



(12) 发明专利

(10) 授权公告号 CN 105793424 B

(45) 授权公告日 2021. 02. 19

(21) 申请号 201480050439.3

(22) 申请日 2014.07.14

(65) 同一申请的已公布的文献号  
申请公布号 CN 105793424 A

(43) 申请公布日 2016.07.20

(30) 优先权数据  
13003539.7 2013.07.12 EP

(85) PCT国际申请进入国家阶段日  
2016.03.11

(86) PCT国际申请的申请数据  
PCT/EP2014/001925 2014.07.14

(87) PCT国际申请的公布数据  
W02015/003816 EN 2015.01.15

(83) 生物保藏信息  
DSM 27004 2013.03.11

(73) 专利权人 亥姆霍兹感染研究中心有限公司  
地址 德国不伦瑞克

(72) 发明人 S·鲍曼 J·赫尔曼 K·莫尔  
H·斯泰因梅茨 K·格斯  
R·拉朱 R·穆勒 R·哈特曼  
M·哈米德 W·A·M·埃尔加

M·莫雷诺 F·吉勒 王亮亮  
A·柯书宁

(74) 专利代理机构 上海一平知识产权代理有限公司 31266

代理人 崔佳佳 马莉华

(51) Int.Cl.  
C12N 15/52 (2006.01)  
C12P 13/02 (2006.01)  
C07C 237/44 (2006.01)

(56) 对比文件  
W0 0180846 A8,2003.03.20  
Jean Euzeby.List of new names and  
new combinations previously effectively,  
but not validly, published.《International  
Journal of Systematic and Evolutionary  
Microbiology》.2007,第57卷第893-897页.  
无.EBI accession no. UniprotKB-  
Q70C52,“Submitted name:Non-ribosomal  
peptide synthase”.《EMBL-EBI》.2004,

审查员 于淼

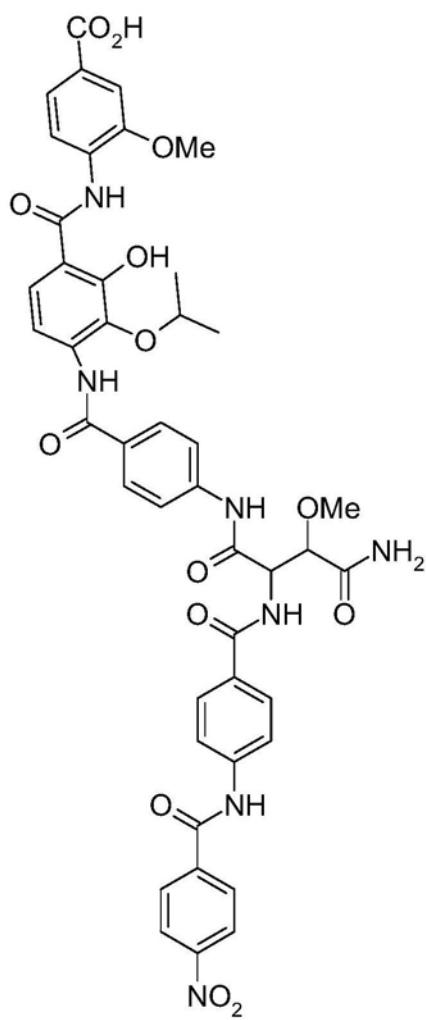
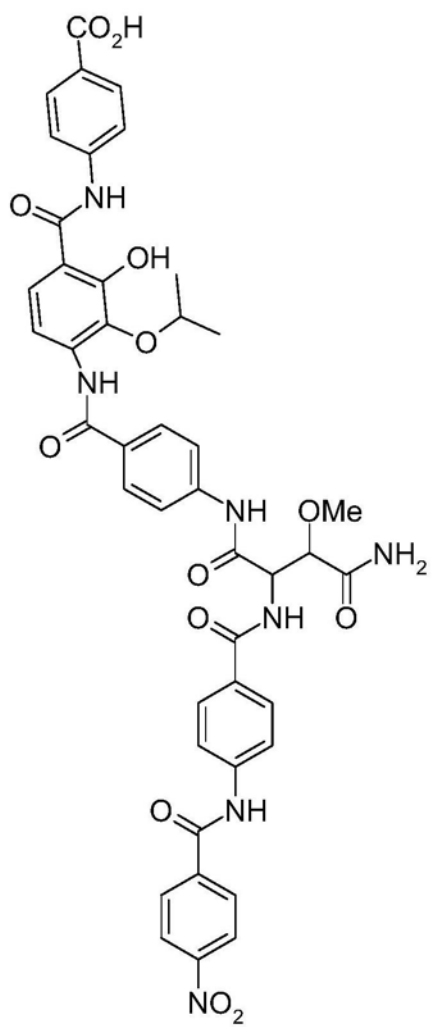
权利要求书6页 说明书88页  
序列表74页 附图13页

(54) 发明名称  
孢囊菌酰胺

(57) 摘要  
本发明提供了一种如式(I)所示的孢囊菌酰  
胺,以及其用于细菌感染的治疗或预防的用途。  
 $R^1-Ar^1-L^1-Ar^2-L^2-Ar^3-L^3-Ar^4-L^4-Ar^5-R^2$  (I)。

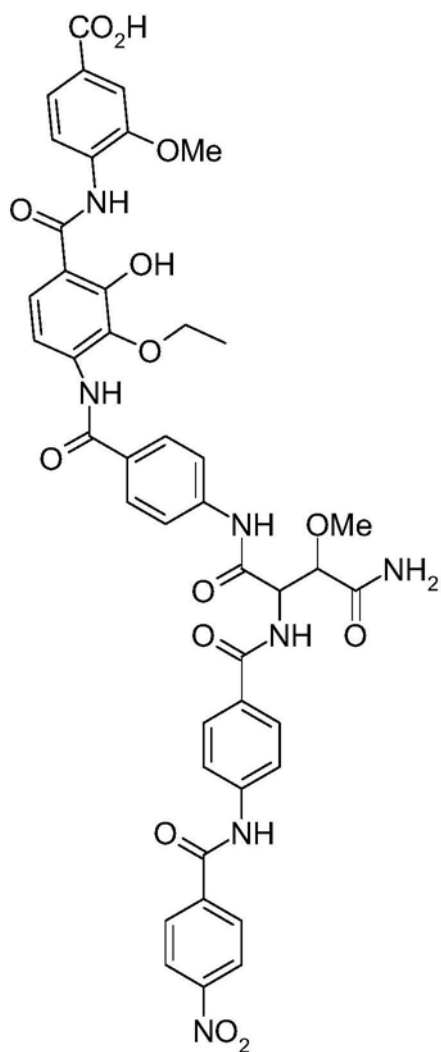
Chemical structures of two diastereomers of a cyclic peptide derivative, labeled 1 and 2. Both structures are cyclic peptides with a 14-membered ring. The backbone consists of a 4-nitrophenyl group, a 4-aminobenzoyl group, a 4-aminobenzoyl group, and a 4-aminobenzoyl group. The side chains are: 1) 4-nitrophenyl, 2) 4-aminobenzoyl, 3) 4-aminobenzoyl, and 4) 4-aminobenzoyl. The structures are shown as two diastereomers, 1 and 2, with different stereochemistry at the chiral centers.



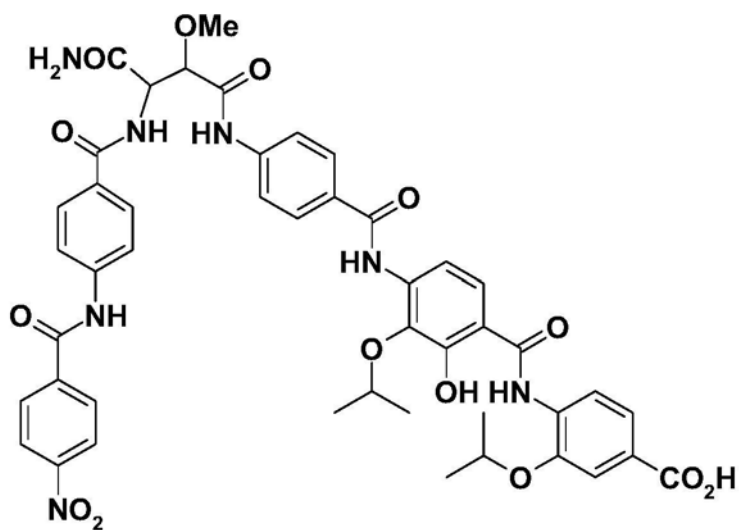


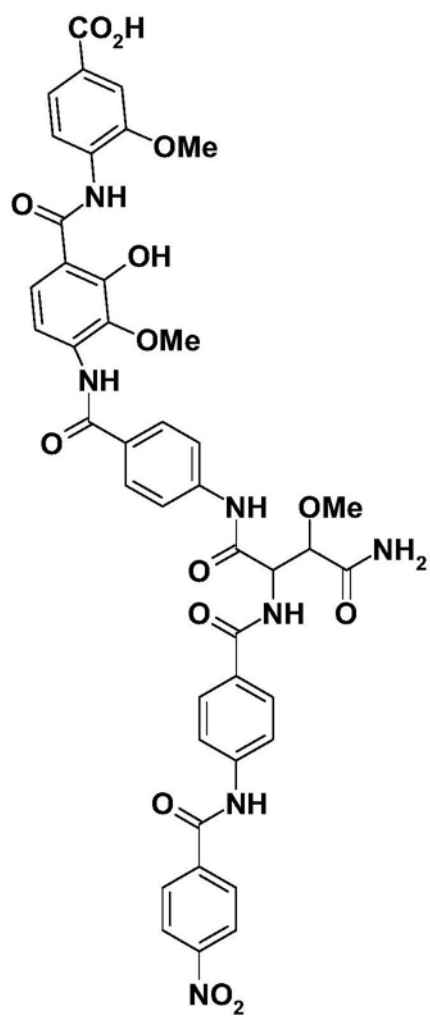
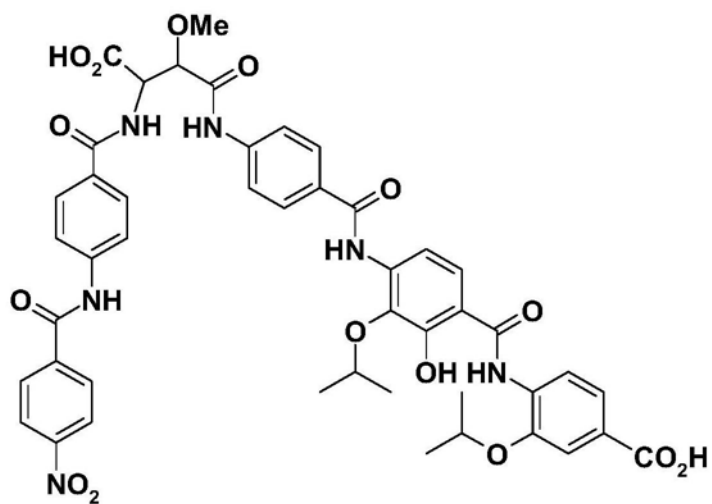
或

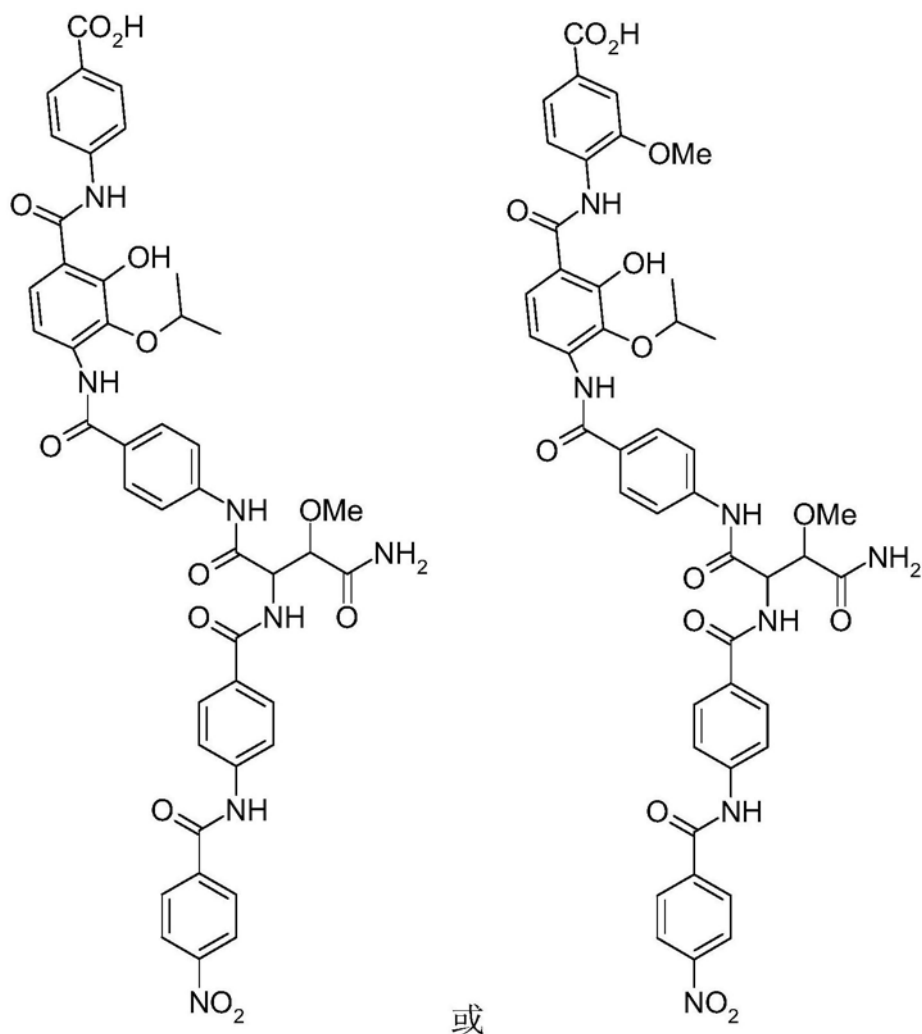




2. 如权利要求1所述的化合物,其特征在于,所述的化合物选自下组:







3. 一种药物组合物, 包括根据前述任一权利要求所述的化合物, 以及任选的一种或多种载体物质和/或一种或多种佐剂。

4. 如权利要求1-2任一所述的化合物的用途, 其特征在于, 用于制备治疗或预防细菌感染的药物组合物。

5. 一种制备如权利要求1所述的化合物的方法, 其特征在于, 所述方法包括以下步骤:

- (a) 培养保藏号为DSM27004的深棕色孢囊杆菌菌株; 和
- (b) 从培养液中分离并保留所述化合物。

## 孢囊菌酰胺

[0001] 孢囊菌酰胺是一种从粘细菌深棕色孢囊杆菌 (*Cystobacter velatus*) (MCy8071; 内部名称: 深棕色孢囊杆菌 (*Cystobacter ferrugineus*)) 中分离出的新天然产物。孢囊菌酰胺表现出良好的抗菌活性 (特别是对于选定的革兰氏阴性菌例如大肠杆菌、绿脓杆菌和鲍曼不动杆菌), 且对革兰氏阳性菌具有广谱的活性。

[0002] 本发明提供了一种式 (I) 化合物, 或其药学上可接受的盐、溶剂化物或水合物, 或其药学上可接受的制剂

[0003]  $R^1-Ar^1-L^1-Ar^2-L^2-Ar^3-L^3-Ar^4-L^4-Ar^5-R^2$

[0004] (I)

[0005] 其中,

[0006]  $Ar^1$  为任选取代的亚苯基, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基;

[0007]  $Ar^2$  为任选取代的亚苯基, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基;

[0008]  $Ar^3$  为任选取代的亚苯基, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基;

[0009]  $Ar^4$  为无, 或任选取代的亚苯基, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基;

[0010]  $Ar^5$  为无, 或任选取代的亚苯基, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基;

[0011]  $L^1$  为键、氧原子、硫原子或如式 NH、CONH、NHCO、COO、OCO、CONR<sup>3</sup>、NR<sup>3</sup>CO、OCONH、NHCOO、NHCONH、OCONR<sup>3</sup>、NR<sup>3</sup>COO、NR<sup>3</sup>CONR<sup>4</sup>、NR<sup>3</sup>、-CNR<sup>3</sup>-、-CO-、-SO-、-SO<sub>2</sub>-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-SO<sub>2</sub>NR<sup>3</sup>-、-NR<sup>3</sup>SO<sub>2</sub>-、-COCH<sub>2</sub>-、-CH<sub>2</sub>CO-、-COCR<sup>3</sup>R<sup>4</sup>-、-CR<sup>3</sup>R<sup>4</sup>CO-、-NHCSNH-、-NR<sup>3</sup>CSNR<sup>4</sup>、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>- 所示的基团, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2 或 3 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基, 或杂亚烃基;

[0012]  $L^2$  为键、氧原子、硫原子或如式 NH、CONH、NHCO、COO、OCO、CONR<sup>3</sup>、NR<sup>3</sup>CO、OCONH、NHCOO、NHCONH、OCONR<sup>3</sup>、NR<sup>3</sup>COO、NR<sup>3</sup>CONR<sup>4</sup>、NR<sup>3</sup>、-CNR<sup>3</sup>-、-CO-、-SO-、-SO<sub>2</sub>-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-SO<sub>2</sub>NR<sup>3</sup>-、-NR<sup>3</sup>SO<sub>2</sub>-、-COCH<sub>2</sub>-、-CH<sub>2</sub>CO-、-COCR<sup>3</sup>R<sup>4</sup>-、-CR<sup>3</sup>R<sup>4</sup>CO-、-NHCSNH-、-NR<sup>3</sup>CSNR<sup>4</sup>、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>- 所示的基团, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2 或 3 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基, 或杂亚烃基;

[0013]  $L^3$  为无, 或键、氧原子、硫原子或如式 NH、CONH、NHCO、COO、OCO、CONR<sup>3</sup>、NR<sup>3</sup>CO、OCONH、NHCOO、NHCONH、OCONR<sup>3</sup>、NR<sup>3</sup>COO、NR<sup>3</sup>CONR<sup>4</sup>、NR<sup>3</sup>、-CNR<sup>3</sup>-、-CO-、-SO-、-SO<sub>2</sub>-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-SO<sub>2</sub>NR<sup>3</sup>-、-NR<sup>3</sup>SO<sub>2</sub>-、-COCH<sub>2</sub>-、-CH<sub>2</sub>CO-、-COCR<sup>3</sup>R<sup>4</sup>-、-CR<sup>3</sup>R<sup>4</sup>CO-、-NHCSNH-、-NR<sup>3</sup>CSNR<sup>4</sup>、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>- 所示的基团, 或任选取代的具有 5 至 6 个环原子 (其中包括 1、2 或 3 个选自下组的杂原子: 氧、硫和氮) 的杂亚芳基, 或杂亚烃基;

[0014]  $L^4$  为无, 或键、氧原子、硫原子或如式 NH、CONH、NHCO、COO、OCO、CONR<sup>3</sup>、NR<sup>3</sup>CO、OCONH、NHCOO、NHCONH、OCONR<sup>3</sup>、NR<sup>3</sup>COO、NR<sup>3</sup>CONR<sup>4</sup>、NR<sup>3</sup>、-CNR<sup>3</sup>-、-CO-、-SO-、-SO<sub>2</sub>-、-SO<sub>2</sub>NH-、-

NHSO<sub>2</sub>-、-SO<sub>2</sub>NR<sup>3</sup>-、-NR<sup>3</sup>SO<sub>2</sub>-、-COCH<sub>2</sub>-、-CH<sub>2</sub>CO-、-COCR<sup>3</sup>R<sup>4</sup>-、-CR<sup>3</sup>R<sup>4</sup>CO-、-NHCSNH-、-NR<sup>3</sup>CSNR<sup>4</sup>-、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>-所示的基团,或任选取代的具有5至6个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的杂亚芳基,或杂亚烷基;

[0015] R<sup>1</sup>为氢原子,卤素原子,羟基,氨基,硫醇基,硝基,如式-COOH-、-SO<sub>2</sub>NH<sub>2</sub>-、-CONH<sub>2</sub>-、-NO<sub>2</sub>或-CN所示的基团,烷基,烯基,炔基,杂烷基,环烷基,杂环烷基,烷基环烷基,杂烷基环烷基,芳基,杂芳基,芳烷基或杂芳基;

[0016] R<sup>2</sup>为氢原子,卤素原子,羟基,氨基,硫醇基,硝基,如式-COOH-、-SO<sub>2</sub>NH<sub>2</sub>-、-CONH<sub>2</sub>-、-NO<sub>2</sub>或-CN所示的基团,烷基,烯基,炔基,杂烷基,环烷基,杂环烷基,烷基环烷基,杂烷基环烷基,芳基,杂芳基,芳烷基或杂芳基;

[0017] R<sup>3</sup>基团各自独立地为氢原子或C<sub>1-6</sub>烷基基团;和

[0018] R<sup>4</sup>基团各自独立地为氢原子或C<sub>1-6</sub>烷基基团。

[0019] 表述烷基是指饱和的直链或支链烷基,它包含1至20个碳原子,优选1至15个碳原子,特别是1至10个(例如1、2、3或4)个碳原子,例如甲基、乙基、丙基、异丙基、正丁基、异丁基、仲丁基、叔丁基、正戊基、异戊基、正己基、2,2-二甲基丁基或正辛基。

[0020] 表述烯基和炔基指至少部分不饱和的直链或支链烷基,包含2至20个碳原子,优选2至15个碳原子,特别是2至10(例如2、3或4)个碳原子,例如乙烯基(乙烯基)、丙烯基(烯丙基)、异丙烯基、丁烯基、乙炔基、丙炔基、丁炔基、乙炔基、炔丙基、异戊二烯基或己-2-烯基。优选地,烯基团具有一个或两个(特别优选一个)双键,且炔基基团具有一个或两个(特别优选一个)三键。

[0021] 此外,术语烷基、烯基和炔基指其中一个或多个氢原子已被替换为卤素原子(优选F或Cl)的基团,诸如,例如,2,2,2-三氯乙基或三氟甲基。

[0022] 表述杂烷基是指烷基、烯基或炔基基团,所述基团中一个或多个(优选1至8个,特别优选为1、2、3或4个)碳原子已被氧、氮、磷、硼、硒、硅或硫原子(优选由氧、硫或氮原子)取代,或被SO或SO<sub>2</sub>基取代。表述杂烷基进一步是指羧酸或由羧酸衍生的基团,如,例如,酰基、酰基烷基、烷氧基羰基、酰氧基、酰氧基烷基、羧基烷基酰胺、或烷氧基羰基氧基。

[0023] 优选地,杂烷基基团含有1至12个碳原子和1至8个选自氧、氮和硫(特别是氧和氮)的杂原子。特别优选地,杂烷基基团含有1至6个(例如,1、2、3或4)个碳原子和1、2、3或4个(特别是1、2或3)选自氧、氮和硫的杂原子(特别是氧和氮)。术语C<sub>1-6</sub>杂烷基指具有1至6个碳原子和1、2或3个选自O,S和/或N(特别是O和/或N)的杂原子的杂烷基基团。术语C<sub>1-4</sub>杂烷基指具有1至4个碳原子和1、2或3个选自O,S和/或N(特别是O和/或N)的杂原子的杂烷基基团。此外,术语杂烷基是指其中一个或多个氢原子已被替换为卤素原子(优选F或Cl)的基团。

[0024] 特别优选的是,表述杂烷基是指如上所定义的(直链或支链)的烷基基团,其中一个或多个(优选1至6;特别优选1、2、3或4)个碳原子已被氧、硫或氮原子所替换。上述基团优选含有1至6个(例如,1、2、3或4)个碳原子和1、2、3或4(特别是1、2或3)选自氧、氮和硫(特别是氧和氮)的杂原子;所述的基团可以优选被一个或多个(优选为1至6;特别优选1、2、3或4)个氟、氯、溴或碘原子、或OH、=O、SH、=S、NH<sub>2</sub>、=NH、N<sub>3</sub>、CN或NO<sub>2</sub>基团所取代。

[0025] 表述亚杂烷基基团指二价的杂烷基基团。

[0026] 杂烷基基团的例子是下式所示的基团:R<sup>a</sup>-O-Y<sup>a</sup>-、R<sup>a</sup>-S-Y<sup>a</sup>-、R<sup>a</sup>-SO-Y<sup>a</sup>-、R<sup>a</sup>-SO<sub>2</sub>-

$Y^a-$ 、 $R^a-N(R^b)-Y^a-$ 、 $R^a-CO-Y^a-$ 、 $R^a-O-CO-Y^a-$ 、 $R^a-CO-O-Y^a-$ 、 $R^a-CO-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CO-Y^a-$ 、 $R^a-O-CO-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CO-O-Y^a-$ 、 $R^a-N(R^b)-CO-N(R^c)-Y^a-$ 、 $R^a-O-CO-O-Y^a-$ 、 $R^a-N(R^b)-C(=NR^d)-N(R^c)-Y^a-$ 、 $R^a-CS-Y^a-$ 、 $R^a-O-CS-Y^a-$ 、 $R^a-CS-O-Y^a-$ 、 $R^a-CS-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CS-Y^a-$ 、 $R^a-O-CS-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CS-O-Y^a-$ 、 $R^a-N(R^b)-CS-N(R^c)-Y^a-$ 、 $R^a-O-CS-O-Y^a-$ 、 $R^a-S-CO-Y^a-$ 、 $R^a-CO-S-Y^a-$ 、 $R^a-S-CO-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CO-S-Y^a-$ 、 $R^a-S-CO-O-Y^a-$ 、 $R^a-O-CO-S-Y^a-$ 、 $R^a-S-CO-S-Y^a-$ 、 $R^a-S-CS-Y^a-$ 、 $R^a-CS-S-Y^a-$ 、 $R^a-S-CS-N(R^b)-Y^a-$ 、 $R^a-N(R^b)-CS-S-Y^a-$ 、 $R^a-S-CS-O-Y^a-$ 、 $R^a-O-CS-S-Y^a-$ ，其中 $R^a$ 为氢原子、 $C_1-C_6$ 烷基、 $C_2-C_6$ 烯基或 $C_2-C_6$ 炔基基团； $R^b$ 为氢原子、 $C_1-C_6$ 烷基、 $C_2-C_6$ 烯基或 $C_2-C_6$ 炔基基团； $R^c$ 为氢原子、 $C_1-C_6$ 烷基、 $C_2-C_6$ 烯基或 $C_2-C_6$ 炔基基团； $R^d$ 为氢原子、 $C_1-C_6$ 烷基、 $C_2-C_6$ 烯基或 $C_2-C_6$ 炔基基团，且 $Y^a$ 为键、 $C_1-C_6$ 亚烷基、 $C_2-C_6$ 亚烯基或 $C_2-C_6$ 亚炔基基团，其中各个杂烷基基团含有至少一个碳原子，并且一个或多个氢原子可被氟或溴原子取代。

[0027] 杂烷基基团的具体例子是甲氧基、三氟甲氧基、乙氧基、正丙氧基、异丙氧基、叔丁氧基、甲氧基甲基、乙氧基甲基、 $-CH_2CH_2OH$ 、 $-CH_2OH$ 、 $-SO_2Me$ 、甲氧基乙基、1-甲氧基乙基、1-乙氧基乙基、2-甲氧基乙基或2-乙氧基乙基、甲氨基、乙氨基、丙氨基、异丙氨基、二甲氨基、二乙氨基、异丙基乙基氨基、甲氨基甲基、乙基氨基甲基、二异丙基氨基乙基、甲硫基、乙硫基、异丙硫基、烯醇醚、二甲基氨基甲基、二甲基氨基乙基、乙酰基、丙酰基、丁酰氧基、乙酰氧基、甲氧羰基、乙氧羰基、丙酰氧基、乙酰氨基或丙酰胺基、羧甲基、羧乙基或羧丙基、N-乙基-N-甲基氨基甲酰基或N-甲基氨基甲酰基。杂烷基基团的进一步例子为腈、异腈、氰酸酯、硫代氰酸酯、异氰酸酯、异硫氰酸酯和烷基腈基。

[0028] 表述环烷基指的是饱和或部分不饱和的(例如，环烯基基团)环基团，它包含一个或多个(优选1或2)环，且含有3至14个环碳原子，优选为3至10(特别为3、4、5、6或7)个环碳原子。表述环烷基进一步指一个或多个氢原子已被氟、氯、溴或碘原子，或由OH、=O、SH、=S、 $NH_2$ 、=NH、 $N_3$ 或 $NO_2$ 基团所取代的基团，因此，例如，环酮，例如，环己酮、2-环己烯酮或环戊酮。环烷基基团进一步具体例子为环丙基、环丁基、环戊基、螺[4,5]癸基、降冰片基、环己基、环戊烯基、环己二烯基、十氢萘基、双环[4.3.0]壬基、四氢化萘、环戊基环己基、氟代环己基或环己-2-烯基。

[0029] 表述杂环烷基是指如上所述的环烷基基团，且所述基团中一个或多个(优选为1、2或3)碳原子已被氧、氮、硅、硒、磷或硫原子(优选为氧、硫或氮原子)，或被 $SO$ 或 $SO_2$ 基取代。杂环烷基基团具有优选1或2个环，各环含有3至10个(特别是3、4、5、6或7个)环原子(优选选自C、O、N和S)。表述杂环烷基进一步是指被氟、氯、溴或碘原子或被OH、=O、SH、=S、 $NH_2$ 、=NH、 $N_3$ 或 $NO_2$ 基团取代的基团。例子为哌啶、脯氨酸(prolinyl)、咪唑烷基、哌嗪基、吗啉基、乌洛托品基(urotropinyl)、吡咯烷基、四氢噻吩基、四氢吡喃基、四氢呋喃基或2-吡啶基，以及内酰胺、内酯、环状酰亚胺和环状酸酐。

[0030] 表述烷基环烷基指同时含有如上文所述的环烷基以及烷基、烯基或炔基的基团，例如烷基环烷基、环烷基烷基、烷基环烯基、烯基环烷基和炔基环烷基基团。烷基环烷基基团优选含有一个环烷基基团(其中包括一个或两个环，各环具有3至10个(特别是3、4、5、6或7)个环碳原子)和一个或两个具有1或2至6个碳原子的烷基、烯基或炔基基团(特别是烷基基团)。

[0031] 表述杂烷基环烷基是指如上所述的，所述基团中一个或多个(优选为1、2或3)碳

原子已被氧、氮、硅、硒、磷或硫原子(优选为氧、硫或氮原子),或SO或SO<sub>2</sub>基取代。杂烷基环烷基基团优选含有1或2个环,各环包括3至10个(特别是3、4、5、6或7)个环原子,并含有一个或两个具有1或2至6个碳原子的烷基、烯基、炔基或杂烷基基团(特别是烷基或杂烷基基团)。所述基团的例子为烷基杂环烷基、烷基杂环烯基、烯基杂环烷基、炔基杂环烷基、杂烷基环烷基、杂烷基杂环烷基和杂烷基杂环烯基,所述的环状基团可以是饱和的或单-、双-或三-不饱和的。

[0032] 表述芳基指包含一个或多个环的芳香基团,各环具有6至14个环碳原子,优选6至10(特别是6)个环碳原子。表述芳基进一步是指被氟、氯、溴或碘原子或被OH、SH、NH<sub>2</sub>、N<sub>3</sub>或NO<sub>2</sub>基团所取代的基团。例子为苯基、萘基、联苯基、2-氟苯基、苯胺基(aniliny1)、3-硝基苯基或4-羟基苯基。

[0033] 表述“杂芳基”指包含一个或多个环的芳香基团,各环含有5至14个环原子,优选为5至10(特别是5或6或9或10)个环原子,且包含一个或多个(优选1、2、3或4)氧、氮、磷或硫原子(优选O、S或N)。表述“杂芳基”进一步是指被氟、氯、溴或碘原子或被OH、SH、NH<sub>2</sub>、N<sub>3</sub>或NO<sub>2</sub>基团取代的基团。例子为吡啶基(例如4-吡啶基)、咪唑基(例如2-咪唑基)、苯基吡咯基(例如3-苯基吡咯基)、噻唑基、异噻唑基、1,2,3-三唑基、1,2,4-三唑基、噁二唑基、噻二唑基、吡啶基、吡唑基、四唑基、吡嗪基、嘧啶基、哒嗪基、噁唑基、异噁唑基、三唑基、四唑基、异噻唑基、吡唑基、吡啶基、苯并咪唑基、苯并噁唑基、苯并异噻唑基、苯并噻唑基、哒嗪基、喹啉基、异喹啉基、吡咯基、嘌呤基、咪唑基、吡啶基、嘧啶基、2,3'-二咪唑基、吡唑基(例如3-吡唑基)和异喹啉基。

[0034] 表述芳烷基是指同时含有如上述的定义的芳基以及烷基、烯基、炔基和/或环烷基基团的基团,如,例如,芳基烷基、芳基烯基、芳基炔基、芳基环烷基、芳基环烯基、烷基芳基环烷基和烷基芳基环烯基。芳烷基的具体例子是甲苯、二甲苯、均三甲苯、苯乙烯、苄基氯、邻-氟甲苯、1H-茚、四氢化萘、二氢化萘、茚满酮、苯基环戊基、枯烯、环己基苯基、茚和二氢化茚。芳烷基基团优选含有一个或两个芳香的环体系(特别是1或2个环),每一个含有6至10个碳原子和一个或两个含有1或2到6个碳原子的烷基、烯基和/或炔基基团,和/或含5或6个环碳原子的环烷基基团。

[0035] 表述杂芳烷基指同时分别具有如上述定义的芳基或杂芳基,并具有烷基、烯基、炔基和/或杂烷基和/或环烷基和/或杂环烷基基团的基团。杂芳烷基基团优选含有一个或两个芳香的环体系(特别是1或2个环),每一个环体系含有5或6至9或10个环碳原子以及一个或两个含有1或2至6个碳原子的烷基、烯基和/或炔基基团,和/或一个或两个含有1至6个碳原子以及1、2或3个选自O、S和N的杂原子的杂烷基基团,和/或一个或两个分别含有5或6个环碳原子的环烷基基团,和/或一个或两个各含有5或6个环原子(其中包括1、2、3或4个氧、硫或氮原子)的杂环烷基基团。

[0036] 例子有芳杂烷基、芳杂环烷基、芳杂环烯基、芳基烷基杂环烷基、芳烯基杂环烷基、芳炔基杂环烷基、芳基烷基杂环烯基、杂芳基烷基、杂芳烯基、杂芳炔基、杂芳杂烷基、杂芳环烷基、杂芳环烯基、杂芳杂环烷基、杂芳杂环烯基、杂芳基烷基环烷基、杂芳基烷基杂环烯基、杂芳杂烷基环烷基、杂芳杂烷基环烯基和杂芳杂烷基杂环烷基基团,上述的环基团可以为饱和的或单-、二-或三-不饱和的。具体例子是四氢异喹啉、苯甲酰基、2-或3-乙基吡啶基、4-甲基吡啶、2-,3-或4-甲氧基苯基、4-乙氧基苯基、2-,3-或4-羧基苯基烷基

基团。

[0037] 如前所述,表述环烷基、杂环烷基、烷基环烷基、杂烷基环烷基、芳基、杂芳基、芳烷基和杂芳烷基也指被氟、氯、溴或碘原子或被OH、=O、SH、=S、NH<sub>2</sub>、=NH、N<sub>3</sub>或NO<sub>2</sub>基团取代的基团。

[0038] 表述“任选取代”指基团可以被氟、氯、溴或碘原子或被OH、=O、SH、=S、NH<sub>2</sub>、=NH、N<sub>3</sub>或NO<sub>2</sub>基团任选地取代。上述表述进一步指可被一个、两个、三个或更多个未取代的C<sub>1</sub>-C<sub>10</sub>烷基、C<sub>2</sub>-C<sub>10</sub>烯基、C<sub>2</sub>-C<sub>10</sub>炔基、C<sub>1</sub>-C<sub>10</sub>杂烷基、C<sub>3</sub>-C<sub>18</sub>环烷基、C<sub>2</sub>-C<sub>17</sub>杂环烷基、C<sub>4</sub>-C<sub>20</sub>烷基环烷基、C<sub>2</sub>-C<sub>19</sub>杂烷基环烷基、C<sub>6</sub>-C<sub>18</sub>芳基、C<sub>1</sub>-C<sub>17</sub>杂芳基、C<sub>7</sub>-C<sub>20</sub>芳烷基或C<sub>2</sub>-C<sub>19</sub>杂芳烷基基团所取代的基团。上述表述进一步指被一个、两个、三个或更多个未取代的C<sub>1</sub>-C<sub>6</sub>烷基、C<sub>2</sub>-C<sub>6</sub>烯基、C<sub>2</sub>-C<sub>6</sub>炔基、C<sub>1</sub>-C<sub>6</sub>杂烷基、C<sub>3</sub>-C<sub>10</sub>环烷基、C<sub>2</sub>-C<sub>9</sub>杂环烷基、C<sub>7</sub>-C<sub>12</sub>烷基环烷基、C<sub>2</sub>-C<sub>11</sub>杂烷基环烷基、C<sub>6</sub>-C<sub>10</sub>芳基、C<sub>1</sub>-C<sub>9</sub>杂芳基、C<sub>7</sub>-C<sub>12</sub>芳烷基或C<sub>2</sub>-C<sub>11</sub>杂芳烷基基团所取代的基团。

[0039] 特别优选地,在基团Ar<sup>1</sup>、Ar<sup>2</sup>、Ar<sup>3</sup>、Ar<sup>4</sup>和Ar<sup>5</sup>中,表述“任选取代的”指基团被一个、两个、三个独立选自下组的基团所取代:卤素原子、羟基、如式-O-C<sub>1-6</sub>烷基(例如,-O-C<sub>1-6</sub>烷基如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、-NH<sub>2</sub>、-NR<sup>5a</sup>R<sup>6a</sup>(其中R<sup>5a</sup>和R<sup>6a</sup>各自独立地为氢原子或烷基基团如C<sub>1-6</sub>烷基基团)、-SO<sub>2</sub>NH<sub>2</sub>、-CONH<sub>2</sub>、-CN、-烷基(例如-C<sub>1-6</sub>烷基、-CF<sub>3</sub>)、-SH、-S-烷基(例如,-S-C<sub>1-6</sub>烷基)。

[0040] 最优选地,在基团Ar<sup>1</sup>、Ar<sup>2</sup>、Ar<sup>3</sup>、Ar<sup>4</sup>和Ar<sup>5</sup>中,表述“任选取代的”指基团被一个、两个、三个独立选自下组的基团所取代:F、Cl、羟基、如式-O-C<sub>1-6</sub>烷基(特别是-O-C<sub>1-4</sub>烷基如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、和-C<sub>1-6</sub>烷基(例如-C<sub>1-4</sub>烷基如-CH<sub>3</sub>或-CF<sub>3</sub>)。

[0041] 特别优选地,在基团Ar<sup>6</sup>中,表述“任选取代的”指基团被一个、两个、三个独立选自下组的基团所取代:卤素原子、羟基、如式-O-烷基(例如,-O-C<sub>1-6</sub>烷基如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、-NH<sub>2</sub>、-NR<sup>5a</sup>R<sup>6a</sup>(其中R<sup>5a</sup>和R<sup>6a</sup>各自独立地为氢原子或烷基基团如C<sub>1-6</sub>烷基基团)、-SO<sub>2</sub>NH<sub>2</sub>、-CONH<sub>2</sub>、-CN、-烷基(例如-C<sub>1-6</sub>烷基、-CF<sub>3</sub>)、-SH、-S-烷基(例如,-S-C<sub>1-6</sub>烷基),和NO<sub>2</sub>。

[0042] 最优选地,在基团Ar<sup>6</sup>中,表述“任选取代的”指基团被一个、两个、三个选自下组的基团所取代:F、Cl、羟基、-NH<sub>2</sub>、-NO<sub>2</sub>、如式-O-C<sub>1-6</sub>烷基(特别是-O-C<sub>1-4</sub>烷基如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、和-C<sub>1-6</sub>烷基(例如-C<sub>1-4</sub>烷基如-CH<sub>3</sub>或-CF<sub>3</sub>)。

[0043] 术语卤素指氟、氯、溴或碘。

[0044] 根据一个优选的实施例,所有本文所述的烷基、烯基、炔基、杂烷基、芳基、杂芳基、环烷基、杂环烷基、烷基环烷基、杂烷基环烷基、芳烷基和杂芳烷基基团可以各自独立地为任选取代的。

[0045] 当芳基、杂芳基、环烷基、烷基环烷基、杂烷基环烷基、杂环烷基、芳烷基或杂芳烷基基团含有多于一个环时,所述的环可以是通过单键或双键互相连接的,或所述的环可以是环合的。

[0046] 由于它们的取代,式(I)的化合物可含有一个或多个手性中心。因此,本发明既包括所有的纯的对映异构体,也包括所有的纯的非对映异构,及其以任意混合比例混合的混合物。此外,本发明还包括通式(I)的化合物的所有顺式/反式异构体及其混合物。本发明



还包括式(I)的化合物的所有互变异构形式。

[0047] 优选地,当 $Ar^4$ 为无, $L^3$ 同样为无。

[0048] 更优选地,当 $Ar^5$ 为无, $L^4$ 同样为无。

[0049] 优选地, $Ar^1$ 为任选取代的1,4-亚苯基,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基。

[0050] 更优选地, $Ar^1$ 为任选取代的1,4-亚苯基。

[0051] 优选地, $Ar^2$ 为任选取代的1,4-亚苯基,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基;

[0052] 更优选地, $Ar^2$ 为任选取代的1,4-亚苯基。

[0053] 优选地, $Ar^2$ 为任选取代的1,4-亚苯基,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基。

[0054] 更优选地, $Ar^3$ 为任选取代的1,4-亚苯基。

[0055] 优选地, $Ar^4$ 为任选取代的1,4-亚苯基,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基。

[0056] 更优选地, $Ar^4$ 为任选取代的1,4-亚苯基。

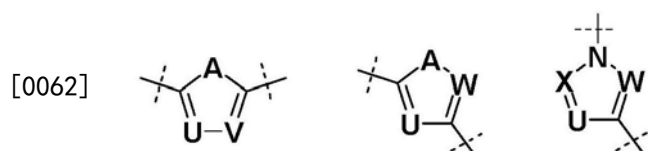
[0057] 优选地, $Ar^5$ 为任选取代的1,4-亚苯基,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基;

[0058] 更优选地, $Ar^5$ 为任选取代的1,4-亚苯基。

[0059] 更优选地, $Ar^4$ 为无。

[0060] 更优选地, $Ar^5$ 为无。

[0061] 术语“具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的1,3-亚杂芳基”特别优选地指以下的基团之一:



[0063] 其中A为O、S或NH;U为N或CH;V为N或CH;W为N或CH;且X为N或CH。

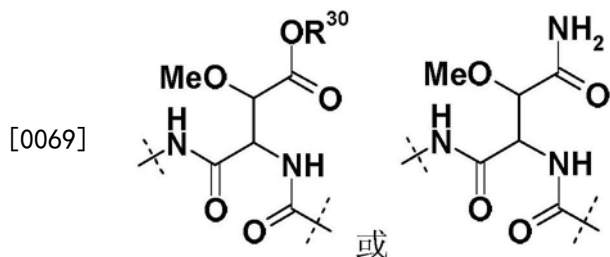
[0064] 更优选地, $L^1$ 为如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>-所示的基团,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的亚杂芳基,其中R<sup>3</sup>和R<sup>4</sup>各自独立地为C<sub>1-6</sub>烷基基团。

[0065] 更优选地, $L^2$ 为如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>-所示的基团,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的亚杂芳基,其中R<sup>3</sup>和R<sup>4</sup>各自独立地为C<sub>1-6</sub>烷基基团。

[0066] 更优选地, $L^3$ 为无,或如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>-所示的基团,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的亚杂芳基,其中R<sup>3</sup>和R<sup>4</sup>各自独立地为C<sub>1-6</sub>烷基基团。

[0067] 更优选地, $L^4$ 为无,或如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-、-NHSO<sub>2</sub>-、-CH=CH-、-CR<sup>3</sup>=CR<sup>4</sup>-所示的基团,或任选取代的具有5个环原子(其中包括1、2或3个选自下组的杂原子:氧、硫和氮)的亚杂芳基,其中R<sup>3</sup>和R<sup>4</sup>各自独立地为C<sub>1-6</sub>烷基基团。

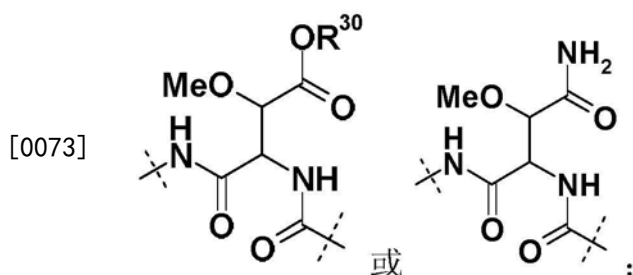
[0068] 更优选地,  $L^1$  为  $\text{NHCOR}^{30}$  (其中氮原子与  $\text{Ar}^2$  相连接), 或如下式所示的基团:



[0070] (其中  $\text{NH}$  基团与  $\text{Ar}^1$  连接), 其中  $\text{R}^{30}$  为氢原子或  $\text{C}_{1-3}$  烷基基团。

[0071] 特别优选地,  $L^1$  为  $\text{NHCOR}^{30}$  (其中氮原子与  $\text{Ar}^1$  相连接)。

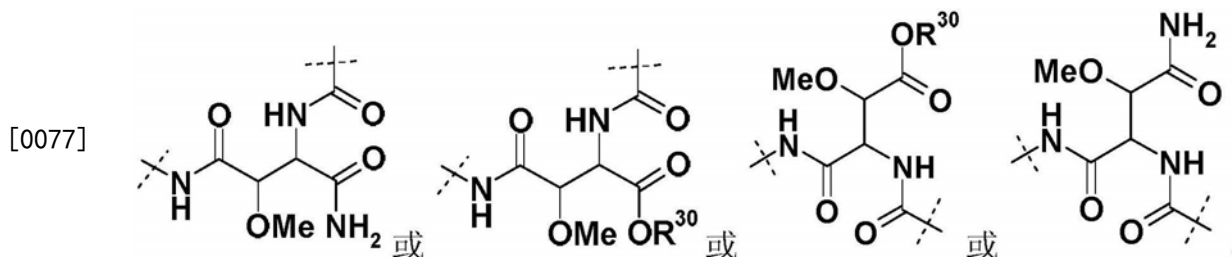
[0072] 更为优选地,  $L^2$  为  $\text{NHCOR}^{30}$  (其中氮原子与  $\text{Ar}^2$  相连接), 或如下式所示的基团:



[0074] (其中  $\text{NH}$  基团与  $\text{Ar}^2$  连接), 其中  $\text{R}^{30}$  为氢原子或  $\text{C}_{1-3}$  烷基基团。

[0075] 特别优选地,  $L^2$  为  $\text{NHCOR}^{30}$  (其中氮原子与  $\text{Ar}^1$  相连接)。

[0076] 更加优选地,  $L^3$  为无或如下式所示的基团:



[0078] (其中  $\text{NH}$  基团与  $\text{Ar}^3$  连接), 其中  $\text{R}^{30}$  为氢原子或  $\text{C}_{1-3}$  烷基基团。

[0079] 更优选地,  $L^4$  为无或  $\text{NHCOR}^{30}$  (其中氮原子与  $\text{Ar}^4$  相连接)。

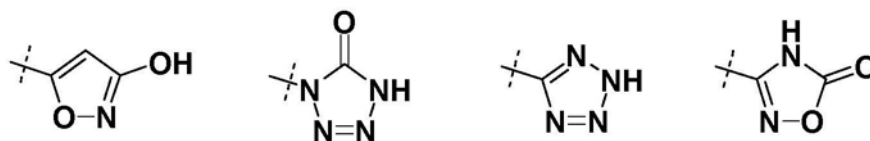
[0080] 更为优选地,  $\text{R}^{30}$  为氢原子。

[0081] 更优选地,  $\text{R}^1$  为氢原子、卤素原子, 或如下式  $-\text{OH}$ 、 $-\text{NH}_2$ 、 $-\text{COOH}$ 、 $-\text{SO}_2\text{NH}_2$ 、 $-\text{CONH}_2$ 、 $-\text{NO}_2$ 、 $-\text{CN}$ 、 $-\text{烷基}$  (例如  $-\text{CF}_3$ )、 $-\text{O-烷基}$ 、 $-\text{O-CO-烷基}$ 、 $-\text{NH-烷基}$ 、 $-\text{NH-CO-烷基}$  所示的基团, 或任选取代的具有 5 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂芳基, 或任选取代的具有 5 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂环烷基基团。

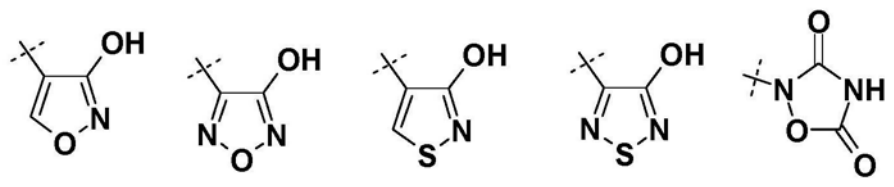
[0082] 更优选地,  $\text{R}^2$  为氢原子、卤素原子, 或如下式  $-\text{OH}$ 、 $-\text{NH}_2$ 、 $-\text{COOH}$ 、 $-\text{SO}_2\text{NH}_2$ 、 $-\text{CONH}_2$ 、 $-\text{NO}_2$ 、 $-\text{CN}$ 、 $-\text{烷基}$  (例如  $-\text{CF}_3$ )、 $-\text{O-烷基}$ 、 $-\text{O-CO-烷基}$ 、 $-\text{NH-烷基}$ 、 $-\text{NH-CO-烷基}$  所示的基团, 或任选取代的具有 5 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂芳基, 或任选取代的具有 5 个环原子 (其中包括 1、2、3 或 4 个选自下组的杂原子: 氧、硫和氮) 的杂环烷基基团。

[0083] 作为基团  $\text{R}^1$  和  $\text{R}^2$  的、优选的任选取代的具有 5 个环原子 (其中包括 1、2、3 或 4 个选自

下组的杂原子:氧、硫和氮)的杂芳基和任选取代的具有5个环原子(其中包括1、2、3或4个选自下组的杂原子:氧、硫和氮)的杂环烷基基团的例子,是羧酸的电子等排体,例如下式基团:



[0084]



[0085] 所有上述基团可以任选地被进一步取代。

[0086] 特别优选地,  $R^1$  为如下式  $-NH_2$ 、 $-NO_2$ 、 $COOR^{11}$ 、或  $-CONR^{12}R^{13}$  所示的基团; 其中  $R^{11}$ 、 $R^{12}$  和  $R^{13}$  各自独立地为氢原子或  $C_{1-6}$  烷基基团; 更优选地,  $R^1$  为如式  $-COOH$  所示的基团。

[0087] 进一步特别优选地,  $R^2$  为如下式  $-NH_2$ 、 $-NO_2$ 、 $COOR^{11a}$ 、或  $-CONR^{12a}R^{13a}$  所示的基团; 其中  $R^{11a}$ 、 $R^{12a}$  和  $R^{13a}$  各自独立地为氢原子或  $C_{1-6}$  烷基基团; 更优选地,  $R^2$  为如式  $-NH_2$  或  $-NO_2$  所示的基团。

[0088] 进一步特别优选地,  $R^1$  为被羟基取代的、具有5个环原子(其中包括1、2、3或4个选自下组的杂原子:氧、硫和氮)的亚杂芳基。

[0089] 进一步特别优选地,  $R^2$  为被羟基取代的、具有5个环原子(其中包括1、2、3或4个选自下组的杂原子:氧、硫和氮)的亚杂芳基。

[0090] 特别优选的是如式 (I) 所示的化合物, 或其药学上可接受的盐、溶剂化物或水合物, 或其药学上可接受的制剂

[0091]  $R^1-Ar^1-L^1-Ar^2-L^2-Ar^3-L^3-Ar^4-L^4-Ar^5-R^2$

[0092] (I)

[0093] 其中,

[0094]  $Ar^1$  为任选取代的1,4-亚苯基;

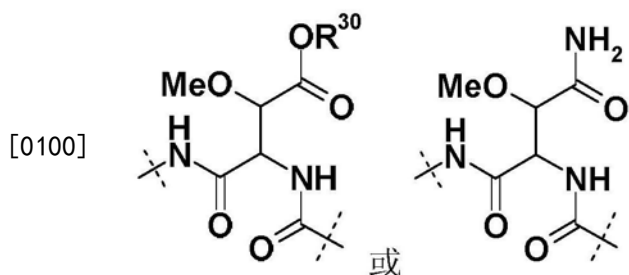
[0095]  $Ar^2$  为任选取代的1,4-亚苯基;

[0096]  $Ar^3$  为任选取代的1,4-亚苯基;

[0097]  $Ar^4$  为无或任选取代的1,4-亚苯基;

[0098]  $Ar^5$  为无或任选取代的1,4-亚苯基;

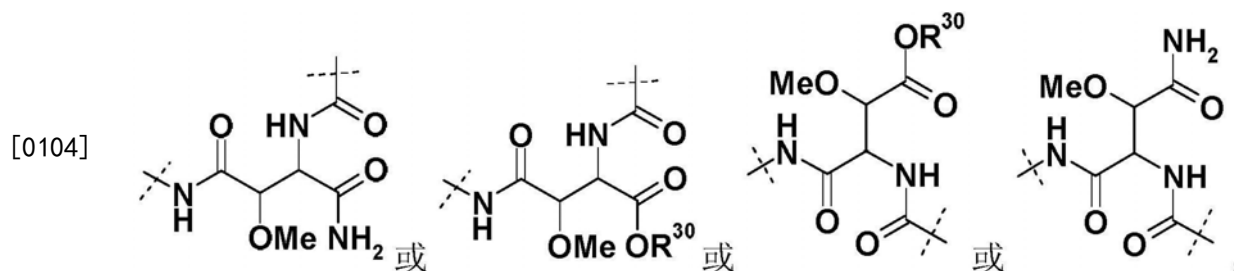
[0099]  $L^1$  为如式  $-CONH-$ 、 $-NHCO-$ 、 $-SO_2NH-$  或  $-NHSO_2-$  所示的基团, 或如下式所示的基团:



[0101] (其中NH基团与Ar<sup>1</sup>相连)；

[0102] L<sup>2</sup>为如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-或-NHSO<sub>2</sub>-所示的基团；

[0103] L<sup>3</sup>为无,或如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-或-NHSO<sub>2</sub>-所示的基团,或如下式所示的基团:



[0105] (其中NH基团与Ar<sup>3</sup>相连)；

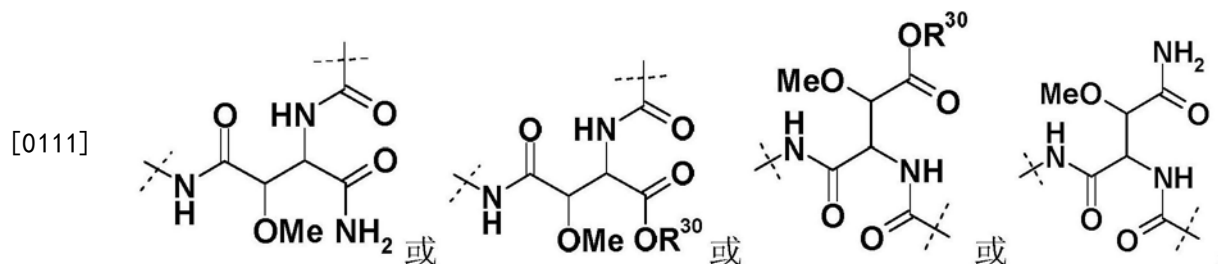
[0106] L<sup>4</sup>为无,或如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-或-NHSO<sub>2</sub>-所示的基团；

[0107] R<sup>30</sup>为氢原子,或C<sub>1-3</sub>烷基基团(特别优选地为氢原子)；

[0108] R<sup>1</sup>为如下式-NH<sub>2</sub>、-NO<sub>2</sub>、COOR<sup>11</sup>、或-CONR<sup>12</sup>R<sup>13</sup>所示的基团;其中R<sup>11</sup>、R<sup>12</sup>和R<sup>13</sup>各自独立地为氢原子或C<sub>1-6</sub>烷基基团(特别优选地,R<sup>1</sup>为如式-COOH所示的基团);和

[0109] R<sup>2</sup>为如下式-NH<sub>2</sub>、-NO<sub>2</sub>、COOR<sup>11a</sup>、或-CONR<sup>12a</sup>R<sup>13a</sup>所示的基团;其中R<sup>11a</sup>、R<sup>12a</sup>和R<sup>13a</sup>各自独立地为氢原子或C<sub>1-6</sub>烷基基团(特别选地,R<sup>2</sup>为如式-NH<sub>2</sub>或-NO<sub>2</sub>所示的基团)。

[0110] 其中,优选地,L<sup>1</sup>为如式-CONH-、-NHCO-、-SO<sub>2</sub>NH-或-NHSO<sub>2</sub>-所示的基团,且L<sup>3</sup>为无,或如下式所示的基团:



[0112] (其中NH基团与Ar<sup>3</sup>相连)。

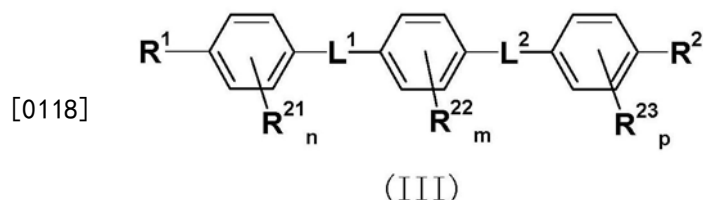
[0113] 更加优选的为如式(II)所示的化合物

[0114] R<sup>1</sup>-Ar<sup>1</sup>-L<sup>1</sup>-Ar<sup>2</sup>-L<sup>2</sup>-Ar<sup>3</sup>-R<sup>2</sup>

[0115] (II)

[0116] 其中Ar<sup>1</sup>、Ar<sup>2</sup>、Ar<sup>3</sup>、L<sup>1</sup>、L<sup>2</sup>、R<sup>1</sup>以及R<sup>2</sup>如上文所述。

[0117] 更加优选的为如式(III)所示的化合物



[0119] 其中,

[0120] n为0、1、2、3或4;

[0121] m为0、1、2、3或4;

[0122] p为0、1、2、3或4；

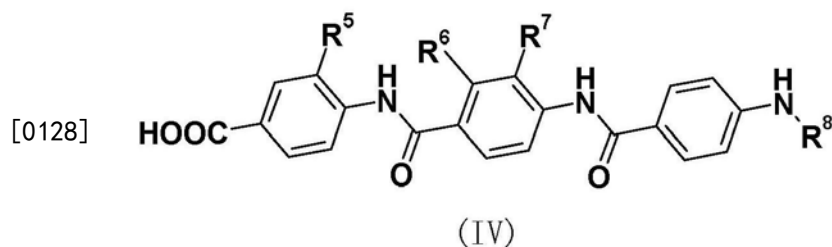
[0123] 基团 $R^{21}$ 各自独立地选自卤素原子、羟基、如式-O-烷基(例如,-O-C<sub>1-6</sub>烷基 如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、-NH<sub>2</sub>、-NR<sup>5a</sup>R<sup>6a</sup>(其中R<sup>5a</sup>和R<sup>6a</sup>各自独立地为氢原子或烷基基团如C<sub>1-6</sub>烷基基团)、-SO<sub>2</sub>NH<sub>2</sub>、-CONH<sub>2</sub>、-CN、-烷基(例如-C<sub>1-6</sub>烷基、-CF<sub>3</sub>)、-SH、-S-烷基(例如,-S-C<sub>1-6</sub>烷基)；

[0124] 基团 $R^{22}$ 各自独立地选自卤素原子、羟基、如式-O-烷基(例如,-O-C<sub>1-6</sub>烷基 如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、-NH<sub>2</sub>、-NR<sup>5a</sup>R<sup>6a</sup>(其中R<sup>5a</sup>和R<sup>6a</sup>各自独立地为氢原子或烷基基团如C<sub>1-6</sub>烷基基团)、-SO<sub>2</sub>NH<sub>2</sub>、-CONH<sub>2</sub>、-CN、-烷基(例如-C<sub>1-6</sub>烷基、-CF<sub>3</sub>)、-SH、-S-烷基(例如,-S-C<sub>1-6</sub>烷基)；

[0125] 基团 $R^{23}$ 各自独立地选自卤素原子、羟基、如式-O-烷基(例如,-O-C<sub>1-6</sub>烷基 如-OMe、-OEt、-O-nPr、-O-iPr、-O-nBu、-O-iBu或-O-tBu)所示的基团、-NH<sub>2</sub>、-NR<sup>5a</sup>R<sup>6a</sup>(其中R<sup>5a</sup>和R<sup>6a</sup>各自独立地为氢原子或烷基基团如C<sub>1-6</sub>烷基基团)、-SO<sub>2</sub>NH<sub>2</sub>、-CONH<sub>2</sub>、-CN、-烷基(例如-C<sub>1-6</sub>烷基、-CF<sub>3</sub>)、-SH、-S-烷基(例如,-S-C<sub>1-6</sub>烷基)；和

[0126]  $R^1$ 、 $R^2$ 、 $L^1$ 和 $L^2$ 如上所述。

[0127] 更加优选的为如式(IV)所示的化合物



[0129] 其中，

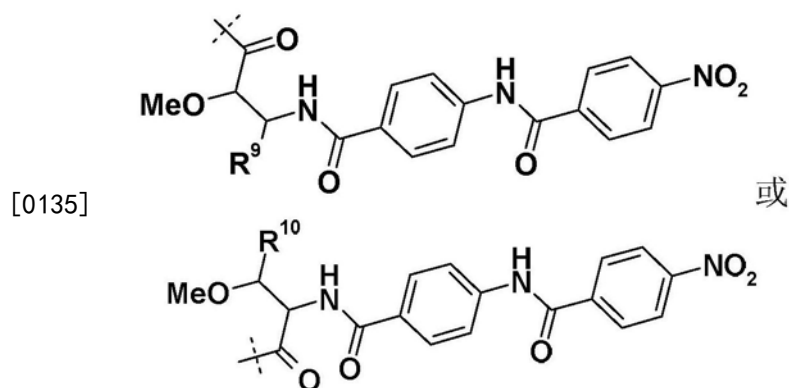
[0130]  $R^5$ 为如式-O-C<sub>1-6</sub>烷基所示的基团；

[0131]  $R^6$ 为羟基基团；

[0132]  $R^7$ 为如式-O-C<sub>1-6</sub>烷基所示的基团；和

[0133]  $R^8$ 为氢原子、烷基、烯基、炔基、杂烷基、环烷基、杂环烷基、烷基环烷基、杂烷基环烷基、芳基、杂芳基、芳烷基或杂芳烷基基团。

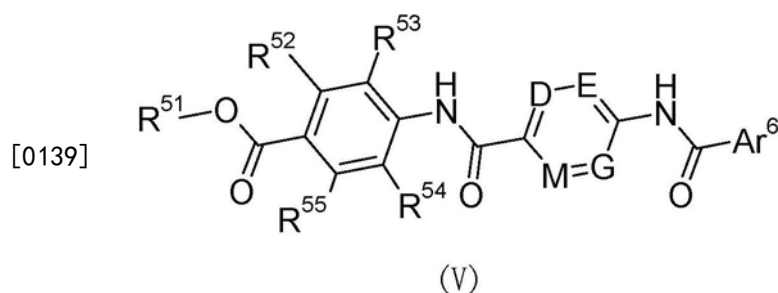
[0134] 更加优选地， $R^8$ 为氢原子或如下式所示的基团：



[0136] 其中 $R^9$ 为COOH或CONH<sub>2</sub>，且 $R^{10}$ 为COOH或CONH<sub>2</sub>。

[0137] 更加优选地， $R^5$ 为如式-O-C<sub>1-4</sub>烷基所示的基团，且 $R^7$ 为如式-O-C<sub>1-4</sub>烷基所示的基团。

[0138] 更加优选的为如式 (V) 所示的化合物,或其药学上可接受的盐、溶剂化物或水合物,或其药学上可接受的制剂:



[0140] 其中,

[0141]  $R^{51}$ 为氢原子,或 $C_{1-6}$ 烷基基团;

[0142]  $R^{52}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0143]  $R^{53}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0144]  $R^{54}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0145]  $R^{55}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0146] D为N或 $CR^{56}$ ;

[0147] E为N或 $CR^{57}$ ;

[0148] G为N或 $CR^{58}$ ;

[0149] M为N或 $CR^{59}$ ;

[0150]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0151]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0152]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;

[0153]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式- $O-C_{1-6}$ 烷基的基团;和

[0154]  $Ar^6$ 为(被一个、两个或更多取代基如 $R^2$ 、 $R^8$ 或 $NHR^8$ )任选取代的苯基、或(被一个、两个或更多取代基如 $R^2$ 、 $R^8$ 或 $NHR^8$ )任选取代的具有5个或6个环原子的杂芳基,其中,所述的环原子中具有1、2、3或4个选自氧、硫和氮的杂原子。

[0155] 特别优选的为如式 (V) 所示的化合物,其中:

[0156]  $R^{51}$ 为氢原子,或 $C_{1-4}$ 烷基基团;

[0157]  $R^{52}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

[0158]  $R^{53}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

[0159]  $R^{54}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

[0160]  $R^{55}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

[0161] D为N或 $CR^{56}$ ;

[0162] E为N或 $CR^{57}$ ;

[0163] G为N或 $CR^{58}$ ;

[0164] M为N或 $CR^{59}$ ;

[0165]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

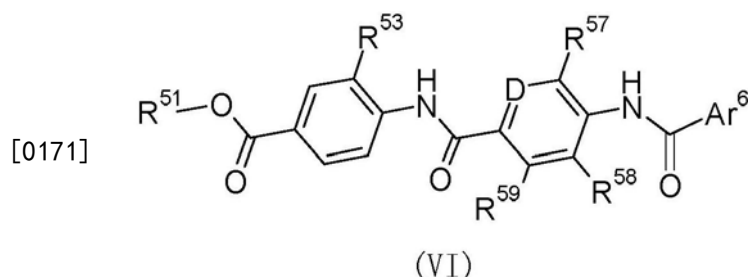
[0166]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;

[0167]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式- $O-C_{1-4}$ 烷基的基团;和

[0168]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团。

[0169] 特别优选地，D、E、G和M中仅有一个或两个（特别是仅有一个）为N。

[0170] 更加优选的为如式(VI)所示的化合物，或其药学上可接受的盐、溶剂化物或水合物，或其药学上可接受的制剂：



[0172] 其中，

[0173]  $R^{51}$ 为氢原子，或 $C_{1-6}$ 烷基基团；

[0174]  $R^{53}$ 为F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团（特别优选地 为如式-O- $C_{1-6}$ 烷基的基团）；

[0175] D为N或 $CR^{56}$ ；

[0176]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；

[0177]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；

[0178]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；

[0179]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；和

[0180]  $Ar^6$ 为（被一个、两个或更多取代基如 $R^2$ 、 $R^8$ 或 $NHR^8$ ）任选取代的苯基，或（被一个、两个或更多取代基如 $R^2$ 、 $R^8$ 或 $NHR^8$ ）任选取代的具有5个或6个环原子的杂芳基，其中，所述的环原子中具有1、2、3或4个选自氧、硫和氮的杂原子。

[0181] 特别优选的为如式(VI)所示的化合物，其中：

[0182]  $R^{51}$ 为氢原子，或 $C_{1-4}$ 烷基基团；

[0183]  $R^{53}$ 为F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团（特别优选地 为如式-O- $C_{1-4}$ 烷基的基团）；

[0184] D为N或 $CR^{56}$ ；

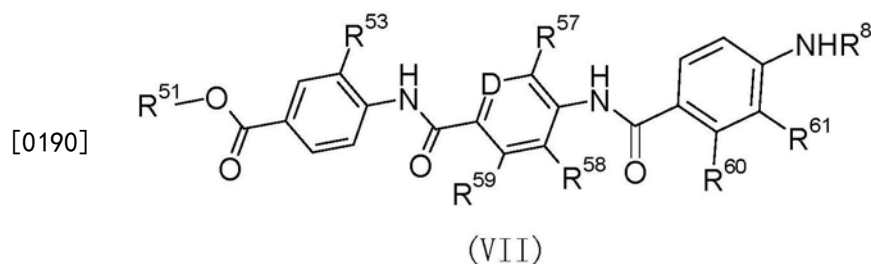
[0185]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；

[0186]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；

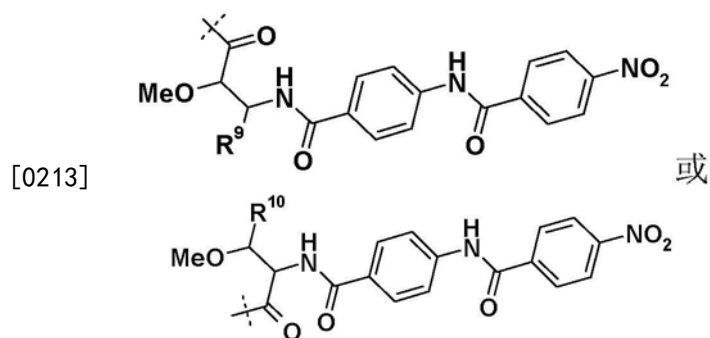
[0187]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；和

[0188]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团。

[0189] 更加优选的为如式(VII)所示的化合物，或其药学上可接受的盐、溶剂化物或水合物，或其药学上可接受的制剂：



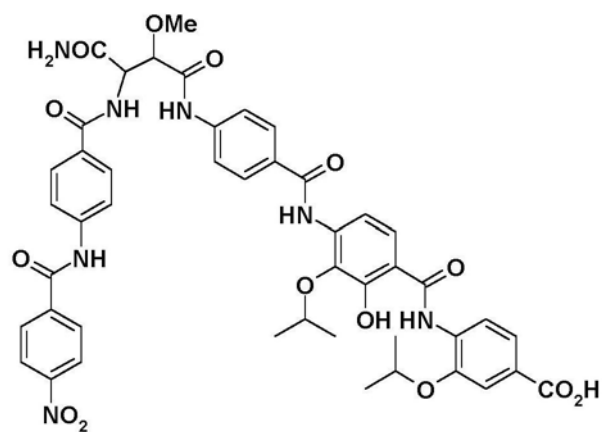
- [0191] 其中，
- [0192]  $R^{51}$ 为氢原子，或 $C_{1-6}$ 烷基基团；
- [0193]  $R^{53}$ 为F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团（特别优选地为如式-O- $C_{1-6}$ 烷基的基团）；
- [0194] D为N或 $CR^{56}$ ；
- [0195]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；
- [0196]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；
- [0197]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；
- [0198]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；
- [0199]  $R^{60}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；
- [0200]  $R^{61}$ 为氢原子、F、Cl、羟基基团、 $C_{1-6}$ 烷基基团或如式-O- $C_{1-6}$ 烷基的基团；和
- [0201]  $R^8$ 为氢原子、烷基、烯基、炔基、杂烷基、环烷基、杂环烷基、烷基环烷基、杂烷基环烷基、芳基、杂芳基、芳烷基或杂芳烷基基团。
- [0202] 特别优选的为如式(VII)所示的化合物，其中：
- [0203]  $R^{51}$ 为氢原子，或 $C_{1-4}$ 烷基基团；
- [0204]  $R^{53}$ 为F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团（特别优选地为如式-O- $C_{1-4}$ 烷基的基团）；
- [0205] D为N或 $CR^{56}$ ；
- [0206]  $R^{56}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；
- [0207]  $R^{57}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；
- [0208]  $R^{58}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；
- [0209]  $R^{59}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；
- [0210]  $R^{60}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团；和
- [0211]  $R^{61}$ 为氢原子、F、Cl、羟基基团、 $C_{1-4}$ 烷基基团或如式-O- $C_{1-4}$ 烷基的基团。
- [0212] 更加优选地， $R^8$ 为氢原子或如下式所示的基团：



- [0214] 其中 $R^9$ 为 $COOH$ 或 $CONH_2$ ，且 $R^{10}$ 为 $COOH$ 或 $CONH_2$ 。
- [0215] 特别优选的为以下化合物：

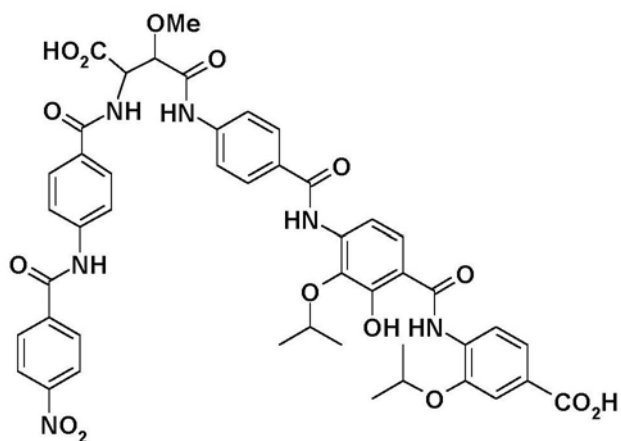


[0216]



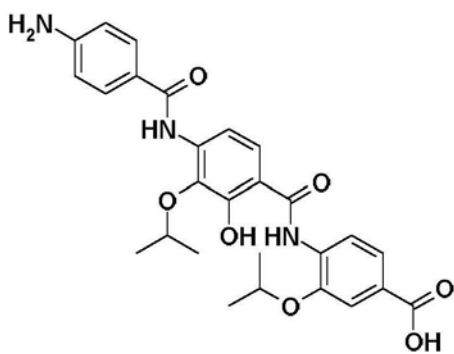
[0217] 孢囊菌酰胺A (1) ;

[0218]



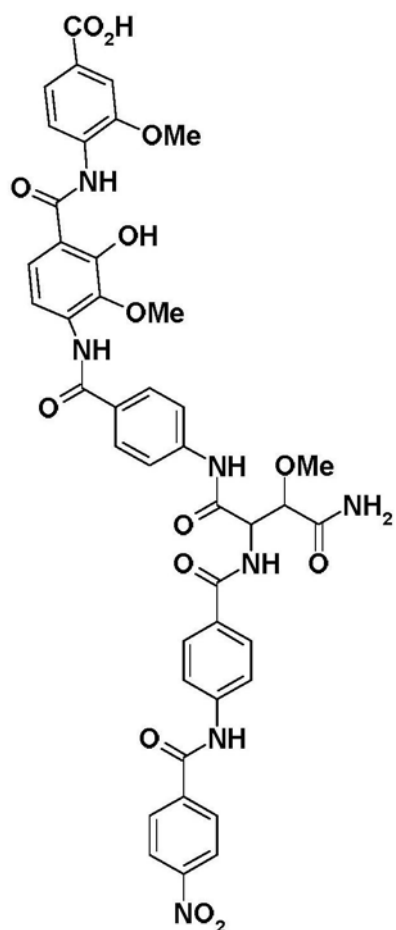
[0219] 孢囊菌酰胺B (2) ;

[0220]



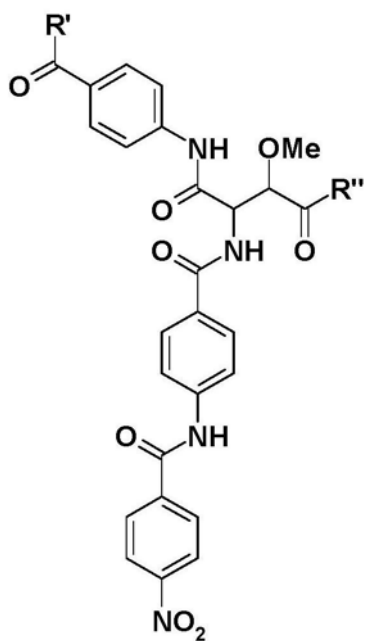
[0221] 孢囊菌酰胺C (3) ;

[0222]



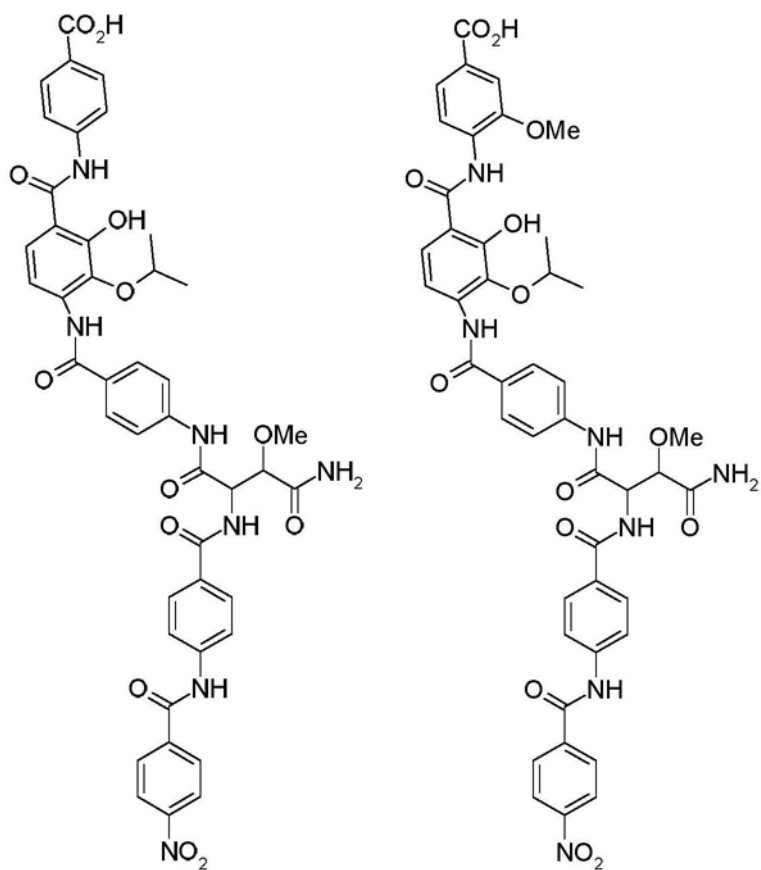
[0223] 孢囊菌酰胺D (4) ; 和

[0224]

[0225] 孢囊菌酰胺E (5) (R' 为NH<sub>2</sub>或OH且R'' 为NH<sub>2</sub>或OH)。

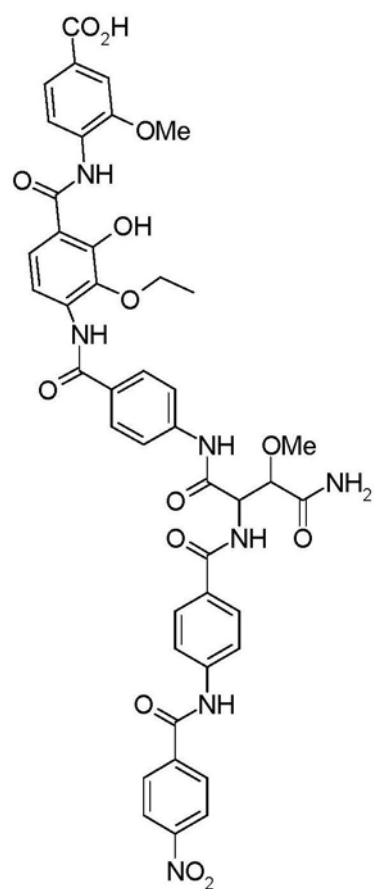
[0226] 此外,特别优选的为以下化合物:

[0227]



[0228] 孢囊菌酰胺F (6) ,孢囊菌酰胺G (7) ;和

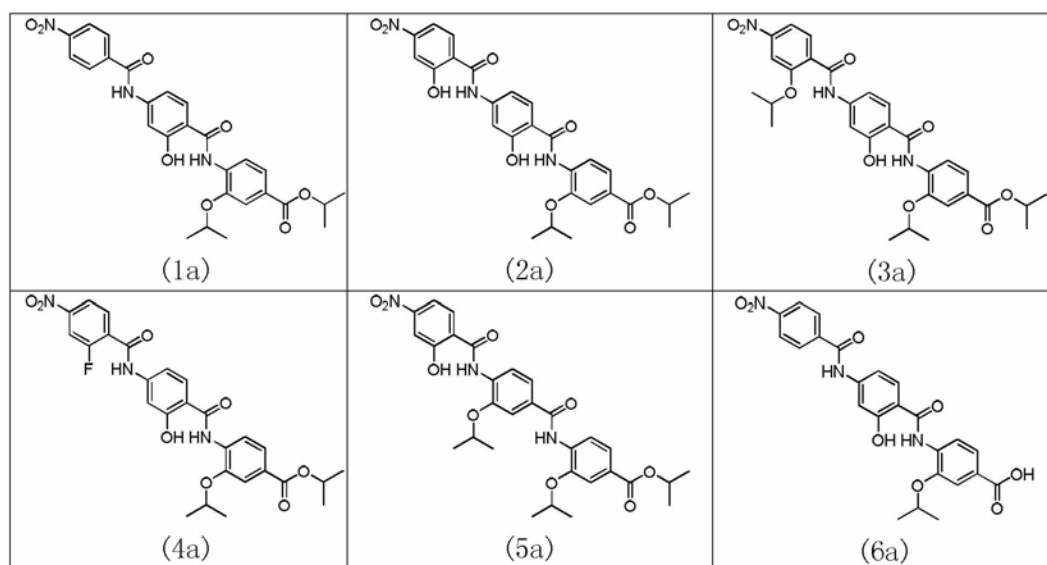
[0229]



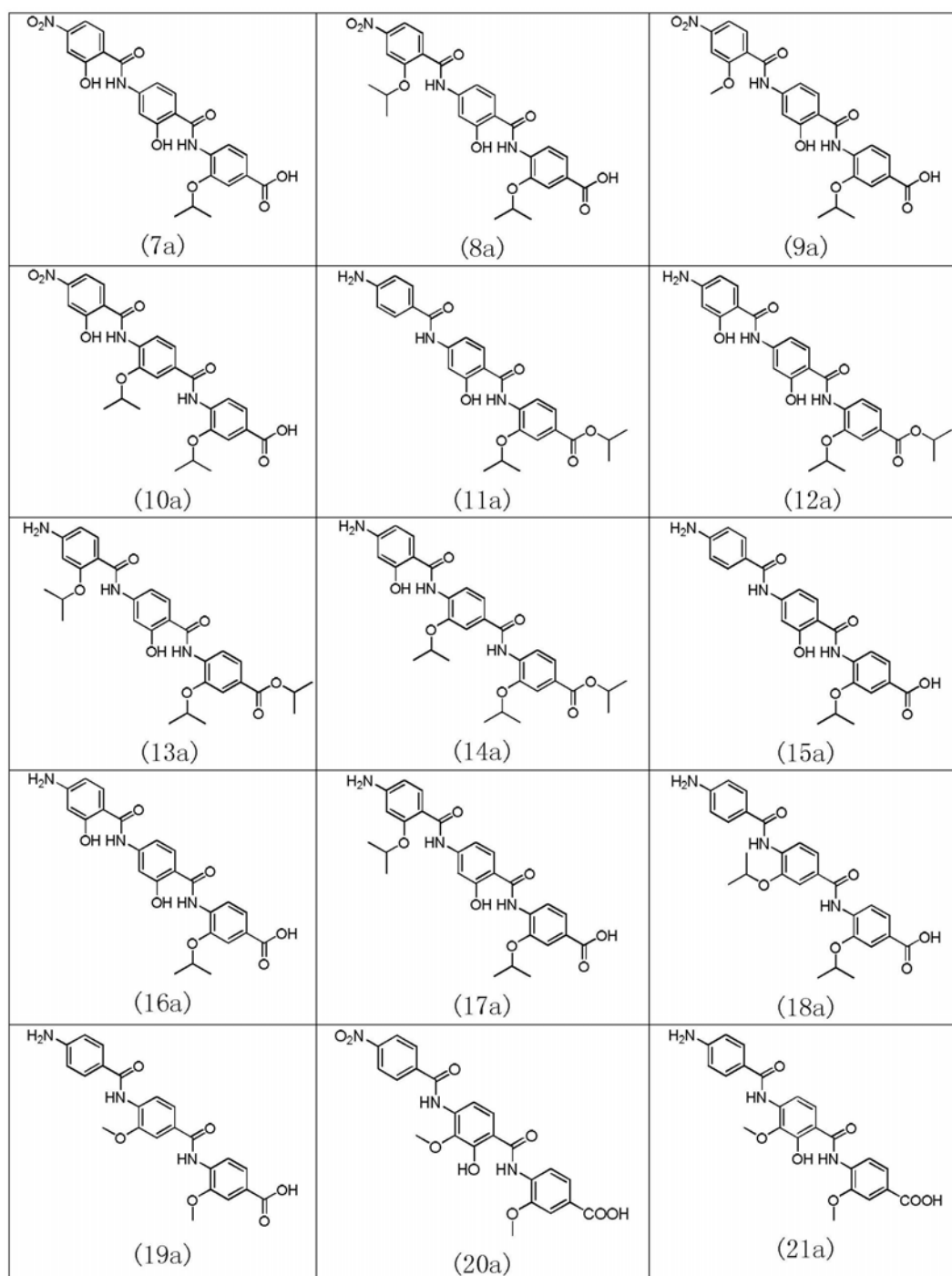
[0230] 孢囊菌酰胺H(8)。

[0231] 此外,特别优选地为以下化合物:

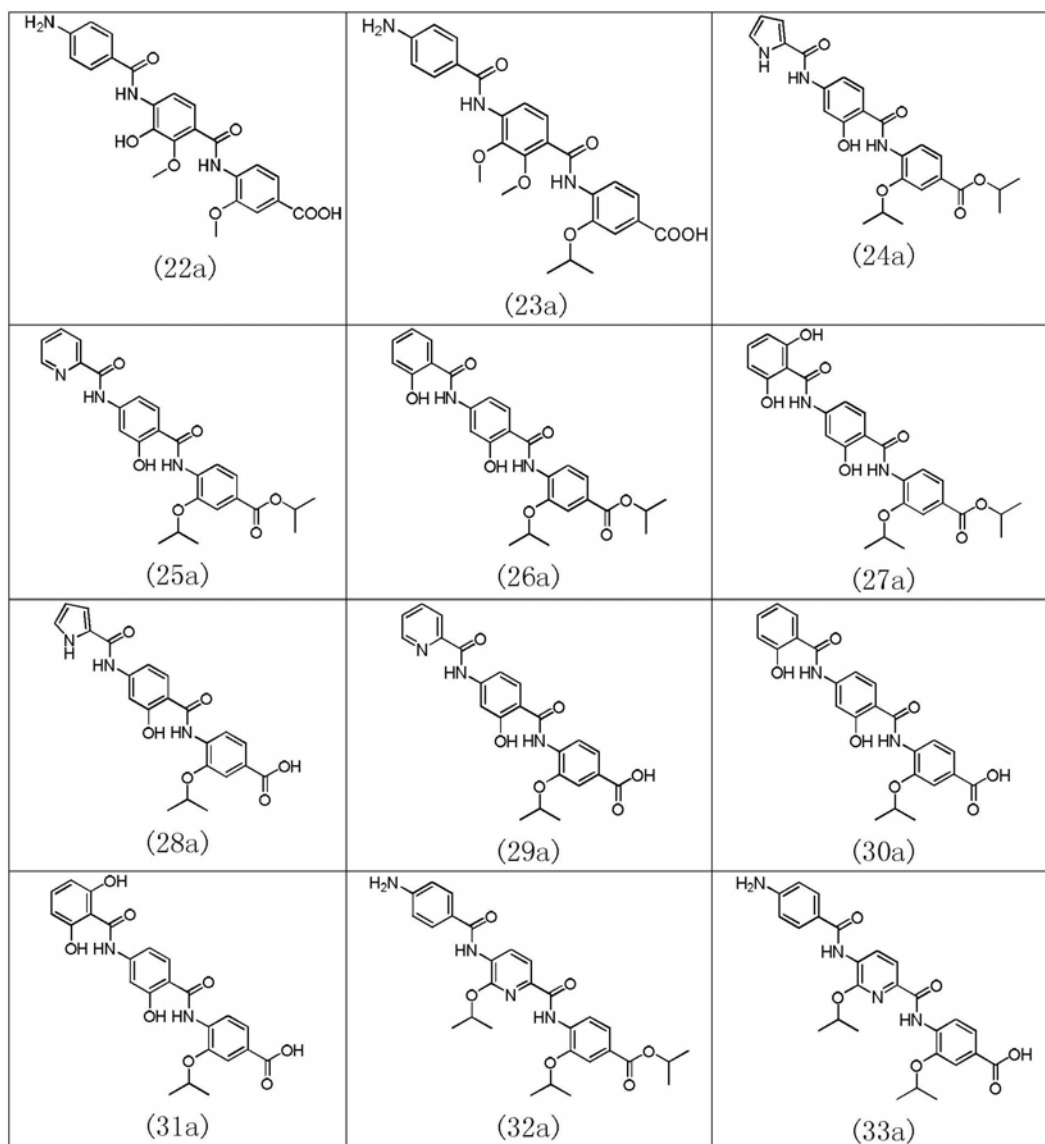
[0232]



[0233]



[0234]



[0235] 本发明还提供了含有本发明所述的一种或多种化合物或其药学上可接受的盐、溶剂合物或水合物,以及任选的一种或多种载体物质和/或一种或多种佐剂的药物组合物。

[0236] 本发明还提供了一种所述的化合物或药物组合物用于治疗 and/或预防细菌感染的用途,特别是由于大肠杆菌、绿脓杆菌、鲍曼不动杆菌或其它革兰氏阴性细菌,以及革兰氏阳性细菌导致的感染。

[0237] 更加优选地,本发明提供了化合物用于治疗 and/或预防细菌感染,特别是由绿脓杆菌和其他革兰氏阴性菌引起的感染。

[0238] 本发明的另一目的是提供一种如本文所述的化合物或药物组合物用于制备治疗和/或预防细菌感染的药物组合物的用途,特别是由选定的革兰氏阴性细菌和革兰氏阳性细菌所引起的感染。

[0239] 具有足够碱性的化合物的药理学上可接受的盐的实例,为生理学上可接受的无机酸如盐酸、氢溴酸、硫酸和磷酸的盐;或有机酸如甲磺酸、对甲苯磺酸、乳酸、乙酸、三氟乙酸、柠檬酸、琥珀酸、富马酸、马来酸和水杨酸的盐。此外,足够酸性的化合物可形成碱金属盐或碱土金属盐,例如钠、钾、锂、钙或镁的盐;铵盐;或有机碱盐,例如甲胺、二甲胺、三甲

胺、三乙胺、乙二胺、乙醇胺、胆碱氢氧化物、甲基葡胺、哌啶、吗啉，三(2-羟乙基)胺、赖氨酸或精氨酸的盐；所有这些也都是本文所述的化合物的盐的进一步实例。本发明所述的化合物可以是溶剂合物，优选水合物。在生产的过程中，或由于最初无水的化合物的吸湿性质导致的结果，可能会发生水合/水化。其溶剂合物和/或水合物可以以例如固体或液体形式存在。

[0240] 本文所述的化合物，其药理学上可接受的盐，溶剂合物和水合物，以及制剂和药物组合物的各自治疗用途，也属于本发明的范围之内。

[0241] 如本发明所述的药物组合物包含至少一种本发明所述化合物和任选的一种或多种载体物质和/或佐剂。

[0242] 如上所述，包含本发明所述的化合物、其溶剂合物，盐或制剂的治疗上有用的药剂，也包括在本发明的范围内。在一般情况下，可通过使用本领域已知的公知的和可接受的方式，将本发明所述的化合物单独施用地或与任何其它治疗剂组合施用。

[0243] 对于口服给药，所述的治疗上有用的药剂可以通过下列途径之一施用：口服，例如作为片剂、锭剂、包衣片剂、丸剂、半固体、软或硬胶囊（例如软或硬明胶胶囊）、水性或油性溶液剂、乳剂、混悬剂或糖浆剂，肠胃外给药包括静脉内、肌肉和皮下注射，例如作为可注射溶液或悬浮液，直肠中作为栓剂，通过吸入或吹入（例如作为粉末制剂，如微晶或作为喷雾剂（例如液体气雾剂）），经皮，例如通过透皮递送系统（TDS），例如包含活性成分的膏药，或鼻内。为了生产这样的片剂、丸剂、半固体、包衣片剂、糖衣丸和硬胶囊例如明胶胶囊，治疗有用的产物可以与药学惰性的无机或有机的赋形剂混合，例如，乳糖、蔗糖、葡萄糖、明胶、麦芽、硅胶、淀粉或其衍生物、滑石、硬脂（stearinic）酸或其盐、脱脂奶粉等。为生产软胶囊，可以使用赋形剂，例如，植物油、石油、动物油或合成油、蜡、脂肪和多元醇。用于生产液体溶液、乳剂或混悬剂或糖浆剂，可以使用以下物质作为赋形剂，例如水、醇、含水盐水、葡萄糖水溶液、多元醇、甘油、脂质、磷脂、环糊精、植物油、石油、动物油或合成油。特别优选的为脂质，而更优选的是磷脂（优选天然来源的；特别优选具有300至350nm之间的粒径），优选在磷酸盐缓冲盐水中（pH值=7-8，优选7.4）。用于栓剂，可以使用的赋形剂是例如，植物油、石油、动物油或合成油、蜡、脂肪和多元醇。用于气雾剂制剂，可以使用对于此目的合适的压缩气体，例如，氧、氮和二氧化碳。该药理学上有用的药剂也可以含有添加剂用于保存、稳定，如UV稳定剂、乳化剂、甜味剂、芳香剂、盐类（用于改变渗透压）、缓冲剂、包衣添加剂和抗氧化剂。

[0244] 通常，口服或肠胃外给药于体重约80公斤的成人的情况，合适的日剂量为约1mg至约10,000mg，优选为约5mg至约1,000mg，尽管在有提示时其上限可超标。日剂量可以为单剂量或分剂量，或对于肠胃外给药，它可以通过连续输注或皮下注射给药。

[0245] 本发明的化合物可以通过发酵（例如，通过菌株MCy8071DSM27004的发酵）或通过本领域技术人员已知的化学合成应用程序来制备。

[0246] 例如，本发明的化合物可以通过以下流程制备：

[0247] 从各个被任选取代的构建单元构建单元（例如Ar<sup>1</sup>、Ar<sup>2</sup>、Ar<sup>3</sup>、Ar<sup>4</sup>和Ar<sup>5</sup>）起始，这些构建单元构建单元可使用本领域技术人员所知的酰氯或偶联试剂互相连接，例如按照下列反应方案：



[0250] 如果 $L^1$ 、 $L^2$ 、 $L^3$ 和/或 $L^4$ 为式-CH=CH-基团(或其他烯炔基团),各个被任选取代的构建单元构建单元(例如 $Ar^1$ 、 $Ar^2$ 、 $Ar^3$ 、 $Ar^4$ 和 $Ar^5$ )可以通过Wittig或 Horner反应互相连接,例如:按照下列反应方案:

[0251]  $R^1-Ar^1-CHO+BrPh_3P-CH_2-Ar^2-L^2-Ar^3-L^3-Ar^4-L^4-Ar^5-R^2$

[0252]  $R^1-Ar^1-CHO+(EtO)_2OPCH_2-Ar^2-L^2-Ar^3-L^3-Ar^4-L^4-Ar^5-R^2$

[0253] 如果 $L^1$ 、 $L^2$ 、 $L^3$ 和/或 $L^4$ 为杂环烷基或杂芳基,各个被任选取代的构建单元构建单元(例如 $Ar^1$ 、 $Ar^2$ 、 $Ar^3$ 、 $Ar^4$ 和 $Ar^5$ )可以使用类似的反应条件互相连接。

[0254] 孢囊菌酰胺生物合成基因簇的鉴定:

[0255] 孢囊菌酰胺生产菌的基因组用鸟枪法测序法进行测序。作为孢囊菌酰胺的主要构建单元构建单元是非蛋白原性的氨基酸对-氨基苯甲酸(PABA),对氨基苯甲酸合酶(查询号NP\_415614)被用作查询项,以识别在Cbv34基因组中推定的孢囊菌酰胺生物合成簇。重要的是,对-氨基苯甲酸合酶的同源物可以被鉴定出(CysD,图12和表A),它在计算机预测的约48KB大的NRPS基因簇(图12,分配:表A)中与非核糖体肽合成酶(CysG,H和K)一起构成一个操纵子。在此NRPS簇中的基因用pfam、NCBI BLAST和phyre2进行分析。除了对氨基苯甲酸合酶的同源物,在该基因簇中还可以发现两个其他的PABA生物合成酶:氨基脱氧分支酸裂合酶(CysI)和3-脱氧-d-阿拉伯-庚酮糖-7-磷酸(DAHP)合成酶(CysN)。DAHP合成酶(CysN)是用于生产莽草酸和分支酸的关键酶。在莽草酸途径的主干线,D-赤藓糖-4-磷酸和磷酸烯醇丙酮酸盐(DAHP合成酶)经由莽草酸转化为分支酸。CysI和CysD使得能够从分支酸直接生物合成PABA。此外,所述的簇包含对氨基苯甲酸N-氧合酶的同源物(CysR)。

[0256] 图12显示了本发明的孢囊菌酰胺的生物合成簇。

[0257] 能够合成选自下组的孢囊菌酰胺:孢囊菌酰胺A、B、C、D、E、F、G和H的重组生物合成簇,其中所述的簇包括如SEQ ID NO:40至73的所有多肽,或其功能性变体。

[0258] 如本文所用的术语“功能性变体”是指具有与本文所述的多肽序列有至少85%,90%,95%或99%序列相同的多肽。一个多肽的“功能性变体”可保留被确认为天然多肽中保守的氨基酸残基,和/或可具有非保守氨基酸残基。氨基酸可以是,相对于天然多肽而言的取代(不同氨基酸),插入或缺失,但是变体相比本文中所描述的多肽具有大致相似的(酶)活性或功能。“功能性变体”可以在自然界中找到,或为其工程化的突变体(重组体)。

[0259] 术语“同源性”指用于衡量序列的相似性或关系的一种序列特性。同源性是通过相同残基的数目除以残基的总数,并将结果乘以100测得的。

[0260] 如本文所用,术语“蛋白质”,“多肽”,“肽”用于定义排成直链的,两个或多个氨基酸残基构成的有机化合物,其中在所述有机化合物的单个氨基酸之间通过肽键(即相邻氨基酸残基之间的酰胺键)连接。按照惯例,一种蛋白质的一级结构的描述是从氨基端(N)开始并朝向羧基端(C)。

[0261] 如本文所用,“包含”,“包括”,“含有”,“特征在于”,以及它们的语法等同词是包含性的或开放式的术语,不排除另外的、未陈述的要素或方法步骤。例如,“包括”等可被解释为包括更具限制性的术语“由……组成”。

[0262] 如本文所用,“由……组成”不包括在权利要求中没有指定的任何元素(元件)、步骤或成分。

[0263] 当商品名在本文中被使用,则独立地包括了商品名的产品配方,仿制药,和商品



名产品的活性药物成分。

[0264] 一般而言,除非另外定义,本文所用的技术和科学术语具有通常本发明所属领域的普通技术人员理解的相同的含义,并与一般的教科书和字典一致。

[0265] 优选地,本发明的NRPS酶是非天然存在的NRPS。本发明的NRPS也可以是杂合型NRPS,它包括来自于两个或更多的NRPS或来自于一个或多个聚酮化合物合成酶(PKSs)的模块、结构域和/或其部分,或其功能性变体。

[0266] 本发明的所述孢囊菌酰胺的生物合成簇,优选包括表A中的元件。

[0267] 表A:本发明的孢囊菌酰胺基因簇。基因和NRPS域的注解对应于SEQ ID NO. 1的基因簇序列。

[0268]

基因簇序列中的位置(bp)						NRPS						
						域	基因簇序列中的位置(bp)			蛋白序列中的位置(aa)		
名称	最小	最大	方向	长度	aa		长度	最小	最大	长度	最小	最大
Orf1	15	845	反向	831	276							
Orf2	912	1148	反向	237	78							
Orf3	1339	1827	反向	489	162							
Orf4	1907	2170	反向	264	87							
Orf5	2347	2796	反向	450	149							
CysT	3035	6838	反向	3804	1267							
CysS	7049	8977	反向	1929	642							
CysR	9086	10087	反向	1002	333							
CysQ	10162	10956	反向	795	264							
CysP	11029	11730	反向	702	233							
CysO	11764	12375	反向	612	203							
CysA	12715	12927	正向	213	70							
CysB	12996	13949	正向	954	317							
CysC	13959	15338	正向	138	45							
CysD	15464	17662	正向	2199	732							
CysE	17749	18480	正向	732	243							
CysF	18503	19540	正向	1038	345							
CysG	19580	25558	正向	5979	1992	AMP-结合域	1451	19694	21145	483	39	521
						PCP 域	209	21221	21430	69	548	616
						缩合_LCL 域	893	21485	22378	297	636	932
						AMP-结合域	1451	22880	24331	483	1101	1583
						PCP 域	215	24404	24619	71	1609	1679
						硫酯酶域	788	24728	25516	262	1717	1978
CysH	25626	28553	正向	2928	975	AMP-结合域	1199	25737	26936	399	38	436
						新结构域类型	332	27231	27563	110	536	645
						AMP 结合域 C-末端	170	28032	28202	56	803	858
						PCP 域	197	28284	28481	65	887	951
CysI	28555	29373	正向	819	272							
CysJ	29392	30375	正向	984	327							

[0269]

CysK	30450	44087	正向	13638	4545	缩合_LCL 域	323	30459	30782	107	4	110
						AMP-结合域	1505	31239	32744	501	264	764
						PCP 域	197	32820	33017	65	791	855
						缩合_LCL 域	893	33072	33965	297	875	1171
						AMP-结合域	1505	34461	35966	501	1338	1838
						PCP 域	197	36042	36239	65	1865	1929
						缩合_LCL 域	890	36285	37175	296	1946	2241
						AMP-结合域	1574	37668	39242	524	2407	2930
						PCP 域	359	39165	39524	119	2906	3024
						缩合_LCL 域	893	39579	40472	297	3044	3340
						AMP-结合域	1505	40968	42473	501	3507	4007
						PCP 域	197	42549	42746	65	4034	4098
CysL	44084	47155	正向	3072	1023	缩合_LCL 域	896	42801	43697	298	4118	4415
						AMP-结合域	1445	45665	47110	481	528	1008

[0270] 本发明还提供了分离的、合成的或重组的用于编码本发明的NRPS的核酸。所述核酸包括包括本发明的NRPS的一部分或全部的核酸,进一步包括调控序列(如启动子和翻译的起始序列和终止序列)的核酸,以及可以进一步包括促进维持在宿主细胞中稳定的序列,即提供复制起点的功能的序列或通过同源重组促进整合入宿主细胞染色体或其它DNA的序列。这些NRPSs可以用作研究工具或作为重组NRPS或PKS簇的模块。

[0271] 优选地,本发明涉及分离的、合成的或重组的核酸,其中包括:

[0272] (i) 编码孢囊菌酰胺生物合成簇的序列,其中所述的序列与SEQ ID NO.1 全长序列具有至少85%、90%、95%、96%、97%、98%、98.5%、99%、或99.5%至100%的同源性;

[0273] (ii) 编码NRPS的序列,其中所述的序列与SEQ ID NO.8、9、12或13具有至少85%、90%、95%、96%、97%、98%、98.5%、99%、或99.5%至100%的同源性;

[0274] (iii) 与(i)或(ii)中的任何核酸的全长序列完全互补的序列;或

[0275] (iv) 编码如SEQ ID NO.46、47、50或51中任一多肽的序列。

[0276] 如本文所用,词语“核酸”或“核酸序列”是指寡核苷酸、核苷酸、多核苷酸、或任何其片段,也指基因组或合成来源的DNA,它可以是单链或双链的,并且可以代表有义链或反义链,可以是天然的或合成来源的。“寡核苷酸”包括单链多脱氧核苷酸,或者两条互补的多脱氧核苷酸链,它们可以是化学合成的。这种合成的寡核苷酸没有5'磷酸,并因此在激酶的存在下不通过ATP添加一个磷酸根,将不会连接于另一个寡核苷酸。合成的寡核苷酸

可以连接于尚未去磷酸化的片段。编码特定多肽或蛋白质的“编码序列”或“核苷酸序列”是这样的核酸序列,它被置于适当调控序列的控制下,并被转录和翻译成多肽或蛋白质。用于实施本发明的核酸可以分离自多种不同来源,被遗传改造,被扩增和/或被表达/被重组。用于核酸操作的技术,如,例如,亚克隆,标记探针(例如,使用Klenow聚合酶的随机引物标记、切口平移、扩增),测序,杂交等,在科学和专利文献中有很好的描述,参见例如,Sambrook编,《分子克隆:实验手册(第二版)》,VoI.1-3,冷泉港实验室,(1989);《分子生物学实验指南(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)》,Ausubel编辑,John Wiley& Sons,Inc.,纽约(1997年);《生物化学与分子生物学实验室技术:核酸探针杂交》,第一部分,理论和核酸合成,Tijssen编辑,Elsevier,纽约(1993)。编码本发明多肽的核酸,与能够指导被翻译的多肽或其片段的分泌的前导序列按合适的读框进行组装。

[0277] 如本文所用,术语“分离的”一词是指材料(例如,核酸、多肽、载体、细胞)与其原始环境例如自然环境(如果它是天然存在的)相分开。例如,天然存在的、位于活动物中的多核苷酸或多肽是未被分离的,而同样的多核苷酸或多肽,当与在自然系统中共存的一些或全部的物质分开时,即为分离的。此类多核苷酸可以是载体的一部分,和/或此类多核苷酸或多肽可以是组合物的一部分,并且仍然是分离的,只要此类载体或组合物不是其天然环境中的一部分。

[0278] 如本文所用,术语“合成的”是指该材料,例如核酸,在体外通过公知的化学合成技术被合成,如描述于例如,Adams (1983) J. Am. Chem. Soc. 105:661;Belousov (1997) Nucleic Acids Res.25:3440-3444;Frenkel (1995) Free Radic. Biol. Med.19:373-380;Blommers (1994) Biochemistry 33:7886-7896;Narang (1979) Meth. Enzymol. 68:90;Brown (1979) Meth. Enzymol. 68:109;Beaucage (1981) Tetra. Lett. 22:1859中。

[0279] 术语“重组的”是指该核酸邻接于在天然环境下并不相邻的“骨架”核酸。本发明的骨架分子包括核酸,如克隆载体和表达载体、自我复制的核酸、病毒、整合核酸和用于维持或操作感兴趣核酸插入序列的其他载体或核酸。从这些核酸产生的本发明重组多肽可被单独分离或克隆,并测试其期望活性。可以使用任何重组表达系统,包括细菌、哺乳动物、酵母、昆虫或植物细胞表达系统。

[0280] 还提供了包含至少一种本发明核酸的载体。所述载体可以是克隆载体、表达载体或人工染色体。

[0281] 如本文所用,术语“载体”是指能够将其连接的另一核酸分子进行转运的核酸。载体(包括克隆载体和表达载体)包含了本发明的核酸或其功能等同物。本发明的核酸可以整合入重组的可复制载体中,例如克隆载体或表达载体中。该载体可用于在相容的宿主细胞中复制该核酸。因此,本发明还提供了制备本发明的多核苷酸的方法,其中通过将本发明的多核苷酸引入可复制载体,将该载体导入相容的宿主细胞,并在能够复制载体的条件下培养所述宿主细胞。载体可以从宿主细胞中回收。合适的宿主细胞在下文中进行了描述。用于插入表达盒或本发明的核酸的载体,可以是可便利地进行重组DNA操作的任何载体,并且载体的选择常常取决于待被引入的宿主细胞。多种用于与原核和真核宿主一起使用的克隆载体和表达载体如Sambrook等人,分子克隆:实验室手册,第二版,冷泉港,纽约(1989)所述。

[0282] 如本发明所述的载体可以是自主复制载体,即,其中作为染色体外的实体存在的

载体,其复制是独立于染色体的复制,例如,质粒。或者,所述的载体可以是当导入宿主细胞时,被整合到宿主细胞基因组中并与其整合的染色体一起复制的载体。

[0283] 一种类型的载体是“质粒”,它指环状双链DNA环,在该环中可连入另外的DNA片段。另一类型的载体是病毒载体,其中额外的DNA片段可以连接入病毒基因组中。某些载体能够在它们所导入的宿主细胞中自主复制(例如,具有细菌复制起点的细菌载体,和游离型(episomal)哺乳动物载体)。其他载体(例如,非游离型哺乳动物载体)被引入宿主细胞后被整合入宿主细胞的基因组中,并由此随着宿主基因组复制。此外,特定的载体能够指导与它们可操作地连接的基因的表达。此类载体在本文中称为“表达载体”。一般而言,用于重组DNA技术的表达载体通常是质粒的形式。术语“质粒”和“载体”在本文中可以互换使用,因为质粒是最常用的载体形式。然而,本发明旨在包括此类其它形式的表达载体,如粘粒,病毒载体(例如,复制缺陷型逆转录病毒,腺病毒和腺相关病毒)及与其发挥等同功能的噬菌体载体。

[0284] 本发明的载体可以体外使用,例如用于生产RNA或用于转染或转化宿主细胞。

[0285] 本发明的载体可包含两种或更多,例如三,四或五种本发明的核酸,例如用于过表达。

[0286] 本发明的重组表达载体包含本发明核酸,其形式适合于核酸在宿主细胞中表达所述核酸,这意味着该重组表达载体包括一个或多个调控序列,所述调控序列可基于用于表达的宿主细胞进行选择,并可操作地连接于待表达的核酸序列。

[0287] 在载体(如表达载体)中,“可操作地连接”是用于指感兴趣的核苷酸序列以允许所述核苷酸序列表达的方式连接于调控序列(例如,在体外转录/翻译系统中,或当载体被引入宿主细胞时在宿主细胞中)。即,术语“可操作地连接”是指其中所述元件处于允许它们以其预期方式发挥功能的并列关系。“可操作地连接”于编码序列的调控序列如启动子、增强子或其它表达调控信号,是用以下方式进行定位:在与控制序列相容的条件下实现该编码序列的表达,或者各序列被排列从而使得它们共同起作用以达到预期目的,例如转录在启动子处开始并持续进行并通过编码多肽的DNA序列。

[0288] 术语“调控序列”或“控制序列”旨在包括启动子、操纵子、增强子、减弱子(attenuator)和其他表达控制元件(例如,多腺苷酸化信号)。此类调节序列描述于例如在Goeddel;基因表达技术:酶学方法,185,学术出版社,圣地亚哥,加利福尼亚(1990)中。

[0289] 术语调控或控制序列,包括在许多类型宿主细胞中指导核苷酸序列组成型表达的序列,以及仅在特定的宿主细胞中指导核苷酸序列表达的序列(如组织特异性的调控序列)。

[0290] 对于给定的宿主细胞的载体或表达构建,可因此包含可操作地依次互相连接的下列元件,按相对于编码本发明多肽的编码链序列的从5'端至3'端方向表示:(i)能够指导编码多肽的核苷酸序列在给定的宿主细胞中的转录的启动子序列;(ii)任选的,能够指导多肽从给定的宿主细胞分泌到培养基中的信号序列;(iii)任选的,编码C-末端、N-末端或内部的,用于纯化、检测或标记多肽的表位标签序列,或上述序列的组合;(iv)编码本发明多肽的本发明的核酸序列;以及还优选地包括(v)转录终止区(终止子),其能够终止位于编码多肽的核苷酸序列下游的转录。特定的被命名的细菌启动子包括lad、lacZ、T3、T7、SP6、K1F、tac、tet、gpt、 $\lambda$ (lambda) $P_R$ 、 $P_L$ 和trp。真核启动子包括CMV即时早期启动,HSV胸苷激酶,早期和晚期SV40,来自逆转录病毒的LTR和小鼠金属硫蛋白-I等启动子。选择合适的载体和

启动子是在普通技术人员的技术水平范围内的。本发明的核苷酸序列的下游可以存在含有一个或多个转录终止位点(例如 终止子)的3' 非翻译区。终止子的来源是不太关键的。终止子可以例如,是对于 编码多肽的DNA序列而言天然的终止子。优选地,终止子是宿主细胞内源性的(在所述宿主细胞中,编码多肽的核苷酸序列是待表达的)。在被转录的区域中,可以存在用于翻译的核糖体结合位点。由构建物表达的成熟转录本的编码部分,包括位于待翻译的多肽的起始处的用于翻译起始的AUG(或在原核生物内为TUG或 GUG)以及位于末端的终止密码子。

[0291] 本发明的多核苷酸的增强表达也可以通过异源调控区的选择实现,例如启动子、分泌前导和/或终止子区域,其可用于增加感兴趣蛋白质的表达和(如果需要的话)从表达宿主分泌的水平 and/或提供本发明多肽表达的诱导型控制。本领域的技术人员可以理解,该表达载体的设计可取决于这些因素:如待转化的宿主细胞的选择,所需的蛋白质表达水平等。如本文所述,载体例如本发明的表达载体可以被引入到宿主细胞中从而产生核酸所编码的蛋白质或肽。

[0292] 本发明的载体,如重组表达载体,可以被设计用于在原核或真核细胞的表达部分或全部的本发明的NRPS。例如,部分或全部的本发明的NRPS可以在细菌细胞如大肠杆菌、芽孢杆菌菌株、昆虫细胞(使用杆状病毒表达载体)、丝状真菌、酵母细胞或哺乳动物细胞中表达。合适的宿主细胞在Goeddel,基因表达技术:酶学方法,185,学术出版社,圣地亚哥,加利福尼亚(1990)中有进一步描述。合适宿主的代表性实例在下文中说明。上述宿主细胞的适当培养基和培养条件是现有技术公知的。

[0293] 如上所述,术语“控制序列”或“调控序列”在本文中定义为包括至少任何一种对于多肽的表达可能是必要的和/或有利的元件。任何控制序列可以是对于 编码多肽的本发明的核苷酸序列而言天然的或外源的。这样的控制序列可包括,但不限于:启动子、前导序列、优化翻译起始序列(如Kozak,1991,J.Biol. Chem.266:19867-19870中所述)、分泌信号序列、原肽序列、多聚腺苷酸化序列、转录终止子。至少,该控制序列通常包括启动子和转录和翻译的终止信号。稳定转化的微生物是指已引入一个或多个DNA片段,从而使得所引入的分子在生长培养过程中被保留、复制和分开。稳定转化可能是由于多个或单个染色体整合,或由于染色体外的元件,例如质粒载体。质粒载体能够指导由特定DNA片段编码的多肽表达。表达可以是组成型或是由诱导型(或阻抑型)启动子所调节,这些启动子使得编码特定多肽的功能性相关的DNA片段被高水平转录。

[0294] 本发明的表达载体还可以包括选择性标记基因,以便选择出被转化的细菌菌株,例如,导致赋予细菌针对药物(例如氯霉素、红霉素、卡那霉素、新霉素、四环素、以及氨基青霉素和其他青霉素衍生物如羧苄青霉素)抗性的基因。可供选择的标记也可包括生物合成基因,例如在组氨酸、色氨酸和亮氨酸生物合成途径中的基因。

[0295] 合适的多核苷酸序列可以通过各种不同程序插入载体中。在一般情况下,多核苷酸序列可以在用适当的限制性内切酶对插入片段和载体进行酶切后,连接到载体的所需位置。或者,可以对插入元件和载体的平末端进行连接。多种克隆技术在Ausubel等人,《目前的分子生物学方法》,John Wiley 503Sons,Inc. 1997以及Sambrook等,《分子克隆实验手册》,第二版,冷泉港实验室出版社(1989)中有公开。多核苷酸序列也可通过同源重组的技术(包括在体外以及在体内重组)进行克隆。这样的方法和其它方法被认为是本领域技术人员的能力范围之内的。所述载体可以是,例如,质粒、病毒颗粒或噬菌体的形式。其它

载体包 括染色体的、非染色体的和合成的多核苷酸序列、SV40的衍生物；细菌质粒、噬 菌体DNA、杆状病毒、酵母质粒、基于质粒和噬菌体DNA组合而衍生的载体，病 毒DNA如牛痘、腺病毒、禽痘病毒和伪狂犬病的组合。

[0296] 本发明还提供了一种工程或重组宿主细胞，即包含本发明的多核苷酸序列作 为异源或非天然多核苷酸(例如，编码孢囊菌酰胺生物合成簇或本发明的NRPS)，或包含本发明的载体的转化细胞。所述宿主细胞可以是本领域技术人员熟悉的任 何宿主细胞，包括原核细胞和真核细胞，例如细菌细胞、真菌细胞、酵母细胞、哺乳动物细胞、昆虫细胞、或植物细胞。

[0297] 优选的哺乳动物细胞包括例如中国仓鼠卵巢 (CHO) 细胞、COS细胞、293细 胞、PerC6细胞、杂交瘤细胞、Bowes黑素瘤细胞或任何小鼠或任何人细胞系。示例性昆虫细胞包括夜蛾属或果蝇属的任何物种，包括果蝇S2和草地夜蛾Sf- 9。示例性的真菌细胞包括任何的曲霉种类。优选的酵母细胞包括，例如，念珠 菌属、汉逊酵母属、克鲁维酵母菌属、毕赤酵母属、酿酒酵母属、裂殖酵母属、或耶氏酵母属菌株的细胞。更优选为乳酸克鲁维酵母、酿酒酵母、多形汉逊酵 母、解脂耶氏酵母、或毕赤酵母。根据本发明，宿主细胞可以是原核细胞。优选 地，所述的原核宿主细胞是细菌细胞。术语“细菌细胞”包括革兰氏阴性和革兰氏阳性以及古细菌的微生物。合适的细菌可选自例如埃希氏杆菌属、鱼腥藻属、柄杆菌属、葡糖杆菌属、光合细菌属、假单胞菌属、副球菌属、芽孢杆菌属、短 杆菌属、棒状杆菌属、根瘤菌属(中华根瘤菌)、黄杆菌属、克雷伯氏菌属、肠杆 菌属、乳杆菌属、乳球菌属、甲基杆菌属、葡萄球菌或链霉菌属。优选地，所述 细菌细胞是选自枯草芽孢杆菌 (*B. subtilis*)、解淀粉芽孢杆菌 (*B. amyloliquefaciens*)、地衣芽孢杆菌 (*B. licheniformis*)、普氏芽孢杆菌 (*B. pumilus*)、巨大芽孢杆菌 (*B. megaterium*)、耐盐芽孢杆菌 (*B. halodurans*)、短 小芽孢杆菌 (*B. pumilus*)、氧化葡糖杆菌 (*G. oxydans*)、新月柄杆菌 CB15 (*Caulobacter crescentus*)、扭脱原单胞菌 (*Methylobacterium extorquens*)、球形红细菌 (*Rhodobacter sphaeroides*)、恶臭假单胞菌 (*Pseudomonas putida*)、zeaxanthinifaciens副球菌 (*Paracoccus zeaxanthinifaciens*)、脱氮副球菌 (*Paracoccus denitrificans*)、大肠杆菌、谷氨酸棒杆菌 (*C. glutamicum*)、肉葡萄球菌 (*Staphylococcus carnosus*)、变青 链霉菌 (*Streptomyces lividans*)、草木樨中华根瘤菌 (*Sinorhizobium meliloti*) 和放射根瘤菌 (*Rhizobium radiobacter*)。合适的宿主的选择是在本领域技术人 员的能力范围内的。

[0298] 所述载体可被使用任何各种不同技术引入到宿主细胞，包括转化、转染、转 导、病毒感染、基因枪或Ti介导的基因转移。具体的方法包括磷酸钙转染、DEAE-葡聚糖介导的转染、脂转染、或电穿孔 (Davis, L., Digner, M., Battey, I., 分子生物学基本方法, (1986))。本发明核酸或载体可被引入到细胞中，以 便由此筛选核酸是否以适合随后表达核酸的方式进入细胞。引入的方法在很大程度上取决于靶细胞类型。示例性的方法包括CaPO<sub>4</sub>沉淀、脂质体融合、脂质转染 (例如, LIPOFECTIN™)、电穿孔、病毒感染等。候选的核酸可以稳定地整合到宿 主细胞的基因组(例如，逆转录病毒引入)，或者可以瞬时或稳定存在于细胞质 (即，通过使用传统的质粒，利用标准的调控序列、选择标记、等等)。因为许多 药 学上重要的筛选需要人或模型哺乳动物的靶细胞，可使用能够转染这些靶细胞 的逆转录病毒载体。

[0299] 在适当情况下，工程宿主细胞可以在常规营养培养基中进行培养，该培养基 可改良以便适合激活启动子、选择转化子或扩增本发明的核酸。在转化合适的宿 主菌株并宿主菌株生长至适当的细胞密度之后，可通过适当的手段(例如，温度 变化或化学诱导)来诱导

选定的启动子,而细胞可以被额外培养一段时间以允许它们以产生其所需的多肽或其片段。细胞可以通过离心收获,通过物理或化学方法破碎细胞,保留得到的粗提取物以用于进一步纯化。用于表达蛋白质的微生物细胞可以通过任何常规方法来破碎,包括冻融循环、超声处理、机械破碎或使用细胞裂解剂。这样的方法是本领域技术人员所熟知的。表达的多肽或其片段可被通过以下方法从重组细胞培养物中回收并纯化,其中包括硫酸铵或乙醇沉淀、酸提取、阴离子或阳离子交换层析、磷酸纤维素层析、疏水相互作用层析、亲和层析、羟基磷灰石层析和凝集素层析。如必要,可采用蛋白质重折叠步骤,以完成所述多肽的构型。如果需要的话,高效液相色谱(HPLC)可用于最终纯化步骤。在宿主细胞中的构建物可用于以常规的方式产生重组序列所编码的基因产物。取决于在重组生产方法中使用的宿主,由含有载体的宿主细胞产生的多肽可以是糖基化或是未糖基化的。本发明的多肽还可以或不包括起始甲硫氨酸的氨基酸残基。无细胞翻译系统也可以用来生产本发明的多肽。无细胞翻译系统可以使用从DNA构建物转录出的mRNA,该构建物包括可操作连接于编码多肽或其片段的核酸的启动子。在一些方面,该DNA构建物在进行体外转录反应之前,可以被线性化。转录的mRNA然后与合适的无细胞翻译提取物(如兔网状细胞提取物)一同温育,以产生所需的多肽或其片段。

[0300] 含有感兴趣的多核苷酸(例如,本发明的核酸)的宿主细胞,可以在常规营养培养基中培养,该培养基可经改良而适合激活启动子、选择转化子或扩增基因。培养条件如温度、pH等,就是先前该选定的用于表达的宿主细胞所用的条件,并且对于普通技术人员是显而易见的。被识别为具有特定酶活性的克隆随后被测序,以确定编码本发明的一部分或全部的NRPS的多核苷酸序列。

[0301] 重组DNA可通过任何手段被引入到宿主细胞中,包括但不限于,质粒、粘粒、噬菌体、酵母人工染色体或其它介导遗传元件转移至宿主细胞中的载体。这些载体可包括复制起点,以及控制该载体和由该载体所携带的遗传因子的复制的顺式作用控制元件。选择性标记可以存在于载体上,以协助识别在其中已引入遗传元件的宿主细胞。用于将遗传元件引入到宿主细胞的方法(例如,克隆)是本领域技术人员公知的。其它克隆方法包括但不限于,遗传物质直接整合入染色体。这可以通过各种手段实现,包括在侧接宿主染色体的同源DNA序列的非复制质粒上克隆本文中描述的遗传元件;一旦将所述重组质粒转入宿主,遗传元件可以通过DNA重组被引入到染色体。如果整合的DNA片段含有可选择的标记,如抗生素抗性,则这样的重组菌株可以回收。或者,该遗传元件可以被直接导入到宿主细胞的染色体中,而不使用非复制型质粒。这可以通过合成产生本发明的遗传元件的DNA片段来实现,其中该DNA片段还包含宿主染色体的同源DNA序列。同样,如果这些合成的DNA片段也包含可选择标记,那么该遗传元件可以被插入到宿主染色体中。

[0302] 所述孢囊菌酰胺生物合成簇或本发明的NRPS在任何上述宿主细胞中可以有利地被表达。因此,本发明提供各种不同的宿主细胞,它们包含一种或多种分离的、合成的或重组的核酸和/或本发明的NRPSs。所述的宿主细胞在合适的条件下培养时,能够产生选自孢囊菌酰胺A、B、C、D、E、F、G和H的孢囊菌酰胺,而在不存在本发明的核酸的情况下这些孢囊菌酰胺原本是不产生的或产生的水平较低。

[0303] 本发明还涉及分离的、合成的或重组的具有SEQ ID NO.40至73任一的氨基酸序列的多肽,或由本发明核酸所编码的氨基酸序列。

[0304] 本发明进一步提供了制备选自下组的孢囊菌酰胺的方法：孢囊菌酰胺A、B、C、D、E、F、G和H，所述方法一般包括：提供本发明的宿主细胞，并在合适的培养基中在合适的条件下培养所述宿主细胞，使得至少一种选自下组的孢囊菌酰胺被生产：孢囊菌酰胺A、B、C、D、E、F、G和H。该方法还可以包括分离选自孢囊菌酰胺A、B、C、D、E、F、G和H中的孢囊菌酰胺的步骤，即从培养液中分离和保留化合物。所述分离步骤可以利用亲和层析、阴离子交换层析、或反相色谱法进行。

## 实施例

[0305] 制备方法

[0306] 生产制备用菌株

[0307] 菌株深棕色孢囊杆菌(*Cystobacter velatus*)MCy8071属于粘细菌目(粘细菌)、孢囊杆菌亚目、孢囊杆菌科、孢囊杆菌属。部分16S rRNA基因序列与公共数据库的序列比较(BLAST,由NCBI(美国国家生物技术信息中心)所提供的“基本的局部比对搜索工具”)显示,与深棕色孢囊杆菌菌株DSM14718相似度100%。

[0308] MCy8071是从1982年采集的中国的土壤样品中,在亥姆霍兹感染研究中心(HZI,前身为GBF)分离出。该菌株于2013年3月保藏在布伦瑞克(Braunschweig)的德国微生物保藏中心(DSM),保藏号为DSM27004。

[0309] 培养

[0310] 菌株MCy8071在酵母琼脂(VY/2:0.5%酿酒酵母,0.14%CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.5μg维生素B<sub>12</sub>/1,1.5%琼脂,pH 7.4),CY-琼脂(酪朊0.3%,酵母提取物 0.1%,CaCl<sub>2</sub>·2H<sub>2</sub>O 0.1%,琼脂1.5%,pH 7.2)和P-琼脂(马考尔蛋白胨(peptone Marcor)0.2%,淀粉0.8%,单细胞蛋白Probione 0.4%,酵母提取物 0.2%,CaCl<sub>2</sub>·2H<sub>2</sub>O 0.1%,MgSO<sub>4</sub> 0.1%,Fe-EDTA 8mg/1,1.5%琼脂,pH 7.5)上充分生长。工作培养物培育于液体介质CY/H(50%CY-培养基+50毫摩尔的Hepes缓冲液,50%的H-培养基:大豆粉0.2%,葡萄糖0.8%,淀粉0.2%,酵母提取物0.2%,CaCl<sub>2</sub>·2H<sub>2</sub>O 0.1%,MgSO<sub>4</sub> 0.1%,Fe-EDTA 8mg/1,Hepes 50mM pH 7.4)中。液体培养物以180rpm在30℃下摇动。为了保存,将2毫升每份的三天的培养物保存在-80℃下。即使经过好几年,在上述的琼脂平板上或在20毫升CY/H介质(于具有栓塞和铝盖的100ml锥形瓶中)中激活也没有问题。一两天后,20毫升培养物可以升级至100毫升。

[0311] 形态描述

[0312] 在液态介质CY/H中两天后,上述菌株MCy8071的杆状细胞具有9.0-14.5μm的长度,以及0.8-1.0μm的宽度。在上述琼脂平板上,菌落是圆形的。在VY/2-琼脂中,菌落是薄而透明的。在VY/2-琼脂中,酵母的退化是可见的。在CY-琼脂中,培养物看起来是透明的橙色。在P-琼脂上,细胞群的产生是独特的,且菌落聚集行为减少。菌落是橙棕色的。P-琼脂中的淀粉被分解。

[0313] MCy8071对以下抗生素耐药:氨苄青霉素、庆大霉素、潮霉素、多粘菌素(polymycin)、杆菌肽、大观霉素、新霉素、和夫西地酸(fusidinic acid)。在头孢菌素和春雷霉素存在下可能有微弱生长,在硫链丝菌素、甲氧苄啶、卡那霉素、和土霉素存在下生长是不可能的(所有抗生素的终浓度调节至50μg ml<sup>-1</sup>)。

[0314] 孢囊菌酰胺A、B、C、D、E、F、G和H的生产



[0315] 菌株在复合介质中生产。它更喜欢在含氮的营养素,如单细胞蛋白(Probion) 和蛋白质分解产物如蛋白胨、胰蛋白胨、酵母提取物、大豆粉和肉提取物等。在此,多个所述的蛋白混合物存在下的生产情况优于使用单一的蛋白质。

[0316] 孢囊菌酰胺是在生长对数期至稳定期之间产生的。经过100升发酵(培养基E) 两天后,产物的量不再增加。

[0317] 孢囊菌酰胺被递送到培养基,并与XAD-吸附树脂结合。XAD用金属筛进行筛分,并在丙酮中洗脱。在不同的生产温度下进行了测试(21℃,30℃,37℃和42℃),而在42℃下是不可能进行生产的。在最大通气量下,最适温度为30℃。

[0318] MCy8071的发酵是在150升发酵罐中进行,其中采用100升培养基E(脱脂奶0.4%,大豆粉0.4%,酵母提取物0.2%,淀粉1.0%,MgSO<sub>4</sub> 0.1%,Fe-EDTA 8mg/l,甘油0.5%;pH为7.4)进行发酵,以及在100升发酵罐中进行,其中用70升培养基M(大豆蛋白胨1.0%,麦芽糖1.0%,CaCl<sub>2</sub>·2H<sub>2</sub>O 0.1%,MgSO<sub>4</sub> 0.1%,Fe-EDTA 8mg/l;pH值7.2),并在30℃下发酵四天。用氢氧化钾(2.5%)和硫酸将pH调节至7.2到7.4之间。搅拌器速度为100-400rpm,用0.05vvm压缩空气进行通气。发酵液中的溶解氧的含量通过搅拌器速度调控至pO<sub>2</sub> 40%。为结合孢囊菌酰胺,1%的吸附树脂被加入到发酵液中。将5升的三天预培养物(分别为E或M-培养基),接种于发酵罐。在发酵过程中的产物经HPLC-MS-分析和针对大肠杆菌的甲醇提取物的系列稀释试验进行检查。该菌株生产孢囊菌酰胺A、B、C、D、E、F、G和H。

[0319] 基因敲除实验

[0320] 为了确认孢囊菌酰胺生物合成基因簇是负责生产孢囊菌酰胺的,进行了敲除(KO)实验,其中CysK(NRPS)和CysL(苯甲酰辅酶A连接酶)被分别敲除。具体而言,使用Taq聚合酶,根据MCy8071基因组DNA,产生CysK和CysL基因的1000bp片段PCR产物。该引物设计成在PCR产物的末端添加3个终止密码子。

[0321] CysL敲除,正向

[0322] TGATTGATTGATCGGCGCGATTTCGGCCTCTGG

[0323] CysL敲除,反向

[0324] TCAATCAATCATCGGGTCGCGGTCTCAGGCTC

[0325] CysK敲除,正向

[0326] TGATTGATTGAAAAACAGTCGGAGGAGTTTCTTGTC

[0327] CysK敲除,反向

[0328] TCAATCAATCAACTCCCAGTGCCCTCAGCCTC

[0329] 将PCR产物进行凝胶纯化,其中使用Nucleospin®凝胶和马谢雷-内格尔(Macherey-Nagel)PCR清理(Clean up)试剂盒,并克隆至pCR2.1-TOPO载体中。该构建物通过热休克整合成到化学感受态大肠杆菌HS996,并在添加有卡那霉素的LB琼脂平板上完成挑选。通过碱裂解质粒制备和EcoRI限制性消化,来筛选具有正确构建物的单菌落。然后对构建物进行测序,以确保序列同源性。

[0330] 用于各KO的正确的构建物被转入非甲基化的化学感受态大肠杆菌SCS110。质粒通过使用Thermo Scientific公司的GeneJET质粒小量制备试剂盒制备,并通过电穿孔导入MCy8071。转化的复制体的选择在添加有卡那霉素的CTT琼脂平板上进行。KO突变体和野生型培养物在吸附树脂(XAD-16)存在下进行平行生长,并对培养物的粗提取物样品进行分析。

[0331] 结果表明,在K0突变体中完全没有孢囊菌酰胺生产,这提示CysK和CysL 对于孢囊菌酰胺的生产是必不可少的。此外,该结果提示了孢囊菌酰胺生物合成 基因簇用于生产孢囊菌酰胺的基本性质。

[0332] 结构分析:

[0333] 孢囊菌酰胺A (1) 的HRESI (+) MS分析返回了准分子离子  $(M+H)^+$ , 其符合分子式  $C_{46}H_{45}N_7O_{14}$ , 需要28个双键当量 (DBE)。 $^{13}C$  NMR (DMSO- $d_6$ ) 数据显示7酯/酰胺 羰基 ( $\delta_c$  163.7 至169.6) 以及进一步的 $30sp^2$ 共振态 ( $\delta_c$  114.2至150.8), 对应于22DBE。对一维和二维NMR数据 (表1) 的考虑, 显示了一套五芳香自旋系统, 其中三个对应于对位取代, 1,3,4-三取代和1,2,3,4-四取代的苯环。一组HMBC 相关性: 从H-6,6' ( $\delta_H$  7.96) 和NH ( $\delta_H$  8.92) 到酰胺羰基C-4 ( $\delta_c$  166.5); 从NH ( $\delta_H$  10.82) 到C-7/7' ( $\delta_c$  119.8) 和第二个酰胺羰基C-10 ( $\delta_c$  164.6); 从H-12/12 ( $\delta_H$  8.20) 到C-10, 建立了两个对位取代的芳香环体系的连接关系 (图1)。对于 $^1H$ 和 COSY NMR数据的进一步检验, 建立了酰胺NH ( $\delta_H$  8.92) 延伸至次甲基H-2 ( $\delta_H$  4.96) 和H-1 ( $\delta_H$  4.70) 的连接关系。低磁场特征H-1 ( $\delta_c$  79.4) 提示了氧取代, 上述特征 被HMBC的从H-1到1-OMe ( $\delta_H$  3.53,  $\delta_c$  59.6) 相关性所证实。还观察到HMBC中从 H-1和H-2到酯/酰胺羰基 ( $\delta_c$  169.6) 的相关性, 这导致指向亚基A的结构 (图1)。

[0334] 对于1,3,4-三取代的苯环, 观察到HMBC中从H-17 ( $\delta_H$  7.58) 到酯/酰胺羰基 C-15 ( $\delta_c$  167.3)、氧季碳C-18 ( $\delta_c$  146.8)、C-19 ( $\delta_c$  133.6) 和C-21 ( $\delta_c$  122.9) 的 相关性。1,2,3,4-四取代苯环的独立的自旋系统, 显示了HMBC中以下相关性: i) 从H-25 ( $\delta_H$  7.82, d, 8.7) 到酯/酰胺羰基C-23 ( $\delta_c$  163.7)、C-27 ( $\delta_c$  136.2) 以及氧 季碳C-29 ( $\delta_c$  150.8); ii) 从H-26 ( $\delta_H$  7.42) 到C-24 ( $\delta_c$  117.3) 和C-28 ( $\delta_c$  139.5) 的相关性, 以及酚羟基 ( $\delta_H$  11.25) 与C-24 和C-28的相关性。三和四取代的苯环 互为对位连接, 这可由HMBC中酰胺NH ( $\delta_H$  10.98) 与C-20 ( $\delta_c$  119.8)、C-18 ( $\delta_c$  146.7) 以及C-23 ( $\delta_c$  163.7) 的相关性得出 (图1)。最后对位取代的 芳族自旋系统 的H-33/33' ( $\delta_H$  8.11, d, 8.3) 和H-34/34' ( $\delta_H$  7.44, d, 8.3), 显示了连接于1, 2,3- 三取代苯环, 这可由HMBC中酰胺NH ( $\delta_H$  9.88) 和H-33/33' 与酰胺羰基C-31 ( $\delta_c$  164.3) 的相关性得出。对COSY数据的额外解读, 显示了两个异丙氧基残留 (H3- 39 ( $\delta_H$  1.38) -H-38 ( $\delta_H$  4.76) -H-40 ( $\delta_H$  1.38)) 和 (H3-42 ( $\delta_H$  1.25) -H-41 ( $\delta_H$  4.30) - H3-43 ( $\delta_H$  1.25))。两个异 丙氧基残留确认与氧季碳C-18 ( $\delta_c$  146.7) 和C-28 ( $\delta_c$  139.5) 相连, 基于ROESY中H-38/H-39 与H-17/NH, 以及H-42/43与NH/29- OH/H-33/33' 的相关性 (图1)。亚基A和B的关系未被建立, 不过基于孢囊菌酰胺 B的结构相似性, 亚基A和B的连接点被推测出。在解释了大部分共振之后,  $N_2O_3H_2$ 和1DBE被留下来用于进一步解释。化合物的UV光谱显示出 $\lambda_{max}$ 为301和 320nm, 这暗示一个偶联系统, 该偶联系统只可能是由硝基官能团对位连接于亚 基A的芳基上所产生的。剩余的MF被调节为生成位于亚基B的1,2,3-取代的芳 环上的羧酸残基 (C-15), 产生了4-氨基-3-异丙氧基苯甲酸部分, 从而导致构成 了孢囊菌酰胺A的平面结构。

[0335] 孢囊菌酰胺B (2) 的HRESI (+) MS分析返回了准分子离子  $(M+H)^+$ , 其符合分子式  $C_{46}H_{44}N_6O_{15}$ , 需要28个双键当量 (DBE)。2的NMR数据 (表2) 与 (1) 高度相似, 并且现在NH ( $\delta_H$  10.19) 和氧次甲基H-1 ( $\delta_H$  4.32) 看到了羰基C-37 ( $\delta_c$  168.6), 这 确认了亚基A和B的连接位点。除了这个, 唯一的变化是羰酰胺现在调整为羧 酸, 这随后被孢囊菌酰胺B二甲酯的生成所证实。

[0336] 孢囊菌酰胺C (3) 的HRESI (+) MS分析返回了准分子离子  $(M+H)^+$ , 其符合分子式  $C_{27}H_{29}N_3O_7$ , 需要15个双键当量 (DBE)。孢囊菌酰胺C的 $^1H$  NMR数据展示了芳 香信号, 这使人

想起孢囊菌酰胺A和B,但它缺乏两组对位取代的芳族单元的芳族共振。COSY数据揭示了存在两组异丙氧基残基连同一组对位取代的芳环体系。基于与,对1D和2D NMR数据的解释(表3,图2)鉴定了孢囊菌酰胺C (3) 与孢囊菌酰胺A和B的右侧部分相似,该部分由3-异丙氧基苯甲酸,2-羟基-3-异丙氧基苯甲酰胺与对氨基苯甲酰胺单元组成。

[0337] 表1.孢囊菌酰胺A (1) 的NMR (700MHz, DMSO-d<sub>6</sub>) 数据

位置	$\delta_{\text{H}}$ , 多重( <i>J</i> , Hz)	$\delta_{\text{C}}$	相关谱 (COSY)	异核多键相关 (HMBC)	旋转性核-核过偶 极偶合影响 (ROESY)
1	4.70, d (6.9)	79.4	2	2, 1-OMe, CO <sub>2</sub> NH <sub>2</sub>	1-OMe, 3
2	4.96, dd (8.2, 6.9)	55.6	1, 3	1, CO <sub>2</sub> NH <sub>2</sub> , 4	1-OMe, 3, 34
3	8.92, d (8.2)		2	4	1, 2, 6'
4		166.5			
5		128.6			
6, 6'	7.96, d (8.6)	128.9	7, 7'	4, 6, 6', 8	3
7, 7'	7.91, d (8.6)	119.8	6, 6'	5, 7, 7'	9
8		142.2			
9	10.82, s			7, 7', 10	7', 12'
10		164.6			
11		140.4			
12, 12'	8.20, d (8.6)	129.5	13, 13'	12, 12', 10, 14	9
13, 13'	8.39, d (8.6)	123.8	12, 12'	11, 13, 13', 14	
14		149.6			
15		167.3			
16		126.2			
[0338] 17	7.58, s	114.2		15, 18, 19, 21,	38, 40
18		146.7			
19		133.6			
20	8.50, d (8.2)	119.8	21	16, 18	21
21	7.60, d (8.2)	122.9	20	15, 17	20
22	10.98, s			18, 20, 23	25, 39
23		163.7			
24		117.3			
25	7.82, d (8.7)	125.2	26	23, 24, 29	22
26	7.42 <sup>a</sup>	116.3	25	27, 28	30
27		136.2			
28		139.5			
29		150.8			
30	9.88, s			26, 27, 31	33, 41, 42, 43
31		164.3			
32		134.0			
33, 33'	8.11, d (8.3)	129.5	34, 34'	31, 33, 33', 35	30, 41, 42, 43
34, 34'	7.44 <sup>a</sup>	125.6	33, 33'	34, 34', 32	1-OMe, 2
35		137.3			

	36	11.53, s			
	37		NO		
	1-OMe	3.53, s	59.6	1	1, 2
	38	4.76, spt (6.0)	72.1	39, 40	17
	39	1.38, d (6.0)	22.1	38	38, 40
[0339]	40	1.38, d (6.0)	22.1	38	38, 39
	41	4.30, spt (6.0)	76.0	42, 43	30, 42, 43
	42	1.25, d (6.0)	22.4	41	41, 43
	43	1.25, d (6.0)	22.4	41	41.42
	<u>CO<sub>2</sub>NH<sub>2</sub></u>		169.6		30, 33'
	29-OH	11.25, s		27, 28	

[0340] <sup>a</sup>重叠信号,\*通过HSQC和HMBC实验支持的分配。

[0341] 表2. 孢囊菌酰胺B (2) 的NMR (700MHz, DMSO-d<sub>6</sub>) 数据

	位置	$\delta_n$ , 多重(J单位 Hz)	$\delta_c$	COSY	HMBC	ROESY
	1	4.31, m <sup>a</sup>	82.0	2	2, 37, <u>CO<sub>2</sub>H</u> , 1-OMe,	2, 3, 36, 1-OMe
	2	5.07, dd (8.1, 5.6)	54.4	1, 3	<u>CO<sub>2</sub>H</u>	1, 1-OMe, 3, 36
	3	8.50 <sup>b</sup>		2	4	1, 2, 6'
	4		166.0			
	5		129.3			
	6, 6'	7.90, m <sup>c</sup>	128.6	7, 7'	6, 6', 8	
	7, 7'	7.90, m <sup>c</sup>	119.8	6, 6'	7, 7'	9
	8		141.7			
	9	10.79, s			7, 7', 10	7', 12'
	10		164.5			
	11		140.5			
[0342]	12, 12'	8.20, d (8.3)	129.6	13, 13'	12, 12', 14, 10	9
	13, 13'	8.38, d (8.3)	123.8	12, 12'	11, 14, 13, 13'	
	14		149.6			
	15		167.2			
	16		125.9			
	17	7.58, s	114.2		15, 18, 19, 21,	38, 40
	18		146.6			
	19		133.5			
	20	8.50 <sup>b</sup> , d (8.4)	119.9	21	16, 18	21
	21	7.59, d (8.4)	123.0	20	15, 17	
	22	10.98, s			20	25, 39
	23		163.9			

	24		116.8			
	25	7.81, d(8.7)	125.2	26	23, 29	22
	26	7.52, d(8.7)	115.6	25	27, 28	30
	27		138.8			
	28		NO			
	29		150.7			
	30	9.62, s		31	33, 33', 26, 41, 43	
	31		164.5			
	32		129.3			
	33, 33'	7.97, d(8.4)	128.6	34, 34'	31, 33, 33'	30, 41, 42, 43
	34, 34'	7.90, m <sup>e</sup>	119.8	33, 33'	34, 34', 32	1-OMe
[0343]	35		141.7			
	36	10.20, s		34, 37	1, 2, 1-OMe	
	37		168.6			
	1-OMe	3.49, s	59.3	1	1, 2, 34, 36	
	38	4.75, spt(6.0)	72.1	39, 40		17
	39	1.38, d(6.0)	22.1	38	38, 40	22
	40	1.38, d(6.0)	22.1	38	38, 39	17
	41	4.30, m <sup>a</sup>	76.1	42, 43		30, 42, 43
	42	1.25, d(6.0)	22.4	41	41, 43	OH
	43	1.25, d(6.0)	22.4	41	41.42	OH, 30, 33'
	CO <sub>2</sub> H		170.7			
	OH	11.22, s		28, 29		

[0344] 表3. 孢囊菌酰胺C(3)的NMR(500MHz, DMSO-d<sub>6</sub>)数据

	位置	$\delta_{\text{H}}$ , 多重 ( $J$ 单位 Hz)	$\delta_{\text{C}}^*$	COSY	HMBC
[0345]	1		167.3		
	2		126.1		
	3	7.57, s	114.1		1, 5
	4		146.8		
	5		133.6		
	6	8.49, d ( $J$ 4)	120.0	7	2, 4
	7	7.58, d ( $J$ 4)	123.0	6	1, 3, 5
	8	10.95, s			6
	9		164.0		
	10		116.0		
	11		150.5		
	12		137.5		
	13		无		

	14	7.65, d (8.7)	114.5	15	10, 12
	15	7.78, d (8.7)	125.3	14	9, 11
	16	9.12, s			14, 17
	17		164.7		
	18		120.4		
	19/19'	7.69, d (8.8)	129.4	20/20'	19/19', 21, 17
[0346]	20/20'	6.62, d (8.8)	113.2	19/19'	18, 20/20'
	21		152.8		
	22	4.75, m	72.0	23/24	
	23/24	1.37, d (6.0)	22.1	22	23/24
	25	4.33, m	75.8	26/27	
	26/27	1.28, d (6.1)	22.5	25	26/27
	OH	11.23, s		25	10

[0347] 无-未观察到,\*通过HSQC和HMBC实验支持的分配。

[0348] 孢囊菌酰胺D (4) 的HRESI (+) MS分析返回了准分子离子 ( $[M+H]^+$ ), 表示 分子式 ( $C_{42}H_{37}O_{14}N_7$ ), 需要28个双键当量。NMR ( $DMSO-d_6$ ) 数据 (表4) 的解释揭示了 磁性相当的芳香质子H-12'/12 ( $\delta_H$  8.17, d, 8.0) 和H-13/13' ( $\delta_H$  8.36, d, 8.0), 它们对应于第一个对位取代的苯环。 $^1H$ - $^1H$  COSY数据的进一步解释反映了两个额外的对位取代苯环的存在, (H-35/35') ( $\delta_H$  7.80, d, 8.1) 和H-36/36' ( $\delta_H$  7.94, d, 8.1); 第二组芳基高度重叠 (H-6/6') 和 (H-7/7') ( $\delta_H$  7.88)。芳香质子的 (H-12/12') 与酰胺羰基C-10 ( $\delta_C$  165.1) 的HMBC相关性判断以及可交换的 (NH) ( $\delta_H$  10.82) 与C-10, C-7/7', 建立了两个对位取代的芳环之间的连接 (图3), 这进一步被NH/H-12和NH/H-7之间ROESY的相关性所证实。COSY数据反映了一个额外的自旋系统, 这基于氧次甲基H-1 ( $\delta_H$  4.08, d, 8.0) 和 $\alpha$ -质子H-2 ( $\delta_H$  4.91, dd, 8.0, 7.7) 到可交换质子 (NH) ( $\delta_H$  8.47)。以下HMBC相关性: (i) 从H-2到三个酰胺羰基 C-4 ( $\delta_C$  166.4)、C-15 ( $\delta_C$  171.8) 和C-32 ( $\delta_C$  169.2); (ii) 从NH ( $\delta_H$  8.48) 到C-4; (iii) 从NH ( $\delta_H$  10.54) 到C-35/35' ( $\delta_C$  119.5), (iv) 从H-6/6'到C-4, 进一步拓展了孢囊菌酰胺D (4) 的部分结构。对于1-D和2-D NMR数据的考虑, 揭示了额外的 1,3,4-三取代和1,2,3,4-四取代的苯环。观察到从芳香质子H-27 ( $\delta_H$  7.55) 和H-29 ( $\delta_H$  7.60) 到羰基C-31 ( $\delta_C$  167.8) 以及季碳C-25 ( $\delta_C$  133.0) 的HMBC相关性, 同时H-30 ( $\delta_H$  8.47, d, 7.0) 和甲氧基信号 ( $\delta_H$  3.96) 被连于与负载有氧的碳C-26 ( $\delta_C$  149.1), 因此揭示了4-氨基-3-甲氧基苯甲酸部分, 其之后被酯化所证实。除此之外, 观察到从可交换质子 (NH) ( $\delta_H$  7.46) 到负载有氧的碳C-1 ( $\delta_C$  80.8), C-18 ( $\delta_C$  141.0) 和芳香碳原子C-22 ( $\delta_C$  116.2) 的HMBC相关性, 其中H-22 ( $\delta_H$  7.48, d, 8.8) 和甲氧基显示了与连接于酰胺羰基C-23 ( $\delta_C$  164.8) 的C-18和H-21 ( $\delta_H$  7.77, d, 8.8) 之间的连接。羟基官能的邻位的甲氧基的存在, 后来被酯化证实 (4a) (图4), 反映了4-氨基-2-羟基-3-甲氧基苯甲酰胺的存在。4-氨基-3-甲氧基苯甲酸和4-氨基-2-羟基-3-甲氧基苯甲酰胺取代基的连接, 被ROESY和从孢囊菌酰胺D二甲酯 (4a) 中观察到的可交换的NH的HMBC相关性所证实。所缺乏的取代基被指定为位于C-14 ( $\delta_C$  150.0) 和羰基C-38。 $\lambda_{max}$  (320nm) 和C-14的低场化学位移揭示了在C-14有硝基取代基, 并且伯胺连接于羰基C-38, 这产生了4的平面结

构。

[0349] 表4.孢囊菌酰胺D (4) 的NMR (700MHz ,DMSO-d<sub>6</sub>) 数据

位置	$\delta_{\text{H}}$ , 多重( <i>J</i> , 单位 Hz)	$\delta_{\text{C}}$	COSY	ROESY	HMBC
1	4. 08, d ( <i>g. Ø</i> )	80. 7	2		32
2	4. 91, dd ( <i>g. Ø, 7. 7</i> )	56. 4	1, 3	33	1, 4, 15, 32
3	8. 47 <sup>a</sup>		2		4
4		166. 4			
5		129. 5			
6/6'	7. 91, m <sup>b</sup>	129. 0	7/7'		4, 8, 6/6'
7/7'	7. 91, m <sup>b</sup>	120. 4	6/6'		5, 7/7'
8		142. 4			
9	10. 82, s			12/12', 7/7'	7, 10
10		165. 1			
11		140. 9			
12/12'	8. 17, d ( <i>g. Ø</i> )	129. 9	13/13'	9	10, 12/12', 14
13/13'	8. 36, d ( <i>g. Ø</i> )	124. 3	12/12'	9	11, 13/13', 14
14		150. 0			
15		171. 8			
16	无				
17		129. 5			
18		141. 0			
19		无			
20		116. 5			
21	7. 77, d ( <i>g. Ø</i> )	125. 8	22		23
22	7. 48, d ( <i>g. Ø</i> )	115. 3	21		18, 20

[0351]	23		164.8			
	24	无				
	25		133.0			
	26		149.1			
	27	7.55, s	111.7			25, 26, 31
	28		126.3			
	29	7.60°, d (7.0)	123.3	30		25, 27, 31
	30	8.47 <sup>a</sup> , d, (7.0)	120.1	29		26, 28
	31		167.8			
	32		169.2			
	33	10.54, s			2, 35/35'	
	34		142.7			
	35/35'	7.80, d, (8.1)	119.5	36/36'	33	35/35', 37
	36/36'	7.94, d, (8.1)	129.3	35/35'		34, 36/36', 38
	37		129.4			
	38		165.5			
	1-OMe	3.30, s	58.4			1
	18-OMe	3.76, s	61.0			18
	26-OMe	3.95, s	56.8			26

[0352] <sup>a,b,c</sup>重叠信号,通过2D HSQC和HMBC实验得到<sup>13</sup>C位移。无-未观察

[0353] 表5. 孢囊菌酰胺D二甲酯 (4a) 的NMR (700MHz, DMSO-d<sub>6</sub>) 数据

[0354]	位置	$\delta_{\text{H}}$ , 多重(J单位 Hz)	$\delta_{\text{C}}$	COSY	ROESY	HMBC
	1	4.10 <sup>a</sup>	80.4	2	3	2
	2	4.92, dd (8.0, 7.8)	56.1	1, 3	3, 33	1, 32
	3	8.50, d (7.8)		2	1, 2, 6/6'	
	4		165.6			
	5		129.4			
	6/6'	7.91, m <sup>b</sup>	128.8	7/7'	3	4, 8
	7/7'	7.91, m <sup>b</sup>	120.1	6/6'		
	8		142.0			
	9	10.82, s			12/12', 7/7'	7/7'
	10		164.8			
	11		140.8			
	12/12'	8.21, d (8.7)	129.7	13/13'	9, 13/13'	10, 12/12', 14
	13/13'	8.39, d (8.7)	124.0	12/12'	12/12'	11, 13/13', 14
	14		149.9			



[0355]	15		无			
	16	9.65, s		18-		38
				OMe, 36/36'		
	17		129.5			
	18		144.7			
	19		152.1			
	20		121.8			
	21	7.88, d (8.8)	126.1	22		19, 23
	22	7.95, d (8.8)	118.9	21		18, 20
	23		162.6			
	24	10.94, s		19-OMe		30
	25		132.8			
	26		148.3			
	27	7.60, s	111.2	26-OMe		25, 29, 31
	28		124.9			
	29	7.67, d (8.6)	123.2	30	30	27
	30	8.61, d (8.6)	119.1	29	29	
	31		166.4			
	32		169.2			
	33	10.59, s		2, 35/35'		
	34		142.8			
	35/35'	7.83, d, (8.1)	119.2	36/36'	33	35/35', 37
	36/36'	7.97, d, (8.1)	129.1	35/35'	16	34, 36/36', 37, 38
	37		129.3			
	38		165.5			
	1-OMe	3.31	58.1			
	18-OMe	3.91, s	61.2	16		18
	19-OMe	4.10 <sup>a</sup> , s	62.0	24		19
	26-OMe	4.05	56.7	27		
	CO <sub>2</sub> Me	3.86, s	52.4			31

[0356] <sup>a, b</sup>重叠信号,通过2D HSQC和HMBC实验得到<sup>13</sup>C位移。无-未观察

[0357] 孢囊菌酰胺E (5) 的HRESI (+) MS分析返回了准分子离子 ([M+H]<sup>+</sup>), 其提示分子式 (C<sub>26</sub>H<sub>23</sub>O<sub>9</sub>N<sub>5</sub>), 需要18个双键当量。<sup>1</sup>H NMR谱图与孢囊菌酰胺D类似, 其主要区别是缺乏对应于4-氨基-3-甲氧基苯甲酸和4-氨基-2-羟基-3-甲氧基苯甲酰基部分的信号。1-D和2-D NMR数据 (表6) 的具体分析, 导致了孢囊菌酰胺E (5) 的平面结构。

[0358] 表6. 孢囊菌酰胺E (5) 的NMR (700MHz, DMSO-d<sub>6</sub>) 数据

	位置	$\delta_{\text{H}}$ , 多重(J单位 Hz)	$\delta_{\text{C}}$	COSY	ROESY	HMBC
[0359]	1	4.08, d(8.2)	80.2	2		1-OMe, 2
	2	4.90, dd(8.2, 7.7)	56.1	1, 3	17	1, 4, 15, 16
	3	8.50, d(7.7)		2	6/6'	4
	4		165.5			
	5		129.2			
	6/6'	7.91, m <sup>a</sup>	128.6	7/7'	3	4, 6/6', 8
	7/7'	7.91, m <sup>a</sup>	120.0	6/6'	9	5, 7/7'
	8		142.0			
	9	10.82, s			7/7', 12/12'	7/7', 10
	10		164.6			
	11		140.5			
	12/12'	8.21, d(8.4)	129.6	13/13'	9	10, 12/12', 14
	13/13'	8.38, d(8.4)	123.9	12/12'		11, 13/13', 14
	14		149.9			
	15		171.2			
	16		168.9			
	17	10.54, s			2, 19/19', 20/20'	16, 19/19'
	18		142.8			
	19/19'	7.77, d(8.2)	119.0	20/20'	17	19/19', 21
	20/20'	7.90, m <sup>a</sup>	130.6	19/19'	17	18, 20/20', 22
	21		125.6			
	22		167.2			
	1-OMe	3.29	58.1			1

[0360] <sup>a</sup>重叠信号,通过2D HSQC和HMBC实验得到<sup>13</sup>C位移。

[0361] 孢囊菌酰胺F(6)的HRESI(+)MS分析返回了准分子离子(M+H)<sup>+</sup>,其符合分子式C<sub>43</sub>H<sub>39</sub>N<sub>7</sub>O<sub>13</sub>,需要28个双键当量(DBE)。NMR(DMSO-d<sub>6</sub>)数据的解释(表7)揭示三组磁性相当的芳族质子,它们可能通过COSY(6/6'和7/7', 12/12'和13/13', 33/33'和34/34')被关联;此外,与所有其他孢囊菌酰胺不同,还揭示了一组磁性相当的芳香质子(26/26'和27/27'),它们也可以通过COSY被关联。上述四组在分子中占有四个对位取代的苯环,而非在所有其他孢囊菌酰胺中发现的三个。仅有一个1,2,3,4-四取代的苯环可以被检测到,其中在芳香质子H-22( $\delta_{\text{H}}$  7.22)与C-18( $\delta_{\text{C}}$  137.1)和C-20( $\delta_{\text{H}}$  114.0)之间的,以及芳香质子H-21( $\delta_{\text{H}}$  7.51)与C-23( $\delta_{\text{C}}$  167.3)之间的HMBC相关性被观察到。质子H-21和H-22可以通过COSY相关性被关联。由于碳C-17、C-19和C-22不能被观察到,所有的肽-片段的HR-MS/MS质量已经建立并且揭示了在各自的片段中存在7个碳、11个质子、1个氮和3个氧,这证实了在这个位置上存在1,2,3,4-取代的对-氨基苯部分(见图1)。HMBC数据进一步证实了H-37( $\delta_{\text{H}}$  4.93)和C-18( $\delta_{\text{C}}$  137.1)的联系。HMBC和COSY数据证实了在分子的两个芳族部分之间相同的连接结构,正如孢囊菌酰胺D中发现的。可交换质子H-9( $\delta_{\text{H}}$  10.82)和C-10( $\delta_{\text{C}}$  163.9)以及C-7/7'( $\delta_{\text{C}}$  119.4)之间,H-3( $\delta_{\text{H}}$  8.49)和C-4( $\delta_{\text{C}}$  165.1)之间,H-31( $\delta_{\text{H}}$  10.56)和C-30( $\delta_{\text{C}}$  168.3)以及C-32( $\delta_{\text{C}}$  141.5)之间,和H-16( $\delta_{\text{H}}$  8.91)与C-36( $\delta_{\text{C}}$  163.1)和C-18( $\delta_{\text{C}}$  137.1)

之间的HMBC相关性,以及H-2 ( $d_H$  4.92) 与可交换质子H-3 ( $d_H$  8.49) 之间的COSY 相关性,以及HRMS片段数据,建立了各个片段之间的系列相关性。硝基基团的位置和芳族链之间的连接基中自由酰胺基的存在,是通过用HR-MS/MS碎片以产生相应片段的总式而建立的。

[0362] 表7:孢囊菌酰胺F (6) 的NMR (700MHz, DMSO- $d_6$ ) 数据

位置	$\delta_H$ 多重 ( $J$ , 单位 Hz)	$\delta_C^*$	COSY	ROESY	HMBC
1	4.10, d (8.08)	79.7	2	1-OMe, 3	1-OMe, 2, 15, 30
2	4.92, dd (4.10, 4.10)	55.9	1, 3	31	1, 4, 15, 30
3	8.49, d (8.14)		2	1	1, 2, 4
4		165.1			
5		128.7			
6/6'	7.91, m <sup>a</sup>	128.1	7/7'		4, 6/6', 8
7/7'	7.91, m <sup>a</sup>	119.4	6/6'	9	5, 7/7'
8		141.6			
9	10.82, s			7/7', 12/12'	7/7', 8, 10
10		163.9			
11		140			
12/12'	8.21, d (8.71)	129.1	13/13'	9	10, 12/12', 14
13/13'	8.39, d (8.71)	123.3	12/12'		11, 13/13'
14		149			
15		170.6			
16	8.91, s			34/34', 38/38'	18, 36
17		无			
18		137.1			
19		无			

	20		114.9		
	21	7.51, d(9.02)	127.5	22	23
	22	7.22, d(9.02)	无	21	18, 20
	23		167.3		
	24	15 非常宽 s			
	25		144.5		
	26/26'	7.78, d(8.57)	118.4	27/27'	26/26', 28
	27/27'	7.86, m <sup>a</sup>	130.1	26/26'	25, 27/27', 29
	28		123.4		
[0364]	29		167.3		
	30		168.3		
	31	10.56, s		2, 33/33'	30, 33/33'
	32		141.5		
	33/33'	7.83, m <sup>a</sup>	118.9	34/34'	33/33', 35
	34/34'	7.87, m <sup>a</sup>	127.5	33/33'	16
	35		129.2		32, 34/34', 36
	36		163.1		
	37	4.93, m <sup>a</sup>	71	38/38'	18
	38/38'	1.21, d(6.18)	22.4	37	16
	1-OMe	3.31, s	57.4	1	1

[0365] <sup>a</sup>重叠信号,无-未观察,\*通过HSQC和HMBC实验支持的分配。

[0366] 孢囊菌酰胺G (7) 的HRESI (+) MS分析返回了准分子离子(M+H)<sup>+</sup>,其符合分子式C<sub>44</sub>H<sub>41</sub>N<sub>7</sub>O<sub>14</sub>,需要28个双键当量(DBE)。由于在DMSO-d<sub>6</sub>中芳香信号重叠,在甲醇-d<sub>4</sub>中获得的NMR数据被用于建立芳香和连接片段的局部结构(表8)。对位取代的苯环可以经由COSY、HSQC和HMBC相关性来确定。通过HSQC、COSY和HMBC相关性建立了1,3,4-三取代苯环(4-氨基-3-甲氧基-苯甲酰胺)和甲氧基取代基(1-OMe, (d<sub>c</sub> 55.2, d<sub>H</sub> 3.50)的结构。由于并非所有1,2,3,4-取代的苯部分的信号均可以在甲醇-d<sub>4</sub>中被检测到,对DMSO-d<sub>6</sub>中测定的NMR数据进行解读,以建立4-氨基-3-异丙基-2-羟基-苯甲酰胺和在孢囊菌酰胺D中所鉴定的芳香部分之间相同的连接结构。C-39(d<sub>c</sub> 74.4)和碳原子C-40/40'(d<sub>c</sub> 22.7)之间的联系通过H-39(d<sub>H</sub> 4.82)和H-40/40'(d<sub>H</sub> 1.31)之间的COSY相关性建立,且1,2,3,4-取代的苯环和H-39(d<sub>H</sub> 4.82)之间的连接,是通过H-39和C-18(d<sub>c</sub> 137.3,在DMSO-d<sub>6</sub>中)之间的HMBC相关性建立的。这个苯部分的结构被DMSO-d<sub>6</sub>中H-22(d<sub>H</sub> 7.04)与C-18(d<sub>c</sub> 137.3)和C-20(d<sub>c</sub> 116.1)之间的HMBC相关性,以及(d<sub>H</sub> 7.45)与C-23(d<sub>c</sub> 165.4)之间的HMBC相关性,以及H-21和H-22之间的COSY相关性进一步证实。整体序列,所述硝基基团的位置和芳香族链之间的连接结构中自由酰胺基团的存在,是在通过HR-MS/MS碎片以产生相应片段的总式而建立的。

[0367] 表8:孢囊菌酰胺G (7) 的NMR (700MHz, 甲醇-d<sub>4</sub>) 数据,包括位置17-23和 39-40/40' (700MHz, DMSO-d<sub>6</sub>) 的数据。

	位置	$\delta_{\text{H}}$ 多重 ( $J$ , 单位 Hz)	$\delta_{\text{C}}^*$	COSY	ROESY	HMBC
[0368]	1	4.17, d(7.45)	82.1	2		1-OMe, 2, 15, 32
	2	5.08, d(7.37)	57.2	1		1, 4, 15, 32
	3	无				
	4		168.9			
	5		130.5			
	6/6'	7.93, m <sup>a</sup>	129.4	7/7'		4, 6/6', 8
	7/7'	7.89, d(8.83)	121.1	6/6'		5, 7/7'
	8		142.9			
	9	无				
	10		166.5			
	11		141.6			
	12/12'	8.16, d(8.77)	129.9	13/13'		10, 12/12', 14
	13/13'	8.38, d(8.74)	124.5	12/12'		11, 13/13'
	14		150.9			
	15		174.4			
	16	无				
	17		139.4			
	18	无	无			
	19		无			
	20		无			
	21	7.74, d(8.83)	125.4	22		23, 17
	22	7.51, 宽 d	无			
	23		168.7			
	24	无				
	25		133.5			
	26		149.9			
	27	7.67, S	112.7			25, 26, 28, 29, 31
	28		131.8			
	29	7.61, d(8.22)	129.9	30		27, 30, 31

	30	8.45, 宽 d	120.5	29	
	31		174.8		
	32		169.5		
	33	无			
	34		142.8		
	35/35'	7.83, d(8.64)	120.8	36/36'	35/35', 37
	36/36'	7.93, m <sup>a</sup>	128.9	35/35'	34, 36/36', 38
	37		131.2		
	38		166.4		
	39	4.82, 水的峰	74.4	40/40'	40
	40/40'	1.31, d(6.13)	22.7	39	39
	1-Ome	3.50, s	55.2		1
[0369]	26-Ome	4.02, s	55.9		26

位置	$\delta_{\text{H}}$ 多重(J, Hz)	$\delta_{\text{C}}^*$	COSY	ROESY	HMBC
17		无			
18		137.3			
19		无			
20		116.1			
21	7.45, d(8.83)	123.9	22		23
22	7.04, d(8.66)	99.7	21		18, 20
23		165.4			
39	5.05, m	69.7	40/40'		18, 40/40'
40/40'	1.17, d(5.98)	22.5	39		39

[0370] <sup>a</sup>重叠信号, 无-未观察, \*通过HSQC和HMBC实验支持的分配。

[0371] 孢囊菌酰胺H(8)的HRESI(+)MS分析返回了准分子离子(M+H)<sup>+</sup>, 其符合分子式C<sub>43</sub>H<sub>39</sub>N<sub>7</sub>O<sub>14</sub>, 需要28个双键当量(DBE)。据发现, 芳香链之间的连接结构与孢囊菌酰胺D中所发现的一致。在DMSO-d<sub>6</sub>中得到的HSQC、HMBC和COSY数据显示出, 三个对位取代苯单元, 这与在孢囊菌酰胺A、B、D、F和G中所发现的相同。对COSY、HSGC和HMBC数据的进一步解释, 显示了相同的1,3,4-三取代苯部分, 它显示出与甲氧基的HMBC相关性, 这与在除了孢囊菌酰胺F之外的所有其他孢囊菌酰胺中发现的情况相同(这通过1-Ome( $\delta_{\text{H}}$  3.27)和C-26( $\delta_{\text{C}}$  147.4)之间的HMBC相关性得以确认)。NMR数据的分析显示了1,2,3,4-取代的苯部分, 这与其他孢囊菌酰胺相一致。显著的改变源自于乙氧基单元的建立, 这是通过亚甲基质子H-39( $\delta_{\text{H}}$  4.17)和甲基基团H-40( $\delta_{\text{H}}$  1.27)之间的COSY相关性, 以及亚甲基基团H-39( $\delta_{\text{H}}$  4.17)和C-18( $\delta_{\text{C}}$  139.5)之间的HMBC相关性得出的, 从而将4-氨基-2-羟基-3-X-苯甲酰胺结构部分的取代方式扩展至: 位置3上X=甲氧基、异丙基或乙氧基。孢囊菌酰胺H的连续序列通过可交换质子H-9( $\delta_{\text{H}}$  10.93)与C-10( $\delta_{\text{C}}$  163.9)和C-7/7'( $\delta_{\text{C}}$  119.6)之间, H-33( $\delta_{\text{H}}$  10.85)与C-32( $\delta_{\text{C}}$  168.7)和C-35/35'( $\delta_{\text{C}}$  118.8)之间, H-16( $\delta_{\text{H}}$  8.91)与C-38( $\delta_{\text{C}}$  163.1)之间, C-18( $\delta_{\text{C}}$  139.5)和C-22( $\delta_{\text{C}}$  100.4)和H-24( $\delta_{\text{H}}$  14.71)与C-20( $\delta_{\text{C}}$  116.1)之间, C-25( $\delta_{\text{C}}$  131.0)、C-26( $\delta_{\text{C}}$  147.4)和C-30( $\delta_{\text{C}}$  118.5)以及H-2( $\delta_{\text{H}}$  4.85)与C-4( $\delta_{\text{C}}$  165.5)之间的

HMBC相关性,以及HR-MS2碎片数据得以建立,HR-MS2碎片数据还使得硝基基团 被定位以及确定了连接部分中的自由酰胺基。

[0372] 表9:孢囊菌酰胺H (8) 的NMR (700MHz, DMSO-d<sub>6</sub>)

位置	$\delta_{\text{H}}$ 多重(J, Hz)	$\delta_{\text{C}}^*$	COSY	ROESY	HMBC
1	4.22, d(8.60)	79.8	2	3, 33	2, 32, 1-OMe
2	4.85, dd(8.42, 8.42)	56.3	1, 3	3, 33	1, 4, 15, 32
3	9.02 s		2		
4		165.5			
5		128.8			
6/6'	7.93 m <sup>a</sup>	128.3	7/7'		4, 6/6', 8
7/7'	7.91 m <sup>a</sup>	119.6	6/6'		5, 7/7'
8		141.7			
9	10.93 s			7/7', 12/12'	
10		163.9			
11		140.3			
12/12'	8.22, d(8.72)	129.4	13/13'		10, 12/12', 14
13/13'	8.38, d(8.72)	123.5	12/12'		11, 13/13'
14		149.2			
15		170.7			
16	8.91 s			22, 39, 40	18, 22, 38
17		无			
18		139.5			
19		无			
20		116.1			
21	7.45, d(8.63)	124.1	22		18, 23
22	6.95, d(8.66)	100.4	21	16	18

[0373]

	23		165.8		
	24	14.71 s		26-OMe, 39	23, 25, 26, 30
	25		131.0		
	26		147.4		
	27	7.46, s	111.1		25, 26, 29, 28, 31
	28		133.9		
	29	7.38, m <sup>a</sup>	121.3	30	27, 28, 30
	30	8.44, d(8.29)	118.5	29	25, 26, 28,
	31		169.9		
[0374]	32		168.7		
	33	10.85 s		1, 2, 35/35'	35/35'
	34		141.9		
	35/35'	7.85, m <sup>a</sup>	118.8	36/36'	37
	36/36'	7.85, m <sup>a</sup>	127.7	35/35'	34, 38
	37		129.5		
	38		163.1		
	39	4.17, q(7.03)	65.4	40	18, 40
	40	1.27, t(7.07)	15.7	39	39
	1-Ome	3.27, s	57.4		1
	26-Ome	3.84, s	55.2		26

[0375] <sup>a</sup>重叠信号,无-未观察,\*通过HSQC和HMBC实验支持的分配。

#### 附图说明:

[0376] 图1:孢囊菌酰胺A (1) 关键的2D NMR相关性 (700MHz, DMSO-d<sub>6</sub>)

[0377] 图2:孢囊菌酰胺C (3) 关键的2D NMR相关性 (500MHz, DMSO-d<sub>6</sub>)

[0378] 图3:孢囊菌酰胺D (4) 关键的2D NMR相关性 (700MHz, DMSO-d<sub>6</sub>)

[0379] 图4:孢囊菌酰胺D二甲酯 (4a) 关键的2D NMR相关性

[0380] 图5:孢囊菌酰胺E (5) 关键的2D NMR相关性

[0381] 图6:孢囊菌酰胺F (6) 关键的2D NMR相关性 (700MHz, DMSO-d<sub>6</sub>)

[0382] 图7:孢囊菌酰胺G (7) 关键的2D NMR相关性 (700MHz, MeOH-d<sub>4</sub>)

[0383] 图8:孢囊菌酰胺H (8) 关键的2D NMR相关性 (700MHz, DMSO-d<sub>6</sub>)

[0384] 孢囊菌酰胺的生物学评价

[0385] 如在表10a/b中所概括,评价了孢囊菌酰胺对于若干微生物和细胞系的作用。所有的衍生物表现出对各种不同大肠杆菌菌株 (包括对抗蔡啶酸 (NAL<sup>R</sup>) 的分离株) 的有效抑制作用。测试衍生物的整体效力按照以下顺序增加 (平均MIC值): CysA1, CysC<CysB<CysA, CysG<CysF。重要的是,在低μg/ml范围内,病原性革兰氏阴性菌株鲍曼不动杆菌和绿脓杆菌也为活性最好的衍生物CysA、CysB、CysG和CysF所抑制,以MIC值表示,仅比参考药物环丙沙星高一个数量级。

[0386] 对革兰氏阳性菌,如粪肠球菌、金黄色葡萄球菌和肺炎链球菌的平均MIC值部分地在低于μg/ml范围,且CysA和CysB的平均效力超过了环丙沙星的效力。



[0387] 另外,它表明孢囊菌酰胺即不抑制酵母,也不抑制哺乳动物细胞的生长。因此,该孢囊菌酰胺没有造成明显的细胞毒性。

[0388] 突变的大肠杆菌菌株对于孢囊菌酰胺的敏感性

[0389] 喹诺酮类药物是一类广泛使用的,针对II型拓扑异构酶、DNA促旋酶和拓扑异构酶IV的抗生素。由此,喹诺酮类耐药性常常是由导致药物靶标改变的染色体基因突变所介导。在GyrA中,喹诺酮类耐药决定区(QRDR)位于氨基酸67和106之间,而氨基酸83(Ser)和87(Asp)是最经常参与的<sup>[1,2]</sup>。在ParC类似的区域中(ParC是喹诺酮类的次要目标),已发现,氨基酸80(Ser)的变化赋予喹诺酮类药物耐药性<sup>[3,4]</sup>。

[0390] 用在gyrA和parC基因具有典型突变的大肠杆菌菌株平板,对孢囊菌酰胺进行筛选(表11)。使用环丙沙星时,MIC值对于单个的gyrA突变提高了大约30个因子(倍)(菌株MI和WT-3.2)。然而,这两种gyrA突变的组合(菌株WT-3)导致了接近临床耐药性(1mg/L)。parC突变(菌株WT-4M2.1)导致环丙沙星的MIC增加了两倍。然而,孢囊菌酰胺的MIC值对于gyrA和parC突变的大肠杆菌菌株没有增加或仅略微增加,这表明相比于环丙沙星中观察到的情况,孢囊菌酰胺可能与gyrA的氨基酸87和83以及ParC的氨基酸80之间的干扰程度较低。

[0391] 高浓度的喹诺酮类药物耐药性,通常是由多个目标位点突变和改变的外排机制相结合所导致的。在体外选定的突变株WT III(marR Δ74bp)不会产生功能性MarR,而MarR的作用是作为marA表达的阻遏子。这反过来,导致MarA和AcrAB的生产过剩,而AcrAB外排泵的过表达是与MAR(多重耐药性)表型相关联的<sup>[5]</sup>。大肠杆菌菌株WT III对于环丙沙星治疗更不敏感,其因子为约4(与大肠杆菌WT相比)。相比之下,孢囊菌酰胺B、F和G的MIC值仍处于μg/ml范围内。值得注意的是,CysF对大肠杆菌菌株WT III的MIC仅较野生型大肠杆菌DSM-1116增加了2个因子(倍),而环丙沙星的MIC增加约10倍。

[0392] 表10a:孢囊菌酰胺(Cys)的抗菌活性

[0393]

测试生物	CysA	CysA1	CysB	CysC
------	------	-------	------	------

[0394]

	MIC [ $\mu\text{g/ml}$ ]			
鲍曼不动杆菌 DSM-30008	7.4	58.9	3.7	32.5
伯克霍尔德氏菌 DSM-16553	> 59	> 59	> 59	> 65
青紫色素杆菌 DSM-30191	> 59	> 59	14.7	16.3
大肠杆菌 DSM-1116	0.9	14.7	1.8	16.3
大肠杆菌 DSM-12242 (NAL <sup>R</sup> )	0.9	29.4	3.7	8.1
大肠杆菌 DSM-26863 ( <i>toIC3</i> )	0.5	7.4	1.8	4.1
大肠杆菌 ATCC35218	0.9	14.7	1.8	16.3
大肠杆菌 ATCC25922	0.5	7.4	0.9	8.1
产气肠杆菌 DSM-30053	> 59	> 59	> 59	> 33
肺炎克雷伯菌 DSM-30104	> 59	> 59	> 59	65
铜绿假单胞菌 PA14	> 59	58.9	14.7	65
铜绿假单胞菌 ATCC27853	> 59	58.9	14.7	65
耻垢分枝杆菌 mc <sup>2</sup> 155 ATCC700084	> 59	> 59	> 59	> 65
枯草芽孢杆菌 DSM-10	0.12	1.8	0.46	2.0
粪肠球菌 ATCC29212	0.06	3.7	0.23	4.1
藤黄微球菌 DSM-1790	0.06	7.4	0.23	4.1
金黄色葡萄球菌 ATCC29213	0.12	14.7	0.12	8.1
肺炎链球菌 DSM-20566	0.23	14.7	0.46	8.1
白色念珠菌 DSM-1665	> 59	> 59	> 59	> 65
异常毕赤酵母 DSM-6766	> 59	> 59	> 59	> 65

[0395]

测试生物	CysF	CysG	CIP
鲍曼不动杆菌 DSM-30008	—	—	0.2
伯克霍尔德氏菌 DSM-16553	—	—	6.4
青紫色素杆菌 DSM-30191	—	—	0.006
大肠杆菌 DSM-1116	0.4	0.9	0.006
大肠杆菌 DSM-12242 (NAL <sup>R</sup> )	—		0.05
大肠杆菌 DSM-26863 ( <i>toIC3</i> )	0.4	0.9	$\leq 0.003$

[0396]	大肠杆菌 ATCC35218	—	—	0.006
	大肠杆菌 ATCC25922	—	—	$\leq 0.003$
	产气肠杆菌 DSM-30053	—	—	0.2
	肺炎克雷伯菌 DSM-30104	—	—	0.025
	铜绿假单胞菌 PA14	3.4	7.1	0.1
	铜绿假单胞菌 ATCC27853	—	—	0.1
	耻垢分枝杆菌 mc <sup>2</sup> 155 ATCC700084	—	—	0.4
	枯草芽孢杆菌 DSM-10	—	—	0.1
	粪肠球菌 ATCC29212	—	—	0.8
	藤黄微球菌 DSM-1790	—	—	1.6
	金黄色葡萄球菌 ATCC29213	—	—	0.1
	肺炎链球菌 DSM-20566	—	—	1.6
	白色念珠菌 DSM-1665	—	—	>6.4
	异常毕赤酵母 DSM-6766	—	—	>6.4

[0397] CIP对照抗生素环丙沙星

[0398] —未测定

[0399] 表10b:孢囊菌酰胺 (Cys) 的细胞毒性

	细胞系和原代细胞	GI <sub>50</sub> [ $\mu$ M]		
		CysA	CysA1	CysB
[0400]	CHO-K1 (中国仓鼠卵巢)	37 -111	> 111	> 111
	HCT-116 (人结肠癌)	—	—	> 50
	HUVEC (人脐静脉内皮细胞)	—	—	> 50

	细胞系和原代细胞	GI <sub>50</sub> [ $\mu$ M]		
		CysC	CysF	CysG
[0401]	CHO-K1 (中国仓鼠卵巢)	ca. 111	> 111	37 -111
	HCT-116 (人结肠癌)	—	—	—
	HUVEC (人脐静脉内皮细胞)	—	—	—

[0402] —未测定

[0403] 表11:孢囊菌酰胺 (Cys) 对大肠杆菌突变菌株的抗菌活性

测试生物体[耐药突变]	CysA	CysA1	CysB	CysC
	MIC [ $\mu\text{g/ml}$ ]			
大肠杆菌 WT	0.5	14.7	1.8	8.1
大肠杆菌 M1 [ <i>gyrA</i> (S83L)]	3.7	29.4	3.7	16.3
大肠杆菌 WT-3.2 [ <i>gyrA</i> (D87G)]	3.7	29.4	3.7	32.5
大肠杆菌 WT-3 [ <i>gyrA</i> (S83L, D87G)]	14.7	> 59	7.4	> 33
大肠杆菌 WT-4 M2.1 [ <i>parC</i> (S80I)]	0.5	14.7	1.8	8.1
大肠杆菌 M1-4 [ <i>gyrA</i> (S83L), <i>parC</i> (S80I)]	0.5	14.7	1.8	16.3
大肠杆菌 WTIII [ <i>marR</i> $\Delta$ 74bp]	14.7	58.9	3.7	65

测试生物体 [耐药突变]	CysF	CysG	CIP
	MIC [ $\mu\text{g/ml}$ ]		
大肠杆菌 WT	—	—	0.013
大肠杆菌 M1 [ <i>gyrA</i> (S83L)]	—	—	0.4
大肠杆菌 WT-3.2 [ <i>gyrA</i> (D87G)]	—	—	0.4
大肠杆菌 WT-3 [ <i>gyrA</i> (S83L, D87G)]	—	—	0.8
大肠杆菌 WT-4 M2.1 [ <i>parC</i> (S80I)]	—	—	0.025
大肠杆菌 M1-4 [ <i>gyrA</i> (S83L), <i>parC</i> (S80I)]	—	—	0.4
大肠杆菌 WTIII [ <i>marR</i> $\Delta$ 74bp]	0.9	3.6	0.05

[0406] CIP对照抗生素环丙沙星

[0407] —未测定

[0408] 基于细胞的测定的实验步骤

[0409] 细胞系和原代细胞。人HCT-116结肠癌细胞(CCL-247)来自于美国典型培养物保藏中心(ATCC),中国仓鼠卵巢CHO-K1细胞(ACC-110)来自于德国微生物和细胞培养物保藏中心(DSMZ)。两种细胞系都在各自保藏者所推荐的条件下进行培养。原代HUVEC(人脐静脉内皮细胞;单个供体)购自PromoCell(海德堡,德国),并在含有如下补充剂的内皮细胞生长培养基(PromoCell)中培养:2%FCS, 0.4%ECGS, 0.1ng/ml EGF, 1ng/ml bFGF, 90 $\mu\text{g/ml}$  肝素, 1 $\mu\text{g/ml}$ 氢化可的松。

[0410] 细菌菌株。在敏感性试验中使用的细菌的野生型菌株,或者是我们菌株收集的一部分,或是从德国微生物和细胞培养保藏中心(DSMZ)或从美国典型培养物保藏中心(ATCC)购得。大肠杆菌菌株WT<sup>[6]</sup>以及大肠杆菌突变株为汉堡大学医药生物和微生物学的P.Heisig教授(博士)惠赠。

[0411] 细胞毒性检测。细胞以 $6 \times 10^3$ 个细胞/每孔接种于96-孔板(Corning CellBind®)上的完全培养基(180 $\mu\text{l}$ )中,并直接用系列稀释的溶解于甲醇的孢囊菌酰胺进行处理。化合物以及内部溶剂对照,一式两份测试5天。在5天孵育后,5mg/ml MTT的PBS(20 $\mu\text{L}$ )溶液被

加入到每孔中,并在37℃下继续孵育2 小时<sup>[7]</sup>。弃去培养基,将细胞用PBS (100μl) 洗涤,然后加入2-丙醇/10N盐酸 (250:1,v/v;100μl) 以溶解甲瓚颗粒。在570nm处的吸光度用酶标仪测定 (EL808,Bio-Tek仪器公司)。

[0412] 敏感性测试。MIC值通过微量稀释试验测定。过夜培养物在适当的生长培养基中稀释,以获得 $10^4$ - $10^6$ cfu/mL的接种物。酵母生长于NYC培养基 (1%植物蛋白胨,1%葡萄糖,50mM HEPES,pH7.0),肺炎链球菌和粪肠球菌在胰蛋白酶大豆 培养基中 (TSB:1.7%蛋白胨酪蛋白,0.3%蛋白胨豆粕,0.25%葡萄糖,0.5%的 NaCl,0.25%K<sub>2</sub>HPO<sub>4</sub>;pH7.3);耻垢分枝杆菌在补充了10%的Middlebrook ADC富 集物和2ml/1甘油的Middlebrook 7H9培养基中。列出的所有其他细菌在米勒 -辛顿培养基 (0.2%牛肉输注固体,1.75%酪蛋白水解物,0.15%的淀粉,pH 7.4) 中培养。孢囊菌酰胺和对照药物被直接加入到在无菌96孔板中的培养物中,一式两份,并制备系列稀释物。微生物生长在微孔板摇床上 (750rpm,30-37℃,18-48h) 中,除了肺炎链球菌,其生长在非振荡条件下 (37℃,5%CO<sub>2</sub>,18小 时)。生长抑制通过目测评估,且MIC被定义为抑制可见生长的化合物的最低浓 度。

[0413] 靶标识别

[0414] 为了测试孢囊菌酰胺的抑制促旋酶的活性,使用商业大肠杆菌旋转酶超螺旋 试剂盒 (Inspiralis)。孢囊菌酰胺A抑制大肠杆菌促旋酶 (20.5nM=1单位) 的表 观IC<sub>50</sub>为6μM。孢囊菌酰胺A1抑制大肠杆菌促旋酶 (20.5nM=1单位) 的表观IC<sub>50</sub>为2.5μM。孢囊菌酰胺D抑制大肠杆菌促旋酶 (20.5nM=1单位) 的表观IC<sub>50</sub>为1 μM。孢囊菌酰胺C抑制大肠杆菌促旋酶 (20.5nM=1单位) 的表观IC<sub>50</sub>为7.7μM。因此,孢囊菌酰胺为细菌DNA促旋酶的新型抑制剂。

[0415] 孢囊菌酰胺A-D在促旋酶抑制活性实验中的IC<sub>50</sub>值:

[0416]	化合物	IC <sub>50</sub> /μM
[0417]	孢囊菌酰胺 A	6±1.4
	孢囊菌酰胺 A1	2.5±0.8
	孢囊菌酰胺 C	7.2±0.74
	孢囊菌酰胺 D	0.7±0.4

[0418] 图9a显示了促旋酶抑制测定的结果。促旋酶反应用不同浓度孢囊菌酰胺A、A1、C和D进行滴定,并通过琼脂糖凝胶电泳进行解析。对于IC<sub>50</sub>测定,超螺旋 质粒的带强度使用 Adobe Photoshop来确定,对[孢囊菌酰胺]进行作图,并用 Hill' s方程拟合。

[0419] 原核DNA促旋酶和拓扑异构酶IV具有高度的同源性,而促旋酶抑制剂通常 显示出拓扑异构酶IV抑制活性<sup>8</sup>。为测试孢囊菌酰胺对于拓扑异构酶IV的影 响,使用了市售的大肠杆菌拓扑异构酶IV试剂盒 (Inspiralis)。

[0420] 孢囊菌酰胺A只在815μM的最高测试浓度下抑制大肠杆菌拓扑异构酶IV的 活性。孢囊菌酰胺A1抑制大肠杆菌拓扑异构酶IV,显示出6.4±1.8μM的IC<sub>50</sub>值。孢囊菌酰胺C只在300μM的最高测试浓度下抑制大肠杆菌拓扑异构酶IV的 活性。孢囊菌酰胺D抑制大肠杆菌拓扑异构酶IV,显示出10±3μM的IC<sub>50</sub>值。

[0421] 孢囊菌酰胺A-D对大肠杆菌拓扑异构酶IV的IC<sub>50</sub>值抑制试验:

[0422]	化合物	IC <sub>50</sub> /μM
	孢囊菌酰胺A	>160
	孢囊菌酰胺A1	6.4±1.8
	孢囊菌酰胺C	>60
	孢囊菌酰胺D	10±3

[0423] 图9b显示了拓扑异构酶IV抑制测定的结果。促旋酶反应用不同浓度A-D 进行滴定,并通过琼脂糖凝胶电泳解析。对于IC<sub>50</sub>测定,超螺旋质粒的带强度使用Adobe Photoshop来确定,对[孢囊菌酰胺]进行作图,并用Hill's方程拟合。

[0424] 原核DNA拓扑异构酶IV和真核拓扑异构酶II具有高度的同源性(IIa型拓扑异构酶),而原核酶抑制剂通常也抑制真核中的对应物<sup>8</sup>。为测试孢囊菌酰胺对于真核拓扑异构酶IV的影响,使用了市售的人拓扑异构酶II试剂盒(Inspiralis)。

[0425] 孢囊菌酰胺A只在815μM的最高测试浓度抑制人拓扑异构酶II的活性。孢囊菌酰胺A1抑制人拓扑异构酶II,显示出9±0.03μM的IC<sub>50</sub>值。孢囊菌酰胺C只在300μM最高测试浓度抑制人拓扑异构酶II的活性。孢囊菌酰胺D抑制人拓扑异构酶II显示出41.2±3μM的IC<sub>50</sub>值。

[0426] 孢囊菌酰胺A-D对人拓扑异构酶II的IC<sub>50</sub>值抑制试验:

[0427]	化合物	IC <sub>50</sub> /μM
	孢囊菌酰胺A	>160
	孢囊菌酰胺A1	9±0.03
	孢囊菌酰胺C	>60
	孢囊菌酰胺D	41.2±3

[0428] 图9c显示了拓扑异构酶II抑制测定的结果。拓扑异构酶II的反应用不同浓度A-D进行滴定,并通过琼脂糖凝胶电泳解析。对于IC<sub>50</sub>测定,超螺旋质粒的带强度使用Adobe Photoshop来确定,对[孢囊菌酰胺]进行作图,并用Hill's方程拟合。

[0429] 除了ATP依赖型IIa型拓扑异构酶如大肠杆菌促旋酶,拓扑酶IV和人拓扑酶II之外,孢囊菌酰胺对于ATP非依赖型人拓扑异构酶I的活性同样被测试。

[0430] 孢囊菌酰胺A-D对人拓扑异构酶I的IC<sub>50</sub>值试验:

[0431]	化合物	IC <sub>50</sub> /μM
	孢囊菌酰胺A	~10
	孢囊菌酰胺A1	~0.7
	孢囊菌酰胺C	~6
	孢囊菌酰胺D	~33.6

[0432] 图9d显示了拓扑异构酶I抑制测定的结果。拓扑I反应用不同浓度A-D进行滴定,并通过琼脂糖凝胶电泳解析。对于IC<sub>50</sub>测定,超螺旋质粒的带强度使用Adobe Photoshop来确定,对[孢囊菌酰胺]进行作图,并用Hill's方程拟合。

[0433] 孢囊菌酰胺A-D的IC<sub>50</sub>(促旋酶)对IC<sub>50</sub>(拓扑异构酶IV)的值的比较:

	IC <sub>50</sub> /μM		比值
	促旋酶	拓扑酶 IV	IC <sub>50</sub> (拓扑酶 IV)/IC <sub>50</sub> (促旋酶)
[0434]	比例		
	孢囊菌酰胺 A	6	~815
	孢囊菌酰胺 A1	2.5	6.4
	孢囊菌酰胺 D	0.7	10
	孢囊菌酰胺 C	7.2	~300
			~42

[0435] 孢囊菌酰胺A和C显示了对于促旋酶分子靶点的强烈倾向(对促旋酶有40–100倍更强的偏好)。A1和D几乎同样好地都靶向螺旋酶和拓扑异构酶IV(对促旋酶有2.6–10倍更强的偏好)。

[0436] 通常,对促旋酶抑制剂有两种描述的抑制模式/结合位点:

[0437] 1.如氟喹诺酮类化合物结合到GyrA DNA复合物,并避免切口的双链DNA的再连接(促旋酶中毒);和

[0438] 2.另一方面氨基香豆素结合于GyrB上的ATP结合口袋(竞争性抑制)<sup>8</sup>。

[0439] 为测试孢囊菌酰胺是否按照这两个抑制模式中任一种,用孢囊菌酰胺D进行DNA/促旋酶复合物的线性化测定(A)和ATP竞争测定(B)。(A)在此,DNA与促旋酶的复合物使用SDS进行捕获,而促旋酶使用蛋白酶K进行消化。如果促旋酶/DNA复合体是由1型的促旋酶抑制剂捕获,这将导致线性化的质粒的形成(因为再连接被抑制)。2型抑制剂-结合或无化合物的样品将不显示出线性化质粒的形成。评估的结果在图10a中显示。环丙沙星(一种已知的促旋酶/DNA稳定剂)和孢囊菌酰胺D显示出,在经蛋白酶K处理后,形成了线性质粒。在未处理的对照组中未显示上述效果。因此,可能似乎孢囊菌酰胺以类似于氟喹诺酮类化合物的方式稳定了共价的GyrA-DNA复合体。(B)在此,标准促旋酶反应被用一恒定量的孢囊菌酰胺D抑制,并用数量增加的ATP进行滴定。如果ATP和孢囊菌酰胺D都竞争结合于促旋酶GyrB亚基上的ATP结合口袋,那么ATP的量增加将导致超螺旋质粒在测定中形成。图10b显示了实验结果。甚至在10mM的最高的ATP浓度下(孢囊菌酰胺浓度的2000倍),促旋酶的活性也没有恢复,这表明ATP结合口袋并不是孢囊菌酰胺的结合位点。这一结果是与线性化的测定结果一致。

[0440] 图11示出了DNA/促旋酶复合线性测定的结果。

[0441] 实验步骤

[0442] 促旋酶超螺旋法

[0443] 为了测试孢囊菌酰胺,使用了市售大肠杆菌旋转酶超螺旋试剂盒(Inspiralis, 诺维奇,英国)。3对于标准反应,将0.5μg松弛型质粒与1单位(~20.5nM)的大肠杆菌促旋酶混合于1×反应缓冲液中(30μl最终体积,参见试剂盒手册),并在37℃下温育30分钟。该反应通过加入含有10%(w/v) SDS的DNA凝胶上样缓冲液进行淬灭。将样品在0.8%(w/v)琼脂糖凝胶进行分离,并使用Roti-凝胶染色(Carl Roth)观察DNA。

[0444] 所有天然产物原液和稀释液在100%DMSO中制备,并加入到超螺旋反应中,得到DMSO终浓度为5%(v/v)。环丙沙星原液和稀释液制备于10mM HCl和50%的DMSO中,并在最后的分析中按1:10使用。

[0445] 在测定中使用以下的天然产物浓度:

[0446] 孢囊菌酰胺A:815.8 $\mu$ M;163 $\mu$ M;80 $\mu$ M,16 $\mu$ M;8 $\mu$ M;1.6 $\mu$ M;0.8 $\mu$ M;0.16 $\mu$ M; 0.08 $\mu$ M; 0.016 $\mu$ M

[0447] 孢囊菌酰胺A1:543.5 $\mu$ M;108.7 $\mu$ M;54 $\mu$ M;10.8 $\mu$ M;5.4 $\mu$ M;1.087 $\mu$ M; 0.54 $\mu$ M;0.108 $\mu$ M;0.054 $\mu$ M;0.0108 $\mu$ M

[0448] 孢囊菌酰胺C:300 $\mu$ M;60 $\mu$ M;30 $\mu$ M;6 $\mu$ M;3 $\mu$ M;0.6 $\mu$ M;0.3 $\mu$ M;0.06 $\mu$ M; 0.03 $\mu$ M;0.006 $\mu$ M

[0449] 孢囊菌酰胺D:347 $\mu$ M;173.5 $\mu$ M;86.75 $\mu$ M;43.38 $\mu$ M;21.69 $\mu$ M;10.84 $\mu$ M; 5.42 $\mu$ M; 2.71 $\mu$ M;1.36 $\mu$ M;0.68 $\mu$ M;0.34 $\mu$ M;0.17 $\mu$ M;0.085 $\mu$ M;0.042 $\mu$ M; 0.021 $\mu$ M;0.0106 $\mu$ M;0.0053 $\mu$ M

[0450] 对照组反应为:在5% (v/v) DMSO存在下进行标准反应,无酶。

[0451] 在室温,DNA不存在情况下,所有反应样品平衡10分钟。然后加入松弛型质粒开始反应。

[0452] 蛋白酶K线性化分析

[0453] 为测试孢囊菌酰胺是否稳定DNA促旋酶和切口的DNA底物之间的共价复合物,进行蛋白酶K线性化测定(见a)。

[0454] 在孢囊菌酰胺D(18 $\mu$ M;1.8 $\mu$ M)存在下,进行标准的促旋酶超螺旋实验。对照反应不含促旋酶,没有抑制或已知的促旋酶/DNA复合物稳定剂环丙沙星(1 $\mu$ M)。反应通过加入1/10体积的10% SDS进行淬灭。为了线性化带切口的DNA促旋酶复合物,将50 $\mu$ g/ml蛋白酶K加入到该反应中,并在37 $^{\circ}$ C下孵育30分钟。将样品用0.8% (w/v) 琼脂糖凝胶进行分离,并使用Roti-凝胶染色(Carl Roth)观察DNA。为了检测线性化质粒带,松弛型质粒用单切割型限制性内切酶Nde I消化。

[0455] 用浓度变化ATP进行解旋酶超螺旋测试

[0456] 为测试孢囊菌酰胺是否与ATP竞争结合于GyrB的ATP结合口袋,进行了不同浓度ATP的标准促旋酶超螺旋测定(见a)。标准反应混合物(1mM ATP)中分别补加ATP(0.5M ATP储备溶液,ATP购自Sigma-Aldrich公司)至ATP终浓度为2.5;5和10mM。所有反应一式三份进行。

[0457] 拓扑异构酶IV的松弛测试

[0458] 为了测试孢囊菌酰胺的抗拓扑异构酶IV活性,使用了市售的大肠杆菌拓扑异构酶IV松弛试剂盒(Inspiralis,诺维奇,英国)。4对于标准反应,将0.5 $\mu$ g 松弛型质粒与1单位( $\sim$ 20.5nM)的大肠杆菌拓扑异构酶IV混合于1 $\times$ 反应缓冲液中(参见试剂盒手册),并在37 $^{\circ}$ C下温育30分钟。该反应通过加入含有10% (w/v) SDS的DNA凝胶上样缓冲液进行淬灭。将样品用0.8% (w/v) 琼脂糖凝胶分离,并使用Roti-凝胶染色(Carl Roth)观察DNA。

[0459] 在测定中使用以下的天然产物浓度:

[0460] 孢囊菌酰胺A:815.8 $\mu$ M;163 $\mu$ M;80 $\mu$ M,16 $\mu$ M;8 $\mu$ M;1.6 $\mu$ M;0.8 $\mu$ M;0.16 $\mu$ M; 0.08 $\mu$ M; 0.016 $\mu$ M

[0461] 孢囊菌酰胺A1:543.5 $\mu$ M;108.7 $\mu$ M;54 $\mu$ M;10.8 $\mu$ M;5.4 $\mu$ M;1.087 $\mu$ M; 0.54 $\mu$ M;0.108 $\mu$ M;0.054 $\mu$ M;0.0108 $\mu$ M

[0462] 孢囊菌酰胺C:300 $\mu$ M;60 $\mu$ M;30 $\mu$ M;6 $\mu$ M;3 $\mu$ M;0.6 $\mu$ M;0.3 $\mu$ M;0.06 $\mu$ M; 0.03 $\mu$ M;0.006 $\mu$ M

[0463] 孢囊菌酰胺D:347 $\mu$ M;173.5 $\mu$ M;86.75 $\mu$ M;43.38 $\mu$ M;21.69 $\mu$ M;10.84 $\mu$ M; 5.42 $\mu$ M;



2.71 $\mu$ M;1.36 $\mu$ M;0.68 $\mu$ M;0.34 $\mu$ M;0.17 $\mu$ M;0.085 $\mu$ M;0.042 $\mu$ M; 0.021 $\mu$ M;0.0106 $\mu$ M;0.0053 $\mu$ M  
[0464] 对照组反应为:在5% (v/v) DMSO存在下进行标准反应,无酶。在室温,DNA 不存在情况下,所有反应样品平衡10分钟。然后加入松弛型质粒开始反应。

[0465] 拓扑异构酶II的松弛测试

[0466] 为了测试孢囊菌酰胺的抗拓扑异构酶II活性,使用了市售的人拓扑异构酶 IV松弛试剂盒(Inspiralis,诺维奇,英国)。4对于标准反应,将0.5 $\mu$ g超螺旋 质粒与1单位( $\sim$ 20.5nM)的大肠杆菌拓扑异构酶II混合于1 $\times$ 反应缓冲液中(参见试剂盒手册),并在37 $^{\circ}$ C下温育30分钟。该反应通过加入含有10% (w/v) SDS的 DNA凝胶上样缓冲液进行淬灭。将样品用0.8% (w/v) 琼脂糖凝胶分离,并使用 Roti-凝胶染色(Carl Roth)观察DNA。

[0467] 在测定中使用以下的天然产物浓度:

[0468] 孢囊菌酰胺A:815.8 $\mu$ M;163 $\mu$ M;80 $\mu$ M,16 $\mu$ M;8 $\mu$ M;1.6 $\mu$ M;0.8 $\mu$ M;0.16 $\mu$ M; 0.08 $\mu$ M; 0.016 $\mu$ M

[0469] 孢囊菌酰胺A1:543.5 $\mu$ M;108.7 $\mu$ M;54 $\mu$ M;10.8 $\mu$ M;5.4 $\mu$ M;1.087 $\mu$ M; 0.54 $\mu$ M;0.108 $\mu$ M;0.054 $\mu$ M;0.0108 $\mu$ M

[0470] 孢囊菌酰胺C:300 $\mu$ M;60 $\mu$ M;30 $\mu$ M;6 $\mu$ M;3 $\mu$ M;0.6 $\mu$ M;0.3 $\mu$ M;0.06 $\mu$ M; 0.03 $\mu$ M;0.006 $\mu$ M

[0471] 孢囊菌酰胺D:347 $\mu$ M;173.5 $\mu$ M;86.75 $\mu$ M;43.38 $\mu$ M;21.69 $\mu$ M;10.84 $\mu$ M; 5.42 $\mu$ M; 2.71 $\mu$ M;1.36 $\mu$ M;0.68 $\mu$ M;0.34 $\mu$ M;0.17 $\mu$ M;0.085 $\mu$ M;0.042 $\mu$ M; 0.021 $\mu$ M;0.0106 $\mu$ M;0.0053 $\mu$ M

[0472] 对照组反应为:无酶,在5% (v/v) DMSO存在下进行标准反应。在室温,DNA不存在情况下,所有反应样品平衡10分钟。然后加入松弛型质粒开始反应。

[0473] 拓扑异构酶I的松弛测试

[0474] 为了测试孢囊菌酰胺的抗拓扑异构酶II活性,使用了市售的人的拓扑异构酶I松弛试剂盒(Inspiralis,诺维奇,英国)。4对于标准反应,将0.5 $\mu$ g松弛 型质粒与1单位( $\sim$ 20.5nM)的智人的拓扑异构酶I混合于1 $\times$ 反应缓冲液中(参见试剂盒手册),并在37 $^{\circ}$ C下温育30分钟。该反应通过加入含有10% (w/v) SDS的 DNA凝胶上样缓冲液进行淬灭。将样品用0.8% (w/v) 的琼脂糖凝胶进行分离,并使用Roti-凝胶染色(Carl Roth)观察DNA。

[0475] 在测定中使用以下的天然产物浓度:

[0476] 孢囊菌酰胺A:815 $\mu$ M;81.5 $\mu$ M;8.15 $\mu$ M

[0477] 孢囊菌酰胺A1:543 $\mu$ M;54.3 $\mu$ M;5.43 $\mu$ M

[0478] 孢囊菌酰胺C:300 $\mu$ M;30 $\mu$ M;3 $\mu$ M

[0479] 孢囊菌酰胺D:277 $\mu$ M;27.2 $\mu$ M;2.77 $\mu$ M

[0480] 对照组反应为:无酶,在5% (v/v) DMSO存在下进行标准反应。在室温,DNA 不存在情况下,所有反应样品平衡10分钟。然后加入松弛型质粒开始反应。

[0481] 定量和分析

[0482] 为了确定IC<sub>50</sub>值,超螺旋的(螺旋酶)或宽松型(拓扑异构酶I,II,IV)质粒 的形成使用Adobe Photoshop(直方图模式)量化。这些值相对于化合物浓度的作图,产生了S形曲线的形状,这是用Hill's方程拟合的(Origin Pro 8.5)。所有确定的IC<sub>50</sub>值是三次独立实验的平均值。

[0483] 参考文献:

[0484] [1] T. Gruger, J. L. Nitiss, A. Maxwell, E. L. Zechiedrich, P. Heisig, S. Seeber, Y. Pommier, D. Strumberg, 抗微生物剂化疗 (Antimicrob. Agents Chemother.) 48, 2004, 4495-4504.

[0485] [2] H. Schedletzky, B. Wiedemann, P. Heisig, 抗微生物化学治疗杂志 (J. Antimicrob. Chemother.) 43, 1999, 31-37.

[0486] [3] A. B. Khodursky, E. L. Zechiedrich, N. R. Cozzarelli, 美国科学院院报 (Proc. Natl. Acad. Sci. USA) 92, 1995, 11801-11805.

[0487] [4] A. Schulte, P. Heisig, 抗微生物化学治疗杂志 (J. Antimicrob. Chemother.) 46, 2000, 1037-1046.

[0488] [5] D. Keeney, A. Ruzin, F. McAleese, E. Murphy, P. A. Bradford, 抗微生物化学治疗杂志 (J. Antimicrob. Chemother.) 61, 2008, 46-53.

[0489] [6] P. Heisig, H. Schedletzky, H. Falkenstein-Paul, 抗微生物剂化疗 (Antimicrob. Agents Chemother.) 37, 1993, 669-701.

[0490] [7] T. Mosmann, 免疫学方法杂志 (J. Immunol. Meth.) 65, 1983, 55-63.

[0491] [8] Pommier, Y.; Leo, E.; Zhang, H.; Marchand, C. 化学和生物学 (Chemistry & Biology) 2010, 17, 421.

[0492] 孢囊菌酰胺A和C的合成

[0493] 首先, 对于孢囊菌酰胺C的合成的描述, 可进一步阐述其他的孢囊菌酰胺。

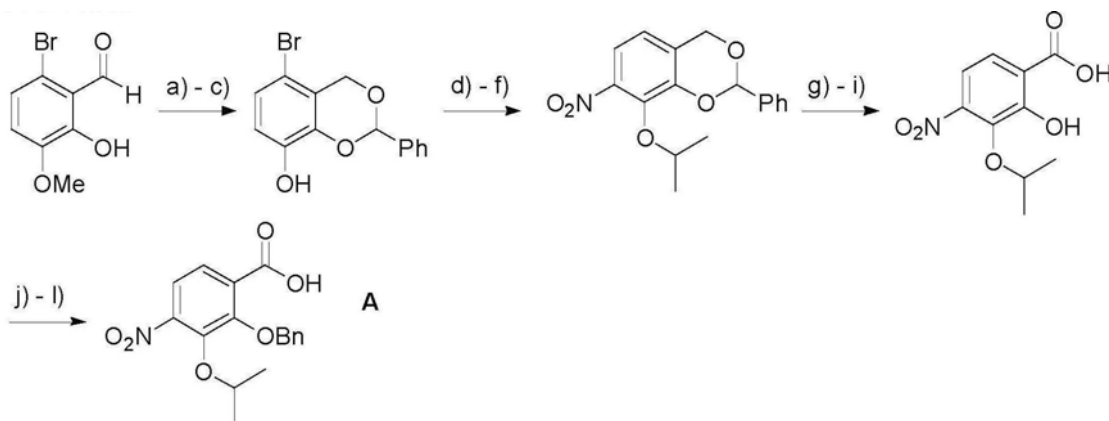
[0494] 1.1. 孢囊菌酰胺C

[0495] 以下方案1和2提供了合成单个芳香构建单元, 随后通过组装这些构建单元生成孢囊菌酰胺C的概述。

[0496] 可选地, 方案1中的步骤e) 可以通过使用其他的醇 ( $R'OH$ ) 替代 $iPrOH$ 进行改变。如果, 例如使用了 $EtOH$ , 则可以制备得到孢囊菌酰胺H的构建单元。同样的也适用于方案1中的第二个反应顺序中的步骤b)。在此,  $iPrOH$ 同样也可以用其他的醇 ( $R'OH$ ) 替代。如果, 例如使用了 $MeOH$ , 则可以制备得到孢囊菌酰胺C、G和H的构建单元。对于孢囊菌酰胺F的制备, 对氨基苯甲酸衍生物如对氨基苯甲酸或相应的N-保护的氨基苯甲酸衍生物, 以及对-硝基苯甲酸被使用以代替构建单元B。

[0497] 方案1: 合成芳烃A和B然后酰胺偶合。

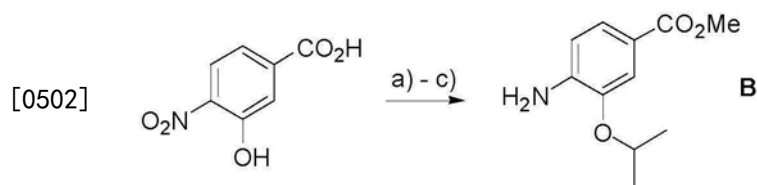
[0498] (中央芳香部分)



[0500] a)  $BBr_3$ ,  $CH_2Cl_2$ ,  $-40^\circ C$ -rt, 17h (95%); b)  $NaBH_4$ , THF,  $-40^\circ C$ -rt, 30min (91%); c)

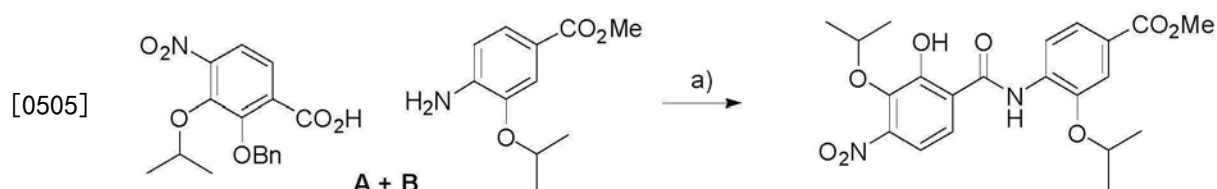
PhCH(OMe)<sub>2</sub>, pTSA.H<sub>2</sub>O, THF, rt, 5天(56%); d) Ni(NO<sub>3</sub>)<sub>2</sub>·5H<sub>2</sub>O, pTsOH.H<sub>2</sub>O, 丙酮, rt, 2.5h(74%); e) <sup>i</sup>PrOH, DEAD, PPh<sub>3</sub>, THF, rt, 17h(85%); f) Pd<sub>2</sub>(dba)<sub>3</sub>, (PhO)<sub>3</sub>P, <sup>i</sup>PrOH, 二氧六环, 80℃, 1.5h(70%); g) 炭酮-1炭酮磺酸, CH<sub>2</sub>Cl<sub>2</sub>/MeOH(1:2), 0℃-rt, 17h(90%); h) MnO<sub>2</sub>, CH<sub>2</sub>Cl<sub>2</sub>, rt, 17h(81%); i) 2-甲基-2-丁烯, NaClO<sub>2</sub>/NaH<sub>2</sub>PO<sub>4</sub>, <sup>t</sup>BuOH, rt, 17h(75%); j) TMSCHN<sub>2</sub>, MeOH/PhMe, 0℃-rt, 30min(57%); k) BnOH, DEAD, PPh<sub>3</sub>, THF, rt, 17h(90%); l) LiOH, THF/H<sub>2</sub>O(1:1), rt, 17h(99%) .

[0501] (末端三取代芳香部分)



[0503] a) TMSCHN<sub>2</sub>, MeOH/PhMe, 0℃-rt, 30min(90%); b) <sup>i</sup>PrOH, DEAD, PPh<sub>3</sub>, THF, rt, 17h(定量); c) Pd/C, MeOH, H<sub>2</sub>atm., rt, 17h(定量) .

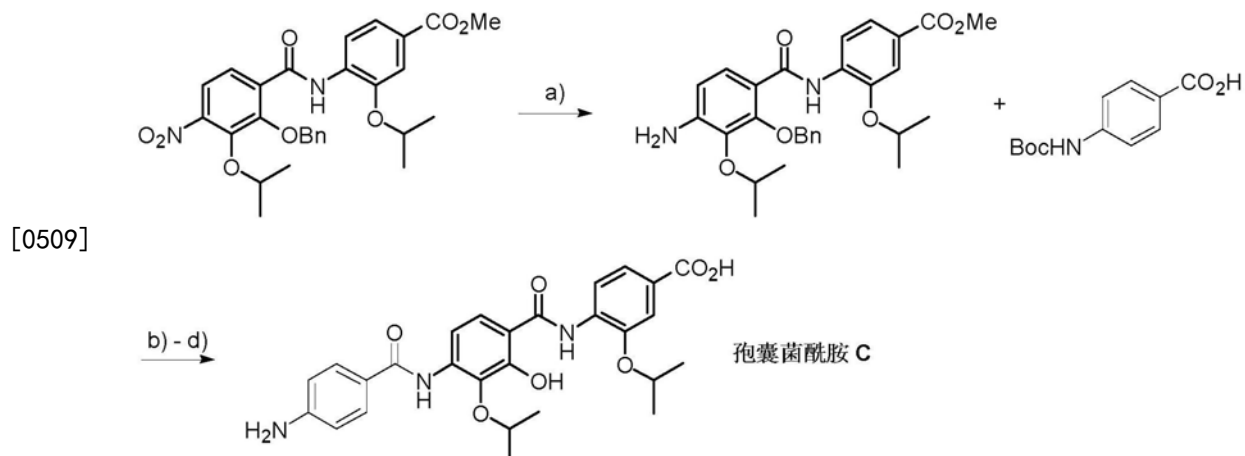
[0504] (耦合芳香部分A和B)



[0506] a) I. A, Goshez's 试剂, CH<sub>2</sub>Cl<sub>2</sub>, 40℃, 3h; II. B, DIPEA, CH<sub>2</sub>Cl<sub>2</sub>, rt, 10min, 然后 I., 40℃, 2天(68%)

[0507] 方案2:最终合成孢囊菌酰胺C。

[0508] 孢囊菌酰胺C(最终合成)



[0510] a) Pd/C, MeOH, H<sub>2</sub>atm., rt, 3h(96%); b) I. 4-Boc-氨基苯甲酸, Goshez' 试剂, CH<sub>2</sub>Cl<sub>2</sub>, rt, 1h; II. B, DIPEA, CH<sub>2</sub>Cl<sub>2</sub>; 然后 I., rt, 1day(7天); c) TFA/CH<sub>2</sub>Cl<sub>2</sub>(10:1), rt, 17h(定量); d) LiOH, THF/H<sub>2</sub>O(1:1), rt, 17h(99%) .

[0511] 1.2孢囊菌酰胺A

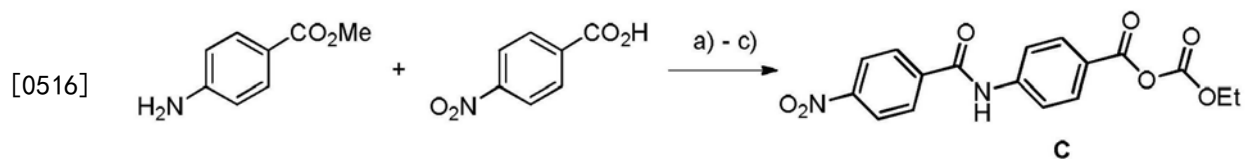
[0512] 更复杂的孢囊菌酰胺由双酰胺(表示为孢囊菌酰胺C), 二芳基酰胺(片段C) 和手性接头元件组成。本节中, 首先报道片段C和手性接头元件, 随后组装所有三个元件从而提

供孢囊菌酰胺A。

[0513] 1.2.1二芳烃C的合成。

[0514] 方案3:活化片段C的合成。

[0515] 片段C



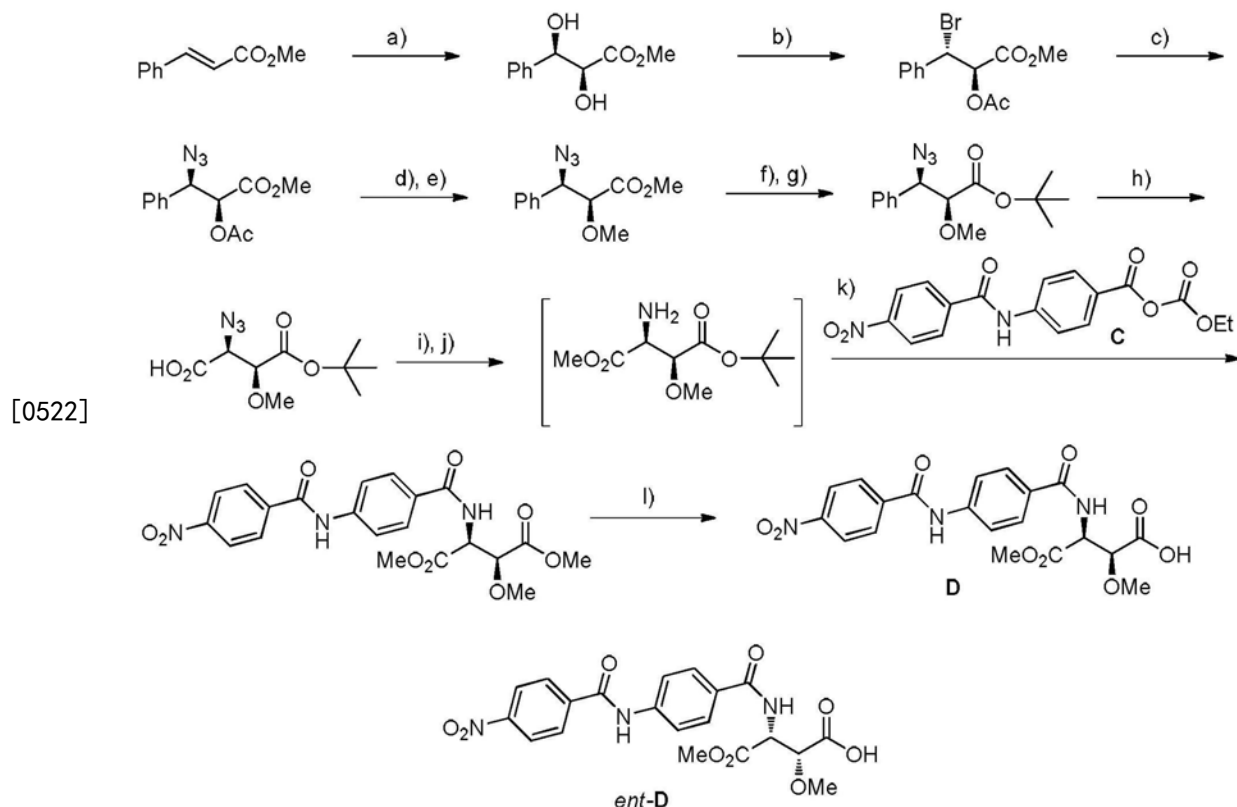
[0517] a)  $P(OMe)_3$ ,  $I_2$ , THF, 3天 (75%) ; b)  $LiOH$ , THF/ $H_2O$  (1:1), rt, 17h (80%) ;

[0518] c) 氯甲酸乙酯,  $Et_3N$ ,  $CH_3CN$ ,  $0^\circ C$  30min (67%)。

[0519] 1.2.2合成连接有二芳烃C的手性构建单元D

[0520] 该合成从肉桂酸甲酯起始,手性通过Sharpless不对称二羟基化引入。该苯环作为第二羧酸的保护基,其不会被氧化释放。最后,构建单元C与自由氨基基团相连接。相应的对映体片段(ent)-D通过使用AD混合物 $\alpha$ 替代AD混合物 $\beta$ 进行制备。

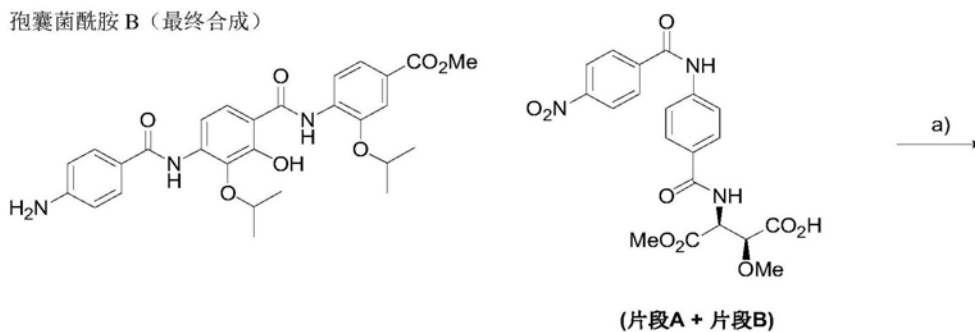
[0521] 方案4:从肉桂酸甲酯起始合成羧酸D。



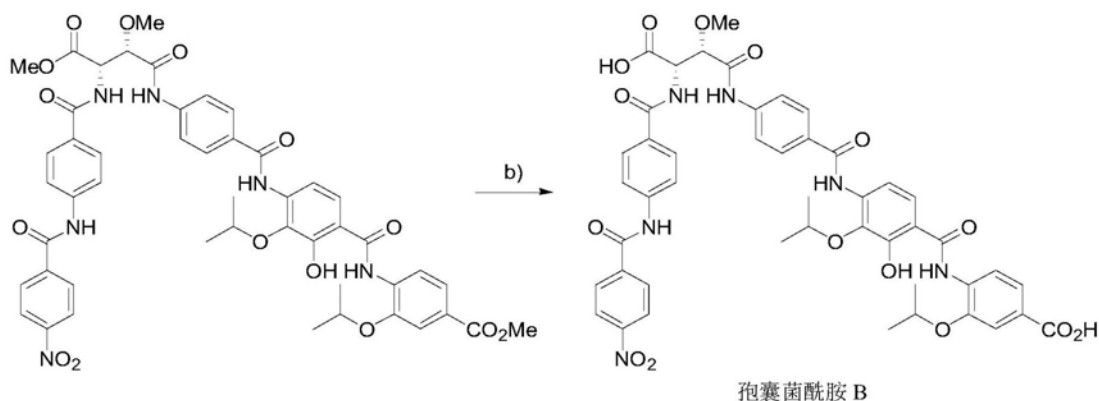
[0523] a) AD混合物 $\beta$ ,  $MeSO_2NH_2$ ,  $tBuOH/H_2O$  (1:1),  $0^\circ C$ , 12h, 然后  $25^\circ C$ , 12h, (79%, ee > 99%) ; b) 33%  $HBr/HOAc$ ,  $45^\circ C$ , 30min., (71%) ; c)  $NaN_3$ , DMF,  $25^\circ C$ , 3h, 然后  $40^\circ C$ , 2h, (89%) ; d)  $KOH$ , THF/ $H_2O$ ; e) 2.  $MeI$ ,  $Ag_2O$ ,  $CaSO_4$  (两步骤74%) ; f)  $KOH$ , THF/ $H_2O$ ; g)  $Me_2N-CH(tBu)_2$ , 甲苯,  $80^\circ C$ , (两步骤87%) ; h)  $RuCl_3 \cdot H_2O$ ,  $NaIO_4$ ,  $CHCl_3/CH_3CN/H_2O$ ,  $70^\circ C$ ; i)  $MeI$ ,  $Ag_2O$ ,  $CaSO_4$ ; j)  $Ph_3P$ , THF/ $H_2O$ ,  $50^\circ C$ , k) DMF (四步骤16%) ; l)  $CF_3CO_2H$ ,  $CH_2Cl_2$ , (定量)。

[0524] 方案5:最终合成孢囊菌酰胺A。

孢囊菌酰胺 B (最终合成)



[0525]



[0526] a) HOAt, EDC·HCl, DIPEA, CH<sub>2</sub>Cl<sub>2</sub>, rt, 17h (75%); b) LiOH, THF/H<sub>2</sub>O (1/1), rt, (95%).

## [0527] 2. 实验

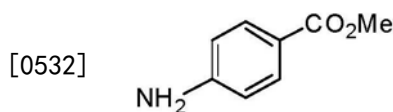
### [0528] 2.1 常规实验信息

[0529] 除非另有说明,所有反应均在烘箱干燥的玻璃器皿中,在氮气气氛下进行。<sup>1</sup>H-NMR谱用Bruker AVS-400,在400MHz下记录,或用Bruker DRX-500,在500 MHz下记录。<sup>13</sup>C-NMR谱用Bruker AVS-400,在100MHz下记录,或用Bruker DRX-500,在125MHz下记录。多重性使用以下缩写描述:s=单峰,d=双峰,t=三重峰,q=四重峰,m=多重峰,b=宽峰。<sup>1</sup>H和<sup>13</sup>C NMR谱的化学位移值通常报告为相对于作为内标的残留溶剂信号的ppm值。重数指在偏共振解耦光谱的共振。这些通过使用无畸变极化转移增强(DEPT)光谱编辑技术,次级脉冲在90°和135°阐明。多重性使用以下缩写报导:s=单重峰(由于季碳),d=双重峰(次甲基),t=三重峰(亚甲基),q=四重峰(甲基)。质谱(EI)在70eV下,将VG Autospec型分光计(质谱仪)、LCT型(ESI)(质谱仪)或Q-TOF型(质谱仪)光谱仪与Waters Aquity超高效液相色谱系统组合使用获得。使用预涂硅胶60F<sub>254</sub>板(Merck,达姆施塔特Darmstadt)进行薄层色谱法分析,在254nm紫外光下显示斑点,或可选地用高锰酸钾、磷钼酸、2,4-二硝基苯酚或对-茴香醛溶液进行染色。四氢呋喃(THF)在氮气下从钠/苯甲酮蒸馏得到。二氯甲烷(CH<sub>2</sub>Cl<sub>2</sub>)使用溶剂纯化系统(SPS)进行干燥。使用供应的市售可得的试剂。制备高效液相色谱法使用Merck Hitachi LaChrom系统(泵L-7150,接口D-7000,二极管阵列检测器L-7450(λ=220-400nm,优选在监测λ=230nm))和柱(在括号中给出的实验部分提到的缩写):Trentec Reprosil-PUR120C18AQ 5μm,250×8mm,保护柱,40×8 mm(C18-SP)。用Merck硅胶60(230-400目)进行快速柱色谱。用于快速色谱法的洗脱剂在使用前蒸馏。使用的SRS

OptiMelt装置测量熔点。旋光度 $[\alpha]$ 用旋光计 341 (Perkin Elmer公司),在波长为589nm下测量,以 $10^{-1}$ 度 $\text{cm}^2\text{g}^{-1}$ 示出。

[0530] 2.2具体步骤

[0531] 4-氨基甲基苯甲酸



[0533] 在烧瓶中加入甲醇(200ml)溶液,并缓慢加入乙酰氯(2.6mL,36.5mmol,1 eq)。然后加入4-氨基苯甲酸(5.00g,36.5mmol),并在室温下将该溶液搅拌7天。减压下除去溶剂,得到米色固体4-氨基甲基苯甲酸酯(5.38g,35.59mmol,定量)。

[0534] 达到熔点前,题述化合物被分解。

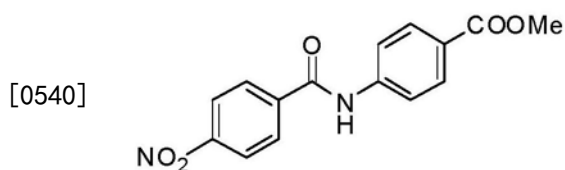
[0535] ATR-IR (净值):  $\tilde{\nu}$  = 2828, 2015, 1724, 1612, 1558, 1508, 1430, 1316, 1280, 1181, 1109, 1072, 1022, 984, 959, 853, 786, 757, 686, 653 $\text{cm}^{-1}$ 。

[0536]  $^1\text{H}$ -NMR (400MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  8.19–8.13 (m, 2H), 7.53–7.48 (m, 2H), 3.93 (s, 3H) ppm.

[0537]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  167.2, 137.0, 132.4, 131.7, 124.2, 53.0 ppm

[0538] HRMS (ESI):  $\text{C}_8\text{H}_{10}\text{NO}_2$  (M+H) $^+$ 计算值:152.0712,测定值:152.0706.

[0539] 4-(4-硝基苯甲酰氨基)苯甲酸甲酯



[0541]  $\text{P}(\text{OMe})_3$  (3.5mL, 29.8mmol)的 $\text{CH}_2\text{Cl}_2$  (100mL)溶液在冰浴中冷却,然后加入 $\text{I}_2$  (7.56g, 29.8mmol)。固体碘完全溶解后,相继加入对硝基苯甲酸(5.52g, 29.8 mmol)和 $\text{Et}_3\text{N}$  (4.70mL, 33.7mmol),溶液在冰浴中搅拌10分钟。加入4-氨基苯甲酸甲酯(3.00gr, 19.9mmol),并将该混合物搅拌10分钟。除去冷却浴后,在室温下,将反应混合物搅拌3天,然后用饱和 $\text{NaHCO}_3$ 水溶液稀释,并用二氯甲烷(3x)萃取。合并的有机相依次用 $\text{H}_2\text{O}$ , 1M  $\text{HCl}$ ,  $\text{H}_2\text{O}$ 和盐水洗涤。合并的有机层用无水 $\text{MgSO}_4$ 干燥,溶剂在真空下浓缩,得到米色固体题述化合物(4.4g, 14.65 mmol, 75%)。

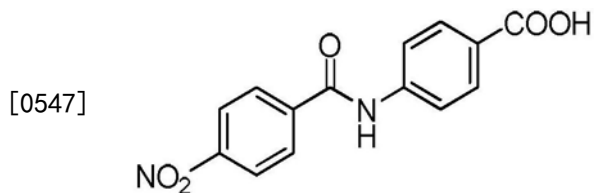
[0542] mp: 245–246 $^{\circ}\text{C}$

[0543]  $^1\text{H}$  NMR (400MHz,  $\text{DMSO}$ )  $\delta$  10.87 (s, 1 $\text{H}_{\text{NH}}$ ), 8.39 (d,  $J$  = 8.8Hz, 2H), 8.20 (d,  $J$  = 8.8Hz, 2H), 7.99 (d,  $J$  = 8.8Hz, 2H), 7.95 (d,  $J$  = 8.8Hz, 2H), 3.84 (s, 3 $\text{H}_{\text{OMe}}$ ) ppm.

[0544]  $^{13}\text{C}$  NMR (100MHz,  $\text{DMSO}$ )  $\delta$  166.2, 164.9, 149.77, 143.6, 140.7, 130.7, 129.8, 125.3, 124.2, 120.2, 52.4 ppm.

[0545] HRMS (ESI):  $\text{C}_{15}\text{H}_{13}\text{N}_2\text{O}_2\text{Na}$  (M+H) $^+$ 计算值:301.0824,测定值:301.0828.

[0546] 4-(4-硝基苯甲酰氨基)苯甲酸



[0548] 4-(4-硝基苯甲酰氨基)苯甲酸甲酯(4.32g,14.38mmol)溶于THF/H<sub>2</sub>O(77/77 mL)的1/1混合物中。然后,加入固体LiOH(5.16g,215.66mmol),并将该体系在室温下搅拌17小时。加入1M HCl直到pH~1,过滤所得的固体,真空下干燥。得到浅黄色固体题述化合物(3.3g,11.54mmol,80%)。

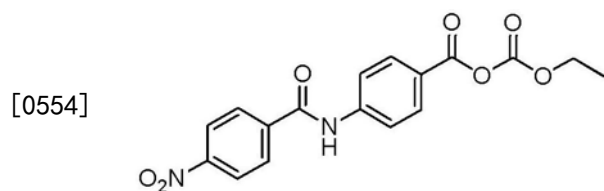
[0549] mp:322-324℃

[0550] <sup>1</sup>H NMR(400MHz,C<sub>6</sub>D<sub>6</sub>) δ10.83(s,1H<sub>CO2H</sub>),8.34(d,J=8.6Hz,1H),8.29(d,J=8.6Hz,1H),8.13(d,J=8.6Hz,1H),8.06(d,J=8.6Hz,1H),7.75(s,1H<sub>NH</sub>) ppm.

[0551] <sup>13</sup>C NMR(100MHz,C<sub>6</sub>D<sub>6</sub>) δ168.2,164.6,162.2,149.7,143.9,141.1,131.1,129.8,123.5,120.4ppm.

[0552] HRMS(ESI):C<sub>14</sub>H<sub>9</sub>N<sub>2</sub>O<sub>5</sub>(M-H)<sup>-</sup>计算值:285.0511,测定值:285.0506.

[0553] (乙基碳酸)4-(4-硝基苯甲酰氨基)苯甲酸酐

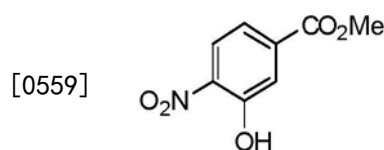


[0555] 搅拌下,在0℃下向4-氨基苯甲酸(1.5g,10.9mmol)和N,N-二甲基苯胺(2.0g,10.9mmol)的丙酮溶液中加入4-硝基苯甲酰氯。然后,将反应混合物温热至室温并再搅拌一小时。将所得的固体过滤并在DMF中通过重结晶纯化,得到4-(4-硝基-苯甲酰基氨基)-苯甲酸(2.75g,88%)。

[0556] 4-(4-硝基-苯甲酰基氨基)-苯甲酸(0.6g,2.1mmol)溶解于14ml CH<sub>3</sub>CN中。然后在0℃下将Et<sub>3</sub>N(0.31ml,2.2mmol)加入。在所得的溶液中加入氯甲酸乙酯。在0℃下搅拌30分钟后,将白色沉淀物过滤并用冷CH<sub>3</sub>CN洗涤,然后在室温下高真空干燥,得到题述酐0.5克,67%。

[0557] <sup>1</sup>H-NMR(400MHz,DMSO,DMSO=2.50ppm):δ=1.33(dd,J=7.2Hz,3H),4.37(q,J=7.2Hz,2H),8.02-8.09(m,4H),8.21(d,J=8.8Hz,2H),8.40(d,J=8.8Hz,2H),11.01(s,1H)。

[0558] 3-羟基-4-硝基苯甲酸甲酯



[0560] 0℃下,TMSCHN<sub>2</sub>(2.0M的Et<sub>2</sub>O溶液,13.20mL,26.48mmol)加入3-羟基-2-硝基苯甲酸(2.50g,13.65mmol)的甲苯/甲醇混合物(81/36mL)溶液中。在0℃下搅拌30分钟后,将溶剂在真空中蒸发,得到油状残余物,将其通过快速色谱纯化(石油醚/乙酸乙酯=9:1),得到标题化合物(2.43g,12.33mmol,90%),为黄色固体。

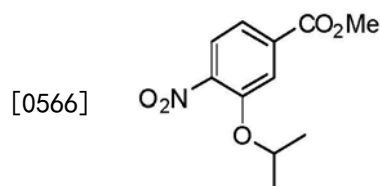
[0561] mp:91-92℃

[0562] <sup>1</sup>H NMR(400MHz,CDCl<sub>3</sub>) δ10.49(s,1H<sub>-OH</sub>),8.17(d,J=8.8Hz,1H),7.83(d,J=1.8Hz,1H),7.61(dd,J=8.8,1.8Hz,1H),3.96(s,3H) ppm.

[0563] <sup>13</sup>C NMR(100MHz,CDCl<sub>3</sub>) δ165.0,154.8,138.1,125.4,121.8,120.74,53.1 ppm.

[0564] HRMS (ESI) :  $C_8H_6NO_5$  (M-H)<sup>-</sup> 计算值:196.0246, 测定值:196.0249.

[0565] 3-异丙氧基-4-硝基苯甲酸甲酯



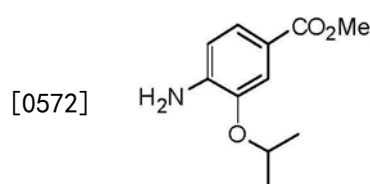
[0567] 3-羟基-4-硝基苯甲酸甲酯 (2.30g, 10.89mmol) 溶解于THF (100mL) 中。加入 <sup>i</sup>PrOH (1.10mL, 14.16mmol) 和PPh<sub>3</sub> (3.90g, 14.70mmol), 混合物搅拌直至所有成份被溶解。然后加入DEAD (2.2M的甲苯溶液, 14.16mmol, 6.50mL), 并将该混合物在室温下搅拌17小时。真空蒸发溶剂, 得到油状残余物, 将其通过快速色谱纯化 (石油醚/乙酸乙酯=95:5), 得到标题化合物 (2.61g, 10.91mmol, 定量), 为黄色油状物。

[0568] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 7.75 (d, J=8.4Hz, 2H), 7.64 (dd, J=8.3, 1.6Hz, 1H), 4.77 (hept, J=6.1Hz, 1H), 3.95 (s, 3H), 1.41 (s, 3H), 1.40 (s, 3H) ppm.

[0569] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 165.5, 150.9, 134.6, 125.2, 121.2, 117.1, 73.2, 52.9, 21.9ppm.

[0570] HRMS (Qtof) :  $C_8H_6NO_5$  (M+Na)<sup>+</sup> 计算值:262.0691, 测定值:262.0700.

[0571] 3-异丙氧基-4-氨基苯甲酸甲酯



[0573] 3-异丙氧基-4-硝基苯甲酸甲酯 (2.60g, 10.87mmol) 溶于MeOH (91.0mL) 中并脱气。加入Pd/C (10%wt., 0.58g, 0.54mmol), 冷却下抽气除去空气。烧瓶用H<sub>2</sub>冲洗, 悬浮液在室温下搅拌17小时。催化剂用硅藻土过滤, 用MeOH洗涤并减压除去溶剂。将粗产物通过快速色谱法 (石油醚/EtOAc=7/3) 纯化。得到3-异丙基-4-氨基苯甲酸甲酯 (2.27g, 10.85mmol, 定量), 为浅橙色固体。

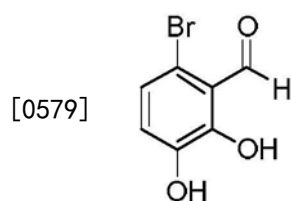
[0574] mp:55-57°C

[0575] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 7.51 (dd, J=8.2, 1.7Hz, 1H), 7.46 (d, J=1.7Hz, 1H), 6.66 (dd, J=8.2, 5.1Hz, 1H), 4.63 (sept, J=5.1Hz, 1H), 3.85 (s, 3H), 1.36 (s, 3H), 1.35 (s, 3H) ppm.

[0576] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 167.5, 144.24, 142.3, 124.0, 119.5, 114.1, 113.5, 70.9, 51.8, 22.3ppm.

[0577] HRMS (ESI) :  $C_{11}H_{16}NO_3$  (M+H)<sup>+</sup> 计算值:210.1130, 测定值:210.1126.

[0578] 6-溴-2,3-二羟基苯甲醛





[0580] 在-30℃下,往6-溴-2-羟基-3-甲氧基苯甲醛(25.0g,108.2mmol)的CH<sub>2</sub>Cl<sub>2</sub>(270mL)溶液中,通过附加的漏斗在45分钟内缓慢加入BBr<sub>3</sub>(1M的CH<sub>2</sub>Cl<sub>2</sub>溶液,200.0mL,200.0mmol)。溶液被升温至室温,搅拌17小时。加入H<sub>2</sub>O,反应混合物继续搅拌30分钟。溶液用EtOAc(3x)萃取,并用H<sub>2</sub>O洗涤。合并的有机层用无水MgSO<sub>4</sub>干燥,过滤,并在真空下浓缩,得到题述化合物(22.16g,102.11 mmol,95%),为黄色固体。

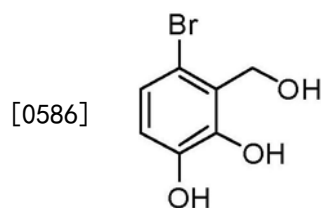
[0581] mp:135-136℃

[0582] <sup>1</sup>H NMR(400MHz,CDCl<sub>3</sub>) δ12.13(d,J=0.5Hz,1H-OH),10.27(s,1H-CHO),7.07(d,J=8.5Hz,1H),7.02(dd,J=8.5,0.5Hz,1H),5.67(s,1H-OH) ppm.

[0583] <sup>13</sup>C NMR(100MHz,CDCl<sub>3</sub>) δ198.4,151.2,145.0,124.4,122.0,117.5,116.1 ppm.

[0584] HRMS(ESI):C<sub>7</sub>H<sub>4</sub>BrO<sub>3</sub>(M-H)<sup>-</sup>计算值:214.3943,测定值:214.9344.

[0585] 4-溴-3-羟基甲基苯-1,2-二醇



[0587] 在-40℃下,6-溴-2,3-二羟基苯甲醛(22.16g,102.10mmol)的THF(650mL)溶液用NaBH<sub>4</sub>(3.86g,102.10mmol)分批(3x)处理。所得的混合物在室温下搅拌30分钟。加入NH<sub>4</sub>Cl的饱和水溶液,混合物再度搅拌10分钟,最后用1M HCl处理。再度搅拌10分钟后,水相用EtOAc(3x)萃取。合并的有机层用无水MgSO<sub>4</sub>干燥并过滤。减压去除溶剂,得到题述化合物(20.27g,92.53mmol,91%),为无色固体。

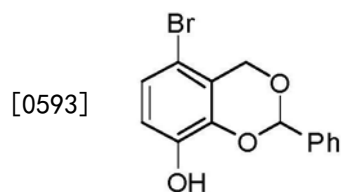
[0588] mp:90-92℃

[0589] <sup>1</sup>H NMR(400MHz,MeOD) δ6.88(d,J=8.5Hz,1H),6.64(d,J=8.5Hz,1H),4.82(s,2H) ppm.

[0590] <sup>13</sup>C NMR(100MHz,MeOD) δ147.1,146.1,126.9,123.9,116.6,114.4,61.1 ppm.

[0591] HRMS(ESI):C<sub>7</sub>H<sub>6</sub>BrO<sub>3</sub>(M-H)<sup>-</sup>计算值:216.9500,测定值:216.9505.

[0592] 5-溴-2-苯基-4H-苯并-[1,3]二氧杂环己烯-8-醇



[0594] 4-溴-3-羟基甲基苯-1,2-二醇(20.27g,92.53mmol)的THF(550mL)溶液用PhCH(OMe)<sub>2</sub>(20.8mL,138.8mmol)和pTSA.H<sub>2</sub>O(0.19g,1.02mmol)处理。混合物在室温下搅拌5天。加入CH<sub>2</sub>Cl<sub>2</sub>,然后相继用5%NaHCO<sub>3</sub>水溶液和盐水洗涤。水相用EtOAc(3x)萃取。合并的有机萃取物用无水MgSO<sub>4</sub>干燥,过滤并减压除去溶剂。通过快速色谱法(石油醚/EtOAc=95/5)纯化,得到5-溴-2-苯基-4H-苯并-[1,3]二氧杂环己烯-8-醇(16.02g,52.16mmol,56%),为无色固体。

[0595] mp:89-91℃

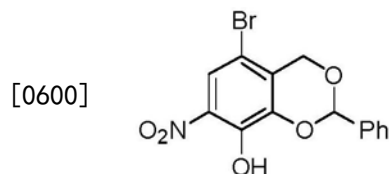
[0596] <sup>1</sup>H NMR(400MHz,CDCl<sub>3</sub>) δ7.62-7.55(m,2H),7.50-7.43(m,3H),7.07(d,J=8.6Hz,

1H), 6.78 (d, J=8.6Hz, 1H), 5.97 (s, 1H), 5.40 (s, 1H-<sub>OH</sub>), 4.99 (s, 2H) ppm.

[0597] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ144.0, 141.8, 136.1, 130.1, 128.8, 126.7, 124.9, 121.0, 115.0, 109.4, 100.0, 67.8 ppm.

[0598] HRMS (ESI): C<sub>14</sub>H<sub>10</sub>BrO<sub>3</sub> (M-H)<sup>-</sup> 计算值: 304.9813, 测定值: 304.9813.

[0599] 5-溴-7-硝基-2-苯基-4H-苯并-[1,3]二氧杂环己烯-8-醇



[0601] 5-溴-2-苯基-4H-苯并[1,3]二氧杂环己烯-8-醇 (6.00g, 19.54mmol; 最大量) 溶解于丙酮 (250mL)。然后加入 Ni(NO<sub>3</sub>)<sub>2</sub>·5H<sub>2</sub>O (5.68g, 19.54mmol) 和 pTSA·H<sub>2</sub>O (3.72g, 19.54mmol)。将混合物在室温下搅拌 2.5 小时。将反应混合物通过硅藻土过滤, 用 CH<sub>2</sub>Cl<sub>2</sub> 洗涤, 真空下浓缩。通过快速色谱法 (干燥负载: SiO<sub>2</sub>+ CH<sub>2</sub>Cl<sub>2</sub>; 石油醚/乙酸乙酯=9:1), 得到题述化合物 (5.08g, 14.43mmol, 74%), 为亮黄色固体。

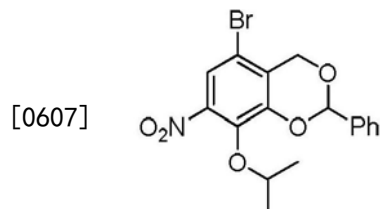
[0602] mp: 154-156°C

[0603] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ10.60 (s, 1H-<sub>OH</sub>), 7.96 (s, 1H), 7.65-7.57 (m, 2H), 7.48-7.42 (m, 3H), 6.02 (s, 1H), 4.99 (s, 2H) ppm.

[0604] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ144.9, 135.5, 133.2, 130.2, 129.0, 128.9, 126.7, 119.2, 109.2, 99.9, 67.4 ppm.

[0605] HRMS (ESI): C<sub>14</sub>H<sub>9</sub>BrNO<sub>5</sub> (M-H)<sup>-</sup> 计算值: 359.9664, 测定值: 349.9660.

[0606] 5-溴-8-异丙氧基-7-硝基-2-苯基-4H-苯并-[1,3]二氧杂环己烯



[0608] 5-溴-7-硝基-2-苯基-4H-苯并-[1,3]-二氧杂环己烯-8-醇 (13.79g, 39.16mmol) 溶解于 THF (429mL) 中。加入 <sup>i</sup>PrOH (4.00mL, 50.91mmol) 和 PPh<sub>3</sub> (13.87g, 52.87 mmol), 搅拌混合物直至所有成份被溶解。缓慢 (通过注射器泵) 加入 DEAD (2.2M 的甲苯溶液, 23.1mL, 50.91mmol), 并将该混合物在室温下搅拌 17 小时。真空蒸发溶剂, 得到油状残余物, 将其通过快速色谱纯化 (石油醚/乙酸乙酯=96:4), 得到标题化合物 (13.08g, 33.18mmol, 85%), 为无色固体。

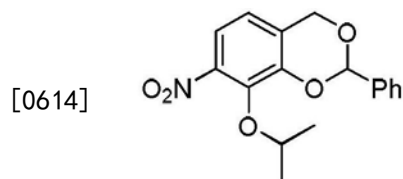
[0609] mp: 87-89°C

[0610] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ7.59 (s, 1H), 7.59-7.54 (m, 2H), 7.50-7.43 (m, 3H), 5.97 (s, 1H), 5.00 (s, 2H), 4.69 (hept, J=6.2Hz, 1H), 1.31 (d, J=6.2Hz, 3H), 1.28 (d, J=6.2Hz, 3H) ppm.

[0611] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ216.8, 149.0, 144.5, 139.9, 135.7, 130.1, 128.8, 126.4, 126.2, 119.8, 112.7, 99.7, 78.1, 67.6, 22.6, 22.4 ppm.

[0612] HRMS (Qtof): C<sub>14</sub>H<sub>9</sub>BrNO<sub>5</sub> (M+Na)<sup>+</sup> 计算值: 416.0110, 测定值: 416.0101.

[0613] 8-异丙氧基-7-硝基-2-苯基-4H-苯并-[1,3]-二氧杂环己烯,73



[0615] 5-溴-8-异丙氧基-7-硝基-2-苯基-4H-苯并-[1,3]二氧杂环己烯72 (4.00g, 10.15 mmol),  $\text{Pd}_2(\text{dba})_3$  (0.93g, 1.01mmol),  $(\text{PhO})_3\text{P}$  (0.53mL, 2.03mmol),  $\text{Cs}_2\text{CO}_3$  (4.30g, 13.19mmol) 和  $i\text{PrOH}$  (4.7mL, 60.88mmol) 溶解于1,4-二氧六环 (28mL) 中。油浴 预热至60℃, 混合物在80℃下搅拌1.5小时。反应混合物通过硅藻土过滤并用  $\text{EtOAc}$  洗涤。合并的有机提取物用无水  $\text{MgSO}_4$  干燥, 并在真空下浓缩。粗产物用 快速色谱纯化 (石油醚/乙酸乙酯=96:4), 得到题述化合物 (2.24g, 7.10mmol, 70%), 为浅黄色固体。

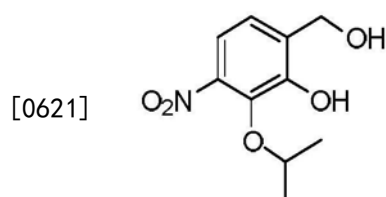
[0616] mp: 80-82℃

[0617]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  7.65-7.55 (m, 2H), 7.51-7.41 (m, 3H), 7.37 (d,  $J=8.5\text{Hz}$ , 1H), 6.81 (d,  $J=8.5\text{Hz}$ , 1H), 6.01 (s, 1H), 5.19 (d,  $J=15.5\text{Hz}$ , 1H), 5.03 (d,  $J=15.5\text{Hz}$ , 1H), 4.71 (hept,  $J=6.2\text{Hz}$ , 1H), 1.32 (d,  $J=6.2\text{Hz}$ , 3H), 1.28 (d,  $J=6.2\text{Hz}$ , 3H) ppm.

[0618]  $^{13}\text{C}$  NMR (100MHz,  $\text{CDCl}_3$ )  $\delta$  147.67, 144.27, 140.55, 136.26, 129.85, 128.72, 126.54, 126.34, 118.82, 116.69, 99.61, 77.71, 66.44, 22.65, 22.41 ppm.

[0619] HRMS (QToF):  $\text{C}_{17}\text{H}_{17}\text{NO}_5\text{Na}$  ( $\text{M}+\text{Na}$ ) $^+$  计算值: 338.1004. 测定值: 338.1003.

[0620] 6-羟甲基-2-异丙氧基-3-硝基苯酚



[0622] 0℃下, 向8-异丙氧基-7-硝基-2-苯基-4H-苯并-[1,3]-二氧杂环己烯 (4.24g, 13.43mmol) 在  $\text{MeOH}$  (102mL) 和  $\text{CH}_2\text{Cl}_2$  (42mL) 的混合物中, 加入樟脑磺酸 (3.12 g, 13.43mmol)。混合物在室温下搅拌17小时。反应混合物用  $\text{Et}_3\text{N}$  淬灭至  $\text{pH} \sim 8$ , 真空下浓缩并用快速色谱纯化 (石油醚/乙酸乙酯=7:3), 得到题述化合物 (2.75 g, 12.09mmol, 90%), 为褐色固体。

[0623] mp: 39-41℃

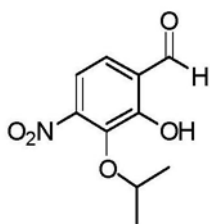
[0624]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  7.46 (d,  $J=7.4\text{Hz}$ , 1H), 7.12 (d,  $J=7.4\text{Hz}$ , 1H), 6.61 (s, 1H-OH), 4.81 (d,  $J=3.5\text{Hz}$ , 2H), 4.39 (hept,  $J=7.4\text{Hz}$ , 1H), 1.36 (s, 3H), 1.35 (s, 3H) ppm.

[0625]  $^{13}\text{C}$  NMR (100MHz,  $\text{CDCl}_3$ )  $\delta$  148.9, 138.5, 132.4, 122.1, 116.5, 79.2, 61.3, 22.5 ppm.

[0626] HRMS (ESI):  $\text{C}_{10}\text{H}_{12}\text{NO}_5$  ( $\text{M}-\text{H}$ ) $^-$  计算值: 226.0715, 测定值: 226.0717.

[0627] 2-羟基-3-异丙氧基-4-硝基苯甲醛

[0628]



[0629] 6-羟甲基-2-异丙氧基-3-硝基苯酚 (2.97g, 13.05mmol) 溶于CH<sub>2</sub>Cl<sub>2</sub> (58mL) 中。然后加入MnO<sub>2</sub> (11.35g, 130.53mmol), 混合物在室温下搅拌17h。混合物用硅藻土过滤, 并用CH<sub>2</sub>Cl<sub>2</sub>洗涤。将溶剂浓缩, 得到标题化合物 (2.38g, 10.57 mmol, 81%), 为棕色油状物。

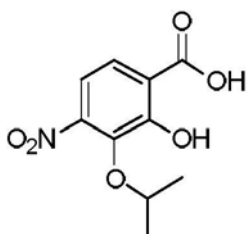
[0630] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 11.44 (s, 1H-CHO), 9.97 (s, 1H-OH), 7.39 (d, J=8.4 Hz, 1H), 7.23 (d, J=8.4Hz, 1H), 4.88 (hept, J=6.2Hz, 1H), 1.33 (s, 3H), 1.32 (s, 3H) ppm.

[0631] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 196.39, 156.53, 149.36, 139.74, 127.28, 122.57, 114.32, 77.42, 77.16, 22.51. ppm.

[0632] HRMS (ESI) : C<sub>10</sub>H<sub>10</sub>NO<sub>5</sub> (M-H)<sup>-</sup> 计算值: 224.0559. 测定值: 224.0535.

[0633] 2-羟基-3-异丙氧基-4-硝基苯甲酸

[0634]



[0635] 2-羟甲基-3-异丙氧基-4-硝基苯甲醛 (2.36g, 10.49mmol) 溶于叔丁醇 (71mL) 中。相继加入2-甲基-2-丁烯 (2M的THF溶液, 36.7mL, 73.45mmol) 以及NaClO<sub>2</sub> (2.85g, 31.48mmol) 和NaH<sub>2</sub>PO<sub>4</sub> (6.32g, 47.22mmol) 的H<sub>2</sub>O溶液 (51mL)。反应混合物在室温下搅拌17小时。加入6M NaOH, 直到pH~10并在真空中浓缩溶剂。加入H<sub>2</sub>O, 有机层用石油醚 (2x) 提取。水层用6M HCl酸化直至pH~1, 并用乙酸乙酯 (3x) 萃取。合并有机萃取物, 用MgSO<sub>4</sub>干燥并过滤。在真空下将溶剂浓缩, 得到标题化合物 (1.90g, 7.87mmol, 75%), 为黑色蜡状。

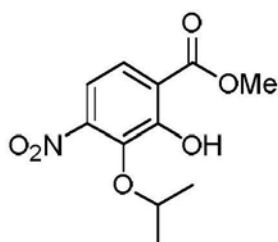
[0636] <sup>1</sup>H NMR (400MHz, MeOD) δ 7.72 (d, J=8.7Hz, 1H), 7.15 (d, J=8.7Hz, 1H), 4.86-4.82 (m, 1H), 1.28 (s, 3H), 1.26 (s, 3H) ppm.

[0637] <sup>13</sup>C NMR (100MHz, MeOD) δ 172.7, 158.0, 140.0, 125.8, 117.4, 113.8, 77.5, 22.6ppm.

[0638] HRMS (ESI) : C<sub>10</sub>H<sub>10</sub>NO<sub>6</sub> (M-H)<sup>-</sup> 计算值: 240.0508, 测定值: 240.0510.

[0639] 2-羟基-3-异丙氧基-4-硝基苯甲酸酯

[0640]



[0641] 0℃下, 将TMSCHN<sub>2</sub> (2.0M的Et<sub>2</sub>O溶液, 0.87mL, 1.75mmol) 加入2-羟基-3-异丙氧基-4-硝基苯甲酸 (0.32g, 1.35mmol) 的甲苯/甲醇混合物 (10.4/2mL) 溶液中。在0℃下搅拌30

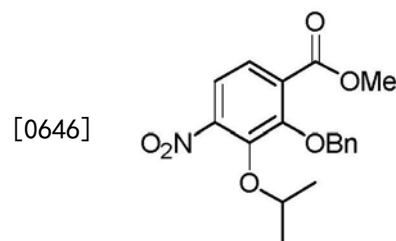
分钟后,在真空中蒸发溶剂,得到油状残余物,将其通过快速色谱纯化( $\text{SiO}_2 \cdot \text{Et}_3\text{N}$ ;石油醚/乙酸乙酯=95:5),得到标题化合物(0.24g,0.94 mmol,57%),为黄色油状物。

[0642]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  11.29 (s, 1H- $\text{OH}$ ), 7.63 (d,  $J=8.8\text{Hz}$ , 1H), 7.12 (d,  $J=8.8\text{Hz}$ , 1H), 4.84 (hept,  $J=6.2\text{Hz}$ , 1H), 4.00 (s, 3H), 1.32 (s, 3H), 1.31 (s, 3H) ppm.

[0643]  $^{13}\text{C}$  NMR (100MHz,  $\text{CDCl}_3$ )  $\delta$  198.2, 188.9, 176.1, 170.0, 157.0, 149.2, 139.8, 123.9, 115.7, 113.4, 77.4, 53.2, 22.5 ppm.

[0644] HRMS (ESI):  $\text{C}_{11}\text{H}_{12}\text{NO}_6$  ( $\text{M}-\text{H}$ ) $^-$  计算值:254.0665, 测定值:254.0666.

[0645] 2-苄氧基-3-异丙氧基-4-硝基苯甲酸酯



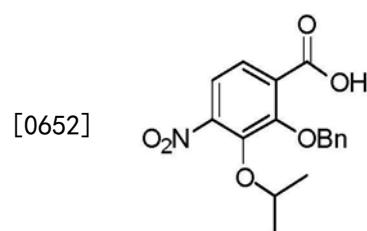
[0647] 2-羟基-3-异丙氧基-4-硝基苯甲酸酯(0.17g,0.69mmol)溶于THF(7.5mL)中。加入BnOH(92.6 $\mu\text{L}$ ,0.89mmol)和 $\text{PPh}_3$ (0.24g,0.93mmol),搅拌混合物直至所有的成份溶解。缓慢(通过注射器泵)加入DEAD(2.2M的甲苯溶液,0.41mL,0.89 mmol),并将该混合物在室温下搅拌17小时。在真空中蒸发溶剂,得到油状残余物,将其通过快速色谱纯化(石油醚/乙酸乙酯=95:5),得到标题化合物(0.20g, 0.58mmol,85%),为无色油状物。

[0648]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  7.53 (d,  $J=8.6\text{Hz}$ , 1H), 7.50 (d,  $J=8.6\text{Hz}$ , 1H), 7.48-7.44 (m, 2H), 7.42-7.35 (m, 3H), 5.14 (s, 2H), 4.74 (hept,  $J=6.2\text{Hz}$ , 1H), 3.86 (s, 3H), 1.28 (s, 3H), 1.26 (s, 3H) ppm.

[0649]  $^{13}\text{C}$  NMR (100MHz,  $\text{CDCl}_3$ )  $\delta$  165.3, 153.4, 148.4, 145.7, 136.4, 130.9, 128.7, 128.7, 128.7, 125.1, 119.3, 78.2, 76.4, 52.8, 22.5 ppm.

[0650] HRMS (QTof):  $\text{C}_{18}\text{H}_{19}\text{NO}_6\text{Na}$  ( $\text{M}+\text{Na}$ ) $^+$  计算值:368.1110, 测定值:368.1112.

[0651] 2-苄氧基-3-异丙氧基-4-硝基苯甲酸



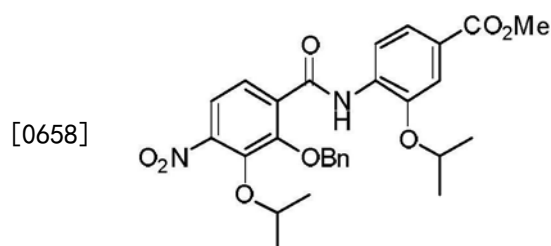
[0653] 2-苄氧基-3-异丙氧基-4-硝基苯甲酸酯(0.23g,0.67mmol)溶于THF/ $\text{H}_2\text{O}$  (3.5/3.5mL)的1/1混合物中。然后,加入固体LiOH(0.16g,6.67mmol),并将该体系在室温下搅拌17小时。水层用1M HCl酸化直至pH~1,并用EtOAc(3x)萃取。合并有机萃取物,用 $\text{MgSO}_4$ 干燥并过滤。在真空下浓缩溶剂,得到标题化合物(0.21g,0.63mmol,95%),为黄色蜡状。

[0654]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  7.91 (d,  $J=8.7\text{Hz}$ , 1H), 7.58 (d,  $J=8.7\text{Hz}$ , 1H), 7.41 (s, 5H), 5.35 (s, 2H), 4.71-4.62 (m, 1H), 1.36 (s, 3H), 1.35 (s, 3H) ppm.

[0655]  $^{13}\text{C}$  NMR (100MHz,  $\text{CDCl}_3$ )  $\delta$  164.3, 152.8, 149.7, 144.7, 134.1, 129.8, 129.4, 129.2, 126.98, 120.0, 79.1, 77.7, 22.5 ppm.

[0656] HRMS (ESI) :  $C_{17}H_{16}NO_6$  (M-H)<sup>-</sup> 计算值: 330.0978, 测定值: 330.0976.

[0657] 4-(2-(苄氧基)-3-异丙氧基-4-硝基苯甲酰胺)-3-异丙基苯甲酸酯



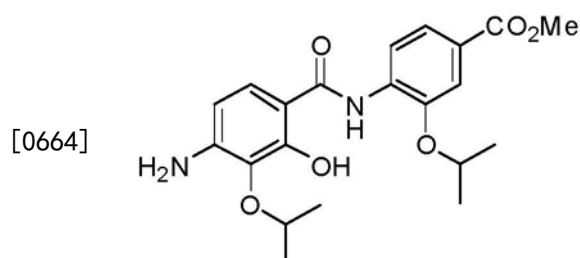
[0659] 40℃下, 2-苄氧基-3-异丙氧基-4-硝基苯甲酸 (51.5mg, 0.16mmol) 溶解于  $CH_2Cl_2$  (8mL) 并与 Ghosez's 试剂 (66.0μL, 0.50mmol) 预活化3小时。3-异丙氧基-4-氨基苯甲酸甲酯 (0.12g, 0.55mmol) 溶解于  $CH_2Cl_2$  (8mL) 中, 并加入 N,N-二异丙基乙胺 (DIPEA) (0.20mL, 1.12mmol)。然后加入含酰氯的溶液, 并在40℃下将反应混合物搅拌2天。然后将溶剂除去并且将粗产物通过制备型HPLC纯化 (RP-18; 运行时间100分钟;  $H_2O/MeCN=100:0 \rightarrow 0:100$ ; tr=80分钟), 得到标题化合物 (56.9mg, 0.11mmol, 68%), 为浅黄色油状物。

[0660]  $^1H$  NMR (400MHz,  $CDCl_3$ )  $\delta$  10.33 (s, 1H-NH), 8.55 (d, J=8.5Hz, 1H), 7.85 (d, J=8.7Hz, 1H), 7.70 (dd, J=8.5, 1.7Hz, 1H), 7.59 (d, J=8.7Hz, 1H), 7.57 (d, J=1.7Hz, 1H), 7.25-7.12 (m, 5H), 5.25 (s, 2H), 4.75-4.67 (m, 1H), 4.67-4.59 (m, 1H), 3.93 (s, 3H), 1.40 (d, J=6.2Hz, 6H), 1.28 (d, J=6.0Hz, 6H) ppm.

[0661]  $^{13}C$  NMR (100MHz,  $CDCl_3$ )  $\delta$  167.0, 161.4, 151.1, 147.9, 146.1, 145.2, 134.1, 132.9, 132.9, 130.0, 129.4, 128.7, 125.79, 125.6, 123.3, 120.1, 119.5, 113.3, 78.9, 77.4, 71.7, 52.3, 22.6, 22.1 ppm.

[0662] HRMS (ESI) :  $C_{28}H_{31}N_2O_8$  (M+H)<sup>+</sup> 计算值: 523.2080, 测定值: 523.2075.

[0663] 4-(4-氨基-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸酯



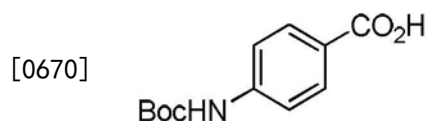
[0665] 4-[2-(苄氧基)-3-异丙氧基-4-硝基苯甲酰基]-3-异丙氧基苯甲酸酯 (7.9mg, 0.015mmol) 溶于 MeOH (0.5mL) 中并脱气。加入 Pd/C (10% wt., 2mg, 0.0014mmol), 冷却下抽气除去空气。烧瓶用  $H_2$  冲洗, 悬浮液在室温下搅拌3小时。催化剂用硅藻土过滤, 用 MeOH 洗涤并在减压下除去溶剂。粗产物用快速色谱纯化 (石油醚/乙酸乙酯=7:3), 得到题述化合物 (5.8g, 0.014mmol, 96%), 为黄色固体。

[0666]  $^1H$  NMR (400MHz,  $CDCl_3$ )  $\delta$  12.21 (s, 1H-OH), 8.81 (s, 1H-NH), 8.49 (d, J=8.5Hz, 1H), 7.69 (dd, J=8.5, 1.8Hz, 1H), 7.58 (d, J=1.7Hz, 1H), 7.07 (d, J=8.8Hz, 1H), 6.28 (d, J=8.7Hz, 1H), 4.80-4.72 (m, 1H), 4.72-4.63 (m, 1H), 4.28 (s, 2H-NH<sub>2</sub>), 3.91 (s, 3H), 1.44 (d, J=6.1Hz, 6H), 1.34 (d, J=6.2Hz, 7H) ppm.

[0667]  $^{13}C$  NMR (100MHz,  $CDCl_3$ )  $\delta$  168.5, 166.9, 156.4, 146.5, 146.0, 132.7, 132.0, 125.1, 123.40, 121.5, 119.1, 113.4, 106.5, 106.3, 77.4, 74.4, 72.0, 52.3, 22.9, 22.4 ppm.

[0668] HRMS (ESI) :  $C_{21}H_{25}N_2O_6$  (M-H)<sup>-</sup> 计算值: 401.1713, 测定值: 401.1716.

[0669] 4-(叔丁氧基羰基氨基) 苯甲酸



[0671] 4-氨基苯甲酸 (1.00g, 7.29mmol) 溶解于1,4-二氧六环 (15mL) 和H<sub>2</sub>O (7mL) 中。Et<sub>3</sub>N (2.0mL, 14.58mmol) 加入到该溶液中, 在室温下将反应混合物搅拌5分钟。然后一次性将二叔丁基二碳酸酯 (3.18g, 14.58mmol) 加入到该溶液中并将反应混合物搅拌24小时。在真空中除去溶剂, 然后将3M盐酸加入到残余物中, 得到白色沉淀。然后过滤该浆体, 并用H<sub>2</sub>O洗涤, 然后在高真空下干燥。用热甲醇重结晶, 得到标题化合物, 为无色固体 (1.63g, 6.85mmol, 94% 产率)。

[0672] mp: 192-194°C.

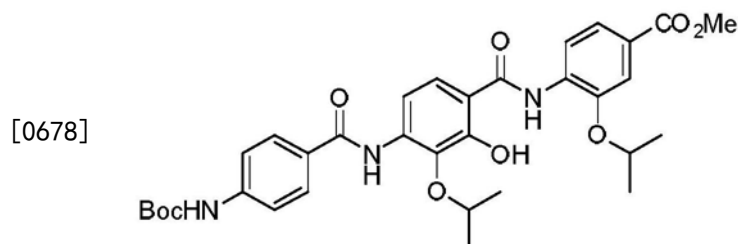
[0673] <sup>1</sup>H NMR (400MHz, DMSO) δ 9.73 (s, 1H-<sub>CO2H</sub>), 7.83 (d, 2H, J=8.9Hz), 7.55 (d, 2H, J=8.9Hz), 1.47 (s, 9H) ppm.

[0674] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 167.1, 152.6, 143.8, 130.4, 124.0, 117.2, 79.7, 28.1 ppm.

[0675] HRMS (ESI) :  $C_{12}H_{15}NNaO_4$  (M+Na)<sup>+</sup> 计算值: 260.0893, 测定值: 260.0897.

[0676] 光谱数据与文献报道一致 (J. Am. Chem. Soc. 2012, 134, 7406-7413)。

[0677] 甲基-4-(4-(4-(叔丁氧羰基)氨基)苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸酯



[0679] 4-(叔丁氧羰基氨基) 苯甲酸 (40.0mg, 0.17mmol) 溶解于CH<sub>2</sub>Cl<sub>2</sub> (8.4mL) 中, 并在室温下与Ghosez's试剂 (22.5μL, 0.17mmol) 预活化2个小时。4-(4-氨基-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸酯 (68.4mg, 0.17mmol) 溶解在 CH<sub>2</sub>Cl<sub>2</sub> (8.4mL) 溶液中, 并加入N,N-二异丙基乙胺 (DIPEA) (59.2μL, 0.34 mmol)。然后加入含酰氯的溶液, 在室温下将反应混合物搅拌1天。然后将溶剂除去, 粗产物通过制备型HPLC纯化 (RP-18; 运行时间100分钟; H<sub>2</sub>O/MeCN = 100:0 → 0:100; tr = 70分钟) 得到浅黄色油状标题化合物 (47.3mg, 0.076mmol, 72%)。

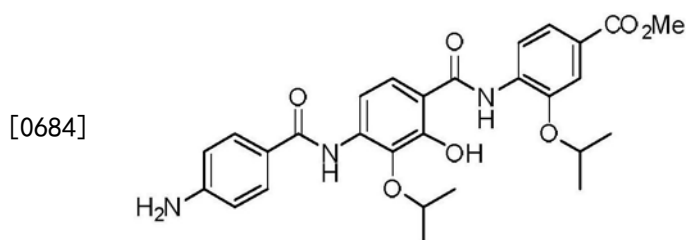
[0680] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 7.98 (d, J=7.5Hz, 2H), 7.78 (d, J=1.4Hz, 1H), 7.72 (dd, J=7.5, 1.4Hz, 1H), 7.69 (s, 1H-NH), 7.68 (d, J=7.3Hz, 3H), 7.56 (d, J=7.5 Hz, 1H), 7.17 (d, J=7.5Hz, 1H), 5.72 (s, 1H-NH), 5.49 (s, 1H-NH), 4.02-3.96 (m, 2H), 3.95 (d, J=3.7Hz, 3H), 1.49 (s, 9H), 1.46 (d, J=5.6Hz, 6H), 1.41 (d, J=5.5Hz, 6H) ppm.

[0681] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 166.89, 166.67, 166.61, 158.88, 154.93, 146.90, 141.47, 135.07, 134.68, 131.70, 130.38, 130.38, 127.26, 127.17, 123.25, 121.40, 120.63, 120.63, 115.87, 114.85, 113.39, 106.06, 80.65, 75.89, 74.13, 52.08, 28.41, 28.41, 28.41,

21.80, 21.80, 21.80, 21.80ppm.

[0682] HRMS (ESI) :  $C_{33}H_{38}N_3O_9$  (M-H)<sup>-</sup> 计算值: 620.2687, 测定值: 620.2689.

[0683] 甲基-4-(4-(4-氨基苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸酯



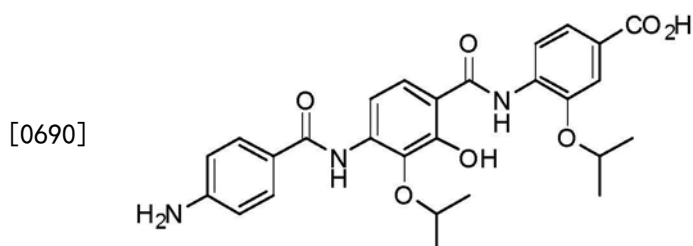
[0685] 甲基-4-(4-(4-(叔丁氧基羰基)氨基)苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸酯 (40.0mg, 0.064mmol) 溶解于二氯甲烷/三氟乙酸的10/1混合物 (1mL) 中, 并在室温下搅拌17小时。减压下除去溶剂, 并在高真空下除去残留的酸, 得到标题化合物 (33.4mg, 0.064mmol, 定量), 为黄色油状物。

[0686] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 7.86 (d, J=1.4Hz, 1H), 7.83 (s, 1H<sub>-NH</sub>), 7.79 (dd, J=7.5, 1.4Hz, 1H), 7.75 (d, J=7.5Hz, 1H), 7.70 (d, J=7.5Hz, 2H), 7.65 (d, J=7.5Hz, 1H), 7.05 (d, J=7.5Hz, 1H), 6.94 (s, 1H<sub>-NH</sub>), 6.75 (d, J=7.5Hz, 2H), 6.09 (s, 1H<sub>-OH</sub>), 4.02-3.97 (m, 1H), 3.95-3.89 (s, 3H), 3.92 (m, 1H), 3.85 (s, 2H<sub>-NH</sub>), 1.47 (d, J=5.7 Hz, 6H), 1.40 (d, J=5.5Hz, 6H) ppm.

[0687] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 166.89, 166.67, 166.61, 158.88, 152.59, 146.90, 135.07, 134.68, 131.70, 130.93, 130.93, 127.17, 123.25, 122.42, 121.40, 115.87, 114.85, 114.35, 114.35, 113.39, 106.06, 75.89, 74.13, 52.08, 21.80, 21.80, 21.80, 21.80ppm.

[0688] HRMS (ESI) :  $C_{28}H_{32}N_3O_7$  (M+H)<sup>+</sup> 计算值: 522.2162, 测定值: 522.2160.

[0689] 孢囊菌酰胺C



[0691] 甲基-4-[4-(4-氨基苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基]-3-异丙氧基苯甲酸酯 (30.0mg, 0.058mmol) 溶于THF/H<sub>2</sub>O (0.3/0.3mL) 的1/1混合物中。然后, 加入固体 LiOH (13.9mg, 0.58mmol), 并将该体系在室温下搅拌17小时。水层用 1M HCl 酸化直至 pH~1, 并用乙酸乙酯 (3x) 萃取。合并有机萃取物, 用 MgSO<sub>4</sub> 干燥并过滤。在真空下将溶剂浓缩, 得到标题化合物 (27.4mg, 0.054mmol, 93%), 为黄色油状。

[0692] <sup>1</sup>H NMR (400MHz, CDCl<sub>3</sub>) δ 7.91 (d, J=1.4Hz, 1H), 7.87 (dd, J=7.5, 1.4Hz, 1H), 7.70 (d, J=7.5Hz, 2H), 7.65 (d, J=7.5Hz, 1H), 7.53 (d, J=7.5Hz, 1H), 7.05 (d, J=7.5Hz, 1H), 6.95 (s, 1H<sub>-NH</sub>), 6.77 (s, 1H<sub>-NH</sub>), 6.75 (d, J=7.5Hz, 2H), 6.12 (s, 1H<sub>-OH</sub>), 3.97-3.89 (m, 2H), 3.85 (s, 2H<sub>-NH</sub>), 1.40 (d, J=5.5Hz, 6H), 1.39 (d, J=5.5Hz, 6H) ppm.

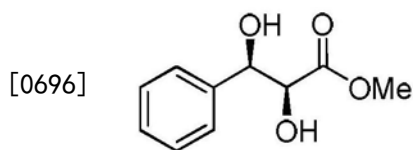
[0693] <sup>13</sup>C NMR (100MHz, CDCl<sub>3</sub>) δ 167.79, 166.67, 166.61, 158.88, 152.59, 149.81,



136.38, 135.07, 134.68, 130.93, 130.93, 125.08, 123.25, 122.80, 122.42, 120.37, 114.35, 114.35, 113.76, 113.39, 106.06, 75.89, 74.13, 21.80, 21.80, 21.80, 21.80 ppm.

[0694] HRMS (ESI) :  $C_{28}H_{32}N_3O_7$  (M+H)<sup>+</sup> 计算值: 508.2006, 测定值: 508.2008.

[0695] (2S, 3R)-2, 3-二羟基-3-苯基丙酸甲酯



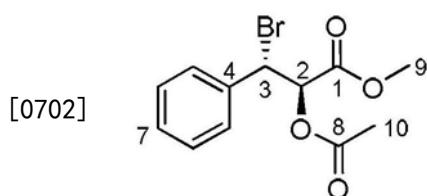
[0697] 25℃下, AD混合物β (20.0g) 溶解于 tBuOH/H<sub>2</sub>O (1:1; 142mL) 混合物中。然后将 CH<sub>3</sub>SO<sub>2</sub>NH<sub>2</sub> (1.36g, 14.3mmol, 1.0eq.) 加入, 反应混合物冷却至 0℃。然后加入肉桂酸甲酯 (2.31g, 14.3mmol, 1.0eq.), 所得的混合物在 0℃ 下猛烈搅拌 16h。在 25℃ 下继续进行搅拌 6h。反应混合物通过加入 Na<sub>2</sub>SO<sub>3</sub> 水溶液 (21.4g, 170mmol, 12.0eq.) 水解, 并继续搅拌 2.5h。反应混合物用乙酸乙酯稀释并分层。水相用 EtOAc (3x) 萃取。合并的有机相用 H<sub>2</sub>O (1x) 洗涤, 并用 Na<sub>2</sub>SO<sub>4</sub> 干燥, 过滤并减压浓缩。通过快速色谱法 (石油醚/乙酸乙酯 = 1:1) 纯化, 得到所需的二醇 (2.21g, 11.3mmol, 79%), 为无色固体。光谱数据与文献报道一致。

[0698]  $R_f = 0.38$  (PE/EtOAc 1:1); m.p. = 84-85℃ (lit: m.p. = 80-81℃);  $[\alpha]_D^{20} = -9.8^\circ$  (c 1.28, CHCl<sub>3</sub>) {文献:  $[\alpha]_D^{26} = -9.8^\circ$  (c 1.07, CHCl<sub>3</sub>)};

[0699] <sup>1</sup>H-NMR (400MHz, CDCl<sub>3</sub>, CHCl<sub>3</sub> = 7.26ppm):  $\delta = 7.42-7.29$  (5H, m, ArH), 5.03 (1H, dd, J = 2.7, 7.2Hz, H-3), 4.38 (1H, dd, J = 2.7, 6.0Hz, H-2), 3.82 (3H, s, H-8), 3.12 (1H, d, J = 6.0Hz, OH-α), 2.76 (1H, d, J = 7.2Hz, OH-β) ppm;

[0700] <sup>13</sup>C-NMR (100MHz, CDCl<sub>3</sub>, CHCl<sub>3</sub> = 77.16ppm):  $\delta = 173.3$  (q, C-1), 140.1 (q, C-4), 128.6 (2C, t, C-6), 128.3 (t, C-7), 126.3 (2C, t, C-5), 74.8 (t, C-2), 74.6 (t, C-3), 53.1 (p, C-8) ppm; HRMS (ESI) : m/z  $C_{10}H_{12}O_4Na$  [M+Na]<sup>+</sup> 计算值: 219.0633; 测量值 219.0633.

[0701] (2R, 3S)-2-乙酰氧基-3-溴-3-苯基丙酸甲酯 (3)



[0703] 在 25℃ 下, 向 (2S, 3R)-2, 3-二羟基-3-苯基丙酸甲酯 (2.15g, 10.9mmol, 1.0eq.) 中逐滴加入 HBr/HOAc (33%; 16.9mL)。所得的混合物加热至 45℃ 并搅拌 30 分钟。然后反应混合物冷却至 25℃, 倒入冰冷的 NaHCO<sub>3</sub> 溶液 (40mL) 中。水相用 Et<sub>2</sub>O (3x) 萃取。合并的有机相用 H<sub>2</sub>O (1x) 和盐水洗涤。然后, 合并的有机相用 Na<sub>2</sub>SO<sub>4</sub> 干燥, 过滤并减压浓缩。通过快速色谱法 (石油醚/乙酸乙酯 = 12.5:1) 纯化, 得到题述化合物 (2.32g, 7.71mmol, 71%), 为无色固体。光谱数据与文献报道一致。

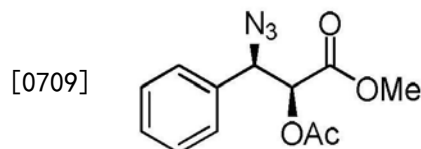
[0704]  $R_f = 0.79$  (PE/EtOAc 1:1); m.p. = 78-82℃ (lit: m.p. = 78-79℃);  $[\alpha]_D^{20} = +89.9^\circ$  (c 1.74, CHCl<sub>3</sub>) {文献:  $[\alpha]_D^{26} = +100.3^\circ$  (c 1.36, CHCl<sub>3</sub>)};

[0705] <sup>1</sup>H-NMR (400MHz, CDCl<sub>3</sub>, CHCl<sub>3</sub> = 7.26ppm):  $\delta = 7.46-7.44$  (2H, m, H-6), 7.36-7.30 (3H, m, H-5, H-7), 5.65 (1H, d, J = 6.3Hz, H-3), 5.35 (1H, d, J = 6.3Hz, H-2), 3.71 (3H, s, H-9), 2.11 (3H, s, H-10) ppm;

[0706]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=77.16\text{ppm}$ ):  $\delta=169.7$  (q, C-1),  $167.5$  (q, C-8),  $136.8$  (q, C-4),  $129.3$  (t, C-7),  $128.7$  (4C, t, C-5, C-6),  $75.4$  (t, C-3),  $52.9$  (p, C-9),  $49.3$  (t, C-2),  $20.6$  (p, C-10) ppm;

[0707] HRMS (ESI) :m/z  $\text{C}_{12}\text{H}_{13}\text{O}_4\text{BrNa}[\text{M}+\text{Na}]^+$  计算值:322.9895;测量值 322.9891.

[0708] (2S,3R)-2-乙酰氧基-3-叠氮基-3-苯基丙酸甲酯



[0710] 在25℃下, (2S,3R)-2-乙酰氧基-3-叠氮基-3-苯基丙酸甲酯 (2.27g, 7.55 mmol, 1.0eq.) 溶解于DMF (27.0mL)。然后加入NaN<sub>3</sub> (1.96g, 30.2mmol, 4.0 eq.), 所得的混合物40℃加热3h。冷却后, 反应混合物冷却至25℃, 并加入 EtOAc。有机层用H<sub>2</sub>O (2x) 洗涤, 然后用盐水 (1x) 洗涤。合并的有机相用Na<sub>2</sub>SO<sub>4</sub>干燥, 过滤并减压浓缩。通过快速色谱法 (石油醚/乙酸乙酯=10:1) 纯化, 得到题述化合物 (1.77g, 6.71mmol, 89%), 为黄色油状物。光谱数据与文献报道一致。

[0711]  $R_f=0.24$  (PE/EtOAc=10:1);  $[\alpha]_D^{20}=-97.8^\circ$  (c 2.3,  $\text{CHCl}_3$ ); {文献:  $[\alpha]_D^{26}=-104.2^\circ$  (c 2.33,  $\text{CHCl}_3$ )};

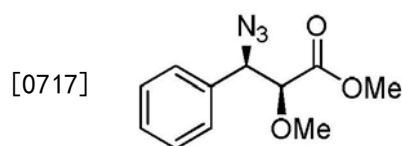
[0712] IR:  $\tilde{\nu}=2955$  (w), 2103 (s, azide), 1747 (s, C=O), 1495 (w), 1454 (m), 1437 (m), 1373 (m), 1210 (s), 1099 (m), 1030 (m), 910 (m), 751 (m), 701 (s)  $\text{cm}^{-1}$ ;

[0713]  $^1\text{H}$ -NMR (400MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=7.26\text{ppm}$ ):  $\delta=7.42-7.33$  (5H, m, ArH),  $5.24$  (1H, d,  $J=4.8\text{Hz}$ , H-2),  $5.07$  (1H, d,  $J=4.8\text{Hz}$ , H-3),  $3.69$  (3H, s, H-9),  $2.14$  (3H, s, H-10) ppm;

[0714]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=77.16\text{ppm}$ ):  $\delta=169.9$  (q, C-1),  $168.0$  (q, C-8),  $134.6$  (q, C-4),  $129.3$  (t, C-7),  $129.0$  (2C, t, C-6),  $127.6$  (2C, t, C-5),  $74.9$  (t, C-2),  $65.4$  (t, C-3),  $52.8$  (p, C-9),  $20.5$  (p, C-10) ppm;

[0715] HRMS (ESI) :m/z  $\text{C}_{12}\text{H}_{13}\text{N}_3\text{O}_4\text{Na}[\text{M}+\text{Na}]^+$  计算值:286.0804;测量值 286.0805.

[0716] (2S,3R)-3-叠氮基-2-甲氧基-3-苯基丙酸甲酯



[0718] 0℃下, (2S,3R)-2-乙酰氧基-3-叠氮基-3-苯基丙酸甲酯 (2.5g, 1.0eq) 溶解于190ml THF中。然后滴加加入KOH (0.5M, 10.0eq) 溶液, 反应混合物在0℃下 搅拌5h。然后将2N HCl水溶液加入到反应混合物中, 水相用乙酸乙酯萃取。将有机相合并, 用硫酸钠干燥, 过滤, 减压浓缩得到粗酸, 将其直接用于下一步骤, 无需进一步纯化。粗产物 (0.5g, 1.0eq) 溶解于17ml碘甲烷中。然后加入  $\text{CaSO}_4$  (2.6g, 8.0eq) 和 $\text{Ag}_2\text{O}$  (1.7g, 3.0eq), 室温下在暗处搅拌所述悬浮液22h。然后过滤该粗混合物并在真空中浓缩得到题述化合物 (70%产率), 可以直接用于下一步骤而无需纯化。

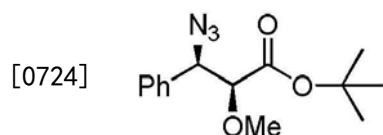
[0719]  $[\alpha]_D^{20}=-143.7^\circ$  (c 1.1,  $\text{CHCl}_3$ );

[0720]  $^1\text{H}$ -NMR (400MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=7.26\text{ppm}$ ):  $\delta=3.44$  (s, 3H),  $3.61$  (s, 3H),  $3.94$  (d,  $J=6.4\text{Hz}$ , 1H),  $4.79$  (d,  $J=6.4\text{Hz}$ , 1H),  $7.35-7.36$  (m, 5H);

[0721]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=77.0\text{ppm}$ ):  $\delta=52.2, 59.1, 66.9, 84.7, 127.7, 128.7, 128.9, 135.1, 170.0$ ;

[0722] HRMS (ESI):  $m/z$   $\text{C}_{11}\text{H}_{13}\text{N}_3\text{O}_3\text{Na}[\text{M}+\text{Na}]^+$  计算值: 258.0855; 测量值 258.0852.

[0723] (2S,3S)-3-叠氮基-2-甲氧基-3-苯基丙酸叔丁酯



[0725] 边搅拌,边往 (2S,3R)-3-叠氮基-2-甲氧基-3-苯基丙酸甲酯 (1.2g, 1.0eq) 的 100 ml THF 溶液中逐滴加入 KOH (0.5M, 10.0eq) 水溶液。反应混合物在室温下搅拌 5h 并通过加入 2N HCl 水解。水相用乙酸乙酯萃取, 合并的有机相经硫酸钠干燥并减压浓缩, 得到羧酸 (1.2g, 产率 98%), 其不经过进一步纯化而用于下一反应。在室温下, 粗酸 (0.3g, 1.0eq) 和 3.9ml 二甲基甲酰胺二叔丁基缩醛 (3.9ml, 12eq) 溶解于 8ml 甲苯中。所得的反应混合物加热至 80℃ 并搅拌 7h。减压下除去溶剂, 粗产物通过快速柱色谱纯化 (石油醚/乙酸乙酯 = 30:1), 得到题述化合物 (0.34g, 89% 收率)。

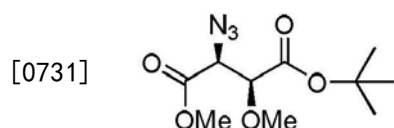
[0726]  $[\alpha]_D^{20} = -113.3^\circ$  (c 1.0,  $\text{CHCl}_3$ );

[0727]  $^1\text{H}$ -NMR (400MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=7.26\text{ppm}$ ):  $\delta=1.26$  (s, 9H), 3.45 (s, 3H), 3.85 (d, J = 7.2Hz, 1H), 4.70 (d, J = 7.2Hz, 1H), 7.34-7.35 (m, 5H);

[0728]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=77.0\text{ppm}$ ):  $\delta=27.7, 58.6, 67.2, 82.3, 85.1, 128.2, 128.6, 128.9, 135.2, 168.5$ ;

[0729] HRMS (ESI):  $m/z$   $\text{C}_{14}\text{H}_{19}\text{O}_3\text{N}_3\text{Na}[\text{M}+\text{Na}]^+$  计算值: 300.1324; 测量值 300.1332.

[0730] (2S,3S)-4-叔丁基-1-甲基-2-叠氮基-3-甲氧基琥珀酸酯

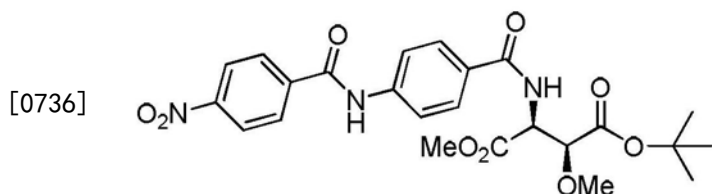


[0732] 在室温下, 边搅拌边向 (2S,3S)-3-叠氮基-2-甲氧基-3-苯基丙酸叔丁酯 (310 mg, 1.0eq) 的 3ml  $\text{CHCl}_3$ , 13ml  $\text{CH}_3\text{CN}$  和 26ml  $\text{H}_2\text{O}$  混合溶液中逐滴加入  $\text{NaIO}_4$  (7.2g, 30eq) 和  $\text{RuCl}_3$  (0.3eq, 69mg)。在回流条件下, 将反应混合物加热 3 小时。冷却至室温后形成白色沉淀。过滤得到固体, 滤液用二乙醚萃取。合并的有机相在减压下浓缩, 得到粗产物。所述的物料溶解于 9ml 碘甲烷中。然后加入  $\text{CaSO}_4$  (1.2g, 8.0eq) 和  $\text{Ag}_2\text{O}$  (778mg, 3.0eq), 室温下在暗处搅拌所述反应混合物 22h。过滤, 滤液在减压下浓缩得到纯品形式的题述化合物, 其可以直接用于下一步骤而无需进一步纯化。

[0733]  $^1\text{H}$ -NMR (400MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=7.26\text{ppm}$ ):  $\delta=1.51$  (s, 3H), 3.48 (s, 3H), 4.15 (d, J = 3.6Hz, 1H), 4.21 (d, J = 4.0Hz, 1H);

[0734]  $^{13}\text{C}$ -NMR (100MHz,  $\text{CDCl}_3$ ,  $\text{CHCl}_3=77.0\text{ppm}$ ):  $\delta=28.1, 53.0, 59.5, 63.4, 81.2, 83.0, 167.7, 168.3$ .

[0735] (2S,3R)-1-叔丁基-4-甲基-2-甲氧基-3-[4-(4-硝基苯甲酰氨基)苯甲酰氨基]琥珀酸酯



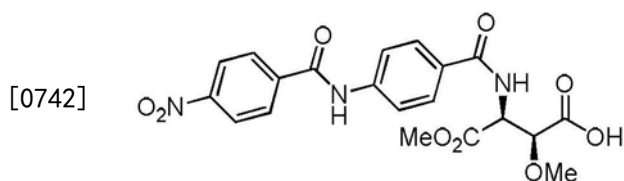
[0737] 粗品混合物 (2S,3S)-4-叔丁基-1-甲基-2-叠氮基-3-甲氧基琥珀酯溶解于12ml THF,然后加入0.5ml水和PPh<sub>3</sub> (881mg,3.0eq)。所得的反应混合物加热至50℃ 并继续进行搅拌12h。然后在减压下除去溶剂,得到粗产物,其足够纯可直接用于下一步骤。粗产物溶解于5ml DMF,在室温下加入(乙基碳酸)4-(4-硝基苯甲酰氨基)苯甲酸酐(481mg,1.2eq)。搅拌20小时后,加入水,水溶液用乙酸乙酯萃取。减压浓缩合并的有机相。通过快速色谱法(石油醚/乙酸乙酯=2:1)纯化,得到题述化合物(81mg,四步收率16%)。

[0738]  $[\alpha]_D^{20} = -11.8^\circ$  (c 1.1, CHCl<sub>3</sub>);

[0739] <sup>1</sup>H-NMR (400MHz, CDCl<sub>3</sub>, CHCl<sub>3</sub>=7.26ppm):  $\delta$ =1.41 (s, 9H), 3.45 (s, 3H), 3.78 (s, 3H), 4.34 (d, J=2.4Hz, 1H), 5.29 (dd, J=2.4, 9.6Hz, 1H), 6.76 (d, J=9.6Hz, 1H), 7.27-7.35 (m, 4H), 8.07 (d, J=8.8Hz, 2H), 8.26 (d, J=8.8Hz, 2H), 8.83 (s, 1H);

[0740] <sup>13</sup>C-NMR (100MHz, CDCl<sub>3</sub>, CHCl<sub>3</sub>=77.0ppm):  $\delta$ =27.9, 52.9, 54.8, 59.1, 79.8, 83.2, 120.1, 123.8, 128.3, 128.7, 129.6, 140.3, 141.1, 149.7, 164.1, 166.9, 168.0, 169.7.

[0741] HRMS (ESI): m/z C<sub>24</sub>H<sub>27</sub>O<sub>9</sub>N<sub>3</sub>Na [M+Na]<sup>+</sup>计算值:524.1645;测量值 524.1647.



[0743] 在室温下,边搅拌边向(2S,3R)-1-叔丁基-4-甲基-2-甲氧基-3-[4-(4-硝基苯甲酰氨基)苯甲酰氨基]琥珀酸酯(74.3mg,0.15mmol)的2.5ml二氯甲烷溶液中加入1.5mL的TFA。搅拌5小时后,向反应混合物中加入水,并用乙酸乙酯萃取。合并的有机相用水洗涤(三次),经硫酸钠干燥,减压下浓缩得到定量收率的标题化合物(65.9mg,定量)。

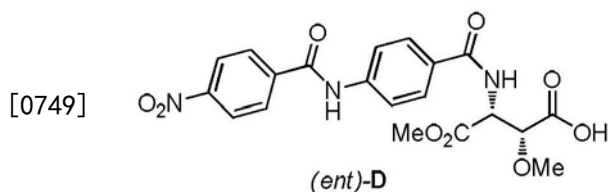
[0744]  $[\alpha]_D^{20} = -16.4^\circ$  (c 1.1, EtOAc);

[0745] <sup>1</sup>H-NMR (400MHz, DMSO, DMSO=2.50ppm):  $\delta$ =3.37 (s, 3H), 3.69 (s, J=3H), 4.34 (d, J=4.4Hz, 1H), 5.09 (dd, J=4.8, 8.8Hz, 1H), 7.89-7.90 (m, 4H), 8.21 (dd, J=2, 6.8Hz, 1H), 8.39 (dd, J=2, 6.8Hz, 1H), 8.55 (d, J=8.8Hz, 1H), 10.8 (s, 1H).

[0746] <sup>13</sup>C-NMR (100MHz, DMSO, DMSO=40.0ppm):  $\delta$ =52.9, 54.8, 58.7, 79.5, 120.0, 124.1, 129.0, 129.2, 129.8, 140.8, 142.2, 149.8, 164.7, 166.6, 170.2, 170.9.

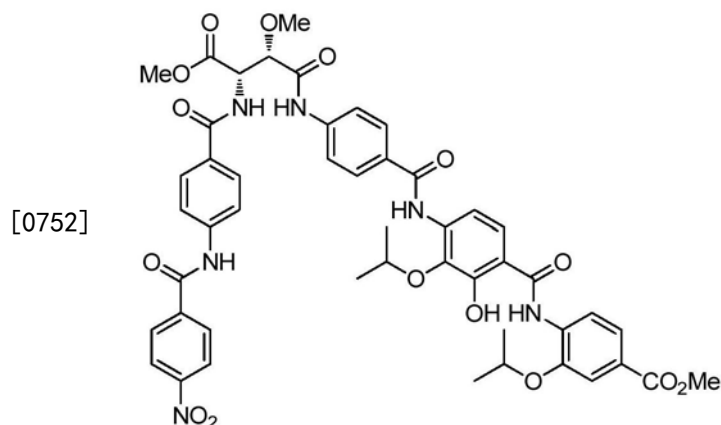
[0747] HRMS (ESI): m/z C<sub>20</sub>H<sub>19</sub>O<sub>9</sub>N<sub>3</sub>Na [M+Na]<sup>+</sup>计算值:468.1019;测量值 468.1016.

[0748] 另一对映体旋光度:



[0750]  $[\alpha]_D^{20} = +13.9^\circ$  (c 1.1, EtOAc);

[0751] 4-(4-(4-((2S,3S)-2,4-二甲氧基-3-(4-(4-硝基苯甲酰氨基)苯甲酰氨基)-4-氧代丁酰胺)苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸甲酯



[0753] 将4-[4-(4-氨基苯甲酰氨基)-2-羟基-3-异丙氧基苯甲酰氨基]-3-异丙氧基苯甲酸甲酯(15.3mg, 0.029mmol)和(2S,3R)-2,4-二甲氧基-3-[4-(4-硝基苯甲酰氨基)苯甲酰氨基]琥珀酸酯(14.2mg, 0.032mmol)溶解于CH<sub>2</sub>Cl<sub>2</sub>(3.4mL)中,并冷却至0℃。然后,加入HOAt(5.9mg, 0.044mmol),DIPEA(7.7μL, 0.044mmol)和EDC·HCl(6.9mg, 0.036mmol)。混合物在0℃至室温下搅拌17小时。在真空中浓缩溶剂,得到油状残余物,将其通过快速色谱纯化(石油醚/乙酸乙酯=94:6),得到标题化合物(20.1mg, 0.021mmol, 73%),为无色油状物。

[0754] <sup>1</sup>H NMR(400MHz, CDCl<sub>3</sub>) δ9.07(s, 1H-OH), 8.37(d, J=7.5Hz, 2H), 8.20(d, J=7.5Hz, 2H), 8.11(s, 1H-NH), 8.02(s, 1H-NH), 8.01(d, J=1.4Hz, 2H), 7.98(d, J=7.5Hz, 2H), 7.90(d, J=1.3Hz, 1H), 7.81(dd, J=7.5, 1.4Hz, 1H), 7.78(d, J=7.4Hz, 1H), 7.69(d, J=7.5Hz, 1H), 7.61(d, J=7.5Hz, 2H), 7.55(s, 1H), 7.54(s, 1H-NH), 7.53(s, 1H), 7.41(d, J=7.5Hz, 1H), 5.72(s, 1H-NH), 5.63(s, 1H-NH), 5.10(d, J=3.8Hz, 1H), 4.76(d, J=3.8Hz, 1H), 4.04-3.98(m, 2H), 3.97(s, J=3.1Hz, 3H), 3.74(s, 3H), 3.32(s, 3H), 1.47(d, J=5.7Hz, 6H), 1.39(d, J=5.7Hz, 6H) ppm.

[0755] <sup>13</sup>C NMR(100MHz, CDCl<sub>3</sub>) δ173.30, 168.15, 168.07, 167.77, 166.93, 166.88, 166.82, 158.83, 151.01, 146.97, 140.78, 139.42, 138.71, 134.97, 134.55, 131.57, 130.00, 130.00, 129.41, 129.41, 129.39, 129.39, 128.12, 127.53, 127.24, 124.17, 124.17, 123.28, 122.61, 122.61, 121.78, 121.78, 121.44, 115.94, 114.88, 113.30, 106.09, 78.00, 75.89, 74.13, 58.51, 56.50, 52.17, 52.08, 21.80, 21.80, 21.80, 21.80 ppm.

[0756] HRMS(ESI): C<sub>48</sub>H<sub>47</sub>N<sub>6</sub>O<sub>15</sub> (M-H)<sup>-</sup>计算值:947.3178,测定值:947.3175.

[0757] 孢囊菌酰胺B

[0758] 4-4-[4-((2S,3S)-2,4-二甲氧基-3-(4-(4-硝基苯甲酰氨基)苯甲酰氨基)-4-氧代丁酰胺)苯甲酰氨基]-2-羟基-3-异丙氧基苯甲酰氨基)-3-异丙氧基苯甲酸甲酯(15.2mg, 0.016mmol)溶于THF/H<sub>2</sub>O(0.2/0.2mL)的1/1混合物中。然后,加入固体LiOH(3.8mg, 0.16mmol),并将该反应混合物在室温下搅拌17小时。水层用1M HCl酸化直至pH~1,并用乙酸乙酯(3x)萃取。合并有机萃取物,用MgSO<sub>4</sub>干燥并过滤。在真空下将溶剂浓缩,得到标题化合物(13.3mg, 0.014mmol, 90%),为黄色蜡状。

[0759] [α]<sub>D</sub><sup>20</sup>=-19.1°(c 1.1, EtOAc)

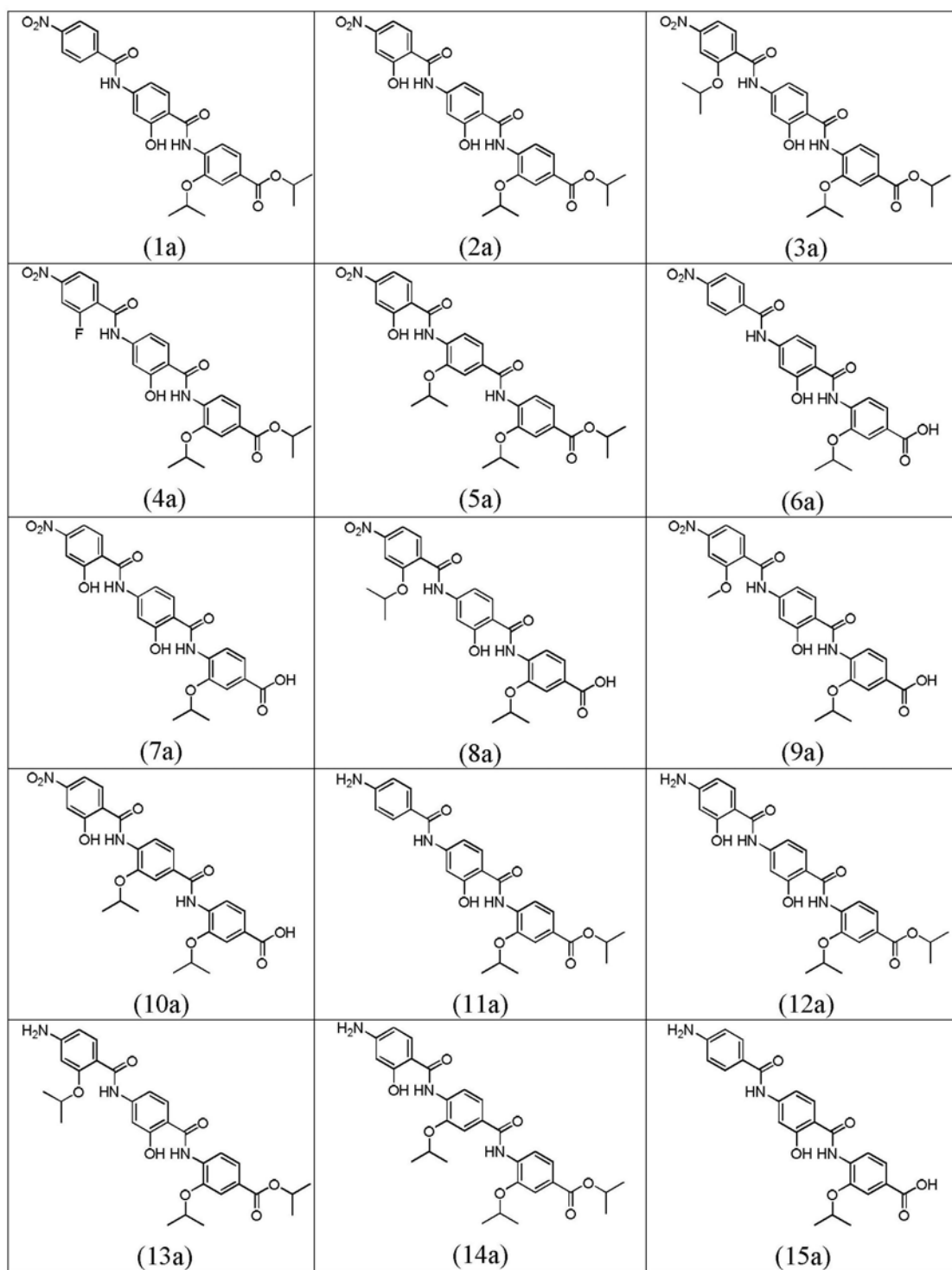
[0760]  $^1\text{H}$  NMR (400MHz,  $\text{CDCl}_3$ )  $\delta$  8.35 (d,  $J=7.5\text{Hz}$ , 2H), 8.15 (d,  $J=7.5\text{Hz}$ , 2H), 8.00 (d,  $J=1.8\text{Hz}$ , 2H), 7.98 (d,  $J=1.8\text{Hz}$ , 2H), 7.90 (d,  $J=1.8\text{Hz}$ , 1H), 7.86 (dd,  $J=7.5, 1.8\text{Hz}$ , 1H), 7.78 (d,  $J=7.5\text{Hz}$ , 1H), 7.65 (s, 1H), 7.63 (d,  $J=7.5\text{Hz}$ , 2H), 7.58 (s,  $1\text{H}_{\text{-NH}}$ ), 7.54 (d,  $J=7.5\text{Hz}$ , 2H), 7.51 (s,  $1\text{H}_{\text{-NH}}$ ), 7.10 (s,  $1\text{H}_{\text{-NH}}$ ), 7.03 (d,  $J=7.5\text{Hz}$ , 1H), 6.35 (s,  $1\text{H}_{\text{-NH}}$ ), 5.57 (s,  $1\text{H}_{\text{-NH}}$ ), 5.42 (s,  $1\text{H}_{\text{-OH}}$ ), 4.93 (s, 1H), 4.70 (s, 1H), 4.01 (hept,  $J=5.6\text{Hz}$ , 1H), 3.95 (hept,  $J=5.6\text{Hz}$ , 1H), 3.38 (s, 3H), 1.48 (s, 6H), 1.47 (s, 6H) ppm.

[0761]  $^{13}\text{C}$  NMR (100MHz  $\text{CDCl}_3$ )  $\delta$  173.30, 169.54, 168.18, 168.07, 167.77, 166.88, 166.82, 158.83, 151.01, 149.88, 140.78, 139.42, 138.71, 136.26, 134.97, 134.55, 130.00, 130.00, 129.41, 129.41, 129.39, 129.39, 128.12, 127.53, 125.15, 124.17, 124.17, 123.28, 122.84, 122.61, 122.61, 121.78, 121.78, 120.41, 113.82, 113.30, 106.09, 77.86, 75.89, 74.13, 58.51, 54.58, 21.80, 21.80, 21.80, 21.80 ppm.

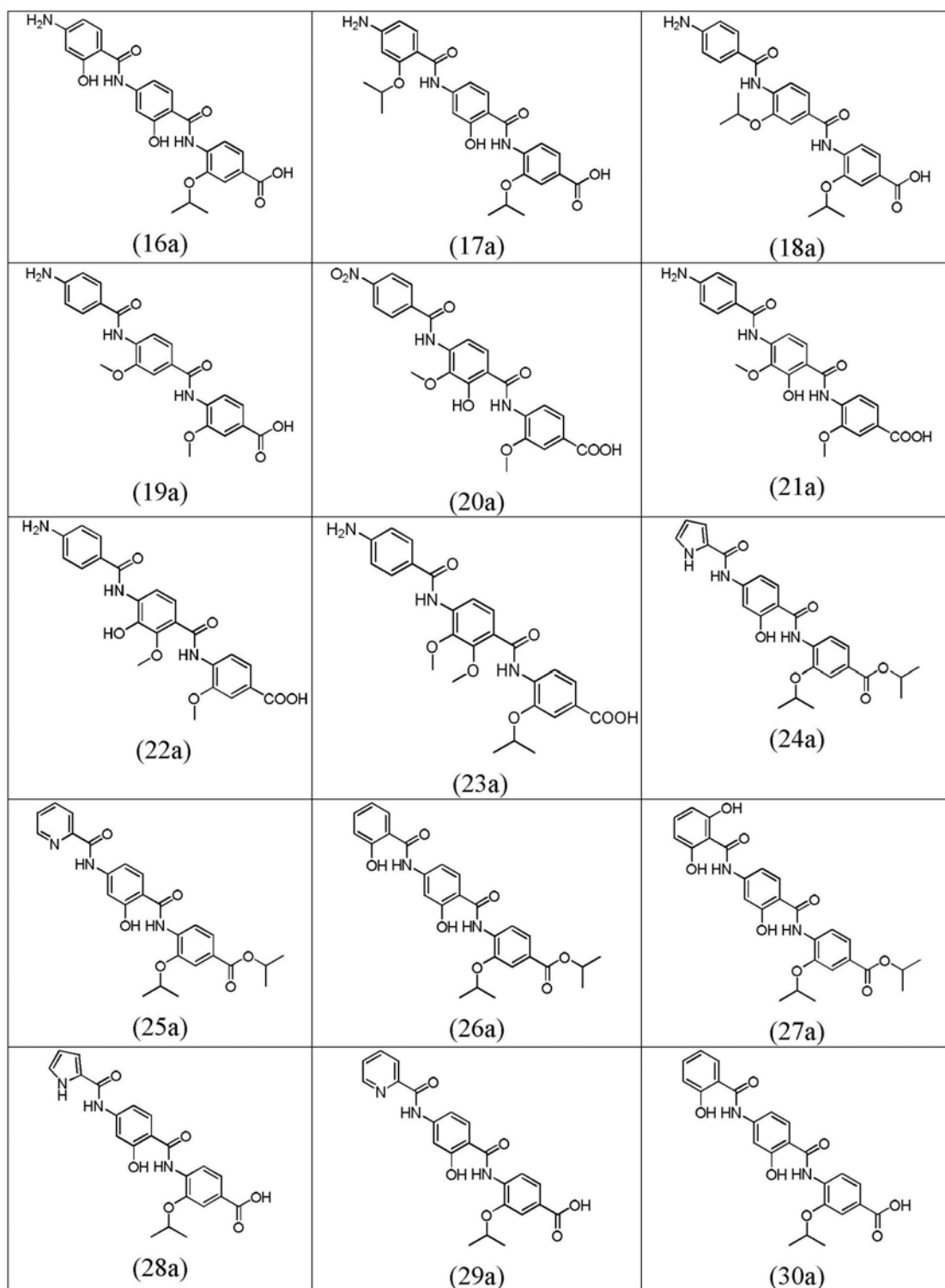
[0762] HRMS (ESI) :  $\text{C}_{46}\text{H}_{43}\text{N}_6\text{O}_{15}$  (M-H) $^-$  计算值: 920.2865, 测定值: 920.2866.

[0763] 孢囊菌酰胺C衍生物的合成

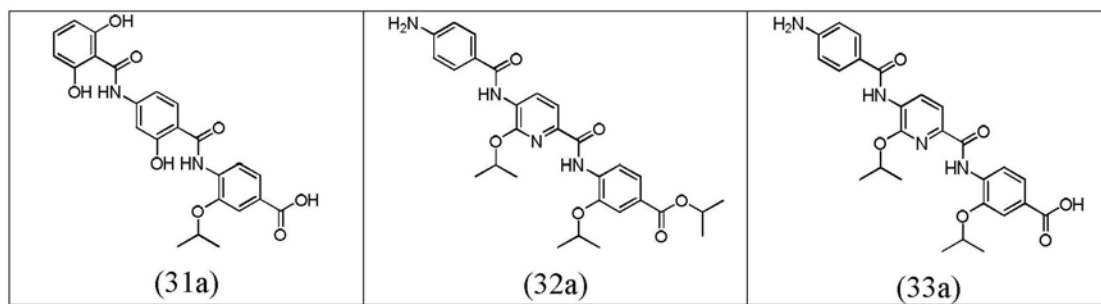
[0764]



[0765]



[0766]

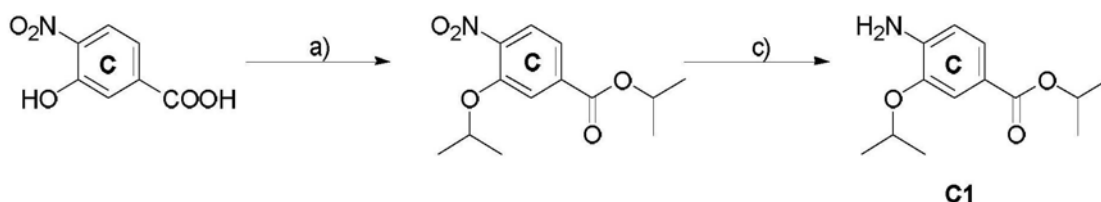




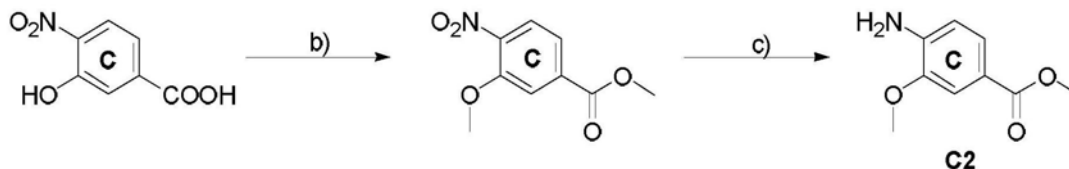
[0767] 1.1.所用的不同单环的合成

[0768] 此处对所述孢囊菌酰胺C衍生物的合成过程中使用的不同的单个环的制备进行说明。

[0769] 环C的制备

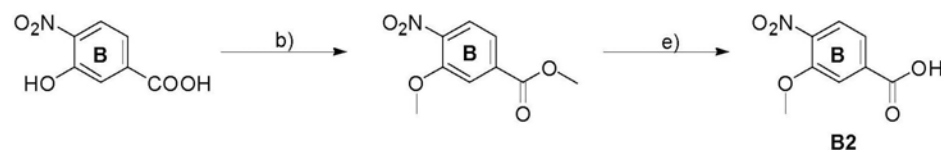
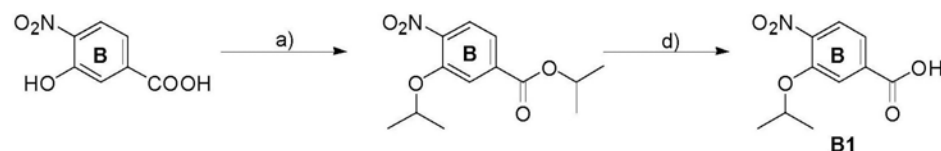


[0770]

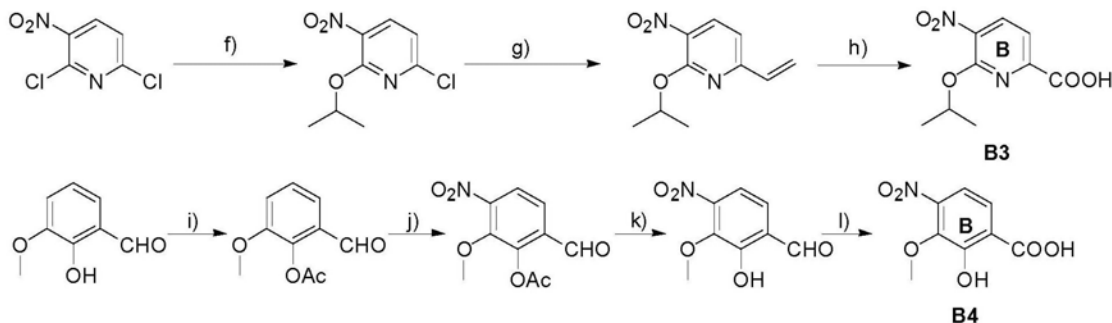


[0771] a)  $\text{BrCH}(\text{CH}_3)_2$ ,  $\text{K}_2\text{CO}_3$ , DMF,  $90^\circ\text{C}$ , 过夜; b)  $\text{SO}_2(\text{OMe})_2$ ,  $\text{K}_2\text{CO}_3$ , DMF,  $90^\circ\text{C}$ , 过夜; c)  $\text{Fe}$ ,  $\text{NH}_4\text{Cl}$ ,  $\text{EtOH}/\text{H}_2\text{O}$ , 回流, 2小时

[0772] 环B的制备

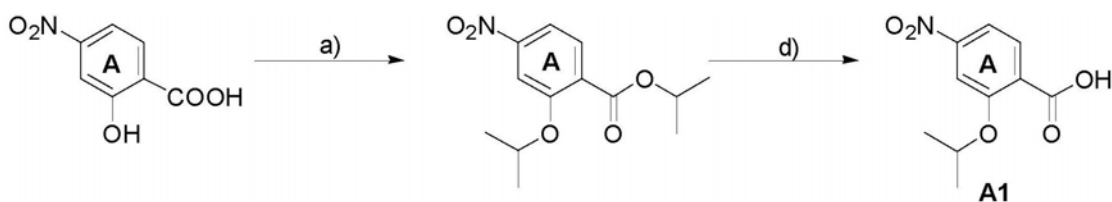


[0773]



[0774] a)  $\text{BrCH}(\text{CH}_3)_2$ ,  $\text{K}_2\text{CO}_3$ , DMF,  $90^\circ\text{C}$ , 过夜; b)  $\text{SO}_2(\text{OMe})_2$ ,  $\text{K}_2\text{CO}_3$ , DMF,  $90^\circ\text{C}$ , 过夜; d)  $\text{NaOH}/\text{MeOH}$ ,  $45^\circ\text{C}$ , 过夜; e)  $\text{KOH}$ ,  $\text{MeOH}/\text{H}_2\text{O}$ ; f)  $i\text{-PrOH}/\text{NaH}$ ; g)  $\text{H}_2\text{C}=\text{CHSn}(\text{Bu})_3$ ,  $\text{Pd}[(\text{Ph})_3\text{P}]_4$ ; h)  $\text{KMnO}_4$ ; i)  $\text{AcCl}/\text{吡啶}$ ; j)  $\text{KNO}_3/\text{TFAA}$ ; k)  $\text{NaOH}$ , l)  $\text{AgNO}_3/\text{NaOH}$

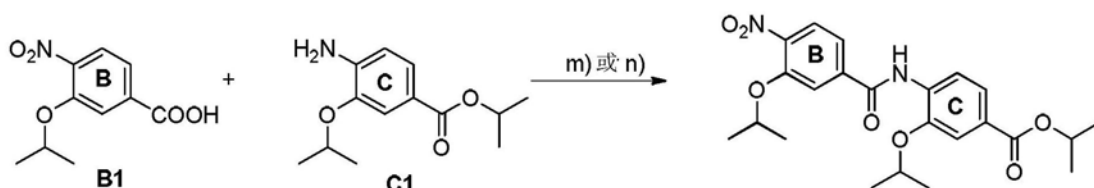
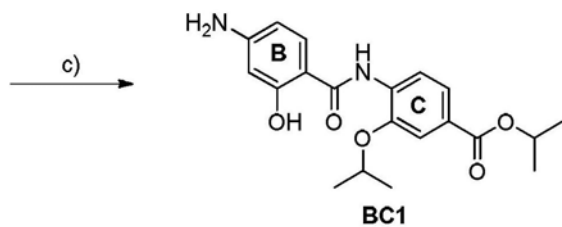
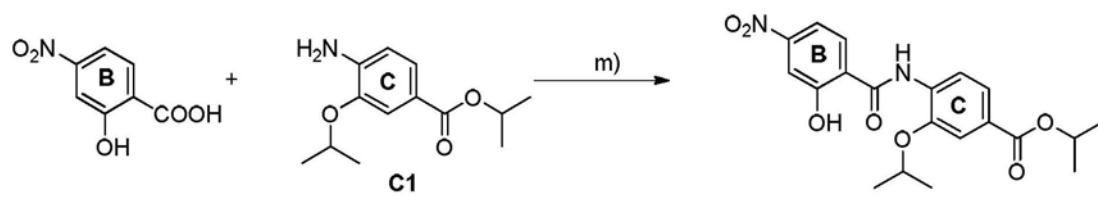
[0775] 环A的制备



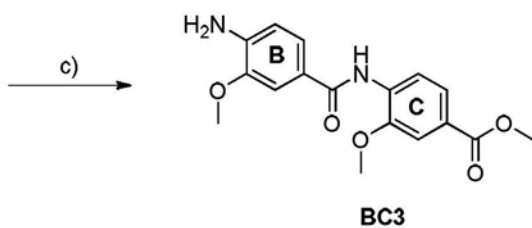
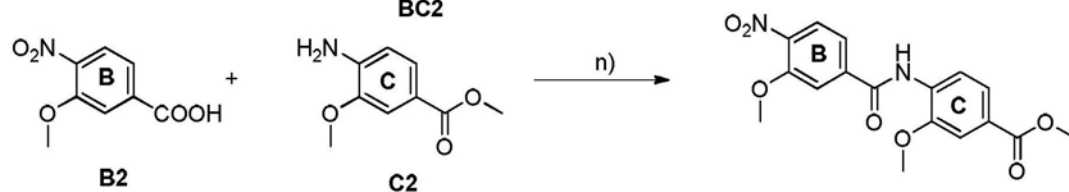
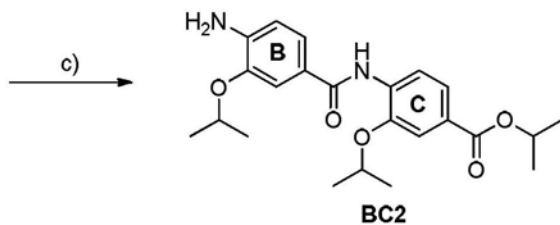
[0776]

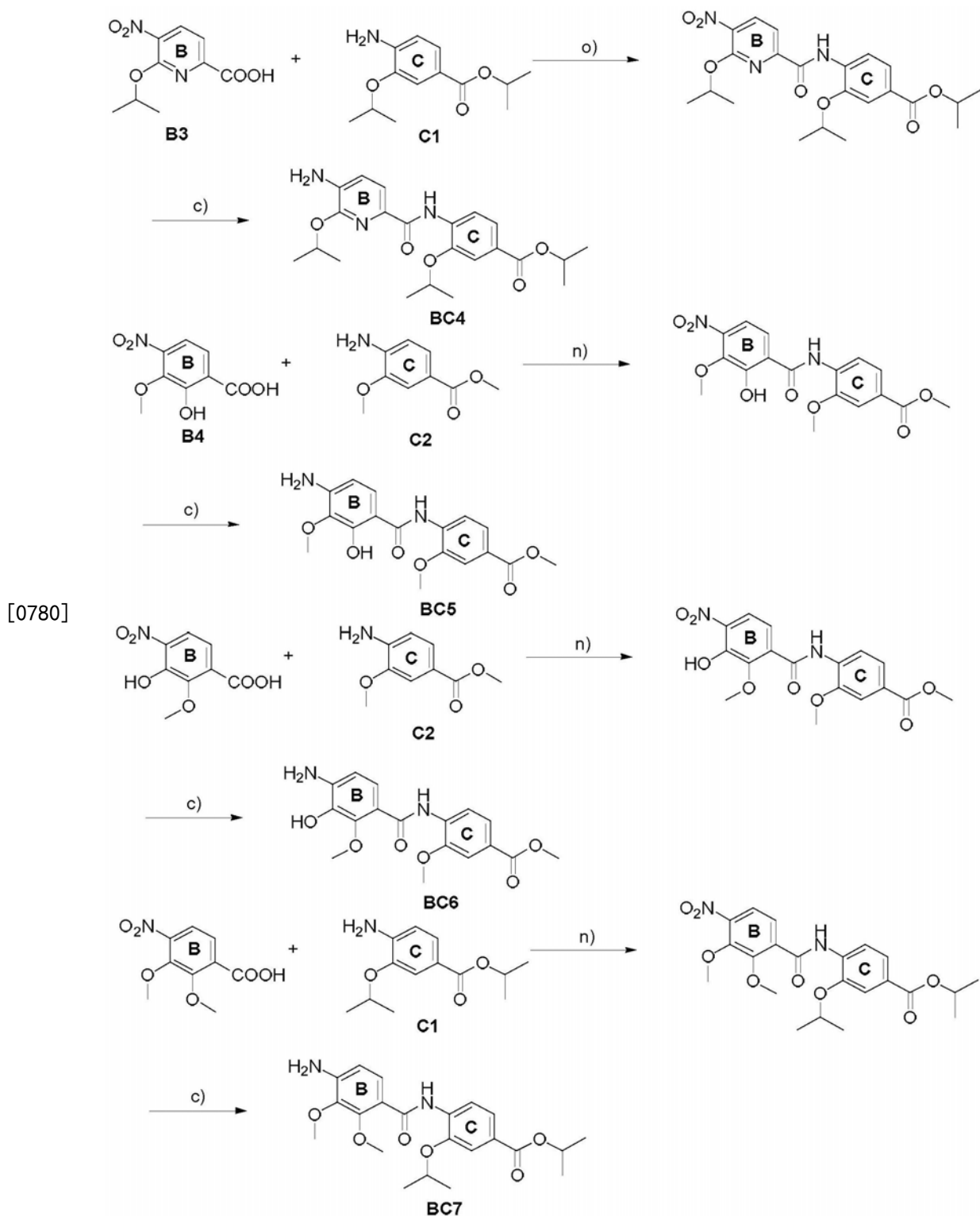
[0777] a)  $\text{BrCH}(\text{CH}_3)_2$ ,  $\text{K}_2\text{CO}_3$ , DMF,  $90^\circ\text{C}$ , 过夜; d)  $\text{NaOH}/\text{MeOH}$ ,  $45^\circ\text{C}$ , 过夜

[0778] 1.2. 环B和C的偶联以提供不同的制备的BC片段



[0779]



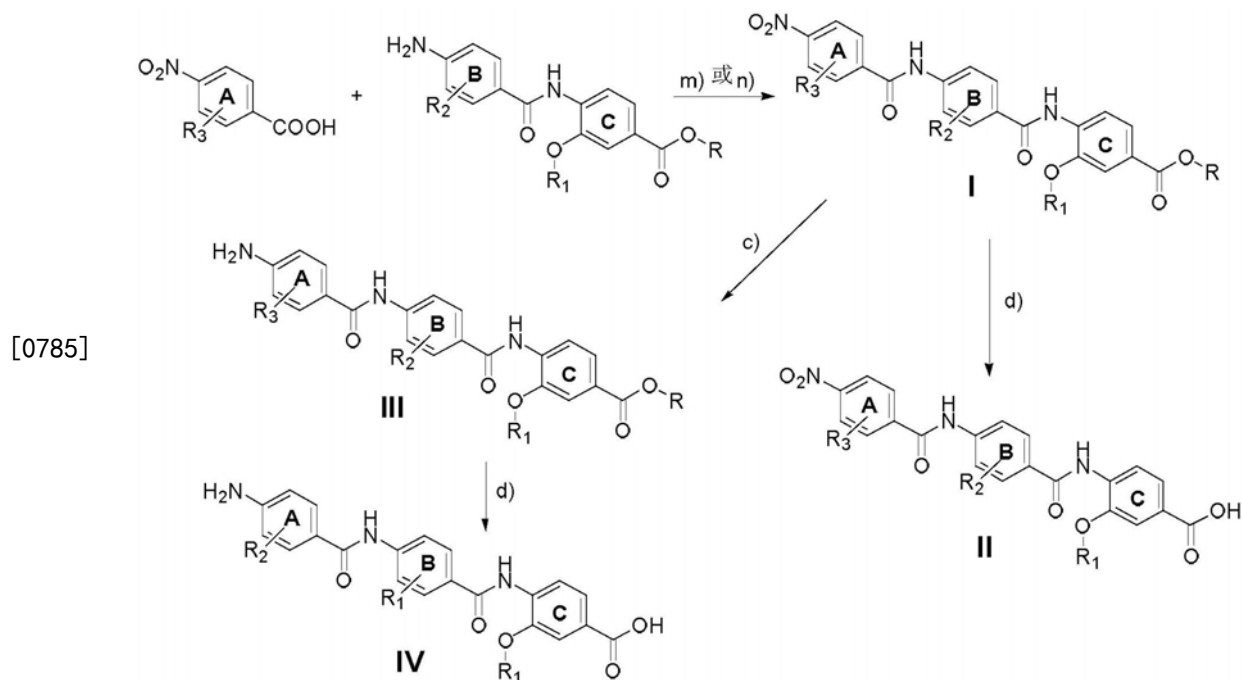


[0781] c)  $\text{Fe}$ ,  $\text{NH}_4\text{Cl}$ ,  $\text{EtOH}/\text{H}_2\text{O}$ , 回流, 2小时; m)  $\text{PCl}_3$ ,  $\text{CH}_2\text{Cl}_2$ , 二甲苯,  $145^\circ\text{C}$ , 2小时; n)  $\text{Cl}_2\text{PPh}_3$ ,  $\text{CHCl}_3$ ; o)

[0782] EDC, HOBT

[0783] 1.3. 环A和BC片段的偶联

[0784] 1.3.1. 环A和BC片段的偶联 (BC1、BC2、BC3、BC5、BC6、BC7) 从而合成孢囊菌酰胺C衍生物 (1a) – (23a)



[0786] c)  $\text{Fe}$ ,  $\text{NH}_4\text{Cl}$ ,  $\text{EtOH}/\text{H}_2\text{O}$ , 回流, 2小时; d)  $\text{NaOH}/\text{MeOH}$ ,  $45^\circ\text{C}$ , 过夜; m)  $\text{PCl}_3$ ,  $\text{CH}_2\text{Cl}_2$ , 二甲苯,  $145^\circ\text{C}$ , 2小时; n)  $\text{Cl}_2\text{PPh}_3$ ,  $\text{CHCl}_3$

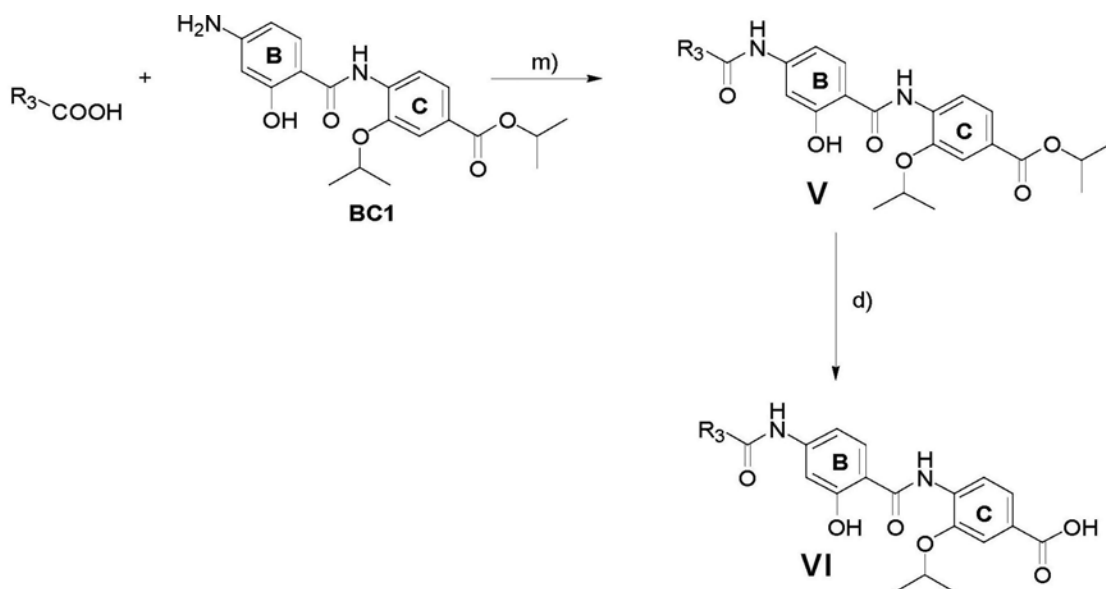
[0787]

化合物	骨架	R	$\text{R}_1$	$\text{R}_2$	$\text{R}_3$
(1a)	I	<i>i</i> Pr	<i>i</i> Pr	2-OH	H
(2a)	I	<i>i</i> Pr	<i>i</i> Pr	2-OH	2-OH
(3a)	I	<i>i</i> Pr	<i>i</i> Pr	2-OH	2-O <i>i</i> Pr
(4a)	I	<i>i</i> Pr	<i>i</i> Pr	2-OH	2-F
(5a)	I	<i>i</i> Pr	<i>i</i> Pr	3-O <i>i</i> Pr	2-OH
(6a)	II	-	<i>i</i> Pr	2-OH	H
(7a)	II	-	<i>i</i> Pr	2-OH	2-OH
(8a)	II	-	<i>i</i> Pr	2-OH	2-O <i>i</i> Pr
(9a)	II	-	<i>i</i> Pr	2-OH	2-OMe
(10a)	II	-	<i>i</i> Pr	3-O <i>i</i> Pr	2-OH
(11a)	III	<i>i</i> Pr	<i>i</i> Pr	2-OH	H
(12a)	III	<i>i</i> Pr	<i>i</i> Pr	2-OH	2-OH
(13a)	III	<i>i</i> Pr	<i>i</i> Pr	2-OH	2-O <i>i</i> Pr
(14a)	III	<i>i</i> Pr	<i>i</i> Pr	3-O <i>i</i> Pr	2-OH
(15a)	IV	-	<i>i</i> Pr	2-OH	H
(16a)	IV	-	<i>i</i> Pr	2-OH	2-OH
(17a)	IV	-	<i>i</i> Pr	2-OH	2-O <i>i</i> Pr
(18a)	IV	-	<i>i</i> Pr	3-O <i>i</i> Pr	H
(19a)	IV	-	Me	3-OMe	H
(20a)	II	-	Me	2-OH, 3OMe	H

[0788]

(21a)	IV	-	Me	2-OH, 3OMe	H
(22a)	IV	-	Me	2-OMe, 3OH	H
(23a)	IV	-	<i>i</i> Pr	2,3-二 OMe	H

[0789] 1.3.2. 环A和BC1片段的偶联从而合成孢囊菌酰胺C衍生物 (24a) - (31a)



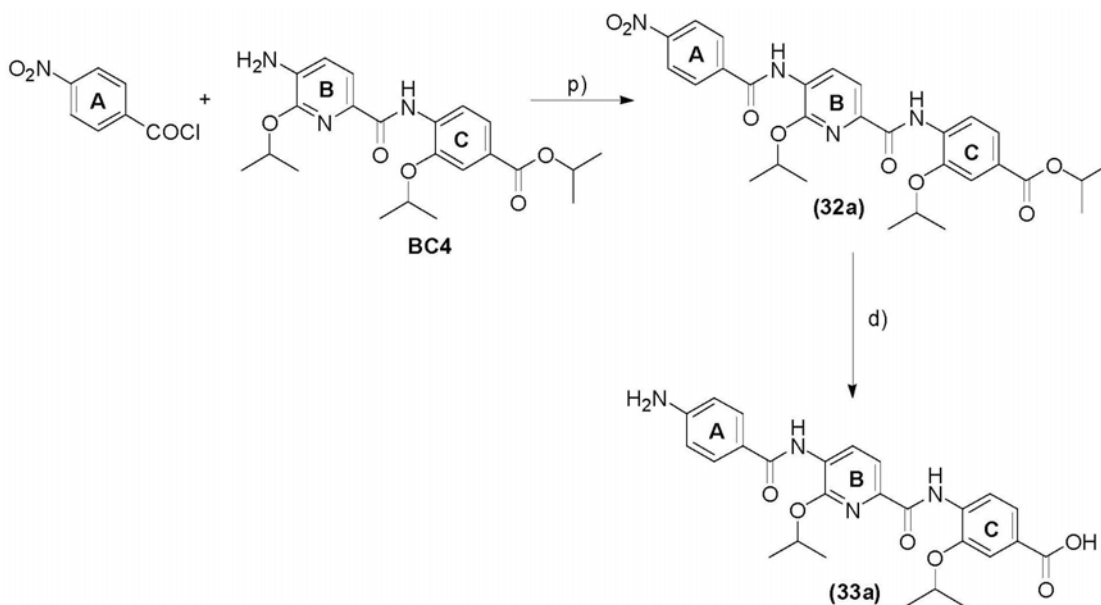
[0790]

[0791] d) NaOH/MeOH, 45°C, 过夜; m)  $PCl_3$ ,  $CH_2Cl_2$ , 二甲苯, 145°C, 2小时

[0792]

化合物	骨架	$R_3$
(24a)	V	
(25a)	V	
(26a)	V	
(27a)	V	
(28a)	VI	
(29a)	VI	
(30a)	VI	
(31a)	VI	

[0793] 1.3.3. 环A和BC4片段的偶联从而合成孢囊菌酰胺C衍生物 (32a) – (33a)



[0794]

[0795] d) NaOH/MeOH, 45°C, 过夜; p) CH<sub>2</sub>Cl<sub>2</sub>, 吡啶, rt, 过夜

[0796] 2. 实验

[0797] 2.1 常规实验信息

[0798] 起始材料和溶剂从供应商处市售购买, 且不进行进一步纯化即被使用。所有 化学产率是指纯化的化合物, 并没有进行优化。反应过程通过TLC硅胶60F<sub>254</sub>铝片, 并通过254nm UV进行可视化。使用硅胶60 Å (40–63μm) 进行快速色谱法。制备RP-HPLC通过包含2767样品管理器、2545二元梯度模块、2998PDA 检测器和3100电子喷雾质谱仪的Waters公司设置进行。使用WatersXBridge柱 (C18, 150×19mm, 5μm) 进行纯化, 二元溶剂系统A和B (A=含0.1%甲酸的水; B=含0.1%甲酸的MeCN) 作为洗脱剂, 流量速率为20mL/min, 并在8分钟施加 60%至95%B的梯度。熔点是在斯图尔特科学熔点测定装置SMP3 (Bibby Sterilin, 英国) 上测定, 并且未校正。NMR光谱是于300K在Bruker DRX-500 (<sup>1</sup>H, 500MHz; <sup>13</sup>C, 126MHz), 或布鲁克傅立叶 (Bruker Fourier) 300 (<sup>1</sup>H, 300MHz; <sup>13</sup>C, 75MHz) 光谱仪上记录。化学位移记录为δ值, 单位为ppm, 参照氘代试剂的卤代残留物为内标 (CDCl<sub>3</sub>: δ=7.26, 77.02; DMSO-d<sub>6</sub>: δ=2.50, 39.99)。分裂模式描述为表观多重性, 并被指定为s (单峰), br s (宽单峰), d (二重峰), dd (二重二重峰), t (三重峰), q (四重峰), m (多重峰)。偶合常数(J) 通过赫兹 (Hz) 表示。在生物测定中使用的所有化合物的纯度为≥95%, 通过LC/MS Finnigan Surveyor MSQ Plue (Thermo Fisher Scientific公司, Dreieich, 德国) 测定。该系统由LC泵, 自动进样器, PDA检测器, 和单四极杆质谱仪, 以及标准的Xcalibur软件工作。RP C18 Nucleodur 100-5 (125×3mm) 柱 (Macherey-Nagel GmbH公司, Dühren, 德国) 用作固定相, 而二元溶剂系统A和B (A=含0.1% TFA的水; B=含0.1% TFA的 MeCN) 作为流动相。在梯度运行中, B的百分比的从初始0分钟的浓度0%, 在 15分钟内增加至100%, 并在100%保持5分钟。注射体积为10μL, 并且流动速率设定为800μL/min。MS (ESI) 分析在3800伏的喷射电压, 350°C的毛细管温度和 10V的源CID下进行。光谱通过100至1000m/z的正离子模式获得, 并在254nm 下紫外光追踪。

[0799] 2.2. 三芳基衍生物 LC/MS数据

[0800]

化合物	LC/MS $m/z$ (ESI+)
(1a)	521.99 [M + H] <sup>+</sup>
(2a)	537.87 [M + H] <sup>+</sup>
(3a)	579.90 [M + H] <sup>+</sup>
(4a)	540.07 [M + H] <sup>+</sup>
(5a)	580.11 [M + H] <sup>+</sup>
(6a)	479.98 [M + H] <sup>+</sup>
(7a)	496.02 [M + H] <sup>+</sup>
(8a)	537.99 [M + H] <sup>+</sup>
(9a)	509.98 [M + H] <sup>+</sup>
(10a)	538.11 [M + H] <sup>+</sup>
(11a)	492.02 [M + H] <sup>+</sup>
(12a)	508.01 [M + H] <sup>+</sup>
(13a)	550.02 [M + H] <sup>+</sup>
(14a)	550.13 [M + H] <sup>+</sup>
(15a)	449.87 [M + H] <sup>+</sup>
(16a)	465.93 [M + H] <sup>+</sup>
(17a)	508.07 [M + H] <sup>+</sup>
(18a)	492 [M + H] <sup>+</sup>
(19a)	435 [M] <sup>+</sup>
(20a)	482 [M + H] <sup>+</sup>
(21a)	452 [M + H] <sup>+</sup>
(22a)	452 [M + H] <sup>+</sup>
(23a)	494 [M + H] <sup>+</sup>
(24a)	466.20 [M + H] <sup>+</sup>
(25a)	478.07 [M + H] <sup>+</sup>
(26a)	493.17 [M + H] <sup>+</sup>
(27a)	509.12 [M + H] <sup>+</sup>
(28a)	423.53 [M + H] <sup>+</sup>
(29a)	436.13 [M + H] <sup>+</sup>

[0801]

(30a)	451.10 [M + H] <sup>+</sup>
(31a)	467.11 [M + H] <sup>+</sup>
(32a)	535 [M + H] <sup>+</sup>
(33a)	493 [M + H] <sup>+</sup>

[0802] 2.3常规合成步骤:

[0803] a) 酸 (25mmol)、异丙基溴 (52mmol) 和碳酸钾 (52mmol) 的100ml DMF溶液在 90℃下

加热过夜。减压下去除多余的DMF,剩余的残留物在水和乙酸乙酯之间分配。将有机层用硫酸钠干燥,然后在减压下除去过量的溶剂,得到纯的产物。

[0804] c) 55℃下,边搅拌边向硝基衍生物(10mmol)的EtOH(60mL)溶液中加入铁粉(2.80g,50mmol),然后加入NH<sub>4</sub>Cl(266mg,5mmol)的水(30mL)溶液。将反应混合物回流1-2小时,然后趁热过滤铁,将滤液在真空下浓缩至干。将残余物用水(30mL)稀释并用NaHCO<sub>3</sub>(饱和水溶液)碱化至pH 7-8。混合物用EtOAc萃取。合并的有机相用盐水洗涤,干燥(MgSO<sub>4</sub>),真空蒸发除去溶剂。将得到的粗物质用正己烷研磨,并通过过滤收集。

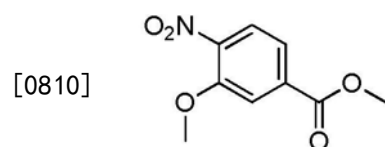
[0805] d) 根据以下报道的步骤<sup>1</sup>进行酯水解。酯(0.1mmol),氢氧化钠1M(3mL)和无水甲醇在45℃下加热过夜。冷却后,将反应混合物酸化至pH 1(3mL,盐酸1 M)并用二氯甲烷(3×150mL)萃取。将有机层用硫酸钠干燥,然后在减压下除去溶剂,得到纯的产物。

[0806] m) 根据以下报道的步骤<sup>2</sup>进行氨基化。煮沸的酸(1mmol)和胺(1mmol)的2.5 ml二甲苯溶液用2M PCl<sub>3</sub>的CH<sub>2</sub>Cl<sub>2</sub>溶液(0.4mmol)进行处理。2小时后,蒸发除去多余溶剂,残留物用柱色谱纯化。

[0807] n) 氮气气氛下,边搅拌边向酸(2mmol),胺(2.4mmol)的无水CHCl<sub>3</sub>(50mL)溶液中加入二氯三苯基膦(3.0g,9mmol)。反应在80℃下加热5h。溶剂通过真空蒸馏除去。残留物通过快速色谱纯化。

[0808] 2.4具体合成步骤:

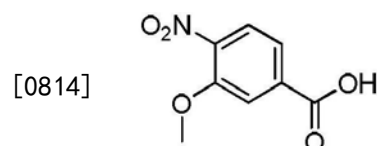
[0809] 3-甲氧基-4-硝基苯甲酸甲酯



[0811] 边搅拌边往3-羟基-4-硝基苯甲酸(9.16g,50mmol)和K<sub>2</sub>CO<sub>3</sub>(15.2g,110 mmol)在DMF(150mL)中的混合物中分批加入硫酸二甲酯(25.2g,200mmol),然后在90℃下将反应搅拌过夜。冷却后,将混合物倒入冰水(400mL),将沉淀物过滤,先后用冷水和正己烷洗涤。

[0812] 产率95%(浅黄色固体),m/z (ESI+) 212[M+H]<sup>+</sup>。

[0813] 3-甲氧基-4-硝基苯甲酸



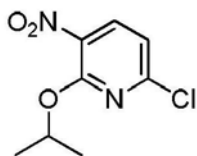
[0815] 边搅拌边往3-甲氧基-4-硝基苯甲酸甲酯(2.11g,10mmol)的MeOH(30mL)溶液中加入KOH(1.68g,30mmol)的水(30mL)溶液。将反应物回流2小时,然后通过真空蒸馏将MeOH蒸发。残留物用水稀释(20mL)。将溶液在冰浴中冷却,用KHSO<sub>4</sub>(饱和水溶液)酸化至pH3-4。析出的固体通过过滤收集,先后用冷水和正己烷洗涤。

[0816] 产率96%(灰白色固体),m/z (ESI+) 198[M+H]<sup>+</sup>。

[0817] 6-氯-2-异丙氧基-3-硝基吡啶



[0818]

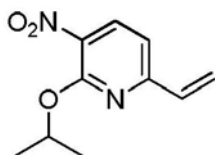


[0819] 边搅拌边往2,6-二氯-3-硝基吡啶 (3.86g, 20mmol) 的甲苯 (30mL) 溶液中加入 异丙醇 (1.44g, 24mmol)。将混合物在0℃下搅拌15分钟。然后在氮气气氛下分批加入NaH (50-60%在矿物油中, 1.22g, 28mmol), 并将该反应物在室温下搅拌过夜。该反应用盐水淬灭, 然后用水稀释并用EtOAc萃取。合并的有机萃取物用盐水洗涤, 干燥 (MgSO<sub>4</sub>), 真空蒸发除去溶剂。将残余物溶于甲苯, 并用快速色谱纯化 (SiO<sub>2</sub>, n-己烷-EtOAc=5:1)。

[0820] 产率70% (微黄色的白色固体), m/z (ESI+) 217 [M+H]<sup>+</sup>。

[0821] 2-异丙氧基-3-硝基-6-乙烯基吡啶

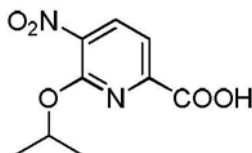
[0822]



[0823] 在氮气气氛中, 边搅拌边向6-氯-2-异丙氧基-3-硝基吡啶 (650mg, 3mmol) 和三丁基(乙烯基)锡 (1.0g, 3.15mmol) 的甲苯溶液 (20mL) 中加入四(三苯基膦)钯 (0) (180mg, 5% eq.)。反应回流过夜。加入盐水后, 将该反应用EtOAc萃取。合并的有机萃取物用盐水洗涤, 干燥 (MgSO<sub>4</sub>), 真空蒸发除去溶剂。粗产物不经纯化直接用于下一步反应。产率90% (黄色液体), m/z (ESI+) 208 [M]<sup>+</sup>。

[0824] 6-异丙氧基-5-硝基吡啶-2-羧酸

[0825]

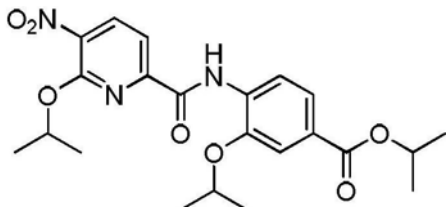


[0826] 边搅拌边向2-异丙氧基-3-硝基-6-乙烯基吡啶 (625mg, 3mmol) 的丙酮 (10mL) 溶液中加入KMnO<sub>4</sub> (1.9g, 12mmol) 的50%的丙酮 (50mL) 水溶液。反应在室温下搅拌24h。加入NaOH 0.5M (5mL), 然后混合物过滤, 滤液在真空下浓缩。残留物在冰浴中冷却, 并小心地用KHSO<sub>4</sub> (饱和水溶液) 酸化至pH 4-5, 然后用EtOAc萃取。合并的有机萃取物用盐水洗涤, 干燥 (MgSO<sub>4</sub>), 真空蒸发除去溶剂。将得到的粗物质用正己烷研磨, 并通过过滤收集。

[0827] 产率75% (米黄色固体), m/z (ESI+) 227 [M+H]<sup>+</sup>。

[0828] 3-异丙氧基-4-[(6-异丙氧基-5-硝基吡啶-2-基)羰基]氨基苯甲酸异丙酯

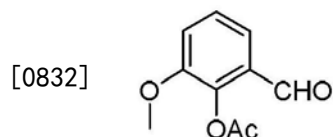
[0829]



[0830] 0℃下, 在氮气气氛中, 边搅拌边往6-异丙氧基-5-硝基吡啶-2-羧酸 (226mg, 1mmol) 和4-氨基-3-异丙氧基苯甲酸异丙酯 (237mg, 1mmol) 的无水CHCl<sub>3</sub> (50 mL) 和DMF (1mL) 混合溶液中加入HOBt (676mg, 5mmol), 然后加入EDC.HCl (958 mg, 5mmol)。反应物在0

℃下搅拌2h,然后在室温下过夜。溶剂通过真空蒸馏除去。将残余物溶于甲苯,并使用快速色谱纯化( $\text{SiO}_2$ , n-己烷-EtOAc=2:1)。产率70% (浅黄色固体),  $m/z$  (ESI+) 446  $[\text{M}+\text{H}]^+$ 。

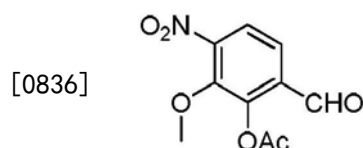
[0831] 2-甲酰基-6-甲氧基苯基乙酸酯



[0833] 边搅拌边向3-甲氧基水杨醛的(4.56g, 30mmol)和吡啶(2.43mL, 30mmol)的DCM(40mL)溶液中滴加加入乙酰氯(2.36g, 30mmol)。反应在室温下搅拌过夜,然后溶剂通过真空蒸馏除去。将残余物在冷的稀HCl溶液中研磨,并过滤,先后用冷水和正己烷洗涤。

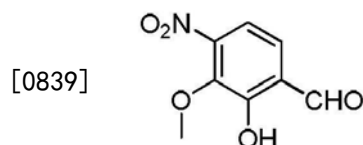
[0834] 产率94% (灰白色固体),  $m/z$  (ESI+) 195  $[\text{M}+\text{H}]^+$ 。

[0835] 6-甲酰基-2-甲氧基-3-硝基苯基乙酸酯



[0837] 边搅拌边向冰冷的2-甲酰基-6-甲氧基苯基乙酸酯(1.94g, 10mmol)和 $\text{KNO}_3$ (1.01g, 10mmol)的 $\text{CHCl}_3$ (15mL)悬浮液中加入三氟乙酸酐(12mL)。反应在冰浴中搅拌2h,然后在室温下过夜。非常小心地用水(50mL)稀释反应物,并用 $\text{CHCl}_3$ 萃取。干燥( $\text{MgSO}_4$ )合并的有机萃取物,真空蒸发除去溶剂。将残余物溶于甲苯,并用快速色谱纯化( $\text{SiO}_2$ , n-己烷-EtOAc=3:1)。产率45% (黄色半固体),  $m/z$  (ESI+) 239  $[\text{M}]^+$ 。

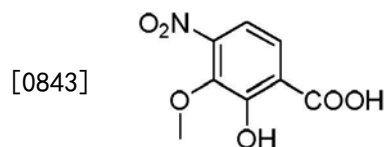
[0838] 2-羟基-3-甲氧基-4-硝基苯甲醛



[0840] 边搅拌边往6-甲酰基-2-甲氧基-3-硝基苯基乙酸酯(957mg, 4mmol)的水(30 mL)悬浊液中加入NaOH(0.8g, 20mmol)。反应回流2h,然后在室温下搅拌过夜。将溶液在冰浴中冷却,用2M HCl酸化至pH3-4。析出的固体通过过滤收集,先后用冷水和正己烷洗涤。

[0841] 产率90% (黄棕色固体),  $m/z$  (ESI+) 197  $[\text{M}]^+$ 。

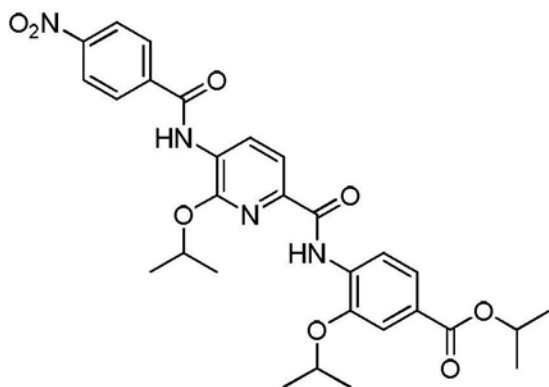
[0842] 2-羟基-3-甲氧基-4-硝基苯甲酸



[0844] 边搅拌边向2-羟基-3-甲氧基-4-硝基苯甲醛(788mg, 4mmol)和NaOH(0.8g, 20mmol)的水(50mL)溶液中分批加入 $\text{AgNO}_3$ (3.4g, 20mmol)。反应回流过夜,然后冷却并通过硅藻土过滤。溶液在冰浴中冷却,并用37% HCl酸化至pH 3-4。析出的固体通过过滤收集,先后用冷水和正己烷洗涤。产率65% (米黄色固体),  $m/z$  (ESI+) 213  $[\text{M}]^+$ 。

[0845] 3-异丙氧基-4-[(6-异丙氧基-5-[(4-硝基苯甲酰基)氨基]吡啶-2-基)羰基]氨基苯甲酸异丙酯

[0846]



[0847] 边搅拌边向4-[[ (5-氨基-6-异丙氧基吡啶-2-基) 羰基]氨基]-3-异丙氧基苯甲酸异丙酯 (207mg, 0.5mmol) 和吡啶 (0.1mL) 的DCM (20mL) 溶液中加入4-硝基苯甲酰氯 (185mg, 1mmol)。反应在室温下搅拌过夜, 然后加入2M HCl (20mL)。混合物依次用DCM和EtOAc萃取。干燥 (MgSO<sub>4</sub>) 合并的有机萃取物, 真空蒸发除去溶剂。将残余物溶于甲苯, 并用快速色谱纯化 (SiO<sub>2</sub>, n-己烷-EtOAc=1:1)。产率80% (黄色晶体), m/z (ESI<sup>+</sup>) 565 [M+H]<sup>+</sup>。

[0848] 5. 参考文献

[0849] 1) Valeria Azzarito, Panchami Prabhakaran, Alice I. Bartlett, Natasha Murphy, Michael J. Hardie, Colin A. Kilner, Thomas A. Edwards, Stuart L. Warriner, Andrew J. Wilson. 2-O-烷基化的对苯酰胺 $\alpha$ -螺旋模拟物: 骨架曲率的作用 (2-O-Alkylated Para-Benzamide $\alpha$ -Helix Mimetics: The Role of Scaffold Curvature.) 有机生物分子化学 (Org. Biomol. Chem.), 2012, 10, 6469.

[0850] 2) Alina Fomovska, Richard D. Wood, Ernest Mui, Jitenter P. Dubey, Leandra R. Ferreira, Mark R. Hickman, Patricia J. Lee, Susan E. Leed, Jennifer M. Auschwitz, William J. Welsh, Caroline Sommerville, Stuart Woods, Craig Roberts, 和 Rima McLeod. 弓形虫的水杨酰苯胺抑制剂 (Salicylanilide Inhibitors of Toxoplasma gondii.), 医学化学杂志 (J. Med. Chem.), 2012, 55 (19), pp 8375-8391.

[0851] 6. 这些化合物的活性

[0852] 根据上述方法测试了这些化合物中的几个对大肠杆菌菌株 (To1C-缺陷型) 的活性。大多数测试的化合物体现出1到320 $\mu$ M的活性 (MIC)。

<110> 亥姆霍兹感染研究中心有限公司

<120> 孢囊菌酰胺

<130> 21191-PCT

<150> EP 13003539.7

<151> 2013-07-12

<160> 73

<170> PatentIn version 3.5

<210> 1

<211> 58456

<212> DNA

<213> 深棕色孢囊杆菌

<220>

<221> misc\_feature

<222> (1).. (58456)

<223> 孢囊菌酰胺生物合成基因簇

<400> 1

[0001]

```

gtagacgccg cggtcagag ggcgggtgcc cagtgttgc agtgggtgcg gtccaggctcg      60
tggccctgca ggcgcagcc gggacaggcg cgcgggtcga tggcgtgctg ccgggtcgcc      120
tgggcgagct ccacggacac gatgcccggt ggcaccgcga ggatgccgta gcccattgatc      180
atcaacaccg aggcgatgaa ctgtccgggc accgtcttgg gcgagaggte tccgtagccc      240
accgtggtca tcgtcacgat ggcccaatac atcccccgcg ggatgtgtgc gaagccgttg      300
gcgcgccctt ccaccatgta caccaccgcg cccatgatga cgaccgtgtc cagcaccgcc      360
cgagggaaga cgatgatctt ccgccgactg gcccgagcgc cggtagagcag cagtcgcc      420
tccccgagga agctggcgag cttgagtacg cgggaagcgc gcagcaggcg gaacaccgc      480
accaccagca gggactgcat gccgggcagc atcaagctca gcaccgaggg caggatggcc      540
agcagatcca ccagcccgaa gaagtcagc gcatagcga gcggccgttt caccgacgc      600
agccgcagca cgtactccag tgtgaagac cgggtgaaac accactcgag gacgcggatg      660
gtctgcccac gcttgacgct gatggactcc acgtctcga gcataccgc gaggacgctg      720
agcacgatgg cccacagcaa tgccacatcg aaggcgcggc ccgccggggt gtccgactcg      780
aagatgattt cgtgcagccg cgcgcggagt cccgacggag cgtctgtctc ggatggatgt      840
ggcacgaggg cagtctagcc ctccacggcg cggcggggcg ggaatgcggt ccgccaccg      900
tgacgcgccg gctactggga gcccgcttg gagctgcgg ggcatgcag ttgccgccgg      960
tcttcttgc cgcccttgtt ggggctccg tgggtgccga actcgccgc gttgcgtgc      1020
ttgggttatt cctcgtcggg ccgcgcctcg cggccccgcg ccgagacgct acccgagtcg      1080
accttcgggt gcagctctc cttgtccgag tgggaatgca gcttctccct gtcttggtcc      1140
tgtgccatcg ggcacctccg ttctctggag gaaacatggg gacggaagac gggagcggct      1200
caggagtgcg gccgcttcgc ggggagggcg ggcgcgcgg cgtctggagg gaaagccgct      1260
gtcgcagatt gggcgttccc tccgcgcga cggaccagcc gcgggaccgg gctcggcgcc      1320
ggcccccgcc aggcgcactc agcgttctt cgcggacttg cgcgcggcgg cggctctcac      1380
ggcgcgcgcc accgttctt tcgccgcgac cgcggcccg gtcaccacct tgcgccgat      1440
gctaccacag agcccggtgg cgtcagctt cttgcgcgcg ggggccttcg gctcggtctc      1500
ctcccgcttg tacgggtag gggccttggc ggtggcctct tgcgtcgcca cttccgggt      1560
gggcttcgcc ttggggcgcg ctttcttcgc cgggtagcgg ggggtcttgg gctccgtgcg      1620
cttcttctcc gcgcgggcgc ggtagcgctt caggtctcgc cgttggcct gtccatagcc      1680
ggttggcgctg tcaggcaggg ccgccttctc ctcttcttg gtgaggcgca tcttgtgttc      1740
ttcattgccg agccggggac ccgagccctt gtaggtctgc gcgggttgcg gcgacaagaa      1800
gggttcgaaa ctgtagggtt ttcccatggc tgtctctcgc ctgcgtgact gggatgtctt      1860
gaagaaataa gtaaggagtg gtccctgatt ttggaatggg cccctctcaa ggcgcctcgc      1920
ggtcccgcta ccaggactct tctcttccc cgtccaggta gcgcaccacg aggcgccgcg      1980
gcttgtgtcg gcgagccagc ctgcgccgcg cgtggatggc ctgggtcacg ttggaatacc      2040
ggccgagctg acggttgccc tcgaatcgga cgtaccaacc ccggccatcc gtcgccacga      2100
tcaccgcgcg tgcgtgtgtg gcggagccca cttegtgctg cttgctctcc ttgcgtcttg      2160
ccgggctcat gaaagcaaac tgtcaaccgc gacggaggtt cgcattgtcc cagggatcag      2220
ggtgcgcgga ttcaggtcgc ccaggagcat gagacggccc cggggacttc aacggcccc      2280
ggagccctcc tgcgtcccg gcacagggtc ctgccagcga gccatcgtgc gccagggagc      2340
gcggcgctag cggctggtgt cgtcttcgga gcccggtcgc ctcatgttgc cgttggagtc      2400
cgtgtgcgcg gagccgtcca cgtcactgcc ggtgccggag ccgccaggt cgtgtccac      2460
gcccgaggcg tcggggctga aaccgcccgc ttactggac acgcgcgagc gcggcgagcc      2520
gcccacgtcg ctaccggagc ccgtgtgtgc actgctcgaa ccgtcgatgt cggcagcacc      2580
tgttccgcgc gtgcctccgc ccgtgtgtgc accaccttcc gtggtgcctc cggctgttcc      2640
ggttccggtc gtctcacttc cggtcattgt gccgccttga gacgtatcgc tcccgcgcgt      2700

```

[0002]

tecttgat	ccagcacgc	ccgtcgtctg	acatcccacg	ccgaagagca	gagccgcgga	2760
caacgcaccc	accaccaccc	ccctgatgtg	tttcatgcgc	tttccctccc	tccagttgga	2820
cacctgtgga	ggctaggaat	ggctccacac	gggtgcattg	gacgtgaaga	cagctccccc	2880
gctcggtgtc	ccactgatgg	tggtctggat	tcttcccttgc	ccctccgagcg	atgaggcacc	2940
ccgtcgtggg	gcgatgggtt	cgaccccggt	ggggctccca	gggcgaggcc	tggcgcgagg	3000
agccgggtgg	cttcgcgcgc	cagacccggt	ccggctactt	ccaggtgtcg	ttgagggtcg	3060
cggctcccga	ggcggggatc	gtggctgagc	ggttggcgcc	gctctcccag	gtgacgttcc	3120
cggagccgct	cttcttgatg	tacttgtatt	cgagggcggt	cgagccgggc	aggetgagcg	3180
tcacgtccca	cttcgggtag	ctggccggag	acaggaggat	ggcggcgccg	gtgttccagt	3240
tgcgagcgcc	ggcatgggta	cccacgaggt	agacgtttctg	tcccacgacg	gtgctggccg	3300
tcacgttgaa	ggtgacggag	gtggccgagg	aggtgctcgt	ggtgacgctc	agggcggtgc	3360
tctgggcgga	ggcattgccc	gcggtgtcgc	gcgcgcgcac	ggtgtagcgg	taggtcgtgc	3420
cggcactcag	gtcgtgtcgc	gtgtagctgg	tggagacggg	tgagcccacc	agcgagccat	3480
cgcggtacac	gtcatatccg	gcgatgcgcg	tggcatccgt	ggaggcgctc	caggagagcg	3540
acacggagga	ggacgtcttg	gacgcgcgcg	tgaggcccga	ggggacggag	ggtgcggtgg	3600
tgtcgagcgc	gggcgcctccg	ctggagacga	cgcgctccct	caccgtggag	gtgcccgcgg	3660
gcaggaggta	gttgttgcce	ttgttgttgt	cccaggtgcc	cttgccatcg	ttgaagacac	3720
actcgtagctg	ggtggcggtt	cccagattga	cgggtgtatt	ggcgtagccc	ggcaccctcg	3780
aggtggccat	gacgttgccc	ggcacggctg	tccacgtgcc	accgccgatg	cggaaagtga	3840
tgtattttgag	ggcgaagtgt	ttgttgaaat	agtagacggt	ggcgcgtgtg	cccgtctggg	3900
tggtagcaga	cagggcgctg	ctgggcgagg	agacattgcc	cgcgcgctcc	cgggcgcgca	3960
cgtgttagctg	gttaggtgggtg	ctcggcgaga	ggccggtatc	cgagttagtc	gtcccctgga	4020
cggacgcgac	ctgggtgcgc	tttcggagca	cctcgtagtt	cgcgacgcgc	tagttgtcgg	4080
tggaggccgt	ccaggcgagg	gccaccgagg	agctcgtcgt	gcccgcgccc	gtcaggcccg	4140
agggaaacga	gggttgggggtg	gtgtccggcg	tcagggtggc	gacgtcagg	gcggtgtctc	4200
ggcgcggacg	gttgcgcgc	gcgtcccggg	cgcgcacggt	gtagctgtat	tgagtgtcgc	4260
gggagaggcc	gtgtcgggta	taggcctgtg	tgggtgtcga	gcccacgtcg	gtgccatcgc	4320
ggaagacgct	atagccgcct	accgccacgt	tgtccgagcg	cgcgctccag	ctcagggtca	4380
ccgagcggtc	cgtcttcgcc	ttcgcggtca	gtcccagagg	gaccgagggc	gcggtgggtg	4440
ccaccacag	gaagggatgc	gttccctgta	ccgtgcggtt	ggggctcctg	atgtacttgc	4500
cgcgccaccg	gttgtaccgg	cccagaccga	cgtagacgga	ctggatctct	ccccgggtga	4560
tgttgccccg	gctgtcgggtg	gcctcgtatg	agtagtcgag	gagctggctg	cggtagttgc	4620
ccaggtagac	gtagtagagg	tcgccgatct	cctgcgcggg	caccttgccc	atgacgggca	4680
ggtaggcggg	ctgccaggaa	acaccattca	tgacaggctt	caggtcgcgc	cgggtgagcg	4740
ggttctggag	ccaggccccc	acgcgggcgc	gatcgatgtt	gggaacaccc	gcggcccttc	4800
gcgcgcggcg	atcatagacc	ttgtgggtgt	tgtcgagcgg	gtcgatgtct	ttgtgggtgt	4860
gcacccggac	gcgggccttg	atggaggaga	tgcgcgtcgc	gtcgtaggcg	taggtgtaga	4920
ggcggaagtg	gttgttgaag	aagtggagcg	tccagccctc	ggacttgtcg	gtgttggcgc	4980
tgcggggggt	gttagggccag	cgttgggccc	accagacgga	ggggcccgctc	ttgtcctggg	5040
cgtatgcctg	ctgcacgtag	ggttgggaga	agtagaggga	ttgattgaag	gacagcgtgg	5100
gcttgacgtt	gtcgtctcgg	ttctcgtcgt	agtagccgaa	gcccgcgagtc	atggcgggca	5160
cgaggaagta	ccaggcgagt	tcgcgggggt	tggcgcgcgc	cgcgccagtcg	ttgttcacgt	5220
cgccttggag	gggaaaggac	atcatccacg	ggttgagctg	gttgcccgtg	tgggtgatct	5280
gcttgtcgat	cgcgggtgggtg	ggcgaccagt	gattgggggtg	cgcgtcgagc	cagatctgct	5340
cggcggtctt	cgcgtagtgt	agggcgccct	ggagcagggc	gaagtgtgcg	tcgaggtagt	5400
gcagccgctg	ctcgaggggag	accgtcatgc	cctcctgcac	gcccgtgagg	ttcgtcttgg	5460
gagagagatt	gaggccgggtg	gcggcggtga	agggcgggaa	ctgacccttc	cagatgccga	5520
agggcagctt	ccagtgtgtc	cactggggat	ccgaggagga	gtcgcgcgtg	tccaccacg	5580
agccgtcctg	gacgtgcacc	acgtcgggtg	agggcggggtg	gtgggtggacg	aggtactcgg	5640
agatgccac	gcactgcacg	ccattggcgc	aggtgacgga	gcggccggtg	taccagggtg	5700
agtcggagcc	ggcgcgtccg	ctcaggttgt	cgcacatcatg	cgcgatgacg	aagaactgcc	5760
gtcggggaa	gaggccctcg	aagctcttga	ggttgacgac	gtcgacggtg	gcctcgcctt	5820
cccagccctc	gagccaggag	ccgttctggt	tgacggggat	gcccgcgacg	cgcgactcgg	5880
cgcgcgtggc	ggggtccacg	tagcgcaccc	agtggggagt	ggaggcgaag	gggtacttgt	5940
tcttgatgac	ctgctgtcgc	tgggccatct	gggcgcctcac	ccaggagccc	acggagctgg	6000
tgttctggag	atcggcgccg	ttgggcgggg	agacgagcgt	gtcggagccc	ggatcgttga	6060
ggtaggggta	gtccttgagg	gtgcgggaga	agtggttgtc	gcccgcgacg	gcccactgca	6120
cgcgcagctt	ggagagggtg	gggatgagge	gtcggagaa	gcccgcgagtcg	gtggggaaga	6180
agcccttggga	ggactggaag	gagccgcga	ggaagtaggg	ctgggcgagc	gtggcgctct	6240
ggtagatgag	atccttgagg	aagttagtcgg	gaccgaccag	gggccccatg	gagtgggtgc	6300
cggtagaagt	gatgagatcc	aggggtcgggt	tgcgcgcggg	ggtgagcagg	gcgctgtagc	6360
ggtccttcca	ggagggcccc	cagttcggat	tgctgtagcc	ggggacgttc	ttcagggtga	6420
cgagatcctg	gacattgttc	accacggcgc	cggacatggt	gacgtgcacc	tggccgggtg	6480
ggccattggt	tttcatgtcc	gaggcgacgc	ttggaggcca	gtacaggtag	gcaccctct	6540
tcgcgttgtg	cgagtaatag	gtgacgaggt	catcgtgcgg	catgggcgcg	cccgatggca	6600
ggtagtattg	gtaattggac	gggggattct	tcttcagggt	gatgacctgc	gcgtcataca	6660
tgtaccggat	ggggcccccgc	gtgggcgttg	acgcgtattg	gcccaggctc	tagtagggcc	6720
gaagattggg	catgtgtgtg	tggtagacgt	gggcgcgcgc	gacttgcgcc	ctggcgggca	6780
gggcacacag	caacagcgcc	gacaggacgg	gcccatacaa	cgccttcaat	cgatgcattg	6840
gggttctctt	ggggtaaagga	ggagcgcacc	ctagtggagc	cgtccggact	ttcctcgttt	6900
ttgatgaaa	aaggatttgc	cgcacgcgcg	caatcgtttg	gcagcagact	ggaacgtcag	6960

[0003]

cgaggagcaa	caacagccac	tggcggcacg	cgcggctctt	ctccagagag	aagagccgcg	7020
cgtgggggag	cgaagccctg	gaggcctgtc	agcccgcgac	ggccacttgt	ggccgcggga	7080
ccggtgtgcg	cgaaggagac	gccgactccc	agaccggaag	tatgcttccc	atcttgtgga	7140
gettgcctc	gcagtaggag	aggttgtcct	cgtacggctt	gttcgccatg	aacggcattt	7200
gagtcgcccg	gtacttcacg	cgcaggctcg	tgttgggagc	gccctcgagc	ccgagtagct	7260
cagacgtgag	ctgcgcgtac	ttttcgggaa	cgcgcaggac	aatggcatcc	ttctccttcg	7320
agacgaccgt	cacgtgtttc	agcaaccctt	cgccattatt	gatgggcgtg	ttgtgtgaca	7380
cgtgcggacg	gttgaggaga	tccatctcga	acaggaactc	cagcagctca	tcgttcacga	7440
tgccccagggt	ttgaatgaac	ccgaagagca	gttggtcgaa	ctctctcgcg	aactcgtgaa	7500
gtgtgacgtg	gaagatgccg	ccgttggcgc	tgaacttcga	ctggctggtc	ccgtcgatca	7560
cgtgtgtgat	gtactgcgtg	taggggtggt	cagggttcgc	cttgacgtac	tcgagaaacg	7620
aggagatcaa	gtccttgaag	gccagccgcc	cctgcttgtc	gaggtacctc	ccgaggaagc	7680
ggagtgcgcg	cagcctgtac	aggtcctgtg	gatgataccc	gaaccgcaca	ccctccttgt	7740
attctcgcg	ggtaacgtcc	ttcgtcgcga	cgcagacctg	cgcctcgtcg	ttcgggtctt	7800
catcatctga	cacctccage	ttgaactcct	cgcgttggtt	gttcateggc	acgttgttga	7860
tgagcaggag	gtgttggtatg	aggatcgcgt	cggcgtcgta	gctcgagagc	ttcccgatgc	7920
ctctcctgaa	ggctctccage	gtctcgcggg	gaagcggcca	gatcatctcc	acgaacgagg	7980
agactcttgt	ggctctcagt	ttctcctgga	ggctcaggta	ggcgtctctc	ttgatgttgc	8040
cgcgttccac	gtctctccage	gtgttcgcgt	ccatcgtctg	gagcagagacc	ggctgggttg	8100
agatcaaaac	ctctctggctc	aggatccgcg	tgatctgcgt	gacccggtca	ggcgagtctt	8160
tcgccgcgct	cagccaaatg	gtgagcggat	agccataact	ccgcttgca	tcggcgatgt	8220
gctgggcgct	ctcaatgtcg	cgggtcagca	tgcgaaatt	cgcgtcggtg	atgaagatgt	8280
agaacgcccc	gtgttggtg	agccagggtga	tctccgcctt	gacccggtcc	atgtcggact	8340
tgaacacgcg	cagattggtc	gccccccccc	agaagcagta	ggtgcactgg	taggggcac	8400
cccggttctg	ctcaaggggc	gcccacacgt	acttctcgt	gtcgaagtag	cttccagggt	8460
aggggagatg	gaccgtgttc	agatcctgga	tgcgcgcttg	gggctcggtc	gtgatcagct	8520
ctccgttccg	gtagaaggag	aggcccttga	cttgccaaag	gtcgggctgg	ggggagcaga	8580
gttcggccag	gtagtccgcg	aaggtatact	caccctcacc	gttcgagagc	accacccgct	8640
cgttgcccg	atccaggtag	tgcgcccggt	ggttcacac	ctgcggacgc	ccaaggatga	8700
ttgtgcgctt	gggcttgcgg	gcggtgagg	tggggagcca	ccgcttcacg	aagcccatgt	8760
tcagacata	gcaagagatc	gcgtagacat	cggcatcgat	cttgttgagc	ttgtcttcga	8820
atcggtcgtc	gttgatcgag	atcgagtgga	tttcgaagct	gcacgactcc	ctgatcaagg	8880
ggtttctgctc	ggccacgcca	cgcattgtagc	cagaggccaa	gggataaacg	ccagagaaga	8940
ccgtcaactc	aatgaatgag	accgcgtggt	tggccatgac	acacgctccc	cgttacctac	9000
aaatttggtat	attgcaacaa	tgttggcggg	caggctagct	gaaaaattta	ctctccggca	9060
ctctcatgtt	cttgggtctc	cgggctcagt	gggcgagcag	cttgaatcgg	cggaacgcct	9120
cgcgcgtcgg	cgcgtgcgac	aggacgtcgc	actgcatcag	cacgtagacc	aggtacccgg	9180
agaggggata	tgcgcgcggg	aaatcctcca	gcggcggctc	cagtagctgc	tcggagagga	9240
tcagctcgat	gtgctcggcc	gctagaccga	gctgctccag	cacgacgcgc	tcgtactccg	9300
ccatcatctt	cgggctgagg	tactcgcgga	tgaactctgg	cagcagctga	ccgatggcga	9360
gccgggcgct	ctctccacac	tcactccaca	ccagcttgag	cacgttcacg	aagaggaccg	9420
cgtggcgccg	ctcgtcgcga	acatggtcct	ccatcacctg	atggagcact	tcgttgaagg	9480
actctccacc	cgtcaggctt	agcagatcct	tggtagtgtg	gttttcccg	atgcagacgg	9540
cgatgatttc	ccagagcccc	tgcagcgtct	cgggcagccg	gtgcttgccg	aaagccatgg	9600
ccctggacag	gtccgtttcc	gttcccaggg	gcagcggctt	gacgcccgtg	cgtgctcga	9660
cttgccgcgt	gaagtcgcgt	gccacgtagg	cgtgatagcg	ctcatcgatg	atgacgggtga	9720
gcgcgtcgtg	cgggatgtcg	tccgggaacg	tgatggcggt	gtgacgggtg	gcgatcttca	9780
tcgccacctc	gttgacggtc	tccgtctcga	agatggcgat	gtcccccacg	aacttgttagg	9840
cggagtggat	gaagaacagc	cgcagcgtct	cgggcgggaag	gttcttcacg	agcgggtggg	9900
tgagcagggc	cgccttggcc	ggcgggtaca	ggtccccgac	cacgtccccc	tcgggcagta	9960
cgcgcggggg	cttgcgtcgt	gtggcggcaa	gcgaatccca	caccaactcc	ttggacttgt	10020
agtcctcgcg	ggaaatcgcg	cccgaacttt	ctgctaccaa	ccctgtcttc	cctgtctctc	10080
cattcactct	ggcttctccg	acggcaccgt	attgctgcat	tgaaggggga	gcgagcgcct	10140
gcgggcgctg	gtcgcgcgcg	ctcagcgtt	gactccgtgc	accaggatc	cctggggcac	10200
accggggggtg	ccgcgtggcg	ccacggggaa	gggcccagcg	ccgagctgct	tcgggatggc	10260
cgcggtggtg	tagacgtggg	ggtcccagcc	aatcggcctt	agaatcacct	cgggctcgtc	10320
ggtgccgaac	tggcgggcga	ggtcgtgcat	caacttgacg	cgcgaggagt	ccaaaatcga	10380
acggccaacg	acgtcgatga	ggacgatgct	ctcgggaacg	ctcagggcgt	tgacacgggc	10440
catgagcagc	gtgacctgcg	cctcggtagg	gtagacgagc	aatccctcga	tgagccacag	10500
ggtggggcacg	cgggacgcga	atccgctttt	cttcagcgcc	gcccggccagt	catcgccag	10560
atcgaccgac	acggcatgct	gtcacattt	cgccgcgacg	ccggtcagct	tcgctcctt	10620
gtcctggagc	acggcgtcgt	ggtcagctc	gaacagccgc	gtgtctcccg	gccaggccaa	10680
acggtagggc	cgggcaccca	ttcccgcgcg	gaggatgacg	atctggcgga	tgcgcggcc	10740
caaccaccag	gtgatctgat	catcgagcca	gcggtgcga	acctcgatgg	cgggagggat	10800
ggcgccctca	cgggcattgc	ggcgcccgag	ctctcgcgac	agcgtgtcac	cggcgagtcg	10860
acgggcaaaag	ggatccccga	acagtgggtt	ggaacgctcg	gtctcaagcg	cgcgcattcc	10920
cgcacaccaa	agtgcctctc	ggcgcatctc	ttgcatatgt	tttatgaccg	ccgcctcgtg	10980
agatgggtta	agggttcggc	aacacgtcaa	ctcgcaacga	cggagcgtc	agcgtccgtg	11040
ctgggattcg	cgaagcgca	acgcgcgcg	ttcgggatcc	tcgcacacgg	cgatcgatc	11100
gccattcgga	agttccatgg	ggccgatgac	gacgcctcgc	gccttgcgga	cgacctccat	11160
cgcgggatcg	agcgcggcga	cgcggaaatg	gaacagccag	tgtgaatgga	cccccttgag	11220



[0004]

```

ccccgccacg tccacgaccg agccggcgct cggctcgtcc gaggccagg tgaactcctg 11280
gtgaaccccc agccgaccaa ggctcgggcg atccgagagc cgcacccga acaggtcgca 11340
atacagggcg gccgtctgtt gcacgttcgc ggcatagagc tgcctccaga ccacctccgg 11400
ctggagcgct ctcgtcgttg ccggtgcgct cggcacggcg aaggtcgccc ctccaggatc 11460
cgaggaggatc gcgacgcgcc cgcgctcgtt cgtcgggtgg gtcgggcca gctgggtcgc 11520
cccgccccc acgaacgagc gcaccgttc atcgacgtcc tcgacgccga cgtaacccag 11580
ccaatggggc ggtgcgcgcg gggcaatcgc ctgctcgggc agcggcacga tgtctcgctt 11640
ggcgggcccc tcaccgaaca gaccggtgta gaacgccctg gcccggggga cgttgggtgt 11700
gcgaactgg agcttgaaga accgtttcat accacgtgac ctctgtaccg ccggggggcc 11760
ggctcagggt gctctagatc cgtcgaccac cattcccaac gccgggcca ggccgacggc 11820
ggtctccacg ccaacacgcg tccctttgac ctggttcgcc ttcgggtcga agatgaaatt 11880
ggtcgatcga gcaaaattgg tcttggtag gacgcagccc tgcattgtgc atccctggcat 11940
atcgggtccc gtgaagtccg actcggcgag gtccacgtca atgaagtgg ctccgcgcgc 12000
ggagcactca acgaaccgcg tcttgcgtag attcaacttc aagaaggagc tgtagcgagc 12060
atcgactgtg tgaactgga cgtccggcat ggttcagagt ccactccagt ccacgcccat 12120
gaggcggggt tctttgaagg tcacgttcgc cagcgcgagc ttctccggtg ccattccgag 12180
gagatcgcat cctcgaata cacaatcctc caggcggtcc cggaccagc ggcttccggg 12240
caacttgcac cgcgggaacg tgcagcgtc gaattccttg ccgagagat cagccgactc 12300
gatcgagaga tcagaaaacg tgacgtcggc gaaaaagtcg ccactttcca gagaggagat 12360
ggagcgggcg gccatatgtc tctatggctg acacgacgag cggccctaa taccagtgcg 12420
tgcgtatagg atccagcaca gtcagtccca tgtcctccaa ccgacatata gtctagatga 12480
gaacaaacta agttgactcc acaacacgtg tgtgccttga atcgagcata actgaactcg 12540
tgccgtgcgc gggccgaatg ccgagtgcgt cggcctgtcc ggaaacgcct gccctgtccg 12600
gagacggccc caactggggc gcacgcgccc tggtaggtcg gtggacgcag ggcaggaagc 12660
gtatttggcc gaaccgaag ttgcggggcc gcgaggtcgc gcaggggaa gacgatgagc 12720
atgaacgggg atcgaaccca gtacgttgtc ttgatcaacg gcgaagagca gtactcgctc 12780
tgccccgtgc acccgcaaat tccgggctgt tggaagaccg ttgggcccga gggaagcaag 12840
gaacagtgtc agtctacat ccaggaggtc tggacggaca tgaggccgaa atcgctacgg 12900
gaagccctga cgcgcagcaa ctgctgatcc cgtgcctcg ggggtccttg taccgccgtc 12960
gtctccagat gaggattgca gcgagggcac aaccaatgag tacgccagca gcaggagcga 13020
agccgtctca tctcggggtg attgaaacgg tgatggctga acctgagctt gaggaggttc 13080
gtacctgac cgtggagagc ggcgacggac ggcagagtac cctctatgag ttcgggtccga 13140
aggacgcgga gaaggtcgtg gtcttgcgcg cctacggagt cactctcttg ctgttggcgc 13200
gactcgcggc gctcctctcc cagcgattcc atgtcttgat ttgggagtcg aggggggtgc 13260
cggactccgc cctccgggtg tatgacacgg accttgggct cgcgcaccag tcaaggcatt 13320
tctccagggt cctcaagcag cagggtctcg aggcgtttca ctctcgtcgc tgggtgcagg 13380
cggcgagctt gcccggtcat gccaccgcca gcggccaggt caagcccgcg acgatgtctt 13440
ggattgcccc ggcggggctg ggttactcgc tggtaagtc cagattcgat cgtatgtcac 13500
tgcctatcta cctcgagatc gagaagcatg gcctgttgca cgcgcgagaag ctccgcagcg 13560
ttctgaacaa atacaatggc gtccccgcga cggcgagaa cgcggcgga aagctgacga 13620
tgcgccattt gcccgaccgc cggatgacat acgtctctc caggtacatg aaggcgtatg 13680
aagacaacag gctcctgcgc aagcaatttg tctcgaccgc gctcgactcg gtcccgacgc 13740
tgccatttca ctcgggggac gacacgtaca gccattctc ggagtccgtt cagctctcga 13800
agctgcatec atccctcgag ctctgcctac tcggtaaggg cggccatctg cagatcttca 13860
acgaccgcgc cactactggc gactacgttc tcggtttcat cgacaccagg gcgtcgagg 13920
ctgccgtccc tgcggtgggc ggagcgtagg gagacaacat gatacttccc aacaacatcg 13980
gccctgacga gcggacgcag ctgcacggc agatctctc gtaccagaag aagttccacg 14040
tgtgtggcg cgagcggggg cccaccgagt tctcgtatcg gcagatgcgc ctctgcacgc 14100
cgaccggggc ggtcagcggc gtggactggg ccgagtacaa gacgatgcgt cccgacgagt 14160
atcgctgggg cctcttcatg gtgcgatgg accaggacga gatcgcttc ggcgaccacc 14220
gtggcaagaa ggcgtgggag gaggttccga gcgaataccg cacgtcgtc ctgcagcaca 14280
tctgcgtgca gcccgacgtg gagaacgcg ccgtcgagca gagccggctg ctgacgcaga 14340
tgccgccgag caaccgggac ctggagaacg tgttccagtt ctctctcgag gaggggcgcc 14400
acacctgggc catggttcc cctctgctcg cccacttcgg tgaggacggg gtcgtcgagg 14460
ccgaagcgct cctggagcgg ctgagcggtg acccgaggaa cccccgttg ctggaggcgt 14520
tcaactatcc gaccgaggac tggctgtccc acttcatgtg gtgcttgcgt gccgaccggg 14580
ttggcaagta ccagatacat gcagtacccg aggtctcgtt cgcctcgttg gcccgggcgg 14640
cgaagtctat gatgttcgag gaaccgtccc acatcgccat gggcgccgtg ggtctggaac 14700
gagtgtctgc caggaccgcc gaggtcacc tgcgtgaggg gacgttcgat acgttccacg 14760
cgggggcgat tccgttccc gttgtccaga agtatctcaa ttattggcg ccgaaggtct 14820
acgacctctt cggaaacgac ggtctcgaac gctcgaaaga actcttcgg gctgggtccc 14880
ggaggccgcg gaatttctgt ggaagcgaat cgcagatcgt tcgcatcgat gagcgcatgg 14940
ggagcgact gaccgtcgtg gaagtggag gggagtgggc gatcaacgcc atcatgagc 15000
gacacttcat cgcgaagtg caaacgtcca ttgatcgtg gaacgccagc ctgcgagcgc 15060
tggcgctcga ctccagttg tacctccctc acgagcgtt cagcaggacc tatggccctt 15120
gcgccgtctt gcccttcgac gtggacggaa aactgctccc ccgcggcacg gaggcgaagc 15180
tcgccgagta ctccccaca cctcgcgaac tcgcgaacgt ccgctcgtg atgcagcggg 15240
agctggctcc cgggcagtag tctcgttga tcgccccgtc cgcgacgcgg ctgagcgcgc 15300
tggttccagtc caggaaacac cccaaggagc acgaatgaaa cgaagccgtc ggatcgttga 15360
cgggagaaga gcgagcagtt cgtgggaacg ggagaggggc tcgccatgag cggcaagctg 15420
ctctctcgta tgtgtccgac accccggaac gacactcat cacatgcgtt gcctcatcat 15480

```

[0005]

cgacaactac	gattcgttca	cctggaatct	ggcggactac	gttgcgcaga	cgttcgggag	15540
cgagccgttg	gtcgtccgca	acgaccagca	tacctggcaa	gaaatcaagg	cettgggctc	15600
cttcgggatgc	atcctgtgtt	ctccgggtcc	gggctcgggtg	accaatccga	aggatttcaa	15660
tgtctcgcga	gacgcgctcg	agcaggatga	gttccccgtg	tttggggctc	gcctgggcca	15720
tcaggggctg	gcgtacatct	acgggggcga	gatcactcac	gctccggttc	cgttccacgg	15780
caggacgtcg	accatctacc	atgacggcac	gggcgtgttt	cagggaactcc	cgccgagctt	15840
cgacgcgggtg	agatatcact	cgttggtcgt	gcggccggag	tcgcttcccg	cgaacctggt	15900
cgtcacgcgt	cggacggaat	gcggcctgat	catgggggtg	cggcacgtga	gtcgcgccaa	15960
gtggggcgctc	cagttccatc	ccgagtcgat	tctgactcgc	cacggcttgc	agctcatctc	16020
caatttccgt	gacagagcgt	accgatacgc	ggggaaagag	gttccgctgc	gccgtcccca	16080
ttcgaactgcc	ggcaacgggtg	tcggcgcagg	tgcgtccagg	cgtgaccgca	gcgccccccg	16140
cacaccggag	cggagaaggg	aacttcagac	gttcaccagg	cggctggcga	cgctctctga	16200
ggccgagacc	gttttctcgg	gcctgtatgc	gggcgcgcag	cactgcttct	ggctcgacag	16260
ccagtcctg	agagaaggga	tatcccggtt	ctccttcacg	ggctcgtgc	cggagggtc	16320
gctgctgacg	tacggcgctg	cggaagcggc	gtcagagggg	ggcgccgagc	ggtacctggc	16380
ggcgctggag	cgggcgctcg	aaagccgtat	cgttggttcg	cccgtggatg	ggctgccatt	16440
cgagtttcat	ggcggttaca	tcggcttcat	gacctacgaa	atgaaggagg	cgtttggggc	16500
cgcagcagag	ttcaagaaca	ctattccgga	cgccttggtg	atgcacgtga	agcggttctc	16560
ggcgcttcgac	cactcgacgc	gagaagtgtg	gctggctgcc	atcgccgagc	tcgaggagag	16620
cgcgagcgctc	ctcgccctga	tggacgagac	cgcgcgacgt	ctgaagtgcg	ttccgcgcgg	16680
cacccgcttcg	ccccagtcce	tgggggttga	atccatctcg	gtatcaatgg	attgtggacg	16740
ggatgactac	ttccgcccca	tcgagcgctg	caaggagaag	atcgctgatg	gggagtccta	16800
cgaggtctgc	ttgacgaacg	gtttctcgtt	cgatctgaag	ctggatcccg	tcgagctgta	16860
cgtgacgatg	cggagaggca	atcccgcccc	gttcggcgct	ttcatcaaga	caggcaagac	16920
ctcgctccctc	agttaccctcc	cggagcgctt	cctgaagggtg	gatgaggatg	ggacgggtcca	16980
ggccaaagcgg	atcgaaggga	cctgcgcgcg	ctctgacgac	cccgcaccgc	acagcacgaa	17040
tgcgcgcggg	ctggcccgct	cggagaaggga	ccggcgaggag	aaactgatga	tcgtggacct	17100
gatgcggaac	gacctcggac	gggtgtccgt	gccgggcagc	gtccatgtct	ccaatctaat	17160
ggacatcgag	agcttcaaga	cggctccatca	gatggctcagc	accgtcgaat	cgaccttgac	17220
gccggagctgc	agcctcgttg	acctcctcgc	cgcggctctc	ccggggggatg	ccatcacccg	17280
ggctcccaag	atccgcacga	tggagatcat	cgatcggtc	gagaagagcc	ctcggggcat	17340
ctactgcggc	acgatcggtc	acctcgggta	caaccggatc	gccgacctga	acatcgccat	17400
ccgcaccttg	tcttacgacg	gcacctcctg	gaagttcggg	gccggcgagg	ccatcaccta	17460
cttgtcacag	cggagggggg	agtttcagga	gatcctgctc	aaggcggaat	ccatcctccg	17520
ggcagtttgg	cgtgtacatca	atggcgcggg	tgtcccttc	gaacccaggt	tgcgcgaccg	17580
ggttctgtgc	ctggaggaga	agccgcgaag	ggtcattcgt	ggccacgggt	cggaattga	17640
tgcagtggag	cctagcgctg	gaagcctacg	tcgagtcgag	acctgcccat	tcgcgcgtca	17700
agcccccagg	gaccatccga	accgcgtgcg	cgtccccggg	gccagtgatg	gattgcgttc	17760
aaaccgcagg	cgcgcgccag	gctcggctc	ttctgcttc	cgtacgcggg	tggcgacgcg	17820
aacatcttcc	gggactgggc	cgcggcgatg	cccagggggg	tcgaggtcct	cgccgttcag	17880
tacccccggc	gggttaccaa	cctggcggtg	ccgcgatca	gcgactgtga	cgagatggcg	17940
tcacaactgc	tggcggtgat	gacgcggtg	cttggcatca	acttcgctt	tttcggccac	18000
agcaatggcg	ccttgatcag	cttcgagggtg	gcgcgaaggc	tcacgacga	actgaagggc	18060
cgcatcgcg	atcaattctc	gtcgcccaag	tcgccccctc	actaccgaa	caacaggagt	18120
aagatcagcg	gcctcaacga	caggacattt	ctccgggcga	tcgggaagat	ggcggttacg	18180
ccccaggaag	tgtcgcagca	cgcgcggctg	atgcagattc	tgttgccaa	actgcgcgcg	18240
agcttcgcgc	tggcgagagc	gtatgtgttt	cgcgccggac	ccacctgac	gtgcgacgtc	18300
agcactctgc	gaggcgagag	cgaccacctg	gtcgacggcg	agttctgtca	gcggttggtc	18360
gagctgacga	cggcgcgccg	gagccagtac	gcaatagatg	gtggccattt	cttcttgaat	18420
tcccacaagt	cgcaggtcgt	ggcgctcgtg	cgagcggcac	tgtttgagtg	tgtgtttag	18480
cgagaaaacg	gattcccaaa	taatgaccgc	tcagaacca	gcctccgctg	tttctttcga	18540
tctcttctac	acgacgggtca	atgcgtacta	cgggaccgcc	gccgtcaagg	cggccatcga	18600
gctcggcggtg	ttcgacgtcg	ttggcgagaa	gggcaagacc	ctggccgaga	tcgcgaaggc	18660
ctgcaacgcg	tcgcccgtg	gcattccgcat	tctctgcggg	ttctcgtgt	cgatcgggtt	18720
cctcaagaat	gggggtgagt	tgtttcttct	cacgcgagag	atggccctgt	ttctggacaa	18780
gaagtcgccc	ggctatctgg	gcggcagcat	tgatttctct	ctgtcgccgt	acatcatgga	18840
cggttccaag	gacctcgcgt	cggttggtgcg	gacggcgag	ttgacgtgc	cggaanaagg	18900
gggtggtggcg	ccagatcctc	cgcagtggtg	gacgttcgcg	cgcgcgatgg	cgcgatgat	18960
gtccctgccca	tcctctctgc	tcgcggaact	ggcggaaccg	caggcgaacc	agccgctcaa	19020
ggtgctcgat	gtcgcgccg	gccacggcct	cttcggcctg	gccatcgccc	agcggaatcc	19080
gaaggcgcat	gtgacgttcc	tcgactggga	aaacgtgcta	caggtggcgc	gcgagaacgc	19140
gacgaaggcg	ggagttctcg	acagggtcga	gttccgcccg	ggagatgcct	tctcgtgga	19200
cttcggcaag	gagctggacg	tcactctctc	gacgaacttc	ttgcatcact	tcgacgagcg	19260
ggcctcgcag	aagatctctca	agaaggccca	cgtgcctctg	aaggaggcg	gccgtgtgct	19320
gacgttcgag	ttcatcgca	acgaggaccg	gacgtgcctc	ccgttgccg	ccaggttcag	19380
catgatgatg	ctcgccacga	cgcgcggcgg	tgagacctac	gcctactccg	atctggagcg	19440
gatgttcaag	aacacgggtt	acgatcaagt	cgagctcaag	gccattctct	cccgatgga	19500
gaaggtcgtc	gtttcgatca	aggcacaagc	gcagctctga	gcaacattca	gcacaatagg	19560
acctctggg	agatttccga	tggccacca	attgtctgac	ttcgcgtccc	tcgactccga	19620
agacgccaac	gtcatctccc	gctcgaacga	gacggggata	tcgctggatc	tgtccaagag	19680
cgtgggtgac	ttgttcaacc	tccaggtcga	gagggcgcct	gacgccacgg	cgtgtctcgg	19740



[0006]

ccgccagggg	cgttgcactt	acggagaact	caaccggcgg	acgaaccage	tcgcgcata	19800
cttgatcgcg	cgaggcgctg	ggccggatgt	tcccgtgggc	gtctgttctg	agegctccgc	19860
cgagcagctc	atcgccatcc	tgggcgtcct	caaggcgggc	gggtgttatg	tcccgttggg	19920
tcgcagtae	cccgcgatt	acatgcagca	ggctctgacg	gaagcccggc	cgcggatggt	19980
ggtgtcgagc	cgggcgctcg	gcgagcgctt	ccgctcgggc	gaggagcaga	tcgtctacct	20040
cgatgacgaa	cagctcctgg	cgcgcgagac	ccgcgaccgc	ccigtgaagg	tgttgccgga	20100
gcagctcgcg	tacgtgatgt	acacgtcggg	ctcgtccgga	gtgccgaagg	gcgtcatggt	20160
gccccatcgc	cagatctcca	actggctgca	tgcactcctg	gcgcgggtgc	cgttcggcga	20220
gaacgaagtg	gtggcccaga	agacgtccac	gtcattcgcc	atctcagtga	aggaactctt	20280
cgcgggattg	gtcgcgggtg	tcccgcaggt	cttcacgacg	gatgcgactg	tccgcgacgt	20340
tgccagcttc	gttcgtgagc	tggagcagtg	gcgcgtcacg	cggctctata	cttttccctc	20400
ccagctggcg	gcgattctct	cgagcgtgaa	tggcgcgtac	gagcgctccc	gctcgtctcg	20460
ccacctgtac	atctcgatcg	agccctgccc	aacagagctg	ctggcgaagc	tccgggcggc	20520
ggcggcggtg	gtcacccctt	ggtagatcta	tggctgcacc	gagatcaacg	acgtcaccta	20580
ctgcgaccca	ggggaccagg	ctggcaacac	gggcttcgtg	ccgatcgggc	ggcccatccg	20640
caacacgcgg	gtgttcgtcc	tcgacgaaga	gctccggatg	gtgcccgctg	gcgcgatggg	20700
ttagatgtac	gtggagagcc	tgagcacggc	gcggggctac	tggggccttc	ccgagttgac	20760
ggcggagcgt	gtcagtccca	accctcacgc	ggaggacggg	tcgcgcctgt	acaagacagg	20820
cgacctgcgc	cgtacctgcg	cggatggttc	cctggagttc	ctcgggcgcc	gggactacga	20880
ggtgaagatc	cgcgggtatc	gcgtggacgt	ccggcaggtc	gagaaggctc	tccgggcgca	20940
tcccgcacat	ctcgagggtg	cgggtggtggg	ctggccgctc	ggcggggcga	atccacaact	21000
ggtcgcctac	gtcgtgccga	gggcgaaggg	ggtctgctcc	atccaggaga	tccgggacta	21060
ccgttcggcg	tccctgcggg	ctacataggt	gccgacgacg	ttccagggtg	tggcggcgct	21120
gccacgtctt	cccaatgaca	aggtggatcg	gttgagcctg	cccgaaccca	aggtggagga	21180
gcagaccgag	gggtacgtgg	cgcctcgcac	ggaaaccgag	aaggtactgg	ccgaaatctg	21240
gagcgcacgt	ctcagccagg	gccgggcgcc	cctgaccctc	ggcgcgacgc	acaacttttt	21300
cgaactggga	ggccattcgc	ttctgcgcgc	ccagatgttc	tcgcggatcc	ggcagaagtt	21360
cgatctcgaa	ctgcccatac	acaccctgtt	cgagaccccc	gtgctggagg	gctttgcgag	21420
cgcctgcgac	gcggctcttg	ccgagcggaa	cggctccggc	cagaggctga	tcagcatgac	21480
ggaccgcggc	caggcgcctc	cgtgttcgca	cgtccaggag	cggctctggt	tcgtgcacga	21540
gcacatgggt	gagcagcgga	gcagctacaa	cgttgccttc	gcctgccaca	tgcgtggcaa	21600
ggggctgtcg	atgccggcgc	tgcgcgcgcg	catcaacggg	ctgggtggctc	gccacgagac	21660
cttgcggacg	acgttcgtcg	tctccgaggg	cggaggagat	cccgtccagc	ggatcgccga	21720
ctccctgttg	atcgagggtc	cgtatataga	ggctgatgcg	tcggaagtcg	cggcccgcac	21780
ggcgggccac	cggggccacg	tgttcgacct	tgcgaagggc	cccctgtctg	agacctcggt	21840
ccgtcgggtg	acgcccgate	accacgtgtt	cttgatgaac	atgcatcaca	tcactctgtg	21900
tgggtgtgtg	atcgacatcc	tgttcgggga	cctctacgag	ttctacaagg	cggccgagac	21960
gggtcgcagc	ccgaacctgc	cggctcctgc	aatccagtat	gccgactact	ccgtgtggca	22020
ggctcagcag	gacctcagca	gtcacctcga	ctactggaag	aagacgctcg	agggctacca	22080
ggaaagggtg	tcgtttccgt	acgacttcgc	ccgcccgtcc	aacaggacct	ggcgtgccgc	22140
gagtgtccgg	caccagttac	cggcggaact	cgcacccgtg	ctgtcggagg	tgagcaagag	22200
ccatcaggcg	acggtgttca	tgaactgtat	ggccagcagc	gcaatcgtgc	tgaaccggtg	22260
cacgggtcgg	gatgtctgtg	gcgtgggtgc	cacggtggcg	ggccgtgacc	acttcgagct	22320
cgagaacctg	attggtttct	tcgtcaacat	cctcgccatc	aggctcgacc	tcagcgggaa	22380
tcccacggcc	gagacggtgc	tgcagcgggc	gcgagcgcag	gtgctggaag	gcatgaagca	22440
tcgcgacctg	ccgttcgagc	acatccttgc	ggcgtctcag	aagcagcgcg	acagcagcca	22500
gattccctcg	gtgccgttga	tgttcgcgca	ccagaacttc	ccgacagtga	cctcgcagga	22560
gcaggggctc	gaccttggta	tcggggagat	cagttttggt	gagcggacga	cgcccaacga	22620
gctcgacatc	cagttcatcg	gcgaggggag	cacgctggag	gtgggtggtc	agtacgcgaa	22680
ggatctgttc	tcgagcgcca	cgatccagcg	gctcatcacg	cacttgcagc	aggtgctgca	22740
gactctcgtg	gacaagccgg	actgcgggct	gacggatttt	ccgctggttg	ccggggacgc	22800
gctcgagggc	gggtgtgtcg	gctccggggg	cgcgacgaag	accggcaagc	tcgacgtgtc	22860
gaagagcccg	gtcgagttgt	tcaacgagcg	ggtagaggcc	tcgcccggac	cggctgcctg	22920
catgggcgcg	gacgggaagc	tgacctaccg	ggagctggac	cgaaggcgca	atcaggtcgc	22980
ccgccacctg	atggggcgag	gggtggggcg	ggagacgcgg	gtgggggtgt	ggttcgagcg	23040
ctcgccggag	ctcgtgtctg	cactcctggg	catactcaag	gcggggggct	gcttcgttcc	23100
gctcgatccg	agctatccgc	aggagtacat	caacaacatc	gtcgccgatg	cgcagccgct	23160
tctgggtgat	tcgagccggg	cgtgggctc	acgctgtca	ctggaggcag	ggcggctggt	23220
gtacctcgat	gacgcgtcgg	cggcgctccac	cgatgcgagc	gatecccgag	tgcgcacga	23280
cccggagcag	ctcatctacg	tcattgtacac	ctccggttcc	accggtctgc	cgaagggggg	23340
gctcgttccc	catcggcaga	tctgaactg	gctgtaccgc	ctgtgggcga	tgggtccctt	23400
cgggcaggac	gaggtgtgtg	cgcagaagac	atccacggcc	ttcgcggtct	cgatgaagga	23460
gctcttcacg	gggtctgtgg	cgggcgtgcc	ccaggtattc	atcgacggca	ccgtggtcaa	23520
ggagcggcg	gccttcgtgc	tcacactgga	gcgatggcgg	gtcacccggc	tgtacacgct	23580
cccgctgcac	ctcgatgcca	tctgttccca	cgtcgacggg	gcggcggagc	gcctcggttc	23640
cctcgcgcat	gtcactctcg	cgggggagcc	gtgcccgttt	gagctgatgg	agaagctgcg	23700
cgagaccctg	ccgtcgtgca	cggcgttggt	caactacggc	tgtaccgagg	tcaacgacat	23760
ctctactcgt	gtcccgaacg	agcagttcca	cagctcgggg	ttcgtgccga	tcggccggcc	23820
catccagttac	accgggcgcg	tgggtctcga	cgacgagctg	cggacgggtc	cgggtgggcat	23880
catggggggag	atttacgtcg	agagcccggg	gacggcgcgg	ggctactgga	ggcagccgga	23940
tttgacggcc	gagcggttca	tcccacaacc	gttcggcgag	ccgggtagcc	gtctctacgc	24000

[0007]

tacgggcgat	atggcgcgat	gccttgagga	tggtctcgtg	gagttcttgg	ggcgccggga	24060
ctacgaggtc	aagatccgtg	gccatccgtg	ggacgtccgc	caggtcgaga	agatccctgc	24120
gagccaccgc	gaagtcctcg	agtcggcggg	gttgggctgg	ccacgggggg	cgaagaaccc	24180
tcagttgctt	gcctacgcgc	ccacgaagcc	gggcgcgtcc	ctgtcgactg	aaaaacgtgcg	24240
ggagttacctg	tcggcccgtt	tgcgcagcta	catgggtgcca	acgtctctacc	agttctctgcc	24300
agcgtgcgcg	cgcctgcccc	atggcaagct	cgaccgtctc	gggctgcccc	atcacaagaa	24360
agtcgaggtg	ggcggcgtct	acgtcgcccc	gcagacgcgc	acggagaagg	tcttggcggg	24420
actgtgggcc	gagtgccctc	agcagggcga	catccccgcg	ccgcaggttg	gccgcttgca	24480
caacttcttc	gacctcggtg	ggcactcgct	gctcgccaat	cgcgtactga	tgcaggtgca	24540
ggggcatttc	gggttcagcc	tgggcacag	tgcgttgctc	ggttctccgg	tgtgaatga	24600
cttcgcggcg	gccatcgaca	aggcgtctcg	gaccgaggag	ccaggcgagg	aaggttcgag	24660
cgacgcacga	gaggtcgctg	cgaaggacac	ctccgtgctc	gtgcgcgtct	ccaccacagg	24720
gacgtgcgcg	agcctgttct	gcgtccatcc	gggtggcggg	caggtccatg	cctaccgcga	24780
gcttgcccag	gggatggaga	agcacgcag	catgtacgcg	ctccagtcgg	aggcgcccg	24840
tgagttcgac	acaatcgaga	ccttggcgcg	cttctacgcc	gatgcgaccc	gcggggtcga	24900
gcccgcaggg	agctaccgtc	tcctcggaatg	gtcttctggt	gggctcatca	ccctggcgat	24960
tgcctcgccg	ctggagcacc	aggcgtcgcc	cgtggagtag	gtgggcctcg	tggattcaaa	25020
gccgaatccg	cgggttcagc	gtgagcgcgc	ctggcgctcg	ctgatcgccg	cgacgaacat	25080
cctggcgccg	atgcgggggc	gcggcttctc	ggtcgcgcg	gtgatgctg	ccgggaagat	25140
cctcgagtcg	cgcggatgga	cggaggagtc	cttcgactcg	gaggggcatg	cggtgttggg	25200
ggagttggct	cgccacttcg	gcataccgtt	cgcgcaagag	tcacgagagt	acctccttgc	25260
cgggttcaag	accacgaagt	actactgttc	gctgttcgct	ggctcaagc	cggcgccgct	25320
cgggcgggag	acgtacctct	atgaggcttc	agagcgggtc	ggagccacct	cgaacgacga	25380
cacgggcgag	tggggggacg	cgttggatcg	caaggccctg	cgggcgaaca	tcgtgcaggt	25440
gcaggaatct	actatactg	tcctgcaggg	agagaacgtg	ctgcaactgg	cggggcggtg	25500
cgcgaagccg	ttgtctcgca	tcgacaactc	gggtggtaacg	aggacgcgag	cttcgtgacc	25560
ctttgcacct	cgggttcgcc	aagaggcaac	aaacgctgat	tcaccggcaa	gggaattccg	25620
tgcagatgga	caatcgagag	atcgcaccca	cccaatcggc	gcgcacgcgt	gatgcgtaca	25680
cggcggtagc	accagccaag	gccgagtatc	cgtcggacgt	ctgtgtgcac	caactgttctg	25740
agtgtcagcg	ggacaggatt	cccagcgcgc	ttgcggcgag	ggcggggaac	gagtccttga	25800
cctaccggga	gctgaacttc	cgggcgaatc	agctcgcccg	gtaccttgtt	gcgaaggcg	25860
tggtcccgcg	aggctcggtg	gccgtgtcta	tgaaccggac	ccctgcgtgt	ctgtgtctac	25920
tgtctcgcat	cataaaggcg	ggcgcggcgt	acgttccggt	ggacgccgga	ttgcccgcca	25980
aacgggttga	ctacattctg	acggacagcg	gcgcgacctg	cgtcctgacc	gacagggaga	26040
cgcgttccct	cctcgacgag	cgcgggtcgc	cttcgacgct	cgtcctgacc	gtggatgac	26100
catccatcta	ttcgggcgag	accagcaacc	tcgggctcgc	tgctgatecc	gagcagcagg	26160
tctactgcac	ctacacctcg	gggttcgacg	gccttcccaa	aggcgtgatg	gtccagcacc	26220
gcgcgctgat	gaactacgtc	tggtgggcga	agaagcagta	cgtcaccgac	gcggtcgaga	26280
gttttgctct	gtactctctg	ttgtcgttcg	acctcacggt	cacctccatc	ttcgttcgc	26340
tgatctccgg	acgtgcacac	gatgtgtacc	cggacctggg	caggagcgtc	cccgtcata	26400
accgggttact	ggaggacaat	aaggctcgatg	tcgtgaagct	cacgccggcc	caccttgcgc	26460
tgctcaggaa	caaggaccta	tcgcaaaacc	ggctgaaagt	gctcactctg	ggaggagagg	26520
acctccagga	ggagacggcg	ggggacgtcc	acaagcggct	ggacggccgg	gcggtgatct	26580
acaacgagta	cgccccacg	gagaccgtcg	tggggtgcat	gattcacccg	tacgaccccg	26640
cgggtggatct	gcacgggtcg	gtgccgattg	gagtggtgat	cgacaacatg	cggatctact	26700
tgctcgacga	ccgtcggcgt	cccgtcaagc	caggagaggt	tggtcgagatt	tacatcgagg	26760
gcagcgtgtg	gtactctggg	tacaaggaca	agcctcaagt	cacggcggac	cacttcatct	26820
ccaatccgtt	cgttgaaggg	gagcgggtgt	acgccagttg	cgacctgggc	cgggtgaatg	26880
agcgcggcgc	gctcgtcttc	ctcgcccgga	aggatttga	gatcaagctg	cgggggtacc	26940
ggatcgagct	gggcgagatc	gagagcgcgc	ttctctctca	tcgggggatc	aaggaaatgca	27000
tcgtcgatct	gaccaagacc	gcgcagagcc	aggccgcgcg	tcagctcacc	tactgcacca	27060
agttgtgctt	ggcgtcgagc	ttcccgaata	cgacgtactc	cgcgcagggg	gtctgcaacc	27120
actgcgagcg	cttcgacaag	taccgcagcg	tcgtcgacga	ctacttcagc	acgatggatg	27180
agctgcagtc	gatcgtcacc	gagatgaaga	gcataccaaa	ctcgaagtac	gactgcacgc	27240
tggcgctcag	cggcggaata	gacagcacgt	atgcactctg	ccggatgac	gaaaccggtg	27300
cccgtgtatt	ggccttcacg	ttggataaac	gctacatctc	ggaggaggcg	aagcagaaca	27360
tcaaccgggt	cgttgcggcg	ctgggagtgg	atcacccgta	ttctctgacc	ggccacatga	27420
aggagatctt	cgtcgacagc	ctgaagcgac	acagcaatgt	gtgcaacggc	tgcttcaaga	27480
ccatctacac	gtttgcgac	aaactggcgc	aggaggtcgg	cgtcaagcac	gtggatcatg	27540
ggttgtcaaa	gggcacaactg	ttcgaaaacg	gcctctcggc	cttgttccgc	acgtcgacct	27600
tcgacaacgc	cgccttcgag	aagagcctcg	tcgacgcgcg	aaagatctac	catcgcatcg	27660
atgatgccgt	gagccgcctg	ctcgacacta	cttcgctcaa	gaacgacaag	gtcatcgaga	27720
acatcaggtt	cgttgacctc	tatcgttatt	gccacgccag	ccgtcaggag	atgtacgact	27780
acatccagga	gagagtcggg	tgggccagcg	cgattgacac	cggcggtcgc	acgaactgtc	27840
tcctcaatga	tgttggcacc	tacgttcaca	acaaggagcg	caggtaccac	aactactccc	27900
tgccctacag	ctgggacgtc	cggatgggcc	acatcagcag	ggaagaggcg	atgagagagc	27960
tcgacgactc	ggccgacatc	gacgtcgaga	gggtcgaggg	catcatcaag	gaccttggct	28020
acagagctgaa	cgaccaggtg	gtgggctcgc	cggaaagccca	gctgggtgcc	tactatgtct	28080
cgcgggagga	gttccccgcg	tcgacgtgc	ggcagttctt	gtcggagatt	ctgcggaggt	28140
acatggttacc	caggtcgttc	gtccagctgg	acagcatccc	gctgacgccc	aatggcaagg	28200
tcaatcgta	ggccctgcgc	aagcctgacc	tgttctggaa	ggccggcacc	gacggacaag	28260

[0008]

ccgcaccccg	aacaccgggtg	gagaagcagt	tggcggagct	gtggaaggag	gtgctgcagg	28320
tcgacagtgt	cgggatccac	gacaacttct	tcgagatggg	cgggcactcg	cttcgggcgc	28380
tcattgtgct	ctacaagatc	gacagtcagt	tcataaagac	gatcagcatc	caggagtctt	28440
cgaaggtecc	caccatcage	gcgtctcgcg	cgcattctcg	cagtgcaccc	gaagcgggtc	28500
cgccagggct	ggcgagggtc	gtcgatcaga	gcgcgcctgc	atacagggga	taacgtgcgc	28560
ttcgtcactg	tcaatggtga	ggactcggca	gtttgctcgg	tgttggtatg	cggactccag	28620
ttcggagatg	gcctgttctga	gacgatgctg	tgtgttgccg	gtgcgccggt	cgacttcccg	28680
gaacactggg	cgcggcttga	tgagggtcgc	cgcggcttgg	gaatcgaaat	cccggacatc	28740
cggcgcgaag	tgaccgctgc	gatcgccagg	tgggggtgctc	ccagggcggt	cgccaagctc	28800
gtcgtcactg	ggggaagcac	ggagcgggga	taccggtcgc	cccccttcgt	ccggccgaac	28860
tggatctca	ccatcacgga	tgccccgaag	tatccgtcgg	cccacgagga	cagaggcgtg	28920
gccgtcaaac	tcigccgaac	gctcgtctcg	ctcgatgacc	cacagctggc	cgggttgaaag	28980
cacctcaaac	ggttgcccca	ggtgctcgcg	aggaggaggt	gggacgacga	gtaccacgat	29040
ggcctgctga	cgaccacgg	tggtcacctc	gtcgagggtt	gcacgagcaa	cctgttcttc	29100
ggtgccgacg	gagccttgag	gacgcccgat	ctgactgcgt	gcggtgtgcg	cggtatcgtg	29160
cggcagaagg	tcctcgacca	ctcgaaaggca	atcgggatcc	gctcgagagt	aaccaccttg	29220
aagctacgag	atctcgaaac	cgcggacgag	gtcttctctga	cgaactctgt	ctacgggatt	29280
gtcctcgttg	ggggcgtcga	tgttatgagg	taccggatag	gtccgacgac	ggcgctttg	29340
ctgaaagacc	tttgccagggt	tgtgtacttt	tgaggctccg	tggaggacgg	tatgaccggt	29400
aatttgata	gcgcggcatg	gcccgtaate	atcacgcctg	gccagcagcc	agcggcgctg	29460
gaggattggg	tcacagcga	ccgtgacgga	ctcagcggcg	agttgaccca	gtgtaagccc	29520
attctctttg	cgagccttcg	tagcaggaa	ggcttcgaga	gcattgccaa	cagcttcttc	29580
gaccggcgcc	tcaactatac	ctatcggtcg	acgccccgta	cggacctggg	gcagaacctc	29640
tacacggcga	cggagtaccc	gaagcagctg	tcgattccgc	agcattgcga	gaacgcctac	29700
cagcgcgact	ggccgatgaa	gctgctgttc	cactgcgtgg	agccggcgag	caaaggcggc	29760
cggagccctt	tggccgacat	gacgaaggta	acggcgatga	tcccccgcca	aatcaaggag	29820
gagttcgcgc	ggaaagaagg	cgggtacgtg	cggaactacc	gtgctggagt	ggatctgcct	29880
tgggaagagg	tgtttggaac	gagcaacaag	gcagagggtg	agaagttctg	cgctcgagaat	29940
ggcatagagt	accactggac	cagagggtgc	ttgaagacca	tccaggtctg	ccaggcgctt	30000
tcttcgcatc	cactcacccg	tgagacgac	tggttcaatc	aggccccact	gtttcacctt	30060
gcccgattgg	accggccttc	acagaagatg	atgctttctc	tcttcgggtg	ggcgggcctc	30120
ccgcgcaact	cgtacttcgg	agacgggtcg	gccatcgggg	gcgacgtcct	cgaccagatc	30180
cgtctccgct	acgaacgcaa	caaggctctg	ttcgagtggc	agaaggacga	cgtgttgctg	30240
atcgacaaca	tgttggttcc	tcacggacga	gatccgttcg	aaggcagccg	gcgggtgctg	30300
gtctgcatg	cggagccgta	ttcggaagtc	cagcggcggg	gattcgccgg	ggcaacgaac	30360
tcaggggcgt	cgttaaggcc	gggtcctgat	gttgtgtcgc	tttcgccgtt	gcgcaaaaca	30420
gtcggaggag	ttctctgtcc	cgaatttcga	tgtctgtgga	gggagagctg	gagggggtacg	30480
aggacgggtt	ggaactcggc	tacgacttcc	cgcggacgtc	gaataggcg	tggagagcgg	30540
cgcagttcca	gctgacatac	ccgccecgag	tggcgaggaa	ggtggcgagg	ctcagccggg	30600
agcagcagtc	cacgtgttcc	atgagcctgg	tggcgagcct	ggcggtgtgt	ttgaaccggt	30660
acacggggcg	cagggacgtg	tgcateggga	cgcgggtggc	gggcccagcg	caggtggggg	30720
cgttggggga	tcctgagcgg	tcacccgtcg	acatctctcc	gctgaggctg	gacctgtcgg	30780
gcgtcccgag	ccttcacgag	gtgctgcgga	ggacgaaggc	ggtggtgctg	gagggtattcg	30840
agcacagggc	gttgccgtgc	cagattccct	tgggtccggt	ggtggtgagg	caccagaact	30900
tcccgatggc	gcgtctggag	ggctggagtg	aggggggtgga	gctgaagaag	ttcgagctgg	30960
cgggggaaag	gacgacggcg	agcgagcagg	actggcagtt	cttcggggac	gggtctctgc	31020
tggagctgag	cctggagtac	gcggcggagc	tgttcagcga	gaagacgggt	aagaggatgg	31080
tggagcacca	ccagcagtg	ctggaggcgc	tggtaggagg	gctggaggag	gtgcggctgc	31140
acgaggtgcg	gctgctgacg	gaggaggagg	aggggctgca	cgggaggttg	aacgacacgg	31200
cgcgagagct	ggaggagcgc	tggagcctgg	cggagacgtt	cgagcgtcag	gtgaggggaga	31260
caccggaggc	ggttgcttgc	gttggcgtgg	aggtggcgac	gggaggggcac	tcgcggccga	31320
cataccggca	gctgacatac	cgcgagctga	atgcgcgagc	caaccaggtg	gcacggaggc	31380
tgagggcact	gggagtgggc	gcggagacac	gggtcgcggt	cttgagcgac	cgctcgccgg	31440
agctgctggt	ggcgatgctg	gcgatattca	aggccggggg	ctgctacgtg	ccggtggacc	31500
cacagtaccc	gggaagctac	atcgagcaga	tactggaggga	tgcggcaccg	caggtggtgt	31560
tgggcaagag	gggaagagcg	gacggggtgc	gggtggatgt	gtggctggag	ctggatggag	31620
cgcaacggct	gacggacgag	gcgttggcgg	cacaggaaga	gggagagctg	gagggggcgg	31680
agaggccgga	gagccagcag	ttggcgtgtt	tgatgtacac	gtcgggctcc	acgggcagac	31740
cgaagggggt	gatgggtccc	tacagccagt	tgcacaactg	gctggaggcg	gggaaggagc	31800
gctcgcctct	cgagcgtggg	gaagtaatgt	tgcagaagac	ggcaatcgcg	ttcgcggtgt	31860
cggtagaagg	gctgctgagc	ggattgtcgg	cgggagtggc	gcaggtgatg	gtgccggaga	31920
cgttggtgaa	ggacagcgtg	gcgctggcgc	aggagataga	gcggtggcgg	gtgacgagaa	31980
tcacactggt	gccatcgca	ctgggagcac	tgttgagggg	ggcgggggaa	gaggcgaagg	32040
ggctgaggtc	gctgaagtac	gtcataacgg	cgggggaggc	actggcgacg	ggggtgaggg	32100
aggaggcgag	gaggaagctg	ccggggcgcg	agttgtgtaa	caactacggg	tgacgggagc	32160
tgaatgacgt	gacttaccac	cccgcgagcg	aggggggagg	ggacacggta	ttcgtgccaa	32220
tcgggcggcc	catcgcgaa	acgcgggtgt	acgtgttgga	cgagcagttg	aggcgggtgc	32280
cggtaggggt	gatgggggag	ttgtatgtgg	acagcgtggg	gatggcgagg	gggtattggg	32340
gccagccagc	gctgacggcg	gagcgttca	tcgcgaaccc	gtacgcgagc	cagccgggag	32400
cagaggttga	cgggacggga	gacatggtga	gggtgctggc	ggacggctcg	ctggagtacc	32460
tggggaggcg	agactacgag	ataaagtgga	gagggcaccg	ggtggacgtg	cgccaggtgg	32520

[0009]

agaagggtgc	gaacgcgc	ccagccatcc	gccaggcggt	ggtgtcggga	tggcgttgg	32580
gctcgagcaa	cgcgcagt	gtggccatcc	tggtgccgca	ggcgggcgcg	acggtggggc	32640
cgcggcaggt	gagggattac	ctggcggagt	cgtgccggcg	gtacatggtg	ccaacgctat	32700
acacggtgtt	ggaggagt	ccgcggctgc	cgaacgggaa	gctggaccgg	ttgtcgctgc	32760
cggagccgga	cctgtcgagc	agccgagagg	agtacgtcgc	gccccacggc	gaggctcgagc	32820
ggaagctggc	ggaaatcttc	ggcaacctcc	tggggctcga	acatgtcggc	gtccacgaca	32880
acttcttcag	cctcggcggg	cactccctcc	tggctgcccc	gctgatttcg	cgcatacggg	32940
cgaccttccc	cgtggaagt	gcgatggcca	cgggtgttcga	gtcccccacg	gtggagcccg	33000
tcgcccgcca	catcgaggag	aagctcaagg	acgagtctcg	ggtccagctc	tccaacgttg	33060
tgcgggtcgc	gcggacgcag	gagattccgc	tctctacct	gcaggagcgg	ctgtggttcg	33120
tgcacgagca	catgaaggag	cagcggacca	gctataacat	cacctggacg	ttgcacttcg	33180
ccggcaaggg	tttctcgggt	gaggcgttgc	ggacggcctt	cgatgagctg	gtggccagac	33240
acgagacact	gcgcacgtgg	ttccaggtgg	gggaggggac	agagcagggc	gtacagggtc	33300
tcggggagcg	ctggctgatg	gagctgcgcg	tgagagaggt	ggcggggacg	gagggtgacgg	33360
cggcaatcaa	tgagatgtcc	cgacaggtct	tcgacttgag	agcgggacgg	ttgctgacgg	33420
cggcggctct	gagggtggcg	gaggatgagc	acatcctcgt	cagcaacatc	caccacatca	33480
tcacggacgg	ctggtcttc	ggggtgatgc	tgcgggagct	gagggagtgt	tacgaggcag	33540
gccgtgcggg	gcagtcacag	gagctgcgcg	cgtgacgggt	gcagtacggc	gactatgcgg	33600
tgtggcagag	gaagcaggac	ctgagcgcgc	acctggcgta	ctggaagggg	aaggtggagg	33660
agtacgagga	cgggttggag	ctgccttacg	acttcccgcg	gacgtcgaat	agggcgtgga	33720
gagcggcgac	gttccagtat	agctaccac	ccgagctggc	gaggaaggtg	gcggagctca	33780
gccggggagca	gcagtcacag	ctgttcatga	gcctgggtggc	gagcctggcg	gtggtgttga	33840
accgggtacac	gggcggccag	gacgtgtgca	tcgggacgac	ggtggcgggc	cgagcgcagg	33900
tggagctgga	gagcctcatc	gggttcttca	tcaacatcct	cccgcctagg	ctggacctgt	33960
cgggcgcctc	gagccttcc	gagggtgcgc	ggaggacgaa	ggcgggtggtg	ctggagggat	34020
tcgagcagca	ggagtgtccg	ttcgagcacc	tgtgaaggcg	gctgaggcgg	cagcgggaca	34080
gcagccagat	tccttgggtg	ccagtgggtg	tgaggcacca	gaacttcccg	atggcgcgtc	34140
tggagggtctg	gagttagggg	gtggagctga	agaagtctga	gctggcgggg	gaaaggacga	34200
cggcgagcga	gcaggactgg	cagttcttcg	gggacgggtc	ctcgctggag	ctgagcctgg	34260
agtagcggcg	ggagctgttc	agcgagaaga	cggtagggag	gatgggtggg	caccaccagc	34320
gagtgctggg	ggcgtgtgtg	gaggggctgg	aggaggggct	gcacgaggtg	cggctgttga	34380
cggaggagga	ggaggggctg	cacgggaggt	tgaacgacac	ggcgcgagag	ctggaggagc	34440
gctggagcct	ggcgggagac	ttcgagcgtc	aggtgaggga	gacaccggag	gcggtggctt	34500
gcgttggcgt	ggaggtggcg	acgggagggc	actcgcggcc	gacataccgg	cagctgacat	34560
accggagcgt	gaatgcgcga	gccaacagg	tgccacggag	gctgaggcca	ctgggagtgg	34620
gcgcggagac	acgggtcgcg	gtcttgagcg	accgctcgcc	ggagctgctg	gtggcgatgc	34680
tggcgatatt	caaggccggg	ggctgctacg	tgcgggtgga	cccacagtac	ccgggacact	34740
acatcgagca	gatattggag	gatgcggcac	cgcaggtggt	gttgggcaag	aggggaagag	34800
cggagctggg	gcgggtggat	gtgtggttgg	agctggatgg	agcgcaacgg	ctgacggacg	34860
aggcgttgge	ggcacaggaa	gagggggagc	tggagggggc	ggagaggccg	gagagccagc	34920
agttggcgtg	tttgatgtac	acgtcgggct	ccacgggcag	gccgaagggg	gtgatgggtc	34980
cgtacagcca	gttgcaaac	tggtcggagg	cggggaaagg	gcgcgcggcg	ctcgagcgtg	35040
gggaagtta	gtttcagaag	acggcaatcg	cgttcgcggg	gtcggtgaag	gagctgttga	35100
gaggattgct	ggcgggagtg	gcgcaggtga	tggtgccgga	gacgctggtg	aaggacagcg	35160
tggcgttgge	gcaggagata	gagcgggtgc	gggtgacgag	aatccacctg	gtgccatcgc	35220
acctgggagc	actgctggag	ggggcggggg	aagaggcgaa	ggggctgagg	tcgctgaagt	35280
acgtcataac	ggcgggggag	gcactggcgc	agggggtgag	ggaggaggcg	aggaggaagc	35340
tgcggggggc	gcagttgtgg	aacaactacg	ggtgcacgga	gctgaatgac	gtgacgtacc	35400
acccgcgcag	cgagggggga	ggggacacgg	tattcgtgcc	aatcgggcgg	cccacgcgca	35460
acacgcgggt	gtacgtgttg	gacgagcagt	tgaggcgggt	gccggtgggg	gtgatggggg	35520
agttgtatgt	ggacagcgtg	gggatggcga	gggggtattg	ggggcagcca	gcgctgacgg	35580
cggagcgtt	cctgcgaac	ccgtacgcga	gccagcccg	agcgaggttg	taccggacgg	35640
gagacatggt	gagggtgctg	gcggacggct	cgttgagta	cctggggagg	cgagactacg	35700
agataaaggt	gagagggcac	cgggtggacg	tgcgccaggt	ggagaagggtg	gcgaacgcgc	35760
atccagccat	ccgccaggcg	gtggtgtcgg	gatggccgtt	gggctcgagc	aacgcgcagt	35820
tgggtggccta	cctgggtccc	caggcggcg	cgacgggtgg	gccgcggcag	gtgagggtt	35880
acctggcgga	gtcgtgccca	gcgtacatgg	tgccaacgct	atacacggtg	ttggaggagt	35940
tgcgcgggtt	gccgaacggg	aagctggacc	ggtgtcgtt	gccggagccg	gacctgtcga	36000
gcagccgagc	ggagtacgtc	gcgccccacg	gcgaggtcga	gcggaagctg	gcggaaatct	36060
tcggcaacct	cctggggctc	gaacatgtcg	gcgtccacga	caacttcttc	agcctcggcg	36120
ggcactccct	cctggtgtcc	caggtggtct	caaggattgg	caaggagctt	ggcactcaga	36180
tctcgatcgc	cgtctgttt	caaaggccca	cgattgaaca	gctctgtgag	ctgattggag	36240
gactggacga	tcagaccag	agggagctcg	ccctcgctcc	gtcggggaac	accgagcgcg	36300
tgtctcgtt	cgcgcaagag	cgcattgtgt	tcctgcacaa	cttcgtcaag	ggcatgcct	36360
acaacacgcc	agggctgcac	cacctgacgg	gtgagctcga	tgtcggcgcg	ctagaaaagg	36420
ccatccgcgc	ggtcatccgt	cgccacgagc	ccctgcggac	gaatttcgtc	gagaaggacg	36480
gggtgctgtc	ccagttgggt	gggacggga	aacgcttccg	cctgaccgtg	actcccatcc	36540
gcgacgagag	cgaggtcgcg	cggctcatgg	aagccgtgat	ccaacgccca	gtcgatctgt	36600
agcgggaggt	gatgatccgg	gcttatctct	accgggtcga	cccgcggaat	cactacctgt	36660
tcaccacat	ccatcacatc	gccttcgatg	gctggctgac	atcgatcttc	taccgtgagc	36720
tggtctcgta	ctacgcgcg	tttctcggcg	gcgaagacag	tccgctgccc	gcgctggaaa	36780



[0010]

tctectatca	ggactatgcc	cgetgggagc	gggcccattt	ccaggacgag	gtgttgccgg	36840
aaaaactgag	gtactggcgg	cagcgctgt	cgggcgctcg	gcccctcgta	cttccgacca	36900
cttaccatcg	gccgcccatc	cagagtttcg	ctggcgccgt	cgtgaacttc	gagatcgatc	36960
gtctcatcac	cgagcggttg	aagacgctgt	tcgccgagtc	gggcaccacg	atgtacatgg	37020
tgttgctcgg	cgcgtttccc	gtggtgctgc	agcgtacttc	cggtcaggac	gacatcttca	37080
tcggctcccc	cgtggcgaa	cggggtcaca	tcagacaga	agggtcgatc	ggcttgttcg	37140
tcacacacct	ggtgatgagg	gtggatgccg	ccgggaatcc	ccgtttcatc	gacctgctgg	37200
cgcgcattca	acggacagcc	atcgatgctt	acgcgaacca	agaagtgcgc	ttcgagaaga	37260
tcgtggacga	ctgcagggtc	gcgagagaca	cggccccgac	tcgcctcggt	caggtcattc	37320
tcacattcca	caacacgcct	cctcaatccg	agctggaa	gcagggggtg	accctcacgc	37380
ggatgccggt	gcacaacggc	acggccaagt	tcgagctctc	catcgacgtc	gcggagacga	37440
gcgcgggtct	aacgggattc	gtggagtacg	cgcgggatct	gttcagcgag	aacttcattc	37500
ggcggtatgat	cggccacctc	gaggtgtgtc	tggacgcggt	cggtcgcgat	ccgcgggcgc	37560
ctatcccatg	gttgccactg	ctcaccgcgc	aggatcagtt	ggacctactg	tcgcggagcg	37620
gccacacage	ccccgcgggt	gaacacgtcg	agttgatccc	tcatacgttc	gagcggcgcg	37680
tcaggagag	ccctcaagcg	attgccctgg	tctgcggtga	cgcgcgcgtc	accctactccg	37740
cgcctcaaccg	cggggccagc	cagattgccc	gccgcctcgc	cgcgcagagg	atcggaaccgg	37800
acacctctgt	tggcctttgc	gcggggcgct	ccatcgagct	ggtctgcggc	gtccttggca	37860
tcttgaagge	gggcggtg	tacgtgcaa	tcgacccac	ctctcgcgc	gaggtgatct	37920
acgacgtcct	gtatgagtcg	aaggtgcggc	atctgttgac	cgcgcgcgc	ctggtcgggg	37980
gactgcgggt	cgatgaccag	gaaatctcgc	tcctggatac	ccccgcggac	ggtgaagggg	38040
acaagcttgc	tggtagccgg	gaggagccac	ctgaccttgg	cgcgggtcgc	ctcactcccg	38100
agtgttggc	gtacgtcaac	ttcacctccg	actccggtgg	ggcgccgagg	ggcatcgccg	38160
tcgcctatgg	ggcgttggt	cgcgggatgg	cgcgcggcca	cgcacagtac	ctggccaatt	38220
ccgcgctacg	tttctcgtg	aaggcgccgc	tcacgttcga	cctggcggtc	gcggagctgt	38280
tccttgatgg	cgtcagcggc	ggcagcctga	gcacccctga	ccccaatgcc	gaccgcgacg	38340
ccctctgctt	cctcgcgcag	gtgcgcaggg	actcgattgg	cgtcctctac	tgcgtccct	38400
ccgaactctc	gacgtggtg	agccacctgg	agcgcgagcg	tgaaggggtg	catgagctga	38460
acacccctccg	gttcattctc	tgcggcgggg	ataccctggc	ggttacgcgc	gtcgagcgtc	38520
tcggggtact	ggtgcggggc	ggccagctcc	cgtgcggct	ggtcaacgtc	tatgggacga	38580
aggagacggg	aatcgcgcg	gggtgtctcg	agtgcgcgct	ggacgcgaac	gacccagcg	38640
ccgaactccc	gccgggacgg	ctctcgcag	agcggatgcc	catcgccggg	ccccccaga	38700
acctgtggtt	ctatgtggtg	caacccaacg	gtggcctggc	tcctgtgggc	atcccggggg	38760
aactgtactg	cggcggcgcg	caactcgccg	acgcccgttt	cggcgacgag	cccacggcga	38820
cccaccccg	cttctcccg	aacccctccc	ggagcggagc	ggagaaggac	tggctgtaca	38880
agacggggga	cctcgtccgc	tggctgcctc	agggggccgt	cgcgcgtggt	agcgcggctc	38940
gggagcgcga	cggagggcgg	gaccacccgc	tcgatccggg	cttcacgcag	gcgcgcagtc	39000
gtcgtgtggt	cattgtccgc	gacgcctggg	tggcctacgt	cccggatcgc	caggacaggg	39060
cccgggttgg	ggcctacgtc	gttctgaagg	agtgcgccgc	ggcgagacgt	gagccgcgcg	39120
aaggggcgga	aacgctgaag	gtcggatca	gcgcgaact	tgggagcag	ttgcggagt	39180
acatgcttcc	ggcgcctac	gtgttcattg	acagcctgcc	gttgacggct	tacgggagga	39240
tcgaccggaa	agccctgccc	gagccggagg	atgaccgcca	cgggtgtagt	gcgatgcct	39300
actgtggccc	cgcgggcgcc	acggagaagg	cactggcgca	catttggcag	caagtgtga	39360
aacgccccca	ggtcggactg	cgcagaca	tttttgagct	gggcgggcac	tcagtggcgg	39420
ccatccaa	ggtgtccgtg	agccggaagc	acctggaggt	cgaagtcccc	ctcagcctga	39480
ttctcgaatc	gccggctcgt	gagccgatgg	cgcgcggcat	cgaagcgcgt	caacagcagg	39540
gccgcagcga	cgcgggtgct	tcgatccatc	gggtggagcg	gaccggaccg	ctgcctctgg	39600
cgtacgttga	ggagaggctg	tggttcgtgc	acgagcacat	gaaggagcag	cggaccagct	39660
ataacatcac	ctggacgttg	cacttcgccg	gcaagggttt	ctcggtggag	gcgttgcgga	39720
cggccttcga	tgcgtggtg	gccagacag	agacactcgc	cacttgggtc	caggttgggg	39780
aggggacaga	gcaggccgta	caggtcatcg	gggagccctg	gtcgatggag	ctgccgtga	39840
gagagggtgc	gggacgggag	gtgacggcgg	caatcaatga	gatgtcccg	caggtcttcg	39900
acttgagagc	gggacgggtg	ctgacggcgg	cggctctgag	ggtggcggag	gatgagcaca	39960
tctctgtcag	caacatccac	cacatcatca	cggacggctg	gtcgttcggg	gtgatgctgc	40020
gggagctgag	ggagtgttac	gagccgcggg	tgcgggggga	gcgagcggag	ctgccgcgcc	40080
tgacgggtga	gtacggcgac	tatgcggtat	ggcagaggaa	gcaggacctg	agcgagcacc	40140
tggcgtactg	gaaggggaag	gtggaggggg	acgaggacgg	gttggagctg	ccgtacgact	40200
tcgcgcggac	gtcgaatagg	gcgtggagag	cggcgacgtt	ccagtatagc	taccaccccg	40260
agctggcgag	gaaggtggcg	gagctcagcc	gggagcagca	gtccacgcgt	ttcatgagcc	40320
tgttggcgag	cctggcggtg	gtgttgaacc	ggtacacggg	cgcgaggac	ctgtgcctcg	40380
ggacgacggt	ggcgggcga	gcgcaggtgg	aactggagag	cctcagcggg	ttcttcata	40440
acatcctccc	gctgagctg	gacctgtcgg	gcgctccgag	ccttcacgag	gtgctgcgga	40500
ggacgaaggt	ggtggtgctg	gagggatctg	agcaccagga	gttgccgttc	gagcacctgc	40560
tgaaggcctg	gagcgccgag	cgggacagca	gccagattcc	ccttggtcca	gtggtggtga	40620
ggcaccagaa	cttcccgatg	gcgcgtctgg	aggcgtggag	tgagggggtg	gagctgaaga	40680
agttcgagct	ggcgggggaa	aggacgacgg	cgcgcgagca	ggactggcag	ttcttcgggg	40740
acgggtcctc	gctggagctg	agcctggagt	acgcggcgga	gctgttcagc	gagaagacgg	40800
tgaggaggat	ggtggagcac	caccaacgag	tgcgtggagg	gctggtggag	gggctggagg	40860
agggcgctga	cgaagtgcgg	ctgctgacgg	aggaggagga	ggggctgcac	gggaggttga	40920
acgacacggc	gcgagagctg	gaggagcgt	ggagcctggc	ggagacgttc	gagcgtcagg	40980
tgaggggagac	accggaggcg	gtggcttgcg	ttggcgtgga	ggtggcgacg	ggagggcact	41040

[0011]

cgcggccgac	ataccggcag	ctgacatacc	ggcagctgaa	tgcgcgagcc	aaccagggtgg	41100
cacggaggct	gagggcactg	ggagtgggcg	cggagacacg	ggtcgcggtc	ttgagcgacc	41160
gctcgccgga	gctgctgggt	gcgatgtctg	cgatattcaa	ggccgggggc	tgctacgtgc	41220
cgggtggacc	acagtaccgc	ggaagctaca	tcgagcagat	actggaggat	gcggcaccgc	41280
agggtgggtt	gggcaagagg	ggaagagcgg	acggggtcgc	ggtaggatgt	tggtcggagc	41340
tgtagtgagc	gcaacggctg	acggacgagg	cgtlggcggc	acaggaagag	ggagagctgg	41400
agggggcgga	gaggccggag	agccagcagt	tggtgtgttt	gatgtacacg	tcgggctcca	41460
cgggcagacc	gaaggggggt	atggtgcctg	acagccagtt	gcacaactgg	ctggaggcgg	41520
ggaaggagcg	ctcgccgctc	gagcgtgggg	aagtaatgtt	gcagaagacg	gcaatcgctg	41580
tcgctgggtc	ggtagaggag	ctgctgagcg	gattgtctgg	gggagtggcg	caggtgatgg	41640
tgccggagac	gctgggtgaag	gacagcgtgg	cgtggcgca	ggagatagag	cgggtggcggg	41700
tgacgagaat	ccacctgggt	ccatcgaccc	tgggagcact	gctggagggg	gcgggggaag	41760
aggcgaagg	gctgaggtcg	ctgaagtacg	tcataacggc	gggggaggca	ctggcgcagg	41820
gggtgaggga	ggagggcagg	agggaagctg	cggggcgcca	gttgtggaac	aactacgggt	41880
gcacggagct	gaatgacgtg	acgtaccacc	ccgcgagcga	ggggggaggg	gacacggtat	41940
tcgtgccaat	cgggcggccc	atcggaaca	cgcgggtgta	cgtgttggac	gagcagttga	42000
ggcgggtgce	ggtgggggtg	atgggggagt	tgtatgtgga	cagcgtgggg	atggcgaggg	42060
ggtattgggg	ccagccagcg	ctgacggcgg	agcgttccat	cgcgaaccg	tacgcgagcc	42120
agcccgagc	gaggttgtac	cggacgggag	acatggtgag	ggtgctggcg	gacggctcgc	42180
tgtagtacct	ggggaggcga	gactacgaga	taaagggtgag	agggcaccgg	gtggacgtgc	42240
gccaggttga	gaaggtggcg	aacgcgcac	cagccatccg	ccaggcggtg	gtgtcgggat	42300
ggcgttgggg	ctcgacaac	gcgcagttgg	tggtctacct	ggtgcccgag	gcgggcgcga	42360
cgggtggggc	gcggcaggtg	agggattacc	tggtggagtc	gctgccagcg	tacatggtgc	42420
caacgctata	cacggtgttg	gaggagtgtg	cgcggttgcc	gaacggggaag	ctggaccggc	42480
tgtcgttgcc	ggagccggac	ctgtcgagca	gccgagagga	gtacgtcgcg	ccccacggcg	42540
agtcgtgagc	gaagctggcg	gaaatcttcg	gcaacctctc	ggggctcgaa	catgtcggcg	42600
tcacagacaa	cttcttcaac	ctcggcgggc	actccctcct	ggttcccag	ctgatttcgc	42660
gcatacgggc	gaccttcgcg	gtggaagtgg	cgatggccac	ggtgttcgag	teccccacgg	42720
tgtagccgct	cgcgcggcac	atcgaggaga	agctcaagga	cagatctcgg	gtccagctct	42780
ccaacgttgt	gccggtcgag	cggacgcagg	agcttccgct	ctcctacctg	caggagaggc	42840
tgtgtgtcgt	gcacgagcac	atgaaggagc	agcggaccag	ctataacgga	acgatcgggc	42900
tcgggttcg	gggtcctctg	tcaatccccg	cgtcaggggc	caccttcac	gatctggctg	42960
cccgtcacga	gagcctgcgc	accgtcttcc	gggtccccga	aggccgcacc	acgcgggtgc	43020
aggtgattct	tgattcgatg	gatctggaca	tcccgggtccg	cgatgcaacc	gaggccgaca	43080
tcattccggg	ctcgatgagc	ctggcgggtg	acatctacga	catggagaag	ggtccgctgt	43140
tcattggttc	ctctttgcgg	ctggccgagg	actcccacgt	tctcctgatg	gggatgcac	43200
acatcgctta	cgacgcattg	tcacagtcca	atgtgatgag	tcgcgatata	aacctgtctt	43260
actcggcgca	cgtgacggga	atcgaggcac	ggcttcccgc	gcttcccac	cagtacggcg	43320
acttctcggt	gtggcagcgc	cagcaggact	tccgtcacca	cctggactac	tgggaagtcca	43380
caactggcgca	ctaccgggat	gatctcgagc	tgcggtatga	ctaccgcggg	ccgcccagcc	43440
ggacatggca	cgcgaccgga	ttcaccttcc	ggatcccgga	tgcactggcg	cgcgcgttcg	43500
ccaggttcaa	tcagteccat	cagtcgacgc	tgttcattggg	gctgctgacc	agcttcgcga	43560
tcgttcgctg	gcctacacac	ggccgggaac	acatctgcat	cggaaacgaca	acggcggggc	43620
gcgccaggtt	ggagttaggag	aaacctgttg	gcttcttcac	caacatcctg	ccgttgcgca	43680
tcattctggc	gggtgacccc	gacatcagcg	agctcatgaa	tcgagcggaag	aagagcgtct	43740
tggggcgctt	cgagcatcaa	gctctgcctg	tcgagcgtct	cctcagtgcc	ctcaacaaac	43800
agcgtgacag	cagccatate	ccgttggttc	ccgtcatgtt	gcgccaccag	aaacttccga	43860
cggcgatgac	cggcaagtgg	gccgatgggtg	tggacatgga	ggatcatcgag	cgcgacgagc	43920
gcacgacgcc	caacgagctg	gacctccagt	tctttggcga	cgacacctac	ttgcatgctg	43980
tcgtcgagtt	ccccgcgcag	ctcttctccg	aggtgaccgt	ccggcgctctg	atgcagcgtc	44040
accagaaggt	catagagtgc	atgtgcgcga	cgtggggggc	tcggtgaacg	tgtctcgtag	44100
gcattccacc	ggctcccacg	acgagccggt	ggccggcgac	gtcgaaactc	gcgtcgggtg	44160
ccccggtgtg	cgggacgctc	attccagcga	gagcgttgaa	gtgctggcgc	gggtggtcgc	44220
gaccgccgag	gagaagtacc	cgggcgtcat	gggcccgatc	cgcacaggag	gcccctggtt	44280
cgccatcccg	ttgacctgcc	cgcgcgggtg	ccggtcggcg	cgatccggcc	tctggtctcg	44340
ggaaactagac	cgtcagggac	agctcctcca	catggtcgcc	tcgtatctgg	cggccgtgca	44400
ccacgtgctg	gtcagcgttc	gcgagcccag	cgccaaactg	ctggaggtgc	tggctctctga	44460
ctcaacaacg	ccatctgggc	tcaaccgggt	cctgaacggc	ctggactccg	tcctggagat	44520
cctggtctac	gggcgcagcg	acctctcctc	gcagcatctc	acgggcggcg	tgcctcccca	44580
cgagatgccc	ttcgtggagg	accgtgagga	gcgcgaggag	caccgggcca	ccgatgtcga	44640
ggccgatgag	gttgtctccg	tctgttcca	accagttgac	tccccagacc	tggcgagget	44700
ggacgcgagc	ctctctcgct	atgacgacga	ggatgccggc	gcggtggggc	gggtcctggg	44760
ggagctcctc	cagccgttcc	tgtctgactc	cgcagagatg	accgtggggc	gaaaggcggt	44820
gagggctcat	caacatcgcc	tgcctgctt	gttgcgagcc	gacagcagag	cggcgaggga	44880
gtcgggttct	gcgcccgcct	tgcgcttggc	gacgaagccc	ggtcggcatt	tcgtcgcgtt	44940
gtcccggaac	accgcccctg	ggctggggaga	caggtcgccc	cacttgctcg	cgcaggggccc	45000
gctctcgcat	ggcgcgtcaa	cggcgctcct	tctgttgcaa	cgggtgctgg	acacgcttat	45060
cgggagcggg	ggactgaagg	accatcgctc	cagctcagag	ctggttggcg	ccgatccacg	45120
gaccgaggcc	gcgtttcggg	cccggactcc	gtggctgggtg	gcggaaacggg	ccgcttcggc	45180
tgcatacaac	gatgcaccgc	gcgtcgacgt	cgtcgtcctg	ttcccggcgg	cacggccgag	45240
cgcgctcgag	ctgcggccag	acagcgtcgt	catcgacctt	tttggcacct	ggagcctgag	45300

[0012]

```

accgcgaccc gaggttctgg cgaagaacat cgtctacgtg cgaggggcct cggctccgtct 45360
cgccggagag gccgtcgtct cgaactccctc ctctcgcccg gatcgagtgg agccggcgct 45420
cctcgaggcg ctctccgggg aactcgacgc ggagggcagt agtgacgggc tcgccacaga 45480
gcaccgcctt gagattggcg gcattcgccg gtcttggggg gagatccgcc gggcggagtg 45540
ggacgccttt cattcgccgc gccgggggga gctggcgagg tttaagggtg cggggcaggt 45600
gaccgcgcgc aatccggggc tcgccagcct gcccgatggg gcgacgaaca tctgcgaata 45660
catcttcgag gaagcgaccc ttccgtcccg ctctgctcctc gtcgatcccc agagcgacca 45720
gtccgcgacc tacgcgagc tcggcgccact ggccgcagcg tacgcgcggc ggtttcgggc 45780
attggggctc cgccaggag agctcgtggc gctcgccggc ccggtaggga ttctgtccgt 45840
cgcgggatg ctgggttgc tccgtggcgg gtgggtcttc gcgcgcctca accacaccgc 45900
ctcgcccggt aacttcgagg cgtatgttag ttccgccagt ccccgccctg tgctccatgc 45960
cgcgtcgacg gtgcgccgc atctgccggt cctgagcagc cggcgatgcg cggaactcgc 46020
gtccttccgt ccgccggacg cgtggacgg cgtggagggg gacgtacccc cctgccagt 46080
gtcacccgaa gcccccgcg tcatgtgtt cacctcgggc tccacggggg ggccgaaggc 46140
agtacgcacc acccaaccgc acttcatac ctgcagtgc aactacgcac cctatgtcgt 46200
cgaactcaga ccggacgac gtgtctatac gccgtccccg acctctctcg cctatggatt 46260
gaacaacttg ctgctgtccc tcagcgcggg ggccacgcac gtgatctcgg tccctcgcaa 46320
gtcacggatg ggtggttgc agatcctcgc gcggaacgaa gtaaccgtgc tcttcgggt 46380
tcccgccgct tataagctga tcatctcgaa gaacgaccgg ggccctgcgt tgccgaagt 46440
gagattgtgc atctctcgt gcgagaagct gccattgaag ctgtatcggg aggcgcgaag 46500
cttcttcage gtgaacgtac tggacgggat cgggtgcacc gaagccatct cgacgttcat 46560
ctcgaccggg ggtggttatg tcgcgcccg gtgcacgggc gtggtgtgct cggggttcga 46620
ggtaacgtg gtgaaccgc gtggcgagct ctgccgggtg ggagaggtgg gcgtcctctg 46680
ggttcggggg ggggcgctga cccggggcta cgtgaacccc cccgatctga cagagaagca 46740
cttcgtggac ggtgtgttca acaccagga catgttcttc atggatgcgg agtaccggt 46800
ctcacacgtg ggcaaggctg gttcgttcat caagatcaat tctgctggt tctcaccgga 46860
gatgatggag tcggtctcgc aatcccatcc agcgggtgaag gagtgtgcg tctgcgtcgt 46920
cattgacgac tacgggttgc caagcccgaa ggcatctatc gtcaccggcg agcatgagcg 46980
ctccgagccg gacgtcgagc acttgtgggc cgagtgcgc gttctgtcga aagagaagct 47040
tgggaaggac cactaccgc atctgttgc gaccataaaa acgcttcccc ggacctccag 47100
cggaaggctg atgcggtcgc aactcgcgaa gctgctcacc agegggcccc catgaatcca 47160
aagtctctcg gaggcctggg ggcaggggtg tgcctgcct ctttgttcca gacggctatg 47220
cggaccgtgc cgtcaagga cgccggctcc ggccagaggg cttgttagac ttgttgccaa 47280
tgtcagctcg caccagaac ttcaatgtca tgggaatcga ctggatgct tctccgcgg 47340
agttcaagg ctccgtccc cgaccacgag ggccggcgaga ggccgtgctc gccggacgga 47400
gatgcttgat ggatatcctg gaccgcgggg atctcgcct ctctgtcatc gtggggccct 47460
gttcatttca cgtaccggtg gcggggctgg actatgcgaa gcgctgcgg aaactcgtg 47520
atgaggttgc cgagaccctg ttctgttgta tgcgcgtgta ctccgaaaag ccgcgcacca 47580
ccacggttgc gaaaggcttc atcaatgacc cgcgcattgga tggctcttcc cacatcgagg 47640
aggcattgga gcggggacgt cgttctcgc tcgacgtggc cgaggagggt ctaccgcgtg 47700
ccaccgagge gctggacccc atcgcgtcgc agtactacgg cgacctcatt tcttgagcgg 47760
ccattggcgc gcgcaccgce gactgcgaga cgcaccgcga gatggcgtec ggcccttcca 47820
cccagtagg cttaagaac ggccacggac gctcgttggg tgcggccgctc aatggcatca 47880
tctccgttcc acaccgcac agcttcttgg ggttgagcga aaatggcgcg tgcgccatca 47940
tcgcacgcgc cgccaacacc tacggccacc tgggtgctcg cgccgggtgt ggccggccca 48000
actacgacgc cgtgtcgggt gcgcttgcgg agaaggcgt tgcgaaggcc aggtaccaca 48060
ccacatcgtt ggtggactgc tctcacgcca actcctggaa gaatcccgag ctccagccgc 48120
tgggtgatgc ggacgtgggt caccagatcc gcgagggcaa ccgctcggtg gtgggcctga 48180
tgatcgagag ctcatcgag gcaggcaacc agcccatccc ggccggacct tcgcaactgc 48240
gtacagcgtg ctcggtcact gatcgatgtg tggactggaa gaccaccgag aagatgctgt 48300
acagcgcgca cgaggagctg ctccacatcc tgcctccgtg caaggtggtg tgaccgccga 48360
gggttgaggt gttggttctt cccagcaggg gttccccggc caggtggcgg cgccgcacgg 48420
cctggtacac gcagcggcgt ttagctttac ggagagctcg ggcccgccgac tgggtgctg 48480
gcgcctgatt caaaggctga tgcgcagacc caccgccgccc tggatggtag gtggagcgac 48540
ggcgatggga ggctgcacct gctgcacct gcgcagggcc ggccaggttga gcgcgaagcg 48600
gaactcgcca ccccgctggc catagttggc ggcgatgaag gcctcgatgc tgagatagcg 48660
cagggccegg tgcgtcacgt ccaaccgtgt gatgaaagaa cggtcagaga ggttgcaccg 48720
gttggacagg atgaagtgg tgtgttccca ggatcccgga ccggacagga acgctagac 48780
ggcgcgctag tgcggccgga ggtagaagg ctgatactgg ccttgaggga tgaggtaggc 48840
gtaggccagc gaacgggat agcccatcga attgtagaag tactcgacgc ccacggtggc 48900
ggtgtcgtc tccgagtagg cgaacgtcca ggtcgcgcgc ccgctacact gcggcgtgta 48960
accctcgggg tagtacgct ctatggggag cgcgccagg tcgggaggca tgccgccatt 49020
gccttggaac tgaccgagca ggtctccgag ggagacacct tggggcatgc ggaacatggg 49080
cgcatccgag ccttcttga gggcgagttc gccgtagatg tcgatggggc cgagcccgga 49140
ggagaggtcg agcccgaagc gggccttgcg gccgtgttgg agcacggcat cgacgccgag 49200
ttcgataggg ccgagcacca cctcgccgcg agcagcgccc ccgacgcggc cgagcgtatt 49260
ggccggggccg gcgttgtcga gcaggccgag gacgtagaag ttccagcctt tcgctccca 49320
ggcactgtgc atcttagca tggtcgcgcc ggtgcgcgtg tccaagaggc cgagcggatc 49380
cctgcgctgg ggcgagagga agtcggtggg gttccagaag cgcgaggtgc cccacttcac 49440
gtgctgcttg ccgacggtga tgaagagctt gtgttccagg tcgaagcgca gccaggcctg 49500
atccaacagc acgaccgat ccgcagcgac gttggaggtg gacgtgctcg tggggacgat 49560

```

[0013]

gccgagggag	cccgccttgc	gggtcggatc	gaaggtgagc	cgteccagca	cgaagccgcg	49620
cagccgctcg	gtggggcggg	catcgaagta	gccgtccacc	agcatggggg	cggagaaggt	49680
ggtgtttctg	aaggacaccc	cttcgtttggc	ctgtgagtag	gcgcgcaggt	agaagcggcc	49740
gecgatcttc	agcggatcct	cgaaggccctc	ctcgggtgctg	aaggcgttgg	tggccgacgg	49800
gccaccgagc	gcctgcgcac	cccggctctg	gggcgtggcg	gagggccttg	cggggggccg	49860
ggtggccgcg	gcgcctctgt	cggccggggg	ggacgcgggg	gtgtcaccga	agagggaact	49920
ctcgtcgggg	cggggcgcat	cggccggagc	gggcttcgtc	tctggagtgt	cggccgccaa	49980
gaggtcgccc	tcgctgggac	gtccttgggc	gagcgcgggc	agcgcggcga	gggacgcggc	50040
cagggccagg	gaggtgcgcg	tgtctatcgg	cttttgcctc	cgaaccaggc	cttgggtgaag	50100
atgttctcct	cgagcgagcg	caggtccacg	ctcttcacga	cgatgacggg	ggagtgggtc	50160
ttctccacct	cgtcatagaa	gcgcattctc	tgcgggtacc	agacgtcggc	cttcttggac	50220
tcgttgaaga	gtttcatcca	cttgggggaag	taggaggtgc	gcattcaggcg	gccggaaaagg	50280
gcgaactcct	ggcgcttgag	gatgtttgtc	gtgtccttct	ccaccacag	gtgtaccacg	50340
gggttagccga	gttccacgtt	cggcttggcg	gtgaggacga	gttccagggt	ggtagaacttg	50400
ccgagtttct	ctctgccctc	gaacttgcca	tcgagctcct	cggccaggcg	cgaactcgtc	50460
aagtccggcg	ggcggtctgc	ggtccggcgg	atacgtctac	gtcgggtgcg	ccggtccccac	50520
ttgccggtgt	tcgggtcgta	gtccagagg	ttcttgtcca	gccgcaggta	gcccctggcc	50580
agcttgcctc	tggcttggc	catgaggatc	atcagctgat	ccttctcgtc	gcgcggtag	50640
acgacggcct	cgcgcacgac	gtctgttttg	tccttctcct	tctgtctgat	atacaccagc	50700
gactttagtg	cgcgcgcgtt	gcgctggcgg	ttgtcagcgc	tctccaggag	cttcttgcac	50760
tcggcggggg	cgggtgaggtc	cgcgcgagcg	gtcggagcgg	ccagcagcag	cgcggcgaaac	50820
agggcgcccg	ggaggttggc	cagggctcatg	gtcgtcacc	gatgtgggtc	atcgccgtga	50880
tgggcttcat	ccgcgcggcg	aggaaagagg	gaatgagcga	gatgaaggtg	gtgcacagcg	50940
tgatgaacgc	gatggctctc	atacaccgatc	cgggcttccac	gatgaggtgg	agcttgtcgg	51000
agaggatgaa	gagctggacg	ggcaccggcca	cggaggggctc	cacggcgttg	atgagcagcg	51060
acacgcccat	gcccaccagg	gcgcaccacg	tgggtcccag	cagtcgcagc	acgagccctc	51120
ccaggaggaa	catcaccagc	acgtaccagc	gtcgcattgcc	gatggcgcgc	aggggtgccga	51180
tttcccgggt	gcgtcgcgg	atggcgatcc	acagggtgtt	catgatgcc	accgcgatga	51240
tgatgacgag	cacgaagatg	aggacgcggg	tgaggggctc	catcgccgac	acggctccact	51300
tgtatgaagg	gattctgttc	tccagttgg	tgatgtccag	cttctgcccc	gtccaggcct	51360
cgcggttgcg	ggcttggaa	ttcatgaaga	aggcccgggg	gtcatgtctc	agcacctgat	51420
aacccaacte	gggcagacgc	ttgtagagcg	gcgcctgcac	gctggggatg	gcgtctatgt	51480
ccttgagggt	gagcatgagg	gcgcggcttg	agtcctcgcg	cagctggtag	agggcgccga	51540
gggtggcggt	gggcaccaag	acgttgaagg	aaactcagcat	gcccacgttg	gcggcgatgg	51600
ccaccacgag	tacgtccacg	gtgttgcgta	tcccgcgcac	gggtggacgc	gagagggtga	51660
cgtgtcacc	caacttgacc	tcgagccgct	tcgcctgctc	gtcgaagagg	aggagggtat	51720
tgggttgcgc	caggtcttcc	aaccgacct	cccgaactg	cagcaccttg	cggatgccag	51780
tctcggccgc	tacgtcgatg	ccgcggatc	ccgtctgcac	ggagccagac	tcgtccacca	51840
acttgaccga	gcgcgcgctg	cgttggacgg	agaagtccag	ctcggggact	tccttgccga	51900
gtctctcgag	cagcttgggg	taggaggtca	ccacgggcgc	agactggccg	gccgtcacct	51960
tgtagaagcc	agccacgttg	acgtgcccg	tcaccagcgt	gggtggcgga	cggagcatcg	52020
tgctcttcat	gccgttggac	aggcccatga	ggatgacgag	caggcccggt	acaccggcga	52080
tggcgcgcgc	gcgcagaaag	gtacggcgct	tgtgggtgcc	caggttgcgc	actgcgatga	52140
ggaggagctg	ttgcattggt	tcactcgtcc	gtctgcacgc	cctggagagg	cgagaccggg	52200
gtcgcgaggt	acgcggggta	gaaggtggag	agggcggaca	ccacgagcac	gatgacgaag	52260
gcgcgccaga	ggtttgacag	gtggagactg	gggaagagcg	gggttcccga	gaagaagaag	52320
tagagcgctt	cgttgcggcg	ggggatgccc	acgtggccga	gcattgttcat	gatggcacct	52380
ccatagcgcg	ctcccagcac	gccgaagacg	agcccccaga	ccaccgtttc	caccagcacc	52440
atgtcagca	cgaacgagcg	ctgcgcgcgc	atggcccaga	gggtgcccac	ctcgcgcacc	52500
cgttgcagcg	tggccatcat	catcgcttg	ttgatgatga	cagcgcgcac	cacgaagatg	52560
atgaagacgg	cgaagttagag	caccagcttg	gcgaccagga	cgaactggcc	gatcgtgccg	52620
gagggcttct	gccaggagat	gatccgcaag	ggtagtttcg	cgtcgtccgc	cgatttccgc	52680
agctcggcca	gggtctgctc	cagcttctcc	ggatgcttca	gcagaaccgc	gggtctgagc	52740
accacgcgcg	tttcgatttc	ctgctgcgtg	tacaccgggg	aggcgagctc	ctcgcgggtc	52800
agcttctggg	cgaagccgct	gagttgcttg	tcctcgtcga	tctggccggc	gggtccctcg	52860
gccaccagcg	aggcgtgcgc	ctgctcgcca	aagagcgccg	tctcggcgctc	ctcgcgcttc	52920
acctgtctga	ccccgttggc	cttctgcagg	cccgcgagct	cggccttctt	ctcagcggtg	52980
agatagccgt	acagctcgcg	gaaggacatc	aggtccagca	ggttgagggc	tcggcgacc	53040
gcggacttct	ccagcccgct	gaactggtag	gtgccgtaga	tcttcaagtt	cacgtctctg	53100
acatagccgg	tgcgcgagaa	tgcgggtgatg	gtgaggtcgt	ccccgatgcg	gatgcggtac	53160
aggctcagca	gcgtcccgag	ctcggagtag	aaactgtggt	agcgcgtgct	gaagttggcg	53220
tcattccatgg	tgaagaagcg	gggcagtagc	ttgccagggt	ccgtctcctg	gctgccccagc	53280
acgcgttgga	gccgtccacc	ggcctgtctc	gtctttaggt	cgtcagagctg	gaagaggatc	53340
tcccgcgtct	gggtcttggt	ctccttcacc	cagcgttgga	gttgcggatc	catcgcgatg	53400
gtcttctggt	tgggtatcacg	cgcctctctg	atgagatcca	accggtgcgc	cgtcttcagc	53460
ttgaagtcgt	tctcgtaggt	gaacttggag	agcatcatgc	cgcgggtccc	cgggggcacc	53520
ggcgtgcctt	ccagatgcgc	catcggtcgc	aaggtctctt	ggaagttgac	caggtcgggtg	53580
ctcatagagc	gcagggacaa	catgtccccg	tccgtcatat	accggggcgat	gcggttctcc	53640
aggaaactga	gcagatcgaa	tggcttctcg	tcgaagtcgc	cccagaagcg	ctcggaacgg	53700
gcgcgggcca	tggcctccgc	gtccgcgggg	tccgtggtct	tgctcgtgat	gatttccctg	53760
cgcgcgttca	tatctctctc	gagcaagggtg	atgatgtgac	gcacatgcgc	ctggaggctg	53820



[0014]

tggtatctgcc	cgcggagttc	gggtgtgtcg	ccctgtgctg	ctttcttcta	gaggtcgcgc	53880
aggegcgcga	aggtcaggte	gatggtgttt	cccagttga	tgaacgtgce	gccggtgccc	53940
atgggcacca	ccgtcttcac	gttgggtgce	tgtgtacca	gttgcttgat	gcgcgagaag	54000
tcateccagc	cgtcaggte	cggttcgcgg	cccatctgcc	cgaagagcga	gagctcgtcc	54060
ttggagttgg	ccaggtacac	ctggagggtg	ccggcgacgc	tgcgataat	gctgcggctc	54120
atcgctcgt	ccacgtgtgc	gacgaggag	ccgccacca	ccaccagcac	ggtgccgaag	54180
aatgatgatc	ctccgatgag	gaggttgatc	ctgtcacga	acaagttgcg	cagggccact	54240
tgagcagga	gcttgagttg	gcccattagt	ggccccctc	gctcacggcc	atgaccttct	54300
ggcctcggc	cggcgtgatg	cggtcgagga	tcttcccgtc	cgcaggcgcc	accacggcgt	54360
tgctgtggc	catcaccttg	gcgtcgttgc	tgagagaagt	gaaggttggt	ccctccttgc	54420
ggttgagctc	cttcacagg	tcgatgatgt	tctggccggt	gacggagtcg	aggttggcgg	54480
tggtctcgtc	ggcgagcacc	agcttgggccc	gggtgacgag	agcgcgcgcc	acggccacgc	54540
gctggcgctg	gctccagac	agctcattgg	ggcgggtgtt	ggcgtgtctc	tcaggcccca	54600
cattgctccg	cagcgtctac	acgcgcgtgc	ggcgtcggga	ggcgttgagc	ttgcgtgca	54660
gcagcagggg	gaactctacg	ttctggaaga	cgtgagcac	cagacagagg	ttgaagctct	54720
ggaagatgaa	gccgatgggt	tgcagccgca	agtgggtgag	ctgcgcgtcg	gtgagcttct	54780
tggtgtcctg	gccatccacg	ctcaccacgc	ccgaggaggc	cgtgtccacg	cagccgatga	54840
gattgagcgc	ctcggcttgg	ccactgcccg	atgggcccgc	gatggagatg	aactctcccg	54900
ggtacacctc	tagcgtcacg	cctcggagtg	cgggcacctg	caccttacct	agggagtacg	54960
tcttggtaac	ctcgggtagg	gagacgatcg	gctgggtgct	gccggggagg	gcagtacact	55020
ggctcatgat	tgtttggate	ctttccgcga	aacggaggga	tggggtgggg	gacgcctggg	55080
agggggggcg	ctcggcgttg	gcgtgcggcg	gacgagggtg	atggcaactg	gtattgaatt	55140
cgcagatgag	cggctccccc	tggtattccc	ccaccggggc	aaaagttgcg	cgttgtcttg	55200
actactggcg	tcaagacatt	gagtcaacgc	cgaaggagag	cgcattccaa	aagaggcagc	55260
gtccatggag	cgaaggcagc	ggcgagtggt	gcattgcgtc	agaggggaaa	acagggtcgg	55320
taggacagag	gaatcgaaac	tcccggggag	atgtctccat	gccccccacc	ggttttgaag	55380
gctgggtgtg	tcagtggggt	tctccctcgg	agattgcac	tggttccact	cggctgtatc	55440
ccagggacgt	aatagggacg	taatcccga	tccgatgggt	gcagcatgcc	gcagaagtcc	55500
gtggggaaat	ggaagggcgg	gcgggtcaag	ctcgtcgatg	gtcggaaagt	gtggctcttc	55560
gagaagatgt	ctcggggggc	cgggttctcg	gtctccttgg	cggctctcaa	cgaggaggac	55620
gcctctggcg	agctggccct	gttcggggag	gaccgggacg	cctacctggc	caaggtgaag	55680
gccgacaggt	cggaggaagt	ccaggcatcc	actgtagccg	gggcagttcc	tctgtcgggg	55740
gatgtggggc	ctcggctcga	tgcgattctt	gtccgggagt	tcttccgaca	cttgaccacg	55800
cggggggcga	cggagggtta	ccggcgggac	gcccgaacct	acctgtcgca	atgggcccag	55860
gttctgagga	gaaggacact	gagtaccctg	agctctctcg	agttgcgcgc	cgccttgagc	55920
caatggccca	cggccaggaa	gatgcggatc	atcacgtcga	agagcttctt	ctcgtggctg	55980
agggaagagg	atcgctcaaa	ggctgtctga	gaccccacgt	tgctccctca	gggtccgccc	56040
cggttcgcgg	agaaggggag	acggggcaag	gggtattcga	tgccccaaat	ggagaagctc	56100
tacgcggcca	aggtctccca	gacggtgagg	gacgtgctgt	gtctgcgggc	caagaccggc	56160
atgcacgact	cggagatcgc	ccgcctggca	tcgggcaagg	gggaactcgc	cgtcgtcaat	56220
gacccctccg	gcattcgcgg	tactgcgcgg	ttctgcaca	agaacggccg	cgttcacatc	56280
ctcagctcgg	atgcccagge	ccttgcgtgc	gcgcagcgge	tcagggttgc	gggcagggcg	56340
cccatcagga	acaccgtccg	gggtatgcgt	cggcgcgcgt	tgggcagtcg	tgggcagtcg	56400
cccatccate	ccagcgagct	ccgccacage	ttaccacctt	gggccacgaa	tgagggccag	56460
gtcgtgaggg	caacccgggg	cggagtgcga	ctcgatgtcg	ttgcctcggg	tcttgcccat	56520
cagtccacac	gggcgaccaa	gaagtcttat	gacgggacgc	aaattccccc	gatgatcacc	56580
gtacgcgcca	agtcgcatca	tcacaggac	ccagcggtag	tgcagctgag	gcgtaaatgc	56640
tcgccggacc	ccgtcgtgac	gagagaggca	gaggcgtgag	acgtccaggc	catcaacctg	56700
gaggtacacg	tgagacgtgc	cggggctcct	ccccgcacct	ccttcgaggt	tgatttctct	56760
tgctcctcgc	attccccctc	ggcctctctg	cgttgccgct	cctgtccact	accaccgaaa	56820
tctctgcggc	tcttcccggt	gacgagtgcg	agtcggcgag	cctgcgcac	gagctgcccg	56880
ctacgcggcg	gggaaagcca	cccgtgggtg	gtctcggctc	aggtctgccc	attcatttcc	56940
gcttcgactc	cgcgtcccaa	cagaagtccc	tgaggattca	ggatcggggc	tggttcgagg	57000
attgggcttt	gggcacagcag	acgtcgtac	tgactctcca	cgacaacctg	gtggctggga	57060
agcgatctga	agtggagggt	tgttcgcgg	atggtgccgc	cccggcgctg	gcttctctcg	57120
tgctccggcg	ctgaggcgag	tgcaccgcac	tgattcagtt	cctcttcaac	cccggtagcc	57180
ctcgccacg	cggtagagct	gggtgaggga	gtagtccagc	aacgattcgc	acgagcgcat	57240
gtagttggtg	gcccgcgcaa	acggcaacca	cacacgggtg	ccgacgagca	cctgggcttc	57300
gtcagatttc	tgccgaggta	tcacattgcc	gcccagggtt	cgcaccagca	cctcgcccaa	57360
gtacgcgcca	atggcgggca	cccggtgcgc	gtcagatgtg	tgcgctcga	acacctcgg	57420
gaagtctcca	tgcagaaact	ggtagtccac	gtcactgaga	gactcgggcg	ttgcctcgaa	57480
gacggagggc	accttcgtgt	gcacagcgcc	cacgaggtgc	tcagccagag	ttctgtaatg	57540
ctcgttagcg	cgtccaggt	tttccacatc	cggatggcga	cgtatccac	cacgtgagag	57600
aggagtggcg	ctacatccgg	gtggaagcag	ggctcaacgg	gagcgagcgc	ggcgctacgc	57660
tcgtgcaggg	ttcgagcacc	cgtgtcgaa	cggaggtccg	gccggaggtg	aacgtgcgcg	57720
cgcgcctgtg	cgtgccgtgc	ctcggcgccc	gcgaagtccg	cagcgggtgg	ccacgtcacc	57780
aggaggtatg	agccattggc	cagttctctc	acccgggtgag	ctggcggtga	cagcatgcgc	57840
tcgcggccca	cggcttccac	caacttgggg	ccgaagacgt	tgagccagaa	gatctcgtag	57900
attctgtcga	accgctctct	ccgtgcggtc	cgcgcacatc	gtccaaaatc	gggcgcacct	57960
gccaacgcgc	tgtcagccac	gctgtgggct	gcggcggtgag	tgaccgggta	gcaagaggcc	58020
caggtgcgta	ccatttctac	gaattggcgg	cagcgtctct	tctcccgcaa	gcgggtgagc	58080

	ggttgaccg tagtcattac gtccaaagcg ggcggaagcg gcggaacca gagatgcagc	58140
	gacatctcca gtgtcgccg ctgtgtgcgg tagagccagg tgtctgtgct tcgttcateg	58200
	cgcgcctcct ccagagcett ccagatattg gctcgggagt atttgagtcg ccgcctgcca	58260
	ctgacgactt ccggcatcca atcgcttcca tatctctcca gcgcctgaa aaatggctcg	58320
	agaaatttct caagcgagc ctgcggatca agcgcaccct caaaagttag ccggagactg	58380
	tcttcgact tcacgtcacc aagccccagc accttcattg aaacaggacc tccactcccg	58440
	gaactgcctt ctcaagt	58456
	<210> 2	
	<211> 213	
	<212> DNA	
	<213> 深棕色孢囊杆菌	
	<220>	
	<221> misc_feature	
	<222> (1).. (213)	
	<223> CysA	
	<400> 2	
	atgagcatga acggggacga agccgagtac gttgtcttga tcaacggcga agagcagtac	60
	tcgtcttgge ccgtgcaccg cgaattccg ggcggttgga agaccgttgg gcccagggga	120
	agcaaggaaa cgtgtcagtc ctacatccag gaggcttgga cggacatgag gccgaaatcg	180
	ctacgggaag cctgacgcg cagcaactgc tga	213
	<210> 3	
	<211> 954	
	<212> DNA	
	<213> 深棕色孢囊杆菌	
	<220>	
	<221> misc_feature	
	<222> (1).. (954)	
	<223> CysB	
[0015]	<400> 3	
	atgagtacgc cagcagcagg agcgaagccg tcctatctcg cgggtattga aacggtagatg	60
	gtcgaacctg agcttgagga gggtcgtac ctgaccgtgg agagcggcga cggacggcag	120
	agtacctctt atgagttcgg tccgaaggac gcggagaagg tcgttggtctt gccgccctac	180
	ggagtcaact tcttgctggg ggccgcgactc gcccggtcc tctccagcg attccatgtc	240
	ttgatttggg agtcaagggg gtgtccggac tccgccatcc cgggtgatga caaggacctt	300
	gggctcgccg accagtcagg gcattttccc gaggtcctca agcagcaggg cttcgaggcg	360
	tttcaactcg teggttggtg tcaggcgccg cagctggccg tgcattgccac cggcagcggc	420
	caggtcaagc cgcggacgat gtcttgatt gccccggcgg ggctgggtta ctgctggtc	480
	aagtccgagt tgcagcatg tgcactgcc atctacctgg agatcgagaa gcatggcctg	540
	ttgcacgccg agaagctcgg caggcttctg aacaaataga atggcgttcc cgcgacggcg	600
	cagaacgcgg cggaaaagct gacgatgcgc catttggccg acccgcgatg gacatacgtc	660
	ttctccaggt acatgaagc gtaggaagc aacaggctcc tgcgaagca atttgtctcg	720
	accgcgctgc actcgggtgc gacgtggcc attcactgcc gggacgacac gtacagccac	780
	ttctcgaggt ccgttcagct ctgaagctg catccatccc tcgagcttcg cctactcggt	840
	aaggcgccgc atctgcagat cttcaacgac cccgccacac tggcggagta cgttctcggt	900
	ttcatcgaca ccaggcgctc gcaggtgcc gctcctgcgg tggcgggagc gtag	954
	<210> 4	
	<211> 1380	
	<212> DNA	
	<213> 深棕色孢囊杆菌	
	<220>	
	<221> misc_feature	
	<222> (1).. (1380)	
	<223> CysC	
	<400> 4	
	atgatacttc ccaacaacat cggcctcgac gagcggacgc agctcgcac gcagatctcc	60
	tcgtaccaga agaagttcca cgtgtggtgg cgcgagcggg ggcccaccga gttcctcgat	120
	cggcagatgc gccttcgcac gccgaccggg gcggtcagcg gcgtggactg ggccgagtac	180
	aagacgatgc gtcccagca gtatcgctgg ggctcttcca tgggtccgat ggaccaggac	240
	gagatgcct tcggcgacca ccgtggcaag aaggcgtggg aggaggttcc gagcgaatac	300
	cgacagctgc tgcgcagca catctgcgtg caggccgacg tggagaacgc cgcgctcgag	360
	cagagccggc tgctgacgca gatggcgccg agcaaccgg acctgagaa cgtgttccag	420
	ttcttctcg aggaggggcg ccacacctgg gccatggttc acctcctgct cggccacttc	480

```

ggtaggacg gggctcgtcga ggccgaagcg ctcttgagc ggctgagcgg tgacccgagg 540
aacccecgct tgcctggagc gttcaactat ccgaccgagg actggctgtc ccacttcatg 600
tggttgccttgc tggccgacgg ggttggcaag taccagatac atgcagtgc cgaggcttcg 660
ttcgccecggt tggcccgggc ggccgaagttc atgatgttcg aggaaccgct ccacatcgcc 720
atggggcgccg tgggtctgga acgagtgtcg gccaggaccg ccgaggtcac cctgcgtgag 780
gggacgttcg atacgttcca cgcggggggc attccgttcc cggttgtcca gaagtatctc 840
aattattggg cgcgaaggt ctacgacctc ttccgaaacg acggctccga acgctcgaac 900
gaactcttcc gggctgggct ccggaggccg cgggaatttcg tgggaagcga atcgcgatc 960
gttcgcacgc atgagcgcat gggcgacgga ctgaccgtcg tgggaagtga aggggagtg 1020
gcgataacg ccattcatgc acgacagttc atcgccgaag tgcaaacgt cattgategc 1080
tggaacgcca gcttcgagc gctgggcgtc gacttccagt tgtacctccc tcacgagcgc 1140
ttcagcagga cctatggccc ctgcgccgt ctgcccttcg acgtggacgg aaaactgtc 1200
ccccgcggca cggaggcgaa gctcgccgag tacttcccca caccctcgca actcgcgaac 1260
gtccgcctcg tgatgcagcg ggagctggct cccgggcagt actctcgtg gatcgccccg 1320
tcgcgacgc ggctgagcgc gctgggtccg ggcaggaaca cgccaagga gcacgaatga 1380

```

<210> 5  
 <211> 2199  
 <212> DNA  
 <213> 深棕色胞囊杆菌

<220>  
 <221> misc\_feature  
 <222> (1).. (2199)  
 <223> CysD

[0016]

```

<400> 5
atgcgttgcc tcatcatcga caactacgat tcgttcacct ggaatctggc ggactacgtt 60
gcgcagacgt tccggagcga gccgttggtc gtccgcaacg accagcatac ctggcaagaa 120
atcaaggcct tgggtcctt cggatgcac ctggtttctc cgggtccggg ctccgtgacc 180
aatccgaagg attcaatgt ctccgagac gcgctcgagc aggatgagtt cccggtgttt 240
ggggtctgcc tgggccatca agggctggcg tacatctacg ggggcgagat cactcacgct 300
ccggttccgt tcaacggcag gacgtcgacc atctaccatg acggcacggg cgtgtttcag 360
ggactccgcg cgagcttcga cgcgggtgaga tactactcgc tggctcgtcg gccggagtgc 420
cttcccgca acctggctgt caccgtcgg acggaatcg gctgateat ggggttgccg 480
cacgtgagtc gccgaagtg gggcgctcag ttccatcccg agtcgattct gactgcgcac 540
ggcttgagc tcatctccaa ttccgtgac gaggcgtacc gatacgccgg gaaagagggt 600
ccgtcgcgc gtcgccatc gactgccgcg aacggtgtcg gcgcagggtc tgccaggcgt 660
gaccgcagcg cccgcgcac accggagcgg agaagggaac ttcagacgtt caccaggcgg 720
ctggcgacgt ctctcgagc cgagaccgtt ttcttgggccc tgatgagggg ccgcgagcac 780
tgcttctggc tcgacagcca gtccgtgaga gaaggatata cccggttctc ctctatgggc 840
tgctgcccgg agggctcgt gctgacgtac ggctcgtcgg aagcggcgct agaggggggc 900
gccgagcggg acctggcggc gctggagcgg gcgctcgaaa gccgtatcgt tgttcgcccc 960
gtggatgggc tgccattcga gtttcatggc ggctacatcg gcttcatgac ctacgaaatg 1020
aaggaggcgt ttggggccgc gacgacgcac aagaacacta ttcccgacgc cttgtggatg 1080
cacgtgaagc ggttccctgc gttcgaccac tcgacgcgag aagtgtggct ggctgccatc 1140
gccgagctcg aggagagcgc gacgctctc gccctggatgg acgagaccgc cgacgctctg 1200
aagtcgcttc cgcgcggcac ccgttcgccc cagtccttgg ggttgaaatc catctcggt 1260
tcaatggatt gtggacggga tgactacttc gccgccatcg agcgtgcaa ggagaagatc 1320
gtcgatgggg agtctacga ggtctgcttg acgaacggtt tctcgttcga tctgaagctg 1380
gateccgtcg agctgtacgt gacgatcgcg agaggcaatc ccgccccgtt cggcgcttcc 1440
atcaagacag gcaagacctg cgtcctcagt acctccccgg ageccttctt gaaggtggat 1500
gaggatggga cgggtccagc caagcccatc aaggggacct gcgcgcgctc tgacgacccc 1560
gccaccgaca gcacgaatgc cgcgcggctg gccgcctcgg agaaggaccg gccggagaa 1620
ctgatgatcg tggacctgat gcggaacgac ctccggacgg tgctcgtgcc gggcagcgtc 1680
catgtctcca atctaattga catcgagagc ttcaagacgg tccatcagat ggtcagcacc 1740
gtcgaatcga ccttgacgcc ggagtgcagc ctctgtgacc tctcgcgcgc ggtcttcccg 1800
gggggatcca tcaccggggc tccaagatc cgcacgatgg agatcatcga tcggctcgag 1860
aagagccctc ggggcatcta ctgcggcagc atcggttacc tcgggtacaa ccggatcgcc 1920
gacctgaaca tcgcatcccg cacttgttc tacgacggca cctcgtgaa gttcgggtgc 1980
ggcggagcca tcacctact gtcacagcgg gagggggagt ttcaggagat cctgtcgaag 2040
gcggaatcca tctcggccc gatttggcag tacatcaatg gcgcgggtgc tcccttcgaa 2100
ccccagttgc gcgaccgggt tctgtgcctg gaggagaagc cgcgaagggt cattcgtggc 2160
cacgggtcgg caattgatgc agtgagacct agcgcgtga 2199

```

<210> 6  
 <211> 732  
 <212> DNA  
 <213> 深棕色胞囊杆菌

<220>

<221> misc\_feature  
<222> (1)..(732)  
<223> CysE

<400> 6  
atgattgcgt tcaaccgcga ggccgggccc aggcctgcgcg tcttctgctt tccgtacgcc 60  
ggtggcgacg cgaacatctt cggggactgg gcccgggcga tgcccgaggg ggtcgaggtc 120  
ctcggcggtc agtaccgccg gcgcgggtacc aacctggcgt tgcccgccgat cagcgactgt 180  
gacgagatgg cgtcacaaact gctggcggtg atgacgccgt tgcctggcat caacttcgct 240  
tttttcggcc acagcaatgg cgccttgatc agcttcgagg tggcgcgaa gctccacgac 300  
gaactgaagg gccgcacgcg gcatcacttc ctgtcggcca agtcgcgccc tcatcaccgc 360  
aacaacagga gtaagatcag cggcctcaac gacgaggact ttctccgggc gatccggaag 420  
atgggcggta cgcgccagga agtgcctcgc gacgcccgcc tgatgcagat tctgctgcca 480  
agactgcgcg cggacttcgc gctcggcgag acgtatgtgt ttccgcccg acccaccctg 540  
acgtgcgacg tcagcatcct gcgaggcgag agcgaccacc tggtcgacgg cgagtctctc 600  
cagcgggtgt ccgagctgac gacgggcggc gcgagccagt acgcaataga tgggtggccat 660  
ttcttctga attcccacaa gtcgcaggtc gtggcgctcg tgcgagcgcc actgcttgag 720  
tgtgtgttgt ag 732

<210> 7  
<211> 1038  
<212> DNA  
<213> 深棕色孢囊杆菌

<220>  
<221> misc\_feature  
<222> (1)..(1038)  
<223> CysF

<400> 7  
atgaccgctc agaaccaage ctcccggttt tctttcgatc tctttacac gacggtcaat 60  
gcgtactacc ggaccgcgcg cgtcaaggcg gccatcgagc tcggcggtgt cgacgtcgtt 120  
ggcgagaaag gcaagaccct ggcccgagatc gcgaaggcct gcaacgcgtc gcccggtggc 180  
atccgcattc tctgcgggtt cctcgtgtcg atcgggttcc tcaagaatgc ggggtgagttg 240  
ttcttctca cgaacttctt ggccctgttt ctggacaaga agtcgcgccg ctatctgggc 300  
ggcagcattg atttcttct gtcgccgtac atcatggacg gcttcaagga cctcgcgtcg 360  
gtggtgcgga cgggcgagtt gacgtgccc gaaaaagggg tgggtggccc agatcatccg 420  
cagtggtgta cgttcgcgcg cgcgatggcg ccgatgatgt ccccgccatc cctcctgctc 480  
gcggaactgg cggaccgcca ggccgaaccag ccgctcaagg tgctcgatgt cgcgcgcggc 540  
cacggcctct tcggcctgce catcgcccag cggaaatcca aggcgcgatgt gacgttctc 600  
gactgggaaa acgtgttaca ggtggcgcg gcgaacgcga cgaaggcggg agttctcgac 660  
agggtcgagt tccgcccggg agatgccttc tccgtggact tcggcaagga gctggacgtc 720  
atctctctga cgaacttctt gcatcacttc gacgagcgcg gctgcgagaa gatcctcaag 780  
aaggccacg ctgcctgaa ggaggcgcg cgtgtgtga cgttcgagtt catcgcaac 840  
gaggaccgga cgtgcctcc gcttgccc acgttcagca tgatgatgct cggcacgacg 900  
ccggcggtg agacctacgc ctactccgat ctggagcgga tgttcaagaa cacgggttac 960  
gatcaagtcg agctcaagge cattctctcc gcgatggaga aggtcgtcgt ttcgatcaag 1020  
ggcaaaagcg agctctga 1038

<210> 8  
<211> 5979  
<212> DNA  
<213> 深棕色孢囊杆菌

<220>  
<221> misc\_feature  
<222> (1)..(5979)  
<223> CysG

<400> 8  
atggccacca aattgtctga cttegcgcgc ctgcactcgc aagacgcca cgtcatctcc 60  
cgctcgaaag agacggggat atcgtctgat ctgtccaaga gcgtgggtga cttgttcaac 120  
ctccaggteg agaggcgccc tgacgccacg gcgtgtctcg gccgccaggg gcgcttgact 180  
tacggagaac tcaaccggtt gacgaaccag ctccgcgcat acctgatcgc gcgagcgctc 240  
gggccggatg ttcccggtgg cgtcctgttc gagecgtcgc ccgagcagct catcgccatc 300  
ctggcgctcc tcaaggcggg cgggtgttat gtcccgttgg atccgcagta ccccgccgat 360  
tacatgcagc aggtctctgac ggacgcccg ccgcggatgg tgggtgcgag ccgggcgcgc 420  
ggcgagcgcc tccgctcggg cgaggagcag atcgtctacc tcgatgacga acagctcctg 480  
gcgcgcgaga cccgcgacce gcctgtgaag gtgttgccgg agcagctcgc gtacgtgatg 540  
tacacgtcgg gctcgtccgg agtgcggaag ggctcatggt tgccccatcg ccagatctc 600  
aactggctgc atgcactcct ggccggggtg ccgttcggcg agaacgaagt ggtggcccag 660

[0018]

```

aagacgtcca cgtcattcgc catctcagtg aaggaaactct tcgcgggatt ggtcgcgggt 720
gtcccgaggc tcttcatega cgtatcgact gtccgcgacg ttgccagett cgttcgtgag 780
ctggagcagt ggccgcgtcac gggcgtctat acttttccct cccagctggc ggccattctc 840
tcgagcgtga atggcgcgta ctagcgcctc cgtctcgtgc gccacctgta catctcgate 900
gagccctgcc caacagagct gctggcgaaag ctccggggcg cgatgccgtg ggtcaccccc 960
tggtaacatc atggctgcac cgagatcaac gacgtcacct actgcgaccc aggggaccag 1020
gctggcaaca cgggcttcgt gccgatcggg cggcccatcc gcaacacgcg ggtgttcgtc 1080
ctcgacgaag agctccggat ggtgcccgtc ggcgcgatgg gtgagatgta cgtggagagc 1140
ctgagcacgg cgcggggccta ctggggcctt cccagttga cggcggagcg gttcatcgcc 1200
aaccctcacg cggaggacgg ttccgcctcg tacaagacag gcgacctgc ccgtacctg 1260
ccgatgggtt ccttgaggtt cctcgggcgc cgggactacg aggtgaagat ccgcgggtat 1320
cgcgtggacg tcggcaggt cgagaaggtc ctccgggcgc atcccgacat cctcgaggtg 1380
gcggtgttgg gctggccgct cggcggggcg aatccacaac tggtcgccta cgtcgtgccg 1440
agggcggaagg gggctgtccc catccaggag atccgggact acctgtcggc gtccctgcgc 1500
gcctacatgg tgcgcagat ctccaggtg ctggcggcgc tgcacgtct tcccaatgac 1560
aaggtggatc ggttgagcct gcccgacccc aaggtggagg agcagaccga ggggtacgtg 1620
ggccttcgca cggaaaccga gaaggtactg gccgaaatct ggagcgacgt cctcagccag 1680
ggccggagcg cggctgacct cggcgcgacg cacaactttt tcgaactggg aggcatttcg 1740
cttctcgcgc cccagatgtt ctccgggact cggcagaagt tcgattctga actgcccata 1800
aacacctgtt tcgagacccc cgtcgtggag ggttttgcga gcgcgctga cgcggtcttt 1860
gccgagcgga acggtccgcg gcagaggtcg atcagatga cggaccgcgg ccaggcgctt 1920
ccgctgtcgc ccgttcagga cggcgtctgg ttctgtcacg agcagatggt cgagcagcgg 1980
agcagctaca acgttgccct cgcctgccac atgcgtggca aggggctgtc gatgccggcg 2040
ctgcgcgcgc ccatcaacgg gctggtggct cggcacgaga ccttgcggac gacgttcgtc 2100
gtctccgagg cgggaggaga tccgctccag cggatccgcg actccctgtg gatcgaggtt 2160
cccttatatg aggtcgtatg gtcggaagtc cggccccgca tggcggccca cgcgggccac 2220
gtgttcgacc ttgcgaaggg cccctgtctg aagacctcgg tcttgcgggt gacgccgat 2280
caccacgtgt tcttgatgaa catcctcac atcattctgt atgggtgttc gatcgacatc 2340
ctgctgcggg acctctacga gtcttacaag gcggccgaga cgggctcgca gccgaacctg 2400
ccggtctcgc caatccagta tgcgactac tccgtgtggc agcgtcagca ggacctcagc 2460
agtcacctgc actactgaa gaagacgtc gagggctacc aggaagggtt gtctctccg 2520
tacgaattcg cccgcccgtc caacaggacc tggcgtgccg cgagtgtccg gcaccagtac 2580
ccggcggaac tcgccacccg tctgtcggag gtgagcaaga gccatcaggc gacggtgttc 2640
atgacgttga tggccagcac ggcaatcgtg ctgaaccggt acacgggtcg ggatgatctg 2700
tgcgtgggtg tcacgggtgc gggccgtgac cacttcgagc tcgagaacct gattggttc 2760
ttctcaaca tcttcgcat caggtctgac ctccagcgga atccacggc cgagacggtg 2820
ctgcagcggg cgcgagcgca ggtcgtggaa ggcatgaagc atccgcacct gccgttcgag 2880
cacatctcgg cggcgctgca gaagcagcgc gacagcagcc agattccctt ggtgccgggtg 2940
atgctcgcgc accagaactt cccgacagt accctcgagg agcaggggct cgacctgggt 3000
atcggggaga tcgagtttgg tgagcggacg acgcccacg agctcgacat ccagttcctc 3060
ggcgagggaa gcacgttggg ggtggtgttc gagtacgcga aggatctgtt ctccgagcgc 3120
acgatecage ggtcctacac gcaattgcag caggtgtcgc agactctcgt ggacaagccg 3180
gacttcgcgg cgtccgattt tccgtgtgtg gccggggacg cgttcgaggg cgggtgttcg 3240
ggctccgggg gcgcgacgaa gaccggcaag ctccagctgt cgaagagccc ggtcagttg 3300
ttcaacgagc ggttagaggg ctcccgagc gcggtcgctt gcatgggcgc ggacggaagc 3360
ctgacctacc gggagcttga ccgaaggccc aatcaggtcg cccgccacct gatggggcga 3420
gggggtgggg gggagacgca ggtgggggtt tggttcagc gctcgcggga cctgtctgtc 3480
gcaactctgc gcatactcaa ggccggggggc tgcctcgtt cgtcgcagcc gactatccg 3540
caggagtaca tcaacaacat cgtcggcgat gcgcagccgc ttctggtgat gtcgagccgg 3600
gcgttgggtt cagcctgtc actggaggca gggcggtctg tgtacctga tgacgcgtg 3660
gcggcgctca ccgatcgag cgateccag gtgcgcatcg acccgagca gctcatctac 3720
gtcatgtaca cctccggttc caccggtctg ccgaaggggg tgcctgttcc ccacggcag 3780
atcctgaact ggctgtaccc gctgtggcg atggtgccct tcgggcagga cgaggtggtg 3840
gcgcagaaga catccacggc ctccggtgct tcgatgaagg agctcttcac gggcgtctg 3900
gcggcgctgc cccaggtatt catcgacggc accgtgttca aggacgcggc ggcttcgtg 3960
ctccacctgg agcgatggcg ggtcacccgg ctgtacacgc tccgctcgca cctcgatgcc 4020
atcctgtccc acgtcgacgg ggcggcgag cgcctgcggt ccttcggcga tgtcactctc 4080
gcggggggag cgtgccctgt tgagctgatg gagaagctgc gcgagacct gccgtcgtg 4140
acggcgtggt tcaactacgg ctgtaccgag gtcaacgaca tctctactg cgtcccgaa 4200
gagcagttcc acagctcggg gttcgtgcgc atcgccggc ccatccagta caccggggc 4260
ctggtgtcgc acgacgagct gggacgggtg ccggtgggca tcatggggga gatttacgtc 4320
gagagcccg ggacggcgcg gggctactgg aggcagccgg atttgacggc cgagcggttc 4380
atccccaaac cgttcggcga gccgggtage cgtctctacc gtacggcgca tatggcgca 4440
tgctttgagg atggctcgt ggagttctt ggccgcgggg actacgaggt caagatccgt 4500
ggccatcgcg tggacgtccg ccaggtcgag aagatcccg cgagccaccc ggaagtctc 4560
gagtcggcgg tgttgggctg gccacggggg gcgaagaacc ctacgttctg tgcctacgcc 4620
gccacgaagc cggcccgctc cctgtcgact gaaaacgtgc gggagtacct gtcggccgc 4680
ttccgacgt acatggtgcc aacgtctac cagttcctgc cagcgtgcc gcgcctgcc 4740
aatggcaagc tcgacgctt cggcgtgcgc gatcacaaga aagtcgaggt gggcgcgctc 4800
tacgtcgcgc cgcagacgcc gacggagaag gtcttggcgg gactgtgggc cgagtgcctc 4860
aagcaggggc acatgcccg gccgcaggtt ggccgcttgc acaacttctt cgacctcggt 4920

```



```

gggcactcgc tgcctgceaa tcgcgtactg atgcagggtgc agcggcattt cggggtcagc 4980
ctgggcatca gtgcgttgtt cggttctccg gtgtctaatg acttcgcggc ggccatcgac 5040
aaggcgctcg ggaccgagga gccaggcgag gaaggttcga gcgacgcacg agaggtcgct 5100
ggaaggaca cctccgtgct cgtgcgcctc tccaccacag ggacgctgcc gagcctgttc 5160
tgcgtccate cgggtggcgg gcaggctccat gcctaccgcg agctcgccca ggcatggag 5220
aagcacgcca gcatgtacgc gctccagtcg gagggcgcgc gtgagttcga cacaatcgag 5280
accttgccgc gcttctacgc cgatcgatc cgcggggctc agcccgacgg gagctaccgt 5340
ctcctcggat ggtctcttgg tgggctcacc accttgccga ttgctcgca gctggagcac 5400
cagggtcgcg ccgtggagta cgtgggcctc gtggattcaa agccaatccc gcggttgccg 5460
ggtgagcgcg gctgggcgtc gctgatcgcg gcgacgaaca tccctggcgc gatcgggggg 5520
cgcggtctct cggctcgcca ggtcgatgct gccgggaaga tccctgagtc gcgcggatgg 5580
acggaggagt ccttcgactc ggaggggcat gcggcgcttg aggagttggc tcggcacttc 5640
ggcatcacgc tcgcgcaaga gtcacggag tactctctgg cccggttcaa gaccacgaag 5700
tactacttgt cgtgttgcgc tggcttcaag ccggcggcgc tcgggcccga gacgtacctc 5760
tatgaggett cagagcgggt cggagccacc tcgaacgacg acacgggcga gtggggggac 5820
gcgctggatc gcaaggccct gcggcgcaac atcgtgcagg tgccaggcaa tcactatact 5880
gtcctgcagg gagagaacct gctcgaactg gcggggcgga tcgccgaagc cttgtctgcg 5940
atcgacaact cggtggtaac gaggacgcga gcttcgtga 5979

```

```

<210> 9
<211> 2928
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(2928)
<223> CysH

```

```

<400> 9
atggacaate gagagatcgc acccacccea tcggcgcgca cgcgtgatgc gtacacggcg 60
gtaccaccag ccaaggccga gtatccgtcg gacgtctgtg tgcaccaact gttcgagttg 120
caggcggaca ggattcccga cgcggttgcg gcgaggcgcg ggaacgagtc cctgacctac 180
cgggagctga acttcgggca gaatcagctc gcccggtacc ttgttgccaa aggcgtggtc 240
ccgcgaggct cgggtggcgt gctgatgaac cggaccctcg cgtgtctggt ctcactgctc 300
gccatcatca aggcgggcgc ggcgtacgtt ccggtggacg ccggtattgc cgcctaacgg 360
gtggactaca ttctgacgga cagcggcgcg acctgcgtcc tgaccgacag ggagacgcgg 420
tcactctcgc acgagccgcg gtcggcttcg acgtctgtca tcgacgtgga tgatccatcc 480
atctattcgg gcgagaccag caacctcggg ctgcgtgtcg atcccgagca gcaggtctac 540
tgcatctaca cctcgggttc gacgggcctt cccaaaggcg tgatggtcca gcaccgcgcg 600
ctgatgaact acgtctgggt ggccgaagaag cagtacgtca ccgacgcggt cgagagtttt 660
gccctgtact cctcgttgtc gttcgacctc acggtaacct ccatctctgt tccgtgacg 720
tcgggacgct gcgacacgag gtacccggag ctgggcgagg acgtcccggt catcaaccgg 780
gtactggagg acaataaggt cgatgtcgtg aagctcacgc cggcccacct tgccctgctc 840
aggaacacgg acctatcgca aagccggctg aaagtgtcga tccctgggagg agaggacctc 900
cgagcggaga cggcggggga cgtccacaag cggctggacg gccggcggtg gatctacaac 960
gagtagcgcc ccacggagac cgtcgtgggg tgcattgatt accgctacga ccccgcggtg 1020
gattctgcag ggctcgggtc gattggagtg ggcatcgaca acatgcggat ctacttgcct 1080
gacgaccgct ggctcctcgt caagccagga gaggttggcg agattacat cggagggcac 1140
ggtgtgaccc tggggtacaa ggacaagcct caagtcacgg cggaccactt catctccaat 1200
ccgtctctgg aaggggagcg gttgtacgcc agtggcgacc tcggccgggt gaatgagcgc 1260
ggcgcgctcg tcttcctcgg ccggaaggat ttgcagatca agctcggggg gtaccggatc 1320
gagctggggc agatcgagag cgcctctctc tccatccgg ggatcaagga atgcacgtc 1380
gattcgacca agaccgcgca gagccaggcc gccgctcagc tcacctactg caccaagtgt 1440
ggtctggcgt cgagcttccc gaatacgacg tactccgcgc aggggggtctg caacctctgc 1500
gaggccttcg acaagtaccg cagcgtcgtc gacgactact tcagcacgat ggatgagctg 1560
cagtcgatcg tcaccgagat gaagagcatc cacaactcga agtacgactg catcgtggcg 1620
ctcagcggcg gaaaagacag cactgatgca ctctgccgga tgatcgaaac cgggtcccgt 1680
gtattggcct tcacgttgga taacggctac atctcggagg aggcgaagca gaacatcaac 1740
cgggtcgttg cccggtggg agtggatcac cgtatctctc cgaccggcca catgaaggag 1800
atctctgctg acagcctgaa ggcacacage aatgtgtgca acggtgctt caagaccatc 1860
tacacgtttg cgatcaacct ggccaggag gtcggcgta agcagctggt catggggttg 1920
tcaaagggce aactgttcca aacgcgcctc tcggccttgt tccgcacgtc gaccttcgac 1980
aacgcgcctc tcgagaagag cctcgtcgac gcgcgaaaga tctaccatcg catcgatgat 2040
gccgtgagcc gccgtctcga cactacttgc gtcgaagaac acaaggatcat cgagaacatc 2100
aggttctgtg acttctatcg ttattgccac gccagccgctc aggagatgta cgactacatc 2160
caggagagag tcgggtgggc caggccgatt gacaccgggc ggtcgacgaa ctgtctctc 2220
aatgatgttg gcatctacgt tcacaacaag gagcgaggt accacaacta ctccctgccc 2280
tacagctggg acgtccgagt ggccacatc agcgatgag agagctcgac 2340
gactcgccgc acatcgacgt cgagagggtc gaggcatca tcaaggacct tggctacgag 2400
ctgaacgacc aggtgtgtgg ctcggcgga gccagctgg tcgctacta tgtctccgcg 2460

```

[0019]

```

gaggagttcc ccgcgtccga cctgcggcag ttctgtctgg agattctgcc ggagtacatg 2520
gtaccacaggt cgttcgtcca gctggacagc atcccgtga cgcaccaatg caaggtcaat 2580
cgtcaggccc tgcgaagcc tgacctgctt cgggaaggccg gcaccgacgg acaagccgca 2640
ccccgaacac cgggtggagaa gcagttggcg gagctgtgga aggaggtgct gcaggtcgac 2700
agtgtcggga tccacgacaa cttcttcgag atgggcgggc actcgttcc ggcgctcatg 2760
ctgctctaca agatcgacag tcagttccat aagacgata gcacccagga gtctctgaag 2820
gtccccacca tcagcgcgt cgcggcgcat ctgcgcagtg acaccgaagc ggtgccgcca 2880
gggctgggcg aggtcgtcga tcagagcgcg cctgcataca ggggataa 2928

```

```

<210> 10
<211> 819
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(819)
<223> CysI

```

```

<400> 10
gtgcgttcg tcaactgcaa tggtagggac tcggcagttt gctcgggtgt ggatcgcgga 60
ctccagttcg gagatggcct gttcgagacg atgctgtgtg ttggcgggtg gccggtcgac 120
ttcccggaac actgggcgcg gcttgatgag ggetgccgcc ggtggggaat cgaatgcccg 180
gacatccggc gcgaagtgc cgtcgcgac gccaggtggg gtgctcccag gccggtcgcc 240
aagctcgtcg taccctgggg aagcacggag cggggatacc ggtgcgcccc ttccgtccgg 300
ccgaactgga tectacacat caccgatgcc ccgaagtatc cgttgcccca cgaggacaga 360
ggcgtggcgg tcaactctg ccgaacgctc gtctcgtcgt atgaccaca gctggccggg 420
ttgaagcacc tcaaccggtt gcccaggtg ctcccgagga gggagtggga cgacgagtac 480
cacgatggcc tgcagaccga ccacggtggt caccctcgtc agggttgcac gagcaacctg 540
ttctcgttg ccgacggagc cttgaggacg cccgatctga ctgcgtgcgg tgtgcgcggt 600
atcgtgcggc agaaggctct cgaccactcg aaggcaatcg ggatccgctg cgaggtaacc 660
accctgaage tacgagatct cgaacacgcg gacgaggtct tctgacgaa ctctgtctac 720
gggattgtgc cggttggtag cgtcgatggt atgaggtacc ggataggtcc gacgacggcg 780
cgtttgtcga aagaccttg ccagggtgtg tacttttga 819

```

[0020]

```

<210> 11
<211> 984
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(984)
<223> CysJ

```

```

<400> 11
atgaccggta atttgatag cgcggcatgg cccgtaatca tcacgcctgg ccagcagcca 60
gcggcgctgg aggatgggt ctcagcgaa cgtgacggac tcgagcgcca gttgaccgag 120
tgtaaggcca ttctcttctg aggtctccgt agcaggaatg gcttcgagag cattgccaac 180
agcttcttcg accggcgctt caactatacc tatcggtcga cgccecgta ggacctgggg 240
cagaacctct acacggcgac ggagtaccgg aagcagctgt cgattccgca gcattgcgag 300
aacgcctacc agcgcgactg gccgatgaag ctgctgttcc actgcgtgga gccggcgagc 360
aaagcgggcc ggacgccctt ggccgacatg acgaaggtaa cggcgatgat ccccgccgaa 420
atcaaggagg agttcgcgcg gaagaaggtc gggtagctgc ggaactaccg tgcaggagt 480
gatctgcctt gggaagaggt gtttggaacg agcaacaagg cagaggttga gaagttctgc 540
gtcgagaatg gcatagagta ccactggacc gaggtggct tgaagaccat ccaggtctgc 600
caggcgttcg cttcgcatec actcaccggt gagacgatct ggttcaatca ggcccacctg 660
tttcaacctt ccgcatggga cccggcttca cagaagatga tgccttctt cttcgggtgag 720
ggcgccctcc cgcgcaactc gtacttcgga gacgggtcgg ccacggggag cgacgtctc 780
gaccagatcc gctccgctta cgaacgcaac aaggtctcgt tcgagtggca gaaggacgac 840
gtgttgcgta tcgacaacat gctggtttct caggacgag atccgttcga aggcagccgg 900
cgggtgctgg tctgcatggc ggagccgtat tcggaagtcc agcggcgggg attcgccggg 960
gcaacgaact caggcgctc gtaa 984

```

```

<210> 12
<211> 13638
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature

```

&lt;222&gt; (1).. (13638)

&lt;223&gt; CysK

&lt;400&gt; 12

atgctgctgg	agggagagct	ggaggggtac	gaggacgggt	tggaactgcc	gtacgacttc	60
ccgcggcagt	cgaatagggc	gtggagagcg	gcgacgttcc	agcatagcta	ccccccgag	120
ctggcgagga	aggtggcgga	gtcagccgg	gagcagcagt	ccacgtgttt	catgagcctg	180
gtggcgagcc	tggcgggtgt	gttgaaccgg	tacacgggcc	gcgaggacgt	gtgcatcggg	240
acgacgggtg	cgggccgagc	gcaggtgggg	gcgttggggg	atctgagcgg	gtccaccgtc	300
gacatcctcc	cgttagaggct	ggacctgtcg	ggcgtccga	gccttcacga	ggtgctgcgg	360
aggacgaagg	cgggtggtgt	ggagggattc	gagcacgagg	cgttgccgtg	ccagattccc	420
ttggtgccgg	tgggtggtgag	gcaccagaac	ttcccgatgg	cgcgtctgga	gggctggagt	480
gaggggggtg	agctgaagaa	gttcgagctg	gcgggggaaa	ggacgacggc	gagcgagcag	540
gactggcagt	tcttcgggga	cgggtccctg	ctggagctga	gcctggagta	cgcggcggag	600
ctgttcacgg	agaagacggt	gaagaggatg	gtggagcacc	accagcgagt	gctggaggcg	660
ctggtggagg	ggctggagga	ggtgcggctg	cacgaggtgc	ggctgctgac	ggaggaggag	720
gaggggctgc	acgggaggtt	gaacgacacg	gcgcgagagc	tggaggagcg	ctggagcctg	780
gcggagacgt	ggtgaggag	ggtagggag	acaccggagg	cggtaggttg	cgttgccgtg	840
gaggtggcga	cgggagggca	ctcgcggccg	acataccggc	agctgacata	ccggcagctg	900
aatgcgcgag	ccaaccaggt	ggcaccgagg	ctgagggcac	tgggagtggg	cgcggagaca	960
cgggtccggg	tcttgagcga	ccgtcccgcc	gagctgtctg	tggcgtgctg	ggcgatattc	1020
aaggccgggg	gctgtacgt	gccggtggac	ccacagtacc	cgggaagcta	catcgagcag	1080
atactggagg	atcgggcacc	gcaggtggtg	ttgggcaaga	ggggaagagc	ggacggggtg	1140
cgggtggatg	tgtggttgga	gctggatgga	gcgcaacggc	tgacggacga	ggcgttgccg	1200
gcacaggaag	agggagagct	ggagggggcg	gagaggccgg	agagccagca	gttggcgtgt	1260
ttgatgtaca	cgctgggctc	cacgggcaga	ccgaaggggg	tgatggtgcc	gtacagccag	1320
ttgcacaaat	ggctggaggc	ggggaaggag	cgtccgccgc	tcgagcgtgg	ggaagtaatg	1380
ttgcagaaga	cgccaatcgc	gttcgcgggt	tcggtgaagg	agctgctgag	cggattgctg	1440
gcgggagtgg	cgcaggtgat	ggtgccggag	acgctgttga	aggacacgct	ggcgttgccg	1500
caggagatag	agcgggtggc	ggtgacgaga	atccacctgg	tgccatcgca	cctgggagca	1560
ctgctgttgg	ggcggggga	agaggcggaag	ggcgtgaggt	cgttgaagta	cgctataacg	1620
gcgggggagg	cactggcgca	gggggtgagg	gaggaggcga	ggaggaagct	gccgggggcg	1680
cagttgttga	acaactacgg	gtgcacggag	ctgaatgacg	tgacgtacca	ccccgcgagc	1740
gagggggggg	gggacacggt	attcgtgcc	atcgggcggc	ccatcgcgaa	cacgcgggtg	1800
tacgtgttgg	acgagcagtt	gagcggggtg	ccggttgggg	tgatggggga	gttgtatgtg	1860
gacagcgtgg	ggatggcgag	gggtatttgg	ggccagccag	cgtgacggc	ggagcgttcc	1920
atcgcgaaac	cgtacgcgag	ccagcccgg	gcgaggttgt	accggacggg	agacatggtg	1980
agggtgctgg	cggacggctc	gctggagtac	ctggggaggc	gagactacga	gataaagggt	2040
agaggggcgc	gggtgtacgt	gcgccaggtg	gagaagggtg	gcaacgcgca	tcacgccatc	2100
cgccaggcgg	tgggtgtcgg	atggccgttg	ggctcgagca	acgcgcagtt	ggtggcctac	2160
ctggtgccgc	aggcgggcgc	gacggtgggg	ccgcggcagg	tgagggatta	cctggcggag	2220
tcgttgcgg	cgtacatggt	gccaacgcta	tacacggtgt	tggaggagtt	gccgcggctg	2280
ccgaagggga	gccttcacgg	gttgcgtcgt	ccggagccgg	acctgtcgag	cagccgagag	2340
gagtagctcg	cgtccacgag	cgaggtcgag	cggaagctgg	cggaaatctt	cggcaacctc	2400
ctggggctcg	aacatgtcgg	cgtccacgac	aacttcttca	gcctcggcgg	gcactccctc	2460
ctggctcccc	agctgatttc	gcgcatacgg	gcgaccttcc	gcgtggaagt	ggcgatggcc	2520
acgggtgttc	agtcceccac	ggtggagccc	ctcgcgccgc	acatcgagga	gaagctcaag	2580
gacgagcttc	gtagggagtt	ctccaacgtt	gtgccggtcg	agcgagcga	ggagattccg	2640
ctctcttacc	tgcaggagcg	gctgtggttc	gtgcacgagc	acatgaagga	gcagcggacc	2700
agctataaca	tcacctggac	gttgcacttc	gccggcaagg	gtttctcggt	ggaggcgttg	2760
cggacggcct	tcgatgagct	ggtggccaga	cacgagacac	tgcgcacgtg	gttccagggtg	2820
ggggagggga	cagagcaggg	cgtacaggtc	atcgggggagc	cctggtcgat	ggagctgccg	2880
ctgagagagg	tggcggggac	ggaggtgacg	gcggcaatca	atgagatgtc	ccgacaggtc	2940
ttcgacttga	gagcgggacg	gttgcgtacg	gcggcggctc	tgagggtggc	ggaggatgag	3000
cacatctctg	tcagcaacat	ccaccacatc	atcacggagc	gctggctggt	cgggggtgatg	3060
ctcggggagc	tgagggagtt	gtacgaggca	gcggtgcggg	ggaagagagc	ggagctgccg	3120
ccgtgacgg	tgcagtacgg	cgactatgct	gtgtggcaga	ggaagcagga	cctgagcgag	3180
cacctggcgt	actggaaggg	gaaggtggag	gagtacgagg	acgggttgg	gctgccgtac	3240
gacttccgc	ggacgtcgaa	tagggcgtgg	agagcggcga	cgttccagta	tagctaccca	3300
cccagctgg	cgaggaaggt	ggcggagctc	agccgggagc	agcagtcac	gctgttcatg	3360
agcctggtgg	cgagcctggc	ggtggtgttg	aaccggtaca	cgggcgcgca	ggacgtgtgc	3420
atcggggacga	cgggtggcggg	ccgagcgcag	gtggagctgg	agagcctcat	cgggttcttc	3480
atcaacatcc	tcccgctgag	gctggacctg	tcgggcgtc	cagaccttca	cgaggtgctg	3540
cggaggacga	agggcgtggt	gctggaggga	ttcgagcacc	agaggttgc	gttcgagcac	3600
ctgctgaagg	cgttgaggcg	gcagcgggag	agcagccaga	ttcccttggt	gccagtgggtg	3660
gtgaggcacc	agaacttccc	gatggcgcgt	ctggagggct	ggagttaggg	ggtggagctg	3720
aagaagtctg	agctggcggg	ggaaggagc	acggcgagcg	agcaggactg	gcagttcttc	3780
ggggacgggt	cctcgttgga	gctgagcctg	gagtacgcgg	cggagctgtt	cagcgagaag	3840
acggtgagga	ggatgggtgga	gcaccacag	cgagtgtctg	agggcgtggt	ggaggggctg	3900
gaggaggggc	tgcacgaggt	cgccgtctctg	acggaggagg	aggaggggct	gcacgggagg	3960
ttgaacgaca	cggcgcgaga	gctggaggag	cgttgaggcc	tggcggagac	gttcgagcgt	4020

[0021]



[0022]

caggtagagg	agacaccgga	ggcgggtggct	tgcgttggcg	tggagggtggc	gacgggagggg	4080
cactcgcggc	cgacataccg	gcagctgaca	taccggcagc	tgaatgcgcg	agccaaccag	4140
gtggcacgga	ggcttagaggg	actgggagtg	ggcgcggaga	cacgggtcgc	ggctcttgagc	4200
gaccgctcgc	cggagctgct	ggtggcgatg	ctggcgatat	tcaaggccgg	gggctgtctac	4260
gtgccgggtgg	accacacagta	cccgggacac	tacatcgagc	agatattgga	ggatgcggca	4320
ccgcagggtgg	tgttgggcaa	gaggggaaga	gcggacgggg	tgcgggtgga	tgtgtggttg	4380
gagctggatg	gagcgcaacg	gctgacggac	gaggcgctgg	cggcacagga	agagggggag	4440
ctggagggggg	cggagaggcc	ggagagccag	cagttggcgt	gtttgatgta	cacgtcgggc	4500
tcacggggca	ggccgaaggg	ggtgatgggt	ccgtacagcc	agttgcacaa	ctggctggag	4560
gcgggggaagg	agcgctcgcc	gctcgagcgt	ggggaagtaa	tgttcagaa	gacggcaatc	4620
gcgttcgcgg	tgtcggtgaa	ggagctgctg	agcggattgc	tggcgggagt	ggcgaggttg	4680
atggtgcggg	agacgtgggt	gaaggacagc	gtggcgctgg	cgcaggagat	agagcggttg	4740
cgggtgacga	gaatccacct	ggtgccatcg	cacctgggag	cactgctgga	gggggcgggg	4800
gaagaggcga	aggggctgag	gtcgtgaag	tacgtcataa	cggcggggga	ggcactggcg	4860
caggggggtga	gggaggaggg	gaggaggaag	ctgccggggg	cgcagttgtg	gaacaactac	4920
gggtgcacgg	agctgaatga	cgtgacgtac	caccccgcga	gcgagggggg	aggggacacg	4980
gtattcgtgc	caatcggggc	gcccacgcg	aacacgcggg	tgtacgtgtt	ggacgagcag	5040
ttgagggcagg	tggcgggtgg	ggtgatgggg	gagttgtatg	tggacagcgt	gggatggcg	5100
aggggggtatt	ggggccagcc	agcgtgacg	gcggagcgct	tcacgcgaa	cccgtacgcg	5160
agccagcccg	gagcgaggtt	gtaccggacg	ggagacatgg	tgagggtgct	ggcgacgggc	5220
tcgctggagt	acctggggag	gcgagactac	gagataaagg	tgagagggca	ccgggtggac	5280
gtgcggccagg	tggagaaggt	ggcgaacgca	catccagcca	tccgccagcc	ggtgtgtctg	5340
ggatggccgt	tgggctcgag	caacgcgcag	ttggtggcct	acctggtgcc	gcaggcgggc	5400
gcacgggtgg	ggccgcggca	ggtgagggat	tacctggcgg	agtcgctgcc	agcgtacatg	5460
gtgccaacgc	tatacacggt	gttggaggag	ttgccgcggt	tgccgaacgg	gaagctggac	5520
cggctgtcgt	ctgcggagcc	ggacctgtcg	agcagccgag	aggagtacgt	cgcgccccac	5580
ggcgaggctcg	agcggaaagct	ggcggaaatc	ttcggcaacc	tcctggggct	cgaacatgtc	5640
ggcgctccacg	acaactctct	cagcctcggc	gggcactccc	tcctggctgc	ccaggtggtc	5700
tcaaggattg	gcaaggagct	tggcactcag	atctcgatcg	ccgatctgtt	tcaaggcccc	5760
acagttgaac	agctctgtga	gctgattgga	ggactggacg	atcagaccca	gagggagctc	5820
gcctcgtcgc	cgctcgggaa	caccgaggcg	gtgctctcgt	tcgcgcgaaga	gcgcatgtgg	5880
ttcttcgaca	acttcgtcaa	gggcatgccc	tacaacacgc	cagggtctga	ccacctgacg	5940
ggtgagctcg	atgtcgcggc	gctagaaaag	gccatccgcg	cggctcatcg	tcgccacgag	6000
cccctgcgca	cgaatttctg	cgagaaggac	gggggtgctg	cccagttggt	ggggacggaa	6060
gaagccttcc	gcttgcagct	gaactccatc	cgcgacgaga	gcgaggtcgc	gcgctcatg	6120
gaagccgtga	tccaaacgcc	agtcgatctg	gagcgggagt	tgatgatecg	ggcttatctc	6180
taccgggtcg	accgcgggaa	tcactacctg	ttaccacca	tccatcacat	cgccttcgat	6240
ggctgggtcga	catcgatctt	ctaccgtgag	ctggtctcgt	actacgccgc	gtttctccgg	6300
cgcgaagaca	gtccgctcgc	cgcgttgga	atctcctatc	aggactatgc	ccgctgggag	6360
cgggcccatt	tccaggacga	ggtgttggcg	gaaaaactga	ggtactggcg	gcagcggtcg	6420
tcgggcgctc	ggccctctgt	acttccgacc	acctaccatc	ggccgcccat	ccagagtttc	6480
ctgtggcgccg	tctgaactt	cgagatcgat	cgtctcatca	ccgagcggtt	gaagacgctg	6540
ttccgcgttc	cggcgccacc	gatgtacatc	gtgttgcctg	gcgcgttctc	cgtgtgtcgt	6600
cagcgtact	ccggtcagga	gcacatctgc	atcggtctcc	ccgtggcgaa	ccggggtcac	6660
atccagacag	aagggtgat	cggcttgctc	gtcaacaccc	tggtgatgag	ggttgatgcc	6720
gccgggaate	ccggttctat	cgacctgctg	gcgcgcattc	aacggacagc	catcgatgct	6780
tacgcgaacc	aagaagtgcc	cttcgagaag	atcgtggacg	acctgcaggt	cgcgagagac	6840
acggcccgat	ctccgctcgt	gcaggctcatt	ctcaacttcc	acaacacgcc	tcctcaatcc	6900
gagctggaa	tgcagggggg	gacctctacg	cggatgccgg	tgcacaacgg	cacggccaag	6960
ttcgagctct	ccatcgacgt	cgcggagacg	agcggcggtc	taacgggatt	cgtggagtac	7020
gcgacggatg	tgttcagcga	gaacttcate	cggcggatga	tcggccacct	caggttggtg	7080
ctggagcgcg	tcgggtcgca	tcccgggcgg	cctatccatg	agttgccact	gctcacccgg	7140
caggatcagt	tggacctact	gtcgcggagc	ggccacacag	cccccgcggt	ggaacacgtc	7200
gagttgatcc	ctcatacgtt	cgagcggcgc	gtccaggaga	gccctcaagc	gattgccctg	7260
gtctgcgggtg	acgagcgctg	caactactcc	gcgtcaacc	gccggggccag	ccagattgcc	7320
cgcgccttgc	gcgcgcagg	gatcggacgc	gacacctcgc	tcgggctttg	cgcggggcgc	7380
tccatcgagc	tggctctcgg	cgtccttggc	atcttgaagg	cgggcggtgc	gtacttgcca	7440
atcgacccca	cctcctcgcc	cgaggtgate	tacgacgtcc	tgtatgagtc	gaaggtgcgg	7500
catctgttga	cagagtcgcg	cctggctggg	ggactgccgg	tcgatgacca	ggaaatcctg	7560
ctcttgata	ccccgcggga	cgttgaaggg	gacaaggctg	ttgctgaccg	ggaggagcca	7620
cctgaacttg	gcgaggtcag	cctcactccc	gagtgccttg	cgtacgtcaa	cttcacctcc	7680
gactccgggtg	gggcgcgcag	gggcatcgcc	gtccgccatg	gggcgctggc	tcgcgggatg	7740
gccgcggccc	acgcacagta	cctggccaat	tccgccgtac	gtttcctgct	gaaggcgccg	7800
ctcagcttgc	accccgcggt	cgcgggagctg	ttccagtggg	tcgtcagcgg	cggcagcctg	7860
agcactctcg	accccaatgc	cgaccgcgac	gcctctgctt	tcctcgcgca	ggtgcgcagg	7920
gactcgattg	gcgtcctcta	ctcgctcccc	tccgaactct	cgacgttggt	gagccacctg	7980
gagcgcgagc	gtgaaagggg	gcatgagctg	aacacctctc	ggttcatctt	ctcggcgggg	8040
gatacccttg	cggttaccgt	cgtcgagcgt	ctcgggggtac	tggtgcgggc	cggccagctc	8100
ccgctgcggc	tgttcaacgt	ctatgggacg	aaggagacgg	gaatcgcgcg	gggttgcttc	8160
gagtgcgcgc	tggacgcgaa	cgaccccgag	gccgaactcc	cgcggggacg	gctctcgcat	8220
gagcggatgc	ccatcgcgcg	gcccgccag	aacctgtggt	tctatgtggt	gcaacccaac	8280

[0023]

gggtgectgg	ctcgttggg	catccgggg	gaactgtacg	tcggcggcgc	gcaactcgcc	8340
gacgcccgtt	tcggcgacga	gcccacggcg	accaccccgc	gttcgtccc	gaaccccttc	8400
cggagcggag	cggagaagga	ctggctgtac	aagacggggg	acctcgtccg	ctggctgcct	8460
cagggggcgc	tcgagctggt	cagcgcggt	cgggagcgcg	acggagcgcg	ggaccaccgg	8520
ctcgatcgcg	gcttcatcga	ggcgcgcgat	cgtcgtgtgg	ccattgtccg	cgacgccgtg	8580
gtggcctacg	tcceggatcg	ccaggacagg	gcccgggttg	tggcctacgt	cgttctgaag	8640
gagtcgcccc	cggcggacgt	ggagcccgcg	gaagggcggg	aaacgctgaa	ggctcggatc	8700
agcgccgaac	ttgggagcac	gttgccggag	tacatgttcc	cggccgccta	cgtgttcatg	8760
gacagccctg	cgttgacggc	ttacgggagg	atcgaccgga	aagccctgcc	cgagccggag	8820
gatgaccgcc	acggtggtag	tgcgatcgcc	tacgtggccc	cgcgcgggcc	cacggagaag	8880
gcactggcgc	acatttgcca	gcaagtgtcg	aaacgcccc	aggtcggact	gcgagacaac	8940
ttctttgagc	tggcgggcca	ctcagtgggc	gccatccaac	tgggtgccgt	gagccgggag	9000
cacctggagg	tcgaagtccc	cctcagcctg	atcttcgaat	cgcgggtcct	ggaggcgatg	9060
gcgcgcggcg	tcgaagcgct	gcaacagcag	ggcgcgacgc	gcgcgggtgc	gtcgatccat	9120
cgggtggagc	ggaccggacc	gctgctctcg	gcgtacgtgc	aggagaggct	gtggttcgtg	9180
cacgagcaca	tgaaggagca	gcggaccagc	tataacatca	cctggacgtt	gcacttcgcc	9240
ggcaagggtt	tctcgttgga	ggcgttgcgg	acggccttcg	atgagctggt	ggccagacac	9300
gagacactgc	ggctgttggt	ccaggtgggg	gaggggacag	agcaggccgt	acaggtcatc	9360
ggggagccct	ggtcgtatga	gctgccgtcg	agagaggtgg	cggggacgga	ggtgacggcg	9420
gcaatcaatg	agatgtcccc	gcaggtcttc	gacttgagag	cgggacgggt	gctgacggcg	9480
gcggtctcga	gggtggcgga	ggatgagcac	atcctcgtca	gcaacatcca	ccacatcacc	9540
acggacggct	ggctgttcgg	ggatgtgctg	cgggagctga	gggagttgta	cgaggcccg	9600
gtcggggggg	agcgagcgga	gctgcccgcc	ctgacgggtc	agtacggcga	ctatcggtta	9660
tggcagagga	agcaggacct	gagcgagcac	ctggcgtact	ggaaggggaa	ggtggagggg	9720
gacgaggacg	gggtggagct	gccgtacgac	ttcccgcgga	cgtcgaatag	ggcgtggaga	9780
gcggcgccgc	tccagtatag	ctaccaccac	gagctggcga	ggaaggtggc	ggagctcagc	9840
cgggagcagc	agtcacacgt	gttcatgagc	ctgggtggcg	gcctggcggt	ggtgttgaa	9900
cggtagacgg	gccgcgagga	cctgtgcacc	gggacgacgg	tggcggggcg	agcgaggtg	9960
gaactggaga	gcctcatcgg	gttcttcacc	aacatcctcc	cgtgagggct	ggacctgtcg	10020
ggcgctccga	gccttcacga	gggtgtcgcg	aggacgaagg	tgggtgtgct	ggagggattc	10080
gagcaccagg	agcttgcgtt	cgagcacctg	ctgaaggcgc	tgaggcgcca	gcgggacagc	10140
agccagattc	ccttggtgcc	agtgggtgtg	aggcaccaga	acttcccgat	ggcgctctcg	10200
gagggcttga	gtgagggggt	ggagctgaag	aagttcgagc	tggcggggga	aaggacgacg	10260
gcgagcgagc	aggactggca	gttcttcggg	gacgggtcct	cgttgagctg	gagcctggag	10320
tacgcgccgc	agctgttcag	cgagaagacg	gtgaggagga	tgggtggagca	ccaccaacga	10380
gtgctggagg	cgttggtgga	ggggctggag	gaggggtgc	acgaagtgcg	gctgctgacg	10440
gaggaggagg	aggggctgca	cgggaggttg	aacgacacgg	cgcgagagct	ggaggagcgc	10500
tggagcctgg	cggagacgtt	cgagcgtcag	gtgagggaga	caccggaggc	ggtggtttgc	10560
gttggcgtgg	aggttgcgac	gggagggcac	tcgcggcgca	cataccggca	gctgacatac	10620
cggcagctga	atgcgcgagc	caaccaggtg	gcacggaggc	tgagggcact	gggagtgggc	10680
gcggagacac	gggtcgcggt	cttgagcgac	cgtcgcgcgg	agctgctggt	ggcgatgctg	10740
gcgatattca	aggeccgggg	ctgctacgtg	cgggtggacc	cacagtaccc	gggaagctac	10800
atcgagcaga	tacttgaggga	tgcggcacgc	caggttggtg	tgggcaagag	gggaagagcg	10860
gacgggggtg	gggttgatgt	gtggctggag	ctggatggag	cgaacggct	gacggacgag	10920
gccttggcgg	cacaggaaga	gggagagctg	gagggggcgg	agaggccgga	gagccagcag	10980
ttggcgtgtt	tgatgtacac	gtcgggtccc	acgggcagac	cgaagggggg	gatggtgccg	11040
tacagccagt	tgcacaactg	gctggaggcg	gggaaggagc	gctcgcgct	cgagcgtggg	11100
gaagtaatgt	tgcagaagac	ggcaatcgcg	ttcgcgggtg	cgggtgaagga	gctgctgagc	11160
ggattgtcgg	cgggagtgge	gcaggtgatg	gtgccggaga	cgttggtgaa	ggacagcgtg	11220
gccttggcgc	aggagataga	gcggtggcgg	gtgacgagaa	tcacctgggt	gccatcgcac	11280
ctgggagcac	tgttgagggg	ggcgggggaa	gaggcgaagg	ggctgaggtc	gctgaagtac	11340
gtcataacgg	cgggggaggg	actggcgacg	ggggtgaggg	aggaggcgag	gaggaaagctg	11400
ccgggggcgc	agttgtggaa	caactacggg	tgcacggagc	tgaatgacgt	gacgtaccac	11460
cccgcgagcg	agggggggagg	ggacacggta	ttcgtgccaa	tcggcgggcc	catcgcgaac	11520
acgcggtgtg	acgtgttgga	cgagcagttg	aggcgggtgc	cgggtggggg	gatgggggag	11580
ttgtatgtgg	acagcgtggg	gatggcgagg	gggtattggg	gccagccagc	gctgacggcg	11640
gagcgcttca	tcgcgaaccc	gtacgcgagc	cagcccggag	cgaggttgta	ccggacggga	11700
gacatggtga	gggtgtggtg	ggacgggtcg	ctggagtacc	tggggaggcg	agactacgag	11760
ataaaggtga	gagggcaccg	ggtggacgtg	cgcaggttgg	agaaggtggc	gaacgcgcct	11820
ccagccatcc	gccaggcggt	ggtgtcggga	tggcgggttg	gctcgagcaa	cgcgcagttg	11880
gtggcctacc	tgggtgccga	ggcgggcgcg	acggtggggc	cgcggcaggt	gagggattac	11940
ctggcggagt	cgttgccagc	gtacatgggt	ccaacgctat	acacggtgtt	ggaggagttg	12000
cccggttgc	cgaacgggaa	gctggaccgg	ctgtcgttgc	cggagccgga	cctgtcgagc	12060
agccgagagg	agtacgtcgc	gccccacggc	gaggtcgagc	ggaagctggc	ggaaatcttc	12120
ggcaacctcc	tggggctcga	acatgtcggc	gtccacgaca	acttcttcaa	cctcggcggg	12180
cactccctcc	tggcttccca	gctgatttcg	cgcatacggg	cgaccttccg	cgtggaagtg	12240
gcgatggcca	cgggtgttcga	gtccccacg	gtggagccgc	tcgcccgcga	catcgaggag	12300
aagctcaagg	acgagctctg	ggtccagctc	tccaacgttg	tgcgggtcga	gcggacgcag	12360
gagcttccgc	tctcctacct	gcaggagagg	ctgtggttcg	tgcacgagca	catgaaggag	12420
cagcggacca	gctataacgg	aacgacggg	ctccggcttc	ggggtcctct	gtcaatcccc	12480
gcgctcaggg	ccaccttcca	cgtcttggtc	gcccgtcacg	agagcctgcg	caccgtcttc	12540

```

cggttccccg aaggccgcac cagcccggtg caggtgatgc ttgatccgat ggatctggac 12600
atcccggtcc gcatgcaaac cagggccgac atcatcccg gcatggatga gctggcgggt 12660
cacatctacg acatggagaa gggtcgcgtg tcatgtgttc gcctcttgcg gctggccgag 12720
gactccacag ttctctgat ggggatgcac cacatcgtct acgacgcagt gtcacagttc 12780
aatgtgatga gtccgatat caacctgctc tactcggcgc acgtgacggg aatcgaggca 12840
cggttccccg cgttcccat ccagtacgcc gacttctcgg ttggcagcgc ccagcaggac 12900
ttcgtcacc acctggacta ctggaagtcc acactggcgc actaccggga tgatctcgag 12960
ctgccgatg actaccgcgc gcccccage cggacatggc acgcgaccgc attcaccttc 13020
cggtatccgg atgcactggc gcgcgcgttc gccaggttca atcagtccca tcagtcgacg 13080
ctgttcatgg ggctgtgac cagcttcggt atcgtgtca ggactacac cggccggaac 13140
gacatctgca tcggaacgac aacggcgggg gcgcgccagt tggagttgga gaacctcgtt 13200
ggcttcttca tcaacatcct gccgttgcgc atcaatctgg cgggtgacc cgcacatcagc 13260
gagctcatga atcgagcga gaagagcgtc ttgggcgcct tcgagcatca agctctgccg 13320
ttcgagcgtc tccctcagtgc cctcaacaaa cagcgtgaca gcagccatat cccgttggtt 13380
cccgatcatg tgcgccacca gaacttcccg acggcgatga ccggcaagtg ggccgatggt 13440
gtggacatgg aggtcatcga gcgcgacgag cgcacgacgc ccaacgagct ggacctccag 13500
ttctttggcg acgacacctt cttgcatgct gtctcagagt tccccgcga gctcttctcc 13560
gaggtgaccg tccgctgct gatgcagcgt caccagaagg tcatagagtt catgtgcgcg 13620
acgttggggg ctcggtga                                     13638

```

```

<210> 13
<211> 3072
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(3072)
<223> CysL

```

```

<400> 13
gtgaacgtgc tegttaggca ttccaccggc tcccacgacg agccgggtggc cggcgacgtc 60
gaactccgcg tccgtggccc cgggtgtgcc gacgtctcatt ccagcgagag cgttgaagtg 120
ctggcgcggg ggctgcggac cgcgcgagg agtaccgcg gcgtcatggg cccgacccgc 180
caggaggcgc cctggttcgc catcccgttg acctgccgc gcggtgcccg gtcggcgcga 240
ttcgacctct ggctcgggga actagaccgt cagggacagc tctccacat ggctgcctcg 300
tatctggcgg ccgtgcacca cgtgctgtgc agcgttcgcg agccagcgc caacgtgtcg 360
gaggtgtcgg tctctgactc aacaacgcca tctgggtcga accggttctt gaacggcctg 420
gactccgctc tggagatcct ggctcacggg cgcagcgacc tctcctcga gcattctacg 480
ggccggctgc ccccgcagca gatgcccttc gtggaggacc gtgaggagcg caggagcac 540
ccggccaccg atgtcgagge cgtgcccgtt gtctccgtcc tgttccaacc agttgacttc 600
ccgagcctgg cagaggtgga cgcgagcctc ctgcgcatg acgacgagga tgccggcgcg 660
ctgggcgggg tctgggggga gctcctccag ccgttctgc tcgactccgc caggatgacc 720
gtggggcgaa agggcggtg ggctcgatcac atctgctgc ctggcttgtt gcgagccgac 780
agcagagcgg cggaggagtc ggttctcgcg cccgccttgc gcttggcgac gaagcccggt 840
cgccatttgc tccggttgtg ccggaacacc gccctgcggc tgggagacag gctgcccac 900
ttgctgcgc agggcccgtc ctgcgatggc gcgtcaacgg cgtcctctt gttgcaacgg 960
gtgctggaca cgttatcgg gagcggggga ctgaaggacc atgcctcac gctcgagctg 1020
gttggccgag atccacggac caggcccgcg ttctggggcc ggactccgtg gctggtggcg 1080
gaacgggcgg cttcggctgc atcaacggat gcaccgcgcg tcgacgtcgt cgtctgttc 1140
ccggcggcac ggccgagcgc gctcgagctg cggccagaca gcgtcgtcat cgacctttt 1200
ggcaccctga cgttgagacc gcgaccgag gttctggcga agaacatcgt ctacgtgcga 1260
ggggcctcgg tccgtctcgc cggagaggcc gtctctcga ctccctcctt cgcgccggat 1320
cgagtggagc cggcgtcctt caggcgctt ctccgggaac tcgacgcgga ggccagtagt 1380
gacgggctcg cccacagaca ccgccttgag attggcgga ttcgcggtt ctggggtgag 1440
atccgccggg cggagtggga cgccttcat tcgcgccgc ggggggagct ggcgaggttt 1500
caggtgtcgg ggcaggtgac cgcgcgaat ccgggctcgc ccagcctgcc cgtggtggcg 1560
acgaacatct ggaatacat cttccgggaa gcgcaccttc gctcggctc gtgctcgtc 1620
gacccacaga gggccagtc cgcgacctac gccagctgc ggcgactggc ggcagcgtac 1680
gcgcggcggt ttctggcatt ggggctcgc caggagagcg tcgtggcgct cgcggcgccg 1740
gatgggattt cgtccgtcgc ggtgatgctg ggttgccttc tggcggggtg ggtcttcgcg 1800
ccgtcaacc acaccgcctc ggccgtgaac ttcgaggcga tgttgagttc cgcagtcctc 1860
cgcttggtgc tccatgcgcg gtgcaggttc gcccgccatc tcccggtcct gagcacgcgg 1920
cgatgcgcgg aactcgcgtc cttcctgccg ccggacgcgc tggacggcgt ggagggggac 1980
gtcaccccc tgccagtgc accggaagcc cccgccgta tctgtttcac ctgggctcc 2040
acgggggggc cgaaggcagt gacgcacacc caccgcgact tcatcacctg cagtcgaac 2100
tacgcacct atgtctcga actcagaccg gacgatcgt tctatacgcc gtccccgacc 2160
ttcttcgctc atgattgaa caacttctgc ctgtccctca gcgcgggggc cagcacgtg 2220
atctcgctc ctcgcaacgg cgggatgggt gtgcgggaga tctcgcgcg gaacgaagta 2280
accgtgctct tcgcggttcc cgcgctctat aagctgatca tctcgaagaa cgaccggggc 2340
ctcggttgc cgaagttag attgtgcatc tctgctggcg agaagctgcc attgaagctg 2400

```

[0024]

	tatcgggagg	cgcaagctt	cttcagcgtg	aacgtactgg	acgggatcgg	gtgcaccgaa	2460
	gccatctega	cgttcacetc	gaaccgggag	agttatgtcg	cgeccgggtg	cacgggcgtg	2520
	gtggtcccg	ggttcgaggt	caagctgggt	aaccgcgtg	gcgagctctg	ccgggtggga	2580
	gaggtggcg	tectctgggt	tcgggtggg	gcgtgaccc	ggggtacgt	gaacgcccc	2640
	gattcgacag	agaagcactt	cgtggacggc	tggttcaaca	cccaggacat	gttcttcattg	2700
	gatgccgagt	accggtctta	caacgtgggc	agggttggtt	cggtcacaa	gatcaattcc	2760
	tgttggttct	caccggagat	gatggagtgc	gtcctgcaat	cccattccagc	ggtgaaggag	2820
	tgtgccgtct	gcgtcgtcat	tgacgactac	gggttgccaa	ggccgaaggc	attcatcgtc	2880
	accggcgagc	atgagcgtc	cgagccggag	ctcgagcact	tgtgggccga	gttgccgcgtt	2940
	ctgtcgaaag	agaagcttgg	gaaggaccac	taccgcgcatc	tgttcgcgac	catcaaaacg	3000
	cttccccgga	cttcagcgg	gaagctgatg	cggtcggaac	tcgcgaagct	gtccaccagc	3060
	gggcccccat	ga					3072
<210>	14						
<211>	117						
<212>	DNA						
<213>	深棕色孢囊杆菌						
<220>							
<221>	misc_feature						
<222>	(1).. (117)						
<223>	CysM						
<400>	14						
	atgaatccaa	agttctctgg	aggcctgggg	gcagggggtgt	gcategcctc	tttgttccag	60
	acggtcattgc	ggaccgtgcc	gctcaaggac	gccggctccg	gcgacagggc	ttgttag	117
<210>	15						
<211>	1074						
<212>	DNA						
<213>	深棕色孢囊杆菌						
<220>							
<221>	misc_feature						
<222>	(1).. (1074)						
<223>	CysN						
<400>	15						
	atgtcgactc	gcaccaagaa	cttcaatgtc	atgggaatcg	actggatgcc	ttcttccgcg	60
	gagtccaage	gacgcgtccc	gcggaccag	cgggcggcag	aggccgtgct	cgcgggacgg	120
	agatgcttga	tggatactct	ggaccgcggg	gatcctcgcc	tcttcgtcat	cgtggggccc	180
	tgtctccattc	acgateccgt	ggcggggctg	gactatgcga	agcggtgcg	gaaactcgt	240
	gatgaggttc	gcgagaccct	gttcgttggtg	atgcgcgtgt	acttcgaaaa	gccgcgcacc	300
	accacgggtt	ggaaaggett	catcaatgac	ccgcgcattg	atggtctttt	ccacatcgag	360
	gagggcattg	agcggggacg	tcgttctctg	ctcgacgtgg	ccgaggaggg	tctaccctct	420
	gccaccgagg	cgttggaccc	categcgtcg	cagtactacg	gcgacctcat	ttcttggacg	480
	gcatttgccg	cgcgcaccgc	cgagtgcgag	acgcaccgcg	agatggcgtc	cggcctttcc	540
	accccatgag	gcttcaagaa	cggcacggac	ggctcgtctg	atgcggccgt	caatggcatc	600
	atctccgctt	cacaccgcga	cagtttctctg	ggggtgagcg	aaaatggcgc	gtgcgccatc	660
	atcgcgcgc	ggggcaacac	ctacggccac	ctgggtctgc	gcggcggtgg	tgggcggccc	720
	aactacgacg	cgtgtctggt	ggcgttgcg	gagaaggcgc	ttgccaaggc	caggtaccgc	780
	accaacatcg	tgggtggactg	ctctcacgcc	aactcctgga	agaatcccga	gctccagccg	840
	ctggtgatgc	gggacgtggt	gcaccagatt	cgcgagggca	accgctcggt	ggtgggcctg	900
	atgatcgaga	gcttcacga	ggcaggcaac	cagcccatcc	cggcggacct	gtcgcaactg	960
	cgtacggct	gctcggtcac	tgatgcattg	gtggactgga	agaccaccga	gaagatgctg	1020
	tacagcgcg	acgaggagct	gctccacatt	ctgccccgta	gcaaggtggc	ttga	1074
<210>	16						
<211>	612						
<212>	DNA						
<213>	深棕色孢囊杆菌						
<220>							
<221>	misc_feature						
<222>	(1).. (612)						
<223>	CysO						
<400>	16						
	atcccccccc	gtccactccc	ctctctggaa	agtggcgact	ttttcgccga	cgtaacgttt	60
	tctgatctct	cgtacgagtc	ggctgatctc	tccggcaagg	aattcgagcg	ctgcacgttc	120
	cggcggttga	agttgccga	aagccgtgg	gtccggagcc	gcctggagga	tttgtatttc	180

	gagggaatgcg atctcctgcg gatgggtaccg gagaagctcg cgctgcgaag cgtgaccttc	240
	aaagacaccc gectcatggg cgtggactgg agtggactcg gaaccatgcc ggacgtccag	300
	ttcgaacagt gcgatctgcg ctacagctcc ttcttgaagt tgaatctacg caagacgcgg	360
	ttcgtttggt gctccgcgcg cgaagccaac ttcattgacg tggacctcgc cgagtcggac	420
	ttcacccgca ccgatatgcc aggatgcacc atgcagggtc gcgtcctcac caagaccaat	480
	tttgcctgat cgaccaatit catcttcgac ccgaaggcga accaggtcaa agggacgcgt	540
	gttggcgtgg agaccgcctg cgcctcgcgc caggcgttgg gaatggtggt cgacggctat	600
	cagacaccct ga	612
<210>	17	
<211>	702	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1).. (702)	
<223>	CysP	
<400>	17	
	atgaaacggt tcttcaagct ccagttgcgc accaccaacg tccccgcggc acgggcgttc	60
	tacacggctc tgttcgggtga gggcgcgcgc aacgcagaca tcgtgccgct gcccgagcag	120
	gcgattgccc gggcgcgcgc cgcctattgg ctgggttacg tcggcgtcga ggacgtcgat	180
	gaagcgggtg gctcgttctg gggcgcgcgc gcgacccagc tcggcccgac ccacccgacg	240
	aacgacggcg ggcgcgtcgc gatcctcgcg gatcctggag gggcgacctt cgcctggcg	300
	acggcaccgg caacgacgag agcgtccag cggaggtgg tctggcagca gctctatgcc	360
	gcgaacgtgc aacgacggcg cgcctcgtat tgcgacctgt tcggatggcg gctctcgat	420
	cgcgcgcgac ttggtgcgct gggggttcac caggagtcca cctggcgtc ggacgagccg	480
	agcgcgggct cggtcgttga cgtggcggcg ctcaaggggg tccattcaca ctggctgttc	540
	catttcgcg tcgccgcgct cgtacccgcg atggaggtcg tccgcaaggc cggagggcgtc	600
	gtcatcgccc ccattggaact tccgaatggc gatcgcatcg ccgtgtgcga ggatccgcaa	660
	cgggcggcgt tcgcgttcg cgaatccagc caccgacgct ga	702
<210>	18	
<211>	795	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1).. (795)	
<223>	CysQ	
<400>	18	
	atgcaagaga tcggccagac ggcactttgg gtggcgggaa tgcgcgcgct tgagaccgag	60
	cgttccaacc cactgttccg ggatcccttt gcccgctgac tcgccggtga caccctcgtc	120
	gaggagctgc ggcgcgcgcaa tgcgggtgag ggcgccatgc ctcccgccat cgaggttcgc	180
	acgcgttggc tcgatgatca gatcacgtg ggggttgggc gcggcatccg ccagatcgtc	240
	atcctcgccg cgggaatgga tgcgcgcgcc taccgtttgg cctggccggg agacacgcgg	300
	ctgttcgagc tcgaccagca cgcctgctc caggacaagg aggcgaagct gaccggcgtc	360
	gcgccgaat gtgagcgaca tgcctgtctg gtcgatctgg ccgatgactg gccggcggcg	420
	ctgaagaaaa gcggattcga tcccgcgctg cccaccctgt ggctcctega gggattgctc	480
	gtctacctca ccgaggcgca ggtcacgtg ctcatggccc gtgtcaacgc cctgagcgtt	540
	cccagagca tcgtcctcat cgacgtcgtt ggccgttcga ttttggactc ctgcgcgtc	600
	aagtgtatgc acgacctcgc ccgccagttc ggcaccgacg agcccaggtt gattctaagg	660
	ccgattggct gggaccccca cgtctacacc accgcggcca tcgggaagca gctcgggcgc	720
	tggcccttcc ccgtggcgcc acgcggcacc cccggtgtgc cccagggata cctggtgcac	780
	ggagtcaagc gctga	795
<210>	19	
<211>	1002	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1).. (1002)	
<223>	CysR	
<400>	19	
	gtgaatggga cgacagggaa gacagggttg gtagcagaaa ggctcgggcg gatttccccg	60

```

agggaactaca agtccaagga gttggtgtgg gattcgettg ccgccacacg cagcaagccc 120
cggcgcgtac tgcggagggg ggacgtggtc gggcacctgt acccgccggc caaggcgccc 180
ctgtcaccce acccgctcat gaagaacctt ccgcccgaga cgtgcggct gttcttcac 240
cactccgctt acaagttcat gggggacatc gccatcttcg agacggagac cgtcaacgag 300
gtggcgatga agatcgccaa cggtcacacg cccatcacgt tccgggacga catcgccac 360
gacgcgctca cgtcatcat cgatgaggcc tatcacgctt acgtggcacg cgacttcattg 420
cggcagatcg agcagcgcac gggcgtcaag ccgctgcccc tgggaacgga aacggacctg 480
tcaggggcca tggctttcgg caagcacccg ctgcccgaga cgtgcacggg gctctgggaa 540
atcatcgccg tctgcatcgg ggaaaacaca ctcaccaagg atctgctgaa cctgacgggt 600
gagaagtcct tcaacgaagt gctccatcag gtgatggagg accatgttcg cgacgagggc 660
cgccacgcgg tctcttcat gaacgtgctc aagctgggtt ggagttagat ggaggagagc 720
gcccggctcg ccatcggtca gctgctgcca gagttcatcc gcgagtacct cagcccgaag 780
atgatggcgg agtacgagcg cgtcgtgctg gacgagctcg gtctageggc cgagcacatc 840
gagcggatcc tctccgagac gtactcggag ccgcgctgg aggatcttcg cgcgcgatat 900
ccccctccg ggtacctggt ctacgtgctg atgcagtgcg acgtcctgtc gcacgcgccg 960
acgcgcgagg cgttcgccg attcaagctg ctgccccact ga 1002

```

```

<210> 20
<211> 1929
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(1929)
<223> CysS

```

[0027]

```

<400> 20
atggccaacc agcgggtcgc attcattgag ttgacggctt tctctggcgt ttatcccttg 60
gcctctggct acatcgctgg cgtggccgag cagaaccctt tgatcaggga gtcgtgcagc 120
ttcgaaatcc actcgatctg catcaacgac gaccgattcg aagacaagct caacaagatc 180
gatgcgatg tctacgcgat ctcttgctat gtctggaaca tgggcttcgt gaagcgggtg 240
ctccccacc tcaccgcccg caagcccaac gcgcacatca tcttggcggg tccgcagggtg 300
atgaaccacg gggcgagta cctggatccg ggcaacgagc gggtggtgct ctgcaacggt 360
gagggtgagt ataccttcgc gaactacctg gccgaactct gctcccccca gcccgacctt 420
ggcaaggtca agggcctctc cttctaccgg aacggagagc tgatcacgac cgagccccaa 480
gcgcgcattc aggatctgaa cacgggtcca tctccctacc tgggaaggcta cttcgacagc 540
gagaagtacg tgtgggcgce ccttgagacg aaccggggat gccctacca gtgcacctac 600
tgtcttggg gggcggcgac caactcgcgc gtgttcaagt ccgacatgga cgggtcaag 660
gaggagatca cctggctcag ccagcacccg gcgttttaca ttttcatac cgacgcgaat 720
ttcggcattc tgaccgcga cattgagatc gccagcaca tcgccgagtg caagcgggag 780
tatggtatc cgtcaccat ttggctgagc gggcggaaga actcgctga ccgggtcag 840
cagatcacgc ggatcctgag ccaggagggt ttgatctcca ccagccggt ctcgctccag 900
acgatggacg cgaacacgct gaagagcgtg aagcgcggca acatcaagga gaggcctac 960
ctgagcctcc aggaagaact gcaccgcagc aagctctcct cgttcgtgga gatgatctgg 1020
ccgttcccc gcgagacgct ggagaccttc agggagggca tcgggaagct ctgcagctac 1080
gagcgcgagc cgatcctcat ccaccacctc ctgctcatca acaacgtgcc gatgaacagc 1140
cagcgcgagg agttcaagct ggagggtgct aatgatgaag acccgaaacg cgaggcgag 1200
gtcgtcgtcg cgacgaagga cgttaccgcg gaggaatata aggagggtgt gcggttcggg 1260
tatcatctca cgagcctgta cagcctgcgc gactccgct tcgtcgggag gtacctgac 1320
aagcaggggc ggctggcctt caaggacttg atctcctcgt tctcggagta ctgcaagcgg 1380
aaccttgacc accctacac gcagtacatc accagcgtga tcgacgggac cagccagtcg 1440
aagttcagcg ccaacggcgg catcttccac gtcacacttc acgagttccg cagagagttc 1500
gaccaactgc tcttcgggtt cattcaaac ctgggcattg tgaacgatga gctgctggag 1560
ttctgttcg agatggtatc cctcaaccgt ccgcacgtgt acagcaaac gcccatcaat 1620
aatggcgagg ggttgetgaa acacgtgacg gtcgtctcga aggagaagga tgccattgtc 1680
ctgcgcgttc ccgaaaagta cgcgcagctc acgtctgagc tactcgggct cgaggcgct 1740
cccagcacga gctgcgcgt gaagtaccgc gggactcaaa tgcggttcac ggcgaaacag 1800
ccgtacgagg acaacctctc ctactgcgag gcgaagctcc acaagatggg aagcatactt 1860
ccggtctggg agtcggcgtt ccttcgcgc acaccggtcc ggcggccaca agtggccgtc 1920
gggggtga 1929

```

```

<210> 21
<211> 3804
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(3804)
<223> CysT

```



<400> 21	
atgcatcgag tgaagccgtt gatagggccc gtctgtgctg cgtgtgtgct gtgtgccctg	60
cccgccaggg cgcagatcgc cgcggccccc gtctaccaca accacatgcc caactctctg	120
gcctactacg acctggggcca ataccgctcc acgcccaccg gcggccccat ccggtacatg	180
tatgacgcgc aggtcatcaa cctgaagaag aatcccccgt ccaattacac atactacctg	240
ccatcgggcg cgcceatgcc gcacgatgac ctctcaactt attactcgca caacgcgaag	300
acgggtgctt acctgtactg gctccaagc gtgcctctgg acatgaaaac caatgcccc	360
accggccagg tgcacgtcac catgtccggc gccgtgggtg acaatgtcca ggatctctgt	420
accctgaaga acgtcccccgt ctacgacaat ccgaactggg gcgcctctct gaaggaccgc	480
tacagcgcgc tgcaccccc cgcgggcaac cgcaccctgg attctatcca cttcaccggc	540
caccactcca tggggccccct ggctgggtccc gactacttcc tcaaggatct catctaccag	600
agcgccacgc tcgcccagcc ctacttctct ggccgctcct tcacgtctct caaggccttc	660
ttccccaccg agcteggtct ctccgagcgc ctatccccca cctctcccaa gctcggcggt	720
cagtggggcg tcctcgagca caaccattt tcccgaccc tcaaggacta cccctacctc	780
aacgatccgg gctccgacac gctcgtctcc ccgcccaacc gcgcgatct ccagaacacc	840
agctccgtgg gctcctgggt gaggcccccag atggcccacg agcagcaggt catcaagaac	900
aagtaccctt tcgctccac tccccactgg gtgcgtacg tggaccccc caccggcgcc	960
gagtcgcgcg tcgtcggcat ccccgtaaac cagaacggct cctggctcga gggtgggaa	1020
ggcgaggcca ccgtcgacgt cgtcaacctc aagagcttgc agggcctcgt tccccagcgg	1080
cagtctctcg tcactcgcca tgatggcgac aactcgagcg gacgcgcgg ctccgactcc	1140
acctggtaca accgcccctc cgtcacctgc gccaatggcg tgcagtgcgt gggcatctcc	1200
gagtaacctg tccaccacac ccccgcctcc accgacgtgg tgcacgtcca ggacggctcg	1260
tgggtgggca cgcgcgactc ctctctgggt cccagtggg accactggaa gctcccttc	1320
ggcatctgga agggctcagtt ccccgcttc aacgcgcga ccggcctcaa tctctctccc	1380
aagacgaacc tcagcggcgt gcaggagggc atgacgggtc cctcgagca cggctggcac	1440
tacctcgagc gcaacttcgc cctgctccag ccgcacctca actacgcga gaccgcgag	1500
catagctggc tcgaacgcga cccaatcac tggtcgccc ccaccgcgat cgacaagcag	1560
atcacccaag cgggcaacca gctcaaccg tggatgatgt cctttccgt caaggcgac	1620
gtgaacaacg actggggggg ggccgccaac cccgcggaac tcgcttggtt ctctctgtg	1680
cccgccatgg actcgggctt cggtacttac gacgagaacc aggacgacaa cgtcaagccc	1740
acgtgtctt tcaatcaatc cctctacttc tccaagcct acgtgcagca gcgcactgce	1800
caggacaaga aggcgccctc cgtctgggtg gccacgcgt ggccctacaa ccccgccgc	1860
gccaaacacc acaagtcgga gggttggaag ctccacttct tcaacaacca cttegcctc	1920
tacacctacg cctacgacgc gacgggcact tcttccatca agggcccggt ccgggtgcac	1980
accacaaga gcatcgaccc gctcgacaa accacaagg tctatgatec ggccggcgcg	2040
aaggcccgcg gtgttcccaa categatccg gcccgctggg gcgcctgggt ggactaccg	2100
ctcacccgcg cgcacctgaa gcctgtcatg aatggtgttt cctggcagcc cgcctacctg	2160
cccgctatgg ccaaggtgcc cgcgcaggag atcgccgacc tctactacgt ctacctgggc	2220
aactaccgcg accagctcct cgaactactac atcgaggcca ccgacagccg gggcaacatc	2280
accgggggag agatccagtc cgtctactgt ggctcgggce ggtacaacct ggtgggcggc	2340
aagtacatcg aggaccccaa cggcacggta cagggaacgc atcccttct cgtgggtggac	2400
accaccgcgc cctcgggtccc ctccggactg accgcgaagg cgaagacgga ccgctcgggtg	2460
acctgagct ggagcgcggc ctccgacaaac gtggcggtga gcggctatga cgtcttccgc	2520
gatggcacgc aggtgggctc gacacccagc accgctata ccgacagcgg cctctccccg	2580
agcaactaat acagctacac cgtgcgcgcc cgggacgcgg cgggcaacgc gctcccccag	2640
agcaccgcgc tgagcgtcgc caccctgacg ccggacaaca cccaccctc gttccctcg	2700
ggcctgacgg cgtcgggcac gacgagctcc tcggtggccc tcgcttggaac ggctccacc	2760
gacaactacg gcgtcgcgaa ctacgaggtg ctccgaaacg gcacccaggt cgcgtccgtc	2820
acggggacga cctactcgga taccggcctc tcgccgagca ccacctacag ctacaccgtg	2880
cgcgcgccgg acccgccggg caatgtctcc tcgccagca cggccctgtc cgtcaccacc	2940
cagacgggca acagcgccac cgtctactat ttcaacaaca acttegcctt caaatacate	3000
cacttccgca tcggcggtgg cacgtggacg accgtgcccg gcaacgtcat ggccacctcc	3060
gaggtgccgg gtaacgcca atacaccgtc aatctgggag cggccacca gctcagtggt	3120
gtcttcaacg atggcaaggg cacctgggac aacaacaagg gcaacaacta cctcctgcc	3180
ggggcaccct ccacggtgaa ggacggcgct gtctccagcg gacgcgccgc gctcgacacc	3240
accgaccctt ccgtccctc ggccctccac gcggcgctca agacgtctct ctccgtgtcg	3300
ctctcctgga gcgcctccac ggatgccagc ggcatcgccg gatatgacgt gtaccgcgat	3360
gctcgtctgg tgggtccacc cgtctccacc agctacacc acagcagctt gactgccggc	3420
acgacctacc gctacaccgt gcgcgcgcgc gacaccggcg gcaatgcctc cgcaccagac	3480
accgacctga gcgtaccac gagcaactcc tcggccaact ccgtcactt caacgtgacg	3540
gccagaccg tcgtgggaca gaacgtctac ctctgggta accatgccgc gctcggcaac	3600
tgaacaccg gcgcgcccat cctcctgtct ccggccagct acccggaagt gacggtgacg	3660
ctcagctgc cggctcgac ggccctcgaa tacaagtaca tcaagaagga cggctccggg	3720
aacgtacctt gggagagcgg gcgcaaccgc tcgaccaaga tccccgcctc ggggaccgcg	3780
acctcaacg acacctgaa gtag	3804
<210> 22	
<211> 831	
<212> DNA	
<213> 深棕色孢囊杆菌	

<220>		
<221>	misc_feature	
<222>	(1)..(831)	
<223>	ORF1	
<400>	22	
gtgccacatc	catccgagca	gagcgcctccg tcgggactcc gggcgcggct gcacgaaatc 60
atcttcgagt	cggacacccc	ggcggggccgc gccttcgatg tggcattgct gtgggccatc 120
gtgctcagcg	tcctcgcggt	gatgctcgag agcgtggagt ccatcagcgt ccagcatggg 180
cagaccatcc	gcgtcctcga	gtgggtgttc accgggctct tcacactgga gtacgtgctg 240
cggtctgtgt	cggtgaaacg	gccgctgcgc tatgcctga gctttctcgg gtggtggat 300
ctgctggcca	tcctgccctc	ggtgctgagc ttgatgctgc ccggcatgca gtccctgctg 360
gtggtgcggg	tggtccgcct	gctgcgcgtc ttccgcgtac tcaagctgc cagcttctc 420
ggggaggcgg	acgtgctgct	caccgcgtc cgggccagtc ggcggaagat catcgtcttc 480
ctcggggcgg	tgtgagcac	ggctgctcgc atgggcggg tgatgtacat ggtggagggg 540
cgcgccaacg	gcttcgacag	catccgcgg gggatgtatt gggccatcgt gacgatgacc 600
acggtgggct	acggagacct	ctcgcccaag acggtgcccg gacagttcat cgcctcgggtg 660
ttgatgatca	tgggctacgg	catcctcgcg gtgcccacgg gcacgtgtc cgtggagctc 720
gccaggcgca	cccggcagca	cgccatcgac ccgcgcgct gtcccggctg cggcctgcag 780
ggccacgacc	tggacgcgca	ccactgcaag cactgcggca ccgcctctg a 831
<210>	23	
<211>	237	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1)..(237)	
<223>	ORF2	
<400>	23	
atggcacagg	accaggacag	ggagaagctg cattccgacg cggacaagga gaggtgcac 60
ccgaaggtcg	actcgggtga	cgctctcgggc cggggccgcg agcggcggcc cgacaggagg 120
taccccaage	agcgaacgc	gggcgagttc ggcacccacg gaggccca caaggcgggc 180
aaggaagacc	ggcggaact	gcatgcccc ggcagctcca aggcgggctc ccagtag 237
<210>	24	
<211>	489	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1)..(489)	
<223>	ORF3	
<400>	24	
atgggaagaa	cctacagttt	cgaacccttc ttgtcgcagc aaccgcgcga gacctacaag 60
ggctcgggtc	cccggtcgg	caatgaagaa cacaagatcg ccctaccaa ggaagaggag 120
aaggcggccc	tgcctgacac	gccacccggc tatggacagg cccacgccga gaccgtgaag 180
cgtaaccgcg	cccgcgcgga	gaagaagcgc acggagccca agacccccgc tacccgggcg 240
aagaaggccg	cccccaagc	gaagcccacc cggaaggtgg cgacgcaaga ggccaccgcc 300
aaggcccta	cccgtaagc	gcgggaggag accgagccga agggccccgc gcgcaagaag 360
ctgagcgcca	cggggtcgt	gggtagcctc gggcgcaagg ttgtgactcg ggccgcggtc 420
gcggcgaa	agaccgtggc	gcgcgccgtg aagaccgccg ccgcgcgcaa gtccgcgaag 480
aagcgtga		489
<210>	25	
<211>	264	
<212>	DNA	
<213>	深棕色孢囊杆菌	
<220>		
<221>	misc_feature	
<222>	(1)..(264)	
<223>	ORF4	
<400>	25	
atgagccgg	caagacgcaa	ggagagcaag cagcacgaag tgggtccgc cacacacgca 60



```

cgccgggtga tcgtggcgac ggatggccgg ggttggtagc tccgattcga gggcaaccgt 120
cagctcgccc ggtattccaa cgtgaccagc gccatccacg gcgggcgcag gctggctcgc 180
cagcacaaag ccgcccggct cgtggtagcg tacctggacg gggaagagga agagtctctg 240
tacggggacc gcgagcgccc ttga 264

```

```

<210> 26
<211> 450
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(450)
<223> ORF5

```

```

<400> 26
atgaaacaca tcaaggcggg ggtgggtggg gcgttgctcg cggtcttgcg ctteggcgctg 60
ggatgtcaga cgacgggcgg tgcctgggaat caaggaaacgg gcgggagcga tacgtctcag 120
ggcggcacca tgaccggaag tgagacgacc ggaaccggaa cgaccggagg caccacggaa 180
ggtgggtgaca ccacgggcgg aggcaccggc ggaaccagtg ctggcgacat cgacgggttcg 240
agcagtggca gcacgggctc cggtagcgac gtgggcggct ccggcggtcc gggcgtgtcc 300
agtgaaccgg gcgggtttcag ccccgacgcc tcgggcgtgg acagcgacct gggcggctcc 360
ggcaccggca gtgacgtgga cggtccggcg agcaccgact ccagcgggcaa catgagcggc 420
acgggctccg aagacgacac cagcgcgtga 450

```

```

<210> 27
<211> 1578
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(1578)
<223> ORF6

```

[0030]

```

<400> 27
atgagcagcg gcacctccct ggccctggcc gcgtccctcg ccgcgctgcc cgcgctcgcc 60
caggagcgte ccagcgaggg cgacctcttc ggccggcgaca ctccagagac gaagcccgt 120
ccggccgatg gcgcccgccc cgacgagagt tccctcttcg gtgacacccc cgcgtccacc 180
ccggccgcac agagcgcggc ggccaccgcg gcccccgcga agccctccgc cagccccag 240
gaccgggatg cgcaggcgct cggtagcccg tcggccacca acgccttcga caccgaggag 300
gcccgtgagg atcgcgtgaa gatcgccggc cgtttctacc tgcgcgcta ctacaggcc 360
aacgaagggg tgccttcag caacaccacc ttctccgccc ccattgctgt ggacggctac 420
ttcgatcccc gcccccagga gcggctgcgc ggtctcgtgc tcggacggct caccctcgat 480
ccgaccgcga aggcgggctc cctcggcacc gtcgccacca gcacgtccac ctccaacgtc 540
gttgcggatc cggctcgtgt gttggatcag gcctggctgc gcttcgacct ggaccacaag 600
ctttctatca ccgtcggcaa gcagcacgtg aagtggggca cctcgcgctt ctggaacccc 660
accgaattcc tctcgcccca gcgcagggat ccgctcgccc tcttggacac gcgcaccggc 720
gcgaccatgc tcaagatgca catgccctgg gagcgaaaag gctggaactt ctacgtctc 780
ggcctgctcg acaacgcgg cccggccaat acgctcgccc gcgtcggggg cgctgctcgc 840
gccgaggtgg tgctcgccca tacggaactc ggctcgatg ccgtgctcca acacggccgc 900
aagccccgct tcgggctcga cctctctcc gggtcggccc ccacgacat ctacggcgaa 960
ctcgcctca agaaggctc ggatcgcccc atgttccgca tgccccaagg tgtctccctc 1020
ggagacctgc tcggtagtt ccaggccaat ggccgcatgc ctcccacct gggcgcgctc 1080
ccatagagg cgtactacc cgagggttac acgcgcgagg tgagcgggcg cgcgacctgg 1140
acgttcgct actcggagag cgacaccgcc accgtggcg tcgagtact ctacaattcg 1200
atgggtatc ccgctcgtg ggctacccc tacctatcc tccaggcca gtatcagccc 1260
ttctacctcg gccggcata cgcgcgctc tacgcgttc tgctcggtcc gggatcctgg 1320
gacaacacca acttcatct gtcaacctg ggaacctct ctgaccttc ttctatcaca 1380
cggttgagc tgacgcacc ggccctgcgc tatctcagca tcgagggctt catcgccgc 1440
aactatggcc agcggggtgg cgagtccgc ttcgcgtca acctccggc cctgcgcatg 1500
ggcgagcagg tgacgcctc catcgccgc gctccacct ccatccaggc cggggtgggt 1560
ctgcgcatcg acctttga 1578

```

```

<210> 28
<211> 786
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature

```

&lt;222&gt; (1).. (786)

&lt;223&gt; ORF7

&lt;400&gt; 28

```

atgacctgc gcaacctctc eggcgccctg ttcgccgcgc tgetgtggc cgetccgacc      60
gtctgcgcgg acctcaccga ccccgccgag atcaagaagc tcctggagac gctcgacaac      120
cgccagcgca acggcgcgga ctacaagtcg ctggtgtata tcgagcagaa ggagaaggac      180
aaaacagacg tcgtgcgcga ggccgtcgtc taccggcgcg acgagaagga tcagctgatg      240
atcctcatga ccaagcccaa gggcgaggcc ggcaagggtc acctgcggct ggacaagaac      300
ctctggagct acgaccgaa caccggcaag tgggaccggc gcaccgagcg tgagcgatc      360
gccggcaccg acagccgccg gcccgacttc gacgagtcgc gcctggccga ggagctcgat      420
ggcaagttcg agggcgagga gaaactcgcc aagttcacca cctggaagct cgtcctcacc      480
gccaagccga acgtggacgt cgcctacccc gtgttacacc tgtgggtgga gaaggacacg      540
aacaacatcc tcaagccca ggagttcgcc ctttcgggcc gcctgatcg cactcctac      600
ttccccaagt ggatgaagct cttcagcgag tccaagaagg ccgagctctg gtaccgcgag      660
gagatgcgct tctatgacga ggtggagaag accaactcca ccgtcatcgt cgtgaagage      720
gtggacctgc gctcgtcga ggagaacatc ttcaccaagg cctggttcga gagcaaaage      780
cgatga                                           786

```

&lt;210&gt; 29

&lt;211&gt; 1302

&lt;212&gt; DNA

&lt;213&gt; 深棕色孢囊杆菌

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1).. (1302)

&lt;223&gt; ORF8

&lt;400&gt; 29

```

atgcaacagc tctctctcat cgcagtgcgc aacctgggca cccacaagcg ccgtacgctt      60
ctgttggggc ggcctatcgc cgggtgtcac gccctgtctg tcatectcat gggcctgtcc      120
aacggcatga aggacacgat gctccgggcc gccaccacgc tggtagccgg gcacgtcaac      180
gtggttggct tctacaaggt gacggccggc cagtctgcgc ccgttggtag ctcctacccc      240
aagctgtctg agcagctgg caaggaagtc cccgagctgg acttctccgt ccagcgcacg      300
cgcggctggg tcaagttggt gacgagctct ggctccgtgc agacgggaat cggcgccatc      360
gacgtagcgg ccgagactgg catccgcaag gtgctgcagt tgcgggaggg tcggttggaa      420
gacctggcgc aacccaatac cctctctcct ttcgacgagc aggcgaagcg gctcgaggtc      480
aaggtgggtg acagcgtcac cctctccgcg tccaccatgc gcgggatcag caacaccgtg      540
gacgtacgtg tggtagccat cgcgcgcaac gtgggcatgc tgagttcctt caacgtcttg      600
gtccccaacg ccaccctcgc cgcctcttac cagctgcgcg aggactccac cggcgccctc      660
atgtccaccc tcaaggacat gagegccatc cccagcgtgc aggcgcgcct ctacaagcgt      720
ctgcceagat tgggttatca ggtgctggag catgaccccc gggccttctt catgaagttc      780
cagaccgtga acccgagggc ctggacgggg cagaagctgg acatcaccaa ctgggaggac      840
gagatctcct tcatcaagtg gaccgtgtcg gcgatggacg ccctcaccgg cgtcctcctc      900
ttcgtgtcgc tcatcatcat cgcgggtgggc atcatgaaca cctgttggat cggcatccgc      960
gagcgcaccc gggaaatcgg caccctgcgc gccatcgcca tgcagcgtg gtactgtctg      1020
gtgatgttcc tcttggagge gctcgtgctc ggactgtctg gcaccacggt gggcgccctc      1080
gtgggcatgg gcgtgtgctt gctcatcaac gccgtggacc cctccgtgcc cgtgcccgtc      1140
cagctcttca tctctctcga caagctccac ctcatcgtga agcccggtc ggtgatgaga      1200
gccatcgctg tcatcacgct gtgcaccacc ttcattctgc tcattccctc ttctctcgcc      1260
gcgcggatga agcccatcac ggcatgcac cacatcggtg ga                                           1302

```

&lt;210&gt; 30

&lt;211&gt; 2106

&lt;212&gt; DNA

&lt;213&gt; 深棕色孢囊杆菌

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;222&gt; (1).. (2106)

&lt;223&gt; ORF9

&lt;400&gt; 30

```

atgggcaaac tcaagctcct gctccaagtg gccctgcgca acttgttctg gacgaggatc      60
aacctctctc tggaggcatc catcttcttc ggcaccgtgc tggtagtggt gggcggtctc      120
ctcgtcgaca gcttggacga ggcatgagc cgcagcatta tcggcagcgt cggcgccac      180
ctccaggtgt actcggccca ctccaaggac gagctctcgc tcttcgggca gatgggcccgc      240
gaaccggacc tgacgcgctt ggaatgactt tcgcgcacca agcaactggt acagcagcac      300
cccaactgta agacggtggt gcccatgggc accggcgcca cgttcatcaa ctcgggaaac      360
accatcgacc tgaccttggc gcgcctgcgc gacctctaca agaaagcagc acaggcgcac      420

```

[0031]

```

acaccegaac tccgcgggca gatccacage ctccaggcgc atgtgcgtca catcatcacc 480
ttgtctgagg aggatatgaa gcggcgagg gaaatcatcg acgacaagac caccgacccc 540
gcggacgcgg aggcacatggc ccgcgccgt tccgaggcct tctggcgga cttcgacgag 600
aagccatteg actcgtcga gtctctggag aaccgcctcg ccccgatat gacggacggg 660
gacatgttgt cccctgccta ttagggacc gacctggta acttcagaa gaccttcgac 720
cgcatgcga tctggaggg cacgcgggtg ccccggggc accgcggcat gatctctcc 780
aagttcact acgagaacga cttcaagctg aagacggcgc accggttga tctcatcaag 840
gaggcgctg ataccaacca caagaccatc gcgatggatc cgaactcca gcctgggtg 900
aaggagaacc agaccagac gcgggagatc ctctccagc tcgacgacct caagacgaag 960
caggcgctg agcggtcca gcgctgtg ggcagccagg agacggacct gggcaagta 1020
ctgcccgcct tcttcccat ggatgacgcc aacttcgaca cgcctacca gcagttctac 1080
tcgagctgg cgacgtgct cgacctgtac cgcacccga tcggggagca cctcaccatc 1140
accgcattct cgcgcaccgg ctatgtcag agcgtgaacg tgaagatcta cggcacctac 1200
cagttcgacg ggtctggaga gtccgggtc gccggagccc tcaacctgct ggacctgatg 1260
tcttccgcg agctgtacgg ctatctcacc gctgagaaga aggcagact cgcgggctg 1320
cagaaggcca gcggggtgca gcaggtgaag cgcgaggacg ccgagacggc gctctttggc 1380
gagcagggca gcgctcgt ggtggccgag gggaccgccg gccagatcga cgaggacaag 1440
caactcgacg gcttcgcca gaagctgcac cgcgaggagc tcgctcccg ggtgtacacg 1500
cagcaggaaa tcgaaagcgg cgtggtgctc agcaccgcgg ttctgtgaa gcacccggag 1560
aagctggagc agaccctggc cgagctcgg aaatcgccgg acgacgcgaa actacccttg 1620
cggatcatct cctggcagaa ggctccggc acgatcgcc agttcgtct ggctgccaag 1680
ctggtcctct agctggagct cttcatcacc ttcgtggtg gcctcgtcat catcaaac 1740
gcgatgatga tggccacgct gcagcgggtg cgcgagggtg gcacctcg gcgcacggc 1800
gcgcagcgt cgttcgtgct gagcatggtg ctggtggaaa cgggtggtg ggggctctc 1860
ttcgcgctg tgggagccgc catgggaggt gccatcatga acatgctcgg ccacgtggg 1920
atcccccg gcaacgagc gctctactt ttctctcgg gaccccgct cttcccaagt 1980
ctccacctgt caaacctcgt ggcggccttc gtcacgtgc tctgtgtgc cgcctctcc 2040
accttctacc ccgctacct cgcgaccgg gtctcgctc tccaggcgat gcagacggac 2100
gagta 2160

```

<210> 31  
 <211> 762  
 <212> DNA  
 <213> 深棕色孢囊杆菌

[0032]

<220>  
 <221> misc\_feature  
 <222> (1)..(762)  
 <223> ORF10

```

<400> 31
atgagccagg tcaatgcct ccccgccagc acccagccga tctctccct caccgaggtt 60
accaagacgt actccctggg taaggtgcag gtgcccgac tccgaggcgt gacgctagag 120
gtgtaccggg gagatgcat ctccatgcc ggcaccatcg gcagtggcaa gacgacggcg 180
ctcaatccta tggctcgt ggacacggcc tctcgggcg tggtagcgt ggatggccag 240
gacaccaaga agctcaccga gcggcagct acccacttgc ggtgcacac catcggttc 300
atcttcaga gctcaacct cgtctgggtg ctacgcgtt tccagaacgt agagtcccc 360
ctgtgtgc agcgaagct caacgcctc gagcgcgca cgcgctgat gacgtgtg 420
gagcaggtgg gcttgagaa gcacgcaaa caccgcccc atgagctgtc tggaggccag 480
cgccagcgg tggccgtggc gcgctctc gtcaccggc ccaagctggt gctcgccgac 540
gagccaccg ccaacctga ctccgtacc ggcagaaca tcatgacct gatgaaggag 600
ctcaaccga aggagggcac cacttctc ttctccacc acgacgcaa ggtgatgac 660
cacgccaacg cgtggtgct cctggcgac gggaagatcc tcgaccgat cagccggcc 720
gaggccaga aggtcatggc cgtgagcag gggggccact aa 762

```

<210> 32  
 <211> 1194  
 <212> DNA  
 <213> 深棕色孢囊杆菌

<220>  
 <221> misc\_feature  
 <222> (1)..(1194)  
 <223> ORF11

```

<400> 32
atgccgaga agttcgtggg gaagtgaag ggcggcgagg tcaagctcgt cgatggtcgg 60
aaggtgtggc tcttcagaaa gatggtctc gggcccggt tctcggtctc cttggcggtc 120
tcaaacgagg aggcagcgt ggccgagctg gccctgttc ggcgcgaccg ggacgctac 180
ctggccaagg tgaaggccga caggtcggag gaagtccagg catccactgt agccggggca 240
gttctctgt cgggggatgt ggggcctcgg ctcatgccc attctgtcc ggagtctctc 300

```

```

cgacacttga cccagcgggg gccaacggag ggttaccggc gggacgccc aacctacctg 360
tcgaatggg ccgaggttct ggccggaagg gacctgagta ccgtcagcct cctcagttg 420
cgccgcgccc tgagccaatg gcccacggcc aggaagatgc ggatcatcac gctcaagagc 480
ttcttctcgt ggctgagggg agaggatcgc ctcaaggctg ctgaagacc cacttggtcc 540
ctcaagggtc cgcccgcggt cgcgggagaag gggagacggg ccaaggggta ttcgatggcc 600
caagtggaga agctctacgc ggccatcgge tcccagacgg tgaggagcgt gctgtgtctg 660
cgggccaaga ccggcatgca cgactcggag atcggccgcc tggcatcggg caagggggaa 720
ctgcgcgtcg tcaatgaccc ctccggcacc gccggtactg cgcggtttct gcacaagaac 780
ggccgcgttc acatcctcag tctggatgcc caggcccttg ctgcgcgca gcggtccag 840
gttcggggca ggccgcccat caggaacacc gtccgggagt ccacgggta tgcgtcggcg 900
cgcatggggc agtcgcccat ccacccagc gagctccgcc acagcttcac cacttgggcc 960
acgaatgagg gccaggtcgt gagggcaacc cggggcggag tgcactcga tgcgttggc 1020
tcggttcttg gccatcagtc cacacgggag accaagaagt tctatgacgg gaccgaaatt 1080
ccccgatga tcaccgtccc gctcaagctg catcatccac aggaccagc ggtgatgcag 1140
ctgaggcgta actgctcgcc ggaccccgtc gtgacgagag aggcagaggc gtga 1194

```

```

<210> 33
<211> 375
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(375)
<223> ORF12

```

```

<400> 33
gtgtctctcg cattccctc cggcctcctg tcgttggcgc tcctgtccac taccaccgaa 60
atctctcgcg ctcttccctg ggacgagtc gagtcggcga gccctgcgat cgagctgcc 120
gtacgccag ggggaaagcc acccgtgggtg tgtctcggtc caggtctgcc cattcattc 180
cgcttcgact ccgcgtcca acagaagtc ctgaggatc aggatcgggg ctggttcgag 240
gattgggctt tgggccagca gacgtctgta ctgactctc acgacaacct ggtggctggg 300
aagcgatctg aagtggaggt gtgcttcgcg gatggtgcg ccccgcgctg cgcttcctc 360
gtgtccggc gtga 375

```

[0033]

```

<210> 34
<211> 339
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(339)
<223> ORF13

```

```

<400> 34
atgcacacga aggtgccctc cgtcttcgag gcaacgccc agtctctcag tgacgtggac 60
taccagttct ggcatgagga ctcccgagg gtgttcgagc ggcagcacat cgacgcgcac 120
gggtgcccc ccattggcgc gtacttgggc gaggtgctg tgcgtaacct gggcggcaag 180
tgatatactc gccagaaact cgacgaggcc caggtgctcg tcggcaaccg tgtgtggttg 240
ccgtttgcgc gggtcaccac ctacatgcgc tcgtgcgaat cgttgcgtga ctactccctc 300
accagctct accgcgtggc cgagcgttac cggggttga 339

```

```

<210> 35
<211> 915
<212> DNA
<213> 深棕色孢囊杆菌

```

```

<220>
<221> misc_feature
<222> (1)..(915)
<223> ORF 14

```

```

<400> 35
atgaagggtc tggggcttgg tgacgtgaag tcggaggaca gtctccggt cacttttgag 60
ggtgcgcttg atccgcaggc tgcgcttgag aaagttctcg agccattttt ccaggcgctg 120
gaggaatatg caggcgattg gatccggaa gtctcagtg gcaggcggcg actcaaatac 180
tcccagacca atatctggaa ggctctggag gagcggcgcg atgaacgaag cacagacacc 240
tggtcttacc gcacacagcg gccgacactg gagatgtcgc tgcattctctg gtttccgcgc 300
cttccgcccc ctttggacgt aatgactacg gtgcaaccgc tcacccgctt cgcggagaag 360
gagcgctgcc gccaatcgt agaaatggta cgcacctggg cctcttgcta cccggtcact 420

```

	cacgcegcag cccacagcgt ggctgacagg gcgttggcag gtgcgcccga ttttggacgc	480
	gatgcgcgga ccgcacggag agacgggttc gacagaatct acgagatctt ctggetcaac	540
	gtcttcggcc ccaagtgtgt ggaagccgtg ggccgcgagc gcattgtgtc caccgcagct	600
	caccgggtgg aggaactgcc caatggctcc atcctcctgg tgacgtggcc caccgtgcg	660
	gacttcggg gcgcgagggc acggcacgca caggcgcgcg cgcacgttca cctccggccg	720
	gacctccgct tcgacacggg gctgcgaacc ctgcacgagc gtagegccgc gctcgtccc	780
	gttgagccct gcttcacccc ggatgtagcg ccaactctct ctcacgttgt ggatagcgtc	840
	gccatccgga tgtggaaaac ctggagcgcg ctaacgagca ttacagaact ctggctgagc	900
	acctcgtggc gctga	915
	<210> 36	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<220>	
	<221> misc_feature	
	<222> (1)..(32)	
	<223> CysL KO For	
	<400> 36	
	tgattgattg atcggcgcga ttccgcctct gg	32
	<210> 37	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
[0034]	<220>	
	<221> misc_feature	
	<222> (1)..(32)	
	<223> CysL KO Rev	
	<400> 37	
	tcaatcaatc atcggtcgc ggtctcaggc tc	32
	<210> 38	
	<211> 37	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<220>	
	<221> misc_feature	
	<222> (1)..(37)	
	<223> CysK KO For	
	<400> 38	
	tgattgattg aaaaacagtc ggaggagttt cttgtcc	37
	<210> 39	
	<211> 32	
	<212> DNA	
	<213> 人工序列	
	<220>	
	<223> 引物	
	<220>	
	<221> misc_feature	
	<222> (1)..(32)	
	<223> CysK KO Rev	

<400> 39  
tcaatcaate aactcccagt gccctcagcc tc

32

<210> 40  
<211> 70  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1).. (70)  
<223> CysA

<400> 40  
Met Ser Met Asn Gly Asp Glu Ala Glu Tyr Val Val Leu Ile Asn Gly  
1 5 10 15  
Glu Glu Gln Tyr Ser Leu Trp Pro Val His Arg Glu Ile Pro Gly Gly  
20 25 30  
Trp Lys Thr Val Gly Pro Lys Gly Ser Lys Glu Thr Cys Gln Ser Tyr  
35 40 45  
Ile Gln Glu Val Trp Thr Asp Met Arg Pro Lys Ser Leu Arg Glu Ala  
50 55 60  
Leu Thr Arg Ser Asn Cys  
65 70

<210> 41  
<211> 317  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1).. (317)  
<223> CysB

[0035]

<400> 41  
Met Ser Thr Pro Ala Ala Gly Ala Lys Pro Ser Tyr Leu Ala Gly Ile  
1 5 10 15  
Glu Thr Val Met Val Glu Pro Glu Leu Glu Val Arg Tyr Leu Thr  
20 25 30  
Val Glu Ser Gly Asp Gly Arg Gln Ser Thr Leu Tyr Glu Phe Gly Pro  
35 40 45  
Lys Asp Ala Glu Lys Val Val Val Leu Pro Pro Tyr Gly Val Thr Phe  
50 55 60  
Leu Leu Val Ala Arg Leu Ala Arg Leu Leu Ser Gln Arg Phe His Val  
65 70 75 80  
Leu Ile Trp Glu Ser Arg Gly Cys Pro Asp Ser Ala Ile Pro Val Tyr  
85 90 95  
Asp Thr Asp Leu Gly Leu Ala Asp Gln Ser Arg His Phe Ser Glu Val  
100 105 110  
Leu Lys Gln Gln Gly Phe Glu Ala Phe His Phe Val Gly Trp Cys Gln  
115 120 125  
Ala Ala Gln Leu Ala Val His Ala Thr Ala Ser Gly Gln Val Lys Pro  
130 135 140  
Arg Thr Met Ser Trp Ile Ala Pro Ala Gly Leu Gly Tyr Ser Leu Val  
145 150 155 160  
Lys Ser Glu Phe Asp Arg Cys Ala Leu Pro Ile Tyr Leu Glu Ile Glu  
165 170 175  
Lys His Gly Leu Leu His Ala Glu Lys Leu Gly Arg Leu Leu Asn Lys  
180 185 190  
Tyr Asn Gly Val Pro Ala Thr Ala Gln Asn Ala Ala Glu Lys Leu Thr  
195 200 205  
Met Arg His Leu Ala Asp Pro Arg Met Thr Tyr Val Phe Ser Arg Tyr  
210 215 220  
Met Lys Ala Tyr Glu Asp Asn Arg Leu Leu Ala Lys Gln Phe Val Ser  
225 230 235 240  
Thr Ala Leu Asp Ser Val Pro Thr Leu Ala Ile His Cys Arg Asp Asp  
245 250 255  
Thr Tyr Ser His Phe Ser Glu Ser Val Gln Leu Ser Lys Leu His Pro  
260 265 270  
Ser Leu Glu Leu Arg Leu Leu Gly Lys Gly Gly His Leu Gln Ile Phe

```

                275                280                285
Asn Asp Pro Ala Thr Leu Ala Glu Tyr Val Leu Gly Phe Ile Asp Thr
                290                295                300
Arg Ala Ser Gln Ala Ala Ala Pro Ala Val Ala Gly Ala
305                310                315

<210> 42
<211> 459
<212> PRT
<213> 深棕色孢囊杆菌

<220>
<221> MISC_FEATURE
<222> (1)..(459)
<223> CysC

<400> 42
Met Ile Leu Pro Asn Asn Ile Gly Leu Asp Glu Arg Thr Gln Leu Ala
1      5      10      15
Arg Gln Ile Ser Ser Tyr Gln Lys Lys Phe His Val Trp Trp Arg Glu
      20      25      30
Arg Gly Pro Thr Glu Phe Leu Asp Arg Gln Met Arg Leu Arg Thr Pro
      35      40      45
Thr Gly Ala Val Ser Gly Val Asp Trp Ala Glu Tyr Lys Thr Met Arg
      50      55      60
Pro Asp Glu Tyr Arg Trp Gly Leu Phe Met Val Pro Met Asp Gln Asp
65      70      75      80
Glu Ile Ala Phe Gly Asp His Arg Gly Lys Lys Ala Trp Glu Glu Val
      85      90      95
Pro Ser Glu Tyr Arg Thr Leu Leu Leu Gln His Ile Cys Val Gln Ala
      100     105     110
Asp Val Glu Asn Ala Ala Val Glu Gln Ser Arg Leu Leu Thr Gln Met
      115     120     125
Ala Pro Ser Asn Pro Asp Leu Glu Asn Val Phe Gln Phe Phe Leu Glu
130     135     140
Glu Gly Arg His Thr Trp Ala Met Val His Leu Leu Leu Ala His Phe
145     150     155     160
Gly Glu Asp Gly Val Val Glu Ala Glu Ala Leu Leu Glu Arg Leu Ser
      165     170     175
Gly Asp Pro Arg Asn Pro Arg Leu Leu Glu Ala Phe Asn Tyr Pro Thr
      180     185     190
Glu Asp Trp Leu Ser His Phe Met Trp Cys Leu Leu Ala Asp Arg Val
      195     200     205
Gly Lys Tyr Gln Ile His Ala Val Thr Glu Ala Ser Phe Ala Pro Leu
210     215     220
Ala Arg Ala Ala Lys Phe Met Met Phe Glu Glu Pro Leu His Ile Ala
225     230     235     240
Met Gly Ala Val Gly Leu Glu Arg Val Leu Ala Arg Thr Ala Glu Val
      245     250     255
Thr Leu Arg Glu Gly Thr Phe Asp Thr Phe His Ala Gly Ala Ile Pro
260     265     270
Phe Pro Val Val Gln Lys Tyr Leu Asn Tyr Trp Ala Pro Lys Val Tyr
275     280     285
Asp Leu Phe Gly Asn Asp Gly Ser Glu Arg Ser Asn Glu Leu Phe Arg
290     295     300
Ala Gly Leu Arg Arg Pro Arg Asn Phe Val Gly Ser Glu Ser Gln Ile
305     310     315     320
Val Arg Ile Asp Glu Arg Met Gly Asp Gly Leu Thr Val Val Glu Val
      325     330     335
Glu Gly Glu Trp Ala Ile Asn Ala Ile Met Arg Arg Gln Phe Ile Ala
      340     345     350
Glu Val Gln Thr Leu Ile Asp Arg Trp Asn Ala Ser Leu Arg Ala Leu
      355     360     365
Gly Val Asp Phe Gln Leu Tyr Leu Pro His Glu Arg Phe Ser Arg Thr
370     375     380
Tyr Gly Pro Cys Ala Gly Leu Pro Phe Asp Val Asp Gly Lys Leu Leu
385     390     395     400
Pro Arg Gly Thr Glu Ala Lys Leu Ala Glu Tyr Phe Pro Thr Pro Arg
      405     410     415
Glu Leu Ala Asn Val Arg Ser Leu Met Gln Arg Glu Leu Ala Pro Gly
      420     425     430

```

[0036]

Gln Tyr Ser Ser Trp Ile Ala Pro Ser Ala Thr Arg Leu Ser Ala Leu  
 435 440 445  
 Val Gln Gly Arg Asn Thr Pro Lys Glu His Glu  
 450 455

<210> 43  
 <211> 732  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC FEATURE  
 <222> (1).. (732)  
 <223> CysD

<400> 43  
 Met Arg Cys Leu Ile Ile Asp Asn Tyr Asp Ser Phe Thr Trp Asn Leu  
 1 5 10 15  
 Ala Asp Tyr Val Ala Gln Thr Phe Gly Ser Glu Pro Leu Val Val Arg  
 20 25 30  
 Asn Asp Gln His Thr Trp Gln Glu Ile Lys Ala Leu Gly Ser Phe Gly  
 35 40 45  
 Cys Ile Leu Val Ser Pro Gly Pro Gly Ser Val Thr Asn Pro Lys Asp  
 50 55 60  
 Phe Asn Val Ser Arg Asp Ala Leu Glu Gln Asp Glu Phe Pro Val Phe  
 65 70 75 80  
 Gly Val Cys Leu Gly His Gln Gly Leu Ala Tyr Ile Tyr Gly Gly Glu  
 85 90 95  
 Ile Thr His Ala Pro Val Pro Phe His Gly Arg Thr Ser Thr Ile Tyr  
 100 105 110  
 His Asp Gly Thr Gly Val Phe Gln Gly Leu Pro Pro Ser Phe Asp Ala  
 115 120 125  
 Val Arg Tyr His Ser Leu Val Val Arg Pro Glu Ser Leu Pro Ala Asn  
 130 135 140  
 Leu Val Val Thr Ala Arg Thr Glu Cys Gly Leu Ile Met Gly Leu Arg  
 145 150 155 160  
 His Val Ser Arg Pro Lys Trp Gly Val Gln Phe His Pro Glu Ser Ile  
 165 170 175  
 Leu Thr Ala His Gly Leu Gln Leu Ile Ser Asn Phe Arg Asp Glu Ala  
 180 185 190  
 Tyr Arg Tyr Ala Gly Lys Glu Val Pro Ser Arg Arg Pro His Ser Thr  
 195 200 205  
 Ala Gly Asn Gly Val Gly Ala Gly Ala Ala Arg Arg Asp Pro Ser Ala  
 210 215 220  
 Arg Arg Thr Pro Glu Arg Arg Arg Glu Leu Gln Thr Phe Thr Arg Arg  
 225 230 235 240  
 Leu Ala Thr Ser Leu Glu Ala Glu Thr Val Phe Leu Gly Leu Tyr Ala  
 245 250 255  
 Gly Arg Glu His Cys Phe Trp Leu Asp Ser Gln Ser Val Arg Glu Gly  
 260 265 270  
 Ile Ser Arg Phe Ser Phe Met Gly Cys Val Pro Glu Gly Ser Leu Leu  
 275 280 285  
 Thr Tyr Gly Ala Ala Glu Ala Ala Ser Glu Gly Gly Ala Glu Arg Tyr  
 290 295 300  
 Leu Ala Ala Leu Glu Arg Ala Leu Glu Ser Arg Ile Val Val Arg Pro  
 305 310 315 320  
 Val Asp Gly Leu Pro Phe Glu Phe His Gly Gly Tyr Ile Gly Phe Met  
 325 330 335  
 Thr Tyr Glu Met Lys Glu Ala Phe Gly Ala Ala Thr Thr His Lys Asn  
 340 345 350  
 Thr Ile Pro Asp Ala Leu Trp Met His Val Lys Arg Phe Leu Ala Phe  
 355 360 365  
 Asp His Ser Thr Arg Glu Val Trp Leu Val Ala Ile Ala Glu Leu Glu  
 370 375 380  
 Glu Ser Ala Ser Val Leu Ala Trp Met Asp Glu Thr Ala Asp Ala Leu  
 385 390 395 400  
 Lys Ser Leu Pro Arg Gly Thr Arg Ser Pro Gln Ser Leu Gly Leu Lys  
 405 410 415  
 Ser Ile Ser Val Ser Met Asp Cys Gly Arg Asp Asp Tyr Phe Ala Ala  
 420 425 430  
 Ile Glu Arg Cys Lys Glu Lys Ile Val Asp Gly Glu Ser Tyr Glu Val

[0037]



435 440 445  
 Cys Leu Thr Asn Gly Phe Ser Phe Asp Leu Lys Leu Asp Pro Val Glu  
 450 455 460  
 Leu Tyr Val Thr Met Arg Arg Gly Asn Pro Ala Pro Phe Gly Ala Phe  
 465 470 475 480  
 Ile Lys Thr Gly Lys Thr Cys Val Leu Ser Thr Ser Pro Glu Arg Phe  
 485 490 495  
 Leu Lys Val Asp Glu Asp Gly Thr Val Gln Ala Lys Pro Ile Lys Gly  
 500 505 510  
 Thr Cys Ala Arg Ser Asp Asp Pro Ala Thr Asp Ser Thr Asn Ala Ala  
 515 520 525  
 Arg Leu Ala Ala Ser Glu Lys Asp Arg Ala Glu Asn Leu Met Ile Val  
 530 535 540  
 Asp Leu Met Arg Asn Asp Leu Gly Arg Val Ser Val Pro Gly Ser Val  
 545 550 555 560  
 His Val Ser Asn Leu Met Asp Ile Glu Ser Phe Lys Thr Val His Gln  
 565 570 575  
 Met Val Ser Thr Val Glu Ser Thr Leu Thr Pro Glu Cys Ser Leu Val  
 580 585 590  
 Asp Leu Leu Arg Ala Val Phe Pro Gly Gly Ser Ile Thr Gly Ala Pro  
 595 600 605  
 Lys Ile Arg Thr Met Glu Ile Ile Asp Arg Leu Glu Lys Ser Pro Arg  
 610 615 620  
 Gly Ile Tyr Cys Gly Thr Ile Gly Tyr Leu Gly Tyr Asn Arg Ile Ala  
 625 630 635 640  
 Asp Leu Asn Ile Ala Ile Arg Thr Leu Ser Tyr Asp Gly Thr Leu Val  
 645 650 655  
 Lys Phe Gly Ala Gly Gly Ala Ile Thr Tyr Leu Ser Gln Pro Glu Gly  
 660 665 670  
 Glu Phe Gln Glu Ile Leu Leu Lys Ala Glu Ser Ile Leu Arg Pro Ile  
 675 680 685  
 Trp Gln Tyr Ile Asn Gly Ala Gly Ala Pro Phe Glu Pro Gln Leu Arg  
 690 695 700  
 Asp Arg Val Leu Cys Leu Glu Glu Lys Pro Arg Arg Val Ile Arg Gly  
 705 710 715 720  
 His Gly Ser Ala Ile Asp Ala Val Glu Pro Ser Ala  
 725 730

[0038]

<210> 44  
 <211> 243  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (243)  
 <223> CysE

<400> 44  
 Met Ile Ala Phe Asn Pro Gln Ala Arg Pro Arg Leu Arg Leu Phe Cys  
 1 5 10 15  
 Phe Pro Tyr Ala Gly Gly Asp Ala Asn Ile Phe Arg Asp Trp Ala Ala  
 20 25 30  
 Ala Met Pro Glu Gly Val Glu Val Leu Gly Val Gln Tyr Pro Gly Arg  
 35 40 45  
 Gly Thr Asn Leu Ala Leu Pro Pro Ile Ser Asp Cys Asp Glu Met Ala  
 50 55 60  
 Ser Gln Leu Leu Ala Val Met Thr Pro Leu Leu Gly Ile Asn Phe Ala  
 65 70 75 80  
 Phe Phe Gly His Ser Asn Gly Ala Leu Ile Ser Phe Glu Val Ala Arg  
 85 90 95  
 Arg Leu His Asp Glu Leu Lys Gly Arg Met Arg His His Phe Leu Ser  
 100 105 110  
 Ala Lys Ser Ala Pro His Tyr Pro Asn Asn Arg Ser Lys Ile Ser Gly  
 115 120 125  
 Leu Asn Asp Glu Asp Phe Leu Arg Ala Ile Arg Lys Met Gly Gly Thr  
 130 135 140  
 Pro Gln Glu Val Leu Asp Ala Arg Leu Met Gln Ile Leu Leu Pro  
 145 150 155 160  
 Arg Leu Arg Ala Asp Phe Ala Leu Gly Glu Thr Tyr Val Phe Arg Pro  
 165 170 175

Gly Pro Thr Leu Thr Cys Asp Val Ser Ile Leu Arg Gly Glu Ser Asp  
 180 185 190  
 His Leu Val Asp Gly Glu Phe Val Gln Arg Trp Ser Glu Leu Thr Thr  
 195 200 205  
 Gly Gly Ala Ser Gln Tyr Ala Ile Asp Gly Gly His Phe Phe Leu Asn  
 210 215 220  
 Ser His Lys Ser Gln Val Val Ala Leu Val Arg Ala Ala Leu Leu Glu  
 225 230 235 240  
 Cys Val Leu

<210> 45  
 <211> 345  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (345)  
 <223> CysF

[0039]

<400> 45  
 Met Thr Ala Gln Asn Gln Ala Ser Ala Phe Ser Phe Asp Leu Phe Tyr  
 1 5 10 15  
 Thr Thr Val Asn Ala Tyr Tyr Arg Thr Ala Ala Val Lys Ala Ala Ile  
 20 25 30  
 Glu Leu Gly Val Phe Asp Val Val Gly Glu Lys Gly Lys Thr Leu Ala  
 35 40 45  
 Glu Ile Ala Lys Ala Cys Asn Ala Ser Pro Arg Gly Ile Arg Ile Leu  
 50 55 60  
 Cys Arg Phe Leu Val Ser Ile Gly Phe Leu Lys Asn Ala Gly Glu Leu  
 65 70 75 80  
 Phe Phe Leu Thr Arg Glu Met Ala Leu Phe Leu Asp Lys Lys Ser Pro  
 85 90 95  
 Gly Tyr Leu Gly Gly Ser Ile Asp Phe Leu Leu Ser Pro Tyr Ile Met  
 100 105 110  
 Asp Gly Phe Lys Asp Leu Ala Ser Val Val Arg Thr Gly Glu Leu Thr  
 115 120 125  
 Leu Pro Glu Lys Gly Val Val Ala Pro Asp His Pro Gln Trp Val Thr  
 130 135 140  
 Phe Ala Arg Ala Met Ala Pro Met Met Ser Leu Pro Ser Leu Leu Leu  
 145 150 155 160  
 Ala Glu Leu Ala Asp Arg Gln Ala Asn Gln Pro Leu Lys Val Leu Asp  
 165 170 175  
 Val Ala Ala Gly His Gly Leu Phe Gly Leu Ala Ile Ala Gln Arg Asn  
 180 185 190  
 Pro Lys Ala His Val Thr Phe Leu Asp Trp Glu Asn Val Leu Gln Val  
 195 200 205  
 Ala Arg Glu Asn Ala Thr Lys Ala Gly Val Leu Asp Arg Val Glu Phe  
 210 215 220  
 Arg Pro Gly Asp Ala Phe Ser Val Asp Phe Gly Lys Glu Leu Asp Val  
 225 230 235 240  
 Ile Leu Leu Thr Asn Phe Leu His His Phe Asp Glu Ala Gly Cys Glu  
 245 250 255  
 Lys Ile Leu Lys Lys Ala His Ala Ala Leu Lys Glu Gly Gly Arg Val  
 260 265 270  
 Leu Thr Phe Glu Phe Ile Ala Asn Glu Asp Arg Thr Ser Pro Pro Leu  
 275 280 285  
 Ala Ala Thr Phe Ser Met Met Met Leu Gly Thr Thr Pro Gly Gly Glu  
 290 295 300  
 Thr Tyr Ala Tyr Ser Asp Leu Glu Arg Met Phe Lys Asn Thr Gly Tyr  
 305 310 315 320  
 Asp Gln Val Glu Leu Lys Ala Ile Pro Pro Ala Met Glu Lys Val Val  
 325 330 335  
 Val Ser Ile Lys Gly Lys Ala Gln Leu  
 340 345

<210> 46  
 <211> 1992  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (1992)  
 <223> CysG  
  
 <400> 46  
 Met Ala Thr Lys Leu Ser Asp Phe Ala Leu Leu Asp Ser Glu Asp Ala  
 1 5 10 15  
 Asn Val Ile Ser Arg Ser Asn Glu Thr Gly Ile Ser Leu Asp Leu Ser  
 20 25 30  
 Lys Ser Val Val Asp Leu Phe Asn Leu Gln Val Glu Arg Ala Pro Asp  
 35 40 45  
 Ala Thr Ala Cys Leu Gly Arg Gln Gly Arg Leu Thr Tyr Gly Glu Leu  
 50 55 60  
 Asn Arg Arg Thr Asn Gln Leu Ala His His Leu Ile Ala Arg Gly Val  
 65 70 75 80  
 Gly Pro Asp Val Pro Val Gly Val Leu Phe Glu Arg Ser Ala Glu Gln  
 85 90 95  
 Leu Ile Ala Ile Leu Gly Val Leu Lys Ala Gly Gly Cys Tyr Val Pro  
 100 105 110  
 Leu Asp Pro Gln Tyr Pro Ala Asp Tyr Met Gln Gln Val Leu Thr Asp  
 115 120 125  
 Ala Arg Pro Arg Met Val Val Ser Ser Arg Ala Leu Gly Glu Arg Leu  
 130 135 140  
 Arg Ser Gly Glu Glu Gln Ile Val Tyr Leu Asp Asp Glu Gln Leu Leu  
 145 150 155 160  
 Ala Arg Glu Thr Arg Asp Pro Pro Val Lys Val Leu Pro Glu Gln Leu  
 165 170 175  
 Ala Tyr Val Met Tyr Thr Ser Gly Ser Ser Gly Val Pro Lys Gly Val  
 180 185 190  
 Met Val Pro His Arg Gln Ile Leu Asn Trp Leu His Ala Leu Leu Ala  
 195 200 205  
 Arg Val Pro Phe Gly Glu Asn Glu Val Val Ala Gln Lys Thr Ser Thr  
 210 215 220  
 Ser Phe Ala Ile Ser Val Lys Glu Leu Phe Ala Gly Leu Val Ala Gly  
 225 230 235 240  
 Val Pro Gln Val Phe Ile Asp Asp Ala Thr Val Arg Asp Val Ala Ser  
 245 250 255  
 Phe Val Arg Glu Leu Glu Gln Trp Arg Val Thr Arg Leu Tyr Thr Phe  
 260 265 270  
 Pro Ser Gln Leu Ala Ala Ile Leu Ser Ser Val Asn Gly Ala Tyr Glu  
 275 280 285  
 Arg Leu Arg Ser Leu Arg His Leu Tyr Ile Ser Ile Glu Pro Cys Pro  
 290 295 300  
 Thr Glu Leu Leu Ala Lys Leu Arg Ala Ala Met Pro Trp Val Thr Pro  
 305 310 315 320  
 Trp Tyr Ile Tyr Gly Cys Thr Glu Ile Asn Asp Val Thr Tyr Cys Asp  
 325 330 335  
 Pro Gly Asp Gln Ala Gly Asn Thr Gly Phe Val Pro Ile Gly Arg Pro  
 340 345 350  
 Ile Arg Asn Thr Arg Val Phe Val Leu Asp Glu Glu Leu Arg Met Val  
 355 360 365  
 Pro Val Gly Ala Met Gly Glu Met Tyr Val Glu Ser Leu Ser Thr Ala  
 370 375 380  
 Arg Gly Tyr Trp Gly Leu Pro Glu Leu Thr Ala Glu Arg Phe Ile Ala  
 385 390 395 400  
 Asn Pro His Ala Glu Asp Gly Ser Arg Leu Tyr Lys Thr Gly Asp Leu  
 405 410 415  
 Ala Arg Tyr Leu Pro Asp Gly Ser Leu Glu Phe Leu Gly Arg Arg Asp  
 420 425 430  
 Tyr Glu Val Lys Ile Arg Gly Tyr Arg Val Asp Val Arg Gln Val Glu  
 435 440 445  
 Lys Val Leu Gly Ala His Pro Asp Ile Leu Glu Val Ala Val Val Gly  
 450 455 460  
 Trp Pro Leu Gly Gly Ala Asn Pro Gln Leu Val Ala Tyr Val Val Pro  
 465 470 475 480  
 Arg Ala Lys Gly Ala Ala Pro Ile Gln Glu Ile Arg Asp Tyr Leu Ser  
 485 490 495  
 Ala Ser Leu Pro Ala Tyr Met Val Pro Thr Ile Phe Gln Val Leu Ala  
 500 505 510

[0040]

[0041]

```

Ala Leu Pro Arg Leu Pro Asn Asp Lys Val Asp Arg Leu Ser Leu Pro
515 520 525
Asp Pro Lys Val Glu Glu Gln Thr Glu Gly Tyr Val Ala Pro Arg Thr
530 535 540
Glu Thr Glu Lys Val Leu Ala Glu Ile Trp Ser Asp Val Leu Ser Gln
545 550 555 560
Gly Arg Ala Pro Leu Thr Val Gly Ala Thr His Asn Phe Phe Glu Leu
565 570 575
Gly Gly His Ser Leu Leu Ala Ala Gln Met Phe Ser Arg Ile Arg Gln
580 585 590
Lys Phe Asp Leu Glu Leu Pro Ile Asn Thr Leu Phe Glu Thr Pro Val
595 600 605
Leu Glu Gly Phe Ala Ser Ala Val Asp Ala Ala Leu Ala Glu Arg Asn
610 615 620
Gly Pro Ala Gln Arg Leu Ile Ser Met Thr Asp Arg Gly Gln Ala Leu
625 630 635 640
Pro Leu Ser His Val Gln Glu Arg Leu Trp Phe Val His Glu His Met
645 650 655
Val Glu Gln Arg Ser Ser Tyr Asn Val Ala Phe Ala Cys His Met Arg
660 665 670
Gly Lys Gly Leu Ser Met Pro Ala Leu Arg Ala Ala Ile Asn Gly Leu
675 680 685
Val Ala Arg His Glu Thr Leu Arg Thr Thr Phe Val Val Ser Glu Gly
690 695 700
Gly Gly Asp Pro Val Gln Arg Ile Ala Asp Ser Leu Trp Ile Glu Val
705 710 715 720
Pro Leu Tyr Glu Val Asp Ala Ser Glu Val Pro Ala Arg Met Ala Ala
725 730 735
His Ala Gly His Val Phe Asp Leu Ala Lys Gly Pro Leu Leu Lys Thr
740 745 750
Ser Val Leu Arg Val Thr Pro Asp His His Val Phe Leu Met Asn Met
755 760 765
His His Ile Ile Cys Asp Gly Trp Ser Ile Asp Ile Leu Leu Arg Asp
770 775 780
Leu Tyr Glu Phe Tyr Lys Ala Ala Glu Thr Gly Ser Gln Pro Asn Leu
785 790 795 800
Pro Val Leu Pro Ile Gln Tyr Ala Asp Tyr Ser Val Trp Gln Arg Gln
805 810 815
Gln Asp Leu Ser Ser His Leu Asp Tyr Trp Lys Lys Thr Leu Glu Gly
820 825 830
Tyr Gln Glu Gly Leu Ser Leu Pro Tyr Asp Phe Ala Arg Pro Ser Asn
835 840 845
Arg Thr Trp Arg Ala Ala Ser Val Arg His Gln Tyr Pro Ala Glu Leu
850 855 860
Ala Thr Arg Leu Ser Glu Val Ser Lys Ser His Gln Ala Thr Val Phe
865 870 875 880
Met Thr Leu Met Ala Ser Thr Ala Ile Val Leu Asn Arg Tyr Thr Gly
885 890 895
Arg Asp Asp Leu Cys Val Gly Ala Thr Val Ala Gly Arg Asp His Phe
900 905 910
Glu Leu Glu Asn Leu Ile Gly Phe Phe Val Asn Ile Leu Ala Ile Arg
915 920 925
Leu Asp Leu Ser Gly Asn Pro Thr Ala Glu Thr Val Leu Gln Arg Ala
930 935 940
Arg Ala Gln Val Leu Glu Gly Met Lys His Arg Asp Leu Pro Phe Glu
945 950 955 960
His Ile Leu Ala Ala Leu Gln Lys Gln Arg Asp Ser Ser Gln Ile Pro
965 970 975
Leu Val Pro Val Met Val Arg His Gln Asn Phe Pro Thr Val Thr Ser
980 985 990
Gln Glu Gln Gly Leu Asp Leu Gly Ile Gly Glu Ile Glu Phe Gly Glu
995 1000 1005
Arg Thr Thr Pro Asn Glu Leu Asp Ile Gln Phe Ile Gly Glu Gly
1010 1015 1020
Ser Thr Leu Glu Val Val Val Glu Tyr Ala Lys Asp Leu Phe Ser
1025 1030 1035
Glu Arg Thr Ile Gln Arg Leu Ile Thr His Leu Gln Gln Val Leu
1040 1045 1050
Gln Thr Leu Val Asp Lys Pro Asp Cys Arg Leu Thr Asp Phe Pro
1055 1060 1065
Leu Val Ala Gly Asp Ala Leu Gln Gly Gly Val Ser Gly Ser Gly

```

[0042]

1070	Gly	Ala	Thr	Lys	Thr	Gly	1075	Lys	Leu	Asp	Val	Ser	1080	Lys	Ser	Pro	Val
1085	Glu	Leu	Phe	Asn	Glu	Arg	1090	Val	Glu	Ala	Ser	Pro	1095	Asp	Ala	Val	Ala
1100	Cys	Met	Gly	Ala	Asp	Gly	1105	Ser	Leu	Thr	Tyr	Arg	1110	Glu	Leu	Asp	Arg
1115	Arg	Ala	Asn	Gln	Val	Ala	1120	Arg	His	Leu	Met	Gly	1125	Arg	Gly	Val	Gly
1130	Arg	Glu	Thr	Arg	Val	Gly	1135	Leu	Trp	Phe	Glu	Arg	1140	Ser	Pro	Asp	Leu
1145	Leu	Val	Ala	Leu	Leu	Gly	1150	Ile	Leu	Lys	Ala	Gly	1155	Gly	Cys	Phe	Val
1160	Pro	Leu	Asp	Pro	Ser	Tyr	1165	Pro	Gln	Glu	Tyr	Ile	1170	Asn	Asn	Ile	Val
1175	Ala	Asp	Ala	Gln	Pro	Leu	1180	Leu	Val	Met	Ser	Ser	1185	Arg	Ala	Leu	Gly
1190	Ser	Arg	Leu	Ser	Leu	Glu	1195	Ala	Gly	Arg	Leu	Val	1200	Tyr	Leu	Asp	Asp
1205	Ala	Leu	Ala	Ala	Ser	Thr	1210	Asp	Ala	Ser	Asp	Pro	1215	Gln	Val	Arg	Ile
1220	Asp	Pro	Glu	Gln	Leu	Ile	1225	Tyr	Val	Met	Tyr	Thr	1230	Ser	Gly	Ser	Thr
1235	Gly	Leu	Pro	Lys	Gly	Val	1240	Leu	Val	Pro	His	Arg	1245	Gln	Ile	Leu	Asn
1250	Trp	Leu	Tyr	Pro	Leu	Trp	1255	Ala	Met	Val	Pro	Phe	1260	Gly	Gln	Asp	Glu
1265	Val	Val	Ala	Gln	Lys	Thr	1270	Ser	Thr	Ala	Phe	Ala	1275	Val	Ser	Met	Lys
1280	Glu	Leu	Phe	Thr	Gly	Leu	1285	Leu	Ala	Gly	Val	Pro	1290	Gln	Val	Phe	Ile
1295	Asp	Gly	Thr	Val	Val	Lys	1300	Asp	Ala	Ala	Ala	Phe	1305	Val	Leu	His	Leu
1310	Glu	Arg	Trp	Arg	Val	Thr	1315	Arg	Leu	Tyr	Thr	Leu	1320	Pro	Ser	His	Leu
1325	Asp	Ala	Ile	Leu	Ser	His	1330	Val	Asp	Gly	Ala	Ala	1335	Glu	Arg	Leu	Arg
1340	Ser	Leu	Arg	His	Val	Ile	1345	Leu	Ala	Gly	Glu	Pro	1350	Cys	Pro	Val	Glu
1355	Leu	Met	Glu	Lys	Leu	Arg	1360	Glu	Thr	Leu	Pro	Ser	1365	Cys	Thr	Ala	Trp
1370	Phe	Asn	Tyr	Gly	Cys	Thr	1375	Glu	Val	Asn	Asp	Ile	1380	Ser	Tyr	Cys	Val
1385	Pro	Asn	Glu	Gln	Phe	His	1390	Ser	Ser	Gly	Phe	Val	1395	Pro	Ile	Gly	Arg
1400	Pro	Ile	Gln	Tyr	Thr	Arg	1405	Ala	Leu	Val	Leu	Asp	1410	Asp	Glu	Leu	Arg
1415	Thr	Val	Pro	Val	Gly	Ile	1420	Met	Gly	Glu	Ile	Tyr	1425	Val	Glu	Ser	Pro
1430	Gly	Thr	Ala	Arg	Gly	Tyr	1435	Trp	Arg	Gln	Pro	Asp	1440	Leu	Thr	Ala	Glu
1445	Arg	Phe	Ile	Pro	Asn	Pro	1450	Arg	Gly	Glu	Pro	Gly	1455	Ser	Arg	Leu	Tyr
1460	Arg	Thr	Gly	Asp	Met	Ala	1465	Arg	Cys	Leu	Glu	Asp	1470	Gly	Ser	Leu	Glu
1475	Phe	Leu	Gly	Arg	Arg	Asp	1480	Tyr	Glu	Val	Lys	Ile	1485	Arg	Gly	His	Arg
1490	Val	Asp	Val	Arg	Gln	Val	1495	Glu	Lys	Ile	Leu	Ala	1500	Ser	His	Pro	Glu
1505	Val	Leu	Glu	Ser	Ala	Val	1510	Leu	Gly	Trp	Pro	Arg	1515	Gly	Ala	Lys	Asn
1520	Pro	Gln	Leu	Leu	Ala	Tyr	1525	Ala	Ala	Thr	Lys	Pro	1530	Gly	Arg	Pro	Leu
1535	Ser	Thr	Glu	Asn	Val	Arg	1540	Glu	Tyr	Leu	Ser	Ala	1545	Arg	Leu	Pro	Thr
1550	Tyr	Met	Val	Pro	Thr	Leu	1555	Tyr	Gln	Phe	Leu	Pro	1560	Ala	Leu	Pro	Arg
1565	Leu	Pro	Asn	Gly	Lys	Leu	1570	Asp	Arg	Phe	Gly	Leu	1575	Pro	Asp	His	Lys
1580	Lys	Val	Glu	Val	Gly	Gly	1585	Val	Tyr	Val	Ala	Pro	1590	Gln	Thr	Pro	Thr
1595							1600						1605				

[0043]

Glu Lys Val Leu Ala Gly Leu Trp Ala Glu Cys Leu Lys Gln Gly  
 1610 1615 1620  
 Asp Met Pro Ala Pro Gln Val Gly Arg Leu His Asn Phe Phe Asp  
 1625 1630 1635  
 Leu Gly Gly His Ser Leu Leu Ala Asn Arg Val Leu Met Gln Val  
 1640 1645 1650  
 Gln Arg His Phe Gly Val Ser Leu Gly Ile Ser Ala Leu Phe Gly  
 1655 1660 1665  
 Ser Pro Val Leu Asn Asp Phe Ala Ala Ala Ile Asp Lys Ala Leu  
 1670 1675 1680  
 Gly Thr Glu Glu Pro Gly Glu Glu Gly Ser Ser Asp Ala Arg Glu  
 1685 1690 1695  
 Val Ala Ala Lys Asp Thr Ser Val Leu Val Pro Leu Ser Thr His  
 1700 1705 1710  
 Gly Thr Leu Pro Ser Leu Phe Cys Val His Pro Val Gly Gly Gln  
 1715 1720 1725  
 Val His Ala Tyr Arg Glu Leu Ala Gln Ala Met Glu Lys His Ala  
 1730 1735 1740  
 Ser Met Tyr Ala Leu Gln Ser Glu Gly Ala Arg Glu Phe Asp Thr  
 1745 1750 1755  
 Ile Glu Thr Leu Ala Arg Phe Tyr Ala Asp Ala Ile Arg Gly Ala  
 1760 1765 1770  
 Gln Pro Asp Gly Ser Tyr Arg Leu Leu Gly Trp Ser Ser Gly Gly  
 1775 1780 1785  
 Leu Ile Thr Leu Ala Ile Ala Arg Glu Leu Glu His Gln Gly Cys  
 1790 1795 1800  
 Ala Val Glu Tyr Val Gly Leu Val Asp Ser Lys Pro Ile Pro Arg  
 1805 1810 1815  
 Leu Ala Gly Glu Arg Gly Trp Ala Ser Leu Ile Ala Ala Thr Asn  
 1820 1825 1830  
 Ile Leu Gly Ala Met Arg Gly Arg Gly Phe Ser Val Ala Glu Val  
 1835 1840 1845  
 Asp Ala Ala Gly Lys Ile Leu Glu Ser Arg Gly Trp Thr Glu Glu  
 1850 1855 1860  
 Ser Phe Asp Ser Glu Gly His Ala Ala Leu Glu Glu Leu Ala Arg  
 1865 1870 1875  
 His Phe Gly Ile Thr Val Ala Gln Glu Ser Ser Glu Tyr Leu Leu  
 1880 1885 1890  
 Ala Arg Phe Lys Thr Thr Lys Tyr Tyr Leu Ser Leu Phe Ala Gly  
 1895 1900 1905  
 Phe Lys Pro Ala Ala Leu Gly Pro Glu Thr Tyr Leu Tyr Glu Ala  
 1910 1915 1920  
 Ser Glu Arg Val Gly Ala Thr Ser Asn Asp Asp Thr Gly Glu Trp  
 1925 1930 1935  
 Gly Asp Ala Leu Asp Arg Lys Ala Leu Arg Ala Asn Ile Val Gln  
 1940 1945 1950  
 Val Pro Gly Asn His Tyr Thr Val Leu Gln Gly Glu Asn Val Leu  
 1955 1960 1965  
 Gln Leu Ala Gly Arg Ile Ala Glu Ala Leu Ser Ala Ile Asp Asn  
 1970 1975 1980  
 Ser Val Val Thr Arg Thr Arg Ala Ser  
 1985 1990

<210> 47  
 <211> 975  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (975)  
 <223> CysH

<400> 47  
 Met Asp Asn Arg Glu Ile Ala Pro Thr Gln Ser Ala Arg Thr Arg Asp  
 1 5 10 15  
 Ala Tyr Thr Ala Val Pro Pro Ala Lys Ala Glu Tyr Pro Ser Asp Val  
 20 25 30  
 Cys Val His Gln Leu Phe Glu Leu Gln Ala Asp Arg Ile Pro Asp Ala  
 35 40 45  
 Val Ala Ala Arg Ala Gly Asn Glu Ser Leu Thr Tyr Arg Glu Leu Asn

[0044]

50	Phe	Arg	Ala	Asn	Gln	Leu	Ala	Arg	Tyr	Leu	Val	Ala	Lys	Gly	Val	Val
65	Pro	Arg	Gly	Ser	Val	Ala	Val	Leu	Met	Asn	Arg	Thr	Pro	Ala	Cys	Leu
				85						90					95	
	Val	Ser	Leu	Leu	Ala	Ile	Ile	Lys	Ala	Gly	Ala	Ala	Tyr	Val	Pro	Val
				100					105					110		
	Asp	Ala	Gly	Leu	Pro	Ala	Lys	Arg	Val	Asp	Tyr	Ile	Leu	Thr	Asp	Ser
				115					120					125		
	Gly	Ala	Thr	Cys	Val	Leu	Thr	Asp	Arg	Glu	Thr	Arg	Ser	Leu	Leu	Asp
				130					135					140		
	Glu	Pro	Arg	Ser	Ala	Ser	Thr	Leu	Val	Ile	Asp	Val	Asp	Asp	Pro	Ser
	145					150					155				160	
	Ile	Tyr	Ser	Gly	Glu	Thr	Ser	Asn	Leu	Gly	Leu	Ala	Val	Asp	Pro	Glu
				165						170					175	
	Gln	Gln	Val	Tyr	Cys	Ile	Tyr	Thr	Ser	Gly	Ser	Thr	Gly	Leu	Pro	Lys
				180						185					190	
	Gly	Val	Met	Val	Gln	His	Arg	Ala	Leu	Met	Asn	Tyr	Val	Trp	Trp	Ala
				195					200					205		
	Lys	Lys	Gln	Tyr	Val	Thr	Asp	Ala	Val	Glu	Ser	Phe	Ala	Leu	Tyr	Ser
	210						215					220				
	Ser	Leu	Ser	Phe	Asp	Leu	Thr	Val	Thr	Ser	Ile	Phe	Val	Pro	Leu	Ile
	225					230					235				240	
	Ser	Gly	Arg	Cys	Ile	Asp	Val	Tyr	Pro	Asp	Leu	Gly	Glu	Asp	Val	Pro
				245						250					255	
	Val	Ile	Asn	Arg	Val	Leu	Glu	Asp	Asn	Lys	Val	Asp	Val	Val	Lys	Leu
				260					265					270		
	Thr	Pro	Ala	His	Leu	Ala	Leu	Leu	Arg	Asn	Thr	Asp	Leu	Ser	Gln	Ser
				275					280					285		
	Arg	Leu	Lys	Val	Leu	Ile	Leu	Gly	Gly	Glu	Asp	Leu	Arg	Ala	Glu	Thr
	290						295					300				
	Ala	Gly	Asp	Val	His	Lys	Arg	Leu	Asp	Gly	Arg	Ala	Val	Ile	Tyr	Asn
	305					310					315				320	
	Glu	Tyr	Gly	Pro	Thr	Glu	Thr	Val	Val	Gly	Cys	Met	Ile	His	Arg	Tyr
				325						330					335	
	Asp	Pro	Ala	Val	Asp	Leu	His	Gly	Ser	Val	Pro	Ile	Gly	Val	Gly	Ile
				340					345					350		
	Asp	Asn	Met	Arg	Ile	Tyr	Leu	Leu	Asp	Asp	Arg	Arg	Arg	Pro	Val	Lys
				355					360					365		
	Pro	Gly	Glu	Val	Gly	Glu	Ile	Tyr	Ile	Gly	Gly	Asp	Gly	Val	Thr	Leu
				370					375				380			
	Gly	Tyr	Lys	Asp	Lys	Pro	Gln	Val	Thr	Ala	Asp	His	Phe	Ile	Ser	Asn
	385					390					395				400	
	Pro	Phe	Val	Glu	Gly	Glu	Arg	Leu	Tyr	Ala	Ser	Gly	Asp	Leu	Gly	Arg
				405						410					415	
	Val	Asn	Glu	Arg	Gly	Ala	Leu	Val	Phe	Leu	Gly	Arg	Lys	Asp	Leu	Gln
				420					425					430		
	Ile	Lys	Leu	Arg	Gly	Tyr	Arg	Ile	Glu	Leu	Gly	Glu	Ile	Glu	Ser	Ala
				435					440					445		
	Leu	Leu	Ser	Tyr	Pro	Gly	Ile	Lys	Glu	Cys	Ile	Val	Asp	Ser	Thr	Lys
	450						455					460				
	Thr	Ala	Gln	Ser	Gln	Ala	Ala	Ala	Gln	Leu	Thr	Tyr	Cys	Thr	Lys	Cys
	465					470					475				480	
	Gly	Leu	Ala	Ser	Ser	Phe	Pro	Asn	Thr	Thr	Tyr	Ser	Ala	Glu	Gly	Val
				485						490					495	
	Cys	Asn	His	Cys	Glu	Ala	Phe	Asp	Lys	Tyr	Arg	Ser	Val	Val	Asp	Asp
				500					505					510		
	Tyr	Phe	Ser	Thr	Met	Asp	Glu	Leu	Gln	Ser	Ile	Val	Thr	Glu	Met	Lys
				515					520					525		
	Ser	Ile	His	Asn	Ser	Lys	Tyr	Asp	Cys	Ile	Val	Ala	Leu	Ser	Gly	Gly
				530					535					540		
	Lys	Asp	Ser	Thr	Tyr	Ala	Leu	Cys	Arg	Met	Ile	Glu	Thr	Gly	Ala	Arg
	545					550					555				560	
	Val	Leu	Ala	Phe	Thr	Leu	Asp	Asn	Gly	Tyr	Ile	Ser	Glu	Glu	Ala	Lys
				565						570					575	
	Gln	Asn	Ile	Asn	Arg	Val	Val	Ala	Arg	Leu	Gly	Val	Asp	His	Arg	Tyr
				580					585					590		
	Leu	Ser	Thr	Gly	His	Met	Lys	Glu	Ile	Phe	Val	Asp	Ser	Leu	Lys	Arg
				595					600					605		
	His	Ser	Asn	Val	Cys	Asn	Gly	Cys	Phe	Lys	Thr	Ile	Tyr	Thr	Phe	Ala
				610					615					620		

[0045]

```

Ile Asn Leu Ala Gln Glu Val Gly Val Lys His Val Val Met Gly Leu
625                      630                      635                      640
Ser Lys Gly Gln Leu Phe Glu Thr Arg Leu Ser Ala Leu Phe Arg Thr
                      645                      650                      655
Ser Thr Phe Asp Asn Ala Ala Phe Glu Lys Ser Leu Val Asp Ala Arg
                      660                      665                      670
Lys Ile Tyr His Arg Ile Asp Asp Ala Val Ser Arg Leu Leu Asp Thr
                      675                      680                      685
Thr Cys Val Lys Asn Asp Lys Val Ile Glu Asn Ile Arg Phe Val Asp
                      690                      695                      700
Phe Tyr Arg Tyr Cys His Ala Ser Arg Gln Glu Met Tyr Asp Tyr Ile
705                      710                      715                      720
Gln Glu Arg Val Gly Trp Ala Arg Pro Ile Asp Thr Gly Arg Ser Thr
                      725                      730                      735
Asn Cys Leu Leu Asn Asp Val Gly Ile Tyr Val His Asn Lys Glu Arg
                      740                      745                      750
Arg Tyr His Asn Tyr Ser Leu Pro Tyr Ser Trp Asp Val Arg Met Gly
                      755                      760                      765
His Ile Ser Arg Glu Glu Ala Met Arg Glu Leu Asp Asp Ser Ala Asp
                      770                      775                      780
Ile Asp Val Glu Arg Val Glu Gly Ile Ile Lys Asp Leu Gly Tyr Glu
785                      790                      795                      800
Leu Asn Asp Gln Val Val Gly Ser Ala Glu Ala Gln Leu Val Ala Tyr
                      805                      810                      815
Tyr Val Ser Ala Glu Glu Phe Pro Ala Ser Asp Leu Arg Gln Phe Leu
                      820                      825                      830
Ser Glu Ile Leu Pro Glu Tyr Met Val Pro Arg Ser Phe Val Gln Leu
                      835                      840                      845
Asp Ser Ile Pro Leu Thr Pro Asn Gly Lys Val Asn Arg Gln Ala Leu
                      850                      855                      860
Pro Lys Pro Asp Leu Leu Arg Lys Ala Gly Thr Asp Gly Gln Ala Ala
865                      870                      875                      880
Pro Arg Thr Pro Val Glu Lys Gln Leu Ala Glu Leu Trp Lys Glu Val
                      885                      890                      895
Leu Gln Val Asp Ser Val Gly Ile His Asp Asn Phe Phe Glu Met Gly
                      900                      905                      910
Gly His Ser Leu Pro Ala Leu Met Leu Leu Tyr Lys Ile Asp Ser Gln
                      915                      920                      925
Phe His Lys Thr Ile Ser Ile Gln Glu Phe Ser Lys Val Pro Thr Ile
                      930                      935                      940
Ser Ala Leu Ala Ala His Leu Gly Ser Asp Thr Glu Ala Val Pro Pro
945                      950                      955                      960
Gly Leu Gly Glu Val Val Asp Gln Ser Ala Pro Ala Tyr Arg Gly
                      965                      970                      975

```

```

<210> 48
<211> 272
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC FEATURE
<222> (1)..(272)
<223> CysI

```

```

<400> 48
Val Arg Phe Val Thr Val Asn Gly Glu Asp Ser Ala Val Cys Ser Val
1                      5                      10                      15
Leu Asp Arg Gly Leu Gln Phe Gly Asp Gly Leu Phe Glu Thr Met Leu
                      20                      25                      30
Cys Val Gly Gly Ala Pro Val Asp Phe Pro Glu His Trp Ala Arg Leu
                      35                      40                      45
Asp Glu Gly Cys Arg Arg Leu Gly Ile Glu Cys Pro Asp Ile Arg Arg
                      50                      55                      60
Glu Val Thr Ala Ala Ile Ala Arg Trp Gly Ala Pro Arg Ala Val Ala
65                      70                      75                      80
Lys Leu Val Val Thr Arg Gly Ser Thr Glu Arg Gly Tyr Arg Cys Ala
                      85                      90                      95
Pro Ser Val Arg Pro Asn Trp Ile Leu Thr Ile Thr Asp Ala Pro Lys
                      100                     105                     110
Tyr Pro Leu Ala His Glu Asp Arg Gly Val Ala Val Lys Leu Cys Arg

```



```

      115      120      125
Thr Leu Val Ser Leu Asp Asp Pro Gln Leu Ala Gly Leu Lys His Leu
      130      135      140
Asn Arg Leu Pro Gln Val Leu Ala Arg Arg Glu Trp Asp Asp Glu Tyr
145      150      155      160
His Asp Gly Leu Leu Thr Asp His Gly Gly His Leu Val Glu Gly Cys
      165      170      175
Thr Ser Asn Leu Phe Leu Val Ala Asp Gly Ala Leu Arg Thr Pro Asp
      180      185      190
Leu Thr Ala Cys Gly Val Arg Gly Ile Val Arg Gln Lys Val Leu Asp
      195      200      205
His Ser Lys Ala Ile Gly Ile Arg Cys Glu Val Thr Thr Leu Lys Leu
210      215      220
Arg Asp Leu Glu His Ala Asp Glu Val Phe Leu Thr Asn Ser Val Tyr
225      230      235      240
Gly Ile Val Pro Val Gly Ser Val Asp Gly Met Arg Tyr Arg Ile Gly
      245      250      255
Pro Thr Thr Ala Arg Leu Leu Lys Asp Leu Cys Gln Gly Val Tyr Phe
      260      265      270

```

```

<210> 49
<211> 327
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC_FEATURE
<222> (1).. (327)
<223> CysJ

```

[0046]

```

<400> 49
Met Thr Gly Asn Leu Asp Ser Ala Ala Trp Pro Val Ile Ile Thr Pro
1      5      10      15
Gly Gln Gln Pro Ala Ala Leu Glu Asp Trp Val Ser Ala Asn Arg Asp
      20      25      30
Gly Leu Glu Arg Gln Leu Thr Glu Cys Lys Ala Ile Leu Phe Arg Gly
      35      40      45
Phe Arg Ser Arg Asn Gly Phe Glu Ser Ile Ala Asn Ser Phe Phe Asp
      50      55      60
Arg Arg Leu Asn Tyr Thr Tyr Arg Ser Thr Pro Arg Thr Asp Leu Gly
65      70      75      80
Gln Asn Leu Tyr Thr Ala Thr Glu Tyr Pro Lys Gln Leu Ser Ile Pro
      85      90      95
Gln His Cys Glu Asn Ala Tyr Gln Arg Asp Trp Pro Met Lys Leu Leu
      100      105      110
Phe His Cys Val Glu Pro Ala Ser Lys Gly Gly Arg Thr Pro Leu Ala
      115      120      125
Asp Met Thr Lys Val Thr Ala Met Ile Pro Ala Glu Ile Lys Glu Glu
      130      135      140
Phe Ala Arg Lys Lys Val Gly Tyr Val Arg Asn Tyr Arg Ala Gly Val
145      150      155      160
Asp Leu Pro Trp Glu Glu Val Phe Gly Thr Ser Asn Lys Ala Glu Val
      165      170      175
Glu Lys Phe Cys Val Glu Asn Gly Ile Glu Tyr His Trp Thr Glu Gly
      180      185      190
Gly Leu Lys Thr Ile Gln Val Cys Gln Ala Phe Ala Ser His Pro Leu
      195      200      205
Thr Gly Glu Thr Ile Trp Phe Asn Gln Ala His Leu Phe His Leu Ser
210      215      220
Ala Leu Asp Pro Ala Ser Gln Lys Met Met Leu Ser Phe Phe Gly Glu
225      230      235      240
Gly Gly Leu Pro Arg Asn Ser Tyr Phe Gly Asp Gly Ser Ala Ile Gly
      245      250      255
Ser Asp Val Leu Asp Gln Ile Arg Ser Ala Tyr Glu Arg Asn Lys Val
      260      265      270
Ser Phe Glu Trp Gln Lys Asp Asp Val Leu Leu Ile Asp Asn Met Leu
275      280      285
Val Ser His Gly Arg Asp Pro Phe Glu Gly Ser Arg Arg Val Leu Val
290      295      300
Cys Met Ala Glu Pro Tyr Ser Glu Val Gln Arg Arg Gly Phe Ala Gly
305      310      315      320

```

Ala Thr Asn Ser Gly Arg Ser  
325

<210> 50  
<211> 4545  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1).. (4140)  
<223> CysK

<220>  
<221> MISC\_FEATURE  
<222> (1).. (4545)  
<223> CysK

<400> 50  
Met Leu Leu Glu Gly Glu Leu Glu Gly Tyr Glu Asp Gly Leu Glu Leu  
1 5 10 15  
Pro Tyr Asp Phe Pro Arg Thr Ser Asn Arg Ala Trp Arg Ala Ala Thr  
20 25 30  
Phe Gln His Ser Tyr Pro Pro Glu Leu Ala Arg Lys Val Ala Glu Leu  
35 40 45  
Ser Arg Glu Gln Gln Ser Thr Leu Phe Met Ser Leu Val Ala Ser Leu  
50 55 60  
Ala Val Val Leu Asn Arg Tyr Thr Gly Arg Glu Asp Val Cys Ile Gly  
65 70 75 80  
Thr Thr Val Ala Gly Arg Ala Gln Val Gly Ala Leu Gly Asp Leu Ser  
85 90 95  
Gly Ser Thr Val Asp Ile Leu Pro Leu Arg Leu Asp Leu Ser Gly Ala  
100 105 110  
Pro Ser Leu His Glu Val Leu Arg Arg Thr Lys Ala Val Val Leu Glu  
115 120 125  
Gly Phe Glu His Glu Ala Leu Pro Cys Gln Ile Pro Leu Val Pro Val  
130 135 140  
Val Val Arg His Gln Asn Phe Pro Met Ala Arg Leu Glu Gly Trp Ser  
145 150 155 160  
Glu Gly Val Glu Leu Lys Lys Phe Glu Leu Ala Gly Glu Arg Thr Thr  
165 170 175  
Ala Ser Glu Gln Asp Trp Gln Phe Phe Gly Asp Gly Ser Ser Leu Glu  
180 185 190  
Leu Ser Leu Glu Tyr Ala Ala Glu Leu Phe Ser Glu Lys Thr Val Lys  
195 200 205  
Arg Met Val Glu His His Gln Arg Val Leu Glu Ala Leu Val Glu Gly  
210 215 220  
Leu Glu Glu Val Arg Leu His Glu Val Arg Leu Leu Thr Glu Glu Glu  
225 230 235 240  
Glu Gly Leu His Gly Arg Leu Asn Asp Thr Ala Arg Glu Leu Glu Glu  
245 250 255  
Arg Trp Ser Leu Ala Glu Thr Phe Glu Arg Gln Val Arg Glu Thr Pro  
260 265 270  
Glu Ala Val Ala Cys Val Gly Val Glu Val Ala Thr Gly Gly His Ser  
275 280 285  
Arg Pro Thr Tyr Arg Gln Leu Thr Tyr Arg Gln Leu Asn Ala Arg Ala  
290 295 300  
Asn Gln Val Ala Arg Arg Leu Arg Ala Leu Gly Val Gly Ala Glu Thr  
305 310 315 320  
Arg Val Ala Val Leu Ser Asp Arg Ser Pro Glu Leu Leu Val Ala Met  
325 330 335  
Leu Ala Ile Phe Lys Ala Gly Gly Cys Tyr Val Pro Val Asp Pro Gln  
340 345 350  
Tyr Pro Gly Ser Tyr Ile Glu Gln Ile Leu Glu Asp Ala Ala Pro Gln  
355 360 365  
Val Val Leu Gly Lys Arg Gly Arg Ala Asp Gly Val Arg Val Asp Val  
370 375 380  
Trp Leu Glu Leu Asp Gly Ala Gln Arg Leu Thr Asp Glu Ala Leu Ala  
385 390 395 400  
Ala Gln Glu Glu Gly Glu Leu Glu Gly Ala Glu Arg Pro Glu Ser Gln  
405 410 415

[0047]

[0048]

Gln Leu Ala Cys Leu Met Tyr Thr Ser Gly Ser Thr Gly Arg Pro Lys  
 420 425 430  
 Gly Val Met Val Pro Tyr Ser Gln Leu His Asn Trp Leu Glu Ala Gly  
 435 440 445  
 Lys Glu Arg Ser Pro Leu Glu Arg Gly Glu Val Met Leu Gln Lys Thr  
 450 455 460  
 Ala Ile Ala Phe Ala Val Ser Val Lys Glu Leu Leu Ser Gly Leu Leu  
 465 470 475 480  
 Ala Gly Val Ala Gln Val Met Val Pro Glu Thr Leu Val Lys Asp Ser  
 485 490 495  
 Val Ala Leu Ala Gln Glu Ile Glu Arg Trp Arg Val Thr Arg Ile His  
 500 505 510  
 Leu Val Pro Ser His Leu Gly Ala Leu Leu Glu Gly Ala Gly Glu Glu  
 515 520 525  
 Ala Lys Gly Leu Arg Ser Leu Lys Tyr Val Ile Thr Ala Gly Glu Ala  
 530 535 540  
 Leu Ala Gln Gly Val Arg Glu Glu Ala Arg Arg Lys Leu Pro Gly Ala  
 545 550 555 560  
 Gln Leu Trp Asn Asn Tyr Gly Cys Thr Glu Leu Asn Asp Val Thr Tyr  
 565 570 575  
 His Pro Ala Ser Glu Gly Gly Gly Asp Thr Val Phe Val Pro Ile Gly  
 580 585 590  
 Arg Pro Ile Ala Asn Thr Arg Val Tyr Val Leu Asp Glu Gln Leu Arg  
 595 600 605  
 Arg Val Pro Val Gly Val Met Gly Glu Leu Tyr Val Asp Ser Val Gly  
 610 615 620  
 Met Ala Arg Gly Tyr Trp Gly Gln Pro Ala Leu Thr Ala Glu Arg Phe  
 625 630 635 640  
 Ile Ala Asn Pro Tyr Ala Ser Gln Pro Gly Ala Arg Leu Tyr Arg Thr  
 645 650 655  
 Gly Asp Met Val Arg Val Leu Ala Asp Gly Ser Leu Glu Tyr Leu Gly  
 660 665 670  
 Arg Arg Asp Tyr Glu Ile Lys Val Arg Gly His Arg Val Asp Val Arg  
 675 680 685  
 Gln Val Glu Lys Val Ala Asn Ala His Pro Ala Ile Arg Gln Ala Val  
 690 695 700  
 Val Ser Gly Trp Pro Leu Gly Ser Ser Asn Ala Gln Leu Val Ala Tyr  
 705 710 715 720  
 Leu Val Pro Gln Ala Gly Ala Thr Val Gly Pro Arg Gln Val Arg Asp  
 725 730 735  
 Tyr Leu Ala Glu Ser Leu Pro Ala Tyr Met Val Pro Thr Leu Tyr Thr  
 740 745 750  
 Val Leu Glu Glu Leu Pro Arg Leu Pro Asn Gly Lys Leu Asp Arg Leu  
 755 760 765  
 Ser Leu Pro Glu Pro Asp Leu Ser Ser Ser Arg Glu Glu Tyr Val Ala  
 770 775 780  
 Pro His Gly Glu Val Glu Arg Lys Leu Ala Glu Ile Phe Gly Asn Leu  
 785 790 795 800  
 Leu Gly Leu Glu His Val Gly Val His Asp Asn Phe Phe Ser Leu Gly  
 805 810 815  
 Gly His Ser Leu Leu Ala Ala Gln Leu Ile Ser Arg Ile Arg Ala Thr  
 820 825 830  
 Phe Arg Val Glu Val Ala Met Ala Thr Val Phe Glu Ser Pro Thr Val  
 835 840 845  
 Glu Pro Leu Ala Arg His Ile Glu Glu Lys Leu Lys Asp Glu Ser Arg  
 850 855 860  
 Val Gln Leu Ser Asn Val Val Pro Val Glu Arg Thr Gln Glu Ile Pro  
 865 870 875 880  
 Leu Ser Tyr Leu Gln Glu Arg Leu Trp Phe Val His Glu His Met Lys  
 885 890 895  
 Glu Gln Arg Thr Ser Tyr Asn Ile Thr Trp Thr Leu His Phe Ala Gly  
 900 905 910  
 Lys Gly Phe Ser Val Glu Ala Leu Arg Thr Ala Phe Asp Glu Leu Val  
 915 920 925  
 Ala Arg His Glu Thr Leu Arg Thr Trp Phe Gln Val Gly Glu Gly Thr  
 930 935 940  
 Glu Gln Ala Val Gln Val Ile Gly Glu Pro Trp Ser Met Glu Leu Pro  
 945 950 955 960  
 Leu Arg Glu Val Ala Gly Thr Glu Val Thr Ala Ala Ile Asn Glu Met  
 965 970 975  
 Ser Arg Gln Val Phe Asp Leu Arg Ala Gly Arg Leu Leu Thr Ala Ala

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

[0050]

Leu	Glu	Ala	Gly	Lys	Glu	Arg	Ser	Pro	Leu	Glu	Arg	Gly	Glu	Val
1520						1525					1530			
Met	Leu	Gln	Lys	Thr	Ala	Ile	Ala	Phe	Ala	Val	Ser	Val	Lys	Glu
1535						1540					1545			
Leu	Leu	Ser	Gly	Leu	Leu	Ala	Gly	Val	Ala	Gln	Val	Met	Val	Pro
1550						1555					1560			
Glu	Thr	Leu	Val	Lys	Asp	Ser	Val	Ala	Leu	Ala	Gln	Glu	Ile	Glu
1565						1570					1575			
Arg	Trp	Arg	Val	Thr	Arg	Ile	His	Leu	Val	Pro	Ser	His	Leu	Gly
1580						1585					1590			
Ala	Leu	Leu	Glu	Gly	Ala	Gly	Glu	Glu	Ala	Lys	Gly	Leu	Arg	Ser
1595						1600					1605			
Leu	Lys	Tyr	Val	Ile	Thr	Ala	Gly	Glu	Ala	Leu	Ala	Gln	Gly	Val
1610						1615					1620			
Arg	Glu	Glu	Ala	Arg	Arg	Lys	Leu	Pro	Gly	Ala	Gln	Leu	Trp	Asn
1625						1630					1635			
Asn	Tyr	Gly	Cys	Thr	Glu	Leu	Asn	Asp	Val	Thr	Tyr	His	Pro	Ala
1640						1645					1650			
Ser	Glu	Gly	Gly	Gly	Asp	Thr	Val	Phe	Val	Pro	Ile	Gly	Arg	Pro
1655						1660					1665			
Ile	Ala	Asn	Thr	Arg	Val	Tyr	Val	Leu	Asp	Glu	Gln	Leu	Arg	Arg
1670						1675					1680			
Val	Pro	Val	Gly	Val	Met	Gly	Glu	Leu	Tyr	Val	Asp	Ser	Val	Gly
1685						1690					1695			
Met	Ala	Arg	Gly	Tyr	Trp	Gly	Gln	Pro	Ala	Leu	Thr	Ala	Glu	Arg
1700						1705					1710			
Phe	Ile	Ala	Asn	Pro	Tyr	Ala	Ser	Gln	Pro	Gly	Ala	Arg	Leu	Tyr
1715						1720					1725			
Arg	Thr	Gly	Asp	Met	Val	Arg	Val	Leu	Ala	Asp	Gly	Ser	Leu	Glu
1730						1735					1740			
Tyr	Leu	Gly	Arg	Arg	Asp	Tyr	Glu	Ile	Lys	Val	Arg	Gly	His	Arg
1745						1750					1755			
Val	Asp	Val	Arg	Gln	Val	Glu	Lys	Val	Ala	Asn	Ala	His	Pro	Ala
1760						1765					1770			
Ile	Arg	Gln	Ala	Val	Val	Ser	Gly	Trp	Pro	Leu	Gly	Ser	Ser	Asn
1775						1780					1785			
Ala	Gln	Leu	Val	Ala	Tyr	Leu	Val	Pro	Gln	Ala	Gly	Ala	Thr	Val
1790						1795					1800			
Gly	Pro	Arg	Gln	Val	Arg	Asp	Tyr	Leu	Ala	Glu	Ser	Leu	Pro	Ala
1805						1810					1815			
Tyr	Met	Val	Pro	Thr	Leu	Tyr	Thr	Val	Leu	Glu	Glu	Leu	Pro	Arg
1820						1825					1830			
Leu	Pro	Asn	Gly	Lys	Leu	Asp	Arg	Leu	Ser	Leu	Pro	Glu	Pro	Asp
1835						1840					1845			
Leu	Ser	Ser	Ser	Arg	Glu	Glu	Tyr	Val	Ala	Pro	His	Gly	Glu	Val
1850						1855					1860			
Glu	Arg	Lys	Leu	Ala	Glu	Ile	Phe	Gly	Asn	Leu	Leu	Gly	Leu	Glu
1865						1870					1875			
His	Val	Gly	Val	His	Asp	Asn	Phe	Phe	Ser	Leu	Gly	Gly	His	Ser
1880						1885					1890			
Leu	Leu	Ala	Ala	Gln	Val	Val	Ser	Arg	Ile	Gly	Lys	Glu	Leu	Gly
1895						1900					1905			
Thr	Gln	Ile	Ser	Ile	Ala	Asp	Leu	Phe	Gln	Arg	Pro	Thr	Ile	Glu
1910						1915					1920			
Gln	Leu	Cys	Glu	Leu	Ile	Gly	Gly	Leu	Asp	Asp	Gln	Thr	Gln	Arg
1925						1930					1935			
Glu	Leu	Ala	Leu	Ala	Pro	Ser	Gly	Asn	Thr	Glu	Ala	Val	Leu	Ser
1940						1945					1950			
Phe	Ala	Gln	Glu	Arg	Met	Trp	Phe	Leu	His	Asn	Phe	Val	Lys	Gly
1955						1960					1965			
Met	Pro	Tyr	Asn	Thr	Pro	Gly	Leu	Asp	His	Leu	Thr	Gly	Glu	Leu
1970						1975					1980			
Asp	Val	Ala	Ala	Leu	Glu	Lys	Ala	Ile	Arg	Ala	Val	Ile	Arg	Arg
1985						1990					1995			
His	Glu	Pro	Leu	Arg	Thr	Asn	Phe	Val	Glu	Lys	Asp	Gly	Val	Leu
2000						2005					2010			
Ser	Gln	Leu	Val	Gly	Thr	Glu	Glu	Arg	Phe	Arg	Leu	Thr	Val	Thr
2015						2020					2025			
Pro	Ile	Arg	Asp	Glu	Ser	Glu	Val	Ala	Arg	Leu	Met	Glu	Ala	Val
2030						2035					2040			
Ile	Gln	Thr	Pro	Val	Asp	Leu	Glu	Arg	Glu	Leu	Met	Ile	Arg	Ala

[0051]

2045	Tyr Leu Tyr Arg Val Asp	2050	Pro Arg Asn His Tyr	2055	Leu Phe Thr Thr
2060	Ile His His Ile Ala Phe	2065	Asp Gly Trp Ser Thr	2070	Ser Ile Phe Tyr
2075	Arg Glu Leu Ala Ala Tyr	2080	Tyr Ala Ala Phe Leu	2085	Arg Arg Glu Asp
2090	Ser Pro Leu Pro Ala Leu	2095	Glu Ile Ser Tyr Gln	2100	Asp Tyr Ala Arg
2105	Trp Glu Arg Ala His Phe	2110	Gln Asp Glu Val Leu	2115	Ala Glu Lys Leu
2120	Arg Tyr Trp Arg Gln Arg	2125	Leu Ser Gly Ala Arg	2130	Pro Leu Val Leu
2135	Pro Thr Thr Tyr His Arg	2140	Pro Pro Ile Gln Ser	2145	Phe Ala Gly Ala
2150	Val Val Asn Phe Glu Ile	2155	Asp Arg Ser Ile Thr	2160	Glu Arg Leu Lys
2165	Thr Leu Phe Ala Glu Ser	2170	Gly Thr Thr Met Tyr	2175	Met Val Leu Leu
2180	Gly Ala Phe Ser Val Val	2185	Leu Gln Arg Tyr Ser	2190	Gly Gln Asp Asp
2195	Ile Cys Ile Gly Ser Pro	2200	Val Ala Asn Arg Gly	2205	His Ile Gln Thr
2210	Glu Gly Leu Ile Gly Leu	2215	Phe Val Asn Thr Leu	2220	Val Met Arg Val
2225	Asp Ala Ala Gly Asn Pro	2230	Arg Phe Ile Asp Leu	2235	Leu Ala Arg Ile
2240	Gln Arg Thr Ala Ile Asp	2245	Ala Tyr Ala Asn Gln	2250	Glu Val Pro Phe
2255	Glu Lys Ile Val Asp Asp	2260	Leu Gln Val Ala Arg	2265	Asp Thr Ala Arg
2270	Ser Pro Leu Val Gln Val	2275	Ile Leu Asn Phe His	2280	Asn Thr Pro Pro
2285	Gln Ser Glu Leu Glu Leu	2290	Gln Gly Val Thr Leu	2295	Thr Arg Met Pro
2300	Val His Asn Gly Thr Ala	2305	Lys Phe Glu Leu Ser	2310	Ile Asp Val Ala
2315	Glu Thr Ser Ala Gly Leu	2320	Thr Gly Phe Val Glu	2325	Tyr Ala Thr Asp
2330	Leu Phe Ser Glu Asn Phe	2335	Ile Arg Arg Met Ile	2340	Gly His Leu Glu
2345	Val Val Leu Asp Ala Val	2350	Gly Arg Asp Pro Arg	2355	Ala Pro Ile His
2360	Glu Leu Pro Leu Leu Thr	2365	Arg Gln Asp Gln Leu	2370	Asp Leu Leu Ser
2375	Arg Ser Gly His Thr Ala	2380	Pro Ala Val Glu His	2385	Val Glu Leu Ile
2390	Pro His Thr Phe Glu Arg	2395	Arg Val Gln Glu Ser	2400	Pro Gln Ala Ile
2405	Ala Leu Val Cys Gly Asp	2410	Glu Arg Val Thr Tyr	2415	Ser Ala Leu Asn
2420	Arg Arg Ala Ser Gln Ile	2425	Ala Arg Arg Leu Arg	2430	Ala Ala Gly Ile
2435	Gly Pro Asp Thr Leu Val	2440	Gly Leu Cys Ala Gly	2445	Arg Ser Ile Glu
2450	Leu Val Cys Gly Val Leu	2455	Gly Ile Leu Lys Ala	2460	Gly Gly Ala Tyr
2465	Val Pro Ile Asp Pro Thr	2470	Ser Ser Pro Glu Val	2475	Ile Tyr Asp Val
2480	Leu Tyr Glu Ser Lys Val	2485	Arg His Leu Leu Thr	2490	Glu Ser Arg Leu
2495	Val Gly Gly Leu Pro Val	2500	Asp Asp Gln Glu Ile	2505	Leu Leu Leu Asp
2510	Thr Pro Ala Asp Gly Glu	2515	Gly Asp Lys Ala Val	2520	Ala Asp Arg Glu
2525	Glu Pro Pro Asp Leu Gly	2530	Glu Val Ser Leu Thr	2535	Pro Glu Cys Leu
2540	Ala Tyr Val Asn Phe Thr	2545	Ser Asp Ser Gly Gly	2550	Ala Pro Arg Gly
2555	Ile Ala Val Arg His Gly	2560	Ala Leu Ala Arg Arg	2565	Met Ala Ala Gly
2570		2575		2580	

[0052]

His	Ala	Gln	Tyr	Leu	Ala	Asn	Ser	Ala	Val	Arg	Phe	Leu	Leu	Lys
2585						2590					2595			
Ala	Pro	Leu	Thr	Phe	Asp	Leu	Ala	Val	Ala	Glu	Leu	Phe	Gln	Trp
2600						2605					2610			
Ile	Val	Ser	Gly	Gly	Ser	Leu	Ser	Ile	Leu	Asp	Pro	Asn	Ala	Asp
2615						2620					2625			
Arg	Asp	Ala	Ser	Ala	Phe	Leu	Ala	Gln	Val	Arg	Arg	Asp	Ser	Ile
2630						2635					2640			
Gly	Val	Leu	Tyr	Cys	Val	Pro	Ser	Glu	Leu	Ser	Thr	Leu	Val	Ser
2645						2650					2655			
His	Leu	Glu	Arg	Glu	Arg	Glu	Arg	Val	His	Glu	Leu	Asn	Thr	Leu
2660						2665					2670			
Arg	Phe	Ile	Phe	Cys	Gly	Gly	Asp	Thr	Leu	Ala	Val	Thr	Val	Val
2675						2680					2685			
Glu	Arg	Leu	Gly	Val	Leu	Val	Arg	Ala	Gly	Gln	Leu	Pro	Leu	Arg
2690						2695					2700			
Leu	Val	Asn	Val	Tyr	Gly	Thr	Lys	Glu	Thr	Gly	Ile	Gly	Ala	Gly
2705						2710					2715			
Cys	Phe	Glu	Cys	Ala	Leu	Asp	Ala	Asn	Asp	Pro	Ser	Ala	Glu	Leu
2720						2725					2730			
Pro	Pro	Gly	Arg	Leu	Ser	His	Glu	Arg	Met	Pro	Ile	Gly	Gly	Pro
2735						2740					2745			
Ala	Gln	Asn	Leu	Trp	Phe	Tyr	Val	Val	Gln	Pro	Asn	Gly	Gly	Leu
2750						2755					2760			
Ala	Pro	Leu	Gly	Ile	Pro	Gly	Glu	Leu	Tyr	Val	Gly	Gly	Ala	Gln
2765						2770					2775			
Leu	Ala	Asp	Ala	Arg	Phe	Gly	Asp	Glu	Pro	Thr	Ala	Thr	His	Pro
2780						2785					2790			
Gly	Phe	Val	Pro	Asn	Pro	Phe	Arg	Ser	Gly	Ala	Glu	Lys	Asp	Trp
2795						2800					2805			
Leu	Tyr	Lys	Thr	Gly	Asp	Leu	Val	Arg	Trp	Leu	Pro	Gln	Gly	Pro
2810						2815					2820			
Leu	Glu	Leu	Val	Ser	Ala	Ala	Arg	Glu	Arg	Asp	Gly	Gly	Gly	Asp
2825						2830					2835			
His	Arg	Leu	Asp	Arg	Gly	Phe	Ile	Glu	Ala	Arg	Met	Arg	Arg	Val
2840						2845					2850			
Ala	Ile	Val	Arg	Asp	Ala	Val	Val	Ala	Tyr	Val	Pro	Asp	Arg	Gln
2855						2860					2865			
Asp	Arg	Ala	Arg	Leu	Val	Ala	Tyr	Val	Val	Leu	Lys	Glu	Ser	Pro
2870						2875					2880			
Ala	Ala	Asp	Val	Glu	Pro	Arg	Glu	Gly	Arg	Glu	Thr	Leu	Lys	Ala
2885						2890					2895			
Arg	Ile	Ser	Ala	Glu	Leu	Gly	Ser	Thr	Leu	Pro	Glu	Tyr	Met	Leu
2900						2905					2910			
Pro	Ala	Ala	Tyr	Val	Phe	Met	Asp	Ser	Leu	Pro	Leu	Thr	Ala	Tyr
2915						2920					2925			
Gly	Arg	Ile	Asp	Arg	Lys	Ala	Leu	Pro	Glu	Pro	Glu	Asp	Asp	Arg
2930						2935					2940			
His	Gly	Gly	Ser	Ala	Ile	Ala	Tyr	Val	Ala	Pro	Arg	Gly	Pro	Thr
2945						2950					2955			
Glu	Lys	Ala	Leu	Ala	His	Ile	Trp	Gln	Gln	Val	Leu	Lys	Arg	Pro
2960						2965					2970			
Gln	Val	Gly	Leu	Arg	Asp	Asn	Phe	Phe	Glu	Leu	Gly	Gly	His	Ser
2975						2980					2985			
Val	Ala	Ala	Ile	Gln	Leu	Val	Ser	Val	Ser	Arg	Lys	His	Leu	Glu
2990						2995					3000			
Val	Glu	Val	Pro	Leu	Ser	Leu	Ile	Phe	Glu	Ser	Pro	Val	Leu	Glu
3005						3010					3015			
Ala	Met	Ala	Arg	Gly	Ile	Glu	Ala	Leu	Gln	Gln	Gln	Gly	Arg	Ser
3020						3025					3030			
Gly	Ala	Val	Ser	Ser	Ile	His	Arg	Val	Glu	Arg	Thr	Gly	Pro	Leu
3035						3040					3045			
Pro	Leu	Ala	Tyr	Val	Gln	Glu	Arg	Leu	Trp	Phe	Val	His	Glu	His
3050						3055					3060			
Met	Lys	Glu	Gln	Arg	Thr	Ser	Tyr	Asn	Ile	Thr	Trp	Thr	Leu	His
3065						3070					3075			
Phe	Ala	Gly	Lys	Gly	Phe	Ser	Val	Glu	Ala	Leu	Arg	Thr	Ala	Phe
3080						3085					3090			
Asp	Glu	Leu	Val	Ala	Arg	His	Glu	Thr	Leu	Arg	Thr	Trp	Phe	Gln
3095						3100					3105			
Val	Gly	Glu	Gly	Thr	Glu	Gln	Ala	Val	Gln	Val	Ile	Gly	Glu	Pro

[0053]

3110	Trp Ser Met Glu Leu Pro	3115	Leu Arg Glu Val Ala	3120	Gly Thr Glu Val
3125	Thr Ala Ala Ile Asn Glu	3130	Met Ser Arg Gln Val	3135	Phe Asp Leu Arg
3140	Ala Gly Arg Leu Leu Thr	3145	Ala Ala Val Leu Arg	3150	Val Ala Glu Asp
3155	Glu His Ile Leu Val Ser	3160	Asn Ile His His Ile	3165	Ile Thr Asp Gly
3170	Trp Ser Phe Gly Val Met	3175	Leu Arg Glu Leu Arg	3180	Glu Leu Tyr Glu
3185	Ala Ala Val Arg Gly Glu	3190	Arg Ala Glu Leu Pro	3195	Pro Leu Thr Val
3200	Gln Tyr Gly Asp Tyr Ala	3205	Val Trp Gln Arg Lys	3210	Gln Asp Leu Ser
3215	Glu His Leu Ala Tyr Trp	3220	Lys Gly Lys Val Glu	3225	Gly Asp Glu Asp
3230	Gly Leu Glu Leu Pro Tyr	3235	Asp Phe Pro Arg Thr	3240	Ser Asn Arg Ala
3245	Trp Arg Ala Ala Thr Phe	3250	Gln Tyr Ser Tyr His	3255	Pro Glu Leu Ala
3260	Arg Lys Val Ala Glu Leu	3265	Ser Arg Glu Gln Gln	3270	Ser Thr Leu Phe
3275	Met Ser Leu Val Ala Ser	3280	Leu Ala Val Val Leu	3285	Asn Arg Tyr Thr
3290	Gly Arg Glu Asp Leu Cys	3295	Ile Gly Thr Thr Val	3300	Ala Gly Arg Ala
3305	Gln Val Glu Leu Glu Ser	3310	Leu Ile Gly Phe Phe	3315	Ile Asn Ile Leu
3320	Pro Leu Arg Leu Asp Leu	3325	Ser Gly Ala Pro Ser	3330	Leu His Glu Val
3335	Leu Arg Arg Thr Lys Val	3340	Val Val Leu Glu Gly	3345	Phe Glu His Gln
3350	Glu Leu Pro Phe Glu His	3355	Leu Leu Lys Ala Leu	3360	Arg Arg Gln Arg
3365	Asp Ser Ser Gln Ile Pro	3370	Leu Val Pro Val Val	3375	Val Arg His Gln
3380	Asn Phe Pro Met Ala Arg	3385	Leu Glu Gly Trp Ser	3390	Glu Gly Val Glu
3395	Leu Lys Lys Phe Glu Leu	3400	Ala Gly Glu Arg Thr	3405	Thr Ala Ser Glu
3410	Gln Asp Trp Gln Phe Phe	3415	Gly Asp Gly Ser Ser	3420	Leu Glu Leu Ser
3425	Leu Glu Tyr Ala Ala Glu	3430	Leu Phe Ser Glu Lys	3435	Thr Val Arg Arg
3440	Met Val Glu His His Gln	3445	Arg Val Leu Glu Ala	3450	Leu Val Glu Gly
3455	Leu Glu Glu Gly Leu His	3460	Glu Val Arg Leu Leu	3465	Thr Glu Glu Glu
3470	Glu Gly Leu His Gly Arg	3475	Leu Asn Asp Thr Ala	3480	Arg Glu Leu Glu
3485	Glu Arg Trp Ser Leu Ala	3490	Glu Thr Phe Glu Arg	3495	Gln Val Arg Glu
3500	Thr Pro Glu Ala Val Ala	3505	Cys Val Gly Val Glu	3510	Val Ala Thr Gly
3515	Gly His Ser Arg Pro Thr	3520	Tyr Arg Gln Leu Thr	3525	Tyr Arg Gln Leu
3530	Asn Ala Arg Ala Asn Gln	3535	Val Ala Arg Arg Leu	3540	Arg Ala Leu Gly
3545	Val Gly Ala Glu Thr Arg	3550	Val Ala Val Leu Ser	3555	Asp Arg Ser Pro
3560	Glu Leu Leu Val Ala Met	3565	Leu Ala Ile Phe Lys	3570	Ala Gly Gly Cys
3575	Tyr Val Pro Val Asp Pro	3580	Gln Tyr Pro Gly Ser	3585	Tyr Ile Glu Gln
3590	Ile Leu Glu Asp Ala Ala	3595	Pro Gln Val Val Leu	3600	Gly Lys Arg Gly
3605	Arg Ala Asp Gly Val Arg	3610	Val Asp Val Trp Leu	3615	Glu Leu Asp Gly
3620	Ala Gln Arg Leu Thr Asp	3625	Glu Ala Leu Ala Ala	3630	Gln Glu Glu Gly
3635		3640		3645	



[0054]

Glu Leu	Glu Gly Ala Glu Arg	Pro Glu Ser Gln Gln	Leu Ala Cys
3650	3655	3660	
Leu Met	Tyr Thr Ser Gly Ser	Thr Gly Arg Pro Lys	Gly Val Met
3665	3670	3675	
Val Pro	Tyr Ser Gln Leu His	Asn Trp Leu Glu Ala	Gly Lys Glu
3680	3685	3690	
Arg Ser	Pro Leu Glu Arg Gly	Glu Val Met Leu Gln	Lys Thr Ala
3695	3700	3705	
Ile Ala	Phe Ala Val Ser Val	Lys Glu Leu Leu Ser	Gly Leu Leu
3710	3715	3720	
Ala Gly	Val Ala Gln Val Met	Val Pro Glu Thr Leu	Val Lys Asp
3725	3730	3735	
Ser Val	Ala Leu Ala Gln Glu	Ile Glu Arg Trp Arg	Val Thr Arg
3740	3745	3750	
Ile His	Leu Val Pro Ser His	Leu Gly Ala Leu Leu	Glu Gly Ala
3755	3760	3765	
Gly Glu	Glu Ala Lys Gly Leu	Arg Ser Leu Lys Tyr	Val Ile Thr
3770	3775	3780	
Ala Gly	Glu Ala Leu Ala Gln	Gly Val Arg Glu Glu	Ala Arg Arg
3785	3790	3795	
Lys Leu	Pro Gly Ala Gln Leu	Trp Asn Asn Tyr Gly	Cys Thr Glu
3800	3805	3810	
Leu Asn	Asp Val Thr Tyr His	Pro Ala Ser Glu Gly	Gly Gly Asp
3815	3820	3825	
Thr Val	Phe Val Pro Ile Gly	Arg Pro Ile Ala Asn	Thr Arg Val
3830	3835	3840	
Tyr Val	Leu Asp Glu Gln Leu	Arg Arg Val Pro Val	Gly Val Met
3845	3850	3855	
Gly Glu	Leu Tyr Val Asp Ser	Val Gly Met Ala Arg	Gly Tyr Trp
3860	3865	3870	
Gly Gln	Pro Ala Leu Thr Ala	Glu Arg Phe Ile Ala	Asn Pro Tyr
3875	3880	3885	
Ala Ser	Gln Pro Gly Ala Arg	Leu Tyr Arg Thr Gly	Asp Met Val
3890	3895	3900	
Arg Val	Leu Ala Asp Gly Ser	Leu Glu Tyr Leu Gly	Arg Arg Asp
3905	3910	3915	
Tyr Glu	Ile Lys Val Arg Gly	His Arg Val Asp Val	Arg Gln Val
3920	3925	3930	
Glu Lys	Val Ala Asn Ala His	Pro Ala Ile Arg Gln	Ala Val Val
3935	3940	3945	
Ser Gly	Trp Pro Leu Gly Ser	Ser Asn Ala Gln Leu	Val Ala Tyr
3950	3955	3960	
Leu Val	Pro Gln Ala Gly Ala	Thr Val Gly Pro Arg	Gln Val Arg
3965	3970	3975	
Asp Tyr	Leu Ala Glu Ser Leu	Pro Ala Tyr Met Val	Pro Thr Leu
3980	3985	3990	
Tyr Thr	Val Leu Glu Glu Leu	Pro Arg Leu Pro Asn	Gly Lys Leu
3995	4000	4005	
Asp Arg	Leu Ser Leu Pro Glu	Pro Asp Leu Ser Ser	Ser Arg Glu
4010	4015	4020	
Glu Tyr	Val Ala Pro His Gly	Glu Val Glu Arg Lys	Leu Ala Glu
4025	4030	4035	
Ile Phe	Gly Asn Leu Leu Gly	Leu Glu His Val Gly	Val His Asp
4040	4045	4050	
Asn Phe	Phe Asn Leu Gly Gly	His Ser Leu Leu Ala	Ser Gln Leu
4055	4060	4065	
Ile Ser	Arg Ile Arg Ala Thr	Phe Arg Val Glu Val	Ala Met Ala
4070	4075	4080	
Thr Val	Phe Glu Ser Pro Thr	Val Glu Pro Leu Ala	Arg His Ile
4085	4090	4095	
Glu Glu	Lys Leu Lys Asp Glu	Ser Arg Val Gln Leu	Ser Asn Val
4100	4105	4110	
Val Pro	Val Glu Arg Thr Gln	Glu Leu Pro Leu Ser	Tyr Leu Gln
4115	4120	4125	
Glu Arg	Leu Trp Phe Val His	Glu His Met Lys Glu	Gln Arg Thr
4130	4135	4140	
Ser Tyr	Asn Gly Thr Ile Gly	Leu Arg Leu Arg Gly	Pro Leu Ser
4145	4150	4155	
Ile Pro	Ala Leu Arg Ala Thr	Phe His Asp Leu Val	Ala Arg His
4160	4165	4170	
Glu Ser	Leu Arg Thr Val Phe	Arg Val Pro Glu Gly	Arg Thr Thr

4175 4180 4185  
 Pro Val Gln Val Ile Leu Asp Ser Met Asp Leu Asp Ile Pro Val  
 4190 4195 4200  
 Arg Asp Ala Thr Glu Ala Asp Ile Ile Pro Gly Met Asp Glu Leu  
 4205 4210 4215  
 Ala Gly His Ile Tyr Asp Met Glu Lys Gly Pro Leu Phe Met Val  
 4220 4225 4230  
 Arg Leu Leu Arg Leu Ala Glu Asp Ser His Val Leu Leu Met Gly  
 4235 4240 4245  
 Met His His Ile Val Tyr Asp Ala Trp Ser Gln Phe Asn Val Met  
 4250 4255 4260  
 Ser Arg Asp Ile Asn Leu Leu Tyr Ser Ala His Val Thr Gly Ile  
 4265 4270 4275  
 Glu Ala Arg Leu Pro Ala Leu Pro Ile Gln Tyr Ala Asp Phe Ser  
 4280 4285 4290  
 Val Trp Gln Arg Gln Gln Asp Phe Arg His His Leu Asp Tyr Trp  
 4295 4300 4305  
 Lys Ser Thr Leu Gly Asp Tyr Arg Asp Asp Leu Glu Leu Pro Tyr  
 4310 4315 4320  
 Asp Tyr Pro Arg Pro Pro Ser Arg Thr Trp His Ala Thr Arg Phe  
 4325 4330 4335  
 Thr Phe Arg Tyr Pro Asp Ala Leu Ala Arg Ala Phe Ala Arg Phe  
 4340 4345 4350  
 Asn Gln Ser His Gln Ser Thr Leu Phe Met Gly Leu Leu Thr Ser  
 4355 4360 4365  
 Phe Ala Ile Val Leu Arg His Tyr Thr Gly Arg Asn Asp Ile Cys  
 4370 4375 4380  
 Ile Gly Thr Thr Thr Ala Gly Arg Ala Gln Leu Glu Leu Glu Asn  
 4385 4390 4395  
 Leu Val Gly Phe Phe Ile Asn Ile Leu Pro Leu Arg Ile Asn Leu  
 4400 4405 4410  
 Ala Gly Asp Pro Asp Ile Ser Glu Leu Met Asn Arg Ala Lys Lys  
 4415 4420 4425  
 Ser Val Leu Gly Ala Phe Glu His Gln Ala Leu Pro Phe Glu Arg  
 4430 4435 4440  
 [0055] Leu Leu Ser Ala Leu Asn Lys Gln Arg Asp Ser Ser His Ile Pro  
 4445 4450 4455  
 Leu Val Pro Val Met Leu Arg His Gln Asn Phe Pro Thr Ala Met  
 4460 4465 4470  
 Thr Gly Lys Trp Ala Asp Gly Val Asp Met Glu Val Ile Glu Arg  
 4475 4480 4485  
 Asp Glu Arg Thr Thr Pro Asn Glu Leu Asp Leu Gln Phe Phe Gly  
 4490 4495 4500  
 Asp Asp Thr Tyr Leu His Ala Val Val Glu Phe Pro Ala Gln Leu  
 4505 4510 4515  
 Phe Ser Glu Val Thr Val Arg Arg Leu Met Gln Arg His Gln Lys  
 4520 4525 4530  
 Val Ile Glu Phe Met Cys Ala Thr Leu Gly Ala Arg  
 4535 4540 4545  
  
 <210> 51  
 <211> 1023  
 <212> PRT  
 <213> 深棕色孢囊杆菌  
  
 <220>  
 <221> MISC\_FEATURE  
 <222> (1).. (1023)  
 <223> CysL  
  
 <400> 51  
 Val Asn Val Leu Ala Arg His Ser Thr Gly Ser His Asp Glu Pro Val  
 1 5 10 15  
 Ala Gly Asp Val Glu Leu Arg Val Gly Gly Pro Gly Val Pro Asp Ala  
 20 25 30  
 His Ser Ser Glu Ser Val Glu Val Leu Ala Arg Trp Leu Arg Thr Ala  
 35 40 45  
 Glu Glu Lys Tyr Pro Gly Val Met Gly Pro Ile Arg Gln Glu Gly Pro  
 50 55 60  
 Trp Phe Ala Ile Pro Leu Thr Cys Pro Arg Gly Ala Arg Ser Ala Arg  
 65 70 75 80

[0056]

Phe Gly Leu Trp Leu Gly Glu Leu Asp Arg Gln Gly Gln Leu Leu His  
 85 90 95  
 Met Val Ala Ser Tyr Leu Ala Ala Val His His Val Leu Val Ser Val  
 100 105 110  
 Arg Glu Pro Ser Ala Asn Val Leu Glu Val Leu Val Ser Asp Ser Thr  
 115 120 125  
 Thr Pro Ser Gly Leu Asn Arg Phe Leu Asn Gly Leu Asp Ser Val Leu  
 130 135 140  
 Glu Ile Leu Ala His Gly Arg Ser Asp Leu Leu Gln His Leu Thr  
 145 150 155 160  
 Gly Arg Leu Pro Pro Asp Glu Met Pro Phe Val Glu Asp Arg Glu Glu  
 165 170 175  
 Arg Glu Glu His Pro Ala Thr Asp Val Glu Ala Asp Ala Val Val Ser  
 180 185 190  
 Val Leu Phe Gln Pro Val Asp Phe Pro Ser Leu Ala Arg Leu Asp Ala  
 195 200 205  
 Ser Leu Leu Ala Tyr Asp Asp Glu Asp Ala Gly Ala Val Gly Arg Val  
 210 215 220  
 Leu Gly Glu Leu Leu Gln Pro Phe Leu Leu Asp Ser Ala Arg Met Thr  
 225 230 235 240  
 Val Gly Arg Lys Ala Val Arg Val Asp His Ile Cys Leu Pro Gly Leu  
 245 250 255  
 Leu Arg Ala Asp Ser Arg Ala Ala Glu Glu Ser Val Leu Ala Pro Ala  
 260 265 270  
 Leu Arg Leu Ala Thr Lys Pro Gly Arg His Phe Val Ala Leu Cys Arg  
 275 280 285  
 Asn Thr Ala Leu Arg Leu Gly Asp Arg Leu Pro His Leu Leu Ala Gln  
 290 295 300  
 Gly Pro Leu Cys Asp Gly Ala Ser Thr Ala Leu Leu Leu Gln Arg  
 305 310 315 320  
 Val Leu Asp Thr Leu Ile Gly Ser Gly Gly Leu Lys Asp His Arg Leu  
 325 330 335  
 Thr Leu Glu Leu Val Gly Ala Asp Pro Arg Thr Glu Ala Ala Phe Arg  
 340 345 350  
 Ala Arg Thr Pro Trp Leu Val Ala Glu Arg Ala Ala Ser Ala Ala Ser  
 355 360 365  
 Thr Asp Ala Pro Arg Val Asp Val Val Val Leu Phe Pro Ala Ala Arg  
 370 375 380  
 Pro Ser Ala Leu Glu Leu Arg Pro Asp Ser Val Val Ile Asp Leu Phe  
 385 390 395 400  
 Gly Thr Trp Ser Leu Arg Pro Arg Pro Glu Val Leu Ala Lys Asn Ile  
 405 410 415  
 Val Tyr Val Arg Gly Ala Ser Val Arg Leu Ala Gly Glu Ala Val Val  
 420 425 430  
 Ser Thr Pro Ser Phe Ala Pro Asp Arg Val Glu Pro Ala Leu Leu Glu  
 435 440 445  
 Ala Leu Leu Arg Glu Leu Asp Ala Glu Ala Ser Ser Asp Gly Leu Ala  
 450 455 460  
 His Glu His Arg Leu Glu Ile Gly Gly Ile Arg Gly Phe Trp Gly Glu  
 465 470 475 480  
 Ile Arg Arg Ala Glu Trp Asp Ala Phe His Ser Arg Arg Arg Gly Glu  
 485 490 495  
 Leu Ala Arg Phe Gln Val Ser Gly Gln Val Thr Ala Ala Asn Pro Gly  
 500 505 510  
 Leu Ala Ser Leu Pro Asp Gly Ala Thr Asn Ile Cys Glu Tyr Ile Phe  
 515 520 525  
 Arg Glu Ala His Leu Arg Ser Gly Ser Cys Leu Val Asp Pro Gln Ser  
 530 535 540  
 Gly Gln Ser Ala Thr Tyr Ala Glu Leu Arg Arg Leu Ala Ala Ala Tyr  
 545 550 555 560  
 Ala Arg Arg Phe Arg Ala Leu Gly Leu Arg Gln Gly Asp Val Val Ala  
 565 570 575  
 Leu Ala Ala Pro Asp Gly Ile Ser Ser Val Ala Val Met Leu Gly Cys  
 580 585 590  
 Phe Leu Gly Gly Trp Val Phe Ala Pro Leu Asn His Thr Ala Ser Ala  
 595 600 605  
 Val Asn Phe Glu Ala Met Leu Ser Ser Ala Ser Pro Arg Leu Val Leu  
 610 615 620  
 His Ala Ala Ser Thr Val Ala Arg His Leu Pro Val Leu Ser Thr Arg  
 625 630 635 640  
 Arg Cys Ala Glu Leu Ala Ser Phe Leu Pro Pro Asp Ala Leu Asp Gly

[0057]

```

        645              650              655
Val Glu Gly Asp Val Thr Pro Leu Pro Val Ser Pro Glu Ala Pro Ala
        660              665              670
Val Met Leu Phe Thr Ser Gly Ser Thr Gly Gly Pro Lys Ala Val Thr
        675              680              685
His Thr His Ala Asp Phe Ile Thr Cys Ser Arg Asn Tyr Ala Pro Tyr
        690              695              700
Val Val Glu Leu Arg Pro Asp Asp Arg Val Tyr Thr Pro Ser Pro Thr
705              710              715              720
Phe Phe Ala Tyr Gly Leu Asn Asn Leu Leu Leu Ser Leu Ser Ala Gly
        725              730              735
Ala Thr His Val Ile Ser Val Pro Arg Asn Gly Gly Met Gly Val Ala
        740              745              750
Glu Ile Leu Ala Arg Asn Glu Val Thr Val Leu Phe Ala Val Pro Ala
        755              760              765
Val Tyr Lys Leu Ile Ile Ser Lys Asn Asp Arg Gly Leu Arg Leu Pro
        770              775              780
Lys Leu Arg Leu Cys Ile Ser Ala Gly Glu Lys Leu Pro Leu Lys Leu
785              790              795              800
Tyr Arg Glu Ala Arg Ser Phe Phe Ser Val Asn Val Leu Asp Gly Ile
        805              810              815
Gly Cys Thr Glu Ala Ile Ser Thr Phe Ile Ser Asn Arg Glu Ser Tyr
        820              825              830
Val Ala Pro Gly Cys Thr Gly Val Val Val Pro Gly Phe Glu Val Lys
        835              840              845
Leu Val Asn Pro Arg Gly Glu Leu Cys Arg Val Gly Glu Val Gly Val
        850              855              860
Leu Trp Val Arg Gly Gly Ala Leu Thr Arg Gly Tyr Val Asn Ala Pro
865              870              875              880
Asp Leu Thr Glu Lys His Phe Val Asp Gly Trp Phe Asn Thr Gln Asp
        885              890              895
Met Phe Phe Met Asp Ala Glu Tyr Arg Leu Tyr Asn Val Gly Arg Ala
        900              905              910
Gly Ser Val Ile Lys Ile Asn Ser Cys Trp Phe Ser Pro Glu Met Met
        915              920              925
Glu Ser Val Leu Gln Ser His Pro Ala Val Lys Glu Cys Ala Val Cys
        930              935              940
Val Val Ile Asp Asp Tyr Gly Leu Pro Arg Pro Lys Ala Phe Ile Val
945              950              955              960
Thr Gly Glu His Glu Arg Ser Glu Pro Glu Leu Glu His Leu Trp Ala
        965              970              975
Glu Leu Arg Val Leu Ser Lys Glu Lys Leu Gly Lys Asp His Tyr Pro
        980              985              990
His Leu Phe Ala Thr Ile Lys Thr Leu Pro Arg Thr Ser Ser Gly Lys
        995              1000              1005
Leu Met Arg Ser Glu Leu Ala Lys Leu Leu Thr Ser Gly Pro Pro
        1010              1015              1020

```

```

<210> 52
<211> 38
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC_FEATURE
<222> (1)..(38)
<223> CysM

```

```

<400> 52
Met Asn Pro Lys Phe Leu Gly Gly Leu Gly Ala Gly Val Cys Ile Ala
1          5          10          15
Ser Leu Phe Gln Thr Val Met Arg Thr Val Pro Leu Lys Asp Ala Gly
        20          25          30
Ser Gly Asp Arg Ala Cys
        35

```

```

<210> 53
<211> 357
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC_FEATURE
<222> (1).. (357)
<223> CysN

<400> 53
Met Ser Thr Arg Thr Lys Asn Phe Asn Val Met Gly Ile Asp Trp Met
1      5      10      15
Pro Ser Ser Ala Glu Phe Lys Arg Arg Val Pro Arg Thr Gln Arg Ala
20     25     30
Ala Glu Ala Val Leu Ala Gly Arg Arg Cys Leu Met Asp Ile Leu Asp
35     40     45
Arg Gly Asp Pro Arg Leu Phe Val Ile Val Gly Pro Cys Ser Ile His
50     55     60
Asp Pro Val Ala Gly Leu Asp Tyr Ala Lys Arg Leu Arg Lys Leu Ala
65     70     75     80
Asp Glu Val Arg Glu Thr Leu Phe Val Val Met Arg Val Tyr Phe Glu
85     90     95
Lys Pro Arg Thr Thr Thr Gly Trp Lys Gly Phe Ile Asn Asp Pro Arg
100    105    110
Met Asp Gly Ser Phe His Ile Glu Gly Met Glu Arg Gly Arg Arg
115    120    125
Phe Leu Leu Asp Val Ala Glu Glu Gly Leu Pro Ala Ala Thr Glu Ala
130    135    140
Leu Asp Pro Ile Ala Ser Gln Tyr Tyr Gly Asp Leu Ile Ser Trp Thr
145    150    155    160
Ala Ile Gly Ala Arg Thr Ala Glu Ser Gln Thr His Arg Glu Met Ala
165    170    175
Ser Gly Leu Ser Thr Pro Val Gly Phe Lys Asn Gly Thr Asp Gly Ser
180    185    190
Leu Asp Ala Ala Val Asn Gly Ile Ile Ser Ala Ser His Pro His Ser
195    200    205
Phe Leu Gly Val Ser Glu Asn Gly Ala Cys Ala Ile Ile Arg Thr Arg
210    215    220
Gly Asn Thr Tyr Gly His Leu Val Leu Arg Gly Gly Gly Gly Arg Pro
225    230    235    240
Asn Tyr Asp Ala Val Ser Val Ala Leu Ala Glu Lys Ala Leu Ala Lys
245    250    255
Ala Arg Leu Pro Thr Asn Ile Val Val Asp Cys Ser His Ala Asn Ser
260    265    270
Trp Lys Asn Pro Glu Leu Gln Pro Leu Val Met Arg Asp Val Val His
275    280    285
Gln Ile Arg Glu Gly Asn Arg Ser Val Val Gly Leu Met Ile Glu Ser
290    295    300
Phe Ile Glu Ala Gly Asn Gln Pro Ile Pro Ala Asp Leu Ser Gln Leu
305    310    315    320
Arg Tyr Gly Cys Ser Val Thr Asp Ala Cys Val Asp Trp Lys Thr Thr
325    330    335
Glu Lys Met Leu Tyr Ser Ala His Glu Glu Leu Leu His Ile Leu Pro
340    345    350
Arg Ser Lys Val Ala
355

```

[0058]

```

<210> 54
<211> 203
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC_FEATURE
<222> (1).. (203)
<223> CysO

```

```

<400> 54
Met Pro Ala Arg Ser Thr Pro Ser Leu Glu Ser Gly Asp Phe Phe Ala
1      5      10      15
Asp Val Thr Phe Ser Asp Leu Ser Ile Glu Ser Ala Asp Leu Ser Gly
20     25     30
Lys Glu Phe Glu Arg Cys Thr Phe Arg Arg Cys Lys Leu Pro Glu Ser
35     40     45
Arg Trp Val Arg Ser Arg Leu Glu Asp Cys Val Phe Glu Gly Cys Asp

```

```

      50              55              60
Leu Leu Arg Met Val Pro Glu Lys Leu Ala Leu Arg Ser Val Thr Phe
65
Lys Asp Thr Arg Leu Met Gly Val Asp Trp Ser Gly Leu Gly Thr Met
      85              90              95
Pro Asp Val Gln Phe Glu Gln Cys Asp Leu Arg Tyr Ser Ser Phe Leu
      100             105             110
Lys Leu Asn Leu Arg Lys Thr Arg Phe Val Gly Cys Ser Ala Arg Glu
      115             120             125
Ala Asn Phe Ile Asp Val Asp Leu Ala Glu Ser Asp Phe Thr Gly Thr
      130             135             140
Asp Met Pro Gly Cys Thr Met Gln Gly Cys Val Leu Thr Lys Thr Asn
145             150             155
Phe Ala Arg Ser Thr Asn Phe Ile Phe Asp Pro Lys Ala Asn Gln Val
      165             170             175
Lys Gly Thr Arg Val Gly Val Glu Thr Ala Val Ala Leu Ala Gln Ala
      180             185             190
Leu Gly Met Val Val Asp Gly Tyr Gln Thr Pro
      195             200

```

<210> 55  
 <211> 233  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (233)  
 <223> CysP

[0059]

```

<400> 55
Met Lys Arg Phe Phe Lys Leu Gln Leu Arg Thr Thr Asn Val Pro Ala
1      5              10              15
Ala Arg Ala Phe Tyr Thr Ala Leu Phe Gly Glu Gly Ala Ala Asn Ala
      20              25              30
Asp Ile Val Pro Leu Pro Glu Gln Ala Ile Ala Arg Gly Ala Pro Ala
      35              40              45
His Trp Leu Gly Tyr Val Gly Val Glu Asp Val Asp Glu Ala Val Arg
      50              55              60
Ser Phe Val Gly Arg Gly Ala Thr Gln Leu Gly Pro Thr His Pro Thr
65              70              75              80
Asn Asp Gly Gly Arg Val Ala Ile Leu Arg Asp Pro Gly Gly Ala Thr
      85              90              95
Phe Ala Val Ala Thr Ala Pro Ala Thr Thr Arg Ala Leu Gln Pro Glu
      100             105             110
Val Val Trp Gln Gln Leu Tyr Ala Ala Asn Val Gln Gln Thr Ala Ala
      115             120             125
Ser Tyr Cys Asp Leu Phe Gly Trp Arg Leu Ser Asp Arg Arg Asp Leu
      130             135             140
Gly Ala Leu Gly Val His Gln Glu Phe Thr Trp Arg Ser Asp Glu Pro
145             150             155             160
Ser Ala Gly Ser Val Val Asp Val Ala Gly Leu Lys Gly Val His Ser
      165             170             175
His Trp Leu Phe His Phe Arg Val Ala Ala Leu Asp Pro Ala Met Glu
      180             185             190
Val Val Arg Lys Ala Gly Gly Val Val Ile Gly Pro Met Glu Leu Pro
      195             200             205
Asn Gly Asp Arg Ile Ala Val Cys Glu Asp Pro Gln Arg Ala Ala Phe
      210             215             220
Ala Leu Arg Glu Ser Ser His Gly Arg
      225             230

```

<210> 56  
 <211> 264  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (264)  
 <223> CysQ

```

<400> 56
Met Gln Glu Ile Gly Gln Thr Ala Leu Trp Val Ala Gly Met Arg Ala
1      5      10      15
Leu Glu Thr Glu Arg Ser Asn Pro Leu Phe Arg Asp Pro Phe Ala Arg
20     25     30
Arg Leu Ala Gly Asp Thr Leu Val Glu Glu Leu Arg Arg Arg Asn Ala
35     40     45
Gly Glu Gly Ala Met Pro Pro Ala Ile Glu Val Arg Thr Arg Trp Leu
50     55     60
Asp Asp Gln Ile Thr Leu Gly Leu Gly Arg Gly Ile Arg Gln Ile Val
65     70     75     80
Ile Leu Ala Ala Gly Met Asp Ala Arg Ala Tyr Arg Leu Ala Trp Pro
85     90     95
Gly Asp Thr Arg Leu Phe Glu Leu Asp His Asp Ala Val Leu Gln Asp
100    105    110
Lys Glu Ala Lys Leu Thr Gly Val Ala Pro Lys Cys Glu Arg His Ala
115    120    125
Val Ser Val Asp Leu Ala Asp Asp Trp Pro Ala Ala Leu Lys Lys Ser
130    135    140
Gly Phe Asp Pro Gly Val Pro Thr Leu Trp Leu Ile Glu Gly Leu Leu
145    150    155    160
Val Tyr Leu Thr Glu Ala Gln Val Thr Leu Leu Met Ala Arg Val Asn
165    170    175
Ala Leu Ser Val Pro Glu Ser Ile Val Leu Ile Asp Val Val Gly Arg
180    185    190
Ser Ile Leu Asp Ser Ser Arg Val Lys Leu Met His Asp Leu Ala Arg
195    200    205
Gln Phe Gly Thr Asp Glu Pro Glu Val Ile Leu Arg Pro Ile Gly Trp
210    215    220
Asp Pro His Val Tyr Thr Thr Ala Ala Ile Gly Lys Gln Leu Gly Arg
225    230    235    240
Trp Pro Phe Pro Val Ala Pro Arg Gly Thr Pro Gly Val Pro Gln Gly
245    250    255
Tyr Leu Val His Gly Val Lys Arg
260

```

[0060]

```

<210> 57
<211> 333
<212> PRT
<213> 深棕色孢囊杆菌

<220>
<221> MISC_FEATURE
<222> (1)..(333)
<223> CysR

```

```

<400> 57
Val Asn Gly Thr Thr Gly Lys Thr Gly Leu Val Ala Glu Arg Ser Gly
1      5      10      15
Ala Ile Ser Pro Arg Asp Tyr Lys Ser Lys Glu Leu Val Trp Asp Ser
20     25     30
Leu Ala Ala Thr Arg Ser Lys Pro Arg Arg Val Leu Pro Glu Gly Asp
35     40     45
Val Val Gly His Leu Tyr Pro Pro Ala Lys Ala Ala Leu Leu Thr His
50     55     60
Pro Leu Met Lys Asn Leu Pro Pro Glu Thr Leu Arg Leu Phe Phe Ile
65     70     75     80
His Ser Ala Tyr Lys Phe Met Gly Asp Ile Ala Ile Phe Glu Thr Glu
85     90     95
Thr Val Asn Glu Val Ala Met Lys Ile Ala Asn Gly His Thr Pro Ile
100    105    110
Thr Phe Pro Asp Asp Ile Arg His Asp Ala Leu Thr Val Ile Ile Asp
115    120    125
Glu Ala Tyr His Ala Tyr Val Ala Arg Asp Phe Met Arg Gln Ile Glu
130    135    140
Gln Arg Thr Gly Val Lys Pro Leu Pro Leu Gly Thr Glu Thr Asp Leu
145    150    155    160
Ser Arg Ala Met Ala Phe Gly Lys His Arg Leu Pro Glu Thr Leu His
165    170    175
Gly Leu Trp Glu Ile Ile Ala Val Cys Ile Gly Glu Asn Thr Leu Thr

```

```

      180      185      190
Lys Asp Leu Leu Asn Leu Thr Gly Glu Lys Ser Phe Asn Glu Val Leu
      195      200      205
His Gln Val Met Glu Asp His Val Arg Asp Glu Gly Arg His Ala Val
      210      215      220
Leu Phe Met Asn Val Leu Lys Leu Val Trp Ser Glu Met Glu Glu Ser
      225      230      235      240
Ala Arg Leu Ala Ile Gly Gln Leu Leu Pro Glu Phe Ile Arg Glu Tyr
      245      250      255
Leu Ser Pro Lys Met Met Ala Glu Tyr Glu Arg Val Val Leu Glu Gln
      260      265      270
Leu Gly Leu Ala Ala Glu His Ile Glu Arg Ile Leu Ser Glu Thr Tyr
      275      280      285
Ser Glu Pro Pro Leu Glu Asp Phe Arg Ala Arg Tyr Pro Leu Ser Gly
      290      295      300
Tyr Leu Val Tyr Val Leu Met Gln Cys Asp Val Leu Ser His Ala Pro
      305      310      315      320
Thr Arg Glu Ala Phe Arg Arg Phe Lys Leu Leu Ala His
      325      330

```

```

<210> 58
<211> 642
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC_FEATURE
<222> (1).. (642)
<223> CysS

```

[0061]

```

<400> 58
Met Ala Asn Gln Arg Val Ala Phe Ile Glu Leu Thr Val Phe Ser Gly
1      5      10      15
Val Tyr Pro Leu Ala Ser Gly Tyr Met Arg Gly Val Ala Glu Gln Asn
      20      25      30
Pro Leu Ile Arg Glu Ser Cys Ser Phe Glu Ile His Ser Ile Cys Ile
      35      40      45
Asn Asp Asp Arg Phe Glu Asp Lys Leu Asn Lys Ile Asp Ala Asp Val
      50      55      60
Tyr Ala Ile Ser Cys Tyr Val Trp Asn Met Gly Phe Val Lys Arg Trp
      65      70      75      80
Leu Pro Thr Leu Thr Ala Arg Lys Pro Asn Ala His Ile Ile Leu Gly
      85      90      95
Gly Pro Gln Val Met Asn His Gly Ala Gln Tyr Leu Asp Pro Gly Asn
      100      105      110
Glu Arg Val Val Leu Cys Asn Gly Glu Gly Glu Tyr Thr Phe Ala Asn
      115      120      125
Tyr Leu Ala Glu Leu Cys Ser Pro Gln Pro Asp Leu Gly Lys Val Lys
      130      135      140
Gly Leu Ser Phe Tyr Arg Asn Gly Glu Leu Ile Thr Thr Glu Pro Gln
      145      150      155      160
Ala Arg Ile Gln Asp Leu Asn Thr Val Pro Ser Pro Tyr Leu Glu Gly
      165      170      175
Tyr Phe Asp Ser Glu Lys Tyr Val Trp Ala Pro Leu Glu Thr Asn Arg
      180      185      190
Gly Cys Pro Tyr Gln Cys Thr Tyr Cys Phe Trp Gly Ala Ala Thr Asn
      195      200      205
Ser Arg Val Phe Lys Ser Asp Met Asp Arg Val Lys Ala Glu Ile Thr
      210      215      220
Trp Leu Ser Gln His Arg Ala Phe Tyr Ile Phe Ile Thr Asp Ala Asn
      225      230      235      240
Phe Gly Met Leu Thr Arg Asp Ile Glu Ile Ala Gln His Ile Ala Glu
      245      250      255
Cys Lys Arg Lys Tyr Gly Tyr Pro Leu Thr Ile Trp Leu Ser Ala Ala
      260      265      270
Lys Asn Ser Pro Asp Arg Val Thr Gln Ile Thr Arg Ile Leu Ser Gln
      275      280      285
Glu Gly Leu Ile Ser Thr Gln Pro Val Ser Leu Gln Thr Met Asp Ala
      290      295      300
Asn Thr Leu Lys Ser Val Lys Arg Gly Asn Ile Lys Glu Ser Ala Tyr
      305      310      315      320

```



[0062]

Leu Ser Leu Gln Glu Glu Leu His Arg Ser Lys Leu Ser Ser Phe Val  
 325 330 335  
 Glu Met Ile Trp Pro Leu Pro Gly Glu Thr Leu Glu Thr Phe Arg Glu  
 340 345 350  
 Gly Ile Gly Lys Leu Cys Ser Tyr Asp Ala Asp Ala Ile Leu Ile His  
 355 360 365  
 His Leu Leu Leu Ile Asn Asn Val Pro Met Asn Ser Gln Arg Glu Glu  
 370 375 380  
 Phe Lys Leu Glu Val Ser Asn Asp Glu Asp Pro Asn Ser Glu Ala Gln  
 385 390 395 400  
 Val Val Val Ala Thr Lys Asp Val Thr Arg Glu Glu Tyr Lys Glu Gly  
 405 410 415  
 Val Arg Phe Gly Tyr His Leu Thr Ser Leu Tyr Ser Leu Arg Ala Leu  
 420 425 430  
 Arg Phe Val Gly Arg Tyr Leu Asp Lys Gln Gly Arg Leu Ala Phe Lys  
 435 440 445  
 Asp Leu Ile Ser Ser Phe Ser Glu Tyr Cys Lys Arg Asn Pro Asp His  
 450 455 460  
 Pro Tyr Thr Gln Tyr Ile Thr Ser Val Ile Asp Gly Thr Ser Gln Ser  
 465 470 475 480  
 Lys Phe Ser Ala Asn Gly Gly Ile Phe His Val Thr Leu His Glu Phe  
 485 490 495  
 Arg Arg Glu Phe Asp Gln Leu Leu Phe Gly Phe Ile Gln Thr Leu Gly  
 500 505 510  
 Met Met Asn Asp Glu Leu Leu Glu Phe Leu Phe Glu Met Asp Leu Leu  
 515 520 525  
 Asn Arg Pro His Val Tyr Ser Asn Thr Pro Ile Asn Asn Gly Glu Gly  
 530 535 540  
 Leu Leu Lys His Val Thr Val Val Ser Lys Glu Lys Asp Ala Ile Val  
 545 550 555 560  
 Leu Arg Val Pro Glu Lys Tyr Ala Gln Leu Thr Ser Glu Leu Leu Gly  
 565 570 575  
 Leu Glu Gly Ala Pro Ser Thr Ser Leu Arg Val Lys Tyr Arg Gly Thr  
 580 585 590  
 Gln Met Pro Phe Met Ala Asn Lys Pro Tyr Glu Asp Asn Leu Ser Tyr  
 595 600 605  
 Cys Glu Ala Lys Leu His Lys Met Gly Ser Ile Leu Pro Val Trp Glu  
 610 615 620  
 Ser Ala Val Pro Ser Arg Thr Pro Val Arg Arg Pro Gln Val Ala Val  
 625 630 635 640  
 Ala Gly

<210> 59  
 <211> 1267  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (1267)  
 <223> CysT

<400> 59  
 Met His Arg Val Lys Pro Leu Ile Gly Pro Val Leu Ser Ala Leu Leu  
 1 5 10 15  
 Leu Cys Ala Leu Pro Ala Arg Ala Gln Ile Ala Ala Ala His Val Tyr  
 20 25 30  
 His Asn His Met Pro Asn Phe Trp Ala Tyr Tyr Asp Leu Gly Gln Tyr  
 35 40 45  
 Ala Ser Thr Pro Thr Gly Gly Pro Ile Arg Tyr Met Tyr Asp Ala Gln  
 50 55 60  
 Val Ile Asn Leu Lys Lys Asn Pro Pro Ser Asn Tyr Thr Tyr Tyr Leu  
 65 70 75 80  
 Pro Ser Gly Ala Pro Met Pro His Asp Asp Leu Val Thr Tyr Tyr Ser  
 85 90 95  
 His Asn Ala Lys Thr Gly Ala Tyr Leu Tyr Trp Pro Pro Ser Val Ala  
 100 105 110  
 Ser Asp Met Lys Thr Asn Ala Pro Thr Gly Gln Val His Val Thr Met  
 115 120 125  
 Ser Gly Ala Val Val Asn Asn Val Gln Asp Leu Val Thr Leu Lys Asn

[0063]

130	135	140
Val Pro Gly Tyr Asp Asn	Pro Asn Trp Gly Ala	Ser Trp Lys Asp Arg
145	150	155
Tyr Ser Ala Leu Leu Thr	Pro Ala Gly Asn Arg Thr	Leu Asp Leu Ile
165	170	175
His Phe Thr Gly His His	Ser Met Gly Pro Leu Val	Gly Pro Asp Tyr
180	185	190
Phe Leu Lys Asp Leu Ile	Tyr Gln Ser Ala Thr	Leu Ala Gln Pro Tyr
195	200	205
Phe Leu Gly Gly Ser Phe	Gln Ser Ser Lys Gly Phe	Phe Pro Thr Glu
210	215	220
Leu Gly Phe Ser Glu Arg	Leu Ile Pro Thr Leu	Ser Lys Leu Gly Val
225	230	235
Gln Trp Ala Val Ile Gly	Asp Asn His Phe Ser	Arg Thr Leu Lys Asp
245	250	255
Tyr Pro Tyr Leu Asn Asp	Pro Gly Ser Asp Thr	Leu Val Ser Pro Pro
260	265	270
Asn Arg Ala Asp Leu Gln	Asn Thr Ser Ser Val	Gly Ser Trp Val Ser
275	280	285
Ala Gln Met Ala His Glu	Gln Gln Val Ile Lys	Asn Lys Tyr Pro Phe
290	295	300
Ala Ser Thr Pro His Trp	Val Arg Tyr Val Asp	Pro Ala Thr Gly Ala
305	310	315
Glu Ser Arg Val Val Gly	Ile Pro Val Asn Gln	Asn Gly Ser Trp Leu
325	330	335
Glu Gly Trp Glu Gly Glu	Ala Thr Val Asp Val	Val Asn Leu Lys Ser
340	345	350
Phe Glu Gly Leu Val Pro	Gln Arg Gln Phe Phe	Val Ile Ala His Asp
355	360	365
Gly Asp Asn Ser Ser Gly	Arg Ala Gly Ser Asp	Ser Thr Trp Tyr Asn
370	375	380
Gly Arg Ser Val Thr Cys	Ala Asn Gly Val Gln	Cys Val Gly Ile Ser
385	390	395
Glu Tyr Leu Val His His	Thr Pro Ala Ser Thr	Asp Val Val His Val
405	410	415
Gln Asp Gly Ser Trp Val	Asp Thr Arg Asp Ser	Ser Ser Asp Pro Gln
420	425	430
Trp His His Trp Lys Leu	Pro Phe Gly Ile Trp	Lys Gly Gln Phe Pro
435	440	445
Ala Phe Asn Ala Ala Thr	Gly Leu Asn Leu Ser	Pro Lys Thr Asn Leu
450	455	460
Ser Gly Val Gln Glu Gly	Met Thr Val Ser Leu	Glu His Gly Trp His
465	470	475
Tyr Leu Glu Arg Asn Phe	Ala Leu Leu Gln Ala	Ala Leu Asn Tyr Ala
485	490	495
Lys Thr Ala Glu Gln Ile	Trp Leu Asp Ala His	Pro Asn His Trp Ser
500	505	510
Pro Thr Thr Ala Ile Asp	Lys Gln Ile Thr His	Thr Gly Asn Gln Leu
515	520	525
Asn Pro Trp Met Met Ser	Phe Pro Val Lys Gly	Asp Val Asn Asn Asp
530	535	540
Trp Ala Gly Gly Ala Asn	Pro Ala Glu Leu Ala	Trp Tyr Phe Leu Leu
545	550	555
Pro Ala Met Asp Ser Gly	Phe Gly Tyr Tyr Asp	Glu Asn Gln Asp Asp
565	570	575
Asn Val Lys Pro Thr Leu	Ser Phe Asn Gln Ser	Leu Tyr Phe Ser Lys
580	585	590
Pro Tyr Val Gln Gln Arg	Ile Ala Gln Asp Lys	Thr Gly Pro Ser Val
595	600	605
Trp Trp Ala Gln Arg Trp	Pro Tyr Asn Pro Gly	Ser Ala Asn Thr Asp
610	615	620
Lys Ser Glu Gly Trp Thr	Leu His Phe Phe Asn	Asn His Phe Ala Leu
625	630	635
Tyr Thr Tyr Ala Tyr Asp	Ala Ser Gly Ile Ser	Ser Ile Lys Ala Arg
645	650	655
Val Arg Val His Thr His	Lys Ser Ile Asp Pro	Leu Asp Asn Thr His
660	665	670
Lys Val Tyr Asp Pro Ala	Ala Arg Lys Ala Ala	Gly Val Pro Asn Ile
675	680	685
Asp Pro Ala Arg Val Gly	Ala Trp Val Asp Tyr	Pro Leu Thr Arg Arg
690	695	700

[0064]

```

Asp Leu Lys Pro Val Met Asn Gly Val Ser Trp Gln Pro Ala Tyr Leu
705              710              715              720
Pro Val Met Ala Lys Val Pro Ala Gln Glu Ile Gly Asp Leu Tyr Tyr
              725              730              735
Val Tyr Leu Gly Asn Tyr Arg Asp Gln Leu Leu Asp Tyr Tyr Ile Glu
              740              745              750
Ala Thr Asp Ser Arg Gly Asn Ile Thr Arg Gly Glu Ile Gln Ser Val
              755              760              765
Tyr Val Gly Ser Gly Arg Tyr Asn Leu Val Gly Gly Lys Tyr Ile Glu
              770              775              780
Asp Pro Asn Gly Thr Val Gln Gly Thr His Pro Phe Leu Val Val Asp
785              790              795              800
Thr Thr Ala Pro Ser Val Pro Ser Gly Leu Thr Ala Lys Ala Lys Thr
              805              810              815
Asp Arg Ser Val Thr Leu Ser Trp Ser Ala Ala Ser Asp Asn Val Ala
              820              825              830
Val Ser Gly Tyr Asp Val Phe Arg Asp Gly Thr Gln Val Gly Ser Ser
              835              840              845
Thr Ser Thr Ala Tyr Thr Asp Ser Gly Leu Ser Pro Ser Thr Gln Tyr
              850              855              860
Ser Tyr Thr Val Arg Ala Arg Asp Ala Ala Gly Asn Ala Ser Ala Gln
865              870              875              880
Ser Thr Ala Leu Ser Val Ala Thr Leu Thr Pro Asp Thr Thr Pro Pro
              885              890              895
Ser Val Pro Ser Gly Leu Thr Ala Ser Gly Thr Thr Ser Ser Ser Val
              900              905              910
Ala Leu Ala Trp Thr Ala Ser Thr Asp Asn Tyr Gly Val Ala Asn Tyr
              915              920              925
Glu Val Leu Arg Asn Gly Thr Gln Val Ala Ser Val Thr Gly Thr Thr
              930              935              940
Tyr Ser Asp Thr Gly Leu Ser Pro Ser Thr Thr Tyr Ser Tyr Thr Val
945              950              955              960
Arg Ala Arg Asp Ala Ala Gly Asn Val Ser Ser Pro Ser Thr Ala Leu
              965              970              975
Ser Val Thr Thr Gln Thr Gly Asn Ser Ala Thr Val Tyr Tyr Phe Asn
              980              985              990
Asn Asn Phe Ala Leu Lys Tyr Ile His Phe Arg Ile Gly Gly Gly Thr
              995              1000              1005
Trp Thr Thr Val Pro Gly Asn Val Met Ala Thr Ser Glu Val Pro
              1010              1015              1020
Gly Tyr Ala Lys Tyr Thr Val Asn Leu Gly Ala Ala Thr Gln Leu
              1025              1030              1035
Glu Cys Val Phe Asn Asp Gly Lys Gly Thr Trp Asp Asn Asn Lys
              1040              1045              1050
Gly Asn Asn Tyr Leu Leu Pro Ala Gly Thr Ser Thr Val Lys Asp
              1055              1060              1065
Gly Val Val Ser Ser Gly Ala Pro Ala Leu Asp Thr Thr Ala Pro
              1070              1075              1080
Ser Val Pro Ser Gly Leu Thr Ala Ala Ser Lys Thr Ser Ser Ser
              1085              1090              1095
Val Ser Leu Ser Trp Ser Ala Ser Thr Asp Ala Ser Gly Ile Ala
              1100              1105              1110
Gly Tyr Asp Val Tyr Arg Asp Gly Ser Leu Val Gly Ser Pro Val
              1115              1120              1125
Ser Thr Ser Tyr Thr Asp Ser Asp Leu Ser Ala Gly Thr Thr Tyr
              1130              1135              1140
Arg Tyr Thr Val Arg Ala Arg Asp Thr Ala Gly Asn Ala Ser Ala
              1145              1150              1155
Gln Ser Thr Ala Leu Ser Val Thr Thr Ser Thr Ser Ser Ala Thr
              1160              1165              1170
Ser Val Thr Phe Asn Val Thr Ala Ser Thr Val Val Gly Gln Asn
              1175              1180              1185
Val Tyr Leu Val Gly Asn His Ala Ala Leu Gly Asn Trp Asn Thr
              1190              1195              1200
Gly Ala Ala Ile Leu Leu Ser Pro Ala Ser Tyr Pro Lys Trp Ser
              1205              1210              1215
Val Thr Leu Ser Leu Pro Gly Ser Thr Ala Leu Glu Tyr Lys Tyr
              1220              1225              1230
Ile Lys Lys Asp Gly Ser Gly Asn Val Thr Trp Glu Ser Gly Ala
              1235              1240              1245
Asn Arg Ser Thr Thr Ile Pro Ala Ser Gly Thr Ala Thr Leu Asn

```

```

1250      1255      1260
Asp Thr Trp Lys
1265

<210> 60
<211> 276
<212> PRT
<213> 深棕色孢囊杆菌

<220>
<221> MISC_FEATURE
<222> (1)..(276)
<223> ORF1

<400> 60
Val Pro His Pro Ser Glu Gln Ser Ala Pro Ser Gly Leu Arg Ala Arg
1      5      10      15
Leu His Glu Ile Ile Phe Glu Ser Asp Thr Pro Ala Gly Arg Ala Phe
20     25     30
Asp Val Ala Leu Leu Trp Ala Ile Val Leu Ser Val Leu Ala Val Met
35     40     45
Leu Glu Ser Val Glu Ser Ile Ser Val Gln His Gly Gln Thr Ile Arg
50     55     60
Val Leu Glu Trp Cys Phe Thr Gly Leu Phe Thr Leu Glu Tyr Val Leu
65     70     75     80
Arg Leu Leu Ser Val Lys Arg Pro Leu Arg Tyr Ala Leu Ser Phe Phe
85     90     95
Gly Leu Val Asp Leu Leu Ala Ile Leu Pro Ser Val Leu Ser Leu Met
100    105    110
Leu Pro Gly Met Gln Ser Leu Leu Val Val Arg Val Phe Arg Leu Leu
115    120    125
Arg Val Phe Arg Val Leu Lys Leu Ala Ser Phe Leu Gly Glu Ala Asp
130    135    140
Val Leu Leu Thr Ala Leu Arg Ala Ser Arg Arg Lys Ile Ile Val Phe
145    150    155    160
Leu Gly Ala Val Leu Ser Thr Val Val Ile Met Gly Ala Val Met Tyr
165    170    175
Met Val Glu Gly Arg Ala Asn Gly Phe Asp Ser Ile Pro Arg Gly Met
180    185    190
Tyr Trp Ala Ile Val Thr Met Thr Thr Val Gly Tyr Gly Asp Leu Ser
195    200    205
Pro Lys Thr Val Pro Gly Gln Phe Ile Ala Ser Val Leu Met Ile Met
210    215    220
Gly Tyr Gly Ile Leu Ala Val Pro Thr Gly Ile Val Ser Val Glu Leu
225    230    235    240
Ala Gln Ala Thr Arg Gln His Ala Ile Asp Pro Arg Ala Cys Pro Gly
245    250    255
Cys Gly Leu Gln Gly His Asp Leu Asp Ala His His Cys Lys His Cys
260    265    270
Gly Thr Ala Leu
275

<210> 61
<211> 78
<212> PRT
<213> 深棕色孢囊杆菌

<220>
<221> MISC_FEATURE
<222> (1)..(78)
<223> ORF2

<400> 61
Met Ala Gln Asp Gln Asp Arg Glu Lys Leu His Ser Asp Ala Asp Lys
1      5      10      15
Glu Arg Leu His Pro Lys Val Asp Ser Gly Asp Val Ser Gly Arg Gly
20     25     30
Arg Glu Arg Arg Pro Asp Glu Glu Tyr Pro Lys Gln Arg Asn Ala Gly
35     40     45
Glu Phe Gly Thr His Gly Gly Pro Asn Lys Gly Gly Lys Glu Asp Arg
50     55     60

```

Arg Gln Leu His Ala Pro Gly Ser Ser Lys Ala Gly Ser Gln  
65 70 75

<210> 62  
<211> 162  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1)..(162)  
<223> ORF3

<400> 62  
Met Gly Arg Thr Tyr Ser Phe Glu Pro Phe Leu Ser Gln Gln Pro Ala  
1 5 10 15  
Gln Thr Tyr Lys Gly Ser Gly Pro Arg Leu Gly Asn Glu Glu His Lys  
20 25 30  
Ile Ala Leu Thr Lys Glu Glu Glu Lys Ala Ala Leu Pro Asp Thr Pro  
35 40 45  
Thr Gly Tyr Gly Gln Ala His Ala Glu Thr Val Lys Arg Tyr Arg Ala  
50 55 60  
Arg Ala Glu Lys Lys Arg Thr Glu Pro Lys Thr Pro Ala Thr Arg Ala  
65 70 75 80  
Lys Lys Ala Ala Pro Lys Ala Lys Pro Thr Arg Lys Val Ala Thr Gln  
85 90 95  
Glu Ala Thr Ala Lys Ala Pro Thr Arg Gln Ala Arg Glu Glu Thr Glu  
100 105 110  
Pro Lys Ala Pro Ala Arg Lys Lys Leu Ser Ala Thr Gly Leu Val Gly  
115 120 125  
Ser Ile Gly Arg Lys Val Val Thr Arg Ala Ala Val Ala Ala Lys Lys  
130 135 140  
Thr Val Ala Arg Ala Val Lys Thr Ala Ala Ala Arg Lys Ser Ala Lys  
145 150 155 160  
Lys Arg

[0066]

<210> 63  
<211> 87  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1)..(87)  
<223> ORF4

<400> 63  
Met Ser Pro Ala Arg Arg Lys Glu Ser Lys Gln His Glu Val Gly Ser  
1 5 10 15  
Ala Thr His Ala Arg Arg Val Ile Val Ala Thr Asp Gly Arg Gly Trp  
20 25 30  
Tyr Val Arg Phe Glu Gly Asn Arg Gln Leu Gly Arg Tyr Ser Asn Val  
35 40 45  
Thr Gln Ala Ile His Gly Gly Arg Arg Leu Ala Arg Gln His Lys Pro  
50 55 60  
Ala Gly Leu Val Val Arg Tyr Leu Asp Gly Glu Glu Glu Ser Trp  
65 70 75 80  
Tyr Gly Asp Arg Glu Ala Pro  
85

<210> 64  
<211> 149  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1)..(149)  
<223> ORF5

```

<400> 64
Met Lys His Ile Lys Ala Val Val Val Gly Ala Leu Ser Ala Ala Leu
1          5          10          15
Leu Phe Gly Val Gly Cys Gln Thr Thr Gly Gly Ala Gly Asn Gln Gly
          20          25          30
Thr Gly Gly Ser Asp Thr Ser Gln Gly Gly Thr Met Thr Gly Ser Glu
          35          40          45
Thr Thr Gly Thr Gly Thr Thr Gly Gly Thr Thr Glu Gly Gly Asp Thr
          50          55          60
Thr Gly Gly Gly Thr Gly Gly Thr Gly Ala Gly Asp Ile Asp Gly Ser
65          70          75          80
Ser Ser Gly Ser Thr Gly Ser Gly Ser Asp Val Gly Gly Ser Gly Gly
          85          90          95
Ser Gly Val Ser Ser Glu Pro Gly Gly Phe Ser Pro Asp Ala Ser Gly
          100          105          110
Val Asp Ser Asp Leu Gly Gly Ser Gly Thr Gly Ser Asp Val Asp Gly
          115          120          125
Ser Gly Ser Thr Asp Ser Ser Gly Asn Met Ser Gly Thr Gly Ser Glu
          130          135          140
Asp Asp Thr Ser Arg
145

```

```

<210> 65
<211> 525
<212> PRT
<213> 深棕色孢囊杆菌

```

```

<220>
<221> MISC FEATURE
<222> (1).. (525)
<223> ORF6

```

[0067]

```

<400> 65
Met Ser Thr Arg Thr Ser Leu Ala Leu Ala Ala Ser Leu Ala Ala Leu
1          5          10          15
Pro Ala Leu Ala Gln Glu Arg Pro Ser Glu Gly Asp Leu Phe Gly Gly
          20          25          30
Asp Thr Pro Glu Thr Lys Pro Ala Pro Ala Asp Ala Pro Arg Pro Asp
          35          40          45
Glu Ser Ser Leu Phe Gly Asp Thr Pro Ala Ser Thr Pro Ala Ala Gln
          50          55          60
Ser Ala Ala Ala Thr Ala Ala Pro Asp Lys Pro Ser Ala Thr Pro Gln
65          70          75          80
Asp Arg Asp Ala Gln Ala Leu Gly Gly Pro Ser Ala Thr Asn Ala Phe
          85          90          95
Asp Thr Glu Glu Ala Val Glu Asp Pro Leu Lys Ile Gly Gly Arg Phe
          100          105          110
Tyr Leu Arg Ala Tyr Ser Gln Ala Asn Glu Gly Val Ser Phe Ser Asn
          115          120          125
Thr Thr Phe Ser Ala Pro Met Leu Val Asp Gly Tyr Phe Asp Ala Arg
          130          135          140
Pro Thr Glu Arg Leu Arg Gly Phe Val Leu Gly Arg Leu Thr Phe Asp
145          150          155          160
Pro Thr Arg Lys Ala Gly Ser Leu Gly Ile Val Pro Thr Ser Thr Ser
          165          170          175
Thr Ser Asn Val Ala Ala Asp Pro Val Val Leu Leu Asp Gln Ala Trp
          180          185          190
Leu Arg Phe Asp Leu Asp His Lys Leu Phe Ile Thr Val Gly Lys Gln
          195          200          205
His Val Lys Trp Gly Thr Ser Arg Phe Trp Asn Pro Thr Asp Phe Leu
          210          215          220
Ser Pro Gln Arg Arg Asp Pro Leu Ala Leu Leu Asp Thr Arg Thr Gly
225          230          235          240
Ala Thr Met Leu Lys Met His Met Pro Trp Glu Ala Lys Gly Trp Asn
          245          250          255
Phe Tyr Val Leu Gly Leu Leu Asp Asn Ala Gly Pro Ala Asn Thr Leu
          260          265          270
Gly Arg Val Gly Gly Ala Ala Arg Ala Glu Val Val Leu Gly His Thr
          275          280          285
Glu Leu Gly Val Asp Ala Val Leu Gln His Gly Arg Lys Pro Arg Phe
          290          295          300

```

Gly Leu Asp Leu Ser Ser Gly Leu Gly Pro Ile Asp Ile Tyr Gly Glu  
 305 310 315 320  
 Leu Ala Leu Lys Lys Gly Ser Asp Ala Pro Met Phe Arg Met Pro Gln  
 325 330 335  
 Gly Val Ser Leu Gly Asp Leu Leu Gly Gln Phe Gln Gly Asn Gly Gly  
 340 345 350  
 Met Pro Pro Asp Leu Gly Ala Leu Pro Ile Glu Ala Tyr Tyr Pro Glu  
 355 360 365  
 Gly Tyr Thr Pro Gln Val Ser Gly Gly Ala Thr Trp Thr Phe Ala Tyr  
 370 375 380  
 Ser Glu Ser Asp Thr Ala Thr Val Gly Val Glu Tyr Phe Tyr Asn Ser  
 385 390 395 400  
 Met Gly Tyr Pro Gly Ser Leu Ala Tyr Pro Tyr Leu Ile Leu Gln Gly  
 405 410 415  
 Gln Tyr Gln Pro Phe Tyr Leu Gly Arg His Tyr Ala Ala Val Tyr Ala  
 420 425 430  
 Phe Leu Ser Gly Pro Gly Ser Trp Asp Asn Thr Asn Phe Ile Leu Ser  
 435 440 445  
 Asn Leu Gly Asn Leu Ser Asp Arg Ser Phe Ile Thr Arg Leu Asp Val  
 450 455 460  
 Thr His Arg Ala Leu Arg Tyr Leu Ser Ile Glu Ala Phe Ile Ala Ala  
 465 470 475 480  
 Asn Tyr Gly Gln Arg Gly Gly Glu Phe Arg Phe Ala Leu Asn Leu Pro  
 485 490 495  
 Ala Leu Arg Met Gly Glu Gln Val Thr Pro Pro Ile Ala Val Ala Pro  
 500 505 510  
 Pro Thr Ile Gln Ala Gly Val Gly Leu Arg Ile Asp Leu  
 515 520 525

<210> 66  
 <211> 261  
 <212> PRT  
 <213> 深棕色孢囊杆菌

[0068]

<220>  
 <221> MISC\_FEATURE  
 <222> (1).. (261)  
 <223> ORF7

<400> 66  
 Met Thr Leu Arg Asn Leu Leu Gly Ala Leu Phe Ala Ala Leu Leu Leu  
 1 5 10 15  
 Ala Ala Pro Thr Ala Arg Ala Asp Leu Thr Asp Pro Ala Glu Ile Lys  
 20 25 30  
 Lys Leu Leu Glu Thr Leu Asp Asn Arg Gln Arg Asn Gly Gly Asp Tyr  
 35 40 45  
 Lys Ser Leu Val Tyr Ile Glu Gln Lys Glu Lys Asp Lys Thr Asp Val  
 50 55 60  
 Val Arg Glu Ala Val Val Tyr Arg Arg Asp Glu Lys Asp Gln Leu Met  
 65 70 75 80  
 Ile Leu Met Thr Lys Pro Lys Gly Glu Ala Gly Lys Gly Tyr Leu Arg  
 85 90 95  
 Leu Asp Lys Asn Leu Trp Ser Tyr Asp Pro Asn Thr Gly Lys Trp Asp  
 100 105 110  
 Arg Arg Thr Glu Arg Glu Arg Ile Ala Gly Thr Asp Ser Arg Arg Ala  
 115 120 125  
 Asp Phe Asp Glu Ser Arg Leu Ala Glu Glu Leu Asp Gly Lys Phe Glu  
 130 135 140  
 Gly Glu Glu Lys Leu Gly Lys Phe Thr Thr Trp Lys Leu Val Leu Thr  
 145 150 155 160  
 Ala Lys Pro Asn Val Asp Val Ala Tyr Pro Val Val His Leu Trp Val  
 165 170 175  
 Glu Lys Asp Thr Asn Asn Ile Leu Lys Arg Gln Glu Phe Ala Leu Ser  
 180 185 190  
 Gly Arg Leu Met Arg Thr Ser Tyr Phe Pro Lys Trp Met Lys Leu Phe  
 195 200 205  
 Ser Glu Ser Lys Lys Ala Asp Val Trp Tyr Pro Gln Glu Met Arg Phe  
 210 215 220  
 Tyr Asp Glu Val Glu Lys Thr Asn Ser Thr Val Ile Val Val Lys Ser  
 225 230 235 240  
 Val Asp Leu Arg Ser Leu Glu Glu Asn Ile Phe Thr Lys Ala Trp Phe

245  
 Glu Ser Lys Ser Arg  
 260  
 <210> 67  
 <211> 433  
 <212> PRT  
 <213> 深棕色孢囊杆菌  
 <220>  
 <221> MISC\_FEATURE  
 <222> (1).. (433)  
 <223> ORF8  
 <400> 67  
 Met Gln Gln Leu Leu Leu Ile Ala Val Arg Asn Leu Gly Thr His Lys  
 1 5 10 15  
 Arg Arg Thr Leu Leu Gly Gly Ala Ile Ala Gly Val Thr Ala Leu  
 20 25 30  
 Leu Val Ile Leu Met Gly Leu Ser Asn Gly Met Lys Asp Thr Met Leu  
 35 40 45  
 Arg Ser Ala Thr Thr Leu Val Thr Gly His Val Asn Val Ala Gly Phe  
 50 55 60  
 Tyr Lys Val Thr Ala Gly Gln Ser Ala Pro Val Val Thr Ser Tyr Pro  
 65 70 75 80  
 Lys Leu Leu Glu Gln Leu Arg Lys Glu Val Pro Glu Leu Asp Phe Ser  
 85 90 95  
 Val Gln Arg Thr Arg Gly Trp Val Lys Leu Val Ser Glu Ser Gly Ser  
 100 105 110  
 Val Gln Thr Gly Ile Gly Gly Ile Asp Val Ala Ala Glu Thr Gly Ile  
 115 120 125  
 Arg Lys Val Leu Gln Leu Arg Glu Gly Arg Leu Glu Asp Leu Ala Gln  
 130 135 140  
 Pro Asn Thr Leu Leu Leu Phe Asp Glu Gln Ala Lys Arg Leu Glu Val  
 145 150 155 160  
 Lys Val Gly Asp Ser Val Thr Leu Ser Ala Ser Thr Met Arg Gly Ile  
 165 170 175  
 Ser Asn Thr Val Asp Val Arg Val Val Ala Ile Ala Ala Asn Val Gly  
 180 185 190  
 Met Leu Ser Ser Phe Asn Val Leu Val Pro Asn Ala Thr Leu Arg Ala  
 195 200 205  
 Leu Tyr Gln Leu Arg Glu Asp Ser Thr Gly Ala Leu Met Leu His Leu  
 210 215 220  
 Lys Asp Met Ser Ala Ile Pro Ser Val Gln Ala Arg Leu Tyr Lys Arg  
 225 230 235 240  
 Leu Pro Glu Leu Gly Tyr Gln Val Leu Glu His Asp Pro Arg Ala Phe  
 245 250 255  
 Phe Met Lys Phe Gln Thr Val Asn Arg Glu Ala Trp Thr Gly Gln Lys  
 260 265 270  
 Leu Asp Ile Thr Asn Trp Glu Asp Glu Ile Ser Phe Ile Lys Trp Thr  
 275 280 285  
 Val Ser Ala Met Asp Ala Leu Thr Gly Val Leu Ile Phe Val Leu Leu  
 290 295 300  
 Ile Ile Ile Ala Val Gly Ile Met Asn Thr Leu Trp Ile Ala Ile Arg  
 305 310 315 320  
 Glu Arg Thr Arg Glu Ile Gly Thr Leu Arg Ala Ile Gly Met Gln Arg  
 325 330 335  
 Trp Tyr Val Leu Val Met Phe Leu Leu Glu Ala Leu Val Leu Gly Leu  
 340 345 350  
 Leu Gly Thr Thr Val Gly Ala Leu Val Gly Met Gly Val Cys Leu Leu  
 355 360 365  
 Ile Asn Ala Val Asp Pro Ser Val Pro Val Pro Val Gln Leu Phe Ile  
 370 375 380  
 Leu Ser Asp Lys Leu His Leu Ile Val Lys Pro Gly Ser Val Met Arg  
 385 390 395 400  
 Ala Ile Ala Phe Ile Thr Leu Cys Thr Thr Phe Ile Ser Leu Ile Pro  
 405 410 415  
 Ser Phe Leu Ala Ala Arg Met Lys Pro Ile Thr Ala Met His His Ile  
 420 425 430  
 Gly

[0069]



<210> 68  
 <211> 701  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1)..(701)  
 <223> ORF9

<400> 68  
 Met Gly Gln Leu Lys Leu Leu Leu Gln Val Ala Leu Arg Asn Leu Phe  
 1 5 10 15  
 Val Ser Arg Ile Asn Leu Leu Ile Gly Gly Ile Ile Phe Phe Gly Thr  
 20 25 30  
 Val Leu Val Val Val Gly Gly Ser Leu Val Asp Ser Val Asp Glu Ala  
 35 40 45  
 Met Ser Arg Ser Ile Ile Gly Ser Val Ala Gly His Leu Gln Val Tyr  
 50 55 60  
 Ser Ala His Ser Lys Asp Glu Leu Ser Leu Phe Gly Gln Met Gly Arg  
 65 70 75 80  
 Glu Pro Asp Leu Ser Ala Leu Asp Asp Phe Ser Arg Ile Lys Gln Leu  
 85 90 95  
 Val Gln Gln His Pro Asn Val Lys Thr Val Val Pro Met Gly Thr Gly  
 100 105 110  
 Ala Thr Phe Ile Asn Ser Gly Asn Thr Ile Asp Leu Thr Leu Ala Arg  
 115 120 125  
 Leu Arg Asp Leu Tyr Lys Lys Ala Ala Gln Gly Asp Thr Pro Glu Leu  
 130 135 140  
 Arg Gly Gln Ile His Ser Leu Gln Ala His Val Arg His Ile Ile Thr  
 145 150 155 160  
 Leu Leu Glu Glu Asp Met Lys Arg Arg Arg Glu Ile Ile Asp Asp Lys  
 165 170 175  
 Thr Thr Asp Pro Ala Asp Ala Glu Ala Met Ala Arg Ala Arg Ser Glu  
 180 185 190  
 Ala Phe Trp Ala Asp Phe Asp Glu Lys Pro Phe Asp Ser Leu Glu Phe  
 195 200 205  
 Leu Glu Asn Arg Ile Ala Pro Tyr Met Thr Asp Gly Asp Met Leu Ser  
 210 215 220  
 Leu Arg Tyr Val Gly Thr Asp Leu Val Asn Phe Gln Lys Thr Phe Asp  
 225 230 235 240  
 Arg Met Arg Ile Val Glu Gly Thr Pro Val Pro Pro Gly His Arg Gly  
 245 250 255  
 Met Met Leu Ser Lys Phe Thr Tyr Glu Asn Asp Phe Lys Leu Lys Thr  
 260 265 270  
 Ala His Arg Leu Asp Leu Ile Lys Glu Ala Arg Asp Thr Asn His Lys  
 275 280 285  
 Thr Ile Ala Met Asp Pro Gln Leu Gln Arg Trp Val Lys Glu Asn Gln  
 290 295 300  
 Thr Gln Thr Arg Glu Ile Leu Phe Gln Leu Asp Asp Leu Lys Thr Lys  
 305 310 315 320  
 Gln Ala Val Glu Arg Leu Gln Arg Val Leu Gly Ser Gln Glu Thr Asp  
 325 330 335  
 Leu Gly Lys Leu Leu Pro Ala Phe Phe Thr Met Asp Asp Ala Asn Phe  
 340 345 350  
 Asp Thr Arg Tyr Gln Gln Phe Tyr Ser Glu Leu Ala Thr Leu Leu Asp  
 355 360 365  
 Leu Tyr Arg Ile Arg Ile Gly Asp Asp Leu Thr Ile Thr Ala Phe Ser  
 370 375 380  
 Arg Thr Gly Tyr Val Gln Ser Val Asn Val Lys Ile Tyr Gly Thr Tyr  
 385 390 395 400  
 Gln Phe Asp Gly Leu Glu Lys Ser Ala Val Ala Gly Ala Leu Asn Leu  
 405 410 415  
 Leu Asp Leu Met Ser Phe Arg Glu Leu Tyr Gly Tyr Leu Thr Ala Glu  
 420 425 430  
 Lys Lys Ala Glu Leu Ala Gly Leu Gln Lys Ala Ser Gly Val Gln Gln  
 435 440 445  
 Val Lys Arg Glu Asp Ala Glu Thr Ala Leu Phe Gly Glu Gln Gly Ser  
 450 455 460  
 Ala Ser Leu Val Ala Glu Gly Thr Ala Gly Gln Ile Asp Glu Asp Lys

[0070]

```

465          470          475          480
Gln Leu Asp Gly Leu Ala Gln Lys Leu His Arg Glu Glu Leu Ala Ser
          485          490          495
Arg Val Tyr Thr Gln Gln Glu Ile Glu Ser Gly Val Val Leu Ser Thr
          500          505          510
Ala Val Leu Leu Lys His Pro Glu Lys Leu Glu Gln Thr Leu Ala Glu
          515          520          525
Leu Arg Lys Ser Ala Asp Asp Ala Lys Leu Pro Leu Arg Ile Ile Ser
          530          535          540
Trp Gln Lys Ala Ser Gly Thr Ile Gly Gln Phe Val Leu Val Ala Lys
545          550          555          560
Leu Val Leu Tyr Phe Ala Val Phe Ile Ile Phe Val Val Ala Leu Val
          565          570          575
Ile Ile Asn Asn Ala Met Met Met Ala Thr Leu Gln Arg Val Arg Glu
          580          585          590
Val Gly Thr Leu Arg Ala Ile Gly Ala Gln Arg Ser Phe Val Leu Ser
          595          600          605
Met Val Leu Val Glu Thr Val Val Leu Gly Leu Val Phe Gly Val Leu
          610          615          620
Gly Ala Ala Met Gly Gly Ala Ile Met Asn Met Leu Gly His Val Gly
625          630          635          640
Ile Pro Ala Gly Asn Glu Ala Leu Tyr Phe Phe Phe Ser Gly Pro Arg
          645          650          655
Leu Phe Pro Ser Leu His Leu Ser Asn Leu Val Ala Ala Phe Val Ile
          660          665          670
Val Leu Val Val Ser Ala Leu Ser Thr Phe Tyr Pro Ala Tyr Leu Ala
          675          680          685
Thr Arg Val Ser Pro Leu Gln Ala Met Gln Thr Asp Glu
          690          695          700

```

```

<210> 69
<211> 253
<212> PRT
<213> 深棕色孢囊杆菌

```

[0071]

```

<220>
<221> MISC_FEATURE
<222> (1)..(253)
<223> ORF10

```

```

<400> 69
Met Ser Gln Val Thr Ala Leu Pro Gly Ser Thr Gln Pro Ile Val Ser
1          5          10          15
Leu Thr Glu Val Thr Lys Thr Tyr Ser Leu Gly Lys Val Gln Val Pro
          20          25          30
Ala Leu Arg Gly Val Thr Leu Glu Val Tyr Pro Gly Glu Phe Ile Ser
          35          40          45
Ile Ala Gly Pro Ser Gly Ser Gly Lys Thr Thr Ala Leu Asn Leu Ile
          50          55          60
Gly Cys Val Asp Thr Ala Ser Ser Gly Val Val Ser Val Asp Gly Gln
65          70          75          80
Asp Thr Lys Lys Leu Thr Glu Arg Gln Leu Thr His Leu Arg Leu His
          85          90          95
Thr Ile Gly Phe Ile Phe Gln Ser Phe Asn Leu Val Ser Val Leu Ser
          100          105          110
Val Phe Gln Asn Val Glu Phe Pro Leu Leu Leu Gln Arg Lys Leu Asn
          115          120          125
Ala Ser Glu Arg Arg Thr Arg Val Met Thr Leu Leu Glu Gln Val Gly
          130          135          140
Leu Glu Lys His Ala Lys His Arg Pro Asn Glu Leu Ser Gly Gly Gln
145          150          155          160
Arg Gln Arg Val Ala Val Ala Arg Ala Leu Val Thr Arg Pro Lys Leu
          165          170          175
Val Leu Ala Asp Glu Pro Thr Ala Asn Leu Asp Ser Val Thr Gly Gln
          180          185          190
Asn Ile Ile Asp Leu Met Lys Glu Leu Asn Arg Lys Glu Gly Thr Thr
          195          200          205
Phe Ile Phe Ser Thr His Asp Ala Lys Val Met Thr His Ala Asn Ala
          210          215          220
Val Val Arg Leu Ala Asp Gly Lys Ile Leu Asp Arg Ile Thr Pro Ala
225          230          235          240

```

Glu Ala Gln Lys Val Met Ala Val Ser Glu Gly Gly His  
245 250

<210> 70  
<211> 397  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>  
<221> MISC\_FEATURE  
<222> (1).. (397)  
<223> ORF11

<400> 70  
Met Pro Gln Lys Phe Val Gly Lys Trp Lys Gly Gly Arg Val Lys Leu  
1 5 10 15  
Val Asp Gly Arg Lys Val Trp Leu Leu Glu Lys Met Val Ser Gly Ala  
20 25 30  
Arg Phe Ser Val Ser Leu Ala Val Ser Asn Glu Glu Asp Ala Leu Ala  
35 40 45  
Glu Leu Ala Leu Phe Arg Arg Asp Arg Asp Ala Tyr Leu Ala Lys Val  
50 55 60  
Lys Ala Asp Arg Ser Glu Glu Val Gln Ala Ser Thr Val Ala Gly Ala  
65 70 75 80  
Val Pro Leu Ser Gly Asp Val Gly Pro Arg Leu Asp Ala Asp Ser Val  
85 90 95  
Arg Glu Phe Leu Arg His Leu Thr Gln Arg Gly Arg Thr Glu Gly Tyr  
100 105 110  
Arg Arg Asp Ala Arg Thr Tyr Leu Ser Gln Trp Ala Glu Val Leu Ala  
115 120 125  
Gly Arg Asp Leu Ser Thr Val Ser Leu Leu Glu Leu Arg Arg Ala Leu  
130 135 140  
Ser Gln Trp Pro Thr Ala Arg Lys Met Arg Ile Ile Thr Leu Lys Ser  
145 150 155 160  
Phe Phe Ser Trp Leu Arg Glu Glu Asp Arg Leu Lys Ala Ala Glu Asp  
165 170 175  
Pro Thr Leu Ser Leu Lys Val Pro Pro Ala Val Ala Glu Lys Gly Arg  
180 185 190  
Arg Ala Lys Gly Tyr Ser Met Ala Gln Val Glu Lys Leu Tyr Ala Ala  
195 200 205  
Ile Gly Ser Gln Thr Val Arg Asp Val Leu Cys Leu Arg Ala Lys Thr  
210 215 220  
Gly Met His Asp Ser Glu Ile Ala Arg Leu Ala Ser Gly Lys Gly Glu  
225 230 235 240  
Leu Arg Val Val Asn Asp Pro Ser Gly Ile Ala Gly Thr Ala Arg Phe  
245 250 255  
Leu His Lys Asn Gly Arg Val His Ile Leu Ser Leu Asp Ala Gln Ala  
260 265 270  
Leu Ala Ala Ala Gln Arg Leu Gln Val Arg Gly Arg Ala Pro Ile Arg  
275 280 285  
Asn Thr Val Arg Glu Ser Ile Gly Tyr Ala Ser Ala Arg Ile Gly Gln  
290 295 300  
Ser Pro Ile His Pro Ser Glu Leu Arg His Ser Phe Thr Thr Trp Ala  
305 310 315 320  
Thr Asn Glu Gly Gln Val Val Arg Ala Thr Arg Gly Gly Val Pro Leu  
325 330 335  
Asp Val Val Ala Ser Val Leu Gly His Gln Ser Thr Arg Ala Thr Lys  
340 345 350  
Lys Phe Tyr Asp Gly Thr Glu Ile Pro Pro Met Ile Thr Val Pro Leu  
355 360 365  
Lys Leu His His Pro Gln Asp Pro Ala Val Met Gln Leu Arg Arg Asn  
370 375 380  
Cys Ser Pro Asp Pro Val Thr Arg Glu Ala Glu Ala  
385 390 395

[0072]

<210> 71  
<211> 124  
<212> PRT  
<213> 深棕色孢囊杆菌

<220>

<221> MISC\_FEATURE  
 <222> (1)..(124)  
 <223> ORF12

<400> 71  
 Val Leu Leu Ala Phe Pro Ser Gly Leu Leu Ser Leu Ala Leu Leu Ser  
 1 5 10 15  
 Thr Thr Thr Glu Ile Ser Ala Ala Leu Pro Val Asp Glu Cys Glu Ser  
 20 25 30  
 Ala Ser Leu Arg Ile Glu Leu Pro Ala Thr Pro Gly Gly Lys Pro Pro  
 35 40 45  
 Val Val Cys Leu Gly Pro Gly Leu Pro Ile His Phe Arg Phe Asp Ser  
 50 55 60  
 Ala Leu Gln Gln Lys Ser Leu Arg Ile Gln Asp Arg Gly Trp Phe Glu  
 65 70 75 80  
 Asp Trp Ala Leu Gly Gln Gln Thr Leu Val Leu Thr Pro His Asp Asn  
 85 90 95  
 Leu Val Ala Gly Lys Arg Ser Glu Val Glu Val Cys Phe Ala Asp Gly  
 100 105 110  
 Ala Ala Pro Ala Cys Ala Ser Phe Val Leu Arg Arg  
 115 120

<210> 72  
 <211> 112  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1)..(112)  
 <223> ORF13

<400> 72  
 Met His Thr Lys Val Pro Ser Val Phe Glu Ala Thr Pro Glu Ser Leu  
 1 5 10 15  
 Ser Asp Val Asp Tyr Gln Phe Trp His Glu Asp Phe Pro Arg Val Phe  
 20 25 30  
 Glu Arg Gln His Ile Asp Ala His Ala Val Pro Ala Ile Gly Ala Tyr  
 35 40 45  
 Leu Gly Glu Val Leu Val Arg Asn Leu Gly Gly Lys Trp Ile Pro Arg  
 50 55 60  
 Gln Lys Leu Asp Glu Ala Gln Val Leu Val Gly Asn Arg Val Trp Leu  
 65 70 75 80  
 Pro Phe Ala Arg Ala His His Tyr Met Arg Ser Cys Glu Ser Leu Leu  
 85 90 95  
 Asp Tyr Ser Leu Thr Gln Leu Tyr Arg Val Ala Glu Arg Tyr Arg Gly  
 100 105 110

<210> 73  
 <211> 304  
 <212> PRT  
 <213> 深棕色孢囊杆菌

<220>  
 <221> MISC\_FEATURE  
 <222> (1)..(304)  
 <223> ORF 14

<400> 73  
 Met Lys Val Leu Gly Leu Gly Asp Val Lys Ser Glu Asp Ser Leu Arg  
 1 5 10 15  
 Leu Thr Phe Glu Gly Ala Leu Asp Pro Gln Ala Ala Leu Glu Lys Val  
 20 25 30  
 Leu Glu Pro Phe Phe Gln Ala Leu Glu Glu Tyr Ala Gly Asp Trp Met  
 35 40 45  
 Pro Glu Val Val Ser Gly Arg Arg Arg Leu Lys Tyr Ser Arg Ala Asn  
 50 55 60  
 Ile Trp Lys Ala Leu Glu Arg Arg Asp Glu Arg Ser Thr Asp Thr  
 65 70 75 80  
 Trp Leu Tyr Arg Thr Gln Arg Pro Thr Leu Glu Met Ser Leu His Leu  
 85 90 95

[0073]

[0074]

Trp	Phe	Pro	Pro	Leu	Pro	Pro	Ala	Leu	Asp	Val	Met	Thr	Thr	Val	Gln
			100					105					110		
Pro	Leu	Thr	Arg	Phe	Ala	Glu	Lys	Glu	Arg	Cys	Arg	Gln	Phe	Val	Glu
		115					120					125			
Met	Val	Arg	Thr	Trp	Ala	Ser	Cys	Tyr	Pro	Val	Thr	His	Ala	Ala	Ala
		130				135					140				
His	Ser	Val	Ala	Asp	Arg	Ala	Leu	Ala	Gly	Ala	Pro	Asp	Phe	Gly	Arg
145				150					155						160
Asp	Ala	Arg	Thr	Ala	Arg	Arg	Asp	Gly	Phe	Asp	Arg	Ile	Tyr	Glu	Ile
			165					170						175	
Phe	Trp	Leu	Asn	Val	Phe	Gly	Pro	Lys	Leu	Val	Glu	Ala	Val	Gly	Arg
		180					185					190			
Glu	Arg	Met	Leu	Ser	Thr	Pro	Ala	His	Arg	Val	Glu	Glu	Leu	Pro	Asn
		195					200					205			
Gly	Ser	Ile	Leu	Leu	Val	Thr	Trp	Pro	Thr	Ala	Ala	Asp	Phe	Ala	Gly
		210				215				220					
Ala	Glu	Ala	Arg	His	Ala	Gln	Ala	Arg	Ala	His	Val	His	Leu	Arg	Pro
225				230					235						240
Asp	Leu	Arg	Phe	Asp	Thr	Val	Leu	Arg	Thr	Leu	His	Glu	Arg	Ser	Ala
			245					250					255		
Ala	Leu	Ala	Pro	Val	Glu	Pro	Cys	Phe	His	Pro	Asp	Val	Ala	Pro	Leu
			260					265					270		
Leu	Ser	His	Val	Val	Asp	Ser	Val	Ala	Ile	Arg	Met	Trp	Lys	Thr	Trp
		275					280					285			
Ser	Ala	Leu	Thr	Ser	Ile	Thr	Glu	Leu	Trp	Leu	Ser	Thr	Ser	Trp	Arg
		290				295					300				

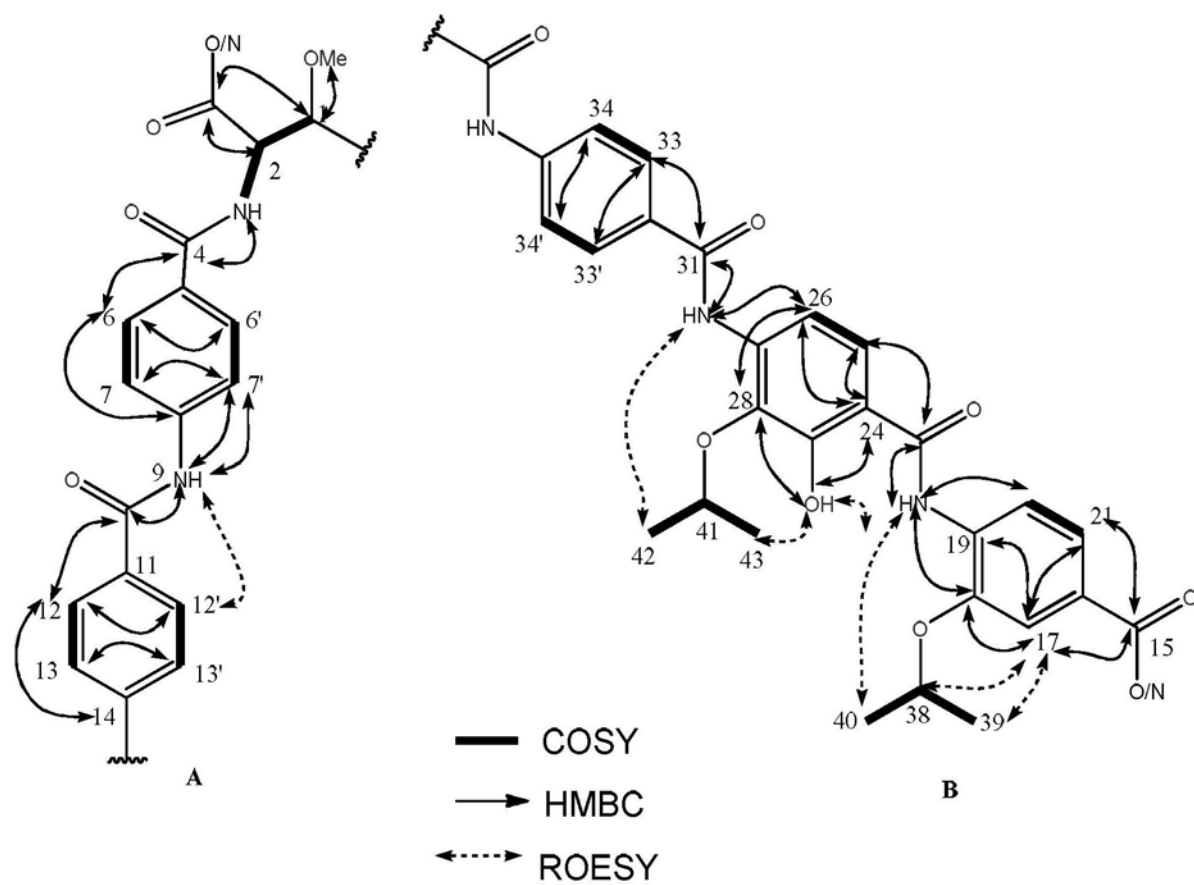


图1

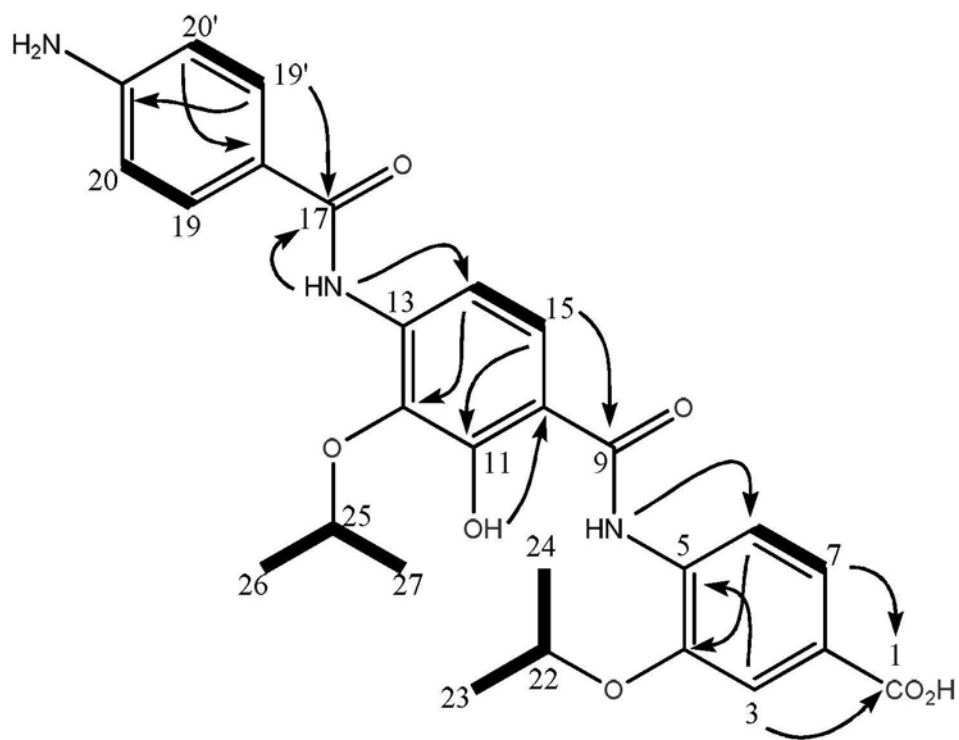


图2

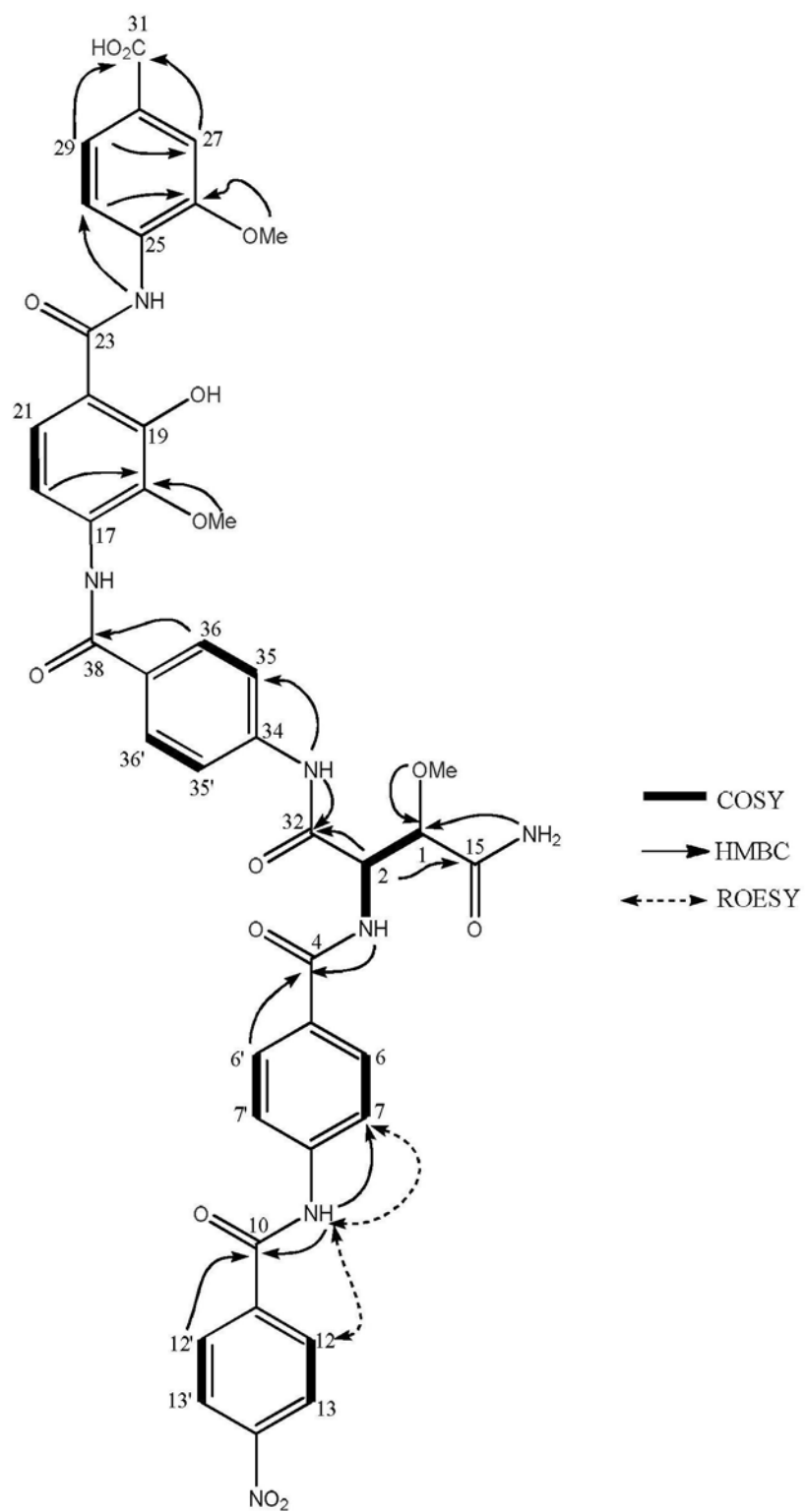


图3



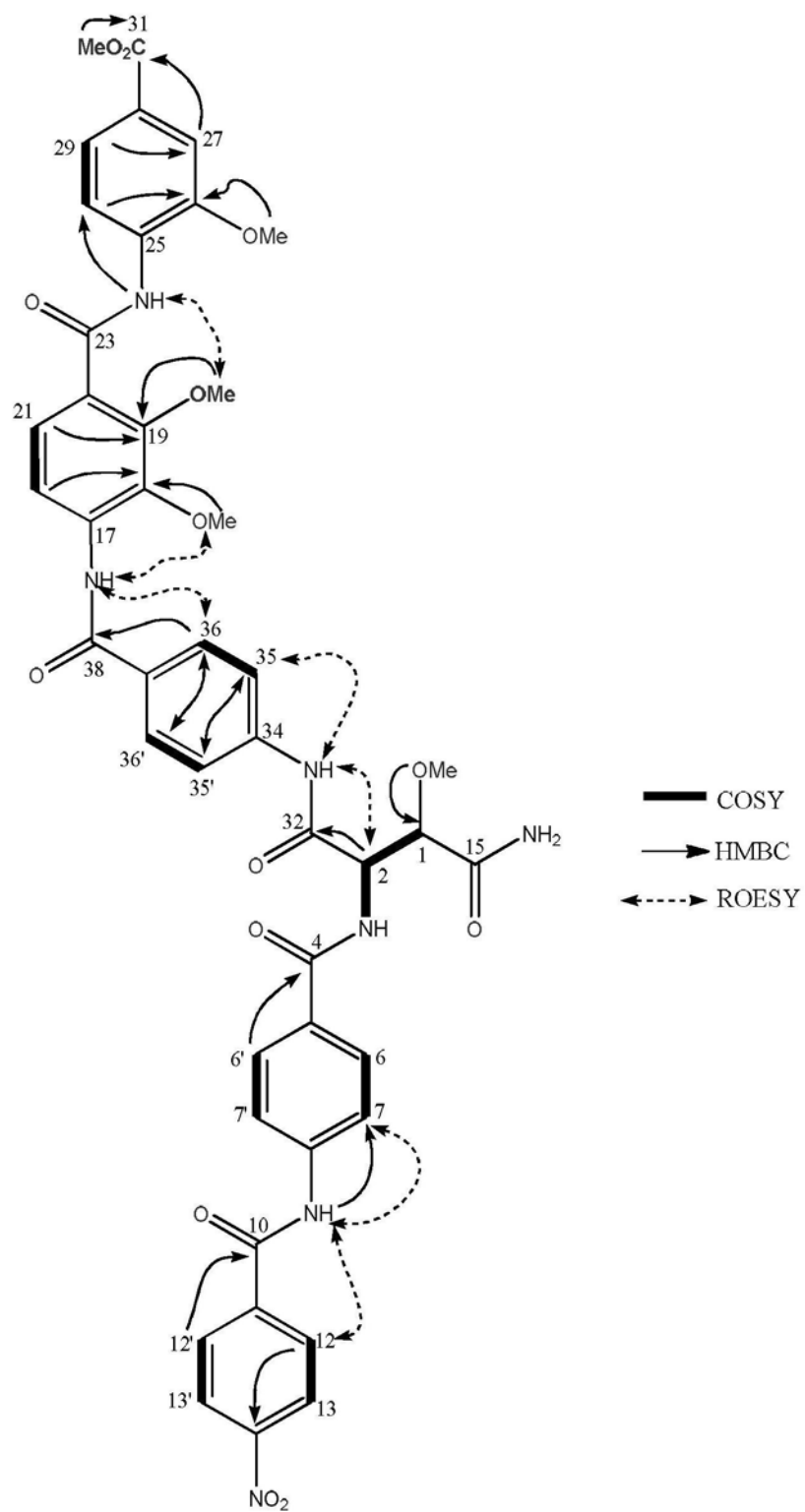


图4

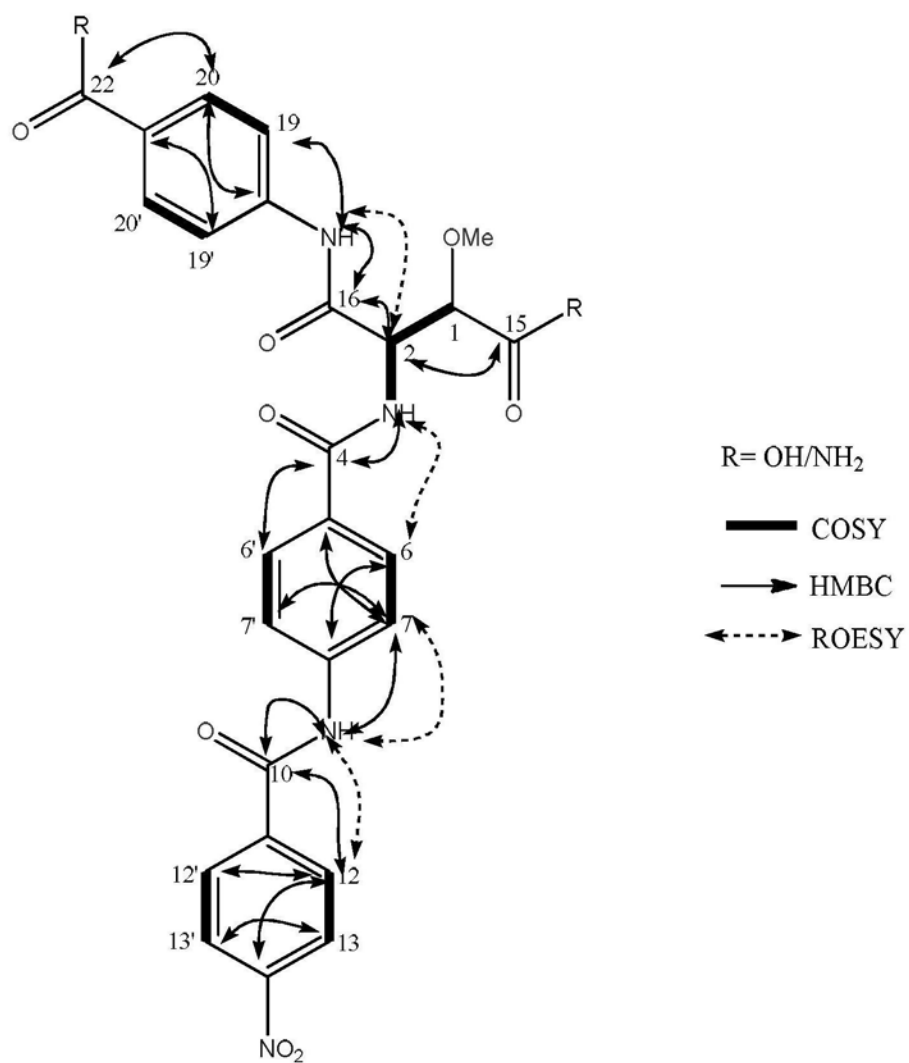


图5

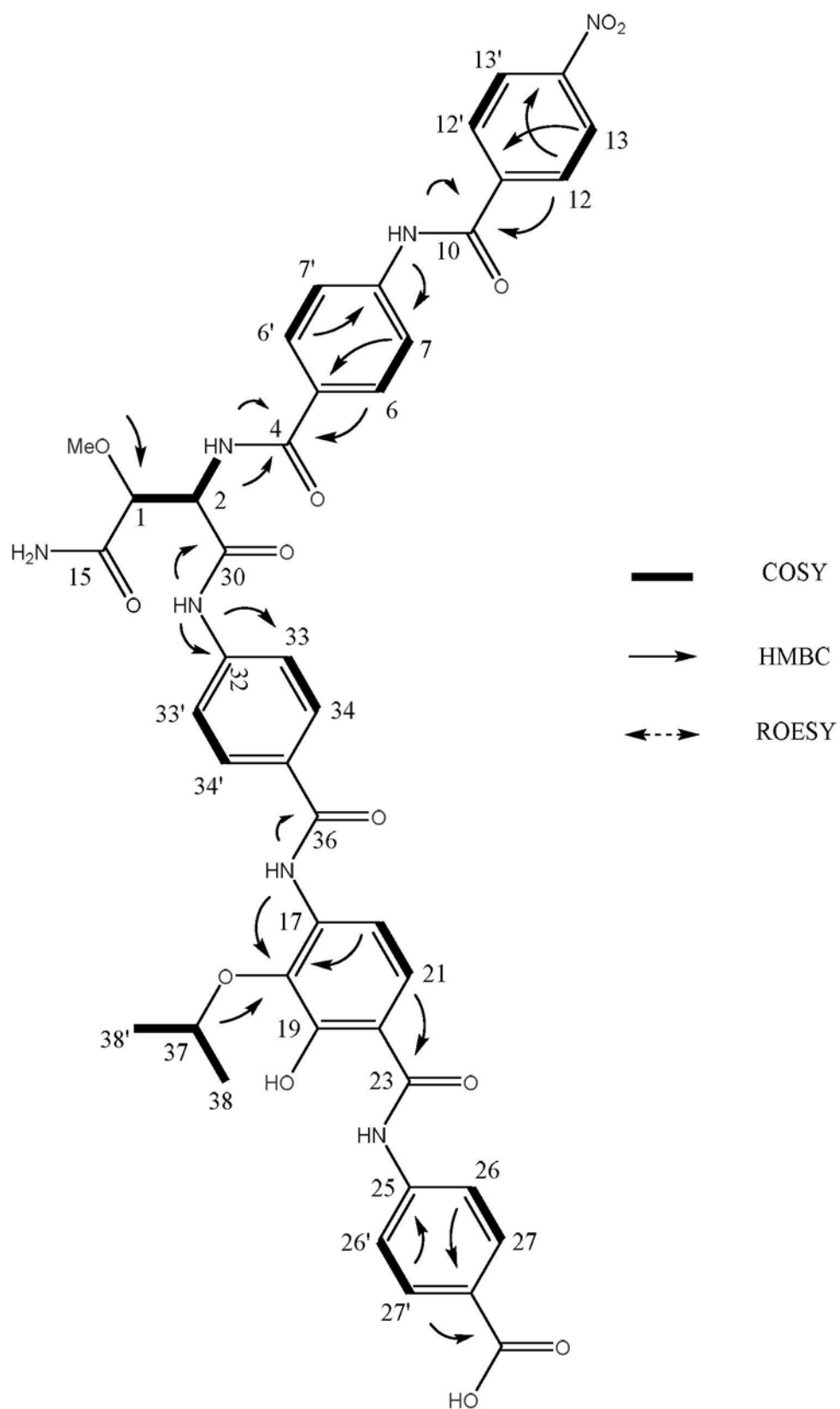


图6

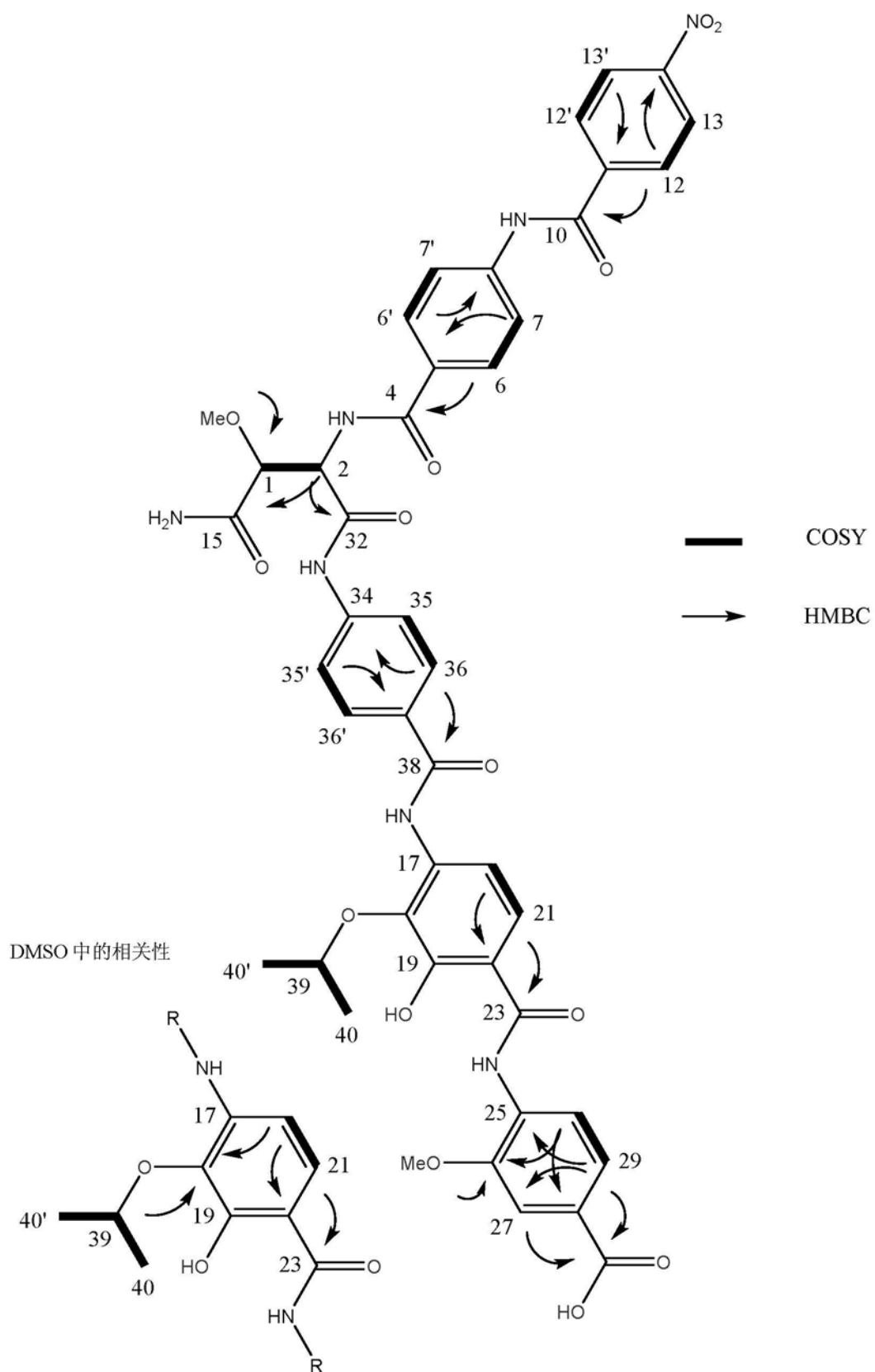


图7

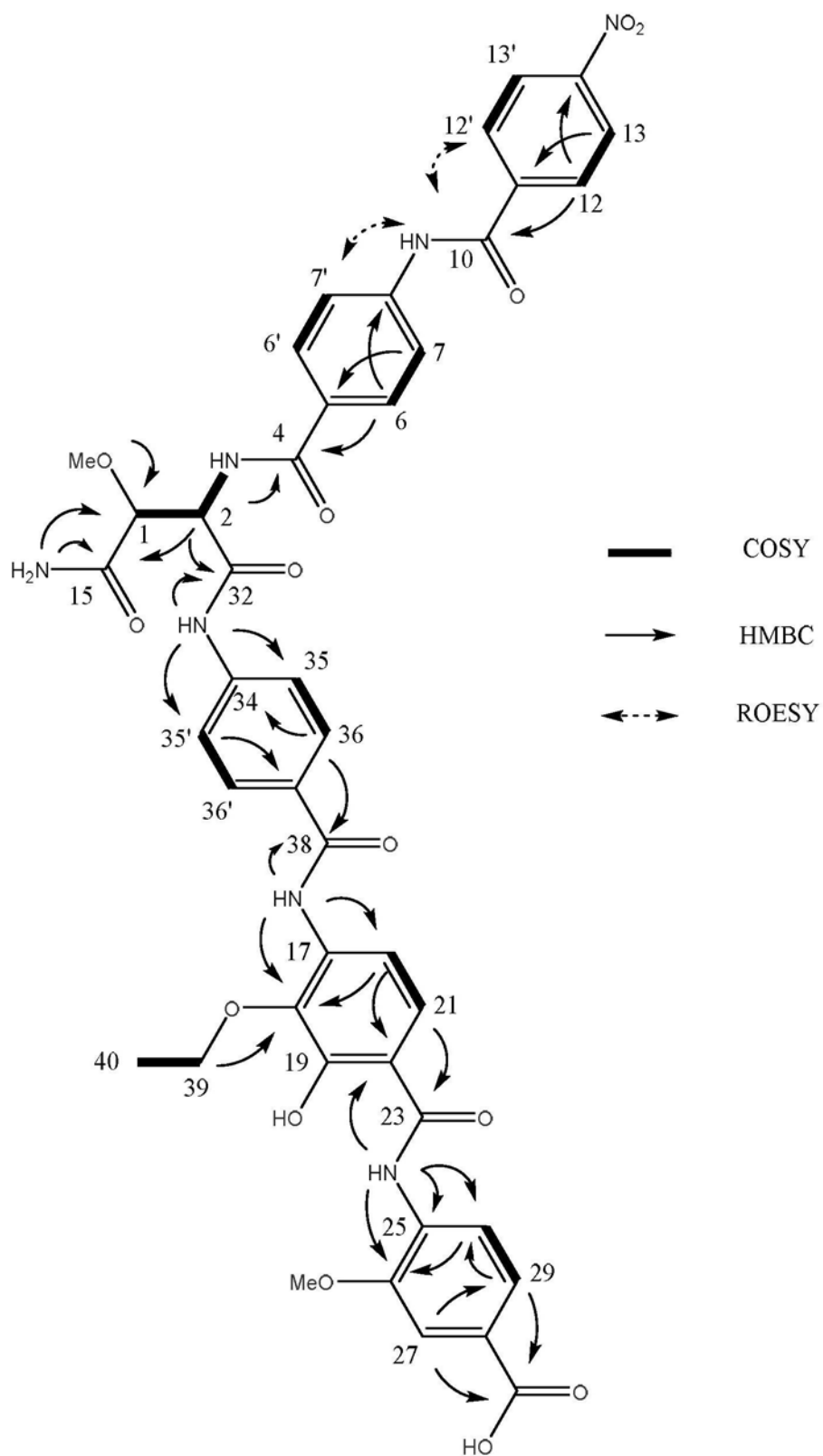


图8

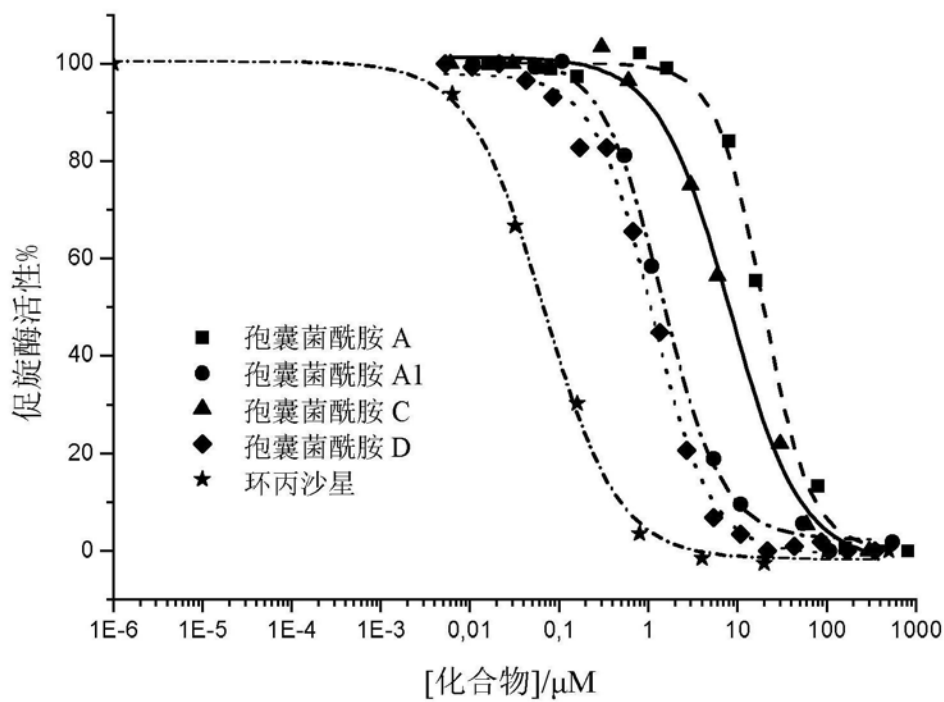


图9a

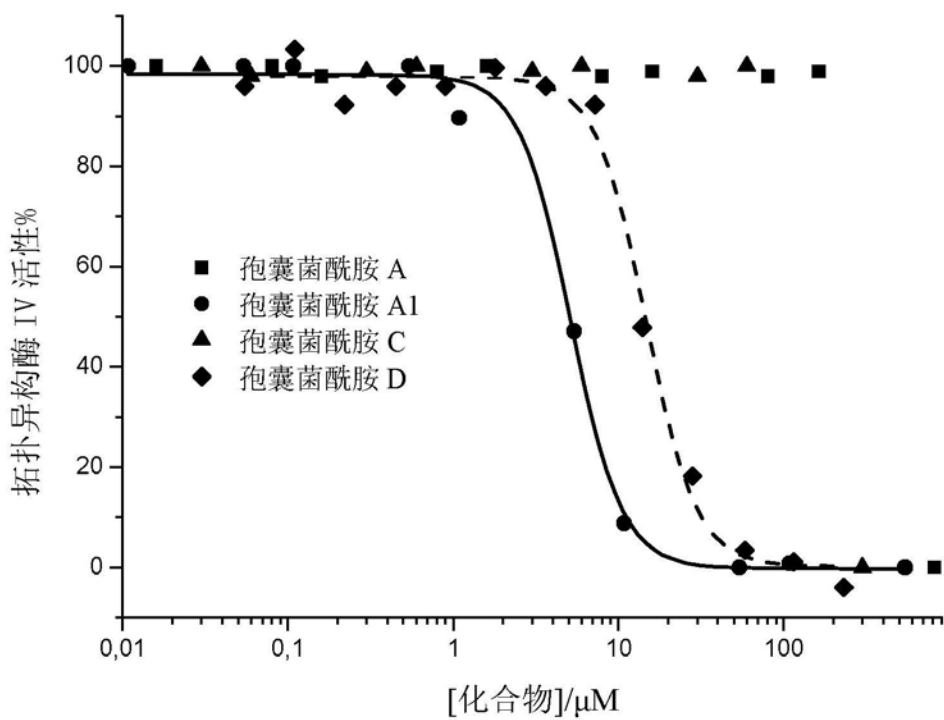


图9b

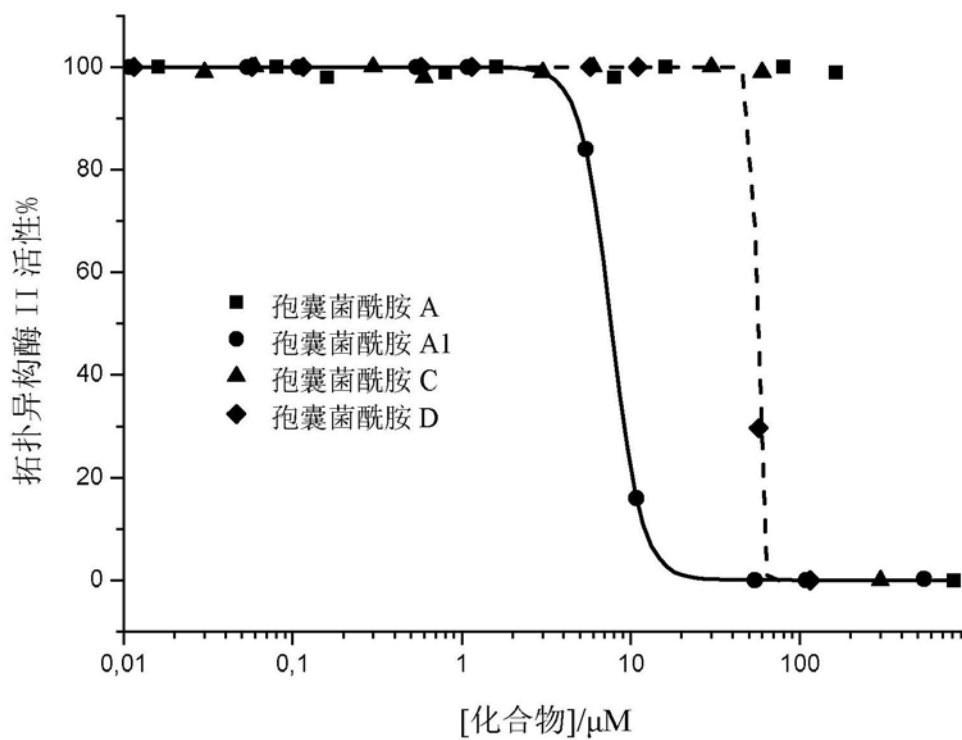


图9c

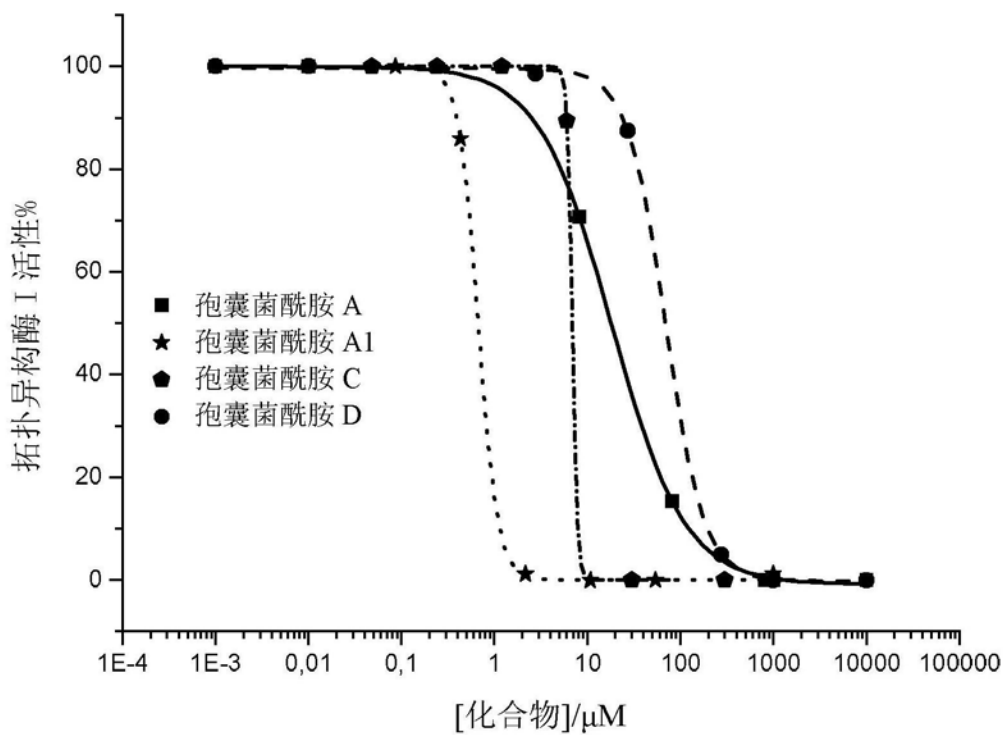


图9d

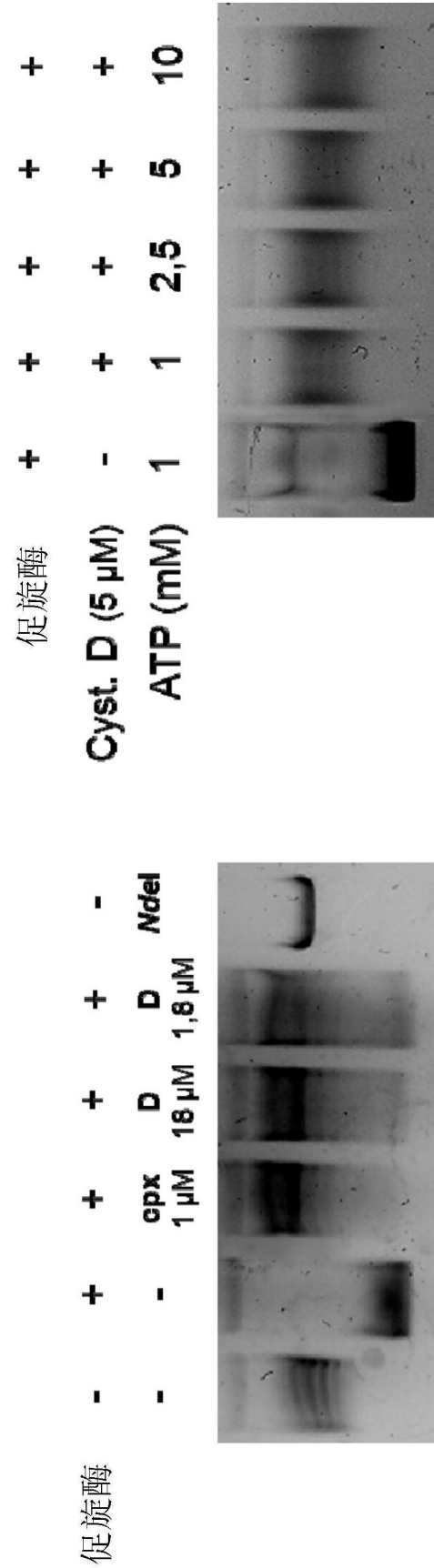


图10a和b



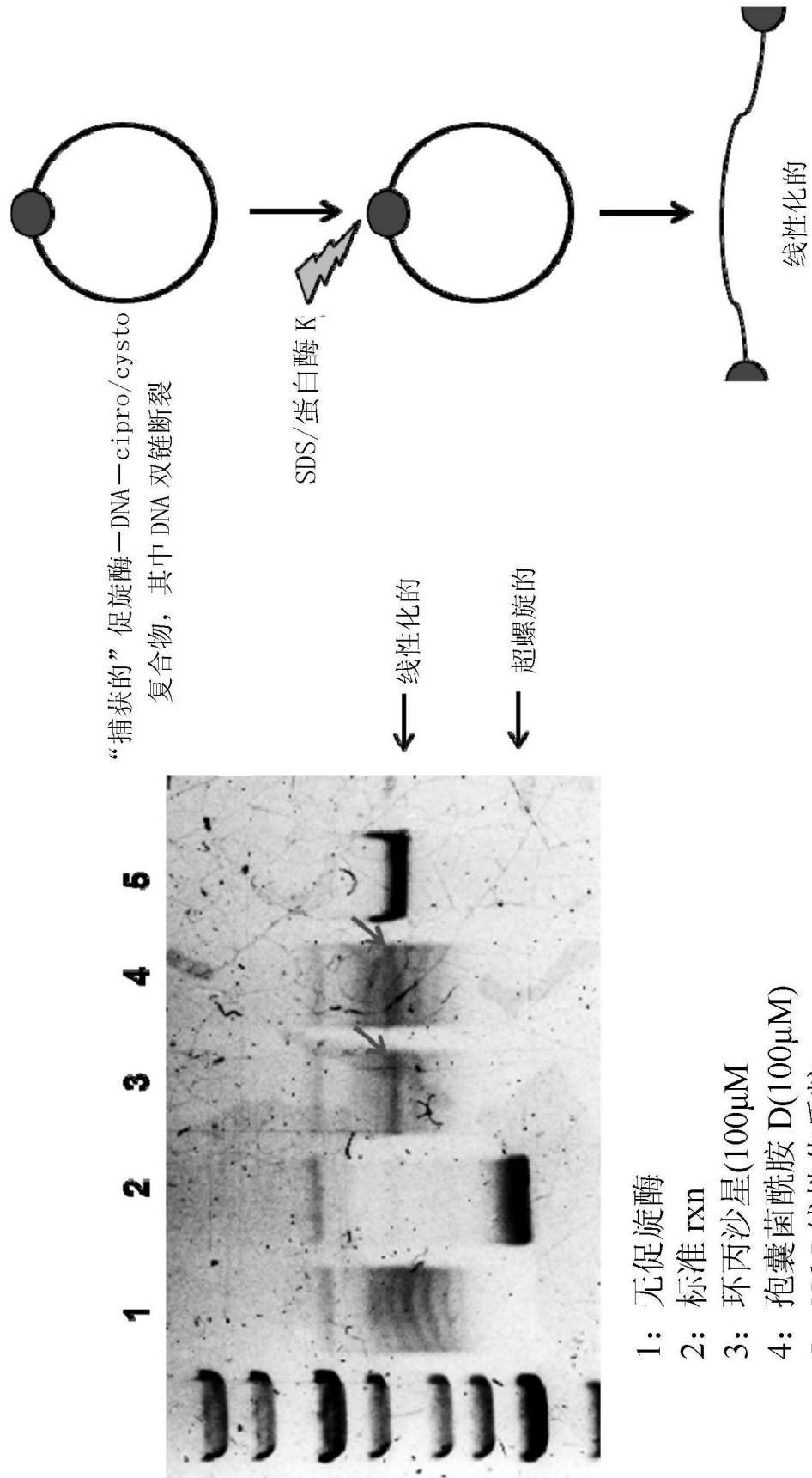


图11

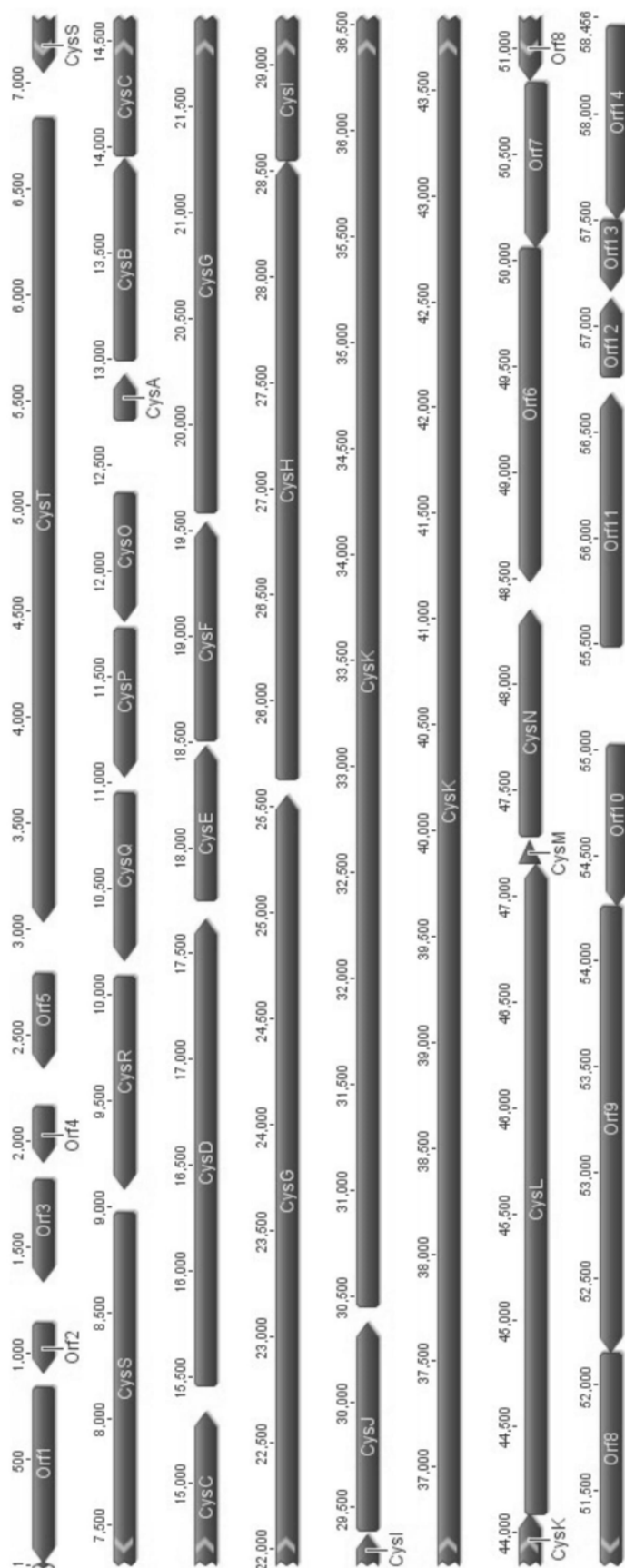


图12