

(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
2025 年 6 月 26 日 (26.06.2025)



(10) 国际公布号
WO 2025/129656 A1

- (51) 国际专利分类号:
G16B 5/00 (2019.01) **C12N 15/81** (2006.01)
- (21) 国际申请号: PCT/CN2023/141107
- (22) 国际申请日: 2023 年 12 月 22 日 (22.12.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 中国科学院深圳先进技术研究院 (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY CHINESE ACADEMY OF SCIENCES) [CN/CN]; 中国广东省深圳市南山区西丽深圳大学城学苑大道1068号 518000 (CN)。
- (72) 发明人: 陈业(CHEN, Ye); 中国广东省深圳市南山区西丽深圳大学城学苑大道1068号 518000 (CN)。黎欣睿(LI, Xinrui); 中国广东省深圳市南山区西

丽深圳大学城学苑大道1068号 518000 (CN)。栗曾理(LI, Cengli); 中国广东省深圳市南山区西丽深圳大学城学苑大道1068号 518000 (CN)。

- (74) 代理人: 深圳五邻知识产权代理事务所(普通合伙)等(SHENZHEN WULIN INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP) et al.); 中国广东省深圳市龙华区民治街道民新社区深国际华南数字谷1#12F 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,

(54) Title: MODULAR ORTHOGONAL EUKARYOTIC TRANSCRIPTION FACTOR MODEL CONSTRUCTION METHOD

(54) 发明名称: 一种模块化正交真核转录因子模型构建方法

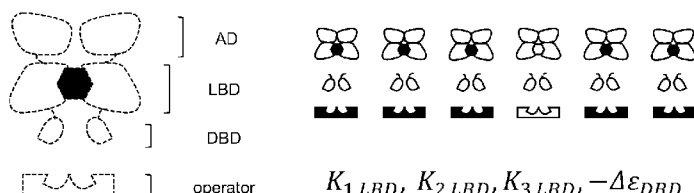


图 1

(57) Abstract: The present invention provides a modular orthogonal eukaryotic transcription factor model construction method, comprising the following steps: S1, establishing the basic architecture DBD-LBD-AD of a transcription factor; S2, constructing a domain screening vector and system to screen effective DBD, LBD and AD; S3, screening out corresponding parameters of each DBD and LBD domain; S4, on the basis of the corresponding parameters of each DBD and LBD domain, characterizing a transcription factor regulation process, and establishing a quantitative model of transcription regulation; predicting the transcription factor under conditions of different DBD and LBD combinations by means of the quantitative model of transcription regulation, to obtain a transcription factor having optimized performance. The modular orthogonal eukaryotic transcription factor model construction method of the present invention, by means of determining the basic architecture of a modular transcription factor, ultimately establishes a quantitative model of transcription regulation, achieving modular domain parameter optimization of the transcription factor, and obtaining a transcription factor having optimized performance.

(57) 摘要: 本发明提供一种模块化正交真核转录因子模型构建方法, 包括以下步骤: S1、确立转录因子基本架构DBD-LBD-AD; S2、构建结构域筛选载体和系统, 筛选得到有效的DBD、LBD和AD; S3、筛选出DBD、LBD各结构域的对应该参数; S4、根据DBD、LBD各结构域的对应该参数来表征转录因子调控过程, 建立转录调控定量模型; 通过转录调控定量模型预测不同DBD和LBD组合情况下转录因子, 获得性能优化的转录因子。本发明模块化正交真核转录因子模型构建方法, 通过确定了模块化转录因子的基本架构, 最终建立了转录调控定量模型, 实现对转录因子进行模块化结构域参数优化, 获得了性能优化的转录因子。

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, 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))。

一种模块化正交真核转录因子模型构建方法

技术领域

本发明属于生物技术领域，具体涉及一种模块化正交真核转录因子模型构建方法。

背景技术

细胞能够接受外界的化学或物理信号，将其转化为生物信号，从而对环境做出响应。转录因子可以作为一种信号转化器，通过识别并结合DNA上特定位点，称为操纵序列，影响RNA聚合酶结合启动子的概率，从而影响基因转录。在感受外界化学或物理信号刺激后，转录因子通过一系列结构改变产生不同的操纵序列结合亲和力，与启动子结合或从启动子上掉落，进而影响下游基因的转录和表达。结合DNA的转录因子必须具有三个特性：①能够与特定的DNA序列结合；②能够响应信号；③能够控制转录（Schleif, R.F., Modulation of DNA binding by gene-specific transcription factors. *Biochemistry*, 2013. 52(39): p. 6755-65.）。这些特性可以分别由结构和功能模块化的不同结构域来执行。根据其执行的功能，转录因子由DNA结合结构域、配体结合结构域、转录激活结构域组成。每一个结构域在行使功能时，可以分别用几个反应方程来描述。每一个反应方程式又可以抽象为一个参数——反应平衡常数。因此，在我们的酿酒酵母宿主中，特定序列的某一结构域可以抽象为一个特定参数，不同参数的不同结构域共同组成了特定转录调控行为的完整转录因子，而转录因子在特定转录系统中的行为又被反应方程所描述。

目前常用于真核细胞转录调控的转录因子根据来源可分为原核天然宿主转录因子和人工合成转录因子两大类。来自原核的LacI、TetR和XylR分别受到IPTG、aTc和xylose的诱导应用于真核细胞中（Chen, Y., et al., Genetic circuit design automation for yeast. *Nature Microbiology*, 2020. 5(11): p. 1349-1360.）。人工合成转录因子常利用锌指、TALE和CRISPR-dCas9作为自定义的DNA结合模块，连接配体结合结构域或转录激活结构域等功能性结构域进行构建[9]。如Khalil基于三个锌指结构结合9bp DNA建立的一系列DBD-operator系统（Khalil, A.S., et al., A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 2012. 150(3): p. 647-58.）。另外，分别将模块化的DBD、LBD、AD组装，形成真核系统中常用的人工转录因子XEV, lexA-ER-VP16（Louvion, J.F., B. Havaux-Copf, and D. Picard, Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*, 1993. 131(1): p. 129-34.）。

复杂基因线路中多个不同诱导系统的使用要求获得诱导信号接受和DNA序列结合两方面均相互正交的多对转录因子。现有的真核转录因子难以满足这一要求。主要原因有：其一，特征良好的常用真核转录系统数量不足，强度不足，需要通过使用多个operator或增加协同性相互作用来进行优化；其二，大多数现有真核转录调控系统中转录因子以整体形式作用，存在本底泄露高表达、诱导强度不足，动态范围较差的现象；其三，真核转录系统之间，或与宿主已有系统之间存在串扰。因此，为了解决这一类问题，需要开发一组通用的可设计的正交真核转录体系。

对于转录调控过程的优化可以分为转录因子本身性能参数的优化和转录调控模式的优化两方面。Khalil 等提出的一系列相互正交的不同调控活性的锌指库对应不同的转录因子结合 DNA 的亲和力参数, 通过使用高亲和力的锌指 DNA 结合模块, 可以对转录因子本身的性能进行优化 (Khalil, A.S., et al., A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 2012. 150(3): p. 647-58.)。另外通过使用多个 operator 来增加潜在的希尔系数, 通过使用 8 个 operator 的输出启动子达到~60 倍的调控倍数, 是单个 operator 启动子情况下的~10 倍 (Khalil, A.S., et al., A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 2012. 150(3): p. 647-58.)。除了对单个转录因子的转录调控过程的优化, 转录调控模式的改变也提供了一个优化思路。2019 年和 2023 年, Khalil 等也通过增加表达一个结合多个锌指转录激活因子的“支架”来增加转录调控的协同性, 从而在单个转录因子低亲和力的情况下对转录调控进行优化 (Bashor, C.J., et al., Complex signal processing in synthetic gene circuits using cooperative regulatory assemblies. *Science*, 2019. 364(6440): p. 593-597.) (Bragdon, M.D.J., et al., Cooperative assembly confers regulatory specificity and long-term genetic circuit stability. *Cell*, 2023. 186(18): p. 3810-3825. e18.)。2018 年 Rob Phillips 等使用 Monod-Wyman-Changeux 模型对转录因子变构调控过程针对转录因子浓度、诱导剂浓度和转录因子-operator 结合能的变化进行了定量描述, 但其变构系统只能通过突变对转录因子-operator 结合能进行调整, 无法实现结构域的模块化和有效替换优化 (Razo-Mejia M, Barnes S L, Belliveau N M, et al. Tuning Transcriptional Regulation through

Signaling: A Predictive Theory of Allosteric Induction. 2017[2023-11-10].DOI:10.1016/j.cels.2018.02.004)。

现有真核转录调控系数中转录因子多以整体形式作用，调控范围较差，具有高本底泄露和低诱导强度；缺少多诱导系统所需的正交性方面的考虑，没有多对同时存在的、相互及与宿主之间不存在交互作用的高诱导范围的转录因子；现有模型以变构型调控系统为基础，无法实现结构域的模块化和有效替换，可用范围较窄。

发明内容

本发明提供一种模块化正交真核转录因子模型构建方法，解决现有技术中真核转录调控以变构型调控为基础，无法实现结构域的模块化和有效替换，可用范围较窄的不足

为实现上述目的，本发明采用以下技术方案：

一种模块化正交真核转录因子模型构建方法，包括以下步骤：

S1、确立由 DNA 结合结构域（DBD）、配体结合结构域（LBD）、转录激活结构域（AD）三种结构域组成的转录因子基本架构 DBD-LBD-AD；

S2、在转录因子基本架构 DBD-LBD-AD 基础上，分别构建 DNA 结合结构域(DBD)、配体结合结构域（LBD）、转录激活结构域（AD）三种结构域的筛选酿酒酵母菌株，然后对筛选酿酒酵母菌株进行筛选测试培养，筛选得到有效的 DBD、LBD 和 AD；

S3、在转录因子基本架构 DBD-LBD-AD 基础上，以筛选出的 DBD、LBD 分别构建用于结构域参数定量测试的定量测试酿酒酵母菌株；对定量测试酿酒酵母菌株进行定量测试培养，获得转录因子表达量下的调控诱导范围和调控倍数；对获得调控诱导范围和调控倍数数据进行拟合，得到筛选出的 DBD、LBD 各结构域的对

应参数；

S4、根据 DBD、LBD 各结构域的对应参数来表征转录因子调控过程，建立转录调控定量模型；通过转录调控定量模型预测不同 DBD 和 LBD 组合情况下转录因子，获得性能优化的转录因子。

本发明中，所述基本架构 DBD-LBD-AD 中以 *E. coli* SOS 调控蛋白 *lexA* 的 1-87aa 作为 DBD，以沼泽红假单胞菌群体感应转录因子 RpaR 的 1-179aa 作为 LBD，以 VP16 作为 AD。

本发明中，S2 中 DNA 结合结构域（DBD）的筛选酿酒酵母菌株，通过以下方式得到：构建 DBD 的筛选转录因子表达载体，构建对应的报告基因表达载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合到酵母中，得到 DBD 筛选酿酒酵母菌株。

进一步地，所述 DBD 的筛选转录因子表达载体以 aTc 诱导的 *ptet* 作为启动子、*tEN02* 作为终止子、用 CCDB 替换转录因子基本架构 DBD-LBD-AD 中的 DBD 结构域得到。

进一步地，所述 DBD 的筛选转录因子表达载体序列如 SEQ ID NO: 1。

进一步地，S2 中构建 DBD 对应报告基因表达载体以对应的报告基因用含有两个对应 DBD 调控的 *operator* 的启动子，控制黄色荧光蛋白 YFP 的表达，以 *tEN02* 作为终止子；将 CCDB 放置在报告基因质粒的启动子位置得到。

本发明中，所述筛选得到有效的 DBD 序列如 SEQ ID NO: 7-18。

本发明中，S2 中配体结合结构域（LBD）的筛选酿酒酵母菌株，通过以下方式得到：构建 LBD 的筛选转录因子表达载体，构建对应的报告基因表达载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合

到酵母中，得到 LBD 筛选酿酒酵母菌株。

进一步地，所述 LBD 的筛选转录因子表达载体以 aTc 诱导的 ptet 作为启动子、tEN02 作为终止子、用 CCDB 替换转录因子基本架构 DBD-LBD-AD 中的 LBD 结构域得到。

进一步地，所述 LBD 的筛选转录因子表达载体序列如 SEQ ID NO: 2。

进一步地，S2 中构建 LBD 对应报告基因表达载体对应的报告基因用含有两个对应 DBD 调控的 operator 的启动子，控制黄色荧光蛋白 YFP 的表达，以 tEN02 作为终止子得到。

本发明中，所述筛选得到有效的 LBD 序列如 SEQ ID NO: 19-31。

本发明中，S2 中转录激活结构域（AD）的筛选酿酒酵母菌株，通过以下方式得到：构建 AD 的筛选转录因子表达载体，构建对应的报告基因表达载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合到酵母中，得到 AD 筛选酿酒酵母菌株。

进一步地，所述 AD 的筛选转录因子表达载体以 aTc 诱导的 ptet 作为启动子、tEN02 作为终止子、用 CCDB 替换转录因子基本架构 DBD-LBD-AD 中的 AD 结构域得到。

进一步地，所述 AD 的筛选转录因子表达载体的序列如 SEQ ID NO: 3。

进一步地，S2 中构建 AD 对应报告基因表达载体对应的报告基因用含有两个对应 DBD 调控的 operator 的启动子，控制黄色荧光蛋白 YFP 的表达，以 tEN02 作为终止子得到。

本发明中，所述筛选得到有效的 AD 序列如 SEQ ID NO: 32-40。

本发明中，筛选测试培养的具体过程如下：将构建的筛选酿酒酵母菌株接种

于对应缺陷型 SD 培养基中培养一段时间，转接至新的 SD 培养基，设置对照组和不同实验组，继续培养，取菌液，稀释后，检测每组荧光蛋白 YFP 表达量，计算相对荧光表达量。

本发明中，S3 中 DNA 结合结构域（DBD）的定量测试酿酒酵母菌株，通过以下方式得到：构建 DBD 的参数定量测试转录因子表达载体，然后构建对应的报告基因载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合到酵母中，得到 DBD 定量测试酿酒酵母菌株。

进一步地，所述 DBD 的参数定量测试转录因子表达载体以 xylose 诱导的 pxyluas 作为启动子、tEN02 作为终止子、用 CCDB 替换 DBD 结构域得到。

进一步地，所述 DBD 的参数定量测试转录因子表达载体序列如 SEQ ID NO: 4。

进一步地，S3 中构建 DBD 对应报告基因表达载体对应的报告基因用含有一个对应 DBD 调控的 operator 的启动子，控制黄色荧光蛋白 YFP 的表达，以 tEN02 作为终止子；将 CCDB 放置在报告基因质粒的启动子位置得到。

本发明中，S3 中配体结合结构域（LBD）的定量测试酿酒酵母菌株，通过以下方式得到：构建 LBD 的参数定量测试转录因子表达载体，然后构建对应的报告基因载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合到酵母中，得到 LBD 定量测试酿酒酵母菌株。

进一步地，所述 LBD 的参数定量测试转录因子表达载体以 xylose 诱导的 pxyluas 作为启动子、tEN02 作为终止子、用 CCDB 替换 LBD 结构域得到。

进一步地，所述 LBD 的参数定量测试转录因子表达载体如 SEQ ID NO: 5。

进一步地，S3 中构建 LBD 对应报告基因表达载体对应的报告基因用含有一个对应 DBD 调控的 operator 的启动子，控制黄色荧光蛋白 YFP 的表达，以 tEN02

作为终止子得到。

本发明中，定量测试培养的具体过程如下：将构建的筛选酿酒酵母菌株接种于对应缺陷型 SD 培养基中培养一段时间，第一次转接至新的 SD 培养基进行诱导范围测试，设置不同 xylose 浓度梯度，获得不同转录因子表达量，在不同 xylose 浓度下分别设置 ± 最大工作浓度诱导剂两组，获得每个转录因子表达量下的调控诱导范围；选取具有合适诱导范围的转录因子表达量，分别进行不同诱导剂浓度梯度下的诱导曲线的测试，含有对应 xylose 浓度的 SD 培养基中继续培养一段时间，设置对照组和不同实验组，第二次转接至含有对应 xylose 浓度下不同诱导剂浓度的 SD 培养基中培养，取菌液，稀释后，检测每组荧光蛋白 YFP 表达量，计算相对荧光表达量和转录因子的调控倍数。

进一步地，每个实验组的相对荧光表达量的计算方法为： $RPU = (YFP_{\text{测试}} - YFP_{\text{CYE72}}) / (YFP_{\text{CYE72/CY637}} - YFP_{\text{CYE72}})$ ；其中，CYE72 是阴性对照，CYE72/CY637 是 ptet 诱导对照。

转录因子的调控倍数的计算方法为： $\text{Fold Change} = RPU_{\text{诱导后}} / RPU_{\text{诱导前}}$ 。

本发明中，所述转录调控定量模型包括非核受体二聚化转录激活模型（不考虑入核过程）和核受体类二聚化转录激活模型（考虑入核过程）；

非核受体二聚化转录激活模型方程如下：

$$I(x) = \frac{c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1}) \cdot I_{\max} + I_0,$$

$$F_1 = -k_B T \left[\ln\left(\frac{L_{\text{tot}}}{2}\right) + \ln\left(\frac{\sqrt{(k_2 I + 1)^2 + 8 L_{\text{tot}} (k_2^2 k_3 I^2 + k_1)} - (k_2 I + 1)}{\sqrt{(k_2 I + 1)^2 + 8 L_{\text{tot}} (k_2^2 k_3 I^2 + k_1)} + (k_2 I + 1)}\right) \right] + \Delta \varepsilon_{DBD};$$

核受体类二聚化转录激活模型方程如下：

$$I(x) = \frac{c_{[LBD^*{}_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*{}_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD^*{}_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*{}_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1})^{-1} \cdot I_{\max} + I_0,$$

$$F_1 = \left[-k_B T \cdot \ln\left(\frac{L_{tot}}{2}\right) \right]$$

$$+ \left[-k_B T \ln\left(\frac{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8 L_{tot} (k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} - (k_2 I + k_* + k'_* k_2 I + 1)}}{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8 L_{tot} (k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} + (k_2 I + k_* + k'_* k_2 I + 1)}}\right) \right]$$

$$+ \Delta \varepsilon_{DBD}$$

其中，F1 因子为描述特定 DBD-LBD 组合下转录因子性能的参数；c 为 LBD 不同形态下的浓度， I_0 为特定序列启动子的本底泄露表达强度， $I_{\max}$ 为特定序列启动子的理论最大激活强度，I 为诱导剂浓度， L_{tot} 为所有形式的转录因子的表达量的单体形式的总和， $\Delta \varepsilon_{DBD}$ 描述 DBD 的行为，K1、K2、K3 描述 LBD 的行为， K_* 和 K'_* 描述转录因子的入核过程。

本发明具有以下有益效果：

(1) 本发明模块化正交真核转录因子模型构建方法，通过确定了模块化转录因子的基本架构，最终建立了转录调控定量模型，对转录因子模块化结构域进行参数化；通过定量模型可实现对转录因子进行模块化结构域参数优化，获得了性能优化的转录因子。

(2) 本发明确定了协同性二聚化过程引入转录调控过程对转录调控的优化效果，建立了各结构域不同参数范围的结构域库，建立了 DBD × operator、LBD × inducer 两个层面均相互正交的结构域库，为复杂基因调控线路的实现提供了工具。

附图说明

下面结合说明书附图和具体实施方式对本发明的技术方案做进一步说明。

图 1 是本发明模块化结构域转录因子的基本架构示意图；

图 2 是本发明构建的 DBD 结构域筛选测试载体和系统示意图；

图 3 是本发明构建的 LBD 结构域筛选测试载体和系统示意图；

图 4 是本发明构建的 AD 结构域筛选测试载体和系统示意图；

图5是本发明构建的DBD结构域定量测试载体；

图6是本发明构建的LBD结构域定量测试载体；

图7是本发明构建的转录因子定量测试载体；

图8是本发明有效DBD测试结果及正交性；

图9是本发明有效LBD测试结果；

图10是本发明有效AD测试结果；

图11是本发明非核受体类转录调控定量模型示意图；

图12是本发明定量测试结果；

图13是本发明LBD $\times$ inducer正交表；

图14是本发明crosstalk诱导曲线；

图15是本发明模型初步拟合的部分参数值；

图16是本发明模型指导优化情况。

具体实施方式

下述实施例中的实验方法，如无特殊说明，均为常规方法，按照本领域内的文献所描述的技术或条件或者按照产品说明书进行。下述实施例中所用的材料、试剂等，如无特殊说明，均可从商业途径得到。

下述实施例实例中所用的golden gate体系及条件如下：

试剂	用量/ μ l	反应条件	
BpiI or BsaI	0.5	37°C	2 mins
T4 ligase	0.5	16°C	3 mins
T4 ligase buffer	1	go to 1 for 25 times	
载体	a(50ng)	50°C	15 mins
片段	b(50ng)	80°C	20 mins
ddH ₂ O	up to 10		

下述实施例中酵母筛选测试培养条件如下：

得到酵母转化子后，将其接种于含500 μ l的对应缺陷型SD培养基的96深孔板中，并置于30°C、800 rpm培养24 h后，以1：200的比例转接至新的SD培养基96深孔板中。与此同时每个实验组设置非诱导--、添加诱导剂脱水四环素aTc+-（工作浓度：100 ng/ml）或最大工作浓度的对应诱导剂-+、或两者均添加++组。培养16 h后，取适量菌液，并用1xPBS溶液稀释后，使用流式分析仪器BD FACSCelesta™ FITC-A通道检测10,000个细胞的荧光蛋白YFP表达量。用FlowJo对获得的数据进行处理，取中位数作为每个样品孔的荧光值。以CYE72/CY671为表达量对照，每个实验组的相对荧光表达量的计算方法为： $RPU = (YFP_{\text{测试}} - YFP_{\text{CYE72}}) / (YFP_{\text{CYE72/CY637}} - YFP_{\text{CYE72}})$ 。

下述实施例中酵母定量测试培养条件如下：

得到酵母转化子后，将其接种于含500 μ l的对应缺陷型SD培养基的96深孔板中，并置于30°C、800 rpm培养24 h后，以1：200的比例第一次转接至新的SD培养基96深孔板中进行诱导范围测试：根据输入启动子诱导曲线设置不同xylose浓度梯度（0、0.2、0.3、0.4、0.5、0.6、0.8、0.9、1、1.5、2、5、10、20mM），以获得不同的转录因子表达量；进一步，在不同xylose浓度下分别设置±最大工

作浓度诱导剂两组。在该一系列设置实验组条件下,将96深孔板置于30℃、800 rpm培养16 h后,第一次取样测试。从而获得每个转录因子表达量下的调控诱导范围。

选取具有合适诱导范围的转录因子表达量(一般情况下设置0.4、0.6、1、2、10mM xylose浓度梯度),分别进行不同诱导剂浓度梯度下的诱导曲线的测试:第一次转接后,30℃、800 rpm培养24 h,分别选取对应浓度的+xylose组,作为母液,以1:200的比例第二次转接至新的SD培养基96深孔板中,同时分别设置这些不同xylose浓度下的不同诱导剂浓度梯度。30℃、800 rpm培养16 h后,第二次取样测试。

对于两次转接,每次转接设置空白对照组CYE72、组成型表达对照组CYE72/CY671;每次测试用输入启动子诱导型表达对照菌CYE72/LXR347进行对应xylose浓度下的输入启动子的表达量标定。对于两次取样测试,取样时,将适量菌液用1xPBS溶液稀释后,使用流式分析仪器BD FACSCelesta™ FITC-A通道检测10,000个细胞的荧光蛋白YFP表达量。用FlowJo对获得的数据进行处理,取中位数作为每个样品孔的荧光值。以CYE72/CY671为表达量对照,每个实验组的相对荧光表达量的计算方法为:
$$RPU = (YFP_{\text{测试}} - YFP_{\text{CYE72}}) / (YFP_{\text{CYE72/CY671}} - YFP_{\text{CYE72}})$$
。从而获得每个转录因子表达量下的调控诱导范围。

实例一:有效DNA结合结构域筛选

1. DNA结合结构域筛选载体LXR66构建

以实验室已有质粒pXJH119 (ptet-lexA_{ec1-87}-RpaR₁₋₁₇₉-VP16- tENO2) 为模板,用引物66-F(5' -GGTTCTAAGGATATCTCTGCTGGAGACATGA-3')和66-R(5' -CATTTTTTATTTATTTTGTAGCTTGATATTCTCTATCAC-3')扩增得到1号PCR片段。以实验室已有的质粒CY386(含有致死基因ccdB)为模板,用引物

Homo-CCDB-66-F(5'

-ACAAAAATAAATAAAAAATGAGGTCTTCTTATATTCCCCAGAACATCAGG

TTAATGG-3')和Homo-CCDB-66-R(5'

-GCAGAGATATCCTTAGAACCAGGTCTTCGGCTTACTAAAAGCCAGATAAC

AGTATGC-3')扩增得到2号PCR片段CCDB致死基因。使用诺维赞ClonExpress®

II One Step Cloning Kit C112试剂盒对2个PCR片段进行同源重组，获得酿酒酵母

转录激活因子DBD筛选载体LXR66，见SEQ ID NO: 1。

2. DNA结合结构域筛选系统构建

根据载体LXR66的序列，通过Golden Gate方法，用BpiI进行酶切，通过两个粘性末端的scar序列：AATG和GGTT，实现不同DBD的CCDB替换。对一系列来源于原核或病毒的转录因子根据其结构特征进行DBD的截取，获得其DBD的氨基酸和核苷酸序列（见SEQ ID NO: 7-18）。不同DBD序列通过基因合成、基因组序列或已有质粒序列获得，通过PCR过程在两端分别加上AATG和GGTT scar以及BpiI识别位点和保护碱基，通过BpiI Golden Gate的方法，构建含有不同DBD的转录因子质粒。对应的，以实验室已有质粒pXJH1为载体，通过分段长引物PCR和overlap PCR获得含有对应DBD调控operator的报告基因质粒。获得所需要的构建之后，通过用限制性内切酶BsaI酶切，得到带转化的酶切片段。利用zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株CYE72感受态，再将酶切片段分两步转至CYE72感受态中，得到酵母转化子。

包含酵母筛选测试系统所需DBD和operator的质粒如表1：

表 1 酵母筛选测试系统所需 DBD 和 operator 的质粒

DBD-operator	转录因子质	报告基因质	Operator序列
--------------	-------	-------	------------

	粒	粒	
lexAec ₁₋₈₇ -lexO	pXJH119	LXR152	TACTGTATGATCATACAGTA
lexAbs ₁₋₉₄ -lexObs	LXR69	LXR74	AGAACATATGTTCG
CI ₁₋₉₄ -O _{CI1}	LXR193	LXR204	TACCACTGGCAGTGGTA
CI434 ₁₋₇₀ -O _{CI434}	LXR194	LXR205	TACAAGAAAGTTTGTT
HKCI ₁₋₈₄ -O _{HKCI}	LXR195	LXR206	TGAACCATAAGTTCA
PurR ₁₋₆₀ -O _{PurR}	LXR196	LXR207	ACGCAAACGTTTGCGT
DeoR ₁₋₉₂ -O _{DeoR}	LXR197	LXR208	TGTTAGAATTCTAACA
lexAfs ₁₋₁₀₄ -lexOfs	LXR190	LXR201	CTGCACAAAGGTGCAC
lexAgs ₁₋₉₁ -lexOgs	LXR234	LXR237	GGTTGACACATGTCAACC
lexAmm ₁₋₁₁₅ -lexOmm	LXR235	LXR238	ACCTAATATTAATAAGGT
lexAxa ₁₋₁₀₁ -lexOxa	LXR236	LXR239	TTAGTAGTAATACTACTAA
RecApact ₁₋₁₀₄ -O _{RecApact}	LXR198	LXR209	TACTAATTAATTTAGT

3. 不同DNA结合结构域对转录调控的影响

获得酵母测试转化子后，通过酵母筛选测试培养得到测试菌液，用BD FACSCelesta™流式分析检测，每个DBD设置非诱导、诱导表达且诱导二聚两个组。筛选到的12个有效的DBD测试结果如图8。如图8所示，含不同operator的启动子存在不同程度的基础表达强度，在测试条件的表达量下激活强度也随不同DBD改变。

实例二：有效配体结合结构域筛选

1. 配体结合结构域筛选载体LXR77构建

以实验室已有质粒pXJH119 (ptet-lexAec₁₋₈₇-RpaR₁₋₁₇₉-VP16- tEN02) 为模板，用引物77-F (5' - GGTAGCCCTAAGAAAAAGAGAAAAGTGG-3')和77-R (5' - GATCATGAGCGGGCTTGCCCA-3')扩增得到1号PCR片段。以实验室已有的质粒CY386

为模板，用引物Homo-CCDB-77-F (5' - GGCCAAGCCCGCTCATGATCAGGTCTTCTTATATTCCCCAGAACATCAG-3') 和 Homo-CCDB-77-R (5' - CTCTTTTCTTAGGGCTACCAGGTCTTCGGCTTACTAAAAGCCAGATAAC-3') 扩增得到2号PCR片段CCDB致死基因。使用诺维赞ClonExpress® II One Step Cloning Kit C112试剂盒对2个PCR片段进行同源重组，获得酿酒酵母转录激活因子配体结合结构域筛选载体LXR77，见SEQ ID NO: 2。

2. 配体结合结构域筛选系统构建

根据载体LXR77的序列，通过Golden Gate方法，用BpiI进行酶切，通过两个粘性末端的scar序列：GATC和GGTA，实现不同LBD的CCDB替换。根据文献调研，将一系列存在化学诱导二聚化的蛋白进行配体结合-二聚化结构域的截取，获得其氨基酸和核苷酸序列（有效LBD序列见SEQ ID NO: 19-31）。不同LBD序列通过基因合成或已有质粒序列获得，通过PCR过程在两端分别加上GATC和GGTA scar以及BpiI识别位点和保护碱基，通过BpiI Golden Gate的方法，构建含有不同LBD的转录因子质粒。需说明的是，大部分LBD以LXR77为载体进行筛选构建，匹配lexAec₁₋₈₇作为DBD。但转录因子的整体表现同时受所有结构域影响，故对于部分LBD也在其他DBD情况下验证得到有效性。筛选获得LBD有效性的构建如表2：

表 2 筛选获得 LBD 有效性的构建情况

质粒编号	质粒包含转录因子	LBD对应诱导剂	筛选测试诱导剂工作浓度
LXR210	lexAec ₁₋₈₇ -acVHH-VP16	Caffeine	100uM
pXJH119	lexAec ₁₋₈₇ -RpaR ₁₋₁₇₉ -VP16	pC-HSL	100uM
LXR111	lexAec ₁₋₈₇ -BjaR ₂₋₁₈₀ -VP16	IV-HSL	100uM

LXR150	lexA _{ec1-87} -CinRori ₂₋₁₇₉ -VP16	3OC14-HSL	10uM
LXR114	lexA _{ec1-87} -LasR ₂₋₁₇₇ -VP16	3OC12-HSL	100uM
LXR333	lexA _{ec1-87} -TraR ₁₋₁₇₄ -VP16	3OC8-HSL	100uM
LXR334	lexA _{ec1-87} -CarRecc ₁₋₁₆₉ -VP16	3OC6-HSL	10uM
LXR384	CI ₁₋₉₄ -SmaR ₁₋₁₇₉ -VP16	C4-HSL	100uM
LXR271	lexA _{ec1-87} -ER ₂₈₂₋₅₉₅ -VP16	β -estradiol	10uM
LXR272	lexA _{ec1-87} -DHBR ₂₈₂₋₅₉₅ -VP16	DHB	100uM
LXR402	lexA _{bs1-94} -PR-VP16	RU486	100uM
LXR404	lexA _{bs1-94} -GR ₄₈₇₋₇₇₇ -VP16	dexamethasone	1mM
LXR405	lexA _{bs1-94} -MR ₆₆₉₋₉₈₄ -VP16	aldosterone	10uM

获得所需要的构建之后，通过用限制性内切酶BsaI酶切，得到带转化的酶切片段。利用zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株CYE72/LXR152感受态，再将酶切片段进行转化，得到酵母转化子。

3. 不同配体结合结构域对转录调控的影响

获得酵母测试转化子后，通过酵母筛选测试培养得到测试菌液，用BD FACSCelesta™流式分析检测，每个LBD设置非诱导、诱导表达、诱导表达且诱导二聚三个组。筛选到的13个有效的LBD测试结果如图9。如图9所示，LBD主要来自三类蛋白：单结构域抗体、原核群体感应转录调控蛋白和哺乳动物核受体。在测试条件的表达量下，不同LBD会导致不同程度的转录因子本底二聚化带来的本底激活强度，不同LBD对应不同的诱导剂浓度，其分别在其诱导剂最大工作浓度下不同LBD能达到的最大激活强度也不同。

实例三：有效转录激活结构域筛选

1. 转录激活结构域筛选载体LXR233构建

以实验室已有质粒pXJH119 (ptet-lexA_{ec1-87}-RpaR₁₋₁₇₉-VP16- tENO2) 为模板, 用引物233-F(5' - TAAAAGCTTTTGATTAAGCCTTCTAGTCCAAAAAAC-3') 和233-R(5' - CACTTTTCTCTTTTTCTTAGGGCTACCATTAC-3') 扩增得到1号PCR片段。以实验室已有的质粒CY386为模板, 用引物Homo-GGS5-CCDB-233-F (5' - CTAAGAAAAAGAGAAAAGTGGGTGGCAGTGGCGGAAGCGGGGGATCAGG TGGTTCTGGAGGGTCCAGGTCTTCTTATATTCCCCAGAACATCAGGTTAAT GG-3') 和Homo-CCDB-233-R (5' - GGCTTAATCAAAAGCTTTTAAGGTCTTCGGCTTACTAAAAGCCAGATAACA GTATGCA-3') 扩增得到2号PCR片段CCDB致死基因。使用诺维赞ClonExpress® II One Step Cloning Kit C112试剂盒对2个PCR片段进行同源重组, 获得酿酒酵母转录激活因子配体结合结构域筛选载体LXR233, 见SEQ ID NO: 3。

2. 转录激活结构域筛选系统构建

通过文献调研, Ariel Erijman等2020年提出一个AD结构和功能预测模型, 预测了一系列具有转录激活功能的氨基酸序列。我们从中挑选出7个预测功能良好的转录激活氨基酸序列, 并且对一些常用的酵母转录激活结构域进行测试, 以获取多个匹配多个转录因子的AD。不同AD序列通过基因合成或已有质粒序列获得, 通过PCR过程在两端分别加上与载体LXR233匹配的GTCC和TAAA scar以及BpiI识别位点和保护碱基, 通过BpiI Golden Gate的方法。

获得所需要的构建之后, 通过用限制性内切酶BsaI酶切, 得到带转化的酶切片段。利用zyzo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株CYE72/LXR152感受态, 再将酶切片段进行转化, 得到酵母转化子。

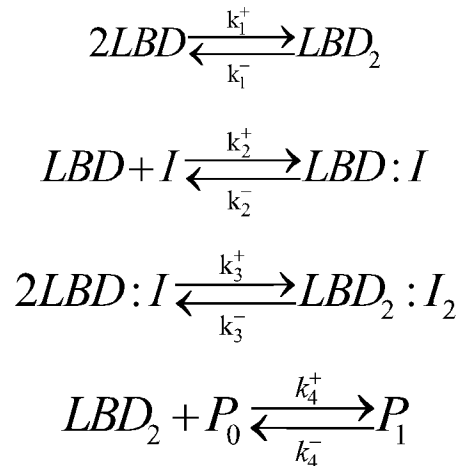
3. 不同转录激活结构域对转录调控的影响

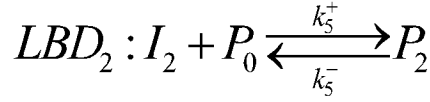
获得酵母测试转化子后，通过酵母筛选测试培养得到测试菌液，用BD FACSCelesta™流式分析检测，每个DBD设置非诱导、诱导表达且诱导二聚两个组。结果如图10，从图10中结果可以看出，不同AD对转录因子的最大激活强度具有不同程度的影响。从筛选的13个AD中，其中有9个有明显的转录激活效果（SEQ ID NO: 32-40），其中在VP16的情况下最大激活强度最强。

实例四：激活型转录调控模型构建

1. 非核受体二聚化转录激活模型（不考虑入核过程）

本方法中的转录调控过程可描述如下（参数见图11）：①输入启动子以一定表达量表达转录因子单体；②转录因子单体可以自发或受到诱导剂作用两种形式不同程度地形成转录因子二聚体；③转录因子二聚体以一定地亲和力识别并结合启动子上的operator序列。用LBD表示转录因子单体蛋白，LBD₂表示两个转录因子单体蛋白形成的二聚体，I表示诱导剂分子，LBD:I表示单体蛋白和诱导剂分子结合的中间形态，LBD₂:I₂表示两个蛋白单体和两个诱导剂分子形成的二聚体，P₀表示非活性状态启动子，P₁和P₂分别表示结合不同途径形成的二聚体的活性状态启动子；该过程可用以下方程表示：





调控过程中，启动子可能状态包括：①空启动子；②只结合RNAP；③同时结合RNAP和转录因子本底二聚体；④同时结合RNAP和转录因子诱导二聚体。③和④分别对应启动子的不同活性状态，根据配分方程，可得到模型方程描述如下：

$$I(x) = \frac{c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1}) \cdot I_{\max} + I_0$$

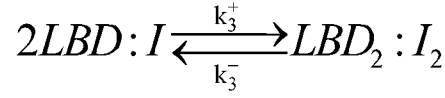
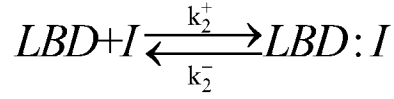
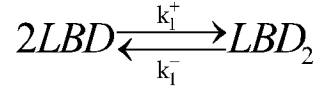
$$F_1 = -k_B T \left[\ln\left(\frac{L_{tot}}{2}\right) + \ln\left(\frac{\sqrt{(k_2 I + 1)^2 + 8 L_{tot} (k_2^2 k_3 I^2 + k_1)} - (k_2 I + 1)}{\sqrt{(k_2 I + 1)^2 + 8 L_{tot} (k_2^2 k_3 I^2 + k_1)} + (k_2 I + 1)}\right) \right] + \Delta \varepsilon_{DBD}$$

其中，F1 因子为描述特定 DBD-LBD 组合下转录因子性能的参数；c 为 LBD 不同形态下的浓度， I_0 为特定序列启动子的本底泄露表达强度， $I_{\max}$ 为特定序列启动子的理论最大激活强度，I 为诱导剂浓度， L_{tot} 为所有形式的转录因子的表达量的单体形式的总和， $\Delta \varepsilon_{DBD}$ （K4、K5）描述 DBD 的行为，K1、K2、K3 描述 LBD 的行为。

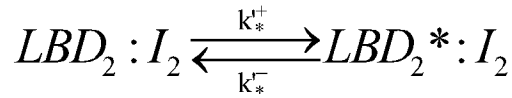
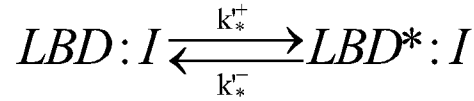
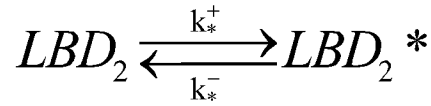
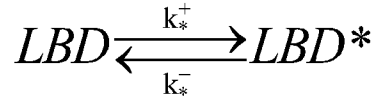
2. 核受体类二聚化转录激活模型（考虑入核过程）

核受体类转录调控过程增加了各类转录因子蛋白的入核过程，用LBD、LBD*分别表示转录因子单体蛋白核外、核内状态，LBD₂、LBD₂*分别表示两个转录因子单体蛋白形成的二聚体核外、核内状态，I表示诱导剂分子，LBD：I、LBD*：I分别表示单体蛋白和诱导剂分子在核外、核内结合的中间形态，LBD₂：I₂、LBD₂*：I₂表示两个蛋白单体和两个诱导剂分子形成的二聚体核外、核内状态，该过程可用以下方程表示：

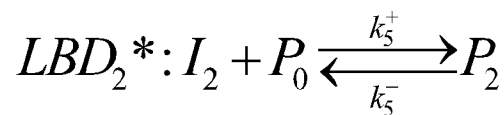
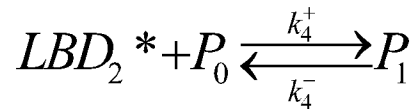
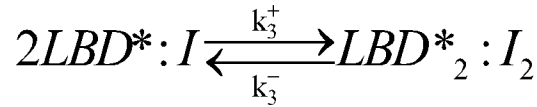
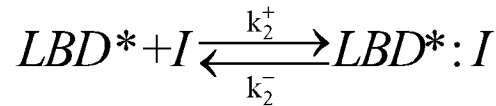
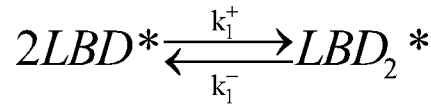
核外：



入核:



核内:



根据配分方程，可得到模型方程描述如下：

$$I(x) = \frac{c_{[LBD^*{}_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*{}_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD^*{}_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*{}_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1})^{-1} \cdot I_{\max} + I_0,$$

$$F_1 = \left[-k_B T \cdot \ln\left(\frac{L_{tot}}{2}\right) \right] + \left[-k_B T \ln\left(\frac{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8 L_{tot} (k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} - (k_2 I + k_* + k'_* k_2 I + 1)}}{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8 L_{tot} (k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} + (k_2 I + k_* + k'_* k_2 I + 1)}}\right) \right] + \Delta \varepsilon_{DBD};$$

其中，F1因子为描述特定DBD-LBD组合下转录因子性能的参数；c为LBD不同形态下的浓度，I₀为特定序列启动子的本底泄露表达强度，I_{max}为特定序列启动子的理论最大激活强度，I为诱导剂浓度，L_{tot}为所有形式的转录因子的表达量的单体形式的总和， $\Delta \varepsilon_{DBD}$ （K4、K5）描述DBD的行为，K1、K2、K3描述LBD的行为，K*和K*'描述转录因子的入核过程。

实例五：结构域参数定量测试

1. DNA结合结构域参数定量测试载体LXR322构建

以已构建质粒LXR319（pxyluas-lexA_{ec1-87}-CarHc-VP16- tEN02）为模板，用引物322-F（5' - CAGTGAGAAGACCTGTAGCCCTAAGAAAAAGAGAAAAGTGGG-3'）和322-R（5' - CAGTACGAAGACTACATTTTTTATTTATTTTGTAGCTTGATATTCTAGTTTGTGG-3'）扩增得到1号PCR片段。以实验室已有质粒pXJH119为模板，用引物RpaR-322-F（5' - CAGTGAGAAGACCTGGTTCTAAGGATATCATTGTGGGTGAAGATCAGCTGTGG-3'）和RpaR-322-R（5' - TAAGCCGAAGACCTCTACCATTACGACGAATCGGTTTCG-3'）扩增得到2号PCR片段，RpaR₁₋₁₇₉ LBD结构域。以实验室已有的质粒CY386为模板，用引物CCDB-322-F（5' - CAGTGAAATGAGGTCTTCTTATATTCCCCAGAACATCAG-3'）和

CCDB-322-R (5' - TAAGCCAACCAGGTCTTCGGCTTACTAAAAGCCAGATAAC-3') 扩增得到3号PCR片段CCDB致死基因。通过BpiI限制性内切酶的Golden Gate技术, 获得酿酒酵母转录激活因子DNA结合结构域参数定量测试载体LXR322, 见SEQ ID NO: 4。

2. DNA结合结构域参数定量测试系统构建

将筛选到的12个DBD通过PCR扩增在两端加上与载体LXR322匹配的AATG和GGTT scar以及BpiI识别位点和保护碱基, 通过BpiI Golden Gate的方法, 构建用于不同DBD参数定量测试的质粒。为了简单化参数测试, 调控启动子均只包含一个operator, 以lexAec₁₋₈₇对应的lex0为例, 启动子序列为:

acaccgggcataactgtatatacacccaggggtctaggTACTGTATGATCATACAGTAaataagagaatttcagattgagagaat**gaaaaaaaaaaaaaaaaa**aggcagtccttgacaatagttatatcacgttgactcaggaagtgggtggaggactgccacat**tataa**tagagtgccagtagcgacttttttcacactcgaaatactcttactactgctctcttgttgtttttatcacttcttgtttcttcttggtaaat**agaatatcaagctacaaaaataaataaaaa**, 加粗字体位置为启动子的polyA、TATA box、TSS和kozak序列, 大写非加粗位置为lex0。通过在此位置替换不同operator, 构建一系列调控启动子质粒。通过长引物PCR和overlap PCR获得两端带有酶切位点的启动子片段, 以实验室已有质粒pXJH1为载体, 通过BpiI Golden Gate的方法, 获得不同DBD定量测试对应的报告基因质粒。用于DBD定量测试的质粒信息如表3:

表 3 用于 DBD 定量测试的质粒信息

质粒编号	所包含的转录因子	质粒编号	所包含的operator
LXR335	CI434 ₁₋₇₀ -RpaR ₁₋₁₇₉ -VP16	LXR352	lexO
LXR327	lexAbs ₁₋₉₄ -RpaR ₁₋₁₇₉ -VP16	LXR349	lexObs
LXR337	lexAmm ₁₋₁₁₅ -RpaR ₁₋₁₇₉ -VP16	LXR353	O _{CI L1}
LXR338	lexAec ₁₋₈₇ -RpaR ₁₋₁₇₉ -VP16	LXR348	O _{CI434}

LXR339	lexA _{α1-101} -RpaR ₁₋₁₇₉ -VP16	LXR354	O _{HKCI}
LXR340	HKCI ₁₋₈₄ -RpaR ₁₋₁₇₉ -VP16	LXR355	O _{PurR}
LXR341	RecA _{pact1-104} -RpaR ₁₋₁₇₉ -VP16	LXR356	O _{DeoR}
LXR342	DeoR ₁₋₉₂ -RpaR ₁₋₁₇₉ -VP16	LXR357	lexO _{fs}
LXR343	PurR ₁₋₆₀ -RpaR ₁₋₁₇₉ -VP16	LXR358	lexO _{gs}
LXR344	CI ₁₋₉₄ -RpaR ₁₋₁₇₉ -VP16	LXR359	lexO _{mm}
LXR345	lexA _{gs1-91} -RpaR ₁₋₁₇₉ -VP16	LXR360	lexO _{xa}
LXR346	lexA _{fs1-104} -RpaR ₁₋₁₇₉ -VP16	LXR361	O _{RecA_{pact}}

获得所需要的构建之后，通过用限制性内切酶BsaI酶切，得到带转化的酶切片段。利用zyzo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株CYE72感受态，再将酶切片段分两步转至CYE72感受态中，得到酵母转化子。

3. 配体结合结构域参数定量测试载体LXR321构建

以已构建质粒LXR319 (pxyluas-lexA_{ec1-87}-CarHc-VP16-tEN02) 为模板，用引物321-F (5' - CAGTGAGAAGACCTGGTAGCCCTAAGAAAAAGAGAAAAGTGG-3') 和321-R (5' - CAGTACGAAGACTACATTTTTTATTTATTTTTGTAGCTTGATATTCTAGTTTGTG-3') 扩增得到1号PCR片段。以已构建质粒LXR36为模板，用引物lexA_{bs1-94}-321-F (5' - CAGTGAGAAGACCTAATGTCTACGAAGCTATCAAAAAGGCAAC-3') 和lexA_{bs1-94}-321-R (5' - TAAGCCGAAGACCTGATATCCTTAGAACCAGGAGATCCCGCCGTGACTTTC-3') 扩增得到2号PCR片段，lexA_{bs1-94} DBD结构域。以实验室已有的质粒CY386为模板，用引物CCDB-321-F (5' - CAGTGATATCAGGTCTTCTTATATTTCCCAGAACATCAG-3') 和CCDB-321-R (5' - TAAGCCTACCAGGTCTTCGGCTTACTAAAAGCCAGATAAC-3') 扩增得到3号PCR片段CCDB致死基因。通过BpiI限制性内切酶的Golden Gate技术，获得酿酒酵母转录激活因子配体结合结构域参数定量测试载体LXR321，见SEQ ID NO: 5。

4. 配体结合结构域参数定量测试系统构建

将筛选到的13个LBD通过PCR扩增在两端加上与载体LXR321匹配的TATC和GGTA scar以及BpiI识别位点和保护碱基，通过BpiI Golden Gate的方法，构建用于不同LBD参数定量测试的质粒。需说明的是，大部分LBD以LXR321为载体进行定量测试构建，匹配lexAb_{S1-94}作为DBD。但考虑到不同DBD对于诱导曲线的影响，部分LBD也在其他DBD情况下进行定量测试。用于LBD定量测试的质粒信息如表4：

表 4 用于 LBD 定量测试的质粒信息

质粒编号	所包含的转录因子	最低诱导剂浓度	最高诱导剂浓度
LXR325	lexAb _{S1-94} -acVHH-VP16	0.1nM	1uM
LXR326	lexAb _{S1-94} -LasR ₂₋₁₇₇ -VP16	1nM	10uM
LXR327	lexAb _{S1-94} -RpaR ₁₋₁₇₉ -VP16	0.1uM	1mM
LXR328	lexAb _{S1-94} -CinRori ₂₋₁₇₉ -VP16	0.1nM	1uM
LXR329	lexAb _{S1-94} -BjaR ₂₋₁₈₀ -VP16	1uM	1mM
LXR330	lexAb _{S1-94} -ER ₂₈₂₋₅₉₅ -VP16	1nM	1uM
LXR331	lexAb _{S1-94} -DHBR ₂₈₂₋₅₉₅ -VP16	10nM	100uM
LXR334	lexAec ₁₋₈₇ -CarRecc ₁₋₁₆₉ -VP16	1nM	10uM
LXR371	lexAb _{S1-94} -TraR ₁₋₁₇₄ -VP16	0.1uM	100uM
LXR384	CI ₁₋₉₄ -SmaR ₁₋₁₇₉ -VP16	10nM	100uM
LXR402	lexAb _{S1-94} -PR-VP16	10nM	100uM
LXR404	lexAb _{S1-94} -GR ₄₈₇₋₇₇₇ -VP16	1uM	1mM
LXR405	lexAb _{S1-94} -MR ₆₆₉₋₉₈₄ -VP16	1nM	10uM

获得所需要的构建之后，通过用限制性内切酶BsaI酶切，得到带转化的酶切片段。利用zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备菌株CYE72/LXR349感受态，并将酶切片段转化入感受态，得到酵母转化子。

5. 结构域参数测量和拟合

本方法中主要对DBD和LBD两个结构域的对应参数 $\Delta\epsilon_{DBD}$ 和K1 K2 K3进行定量测试。获得酵母测试转化子后，通过酵母定量测试培养方法得到测试菌液，用BD FACSCelesta™流式分析检测，用FlowJo进行数据处理，测试数据结果如图12。将获得的测试数据使用模型进行拟合，初步获得不同DBD和LBD的对应参数。

实例六：配体结合结构域结合配体正交性测试

1. 正交表测量

选择除了PR、GR、MR之外的其他LBD的有效转录因子构建，用酵母定量测试培养方法获取酵母菌液。设置RPU=0.67和RPU=1.24两个转录因子表达量，分别在两个表达量下对所有LBD和诱导剂进行全排列正交性测试。其中，在添加诱导剂时，对每个LBD设置所有诱导剂最大工作浓度的诱导组和非诱导表达组，从而获得特定表达量下每个LBD响应不同诱导剂的诱导范围和调控倍数。用于正交表测试的质粒信息和诱导剂信息如表5：

表 5 用于正交表测试的质粒信息和诱导剂信息

质粒编号	所包含的转录因子	天然诱导剂	诱导剂浓度
LXR386	HKCI ₁₋₈₄ -acVHH-VP16	Caffeine	1uM
LXR330	lexA _{bs1-94} -ER ₂₈₂₋₅₉₅ -VP16	β -estradiol	1uM
LXR331	lexA _{bs1-94} -DHBR ₂₈₂₋₅₉₅ -VP16	DHB	100uM
LXR327	lexA _{bs1-94} -RpaR ₁₋₁₇₉ -VP16	pC-HSL	100uM
LXR329	lexA _{bs1-94} -BjaR ₂₋₁₈₀ -VP16	IV-HSL	100uM
LXR381	lexA _{mm1-115} -CinRori ₂₋₁₇₉ -VP16	3OC14-HSL	1uM
LXR326	lexA _{bs1-94} -LasR ₂₋₁₇₇ -VP16	3OC12-HSL	10uM
LXR371	lexA _{bs1-94} -TraR ₁₋₁₇₄ -VP16	3OC8-HSL	100uM
LXR334	lexA _{ec1-87} -CarRecc ₁₋₁₆₉ -VP16	3OC6-HSL	1uM

LXR384	CI ₁₋₉₄ -SmaR ₁₋₁₇₉ -VP16	C4-HSL	100uM
--------	---	--------	-------

分别在两个表达量下，用测试所获得的所有LBD对所有诱导剂响应的调控倍数取对数，制作10×10的正交表，如图13。正交表每个格子的颜色对应调控倍数的对数的数值，数值越大，颜色越深。仅有对角线深色表示无交互作用的完全正交的情况。如图13所示，crosstalk存在于本方法所选的LBD和诱导剂之中，包括弱于天然响应的可优化crosstalk（如CinRori₂₋₁₇₉对30C12-HSL）和强于天然响应的不可优化crosstalk（如TraR₁₋₁₇₄对30C12-HSL）两种。

2. Crosstalk诱导曲线测试和参数拟合

对于正交表测试得到的crosstalk的可优化部分，通过酵母定量测试培养方法，获得对应的LBD对诱导剂的诱导曲线，如图14。该测试结果提供对应LBD对于诱导剂的响应调控参数K2'、K3'的拟合数据。

实例七：模型指导理性优化

1. 结构域模块化转录因子表达载体LXR370构建

以已构建质粒LXR319（pxyluas-lexA_{ec1-87}-CarHc-VP16- tEN02）为模板，用引物370-F（5'-CAGTGAGAAGACCTGGTAGCCCTAAGAAAAAGAGAAAAGTGG-3'）和370-R（5'-CAGTACGAAGACTACATTTTTTATTTATTTTGTAGCTTGATATTCTAGTTTGTG-3'）扩增得到1号PCR片段。以实验室已有的质粒CY386为模板，用引物CCDB-370-F（5'-CAGTGAAATGAGGTCTTCTTATATTTCCCCAGAACATCAG-3'）和CCDB-370-R（5'-TAAGCCTACCAGGTCTTCGGCTTACTAAAAGCCAGATAAC-3'）扩增得到2号PCR片段CCDB致死基因。通过BpiI限制性内切酶的Golden Gate技术，获得酵母转录激活因子配体结合结构域参数定量测试载体LXR370，见SEQ ID NO: 6。

2. 模型验证

挑选模型输出中14个不同转录因子及其对应表达量，进行模型输出的实验测试验证。以LXR370为载体，分别通过PCR获得两端带有相应scar和linker的DBD作为1号PCR片段、LBD作为2号PCR片段，通过Bpi I Golden Gate的方法，获得以VP16作为AD、不同DBD和LBD组合情况下的酵母转录激活因子构建，用于组合测试和参数化模型验证。获得相关转录因子质粒后，通过用限制性内切酶BsaI酶切，得到带转化的酶切片段。利用zymo Frozen-EZ Yeast Transformation II Kit™ 酵母转化试剂盒制备含有对应DBD调控启动子的酵母感受态，并将酶切片段转化入感受态，得到酵母转化子。模型验证组合测试相关的质粒信息、转化感受态以及转录因子表达量如表6：

表6 模型验证组合测试相关的质粒信息、转化感受态以及转录因子表达量

质粒编号	所包含的转录因子	转录因子表达量
LXR373	lexAxa1-101-LasR2-177-VP16	1.88
LXR374	lexAfs1-104-LasR2-177-VP16	4.34
LXR375	RecApact1-104-DHBR282-595-VP16	3.08
LXR376	lexAgS1-91-DHBR282-595-VP16	6.94
LXR377	PurR1-60-ER282-595-VP16	3.08
LXR378	DeoR1-92-ER282-595-VP16	6.94
LXR379	CI4341-70-BjaR2-180-VP16	1.88
LXR380	lexAxa1-101-BjaR2-180-VP16	6.94
LXR381	lexAmm1-115-CinRori2-179-VP16	0.35
LXR382	lexAec1-87-CinRori2-179-VP16	1.57
LXR383	CI4341-70-SmaR1-179-VP16	0.99
LXR384	CI1-94-SmaR1-179-VP16	0.99
LXR385	lexAmm1-115-acVHH-VP16	0.21
LXR386	HKCI1-84-acVHH-VP16	0.35

获得酵母转化子后，通过酵母定量测试培养方法获取酵母菌液，在模型给出的表达量下进行诱导曲线的测试，随后于模型理论输出结果进行对比，部分拟合结果如图15。如图15所示，实验数据基本能够符合理论预测， $R^2 > 0.84$ ，模型的预测输出准确性较好。

3. 模型指导优化

模型描述了酿酒酵母中特定转录激活因子的转录调控性能，如图16a所示。一般来说，具体表现为：①特定DBD对应的operator的序列作为启动子序列的一部分，决定了启动子确定的理论上的最小泄露和最大转录激活强度；②LBD的本底二聚化程度 k_1 决定了特定转录因子表达量下由本底二聚化造成的泄露程度，即该表达量下诱导曲线的最小值；③LBD的诱导二聚化能力 k_2 和 k_3 决定了其转录调控的敏感性和其能够实现的转录激活强度，即曲线的EC50和最大激活；总的来说LBD的参数确定了曲线形状；④DBD与operator结合的亲和力 k_4 和 k_5 和转录因子表达量则使得诱导曲线在坐标系中平移，亲和力越强、转录因子表达量越高，诱导曲线越倾向于朝左上移动。据此，对于响应特定诱导剂小分子的特定LBD，我们可以通过改变DBD-operator和转录因子表达量来实现对于本底泄露表达强度、激活表达强度、正交性等的优化。如图16b，在相同的转录因子表达量下，以LasR₂₋₁₇₇为LBD，更换不同DBD，可以获得本底二聚泄露表达、诱导激活表达不同的转录因子，从而可以实现诱导范围的优化，在lexA_{x_{a1-101}} DBD下实现最优调控。另外，如图16c和d，ER₂₈₂₋₅₉₅能够响应自身诱导剂 β -estradiol，同时对DHBR₂₈₂₋₅₉₅的诱导剂DHB存在交互作用。可以通过改变表达量来使得ER₂₈₂₋₅₉₅对两种诱导剂的响应在一个相对正交的范围内（激活强度相差>10倍），在lexA_{b_{S1-94}} DBD情况下，可以通过将表达量控制在0.1~0.5 RPU范围内，实现低本底、高激活、相对正交。进一步替换

DBD进行优化，在DeoR₁₋₉₂ DBD情况下，可以实现在一个比较合适的转录因子表达量下（1~7 RPU）实现对自身诱导剂诱导范围良好、并且相互正交的情况。由此可以实现模型指导下，对于转录因子调控的泄露、本底表达、激活强度、动力学范围、对诱导剂敏感性以及多转录因子正交性等多指标的优化。

序列：

SEQ ID NO: 1 LXR66:

```
gggtctcgaccagaagaacagtaaaataaagcaaggtagctgaaattaatatttttaaatgggtctaaccgatgccgaagaactgcgcagtcgggtataacgtct
gacatgtcctttttgatttggaatccaaccactcaagtgactctgttcatttactttgcgaaaaatatacccacaaattgcccatcgaaagtgaatcgaaaccacctcaga
ctggcaccgacaagcaagattatacagacagagtactttatactacgtacgttaagtcgaagcaagggtttcttatttactgaacgggtaagagatctgggccggct
tgccaagatgcaaacgaataagatttcaaagttgcatttgccttagccgtctgacaccattggctatttggatatttatattgacttctgtgtacattgatcacatcgactgt
tctattggcaaatgaaccacgggcattgactattttcaggttactactatatattatcatcacgggcaaggattgtagcttgcctatgctgcgtcgaggagatatttactttat
tattctagtttttacagttatttattaattaatttttatatgcacgcaataaaaagctctatatttaagttctttatttattaatacattttcctctacgagctgtcaccggatgtgctt
tcgggtctgatgagtcggtgaggacgaacagcctctacaaataattttgttaagagcagggtgttcctatggccgtgctgatggtggggggcgtcgggcggtgaaaccg
gggttcggagcggcagggggtttttgttttttttttcggtataaaaatccctatcagtgatagagattttcttttttttttttccctatcagtgatagagaatatcaagct
acaaaaataataaaaaATGAGGTCTTCTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGGTGTGCTTTT
CGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAAACCGGCACACTGGCCATATCGGTGGT
CATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACA
GCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCCGGGCGTGTCAATAATCACTCTGTAC
ATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTAACTGCATTTACACAGCCC
CTGTTCTCGTCAGCAAAAGAGCCGTTTCAATTAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATT
TTCCGCTTTCCAGCGTTTCGGCACGCAGACGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGG
AGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTGCTGTCAACTGTCACTGTAATACGCT
GCTTCATAGCATACCTCTTTTTGACATACTTCGGGTATACATATCAGTATATATTCTTATACCGCAAAA
ATCAGCGCGCAAATATGCATACTGTTATCTGGCTTTTAGTAAGCCGAAGACCTgggtctaaggatatctctgtggag
acatgagagctgccaacctttggccaagcccgtcatgatcATTGTGGGTGAAGATCAGCTGTGGGGTCGTCGTACACTGGAA
TTTGTTGATAGCGTTGAACGTCTGGAAGACCGGCACTGATTAGCCGTTTTGAAAGCCTGATTGCAAG
CTGTGGTTTTACCGCCTATATCATGGCAGGTCTGCCGAGCCGTAATGCCGGTCTGCCGGAAGTACCC
TGGCAAATGGTTGGCCTCGTGATTGGTTTGATCTGTATGTTAGCGAAAACCTTAGCGCAGTTGATCCG
GTTCCGCGTTATGGTGCAACCACTGTTTCATCCGTTTGTGGAGTGATGCACCGTATGATCGTGACCGT
GATCAGGCAGCACATCGTGTTATGACCCGTGCAGCAGAATTTGGTCTGGTTGAAGGTTATTGTATTCC
GCTGCATTACGATGATGGTAGCGCAGCAATTAGTATGGCAGGTGAAGATCCTGATCTGAGTCCGGCA
GCCCCGTGGTGTAAATGCAGCTGGTTAGCATTTATGCACATAGCCGTCTGCGTGTACTGAGCCGTCCGAA
ACCGATTTCGTCTGTAATggtagccctaagaaaaagagaaaagtggtggcagtgccggaagcgggggatcagggtggttctggaggggtccGATAT
CATGCAGGATCTGCCGGGCAACGATAACAGCACCGCGGGCGCGCGTACGAAAAACAATTACGGGTCT
ACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCGAAGAGGCGGGGCTGGCGGCTCCGC
```

GCCTGTCCTTTCTCCCCGCGGGACACACGCGCAGACTGTCGACGGCCCCCCCCGACCGATGTCAGCCTG
GGGGACGAATTGCACTTAGACGGCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAGACGATTTTCG
ATCTGGACATGTTGGGGGACGGGGATTCCCCGGGTCCGGGATTTACCCCCACGACTCCGCCCCCTAC
GGCGCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGAATTGACGAGTA
CGGTGGGAAGCTTTAAtaaaagcttttgattaagccttctagtccaaaaacacgtttttgtcatttatttcttagaatagttagttattcattttatagt
cacgaatgtttatgattctatatagggttgcaacaagcatttttcatttatgttaaaacaatttcaggttacccttttctgcttgggtgacgcgtgtatccgccgcctttt
ggtcacccatgtatttaattgcataaataattcttaaaagtgagctagtctatttctattacatacctctcatttctcatttctcctgaccagctgagtttcatggcttgaattact
atgtgcggggtaattgtgctgtcgtagccttggtgaagtatttggcacaaggctctaggccctgttgaggagtgagcttgccttgcctccgccgggtcacccggcca
gagacatggaggccagaataccctccttgacagcttgacgtgagcagctcaggggcatgatgtgactgtcgtccgtacatttagccatacatcccatgtataatca
ttgcatccatacattttgatggccgcacggcggaagcaaaaattacggctcctcgtgcagacctgcgagcagggaacgctccccctcacagacgcgttgaattgtc
cccacgccgcgccctgtagagaaataaaaaggttaggttggcactgaggttcttctatatacttcttttaaaatctttaggatacagttctcatcacatccga
acataaacaacaatgtctaagaatactgttgcctaccgggtgatcagctcggtaagaagttactgacgaagctattaaggcttgaatgccattgtgaagtcctccag
aaattaagttaattccaacatcacttgatcgggggtgtgctcctgatgccactggcactcctttaccagatgaagctctagaagccttaagaaagccgatgtgtctt
actaggtgctgttggtggtcctcaaatgggtgacggcgagcttagaccagaacaaggctctattgaagatcagaaggaattgggtctatacgccaacttgaggccatgt
aacttgcctctgattcttactagatcttctccttgaagcctgaatatgcaaaagggtaccgatttctgctgcttagagaattgggtggttatctacttgggtgaaagaaa
gaagatgaaggtgacggagttgcttgggtctgagaaatacagtttctgaagtcaagaattacaagaatggctgcttcttggcattgcaacaaaaccaccatta
ccaatctggtcacttgacaaggctaactgttgcctcttccagattgtggagaaagactgtgaagaaacctcaagactgagttcccacaattaactgttcagccaat
tgatcgattctgctgctatgatttgggttaaatccaactaagctaaacgggtgtgttattaccaacaacatgttgggtatattatctccagtaagcctctgttattccaggtt
cttgggttattacccttctgcatctctagcttccctacctgacactaacaaggcattcgggtttgacgaacctgtcatgggtctgcccagattaccagcaacaaggtaa
cccaattgctaccatcttatctgacgctatgatgttgaagttatccttggatttgggtgaagaaggtagggctcttgaagaagctgttagaatgtcttggatgcaggtgcag
aaccggtagccttgggtgttctaaacttaccactgaggttggcgatgctatcgccaaggctgtcaaggaaatcttggcttaaacagtagtacaataaaaagattcttgttt
caagaactgtcattttagatgttttttatgtttagtttctatttaataaatgttagcgtgatttatatttttgcctcgacatcatctgccagatgcgaagttaagtgcg
cagaaagtaatatcatgcgtcaatcgtatgtgaatgctggctgctatactgctgctgattcgatactaacgccgccatccagtgctgacctccattactcgatcgcgtgcg
gtgcgacacgaaattacaaaatggaatatgtcatagggtagacgaaactatatacgcaatctacatacatttatcaagaaggagaaaaaggaggatgtaaaggataca
ggtaagcaaatgataactaatggctcaacgtgataaggaaaaagaattgcacttaacattaatattgacaaggaggagggcaccacacaaaaagttagggtgaacagaa
aatcatgaaactatgattcctaatttatattggaggattttcttaaaaaaaaaaaaaatacaacaataaaaaaacactcaatgacctgacctttagtgagggttaagtcaat
accttctgaaccatttccataatgggtgaaagttccctcaagaattttactctgcagaaacggccttaacgacgtagtcgaacctcctcttcagtaactaaatctaccaatacca
aatctgatggaagaatgggctaattgcatcatccttaccacgcgagagacctatcagctcactcaaaaggcggtatcaccagaaatcaggggataacgcagga
aagaacatgtgagcaaaaaggccagcaaaaaggccaggaaccgtaaaaaggccgcgttgcgttgcgttttccataggctccgccccctgacgagcatcacaataatcg
acgtcaagttagagggtggcgaaacccgacaggactataaagataccaggcggttccccctggaagctccctcgtgcgtctcctgttccgacctgccgttaccgga
tacctgtccgcttctccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgtccaagctgggtgtgtgcacgaac
ccccgttcagcccgacctgtgccttaccggtaactatcgtcttgagcccaacccggtaagacacgacttatgccactggcagcagccactggtaacaggattag
cagagcgaggatgttaggggtgtacagagttctgaagtggtggcctaactacggctacactagaagaacagtaattgggtatctgcgtctgtgaagccagttacctt
cggaagaaagagttgtagcttctgacgggcaaaacaccacgctggtgagcgggtgtttttgtttgcaagcagcagattacgcgcagaaaaaaggatctcaagaa
gatcctttgatctttctacgggtgtgacgctcagtggaacgaaaactcagtttaagggttttggctatgagattatcaaaaaggatcttccactagatccttttaataaa
aatgaagtttaaaatcaatctaaagtatataatgagtaaaacttggctgacagagttctgaggtcattactggatctatcaacagcagtcgaagcgagctcgatcaaatlacg
ccccgacctgacctatcgcagtagttgttaattcattaagcattctgccgacatggaagccatcacaacggcatgatgaacctgaatcgccagcgccatcagcacc
ttgtcgcttgcgtataatatttggccatggtgaaaacggggcggaagaagttgtccatattggccacgtttaaatcaaaactggtgaaactcaccagggttggtgag

acgaaaaacatatctcaataaaccccttagggaaataggccaggtttaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaatcgtcgtggtatt
 cactccagagcgcgatgaaaacgtttcagtttgcctcatggaaaacgggtgtaacaagggtgaacactatcccatatcaccagctcaccgcttttcattgccatacgaattccg
 gatgagcattcatcagggcgggaagaatgtgaataaaggccggataaaacttgtgcttattttctttacggcttttaaaaggccgtaatatccagctgaacggctggtta
 taggtacattgagcaactgactgaaatgcctcaaatgttctttacgatgccattgggatatacaacgggtggtatataccagtgattttttctccattttagcttcttagctct
 gaaaatctcgataactcaaaaaatacggccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgccgatcaactcgcgcgttggccacctgacgtctaa
 gaaaaggaaatattcagcaatttgcctgcccgaagaaggccaccctgaaggtagacc

SEQ ID NO: 2 LXR77:

ggctcgcaccagaagaacagtaaaataagcaaggtagctgaaattaatatttttaaatggttctaaccgatgccgaagaactgcgcagtcgggtataacgtctgacat
 gtcttttttgaattggaatccaaccactcaagtactctgttcttttgcgaaaaatataccacaaatggccatcgaaagtgaatcgaaaccacctcagactggc
 accgacaaagcaagattatacagacagagtagctttatacgtaccgttaagtctcaagcaaagggttttcttattactgaacgggtaaaagagtatctgggcccgttgcga
 agatgcaaacggaataagtatttcaagtgtcatttgccttagccgtctgacaccattggctatttggatatttatattgactttcgtgtacattgacacatgactgttctatt
 ggcaaatgaaccacgggcattgactattttcaggttactactatataattatcatcacgggcaaggattgtagcttgcctatgctgcgcggaggagatatttacttttatttct
 agttttttacagttatttattaattaattttttatgcacgaataaaaaagctataatttaagttctttttatttataatacatttctctacagactgtcaccggatgtgcttccg
 gtctgatgagtcctgaggacgaacagcctctacaataattttgtttaagagcaggtgttcatggccgtgcgtatgatgtggggggctcggggtgaaaccggggt
 tcggagcggcagggtttttttgttttttttttgcggataaaaaatccctatcagtgatagagattttctttttttttttccctatcagtgatagagaatacaagctacaa
 aaataaataaaaaATGAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGATCTCATCCGTGATCACATCAG
 CCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCG
 GCTGAAGAACATCTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCG
 GGATTTCGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACC
 AGGTTCTAAGGATATCTCTGCTGGAGACATGAGAGCTGCCAACCTTTGGCCAAGCCCGCTCATGATCa
 ggtcttTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTTGATGTCATTTTCGCGGTGGCTGAGATCA
 GCCACTTCTTCCCCGATAACGGAAACCGGCACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTC
 ATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCC
 AGGGGGATCACCATCCGTGCCCCGGGCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGAT
 AACGGCTCTCTCTTTTATAGGTGTAAACCTTAAACTGCATTTACCAGCCCCCTGTTCTCGTCAGCAAAA
 GAGCCGTTTCAATTCAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCCAGCGTTTCG
 GCACGCAGACGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTGACATCATATAT
 GCCTTGAGCAACTGATAGCTGTCGCTGTCAACTGTCACTGTAATACGCTGCTTCATAGCATACCTCTTT
 TTGACATACTTCGGGTATACATATCAGTATATATTCTTATACCGCAAAAATCAGCGCGCAAATATGCA
 TACTGTTATCTGGCTTTTAGTAAGCCgaagacctggtagccctaagaaaaagagaaaagtgggtggcagtgccggaagcgggggatcaggt
 ggttctggagggtccGATATCATGCAGGATCTGCCGGGCAACGATAACAGCACCGCGGGCGCGGTACGAAA
 AACAATTACGGGTCTACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCGAAGAGGCGG
 GGCTGGCGGCTCCGCGCCTGTCTTTCTCCCCGCGGGACACACGCGCAGACTGTCGACGGCCCCCCCCG
 ACCGATGTCAGCCTGGGGGACGAATTGCACTTAGACGGCGAGGACGTGGCGATGGCGCATGCCGACG
 CGCTAGACGATTTTCGATCTGGACATGTTGGGGGACGGGGATTCCCCGGGTCCGGGATTTACCCCCAC
 GACTCCGCCCCCTACGGCGCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCT
 TGGAATTGACGAGTACGGTGGGAAGCTTTAAtaaaagcttttgattaagccttctagtccaaaaaacacggtttttgtcatttatttcattttctta

gaatagtttagtttattcattttatagtcacgaatgtttatgattctatataggggtgcaacaagcatttttattttatgttaaaacaatttcagggttaccttttattctgcttgggt
gacgcgtgtatccgcccgtcttttggtcacccatgtatttaattgcataataattcttaaaagtgagctagctatfttctatttaccatctctcatttctcatttctcctgacc
agctgagtttcatggctgaattactatgtcggggtaatgtggcttgcctagccttgggtgaagtatttggcacaaggcttagggcctgttgaggagtgacgttgccttg
tccccgcccgtcaccggccagcagcatggaggccagaataccctccttgacagcttgcagctcagggcatgatgtgactgtcggcgtacatttag
cccatacatccccatgtataatcatttgcaccatacattttgatggccgcacggcggaagcaaaaattacggctcctcgtcgcagacctgcgagcagggaacgctcc
cctcacagacgcgttgaattgtccccacggcgccctgtagagaaataaaaaagggttaggatttggcactgaggttcttcttcatatacttctttaaacttctgctagg
atacagttctcacatcacatccgaacataaacaacaatgtctaagaatacgttgcctaccgggtgatcacgtcggtaagaagttactgacgaagctattaaggcttga
atgccattgtgaagtcgctccagaaatfaagttcaatttccaacatcattgatcgggggtgctgccatcgtaccactggcactccttaccagatgaagctctagaagc
ctctaagaaagccgatgctgttactagggtgctgttgggtggtccaaaatggggtagggcgagtagaccagaacaaggctattgaagatcagaaaggaattgggt
ctatacgccaactgagggcatgtaacttcttctgatttcttactagatcttctccttgaagcctgaatatgcaaaagggtaccgatttctcgtcgttagagaattgggtgg
gggtatctacttgggtgaaagaaaagaagatgaagggtgacggagttgctgggttctgagaaatacagtttctgaagttcaagaattacaagaatggctgcttcttgg
cattgcaacaaaaccaccattaccaatctgtgacttgcagaggctaacgtgcttgccttccagattgtggagaaagactgttgaagaaaccatcaagactgagttcc
cacaattaactgttcagcaccatgtatcgttctgctgtatgatttgggttaatacacaactaagctaaacgggtgttgttattaccaacaacatgttgggtgatattatccg
atgaagcctctgttattccagggttcttgggttattaccctctgcactctagcttccctacctgacactaacaaggcattcgggttgcagaaacctgcatggttctgcccc
gatttaccagcaacaagggttaaccaattgctaccatcttctgcagctatgatgttgaagttatccttggatttgggtgaagaaggtagggctcttgaagaagctgttaga
aatgtcttggatgcagggtgcagaaccgggtgaccttgggtggttctaactctaccactgaggttggcgatgctatcgccaaggctgtcaaggaaacttggcttaaacagta
ctgacaataaaaagattctgtttcaagaactgtcatttgtatagtttttatattgtatgttctattttaatcaaatgttagcgtgatttatatttttgcctcgacatcatctg
cccagatgcgaagttaagtcgcagaaagtaatacatcgctcaatcgatgtgaatgctggtcgtatactgctgtcgttgcgatactaacgccccatccagtgctgcac
ctccattactcgtatcgtatcgttgcgacacgaaattacaaaatggaatatgttcatagggtagacgaaactatatacgcaatctacatacattatcaagaaggagaaa
aaggaggtgtaaggaatacaggtgaagcaaatgatactaatggctcaacgtgataaggaaaaagaattgcacttaacattaatattgacaaggaggaggccaccac
acaaaaagttagggtlaacagaaaatcatgaaactatgattcctaatttatattggaggatttctctaaaaaaaaaaaaatacaacaataaaaaaacactcaatgacctg
accatttggatggagtttaagtcataacttcttgaaccatttccataatgggtgaaagtccctcaagaatttactctgtcagaacggccttaacgacgtatgcgacctctc
ttcagtactaaatctaccaataccaaatctgatggaagaatgggctaagcatcatccttaccagcgagagacctatcagctcactcaaaggcggtatacgggtatcca
cagaatcaggggataacgcaggaaagaacatgtgacaaaaggccagcaaaaggccaggaaccgtaaaaaggccgctgttgcgtggcgttttccatagggtccgccc
ccctgacgagcatcacaataatcgacgtcaagtcagaggtggcgaacccgacaggactataaagataccaggcgttccccctggaagctccctcgtgcgtctcc
tgttccgacctgcccgttaccggatacctgtccgccttctcccttccgggaagcgtggcgcttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttccgt
ccaagctgggctgtgtgcacgaacccccgttcagcccagccgtgcgccttatccggtaactatcgtttagcccaacccggtaagacacgactatcggcactggc
agcagccactggtaacaggattagcagagcgaggtatgtagcgggtgtacagagttcttgaagtggtggcctaactacggctacactagaagaacagtttgggtatct
gcgctctgtgaagccagttaccttcgaaaaagagttggtagctcttgcctggcaacaaccaccgctggtagcgggtgttgggttttgggttgaagcagcagattacgc
gcagaaaaaaaggatctcaagaagatccttctgatcttctacggggctcgtacgctcagtggaacgaaaactcacgttaagggttgggtcatgagattacaaaaagga
tcttcacctagatcctttaaattaaaaatgaagtttaaatcaatctaaagtatataatgagtaaaacttggctgtacagagttctgaggtcattactggatctatcaacagcagtc
caagcgagctcgatatcaaatcagccccgcctgccactcgcagctactgttgaattcattaagcattctgccgacatggaagccatcacaacggcatgatgaac
ctgaatcgccagcggcatcagcaccctgtgccttgcgtataatfttcccatgggtgaaaacggggcggaagaagtgtccatattggccacgtttaaataaaaactgggt
gaaactcaccaggggattggctgagacgaaaaacataattctcaataaaccctttagggaaataggccaggttttaccgtaaacgccacatcttgcgaatatatgtgtag
aaactgccggaaatcgtcgtggttactccagagcgatgaaaacgttgcagtttgcctatggaacgggtgaacaagggtgaacactatcccatatcaccagctcac
cgtcttcttattgccaatcagaaatccggatgagcattatcaggcgggcaagaalgtgaataaaggccgataaaaacttgccttattttcttaccggctttaaaggcc
gtaatatccagctgaacggctgtggtataggtacattgagcaactgactgaaatgcctcaaatgttcttaccgatgccattgggatatatcaacgggtgtatatccagtgatt
ttttctccattttagcttcttagctcctgaaaaatctcgataactcaaaaatcagccccgtagtgatcttatttatttggtgaaagttggaacctcttaccgtgcccgatcaac

tcgcgcgtttgccacctgacgtctaagaaaaggaatattcagcaatttgcctgcccgaagaaaggccaccgctgaaggtgagcc

SEQ ID NO: 3 LXR233;

ggctcgcacccagaagaacagtaaaataaagcaaggtacgtgaaaltaatattttaaaggttctaaccgatgccgaagaactgcgcagtcgggtataacgtctgacat
gtcctttttgatttgaatccaaccactcaagtactctgttcatttactttgcgaaaaatatacccacaaattgcccacgaaagtgaatcgaaaccaccttcagactggc
accgacaaagcaagattatacagacagagtactttatacgtaccgttaagtcacagcaaggggtttcttatttactgaacgggtaagagtatctgggccggcttgcca
agatgcaaacgaataagatttcaaaagtgcatttgccttagccgctcgtgacaccattggctatttggatatttataattgactttcgtgtacattgatcacatcgactgttctatt
ggcaaatgaaccacgggcattgactattttcaggttactactatataattatcatcacgggcaaggattgtagcttgcctatgctgcgaggagatatttacttttattattct
agtttttacagttatttattaattaattttttatagcatgcgaataaaaagtcataatttaagttcttttatttataatacatttctctacagagctgtcaccggatgtgcttccg
gtctgatgagtcgtaggacgaacagcctctacaataattttgttaagagcaggtgttcattgcccgtgctatgatgtggggggctcgggctgtgaaaccggggt
tcggagcggcaggggggtttttgttttttttttgcggtataaaaatccctatcagtgatagagatttcttttttttttttccctatcagtgatagagaatacaagctacaa
aaataaataaaaaATGAAAGCGTTAACGGCCAGGCAACAAGAGGTGTTTGATCTCATCCGTGATCACATCAG
CCAGACAGGTATGCCGCCGACGCGTGCGGAAATCGCGCAGCGTTTGGGGTTCCGTTCCCCAAACGCG
GCTGAAGAACATCTGAAGGCGCTGGCACGCAAAGGCGTTATTGAAATTGTTTCCGGCGCATCACGCG
GGATTCTGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACC
AggttctaaggatatctctgtgagacatgagagctgccaacctttggccaagcccgtcatgacATTGTGGGTGAAGATCAGCTGTGGGG
TCGTCTGACTGGAATTTGTTGATAGCGTTGAACGTCTGGAAGCACCGGCACTGATTAGCCGTTTTG
AAAGCCTGATTGCAAGCTGTGGTTTTACCGCCTATATCATGGCAGGTCTGCCGAGCCGTAATGCCGGT
CTGCCGGAAGTACCCTGGCAAATGGTTGGCCTCGTGATTGGTTTGATCTGTATGTTAGCGAAAATT
TAGCGCAGTTGATCCGGTTCCGCGTTATGGTGCAACCACTGTTTCATCCGTTTGTGGAGTGATGCAC
CGTATGATCGTGACCGTGATCAGGCAGCACATCGTGTTATGACCCGTGCAGCAGAATTTGGTCTGGTT
GAAGGTTATTGTATTCCGCTGCATTACGATGATGGTAGCGCAGCAATTAGTATGGCAGGTGAAGATCC
TGATCTGAGTCCGGCAGCCCGTGGTGTAATGCAGCTGGTTAGCATTATGCACATAGCCGTCTGCGTG
TACTGAGCCGTCCGAAACCGATTTCGTCGTAATggtagccctaagaaaaagagaaaagtgggtggcagtgggcgaagcgggggatca
gggtgttctggagggtccaggtcttTTATATTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGTCATTTTCGCGGT
GGCTGAGATCAGCCACTTCTTCCCCGATAACGGAAACCGGCACACTGGCCATATCGGTGGTCATCATG
CGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACG
TGCACTGGCCAGGGGATCACCATCCGTGCCCCGGGCGTGTCAATAATATCACTCTGTACATCCACAA
ACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTAAACTGCATTTCACCAGCCCCCTGTTCTCG
TCAGCAAAAGAGCCGTTCATTTCAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTT
CCAGCGTTCGGCACGCAGACGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTGA
CATCATATATGCCCTTGAGCAACTGATAGCTGTCGCTGTCAACTGTCACTGTAATACGCTGCTTCATAG
CATACTCTTTTTGACATACTTCGGGTATACATATCAGTATATATTCTTATACCGCAAAAATCAGCGCG
CAAATATGCATACTGTTATCTGGCTTTTAGTAAGCCGAAGACCTTAAaagcttttgattaagccttctagtcaaaaaacacg
ttttttgtcattttattctttagaatagttttagttattcattttatagtcacgaatgtttatgattctatatagggtgcaaacaagcattttcattttatgttaaaacaatttcag
gtttaccctttattctgcttgggtgacgcgtgtatccgcccgtcttttggtcacccatgtatttaattgcalaataattcttaaaagtggagctagctatttctatttacatcct
ctcatttctcatttctcctgaccagctgagttcatggcttgaattactatgtcggggtaattgtggctgtcgtagccttgggtgaagtatttggcacaaggtctagggcctgt
tgaggagtgacagcttgccttgcctcccgccgggtcaccggccagcgacatggaggccagaataccctccttgacagcttgcagctgcgcagctcagggcatgat

gtgactgtgccccgacatttagcccatatccccatgtataatcatttgcacccatacattttagtgccgcacggcggaagcaaaaattacggctcctcgctgcagac
ctgcgagcagggaacgctccccacagacggtgaattgtccacgcccgcgccccgtagagaaatataaaagggttaggtattgcccactgagggtctcttctcatat
acttcttttaaaatttgetaggatacagttctcacatcacatccgaacataaacaacatgtctaagaatcgttgcctaccgggtgatcacgtcggttaaagaagtact
gacgaagctattaaggcttgaatgccattgctgaagtcgctccagaaftaagtccaattccaacatcattgatcgggggtgctgccatcgatccactggcactcctt
accagatgaagctctagaagcctctaagaaagccgatgctgtcttactaggtgctgttggtggtccaaaatggggtagggcgagtagaccagaacaaggctattg
aagatcagaaggaattgggtctatacgccaactgaggccatgtaactttgcttctgattcttactagatcttctccttgaagcctgaatatgaaaagggtaccgatttcg
tcgtcgttagagaattgggtggtatcttgggtgaaagaaaagaaggaaggtgacggaggtgcttgggattctgagaatacagtgcttctgaagttcaagaat
tacaagaatggctgcttctggcattgcaacaaaaccaccattaccaatctggtcacttgacaaggetaacgtgcttgcctctccagattgggagaagactgttgaa
gaaaccatcaagactgagttcccaataaactgttcagcaccaattgatcgattctgctgctatgatttgggttaaatcaccacaaagctaaacgggtgtgtattaccaaca
acatgttgggtgataattatccgatgaagcctctgttattccaggtcttgggttattacctctgcatctctagcttccctacctgacactaacaaggcattcggttgtagca
accatgtcatggttctgccccagatttaccagcaacaagggttaaccaatgtctaccatcttatctgcagctatgatgtgaagttatccttggttgggtgaagaaggtag
ggctcttgaagaagctgttagaaatgcttggatgcaggtgtcagaaccgggtgacctgggtggttctaactctaccactgaggttggcgatgctatcgccaaggctgtcaa
ggaaatcttggcttaaacagttactgacaataaaaagattctgtttcaagaactgtcattgtatagtttttataattgtagtgttctattttaatcaaatgttagcgtgatttatat
tttttctgctcgacatcatctgcccagatgcgaagttaagtcgcgagaaataatcatcgctcaatcgtatgtgaatgctggtcgtctatactgctgctgattcgatacta
acggccgcatccagtgctgacctccattactgctatcgatgctgggtgcgacacgaaattacaaaatggaatatgttcatagggtagacgaaactatatacgcaatctaca
tacatttatcaagaaggagaaaaaggaggtgtaaaaggaatacaggtaagcaaaattgatactaaggtcacaacgtgataaggaaaaagaattgcactttaaactaatatt
gacaaggaggaggccaccacacaaaaagttaggtgaacagaaaatcatgaaactatgattcctaatttatattggaggattttcttaaaaaaaaaatacaacaa
ataaaaaacacacaaatgacctgacctttagtgagggttaagtaacactcttgaaccatttccataatggtgaaagtccctcaagaatttactctgacagaaacggcct
taacgacgtagtcgacctcctctcagttactaaatctaccaataccaaatctgatggaagaatgggctaattgcacatccttaccagcgagagacctatcagctcactcaa
aggcggtataacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaacgtaaaaggccgctgtgctgg
cgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaaccgacaggactataaagataccaggcggttccccctg
gaagctccctctgctgctctctgttccgacctgcccgttaccggataacctgtccgcttctccttccgggaagcgtggcgcttctcagctcagctgtaggtatctc
agttcgggtaggtcgttcgtccaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgcttatccggtaactatcgtcttgagcccaaccggtaag
acacgactatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgtacagagttcttgaagtgttggcctaactacggctacact
agaagaacagtatgttggatctgctgctctgtgaagccagttaccttggaaaaagagttggtagctcttgatccggcaacaaaccaccgctggtagcgggtgtttttg
ttgcaagcagcagattacgcgcagaaaaaaggatctcaagaagatcctttagcttcttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttgg
tcatgagattatcaaaaaggatcttccatagatccttttaaaataaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggctgacagagttctgaggtcatta
ctggatctatcaacagcagtcgaagcagctcgatatcaaaattacgccccgcccgtccactcatcgagctactgttgaattcattaaagcattctgccgacatggaagccat
cacaacggcatgatgaacctgaatcgccagcggcatcagcacctgtcgccttgcgtataatattgcccattggtgaaaacggggcggaagaagttgccatattggc
cacgtttaaataaaaactggtgaaactacccagggttggctgagacgaaaaacataattcctaataaacctttagggaaataggccaggtttaccgtaaacgcca
catcttgcgaatatatgtgtagaaactgcccggaaatcgtcgtggtattcactccagagcgatgaaaacgttccagtttgcctatggaaaacgggtgaacaagggtgaacac
tatcccatatcaccagctcaccgtcttctattgccatacgaaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaactgtgcttattttct
ttacggctttaaaggccgtaatatccagctgaacggctcgtgttataggfacattgagcaactgactgaaatgcctcaaaatgttctttacgatccattgggatatatcaa
cggtgggtatattcagtgatttttctcatttttagcttcttagctcctgaaaaatctcgataactcaaaaatagcccggtagtgatcttatttcattatggtgaaagttggaac
ctcttacgtgcccgatcaactcgcgcgttggccacctgacgtctaagaaaaggaatattcagcaatttggccgtgccgaagaaaggccaccgctgaaggtgagcc

SEQ ID NO: 4 LXR322:

agaacagtaaaataaagcaagggtacgtgaaattaatatttttaaatggttctaaccgatgccgaagaactgcgcagtcgggtataacgtctgacatgtcctttttgatttgg

aatccaaccactcaagtgactctgttcatttactttgcgaaaaataaccacaaattgcccatcgaaagtgaatcgcaaaccaccttcagactggcaccgacaaagcaaa
gattatacagacagagtactttatacgtaccgttaagtctcaagcaaagggtttcttatttactgaacgggtaaagagtatctggcccggttccaagatgcaaaccgaa
taagtatttcaaaagttgcattgccttagccgtcctgacaccattggctatttggatatttataattgactttcgtgtacattgatcacatcgactgttctattggcaaatgaaccac
gggcattgactattttcaggttactactatataattatcatcacgggcaaggattgtagcttgcgtatgctgctggaggagatatttacttttatttctagtttttacagtattt
attaattaattattttatgcatgcgaataaaaaagtctatatattaagttcttttatttataatacattttctctacgactgtcacccggatgtgctttccggtctgatgagtcctgt
aggacgaaacagccctacaaaataatttgttaagagcaggtgttcatggccgtgcgtatgatgtgggggctcgggcgttgaaccgggggtcggagcggcaggg
ggttttcaacggcctagcatgtgattaattaatttttgtttttttgcagtataaaaaagtagttgtttaaacaacaaactttttcatttctttgtttcccttctcttcttttagtta
gtttgttaaacaacaaactagaatatcaagctacaaaataaataaaaaATGAGGTCTTCTTATATTCCCCAGAACATCAGGTTAAT
GGCGTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAAACCGGCA
CACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGCACCACCGGGTAAAGTTCA
CGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCCGGGCGTGT
CAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTA
AACTGCATTTACCAAGCCCCCTGTTCTCGTCAGCAAAAAGAGCCGTTCAATTTCAATAAACCGGGCGACCT
CAGCCATCCCTTCTGATTTTCCGCTTTCAGCGTTTCGGCACGCAGACGACGGGCTTCATTCTGCATGG
TTGTGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTCGCTGTCAA
CTGTCACTGTAATACGCTGCTTCATAGCATACTCTTTTTGACATACTTCGGGTATACATATCAGTATA
TATTCTTATACCGCAAAAATCAGCGCGCAAATATGCATACTGTTATCTGGCTTTTAGTAAGCCgaagacct
ggttctaaggatacATTGTGGGTGAAGATCAGCTGTGGGGTCGTCGTACACTGGAATTTGTTGATAGCGTTGA
ACGTCTGGAAGCACCGGCACTGATTAGCCGTTTTGAAAGCCTGATTGCAAGCTGTGGTTTTACCGCCT
ATATCATGGCAGGTCTGCCGAGCCGTAATGCCGGTCTGCCGGAAGTACCCTGGCAAATGGTTGGCCT
CGTGATTGGTTTTGATCTGTATGTTAGCGAAAACCTTAGCGCAGTTGATCCGTTCCGCGTTATGGTGC
AACCCTGTTTCATCCGTTTGTGGAGTGATGCACCGTATGATCGTGACCGTGATCAGGCAGCACATC
GTGTTATGACCCGTGCAGCAGAATTTGGTCTGGTTGAAGGTTATTGTATTCCGCTGCATTACGATGAT
GGTAGCGCAGCAATTAGTATGGCAGGTGAAGATCCTGATCTGAGTCCGGCAGCCCGTGGTGTAAATGC
AGCTGGTTAGCATTTATGCACATAGCCGTCTGCGTGTACTGAGCCGTCCGAAACCGATTCTGTCGTAAT
ggtagccctaagaaaaagagaaaagtggtggcagtgggcgaagcgggggatcaggtggttctggagggtccGATATCATGCAGGATCTGCCG
GGCAACGATAACAGCACCGCGGGCGCGGTACGAAAAACAATTACGGGTCTACCATCGAGGGCCTGC
TCGATCTCCCGGACGACGACGCCCCCGAAGAGGCGGGGCTGGCGGCTCCGCGCCTGTCTTTCTCCCC
GCGGGACACACGCGCAGACTGTCGACGGCCCCCCCCGACCGATGTCAGCCTGGGGGACGAATTGCACT
TAGACGGCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAGACGATTTGATCTGGACATGTTGGG
GGACGGGGATTCCCCGGGTCCGGGATTTACCCCCACGACTCCGCCCCCTACGGCGCTCTGGATATGG
CCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGTGGGAAGCTTTAA
taaaagctttgattaagcctctagtcacaaaaacacgtttttgtcatttatttcttctttagaatagtttagttattcattttatagtcacgaatgtttatgattctatatagggt
tgcaaacagcatttttatttattgtaaaacaatttcaggtttacccttttattctgcttgggtgacgcgtgtatccgccgctctttgttcacccatgtatttaattgcataata
attcttaaaagtgagctagctatttctatttatacctctcatttctcatttctcctgaccagctgagtttcatggcttgaattactatgtcggggtaattgtggcttgcgtg
gccttggtagaattttggcacaaggtctagggcctgttgaggagtgacgcttgccttgtccccgccgggtcaccggccagcagacatggaggccagaataccctc
cttgacagcttgacgtgcgcagctcaggggcatgatgtgactgtcgccgtacatttagccatacatcccatgtataatcatttgcatacatctttgatggccgcac
ggcgcggaagcaaaaattacggctctcgtctgcagacctgcgagcagggaaacgtccctcacagacgcgttgaattgtccccacgccgcgccctgtagagaaata

taaaagggttaggatttgcactgaggttcttcttcatatacttctttaaacttctgctaggatacagttctcacatcacatccgaacataaacaacaatgtctaagaatatcgt
tgtctaccgggtgatcagctcggttaaagaagtactgacgaagcttaaggcttgaatgccattgctgaagtcctccagaaattaagttaattccaacatcacttga
tcgggggtgctgccatcgatgccactggcactcctttaccagatgaagctctagaagcctctaagaaagccgatgctgttactaggtgctgttggtggtccaaaatgg
ggtagcgggcgcaglttagaccagaacaaggctattgaagatcagaaggaattgggtctatacggcaacttgaggccatgtaacttctgttctgattcttactagatcttct
tccttgaagcctgaatatgcaaaagggtaccgatttctgctgctgtagagaattgggttggtggtatctacttgggtgaagaaaagaagatgaaggtgacggaggtgcttgg
gattctgagaatacagtggttctgaagttcaagaattacaagaatggctgcttcttggcattgcaacaaaaccaccattaccaatctgggtcacttgacaaggctaactg
gcttgcctctccagattgtggagaagactgttgaagaaccatcaagactgagttcccacaattaactgttcagcaccatgtatcgattctgctgctatgatttggtaa
atcaccaactaagctaaacgggtgttattaccaacaacatgttgggtgatattatctccgatgaagcctctgttattccagggtcttgggttattacctctgcatctctagct
tcctacctgacactaacaaggcattcgggtgtacgaaccatgtcatggttctgccccagattaccagcaacaagggttaaccaatgtaccatcttatctgcagctat
gatgtgaagttatcttggatttgggtgaagaaggtagggctcttgaagaagctgttagaaatgcttggatgcaggtgtcagaaccgggtgaccttgggtggttctaactta
ccactgaggttggcgtatctgccaaggctgtcaaggaaatctggcttaaacagtactgacaataaaaagattcttgtttcaagaactgtcattgtatagtttttata
ttgtagtgttctttaaatacaaatgttagcgtgatttatatttttgcctcgacatcatctgccagatgcgaagttaagtgcgcagaaagtaatatcatcgctcaatcgtat
gtgaatgctggctgctatactgctgctgattcgatactaacgccgccatccagtgctgacctccattactcgtatcgatgctgggtgcgacacgaattacaaaatggaata
tgttcatagggtagacgaactatatacgaactacatacattatcaagaaggagaaaaaggaggatgtaagggaatacaggttaagcaaatgatactaatggctcaa
cgtgataaggaaaaagaattgcactttaacattaattgacaaggaggaggccaccacacaaaaagttagggttaacagaaaatcatgaaactatgattcctaatttatat
attggaggattttcttaaaaaaatacaacaaataaaaaacactcaatgacctgaccttggatggagttfaagtaacatcttcttgaaccttcccataatgggtg
aaagtccctcaagaattttactctgtcagaaacggccttaacgacgtagtcgacctctctcagtaactaaatctaccaataccaaatctgatggaagaatgggctaatagca
tcactcttaccagcgagagacctatcagctcactcaaggcggtaatacgggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagc
aaaaggccaggaaccgtaaaaaggccgcgttctgctggcgttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaac
ccgacaggactataagataccaggcgttccccctggaagctccctgctgctctctctgttccgacctgcccgttaccggataacctgtccgcttcttcccttcggga
agcgtggcgcttctcatagctcacgctgtaggtatctcagttcgggttaggtcgttgcctcaagctgggctgtgtgcacgaacccccgttcagcccagccgctgcgc
cttatccggtaactatcgtcttgagcccaaccggtaagacacgactatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtagcggtgct
acagagttctgaagtgggtggcctaactacggctacactagaagaacagtatttgggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgat
ccggcaaaacaccaccgctgtagcgggtgttttttggcaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttctgatctttctacggggtctg
acgctcagtggaacgaaaactcacgttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatcctttaaataaaaaatgaagtttaaatcaatctaagtat
atatgagtaaaacttggtctgacagagttctgaggtcattactggatctatcaacagcagtcgaagcgagctcgatatcaaatcagccccgcctgccactcctgcagta
ctgttgaattcaatgaacattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgccagcggcatcagcaccttctgccttgcgtataatattgcc
atggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaataaaaactgggtgaactcaccagggttgggtgagacgaaaaacataattctcaataaacc
cttagggaaataggccaggtttaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgttc
agtttgcctatggaacgggtgaacaagggtgaacactatccatataccagctcaccgtcttcttgcatacgaattccggatgagcattcatcagcgggcaag
aatgtgaataaaggccggataaaactgtgcttttttcttaccggtctttaaaggccgtaatatccagctgaacggctgtgtataggtacattgagcaactgactgaaa
tgctcaaaatgttcttaccatgcccattgggatatacaacgggtgttatccagtgattttttctcatttttagcttctgctgaaaatctcgataactcaaaaatac
gcccggtagtgatcttattcattatgggtgaaagttggaacctcttactgcccgatcaactcgcgcgttggccacctgacgtctaagaaaaggaaatcagcaatttgc
gtgccgaagaaaggcccccgtgaagggtgagccggtctcgaccaga

SEQ ID NO: 5 LXR321:

agaacagtaaaataaagcaagggtacgtgaataatatttttaaatgggttaaccgatgccgaagaactgcgcagtcgggtataacgtctgacatgtcctttttgatttgg
aatccaaccactcaagtgactctgttctatttcttgcgaaaaatataccacaaattgcccatcgaaagtgaatcgcaaacaccttcagactggcaccgacaaagcaaa

gattatacagacagagtactttatacgtaccgttaagtctcaagcaaagggttttctatttactgaacgggtaagagtatctggccggcttgccaagatgcaaaccgaa
taagtatttcaaaagttgcatttgccctgagccgtcctgacaccattggctatttggatattttattgactttcgtgtacattgatcacatcgactgttctattggcaaatgaaccac
gggcattgactattttcaggttactactatattatcatcacgggcaaggattgtagcttgcctatgctgcggaggagatatttacttttattattctagtttttacagttattt
attaattaattttttatatgcatgcgaataaaaaagctatatttaagttctttttatttataatacattttctctacgagctgtcacccggatgtgctttccggctgatgagtcggtg
aggacgaaacagcctctacaaataattttgttaagagcaggtgttcattggccgtgcgtatgatgtggggggctcgggcgttgaaaccgggggtcggagcgccaggg
ggttttcaacggcctagcatgtgattaattaattttgttttttttgcagtataaaaaagttagttgttttaacaacaaactttttcattttttgtttcccttctcttttagtta
gtttgttttaacaacaaactagaatatcaagctacaaaaataataaaaaATGTCTACGAAGCTATCAAAAAGGCAACTTGATATCCT
CCGTTTTTATTAAAGCAGAGGTTAAATCAAAAGGATATCCGCCTTCCGTGAGAGAGATCGGAGAGGCT
GTCGGGCTTGCGTCCAGTTCTACTGTCCACGGCCATTTGGCCCGTTTGGAACAAAAGGGCTGATCAG
ACGAGATCCGACAAAACCAAGAGCAATAGAAATTCTTGATGAAGAAGTAGACATTCCGCAAGCCA
GGTCGTTAATGTACCGGTCATCGGGAAAGTCACGGCGGGATCTCCTggttctaaggatataAGGTCTTCTTATA
TTCCCCAGAACATCAGGTTAATGGCGTTTTTGATGTCAATTTTCGCGGTGGCTGAGATCAGCCACTTCTT
CCCCGATAACGGAAACCGGCACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATA
TGCACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCA
CCATCCGTCGCCCCGGGCGTGTCAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCT
CTTTTATAGGTGTAAACCTTAAACTGCATTTACACCAGCCCCGTGTTCTCGTCAGCAAAAAGAGCCGTTCA
TTTCAATAAACCGGGCGACCTCAGCCATCCCTTCCTGATTTTCCGCTTTCAGCGTTCGGCACGCAGA
CGACGGGCTTCATTCTGCATGGTTGTGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGC
AACTGATAGCTGTCGCTGTCAACTGTCACTGTAATACGCTGCTTCATAGCATACTCTTTTTTGACATAC
TTCGGGTATACATATCAGTATATATTCTTATACCGCAAAAATCAGCGCGCAAAATATGCATACTGTTAT
CTGGCTTTTAGTAAGCCGAAGACCTggtagccctaagaaaaagagaaaagtggtggcagtggcggaagcgggggatcaggtggttctgga
gggtccGATATCATGCAGGATCTGCCGGGCAACGATAACAGCACCGCGGGCGCGGTACGAAAAACAAT
TACGGGTCTACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCGAAGAGGCGGGGCTGG
CGGCTCCGCGCCTGTCTTTCTCCCCGCGGGACACACGCGCAGACTGTGACGCGCCCCCCCCGACCGAT
GTCAGCCTGGGGGACGAATTGCACTTAGACGGCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAG
ACGATTTTCGATCTGGACATGTTGGGGGACGGGGATTCCCCGGGTCCGGGATTACCCCCACGACTCC
GCCCCCTACGGCGCTCTGGATATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGAAT
TGACGAGTACGGTGGGAAGCTTTAAtaaaagcttttgattaagccttctagtcaaaaaacacgttttttgtcatttatttctttagaatagttta
gtttattcattttatagtcacgaatgttttatgattctatataggttgcaacaagcatttttattttatgttaaaacaatttcaggtttacctttattctgcttgggtgacgcgtgt
atccgcccgctcttttggtcacccatgtatttaattgcataaataattcttaaaagtgagctagtctatttctatttacaacacctctcatttctcatttctcctgaccagctgagttt
catggcttgaattactatgtgcggggaatgtggcttgtcgtagccttggtgaagtatttggcacaaggtctagggcctgttgaggagtgacagcttgccttgcctcccgccg
ggtcacccggccagcgacatggaggccagaataacctcttgcagcttgcagctgcgcagctcaggggcatgatgtgactgtcggcgtacatttagccatacatc
cccatgtataatcatttgcatacatatttgcagcgcacggcgcgaagcaaaaattacggctcctcgtgcagacctgcgagcagggaaacgctccctcacaga
cgcggtgaattgtcccccagccgcgcccctgtagagaaataaaaaggttaggatttggcactgaggttcttcttcatatacttcttttaaaatcttgcagatacagttct
cacatcacatccgaacataaacaacaatgtctaagaataatcgttgcctaccgggtgatcacgtcggttaagaagtactgacgaagctattaaggtcttgaatgccattgc
tgaagtcgclccagaataaagttcaatttccaacatcacttgatcgggggtgctgccaatcgatgccactggcactccttaccagatgaagcctctagaagcctctaaagaa
gccgatgctgtcttactaggtgctgttgggtgggtcaaaatggggtacgggcgcagtttagaccagaacaaggtctattgaagatcagaaggaattgggtctatagcca
acttgaggccatgtaactttgcttctgattcttactagatcttctccttgaagcctgaatatgcaaagggtaccgattcgtcgtcgttagagaattggttgggtggtatctactt

tggtgaaagaaaagaagatgaaggtgacggaggtgcttgggattctgagaaatacagtgctcctgaagttcaagaattacaagaatggctgcttcttggcattgcaaca
aaaccaccattaccaatctggctacattgacaaggctaacgtgcttgccctctccagattgtggagaaagactgttgaagaaccatcaagactgagttccacaaftaact
gttcagcaccaattgatcgattctgctgctatgatttggtaaatcaccaactaagctaaacgggtgtgtattaccaacaacatgttgggtgatattatctccgatgaagcctc
tgttattccaggttcttgggttattaccttctgcatctctagcttccctacctgacactaacaaggcattcggttgtacgaaccatgcatggttctgccccagattaccagc
aaacaaggftaaccaattgctaccatcttatctgcagctatgatgttgaagttatccttggatttgggtgaagaaggtagggctcttgaagaagctgttagaaatgtcttggga
tgcaggtgtcagaaccgggtgaccttgggtggttctaactctaccactgaggttggcgatgctatcgccaaggctgtcaaggaaatcttggcttaaacagctactgacaataaa
aagattcttgtttcaagaacttgcatttgtatagttttttatattgtagtgttctattttaaatacaaatgttagcgtgatttataatttttgcctcgacatcatctgccagatgcg
aagttaagtgcgcagaaagtaatatcatgcgtcaatcgatgtgaatgtgtgctgctatactgctgctgattcgatactaacgccgccatccagtgctgacctccattactc
gtatcgcatgtcgggtgcgacacgaaattacaaaaaggaatatgttcatagggtagacgaaactatatacgcaatctacatacatattatcaagaaggagaaaaaggaggat
gtaaaggataacaggtgaagcaaatgatactaatggctcaacgtgataaggaaaaagaattgcactttaacattaattgacaaggaggaggccaccacacaaaaagtt
aggtgtaacagaaaatcatgaaactatgattcctaatttatatttggaggattttcttaaaaaaaaaatacaacaaaaaaacactcaatgacctgaccatttgatg
gagtttaagtcaataccttcttgaaccatttcccataatgtgaaagtccctcaagaattttactctgtcagaaacggccttaacgacgtatgcgacctctctcagtactaa
atctaccaataccaaatctgatggaagaatgggctaattgcatcatccttaccagcgagagacctatcagctcactcaaggcggtaatacggttatccacagaatcagg
ggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttctgtggcggttttccataggtcggccccctgacgag
catcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcggttccccctggaaagctccctctgctgctctctgttccgacc
tgccgcttacccgatactgtccgcttctcccttccgggaagcggtggcgcttctcatagctcagctgttaggtatctcagttcgggtgtaggtcgttctcctcaagctggg
ctgtgtgcacgaacccccgttcagcccgaccgctgcgccttatccggttaactatcgtctttagcccaacccggtaagacacgacttatccactggcagcagccact
ggtaacaggattagcagagcgaggtatgtaggcgggtgtacagagttcttgaagtgggtggcctaactacggctacactagaagaacagatttgggtatctgcgctctgt
gaagccagttaccttcgaaaaagagttggtagctcttgatccggcaacaaaccaccgctggtagcgggtgtttttgttgaagcagcagattacgcgcagaaaaa
aaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcaggttaagggttttgggtcatgagattatcaaaaaggatcttcacctag
atccttttaaataaaaatgaagttttaaatcaatctaaagtatatatgagtaaaacttggctgtacagagttctgaggtcattactggatctatcaacagcagtcgaagcgagct
cgatatcaaatcagccccgcccgtccactcatcgagttgttgaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgcca
gcggcatcagcacttgtgccttgcgtataatattgcccatggtgaaacggggggaagaagttgtccatattggccacgtttaaataaaaactggtgaaactcacc
agggttgggtgagacgaaaaacatattctcaataaacctttagggaaatagggcaggtttaccgtaacacgccacatctgcgaatatatgtgtagaactgccgga
aatcgtcgtggtattcactccagagcgatgaaaacgttcagtttgcctatggaacgggtgaacaagggtgaacactatccatcaccagtcaccgtcttcttgc
catacgaaatccggatgagcattcatcagcggggcaagaatgtgaataaaggccggataaaacttgccttattttttacggctttaaaggccgtaatatccagct
gaacggctgtgttataggctacattgagcaactgactgaaatgcctcaaatgttcttaccgatgccattgggatatacaacgggtgttatccagtgattttttctcatttta
gttctctagctcctgaaaatctcgataactcaaaaaatagcccggtagtgatcttattcattatggtgaaagttggaacctttacgtgcccgatcaactcgcgcgttgc
cacctgacgtctaagaaaaggaattatcagcaatttgcctgccgaagaaaggccaccgtgaaggtgagccgggtctgacccaga

SEQ ID NO: 6 LXR370:

agaacagtaaaaataaagcaaggtacgtgaaattaatatttttaaatggttctaaccgatgccgaagaactgcgcagtcgggtataacgtctgacatgtcctttttgatttgg
aatccaaccactcaagtgactctgttctatttcttgcgaaaaataaccacaaatgcccatcgaaagtgaatcgcaaacaccttcagactggcaccgacaaagcaaa
gattatcacagacagagtatttatcgtacggttaagtctcaagcaaaagggtttcttatttactgaacgggtgaagaagtatctgggcccggcttgcgaagatgcaaacggaa
taagtatttcaaaagttgcatttgccttagccgtcctgacaccattggctatttggatattttatattgacttctgtgtacattgatcacatcgactgttctatttggcaaatgaaccac
gggcattgactattttcagggttactatataattatcatcacgggcaaggattgtagcttgcctatgctgcggaggagataatttacttttatttattctagttttttacagttatt
attaattaattttttatagcatgcgaataaaaagctatatttaagtcttttatttataatacatcttctctacgagctgtcacccggatgtgcttccgggtctgatgagtcggtg
aggacgaaacagcctctacaataattttgttaagagcaggttgttcatggccgtgctgtatgatgtggggggctcggcggttgaacccggggttcggagcgccaggg

ggttttcaacggcctagcatgtgattaatfaattatgttttttttgcagtataaaaaagtagttgtttaacaacaaacfttttcatftttttgttcccccttctcttttagtta
gtttgtttaacaacaaactagaatatcaagctacaaaaataataaaaaATGAGGTCTTCTTATATTCCCCAGAACATCAGGTTAAT
GGCGTTTTTGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAAACCGGCA
CACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTTCATCCCCGATATGCACCACCGGGTAAAGTTCA
CGGGAGACTTTATCTGACAGCAGACGTGCACTGGCCAGGGGGATCACCATCCGTCGCCCCGGGCGTGT
CAATAATATCACTCTGTACATCCACAAACAGACGATAACGGCTCTCTCTTTTATAGGTGTAAACCTTA
AACTGCATTTACCAGCCCCTGTTCTCGTCAGCAAAAAGAGCCGTTTCAATAAACCGGGCGACCT
CAGCCATCCCTTCTGATTTTCCGCTTTCCAGCGTTTCGGCACGCAGACGACGGGCTTCAATTCTGCATGG
TTGTGCTTACCAGACCGGAGATATTGACATCATATATGCCTTGAGCAACTGATAGCTGTCGCTGTCAA
CTGTCACTGTAATACGCTGCTTCATAGCATACTCTTTTTGACATACTTCGGGTATACATATCAGTATA
TATTCTTATACCGCAAAAATCAGCGCGCAAATATGCATACTGTTATCTGGCTTTTAGTAAGCCGAAGA
CCTggtagccctaagaaaaagagaaaagtggtggcagtgccggaagcgggggatcaggtggttctggagggtccGATATCATGCAGGATCTG
CCGGGCAACGATAACAGCACCCGCGGGCGCGGTACGAAAAACAATTACGGGTCTACCATCGAGGGC
CTGCTCGATCTCCCGGACGACGACGCCCCCGAAGAGGCGGGGCTGGCGGCTCCGCGCCTGTCCTTTCT
CCCCGCGGGACACACGCGCAGACTGTCGACGGCCCCCCCCGACCGATGTCAGCCTGGGGGACGAATTG
CACTTAGACGGCGAGGACGTGGCGATGGCGCATGCCGACGCGCTAGACGATTTTCGATCTGGACATGT
TGGGGGACGGGGATTCCCCGGGTCCGGGATTTACCCCCACGACTCCGCCCCCTACGCGCCTCTGGAT
ATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGAATTGACGAGTACGGTGGGAAGCT
TTAAtaaaagctttgattaagcctttagtceaaaaaacgftttttgtcatttattcattttctagaatgttagtttattcattttatagtcacgaatgtttatgattctat
atagggttgcaacaagcatttttcatttatgttaaaacaattcaggtttaccttttattctgcttggtgacgcgtgatccgcccgctcttttggtcaccatgtatttaattg
cataaataattcttaaagtgagctagcttatttcatatacctctcatttctcatttctcctgaccagctgagtttcatggttgaattactatgtcggggtaattgtggc
ttgtcgtagccttggtgaagtatttggcacaaggtctagggcctgttgaggagtgtagcgttgccttgccccgcgggtcaccggccagcgacatggaggccaga
ataccctccttgacagcttgacgtgcgcagctcaggggcatgatgtgactgtcggcgtagcatttagccatacatcccatgtataatcattgcatccatacatttgatg
ggcgacggcgcaagcaaaaattacggctcctcgtcgcagacctgcgagcagggaacgctccccacagacgcgttgaattgccccacggcgccccctgta
gagaaatataaaagggttaggattgccactgaggttcttcttcatatacttctttaaactctgtaggatacagttctcacatcacatccgaacataaacaacatgtctaa
gaatatcgttgcctaccgggtgatcacgtcggttaagaagttactgacgaagctattaaggtctgaatgccattgctgaagtccttcagaaattaagtcaattccaac
atcattgatcgggggtgctgccatcgatgccactggcactccttaccagatgaagctctagaagcctctaagaaagccgatgctgttactaggtgctgttggtggtc
caaatggggtagggcgagtttagaccagaacaaggtctattgaagatcagaaggaattgggtctatagccaacttgaggcatgtaactttgcttctgattcttact
agatcttctccttgaagcctgaatatgcaagggtaccgatttctcgtcgttagagaattgggtggtggtatctactttggtgaaagaaaagaagatgaaggtgacgga
gttgcttgggattctgagaaatacagtgctcctgaagttcaagaattacaagaatggctgcttcttggcattgcaaaaaaccaccattaccaatctggcacttgacaa
ggctaacgtgcttgcctctccagattgtggagaaagactgtgaagaaccatcaagactgagttcccacaattaactgttcagcaccaattgatcgattctgctgtatga
tttgggtaaatcaccaactaagctaaacgggtgttattaccaacaacatgttgggtgatattatctccgatgaagcctctgtattccaggttcttgggttattaccttctgca
tctctagcttccctacctgacactaacaaggcatcgggttgcgaaccatgtcatgggtctgccccagattaccagcaacaagggttaaccaattgtaccatcttatct
gcagctatgatgtgaagttatccttgatttgggtgaagaaggtagggctcttgaagaagctgttagaaatgtcttgatgcaggtgtcagaaccgggtgaccttgggtggt
ctaactctaccactgaggttggcgatgctatcgcaaggctgtcaaggaaacttggcttaaacagtactgacaataaaaagattctgtttcaagaactgtcatttgata
gttttttatattgtagttgttctattttaaataatgttagcgtgatttataatttttccgctcgacatcatctgccagatgcgaagttaaagtcgcagaaagtaatatcatgcgt
caatcgatgtgaatgctggctgctatctgctgtcgaatcgatactaacgcgcgcatccagtgctgacctccattactcgatcgatgctgggtgcgacacgaataacaa
aatggaatatgttcatagggtagacgaactatatacgcaatctacatacatattatcaagaaggagaaagggaggtgtaaaaggaatacaggtgaagcaaatgatacta

atggctcaacgtgataaggaaaaagaattgcactttaacattaattgacaaggaggaggcaccacacaaaaagttagggtgaacagaaaatcatgaaactatgattc
ctaatttatattggaggattttcttaaaaaaaaaaatacaacaaataaaaaaactcaatgacctgaccatttgatggagttaagtaataccttctgaaccatttccc
ataatgggtgaaagttccctcaagaattttactctgtcagaaacggccttaacgacgtagtcgacctctcttcagtactaaatctaccaataccaaatctgatggaagaatgg
gctaatgcatcatccttaccagcgagagacctatcagctcactcaaaaggcggtatccacagaatcaggggataacgcaggaaagaacatgtgagcaaa
aggccagcaaaaggccaggaaccgtaaaaaggccggtgtgctggcggttttccataggtccgccccctgacgagcatcacaaaaatcgacgtcaagtcagaggt
ggcgaaaccggacaggactataagataaccaggcggttccccctggagctccctctgctgctctctgttccgacctgcccgttaccggatacctgtccgctttctc
ccttcgggaagcggtggcggtttctcatagctcagcgtgtaggtatctcagttcgggtgtaggtcgttctccaagctgggctgtgtgcagcaacccccgttcagccga
ccgctgcgccttatccggttaactatcgtcttgagcccaaccggtaagacacgacttatccactggcagcagccactggttaacaggattagcagagcgaggtatgta
ggcggtgtacagagttctgaagtgggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgtgaagccagttaccttcgaaaaagagttggt
agctcttgatccggcaacaaaccaccgctggtagcgggtgtttttgttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttcta
cgggggtctgacgtcagtggaacgaaaactcacgttaagggttttggctatgagattatcaaaaaggatcttcacctagatcctttaaftaaaaatgaagtttaaatca
atctaagtatatagtaaaacttggtctgacagagttctgaggtcattactggatctatcaacagcagtcgaagcagctcgatatcaaatcagccccgcctgccact
catcgcagtactgttgaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatgccagcggcatcagcacttgcgccttgcgtat
aatatttgcccatggtgaaaaacggggcggaagaagttgtccatttgccacggttaaatcaaaaactggtgaaactcaccagggttggtgagacgaaaaacatattc
tcaataaaccttttagggaaataggccaggtttaccgtaacacgccacatcttgcgaatatatgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgat
gaaaacgtttcagtttgcctatggaaaacgggtgtaacaagggtgaacactatcccatatcaccagctcaccgtcttctcattgccatacgaattccggatgagcattcatca
ggcgggcaagaatgtgaataaaggccggataaaactgtgcttattttctttacggcttttaaaaaggccgtaatatccagctgaacgggtcgtgtataggtacattgagca
actgactgaaatgcctcaaaatgttctttacgatgccattgggtatatacaacgggtgtatatccagtgattttttctccatttttagcttcttagctctgaaaaatctcgataact
caaaaaatacgcgggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaactcgcgcgtttgccacctgacgtctaagaaaaggaatattca
gcaatttgcccggtgccgaagaaggcccccgtgaagggtgagccggtctcgaccaga

SEQ ID NO: 7 lexAec₁₋₈₇ AAs:

MSKALTARQQEVFDLIRDHISQTGMPPTRAEIAQRLGFRSPNAAEEHLKALARKGVIEIVSGASRGIRLLQE
EEEGLPLVGRVAAGEP

SEQ ID NO: 8 lexAbs₁₋₉₄ AAs:

MSTKLSKRQLDILRFIKAIEVKSKGYPPSVREIGEAVGLASSSTVHGHLLARLETKGLIRRDPTKPRAIEILDEE
VDIPQSQVVNVPVIGKVTAGSP

SEQ ID NO: 9 CI₁₋₉₄ AAs:

MSTKKKPLTQEQLDARRLKAIYEKKKKNELGLSQESVADKMGMGQSGVGALFNGINALNAYNAALLAKI
LKVSVEEFSPSIAREIYEMYEAVSM

SEQ ID NO: 10 CI₄₃₄₋₇₀ AAs:

MSISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRPRFLPELASALGVSDWLLNGTSDSNVR

SEQ ID NO: 11 HKCI₁₋₈₄ AAs:

MVQQKERETFSQRLALACDKAGLPLHGRQADLAVRLKVTPKAISKWFNGESIPRKDKMESLASVLGTTA

AYLHGYADDDGITVN

SEQ ID NO: 12 PurR₁₋₆₀ AAs:

MATIKDVAKRANVSTTTVSHVINKTRFVAEETRNAVWAAIKELHYSPSAVARSLKVNHTK

SEQ ID NO: 13 DeoR₁₋₉₂ AAs:

METRREERIGQLLQELKRSDKLHLKDAAALLGVSEMTIRRDNLNHSAPVVLLGGYIVLEPRSASHYLLSDQ
KSRLVEEKRRAAKLAATLVEP

SEQ ID NO: 14 lexA₅₁₋₁₀₄ AAs:

MSENTNEKRKEMTARQEEIYEYIKKYSKENHMPPTVREIGNHFDISSTNGVRSILAALIKKGYINRSPRLSR
GIEILSDDKESKEVASNTIEIPIVGRVAAGTP

SEQ ID NO: 15 lexA₅₁₋₉₁ AAs:

MSETLTSRQRTVLEFITAHVDRHGYPTMREIARHLNVNGTLGVAKHLEALARKGYLQREPGNSRGITLT
GQTRHTPTVSLPVIGVVRAGVP

SEQ ID NO: 16 lexA_{mm1-115} AAs:

MSTDREKEHKMKPGRRPTEGMTSPQERIMQVIRNAIDRDGLPPTVKEIAEALGMKTPSAHEQVQKLVKKGF
IRRTPRKARSIEIVEQSPEEEPVEKKPDVARLVPVPIIGEVAAGIP

SEQ ID NO: 17 lexA_{α1-101} AAs:

MSDLTDTQQAILALIAERIDADGVPPSQTEIARAFGFKGIRAAQYHLEALEHAGAIRRVPGQARGIRLAGQ
GAQTRTAPVSEVARDDVLRLPVLGRVAAGLP

SEQ ID NO: 18 RecA_{pact1-104} AAs:

MQLVTSNIRHLRKHAGFTQAQLAEKLEIKRSLVGAYEEGRAEPLSTLVNVAKLFSVSLDDLVTDLQRE
GAAGGADRSSGGNLRVLAITVDQDEKENIELVPY

SEQ ID NO: 19 acVHH AAs:

GSQVQLVESGGGLVQAGGSLRLSCTASGRTGTIYSMAWFRQAPGKEREFATVGVSSGITYYMDSVKGR
FTISRDKGKNTVYLQMDSLKPEDTAVYYCTATRAWSVGVDYWGQGTQVTVSS

SEQ ID NO: 20 RpaR₁₋₁₇₉ AAs:

IVGEDQLWGRRTLEFVDSVERLEAPALISRFESLIASCGFTAYIMAGLPSRNAGLPELTLANGWPRDWFDL
YVSENFSAVDPVPRYGATTVHPFVWSDAPYDRDRDQAAHRVMTRAAEFGLVEGYCIPLHYDDGSAAISM
AGEDPDLSPAARGVMQLVSIYHSRLRVLSRPKPIRRN

SEQ ID NO: 21 BjaR₂₋₁₈₀ AAs:

SAVDYGREALDFIEGLGVYRKVPDAMNALEAAFGFRGFETIIVTGLPNPDQRFAQMVLAKRWPAGWFNL
YTQNNYDRFDPVVRLCRQSVNPFWESEAPYDAELEPSAAEVMNRAGDFRMSRGFIVPIHGLTG YEAAVSL
GGVHLDLNPRSKPALHLMAMYGFDHIRRLLEPTYPSTRL

SEQ ID NO: 22 CinRori₂₋₁₇₉ AAs:

IENTYSEKFESAFEQIKAAANVDAAIRILQAEYNLDFVTYHLAQTIAASKIDSPFVRTTYPDAWVSRYLLNSY
VKVDPIVKQGFERQLPFDWSEVEPTPEAYAMLVDAQKHGIGGNGYSIPVADKAQRRALLSLNARIPADEW
TELVRRCRNEWIEIAHLIHRKAVYELHGENDPVPAL

SEQ ID NO: 23 LasR₂₋₁₇₇ AAs:

ALVDGFLELERSSGKLEWSAILQKMASDLGFSKILFGLLPKDSQDYENAFIVGNYPAAWREHYDRAGYAR
VDPTVSHCTQSVLPFWEPSIYQTRKQHEFFEEASAAGLVYGLTMPLHGARGELGALSLSVEAENRAEANR
FMESVLPTLWMLKDYALQSGAGLAFEHPVSKPVVL

SEQ ID NO: 24 TraR₁₋₁₇₄ AAs:

MQHWLDKLTDLAAIEGDECILKTGLADIADHFGFTGYAYLHIQHRHITAVTNYHRQWQSTYFDKKFEAL
DPVVKRARSRKHIFTWSGEHERPTLSKDERAFYDHASDFGIRSGITIPKTANGFMSMFTMASDKPVIDLDR
EIDAVAAAATIGQIHARISFLRTTPTAEDAAWL

SEQ ID NO: 25 CarRecc₁₋₁₆₉ AAs:

MDHEIHSFIKRLKGVGDVWFSYFMMSKNSTSQPYIISNYPEAWMKEYIKKEMFLSDPIIVASLARITPFSW
DDNDIVTLRAKNQDVFISVQHDISSGYTFVLHDHNNVATLSIANHLEDANFEKCMKNHENDLQMLLV
NVHEKVMAYQRAINDQDNPPDNSRNALL

SEQ ID NO: 26 SmaR₁₋₁₇₉ AAs:

MSNSFFNNTSINISIKNYLEKNLKVFNNIKYAYAIMNKKNPNDFAIISNRMEWFDFYTKNNLQFIDPVLITAS
CCFTPFLWDENIMISSGLKMPKIFNMAKNYDVINGYTFVLHDHNNLVLVLSIIMDKSCDDDIEKIIVDKKN
DLQMLLLTTHEKLITLYQEINDTHQFNKKNQKEIL

SEQ ID NO: 27 ER₂₈₂₋₅₉₅ AAs:

SAGDMRAANLWPSPLMIKRSKKNLALSALTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLAD
RELVHMINWAKRVPGFVDLTLHDQVHLLCAWLEILMIGLVWRSMEHPVKLLFAPNLLLDNRNQGKCV
MVEIFDMLLATSSRFRMMNLQGEFVCLKSIILLNSGVYTFLLSSTLKSLEEKDHIHRVLDKITDTLIHLMAK
AGLTLQQQHQRQAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRLHAPTSRGGASVE
ETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV

SEQ ID NO: 28 DHBR₂₈₂₋₅₉₅ AAs:

SAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLITNMD
RELVHMINWAKRVPGFVDLTLHDQVHLLCAWLEILQIGLVWRSMHPVKLLFAPNLLDRNQGKCVEG
MVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAK
AGLTLQQQHQRQAQLLLILSHIRHMSNKSMEHLDSMKCKNVVPLYDLLLEMLDAHRLHAPTSRGASVE
ETDQSHLATAGSTSSSHSLQKYYITGEAEGFPATV

SEQ ID NO: 29 PR AAs:

AGQGDPVRVVRALDAVALPQPVGVPNESQALSQRFTFSPGQDIQLIPPLINLLMSIEPDVIYAGHDNTKPDT
SSSLLTSLNQLGERQLLSVVKWSKSLPGFRNLHIDDQITLIQYSWMSLMVFGLGWRSYKHVSGQMLYFAP
DLILNEQRMKESFYSLCLTMWQIPQEFVKLQVSQEEFLCMKVLLLLNTIPLEGLRSQTQFEEMRSSYIRELI
KAIGLRQKGVVSSSQRFYQLTKLLDNLHDLVKQLHLYCLNTFIQSRALSVEFPPEMMSEVIAGSTASGRAGA
PPTDVSLGDELHLDGEDVAMAH

SEQ ID NO: 30 GR₄₈₇₋₇₇₇ AAs:

NLEARKTKKKIKIGIQQATTGVSQETSENPKNKTIVPATLPQLTPTLVSLLEVIEPEVLYAGYDSSVPDSTWR
IMTTLNMLGGRQVIAAVKWAKAIPGFRNLHLDQMTLLQYSWMFLMAFALGWRSYRQPSANLLCFAPD
LIINEQRMTLPCMYDQCKHMLYVSSELHRLQVSYEEYLCMKTLLLLSSVPKDGLKSQELFDEIRMTYIKEL
GKAIVKREGNSSQNWQRFYQLTKLLDSMHEVVENLLNYCFQTFLDKTMSEIPEMLAEIITNQIPKYSNGN
IKKLLFHQK

SEQ ID NO: 31 MR₆₆₉₋₉₈₄ AAs:

NLGARKSKKLKGLKGIHEEQPQQQPPPPPPPPQSPEEGTTYIAPAKEPSVNTALVPQLSTISRALTSPVMV
LENIEPEIVYAGYDSSKPDTAENLLSTLNRLAGKQMIQVVKWAKVLPGFKNLPLEDQITLIRYSWMCLSSF
ALSWRSYKHTNSQFLYFAPDLVFNEEKMHQSAMYELCQGMHQISLQFVRLQLTFEEYTIMKVLLLLSTIP
KDGLKSQAAFEEMRTNYIKELRKMVTKCPNNSGQSWQRFYQLTKLLDSMHDLVSDLLEFCFYTFRESHA
LKVEFPAMLVEIISDQLPKVESGNAKPLYFHRK

SEQ ID NO: 32 D1+ AAs:

STKGVALDDVEGDDDFYMFDDIIASRSTG

SEQ ID NO: 33 D2+ AAs:

NNGTSIEGCSENIDVVLEADLDLPLDYLV

SEQ ID NO: 34 E1+ AAs:

RSARDDGVGEEEEDDYGVYYDFDEWFDENS

SEQ ID NO: 35 haB112 AAs:

DIMQDLPGNDNSTAGEFPGITLRIQETDMLYKGDITLYLDWLEDGIAELVFDAPGSVNKLDTAVASLGAEIG

VLEQQSDLIWETLTVKDAKVNFDSGLEKFEEAIPSAADD FDPVAERRSSGEFRAERHSGGTDLCKFL

SEQ ID NO: 36 GAL4 AAs:

NFNQSGNIADSSLSFTFTNSSNGPNLITTQTN SQALSQPIASSNVHDNFMNNEITASKIDDGNN SKPLSPGWT
DQTAYNAFGITTGMFNTTTMDDVYNYLFDD EDDTPPNPKKE

SEQ ID NO: 37 rTA AAs:

RDSREGMFLPKPEAGSAISDVFE GREVCQPKRIRPFHPPGSPWANRPLPASLAPTPTGPVHEPVGSLTPAPV
PQPLDPAPAVTPEASHLLED PDEETSQAVKALREMA DTVIPQKEEAICGQMDLSHPPPRGHLDELTTTLE
SMTEDLNLDSPLTPELNEILD TFLNDECLLHAMHISTGLSIFDTSLF

SEQ ID NO: 38 GCN4 AAs:

MFEYENLEDNSKEWTS LFDNDIPVTTDDVSLADKAIES

SEQ ID NO: 39 Msn2 AAs:

MTVDHDFNSEDILFPIESMSSIQYVENNNPNNINNDVIPYSLDIKNTVLD SADLNDIQNQETSLNLGLPPLSF
DSPLPVTETIPSTTDNSLHLKADSNKNRDARTIENDSEIKSTNNASGSGANQYTTLTSPYPMNDILYNMNNP
LQSPSPSSVPQNPTINPPINTASNETNLSPQTSNGNETLISPRAQQHTSIKDNRLSLPNGANSNL FIDTNPNNL
NEKLRNQLNSDTNSYSNSISNSNSNSTGNLNSSYFNSLNIDSMLDDYVSSDLLLNDDDDDTNLSRRRFSDVI
TNQEL

SEQ ID NO: 40 VP16 AAs:

MIMQDLPGNDNSTAGARTKNNYGSTIEGLLDLPDDDAPEEAGLAAPRLSFLPAGHTRRLSTAPPTDVSLG
DELHLDGEDVAMAHADALDDFDLDM LGDGDSPPGPGFTPHDSAPYGALDMADFEFEQMFTDALGIDEYG
GKL

以上所揭露的仅为本发明较佳实施例而已，当然不能以此来限定本发明之权利范围，因此依本发明权利要求所作的等同变化，仍属本发明所涵盖的范围。

权利要求书

1. 一种模块化正交真核转录因子模型构建方法，包括以下步骤：

S1、确立由 DNA 结合结构域、配体结合结构域、转录激活结构域三种结构域组成的转录因子基本架构 DBD-LBD-AD；

S2、在转录因子基本架构 DBD-LBD-AD 基础上，分别构建 DNA 结合结构域、配体结合结构域、转录激活结构域三种结构域的筛选酿酒酵母菌株，然后对筛选酿酒酵母菌株进行筛选测试培养，筛选得到有效的 DBD、LBD 和 AD；

S3、在转录因子基本架构 DBD-LBD-AD 基础上，以筛选出的 DBD、LBD 分别构建用于结构域参数定量测试的定量测试酿酒酵母菌株；对定量测试酿酒酵母菌株进行定量测试培养，获得转录因子表达量下的调控诱导范围和调控倍数；对获得调控诱导范围和调控倍数数据进行拟合，得到筛选出的 DBD、LBD 各结构域的对应该参数；

S4、根据 DBD、LBD 各结构域的对应该参数来表征转录因子调控过程，建立转录调控定量模型；通过转录调控定量模型预测不同 DBD 和 LBD 组合情况下转录因子，获得性能优化的转录因子。

2. 根据权利要求 1 所述模块化正交真核转录因子模型构建方法，其特征在于，所述基本架构 DBD-LBD-AD 中以 *E. coli* SOS 调控蛋白 *lexA* 的 1-87aa 作为 DBD，以沼泽红假单胞菌群体感应转录因子 *RpaR* 的 1-179aa 作为 LBD，以 VP16 作为 AD。

3. 根据权利要求 2 所述模块化正交真核转录因子模型构建方法，其特征在于，S2 中 DNA 结合结构域的筛选酿酒酵母菌株，通过以下方式得到：构建 DBD 的筛选转录因子表达载体，构建对应的报告基因表达载体，对报告基因表达载体进行酶切，得到转录因子表达片段，将转录因子表达片段整合到酵母中，得到 DBD 筛选酿酒酵母菌株。

4. 根据权利要求3所述模块化正交真核转录因子模型构建方法,其特征在于,所述 DBD 的筛选转录因子表达载体序列如 SEQ ID NO: 1; 所述筛选得到有效的 DBD 序列如 SEQ ID NO: 7-18。

5. 根据权利要求3或4所述模块化正交真核转录因子模型构建方法,其特征在于,S2中配体结合结构域的筛选酿酒酵母菌株,通过以下方式得到:构建 LBD 的筛选转录因子表达载体,构建对应的报告基因表达载体,对报告基因表达载体进行酶切,得到转录因子表达片段,将转录因子表达片段整合到酵母中,得到 LBD 筛选酿酒酵母菌株。

6. 根据权利要求5所述模块化正交真核转录因子模型构建方法,其特征在于,所述 LBD 的筛选转录因子表达载体序列如 SEQ ID NO: 2; 所述筛选得到有效的 LBD 序列如 SEQ ID NO: 19-31。

7. 根据权利要求1所述模块化正交真核转录因子模型构建方法,其特征在于,S3中DNA结合结构域的定量测试酿酒酵母菌株,通过以下方式得到:构建 DBD 的参数定量测试转录因子表达载体,然后构建对应的报告基因载体,对报告基因表达载体进行酶切,得到转录因子表达片段,将转录因子表达片段整合到酵母中,得到 DBD 定量测试酿酒酵母菌株。

8. 根据权利要求7所述模块化正交真核转录因子模型构建方法,其特征在于,所述 DBD 的参数定量测试转录因子表达载体序列如 SEQ ID NO: 4。

9. 根据权利要求7所述模块化正交真核转录因子模型构建方法,其特征在于,S3中配体结合结构域的定量测试酿酒酵母菌株,通过以下方式得到:构建 LBD 的参数定量测试转录因子表达载体,然后构建对应的报告基因载体,对报告基因表达载体进行酶切,得到转录因子表达片段,将转录因子表达片段整合到酵母中,

得到 LBD 定量测试酿酒酵母菌株。

10. 根据权利要求 9 所述模块化正交真核转录因子模型构建方法，其特征在于，所述 LBD 的参数定量测试转录因子表达载体如 SEQ ID NO: 5。

11. 根据权利要求 1 所述模块化正交真核转录因子模型构建方法，其特征在于，所述转录调控定量模型包括非核受体二聚化转录激活模型和核受体类二聚化转录激活模型；

非核受体二聚化转录激活模型方程如下：

$$I(x) = \frac{c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1}) \cdot I_{\max} + I_0,$$

$$F_1 = -k_B T \left[\ln\left(\frac{L_{tot}}{2}\right) + \ln\left(\frac{\sqrt{(k_2 I + 1)^2 + 8L_{tot}(k_2^2 k_3 I^2 + k_1)} - (k_2 I + 1)}{\sqrt{(k_2 I + 1)^2 + 8L_{tot}(k_2^2 k_3 I^2 + k_1)} + (k_2 I + 1)}\right) \right] + \Delta \varepsilon_{DBD};$$

核受体类二聚化转录激活模型方程如下：

$$I(x) = \frac{c_{[LBD^*_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}}{1 + c_{[LBD^*_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}} + c_{[LBD^*_2:I_2]} \cdot e^{-\beta \Delta \varepsilon_{DBD}}} \cdot I_{\max} + I_0 = (1 + e^{-\beta F_1})^{-1} \cdot I_{\max} + I_0,$$

$$F_1 = \left[-k_B T \cdot \ln\left(\frac{L_{tot}}{2}\right) \right]$$

$$+ \left[-k_B T \ln\left(\frac{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8L_{tot}(k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} - (k_2 I + k_* + k'_* k_2 I + 1)}{\sqrt{(k_2 I + k_* + k'_* k_2 I + 1)^2 + 8L_{tot}(k_1 + k_3 k_2^2 I^2 + k_1 k_*^2 + k_3 k_2^2 k_*'^2 I^2)} + (k_2 I + k_* + k'_* k_2 I + 1)}\right) \right]$$

$$+ \Delta \varepsilon_{DBD};$$

其中，F1 因子为描述特定 DBD-LBD 组合下转录因子性能的参数；c 为 LBD 不同形态下的浓度，I₀ 为特定序列启动子的本底泄露表达强度，I_{max} 为特定序列启动子的理论最大激活强度，I 为诱导剂浓度，L_{tot} 为所有形式的转录因子的表达量

的单体形式的总和， $\Delta\mathcal{E}_{DBD}$ 描述 DBD 的行为，K1、K2、K3 描述 LBD 的行为，K*和 K*'描述转录因子的入核过程。

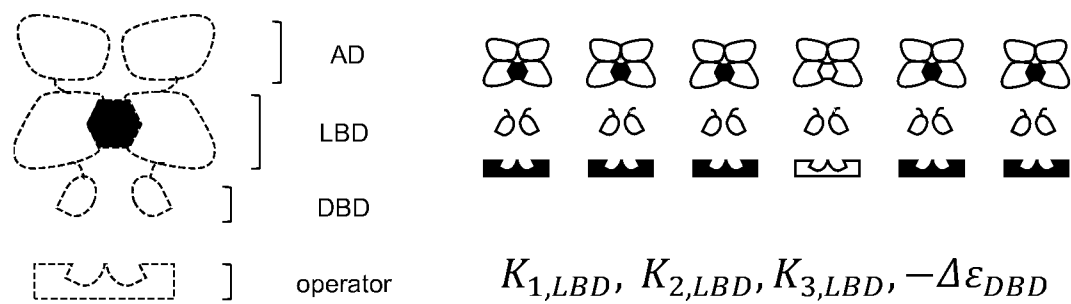


图 1

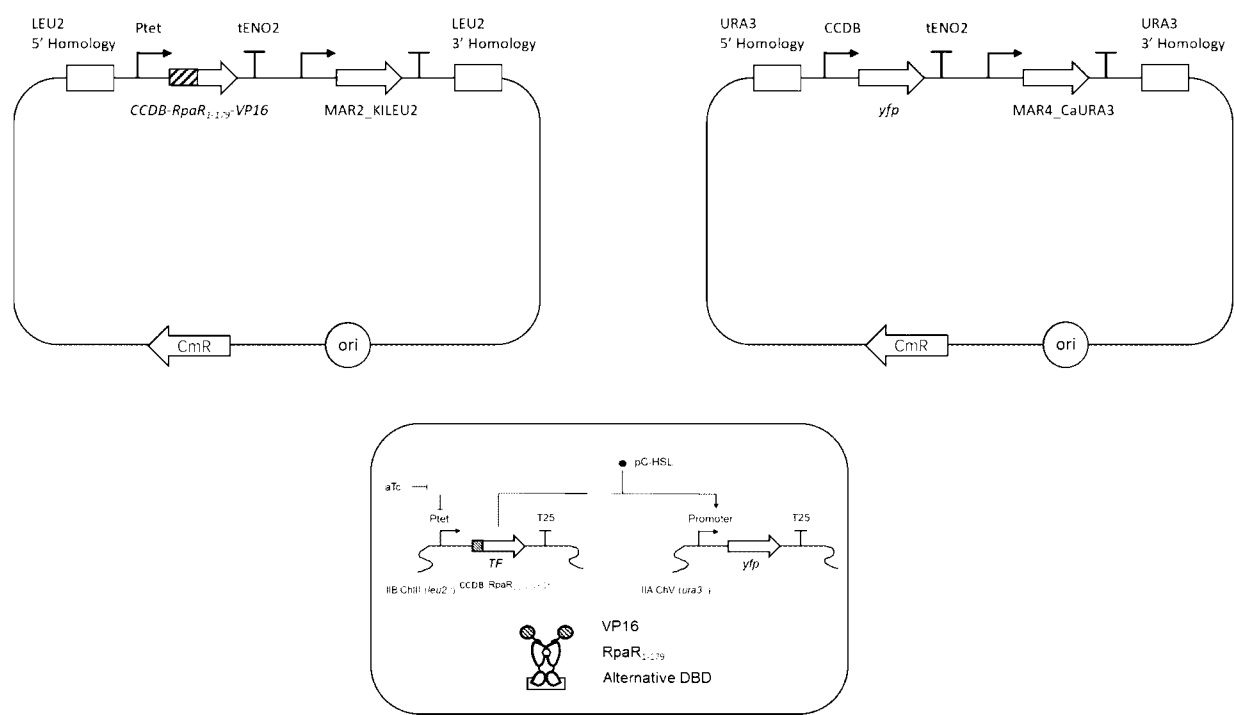


图 2

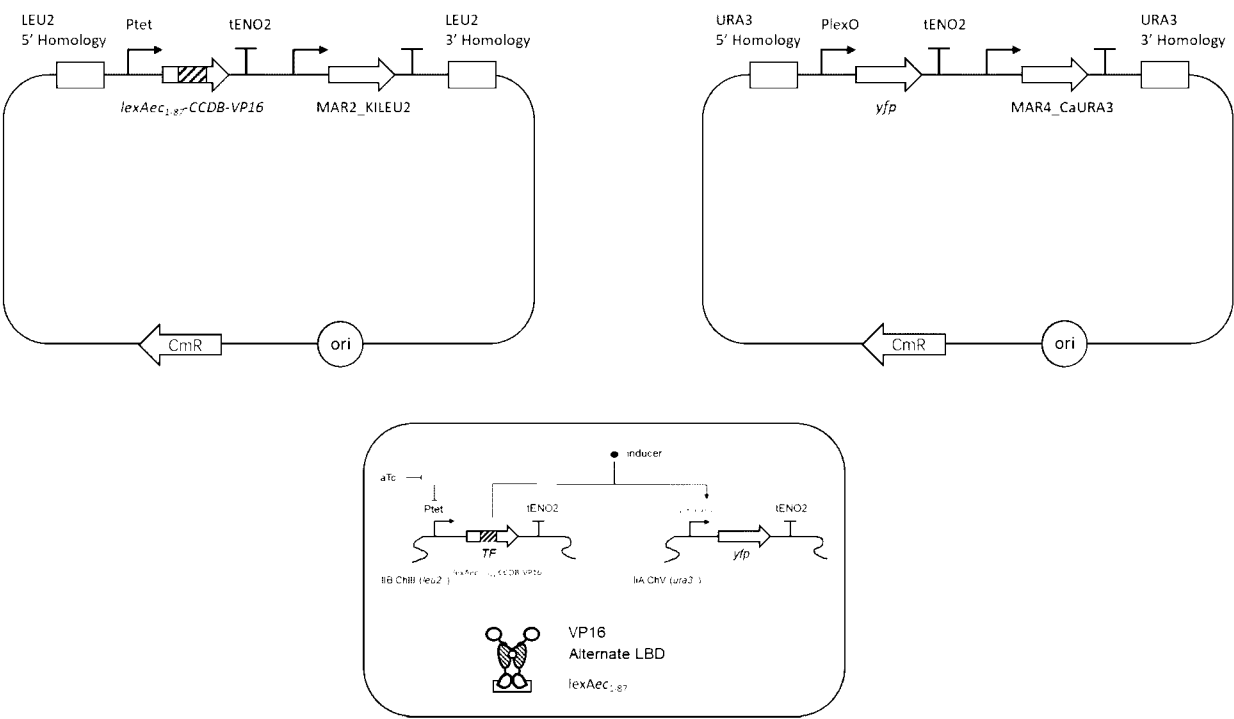


图 3

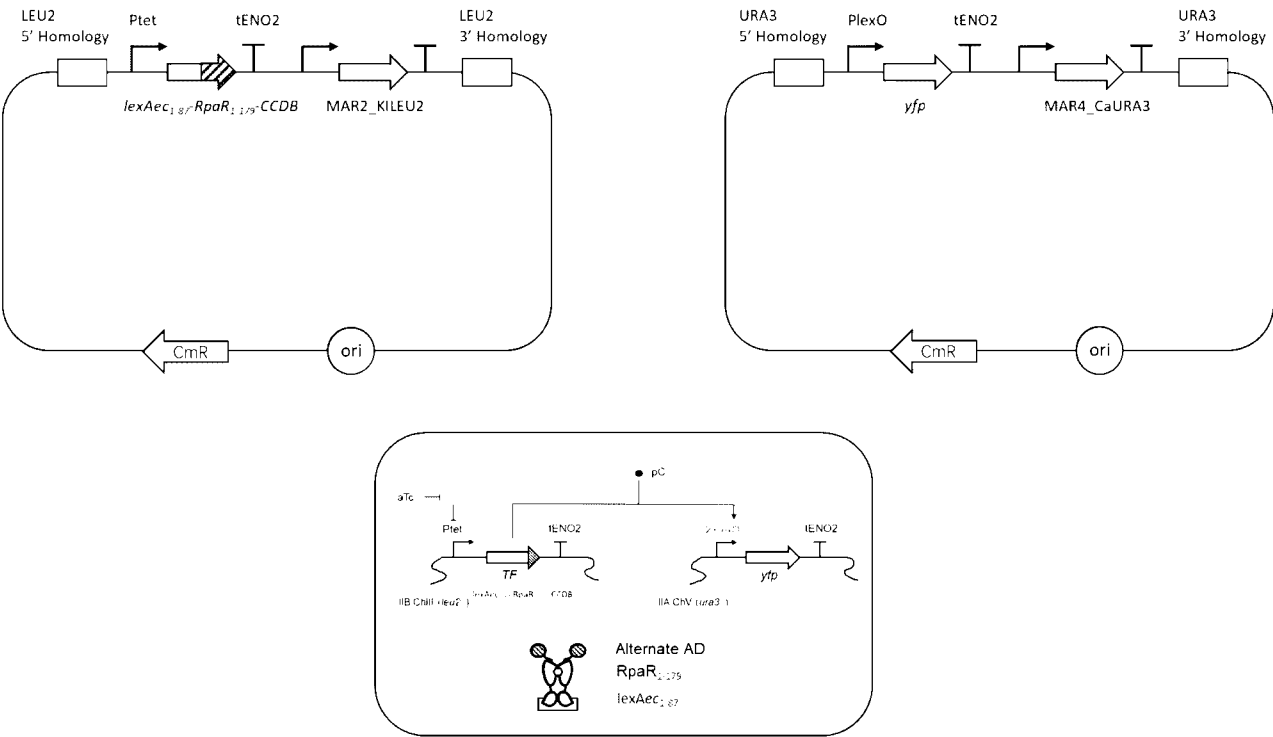


图 4

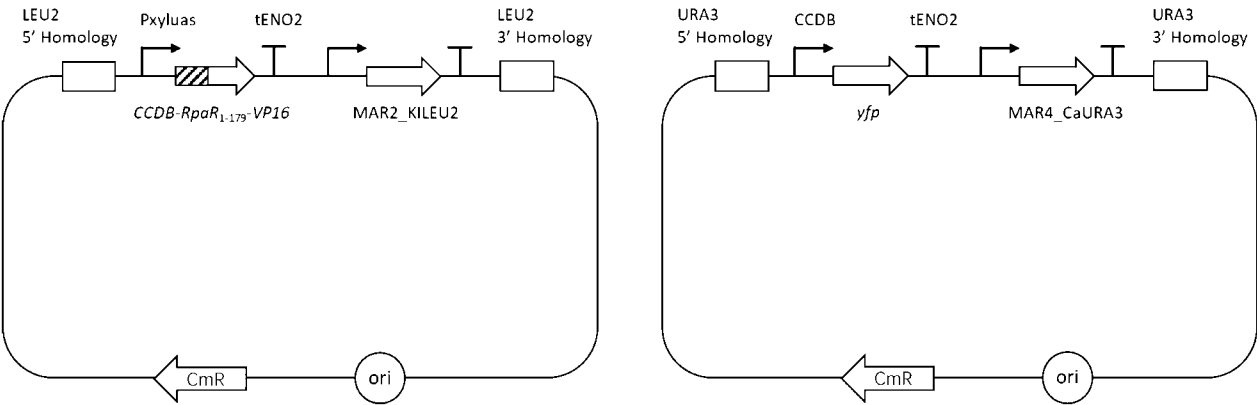


图 5

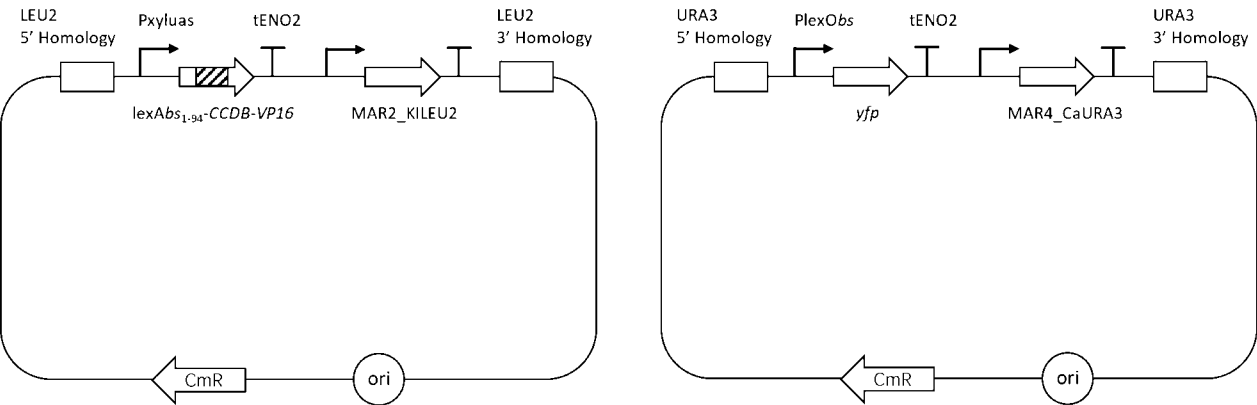


图 6

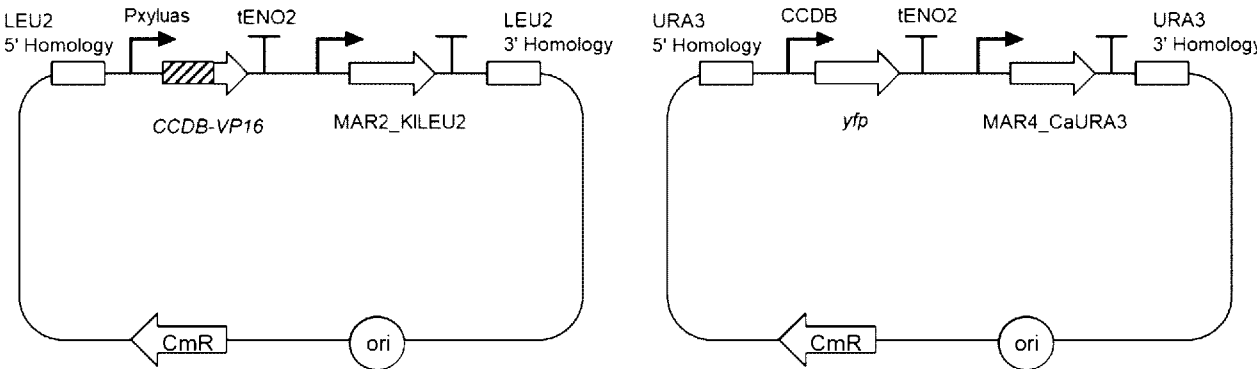


图 7

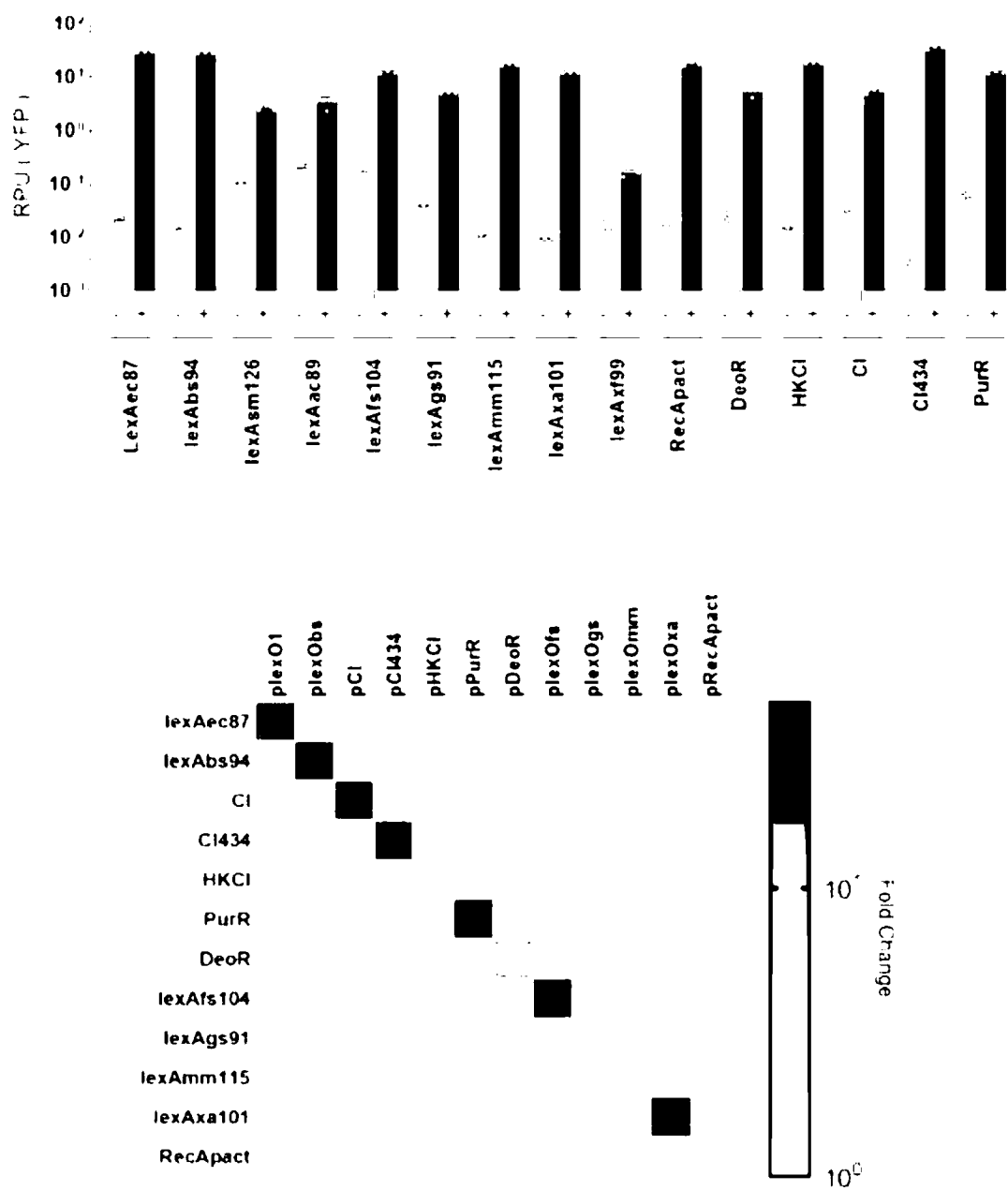


图 8

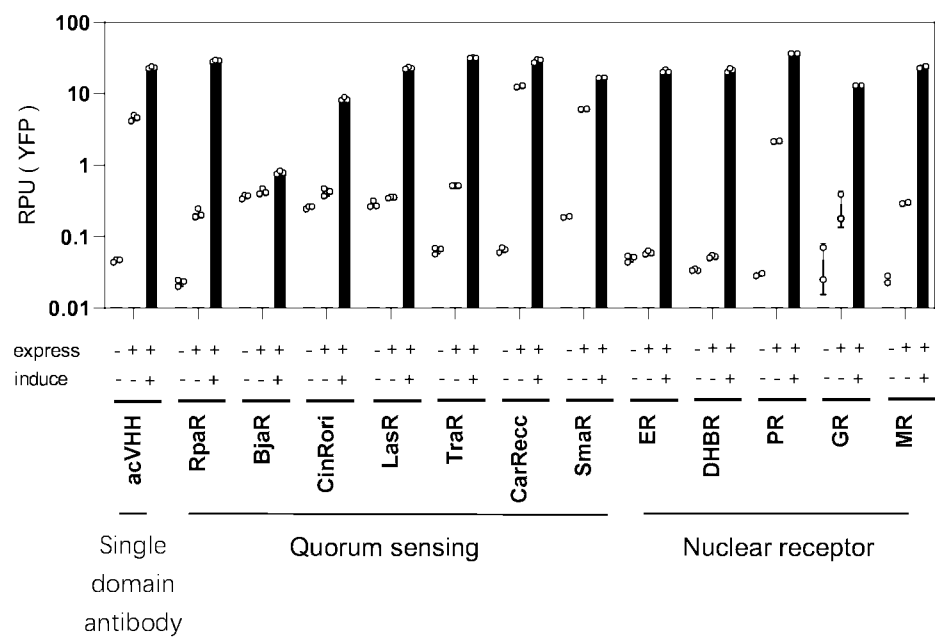


图 9

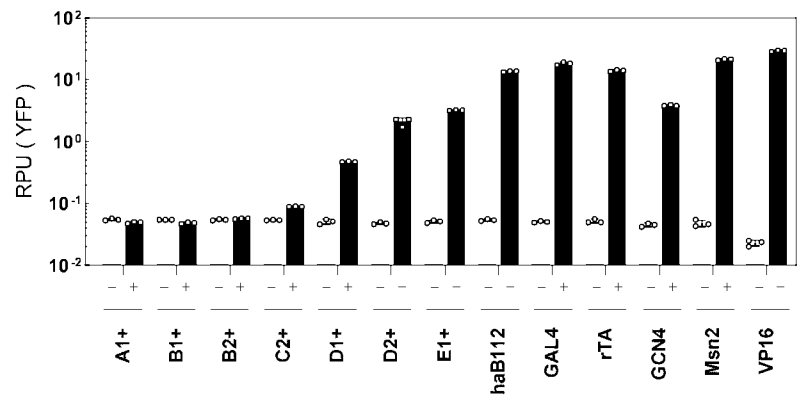


图 10

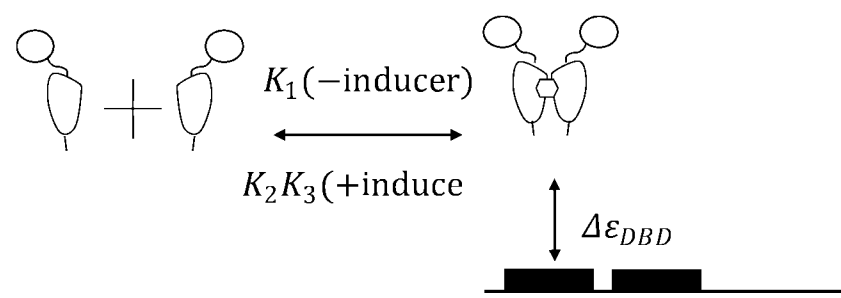


图 11

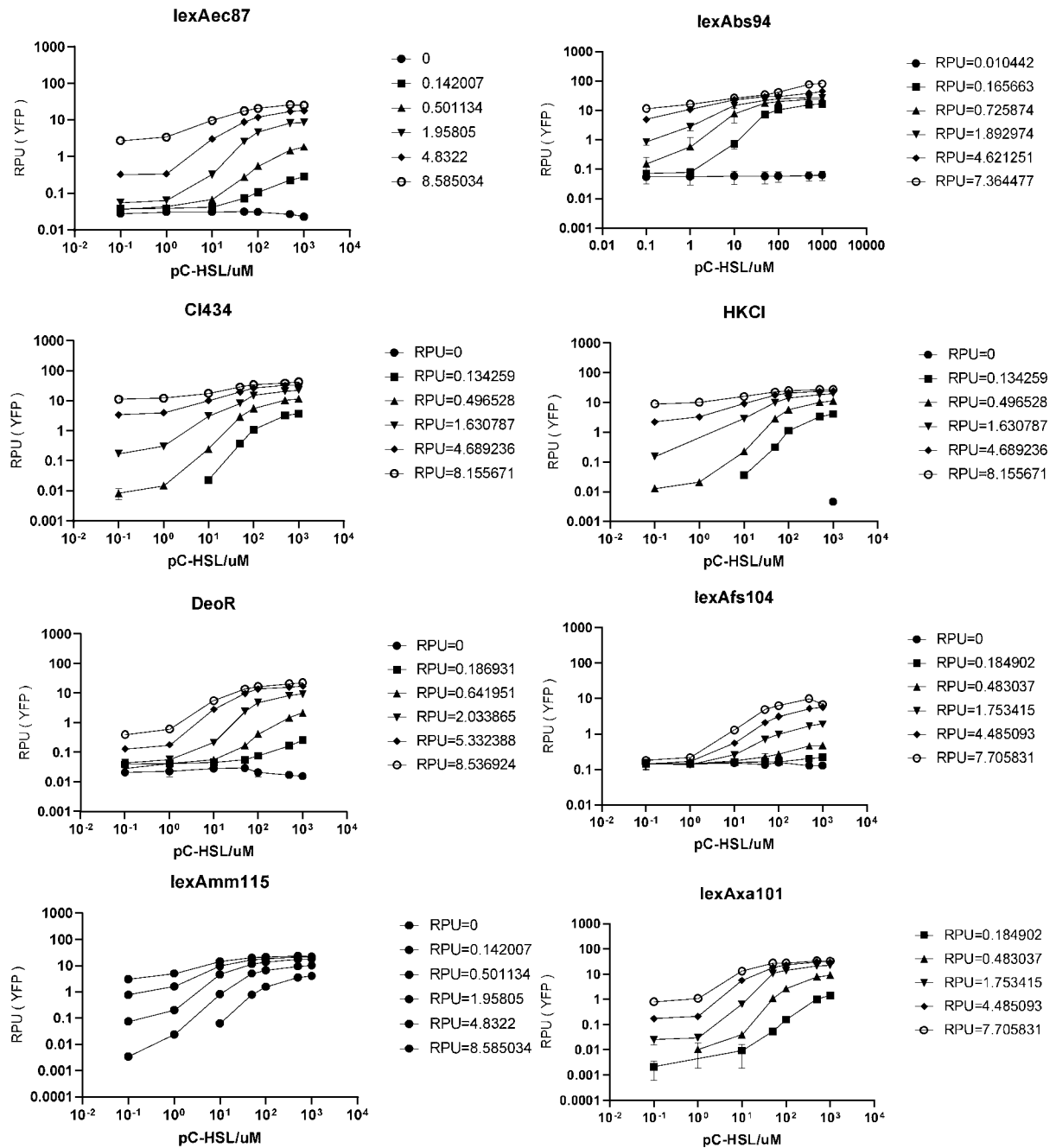


图 12

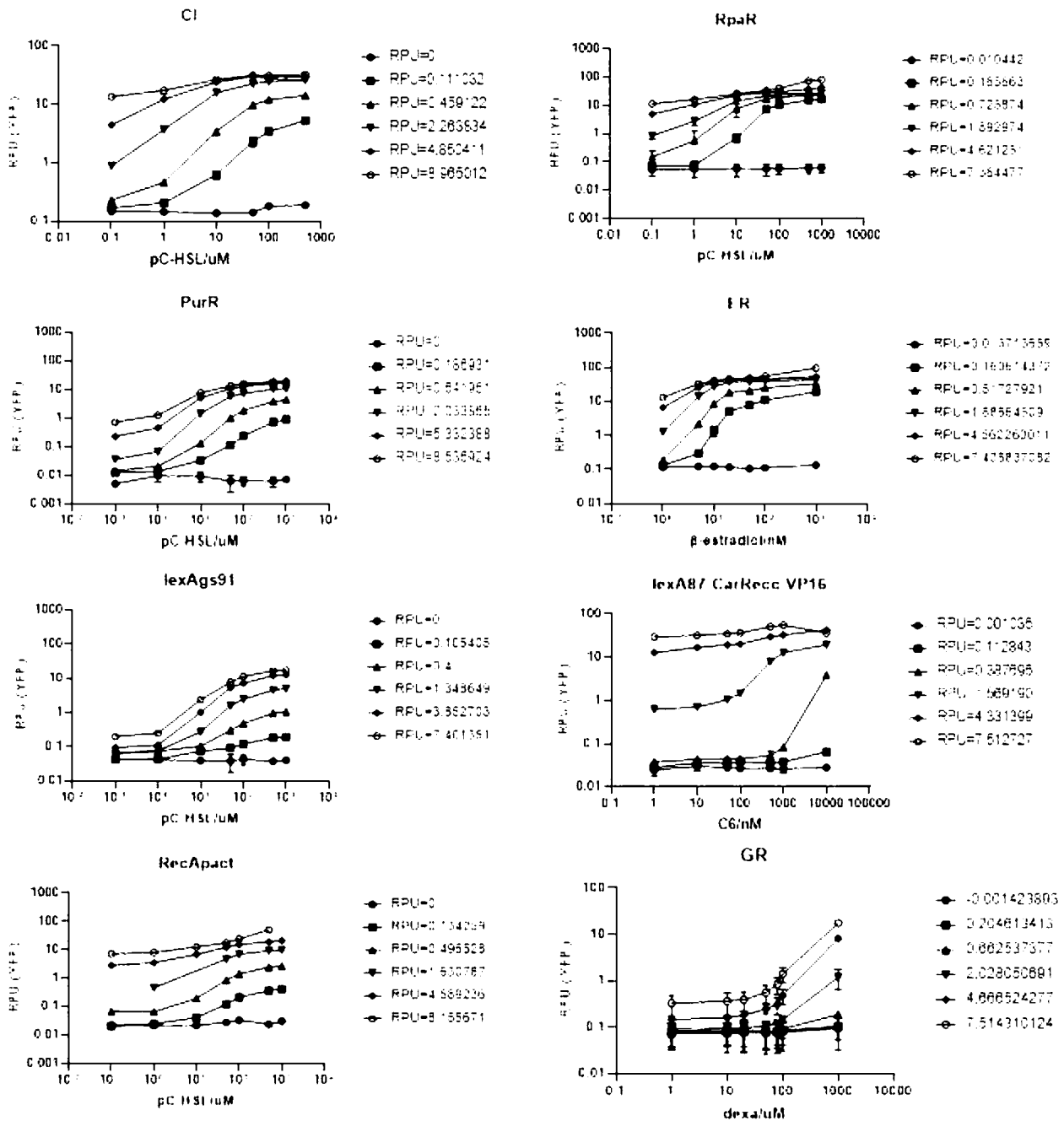


图 12 (续)

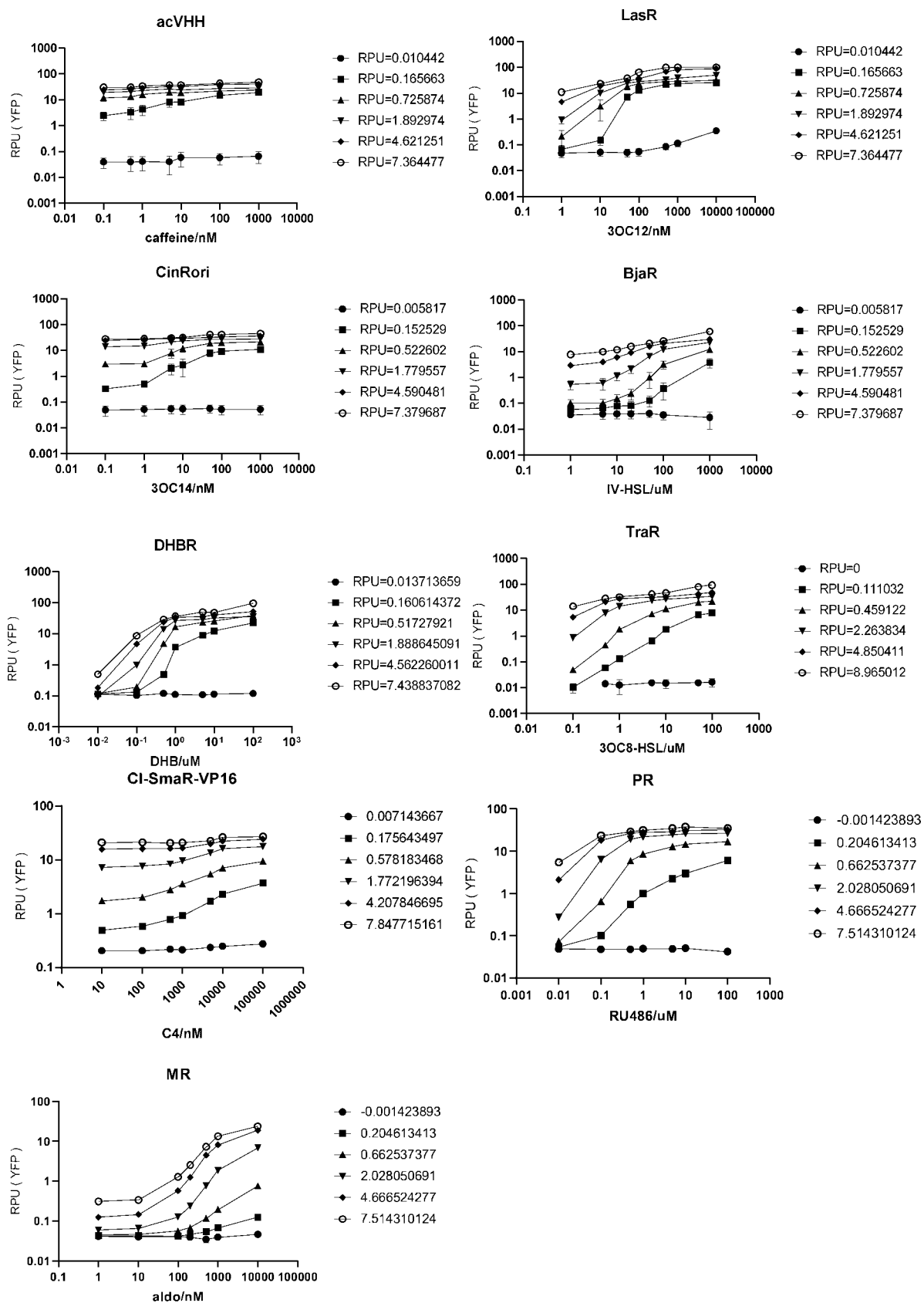
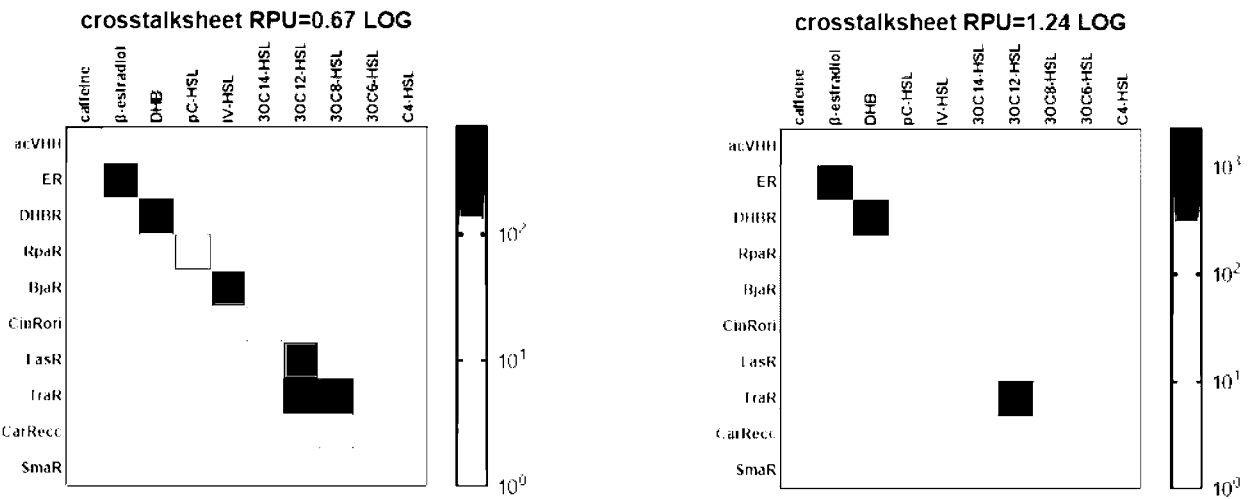


图 12 (续)



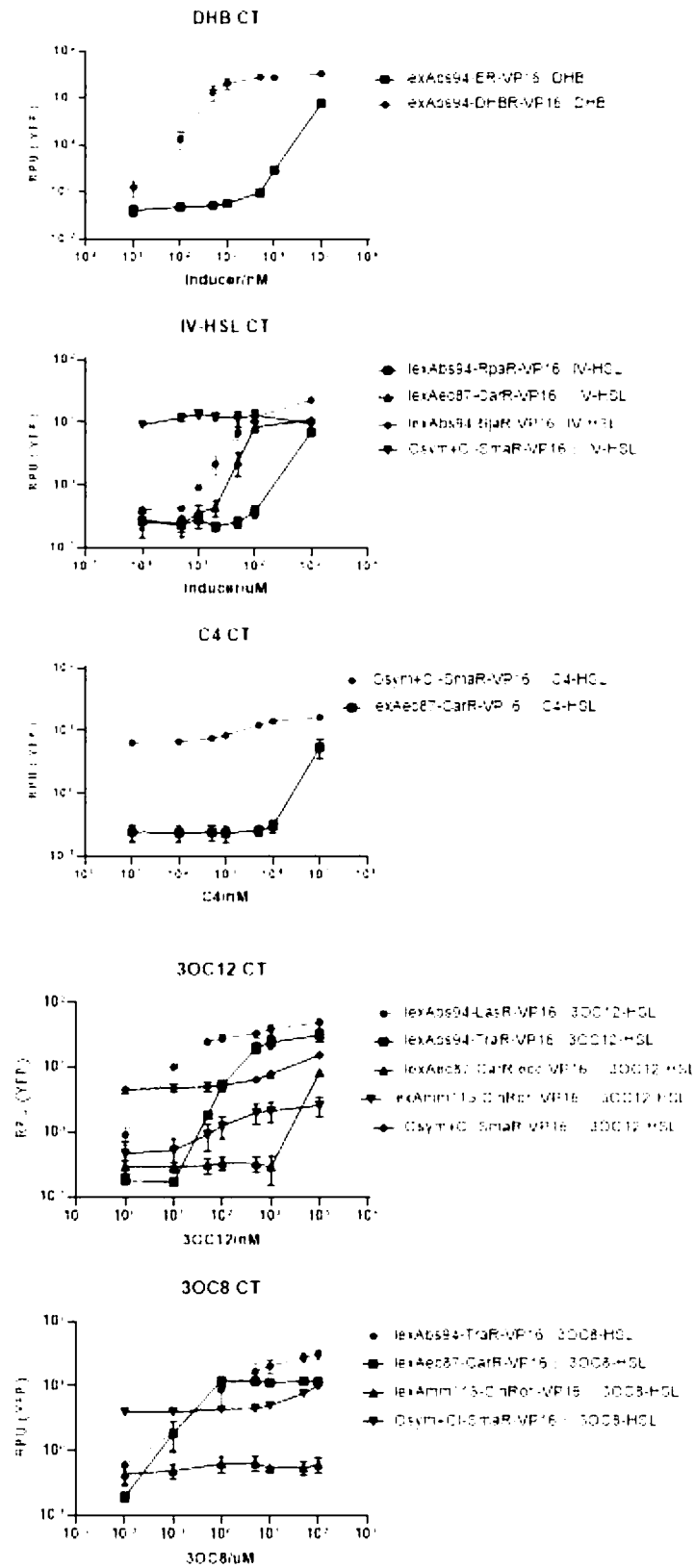


图 14

	CI434	LexAs5	LexAs17	LexA87	LexAs23	HKCI	RecApact	DeoR	PurR	LexAs15	LexAs14	CI OLI	CI OR1	CI Osym
kd	6.216348	40.97705	5.374494	0.745553	0.945353	5.066513	2.956591	0.407056	1.130833	0.313528	0.052414	30.08698	10.36936	10.53283
	acVHH	LasR	RpaR	CinRori	BjaR	CarR	SmaR	TraR	ER	DHBR	PR	GR	MR	
k1	0.010479	0.000171	0.000284	0.004922	0.000116	0.393133	0.008113	0.00011	4.59E-15	4.22E-22	1.98E-05	6.51E-06	8.49E-06	
k2	58.86022	0.145665	0.006111	71.69633	0.005969	1.93E-12	0.50421	0.179112	3.759987	0.039591	0.641768	1.017721	2.153996	
k3	0.746785	737.5151	0.899894	0.16094	0.015982	26.41891	0.029022	0.061774	4.282335	13.17104	0.105259	6.66E-05	0.002019	

图 15

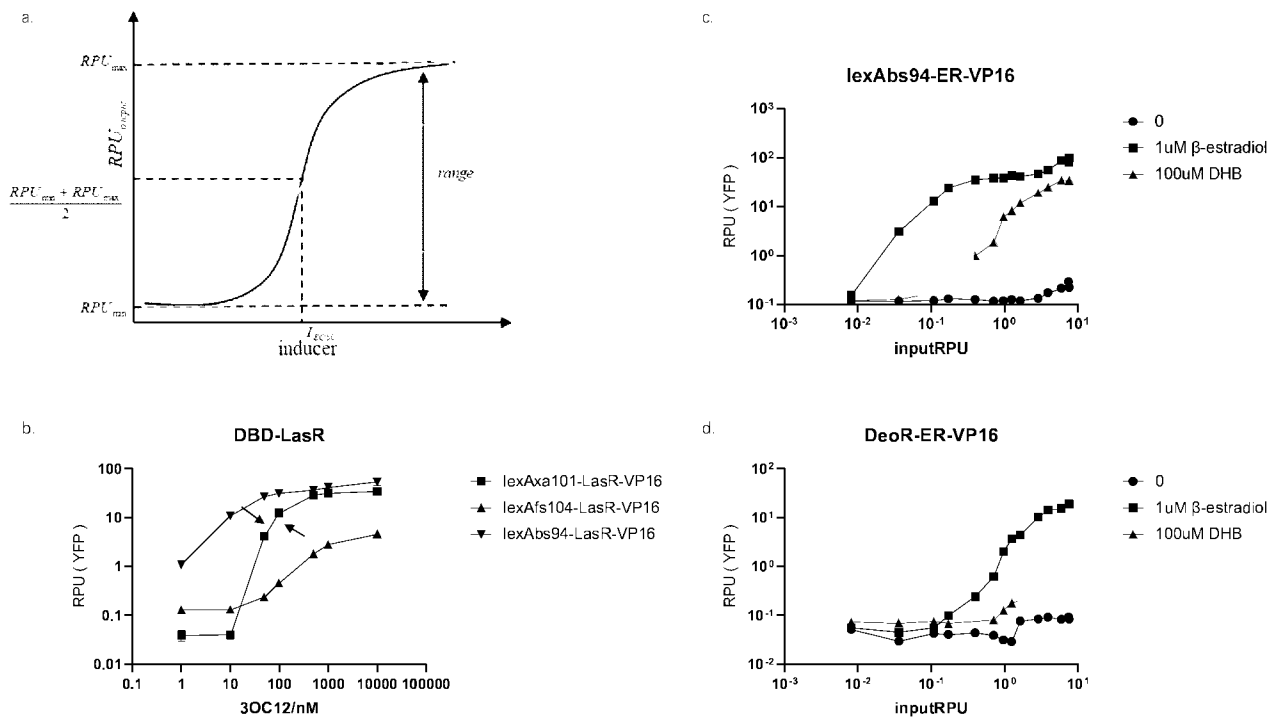


图 16

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/141107

A. CLASSIFICATION OF SUBJECT MATTER

G16B5/00(2019.01)i; C12N15/81(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)

IPC:G16B,C12N

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)

VEN, CNABS, CNTXT, WOTXT, USTXT, EPTXT, CNKI, IEEE: 转录因子, 蛋白, 诱导, 结构域, DNA, DBD, LBD, AD, 调控, 模型, 优化, 定量, 拟合, transcription factor, protein, induction, domain, regulation, model, optimization, quantification, fitting

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104419718 A (TIANJIN UNIVERSITY) 18 March 2015 (2015-03-18) description, paragraphs 0007-0098	1-11
A	CN 104630258 A (JIANGNAN UNIVERSITY) 20 May 2015 (2015-05-20) entire document	1-11
A	CN 108410902 A (QILU UNIVERSITY OF TECHNOLOGY et al.) 17 August 2018 (2018-08-17) entire document	1-11
A	CN 112102876 A (XI'AN JIAOTONG UNIVERSITY) 18 December 2020 (2020-12-18) entire document	1-11
A	CN 115873125 A (SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY, CHINESE ACADEMY OF SCIENCES) 31 March 2023 (2023-03-31) entire document	1-11
A	US 2021269811 A1 (BOEHRINGER INGELHEIM RCV GMBH & CO., KG. et al.) 02 September 2021 (2021-09-02) entire document	1-11



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

09 July 2024

Date of mailing of the international search report

12 July 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
Information on patent family members

International application No.

PCT/CN2023/141107

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	104419718	A	18 March 2015	None			
CN	104630258	A	20 May 2015	None			
CN	108410902	A	17 August 2018	US	2019225973	A1	25 July 2019
CN	112102876	A	18 December 2020	None			
CN	115873125	A	31 March 2023	None			
US	2021269811	A1	02 September 2021	WO	2020002494	A1	02 January 2020
				EP	3814491	A1	05 May 2021
				CA	3103988	A1	02 January 2020
				AU	2019294515	A1	14 January 2021
				SG	11202012529	A	28 January 2021
				JP	2021528985	A	28 October 2021
				KR	20210032972	A	25 March 2021
				CN	112955547	A	11 June 2021

A. 主题的分类

G16B5/00(2019.01)i; C12N15/81(2006.01)i

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:G16B,C12N

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))

VEN,CNABS,CNTEXT,WOTXT,USTXT,EPTXT,CNKI,IEEE:转录因子, 蛋白, 诱导, 结构域, DNA, DBD,LBD,AD, 调控, 模型, 优化, 定量, 拟合, transcription factor, protein, induction, domain, regulation, model, optimization, quantifi-cation, fitting

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 104419718 A (天津大学) 2015年3月18日 (2015 - 03 - 18) 说明书第0007-0098段	1-11
A	CN 104630258 A (江南大学) 2015年5月20日 (2015 - 05 - 20) 全文	1-11
A	CN 108410902 A (齐鲁工业大学 等) 2018年8月17日 (2018 - 08 - 17) 全文	1-11
A	CN 112102876 A (西安交通大学) 2020年12月18日 (2020 - 12 - 18) 全文	1-11
A	CN 115873125 A (中国科学院深圳先进技术研究院) 2023年3月31日 (2023 - 03 - 31) 全文	1-11
A	US 2021269811 A1 (BOEHRINGER INGELHEIM RCV GMBH & CO KG 等) 2021年9月2日 (2021 - 09 - 02) 全文	1-11

☐ 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7月9日

国际检索报告邮寄日期

2024年7月12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

张明霞

电话号码 (+86) 010-53961517

PCT/ISA/210 表(第2页) (2022年7月)

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2023/141107

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	104419718	A	2015年3月18日	无			
CN	104630258	A	2015年5月20日	无			
CN	108410902	A	2018年8月17日	US	2019225973	A1	2019年7月25日
CN	112102876	A	2020年12月18日	无			
CN	115873125	A	2023年3月31日	无			
US	2021269811	A1	2021年9月2日	WO	2020002494	A1	2020年1月2日
				EP	3814491	A1	2021年5月5日
				CA	3103988	A1	2020年1月2日
				AU	2019294515	A1	2021年1月14日
				SG	11202012529	A	2021年1月28日
				JP	2021528985	A	2021年10月28日
				KR	20210032972	A	2021年3月25日
				CN	112955547	A	2021年6月11日