

(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
2024 年 12 月 12 日 (12.12.2024)



(10) 国际公布号
WO 2024/250157 A1

(51) 国际专利分类号:

C12N 15/81 (2006.01) *G16B 20/30* (2019.01)
C12N 15/113 (2010.01) *G16B 30/00* (2019.01)
C12N 1/19 (2006.01) *C12R 1/865* (2006.01)
G16B 5/00 (2019.01)

(21) 国际申请号: PCT/CN2023/098388

(22) 国际申请日: 2023 年 6 月 5 日 (05.06.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳先进研究院 (SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY) [CN/CN]; 中国广东省深圳市南山区西丽街道大学城学苑大道1068号, Guangdong 518000 (CN)。

(72) 发明人: 陈业 (CHEN, Ye); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518000 (CN)。熊婧卉 (XIONG, Jinghui); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518000 (CN)。栗曾理 (LI, Zengli); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518000 (CN)。郭淑慧 (GUO, Shuhui); 中国广东省深圳市南山区西丽大学城学苑大道 1068 号, Guangdong 518000 (CN)。

(74) 代理人: 深圳五邻知识产权代理事务所 (普通合伙) (SHENZHEN WULIN INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市龙华区龙华街道清华社区和平东路幸福城商业大厦A1106, Guangdong 518000 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,

(54) Title: FULLY SYNTHETIC YEAST INDUCIBLE PROMOTER AND CONSTRUCTION METHOD THEREFOR

(54) 发明名称: 一种全合成的酵母诱导型启动子及其构建方法

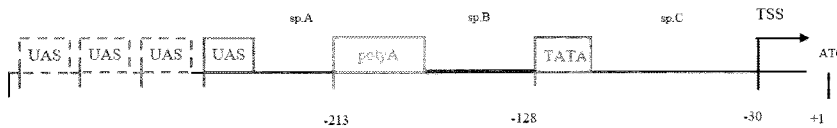


图 1

(57) Abstract: Provided are a fully synthetic yeast inducible promoter and a construction method therefor. By means of modularizing the promoter architecture, the present method splits the upstream activation region and core promoter region thereof, and uses rational design, computer model assistance and high-throughput library construction to screen spacer sequences. Promoter architecture using *Saccharomyces cerevisiae* as the chassis microorganism is explored, and a fully synthetic non-natural yeast inducible promoter sequence is obtained. The problems of there being an insufficient number of well-characterized promoters due to high background leakage of promoters, insufficient induction strength and limited types, or of homologous recombination in *Saccharomyces cerevisiae* easily occurring, are solved.

(57) 摘要: 提供了一种全合成的酵母诱导型启动子及其构建方法, 该方法通过将启动子的构架模块化, 对其上游激活区域和核心启动子区域进行拆分, 并利用理性设计、计算机模型辅助和高通量建库等方法对间隔序列进行筛选, 探索了以酿酒酵母作为底盘微生物的启动子的架构模式, 获得了全合成的非天然的酵母诱导型启动子序列, 解决了由于启动子本底泄露高、诱导强度不足、种类少导致的特征良好启动子数量不足, 或是酿酒酵母体内易发生同源重组的问题。

BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, 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))。

一种全合成的酵母诱导型启动子及其构建方法

技术领域

本发明涉及合成生物学技术领域，具体涉及一种全合成的酵母诱导型启动子及其构建方法。

背景技术

随着合成生物学的迅速发展，酿酒酵母作为底盘微生物表达代谢途径相关的酶，并用于合成天然产物或大众化合物的应用相继出现。例如，通过改造酿酒酵母启动子减缓或加速某些代谢途径，使得产物有更好的产量或纯度等。

目前，酿酒酵母已广泛应用于代谢工程和基因线路设计中，而启动子作为基因调控的关键元素，在这些应用中发挥了重要作用。在代谢工程中，通过使用诱导型启动子可以精确控制目标基因的表达水平，实现代谢途径的平衡和优化，或是建立基因回路以控制细胞行为。在基因线路设计中，启动子可以被用来构建生物传感器，响应外界信号并输出需要的数据。

依据酿酒酵母启动子的作用方式可将其分为组成型启动子和诱导型启动子。其中组成型启动子的活性相对稳定，不受刺激而改变，比如 pTDH3 和 pCYC1。大多数情况下，该类启动子的活性大小与菌株生长速率是相偶联的，和环境中的碳源含量有关。诱导型启动子的活性则依赖于化学或物理刺激，通过定时定量地添加某种刺激可以增强或减弱启动子的活性并改变基因的转录水平。用于刺激的常见诱导剂有半乳糖、无机磷酸盐、铜等，常用的诱导型启动子有 pGAL1、pPHO5、pCUP1 等。由于诱导性启动子能避免外源基因表达对早期酿酒酵母生长的影响，使得其在代谢工程中被广泛使用，此外，诱导型启动子也会影

响生物传感器的灵敏度和整体性能。

然而现有内源性的诱导型启动子通常难以满足优化代谢通量的要求，主要原因如下：其一，现有诱导型启动子的动态范围较差，且往往与酵母原生的调控不正交，因此特征良好的诱导型启动子数量不足。其二，同一启动子或核心启动子区域的多次使用容易产生同源重组，进而导致酿酒酵母合成途径的不稳定。例如，pGAL1 经常被重复用于不同酶的表达以构建代谢通路，为弥补启动子强度的不足必须增加基因拷贝数，而使用过多的半乳糖诱导启动子可能会耗尽转录激活子 Gal4p，进而干扰半乳糖的代谢；其三，作为优良生物传感器的组成部分，大多数天然诱导型启动子存在缺陷，如本底泄露表达过高、诱导强度不够等。其四，外源转入子的设计仍不明确。因此，需要开发更多的诱导合成启动子来提高目标产物的产量，减少宿主适应度损失的干扰等。

随着基因组学的快速发展和机器学习在合成生物学中的应用，有研究通过构建启动子文库并用卷积神经网络（CNN）技术建立数学模型，以输入序列预测启动子活性，但是这种建立启动子文库的筛选方法工作量极大，且有用的启动子数量少。相比之下，利用基因挖掘技术挖掘新的天然启动子发现了许多有用的序列，但基因组所包含的诱导型启动子仍有限。

此外，用于新型酿酒酵母启动子开发的技术还包括位点选择性突变技术和杂交启动子技术等。其中，位点选择性突变技术通过建立微调强度的启动子文库调控目标基因的表达，如易错 PCR（ep-PCR）、饱和诱变等，但获得的启动子具有相似的序列，可能导致序列不稳定，发生同源重组。杂交启动子技术则采用组装替换的策略，将所需功能的各种结果部件连接在一起，或是替换核心启动子区域，获得具有多种功能的新的杂交启动子。例如，将上游激活序列多

聚化，创建带有重复操作符的启动子，以招募更多的转录因子进行相互作用。有研究人员将上游激活序列 ZF 串联，发现当包含八个串联上游激活序列 ZF 时，系统的转录输出增加了近 60 倍，提高了启动子的强度。

虽然上述部分方法可以用于构建非天然的诱导型启动子，但所构建的启动子的部分序列仍源于酿酒酵母的天然序列，存在一定的局限性。一是，特征良好的酿酒酵母启动子数量不多，比如启动子强度不足、种类少、同源性强等；二是，启动子不适合基因线路设计，比如序列长、响应范围小、本底表达过高等。因此，开发一种高效全合成的诱导型启动子架构的设计方案显得尤为重要。

发明内容

本发明的目的在于克服现有技术不足，解决上述背景技术中的至少一种缺陷。

为实现上述目的，本发明提供以下技术方案：

本发明的第一方面，提供一种全合成的酵母诱导型启动子的构建方法包括：确定所述启动子的基本架构，所述启动子从 5'→3'依次包括操作性连接的：上游激活序列、第一间隔序列、核小体不利序列、第二间隔序列、TATA 盒、第三间隔序列、转录起始位点；利用理性设计方法、模型建库方法和/或高通量建库方法筛选所述第一间隔序列、所述第二间隔序列和所述第三间隔序列的序列碱基；以及组合所述第一间隔序列、所述第二间隔序列和所述第三间隔序列，获得所述启动子。

在一些实施例中，所述筛选第一间隔序列、第二间隔序列和第三间隔序列的序列碱基包括：通过理性设计方法筛选所述第一间隔序列的序列碱基；通过模型建库方法筛选所述第二间隔序列的序列碱基；和通过高通量建库方法筛选

所述第三间隔序列的序列碱基。

进一步地，所述通过模型建库方法筛选第二间隔序列的序列碱基包括：优化转录模型并随机生成含所述第二间隔序列的启动子片段；利用优化后的转录模型预测含所述第二间隔序列的启动子片段的强度；以及筛选含所述第二间隔序列的目标启动子。

进一步地，所述优化转录模型包括：向转录模型中输入序列组合，所述序列组合由不同长度的序列摆放在不同位置得到；以及以模型预测准确性为标准筛选所述序列组合，获得优化的序列位置及长度。

进一步地，所述通过高通量建库方法筛选第三间隔序列的序列碱基包括：分析酶切位点并设计简并引物，通过扩增获得含所述第三间隔序列的启动子库；利用所述含第三间隔序列的启动子库构建重组载体；以及将所述重组载体转化至宿主并筛选含所述第三间隔序列的目标启动子。

本发明的第二方面，提供一种全合成的酵母诱导型启动子，所述全合成的酵母诱导型启动子由所述全合成的酵母诱导型启动子的构建方法获得。

在一些实施例中，所述启动子中上游激活序列的数量为 1、2、3 或 4。

本发明的第三方面，提供一种重组载体，所述重组载体包括所述全合成的酵母诱导型启动子。

本发明的第四方面，提供一种重组菌，所述重组菌的基因组中包含所述重组载体。

本发明的第五方面，提供所述全合成的酵母诱导型启动子、所述重组载体或所述重组菌在调节酿酒酵母代谢中的应用。

与现有技术相比，本发明的有益效果是，

(1) 本发明提供的全合成的酵母诱导型启动子及其构建方法能够解决代谢工程和细胞线路设计中，由于启动子强度不足、种类少导致的特征良好启动子数量不足的问题。

(2) 本发明提供的全合成的酵母诱导型启动子及其构建方法可以避免序列间高度的相似性，进而解决酿酒酵母体内易发生同源重组的问题。

(3) 本发明提供的全合成的酵母诱导型启动子及其构建方法应用于诱导系统中，可以筛选出一系列本底表达低、诱导倍数高的合成型启动子，从而应用于基因线路设计。

(4) 本发明提供的全合成的酵母诱导型启动子构建方法可用于不同的诱导系统，提高了整体的适应性。

(5) 本发明提供的全合成的酵母诱导型启动子构建方法通过模型设计得到了一系列不同强度的间隔序列，相比于传统高通量建库和筛选方法，这种方法具有节约时间和成本的优势。

附图说明

图 1 为一种实施例中的酿酒酵母全合成的诱导型启动子设计的基本架构图；

图 2 为一种实施例中的酿酒酵母的木糖诱导型启动子的基本架构图；

图 3 为一种实施例中的标准测试载体示意图；

图 4 为一种实施例中的高通量测试载体示意图；

图 5 为一种实施例中的标准测试系统示意图；

图 6 为一种实施例中的不同间隔序列 sp.A 对启动子强度的影响图，其中图 6a 和 6b 分别表示启动子相对强度 RPU 情况和诱导倍数 Fold change 情况；

图 7 为一种实施例中的启动子模型设计的技术路线图；

图 8 为一种实施例中的不同间隔序列 sp.B 对启动子强度的影响图, 其中图 8a 和 8b 分别表示启动子相对强度 RPU 情况和诱导倍数 Fold change 情况;

图 9 为一种实施例中的间隔序列 sp.B 启动子强度的预测值和实测值比较图;

图 10 为一种实施例中的高通量测试系统示意图;

图 11 为一种实施例中的不同间隔序列 sp.C 对启动子强度的影响图, 其中图 11a-11c 分别表示间隔序列为 sp.CI、 sp.CII 和 sp.CIII 的情况;

图 12 为一种实施例中的不同间隔序列 sp.C 对启动子强度的影响图, 其中图 12a 和 12b 分别表示启动子相对强度 RPU 情况和诱导倍数 Fold change 情况;

图 13 为一种实施例中的不同序列对启动子强度的影响图, 其中图 13a 和 13b 分别表示启动子相对强度 RPU 情况和诱导倍数 Fold change 情况。

具体实施方式

下面将结合具体实施方式对本专利的技术方案作进一步详细地说明, 应该指出, 以下详细说明都是示例性的, 旨在对本申请提供进一步的说明。除非另有指明, 本文使用的所有技术和科学术语具有与本申请所属技术领域的普通技术人员通常理解的含义。

本申请中涉及的 5'→3'是指 DNA 或 RNA 分子的方向性。其中 5'端是指 DNA 或 RNA 的末端第五个碳原子与一个磷酸基团连接的位置, 3'端是指 DNA 或 RNA 的末端第三个碳原子上的氧原子, 这个氧原子通常与下一个核苷酸的磷酸基团连接在一起。因此, 5'端与 3'端的不同组成方式决定了 DNA 和 RNA 分子的方向性, 即从 5'端到 3'端是分子的末端到末端的方向。在 DNA 的复制、转录和翻译等过程中, 5'→3'方向更多地被用作描述 DNA 链的延伸方向和 RNA 链的合成方向。比如, 在 DNA 的复制过程中, DNA 聚合酶将以单链模板为基础, 从模

板的 3'端到 5'端进行读取，并沿着 3'→5'方向合成新的 DNA 链，即从 5'端到 3'端进行延伸。类似地，RNA 聚合酶在转录 RNA 分子时也是从 3'端到 5'端阅读单链 DNA 模板，合成 RNA 链的方向是 5'→3'。

启动子是指位于真核基因转录起始点附近的一系列 DNA 序列，可以与 RNA 聚合酶 II、转录因子以及其他调控因子结合，用于启动基因的转录。通常来讲，酵母启动子包括核小体和开放的染色质区域，其中核小体是一个直径约 10 纳米，在染色体上形成高次级结构的复合蛋白质分子，由两对酸性蛋白质和两对碱性蛋白质组成。核小体占据的比例会影响染色质的紧密程度和可访问性，低占用率的核小体有助于转录因子与调控 DNA 序列的结合，以调节启动子活性并促进转录。相对而言，高占用率的核小体会限制启动子区域的可访问性，从而影响基因表达。多聚腺核苷酸/多聚胸腺嘧啶（polyA/polyT）作为核小体不利序列，在真核生物中普遍存在，其长度一般在 10-20bp 之间。这些序列可以帮助维护染色质的稳定性，并在天然的酿酒酵母启动子中很常见，如 pTDH3、pHIS3、pADH2 等。

启动子染色质结构中的开放染色质区域通常包括上游元件序列（The Upstream Element）和核心启动子区域（The Core Promoter），这两个区域之间的距离可能是几百个碱基对或更多。其中，上游元件通过招募转录因子（TF）进行基因调控，当转录因子与上游元件序列中的转录因子结合位点（TFBS）相结合，启动子即可调控基因的表达。依据上游元件对基因表达强度的影响情况，可以将其分为上游激活序列（UAS）和上游抑制序列（URS），前者使基因表达增强，后者则减弱基因表达。核心启动子则独立于任何调控，直接与 RNA 聚合酶 II（pol-II）和通用转录因子相互作用形成起始前复合体（PIC），启动转录。作

为携带启动转录所需最小信息的区域，核心启动子一般包含 TATA 盒（TATA box）和转录起始位点（TSS）。分析这些启动子的序列结构，有助于了解它们的调控机制，这对于酿酒酵母的代谢工程和基因线路设计都非常重要。

本发明提供一种全合成的酵母诱导型启动子的构建方法，解决了代谢工程和细胞线路设计中，由于启动子本底泄露过高、诱导强度不足、种类少导致的特征良好启动子数量不足的问题。此外，该方法可以避免序列间高度的相似性，进而解决了酿酒酵母体内易发生同源重组的问题。

该全合成的酵母诱导型启动子的构建方法包括，确定所述启动子的基本架构，所述启动子从 5'→3'依次包括操作性连接的：上游激活序列、第一间隔序列、核小体不利序列、第二间隔序列、TATA 盒、第三间隔序列、转录起始位点；利用理性设计方法、模型建库方法和/或高通量建库方法筛选所述第一间隔序列、所述第二间隔序列和所述第三间隔序列的序列碱基；以及组合所述第一间隔序列、所述第二间隔序列和所述第三间隔序列，获得所述启动子。

本发明还提供一种全合成的酵母诱导型启动子，所述全合成的酵母诱导型启动子由所述全合成的酵母诱导型启动子的构建方法获得。

本发明还提供一种重组载体，所述重组载体包括所述全合成的酵母诱导型启动子。

本发明还提供一种重组菌，所述重组菌的基因组中所述的重组载体。

本发明还提供所述全合成的酵母诱导型启动子、所述重组载体或所述重组菌在调节酿酒酵母代谢中的应用。

下面通过具体实验实施例来对本发明作进一步的阐述。

实施例 1：酿酒酵母的木糖诱导型启动子 pxlnR 基本架构的确定

(1) 基本元件的确定。经文献调研确定了启动子上的上游激活序列 UAS、多聚腺核苷酸 polyA、TATA 盒 TATA box 以及转录起始位点 TSS。

(2) 酿酒酵母启动子基本架构的确定。如图 1 所示，上游激活序列 UAS 与多聚腺核苷酸 polyA 之间的最优距离为 24bp；多聚腺核苷酸 polyA 与 TATA 盒 TATA box 之间的最优距离为 62bp；TATA box 与转录起始位点 TSS 之间的序列来源于启动子 pADH2，此时本底强度最弱；转录起始位点 TSS 后的序列替换为 kozak 序列，有利于启动子诱导后的高表达，其中 kozak 序列的核苷酸排列为 5'-TAAATAAAAA-3' (SEQ ID NO:1)。

(3) 酿酒酵母的木糖诱导型启动子 pxlnR 基本架构的确定。如图 2 所示，在步骤 (2) 的基本架构上引入真核转录因子 xlnR 的诱导系统，即将木糖操纵子 xlnO 放在上游激活序列 UAS 处，使启动子由木糖诱导激活，其中 xlnO 序列的核苷酸排列为 5'-gaatttaggctaaagaaagatc-3' (SEQ ID NO:2)。所得酿酒酵母的木糖诱导型启动子 proxlnR 的序列如 SEQ ID NO:3 所示，作用于上游激活序列 xlnO 的转录因子 xlnR 的序列如 SEQ ID NO:4 所示。

需要理解的是，本实施例以木糖操纵子 xlnO 为例介绍本申请的具体实施方式，但在现有传感器和接收器一套的回路构件中，还可选择 lexO、RpaO 等与转录因子结合的序列作为上游激活序列 UAS，也就是说，本实施例除了可以应用于木糖诱导的 xlnR 系统外，还可应用与 β -雌二醇诱导的 Er 系统等其他诱导系统中。此外，上游激活序列 UAS 的个数也可改变，如设置 2、3 或 4 个。

实施例 2：标准测试载体 pXJH1 的构建

如图 3 所示，将黄色荧光蛋白 YFP 作为报告基因，tENO2 作为终止子，并在黄色荧光蛋白 YFP 前上放置用于筛选启动子的致死基因 CCDB，构建标准测试载体 pXJH1，从而建立启动子强度检测系统，其中标准测试载体 pXJH1 的序列如 SEQ ID NO:5 所示。其中，MAR4_CaURA3 表示尿嘧啶，用于筛选可以在缺尿嘧啶的酵母缺陷型培养基（SD-Δura）中存活的酵母单克隆；ori 表示原核生物基因质粒的复制起始位点，用于 DNA 的复制；Chl^R 表示氯霉素，用于筛选可以在氯霉素抗性板上存活的大肠单克隆。

具体来讲，如表 1 所示，以质粒 CY386 为模板，以 pXJH1-1-f/pXJH1-1-r 为引物对扩增，得到 1 号 PCR 片段；以质粒 PY114 为模板，以 pXJH1-2-f/pXJH1-2-r 为引物对扩增，得到 2 号 PCR 片段，即基因的表达框；以质粒 CY637 576int 为模板，以 pXJH1-3-f/pXJH1-3-r 为引物对扩增，得到 3 号 PCR 片段，即黄色荧光蛋白 YFP 和终止子 T25；以质粒 CY386 为模板，以 pXJH1-4-f/pXJH1-4-r 为引物对扩增，得到 4 号 PCR 片段，即致死基因 CCDB。而后利用 ClonExpress Ultra One Step Cloning Kit 试剂盒将四个 PCR 片段同源重组，得到重组载体 pXJH1，其中试剂盒购于诺维赞公司。

表 1 引物序列表

PCR 模板	引物名称	引物序列 (5'→3')	SEQ ID NO:
质粒 CY386	pXJH1-1-f	AACGCcgccatccagtgtcgagagtagagcacttgaat cca	6
	pXJH1-1-r	ttTTAGTAAGCCGAAGACCTAAAACCC CCTGGCGCTCCG	7
质粒 PY114	pXJH1-2-f	CACTAGTGCAGTGCAGTACAAGCTTG CCTTGTCCCCGCCGGGTC	8
	pXJH1-2-r	CAGTggattcaagtgtctactcTCGACACTGG ATGGCGGCGTTAGTAT	9
质粒 CY637 576int	pXJH1-3-f	GTTTctggggaatataagaagacctAATGTCTAA AGGTGAAGAATTAT	10
	pXJH1-3-r	CGGCGGGGACAAGGCAAGCTTGTAC	11

		TGCAGTGCCTAGTGG	
质粒 CY386	pXJH1-4-f	tCGGAGCGCCAGGGGGTTTTAGGTCT TCGGCTTACTAA	12
	pXJH1-4-r	attcttcaccttagacattAGGTCTTCTTATATTC CCCAGA	13

需要理解的是，本实施例以酿酒酵母内源终止子 tENO2 为例介绍本申请的具体实施方式，但还可选择其他来源的终止子，所选终止子的具体种类包括但不限于 TTPI1P、TFBA1、THXT7、TPGI1、TCYC1、TTEF1、TADH1、TGPM1。

实施例 3：高通量测试载体 pXJH200、pXJH201 和 pXJH203 的构建

(1) 间隔序列的划分。如图 2 所示，将所有特征序列之间的间隔序列划分并分别命名。其中，xlnR 与 polyA 之间的间隔序列为间隔序列 sp.A，polyA 与 TATA box 之间的间隔序列为间隔序列 sp.B，TATA box 与 TSS 之间的间隔序列为间隔序列 sp.C。此外，将长度为 90bp 的间隔序列 sp.C 平均分为三份，分别命名为 sp.CI、sp.CII 和 sp.CIII。

(2) 高通量测试载体 pXJH200 的构建。如图 4 所示，将黄色荧光蛋白 YFP 作为报告基因，tENO2 作为终止子，红色荧光蛋白 RFP 作为参考基因，并在原始启动子 pxlnR 的间隔序列 sp.CI 处放置致死基因 CCDB，构建间隔序列 sp.CI 的高通量测试载体 pXJH200，其核苷酸序列如 SEQ ID NO:14 所示。其中，致死基因 CCDB 用于筛选不同间隔序列 sp.CI；yfp 表示黄色荧光蛋白；MAR2_K1LEU2 表示亮氨酸，用于筛选可以在缺亮氨酸的酵母缺陷型培养基（SD-Δleu）中存活的酵母单克隆；mRuby2 表示红色荧光蛋白 RFP，用于检测整个基因表达框是否整合到酵母基因组上；ori 表示原核生物基因质粒的复制起始位点，

用于 DNA 的复制；Chl^R 表示氯霉素，用于筛选可以在氯霉素抗性板上存活的大肠单克隆。

具体来讲，以实施例 2 中成功构建的质粒 pXJH1 为模板，以 pXJH200-1-f/pXJH200-1-r 为引物对扩增，得到 1 号 PCR 片段；以质粒 pXJH1 为模板，以 pXJH200-2-f/pXJH200-2-r 为引物对扩增，得到 2 号 PCR 片段；以质粒 pXJH123 为模板，以 pXJH200-3-f/pXJH200-3-r 为引物对扩增，得到 7 号 PCR 片段。而后利用 ClonExpress Ultra One Step Cloning Kit 试剂盒将三个 PCR 片段同源重组转化，得到重组载体 pXJH200，其中试剂盒购于诺维赞公司。

(3) 高通量测试载体 pXJH201 的构建。与步骤 (2) 同理，构建筛选间隔序列 sp.CII 的高通量载体 pXJH201，其核苷酸序列如 SEQ ID NO:15 所示。

具体来讲，以实施例 2 中成功构建的质粒 pXJH1 为模板，以 pXJH200-1-f/pXJH201-1-r 为引物对扩增，得到 3 号 PCR 片段；以质粒 pXJH1 为模板，以 pXJH201-2-f/pXJH202-2-r 为引物对扩增，得到 4 号 PCR 片段；以质粒 pXJH123 为模板，以 pXJH201-3-f/pXJH200-3-r 为引物对扩增，得到 8 号 PCR 片段。而后利用 ClonExpress Ultra One Step Cloning Kit 试剂盒将三个 PCR 片段同源重组转化，得到重组载体 pXJH201，其中试剂盒购于诺维赞公司。

(4) 高通量测试载体 pXJH202 的构建。与步骤 (2) 同理，构建筛选间隔序列 sp.CIII 的高通量载体 pXJH202，其核苷酸序列如 SEQ ID NO:16 所示。

具体来讲，以实施例 2 中成功构建的质粒 pXJH1 为模板，以 pXJH200-1-f/pXJH202-1-r 为引物对扩增，得到 5 号 PCR 片段；以质粒 pXJH1 为模板，以 pXJH202-2-f/pXJH202-2-r 为引物对扩增，得到 6 号 PCR 片段；以质粒 pXJH123 为模板，以 pXJH202-3-f/pXJH200-3-r 为引物对扩增，得到 9 号 PCR 片段。而后

利用 ClonExpress Ultra One Step Cloning Kit 试剂盒将三个 PCR 片段同源重组转化，得到重组载体 pXJH202，其中试剂盒购于诺维赞公司。

另，步骤（2）-（4）中所涉及引物序列如表 2 所示。

表 2 引物序列表

PCR 模板	引物名	引物序列 (5'→3')	SEQ ID NO:
质粒 pXJH1	pXJH200-1-f	AGAGCACTTGAATCCACTGCCCC	17
	pXJH200-1-r	TATTTATATGTGGCAGTCTCCACCACCACT	18
	pXJH200-2-f	AGTGGTGGTGGAGACTGCCACATATAAATAA TGTCTTCGGCTTACTAAAAGCCAGATAACA	19
	pXJH200-2-r	AAGAAACAAGAAGTGATAAAAACAACAAG AGAGCAGTAGTAAGAGTATTTAAGTCTTCTT ATATTCCCCAGAACATCAGGTAA	20
	pXJH201-1-r	CGAGTGTGAAAAAAGTCGCTACTGGCA	21
	pXJH201-2-f	TGCCAGTAGCGACTTTTTTCACACTCGATGT CTTCGGCTTACTAAAAGCCAGATAACA	22
	pXJH202-2-r	ATTCTATTTACCAAGAAGAAACAAGAAGTGA TAAAAAGTCTTCTTATATTCCCCAGAACATC AGGTAA	23
	pXJH202-1-r	AACAACAAGAGAGCAGTAGTAAGAGTATTT CGAGT	24
	pXJH202-2-f	ACTCGAAATACTCTTACTACTGCTCTCTTGTT GTTATGTCTTCGGCTTACTAAAAGCCAGATA ACA	25
	pXJH202-2-r	AAGTCTTCTTATATTCCCCAGAACATCAGGT AA	26
质粒 pXJH1 23	pXJH200-3-f	GCTCTCTTGTTGTTTTTATCACTTCTTGTTTC TTCTTGGTAAATAGAATATCAAGCTACAAAA ATAAATAAAAAAATGTCTAAAG	27
	pXJH200-3-r	ggggcagtggattcaagtgtctt	28
	pXJH201-3-f	CTTTTATCACTTCTTGTTTCTTCTTGGTAAA TAGAATATCAAGCTACAAAAATAAATAAAAA AATGTCTAAAG	29
	pXJH202-3-f	TTAACCTGATGTTCTGGGGAATATAAGAAGA CTTAGAATATCAAGCTACAAAAATAAATAAA AAAATGTCTAAA	30

实施例 4：间隔序列 sp.A 的理性设计及启动子强度测试

(1) 间隔序列 sp.A 的理性设计。随机生成和原始间隔序列 sp.A 等长的 24 bp 的序列并进行筛选，得到通过理性设计方法设计的一系列间隔序列 sp.A mutants，从而验证不同的间隔序列 sp.A 序列对启动子强度的影响。

具体来讲，考虑到转录因子结合位点 xlnO 和 polyA 之间的间隔序列 sp.A 不属于核心启动子区域，认为其序列的特异性对启动子影响不大，故设定如下筛选条件：其一，GC%含量不宜过高，应保持在 25-50%之间；其二，序列中的连续碱基 A 的长度不超过 5 个。在 random sequence 网页上生成并筛选后的 sp.A 序列如表 3 所示，其中原始序列为 sp.A0。

表 3 间隔序列 sp.A 理性设计表

编号	序列碱基	SEQ ID NO:
sp.A0	AATAAGAGAATTTTCAGATTGAGAG	31
sp.A6	CACCTTCTATTGGATGGGAATGAT	32
sp.A7	AACATATGAGGAAAGGCCCCAAAA	33
sp.A8	TCCAAAGTCACAAGTCCCAAAGAA	34
sp.A9	CAAGGTAATTGAACGAGCATGAAC	35
sp.A10	ATAATAGGCATCGAATTA ACTATT	36
sp.A11	AAAGTTCCTCGAAGGATAGCAGAA	37
sp.A12	AACTAATGGCTACAATGAGTCTAG	38
sp.A13	ATGGTTGTGTGAAGGTTGATCCAA	39
sp.A14	AATCAACCACTACACTACAAGAAT	40
sp.A15	GAGTTATACTGAGAGCATATGATA	41
sp.A16	GTCATGAGTATATGATGAGTGATA	42
sp.A17	AGGATTATGAGAAGAACCTAGTGG	43
sp.A18	ATAATAGGGCTACAATACAAGAAT	44
sp.A19	ATAATAGGACTACACTGAGTCTAG	45
sp.A20	AACTAATGCATCGAATACAAGAAT	46
sp.A21	AACTAATGACTACACTTA ACTATT	47
sp.A22	AATCAACCGCTACAATTA ACTATT	48
sp.A23	AATCAACCCATCGAATGAGTCTAG	49

sp.A24	ATAATAGGTTGAACGAACAAGAAT	50
sp.A25	ATAATAGGGCTACAATGCATGAAC	51

(2) 含不同 sp.A 序列的启动子的构建。设计含 sp.A 序列的长引物进行无模板 PCR 扩增，得到一系列不同的含 sp.A 序列的 PCR 片段，并将其作为第一个 PCR 片段。以质粒 pXJH120 为模板，设计引物对扩增，得到第二个 PCR 片段。

通过 overlap PCR 融合第一个 PCR 片段和第二个 PCR 片段，得到含 sp.A 序列的完整的启动子序列，并利用 golden gate 技术，将实施例 2 中成功构建的载体 pXJH1 上的致死基因 CCDB 替换成上述完整的启动子序列，其中 golden gate 体系及条件如表 4 所示。最后通过大肠转化、测序，得到测序正确且含不同 sp.A 序列启动子的质粒 pXJH318-pXJH337。

表 4 golden gate 体系及条件表

试剂	用量 (μL)	反应条件	
BpiI	0.5	37°C	2 mins
T4 ligase	0.5	16°C	3 mins
T4 ligase buffer	1	Go to 1 for 24 times more	
载体	a (50 ng)	50°C	15 mins
片段	b (50 ng)	80°C	20 mins
ddH ₂ O	8-a-b	/	/
UP to	10	/	/

(3) 标准测试系统的构建。如图 5 所示，用限制性内切酶 BsaI 酶切步骤 (2) 所构建的测试质粒 pXJH318 到 pXJH337，得到带转化的酶切片段，这些片段包含 ura3 同源臂以及由步骤 (2) 中通过 overlap PCR 融合得到的完整测试启动子、黄色荧光蛋白 YFP、终止子 tENO1 构成的 YFP 表达框。其中，xyl 表示木糖，用作诱导剂；xlnR 表示作用于上游激活序列 xlnO 的转录因子；P_{test} 表示测试的启动子；aTc 表示脱水四环素盐酸盐，用作诱导剂；P_{tet} 表示启动子，添加诱导剂 aTc 后，P_{tet} 启动子开始作用，下游基因开始表达；TetR 表示在背景菌株中常

表达的四环素抑制因子，其作用于 P_{tet} 启动子，阻止下游基因的表达；ChXV(his3::)表示整合至 ChXV 染色体的组氨酸位点；IIA.ChV(ura3::)表示整合至 IIA.ChV 染色体的组氨酸位点；yfp 表示黄色荧光蛋白。

而后利用 zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株感受态 sXJH110，具体来讲，将上述酶切纯化后得到的片段整合至酿酒酵母染色体 IIA.ChV 的 ura3 位点处，并将含 xlnR 表达框的片段整合至酿酒酵母染色体 ChXV 的 His3 位点处，得到酵母转化子。

将酵母转化子接种于含 500 μ L 的对应缺陷型培养基的 96 深孔板中，并置于 30°C、800rpm 的培养箱中培养。培养 24-48h 后，以 1:200 的比例转接至新的缺陷型培养基的 96 深孔板中培养，并将工作浓度为 100ng/mL 的脱水四环素 atc 和工作浓度为 10mM 的木糖 xyl 作为诱导剂添加至其中。作为对照，每次试验还应添加酵母菌株 CYE72、CYE72/CY671 和 CYE72/CY637，其中转接酵母菌株 CYE72/CY637 需添加工作浓度为 100ng/mL 的脱水四环素 atc。培养菌株 16h 后，用适当量的 1xPBS 缓冲液稀释菌液，得到待检测的测试菌液。

(4) 不同间隔序列 sp.A 对启动子强度的影响测试。利用流式分析仪器 BD FACSCelesta™ 的 HTS 模式检测菌液中黄色荧光蛋白 YFP 的表达量，其中每个样品检测 10,000 个细胞的荧光，黄色荧光蛋白 YFP 用 FITC-A 通道进行测量。所得数据通过 FlowJo 进行处理，得到荧光的中位数 Flu（单位为 au）。其中，相对启动子强度 $RPU = (YFP_{\text{测试}} - YFP_{\text{CYE72}}) \div (YFP_{\text{CYE72/cy671}} - YFP_{\text{CYE72}})$ ，诱导倍数 $\text{Fold change} = (YFP_{\text{诱导后}} - YFP_{\text{CYE72}}) \div (YFP_{\text{本底}} - YFP_{\text{CYE72}})$ 。

结果：图 6 展示了含不同间隔序列 sp.A 的启动子诱导后的相对强度和相应的诱导倍数情况。其中，图 6a 的纵坐标表示通过黄色荧光蛋白 YFP 表达量数据

处理后得到的相对启动子强度 RPU，横坐标表示启动子所含的具体间隔序列 sp.A，横坐标所示负号对应的灰色柱表示启动子本底表达情况，横坐标所示正号对应的黑色柱表示启动子诱导后的表达情况；图 6b 的纵坐标表示诱导倍数 Fold change，横坐标表示启动子所含的具体间隔序列 sp.A。

可以看出，理性设计得到的不同间隔序列 sp.A 对诱导型启动子的强度有所影响。一般情况下，启动子本底的强度与诱导后启动子的强度在一定程度上呈正相关，比如，含 sp.A7、sp.A8、sp.A15 等间隔序列的诱导后启动子的强度较强，相应地其启动子本底表达较高；启动子本底强度降低，诱导后的启动子强度也随之下降。此外，含 sp.A9、sp.A10、sp.A12 等间隔序列的启动子本底比对照组 sp.A0 更低，诱导倍数变化更大，更适用于细胞线路设计。

实施例 5：间隔序列 sp.B 的计算模型设计及启动子强度测试

本实施例通过改进现有酵母启动子模型，开发了一种基于目标输出生成序列的方法。如图 7 所示，在现有酵母启动子模型的基础上，结合湿实验数据进行优化以计算可诱导酵母启动子的强度，并进行模型准确性验证。随机生成序列，并使用上述计算模型预测启动子强度，从而筛选目的序列，即通过模型设计得到了一系列不同强度的间隔序列 sp.B。相比于传统高通量建库和筛选方法，这种方法具有节约时间和成本的优势。

(1) 计算模型的建立。

首先进行算法分析。假设启动子中某一转录因子 x 所有可能的结合位点总和为 BS_{xsi} 。其中， i 表示位置， s 表示链， C_x 表示固定转录因子浓度， K_{dxsi} 表示最小结合平衡常数，该平衡常数通过计算转录因子 TF 基序的位置权重矩阵 PW

M 得到。

$$BS_{xsi} = \frac{1}{1 + \frac{K_{dxi}}{C_x}}$$

计算该酵母启动子上所有转录因子结合总和 DB_x ，即与所有结合位点结合的总和。

$$DB_x = \sum_{i,s} BS_{xsi}$$

预测启动子的可及性。其中， $DB \times P$ 表示没有染色质情况下 DNA 结合转录因子 TF 的加权估计值， C_p 和 e 为常数，不同染色质背景下的 Ω 值不同。

$$\Omega = \frac{1}{1 + e^{-(DB \times P + C_p)}}$$

考虑到核小体会阻止转录因子 TF 的结合，所以用 Ω 对不同染色质背景下转录因子 TF 结合启动子的预估值进行缩放。

$$CB = DB \times \Omega$$

预测表达水平 EL 为预估值对学习到的影响的加权值加上一个常数，预测的表达水平在在线性空间为乘法，在对数空间为加法。

$$EL = CB \times A + c_e$$

其次，进行模型训练。使用介质为葡萄糖的数据集训练模型，数据集主要涵盖了以下内容：一是酵母菌细胞在葡萄糖介质下的转录组数据，包括基因编号和基因的表达量数据；二是酵母菌在不同浓度的葡萄糖介质下的生长曲线数据，包括时间、不同浓度的葡萄糖介质下的吸光度（Optical Density, OD 值）数据。通过 tensorflow 框架实现转录模型，并使用 AdamOptimizer 最小化预测和测量表达水平直接的均方误差。此外，分批次学习 1024 个启动子序列，先学习

转录因子 TF 的活性和增效，再学习转录因子 TF 的浓度，之后学习特定位置的活性参数。模型训练的序列长度为 110bp，实际输入序列为 80bp，超过 80bp 的序列删去超出部分，不足 80bp 的序列用随机序列补齐。在模型训练阶段进一步筛选数据集，构建更加贴合 sp.B 特点的训练集，可以提升模型针对 sp.B 序列特点的学习能力。此外，对模型的部分参数进行调整修改，如学习率和训练批次等，可以提升模型预测准确性。

最后，进行模型优化。序列排序以及组成会显著影响模型的预测准确率，一般来讲，转录因子 TF 的结合在 Azf1、Mgal、Mot3、Skn7、Ume6 以及 poly-A 基序上具有明显的位置偏向性。使用 pTpA 启动子支架模型作为基础模型进行模型训练时，通常会在输入序列的上游和下游分别补充 17bp 和 13bp 的侧翼序列，使得实际送入模型的序列长度为 110bp，其中 17bp 的左侧翼序列为 TGC ATTTTTTTCACATC (SEQ ID NO:52)，13bp 的右侧翼序列为 GGTTACGGCT GTT (SEQ ID NO:53)。参考模型训练数据集的特征，发现数据集为 TATA—box 上游序列，然后结合课题组实际使用的启动子架构，重点分析 TATA-box 上游序列，细化基序组成部分，对待预测的序列组成进行平铺排列组合，分别预估其转录因子 TF 结合情况，筛选得到 sp.B+中性序列以及 poly-A+spacer1 的最佳组合。此外，模型用已发表的其他酵母启动子数据进行验证。

在进一步的模型优化阶段，为研究序列长度和顺序，分析基序的位置偏向性，本方案以不同长度序列摆放在不同位置得到的组合作为输入，筛选预测效果最好的、最贴合模型算法的位置排序以及对应的序列长度，提升了模型的预测准确率。而后，利用训练好的模型对随机生成的序列输出进行预测并排序，可以减少高通量实验的人力物力财力以及节约时间成本。

(2) 间隔序列 sp.B 的计算模型设计。将随机生成的指定长度的目标序列放入步骤 (1) 得到的计算模型中进行预测, 得到一系列低本底、高诱导的理想序列, 并筛选结果最佳的前 n 个序列进行后续试验验证。本实施例经筛选得到的 sp.B 序列以及模型计算的输出结果如表 5 所示。

表 5 间隔序列 sp.B 计算模型设计表

编号	SEQ ID NO:	序列碱基	预测表达(输出)
sp.B0	54	GGCAGAGGAGAGCATAGAAATGGGGTTCACT TTTTGGTAAAGCTATAGCATGCCTATCACA	4.3007493
sp.B1	55	GGCAGTCTTGGACAATAGTTATATCACGTTGA CTCAGGAAGTGGTGGTGGAGACTGCCACA	3.7348356
sp.B2	56	GGCAGCGCGTTTCTTTCCCTGCCCTAGCATGA AGACCTAGCCTTTGTATCTGTCGTACACA	11.07776
sp.B3	57	GGCAGCTTTCGATCCGAGTGCCTCAACTGTT CACCCCTTGCTATAACACCCAGTCTACACA	10.589281
sp.B4	58	TGCGTACCCTTATACCCTTCTTTGCATCTGTCC AATCCTTAGTGAACAATGATGGGTAAGC	3.4995785
sp.B5	59	TCGTTGCCAGAAGCTCTGTGCAATTAATGAA GGACTCATGTCACCTACACCATCGGGGATGA	3.439314
sp.B6	60	GGGACCTTTAGCAAACGTCGGTTCCGCAGAG AATGACGTCAAGACGATAGCGCCGCGGGAC	9.397211
sp.B7	61	GTGTCACCTAGATGAAGAGATGCATACCCTTAT ACAGAACCCTCATAGTTGTAAGTTGCACA	3.525889
sp.B8	62	AGCATAGTGTCAATTACACTGATACACTAGGT TTTAACCTACCTTCACCACTTGTAAGGA	3.6678185
sp.B9	63	CTTCGGCTGGGATACTCCATGCTCGATTATCA AATCTGCCTCCACACTTGTTTCAGTTGGGC	4.4174633
sp.B10	64	AAATGTTAACATGGCGCCCAGTATCTACCGCA CGCGGCTGCTTGGTTTAGCCGCTTGAGAAT	8.120921
sp.B11	65	AGACTACTGGTAGGCTAGCTGTTAGTACTCCT CCTACTGCCAACTACTATGACAGTGTTTC	3.6032057
sp.B12	66	GGCAATCCCCACAATGGGTCCCTGAAGGCGT TTGACGAAGTTGTCTGCAGAGCTGTTTCACA	5.317227

sp.B13	67	GGCAAGTAACAAGTAGTTTGTTCGATCAGTC CCTGGTTCGCAAGCAAGTGGAGGGCGATCAC A	4.2635784
sp.B14	68	GGCAATCTGCTACAGGAAGTTATTACGGCAA CGGCCTGAAGCTCCGATGAGCACTAATCACA	5.9438214
sp.B15	69	GGCAAAATGTAGTCCGCTTCTTGGCGAAAGG TACTGCACTTCAACCATTGCACAAGATCACA	10.232451
sp.B16	70	GGCAAATGCTCATCGGGTCCGTCAGGCCATG TGGAGTACGCAGGCCTCTCGGGCAGATTCAC A	12.278882
sp.B17	71	GGCAATCGGCAAGAAAGATTTCTATCACTAAT GCTCAAGAGAAGCTTCCACGGCATGTCACA	5.8025684
sp.B23	72	GGCAGTCTTGGACAATAGTGCAATTAATGAAT ACAGAACCCTCATCACCACTTGTAACACA	3.5735002
sp.B24	73	GGCAGGCCAGAAGCTCTGATGCATACCCTTA TTTTAACCTACCTTGGTGGAGACTGCCACA	4.60665
sp.B25	74	GGCAGACTAGATGAAGAGCTGATACACTAGG ACTCAGGAAGTGGTACACCATCGGGGCACA	3.467486
sp.B26	75	GGCAGAGTGTCAATTACATTATATCACGTTGG GACTCATGTCACTAGTTGTAAGTTGCACA	3.6589136
sp.B27	76	GGCAGTCTTGGACAATAGCTGATACACTAGG TACAGAACCCTCATACACCATCGGGGCACA	3.816956
sp.B28	77	GGCAGACTAGATGAAGAGTTATATCACGTTGT TTTAACCTACCTTACACCATCGGGGCACA	3.9463234
sp.B29	78	GGCAGAGTGTCAATTACAATGCATACCCTTAG GACTCATGTCACTGGTGGAGACTGCCACA	3.5313206

(3) 含不同 sp.B 序列的启动子的构建。利用 PCR 技术，将 pxlnR 上的原始序列 sp.B0 替换为表 6 所示的一系列 sp.B 间隔序列，得到含 sp.B 序列的完整的启动子序列。

具体来讲，先将启动子分为 sp.B 上游、sp.B 和 sp.B 下游，并以 pxlnR 为模板，设计不同的引物对扩增，得到 sp.B 上游片段和 sp.B 下游片段；设计含 sp.B 序列的长引物进行无模板 PCR 扩增，得到 sp.B 片段；而后以 sp.B 上游、sp.B 和 sp.B 下游为模板，用引物对扩增，得到完整的含 sp.B 序列的启动子。

利用 golden gate 技术, 将实施例 2 中成功构建的载体 pXJH1 上的致死基因 CCDB 替换成上述完整的启动子序列, 其中 golden gate 体系及条件与实施例 4 步骤 (2) 中表 5 所示内容类似。最后通过大肠转化、测序, 得到测序正确且含不同 sp.B 序列启动子的质粒。

(4) 标准测试系统的构建。与实施例 4 的步骤 (3) 类似, 通过本实施例步骤 (3) 所构建的 24 个测试质粒得到带转化的酶切片段, 后经转化得到酵母转化子并进行培养, 获得待检测的测试菌液。

(5) 不同间隔序列 sp.B 对启动子强度的影响测试。与实施例 4 的步骤 (4) 类似, 通过流式分析检测本实施例步骤 (4) 所得菌液的荧光蛋白表达量, 并进行数据处理。

结果: 图 8 展示了含不同间隔序列 sp.B 的启动子诱导后的相对强度和相应的诱导倍数情况。其中, 图 8a 的纵坐标表示通过黄色荧光蛋白 YFP 表达量数据处理后得到的相对启动子强度 RPU, 横坐标表示启动子所含的具体间隔序列 sp.B, 横坐标所示负号对应的灰色柱表示启动子本底表达情况, 横坐标所示正号对应的黑色柱表示启动子诱导后的表达情况; 图 8b 的纵坐标表示诱导倍数 Fold change, 横坐标表示启动子所含的具体间隔序列 sp.B。

可以看出, 计算模型设计得到的不同间隔序列 sp.B 对诱导型启动子的强度的影响不同。sp.B1、sp.B5、sp.B8、sp.B23 等序列的本底表达低于原始序列 sp.B0, 同时诱导倍数也远高于原始启动子, 因此通过本实施例筛选得到的序列的性能优于原始启动子 pxlnR。此外, sp.B1、sp.B5、sp.B8、sp.B23 等序列本底的相对强度低于 0.01, 更适用于细胞线路设计。

(6) 模型设计的匹配度检验。通过线性回归分析确定实际测试数据与计算

机模型预测结果的相关性。

结果：图 9 展示了实际测试数据与计算机模型预测结果的匹配情况，其中横坐标表示实际测试的相对启动子强度 RPU，纵坐标表示相对启动子强度 RPU 的预测值。可以看出，实测值与预测值的 R 值达到了 0.821，这说明线性拟合性较好，模型预测较为准确，因此，利用本实施例所建立的计算模型预测 sp.B 序列是可行的。

实施例 6：间隔序列 sp.C 的高通量建库设计及启动子强度测试

(1) 间隔序列 sp.C 的高通量建库设计。将间隔序列 sp.C 等分为 sp.CI、sp.CII 和 sp.CIII，并利用随机引物设计和 golden gate 技术获得一系列不同的上述三种间隔序列的高通量载体库，后续将放置在高通量测试系统中测试。

具体来讲，根据核心启动子区域的序列特征，设计三对兼并引物，用于 sp.CI、sp.CII 和 sp.CIII 间隔序列建库：以 primer-pXJH200-spacer2a-f/XJH-HT-1-r 引物对扩增得到 spcI 的随机片段 1。以 primer-pXJH200-spacer2b-f/XJH-HT-1-r 引物对扩增得到 spcI 的随机片段 2。以 primer-pXJH200-spacer2c-f/XJH-HT-1-r 引物对扩增得到 spcI 的随机片段 3。其中，尽量选择随机碱基 W/Y 进行引物设计，且间隔序列 sp.CIII 的引物设计尽量减少 A 碱基的含量，以避免提前转录。

引物序列如下：

primer-pXJH200-spacer2a-f: 5'-GACCCTGAAGACAAAATANSNSYSYWSS
NSYYNWYYWYNWYWNNYYNWAAATATGTCTTCCATGACAGAGTGCCTG
GTGATGTTAATGGTCACAA-3' (SEQ ID NO:79) ;

primer-pXJH200-spacer2b-f: 5'-GACCCTGAAGACAACCTCGWWYYWYY

NNWNWYYNNNNWNNYYNWNWYWWTTTAATGTCTTCCATGACAGAGTGC
CTGGTGATGTTAATGGTCACAA-3' (SEQ ID NO:80) ;

primer-pXJH200-spacer2c-f: 5'-GACCCTGAAGACAATGTTWWYYWWYY
NNWWWYYNNNNYYWWNNWWYWWAGAAATGTCTTCCATGACAGAGTG
CCTGGTGATGTTAATGGTCACAA-3' (SEQ ID NO:81) ;

XJH-HT-1-r: 5'-TTGTGACCATTAACATCACCAGGCACT-3' (SEQ ID NO:
82) 。

利用优化后的 golden gate 技术，将实施例 3 中成功构建的载体 pXJH200 上的致死基因 CCDB 替换成随机片段 1，得到反应产物，其中 golden gate 体系及条件如表 6 所示。上述反应产物经常规大肠转化，转到感受态 trans10 中，得到大肠转化子。随后，把相应抗性的培养基滴在转化板上，即向平板上滴加含氯霉素的 LB 培养基，并用涂布棒将大肠转化子全部挂到培养基中。最后进行常规质粒提取步骤，得到一个带有不同间隔序列 sp.CI 的新的启动子质粒库 library. s p.CI muts。

表 6 golden gate 体系及条件表

试剂	用量 (μL)	反应条件	
BpiI	2.5	37°C	5 mins
T4 ligase	2.5	16°C	5 mins
T4 ligase buffer	5	Go to1 for 49 times more	
载体	a (300 ng)	50°C	15 mins
片段	b (1500 ng)	80°C	20 mins
ddH ₂ O	40-a-b	/	/
UP to	50	/	/

同理，将实施例 3 中成功构建的载体 pXJH201、pXJH202 上的致死基因 CCDB 分别替换成随机片段 2、随机片段 3，经转化、质粒提取后，得到带有不同

间隔序列 sp.CII 的新的启动子质粒库 library. sp.CII muts 和不同间隔序列 sp.CII I 的新的启动子质粒库 library. sp.CIII muts。

(2) 高通量测试系统的构建。如图 10 所示，用限制性内切酶 BsaI 酶切步骤 (1) 所构建的质粒库 library. sp.CI muts，得到带转化的酶切片段，这些片段包含 leu2 同源臂两端的整个表达盒。其中，xyl 表示木糖，用作诱导剂；xlnR 表示作用于上游激活序列 xlnO 的转录因子；P_{test} 表示测试的启动子；aTc 表示脱水四环素盐酸盐，用作诱导剂；P_{tet} 表示启动子，添加诱导剂 aTc 后，P_{tet} 启动子开始作用，下游基因开始表达；TetR 表示在背景菌株中常表达的四环素抑制因子，其作用于 P_{tet} 启动子，阻止下游基因的表达；ChXV(his3::) 表示整合至 ChX V 染色体的组氨酸位点；IIB.ChIII(leu2::) 表示整合至 IIB.ChIII 染色体的亮氨酸位点；YFP 表示黄色荧光蛋白；MAR2_K1LEU2 表示亮氨酸，用于筛选可以在缺亮氨酸的酵母缺陷型培养基 (SD-Δleu) 中存活的酵母单克隆；RFP 表示红色荧光蛋白；RPU standard cassette 表示表达红色荧光蛋白 RFP 的标准表达盒，用于检测整个表达框是否整入亮氨酸位点。

而后利用 zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株感受态 sXJH110，具体来讲，菌株 sXJH110 中含有整合在 His3 位点处的含 xlnR 表达框，将上述酶切纯化后得到的片段整合至酿酒酵母染色体 IIB.Ch III 的 leu2 位点处，即可得到酵母转化子。

从转化板上随机挑选 96 个酵母转化子，并将其接种于含 500μL 的缺组氨酸和亮氨酸的酵母缺陷型培养基 (SD-ΔHis-ΔLeu 培养基) 的 96 深孔板中，置于 30°C、800rpm 的培养箱中培养。培养 24-48h 后，以 1:200 的比例转接至新的缺陷型培养基的 96 深孔板中培养，并将工作浓度为 100ng/mL 的脱水四环素 atc

和工作浓度为 10mM 的木糖 xyl 作为诱导剂添加至其中。作为对照，每次试验还应添加酵母菌株 CYE72、CYE72/CY671 和 CYE72/CY637，其中转接酵母菌株 CYE72/CY637 需添加工作浓度为 100ng/mL 的脱水四环素 atc。培养菌株 16h 后，用适当量的 1xPBS 缓冲液稀释菌液，得到间隔序列 sp.CI 的高通量测试菌株。

同理，分别酶切步骤（1）所构建的质粒库 library. sp.CII muts、library. sp.CIII muts，并进行转化和常规流式诱导培养，得到间隔序列 sp.CII、sp.CIII 的高通量测试菌株。

（3）不同质粒库中启动子强度的高通量测试。利用流式分析仪器 BD FAC SCelesta™ 的 HTS 模式检测菌液中黄色荧光蛋白 YFP 和红色荧光蛋白 RFP 的表达量，其中每个样品检测 10,000 个细胞的荧光，黄色荧光蛋白 YFP 和红色荧光蛋白 RFP 分别用 FITC-A 和 PE Texas-A 通道进行测量。选取同时表达红色荧光蛋白 RFP 和黄色荧光蛋白 YFP 的菌株数据进行处理，以了解红色荧光蛋白 RFP 恒定表达时黄色荧光蛋白 YFP 的表达情况。

结果：图 11 展示了不同间隔序列对启动子强度的影响，其中图 11a-11c 分别表示间隔序列 sp.CI、sp.CII 和 sp.CIII 的情况，纵坐标均为诱导强度，即诱导后的黄色荧光蛋白 YFP 的表达量，横坐标为转化板上挑选的不同的酵母菌株，虚线则表示对照菌株的情况，即启动子为 pxlnR 时的诱导强度。

可以看出，在同一质粒库中添加诱导剂后，不同酵母菌株的黄色荧光蛋白 YFP 的表达量不同。其中，库 library. sp.CI muts 和库 library. sp.CII muts 中菌株的黄色荧光蛋白 YFP 表达强度均集中在 4k-6.7k a.u 之间，其中库 library. sp.CI muts 中比对照菌株诱导后强度更高的菌株有 7 个，库 library. sp.CII muts 中比对照菌株诱导后强度更高的菌株有 14 个。与库 library. sp.CI muts 和库 library.

sp.CII muts 不同的是, 库 library. sp.CIII muts 中的菌株诱导后强度呈两极分化: 一部分菌株黄色荧光蛋白 YFP 的表达强度集中在 3.6k-6.6k a.u, 另一部分则低于 200a.u, 这种大幅度的启动子强度变化可能与转录起始位点的改变有关。此外, 库 library. sp.CIII muts 中比对照菌株诱导后强度更高的菌株有 16 个。

挑选三个库中诱导后黄色荧光蛋白 YFP 表达量高的 15 株菌株进行本底表达测试, 发现这三个库中酵母菌株黄色荧光蛋白 YFP 的本底表达量均集中在 80-150 a.u 之间, 略高于对照菌株黄色荧光蛋白 YFP 的本底表达量。

与此同时, 以上述菌株的基因组作为模板, seq1-f/sequence-r 为引物对扩增, 得到片段后进行测序, 结果如表 7 所示。引物序列如下:

seq1-f: 5'-CATTTTCCTCTACGAGCTGTCACCG-3' (SEQ ID NO:83);

sequence-r: 5'-GGCCATGGAACTGGCAATTTACCA-3' (SEQ ID NO:84)。

表 7 不同间隔序列 sp.CI muts、sp.CII muts 和 sp.CIII muts 的碱基序列表

序列名称	序列碱基	SEQ ID NO:
sp.CI mut1	TGGGTGTTGCTGCTGATCTCTTCACTTTCT	85
sp.CI mut2	GGCCCGCTGGACTCCACTTTTACAATTCTGA	86
sp.CI mut3	GAGTGCCAGCAGTGGCTTTTTTCTCGTTTCG	87
sp.CI mut4	GGGGCGCTCGACCCGATTATTTCTGCTTTA	88
sp.CI mut5	GGAGCGTTCGAGTCGTCTTTAATAGATCAT	89
sp.CI mut6	GGGGCGCTGGTGTTCATCCACTACATTTTCGT	90
sp.CI mut7	GCCCTGTACGTGCCCACCATTTCTCGCCCA	91
sp.CI mut8	CCGCTGCTGGTCCCGTCTACTTTACATCTT	92
sp.CI mut9	GGGGTGTTGCCGCTCTTTATCTCTCGCTGA	93
sp.CI mut10	CCGTGCCCAGGTGCTTTTTTCCTCAAGCCTA	94
sp.CI mut11	GGGCCGTTGGGGCTGACTACTTCAATTCTGA	95
sp.CI mut12	GGCGTGTAAGGACTCATCCTCTACTTGTTTT	96
sp.CI mut13	CCACCGTTGCTCTCCTTCTTCTCACGTCGT	97
sp.CI mut14	GGGGTGCTGGTGTCTCTCCTTGACTGTCTCA	98
sp.CI mut15	GGGGTCCTGCGGTTCTTCATCTTTGCTCTT	99
sp.CII mut1	AATTTTTTGGTCTTTTAGTGGTTGTCTTAA	100
sp.CII mut2	ATACCTTTGCTCTTTTCCTATTCGACTTTT	101

sp.CII mut3	TTATCATCTGTATCCTGATGCCCCAATCAT	102
sp.CII mut4	ATTCTTCCATAAATCTTGTGATCGTCATTT	103
sp.CII mut5	TCGTTTCCTCTTGTAATCCGCTTTTTGTCTCTT	104
sp.CII mut6	AATCCACCTTAATTCAGATATCTCTGTCAT	105
sp.CII mut7	TTTTTACTCCTTACTCGTAGGTCGTGATTA	106
sp.CII mut8	ATACCTTCCAAAATCCGAAAGTCCACATTA	107
sp.CII mut9	TTTCTTTTAGATTCTGGTTCTTTGATATTT	108
sp.CII mut10	AAATCATCTATGACCTAATAATTTAGTTTT	109
sp.CII mut11	GTTTTCTTTTGTGTTTCGCTGACTGTGTCTT	110
sp.CII mut12	TATCTTCCAATTTCTTCGTCGCTCTTTCTA	111
sp.CII mut13	TTACTTCCTATCTTTGCGTTGTTTACTCTA	112
sp.CII mut14	TATCTTTTCTAGTCCGGCTACTCAATACTT	113
sp.CII mut15	AATTCACTTCTGTTTACTTGGCCCTCTTTT	114
sp.CIII mut1	TTCTTTTCCGTTTTTAAAACTTAAGTATTT	115
sp.CIII mut2	TTCTTACCGGAATTTAAGCTTTTCTTTTTT	116
sp.CIII mut3	TTCCATCCAAATTTTGCTTCTATTTTTTCAT	117
sp.CIII mut4	ATCCATTTTGAATTTACAGTCATCCAACTT	118
sp.CIII mut5	ATTCTACCTAATTCCTACTCATACAATA	119
sp.CIII mut6	TTCCTTTCTCTAACTGGGGCCAACATCAT	120
sp.CIII mut7	AATCAACTCATTACCTATCCCTAAGTATAA	121
sp.CIII mut8	AATCTATTCCTAATTGGGGCTATCGAATAT	122
sp.CIII mut9	TTCTTTTTTGGTATTCTGAGTCAAGCTTCTT	123
sp.CIII mut10	AACCAATCCATTATCAGAACTTAGAAATTA	124
sp.CIII mut11	TTCCTACTGGTTATCAAGTTTAAGGATCAT	125
sp.CIII mut12	TTCTTTTTTATTTACTATGACTAAGGTATTA	126
sp.CIII mut13	TTCTTATCTTTTATTGGCTTCTTGTAACCTT	127
sp.CIII mut14	TATCTTCCAGATTCCGCGATCTAAAAATTA	128
sp.CIII mut15	TATCATTTAGTAACTGCATCTTAACAATTT	129
sp.CIII mut16	TTCTATTCGTAACTGGCTTCTTGCTTCAA	130

(4) 含不同 sp.C 序列的启动子的构建。选取 library. sp.CI muts、library. sp.CII muts 以及 library. sp.CIII muts 的菌株，并将其序列组合在一起，构建新间隔序列的 sp.C muts。构建方式既可以是，从三个库中选取本底低但诱导后表达量高的菌株进行组合，也可以是，随机组合已有的间隔序列 sp.C，所获得的 sp.C 序列如表 8 所示，其中原始序列为 sp.C0。

表 8 间隔序列 sp.C 高通量建库设计表

序列名称	序列碱基	SEQ ID NO:
sp.C0	GAGTGCCAGTAGCGACTTTTTTCACACTCGAAATACTCTTA CTACTGCTCTCTTGTTGTTTTATCACTTCTTGTTTCTTCTTG GTAAAT	131
sp.C1	GGCCCGCTGGACTCCACTTTTACAATTCTGAATACCTTTGCT CTTTTCCTATTCGACTTTTTTCTTTTCCGTTTTTAAACTTA AGTATTT	132
sp.C2	GGCCCGCTGGACTCCACTTTTACAATTCTGAATACCTTTGCT CTTTTCCTATTCGACTTTTATCCATTTTGAATTTACAGTCATC CAACTT	133
sp.C3	GGCCCGCTGGACTCCACTTTTACAATTCTGAATACCTTTGCT CTTTTCCTATTCGACTTTTTTCTTATCTTTTATTGGCTTCTTG TAACTT	134
sp.C4	GGCCCGCTGGACTCCACTTTTACAATTCTGATCGTTTCCTCT TGTAATCCGCTTTTTGTCTCTTTTCTTTTCCGTTTTTAAAC TTAAGTATTT	135
sp.C5	GGCCCGCTGGACTCCACTTTTACAATTCTGATCGTTTCCTCT TGTAATCCGCTTTTTGTCTCTTATCCATTTTGAATTTACAGT CATCCAACCTT	136
sp.C6	GGCCCGCTGGACTCCACTTTTACAATTCTGATCGTTTCCTCT TGTAATCCGCTTTTTGTCTCTTTTCTTATCTTTTATTGGCTTC TTGTAACTT	137
sp.C7	GGCCCGCTGGACTCCACTTTTACAATTCTGATTTCTTTTAGAT TCTGGTTCTTTGATATTTTCTTTTCCGTTTTTAAACTTAA GTATTT	138
sp.C8	GGCCCGCTGGACTCCACTTTTACAATTCTGATTTCTTTTAGAT TCTGGTTCTTTGATATTTATCCATTTTGAATTTACAGTCATCC AACTT	139
sp.C9	GGCCCGCTGGACTCCACTTTTACAATTCTGATTTCTTTTAGAT TCTGGTTCTTTGATATTTTCTTATCTTTTATTGGCTTCTTGT AACTT	140
sp.C10	GGCCCGCTGGACTCCACTTTTACAATTCTGAAATTCACCTTCT GTTTACTTGGCCCTCTTTTTTCTTTTCCGTTTTTAAACTTA AGTATTT	141
sp.C11	GGCCCGCTGGACTCCACTTTTACAATTCTGAAATTCACCTTCT GTTTACTTGGCCCTCTTTTATCCATTTTGAATTTACAGTCAT CCAACCTT	142

sp.C12	GGCCCGCTGGACTCCACTTTTACAATTCGAAATTCACTTCT GTTTACTTGGCCCTCTTTTTTCTTATCTTTTATTGGCTTCTTG TAACTT	143
sp.C13	GAGTGCCAGTACTCCACTTTTACAATTCGAAAATACTCTTT TGTAATCCGGCCCTCTTTTTTTATCACTTTTATTGGCTTCTT GTAACCTT	144
sp.C14	GGCCCGCTGGACTCCACTTTTTCACACTCGTCGTTTCCTCT GTTTACTTGTCTTGTTGTTTTCTTATCTTTTATTGGCTTCTTG GTAAAT	145
sp.C15	GGCCCGCTGGAGCGACTTTTACAATTCGAAATTCACTTCT TGTAATCCGGCCCTCTTTTTTCTTATCTTTTATTGGCTTCTTG TAACTT	146
sp.C16	GGCCCGCTGGACTCCACTTTTTCACACTCGAATTCACTTCT TGTAATCCGTCTTGTTGTTTTCTTATCTTTTATTGGCTTCTTG GTAAAT	147
sp.C17	TCTGCTCCTTGAAGGCGGTATTCCCAATTTAGAGTTCTTC CGAGTAATTGTATTGATAAGTTCTCCTTTAGTCTAATTGATG TTAAATT	148
sp.C18	CCTGTGACATATGCTAGTGTCTACACGGTGGAGTTCATGTT CACAATATACGTTGATTAGCTTTTTACATAGTAGACTTCTAA TCTATTG	149
sp.C19	CTGTAGGTTTCGATTCACTCTCACCTGTGGTGTTCGGTCTTT ATAGTGTATTAGTTATCCGTTATTTACTTTATCTCGATGCTAC TTTAA	150
sp.C20	CTTTCCCATGGTTCAGTTACCTCAGACTCCGTACATAGTATT CCATTTCAATTAGGTCCTATAGTACCATATTTCTTCCTTAGTCA AAATT	151
sp.C21	CCACCGTTGCTCTCCTTCTTCTCACGTCGTATACCTTTGCTC TTTTCCTATTCGACTTTTTTCTTTTCCGTTTTTAAAACTTAA GTATTT	152
sp.C22	CCACCGTTGCTCTCCTTCTTCTCACGTCGTATACCTTTGCTC TTTTCCTATTCGACTTTTATCCATTTTGAATTTACAGTCATCC AACTT	153
sp.C23	CCACCGTTGCTCTCCTTCTTCTCACGTCGTATACCTTTGCTC TTTTCCTATTCGACTTTTTTCTTATCTTTTATTGGCTTCTTGT AACTT	154
sp.C24	CCACCGTTGCTCTCCTTCTTCTCACGTCGTTTCGTTTCCTCTT GTAATCCGCTTTTTGTCTCTTTTCTTTTCCGTTTTTAAAACT TAAGTATTT	155

sp.C25	CCACCGTTGCTCTCCTTCTTCTCACGTCGTTTCGTTTCCTCTT GTAATCCGCTTTTTGTCTCTTATCCATTTGAATTTACAGTC ATCCAACCTT	156
sp.C26	CCACCGTTGCTCTCCTTCTTCTCACGTCGTTTCGTTTCCTCTT GTAATCCGCTTTTTGTCTCTTTTCTTATCTTTTATTGGCTTCT TGTAACCTT	157
sp.C27	CCACCGTTGCTCTCCTTCTTCTCACGTCGTTTCTTTTAGAT TCTGGTTCTTTGATATTTTCTTTTCCGTTTTTAAACTTAA GTATTT	158
sp.C28	CCACCGTTGCTCTCCTTCTTCTCACGTCGTTTCTTTTAGAT TCTGGTTCTTTGATATTTATCCATTTTGAATTTACAGTCATCC AACTT	159

利用 PCR 技术，将 pxlnR 上的原始序列 sp.C0 替换为表 10 所示的一系列 sp.C 间隔序列，得到含 sp.C 序列的完整的启动子序列。具体来讲，先将启动子分为 sp.C 上游和 sp.C，并以 pGSH7 为模板，用引物对扩增，得到 sp.C 上游片段；然后设计含 sp.C 序列的长引物进行无模板 PCR 扩增，得到 sp.C 片段。

利用 golden gate 技术，将实施例 2 中成功构建的载体 pXJH1 上的致死基因 CCDB 替换成上述完整的启动子序列，其中 golden gate 体系及条件与实施例 4 步骤（2）中表 5 所示内容类似。最后通过大肠转化、测序，得到测序正确且含不同 sp.C 序列启动子的质粒。

（5）标准测试系统的构建。与实施例 4 的步骤（3）类似，通过本实施例步骤（4）所构建的 28 个测试质粒得到带转化的酶切片段，后经转化得到酵母转化子并进行培养，获得待检测的测试菌液。

（6）不同间隔序列 sp.C 对启动子强度的影响测试。与实施例 4 的步骤（4）类似，通过流式分析检测本实施例步骤（5）所得菌液的荧光蛋白表达量，并进行数据处理。

结果：图 12 展示了含不同间隔序列 sp.C 的启动子诱导后的相对强度和相应

的诱导倍数情况。其中，图 12a 的纵坐标表示通过黄色荧光蛋白 YFP 表达量数据处理后得到的相对启动子强度 RPU，横坐标表示启动子所含的具体间隔序列 sp.C，横坐标所示负号对应的灰色柱表示启动子本底表达情况，横坐标所示正号对应的黑色柱表示启动子诱导后的表达情况；图 12b 的纵坐标表示诱导倍数 Fold change，横坐标表示启动子所含的具体间隔序列 sp.C。

可以看出，高通量建库设计得到的不同间隔序列 sp.C 对诱导型启动子的强度有所影响。当启动子本底的相对启动子强度 RPU 分布在 0.01-1 之间时，诱导后的相对启动子强度 RPU 分布在 7-16 之间，其中含 sp.C6、sp.C13-sp.C24 序列的启动子的本底相对强度比对照组 sp.C0 更低，含 sp.C20 序列启动子的本底表达低于 0.001。此外，大部分启动子的诱导倍数 fold change 大于原始启动子 pxlnR，其中含 sp.C18 序列的启动子的诱导倍数超过 500。

实施例 7：非天然合成型启动子的构建及启动子强度测试

(1) 合成型启动子的构建。按本底黄色荧光蛋白 YFP 表达量的大小，对实施例 4-6 中获得的间隔序列 sp.A、sp.B 和 sp.C 进行筛选，而后参考启动子 pxlnR 的构建，以 xlnR 为上游激活序列 UAS，随机组合三种间隔序列，得到一系列全新的木糖诱导型启动子序列，其组成架构如表 9 所示，具体序列见 SEQ ID NO:160-SEQ ID NO:182。

表 9 新的启动子序列及名称

质粒名称	启动子名称	间隔序列 sp.A 编号	间隔序列 sp.B 编号	间隔序列 sp.C 编号
pXJH360	p205	sp.A10	sp.B1	sp.C20
pXJH361	p206	sp.A20	sp.B25	sp.C18
pXJH362	p207	sp.A9	sp.B7	sp.C19
pXJH363	p208	sp.A14	sp.B5	sp.C21
pXJH364	p209	sp.A24	sp.B23	sp.C22

pXJH365	p210	sp.A18	sp.B8	sp.C15
pXJH366	p211	sp.A22	sp.B27	sp.C16
pXJH367	p212	sp.A21	sp.B11	sp.C23
pXJH368	p213	sp.A12	sp.B17	sp.C24
pXJH369	p214	sp.A6	sp.B14	sp.C7
pXJH370	p215	sp.A7	sp.B15	sp.C8
pXJH371	p216	sp.A21	sp.B1	sp.C15
pXJH372	p217	sp.A14	sp.B1	sp.C16
pXJH373	p218	sp.A9	sp.B1	sp.C24
pXJH374	p219	sp.A14	sp.B25	sp.C24
pXJH375	p220	sp.A21	sp.B25	sp.C15
pXJH376	p221	sp.A22	sp.B26	sp.C22
pXJH377	p222	sp.A20	sp.B25	sp.C6
pXJH379	p224	sp.A12	sp.B7	sp.C16
pXJH380	p225	sp.A14	sp.B8	sp.C19
pXJH381	p226	sp.A9	sp.B27	sp.C23
pXJH382	p227	sp.A24	sp.B23	sp.C21

启动子 p205-227 是通过 PCR 方式获得的，具体来讲，先将启动子分为上、中、下三个片段，即片段 1、片段 2 和片段 3，然后对片段 1 和片段 3 进行有模板的 PCR 扩增，对片段 2 进行无模板 PCR 扩增，最后以片段 1、片段 2 和片段 3 为模板，用引物对扩增，得到完整的启动子。

利用 golden gate 技术，将实施例 2 中成功构建的载体 pXJH1 上的致死基因 CCDB 替换成上述完整的启动子序列，其中 golden gate 体系及条件与实施例 4 步骤（2）中表 5 所示内容类似。最后通过大肠转化、测序，得到测序正确的质粒。

（2）标准测试系统的构建。与实施例 4 的步骤（3）类似，通过本实施例步骤（1）所构建的 23 个测试质粒得到带转化的酶切片段，后经转化得到酵母转化子并进行培养，获得待检测的测试菌液。

（3）启动子强度测试。与实施例 4 的步骤（4）类似，通过流式分析检测

本实施例步骤（3）所得菌液的荧光蛋白表达量，并进行数据处理。

结果：图 13 展示了不同启动子诱导后的相对强度和相应的诱导倍数情况。其中，图 13a 的纵坐标表示通过黄色荧光蛋白 YFP 表达量数据处理后得到的相对启动子强度 RPU，横坐标表示启动子所含的具体启动子序列，灰色柱表示启动子本底表达情况，黑色柱表示启动子诱导后的表达情况；图 13b 的纵坐标表示诱导倍数 Fold change，横坐标表示启动子所含的具体启动子序列。

可以看出，本方案所构建的新启动子大部分表现出了良好特征。启动子本底表达量较低，其相对启动子强度 RPU 基本低于 0.1，启动子 p205 和 p206 本底的相对强度 RPU 更是小于 0.001，诱导后的相对启动子强度 RPU 均在 4~12 之间。与此同时，除启动子 p215 以外，新启动子的诱导倍数基本上大于原始启动子 pxlnR 的诱导倍数，其中启动子 p208 和启动子 p223 的诱导倍数更是高于 600。

综上所述，本发明提供一种全合成的酵母诱导型启动子及其构建方法。该方法通过将启动子的构架模块化，对其上游激活区域和核心启动子区域进行拆分，并利用理性设计、计算机模型辅助和高通量建库等方法对间隔序列进行筛选，探索了以酿酒酵母作为底盘微生物的启动子的架构模式，获得了全新的非天然的酵母诱导型启动子序列，解决了由于启动子强度不足、种类少导致的特征良好启动子数量不足，或是酿酒酵母体内易发生同源重组的问题。

以上所述仅是本发明的一些实施方式。对于本领域技术人员来说，在不脱离本发明创造构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。

权 利 要 求 书

1. 一种全合成的酵母诱导型启动子的构建方法，其特征在于，包括：

确定所述启动子的基本架构，所述启动子从 5'→3'依次包括操作性连接的：上游激活序列、第一间隔序列、核小体不利序列、第二间隔序列、TATA 盒、第三间隔序列、转录起始位点；

利用理性设计方法、模型建库方法和/或高通量建库方法筛选所述第一间隔序列、所述第二间隔序列和所述第三间隔序列的序列碱基；以及

组合所述第一间隔序列、所述第二间隔序列和所述第三间隔序列，获得所述启动子。

2. 根据权利要求 1 所述的全合成的酵母诱导型启动子的构建方法，其特征在于，所述筛选第一间隔序列、第二间隔序列和第三间隔序列的序列碱基包括：

通过理性设计方法筛选所述第一间隔序列的序列碱基；

通过模型建库方法筛选所述第二间隔序列的序列碱基；和

通过高通量建库方法筛选所述第三间隔序列的序列碱基。

3. 根据权利要求 2 所述的全合成的酵母诱导型启动子的构建方法，其特征在于，所述通过模型建库方法筛选第二间隔序列的序列碱基包括：

优化转录模型并随机生成含所述第二间隔序列的启动子片段；

利用优化后的转录模型预测含所述第二间隔序列的启动子片段的强度；以及
筛选含所述第二间隔序列的目标启动子。

4. 根据权利要求 3 所述的全合成的酵母诱导型启动子的构建方法，其特征在于，所述优化转录模型包括：

向转录模型中输入序列组合，所述序列组合由不同长度的序列摆放在不同位置得到；以及

以模型预测准确性为标准筛选所述序列组合，获得优化的序列位置及长度。

5. 根据权利要求 3 或 4 所述的全合成的酵母诱导型启动子的构建方法，其特征在于，所述通过高通量建库方法筛选第三间隔序列的序列碱基包括：

分析酶切位点并设计简并引物，通过扩增获得含所述第三间隔序列的启动子

库；

利用所述含第三间隔序列的启动子库构建重组载体；以及

将所述重组载体转化至宿主并筛选含所述第三间隔序列的目标启动子。

6. 一种全合成的酵母诱导型启动子，其特征在于，所述全合成的酵母诱导型启动子由权利要求 1-5 中任一项所述的全合成的酵母诱导型启动子的构建方法获得。

7. 根据权利要求 6 所述的全合成的酵母诱导型启动子，其特征在于，所述启动子中上游激活序列的数量为 1、2、3 或 4。

8. 一种重组载体，其特征在于，所述重组载体包括权利要求 6 或 7 所述的全合成的酵母诱导型启动子。

9. 一种重组菌，其特征在于，所述重组菌的基因组中包含权利要求 8 所述的重组载体。

10. 权利要求 6 或 7 所述的全合成的酵母诱导型启动子、权利要求 8 所述的重组载体或权利要求 9 所述的重组菌在调节酿酒酵母代谢中的应用。

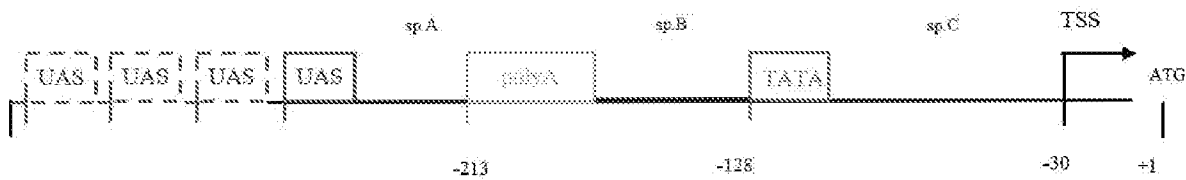


图 1

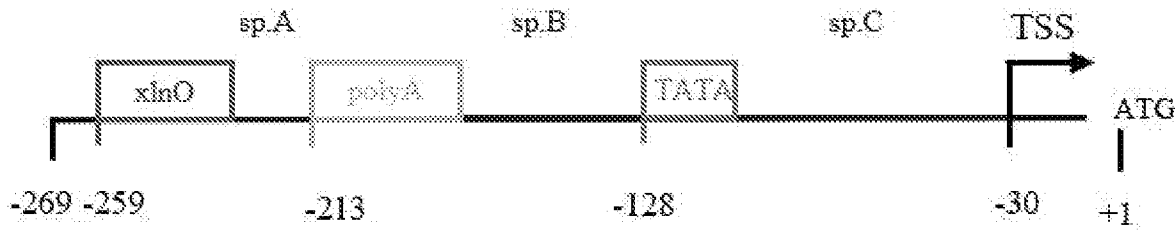


图 2

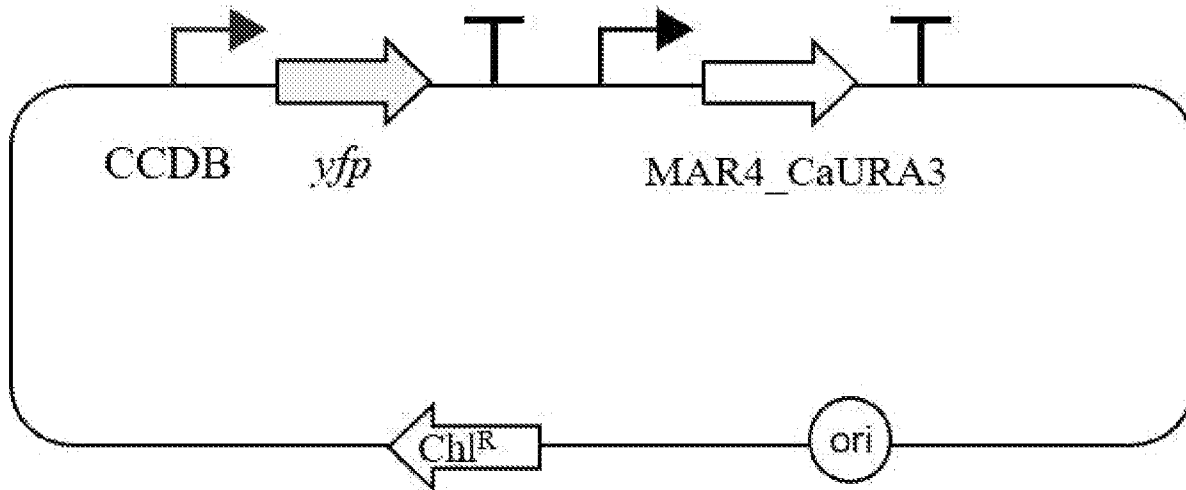


图 3

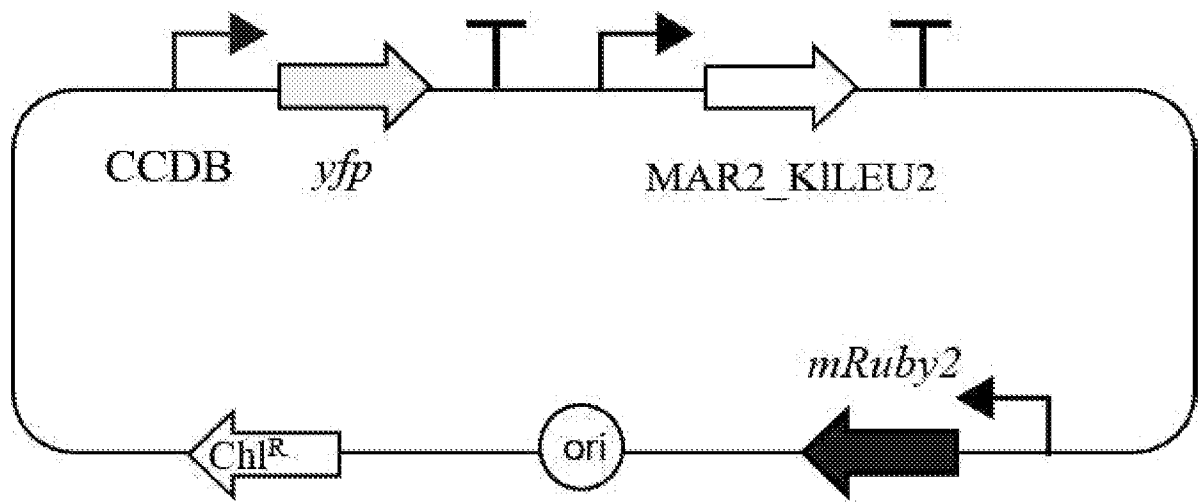


图 4

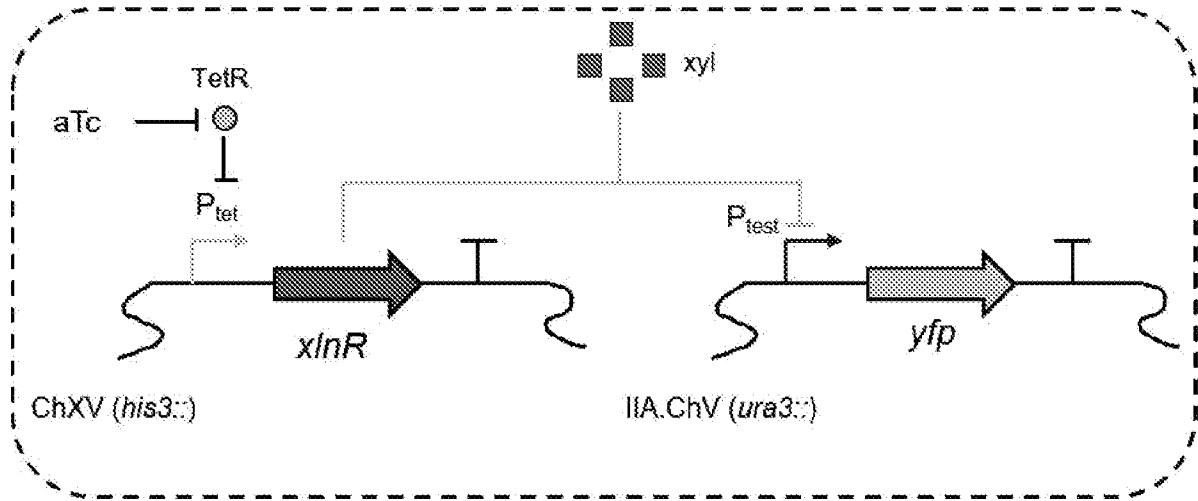


图 5

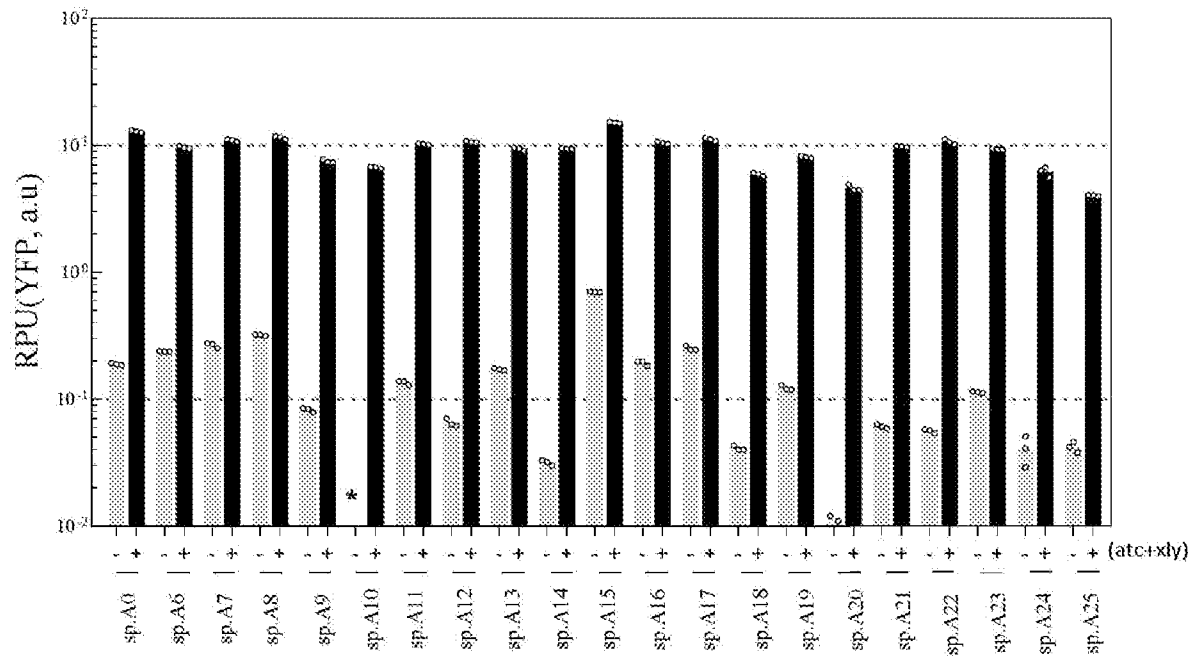


图 6a

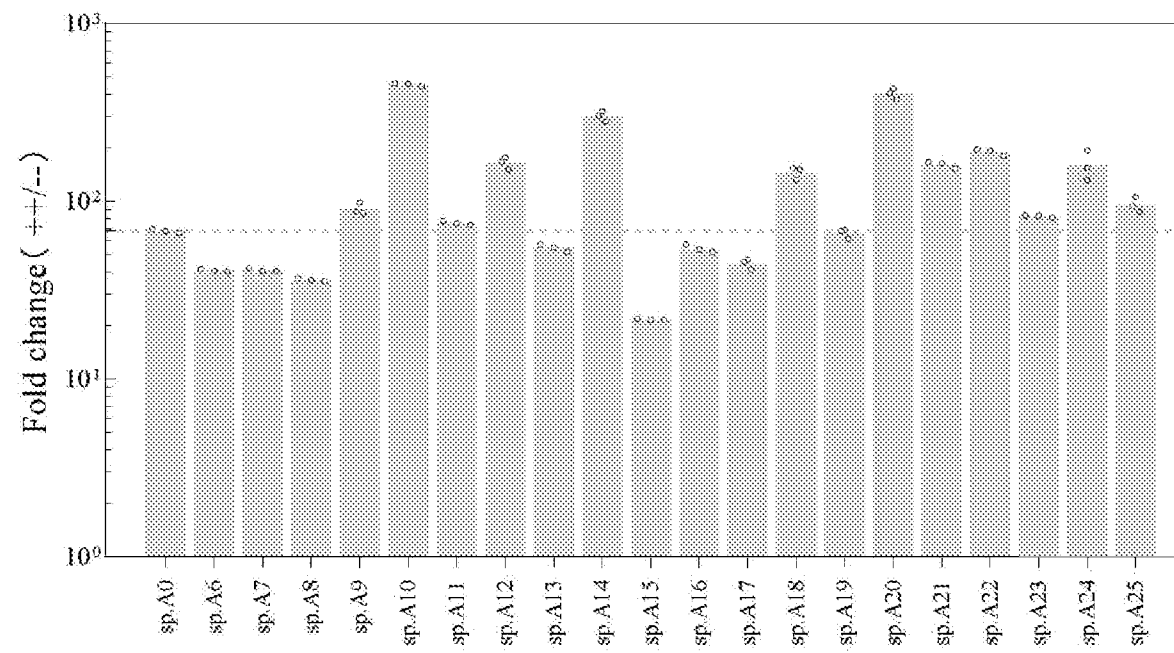


图 6b

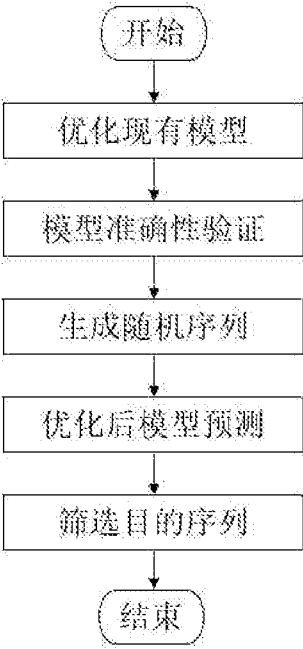


图 7

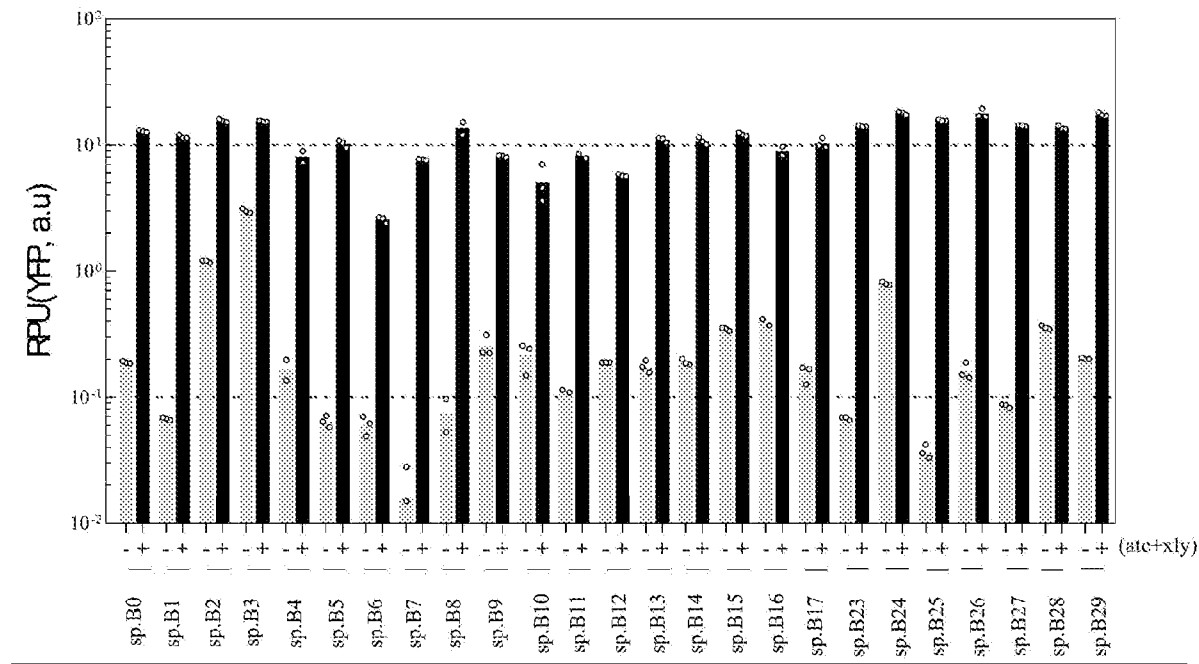


图 8a

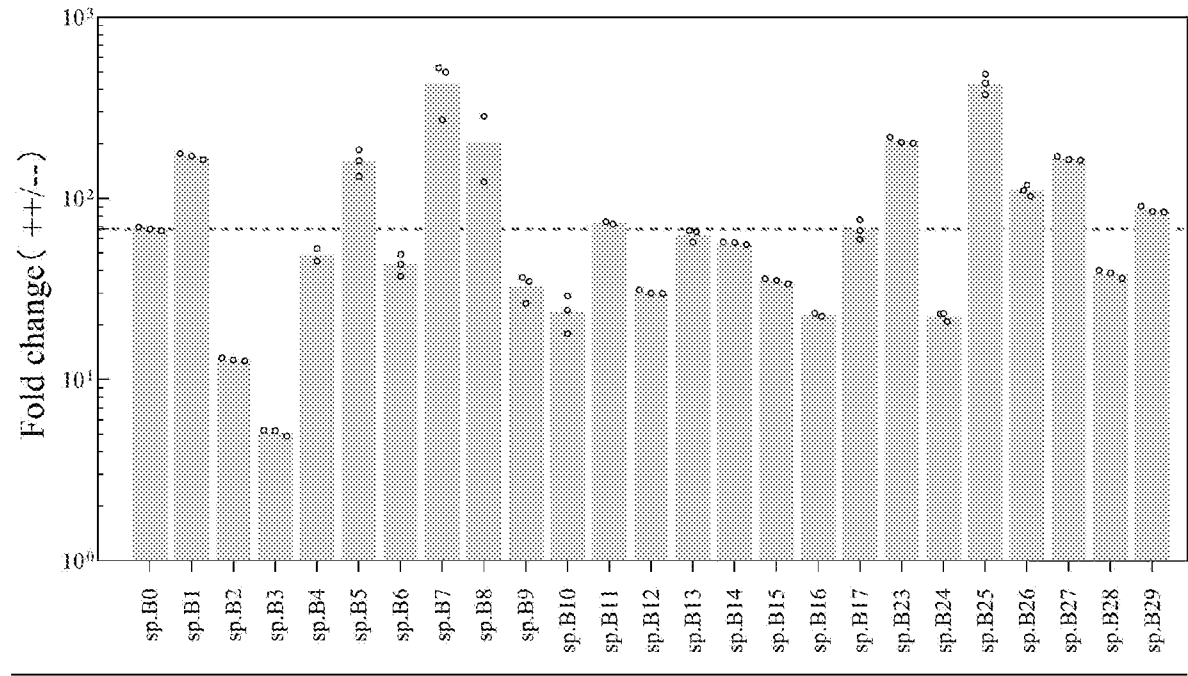
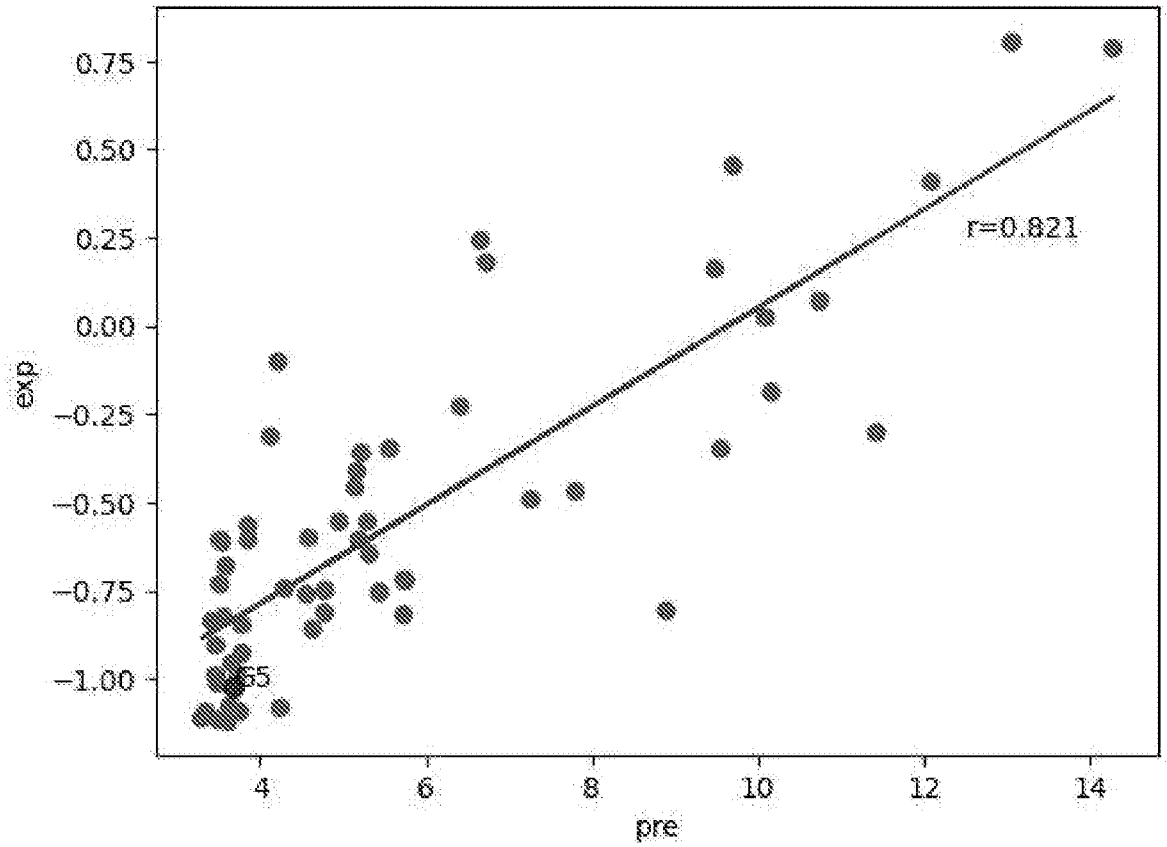


图 8b



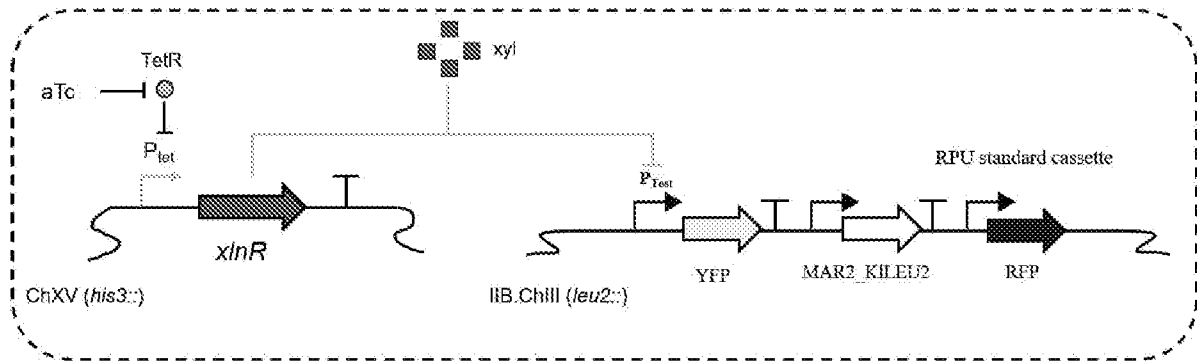


图 10

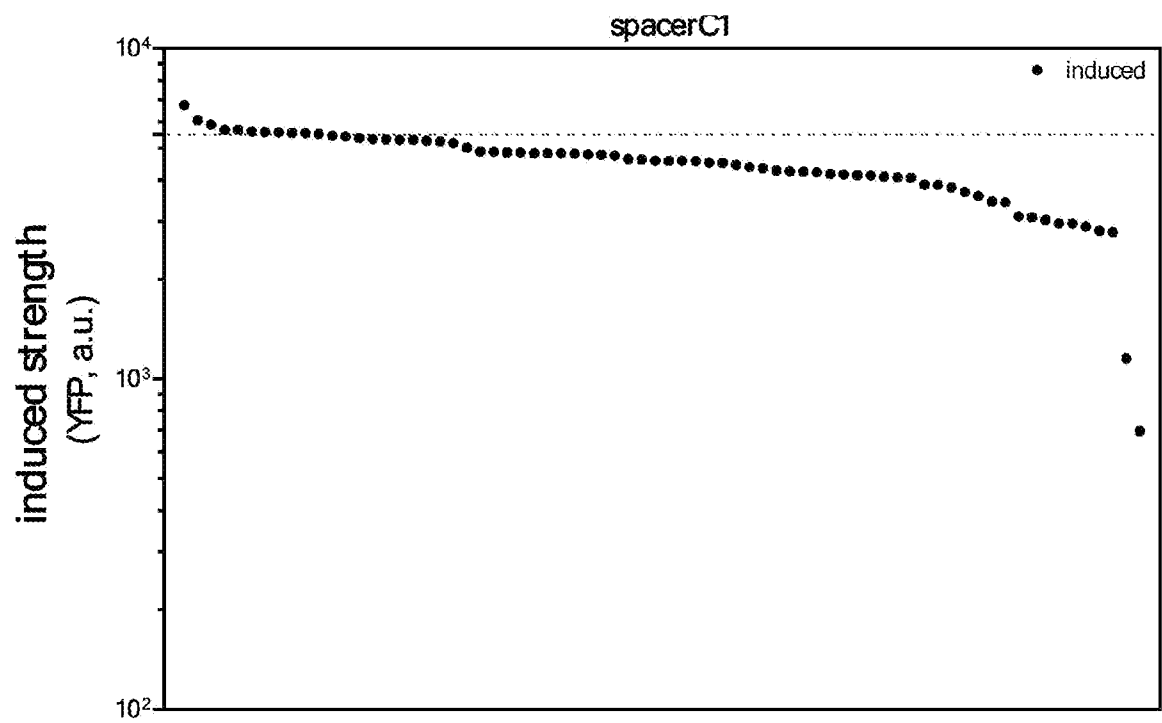


图 11a

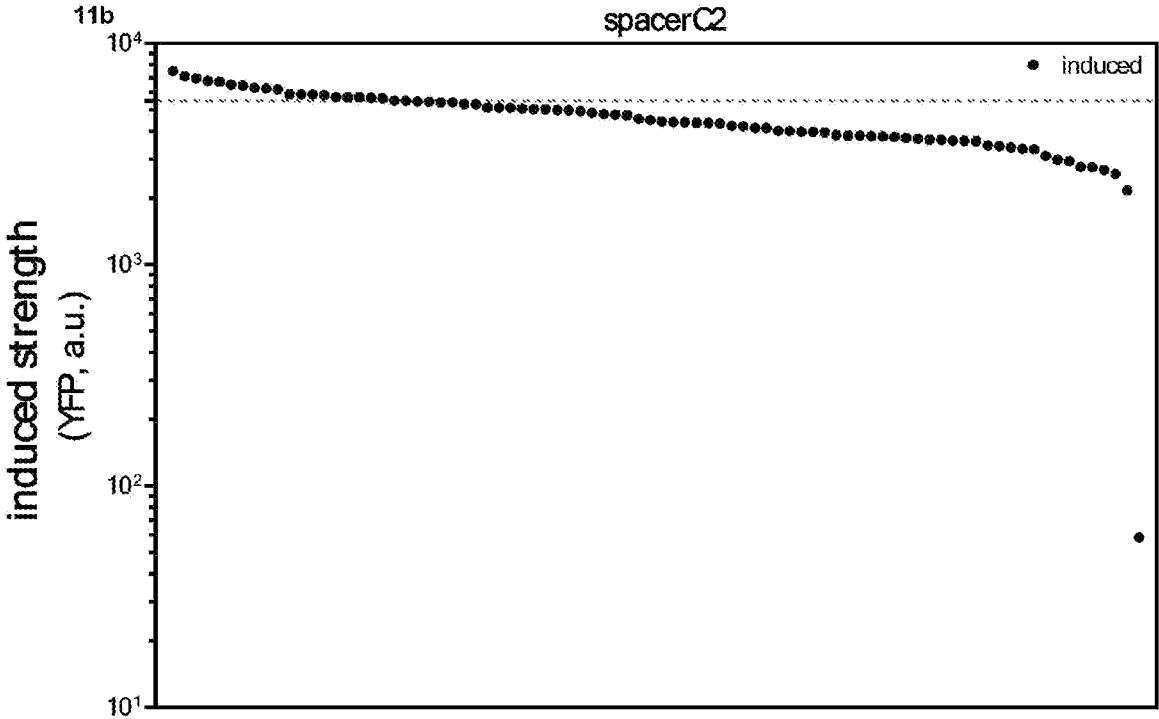


图 11b

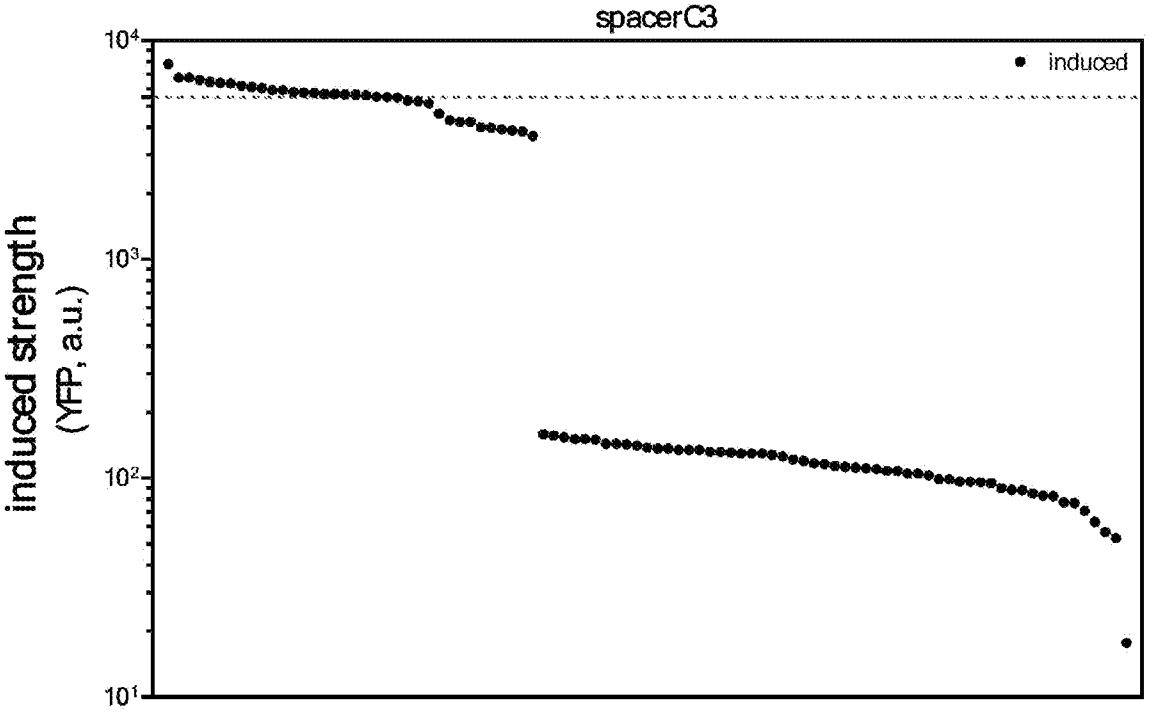


图 11c

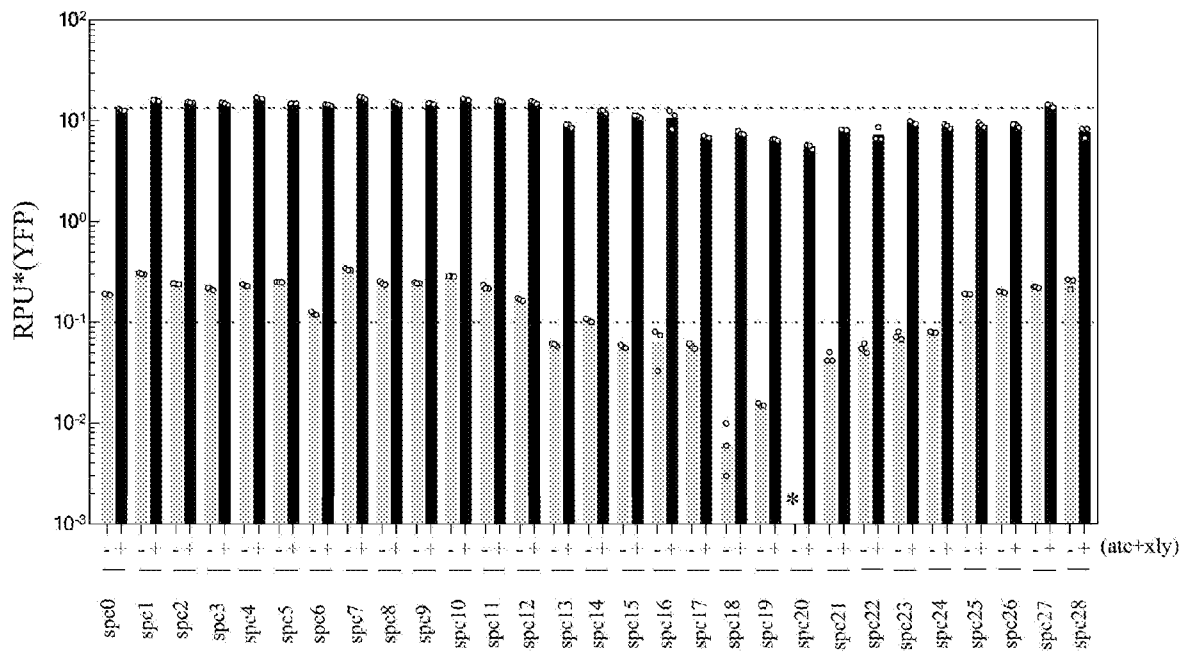


图 12a

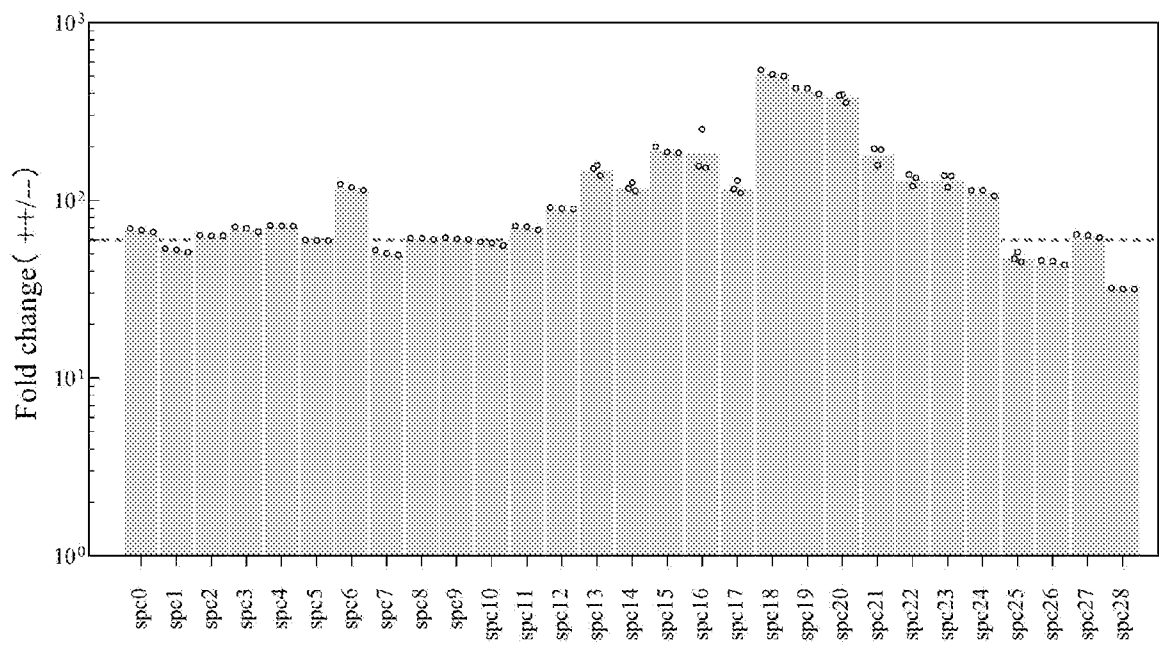


图 12b

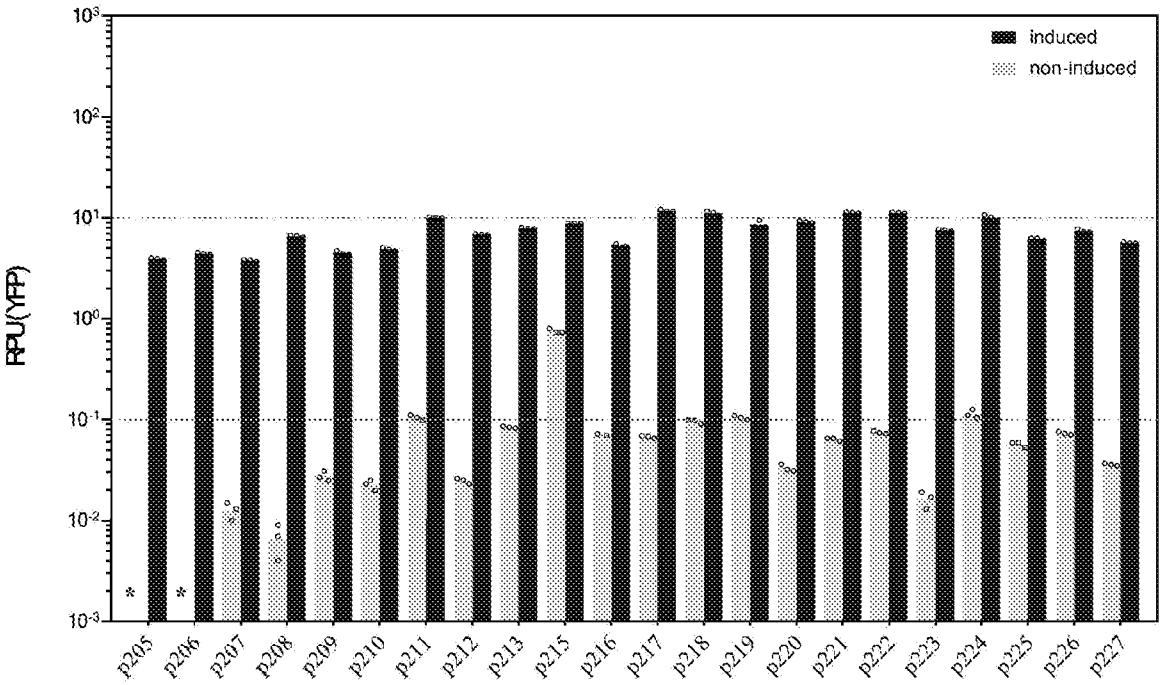


图 13a

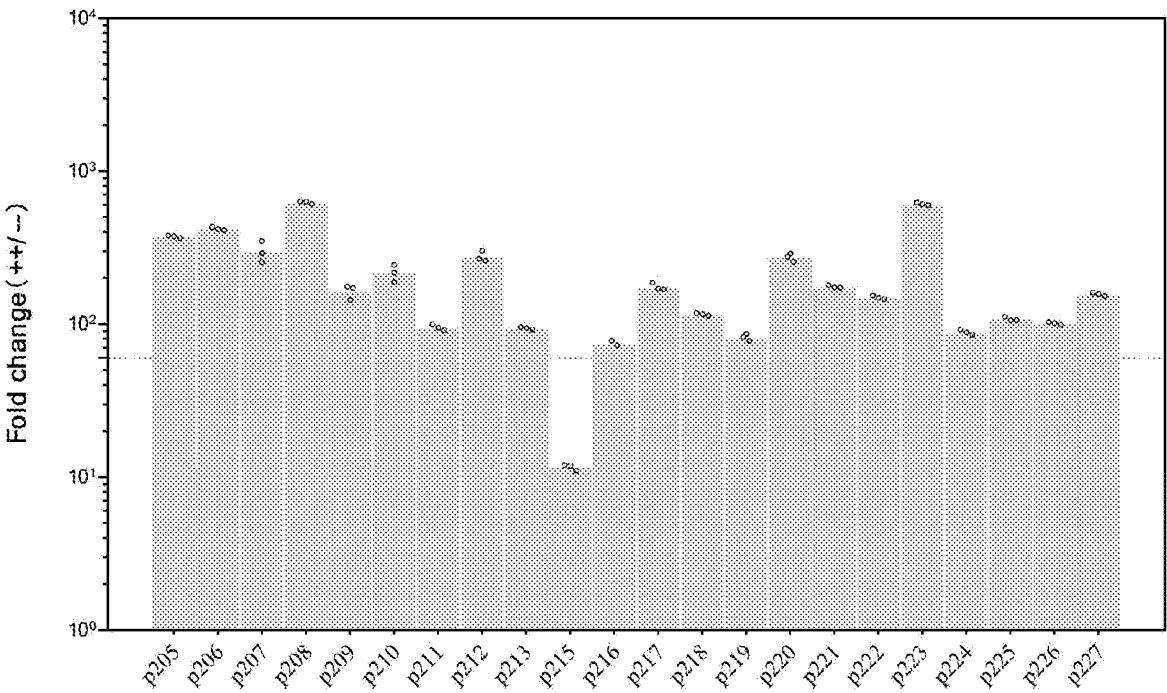


图 13b

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/098388

A. CLASSIFICATION OF SUBJECT MATTER

C12N15/81(2006.01)i; C12N15/113(2010.01)i; C12N1/19(2006.01)i; G16B5/00(2019.01)i; G16B20/30(2019.01)i;
G16B30/00(2019.01)i; C12R1/865(2006.01)n

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)

IPC:C12N,G16B,C12R

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)

DWPI, SIPOABS, CNABS, USTXT, JPTXT, EPTXT, WOTXT, CNTXT, CNKI, ISI WEB OF SCIENCE, PUBMED: 酵母, 诱导型启动子, 合成, Yeast, Promoter, padh2, Spacer, Inducible Promoter, Synthetic; GenBank, EMBL, DDBJ, 中国专利生物序列检索系统, China Patents Biological Sequence Search System: 关于SEQ ID NOs: 1-159的检索, search for SEQ ID NOs: 1-159

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107236732 A (QINGDAO BAIHUIZHIYE BIOTECHNOLOGY CO., LTD.) 10 October 2017 (2017-10-10) see claims 1-6	1-10
A	US 5063154 A (WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH) 05 November 1991 (1991-11-05) see claims 1-33	1-10
A	US 2019062730 A1 (WISCONSIN ALUMNI RESEARCH FOUNDATION) 28 February 2019 (2019-02-28) see claims 1-17	1-10
A	US 2017275635 A1 (LELAND STANFORD JUNIOR UNIVERSITY) 28 September 2017 (2017-09-28) see claims 1-18	1-10
A	Benjamin J. Kotopka et al. "Model-driven generation of artificial yeast promoters" <i>Nature Communications</i> , Vol. 11, No. 2113, 31 December 2020 (2020-12-31), pages 1-12, see abstract, "Results" part, and supplementary figures 1 and 4	1-10



Further documents are listed in the continuation of Box C.



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
 “D” document cited by the applicant in the international application
 “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

02 March 2024

Date of mailing of the international search report

04 March 2024

Name and mailing address of the ISA/CN

**China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088**

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/098388

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.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/098388

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107236732	A	10 October 2017	None			
US	5063154	A	05 November 1991	None			
US	2019062730	A1	28 February 2019	US	10745693	B2	18 August 2020
US	2017275635	A1	28 September 2017	US	2020291411	A1	17 September 2020
				US	11795464	B2	24 October 2023
				US	10612032	B2	07 April 2020

A. 主题的分类

C12N15/81(2006.01)i; C12N15/113(2010.01)i; C12N1/19(2006.01)i; G16B5/00(2019.01)i; G16B20/30(2019.01)i; G16B30/00(2019.01)i; C12R1/865(2006.01)n

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

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:C12N,G16B,C12R

包含在检索领域中的除最低限度文献以外的检索文献

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))
DWPI, SIPOABS, CNABS, USTXT, JPTXT, EPTXT, WOTXT, CNTXT, CNKI, ISI WEB OF SCIENCE, PUBMED:酵母, 诱导型启动子, 合成, Yeast, Promoter, padh2, Spacer, Inducible Promoter, Synthetic GenBank+EMBL+DDBJ+中国专利生物序列检索系统:关于SEQ ID NOs:1-159的检索

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	CN 107236732 A (青岛百慧智业生物科技有限公司) 2017年10月10日 (2017 - 10 - 10) 参见权利要求1-6	1-10
A	US 5063154 A (WHITEHEAD BIOMEDICAL INST) 1991年11月5日 (1991 - 11 - 05) 参见权利要求1-33	1-10
A	US 2019062730 A1 (WISCONSIN ALUMNI RES FOUND) 2019年2月28日 (2019 - 02 - 28) 参见权利要求1-17	1-10
A	US 2017275635 A1 (UNIV LELAND STANFORD JUNIOR) 2017年9月28日 (2017 - 09 - 28) 参见权利要求1-18	1-10
A	Benjamin J. Kotopka等. "Model-driven generation of artificial yeast promoters" NATURE COMMUNICATIONS, 第11卷, 第2113期, 2020年12月31日 (2020 - 12 - 31), 第1-12页, 参见摘要, "Results" 部分, 补充附图1,4	1-10

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

★ 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
“D” 申请人在国际申请中引证的文件
“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, 但为了理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

国际检索实际完成的日期
2024年3月2日

国际检索报告邮寄日期
2024年3月4日

ISA/CN的名称和邮寄地址
中国国家知识产权局
中国北京市海淀区蓟门桥西土城路6号 100088

授权官员
王启扬
电话号码 (+86) 010-62411211

PCT/ISA/210 表(第2页) (2022年7月)

第I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表进行的:

- a. ☒ 作为国际申请的一部分提交的:
- b. ☐ 为国际检索的目的在国际申请日之后提交(细则13之三.1(a)),
☐ 附有说明序列表不超出所提交国际申请公开范围的声明。

2. ☐ 本报告是在没有收到符合WIPO ST.26标准的序列表的情况下,考虑了国际申请中披露的任何核苷酸和/或氨基酸序列,在可进行有意义检索的范围内做出的。

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/CN2023/098388

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	107236732	A	2017年10月10日	无			
US	5063154	A	1991年11月5日	无			
US	2019062730	A1	2019年2月28日	US	10745693	B2	2020年8月18日
US	2017275635	A1	2017年9月28日	US	2020291411	A1	2020年9月17日
				US	11795464	B2	2023年10月24日
				US	10612032	B2	2020年4月7日