



(12)发明专利申请

(10)申请公布号 CN 110551700 A

(43)申请公布日 2019.12.10

(21)申请号 201810554584.5

(22)申请日 2018.05.31

(71)申请人 康码(上海)生物科技有限公司

地址 201321 上海市浦东新区芙蓉花路500
弄8号楼3楼(72)发明人 郭敏 王海鹏 姜灵轩 娄旭
许乃庆 代田纯 李海洋 邓蜜妮
于雪

(51)Int.Cl.

C12N 9/04(2006.01)

C12N 15/53(2006.01)

C12P 21/02(2006.01)

权利要求书1页 说明书29页
序列表51页 附图14页

(54)发明名称

ADH蛋白家族突变体及其应用

(57)摘要

本发明涉及ADH蛋白家族突变体及其应用。具体地,本发明提供一种ADH蛋白的突变体。相比野生型ADH蛋白,本发明的ADH蛋白突变体可显著(i)提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和得率;和/或(ii)降低突变蛋白与Ni介质的结合能力。

K1ADH4	1	NFRLARAQTALANKASVSRGFLRLNSSPT	ETQKGVIFYENG	Q	YKD	VVPKPK	NE
K1ADH3	1	MLRLTSARSIVSPLRKGAFG.SIRTLAT	ETQKGVIFYENG	Q	YKD	VVPKPK	NE
K1ADH1	1	MAAS	ETQKGVIFYENG	Q	YKD	VVPKPK	NE
K1ADH2	1	MAAS	ETQKGVIFYENG	Q	YKD	VVPKPK	NE
K1ADH4	61	LINVKYSQVCHFDIANKGDMPLF	KLPLVGGREGAG	VVA	GENVR	IGG	AGIKW
K1ADH3	60	LINVKYSQVCHFDIANKGDMPLF	KLPLVGGREGAG	VVA	GENVR	IGG	AGIKW
K1ADH1	36	LINVKYSQVCHFDIANKGDMPLF	KLPLVGGREGAG	VVA	GENVR	IGG	AGIKW
K1ADH2	34	LINVKYSQVCHFDIANKGDMPLF	KLPLVGGREGAG	VVA	GENVR	IGG	AGIKW
K1ADH4	121	LNGSCMSCE	CEESNC	ADLSGYIH	GSFQQYATADAVQAA	IF	QTDLAS
K1ADH3	120	LNGSCMSCE	CEESNC	ADLSGYIH	GSFQQYATADAVQAA	IF	QTDLAS
K1ADH1	96	LNGSCMSCE	CEESNC	ADLSGYIH	GSFQQYATADAVQAA	IF	QTDLAS
K1ADH2	94	LNGSCMSCE	CEESNC	ADLSGYIH	GSFQQYATADAVQAA	IF	QTDLAS
K1ADH4	181	CAGVTYKALR	A	L	K	G	VVAISGA
K1ADH3	180	CAGVTYKALR	A	L	K	G	VVAISGA
K1ADH1	156	CAGVTYKALR	A	L	K	G	VVAISGA
K1ADH2	154	CAGVTYKALR	A	L	K	G	VVAISGA
K1ADH4	241	GGE	F	D	T	R	K
K1ADH3	240	GGE	F	D	T	R	K
K1ADH1	216	GGE	F	D	T	R	K
K1ADH2	214	GGE	F	D	T	R	K
K1ADH4	301	A	V	K	S	V	V
K1ADH3	300	A	V	K	S	V	V
K1ADH1	276	A	V	K	S	V	V
K1ADH2	274	A	V	K	S	V	V
K1ADH4	361	T	C	E	E	G	R
K1ADH3	360	T	C	E	E	G	R
K1ADH1	336	T	C	E	E	G	R
K1ADH2	334	T	C	E	E	G	R

1. 一种乙醇脱氢酶 (ADH) 突变蛋白, 其特征在于, 所述的突变蛋白为非天然蛋白, 并且所述突变蛋白在野生型的乙醇脱氢酶 (ADH) 的第45-149位氨基酸中的至少1个组氨酸 (H) 发生突变。

2. 如权利要求1所述的突变蛋白, 其特征在于, 所述突变不包括野生型的乙醇脱氢酶 (ADH) 第67位、69位、93位和/或94位组氨酸 (H)。

3. 如权利要求1所述的突变蛋白, 其特征在于, 所述组氨酸可各自独立地突变为碱性氨基酸。

4. 如权利要求1所述的突变蛋白, 其特征在于, 所述组氨酸可各自独立地突变为选自下组的一种或多种氨基酸: 赖氨酸 (K)、天冬酰胺 (N)、谷氨酰胺 (Q)、精氨酸 (R)、或其组合。

5. 一种多核苷酸, 其特征在于, 所述的多核苷酸编码权利要求1所述的突变蛋白。

6. 一种载体, 其特征在于, 所述的载体含有权利要求5所述的多核苷酸。

7. 一种宿主细胞, 其特征在于, 所述宿主细胞含有权利要求6所述的载体, 或其基因组中整合有权利要求5所述的多核苷酸。

8. 一种产生权利要求1所述的乙醇脱氢酶 (ADH) 突变蛋白的方法, 其特征在于, 包括步骤:

在适合表达的条件下, 培养权利要求7所述的宿主细胞, 从而表达乙醇脱氢酶 (ADH) 突变蛋白; 和/或

分离所述乙醇脱氢酶 (ADH) 突变蛋白。

9. 一种酶制剂, 其特征在于, 所述酶制剂包含权利要求1所述的乙醇脱氢酶 (ADH) 突变蛋白。

10. 一种权利要求1所述的突变蛋白的用途, 其特征在于, 所述的突变蛋白用于制备 (i) 提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率; 和/或 (ii) 降低突变蛋白与Ni介质的结合能力的酶制剂。

ADH蛋白家族突变体及其应用

技术领域

[0001] 本发明涉及生物技术领域,具体地涉及ADH蛋白家族突变体及其应用。

背景技术

[0002] 蛋白质分离纯化,是指通过在目的蛋白的一端融合多聚标签或目的蛋白的特定性状,将混合物、抽提物、细胞破碎液或反应产物等,经由层析(chromatography)方法达到分离其他物质、纯化目的蛋白的过程。亲和层析是指将具有亲和力的两个分子中的一个固定在不溶性基质上,利用分子间亲和力的特异性和可逆性,对另一个分子进行分离纯化的方法。常见的用来亲和层析的多聚标签主要有组氨酸(Histidine, His),谷胱甘肽S转移酶(Glutathione S-transferase, GST)和麦芽糖结合蛋白(Maltose Binding Protein, MBP)等。固定化金属离子亲和层析(immobilized metal-chelating affinity chromatography, IMAC)的原理主要是基于蛋白表面的氨基酸残基与金属离子能形成不同的亲和力,主要分为三种:静电吸引、共价结合与配位键结合。当蛋白表面含有组氨酸、半胱氨酸或色氨酸,对应的咪唑基、巯基或吡咯基便可与金属离子形成配位键。该方法凭借表达方便、成本低廉以及对目的蛋白性质影响小等优点,已成为常用的蛋白纯化手段。

[0003] 然而,在实际应用过程中,该纯化方式也存在一定的问题。以Ni-NTA树脂为例,有些非目的蛋白在其三维结构表面也存在若干不连续的组氨酸残基,导致该蛋白对Ni-NTA树脂也有一定程度的结合,从而使最终目的蛋白的纯度受到干扰。

[0004] 因此,本领域迫切需要发明一种改造非目的蛋白组氨酸残基的方法,使其不能与Ni离子亲和介质,如Ni-NTA树脂相结合,达到提升目的蛋白纯度的效果。

发明内容

[0005] 本发明的目的是提供一种改造非目的蛋白组氨酸残基的方法,使其不能与Ni离子亲和介质,如Ni-NTA树脂相结合,达到提升目的蛋白纯度的效果。

[0006] 本发明第一方面提供了一种乙醇脱氢酶(ADH)突变蛋白,所述的突变蛋白为非天然蛋白,并且所述突变蛋白在野生型的乙醇脱氢酶(ADH)的第45-149位氨基酸中的至少1个组氨酸(H)发生突变。

[0007] 在另一优选例中,所述突变不包括野生型的乙醇脱氢酶(ADH)第67位、69位、93位和/或94位组氨酸(H)。

[0008] 在另一优选例中,与野生型的乙醇脱氢酶(ADH)相比,所述乙醇脱氢酶(ADH)突变蛋白与Ni介质的结合能力降低了10%,较佳地,降低了25%,更佳地,降低了50%,更佳地,降低了80%,最佳地,降低了100%。

[0009] 在另一优选例中,所述组氨酸可各自独立地突变为碱性氨基酸。

[0010] 在另一优选例中,所述组氨酸各自独立突变的类型可以相同,可以不同。

[0011] 在另一优选例中,所述组氨酸可各自独立地突变为选自下组的一种或多种氨基酸:赖氨酸(K)、天冬酰胺(N)、谷氨酰胺(Q)、精氨酸(R)、或其组合。

[0012] 在另一优选例中,所述突变蛋白来源于*K. lactis*。

[0013] 在另一优选例中,所述乙醇脱氢酶 (ADH) 包括ADH1、ADH2、ADH3和/或ADH4蛋白。

[0014] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶 (ADH1) 的对应于SEQ ID NO.: 1的选自下组的一个或多个核心氨基酸发生突变:

第47位组氨酸 (H) ;

第51位组氨酸 (H) ;

第124位组氨酸 (H) 。

[0015] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶 (ADH2) 的对应于SEQ ID NO.:26的选自下组的一个或多个核心氨基酸发生突变:

第45位组氨酸 (H) ;

第49位组氨酸 (H) ;

第122位组氨酸 (H) 。

[0016] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶 (ADH3) 的对应于SEQ ID NO.:27的选自下组的一个或多个核心氨基酸发生突变:

第71位组氨酸 (H) ;

第75位组氨酸 (H) ;

第148位组氨酸 (H) 。

[0017] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶 (ADH4) 的对应于SEQ ID NO.:28的选自下组的一个或多个核心氨基酸发生突变:

第72位组氨酸 (H) ;

第76位组氨酸 (H) ;

第149位组氨酸 (H) 。

[0018] 在另一优选例中,所述第47位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、天冬酰胺 (N)、谷氨酰胺 (Q)、精氨酸 (R)、或其组合,较佳地,赖氨酸 (K)、天冬酰胺 (N) 和/或精氨酸 (R),更佳地,赖氨酸 (K)、和/或天冬酰胺 (N) 。

[0019] 在另一优选例中,所述第51位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、精氨酸 (R)、或其组合,较佳地,赖氨酸 (K)、和/或精氨酸 (R),更佳地,精氨酸 (R) 。

[0020] 在另一优选例中,第124位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、精氨酸 (R)、或其组合,较佳地,赖氨酸 (K) 。

[0021] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶 (ADH) 的对应于SEQ ID NO.: 1的选自下组的核心氨基酸发生突变:

第47位组氨酸 (H) ;和

第124位组氨酸 (H) 。

[0022] 在另一优选例中,第45位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、精氨酸 (R)、天冬酰胺 (N)、谷氨酰胺 (Q)、或其组合。

[0023] 在另一优选例中,第49位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、精氨酸 (R)、或其组合。

[0024] 在另一优选例中,第122位组氨酸 (H) 突变为选自下组的一种或多种氨基酸:赖氨酸 (K)、精氨酸 (R)、或其组合。

[0025] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.:26的选自下组的核心氨基酸发生突变:

第122位组氨酸(H);和
第45位组氨酸(H)。

[0026] 在另一优选例中,第71位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、天冬酰胺(N)、谷氨酰胺(Q)、或其组合。

[0027] 在另一优选例中,第75位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0028] 在另一优选例中,第148位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0029] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.:27的选自下组的核心氨基酸发生突变:

第148位组氨酸(H);和
第71位组氨酸(H)。

[0030] 在另一优选例中,第72位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、天冬酰胺(N)、谷氨酰胺(Q)、或其组合。

[0031] 在另一优选例中,第76位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0032] 在另一优选例中,第149位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0033] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.:28的选自下组的核心氨基酸发生突变:

第149位组氨酸(H);和
第72位组氨酸(H)。

[0034] 在另一优选例中,所述的突变选自下组:H47K;H47N;H47Q;H47R;H51K;H51R;H124K;H124R;H47K和H124K;H47N和H124K;H47Q和H124K;H47R和H124K。

[0035] 在另一优选例中,所述突变选自下组:H47K;H47N;H51R;H124R;H47K和H124K;H47N和H124K。

[0036] 在另一优选例中,所述突变选自下组:H122K;H122R;H45K;H45R;H45N;H45Q;H49K;H49R;H122K和H45K;H122K和H45R;H122K和H45N;H122K和H45Q。

[0037] 在另一优选例中,所述突变选自下组:H148K;H148R;H71K;H71R;H71N;H71Q;H75K;H75R;H148K和H71K;H148K和H71R;H148K和H71N;H148K和H71Q。

[0038] 在另一优选例中,所述突变选自下组:H149K;H149R;H72K;H72R;H72N;H72Q;H76K;H76R;H149K和H72K;H149K和H72R;H149K和H72N;H149K和H72Q。

[0039] 在另一优选例中,所述乙醇脱氢酶(ADH)突变蛋白的氨基酸序列选自下组:

(1) 氨基酸序列如SEQ ID NO.:2-13任一所示的多肽。

[0040] (2) 将SEQ ID NO.: 2-13任一所示氨基酸序列经过一个或几个,优选1-20个、更优选1-15个、更优选1-10个、更优选1-8个、更优选1-3个、最优选1个氨基酸残基的取代、缺失或添加而形成的,具有(1)所述多肽功能的由SEQ ID NO.:2-13所示氨基酸序列的多肽衍生

的多肽。

[0041] 在另一优选例中,所述突变蛋白的氨基酸序列如SEQ ID NO.:2-13所示。

[0042] 在另一优选例中,所述突变蛋白的氨基酸序列与SEQ ID NO.: 2-13所示的序列具有至少70%,优选至少75%、80%、85%、90%,更优选至少95%、96%、97%、98%、99%以上的序列相同性。

[0043] 在另一优选例中,所述的突变蛋白除所述突变(如47、51和/或124位)外,其余的氨基酸序列与SEQ ID NO.: 1所示的序列相同或基本相同。

[0044] 在另一优选例中,所述的突变蛋白除所述突变(如45、49和/或122位)外,其余的氨基酸序列与SEQ ID NO.: 26所示的序列相同或基本相同。

[0045] 在另一优选例中,所述的突变蛋白除所述突变(如71、75和/或148位)外,其余的氨基酸序列与SEQ ID NO.: 27所示的序列相同或基本相同。

[0046] 在另一优选例中,所述的突变蛋白除所述突变(如72、76和/或149位)外,其余的氨基酸序列与SEQ ID NO.: 28所示的序列相同或基本相同。

[0047] 在另一优选例中,所述的基本相同是至多有50个(较佳地为1-20个,更佳地为1-10个、更佳地1-5个)氨基酸不相同,其中,所述的不相同包括氨基酸的取代、缺失或添加,且所述的突变蛋白仍具有(i)提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或(ii)降低突变蛋白与Ni介质的结合能力。

[0048] 在另一优选例中,与SEQ ID NO.:1、26、27或28所示序列的同源性至少为80%,较佳地至少为85%或90%,更佳地至少为95%,最佳地至少为98%或99%。

[0049] 本发明第二方面提供了一种多核苷酸,所述的多核苷酸编码本发明第一方面所述的突变蛋白。

[0050] 在另一优选例中,所述多核苷酸选自下组:

- (a) 编码如SEQ ID NO.:2-13任一所示多肽的多核苷酸;
- (b) 序列如SEQ ID NO.:14-25任一所示的多核苷酸;
- (c) 核苷酸序列与野生型ADH1蛋白的核苷酸序列所示序列的同源性 $\geq 95\%$ (较佳地 $\geq 98\%$),且编码如SEQ ID NO.:2-13任一所示多肽的多核苷酸;
- (d) 与(a)-(c)任一所述的多核苷酸互补的多核苷酸。

[0051] 在另一优选例中,所述的多核苷酸在所述突变蛋白的ORF的侧翼还额外含有选自下组的辅助元件:信号肽、分泌肽、标签序列(如6His)、或其组合。

[0052] 在另一优选例中,所述的多核苷酸选自下组:DNA序列、RNA序列、或其组合。

[0053] 本发明第三方面提供了一种载体,所述的载体含有本发明第二方面所述的多核苷酸。

[0054] 在另一优选例中,所述载体包含一个或多个启动子,所述启动子可操作地与所述核酸序列、增强子、转录终止信号、多腺苷酸化序列、复制起点、选择性标记、核酸限制性位点、和/或同源重组位点连接。

[0055] 在另一优选例中,所述载体包括质粒、病毒载体。

[0056] 在另一优选例中,所述的病毒载体选自下组:腺相关病毒(AAV)、腺病毒、慢病毒、逆转录病毒、疱疹病毒、SV40、痘病毒、或其组合。

[0057] 在另一优选例中,所述载体包括表达载体、穿梭载体、整合载体。

[0058] 本发明第四方面提供了一种宿主细胞,所述宿主细胞含有本发明第三方面所述的载体,或其基因组中整合有本发明第二方面所述的多核苷酸。

[0059] 在另一优选例中,所述的宿主细胞为真核细胞,如酵母细胞、植物细胞或哺乳动物细胞(包括人和非人哺乳动物)。

[0060] 在另一优选例中,所述的宿主细胞为原核细胞,如大肠杆菌。

[0061] 在另一优选例中,所述酵母细胞选自下组的一种或多种来源的酵母:毕氏酵母、克鲁维酵母、或其组合;较佳地,所述的酵母细胞包括:克鲁维酵母,更佳地为马克斯克鲁维酵母、和/或乳酸克鲁维酵母。

[0062] 在另一优选例中,所述宿主细胞选自下组:大肠杆菌、麦胚细胞,昆虫细胞, SF9、Hela、HEK293、CHO、酵母细胞、或其组合。

[0063] 本发明第五方面提供了一种产生本发明第一方面所述的乙醇脱氢酶 (ADH) 突变蛋白的方法,包括步骤:

在适合表达的条件下,培养本发明第四方面所述的宿主细胞,从而表达乙醇脱氢酶 (ADH) 突变蛋白;和/或

分离所述乙醇脱氢酶 (ADH) 突变蛋白。

[0064] 本发明第六方面提供了一种酶制剂,所述酶制剂包含本发明第一方面所述的乙醇脱氢酶 (ADH) 突变蛋白。

[0065] 在另一优选例中,所述的酶制剂包括注射剂、和/或冻干制剂。

[0066] 本发明第七方面提供了一种本发明第一方面所述的突变蛋白的用途,所述的突变蛋白用于制备 (i) 提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或 (ii) 降低突变蛋白与Ni介质的结合能力的酶制剂。

[0067] 在另一优选例中,所述的Ni选自下组:镍元素、镍离子、游离镍离子、结合有镍的层析介质、 Ni^+ 、Ni-beads、Ni-NTA、或其组合。

[0068]

应理解,在本发明范围内中,本发明的上述各技术特征和在下文(如实施例)中具体描述的各技术特征之间都可以互相组合,从而构成新的或优选的技术方案。限于篇幅,在此不再一一累述。

[0069]

附图说明

[0070] 图1显示了通过改造*K. lactis*基因组中ADH1以消除其与Ni介质结合的设计路线图。

[0071] 图2是*Kl*ADH家族蛋白ADH1-4的序列同源性的对比。其中黑色区域代表序列高度同源。

[0072] 图3是*Kl*ADH1与*Sc*ADH1的序列同源性的对比,二者序列同源性为84.9%。

[0073] 图4是以2HCY为模板对*Kl*ADH4进行的同源建模,二者均为四聚体,每个单体的6个His聚集区分别用方框表示。

[0074] 图5是*Sc*ADH1四聚体结构图,A和B两个亚基通过NADH/NAD⁺结合域构成二聚体,两个AB二聚体组成“背靠背”四聚体。

[0075] 图6是ScADH1中H66周围氨基酸结构图,H66在两种ADH1的构象中均参与螯合具有催化作用的 Zn^{2+} 。A图表示结合了NAD的封闭构象,B图表示未结合NAD的开放构象。

[0076] 图7是ScADH1中H48周围氨基酸结构图,H48参与三元复合体的形成,影响底物的质子转移过程。

[0077] 图8是ScADH1中H44周围氨基酸结构图,H44咪唑环上的ND1可能参与 NAD^+/NADH 的结合。

[0078] 图9是pKMCas9_StR_K1ADH1质粒图谱。该质粒带有tRNA-Tyr启动子和SNR52终止子,并带有kana筛选标记。

[0079] 图10是pKMD1- Δ K1ADH1质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0080] 图11是pKMD1-K1ADH1-H47K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0081] 图12是pKMD1-K1ADH1-H47N质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0082] 图13是pKMD1-K1ADH1-H47Q质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0083] 图14是pKMD1-K1ADH1-H47R质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0084] 图15是pKMD1-K1ADH1-H51K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0085] 图16是pKMD1-K1ADH1-H51R质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0086] 图17是pKMD1-K1ADH1-H124K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0087] 图18是pKMD1-K1ADH1-H124R质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0088] 图19是pKMD1-K1ADH1-H47KH124K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0089] 图20是pKMD1-K1ADH1-H47NH124K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0090] 图21是pKMD1-K1ADH1-H47QH124K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0091] 图22是pKMD1-K1ADH1-H47RH124K质粒图谱。homologous arm 1和homologous arm 2分别为K1ADH1基因的ORF上、下游各800 bp左右的基因序列,质粒带有Amp筛选标记。

[0092] 图23是经由Ni-NTA纯化的聚丙烯酰胺凝胶分析。如图所示,K1ADH1的基因敲除和不同位置的His突变可显著降低K1ADH1与Ni-NTA的亲合性。目的条带大小约为37KD,从左至右分别是marker,K. lactis野生型细胞株WT,K1ADH1敲除细胞株 Δ ADH1,K1ADH1单突变细胞株H47K,H47N,H47R,H51K,H124K,H124KH47K和H124KH47N。

[0093] 图24是野生型细胞株WT、K1ADH1基因敲除细胞株 Δ ADH1和K1ADH1单突变细胞株

H47K的无细胞转录翻译活性的分析。如图所示,从左至右分别是 Δ ADH1, *Kl*ADH1 H47K, *K. lactis*野生型细胞株WT和阴性对照。

[0094] 图25是野生型细胞株WT和*Kl*ADH1单突变细胞株H47K (Mut) 的无细胞转录翻译制备出的目标蛋白质的纯化分析。如图所示,从左至右分别是*Kl*ADH1单突变细胞株H47K (Mut) 和 *K. lactis*野生型细胞株WT。A图是没有加入外源目标蛋白时*Kl*ADH1与Ni-NTA的亲和性分析。B图是在*Kl*ADH1单突变细胞株H47K的无细胞转录翻译体系中加入目标蛋白N-His8-eGFP时eGFP的纯化分析。C图是在*Kl*ADH1单突变细胞株H47K的无细胞转录翻译体系中加入目标蛋白N-His8-Strep-eGFP时Strep-eGFP的纯化分析。

[0095] 图26总结了通过*Kl*ADH1基因敲除和定点突变来消除其与Ni介质结合的能力,提高目的蛋白的纯度和特异性。

[0096]

具体实施方式

[0097] 本发明人经过广泛而深入的研究,意外地获得一种ADH家族蛋白突变体。相比野生型ADH家族蛋白,本发明的ADH家族蛋白的突变体可显著 (i) 提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或 (ii) 降低突变蛋白与Ni介质的结合能力。在此基础上,本发明人完成了本发明。

[0098]

术语

为了可以更容易地理解本公开,首先定义某些术语。如本申请中所使用的,除非本文另有明确规定,否则以下术语中的每一个应具有下面给出的含义。在整个申请中阐述了其它定义。

[0099] 术语“约”可以是指在本领域普通技术人员确定的特定值或组成的可接受误差范围内的值或组成,其将部分地取决于如何测量或测定值或组成。例如,如本文所用,表述“约100”包括99和101和之间的全部值(例如,99.1、99.2、99.3、99.4等)。

[0100] 如本文所用,术语“含有”或“包括(包含)”可以是开放式、半封闭式和封闭式的。换言之,所述术语也包括“基本上由...构成”、或“由...构成”。

[0101] 序列同一性(或同源性)通过沿着预定的比较窗(其可以是参考核苷酸序列或蛋白的长度的50%、60%、70%、80%、90%、95%或100%)比较两个对齐的序列,并且确定出现相同的残基的位置的数目来确定。通常地,这表示为百分比。核苷酸序列的序列同一性的测量是本领域技术人员熟知的方法。

[0102] 如本文使用的,术语“受试者”、“需要的对象”指任何哺乳动物或非哺乳动物。哺乳动物包括但不限于人类、脊椎动物诸如啮齿类、非人类灵长类、牛、马、狗、猫、猪、绵羊、山羊、长颈鹿、鹿、骆驼、羚羊、野兔和家兔。

[0103]

ADH家族蛋白

乙醇脱氢酶ADH家族蛋白,大量存在于人和动物肝脏、植物及微生物细胞之中,作为生物体内主要短链醇代谢的关键酶,它在很多生理过程中起着重要作用。例如,在人和哺乳动物体内,乙醇脱氢酶与乙醛脱氢酶(ALDH)构成了乙醇脱氢酶系,参与体内乙醇代谢。

[0104] ADH家族蛋白包括ADH1、ADH2、ADH3和/或ADH4蛋白。

[0105] 如图2所示,本发明首次发现,ADH2(SEQ ID NO.:26中的45位-122位),ADH3(SEQ ID NO.:27中的71-148位),ADH4(SEQ ID NO.:28位中的72-149位)与ADH1(SEQ ID NO.:1中47-124位)高度同源(约98.7%),并且,ADH2(H45, H49,H122),ADH3(H71, H75,H148)和ADH4(H72, H76,H149)分别对应于ADH1中的H47,H51和H124,即ADH1、ADH2、ADH3、ADH4的突变位点之间的对应关系如下所示。

ADH 家族蛋白	突变位点		
ADH1	第 47 位	第 51 位	第 124 位
ADH2	第 45 位	第 49 位	第 122 位
ADH3	第 71 位	第 75 位	第 148 位
ADH4	第 72 位	第 76 位	第 149 位

野生型ADH家族蛋白

如本文所用,“野生型ADH家族蛋白”是指天然存在的、未经过人工改造的ADH家族蛋白,其核苷酸可以通过基因工程技术来获得,如基因组测序、聚合酶链式反应(PCR)等,其氨基酸序列可由核苷酸序列推导而得到。所述野生型ADH家族蛋白的来源为*K. lactis*野生型细胞株。野生型ADH家族蛋白包括ADH1、ADH2、ADH3和/或ADH4蛋白。

[0106] 在本发明的一个优选例中,所述野生型ADH家族蛋白(如ADH1、ADH2、ADH3、ADH4)的氨基酸序列分别如SEQ ID NO.:1、26、27或28所示。

[0107]

ADH家族蛋白的突变体及其编码核酸

如本文所用,术语“突变蛋白”、“本发明突变蛋白”、“本发明ADH突变蛋白”、“本发明突变的ADH蛋白”、ADH突变体”、“ADH家族蛋白的突变体”均可互换使用,均指非天然存在的突变的ADH蛋白,并且所述突变蛋白在野生型的乙醇脱氢酶(ADH)的第45-149位氨基酸中的至少1个组氨酸(H)发生突变。

[0108] 并且,在本发明中,所述突变不包括野生型的乙醇脱氢酶(ADH)第67位、69位、93位和/或94位组氨酸(H)。

[0109] 在本发明中,所述组氨酸可各自独立地突变为碱性氨基酸,并且所述组氨酸各自独立突变的类型可以相同,可以不同。

[0110] 在一优选实施方式中,所述组氨酸突变为选自下组的一种或多种氨基酸:赖氨酸(K)、天冬酰胺(N)、谷氨酰胺(Q)、精氨酸(R)、或其组合。

[0111] 在一优选实施方式中,本发明的所述突变的ADH蛋白为对SEQ ID NO.: 1、26、27或28所示蛋白进行人工改造的蛋白。

[0112] 其中,所述的突变蛋白含有与活性相关的核心氨基酸,且所述核心氨基酸中至少有一个是经过人工改造的;并且本发明突变蛋白具有显著(i)提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或(ii)降低突变蛋白与Ni介质的结合能力的活性。

[0113] 术语“核心氨基酸”指的是基于SEQ ID NO.: 1、26、27或28,且与SEQ ID NO.: 1、26、27或28同源性达至少80%,如84%、85%、90%、92%、95%、98%的序列中,相应位点是本文所述

的特定氨基酸,如基于SEQ ID NO.: 1所示的序列,核心氨基酸为:

第47位组氨酸(H);和/或

第51位组氨酸(H);和/或

第124位组氨酸(H);

如基于SEQ ID NO.: 26所示的序列,核心氨基酸为:

第45位组氨酸(H);和/或

第49位组氨酸(H);和/或

第122位组氨酸(H);

如基于SEQ ID NO.: 27所示的序列,核心氨基酸为:

第71位组氨酸(H);和/或

第75位组氨酸(H);和/或

第148位组氨酸(H);

如基于SEQ ID NO.: 28所示的序列,核心氨基酸为:

第72位组氨酸(H);和/或

第76位组氨酸(H);和/或

第149位组氨酸(H);

且对上述核心氨基酸进行突变所得到的突变蛋白具有显著(i)提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或(ii)降低突变蛋白与Ni介质的结合能力的活性。

[0114] 在另一优选例中,所述第47位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、天冬酰胺(N)、谷氨酰胺(Q)、精氨酸(R)或其组合,较佳地,赖氨酸(K)、天冬酰胺(N)和/或精氨酸(R),更佳地,赖氨酸(K)、和/或天冬酰胺(N)。

[0115] 在另一优选例中,所述第51位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)或其组合,较佳地,赖氨酸(K)、和/或精氨酸(R),更佳地,精氨酸(R)。

[0116] 在另一优选例中,第124位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)或其组合,较佳地,赖氨酸(K)。

[0117] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.: 1的选自下组的核心氨基酸发生突变:

第47位组氨酸(H);和

第124位组氨酸(H)。

[0118] 在另一优选例中,第45位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、天冬酰胺(N)、谷氨酰胺(Q)、或其组合。

[0119] 在另一优选例中,第49位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0120] 在另一优选例中,第122位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0121] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.: 26的选自下组的核心氨基酸发生突变:

第122位组氨酸(H) 和

第45位组氨酸(H)。

[0122] 在另一优选例中,第71位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、天冬酰胺(N)、谷氨酰胺(Q)、或其组合。

[0123] 在另一优选例中,第75位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0124] 在另一优选例中,第148位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0125] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.:27的选自下组的核心氨基酸发生突变:

第148位组氨酸(H);和

第71位组氨酸(H)。

[0126] 在另一优选例中,第72位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、天冬酰胺(N)、谷氨酰胺(Q)、或其组合。

[0127] 在另一优选例中,第76位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0128] 在另一优选例中,第149位组氨酸(H)突变为选自下组的一种或多种氨基酸:赖氨酸(K)、精氨酸(R)、或其组合。

[0129] 在另一优选例中,所述突变蛋白在野生型的乙醇脱氢酶(ADH)的对应于SEQ ID NO.:28的选自下组的核心氨基酸发生突变:

第149位组氨酸(H);和

第72位组氨酸(H)。

[0130] 应理解,本发明突变的ADH蛋白中的氨基酸编号基于野生型ADH蛋白(优选地,SEQ ID NO.: 1、26、27或28)作出。当某一具体突变蛋白与SEQ ID NO.: 1、26、27或28所示序列的同源性达到80%或以上时,突变蛋白的氨基酸编号可能会有相对于SEQ ID NO.: 1、26、27或28的氨基酸编号的错位,如向氨基酸的N末端或C末端错位1-5位,而采用本领域常规的序列比对技术,本领域技术人员通常可以理解这样的错位是在合理范围内的,且不当由于氨基酸编号的错位而使同源性达80%(如90%、95%、98%)的、具有相同或相似活性的突变蛋白不在本发明突变蛋白的范围内。

[0131] 本发明突变蛋白是合成蛋白或重组蛋白,即可以是化学合成的产物,或使用重组技术从原核或真核宿主(例如,细菌、酵母、植物)中产生。根据重组生产方案所用的宿主,本发明的突变蛋白可以是糖基化的,或可以是非糖基化的。本发明的突变蛋白还可包括或不包括起始的甲硫氨酸残基。

[0132] 本发明还包括所述突变蛋白的片段、衍生物和类似物。如本文所用,术语“片段”、“衍生物”和“类似物”是指基本上保持所述突变蛋白相同的生物学功能或活性的蛋白。

[0133] 本发明的突变蛋白片段、衍生物或类似物可以是(i)有一个或多个保守或非保守性氨基酸残基(优选保守性氨基酸残基)被取代的突变蛋白,而这样的取代的氨基酸残基可以是也可以不是由遗传密码编码的,或(ii)在一个或多个氨基酸残基中具有取代基团的突变蛋白,或(iii)成熟突变蛋白与另一个化合物(比如延长突变蛋白半衰期的化合物,例如聚乙二醇)融合所形成的突变蛋白,或(iv)附加的氨基酸序列融合到此突变蛋白序列而形

成的突变蛋白(如前导序列或分泌序列或用来纯化此突变蛋白的序列或蛋白原序列,或与抗原IgG片段的形成的融合蛋白)。根据本文的教导,这些片段、衍生物和类似物属于本领域熟练技术人员公知的范围。本发明中,保守性替换的氨基酸最好根据表I进行氨基酸替换而产生。

[0134] 表I

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala
His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

本发明的活性突变蛋白具有显著(i)提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或(ii)降低突变蛋白与Ni介质的结合能力的活性。

[0135] 在本发明中,所述突变蛋白如本发明第一方面所述,优选地,所述的突变蛋白为如SEQ ID NO.:2-13任一所示的氨基酸序列。

[0136] 应理解,本发明突变蛋白与SEQ ID NO.:2-13任一所示的序列相比,通常具有较高的同源性(相同性),优选地,所述的突变蛋白与SEQ ID NO.:2-13任一所示序列的同源性至少为80%,较佳地至少为85%-90%,更佳地至少为95%,最佳地至少为98%。

[0137] 此外,还可以对本发明突变蛋白进行修饰。修饰(通常不改变一级结构)形式包括:体内或体外的突变蛋白的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化,如那些在突变蛋白的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的突变蛋白。这种修饰可以通过将突变蛋白暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸,磷酸丝氨酸,磷酸苏氨酸)的序列。还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的突变蛋白。

[0138] 术语“编码突变蛋白的多核苷酸”可以是包括编码本发明突变蛋白的多核苷酸,也可以是还包括附加编码和/或非编码序列的多核苷酸。

[0139] 本发明的多核苷酸可以是DNA形式或RNA形式。在另一优选例中,所述核苷酸为DNA。DNA形式包括cDNA、基因组DNA或人工合成的DNA。DNA可以是单链的或是双链的。DNA可以是编码链或非编码链。编码成熟多肽的编码区序列可以与编码SEQ ID NO.:2-13任一所示的多肽的序列相同或者是简并的变异体。如本文所用,“简并的变异体”在本发明中是指编码具有SEQ ID NO.:2-13任一所示的多肽,但相应编码区序列有差别的核酸序列。

[0140] 本发明还涉及上述多核苷酸的变异体,其编码与本发明有相同的氨基酸序列的多肽或突变蛋白的片段、类似物和衍生物。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知的,等位变异体是一个多核苷酸的替换形式,它可能是一个或多个核苷酸的取代、缺失或插入,但不会从实质上改变其编码的突变蛋白的功能。

[0141] 核酸序列可以是DNA、RNA、cDNA或PNA。核酸序列可以是基因组的、重组的或合成的。核酸序列可以是分离的或纯化的。核酸序列可以是单链或双链的。优选地,核酸序列将编码如本文描述的光敏蛋白。核酸序列可以通过克隆衍生,例如使用包括限制性酶切、连接、凝胶电泳的标准的分子克隆技术,例如在Sambrook 等Molecular Cloning:A laboratory manual ,Cold Spring Harbour Laboratory Press)中描述的。核酸序列可是分离的,例如使用PCR技术分离的。分离意指从任何杂质和从被自然地发现与其来源中的核酸序列缔合的其他核酸序列和/或蛋白中分离核酸序列。优选地,其还将不含细胞材料、培养基或来自纯化/生产过程的其他化学物质。核酸序列可以是合成的,例如通过直接的化学合成产生。核酸序列可以作为裸露的核酸被提供,或可与蛋白或脂质复合提供。

[0142] 本发明还涉及与上述的序列杂交且两个序列之间具有至少50%,较佳地至少70%,更佳地至少80%相同性的多核苷酸。本发明特别涉及在严格条件(或严紧条件)下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中,“严格条件”是指:(1)在较低离子强度和较高温度下的杂交和洗脱,如 $0.2 \times \text{SSC}$,0.1%SDS,60℃;或(2)杂交时加有变性剂,如50%(v/v)甲酰胺,0.1%小牛血清/0.1% Ficol1,42℃等;或(3)仅在两条序列之间的相同性至少在90%以上,更好是95%以上时才发生杂交。

[0143] 本发明的突变蛋白和多核苷酸优选以分离的形式提供,更佳地,被纯化至均质。

[0144] 本发明多核苷酸全长序列通常可以通过PCR扩增法、重组法或人工合成的方法获得。对于PCR扩增法,可根据本发明所公开的有关核苷酸序列,尤其是开放阅读框序列来设计引物,并用市售的cDNA库或按本领域技术人员已知的常规方法所制备的cDNA库作为模板,扩增而得有关序列。当序列较长时,常常需要进行两次或多次PCR扩增,然后再将各次扩增出的片段按正确次序拼接在一起。

[0145] 一旦获得了有关的序列,就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体,再转入细胞,然后通过常规方法从增殖后的宿主细胞中分离得到有关序列。

[0146] 此外,还可用人工合成的方法来合成有关序列,尤其是片段长度较短时。通常,通过先合成多个小片段,然后再进行连接可获得序列很长的片段。

[0147] 目前,已经可以完全通过化学合成来得到编码本发明蛋白(或其片段,或其衍生物)的DNA序列。然后可将该DNA序列引入本领域中已知的各种现有的DNA分子(或如载体)和细胞中。此外,还可通过化学合成将突变引入本发明蛋白序列中。

[0148] 应用PCR技术扩增DNA/RNA的方法被优选用于获得本发明的多核苷酸。特别是很难从文库中得到全长的cDNA时,可优选使用RACE法(RACE-cDNA末端快速扩增法),用于PCR的引物可根据本文所公开的本发明的序列信息适当地选择,并可用常规方法合成。可用常规方法如通过凝胶电泳分离和纯化扩增的DNA/RNA片段。

[0149] 在本发明中,所述ADH蛋白突变体的DNA编码序列为SEQ ID NO.:14-25任一所示的核苷酸序列

表达载体和宿主细胞

本发明也涉及包含本发明的多核苷酸的载体,以及用本发明的载体或本发明突变蛋白编码序列经基因工程产生的宿主细胞,以及经重组技术产生本发明所述多肽的方法。

[0150] 通过常规的重组DNA技术,可利用本发明的多聚核苷酸序列可用来表达或生产重组的突变蛋白。一般来说有以下步骤:

(1).用本发明的编码本发明突变蛋白的多核苷酸(或变异体),或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞;

(2).在合适的培养基中培养的宿主细胞;

(3).从培养基或细胞中分离、纯化蛋白质。

[0151] 本发明中,编码突变蛋白的多核苷酸序列可插入到重组表达载体中。术语“重组表达载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、植物细胞病毒、哺乳动物细胞病毒如腺病毒、逆转录病毒或其他载体。只要能在宿主体内复制和稳定,任何质粒和载体都可以用。表达载体的一个重要特征是通常含有复制起点、启动子、标记基因和翻译控制元件。

[0152] 本领域的技术人员熟知的方法能用于构建含本发明突变蛋白编码DNA序列和合适的转录/翻译控制信号的表达载体。这些方法包括体外重组DNA技术、DNA合成技术、体内重组技术等。所述的DNA序列可有效连接到表达载体中的适当启动子上,以指导mRNA合成。这些启动子的代表性例子有:大肠杆菌的lac或trp启动子; λ 噬菌体PL启动子;真核启动子包括CMV立即早期启动子、HSV胸苷激酶启动子、早期和晚期SV40启动子、反转录病毒的LTRs和其他一些已知的可控制基因在原核或真核细胞或其病毒中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子。

[0153] 此外,表达载体优选地包含一个或多个选择性标记基因,以提供用于选择转化的宿主细胞的表型性状,如真核细胞培养用的二氢叶酸还原酶、新霉素抗性以及绿色荧光蛋白(GFP),或用于大肠杆菌的四环素或氨苄青霉素抗性。

[0154] 包含上述的适当DNA序列以及适当启动子或者控制序列的载体,可以用于转化适当的宿主细胞,以使其能够表达蛋白质。

[0155] 宿主细胞可以是原核细胞(如大肠杆菌),或是低等真核细胞,或是高等真核细胞,如酵母细胞、植物细胞或哺乳动物细胞(包括人和非人哺乳动物)。代表性例子有:大肠杆菌、麦胚细胞,昆虫细胞,SF9、HeLa、HEK293、CHO、酵母细胞等。在本发明的一个优选实施方式中,选择酵母细胞(如毕氏酵母、克鲁维酵母、或其组合;较佳地,所述的酵母细胞包括:克鲁维酵母,更佳地为马克斯克鲁维酵母、和/或乳酸克鲁维酵母)为宿主细胞。

[0156] 本发明的多核苷酸在高等真核细胞中表达时,如果在载体中插入增强子序列时将会使转录得到增强。增强子是DNA的顺式作用因子,通常大约有10到300个碱基对,作用于启动子以增强基因的转录。可举的例子包括在复制起始点晚期一侧的100到270个碱基对的

SV40增强子、在复制起始点晚期一侧的多瘤增强子以及腺病毒增强子等。

[0157] 本领域一般技术人员都清楚如何选择适当的载体、启动子、增强子和宿主细胞。

[0158] 用重组DNA转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时,能吸收DNA的感受态细胞可在指数生长期后收获,用CaCl₂法处理,所用的步骤在本领域众所周知。另一种方法是使用MgCl₂。如果需要,转化也可用电穿孔的方法进行。当宿主是真核生物,可选用如下的DNA转染方法:磷酸钙共沉淀法,常规机械方法如显微注射、电穿孔、脂质体包装等。

[0159] 获得的转化子可以用常规方法培养,表达本发明的基因所编码的多肽。根据所用的宿主细胞,培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后,用合适的方法(如温度转换或化学诱导)诱导选择的启动子,将细胞再培养一段时间。

[0160] 在上面的方法中的重组多肽可在细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要,可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法的例子包括但不限于:常规的复性处理、用蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超处理、超离心、分子筛层析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

[0161]

本发明的主要优点包括:

(1) 本发明通过对细胞中与Ni介质结合成分的质谱鉴定,确定了能够与Ni介质结合的细胞内蛋白的种类;

(2) 本发明通过对细胞基因组中ADH家族蛋白进行系统分析,首次提供了一系列ADH家族蛋白中可能与Ni介质结合的组氨酸位点。

[0162] (3) 本发明对ADH1进行基因敲除并对ADH1中一系列可能与Ni介质结合的组氨酸位点进行定点突变,首次得到了多株可显著降低ADH1与Ni介质结合能力的菌株。

[0163] (4) 本发明通过对上述多个细胞株进行无细胞体系转录翻译活性分析,发现相对于基因敲除,ADH1的定点突变对无细胞体系转录翻译活性的影响较小,并首次得到了一株既不影响无细胞体系转录翻译活性,又显著降低ADH1与Ni介质结合能力的细胞株。

[0164] (5) 本发明以乳酸克鲁维酵母(*Kluyveromyces lactis*, K. *lactis*)为实施例,但同样的设计和分析、实验方法也适用于其他原核细胞,真核细胞,酵母,人源细胞,HeLa, CHO, HEK293, *Saccharomyces cerevisiae* 等细胞。

[0165] (6) 本发明首次发现,对ADH家族蛋白的核心氨基酸进行突变,会显著(i) 提高体外无细胞合成体系中外源蛋白的表达纯度、效率、和/或得率;和/或(ii) 降低突变蛋白与Ni介质的结合能力。

[0166] 下面结合具体实施例,进一步阐述本发明。应理解,这些实施例仅用于说明本发明而不适用于限制本发明的范围。下列实施例中未注明具体条件的实验方法,通常按照常规条件,例如Sambrook等人, 分子克隆:实验室手册(New York: Cold Spring Harbor Laboratory Press, 1989)中所述的条件,或按照制造厂商所建议的条件。除非另外说明,否则百分比和份数是重量百分比和重量份数。

[0167] 除非有特别说明,否则本发明实施例中的试剂和材料均为市售产品。

[0168] 本发明以乳酸克鲁维酵母(*Kluyveromyces lactis*, *K. lactis*)为实施例,但同样的设计和分析、实验方法也适用于其他原核细胞,真核细胞,酵母,人源细胞,HeLa, CHO, HEK293, *Saccharomyces cerevisiae* 等细胞。

[0169]

通用方法

本发明提供了一种设计和方法,即通过Ni纯化和质谱分析,鉴定Ni结合蛋白,并以此对基因组中ADH家族蛋白进行点突变,降低其与Ni介质结合的能力,从而提高经由Ni纯化蛋白的效率、得率和纯度等优点。包括以下步骤:

(1) ADH家族蛋白的鉴定分析,方法如下:

A、经由Ni纯化的蛋白质聚丙烯酰胺凝胶电泳分析;

B、非特异性条带(35–50 kDa)的质谱分析;

C、质谱分析结果的肽段比对,鉴定主要非特异性条带为ADH蛋白;

(2) *K. lactis* ADH家族蛋白与Ni介质潜在结合位点分析,方法如下:

A、*K. lactis*中ADH1与*Saccharomyces cerevisiae*中ADH1蛋白进行序列比对;

B、以ScADH为模板进行同源建模,并找出KlADH中的His富集区域位点,如ADH1的H47, H51, H124; ADH2的H45, H49, H122; ADH3的H71, H75, H148; ADH4的H72, H76, H149;

C、对该区域内的His位点进行可改造性分析,如是否影响ADH的催化活性,并最终选定替代氨基酸,如ADH1的定点突变选自下组:H124K, H124R, H47K, H47R, H47N, H47Q, H51K, H51R, H124KH47K, H124KH47R, H124KH47N, H124KH47Q或其组合;ADH2的定点突变选自下组:H122K, H122R, H45K, H45R, H45N, H45Q, H49K, H49R, H122KH45K, H122KH45R, H122KH45N, H122KH45Q或其组合;ADH3的定点突变选自下组:H148K, H148R, H71K, H71R, H71N, H71Q, H75K, H75R, H148KH71K, H148KH71R, H148KH71N, H148KH71Q或其组合;ADH4的定点突变选自下组:H149K, H149R, H72K, H72R, H72N, H72Q, H76K, H76R, H149KH72K, H149KH72R, H149KH72N, H149KH72Q或其组合;

(3) Cas9-gRNA克隆载体的构建,构建方法如下:

A、针对特定基因的序列,分别设计引导该基因切割的gRNA序列;

B、将上述gRNA序列重组至含有Cas9的载体中,得到gRNA和Cas9共表达的第一载体;

(4) 构建敲除特定基因的供体DNA,方法如下:

A、从基因数据库中下载特定基因核苷酸序列,用基因上下游序列各800 bp构建第二载体;

B、利用引物M13F和M13R对第二载体进行扩增,获得的PCR产物进行乙醇沉淀浓缩,最终获得供体DNA;

(5) 构建特定基因组氨基酸残基点突变的供体DNA,方法如下:

A、从基因数据库中下载特定基因核苷酸序列,用目的基因和上下游序列各800bp构建第二载体,同时对部分氨基酸残基进行相应的点突变,并对gRNA进行同义突变;

B、利用引物M13F和M13R对第二载体进行扩增,获得的PCR产物进行乙醇沉淀浓缩,最终获得供体DNA;

(6) 获得敲除或点突变特定基因的细胞株,方法如下:

A、将第一载体和供体DNA同时转化到感受态细胞中;

B、筛选出单克隆细胞进行扩大培养,抽提基因组,设计引物对ADH基因及同源臂进行扩增,扩增的PCR产物经测序验证;

(7) 检测基因敲除或点突变菌株和Ni介质的结合能力,方法如下:

A、将得到的基因敲除或点突变菌株进行扩大培养,并进行细胞破碎,得到细胞裂解液;

B、对细胞裂解液通过Ni介质进行纯化,并对纯化产物进行聚丙烯酰胺凝胶电泳分析;

(8) 无细胞体外蛋白质合成体系

利用基因改造后的细胞株制备成酵母细胞裂解液,加入到蛋白体外翻译体系中,20-30℃的环境中,静置反应2-12 h,利用多功能酶标仪(Perkin Elmer)读数,检测增强型绿色荧光蛋白活性。

[0170] 实施例1 *K. lactis*中ADH家族基因的分析 and 特异性改造

1.1 *K. lactis*中ADH家族基因的分析

当用Ni介质对*K. lactis* 中表达的含有His标签的目的蛋白进行亲和纯化时,在35-50 kDa处有明显的非特异性条带。质谱结果表明该非特异性蛋白主要来源于*K. lactis* 中的ADH家族蛋白(表1)。ADH是一种依赖NAD(P)的氧化还原酶,几乎存在于所有生物体中,它们催化伯醇和仲醇的可逆氧化分别生成醛和酮。*K. lactis*中的ADH家族蛋白包括*K1ADH1*、*K1ADH2*、*K1ADH3*、*K1ADH4*四种,均由染色体DNA编码。其中,ADH1基因序列(SEQ ID NO.164),KEGG数据库编码为*K1LA0F21010g*,位于F染色体;ADH2基因序列(SEQ ID NO.29),KEGG数据库编码为*K1LA0F18260g*,位于F染色体;ADH3基因序列(SEQ ID NO.30),KEGG数据库编码为*K1LA0B09064g*,位于B染色体;ADH4基因序列(SEQ ID NO.31),KEGG数据库编码为*K1LA0F13530g*,位于F染色体。其中ADH1和ADH2在细胞溶质中发挥功能,ADH3和ADH4在线粒体中发挥功能,相对分子量分别为37.3 kDa、37.1 kDa、39.6 kDa、40.2 kDa。这四种ADH的氨基酸序列同源性很高,三级结构也很接近。

[0171] 表1 质谱鉴定Ni介质纯化非特异性蛋白所得肽段及所指向的

K. lactis 内源蛋白

peptide sequence	target protein in <i>K.lactis</i>	SEQ ID NO.:
EAIDFFSR	ADH1	32
SNGTVVLVGLPR	ADH1&ADH2	33
SISIVGSYVGNR	ADH1	34
DLGGEYFIDFTK	ADH1	35
GVIFYENGELQYK	ADH1&ADH2	36
LPLVGGHEGAGVVVAMGENVK	ADH1&ADH2&ADH3 (mitochondria)	37
APIHVVLSELPSIYEK	ADH1	38
GGAHGVINVSVEFAIEQSTNYVR	ADH1&ADH2	39
AGDWVAISGAAGGLGSLAVQYAK	ADH1&ADH3 (mitochondria)	40
GVIFYENGK1EYK	ADH3&ADH4 (mitochondria)	41
SDVFNQVVK	ADH1	42
NIPEEVIEATK	ADH1	43
ANEILINVK	ADH4	44

由于*K. lactis* 中ADH家族蛋白的晶体结构还未知,我们采用了同源建模的方式预测

其空间三级结构。*K. lactis* 中*KlADH1*与酿酒酵母(*Saccharomyces cerevisiae*)中*ScADH1* (PDB编号4W6Z)的一级序列比对表明两者同为四聚体,序列同源性84.9%。以*ScADH1*为模板进行的同源建模显示,*K. lactis* 中*KlADH1*的每个单体共含有6个His,分别位于47位、51位、69位、124位、243位、321位。整个四聚体共含有4个His富集区,每个单体除第5个His相对独立外,另5个His空间靠近,其中空间上(同时也是序列上)最为靠近的是每个单体的前3~4个His。

[0172] 由于*K. lactis*中的四种ADH高度同源,且有5个His高度保守,所以*KlADH2*,*KlADH3*和*KlADH4*的空间结构也与*KlADH1*相近,即每种ADH蛋白的前4个His可能在空间上成簇,从而与Ni介质产生亲和作用。

[0173]

1.2 *K. lactis*中ADH家族基因的特异性改造

鉴于在消除ADH与Ni介质的亲和作用的同时,需要尽可能地减少对ADH活性的影响,所以分析*KlADH*中His的功能成为关键。通过大量比对和筛选发现,*ScADH1*蛋白与*KlADH1*同源的6个His中,H44、H48、H66、H121由于与*KlADH1*中成簇的4个His同源而成为本发明分析的重点。由于H66在*ScADH1*的两种构象中均参与螯合具有催化作用的 Zn^{2+} ,如图6所示,可能对ADH的催化活性有重大影响,故而没有成为本发明的改造靶点。而H48虽不与辅酶结合,但参与三元复合体的形成,影响底物的质子转移过程,如图7所示。本发明的研究发现,H44咪唑环上的ND1可能与 NAD^+ 的O2A形成氢键,从而参与了 NAD^+ / $NADH$ 的结合,H44R可以将 NAD^+ 和 $NADH$ 的解离常数降低2-4倍,并且使转换数减少4-6倍。此外,H121的侧链可与T130形成氢键。

[0174] 基于上述研究,本发明将对*ScADH1*中对ADH1功能影响较小的三个His位点(分别为H44,H48和H121)突变为功能相近的氨基酸,最终在*KlADH1*上设计的点突变方案为H124K, H124R, H47K, H47R, H47N, H47Q, H51K, H51R, H124KH47K, H124KH47R, H124KH47N, H124KH47Q。

[0175] 由于*K. lactis*中ADH2,ADH3,ADH4与ADH1高度同源(图2),即ADH2 (H45, H49, H122),ADH3 (H71, H75,H148)和ADH4 (H72, H76,H149)分别对应于ADH1中的H47,H51和H124。

[0176] 因此在本发明中,以ADH1为例来减少ADH1对Ni介质的结合,具体表现在ADH1基因的敲除及H47,H51和H124的定点突变。ADH2,ADH3和ADH4也可在对应的位点用类似ADH1的方法进行改造。

[0177]

实施例2 通过CRISPR/Cas9靶向敲除ADH1基因

2.1 *KlADH1* CRISPR gRNA序列确定

根据在*KlADH1*设计的点突变,选择PAM序列(NGG),并确定对应的gRNA序列。本实施例中gRNA选择的原则为:GC含量适中(40%-60%),避免poly T结构的存在。在本实施例中,*KlADH1* gRNA1序列为TGGGTGAAAACGTCAAGGGC (SEQ ID NO.:45)。

[0178] 质粒构建及转化方法如下:使用引物pKM Cas9-*KlADH1*-gRNA1-PF: TGGGTGAAAACGTCAAGGGCGTTTGTAGAGCTAGAAATAGC (SEQ ID NO.:46)和pCas9-*KlADH1*-gRNA1-PR:GCCCTTGACGTTTTACCCAAAAGTCCCATTCGCCACCCG (SEQ ID NO.:47),以pCAS质粒为模板,进行PCR扩增。取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后

37 °C温浴3 h。取Dpn I处理后产物10 μL加入50 μL DH5α感受态细胞中,冰上放置30 min,42 °C热激45 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Kan抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取2个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMCas9_StR_KIADH1。

[0179] 2.2 供体 DNA质粒构建及扩增

为了便于线性供体DNA的保存及扩增,本实施例将供体DNA插入到pKMD1质粒中,并通过PCR扩增得到线性供体DNA序列。

[0180] 以pKMD1质粒为模板,以引物pKMD1-PF:ATCGTCGACCTGCAGGCATG (SEQ ID NO.:48)和pKMD1-PR:ATCTCTAGAGGATCCCCGGG (SEQ ID NO.:49)进行PCR扩增,取扩增产物17 μL,加入1 μL Dpn I,2 μL 10 × digestion buffer,混匀后37 °C温浴3 h,获得质粒骨架线性片段pKMD1-T。

[0181] 2.2.1 构建供体质粒pKMD1-ΔKIADH1

以乳酸克鲁维酵母基因组DNA为模板,用引物KIADH1-HR1-PF:TCGAGCTCGGTACCCGGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:50)和KIADH1-HR1-PR:GAAGAGATTTTCAATTTATCTTTTTTTTAGTATAGAGTTTGTGTGTTTAAAGCTTG (SEQ ID NO.:51)进行PCR扩增,产物名为ΔKIADH1-F1;以乳酸克鲁维酵母基因组DNA为模板,用引物KIADH1-HR2-PF:CAAACCTATACTAAAAAAGATAAATGAAATCTCTTCCGCATTCAAGTCATGAC (SEQ ID NO.:

52)和KIADH1-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAA CGGGAC (SEQ ID NO.:53)进行PCR扩增,产物名为ΔKIADH1-F2。

[0182] 将扩增产物ΔKIADH1-F1、ΔKIADH1-F2、pKMD1-T各1 μL,加入5 μL Cloning Mix (试剂盒名为Transgene pEASY-Uni Seamless Cloning and Assembly Kit,来源于全式金公司,下同)与2 μL水,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞(来源于全式金公司,下同)中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-ΔKIADH1。

[0183] 2.3 *K. lactis*电转化

从-80 °C 冰箱取出感受态,冰上融化,加入400 ng gRNA&Cas9质粒(或gRNA/cas9 片段)和1000 ng Donor DNA 片段,混匀后全部转入电击杯中,冰浴 2 min;将电击杯放入电转仪中进行电击(参数为1.5 kV,200 Ω,25 μF);电击结束后立即加入 700 μL YPD,30 °C,200 rpm 摇床孵育 1~3 h;取 2~200 μL 接种于YPD(含G418抗性)平板,30 °C 培养2~3天至单菌落出现。

[0184] 2.4 阳性鉴定

在细胞转化后的平板上挑取12-24个单克隆,以细胞为模板,用鉴定引物 ADH1-CF:GTGATGGAACACGGGAATAG (SEQ ID NO.:54)和ADH1-CR:CACATATACCTTGGCAGTAG (SEQ ID NO.:55)对样品进行PCR检测。PCR结果阳性并经测序鉴定的细胞株,确定为阳性细胞株并分别命名为ΔADH1。

[0185]

实施例3 通过CRISPR/Cas9定点突变ADH1 H47位点

3.1 *KIADH1* CRISPR gRNA序列确定

根据在*KIADH1*设计的点突变,选择PAM序列(NGG),并确定对应的gRNA序列。本实施例中gRNA选择的原则为:靠近设计的突变位点,GC含量适中(40%-60%),避免poly T结构的存在。在本实施例中,*KIADH1* gRNA1序列为TGGGTGAAAACGTCAAGGGC(SEQ ID NO.:56)。

[0186] 质粒构建及转化方法如下:使用引物pCas9-*KIADH1*-gRNA1-PF:TGGGTGAAAACGTCAAGGGCGTTTGTAGAGCTAGAAATAGC(SEQ ID NO.:57)和pCas9-*KIADH1*-gRNA1-PR:GCCCTTGACGTTTTCACCCAAAAGTCCCATTCGCCACCCG(SEQ ID NO.:58),以pCAS质粒为模板,进行PCR扩增。取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h。取Dpn I处理后产物10 μ L加入50 μ L DH5 α 感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激45 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Kan抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取2个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMCas9_StR_*KIADH1*。

[0187] 3.2 供体 DNA质粒构建及扩增

为了便于线性供体DNA的保存及扩增,本实施例将供体DNA插入到pKMD1质粒中,并通过PCR扩增得到线性供体DNA序列。

[0188] 以pKMD1质粒为模板,以引物pKMD1-PF:ATCGTCGACCTGCAGGCATG(SEQ ID NO.:59)和pKMD1-PR:ATCTCTAGAGGATCCCCGGG(SEQ ID NO.:60)进行PCR扩增,取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h,获得质粒骨架线性片段pKMD1-T。

[0189] 3.2.1 构建供体质粒pKMD1-*KIADH1*-H47K

以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-HR1-PF:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG(SEQ ID NO.:61)和*KIADH1*-H47K-PR:aGCaTGaAggTCaGtTtACAGACACCGGAGTACTTGACG(SEQ ID NO.:62)进行PCR扩增,产物名为*KIADH1*-H47K-F1;以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-H47K-PF:GTaAaACtGAccTtCaTGCtTGGAAGGGTGACTGGCCTTTG(SEQ ID NO.:63)和*KIADH1*-gRNA1-m-PR:CCAaCCtTTaACaTtTcTCCCATAGCAACAACGACACCAG(SEQ ID NO.:64)进行PCR扩增,产物名为*KIADH1*-H47K-F2;以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-gRNA1-m-PF:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG(SEQ ID NO.:65)和*KIADH1*-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC(SEQ ID NO.:66)进行PCR扩增,产物名为*KIADH1*-H47K-F3。

[0190] 将扩增产物*KIADH1*-H47K-F1、*KIADH1*-H47K-F2、*KIADH1*-H47K-F3、pKMD1-T各1 μ L,加入5 μ L Cloning Mix 与1 μ L水,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*KIADH1*-H47K。

[0191] 3.2.2 构建供体质粒pKMD1-*KIADH1*-H47N

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:67)和*K1ADH1-H47N-PR*:aGCaTGaAggTCaGTaTtACAGACACCGGAGTACTTGACG (SEQ ID NO.:68)进行PCR扩增,产物名为*K1ADH1-H47N-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H47N-PF*:GTaAtACtGAccTtCaTGcTtGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:69)和*K1ADH1-gRNA1-m-PR*:CCAaCCtTTaACaTtTcTtCCCATAGCAACAACGACACCAG (SEQ ID NO.:70)进行PCR扩增,产物名为*K1ADH1-H47N-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:71)和*K1ADH1-HR2-PR*:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:72)进行PCR扩增,产物名为*K1ADH1-H47N-F3*。

[0192] 将扩增产物*K1ADH1-H47N-F1*、*K1ADH1-H47N-F2*、*K1ADH1-H47N-F3*、pKMD1-T各1 μL,加入5 μL Cloning Mix与1 μL水,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H47N*。

[0193] 3.2.3构建供体质粒pKMD1-*K1ADH1-H47Q*

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:73)和*K1ADH1-H47Q-PR*:aGCaTGaAggTCaGTtTGACAGACACCGGAGTACTTGACG (SEQ ID NO.:74)进行PCR扩增,产物名为*K1ADH1-H47Q-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H47Q-PF*:GTCAaACtGAccTtCaTGcTtGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:75)和*K1ADH1-gRNA1-m-PR*:CCAaCCtTTaACaTtTcTtCCCATAGCAACAACGACACCAG (SEQ ID NO.:76)进行PCR扩增,产物名为*K1ADH1-H47Q-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:77)和*K1ADH1-HR2-PR*:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:78)进行PCR扩增,产物名为*K1ADH1-H47Q-F3*。

[0194] 将扩增产物*K1ADH1-H47Q-F1*、*K1ADH1-H47Q-F2*、*K1ADH1-H47Q-F3*、pKMD1-T各1 μL,加入5 μL Cloning Mix与1 μL水,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H47Q*。

[0195] 3.2.4构建供体质粒pKMD1-*K1ADH1-H47R*

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:79)和*K1ADH1-H47R-PR*:aGCaTGaAggTCaGTtctACAGACACCGGAGTACTTGACG (SEQ ID NO.:80)进行PCR扩增,产物名为*K1ADH1-H47R-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H47R-PF*:GTagaACtGAccTtCaTGcTtGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:81)和*K1ADH1-*

gRNA1-m-PR:CCAAcCtTTaACaTTcTCtCCCATAGCAACAACGACACCAG (SEQ ID NO.:82) 进行PCR扩增,产物名为*KlADH1*-H47R-F2;以乳酸克鲁维酵母基因组DNA为模板,用引物*KlADH1*-gRNA1-m-PF:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:83) 和*KlADH1*-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:84) 进行PCR扩增,产物名为*KlADH1*-H47R-F3。

[0196] 将扩增产物*KlADH1*-H47R-F1、*KlADH1*-H47R-F2、*KlADH1*-H47R-F3、pKMD1-T各1 μ L,加入5 μ L Cloning Mix与1 μ L水,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*KlADH1*-H47R。

[0197] 3.3 *K. lactis*电转化(同实施例2.3)

3.4 阳性鉴定

在细胞转化后的平板上挑取12-24个单克隆,以细胞为模板,用鉴定引物 ADH1-m-CF: TGGGAGAGAATGTAAAGGT (SEQ ID NO.:85) 和ADH1-m-CR: TGACGGTCGTAACTAAGAT (SEQ ID NO.:86) 对样品进行PCR检测。PCR结果阳性并经测序鉴定的细胞株,确定为阳性细胞株并分别命名为ADH1 H47K, ADH1 H47N, ADH1 H47Q和ADH1 H47R。

[0198]

实施例4 通过CRISPR/Cas9定点突变ADH1 H51位点

4.1 *KlADH1* CRISPR gRNA序列确定

根据在*KlADH1*设计的点突变,选择PAM序列(NGG),并确定对应的gRNA序列。本实施例中gRNA选择的原则为:靠近设计的突变位点,GC含量适中(40%-60%),避免poly T结构的存在。在本实施例中,*KlADH1* gRNA1序列为TGGGTGAAAACGTCAAGGGC (SEQ ID NO.:87)。

[0199] 质粒构建及转化方法如下:使用引物 pCas9-*KlADH1*-gRNA1-PF: TGGGTGAAAACGTCAAGGGCGTTTATAGCTAGAAATAGC (SEQ ID NO.:88) 和pCas9-*KlADH1*-gRNA1-PR: GCCCTTGACGTTTTCACCCAAAAGTCCCATTCGCCACCCG (SEQ ID NO.:89),以pCAS质粒为模板,进行PCR扩增。取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h。取Dpn I处理后产物10 μ L加入50 μ L DH5 α 感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激45 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Kan抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取2个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMCas9-StR-*KlADH1*。

[0200] 4.2 供体 DNA质粒构建及扩增

为了便于线性供体DNA的保存及扩增,本实施例将供体DNA插入到pKMD1质粒中,并通过PCR扩增得到线性供体DNA序列。

[0201] 以pKMD1质粒为模板,以引物pKMD1-PF: ATCGTCGACCTGCAGGCATG (SEQ ID NO.:90) 和pKMD1-PR: ATCTCTAGAGGATCCCCGGG (SEQ ID NO.:91) 进行PCR扩增,取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h,获得质粒骨架线性片段pKMD1-T。

[0202] 4.2.1 构建供体质粒pKMD1-*KlADH1*-H51K

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:92)和*K1ADH1-H51K-PR*:CCAAgCtTtaAggTCaGTaTGACAGACACCGGAGTACTTGACG (SEQ ID NO.:93)进行PCR扩增,产物名为*K1ADH1-H51K-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H51K-PF*:GTCAtACtGAccTtaAaGCtTGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:94)和*K1ADH1-gRNA1-m-PR*:CCAAcCtTTaACaTTcTCtCCCATAGCAACAACGACACCAG (SEQ ID NO.:95)进行PCR扩增,产物名为*K1ADH1-H51K-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:96)和*K1ADH1-HR2-PR*:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:97)进行PCR扩增,产物名为*K1ADH1-H51K-F3*。

[0203] 将扩增产物*K1ADH1-H51K-F1*、*K1ADH1-H51K-F2*、*K1ADH1-H51K-F3*、pKMD1-T各1 μ L,加入5 μ L Cloning Mix与1 μ L水,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H51K*。

[0204] 4.2.2构建供体质粒pKMD1-*K1ADH1-H51R*

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:98)和*K1ADH1-H51R-PR*:CCAAgCtctaAggTCaGTaTGACAGACACCGGAGTACTTGACG (SEQ ID NO.:99)进行PCR扩增,产物名为*K1ADH1-H51R-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H51R-PF*:GTCAtACtGAccTtagaGCtTGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:100)和*K1ADH1-gRNA1-m-PR*:CCAAcCtTTaACaTTcTCtCCCATAGCAACAACGACACCAG (SEQ ID NO.:101)进行PCR扩增,产物名为*K1ADH1-H51R-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:102)和*K1ADH1-HR2-PR*:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:103)进行PCR扩增,产物名为*K1ADH1-H51R-F3*。

[0205] 将扩增产物*K1ADH1-H51R-F1*、*K1ADH1-H51R-F2*、*K1ADH1-H51R-F3*、pKMD1-T各1 μ L,加入5 μ L Cloning Mix与1 μ L水,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H51R*。

[0206] 4.3 乳酸克鲁维酵母电转化(同实施例2.3)

4.4 阳性鉴定

在细胞转化后的平板上挑取12-24个单克隆,以细胞为模板,用鉴定引物*ADH1-m-CF*:TGGGAGAGAATGTAAAGGT (SEQ ID NO.:104)和*ADH1-m-CR*:TGACGGTCGTTAACTAAGAT (SEQ ID NO.:105)对样品进行PCR检测。PCR结果阳性并经测序鉴定的细胞株,确定为阳性细胞株并分别命名为ADH1 H51K和ADH1 H51R。

[0207]

实施例5 通过CRISPR/Cas9定点突变ADH1 H124位点

5.1 *KIADH1* CRISPR gRNA序列确定

根据在*KIADH1*设计的点突变,选择PAM序列(NGG),并确定对应的gRNA序列。本实施例中gRNA选择的原则为:靠近设计的突变位点,GC含量适中(40%-60%),避免poly T结构的存在。在本实施例中,*KIADH1* gRNA1序列为TGGGTGAAAACGTCAAGGGC(SEQ ID NO.:106)。

[0208] 质粒构建及转化方法如下:使用引物pCas9-*KIADH1*-gRNA1-PF:TGGGTGAAAACGTCAAGGGCGTTTGTAGAGCTAGAAATAGC(SEQ ID NO.:107)和pCas9-*KIADH1*-gRNA1-PR:GCCCTTGACGTTTTACCCAAAAGTCCCATTGCGCCACCCG(SEQ ID NO.:108),以pCAS质粒为模板,进行PCR扩增。取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h。取Dpn I处理后产物10 μ L加入50 μ L DH5 α 感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激45 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Kan抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取2个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMCas9_StR_*KIADH1*。

[0209] 5.2 供体 DNA质粒构建及扩增

为了便于线性供体DNA的保存及扩增,本实施例将供体DNA插入到pKMD1质粒中,并通过PCR扩增得到线性供体DNA序列。

[0210] 以pKMD1质粒为模板,以引物pKMD1-PF:ATCGTCGACCTGCAGGCATG(SEQ ID NO.:109)和pKMD1-PR:ATCTCTAGAGGATCCCCGGG(SEQ ID NO.:110)进行PCR扩增,取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h,获得质粒骨架线性片段pKMD1-T。

[0211] 5.2.1构建供体质粒pKMD1-*KIADH1*-H124K

以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-HR1-PF:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG(SEQ ID NO.:111)和*KIADH1*-gRNA1-m-PR:CCAaCCtTTaCaTtTcTcCCCATAGCAACAACGACACCAG(SEQ ID NO.:112)进行PCR扩增,产物名为*KIADH1*-H124K-F1;以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-gRNA1-m-PF:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG(SEQ ID NO.:113)和*KIADH1*-H124K-PR:CCaTcTtTtGTaTaTCCactCAAGTCAGTTCTGGACAGTTGG(SEQ ID NO.:114)进行PCR扩增,产物名为*KIADH1*-H124K-F2;以乳酸克鲁维酵母基因组DNA为模板,用引物*KIADH1*-H124K-PF:CTTGagtGGaTAtACaaAaGAtGGTTCTTTCCAACAATACGCTACTGC(SEQ ID NO.:115)和*KIADH1*-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTGACGATCTTATACTGGGTAGATTTCAAACGGGAC(SEQ ID NO.:116)进行PCR扩增,产物名为*KIADH1*-H124K-F3。

[0212] 将扩增产物*KIADH1*-H124K -F1、*KIADH1*-H124K -F2、*KIADH1*-H124K -F3、pKMD1-T各1 μ L,加入5 μ L Cloning Mix与1 μ L水,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*KIADH1*-H124K。

[0213] 5.2.2构建供体质粒pKMD1-*KIADH1*-H124R

以乳酸克鲁维酵母基因组DNA为模板,用引物 *K1ADH1-HR1-PF*: TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:117) 和 *K1ADH1-gRNA1-m-PR*: CCAaCCtTTaACaTTcTcTCCCATAGCAACAACGACACCAG (SEQ ID NO.:118) 进行PCR扩增,产物名为*K1ADH1-H124K-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物 *K1ADH1-gRNA1-m-PF*: GGGaGAgAAAtGTtAAaGGtTGGAAGATTGGTGAAGTTCGCTGG (SEQ ID NO.:119) 和 *K1ADH1-H124R-PR*: CCaTCtcttGTaTAtCCactCAAGTCAGCTTCTGGACAGTTGG (SEQ ID NO.:120) 进行PCR扩增,产物名为*K1ADH1-H124R-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物 *K1ADH1-H124R-PF*: CTTGagtGGaTAtACaagaGAtGGTTCTTTCCAACAATACGCTACTGC (SEQ ID NO.:121) 和 *K1ADH1-HR2-PR*: TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:122) 进行PCR扩增,产物名为*K1ADH1-H124R-F3*。

[0214] 将扩增产物*K1ADH1-H124R-F1*、*K1ADH1-H124R-F2*、*K1ADH1-H124R-F3*、pKMD1-T各1 μL,加入5 μL Cloning Mix与1 μL水,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H124R*。

[0215] 5.3 *K. lactis*电转化(同实施例2.3)

5.4 阳性鉴定

在细胞转化后的平板上挑取12-24个单克隆,以细胞为模板,用鉴定引物 *ADH1-m-CF*: TGGGAGAGAATGTAAAGGT (SEQ ID NO.:123) 和 *ADH1-m-CR*: TGACGGTCGTAACTAAGAT (SEQ ID NO.:124) 对样品进行PCR检测。PCR结果阳性并经测序鉴定的细胞株,确定为阳性细胞株并分别命名为*ADH1 H124K*和*ADH1 H124R*。

[0216]

实施例6 通过CRISPR/Cas9同时定点突变*ADH1 H47*和*H124*位点

6.1 *K1ADH1* CRISPR gRNA序列确定

根据在*K1ADH1*设计的点突变,选择PAM序列(NGG),并确定对应的gRNA序列。本实施例中gRNA选择的原则为:靠近设计的突变位点,GC含量适中(40%-60%),避免poly T结构的存在。在本实施例中,*K1ADH1* gRNA1序列为TGGGTGAAAACGTCAAGGGC (SEQ ID NO.:125)。

[0217] 质粒构建及转化方法如下:使用引物 *pCas9-K1ADH1-gRNA1-PF*: TGGGTGAAAACGTCAAGGGCGTTTGTAGAGCTAGAAATAGC (SEQ ID NO.:126) 和 *pCas9-K1ADH1-gRNA1-PR*: GCCCTTGACGTTTTTACCCAAAAGTCCCATTGCCACCCG (SEQ ID NO.:127),以pCAS质粒为模板,进行PCR扩增。取扩增产物17 μL,加入1 μL Dpn I,2 μL 10 × digestion buffer,混匀后37 °C温浴3 h。取Dpn I处理后产物10 μL加入50 μL DH5α感受态细胞中,冰上放置30 min,42 °C热激45 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Kan抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取2个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMCas9-StR-*K1ADH1*。

[0218] 6.2 供体DNA质粒构建及扩增

为了便于线性供体DNA的保存及扩增,本实施例将供体DNA插入到pKMD1质粒中,并通过PCR扩增得到线性供体DNA序列。

[0219] 以pKMD1质粒为模板,以引物pKMD1-PF:ATCGTCGACCTGCAGGCATG (SEQ ID NO.:128)和pKMD1-PR:ATCTCTAGAGGATCCCCGGG (SEQ ID NO.:129)进行PCR扩增,取扩增产物17 μ L,加入1 μ L Dpn I,2 μ L 10 $\times$ digestion buffer,混匀后37 $^{\circ}$ C温浴3 h,获得质粒骨架线性片段pKMD1-T。

[0220] 6.2.1构建供体质粒pKMD1-K1ADH1-H47KH124K

以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-HR1-PF:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:130)和K1ADH1-H47K-PR:aGCATGaAggTCaGTtTtACAGACACCGGAGTACTTGACG (SEQ ID NO.:131)进行PCR扩增,产物名为K1ADH1-H47KH124K-F1;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-H47K-PF:GTaAaACtGAccTtCaTGCtTGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:132)和K1ADH1-gRNA1-m-PR:CCAaCCtTTaACaTTcTcTCCCATAGCAACAACGACACCAG (SEQ ID NO.:133)进行PCR扩增,产物名为K1ADH1-H47KH124K-F2;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-gRNA1-m-PF:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:134)和K1ADH1-H124K-PR:CCaTCtTtGTaTaTCCactCAAGTCAGCTTCTGGACAGTTGG (SEQ ID NO.:135)进行PCR扩增,产物名为K1ADH1-H47KH124K-F3;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-H124K-PF:CTTGagtGGaTaTACaaAaGAtGGTTCTTTCCAACAATACGCTACTGC (SEQ ID NO.:136)和K1ADH1-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:137)进行PCR扩增,产物名为K1ADH1-H47KH124K-F4。

[0221] 将扩增产物K1ADH1-H47KH124K-F1、K1ADH1-H47KH124K-F2、K1ADH1-H47KH124K-F3、K1ADH1-H47KH124K-F4、pKMD1-T各1 μ L,加入5 μ L Cloning Mix,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-K1ADH1-H47KH124K。

[0222] 6.2.2构建供体质粒pKMD1-K1ADH1-H47NH124K

以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-HR1-PF:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:138)和K1ADH1-H47N-PR:aGCATGaAggTCaGTaTtACAGACACCGGAGTACTTGACG (SEQ ID NO.:139)进行PCR扩增,产物名为K1ADH1-H47NH124K-F1;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-H47N-PF:GTaAtACtGAccTtCaTGCtTGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:140)和K1ADH1-gRNA1-m-PR:CCAaCCtTTaACaTTcTcTCCCATAGCAACAACGACACCAG (SEQ ID NO.:141)进行PCR扩增,产物名为K1ADH1-H47NH124K-F2;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-gRNA1-m-PF:GGGaGAgAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:142)和K1ADH1-H124K-PR:CCaTCtTtGTaTaTCCactCAAGTCAGCTTCTGGACAGTTGG (SEQ ID NO.:143)进行PCR扩增,产物名为K1ADH1-H47NH124K-F3;以乳酸克鲁维酵母基因组DNA为模板,用引物K1ADH1-H124K-PF:CTTGagtGGaTaTACaaAaGAtGGTTCTTTCCAACAATACGCTACTGC (SEQ ID NO.:144)和K1ADH1-HR2-PR:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:145)进

行PCR扩增,产物名为*K1ADH1-H47NH124K-F4*。

[0223] 将扩增产物*K1ADH1-H47NH124K-F1*、*K1ADH1-H47NH124K-F2*、*K1ADH1-H47NH124K-F3*、*K1ADH1-H47NH124K-F4*、pKMD1-T各1 μL,加入5 μL Cloning Mix,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H47NH124K*。

[0224] 6.2.3构建供体质粒pKMD1-*K1ADH1-H47QH124K*

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:146)和*K1ADH1-H47Q-PR*:aGCaTGaAggTCaGTtTGACAGACACCGGAGTACTTGACG (SEQ ID NO.:147)进行PCR扩增,产物名为*K1ADH1-H47QH124K-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H47Q-PF*:GTCAaACtGAccTtCaTGcTtGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:148)和*K1ADH1-gRNA1-m-PR*:CCAaCCtTTaCaTTcTCtCCCATAGCAACAACGACACCAG (SEQ ID NO.:149)进行PCR扩增,产物名为*K1ADH1-H47QH124K-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAGAAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:150)和*K1ADH1-H124K-PR*:CCaTCtTtGTaTaTCCactCAAGTCAGCTTCTGGACAGTTGG (SEQ ID NO.:151)进行PCR扩增,产物名为*K1ADH1-H47QH124K-F3*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H124K-PF*:CTTGagtGGaTAtACaaAaGAtGGTTCTTTCCAACAATACGCTACTGC (SEQ ID NO.:152)和*K1ADH1-HR2-PR*:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:153)进行PCR扩增,产物名为*K1ADH1-H47QH124K-F4*。

[0225] 将扩增产物*K1ADH1-H47QH124K-F1*、*K1ADH1-H47QH124K-F2*、*K1ADH1-H47QH124K-F3*、*K1ADH1-H47QH124K-F4*、pKMD1-T各1 μL,加入5 μL Cloning Mix,混匀后50 °C水浴1 h。水浴结束后置于冰上2 min,将10 μL反应液全部加入50 μL Trans-T1感受态细胞中,冰上放置30 min,42 °C热激30 s后,加入1 mL LB液体培养基37 °C振荡培养1 h,涂布于Amp抗性LB固体培养,37 °C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1-*K1ADH1-H47QH124K*。

[0226] 6.2.4构建供体质粒pKMD1-*K1ADH1-H47RH124K*

以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-HR1-PF*:TCGAGCTCGGTACCCGGGATCCTCTAGAGATCATGGACAATACGTTACCGAGATGAGG (SEQ ID NO.:154)和*K1ADH1-H47R-PR*:aGCaTGaAggTCaGTtctACAGACACCGGAGTACTTGACG (SEQ ID NO.:155)进行PCR扩增,产物名为*K1ADH1-H47RH124K-F1*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-H47R-PF*:GTagaACtGAccTtCaTGcTtGGAAGGGTGACTGGCCTTTG (SEQ ID NO.:156)和*K1ADH1-gRNA1-m-PR*:CCAaCCtTTaCaTTcTCtCCCATAGCAACAACGACACCAG (SEQ ID NO.:157)进行PCR扩增,产物名为*K1ADH1-H47RH124K-F2*;以乳酸克鲁维酵母基因组DNA为模板,用引物*K1ADH1-gRNA1-m-PF*:GGGaGAGAAAtGTtAAaGGtTGGAAGATTGGTGACTTCGCTGG (SEQ ID NO.:158)和*K1ADH1-H124K-PR*:CCaTCtTtGTaTaTCCactCAAGTCAGCTTCTGGACAGTTGG (SEQ ID NO.:159)进行PCR扩增,产物名为*K1ADH1-H47RH124K-F3*;以乳酸克鲁维酵母基因组DNA为模板,用引

物 $KlADH1-H124K-PF$:CTTGagtGGaTAtACaaAaGAtGGTTCTTTCCAACAATACGCTACTGC (SEQ ID NO.: 160) 和 $KlADH1-HR2-PR$:TGCCAAGCTTGCATGCCTGCAGGTCGACGATCTTATACTGGGTAGATTTCAAACGGGAC (SEQ ID NO.:161) 进行PCR扩增,产物名为 $KlADH1-H47RH124K-F4$ 。

[0227] 将扩增产物 $KlADH1-H47RH124K-F1$ 、 $KlADH1-H47RH124K-F2$ 、 $KlADH1-H47RH124K-F3$ 、 $KlADH1-H47RH124K-F4$ 、pKMD1-T各1 μ L,加入5 μ L Cloning Mix,混匀后50 $^{\circ}$ C水浴1 h。水浴结束后置于冰上2 min,将10 μ L反应液全部加入50 μ L Trans-T1感受态细胞中,冰上放置30 min,42 $^{\circ}$ C热激30 s后,加入1 mL LB液体培养基37 $^{\circ}$ C振荡培养1 h,涂布于Amp抗性LB固体培养,37 $^{\circ}$ C倒置培养至单克隆长出。挑取6个单克隆在LB液体培养基中振荡培养,PCR检测阳性并测序确认后,提取质粒保存,命名为pKMD1- $KlADH1-H47RH124K$ 。

[0228] 6.3 $K. lactis$ 电转化(同实施例2.3)

6.4 阳性鉴定

在细胞转化后的平板上挑取12-24个单克隆,以细胞为模板,用鉴定引物 $ADH1-m-CF$:TGGGAGAGAATGTAAAGGT (SEQ ID NO.:162) 和 $ADH1-m-CR$:TGACGGTCGTTAACTAAGAT (SEQ ID NO.:163) 对样品进行PCR检测。PCR结果阳性并经测序鉴定的细胞株,确定为阳性细胞株并分别命名为 $ADH1-H47KH124K$ 、 $ADH1-H47NH124K$ 、 $ADH1-H47QH124K$ 和 $ADH1-H47RH124K$ 。

[0229]

实施例7 $KlADH1$ 改造细胞株的内源蛋白对Ni介质结合能力的分析

7.1 蛋白纯化体系的试剂配制

Binding buffer: 0.5 M NaCl, 20 mM Tris-HCl, 5 mM咪唑,pH 7.9。

[0230] Wash buffer: 0.5 M NaCl, 20 mM Tris-HCl, 60 mM咪唑,pH 7.9。

[0231] 7.2 $KlADH1$ 改造细胞株内源蛋白的纯化和Ni介质结合能力的分析

用binding buffer对Ni-NTA beads进行平衡;加入适量 $KlADH1$ 改造细胞株的细胞裂解液,4 $^{\circ}$ C孵育60 min;用wash buffer洗涤beads;去除wash buffer后加入少许loading buffer于96 $^{\circ}$ C处理5 min。取上述样品进行聚丙烯酰胺凝胶电泳分析。

[0232]

实施例8 体外蛋白质合成体系

8.1 体外蛋白质合成体系的储存液配制

终浓度为22 mM pH为7.4的4-羟乙基哌嗪乙磺酸,30-150 mM醋酸钾,1.0-5.0 mM醋酸镁,1.5-4 mM核苷三磷酸混合物(腺嘌呤核苷三磷酸、鸟嘌呤核苷三磷酸、胞嘧啶核苷三磷酸和尿嘧啶核苷三磷酸),0.08-0.24 mM的氨基酸混合物(甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、苯丙氨酸、脯氨酸、色氨酸、丝氨酸、酪氨酸、半胱氨酸、蛋氨酸、天冬酰胺、谷氨酰胺、苏氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸和组氨酸),25 mM磷酸肌酸,1.7 mM二硫苏糖醇,0.27 mg/mL磷酸肌酸激酶,0.027-0.054 mg/mL T7 RNA 聚合酶,1% - 4% 的聚乙二醇,0.5% - 2% 的蔗糖,最后加入50 %体积的细胞提取物。

[0233] 8.2 体外蛋白质合成反应

将上述的反应体系放置在20-30 $^{\circ}$ C的环境中,静置反应2-12 h;

增强型绿色荧光蛋白(eGFP)活性测定:反应若干小时后,在96孔白板或384孔白板中加入10 μ L反应液,立即放置于Envision 2120多功能酶标仪(Perkin Elmer),读数,检测增

强型绿色荧光蛋白活性,相对荧光单位值(Relative Fluorescence Unit,RFU)作为活性单位,如图24所示。

[0234]

实施例9 体外蛋白质合成纯化

9.1 体外蛋白质合成

鉴于相较于野生型来说, Δ ADH1细胞株的无细胞体系转录翻译活性明显降低,而ADH1 H47K细胞株无明显变化,ADH1 H47K因此成为我们研究体外蛋白合成材料。在ADH1 H47K细胞株的无细胞转录翻译体系中加入N-His8-eGFP和N-His8-Strep-GFP的模板,将上述的反应体系放置在20-30 °C的环境中,静置反应2-12 h。

[0235] 9.2 体外蛋白质纯化

用binding buffer对Ni-NTA beads进行平衡;加入适量上述反应体系,4 °C孵育60 min;用wash buffer洗涤beads;去除wash buffer后加入少许loading buffer于96 °C处理5 min。取上述样品进行聚丙烯酰胺凝胶电泳分析。结果如图25所示。

[0236] 本发明实施例结果表明:

本发明实施例7的结果表明:与野生型细胞株相比,本发明的改造细胞株与Ni介质的亲和力均有明显降低。

[0237] 其中,细胞株如 Δ ADH1,ADH1 H47K,ADH1 H47N,ADH1 H51R,ADH1 H124K,ADH1 H47KH124K和ADH1 H47NH124K与Ni介质的亲和力最低。

[0238] Δ ADH1,ADH1 H47K,ADH1 H47N,ADH1 H51R,ADH1 H124K,ADH1 H47KH124K和ADH1 H47NH124K与Ni介质的亲和力相近,本发明从中选取 Δ ADH1和ADH1 H47K进行无细胞体系转录翻译活性的分析。

[0239] 结果如图24所示,野生型无细胞体系转录翻译活性的相对荧光单位值为442 RFU, Δ ADH1的相对荧光单位值为12 RFU,相对于野生型显著降低,说明ADH1基因敲除对无细胞体系转录翻译活性有较大的负面影响。而ADH1 H47K的相对荧光单位值为419 RFU,相对于野生型来说没有明显变化。

[0240] 采用相同方法,对其他突变细胞株进行无细胞体系转录翻译活性的分析,结果表明,其他突变细胞株的无细胞体系转录翻译活性的相对荧光单位值为300-400 RFU,与野生型的无细胞体系转录翻译活性的相对荧光单位值相当。

[0241] 此外,本发明对外源蛋白N-His8-eGFP和N-His8-Strep-eGFP在 $K1\Delta$ ADH1 H47K的无细胞转录翻译体系中的表达纯度进行了分析,结果如图25所示,相对于野生型细胞株WT, $K1\Delta$ ADH1 H47K细胞株在37 kD处所示的ADH1蛋白条带明显减弱,说明外源蛋白N-His8-eGFP和N-His8-Strep-eGFP的纯度在 $K1\Delta$ ADH1 H47K的无细胞转录翻译体系中得到明显提升。

[0242] 对其他突变细胞株用同样方法进行纯度分析,结果表明,相对于野生型细胞株WT,外源蛋白N-His8-eGFP和N-His8-Strep-eGFP的纯度在其他突变细胞株的无细胞转录翻译体系中也得到明显提升。

[0243] 以上所有结果说明,本发明对ADH家族蛋白进行的定点突变既可以降低其与Ni介质结合的能力,又不影响该菌株的无细胞体系转录翻译活性,更可有效提高外源蛋白在该细胞株无细胞体系转录翻译活性中的表达纯度。

[0244] 在本发明提及的所有文献都在本申请中引用作为参考,就如同每一篇文献被单独

引用作为参考那样。此外应理解,在阅读了本发明的上述讲授内容之后,本领域技术人员可以对本发明作各种改动或修改,这些等价形式同样落于本申请所附权利要求书所限定的范围。

序列表

<110> 康码(上海)生物科技有限公司

<120> ADH蛋白家族突变体及其应用

<130> P2018-0635

<141> 2018-05-31

<160> 164

<170> SIPOSequenceListing 1.0

<210> 1

<211> 350

<212> PRT

<213> Kluyveromyces lactis

<400> 1

```

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

```

Lys	Ala	Lys	Leu	Phe	Lys	Asp	Leu	Gly	Gly	Glu	Tyr	Phe	Ile	Asp	Phe
210						215					220				
Thr	Lys	Ser	Lys	Asn	Ile	Pro	Glu	Glu	Val	Ile	Glu	Ala	Thr	Lys	Gly
225					230					235					240
Gly	Ala	His	Gly	Val	Ile	Asn	Val	Ser	Val	Ser	Glu	Phe	Ala	Ile	Glu
				245					250					255	
Gln	Ser	Thr	Asn	Tyr	Val	Arg	Ser	Asn	Gly	Thr	Val	Val	Leu	Val	Gly
			260					265					270		
Leu	Pro	Arg	Asp	Ala	Lys	Cys	Lys	Ser	Asp	Val	Phe	Asn	Gln	Val	Val
		275					280				285				
Lys	Ser	Ile	Ser	Ile	Val	Gly	Ser	Tyr	Val	Gly	Asn	Arg	Ala	Asp	Thr
290					295					300					
Arg	Glu	Ala	Ile	Asp	Phe	Phe	Ser	Arg	Gly	Leu	Val	Lys	Ala	Pro	Ile
305				310					315						320
His	Val	Val	Gly	Leu	Ser	Glu	Leu	Pro	Ser	Ile	Tyr	Glu	Lys	Met	Glu
			325					330					335		
Lys	Gly	Ala	Ile	Val	Gly	Arg	Tyr	Val	Val	Asp	Thr	Ser	Lys		
		340					345					350			

<210> 2
 <211> 350
 <212> PRT
 <213> 人工序列 (artificial sequence)
 <400> 2

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

Gln	Gln	Tyr	Ala	Thr	Ala	Asp	Ala	Val	Gln	Ala	Ala	Lys	Ile	Pro	Val
130						135					140				
Gly	Thr	Asp	Leu	Ala	Glu	Val	Ala	Pro	Val	Leu	Cys	Ala	Gly	Val	Thr
145					150					155					160
Val	Tyr	Lys	Ala	Leu	Lys	Ser	Ala	Asn	Leu	Lys	Ala	Gly	Asp	Trp	Val
				165					170					175	
Ala	Ile	Ser	Gly	Ala	Ala	Gly	Gly	Leu	Gly	Ser	Leu	Ala	Val	Gln	Tyr
			180					185						190	
Ala	Lys	Ala	Met	Gly	Tyr	Arg	Val	Leu	Gly	Ile	Asp	Ala	Gly	Glu	Glu
			195					200					205		
Lys	Ala	Lys	Leu	Phe	Lys	Asp	Leu	Gly	Gly	Glu	Tyr	Phe	Ile	Asp	Phe
			210				215						220		
Thr	Lys	Ser	Lys	Asn	Ile	Pro	Glu	Glu	Val	Ile	Glu	Ala	Thr	Lys	Gly
225					230					235					240
Gly	Ala	His	Gly	Val	Ile	Asn	Val	Ser	Val	Ser	Glu	Phe	Ala	Ile	Glu
				245						250					255
Gln	Ser	Thr	Asn	Tyr	Val	Arg	Ser	Asn	Gly	Thr	Val	Val	Leu	Val	Gly
				260					265					270	
Leu	Pro	Arg	Asp	Ala	Lys	Cys	Lys	Ser	Asp	Val	Phe	Asn	Gln	Val	Val
			275						280					285	
Lys	Ser	Ile	Ser	Ile	Val	Gly	Ser	Tyr	Val	Gly	Asn	Arg	Ala	Asp	Thr
			290				295						300		
Arg	Glu	Ala	Ile	Asp	Phe	Phe	Ser	Arg	Gly	Leu	Val	Lys	Ala	Pro	Ile
305					310					315					320
His	Val	Val	Gly	Leu	Ser	Glu	Leu	Pro	Ser	Ile	Tyr	Glu	Lys	Met	Glu
				325						330					335
Lys	Gly	Ala	Ile	Val	Gly	Arg	Tyr	Val	Val	Asp	Thr	Ser	Lys		
			340						345					350	

<210> 3

<211> 350

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 3

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

Asp	Leu	His	Ala	Trp	Lys	Gly	Asp	Trp	Pro	Leu	Pro	Thr	Lys	Leu	Pro
50						55					60				
Leu	Val	Gly	Gly	His	Glu	Gly	Ala	Gly	Val	Val	Val	Ala	Met	Gly	Glu
65					70				75					80	
Asn	Val	Lys	Gly	Trp	Lys	Ile	Gly	Asp	Phe	Ala	Gly	Ile	Lys	Trp	Leu
				85				90						95	
Asn	Gly	Ser	Cys	Met	Ser	Cys	Glu	Tyr	Cys	Glu	Leu	Ser	Asn	Glu	Ser
			100					105					110		
Asn	Cys	Pro	Glu	Ala	Asp	Leu	Ser	Gly	Tyr	Thr	His	Asp	Gly	Ser	Phe
		115					120					125			
Gln	Gln	Tyr	Ala	Thr	Ala	Asp	Ala	Val	Gln	Ala	Ala	Lys	Ile	Pro	Val
	130					135					140				
Gly	Thr	Asp	Leu	Ala	Glu	Val	Ala	Pro	Val	Leu	Cys	Ala	Gly	Val	Thr
145					150					155					160
Val	Tyr	Lys	Ala	Leu	Lys	Ser	Ala	Asn	Leu	Lys	Ala	Gly	Asp	Trp	Val
				165					170					175	
Ala	Ile	Ser	Gly	Ala	Ala	Gly	Gly	Leu	Gly	Ser	Leu	Ala	Val	Gln	Tyr
			180					185						190	
Ala	Lys	Ala	Met	Gly	Tyr	Arg	Val	Leu	Gly	Ile	Asp	Ala	Gly	Glu	Glu
		195						200					205		
Lys	Ala	Lys	Leu	Phe	Lys	Asp	Leu	Gly	Gly	Glu	Tyr	Phe	Ile	Asp	Phe
	210					215						220			
Thr	Lys	Ser	Lys	Asn	Ile	Pro	Glu	Glu	Val	Ile	Glu	Ala	Thr	Lys	Gly
225					230					235					240
Gly	Ala	His	Gly	Val	Ile	Asn	Val	Ser	Val	Ser	Glu	Phe	Ala	Ile	Glu
				245						250				255	
Gln	Ser	Thr	Asn	Tyr	Val	Arg	Ser	Asn	Gly	Thr	Val	Val	Leu	Val	Gly
			260						265				270		
Leu	Pro	Arg	Asp	Ala	Lys	Cys	Lys	Ser	Asp	Val	Phe	Asn	Gln	Val	Val
		275						280					285		
Lys	Ser	Ile	Ser	Ile	Val	Gly	Ser	Tyr	Val	Gly	Asn	Arg	Ala	Asp	Thr
	290					295					300				
Arg	Glu	Ala	Ile	Asp	Phe	Phe	Ser	Arg	Gly	Leu	Val	Lys	Ala	Pro	Ile
305					310					315					320
His	Val	Val	Gly	Leu	Ser	Glu	Leu	Pro	Ser	Ile	Tyr	Glu	Lys	Met	Glu
				325						330				335	
Lys	Gly	Ala	Ile	Val	Gly	Arg	Tyr	Val	Val	Asp	Thr	Ser	Lys		
			340					345					350		

<210> 4

<211> 350

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 4

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

275	280	285
Lys Ser Ile Ser Ile Val Gly	Ser Tyr Val Gly	Asn Arg Ala Asp Thr
290	295	300
Arg Glu Ala Ile Asp Phe Phe	Ser Arg Gly Leu Val Lys	Ala Pro Ile
305	310	315
His Val Val Gly Leu Ser Glu	Leu Pro Ser Ile Tyr Glu	Lys Met Glu
325	330	335
Lys Gly Ala Ile Val Gly Arg	Tyr Val Val Asp Thr	Ser Lys
340	345	350
<210> 5		
<211> 350		
<212> PRT		
<213> 人工序列 (artificial sequence)		
<400> 5		
Met Ala Ala Ser Ile Pro Glu Thr	Gln Lys Gly Val Ile Phe Tyr Glu	
1	5	10
Asn Gly Gly Glu Leu Gln Tyr Lys	Asp Ile Pro Val Pro Lys Pro Lys	
20	25	30
Ala Asn Glu Leu Leu Ile Asn Val	Lys Tyr Ser Gly Val Cys Arg Thr	
35	40	45
Asp Leu His Ala Trp Lys Gly Asp	Trp Pro Leu Pro Thr Lys Leu Pro	
50	55	60
Leu Val Gly Gly His Glu Gly Ala	Gly Val Val Val Ala Met Gly Glu	
65	70	75
Asn Val Lys Gly Trp Lys Ile Gly	Asp Phe Ala Gly Ile Lys Trp Leu	
85	90	95
Asn Gly Ser Cys Met Ser Cys Glu	Tyr Cys Glu Leu Ser Asn Glu Ser	
100	105	110
Asn Cys Pro Glu Ala Asp Leu Ser	Gly Tyr Thr His Asp Gly Ser Phe	
115	120	125
Gln Gln Tyr Ala Thr Ala Asp Ala	Val Gln Ala Ala Lys Ile Pro Val	
130	135	140
Gly Thr Asp Leu Ala Glu Val Ala	Pro Val Leu Cys Ala Gly Val Thr	
145	150	155
Val Tyr Lys Ala Leu Lys Ser Ala	Asn Leu Lys Ala Gly Asp Trp Val	
165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly	Leu Gly Ser Leu Ala Val Gln Tyr	
180	185	190
Ala Lys Ala Met Gly Tyr Arg Val	Leu Gly Ile Asp Ala Gly Glu Glu	

195					200					205						
Lys	Ala	Lys	Leu	Phe	Lys	Asp	Leu	Gly	Gly	Glu	Tyr	Phe	Ile	Asp	Phe	
210					215					220						
Thr	Lys	Ser	Lys	Asn	Ile	Pro	Glu	Glu	Val	Ile	Glu	Ala	Thr	Lys	Gly	
225					230					235					240	
Gly	Ala	His	Gly	Val	Ile	Asn	Val	Ser	Val	Ser	Glu	Phe	Ala	Ile	Glu	
245					250					255						
Gln	Ser	Thr	Asn	Tyr	Val	Arg	Ser	Asn	Gly	Thr	Val	Val	Leu	Val	Gly	
260					265					270						
Leu	Pro	Arg	Asp	Ala	Lys	Cys	Lys	Ser	Asp	Val	Phe	Asn	Gln	Val	Val	
275					280					285						
Lys	Ser	Ile	Ser	Ile	Val	Gly	Ser	Tyr	Val	Gly	Asn	Arg	Ala	Asp	Thr	
290					295					300						
Arg	Glu	Ala	Ile	Asp	Phe	Phe	Ser	Arg	Gly	Leu	Val	Lys	Ala	Pro	Ile	
305					310					315					320	
His	Val	Val	Gly	Leu	Ser	Glu	Leu	Pro	Ser	Ile	Tyr	Glu	Lys	Met	Glu	
325					330					335						
Lys	Gly	Ala	Ile	Val	Gly	Arg	Tyr	Val	Val	Asp	Thr	Ser	Lys			
340					345					350						

 $\langle 210 \rangle$ 6

350 $\langle 211 \rangle$

<212> PRT

〈213〉 人工序列 (artificial sequence)

 $\langle 400 \rangle$ 6

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

115	120	125
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val		
130	135	140
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr		
145	150	155
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val		
165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr		
180	185	190
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225	230	235
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
245	250	255
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		
260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
340	345	350

<210> 7

<211> 350

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 7

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

35	40	45
Asp Leu Arg Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro		
50	55	60
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu		
65	70	75
Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu		
85	90	95
Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser		
100	105	110
Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe		
115	120	125
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val		
130	135	140
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr		
145	150	155
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val		
165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr		
180	185	190
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225	230	235
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
245	250	255
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		
260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
340	345	350

<210> 8

<211> 350

<212> PRT

<213> 人工序列 (artificial sequence)

<400> 8

```

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

```

Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val
 275 280 285
 Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr
 290 295 300
 Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile
 305 310 315 320
 His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu
 325 330 335
 Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys
 340 345 350
 <210> 9
 <211> 350
 <212> PRT
 <213> 人工序列 (artificial sequence)
 <400> 9
 Met Ala Ala Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu
 1 5 10 15
 Asn Gly Gly Glu Leu Gln Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys
 20 25 30
 Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys His Thr
 35 40 45
 Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro
 50 55 60
 Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu
 65 70 75 80
 Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu
 85 90 95
 Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser
 100 105 110
 Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr Arg Asp Gly Ser Phe
 115 120 125
 Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val
 130 135 140
 Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr
 145 150 155 160
 Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val
 165 170 175
 Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr
 180 185 190

Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225	230	235
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
	245	250
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		
	260	265
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
	275	280
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
	325	330
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
	340	345
		350
<210> 10		
<211> 350		
<212> PRT		
<213> 人工序列 (artificial sequence)		
<400> 10		
Met Ala Ala Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu		
1	5	10
Asn Gly Gly Glu Leu Gln Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys		
	20	25
Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys Lys Thr		
	35	40
Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro		
	50	55
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu		
65	70	75
Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu		
	85	90
Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser		
	100	105
		110

Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr Lys Asp Gly Ser Phe			
115	120	125	
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val			
130	135	140	
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr			
145	150	155	160
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val			
	165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr			
	180	185	190
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu			
	195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe			
	210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly			
225	230	235	240
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu			
	245	250	255
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly			
	260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val			
	275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr			
	290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile			
305	310	315	320
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu			
	325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys			
	340	345	350
<210> 11			
<211> 350			
<212> PRT			
<213> 人工序列 (artificial sequence)			
<400> 11			
Met Ala Ala Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu			
1	5	10	15
Asn Gly Gly Glu Leu Gln Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys			
	20	25	30

Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys Asn Thr		
35	40	45
Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro		
50	55	60
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu		
65	70	75
Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu		
85	90	95
Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser		
100	105	110
Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr Lys Asp Gly Ser Phe		
115	120	125
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val		
130	135	140
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr		
145	150	155
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val		
165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr		
180	185	190
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225	230	235
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
245	250	255
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		
260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		

340	345	350
<210> 12		
<211> 350		
<212> PRT		
<213> 人工序列 (artificial sequence)		
<400> 12		
Met Ala Ala Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu		
1 5 10 15		
Asn Gly Gly Glu Leu Gln Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys		
20 25 30		
Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys Gln Thr		
35 40 45		
Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro		
50 55 60		
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu		
65 70 75 80		
Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu		
85 90 95		
Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser		
100 105 110		
Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr Lys Asp Gly Ser Phe		
115 120 125		
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val		
130 135 140		
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr		
145 150 155 160		
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val		
165 170 175		
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr		
180 185 190		
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195 200 205		
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210 215 220		
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225 230 235 240		
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
245 250 255		
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		

260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
340	345	350
<210> 13		
<211> 350		
<212> PRT		
<213> 人工序列 (artificial sequence)		
<400> 13		
Met Ala Ala Ser Ile Pro Glu Thr Gln Lys Gly Val Ile Phe Tyr Glu		
1	5	10
Asn Gly Gly Glu Leu Gln Tyr Lys Asp Ile Pro Val Pro Lys Pro Lys		
20	25	30
Ala Asn Glu Leu Leu Ile Asn Val Lys Tyr Ser Gly Val Cys Arg Thr		
35	40	45
Asp Leu His Ala Trp Lys Gly Asp Trp Pro Leu Pro Thr Lys Leu Pro		
50	55	60
Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met Gly Glu		
65	70	75
Asn Val Lys Gly Trp Lys Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu		
85	90	95
Asn Gly Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser		
100	105	110
Asn Cys Pro Glu Ala Asp Leu Ser Gly Tyr Thr Lys Asp Gly Ser Phe		
115	120	125
Gln Gln Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Lys Ile Pro Val		
130	135	140
Gly Thr Asp Leu Ala Glu Val Ala Pro Val Leu Cys Ala Gly Val Thr		
145	150	155
Val Tyr Lys Ala Leu Lys Ser Ala Asn Leu Lys Ala Gly Asp Trp Val		
165	170	175
Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser Leu Ala Val Gln Tyr		

180	185	190
Ala Lys Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Ala Gly Glu Glu		
195	200	205
Lys Ala Lys Leu Phe Lys Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe		
210	215	220
Thr Lys Ser Lys Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys Gly		
225	230	235
Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu		
245	250	255
Gln Ser Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly		
260	265	270
Leu Pro Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Asn Gln Val Val		
275	280	285
Lys Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr		
290	295	300
Arg Glu Ala Ile Asp Phe Phe Ser Arg Gly Leu Val Lys Ala Pro Ile		
305	310	315
His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu Lys Met Glu		
325	330	335
Lys Gly Ala Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
340	345	350

<210> 14

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 14

```

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtggtgaa 60
ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtaa aactgacctt catgcttgga agggtgactg gcctttgcc 180
accaaattgc cattagttgg tggtcacgaa ggtgctgggtg tcgttggtgc tatgggagag 240
aatgttaaag gttggaagat tggtgacttc gctgggtatca aatggttgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
ggttacactc acgacggttc tttccaacaa tacgtactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttgggtgaa gttgtccag tgctatgtgc tgggtgtcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
gctgctgggtg gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
ttgggtatcg atgctggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780

```

tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcggtg gtttgtccga actaccatcc atctacgaaa agatggaaaa gggtagctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 15

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 15

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtaggtgaa 60
 ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtaa tactgacctt catgcttgga agggtagctg gcctttgcca 180
 accaaattgc cattagttag tggtcacgaa ggtgctgggt tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tggtagcttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
 gggtacactc acgacgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
 aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
 gctgctgggt gtctaggctt tctagctgtc caatacgcca aggccatggg ttacagagtg 600
 ttgggtatcg atgctgggtg agaaaaggct aagttgttca aggacttggg tggtagaatac 660
 ttcatgtgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcggtg gtttgtccga actaccatcc atctacgaaa agatggaaaa gggtagctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 16

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 16

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtaggtgaa 60
 ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtca aactgacctt catgcttgga agggtagctg gcctttgcca 180
 accaaattgc cattagttag tggtcacgaa ggtgctgggt tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tggtagcttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360

ggttacactc acgacgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
 aagatccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccgggtg actgggtcgc catctctggt 540
 gctgctgggtg gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagt 600
 ttgggtatcg atgctgggtga agaaaaggct aagttgttca aggacttggg tgggtgaatac 660
 ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacgggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 17

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 17

atggccgcat ctatcccaga aactcaaaaag ggtgttatct tctacgaaaa cggtggtgaa 60
 ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtag aactgacctt catgcttggg aggggtgactg gcctttgcc 180
 accaaattgc cattagttgg tggtcacgaa ggtgctgggtg tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tgggtgacttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
 ggttacactc acgacgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
 aagatccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccgggtg actgggtcgc catctctggt 540
 gctgctgggtg gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagt 600
 ttgggtatcg atgctgggtga agaaaaggct aagttgttca aggacttggg tgggtgaatac 660
 ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacgggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 18

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 18

```

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtggtgaa 60
ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtca tactgacctt aaagcttgga agggtgactg gcctttgcc 180
accaaattgc cattagttgg tggtcacgaa ggtgctggtg tcgttggtgc tatgggagag 240
aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
ggttacactc acgacgggtc tttccaacaa tacgtactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
gctgctggtg gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
ttgggtatcg atgctggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
gtcggtagat acgtcgtcga cacttcaaaa taa 1053

```

<210> 19

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 19

```

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtggtgaa 60
ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtca tactgacctt agagcttgga agggtgactg gcctttgcc 180
accaaattgc cattagttgg tggtcacgaa ggtgctggtg tcgttggtgc tatgggagag 240
aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
ggttacactc acgacgggtc tttccaacaa tacgtactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
gctgctggtg gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
ttgggtatcg atgctggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900

```

agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa gggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 20

<211> 1053

<212> DNA

<213> 人工序列(artificial sequence)

<400> 20

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtgggtgaa 60
 ttgcaatata aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtca caccgatttg cacgcatgga agggtgactg gcctttgcc 180
 accaaattgc cattagttagg tggtcacgaa ggtgctgggt tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
 ggatatacaa aagatgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
 aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
 gctgctgggt gtctagggtc tctagctgtc caatacgcca aggccatggg ttacagagt 600
 ttgggtatcg atgctgggtg agaaaaggct aagttgttca aggacttggg tggatgaatac 660
 ttcatgtgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa gggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 21

<211> 1053

<212> DNA

<213> 人工序列(artificial sequence)

<400> 21

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtgggtgaa 60
 ttgcaatata aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtca caccgatttg cacgcatgga agggtgactg gcctttgcc 180
 accaaattgc cattagttagg tggtcacgaa ggtgctgggt tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
 ggatatacaa gagatgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
 aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480

gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
 gctgctggtg gtctagggtc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
 ttgggtatcg atgctggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
 ttcatgtgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgag 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 22

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 22

atggccgcat ctatcccaga aactcaaaaag ggtgttatct tctacgaaaa cggtggtgaa 60
 ttgcaataca aggacatccc agttccaaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtaa aactgacctt catgcttggg aggggtgactg gcctttgcca 180
 accaaattgc cattagtgtg tggtcacgaa ggtgctggtg tcgttgttgc tatgggagag 240
 aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
 ggatatacaa aagatgggtc tttccaacaa tacgtactg ctgatgccgt tcaagctgcc 420
 aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccggtg actgggtcgc catctctggt 540
 gctgctggtg gtctagggtc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
 ttgggtatcg atgctggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
 ttcatgtgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgag 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 23

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 23

atggccgcat ctatcccaga aactcaaaaag ggtgttatct tctacgaaaa cggtggtgaa 60

```

ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtaa tactgacctt catgcttgga agggtgactg gcctttgcc 180
accaaattgc cattagtgg tggtcacgaa ggtgctgggt tcgttggtgc tatgggagag 240
aatgttaaag gttggaagat tggtagcttc gctgggtatca aatgggtgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
ggatatacaa aagatgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttgggtgaa gttgctccag tgctatgtgc tgggtgcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccgggt actgggtcgc catctctggt 540
gctgctgggt gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagt 600
ttgggtatcg atgctgggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgag 840
tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
gtcggtagat acgtcgtcga cacttcaaaa taa 1053

```

<210> 24

<211> 1053

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 24

```

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cggtgggtgaa 60
ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtca aactgacctt catgcttgga agggtgactg gcctttgcc 180
accaaattgc cattagtgg tggtcacgaa ggtgctgggt tcgttggtgc tatgggagag 240
aatgttaaag gttggaagat tggtagcttc gctgggtatca aatgggtgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
ggatatacaa aagatgggtc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttgggtgaa gttgctccag tgctatgtgc tgggtgcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccgggt actgggtcgc catctctggt 540
gctgctgggt gtctaggttc tctagctgtc caatacgcca aggccatggg ttacagagt 600
ttgggtatcg atgctgggtga agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgag 840
tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020

```

gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 25

<211> 1053

<212> DNA

<213> 人工序列(artificial sequence)

<400> 25

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaaa cgggtggtgaa 60
 ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
 aagtactccg gtgtctgtag aactgacctt catgcttgga agggtgactg gcctttgcca 180
 accaaattgc cattagttgg tggtcacgaa ggtgctgggt tcgttggtgc tatgggagag 240
 aatgttaaag gttggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
 atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgagt 360
 ggatatacaa aagatgggtc tttccaacaa tacgtactg ctgatgccgt tcaagctgcc 420
 aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgtcacc 480
 gtttacaagg ccctaaaatc cgccaacttg aaggccgggt actgggtcgc catctctggt 540
 gctgctgggt gtctagggtc tctagctgtc caatacgcca aggccatggg ttacagagt 600
 ttgggtatcg atgctgggtg agaaaaggct aagttgttca aggacttggg tgggtgaatac 660
 ttcatgtatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
 ggtgtcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
 tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
 tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
 agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
 cacgtcgttg gtttgtccga actaccatcc atctacgaaa agatggaaaa gggtgctatc 1020
 gtcggtagat acgtcgtcga cacttcaaaa taa 1053

<210> 26

<211> 348

<212> PRT

<213> *Kluyveromyces lactis*

<400> 26

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

Lys Gly Trp Asn Ile Gly Asp Phe Ala Gly Ile Lys Trp Leu Asn Gly		
85	90	95
Ser Cys Met Ser Cys Glu Tyr Cys Glu Leu Ser Asn Glu Ser Asn Cys		
100	105	110
Pro Asp Ala Asp Leu Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln		
115	120	125
Tyr Ala Thr Ala Asp Ala Val Gln Ala Ala Arg Ile Pro Lys Gly Thr		
130	135	140
Asp Leu Ala Glu Val Ala Pro Ile Leu Cys Ala Gly Val Thr Val Tyr		
145	150	155
Lys Ala Leu Lys Ser Ala Asp Leu Lys Ala Gly Asp Trp Val Ala Ile		
165	170	175
Ser Gly Ala Cys Gly Gly Leu Gly Ser Leu Ala Ile Gln Tyr Ala Lys		
180	185	190
Ala Met Gly Tyr Arg Val Leu Gly Ile Asp Thr Gly Ala Glu Lys Ala		
195	200	205
Lys Leu Phe Lys Glu Leu Gly Gly Glu Tyr Phe Val Asp Tyr Ala Val		
210	215	220
Ser Lys Asp Leu Ile Lys Glu Ile Val Asp Ala Thr Asn Gly Gly Ala		
225	230	235
His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile Glu Gln Ser		
245	250	255
Thr Asn Tyr Val Arg Ser Asn Gly Thr Val Val Leu Val Gly Leu Pro		
260	265	270
Arg Asp Ala Lys Cys Lys Ser Asp Val Phe Thr Gln Val Val Lys Ser		
275	280	285
Val Ser Ile Val Gly Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu		
290	295	300
Ala Leu Asp Phe Phe Ala Arg Gly Leu Val His Ala Pro Ile Lys Ile		
305	310	315
Val Gly Leu Ser Glu Leu Ala Asp Val Tyr Asp Lys Met Val Lys Gly		
325	330	335
Glu Ile Val Gly Arg Tyr Val Val Asp Thr Ser Lys		
340	345	

<210> 27

<211> 374

<212> PRT

<213> Kluyveromyces lactis

<400> 27

Met	Leu	Arg	Leu	Thr	Ser	Ala	Arg	Ser	Ile	Val	Ser	Pro	Leu	Arg	Lys
1				5					10					15	
Gly	Ala	Phe	Gly	Ser	Ile	Arg	Thr	Leu	Ala	Thr	Ser	Val	Pro	Glu	Thr
			20					25					30		
Gln	Lys	Gly	Val	Ile	Phe	Tyr	Glu	Asn	Gly	Gly	Lys	Leu	Glu	Tyr	Lys
		35					40					45			
Asp	Ile	Pro	Val	Pro	Lys	Pro	Lys	Pro	Asn	Glu	Ile	Leu	Ile	Asn	Val
	50					55					60				
Lys	Tyr	Ser	Gly	Val	Cys	His	Thr	Asp	Leu	His	Ala	Trp	Lys	Gly	Asp
65					70				75					80	
Trp	Pro	Leu	Pro	Thr	Lys	Leu	Pro	Leu	Val	Gly	Gly	His	Glu	Gly	Ala
				85					90					95	
Gly	Val	Val	Val	Ala	Met	Gly	Glu	Asn	Val	Lys	Gly	Trp	Asn	Ile	Gly
			100					105					110		
Asp	Phe	Ala	Gly	Ile	Lys	Trp	Leu	Asn	Gly	Ser	Cys	Met	Ser	Cys	Glu
		115					120					125			
Tyr	Cys	Glu	Leu	Ser	Asn	Glu	Ser	Asn	Cys	Pro	Asp	Ala	Asp	Leu	Ser
	130					135					140				
Gly	Tyr	Thr	His	Asp	Gly	Ser	Phe	Gln	Gln	Tyr	Ala	Thr	Ala	Asp	Ala
145					150				155					160	
Val	Gln	Ala	Ala	Arg	Ile	Pro	Lys	Gly	Thr	Asp	Leu	Ala	Glu	Val	Ala
				165					170					175	
Pro	Ile	Leu	Cys	Ala	Gly	Val	Thr	Val	Tyr	Lys	Ala	Leu	Lys	Ser	Ala
		180						185				190			
Asn	Leu	Lys	Ala	Gly	Asp	Trp	Val	Ala	Ile	Ser	Gly	Ala	Ala	Gly	Gly
	195					200					205				
Leu	Gly	Ser	Leu	Ala	Val	Gln	Tyr	Ala	Lys	Ala	Met	Gly	Tyr	Arg	Val
	210					215					220				
Val	Gly	Ile	Asp	Gly	Gly	Glu	Glu	Lys	Gly	Lys	Leu	Val	Lys	Gln	Leu
225				230					235					240	
Gly	Gly	Glu	Ala	Phe	Val	Asp	Phe	Thr	Lys	Thr	Lys	Asp	Met	Val	Ala
				245					250				255		
Glu	Ile	Gln	Glu	Ile	Thr	Asn	Gly	Gly	Pro	His	Gly	Val	Ile	Asn	Val
			260				265					270			
Ser	Val	Ser	Glu	Ala	Ala	Met	Asn	Ala	Ser	Thr	Gln	Phe	Val	Arg	Pro
	275						280					285			
Thr	Gly	Thr	Val	Val	Leu	Val	Gly	Leu	Pro	Ala	Gly	Ala	Val	Ile	Lys
	290					295					300				
Ser	Glu	Val	Phe	Ser	His	Val	Val	Lys	Ser	Ile	Asn	Ile	Lys	Gly	Ser

305	310	315	320
Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu Ala Ile Asn Phe Phe Ala			
	325	330	335
Asn Gly His Val His Ser Pro Ile Lys Val Val Gly Leu Ser Glu Leu			
	340	345	350
Pro Lys Val Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly Arg Tyr			
	355	360	365
Val Val Asp Thr Ser Asn			
370			
<210> 28			
<211> 375			
<212> PRT			
<213> Kluyveromyces lactis			
<400> 28			
Met Phe Arg Leu Ala Arg Ala Gln Thr Ala Leu Ala Asn Lys Ala Ser			
1	5	10	15
Val Ser Arg Ser Phe Leu Arg Leu Asn Ser Ser Phe Ala Ile Pro Glu			
	20	25	30
Thr Gln Lys Gly Val Ile Phe Tyr Glu Asn Gly Gly Lys Leu Glu Tyr			
	35	40	45
Lys Asp Leu Pro Val Pro Lys Pro Lys Ala Asn Glu Ile Leu Ile Asn			
	50	55	60
Val Lys Tyr Ser Gly Val Cys His Thr Asp Leu His Ala Trp Lys Gly			
65	70	75	80
Asp Trp Pro Leu Pro Val Lys Leu Pro Leu Val Gly Gly His Glu Gly			
	85	90	95
Ala Gly Ile Val Val Ala Lys Gly Glu Asn Val Lys Asn Phe Glu Ile			
	100	105	110
Gly Asp Tyr Ala Gly Ile Lys Trp Leu Asn Gly Ser Cys Met Ser Cys			
	115	120	125
Glu Leu Cys Glu Gln Gly Tyr Glu Ser Asn Cys Leu Gln Ala Asp Leu			
	130	135	140
Ser Gly Tyr Thr His Asp Gly Ser Phe Gln Gln Tyr Ala Thr Ala Asp			
145	150	155	160
Ala Val Gln Ala Ala Gln Ile Pro Lys Gly Thr Asp Leu Ala Glu Ile			
	165	170	175
Ala Pro Ile Leu Cys Ala Gly Val Thr Val Tyr Lys Ala Leu Lys Thr			
	180	185	190
Ala Asp Leu Lys Pro Gly Gln Trp Val Ala Ile Ser Gly Ala Ala Gly			

195	200	205
Gly Leu Gly Ser Leu Ala Val Gln Tyr Ala Lys Ala Met Gly Leu Arg		
210	215	220
Val Leu Gly Ile Asp Gly Gly Asp Gly Lys Glu Glu Leu Phe Lys Gln		
225	230	235
Cys Gly Gly Glu Val Phe Ile Asp Phe Arg Lys Ser Lys Asp Met Val		
245	250	255
Ala Asp Ile Gln Glu Ala Thr Asn Gly Gly Pro His Gly Val Ile Asn		
260	265	270
Val Ser Val Ser Glu Ala Ala Ile Ser Met Ser Thr Glu Tyr Val Arg		
275	280	285
Pro Thr Gly Val Val Val Leu Val Gly Leu Pro Ala Asp Ala Tyr Val		
290	295	300
Lys Ser Glu Val Phe Ser His Val Val Lys Ser Ile Ser Ile Lys Gly		
305	310	315
Ser Tyr Val Gly Asn Arg Ala Asp Thr Arg Glu Ala Thr Asp Phe Phe		
325	330	335
Thr Arg Gly Leu Val Lys Ser Pro Ile Lys Ile Ile Gly Leu Ser Glu		
340	345	350
Leu Pro Glu Ala Tyr Glu Leu Met Glu Gln Gly Lys Ile Leu Gly Arg		
355	360	365
Phe Val Val Asp Thr Tyr Lys		
370	375	

<210> 29

<211> 1047

<212> DNA

<213> Kluyveromyces lactis

<400> 29

```

atgtccattc ctgaaactca aaagggtggt atctttttacg aaaacggtgg tgaattgcaa 60
tacaaggaca ttccagttcc aaagccaaag gccaacgaac ttttgatcaa cgtcaagtac 120
tccggtgtct gtcacaccga tttgcacgca tggaagggtg actggccttt gccaaccaaa 180
ttgccattag ttggtggtca cgaagggtgct ggtgtcgttg ttgctatggg tgaaaatgtc 240
aagggtgga acattggtga ctttgctggt atcaaagtgt tgaacggttc ttgtatgtcc 300
tgtgaatact gtgaattgtc caatgaatcc aactgtcccc atgctgactt gtctggttac 360
accacgatg gttctttcca acaatatgct accgctgatg ccgttcaagc tgctagaatt 420
ccaaagggtg ccgatttggc tgaagttgct ccaattctat gtgccggtgt taccgtttac 480
aaggtcttaa aatctgctga cttgaaggcc ggtgactggg ttgccatttc cgggtgcctgt 540
ggtggtctag gttctttggc tatccaatac gccaaaggcta tgggttacag agtcttgggt 600
attgataccg gtgctgaaaa ggctaagttg ttcaaggagc taggtggtga atacttcgtc 660

```

gattatgccg tctctaagga ttgattaaa gaaattgttg acgtactaa cggtggtgcc 720
 cacggtgtca ttaacgtctc tgtctccgaa ttgctatag aacaatccac caattacgtt 780
 agatcaaacg gtaccgttgt attggtcggc ctaccaaggc acgccaatg taagtctgat 840
 gtctttacac aagttgtcaa atcggctctc attgttggtt cttacgtcgg taacagagct 900
 gacaccagag aagctctaga tttcttcgca agaggtttag tgcattgctc aattaagata 960
 gtcggattat ctgaattggc agatgtttat gacaagatgg tcaagggtga aatcgttggt 1020
 agatacgttg tcgacacctc aaaataa 1047

<210> 30

<211> 1125

<212> DNA

<213> *Kluyveromyces lactis*

<400> 30

atgttgagat tgacttccgc cagatccatt gtttcccat tgcgtaaggc tgcttttggc 60
 tccatcagaa ccttagctac ctctgtgcca gaaacccaaa aggtgttat tttctatgag 120
 aatggtggtg aattggaata caaggacatt ccagttccaa agccaaagcc aaatgaaatc 180
 ttgatcaacg tcaagtactc cgggtgtgtg cataccgatt tgcacgcatg gaagggtgac 240
 tggccattgc caaccaagtt gccattggc ggtggtcacg aaggtgctgg tgcgttggc 300
 gctatgggtg aaaacgtcaa gggctggaac attggtgact ttgcgggtat caaatggttg 360
 aacggttctt gtatgtcctg tgaatactgt gaattgtcca atgaatccaa ctgtccagat 420
 gctgacttgt ctggttacac ccacgatggc tctttccaa aatacgtac cgcagatgct 480
 gttcaagctg ccagaattcc aaagggtacc gatttggtg aagttgctcc aattctatgt 540
 gccggtgtta ctgtttacaa ggctttgaaa agtgctaact tgaaggctgg tgactgggtt 600
 gccatctctg gtgctgctgg tggcttaggt tctctagctg tccaatacgc caaggccatg 660
 gggttacagag tcgttggtat cgacggtggc gaagaaaagg gtaagttggc caagcaattg 720
 ggtggtgaag cctttgttga tttcaccaa accaaggaca tgggtgctga aatccaagaa 780
 atcaccaacg gtggtccaca cgggtgcatt aacgtctctg tttctgaagc tgccatgaac 840
 gcttccactc aattcgtcag accaactggc actgtcgtat tggtcggtt gccagctggt 900
 gccgtcatca agtccgaagt cttctccac gtcgttaagt ctattaacat caagggttct 960
 tacgtcggta acagagctga caccagagaa gctatcaact tcttcgctaa cggtcacgtc 1020
 cactctccaa tcaaggttgt tggtttgtcc gaactacaa aggtttacga attgatggaa 1080
 caaggtaaga ttttgggtag atacgttggt gacacctcca actag 1125

<210> 31

<211> 1128

<212> DNA

<213> *Kluyveromyces lactis*

<400> 31

atgttcagat tagcccggtc ccaaactgct ctacgaaca aagcttctgt ttccagaagc 60
 tttttgagat taaactcttc tttcgtatc ccagaaactc aaaagggtgt tatcttctac 120
 gaaaatgggtg gtaagttgga atacaaggat ttgccagttc caagccaaa ggctaacgaa 180

attttgatta acgftaagta ctccggtgtt tgtcacaccg atttgcacgc ctggaagggt 240
 gactggccat tgccagttaa attgccatta gtcggtggtc acgaagggtgc tggatatcgtt 300
 gttgccaagg gtgaaaacgt taagaacttc gaaattgggtg attacgctgg tatcaagtgg 360
 ttgaacgggtt cttgtatgtc ttgtgaattg tgtgaacaag gttacgaatc taactgtttg 420
 caagctgact tgtctgggta caccatgac ggttctttcc aacaatacgc taccgctgat 480
 gccgttcaag ctgctcaaatt tccaaagggt accgatttgg ctgaaatcgc cccaatcttg 540
 tgtgccgggtg ttaccgttta caaggctttg aagaccgctg acttgaaacc aggccaatgg 600
 gttgctatct ccggtgctgc cgggtggtta ggttcccttg ctgtgcaata cgccaaggcc 660
 atgggtttga gaggcttagg tattgacggg ggtgatggta aggaagaatt gttcaagcaa 720
 tgtgggtggg aagtcttcat cgacttcaga aagtctaagg acatgggtgc cgatatccaa 780
 gaagctacca acggtgggtc tcacgggtgc attaacgtct cgtctccga ggctgctatc 840
 tccatgtcta ccgaatacgt tagaccaacc ggtgttggtt tcttggtcgg ttgcccagct 900
 gacgcttacg tcaagtccga agtcttctcc cagtcgtta agtccatctc catcaagggt 960
 tcttacgtcg gtaacagagc tgacaccaga gaagccaccg acttcttcac cagaggtttg 1020
 gtcaagtctc caatcaagat catcgggtcta tctgaattgc cagaagctta tgaactaatg 1080
 gaacaaggta agatcttggg tagattcgtc gttgacactt acaaataa 1128

<210> 32

<211> 8

<212> PRT

<213> *Kluyveromyces lactis*

<400> 32

Glu Ala Ile Asp Phe Phe Ser Arg

1 5

<210> 33

<211> 12

<212> PRT

<213> *Kluyveromyces lactis*

<400> 33

Ser Asn Gly Thr Val Val Leu Val Gly Leu Pro Arg

1 5 10

<210> 34

<211> 12

<212> PRT

<213> *Kluyveromyces lactis*

<400> 34

Ser Ile Ser Ile Val Gly Ser Tyr Val Gly Asn Arg

1 5 10

<210> 35

<211> 12

<212> PRT

<213> Kluyveromyces lactis

<400> 35

Asp Leu Gly Gly Glu Tyr Phe Ile Asp Phe Thr Lys

1 5 10

<210> 36

<211> 14

<212> PRT

<213> Kluyveromyces lactis

<400> 36

Gly Val Ile Phe Tyr Glu Asn Gly Gly Glu Leu Gln Tyr Lys

1 5 10

<210> 37

<211> 21

<212> PRT

<213> Kluyveromyces lactis

<400> 37

Leu Pro Leu Val Gly Gly His Glu Gly Ala Gly Val Val Val Ala Met

1 5 10 15

Gly Glu Asn Val Lys

20

<210> 38

<211> 17

<212> PRT

<213> Kluyveromyces lactis

<400> 38

Ala Pro Ile His Val Val Gly Leu Ser Glu Leu Pro Ser Ile Tyr Glu

1 5 10 15

Lys

<210> 39

<211> 24

<212> PRT

<213> Kluyveromyces lactis

<400> 39

Gly Gly Ala His Gly Val Ile Asn Val Ser Val Ser Glu Phe Ala Ile

1 5 10 15

Glu Gln Ser Thr Asn Tyr Val Arg

20

<210> 40

<211> 23

<212> PRT

<213> *Kluyveromyces lactis*

<400> 40

Ala Gly Asp Trp Val Ala Ile Ser Gly Ala Ala Gly Gly Leu Gly Ser

1

5

10

15

Leu Ala Val Gln Tyr Ala Lys

20

<210> 41

<211> 14

<212> PRT

<213> *Kluyveromyces lactis*

<400> 41

Gly Val Ile Phe Tyr Glu Asn Gly Gly Lys Leu Glu Tyr Lys

1

5

10

<210> 42

<211> 9

<212> PRT

<213> *Kluyveromyces lactis*

<400> 42

Ser Asp Val Phe Asn Gln Val Val Lys

1

5

<210> 43

<211> 11

<212> PRT

<213> *Kluyveromyces lactis*

<400> 43

Asn Ile Pro Glu Glu Val Ile Glu Ala Thr Lys

1

5

10

<210> 44

<211> 9

<212> PRT

<213> *Kluyveromyces lactis*

<400> 44

Ala Asn Glu Ile Leu Ile Asn Val Lys

1

5

<210> 45

<211> 20

<212> DNA

<213> 人工序列 (artificial sequence)
<400> 45
tgggtgaaaa cgtcaagggc 20
<210> 46
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 46
tgggtgaaaa cgtcaagggc gttttagagc tagaaatagc 40
<210> 47
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 47
gcccttgacg ttttcaccca aaagtcccat tcgccaccgc 40
<210> 48
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 48
atcgctgacc tgcaggcatg 20
<210> 49
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 49
atctctagag gatccccggg 20
<210> 50
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 50
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 51
<211> 53
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 51
gaagagattt catttatctt tttttagtat agagtttgtg tgtttaaagc ttg 53

<210> 52
<211> 55
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 52
caaactctat actaaaaaaaa gataaatgaa atctcttccg cattcaagtc atgac 55
<210> 53
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 53
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 54
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 54
gtgatggaac acgggaatag 20
<210> 55
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 55
cacatatacc ttggcagtag 20
<210> 56
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 56
tgggtgaaaa cgtcaagggc 20
<210> 57
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 57
tgggtgaaaa cgtcaagggc gttttagagc tagaaatagc 40
<210> 58
<211> 40
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 58
gcccttgacg ttttcaccca aaagtcccat tcgccaccg 40
<210> 59
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 59
atcgctgacc tgcaggcatg 20
<210> 60
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 60
atctctagag gatccccggg 20
<210> 61
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 61
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 62
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 62
agcatgaagg tcagttttac agacaccgga gtacttgacg 40
<210> 63
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 63
gtaaaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 64
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 64
ccaaccttta acattctctc ccatagcaac aacgacacca g 41

<210> 65
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 65
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 66
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 66
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 67
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 67
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 68
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 68
agcatgaagg tcagtattac agacaccgga gtacttgacg 40
<210> 69
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 69
gtaatactga ccttcatgct tggaagggtg actggccttt g 41
<210> 70
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 70
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 71
<211> 42
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 71
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 72
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 72
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 73
<211> 58
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 73
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgag 58
<210> 74
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 74
agcatgaagg tcagtttgac agacaccgga gtacttgacg 40
<210> 75
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 75
gtcaaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 76
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 76
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 77
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 77
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42

<210> 78
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 78
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 79
<211> 58
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 79
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgag 58
<210> 80
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 80
agcatgaagg tcagttctac agacaccgga gtacttgacg 40
<210> 81
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 81
gtagaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 82
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 82
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 83
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 83
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 84
<211> 59
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 84
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 85
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 85
tgggagagaa tgttaaaggt 20
<210> 86
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 86
tgacggtcgt taactaagat 20
<210> 87
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 87
tgggtgaaaa cgtcaagggc 20
<210> 88
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 88
tgggtgaaaa cgtcaagggc gttttagagc tagaaatagc 40
<210> 89
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 89
gcccttgacg ttttcaccca aaagtcccat tcgccaccgc 40
<210> 90
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 90
atcgtcgacc tgcaggcatg 20

<210> 91
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 91
atctctagag gatccccggg 20
<210> 92
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 92
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 93
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 93
ccaagcttta aggtcagtat gacagacacc ggagtacttg acg 43
<210> 94
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 94
gtcatactga ccttaaagct tggaagggtg actggccttt g 41
<210> 95
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 95
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 96
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 96
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 97
<211> 59
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 97
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 98
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 98
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 99
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 99
ccaagctcta aggtcagtat gacagacacc ggagtacttg acg 43
<210> 100
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 100
gtcatactga ccttagagct tggaagggtg actggccttt g 41
<210> 101
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 101
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 102
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 102
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 103
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 103
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59

<210> 104
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 104
tgggagagaa tgttaaaggt 20
<210> 105
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 105
tgacggtcgt taactaagat 20
<210> 106
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 106
tgggtgaaaa cgtcaagggc 20
<210> 107
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 107
tgggtgaaaa cgtcaagggc gttttagagc tagaaatagc 40
<210> 108
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 108
gcccttgacg ttttcaccca aaagtcccat tcgccaccg 40
<210> 109
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 109
atcgtcgacc tgcaggcatg 20
<210> 110
<211> 20
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 110
atctctagag gatccccggg 20
<210> 111
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 111
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 112
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 112
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 113
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 113
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 114
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 114
ccatcttttg tatatccact caagtcagct tctggacagt tgg 43
<210> 115
<211> 48
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 115
cttgagtgga tatacaaaaag atggttcttt ccaacaatac gctactgc 48
<210> 116
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 116
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59

<210> 117
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 117
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 118
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 118
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 119
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 119
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 120
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 120
ccatctcttg tatatccact caagtcagct tctggacagt tgg 43
<210> 121
<211> 48
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 121
cttgagtgga tatacaagag atggttcttt ccaacaatac gctactgc 48
<210> 122
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 122
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 123
<211> 20
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 123
tgggagagaa tgttaaaggt 20
<210> 124
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 124
tgacggtcgt taactaagat 20
<210> 125
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 125
tgggtgaaaa cgtcaagggc 20
<210> 126
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 126
tgggtgaaaa cgtcaagggc gtttagagc tagaaatagc 40
<210> 127
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 127
gcccttgacg ttttcaccca aaagtcccat tcgccaccg 40
<210> 128
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 128
atcgctgacc tgcagggcatg 20
<210> 129
<211> 20
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 129
atctctagag gatccccggg 20

<210> 130
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 130
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 131
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 131
agcatgaagg tcagttttac agacaccgga gtacttgacg 40
<210> 132
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 132
gtaaaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 133
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 133
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 134
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 134
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 135
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 135
ccatcttttg tatatccact caagtcagct tctggacagt tgg 43
<210> 136
<211> 48
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 136
cttgagtgga tatacaaaag atggttcttt ccaacaatac gctactgc 48
<210> 137
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 137
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 138
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 138
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 139
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 139
agcatgaagg tcagtattac agacaccgga gtacttgacg 40
<210> 140
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 140
gtaatactga ccttcatgct tggaagggtg actggccttt g 41
<210> 141
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 141
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 142
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 142
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42

<210> 143
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 143
ccatctttttg tataatccact caagtcagct tctggacagt tgg 43
<210> 144
<211> 48
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 144
cttgagtgga tatacaaaaag atggttcttt ccaacaatac gctactgc 48
<210> 145
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 145
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 146
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 146
tcgagctcgg taccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 147
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 147
agcatgaagg tcagtttgac agacaccgga gtacttgacg 40
<210> 148
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 148
gtcaaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 149
<211> 41
<212> DNA

<213> 人工序列 (artificial sequence)
<400> 149
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 150
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 150
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 151
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 151
ccatcttttg tatatccact caagtcagct tctggacagt tgg 43
<210> 152
<211> 48
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 152
cttgagtgga tatacaaaag atggttcttt ccaacaatac gctactgc 48
<210> 153
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 153
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 154
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 154
tcgagctcgg tacccgggga tcctctagag atcatggaca atacgttacc gagatgagg 59
<210> 155
<211> 40
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 155
agcatgaagg tcagttctac agacaccgga gtacttgacg 40

<210> 156
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 156
gtagaactga ccttcatgct tggaagggtg actggccttt g 41
<210> 157
<211> 41
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 157
ccaaccttta acattctctc ccatagcaac aacgacacca g 41
<210> 158
<211> 42
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 158
gggagagaat gttaaagggtt ggaagattgg tgacttcgct gg 42
<210> 159
<211> 43
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 159
ccatcttttg tatatccact caagtcagct tctggacagt tgg 43
<210> 160
<211> 48
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 160
cttgagtgga tatacaaaaag atggttcttt ccaacaatac gctactgc 48
<210> 161
<211> 59
<212> DNA
<213> 人工序列 (artificial sequence)
<400> 161
tgccaagctt gcatgcctgc aggtcgacga tcttatactg ggtagatttc aaacgggac 59
<210> 162
<211> 20
<212> DNA

<213> 人工序列 (artificial sequence)

<400> 162

tgggagagaa tgttaaaggt 20

<210> 163

<211> 20

<212> DNA

<213> 人工序列 (artificial sequence)

<400> 163

tgacggtcgt taactaagat 20

<210> 164

<211> 1053

<212> DNA

<213> *Kluyveromyces lactis*

<400> 164

atggccgcat ctatcccaga aactcaaaag ggtgttatct tctacgaaa cggtggtgaa 60
ttgcaataca aggacatccc agttccaaag ccaaaggcta acgaactttt gatcaacgtc 120
aagtactccg gtgtctgtca caccgatttg cacgcatgga agggtgactg gcctttgcca 180
accaaattgc cattagtgg tggtcacgaa ggtgctgggtg tcgttggtgc tatgggtgaa 240
aacgtcaagg gctggaagat tggtgacttc gctggtatca aatggttgaa cggttcttgt 300
atgtcctgtg aatactgtga attgtccaac gaatccaact gtccagaagc tgacttgtcc 360
ggttacactc acgacggttc tttccaacaa tacgctactg ctgatgccgt tcaagctgcc 420
aagatcccag tcggtactga cttggctgaa gttgctccag tgctatgtgc tgggtgcacc 480
gtttacaagg ccctaaaatc cgccaacttg aaggccgggtg actgggtcgc catctctggt 540
gctgctgggtg gtctagggtc tctagctgtc caatacgcca aggccatggg ttacagagtg 600
ttgggtatcg atgctgggtg agaaaaggct aagttgttca aggacttggg tggatgaatac 660
ttcattgatt tcaccaagtc caagaacatc ccagaagaag tcatcgaagc taccaagggt 720
ggtgctcacg gtgtcatcaa cgtctctgtc tccgaattcg ctatcgaaca atctaccaac 780
tacgtcagat ctaacggtac cgtcgtattg gtcggtctac caagagacgc caagtgtgaa 840
tccgatgtct ttaaccaagt tgtgaagtcc atctccattg tcggttctta cgtcggtaac 900
agagctgaca ccagagaagc cattgacttc ttctccagag gtctagtcaa ggctccaatc 960
cacgtcggtg gtttgtccga actaccatcc atctacgaaa agatggaaaa ggggtgctatc 1020
gtcggtagat acgtcgtcga cacttcaaaa taa 1053

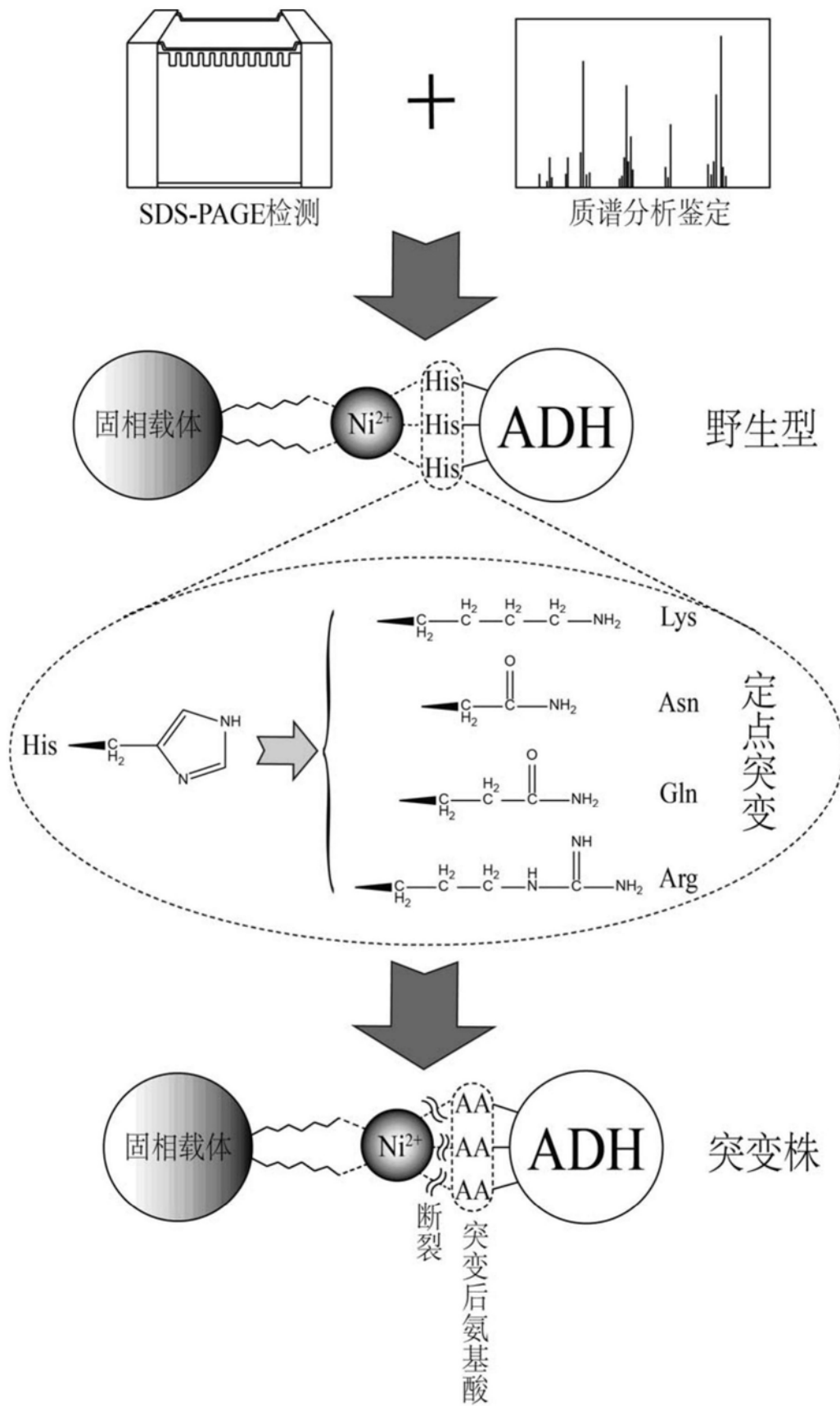


图1

KlADH4	1	MFRLLARAQTALANKASVSRSLRLNSSFAIPETQKGVIFYENGGLLEYKDIIPVPKPKANE
KlADH3	1	MLRLTSARSIVSPLRKAGFG.SIRTLATSVIPETQKGVIFYENGGLLEYKDIIPVPKPKPNE
KlADH1	1	MAASIPETQKGVIFYENGGLLEYKDIIPVPKPKANE
KlADH2	1	MSTIPETQKGVIFYENGGLLEYKDIIPVPKPKANE
KlADH4	61	ILINVKYSGVCHTDLHAWKGDWPLPVKLPLVGGHEGAGIVVAKGENVKNFIEIGDYAGIKW
KlADH3	60	ILINVKYSGVCHTDLHAWKGDWPLPTKLPLVGGHEGAGVVVAMGENVKGNINIGDFAGIKW
KlADH1	36	ILINVKYSGVCHTDLHAWKGDWPLPTKLPLVGGHEGAGVVVAMGENVKGNINIGDFAGIKW
KlADH2	34	ILINVKYSGVCHTDLHAWKGDWPLPTKLPLVGGHEGAGVVVAMGENVKGNINIGDFAGIKW
KlADH4	121	LNGSCMSCEYCEQGYESNCLQADLSGYTHDGSFQQYATADAVQAAQIPKGTDLAEIAPIL
KlADH3	120	LNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAAARIPKGTDLAEVAPIL
KlADH1	96	LNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAAKIPVGTDLAEVAPIL
KlADH2	94	LNGSCMSCEYCELSNESNCPDADLSGYTHDGSFQQYATADAVQAAARIPKGTDLAEVAPIL
KlADH4	181	CAGVTVYKALKTADLKFGQWVAISGAAGGLGSLAVQYAKAMGLRVIGIDGGGKEELFKQ
KlADH3	180	CAGVTVYKALKSANLKAGDWVAISGAAGGLGSLAVQYAKAMGYRVIGIDGGEKGLVKQ
KlADH1	156	CAGVTVYKALKSANLKAGDWVAISGAAGGLGSLAVQYAKAMGYRVIGIDAGEEKAKEKD
KlADH2	154	CAGVTVYKALKSADLKAGDWVAISGAAGGLGSLAVQYAKAMGYRVIGIDTGAEKAKEKE
KlADH4	241	CGGEVFIDFRKSKDMVADIQEATNGGHHGVINVSVEAAISMSTEVVRPTGVVVLVLGLPA
KlADH3	240	LGGEAFVDFITKTKDMVAEIQEITNGGHHGVINVSVEAAAMNASTQVRPTGTIVVLVLGLPA
KlADH1	216	LGGEYFIDFTKSKNIPPEVIEATKGGHHGVINVSVEFAIEQSTNVVRSNGTIVVLVLGLR
KlADH2	214	LGGEYFVDYAVSKDLIKEIVDATNGGHHGVINVSVEFAIEQSTNVVRSNGTIVVLVLGLR
KlADH4	301	DAYVKSEVFSHVVKSESIKGSYVGNRADTREATDFFTRGLVKSPIKIIIGLSELPEAYELM
KlADH3	300	GAVIKSEVFSHVVKSEINIKGSYVGNRADTREATINFFANGHVHSPIKVVGLSELPKVYELM
KlADH1	276	DAKCKSDVFNQVVKSSIVGSYVGNRADTREATIDFFSRGLVKAPIIHVVGSELPSIYELM
KlADH2	274	DAKCKSDVFTQVVKSSIVGSYVGNRADTREATLDFFRGLVKAPIIKVGLSELADVYDKM
KlADH4	361	EQGRILGRFVVDITYK
KlADH3	360	EQGRILGRYVVDITSN
KlADH1	336	EKGATVGRYVVDITSK
KlADH2	334	VKGEIVGRYVVDITSK

图2

KlADH1	1	10	20	30	40	50	60	70
4w6r_ScADH1	1	10	20	30	40	50	60	70
KlADH1	80	90	100	110	120	130	140	150
4w6r_ScADH1	80	90	100	110	120	130	140	150
KlADH1	160	170	180	190	200	210	220	
4w6r_ScADH1	160	170	180	190	200	210	220	
KlADH1	230	240	250	260	270	280	290	300
4w6r_ScADH1	230	240	250	260	270	280	290	300
KlADH1	310	320	330	340	350	360	370	380
4w6r_ScADH1	310	320	330	340	350	360	370	380
KlADH1	390	400	410	420	430	440	450	
4w6r_ScADH1	390	400	410	420	430	440	450	
KlADH1	460	470	480	490	500	510	520	530
4w6r_ScADH1	460	470	480	490	500	510	520	530
KlADH1	540	550	560	570	580	590	600	
4w6r_ScADH1	540	550	560	570	580	590	600	

图3



图4

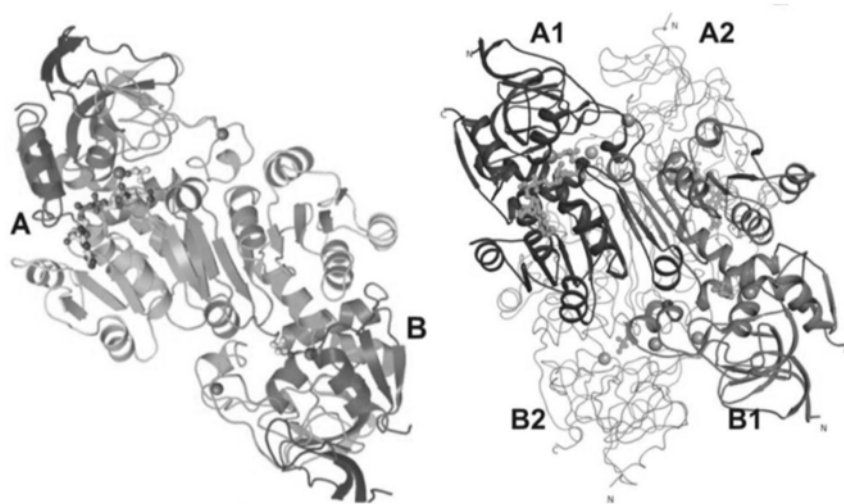


图5

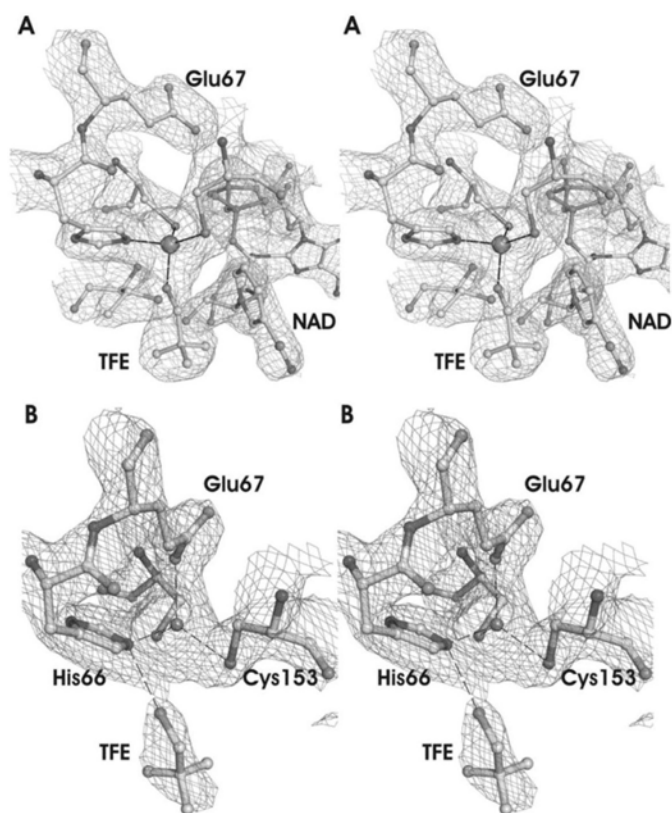


图6

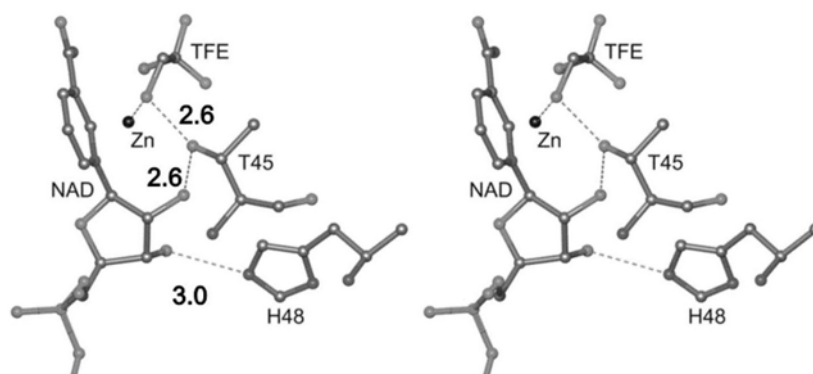


图7

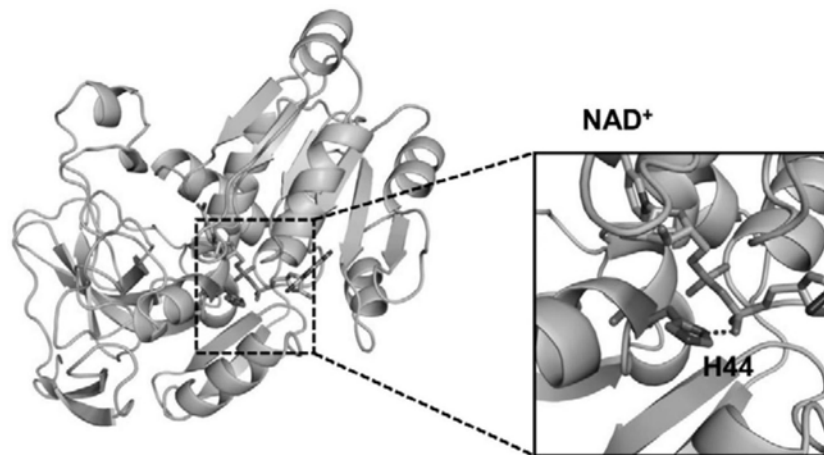


图8

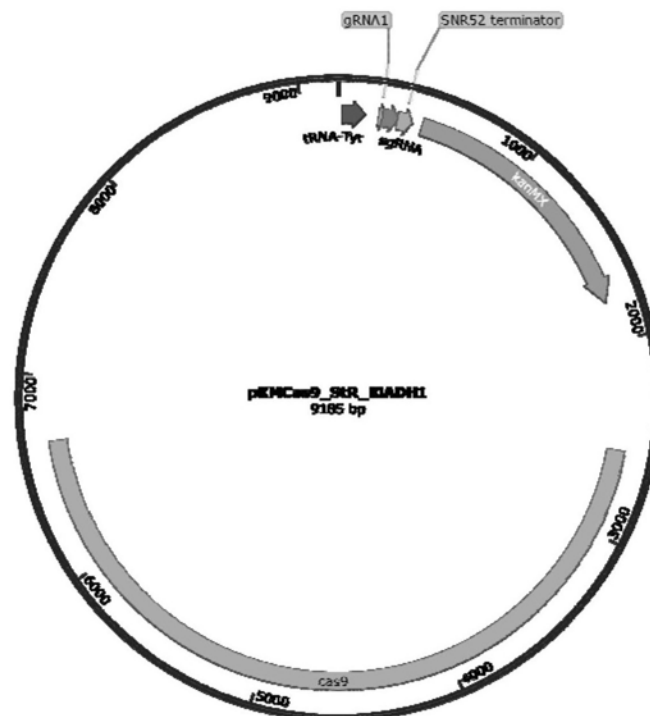


图9

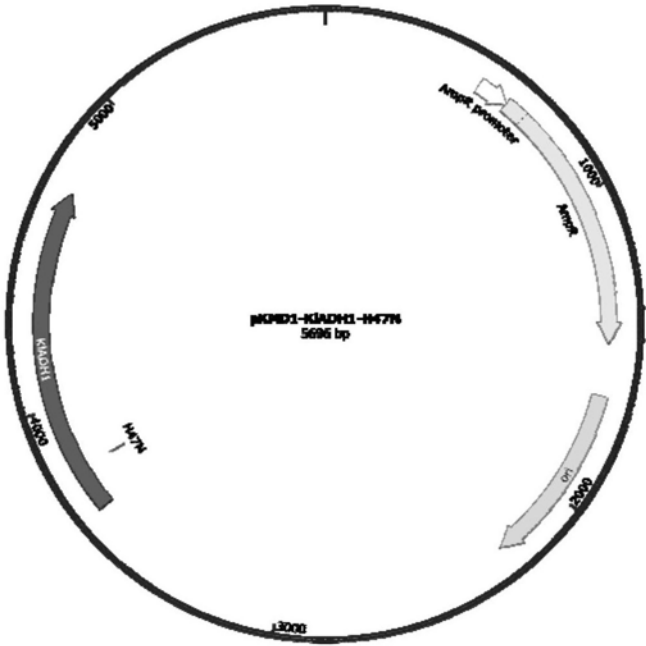


图12

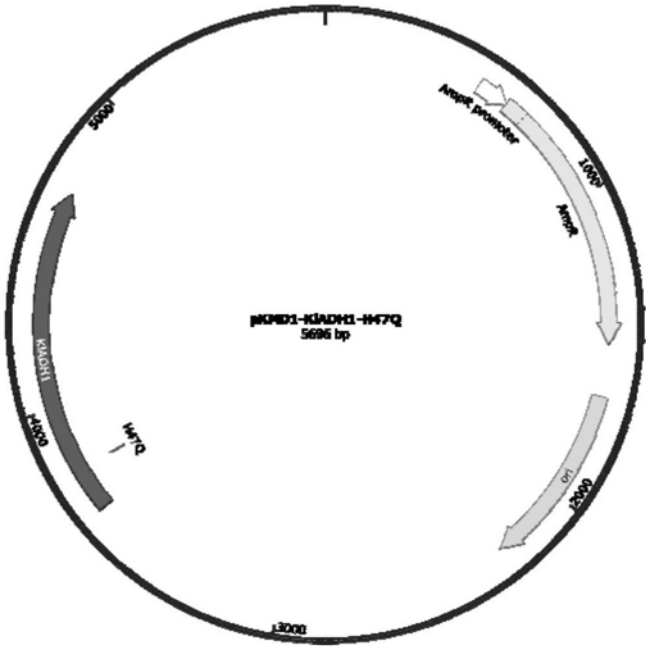


图13

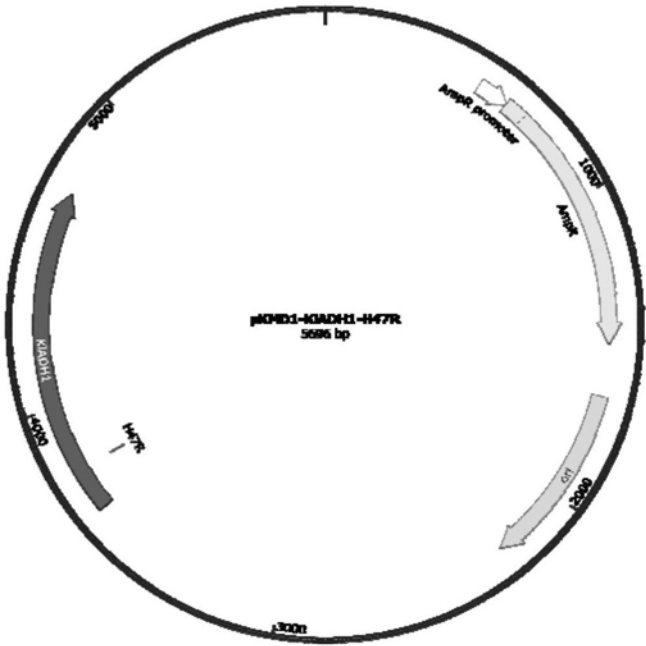


图14

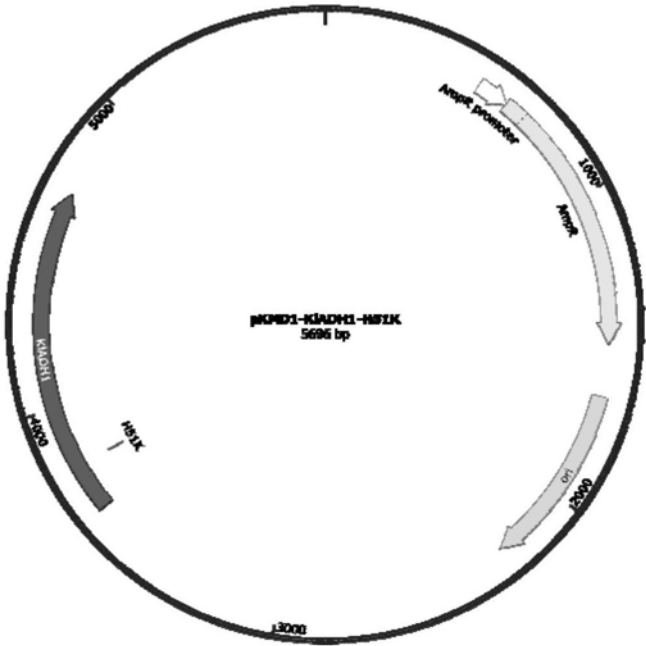


图15

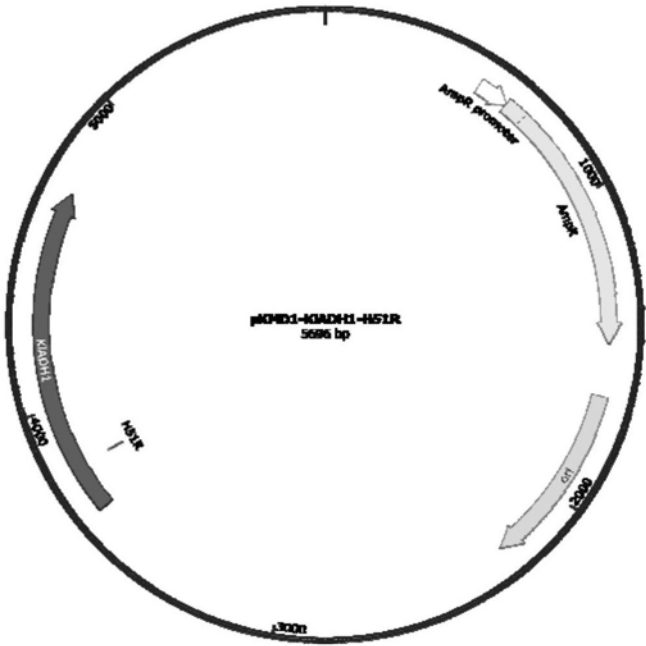


图16

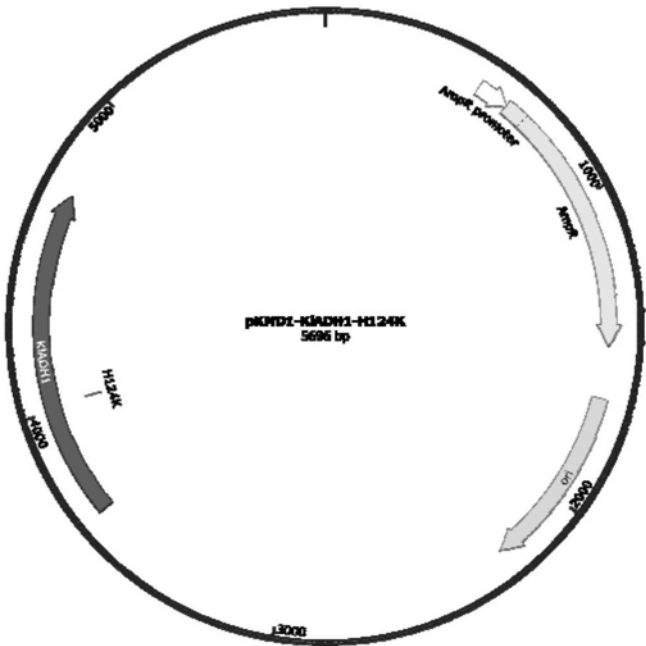


图17

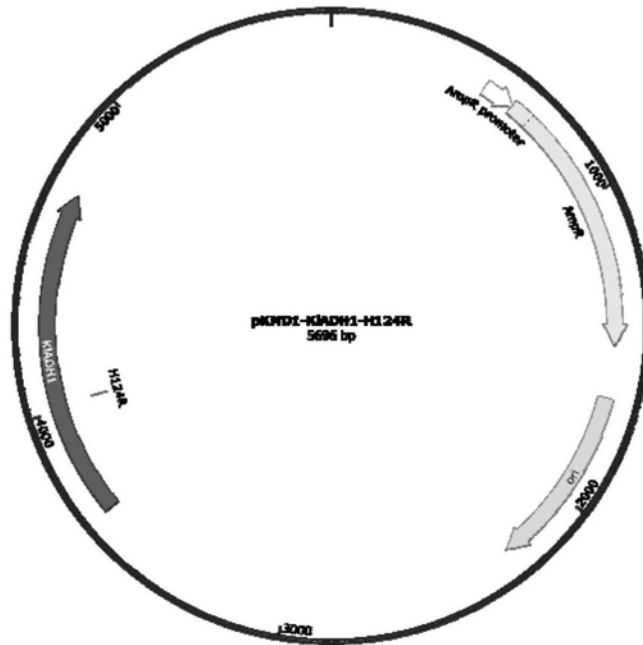


图18

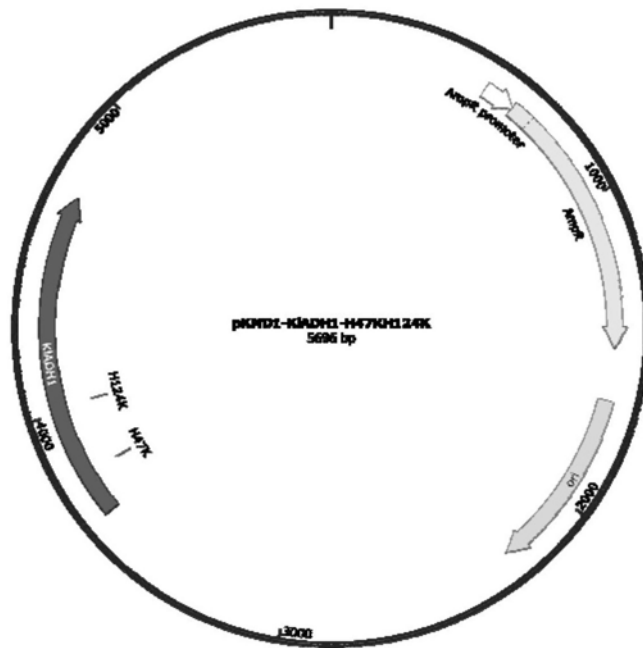


图19

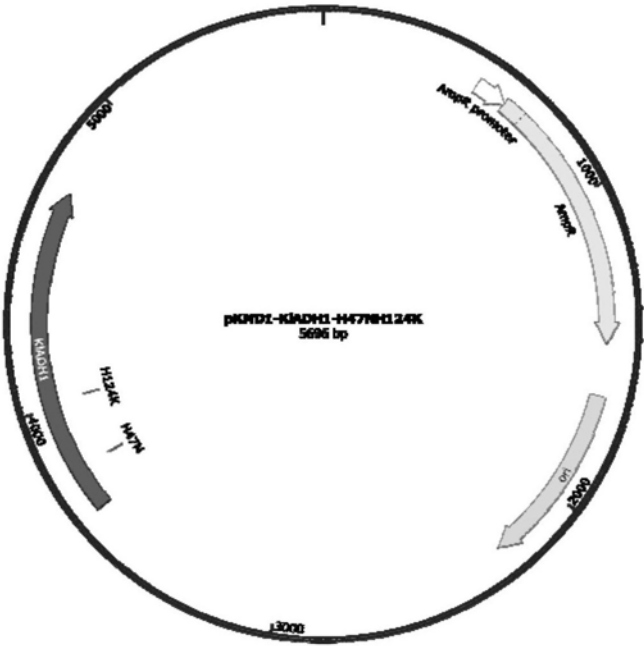


图20

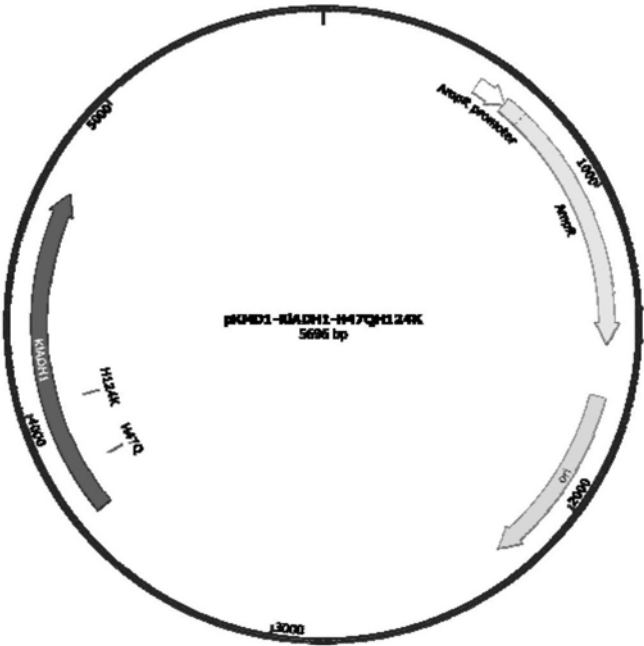


图21

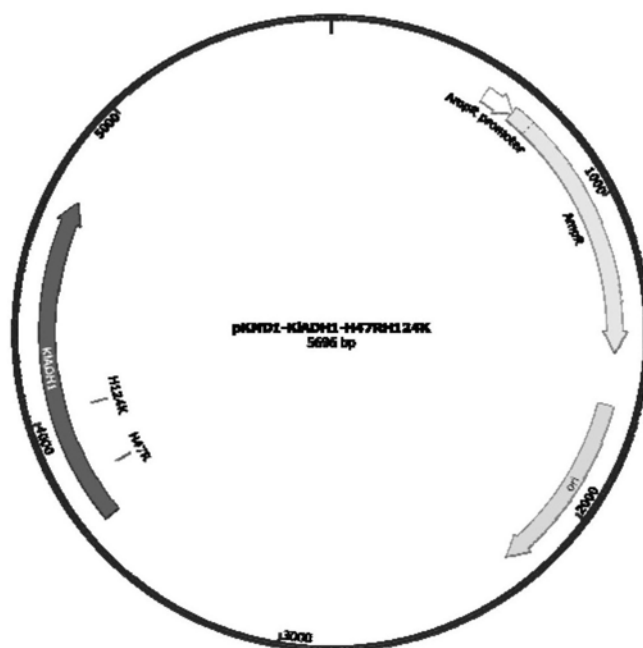


图22

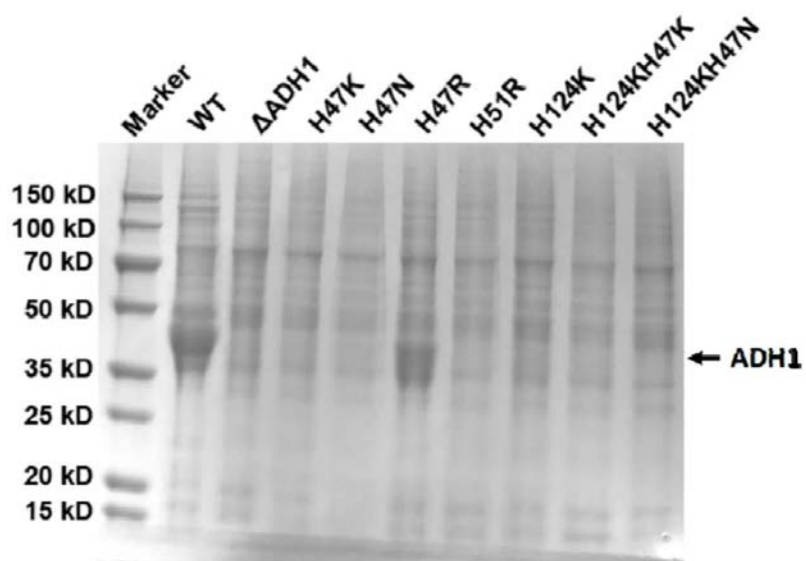


图23

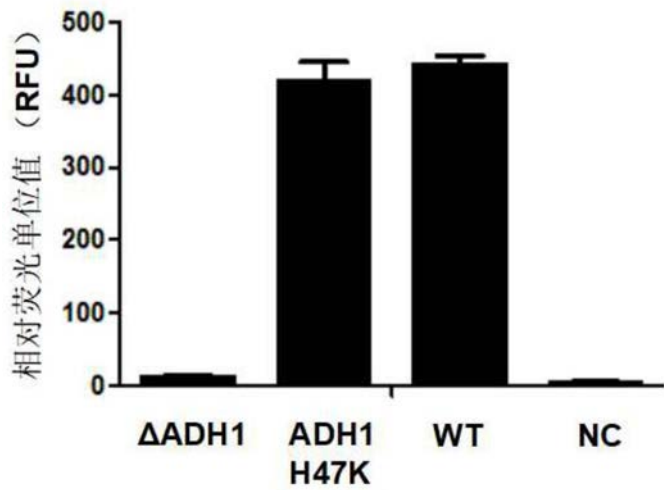


图24

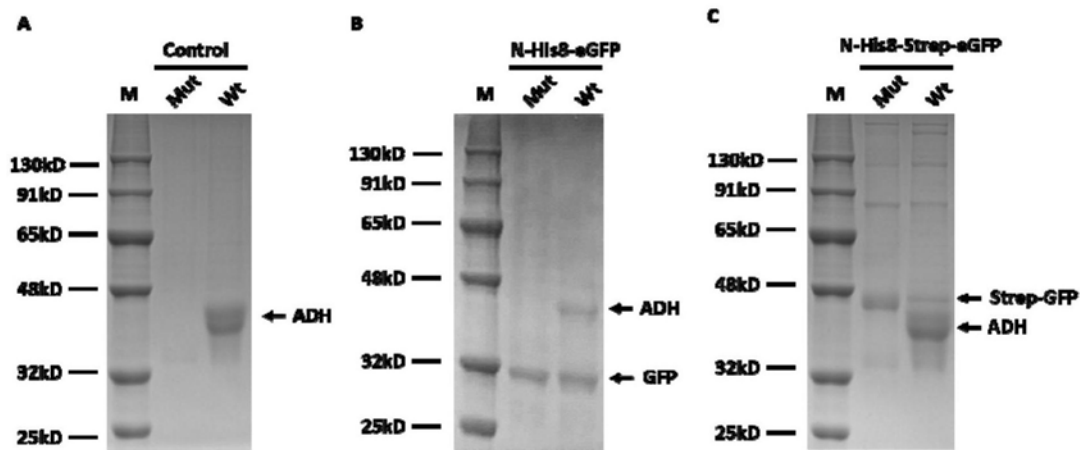


图25

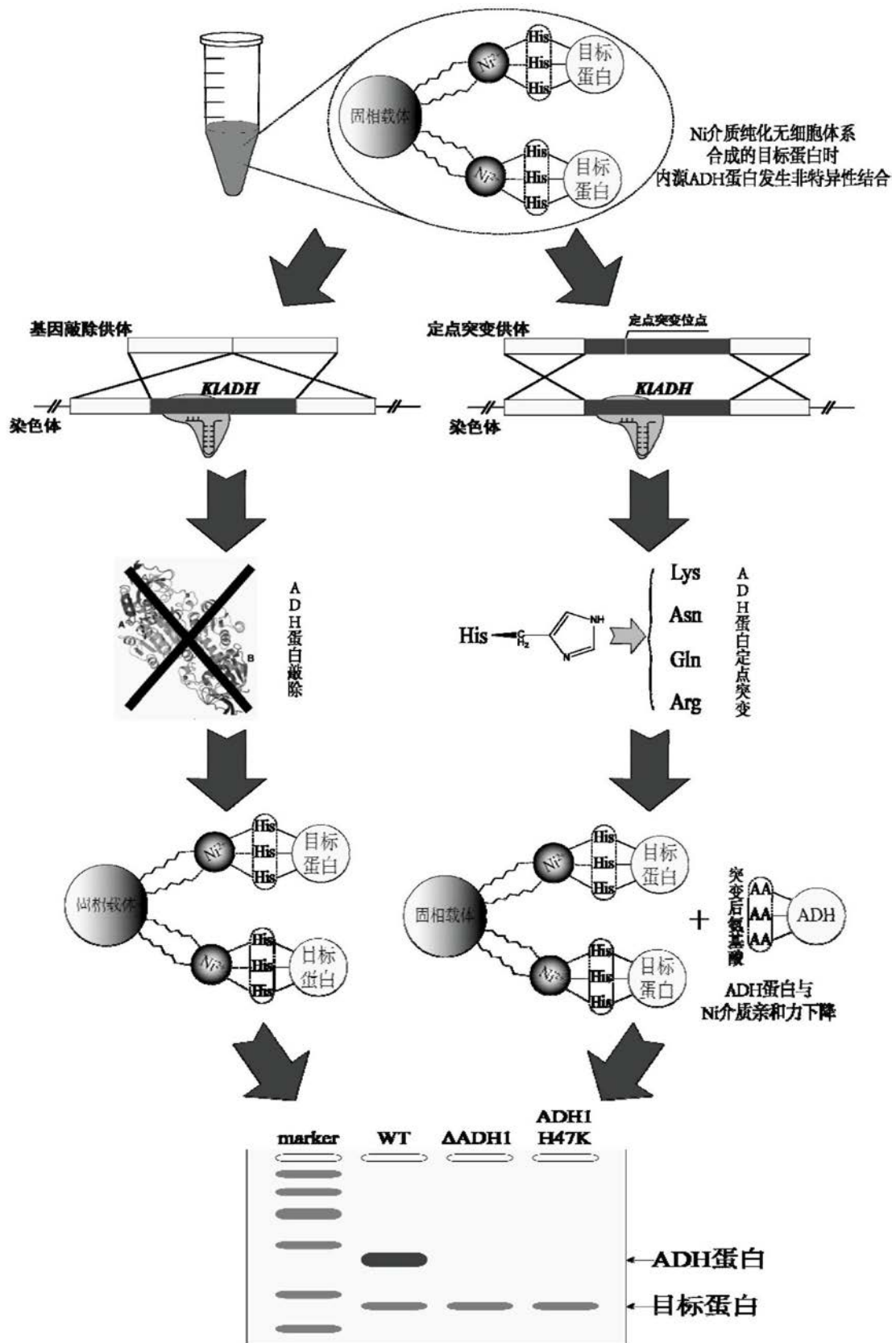


图26