所说的 TBA 识别的完美的靶结合区 TBR; 和

(B) 在 PNA 中包含进可以结合辅助核酸 BNA 的单链核酸 1/2BBR, 所说辅助核酸含有单链互补 1/2 BBR, 当其和 PNA 中的 1/2 BBR 杂交时, 形成可以和标记的辅助结合装配体 BBA 结合的 BBR。

12. 一种靶结合装配体 TBA 或者辅助结合装配体 BBA, 其包含至少一个核酸识别单位, 其中可以也可以不包含选自下组的一个或所有序列: 接头序列、装配序列、不对称序列、核定位信号序列(NLS)和 OSA。

13. 权利要求 12 的 TBA, 其中所说的核酸识别单位选自下组: NF-kB 结合单位、SP1 结合单位、TATA 结合单位、人乳头状瘤病毒 E2 结合单位、HPV LTR 结合单位、HIV LTR 结合单位和 Tat。

14. 权利要求 13 的 TBA, 其中所说的核酸识别单位具有选自下组的序列: SEQ ID NO. 63、SEQ ID NO. 64、SEQ ID NO. 65、SEQ ID NO. 66、SEQ ID NO. 67、SEQ ID NO. 68、SEQ ID NO. 69、SEQ ID NO. 70、SEQ ID NO. 71、SEQ ID NO. 72、SEQ ID NO. 73、SEQ ID NO. 74、SEQ ID NO. 75、SEQ ID NO. 76、SEQ ID NO. 77、SEQ ID NO. 78、SEQ ID NO. 79、SEQ ID NO. 80、SEQ ID NO. 81、SEQ ID NO. 82、SEQ ID NO. 83、SEQ ID NO. 84、SEQ ID NO. 93、SEQ ID NO. 94、SEQ ID NO. 95、SEQ ID NO. 96、SEQ ID NO. 97、SEQ ID NO. 98 和 SEQ ID NO. 118。

15. 权利要求 12 的 TBA, 其中所说的接头序列是寡肽, 此寡肽不干扰核酸识别单位的核酸识别功能, 并且提供核酸识别单位与剩余 TBA 间隔的稳定性和控制。

16. 权利要求 15 的 TBA, 其中所说的接头序列是来源于结构蛋白区域间一级序列的寡肽序列。

17. 权利要求 12 的 TBA, 其中所说的装配序列是一种寡肽序列, 此寡肽序列指导核酸识别单位的折叠和缔合。

18. 权利要求 17 的 TBA, 其中所说的装配序列来源于 $\lambda$ 噬菌体 *cro* 蛋白或 CI 蛋白, 并且选自下组: SEQ ID NO. 104、SEQ ID NO. 105、SEQ ID NO. 106、SEQ ID NO. 107 和 SEQ ID NO. 108。

19. 权利要求 12 的 TBA, 其中所说的不对称序列指导核酸识别序列和装配序列以预定的顺序缔合。

20. 权利要求 17 的 TBA, 其中所说的不对称序列来源于胰岛素、促性腺激素、FSH、HCG、LH、ACTH 或松驰素。

21. 权利要求 20 的 TBA, 其中所说的不对称序列选自下组: SEQ ID NO. 85、SEQ ID NO. 86、SEQ ID NO. 87、SEQ ID NO. 88、SEQ ID NO. 89、SEQ ID NO. 90、SEQ ID NO. 91 和 SEQ ID NO. 92。

22. 权利要求 12 的 TBA, 其中所说的 NLS 是一种寡肽, 此寡肽指导与所说的 NLS 相关的蛋白质或复合体迁移和摄入到细胞核中。

23. 权利要求 22 的 TBA, 其中所说的 NLS 选自下组: SEQ ID NO. 72 和 SEQ ID NO. 103。

24. 权利要求 12 的 TBA, 其是 HIV Detect I-IV 或 HPV Detect I-IV。

25. 权利要求 12 的 TBA, 其具有选自下组的序列: SEQ ID NO. 109、SEQ ID NO. 110、SEQ ID NO. 111、SEQ ID NO. 112、SEQ ID NO. 113、SEQ ID NO. 114、SEQ ID NO. 115 和 SEQ ID NO. 116。

26. 一种使用权利要求 12 的 TBA 结合靶核酸样品中特定的核酸序列的方法, 该方法包括:

(A) 裂解靶核酸样品中的核酸;

(B) 在杂交条件下, 使裂解的核酸与和特定的兴趣核酸互补的探针核酸接触, 其中在所说的的探针核酸与所说的特定的兴趣核酸序列杂交时, 形成所说的 TBA 特异性地结合到其中的靶结合区。

27. 权利要求 26 的方法, 其中所说的探针核酸除了具有和所说的特定的兴趣核酸互补的序列外, 还具有附加序列, 此附加序列可与辅助核酸结合, 形成辅助结合位点, 标记的辅助结合装配体可以结合到此结合位点上, 提供显示和扩增探针核酸和兴趣靶核酸序列结合的信号。

28. 一种使用权利要求 12 的 TBA 的方法, 其中将预防和治疗有效量的所说 TBA 施用给需要此种治疗的病人, 该方法包括以提纯的蛋白复合体的形式或以重组载体(进入病人体内后可以表达 TBA)的形式施用 TBA, 以便 TBA 和特定的核酸序列结合达到所期望的治疗和预防结果。

29. 权利要求 28 方法, 其中所说的 TBA 选自下组: SEQ ID NO. 109、SEQ ID NO. 110、SEQ ID NO. 111、SEQ ID NO. 112、SEQ ID NO. 113、SEQ ID NO. 114、SEQ ID NO. 115、SEQ ID NO. 116, 并且所说的病人是由 HIV 和 HPV 感染的病人。

30. 权利要求 26 的方法, 该方法还包括以下步骤:

(C) 监测靶核酸样品中核酸移动性的改变, 以便 TBA 和样品中特定的核酸片段结合改变此片段的移动性作为核酸大小的函数。

31. 一种用于检测样品中具有特定序列组成的核酸的诊断或法医试验试剂盒, 该试剂盒包含:

(A) 第一个核酸探针, 其互补于具有特定序列组成的核酸, 此核酸在试验样品中的存在与否是所要确定的, 其中所说的第一核酸探针和所说的具有特定序列组成的核酸杂交时, 形成第一个核酸结合蛋白的结合位点, 其



中所说的第一核酸探针还含有和第二核酸探针互补的附加序列;

(B) 第一核酸结合蛋白, 其对由所说的第一核酸探针和所说的具有特定序列组成的核酸杂交形成的双链体是特异性的;

(C) 第二核酸探针, 其互补于所说的第一核酸探针中所说附加序列, 当所说的第一和第二核酸探针杂交时, 形成第二个核酸结合蛋白的结合位点。

(D) 第二个核酸结合蛋白, 其特异性地结合由所说第一核酸探针和所说第二核酸探针杂交形成的双链体, 其中所说的第二核酸结合蛋白由可检测的标记标记。

32. 权利要求 31 的诊断和法医试验试剂盒, 其中所说的第一核酸探针与 HIV LTR 互补, 由此当所说第一核酸探针和 HIV LTR 杂交时, 形成 NF-KB 或其亚单位 SP1、TATA 结合蛋白、HIV-Detect I, II, III 或 IV、或者 HIV-Lock 的结合位点。

33. 权利要求 32 的诊断或法医试验试剂盒, 其中所说的第一核酸结合蛋白是 NF-KB 或其亚单位 SP1、TATA 结合蛋白、HIV-Detect I, II, III 或 IV、或者 HIV-Lock。

34. 权利要求 33 的诊断或法医试验试剂盒, 其中所说的第一核酸探针除了和 HIV LTR 互补外还包含噬菌体 $\lambda$  左和右操纵子的编码序列, 所说的第二核酸探针包含和所说的第一核酸探针中噬菌体 $\lambda$  左或右操纵子序列互补的序列, 由此当所说的第一和所说的第二核酸探针杂交时, 噬菌体 $\lambda$  CI 抑制子蛋白, 噬菌体 $\lambda$  *cro* 蛋白或它们的衍生物或类似物的结合位点。

35. 权利要求 34 的诊断或法医试验试剂盒, 其中所说的第二核酸结合蛋白是噬菌体 $\lambda$  CI 抑制子蛋白, 噬菌体 $\lambda$  *cro* 蛋白或它们的衍生物或类似物。

36. 一种组合物, 该组合物包含 HIV-Lock 或编码 HIV-Lock 的重组载体和药物上可接受的载体。

37. 一种特异性地将核酸结合蛋白结合到与病理状态相关的核酸上的方法, 该方法包括:

(A) 选择存在于病理状态相关核酸序列中的特定构型的核酸结合蛋白序列作为靶序列, 用以设计将与样品中那些特定构型的核酸序列(如果存在)

杂交的核酸探针，并且保证在所说的核酸探针和选为靶的所说特定构型的核酸序列杂交时，形成可获得的核酸结合蛋白的结合位点；

(B) 选择核酸结合蛋白，此蛋白特异性地和与病理状况相关的选择的特定构型的核酸结合蛋白序列结合，但此蛋白不和与所说病理状况不相关的序列结合。

(C) 将所说的核酸探针与怀疑含有存在于病理状态相关核酸序列中的所说特定构型的核酸结合蛋白序列的试验样品杂交；

(D) 使所说的核酸结合蛋白和(B)中形成的任何一个杂交体接触。

(E) 检测所说核酸结合蛋白和所说杂交体的任何结合。

38. 权利要求 37 的方法，其中所说的特定构型的核酸结合蛋白序列选自病理状态发展的关键步骤或控制点。

39. 权利要求 9 的方法，其中所说的方法以自动方式进行。

40. 权利要求 39 的方法，其中所说的方法是在 Abbott Laboratories IMx 仪器上进行的。

41. 权利要求 9 的方法，其是在微量滴定板上进行的。

42. 一种用于扩增经将权利要求 1 的 PNA 和 TNA 结合产生的信号的方法，该方法包括将 BNA 结合到 PNA-TNA 杂交体上和将标记的 BBA 结合到 BNA 上。

43. 一种用于装配核酸结合复合体的方法，该方法包括用不对称序列指导核酸结合复合体组份的缔合或者非缔合。

44. 一种用于装配核酸结合复合体的方法，该方法包括使用来源于噬菌体  $\lambda$  *cro* 或 *CI* 的装配序列来装配核酸结合复合体的缔合组份。

45. 一种使用装配，不对称，或先导序列来装配多聚体蛋白复合体的方法，该方法包括将待掺入到多聚体蛋白复合体中的亚单位连接到所说的装配，不对称或先导序列上，以及回收所说的多聚体复合体。

46. 一种组合物，该组合物包含 SEQ ID NO. 105、SEQ ID NO. 106 或 SEQ ID NO. 108。

47. 一种编码权利要求 12 的 TBA 或 BBA 的核酸。

48. 权利要求 12 的 TBA 或者编码所说的 TBA 的核酸，其中所说的 TBA 的氨基酸序列选自下组：组合 A、组合 B 和组合 C，所说的各组合如以下所示：

组合	连接的序列组
A	I + II + III
B	IV + V + III
C	IV + III

其中 I-V 组所包括的序列选自:

组	所选自的序列
I	SEQ ID NO. 85-92 的任何序列
II	Met Ser, 连接到 SEQ ID NO. 104-106 的任何序列上, 其各自连接到 SEQ ID NO. 99 上。
III	连接到 SEQ ID NO. 75-84 或 94-98 之任一上的 SEQ ID NO. 100, 连接到 SEQ ID NO. 74 或 SEQ ID NO. 93 上的 SEQ ID NO. 101; 或连接到 SEQ ID NO. 74 或 SEQ ID NO. 93 上的 SEQ ID NO. 102; 或者 SEQ ID NO. 72, 103, 73, 或 63-71 的任何序列。
IV	SEQ ID NO. 104-108 的任何序列
V	SEQ ID NO. 99。

49. 一种用于在体内或原位装配多聚体 TBA 方法, 该方法包括使用共价或非共价连接的蛋白或双层载体将组份 TBA 引入到细胞中, 或者通过将编码 TBA 组份的核酸引入到细胞中来将组份 TBA 引入到细胞中, 所说的 TBA 组份每个都含有 DNA 识别单位, 装配序列, 不对称序列, 核定位信号序列和可有可无的接头序列, 由此当经由各组份 TBA 的 DNA 识别单位邻近地结合到在细胞核中或细胞其它区域遇到的核酸序列上时, 组份表达的 TBA 通过所说的装配和不对称序列装配成多聚体 TBA。

50. 一种用于鉴别核酸结合分子以便制备靶结合装配体或辅助结合装配体的方法, 该方法包括:

- (A) 获得含有靶核酸的样品;
- (B) 破碎此样品以便在样品中暴露核酸并减少所说核酸大小的复杂性;
- (C) 使第一小份破碎的核酸和对照缓冲介质接触; 使第二小份破碎的核酸和含有已知组成的核酸结合分子的对照缓冲介质接触;

(D) 分析所说的两个小份，以鉴别和与对照缓冲介质接触的小份相比在与靶结合分子接触的小份中已改变行为的片段；

(E) 鉴别和分离在和核酸结合分子接触时确实显示出改变的行为的，测定所说核酸片段的序列以确定是否存在已知的核酸结合分子基序或者直接鉴别结合到核酸上的核酸结合分子；和

(F) 用装配、不对称核定位以及可有可无的接头序列合成含有产生改变的行为的核酸结合分子的 TBA。

51. 一种用于鉴别样品中特定核酸序列方法，该方法包括：

(A) 破碎所说样品中的核酸以暴露核酸和降低核酸大小的复杂性；

(B) 使 TBA 和样品接触，所说的 TBA 含有两个或多个核酸结合组份，每个组份具有相对弱的对 TBR 中的核酸识别单位的结合性，但是在组合时其提供对完整的 TBR 的强的结合性；和

(C) 排除任何由 TBA 和含有单独识别单位的表亲核酸结合产生的“交叉对话”，其包括将所说样品和过量的核酸结合组份接触，该组份对含有单个识别单位的表亲核酸具有相对强的亲和力，但对于含有两个或多个核酸结合组份的完整 TBR 的结合而言，具有较 TBA 亲和力相对弱的结合力。

## 用特定序列组合体检测核酸的方法

### 发明背景

#### 1. 发明所属技术领域

本发明提供了用于结合、检测和扩增样品中特定靶核酸序列检测结果(即使在有密切相关但不同的核酸存在下)的具有真实性和准确性的方法和组合体(composition)。所说的结合可以涉及将特定分子陪伴和装配成靶结合装配体,该装配体特异性地与由探针核酸和靶核酸序列杂交形成的靶结合区结合。所说的扩增可以涉及将特定分子陪伴和/或装配成辅助(boosters)结合装配体,该装配体特异性地与由辅助核酸和探针核酸、靶核酸序列或其它辅助核酸杂交形成的辅助靶结合区结合。本发明也提供了一种涉及发夹核酸的方法和组合体,其提供来控制用于扩增中的特异性或非特异性伸延的辅助核酸和辅助结合装配体的大小。所说的检测涉及提供一种或几种检测标记,包括放射性、光或荧光发射酶以其它可检测的或可产生信号分子,这些标记与探针核酸、靶结合装配体、辅助核酸、辅助核酸装配体或者发夹核酸相关联。本发明也提供了一种用于从生物体分离核酸片段的方法,所说核酸片段具有产生探针核酸和对所说片段和/或生物体唯一的TBA的TBA组份结合位点。本发明还提供了靶结合装配体在预防和治疗上的用途和用于这种用途的组合物(composition)。

#### 2. 背景和相关技术的描述

在越来越多的场合,从样品中检测包含特定序列的核酸(此后称为靶核酸(TNA))是十分重要的。需要能够用最少的步骤,最简单的组份检测TNA,并排除其它类似的但不同的核酸(此后称为表亲核酸(CNA))的干扰。不需要扩增或其它检测后的步骤就可以检测特定的TNA并可以排除任何和所有表亲分子是合乎需要的。

尽管具有许多用固定化的或者标记的核酸作为探针检测TNA的方法,但是使用目前已知手段来区分与探针核酸(PNA)结合的TNA和与PNA结合的

CNA。例如，CNA和PNA之间的一个或多个错配碱基仍可以导致无法和TNA-PNA杂交区分开来的CNA-PNA杂交。因此杂交本身并不是一个可用来表明PNA已经和唯一的TNA杂交的理想指示。

具有许多场合，其中PNA会被用来试图确定在含有CNA的样品中TNA存在与否。如果没有附加的证实，在这种情况下PNA和任何CNA杂交会限制PNA用来检测TNA的应当具有的诊断值。而且，可以从含有许多CNA拷贝的样品中检测和定位低拷贝数的TNA而不需要产生额外的TNA拷贝是合乎需要的。可以证实CNA的存在(不依赖于TNA)而不需要样品中的TNA和CNA也是合乎需要的。

另外，能够扩增信号即使是特定TNA-PNA低频率杂交的信号是合乎需要的。为此目的，一种聚合标记的多个拷贝(此后称为辅助核酸(BNA))到TNA-PNA上的方法将是合乎需要的。

本发明提供了达到以上所期望的目的的方法，正如如下所综述的，本发明的组合体和方法在现有技术中还没有被报道过或者提示过。Keller, H., M. M. Manak (1989)的DNA探针一书提供了本领域核酸检测状态的全面的和广泛的回顾。

一种用化学方法检测碱基错配以决定PNA杂交于CNA而不是TNA的方法已有报道。在Ford等的美国专利4,794,075中，讨论了一种把含有单一碱基错配的DNA片段和其完全配对的类似物区分开来的方法。双链片段中的单链区域被碳化二亚胺修饰，碳化二亚胺与未配对的鸟嘌呤(G)和胸腺嘧啶(T)残基反应。线性双链DNA分子则不和碳化二亚胺反应，但含有单个碱基错配的DNA则定量地反应。在和碳化二亚胺反应后，将DNA分子在高浓度的聚丙烯酰胺凝胶上分级分离，以便区分修饰和非修饰的片段。Ford等人使用这一技术来定位并纯化DNA序列(由于表型变异和遗传疾病产生了差异)。尽管这一方法在跟踪遗传物质的变异上有用，但这一方法需要许多步骤，需要昂贵的组份，而且这一方法不能提供直接的手段来确定PNA已杂交到样品中的除CNA之外的TNA上。

有些尝试试图确定至少一部份PNA和其它核酸之间的杂交是互补的。一种方法涉及检测转录产物，如果PNA杂交到核酸上足以被从包含在探针中的启动子转录，则产生该转录产物。在授给Dattagupta的美国专利

5, 215, 899 中公开了如何经过发夹探针扩增特定核酸序列, 所说发夹探针在杂交并连接到靶序列上后能够被转录。所说探针包含单链的自身互补的序列, 其在杂交条件下形成具有功能性启动子区的发夹结构, 该探针还包含从发夹序列 3' 末端延伸的单链探针序列。一旦与互补于探针序列的靶序列杂交并且杂交的靶序列的 3' 末端和发夹探针的 5' 末端连接, 在合适的 RNA 聚合酶和适当的核糖核苷三磷酸 (rNTPs) 存在下, 所说靶序列就成为可转录的。通过将所需 TNA 序列与探针杂交, 将 TNA 与 PNA 连接, 将 RNA 聚合酶和 rNTPs 添加到分离的杂交体中, 并使转录进行直到积累所需量的 RNA 转录产物来完成扩增。所说方法一般来说和具体地说涉及使用由单链未配对末端形成的发夹 DNA 来退火靶序列。当靶序列结合后, 就能产生 RNA 转录产物。因此, 所说方法涉及检测二级转录产物而不是用核酸结合装配体直接固定化和/或定位靶序列。CNA 易于结合探针, 互补性的缺乏并不一定干扰 CNA-PNA 杂交体的形成, 所述杂交体会支持不需要的转录产物的形成。

如果互补性的缺乏干扰被限制性内切酶切割的杂交体 CNA-PNA 对的敏感性, 结合到 PNA 上的 CNA 也可被检测。在 Urdea 的美国专利 5, 118, 605 和 4, 775, 619 提供了一些新的检测核酸分析物的方法。这些方法使用具有基本上同源于分析物中兴趣序列的寡核苷酸的多核苷酸。在预定的严格条件下, 杂交的存在与否决定标记从支持物上的释放。多种方法被用来将标记结合到支持物上。当单链或双链裂解后, 标记便可以从支持物释放, 标记的释放可以被检测, 作为样品中是否存在特定的多核苷酸序列的指示。但这一技术具有其缺点, 因为尽管具有错, 只要这一错配在内切酶识别位点之外, CNA-PNA 对仍然可被限制性内切酶切割, 此缺点将导致鉴别 CNA-PNA 杂交体的失败。

另一种方法用分支的 DNA 探针来检测核酸, 授给 Urdea 的美国专利 5, 124, 246 公开了在生化分析中作为扩增子有用的直链或分支寡核苷酸多聚体, 其包含: ① 至少一个和兴趣单链寡核苷酸序列 (TNA) 互补的第一单链寡核苷酸单位 (PNA), 和 ② 多重的第二单链寡核苷酸单位, 其互补于单链标记的寡核苷酸。尽管描述了扩增的夹心核酸杂交和使用多聚体的免疫测定, 但由于 PNA-CNA 间的杂交仍可以发生并导致产生不需要的信号, 该方

法具有局限性。

除了用来鉴别 TNA 的方法外, 用来扩增这一 DNA 的方法也以公开。在授给 Urdea 的美国专利 5, 200, 314 中, 具有分析物序列 (TNA) 的分析物多核苷酸链通过将所说的分析物多核苷酸在杂交条件下与捕获探针 (PNA) 接触在含有多核苷酸的样品中进行检测, 所述捕获探针具有特异于 TNA 的第一结合配偶体 (partner), 和特异于固相第三结合配偶体的第二结合序列。然后经结合配偶体之间的特异性结合将所形成的双链固定化, 并从结合的物种上分离非结合的寡核苷酸。分析物多核苷酸可以也可以不从固相转移, 然后经 PCR 扩增。每个 PCR 引物都具有可以和分析物多核苷酸杂交的多核苷酸区, 至少一个引物还含有可以和固相结合配偶体结合的附加结合配偶体, 扩增的产物经结合配偶体间的特异性结合从反应物混合物分离。并检测扩增产物。尽管 (通过 PCR) 证实特定的核酸已和 PNA 杂交是可能的, 但此种证实费用昂贵并且涉及多个步骤。

至于牵涉到双链核酸和 DNA 结合蛋白相互作用的报道, 用含有单一蛋白结合位点的固定化 DNA 序列用于提纯此单一蛋白的方法已有描述。授给 Kawaguchi 的美国专利 5, 122, 600 公开了包含 DNA 链和载体的 DNA-固定化的小球以及用此小球分离蛋白的方法, 所述 DNA 链具有特异性结合特定蛋白质的碱基序列, 所述载体具有不大于  $50 \mu\text{m}$  不少于  $0.01 \mu\text{m}$  的粒子大小, 并不吸附任何蛋白质, 所说的载体和所说的 DNA 链由化学键相互连接。由于这是一种蛋白质纯化方法, 它既没有公开检测 TNA 的方法, 也没有披露为了检测和定位特定 TNA 序列而使多于一种以上的蛋白质结合到双链核酸上的方法。

在 EP0 453 301 中, 描述了一种用来检测样品中多核苷酸靶的方法。在此方法中, TNA 中的序列由第一和第二 PNA 杂交到 TNA 上来检测。所述第一和第二 PNA 各自具有一段预先形成的双链序列或含有由链延伸形成的双链, 其可以和核苷酸序列特异结合蛋白结合。没有公开也没有指出仅基于 PNA 与 TNA 之间形成双链, 将核苷酸特异性结合蛋白结合到 TNA 和 PNA 之间形成的双链上的方法。

在美国专利 4, 556, 643 中, 公开了一种非放射检测样品中的特定核苷酸序列的方法, 其涉及含有 DNA 结合蛋白特异序列的探针的杂交。但这一公



开既没有教导也没有提示这样一种方法，即仅基于存在于 PNA 中的序列与存在于 TNA 中的序列之间形成双链，将核苷酸特异性结合蛋白结合到 TNA 和 PNA 之间形成的双链上。

### 发明概要

本发明公开了一种经采用探针核酸(PNA)检测特定靶核酸(TNA)序列的方法，所述探针在和 TNA 杂交后，可以和靶结合装配体(TBA)结合。每一个 TBA 结合至少一个 PNA-TNA 杂交对特异性区，靶结合区(TBR)。TBA 由一个或多个分子组成，一个或多个此种分子以特异性和序列或构象依赖性方式结合到 TBR 上。TBA 可含有一个或多个先导(piloting)序列，叫做“PILOTS”或“不对称序列”，其装配并限制 TBA 的核酸结合部份到特定的区域。PILOTS 的作用是按照事先决定的方式形式装配特定核酸识别单位或其它先导物(pilots) (特定核酸识别单位连接到其上)成 TBA。TBA 也可含有一个或多个能够锚定或定位 TBA 的分子。本发明也公开了具有独特特征的新 TBA，所述特征出人意料地不仅使 TBA 作为诊断工具有用，也使其作为预防和治疗化合物有用。本发明所公开的是有关利用 PNA、TBR、TBA 和 TBA PILOTS 的方法和组合体，包括用它们作为诊断及法医试验试剂盒的组份，和用新 TBA 作为预防和治疗剂。

除了 TNA 特异性序列外，PNA 还可以含有一种或几种可以和辅助核酸(BBR)中 1/2 BBR 互补杂交的序列 1/2BBR。通过加入 BNA 到存在于 PNA 中起始 1/2BBR 上的杂交，以 PNA-BNA 的形式，然后以 BNA-BNA 杂交体的形式进行 PNA 的延伸。这些延伸产物可以含有一个或多个辅助结合区(BBR)，每一个 BBR 都具有结合辅助结合装配体(BBA)的能力。BBA 由分子组成，其中的一个或多个分子可以以特异性和序列或构象依赖性方式和 BBR 结合。BBA 可含有一个或多个先导序列，叫做“PILOTS”或“不对称序列”，其装配并限制 TBA 的核酸结合部份到特定的区域。PILOTS 的作用是按照事先决定的方式形式装配特定核酸识别单位或其它先导物(pilots) (特定核酸识别单位连接到其上)成 BBA。BBA 也可含有锚定或定位 BBA 的分子，或者使得可以检测结合的 BBA 并从而检测 TBA-TNA-PNA 复合体(它们依次结合于其上)。本发明所公开的是有关利用 1/2 BBR、BNA、BBR、BBA 和 BBA PILOTS 的方法和组合体，包括用它们作为诊断及法医试验试剂盒的组份。

本发明公开了使用发夹核酸(HNA)作为加帽结构的方法和组合体。所述HNA含有自我杂交区和1/2 BBR单链区,在杂交条件下,其可和PNA中的1/2 BBR或者已结合到PNA上的BNA中的1/2 BBR杂交,终止BNA延伸到PNA或其它BNA上。

本发明公开了用于检测、定位和区分特定核酸序列的试验方法和产生包含PNA、TBA、TBR、BNA、BBR、BBA和HNA的试验试剂盒的方法和组合体,所述核酸序列包括在人细胞、人免疫缺损病毒(HIV)、人乳头状瘤病毒(HPV)和包括病毒和细胞在内的其它含核酸的体系中发现的核酸序列。

因此,本发明的一个目的是提供用于结合、检测和扩增检测样品中特定靶核酸序列(即使在有密切相关但不同的核酸存在下)的具有真实性和准确性的方法和组合体。

因此,本发明的一个目的是提供用于制备特异地和由探针核酸序列和靶核酸序列杂交所形成的靶结合区结合的靶结合装配体的方法和组合体。

本发明的另一个目的是提供用于制备特异地和由辅助核酸序列和探针核酸、辅助核酸和发夹核酸杂交所形成的辅助结合区结合的辅助结合装配体的方法和组合体。

本发明的另一目的是提供涉及发夹核酸的方法和组合体,所述发夹核酸可以控制用于扩增PNA-TNA杂交事件中的特异或非特异延伸的辅助核酸和辅助结合装配体的大小。

本发明的另一目的是提供用于选择、装配和陪伴特定核酸分子(每一个都具有核酸结合的辨别能力)成为靶和辅助结合装配体的方法和组合体。

本发明的另一目的是提供用于用辅助结合装配体和辅助核酸扩增检测结合到靶结合区上的靶结合装配体的方法和组合体。

本发明的另一目的是提供使得可以使用一种或多种可检测标记的方法和组合体,所述标记包括但不限于放射性标记,发光,荧光、酶促或其它信号产生分子。这些标记物在与探针核酸、靶结合装配体、辅助结合装配体、辅助核酸或发夹核酸相关的场合下使用。

本发明的另一目的是提供用于从生物体分离核酸片段的方法,所述核酸片段具有产生探针核酸和对所述片段或生物体唯一的TBA的TBA组份结合位点。

## 附图简要说明

下列说明包含在图 1 中：图 1-I 是含有 1/2 TBR 和 1/2 BBR 的 PNA，其是互补于 TNA 的单链序列；图 1-IIA 是 TNA，图 1-I 的组份添加于其上，在杂交条件下，其结合 PNA 形成图 1-IIIA 的组份 PNA-TNA 杂交体，该杂交体含有至少一个 TBR；图 1-IVA 是 BNA，图 1-IIIA 的组份添加于其上，在杂交条件下，其结合图 1-IIIA 的 1/2 BBR 形成图 PNA-BNA 杂交体，该杂交体含有图 1-VA 所示的 BBR。

图 1-IIB 是 BNA，图 1-I 的组份添加于其上，在杂交条件下，其结合 PNA 形成图 1-IIIB 的组份 PNA-TNA 杂交体，该杂交体含有 BBR；图 1-VIA 是 TNA，图 1-IIIB 的组份添加于其上，在杂交条件下，其结合图 1-IIIB 的 1/2 TBR 形成 PNA-BNA 杂交体，该杂交体含有图 1-VB 所示的 TBR。

图 1-IIC 是 HNA，图 1-I 的组份添加于其上，在杂交条件下，其结合 PNA 形成图 1-IIIC 的组份 PNA-HNA 杂交体，该杂交体含有 BBR；图 1-VIC 是 TNA，图 1-IIIC 的组份添加于其上，在杂交条件下，其结合图 1-IIIC 的 1/2 TBR 形成 PNA-BNA 杂交体，该杂交体含有图 1-VC 所示的 BBR。

所说的形成 TBR 和 BBR 的杂交体在本发明中是有用的，如图 1 所示，PNA 和 BNA 可以包含非连接的支持物和/或指示剂 (OSA)，或者连接的支持物或其它定位方式和/或指示剂，所述定位方式包括但不限于连接到小珠上，聚合物上和表面上；

图 2A 是 BNA 聚合到 PNA 上和由 HNA 加帽的策略的示意图。

图 2B 是经聚合 BNA 和加帽 HNA 来扩增 PNA-TNA 信号的策略的示意图。

图 2C-D 为扩增 PNA-TNA 信号的附加策略，其是通过 BNA 聚合及 HNA 加帽。

图 3 是显示使用每个 BNA 含有多个 1/2 BBR 的 BNA 的示意图。

图 4A 是显示 TBA 和 BBA 与 TBR 和 BBR 结合，以及 TBA 辨别 TNA 和 CNA 的能力的示意图。按照这一实施方案，如果 TBA 固定在小珠、微量滴定平板表面或任何其它这样的表面上，则只有复合体例如复合体 X 被保留和检测，而诸如复合体 XI 之类的复合体不被保留和检测。

图 4B 是解释类似图 4A 所示的但在发生的顺序上略有不同那些事件的示意图。

图 4C 是解释类似图 4A 所示的但在发生的顺序上略有不同那些事件的示意图。

图 5 是解释含有一个 1/2 TBR 和不含 1/2 BBR 的 PNA 到含有高达五个 1/2 TBR 和一个 1/2 BBR 的 PNA 的示意图。每一数码 (I, II, III, IV,

V)的(A)和(B)成员形成一组,其在杂交到TNA上时,给((A)成员)或不给((B)成员)TBR提供可利用的1/2 BBR(用于经杂交到具有互补1/2 BBR的BNA上的扩增)。

图6A是解释含两个1/2 TBR的特定的TNA的示意图,当与PNA结合时,其形成两个密切相关的TBR,这两个TBR可以和两个TBA结合。一个1/2 BBR也被提供来供扩增使用。

图6B是显示与与图6A中相同的事件的示意图,所不同的是使用双TBA,以便在正常细胞样品中出现的单个TBR之间的区别可以与异常的双TBR相区别。

图6C是显示与图6A相同的方案的示意图,所不同的是TNA中的5个TBR被辨明。每一TBR可以与相同或不相同的TBA结合。TBA可以用不同的标记物标记,以便确定TNA中存在五个位点。

图6D是显示与图6C相似的同事件的示意图,所不同的是显示了双TBA,扩展了图6B所示的对双TBA的使用。图6A,6B,6C和6D中的II项是HIV单链DNA或RNA。

图6E是显示与图6C相似的同事件的示意图,所不同的是显示了双TBA,扩展了图6B所示的对双TBA的使用。图6A,6B,6C和6D中的II项是HIV单链DNA或RNA。

图7显示作为TNA的HIV LTR和两个PNA以及用此PNA检测TNA的策略。

图8是本发明的一个实施方案的图解:其中靶结合装配体用于结合杂交体TNA-PNA。辅助结合装配体用于结合聚合的BNA。

图9是模(modular)TBA的图解,其中,装配序列,接头序列和不对称序列用于陪伴核酸识别位点一起形成TBA。

图10显示在HIV-特异性序列检测中有用的模TBA。

图11显示在人乳头瘤病毒序列检测中有用的模TBA。各E2单位实际上是E2 DNA结合部分的双体。

图12A是显示由于TBA的结合,TNA分级分离和迁移率变化的图解。

图12B是显示由于除TBA外的BBA的结合,TNA分级分离和迁移率增加的变化变化的图解。

图13显示检测缺失序列的策略。使用这一策略的例子是人类乳头瘤的整合测定。

图14显示经采用核酸识别单位、接头、装配和不对称序列装配高级TBA,以便形成对HIV LTR中结合位点特异性的各种靶结合装配体。

图 15 显示经采用 DNA 识别单位、接头、装配和不对称序列装配高级 TBA，以便形成对 HPV 基因组中结合位点特异性的各种靶结合装配体。

图 16 显示由使用复合 TBA 获得的区别，内源竞争靶结合分子消除 TBA 结合到表亲核酸上的(但不是从含有多于一个被 TBA 识别的位点的合适的取向的 TNA 上消除)能力。

图 17 显示 TBA 被特异性地靶引以结合到序列错配位点上(并且优先结合在不含有所有的靶错配的表亲位点上的那些位点)的能力。

### 序列简要描述

SEQ ID NO. 1 相应于图 5-IA-1，显示第一类 MHC NF-kB 结合位点。

SEQ ID NO. 2 相应于图 5 (IA)，显示 B2 微球蛋白 NF-kB 结合位点。

SEQ ID NO. 3 相应于图 5 (IA)，显示 K 免疫球蛋白 NF-kB 结合位点。

SEQ ID NO. 4 相应于图 5 (IA)，显示一个 HIV NF-kB 结合位点。

SEQ ID NO. 5 相应于图 5 (IA)，显示一个 HIV NF-kB 结合位点。

SEQ ID NO. 6 相应于图 5 (IA)，显示 c-myc NF-kB 结合位点。

SEQ ID NO. 7 相应于图 5 (IIA)，显示双 HIV NF-kB 结合位点。

SEQ ID NO. 8 相应于图 5 (IIA)，显示双 HIVNF-kB 结合位点。

SEQ ID NO. 9-16 相应于图 5 (IIA)，显示双结合位点，其一为 HIV NF-kB 结合位点，另一为 HIV SP1 结合位点。

SEQ ID NO. 17-18 相应于图 5 (IIA)，显示双 HIV SP1 结合位点。

SEQ ID NO. 19-31 相应于图 5 (IIIA)，显示双 HIV NF-kB 结合位点和 HIV SP1 结合位点。

SEQ ID NO. 32-33 相应于图 5 (IVA)，显示四结合位点，其中两个位点为 HIV NF-kB 结合位点，另两个为 HIV SP1 结合位点。

SEQ ID NO. 34 相应于图 (VIA)，显示 5 个结合位点，其中两个为 HIV NF-kB 结合位点，另三个为 HIV-SP1 结合位点。

SEQ ID NO. 35 是 1/2 BBR 的例子，在此情况下 OL1，OL2 和 OL3 元件是  $\lambda$  噬菌体左操纵子，包括插入序列。

SEQ ID NO. 36 是 1/2 BBR 的例子，在此情况下 OR3，OR2，和 OR1 元件是  $\lambda$  噬菌体右操纵子，包括插入序列。

SEQ ID NO. 37 是 HIV LTR。

SEQ ID NO. 38 是互补于 HIV LTR 之 PNA 的 PNA。

SEQ ID NO. 39 是互补于 HIV LTR 的不同于 SEQ ID NO. 38 的 PNA 的 PNA。

SEQ ID NO. 40 是互补于 HIV LTR 的一部分的 PNA，其也含有 1/2 BBR 和使 BNA 聚合到 PNA 上的悬挂序列。

SEQ ID NO. 41 是互补于 SEQ ID NO. 40 1/2 BBR 的 BNA。

SEQ ID NO. 42 是一个 BNA，它将聚合到 SEQ ID NO. 41 BNA 上，并且其和 SEQ ID NO. 40 和 41 一起产生一个 PstI 识别位点。

SEQ ID NO. 43 是一个 BNA，其互补于 SEQ ID NO. 42 BNA，并且其完成一 BamHI 识别位点。

SEQ ID NO. 44 是一 HNA，其具有 BamHI 识别位点(它会与另一由 SEQ ID NO. 42 和 43 产生的识别位点杂交，产生聚合物)。

SEQ ID NO. 45 是另一 PNA，与 SEQ ID NO. 40 相似，互补于 HIV LTR 的一部分。但不是与 SEQ ID NO. 40 相同的序列。SEQ ID NO. 45 也编码一 1/2 BBR 序列和一突出端，其使得用 SphI 识别位点开始 BNA 的聚合。

SEQ ID NO. 46-62 是人类乳头瘤病毒(HPV)特定 PNA，一旦与 HPV 序列杂交，即形成和 HPV DNA 结合蛋白结合的 TBR。

SEQ ID NO. 63-71 是整合进 TBA 中的 NF- $\kappa$ B DNA 识别单位。

SEQ ID NO. 72 是一核定位序列。

SEQ ID NO. 73 是一 SP1 序列识别单位。

SEQ ID NO. 74 是一 TATA 结合蛋白识别单位。

SEQ ID NO. 75-84 是乳头瘤 E2 DNA 的识别单位。

SEQ ID NO. 85-92 是不对称序列。

SEQ ID NO. 93 是一 arabidopsis TATA 结合蛋白识别单位。

SEQ ID NO. 94 是一 HPV-16-E2-1 DNA 结合蛋白识别单位。

SEQ ID NO. 95 是一 HPV-16-E2-2 DNA 结合蛋白识别单位。

SEQ ID NO. 96 是一 HPV-18-E2 DNA 结合蛋白识别单位。

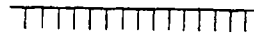
SEQ ID NO. 97 是一 HPV-33-E2 DNA 结合蛋白识别单位。

SEQ ID NO. 98 是一牛乳头瘤 E2 DNA 结合蛋白识别单位。

SEQ ID NO. 99-102 是例证性的接头序列。

- SEQ ID NO. 103 是一例证性核定位信号序列。
- SEQ ID NO. 104-108 是例证性陪伴序列。
- SEQ ID NO. 109-116 是例证性装配 TBA 序列。
- SEQ ID NO. 117 是共有 NF-kB 结合位点。
- SEQ ID NO. 118 是 HIV Tat 氨基酸序列。

缩 写



单链核酸



双链核酸



核酸上的结合区



无支持物和指示剂，或固体支持或其它定位方式，包括但不限于连接到小珠上，聚合物上和表面上和/或指示剂上=OSA。

- BBA 辅助结合装配体
- BBR 辅助结合区
- BNA 辅助核酸
- CNA 表亲核酸
- 1/2 BBR 单链区，当其与 HNA 或 BNA 的互补序列杂交时，可以结合 BBA
- 1/2 TBR PNA 之单链区，当其与 TNA 的互补序列杂交时，可以结合 TBA
- OSA 可有可无的支持物或连接
- PNA 探针核酸
- TBA 靶结合装配体
- TBR 靶结合区
- TNA 靶核酸

## HNA 发夹核酸

## 定 义

应当理解, 在以下所公开的内容中, 当提到靶结合装配体(TBA)、辅助结合装配体(BBA)、DNA结合蛋白、核酸结合蛋白或RNA结合蛋白这样一些术语时, 意指包含结合到DNA或RNA靶核酸序列(TNA)上的分子的组合体, 而不考虑其所来源的结合分子的类别的特殊性。这样, 例如, 采用来与人类免疫缺损病毒序列结合的TBA可以最类似于NF-KB转录因子, 该因子典型地与DNA结合。但是, 如本文所用的, 人们应该理解可以使TBA最适地用于和特定序列组合体或构象的RNA序列结合。

本发明公开的检测方法的真实性在很大程度上取决于TBA和BBA与特定核酸基元的选择性结合。从本文的公开内容应该理解, TNA的TBA及BBA和相关序列的区别(表亲核酸或CNA)的基础可以是形成精确核酸探针(PNA)-靶核酸(TNA)杂交体片断(PNA-TNA杂交体)。然而, 区别的基础也可以仅仅是特定构象的形成, 同时可以不需要PNA-TNA杂交体中错配碱基对的完全缺失。因此, 应该完全理解TBA或BBA操作的基础完全依靠TNA-PNA杂交体独有的任何特性, 这种特性不同于任何PNA-CNA杂交体显示的特性, PNA-CNA杂交体可以在和给定的PNA接触的试验样品中形成。

## 发明的详细描述

本发明提供了通过利用掺入特定核酸结合蛋白的靶结合装配体(TBA)特异性地鉴定样品中靶核酸(TNA)的方法。通过利用对给定TNA序列特异性的探针核酸(PNAs)和对所说的双螺旋靶结合区(TBR)特异性的TBA(靶结合区是在杂交体TNA-PNA序列形成的基础上形成的), 形成稳定的TBA-TNA-PNA复合体。此外, 通过提供PNA中的特异的可扩增序列, 除特异性地对由TBA识别的TBR的形成有贡献的序列外, 检测PNA和TNA的结合并扩增这种检测。为了这个目的, 利用了大量的核酸扩增系统的任何一个, 包括聚合酶链式反应或利用分枝DNA技术, 其每一个分支上都含有可检测的标记。具体地说, 本文描述了新型扩增方法, 在这种方法中, PNA的可扩增的部分含有这样的序列, 辅助核酸(BNA)在其上可以发生聚合。依据每一个BNA-PNA杂交体的形成, 形成了辅助结合区, 其和辅助结合装配体特异性结合。如果存在可检测标记, TNA-PNA提供必要的无条件的原始TNA-PNA结合事件



的扩增。

按照本发明，应把 TNA 理解为含有特定核酸序列。所说的 TBA 应理解为可特异而紧密地和形成的 TNA-PNA 杂交体结合的任一分子装配体。TBA 包括一个或多个分子，其序列对结合到 TBR 上是足够的。已知的 DNA 结合区，既可以直接用作所说的 TBA 的组分，也可以根据本文提供的教导进行修饰。最容易得到的分子是 DNA 结合蛋白的 DNA 结合区。已知许多 DNA 或 RNA 结合蛋白既可以直接用作已知的未修饰的蛋白，或所说的 TBA 也可以是根据本文提供的特殊教导修饰后核酸结合蛋白。在后一种情况下，所需要的特殊修饰应该包括优化结合亲和力、去除不必要的活性(如核酸酶活性和在其它和 TBA 组分有亲和力的分子存在的情况下所说的 TBA 的识别力)、优化靶序列超过密切相关的序列的选择性、优化稳定性。

根据本发明可以利用的 DNA 结合蛋白的例子是转录因子 NF- $\kappa$ B(p50 和 p65)的结合蛋白，NF-IL6，NF-AT，rel，TBP，乳头状瘤病毒 E2 蛋白，sp1，噬菌体 $\lambda$ 的 cro 和 C1 阻遏蛋白和其它已知的蛋白(其 DNA 结合蛋白已经分离、克隆、测序、定性)。此外，任何其它的 DNA 结合蛋白或作为与 TBR 杂交体或 BBR 结合的充要条件的部分蛋白都包括在内。这包括蛋白或具有改变的 DNA 结合活性的野生型蛋白或蛋白的部分，以及产生的具有改变的 DNA 结合特异性的蛋白，例如，DNA 结合识别螺旋从一个蛋白到另一个蛋白的互相交换。此外，具有核酸结合及其它核酸功能的蛋白，如限制性内切酶，也可以用作所说的核酸结合的功能物。在 DNA-RNA 杂交体及 RNA-RNA 杂交体中和靶区结合的蛋白都包括在内(见，例如 Shi 1995, DeStefano 993, Zhu 1995, Gonzalcs 1994, Salazar1993, Jaishree 1993, Wang 1992, Roberts 1992, Kainz1992, Salagar 1993(6))。利用陪伴结合装配体部分的分子可以构建结合装配体，以便获得特定组分的组合(combination)和几何构象。这一分子指定为 PILOT。Pilot 可以由蛋白质或任何有机或无机材料组合而成，这些有机或无机材料的组合在 TBA 或 BBA 成员之间获得组合选择和/或诱导特定几何构象。陪伴分子是一稳定构架，TBA 或 BBA 依靠它组建，这样提供了 TBA 或 BBA 的正确构象，同时除去了天然产生的核酸结合蛋白的不需要的特性。此实施方案的具体例子是，以修饰的噬菌体 $\lambda$  cro 蛋白作为陪伴分子，提供了所说的多效转录因子 NF- $\kappa$ B 的修饰体。每一

个 NF-kB 结合二聚体保留对 NF-kB 结合位点的微小亲和力，同时结合装配体显示了有利于制造、稳定和特异的特征。

鉴于以上所述，下文详细地描述了本发明的各方面和实施方案。

1. 探针核酸(PNA)及其制备: 本发明的 PNA 至少包括三个连接在一起的主要部分: 参见附图 1(I), 所说的 PNA 的第一部分是一个或多个碱基序列, 指定为“1/2 TBR”。参见图 1(I 和 IIa), PNA 中的 1/2 TBR 与样品中兴趣序列互补, TNA 包含一个 1/2 TBR。参见附图 1(IIIa), 在杂交条件下, 当 TNA 加入到 PNA 中时, 形成了含有一个 TBR 的 PNA-TNA 杂交体。参见附图 1(I), 所述 PNA 的第二部分是一碱基序列, 指定为“1/2 BBR”。参见图 1(I, IIb, IIc 和 IVa), PNA 的 1/2 BBR 与 BNA 或 HNA 中的 1/2 BBR 互补。参见附图 1(IIIb, IIIc 和 Va), 在杂交条件下向 PNA 中加入 BNA 或 HNA, 分别形成含有 BBR 的 PNA-BNA 杂交体或 PNA-HNA 杂交体。参见附图 1(I), 所述 PNA 的第三部分是 OSA, 由方框中一个圆圈代表。OSA 不是支持物和/或指示剂、或者是固体支持物或其它的定位方式和/或指示剂, 所述定位方式包括但不限于连接到小珠, 聚合物和表面(这些物质共价或非共价地结合到 PNA 上, 但特异性地与 PNA 相关)上。OSA 可以是帮助分离和/或定位的原子或分子, 如可通过各种物理手段检测的固体支持物结合基或标记, 这些物理方法包括但不限于释放的粒子或光子的吸收或成象。将指示剂连在寡聚核苷酸上或用固体支持物固定寡聚核苷酸的方法是本领域公知的(见 Keller 和 Manak, *Supra*, 本文一并参考)。

本发明所说的 PNA 可用任何合适的方法制备。通常这种方法包括寡聚核酸的合成和可复制载体的克隆。核酸合成方法在本领域是广为人知的。在合成或克隆时, 有必要进行链纯化和分离以便该产品用作纯化的 PNA。制备 RNA 探针的方法是公知的(参见 Blais(1993), Blais(1994)的例子, 该方法使用体外转录, 采用结合了 T7 RNA 聚合酶启动子 PCR 反应)。

本领域技术人员应理解所说的 PNA 的长度和特定序列依赖于 TNA 中被检测的长度和序列, 以及获得紧密特异结合所使用的特定 TBA 的结构(参见以下有关 TBA 的讨论)。一般来说, PNA 的序列长度在 10 到 300 个核苷酸之间是足够的, 对本文所列举的实施方案, 需要具有 15-100 个核苷酸的长度。

应该理解可以组建这种 PNA, 目的是使之含有多于一个的 1/2 TBR 和产

生一个或多个 TBA、相同或不同的 TBR，以及在 PNA 和 TNA 杂交体基础上被新型双螺旋或多螺旋 TBA (参见以下有关这些新 TBA 的描述) 识别的复合体 TBR。图 5 具体说明了含有一个或多个 1/2 TBR 的具体 PNA。相应于图 5 (Ia, IIa, IIIa, IVc 和 IIa) 显示的 1/2 TBR 的具体序列见 SEQ ID NOS. 1-34 (见前面描述的序列)。

如图 2a 和图 2b 所示, 包括 1/2 TBR 的 PNA 可以和一个或多个 BNA 杂交 (见以下描述), 同时这种聚合 BNA 可以是任何需要的可能的长度, 此链用于 TNA-PNA 杂交事件的扩增。优选的是, 在这种 PNA 中存在约 0 到约 10 个 1/2 BBR。

如图 6a 和 6b 所示, 所说的 PNA 可以包含几个相同或不同的 1/2 TBR, 其可以与 TNA 中的几个 1/2 TBR 杂交。每次一个 PNA 中的 1/2 TBR 与 TNA 中 1/2 TBR 配对, 即形成一个靶结合区 (TBR), 其可以结合一个 TBA。并且, 没有必要所有的 TBR 都在单链的连续的 PNA 上。这样, 在本发明的一个实施方案中, 采用两种不同的 PNA 来检测特定的 TNA 序列。作为本发明这方面的一个说明, 图 7 是人体免疫缺陷病毒 (HIV) 的长末端重复序列 (LTR) 的代表。本领域已经知道, HIV LTR 包括两个非常接近 NF- $\kappa$ B 结合位点和三个 SP1 结合位点, 其中的 NF- $\kappa$ A 和 SP1 已知是 DNA 结合蛋白。图 7 包括两个 PNA, PNA1 (SEQ ID NO. 38) 和 PNA2 (SEQ ID NO. 39), 每个都与所示的 TNA (SEQ ID NO. 37) 的反链互补, 其包含两个 HIV LTR 的 NF- $\kappa$ B 结合位点和三个 SP1 结合位点。根据本发明的这一方面, PNA1 与图 7 所表示的 TNA 下划线的用 “+” 标记的碱基特异杂交, PNA2 与图 7 所表示的 TNA 下划线的用 “=” 标记的碱基特异杂交。PNA1 或 PNA2 都包含可以与 BNA 的 1/2 BBR 序列 (见下文) 杂交的序列 (序列上用 “#” 或 “\*” 标记)。此外, PNA1 或 PNA2 都可以用 OSA 进行不同的标记, 比如用荧光素或罗丹明等荧光团, 其可以让两种探针都与 TNA 确切地结合。如果只发现一个标记或两个标记都没发现, 则推断这个 TNA 在检测的样品中不存在。

图 7 所示实施方案的另一个方面, 显示了一种改变本测定特异性的方法。通过改变 PNA1 和 PNA2 之间缺口的长度, 这样就改变了仍未杂交的 TNA 区, 实施本发明, 可以改变这种测定的分辨性。

为了更清楚地举例说明本发明的这个方面, 有必要强调所说的 TBR 可以

具有螺旋结构。这样，当 PNA1 在螺旋的一面产生出 TBR，PNA2 在螺旋的同一侧或不同侧产生出又一 TBR，视两个 TBR (图 7 下划线部分) 中间部位的距离而定。如果每个结合位点中间都是相距 10.5 碱基的完整产物，TBR 就会在螺旋的同一侧；如果是相距 10.5 碱基非完整产物，那么 TBR 就会在螺旋的两侧。在这种方式中，任何由 TBA 识别 PNA1 TBR 以及 TBA 识别 PNA2 TBR 在结合上的协调性都是可以控制的 (见 Hochschild, A., M. Ptashne [1986] 细胞 44: 681-687, 显示出噬菌体  $\lambda$  阻遏物和一个 DNA 螺旋上彼此定位在不同距离的两个不同操纵子位点的结合效应)。如 Perkin 等 ([1993] *EMBO J.* 12: 3551-3558) 所述, NF- $\kappa$ B 和 SP1 位点的协调对 HIV LTR 的激活是必要的。但是, 从本发明的目的来说, 可以通过提供能和两种位点自发结合的单一、新型结合蛋白, 来利用 HIV LTR 上的两个 NF- $\kappa$ B 和三个 SP1 结合位点标志, 但必须在两个位点的间距几何学上可行的条件下。这一点受所选择的 TBA 和使用的 PNA 的结构限制。这样, 在图 7 所举的实施例中, 可以使用两个探针, 这两个探针之间有一个足够大的单链 DNA 区, 即使 NF- $\kappa$ B 和 SP1 结合位点在螺旋的两个不同侧面, 两个探针间的这个单链区提供一个足够灵活的“铰链区 (hinge)”, 使 DNA 可以同时弯曲和扭曲以容纳几何构型的 TBA。另外, 可以通过缩短探针间距离来设计一个更严格的检测方法, 这样 DNA 只能弯曲, 不能扭曲。最后, 探针间距可以很近, 或使用单一 PNA, 所以 DNA 仅能弯曲, 无法扭曲。这样, 对任何给定的需要的靶核酸间的分辨程度, 此图举例说明了并使得可以产生检测系统, 所述靶核酸具有相似的序列, 但这些序列具有不同的并列位置。

作为一种诊断和法医学鉴定 HIV 的试剂盒, 本领域的技术人员应该懂得前面所述的关于本发明的各方面允许选择诊断和法医学鉴定试剂盒的组份以使其在任何时候与当时所知的流行爱滋病毒菌株或其它病原菌或疾病状况相匹配。本领域的技术人员还应该意识到, 检测爱滋病的感染不是本发明的独有用途; 由于爱滋病毒基因的突变性, 对于这样的诊断, 基因的突变是最复杂的实验环境因素之一。恰恰是在这种易变的条件中, 本发明方法的灵活性连同其能区别非常紧密相关的序列的能力是最值得注意的。在突变较少的环境中, 不必实施本发明的那些复杂方法, 这些方法是经得起检验的。这样, 用于诊断乳头状瘤病毒感染的试剂盒中, 所有 TBA-

TBR 相互作用的鉴别特性都可以得到, 其也具有利用 BNAs 和 BBA 来扩增信号的能力, 但是也可以利用一个单一简单的 PNA, 如 SEQ ID NO. 46-62 中任何一个, 其鉴别独有的乳头状瘤病毒序列, 这种序列和 TBA 的结合也是公知的, 例如乳头状瘤病毒 E2 蛋白或其变形 DNA 结合蛋白(参见 Hegde 等 [ 1992 ] 自然 359: 505-512; Monini 等 [ 1991 ] 病毒学杂志 65: 2124-2130)。

在使用本方法检测一个特定 TNA 以估价是否存在特定核酸, 这种核酸与黑素瘤、肝细菌瘤、乳腺癌、子宫颈癌、肺癌、直肠癌、前列腺癌、胰腺癌或卵巢癌相关, TNA 可以从怀疑含有癌细胞的器官和液体的活体组织材料中获得。为了检测基因的缺失, TNA 可以从含有感染细胞的病人样品中获得。为了检测发酵中的污染物和在生产食品, 化学试剂或生物工程产品或废物生物处理过程中的产品, TNA 样品可以从发酵或处理过程中不同阶段获得。为了检测食品和药品的病原体或污染物, TNA 样品可以从如下方面获得: 食品或药品; 清理食物所用的工具或与食物相接触的表面; 与食物接触的液体; 加工材料; 与食品、药物或生物样品的生产相关或与其接触的液体或其它物质, 这些物质从与食品、药物或其它产品相接触的物质中获得。

## 2. 辅助核酸(BNA), 辅助结合区(BBR)和它们的制备

本发明的 BNA 包括至少一个或多个与 OSA 结合的 1/2 BBR。所说的 1/2 BBR 可以与包含在 PNA、其它 BNA 或 HNA 中的 1/2 BBR 杂交。

参考附图 1(I、IIb 和 IIIb), 最简单的 BNA 包括二部分。参见附图 1(IIb), 最简单 BNA 的第一部分是一段与所说的 PNA 中命名为“1/2 BBR”的序列互补的碱基序列。参考附图 1(IIb), 最简单 BNA 的第二部分是 OSA。OSA 不是支持物和/或指示剂、或者是固体支持物或其它的定位方式和/或指示剂, 所述定位方式包括但不限于连接到小珠, 聚合物和表面(这些物质共价或非共价地结合到 PNA 上, 但特异性地与 PNA 相关)上。

参考附图 2a(II 和 III), 所说的 BNA 可以含有一个以上的 1/2 BBR 序列。图 3(II)所示的 BNA 包括一段与图 3(I)所示的 PNA 及两段其它 1/2 BBR 序列互补的序列。图 3(III)所示的 BNA 含有两个 1/2 BBR 序列, 它们与图 3(II)所示的 BNA 上两个 1/2 BBR 序列互补, 加至“n”个额外 1/2 BBR 用于和

额外的 BNA 聚合。

在杂交的条件下，图 3(II) 所示的 BNA，当与图 3(I) 所示的 PNA 组合在一起时，产生了一个如图 3(IVa) 所示 PNA-BNA 杂交体，这一杂交体含有一个 BBR 和具有二额外 1/2 BBR 或“辅助”序列的非杂交突出端(extension)。通过所说的杂交产生的 BBR 在序列上可以相同，相似或不同。通过所说的杂交而产生的 BBR 可以结合相同的、相似或不同的 BBA(见下面)。BNA 的制备和 PNA 的制备相似。

在杂交条件下，图 3(IVb) 所示的 BNA-BNA 杂交体，当与图 3(Vb) 所示的 PNA 结合后，产生如图 3(VI) 所示的 PNA-BNA 杂交体，其包含一个 BBR，两个额外的含有 BBR 的 BNA-BNA 杂交体，和一个带有额外 1/2 BBR 序列的非杂交突出端，一个“辅助”序列。通过所说的杂交产生的 BBR 可以在序列上相同，相似或不同。通过所说的杂交产生的 BBR 可以结合相等，相似或不同的 BBA(见下面)。BNA 可以以类似于制备 PNA 的方式制备。

### 3. 靶核酸(TNA)和它们的制备

按照本方法，检测和扩增通过检测特定 TNA 而产生的信号的第一步是在一个合适的混和液中将 PNA 与这一靶物质杂交。这种杂交是在本领域公知的合适的条件下完成的。

可能或已知含有预计的 TNA 的样品可以由各种来源获得。它可以是生物样品、食物或农业样品、环境样品等等。使用本方法检测用于医疗诊断或法医学检测的特定 TNA 时，TNA 可以从活体组织样品，体液或尿液、血液、牛奶、脑脊髓液、唾液、颌下腺液、粪便、肺抽提液、喉或生殖器官样品等的提取液。另外，检测可以在体内进行(见例 Embteton 1993, Patterson 1993, Adams 1994)。

因此，对脊椎动物(包括哺乳动物和人)或者对下列任何一种或所有兴趣微生物特异性的 PNA 可以按照本发明的方法设想和使用：

棒杆菌属

白喉棒杆菌

芽孢杆菌属

苏云金芽孢杆菌

肺炎球菌属

- 肺炎双球菌
  - 链球菌属
    - 酿脓链球菌
    - 唾液链球菌
  - 葡萄球菌属
    - 金黄色葡萄球菌
    - 白色葡萄球菌
  - 假单胞菌属
    - Pseudomonas stutzen*
  - 奈瑟氏球菌属
    - 脑膜炎奈瑟氏球菌
    - 淋病奈瑟氏球菌
  - 肠杆菌属
    - 大肠杆菌
    - 产气气杆菌
    - 肺炎克雷伯氏菌
    - 伤寒沙门氏菌
    - 猪霍乱沙门氏菌
    - 鼠伤寒沙门氏菌
    - 痢疾志贺氏菌
    - 施氏志贺氏菌
    - 阿糖迟酵志贺氏菌
    - 弗氏志贺氏菌
    - 鲍氏志贺氏菌
    - 索氏志贺氏菌
  - 其他肠杆菌
    - 普通变形菌
    - 奇异变形菌
    - 摩氏变形菌
    - 铜绿假单胞菌
- 大肠肝菌状的细胞
- 沙门氏菌
- 志贺氏菌
- 变形杆菌类

粪产碱菌

霍乱弧菌

嗜血杆菌-博德特氏菌属组

流感嗜血菌, 杜克雷氏嗜血杆菌

*Hemophilus hemophilus*

埃及嗜血菌

副流感嗜血菌

百日咳博德特氏菌

巴斯德氏菌属

鼠疫巴斯德氏菌

土拉巴斯德氏菌

布鲁氏菌属

马尔他布鲁氏菌

流产布鲁氏菌

猪布鲁氏菌

需氧芽孢形成细菌

炭疽芽孢杆菌

枯草芽孢杆菌

巨大芽孢杆菌

蜡状芽孢杆菌

厌氧芽孢形成细菌

肉毒梭菌

破伤风梭菌

产气荚膜梭菌

诺氏梭菌

败毒梭菌

溶组织梭菌

第三梭菌

双酶菌

生孢梭菌



分枝杆菌属

结核分枝杆菌

牛分枝杆菌

鸟分枝杆菌

麻疯分枝杆菌

副结核分枝杆菌

放线菌属(类真菌细菌)

衣氏放线菌

牛型放线菌

内氏放线菌

星状诺卡菌

巴西诺卡氏菌

螺旋体

苍白密螺旋体

极细密螺旋体

斑点病密螺旋体

减少螺菌

念珠状链杆菌

回归热疏螺旋体

出血黄疸钩端螺旋体

犬钩端螺旋体

锥虫

支原体

肺浆支原体

其他病原体

单核细胞增生利斯特氏菌

猪丹毒丹毒丝菌

念珠状链杆菌

肉芽多诺万氏菌

杆状巴尔通氏体

## 立克次氏体族 (类细菌寄生虫)

普氏立克次氏体

莫赛氏立克次氏体

立氏立克次氏体

康纳次氏立克次氏体

澳大利亚立克次氏体

西伯利亚立克次氏体

螨立克次氏体

*Rickettsia tsursugamushi*

伯耐梯氏立克次氏体

五日热立克次氏体

## 衣原体属 (不能归类的寄生细菌/病毒)

衣原体属病原体 (命名不确定)

## 真菌

新型隐球酵母

皮炎芽生菌

荚膜组织胞浆菌

粗球孢子菌

巴西芽生菌

白假丝酵母

烟曲霉

丛生毛霉 (伞枝犁头霉)

米根霉

少根根霉

藻状菌纲

变黑色根霉

申克氏孢子丝菌

*Flonsecaea pedrosoi**Fonsecaea compact**Fonsecacae dermatidis*

腐肉枝孢

疣状瓶霉

构巢曲霉

足肿马杜拉放线菌

灰色马杜拉放线菌

波伊德氏霉样真菌

甄氏瓶霉

石膏状小孢子菌

须发癣菌

*Keratinomyces ajelloi*

狗小孢霉

深红色发癣菌

头癣小孢霉

病毒

腺病毒

疱疹病毒

单纯疱疹

水痘(小鸡痘)

带状疱疹(Shingles)

病毒B

细胞肥大病毒

痘病毒

天花(天花)

牛痘

牛囊尾蚴痘病毒

副牛痘

触染性软疣

细小核糖核酸病毒

脊髓灰质炎病毒

柯萨奇病毒

Echoviruses

## 鼻病毒

### 粘液病毒

流感 (A, B 和 C)

副流感 1

流行性腮腺炎病毒

纽卡斯尔疾病病毒

麻疹病毒

牛痘病毒

犬瘟热病毒

呼吸合胞体病毒

风疹病毒

### 节肢介体病毒

东方马脑炎病毒

西方马脑炎病毒

Sindbis 病毒

Chikungunyavirus

赛姆利基森林病毒

Mayora 病毒

圣路易斯脑炎病毒

加利福尼亚脑炎病毒

科罗拉多蜱热病毒

黄热病病毒

Dengue 病毒

### 呼吸道肠道病毒

呼吸道肠道病毒 13 型

### 逆转录病毒

人类免疫缺损病毒 (HIV)

人类 T-细胞淋巴营养病毒 III (HTLV )

### 肝炎

肝炎 A 病毒

肝炎 B 病毒

肝炎非 A-非 B 病毒

肝炎 C, D, E

肿瘤病毒

劳舍尔氏白血病毒

Gross 病毒

马隆尼氏白血病毒

人类乳头瘤病毒

本领域的技术人员会理解, 一般需要以这样一种方式来处理可能含有特定 TNA 样品, 即产生和所说的 PNA 容易杂交的片断。有必要对实验样品进行处理, 以便完成用于杂交的 TNA 的释放或提取用于杂交的 TNA, 如通过把血液样品或其它细胞露置于低渗透压条件中, 或者采取别的更为剧烈的方法破碎样品。当认为 TNA 是以双链形式存在时, 自然需要通过本领域公知的方法分开双链, 使 TNA 能以单链形式进行杂交, 这些方法包括但不限于加热或有限度地露置在碱环境中, 可以通过另外加入单链 PNA 而中和碱并使得可以发生杂交。制备 RNA 靶的方法是公知的(参见 Waterhouse1993 和 Mitchell1992)。

含有 TNA 的核酸片段通常需要降低样品的粘度并增加 TNA 和 PNA 的可接触性。通过本领域任意的和特定的方法得到这种片断。这样, 例如, 已知以特定的频率在分析的特异基因组中切割的核酸酶, 可以用来产生一种已知平均分子量的片段。此外, 其它的核酸酶, 磷酸二酯酶, 核酸外切酶和内切酶, 物理切碎或超声波处理均是用于此目的的可行方法。这些方法是本领域公知的。为了获得 DNA 片段, 一般优选使用限制性内切酶。然而也可以通过各种化学方法获得 DNA 片断, 例如使用下列类型的试剂: EBTA-FE(II) (见 Stroebel 等(1988), 美国化学会杂志, 110: 7921, 或 Dervan(1986) 科学, 232: 464); CU(II)-菲咯啉(见 Chen 和 Sigman(1987), 科学, 237: 1197); IIS 级限制性内切酶(见 Kim 等(1988), 科学, 240: 504); 杂交 DNA 酶(见 Corey 等.(1989), 生物化学, 28: 8277); 争光霉素(见 Umezawa 等, (1986), 抗生素杂志, (Tokyo) Ser. A. 19: 200); neocarzinotatin(见 Goldberg 等.(1981), 癌

症研究第二届 Bristol-Myers 会议, 学术出版社, 纽约, p.163), 以及 methidiumpropyl-EDTA-Fe(II) (见 Hertzberg 等. (1982). 美国化学会杂志, 104: 313)。一般需要用蛋白酶处理除去蛋白质, 有效地除去核酸样品之中的蛋白又不损伤核酸的方法也是本领域公知的。

本发明的 TNA 应该有足够长度, 以便在 TBR 两侧有足够数量的双链杂交体, 这样, TBA 能够免受不连接的片段末端的干扰而进行结合。一般地, 大约 10 到大约 100, 000 个核苷酸范围内的片断, 优选的是大约 20 到大约 1000 个核苷酸范围内的片断, 作为 TNA 片段的平均大小。能被检测的特定 TNA 序列的例子是和本文描述的 PNA 序列互补的序列, 这种 PNA 序列用于检测正常细胞的、不正常细胞的(如有激活的致癌基因、整合的外源基因、遗传学上具有缺陷的基因)、以及病原特异性的核酸序列, 由于这种序列的特异性核酸结合蛋白是已知的, 或可根据本文描述的方法产生。参见图 7, 一种和 HIV 相关的特异 TNA 示于 SEQ ID NO. 37 中。

#### 4. 用于 BNA 的 PNA 的突出端及其制备和信号扩增

在杂交条件下, 可以加入 BNA, 其与 PNA、PNA-BNA 杂交体、BNA 和/或 BNA-BNA 杂交体杂交。可以以非媒介聚合物方式或以媒介的方式用已知顺序的 BNA 完成前面所述的加入。

参见图 2a, 给出了一个简单的辅助。通过加入如 2a (Ib 和 Ic) 所示的两种 BNA 产生了辅助聚合物, BNA 在杂交条件下与 PNA 结合, 形成 PNA-BNA-BNA 杂交体, 这种杂交体含有 PNA 和“辅助”突出端(如图 2a (IIa, IIb, IIc 和 IId) 所示), 留下至少一个未配对的 1/2 BBR 序列。每个未配对的 1/2 BBR 序列(如图 2a (IIa, IIb, IIc 和 IId) 所示), 可与添加的 BNA 杂交形成“辅助”突出端。每个未配对的 1/2 BBR 序列(如图 2a (IIa, IIb, IIc 和 IId) 所示), 又可以与添加的 HNA(如图 2a (IIIa 和 IIIb)) 杂交。不能与添加的 BNA 杂交的 HNA, 其杂交作用覆盖了 BNA 加入到 PNA 上, 如图 2a (IVa, IVb, IVc 和 IVd) 所示。

参照 2b 所示, 有可能控制和指定 PNA 突出端的顺序和组份。如果需要单一的 BBR, 可把包含一个与 PNA 中的 1/2 BBR 互补的序列的 HNA 加入到 PNA 中, 这样就可产生一个单独的 BBR 并“覆盖”任何 PNA “辅助”突出端。如果把添加的 BBR 加入到 PNA 中, 就可以形成 PNA 的受控制的突出端。

参照 2b 所示, 给出了一个简单的辅助。媒介(vectorial)聚合物的突出端可采用加入对 PNA (如图 2B (IA 和 IB) 所示) 特异性 BNA 完成, 这种 BNA 可在杂交的条件下与 PNA 杂交, 形成 PNA-BNA-BNA 杂交体, 由 PNA 和“辅助”突出端组成。如果用 OSA 标记, 这种突出端可提供用于极大地扩增任一信号的方法, 这种信号是基于样品中 PNA 和 TNA 结合而产生的。并且, 通过聚合物中的标记的 BBA 与 BBR 结合, 还可完成附加扩增。

可以用许多方法制备 BNA, 包括, 如通过化学合成方法或重组 DNA 产生方法。后一种方法可以简单廉价地产生无限量的 BNA, 如通过原核生物(如大肠杆菌)中的质粒 DNA 产生, 这种质粒 DNA 含有多重复的特定 BNA 序列, 其两侧是具有悬突末端的限制性酶切位点。通过这种方式, 例如, 可以以实质上无限量的拷贝产生噬菌体  $\lambda$  左侧或右侧的操纵位点、其它 DNA、其它已知和特定 BBA 特异且紧密结合的核酸序列, 如 DNA 或 RNA 结合蛋白序列, 每一个拷贝的两侧是 *EcoRI*, *PstI*, *BamHI* 或其它许多常见的限制性内切酶位点的任何一个。此外, 重复位点的聚合物可由不存在于聚合物内部的独有内切酶位点切除。可以通过本领域公知的方法产生大量的 pBR322、pUC 质粒, 或其它带有这些序列多拷贝的质粒, 在聚合位点两侧用限制性内切酶切开质粒, 通过层析方法或本领域其它任何方便易行的方法分离自由的多拷贝操纵子。BNA 在使用之前, 分离成单链, 然后聚合在一个编码操纵子的单链互补拷贝的 PNA 上(即 1/2 BBR)。通过利用质粒上的每个重复聚合物两侧的不同限制性内切酶, 把 BNA 媒介聚合在 PNA 上, 以产生多拷贝的 BNA。此外, 可以通过 BNA 聚合物一端或两端的悬突端杂交 BNA 聚合物和所说的 PNA, 这样就不需要分离双链和对 BNA 片段进行退火处理。以上提到的特异 BNA 序列的例子见序列描述部分, 为 SEQ ID NO. 35-36。为了稳定 BNA 聚合物, 可用 DNA 连接酶共价连接杂交的 BNA。

#### 5. 发夹核酸(HNA)及其制备

本发明的 HNA 至少包括两个主要的组成部分: 一个单链的序列, 其与 1/2 BBR 互补; 和一个双链的核酸区, 其在杂交条件下通过 HNA 内部的自我互补序列的自我连接而形成。参见附图 1 (IIc), 可以构建 HNA 中的 1/2 BBR 以使其与 PNA 中的 1/2 BBR 序列互补。参见附图 1 (I, IIc 和 IIIc), 在杂交条件下, 把前述的 HNA 加入到 PNA 中, 其则形成含有 BBR 的 PNA-HNA 杂交

体。参见附图 1 (IIIc, IVc 和 Vc), 在杂交条件下, PNA-HNA 杂交体在加入 TNA 之后, 可形成含有一个 TBR 和 BBR 的 TNA-PNA-HNA 杂交体。

如图 2a 和 2b 所示, 可以用 HNA “覆盖”或终止加入的相对于 PNA 的 BNA 的突出端。图 2a (Ib 和 Ic) 中所示的两个 BNA 分子可缔合形成图 3 (IVb) 所示的杂交体, 或者直接单独与 PNA 杂交 (如图 2a (Ia-c, IIa-d) 所示)。两个 HNA 分子 (图 2a (IIIa 和 IIIb 所示) 可以终止 BNA 与其它 BNA 的杂交, 这些 BNA 从 PNA 延伸 (图 2a (IVa-d))。图 2a (IIa) 所示的 HNA 可终止图 2a (IIb 和 IIc) 所示的 PNA-BNA 杂交体的形式和任何一种带有一条单链 1/2 BBR 的 PNA-BNA 杂交体, 这种单链 1/2 BBR 与图 2a (IIIa) 所示的 HNA 中 1/2 BBR 序列互补。同样地, 图 2a (IIb) 所示的 HNA 可终止图 2a (IIa 和 IIc) 所示的 PNA-BNA 杂交体和任何一种带有单链 1/2 BBR 的 PNA-BNA 杂交体, 这种 1/2 BBR 与图 2a (IIIb) 所示的 HNA 中 1/2 BBR 序列互补。

构建 HNA, 其能终止由连续加入 BNA 到 PNA 中而构建的 PNA-BNA 杂交体的形成 (见图 2b)。图 2b (Ia、IIIa、Va 和 VIIa) 中所示的单链 1/2 BBR 序列 (见) 与图 2b (Ib、IIIb、Vb 和 VIIb) 中所示的单链 1/2 BBR 序列特异性互补, 并可以产生独有的加帽的 PNA-BNA-HNA 杂交体 (图 2b (Ic、IIIc、Vc 和 VIIc))。

HNA 中的自互补序列和连接自互补发夹序列的环状序列, 只要其实质上不由 HNA 干扰或抑制单链 1/2 BBR (包含部分 HNA) 的存在或选择性地结合 BBA 或 TBA, 就可以有任何的组成和任意长度。应该选择环状序列以使环的形成不阻碍发夹的形成。在此项应用中, 提供的有用的 HNA 例子是 SEQ ID NO. 44 (见前面的序列描述部分)。

## 6. 靶结合装配体 (TBA) 及其制备

TBA 可以是与由特定 TNA 和 PNA 杂交形成的 TBR 结合的任何一种物质。条件是 TBA 必须具有至少下列特点:

(a) TBA 必须以以对兴趣 TBR (s) 高度特异性的方式与 TBR (s) 结合。也就是说, TBA 必须区分存在于 TNA-PNA 杂交体中的 TBR 和由 PNA-CNA 杂交体形成的类似的双链序列。TBA 必须与 PNA-CNA 杂交体有足够低的亲和力, 这样通过洗脱 TBA-TNA-PNA 复合体, 除去 PNA-CNA 杂交体而不除去 PNA-TNA 杂交体。



(B) TBA 必须与由 TNA 与 PNA 杂交形成的 TBR 极易结合。一般认为结合亲和力和在  $10^{-5}$  到  $10^{-12}$  或更高的范围内是足够的。在以下的描述中, 在某些情况下, 需要利用对特定 TBR 具有非常低的亲和力的特异 TBA, 但是当提供多聚 TBR 的特异构型时, TBA 的亲和力则极大地增加, 这样对于每个 TBR 的 TBA 亲和力的平方则是有关这个特异 TBA 的亲和力。

在 TBA 形成中有用的 DNA 结合组份的例子包括但不限于: NF- $\kappa$ B、乳头状瘤病毒 E2 蛋白、转录因子 SP1、无活性的限制性内切酶、抗体等。本领域认为每种这样的蛋白都含有和特异核酸序列结合的序列, 它们之间的亲和力是已知的。自然, 本发明的方法不限于这些已知 DNA 结合蛋白及其片断的使用。从本公开内容可知, 对于本领域一般技术人员很明显, 可以容易地把本发明的方法施用到新型 TBA 的使用中, 这种 TBA 至少需要具有上面所说的特性。比如, 在 W092 / 20698 中, 描述了包含寡聚核苷酸偶联物的序列特异性 DNA 结合分子, 这种寡聚核苷酸偶联物是由 DNA 结合药物和三螺旋体形成的寡聚核苷酸共价连接而形成的。可以用本公开内容的方法产生新型 TBA, 以便按照本文的方法使用, 条件是这样形成的 TBA 可以达到以上标准。此外, 美国专利 5, 096, 815, 5, 198, 346 和 W088/06601 中的方法(本文一并参考)也可以用来产生新型 TBA, 以便按照本文的方法使用(见 Blais(1994)的例子)。

不管 TBA 是一种蛋白质, 或是蛋白质复合体, 都可用本领域中很多方法路线的任何一种来产生 TBA。TBA 实际上可以从其自然产生的环境中分离, 或者如这种方法不可行, 则可用标准的分子生物学技术产生。这样, 以 NF- $\kappa$ B 为例子, 利用 p50 或 p65 亚单位的 DNA 结合部位, 可以按照本领域已知的重组方法产生这个结合装配体(见 Ghosh(1990), 细胞, 62:1019-1029, 描述了 NF- $\kappa$ B 的 p50 DNA 结合亚单位的克隆及这种蛋白对 *rel* 和 *dorsal* 的同源性)。

按照本发明, 已知许多 DNA 和其它核酸的结合蛋白都可用作或用于 TBA。一旦任何 DNA、RNA:DNA、RNA 或其它核酸结合蛋白的氨基酸序列已知, 就可通过合成的方法制备编码这段蛋白的合适的 DNA 序列, 或可以利用来源于合适组织的编码这种蛋白的 mRNA 的 cDNA 拷贝。并且可以获得编码这种蛋白的基因组拷贝, 同时用本领域已知的方法切除内含子。另外,

也可化学合成 TBA。

一旦得到了编码序列，就可利用点直接突变改变氨基酸序列而产生突变的核酸结合蛋白，这种蛋白比原核酸结合蛋白质具有更需要的结合特性。例如，可以改变 NF- $\kappa$ B 的 DNA 结合部分的氨基酸顺序，使之成为 NF- $\kappa$ B' 分子，该分子能更紧密地结合在 NF- $\kappa$ B 的结合位点上(见下面 HIV-检测和 HIV-锁定(Lick)的例子)。

为了进一步了解本发明的这个方面，考虑到了以下几点。以 NF- $\kappa$ B 为例子，TBA 可以利用自然产生的 NF- $\kappa$ B 分子制备。但是，由于该分子在细胞中的存在量极微，并且已克隆了这种 DNA 结合蛋白的亚单位，所以通过重组 DNA 的方法制备大量的复合体是可行的，并且在这种蛋白上已经完成了复合体的制备(见 Ghosh (1990), 细胞, 62:1019-1029)。NF- $\kappa$ B 是一种和免疫、炎症和生长调节反应相关的基因多效性诱导物，这种反应是对原初病原(病毒、细菌或逆境)竞争或次生病原(炎症状细胞素)竞争的反应。NF- $\kappa$ B 是一种二聚体 DNA 结合蛋白，包括 p50 和 p65 两个亚单位，两者都可以与特定 DNA 序列接触并结合。在非活跃时，NF- $\kappa$ B 存在于细胞质中，与一个特定抑制物(I- $\kappa$ B)形成一个细胞质的异源三聚体。在活化时，抑制物解离，p50-p65 二聚体通过一个特定细胞核定位信号(NLS)重新定位于细胞核中，在此与 DNA 结合，作为大量基因的转录激活体(见 Grimm 和 baeuerle(1993) 生物化学杂志, 290:297-308, 本文是有关 NF- $\kappa$ B 的综述文章)。

p50-p65 二聚体与共有序列 GGGAMTNYCC (SEQ ID NO.117)的配对序列以微微摩尔级亲和力结合，其精确序列的不同亲和力稍有不同。值得注意的是也观察到了 p50 和 p65 的同源二聚体的发生。这些同源二聚体具有不同的生化特性以及和结合序列不同的亲和力，结合序列存在于上面的共有序列中或和其相似。这样，根据 TBA 需要的结合特性，p50-p65 异源二聚体，p50-p50 同源二聚体，或 p65-p65 同源二聚体都可以利用。

图 9 以图解的形式说明了一种方法，利用这种方法可以产生各种各样的新型 TBA，以便根据本发明的方法进行使用。TBA 的核酸识别单位可以通过“陪伴”与类似的或不同的 TBA 核酸识别单位组装并缔合起来。所说的陪伴是一种结构，在这种结构上，可组装各种 TBA 识别因素，同时其能在

核酸识别位点上提供各种需要的特性。所说的陪伴由能提供装配序列的序列组成。以便将同样或不同的核酸识别部分拉近并彼此稳定地连接。这样，例如，在指定的一种 TBA 与 NF- $\kappa$ B TBR 紧密结合的情况下，通过提供噬菌体  $\lambda$  *cro* 序列做为组装序列而装配 TBA，这种序列和 NF- $\kappa$ B p50 或 p65 的核酸结合序列连接。p50 或 p65 的核酸结合序列与 *cro* 序列连接，其连接位点在 *cro* 的羧基端或氨基端，或者在 p50 或 p65 的核酸识别位点的羧基端或氨基端。可以提供也可以不提供连接序列，以便在核酸结合位点上存在用于合适 TBR 结合的适当间距。

装配序列，例如上面所举的 *cro* 和 CI 序列 (SEQ ID NO. 104-108)，包含任何能够自然而牢固地与其它序列结合的寡聚小肽。对于 *cro*，公知 *cro* 二聚体可与噬菌体  $\lambda$  的操纵子位点结合 (Anderson 等 (1981)，自然，290: 754-758; Harrison 和 asgarwal (1990)，生物化学年评，59: 933-969)。 *cro* 的单体彼此紧密且特异地结合。这样，通过 *cro* 序列和 DNA 识别单位序列连接，即可产生牢固而紧密的缔合。

可有可无的接头序列包含任何氨基酸序列，这种氨基酸序列不干扰 TBA 的装配或核酸的结合，并且不易发生变化以便从完整的 TBA 中释放出核酸识别单位。接头序列与其它结合装配组份应共价联接，这一点是需要的但不是必要的。这种缔合应该是特异性的，这样有利于结合装配体的装配和制造。这种序列的例子包括但不限于那些在结构蛋白中发现的连接各种区域的公知的序列。比如，在噬菌体  $\lambda$  阻抑蛋白中，在 DNA 结合区和二聚体区之间就有一个连接序列用于此目的。已知还有很多其它的这样序列，其精确的顺序不限于此发明，只要常规实验能保证其稳定性和对靶核酸结合的不干扰性。本文提供了这些序列的例子，为 Met Ser 和 SEQ ID NO. 99-102。在这些接头中插入特异性的已知的蛋白酶解位点也是本发明的组成部分。在接头序列中存在这类位点有助于进行制造，并且允许在陪伴框架上装配不同的分子。

除核酸识别单位，可有可无的接头序列，装配序列外，本发明的新 TBA 还可有或可无不对称或 PILOT TNA 序列以及一个或多个 OSA 单位。这些不对称序列用于促进或阻止某种需要或不需要的缔合。例如，当需要带有同源二聚体的 p50 DNA 识别单位 TBA 时，需要不对称序列打断自然条件下较

牢固的 NF- $\kappa$ B 的 p50 亚单位与 p65 亚单位的缔合，但又不干扰装配序列使 p50 亚单位聚合。本文的例子是 SEQ ID NO. 85-92 和 SEQ ID NO. 105 和 106。

在不同的构型中，NF- $\kappa$ B p50 亚单位可和转录因子 SP1 DNA 识别单位紧密缔合。在某些情况下希望这种结合，即 NF- $\kappa$ B/SP1 结合基元是很有价值的时候，如在 HIN LTR 中，至少六个 DNA 结合蛋白识别位点，两个 NF- $\kappa$ B，三个 SP1，和一个 TATA 位点的结合部分已经证明是存在的。已知第二个 NF- $\kappa$ B 和第一个 SP1 位点对 HIV 转录调节是非常重要的 (Perkins 等. (1993) *Embo J.* 12: 3551-3558)，这种特定 TBA 构型不仅用于 HIV 的诊断，而且可用作 HIV 感染的治疗剂或预防剂 (见下文)。同样地，按照本发明也可有用于人类乳头状瘤病毒的长控制区 (LCR) 作为探针探测的主控制区。

由于能够缔合的不同组份和弹夹方式，按照本发明的 TBA 制备方法，实质上可以不受限制地产生各种 TBA。图 10 中，列举了一系列不同的分子，即是“HIV-诊断 I-IV”，其中的“CHAP”代表陪伴、“NF $\kappa$ B”代表 NF- $\kappa$ B 亚单位、“SP1”代表 SP1 转录因子的 DNA 识别单位、“TATA”代表 TATA 序列 DNA 结合蛋白 (TBP) 的 DNA 识别单位的二聚体，也称为 TATA 结合蛋白，或 TBP。下文进一步列举了这些构型，其也是本发明的组成部分。

在另一种构型中，图 9 所示的模型结构是用于监测，处理和预防一种完全不同的病原。图 11 中，与上面描述的“HIV-诊断 I-IV”分子的方式相似，产生了一系列“HPV-诊断 I-IV”分子。在这种产品中，利用了人类乳头状瘤病毒 (HPV) 的 E2 蛋白的 DNA 结合的有利特性。此外，通过提供特异 DNA 识别单位来利用 SP1 和 TBP，用这种特异 DNA 识别单位结合 HPV 基因组中的这些序列。在用于监测 HPV 感染的 E2 特定 TBA 的形成过程中，希望利用 SEQ ID NO. 75-84 或 93-98 中的任何一种作为 E2 DNA 识别单位。含有牛 E2 二聚体和人 E2 二聚体的 DNA 结合区也许是特别有用的。

以上提到的各种序列可以通过使用纯寡聚核苷酸起动材料进行化学连接，或者通过提供重组核酸进行连接，这些重组核酸通过各种遗传密码编码各种亚单位。在重组产生过程中，把 *cro* 编码序列和核酸识别单位序列连在一起形成 TBA 是非常有利的，因为 *cro* 不仅在陪伴中充当装配序列，而且可以引导核酸识别单位进行适当的折叠。本文列举的陪伴序列见 SEQ ID NO. 104-108。并且，在希望得到含有多个结合位点的高级结构的过程中，

如 NF-kB/NF-kB/SP1/SP1/SP1 TBA 的五聚体, 正确地设计不对称序列使得能够制造这种结构。

以前面所述的方法, 制备以高亲和力和其同类结合位点结合的 TBA。比如: 图 10 中 TBA 的 NF-kB 的 DNA 结合组份是计划以  $10^{-8}$  和  $10^{-12}$  摩尔的亲和力结合 HIV-LTR 的。DNA 结合位点有用的序列见 SEQ ID NO. 63-71, 73-84, 93-98 和 104-108, 下文还要引用。

由于前面使用装配序列或不对称序列的直接装配核酸结合蛋白的描述, 本领域的技术人员会认识到本发明提供了装配蛋白结构的一般实用的方法。通过进一步的阐述和举例的方式、利用需要的结构的装配体中的抗体-抗原表位的相互作用的方式, 进一步验证本了方法施用的普遍性。更详细地说, 通过把 NF-kB p50 亚单位连到一个抗原上, 如环形的(酶二硫键)黑色素细胞刺激激素(MSH), 抗原装配 DNA 结合蛋白结构。这种原-MSH 分子可以通过抗 MSH 的抗体结合, 形成一个新型核酸结合装配体, 这个过程中抗原和抗体充当装配序列。

图 9 的模式结构表明可以用各种组份的不同结合体来装配大量不同的 TBA。这样, 一般结构的代表实例见 SEQ ID NO. 119-116。

### 7. 辅助结合装配体(BBA)及其制备

BBA 可以是和特定的 BBR 结合的任何物质, 这种 BBR 是通过特定的 PNA 与 BNA 杂交而形成的, 包括多个 BNA(上至“n”且包括“n”的 BNA, 即  $BNA_n$ , 其中的“n”在理论上从 0 到无穷大, 一般是在 0 到 100 之间)聚合在 PNA 上以用于信号扩增, 条件是 BBA 必须具有至少以下两个特征:

(a) BBA 必须以某种方式与 BBR 结合, 其对兴趣 BBR 是高度特异性。也就是说, BBA 必须区分存在于 PNA-BNA 杂交体中的 BBR 和 BNA-CNA 杂交体中类似的双链序列或其它的 CNA。因此, 在 PNA- $BNA_n$  或 PNA- $BNA_n$ -HNA 杂交体的产生过程中, 当甚至只有一个碱基错配, 或构象差异(有或无碱基误配), 这种 BBA 必须以足够低的亲和力和杂交体聚合, 以便在洗脱 TBA-TNA-PNA-BNA 复合体时, 从 CNA 序列中除去了 BBA 而不是 BBR 序列。

(b) BBA 必须极易结合 BBR。在  $10^{-5}$  到  $10^{-9}$  或更高范围内的结合力一般认为是足够的。

BBA 的例子包含但不限于 *cro* 和噬菌体  $\lambda$  阻抑蛋白 CI。此外美国专利

4, 556, 643, 本文一并参考, 其介绍了其它的 DNA 序列和特异结合蛋白, 如阻遏子, 组蛋白, DNA 修饰酶类, 及分解代谢产物基因的激活蛋白。也见 EP0453301, 本文一并参考, 其提议了一大群核苷酸序列特异结合蛋白 (NSSBPs), 例如四环素阻遏物, 乳糖阻遏物以及色氨酸阻遏物。本领域已经认识到这些 BBA 的每一个都和特定的、已知的核酸序列结合, 它们之间的亲和力也是已知的。当然, 本发明的方法不限于使用这些已知的 BBA。从本发明的公开内容可知, 普通技术人员可以容易地施用新型至少具有本方法以上提到的所需特性的 BBA 的用途。

按照本发明的这方面, 有用的新型 BBA 的例子包括基于已知的 DNA 或 RNA 或 DNA: RNA 结合蛋白基元的新蛋白质, 例如 *cro* 或入 CI 阻抑蛋白。优选的是对这些蛋白进行一些修饰, 以提高本发明这些组分的操作性。这样, 对一个分析来说, 以希望加入高浓度的 *cro*。*cro* 的一个缺点是在高浓度时, *cro* 与其靶 DNA 的结合将与 *cro-cro* 间的相互作用竞争。这样, 例如, 可以产生一种没有这种缺点的陪伴的或突变的 *cro*。这种改变了的陪伴的例子见 SEQ ID NO. 105-106 和 108。本领域已知的方法, 例如用多样的核酸群体产生新型靶结合蛋白及选择噬菌体和特定的、预先靶结合(例如, 所谓的噬菌斑展示技术, 参见上面有关新型 TBA 产生的讨论)的方法, 都可以用来产生这种新型 BBA 以及前面提到的新型 TBA。

无论 BBA 是一个蛋白质或蛋白质复合体, 应该认识到可以利用本领域许多常规方法的任何一种来制备这种 BBA。BBA 事实上可从其自然产生的条件中分离, 或者, 如果这样不可行, 可用标准的分子生物学技术来制备。例如, *cro* 蛋白序列是已知的, 噬菌体  $\lambda$  的任何分子克隆可用于获得编码 *cro* 的适当的核酸以用于重组体产生。此外, 如果利用不同的 TBA 和 TBR 及 BBR 结合, 本文描述的 TBA 也可用作 BBA。

#### 8. BBA 和 BBR 用来定位和扩增 PNA-TNA-TBA 复合体定域的用途(见图 8)

在本发明的一个实施方案中, 使用了高度特异和极紧密结合的包含核酸结合组份的 TBA, 来产生可扩增的核酸夹层分析。按照本实施方案的一个方面, 用第一个 TBA 包裹一个固体支持物以产生固定化的 TBA。在溶液中, PNA 和 TNA 在杂交条件下接触, 然后与固定化的 TBA 接触。只有那些能够形成由固定化 TBA 识别的特异 TBR 的 PNA-TNA 相互作用才能在洗脱后保留下

来, 该固体表面结合了 TBA-TBR 复合体。

结合的 TBR 的检测通过在杂交条件下辅助核酸、BNA 和存在于 PNA 上的 1/2 BBR 结合来完成。以这种方式, 即使只有单一的 TBA-TBR 复合体结合在固定化 TBA 上, 也可以通过聚合众多的 BNA 在固定化 TNA 上来产生一个扩增产的信号。每一个和 TNA 结合的 BNA 产生一个 BBR, 它可和 BBA 结合, 就象 TBA 固定在固体表面一样, 可以选择 BBA 使它们能非常紧密且特异性地和特定核酸结构结合。这样, 按照这一实施方案, 固定化的 TBA 可以包含 NF-kB 的 DNA 结合蛋白, 它非常特异地且紧密地与 NF-kB 结合位点结合, 这一结合位点是基于 TNA 与 PNA 杂交而形成的。

因为, 公知在正常人的基因组与人类免疫缺陷病毒(HIV)的长末端重复序列中都存在 NF-kB 结合位点, 本发明提供了一种能区别“正常”人类的结合位点和由于 HIV 感染而在细胞中出现的结合位点的方法。因此, 在一个设计来确定人 DNA 样品中是否存在 HIV DNA 的测试中, 可以把 HIV NF-kB 结合位点看成 TNA, 正常人的 NF-kB 结合位点看成 CNA。按照本发明的方法, 区别 TNA 和 CNA 可以通过利用如下事实来完成, 即在 HIV LTR 中, 有两个 NF-kB 结合位点, 后面紧接着三个 SP1 位点(见 Koken 等. [1992] 病毒学 191: 968-972), 而发现具有同样序列的细胞 NF-kB 结合位点不是一前一后的。

在 TNA 包含不只一个 1/2 TBR 时, 希望实施 TBA 的诊断和预防应用时, 可以使用不只一个的 TBA, 每个 TBA 都具有与 TNA-PNA 复合体中的 TBR 结合的能力。在这种情况下, 选择对 TBR 亲和力小于野生型 DNA 或 RNA 结合区的 DNA 或 RNA 结合区是有利的, 这种结合区作为所说的 TBA 的组分。假设那些与众多 TBR 结合的 TBA 既可以在与它们的 TBR 结合之前聚合在一起, 也可以在与它们的 TBR 结合之后聚集在一起, 在其它非靶基因组中, 除非 TBR 具有空间上结合已装配的 TBA 复合体的能力, 单个的 TBA 将不会阻碍相应的 TBR。在这儿特别声明的作为本发明一个部分的多聚 TBA 复合体的一个特征是: 预计这种多聚复合体与 TNA 内单个位点的亲和力极大地减小。然而, 由于其结合能力相对于单个的 TBA 极大地增加, 预计这种 TBA 复合体不以任何单个的 TBR 与活体内相应蛋白质之间的结合发生竞争, 但却和含有 TBR 的 PNA-TNA 杂交体的序列紧密结合, 这种 TBR 用于 TBA 中每个装

配的核酸结合组份。装配 TBA 复合体并调整单个 TBA 中的接头以便包含在 TBA 复合体中的核酸结合区到达这些靶并与之结合。

一旦 TNA-PNA 杂交体形成并与固定化 TBA 接触, 未结合上的核酸则可从固定化表面洗脱掉并可检测固定化的杂交体。可以用几种方法中的任何一个来完成这个过程。本发明的一个方面是, PNA 用一种 OSA 标记, 如放射性核素, 有色颗粒, 或有能力产生带色反应产物的酶。并且, 这种 PNA 除了具有一个或多个 1/2 TBR 外, 也可以含有至少一个 1/2 BBR。对这种 1/2 BBR 序列进行选择以便其与 BNA 中独有的 1/2 BBR 序列互补。在上面描述的实施例中, 例如, 其中的 TBA 是 NF-kB 以及在 TNA-PNA 杂交体上形成的 TBR 是一个或多个 NF-kB 的结合位点时, 1/2 BBR 可提供噬菌体  $\lambda$  操纵子左端或右端的可杂交(即是单链的, 互补的)的序列(见, 例如, Ptashne [1982], 科学的美国 (Scientific American) 247:128-140, 以及该文关于这些操纵子序列所引用的参考文献)。这些序列可以以载体的形式聚合在 PNA 1/2 BBR 上(见图 2 和 3), 提供“n”个 BBR, 每个 BBR 形成一个 *cro* 结合位点。以酶法、放射性或其它方法标记的 *cro* 与 TBA-TNA-PNA-(BNA)<sub>n</sub> 复合体接触。以这种方式, 可产生一个高度选择性的和扩增的信号。用含有单个 1/2 TBR 的 PNA 产生的信号表明分析的成功, 完成了 TBA-TBR 的结合及 BNA 的聚合结果从细胞位点产生信号(例如, 从 CNA)。当使用双聚 TBA 时, 信号的缺乏表明在 TNA 中没有 HIV LTRs, 因为不存在双 NF-kB 结合位点。另一方面, 使用双聚 NF-kB 时信号的出现表明有 HIV 感染。本发明此实施例前面所述的一个具体例子, 见例 6 所描述的 HIV 检测试剂盒。

自然, 本领域的技术人员应该认识到前面所述的描述可以在选择 PNA, TNA, TBA, BNA 和 BBA 有几种修改。并且, 在非 HIV 的其它系统中, 本领域的技术人员应意识到以上描述的一般方法可同样使用。然而, 其它的使用可以比以上描述的方法更简单, 因为使用的 TBA 可以不识别任何正常细胞位点因而采取双聚合法, 或其它能区别 TNA 和 CNA 方法都可以。在设计用于其它系统的探针和结合装配体时, 本领域技术人员应遵守下面的原则和注意事项。

在以上所述的实施方案中, 要求用 NF-kB 蛋白的 DNA-结合部分作为 TBA 和用 NF-kB 识别结合组份作为 TBR, 是因为这些组份形成一个重要的 HIV



复制的“控制点”。也就是说，已知 HIV 需要在其复制生命周期中用 NF- $\kappa$ B 作为一个关键组份。按照本文描述的方法，可以选择其它病原体的类似控制点并作为检测的基础。

从以上描述的本发明的一般特性和操作方式，本领域的技术人员应认识到有很多实践本发明的具体方式。例如，本发明的方法适用于美国专利 4690691 和 5310650 ('691 和 '650 专利)描述的使用层析测试盒的方法和装置。在那些专利中，利用孔状培养基固定 TNA 或捕获探针，利用一种溶液传输流动相到“捕获区”，如果固定的是 TNA，流动相中则含有标记的 PNA；如果固定的是捕获探针，流动相则含有 TNA。通过直接固定或捕获，一旦 TNA 结合到捕获区，在捕获区内色谱分析标记的 PNA 同时检测任何结合的标记。

本发明采用这种系统，改进了捕获区的子结合装配体的利用，因而，捕获的只是完美相配的 TBR 序列或其它代表核酸构型的 TBR，这种 TBR 和 TNA-PNA 双螺旋体中的 TBA 特异结合，其中 TNA 与 CNA 的区别依靠前面描述的 TBA 灵敏区别 TNA 与 CNA 的方法。

一旦 TNA-PNA 杂交体与固定化的 TBA 结合，就可以通过加入 BNA 或捕获区层析分析的 BNA 而扩增信号。最后，可以通过加入 BNA 或捕获区层析分析的 BNA 而进一步扩增信号。以这种方式，通过提供增加那些专利所述方法的特异性和灵敏性的额外能力(通过扩增)，本发明提高了专利 '691 与 '650 所述分析各步的简便性。为了显示该方法的细节和该文本所述特殊操作条件的教导(该条件下，本发明的组分和方法也是适用的)，专利 '691 与 '650 的公开内容本文一并参考。

本领域的技术人员也会认识到本发明的方法适合用于微孔滴板或自动分析。因此，并入本发明方法的仪器的用途自然在本公开内容和有关权利要求范围内。这样，例如，本发明可采用的仪器，如 Abbott Laboratories (Cabbott Park, IL) 的 IMx 桌面分析仪。设计 IMX 目前是用来进行荧光极化免疫分析(MEZA, 见 Kier [1983] KCLA 3:13-15)和微粒酶联免疫分析(MEZA, 见实验医学, Vol. 20, NO. 1, January 1989, PP. 47-49)。用本发明的 TBA 作为捕获分子，这种捕获分子包裹在悬浮于溶液中的亚微粒(平均 $<0.5 \mu\text{m}$ )大小的微粒上，可以容易地把 MEZA 法转变成核酸检测的

方法。将TBA包裹的微粒转入反应细胞中。然后IMx将样品(杂合的PNA-TNA)转入反应细胞中,与TBA形成一个复合体。经过一段适当时间的培养,将溶液转移到惰性的玻璃纤维基质上,颗粒物与其具有强的亲和力因而微粒粘在其上。在通过玻璃纤维基质过滤反应液之前或之后加入BNA和BBA,或依据TNA-PNA杂交体的特定结构,使用另外的信号扩增和检测方法。对固定化的复合体进行洗脱,未结合上的物质通过玻璃纤维基质流出来。

用碱性磷酸酶标记BBA或其它形式(放射性的,酶学的,荧光的)标记BBA的方法检测结合的复合体。在碱性磷酸酶标记BBA的情况下,可以加入荧光底物4-甲基umbelliferyl磷酸盐或类似的试剂。另外,可以通过直接以这种试剂或类似试剂标记BBA的方法而省去酶。不管哪种情况,荧光或其它信号都与PNA-TNA杂交体的数量成比例。

用IMx制造商所描述的前表面荧光计的方法来检测基质表面的荧光。通过一些小的调整,这一方法可以由常规实验来完善仪器,例如Imx用于核酸杂交和核酸-TBA相互作用的检测,本发明完全适用于TNA样品的自动分析。

#### 9. 本发明的其它诊断应用

虽然,从前面的描述可知能以多种不同的方式利用本发明,但本发明另外众多的应用也是相当有价值的,例如,在移动延迟系统中的应用。

在本发明的这个实施例中,依下列各项对公知的电泳迁移率变动分析(EMSA)进行了改进(见图12a和12b)。

通过随机切割或通过特定的限制性核酸内切酶处理的方法把DNA样品切成片段。然后将DNA样品分成相等的两份,将特定的TNA加入到第一份中但不加入第二份中。然后将两等分试样在聚丙烯酰胺或琼脂糖凝胶上电泳。比较两份样品的DNA带型(通过溴化乙锭结合或通过电泳前放射性标记的方法显示)。具有TBA特异结合位点的DNA片段在通过电泳介质迁移时受阻。用合适的TBA,任何数量的DNA或其它核酸都可以以此种方式追踪。

在以上描述的EMSA的改进中,TNA的片断与PNA杂交并被分离在第一部分中。分离的DNA与适当的TBA反应同时注意到DNA片段的迁移率发生变化。如上文描述,加入BBA可能会增加这种阻滞(见Vijg及其所引用的有关已知的二维核酸电泳技术的文献,对于这种技术可直接应用本方法)。

## 10. 治疗应用

由于本文描述的新型 TBA 具有紧密且选择性的核酸结合特征, 所以考虑到这些化合物除了诊断上的应用外, 还可用于治疗。这样, 由于在临近的组合中存在 NF-kB p50 和 SP1 DNA 识别单位(见图 10, HIV-检测 II), 具有紧密及特异地与 HIV-LTR 结合的特性的 TBA, 可以用来结合到 HIV-LTR 上而阻止 HIV 基因组的某些关键因素的转录。TBA 装配序列独有的这种特征允许重组载体把编码 TBA 的 DNA 导入细胞并把表达的序列正确地折叠。一旦进入细胞, p50 亚单位上的细胞核定位信号将引导 TBA 进入细胞核并紧密地与任何整合的 HIV 的 LTR 结合, 结果有效地控制了病原。在预防模式中, 对有可能暴露在 HIV 中的人施用足够剂量的 TBA 或一个可以表达 TBA 的重组载体, 这样阻断任何可能侵入人体的 HIV。在这种模式中, TBA 的使用类似于用特异性免疫球蛋白的被动保护。在治疗或预防模式, 使用 NLS 序列替代用于诊断模式中的 OSA。例举的 NLS 序列见 SEQ ID NO. 72 和 103(另见 Heinzinger 1994 和 bukinsky 1993, 分别描述了 HIV Vpr 和 gag 蛋白质的 NLS 序列)。在任何情况下, 以药学上可接受的载体施用 TBA, 如无菌盐溶液, 与磷脂体结合或以重组载体的形式, 优选的方法应该能指导 TBA 在选择性的细胞类型中表达, 或取决于蛋白质的转递系统。

### II. 本发明的实施方案

由于前面的描述和下面的实施例, 本领域的技术人员会理解本文描述了并使得可以实施本发明的各种实施方案, 这些实施方案包括:

#### 1. 一种探针核酸(PNA), 该核酸包含:

(A) 单链序列 1/2 TBR, 该序列在杂交条件下可以和存在于靶核酸(TNA)中的 1/2 TBR 形成杂交体 TBR;

(B) 单链序列 1/2 BBR, 该序列在杂交条件下可以和辅助核酸(BNA)中的 0-10 个 1/2 BBR 形成杂交体 BBR; 和

(C) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

其中所说的 TBR 可以以高亲和力和 TBA 结合, 所说的 TBA 是一种能够区分配对的 TBR 和具有未配对的核苷酸的 TBR 的物质, 并且, 所说的 BBR 可以

以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分配对的 BBR 和具有未配对的核苷酸的 BBR 的物质。这一实施方案包括 TRR (其是核酸结合蛋白识别位点, 例如 HIV LTR) 在其它病原体中的核酸结合蛋白识别位点, 其中某些以上已说明。本发明的这一实施方案的 PNA 可以产生 TBR, 后者是存在于病原体基因组中的核酸结合蛋白识别位点或者是与人基因组中的病理状态或发酵过程中污染物相关的结合位点。

2. 一种辅助核酸(BNA), 该核酸包含:

(A) 1/2 BBR, 其具有和 PNA 或另外的 BNA 中 1/2 BBR 序列互补的序列, 并且其在杂交条件下可以和 PNA 形成杂交体 BBR;

(B) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上; 和

(C) 附加的杂交位点 1/2 BBR, 用来和附加的 BNA 杂交;

其中所说的 BBR 可以以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分配对的 BBR 和具有未配对的核苷酸的 BBR 的物质。

3. 一种发夹核酸(HNA), 该核酸包含单链序列 1/2 BBR, 该序列在杂交条件下能够形成发夹结构, 并且同时和 BNA 结合形成能够结合 BBA 的 BBR, 其中所说的 BBR 可以以高亲和力和 BBA 结合, 所说的 BBA 是一种能够区分完美的 BBR 和具有未配对的核苷酸的 BBR 的物质。

4. 一种用于检测特定 TNA 序列的方法, 该方法包括以下步骤:

(A) 使所说的 TNA 与权利要求 1 的 PNA 杂交;

(B) 使所说的 TNA 与含有 1/2 BBR 的 BNA 杂交, 所说 1/2 BBR 的序列和 PNA 中的 1/2 BBR 序列互补;

(C) 将含有 TBR 和 BBR 的步骤(A)和(B)的产物加入到含有 TBA 的表面、液体或其它介质中;

(D) 将 BBA 加入到步骤 C 的混合物中, 其中所说的 BBA 包含:

(I) 能够选择性地结合 BBR 的分子或分子的一部分;

(II) 可检测的指示剂; 和

(E) 检测连接到 BBA 上的指示剂所产生的信号。这一方法包括使用蛋白质指示剂包括可以催化反应导致产生有色反应产物的酶, 也包括象放射性

核素, 有色小球之类的指示剂。

5. 一种用于检测样品中特定靶核酸 TNA 存在的方法, 该方法包括以下步骤:

(A) 使所说的样品和探针核酸 PNA 接触, 如果样品中存在所说的 TNA, 则 PNA 与 TNA 杂交形成靶结合区 TBR, 此 TBR 可以和靶结合装配体 TBA 结合; 和

(B) 使已经和所说的 PNA 接触的样品和 TBA 接触, 此 TBA 可以结合任何由所说的 PNA 与样品中所说的 TNA 杂交形成的 TBR。

6. 一种具有高灵敏性和特异性的检测和定位特定核酸序列的方法, 该方法包括:

(A) 将含有 1/2 BBR 和 1/2 TBR 的 PNA 加入到含有或怀疑含有包含 1/2 TBR 序列的 TNA 样品中, 以形成具有靶结合区的复合体 TBR, 此 TBR 由分别存在于 PNA 和 TNA 中的互补 1/2 TBR 的杂交形成;

(B) 将步骤 A 中形成的 TBR 结合到固定化的 TBA 上, 以形成 TBA-TNA-PNA 复合体;

(C) 将包含辅助结合区 1/2 BBR 的辅助核酸 BNA 加入到步骤 B 中形成的复合体中, 以便 BNA 中的 1/2 BBR 与存在于 PNA 中的 1/2 BBR 序列杂交, 或者加入到存在于已经和 PNA 结合的 BNA 中的 1/2 BBR 上形成 BBR, 以便形成 TBA-TNA-PNA-(BNA)<sub>n</sub> 复合体;

(D) 将含有 1/2 BBR 序列的发夹核酸 HNA 加入到步骤 C 中形成的复合体中, 以便 HNA 中的 1/2 BBR 和任何存在于步骤 C 的复合体的 BNA 中的可获得的 1/2 BBR 序列杂交, 由此在步骤 C 的 TBA-TNA-PNA-(BNA)<sub>n</sub> 复合体上加帽使 BNA 延伸, 形成 TBA-TNA-PNA-(BNA)<sub>n</sub>-HNA 复合体;

(E) 将和指示剂部分连接的辅助结合装配体加入到步骤 D 中所形成的 TBA-TNA-PNA-(BNA)<sub>n</sub>-HNA 复合体中, 以便形成 TBA-TNA-PNA (BNA-BBA)<sub>n</sub>-HNA 复合体; 和

(F) 检测由连接到步骤 (E) 中 TBA-TNA-PNA-(BNA-BBA)<sub>n</sub>-HNA 复合体中的 TBA, PNA, BNA, BBA 或 HNA 上的指示剂所产生的信号;

其中所说的 TNA 包含:

(i) 一个或多个特定 1/2 TBR 核酸序列, 此特定序列在特定样品中的存

在与否有待证实;

所说的PNA包含:

(i) 单链序列 1/2 TBR, 该序列在杂交条件下可以和存在于靶核酸(TNA)中的 1/2 TBR 形成杂交体 TBR;

(ii) 单链序列 1/2 BBR, 该序列在杂交条件下可以和辅助核酸(BNA)中的 0-10 个 1/2 BBR 形成杂交体 BBR; 和

(iii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

所说的BNA包含:

(i) 1/2 BBR, 其具有和PNA或另外的BNA中 1/2 BBR 序列互补的序列, 并且其在杂交条件下可以和PNA形成杂交体 BBR;

(ii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上; 和

(iii) 附加的其它BNA的杂交位点 1/2 BBR;

(iv) 序列 1/2 BBR, 其可以与已经和PNA杂交的BNA杂交;

所说的BBA包含:

(i) 能够选择性地结合BBR的分子或分子的一部分; 和

(ii) OSA, 其不是连接的支持物和/或指示剂, 或者是连接的支持物或其它定位方式和/或指示剂, 所述的定位方式包括但不限于连接到小珠上, 聚合物上和表面上;

并且所说的TBA包含

(i) 能够选择性地结合TBR的分子或分子的一部分; 和

(ii) 无连接的支持物和/或指示剂, 或者连接的支持物或其它定位方式, 包括, 但不限于连接到小珠上, 聚合物上和表面上和/或指示剂上;

7. 一种用于检测靶多核苷酸存在的改进的固相杂交方法, 所述的固相杂交方法涉及将靶多核苷酸(如果存在于样品中)直接或经媒介俘获结构固定化在固相的俘获位点上; 在所说的固定化过程之前、之中或之后将一种可检测的标记连接到所说的靶多核苷酸(如果存在)上, 并且在俘获位点上

检测所说的标记(如果具有任何所说的标记存在), 所述改进包括:

(A) 使用靶结合装配体 TBA 作为达到固定化所说的靶多核苷酸的手段, 其中所说的 TBA 只结合到由特定的探针核酸 PNA 和所说的靶核酸形成的独特的杂交体上, 以便形成可以被所说的 TBA 识别的完美的靶结合区 TBR; 和

(B) 在 PNA 中包含进可以结合辅助核酸 BNA 的单链核酸 1/2BBR, 所说辅助核酸含有单链互补 1/2 BBR, 当其和 PNA 中的 1/2 BBR 杂交时, 形成可以和标记的辅助结合装配体 BBA 结合的 BBR。

8. 一种靶结合装配体 TBA, 其包含一个或多个核酸识别单位、接头序列、装配序列、不对称序列、核定位信号序列(NLS)和 OSA。所说的核酸识别单位可以是 NF-kB 结合单位、SP1 结合单位、TATA 结合单位、人乳头状瘤病毒结合单位、HIV LTR 结合单位或特定序列(其检测是所需的, 并且其可以通过与 TBA 特定缔合完成)的任何其它片段的结合单位。这样的核酸识别单位包括但不限于这里作为举例的这样一些: SEQ ID NO. 63、SEQ ID NO. 64、SEQ ID NO. 65、SEQ ID NO. 66、SEQ ID NO. 67、SEQ ID NO. 68、SEQ ID NO. 69、SEQ ID NO. 70、SEQ ID NO. 71、SEQ ID NO. 72 和 SEQ ID NO. 73。接头序列, 例如寡肽, 不干扰核酸识别单位的核酸识别功能, 并且提供核酸识别单位与剩余 TBA 间隔的稳定性和控制。这种接头序列是本领域已知的, 包括但不限于来源于结构蛋白区域间一级序列的寡肽序列。装配序列包括寡肽序列, 此寡肽序列指导核酸识别单位的折叠和缔合。这种序列优选的例子是来源于 $\lambda$ 噬菌体 *cro* 蛋白寡肽。所说的不对称序列指导核酸识别序列和装配序列以预定的顺序缔合。这种不对称序列的例子是来源于胰岛素、松弛素、促性腺激素、FSH、HCG、LH、ACTH 的序列, 包括但不限于 SEQ ID NO. 85-92。参照图 14 和 15, SEQ ID No. 85 是“ A ”序列, SEQ ID No. 86 是“ B ”序列; SEQ ID No. 87 是“ A ”序列, SEQ ID No. 88 是“ B ”序列; SEQ ID No. 89 是人松弛素“ A ”序列, SEQ ID No. 90 是人松弛素“ B ”序列; SEQ ID No. 91 是 skate “ A ”序列, SEQ ID No. 92 是 skate “ B ”序列。此外, 所说的 TBA 可以含有核定位序列 NLS, 其指导与所说的 NLS 相关的蛋白质或复合体迁移和摄入到细胞核中。这样的 NLS 序列的例子是 SEQ ID NO. 72 和

SEQ ID NO. 103。TBA 优选的具体例子是 HIV Detect I-IV 或 HPV Detect I-IV，和 SEQ ID NO. 109-116。

9. 使用本发明的新 TBA 的方法，包括但不限于使用 TBA 结合靶核酸样品中特定的核酸序列的方法，该方法包括：

(A) 裂解靶核酸样品中的核酸；

(B) 在杂交条件下，使裂解的核酸与和特定的兴趣核酸互补的探针核酸接触，其中在所说的的探针核酸与所说的特定的兴趣核酸序列杂交时，形成所说的 TBA 特异性地结合到其中的靶结合区。

在这一方法中，所说的探针核酸除了具有和所说的特定的兴趣核酸互补的序列外，还具有附加序列，此附加序列可与辅助核酸结合，形成辅助结合位点，标记的辅助结合装配体可以结合到此结合位点上，提供显示和扩增探针核酸和兴趣靶核酸序列结合的信号。

本发明不需要裂解靶核酸的另一方面涉及将预防和治疗有效量的所说 TBA 施用给需要此种治疗的病人，该方法包括以提纯的蛋白复合体的形式或以重组载体（进入病人体内后可以表达 TBA）的形式施用 TBA，以便 TBA 和特定的核酸序列结合达到所期望的治疗和预防结果。这可以包括提供可由常规实验确定的足以阻止病原体活性感染的剂量。当以指导 TBA 体内表达和折叠的重组表达载体提供时，重组核酸的剂量可以实质上较低，特别是以非病原体病毒载体形式提供时。使用 TBA 的方法也包括监测靶核酸样品中核酸移动性的改变，以便 TBA 和样品中特定的核酸片段结合改变此片段的移动性作为核酸大小的函数。所述方法的这一方面提供了为特定的畸变（例如可能在与新陈代谢相关的过程中发现的）分析核酸片段的有用的方法。

10. 诊断或法医试验试剂盒，该试剂盒在确定感染的存在、疾病的易感性或包含特定核酸源的样品。

11. 一种用于在体内装配多聚体 TBA 方法，该方法包括将组份 TBA 引入到细胞中，所说的 TBA 组份每个都应含有核酸识别单位、装配序列、不对称序列和核定位信号序列，如果足迹法实验表明需要接头序列来获得多聚体 TBA 的最适的几何学，则接头序列也可以包含在其中。当体内表达各组份 TBA 和经由各组份 TBA 的 DNA 识别单位邻近地结合到在细胞核中或细胞其它区域遇到的核酸序列上时，组份表达的 TBA 通过所说的装配和不对称



序列被装配成多聚体 TBA。如上所述, 这样的多聚体 TBA 具有以高的亲和力特异性地结合到特定靶序列中的 TBR 上但根本不结合到表亲核酸上或者以非常低的亲和力结合到表亲核酸上的优点。

本发明以上的描述会使本领域技术人员明白本发明的优选的实施方案和最佳方式。不将发明主题限制在此后给出的具体例子中, 以下实施例提供来指定本领域技术人员实施本发明的方法。在 Sambrook, Fritsch 和 Maniatis (1989) 分子克隆: 实验室手册, 第二版, 冷泉港实验室出版社, 冷泉港, NY 中描述了标准的重组 DNA 技术, 没有描述较近期的内容, 因为现在这些已经完全在本领域技术人员的知识范围内。

#### 实施例 1-PNA 的制备及标记

可以通过本领域已知的各种方法制备核酸探针(PNA)。这样, 限定序列的单链多核苷酸的 PNA 可以按照 Merrifield 的固相化学合成方法进行制备。可以利用通过商业途径可以得到的技术经自动合成的方法制备 PNA, 例如使用 Biosystems、ABI 和其它制造者生产和销售的树脂和仪器。另外, 通过已知的重组 DNA 方法, 体内合成特定的 PNA 序列, 例如通过将双链 PNA 克隆进能在大肠杆菌中复制的载体中, 则可以制备大量的双链 PNA。可以把 PNA 聚合物克隆进这种载体, 这样对于每摩尔的载体, 通过用位于 PNA 序列两侧的限制片断消化载体, 可以释放几摩尔的 PNA。继合成或重组产生之后, 通过本领域公知的方法纯化 PNA, 如通过凝胶电泳或高压液相色谱技术。如果 PNA 以双链形式产生, 在其用于检测靶核酸序列的杂交分析之前, 用加热或本领域已知的其它方法将 PNA 的双链分离。

选择 PNA 中的特定碱基序列, 这些序列反映了 TNA 中被检测的序列, 但条件是, 根据本发明, PNA 含有一个 1/2 TBR 序列, 此序列是基于 PNA 和 TNA 杂交形成 TBR 的之一。由于本领域中实际上有无限量的这样的序列, 所以对于任何给定的应用, 技术人员能够选择到合适的 PNA 序列。HIV LTR 序列就是这样一个序列, 其在编码部分 LTR 的 PNA 与编码 HIV LTR 的 TNA 杂交后, 形成可以与 NF-kB 或 SP1 DNA 结合蛋白结合的 TBR。

除了在杂交基础上形成 TBR 的序列, PNA 还可以含有 1/2 BBR。这一序列是指在和辅助核酸 BNA 杂交后可以形成 BBR 的序列, BBR 序列能结合 BBA。优选的 BBA 是对 BBR 序列有高度亲合力的 DNA 结合蛋白。

在这个具体的例子中,杂交发生在有一个1/2 TBR的PNA(见SEQ ID NO. 4)和该序列3'端的1/2 TBR(见SEQ ID NO. 35)之间。编码这些序列的PNA在使用中可以不标记或者用同位素如 $P^{32}$ ,  $S^{35}$ 标记或按照本领域已知的方法用类似的同位素标记。另外, PNA也可以和直径0.01到10微米的胶粒结合,将这些胶粒染色以便于视觉检测。这种标记形成如说明书所描述的OSA。这种探针与HIV LTR序列杂交形成了结合NF- $\kappa$ B的TBR。此外, PNA与具有互补1/2 BBR的BNA杂交形成噬菌体 $\lambda$ 的左操纵子,这种操纵子或者与*cro*或与 $\lambda$ 阻抑蛋白结合。

以上面描述类似的方法,使用PNA,其中所说的1/2 TBR是SEQ ID NO. 5或SEQ ID NOS. 7-34的任何一个,另外一个1/2 BBR,如SEQ ID NO. 35或36,在前一个1/2 TBR的3'端或者5'端。

#### 实施例2-BNA的制备和标记

与实施例1中所描述的制备和标记PNA的方法相似,按照本领域已知的方法制备和标记BNA。美国专利4,556,643所描述的方法,本文一并参考(具体参见例1),可以通过将其克隆到一个可复制的载体中而大量产生编码特异核酸结合序列的核酸序列。并且,和其所公开的内容相似,以这种方式,可以共线性产生1/2 TBR和1/2 BBR,但不同的是,按照本发明,1/2 TBR自身形成一个核酸结合组份的识别位点,而1/2 BBR,尽管形成一个核酸结合组份识别位点,但还可以通过把BNA聚合在结合PNA的TNA上,而提供了扩增基于这种1/2 TBR和TNA中的互补序列结合而产生的信号的方法。为了实现这种扩增,给如SEQ ID NO. 35的序列(这种序列能够编码噬菌体 $\lambda$ 左操纵子)装配额外的序列,这样在BNA和PNA杂交的基础上,在BNA的一端或两端就形成了突出端序列。

BNA媒介聚合在TNA上的具体例子见SEQ ID NO. 40-43。在这一例子中,SEQ ID NO. 40编码2个1/2 TBR,这2个1/2 TBR与TNA上的2个1/2 TBR杂交而形成2个NF- $\kappa$ B结合位点,同时还提供了一个噬菌体左操纵子1/2 BBR,该序列的3'端是一个限制酶*Pst*I的识别位点。加入BNA(SEQ ID NO. 41)及与PNA中的1/2 BBR互补的1/2 BBR(SEQ ID NO. 40),就形成一个完整的BBR同时形成一个完整的*Pst*I识别位点,留下一个4碱基的突出端用于和额外的BNA杂交。因此,加入3'端有一个4碱基序列的SEQ ID NO. 42,

其与 SEQ ID NO. 4 和 41 杂交后剩下的 4 个碱基的突出端互补。此外, 给 SEQ ID NO. 42 在 5' 端装配一个 5 碱基的序列, 该序列是 *Bam*HI 识别位点的组成部分。通过加入 SEQ ID NO. 43 的 BNA, 可进一步扩展 BNA 生长聚合体, SEQ ID NO. 43 的 BNA 和 SEQ ID NO. 42 互补, 形成一个完整的 BBR 同时形成完整的 *Bam*HI 识别位点, 并且, 留下一个还可和具有互补序列的 BNA 杂交的 4 碱基的突出端。通过这种方式, 可以广泛的杂交 BNA, 结果极大地扩增了单一 PNA 与 TNA 杂交作用的信号。

和实施例 1 中所述的 PNA 一样, 利用 BNA 可以以不标记形式或按照本领域已知的和实施例 1 所描述的方法对其进行标记。产生 BNA 聚合物时, 向 PNA-TNA 复合体直接加入 BNA 而形成 BNA 聚合体是值得欣赏的, 而不是通过向 PNA-TNA 复合体中顺序添加 BNA。形成这种 BNA 聚合物的一个简易方法包含重组产生载体, 在该载体中给 BNA 聚合物在其任何一端提供一个独有的限制酶切位点。把含有多个 BBR 的多聚体 BNA 从载体上切下来, 并和基于 PNA 和 TNA 杂交保留在 PNA 上的单链 1/2 BBR 杂交。通过在 PNA 上提供和 BNA 聚合物上的突出端互补的单链序列来完成这一过程, 其中突出端是 BNA 聚合物从产生载体上切割下来时产生的。

### 实施例 3-HNA 的产生和它们用于加帽 BNA 聚合物的用途

按照本领域用于产生核苷酸的已知方法, 如例 1 和例 2 中产生 PNA 和 BNA 的方法产生本发明的 HNA。但是, 在生产 HNA 的过程中, 特异设计 HNA 的序列, 以便 HNA 的实质性部分形成一个自我配对的回文序列而形成一个发夹结构, 同时, 在剩余的单链部分形成足够的碱基以便能与例 2 所描述的 BNA 的生长链的单链序列杂交。

在这一实施例中, 提供 HNA (SEQ ID NO. 44) 以加帽例 2 中在加入 BNA (SEQ ID NO. 43) 之后, 和 PNA 杂交的 BNA 的突出端。因为 SEQ ID NO. 44 除了具有一个形成一个稳定发卡结构的回文序列, 还在 HNAs 的 5' 端有一段序列, 这段序列完善通过 SEQ ID NO. 42 和 43 的杂交而形成的 *Bam*HI 序列, 所以上述过程能够完成。自然, 在说明本发明的过程中, 只加入 3 个 BNA 而终止聚合物是为了简便说明的目的。如上所述, 这一聚合过程实质上可以无限地连续扩增 PNA-TNA 杂交过程中的信号。一旦 HNA 与 BNA 的生长链杂交, 则加帽该聚合物, 并且失去聚合物的突出端。

#### 实施例 4-TBA 和 BBA 的制备、标记和其固定

根据本发明可以使用的 TBA 和 BBA 包括任何特异性地和 PNA、TNA 和 BNA 杂交后形成的 TBR 和 BBR 结合的物质。使用 DNA 结合蛋白形成这种物质的一例。

在这个例子中，所说的 TBA 是 DNA 结合部分 p50 的二聚体，所说的 BBA 是  $\lambda$  *cro* 蛋白。这些蛋白可以通过本领域已知的方法生产。已经克隆了编码这些蛋白的基因。这样，依据本领域已知的方法能够重组产生这些蛋白并对其进行纯化。并且可按照本领域已知的方法标记这些蛋白，或者通过放射性同位素标记（如放射性碘），或者用酶标记（如  $\beta$ -半乳糖酶或辣根过氧化物酶），或者通过荧光染料标记（如荧光素或罗丹明）。另外，可以把 TBA 和 BBA 中的任何一个或两个固定在固相表面，如微量滴定板表面或微粒的表面（直径介于 0.01 至 10 微米的彩色粒子）。TBA 和 BBA 的标记可以相同也可以不同。

在这一例子中，含有二聚 p50 DNA 结合区的 TBA 用罗丹明标记，而 BBA，*cro* 蛋白，用荧光素标记。因此，经过和本专利公开内容和前面与下面的例子所描述的 PNA、TNA、BNA 和 HNAs 的杂交，核酸杂交体（如果形成），和过量标记的 TNA 和 *cro* 接触。这些标记的荧光信号通过已知的方法来测量，二种信号的检测结果都表明 TNA 上存在 1/2 TBR 序列。NF- $\kappa$ B 和 *cro* 的荧光产生的差异信号能够量度 BNA 在 PNA-TBA 杂交体上的聚合引起的信号扩增程度。按照本发明的方法，预计扩增可从一倍至一千倍。

#### 实施例 5-两个 PNA 和一个 TNA 的杂交以及 TNA 和 CNA 的区别

所用的 PNA，PNA1 (SEQ ID NO. 40) 和 PNA2 (SEQ ID NO. 45)，以十倍于实验样品中 TNA 的摩尔浓度的量使用。在该实施例中，一个分离的双链 HIV LTR，其中一条链具有图 7 所示的 SEQ ID NO. 37 序列，另一条链与图 7 所示序列互补，用作 TNA。在此实施例中也使用了一个分离的双链 CNA，其中的一条链与 SEQ ID NOS. 37 的序列相同，除了在图 7 所示的第一个 NF- $\kappa$ B 结合位点的中心，即图 7 的 1 位，“A”代替“T”；因此互补链在那个位点上和 SEQ ID NOS. 40 的 PNA 错配。

把 SEQ ID NOS. 40 和 SEQ ID NOS. 45 的序列都加入到不同的反应中，第一个序列包括上述的 TNA，而第二个序列包括上述的 CNA。把样品溶解在

合适的杂交缓冲液中，例如 10mM Tris (pH7.5)、1mM EDTA。把样品加热到 90 °C 保持大约 5 分钟，使样品中双链的 TNA 和 CNA 链分离，然后冷却样品使 PNA，TNA 和 CNA 的链退火。

杂交结束后，可以根据已知的方法检测杂交，如根据已知的方法计算基于碱基组份的  $t_{1/2}$  和退火温度，SEQ ID NOS. 40 的 PNA 和加入的 BNA 多聚合(和实施例 2 一样)，SEQ ID NOS. 45 的 PNA2 探针与始于 Sph1 识别位点突出端的 BNA 进行多聚合。在加入 BNA 并短时间杂交后，不同的样品分别加入用共价固定的 NF-kB 包被的颗粒上，允许 NF-kB 和 TNA 及 CNA 样品中形成的任何 TBR 结合。大约结合 15 分钟后，用大约 3 倍体积的合适洗脱缓冲液洗脱样品两次，所用的缓冲液如 10mM Tris (pH7.5)、100mM 氯化钠，或已确定不会影响 NF-kB 或噬菌体  $\lambda$  CI 阻抑蛋白结合活性的另外缓冲液。每次洗脱后，颗粒都通过重力或短时间离心而沉降下来。这样就去除了那些没有正确地经 PNA1 和 TNA 杂交而形成的 NF-kB 结合位点的核酸。

最后一次洗脱后，把标记的噬菌体  $\lambda$  的 CI 阻抑蛋白加到每一个样品中，这些噬菌体  $\lambda$  CI 阻抑蛋白用放射性同位素标记(如放射性碘)、或用酶标记(如辣根过氧化物酶)、或者用带颜色的颗粒、或者用萤光标记物标记。然后用几倍体积(约 2 倍)的合适洗脱缓冲液洗脱样品几次(约 3 次)，缓冲液如 10mM Tris, pH7.5、100mM 氯化钠，或已确定不会影响 NF-kB 或噬菌体  $\lambda$  CI 阻抑蛋白结合活性的另外缓冲液。每次洗脱后，颗粒都通过重力或短时间离心而沉降下来。最后一次沉降或离心之后，通过检测结合的放射性量、酶分析中释放的颜色、结合的颗粒的颜色、或萤光来定量结合的标记物。另外，也可以加入一种抗-CI 的抗体并使用一种标准的三明治式酶联免疫测试或放射免疫测试来检测结合的阻抑蛋白。另外，作为阴性对照(背景)，所有的上述操作都在另一种样品中依次进行，这种样品中的颗粒不含固定的 NF-kB。

前面检测的结果是，对照和含 CNA 的样品都具有相似的低度信号而含 TNA 的样品则具有远远高于背景的信号。

## 实施例 6-A. 检测 HIV 的试剂盒

### A. 试剂盒组成:

#### 1. 微量滴定板

2. 在 tris 缓冲液盐水中重组产生的 NF-kB 溶液 1mg/ml
3. 含有单链 HIV PNA 的管(预先混合的编码 2 个 NF-kB1/2 结合位点的寡核苷酸混合物, 即 SEQ. ID. No. 7 和 8 的混合物)
4. 含有单链人基因组 PNA (SEQ ID NO. 1) 的管
5. 含有核酸酶(*Pst*I) 的管
6. 含有蛋白酶的管
7. 含有预先聚合的 BNA s, 100 个重复单位的噬菌体  $\lambda$  O<sub>r</sub> 的管, 其被 HNA 加帽但具有自由的 1/2 BBR 以便结合 PNA-TNA 杂交体
8. 含有结合 *cro* 的辣根过氧化物酶(hrp) 的管
9. 含有 hrp 染色底物的管
10. Tris 缓冲液盐水 100ml
11. 小刀
12. 反应管 A、B、C, 每个管含有 250ml 蒸馏水
13. 医用滴管

#### B. 检测方法

(a) 用在 tris 缓冲液盐水中重组产生的 NF-kB 溶液(项 2) 1mg/ml, 浸没微量滴定板(项 1), 4 °C 下旋转过夜。

(b) 用小刀(项 11) 刺破手指取出三滴试验者的血, 在 A、B、C 反应管(项 12) 中各悬浮一滴血。

(c) 在每个管中用医用滴管(项 12) 悬浮一滴蛋白酶溶液(项 6), 振动管, 静置 5 分钟。

(d) 用医用滴管在管 A-C 的每个管中滴一滴核酸酶溶液(项 5), 振动管, 然后静置 10 分钟。

(e) 把一滴项 3 的溶液加入管 A 中(试验样品); 一滴项 4 的溶液滴入管 B 中(阳性对照); 一滴盐水(项 12) 加入管 C 中(阴性对照)。将管在热水中加热至 50 °C, 然后经一小时冷却至室温。

(f) 杂交可在步骤(d) 中发生时, 从表面和微量滴定板上除去多余的蛋白(步骤(a) 中), 用 Tris 缓冲液盐水冲洗盘子(管 10)。

(g) 从步骤(e) 始把管 A-C 中的样品转移到有 3 个微孔的微量滴定板中, 放置 1 个小时, 不断摇动。

(h) 用 Tris 缓冲液盐水冲洗含有管 A-C 内容物的微量滴定板微孔, 再倒空。

(i) 把一滴项 7 的溶液加入到每个微孔中, 与结合在盘上的任何 1/2 BBR 杂交, 一小时后, 再用 Tris 缓冲液盐水冲洗三次。

(j) 把一滴项 8 的溶液加入到每个微孔中, 使 *cro* 可以与任何附着的 BNA s 结合, 10 分钟后, 然后再用 1 毫升 Tris 缓冲液盐水洗脱 5 次。

(k) 在每个微孔中加入一滴 hrp 底物, 颜色反应即可发生。

### C. 结果

如果微孔 A 和 B 都显出颜色反应, 而微孔 C 中无颜色反应, 检验结果有效, 被检对象感染了爱滋病。如果仅仅微孔 A 显示颜色反应, 或者微孔 C 显出颜色反应, 则检验的操作不正确, 结果无效。如果微孔 A 和 C 不显示颜色反应但是微孔 B 显示颜色反应, 则该检验有效, 该个体没有感染爱滋病。

### 实施例 7-各种新型 TBA 的产生

根据本发明使用的新型 TNA 按如下过程制备:

(a) NFkB / NF-kB (HIV-Detect I). 把编码任意一种 SEQ ID NOS. 63-71 或一种类似于 NF-kB DNA 结合蛋白的核酸以读框形式融入到另一个编码装配序列, 如 *cro* 的核酸序列上, 这样 NF-kB DNA 识别序列可在 *cro* 序列的氨基端或者羧基端编码。在 NF-kB 序列和 *cro* 序列之间可以提供也可以不提供一个接头序列。在 *cro* 的另一端, 可以提供也可以不提供一个核定位序列, 例如 SEQ ID NOS. 72。并且, 在 *cro* 末端可以提供也可以不提供不被 NF-kB 识别序列利用的不对称序列。完整的 TNA 序列的例子如下所示。

(b) NF-kB / SP1 (HIV- Detect II). 以上文(a)中描述的相似方式, 制备一个编码一个 NF-kB 识别区的重组编码序列。在另一个构建中, 包括编码 SP1 DNA 识别部分的编码序列而不是 SEQ ID NOS. 63-72。这一序列编码 SEQ ID NOS. 73 的序列的所有部分或功能部分, 这部分是具有 DNA 结合功能的 SP1 转录因子的组成部分(参见 Kadonaga 等 [1987] 细胞 51:1079-1090)。把 NF-kB 编码载体和 SP1-编码载体共转染到一个适合的表达系统, 这种表达系统是本领域公知的。在 SP1 和 NF-kB 识别单位通过陪伴装配之后, 加入单体的 NF-kB 识别单位以完成 NF-kB 识别二聚体的聚合。不对称

序列阻止 NF- $\kappa$ B 或 SP1 二聚体的形成, 相反却指导 NFKB-SP1 异源二聚体的形成(即, 爱滋病-检测 II), 然后通过本领域已知的方法把这种异源二聚体从表达系统(哺乳动物或细菌细胞)中分离出来。

(c) SP1/SP1 TBA (HIV- Detect III). 如上文(b)所述, 制备 SP1-编码的 TBA 构建体。但是, 只有把构建体转染到表达系统时, 才具有允许 SP1-SP1 二聚体形成的不对称序列。

(d) SP1-TATA (HIV- Detect IV). 如上文(b)所述, 产生 SP1-编码的 TBA 重组体。此外, 用和 SP1 TBA-编码的构建体中的序列互补的不对称序列制备了编码具有结合序列的 TBA 的重组序列, SEQ ID NO. 74 或其它编码 TATA 结合位点的序列。通过标准的方法把这些构建体共转染并把异源二聚体分离出来, 这些方法包括在具有适当的 SP1-TATA 靶结合区的 DNA 柱上的亲和纯化法。

(e) SP1-E2 (HPV- Detect I). 按上文(b)所述的方法制备 SP1-编码的构建体, 通过利用编码 SEQ ID NO. 75-84 和 94-98 任何一个的序列(这些序列是乳头状瘤病毒 E2 DNA 识别位点(见 Hegde 等) [1992] 自然, 359: 505-512) 或其它识别位点), 制备一种 E2 TBA-编码的构建体并把其和编码 SP1 TBA 的构建体共转化或共转染。在 E2-SP1 识别单位通过陪伴装配之后, 加入单体 E2 识别单位以形成完整的 E2 识别二聚体。按照已知的方法把异源二聚体分离出来。

(f) E2-E2 (HPV- Detect II) 如上文(e)所述, 制备 E2 TBA-编码的构建体, 除了包含不对称序列, 这种序列允许 E2 二聚体的形成。然后把表达的二聚体用已知的方法分离出来, 这些方法包括在 DNA 亲和柱上用于二聚的 E2 的结合位点的亲和纯化方法。

(g) E2-TATA (HPV- Detect IV) 如上文(e)和(d)所述, 分别制备结合 TBA 的 E2 和 TATA, 除了包含不对称序列, 这些序列能促进异源二聚体而不是同源二聚体的形成。然后, 使这些构建体共表达并分离异源二聚体。

(h) TATA-TATA (HPV- Detect IV) 如(a)和(d)中所述, 通过使用促进同源二聚体形成的不对称序列制备和 TBA-编码的构建体结合的 TATA, 同时分离同源二聚体。

(i) 其它 TBA 如上文所述关于 HIV 和 HPV 的 TBA, 任何关于病原体或疾



病状态的 TBA 都可通过鉴定特定 DNA 结合蛋白, 以及通过使用合适接头、装配体和不对称序列而形成表达构建体来制备。

#### 实施例 8

以实施例 5 所述的类似分析方式, 通过使用实施例 6 中制备的双链 NF- $\kappa$ B-SP1 结合蛋白而产生一种更严格的分析方法。因此, 可以加长图 7 所示和实施例 5 所使用的探针以缩小探针间距离从而减少 TNA 中 DNA 的自由度。

#### 实施例 9-“高级” TBA 的产生

通过恰当使用不对称序列, 产生 TBA, 这些 TBA 是特定 DNA 识别单位的二聚体, 三聚体, 四聚体, 五聚体或六聚体。在这一方式中, 产生一个六聚体 TBA, 先要通过使用能够促进二聚体形成的不对称序列产生一个 NF- $\kappa$ B p50 二聚体 TBA。此外, 这个不对称序列可以促使这个 p50 二聚体与一个 SP1-SP1 二聚体四聚化。最后, 另外加入不对称序列指导具有核定位序列的二聚体参入而形成六聚体。这个过程是通过加入, 例如, 胰岛素的不对称序列, 其实质上形成六聚体。SEQ ID NOS. -85(A) 和 86(B)、87(A) 和 88(B)、89(A) 和 90(B)、以及 91(A) 和 92(B) (参与图 13 和 14) 指导这一六聚体的形成。

因为其对 HIV-LTR 有非常高的亲合力, 可以通过使用一个多聚的 TBA 来产生此六聚体, 具有这种结构并用于此目的多聚体化合物本文称为“HIV-Lock”。

最佳 HIV-Lock 是通过结合在 HIV-LTR 的 TBR 上的 TBA 的足迹法(按照本领域公知的方法)来定义的, 用这种方法来证实对多聚的 TBA 复合体的形成具有贡献的每个 DNA 结合蛋白的结合亲合力, 参照能衍生出 TBA 的 DNA 结合识别单位的任意自然靶序列(如 CNA)的亲合力而下移。任何对 HIV TBA 亲合力的并发丢失都可通过下述的聚合物形成而得到极大地补偿。

TBA 每个组份与其 TBR 的结合和通过不对称序列形成二聚体的装配过程存在竞争。这种竞争可以通过在每个 TBA 组份中调整陪伴和不对称序列之间的接头而解除, 这样就解开了这些竞争过程。所造成的 TBA 不对称性和装配组份在扩散方向上的降低(增加了有效浓度)导致了多聚复合体有效地形成。

在足迹的基础上, 调整接头的长度和组成以便更好地区别靶 HIV 序列和

自然序列。在这种方式下，虽然每个 TBA 组份将对 CNA 和 TBR 序列具有低亲合力，但聚合的复合体将对现已扩展的多聚复合体识别的 TBR 具有更高的亲合力(是 TBA 聚合物每个组份识别的 TBR 的亲合力的平方)，同时对 CNA 仍具有低亲合力。以同样方式，制备除 HIV-Lock 以外的其它多聚的 TBA 复合体。

以这种方式产生的 TBA 包括如下序列，它们可以通过连接蛋白亚单位或编码这些亚单位的核酸序列而装配起来，如下所示：

组合(set)	连接的序列组
A	I + II + III
B	IV + V + III
C	IV + III

其中 I-V 组所包括的序列选自：

组	所选自的序列
I	SEQ ID NO. 85-92 的任何序列
II	Met Ser，连接到 SEQ ID NO. 104-106 的任何序列上，其各自连接到 SEQ ID NO. 99 上。
III	连接到 SEQ ID NO. 75-84 或 94-98 之任一上的 SEQ ID NO. 100，连接到 SEQ ID NO. 74 或 SEQ ID NO. 93 上的 SEQ ID NO. 101；或连接到 SEQ ID NO. 74 或 SEQ ID NO. 93 上的 SEQ ID NO. 102；或者 SEQ ID NO. 72, 103, 73, 或 63-71 的任何序列。
IV	SEQ ID NO. 104-106 的任何序列
V	SEQ ID NO. 99。

这样的 TBA 的具体例子是 SEQ ID NO. 109-116，按如下装配：

组合	SEQ ID NO.	连接的序列
A	109	85+ 蛋 氨 酸 丝 氨 酸 +104+99+100+94

A	110	85+蛋氨酸丝氨酸+104+99+72
A	111	86+ 蛋 氨 酸 丝 氨 酸 +105+99+102+74
A	112	86+蛋氨酸丝氨酸+106+99+73
A	113	89+蛋氨酸丝氨酸+106+99+63
C	114	106+64
C	115	105+64
B	116	106+99+73

以这种方式，通过在合适的不对称序列、装配序列和 DNA 的识别单位之间进行选择，可以形成许多不同的 TBA。并且，这些系列，例如 SEQ ID NO. 114 和 115，将和非 SEQ ID NO. 114 或 115 的二聚体的其它序列结合，由于在突变装配序列中的电性排斥作用不会形成二聚体 (SEQ ID NO. 104 是 *cro*；SEQ ID NO. 105 是一个新型突变的，负电性的 *cro*，SEQ ID NO. 106 是一个新型突变的正电性的 *cro*)。

当然，对于给定的这些 TBA 的氨基酸顺序，本领域的一般技术人员可以生产出编码这些序列的核酸克隆，而这些重组克隆自然是本发明的组成部分。

#### 实施例 10-用“HIV-LOCK”检测 HIV

将与实施例 6 非常相似方法，用实施例 9 产出的“HIV-LOCK”作为 TBA，试剂 2，得到类似结果。

#### 实施例 11-用“HIV-LOCK”对捐献血样进行 HIV 检测

当不限制待测血的质量时，检验捐献的血样是否有 HIV 污染，使用类似实施例 6 的检验，但是对 A-C 的每个管，把约 5 毫升血样在桌式离心机上离心沉淀下来。其它试剂用量根据需要按比例加大以便于处理样品中大量的 TNA。

#### 实施例 12-作为抗 HIV 治疗剂的“HIV-LOCK”

把按照实施例 9 产生的“HIV-LOCK”在脂质体中制成 1mg / ml 的溶液，然后将其静脉注射到通过检测已确认感染 HIV 的个体中。注射剂量为每千

克体重约 0.1mg 至 100mg 的“HIV-LOCK”，24 小时后，监测病人血清中的 HIV p24 的浓度。有必要经常重复这种治疗，如血清中 p24 水平上升时。

#### 实施例 13-HIV-TBA 构建体作为治疗剂的使用

使用重组逆转录病毒或类似载体把一个编码结合 TBA 的 HIV-LTR 的构建体传送到感染的病人身上。载体编码陪伴，如 *cro*，和用于 p50 结合部分的 DNA 序列。同样的载体还编码一个陪伴，SP1 TBA 在其上折叠。提供不对称序列，以便当 p50-TBA 和 SP1-TBA 共同在一个 HIV 感染的细胞中进行体内表达后，这些 TBA 之间立即进行结合，同时防止任何在 p50 DNA 结合区与体内 p50 或 p65 单体之间的结合。在 TBA 中也提供了 NLS，以便一旦形成二聚体，TBA 快速定位至细胞核中并与整合的 HIV 序列特异性结合，于是防止了任何从那个位点的转录。

为了此目的，需要选择编码 DNA 结合区的序列以便把表达的单体装配在不结合到人体自然序列的 TBA 上。这样，只有当 TBA 的组分和其靶序列结合后，TBA 的所有组分间才能结合，形成一个能够紧密地特异性地结合 HIV-LTR 的复合体。

#### 实施例 14-用于人类乳头状瘤病毒的诊断试剂盒

用于人类乳头状瘤病毒的这种诊断剂利用了已知的有差异的良性和恶性 HPV，提供了一个能表明病人体内恶性肿瘤的敏感性的检测手段。人类乳头状瘤病毒是一组与高等脊髓动物良性扁形上皮细胞肿瘤相关的小 DNA 病毒。至少已发现了 27 种不同类型的 HPV；其中许多和特定的医疗损害相关。其中有 5 种，HPV-6，HPV-11，HPV-16，HPV-18 和 HPV-33 和人生殖道的损伤相关。总的来说，已经发现 HPV-6 和 HPV-11 的 DNA 与生殖道的良性损伤相关。已经发现 HPV-16，HPV-18 和 HPV-33 与预恶性和恶性损伤相关，并且能在大多数由子宫颈肿瘤细胞系产生的细胞中转录。HPV-16，HPV-18 和 HPV-33 可能仅是一大系列与恶性人子宫颈肿瘤相关的 HPV DNAs 的 2 个成员。

动物模型表明良性乳头状瘤病毒损伤可以在一个辅致癌物质存在的情况下发展成为恶性损伤。已经在子宫颈肿瘤的转移瘤中发现了 HPV DNA。在恶性子宫颈肿瘤损伤中，HPV DNA 一般整合在人基因组上，但也有一些染色体外的 HPV DNA 存在。HPV 的整合形成原病毒，一般导致病毒 E2 开放

阅读框架(ORF)的破裂。尽管使 E2 开放阅读框架破裂,对几个子宫颈癌细胞系的检查显示了其的转录活性并整合了 HPV-16 和 HPV-18。检测存在于人子宫颈癌细胞系 SiHa 和 CaSKi 中的 HPV-16 基因组时,发现 HIV-16 的整合存在差异。在 SiHa 细胞系中,单一 HPV-16 基因组的整合发生在碱基 3132 和 3384 之上, E1 和 E2 ORFs 的破坏是缺失了 0.3KB 的序列。另外 50 个碱基在 HPV-16 DNA 上的缺失,造成了 E2 和 E4 ORFs 的融合。HPV-16 DNA 的 5' 部分,包括损坏了的 E2 ORF, 连接到连续的人右端序列上。另外,在 E1 ORFs 中间的 1138 位上检测到了单一添加的鸟嘌呤。碱基的添加使 E1a 和 E1b ORFs 融合成一个单一 E1 ORFs。

可以在基因库中得到 HPV-16 的完整基因组,保藏号为 K02118;可以在基因库中得到 HPV-33 的完整基因组,保藏号为 M12732;可以在基因库中得到 HPV-18 的完整基因组,保藏号为 X05015。

作为初始筛选,对于一个给定的子宫颈活体检查样品的 HPV 感染事实的确定是通过使用,例如 PNA SEQ ID NO. 46-53 的任何一个或全部和一个上述的 E2 TBA 的简单“是/不是”的分析类型建立起来。(即片断 DNA, 结合 PNA, 用 TBA 固定, 和用 BNA 和 BBA 进行检测)。

一旦发现活组织检查的样品呈现 HPV 阳性反应,通过分析病毒在人基因组的整合状态来获得关于 HPV 恶化可能的额外信息。

1. 把宫颈活组织检查样品的 DNA 切成片断,杂交到具有 SEQ ID NO. 60 序列的阻断探针上。这一探针和未切出 0.3kb 片段的所有 DNA 片段结合。

2. 将活组织检查样品中的 DNA 暴露给具有 SEQ ID NO. 61 序列的 PNA。该探针只和缺失 0.3kb 片段的 DNA 片段结合(如果存在大的缺失片段,阻碍探针将阻碍其环出)。

3. 具有 SEQ ID NO. 62 序列的 PNA 与 SEQ ID NO. 41 序列杂交形成一个 BBR, 这种 BBR 再与作为 BBA 的 *cro* 或  $\lambda$ CI 阻抑蛋白结合,留下一条能与 SEQ ID NO. 61 序列上 TATA 位点杂交的单链段。这有助于在缺失的大段序列 5' 端形成 TBR。

4. 用具有 TATA 结合蛋白 DNA 识别单位的 TBA 固定 TBR。

5. 通过加入上述的 BNA 和 BBA 来检测结合的片断。

该分析的检测信号表明存在于 TNA 的 HPV 上有大片段 DNA 缺失。既然这

种缺失与恶性化有关, 此分析方法使得能深入了解 HPV 感染恶化的可能。通过进行基于这 52-碱基片段缺失的类似分析可以证实这一结论, 这段碱基片断和 HPV-诱导的恶性化有关。

可以选择用于 TBA 的 TBP 识别单位来用于此分析, 如, 从 SEQ ID NO. 70 和 SEQ ID NO. 93 中进行选择。

#### 实施例 15-重组 HIV-LOCK™ 的产生

**第一阶段** 制备用以产生 HIV-LOCK™ 的 DNA。通过使用 MutaGene 的噬菌粒试剂盒来完成自然发生并已克隆的需要修饰的 HIV-LOCK™ 的组份。修改流程包括使用一个含有 HIV-Lock™ 的每一个结合组份的 Blue-script 质粒。把这些质粒转化到感受态细胞中, 培养含有尿嘧啶的噬菌粒。提取单链 DNA 并作为突变链的模板。合成包括所需突变(突变中包含并入一个新的限制酶切位点)的寡聚核苷酸, 然后再用多聚核酸激酶和 ATP 处理。激酶处理过的寡聚核苷酸退火后和单链模板结合, 按照 MutaGene 的程序(只是测序酶 2.0 作为聚合酶)合成和连接一条突变链。使用 <sup>32</sup>P 末端标记的与所导入的突变互补的核酸和通过分离这种质粒 DNA 及用导入的限制酶切位点鉴别突变体来筛选文库。也可以通过测序酶试剂盒进行测序来证实突变。用多面体启动子把 HIV-Lock™ DNA 克隆到杆状病毒的表达系统中。

**第二阶段** 使用杆状病毒产生 HIV-Lock™ 蛋白。把 Sf-9 细胞培养至预定的密度(约  $1 \times 10^6$  细胞/ml, Log 指数期), 用含有 HIV-Lock™ 构件的杆状病毒感染培养的细胞, 然后收集含有 HIV-Lock™ 的重组蛋白。在大规模产生中, 细胞培养从烧瓶扩大到旋转器而后到生物反应器。感染后的第 12, 24, 36 和 48 小时收集细胞以便提取蛋白。在整个过程中监测细胞存活指数。

**第三阶段—纯化 HIV-Lock™ 蛋白。** 为了方便下面的纯化, 首先通过最大限度超速沉降离心把要收集的蛋白从微粒中分离出来。然后离心的产物经过滤消毒。提取物在 4℃ 下以 40,000rpm 的速度离心约 30 分钟, 等分试样再通过抗一种 HIV-Lock™ 组份的多克隆兔抗体进行免疫沉淀。免疫沉淀的蛋白在 10% SDS PAGE 凝胶上电泳。

**第四阶段—针对于 HIV DNA 的 HIV-Lock™ 蛋白的检测。** 通过使用含有 HIV 长末端重复序列组份的寡聚核苷酸探针和含有和 Kappa 轻链和微球蛋白调

节相关的 NF $\kappa$ B 结合的 DNA 的片段, 进行迁移率变动分析。寡聚核苷酸退火后和其互补链结合并且末端用  $^{32}$ P-ATP 标记。

室温下通过在缓冲液混合少量 ( $10^{-15}$ M) 放射性标记的 HIV LTR DNA 和稍大量的 HIV-Lock™, 10 分钟进行足迹法检测。在加入蛋白之前加入二巯基乙醇。加入 EDTA 铁(二价)、过氧化氢和抗坏血酸钠然后培养反应混合物。加入熄灭剂, 产物用变性胶电泳方法分析。这是依据不同蛋白浓度而操作的。把得到的凝胶用磷酸图像扫描仪作图, 分析得到的高分辨率图象资料, 提炼出与细胞 DNA 相关的 HIV-Lock™ 与 HIV DNA 的结合亲和力。

可以使用多种设计和检测重复以发现 HIVLock™ 和其它用于 HIV 及其它生物体的 TBA 的结合。这一过程使设计结合装配体成为可能, 这种结合装配体不和野生型蛋白竞争基因组样品中的单一结合位点。用于其它生物体的 TBA 以及用于这些生物体体内序列的 TNA 可以通过使用前面所述的方法来产生。当要产生所有核酸的 TBR 时(包括 DNA-DNA、DNA-RNA 和 RNA-RNA 杂交体以及这些杂交体的组合), 这种方法是有效的。

实施例 16-为了产生本发明的 TBA 和 BBA, 鉴别核酸结合分子的方法

在本发明的方法中, 靶结合装配体和辅助结合装配体的装配是通过鉴别核酸结合分子, 和以一种能得到区别特定靶序列和非常相关序列的 TBA 的方式连接该分子的核酸结合部分。一种鉴定核酸结合分子的方法包括如下步骤:

1. 获得含有靶核酸的生物样品。这种样品可以是, 例如, 感染病原体的一种生物体或组织提取物。

2. 破碎样品以使核酸暴露出来, 同时降低样品中核酸大小的复杂程度。

3. 把第一等分切成片断的核酸加入到对照缓冲液介质中, 再把第二等分的样品加入到含有已知分布型的核酸结合分子的对照缓冲液介质中。

4. 分析两个等分试样以鉴定含有靶结合分子的试样相对于对照试样, 行为发生改变的片断。这个过程可以通过单相凝胶电泳, 双相凝胶电泳, 高压液相色谱, 纸色谱或任何可以揭示和核酸结合分子结合的核酸片断(相对于没有结合的核酸片断)的不同行为的方法来完成。

5. 鉴定和分离那些当和核酸结合分子相接时行为的确发生改变的核酸片断和, 或对核酸结合片断进行测序以确定是否存在已知的核酸结合分子

基元，或直接鉴定结合在核酸片断上的核酸结合分子。后者可以通过，例如，把电泳核酸的双相框架和能够结合不同核酸结合分子的不同标记的抗体接触来完成。

在此方法中，优选使用核酸基元作为诊断或治疗的目的，其中的靶核酸有不止一个可利用的核酸结合分子靶。以这种方式，可以产生复杂的靶结合装配体，利用邻近的不同核酸结合分子类型以增强由该个体鉴定的核酸结合组份装配的 TBA 的特异性。然后把核酸结合分子的各个核酸结合部分装配成一个上述的完整的 TBA，例如 HIV-Lock™。

### 实施例 17-鉴定样品中特定 RNA 序列的方法

按照本发明的方法和组合体可以特异性地鉴定任何核酸序列。样品中靶 HIV RNA 的鉴定是通过获得含有 HIV RNA 的病人血样或其它生物液体或提抽物样品，然后检测 TAR 的结合位点来完成的。Tat 是一个 HIV 复制的正调节因子，其和 HIV RNA 的 TAR 区结合。以最小的自然生成机率，完全具有活性的 HIV-Tat 形式含有 72 个氨基酸，本文的 SEQ Id NO.118。Tat 至少具有 2 个功能区，和反式激活来源于 HIV 长末端重复序列(HIV LTR)的基因表达。Tat 和 TAR 中的自杂交序列形成的 RNA 茎环结构结合，茎环结构恰在 HIV LTR 的 5' 端。HIV TAR RNA 形成一个两核苷酸突起和两个茎环结构(Rhim 等 1994 病毒学: 202, 202-211)。Tat (SEQ Id.118) 结合在比 Tat 突变体具有更低的亲合力的结构上，Tat 的这种突变体，其中 58 位丙氨酸突变为苏氨酸或 65 位组氨酸突变为天冬氨酸。(Derse 等 1993, 病毒学: 194, 530-536)。本发明中这些过程的使用如下：

1. 切断生物样品使核酸暴露出来，同时降低核酸大小复杂性。
2. 将 TBA 与鉴定 TAR 结合蛋白序列的杂交体和 HIV 基因组中邻近端的侧面序列的样品接触。用于此目的 TBA 是装配到利用 Tat 作为 HIV RNA 特异性结合分子的 *cro* 陪伴上。为了提供特异性，例如 HIV TAR 位点和与其紧密相关的由于其它病原体如巨细胞病毒的存在而出现的 TAR 位点间的交叉作用，TBA 还具有一个识别 DNA-RNA 杂交体此靶的抗体组份，这种靶结合区是在探针核酸和 HIV LTR RNA 结合后产生的。
3. 通过用过量的具有更高亲合力的 Tat 异变体(58 位丙氨酸突变为苏氨酸或 65 位组氨酸突变为天冬氨酸)和反应物接触，消除由于 Tat 结合到由



污染物(表亲 RNAs)如 CMV TAR 序列而形成的 HIV RNA 的 TAR 区所引起的交叉作用。通过这种方法,由于 TBA 结合到一个表亲 RNA 而形成的单一结合过程就会被来自 Tat 异变体的核酸样品竞争下来。另一方面,通过恰当选择由抗体和 Tat 引起的双重结合的亲合力,不会从真正的靶序列上把 TBA 置换下来。图 16 具体说明了此过程。同样方法的另一方面, TBA 可以是这样一种物质,其上可以使用能够识别核酸片断的抗体,而不是使用 Tat 的突变体,并且使用的 TBA 可以是双抗体 TBA。

本方法的另一种形式,可以使用和 HIV LTR RNA 杂交的核酸探针。因此,可以产生 LTR sp1 位点的双链片断,用作靶结合区的一部分。HIV RNA 区在位于 LTR 5' 端但和其更靠近的 TAR 区的侧面。把含有 Tat 和 2 个 Sp1 结合位点的 TBA 装配以提供 Tat 结合 TAR 以及 Sp1 结合 Sp1 的结合位点。然后通过加入适当的 BNA、BBA 和 HNAs 而进行扩增和检测。在另外一种情况下,可以使用 Seq. ID. 38 和 Seq. ID. 39(见图 7)的 PNA。使用一个含有一个或多个 Sp1 结合单位和一个抗体单位的 TBA,用来结合由样品 RNA 和 Seq. ID. 38 PNA 而产生的 DNA-RNA 杂交体。然后加入适当的 BNA、BBA 和 HNAs 以便扩增信号。

自然,本领域的技术人员会认识到可以使用其它 TBA 和 TNA 的组合来优化本文所列举的方法。

应该理解,本文所描述的实施例和实施方案只是为了具体说明本发明,本领域的技术人员可以根据其进行修改和变化,并且这些修改和变化包含在本申请的精神和范围内以及所附的权利要求书的范围内。应该理解本文所提供的序列只是作为示范,并且其它受到这些序列提示的类似序列也可以在本发明的方法中使用。应该理解尽管本文提供的序列可以设计成线性,但可以以环状或其它替换性形式使用,同时,尽管这些序列不是作为反义链而设计的,但是可以以编码或非编码形式或者和编码或非编码互补序列结合的形式使用。

## 序列表

## (1)一般信息:

## (i)申请人:

申请人名称:THE GENE POOL, INC.

街道地址:300 Queen Anne Ave.N., Suite 392

城市:Seattle

州/省:华盛顿

国家:美国

邮编:98109-4599

电话号码:(206)526-8617 传真号:

## (ii)发明名称:用特异性序列组合物检测核酸的方法

## (iii)序列数:118

## (iv)联系地址:

(A)联系人:Saliwanchik & Saliwanchik

(B)街道:2421 N.W. 41st St., Suite A-1

(C)城市:Gainesville

(D)州:佛罗里达

(E)国家:美国

(F)邮编:32606

## (v)计算机可读形式:

(A)介质类型:软盘

(B)计算机:IBM PC 兼容机

(C)操作系统:PC-DOS/MS-DOS

(D)软件:PatentIn Release #1.0, 版本#1.25

(vi) 目前申请资料:

(A) 申请号:

(B) 申请日:

(C) 分类:

(viii) 律师/代理人信息:

(A) 姓名: Bencen, Gerard H

(B) 登记号: 35, 746

(C) 参考/证书号: GP-100.C1

(ix) 电讯信息:

(A) 电话: (904)375-8100

(B) 传真: (904)372-5800

(2) SEQ ID NO:1 信息:

(i) 序列特征:

(A) 长度: 13 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:1:

**TGGGGATTCC CCA**

(2) SEQ ID NO:2 信息:

(i) 序列特征:

(A) 长度: 13 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 2:

**AAGGGACTTT CCC**

13

(2) SEQ ID NO: 3 信息:

(i) 序列特征:

(A) 长度: 13 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 3:

**AGGGGACTTT CCG**

13

(2) SEQ ID NO: 4 信息:

(i) 序列特征:

(A) 长度: 15 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 4:

**GCTGGGGACT TTCCA**

(2) SEQ ID NO: 5 信息:

15

(i) 序列特征:

(A) 长度: 15 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 5:

**ACAAGGGACT TTCCG**

(2) SEQ ID NO: 6 信息:

15

(i)序列特征:

(A)长度:13 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:6:

**CCGGGTTTTTC CCC**

13

(2)SEQ ID NO:7 信息:

(i)序列特征:

(A)长度:27 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:7:

**AAGGGACTTT CCGCTGGGGA CTTTCCA**

27

(2)SEQ ID NO:8 信息:

(i) 序列特征:

(A) 长度: 27 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 8:

**AAGGGACTTT CCGCTGGGGA CTTTCCG**

(2) SEQ ID NO: 9 信息:

27

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 9:

**GCTGGGGACT TTCCAGGGAG GCCTGG**

(2) SEQ ID NO: 10 信息:

26

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 10:

**GCTGGGGACT TTCCAGGGGA GGTGTG**

26

(2) SEQ ID NO: 11 信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 11:

**GCTGGGGACT TTCCGGGGAG CGTGGC**

26

(2) SEQ ID NO: 12 信息:



(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 12:

**GCTGGGGACT TTCCGGGGAG GCGCGG**

26

(2) SEQ ID NO: 13 信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 13:

**GCTGGGGACT TTCCAGAGAG GCGTGG**

26

(2) SEQ ID NO: 14 信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 14:

**GCTGGGGACT TTCCAGGGGA GCGTG**

26

(2) SEQ ID NO: 15 信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 15:

**GCTGGGGACT TTCCAGGGAG GCGTGG**

26

(2) SEQ ID NO: 16 信息:

(i) 序列特征:

(A) 长度: 26 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 16:

**GCTGGGGACT TTCCAGGGAG GCTGCC**

26

(2) SEQ ID NO: 17 信息:

(i) 序列特征:

(A) 长度: 33 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 17:

**TTTCCAGGGA GGCGTGGCCT GGGCGGGACT GGG**

33

(2) SEQ ID NO: 18 信息:

(i) 序列特征:

(A) 长度: 33 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 18:

**CGTGGCCTGG GCGGGACTGG GGAGTGGCGT CCC**

33

(2) SEQ ID NO: 19 信息:

(i) 序列特征:

(A) 长度: 45 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 19:

**CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCCT**

45

(2) SEQ ID NO: 20 信息:

(i) 序列特征:

(A) 长度: 46 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 20:

**CAGCAAGGGA CTTTCCGCTG GGGACTTCC AGGGGAGGTG TGGCCT**

46

(2) SEQ ID NO: 21 信息:

(i) 序列特征:

(A) 长度: 46 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 21:

**CATCAAGGGA CTTTCCGCTG GGGACTTCC AGGGGAGGTG TGGCCT**

46

(2) SEQ ID NO: 22 信息:

(i) 序列特征:

(A) 长度: 46 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 22:

**CAACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGGAGGTG TGGCCT**

46

(2) SEQ ID NO: 23 信息:

(i) 序列特征:

(A) 长度: 45 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 23:

**CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCAT**

45

(2) SEQ ID NO: 24 信息:

(i) 序列特征:

(A) 长度: 44 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 24:

**CTACAAGGGA CTTTCCGCTG GGGACTTTC GGGGAGCGTG GCCT**

44

(2) SEQ ID NO: 25 信息:

(i) 序列特征:

(A) 长度: 44 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 25:

**CTACAAGGGA CTTTCCGCTG GGGACTTTC GGGGAGGCCG GCCT**

44

(2) SEQ ID NO: 26 信息:

(i) 序列特征:

(A) 长度: 45 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 26:

**CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGAGAGGCGT GGACT**

45

(2) SEQ ID NO: 27 信息:

(i) 序列特征:

(A) 长度: 46 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 27:

**CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGGAGGCG TGGACT**

46

(2) SEQ ID NO: 28 信息:



(i) 序列特征:

(A) 长度: 46 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 28:

**CTACAGGGGA CTTTCCGCTG GGGACTTTC AGGGAGGCGT GGGGAG**

46

(2) SEQ ID NO: 29 信息:

(i) 序列特征:

(A) 长度: 43 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 29:

**CTACAGGGGA CTTTCCGCTG GGGACTTTC AGGGAGGCTG CCT**

43

(2) SEQ ID NO: 30 信息:

(i) 序列特征:

(A) 长度: 48 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 30:

**CTTCCGCTG GGGACTTTC AGGGAGGCGT GGCCTGGGCG GGA**

48

(2) SEQ ID NO: 31 信息:

(i) 序列特征:

(A) 长度: 45 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 31:

**TTTCCAGGGA GGCCTGGCCT GGGCGGACT GGGGAGTGGC GTCCC**

45

(2) SEQ ID NO: 32 信息:

(i) 序列特征:

(A) 长度: 59 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 32:

CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCCTGGGCG GGACTGGGG 59

(2) SEQ ID NO: 33 信息:

(i) 序列特征:

(A) 长度: 59 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 33:

TTTCCGCTGG GGACTTTCCA GGGAGGCGTG GCCTGGGCGG GACTGGGGAG TGGCGTCCC 59

(2) SEQ ID NO: 34 信息:

(i) 序列特征:

(A) 长度: 70 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 34:

```
CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCCTGGGCG GGA
```

60

```
CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCCTGGGCG GGA
```

70

(2) SEQ ID NO: 35 信息:

(i) 序列特征:

(A) 长度: 61 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 35:

---

TATCACCGCC AGTGGTATTT ATGTCAACAC CGCCAGAGAT AATTTATCAC CGCAGATGGT 60  
T 61

## (2)SEQ ID NO:36 信息:

## (i)序列特征:

(A)长度:64 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

## (ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

## (xi)序列描述:SEQ ID NO:36:

TATCACCGCA AGGGATAAAT ATCTAACACC GTGCGTGTTG ACTATTTTAC CTCTGGCGGT 60  
GATA 64

## (2)SEQ ID NO:37 信息:

## (i)序列特征:

(A)长度:70 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

## (ii)分子类型:cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 37:

```
CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGCGT GGCCTGGGCG GGACTGGGGA      60
GTGGCGTCCC                                         70
```

(2) SEQ ID NO: 38 信息:

(i) 序列特征:

(A) 长度: 37 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 38:

```
CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGG                                         37
```

(2) SEQ ID NO: 39 信息:

(i) 序列特征:

(A) 长度: 22 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:39:

**CGGGACTGGG GAGTGGCGTC CC**

(2)SEQ ID NO:40 信息:

22

(i)序列特征:

(A)长度:103 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:40:

**CTACAAGGGA CTTTCCGCTG GGGACTTTCC AGGGAGGTAT CACCGCCAGT GGTATTTATG**

60

**TCAACACCGC CAGAGATAAT TTATCACCGC AGATGGTTCT GCA**

103

(2)SEQ ID NO:41 信息:

(i)序列特征:

(A)长度:62 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)MÅ

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:41:

<b>GAACCATCTG CGGTGATAAA TTATCTCTGG CGGTGTTGAC ATAAATACCA CTGGCGGTGA</b>	60
<b>TA</b>	62

(2)SEQ ID NO:42 信息:

(i)序列特征:

(A)长度:71 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:42:

<b>GATCCAACCA TCTGCGGTGA TAAATTATCT CTGGCGGTGT TGACATAAAT ACCACTGGCG</b>	60
<b>GTGATACTGC A</b>	71

(2)SEQ ID NO:43 信息:



(i) 序列特征:

(A) 长度: 63 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 43:

```
GTATCACCCGC CAGTGGTATT TATGTCAACA CCGCCAGAGA TAATTTATCA CCGCAGATGG      60
TTG                                          63
```

(2) SEQ ID NO: 44 信息:

(i) 序列特征:

(A) 长度: 21 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 44:

```
GATCCGGGGG GATACCCCC G
```

21

## (2)SEQ ID NO:45 信息:

## (i)序列特征:

(A)长度:91个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

## (ii)分子类型:cDNA

## (iii)假拟:无

## (iv)反义:否

## (xi)序列描述:SEQ ID NO:45:

```

CGGGACTGGG GAGTGGCGTC CCTATCACCG CAAGGGATAA ATATCTAACA CCGTGCGTGT      60
TGACTATTTT ACCTCTGGCG GTGATAGCAT G                                          91

```

## (2)SEQ ID NO:46 信息:

## (i)序列特征:

(A)长度:53个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

## (ii)分子类型:cDNA

## (iii)假拟:无

## (iv)反义:否

(xi) 序列描述: SEQ ID NO:46:

CTAAGGGCGT AACCGAAATC GGTGAACCG AAACCGGTTA GTATAAAGC AGA

53

(2) SEQ ID NO:47 信息:

(i) 序列特征:

(A) 长度: 54 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:47:

AAAAGGGAGT AACCGAAAAC GGTCGGGACC GAAAACGGTG TATATAAAG ATGT

54

(2) SEQ ID NO:48 信息:

(i) 序列特征:

(A) 长度: 54 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:48:

AGTAGGGTGT AACCGAAAGC GGTTCAACCG AAAACGGTGC ATATATAAAG CAAA

54

(2) SEQ ID NO:49 信息:

(i) 序列特征:

(A) 长度: 24 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:49:

GCTTCAACCG AATTCGGTTG CATG

24

(2) SEQ ID NO:50 信息:

(i) 序列特征:

(A) 长度: 24 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi)序列描述:SEQ ID NO:50:

TGTGCAACCG ATTCGGTTG CCTT

24

(2)SEQ ID NO:51 信息:

(i)序列特征:

(A)长度:24 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:51:

TATGCAACCG AAATAGGTTG GGCA

24

(2)SEQ ID NO:52 信息:

(i)序列特征:

(A)长度:24 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:52:

TGCCTAACCG TTTTCGGTTA CTTG

(2)SEQ ID NO:53 信息:

24

(i)序列特征:

(A)长度:24个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:53:

GGACTAACCG TTTTAGGTCA TATT

(2)SEQ ID NO:54 信息:

24

(i)序列特征:

(A)长度:52个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi) 序列描述: SEQ ID NO: 54:

**GACGACTATC CAGCGACCAA GATCAGAGCC AGACACCGGA AACCCCTGCC AC**

52

(2) SEQ ID NO: 55 信息:

(i) 序列特征:

(A) 长度: 53 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 55:

**GACGACACGG TATCCGCTAC TCAGCTTGTT AACAGCTAC AGCACACCCC CTC**

53

(2) SEQ ID NO: 56 信息:

(i) 序列特征:

(A) 长度: 60 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 56:

GACGACGACC TGCAGACACC ACAGACACCG CCCAGCCCCT TACAAAGCTG TTCTGTGCAG 60

(2) SEQ ID NO: 57 信息:

(i) 序列特征:

(A) 长度: 68 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO: 57:

CATACCAAAG CCGTCGCCTT GGGCACCGAA GAAACACAAC CACTAAGTTG TTGCACAGAG 60

ACTCAGTG 68

(2) SEQ ID NO: 58 信息:

(i) 序列特征:

(A) 长度: 77 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无



(iv)反义:否

(xi)序列描述:SEQ ID NO:58:

TAATGTAATT GATTGTAATG ACTCTATGTG CAGTACCAGT ACCGTATTCC AGCACCGTGT	60
CCGTGGGCAC CGCAAAG	77

(2)SEQ ID NO:59 信息:

(i)序列特征:

(A)长度:80 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:59:

ACAGACAACG ATAACCGACC ACCACAAGCA GCGGCCAAAC ACCCCGCCTT GGACAATAGA	60
ACAGCACGTA CTGCAACTAA	80

(2)SEQ ID NO:60 信息:

(i)序列特征:

(A)长度:266 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:60:

```

CATATGCAAT ACAATGCATT ATACAAACTG GACACATATA TATATTTGTG AAGAAGCATC      60
AGTAACTGTG GTAGAGGGTC AAGTTGACTA TTATGGTTTA TATTATGTTT ATGAAGGAAT      120
ACGAACATAT TTTGTGCAGT TTAAAGATGA TGCAGAAAAA TATAGTAAAA ATAAAGTATG      180
GGAAGTTCAT GCGGGTGGTC AGGTAATATT ATGTCCTACA TCTGTGTTTA GCAGCAACGA      240
AGTATCCTCT CCTGAAATTA TTAGGC                                           266

```

(2) SEQ ID NO:61 信息:

(i) 序列特征:

(A) 长度: 95 个碱基对

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑学: 线性

(ii) 分子类型: cDNA

(iii) 假拟: 无

(iv) 反义: 否

(xi) 序列描述: SEQ ID NO:61:

```

AGGATGTATA AAAAAACATG GATATACAGT GGAAGTGCAG TTTGATGGAG ACATATGCTA      60
TTAGGCAGCA CTTGGCCAAC CACCCCGCCG CGACC                                           95

```

## (2)SEQ ID NO:62 信息:

## (i)序列特征:

- (A)长度:81 个碱基对
- (B)类型:核酸
- (C)链型:双链
- (D)拓扑学:线性

## (ii)分子类型:cDNA

## (iii)假拟:无

## (iv)反义:否

## (xi)序列描述:SEQ ID NO:62:

```

CATGTTTTTT TATACATCCA TATCACCGCC AGTGGTATTT ATGTCAACAC CGCCAGAGAT      60
AATTTATCAC CGCAGATGGT T                                                    81

```

## (2)SEQ ID NO:63 信息:

## (i)序列特征:

- (A)长度:322 个氨基酸
- (B)类型:氨基酸
- (D)拓扑学:线性

## (ii)分子类型:肽

## (iii)假拟:无

## (iv)反义:否

## (v)片段类型:内部

## (xi) 序列描述: SEQ ID NO:63:

```

Met Ala Asp Asp Asp Pro Tyr Gly Thr Gly Gln Met Phe His Leu Asn
1           5           10           15

Thr Ala Leu Thr His Ser Ile Phe Asn Ala Glu Leu Tyr Ser Pro Glu
20           25           30

Ile Pro Leu Ser Thr Asp Gly Pro Tyr Leu Gln Ile Leu Glu Gln Pro
35           40           45

Lys Gln Arg Gly Phe Arg Phe Arg Tyr Val Cys Glu Gly Pro Ser His
50           55           60

Gly Gly Leu Pro Gly Ala Ser Ser Glu Lys Asn Lys Lys Ser Tyr Pro
65           70           75           80

Gln Val Lys Ile Cys Asn Tyr Val Gly Pro Ala Lys Val Ile Val Gln
85           90           95

Leu Val Thr Asn Gly Lys Asn Ile His Leu His Ala His Ser Leu Val
100          105          110

Gly Lys His Cys Glu Asp Gly Val Cys Thr Val Thr Ala Gly Pro Lys
115          120          125

Asp Met Val Val Gly Phe Ala Asn Leu Gly Ile Leu His Val Thr Lys
130          135          140

Lys Lys Val Phe Glu Thr Leu Glu Ala Arg Met Thr Glu Ala Cys Ile
145          150          155          160

Arg Gly Tyr Asn Pro Gly Leu Leu Val His Ser Asp Leu Ala Tyr Leu
165          170          175

Gln Ala Glu Gly Gly Gly Asp Arg Gln Leu Thr Asp Arg Glu Lys Glu
180          185          190

Ile Ile Arg Gln Ala Ala Val Gln Gln Thr Lys Glu Met Asp Leu Ser
195          200          205

Val Val Arg Leu Met Phe Thr Ala Phe Leu Pro Asp Ser Thr Gly Ser
210          215          220

Phe Thr Arg Arg Leu Glu Pro Val Val Ser Asp Ala Ile Tyr Asp Ser
225          230          235          240

Lys Ala Pro Asn Ala Ser Asn Leu Lys Ile Val Arg Met Asp Arg Thr
245          250          255

Ala Gly Cys Val Thr Gly Gly Glu Glu Ile Tyr Leu Leu Cys Asp Lys
260          265          270

Val Gln Lys Asp Asp Ile Gln Ile Arg Phe Tyr Glu Glu Glu Glu Asn
275          280          285

```

```

Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser Pro Thr Asp Val His
 290                               295                               300
Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys Tyr Lys Asp Val Asn
 305                               310                               315                               320
Ile Thr

```

## (2)SEQ ID NO:64 信息:

(i)序列特征:

(A)长度:325 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:64:

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

Glu Glu Asn Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser Pro Thr  
 290 295 300  
 Asp Val His Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys Tyr Lys  
 305 310 315 320  
 Asp Ile Asn Ile Thr  
 325

## (2)SEQ ID NO:65 信息:

(i) 序列特征:

(A) 长度:268 个氨基酸

(B) 类型:氨基酸

(D) 拓扑学:线性

(ii) 分子类型:肽

(iii) 假拟:无

(iv) 反义:否

(v) 片段类型:内部

(xi) 序列描述:SEQ ID NO:65:

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

```

Ser Leu Val Thr Lys Asp Pro Pro His Gly Pro His Pro His Glu Leu
      85                                90                                95
Val Gly Arg His Cys Gln His Gly Tyr Tyr Glu Ala Glu Leu Ser Pro
      100                                105                                110
Asp Arg Ser Ile His Ser Phe Gln Asn Leu Gly Ile Gln Cys Val Lys
      115                                120                                125
Lys Arg Glu Leu Glu Ala Ala Val Ala Glu Arg Ile Arg Thr Asn Asn
      130                                135                                140
Asn Pro Phe Asn Val Pro Met Glu Glu Arg Gly Ala Glu Tyr Asp Leu
      145                                150                                155                                160
Ser Ala Val Arg Leu Cys Phe Gln Val Trp Val Asn Gly Pro Gly Gly
      165                                170                                175
Leu Cys Pro Leu Pro Pro Val Leu Ser Gln Pro Ile Tyr Asp Asn Arg
      180                                185                                190
Ala Pro Ser Thr Ala Glu Leu Arg Ile Leu Pro Gly Asp Arg Asn Ser
      195                                200                                205
Gly Ser Cys Gln Gly Gly Asp Glu Ile Phe Leu Leu Cys Asp Lys Val
      210                                215                                220
Gln Lys Glu Asp Ile Glu Val Arg Phe Trp Ala Glu Gly Trp Glu Ala
      225                                230                                235                                240
Lys Gly Ser Phe Ala Ala Ala Asp Val His Arg Gln Val Ala Ile Val
      245                                250                                255
Phe Arg Thr Pro Pro Phe Arg Glu Arg Ser Leu Arg
      260                                265

```

(2)SEQ ID NO:66 信息:

(i)序列特征:

(A)长度:263 个氨基酸



(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:66:

```

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

```

---

Gly Asp Glu Ile Phe Leu Leu Cys Asp Lys Val Gln Lys Glu Asp Ile  
 210 215 220

Glu Val Tyr Phe Thr Gly Pro Gly Trp Glu Ala Arg Gly Ser Phe Ser  
 225 230 235 240

Gln Ala Asp Val His Arg Gln Val Ala Ile Val Phe Arg Thr Pro Pro  
 245 250 255

Tyr Ala Asp Pro Ser Leu Gln  
 260

(2)SEQ ID NO:67 信息:

(i)序列特征:

(A)长度:263 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:67:

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

## (2)SEQ ID NO:68 信息:

## (i)序列特征:

(A)长度:299个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:68:

```

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

```

Pro His Pro His Asn Leu Val Gly Lys Glu Gly Cys Lys Lys Gly Val  
 115 120 125  
 Cys Thr Leu Glu Ile Asn Ser Glu Thr Met Arg Ala Val Phe Ser Asn  
 130 135 140  
 Leu Gly Ile Gln Cys Val Lys Lys Lys Asp Ile Glu Ala Ala Leu Lys  
 145 150 155 160  
 Ala Arg Glu Glu Ile Arg Val Asp Pro Phe Lys Thr Gly Phe Ser His  
 165 170 175  
 Arg Phe Gln Pro Ser Ser Ile Asp Leu Asn Ser Val Arg Leu Cys Phe  
 180 185 190  
 Gln Val Phe Met Glu Ser Glu Gln Lys Gly Arg Phe Thr Ser Pro Leu  
 195 200 205  
 Pro Pro Val Val Ser Glu Pro Ile Phe Asp Lys Lys Ala Met Ser Asp  
 210 215 220  
 Leu Val Ile Cys Arg Leu Cys Ser Cys Ser Ala Thr Val Phe Gly Asn  
 225 230 235 240  
 Thr Gln Ile Ile Leu Leu Cys Glu Lys Val Ala Lys Glu Asp Ile Ser  
 245 250 255  
 Val Arg Phe Phe Glu Glu Lys Asn Gly Gln Ser Val Trp Glu Ala Phe  
 260 265 270  
 Gly Asp Phe Gln His Thr Asp Val His Lys Gln Thr Ala Ile Thr Phe  
 275 280 285  
 Lys Thr Pro Arg Tyr His Thr Leu Asp Ile Thr  
 290 295

(2)SEQ ID NO:69 信息:

(i)序列特征:

(A)长度:261个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:69:

```

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

```

```

His Asn Ile Asp Glu Tyr Asp Leu Asn Val Val Arg Leu Cys Phe Gln
145                150                155                160
Ala Phe Leu Pro Asp Glu His Gly Asn Tyr Thr Leu Ala Leu Pro Pro
                165                170                175
Leu Ile Ser Asn Pro Ile Tyr Asp Asn Arg Ala Pro Asn Thr Ala Glu
                180                185                190
Leu Arg Ile Cys Arg Val Asn Lys Asn Cys Gly Ser Val Lys Gly Gly
                195                200                205
Asp Glu Ile Phe Leu Leu Cys Asp Lys Val Gln Lys Asp Asp Ile Glu
                210                215                220
Val Arg Phe Val Leu Gly Asn Trp Glu Ala Lys Gly Ser Phe Ser Gln
225                230                235                240
Ala Asp Val His Arg Gln Val Ala Ile Val Phe Arg Thr Pro Pro Phe
                245                250                255
Leu Gly Asp Ile Thr
                260

```

## (2)SEQ ID NO:70 信息:

(i) 序列特征:

(A) 长度: 262 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO:70:

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



## (2)SEQ ID NO:71 信息:

## (i)序列特征:

(A)长度:314 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

## (ii)分子类型:肽

## (iii)假拟:无

## (iv)反义:否

## (v)片段类型:内部

## (xi)序列描述:SEQ ID NO:71:

Met Ser Asn Lys Lys Gln Ser Asn Arg Leu Thr Glu Gln His Lys Leu  
1                   5                   10                   15

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



## (xi) 序列描述: SEQ ID NO:73:

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

## (2) SEQ ID NO:74 信息:

(i) 序列特征:

(A) 长度: 181 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:74:

```

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

```

## (2)SEQ ID NO:75 信息:

(i)序列特征:

(A)长度:85个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:75:

```

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

```



(A)长度:84 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:77:

```

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

```

(2)SEQ ID NO:78 信息:

(i)序列特征:

(A)长度:84 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽



(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO: 78:

```

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

```

(2) SEQ ID NO: 79 信息:

(i) 序列特征:

(A) 长度: 84 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

## (xi) 序列描述:SEQ ID NO:79:

```

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

```

## (2)SEQ ID NO:80 信息:

(i) 序列特征:

(A) 长度:84 个氨基酸

(B) 类型:氨基酸

(D) 拓扑学:线性

(ii) 分子类型:肽

(iii) 假拟:无

(iv) 反义:否

(v) 片段类型:内部

## (xi) 序列描述:SEQ ID NO:80:

```

Thr Pro Ile Val Gln Leu Gln Gly Asp Ser Asn Cys Leu Lys Cys Phe
1          5          10          15
Arg Tyr Arg Leu Asn Asp Lys Tyr Lys His Leu Phe Glu Leu Ala Ser
20          25          30

```

```

Ser Thr Trp His Trp Ala Ser Pro Glu Ala Pro His Lys Asn Ala Ile
   35                               40                               45
Val Thr Leu Thr Tyr Ser Ser Glu Glu Gln Arg Gln Gln Phe Leu Asn
   50                               55                               60
Ser Val Lys Ile Pro Pro Thr Ile Arg His Lys Val Gly Phe Met Ser
   65                               70                               75                               80
Leu His Leu Leu

```

## (2)SEQ ID NO:81 信息:

(i)序列特征:

(A)长度:84 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:81:

```

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

```

## (2)SEQ ID NO:82 信息:

## (i)序列特征:

- (A)长度:80 个氨基酸  
 (B)类型:氨基酸  
 (D)拓扑学:线性

## (ii)分子类型:肽

## (iii)假拟:无

## (iv)反义:否

## (v)片段类型:内部

## (xi)序列描述:SEQ ID NO:82:

```

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

```

## (2)SEQ ID NO:83 信息:

## (i)序列特征:

(A)长度:80 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:83:

```

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

```

(2)SEQ ID NO:84 信息:

(i)序列特征:

(A)长度:81 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:84:

```

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

```

(2)SEQ ID NO:85 信息:

(i)序列特征:

(A)长度:21个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部



(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:87:

```
Gly Ile Val Glu Gln Cys Cys Ala Ser Val Cys Ser Leu Tyr Gln Leu
1           5           10           15
Glu Asn Tyr Cys Asn
                20
```

(2)SEQ ID NO:88 信息:

(i)序列特征:

(A)长度:30 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:88:

```
Phe Val Asn Gln His Leu Cys Gly Ser His Leu Val Glu Ala Leu Tyr
1           5           10           15
```



Leu Val Cys Gly Glu Arg Gly Phe Phe Tyr Thr Pro Lys Thr  
 20 25 30

## (2)SEQ ID NO:89 信息:

(i)序列特征:

(A)长度:24 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:89:

Gln Leu Tyr Ser Ala Leu Ala Asn Lys Cys Cys His Val Gly Cys Ile  
 1 5 10 15  
 Lys Arg Ser Leu Ala Arg Phe Cys  
 20

## (2)SEQ ID NO:90 信息:

(i)序列特征:

(A)长度:33 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO:90:

Asp Ser Trp Met Glu Glu Val Ile Lys Ile Cys Gly Arg Glu Leu Val  
1                   5                   10                   15

Arg Ala Gln Ile Ala Ile Cys Gly Met Ser Thr Trp Ser Lys Arg Ser  
          20                   25                   30

Leu

(2) SEQ ID NO:91 信息:

(i) 序列特征:

(A) 长度: 24 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO:91:

Glu Glu Lys Met Gly Thr Ala Lys Lys Cys Cys Ala Ile Gly Cys Ser  
1                   5                   10                   15

Thr Glu Asp Phe Arg Met Val Cys  
          20

## (2)SEQ ID NO:92 信息:

## (i)序列特征:

(A)长度:40 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

## (ii)分子类型:肽

## (iii)假拟:无

## (iv)反义:否

## (v)片段类型:内部

## (xi)序列描述:SEQ ID NO:92:

```

Arg Pro Asn Trp Glu Glu Arg Ser Arg Leu Cys Gly Arg Asp Leu Ile
1           5           10           15
Arg Ala Phe Ile Tyr Leu Cys Gly Gly Thr Arg Trp Thr Arg Leu Pro
                20           25           30
Asn Phe Gly Asn Tyr Pro Ile Met
          35           40

```

## (2)SEQ ID NO:93 信息:

## (i)序列特征:

(A)长度:182 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

## (ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:93:

```

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

```

## (2)SEQ ID NO:94 信息:

## (i)序列特征:

(A)长度:84个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

## (ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

## (xi)序列描述:SEQ ID NO:94:

```

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

```

## (2)SEQ ID NO:95 信息:

## (i)序列特征:

(A)长度:84个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:95:

```

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

```

(2)SEQ ID NO:96 信息:

(i)序列特征:

(A)长度:83个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:96:

```

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

```

(2)SEQ ID NO:97 信息:

(i)序列特征:

(A)长度:84个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:97:

```

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

```

## (2) SEQ ID NO:98 信息:

(i) 序列特征:

(A) 长度:89 个氨基酸

(B) 类型:氨基酸

(D) 拓扑学:线性

(ii) 分子类型:肽

(iii) 假拟:无

(iv) 反义:否

(v) 片段类型:内部

(xi) 序列描述:SEQ ID NO:98:

```

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

```



Arg Gln Gly Gln Ala Gln Ile Leu Ile Thr Phe Gly Ser Pro Ser Gln  
50 55 60

Arg Gln Asp Phe Leu Lys His Val Pro Leu Pro Pro Gly Met Asn Ile  
65 70 75 80

Ser Gly Phe Thr Ala Ser Leu Asp Phe  
85

(2)SEQ ID NO:99 信息:

(i)序列特征:

(A)长度:7个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:C-末端

(xi)序列描述:SEQ ID NO:99:

Ser Asn Lys Lys Thr Thr Ala  
1 5

(2)SEQ ID NO:100 信息:

(i)序列特征:

(A)长度:4个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:100:

Asn Ser Asn Thr  
1

(2)SEQ ID NO:101 信息:

(i)序列特征:

(A)长度:4个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:内部

(xi)序列描述:SEQ ID NO:101:

Ser Gly Asn Thr  
1

(2)SEQ ID NO:102 信息:

(i) 序列特征:

(A) 长度: 6 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO: 102:

Ser Ser Gly Ser Ser Gly  
1 5

(2) SEQ ID NO: 103 信息:

(i) 序列特征:

(A) 长度: 15 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑学: 线性

(ii) 分子类型: 肽

(iii) 假拟: 无

(iv) 反义: 否

(v) 片段类型: 内部

(xi) 序列描述: SEQ ID NO: 103:

Cys Tyr Pro Glu Ile Lys Asp Lys Glu Glu Val Gln Arg Lys Arg  
 1                    5                    10                    15

## (2)SEQ ID NO:104 信息:

(i)序列特征:

(A)长度:66 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:104:

Met Glu Gln Arg Ile Thr Leu Lys Asp Tyr Ala Met Arg Phe Gly Gln  
 1                    5                    10                    15

Thr Lys Thr Ala Lys Asp Leu Gly Val Tyr Gln Ser Ala Ile Asn Lys  
 20                    25                    30

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

Ser Val Tyr Ala Glu Glu Val Lys Pro Phe Pro Ser Asn Lys Lys Thr  
 50                    55                    60

Thr Ala  
 65

## (2)SEQ ID NO:105 信息:

(i)序列特征:

(A)长度:66 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:105:

```

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

```

(2)SEQ ID NO:106 信息:

(i)序列特征:

(A)长度:66 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:106:

```

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

```

(2)SEQ ID NO:107 信息:

(i)序列特征:

(A)长度:96个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:肽

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

## (xi) 序列描述:SEQ ID NO:107:

```

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

```

## (2)SEQ ID NO:108 信息:

## (i) 序列特征:

(A) 长度:96 个氨基酸

(B) 类型:氨基酸

(D) 拓扑学:线性

## (ii) 分子类型:肽

(iii) 假拟:无

(iv) 反义:否

(v) 片段类型:N-末端

## (xi) 序列描述:SEQ ID NO:108:

```

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

```

## (2)SEQ ID NO:109 信息:

(i)序列特征:

(A)长度:180 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:109:



```

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

```

## (2)SEQ ID NO:110 信息:

(i)序列特征:

(A)长度:113 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:110:

```

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

```

(2)SEQ ID NO:111 信息:

(i)序列特征:

(A)长度:292 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:111:

```

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

```

Ile Phe Val Ser Gly Lys Val Val Leu Thr Gly Ala Lys Val Arg Ala  
260 265 270

Glu Ile Tyr Glu Ala Phe Glu Asn Ile Tyr Pro Ile Leu Lys Gly Phe  
275 280 285

Arg Lys Thr Thr  
290

(2)SEQ ID NO:112 信息:

(i)序列特征:

(A)长度:273 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:112:

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



His Leu Asn Thr Ala Leu Thr His Ser Ile Phe Asn Ala Glu Leu Tyr  
 115 120 125

Ser Pro Glu Ile Pro Leu Ser Thr Asp Gly Pro Tyr Leu Gln Ile Leu  
 130 135 140

Glu Gln Pro Lys Gln Arg Gly Phe Arg Phe Arg Tyr Val Cys Glu Gly  
 145 150 155 160

Pro Ser His Gly Gly Leu Pro Gly Ala Ser Ser Glu Lys Asn Lys Lys  
 165 170 175

Ser Tyr Pro Gln Val Lys Ile Cys Asn Tyr Val Gly Pro Ala Lys Val  
 180 185 190

Ile Val Gln Leu Val Thr Asn Gly Lys Asn Ile His Leu His Ala His  
 195 200 205

Ser Leu Val Gly Lys His Cys Glu Asp Gly Val Cys Thr Val Thr Ala  
 210 215 220

Gly Pro Lys Asp Met Val Val Gly Phe Ala Asn Leu Gly Ile Leu His  
 225 230 235 240

Val Thr Lys Lys Lys Val Phe Glu Thr Leu Glu Ala Arg Met Thr Glu  
 245 250 255

Ala Cys Ile Arg Gly Tyr Asn Pro Gly Leu Leu Val His Ser Asp Leu  
 260 265 270

Ala Tyr Leu Gln Ala Glu Gly Gly Gly Asp Arg Gln Leu Thr Asp Arg  
 275 280 285

Glu Lys Glu Ile Ile Arg Gln Ala Ala Val Gln Gln Thr Lys Glu Met  
 290 295 300

Asp Leu Ser Val Val Arg Leu Met Phe Thr Ala Phe Leu Pro Asp Ser  
 305 310 315 320

Thr Gly Ser Phe Thr Arg Arg Leu Glu Pro Val Val Ser Asp Ala Ile  
 325 330 335

Tyr Asp Ser Lys Ala Pro Asn Ala Ser Asn Leu Lys Ile Val Arg Met  
 340 345 350

Asp Arg Thr Ala Gly Cys Val Thr Gly Gly Glu Glu Ile Tyr Leu Leu  
 355 360 365

Cys Asp Lys Val Gln Lys Asp Asp Ile Gln Ile Arg Phe Tyr Glu Glu  
 370 375 380

Glu Glu Asn Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser Pro Thr  
 385 390 395 400

Asp Val His Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys Tyr Lys  
 405 410 415

Asp Val Asn Ile Thr  
420

(2)SEQ ID NO:114 信息:

(i)序列特征:

(A)长度:391 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:114:



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

```

Asp Ser Thr Gly Ser Phe Thr Arg Arg Leu Glu Pro Val Val Ser Asp
290                               295                               300

Ala Ile Tyr Asp Ser Lys Ala Pro Asn Ala Ser Asn Leu Lys Ile Val
305                               310                               315                               320

Arg Met Asp Arg Thr Ala Gly Cys Val Thr Gly Gly Glu Glu Ile Tyr
325                               330                               335

Leu Leu Cys Asp Lys Val Gln Lys Asp Asp Ile Gln Ile Arg Phe Tyr
340                               345                               350

Glu Glu Glu Glu Asn Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser
355                               360                               365

Pro Thr Asp Val His Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys
370                               375                               380

Tyr Lys Asp Ile Asn Ile Thr
385                               390

```

(2)SEQ ID NO:115 信息:

(i)序列特征:

(A)长度:391 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:115:

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

```

Asp Ser Thr Gly Ser Phe Thr Arg Arg Leu Glu Pro Val Val Ser Asp
 290                               295                               300

Ala Ile Tyr Asp Ser Lys Ala Pro Asn Ala Ser Asn Leu Lys Ile Val
305                               310                               315                               320

Arg Met Asp Arg Thr Ala Gly Cys Val Thr Gly Gly Glu Glu Ile Tyr
                               325                               330                               335

Leu Leu Cys Asp Lys Val Gln Lys Asp Asp Ile Gln Ile Arg Phe Tyr
                               340                               345                               350

Glu Glu Glu Glu Asn Gly Gly Val Trp Glu Gly Phe Gly Asp Phe Ser
                               355                               360                               365

Pro Thr Asp Val His Arg Gln Phe Ala Ile Val Phe Lys Thr Pro Lys
 370                               375                               380

Tyr Lys Asp Ile Asn Ile Thr
385                               390

```

## (2)SEQ ID NO:116 信息:

## (i)序列特征:

(A)长度:241个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

## (ii)分子类型:蛋白质

## (iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:116:

Met Arg Gln Arg Ile Thr Leu Lys Asp Tyr Ala Met Arg Phe Gly Gln  
 1 5 10 15

Thr Lys Thr Ala Lys Asp Leu Gly Val Tyr Gln Ser Ala Ile Asn Lys  
 20 25 30

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

Ser Val Tyr Ala Glu Glu Val Lys Pro Phe Pro Ser Asn Lys Lys Thr  
 50 55 60

Thr Ala Ser Asn Lys Lys Thr Thr Ala Gly Asp Pro Gly Lys Lys Lys  
 65 70 75 80

Gln His Ile Cys His Ile Gln Gly Cys Gly Lys Val Tyr Gly Lys Thr  
 85 90 95

Ser His Leu Arg Ala His Leu Arg Trp His Thr Gly Glu Arg Pro Phe  
 100 105 110

Met Cys Thr Trp Ser Tyr Cys Gly Lys Arg Phe Thr Arg Ser Asp Glu  
 115 120 125

Leu Gln Arg His Lys Arg Thr His Thr Gly Glu Lys Lys Phe Ala Cys  
 130 135 140

Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp His Leu Ser Lys His  
 145 150 155 160

Ile Lys Thr His Gln Asn Lys Lys Gly Gly Pro Gly Val Ala Leu Ser  
 165 170 175

Val Gly Thr Leu Pro Leu Asp Ser Gly Ala Gly Ser Glu Gly Ser Gly  
 180 185 190

Thr Ala Thr Pro Ser Ala Leu Ile Thr Thr Asn Met Val Ala Met Glu  
 195 200 205

Ala Ile Cys Pro Glu Gly Ile Ala Arg Leu Ala Asn Ser Gly Ile Asn  
 210 215 220

Val Met Gln Val Ala Asp Leu Gln Ser Ile Asn Ile Ser Gly Asn Gly  
 225 230 235 240

Phe

(2)SEQ ID NO:117 信息:

(i)序列特征:

(A)长度:10 个碱基对

(B)类型:核酸

(C)链型:双链

(D)拓扑学:线性

(ii)分子类型:cDNA

(iii)假拟:无

(iv)反义:否

(xi)序列描述:SEQ ID NO:117:

(2)SEQ ID NO:118 信息:

(i)序列特征:

(A)长度:72 个氨基酸

(B)类型:氨基酸

(D)拓扑学:线性

(ii)分子类型:蛋白质

(iii)假拟:无

(iv)反义:否

(v)片段类型:N-末端

(xi)序列描述:SEQ ID NO:118:

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

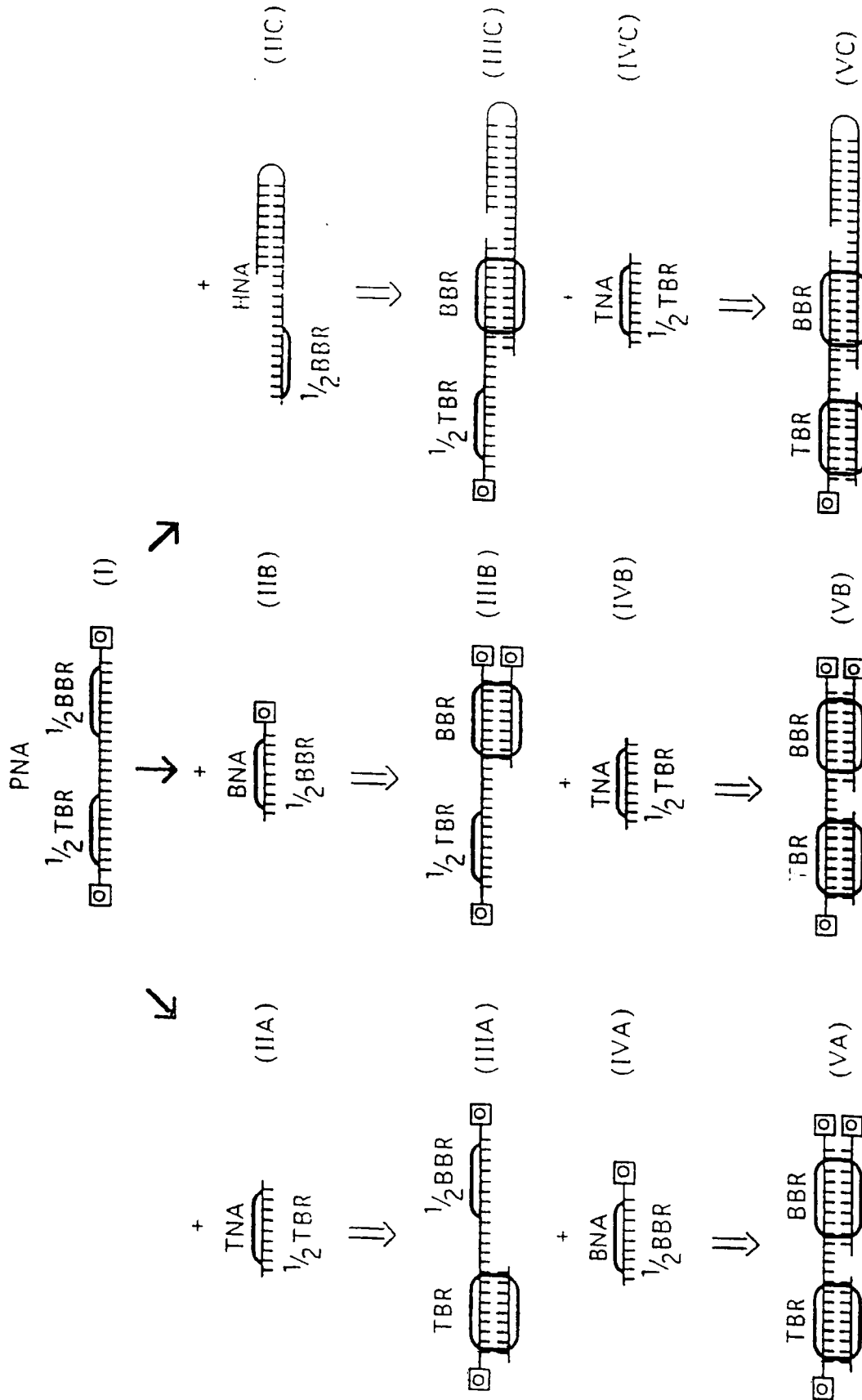


图 1



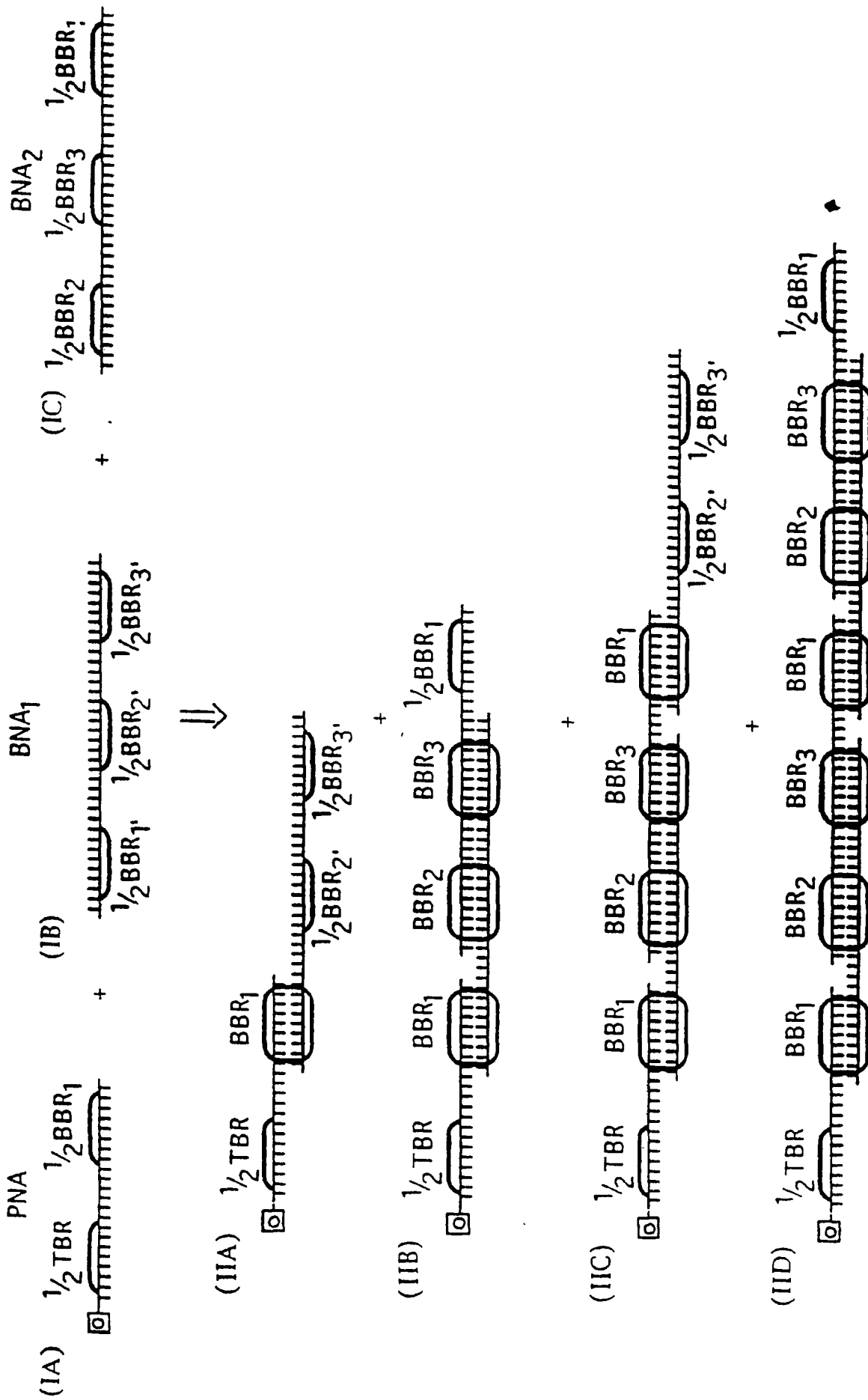


图 2A

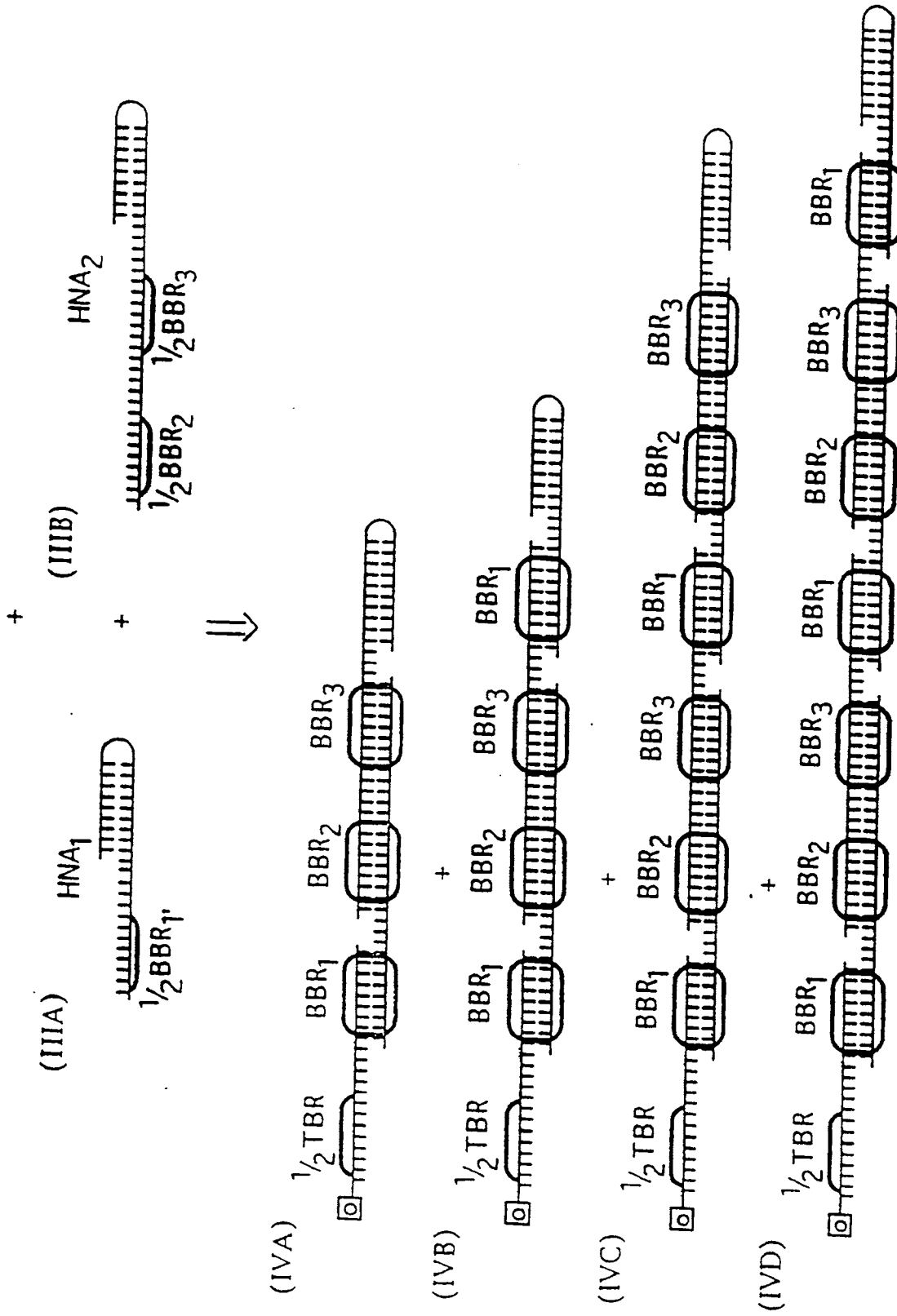


图 2B

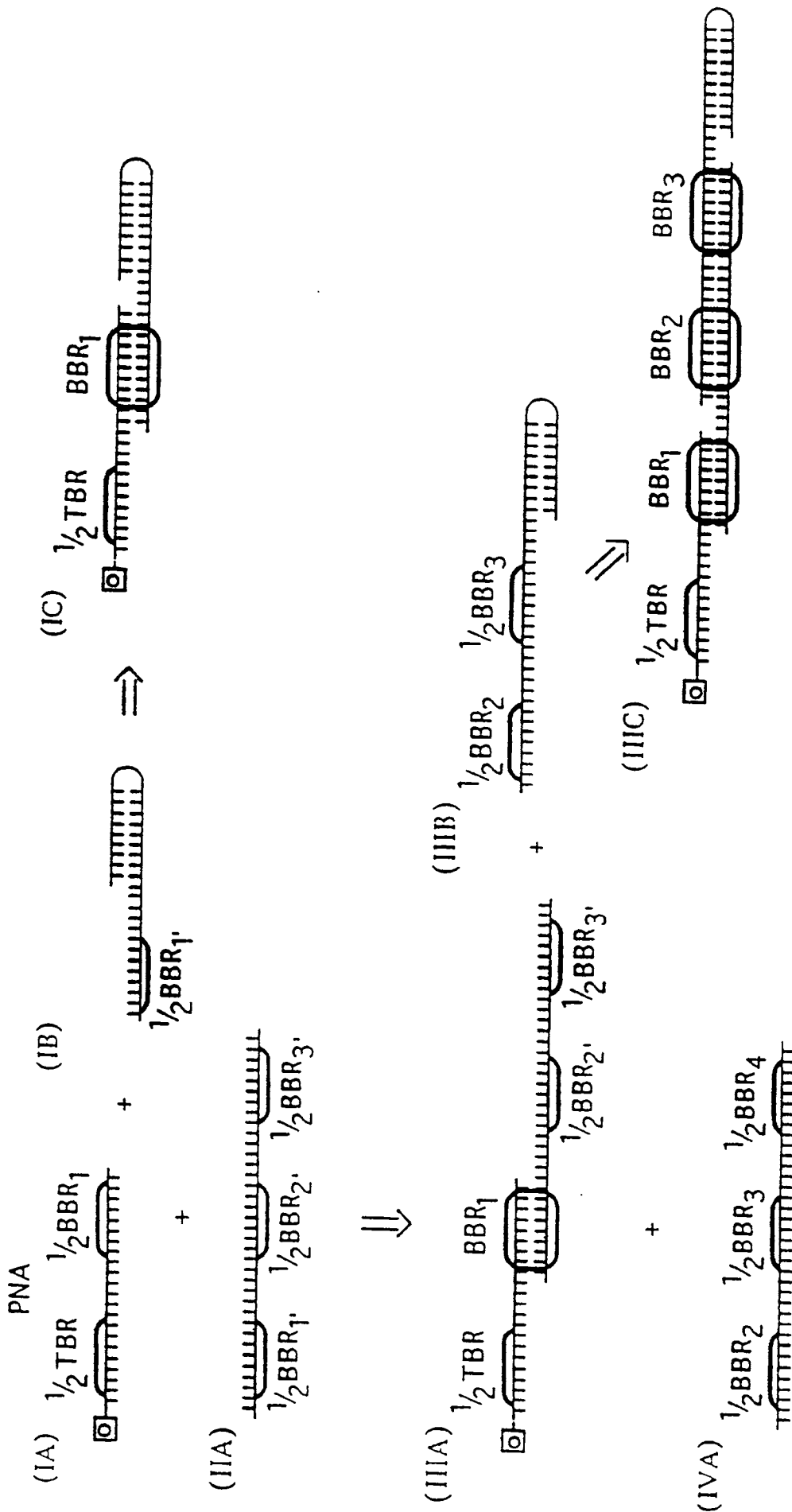


图 2C

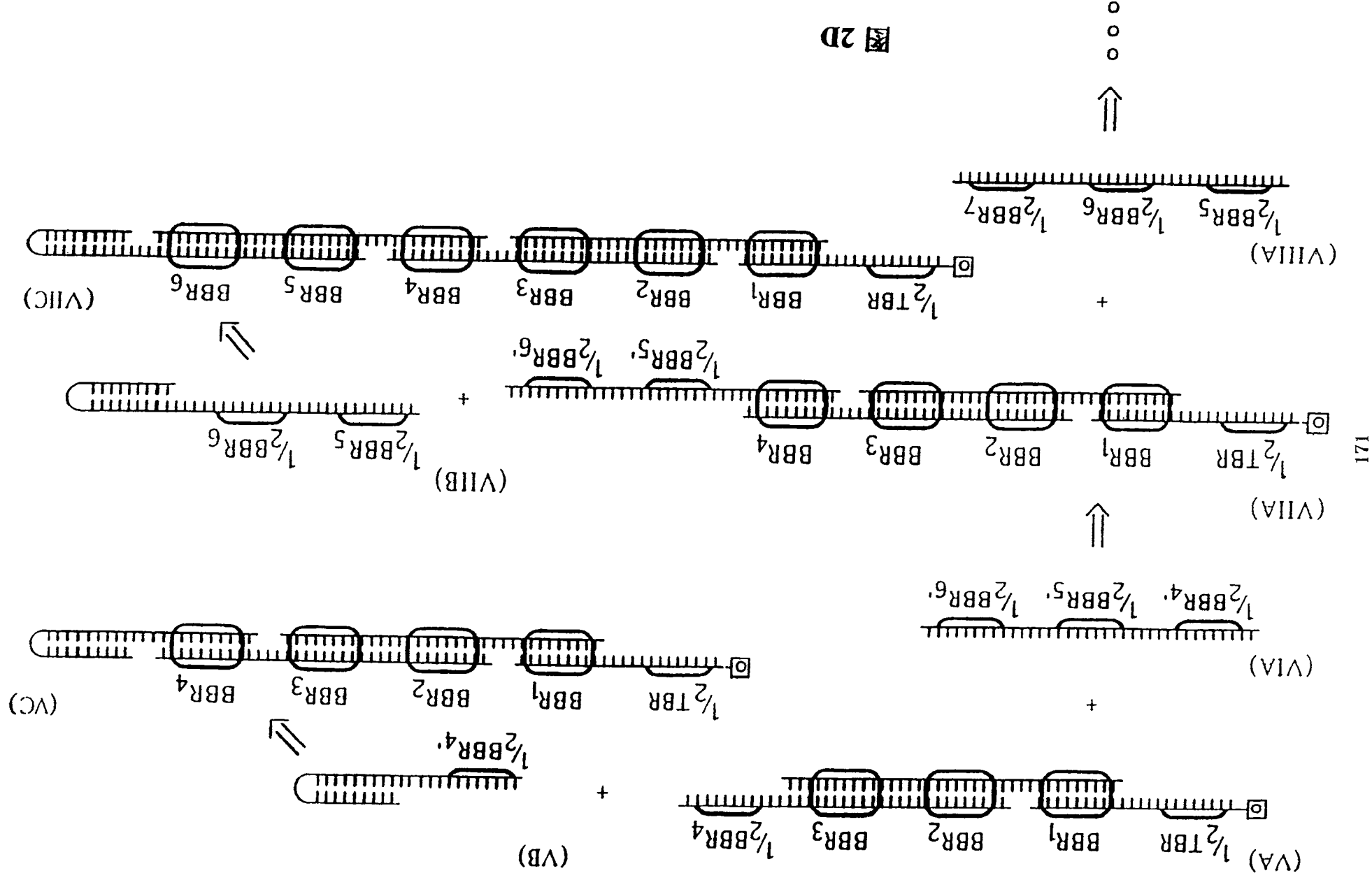


图 2D

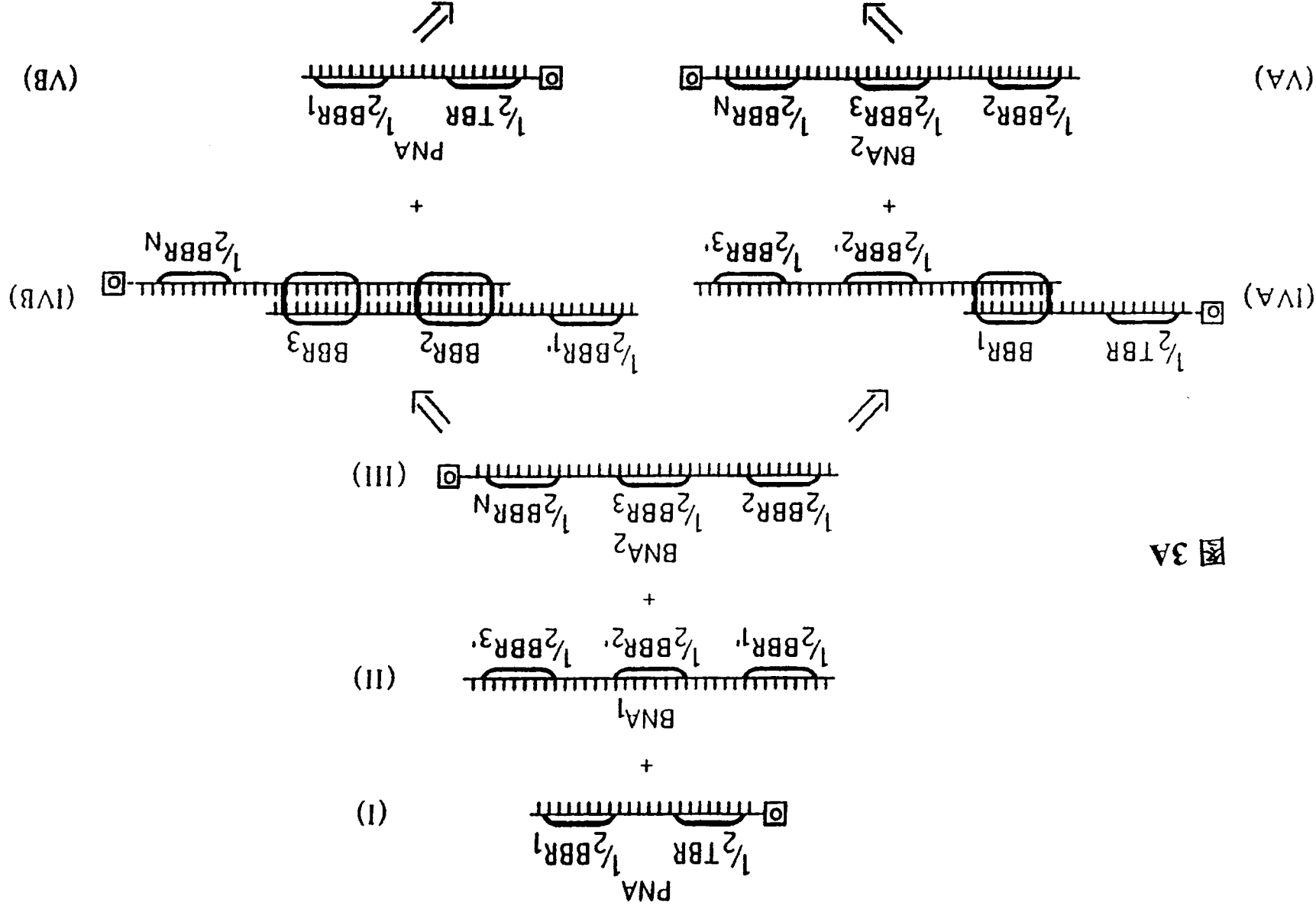
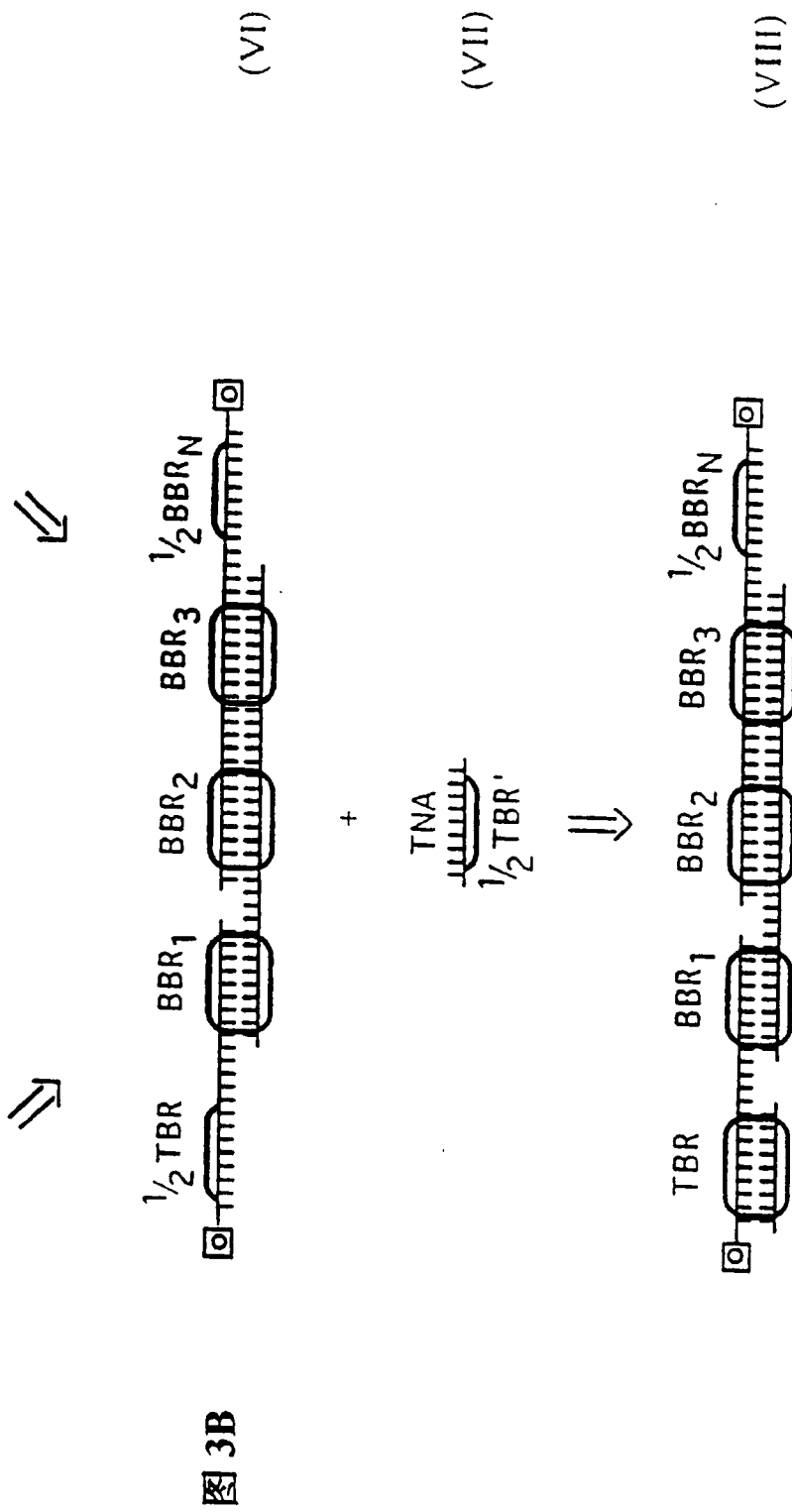


图 3A



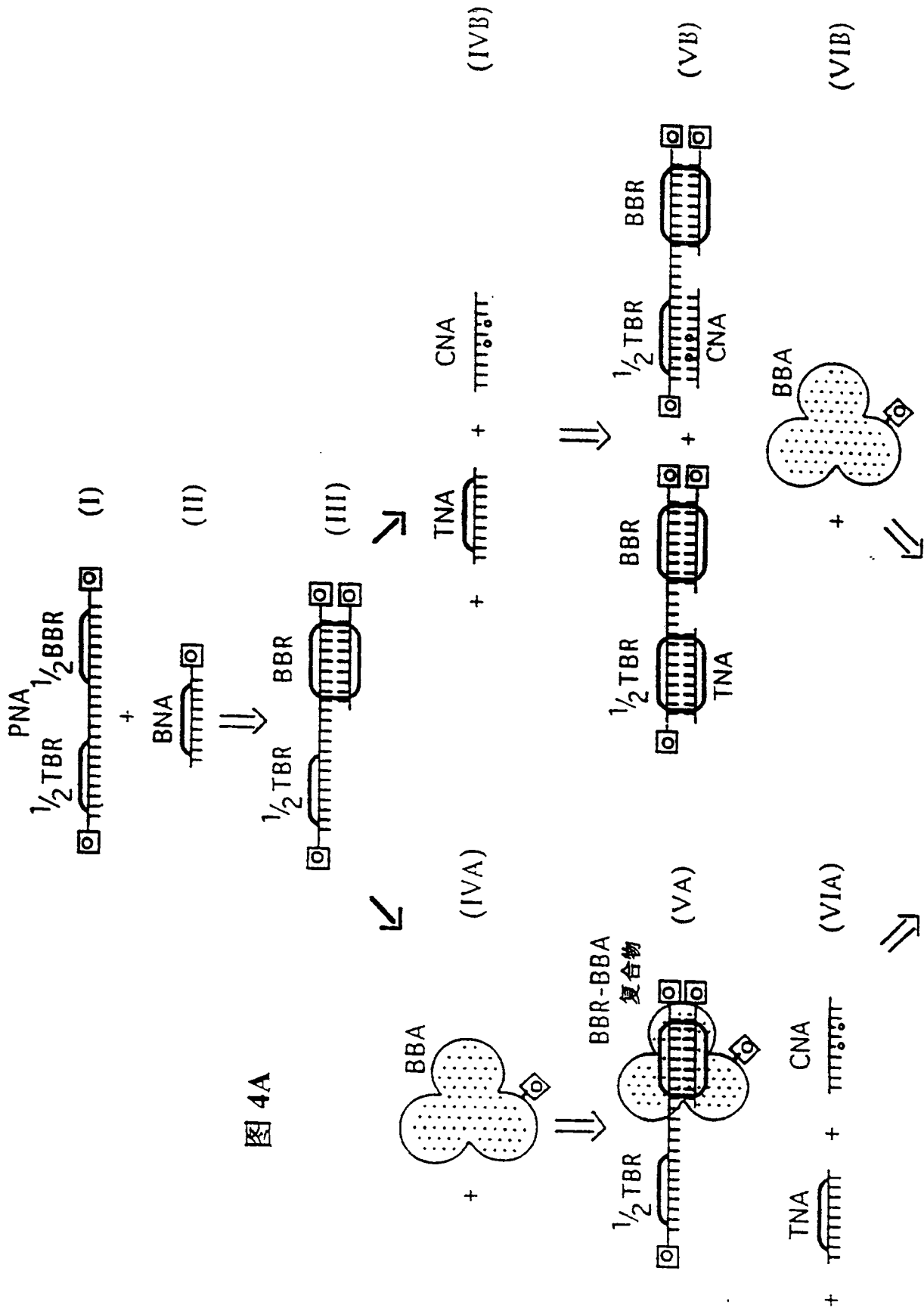


图 4A

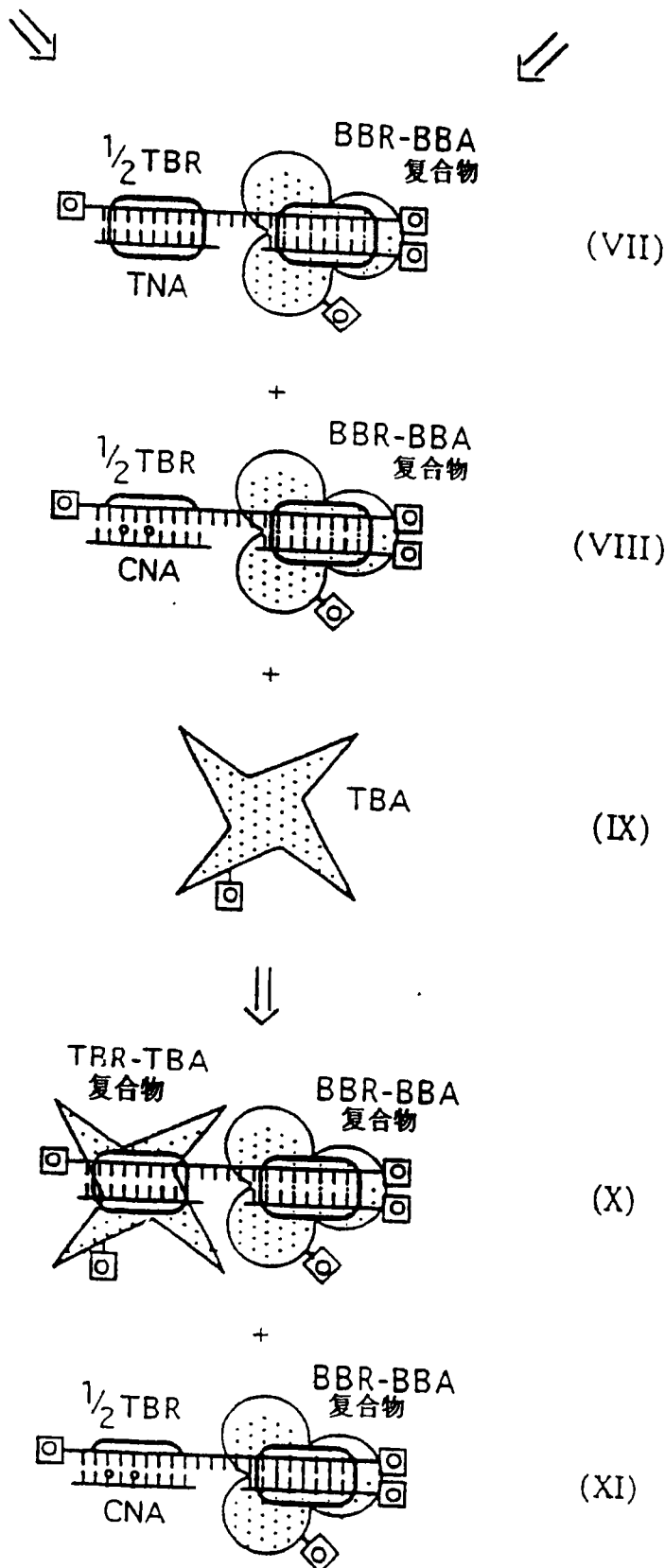


图 4B



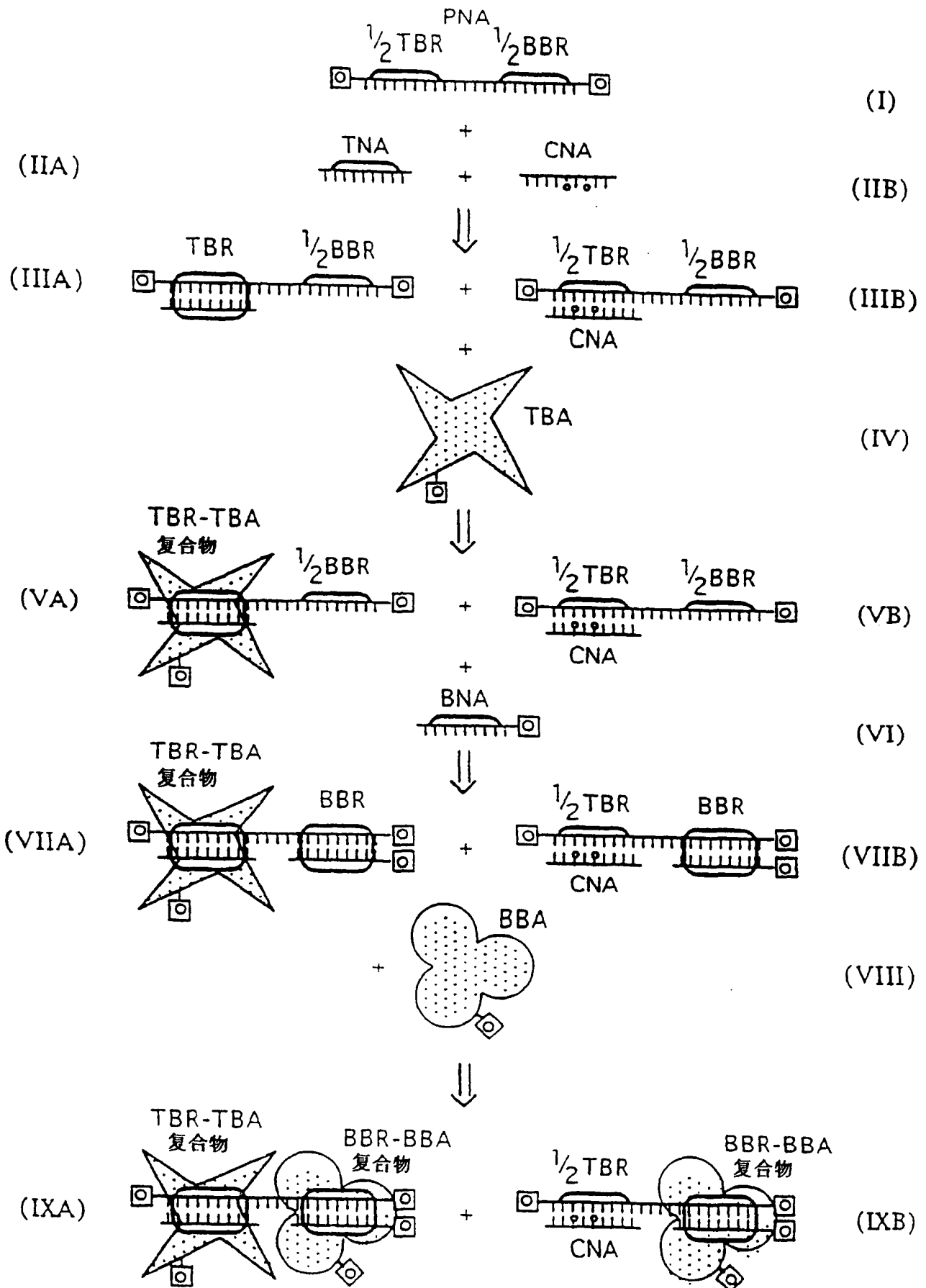


图 4C

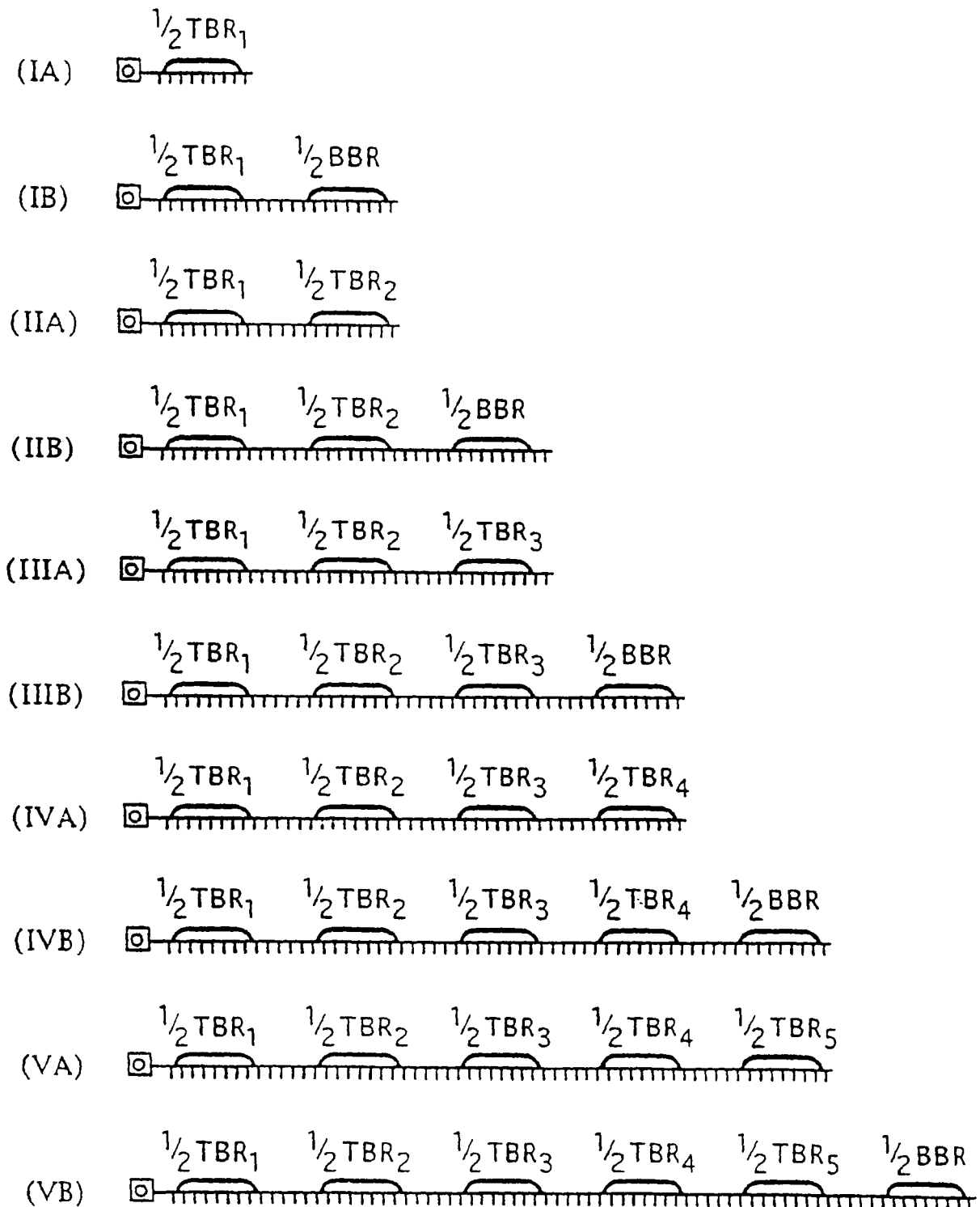
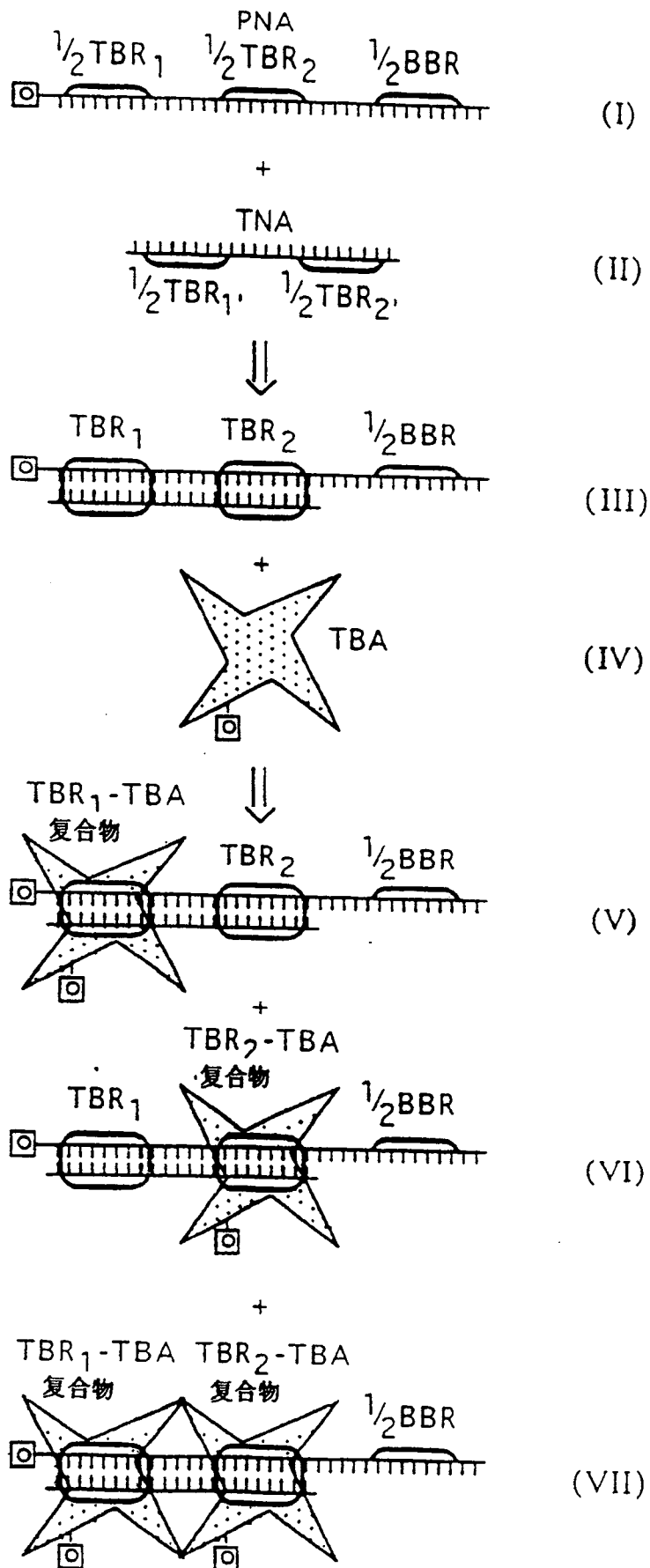


图 5

图 6A



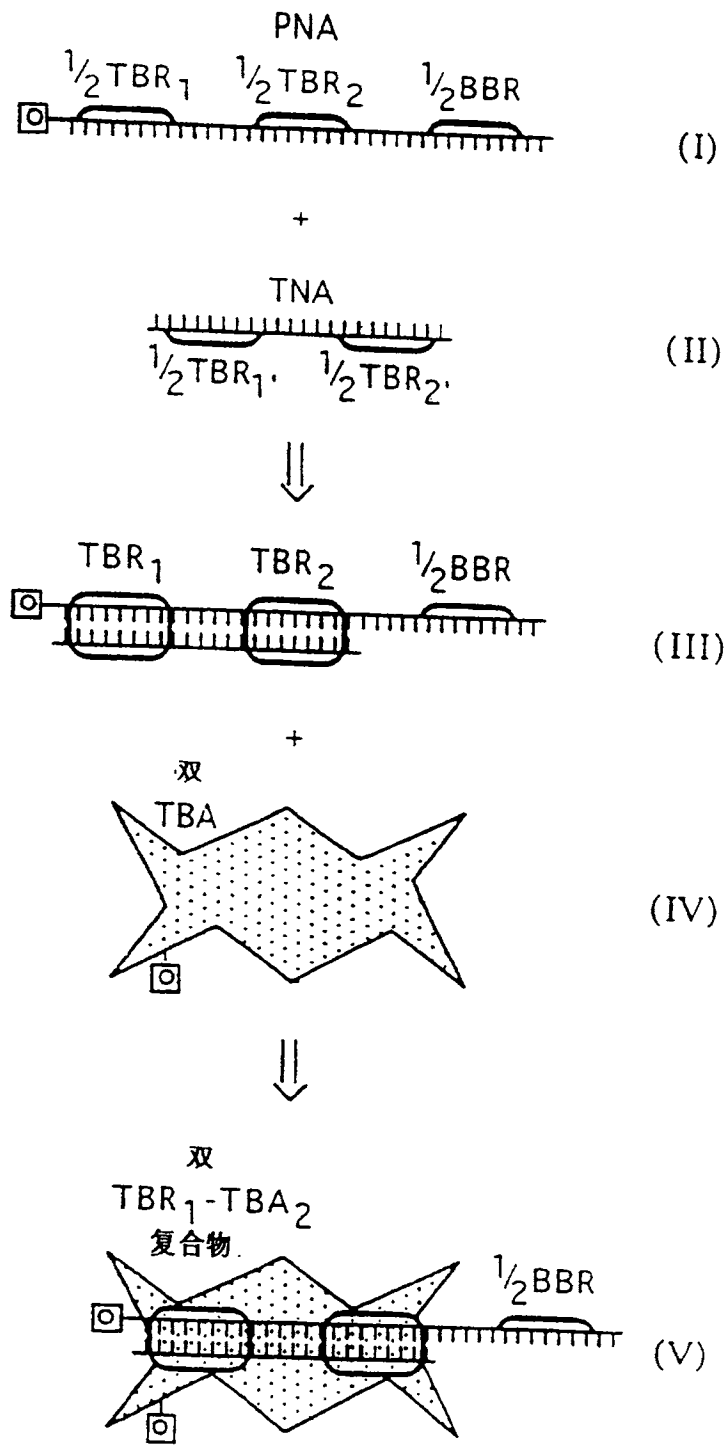


图 6B

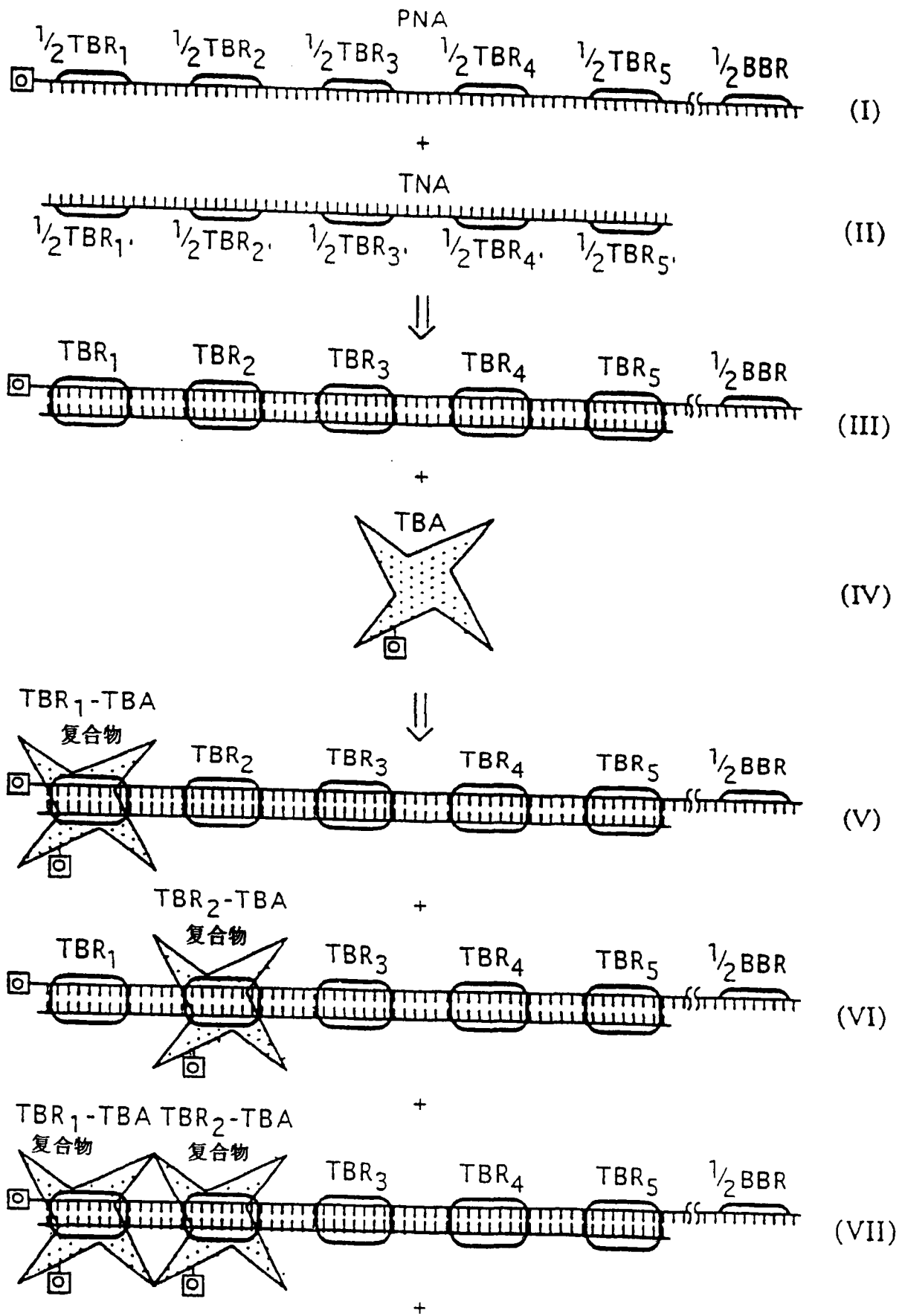


图 6C

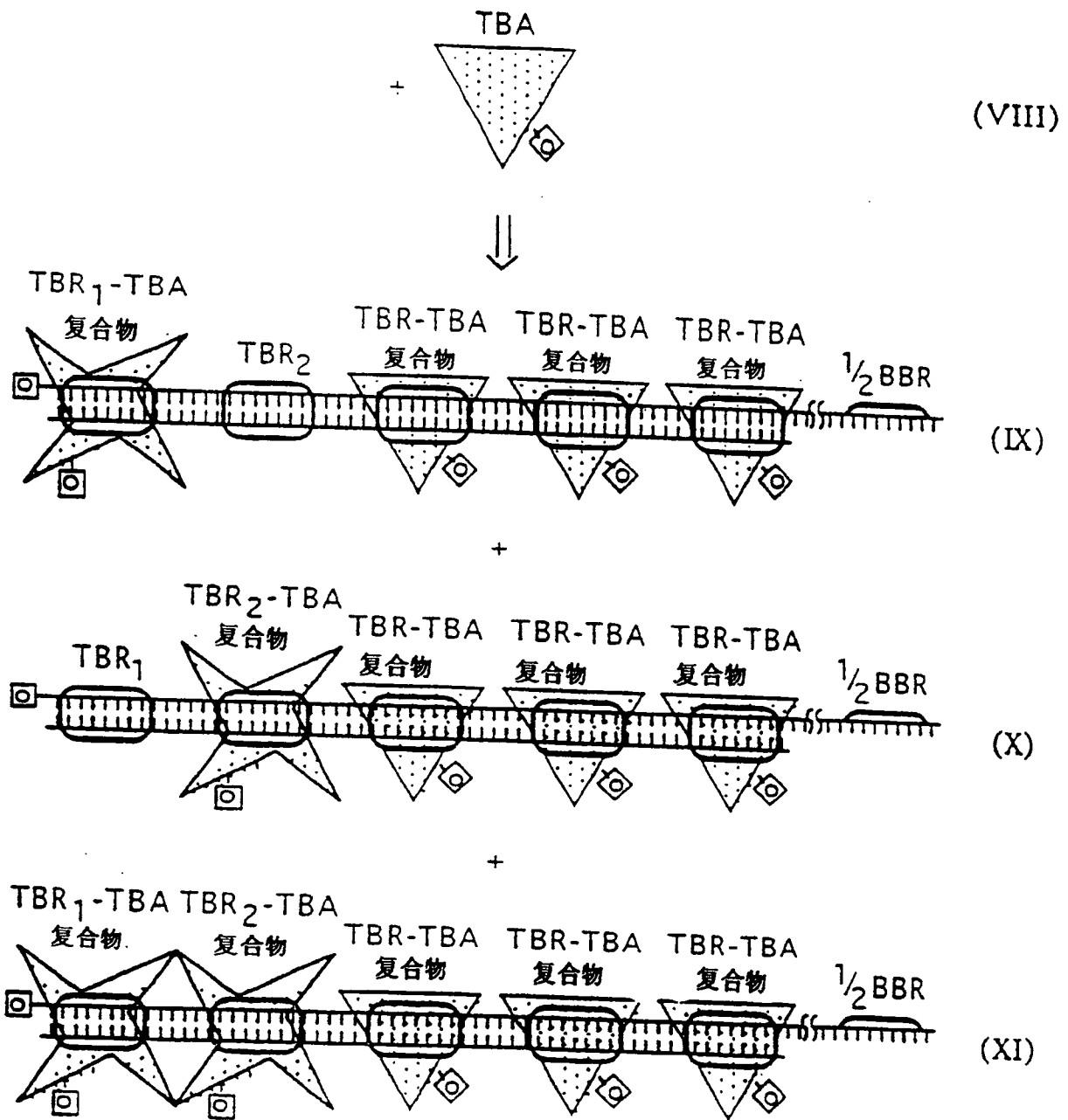


图 6D

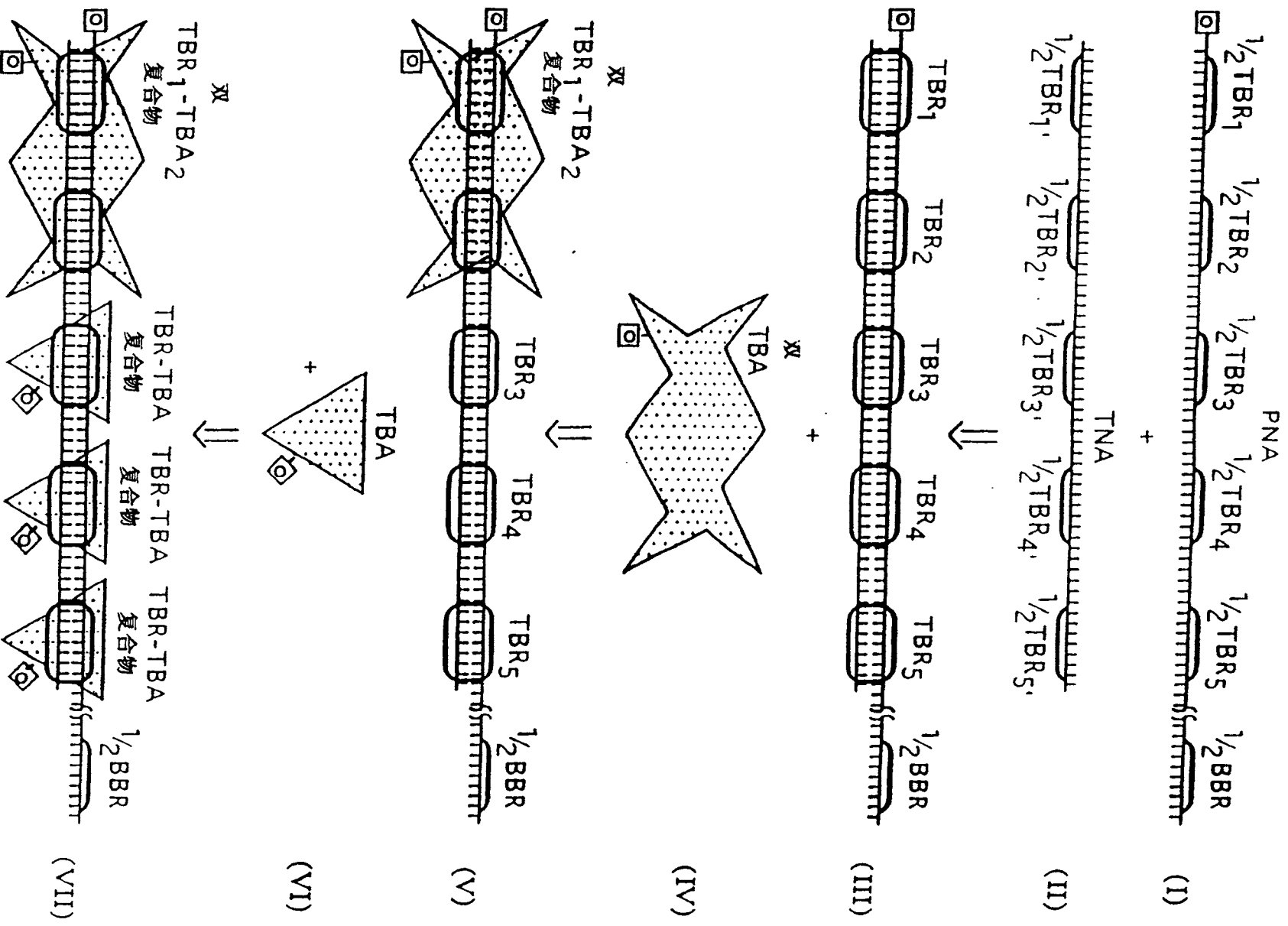


图 6E

SEQ. ID: 37:  
 123456789012345678901234567890123456789012345678901234567890  
 CTACAAGGACITTCGGCTGGGACITTCAGGGAGGGGTTGGCCCTGGCGGACTGGGGAGTGGCGTCCC  
 ++++++NF-kB NF-kB SP1 SP1 SP1  
 =====

HIV 试验药盒 PNA1 ( +++ 来自以上 ), SEQ. ID: 38;

\*\*\*CTACAAGGACTTCCGCTGGGACTTCCAGGGAGG\*\*\*

HIV 试验药盒 PNA2 ( === 来自以上 ), SEQ. ID: 39;

###CGGGACTGGGAGTGGCGTCCC###

PNA2 中的粘端序列与由下列形成的操纵基因 DNA 的末端之一互补:

###OL1-OL2-OL3  
 OL1'-OL2'-OL3'\*\*\*

或

###OR3-OR2-OR1  
 OR3'-OR2'-OR1'\*\*\*

图 7



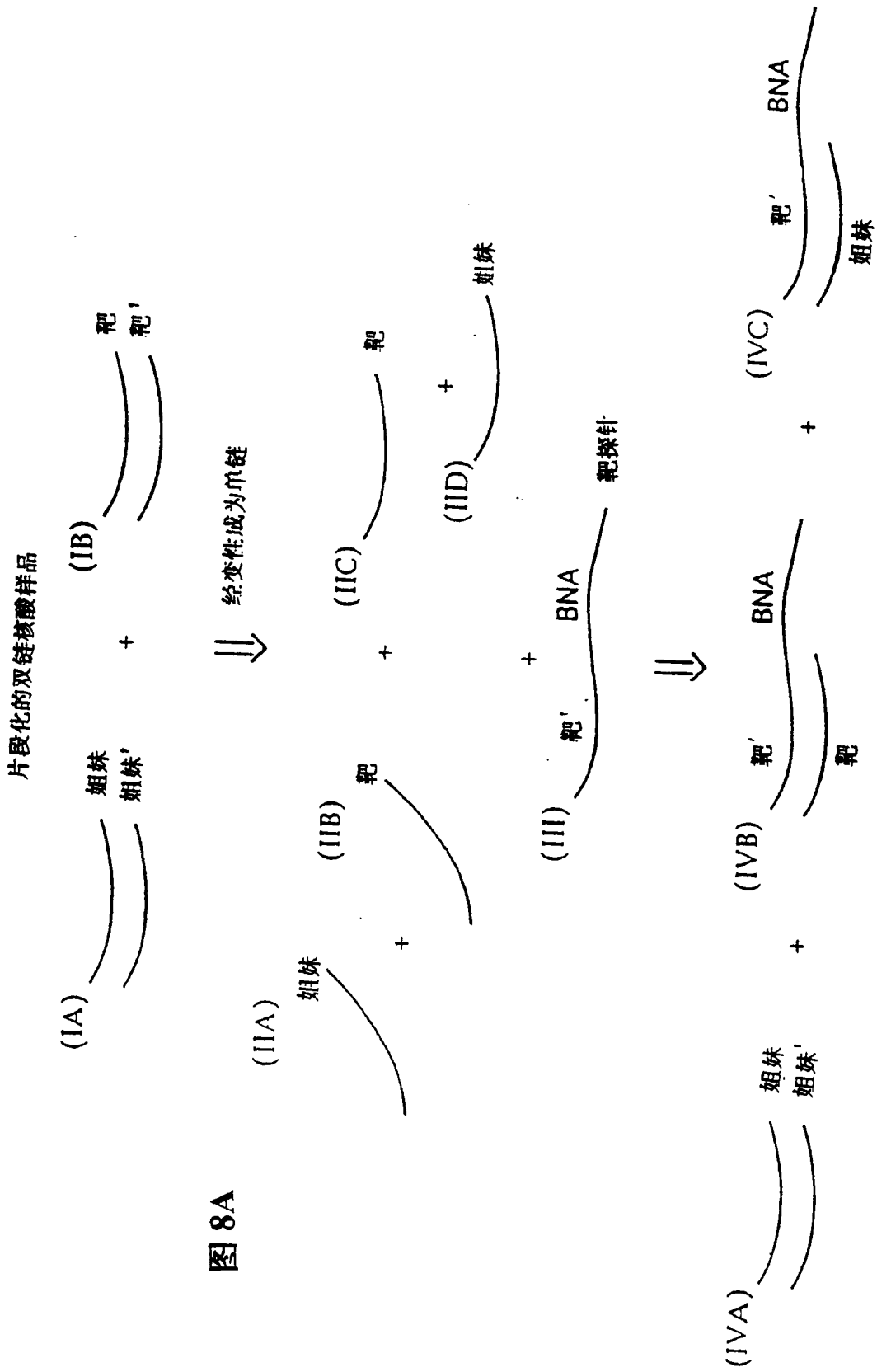


图 8A

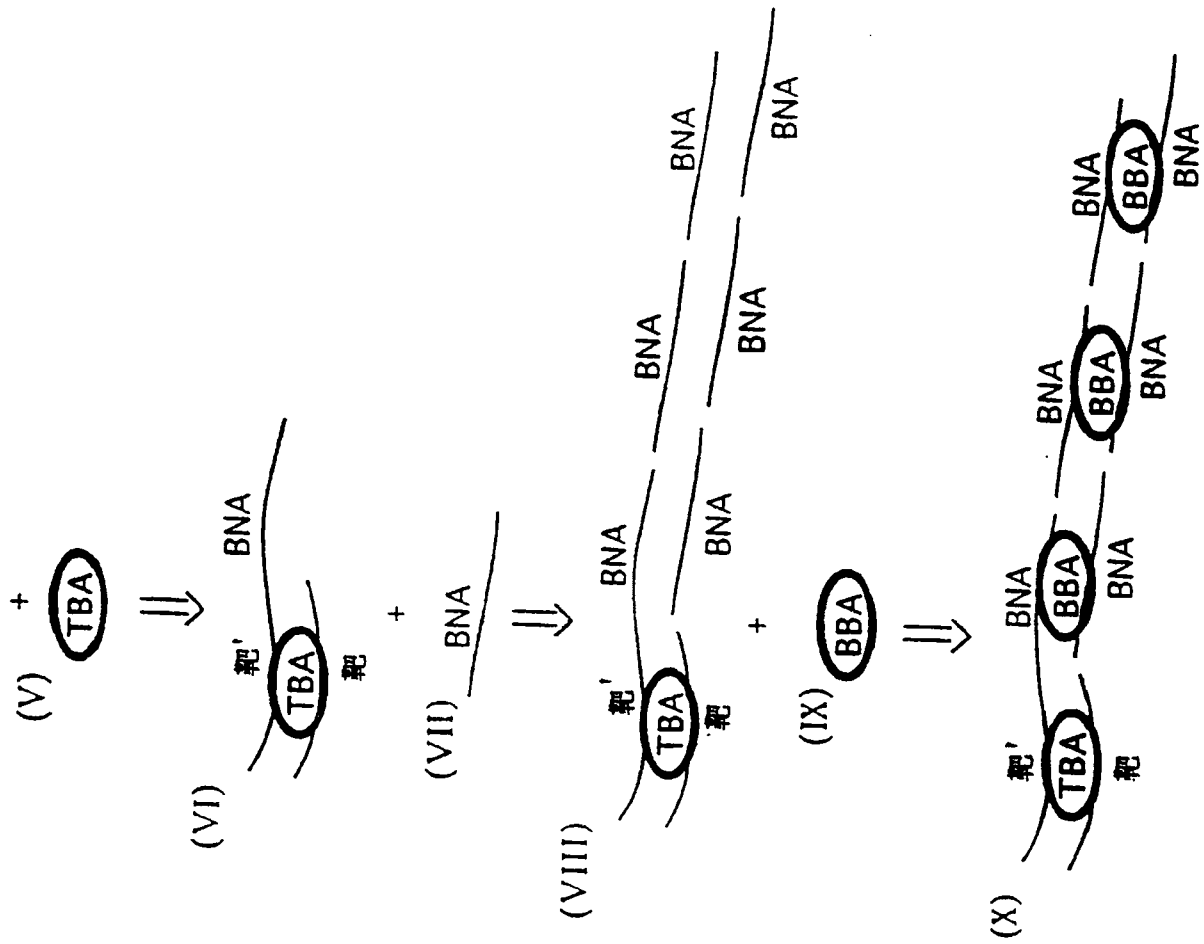


图 8B

TBA: 靶结合组装

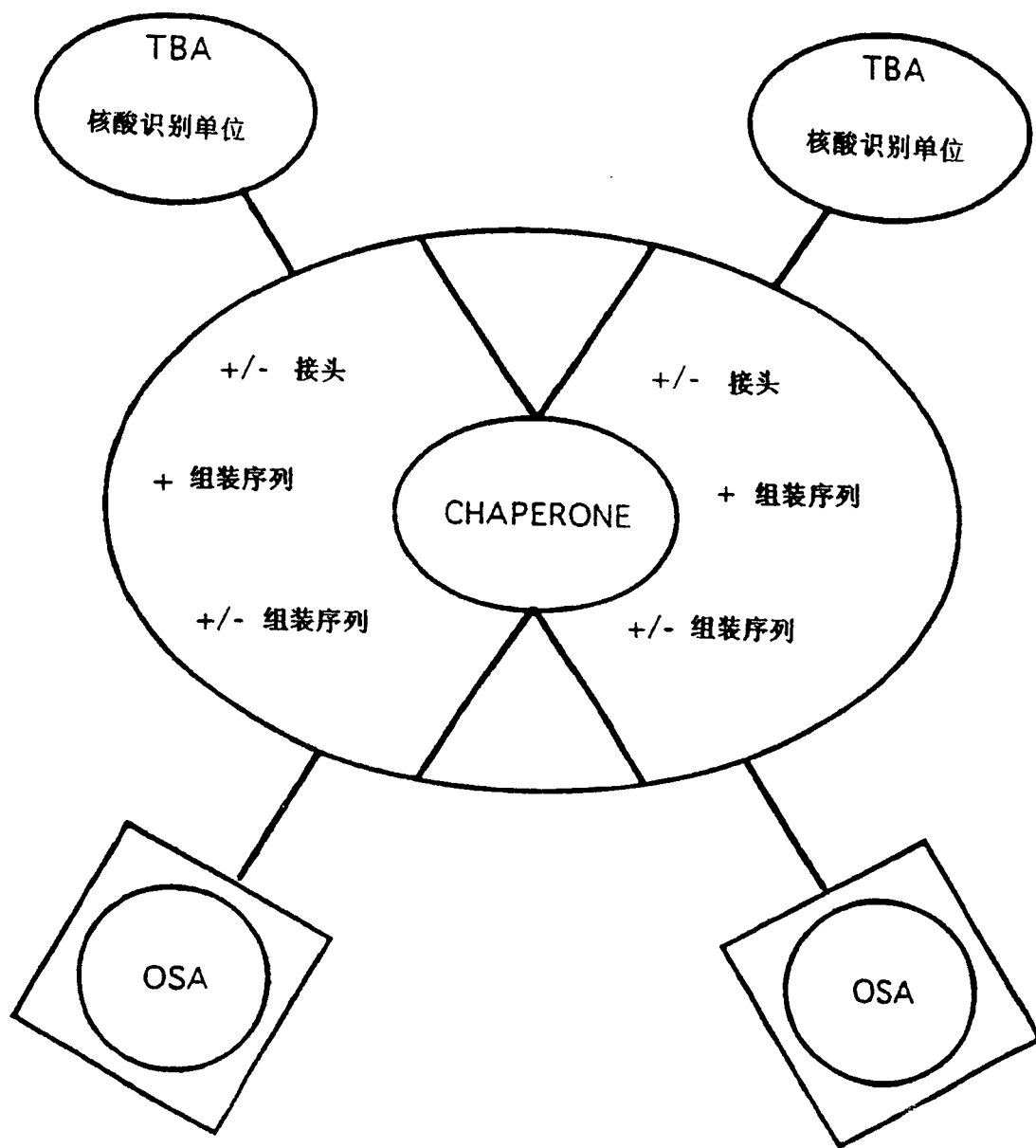
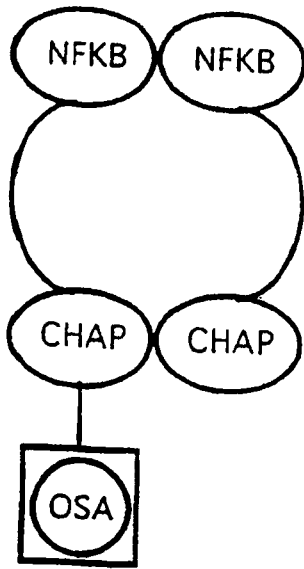
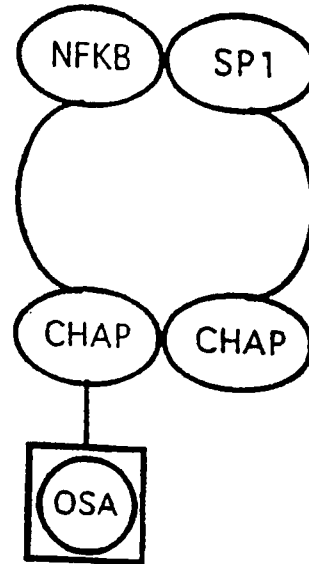


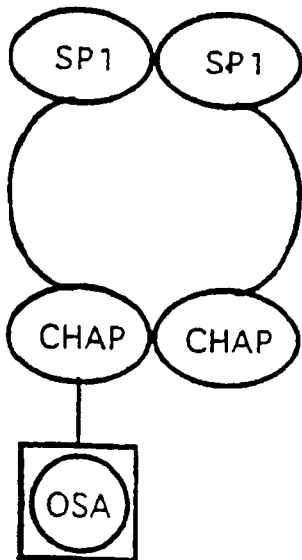
图 9



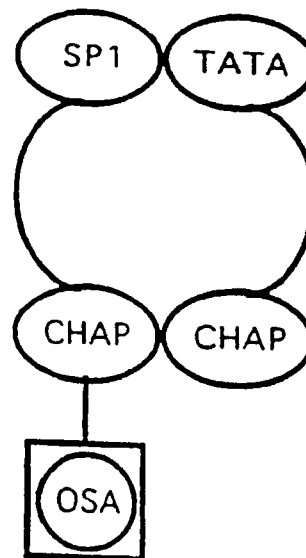
检测 I



检测 II

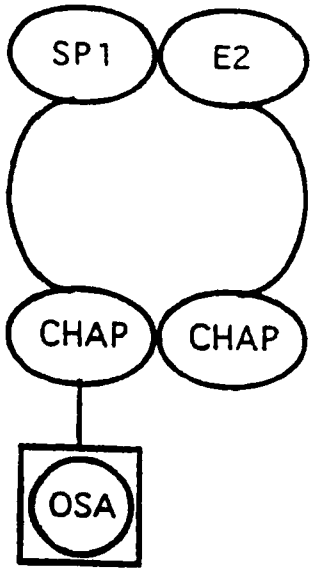


检测 III

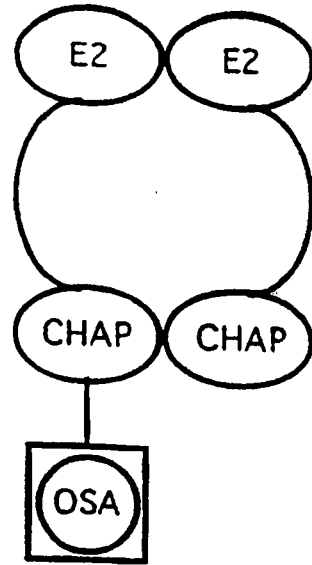


检测 IV

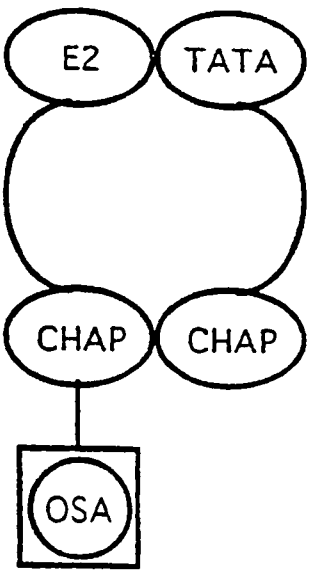
图 10



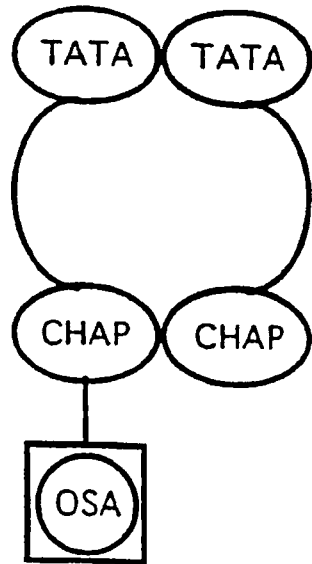
检测 I



检测 II



检测 III



检测 IV

图 11

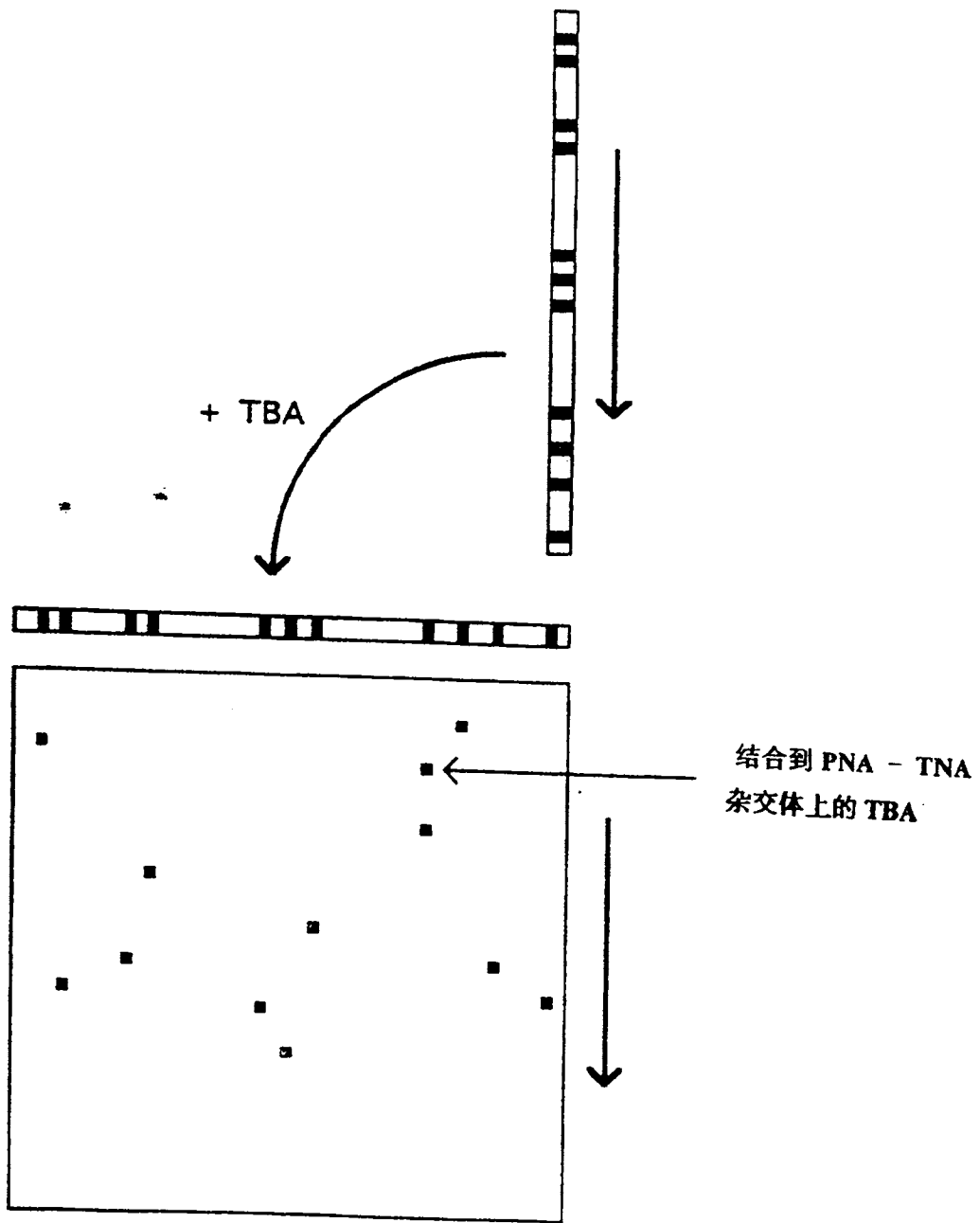


图 12A

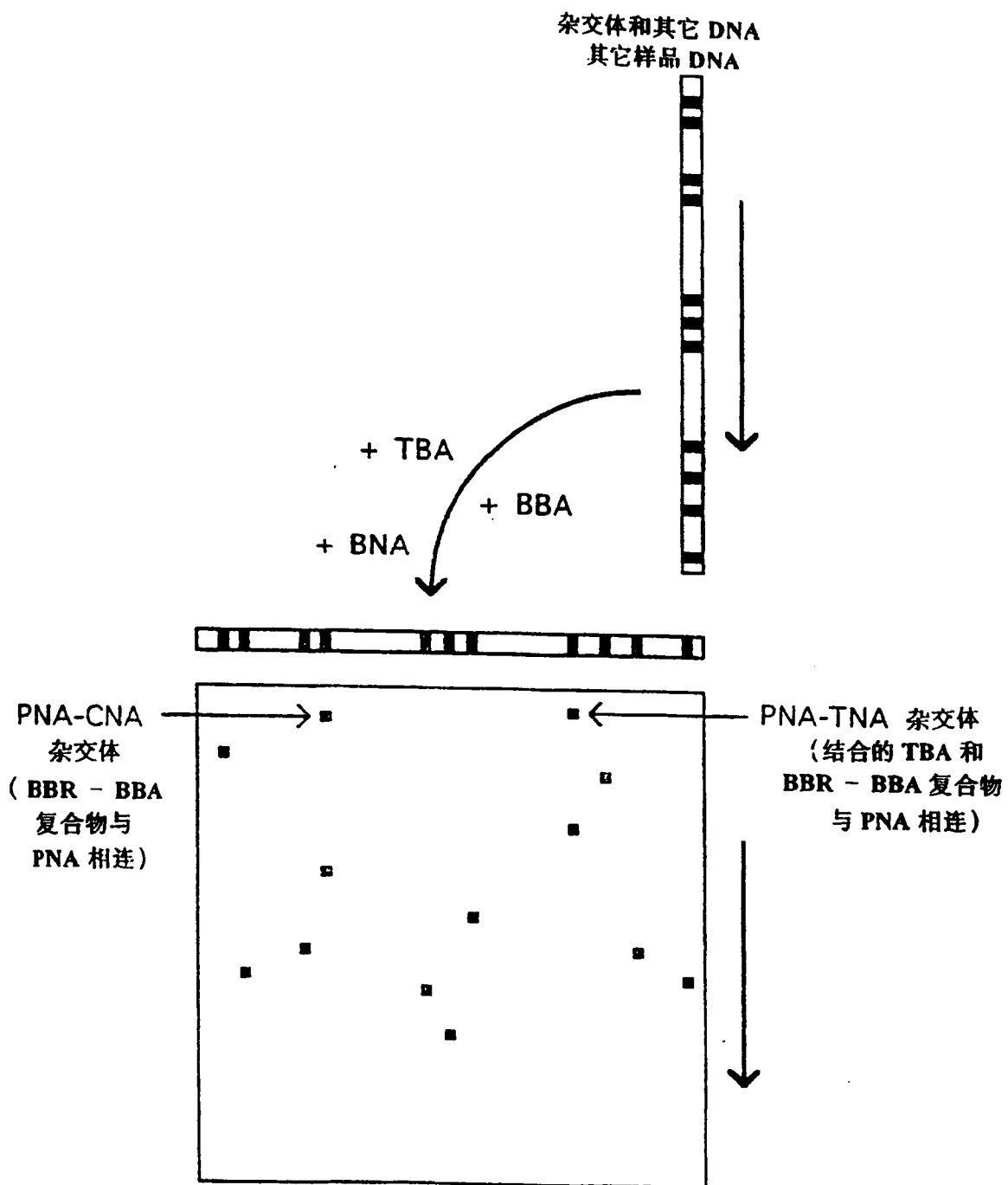


图 12B

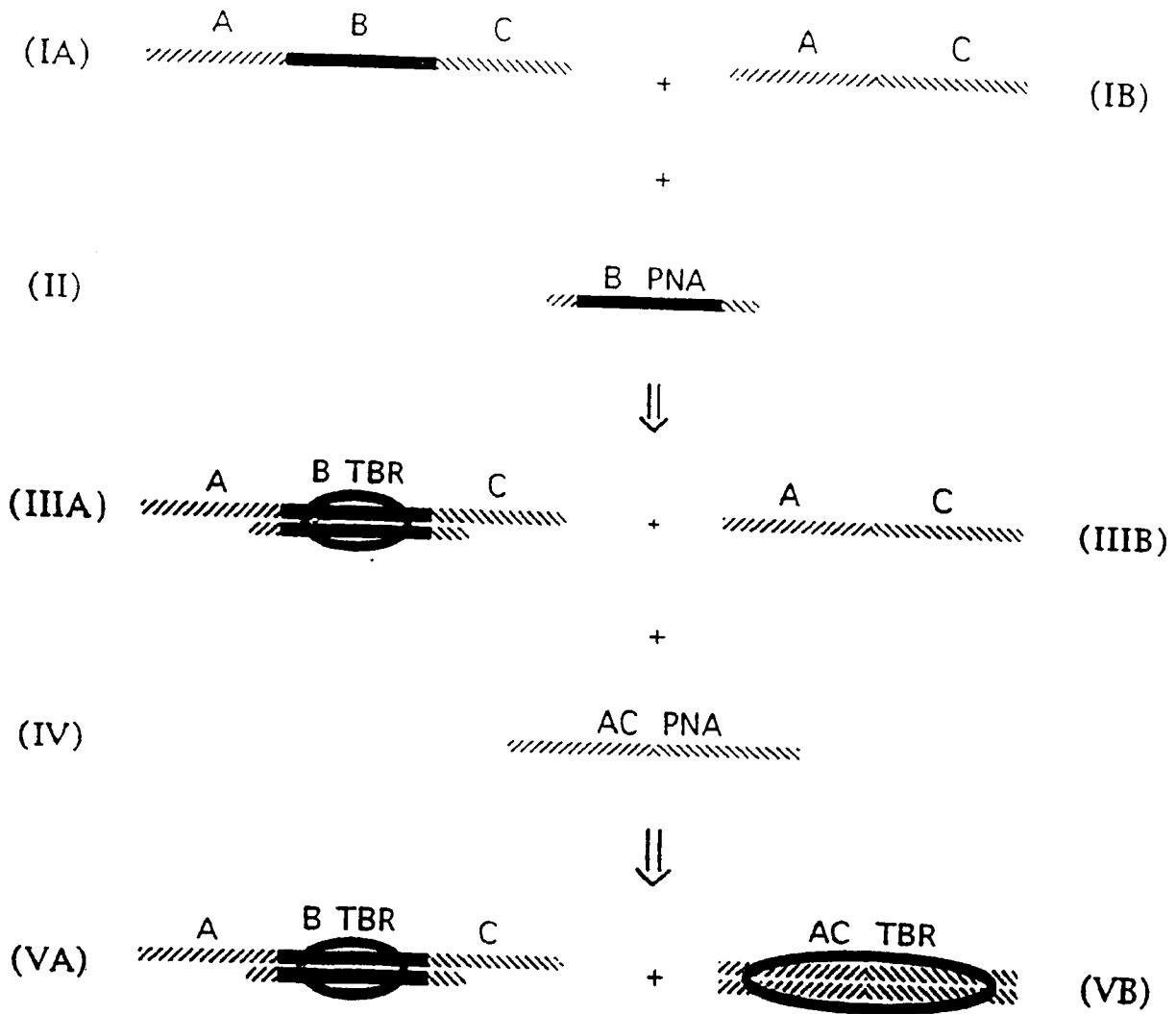


图 13



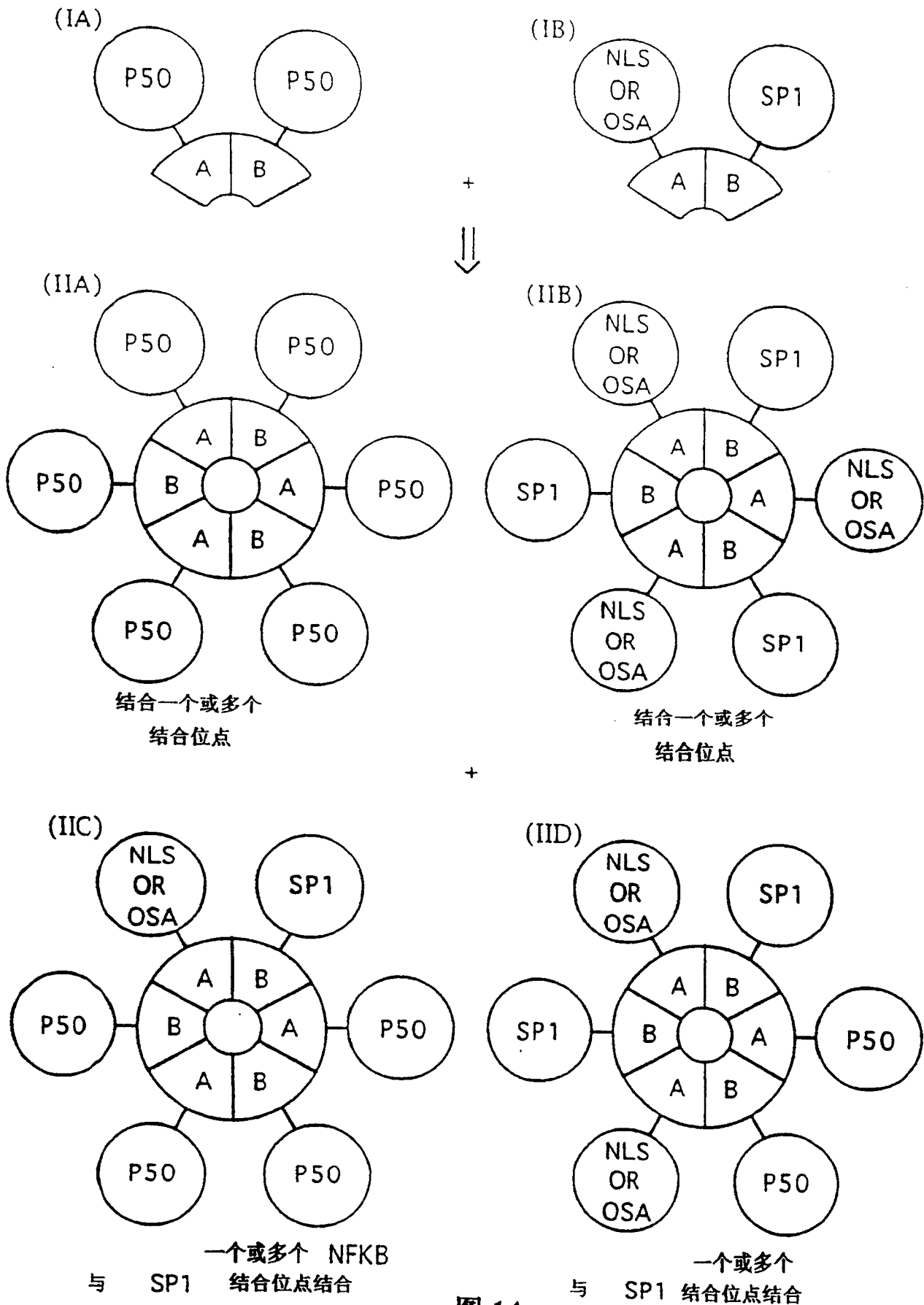


图 14

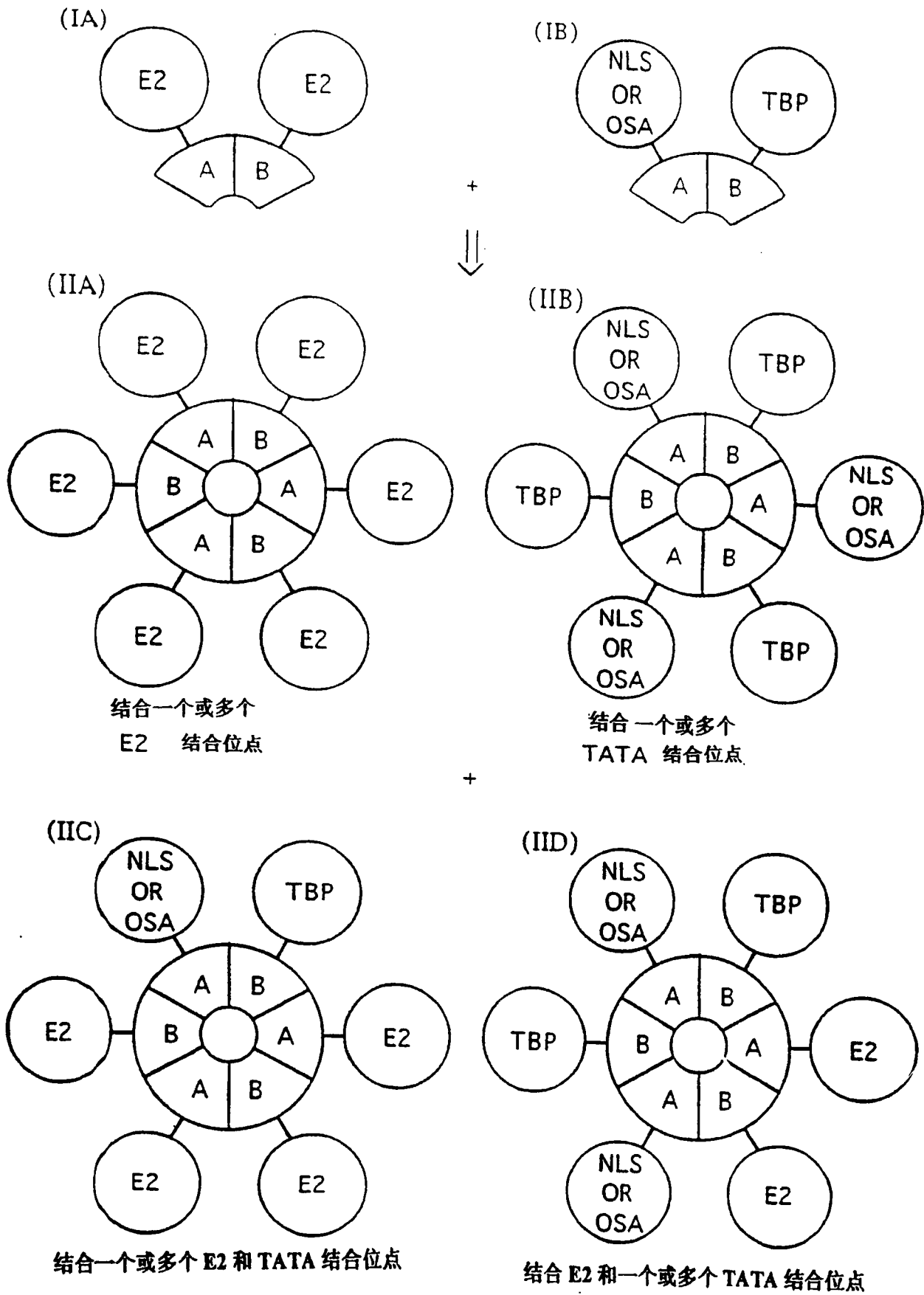


图 15

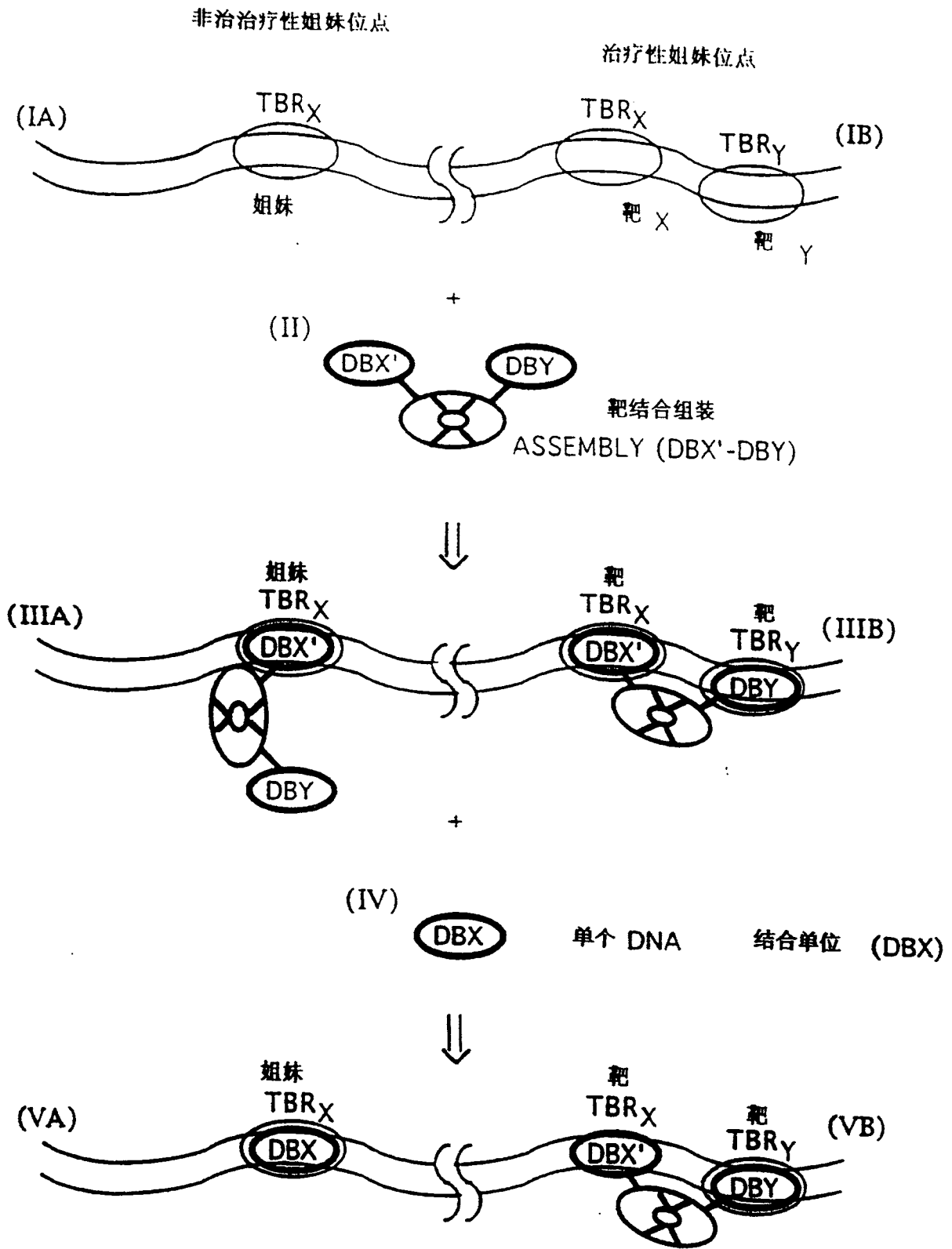


图 16

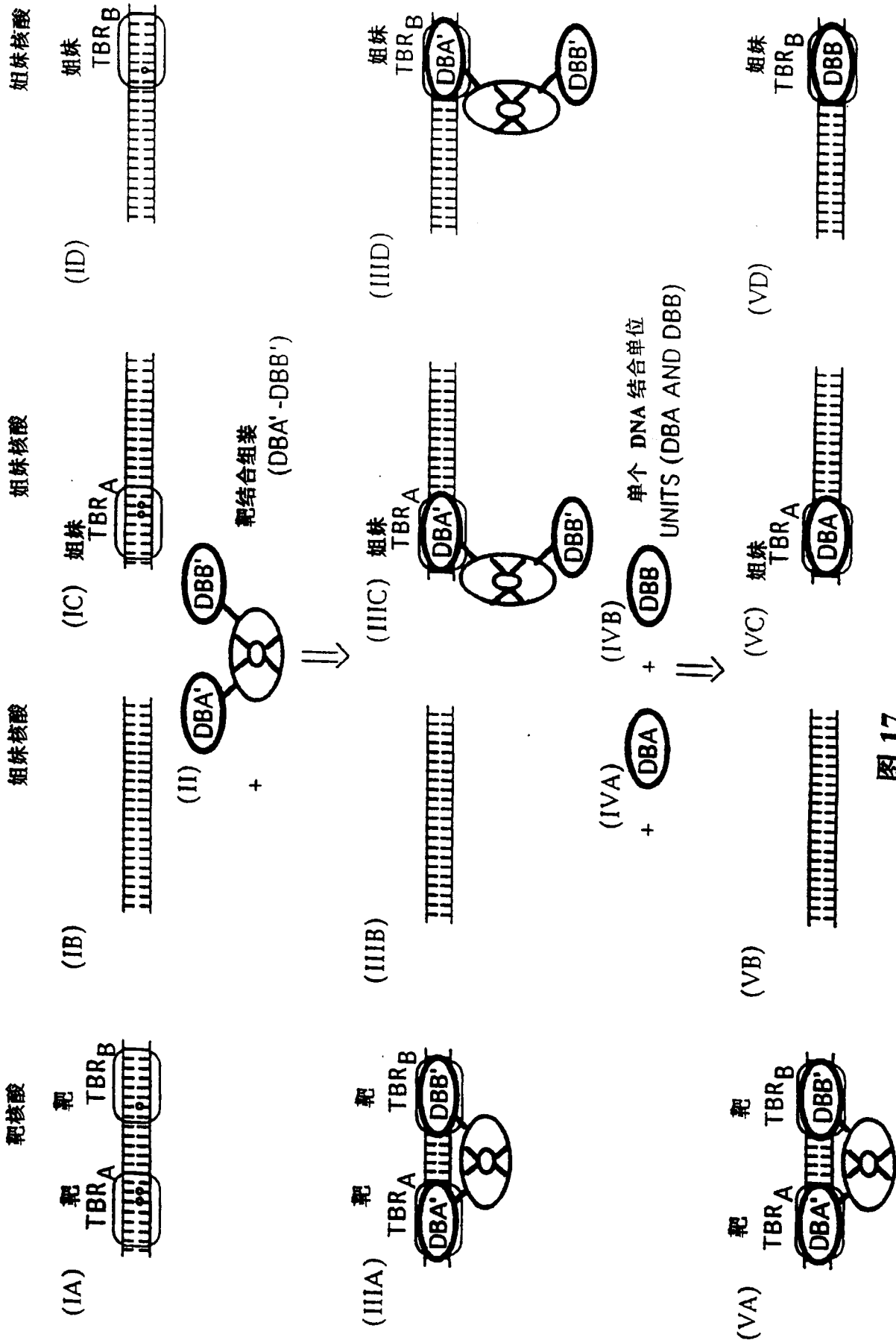


图 17