

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2022年10月13日 (13.10.2022)



(10) 国际公布号
WO 2022/213253 A1

(51) 国际专利分类号:
C12N 15/55 (2006.01) *C12Q 1/6869* (2018.01)
C12N 9/14 (2006.01) *C12M 1/34* (2006.01)
G01N 33/48 (2006.01)

(21) 国际申请号: PCT/CN2021/085609

(22) 国际申请日: 2021年4月6日 (06.04.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 成都齐碳科技有限公司 (QITAN TECHNOLOGY LTD., CHENGDU) [CN/CN]; 中国四川省成都市中国(四川)自由贸易试验区成都高新区天府五街200号7栋A区2楼, Sichuan 610041 (CN)。

(72) 发明人: 张周刚 (ZHANG, Zhougang); 中国四川省成都市中国(四川)自由贸易试验区成都高新区天府五街200号7栋A区2楼, Sichuan 610093 (CN)。李文 (LI, Wen); 中国四川省成都市中国(四川)自由贸易试验区成都高新区天府五街200号7栋A区2楼, Sichuan 610093 (CN)。王艳双 (WANG, Yanshuang); 中国四川省成都市中国(四川)自由贸易试验区成都高新区天府五

街200号7栋A区2楼, Sichuan 610093 (CN)。王慕旻 (WANG, Muyang); 中国四川省成都市中国(四川)自由贸易试验区成都高新区天府五街200号7栋A区2楼, Sichuan 610093 (CN)。

(74) 代理人: 北京东方亿思知识产权代理有限公司 (BEIJING EAST IP LTD.); 中国北京市东城区东长安街1号东方广场东方经贸城东2座1601室, Beijing 100738 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,

(54) Title: MODIFIED PRP43 HELICASE AND USE THEREOF

(54) 发明名称: 经修饰的Prp43解旋酶及其用途

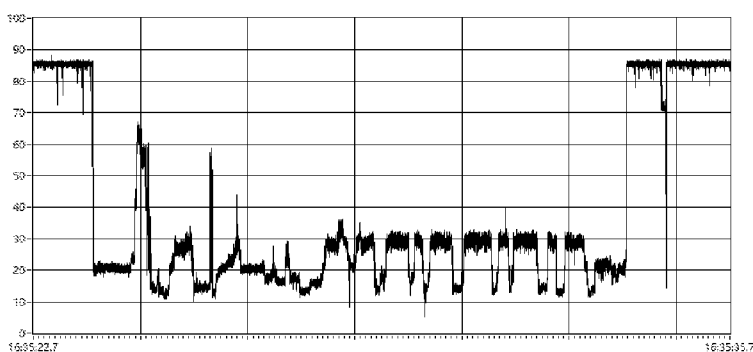


图 7

(57) Abstract: A modified Prp43 helicase and use thereof. The ATP hydrolytic activity or unwinding activity of the Prp43 helicase is enhanced due to the introduction of mutation and/or the introduction of accessory proteins, and the binding to a target polynucleotide can be maintained for a long time, thereby allowing the enzyme to continuously and stably control the moving speed of the polynucleotide at a suitable rate required by sequencing, and can be used for nanopore sequencing.

(57) 摘要: 一种经修饰的Prp43解旋酶及其用途。由于突变的引入和/或辅助蛋白的引入而增强了Prp43解旋酶的ATP水解活性或解旋活性, 能长时间保持与目标多核苷酸的结合, 从而允许酶以测序要求的合适速率、持续稳定地控制多核苷酸的移动速度, 可以用于纳米孔测序。



WO 2022/213253 A1

AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布：

- 包括国际检索报告(条约第21条(3))。
- 包括说明书序列表部分(细则5.2(a))。

经修饰的 Prp43 解旋酶及其用途

5 技术领域

本申请涉及核酸测序技术。

背景技术

10 纳米孔测序技术是第三代核酸测序技术，它是通过记录 DNA/RNA 链穿过纳米孔时不同碱基产生的不同电信号来获得 DNA/RNA 序列信息。纳米孔测序技术所面临的挑战之一是 DNA/RNA 分子通过纳米孔的速度往往太快、超过仪器的分辨率，因此难以获得反映序列信息的准确的电信号。因此，如何控制或减缓 DNA/RNA 分子通过纳米孔的速度对于提高纳米孔测序技术的准确率是至关重要的。

15 目前，新兴的一种表征多核苷酸的方法包括跨膜孔、解旋酶与多核苷酸的接触及相互作用，从而解旋酶控制目标多核苷酸穿过纳米孔的运动，以增加多核苷酸在纳米孔处的停留时间。

例如，专利 WO2013057495A3 和 US20150191709A1 公开了一种新的表征目标多核苷酸的方法，所述的方法使用孔和 Hel308 解旋酶或能结合目标多核苷酸内部核苷酸的分子马达。该发明所述的解旋酶或分子马达可以有效控制目标多核苷酸穿过所述孔的运动。另外，专利 US20150065354A1 和 US9617591B2 公开了一种使用 XPD 解旋酶表征目标多核苷酸的方法，所述方法利用孔和 XPD 解旋酶。该发明所述的 XPD 解旋酶可以控制目标多核苷酸穿过所述孔的运动。此外，专利 US 2016O257942A1 和
25 US20180179500 A1 则披露了 T4 噬菌体来源 Dda 解旋酶及其某些同源蛋白经过修饰后可以应用于多核苷酸过孔测序。

自然界中，解旋酶可分为六大超家族(Superfamily,SF),其中 SF1 和 SF2 超家族解旋酶以单体形式发挥移位和解旋功能，SF3- SF6 家族则以多聚体形式发挥作用。纳米孔测序应用中，单体形式作用的解旋酶应用更简便、表

现均一性也会更强。SF1 和 SF2 超家族的解旋酶根据蛋白序列同源性、结构域排列、底物结合形式和特异性、解旋的极性（5'-3'方向和 3'-5'方向）以及解旋或移位机制等特性分类为不同的家族，SF1 超家族包含 UvrD/Rep 家族解旋酶、类 Upf1 家族解旋酶和类 Pif1 家族解旋酶，SF2 超家族包含
5 Rad3/XPD 家族解旋酶、类 Ski2 家族解旋酶、DEAH/RHA 家族解旋酶、NS3/NPH-II 家族解旋酶、DEAD 盒（DEAD-Box）家族解旋酶、类 RIG-I 家族解旋酶、类 RecQ 家族解旋酶、类 RecG 家族解旋酶、Swi/Snf 家族解旋酶和 TIR 家族解旋酶。RecD 和 T4 噬菌体来源的 Dda 解旋酶同属于 SF1 超家族的类 Pif1 家族，底物倾向于单链 DNA，以 5'-3'方向进行移位和解旋，
10 常见的同属于该家族的解旋酶还包括 Pif1 解旋酶、TrwC 解旋酶等；*Methanococcoides burtonii* 菌株来源的 Hel308 解旋酶（如 US20150191709A1 中所披露）属于 SF2 超家族的类 ski2 家族解旋酶，能同时以单链 DNA 或 RNA 为底物，以 3'-5'方向极性移位或解旋双链核酸，常见的同属于该家族的解旋酶还包括 ski2 解旋酶、Brr2 解旋酶、Mtr4 解旋酶等；XPD 解旋酶则
15 属于 SF2 超家族的 Rad3/XPD 家族解旋酶，特异性结合单链 DNA，以 5'-3'极性方向移位或解旋双链核酸，常见的同属于该家族的酶还包括 Rad3 解旋酶等。

虽然现有技术中公开了多种可用于纳米孔测序技术的解旋酶，但是每种解旋酶各有其优缺点和各自的适用环境，这些解旋酶仍然难以满足科学研究和医疗技术等对于核酸测序技术在多方面提出的更苛刻的要求，因此
20 仍然存在对于可以用于核酸纳米孔测序技术的新型解旋酶的需要，以便改善纳米孔测序技术的适用性、准确性和灵敏度等。

发明内容

25 发明人发现 Prp43 解旋酶，尤其是经修饰的 Prp43 解旋酶可以控制多核苷酸分子移动穿过纳米孔，从而可以用于纳米孔测序技术。

因此，本申请的第一方面涉及一种经修饰的 Prp43 解旋酶，包括 RecA1 结构域、RecA2 结构域和 Ratchet 结构域，相对于对应的野生型 Prp43 解旋酶或其片段而言所述经修饰的 Prp43 解旋酶包括在选自 RecA1 结构域、

RecA2 结构域、Ratchet 结构域的至少一个结构域中引入的 1、2、3、4、5、6、7、8、9、10、11 个或更多个半胱氨酸的插入或置换，和/或 1、2、3、4、5、6、7、8、9、10、11 个或更多个非天然氨基酸的插入或置换。

5 本申请的第二方面涉及一种蛋白构建体，其包括本申请第一方面所述的经修饰的 Prp43 解旋酶，以及在所述 Prp43 解旋酶的 C 端或 N 端融合的辅助激活蛋白 Paf1 的 G-Path 结构域或含有 G-Path 结构域的 Paf1 片段。

本申请的第三方面涉及一种编码本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体的核酸。

10 本申请的第四方面涉及一种包含本申请的第三方面所述核酸的表达载体。

本申请的第五方面涉及本申请第三方面所述的核酸或包含本申请的第四方面所述表达载体的宿主细胞。

15 本申请的第六方面涉及一种制备本申请的第二方面所述的蛋白构建体的方法，包括：提供 SEQ ID NO: 1 的多肽或其变体和 SEQ ID NO: 26 的多肽或其变体，在 SEQ ID NO: 1 的多肽或其变体中引入至少一个半胱氨酸残基和/或至少一个非天然氨基酸，然后在所得多肽的 C 端或 N 端融合 SEQ ID NO: 26 的多肽或其变体，形成所述蛋白构建体。

20 本申请的第七方面涉及一种制备本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体的方法，包括：包括培养本申请第五方面所述的宿主细胞，并进行诱导表达，然后纯化所得表达产物。

本申请的第八方面涉及一种控制多核苷酸分子移动的方法，包括将所述多核苷酸分子与本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体接触。

25 本申请的第九方面涉及一种表征目标多核苷酸的方法，所述的方法包括：

(a) 将目标多核苷酸与本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体接触，使得所述 Prp43 解旋酶或蛋白构建体控制所述目标多核苷酸移动穿过一纳米孔； (b) 获取目标多核苷酸中的核苷酸与所述纳米孔相互作用时的一个或多个特征，从而表征所述目标

多核苷酸。

本申请的第十方面涉及本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体在表征目标多核苷酸或控制目标多核苷酸穿过孔的移动中的用途。

5 本申请的第十一方面涉及一种用于表征目标多核苷酸的分析装置，所述的分析装置包含一个或多个纳米孔、一个或多个本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体、以及一个或多个容器。

本申请的第十二方面涉及一种形成表征目标多核苷酸的传感器的方法，
10 包括提供纳米孔，和在所述纳米孔和本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体之间形成复合物。

本申请提供了一种可用于核酸纳米孔测序的新型 Prp43 解旋酶突变体或其构建体，由于突变的引入和/或辅助蛋白的引入而增强了其 ATP 水解活性或解旋活性，并且/或者能长时间保持与目标多核苷酸的结合，从而允许
15 持续稳定地控制多核苷酸的移动速度。因此，本申请的 Prp43 解旋酶突变体或其构建体可以在测序要求的合适速率下持续地控制目标多核苷酸穿过孔的移动，进而改善纳米孔测序的通量和准确率。

具体实施方式

20 定义

为了更清楚地解释本发明的实施方式，本文中使用了某些学术术语和专有名词。除非在本文中进行了明确定义，所有这些术语和名词应当被理解为具有本领域技术人员所通常理解的含义。为了更清楚起见，对于本文中使用的某些术语进行了以下定义。

25 术语“多肽”是指包含通过肽键连接的氨基酸残基并含有多于五个氨基酸残基的分子。多肽通常可以包含 20 个或更多个氨基酸，优选地包含 50 个或更多个氨基酸，或者包含 100 个或更多个氨基酸。在本文中，术语“蛋白质”、“蛋白”与术语“多肽”被认为具有相同的含义；因此，术语“蛋白质”、“蛋白”和“多肽”可互换使用。可任选地修饰（例如，糖基化、

磷酸化、酰化、法尼基化、异戊烯基化、磺化等)多肽以增加其官能性或活性。在某些条件下、在特定底物存在下表现出活性的多肽可称为“酶”。应当理解,由于遗传密码的简并性,可以产生编码给定多肽的多种核苷酸序列。

5 本文所述的“核酸”是脱氧核糖核酸(DNA)和核糖核酸(RNA)的总称,是由许多核苷酸单体聚合成的生物大分子化合物。在本文中,术语“核酸”与术语“多核苷酸”被认为具有相同的含义;因此,术语“核酸”与术语“多核苷酸”可互换使用。

核苷酸单体由五碳糖、磷酸基和含氮碱基组成。如果五碳糖是核糖,则形成的聚合物是RNA;如果五碳糖是脱氧核糖,则形成的聚合物是DNA。
10 核苷酸中的含氮碱基可以包括但不限于:腺嘌呤(A)、鸟嘌呤(G)、胸腺嘧啶(T)、尿嘧啶(U)和胞嘧啶(C)。所述核苷酸可以是天然存在的或人工合成的。因此,本文所述的“核苷酸”包括但不限于:腺苷单磷酸(AMP)、鸟苷单磷酸(GMP)、胸苷单磷酸(TMP)、尿苷单磷酸(UMP)、胞嘧啶核苷单磷酸(CMP)、环状腺苷单磷酸(cAMP)、环状鸟苷单磷酸(cGMP)脱氧腺苷单磷酸(dAMP)、脱氧鸟苷单磷酸(dGMP)、脱氧胸苷单磷酸(dTMP)、
15 脱氧尿苷单磷酸(dUMP)和脱氧胞苷单磷酸(dCMP)。优选的,所述核苷酸选自AMP、TMP、GMP、CMP、UMP、dAMP、dTMP、dGMP或dCMP。

在本申请中,多肽或多肽结构域的“片段”是指在所述多肽或多肽结构域的氨基和/或羧基末端有一个或多个(例如,几个、几十个或100个等等)
20 氨基酸残基缺失的多肽或多肽结构域,但该片段仍保留有所期望的活性。例如,Prp43解旋酶的片段表示在野生型Prp43的氨基和/或羧基末端有一个或多个(例如,1-5个、1-10个、1-20个、1-50个、1-100个、1-150个、1-200个,或者例如20个、30个、40个、50个、60个、70个、80个或90个)氨基酸残基缺失但仍保留有解旋酶活性的多肽序列。

25 通常,多肽或结构域的片段包含其原始序列长度的至少50%、55%、60%、65%、70%、75%、80%、85%、90%、92%、94%、96%、98%或99%的长度。在本申请中,取决于原始多肽或结构域的长度,多肽或结构域的片段包含至少50个氨基酸,例如至少60个氨基酸、至少70个氨基酸、至少80个氨基酸、至少90个氨基酸、至少100个氨基酸、至少150个氨基酸、

至少 200 个氨基酸、至少 250 个氨基酸、至少 300 个氨基酸、至少 350 个氨基酸、至少 400 个氨基酸、至少 500 个氨基酸、至少 650 个氨基酸、或至少 700 个氨基酸。在本申请中，多肽或结构域的片段也可以包含少于 700 个氨基酸，例如少于 600 个氨基酸、少于 500 个氨基酸、少于 400 个氨基酸、少于 300 个氨基酸、少于 200 个氨基酸、或少于 100 个氨基酸。

术语“表达”包括参与多肽产生的任何步骤，包括但不限于转录、转录后修饰、翻译、翻译后修饰和分泌。

“表达载体”包含编码多肽的多核苷酸，该多核苷酸可操作地连接至适当的控制序列（例如启动子，以及转录和翻译终止信号）以用于在体外表达和/或翻译。表达载体可以是任何载体（例如，质粒或病毒），该表达载体可以方便地经历重组 DNA 程序并且可以引起多核苷酸的表达。载体的选择将通常取决于载体与要导入载体的细胞的相容性。载体可以是线性或闭环质粒。载体可以是自主复制载体，即这样的载体，所述载体作为染色体外实体存在，所述载体的复制独立于染色体复制，为例如质粒、染色体外元件、微型染色体或人工染色体。或者，载体可以是这样的载体，所述载体当被引入宿主细胞时整合到基因组中并与其所整合到的染色体一起复制。整合克隆载体可以整合在宿主细胞的染色体中的随机或预定靶基因座处。载体系统可以是单一载体或质粒或两种或更多种载体或质粒，所述载体或质粒一起含有待引入宿主细胞基因组的总 DNA，或转座子。

本文所用的术语“控制序列”是指在特定生物体内或体外参与编码序列表达的调节的组分。控制序列的示例是转录起始序列、终止序列、启动子、前导序列、信号肽、前肽、前原肽或增强子序列；夏因-达尔加诺序列（Shine-Delgarno sequence）、阻遏物或激活物序列；有效的 RNA 处理信号，例如剪接和多腺苷酸化信号；稳定化细胞质 mRNA 的序列；增强翻译效率的序列（例如，核糖体结合位点）；增强蛋白质稳定性的序列；以及当需要时，增强蛋白质分泌的序列。

如本文所定义的“宿主细胞”是适用于遗传操纵并且可以在可用于生产目标产物（如本申请所述的 Prp43 解旋酶）的生物体。宿主细胞可以是在自然界中发现的宿主细胞，或来源于亲本宿主细胞的遗传操纵或经典诱变

后的宿主细胞。有利地，宿主细胞是重组宿主细胞。宿主细胞可以是原核、古细菌或真核宿主细胞。原核宿主细胞可以是但不限于细菌宿主细胞。真核宿主细胞可以是但不限于酵母、真菌、变形虫、藻类、植物、动物、或昆虫宿主细胞。

5 针对核酸或蛋白质（或酶）使用时，术语“重组”是指该核酸或蛋白质（或酶）与其天然形式相比，已经通过人工干预进行了序列修饰。当涉及细胞（例如宿主细胞）时，术语“重组”表示该细胞的基因组如果与其天然形式相比则已经通过人工干预进行了序列修饰。在本文中，术语“重组”与“经修饰”被认为同义。

10 针对经修饰的多肽或酶使用时，术语“置换”表示相应的野生型多肽或酶中存在的天然氨基酸残基被另一种氨基酸残基代替。在本文中，术语“氨基酸置换”与“氨基酸取代”被认为同义。

本文所用的术语“变体”或“突变体”具有相同含义，可互换地使用。它们可以指多肽或核酸。变体是指相对于参考序列（通常为核酸或多肽的野生形式）存在一个或多个位置处的置换、插入、缺失、截短、颠换等。可以
15 通过例如位点饱和诱变、扫描诱变、插入诱变、随机诱变、定点诱变和定向进化以及本领域技术人员已知的各种其他重组方法来产生变体。核酸的变异基因可以通过本领域已知的技术人工合成。

“成熟多肽”在本文中定义为这样的多肽，所述多肽处于其最终形式下
20 并且在将 mRNA 翻译成多肽并对所述多肽进行翻译后修饰后获得。翻译后修饰包括 N-末端加工、C-末端截短、糖基化、磷酸化，以及通过切割去除前导序列（诸如信号肽和/或前肽）。

两个多肽序列或核酸序列之间的相似性可以用其同源性来表示。在本文中，两个序列之间的“同一性”或“同源性”被认为具有相同含义，在本文中可互换使用。为了确定两个氨基酸序列或两个核酸序列的序列同源性
25 或序列同一性的百分比，对序列进行比对以实现最佳匹配，序列同一性是两个序列之间在比对区域上相同匹配的百分比。两个氨基酸序列之间或两个多核苷酸序列之间的序列同源性百分比可以使用公知的算法来确定，例如用于比对两个序列的 Needleman 和 Wunsch 算法 (Needleman, S. B. 和 Wunsch,

C. D. (1970) J. Mol. Biol. 48, 443-453)。例如可以使用来自 EMBOSS 程序包的 NEEDLE 程序来进行。本领域技术人员将理解，当使用不同的算法或特定算法的不同参数时，可能会产生略微不同的结果，但是两个序列之间的同一性百分比不会显著改变。

5 本申请所述的“开口”是指野生型 Prp43 解旋酶本身带有的多核苷酸结合结构域的开口，也可以指与 Prp43 解旋酶结合的多核苷酸结合部分的开口，所述的开口为使得多核苷酸与 Prp43 解旋酶解离的开口，并且该开口可以不是一直存在的，但是至少在一种构象状态下包含至少一个开口。本申请所述的“经修饰的 Prp43 解旋酶”或者包含经修饰的 Prp43 解旋酶的构建体
10 含有一个或多个开口。经过修饰 Prp43 解旋酶，使得解旋酶的同一体上有两个或多个部分连接以减小开口的大小。

 本申请所述的“一个以上”、“至少一个”、“一个或多个”或“一个或更多个”包括：一个、两个、三个、四个、五个、六个、七个、八个、九个、十个、十一个、十二个、或更多个等等。

15 本申请所述的“两个以上”或“两个或多个”或“两个或以上”包括：两个、三个、四个、五个、六个、七个、八个、九个、十个、十一个、十二个、或更多个等等。

 本申请所述的“多个”包括但不限于：三个、四个、五个、六个、七个、八个、九个、十个、十一个、十二个、或更多个等等。

20 本申请所述的“和/或”包括择一列出的项目以及任何数量的项目组合。

 本申请所述的“包括”、“含有”或“包含”是开放式的描述，表示含有所描述的指定成分或步骤，以及不会产生实质影响的其他指定成分或步骤。特别地，当上述术语用于描述蛋白质或核酸的序列时，表示所述蛋白质或核酸既可以是所述序列组成，或者也可以是在所述蛋白质或核酸的一端或
25 两端可以具有额外的氨基酸残基或核苷酸，但所述蛋白质或核酸仍然具有本申请所述的活性（例如其控制多核苷酸移动的能力等等）。

Prp43 解旋酶

Prp43 解旋酶是一种已知的解旋酶，其结构和功能在现有技术中已有研究和报道，例如可参见 Marcel J. Tauchert et.al, “Structural and functional analysis of the RNA helicase Prp43 from the thermophilic eukaryote *Chaetomium thermophilum*”, *Acta Cryst.*, 2016, F72, 112–120。但是，未见有使用 Prp43 解旋酶用于纳米孔测序或控制多核苷酸分子移动通过纳米孔的报道。

Prp43 解旋酶属于 SF2 超家族的 DEAH/RHA 解旋酶，能结合单链 DNA 或 RNA，以 3’-5’方向进行移位或解旋双链 DNA 或 RNA 核酸，同属于该家族的解旋酶还包括 Prp22 解旋酶、Prp2 解旋酶、MLE 解旋酶、DHX9 解旋酶等。

Prp43 解旋酶的 3D 结构和其组成结构域已经在现有技术中被阐明。例如，图 1 是嗜热毛壳菌 (*Chaetomium thermophilum*) 来源的 Prp43 解旋酶 (SEQ ID NO: 1) 的 3D 结构示意图。Prp43 解旋酶除了解旋酶常有的 RecA1 (P97-R273) 和 RecA2 (T274-T458) 两个核心结构域之外，还包含若干结构域：N 端结构域 (M1-L96)，C 端的 WH 结构域 (Y459-P526)、Ratchet 结构域 (L527-V640) 和 OB 结构域 (S641-A764) 等。RecA1 和 RecA2 包含 7 个保守的基序 (Motifs)，其中 Ia (TQPRRVAA)、Ib (TDGQLLR) 和 IV (LLFLTG) 与底物核酸相互作用，基序 I (GSGKT)、II (DEAH)、V (TNIAETSLT) 和 VI (QRAGRAGR) 则与核苷酸的结合有关，而基序 III (SAT) 则耦联核苷酸的水解和核酸的移位或解旋。

Prp43 解旋酶在 RecA1 和 RecA2 结构域的顶端富含带正电荷氨基酸，与 C 端的 WH 结构域、Ratchet 结构域和 OB 结构域一起形成环绕单链 DNA 或 RNA 的通道。

发明人发现虽然 Prp43 解旋酶与单链 DNA 或 RNA 具有较强的亲和力，但该过程仍然是一个热力学动态平衡过程，并不能完全保持控制目标核酸的过孔移动，尤其是目标核酸的长度较长时，比如 1000 个碱基长度、5000 个碱基长度、10000 个碱基长度、100000 个碱基长度或更长碱基长度的核酸。发明人发现可以通过对 Prp43 解旋酶进行修饰，保证酶与核酸的结合并持续控制核酸穿过纳米孔。具体地，发明人发现，在 Prp43 解旋酶的 RecA1 结构域、RecA2 结构域和/或 Ratchet 结构域中引入一个或多个半胱氨酸或

非天然氨基酸，可以减小 Prp43 解旋酶的多核苷酸结合结构域的开口大小，进而改善其目标核酸的结合能力。

因此，本申请的第一方面涉及一种经修饰的 Prp43 解旋酶，包括 RecA1 结构域、RecA2 结构域和 Ratchet 结构域，相对于对应的野生型 Prp43 解旋酶或其片段而言所述经修饰的 Prp43 解旋酶包括在选自 RecA1 结构域、RecA2 结构域、Ratchet 结构域的至少一个结构域中引入的 1、2、3、4、5、6、7、8、9、10、11 个或更多个半胱氨酸的插入或置换，和/或 1、2、3、4、5、6、7、8、9、10、11 个或更多个非天然氨基酸的插入或置换。

优选的，可以在下列任意一组中引入一个至少一个半胱氨酸残基和/或至少一个非天然氨基酸：

- (a) RecA1 结构域；
- (b) RecA2 结构域；
- (c) Ratchet 结构域；
- (d) RecA1 结构域和 Ratchet 结构域；
- (e) RecA2 结构域和 Ratchet 结构域。

本申请所述的非天然氨基酸包括但不限于：4-叠氮基-L-苯丙氨酸(Faz)，4-乙酰基-L-苯丙氨酸，3-乙酰基-L-苯丙氨酸，4-乙酰乙酰基-L-苯丙氨酸，O-烯丙基-L-酪氨酸，3-(苯基硒烷基)-L-丙氨酸，O-2-丙炔-1-基-L-酪氨酸，4(二羟基硼基)-L-苯丙氨酸，4-[(乙基硫烷基)羰基]-L-苯丙氨酸，(2S)-2-氨基-3-{4-[(丙烷-2-基硫烷基)羰基]苯基}丙酸，(2S)-2-氨基-3-{4-[(2-氨基-3-硫烷基丙酰基)氨基]苯基}丙酸，O-甲基-L-酪氨酸，4-氨基-L-苯丙氨酸，4-氰基-L-苯丙氨酸，3-氰基-L-苯丙氨酸，4-氟-L-苯丙氨酸，4-碘-L-苯丙氨酸，4-溴-L-苯丙氨酸，O-(三氟甲基)酪氨酸，4-硝基 L-苯丙氨酸，3-羟基-L-酪氨酸，3-氨基-L-酪氨酸，3-碘-L-酪氨酸，4-异丙基-L-苯丙氨酸，3-(2-萘基)-L-丙氨酸，4-苯基-L-苯丙氨酸，(2S)-2-氨基-3-(萘-2-基氨基)丙酸，6-(甲基硫烷基)正亮氨酸，6-氧-L-赖氨酸，D-酪氨酸，(2R)-2-羟基-3-(4-羟基苯基)丙酸，(2R)-2-氨基辛酸酯 3-(2, 2'-二吡啶-5-基)-D-丙氨酸，2-氨基-3-(8-羟基-3-喹啉基)丙酸，4-苯甲酰-L-苯丙氨酸，S-(2-硝基苄基)半胱氨酸，(2R)-2-氨基-3-[(2-硝基苄基)硫烷基]丙酸，(2S)-2-氨基-3-[(2-硝基苄基)氧基]丙酸，O-(4，

5-二甲氧基-2-硝基苄基)-L-丝氨酸, (2S)-2-氨基-6-({[(2-硝基苄基)氧基]羰基}氨基)己酸, O-(2-硝基苄基)-L-酪氨酸, 2-硝基苯丙氨酸, 4-[(E)-苯基二氮烯基]-L-苯丙氨酸, 4-[3-(三氟甲基)-3H-吡啶基-3基]-D-苯丙氨酸, 2-氨基-3-[[5-(二甲基氨基)-1-萘基]磺酰基氨基]丙酸, (2S)-2-氨基-4-(7-羟基-2-氧-2H-色烯-4-基)丁酸, (2S)-3-[(6-乙酰基萘-2-基)氨基]-2-氨基丙酸, 4(羧基甲基)苯丙氨酸, 3-硝基-L-酪氨酸, O-硫基-L-酪氨酸, (2R)-6-乙酰氨基-2-氨基己酸酯, 1-甲基组氨酸, 2-氨基壬酸, 2-氨基癸酸, L-同质半胱氨酸, 5-硫烷基正缬氨酸, 6-硫烷基-L-正亮氨酸, 5-(甲基硫烷基)-L-正缬氨酸, N6-[(2R, 3R)-3-甲基-3, 4-二氢-2H-吡咯-2-基]羰基}-L-赖氨酸, N6-[(苄基氧基)羰基]赖氨酸, (2S)-2-氨基-6-[(环戊基羰基)氨基]己酸, N6-[(环戊基氧基)羰基]-L-赖氨酸, (2S)-2-氨基-6-[(2R)-四氢呋喃-2-基羰基]氨基}己酸, (2S)-2-氨基-8-[(2R, 3S)-3-乙炔基四氢呋喃-2-基]-8-氧基辛酸, N6-(叔丁氧基羰基)-L-赖氨酸, (2S)-2-羟基-6-({[(2-甲基-2-丙烷基)氧基]羰基}氨基)己酸, N6-[(烯丙氧基)羰基]赖氨酸, (2S)-2-氨基-6-({[(2-叠氮苄基)氧基]羰基}氨基)己酸, N6L-脯氨酰基-L-赖氨酸, (2S)-2-氨基-6-[(丙-2-炔-1-基氧基)羰基]氨基}己酸或 N6-[(2-叠氮乙氧基)羰基]-L-赖氨酸。

本申请中的“Prp43 解旋酶”应该按照其广义来理解, 并认为涵盖了 Prp43 解旋酶 (例如 SEQ ID NO: 1) 的同源蛋白。通常, 只要一种酶具有 DNA/RNA 解旋活性, 含有 RecA1 结构域、RecA2 结构域和/或 Ratchet 结构域, 且与 SEQ ID NO: 1 具有至少 30% 的同源性, 例如至少 35%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或至少 99.9% 的同源性, 则可以认为其属于 Prp43 解旋酶。因此, 一些被称为 HrpA 解旋酶或 HrpB 解旋酶的解旋酶 (例如表 1 中所列出的那些) 也被认为落入本申请中的“Prp43 解旋酶”范围内。

本申请所述的 Prp43 解旋酶可以是各种常规来源的 Prp43 解旋酶, 例如所述 Prp43 解旋酶可以来源于 *Chaetomium thermophilum*、*Bathycoccus prasinus*、*Uncultured bacterium*、*Archaeon*、*Parcubacteria*、*Sorangium cellulosum*、*Candidatus Sungbacteria*、*Mycolicibacterium chitae*、*Parcubacteria*、

Thermodesulforhabdus norvegica、*Deltaproteobacteria*、*Puniceicoccales*、*Desulfobacterium vacuolatum* 或 *Desulfobacter sp.* 或来源于病毒基因组 (viral metagenome) 等。表 1 给出了一些可以用于本申请的同源 Prp43 解旋酶的示例, 但是本申请的 Prp43 解旋酶并不限于这些例子。

5 表 1: ctPrp43 同源蛋白示例

SEQ ID	名称	功能	来源	长度 (Lengths)	同一性 (Identities)	NCBI 编号
SEQ ID NO: 1	ctPrp43	ATP-dependent RNA helicase	Chaetomium thermophilum	764	100%	XP_00669 1112.1
SEQ ID NO: 2	bpPrp43	ATP-dependent RNA helicase	Bathycoccus prasinus	711	64.69%	XP_00751 1797.1
SEQ ID NO: 3	unPrp43	ATP-dependent RNA helicase	uncultured bacterium	533	59%	L7RXK7
SEQ ID NO: 4	vmPrp43	ATP-dependent RNA helicase	viral metagenome	635	53.18%	QHU0768 1.1
SEQ ID NO: 5	arPrp43	ATP-dependent RNA helicase	archaeon	628	41.90%	A0A482T 3B5
SEQ ID NO: 6	paPrp43	ATP-dependent RNA helicase	Parcubacteria group bacterium GW2011	740	41.81%	A0A0G1F X53
SEQ ID NO: 7	scHrpA	ATP-dependent RNA helicase	Sorangium cellulosum	1230	40.25%	A0A150R F04
SEQ ID NO: 8	csHrpA	ATP-dependent RNA helicase	Candidatus Sungbacteria bacterium	767	39.76%	A0A1G2K 8K0
SEQ ID NO: 9	mcHrpA	ATP-dependent RNA helicase	Mycolicibacterium chitae	1311	39.61%	A0A3S4T QL2
SEQ ID NO: 10	paHrpB	ATP-dependent RNA helicase	Parcubacteria group bacterium ADurb.Bin305	1305	39.42%	A0A1V5R PG5
SEQ ID NO: 11	tnHrpA	ATP-dependent RNA helicase	Thermodesulforhabdu s norvegica	703	38.14%	A0A7C0 WSW3
SEQ ID NO: 12	dbHrpA	ATP-dependent RNA helicase	Deltaproteobacteria bacterium	1277	38.06%	A0A661LI X4
SEQ ID NO: 13	pbHrpA	ATP-dependent RNA helicase	Puniceicoccales bacterium CK1056	1255	37.95%	A0A6B2L XC5
SEQ ID NO: 14	dvHrpA	ATP-dependent RNA helicase	Desulfobacterium vacuolatum DSM 3385	1331	36.99%	A0A1W2 CQF0
SEQ ID NO: 15	dsHrpA	ATP-dependent RNA helicase	Desulfobacter sp.	1311	36.67%	A0A357Y TE1

最优的，申请所述的 Prp43 解旋酶是来源于嗜热毛壳菌（*Chaetomium thermophilum*）的。

因此，在一些优选实施方式中，本申请提供了一种经修饰的 Prp43 解旋酶，所述 Prp43 解旋酶包含 SEQ ID NO: 1 或其片段的变体，所述变体包括
5 在 RecA1 结构域、RecA2 结构域和/或 Ratchet 结构域中引入的 1、2、3、4、5、6、7、8、9、10、11 个或更多个半胱氨酸的插入或置换，和/或 1、2、3、4、5、6、7、8、9、10、11 个或更多个非天然氨基酸的插入或置换。

在一些优选实施方式中，所述变体包括在对应于 SEQ ID NO: 1 的 M157、Q161、D165、F181、E182、N183、R324、L328、E332、R335、L351、P352、
10 P353、H354、D321、E320、R358、P563、A564、N565、D603、K605、K606、H609、Y615、R616、S619、N623、A626 或 K630 中的任一个或两个以上的位置上引入至少一个半胱氨酸残基和/或至少一个非天然氨基酸残基。更优选地，所引入的半胱氨酸残基或非天然氨基酸残基位于对应于 SEQ ID NO: 1 的 F181、P352、S619 或 N623 中的任一个或两个以上的位置。

已经发现，将野生型 Prp43 解旋酶的 N 端结构域去除，更有利于控制核苷酸的移动。因此，在一些优选实施方式中，所述经修饰的 Prp43 解旋酶包括 N 端结构域的去除，优选地去除 N 端的位置 1 开始的至少 96 个、至少 90 个、至少 80 个、至少 70 个、至少 60 个、至少 50 个、至少 40 个或至少 30 个残基。对于 SEQ ID NO: 1 而言，优选地去除 M1-N60，即优选使
20 用 SEQ ID NO: 1 的 T61-A764 片段，并在此基础上引入一个或更多个半胱氨酸的插入或置换，和/或一个或更多个非天然氨基酸的插入或置换。

为提高本申请所述的 Prp43 解旋酶与目标多核苷酸结合的稳定性，降低从目标多核苷酸上解脱的能力，还可以引入 2 个或 2 个以上的半胱氨酸残基或非天然氨基酸残基，且在引入的半胱氨酸与半胱氨酸之间相互连接、
25 在引入的非天然氨基酸与非天然氨基酸之间相互连接、在引入的半胱氨酸与非天然氨基酸之间相互连接、在引入的半胱氨酸与天然氨基酸之间相互连接，或者在引入的非天然氨基酸与天然氨基酸之间相互连接。

优选的，可以使任何数目和组合的两个以上引入的半胱氨酸与非天然氨基酸相互连接。例如，可以使 2、3、4、5、6、7、8 或更多个半胱氨酸和

/或非天然氨基酸相互连接。一个或多个半胱氨酸可以与一个或多个半胱氨酸连接。一个或多个半胱氨酸可以与一个或多个非天然氨基酸诸如 Faz 连接。一个或多个非天然氨基酸诸如 Faz 可以与一个或多个非天然氨基酸诸如 Faz 连接。一个或多个半胱氨酸可以与一个或多个解旋酶上的天然氨基酸连接。一个或多个非天然氨基酸诸如 Faz 可以与一个或多个解旋酶上的天然氨基酸连接。

5 优选的,所述的连接可以是任何连接方式,包括暂时连接或者永久的连接方式,例如共价连接或氢键连接或静电相互作用或 $\pi-\pi$ 相互作用或疏水相互作用等。在本发明的另一个具体实施方式中,所述的连接可以是永久的,例如共价连接。可以采用化学交联剂进行共价连接,其长度可以从一个碳(碳酰氯型连接器)到多个埃变化。例如马来酰亚胺、活性酯、琥珀酰亚胺、叠氮化物、烷烃、烯烃、炔烃(诸如二苯并环辛炔醇(DIBO 或 DBCO), 二氟环炔烃和线性炔烃)等。又例如聚乙二醇(PEGs)、多肽、多糖、脱氧核糖核酸(DNA)、肽核酸(PNA)、苏糖核酸(TNA)、甘油核酸(GNA)、饱和的和饱和的炔或聚酰胺等等的线性分子,又例如 TMAD 等等的催化试剂,可以通过-S-S 键进行连接。

在本发明的某些具体实施方式中,采用 TMAD 催化剂使得 F181 和 N623 位置或 P352 和 S619 位置引入的半胱氨酸残基与半胱氨酸残基之间共价连接。

20 在一些优选实施方式中,所述经修饰的 Prp43 解旋酶还包括对于一个或多个半胱氨酸残基的置换,更优选为对应于 SEQ ID NO: 1 的 C148、C214、C303、C323、C377、C441、C508、C543、C608 的一个或多个半胱氨酸残基被置换,更优选地半胱氨酸残基被置换为丙氨酸、甘氨酸、缬氨酸、异亮氨酸、亮氨酸、苯丙氨酸、酪氨酸、丝氨酸、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、组氨酸、蛋氨酸、色氨酸、谷氨酰胺、天冬酰胺或脯氨酸残基。

25 在一些优选实施方式中,为了进一步使所述 Prp43 解旋酶具有持续、稳定地以一定速率控制多核苷酸移动的能力,所述经修饰的 Prp43 解旋酶还进一步包含选自以下组的一个或多个氨基酸修饰:

(a)一个或多个与核苷酸相互作用的氨基酸被置换；

(b)一个或多个与 NTP 和/或二价金属离子（如 Mg^{2+} ）结合相关的氨基酸被置换；

(c)一个或多个与跨膜孔相互作用的氨基酸被置换；

5 (d)降低 Prp43 解旋酶的表面的负电荷的进一步修饰。

优选的，所述与核苷酸相互作用且被置换的氨基酸包括但不限于：对应于 SEQ ID NO: 1 的 R152、R153、R180、T195、Q198、R201、E316、E317、G349、T381、N382、K403、K405、L416、P526、P557、R562、Q558、H688、P689、T708、K710、Y712、R714。进一步优选的，至少一个与单链 DNA、
10 RNA 或双链 DNA、RNA 中一个或多个核苷酸的磷酸基团相互作用的氨基酸被置换。

优选的，所述一个或多个与 NTP 和/或二价金属离子（如 Mg^{2+} ）结合相关的氨基酸包括但不限于：对应于 SEQ ID NO: 1 的 T126、D218、S387、E219、R432、R435、T121、K125、T127、T389、R162、D391、F360。

15 优选的，所述一个或多个与跨膜孔相互作用的氨基酸包括但不限于：对应于 SEQ ID NO: 1 的 C303、E336、D288、R287、E286、E284、E291。

进一步优选的，用包含较大侧链的氨基酸取代至少一个与单链 DNA、RNA 或双链 DNA、RNA 中一个或多个核苷酸的糖和/或碱基相互作用的氨基酸。所述较大侧链包括增加数目的碳原子，具有增加的长度，增加的分子
20 体积和/或具有增加的范德华体积。所述较大侧链增加了所述至少一个氨基酸与所述单链或双链 DNA 中一个或多个核苷酸之间的(i)静电相互作用；(ii)氢键和/或(iii)阳离子- π 相互作用。所述较大侧链的氨基酸不是丙氨酸(A)、半胱氨酸(C)、甘氨酸(G)、硒代半胱氨酸(U)、甲硫氨酸(M)、天冬氨酸(D)或谷氨酸(E)。

25 优选的，所述的 Prp43 解旋酶进一步被修饰降低其表面的负电荷。所述的 Prp43 解旋酶还包含增加净正电荷的取代。优选的，所述的 Prp43 解旋酶还包含对表面带负电的氨基酸、极性或非极性氨基酸进行取代或修饰。进一步优选的，所述的取代包括带正电的氨基酸、不带电荷的氨基酸取代带负电的氨基酸、不带电荷的氨基酸、芳香族氨基酸、极性或非极性氨基酸。其中，

所述的带正电的氨基酸、不带电荷的氨基酸、极性、非极性氨基酸或芳香族氨基酸可以是天然的或非天然的氨基酸，其可以是人工合成的或者经过修饰的天然氨基酸。

本申请所述所述的 Prp43 解旋酶经修饰后，可以与对应的野生型的
5 Prp43 解旋酶的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少 60%、
至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少
93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%
或至少 99.9%的同源性。

在一些更优选的实施方式中，所述 Prp43 解旋酶为 SEQ ID NO: 1 的变
10 体（即来源于 *Chaetomium thermophilum*），且所述 SEQ ID NO: 1 的变体包
括在 SEQ ID NO: 1 的 F181 和/或 N623 位置上引入至少一个半胱氨酸残基
和/或至少一个非天然氨基酸；或者所述 SEQ ID NO: 1 的变体包括在 SEQ
ID NO: 1 的 P352 和/或 S619 位置上引入至少一个半胱氨酸残基和/或至少
一个非天然氨基酸。

15 在一些更优选的实施方式中，所述 Prp43 解旋酶为 SEQ ID NO: 1 的变
体（即来源于 *Chaetomium thermophilum*），且所述 SEQ ID NO: 1 的变体还
包括 SEQ ID NO: 1 的至少一个或多个半胱氨酸被取代。取代的氨基酸可以
为丙氨酸、甘氨酸、缬氨酸、异亮氨酸、亮氨酸、苯丙氨酸、酪氨酸、丝氨
酸、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、组氨酸、蛋氨酸、色氨
20 酸、谷氨酰胺、天冬酰胺、脯氨酸。优选的，所述的一个或多个被取代的半
胱氨酸为 C148、C214、C303、C323、C377、C441、C508、C543、C608。

在一些更优选的实施方式中，所述 Prp43 解旋酶为 SEQ ID NO: 1 的变
体（即来源于 *Chaetomium thermophilum*），且所述 SEQ ID NO: 1 的变体去
除 N 端结构域的 M1 至 N60 序列，进一步优选的去掉 N 端的 M1 至 L96 序
25 列。本发明的一个具体案例中是去除了 N 端结构域 M1 至 N60 序列的解旋
酶。

在一些更优选的实施方式中，所述 Prp43 解旋酶为 SEQ ID NO: 1 的变
体（即来源于 *Chaetomium thermophilum*），且所述 SEQ ID NO: 1 的变体与
SEQ ID NO: 1 的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少

60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或至少 99.9%的同源性。更优选地，所述 Prp43 解旋酶为经修饰的 SEQ ID NO: 1 的 T61-A764 片段。

5 在一些更优选的实施方式中，所述 Prp43 解旋酶为经修饰的 SEQ ID NO: 1 的 T61-A764 片段（来源于 *Chaetomium thermophilum*），且所述修饰为 F181C/N623C/C508S 或 P352C/S619C/C508S。

另外，本申请所述的 Prp43 解旋酶可以被修饰以助于鉴定或纯化，例如通过添加组氨酸残基(His 标签)，天冬氨酸残基(asp 标签)，链霉亲和素标签，10 Flag 标签，SUMO 标签，GST 标签或 MBP 标签，或通过添加信号序列以促进它们从细胞中分泌，该细胞中的多肽不天然地含有该信号序列。引入遗传标签的替换方式是通过化学反应将标签连到 Prp43 解旋酶上的天然或人工位点。

本申请所述的 Prp43 解旋酶可以是 Prp43 解旋酶寡聚体形式，所述的15 Prp43 解旋酶寡聚体包含一个或多个的本申请所述的 Prp43 解旋酶。

在一些实施方式中，所述的 Prp43 解旋酶寡聚体还可以包含野生型 Prp43 解旋酶或其他类型的解旋酶。其中，所述的其他类型的解旋酶可以为 Hel308 解旋酶、XPD 解旋酶、Dda 解旋酶、RecD2 解旋酶、TraI 解旋酶或 TrwC 解旋酶等等。

20 优选的，所述的 Prp43 解旋酶与野生型 Prp43 解旋酶之间、Prp43 解旋酶与 Prp43 解旋酶之间、野生型 Prp43 解旋酶与野生型 Prp43 解旋酶、Prp43 解旋酶与其他类型解旋酶之间或者野生型 Prp43 解旋酶与其他类型解旋酶之间，可以通过头对头、尾对尾或者头对尾的方式连接或排列。

25 优选的，所述的 Prp43 解旋酶寡聚体包含两个以上的本申请所述的 Prp43 解旋酶，其中，所述的 Prp43 解旋酶可以是不同的或者相同的。

蛋白构建体

在生理功能中，Prp43 解旋酶参与了前体 mRNA 加工过程中由 U2.U5.U6snRNPs 组成的内含子剪接体的解离，该过程中酶功能的发挥需要

与两个含有富含甘氨酸基序 (G-Path motif) 的辅助蛋白 Ntr1 和 Ntr2 相互作用, 以激活其 ATP 水解活性和解旋活性; Prp43 解旋酶也参与了核糖体合成过程, 以帮助 18S 和 25S 的 rRNAs 前体的成熟, 该过程也需要富含 G-Path 基序的蛋白 Paf1 和 Gno1 蛋白的激活。

5 Prp43 解旋酶在生理功能条件下需要含有 G-Path 结构域的辅助蛋白激活其 ATP 水解活性和解旋活性。虽然在缺少辅助蛋白的条件下该酶具有微弱的活性, 更优地是在辅助激活蛋白的存在下其 ATP 水解和解旋活性更强。特别地, 发明人发现含有 G-Path 结构域辅助蛋白的单独部分片段仍具有激活功能。

10 因此, 在本申请的第二方面, 提供了一种蛋白构建体, 其包括本申请第一方面所述的经修饰的 Prp43 解旋酶, 以及在所述 Prp43 解旋酶的 C 端或 N 端融合的辅助激活蛋白 Paf1 的 G-Path 结构域或含有 G-Path 结构域的 Paf1 的片段。该蛋白构建体也可以视为一种融合蛋白。

这种经修饰的 Prp43 解旋酶构建体, 由于在 Prp43 解旋酶的 C 端或 N
15 端融合了辅助激活蛋白 Paf1 或其同源蛋白的 G-Path 结构域或含有 G-Path 结构域的片段, 使修饰后的解旋酶 ATP 水解和或解旋活性明显增强, 更有利于控制纳米孔核酸测序中核酸的过孔移动。

所述蛋白构建体中, 经修饰的 Prp43 解旋酶的数目可以为一个或多个。

20 所述蛋白构建体中, 所述辅助激活蛋白 Paf1 可以是本领域常规使用的各种来源的 Paf1 蛋白, 例如可以是来源于 *Chaetomium thermophilum var. thermophilum*、*Thermothielavioides terrestris*、*Thermothelomyces thermophilus*、*Podospora anserina*、*Neurospora tetrasperma*、*Coniochaeta sp.*、*Monosporascus sp.*、*Hypoxylon sp.*、*Madurella mycetomatis* 或 *Coniochaeta pulveracea* 的 Paf1。

25 表 2 给出了一些可以用于本申请的 Prp43 解旋酶构建体的同源 Paf1 蛋白的示例, 但是本申请的 Paf1 蛋白并不限于这些例子。

表 2: ctPfa1 同源蛋白

SEQ ID	名称	来源	长度 (Lengths)	同一性 (Identities)	NCBI 编号	注释
SEQ ID NO: 16	ctPfa1	Chaetomium thermophilum var. thermophilum DSM 1495	742	100%	XP_006695185	hypothetical protein CTHT_0048220
SEQ ID NO: 17	ttnPfa1	Thermothielavioides terrestris NRRL 8126	709	50.07%%	XP_003654107.1	uncharacterized protein
SEQ ID NO: 18	ttPfa1	Thermothelomyces thermophilus ATCC 42464	700	47.44%	XP_003665422.1	uncharacterized protein
SEQ ID NO: 19	paPfa1	Podospora anserina	718	45.66%	XP_001929716.1	uncharacterized protein
SEQ ID NO: 20	ntPfa1	Neurospora tetrasperma FGSC 2508	657	43.30%	XP_009853333.1	hypothetical protein
SEQ ID NO: 21	csPfa1	Coniochaeta sp. 2T2.1	679	36.97%	KAB5518794.1	hypothetical protein
SEQ ID NO: 22	maPfa1	Monosporascus sp. GIB2	688	35.80%	RYP03421.1	hypothetical protein
SEQ ID NO: 23	hsPfa1	Hypoxyylon sp. EC38	667	37.52%	OTA65903.1	hypothetical protein
SEQ ID NO: 24	mmPfa1	Madurella mycetomatis	682	45.19%	KXX74671.1	Protein SQS1
SEQ ID NO: 25	cpPfa1	Coniochaeta pulveracea	718	36.95%	RKU41729.1	squalene synthetase-like protein

优选的，所述 G-path 结构域序列为上述 Pfa1 辅助蛋白或其同源蛋白中对应 SEQ ID NO: 16 的 K662-G742 片段（即 SEQ ID NO: 26 的序列）的序列或其变体的序列。

5 在一些优选的实施方式中，在所述蛋白构建体中，所述辅助激活蛋白 Paf1 的氨基酸序列为 SEQ ID NO: 16 或与 SEQ ID NO: 16 的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或至少 99.9%的同源性的变体的氨基酸序列，并且所述辅助激活蛋白 Paf1 具有激活 Prp43 解旋酶的功能。

在一些更优选的实施方式中,所述 Paf1 的 G-Path 结构域为 SEQ ID NO: 16 的 K662-G742 片段。

在最优选的实施方式中,在所述蛋白构建体中,所述 Prp43 解旋酶包括 SEQ ID NO: 1 的序列或其变体,所述 Pfa1 辅助激活蛋白包括 SEQ ID NO: 16 序列或其变体、或 SEQ ID NO: 16 的 G-Path 结构域序列 SEQ ID NO: 26 (对应 SEQ ID NO: 16 序列的 K662-G742 片段) 或其变体。

本申请所述的蛋白构建体可以被修饰以助于鉴定或纯化,例如通过添加组氨酸残基(His 标签),天冬氨酸残基(asp 标签),链霉亲和素标签,Flag 标签,SUMO 标签,GST 标签或 MBP 标签或 Strep TagII 标签,或通过添加信号序列以促进它们从细胞中分泌,该细胞中的多肽不天然地含有该信号序列。引入遗传标签的替换方式是通过化学反应将标签连到蛋白构建体上的天然或人工位点。

核酸

本申请的第三方面,提供了一种核酸,所述的核酸编码本申请第一方面所述的 Prp43 解旋酶和/或本申请第二方面所述的蛋白的构建体。

表达载体

本申请的第四个方面,提供了一种表达载体,所述的表达载体包含本申请第三方面所述的核酸。优选的,所述的核酸可操作的连接至表达载体中的调控元件,其中所述的调控元件优选为启动子。在本申请的一些具体实施方式中,所述的启动子选自 T7、trc、lac、ara 或 λ L。优选的,所述的表达载体包括但不限于质粒、病毒或噬菌体。

本领域技术人员已知有多种用于将核酸插入核酸构建体或表达载体中的方法,参见例如 Sambrook 和 Russell, Molecular Cloning: A Laboratory Manual, 第 3 版, CSHL Press, Cold Spring Harbor, NY, 2001。

宿主细胞

本申请的第五方面，提供了一种宿主细胞，所述的宿主细胞包含本申请第三方面所述的核酸或本申请第四方面所述的表达载体。优选的，所述的宿主细胞包括但不限于大肠杆菌。在本申请的一个具体实施方式中，所述的宿主细胞选自 BL21 (DE3)、JM109 (DE3)、B834 (DE3)、TUNER、C41
5 (DE3)、Rosetta2 (DE3)、Origami、Origami B 等等。

Prp43 解旋酶或蛋白构建体的制备方法

本申请的第六方面涉及一种制备本申请的第二方面所述的蛋白构建体的方法，包括：提供 SEQ ID NO: 1 的多肽或其变体和 SEQ ID NO: 26 的多肽或其变体，在 SEQ ID NO: 1 的多肽或其变体中引入至少一个半胱氨酸残基和/或至少一个非天然氨基酸，然后在所得多肽的 C 端或 N 端融合 SEQ
10 ID NO: 26 的多肽或其变体，形成所述蛋白构建体。

本申请的第七方面涉及一种制备本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体的方法，包括：包括培养
15 本申请第五方面所述的宿主细胞，并进行诱导表达，然后纯化所得表达产物。

遗传改造技术，诸如酶在宿主细胞中的过表达、对宿主细胞的遗传修饰、或杂交技术是本领域中已知的方法，诸如在 Sambrook 和 Russel (2001) "Molecular Cloning: A Laboratory Manual (第 3 版)", Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press 或 F. Ausubel 等人编著，
20 "Current protocols in molecular biology", Green Publishing and Wiley Interscience, New York (1987)中描述的那些技术。

例如，在本申请的一个具体实施方式中，经修饰的 Prp43 解旋酶的制备方法包括：根据本申请所述 Prp43 解旋酶和或辅助激活蛋白或激活结构域的氨基酸序列，获得编码 Prp43 解旋酶的核酸序列，酶切连接至表达载体后
25 转化至大肠杆菌中，诱导表达和纯化，获得所述 Prp43 解旋酶。

Prp43 解旋酶或蛋白构建体的应用

本申请的 Prp43 解旋酶或蛋白构建体可以用于控制多核苷酸分子移动或是用于表征目标多核苷酸。

本发明所述的 Prp43 解旋酶是一种在链测序过程中控制目标多核苷酸移动的有用工具，当提供了促进移动的常规必要组分时，Prp43 解旋酶沿着 DNA 或 RNA 以 3'-5' 的方向移动，但 DNA 或 RNA 在孔中的定向（取决于 DNA 或 RNA 的哪个末端被捕获）意味着 Prp43 解旋酶可以用于逆着所施加的场的方向或顺着施加的场的方向将 DNA 或 RNA 移进孔。通过在野生型 Prp43 解旋酶中引入半胱氨酸残基和/或至少一个非天然氨基酸，可以有效减少 Prp43 解旋酶或构建体上多核苷酸结合结构域或多核苷酸结合部分开口的大小或开闭，以及目标多核苷酸被解开的开口的大小或开闭，从而显著降低 Prp43 解旋酶从目标多核苷酸上解脱的能力，提高控制目标多核苷酸通过孔的能力。通过在野生型 Prp43 解旋酶或者修饰后的 Prp43 解旋酶的 C 端或 N 端融合 G-path 结构域或含有 G-Path 结构域的多肽，可以有效提升修饰后的 Prp43 解旋酶的 ATP 水解活性或者解旋活性，从而提高控制目标多核苷酸通过孔的能力。

本申请的第八方面涉及一种控制多核苷酸分子移动的方法，包括将所述多核苷酸分子与本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体接触。

优选的，所述的控制多核苷酸移动为控制多核苷酸穿过孔的移动。所述的孔为纳米孔，所述的纳米孔为跨膜孔。该孔可以是天然的或人造的，包括但不限于生物孔、固态孔或生物与固态杂交的孔。优选的，所述的方法可以包含一个或多个的 Prp43 解旋酶共同控制多核苷酸的移动。

本申请的第九方面涉及一种表征目标多核苷酸的方法，所述的方法包括：

(a) 将目标多核苷酸与本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体接触，使得所述 Prp43 解旋酶或蛋白构建体控制所述目标多核苷酸移动穿过一纳米孔； (b) 获取目标多核苷酸中的核苷酸与所述纳米孔相互作用时的一个或多个特征，从而表征所述目标多核苷酸。

优选的，重复步骤(a)和(b)一次或多次。

优选的，所述的方法中可以使用任意数量的本申请所述的 Prp43 解旋

酶。优选可以为一个或多个，更优选为 1、2、3、4、5、6、7、8、9 个或更多个。其中，所述的两个以上本申请所述的 Prp43 解旋酶可以相同或不同。也可以包含野生型 Prp43 解旋酶或者其他类型的解旋酶。进一步的，两个以上解旋酶之间可以连接或者只是通过分别结合在多核苷酸上而排列发挥控制多核苷酸移动的功能。

优选的，所述的方法还包括横跨与所述解旋酶或构建体，和目标多核苷酸接触的孔施加电势差的步骤。

优选的，所述的孔是允许水合离子在施加的电势的驱动下从膜的一侧流向膜的另一层的结构。进一步优选的，所述的孔为纳米孔，所述的纳米孔为跨膜孔。所述跨膜孔为目标多核苷酸的移动提供了通道。进一步优选的，所述的孔选自生物孔、固态孔或生物与固态杂交的孔。

在一些具体实施方式中，所述的孔包括但不限于衍生自耻垢分枝杆菌孔蛋白 A、耻垢分枝杆菌孔蛋白 B、耻垢分枝杆菌孔蛋白 C、耻垢分枝杆菌孔蛋白 D、溶血素、胞溶素、白细胞介素、外膜孔蛋白 F、外膜孔蛋白 G、外膜磷脂酶 A、WZA 或奈瑟氏菌自转运脂蛋白等等。

所述的膜可以为任何现有技术中存在的膜，优选为两性分子层，即一种由具有至少一个亲水性部分和至少一个亲脂性或疏水性部分的两性分子诸如磷脂质形成的层，两性分子可以是合成的或天然存在的。进一步优选的，所述的膜为脂质双层膜。所述的目标多核苷酸可以使用任何已知的方法连接到膜上。如果膜是两性分子层，如脂质双分子层，所述多核苷酸优选通过在所述膜中存在的多肽或通过所述膜中存在的疏水锚被连接到该膜上。其中，疏水锚优选为脂质、脂肪酸、甾醇、碳纳米管或氨基酸。

优选的，当在孔施加一种力（如电压），目标多核苷酸通过孔的速率被 Prp43 解旋酶或构建体所控制，从而获得一种可识别的稳定的电流水平，用于确定目标多核苷酸的特征。

优选的，所述的目标多核苷酸为单链、双链或至少一部分是双链的。

进一步优选的，所述的目标多核苷酸可以通过标签、间隔物、甲基化、氧化或损伤的方式进行修饰。

在本申请的一个具体实施方式中，所述的目标多核苷酸为至少一部分

是双链的。其中所述的双链部分构成 Y 衍体结构，所述的 Y 衍体结构包含优先螺入所述孔的前导序列。

进一步优选的，所述的目标多核苷酸的长度可以为 10-100000 个碱基或更多个碱基。

5 在本申请的一个具体实施方式中，所述的目标多核苷酸的长度可以为至少 10 个、至少 50 个、至少 100 个、至少 200 个、至少 300 个、至少 400 个、至少 500 个、至少 1000 个、至少 2000 个、至少 5000 个、至少 10000 个、至少 50000 个或至少 100000 个碱基等等。

优选的，所述的解旋酶结合到单链多核苷酸的内部核苷酸中。

10 优选的，所述的目标多核苷酸为 DNA 或 RNA。

优选的，当所述的目标多核苷酸为 RNA 时，为提高要被测序的 RNA 穿过孔的能力和效率，将 RNA 修饰为包含非 RNA 多核苷酸。

优选的，RNA 修饰的步骤包含将 DNA 前导区与待测 RNA 的 3' 末端连接。还包括将待测 RNA 反转录的步骤

15 优选的，所述的一个或多个特征选自目标多核苷酸的来源、长度、同一性、序列、二级结构或目标多核苷酸是否被修饰。进一步优选的，所述的一个或多个特征通过电测量和/或光学测量进行。

进一步优选的，通过电测量和/或光测量产生电信号和/或光信号，而每种核苷酸对应一种信号水平，继而将电信号和/或光信号转化为核苷酸的特征。
20

在本申请的一个具体实施方式中，所述的电测量包括但不限于电流测量、阻抗测量、隧道测量、风洞测量或场效应晶体管（FET）测量等等。

本申请所述的电信号选自电流、电压、隧穿、电阻、电位、电导率或横向电测量的测量值。

25 在一些具体实施方式中，所述的电信号为穿过所述孔的电流。

优选的，所述的表征还包括应用改进型维特比算法。

本申请的第十方面涉及本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体在表征目标多核苷酸或控制目标多核苷酸穿过孔的移动中的用途。

传感器和分析装置

本申请的第十一方面涉及一种用于表征目标多核苷酸的分析装置，所述的分析装置包含一个或多个纳米孔、一个或多个本申请第一方面所述的
5 经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体、以及一个或多个容器。

优选的，所述的分析装置选自试剂盒、装置或传感器。

进一步优选的，所述的分析装置是试剂盒，所述的试剂盒中还包括包含
10 脂质双层的芯片。所述的孔横跨脂质双层。本申请所述的试剂盒包含一个或多个脂质双层，每个脂质双层包含一个或多个所述的孔。本申请所述的试剂盒还包括实施表征目标多核苷酸的试剂或装置。优选的，所述的试剂包括缓冲剂、PCR 扩增所需的工具。

本申请的第十二方面涉及一种形成表征目标多核苷酸的传感器的方法，
15 包括提供纳米孔，和在所述纳米孔和本申请第一方面所述的经修饰的 Prp43 解旋酶或本申请第二方面所述的蛋白构建体之间形成复合物。

下面结合附图和具体实施例对本申请的实施方式作进一步的解释和说明。这些实施例仅仅是为了解释和说明本申请的各个方面，而不能被理解为本申请的范围的限制。

附图说明

20 图 1 示出了来自于嗜热毛壳菌(*Chaetomium thermophilum*)的 N 端(M1-N60)截短的野生型 Prp43 解旋酶 (SEQ ID NO: 1) 的 3D 结构示意图。

图 2 示出了 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶、经修饰的 Prp43 解旋酶 Prp43-2(F181C/N623C/C508S)、经修饰的 Prp43 解旋酶 Prp43-
25 3 (P352C/S619C/C508S)、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 和 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-3 (P352C/S619C/C508S) 的单链 DNA 依赖 ATP 水解活性检测。

图 3 示出了 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶、经修饰的

Prp43 解旋酶 Prp43-2(F181C/N623C/C508S)、经修饰的 Prp43 解旋酶 Prp43-3 (P352C/S619C/C508S)、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 和 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-3 (P352C/S619C/C508S) 的单链 RNA 依赖 ATP 水解活性检测。

图 4 示出了低盐条件下 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶或 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 与单链 DNA 的亲和力曲线。

图 5 示出了 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP 和 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 的凝胶迁移实验结果。其中, 泳道 1 是 T44-37-FAM 底物, 泳道 2 是野生型 Prp43 解旋酶和 T44-37-FAM 底物结合的复合物, 泳道 3 是野生型 Prp43 解旋酶和 T44-37-FAM 底物结合后进行 TMAD 催化处理的产物, 泳道 4 是 Prp43-GP 解旋酶和 T44-37-FAM 底物结合的复合物, 泳道 5 是 Prp43-GP 解旋酶和 T44-37-FAM 底物结合后进行 TMAD 催化处理的产物, 泳道 6 是 Prp43-GP-2 解旋酶突变体和 T44-37-FAM 底物结合的复合物, 泳道 7 是 Prp43-GP-2 解旋酶突变体和 T44-37-FAM 底物结合后进行 TMAD 催化处理的产物。

图 6 示出了 DNA 构建体 X 的示意图, 其中 A 区对应序列 SEQ ID NO: 32 其 5' 末端连接到 4 个 iSpC3 间隔区 (B 区), 该间隔区连接到 C 区对应序列 SEQ ID NO: 33 的 3' 末端, C 区序列的 5' 末端连接到 D 区对应序列 SEQ ID NO: 34, 该构建体的 E 区对应序列 SEQ ID NO: 35 与 F 区对应序列 SEQ ID NO: 36 (其具有 3' 胆固醇系链) 杂交。

图 7 示出了 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 控制 DNA 构建体 X 通过 MspA 纳米孔移动时的电流轨迹示例 (y 轴坐标为电流(pA, 0 到 100), x 轴坐标为时间(h:m:s))。

图 8 说明: 示出了 RNA 构建体 Y 示意图, 其中 SEQ ID NO: 37 (标记为 D) 其 3' 末端连接到 20 个 iSpC3 间隔区 (标记为 A), 其 5' 末端连接到 4 个 iSpC3 间隔区 (标记为 B), 该间隔区连接到 SEQ ID NO: 38 (标记

为 C) 的 3' 末端, 该构建体的 SEQ ID NO: 39 (标记为 E) 区域与 SEQ ID NO: 40 (标记为 F, 其具有 3' 胆固醇系链) 杂交。

图 9 示出了 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-2 (F181C/N623C/C508S) 控制 RNA 构建体 Y 穿过 MspA 纳米孔的电流轨迹示例 (y 轴坐标为电流(pA, 0 到 100), x 轴坐标为时间(h:m:s))。

实施例

以下各实施例中未具体注明的操作细节可以参考本文为所引用的参考文献, 所采用的实验试剂和仪器设备均为常规商业可得的试剂或仪器。

10

实施例 1

野生型 Prp43 解旋酶和经修饰的解旋酶 Prp43 以及蛋白构建体均采用标准的分子生物学方法进行制备, 其原理和操作过程为本领域技术人员所熟知 (参见本文为所引用的参考文献)。

15

20

25

N 端截短的野生型 Prp43 解旋酶 (即 T61-A764 片段): 将 N 端截短 Prp43 解旋酶 T61-A764 片段 (对应于 SEQ ID NO: 1 的 Prp43 解旋酶氨基酸序列去除了 N 端结构域的 M1 至 N60 片段) 对应的核酸序列 (SEQ ID NO: 28, 由金斯瑞生物科技股份有限公司提供) 通过酶切连接方式连接至载体 pGS-21a (金斯瑞生物科技股份有限公司, 货号 SD0121) 中, 经测序验证正确后转化至表达感受态宿主细胞 BL21(DE3) (北京全式金生物技术有限公司, 货号 CD601-02) 中。从平板上挑选单克隆接种至 100ml 氨苄抗性的液体 LB 培养基, 37°C 过夜培养后第二天转接至大瓶培养基中扩大培养, 待 OD600 达到 0.4-0.8 左右时, 添加终浓度为 0.5mM 的异丙基-β-D-硫代半乳糖苷 (IPTG), 16°C 过夜诱导表达 12 小时左右, 低温离心收集的菌体经过裂解缓冲液重悬后高压均质破碎, 高速离心收集上清液, 进行后续的蛋白层析纯化, 具体包括镍离子亲和层析、离子交换层析和分子筛分离, 目标蛋白经酶切切除 GST 标签后经过镍离子亲和层析柱后收集流穿的目的蛋白, 切除 GST 标签后的目标蛋白使用 SDS-PAGE 凝胶电泳检测。切除标签后的截短 Prp43 蛋白 (去除了 N 端结构域的 M1 至 N60) 经过 SDS-PAGE 检测,

显示目的蛋白大小正确，可用于后续的检测和分析。

N端截短 Prp43 解旋酶 T61-A764 片段融合 GP 结构域蛋白突变体 Prp43-GP-2 (F181C/N623C/C508S) (即 SEQ ID NO: 27): 按照 N 端截短 Prp43 解旋酶 T61-A764 片段的相同制备方法进行制备, 只是将起始序列由对应于 N 端截短 Prp43 解旋酶 T61-A764 片段的核酸序列 (SEQ ID NO: 28) 替换为 SEQ ID NO: 30。切除标签后的蛋白构建体 Prp43-GP-2, 经过 SDS-PAGE 检测, 显示目的蛋白大小正确, 可用于后续的检测和分析。

按照上述相同方法, 采用不同的起始核酸序列, 分别制备经修饰的 N 端 (M1-N60) 截短的 Prp43 解旋酶和蛋白构建体: Prp43-2 (F181C/N623C/C508S)、经修饰的 Prp43 解旋酶 Prp43-3 (P352C/S619C/C508S)、N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP 和 N 端 (M1-N60) 截短的蛋白构建体 Prp43-GP-3 (P352C/S619C/C508S)。其中所用的启示核酸序列分别如下表 3 中所示。

表 3: 实施例 2 中使用的蛋白或蛋白构建体及其制备

描述	简写	用于制备的起始核酸序列	起始核酸提供商
N 端截短的野生型 Prp43 解旋酶	Prp43	SEQ ID NO: 28	金斯瑞生物科技股份有限公司
经修饰的 N 端截短的 Prp43 解旋酶 (F181C/N623C/C508S)	Prp43-2	SEQ ID NO: 41	金斯瑞生物科技股份有限公司
经修饰的 N 端截短的 Prp43 解旋酶 (P352C/S619C/C508S)	Prp43-3	SEQ ID NO: 42	金斯瑞生物科技股份有限公司
N 端截短的 Prp43 解旋酶蛋白构建体	Prp43-GP	SEQ ID NO: 43	金斯瑞生物科技股份有限公司
经修饰的 N 端截短的 Prp43 解旋酶蛋白构建体 (F181C/N623C/C508S)	Prp43-GP-2	SEQ ID NO: 30	金斯瑞生物科技股份有限公司
经修饰的 N 端截短的 Prp43 解旋酶蛋白构建体 (P352C/S619C/C508S)	Prp43-GP-3	SEQ ID NO: 44	金斯瑞生物科技股份有限公司

实施例 2

本实施例对 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶、经修饰的 Prp43 解旋酶 Prp43-2 (F181C/N623C/C508S)、经修饰的 Prp43 解旋酶 Prp43-3 (P352C/S619C/C508S)、N 端截短的蛋白构建体 Prp43-GP、N 端截短的蛋

白构建体 Prp43-GP-2 (F181C/N623C/C508S) 和 N 端截短的蛋白构建体 Prp43-GP-3 (P352C/S619C/C508S) 在结合或孵育单链 DNA 或单链 RNA 底物时的 ATP 水解活性进行了测试。

(1) 材料与方法

- 5 本实施例使用吸光光度法对 Prp43 解旋酶的 ATPase 水解活性进行检测。具体步骤是，准备含有 50uM 磷酸盐的预混溶液，将 50uL 的磷酸盐标准溶液移入 950 uL 的超纯水中，给管道编号。

表 4： 标准品制备

#	Premix + water	Final Vol (uL)	Phosphate Conc (uM)	pmoles Phosphate in 50 uL
1	200 L + 0 L	200	50	2,500
2	150 L + 50 L	200	37.5	1875
3	125 L + 75 L	200	31.25	1,562.5
4	100 L + 100 L	200	25	1250
5	50 L + 150 L	200	12.5	625
6	0 L + 200 L	200	0	0

- 10 在 96 孔板的重复孔中加入 25nM Prp43 解旋酶样品，加入 0.5nM M13ssDNA，用测试 Buffer (10mM HEPES,600mM KCL,5mM Mg²⁺)将样品加入 10 uL 的最终体积，30°C反应 30 min，加入终浓度为 1mM TMAD，30°C反应 30 min。在重复的孔中加入 10 uL(10mM HEPES,50mM KCL,5mM Mg²⁺)缓冲液作为阴性对照。高水平的磷酸盐会导致样品背景，要更正此背景。反应混合物加入后，立即在每个背景空白孔中加入 160 uL 工作试剂，使反应
- 15 停止。不需要进行最初的 30 min 的孵化，然后可以从样本读数中减去背景空白读数。根据表 4、表 5 的方案设置反应组合。每个样品、背景空白或阴性对照反应需要 70uL 的反应混合物。

表 5： 样品制备

试剂	样品， 背景空白及阴性对照
测试Buffer (10mM HEPES,600mM KCl,5mM Mg ²⁺)	66 uL
4 mM ATP	4 uL

20

每孔加 70 uL 反应混合液，空白背景，阴性对照。不要添加到标准品中。

在室温下培养反应 30 min。在每个孔中加入 160uL 工作试剂，在室温下再孵育 15 min，终止酶反应，生成比色产物。酶标仪读取所有样品、标准品和对照品的 600-660 nm 处的吸光度[620 nm 处的最大吸光度(A620)]。

(2) 结果

5 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶和修饰的 Prp43 解旋酶或蛋白构建体在分别和 DNA 或 RNA 结合后的 ATP 水解活性如图 2 和图 3 所示。从图 2、图 3 可知：在 Prp43 解旋酶或突变体的 C 端融合了 G-Path 激活结构域之后，该酶的 ATP 水解活性得到了明显的提升；在 Prp43 解旋酶中引入两个半胱氨酸后，同样改善了 ATP 水解活性。

10

实施例 3

本实施例使用荧光偏振方法对 N 端 (M1-N60) 截短的野生型 Prp43 解旋酶或经修饰的蛋白构建体 Prp43-GP 和 Prp43-GP-2(F181C/N623C/C508S) 在单链 DNA 的亲和力进行了测试。

15

(1) 材料与方法

N 端 (M1-N60) 截短的野生型解旋酶或修饰后的解旋酶按如下浓度梯度进行稀释：800nM, 400nM, 200nM, 100nM, 50nM, 25nM, 12.5nM, 6.25nM, 3.125nM, 1.56nM, BLANK, 酶与 10nM 单链 DNA 底物在结合 Buffer (10mM HEPES, 50mM KCl, 5% Glycerol, pH7.0) 中孵育 20min 后，在 530nm 激发光和 560nm 发射光下读取其偏振值并拟合绘制亲和力曲线，每个酶浓度设置
20 三个重复。

(2) 结果

拟合结果如图 4 所示，N 端 (M1-N60) 截短的 Prp43 解旋酶在 C 端融合了 G-Path 结构域，即 Prp43-GP 解旋酶，或在 Prp43-GP 基础上进行定点
25 突变的修饰酶 Prp43-GP-2 (F181C/N623C/C508S) 与单链 DNA 的亲和力在低盐条件下与野生型没有明显的差别。

实施例 4

本实施例通过凝胶迁移实验来检测 N 端 (M1-N60) 截短的野生型 Prp43

解旋酶或经修饰的蛋白构建体 Prp43-GP 和 Prp43-GP-2 (F181C/N623C/C508S) 结合 DNA 的情况, 包括 TMAD 催化剂催化突变体中突变体位点 F181C 和 N623C 之间形成二硫键后对核酸结合力的增强作用。

5 (1) 材料和方法

本实验条件如下, 在 Buffer (10mM HEPES, 50mM KCl, PH7.0) 中加入 30nM 的 FAM 荧光基团标记的单链多聚胸腺嘧啶底物 T44-37-FAM, 然后分别加入使用终浓度为 120nM 的野生型 Prp43 解旋酶和修饰修饰的 Prp43-2、Prp43-GP-2 解旋酶, 在 30°C 孵育 1.5h; 使用终浓度为酶的 1000 倍 TMAD 交联剂进行催化突变位点半胱氨酸的交联, 30°C 孵育 1.5h。

(2) 结果

凝胶迁移实验结果如图 5 所示, 野生型 Prp43 解旋酶与 DNA 结合后在电泳条件下酶与核酸脱落情况较严重, 修饰体 Prp43-GP 解旋酶与 DNA 结合后在电泳条件下酶与核酸脱落相比野生型 Prp43 解旋酶稍微好点, 而修饰体突变体 Prp43-GP-2 与 DNA 结合效果较好, 无论经过 TMAD 处理与否都没有明显的酶从核酸上脱落现象。

实施例 5

本实施例显示了 N 端 (M1-N60) 截短的修饰体突变体解旋酶 Prp43-GP-2 (F181C/N623C/C508S) 控制 DNA 构建体 X 穿过 MspA 纳米孔的运动。

(1) 材料和方法

制备如图 6 所示的 DNA 构建体 X: A 区对应序列 (SEQ ID NO: 32) 的 5' 末端连接到 4 个 iSpC3 间隔区 (B 区), 该间隔区连接到 C 区对应序列 (SEQ ID NO: 33) 的 3' 末端, C 区序列的 5' 末端连接到 D 区对应序列 (SEQ ID NO: 34), 该构建体的 E 区对应序列 (SEQ ID NO: 35) 与 F 区对应序列 (SEQ ID NO: 36, 其具有 3' 胆固醇系链) 杂交。将浓度为 10uM 的 A、B、C、D 区段合成连接在一起的片段, 与 E 片段、F 片段按照 1:1:1 比例加入到退火缓冲液 (10mM Tris, pH7.0, 50mM NaCl) 中, 按照 98°C 10min, -0.1°C/0.6s, 300 个循环, 65°C 5min, -0.1°C/0.6s, 400 个循环的流程进行退火处理 (其中, A、B、C、D、E、F 片段由生工生物工程 (上海) 股份有限

公司提供)。

将制备的 DNA 构建体 X 和修饰体突变体解旋酶 Prp43-GP-2 (F181C/N623C/C508S) 或 N 端截短的野生型 Prp43-GP 在 25°C 的缓冲液 (10mM HEPES, pH 8.0, 50mM NaCl, 5% 甘油) 中一起预孵育 30 分钟, 加入 1000 倍于解旋酶浓度的 TMAD 催化剂室温孵育处理 30 分钟。由嵌入 1,2-二乙醇酰基-甘油-3-胆碱磷酸脂质双分子层的 MspA 纳米孔 (MspA 蛋白序列为 SEQ ID NO: 31, 按照 Michael Faller et al., “The Structure of a Mycobacterial Outer-Membrane Channel”, *Science* 303, 1189 (2004); DOI: 10.1126/science.1094114 所述进行制备) 获得电测量信号。通过 Montal-Mueller 技术, 在 PTFE 膜上的 ~25 μ m 直径孔穴形成双分子层, 隔开两个约 100 μ L 的缓冲溶液。所有实验在所述缓冲液中进行。使用装配有数字转换器的放大器测定单通道电流。将 Ag/AgCl 电极连接到所述缓冲液中使得顺式隔间连接到放大器的接地端, 并且反式隔间连接到活性电极。

在所述双分子层实现单孔之后, 将 DNA 多核苷酸和修饰体突变体解旋酶 Prp43-GP-2 (F181C/N623C/C508S) 或 N 端截短的野生型 Prp43-GP 的复合物添加到电生理学室的顺式隔间的 70 μ L 缓冲液中以引发解旋酶-DNA 复合物在所述纳米孔的捕获。根据需要向所述顺式隔间添加二价金属 (5mM MgCl₂) 和 NTP (2.86 μ M ATP) 激活解旋酶 ATP 酶活性。实验在 +180mV 的恒定电势下实施。

(2) 结果

结果显示 DNA 构建体 X 的移动被解旋酶 Prp43-GP-2 (F181C/N623C/C508S) 控制, 结果见图 7, Prp43-GP-2 解旋酶控制了接近 200bp 的 DNA 构建体穿过所述纳米孔的移位。相对应地, N 端截短的野生型 Prp43 (T61-A764 片段) 或构建体 Prp43-GP 则难以获得构建体 X 的 A/B/C/D 片段通过纳米孔所产生的的持续电流信号。

实施例 6

本实施例显示了 N 端 (M1-N60) 截短的修饰体突变体解旋酶 Prp43-GP-2 (F181C/N623C/C508S) 控制 RNA 构建体 Y 穿过 MspA 纳米孔的运动。

(1) 材料与amp;方法

制备如图 8 所示的 RNA 构建体：将 D 区对应序列 (SEQ ID NO: 37) 的 3'末端连接到 20 个 iSpC3 间隔区 (A 区)，其 5'末端连接到 4 个 iSpC3 间隔区 (B 区)，该间隔区连接到 C 区对应序列 (SEQ ID NO: 38) 的 3'末端，该构建体的 E 区对应序列 (SEQ ID NO: 39) 与 F 区对应序列 (SEQ ID NO: 40) 杂交。将浓度为 10uM 的 A、B、C、D 区段合成连接在一起的片段，与 E 片段、F 片段按照 1:1:1 比例加入到退火缓冲液 (10mM Tris, pH7.0, 50mM NaCl) 中，按照 98°C 10min, -0.1°C/0.6s, 300 个循环, 65°C 5min, -0.1°C/0.6s, 400 个循环的流程进行退火处理 (其中，A、B、C、D、E、F 片段由生工生物工程 (上海) 股份有限公司提供)。

将制备的 RNA 构建体和 Prp43-GP-2 或 N 端截短的野生型 Prp43-GP 在 30°C 的缓冲液 (10mM HEPES, pH 7.0, 50mM NaCl) 中一起预孵育 30 分钟。由嵌入 1,2-二乙醇酰基-甘油-3-胆碱磷酸脂质双分子层的 MspA 纳米孔 (MspA 蛋白序列为 SEQ ID NO: 31, 按照 Michael Faller et al., "The Structure of a Mycobacterial Outer-Membrane Channel", *Science* 303, 1189 (2004); DOI: 10.1126/science.1094114 所述进行制备) 获得电测量信号。通过 Montal-Mueller 技术, 在 PTFE 膜上的 ~25μm 直径孔穴形成双分子层, 隔开两个约 100μL 的缓冲溶液。所有实验在所述缓冲液中进行。使用装配有数字转换器的放大器测定单通道电流。将 Ag/AgCl 电极连接到所述缓冲液中使得顺式隔间连接到放大器的接地端, 并且反式隔间连接到活性电极。

在所述双分子层实现单孔之后, 将 RNA 多核苷酸构建体和 Prp43-GP-2 解旋酶或 N 端截短的野生型 Prp43-GP 添加到电生理学室的顺式隔间的 70μL 缓冲液中以引发解旋酶-RNA 复合体在所述纳米孔的捕获。根据需要向所述顺式隔间添加二价金属 (5mM MgCl₂) 和 NTP (5mM ATP) 激活解旋酶 ATP 酶活性。实验在 +180mV 的恒定电势下实施。

(2) 结果

结果显示 RNA 构建体被 Prp43-GP-2 解旋酶控制的 RNA 移动, Prp43-GP-2 解旋酶控制的 RNA 移动的结果见图 9。Prp43-GP-2 解旋酶控制的 RNA 移动为 3 秒长并对应于接近 30bp 的 RNA 构建体穿过所述纳米孔的移位。

相对应地，N 端截短的野生型 Prp43（T61-A764 片段）或 N 端截短的构建体 Prp43-GP 则难以获得构建体 Y 的 A/B/C/D 片段通过纳米孔所产生的的持续电流信号。

- 5 本文中描述了本发明的优选实施方式和具体实施例，但是这些实施方式和实施例仅作为示例提供，而非用来限制本发明。在不脱离本发明的情况下，本领域技术人员现在将想到许多变化、改变和替换。因此，本发明还应涵盖任何此类替代方案、修改、变体或等效形式。

权利要求书

1. 一种经修饰的 Prp43 解旋酶，包括 RecA1 结构域、RecA2 结构域和 Ratchet 结构域，相对于对应的野生型 Prp43 解旋酶或其片段而言所述经修饰的 Prp43 解旋酶包括在选自 RecA1 结构域、RecA2 结构域、Ratchet 结构域的至少一个结构域中引入的 1、2、3、4、5、6、7、8、9、10、11 个或多个半胱氨酸的插入或置换，和/或 1、2、3、4、5、6、7、8、9、10、11 个或多个非天然氨基酸的插入或置换。

2. 如权利要求 1 所述的经修饰的 Prp43 解旋酶，其中所述引入的半胱氨酸残基或非天然氨基酸残基位于对应于 SEQ ID NO: 1 的 M157、Q161、D165、F181、E182、N183、R324、L328、E332、R335、P353、L351、P352、H354、D321、E320、R358、P563、A564、N565、D603、K605、K606、H609、Y615、R616、S619、N623、A626 或 K630 中的任一个或两个以上的位置，优选地位于对应于 SEQ ID NO: 1 的 F181、P352、S619 或 N623 中的任一个或两个以上的位置。

3. 如权利要求 1 或 2 所述的经修饰的 Prp43 解旋酶，其中所述野生型 Prp43 解旋酶的片段是 Prp43 解旋酶的 N 端结构域被去除后所得片段，优选地去除 N 端的位置 1 开始的至少 96 个、至少 90 个、至少 80 个、至少 70 个、至少 60 个、至少 50 个、至少 40 个或至少 30 个残基。

4. 如权利要求 1 至 3 任一项所述的经修饰的 Prp43 解旋酶，其中所述经修饰的 Prp43 解旋酶还包括对于一个或多个半胱氨酸残基的置换，优选为对应于 SEQ ID NO: 1 的 C148、C214、C303、C323、C377、C441、C508、C543、C608 的一个或多个半胱氨酸残基被置换，更优选地半胱氨酸残基被置换为丙氨酸、甘氨酸、缬氨酸、异亮氨酸、亮氨酸、苯丙氨酸、酪氨酸、丝氨酸、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、组氨酸、蛋氨酸、色氨酸、谷氨酰胺、天冬酰胺或脯氨酸残基。

5. 如权利要求 1 至 4 任一项所述的经修饰的 Prp43 解旋酶，其中所述被引入的半胱氨酸残基和非天然氨基酸残基总个数为 2 个或以上，并且在至少一个被引入的半胱氨酸或非天然氨基酸残基与另一个被引入的半胱氨酸

酸或非天然氨基酸残基之间形成相互连接。

6. 如权利要求 5 所述的经修饰的 Prp43 解旋酶，其中所述连接选自共价连接或氢键连接或静电相互作用或 π - π 相互作用或疏水相互作用等，优选共价连接。

5 7. 如权利要求 6 所述的经修饰的 Prp43 解旋酶，其中所述共价连接是 -S-S 键或者是通过选自碳酰氯、马来酰亚胺、活性酯、琥珀酰亚胺、叠氮化物、烷烃、烯烃、炔烃、聚乙二醇(PEGs)、多肽、多糖、脱氧核糖核酸(DNA)、肽核酸(PNA)、苏糖核酸(TNA)、甘油核酸(GNA)、聚酰胺或 TMAD 的交联剂或催化剂实现的共价连接。

10 8. 如权利要求 1 至 7 任一项所述的经修饰的 Prp43 解旋酶，其中所述经修饰的 Prp43 解旋酶还包含选自以下组的一个或多个氨基酸修饰：

- (a) 一个或多个与核苷酸相互作用的氨基酸被置换；
- (b) 一个或多个与 NTP 和/或二价金属离子结合相关的氨基酸被置换；
- (c) 一个或多个与跨膜孔相互作用的氨基酸被置换；
- 15 (d) 降低 Prp43 解旋酶的表面上的负电荷的进一步修饰。

9. 如权利要求 1 至 8 任一项所述的经修饰的 Prp43 解旋酶，其来源于 *Chaetomium thermophilum*、*Bathycoccus prasinus*、*Uncultured bacterium*、*Archaeon*、*Parcubacteria*、*Sorangium cellulosum*、*Candidatus Sungbacteria*、*Mycolicibacterium chitae*、*Parcubacteria*、*Thermodesulforhabdus norvegica*、
20 *Deltaproteobacteria*、*Puniceicoccales*、*Desulfobacterium vacuolatum* 或 *Desulfobacter sp.* 或来源于病毒基因组 (viral metagenome)。

10. 如权利要求 1 至 8 任一项所述的经修饰的 Prp43 解旋酶，其中所述野生型 Prp43 解旋酶是选自具有以下序列之一的 Prp43 解旋酶：SEQ ID NO: 1、SEQ ID NO: 2、SEQ ID NO: 3、SEQ ID NO: 4、SEQ ID NO: 5、SEQ ID
25 NO: 6、SEQ ID NO: 7、SEQ ID NO: 8、SEQ ID NO: 9、SEQ ID NO: 10、SEQ ID NO: 11、SEQ ID NO: 12、SEQ ID NO: 13、SEQ ID NO: 14、SEQ ID NO: 15。

11. 如权利要求 1 至 8 任一项所述的经修饰的 Prp43 解旋酶，其与对应的野生型的 Prp43 解旋酶的氨基酸序列具有至少 30%、至少 40%、至少 50%、

至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或至少 99.9%的同源性。

12. 如权利要求 1 至 8 任一项所述的经修饰的 Prp43 解旋酶，其来源于 *Chaetomium thermophilum*，优选地其与 SEQ ID NO: 1 的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%或至少 99.9%的同源性。

13. 如权利要求 1 至 3 任一项所述的经修饰的 Prp43 解旋酶，其来源于 *Chaetomium thermophilum*，且所述引入的半胱氨酸残基或非天然氨基酸残基位于对应于 SEQ ID NO: 1 的 F181、P352、S619 或 N623 中的任一个或多个位置。

14. 如权利要求 13 所述的经修饰的 Prp43 解旋酶，其为经修饰的 SEQ ID NO: 1 的 T61-A764 片段，且所述修饰选自 F181C/N623C/C508S 和 P352C/S619C/C508S。

15. 如权利要求 1 至 14 任一项所述的经修饰的 Prp43 解旋酶，其为寡聚体形式，包含一个或多个如权利要求 1 至 12 任一项所述的经修饰的 Prp43 解旋酶。

16. 一种蛋白构建体，其包括如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶，以及在所述 Prp43 解旋酶的 C 端或 N 端融合的辅助激活蛋白 Paf1 的 G-Path 结构域或含有 G-Path 结构域的 Paf1 片段。

17. 如权利要求 16 所述的蛋白构建体，其中包含一个或多个所述经修饰的 Prp43 解旋酶。

18. 如权利要求 16 或 17 所述的蛋白构建体，其中所述辅助激活蛋白 Paf1 是来源于 *Chaetomium thermophilum var. thermophilum*、*Thermothielavioides terrestris*、*Thermothelomyces thermophilus*、*Podospira anserina*、*Neurospora tetrasperma*、*Coniochaeta sp.*、*Monosporascus sp.*、*Hypoxylon sp.*、*Madurella mycetomatis* 或 *Coniochaeta pulveracea* 的 Paf1。

19. 如权利要求 16 或 17 所述的蛋白构建体，其中所述辅助激活蛋

白 Paf1 的氨基酸序列选自 SEQ ID NO: 16、SEQ ID NO: 17、SEQ ID NO: 18、SEQ ID NO: 19、SEQ ID NO: 20、SEQ ID NO: 21、SEQ ID NO: 22、SEQ ID NO: 23、SEQ ID NO: 24、或 SEQ ID NO: 25，或具有与 SEQ ID NO: 16、SEQ ID NO: 17、SEQ ID NO: 18、SEQ ID NO: 19、SEQ ID NO: 20、SEQ ID NO: 21、SEQ ID NO: 22、SEQ ID NO: 23、SEQ ID NO: 24 或 SEQ ID NO: 25 其中之一的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或至少 99.9% 的同源性的变体的氨基酸序列，并且所述辅助激活蛋白 Paf1 具有激活 Prp43 解旋酶的功能。

20. 如权利要求 16 或 17 所述的蛋白构建体，其中所述 Paf1 的 G-Path 结构域为 SEQ ID NO: 16 序列的 K662-G742 片段（SEQ ID NO: 26）或与 SEQ ID NO: 26 的氨基酸序列具有至少 30%、至少 40%、至少 50%、至少 60%、至少 70%、至少 80%、至少 85%、至少 90%、至少 91%、至少 92%、至少 93%、至少 94%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99% 或至少 99.9% 的同源性的变体的氨基酸序列，并且所述变体具有激活 Prp43 解旋酶的功能。

21. 如权利要求 16 所述的蛋白构建体，其中所述 Prp43 解旋酶为 SEQ ID NO: 1 的 T61-A764 片段，并且在对应于 SEQ ID NO: 1 的 F181、P352、S619 或 N623 中的任一个或两个以上的位置处具有引入的 1 个或多个半胱氨酸的插入或置换和/或非天然氨基酸的插入或置换，且所述辅助激活蛋白 Paf1 的氨基酸序列为 SEQ ID NO: 16。

22. 如权利要求 16 所述的蛋白构建体，其中所述 Prp43 解旋酶为 SEQ ID NO: 1 的 T61-A764，并且还具有选自 F181C/N623C/C508S 和 P352C/S619C/C508S 的修饰，且所述 Prp43 解旋酶的 C 端与氨基酸序列为 SEQ ID NO: 26 的多肽融合。

23. 一种编码如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体的核酸。

24. 如权利要求 23 所述的核酸，其中所述核酸被包含在选自于质粒、

病毒或噬菌体的载体中。

25. 一种包含如权利要求 23 所述核酸的表达载体。

26. 如权利要求 25 所述的表达载体,其中所述表达载体选自于质粒、病毒或噬菌体。

5 27. 如权利要求 25 或 26 所述的表达载体,其中所述表达载体进一步包含用于控制所述核酸的表达的调控元件。

28. 如权利要求 27 所述的表达载体,其中所述调控元件是与所述核酸可操作地连接的启动子。

10 29. 如权利要求 28 所述的表达载体,其中所述启动子选自 T7、trc、lac、ara 或 λ L。

30. 包含如权利要求 23 或 24 所述的核酸或包含如权利要求 25 至 28 任一项所述表达载体的宿主细胞。

31. 如权利要求 30 所述的宿主细胞,其为大肠杆菌(*Escherichia coli.*)。

15 32. 一种制备如权利要求 16 至 22 任一项所述的蛋白构建体的方法,包括:提供 SEQ ID NO: 1 的多肽或其变体和 SEQ ID NO: 26 的多肽或其变体,在 SEQ ID NO: 1 的多肽或其变体中引入至少一个半胱氨酸残基和/或至少一个非天然氨基酸,然后在所得多肽的 C 端或 N 端融合 SEQ ID NO: 26 的多肽或其变体,形成所述蛋白构建体。

20 33. 一种制备如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体的方法,包括:包括培养如权利要求 30 或 31 所述的宿主细胞,并进行诱导表达,然后纯化所得表达产物。

25 34. 一种控制多核苷酸分子移动的方法,包括将所述多核苷酸分子与如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体接触。

35. 如权利要求 34 所述的控制多核苷酸分子移动的方法,其中所述多核苷酸分子被控制穿过纳米孔,所述的纳米孔为跨膜孔。

36. 如权利要求 35 所述的控制多核苷酸分子移动的方法,其中所述跨膜孔选自蛋白孔、固态孔或生物与固态杂交的孔,优选地所述的蛋白孔选

自于耻垢分枝杆菌孔蛋白 A、耻垢分枝杆菌孔蛋白 B、耻垢分枝杆菌孔蛋白 C、耻垢分枝杆菌孔蛋白 D、溶血素、胞溶素、白细胞介素、外膜孔蛋白 F、外膜孔蛋白 G、外膜磷脂酶 A、WZA 或奈瑟氏菌自转运脂蛋白。

37. 一种表征目标多核苷酸的方法，所述的方法包括：

5 (a) 将目标多核苷酸与如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体接触，使得所述 Prp43 解旋酶或蛋白构建体控制所述目标多核苷酸移动穿过一纳米孔；

(b) 获取目标多核苷酸中的核苷酸与所述纳米孔相互作用时的一个或多个特征，从而表征所述目标多核苷酸。

10 38. 如权利要求 37 所述的表征目标多核苷酸的方法，其中所述方法进一步包括跨所述纳米孔施加电势差的步骤。

39. 如权利要求 37 或 38 所述的表征目标多核苷酸的方法，其中所述方法使用一个或多个所述 Prp43 解旋酶或蛋白构建体。

15 40. 如权利要求 37 或 38 所述的表征目标多核苷酸的方法，其中所述的纳米孔为跨膜孔，所述跨膜孔选自蛋白孔、固态孔或生物与固态杂交的孔，优选地所述的蛋白孔选自于耻垢分枝杆菌孔蛋白 A、耻垢分枝杆菌孔蛋白 B、耻垢分枝杆菌孔蛋白 C、耻垢分枝杆菌孔蛋白 D、溶血素、胞溶素、白细胞介素、外膜孔蛋白 F、外膜孔蛋白 G、外膜磷脂酶 A、WZA 或奈瑟氏菌自转运脂蛋白。

20 41. 如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体在表征目标多核苷酸或控制目标多核苷酸穿过孔的移动中的用途。

25 42. 一种用于表征目标多核苷酸的分析装置，所述的分析装置包含一个或多个纳米孔、一个或多个如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体、以及一个或多个容器。

43. 如权利要求 42 所述的用于表征目标多核苷酸的分析装置，所述分析装置中还包括包含脂质双层的芯片，其中所述的纳米孔横跨脂质双层。

44. 如权利要求 42 或 43 所述的用于表征目标多核苷酸的分析装置，

所述分析装置中还包括缓冲剂、PCR 扩增试剂。

45. 如权利要求 42、43 或 44 所述的用于表征目标多核苷酸的分析装置，所述的分析装置是试剂盒或传感器。

46. 一种形成表征目标多核苷酸的传感器的方法，包括提供纳米孔，
5 和在所述纳米孔和如权利要求 1 至 15 任一项所述的经修饰的 Prp43 解旋酶或如权利要求 16 至 22 任一项所述的蛋白构建体之间形成复合物。

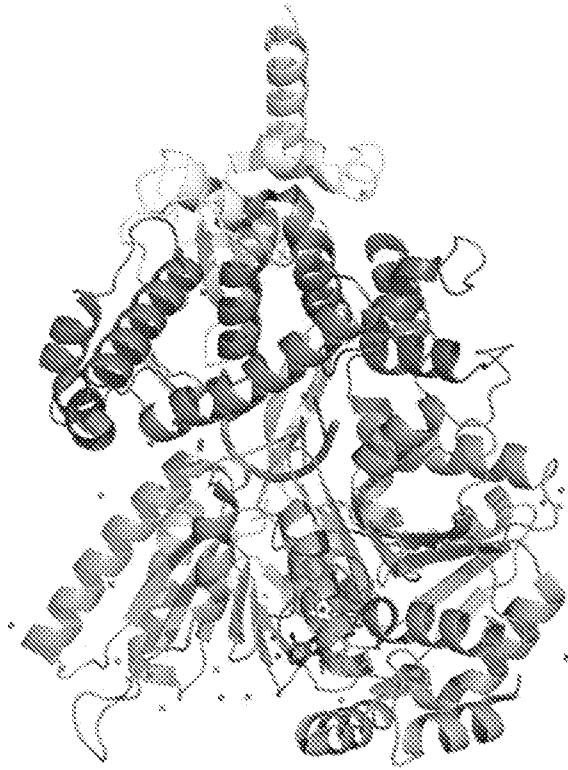


图 1

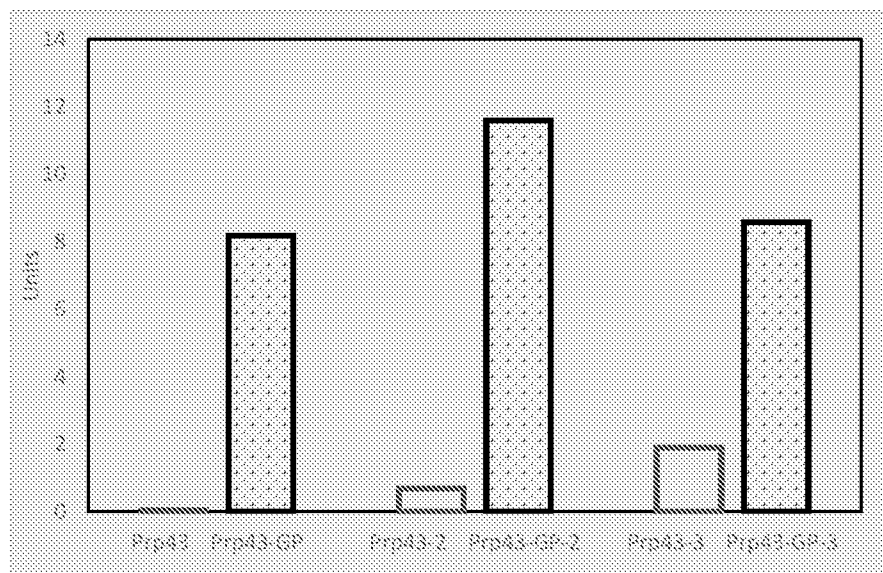


图 2

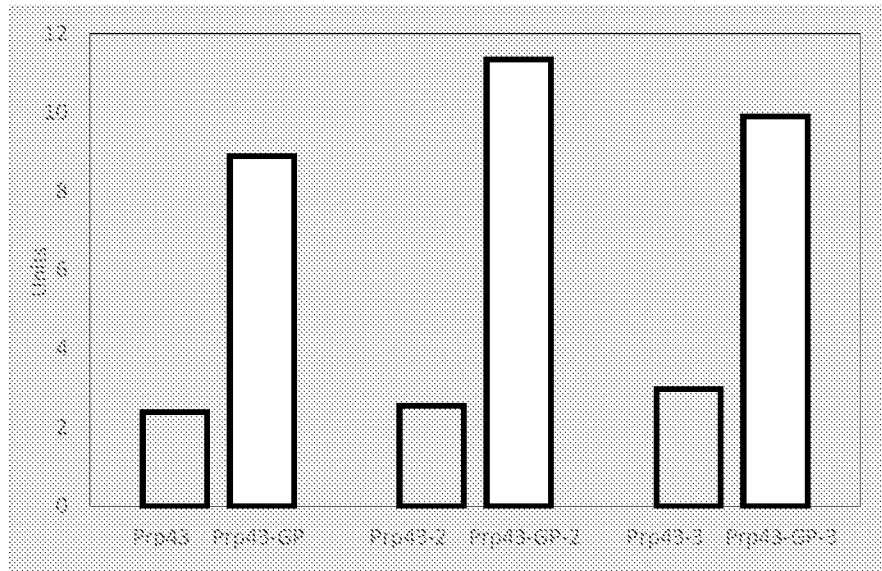


图 3

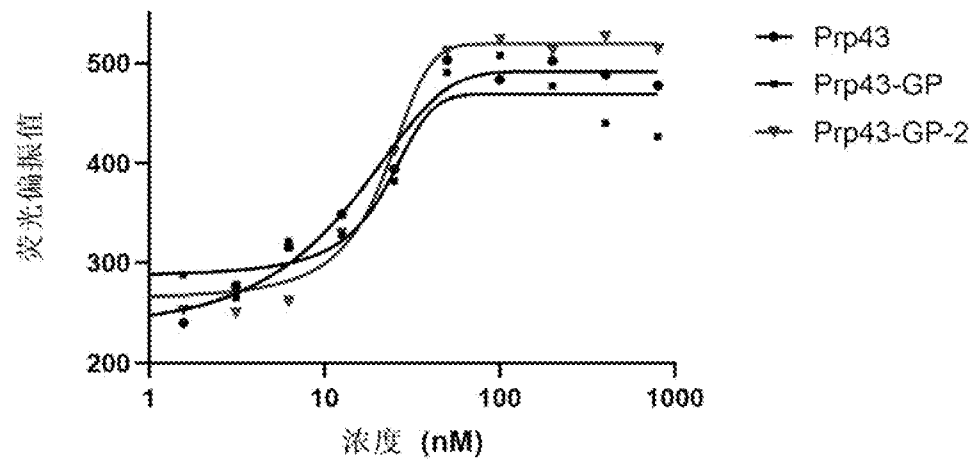


图 4

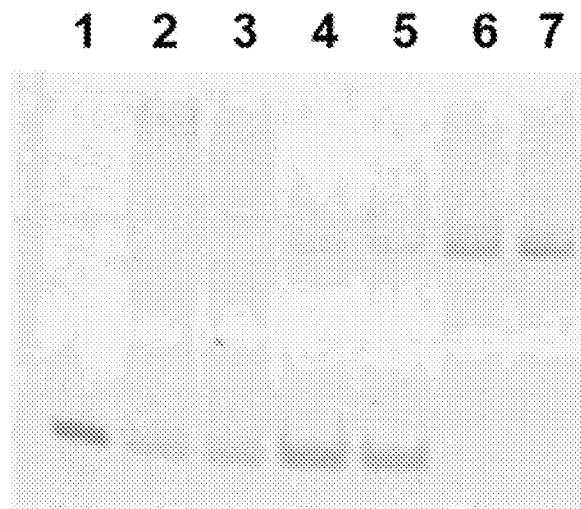


图 5

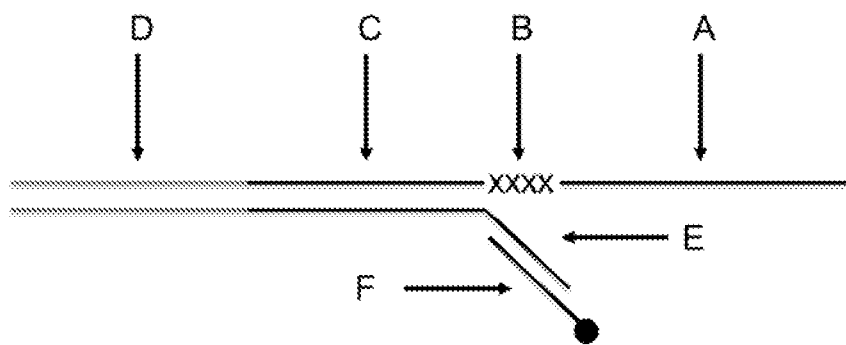


图 6

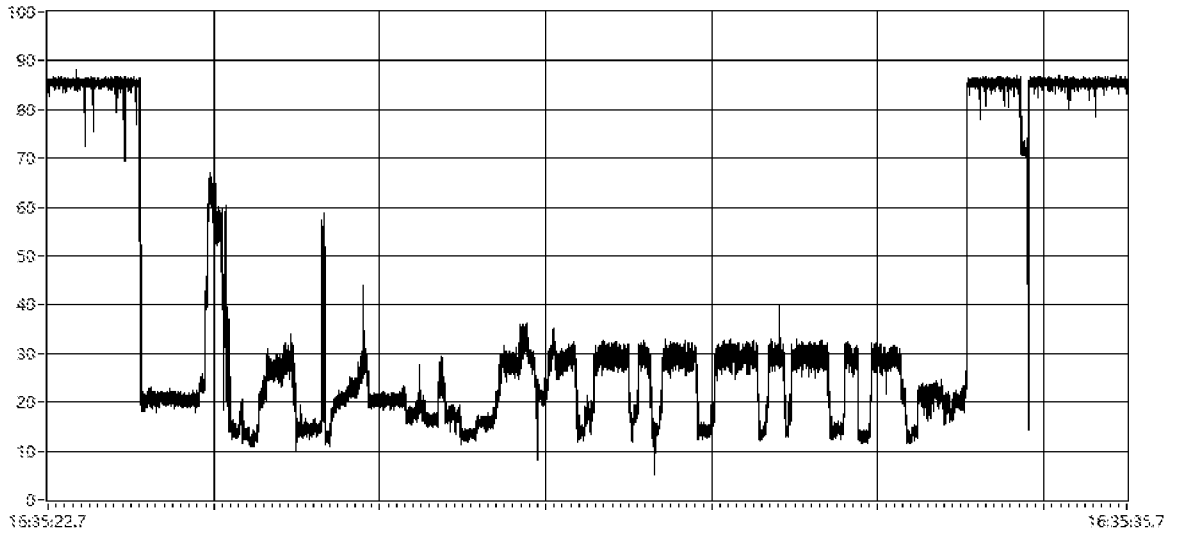


图 7

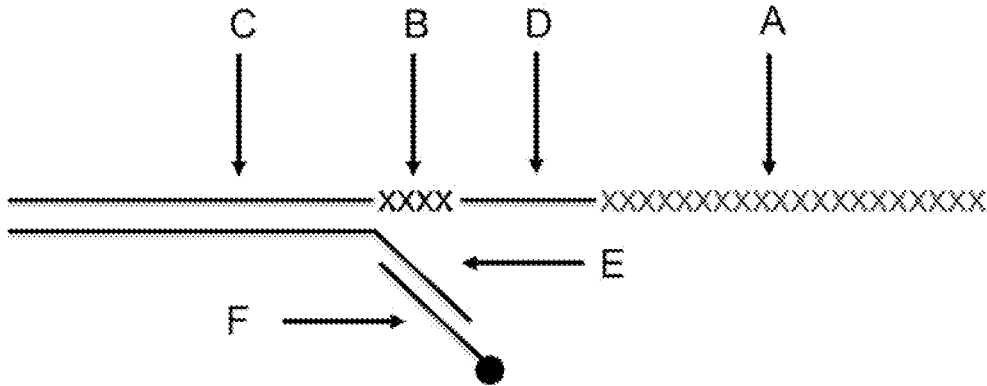


图 8

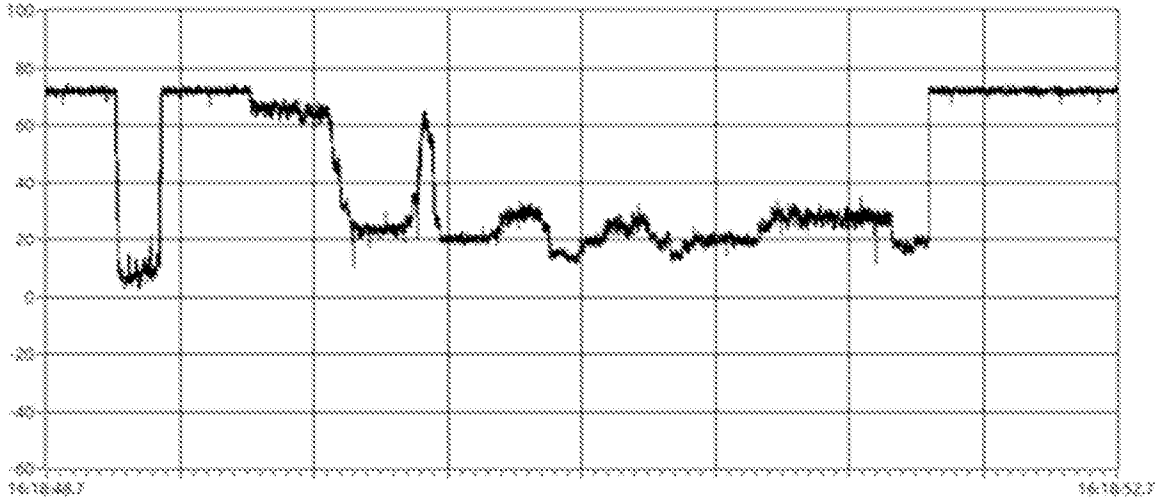


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2021/085609

A. CLASSIFICATION OF SUBJECT MATTER		
C12N 15/55(2006.01)i; C12N 9/14(2006.01)i; G01N 33/48(2006.01)i; C12Q 1/6869(2018.01)i; C12M 1/34(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) C12N,G01N,C12Q,C12M		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNTXT; CNABS; ENTXT; CJFD; VEN, ENTXTC, ISI WEB OF SCIENCE, PUBMED, CNKI: PRP43, Cysteine, reca, ratchet, paf1, dhx15, DEAH-box RNA helicase, mutation, reca2, pore, reca1, DEAH, DHX15, Chaetomium thermophilum, 突变, 半胱氨酸, 解旋酶, DEAH盒RNA解旋酶; GenBank+Embl+中国专利生物序列检索系统: 关于SEQ ID Nos: 1-16, 26的检索, GenBank+Embl+China Patent Biological Sequence Retrieval System: search for SEQ ID Nos: 1-16,26		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARCEL, J. T. et al. "Structural Insights into the Mechanism of the DEAH-box RNA Helicase Prp43" <i>eLife</i> , Vol. 6, 16 January 2017 (2017-01-16), ISSN: 2050-084X, see page 7, last paragraph, line 4, page 8, last paragraph, page 19, paragraph 2, supplement 1 of figure 4, table 3, and first paragraph of "Result" part	1-33
Y	MARCEL, J. T. et al. "Structural Insights into the Mechanism of the DEAH-box RNA Helicase Prp43" <i>eLife</i> , Vol. 6, 16 January 2017 (2017-01-16), ISSN: 2050-084X, see page 7, last paragraph, line 4, supplement 1 of figure 4, first paragraph of "Result" part	34-46
Y	CN 112147185 A (TSINGHUA UNIVERSITY) 29 December 2020 (2020-12-29) see claims 1-22	34-46
Y	CN 109266537 A (SHOUDU BIOTECH SUZHOU CO., LTD.) 25 January 2019 (2019-01-25) see claims 1-12	34-46
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 24 December 2021		Date of mailing of the international search report 04 January 2022
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China Facsimile No. (86-10)62019451		Authorized officer Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TANAKA, N. et al. "Mutations in PRP43 that uncouple RNA-dependent NTPase Activity and pre-mRNA Splicing Function" <i>Biochemistry</i> , Vol. 45, No. 20, 23 May 2006 (2006-05-23), ISSN: 0006-2960, pp. 6510-6521, see entire document	1-46
A	陈飞龙等人 (CHEN, Feilong et al.). "RNA聚合酶II相关因子1(Paf1)复合物的结构和功能研究进展 (Research Progress in Structure and Function of Polymerase-associated Factor 1 Complex)" <i>中国科学: 生命科学 (Science in China: Life Sciences)</i> , Vol. 49, No. 9, 31 December 2019 (2019-12-31), ISSN: 1674-7216, pp. 1143-1154, see entire document	1-46
A	CN 108676785 A (TIANJIN LAKESIDE POWERGENES SCIENCE DEVELOPMENT CO., LTD.) 19 October 2018 (2018-10-19) see entire document	1-46

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. forming part of the international application as filed:
 - in the form of an Annex C/ST.25 text file.
 - on paper or in the form of an image file.
 - b. furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. furnished subsequent to the international filing date for the purposes of international search only:
 - in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 - on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2021/085609

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112147185 A	29 December 2020	None	
CN 109266537 A	25 January 2019	None	
CN 108676785 A	19 October 2018	None	

国际检索报告

国际申请号

PCT/CN2021/085609

<p>A. 主题的分类</p> <p>C12N 15/55 (2006.01)i; C12N 9/14 (2006.01)i; G01N 33/48 (2006.01)i; C12Q 1/6869 (2018.01)i; C12M 1/34 (2006.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																	
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>C12N, G01N, C12Q, C12M</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNXTX;CNABS;ENTXT;CJFD;VEN, ENTXTC, ISI WEB OF SCIENCE, PUBMED, CNKI:PRP43, Cysteine, reca, ratchet, paf1, dhx15, DEAH-box RNA helicase, mutation, reca2, pore, reca1, DEAH, DHX15, Chaetomium thermophilum, 突变, 半胱氨酸, 解旋酶, DEAH盒RNA解旋酶; GenBank+Embl+中国专利生物序列检索系统: 关于SEQ ID Nos: 1-16, 26的检索</p>																	
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 第8页最后一段, 第19页第2段, 图4的补充1, 表3, "Result" 部分第一段</td> <td>1-33</td> </tr> <tr> <td>Y</td> <td>Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 图4的补充1, "Result" 部分第一段</td> <td>34-46</td> </tr> <tr> <td>Y</td> <td>CN 112147185 A (清华大学) 2020年12月29日 (2020 - 12 - 29) 参见权利要求1-22</td> <td>34-46</td> </tr> <tr> <td>Y</td> <td>CN 109266537 A (首度生物科技苏州有限公司等) 2019年1月25日 (2019 - 01 - 25) 参见权利要求1-12</td> <td>34-46</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 第8页最后一段, 第19页第2段, 图4的补充1, 表3, "Result" 部分第一段	1-33	Y	Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 图4的补充1, "Result" 部分第一段	34-46	Y	CN 112147185 A (清华大学) 2020年12月29日 (2020 - 12 - 29) 参见权利要求1-22	34-46	Y	CN 109266537 A (首度生物科技苏州有限公司等) 2019年1月25日 (2019 - 01 - 25) 参见权利要求1-12	34-46
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 第8页最后一段, 第19页第2段, 图4的补充1, 表3, "Result" 部分第一段	1-33															
Y	Marcel J Tauchert等人. "Structural insights into the mechanism of the DEAH-box RNA helicase Prp43" eLife, 第6卷, 2017年1月16日 (2017 - 01 - 16), ISSN: 2050-084X, 参见第7页最后一段第4行, 图4的补充1, "Result" 部分第一段	34-46															
Y	CN 112147185 A (清华大学) 2020年12月29日 (2020 - 12 - 29) 参见权利要求1-22	34-46															
Y	CN 109266537 A (首度生物科技苏州有限公司等) 2019年1月25日 (2019 - 01 - 25) 参见权利要求1-12	34-46															
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件</p>																	
<p>国际检索实际完成的日期</p> <p>2021年12月24日</p>		<p>国际检索报告邮寄日期</p> <p>2022年1月4日</p>															
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451</p>		<p>授权官员</p> <p>王启扬</p> <p>电话号码 62411211</p>															

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	Naoko Tanaka等人. "Mutations in PRP43 that uncouple RNA-dependent NTPase Activity and pre-mRNA Splicing Function" Biochemistry, 第45卷, 第20期, 2006年5月23日 (2006 - 05 - 23), ISSN: 0006-2960, 第6510 - 6521页, 参见全文	1-46
A	陈飞龙等人. "RNA聚合酶 II 相关因子1 (Paf1) 复合物的结构和功能研究进展" 中国科学: 生命科学, 第49卷, 第9期, 2019年12月31日 (2019 - 12 - 31), ISSN: 1674-7216, 第1143-1154页, 参见全文	1-46
A	CN 108676785 A (天津市湖滨盘古基因科学发展有限公司) 2018年10月19日 (2018 - 10 - 19) 参见全文	1-46

第1栏 核苷酸和/或氨基酸序列(续第1页第1. c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列, 国际检索是基于下列序列列表进行的:
- a. 作为国际申请的一部分提交的:
- 附件C/ST. 25文本文件形式
 - 纸件或图形文件形式
- b. 根据细则13之三. 1(a) 仅为国际检索目的以附件C/ST. 25文本文件形式与国际申请同时提交的:
- c. 仅为国际检索目的在国际申请日之后提交的:
- 附件C/ST. 25文本文件形式(细则13之三. 1(a))
 - 纸件或图形文件形式(细则13之三. 1(b)和行政规程第713段)
2. 另外, 在提交/提供了多个版本或副本的序列列表的情况下, 提供了关于随后提交的或附加的副本中的信息与申请时提交的作为申请一部分的序列列表的信息相同或未超出申请时提交的申请中的信息范围(如适用)的所需声明。
3. 补充意见:

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2021/085609

检索报告引用的专利文件	公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN 112147185 A	2020年12月29日	无	
CN 109266537 A	2019年1月25日	无	
CN 108676785 A	2018年10月19日	无	