[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 200680049801.0

[43] 公开日 2009 年 2 月 11 日

[11] 公开号 CN 101365805A

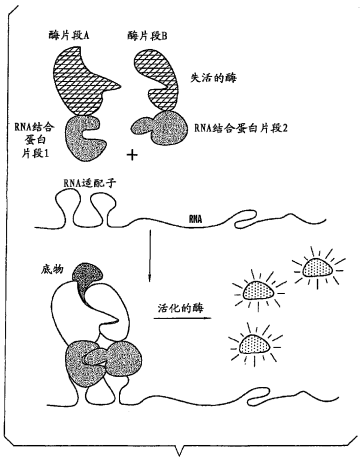
[22] 申请日 2006.10.27
[21] 申请号 200680049801.0
[30] 优先权
[32] 2005.10.27 [33] US [31] 60/730,746
[86] 国际申请 PCT/US2006/042210 2006.10.27
[87] 国际公布 WO2007/050979 英 2007.5.3
[85] 进入国家阶段日期 2008.6.27
[71] 申请人 波士顿大学董事会
地址 美国马萨诸塞州
[72] 发明人 N·布劳德 C·R·坎托
M·伯顿 V·V·德米达夫
[74] 专利代理机构 北京泛华伟业知识产权代理有限公司
代理人 曹津燕 郭广迅
权利要求书 5 页 说明书 64 页 附图 37 页

[54] 发明名称

使用蛋白质互补在体内实时检测核酸

[57] 摘要

本发明涉及使用实时蛋白互补方法体内检测核酸分子如 RNA 分子的方法。 本发明进一步涉及高灵敏度实时检测活细胞中核酸，例如 RNA 的方法，该方法使用了本发明的新的裂解生物分子接合物。



1. 用于实时检测核酸的方法，包括:

a. 表达编码检测子构建体的核酸序列，其中检测子构建体包含接合到核酸结合基序上的第一多肽片段，以及至少一种接合到核酸结合基序上的其它多肽片段，其中两种多肽片段结合形成活化状态的检测子蛋白，其中在与活化的野生型蛋白相比时，两个片段处于活化的且构象正确的形态;以及

b. 表达编码报道子构建体的核酸序列，其中报道子构建体包含目标核苷酸和核酸结合序列，其中核酸结合序列由检测子构建体中的两种或更多种核酸结合基序识别；以及其中多肽片段的两种或更多种核酸结合基序与核酸结合序列结合实时重构成活性检测子蛋白；以及

c. 用于检测重构的检测子蛋白的方法。

2. 权利要求 1 的方法，其中所述检测是在体外或体内进行的。

3. 权利要求 1 的方法，其中所述检测是在体内进行的。

4. 权利要求 1 的方法，其中所述核酸是 RNA。

5. 权利要求 1 的方法，其中所述核酸是 DNA。

6. 权利要求 1、3 或 4 的方法，其中与所述检测子蛋白的第一多肽片段相结合的所述核酸结合基序是全长基序的一部分，该基序的余下部分与检测子蛋白的至少一种其它多肽片段相结合。

7. 权利要求 1、3、4 或 6 的方法，其中与所述检测子蛋白的第一多肽片段相结合的所述核酸结合基序是全长基序，该基序与结合到所述检测子蛋白的至少一种其它多肽片段的核酸结合基序是无关的。

8. 权利要求 7 的方法，其中所述核酸结合基序包含多结构域核酸结合分子的结构域。

9. 权利要求 1、3 或 4 的方法，其中所述检测子蛋白是荧光蛋白。

10. 权利要求 9 的方法，其中所述荧光蛋白选自:绿色荧光蛋白(GFP);增强型绿色荧光蛋白（EGFP）；绿色荧光蛋白样蛋白（GFP 样）；黄色荧光蛋白（YFP）；增强型黄色荧光蛋白（EYFP）；蓝色荧光蛋白（BFP）；增强型蓝色荧光蛋白（EBFP）；青色荧光蛋白（CFP）；增强型青色荧光蛋白（ECFP）；红色荧光蛋白（dsRED）以及它们的变型和片段。

11. 权利要求 10 的方法，其中所述荧光蛋白是 EGFP。

12. 权利要求 11 的方法，进一步包括位于第一和第二 EGFP 片段之间的裂解产物。

13. 权利要求 11 的方法，其中所述 EGFP 的第一片段为第 1 位氨基酸到大约第 158 位氨基酸，并且其中所述 EGFP 的第二片段为大约第 159 位氨基酸到第 239 位氨基酸。

14. 权利要求 1 的方法，其中所述检测子蛋白是酶。

15. 权利要求 13 的方法，其中所述酶选自：β-半乳糖苷酶、β-内酰胺酶、β-葡萄糖苷酶、β-葡萄糖醛酸酶、氯霉素乙酰转移酶、二氢叶酸还原酶（DHFR）。

16. 权利要求 14 的方法，其中所述酶是β-内酰胺酶。

17. 权利要求 15 的方法，其中所述β-内酰胺酶的第一片段为大约第 24 位氨基酸到第 197 位氨基酸，并且其中所述β-内酰胺酶的第二片段为大约第 198 位氨基酸到第 240 位氨基酸。

18. 权利要求 1 的方法，其中所述核酸结合基序是蛋白质。

19. 权利要求 1、3 或 4 的方法，其中所述核酸结合基序蛋白被分成两个或更多个片段，其中一个片段接合到所述第一检测子多肽片段上，并且余下的片段接合到一种或更多种互补多肽片段上，并且其中的片段是：（a）处于活化构象的；（b）非自身活化的；（c）互补的，以便通过结合到报道子构建体中的核酸结合序列上实时重构成活性检测子蛋白。

20. 权利要求 1、3 或 4 的方法，其中所述核酸结合基序是 MS2 外壳蛋白且所述核酸结合序列是 RNA 茎环，或者所述核酸结合基序是 TAR 且所述核酸结合序列是 Tat BIV-1 茎环，或者所述核酸结合基序是 G3/C3 茎环的 3 个重复序列且所述核酸结合序列是转录活化剂，或者所述核酸结合基序是特异性针对于 eIF4A 的适配子且所述核酸结合序列是 eIF4a，或者所述核酸结合基序是特异性针对于报道子构建体中存在的核酸结合序列的适配子，或者所述报道子构建体中存在的核酸结合序列是适配子标签。

21. 权利要求 1、3 或 4 的方法，其中所述用于检测报道子蛋白的方法包括定量或定性方法。

22. 权利要求 1 的方法，其中所述方法进一步包括：（a）检测生物样品中检测子蛋白的基线信号；（b）改变分析条件使得所述报道子构建体的核酸结合序列与接合到所述检测子蛋白的多肽片段上的核酸结合基序的接触发生改变，（c）立即检测生物样品中活性检测子蛋白的变化，其中信

号的变化显示了目标核苷酸的变化。

23. 权利要求 1 的方法，其中所述方法选自荧光显微镜、共聚焦显微镜、电子显微镜、荧光活化细胞分选仪和肉眼观察。

24. 根据权利要求 1 的方法，其中通过用构建体转化或转染细胞而在细胞中表达检测子构建体和报道子构建体。

25. 根据权利要求 1 的方法，其中通过将编码检测子蛋白和核酸结合基序的两种核酸进行符合读码框的融合而将检测子蛋白接合到核酸结合基序上，这样就产生了融合产物。

26. 根据权利要求 25 的方法，其中所述融合产物包括蛋白酶位点或标签以辅助纯化。

27. 根据权利要求 26 的方法，其中所述标签是 6-His 标签或者谷胱甘肽-S-转移酶标签或者肽表位。

28. 包含 DNA 序列的质粒，该 DNA 序列编码至少以下一种:

（i）权利要求 1 的检测子构建体;

（ii）权利要求 1 的报道子构建体；或者

（iii）（i）中的检测子构建体和（ii）中的报道子构建体。

29. 转基因生物，在其基因组中具有适当整合的至少一种权利要求 28 的 DNA 序列。

30. 转基因生物，在其基因组中具有适当整合的以下核酸序列的功能性基因组表达:

a. 编码检测子构建体的核酸序列，其中检测子构建体包含接合到核酸结合基序上的第一多肽片段，以及至少一种接合到核酸结合基序上的其它多肽片段，其中两种多肽片段结合形成活化状态的检测子蛋白，其中在与活化的野生型蛋白相比时，两个片段处于活化的且构象正确的形态；和/或

b. 编码报道子构建体的核酸序列，其中报道子构建体包含目标核苷酸和核酸结合序列，其中核酸结合序列由检测子构建体中的两种或更多种核酸结合基序识别；以及其中多肽片段的两种或更多种核酸结合基序与核酸结合序列结合实时重构成活性检测子蛋白。

31. 权利要求 29 或 30 的转基因生物，其中所述转基因生物选自小鼠、斑马鱼、线虫、酵母、细菌、哺乳动物细胞、原代细胞和次代细胞。

32. 用于检测个体中的疾病或障碍的方法，包括:

a. 将来自个体的 DNA 或 RNA 与权利要求 1 所述的检测子构建体相接触，其中所述核酸结合基序特异性针对于特定疾病或障碍；以及

b. 检测检测子构建体的信号变化，其中检测到的检测子构建体中的信号变化显示了疾病或障碍的存在。

33. 权利要求 32 的方法，其中所述疾病是病原体。

34. 权利要求 33 的方法，其中所述病原体选自病毒；流感，细菌，真菌，寄生虫和酵母。

35. 权利要求 32 的方法，其中所述 DNA 或 RNA 是倾向于遗传疾病的。

36. 权利要求 1 或 3 的方法，其中对所述检测子蛋白片段进行设计使得它们在重构时被立即活化。

37. 权利要求 1 的方法，其中在与活化的野生型蛋白相比时，所述检测子多肽蛋白片段处于活化的且构象正确的形态，其中在靶核酸存在下，互补的检测子多肽蛋白片段实时重构成检测子蛋白和信号表型。

38. 编码检测子构建体和报道子构建体的核酸片段在实时检测核酸中的用途，其中：

（i）检测子构建体包含检测子蛋白片段，在与活化的野生型蛋白相比时，其处于活化的且构象正确的形态；并且其中它们被接合到核酸结合基序上，且至少一种其它多肽片段被接合到核酸结合基序上；以及

（ii）报道子构建体包含目标核苷酸和核酸结合序列，其中核酸结合序列由检测子构建体中的两种或更多种核酸结合基序识别；并且其中多肽片段的两种或更多种核酸结合基序与核酸结合序列结合实时重构成活性的检测子蛋白。

39. 权利要求 38 的核酸片段的用途，其中所述实时检测核酸是在体内检测的。

40. 包含质粒的试剂盒，该质粒包含 DNA 序列，该 DNA 序列编码至少以下一种：

（i）权利要求 1 的检测子构建体；

（ii）权利要求 1 的报道子构建体；或者

（iii）（i）中的检测子构建体和（ii）中的报道子构建体。

41. 权利要求 40 的试剂盒，其中所述检测子构建体包含权利要求 9 或 10 的荧光蛋白或者权利要求 15 的酶。

42. 权利要求 40 的试剂盒，其中所述报道子构建体包含权利要求 20 的适配子。

43. 权利要求 1 的方法，其中所述检测是在选自：成纤维细胞、神经细胞、卵母细胞、肿瘤细胞、病毒感染的哺乳动物细胞、表皮细胞、细菌细胞和酵母细胞的细胞中进行的。

44. 权利要求 43 的方法，其中所述细胞是经基因修饰的细胞。

使用蛋白质互补在体内实时检测核酸

相关申请的交叉引用

本申请依据 35 U.S.C. 119（e）要求 2005 年 10 月 27 日提交的 No.60/730,746 号美国临时专利申请的权益，并将其内容整体引入本发明作为参考。

发明领域

本发明涉及用于体内检测核酸的组合物和方法。更优选地，组合物和方法能够在体内灵敏地且实时地检测 RNA。

发明背景

RNA 是广义上被认定为基因表达的多步骤过程的积极参与者，该过程包括 RNA 在细胞核内的转录和加工、由细胞核输出、经细胞质转运以及在核糖体内翻译。此外，非编码 RNA——类日益增多的 RNA 分子，参与了多种转录后和翻译后事件，这些事件涉及所有的细胞大分子、蛋白质、DNA 和 RNA: RNA 编辑、RNA 修饰、DNA 甲基化和蛋白修饰（Kiss, 2002, Mattick, 2004; Huttenhofer 等, 2005）。为了发挥这样的多种功能，RNA 必须在正确时间处于正确的细胞位置上。换句话说，RNA 分子在细胞中的空间和时间定位已经成为目前细胞生物学中的一项重要机制。

已经认识到 RNA 的定位不仅调节蛋白质的合成，而且形成了形态发生素梯度，并且决定了细胞系和细胞器的形成（综述参见 Kloc 等, 2002）。如果 RNA 没有适当地发挥作用，其可能造成各种形式的病状。RNA 的动态性和不稳定性及其在多种功能中的重要作用都与它的运动和定位有关，这些都已经引起了兴趣和挑战。

目前，体内 RNA 定位领域中所面临的困难的一个例子是缺少适当的方法来评估 poly（A）+-mRNA 在细胞核中的扩散系数。已经报道的数值相差两个数量级，即 0.03-0.6 μ^2/秒和 9 μ^2/秒（Politz 等, 1998, Politz 等, 1999, Molenaar 等, 2004）。因此，为了更好地了解各种 RNA 在体内的功能和行为，显然是迫切需要新的方法来研究它们在活细胞中的定位和运

动。

为了观察细胞中的RNA，应该对其进行选择性标记。有人已经建议并使用了用于体内RNA标记的不同策略（综述参见Pederson, 2001）。它们中的大多数使用了不同的RNA特异性杂交探针，并使用显微注射或脂转染法将探针输送到细胞核或细胞质中。另一种可能是预先标记RNA并将其注射到细胞中。Kool和合作者们成功地使用自连接淬灭探针（Self-ligating quenching probes）在细菌细胞中进行RNA杂交（Sando & Kool, 2002）。荧光2'-O-甲基-RNA探针已经显示出比未甲基化的寡核苷酸更加稳定，并且被几个小组所采用（Carmo-Fonseca等, 1999; Molenaar等, 2001; Molenaar等, 2004）。分子信标和笼锁荧光素（caged-fluorescein）标记的反义寡核苷酸提供了先进的RNA杂交探针的变体（综述参见Politz, 1999, Pederson, 2001）。分子信标和笼锁探针的优点在于它们只有在分别与靶点杂交或者被照射时才产生信号；这就能获得较低的背景并因此获得更好的灵敏度（Sokol等, 1998; Perlette &Tan, 2001; Matsuo, 1998; Tsuji等, 2001; Sei-Iida等, 2000）。

这种杂交策略最大局限性在于杂交灵敏度低，这是由于细胞中RNA的浓度低。在大多数情况下，只有高丰度的RNA种类才能被检测到（β-肌动蛋白mRNA、c-fos mRNA、碱性成纤维细胞生长因子RNA或者总poly（A）-RNA）。使用寡核苷酸探针在体内进行RNA检测的另一个障碍是它们会在细胞核中快速积累（Tsuji等, 2000; Molenaar等, 2001）。

研究体内RNA的另一种策略是利用RNA结合蛋白和荧光蛋白之间的融合，在RNA靶点中引入蛋白结合标签。两个小组使用了基于MS2外壳蛋白和相应的RNA基序(motif)之间的相互作用的系统(Bertrand等, 1998; Beach等., 1999; Beach & Bloom, 2001），另一个小组使用了基于U1A剪接蛋白（splicing protein）的系统（Takizawa & Vale, 2000）。这种方法寄希望于能够对酵母细胞（Bertrand等, 1998; Beach等, 1999; Corral-Debrinski等, 2000; Beach & Bloom, 2001）和神经元（Rook等, 2000）中的特定RNA的运动进行实时监测。这项技术代表了将荧光蛋白与细胞中的RNA研究相结合的首次尝试。然而，实质性的限制在于功能完整的荧光蛋白的高背景信号使得实际监测很困难，该荧光蛋白是融合到RNA结合蛋白上的。

进一步来说，体内RNA分析主要依赖于荧光检测方法，其中RNA特异性荧光探针要么被输送到细胞中（例如分子信标），要么在细胞内合成

（例如融合到 RNA 结合蛋白，例如 MS2 外壳蛋白上的增强型绿色荧光蛋白，EGFP）[综述参见（14、15）]。然而，这些方法都有严重的缺点和局限性。预先标记的寡核苷酸检测子探针必须被修饰成抗核酸酶的，而且必须使用侵入性技术将其输送到细胞中（16）。第二种使用 RNA 结合蛋白的方法涉及采用串联形式的多拷贝（高达 96 个）RNA 结合蛋白识别序列来修饰目标 RNA，同时将 RNA 结合蛋白融合到全长荧光蛋白上。结果，融合蛋白的多亚基大复合物就被组装到目标 RNA 上（17-20），这可能损害细胞中 RNA 的正常运动和迁移。在细菌中也存在由全长荧光蛋白导致的高背景荧光，该荧光蛋白易于聚集，并且聚集体可能与 RNA-蛋白复合物相混淆。在真核细胞中，荧光蛋白和 RNP 复合物分开处在不同区室中在某些情况下这是有利的（21），但是一般来说全长荧光蛋白所导致的高荧光背景限制了该方法的灵敏度。

因此，目前所有的 RNA 标记方法都受到高非特异性背景和缺少信号放大的困扰；因此它们中的大多数受限制于种类丰富的 RNA 的检测。此外，实时检测 RNA 的需求也很大。

发明概述

本发明的发明人发现了使用实时蛋白质互补方法在体内检测核酸分子，如 RNA 分子的方法。特别地，本发明提供了具有高信号:背景比的体内检测核酸，例如 RNA 的方法，由此能够以高灵敏度来检测 RNA。此外，本发明的方法提供了采用非侵入性方法检测活细胞中的 DNA 和 RNA 的方法和组分（components）。更特别地，本发明的组合物包含检测子分子（detector molecule），该检测子分子包含连接到核酸结合基序上的被裂解成两个或更多个多肽片段的检测子蛋白，它们可以独立或相互配合结合单一位点。检测子多肽蛋白片段与具有功能活性的检测子蛋白的重新结合仅仅发生在靶核酸，如 RNA 或 DNA 存在的情况下，该靶核酸已经被修饰成包含报道子核酸结合序列（例如适配子），使得核酸结合基序能够与报道子核酸序列上的其同源结合位点相互作用。这种相互作用聚集了检测子蛋白片段从而能够立即检测信号。检测子分子多肽特别是检测子蛋白组分处于活化构型，但其单独时并不会被活化，而是一旦重组它们就立即形成活性蛋白，从而在靶核酸存在时实时产生信号。在一个实施方案中，靶核酸是 RNA。并且，一个实施方案包括了作为检测子分子的核酸结合基序组分

的 RNA 结合蛋白或肽。

如本发明所述，“检测子构建体”是指编码包含上述检测子蛋白（带荧光的或酶的）的检测子分子的核酸序列，该检测子蛋白被裂解成两个或更多个活化的多肽片段，每个片段都与核酸结合基序接合（conjugated）。经接合的核酸结合基序与靶核酸结合，从而将多肽片段聚集在一起并重构成完全活化的蛋白，就可以被立即检测。

本发明所述的“报道子构建体”是指编码报道子分子的核酸序列，该报道子分子包含上述核酸，例如编码目标基因的 RNA 或 DNA，以及靶核酸结合序列，例如适配子。核酸结合序列可以被检测子构建体的核酸结合基序所识别。核酸结合序列可以是核酸、蛋白、适配子或适配子标签。

在一个实施方案中，检测子蛋白是荧光分子，例如 EGFP。在本发明的一个实施方案中，荧光报道子是被裂解成α片段（大约第 1-158 位氨基酸）和β片段（大约第 159-239 位氨基酸）的 EGFP。α片段包含成熟的完全成形的发色团，它单独并不产生荧光，而是为产生荧光备用的，当其与β片段配对时就立即产生荧光。立即产生荧光就能够进行体内实时检测 RNA，这是目前的现有技术所无法实现的。重要的是，α和β片段在没有靶核酸，例如靶 RNA 的情况下是不会重新结合或者发荧光的。

在另一个实施方案中，荧光蛋白是绿色荧光蛋白（GFP）或增强型绿色荧光蛋白（EGFP）。在备选的实施方案中，荧光蛋白是黄色荧光蛋白（YFP）、增强型黄色荧光蛋白（EYFP）、蓝色荧光蛋白（BFP）、增强型蓝色荧光蛋白（EBFP）、青色荧光蛋白（CFP）、增强型青色荧光蛋白（ECFP）或者红色荧光蛋白（dsRED）或者上文列举的荧光蛋白的任何其它天然或基因工程改造的荧光蛋白。在更具体的实施方案中，重构的荧光蛋白可以由上文列举的荧光蛋白片段的混合物或任意上文列举的荧光蛋白的组合所组成。

备选的检测子蛋白是能够放大信号的酶。该酶可以是例如β-半乳糖苷酶、β-内酰胺酶、β-葡萄糖苷酶、β-葡萄糖醛酸酶、氯霉素乙酰转移酶。在特定实施方案中，检测子酶是β-内酰胺酶。

在本发明的一个重要实施方案中，组合物和方法提供了能够在体内实时检测 RNA 的报道子分析方法。在一个实施方案中，制备稳定表达检测子构建体的目标细胞。检测子构建体表达检测子分子，在没有靶 RNA 的情况下，检测子分子的检测子蛋白不产生信号，从而不发荧光或者显示出

活性。细胞也能表达报道子构建体，该构建体编码包含用于检测子分子的核酸结合基序的结合位点的报道子分子以及目标基因。这种报道子构建体的表达可以被诱导。当这种报道子分子在细胞中表达时，检测子分子将结合靶 RNA 并促进检测子蛋白的重构以及活性蛋白和信号的产生。这就能够在体内实时检测 RNA 的表达，在图 10-11 和 18-19 以及实施例 9 到 11 中进行了举例说明。

在备选的实施方案中，目标细胞能够表达包含报道子分子的双顺反子核酸序列，其中检测子构建体是以组成型表达的，报道子构建体的表达与目标启动子可操作地相关联。在相关的实施方案中，报道子构建体（可操作地连接到目标启动子上）和检测子构建体处在分开的核酸序列上。当启动子在细胞中被某些刺激物活化时，靶核酸序列的结合位点就变得可用了，检测子分子的核酸结合蛋白的结合促进了检测子多肽片段的结合以及能够被立即检测到的活性检测子蛋白的形成。该实施方案能够在体内实时地研究启动子功能。例如，能够通过转染质粒和分析任何初始检测子构建体的活性来实时研究基因的调节，该质粒表达驱动 RNA 结合位点的启动子。在相关实施方案中，i）将细胞与一种或多种化合物相接触；或者 ii）使细胞经受各种可能改变启动子活性的情况，从而增加、减少或者不改变 RNA 的转录。这种变化通过检测子构建体立即检测到。

在相似的实施方案中，本发明的方法可以采用“tet-on”或“tet-off”系统。例如，可以制备稳定表达本发明的检测子构建体的目标细胞。此外，细胞还可以稳定表达 tet-on 或 tet-off 可调节报道子的构建体，使得检测子分子核酸结合基序的靶结合位点和目标基因的转录受到控制。例如，在 tet-off 系统中，细胞将立即表达目标基因，报道子构建体将被活化并可以立即检测到。一旦加入 tet，报道子构建体（包含目标基因和靶核酸结合位点）的转录就停止了，只有已经转录的转录物能够被实时分析。在这种 tet-off 的例子中，我们能够通过检测衰减的报道子荧光或活性来在体内实时分析 RNA 的稳定性，例如它的半衰期。

在备选实施方案中，使用”tet-on”构建体能够在活细胞中稳定表达目标基因，因此将 tet 加入到系统中就引发了报道子构建体的转录。然后，本发明的检测子分子能够结合它的靶 RNA 并且可以立即检测到。在该实施方案中，可以如进行实验的人员所要求的那样实时研究 RNA 定位。

在本发明的另一个实施方案中，报道子分析是在生物体内进行的。例

如但不限于动物、非人动物、小鼠、斑马鱼、线虫或蛙，它们可以经处理以表达检测子构建体和目标基因，而目标基因上带有将结合到检测子构建体上的 RNA 结合位点。两种构建体在生物中的表达使得能够通过从第 1 天的胚胎直到成年期进行取样从而对整个发育期间的目标转录物进行分析。在一个实施方案中，生物是哺乳动物。在一个实施方案中，哺乳动物是小鼠，在另一实施方案中，生物是转基因生物，例如但不限于转基因小鼠。

可选择地，本发明的组合物和方法可用于类似于原位杂交的体外实时检测 RNA。在这种实施方案中，将表达包含目标 RNA 的报道子构建体的细胞进行透性化（permeablized），该目标 RNA 带有将结合到检测子分子的核酸结合基序上的 RNA 结合序列，并且将本发明的检测子构建体给予细胞。如果目标 RNA 被表达，检测子构建体将结合到 RNA 结合序列上，就能够确定 RNA 的定位和丰度。该系统相对于传统原位杂交技术的优点在于能够检测低丰度的 RNA（这是由于采用酶报道子的检测子构建体所获得的信号放大）并且能够迅速完成分析且没有放射性。立即进行检测子构建体检测是非常值得期待的。目前，进行原位杂交（例如利用放射性）的传统方法在知道实验是否成功之前可能耗时数周到数月。此外，传统原位杂交中使用的 RNA 探针很难制备，这是由于 RNA 容易降解以及使用了放射性。本发明的方法无需使用放射性或 RNA 探针。检测子构建体很容易例如由细菌来制备，正如本领域技术人员已知的。这种分析方法类似于免疫组织化学或免疫细胞化学技术，其通过使用抗体来体外检测蛋白。这里检测的是 RNA。

在备选实施方案中，本发明的组合物和方法用于实时检测核酸，特别是体内的 RNA 和 DNA，当在体外的细胞中操作 RNA 和蛋白构建体时，例如在试管中就涉及到本发明的应用，作为非限制性的例子，检测离体（ex vivo）细胞中的 RNA。

此外，本发明的组合物和方法可以用于实时体内检测 DNA 的实施方案。例如，本发明的检测子构建体可以包含检测子蛋白（荧光的或酶的），它被裂解成至少两个失活的多肽片段，每个片段都与核酸结合基序结合。当靶核酸，例如适配子的存在使多肽片段聚集在一起时，多肽片段就形成了完全活化的检测子蛋白，并能够被立即检测到。例如，本发明的方法能够实时检测基因组中各种位点的复制。与检测子构建体中的检测子蛋白相

结合的核酸结合基序是特异性针对于 RNA 或 DNA 的各个区域的，并且那些位点的复制能够在体内实时检测。

类似地，我们可以使用本发明的方法和组合物在体内实时检测染色体。例如，检测子构建体可以包含与端粒末端转移酶结合蛋白相结合并在细胞中表达的检测子蛋白。这样，在细胞的生命周期中都能够实时检测端粒末端转移酶。本发明记载的组合物和方法的优点在于通过提供端粒末端转移酶结合蛋白提高了特异性，该端粒末端转移酶蛋白被切成两半，每一半都与检测子蛋白的一半结合。端粒末端转移酶结合蛋白的两半协同结合到单一结合位点上确保了在靶向识别之前几乎不存在检测子活性。

在另一个实施方案中，本发明提供了适用于本发明方法，特别是体内检测 RNA 的试剂盒。在一个实施方案中，试剂盒包括检测子构建体和可诱导的报道子构建体。在一个实施方案中，报道子构建体和检测子构建体处在分开的核酸分子上。在另一实施方案中，它们处在双顺反子核酸分子的两部分上。在另一实施方案中，加入实验人员感兴趣的基因从而对报道子分子进行修饰，在另一实施方案中，报道子分子的表达与 tet-on 或 tet-off 启动子可操作地关联起来。在另一个变型的实施方案中，实验人员可以将包含靶核酸序列的报道子构建体修饰成由实验人员感兴趣的启动子调节。在每个这些实施方案中，试剂盒都包含用于检测子分子中的检测子蛋白的裂解多肽的选择，并且还包含用于检测来自报道子蛋白的信号的说明书和试剂。在一些实施方案中，试剂盒还包含适用于捕获和/或检测样品中靶核酸的存在或数量的试剂。用于检测靶核酸的存在和/或数量的试剂可以包括酶活性试剂或者具有针对组装蛋白的特异性抗体。抗体可以被标记。

蛋白互补和酶报道子的组合能够获得实质上降低的背景和信号放大。这形成了能够分析中等丰度的核酸的体内 RNA 检测技术。

在更具体的实施方案中，本发明提供了包含用于检测体内 RNA 的工具的试剂盒。

本发明的其它方面在下文中公开。

附图简述

图 1：研究适配子对酿酒酵母中基因表达的影响的转录系统（Blake 等, 2003）。

图 2：在标记基因的 3'末端插入适配子未影响酵母中的基因表达。将

适配子标签插入 EGFP 的 3'末端（见图 1 对质粒的描述）。在 0.2%半乳糖和 40 ng/ml ATc 存在时，将培养物培养到对数期。数字 1-4 对应于可诱导构建体的四种不同染色体整合。通过 FACS 分析了每种培养物中的 50,000 个细胞。

图 3：用于该项研究的含有 EGFP 和 eIF4A 片段的融合蛋白的略图。将真核起始因子裂解成两个结构域，并且每个结构域都与 EGFP 的两个片段融合以产生 EGFP（A）-eIF4A（F1）和 EGFP（B）-eIF4A（F1）。数字表示包括在每个片段中的氨基酸数量范围。（N,N-末端；C,C-末端）。

图 4：在没有适配子标签时，细菌细胞中两种嵌合蛋白的表达不会引起重构而产生荧光。EGFP（A）-eIF4A（F1）和 EGFP（B）-eIF4A（F1）在 BL21（DE3）细胞中共表达。用 1mM IPTG 在 37℃下诱导培养物 3 小时，通过荧光流式细胞仪监测复合物的形成。将表达全长 EGFP 的 BL21（DE3）细胞用作诱导的阳性对照（bEGFP），A-F1 和 B-F1 的表达被认为是阴性对照。

图 5：采用两种相邻的适配子标签以及与同源适配子独立相互作用的两种肽进行的体内蛋白互补的示意图。裂解的标记蛋白是灰色的，RNA 结合肽（探针蛋白）是橙色的。（图 5B）分别通过针对 HIV Tat 蛋白的 CQ 肽（SEQ ID NO:2）；HIV rev 肽（SEQ ID NO:1）和 HTLV Rex 肽（SEQ ID NO:3）的体外筛选发现了三种适配子（参见表 1 的 Kd 和参考）。这些结构或相似结构可以作为备选设计中的 RNA 标签用于在我们的方案中（与裂解蛋白的方法相反）。

图 6：蛋白互补分析方法得到了用于体内 RNA 研究的核酸相互作用的验证的略图。将具有酶活性或荧光性质的蛋白裂解成两个无活性的部分（称为 α 亚基和 β 亚基）。将这些部分作为与另一蛋白的嵌合体在体内表达，而另一蛋白具有 RNA 结合域。理想的 RNA 结合蛋白将包括两个部分，它们本身是无活性的，但是在一起时是有活性的。在含有可以被 RNA 结合蛋白所识别的基序的 RNA 存在时，标记蛋白的活性被恢复。如果蛋白是酶，信号就被放大了。如果蛋白是发荧光的，就没有信号放大，但也没有背景，因为荧光完全取决于靶 RNA 的存在。

图 7：体内检测 RNA 的示意图。在体内合成靶 RNA 或者将靶 RNA 注射到细胞中。在体内合成两种蛋白的嵌合体或者将两种蛋白的嵌合体转染到细胞中。每种嵌合体含有标记蛋白部分以及特异性结合靶 RNA 中的

特殊二级结构（或基序）的结构域。在携带可以由 RNA 结合蛋白特异性识别的标签的靶 RNA 存在时，蛋白标记被组装起来，依据其活性（荧光或酶活性）进行检测。

图 8：本发明的一个实施方案的示意图，其中在体内实时分析 RNA 表达。目标基因和 RNA 结合位点被编码在质粒上并转染到含有稳定整合了本发明的报道子构建体的细胞中。当基因（和 RNA 结合位点）表达时，报道子构建体就被活化了。

图 9：裂解的 EGFP 的蛋白互补检测活细菌细胞中的 RNA。（a）分别含有裂解的 eIF4A 和裂解的 EGFP 片段的两种融合蛋白的表达不产生荧光信号。（b）全长 EGFP 的表达使 *E. coli* 细胞产生均匀的荧光。（c）两种融合蛋白和带有适配子的 RNA 转录物的共表达形成了经常定位于细胞极的荧光信号。最上面一排：*E. coli* 中表达的分子构建体。第二排：表达互补复合物组分的质粒。第三排：由 FACS 获得的表达 EGFP 互补复合物的细胞的荧光分布；黑色，IPTG 诱导前；红色，IPTG 诱导后。最下面一排：表达互补复合物相应组分的 *E. coli* 细胞的荧光显微照片。

图 10：单个细菌中转录的动态。（a）对表达 RNA 适配子-eIF4A 互补复合物的 *E. coli* 细胞以 30 分钟间隔随时间（Time- lapse）拍摄的荧光显微照片。实验期间的相差显微照片没有变化。

图 11：活细胞中 RNA 检测所促成的荧光动态。（b）不同的单个细胞中的荧光变化的动态。（c）视野中的所有细胞的总荧光变化的动态。（d）沿细胞长轴测定的单个细菌细胞中的实时荧光分布的变化。

图 12：表达不含适配子序列的 RNA 的细胞未显示出高荧光分布，而在 lacZ 转录物存在时，由 FACS 获得了表达互补复合物蛋白组分的细胞的高荧光分布。黑色，IPTG 诱导前；红色，IPTG 诱导后。为了对比，表达含有适配子序列的 RNA 和互补蛋白的细胞的荧光分布以蓝色显示。

图 13：表达来自较高拷贝载体（~100 拷贝）的含适配子的靶 RNA 细胞显示出较高荧光。（a）表达融合蛋白和含有适配子序列的 RNA 转录物的质粒的示意图。（b）由 FACS 获得的表达 EGFP 互补复合物的细胞的荧光分布。黑色，IPTG 诱导前；红色，IPTG 诱导后。为了对比，带有适配子的细胞的荧光分布显示为蓝色，所述适配子是由较低拷贝数（~40）质粒表达的。

图 14：表达全长 EGFP 的细胞的荧光光谱（a）与含有裂解的 EGFP

的重构核蛋白复合物的荧光光谱（b）是不同的。将细胞悬浮液稀释到光密度为 0.5（OD_{600}=0.5），使用 F-2500 荧光计（Hitachi）记录荧光。具有同样光密度的未诱导细胞用作散射的对照。

图 15：新分裂的细胞不发出荧光。（a）在所示的时间点获得的相差照片。（b）相应的荧光照片。

图 16：基于二元肽-RNA 适配子相互作用设计荧光蛋白互补。EGFP 的两个片段α和β，与两个病毒肽 HIV-1 Rex 肽和噬菌体 λN 肽融合。在带有两个相应适配子的 RNA 转录物存在时，两种肽与同源适配子相互作用并将裂解的 EGFP 的两个片段聚集在一起。EGFP 的重组装导致强烈荧光的出现。

图 17：基于二元适配子/肽相互作用的裂解 EGFP 的蛋白互补检测活细菌细胞中的 RNA。低浓度 IPTG（0.01 mM）能够从背景中区分信号。A，1.00 mM IPTG 诱导的细胞的荧光分布。B，降低浓度的 IPTG 诱导的细胞的荧光分布。

图 18：具有不同荧光信号定位的 E. coli 细胞。

图 19：单个细菌中 RNA 定位和浓度变化的动态。（a）表达互补复合物的 *E. coli* 细胞以 30 分钟间隔的随时间拍摄的荧光显微照片。实验持续期间的相差显微照片没有变化。（b）在两个单个细胞中测定的总荧光变化。（c）沿细胞长轴测定的单个细菌细胞中荧光分布的实时变化。

图 20A 和 20B：由 DNA 双链形成验证的功能性 EGFP 重组装。20A：实验略图。用 N-[6-(生物素氨基)己基]-3'-(2'-吡啶二硫)丙酰胺（HPDP-biotin, Pierce）分别对 C 末端和 N 末端上带有 Cys 残基的纯化 EGFP 的α和β片段进行生物素化，纯化并与等摩尔量的链霉亲和素孵育。然后将这些复合物分别与等摩尔量的两种互补的 21-nt 长的寡核苷酸孵育。通过凝胶转移（gel shift）分析检验每个步骤中复合物的收率，发现接近 100%。然后将两种蛋白-寡核苷酸嵌合体以等摩尔浓度混合。1B：天然 EGFP（上面）和带有附加的 12 bp DNA 双链的重组装 EGFP（下面）的荧光发射光谱，在 480 nm 激发。在含有 150 mM NaCl, 1 mM EDTA，pH 7.4 的磷酸钠缓冲液中测定荧光光谱。

图 21：Mg2+离子对天然的和重构的 EGFP 的荧光的不同影响。

图 22A 和 22B: 图 22A: 当带有附加的互补 21-nt 长的寡核苷酸的 EGFP 的α和β片段混合在一起时，荧光快速增加的动态。图 22B： S65T 发色

团形成的动力学途径，附带所有步骤的t1/2（根据Zimmer, 2002）。

图23A-4B: 折叠的EGFPα片段可以含有预先形成的发色团。图23A: 根据离散分子动力学（DMD）模拟（Dokholyan，未出版），10种比对的α片段典型折叠结构的骨架示意图。形成发色团的氨基酸以蓝色显示。注意片段的N末端主要部分的紧密堆积，以及由于与分子其余部分的接触量少而形成的余下片段的非常柔性的C末端部分。图23B: 折叠的α结构域与全长的EGFP的比对。全长EGFP为黄色，α结构域为蓝色。形成发色团的残基（#62-70）显示为红色。图23C: 与全长EGFP的氨基酸相比的EGFP的α片段中每个残基的平均均方根偏差（RMSD）。RMSD值是根据蛋白折叠时低温下的DMD模拟计算的。形成发色团的氨基酸处在阴影区域中，它们的偏差≤2Å。

图24A-24D: EGFP的α片段含有预先形成的发色团。图24A: EGFP的吸收光谱，图24B: EGFP的α和β片段的吸收光谱，图24C: EGFP的荧光光谱（蓝色-激发，粉红色-发射），图24D: α片段的荧光光谱（蓝色-激发，粉红色-发射）。所有蛋白均以等摩尔浓度进行光谱分析。

图25-25B: 通过部分蛋白酶解从GFP中分离的含有发色团的肽的吸收光谱（图25A）以及在酸性/中性pH下化学合成的发色团的吸收光谱（图25B）（根据Niwa等, 1996）。

图26: 核酸相互作用的两种可能排列，它们作为对蛋白互补分析的支持。

图27: 作为嵌合体在E. coli中表达的带有自裂解蛋白内含子（intein）的EGFP β片段（β-cys）的纯化。泳道6中的蛋白是蛋白内含子自裂解后获得的纯EGFP β亚基。

图28: 作为嵌合体在E. coli中表达的带有自裂解蛋白内含子的EGFP α片段的纯化。泳道5中的蛋白是纯EGFP α亚基。

图29: 蛋白与含SH的寡核苷酸的接合几乎是100%完全的。使用非变性PAGE进行分析。接合之后，蛋白被连接到带有负电荷的寡核苷酸上。因此，修饰的蛋白比未修饰的蛋白移动更快（对比泳道1和2，以及3和4）。如果寡核苷酸是生物素化的，就能够形成与链霉亲和素的复合物。这种复合物也比单独的链霉亲和素移动得更快（对比泳道6和7）。根据该数据我们得出结论，寡核苷酸与蛋白的偶联效率接近100%。

图30: 组合的α+A和β+B的荧光显示了活性EGFP的重构，而没有

寡核苷酸时就没有活性 EGFP。

图 31：恢复蛋白的酶学活性的蛋白互补的示意图。原理与图 6 相似。酶的重组装取决于 RNA 与 RNA 结合蛋白的相互作用。活化的重组装酶分解它的底物，这就产生了显色产物或荧光产物。由于蛋白的酶活性，因此信号被放大了。

发明详述

本发明人已经发现了用于快速检测活细胞中的 RNA 的新方法，它能够以高信号:背景比检测出 RNA，使得检测灵敏度提高。本发明包括用于在体外和体内实时地、灵敏地检测核酸，例如 RNA 和 DNA 的组合物和方法。特别地，本发明包括使用检测子分子的方法，其中包括由检测子蛋白裂解成的至少两个多肽片段，所述多肽片段结合到核酸结合基序，例如 RNA 结合基序上，它们可以独立地或配合地起作用以结合到单一核酸结合位点上。只有在靶核酸分子存在时，作为 RNA 结合基序与核酸上它们的同源结合位点的相互作用的结果，才会发生检测子蛋白片段重新结合成功能性的检测子蛋白的情况。这种相互作用将会把检测子蛋白的互补多肽片段聚集在一起用来进行信号检测。

本发明基于蛋白互补提供了用于体内观察核酸如 RNA 的创新方法，这是指两个或更多个多肽蛋白片段重新结合成活性蛋白。特别地，裂解的多肽处于活性构型，然而单独是无活性的，因此它们与互补多肽片段的结合就立即形成活性检测子蛋白。在这种情况下，互补只有在其它的核酸/蛋白相互作用的支持下才发生。在一个实施方案中，描述了高亲和力和高特异性的蛋白/RNA 适配子的相互作用。适配子被表达为靶 RNA 中的识别标签，同时合成核酸结合基序作为融合到裂解的检测子蛋白片段上的两个无活性片段。核酸结合基序片段与适配子的相互作用将检测子蛋白片段聚集在一起，这形成了细胞内的重组装检测子蛋白的酶活性或荧光。

在一个实施方案中，本发明描述的方法可以用于核酸的实时检测。描述了包含编码检测子分子的核酸序列的检测子构建体，检测子分子包含一种接合了核酸结合基序的检测子蛋白的多肽片段以及至少一种接合到核酸结合基序上的检测子蛋白的其它多肽片段。此外，分析方法利用了报道子构建体，它编码包含目标核苷酸和核酸结合序列的报道子分子。核酸结合序列被检测子分子的核酸结合基序所识别。这种识别使得检测子蛋白能

够被重构并且立即地被实时检测。检测子蛋白在没有靶核酸时没有活性，因此它是相当灵敏的并且显示出低背景。此外，对检测子蛋白的裂解进行设计以使得在靶核酸序列存在时，一旦识别活性就可以立即被检测到。

在本发明的一个实施方案中，核酸是 RNA。选自 mRNA 和 miRNA 的任何 RNA 都可以被检测。备选地，核酸是 DNA。

本发明的分析方法被描述为核酸结合基序。在一个实施方案中，基序可以以一种方式与检测子蛋白结合，该方式中核酸结合基序被裂解成多肽片段，一个片段与检测子蛋白的一个片段接合，余下的核酸多肽片段与检测子蛋白的一个或更多个其它片段接合。在该实施方案中，基序的结合是协同的，每个片段都必须存在并结合到一种核酸结合序列上。

在备选实施方案中，与检测子蛋白的一个多肽片段结合的核酸结合基序可以是与其它核酸结合基序无关的全长基序，其它核酸结合基序是与检测子蛋白的一个或更多个其它片段接合的。例如，核酸结合蛋白可以是小的多结构域核酸结合蛋白，因此每个结构域都与检测子蛋白的一个或更多个片段结合。在该实施方案中，核酸结合基序与报道子分子中它们的同源核酸序列（或天然核酸）的结合是独立的。

细胞中的构建体表达

在本发明的一个实施方案中，检测子和报道子分子在体内细胞中表达。为实现这一点，可以根据本领域技术人员已知的任何适宜的方法，例如转化或转染将编码检测子和报道子分子的核酸序列插入细胞中。在一个实施方案中，构建体可以例如位于单个核酸序列上。在备选实施方案中，例如，它们可以处在两个构建体上（一个包含报道子构建体，并且一个包含检测子构建体）。这样的构建体可以被共转化或共转染到细胞中。不受理论的限制，有人认为共转化/共转染的构建体可以在转化/转染期间重新结合，结果是两者被整合到细胞基因组 DNA 的相同位点上。

备选地，检测子和报道子构建体可以单独地或者组合地在细胞中稳定表达。产生高水平重组蛋白的稳定克隆是在用编码所需目标基因和显性遗传标记的表达载体转染细胞后获得的。稳定整合到各种类型细胞中的方法是本领域技术人员已知的。

在一个实施方案中，核酸可以编码裂解的多肽片段（包括检测子蛋白和核酸结合基序），它们重构形成活性检测子蛋白。在一个这样的实施方案中，裂解的多肽片段可以，例如由内核糖体进入位点（IRES）连接。IRES

能够由两个或更多个分离的基因形成单一的转录物，由于转录物上额外核糖体进入位点的存在，该转录物可以被翻译成相应的分离的产物。在备选实施方案中，裂解的多肽片段和/或检测子蛋白和/或核酸结合基序可以由一条核酸编码，在每个核酸序列之间有可裂解的位点以使所需的多肽分离。作为非限制性例子，核酸可以包含编码活性检测子蛋白的序列，该检测子蛋白包含至少两种核酸结合基序，其中编码检测子蛋白的序列包含可裂解的位点从而能够产生检测子蛋白的裂解多肽片段，每个片段都包含核酸结合基序。在这样的实施方案中，通过本领域技术人员已知的任何手段可裂解的位点能够将裂解的多肽片段分离，例如但不限于酶切、化学裂解；热裂解、酸裂解、辐射裂解、光裂解等等。此外，检测子构建体和报道子构建体也可以在单个构建体中编码，该构建体具有由 IRES 连接的分离的组分。

将核酸序列引入细胞的方法是本领域技术人员已知的，包括例如可以将检测子和报道子构建体通过多种手段，包括载体、病毒载体和非病毒方法引入到细胞中。非病毒方法包括但不限于融合、电穿孔、基因枪（biolistic）、转染、脂转染、原生质体融合、磷酸钙转染、显微注射、压力进入、裸 DNA 等等，或者任何本领域技术人员已知的任何其它方法。

术语“载体”与“质粒”可交换使用，是指一种核酸分子，其能够将另一种核酸输送到与其连接的核酸分子上。载体被可操作地连接到基因和/或核酸序列上，而该载体能够引导所述基因和/或核酸序列表达，这样的载体在本发明中被称为“表达载体”。通常，用于重组 DNA 技术的表达载体经常是“质粒”形式的，质粒是指环状的双链 DNA 环，它们作为载体形式时不与染色体结合。其它的表达载体可以用于本发明的不同实施方案中，例如但不限于质粒、附加体（episome）、噬菌体或病毒载体，这些载体可以整合到宿主基因组中或者在特定细胞中自主复制。也可以使用本领域技术人员已知的具有等价功能的其它形式的表达载体。表达载体包括用于稳定表达或瞬时表达编码 DNA 的表达载体。

在一个实施方案中，检测子和报道子构建体可以通过同源重组引入，其中希望将构建体整合到特定位点上。例如，我们可以缺失和/或用本发明的核酸构建体取代相同位点或者其它位点上的内源基因。对于同源重组来说，将核酸克隆到特定载体中，包括但不限于Ω或O载体，参见例如 Thomas 和 Capecchi, cell,（1987），51;503-512, Mansour 等，nature,（1988）

336;348-352; 以及 Joyner 等, nature（1989）338;153-156。

包含用于原核或真核表达的有用元件，如细菌或酵母复制起点、可选择和可扩增的标记、启动子/增强子元件以及可用于制备核酸构建体的储备并用于进行转染的哺乳动物表达控制元件等等的载体都是本领域中公知的，并且许多可以商购。

在一些实施方案中，可以使用病毒载体来引入检测子构建体和核酸构建体的核酸序列，并驱动它们的表达。病毒载体是指使用病毒或病毒相关的载体作为核酸构建体进入细胞的载体。可以将构建体整合并包装到不复制的缺陷型病毒基因组中，象腺病毒、腺相关的病毒（Adeno-associated virus，AAV）或单纯疱疹病毒（HSV）或者其它病毒，包括反转录病毒和慢病毒（lentiviral）载体，用于感染或转导到细胞中。载体可以被引入到细胞基因组中，也可以不被引入。如果需要，构建体可以包括用于转染的病毒序列。备选地，可以将构建体引入到能够进行附加型复制的载体中，例如 EPV 和 EBV 载体。

对于具有作为核酸结合部分的多肽的检测子构建体的裂解多肽片段来说，整个裂解的多肽片段和核酸结合部分分子可以由单个构建体编码，包括多肽部分、接头和核酸结合部分多肽。该构建体可以在细胞中表达或者被显微注射到细胞中。也可以将这些构建体用于体外检测目标核酸。

适配子

适配子是相当短的 RNA 或 DNA 寡核苷酸，它结合配体，是采用被称为 SELEX（指数富集配体的系统进化，systematic evolution of ligands by exponential enrichment）的筛选程序在体外分离到的（Tuerk & Gold, 1990; Ellington & Szostak, 1990）。因为筛选程序是由配体结合所驱动的，所以适配子以高亲和性结合它们的配体并折叠成最适于配体结合的二级结构（Herman & Patel, 2000）。在这个方面，适配子类似于抗体，都是选择性结合来自复杂的化学或生物学混合物的相应配体。

适配子在本发明的方法中是特别有用的。例如，可以构建检测子构建体以便核酸结合基序包含适配子结合序列，而报道子构建体包含适配子。相反地，检测子构建体可以包含适配子，而报道子构建体可以包含适配子结合序列。

这样，在本发明的一个实施方案中，使用编码包含适配子的 RNA 寡核苷酸的载体从而能够表达出不带有能实质性干扰适配子功能的侧翼序

列的适配子，这样适配子在细胞中表达后就能够起作用。在一个这样的实施方案中，这是采用含有两侧是两种自体裂解核酶的适配子的 RNA 寡核苷酸来实现的。如美国专利公开文本 20050222400 所证实的，当寡核苷酸被表达时，一旦自体裂解核酶裂解就会释放出适配子，释放的适配子具有短的侧翼序列，那是自体裂解核酶的残留，这不大可能干扰适配子序列的正确折叠，这与现有技术的方法是相反的。不受限于任何特定机理，我们相信这些短的侧翼序列不会实质性干扰适配子的功能。

由于预计 RNA 寡核苷酸在一些实施方案中是根据 DNA 模板的转录物，因此 RNA 寡核苷酸可以是能够通过转录形成的任何长度。其它元件如额外的适配子、核酶、编码区、前导序列等等也可以是 RNA 寡核苷酸的一部分，只要那些元件不会实质性干扰适配子或自体裂解核酶的功能。

这样的实施方案中的适配子寡核苷酸可以是现在已知的或者以后开发的任何有用的的适配子。适配子可以以细胞中的靶点作为靶向。这样的靶点的一种类型是细胞组分（即天然组分）。然而，在备选实施方案中，适配子以细胞中的工程化组分为靶向。预计能够被适配子结合有效影响的任何细胞组分都可以作为本发明适配子的潜在靶点。这些细胞组分的非限制性例子包括蛋白，如酶、结构蛋白、离子通道蛋白、电子传递蛋白、核酶组分、脂蛋白和转录因子；蛋白多糖；糖蛋白；多糖；核酸；脂；以及小分子如甾体。

设计和合成适配子和适配子结合序列的方法是本领域技术人员已知的。

适配子-核酸结合蛋白配对。

适配子-蛋白配对的任何组合都可以用于本发明。例如，在一个实施方案中，检测子分子的核酸结合蛋白可以是转录因子或者参与 RNA 剪接的蛋白。检测子分子的例子是 eIF4A，它结合到报道子分子的 58 个核苷酸（nt）的适配子（Oguro 等, 2003）（参见实施例 1-8）。在其它实施方案中，核酸结合蛋白和适配子配对包括二元适配子-肽配对，如表 1 和实施例 11 所示。此外，包括在本发明中的适配子和蛋白配对的其它例子，特别是 RNA-蛋白伙伴（partner），包括但不限于；（i）MS2 外壳蛋白-RNA 茎环（Sawata &Taira, 2003; Valegard 等, 1997）；（ii）TAR-Tat BIV-1 茎环（Roy 等, 1990; Comolli 等, 1998）；（iii）表现出转录活化剂作用的 G3/C3 茎环的 3 个重复序列（Jarrell & Ptashne, 2003）；（iv）适配子的 3 个重复序列

（Yamamoto 等，2000）；（v）eIF4A-87-nt 长的适配子，它可被修饰成 58 个碱基，最佳 Kd = 27 nM（Oguro 等，2003）。

核酸结合基序。

核酸结合基序可以是能够连接到另一分子，如多肽上的任何多肽或蛋白或肽分子，它们能够以相近的距离结合到靶核酸上。

本发明的其它实施方案提供的核酸结合部分是多肽。多肽可以是针对靶核酸具有高亲和力的任何多肽。在该实施方案中，靶核酸可以是双链的、三链的或者单链的 DNA 或 RNA。在一些实施方案中，多肽是小于 100 个氨基酸的肽或全长的蛋白。多肽对靶核酸的亲和力可以处于低至纳摩尔、高至皮摩尔的范围内。多肽可以包括含有锌指的多肽，要么是天然的，要么是通过推理或筛选方法设计的。锌指的例子包括 Zif 2g8、Sp1、Gfi-1 的指 5、YY1 的指 3、CF2II 的指 4 和 6 以及 TTK 的指 2（PNAS（2000）97: 1495-1500; J Biol Chem(20010 276（21）: 29466-78; Nucl Acids Res（2001）29（24）: 4920-9; Nucl Acid Res（2001）29（11）: 2427-36）。其它多肽包括通过体外筛选获得的多肽，它们结合到特定的核酸序列上。这种适配子的例子包括血小板衍生的生长因子（PDGF）（Nat Biotech（2002）20:473-77）和凝血酶（Nature（1992）355: 564-6）。其它多肽是体外与 DNA 三聚体结合的多肽；例子包括异核核糖核酸颗粒（hnRNP）蛋白的成员，如 hnRNP K、L、E1、A2/B1 和 I（Nucl Acids Res（2001）29（11）: 2427-36）。

每个裂解的多肽核酸基序多肽的核酸结合部分可以是能够结合到靶核酸上的任何分子。在一些实施方案中，核酸结合部分包括核酸、核酸类似物和多肽。在一个实施方案中，核酸结合部分是寡核苷酸。活化的裂解多肽片段的特定配对的核酸结合部分可以是相同种类的分子，例如寡核苷酸，或者它们可以是不同的，例如包含活性蛋白的配对中的一种裂解多肽可以含有寡核苷酸核酸结合部分，而配对中的其它成员可以含有多肽核酸结合部分。

检测子蛋白

检测子蛋白（它是检测子分子的一部分）的裂解多肽片段可以是距离相近时进行结合形成活性蛋白的任何多肽，它可以通过能够识别组装的活性蛋白而不是单独的多肽的任何手段进行检测。例如，两个多肽可以重新结合形成具有酶活性的蛋白，形成具有显色活性或荧光活性蛋白，或者形成由抗体识别的蛋白。此外，对它们进行设计以使它们处于活性状态并准

备（即准备就绪状态）重构成活性蛋白，从而使体外和体内使用蛋白互补时通常可见的任何滞后时间最小化。

在一个实施方案中，检测子分子的活化裂解多肽片段是荧光蛋白。在这样的实施方案中，一种裂解的荧光蛋白片段是活性的，因为该片段之一含有成熟的预先形成的发色团，其准备就绪，一旦与它的同源活化的裂解荧光片段发生互补就立即发出荧光。

在本发明的优选实施方案中，检测子蛋白是荧光蛋白，作为非限制性的例子是增强型绿色荧光蛋白（EGFP）。备选地，荧光蛋白可以是黄色荧光蛋白（YFP）、增强型黄色荧光蛋白（EYFP）、蓝色荧光蛋白（BFP）、增强型蓝色荧光蛋白（EBFP）、青色荧光蛋白（CFP）、增强型青色荧光蛋白（ECFP）或红色荧光蛋白（dsRED）或者它们的任何其它天然的或基因工程化的片段，其中重构的荧光蛋白的一个片段含有成熟的预先形成的发色团。所有上文提到的荧光蛋白及形成发荧光的荧光蛋白的它们的片段都包括在内用于本发明。那些本领域技术人员已知的荧光蛋白以及它们的片段和基因工程化的蛋白也包括在内。

备选地，检测子蛋白是能够用于信号放大的酶。酶可以是例如β-内酰胺酶、β-半乳糖苷酶、β-葡萄糖苷酶、β-葡萄糖醛酸酶、氯霉素乙酰转移酶。在特定实施方案中，检测子酶是β-内酰胺酶。

在本发明的另一个实施方案中，对检测子蛋白片段进行设计以使得它们在重构时被立即活化。此外，对它们进行设计使得它们准备重构，以便将使用荧光蛋白时通常可见的任何滞后时间最小化。

在一些实施方案中，与含有成熟发色团的裂解荧光片段相结合的同源非荧光多肽片段可以包含超过一个的活性非荧光片段。这种活化的非荧光多肽通常是由一种荧光蛋白的编码核苷酸序列在适当位点上裂解，并且独立地表达每个核苷酸序列片段而形成的。活化的裂解荧光蛋白片段可以单独表达或者与一个或更多个蛋白融合伙伴融合表达。

在本发明的一个实施方案中，重构的活性蛋白包含活化的裂解 EGFP 片段，其中第一片段是 EGFP 的 N 末端片段，其包含从第 1 位氨基酸到大约第 158 位氨基酸的氨基酸连续延伸。可以将 C 末端半胱氨酸加到该片段上以辅助各种核酸结合基序表达后的接合。另一个活化的裂解 EGFP 片段是从大约第 159 位氨基酸到第 239 位氨基酸的氨基酸连续延伸。也可加入 N 末端半胱氨酸。第 1 位氨基酸是指 EGFP 的第一个氨基酸。第 239 位氨

基酸是指 GFP 的最后一个氨基酸。所有的残基都是根据野生型 A. victoria GFP（GenBank 登录号 no. M62653; SEQ ID NO 7）的编号进行编号的，编号也适用于同源序列中相当的位置。这样，当用截短的 GFP（与野生型 GFP 相比）或者用带有额外氨基酸的 GFP 进行研究时，编号必须相应地改变。

在备选实施方案中，重组装的荧光蛋白可以包含来自不同的且光谱学截然不同的荧光蛋白的活化裂解荧光片段。重构的活性荧光蛋白可能具有截然不同的和/或独特的光谱特性，这取决于用于互补的活化裂解荧光片段。例如，通过将用于多色生物分子荧光互补（多色 BiFC）的不同荧光蛋白片段进行重构，已经实现了多色荧光互补（参见 Hu 等，Nature Biotechnology, 2003; 21; 539-545; Kerppola, 2006, 7;449-456, Hu 等，蛋白-蛋白相互作用（P. Adams 和 E. Golemis 编辑），冷泉港实验室出版社. 2005, 将其完整引入本发明作为参考）。采用多种荧光蛋白的活化裂解荧光片段以获得多色实时荧光，这包括在本发明的应用中，其中一个片段含有预先形成的成熟发色团。

在一个实施方案中，荧光蛋白可以通过流式细胞仪、荧光板读数仪、荧光计、显微镜、荧光共振能量转移（FRET）、通过肉眼或通过本领域技术人员已知的其它方法进行检测。在备选实施方案中，使用荧光活化细胞分选仪（FACS）或随时间拍摄显微镜通过流式细胞术检测荧光。

在另一个例子中，标记是产生显色产物/荧光产物的酶。

在本发明的另一个实施方案中，活化的裂解多肽片段以近距离结合形成组装的活性酶，这可以使用酶性活分析进行检测。优选地，通过显色反应或荧光反应来检测酶活性。在一个优选实施方案中，酶是二氢叶酸还原酶（DHFR）或 β-内酰胺酶。

在另一个实施方案中，酶是二氢叶酸还原酶（DHFR）。例如 Michnick 等已经开发了包括 DHFR 的 N 末端和 C 末端片段的“蛋白互补分析”，这些片段单独时没有任何酶活性，但是距离接近时就形成了功能性酶。参见例如美国专利 Nos. 6,428,951、6,294,330 和 6,270,964，它们被引入本发明作为参考。检测 DHFR 活性的方法是现有技术中公知的，包括显色方法和荧光方法。

在备选实施方案中，可以使用其它的裂解多肽。例如，催化底物转变成可检测的产物的酶。几种用于重组装裂解多肽的这样的系统包括但不限

于β-半乳糖苷酶（Rossi 等, 1997, PNAS, 94;8405-8410）；二氢叶酸还原酶（DHFR）（Pelletier 等, PNAS, 1998; 95;12141-12146）；TEM-1 β-内酰胺酶（LAC）（Galarneau at al, Nat. Biotech. 2002; 20;619-622）和萤火虫荧光素酶（Ray 等, PNAS, 2002, 99;3105-3110 以及 Paulmurugan 等, 2002; PNAS, 99;15608-15613）的重组装。例如，已经使用裂解的β-内酰胺酶来检测双链 DNA（参见 Ooi 等, Biochemistry, 2006; 45;3620-3525）。使用活化的裂解多肽片段用于实时信号检测包括在本发明的应用中，其中片段具有充分折叠的成熟构象，从而能够在互补发生时快速检测信号。

在一个实施方案中，可以使用能够放大信号的检测子蛋白，例如带有细胞可渗透的荧光前体β-内酰胺酶底物，头孢菌素β-内酰胺（CCF2）的β-内酰胺酶系统（Zlokarnik 等, 1998; Galarneau 等, 2002; Wehrman 等, 2002），它能够放大信号约 100 倍（Zlokarnik 等, 1998; Galarneau 等, 2002）。

在本发明的另一个实施方案中，活化的裂解多肽片段的结合能够形成含有不连续表位的组装蛋白，它可以通过使用特异性识别组装蛋白上的不连续表位，但不识别任一个体多肽上存在的部分表位的抗体来检测。这样的不连续表位的一个例子见于 HIV 的 gp120 中。这些以及其它这样的衍生物都能够很容易地由本领域普通技术人员根据公知技术来制备，对不通过自身蛋白片段来识别组装蛋白的抗体进行筛选。

在本发明的另一实施方案中，活化的裂解多肽可以是相互作用形成组装蛋白的分子。例如，分子可以是蛋白片段，或者二聚体或多聚体的亚基。

核酸序列和编码目标裂解多肽片段的密码子可以进行优化，例如将密码子转换成所需系统中优选使用的密码子。例如在哺乳动物中。用于在非哺乳动物细胞中蛋白表达的最佳密码子也是现有技术中已知的，当宿主细胞是非哺乳动物细胞（例如在昆虫细胞中）时可以使用。

本发明的活化裂解多肽可以包含所需的任何额外修饰。例如，在一个实施方案中，活化的裂解多肽也可以包含柔性接头，它被偶联到核酸结合部分上。

可以通过本领域技术人员已知的任何手段将检测子蛋白接合到核酸结合基序上。在一个非限制性的例子中，通过基因融合将检测子蛋白接合到核酸结合基序上。本发明使用的术语“接合”是指将两个或更多个蛋白连接在一起形成一个整体。可以通过接头、化学修饰、肽接头、化学接头、共价键或非共价键、或蛋白融合或者通过本领域技术人员已知的任何手段

将蛋白连接在一起。连接可以是永久的或者可逆的。在一些实施方案中，可以包括几个接头以便利用接合物中的每个接头和每个蛋白的所需要的性质。考虑单独使用或与其它接头一起使用的柔性接头以及提高接合物溶解性的接头都包括在本发明中。可以通过表达编码接头的 DNA 将肽接头与接合物中的一个或多个蛋白相连接。接头可以是酸可裂解的、光可裂解的以及热敏感性接头。

术语“融合蛋白”是指两个或更多个蛋白的重组蛋白。融合蛋白的产生可以通过例如将编码一种蛋白的核酸序列连接到编码另一种蛋白的核酸上，以便它们构成单个开放读码框，能够在细胞中翻译成包括所有预期蛋白的单个多肽。蛋白排列的顺序可以变化。作为非限制性的例子，将编码裂解检测子多肽片段的核酸序列按照读码框融合到编码核酸结合基序的核酸的 5’或 3’末端。这样，一旦基因表达，裂解的检测子蛋白就被功能性表达，包含融合到 N 末端或 C 末端的核酸结合基序。对检测子和/或荧光蛋白进行修饰以便检测子和/或荧光蛋白的功能性实质上不受核酸结合基序与检测子和/或荧光蛋白的融合的影响。在一个实施方案中，将 EGFP 的裂解多肽片段按照读码框在羧基末端与裂解的 eIF4A 片段或 RNA 结合肽相融合。

术语“接头”是指通过除了形成融合蛋白之外的手段连接两个或更多个蛋白的任何手段。接头可以是共价接头或者非共价接头。共价接头的例子包括共价键或者共价连接到一个或多个待连接的蛋白上的接头部分。接头也可以是非共价键，例如通过金属中心，如铂原子形成的有机金属键。对于共价连接来说，可以使用如氨基的各种官能团，包括碳酸衍生物，醚，酯，包括有机酯和无机酯，氨基，尿烷，脲等等。例如，为了进行连接，可以通过氧化、羟基化、取代、还原等等对核酸结合基序和/或荧光蛋白进行修饰以提供偶联位点。不会明显削减核酸结合基序和/或检测子蛋白，例如荧光蛋白的功能的修饰将是所期望的。

表达载体

为了进行用于本发明的检测子和报道子构建体的重组表达，可以使用常规技术来完成载体的构建，这些技术对于本领域普通技术人员来说不需要详细解释。然而为了回顾一下，普通技术人员可能希望查阅 Maniatis 等的分子克隆：实验指南，冷泉港实验室（NY 1982）。

简要地说，编码检测子分子和报道子分子的核酸序列的构建采用了标

准连接技术。为了分析从而确认产生的核酸构建体中的序列是正确的，可以采用连接混合物来转染/转导宿主细胞，并且可以通过适合的抗生素抗性选择成功地发生遗传改变的细胞。

由转染/转导的细胞制备载体，通过限制性进行分析和/或通过例如 Messing 等（Nucleic Acids Res., 9: 309-,1981）的方法、Maxam 等（Methods in Enzymology, 65: 499, 1980）的方法、Sanger 的双脱氧法或者本领域技术人员已知的其它合适方法进行测序。

在转录、翻译或者翻译后水平上控制基因的表达。转录起始是基因表达中的早期事件和关键事件。这取决于启动子和增强子序列，并受到与这些序列相互作用的特定细胞因子的影响。许多原核基因的转录单元由启动子以及某些情况下的增强子或调节元件组成（Banerji 等, Cell 27: 299（1981）;Corden 等, Science 209: 1406（1980）; 以及 Breathnach 和 Chambon, Ann. Rev. Biochem. 50: 349（1981））。对于反转录病毒来说，控制元件参与了位于长末端重复序列（LTR）中的反转录病毒基因组的复制（Weiss 等编辑，肿瘤病毒分子生物学: RNA 肿瘤病毒，冷泉港实验室（NY 1982））。莫洛尼（Moloney）氏小鼠白血病病毒（MLV）和劳氏（Rous）肉瘤病毒（RSV）的 LTR 含有启动子和增强子序列（Jolly 等, Nucleic Acids Res. 11: 1855（1983）; Capecchi 等, 增强子和真核基因表达, Gulzman 和 Shenk 编辑, 第 101-102 页, 冷泉港实验室（NY 1991））。其它有效的启动子包括源自巨细胞病毒（CMV）的启动子和其它野生型病毒启动子。

除非另有指明，本发明的实施采用细胞生物学、细胞培养、分子生物学、转基因生物学、微生物学、重组 DNA 和免疫学的常规技术，它们都在现有技术的范围内。这些技术在文献中有详细说明。参见例如 Sambrook, Fritsch 和 Maniatis 编辑的分子克隆实验指南第二版（冷泉港实验室出版社: 1989）; DNA 克隆，第 I 卷和第 II 卷（D.N. Glover 编辑，1985）; 寡核苷酸合成（MJ. Gait 编辑，1984）; Mullis er al 的美国专利 No: 4,683,195; 核酸杂交（B.D. Hames & S.J. Higgins 编辑，1984）; 转录和翻译（B.D. Hames & S.J. Higgins 编辑，1984）; 动物细胞培养（R.I. Freshley, Alan R. Liss, Inc., 1987）; 固定化细胞和酶（IRL Press，1986）; B. Perbal，分子克隆实用指南（1984）; 论文，酶学方法（Academic Press, Inc., N.Y.）; 用于哺乳动物细胞的基因转移载体（J.H. Miller 和 M.P. Calos 编辑，1987，冷泉港实验室; 酶学方法，第 154 和 155 卷（Wu 等编辑），细胞中的免疫化学方法和

分子生物学（Mayer和 Walker 编辑，Academic Press, London, 1987）；实验免疫学手册，第 I-IV 卷（D.M. Weir 和 C.C. Blackwell 编辑，1986）；操作小鼠胚胎（冷泉港实验室出版社：冷泉港，N.Y.，1986）。

关于这些蛋白接合物或融合蛋白以及调节序列的设计、组装、引入质粒中以及构建体的转化方面的其它背景信息和对实验人员的一般性指导可以在以下公开的国际专利申请中获得：WO94/18317；WO95/02684；WO95/24419 和 WO 96/41865，其内容引入本发明作为参考。

许多非病毒启动子的启动子和增强子区域也已经被记载（Schmidt 等，Nature 314: 285（1985）; Rossi 和 decrombrugghe, Proc. Natl. Acad. Sci. USA 84: 5590-5594（1987））。用于维持和增加休眠细胞中的转基因表达的方法包括使用包括I型胶原的胶原（1和2）（Prockop和Kivirikko, N. Eng. J. Med. 311 : 376（1984）；Smith 和 Niles, Biochem. 19: 1820（1980）；de Wet 等, J. Biol. Chem., 258: 14385 1983））、SV40 和 LTR 启动子在内的启动子。

根据本发明的一个实施方案，启动子是选自泛素启动子（ubiquitin promoter）、CMV 启动子、JeT 启动子、SV40 启动子、延伸因子 1α 启动子（EF1-α）、鸡β-肌动蛋白、PGK、MT-1（金属硫蛋白）的组成型启动子。

本发明还包括诱导型/抑制型启动子。这些构建体的非限制性例子是"Tet-On"、"Tet-Off"以及雷帕霉素诱导的启动子，它们包括在本发明中。

除了使用病毒启动子和非病毒启动子来驱动转基因表达之外，还可以使用增强子序列来提高转基因表达的水平。增强子不但能够提高其天然基因的转录活性，而且能够提高一些外源基因的转录活性（Armelor, Proc. Natl. Acad. Sci. USA 70 : 2702（1973））。例如在本发明中，使用带有胶原启动子 2（I）的胶原增强子序列来增加转基因表达。此外，也可以使用 SV40 病毒中发现的增强子元件来增加转基因表达。该增强子序列由 72 个碱基对重复序列组成，如 Grass 等的 Proc. Natl. Acad. Sci. USA 78: 943（1981）; Benoist 和 Chambon, Nature 290: 304（1981），以及 Fromm 和 Berg, J. Mol. Appl. Genetics, 1 : 457（1982）所述，它们全部引入本发明作为参考。当这个重复序列与各种启动子串联存在时，它能够增加许多不同病毒基因和细胞基因的转录（Moreau 等, Nucleic Acids Res. 9 : 6047（1981））。

使用细胞因子调节启动子活性也可以增加转基因表达，从而获得长期稳定的表达。已经报道过几种调节胶原 2（I）和 LTR 启动子的转基因表达

的细胞因子（Chua 等, connective Tissue Res., 25: 161-170（1990）; Elias 等, Annals N. Y. Acad. Sci., 580 : 233- 244（1990）; Seliger 等, J. Immunol. 141: 2138-2144（1988）以及 Seliger 等, J. Virology 62: 619-621（1988））。例如，转化生长因子（TOF）、白细胞介素（IL）-I 和干扰素（INF）下调由各种启动子如 LTR 驱动的转基因表达。肿瘤坏死因子（TNF）和 TGF 1 上调这些表达，它们可以被用于控制由启动子驱动的转基因表达。证明可能有用的其它细胞因子包括碱性成纤维细胞生长因子（bFGF）和表皮生长因子（EGF）。

在一个实施方案中，也可以使用带有胶原增强子序列（Coll（E））的胶原启动子来增加转基因表达，这是通过进一步抑制对载体的任何免疫应答实现的，尽管脑处于免疫保护状态，所述免疫应答还可能在经处理的脑中产生。

在另一个实施方案中，载体可以进一步包含序列，如编码 Cre 重组酶蛋白的序列以及 LoxP 序列。确保检测子和/或报道子构建体的瞬时表达的另一种方式是通过使用 Cre-LoxP 系统，该系统造成插入的 DNA 序列的部分切除，这要么是在将 Cre 重组酶给予细胞时（Daewoong 等, Nature Biotechnology 19:929-933），要么是通过把编码重组酶的基因引入病毒构建体中（Pluck, Int J Exp Path, 77:269-278）时发生。在病毒构建体中引入重组酶基因，以及 LoxP 位点和结构基因（目前情况下是神经营养因子）经常造成大约为期五天的结构基因表达。

转基因生物

在本发明的一个实施方案中，包括了表达用于体内实时检测核酸的报道子和检测子构建体的转基因生物。转基因生物或动物可以在其所有的细胞中携带转基因，或者是在一些细胞而不是全部细胞中携带转基因，即嵌合体动物。转基因可以作为单个转基因进行整合，或者串联整合，例如头对头串联，或头对尾串联，或尾对尾串联，或者作为多拷贝整合。二体、三体或多体转基因动物可以优选包含至少两个或更多个转基因。在优选实施方案中，动物包含检测子构建体，该构建体包含可操作地连接到核酸结合基序上的 EGFP 基因的两半以及编码核酸结合序列的转基因和目标基因。

当使用编码蛋白的一种或多种基因作为转基因时，希望将基因可操作地连接到适当的调节元件上，这将使得转基因能够表达。调节元件，例如

启动子，增强子，（例如诱导型或组成型），或者多聚腺苷酸信号都是现有技术中公知的。调节序列可以是内源性调节序列，即来自同一种类动物的调节序列，该调节序列作为转基因被引入所述种类动物中。调节序列也可以是用作转基因的基因天然调节序列。

本发明描述的转基因构建体可以包括 DNA 序列下游的 3'非翻译区。这样的区域能够稳定表达系统的 RNA 转录物，从而增加出自表达系统的所需蛋白的产量。在 3'非翻译区中，用于本发明构建体的是提供 polyA 信号的序列。这些序列可以源自例如 SV40 小 t 抗原或者现有技术中公知的其它 3'非翻译序列。3'非翻译区的长度不是关键的，但是它的 polyA 转录物的稳定效应对于稳定表达序列的 RNA 看来是重要的。

转基因构建体也可以包括启动子和编码信号序列的 DNA 序列之间的 5'非翻译区。这样的非翻译区可以来自获取启动子的相同控制区或者可以来自不同的基因，例如它们可以源自其它合成的、半合成的或者天然的来源。

反义核酸也可以用于本发明的转基因构建体。例如，可以将反义多核苷酸序列（与 DNA 编码链互补的）引入到细胞中以减少“正常”基因的表达。该方法利用例如反义核酸、核酶或者三链体试剂（triplex agents）来阻断特定 mRNA 的转录或翻译，要么是通过用反义核酸或三链体试剂掩蔽 mRNA，要么是通过用核酶将它裂解。备选地，所述方法包括给予模拟基因产物的作用或效应的试剂或者阻断基因作用的试剂。使用反义方法改变体外基因翻译是现有技术中公知的（参见例如 Marcus-Sekura, 172 ANAL. BIOCHEM. 289-95, 1988）。

本发明描述的转基因构建体可以被插入到任何合适的质粒、噬菌体或病毒载体中用于扩增，从而可以使用现有技术中已知的方法进行增殖，如 Maniatis 等（分子克隆：实验指南，冷泉港，N. Y., 1989）中描述的方法。可以将构建体制备成较大质粒的一部分，这使得能够按照现有技术已知的有效方式克隆和筛选构建体。构建体可以位于质粒上适宜的限制性位点之间，以便它们可以从余下质粒序列上容易地分离，从而被引入到所需哺乳动物中。

本发明的方法也涉及转基因动物或细胞的生产，其中通过使用反转录病毒载体将外源 DNA 引入到卵母细胞中。通过利用病毒感染过程，反转录病毒载体可以有效地用于将基因转移到宿主细胞中（Kim 等, 4 ANIM.

BIOTECHNOL. 53-69, 1993; Kim 等, 35 MOL. REPROD. DEV. 105-13, 1993; Haskell 和 Bowen, 40 MOL. REPROD. DEV. 386-90, 1995; Chan 等, 95 PROC. NATL. ACAD. SCI. USA 14028-33, 1998; Krimpenfort 等, 1991; Bowen 等, 50 BIOL. REPROD. 664-68, 1994; Tada 等, 1 TRANSGENICS 535-40, 1995)。

应用

本发明的方法可以被用于本领域技术人员已知的多种应用。本发明的一个重要的实施方案是用于体内实时检测 RNA 的报道子分析方法。在一个这样的实施方案中，目标细胞稳定地表达检测子分子，该检测子分子在没有靶 RNA 时不产生来自检测子蛋白的信号，因而不发出荧光或者显示出活性。细胞也可表达诱导型报道子分子，该报道子分子包含检测子分子的核酸结合基序的结合位点以及目标基因，当目标基因在细胞中表达时，检测子分子将结合靶 RNA 并促进检测子蛋白重构以及产生活性蛋白和信号。这样实现了体内实时检测 RNA 表达，在图 10-11 和 18-19 以及实施例 9-11 中进行了举例说明。

在备选实施方案中，目标细胞可以表达与目标启动子可操作地连接的报道子构建体。当启动子在细胞中被某些刺激物活化时，针对靶核酸序列的结合位点就成为可用的了，检测子分子的核酸结合蛋白的结合有助于检测子多肽片段的结合，以及能被立即检测的活性检测子蛋白的形成。该实施方案能够在体内实时研究启动子的功能。例如，可以通过转染质粒并分析任意的原始检测子构建体活性来实时研究基因的调节，其中质粒表达驱动 RNA 结合位点的启动子。在相关实施方案中，将细胞 i）与一种或多种化合物接触；或者 ii）经历各种可能改变启动子活性的状况，从而增加、减少或者不引起 RNA 转录的变化。通过检测子构建体立即检测出这种变化。

备选地，本发明的方法提供了体内检测靶核酸存在的方法，即使是在没有检测子构建体的情况下。在这样的实施方案中，例如核酸，例如 DNA 和/或 RNA，包含与检测子分子的核酸结合基序结合的靶核酸序列。在这样的实施方案中，根据检测子分子的核酸组分的设计，能够在体内检测天然的或未修饰的核酸。

在类似的实施方案中，本发明的方法可以使用"tet-on"或"tet-off"系统。例如，可以制备稳定表达本发明的检测子构建体的目标细胞。此外，细胞

也可以稳定表达 tet-on 或 tet-off 可调节报道子构建体，而它能够控制针对检测子分子核酸结合基序的靶结合位点和目标基因的转录。例如，在 tet-off 系统中，细胞将立即表达目标基因，报道子构建体将被活化并可以被立即检测到。当加入 tet 时，报道子构建体（包含目标基因和靶核酸结合位点）的转录就停止了，只有已经转录的转录物才能被实时分析。在该 tet-off 的例子中，我们能够通过检测减弱的报道子荧光或活性从而实时分析体内 RNA 的稳定性，如它的半衰期。

在备选实施方案中，使用“te-on”构建体能够在活细胞中稳定表达目标基因，这样将 tet 加入系统中就启动了报道子构建体的转录。然后，本发明的检测子分子能够结合它的靶 RNA 并且可以被立即检测到。在该实施方案中，可以按照进行实验的人员所要求的那样实时研究 RNA 定位。

在本发明的另一个实施方案中，报道子分析是在生物体内进行的。例如但没有限制，可以制备动物、非人动物、小鼠、斑马鱼、线虫或蛙来表达检测子构建体和目标基因，目标基因上带有将结合到检测子构建体上的 RNA 结合位点。两种构建体在生物体中的表达实现了通过从第 1 天的胚胎直到成年期进行取样，从而在整个发育期间分析目标转录物。在一个实施方案中，生物为哺乳动物。在一个实施方案中，哺乳动物是小鼠，在另一个实施方案中，生物是转基因生物，例如但不限于转基因小鼠。

备选地，本发明的组合物和方法用于类似于原位杂交，例如荧光原位杂交（FISH）的体内实时检测 RNA 的方法。在这样的实施方案中，将表达包含目标 RNA 的报道子构建体的细胞进行透化，该目标 RNA 带有将结合到检测子分子的核酸结合基序上的 RNA 结合序列，将本发明的检测子构建体给予细胞。如果表达了目标 RNA，检测子构建体将与 RNA 结合序列结合，就能够确定 RNA 的定位和丰度。该系统相对于常规原位杂交技术的优点在于能够检测低丰度的 RNA(归因于采用酶报道子的检测子构建体所得到的信号放大）以及能够迅速完成分析且没有放射性。立即进行检测子构建体的检测是相当值得期待的。目前，进行原位杂交的常规方法（例如带有放射性的）在知道实验是否成功之前可能甚至耗时数周到数月。此外，常规原位杂交中使用的 RNA 探针很难制备，这是因为 RNA 容易降解以及要应用放射性。在本发明的方法中，无需放射性或 RNA 探针。检测子构建体很容易由例如细菌来制备，这是本领域技术人员已知的。这种分析方法类似于免疫组织化学或免疫细胞化学技术，它们使用抗体来体外检

测蛋白。这里是检测 RNA。

此外，本发明的组合物和方法可以用于体内实时检测 DNA 的实施方案。例如，本发明的检测子构建体可以包含检测子蛋白（荧光的或酶的），它被裂解成至少两个失活的多肽片段，每个片段都与 DNA 结合基序结合。靶核酸的存在将多肽片段聚集在一起，在这种情况下靶核酸是 DNA，多肽片段就形成了具有完全活化的检测子蛋白，能够被立即检测到。例如，本发明的方法能够实时检测基因组中各种位点的复制。与检测子构建体中的检测子蛋白相结合的 DNA 结合基序是特异性针对 DNA 的各个区域的，可以在体内实时检测那些位点的复制。

类似地，我们能够使用本发明的方法和组合物在体内实时检测染色体。例如，检测子构建体可以包含与端粒末端转移酶结合蛋白相结合的并在细胞中表达的检测子蛋白。这样，在细胞的生命期中都能够实时检测端粒末端转移酶。本发明记载的组合物和方法的优点在于通过提供端粒末端转移酶结合蛋白提高了特异性，该端粒末端转移酶被平分成两半，每一半都与检测子蛋白的一半结合。端粒末端转移酶结合蛋白的两半协同结合到单一结合位点上，这就确保了靶向识别之前几乎没有检测子活性。

在本发明的另一个实施方案中，公开了实时检测 PCR 产物的方法。使用本发明的方法和组合物，PCR 反应产物的检测具有比目前可用的方法更高的特异性并且是实时的。在该实施方案的实施例中，PCR 引物的设计使得反应的扩增产物中掺入能够被检测子构建体上存在的核酸结合基序特异性识别的适配子。在扩增进行时，能够实时检测检测子蛋白。本发明提供了相对于目前使用的技术，如 SYBR 绿的优点，这是因为检测子构建体是特异性针对 PCR 产物的，将不会检测模板 DNA。使用该实施方案，能够通过在进行 PCR 实验之前将 RNA 转变成 cDNA 来实时检测 RNA。在相关实施方案中，检测子分子的核酸结合基序组分促进了 PCR 产物存在时的裂解检测子分子的重组装，提供了用于体内免疫 PCR 的新方法。而且，在另一个实施方案中，检测子分子的核酸结合组分能够促进免疫 RCA（滚环扩增）方法中核酸存在时裂解检测子分子的重组装，从而产生信号，导致了体内的高信号放大。

在本发明的另一实施方案中，公开了用于检测个体中多态性、突变或异常基因表达的方法。例如，本发明的分析方法可以被用于检测个体的 DNA 和/或 RNA 以诊断某些疾病、障碍或者易患这些疾病和障碍的体质。

因此，如果检测子构建体能够设计成带有核酸结合基序的话，本发明的方法就能够用于各种疾病和障碍的诊断和预后，其中核酸结合基序是特异性针对于患有疾病或障碍的个体的基因组中可能存在的某些序列的。

在相关的实施方案中，本发明能够用于实时检测个体中的基因突变、多态性或畸变。作为说明性例子，适配子能够标记或者暂时结合到特定的突变或多态性位点，这可以通过本发明的裂解多肽分子进行检测，该多肽分子被设计成便于检测子分子的核酸结合基序识别所结合的适配子，而该适配子结合到包含我们试图检测的特定突变、多态性或畸变的目标基因上。备选地，可以使用分子库（pool），由此可以检测多种突变、多态性或畸变。如果核酸结合基序识别所结合的靶核酸，如与基因置换和/或改变相关的适配子，就形成了检测子蛋白的蛋白互补并能够用于敏感性检测，这归因于立即产生了信号和/或荧光。

在一个重要的实施方案中，分子可以被用于实时检测病原体。在一个实施方案中，本发明的分子可以被用于检测病原体核酸序列和/或核酸序列中的畸变的存在，该畸变是病原体和/或病原体核酸存在的结果。病原体可以是病毒感染、真菌感染、细菌感染、寄生虫感染以及其它感染性疾病。病毒可以选自 1 型单纯疱疹病毒、2 型单纯疱疹病毒、巨细胞病毒、EB 病毒、水痘-带状疱疹病毒、人疱疹病毒 6、人疱疹病毒 7、人疱疹病毒 8、天花病毒、水泡性口膜炎病毒、甲型肝炎病毒、乙型肝炎病毒、丙型肝炎病毒、丁型肝炎病毒、戊型肝炎病毒、鼻病毒、冠状病毒、甲型流感病毒、乙型流感病毒、麻疹病毒、多瘤病毒、人类乳头状瘤病毒、呼吸道合胞病毒、腺病毒、柯萨奇病毒、登革热病毒、腮腺炎病毒、脊髓灰质炎病毒、狂犬病毒、劳氏肉瘤病毒（Rous sarcoma virus）、黄热病毒、埃博拉病毒、马尔堡病毒（Marburg virus）、拉沙热病毒、东方马脑炎病毒、日本脑炎病毒、圣路易脑炎病毒（St. Louis Encephalitis virus）、墨累谷脑炎病毒（Murray Valley fever virus）、西尼罗河病毒、裂谷热病毒（Rift Valley fever virus）、轮状病毒 A 型、轮状病毒 B 型、轮状病毒 C 型、辛德毕斯病毒（Sindbis virus）、猴免疫缺陷病毒、人 T 细胞白血病病毒 1 型、汉坦病毒、风疹病毒、猴免疫缺陷病毒、人免疫缺陷病毒 1 型以及人免疫缺陷病毒 2 型。

靶核酸也可以用于食品、饮料、水、药品、个人护理品、乳制品或者环境样品中细菌和真核生物的检测。优选的饮料包括汽水、瓶装水、果汁、啤酒、葡萄酒或者酒类产品。对于制造或贮藏食品、饮料、水、药品、个

人护理品、乳制品或者环境样品时所使用的原材料、设备、产品或者工艺的分析而言，所开发的分析方法将是特别有用的。

在本发明的另一个相关实施方案中，活化的裂解多肽片段经组装形成含有不连续表位的组装检测子蛋白，它可以通过使用特异性识别组装蛋白上的不连续表位但是不识别任何个体多肽上存在的局部表位的抗体来检测。一个这样的不连续表位的例子发现于 HIV 的 gp120 中。这些以及其它这样的衍生物都能够很容易地由本领域普通技术人员根据公知技术进行制备，并且对不通过自身的任何蛋白片段来识别组装蛋白的抗体进行筛选。

靶核酸可以是人源的。靶核酸可以是 DNA 或 RNA。靶核酸可以是游离在溶液中的或者固定到固体载体上的。

在一个实施方案中，靶核酸是特异性针对遗传疾病的或者是特异性针对易患遗传疾病的体质的。所述疾病可以是，例如 β-地中海贫血，镰状细胞贫血或者凝血因子 V Leiden，遗传疾病象囊性纤维化病（CF），癌症相关靶点如 p53 和 p10，或者乳腺癌易感的 BRC-1 和 BRC-2。在另一个实施方案中，可以研究与亲子鉴定、身份确认或者犯罪调查相关的分离的染色体 DNA。

在另一个实施方案中，本发明提供了适用于本发明的方法，特别是体内检测 RNA 的试剂盒。在一个实施方案中，试剂盒包含检测子构建体和诱导型报道子构建体。在一个实施方案中，报道子构建体和检测子构建体处在分开的核酸分子上。在另一个实施方案中，它们都由双顺反子核酸分子编码。在相关的实施方案中，裂解多肽片段由 IRES 位点分开。在另一个实施方案中，多肽片段被编码成一种包含可裂解位点的核酸，而且它可以被裂解形成分开的多肽片段，这是通过本领域技术人员已知的手段实现的，例如但不限于酶切；化学裂解、光裂解、酸裂解、热裂解等等。在另一个实施方案中，可以通过加入实验人员的目标基因来修饰报道子分子，在另一个实施方案中，报道子分子的表达是与 tet-on 或 tet-off 启动子可操作地相关联的。在另一个变型的实施方案中，实验人员可以修饰包含靶核酸序列的报道子构建体以使其被实验人员的目标启动子调节。在每个这样实施方案中，试剂盒都包含用于检测子分子中的检测子蛋白的裂解多肽的选项，也包含用于检测来自检测子蛋白的信号的说明书和试剂。在一些实施方案中，试剂盒也包含适用于捕获和/或检测样品中靶核酸的存在或数量

的试剂。用于检测靶核酸的存在和/或数量的试剂可以包括酶活性试剂或者特异性针对组装蛋白的抗体。抗体可被标记。

蛋白互补和酶报道子的组合能够实质性降低背景和放大信号。这形成了能够分析低丰度、中等丰度和高丰度核酸的体内 RNA 检测技术。

在另一个实施方案中，本发明提供了适于体内检测靶核酸存在和/或数量的试剂盒。试剂盒至少包含偶联到第一分子上的第一探针和偶联到第二分子上的第二探针，其中探针能够结合到靶核酸的杂交序列上。优选地，探针是在小瓶中的。试剂盒也包含适用于捕获和/或检测样品中靶核酸存在或数量的试剂。用于检测靶核酸存在和/或数量的试剂可以包括酶学活性试剂或者特异性针对组装蛋白的抗体。抗体可以被标记。这样的试剂盒可以任选地包括进行 RCA 反应、免疫 RCA、免疫 PCR 所需的试剂，如 DNA 聚合酶、DNA 聚合酶辅助因子以及脱氧核糖核酸-5'-三磷酸。任选地，试剂盒也可以包括各种多核苷酸分子，DNA 或 RNA 连接酶，限制性内切酶，反转录酶，末端转移酶，各种缓冲液和试剂，抗体以及 RNase 和核酸酶活性的抑制剂。这些组分是在容器如小瓶中的。试剂盒也可以包括进行阳性和阴性对照反应所必需的试剂以及说明书。用于给定反应的试剂的最佳量可以由技术人员在目前公开内容的帮助下很容易地确定。

在另一个实施方案中，本发明的方法可以被用于多个核酸靶点同时进行蛋白互补。例如，具有相关的不同核酸结合基序的互补裂解多肽片段的蛋白互补。一种靶核酸的存在将有助于一种活性裂解多肽片段配对的蛋白互补，而另一种靶点的存在将有助于另一对活化的裂解多肽片段的蛋白互补，这产生了不同的活性蛋白和可检测的信号。在一个这样的实施方案中，多个靶核酸能够被同时检测。在备选实施方案中，可以通过实时蛋白互补来监测靶核酸如 RNA 和 DNA 的同时检测。

在相关实施方案中，多个蛋白互补使用来自不同荧光蛋白的裂解荧光蛋白片段。在相关实施方案中，本发明的方法使得能够实时检测和鉴别各种其它推定的但并不相同的核酸靶点之中的特定靶核酸（参见 Hu 等，Nature Biotechnology, 2003;21;539-545; Kerppola, 2006, 7;449-456, Hu 等，蛋白-蛋白互补(P. Adams 和 E. Golemis 编辑)，冷泉港实验室出版社，2005，将它们全部引入本发明作为参考）。

定义

除非另有说明，本发明使用的下列术语和短语应具有下述含义：

本发明使用的术语“检测子构建体”应包括编码检测子蛋白和核酸结合基序的核酸序列，其中包含检测子分子。

本发明使用的术语“报道子构建体”应包括目标核酸（例如 DNA 或 RNA）以及靶核酸序列（例如 RNA 适配子），其中包含报道子分子。

术语“真核生物”或“真核生物”应包括动物、植物和原生生物界中的所有生物，包括原生动物、真菌、酵母、绿藻、单细胞植物、多细胞植物，以及所有动物，脊椎动物和无脊椎动物。该术语不包括细菌或病毒。“真核细胞”应包括单个的“真核细胞”以及多个“真核细胞”，包含源自真核生物的细胞。

术语“脊椎动物”应包括单个“脊椎动物”以及多个“脊椎动物”，包括哺乳动物和鸟类，以及鱼、爬虫类和两栖动物。

术语“哺乳动物”应包括单个“哺乳动物”和多个“哺乳动物”，包括但不限于人类；灵长类如猿、猴、猩猩和黑猩猩；犬科动物如狗和狼；猫科动物如猫、狮子和虎；马科动物如马、驴和斑马，食用动物如奶牛、猪和绵羊；有蹄类动物如鹿和长颈鹿；啮齿类动物如小鼠、大鼠、仓鼠和豚鼠；以及熊。优选地，哺乳动物是人类受试者。

本发明使用的术语“体内”应包括存在于生物体中的活细胞。当谈及生物体外存在的活细胞时，使用典型的术语“离体”。活细胞可以是形成任何生物体的任何细胞，包括多细胞生物体和单细胞生物体，如酵母和细菌。

本发明使用的术语“侵入性”就检测活细胞中的核酸而言，是指使用侵入性方法检测核酸的方法，例如脂转染和显微注射。因此术语“非侵入性”就是指不采用这样的侵入性方法来检测活细胞中的核酸的方法。

术语“组织培养”或“细胞培养”或“培养（culture）”或培养（culturing）”是指在能够维持细胞结构、维持细胞功能、进一步分化或者所有这三者的条件下体外维持植物或动物组织或细胞，或使植物或动物组织或细胞在体外生长。“原代组织细胞”是直接取自组织的细胞，即在生物体内行使同样功能的相同种类的细胞群。当将组织细胞接种到培养板上时，用蛋白水解酶胰蛋白酶处理这些组织细胞，例如将它们分离形成长成或维持细胞结构的单个原代组织细胞。由组织培养物中的原代细胞增殖产生的细胞培养物被称为“次代细胞培养物”。大多数次代细胞进行有限次数的分裂，然后死亡。然而少数次代细胞可以度过这个“危机期”，此后它们能够无限

增殖形成连续的“细胞系”。培养细胞的液体培养基在本发明中被称为“培养基（culture medium）”或“培养基（culture media）”。在细胞培养期间，所需要的分子例如免疫球蛋白分子已经分泌到培养基中，在本发明中该培养基被称为“条件培养基”。

细胞也可以不是指特定的受试者细胞，而是指这种细胞因为某些修饰或环境影响，例如分化而形成的后代或者潜在后代，这样后代可能实际上不同于亲本细胞，但是其仍然包括在本发明的范围中。

本发明中使用的细胞可以是例如体外或离体培养的细胞，正如本发明实施例中所显示的。例如，细胞在培养基中体外培养，并且可以向培养基中加入外界刺激。备选地，对于离体培养的细胞来说，可以从受试者获取细胞，其中受试者是健康的和/或感染疾病的。作为非限制性的例子，可以通过活组织检查或本领域技术人员已知的其它外科手术手段获取细胞。

术语“IRES”是指内核糖体进入位点（见 Kozak（1991）J. Biol. Chem. 266'19867-70），是编码共有核糖体（consensus ribosome）结合位点的序列，能够被起始密码子和/或调节序列的 5’端或者其它编码基因或标记基因的核酸序列的 5’端即刻插入，以增强下游核酸序列的表达。进行这种修饰的期望或需要可以根据经验确定。

术语“多核苷酸”是指核酸或构建体中存在的任意一种或多种核酸区段或者核酸分子，例如 DNA 或 RNA 片段。“编码目标基因的多核苷酸”是指包含这种多肽的编码区的多核苷酸。此外，多核苷酸可以编码调节序列，如启动子或转录终止子，或者可以编码多肽或蛋白的特定元件，如分泌信号肽或功能结构域。

“核苷酸”是聚合核酸如 DNA 或 RNA 中的单体单元，其由三种不同的子部分或部分组成：糖、磷酸和碱基（Blackburn, M., 1996）。当作为双链体的一部分时，核苷酸也被称为“碱基”或“碱基对”。最常见的天然形成的碱基是腺嘌呤（A）、鸟嘌呤（G）、胞嘧啶（C）和胸腺嘧啶（T），它们具有以序列特异性方式将一种核酸链结合到另一核酸链上的氢键功能。“核苷”是指没有磷酸基的核苷酸。在 DNA 和 RNA 中，核苷单体通过磷酸二酯键连接，其中本发明使用的术语“磷酸二酯键”是指磷酸二酯键，包括其磷酸类似物的键，包括结合的反离子，例如 IT'、NW、Na'等等。

“多核苷酸”或“寡核苷酸”是指天然核苷酸单体或其类似物的线性聚合物，包括双链和单链的脱氧核糖核酸“DNA”、核糖核酸“RNA”等

等。换句话说，“寡核苷酸”是脱氧核糖核酸链或核糖核酸链，它们分别是包含脱氧核糖核酸（DNA）和核糖核酸（RNA）的结构单元。典型的多核苷酸处在几个单体单元，例如 8-40 个到几千个单体单元的长度范围内。只要 DNA 多核苷酸由字母序列来表示，例如“ATGCCTG”，那么应理解为核苷酸是从左到右为 5’→3’方向，“A”表示脱氧腺苷，“C”代表脱氧胞苷，“G”代表脱氧鸟苷，“T”代表胸苷，除非另有说明。

“沃森/克里克碱基配对”和“沃森/克里克互补”是指特定核苷酸及其类似物的配对模式，其通过氢键结合到一起，例如 A 与 T 和 U 配对，G 与 C 配对。特定碱基配对的行为是“杂交（hybridization）”或“杂交（hybridizing）”。当两种或更多种核酸或核酸类似物的互补链进行碱基配对就形成了杂种。

“检测”是指根据检测标记的性质来检测、观察或测定构建体。

术语“修饰的碱基”是指 DNA 和 RNA 中发现的天然形成的碱基 AGC、T、U 的配对碱基衍生物。

术语“启动子”是指足以引导转录的最小核苷酸序列。本发明也包括足以提供由细胞类型特异性、组织特异性控制的或者被外部信号或试剂诱导的启动子依赖的基因表达的那些启动子元件；这些元件可以位于天然基因的 5’或 3’区，或者位于内含子中。术语“诱导型启动子”是指 RNA 聚合酶结合速率和转录起始速率可以被外部刺激物调节的启动子。术语“组成型启动子”是指 RNA 聚合酶结合速率和转录起始速率不变且相对独立于外部刺激物的启动子。术语“瞬时调节启动子”是指在发育期间的特定时间调节 RNA 聚合酶结合速率和转录起始速率的启动子。所有这些启动子类型都包括在本发明中。

本发明使用的“启动子”或“启动子区”或“启动子元件”在本发明中可以互换使用，是指对可操作地连接到其上的核酸序列转录进行控制的核酸序列区段，典型的是 DNA 或 RNA 或其类似物，但不限于此。启动子区包括足以进行 RNA 聚合酶识别、结合和转录起始的特定序列。启动子区的这个部分被称为启动子。此外，启动子区包括调节 RNA 聚合酶的这种识别、结合和转录起始活性的序列。这些序列可以是顺式作用的，或者可以是响应反式作用因子的。根据调节性质，启动子可以是组成型或调节型的。

术语“组成型活性启动子”是指在给定细胞中始终表达的基因的启动

子。用于哺乳动物细胞的示例性启动子包括巨细胞病毒（CMV）启动子，用于原核细胞的包括噬菌体 T7 和 T3 启动子等等。

术语“可操作地连接”或“可操作地结合”在本发明中可以互换使用，是指核酸序列与核苷酸调节序列，如启动子、增强子、转录和翻译终止位点以及其它信号序列之间的功能性关系。例如，核酸序列，典型的是 DNA，与调节序列或启动子区可操作的连接是指 DNA 和调节序列或启动子之间的物理性和功能性关系，其通过特异性识别、结合和转录 DNA 的 RNA 聚合酶使得这种 DNA 的转录从调节序列或启动子开始。为了优化表达和/或体外转录，为了在表达核酸或 DNA 的细胞类型中表达核酸或 DNA 而修饰调节序列可能是必需的。对这种修饰的预期或需要可以根据经验确定。

术语“接合”是指两个或更多个蛋白结合在一起形成一个整体的连接。蛋白质可以通过接头、化学修饰、肽接头、化学接头、共价键或非共价键、或蛋白融合或者通过本领域技术人员已知的任何手段连接在一起。连接可以是永久的或者可逆的。在一些实施方案中，可以包括几个接头以便利用接合物中的每个接头和每个蛋白的所需性质。打算单独使用或与其它接头一起使用的柔性接头和提高接合物溶解性的接头都被包括在本发明中。可以通过表达编码接头的 DNA 而将肽接头与接合物中的一个或多个蛋白相连接。接头可以是酸可裂解的、光可裂解的以及热敏感性接头。在一些实施方案中，“接合”或“接合的”也包括共价的、离子的或疏水的相互作用，由此将分子的各个部分结合在一起并保持接近。

在本发明中，“适配子”是指人工设计的来强力地和特异性地与特定靶蛋白相结合的核酸配体。“调节适配子”是指适配子的核心部分经过设计的适配子，这样其具有低 Tm 值，并且被分成两个短的寡核苷酸链，只有在存在靶蛋白时才结合形成双链。

实施例

实施例 1

由核酸相互作用促成的蛋白互补方法。

为了表明由核酸相互作用促成的快速蛋白互补的可行性，在体外进行了几个实验。图 20a、20b 显示了本发明讨论的核酸相互作用是如何进行的。在这些实验中，选择增强型绿色荧光蛋白（EGFP）作为标记蛋白是出于许多原因的。首先，根据特征性荧光的存在很容易确定 EGFP 的活性。

其次，GFP 家族的荧光蛋白已经被成功地用作几种蛋白-蛋白互补研究中的标记蛋白或检测子蛋白，例如 Ozawa 等, 2000, Ozawa 等, 2001 a,b; Ghosh 等, 2000; Hu 等, 2002; Hu & Kerppola, 2003; Remy & Michnick, 2004; Magliery 等, 2005，它们都论证了成功裂解 EGFP 的方案。

这些研究显示，EGFP 的第 153-161 位氨基酸之间的环是用于裂解的便利部位：蛋白插入该环中不影响体内蛋白的折叠和发色团的形成（Ozawa 等, 2000, Ozawa 等, 2001 a,b; Ghosh 等, 2000; Hu 等, 2002; Hu & Kerppola, 2003; Remy & Michnick, 2004; Magliery 等, 2005）。而且重要的是，当两个片段在 E.coli 中共表达时，没有显示出荧光团或发色团的自发性自组装（Gosh 等, 2000; Hu 等, 2002; Magliery 等, 2005）。这样，为了证实核酸的相互作用可以引导 GFP 的两个裂解多肽片段结合以形成活性荧光蛋白，我们设计了两个与互补寡核苷酸接合的 EGFP 多肽片段，以证实偶联的寡核苷酸形成的双链 DNA 促进了 EGFP 片段的重新结合，进而形成活性 EGFP 蛋白并发出强烈荧光。

在这些研究中，在第 158 位将 EGFP 遗传裂解成两个部分，称为 α 片段和 β 片段。为了与寡核苷酸偶联，分别在 C 末端和 N 末端采用额外的半胱氨酸残基对 α-EGFP 片段和 β-EGFP 片段进行设计。它们在 E.coli 中表达为带有蛋白内含子的 C 末端融合蛋白，应用蛋白内含子自裂解化学（NE Biolabs）进行纯化。使用便利的 Cys 靶向的生物素-HPDP 化学（Pierce）将纯化的蛋白在体外定量生物素化到接合的寡核苷酸上。首先采用链霉亲和素标记生物素化的蛋白，接着用 5'或 3'末端带有生物素的寡核苷酸标记。这样，很容易通过四聚链霉亲和素分子将在不同末端上带有生物素的互补的 21 nt 长的寡核苷酸添加到 EGFP 的 α 片段和 β 片段上（图 20a）。当这些分子嵌合体以等摩尔量组合时，发出强烈的荧光，表明具有类似于 EGFP 发射光谱的荧光团的形成（图 20b）。在对照实验中，将与链霉亲和素融合但不带有互补寡核苷酸的生物素化的 EGFP 的 α 片段和 β 片段混合在一起，没有检测到明显的荧光（图 20b）。荧光恢复随实验不同而稍微不同，最大恢复接近完整折叠 EGF 的荧光的 100％。因此，在大多数情况下，通过可控的蛋白互补能够恢复 100％的 GFP 荧光。

由带有附加的寡核苷酸的裂解 EGFP 片段重构获得的重构 EGFP 的荧光光谱稍微不同于天然 EGFP 的荧光光谱。首先，与天然 EGFP（488/ 507 nm, Zimmer, 2002）相比，重构蛋白的发射和激发最大值（490/524）发生

了红移（图 20b）。这可以由重构蛋白β-折叠结构的构象中存在可能的微小差异来解释。其次，一旦加入 Mg2+离子，重构复合物与 EGFP 之间的差异就明显了。一旦加入 2 mM Mg2+，天然蛋白的荧光降低大约 30％，而重构 EGFP 的荧光首先增加了大约 30％，然后逐渐降低（图 21）。Mg2+的不同影响可以由附加到重构 EGFP 上的双链 DNA 的存在来解释。实际上，DNA 双链象夹钳（pincer grip）一样将 EGFP 的两个片段结合在一起，因为加入 Mg2+离子使得 DNA 双链更稳定，这可能导致重组装的 EGFP 稳定性更高，结果导致初始的荧光增加。相反地，对于天然荧光蛋白来说，已知发色团附近存在二价金属阳离子会淬灭荧光（Richmond 等，2000; Zimmer, 2002）。因此，重构 EGFP 的荧光逐渐降低是两个过程的结果：由 Mg2+离子存在使 DNA 双链稳定性更高导致的复合物稳定化以及荧光淬灭。

实施例 2

检测用于可控的快速蛋白互补的蛋白。

在实施例 1 中，蛋白互补中的标记蛋白或检测子蛋白是荧光蛋白，即增强型绿色荧光蛋白（EGFP），它是水母 Aequorea victoria GFP 的双突变体（F64L, S65T）。EGFP 和其它相关的荧光蛋白在第 154-158 位残基处的裂解已经被成功地用于几种被设计用来测试体内蛋白/蛋白相互作用的研究中（Ghosh 等，2000; Hu & Kerppola, 2003; Remy & Michnick 2004; Magliery 等, 2005）。这些研究表明由片段在体内重组装成活性 EGFP 并不是自发发生的，而是需要额外的蛋白/蛋白相互作用（Maglieri 等, 2005）。此外，已经表明 EGFP 重组装对相互作用的蛋白的大小具有相当大的耐受性（Ghosh 等, 2000; Hu & Kerppola, 2003; Remy & Michnick 2004; Magliery 等, 2005）。

为了将蛋白互补用于 RNA 定位的研究，特别重要的是片段重组装的动力学要足够快。在我们的体外初步研究中，非常迅速的活化裂解 EGFP 片段的互补引起了荧光增加（图 22A）。这可能是因为 EGFP 的α片段含有预先形成的成熟发色团，因而处于活化构象并且一旦与互补的 EGFPβ片段结合就能够立即形成活性荧光蛋白。这种 EGFP 重组装的快速动力学的独特性质使得监测靶 RNA 在体内的快速运动成为可能。为此，我们由一个质粒表达了互补系统的两个蛋白组分，并且在 2-3 小时后诱导带有适

配子标签的靶 RNA 的合成。这样，我们就能够监测特定 RNA 的运动。

使用另一种标记蛋白，例如黄色荧光蛋白的变体 Venus（F64L，M153T，V163 A，S175G）也是可能的，因为它具有比 EGFP 更快的荧光团成熟速度（K_{OX}= 8×10^{-3} s^{-1} 比 1.5×10^{-4} s^{-1}）并具有更高的量子产率（Nagai 等. 2002）。此外，该蛋白对光漂白更稳定，它的光谱特性使得能够更好地与细胞蛋白的自发荧光相区别。

在另一个实施例中，标记是产生生色产物/荧光产物的酶。

为了开发具有信号放大的蛋白互补系统，我们利用了带有促荧光的β-内酰胺酶底物，头孢菌素β-内酰胺（CCF2）的β-内酰胺酶系统（Zlokarnik 等, 1998; Galarneau 等, 2002; Wehrman 等, 2002），该底物是细胞可渗透的。据显示，这种酶系统能够获得大约 100 倍的信号放大（Zlokarnik 等, 1998; Galarneau 等, 2002）。定量体内分析显示，16 小时后低至每个 Jurkat 细胞 50 个β-内酰胺酶分子能够在背景之上被检测到。通过减少暴露于底物的时间，定量测得 20,000 个β-内酰胺酶分子。这种灵敏度实质上比用荧光蛋白能够获得的灵敏度更高，后者的检测需要 10^5 到 10^6 个靶分子/细胞在背景自发荧光之上（Zlokarnik 等, 1998）。

因此，具有酶活性的标记蛋白将产生比荧光蛋白更高的信号。然而，在酶学方法中可能要部分损失空间分辨率，这归因于小分子量显色反应产物的扩散。因此，根据特定实验的需要，可以使用这两种系统（荧光的或酶学的）中的任何一种。

为了证明互补核酸相互作用辅助的蛋白互补的可能性，我们选择增强型绿色荧光蛋白（EGFP）分子作为模型，将其切成两个片段，1-158 和 159-234 aa。使用含有全长 EGFP-1 基因的质粒（Clontech, PaloAlto, CA）经 PCR 获得相应的基因。将两个片段设计成在 C 末端（αCys-片段，1-158）和 N 末端（βCys-片段，159-234 aa）带有额外的 Cys 残基。在 TWIN-1（New England Biolabs, MA）载体中克隆 PCR 产物，作为 Ssp DNAB 蛋白内含子的 C 末端融合蛋白。所有构建体的结构经测序验证。

EGFP 的α片段包含预先形成的发色团。尽管有一些小的光谱差别，通过基于核酸的蛋白互补还是能够恢复至高达 100％的荧光。以 DNA 为模板的 EGFP 重组装的动力学快得惊人，其 t 1/2 为～100 秒（见图 22），接近于 EGFP 由带有成熟发色团的变性蛋白复性的动力学（Reid & Flynn, 1997, Zimmer, 2000）。众所周知，荧光蛋白中典型的发色团成熟需要数小

时（Zimmer, 2000, Fig.22b）。根据快速动力学数据，本发明的方法能够产生含有成熟发色团的EGFPα片段。由于α片段足够大，其含有从N末端起的158个氨基酸（根据EGFP中的总共239个氨基酸），因此它具有形成完全发色团的潜能。

分子模型印证了EGFP的α片段中存在完全形成的成熟发色团。在图23a中，α片段的结构构象的示意图表明α片段折叠成了非常紧密的结构（除了它悬出来的C末端部分），α片段中形成发色团的氨基酸的空间坐标与全长EGFP中的空间坐标是非常接近的（图23b、c）。这也符合荧光蛋白多肽链的特定结构，其在中央螺旋中有剧烈的~800的弯曲，导致了形成发色团的氨基酸T66、Y67和G68很接近(Barondeau等, 2003; Donnely等, 2001)。尽管完全形成的α片段中的发色团缺乏与C末端β片段中的氨基酸的许多重要的接触，这些重要接触存在于全长的EGFP中，而且α片段中的发色团暴露于溶剂中，因此缺乏显示强烈荧光的能力，这种能力是完整蛋白中的生发团的特征。

为了直接确定α片段中促荧光发色团的存在，我们单独分析了β片段的吸收光谱和荧光光谱，并将它们与全长EGFP的光谱相比较（图24）。作为对照，也记录了EGFP的β片段和两种非荧光蛋白（链霉亲和素和胰凝乳蛋白酶原）的光谱。α-和β-EGFP片段的吸收光谱显示在图24b中。正如我们所见的，它们都没有490 nm下的天然EGFP中可见的特征性最大值（图24b）。然而，α片段的特征是在300-400 nm的区域中具有高的多的吸收值，这是β片段（图24b，插图）或者任何非荧光蛋白（未显示）中所没有的。α片段的荧光光谱（图24d）清楚地显示了特征性的激发光谱中360 nm下的最大值以及弱发射光谱中460 nm下的最大值。这些光谱与全长EGFP的光谱（图24c）是完全不同的，然而它们与合成的发色团的光谱以及通过部分蛋白酶解从GFP中分离的含有发色团的短肽的光谱是一致的(图25)(Niwa等, 1996)。要注意的是，暴露于溶剂的发色团(如同它处在α片段中)应该吸收300-40 nm的光而不是450-500 nm的光(如同在全长EGFP中)。在独立的实验中，当裂解的EGFP片段在活的E.coli体内表达时，裂解的EGFP片段产生了不同于完全形成的天然结构EGFP的光谱（图25）。因此以这种方式产生的EGFP的α和β片段是活化形式的，尽管单独的是无荧光的或者失活的，但是它们能够重新结合以形成活性荧光蛋白。

活化的互补EGFP片段的这种重新结合可以是在体外和体内的，核酸相互作用能够促进EGFP片段的互补。荧光恢复或重构的快速动力学是与处于活化状态的片段一致的，只有在与相应的互补片段重构时才是有活性的。特别地，EGFP 的α片段含有成熟发色团，因而是活化状态的，然而由于成熟发色团的淬灭，所以单独的是失活的。含有 81 个 C 末端氨基酸的 EGFP β片段的加入导致了紧密的两部分蛋白结构的形成，将发色团与外界隔离，产生了迅速而强烈的荧光显色（只要两个蛋白片段通过结合的DNA-DNA 或其它核酸互补相互作用而靠在一起）。

这些相当意外的结果对于开发体内快速蛋白互补的方法来说是非常重要的。已经含有成熟发色团的 EGFP 的α片段以及α片段与β片段的重新结合的使用导致了活性 EGFP 蛋白的形成，这是在几秒之内发生的。表达这些裂解的 EGFP 或裂解的荧光片段，其中含有预先形成的成熟发色团的一个片段可以在 RNA 组分诱导之前进行表达。RNA 组分的 RNA 合成的诱导将促进片段的快速蛋白互补并增加荧光。这应当使得我们能够从RNA 表达早期开始跟踪特定 RNA 的命运。

结果显示，带有互补寡核苷酸的两个 EGFP 片段的孵育导致荧光发射光谱（激发最大值 490 nm）增加，其最大值在 524 nm 处，这是 EGFP 光谱的特征。当 EGFP 片段在没有与互补核苷酸混合在一起时，没有发现荧光光谱变化。

实施例 3

检测子蛋白和核酸结合蛋白的接合。

在这个实验中，我们验证了在没有相互作用的适配子序列时，两个切开的蛋白嵌合体的表达将不会产生重构的荧光。这个实验在 E.coli 中进行。通过 PCR 从质粒 pEGFP（Clontech）中获得 EGFP 基因片段。EGFP 基因的裂解发生在与体外实验相同的位置（1-158，159-239）。采用含有全长 eIF4A 的质粒（pGEX-4A1）。通过 PCR 获得 eIF4A 的 F1 和 F2 片段，根据 Oguro 等的方法进行裂解（见图 3）。将带有 SG 接头的两个融合蛋白插入到为两种信使共表达而构建的两个质粒 pETDuet 1 和 pACYDuet-1（Novagen）中。这两个质粒具有不同的复制起点和不同的选择标记。得到的质粒 pMB12 和 pMB13 在 E.coli 菌株 BL21（DE3）（Novagen）中共表达。在 pMB12 中，通过$(Ser\text{-}Gly)_5$肽接头将被称为 A 的 EGFP 的前 158

个氨基酸连接到被称为 F1 的真核起始因子蛋白 4A（eIF-4A）的前 215 个氨基酸上。类似地，在 pMB13 中，将被称为 B 的含有 EGFP 的 C 末端 92 个氨基酸的肽连接到被称为 F2 的 eIF-4A 的另外一半上（图 3）。两种构建体都在 T7 启动子的控制之下，用 1 mM IPTG 进行诱导。通过 FACS 测定荧光水平，它们比得上未经诱导的培养物的荧光水平（图 4）。作为该实验中的阳性对照，我们在载体 pTWIN（NEB）中表达了全长 EGFP，并且作为阴性对照—融合到 eIF4A 的相同 F1 片段的 EGFP 的片段（A-F1，B-F1）（图 4）。

在另一实验中，在被称为 pMP33 的相同质粒中表达 A-EGFP-F1-eIF-4A 和 B-EGFP-F2-eIF-4A 核酸，它们是在 E.coli 中表达的。PMB33 质粒是 pACYCDuet1（Novagen）的衍生物，是表达两种嵌合的裂解 EGFP-eIF-4A 片段蛋白的双顺反子质粒，该蛋白包含标记蛋白（例如 EGFP 或酶）的片段以及 RNA 结合蛋白（例如 MS2 外壳蛋白或 eIF4a）的片段。表达的嵌合的裂解片段在没有靶 RNA 时不会重新组合（图 9a），但是在靶 RNA 存在时就迅速重新结合形成功能性 EGFP 标记蛋白（图 9c）。

因此，RNA 结合蛋白与靶 RNA 的核酸相互作用能够促进体外和体内的快速蛋白互补以及来自裂解 EGFP 片段的荧光以快速动力学恢复。这可以通过所结合的 21-bp 寡核苷酸或者 RNA 结合蛋白与核酸的互补结合来促成。

实施例 4

适配子-核酸结合蛋白配对的选择。

在本发明的一个实施方案中，蛋白互补是基于对标记到目标核酸序列例如 RNA 上的适配子的检测。适配子可以通过接合到检测子蛋白上的核酸结合蛋白来检测。特别地，核酸结合蛋白被分成片段，各个片段接合到检测子蛋白的多肽片段上。几个适配子/蛋白的相互作用的参数对于成功地开发基于核酸的蛋白互补来说是重要的。（i）RNA 适配子和 RNA 结合蛋白之间的结合亲和力应当足够高以便为标记蛋白的重组装提供能量。（ii）适配子的长度不应该太长，否则将适配子引入 RNA 就可能导致其表达和性能的改变。（iii）RNA 结合蛋白应当是单体、小的多肽，优选由两个结构域组成，它们在分开时是失活的，但是当一起表达时能够结合适配子。

eIF4A 是真核起始因子，因此是真核翻译系统中普遍存在的组分。它

在酵母中的天然浓度高（50 nM），比得上另一种丰富的细胞蛋白，肌动蛋白的浓度（Duncan 等, 1987）。它是 29 kD 的小蛋白，其作为 ATP 依赖性 RNA 解旋酶在与其它起始因子（4B、4H、4G、4F）（Kapp & Lorsch, 2004）的复合物中起作用。eIF4A 由两个结构域组成，它的晶体结构类似于哑铃结构：紧密的 N 末端和 C 末端结构域由柔性的 11 aa 长的接头连接（Johnson & McKay, 1999; Benz 等, 1999; Caruthers 等, 2000）。这些结构特征使得这个蛋白成为切开结构域的方便且有用的工具。此外，最近的研究中已经分离了以纳摩尔级亲和力结合 eIF4A 的强结合寡核苷酸或适配子（Oguro 等, 2003）。Oguro 等表明强结合的适配子序列可以短至 56 nt，发现 eIF4A 的切开的结构域不能结合适配子，而全长蛋白能够以高亲和力结合适配子（Oguro 等, 2003）。

由于这些特点，eIF4A 是用于蛋白互补的良好候选物，并被用于我们的系统以驱动蛋白互补。我们使用的 eIF4A 被裂解成两个片段，其中每个片段都融合到 EGFP 检测子蛋白的片段上。在 RNA 适配子存在时，eIF4A 的重组装就会因此将 EGFP 蛋白的两个片段聚集在一起。适配子-eIF4A 相互作用的亲和力与用于监测细胞中 RNA 的 MS2 外壳蛋白/MS2 RNA 相互作用的亲和力处在相同范围内（Bertrand 等, 1998; Beach 等, 1999; Beach & Bloom, 2001; Rook 等, 2000）。

备选地，互补配对的设计可以包括两个不同的 RNA 适配子和附加到标记蛋白片段上的两个不同蛋白（图 5）。这种情况下的蛋白或肽将从病毒蛋白列表中选择，为此已经体外分离了具有高结合亲和力的适配子（见表 1）。

表 1. 具有强结合亲和力的 RNA 适配子/肽配对

肽	肽序列	适配子序列	参考文献
HIV-1 Rex	MPKTRRRPRRSQRKRP（SEQ ID NO.4）	UAGGCGACGGUACGCAAGUACUCUUGCGCCGGCCUA（SEQ ID NO.3）	Baskerville 等 1999
噬菌体 λN	MDAQTRRRERRAEKQAQWKAAN（SEQ ID NO.5）	GGATCCGGGCCCUGAAGAAGGGCCCUUUCCUUU（SEQ ID NO.8）	Baron-Benhanmou 等, 2004

HTLV-1 Rev	TRQARRNRRRRWRERQR （SEQ ID NO.6）	GGCUGGACUCGUACUUCGGUACUGGAGAAACAGCC（SEQID NO.1）	Ye 等. 1996
HIV-1 Tat 蛋白的 CQ 肽	CFLKKGLGISYGRKKRRQRRTAPYDSKNHQDPTPEQ （SEQ ID NO. 7）	GGAGCUUGAUCCCGGAAACGGUCGAUCGCUCC （SEQ ID NO.2）	Yamamot 等, 2000

该备选设计的优点在于两个不同蛋白或肽将不大可能彼此相互作用，这不同于将一种蛋白裂解成两个片段。在裂解蛋白的情况下，切开的蛋白片段仍然可能自发地重组装。选择适配子/蛋白配对的主要参数是复合物的稳定性（Kd）：相互作用越强，标记蛋白互补的概率就越高。我们推断，相似大小的蛋白（或肽）对于活性互补复合物的形成来说是有利的，这是因为组分的对称几何形状。

人工适配子的体外选择能够发现高度特异性结合短肽和只结合同源肽的 RNA 结构。同时，含有那些短肽的更大蛋白与不同适配子交叉反应（Herman & Patel, 2000）。例如，已知 Rex 蛋白与 Rev 应答元件的一部分结合，能够从功能上替代 Rev（Bogerd 等, 1991; Rimsky 等, 1988）；然而，体外筛选导致了不与 Rev 肽相互作用的抗 Rex 特异性适配子的分离（Baskerville 等, 1999）。靶 RNA 也可以被设计为含有结合蛋白的结构域。这种靶 RNA 也将由相应的质粒表达。

表 1 显示了具有强结合亲和力的几种适配子和肽配对，它们可以用作蛋白互补中的 RNA 标签。备选地，可以使用潜在的 RNA-蛋白伙伴的另外的例子，它们包括在本发明中，选自包括但不限于以下列表：

（i）MS2 外壳蛋白-RNA 茎环（Sawata &Taira, 2003; Valegard 等, 1997）。

（ii）TAR-Tat BIV-1 茎环（Roy 等, 1990; Comolli 等, 1998）。

（iii）经显示作为转录激活物起作用的 G3/C3 茎环的 3 个重复序列（Jarrell & Ptashne, 2003）。

（iv）适配子的 3 个重复序列（Yamamoto 等, 2000）。

（v）eIF4A-87-nt 长的适配子，可以被修饰为 58 个碱基，最佳 Kd = 27 nM（Oguro 等, 2003）。

实施例 5

由适配子-核酸结合蛋白相互作用引导的蛋白互补。

在本发明的一个实施方案中，实时蛋白互补方法是基于将适配子标签引入到目标 RNA 中，它将被蛋白复合物识别，该蛋白复合物由两个蛋白嵌合体组成，每个嵌合体含有标记蛋白和 RNA 结合蛋白的片段。为了可靠地使用该系统，适配子标签的引入应当不干扰 RNA 合成和性能。

为了显示使用蛋白互补监测目标 RNA，将适配子加到目标基因中。为了显示适配子不影响真核基因表达，我们在酿酒酵母菌株 YPH500（MAT α，ura3-52，lys2-801 琥珀色，ade2-101 赭石色 trp1-Δ63，his3-Δ200，leu2-Δ1）中的报道子基因的终止密码子后面紧接着引入 64 nt 的寡核苷酸（Oguro 等，2003）。我们使用可诱导系统，其中报道子基因，例如 EGFP 受人工 Tet-应答的 GAL1 启动子（Blake 等，2003）的控制。在该系统中，由 GAL10 启动子表达 TetR 基因以便在存在半乳糖时，TetR 介导 EGFP 的阻遏（图 1A）。这种阻遏可以通过加入诱导物脱水四环素（ATc）来解除，它直接结合到 TetR 上（图 1B）。然后通过流式细胞术分析（FACS），在含有染色体整合的可诱导转录系统的细胞中，对半乳糖和 ATc 存在时的 EGFP 表达定量。我们没有观察到表达未修饰 EGFP 的细胞与在该基因的 3'端插入适配子的细胞之间的荧光差别（图 2）。因此，带有适配子序列的基因的 3'端标记似乎不会影响蛋白水平的基因表达。我们合理地假定 RNA 水平也不会受到该适配子标签存在的影响。

文献分析显示适配子序列可以被引入 5'-或 3'-端非翻译 RNA 区而不干扰 RNA 合成（Hanson 等，2003; Nickens 等，2003），而它能够影响翻译效率，特别是如果被引入的适配子与一些翻译调节物，例如四环素相互作用时（Hanson 等，2003）。我们的初步结果证实了这些数据，显示将针对 eIF4A 的适配子引入 3'端-非翻译区不会影响标记蛋白 EGFP 在酵母中的表达，这表明 EGFP mRNA 水平也不会受影响。

为了使系统更灵活，我们可以通过测试 5'端-非翻译区是否也能用于适配子引入来扩展这些实验。例如，使用 EGFP 作为标记蛋白，根据适配子标签的位置体内测定其表达。适配子的影响将通过使用荧光细胞分选仪（FACS）测定细胞总荧光以及通过 Northern 印迹分析来跟踪，这将直接估计 RNA 的转录水平。适配子标签的最可能的掺入部位仍是 3'端-非翻译区。

引入标签对 RNA 运动和定位的影响也是要考虑的。mRNA 的 3'-UTR 含有决定 mRNA 定位的调节元件(综述,参见 Hesketh, 2004)。例如,ASH1 mRNA 的 3' UTR 编码足够用于 mRNA 在芽中转运和锚定的信号。然而在许多情况下,在信使其它部分中的序列对于正确 RNA 定位来说也是必需的(Chartrand 等, 1999, Gonzalez 等, 1999)。例如,使用不去除信使任何部分的最短的可能适配子标签。在这样做的时候,维持了所有调节 RNA 序列以保留所有的 RNA/蛋白相互作用,并形成 RNA 的完整定位。假设荧光蛋白(或 β-内酰胺酶)以及与 RNA 中与适配子相互作用的小肽是小尺寸的,那么重组装蛋白复合物在 RNA 模板上的定位具有相当大耐受性。尽管如此,仍然在对照实验中通过与特异性针对于给定 mRNA 的探针原位杂交,直接检验了修饰 RNA 的正确定位。

实施例 6

用于二嵌合体蛋白等摩尔共表达的系统的开发,每个嵌合蛋白都含有荧光标记蛋白的片段和 RNA 结合蛋白的片段。

为了成功实施本发明的实时蛋白互补方法,协同合成(coordinated synthesis)了两个多肽蛋白片段,包含活化标记蛋白的每个片段都接合到 RNA 结合蛋白上,它们以等摩尔量形成并准备重组。以前的蛋白互补研究显示,如果由具有不同拷贝数的质粒合成两个融合蛋白,就检测不到 GFP 的荧光恢复(Magliery 等, 2005)。因此,我们开发了进行蛋白片段的等摩尔合成的系统以及用于 RNA 组分独立合成的可诱导系统。

为了等摩尔量表达二嵌合体蛋白,我们使用了能够使几个信使共表达的表达载体(pETDuet 或 pACYCDuet 载体, Novagen)。这些质粒具有两个 T7 启动子和两个多克隆位点,因此每个质粒都能够支持两个信使的表达。同时这些质粒中的复制起点是不同的,使它们能够共表达。我们在一个质粒中放置二嵌合体蛋白,在不同质粒中放置带标签的 RNA; 以便 RNA 表达能够被独立地诱导。将使用独立诱导 RNA 表达的 pBAD 质粒。用抗 EGFP 抗体(Clontech)经 Western 印迹分析检验嵌合蛋白的表达水平,而 RNA 合成通过 Northern 印迹检验。

这些实验在 E.coli 中进行,通过使用 FACS 测定细胞总荧光来监测蛋白互补。作为对照,在没有 RNA 组分时表达二蛋白嵌合物的细胞背景未显示出荧光(图 9A),而在存在适配子时它们的表达则产生了荧光(图 9C),

证实了适配子/RNA 结合蛋白/肽相互作用的特异性。

实施例 7

β-内酰胺酶用作裂解多肽检测子蛋白。

两个独立小组已经使用了 A 类 β-内酰胺酶作为蛋白互补分析中的裂解标记蛋白（Wehrman 等, 2002; Galarneau 等, 2002），这可能是因为这类酶具有几个有吸引力的特征。首先，这些蛋白是相对小的单体酶，它们的晶体结构是公知的（Jelsch 等, 1993）。β-内酰胺酶在细菌和真核细胞中都能表达；重要的是真核细胞不具有内源性内酰胺酶活性。采用体内 β-内酰胺酶做为有力工具的另一重要因素是可以使用由 Tsien 小组开发的细胞透性的荧光底物（Zlokarnik 等, 1997）。

上述两个小组已经表明 β-内酰胺酶可以在 196-198 位残基处切开，在附加的相互作用蛋白的存在下，这些片段能够彼此互补（Wehrman 等，2002; Galarneau 等, 2002）。这个位点（第 196-198 位氨基酸）位于催化中心的相反部位，不显示周期性二级结构。Blau 的小组证明 Asp-Gly-Arg 三肽掺入到 β-内酰胺酶 N 末端片段的 C 末端会增加酶活性高达 10,000 倍（Wehrman 等, 2002）。因此，基于 β-内酰胺酶活性的蛋白互补能够放大信号大约两个数量级，能够在蛋白诱导后数分钟内检测到信号（Wehrman 等, 2002）。

应当强调的是，基于酶活性重构的蛋白互补就灵敏度而言是有前景的；然而由于信号扩散而可能损失空间分辨率。因此，这种分析形式应当优选应用于空间分辨率不是那么重要的时候，以及例如低丰度 RNA 的检测是重要的场合。

在采用 β-内酰胺酶作为原核和真核细胞中的体内 RNA 检测分析的检测子蛋白的方法中，可以将 β-内酰胺酶裂解成活化的多肽片段，正如 Blau 的小组所表明的（Wehrman 等, 2002）。在 E.coli 中克隆两个片段，一个片段由第 24 到 197 位氨基酸残基组成（α片段，缺少周质分泌信号系统从而将酶保留在细胞内），第二个片段由第 198 到 240 位残基组成（β片段）。首先从 pUC19 中 PCR 扩增 α 和 β 片段，并且经 PCR 将三肽 NGR 加到 α 片段的 C 末端。用多肽接头将编码 α 片段的 PCR 产物克隆到 eIF4A 的 F1 片段的下游，同时将 β-内酰胺酶的 β 片段克隆到 eIF4A 的 F2 片段的下游。

对于原核表达来说，将嵌合的裂解 β-内酰胺酶-eIF4A 接合蛋白引入

到 pCDF-duet 载体（Novagen）中，同时由 pACYCD 质粒表达带有 eIF4A 结合适配子的 RNA 靶点。将所得到的质粒在 E.coli 菌株 BL21（DE3）（Novagen）中共表达。

接合到 eIF4A 多肽片段上的β-内酰胺酶在真核细胞中的表达是在酿酒酵母中进行的。用于在酵母中表达含有β-内酰胺酶和 eIF4A 片段[β-内酰胺酶（A）-（F1）和β-内酰胺酶（B）-（F2）]的嵌合蛋白的质粒将是表达载体 pDB20（Becker 等, 1991）的衍生物，它经修饰以表达在其 N 末端与 HA 或 myc 表位标签融合的蛋白。这将使组成型蛋白能够由 ADH1 强启动子进行表达并通过 Western 印迹分析基因的表达。对质粒 pRS4D1（Blake 等 2003）进行修饰以表达在基因的 3'末端带有 64 nt 长的适配子标签的靶 RNA（代替 EGFP）。该质粒在菌株 YPH500 基因组中的整合将使得一旦将半乳糖和脱水四环素加到培养基中时就能够调节标记的靶 RNA 的表达（Blake 等 2003）。

当没有β-内酰胺酶活性时，香豆素在 409 nm 的激发将引起 FRET，并且将在 520nm 处出现发射（绿色荧光）。由于适配子与裂解 eIF4A 片段之间的相互作用而产生了β-内酰胺酶活性，它将打开β-内酰胺环并且荧光素将分裂，在这种情况下将不发生 FRET，将观察到 447 nm 处的发射（蓝色荧光）（Zlokarnik 等 1998）。

实施例 8

通过体外和体内快速蛋白互补检测 RNA 分子。

在下面的实施例中，我们开发了在活细胞中检测 RNA 的强效方法，它具有超出目前检测水平的灵敏度。实时快速动力学蛋白互补与信号放大组合在一起就产生了能够分析中等丰度核酸的检测技术，其中蛋白互补采用了活化的裂解荧光片段（见实施例 2），它能够实质上快速重构活性检测子蛋白并降低背景，其中信号放大是通过使用酶的步骤而引进的。

将具有酶活性或荧光性质的蛋白裂解成两部分（称为α和β亚基）。这些部分作为与核酸结合蛋白的嵌合物在体内表达，该核酸结合蛋白具有适配子结合的结构域。理想的核酸结合蛋白将由两部分组成，它们单独时是失活的但是一起时就是有活性的。在含有可以由核酸结合蛋白识别的基序的 RNA 存在时，检测子蛋白的活性立即恢复。如果蛋白检测子是酶，信号就被放大。如果蛋白是发荧光的，就没有信号放大，但是也没有背景，

因为荧光完全取决于靶 RNA 的存在。

我们制备了用于在 *E.coli* BL21（DE）3 细胞中共表达融合蛋白和含有适配子的 RNA 转录物的构建体（图 9）。我们通过由丝氨酸和甘氨酸残基组成的柔性多肽接头，将 EGFP 片段（第 1-158 位残基）的 C 末端融合到含有第 1-215 位残基的 eIF4A 片段的 N 末端。我们在载体 pACYCDuet-1（Novagen）的第一多克隆位点（MCS）上克隆该融合物，该载体是设计用来由两个 T7 启动子表达两个开放读码框的。类似地，我们通过柔性多肽接头将 EGFP 片段的 C 末端（第 159-238 位残基）融合到含有第 216-406 位氨基酸的 eIF4A 片段的 N 末端。在载体 pACYCDuet-1 的第二 MCS 上克隆该融合物以产生能够以大致等摩尔量表达两个融合蛋白的构建体。载体 pETDuet-1（Novagen）用于表达 360 nt 长的 T7 转录物，它含有串联的两个拷贝的与 eIF4A 相互作用的适配子序列。这个小信使含有 33 nt 的前导序列，接着是两个拷贝的适配子序列以及大约 200 nt 的抗核酸酶的 T7 终止序列。

将表达完整的互补复合物和适宜对照的 *E.coli* 细胞在诱导物，异丙基-β-D-硫代半乳糖苷（IPTG）存在下室温培养，用于蛋白和 RNA 的共表达。当培养物的光密度达到大约 0.5（OD_{600} = 0.5）时，通过荧光活化细胞分选仪（FACS）来分析它们（图 21）。互补融合蛋白与含有适配子的 RNA 转录物的共表达导致平均荧光增加 10-20 倍（图 9C）。然而，在没有含有适配子的转录物时，带有互补融合蛋白的 *E.coli* 细胞未显示高于背景水平的荧光（图 9A）。更重要的是，当细胞表达互补复合物的蛋白组分与未标记的转录物时，我们没有看见荧光明显增加（图 12）。当 T7 转录物有或没有核糖体结合位点时，荧光产量没有差别。这表明信使的翻译不干扰它的检测。

实施例 9

体内 RNA 分子的定量。

表达完整 EGFP 的细胞的荧光比带有互补系统的细胞的荧光高大约 30-50 倍（图 9B 和 9C）。这个差别并不令人惊讶，因为与重组装 EGFP 相比应该有更大量的全长 EGFP，其浓度由 RNA 浓度决定。已知每个 RNA 分子都通过核糖体循环几次，这导致了蛋白与 RNA 的摩尔比大于 1。

接着我们利用 EGFP 标准珠（BD Biosciences Clontech）来评估每个细

胞中重组装 EGFP 分子的绝对平均数，从而评估我们的检测系统的效率。我们发现表达来自具有大约 40 个拷贝数的 pET 载体的 RNA 的细胞产生了 500-600 个重组装 EGFP 分子，其相当于 1-2 μM RNA，如果细胞体积是 1.41 $\times 10^{-15}$L 的话（Lee,等, 2005）。该 RNA 浓度与由具有 50-70 个拷贝/细胞的载体在 *E.coli* 中获得的实验结果很好地关联（Lee,等, 2005）。期望的结果和实验结果之间的一致性表明实际上可以通过重组装蛋白复合物检测到含有适配子的所有 RNA 分子。我们通过增加带有 RNA 适配子的质粒的拷贝数（40-100 个拷贝/细胞）来检验这个假设，观察到细菌细胞显示了与原始构建体相比更高的荧光（图 13）。该结果支持了我们的结论，即我们的互补系统能够检测到几乎所有 RNA 分子。

然后，我们比较了带有重组装 EGFP 的细胞的荧光光谱与表达全长蛋白的细胞的荧光光谱。我们在这些实验中发现重组装复合物的最大激发在 470 nm 处（相对于天然 EGFP 的 490 nm），最大发射在大约 520 nm 处（相对于天然 EGFP 的 508 nm）（图 14）。这个差别可以由天然 EGFP 与重组装 EGFP-RNA 复合物之间蛋白构象的可能的变化来解释。这个结果也支持了细胞中荧光信号归因于形成了核蛋白复合物的想法，因为在我们的体外实验中，我们也发现了由附加双链寡核苷酸重组装的裂解 EGFP 的相似的红移发射光谱（最大 524 nm）（Demidov 等, 2006）。

实施例 10

细菌细胞中荧光在时间和空间上的波动。

使用落射荧光显微镜（epifluorescence microscope）和 B&W 照相机观察表达 RNA 标记系统的细胞（图 9c 和 11）。同时，记录微分干涉相差图像以分析细胞数量和形状。为了达到足够用来形成复合物和产生荧光的时间，将细胞室温培养过夜。因此，当很少或没有分裂发生时，大多数细胞都处于静止期。在某些情况下，我们确实观察到了细胞分裂，在所有情况下新分裂的细胞都不发荧光（图 15）。

融合蛋白和含有适配子的RNA转录物的共表达产生了带荧光的细胞，荧光亮点位于细胞的一极或两极（图 10）。相反，全长 EGFP 的表达产生了带强烈荧光的细胞，荧光均匀分布在整个细胞中（图 9b）。同时，在没有适配子表达时，融合蛋白的表达不会导致明显的荧光产生（图 9a），这证实了我们前面的流式细胞术结果。

时移显微术（Time-lapse microscopy）显示了带荧光粒子的几个显著特征以及细胞总荧光的变化。图 10 显示了在一个典型实验中以 30 分钟间隔拍摄的一系列图像。同一视野中的细胞显示了不同振幅的同步波动（图 11c）。每个细胞中的总荧光在前两个小时内逐渐减少，但是随后又增加了。同时，细胞总荧光的减少导致了细胞极处出现强荧光粒子。有趣的是，几个细胞在实验过程中产生荧光（图 11b，180 分钟以及以后）。这些变化的动力学是随着不同实验而变化的，但是荧光随时间的增加和减少的总体模式始终是一样的。

为了证明细胞荧光的变化实际上显示了 RNA 的动力学，我们从细胞培养物中提取了总 RNA，使用实时竞争性 PCR（rcPCR）结合 MALDI-TOF MS 检测来确定适配子和 mreB RNA 的绝对浓度，mreB RNA 是用于标准化的管家基因。实时竞争性 PCR 利用连续稀释的 DNA 竞争剂作为目标基因在扩增之前的内标。使用 MALDI-TOF MS，在 1 或 2 nt 碱基的延伸反应中利用了竞争剂和目标基因之间的单个碱基差异来对延伸产物的丰度定量。

适配子 mRNA 水平分析显示了统计学上的明显波动，而对照基因 mreB 保持不变（表 2）。数据显示在 1 小时标记处，适配子转录物达到峰值，然后在 120 到 150 分钟处迅速降回到零时间点的基线水平。接着是 170 分钟处的另一转录峰值，最后是急剧的下降，然后恢复到初始的零时间点的适配子 mRNA 水平。

表 2. 适配子 mRNA 转录物水平和 TITAN（Elvidge 等, 2006）统计

时间（分钟）	等价点（M）		倍数	P 值	R^2
	mreB	适配子			
0	2.67E-14	4.52E-11	1.00	-	0.975
60	2.45E-14	5.49E-11	1.32	0.033	0.986
120	2.57E-14	4.76E-11	1.09	0.260	0.990
150	2.42E-14	4.47E-11	1.09	0.243	0.969
170	2.54E-14	5.33E-11	1.24	0.052	0.989
210	2.69E-14	3.73E-11	0.82	0.081	0.988
240	2.06E-14	3.51E-11	1.01	0.496	0.987
270	2.48E-14	4.14E-11	0.98	0.452	0.984

浓度值是针对于从每个细胞培养平板提取的 RNA 样品的。倍数值相对于管家对照基因 mreB 进行了调整，并且是相对于零时间点基线来计算的。bootstrap P 值是对其表达不同于基线的基因的置信度测量值。所有被分析样品的残值（residual value）和失拟值（lack-of-fit value）均<0.05；是优质分析的象征。

用于实施例 9 和本实施例的应用技术可以被用于检测 RNA 定位和运动。这种新技术的这样的应用的一个例子是我们能够研究 ASH1 mRNA。该 RNA 编码抑制 HO 核酸内切酶转录和阻断子细胞中的交配型互换的细胞命运决定子。通过不同方法都已经证明它定位在出芽酵母细胞的芽尖上。ASH1 mRNA 被充分研究了；因此它可以被用作模型实验来验证活细胞中 RNA 定位和运动分析的灵敏度和特异性。

实施例 11

利用二元适配子-肽的相互作用体内检测 RNA。

在实验 9 和 10 中，我们显示了体内对 RNA 的监测，其中组合使用了荧光蛋白互补以及 RNA 蛋白与 RNA-适配子的高亲和力相互作用。在这些实验中，RNA 结合蛋白是由哑铃型结构构成的真核起始因子 4A（eIF4A）。将 eIF4A 切成两个片段，并将每个片段融合到增强型绿色荧光蛋白（EGFP）的裂解片段上。两个互补融合蛋白和含有 eIF4A 特异性适配子的转录物的共表达引起了细菌中 EGFP 荧光的恢复。该方法的主要优点在于荧光信号仅由靶 RNA 决定，换句话说，没有 RNA 就没有可检测的信号。此外，相当小的蛋白复合物组装在靶 RNA 上，这样就不会干扰 RNA 功能和定位。

在实验 11 中，我们提供了关于开发用于 RNA 观察的备选实施方案的数据，其中适配子结合部分是两个短肽。在该方法中，将两个不同的 RNA 适配子作为标签加到目标 RNA 上，并由与裂解 EGFP 片段融合的两个病毒短肽来识别（图 16）。进行这些修饰有几个原因。首先，组装在靶 RNA 上的蛋白复合物与含有重组装 eIF4A 的蛋白复合物相比实质上更小。通常，检测工具越小，干扰功能的可能性就越低。第二，eIF4A 是真核细胞中的翻译机器的组分，在细菌蛋白中也有其相近的同源物。因此，它的过表达有可能会干扰正常细胞的功能。同时，病毒短肽在细菌或真核细胞中没有同源蛋白，因此它们在细胞中的表达对于细胞功能来说更可能是中性的。最后，识别复合物的 RNA 的备选设计给新方法增加了更多灵活性，显示

了它的普适性。

用于体内 RNA 检测的互补复合物的设计根据是二元适配子-肽的相互作用。根据我们的计划，为了检测活细胞中的 RNA，提供的 RNA 应当具有能够与两个肽相互作用的两个适配子序列。每个肽应当在细胞中表达为与裂解的增强型绿色荧光蛋白（EGFP）的两个片段之一的融合物（图 16）。为了使这种系统运行，要考虑几个问题；首先，每个所选择的相互作用的肽/适配子配对的亲和力应当高并且两个配对都要以相当的亲和力进行结合。第二，两个适配子/肽配对之间应当没有交叉反应性，换句话说，每种相互作用的特异性应是足够高的。第三，肽之间应当没有相互作用，否则可能将标记蛋白片段聚集在一起从而增加了非特异性背景。其次，肽的长度应当是能够使互补复合物对称组装的相当长度。最后，两个适配子的位置应当使得它们中的任何一个都能够与相应的肽独立地相互作用。这可以使用位于适配子之间的柔性接头来实现。

带有富含精氨酸基序（ARM）的肽是决定许多 RNA 结合蛋白与相应 RNA 靶点相互作用的高亲和性和高特异性的片段（22）。这些肽的共同特征是精氨酸占优势，而没有其它相似性。最近，目的在于了解 ARM-肽与相应 RNA 的相互作用机制的研究得出结论，精氨酸位置的特定模式和肽骨架的柔性是相应 RNA 配体特异性结合的原因（23）。然而，许多 ARM-肽表现出“变色龙样”行为，并与许多靶 RNA 结合，尽管亲和力较低（24）。记住了这一点，我们从 ARM 病毒肽（25-27）中选择 RNA 结合蛋白，但是在将它们应用到蛋白互补之前，测试了它们与相应 RNA 的交叉反应性。

选择适配子-肽配对的体外实验。根据可获得的数据，我们选择了三种肽/适配子配对；（i）HIV Hex；（ii）噬菌体 λN；以及（iii）HTLV-1 Rev（见表 1），在体外测试了它们与同源伙伴的结合亲和力以及与其它两种 RNA 适配子的交叉反应性。在 RNA 适配子的固定浓度使得随着 ARM-肽浓度增加能够形成复合物的条件下，我们使用了非放射性凝胶转移分析。我们对未结合的 RNA 适配子和转移的复合物进行定量（表 1）。结果显示 λN 肽和 HIV-1 Rex 肽显示出高特异性且不与非匹配 RNA 适配子交叉反应。同时，HTLV-1 Rev 肽未显示出与两种非匹配 RNA 适配子的交叉反应性。根据这些结果，我们得出结论：λN 肽和 HTLV-1 的 Rex 肽以及它们相应的适配子提供了用于我们的蛋白互补复合物中的最佳配对（见表 3）。

适配子 / 肽	HIV-1 Rex 适配子	λN 适配子	HTLV-1 Rev
HIV-1 Rex 肽	+	-	-
噬菌体 λN 肽	-	+	-
HTLV-1 Rev 肽	+	+	+

表 3. 三种 RNA 适配子/肽配对之间的交叉反应

利用二元肽/适配子的相互作用检测细菌活细胞中的 RNA 转录物。克隆融合蛋白和含有适配子的 RNA 转录物，在 *E.coli* BL21（DE）3 细胞中表达（图 17）。通过由丝氨酸和甘氨酸残基组成的柔性接头将 EGFP 片段（第 1-158 位残基）C 末端融合到 Rex 肽（16 aa）的 N 末端。类似地，也通过多肽接头将第二 EGFP 片段（第 159-238 位残基）的 N 末端融合到 λN 肽（22 aa）的 C 末端。通过多步 PCR 合成编码两个融合蛋白的完整 DNA 插入片段和它们之间的 T7 启动子区域，并克隆到载体 pACYCDuet-1 的 *Nco*I 和 *Avr*II 位点之间以产生能够共表达两个融合蛋白的构建体。使用载体 pETDuet-1（Novagen）来表达 230 nt 长的 T7 转录物，其含有两个由 5 或 10 个 dT 残基连接的适配子序列。该信使包括前导序列，接着是两个适配子序列，以及大约 200 nt 的抗核酸酶的 T7 终止序列。

在诱导物异丙基-β-D-硫代半乳糖苷（IPTG）存在时，将表达完整互补复合物和合适对照的 *E.coli* 细胞室温培养过夜用于蛋白和 RNA 的共表达。当培养物的光密度达到大约 0.5（OD_{600} = 0.5）时，通过荧光活化细胞分选仪（FACS）对它们进行分析（图 17）。将表达完整互补复合物的 *E.coli* 细胞的荧光与只表达两个融合蛋白的细胞的荧光以及与表达两个融合蛋白加上两个适配子的错误组合（由 10 个 dT 核苷酸连接的两个相同 Rex 肽结合的适配子）的细胞的荧光进行比较。

结果显示在没有 RNA 表达时，1mM IPTG 的诱导产生了高细胞荧光（图 17A），表达正确适配子序列和错误适配子序列的细胞在荧光分布方面也没有差别（图 17A）。我们认为用 1mM IPTG 诱导的 T7 启动子表达的融合蛋白浓度太高，甚至在没有 RNA 靶时就进行重组装。如果这个意见是正确的，那么降低 IPTG 浓度就会导致背景降低。实际上，IPTG 浓度降低 10 倍就能够将只表达融合蛋白的细胞与表达带有两个正确 RNA 适配子的蛋白的细胞的荧光分布分开。然而，特异性仍然不高，不足以区别正

确适配子序列和错误适配子序列。最后，IPTG 浓度降低到 0.01mM 就能够将带有正确和错误适配子序列的细胞的荧光分布区分开。在这些优化条件下，表达完整互补复合物的细胞的平均荧光超出背景荧光（没有 RNA 成分）10-15 倍，带有正确适配子序列的细胞显示出比带有错误适配子序列的细胞高 4-5 倍的荧光。

荧光变化动态。使用荧光显微镜分析在优化条件下表达完整互补复合物的细菌细胞。观察到不同类型的荧光分布（图 18）。在大多数细胞中，荧光在细胞极处最高，与由 eIF4A-适配子相互作用触发的蛋白互补实验中获得的结果相似。一些细胞（大约 10%）具有位于细胞中部的荧光粒子，这与 Golding 和 Cox 所报道的结果（18）相似。

时移成像展示了荧光在时间和空间上的变化，这与基于 eIF4A 的互补系统的结果也是相似的。图 18 显示了在一个代表性实验中以 30 分钟间隔拍摄 4 小时的一系列图像。细胞荧光显示出波动变化，在不同细胞中的波动是同步的（图 19b）。

采用二元适配子-肽相互作用和蛋白互补系统获得的结果在几个方面与使用裂解起始因子 4A（eIF4A）与相应适配子的相互作用的系统所获得的结果是非常相似的。首先，荧光粒子在绝大多数细胞中都定位在细胞极处，这是两种方法的特征。第二，荧光随时间和细胞空间的动态变化也与基于 eIF4A 的系统相似。最后，在两个系统中都能够看到细胞之间荧光的同步变化。这些结果暗示，用于蛋白互补复合物的 RNA-蛋白结合蛋白或肽的性质对于系统运行来说是不重要的，其它裂解蛋白或短肽可以用于相似的应用中。有关使用基于 MS2 外壳蛋白的系统进行活细胞 RNA 观察的已发表结果也支持该结论（18，20）。

有趣的而同时指出的是，两种互补设计的比较显示了荧光信号强度和信号/背景比方面的差异。由于一些我们不完全理解的原因，表达融合到裂解 EGFP 上的短肽的细胞的荧光要比表达与 EGFP 融合的 eIF4A 片段的细胞的荧光高的多。我们假定 ARM-肽比中性 eIF4A 片段更易于通过第三方分子进行结合，原因在于 ARM-肽的正电荷以及它们与带负电荷蛋白相互作用的能力。这导致了 EGFP 的重组装和更高的背景。为了降低背景，我们降低了 IPTG 浓度，发现了信号/背景比为 10-20 倍的条件，这是与 eIF4A 系统相同的范围。此外，在这些条件下也实现了正确和错误适配子序列之间的区分。

材料和方法。

构建体和菌株。

pMB33 和 pMB38 是 pACYCDuet-1（Novagen）的衍生物，相互作用的蛋白片段和 EGFP 的 ORF 被分别克隆到其中。将 eIF4A 片段 1（F1: 1-215 aa）克隆到 pACYCDuet-1（pMB09）中；片段 2（F2: 216-406 aa）克隆到 pETDuet-1（Novagen）（pMB11）中，这是通过 PCR 扩增质粒 pGEX-4AI（由 Dr. Chris Proud 惠赠）的小鼠 eIF4A 蛋白完成的。类似地，由 pEGFP（Clonetech）PCR 扩增 EGFP 片段，α（A: 1-158 aa）和 β（B: 159-238 aa），并分别克隆到 pACYCDuet-1（pMB08）和 pETDuet-1（pMB10）中。根据 Vasl 等, 2004.（详情参见补充方法）制备了嵌合基因（pMB12），其中通过 10 个氨基酸的柔性多肽接头（Gly-Ser-Ser-Gly-Ser-Ser-Gly-Ser-Gly-Ser）将 EGFP 片段 A 的 C 末端融合到 eIF4A 片段 F1 的 N 末端。使用相似的方案制备融合物 B-F2（pMB13）。如上所述，将 B-F2 克隆到 pACYCDuet-1（pMB33）的衍生物 pMB12 中（Geiser 等, 2001）。表达含有 eIF4A 相互作用的适配子序列（58nt 长）的 T7 转录物的 pMB23 是 pETDuet-1（Novagen）的衍生物。pMB42 是 pRSFDuet（Novagen）的衍生物，也表达 eIF4A 相互作用的适配子序列（详情参见补充方法）。将 *E.coli* 菌株 XL10-GoId（Tet^r Δ（*mcrA*）*183*Δ（*mcrCB-hsdSMR-mrr*）*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac*Hte [F' *proAB lacI*qZΔ*M15* Tn*10*（Tetr）Amy Camr]）和 XL10-GoId Kan^r（TetrΔ（*mcrA*）*183* Δ（*mcrCB-hsdSMR-mrr*）*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*qZΔ*M15* Tn*10*（Tetr）Tn5（Kanr）Amy）（Stratagene）用于克隆目的。将 BL21（DE）3（*E. coli* B F^- *dcm ompT hsdS*（r_B^- m_B^-）*gal λ*（DE3））（Stratagene）用于融合蛋白和靶 RNA 的表达。

根据 Vasi 等（Vasi 等, 2004）描述的方法制备含有 EGFP 和 eIF4A 片段的融合蛋白。设计并从 Integrated DNA Technology（Coraliville, IA）购买了两组 5'-磷酸化的寡核苷酸：1. 5'-pCGAAGA TCCAGAGGA TCCCTGCTTGTCGGCCATGATATAG-3'（SEQ ID NO.9）2. 5'-pGGTTCTGGTAGCATGGAGCCGGAAGGCGTCATCGA-3'（SEQ ID NO. 10），3. 5'-pCGAAGA TCCAGAGGA TCCCTTGTACAGCTCGTCC ATGCC-3'（SEQ ID NO. 11），4. 5'-pGGTTCTGGTAGCATTCGGATTCTTGTCAAGAAGGA-3'（SEQ ID

NO.12）。寡核苷酸中下划线区域对应于 A 片段的 3'末端（SEQ ID NO.9）；F1 片段的起点（SEQ ID NO.10）；片段 B 的 3'末端（SEQ ID NO.11）；以及 F2 片段的起点（SEQ ID NO.12）。余下的寡核苷酸序列对应于肽接头 GSSGSS（SEQ ID NO. 9 和 11）以及 GSGS（SEQ ID NO.10 和 12）的编码序列。用 XhoI 和 EcoNI 限制性内切酶切割质粒 pMBO8（携带 EGFP 的片段 A）和 pMBO9（携带 eIF4A 的片段 F1）；这些限制性位点位于寡核苷酸退火位点的 5'端。使用 Expand Long Template FOR 系统（Roche Diagnostics），利用这些线性化质粒和寡核苷酸 1 和 2 进行 PCR 反应，反应按照如下程序进行：94℃、3 分钟；94℃、30 秒、30 个循环，61℃、30 秒以及 72℃、10 分钟，72℃下以最终的 11 分钟延伸步骤结束。然后用 *Dpn*1 酶处理 PCR 混合物以去除甲基化模板 DNA。凝胶纯化大约 5～5 kbp 的 PCR 产物，使用 1U 的 T4 DNA 连接酶（New England Biolabs）进行连接。将连接产物转化到 XL-10 感受态细胞（Stratagene）中。分离携带 A-F1（pMB12）的新嵌合质粒，经测序确认。使用寡核苷酸#3 和#4 按照相似的方案制备载体 pETDuet-1（pMBI3）中的嵌合基因 B-F2。

将两种嵌合蛋白克隆到 pACYCDuet-1 的两个 MOS 中。为了保持两种嵌合基因表达的水平相近，接着根据所记载的方案（Geiser 等，2001）将 B-F2 片段整合到已经携带 A-F11 的 pMB12 的第二 MCS 中。简要地说，由侧翼带有 5'和 3'载体序列的质粒 pMBI3 对 B-F2 进行 PCR 扩增。侧翼序列直接对应于 pMB12 插入位点上游和下游的 20 bp 的 DNA。用受体载体和 B-F2 片段进行 PCR 反应，B-F2 片段将利用 20 bp 侧翼同源性退火到未切割的载体上。PCR 程序包括 95℃下 30 秒的变性步骤，接着是 18 个循环，95℃下 30 秒、55℃下 30 秒和 68℃下 8-10 分钟，使用 *Pfu* turbo DNA 聚合酶。用 *Dpn1* 酶处理扩增产物 3 小时以去除最初的甲基化模板 DNA。用一份纯化产物转化 XL-10 感受态细胞。分离携带两种蛋白片段（A-F1 和 B-F2）的质粒，经测序确认（pMB33）。还制备构建体 A-F1 和 B-F2 作为阴性对照（pMB54）。最后，将全长 EGFP 克隆到载体 pACYC 中作为 EGFP 表达的阳性对照（pMB38）。

克隆二元适配子-肽嵌合物。在第 158 和 159 位氨基酸残基之间将 EGFP 裂解成两个无荧光片段，分别称为 α-EGFP 和 β-EGFP。通过 10-aa 的柔性多肽接头（Gly-Ser-Ser-Gly-Ser-Ser-Gly-Ser-Gly-Ser）将 HTLV-1 Rex 肽融合到 α-EGFP 的 C 末端；通过同样的 10-aa 接头将噬菌体 λN 肽融合

到β-EGFP的N末端(图18)。进行多步PCR制备编码两个融合蛋白加上它们之间的T7启动子的DNA片段。将DNA构建体插入到载体pACYCDuet-1(Novagen)的限制性位点*Nco*I和*Avr*II之间，该插入序列位于第一T7启动子区域之后和T7终止子区域之前(图17)。这样，由一个pACYCDuet-1质粒表达了两个蛋白嵌合物以确保表达等摩尔量的两种融合物。使用*E.coli*菌株XL10-GoId Kan^r(Tetr Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB lacI*qZΔ*M15* Tn10(Tetr)Tn5(Kanr)Amy](Stratagene)进行克隆。所有构建体都经测序验证。

适配子克隆。设计编码两种RNA适配子的DNA序列，其中一种结合λN肽，另一种结合HTLV-1 Rex肽。通过10个胸腺嘧啶分开适配子，并在末端加入限制性位点*Xba*I和*Avr*II。定制合成相应的DNA模板，PCR扩增，在T7启动子控制下插入到pETDuet-1载体(Novagen)的*Xba*I和*Avr*II限制性位点之间。pETDuet-1和pACYCDuet-1载体在共表达中是相容的。作为阴性对照，类似地合成编码两种相同RNA适配子序列的DNA序列。使用*E.coli*菌株XL10-Gold进行该克隆。所有构建体都经测序验证。

培养条件和诱导。用pMB33(表达嵌合蛋白)和pMB23(表达靶RNA)共转化BL21(DE)3细胞。在*E. coli* BL21(DE)3(B F$^-$ dcm ompT hsdS(r_B^- m_B^-) gal λ(DE3))(Stratagene)中共表达编码EGFP互补复合物的蛋白和RNA组分的两个质粒。作为阴性对照，用含有两种相同适配子的质粒转化细胞。作为另一个阴性对照，用不含适配子插入片段的pETDuet-1质粒转化细胞。首先将转化细胞的单个菌落在37℃下在添加了抗生素的LB培养基中培养3-4小时。37℃孵育之后，将培养物加到含有诱导物异丙基-β-D-硫代半乳糖苷(IPTG; 1 mM或0.01 mM，用于二元适配子-肽相互作用)的新鲜培养基中稀释300倍，室温培养过夜。培养物的光密度在检查之时介于0.4到0.6之间(OD_{600}=0.4到0.6)。在BL21(DE3)pLys感受态细胞(Stratagene, CA)中表达蛋白。用0.35 mM IPTG进行诱导，将细胞在37℃下培养4小时。收获细胞，在含有50 mM Tris-HCl、pH 8.0、25%蔗糖、1 mM EDTA、10 mM DTT和0.1%叠氮化钠的缓冲液中清洗并超声破碎(3次，每次30秒)。用同样的缓冲液清洗包涵体1次，用含有50 mM Tris- HCl、pH 8.5、100 mM NaCl、0.5% triton X100、1 mM EDTA、1 mM DTT、0.1%叠氮化钠的缓冲液清洗3次，然后溶解在含有8 M尿素、

25 mM MES、pH 8.5、10 mM EDTA 和 0.1 mM DTT 的缓冲液中。通过在不含尿素的同样缓冲液中逐滴稀释使蛋白再折叠，加到用缓冲液（50 mM tris-HCl、pH 8.5、0.15 M NaCl、1 mM EDTA、0.1 % Triton X-100、1 mM PMSF）平衡过的几丁质亲和层析树脂（New England Biolabs, MA）中。用大量同样的缓冲液清洗柱子。然后在含有 50 mM tris-HCl、pH 7.0、150 mM NaCl、1 mM EDTA、1 mM DTT、Triton X-100 0.1 %、1 mM PMSF 的缓冲液中平衡柱子，置于 40℃下达 24-48 小时进行蛋白内含子切割。收集洗脱液，通过 SDS-PAGE 和 Coomasie Plus Protein 染色（Pierce, IL）分析蛋白浓度和纯度（见图 15 和 16）。所有蛋白都以相似的方式进行纯化，除了一种情况：在分离 EGFP 的 α 亚基的几丁质层析步骤中，由 PBS 缓冲液代替 Tris-HCl 缓冲液。

在一些实施例（实施例 1 和 2）中，通过在不含尿素的同样缓冲液中逐滴稀释使蛋白重新折叠，并加到用缓冲液（50 mM tris-HCl、pH 8.5、0.15 M NaCl、1 mM EDTA、0.1 % Triton X-100、1 mM PMSF）平衡过的几丁质亲和层析树脂（New England Biolabs, MA）中。用大量同样的缓冲液清洗柱子。然后在含有 50 mM tris-HCl、pH 7.0、150 mM NaCl、1 mM EDTA、1 mM DTT、Triton X-100 0.1 %、1 mM PMSF 的缓冲液中平衡柱子，置于 40℃下达 24-48 小时进行蛋白内含子切割。收集洗脱液，通过 SDS-PAGE 和 Coomasie Plus Protein 染色（Pierce, IL）分析蛋白浓度和纯度（见图 15 和 16）。所有蛋白以相似的方式进行纯化，除了一种情况：在分离 EGFP 的 α 亚基的几丁质层析步骤中，由 PBS 缓冲液代替 Tris-HCl 缓冲液。在这些实施例中，从 IDT DNA Technologies 购买了两条 20 nt 长的互补寡核苷酸，它们的 5'和 3'末端带有 SH 基团。用含有 10 mM DTT 的缓冲液孵育，去除它们的保护帽，通过 G25 柱凝胶过滤进行脱盐，在催化剂 3 mM 硫酸铜和 9 mM 1,10-啡咯啉存在下与等摩尔量的蛋白片段在暗处于 370℃下偶联 30 分钟（Lee 等, 1994）。偶联效率接近 100%（见图 29）。将含有两半 EGFP 和互补寡核苷酸的偶联嵌合物以等摩尔量在透析管中混合，用含有 50 mM tris-HCl、pH 8.5、0.5 M NaCl、5 mM $MgCl_2$ 和 20 μM PEG（6000 MW）的缓冲液进行透析。使用 Hitachi 荧光分光光度计 F-2500 获取 4 小时中的荧光光谱（图 30）。

在有合适抗生素（氯霉素和/或链霉素）的 LB 培养基中培养细胞，用 0.2 mM IPTG 在生色底物头孢硝基噻吩（nitrocefin）存在下进行诱导。阴

性对照将不表达适配子序列、破坏的适配子序列以及没有结合适配子所必需的成分之一的融合蛋白。使用 NanoDrop ND-1000 分光光度计测定无细胞上清液在 490 nm 处的吸光度。

流式细胞术。用带有 488 nm 氩激发激光器和 515-545 nm 发射滤光片（FL1）的 Becton-Dickinson FACS 流式细胞仪获取荧光测定值。在分析之前用 1×PBS 清洗细胞一次。获取测定值一直到收集到 100,000 个细胞。

荧光显微镜、成像和数据分析。将培养物中的细菌细胞固定在盖玻片和薄板之间，该薄板由 0.8%琼脂糖的 1×PBS 溶液制成。用配备了落射荧光系统 X-Cite 120 的 Nikon Eclipse 80i 倒置显微镜在室温下进行显微镜检查。使用带有由 IPLab v.3.7 软件（Scanalytics, Inc）控制的 100×放大倍数物镜的 B&W 数码相机（12 位；20 mHz），以 150-300 ms 的曝光时间拍摄图像。使用 ND4 滤光片减少细胞的光损伤。使用 ImageJ 1.36 B 软件（Wayne Rasband, NIH）进行图像处理。将通过显微镜获得的荧光照片以 JPEG 格式读取到 ImageJ 中，转变成 8 位类型。手动调整每张照片阈值。阈值上限自动设定到最大观测值，根据经验设定阈值下限以使作为对象的背景中没有像素。为获得细胞总荧光，选择了选项“分析粒子”。输出项包括被确定目标的像素面积，平均、最小和最大灰度级。通过将平均值减去最小值（灰度级/像素）再乘以面积（像素）得到积就获得了细胞的总荧光。该计算方法减去背景。以相似的方法计算单个细胞中的总荧光变化动态，这次是通过将矩形分割工具（rectangular segment tool）应用到目标细胞上。对于沿细胞的荧光分布的定量来说，测定沿细菌细胞长轴的荧光图谱，在 Microsoft Excell 中计算表面积峰值。每个细胞测定 3 到 4 次，取结果平均值。类似地，对每个细胞周围的背景荧光定量，并从细胞荧光图谱中将其扣除。

荧光测定和细胞成像：用 Becton-Dickinson 荧光活化细胞分选仪（FACSCalibur，带有 488 nm 氩激发激光器）测定细胞总荧光，使用 Excel 软件分析。使用带有荧光和微分干涉相差显微镜（DIC）之间自动切换的共聚焦显微镜系统进行细胞成像。将放大×150 的样品照片直接发送到冷却的、慢扫描 CCD 成像设备（C4880; Hamamatsu Photonics, Bridgewater, NJ）上。设定计算机控制（MetaMorph 2.5 软件；Universal Imaging Corp., West Chester, PA）的显微镜执行获取方案，拍摄 1μm 轴向步距的荧光照片和相应于中心荧光图像的单个 DIC 照片。对于 EGFP 监测来说，将使用带

有 488 nm 激发光线的氩激光器，发射窗将设定在 500-540 nm 之间。为产生几个时间点的合成照片，我们将使用 Metamorph 软件包（Universal Imaging Corporation）的 3D 重建特性。通过每 1-2 分钟或者如果需要的话可以更短时间拍摄图像来测定 RNA 速度，逐点跟踪荧光 RNA 点。

实时竞争性 PCR 设计、扩增和延伸。用 0.5 μg 随机六核苷酸和 AMV 反转录酶（Promega）以 20 μL 总体积在 42℃下对细胞培养物中总 RNA 样品进行反转录。使用 Sequenom 的 Assay Designer 软件设计以及从 Integrated DNA Technology（Coralville, IA）获得引物和竞争剂。使用 100 nM 的 PCR 引物、不同浓度的竞争剂、2.75 mM 的 $MgCl_2$、200 μM dNTP 和 5 μL 的 0.1 U HotStar Taq DNA 聚合酶（Qiagen）按照下面的 PCR 条件进行 cDNA 扩增：95℃热启动 15 分钟，接着是 45 个循环，95℃、30 秒，56℃、1 分钟、然后是 72℃、1 分钟，最后保持在 72℃、7 分钟。PCR 扩增之后，用 0.04 U 虾碱性磷酸酶 SAP（Sequenom）处理产物，将来自扩增循环的未使用 dNTP 失活，37℃下 20 分钟，接着是 85℃热失活 5 分钟。对于延伸循环来说，将终浓度为 1.2 μM 的延伸引物和 0.6 U 的 ThermoSequenase（Sequenom）加到含有终止混合物的 9 μL 总反应物中，终止混合物含有每个反应每个碱基 50 μM 的特定二脱氧核苷酸和脱氧核苷酸。延伸条件包括 94℃下保持 2 分钟，如下的 75 个循环：94℃、5 秒，52℃、5 秒和 72℃、5 秒。

MALDI-TOF MS 和定量分析。在 MALDI-TOF MS 分析之前，使用 SpectroCLEAN 树脂和 16 μl 水去除反应中的盐。使用 MassARRAY 系统（Sequenom）进行 ASV 分析，其中使用 SpectroPOINT 纳米级分配器（Sequenom）将大约 10 nL 的终产物分配到 384 平板型 MALDI-TOF MS SpectroCHIP 上。使用 TITAN（Elvidge 等, 2005）软件以默认值分析质谱数据。

培养基：将在 YPD（2%葡萄糖，1%酵母提取物，2%蛋白胨）中培养野生型酵母细胞。将在选择性合成缺陷型培养基（dropout media）（0.67%酵母氮源，2%葡萄糖）培养用质粒转化的细胞，该培养基缺乏色氨酸、尿嘧啶或亮氨酸。用于该研究的菌株是 YPH500（MATα，ura3-52，lys2-801 琥珀色，ade2-101 赭石色 trp1-Δ63，his3-Δ200，leu2-Δ1）（Johnsson &Varshavsky, 1994）的衍生物。

接合的检测子-核酸结合蛋白在酵母（真核生物）中的表达。用于在酵

母中表达嵌合蛋白 EGFP（A）-eIF4A（F1）和 EGFP（B）-eIF4A（F2）的质粒是表达载体 pDB20（Becker 等, 1991）的衍生物，该表达载体经修饰用来表达 N 末端融合了 HA 或 myc 表位标签的蛋白。这将能够由 ADH1 强启动子进行组成型蛋白表达并能够通过 Westren 印迹分析基因表达。修饰质粒 pRS4D1（Blake 等 2003）以表达完整的 ASH1 信使（代替 EGFP），它在基因的 3'末端处带有 64 nt 长的适配子标签（在第 4 个 zip-code 之后）。这个质粒整合到菌株 YPH500 的基因组中将使得能够在将半乳糖和脱水四环素加到培养基中时调节被标记的 ASH1 基因的表达。这样，我们将跟踪完整 ASH1 信使的定位，与带有 ASH1 的 3'定位序列的报道子转录物相反（Bertrand 等, 1998; Breach 和 Bloom, 2001）。这可能为我们提供有关 RNA 定位机制的新见解。

用消解酶处理酵母细胞，然后在含 2 μM CCF2/AM 的 DMEM 存在下以 3×10^5 个细胞/ml 的浓度培养 1 小时。用 PBS 缓冲液清洗细胞，使用荧光显微镜 Nikon Eclipse 80i 进行观察，设定为 405 nm 处激发和 440-450 nm 处发射的滤光片。也用荧光活化细胞分选仪（FACSaria，407 nm 处激发激光，Becton-Dickinson）测定细胞荧光。

体外凝胶转移分析。为进行体外凝胶转移分析，我们购买了表 1 所列的定制合成的肽和 RNA 适配子。首先在缓冲液（50 mM pH 8.0 Tris-HCl、50 mM KCl）中 95℃下加热 3 分钟使 RNA 变性，然后慢慢冷却到室温使 RNA 复性。将复性 RNA 和肽在普通缓冲液（50 mM Tris-HCl pH 8.0、50 mM KCl）中在 30℃下混合 15 分钟。使用 15%TBE 凝胶进行凝胶电泳，分析肽-RNA 适配子复合物。采用溴乙锭对 RNA 适配子和肽-RNA 适配子复合物染色。

提出这些实施例是为了向本领域普通技术人员提供关于如何制造和使用本发明的完整的公开内容和说明，并不打算对本发明人所认可的他们的发明范围进行限制，它们也不会是本发明能够实施的实验和实施例的仅有的代表。本领域技术人员将理解的是，根据现在的公开内容，可以在例举的特定实施方案和实施例中进行许多不脱离发明预期范围的修改和变化。预期将所有这些修改都包括在所附权利要求的范围之内，应明白的是，这些方法可以应用于许多系统中进行体内 RNA 观察。

参考文献

将这里和整个申请中引用的参考文献都完整引入这里作为参考。

1. Comolli LR, Pelton JG, Tinoco I Jr.（1998）Mapping of a protein-RNA kissing hairpin interface: Rom and Tar-Tar*. *Nucleic Acids Res.* 26（20）:4688-95.

2. Lee, G. F., Burrows G. G., Lebert M. R., Dutton D. P. & Hazelbauer G. L.（1994） *J. Biol Chem.* 269, 29920-29927.

3. Oguro A, Ohtsu T, Svitkin YV, Sonenberg N, Nakamura Y.（2003）RNA aptamers to initiation factor 4A helicase hinder cap-dependent translation by blocking ATP hydrolysis. RNA. 9（4）:394-407.

4. Roy S, Delling U, Chen CH, Rosen CA, Sonenberg N.（1990）A bulge structure in HIV-I TAR RNA is required for Tat binding and Tat-mediated trans-activation. *Genes Dev.* 4（8）: 1365-73.

5. Saha S, Ansari AZ, Jarrell KA, Ptashne M, Jarell KA（2003） RNA sequences that work as transcriptional activating regions. *Nucleic Acids Res.* 31（5）: 1565-70.

6. Sawata SY, Taira K.（2003） Modified peptide selection in vitro by introduction of a protein-RNA interaction. *Protein Eng.* 16（12）:1115-24.

7. Valegard K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L.（1997） The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein. *J MoI Biol.* 270（5）:724-38.

8. Yamamoto R, Katahira M, Nishikawa S, Baba T, Taira K, Kumar PK（2000） A novel RNA motif that binds efficiently and specifically to the Ttat protein of HIV and inhibits the trans-activation by Tat of transcription in vitro and in vivo. *Genes Cells.* 5（5）:371-88.

9. Le, T.T., *et al.* Real-time RNA profiling within a single bacterium. *Proc Natl Acad Sci USA.* 102, 9160-9164（2005）.

10. Demidov, V. V. *et al.* Fast complementation of split fluorescent protein triggered by DNA hybridization. *Proc Natl Acad Sci USA.* 103, 2052-2056（2006）.

11. Elvidge, G.P., Price, T. S., Glenny, L., Ragoussis, J. Development and

evaluation of real competitive PCR for high-throughput quantitative applications. *Anal Biochem.* 339, 231-241, (2005).

12. Vasl, J., Panter, G., Bencina, M. & Jerala, R. Preparation of chimeric genes without subcloning. Biotechniques. 37, 726, 728, 730 (2004).

13. Geiser, M., Cebe, R., Dreweiio, D. & Schmitz, R. Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. Biotechniques. 31,88-90,92 (2001).

14. Pederson, T. The molecular cytology of gene expression: fluorescent RNA as both a stain and tracer *in vivo*. *Eur J Histochem.* 48, 57-64 (2004).

15. Singer RH. RNA localization: visualization in real-time. *Curr Biol.* 2003, 13 (17):R673-675.

16. Bratu, D.P., Cha, B. J., Mhlanga, M.M., Kramer, F.R. & Tyagi, S. Visualizing the distribution and transport of mRNAs in living cells. *Proc Natl Acad Sci U S A.* 100, 13308-13313 (2003).

17. Bertrand, E. *et al.* Localization of ASHl mRNA particles in living yeast. *Mol Cell.* 2, 437-445 (1998).

18. Golding, I. & Cox, E.C. RNA dynamics in live Escherichia coli cells. *Proc Natl Acad Sci USA.* 101, 11310-11315 (2004).

19. Golding, I, Paulsson, J., Zawilski, S.M. & Cox, E.C. Real-time kinetics of gene activity in individual bacteria. *Cell.* 123, 1025-1036 (2005).

20. Chubb JR, Trcek T, Shenoy SM, Singer RH. (2006) Transcriptional pulsing of a developmental gene. *Curr Biol.* 16, 1018-1025.

21. Shav-Tal Y, Darzacq X, Singer RH. Gene expression within a dynamic nuclear landscape. *EMBO J.* 2006, 25 (15):3469-3479.

22. Smith CA, Chen L, Frankel AD. Using peptides as models of RNA-protein interactions. *Methods Enzymology* 318, 423-428, 2000.

23. Bayer TS, Booth LN, Knudsen SM, Ellington AD. Arginine-rich motifs present multiple interfaces for specific binding by RNA. RNA. 2005, 11 (12):1848-1857.

24. Smith CA, Calabro V, Frankel AD. An RNA-binding chameleon *Mol. Cell*, 2000, 6, 1067-1076.

25. Baskerville, S., Zapp, M., and Ellington, A.D. 1999. Anti-Rex

aptamers as mimics of the Rex-binding element. *J Virol.* 73（6）:4962-4971.

26. Ye, X., Gorin, A., Ellington, A.D., and Patel, DJ. 1996. Deep penetration of an alpha-helix into a widened RNA major groove in the HIV-1 rev peptide-RNA aptamer complex. *Nat Struct Biol.* 3（ 12）: 1026- 1033.

27. Baron-Benhamou J, Gehring NH, Kulozik AE, Hentze MW. Using the lambda N peptide to tether proteins to RNAs. *Methods Mol Biol.* 2004,257:135-154.

28. Michnick, S. W. Protein fragment complementation strategies for biochemical network mapping. *Curr Opin Biotechnol.* 14, 610-617,（2003）.

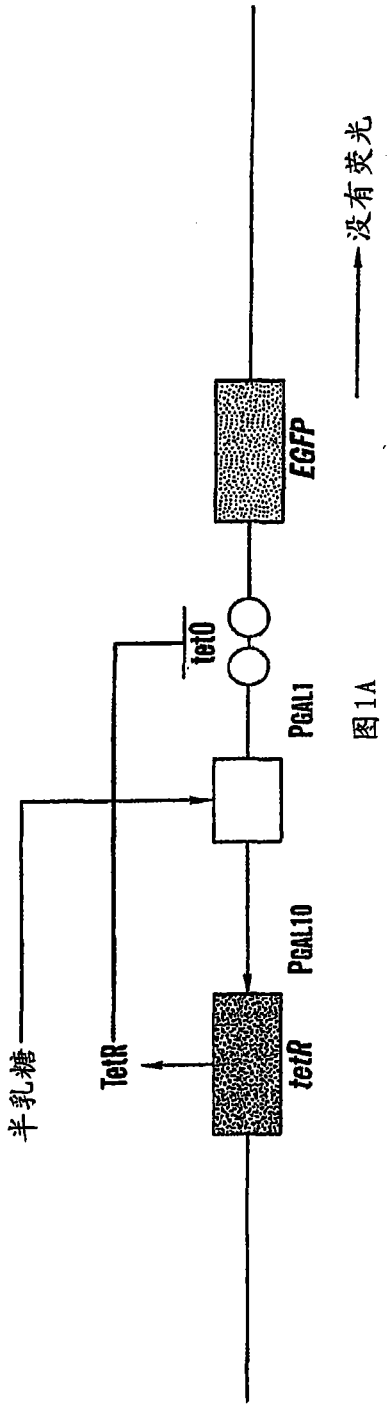

半乳糖
TetR
tetR
PGAL10
PGAL1
tetO
EGFP
没有荧光

图1A

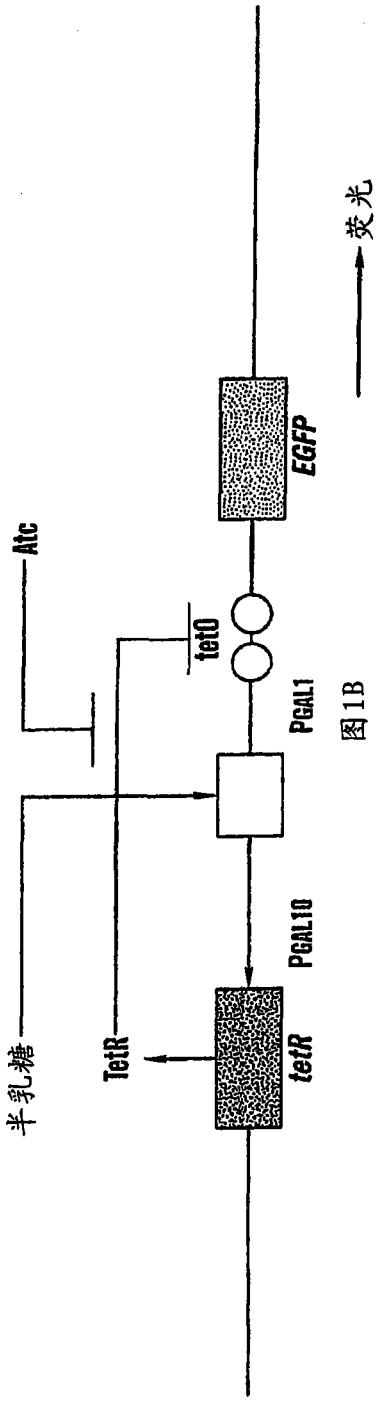

半乳糖
TetR
tetR
PGAL10
PGAL1
tetO
Atc
EGFP
荧光

图1B

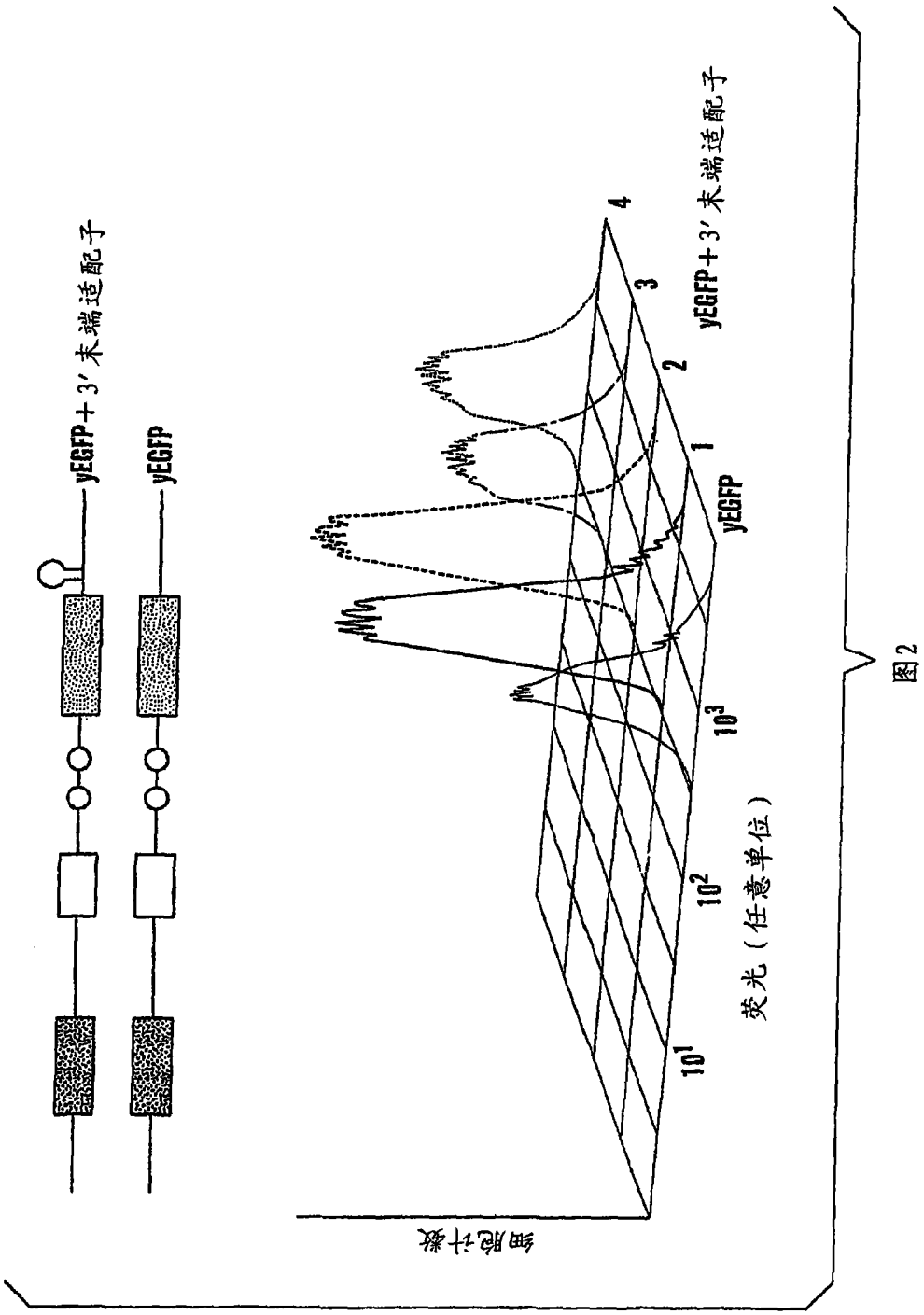

yEGFP+3′末端适配子
yEGFP
yEGFP+3′末端适配子
yEGFP
1
2
3
4
10^1
10^2
10^3
荧光（任意单位）
细胞计数

图2

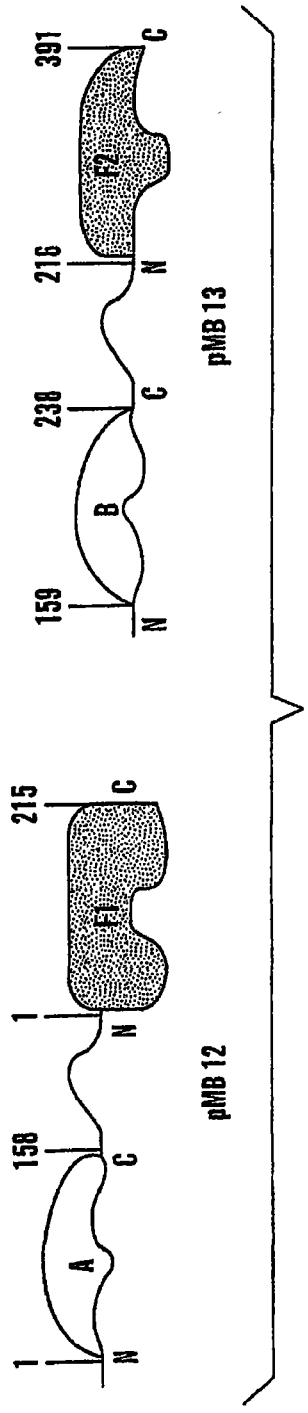


图3

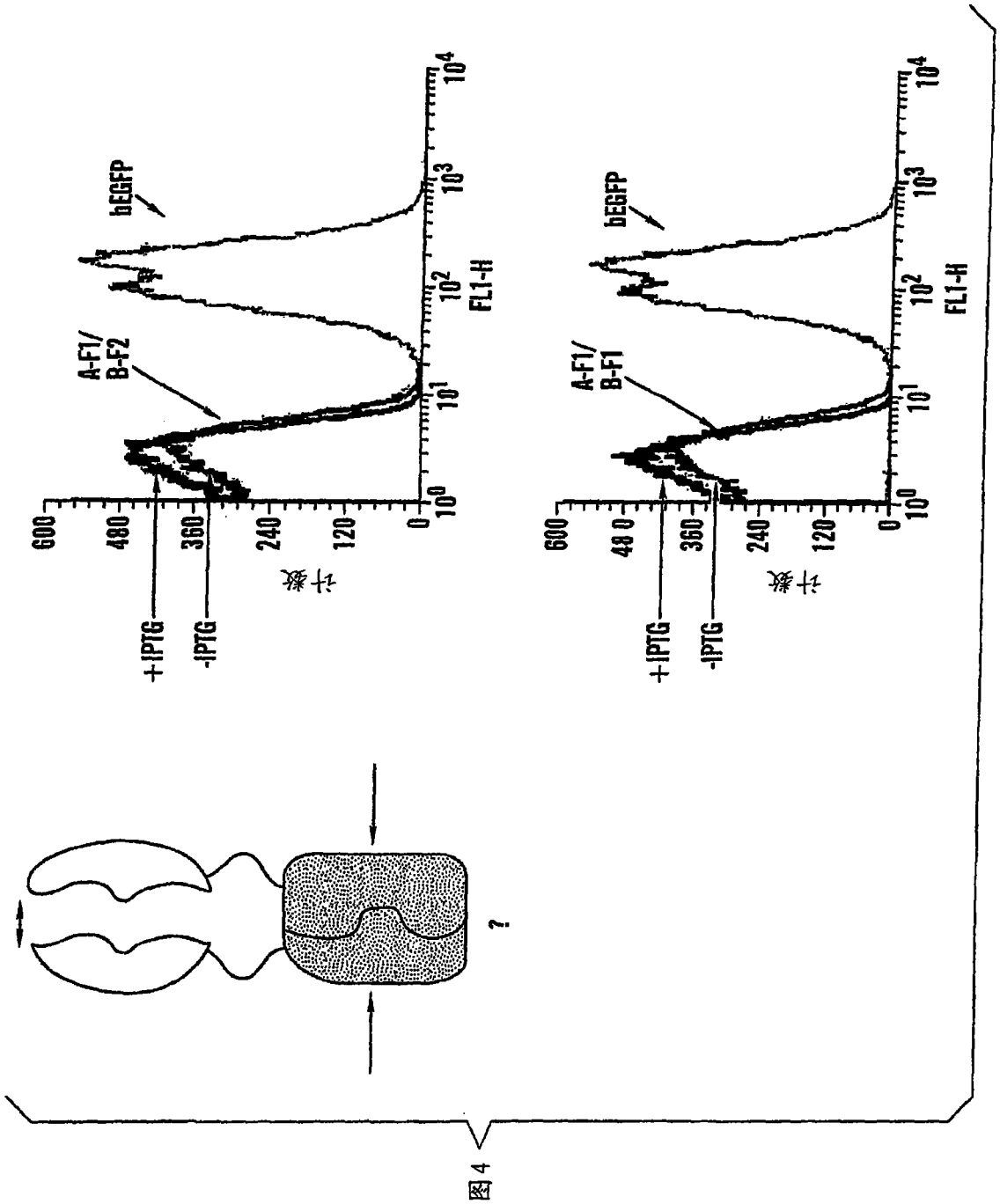

bEGFP
A-F1/
B-F2
FL1-H
计数
+IPTG
-IPTG
bEGFP
A-F1/
B-F1
FL1-H
计数
+IPTG
-IPTG
?

图4

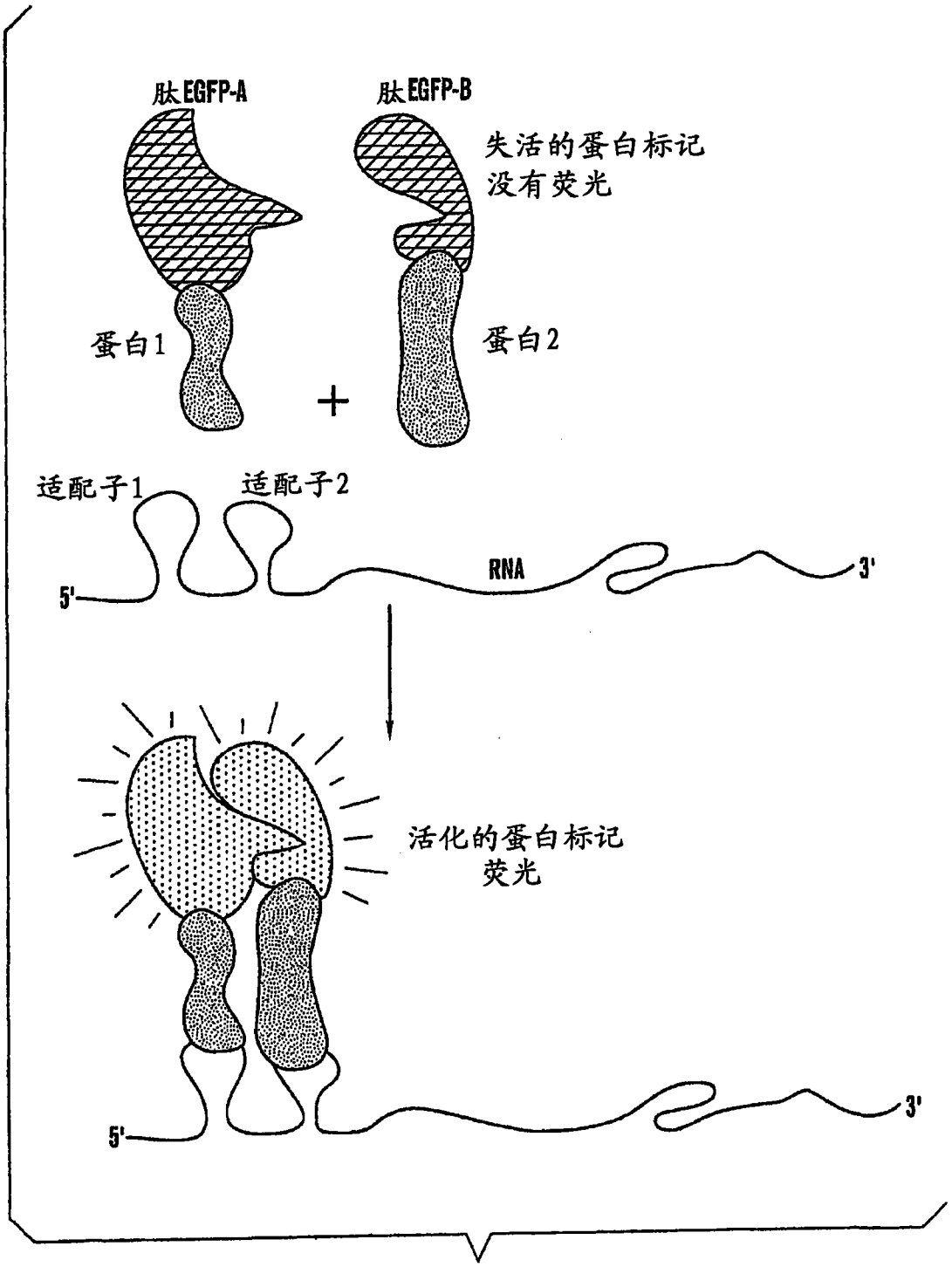

肽EGFP-A
肽EGFP-B
失活的蛋白标记
没有荧光
蛋白1
蛋白2
+
适配子1
适配子2
RNA
5'
3'
活化的蛋白标记
荧光
5'
3'

图5A

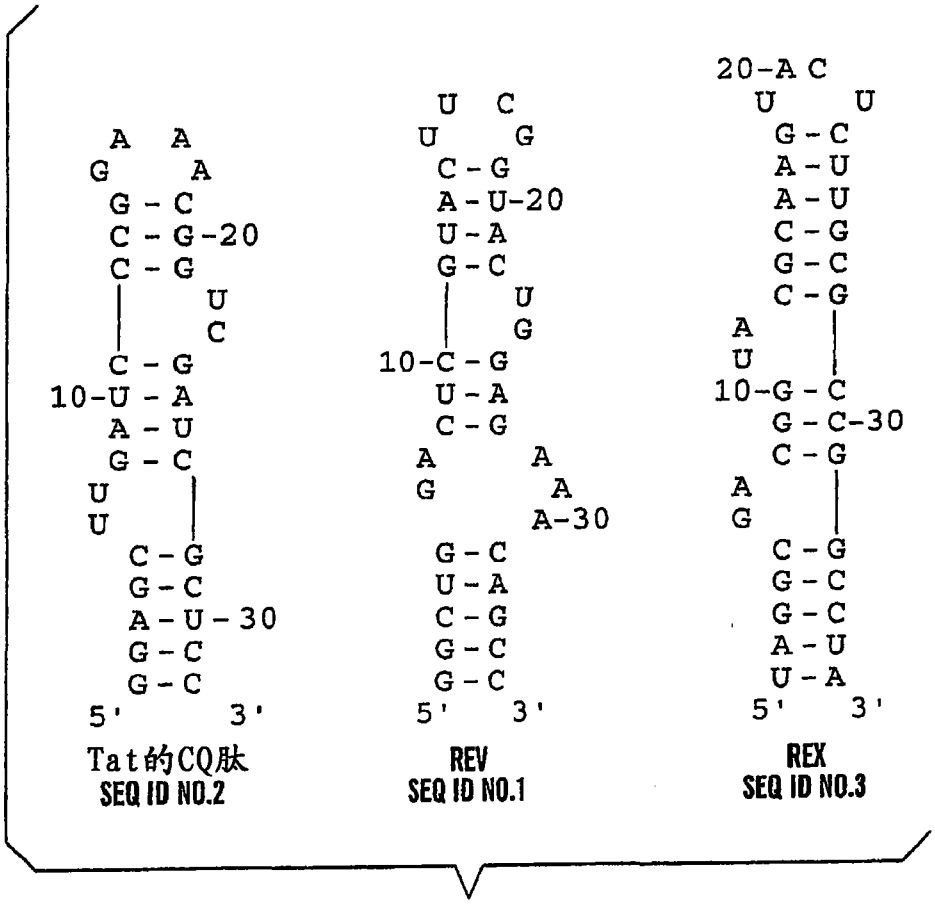

A A
G A
G-C
C-G-20
C-G
U
C
C-G
10-U-A
A-U
G-C
C-G
G-C
A-U-30
G-C
G-C
5' 3'
Tat的CQ肽
SEQ ID NO.2
U C
U G
C-G
A-U-20
U-A
G-C
U
G
10-C-G
U-A
C-G
A A
G A
A-30
G-C
U-A
C-G
G-C
G-C
5' 3'
REV
SEQ ID NO.1
20-A C
U U
G-C
A-U
A-U
C-G
G-C
C-G
A
U
10-G-C
G-C-30
C-G
A
G
C-G
G-C
G-C
A-U
U-A
5' 3'
REX
SEQ ID NO.3

图5B

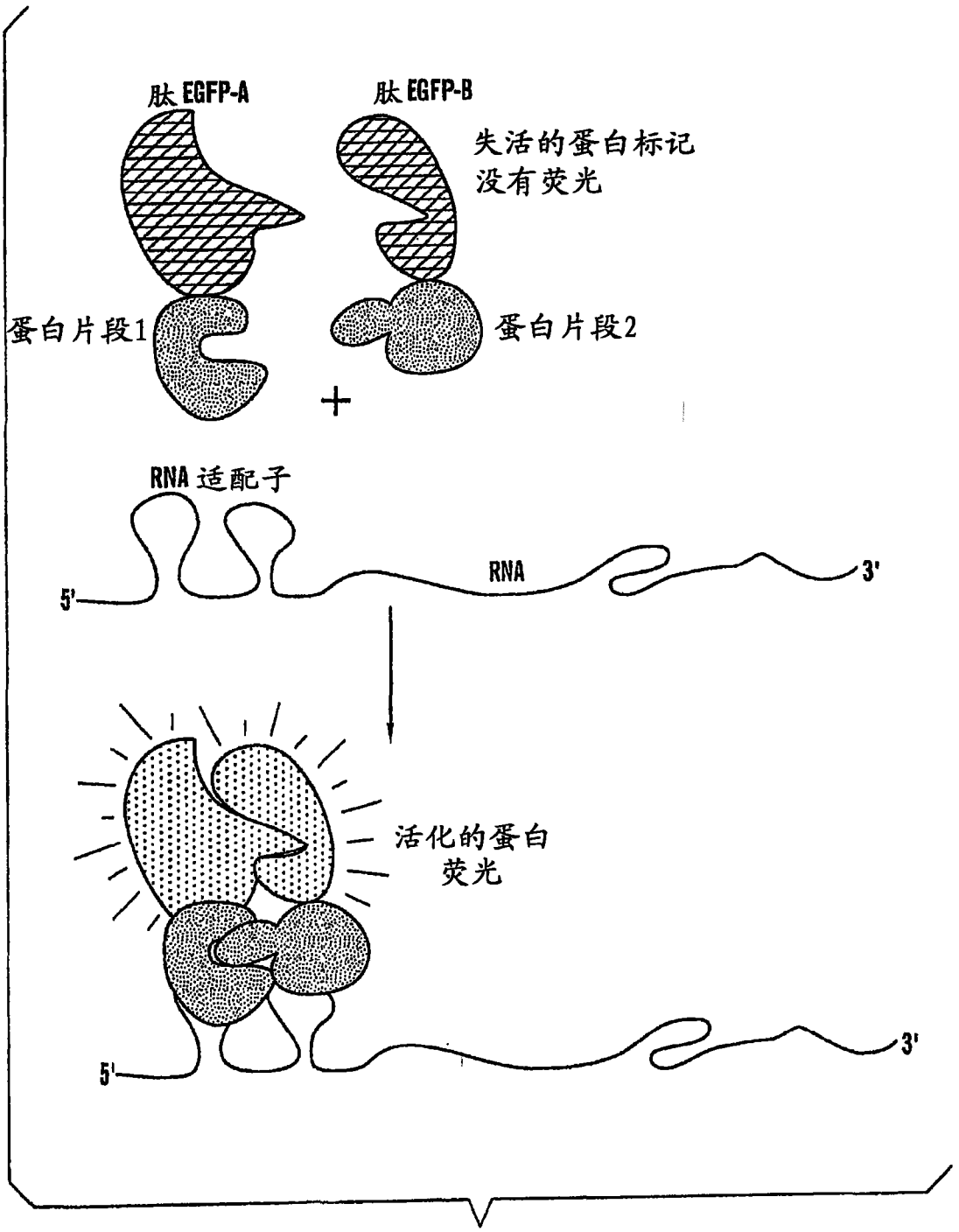

肽 EGFP-A
肽 EGFP-B
失活的蛋白标记
没有荧光
蛋白片段1
蛋白片段2
+
RNA 适配子
RNA
5'
3'
活化的蛋白
荧光
5'
3'

图6

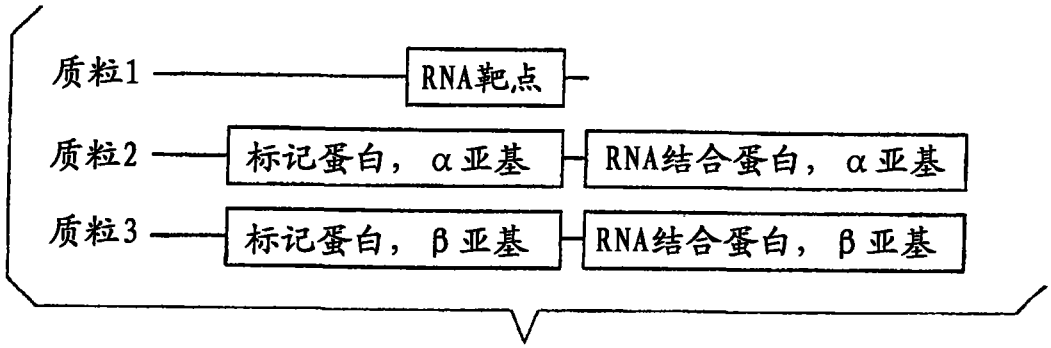

质粒1
RNA靶点
质粒2
标记蛋白，α亚基
RNA结合蛋白，α亚基
质粒3
标记蛋白，β亚基
RNA结合蛋白，β亚基

图7

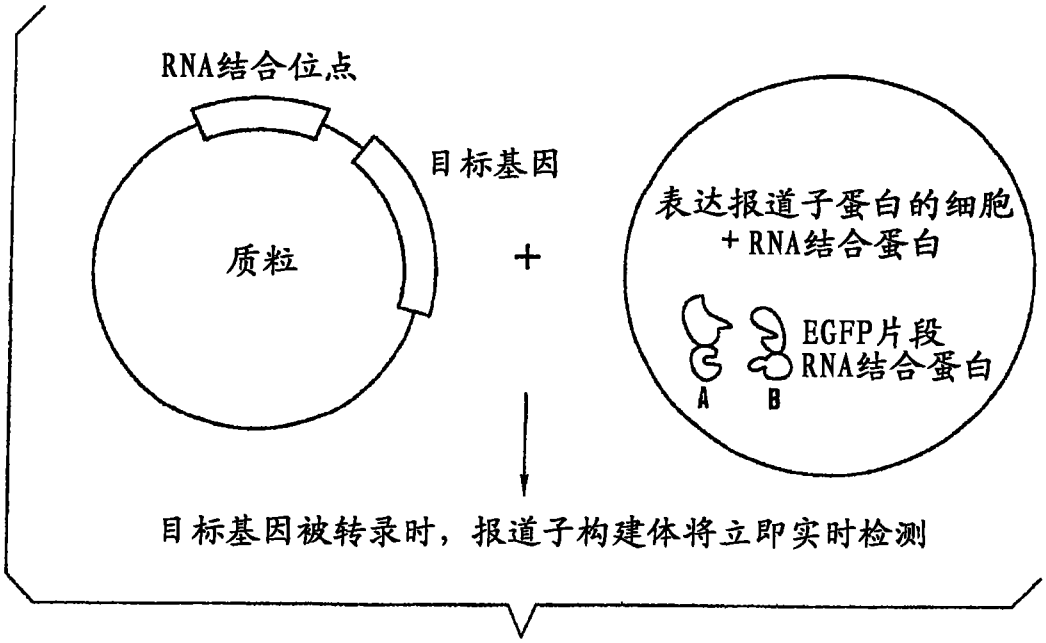

RNA结合位点
目标基因
质粒
+
表达报道子蛋白的细胞
+RNA结合蛋白
EGFP片段
RNA结合蛋白
A
B
目标基因被转录时，报道子构建体将立即实时检测

图8

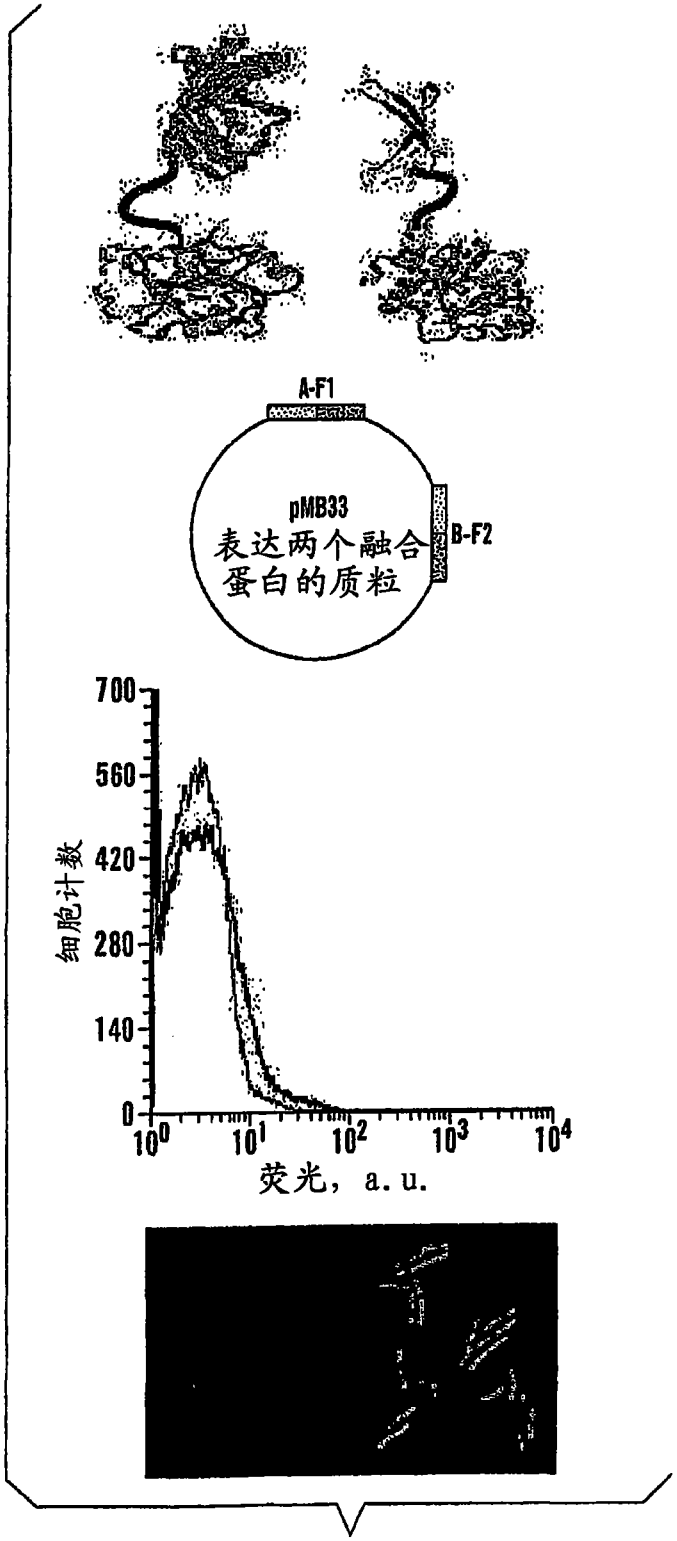

A-F1
pMB33
表达两个融合
蛋白的质粒
B-F2
700
560
420
280
140
0
细胞计数
10^0
10^1
10^2
10^3
10^4
荧光，a.u.

图9A

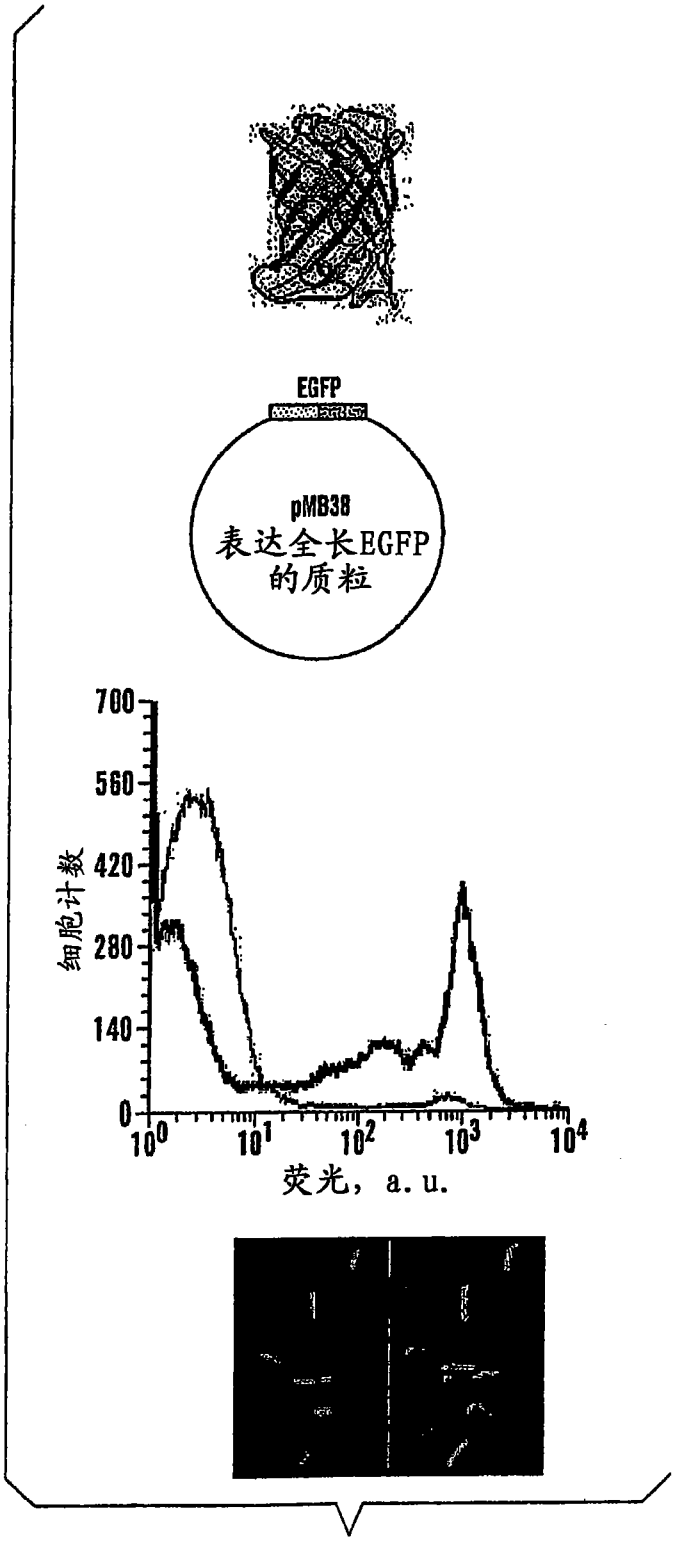

EGFP
pMB38
表达全长EGFP
的质粒
700
560
420
280
140
0
细胞计数
10^0
10^1
10^2
10^3
10^4
荧光，a.u.

图9B

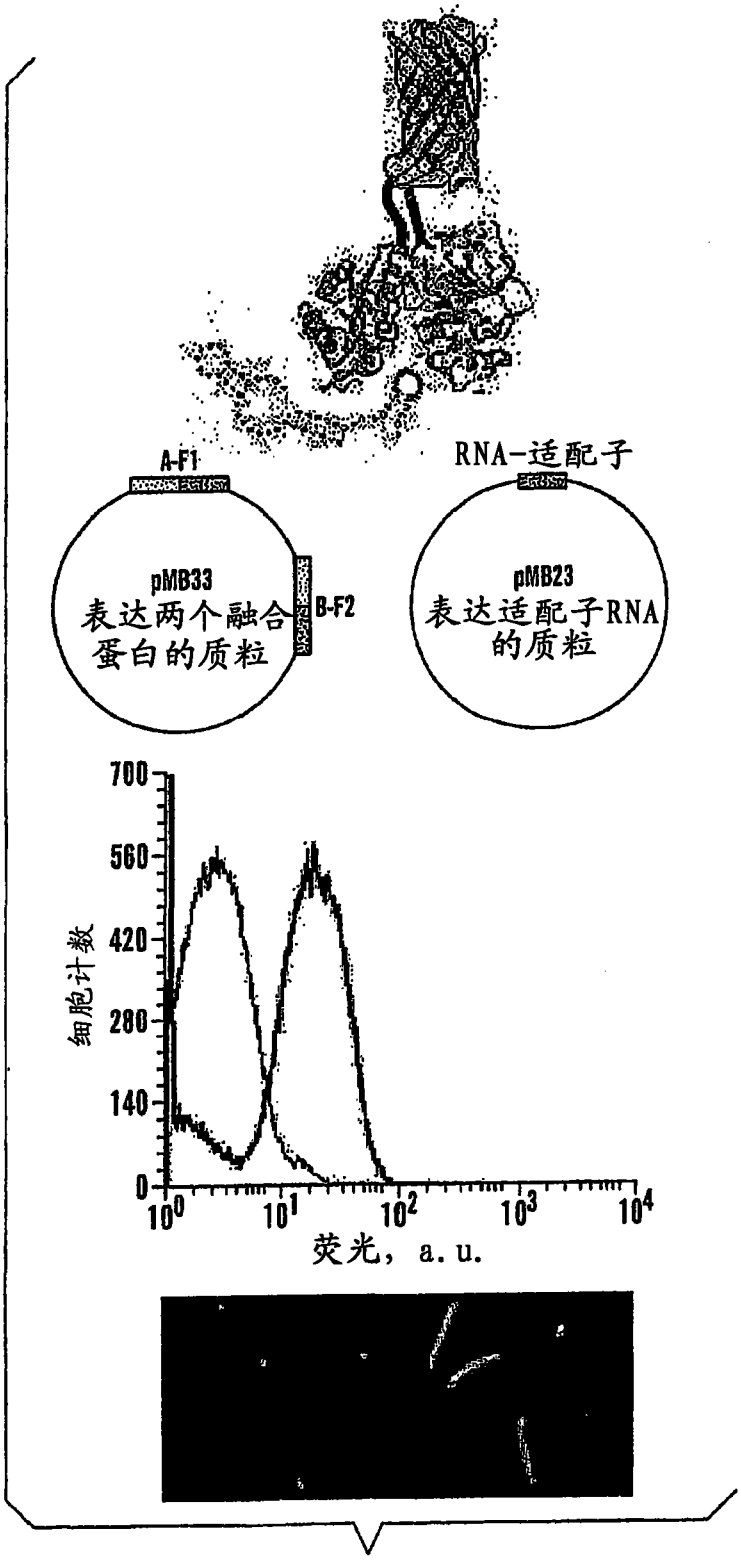

A-F1
RNA-适配子
pMB33
表达两个融合
蛋白的质粒
B-F2
pMB23
表达适配子RNA
的质粒
700
560
420
280
140
0
细胞计数
10^0
10^1
10^2
10^3
10^4
荧光，a.u.

图9C

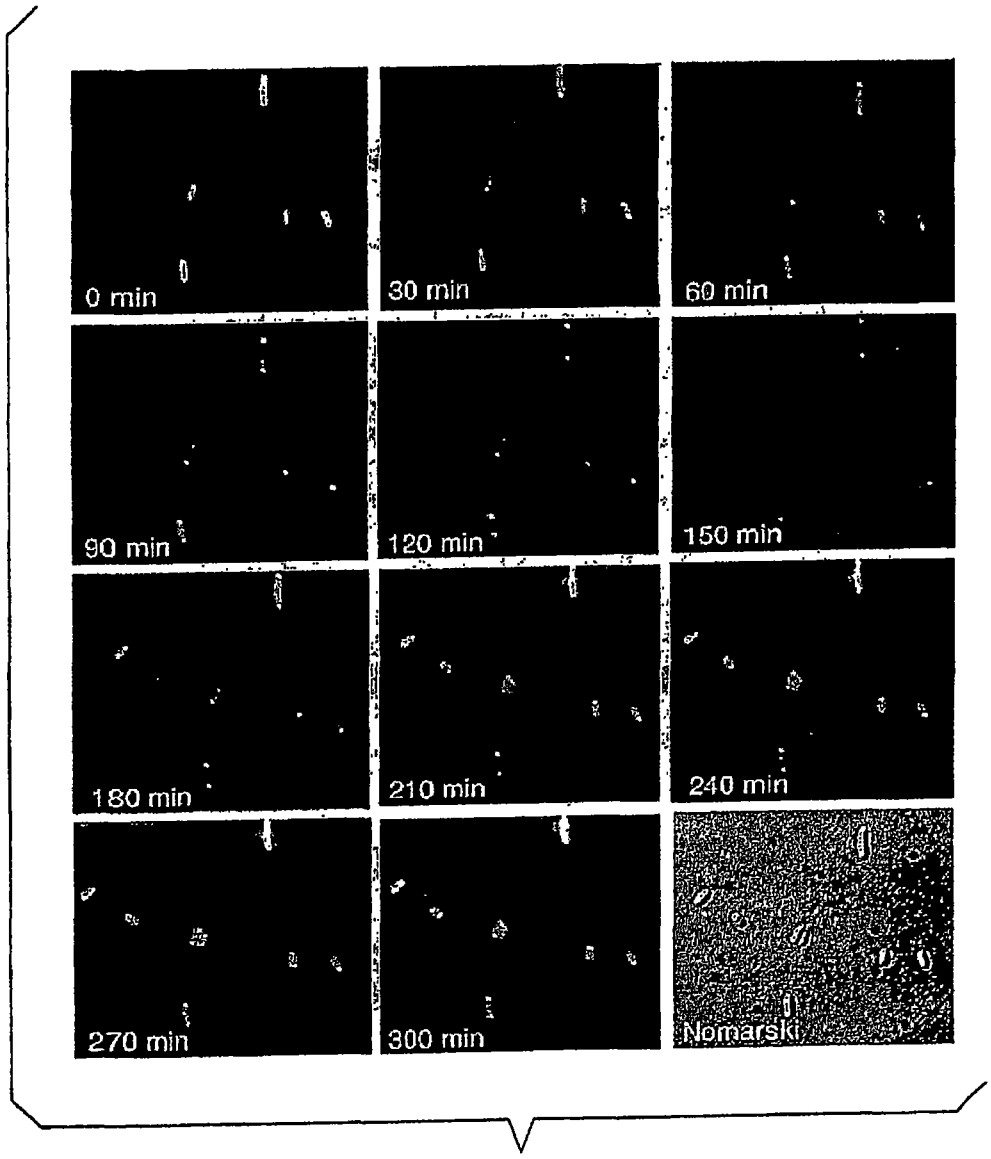

0 min
30 min
60 min
90 min
120 min
150 min
180 min
210 min
240 min
270 min
300 min
Nomarski

图10

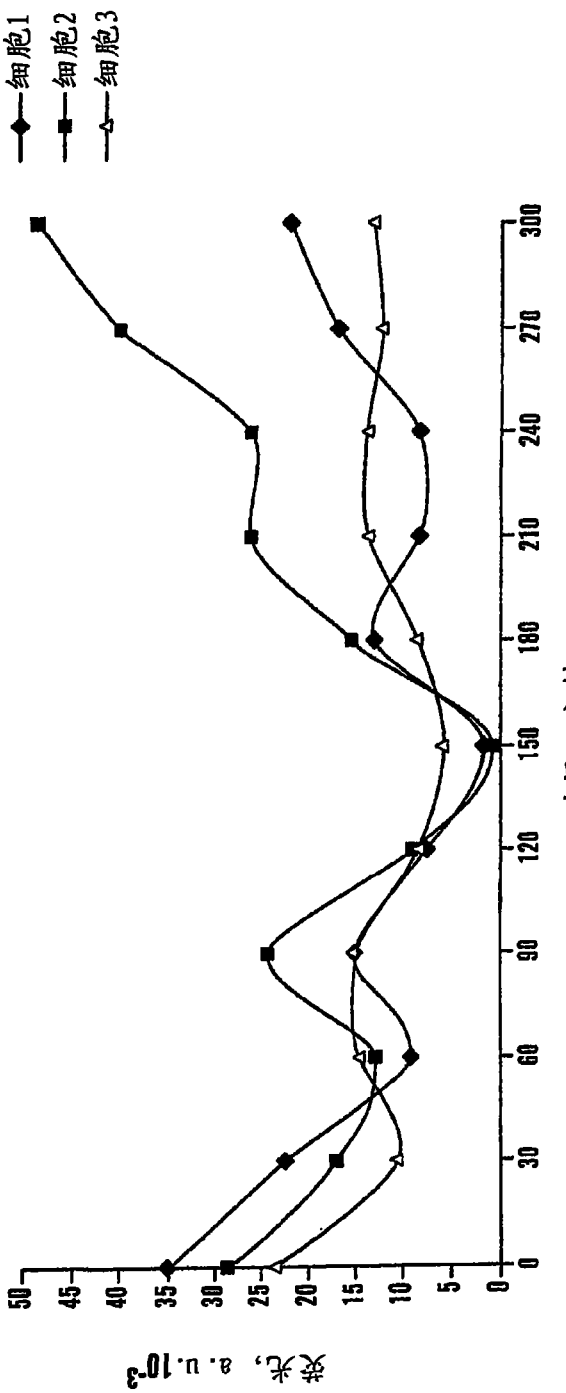

细胞1
细胞2
细胞3
荧光，a.u.10-3
0
5
10
15
20
25
30
35
40
45
50
0
30
60
90
120
150
180
210
240
270
300
时间，分钟

图11A

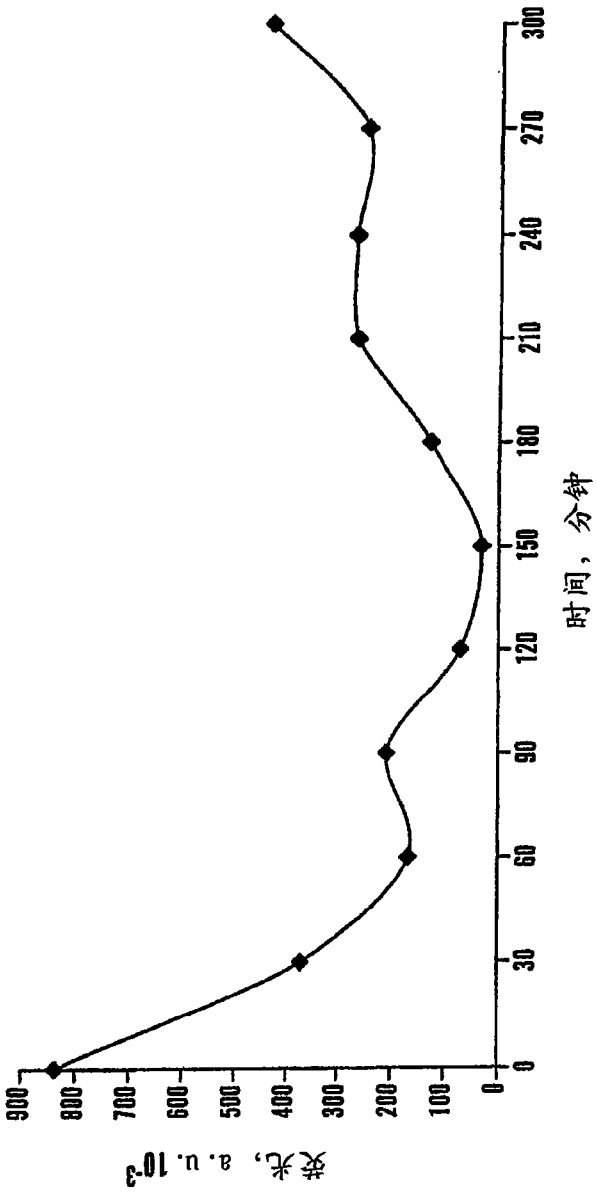

荧光，a.u.10-3
900
800
700
600
500
400
300
200
100
0
0
30
60
90
120
150
180
210
240
270
300
时间，分钟

图11B

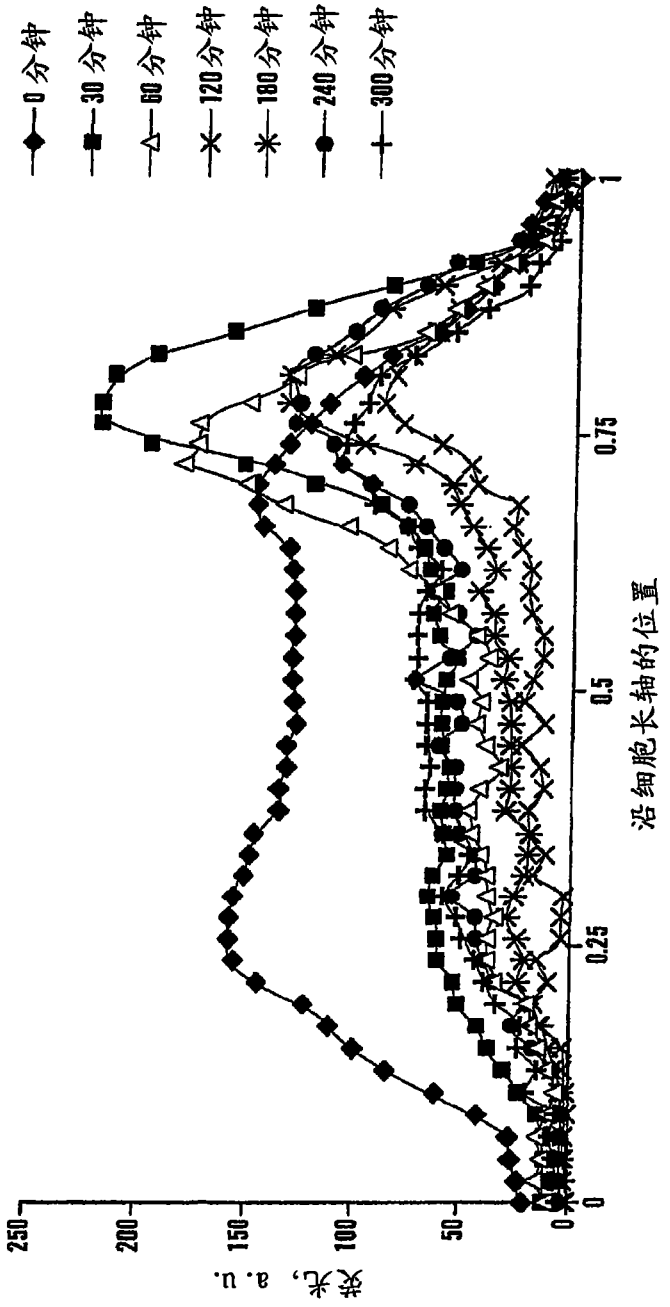

0 分钟
30 分钟
60 分钟
120 分钟
180 分钟
240 分钟
300 分钟
荧光, a.u.
250
200
150
100
50
0
0
0.25
0.5
0.75
1
沿细胞长轴的位置

图11C

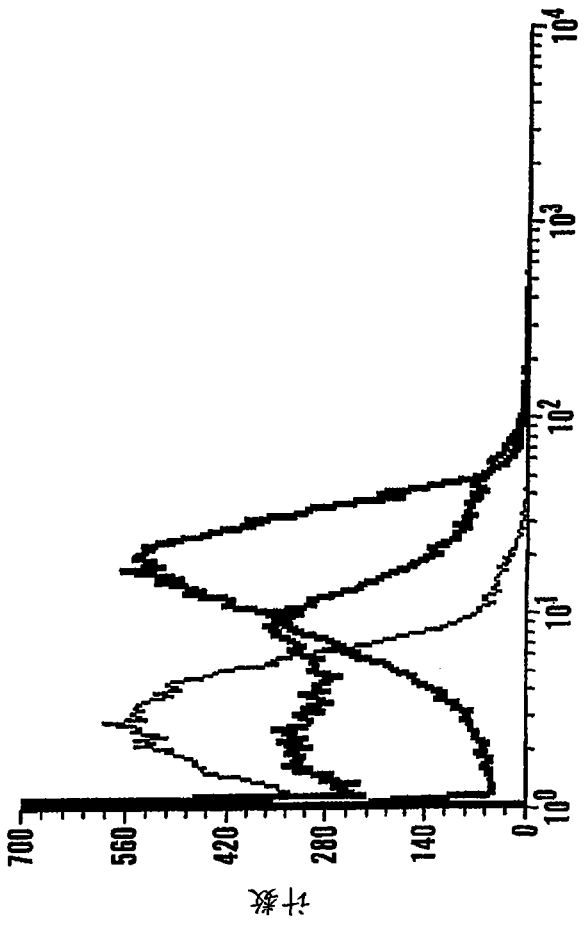


图12

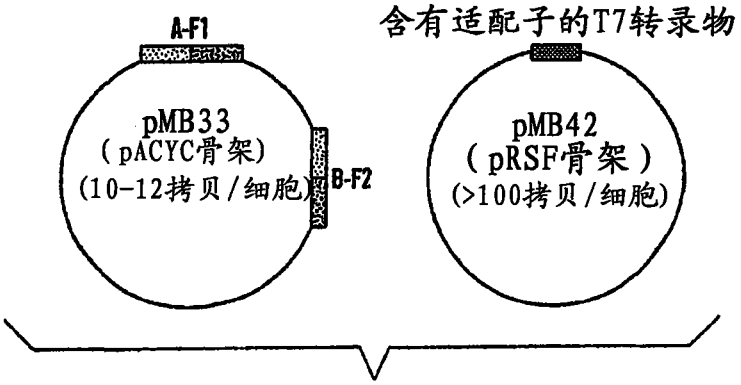


图13A

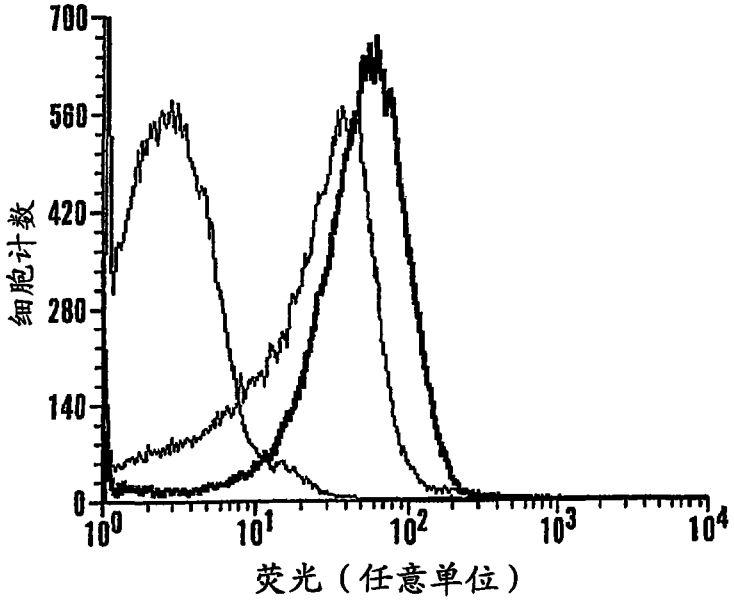


图13B



图13C

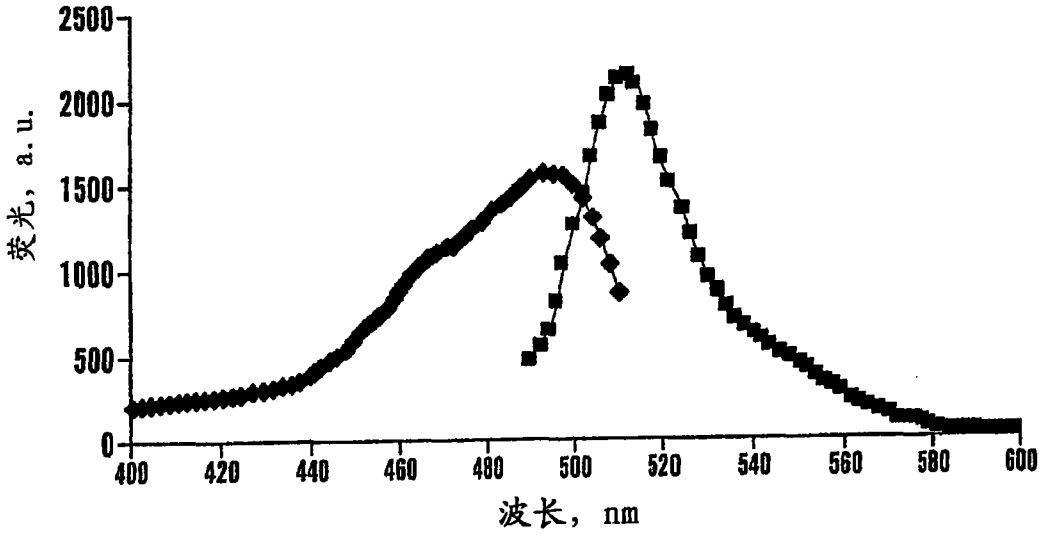

2500
2000
1500
1000
500
0
荧光，a.u.
400
420
440
460
480
500
520
540
560
580
600
波长，nm

图14A

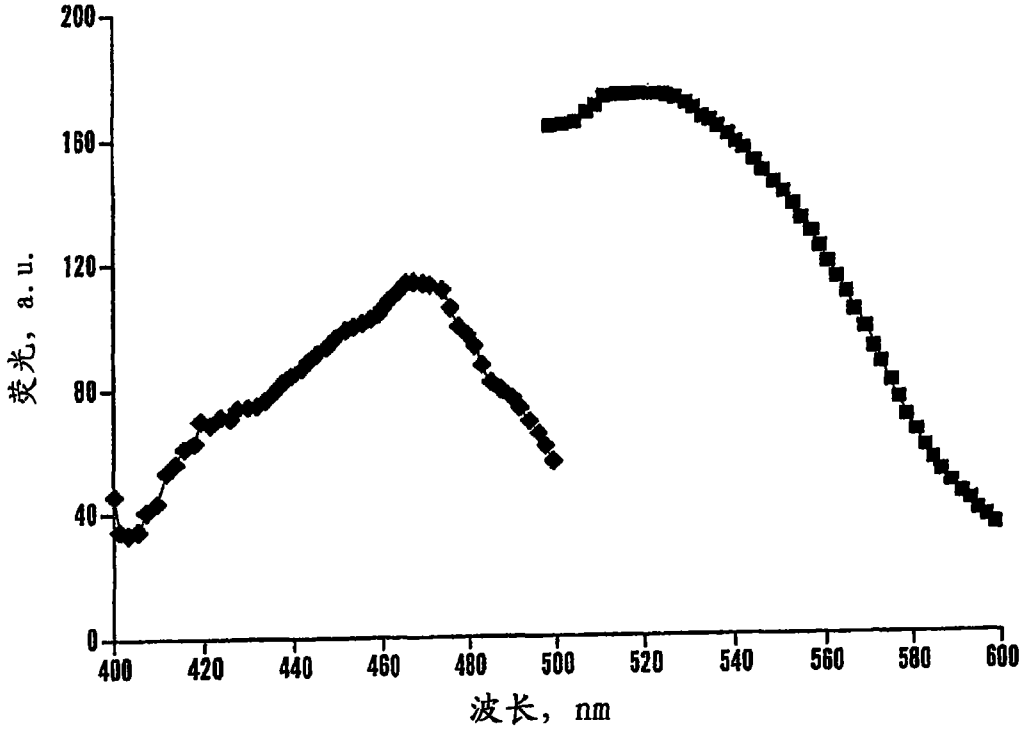

200
160
120
80
40
0
荧光，a.u.
400
420
440
460
480
500
520
540
560
580
600
波长，nm

图14B

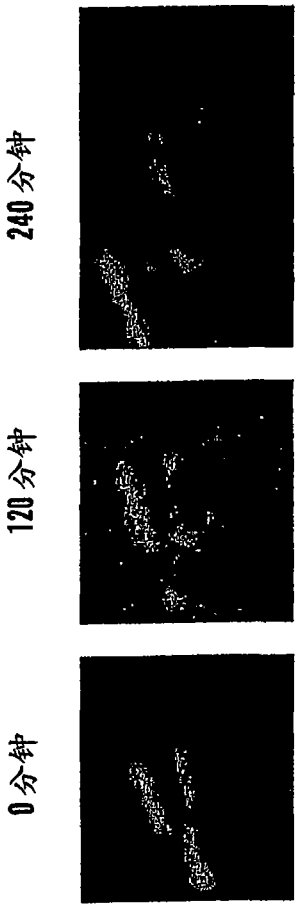


图15A



图15B

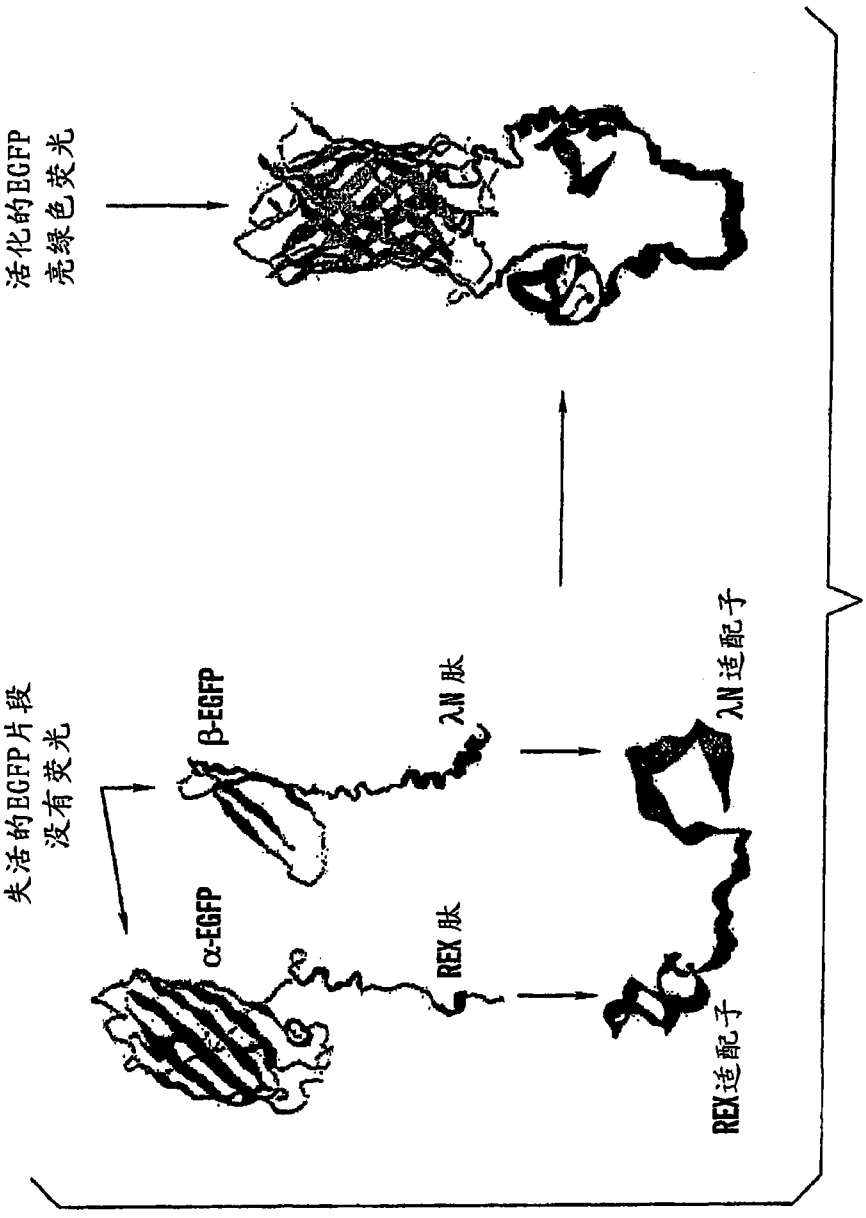

失活的EGFP片段
没有荧光
α-EGFP
β-EGFP
REX 肽
λN 肽
REX适配子
λN适配子
活化的EGFP
亮绿色荧光

图16

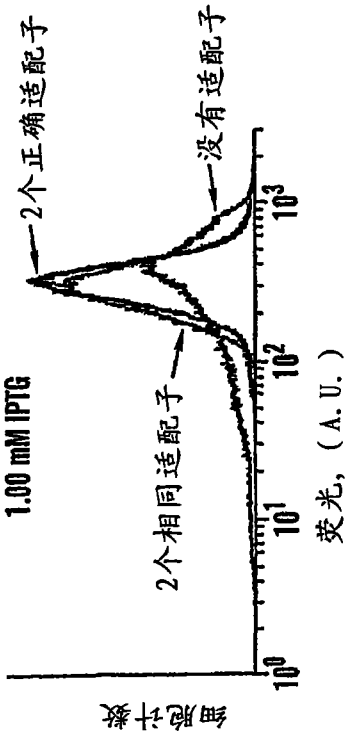

1.00 mM IPTG
2个正确适配子
没有适配子
2个相同适配子
细胞计数
荧光，（A.U.）

图17A

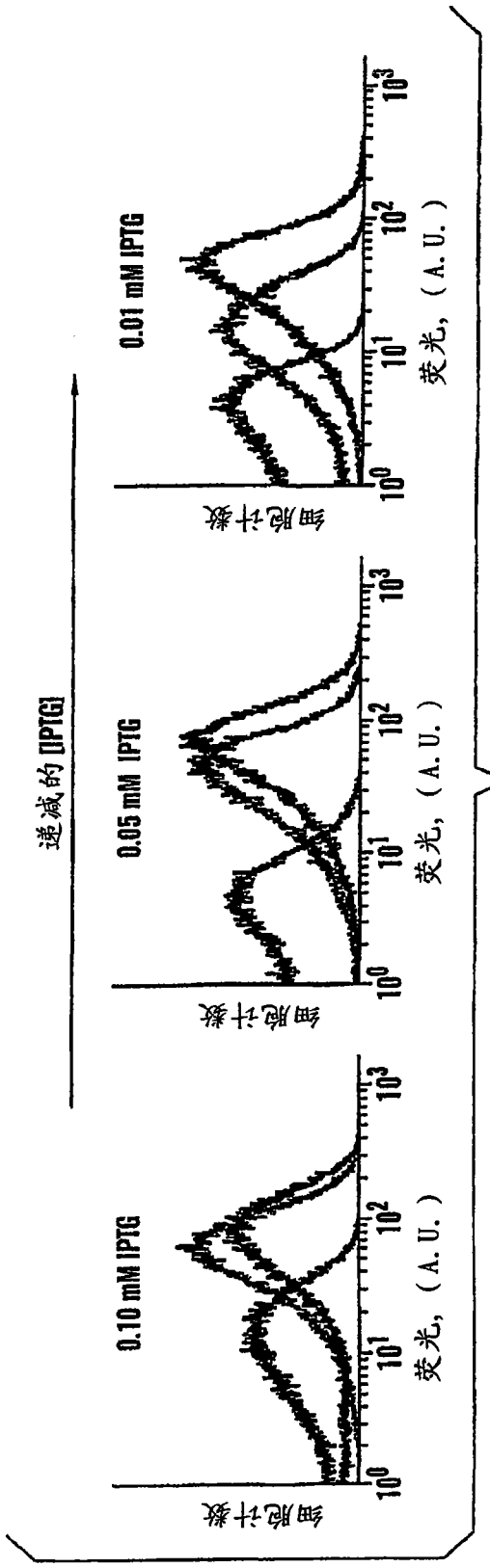

递减的IPTG
0.10 mM IPTG
0.05 mM IPTG
0.01 mM IPTG
细胞计数
荧光，（A.U.）

图17B

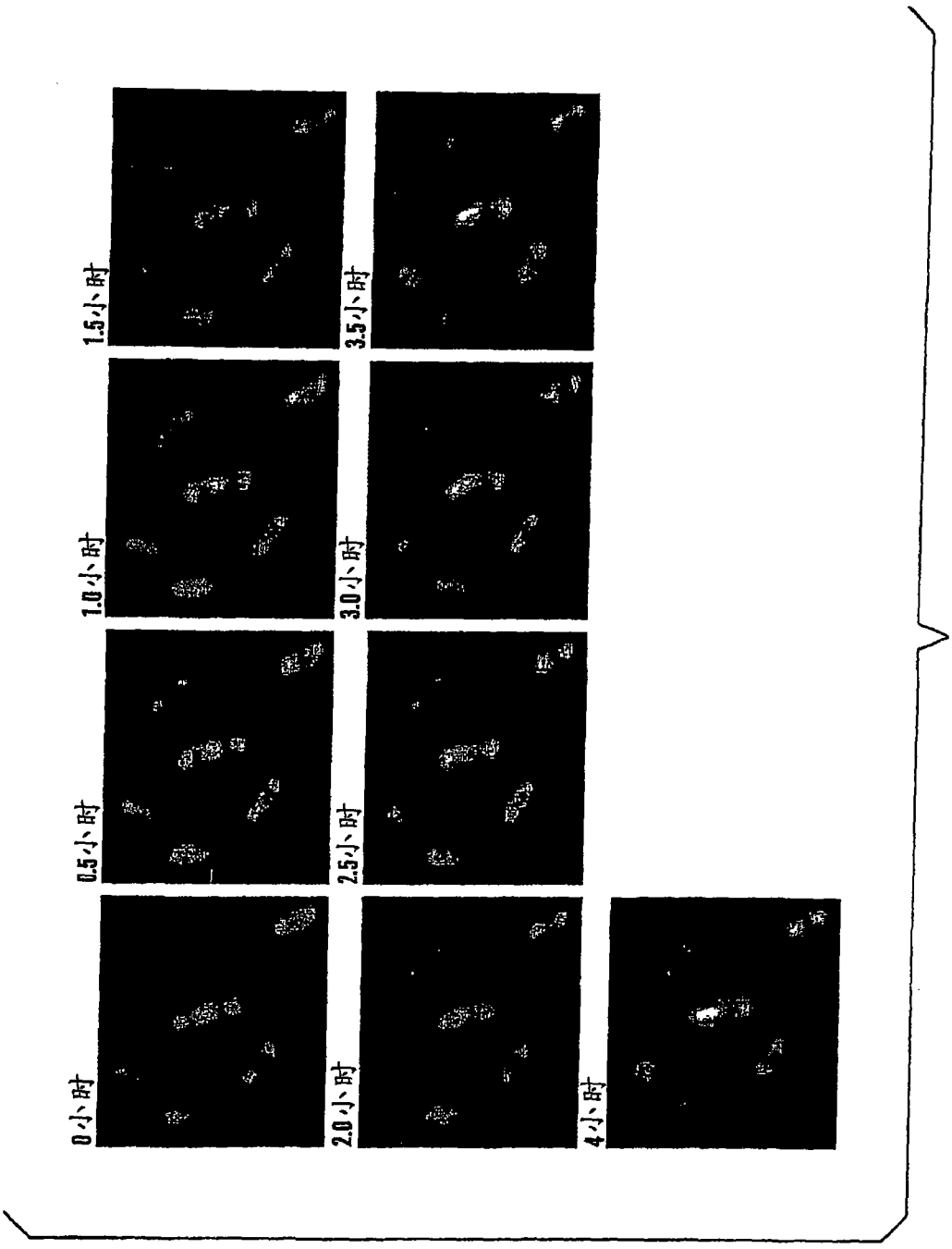


图18

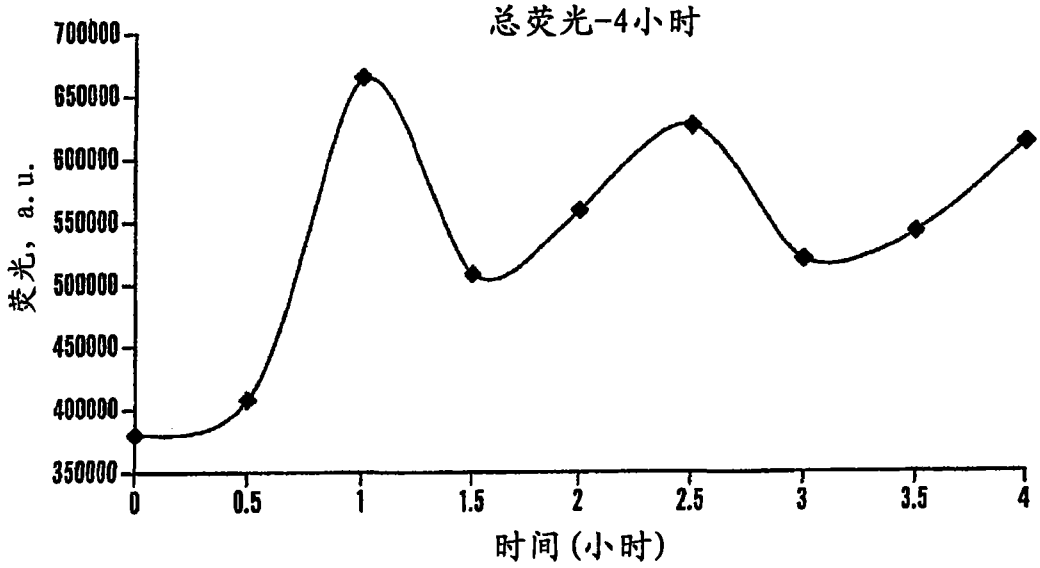

总荧光-4小时
700000
650000
600000
550000
500000
450000
400000
350000
荧光，a.u.
0
0.5
1
1.5
2
2.5
3
3.5
4
时间(小时)

图19A

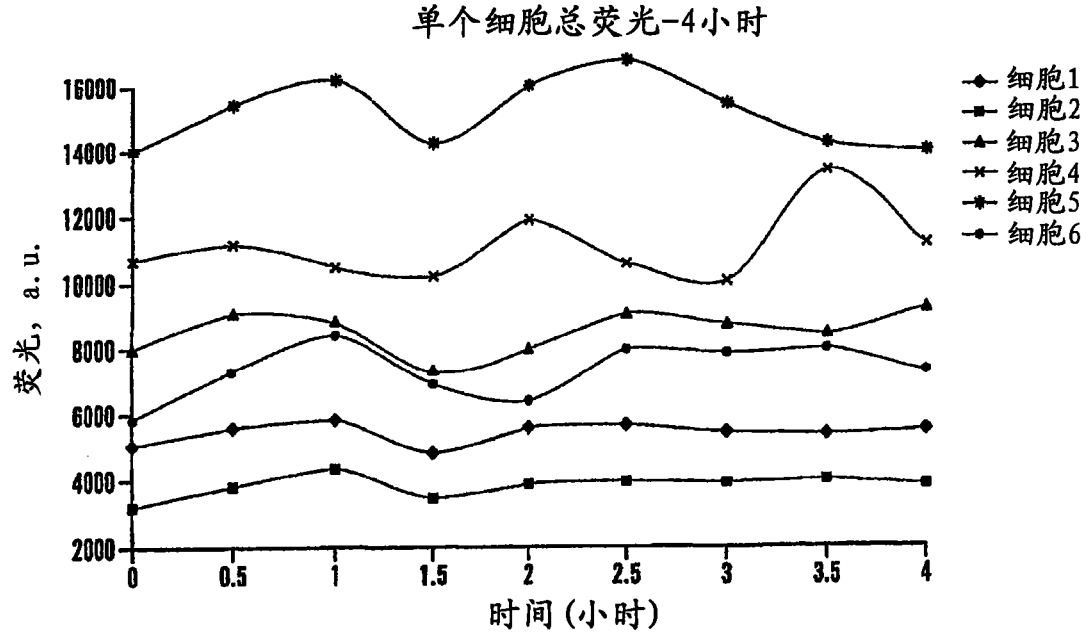

单个细胞总荧光-4小时
16000
14000
12000
10000
8000
6000
4000
2000
荧光，a.u.
0
0.5
1
1.5
2
2.5
3
3.5
4
时间(小时)
细胞1
细胞2
细胞3
细胞4
细胞5
细胞6

图19B

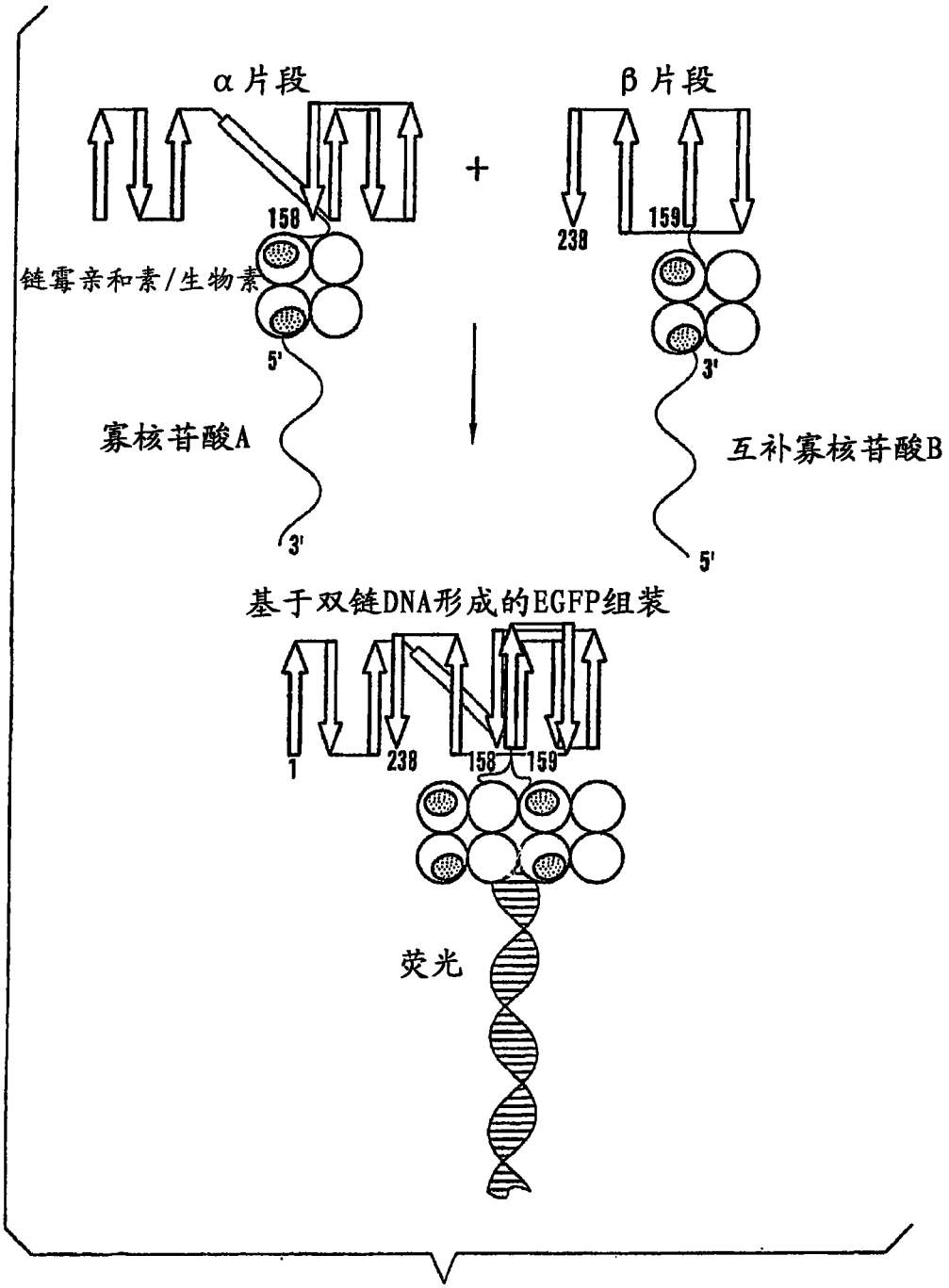

α片段
β片段
+
158
159
238
链霉亲和素/生物素
5'
3'
寡核苷酸A
互补寡核苷酸B
3'
5'
基于双链DNA形成的EGFP组装
1
238
158
159
荧光

图20A

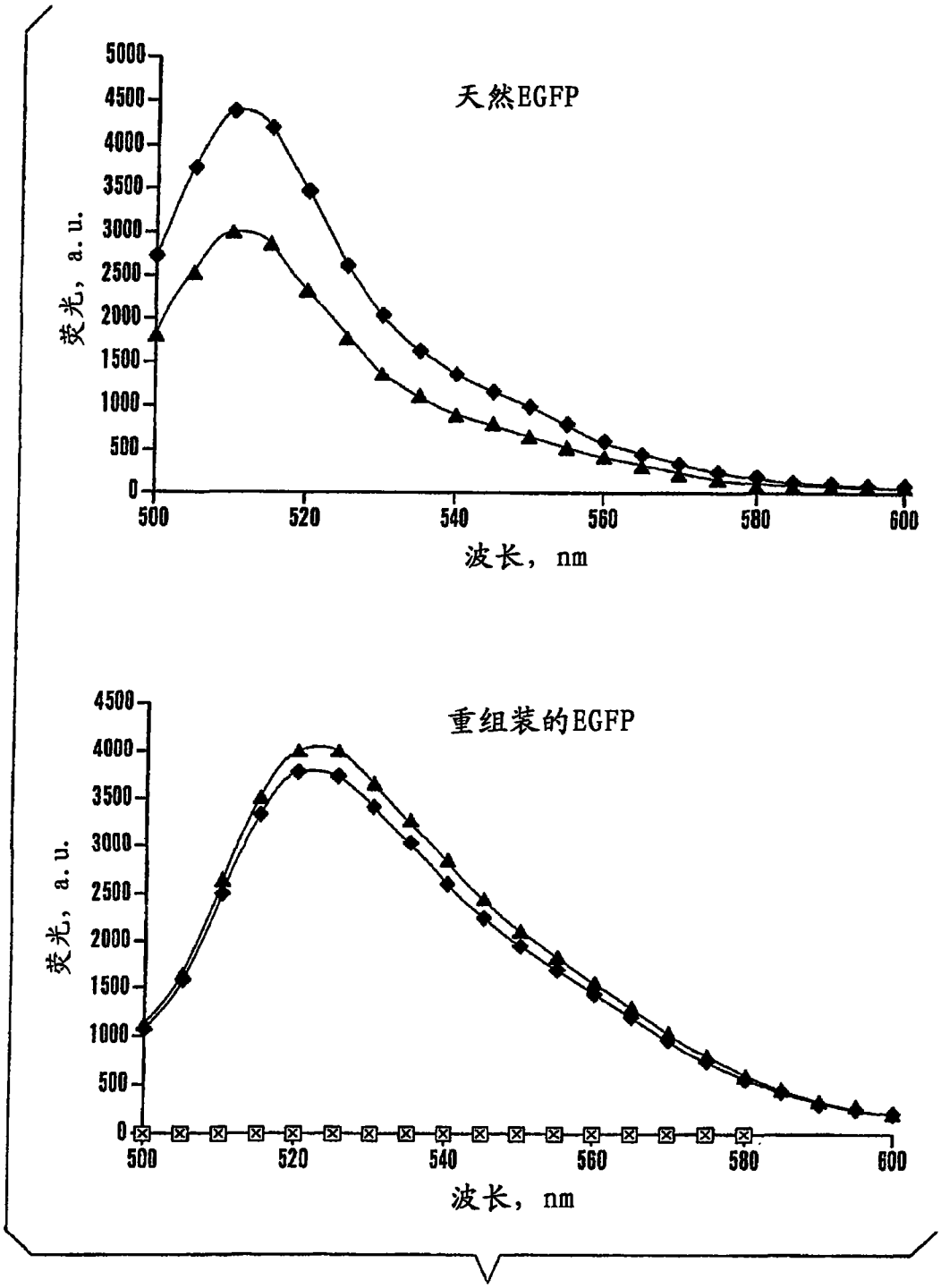

天然EGFP
荧光，a.u.
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
500
520
540
560
580
600
波长，nm
重组装的EGFP
荧光，a.u.
4500
4000
3500
3000
2500
2000
1500
1000
500
0
500
520
540
560
580
600
波长，nm

图20B

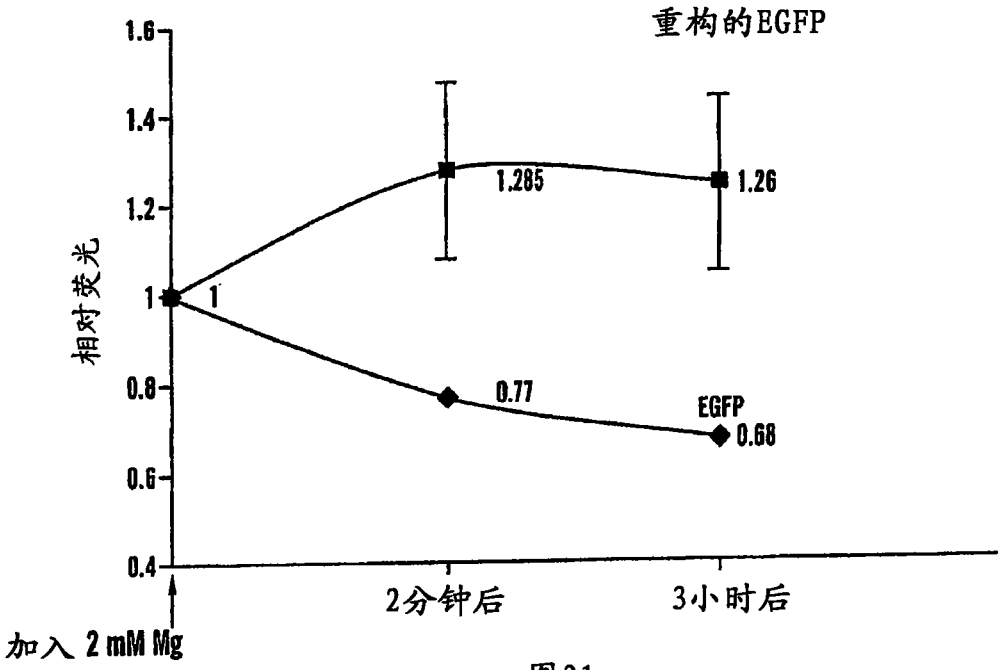

重构的EGFP
1.6
1.4
1.2
1
0.8
0.6
0.4
相对荧光
1.285
1.26
1
0.77
EGFP
0.68
2分钟后
3小时后
加入 2 mM Mg

图21

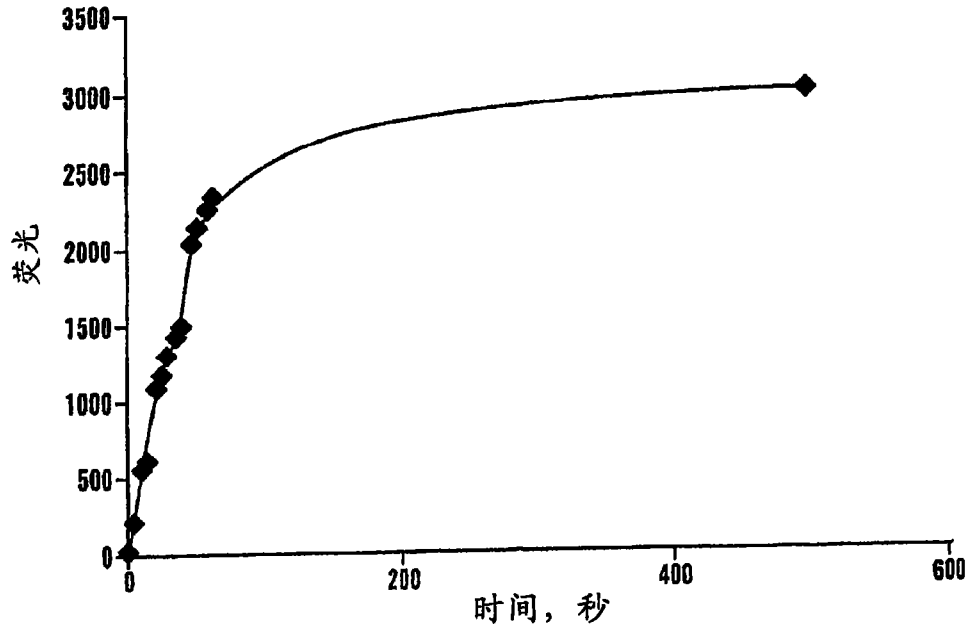

3500
3000
2500
2000
1500
1000
500
0
荧光
0
200
400
600
时间，秒

图22A

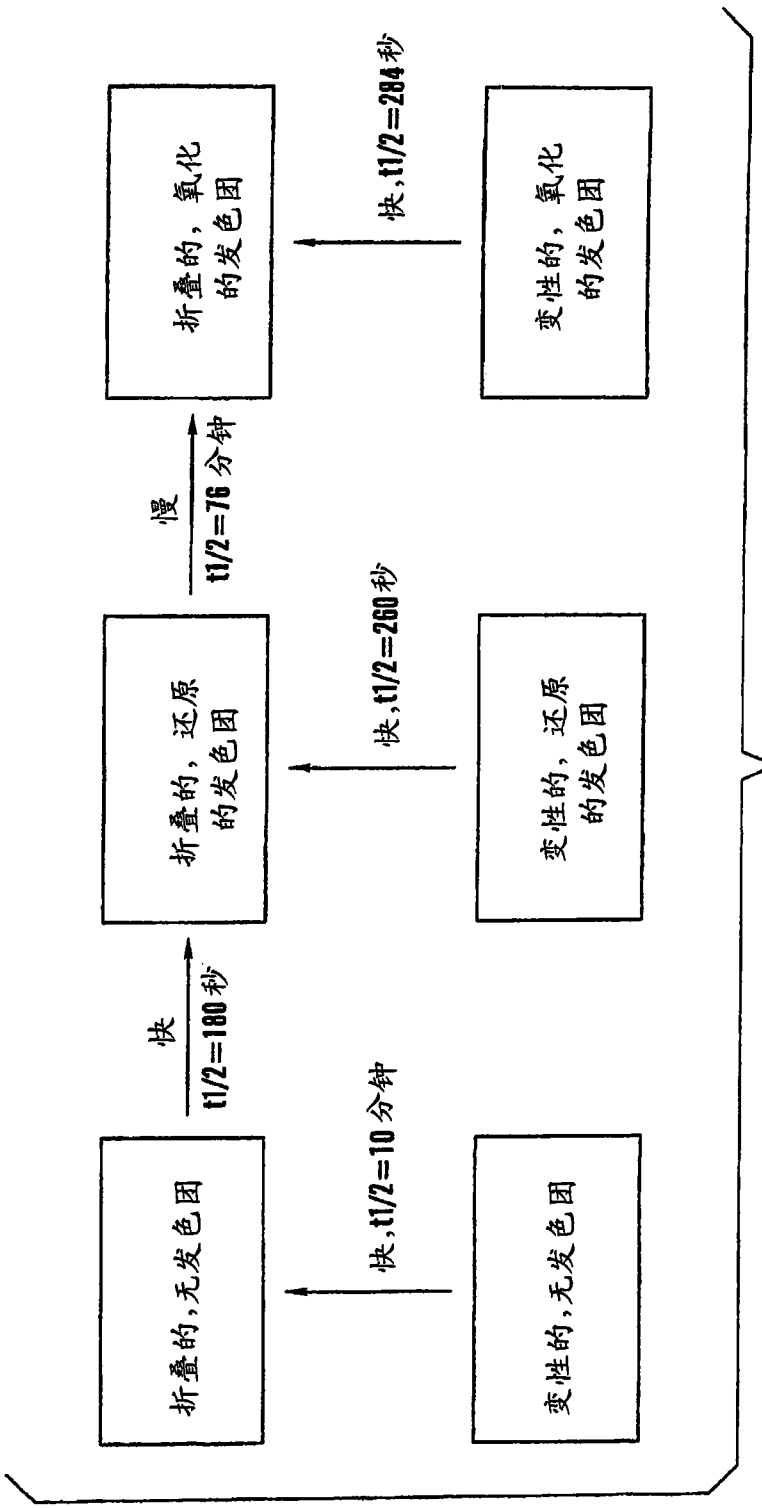

折叠的，无发色团
变性的，无发色团
快
t1/2=180秒
快，t1/2=10分钟
折叠的，还原
的发色团
变性的，还原
的发色团
慢
t1/2=76分钟
快，t1/2=260秒
折叠的，氧化
的发色团
变性的，氧化
的发色团
快，t1/2=284秒

图22B

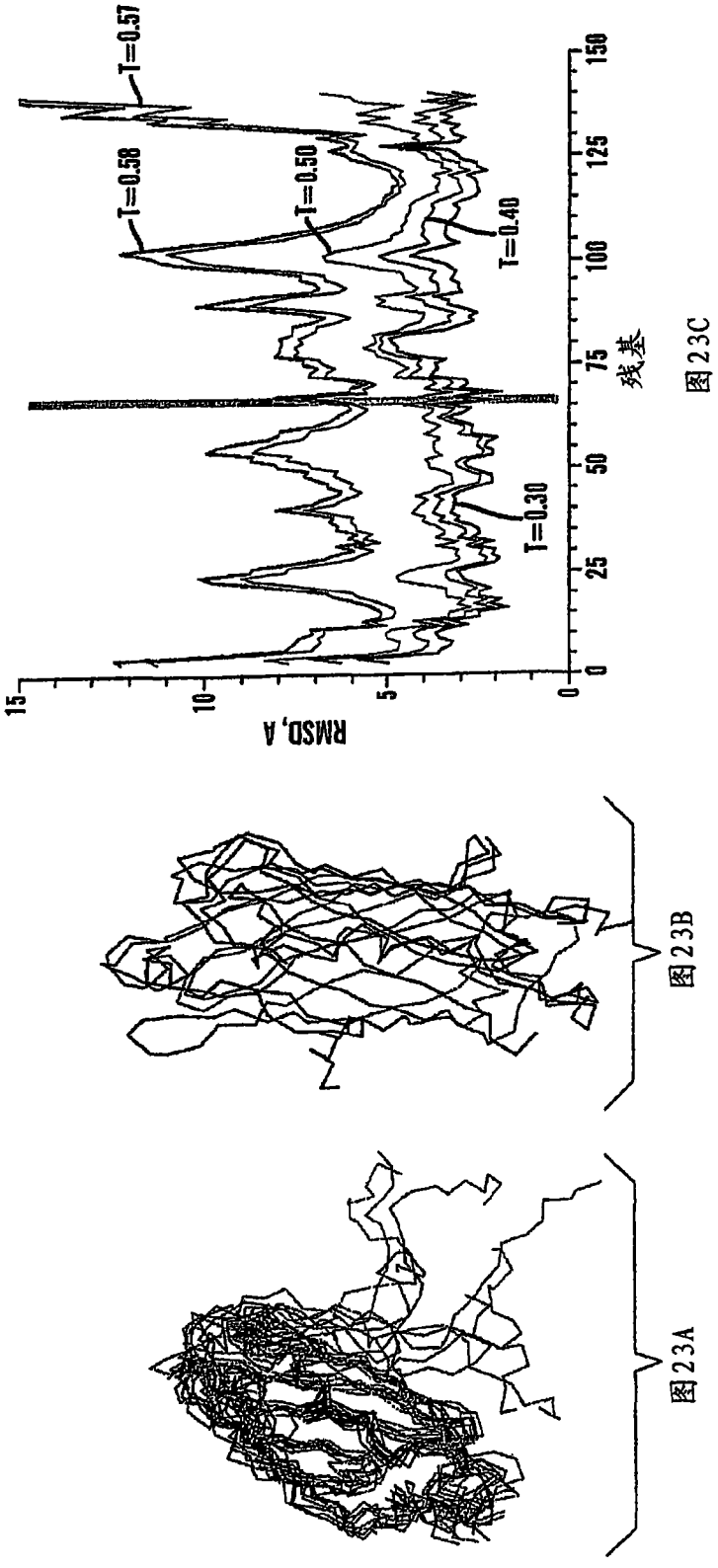


图23C

图23B

图23A

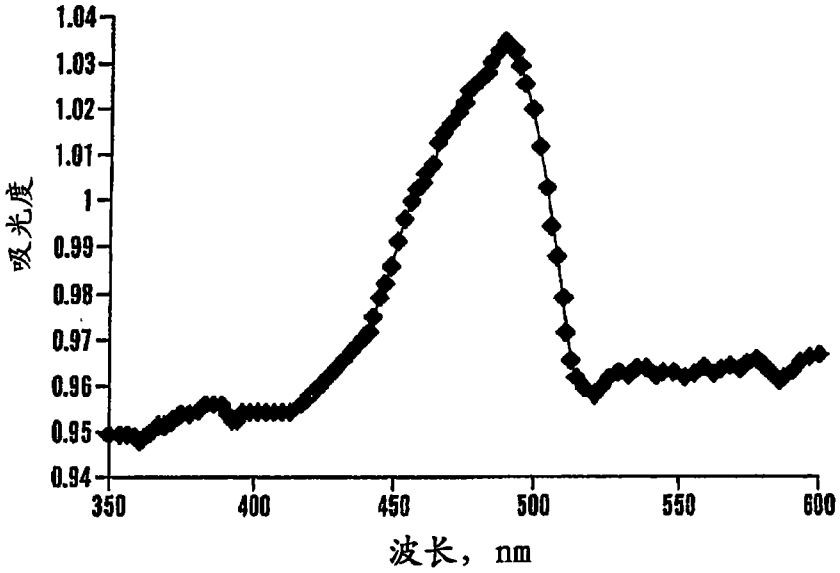

1.04
1.03
1.02
1.01
1
0.99
0.98
0.97
0.96
0.95
0.94
吸光度
350
400
450
500
550
600
波长，nm

图24A

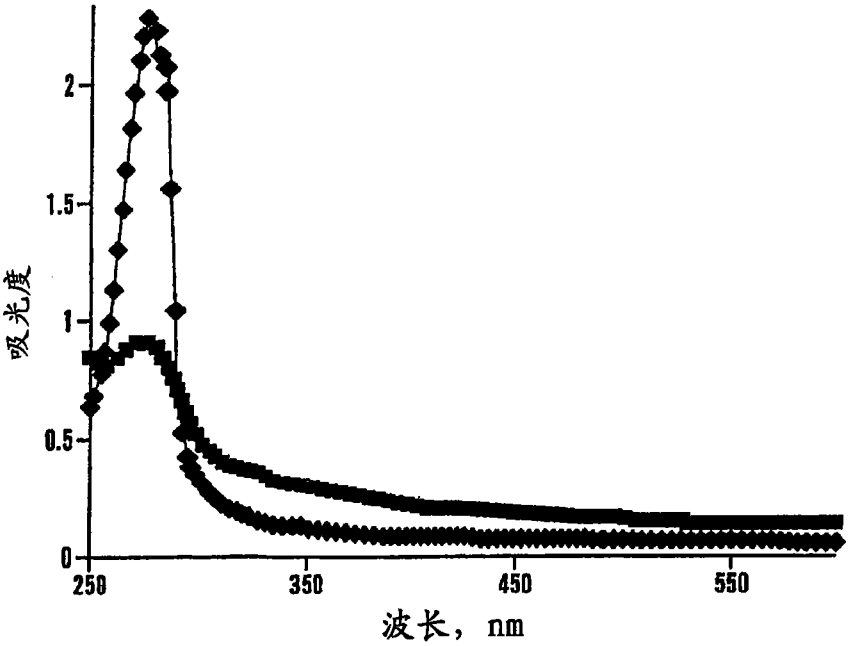

2
1.5
1
0.5
0
吸光度
250
350
450
550
波长，nm

图24B

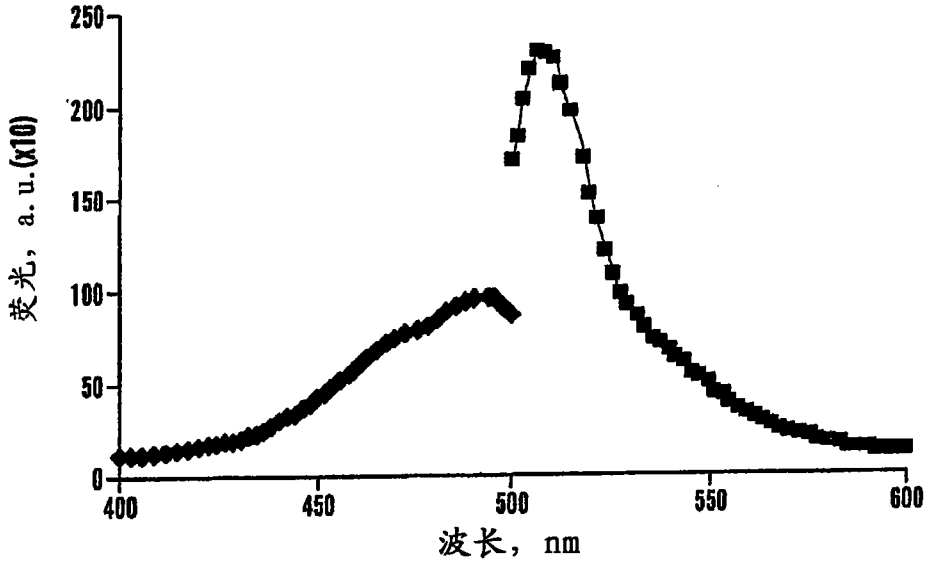

250
200
150
100
50
0
荧光，a.u.(x10)
400
450
500
550
600
波长，nm

图24C

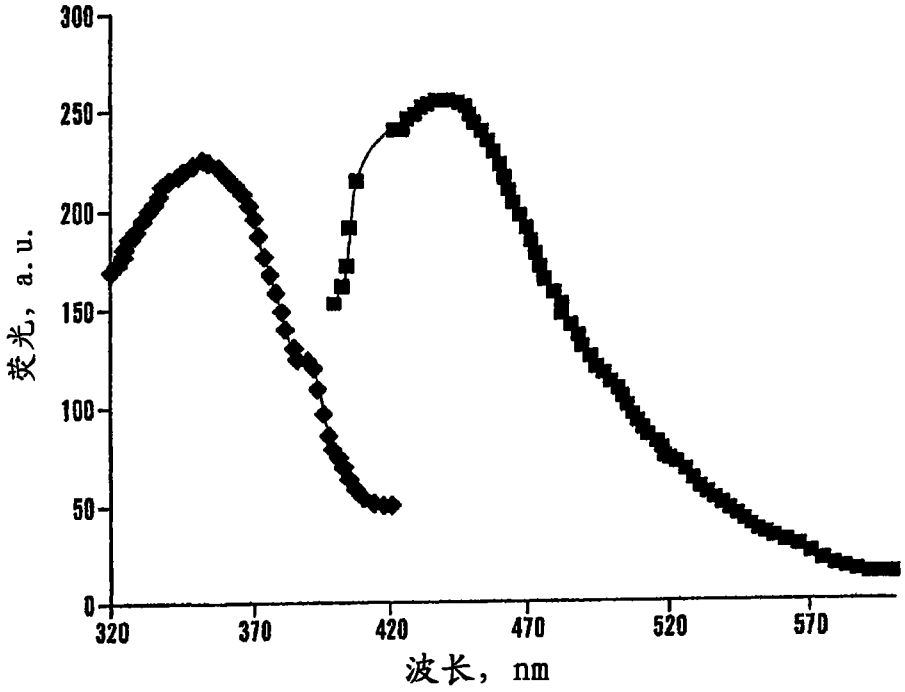

300
250
200
150
100
50
0
荧光，a.u.
320
370
420
470
520
570
波长，nm

图24D

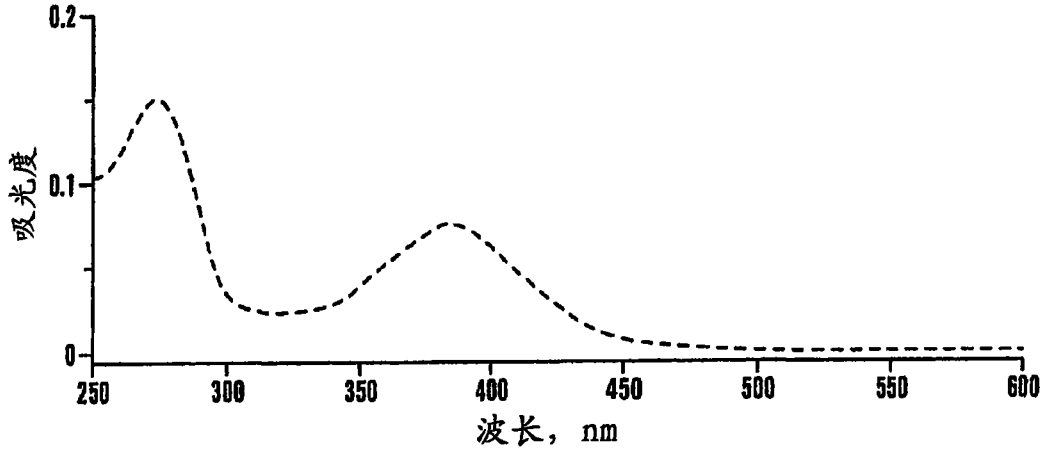


图25A

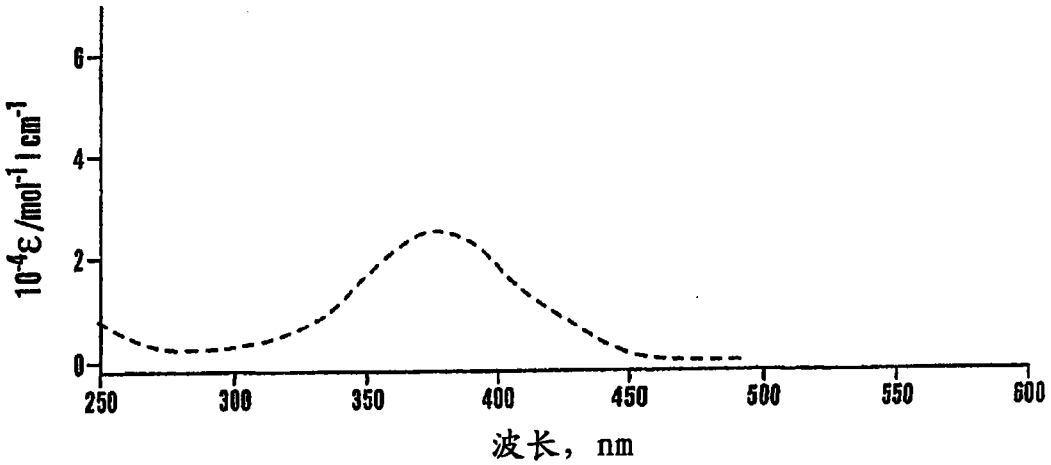


图25B

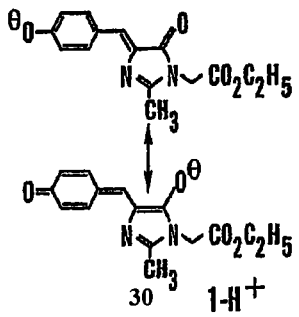


图25C

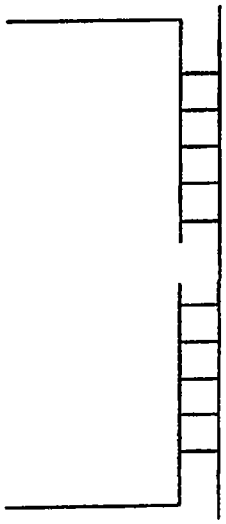


图26B

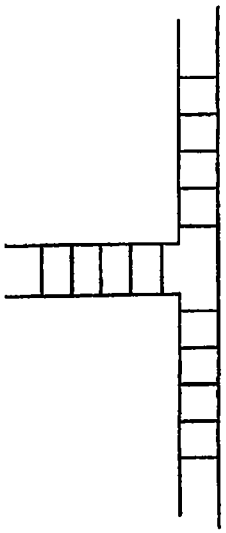


图26A

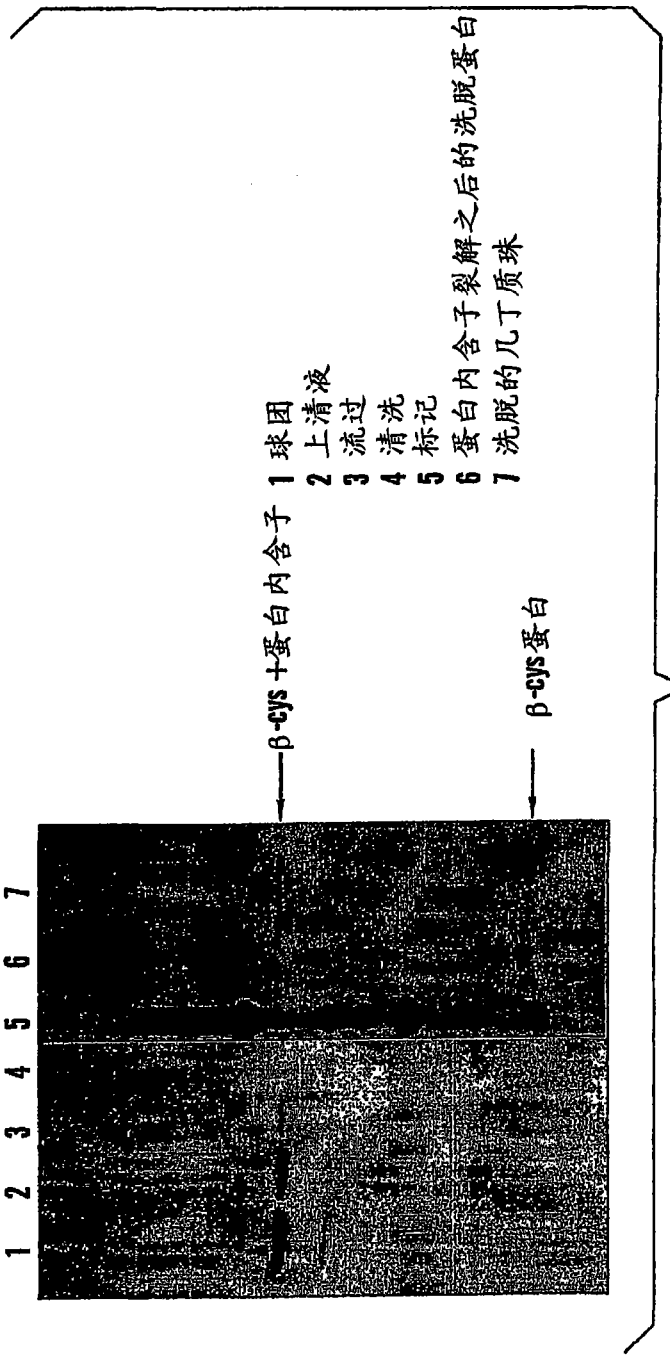

1 2 3 4 5 6 7
β-cys +蛋白内含子
β-cys蛋白
1 球团
2 上清液
3 流过
4 清洗
5 标记
6 蛋白内含子裂解之后的洗脱蛋白
7 洗脱的几丁质珠

图27

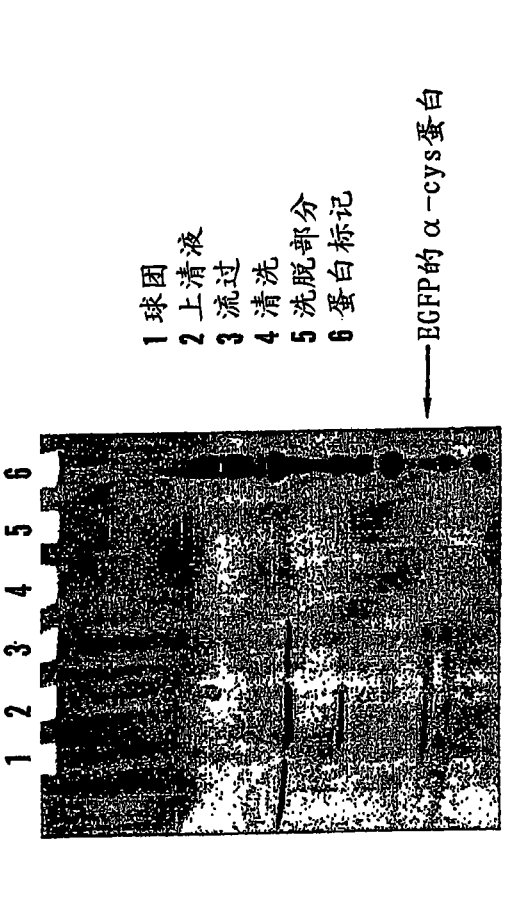


图28

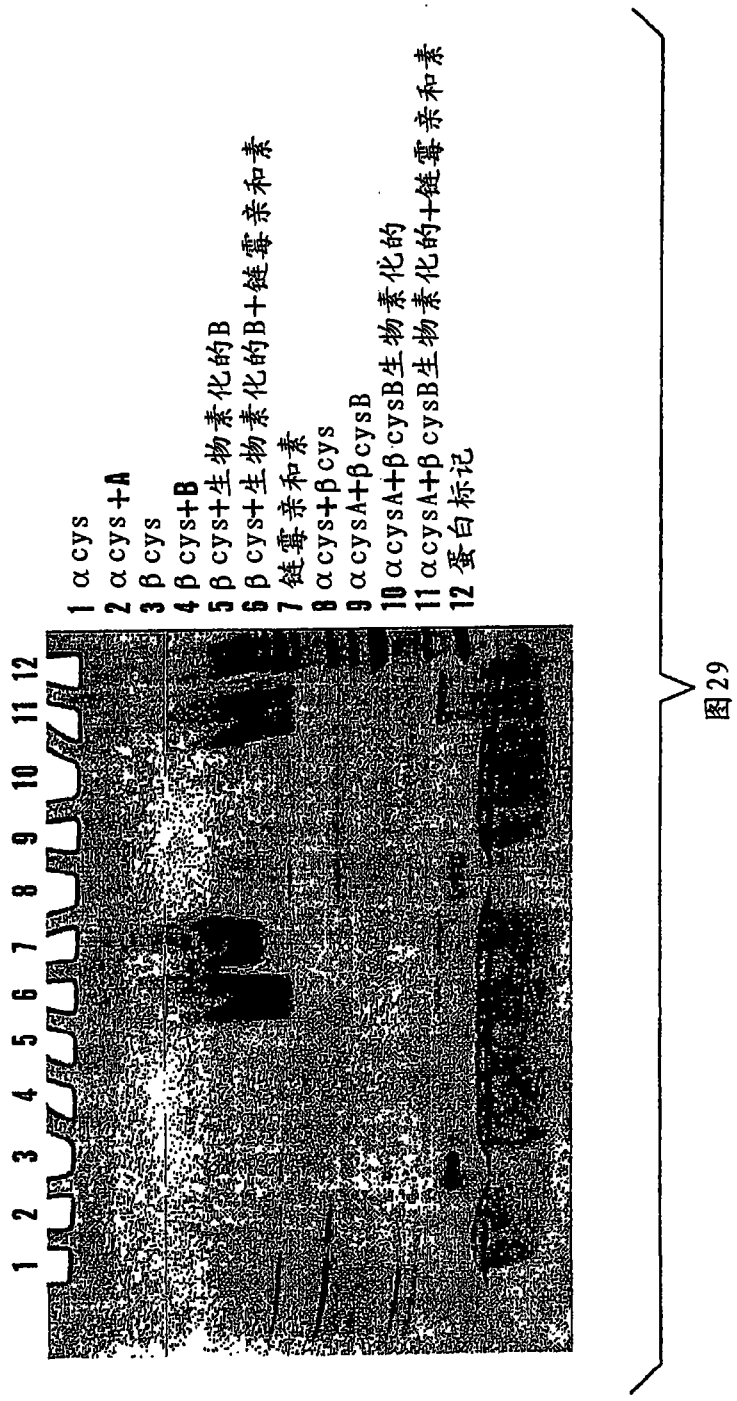

1 2 3 4 5 6 7 8 9 10 11 12
1 α cys
2 α cys +A
3 β cys
4 β cys+B
5 β cys+生物素化的B
6 β cys+生物素化的B+链霉亲和素
7 链霉亲和素
8 α cys+β cys
9 α cysA+β cysB
10 α cysA+β cysB生物素化的
11 α cysA+β cysB生物素化的+链霉亲和素
12 蛋白标记

图29

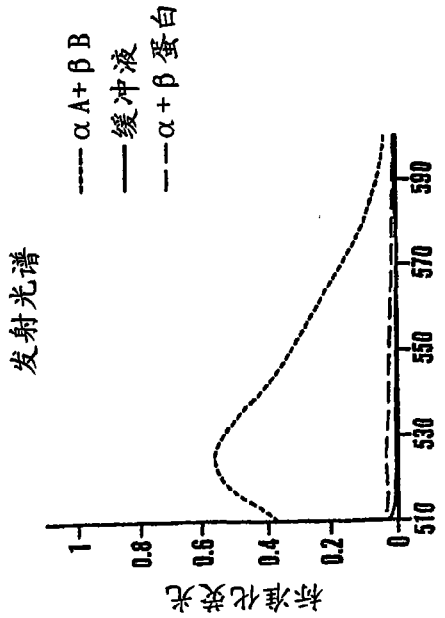

发射光谱
αA+βB
缓冲液
α+β蛋白
标准化荧光
1
0.8
0.6
0.4
0.2
0
510
530
550
570
590
波长(nm)

图30B

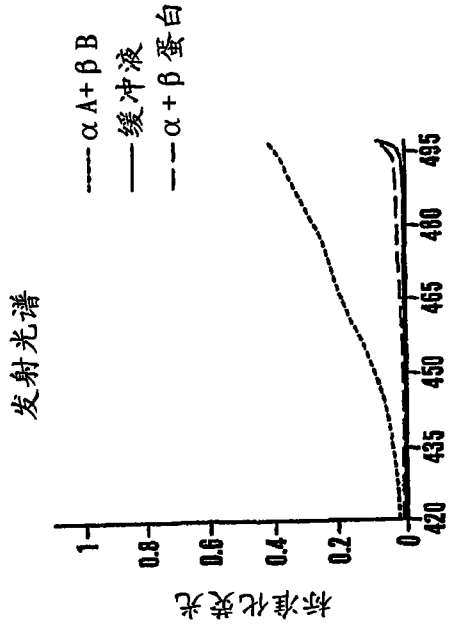

发射光谱
αA+βB
缓冲液
α+β蛋白
标准化荧光
1
0.8
0.6
0.4
0.2
0
420
435
450
465
480
495
波长(nm)

图30A

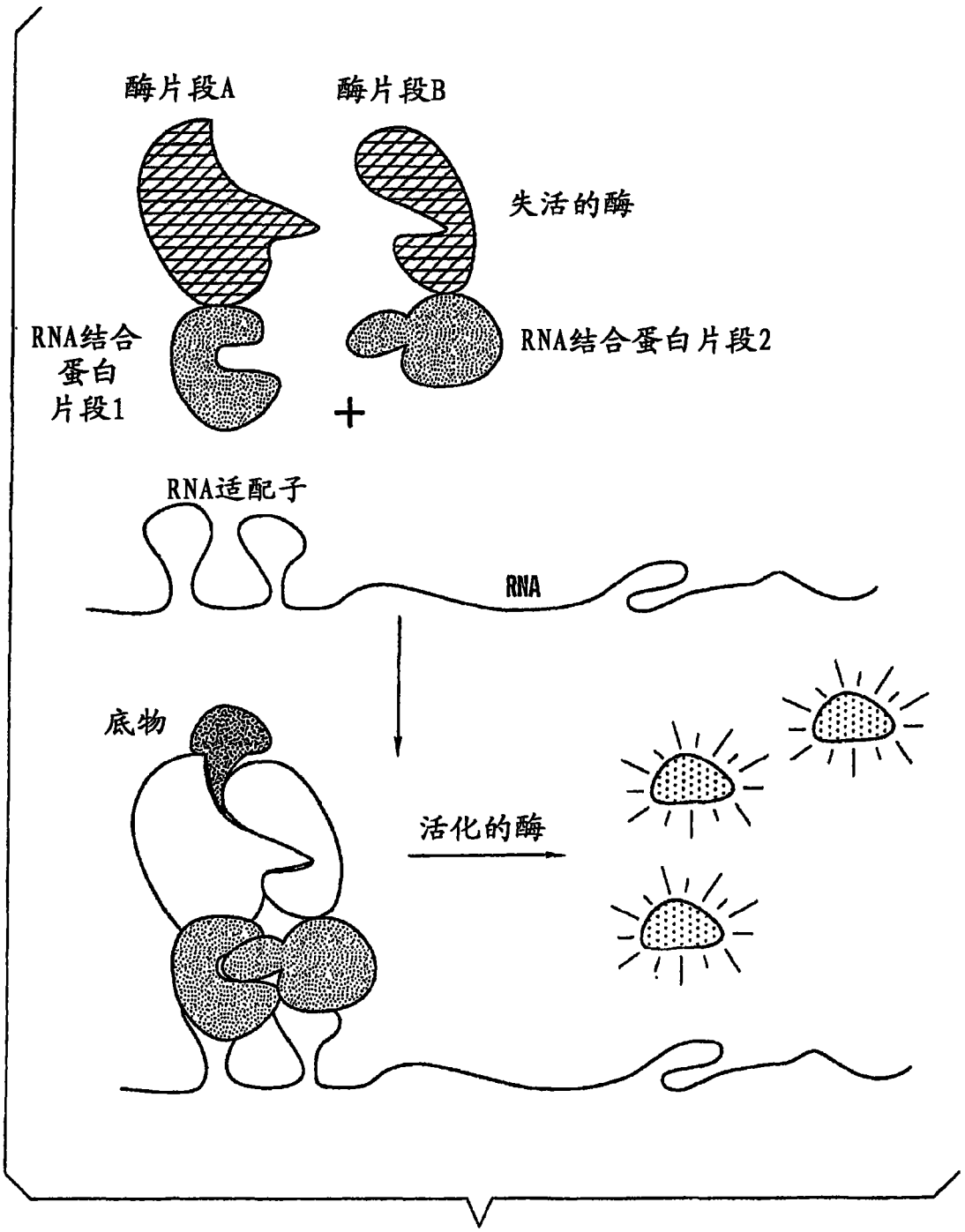


图31