

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第5968784号  
(P5968784)

(45) 発行日 平成28年8月10日 (2016. 8. 10)

(24) 登録日 平成28年7月15日 (2016. 7. 15)

|                                 |                    |
|---------------------------------|--------------------|
| (51) Int. Cl.                   | F I                |
| <b>C 1 2 N</b> 1/21 (2006. 01)  | C 1 2 N 1/21 Z N A |
| <b>C 1 2 P</b> 21/02 (2006. 01) | C 1 2 P 21/02      |
| C 1 2 N 15/09 (2006. 01)        | C 1 2 N 15/00 A    |
| C 1 2 N 9/10 (2006. 01)         | C 1 2 N 9/10       |
| <b>C O 8 B</b> 37/00 (2006. 01) | C O 8 B 37/00 P    |
| 請求項の数 23 (全 129 頁) 最終頁に続く       |                    |

|               |                               |           |                          |
|---------------|-------------------------------|-----------|--------------------------|
| (21) 出願番号     | 特願2012-539870 (P2012-539870)  | (73) 特許権者 | 510225155                |
| (86) (22) 出願日 | 平成22年11月16日 (2010. 11. 16)    |           | グリコヴァキシシ アーゲー            |
| (65) 公表番号     | 特表2013-511520 (P2013-511520A) |           | スイス国, シーエイチー 8 9 5 2 シュリ |
| (43) 公表日      | 平成25年4月4日 (2013. 4. 4)        |           | ーレン, グラーベンシュトラーセ 3       |
| (86) 国際出願番号   | PCT/US2010/002980             | (74) 代理人  | 100091096                |
| (87) 国際公開番号   | W02011/062615                 |           | 弁理士 平木 祐輔                |
| (87) 国際公開日    | 平成23年5月26日 (2011. 5. 26)      | (74) 代理人  | 100118773                |
| 審査請求日         | 平成25年11月15日 (2013. 11. 15)    |           | 弁理士 藤田 節                 |
| (31) 優先権主張番号  | 61/272, 931                   | (74) 代理人  | 100122389                |
| (32) 優先日      | 平成21年11月19日 (2009. 11. 19)    |           | 弁理士 新井 栄一                |
| (33) 優先権主張国   | 米国 (US)                       | (74) 代理人  | 100111741                |
|               |                               |           | 弁理士 田中 夏夫                |
|               |                               | (74) 代理人  | 100169971                |
|               |                               |           | 弁理士 菊田 尚子                |
|               |                               | 最終頁に続く    |                          |

(54) 【発明の名称】 原核細胞における免疫原性多糖体生成の生合成システム

(57) 【特許請求の範囲】

【請求項 1】

組換え原核細胞生合成システムであって：

- (i) ウンデカプレニルピロリン酸上でN-アセチルガラクトサミンを合成するエピメラーゼをコードする核酸であり、配列番号:1と少なくとも90%同一である、前記核酸；
- (ii) 脂質キャリア上でオリゴ糖又は多糖を構築することができるグリコシルトランスフェラーゼをコードする少なくとも1つの核酸；
- (iii) カンピロバクター・ジェジュニに由来するオリゴサッカリルトランスフェラーゼをコードする核酸；及び
- (iv) 1つ以上の導入されたコンセンサス配列D/E-X-N-Z-S/T（式中、X及びZは、プロリン以外の任意の天然のアミノ酸であり得る）を含むタンパク質をコードする核酸；  
を含み、前記グリコシルトランスフェラーゼにより構築されたオリゴ糖は、前記オリゴサッカリルトランスフェラーゼにより脂質キャリアからタンパク質へ転移され得る、前記組換え原核細胞生合成システム。

【請求項 2】

前記エピメラーゼをコードしている核酸が、配列番号:1からなる、請求項 1 記載の組換え原核細胞生合成システム。

【請求項 3】

前記タンパク質が、緑膿菌エキソタンパク質である、請求項 1 記載の組換え原核細胞生合成システム。

## 【請求項 4】

前記オリゴ糖又は多糖が、グラム陰性菌に由来する、請求項 1 記載の組換え原核細胞生合成システム。

## 【請求項 5】

前記オリゴ糖又は多糖が、シゲラ・フレックスネリに由来する、請求項 4 記載の組換え原核細胞生合成システム。

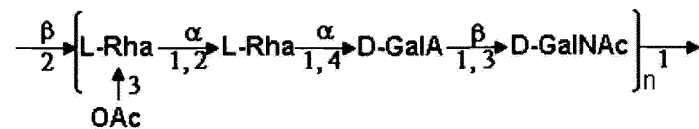
## 【請求項 6】

前記オリゴ糖又は多糖が、シゲラ・フレックスネリ 6 に由来する、請求項 5 記載の組換え原核細胞生合成システム。

## 【請求項 7】

前記オリゴ糖又は多糖が、下記構造を含む、請求項 1 記載の組換え原核細胞生合成システム：

## 【化 1】



。

## 【請求項 8】

前記オリゴ糖又は多糖が、大腸菌に由来する、請求項 4 記載の組換え原核細胞生合成システム。

## 【請求項 9】

前記オリゴ糖又は多糖が、大腸菌 0157 に由来する、請求項 8 記載の組換え原核細胞生合成システム。

## 【請求項 10】

前記オリゴ糖又は多糖が、構造  $\alpha$ -D-PerNAc- $\alpha$ -L-Fuc- $\beta$ -D-Glc- $\alpha$ -D-GalNAc を含む、請求項 1 記載の組換え原核細胞生合成システム。

## 【請求項 11】

N 結合型グリコシル化タンパク質を産生することができる宿主細胞を製造する方法であって：

a.) 下記を、宿主生物へ導入する工程：

i.) ウンデカプレニルピロリン酸上で N-アセチルガラクトサミンを合成するエピメラーゼをコードする核酸であり、配列番号:1 と少なくとも 90% 同一である、前記核酸；

ii.) 脂質キャリア上でオリゴ糖又は多糖を構築することができるグリコシルトランスフェラーゼをコードする少なくとも 1 つの核酸；

iii.) カンピロバクター・ジェジュニに由来するオリゴサッカリルトランスフェラーゼをコードする核酸；及び

iv.) 1 つ以上の導入されたコンセンサス配列 D/E-X-N-Z-S/T (式中、X 及び Z は、プロリン以外の任意の天然のアミノ酸であり得る) を含むタンパク質をコードする核酸；

を含む、前記方法。

## 【請求項 12】

前記エピメラーゼをコードしている核酸が、配列番号:1 からなる、請求項 11 記載の方法。

## 【請求項 13】

前記タンパク質が、緑膿菌エキソタンパク質である、請求項 11 記載の方法。

## 【請求項 14】

前記オリゴ糖又は多糖が、グラム陰性菌に由来する、請求項 11 記載の方法。

## 【請求項 15】

前記オリゴ糖又は多糖が、シゲラ・フレックスネリに由来する、請求項 14 記載の方法

10

20

30

40

50

。

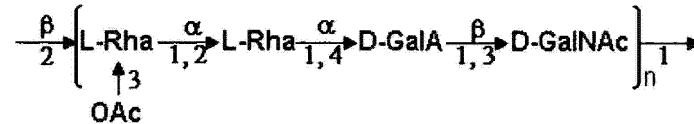
## 【請求項 1 6】

前記オリゴ糖又は多糖が、シゲラ・フレックスネリ6に由来する、請求項 1 5 記載の方法。

## 【請求項 1 7】

前記オリゴ糖又は多糖が、下記構造を有する、請求項 1 1 記載の方法：

## 【化 2】



10

。

## 【請求項 1 8】

前記オリゴ糖又は多糖が、大腸菌に由来する、請求項 1 4 記載の方法。

## 【請求項 1 9】

前記オリゴ糖又は多糖が、大腸菌0157に由来する、請求項 1 8 記載の方法。

## 【請求項 2 0】

前記オリゴ糖又は多糖が、構造  $\alpha$ -D-PerNAc- $\alpha$ -L-Fuc- $\beta$ -D-Glc- $\alpha$ -D-GalNAcを含む、請求項 1 1 記載の方法。

20

## 【請求項 2 1】

前記エピメラーゼをコードする核酸が、配列番号:1と少なくとも95%同一である、請求項 1 記載の組換え原核細胞生合成システム。

## 【請求項 2 2】

前記エピメラーゼをコードする核酸が、配列番号:1と少なくとも95%同一である、請求項 1 1 記載の方法。

## 【請求項 2 3】

請求項 1 1 記載の方法によって製造された宿主生物を、少なくとも1種のN-グリコシル化タンパク質が生成されるまで培養することを含む、N結合型グリコシル化タンパク質を生成する方法。

30

## 【発明の詳細な説明】

## 【技術分野】

## 【0 0 0 1】

(関連出願の相互参照)

本出願は、2009年11月19日に出版され、その全体が引用により本明細書中に組み込まれている、米国特許仮出願第61/272,931号の米国特許法第119条(e)の下での利益を主張するものである。

## 【0 0 0 2】

(発明の分野)

本発明は、ワクチン製造のための生合成システム及びタンパク質の使用に関する。加えて本発明は、還元末端に特定された単糖を持つオリゴ糖又は多糖の合成を開始するエピメラーゼを有する、組換え原核細胞生合成システムに関する。本発明は更に、発現システムにおいてグリカンで生成されたN-グリコシル化タンパク質、及び免疫原性グリカンを含む該N-グリコシル化タンパク質から製造されたバイオコンジュゲートワクチンに関し、且つN-グリコシル化タンパク質の製造方法を提供する。

40

## 【背景技術】

## 【0 0 0 3】

(発明の背景)

糖タンパク質は、1個以上の共有結合された糖ポリマーを有するタンパク質である。N-結合型タンパク質のグリコシル化は、真核生物の小胞体において生じる本質的且つ保存さ

50

れたプロセスである。それは、タンパク質フォールディング、オリゴマー化、安定性、品質管理、選別、並びに分泌タンパク質及び膜タンパク質の輸送にとって重要である(Helenius, A.及びAebi, M.の文献、(2004)、「小胞体におけるN-結合型グリカンの役割(Roles of N-linked glycans in the endoplasmic reticulum)」、Annu. Rev. Biochem. 73, 1019-1049)。

#### 【0004】

タンパク質のグリコシル化は、タンパク質の免疫原性、安定性及び半減期への重大な影響を有する。加えてグリコシル化は、例えばタンパク質のグリコシル化された部分と相互作用する固相に結合されたレクチンリガンドによるアフィニティクロマトグラフィーなどの、クロマトグラフィーによるタンパク質の精製を補助することができる。従って、生物学的及び薬学的に有用なグリコシル化パターンを提供するために、真核細胞において組換えにより多くのグリコシル化タンパク質を生成する手法が確立されている。

10

#### 【0005】

WO 2003/07467(Aebiら)は、細菌である食物媒介性病原体カンピロバクター・ジェジュニ(*Campylobacter jejuni*)が、そのタンパク質をN-グリコシル化することができ、このことはある種の古細菌を除いた公知の原核生物の中で独自の特徴であることを明らかにした。このグリコシル化に必要とされる機序は、いわゆるpgl遺伝子座にクラスター化されている12の遺伝子によりコードされている。N-グリコシル化の破壊は、C.ジェジュニの侵入及び病原性に影響を及ぼすが、しかし大抵の真核生物においては致命的ではない(Burda P.及びM. Aebiの文献、(1999)、「N-結合型グリコシル化のドリコール経路(The dolichol pathway of N-linked glycosylation)」、Biochem Biophys Acta, 1426(2):239-57)。大腸菌(*E. coli*)においてpgl遺伝子座及び受容体糖タンパク質を組換えにより同時に発現させることにより、C.ジェジュニタンパク質のN-グリコシル化を再構成することが可能である(Wackerらの文献、(2002)、「カンピロバクター・ジェジュニにおけるN-結合型グリコシル化及び大腸菌へのその機能転移(N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*)」、Science, 298, 1790-1793)。

20

#### 【0006】

N-グリカン類は、タンパク質内のコンセンサス配列に結合したグリカンをも有する。タンパク質内の公知のN-グリコシル化コンセンサス配列は、原核生物において組換え標的タンパク質をN-グリコシル化することを可能にする。そのような生物は、グリカンをタンパク質のコンセンサス配列に転移する酵素であるオリゴサッカリルトランスフェラーゼ("OT"; "OTase")、例えばC.ジェジュニのオリゴサッカリルトランスフェラーゼなどを含む。

30

#### 【0007】

WO 2003/07467(Aebiら)は、(i)脂質キャリア上のオリゴ糖の構築のための特異的グリコシルトランスフェラーゼ、(ii)コンセンサス配列"N-X-S/T"を含む組換え標的タンパク質(式中、Xはプロリン以外の任意のアミノ酸である)、及び(iii)例えば、該オリゴ糖を標的タンパク質のコンセンサス配列に共有結合する、C.ジェジュニのオリゴサッカリルトランスフェラーゼなどのオリゴサッカリルトランスフェラーゼ：をコードしている核酸が導入された原核生物を教示している。前記原核生物は、特定のグリコシルトランスフェラーゼの種類により規定される特定の構造を持つN-グリカンを生産する。

40

#### 【0008】

WO 2006/119987(Aebiら)は、タンパク質に加え、インビボで原核生物における効率的N-グリコシル化によりタンパク質を生産する手段及び方法を説明している。これは更に、該タンパク質の免疫原性、安定性、生物学的、予防的及び/又は治療的活性を改変するための、N-グリカンの組換えタンパク質への効率的導入、並びにその表面上に発明の組換えN-グリコシル化タンパク質を効率的にディスプレイする宿主細胞の条件を説明している。加えてこれは、以下のN-グリコシル化された最適化アミノ酸配列を1つ以上含む組換えN-グリコシル化タンパク質を説明しており：

D/E-X-N-Z-S/T (最適化コンセンサス配列)、  
(式中、X及びZは、Pro以外の任意の天然アミノ酸であり得る)、且つ該N-グリコシル化さ

50

れた部分的アミノ酸配列の少なくとも1つが導入されている。この特定の部分的アミノ酸配列(最適化コンセンサス配列)のタンパク質への導入は、これらの導入された位置でオリゴサッカリルトランスフェラーゼにより効率的にN-グリコシル化されたタンパク質をもたらす。

#### 【0009】

様々な多糖体の生合成が、細菌細胞において保存されている。多糖体は、規定された特異性を持つ様々なグリコシルトランスフェラーゼにより、細胞質膜で、共通の前駆体(活性化された糖ヌクレオチド)からキャリア脂質上に構築される。リポ多糖("LPS")は、例えば赤痢菌種、シュードモナス種、及び大腸菌(ExPEC、EHEC)などのグラム陰性菌においてのみ提供される。

10

#### 【0010】

LPSの合成は、膜の細胞質側で、単糖がキャリア脂質ウンデカプレニルリン酸("Und-P-P")へ付加されることから始まる。抗原は、様々なグリコシルトランスフェラーゼにより活性化された糖ヌクレオチドから単糖が逐次付加されることにより作製され、その脂質-連結された多糖は、フリッパーゼにより膜を通じて反転される。この抗原-反復単位は、酵素反応により重合される。次にこの多糖は、LPSを形成するリガーゼWaaLによりリポドAへ転移され、LPSがその表面に輸送されるのに対し、莢膜多糖は、重合及び表面への輸送後、キャリア脂質から放出される。これらの多糖の生合成経路は、インビボにおけるLPSバイオコンジュゲートの生成、タンパク質キャリアへのペリプラズム内の多糖の捕獲を可能にする。

20

#### 【0011】

そのように合成されたオリゴ糖又は多糖(すなわち、糖残基)及びタンパク質(すなわち、タンパク質キャリア)の複合体は、多くの細菌感染を防御するためのコンジュゲートワクチンとして使用することができる。コンジュゲートワクチンは、細菌感染を防御するためにうまく使用されている。多糖はT-細胞非依存性免疫原であるので、タンパク質キャリアに対する抗原性多糖の複合が、防御的記憶反応に必要とされる。多糖に加えタンパク質キャリアにおける反応基の活性化を用いる、様々な化学的方法により、多糖はタンパク質キャリアに複合されている。

#### 【0012】

コンジュゲートワクチンは、小児に細菌感染に対する防御のために投与することができ、また成人に長期の持続免疫応答を提供することもできる。WO 2009/104074(Fernandezら)の作製体は、動物においてIgG応答を生じることがわかっている。ヒトにおける赤痢菌O-特異的多糖-タンパク質コンジュゲートワクチンに対するIgG応答は、ヒトにおける免疫防御に相関していることがわかっている(Passwell, J.H.らの文献、「イスラエルにおける成人での改良された赤痢菌O-特異的多糖-タンパク質コンジュゲートワクチンの安全性及び免疫原性(Safety and Immunogenicity of Improved Shigella O-Specific Polysaccharide-Protein Conjugate Vaccines in Adults in Israel)」、Infection and Immunity, 69(3):1351-1357 (2001年3月))。多糖(すなわち、糖残基)は、糖特異的である短期間の免疫応答を誘発すると考えられる。実際ヒトの免疫系は、O-抗原及び莢膜多糖などの、細菌の特異的多糖表面構造に対し強力な反応を生じる。しかし多糖に対する免疫応答はIgM依存型であるので、この免疫系は記憶を発生しない。多糖を運搬するタンパク質キャリアの免疫系は記憶を呈するので、この免疫系は、T-細胞依存型であり、且つ長期の防御を提供するIgG応答を誘発する。

30

40

#### 【0013】

大腸菌0157は、溶血性尿毒症症候群の最近の全症例のおよそ2/3の原因である腸管出血性株であり、ヒトの健康上重大な懸念がある(Law, D.の文献、(2000) J. App. Microbiol., 88, 729-745; Wang, L.及びReeves, P. R.の文献、(1998) Infect. Immun., 66, 3545-3551)。

#### 【0014】

大腸菌(*Escherichia coli*)0157株は、反復する四糖体単位(4-N-アセチルペロサミン→

50

フコース→グルコース→GalNAc)( $\alpha$ -D-PerNAc- $\alpha$ -L-Fuc- $\beta$ -D-Glc- $\alpha$ -D-GalNAc)を含むO-抗原を産生する(Perry, M. B., MacLean, L.及びGriffith, D. W.の文献、(1986) *Biochem. Cell Biol.*, 64, 21-28)。この四糖体は、ウンデカプレニルピロリン酸上で予め構築される。この大腸菌細胞エンベロープは、内側細胞質膜、圧負(stress-bearing)ペプチドグリカン層、並びにリン脂質の内側単層及び細菌LPSで構成された外側単層からなる非対称外膜を含む。LPSは、3つの構成成分、リポDAアンカー、3-デオキシ-D-マンノ-オクト-2-ウロソニン酸-含有コア、及びO-抗原領域を含む(参照: Raetz, C. R. H.及びWhitfield, C.の文献、(2002) *Annu. Rev. Biochem.*, 71, 635-700; Whitfield, C.の文献、(2006) *Annu. Rev. Biochem.* 75, 39-68; Samuel, G.及びReeves, P. R.の文献、(2003) *Carbohydrate Research*, 338, 2503-2519; 及び、細菌LPSのO-抗原の構築に関する総説についてこれらの参考文献を参照されたい)。

10

#### 【0015】

細菌LPSのO-抗原の構成成分は、単独の反復する単糖から構成されたホモポリマー、又は、3~6の糖単位の10~30の反復を含むヘテロポリマーのいずれかであることができる、巨大であり、極めて多様な多糖体である(Reeves, P. R., Hobbs, M., Valvano, M. A., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C. R. H., 及びRick, P. D.の文献、(1996) *Trends Microbiol.*, 4, 495-503)。従ってO-抗原は、その細菌の細胞表面の支配的な特徴であり、且つ毒性及び病原性の重要な決定因子を構成する(Law, D.の文献、(2000) *J. App. Microbiol.*, 88, 729-745; Spears, K. J., Roe, A. J.及びGally, D. L.の文献、(2006) *FEMS Microbiol. Lett.*, 255, 187-202; Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Wang, Q., Reeves, P. R.及びWang, L.の文献、(2008) *FEMS Microbiol. Rev.*, 32, 627-653; Stenutz, R., Weintraub, A.及びWidmalm, G.の文献、(2006) *FEMS Microbiol. Rev.*, 30, 382-403)。独自のO-抗原構造に寄与した180よりも多い個別のO-血清型を有する大腸菌株が、同定されている(Stenutz, R., Weintraub, A.及びWidmalm, G.の文献、(2006) *FEMS Microbiol. Rev.*, 30, 382-403)。

20

#### 【0016】

O-抗原反復単位は、ウンデカプレニルピロリン酸に結合された内側膜の細胞質ゾル面上に予め構築されている。脂質-結合された反復単位は、内側膜のペリプラズム表面へと横断して広がり(フリップ-フロップ)、外膜へ輸送され且つLPSに結合する前に重合される。ほとんどのヘテロポリマー性O-抗原反復単位は、それらの還元末端にN-アセチルグルコサミン("GlcNAc")又はN-アセチルガラクトサミン("GalNAc")のいずれかを有する。

30

#### 【0017】

脂質中間体の生合成は、WecAにより触媒されたGlcNAc-P又はGalNAc-Pのそれらの各糖ヌクレオチド誘導体からウンデカプレニルリン酸("Und-P")への転移により開始されると仮定されている(Samuel, G.及びReeves, P. R.の文献、(2003) *Carbohydrate Research*, 338, 2503-2519; Alexander, D. C.及びValvano, M. A.の文献、(1994) *J. Bacteriol.*, 176, 7079-7084; Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Tolvanen, P.及びSkurkik, M.の文献、(1997) *Mol. Microbiol.* 23, 63-76; Amor, P. A. 及びWhitfield, C.の文献、(1997) *Mol. Microbiol.* 26 (145-161); Wang, L. 及びReeves, P. R.の文献、(1998) *Infect. Immun.* 66, 3545-3551)。WecAのGlcNAc-ホスホトランスフェラーゼ活性の特性及び特異性は特徴付けられているが(Rush, J. S., Rick, P. D. 及びWaechter, C. J.の文献、(1997) *Glycobiology*, 7, 315-322)、WecAは、GalNAc-P-P-Undの合成を触媒するという結論は、遺伝子研究を基にしていた(Wang, L. 及びReeves, P. R.の文献、(1998) *Infect. Immun.* 66, 3545-3551)。そのような初期の遺伝子研究は、脂質-結合された四糖体中間体の生合成は、WecAにより触媒されたGalNAc-PのUDP-GalNAcからUnd-Pへの酵素転移により開始されることを指摘した(Wang, L. 及びReeves, P. R.の文献、(1998) *Infect. Immun.* 66, 3545-3551)。しかし、WecAはGalNAc-P供与体としてUDP-GalNAcを利用することを明らかにする直接の酵素学的証拠は存在しない。

40

#### 【0018】

50

更に、大腸菌055のgne及びgne1遺伝子は、UDP-GlcNAc 4-エピメラーゼをコードしていることが先に提唱された(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.及びReeves, P.R.の文献、(2002) J. Bacteriol. 184, 2620-2625; Guo, H., Yi, W., Li, L.及びWang, P. G.の文献、(2007) Biochem. Biophys. Res. Commun., 356, 604-609)。先の報告は、大腸菌055由来(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.及びReeves, P.R.の文献、(2002) J. Bacteriol. 184, 2620-2625)、並びに大腸菌086由来の(Guo, H., Yi, W., Li, L.及びWang, P. G.の文献、(2007) Biochem. Biophys. Res. Commun., 356, 604-609)、同じ遺伝子ファミリー内のZ3206遺伝子と100%同一である2つの遺伝子、各々、大腸菌055 gne及び大腸菌086 gne1を同定した。

【0019】

10

従って当業者は、Z3206遺伝子も、UDP-GlcNAc/UDP-GalNAcエピメラーゼをコードしていると考えられている。

【発明の概要】

【0020】

(発明の簡単な概要)

現在驚くべきことに、大腸菌0157においてZ3206遺伝子によりコードされたエピメラーゼは、ウンデカプレニルピロリン酸上のN-アセチルガラクトサミン("GalNAc")を合成する反応を触媒し、このことはオリゴ糖又は多糖の形成を開始することが発見された。

【0021】

一態様において、本発明は、ウンデカプレニルピロリン酸上でGalNAcを合成するエピメラーゼを含む、多糖の全て又は一部を生成する組換え原核細胞生合成システムに関する。本発明は更に、還元末端にGalNAcを有する多糖の全て又は一部を合成するグリコシルトランスフェラーゼを含み、且つ更に還元末端にGalNAcを有する抗原性多糖の全て又は一部を合成するグリコシルトランスフェラーゼを含んでいる。

20

【0022】

別の態様において、本発明は、ウンデカプレニルピロリン酸上でGalNAcを生成するエピメラーゼに関し、且つ更なる態様において、本エピメラーゼは、Z3206遺伝子によりコードされている。

【0023】

追加の態様において、本発明は、オリゴサッカリルトランスフェラーゼをコードしているヌクレオチド配列；タンパク質キャリアをコードしているヌクレオチド配列；少なくとも1種の細菌由来の少なくとも1つのオリゴ糖又は多糖の遺伝子クラスターであり、ここで該多糖は、還元末端にGalNAcを含むもの；及び、エピメラーゼをコードしている核酸配列：を含む、N-グリコシル化タンパク質の生成のための発現システムに関する。

30

【0024】

より更なる態様において、本発明は、GlcNAc-P-P-UndをGalNAc-P-P-Undに変換するエピメラーゼをコードしているZ3206遺伝子を含む、組換え原核細胞生合成システムに関する。

【0025】

更に追加の態様において、本発明は、GlcNAc-P-P-UndをGalNAc-P-P-Undに変換するエピメラーゼをコードしている大腸菌055 gne遺伝子又は大腸菌086 gne1遺伝子を含む、組換え原核細胞生合成システムに関する。

40

【0026】

更に別の態様において、本発明は、少なくとも1つの導入されたコンセンサス配列D/E-X-N-Z-S/T(式中、X及びZはプロリン以外の任意の天然のアミノ酸であり得る)、及び還元末端にN-アセチルガラクトサミンを有するグリカン：を含む、N-グリコシル化タンパク質に関する。

【0027】

より別の態様において、本発明は、少なくとも1つの導入されたコンセンサス配列D/E-X-N-Z-S/T(式中、X及びZはプロリン以外の任意の天然のアミノ酸であり得る)；還元末端に

50

N-アセチルガラクトサミンを有する、免疫原性グリカン：を有するN-グリコシル化タンパク質；並びに、アジュバントを含有する、バイオコンジュゲートワクチンに関する。

【0028】

追加の態様において、本発明は、還元末端にGalNAcを含む少なくとも1種の細菌由来の少なくとも1つのオリゴ糖又は多糖を構築するグリコシルトランスフェラーゼ；タンパク質キャリア；オリゴサッカリルトランスフェラーゼ；及び、エピメラーゼ：をコードしている核酸を含む宿主細胞において、N-結合型グリコシル化タンパク質を生成する方法に関する。

【0029】

更なる態様において、本発明は、バイオコンジュゲートワクチンを調製するための生合成システム及びタンパク質の使用に関する。

10

【0030】

追加の態様において、本発明は、単糖、オリゴ糖及び多糖を生成する方法に関し、より更なる態様において、本発明は、抗原性グリカン及びN-グリコシル化タンパク質を生成する方法に関する。

【図面の簡単な説明】

【0031】

【図1】図1は、大腸菌0157由来の膜画分による、 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ 合成の時間経過を示している。大腸菌0157株由来の膜画分は、 $\text{UDP-}[^3\text{H}]\text{GlcNAc}$ と共に、 $37^\circ\text{C}$ で指定された時間インキュベートした。 $[^3\text{H}]$ 脂質生成物を抽出し、 $[^3\text{H}]\text{GlcNAc}$ の $[^3\text{H}]\text{GlcNAc-P-P-Und}(\bigcirc)$ 及び $[^3\text{H}]\text{GalNAc-P-P-Und}(\bullet)$ への取り込みを、実施例2に記載のようにアッセイした。

20

【0032】

【図2】図2は、 $\text{GlcNAc-P-P-Und}$ からの $\text{GalNAc-P-P-Und}$ の形成に関する提唱された生合成経路を示す。

【0033】

【図3】図3は、大腸菌0157株由来の膜画分により合成された $[^3\text{H}]\text{GalNAc-P-P-Und}$ の精製及び特徴決定を示す。大腸菌0157由来の膜画分は、 $\text{UDP-}[^3\text{H}]\text{GlcNAc}$ と一緒にインキュベートし、且つ $[^3\text{H}]\text{GalNAc}$ 脂質を、実施例3に記載のように精製した。図3Aは、DEAE-セルロース上での精製後の、ホウ酸含浸シリカゲルG(Quantum 1)上の $[^3\text{H}]\text{HexNAc}$ 脂質の分取薄層クロマトグラムを示す。図3Bは、パネルAの分取プレートからの回収後の、ホウ酸含浸シリカゲルG(Baker Si250)上の精製された $[^3\text{H}]\text{GalNAc-P-P-Und}$ の薄層クロマトグラフィーを示す。図3Cは、図3Bで精製された $[^3\text{H}]\text{GalNAc-P-P-Und}$ の弱酸加水分解後に回収された $[^3\text{H}]$ -アミノ糖の、下降ペーパークロマトグラム(ホウ酸含浸ワットマン濾紙No.1)を示す。図3Dは、 $\text{NaBH}_4$ による図3C由来の $[^3\text{H}]$ アミノ糖の還元により生成された $[^3\text{H}]\text{HexNAc-アルジトール}$ の下降ペーパークロマトグラム(ワットマンNo.3MM)を示す。

30

【0034】

【図4】図4は、pMLBAD:Z3206による形質転換後の大腸菌21546細胞及び大腸菌21546細胞の代謝標識を示す。大腸菌21546(図4A)及び大腸菌21546:pMLBAD/Z3206(図4B)は、 $[^3\text{H}]\text{GlcNAc}$ により $37^\circ\text{C}$ で5分間代謝標識した。 $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ を抽出し、水溶性夾雑物を除去し、且つ実施例3に記載のようなホウ酸含浸シリカゲルプレート(Baker Si250)上の薄層クロマトグラフィーにより分離した。Bioscanクロマトスキャナーを用い、放射性脂質を検出した。 $\text{GalNAc-P-P-Und}$ 及び $\text{GlcNAc-P-P-Und}$ のクロマトグラフィー位置は、矢印で示した。

40

【0035】

【図5】図5は、大腸菌株由来の膜画分の $\text{UDP-}[^3\text{H}]\text{GlcNAc}$ と一緒にインキュベーションにより形成された $[^3\text{H}]\text{GlcNAc}/\text{GalNAc-P-P-Und}$ の薄層クロマトグラフィーを示す。大腸菌K12株(図5A)、0157株(図5B)、21546株(図5C)、及び21546:pMLBAD/Z3206株(図5D)由来の膜画分を、 $\text{UDP-}[^3\text{H}]\text{GlcNAc}$ と一緒に、 $37^\circ\text{C}$ で10分間インキュベーションし、且つ $[^3\text{H}]$ 脂質生成物を抽出し、分配により水溶性夾雑物を除去し、実施例3に記載のようにホウ酸含浸シリ

50

カゲルプレート(Baker Si250)上の薄層クロマトグラフィーにより分離した。GalNAc-P-P-Und及びGlcNAc-P-P-Undのクロマトグラフィー位置は、矢印で示した。

【0036】

【図6】図6は、UMPと一緒にインキュベーションによるGlcNAc-Pの放出を示す。大腸菌21546:Z3206由来の膜画分を、GlcNAc-P-P-Undを酵素標識するために、UDP-[<sup>3</sup>H]GlcNAcと一緒に37℃で10分間プレインキュベーションし(図6A)、引き続き1mM UMPと一緒に、1分間(図6B)又は2分間(図6C)のいずれかの期間、2回目のインキュベーションした。指定されたインキュベーション期間の後、[<sup>3</sup>H]GlcNAc/GalNAc-P-P-Undを抽出し、実施例3に記載のようなホウ酸含浸シリカゲルプレート(Baker Si250)上の薄層クロマトグラフィーにより分離した。GalNAc-P-P-Und及びGlcNAc-P-P-Undのクロマトグラフィー位置は、矢印で示した。

10

【0037】

【図7】図7は、Z3206を発現している21546株由来の膜により触媒された、外因性[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの、関連する[<sup>3</sup>H]HexNAc-P-P-Und生成物への変換を示す。大腸菌21546株(図7B及び図7E)及び21546:pMLBAD/Z3206株(図7C及び図7F)由来の膜画分を、精製した[<sup>3</sup>H]GlcNAc-P-P-Und(図7A、図7B、及び図7C)又は[<sup>3</sup>H]GalNAc-P-P-Und(図7D、図7E、及び図7Fのパネル)(1%Triton X-100中に超音波処理により分散された)と一緒に、37℃で1分間インキュベートした。[<sup>3</sup>H]GlcNAc/GalNAc-P-P-Undを抽出し、ホウ酸含浸シリカゲルプレート(Baker Si250)上の薄層クロマトグラフィーにより分離し、実施例3に記載のようなBioscan AR2000ラディオクロマトスキャナーにより検出した。

20

【0038】

【図8】図8は、グリコシル化されない及びグリコシル化されたAcrAタンパク質のSDS-PAGE分析を示す。pMLBAD:Z3206(レーン1)、pMLBAD:gne(レーン2)又はベクター対照pMLBAD(レーン3)を補完したAcrA発現プラスミド及びpglオペロンΔgneを保持する大腸菌DH5α細胞から調製したペリプラズム抽出物を、10%SDS-PAGEにより分離し、ニトロセルロースメンブレンに移した。AcrA及びそのグリコシル化された形態は、抗AcrA抗血清により検出した。グリコシル化されないAcrA(AcrA)及びグリコシル化されたAcrA(gAcrA)に対応するバンドの位置を示した。

【0039】

【図9】図9は、Liu Bらの文献(「赤痢菌O-抗原の構造及び遺伝子(Structure and genetics of Shigella O antigens)」、FEMS Microbiology Review, 2008. 32: p.27)により同定された遺伝子を示す。

30

【0040】

【図10】図10は、S.フレックスネリ6 O抗原を合成するために必要とされる遺伝子を含むDNA領域を示すスキームである。

【0041】

【図11】図11は、大腸菌におけるS.フレックスネリ6のO抗原の発現を示す。LPSは、銀染色によるか、又はニトロセルロースメンブレンへの転移及びS.フレックスネリ6に対する抗体による検出のいずれかにより可視化した。

【0042】

40

【図12】図12は、O抗原のHPLCを示す。S.フレックスネリ-Z3206を含む大腸菌細胞(SCM3)、S.フレックスネリ+Z3206を含む大腸菌細胞(SCM3)、又は空の大腸菌(SCM3)細胞のLLO分析。

【0043】

【図13】図13は、EPA、pglB及びS.フレックスネリ6のO-抗原+/-Z3206を発現している大腸菌細胞由来のニッケル精製したタンパク質のウェスタンブロットを示す。

【発明を実施するための形態】

【0044】

(発明の詳細な説明)

本発明は、還元末端にN-アセチルガラクトサミンを有するオリゴ糖又は多糖を合成する

50

エピメラーゼ、及びグリカンの還元末端にN-アセチルガラクトサミンを有するN-グリコシル化タンパク質：をコードしている核酸を含む、組換え原核細胞生合成システムを包含している。

#### 【0045】

用語「部分的アミノ酸配列」は、「最適化コンセンサス配列」又は「コンセンサス配列」とも称される。最適化コンセンサス配列は、通常のコンセンサス配列"N-X-S/T"よりもはるかに効率的に、オリゴサッカリルトランスフェラーゼ("OST"、"OTase")によりN-グリコシル化される。

#### 【0046】

概して、用語「組換えN-グリコシル化タンパク質」とは、該タンパク質をコードしている核酸を天然には含まない宿主細胞において生成される任意のポリペプチド又はオリゴペプチドをいう。本発明の状況において、この用語は、例えば、エシェリキア(*Escherichia*)種、カンピロバクター(*Campylobacter*)種、サルモネラ(*Salmonella*)種、シゲラ(*Shigella*)種、ヘリコバクター(*Helicobacter*)種、シュドモナス(*Pseudomonas*)種、バチルス(*Bacillus*)種などの原核宿主細胞において、並びに更なる実施態様においては、大腸菌、カンピロバクター・ジェジュニ、サルモネラ・ティフィムリウスなどにおいて、組換えにより生成されたタンパク質をいい、ここで該タンパク質をコードしている核酸は、該宿主細胞に導入されており、且つこのコードされたタンパク質は、OTaseにより、N-グリコシル化され、該転移酵素は該宿主細胞において天然に生じるか又は細胞へ組換えにより導入されている。

#### 【0047】

アミノ酸に関する国際的に承認された1文字コードに従い、省略形D、E、N、S及びTは、各々、アスパラギン酸、グルタミン酸、アスパラギン、セリン、及びトレオニンを意味する。

#### 【0048】

本発明のタンパク質は、タンパク質に導入され且つN-グリコシル化された最適化コンセンサス配列D/E-X-N-Z-S/Tを1つ以上含む。従って本発明のタンパク質は、同じく最適化コンセンサス配列を含むが追加の(導入された)最適化コンセンサス配列は全く含まない天然のC.ジェジュニのN-糖タンパク質とは異なる。

#### 【0049】

最適化コンセンサス配列の導入は、1個以上のアミノ酸の付加、欠失及び/又は置換により達成することができる。最適化コンセンサス配列の導入を目的とした1個以上のアミノ酸の付加、欠失及び/又は置換は、化学合成戦略により達成することができ、これは本発明を考慮し、固相支持型化学ペプチド合成のように、当業者には周知であろう。あるいは、より大きいポリペプチドに好ましいことだが、本発明のタンパク質は、本発明に照らして当該技術分野において標準技術である組換え技術により調製することができる。

#### 【0050】

本発明のタンパク質は、高い効率で且つ任意の宿主において生成され得るという利点を有する。本発明の一実施態様において、宿主は、カンピロバクター種に由来する、例えばC.ジェジュニに由来する機能性pglオペロンを含む。更なる実施態様において、本発明の実施のためのカンピロバクター種由来のオリゴサッカリルトランスフェラーゼは、カンピロバクター・コリ(*Campylobacter coli*)又はカンピロバクター・ラリ(*Campylobacter lari*)に由来している。本発明を考慮し、オリゴサッカリルトランスフェラーゼは、当業者には明らかであろう。例えば、オリゴサッカリルトランスフェラーゼは、Szymanski, C.M. 及びWren, B.W.の文献「粘膜病原菌におけるタンパク質グリコシル化(Protein glycosylation in bacterial mucosal pathogens)」, Nat. Rev. Microbiol., 3:225-237 (2005)などの参考文献において明らかにされている。前記原核細胞宿主がカンピロバクター種、又は例えばC.ジェジュニである場合、機能性pglオペロンは天然に存在し得る。しかし、当該技術分野において先に明らかにされ、かつ前述のように、pglオペロンは、細胞に転移され、且つ該新規細胞の環境において機能性を維持する。

## 【0051】

用語「カンピロバクター種、好ましくはC.ジェジュニ由来の機能性のpglオペロン」とは、カンピロバクター種、例えばC.ジェジュニの機能性のオリゴサッカリルトランスフェラーゼ(OTase)、及び脂質キャリア上のオリゴ糖の構築が可能である1種以上の特異的グリコシルトランスフェラーゼをコードしている核酸のクラスターをいい、ここで該オリゴ糖は、OTaseにより、脂質キャリアから、1つ以上の最適化アミノ酸配列：D/E-X N-Z-S/Tを有する標的タンパク質へ転移され得ることを意味する。本発明の状況において用語「カンピロバクター種、好ましくはC.ジェジュニ由来の機能性のpglオペロン」は、必ずしも、単独の転写単位としてのオペロンをいうものではないことが理解されるべきである。この用語は、1つの宿主細胞における組換えタンパク質のN-グリコシル化のための機能性構成成分の存在のみを必要とする。これらの構成成分は、1つ以上の個別のmRNAとして転写されてもよく、且つ一緒に又は個別に調節されてもよい。例えばこの用語は、1つの宿主細胞においてゲノムDNA及びプラスミドに配置された機能性構成成分も包含している。効率化を目的として、一実施態様においては、機能性のpglオペロンの全ての構成成分が調節され、且つ同時に発現される。

10

## 【0052】

オリゴサッカリルトランスフェラーゼは、一部の実施態様においてはカンピロバクター種を起源とし、且つ別の実施態様においてはC.ジェジュニを起源とすることができる。追加の実施態様において、オリゴサッカリルトランスフェラーゼは、例えば、ウォリネラ(Wolinella)種及び真核生物など、オリゴサッカリルトランスフェラーゼを有することが当業者に公知である他の生物を起源とすることができる。

20

## 【0053】

脂質キャリア上にオリゴ糖を構築することが可能である1つ以上の特異的グリコシルトランスフェラーゼは、宿主細胞を起源とするか又は該宿主細胞へ組換えにより導入されることができ、その唯一の機能上の制限は、該グリコシルトランスフェラーゼにより構築されたオリゴ糖は、OTaseにより、脂質キャリアから1つ以上の最適化コンセンサス配列を有する標的タンパク質へ転移され得ることである。従って、特異的グリコシルトランスフェラーゼを天然に含み及び/又は該宿主に天然に存在する特異的グリコシルトランスフェラーゼを交換する宿主細胞の選択に加え、異種の特異的グリコシルトランスフェラーゼの導入は、当業者が、本発明のタンパク質において最適化N-グリコシル化コンセンサス部位に結合されたN-グリカンを変動することを可能にする。

30

## 【0054】

上記の結果として、本発明は、本発明のタンパク質上のN-グリカン-パターンの個別のデザインを提供する。従ってこれらのタンパク質は、生物学的、医薬的及び精製上の必要性に適するようにそれらのN-グリカンパターンを個別化することができる。

## 【0055】

本発明の実施態様において、前記タンパク質は、該N-グリコシル化された最適化アミノ酸配列を1つ、しかし同じく少なくとも2つ、少なくとも3つ又は少なくとも5つなど1つよりも多くも含むことができる。

## 【0056】

本発明のタンパク質内の1つ以上のN-グリコシル化された最適化アミノ酸配列の存在は、それらの免疫原性を高める上で、それらの安定性を増大する上で、それらの生物活性に影響を及ぼす上で、それらの生物学的半減期を延長する上で、及び/又はそれらの精製を簡略化する上で利点であり得る。

40

## 【0057】

本最適化コンセンサス配列は、X及びZの位置に、プロリン以外の任意のアミノ酸を含むことができる。用語「任意のアミノ酸」とは、OTaseにより最適化コンセンサス配列をN-グリコシル化させることが依然可能である、一般的な及び稀少な天然のアミノ酸に加え、合成のアミノ酸誘導体及び類似体を包含することを意味する。X及びZについて、天然の一般的な及び稀少なアミノ酸が好ましい。X及びZは、同じ又は異なることができる。

50

## 【0058】

本発明のタンパク質中の各最適化コンセンサス配列に関して、X及びZは異なってもよいことに留意されたい。

## 【0059】

OTaseによる転移のために脂質キャリア上にオリゴ糖を構築する場合、最適化コンセンサス配列に結合されたN-グリカン、特異的グリコシルトランスフェラーゼ及びそれらの相互作用により決定されるであろう。本発明を考慮し、当業者は、所望の宿主細胞に存在する特異的グリコシルトランスフェラーゼの種類及び量を変動することにより、N-グリカンをデザインすることができるであろう。

## 【0060】

本明細書において使用される「単糖」とは、1個の糖残基をいう。「オリゴ糖及び多糖」とは、2個以上の糖残基をいう。本明細書において使用される用語「グリカン」とは、単糖、オリゴ糖又は多糖をいう。「N-グリカン」は、本明細書において、N-グリコシド結合を介してタンパク質内のアスパラギン残基のε-アミド窒素に連結された変動可能な組成の単糖、オリゴ糖又は多糖と定義される。ある実施態様において、OTaseにより転移されたN-グリカンは、グラム陰性菌又はグラム陽性菌の細胞質膜に存在するウンデカプレノールピロリン酸("Und-P-P")脂質-アンカー上に構築される。これらは、O抗原、O多糖及びペプチドグリカンの合成に参与している(Bugg, T. D.及びBrandish, P. E.の文献、(1994)、*From peptidoglycan to glycoproteins: common features of lipid-linked oligosaccharide biosynthesis*)、FEMS Microbiol Lett, 119, 255-262; Valvano, M. A.の文献、(2003)、*「O-特異的リポ多糖体の輸送(Export of O-specific lipopolysaccharide)」*、Front Biosci, 8, s452-471)。

## 【0061】

脂質-連結された反復する四糖体(4-N-アセチルペロサミン→フコース→グルコース→GalNAc)の生合成は、WecAによるGalNAc-P-P-Undの形成により開始されるかどうかを決定するために、研究を行った。大腸菌K12株、O157株、及びPR4019株、WecA-過剰発現株由来の膜画分がUDP-[<sup>3</sup>H]GalNAcと一緒にインキュベートされる場合、[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの酵素合成はどちらも検出されなかった。しかし、O157株由来の膜画分がUDP-[<sup>3</sup>H]GlcNAcと一緒にインキュベートされる場合、[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの化学特性及びクロマトグラフィー特性を持つ2つの酵素標識された生成物が観察され、O157株は、GlcNAc-P-P-Und及びGalNAc-P-P-Undの相互変換を可能にするエピメラーゼを含むことを確認した。外因性[<sup>3</sup>H]GlcNAc-P-P-Undは、O157株由来の膜と一緒にインキュベートされる場合、[<sup>3</sup>H]GalNAc-P-P-Undへ変換されたことを示すことによって、エピメラーゼの存在が確認された。O157株が[<sup>3</sup>H]GlcNAcにより代謝標識される場合、[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの両方が検出された。大腸菌21546株のZ3206遺伝子による形質転換は、これらの細胞がインビボ及びインビトロにおいてGalNAc-P-P-Undを合成することを可能にした。このエピメラーゼ反応の可逆性は、O157株由来の膜が外因性[<sup>3</sup>H]GalNAc-P-P-Undと一緒にインキュベートされる場合、[<sup>3</sup>H]GlcNAc-P-P-Undが再形成(reform)されたことを示すことによって、明らかにされた。大腸菌におけるカンピロバクター・ジェジュニN-グリコシル化システムの発現においてZ3206はgne遺伝子の喪失を補完することができないことは、これがUDP-GlcNAc/UDP-GalNAcエピメラーゼとしては機能しないことを示した。これらの結果を基に、GalNAc-P-P-Undは、大腸菌O157におけるWecAによるGlcNAc-P-P-Undの形成後に、GlcNAc-P-P-Undエピメラーゼにより可逆的に合成されることが確認された。

## 【0062】

GalNAc-P-P-Und合成は、WecAによるよりもむしろ、いくつかのこれまで未知の機構により触媒されることを確認するために、大腸菌O157のO-抗原サブユニット構築の開始反応を調べた。本明細書に提示された証拠は、GalNAc-P-P-Undは、WecAにより触媒されたUDP-GalNAcからのGalNAc-P転移により合成されないが、しかし大腸菌O157のZ3206遺伝子により

10

20

30

40

50

コードされたエピメラーゼにより触媒されたGlcNAc-P-P-Undの4-OHの可逆的エピマー化により合成されることを示している。

#### 【0063】

従って本発明は、GalNAc-P-P-Undの合成のための新規生合成経路に加え、重要な細菌細胞の表面構成成分の構築のための新規生合成経路を包含している。本発明の更なる実施態様は、抗微生物薬の新規標的としての細菌エピメラーゼを含む。

#### 【0064】

大腸菌0157は、反復する四糖体構造(4-N-アセチルペロサミン→フコース→グルコース→3fGalNAc)を持つO-抗原を合成する。以前の遺伝子研究(Wang, L.及びReeves, P. R.の文献、(1998) Infect. Immun. 66, 3545-3551)とは対照的に、本明細書においては、脂質-連結された四糖体中間体の生合成は、WecAにより触媒されたUDP-GalNAcからUnd-PへのGalNAc-Pの酵素転移により開始されなかったことが示された。本明細書において説明され、ホモロジー検索により得られ、その後遺伝子、酵素及び代謝標識の実験の結果により確認された本発明は、WecAは、基質としてUDP-GalNAcを利用しないこと、しかしWecAは、0157株のZ3206遺伝子によりコードされたエピメラーゼによりその後GalNAc-P-P-Undへ可逆的に変換されるGlcNAc-P-P-Undの合成に必要であることを明らかにしている。

#### 【0065】

本発明のZ3206遺伝子は、それらの還元末端にGalNAc残基を含む表面O-抗原反復単位を生成するいくつかの菌株に存在する遺伝子ファミリーに属する(表1)。Z3206遺伝子配列は、配列番号:1に示されている。先行する報告は、Z3206遺伝子と100%同一である、大腸菌055由来(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.及びReeves, P.R.の文献、(2002) J. Bacteriol. 184, 2620-2625)及び大腸菌086由来(Guo, H., Yi, W., Li, L.及びWang, P. G.の文献、(2007) Biochem. Biophys. Res. Comm., 356, 604-609)の2つの遺伝子、各々、大腸菌055 *gne*及び大腸菌086 *gne1*を報告している(表1)。大腸菌055 *gne*遺伝子配列は配列番号:3として示され、大腸菌086 *gne1*遺伝子配列は配列番号:5として示されている。

#### 【表1】

表1

O-抗原鎖を発現している細菌株におけるZ3206遺伝子と還元末端のGalNAcとの相関

|                                      | Z3206 との%同一性 | O-抗原反復単位の還元末端のGalNAc |
|--------------------------------------|--------------|----------------------|
| 大腸菌 O55 <i>gne</i> (配列番号: 3)         | 100          | あり                   |
| 大腸菌 O86 <i>gne1</i> (配列番号: 5)        | 100          | あり                   |
| シゲラ・ボイディO18 <i>gne</i> (配列番号:7)      | 88           | あり                   |
| サルモネラ・エンテリカ O30 <i>gne</i> (配列番号: 9) | 94           | あり                   |
| C.ジェジュニ <i>gne</i> (配列番号:11)         | 21           | なし                   |
| 大腸菌 K12 <i>galE</i> (配列番号: 13)       | 27           | なし                   |
| 大腸菌 O86 <i>gne2</i> (配列番号: 15)       | 18           | あり                   |

#### 【0066】

従って、本出願人らは、大腸菌055 *gne*及び大腸菌086 *gne1*は、各々、055株及び086株において、GlcNAc-P-P-UndのGalNAc-P-P-Undへの変換が可能であるエピメラーゼもコード

しており、これらは還元末端にGalNAcを持つO-抗原反復単位をも生成することを結論づけた(表1)。

#### 【0067】

本研究における2つの実験的アプローチは、Z3206タンパク質は、0157株におけるUDP-GlcNAcのUDP-GalNAcへのエピマー化を触媒しないことを指摘している。第一に、0157株由来の膜が $[^3\text{H}]$ UDP-GalNAcと一緒にインキュベートされる場合、 $[^3\text{H}]$ GlcNAc-P-P-Und及び $[^3\text{H}]$ GalNAc-P-P-Undはどちらも検出されなかった(表3)。Z3206が、 $[^3\text{H}]$ UDP-GalNAcの $[^3\text{H}]$ UDP-GlcNAcへの変換を触媒するならば、 $[^3\text{H}]$ GlcNAc-P-P-Undが観察されるべきであることが予想される。第二に、本出願人らは、赤血球凝集素-タグ付きZ3206は、UDP-GalNAc-依存性C.ジェジュニN-グリコシル化リポーターシステムを補完することが不可能であることを示した(図8)。

10

#### 【0068】

055株由来の大腸菌055 gne遺伝子(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.及びReeves, P.R.の文献、(2002) J. Bacteriol. 184, 2620-2625)も同じく、粗抽出物をUDP-GalNAcと一緒にインキュベートし、酸加水分解後のp-ジメチルアミノベンズアルデヒドとの反応性の増加を測定し、UDP-GlcNAcへの変換を間接的にアッセイすることにより、エピメラーゼ活性についてアッセイした。両方の研究において、本生成物の形成は、糖ヌクレオチド最終生成物の限定的特徴ではなく、p-ジメチルアミノベンズアルデヒドとの反応性の変化を基にした。90%純粋なポリヒスチジン-タグ付き大腸菌086 gne1も同じく、組合せアッセイにおいてGne2と比べ、低レベルのUDP-グルコースエピメラーゼ活性を有することが示された。

20

#### 【0069】

従って、本発明の実施態様は、GlcNAc-P-P-UndをGalNAc-P-P-Undへ変換する、Z3206遺伝子、大腸菌055 gne遺伝子又は大腸菌086 gne1遺伝子を含む、組換え原核細胞生合成システムに関する。

#### 【0070】

追加の非還元末端GalNAcのためのグリコシル供与体として恐らくUDP-GalNAcを必要とする2個のGalNAc残基を含むO-抗原を合成する大腸菌086は、O-抗原遺伝子クラスター内にgne2と称される追加のGlcNAc 4-エピメラーゼ遺伝子も有することは重要である(Guo, H., Yi, W., Li, L.及びWang, P. G.の文献、(2007) Biochem. Biophys. Res. Commun., 356, 604-609)。この追加のエピメラーゼ遺伝子は、コラン酸遺伝子クラスターのgalE遺伝子と高い相同性を有し、且つUDP-GalNAcを合成することが可能であるUDP-GlcNAc 4-エピメラーゼであるように考えられる。

30

#### 【0071】

Z3206遺伝子は、GalNAcで始まる大腸菌O-血清型において高度に保存されるようである。最近の研究において、確立されたO-抗原反復単位構造を有する62種の大腸菌株が、大腸菌0157 Z3206遺伝子の特異的に検出するようにデザインされたヌクレオチドプライマーを使用するポリメラーゼ連鎖反応を基にした方法により、Z3206の発現についてスクリーニングされた(Wang, L., Huskic, S., Cisterne, A., Rothmund, D.及びReeves, P.R.の文献、(2002) J. Bacteriol. 184, 2620-2625)。この研究において、Z3206は、GalNAcを含むことがわかっている22種の大腸菌株中16種において、及びGalNAcを欠く40種の菌株のわずかに4種において、検出された。更に、UDP-GlcNAc 4-エピメラーゼ活性を持つ代替エピメラーゼ(大腸菌0113のGalE遺伝子)を検出するようにデザインされたプライマーによる、22種のGalNAc-含有株の同様のスクリーニングは、この遺伝子を保持する菌株を検出せず、このことはZ3206は、大腸菌のO-抗原反復単位における還元末端GalNAcの存在に最も一般的に関連づけられたGlcNAc 4-エピメラーゼ遺伝子であることを示している。

40

#### 【0072】

様々なウェブ-ベースのトポロジー予測アルゴリズムによるZ3206タンパク質配列の分析は、Z3206タンパク質は高度に疎水性ではないことを示している。大部分のトポロジー予測アルゴリズムは、Z3206は可溶性37kDaタンパク質であることを示しているが、TMPredは

50

、1個の弱いN-末端膜貫通ヘリックスを予測した(Hofmann, K.及びStoffel, W.の文献、(1993) Biol. Chem. Hoppe-Seyler, 374, 166 (要約))。しかし赤血球凝集素-タグ付きZ3206を発現している大腸菌細胞由来の細胞画分のSDS-PAGE後のウェスタンブロットは、このタグ付きタンパク質は、それらの細胞の低張溶解後に、微粒子画分に会合されることを明確に示している。予備実験は、前記タンパク質は、膜画分の1M KClとのインキュベーション後に微粒子画分に会合され続けるが、0.1%Triton X-100とのインキュベーションにより活性型で可溶化されることを示している。

#### 【0073】

大腸菌0157のZ3206は、NAD(P)結合ポケットと一致する、GXXGXXGモチーフ(Rossmannフォールド)(Allard, S. T. M., Giraud, M. F., 及びNaismith, J. H.の文献、(2001) Cell. Mol. Life Sci. 58, 1650-1655)、並びにプロトン引き抜き及び供与に参与した、保存されたSX<sub>24</sub>YX<sub>3</sub>K配列(Field, R. A.及びNaismith, J. H.の文献、(2003) Biochemistry, 42, 7637-7647)を含むオキシド-レダクターゼの短鎖デヒドロゲナーゼ/レダクターゼファミリーと著しい配列相同性を有する。短鎖デヒドロゲナーゼ/レダクターゼファミリーの別のメンバーであるUDP-Glc 4-エピメラーゼの結晶構造を基にした分子モデリングは、ヒドリド引き抜き後、4-ケト中間体が、UDPのβリン酸の周りを回転し、ケト中間体の反対面を提示し、且つその反対面からのヒドリドの再挿入を可能にし、その結果4位の炭素におけるヒドロキシルの立体配置を反転することを示唆している。これらの保存された配列の存在は、Z3206は恐らく類似した機構を介し機能することを示唆している。図7に見られるように、本エピメラーゼ生成物の平衡分布は、GlcNAc-P-P-Undの形成を支持しているように見えるが、O-抗原反復単位の構築のためのGalNAc-P-P-Undの利用は、質量作用により、GalNAc-P-P-Undの方向にエピマー化反応を駆動するであろう。

#### 【0074】

ポリイソプレノイド脂質中間体のグリコシル部分のエピマー化は、実際には広くは報告されていない。1つの先行する研究において、マイコバクテリアにおけるアラビノガラクトサン生合成のアラビノシル供与体であるアラビノシル-P-デカプレノールを形成するリボシル-P-デカプレノールの2-エピマー化が報告された(Mikusova, K., Huang, H., Yagi, T., Holsters, M., Vereecke, D., D'Haeze, W., Scherman, M. S., Brennan, P. J., McNeil, M. R., 及びCrick, D. C.の文献、(2005) J. Bacteriol. 187, 8020-8025)。アラビノシル-P-デカプレノールは、2種のマイコバクテリアタンパク質Rv3790及びRv3791を必要とする、2工程酸化/還元反応を介して形成される。エピマー化は、NAD及びNADPの添加により若干刺激されるが、Rv3790もRv3791も、短鎖デヒドロゲナーゼ/レダクターゼファミリーの特徴であるRossmannフォールド又はSX<sub>24</sub>YXXXXモチーフのいずれも含まない(Allard, S. T. M., Giraud, M.-F.及びNaismith, J. H.の文献、(2001) Cell. Mol. Life Sci. 58, 1650-1655; Field, R. A.及びNaismith, J. H.の文献、(2003) Biochemistry 42, 7637-7647)。

#### 【0075】

まとめると、GlcNAc-P-P-Undのエピマー化によるGalNAc-P-P-Undの形成に関する新規生合成経路が説明されている。

#### 【0076】

いくつかの抗生物質が、GlcNAc-P-P-Undの合成を阻害することが示されているが、それらはタンパク質N-グリコシル化経路を開始するドリコール-連結された中間体であるGlcNAc-P-P-ドリコールの合成も妨害するので、それらの有用性は限定されている。GlcNAc-P-P-ドリコールは、細菌の糖脂質中間体GlcNAc-P-P-Undの構造上関連した哺乳動物の対応物であるが、真核細胞においてはGlcNAc-P-P-ドリコールをGalNAc-P-P-ドリコールに変換する類似したエピマー化反応の証拠は存在しない。従ってこれは、還元末端にGalNAcを含む表面O-抗原が病理発生プロセスに参与している菌株において、細菌のエピメラーゼを阻害することによって、O-抗原合成が、潜在的に妨害される可能性を生じている。

#### 【0077】

本発明の実施態様は、大腸菌0157において、GlcNAc-P-P-Und(N-アセチルグルコサミニ

10

20

30

40

50

ルピロホスホリルウンデカプレノール)を、GalNAc-P-P-Und(N-アセチルガラクトサミニルピロホスホリルウンデカプレノール)に変換するエピメラーゼに参与している。なお更に本発明の例証的態様は、還元末端にGalNAcを有する脂質-結合した反復する四糖体の合成の開始に参与している。

#### 【0078】

本発明の別の態様の基礎は、カンピロバクター・ジェジュニは、一般的N-連結されたタンパク質グリコシル化システムを含むことの発見を含む。C.ジェジュニの様々なタンパク質は、七糖類により修飾されることが示されている。この七糖類は、特異的グリコシルトランスフェラーゼにより触媒されたヌクレオチド活性化された単糖の段階的添加により、内側膜の細胞質側で、キャリア脂質ウンデカプレニルピロリン酸上に構築される。脂質-10  
連結されたオリゴ糖は、次にフリッパーゼ、例えばPglKにより、ペリプラズム間隙へとフリップ-フロップする(横断して広がる)。N-連結されたタンパク質グリコシル化の最終工程において、オリゴサッカリルトランスフェラーゼ(例えば、PglB)は、キャリア脂質からコンセンサス配列D/E-X-N-Z-S/T内のアスパラギン(Asn)残基への、オリゴ糖の転移を触媒する(式中、X及びZは、Pro以外の任意の天然のアミノ酸であり得る)。七糖体に関するグリコシル化クラスターは、大腸菌へ巧く転移され、且つカンピロバクターのN-連結された糖タンパク質が生成されている。

#### 【0079】

PglBは、脂質-連結された糖基質に対し厳密な特異性を有さないことが明らかにされた。ウンデカプレニルピロリン酸上に構築された抗原性多糖は、ペリプラズム内のPglBにより20  
捕獲され、且つタンパク質キャリアへ転移される(Feldmanの文献、2005; Wacker, M.らの文献、「細菌及び真核生物のシステムの共通の転移機序を示唆する細菌オリゴサッカリルトランスフェラーゼの基質特異性(Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems)、Proc Natl Acad Sci USA, 2006. 103(18): p.7088-93)。この酵素が還元末端にN-アセチル化されたヘキソサミンを含む場合、これはウンデカプレニルピロリン酸(UPP)連結されたオリゴ糖の多様なアレイも転移するであろう。pglBヌクレオチド配列及びPglBアミノ酸配列は、W02009/104074に公開されている。

#### 【0080】

従って本発明の一実施態様は、1つ以上の導入されたコンセンサス配列D/E-X-N-Z-S/T(30  
式中、X及びZは、プロリン以外の任意の天然のアミノ酸であり得る)；及び、還元末端に、N-アセチルガラクトサミンを有し、N-グリコシド結合により1以上の該導入されたコンセンサス配列の各々にN-連結されたオリゴ糖又は多糖：を含む、組換えN-グリコシル化タンパク質に参与している。

#### 【0081】

更なる実施態様において、本発明は、ウンデカプレニルピロリン酸上にN-アセチルガラクトサミン("GalNAc")を合成するエピメラーゼを含む、多糖の全て又は一部を生成するための組換え原核細胞生合成システムに関する。更なる実施態様において、これらの多糖の35  
全て又は一部は抗原性である。

#### 【0082】

別の実施態様において、本発明は、ウンデカプレニルピロリン酸上でGalNAcを合成する40  
エピメラーゼ；及び、還元末端にGalNAcを有する多糖を合成するグリコシルトランスフェラーゼ：を含む、組換え原核細胞生合成システムに関する。

#### 【0083】

本発明の実施態様は、ウンデカプレニルピロリン酸上でGalNAcを合成するエピメラーゼ、及び、多糖を合成するグリコシルトランスフェラーゼ：を含む組換え原核細胞生合成システムを更に含み、ここで該多糖は、下記構造を有し： $\alpha$ -D-PerNAc- $\alpha$ -L-Fuc- $\beta$ -D-Glc- $\alpha$ -D-GalNAc；且つ、ここでGalNAcは該多糖の還元末端に存在する。

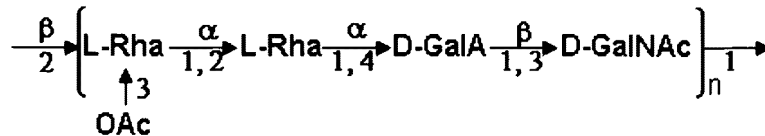
#### 【0084】

本組換え原核細胞生合成システムは、様々な起源の単糖、オリゴ糖又は多糖を生成する50

ことができる。本発明の実施態様は、様々な起源のオリゴ糖及び多糖に関する。そのようなオリゴ糖及び多糖は、原核細胞又は真核細胞を起源とすることができる。原核細胞起源のオリゴ糖又は多糖は、グラム陰性菌又はグラム陽性菌に由来してよい。本発明の一実施態様において、オリゴ糖又は多糖は、大腸菌に由来している。更なる本発明の態様において、該オリゴ糖又は多糖は、大腸菌0157に由来している。別の実施態様において、該オリゴ糖又は多糖は、下記構造を含む： $\alpha$ -D-PerNAc- $\alpha$ -L-Fuc- $\beta$ -D-Glc- $\alpha$ -D-GalNAc。本発明の更なる実施態様において、オリゴ糖又は多糖は、シゲラ・フレックスネリに由来している。より更なる実施態様において、オリゴ糖又は多糖は、シゲラ・フレックスネリ6に由来している。より更なる態様において、該オリゴ糖又は多糖は、下記構造を含む。

【化1】

10



【0085】

本発明の実施態様は、様々な起源のタンパク質を更に含む。そのようなタンパク質は、原核生物及び真核生物について未変性のタンパク質を含む。タンパク質キャリアは、例えばAcrA、又はタンパク質グリコシル化のためのコンセンサス配列、すなわちD/E-X-N-Z-S/T(式中、X及びZは、プロリン以外の任意のアミノ酸であり得る)(例えば、改変された外毒素緑膿菌(*Pseudomonas aeruginosa*)("EPA"))を含むように改変されたタンパク質キャリアであることができる。本発明の一実施態様において、タンパク質は、緑膿菌EPAである。

20

【0086】

本発明の更なる態様は、N-グリカンの還元末端にGalNAcを有する新規バイオコンジュゲートワクチンに関する。追加の本発明の実施態様は、ウンデカプレニルピロリン酸上にGalNAcを生成するエピメラーゼを含む組換え細菌細胞を使用する、そのようなバイオコンジュゲートワクチンを製造する新規アプローチに参与している。一実施態様において、バイオコンジュゲートワクチンは、細菌疾患を治療又は予防するために使用することができる。更なる実施態様において、バイオコンジュゲートワクチンは、癌又は他の疾患に関する治療的及び/又は予防的可能性を有し得る。

30

【0087】

典型的ヒトに関するワクチン接種用量は、約1~25  $\mu$ g、好ましくは約1  $\mu$ g~約10  $\mu$ g、最も好ましくは約10  $\mu$ gである。任意に、本発明のバイオコンジュゲートワクチンなどのワクチンは、アジュバントを含有する。

【0088】

追加の実施態様において、本発明は、オリゴサッカリルトランスフェラーゼをコードしているヌクレオチド配列；タンパク質キャリアをコードしているヌクレオチド配列；少なくとも1種の細菌由来の少なくとも1つの多糖の遺伝子クラスターであり、ここで多糖が還元末端にGalNAcを含むもの；及び、エピメラーゼをコードしている核酸配列：を含有する、少なくとも1種の細菌に対するバイオコンジュゲートワクチンを製造するための発現システムに関する。更なる実施態様において、多糖の遺伝子クラスターは、抗原性多糖をコードしている。

40

【0089】

なお更なる実施態様において、本発明は、オリゴサッカリルトランスフェラーゼをコードしているヌクレオチド配列；少なくとも1つの挿入されたコンセンサス配列D/E-X-N-Z-S/Tであり、ここでX及びZは、プロリン以外の任意の天然のアミノ酸であり得るものを含むタンパク質キャリアをコードしているヌクレオチド配列；少なくとも1種の細菌由来の少なくとも1つの多糖の遺伝子クラスターであり、ここで多糖が還元末端にGalNAcを含むもの；及び、Z3206遺伝子：を含有する、少なくとも1種の細菌に対するバイオコンジュゲートワクチンを製造するための発現システムに関する。更なる実施態様において、この多糖

50

の遺伝子クラスターは、抗原性多糖をコードしている。

【0090】

更に別の実施態様において、本発明は、タンパク質キャリア；該タンパク質キャリアに連結された少なくとも1つの免疫原性多糖鎖であり、ここで該多糖が還元末端にGalNAcを有し、且つ更に該GalNAcがタンパク質キャリアに直接連結されているもの；及び、アジュバント：を含有する、バイオコンジュゲートワクチンに関する。

【0091】

更に追加の実施態様において、本発明は、少なくとも1つの挿入されたコンセンサス配列D/E-X-N-Z-S/T(式中、X及びZは、プロリン以外の任意の天然のアミノ酸であり得る)を含むタンパク質キャリア；該タンパク質キャリアに連結された、少なくとも1種の細菌由来の少なくとも1つの免疫原性多糖であり、ここで少なくとも1つの免疫原性多糖が、タンパク質キャリアに直接連結されたGalNAcを還元末端に含むもの；及び、任意にアジュバント：を含有する、バイオコンジュゲートワクチンに関する。

10

【0092】

本発明の別の実施態様は、グリコシルトランスフェラーゼの使用を通じて、組換え生物において還元末端にGalNAcを有する多糖を構築する工程；該GalNAcを、該組換え生物における1種以上の標的タンパク質のアスパラギン残基へ連結する工程であって、ここで該1種以上の標的タンパク質が1つ以上のT-細胞エпитープを含む工程：を含む、バイオコンジュゲートワクチンの製造方法に関する。

【0093】

更なる実施態様において、本発明は、改変された原核生物を生じるために、原核生物へ標的タンパク質のN-グリコシル化を実行する代謝装置をコードしている遺伝情報を導入する工程であり；ここで1つ以上の組換え標的タンパク質の発現に必要とされる遺伝情報が、該原核生物に導入され；ここで大腸菌0157株エピメラーゼの発現に必要とされる遺伝情報が、該原核生物に導入され；並びに、ここで代謝装置は、脂質キャリア上の還元末端にGalNAcを有する多糖を構築する型のグリコシルトランスフェラーゼ、及びオリゴサッカリルトランスフェラーゼを含み、このオリゴサッカリルトランスフェラーゼは、多糖のGalNAcを標的タンパク質のアスパラギン残基に共有的に連結し、且つこの標的タンパク質は、少なくとも1つのT-細胞エピトープを含むものである工程；改変された原核生物の培養物を生じる工程；並びに、培養培地からグリコシル化タンパク質を得る工程：を含む、バイオコンジュゲートワクチンの製造方法に関する。

20

30

【0094】

本発明の更なる態様は、医薬組成物に関する。本発明の追加の態様は、本発明の少なくとも1つのN-グリコシル化タンパク質を含有する医薬組成物に参与している。本明細書の開示に照らして、タンパク質を含有する医薬品の調製は、当該技術分野において周知であろう。本発明のなお更なる態様は、GlcNAc-P-P-UndをGalNAc-P-P-Undへ変換するエピメラーゼを阻害する抗生物質を含有する医薬組成物に関する。好ましい実施態様において、本発明の医薬組成物は、医薬として許容し得る賦形剤、希釈剤及び/又はアジュバントを含有する。

【0095】

好適な賦形剤、希釈剤及び/又はアジュバントは、当該技術分野において周知である。賦形剤又は希釈剤は、活性成分の溶媒又は媒体として役立つ固体、半固体又は液体の物質であることができる。組成物調製の分野の当業者は、選択された製品の具体的特徴、治療される疾患又は状態、疾患又は状態の病期、及び他の関連状況に応じ適切な投与の形状及び様式を容易に選択することができる(「レミントン薬科学(Remington's Pharmaceutical Sciences)」、Mack Publishing Co. (1990))。医薬として許容し得る希釈剤又は賦形剤の割合及び性質は、選択された医薬活性化合物の溶解度及び化学特性、選択された投与経路、並びに標準の薬務により決定される。医薬品調製品は、経口、非経口又は局所使用に適合されることができ、且つ錠剤、カプセル剤、坐剤、液剤、懸濁剤などの形状で患者へ投与されることができる。本発明の医薬活性化合物は、それら自身が有効である一方、安

40

50

定性、結晶化の簡便性、溶解度の上昇などの目的で、酸付加塩又は塩基付加塩などのそれらの医薬として許容し得る塩の形状で製剤され且つ投与されることができる。

#### 【0096】

特定のヌクレオチド配列又はアミノ酸配列に留意する場合には、本発明は、留意された配列と同じ機能性を依然体現している相同配列を包含することは理解されるであろう。本発明の実施態様において、そのような配列は少なくとも85%相同である。別の実施態様において、そのような配列は少なくとも90%相同である。なお更なる実施態様において、そのような配列は少なくとも95%相同である。

#### 【0097】

2つのヌクレオチド配列又はアミノ酸配列の間の同一性の割合の決定は、当業者に公知である。

10

#### 【0098】

以下に列挙された配列において説明されたもののような、本明細書に記載された核酸配列は、単なる例証であり、これらの配列は様々な方法で組み合わせることができることは当業者には明らかであろう。

#### 【0099】

本発明の追加の実施態様は、核酸の変異形を含む。核酸の変異形(例えば、コドン-最適化核酸)は、配列番号:1、配列番号:3、配列番号:5、配列番号:7、配列番号:9、配列番号:11、配列番号:13、配列番号:15、配列番号:17、配列番号:18、配列番号:19、配列番号:20、配列番号:21、配列番号:22、配列番号:23、配列番号:24、配列番号:27、配列番号:28又は配列番号:29と実質的に同一、すなわち、少なくとも80%同一、例えば、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%又は99.5%同一である。配列番号:1、配列番号:3、配列番号:5、配列番号:7、配列番号:9、配列番号:11、配列番号:13、配列番号:15、配列番号:17、配列番号:18、配列番号:19、配列番号:20、配列番号:21、配列番号:22、配列番号:23、配列番号:24、配列番号:27、配列番号:28又は配列番号:29を含む配列の核酸変異形は、配列番号:1、配列番号:3、配列番号:5、配列番号:7、配列番号:9、配列番号:11、配列番号:13、配列番号:15、配列番号:17、配列番号:18、配列番号:19、配列番号:20、配列番号:21、配列番号:22、配列番号:23、配列番号:24、配列番号:27、配列番号:28又は配列番号:29、若しくはそれらの一部を含む配列から、1個以上のヌクレオチド(例えば、2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、125、150、175又は200個のヌクレオチド)の置換、変異(variation)、修飾、交換、欠失及び/又は付加を伴う核酸を含む。

20

30

#### 【0100】

例えば、本発明の実施態様において、そのような変異形は、GlcNAc-P-P-UndをGalNAc-P-P-Undへ変換するエピメラーゼをコードしている核酸であって、並びにi)例えば大腸菌などの宿主細胞において発現され、且つii)配列番号:1、配列番号:3、配列番号:5、配列番号:7若しくは配列番号:9、又はそれらの一部と実質的に同一である核酸を含む。

#### 【0101】

本明細書に記載された核酸は、組換えDNA及び合成(例えば化学的に合成された)DNAを含む。核酸は、二本鎖又は一本鎖であることができる。一本鎖の核酸である場合、核酸はセンス鎖又はアンチセンス鎖であることができる。核酸は、オリゴヌクレオチド類似体又は誘導体を用いて合成することができる。

40

#### 【0102】

本明細書に記載された核酸を含むプラスミドは、発現のために宿主細胞へと形質導入又は形質転換されることができる。形質導入及び形質転換の技術は、当業者に公知である。

#### 【0103】

本明細書において言及された全ての刊行物は、それらの全体が引用により組み込まれている。本明細書において使用される用語「又は」は、適宜組み合わせることができる代替を意味し；すなわち、用語「又は」は、個別に列挙された代替の各々に加え、それらの組

50

合せを含むと理解されるべきである。本明細書において使用されるように、文脈が別に明確に指摘しない限りは、単数形「1つ(a、an、及びthe)」などの単数の言及は、複数を含み、且つ複数の言及は単数を含む。

#### 【0104】

本発明は、本発明の組成物及び方法に加え、その利用性を更に説明している下記実施例を参照し、更に規定される。当業者には、本発明の範囲内である組成物及び方法の両方の改変を実施することができることは明らかであろう。

#### 【実施例】

#### 【0105】

(実施例)

(細菌株及びプラスミド)－大腸菌PR4019株(Rush, J. S., Rick, P. D.及びWaechter, C. J.の文献、(1997) Glycobiology, 7, 315-322)、及びPR21546株(Meier-Dieter, U., Starman, R., Barr, K., Mayer, H.及びRick, P. D.の文献、(1990) J. Biol. Chem., 265, 13490-13497)は、Dr. Paul Rick(ベセスダ、MD)のご厚意により贈られ、並びに大腸菌0157:H45(Stephan, R., Borel, N., Zweifel, C., Blanco, M., 及びBlanco, J.E.の文献、(2004) BMC Microbiol 4:10)は、Dr. Claudio Zweifel(チューリッヒ大学獣医学研究所)から贈答されたものであった。大腸菌DH5α (Invitrogen社)を、クローニング実験及びタンパク質グリコシル化分析のための宿主として使用した。使用したプラスミドを、表2に列挙した。

#### 【表2】

表2

#### 実施例で使用したプラスミド

| プラスミド                         | 説明                                                                     | 参考文献                           |
|-------------------------------|------------------------------------------------------------------------|--------------------------------|
| pMLBAD                        | クローニングベクター、Tnp <sup>R</sup>                                            | Lefebvre 及び Valvano の文献 (2002) |
| pMLBAD:Z3206<br>(配列番号:<br>23) | pMLBAD中のZ3206、Tnp <sup>R</sup> 、アラビノース-誘導性<br>プロモーターにより制御された発現         | 本明細書の<br>実施例                   |
| pMLBAD:gne<br>(配列番号:<br>24)   | pMLBAD中のgne、Tnp <sup>R</sup> 、アラビノース-誘導性<br>プロモーターにより制御された発現           | 本明細書の<br>実施例                   |
| pACYCpgl                      | C.ジェジュニpglクラスター、Cm <sup>R</sup>                                        | Wacker<br>らの文献<br>(2002)       |
| pACYCgne::kan                 | gne内にkanカセットを含むC. ジェジュニ pgl<br>クラスター、Cm <sup>R</sup> 、Kan <sup>R</sup> | Lintonらの<br>文献(2005)           |
| pWA2                          | pBR322 におけるTetプロモーター制御下の可溶性<br>ペリプラズムヘキサ-His-タグ付きAcrA、Amp <sup>R</sup> | Feldman<br>らの文献<br>(2005)      |

#### 【0106】

(材料)－[1,6-<sup>3</sup>H]GlcNAc(30Ci/mmol)、UDP-[1-<sup>3</sup>H]GlcNAc(20Ci/mmol)及びUDP-[6-<sup>3</sup>H]GlcNAc(20Ci/mmol)は、American Radiolabeled Chemicals社(セントルイス、MO)から入手した。Quantum 1シリカゲルG薄層プレートは、Quantum Industries社(フェアフィールド、NJ)の製品であり、且つBaker Si250シリカゲルGプレートは、Mallinckrodt Chemical Works社により製造されている。酵母エキス及びバクトペプトンは、BD Biosciences社の製品であった。他の化学物質は全て、標準の商業的供給業者から入手した。トリメトプリム(50 µg/ml)、クロラムフェニコール(20 µg/ml)、アンピシリン(100 µg/ml)、及びカナマイ

シン(50 µg/ml)は、必要に応じ培地に添加した。

【0107】

(組換えプラスミドの作製)－大腸菌DH5α株を、DNAクローニング実験に使用し、作製されたプラスミドを、DNA配列決定により検証した。Z3206遺伝子を、大腸菌0157:H45から、オリゴヌクレオチドZ3206-Fw及びZ3206-RvHA

【化2】

AAACCCGGGGATGAACGATAACG TTTTGCTC (配列番号: 17) 及び

AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTACTCAGAAACAA

ACGTTATGTC (配列番号: 18)

10

；(制限部位には下線が付けられている)を用いるPCRにより、増幅した。PCR断片を、SmaI及びXbaIにより消化し、SmaI-XbaI切断されたpMLBADベクター(Lefebvre, M. D. 及びValvan o M. A. の文献、(2002) Appl Environ Microbiol 68: 5956-5964)にライゲーションした。これは、C-末端に赤血球凝集素タグを持つZ3206をコードしているプラスミドpMLBAD:Z3206(配列番号:23)を生じた。

【0108】

gne遺伝子は、オリゴヌクレオチドgne-Fw及びgne-RV

【化3】

AAACCATGGATGAAAATTCTTATTAGCGG (配列番号:19) 及び

AAATCTAGATTAAGCGTAATCTGGAACATCGTATGGGTAGCACTGTTTTTC

CCAATC (配列番号: 20)

20

；(制限部位には下線が付けられている)により、カンピロバクター・ジェジュニpglクラスターをコードしているpACYCpgl(Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M. の文献、(2002) Science 298, 1790-1793)から増幅した。PCR産物を、NcoI及びXbaIで消化し、pMLBADの同じ部位にライゲーションし、C-末端に赤血球凝集素タグを持つGneをコードしているプラスミドpMLBAD:gne(配列番号:24)を生成した(表2)。

30

【0109】

(増殖条件、タンパク質発現及び免疫検出)－大腸菌株を、Luria-Bertani培地(1%酵母エキス、2%バクト-ペプトン、0.6%NaCl)において、37℃で激しく振盪しながら培養した。A<sub>600</sub>が0.05～0.4となるまで増殖させた大腸菌細胞へ、最終濃度0.02～0.2% (w/v)となるようアラビノースを添加することにより、アラビノース誘導性発現を実現した。同量のアラビノースを、誘導後5時間で再度添加し、インキュベーションを4～15時間継続した。

【0110】

(分析手順)－タンパク質濃度は、Pierce Biotechnology社の公報「BCAタンパク質アッセイのための試料からの干渉物質の除去(Eliminate Interfering Substances from Samples for BCA Protein Assay)」に従い、デオキシコール酸及びトリクロロ酢酸による膜タンパク質の沈殿後、BCAタンパク質アッセイ(Pierce社)を使用し決定した。試料は、Packard社Tri-Carb 2100TR液体シンチレーションスペクトロメーターにおいて、1%SDS 0.5ml及びEconosafe Economical Biodegradable Counting Mixture(Research Products International社、マウントプロスペクト、IL)4mlの添加後、シンチレーションスペクトロメーターにより、放射能について分析した。

40

【0111】

(実施例1: GlcNAc-P-P-Und 4-エピメラーゼをコードしている大腸菌0157遺伝子の同定)

本出願人らは、本明細書において、GalNAc-P-P-Undが、これまで未知の4-エピメラーゼの作用により触媒されるGlcNAc-P-P-Undの4-OHのエピマー化により形成される新規合成経路の驚くべき発見を記載する。この経路において、GlcNAc-P-P-Undは、WecAにより触媒

50

されるUDP-GlcNAcからのGlcNAc-Pの転移により合成され、次にGlcNAc-P-P-Undは、GlcNAc-P-P-Und-4-エピメラーゼによりGalNAc-P-P-Undへエピマー化され、これはこれまで未知の経路である(図2)。

#### 【0112】

GlcNAc-P-P-Und 4-エピメラーゼの候補をコードしている遺伝子は、DNAホモロジー検索により同定された。ホモロジー検索は、<http://blast.ncbi.nlm.nih.gov/Blast.cgi>において認められる、米国国立医学図書館(NLM)データベースを用いた。還元末端にGalNAcを有するO-抗原反復単位をコードしている様々な細菌のゲノム配列を検索した。還元末端にGalNAcを含む反復単位を持つ第一群、反復単位に末端GalNAcを欠いている第二群を比較し、可能性のあるエピメラーゼを同定した。これらの分類基準を用い、Z3206を、候補GlcNAc-P-P-Und 4-エピメラーゼとして同定した(表1)。

10

#### 【0113】

GalNAcを含むO-抗原反復単位を有する大腸菌株に存在するGlcNAc 4-エピメラーゼ遺伝子は、表1に示されたように2つのホモロジー群に分けることができる。驚くべきことに、第一のホモロジー群(gne1を含む)は、O-抗原反復単位上の開始の糖としてのGalNAcの存在に明確に相関していることを発見した。更に驚くべきことに、第二群(gne2を含む)は、UDP-GlcエピメラーゼGaleと高度の類似性を示し、且つこれはGalNAcを持つO-抗原反復単位合成を開始しない大腸菌株において認められることを発見した。大腸菌0157におけるZ3206は、gne1と高度の相同性を持つ遺伝子であるが、これは候補GlcNAc-P-P-Und 4-エピメラーゼとして同定された。Z3206遺伝子はO-抗原クラスターのgalFとコラン酸クラスターに属するwcaMの間に存在するので、Z3206遺伝子のゲノム位置は、この経路における役割と一致している。

20

#### 【0114】

実施例2～11に説明された研究は、GalNAc-P-P-Undの形成を触媒するものとしてのGlcNAc 4-エピメラーゼ(大腸菌0157のZ3206)の同定を含む、前記発見を更に確認している。

#### 【0115】

(実施例2：UDP-GalNAcは大腸菌WecA(GlcNAc-ホスホトランスフェラーゼ)の基質でない)

大腸菌WecAは、GalNAc-P-P-Undを形成するためのGalNAc-P供与体としてUDP-GalNAcを利用するかどうかを決定するために、GalNAc-P-P-Undの合成により恐らく開始される還元末端にGalNAcを持つ四糖体O-抗原反復単位を合成する大腸菌K12株、PR4019株、WecA-過剰発現株、及び0157株由来の膜画分を、UDP-[<sup>3</sup>H]GalNAcと一緒にインキュベートした。

30

#### 【0116】

(大腸菌膜の調製)－細菌細胞を、1,000×gで10分間の遠心分離により収集し、氷冷したリン酸-緩衝食塩水中で1回、冷水で1回、及び10mMトリス-HCl(pH7.4)、0.25Mショ糖で1回洗浄した。これらの細胞を、0.2mg/mlリゾチームを含有する10mMトリス-HCl(pH7.4)、0.25Mショ糖、10mM EDTAの中に、密度～200A<sub>600</sub>単位/mlとなるよう再懸濁し、30℃で30分間インキュベートした。細菌細胞を、1,000×gで10分間の遠心分離により回収し、40倍容量の氷冷した10mMトリス-HCl(pH7.4)中に迅速に再懸濁し、且つ氷上に放置した。10分後、細胞を、ぴったりはまったDounceホモジナイザーを用い15ストロークさせてホモジナイズし、0.1mMフッ化フェニルメチルスルホン及びショ糖を補充し、最終濃度を0.25Mとした。破壊されていない細胞を、1,000×gで10分間の遠心分離により除去し、細胞エンベロープを、40,000×gで20分間の遠心分離により回収した。膜画分を、10mMトリス-HCl(pH7.4)、0.25Mショ糖、1mM EDTA中に再懸濁し、且つ40,000×gで再度沈降させ、同じ緩衝液中に、タンパク質濃度～20mg/mlとなるよう再懸濁した。膜画分を、必要になるまで、-20℃で貯蔵した。

40

#### 【0117】

(インビトロ大腸菌膜における[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの生合成のアッセイ)－GlcNAc-P-P-Und及びGalNAc-P-P-Undの合成のための反応混合物は、総容積0.05ml中に50mMトリス-HCl(pH8)、40mM MgCl<sub>2</sub>、5mMジチオスレイトール、5mM 5'AMP、大腸菌膜画分(50～200 μg膜タンパク質)、及び5 μM UDP-[<sup>3</sup>H]GlcNAc/GalNAcのいずれか(500～25

50

00dpm/pmol)を含有した。37℃でインキュベーションした後、40倍容量のCHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)を添加することにより、反応を停止し、且つ[<sup>3</sup>H]HexNAc-P-P-ウンデカプレノールを含有する総脂質抽出物を、先に説明されたように調製した(Waechter, C. J., Kennedy, J. L.及びHarford, J. B.の文献、(1976) Arch. Biochem. Biophys. 174, 726-737)。分配後、有機相を、窒素流下で乾燥し、CHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1) 1ml中に再溶解し、且つアリコート(0.2ml)を取り出し、シンチレーションバイアル中で乾燥し、Packard社Tri-Carb 2100 TR液体シンチレーションスペクトロメーターにおいて液体シンチレーション分光法により放射能を分析した。[<sup>3</sup>H]GlcNAc-P-P-Und又は[<sup>3</sup>H]GalNAc-P-P-Undの合成の割合を決定するために、脂質抽出物を、窒素流下で乾燥し、少量のCHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)中に再溶解し、10cm×20cmのホウ酸含浸したBaker Si250シリカゲルプレート上にスポットし、且つこのプレートを、CHCl<sub>3</sub>、CH<sub>3</sub>OH、H<sub>2</sub>O、0.2Mホウ酸ナトリウム(65:25:2:2)で呈色した。個別の糖脂質を、Bioscan AR2000画像スキャン装置(Bioscan社、ワシントンD.C.)により検出した。各糖脂質の生合成の割合は、[<sup>3</sup>H]GlcNAc/GalNAc-P-P-Und中の放射能の総量に、個別の[<sup>3</sup>H]糖脂質の百分率を乗することにより、算出した。

#### 【0118】

様々な大腸菌株(K12、PR4019及びO157)由来の膜画分を、UDP-[<sup>3</sup>H]GlcNAc又はUDP-[<sup>3</sup>H]GalNAcのいずれかと一緒にインキュベートし、[<sup>3</sup>H]GlcNAc/GalNAc-P-P-Undへの取り込みを、前述のように決定した。表3に認められるように、非標識の糖脂質が、UDP-[<sup>3</sup>H]GalNAcとのインキュベーション後に検出され、膜画分をUDP-[<sup>3</sup>H]GlcNAcと一緒にインキュベートした場合には、GlcNAc-P-P-Undのみが検出可能であった。

#### 【表3】

表3

基質としてUDP-[<sup>3</sup>H]GlcNAc又はUDP-[<sup>3</sup>H]GalNAcのいずれかを使用する大腸菌膜画分における[<sup>3</sup>H]GlcNAc/GalNAc-P-P-ウンデカプレノールの合成

| 膜給源    | 添加された糖ヌクレオチド                | 形成された[ <sup>3</sup> H]糖脂質   |                             |
|--------|-----------------------------|-----------------------------|-----------------------------|
|        |                             | GlcNAc-P-P-Und<br>(pmol/mg) | GalNAc-P-P-Und<br>(pmol/mg) |
| K12    | UDP-[ <sup>3</sup> H]GlcNAc | 6.4                         | <0.01                       |
| K12    | UDP-[ <sup>3</sup> H]GalNAc | <0.01                       | <0.01                       |
| PR4019 | UDP-[ <sup>3</sup> H]GlcNAc | 44                          | <0.01                       |
| PR4019 | UDP-[ <sup>3</sup> H]GalNAc | <0.01                       | <0.01                       |
| O157   | UDP-[ <sup>3</sup> H]GlcNAc | 1.5                         | 0.5                         |
| O157   | UDP-[ <sup>3</sup> H]GalNAc | <0.01                       | <0.01                       |

#### 【0119】

更に、PR4019株、WecA-過剰発現株由来の膜とのインキュベーションへの外因性Und-Pの添加も、O157細胞由来の細胞質ゾル画分の添加も、UDP-GalNAcからのGalNAc-P-P-Undの形成を生じなかった。これらの結果は、UDP-GalNAcは、WecAにとっての基質ではないことを明らかにし、且つGalNAc-P-P-Undは、別の機序により形成されることを示唆している。

#### 【0120】

K12株由来の膜を、UDP-[<sup>3</sup>H]GlcNAcと一緒にインキュベートする場合、予想されたように[<sup>3</sup>H]GlcNAc-P-P-Undが合成された(Rush, J. S., Rick, P. D.及びWaechter, C. J.の文献、(1997) Glycobiology, 7, 315-322)。しかし、O157株由来の膜を、[<sup>3</sup>H]GlcNAc-P-P-Undに加え、UDP-[<sup>3</sup>H]GlcNAcと一緒にインキュベートする場合、[<sup>3</sup>H]GalNAc-P-P-Undであることが示された第二の標識された脂質(下記参照)が認められた。これら2種の糖脂質の形成の時間経過を試験する場合、放射能の[<sup>3</sup>H]GlcNAc-P-P-Undへの取り込み(図1、○)は、

より迅速に生じ、且つ前駆体-生成物の関係が同等である(図2)[<sup>3</sup>H]GalNAc-P-P-Undへの取り込み(図1、●)よりもより高度であった。

#### 【0121】

大腸菌0157膜は、GalNAc-P-P-Undの合成のためのGalNAc-P供与体としてUDP-GalNAcを利用しないという知見は、図2に図示されたGalNAc-P-P-Undの形成に関する生合成経路を確認する一例である。このスキームにおいて、GlcNAc-P-P-Undは、WecAにより触媒された、UDP-GlcNAcからのGlcNAc-Pの転移により形成され、その後GlcNAc-P-P-Undは、これまで未知の4-エピメラーゼの作用によりエピマー化され、GalNAc-P-P-Undを生成する。

#### 【0122】

(実施例3：大腸菌0157株由来の膜画分によりインビトロにおいて形成された[<sup>3</sup>H]GalNAc-P-P-Undの特徴決定) 10

GalNAc-P-P-Undとして図1において検出された追加の0157-特異的糖脂質生成物と同じく、これは弱アルカリメタノリシス(トルエン/メタノール1:3、0.1N KOH含有、0℃、60分間)に対して安定であり、CHCl<sub>3</sub>/CH<sub>3</sub>OH/H<sub>2</sub>O(10:10:3)中に平衡とされたDEAE-セルロースに保持され、且つ[<sup>3</sup>H]GlcNAc<sub>1-2</sub>-P-P-Dolについて先に報告されたように(Waechter, C. J.及びHarford, J. B.の文献、(1977) Arch. Biochem. Biophys. 181, 185-198)、20mM酢酸アンモニウムを含有するCHCl<sub>3</sub>/CH<sub>3</sub>OH/H<sub>2</sub>O(10:10:3)で溶離された。

#### 【0123】

[<sup>3</sup>H]GalNAc-P-P-Undは、ホウ酸含浸シリカゲルG上の薄層クロマトグラフィーにより(Kean, E. L.の文献、(1966) J. Lipid Res. 7, 449-452)、[<sup>3</sup>H]GlcNAc-P-P-Undから明確に 20  
分離され、且つ図3A及び図3Bに示されたように、分取TLCにより精製された。

#### 【0124】

(ホウ酸含浸薄層プレート及びワットマン濾紙No.1の調製)－シリカゲル薄層プレートは、Keanの文献(Kean, E. L.、(1966) J. Lipid Res. 7, 449-452)に説明されたように、これらのプレートを95%メタノールを溶媒とする2.5%Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub>・10H<sub>2</sub>O中に短時間含浸することにより、ホウ酸ナトリウムで含浸した。ホウ酸含浸TLCプレートを、室温で一晩乾燥し、使用時までDrierite上の真空デシケーター内で貯蔵した。クロマトグラフィーの直前に、これらのプレートを、100℃に短時間(～10-15分間)加熱することにより活性化した。ワットマン濾紙No.1の20cm×30cmシートを0.2MのNa<sub>2</sub>B<sub>4</sub>O<sub>7</sub>・10H<sub>2</sub>O中に浸漬することにより、ワットマン濾紙No.1をホウ酸ナトリウムで含浸した。Cardini及びLeloirの文献(Cardini, C. E.及びLeloir, L. F.、(1957) J. Biol. Chem. 225, 317-324)に記載されているように、ワットマン濾紙No.1シートを、ワットマン濾紙No.3MMの2枚のシートの間でしっかり 30  
圧迫し、室温で数日間かけて乾燥させた。

#### 【0125】

(インビトロ反応において形成されたグリカン生成物の特徴決定)－個別の糖脂質([<sup>3</sup>H]GalNAc-P-P-Und及び[<sup>3</sup>H]GlcNAc-P-P-Und)のグリカンを、弱酸加水分解による放出後に、下降ペーパークロマトグラフィーにより特徴決定した。GlcNAc/GalNAc脂質は、コニカルスクリュウキャップ付きチューブ内で窒素流下で乾燥し、0.01M HClの0.2ml中で、100℃で15分間加熱した。加水分解後、試料を、AG50WX8(H<sup>+</sup>)の0.4ml及びAG1X8(酢酸塩型)0.4mlを含む、0.8mlの混合床型イオン交換カラムに適用し、且つ水1.5mlで溶離した。溶離液を、 40  
窒素流れ下で乾燥し、少量のH<sub>2</sub>O(0.02ml)中に再溶解し、ホウ酸含浸ワットマン濾紙No.1の30cm細長片上にスポットし、且つ40～50時間かけて、ブタノール/ピリジン/水(6:4:3)により下降モードで展開した。乾燥後、濾紙細長片を、1cmの带状に切断し、シンチレーションスペクトロメーターにより放射能を分析した。GlcNAc及びGalNAc標準を、アニリン-ジフェニルアミン浸漬試薬を用いて検出した(Schwimmer, S.及びBenvenue, A.の文献、(1956) Science, 123, 543-544)。

#### 【0126】

グリカン生成物は、先に説明されたような、弱酸加水分解後の、0.1M NaOHを溶媒とする0.1M NaBH<sub>4</sub>(最終容積0.1ml)による還元により、それらの対応するアルジトールへ変換された。室温で一晩インキュベーションした後、氷酢酸数滴により反応をクエンチし、且 50

つ窒素流下で、酢酸1滴を含有するメタノールで数回乾燥した。アルジトールを水に溶解し、AG50WX8(H<sup>+</sup>)及びAG1X8(酢酸塩)の0.5mlカラム上の通過により脱塩し、窒素下で乾燥し、且つワットマン濾紙No.3MMの30cm細長片上にスポットした。ワットマンNo.3MM細長片を、酢酸エチル、ピリジン、0.1Mホウ酸(65:25:20)により下降モードで一晩展開し、乾燥し、1cmの带状に切断し、シンチレーション分光法により放射能を分析した。GlcNAcitol及びGalNAcitol標準を、過ヨウ素酸-ベンジジン浸漬手法(Gordon, H. T., Thornburg, W. 及びWerum, L. N.の文献、(1956) Anal. Chem. 28, 849-855)の変法を用いて可視化した。これらの濾紙細長片を、アセトン、0.1M NaIO<sub>4</sub>(95:5)中に浸漬し、3分間風乾させ、その後アセトン/酢酸/H<sub>2</sub>O/o-トリジン(96:0.6:4.4:0.2g)中に浸漬した。cis-ジオールを含有するアルジトールは、青色のバックグラウンド上に黄色スポットとして染まった。

10

#### 【0127】

(糖脂質の質量分析("MS"))—精製された糖脂質を、ABI Turbo Vエレクトロスプレーイオン給源を備えたABI/MDS Sciex 4000 Q-Trapハイブリッド三連四重極直列式イオントラップ型質量分析計を用いて分析した(ABI/MDS-Sciex社、トロント、カナダ)。簡単に述べると、試料を、経験的に決定されたイオン源設定で10 µl/分で注入し、且つMS/MS(二次での質量分光法)情報を、直列式イオントラップ様式での分子イオンのフラグメンテーションにより得た。

#### 【0128】

糖脂質を、弱酸(0.01N HCl、100℃、15分間)で処理する場合、水溶性生成物を、ホウ酸含浸したワットマン濾紙No.1による下降ペーパークロマトグラフィー上での [<sup>3</sup>H]GalNAcとの同時クロマトグラフィーにかけた(図3C)。加えて、標識された糖が還元された場合、これは [<sup>3</sup>H]アルジトール、GalNAc-OHへ変換された(図3D)。更に、陰イオンMS分析は、m/z=128の[M-H]<sup>-</sup>イオンを生じ、これはGalNAc-P-P-Undと予想され、且つMS/MS娘イオンスペクトルは、m/z=907に突出したイオンを示し、これはP-P-Undを含む糖脂質と予想された(Guan, Z., Breazeale, S. D.及びRaetz, C. R.の文献、(2005) Anal. Biochem. 345, 336-339)。GalNAc-P-P-Undとして0157株により形成された糖脂質生成物の同定は、外因性GlcNAc-P-P-Undからのその形成によっても裏付けられる(実施例7参照)。

20

#### 【0129】

(実施例4: Z3206遺伝子を発現している大腸菌細胞における [<sup>3</sup>H]GlcNAcによる [<sup>3</sup>H]GalNAc-P-P-Und(インビボ)の代謝標識)

30

大腸菌0157 Z3206遺伝子の発現は、細胞にGalNAc-P-P-Und合成を可能にするかどうかを調べるために、Z3206遺伝子を発現している大腸菌株21546(Meier-Dieter, U., Starman, R., Barr, K., Mayer, H.及びRick, P. D.の文献、(1990) J. Biol. Chem., 265, 13490-13497)を、 [<sup>3</sup>H]GlcNAcで代謝的に標識し、且つ [<sup>3</sup>H]GlcNAc/GalNAc-P-P-Und形成について分析した。

#### 【0130】

(細菌細胞の代謝標識)—大腸菌細胞を、Luria-Bertani培地において、37℃で激しく振盪しながら、A<sub>600</sub>が0.5~1となるまで培養した。 [<sup>3</sup>H]GlcNAcを、最終濃度1 µCi/mlとなるよう添加し、インキュベーションを37℃で5分間継続した。放射標識の糖脂質への取り込みは、破碎した氷0.5gm/mlの添加により終結させ、培養物を完全に混合した。細菌細胞を、4000×gで10分間の遠心分離により回収し、上清を廃棄した。これらの細胞を、氷冷したリン酸-緩衝食塩水で2回洗浄し、(細胞ペレットの)10倍容量のメタノール中で激しくボルテックス混合することにより再懸濁し、プローブソニケーターにより全出力の40%で短時間超音波処理した。超音波処理後、20倍容量のクロロホルムを添加し、この抽出物を激しく混合し、且つ室温で15分間静置させた。不溶性物質を、遠心分離により沈降させ、ペレットを、少量のCHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)により2回再抽出した。その後一緒にした有機抽出液を、以下に説明したように処理した。

40

#### 【0131】

(GlcNAc-P-P-Und及びGalNAc-P-P-Undの精製)—GlcNAc/GalNAc-P-P-Undを、CHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)により抽出し、別所に記載したように分配することにより水溶性物質を含まないよ

50

うにした(Waechter, C. J., Kennedy, J. L.及びHarford, J. B.の文献、(1976) Arch. Biochem. Biophys. 174, 726-737)。その後有機抽出液を、窒素流下で乾燥し、0.1N KOHを含有するトルエン/メタノール(1:3)中の0℃で60分間の脱アシル化により、バルクグリセロリン脂質を破壊した。この脱アシル化反応を、酢酸で中和し、4倍容量のCHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)により希釈し、且つ1/5倍容量の0.9%NaClにより洗浄した。有機(下側)相を、1/3倍容量のCHCl<sub>3</sub>、CH<sub>3</sub>OH、0.9%NaCl(3:48:47)で洗浄し、水相を廃棄した。有機相を、この有機相中に残留する水相を収容するのに十分量のメタノールで希釈し、CHCl<sub>3</sub>/CH<sub>3</sub>OH(2:1)により平衡としたDEAE-セルロースカラム(5ml)に装加した。このカラムを、20倍カラム容量のCHCl<sub>3</sub>/CH<sub>3</sub>OH/H<sub>2</sub>O(10:10:3)で洗浄し、その後20mM酢酸アンモニウムを含有するCHCl<sub>3</sub>/CH<sub>3</sub>OH/H<sub>2</sub>O(10:10:3)で溶離した。画分(2ml)を収集し、ホウ酸含浸シリカプレート上の薄層クロマトグラフィー(前記)による分離後、放射能、又はアニスアルデヒド噴霧試薬を使用するGlcNAc/GalNAc-P-P-Und(Dunphy, P. J., Kerr, J. D., Pennock, J. F., Whittle, K. J., 及びFeeney, J.の文献、(1967) Biochim. Biophys. Acta 136, 136-147)のいずれかについてモニタリングした。

10

### 【0132】

UDP-ManNAcA合成の突然変異は、腸内細菌共通抗原の合成のためのGlcNAc-P-P-Undの利用の阻止を招くので、大腸菌21546株を、Z3206発現試験のための宿主として選択した。大腸菌21546は、大腸菌K12から誘導されるので、同様にこれもO-抗原反復を合成せず(Stevenson, G., Neal, B., Liu, D., Hobbs, M., Packer, N. H., Batley, M., Redmond, J. W., Lindquist, L.及びReeves, P.の文献、(1994) J. Bacteriol., 176, 4144-4156)、その結果より大量のGlcNAc-P-P-Undが、GalNAc-P-P-Undへの変換のために蓄積する。菌株21546及びZ3206遺伝子を発現する形質転換体を、[<sup>3</sup>H]GlcNAcで標識し、これらの放射標識された脂質を、ホウ酸含浸シリカゲルプレート上の薄層クロマトグラフィーにより分析し、親株(図4A)は、わずかに1種の標識された脂質GlcNAc-P-P-Undを合成した。しかしZ3206遺伝子を発現している21546細胞は(図4B)、追加の標識された脂質も合成し、これはGalNAc-P-P-Undであることが示された。

20

### 【0133】

(実施例5: Z3206遺伝子を発現している大腸菌細胞由来の膜画分はインビトロにおいてGalNAc-P-P-Undを合成する)

大腸菌0157のZ3206遺伝子によりコードされたタンパク質が、GalNAc-P-P-Undの合成を触媒したことを裏付けるために、Z3206遺伝子を発現している大腸菌細胞由来の膜画分を、[<sup>3</sup>H]UDP-GlcNAcと一緒にインキュベートし、且つ[<sup>3</sup>H]糖脂質生成物を、図5に示したようにホウ酸含浸シリカゲルプレート上で薄層クロマトグラフィーにより分析した(クロマトグラフィー調製及び特徴決定の方法は実施例3に説明されている)。大腸菌K12又は宿主株大腸菌21546細胞由来の膜画分を、UDP-[<sup>3</sup>H]GlcNAcと一緒にインキュベートした場合、[<sup>3</sup>H]GlcNAc-P-P-Undのみ認められた(図5A及び図5C)。しかし、Z3206を発現している大腸菌0157及び大腸菌21546由来の膜画分は、GalNAc-P-P-Undも形成した(図5B及び図5D)。

30

### 【0134】

(実施例6: GlcNAc-P-P-Undの形成はUMPの存在下で逆転されるがGalNAc-P-P-Undは逆転されない)

40

GalNAc-P-P-UndはGlcNAc-P-P-Undから合成され、且つグリコシル供与体としてUDP-GalNAcを使用するWecAの作用によっては合成されないことの更なる証拠を提供するために、UMPによる内在性の予備標識された[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの放出の作用を試験した。WecAにより触媒されるGlcNAc-ホスホトランスフェラーゼ反応は、過剰なUMPの添加により自在に可逆性となり、UDP-GlcNAcを再合成し、且つUnd-Pを放出する。

### 【0135】

この実験において、Z3206を発現している大腸菌21546株由来の膜画分を、UDP-[<sup>3</sup>H]GlcNAcで10分間予備標識し、引き続き1mM UMPを添加し、残存する各標識された糖脂質の量を測定した。図6Aに図示された結果は、10分間の標識期間の終了時の、[<sup>3</sup>H]GlcNAc-P-P-Und及び[<sup>3</sup>H]GalNAc-P-P-Undの相対量を示している。1mM UMPと一緒に1分間のインキュベシ

50

オン後、 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ のかなりの喪失が認められたのに対し、 $[^3\text{H}]\text{GalNAc-P-P-Und}$ のピークは比較的变化がなかった(図6B)(クロマトグラフィー調製及び特徴決定の方法は実施例5に説明されている)。この知見は、表3の結果と一致しており、WecAは、UDP-GalNAcからGalNAc-P-P-UndへのGalNAc-Pの転移を触媒しないことを指摘している。UMPとの2分間のインキュベーション期間中(図6C)に、GlcNAc-P-P-Undの喪失は遅くなり、 $[^3\text{H}]\text{GalNAc-P-P-Und}$ のピークのわずかな低下が存在することは注目に値し、これは $[^3\text{H}]\text{GalNAc-P-P-Und}$ は、エピメラーゼ反応の逆転により $[^3\text{H}]\text{GlcNAc-P-P-Und}$ プールと再度平衡となることを示唆している(実施例7参照)。

#### 【0136】

(実施例7：Z3206を発現している大腸菌細胞由来の膜により触媒された外因性の精製された $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 及び $[^3\text{H}]\text{GalNAc-P-P-Und}$ の相互変換)

10

GlcNAc-P-P-Und及びGalNAc-P-P-Undが、Z3260を発現している大腸菌細胞由来の膜画分により直接相互変換され得ることの直接の証拠を提供するために、精製された $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 及び $[^3\text{H}]\text{GalNAc-P-P-Und}$ を、外因性基質として試験した。

#### 【0137】

精製された $[^3\text{H}]\text{GlcNAc-P-P-Und}/[^3\text{H}]\text{GalNAc-P-P-Und}$ は、実施例4(細菌細胞の代謝標識並びにGlcNAc-P-P-Und及びGalNAc-P-P-Undの精製)のように調製した。 $[^3\text{H}]\text{HexNAc-P-P-Und}$ デカプレノール(2000 dpm/pmol、1% Triton X-100中に分散された、最終濃度0.1%)を、実施例2のインビトロにおける大腸菌膜中の $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 及び $[^3\text{H}]\text{GalNAc-P-P-Und}$ 合成のアッセイのように大腸菌膜と一緒にインキュベートした。

20

#### 【0138】

予備実験からは、エピメラーゼは、外因性 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ を、Triton X-100、CHAPS(3-[(3-コラミドプロピル)ジメチルアンモニオ]-1-プロパンスルホン酸)、Nonidet P-40、又はオクチルグルコシド中に分散された反応混合物に添加した場合に、活性化され、且つpH7~8.5の範囲に最適pHを示したことが認められた。膜画分と一緒にインキュベーション前の精製された $[^3\text{H}]\text{GlcNAc-P-P-Und}$ 及び $[^3\text{H}]\text{GalNAc-P-P-Und}$ のクロマトグラフィー移動度を、図7A及び図7Dに示している。図7B及び図7Eに示されるように、糖脂質は、大腸菌21546由来の膜画分と一緒にインキュベーションにより影響を受けなかった。しかし、精製された糖脂質のZ3206を発現している大腸菌21546由来の膜画分と一緒にインキュベーションは、外因性 $[^3\text{H}]\text{GlcNAc-P-P-Und}$ の $[^3\text{H}]\text{GalNAc-P-P-Und}$ への変換(図7C)及び $[^3\text{H}]\text{GalNAc-P-P-Und}$ の $[^3\text{H}]\text{GlcNAc-P-P-Und}$ への変換(図7F)を触媒した。これらの結果は、GlcNAc-P-P-Und及びGalNAc-P-P-Undは、Z3206を発現している大腸菌株において酵素的に相互変換することができることを直接明らかにする。

30

#### 【0139】

(実施例8：大腸菌Z3206はUDP-GlcNAc 4-エピメラーゼではない)

Z3206はUDP-GalNAcの形成を触媒することができるかどうかを決定するために、C. ジェジュニ由来のN-グリコシル化装置を、大腸菌において発現した。このリポーターシステムにおいて、標的タンパク質AcrAのグリコシル化は、機能的Gne UDP-Glc/UDP-GlcNAcエピメラーゼ(Bernatchez, S., Szymanski, C.M., Ishiyama, N., Li, J., Jarrell, H.C., Lau, P.C., Berghuis, A.M., Young, N.M., Wakarchuk, W.W.の文献、(2005) J. Biol. Chem. 280, 4792-4802)を含む、pgl遺伝子座の存在に左右された(Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M.の文献、(2002) Science, 298, 1790-1793)。pglクラスターがgneの欠失を含む場合に、AcrAのグリコシル化が失われた(Linton, D., Dorrell, N., Hitchen, P.G., Amber, S., Karlyshev, A.V., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M.及びWrenn, B.W.の文献、(2005) Mol Microbiol. 55, 1695-1703)。pglオペロンΔgneの存在下でAcrA-グリコシル化を回復するZ3206の能力は、Gne(pMLBAD:gne)又はZ3206(pMLBAD:Z3206)のいずれかにより補完されたpgl遺伝子座ΔgneによりAcrA(pWA2)を発現することで、インビボにおいて調べた。

40

#### 【0140】

50

総大腸菌細胞抽出物を、SDS負荷緩衝液100  $\mu$ l中に再懸濁された1 A<sub>600</sub>単位と等しい濃度で細胞を用いる免疫検出分析のために調製した(Laemmli, U.の文献、(1970) Nature, 27, 680-685)。10  $\mu$ lのアリコートをし、10%SDS-PAGEの上に装加した。大腸菌細胞のペリプラズム抽出物を、リゾチーム処理により調製し(Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., Aebi, M.の文献、(2005) Proc Natl Acad Sci USA, 102, 3016-3021)、且つ最終試料10  $\mu$ l(細胞の0.2 A<sub>600</sub>単位に相当)をSDS-PAGEにより分析した。ニトロセルロースメンブレンにブロットした後、試料を、特異的抗血清により免疫染色した(Aebi, M., Gassenhuber, J., Domdey, H., 及びte Heesen, S.の文献、(1996) Glycobiology, 6, 439-444)。抗-AcrA抗体(Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S. M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wrenn, B.W., Aebi, M.の文献、(2002) Science, 298, 1790-1793)を使用した。抗-ウサギIgG-HRP(Bio-Rad社)を、二次抗体として使用した。検出は、ECL(商標)ウェスタンブロッティング検出試薬(Amersham Biosciences社)により実行した。

10

#### 【0141】

図8に示されたように、グリコシル化されたタンパク質は、グリコシル化されない形よりもより遅く移動し、これは、pgl遺伝子座Δgneを発現している細胞がGneにより補完された場合(レーン2)にのみ、形成された。Z3206は、リポーター糖タンパク質のグリコシル化を回復することができなかった(図8、レーン1)。従って、Z3206は、Gne依存性グリコシル化システムにおいてAcrAのグリコシル化を補完しない。Gne及び膜-会合されたZ3206の発現は、免疫検出により確認された。

20

#### 【0142】

(実施例9：シゲラ・フレックスネリ6+/−Z3206 LPSの分析)

図9において、シゲラ・フレックスネリ6のO-抗原の生合成に必要ないくつかの遺伝子；ヌクレオチド糖前駆体の生合成のための酵素をコードしている遺伝子；グリコシルトランスフェラーゼをコードしている遺伝子；O抗原プロセッシングタンパク質をコードしている遺伝子；及び、O-アセチル化に寄与するタンパク質をコードしている遺伝子：が示された。O抗原の構造は、Dmitriev, B.A.らにより解明されている(Dmitriev, B.A.らの文献、「赤痢菌の菌体抗原(Somatic Antigens of Shigella)」、Eur J Biochem, 1979, 98: p.8；Liu Bらの文献、「赤痢菌O抗原の構造及び遺伝子(Structure and genetics of Shigella O antigens)」、FEMS Microbiology Review, 2008, 32: p.27)。

30

#### 【0143】

シゲラ・フレックスネリ6のO-抗原の生合成に必要な全ての遺伝子を確定するために、ゲノムライブラリーを作製した。

#### 【0144】

(シゲラ・フレックスネリ6ゲノムDNAのクローニング)ー細菌からのDNA単離に関するプロトコールに従い、シゲラ・フレックスネリ6ゲノムDNAを、Macherey-Nagel NucleoSpin(登録商標)組織キットを用いて単離した。DNAは、5つのシゲラ・フレックスネリ6の各2mlの一晩培養物から単離し、最終溶離は、溶離緩衝液(5mMトリス/HCl, pH8.5)100  $\mu$ lで行った。溶離された画分をプールし、イソプロパノールにより沈殿し、且つ最終ペレットを、TE緩衝液52  $\mu$ l中に再懸濁し、その総容量に、CopyControl商標Fosmidライブラリー作製キット(EPICENTRE)により供されたプロトコールに従い、末端修復(end-repair)を施した。このキットのプロトコールに記載されているように、末端修復されたDNAは、1×TAE緩衝液で泳動する1%低融点アガロースゲル上で精製し、回収し、エタノールで沈殿させた。沈殿したDNAの再懸濁は、TE緩衝液7  $\mu$ l中で行い、そのDNA 0.15  $\mu$ lは、EPICENTREプロトコールに従いpCC1FOS(配列番号:27)にライゲーションさせた。このライゲーション生成物のファージへのパッケージングは、プロトコールに従い行い、且つパッケージされたファージは、ファージ希釈用緩衝液において1:1で希釈し、その10  $\mu$ lを、EPICENTREに記載されたように予め増殖したEPI300-T1細胞100  $\mu$ lの感染に使用した。細胞(110  $\mu$ l)を、1プレート当たりおよそ100コロニーで、6回播種し、その結果6つのプレートは、シゲラ・フレ

40

50

ックスネリ6ゲノムライブラリー全体を含んだ。プレートを、コロニープロティングにより呈色し、且つ陽性／陰性コロニーを、ウェスタンブロット及び銀染色した。

#### 【0145】

(コロニープロティング)－コロニーブロットのために、ニトロセルロースメンブレンを、固形寒天プレート上に搭載し、取り外し、1×PBSTで3回洗浄し、同じ様式で処理した。このメンブレンを、最初に10%ミルクにより室温で1時間ブロックし、その後これを1%ミルク(PBST中)2ml中で、抗-VI型抗血清(一次抗体)と一緒に、室温で1時間インキュベーションした。PBST中で3回、各回10分間洗浄した後、メンブレンを、1%ミルク(PBST中)2ml中の二次抗体1:20000ペルオキシダーゼ複合したヤギ-抗-ウサギIgG(BioRad社)と、室温で更に1時間インキュベーションした。最後にPBSTで3回洗浄(各回10分間)した後、メンブレンを、SuperSignal(登録商標)West Dura Extended Duration Substrate Kit(Thermo Scientific社)により提供されるルミノールと過酸化水素緩衝液の1:1混液により、UVP Chemi Doc Imaging Systemにおいて呈色した。

10

#### 【0146】

S.フレックスネリ6ゲノムライブラリー生成後のS.フレックスネリ6抗血清と反応するクローンを、rmlBからwfbZに達するLiuらにより先に配列決定された領域(Liuらの文献、2008)のプライマーウォーキングにより配列決定した(図9)。プライマーrmlB\_rev及びwfbZ\_fw d(S.フレックスネリ-Z3206)は、rmlB及びwfbZとアニーリングさせ、且つそれぞれwcaM及びhisI/Fに達するまで(S.フレックスネリ+Z3206)の、クローンの挿入断片の配列決定に使用した(図10)。

20

#### 【0147】

O抗原合成は、Z3206を欠いている(従って、und-GlcNAcのund-GalNAcへのエピマー化を妨害する)クローンにおいて維持されるかどうかを確立するために、2種のプラスミドを作製し(配列番号:28及び配列番号:29)(図10)、大腸菌細胞に形質転換し、且つ銀染色及びウェスタンブロットにより分析した。

#### 【0148】

図11に示されるように、LPSは、大腸菌細胞+Z3206又は-Z3206において生成された。O抗原は、Z3206がなくとも生成され得るが、生成収量は低く、このことはエピメラーゼ(Z3206)を伴わない多糖生成の効率はより低いことを指摘している。

#### 【0149】

(実施例10: S.フレックスネリ6+/−Z3206 LL0の分析)

(C18カラムクロマトグラフィーによるウンデカプレノール-PP-0抗原の精製)－S.フレックスネリ抗原+/−Z3206を発現している大腸菌細胞を、ペレット化し、0.9%NaCl 50mlで1回洗浄し、最終ペレットを、一晩凍結乾燥した。ペレットを、85~95%メタノール30mlで1回洗浄し、10:10:3のクロロホルム-メタノール-水(v/v/v)で再抽出し、且つこの抽出物を、水の添加により2相ブライ/ダイアースシステムに変換し、最終比10:10:9(C:M:W)とした。相を遠心分離により分離し、上側水相を、メタノール10mlで馴化させ、且つ3:48:47(C:M:W) 10mlで平衡としたC18 Sep-Pakカートリッジ上に各々添加した。添加後、カートリッジを、3:48:47(C:M:W)10mlで洗浄し、10:10:3(C:M:W)5mlで溶離した。これらの添加物の20 OD試料を、流通させ(flow-through)、洗浄し、且つC18カラムの溶離物をEppendorf Concentrator Plusにおいて乾燥し、メタノール250 µlで洗浄し、再蒸発させ、ddH<sub>2</sub>O 30 µlで更に3回洗浄した。

40

#### 【0150】

(糖脂質加水分解)－前記C18カラム洗浄液からの糖脂質試料を、乾燥試料をn-プロパノール:2Mトリフルオロ酢酸(1:1)2ml中に溶解し、50°Cで15分間加熱し、N<sub>2</sub>下で蒸発乾固することにより、加水分解した。

#### 【0151】

(2-アミノ安息香酸によるオリゴ糖標識及びHPLC)－標識はBiggeらの方法(Biggeの文献、1995)に従い実行し、且つグリカンクリーンアップを、Merryら(2002)により説明されたペーパーディスク法(Merryらの文献、2002)を用いて行った。2-アミノ安息香酸(2-AB)標識

50

されたグリカンの分離を、Royleらの方法に従う(Royleの文献、2002)が3溶媒システムに変更したGlycoSep-N順相カラムを用いるHPLCにより行った。溶媒Aは、80%アセトニトリル中の10mMギ酸アンモニウム(pH4.4)であった。溶媒Bは、40%アセトニトリル中の30mMギ酸アンモニウム(pH4.4)であった。溶媒Cは、0.5%ギ酸であった。カラム温度は30℃であり、且つ2-AB標識されたグリカンは蛍光により検出した( $\lambda_{ex}=330nm$ 、 $\lambda_{em}=420nm$ )。勾配条件は、流量0.4ml/分で160分間かけ100%Aから100%Bへの直線勾配、それに続く流量を1ml/分に増し、2分間かけ100%Bから100%Cへであった。カラムを、100%Cで5分間洗浄し、2分間かけて100%Aに戻し、100%Aを流量1ml/分で15分間流し、その後流量0.4ml/分で5分間に戻した。全ての試料は、水中に注入した。

**【0152】**

10

Z3206を持つ(配列番号:29)又は持たない(配列番号:28)S.フレックスネリ0抗原を発現するプラスミドを、SCM3細胞に形質転換した(図10)。遅い溶離容量での追跡は、S.フレックスネリ0抗原+/-Z3206を含む2つの試料の曲線の間の差を示している(図12)。この溶離パターンの差は、エピメラーゼ(Z3206)の存在に応じてGlcNAc又はGalNAcという還元末端に異なる単糖を保持する異なるオリゴ糖構造により説明することができる。

**【0153】**

(実施例11：S.フレックスネリ6+/-Z3206から生成されたバイオコンジュゲートの生成及び特徴決定によるpglB特異性の分析)

PglBは、還元末端にGlcNAc(S.フレックスネリ6の0-抗原)を有するオリゴ糖をキャリアタンパク質EPAに転移するかどうかを評価するために、EPA(配列番号:25)、PglB(配列番号:26)及びS.フレックスネリ6の0-抗原+/-Z3206(配列番号:29/配列番号:28)を発現している大腸菌細胞由来のニッケル精製した抽出物を、抗EPA抗体及び抗VI型抗体を用いるウェスタンブロットにより分析した。還元末端にGalNAcを有する及び有さないS.フレックスネリ6の0抗原を、PglBにより、EPAへ転移し、抗EPA及び抗VI抗血清により検出した(図13)。

20

**【0154】**

0抗原は、依然生成され且つ検出されたが、生成収量はより低く、このことはエピメラーゼを伴わない多糖生成の効率がより低いことを示している。

**【0155】**

本発明は、それらの実施態様を参照し具体的に示され且つ説明されているが、当業者は、請求項に包含された本発明の範囲を逸脱しない限りは、それらの形態及び詳細における様々な変更を行うことができることを理解するであろう。本発明の範囲内に収まるとして当業者に理解されるそのような様々な変更は、特に、還元末端にGalNAcを持つ大腸菌及びシゲラ・フレックスネリ由来のもの以外のグリカンを含むN-グリコシル化タンパク質及びバイオコンジュゲートを含んでいる。

30

**【0156】**

(配列表)

出願人：GlycoVaxyn AG

**【0157】**

表題：原核細胞における免疫原性多糖体生成の生合成システム

**【0158】**

40

配列番号の数：29

**【0159】**

配列番号:1—大腸菌0157 Z3206のヌクレオチド配列

**【0160】**

長さ：993

**【0161】**

種類：DNA

**【0162】**

生物：大腸菌0157

**【0163】**

50

配列：

【化 4】

ATGAACGATAACGTTTTGCTCATAGGAGCTTCCGGATTCGTAGGAACCCGACTACTTGAAACGG  
CAATTGCTGACTTTAATATCAAGAACCTGGACAAACAGCAGAGCCACTTTTATCCAGAAATCAC  
ACAGATTGGCGATGTTTCGCGATCAACAGGCACTCGACCAGGCGTTAGTCGGTTTTGACACTGTT  
GTACTACTGGCAGCGGAACACCGCGATGACGTCAGCCCTACTTCTCTCTATTATGATGTCAACG  
TTCAGGGTACCCGCAATGTGCTGGCGGCCATGGAAAAAATGGCGTTAAAAATATCATCTTTAC  
CAGTTCGGTTGCTGTTTATGGTTTGAACAAACACAACCCTGACGAAAACCATCCACACGACCCT  
TTCAACCACTACGGCAAAAGTAAGTGGCAGGCAGAGGAAGTGCTGCGTGAATGGTATAACAAAG  
CACCAACAGAACGTTTCAATTAACCATCATCCGTCCTACCGTTATCTTCGGTGAACGCAACCGCGG  
TAACGTCTATAACTTGCTGAAACAGATCGCTGGCGGCAAGTTTATGATGGTGGGCGCAGGGACT  
AACTATAAGTCCATGGCTTATGTTGGAACATTGTTGAGTTTATCAAGTACAACTGAAGAATG  
TTGCCGCAGGTTATGAGGTTTATAACTACGTTGATAAGCCAGACCTGAACATGAACCAGTTGGT  
TGCTGAAGTTGAACAAAGCCTGAACAAAAAGATCCCTTCTATGCACTTGCCCTACCCACTAGGA  
ATGCTGGGTGGATATTGCTTTGATATCCTGAGCAAAATTACGGGCAAAAAATACGCTGTCAGCT  
CAGTGCGCGTGAAAAAATTCTGCGCAACAACACAGTTTGACGCAACGAAAGTGCATTCTTCAGG  
TTTTGTGGCACCGTATACGCTGTCGCAAGGTCTGGATCGAACACTGCAGTATGAATTCGTTTCAT  
GCCAAAAAAGACGACATAACGTTTGTCTCTGAG

10

20

【0 1 6 4】

配列番号:2-Z3206のアミノ酸配列

【0 1 6 5】

長さ: 331

【0 1 6 6】

種類: PRT

30

【0 1 6 7】

生物: 大腸菌0157

【0 1 6 8】

配列:

【化 5】

MNDNVLLIGASGFVGRLLLETAIADFNKLNLDKQQSHFYPEITQIGDVRDQQALDQALV  
GFDTVVLLAAEHRDDVSPTSLLYYDVNVQGTRNVLAAMEKNGVKNIIFTSSVAVYGLNKHNPDEN  
HPHDPFNHYGKSKWQAEVLREWYNKAPTERSLTIIRPTVIFGERNRGNVYNLLKQIAGGKFMM  
VGAGTNYKSMAYVGNIVEFIKYKLKNVAAGYEVYNYVDKPDLMNQLVAEVEQSLNKKIPSMHL  
PYPLGMLGGYCFDILSKITGKKYAVSSVRVKKFCATTQFDATKVHSSGFVAPYTLSQGLDRTLQ  
YEFVHAKKDDITFVSE

40

【0 1 6 9】

配列番号:3-大腸菌055 gneのヌクレオチド配列

【0 1 7 0】

遺伝子座 AF461121\_1 BCT 02-MAY-2002

【0 1 7 1】

定義 (UDP-GlcNAc 4-エピメラーゼGne[大腸菌])

50

【0172】

寄託番号 AAL67550

【0173】

長さ: 993

【0174】

種類: DNA

【0175】

生物: 大腸菌055

【0176】

配列:

10

【化6】

```
ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCG TAGGAACCCG
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC
AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTGCTACT
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC
GTTAAAAATA TCATCTTTAC CAGTTCCGTT GCTGTTTATG GTTTGAACAA
ACACAACCTT GACGAAAACC ATCCACACGA TCCTTTCAAC CACTACGGCA
AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA
CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA
ACGGAACCGC GGTAACGTCT ATAAC TTGCT GAAACAGATC GCTGGCGGCA
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT
GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG
TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCAGT
TG GTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA
ATTTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG
```

20

30

【0177】

配列番号: 4-大腸菌055 UDP-GlcNAc 4-エピメラーゼGneのアミノ酸配列

40

【0178】

遺伝子座 AF461121\_1

【0179】

定義 (UDP-GlcNAc 4-エピメラーゼGne[大腸菌])

【0180】

寄託番号 AAL67550

【0181】

長さ: 331個のアミノ酸、線状

【0182】

50

種類：PRT

【0183】

生物：大腸菌055

【0184】

配列：

【化7】

mndnvlliga sgfvgrlle taiadfnikn ldkqqshfyp eitqigdvrd  
qqaldqalag fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng  
vkniiftssv avyglkhnp denhphdpfn hygkskwqae evlrewynka  
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv  
gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq slnkkipsmh  
lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf  
vapytllsqgl drtlqyefvh akkdditfvs e

10

【0185】

配列番号:5-大腸菌086 gne1のヌクレオチド配列

【0186】

遺伝子座 AA037706 BCT 06-DEC-2005

20

【0187】

定義 UDP-GlcNAc C4-エピメラーゼ[大腸菌086]

【0188】

寄託番号 AA037706

【0189】

長さ:993

【0190】

種類：DNA

【0191】

生物：大腸菌086

30

【0192】

配列：

## 【化 8】

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCTG TAGGAACCCG  
 ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC  
 AGCAGAGCCA CTTTTATCCA GAAATCACAC AGATTGGTGA TGTTCGTGAT  
 CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT  
 GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG  
 TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC  
 GTTAAAAATA TCATCTTTAC CAGTTCGGTT GCTGTTTATG GTTTGAACAA  
 ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTC AAC CACTACGGCA  
 AAAGCAAGTG GCAGGCGGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA  
 CCAACAGAAC GTTCATTAAC TATCATCCGT CCTACCGTTA TCTTCGGTGA  
 ACGCAACCGC GGTAACGTCT ATAAC TTGCT GAAACAGATC GCTGGCGGCA  
 AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT  
 GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCGAG  
 TTACGAGGTT TATAACTACG TTGATAAGCC AGACCTGAAC ATGAACCGT  
 TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC  
 TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG  
 CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT  
 TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT  
 GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA  
 ATTCGTCCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

10

20

## 【0193】

30

配列番号:6-大腸菌086 UDP-GlcNAc C4-エピメラーゼのアミノ酸配列

## 【0194】

遺伝子座 AA037706

## 【0195】

定義 UDP-GlcNAc C4-エピメラーゼ[大腸菌086]

## 【0196】

寄託番号 AA037706

## 【0197】

長さ:331個のアミノ酸、線状

## 【0198】

種類:PRT

40

## 【0199】

生物:大腸菌086

## 【0200】

配列:

## 【化 9】

mndnvlīga sgfvgrlle taiadfnikn ldkqqshfyp eitqigdvrd  
qqaldqalag fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng  
vkniiftssv avyglkhnp denhphdpfn hygkskwqae evlrewynka  
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv  
gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq slnkkipsmh  
lpypplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf  
vapytllsqgl drtlqyefvh akkdditfvs e

10

## 【0201】

配列番号:7-シゲラ・ボイディ018 gneのヌクレオチド配列

## 【0202】

遺伝子座 ACD09753 BCT 05-MAY-2008

## 【0203】

定義 UDP-N-アセチルグルコサミン4-エピメラーゼ[シゲラ・ボイディCDC 3083-94]

## 【0204】

寄託番号 ACD09753

20

## 【0205】

長さ: 993

## 【0206】

種類: DNA

## 【0207】

生物: シゲラ・ボイディ018

## 【0208】

配列:

## 【化 1 0】

ATGAACGATA ACGTTTTGCT CATAGGAGCT TCCGGATTCTG TAGGAACCCG  
ACTACTTGAA ACGGCAATTG CTGACTTTAA TATCAAGAAC CTGGACAAAC  
AGCAGAGCCA TTTTATCCA GCAATCACAC AGATTGGCGA TGTTCGTGAT  
CAACAGGCAC TCGACCAGGC GTTAGCCGGT TTTGACACTG TTGTACTACT  
GGCAGCGGAA CACCGCGATG ACGTCAGCCC TACTTCTCTC TATTATGATG  
TCAACGTTCA GGGTACCCGC AATGTGCTGG CGGCCATGGA AAAAAATGGC  
GTTAAAAATA TCATCTTTAC CAGTTCGGTT GCTGTTTATG GTTTGAACAA  
ACACAACCCT GACGAAAACC ATCCACACGA CCCTTTCAAC CACTACGGCA  
AAAGTAAGTG GCAGGCAGAG GAAGTGCTGC GTGAATGGTA TAACAAAGCA  
CCAACAGAAC GTTCATTAAC CATCATCCGT CCTACCGTTA TCTTCGGTGA  
ACGCAACCGC GGTAACGTCT ATAACCTTGCT GAAACAGATC GCTGGCGGCA  
AGTTTATGAT GGTGGGCGCA GGGACTAACT ATAAGTCCAT GGCTTATGTT  
GGAAACATTG TTGAGTTTAT CAAGTACAAA CTGAAGAATG TTGCCGCAGG  
TTATGAGGTT TATAACTATG TTGATAAGCC AGACCTGAAC ATGAACCGT  
TGGTTGCTGA AGTTGAACAA AGCCTGAACA AAAAGATCCC TTCTATGCAC  
TTGCCTTACC CACTAGGAAT GCTGGGTGGA TATTGCTTTG ATATCCTGAG  
CAAAATTACG GGCAAAAAAT ACGCTGTCAG CTCTGTGCGC GTGAAAAAAT  
TCTGCGCAAC AACACAGTTT GACGCAACGA AAGTGCATTC TTCAGGTTTT  
GTGGCACCGT ATACGCTGTC GCAAGGTCTG GATCGAACTC TGCAGTATGA  
ATTCTGTCAT GCCAAAAAAG ACGACATAAC GTTTGTTTCT GAG

10

20

## 【0 2 0 9】

30

配列番号:8-シゲラ・ボイディ018 UDP-N-アセチルグルコサミン4-エピメラーゼのアミノ酸配列

## 【0 2 1 0】

遺伝子座 ACD09753

## 【0 2 1 1】

定義 UDP-N-アセチルグルコサミン4-エピメラーゼ[シゲラ・ボイディCDC 3083-94]

## 【0 2 1 2】

寄託番号 ACD09753

## 【0 2 1 3】

長さ:331個のアミノ酸、線状

40

## 【0 2 1 4】

種類:PRT

## 【0 2 1 5】

生物:シゲラ・ボイディ018

## 【0 2 1 6】

配列:

## 【化 1 1】

mndnvlliga sgfvgtlile taiadfnikn ldkqqshfyp aitqigdvrd  
qqaldqalag fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng  
vkniiftssv avyglkhnp denhphdpfn hygkskwqae evlrewynka  
ptersltiir ptvifgernr gnvynllkqi aggkfmmvga gtnyksmayv  
gnivefikyk lknvaagyev ynyvdkpdlm mnqlvaeveq slnkkipsmh  
lpyplgmlgg ycfdilskit gkkyavssvr vkkfcattqf datkvhssgf  
vapytlsqgl drtlqyefvh akkdditfvs e

10

## 【0 2 1 7】

配列番号:9ーサルモネラ・エンテリカ030 gneのヌクレオチド配列

## 【0 2 1 8】

遺伝子座 AAV34516 BCT 25-OCT-2004

## 【0 2 1 9】

定義 UDP-GlcNAc 4-エピメラーゼ[サルモネラ・エンテリカ亜種サラマエ血清型Greens  
ide]

## 【0 2 2 0】

寄託番号 AAV34516

20

## 【0 2 2 1】

長さ: 993

## 【0 2 2 2】

種類: DNA

## 【0 2 2 3】

生物: サルモネラ・エンテリカ030

## 【0 2 2 4】

配列:

## 【化 1 2】

ATGAACGATA ACGTTTTGCT CATTGGTGCT TCCGGATTCTG TAGGAACCCG  
 ACTCCTTGAA ACGGCAGTGG ATGATTTTAA TATCAAGAAC CTGGATAAAC  
 AGCAAAGCCA TTTCTACCCA GAGATTACAC ACATTGGCGA TGTTCGTGAC  
 CAACAAATCC TTGACCAGAC GTTGGTGGGT TTTGACACCG TAGTACTATT  
 GGCTGCGGAG CATCGTGATG ATGTTAGTCC TACCTCGCTT TATTATGATG  
 TCAACGTCCA GGAACGCGT AATGTACTGG CGGCGATGGA AAAAAATGGT  
 GTAAAAATA TCATTTTTAC CAGTTCGGTT GCAGTTTATG GACTCAACAA  
 GAAAAATCCT GACGAAACGC ACCCTCACGA TCCCTTTAAT CATTACGGAA  
 AAAGTAAATG GCAAGCAGAA GAAGTTCTGC GTGAGTGGCA TGCTAAAGCG  
 CCGAATGAGC GTTCTTTGAC CATAATTCGT CCTACCGTTA TTTTCGGGGA  
 GCGTAACCGC GGTAATGTAT ACAATCTCTT GAAACAGATC GCTGGTGGTA  
 AATTTGCGAT GGTGGTCCG GGAATAACT ATAAATCAAT GGCTTATGTT  
 GGTAATATCG TTGAGTTTAT CAAATTCAA CTCAAGAATG TTACGGCGGG  
 CTATGAAGTT TATAATTATG TTGATAAACC TGATCTGAAT ATGAATCAAT  
 TGGTTGCTGA AGTAGAGCAG AGCCTGGGCA AAAAAATACC ATCGATGCAC  
 CTTCCATATC CATTAGGTAT GCTGGGGGGT TACTGTTTCG ATATCCTGAG  
 CAAAGTAACG GGCAAGAAGT ACGCTGTAAG TTCGGTTCGT GTTAAAAAAT  
 TCTGTGCGAC AACGCAGTTT GATGCAACAA AAGTGCATTC TTCTGGTTTT  
 GTTGCGCCAT ACACCTTATC TCAGGGGTTG GATCGTACAC TGCAATATGA  
 ATTTGTTTCA GCAAAGAAAG ATGACATTAC ATTCGTTTCA GAG

10

20

## 【0 2 2 5】

配列番号:10ーサルモネラ・エンテリカ030 UDP-GlcNAc 4-エピメラーゼのアミノ酸配列

30

## 【0 2 2 6】

遺伝子座 AAV34516

## 【0 2 2 7】

定義 UDP-GlcNAc 4-エピメラーゼ[サルモネラ・エンテリカ亜種サラマエ血清型Greenside]

## 【0 2 2 8】

寄託番号 AAV34516

## 【0 2 2 9】

長さ:331個のアミノ酸、線状

## 【0 2 3 0】

種類:PRT

40

## 【0 2 3 1】

生物:サルモネラ・エンテリカ030

## 【0 2 3 2】

配列:

## 【化 1 3】

mndnvllica sgfvgrlle tavddfnikn ldkqqshfyp eithigdvr  
qqildqtlvg fdtvllaae hrddvsptsl yydvnvqgtr nvlaamekng  
vkniiftssv avyglkknk dethphdpfn hygkskwqae evlrewhaka  
pnrsltiir ptvifgernr gnvynllkqi aggkfamvgp gtnyksmayv  
gnivefikfk lknvtagyev ynyvdkpdlm mnqlvaeveq slgkkipshm  
lpyplgmlgg ycfdilskvt gkkyavssvr vkkfcattqf datkvhssgf  
vapytllsqgl drtlqyefvh akkdditfvs e

10

## 【0 2 3 3】

配列番号:11-C.ジェジュニgneのヌクレオチド配列

## 【0 2 3 4】

遺伝子座 YP\_002344524 BCT 14-SEP-2010

## 【0 2 3 5】

定義 UDP-GlcNAc/Glc 4-エピメラーゼ[カンピロバクター・ジェジュニ亜種ジェジュニ

## 【0 2 3 6】

寄託番号 YP\_002344524

## 【0 2 3 7】

長さ:987

## 【0 2 3 8】

種類:DNA

## 【0 2 3 9】

生物:C.ジェジュニ

## 【0 2 4 0】

配列:

20

## 【化 1 4】

ATGAAAATTCTTATTAGCGGTGGTGCAGGTTATATAGGTTCTCATACTTTAAGACAATT  
TTTAAAAACAGATCATGAAATTTGTGTTTTAGATAATCTTTCTAAGGGTTCTAAAATCG  
CAATAGAAGATTTGCAAAAAACAAGAGCTTTTAAATTTTTCGAACAAGATTTAAGTGAT  
TTTCAAGGCGTAAAAGCATTGTTTGAGAGAGAAAAATTTGACGCTATTGTGCATTTTGC  
AGCAAGCATTGAAGTTTTTGAAAGTATGCAAAATCCTTTAAAATATTATATGAACAACA  
CTGTTAATACGACAAATCTCATCGAACTTGTTTGCAAACTGGAGTGAATAAATTTATA  
TTTTCTTCAACGGCGGCCACTTATGGCGAACCACAACTCCCGTTGTGAGCGAAACAAG  
TCCTTTAGCACCTATTAATCCTTATGGGCGTAGTAAGCTTATGAGTGAAGAAGTTTTGC  
GTGATGCAAGTATGGCAAATCCTGAATTTAAGCATTGTATTTTAAGATATTTTAATGTT  
GCAGGTGCTTGTATGGATTATACTTTAGGACAACGCTATCCAAAAGCGACTTTGCTTAT  
AAAAGTTGCAGCTGAATGTGCCGCAGGAAAACGTGATAAACTTTTCATATTTGGCGATG  
ATTATGATACAAAAGATGGTACTTGCATAAGAGATTTTATCCATGTAGATGATATTTCA  
AGTGACATTTAGCGGCTTTGGATTATTTAAAAGAGAATGAAAGCAATGTTTTTAATGT  
AGGTTATGGACATGGTTTTAGCGTAAAAGAAGTGATTGAAGCGATGAAAAAAGTTAGCG  
GAGTGGATTTTAAAGTAGAACTTGCCCCACGCCGTGCGGGTGATCCTAGTGATTGATT  
TCTGATGCAAGTAAAATCAGAAATCTTACTTCTTGGCAGCCTAAATATGATGATTTAGA  
GCTTATTTGTAAATCTGCTTTTGATTGGGAAAAACAGTGTTAA

10

20

## 【 0 2 4 1】

配列番号:12-C.ジェジュニUDP-GlcNAc/Glc 4-エピメラーゼのアミノ酸配列

## 【 0 2 4 2】

遺伝子座 YP\_002344524

## 【 0 2 4 3】

定義 UDP-GlcNAc/Glc 4-エピメラーゼ[カンピロバクター・ジェジュニ亜種ジェジュニ

## 【 0 2 4 4】

寄託番号 YP\_002344524

## 【 0 2 4 5】

長さ:328個のアミノ酸、線状

## 【 0 2 4 6】

種類:PRT

## 【 0 2 4 7】

生物:C.ジェジュニ

## 【 0 2 4 8】

配列:

30

40

## 【化 1 5】

mkilisggag yigshtlrqf lktdheicvl dnlskgskia iedlqktraf  
kffeqdlsdf qgvkalfere kfdaivhfaa sievfesmqn plkyymnntv  
nttnlietcl qtgvnkfifs  
staatygepq tpvvsetspl apinpygrsk lmseevlrda smanpefkhc  
ilryfnvaga cmdytlgqry pkatlilikva aecaagkrdk lfifgddydt  
kdgtcirdfi hvddissahl  
aaldylkene snvfnvgygh gfsvkeviea mkkvsgvdfk velaprragd  
psvlisdask irnltswqpk yddlelicks  
afdwekqc

10

## 【0 2 4 9】

配列番号:13ー大腸菌K12 galeのヌクレオチド配列

## 【0 2 5 0】

遺伝子座 AP\_001390 BCT 30-APR-2010

## 【0 2 5 1】

定義 UDP-ガラクトース-4-エピメラーゼ[大腸菌K-12株W3110亜株]

## 【0 2 5 2】

寄託番号 AP\_001390

## 【0 2 5 3】

長さ:1,017

## 【0 2 5 4】

種類:DNA

## 【0 2 5 5】

生物:大腸菌K12

## 【0 2 5 6】

配列:

20

## 【化 1 6】

ATGAGAGTTCTGGTTACCGGTGGTAGCGGTTACATTGGAAGTCATACCTGTGTGCAA  
T TACTGCAAAACGGTCATGATGTCATCATTCTTGATAACCTCTGTAACAGTAAGCGC  
AGCGTACTGCCTGTTATCGAGCGTTTAGGCGGCAAACATCCAACGTTTGTGTAAGGC  
GATATTCGTAACGAAGCGTTGATGACCGAGATCCTGCACGATCACGCTATCGACAC  
CGTGATCCACTTCGCCGGGCTGAAAGCCGTGGGCGAATCGGTACAAAAACCGCTGG  
AATATTACGACAACAATGTCAACGGCACTCTGCGCCTGATTAGCGCCATGCGCGCC  
GCTAACGTCAAAAACCTTTATTTTTAGCTCCTCCGCCACCGTTTATGGCGATCAGCCC  
AAAATTCCATACGTTGAAAGCTTCCCGACCGGCACACCGCAAAGCCCTTACGGCAA  
AAGCAAGCTGATGGTGGAACAGATCCTCACCGATCTGCAAAAAGCCCAGCCGGACT  
GGAGCATTGCCCTGCTGCGCTACTTCAACCCGGTTGGCGCGCATCCGTCGGGCGATA  
TGGGCGAAGATCCGCAAGGCATTCCGAATAACCTGATGCCATACATCGCCCAGGTT  
GCTGTAGGCCGTCGCGACTCGCTGGCGATTTTTGGTAACGATTATCCGACCGAAGAT  
GGTACTGGCGTACGCGATTACATCCACGTAATGGATCTGGCGGACGGTCACGTCGT  
GGCGATGGAAAACTGGCGAACAAGCCAGGCGTACACATCTACAACCTCGGCGCTG  
GCGTAGGCAACAGCGTGCTGGACGTGGTTAATGCCTTCAGCAAAGCCTGCGGCAAA  
CCGGTTAATTATCATTTTTGCACCGCGTCGCGAGGGCGACCTTCCGGCCTACTGGGCG  
GACGCCAGCAAAGCCGACCGTGAAGTGAAGTGGCGCGTAACGCGCACACTCGATGA  
AATGGCGCAGGACACCTGGCACTGGCAGTCACGCCATCCACAGGGATATCCCGATT  
AA

10

20

## 【0 2 5 7】

配列番号:14ー大腸菌K12 UDP-ガラクトース-4-エピメラーゼのアミノ酸配列

## 【0 2 5 8】

遺伝子座 AP\_001390

## 【0 2 5 9】

定義 UDP-ガラクトース-4-エピメラーゼ [大腸菌K-12株W3110亜株]

## 【0 2 6 0】

寄託番号 AP\_001390

## 【0 2 6 1】

長さ:338個のアミノ酸、線状

## 【0 2 6 2】

種類:PRT

## 【0 2 6 3】

生物:大腸菌K12

## 【0 2 6 4】

配列:

30

40

【化 1 7】

mrvtltvggsgyigshtcvqllqnghdviildnlcnskrsvlpvierlggkhptfvegdi  
rnealmteilhdhaidtvihfaglkavgesvqkpleyydnnvnngtlrlisamraanvkn  
fifsssatvygdqpkipyvesfptgtpqspygksklmveqiltdlqkaqpdwsiallry  
fnpvgahpsgdmgedpqqgipnnlmpyiaqvavgrrdslaiifgndyptedgtgvrdyihv  
mdladghvvameklankpgvhiynlgagvgnsvladvnafskacgkpvnyhfaprrregd  
lpaywadaskadrelnwrvtrtldemaqdtwhwqsrhpqgydp

10

【0 2 6 5】

配列番号:15ー大腸菌086 gne2のヌクレオチド配列

【0 2 6 6】

遺伝子座 AAV85952 BCT 27-MAR-2005

【0 2 6 7】

定義 Gne[大腸菌086]

【0 2 6 8】

寄託番号 AAV85952

【0 2 6 9】

長さ:1,020

20

【0 2 7 0】

種類:DNA

【0 2 7 1】

生物:大腸菌086

【0 2 7 2】

配列:

## 【化 1 8】

ATGGTGATTT TCGTAACAGG CGGTGCAGGA TATATTGGAT CCCATACCAT  
 ACTTGAGTTA CTTAATAATG GTCATGATGT CGTTTCGATA GATAATTTTG  
 TCAATTCCTC TATAGAATCA TTAAAAAGAG TAGAGCAAAT AACTAATAAG  
 AAAATTATTT CTTATCAAGG TGATATCCGT GATAAAAATC TACTTGATGA  
 GATTTTTTCA AGACACCATA TCGATGCTGT AATTCACTTT GCATCGTTAA  
 AATCTGTAGG TGAGTCTAAG TTAAAGCCCT TAGAGTATTA TTCTAATAAT  
 GTTGGTGGAA CTTTAGTATT ACTTGAATGC ATGAAGAGAT ATAACATTAA  
 TAAAATGATA TTTAGCTCTT CTGCTACTGT TTATGGGAGT AACAGTATCC  
 CTCCCCATAC GGAAGATAGA CGAATTGGTG AAAC TACAAA CCCATATGGG  
 ACATCGAAAT TTATAATAGA AATAATTTTG AGTGATTATT GTGATAGTGA  
 TAATAATAAA TCAGTAATTG CACTGCGTTA CTTTAATCCA ATCGGAGCAC  
 ATAAGTCCGG GATGATTGGT GAAAATCCTA ACGGGATCCC TAATAATCTG  
 GTTCCTTATA TATCTAAAGT TGCACAAAAT CAACTTCCTG TATTAAATAT  
 TTATGGCAAC GATTATCCAA CTAAAGATGG TACAGGAGTA AGAGACTATA  
 TACATGTCTG TGATTTGGCT AAAGGGCATG TTAAAGCATT AGAATATATG  
 TTTTAAATG ATGTCAATTA TGAAGCTTTT AATTTAGGTA CTGGTCAAGG  
 TTATTCTGTT TTAGAGATTG TAAAAATGTT TGAGATAGTC ACTAAAAAGA  
 GTATACCTGT TGCTATTTGT AATAGACGTG AGGGGGATGT TGCGGAGTCA  
 TGGGCGTCTG CTGATTTGGC ACATAAAAAG CTTTCCTGGA AAGCGGAAAA  
 AAATTTGAAA GAAATGATCG AAGATGTATG GCGTTGGCAA ACAAACAATC  
 CAAATGGATA TAAAAAATAA

10

20

30

## 【0 2 7 3】

配列番号:16－大腸菌086 Gneのアミノ酸配列

## 【0 2 7 4】

遺伝子座 AAV85952

## 【0 2 7 5】

定義 Gne[大腸菌086]

## 【0 2 7 6】

寄託番号 AAV85952

## 【0 2 7 7】

長さ:339個のアミノ酸(gne2)、線状

40

## 【0 2 7 8】

種類:PRT

## 【0 2 7 9】

生物:大腸菌086

## 【0 2 8 0】

配列:

## 【化 1 9】

mvifvtggag yigshtilel lnngdhvvs i dnfvnssies lkrveqitnk  
kiisyqgdir dknlldeifs rhhidavihf aslksvgesk lkpleyysnn  
vggtlvlllec mkryninkmi fsssatvygs nsipphtedr rigettnpyg  
tskfiieiiil sdycdsdnnk svialryfnp igahksgmig enpngipnnl  
vpyiskvaqn qlpvl niygn dyptkdgtgv rdyihvcdla kghvkaleym  
flndvnyeaf nlgtgggysv leivkmfeiv tkksipvaic nrregdvaes  
wasadlahkk lswkaeknlk emiedvwrwq tnnpngykk

10

## 【0 2 8 1】

配列番号:17-Z3206の末端をコードしている合成オリゴヌクレオチドZ3206-Fw(プライマー)のヌクレオチド配列；制限部位には下線をつけている

## 【0 2 8 2】

長さ：30

## 【0 2 8 3】

種類：DNA

## 【0 2 8 4】

配列：

20

## 【化 2 0】

AAACCCGGGATGAACGATAACGTTTTGCTC

## 【0 2 8 5】

配列番号:18-赤血球凝集素タグ(HAタグ)をつけたZ3206の末端をコードしている合成オリゴヌクレオチドZ3206-RvHA(プライマー)のヌクレオチド配列；制限部位には下線をつけている

## 【0 2 8 6】

長さ：60

## 【0 2 8 7】

種類：DNA

## 【0 2 8 8】

生物：

## 【0 2 8 9】

配列：

## 【化 2 1】

AAATCTAGATTAAGCGTAATCTGGAAC

ATCGTATGGGTACTCAGAAACAAACGTTATGTC

40

## 【0 2 9 0】

配列番号:19-制限部位には下線をつけた、合成オリゴヌクレオチドgne-Fw(プライマー)のヌクレオチド配列

## 【0 2 9 1】

長さ：29

## 【0 2 9 2】

種類：DNA

## 【0 2 9 3】

生物：

## 【0 2 9 4】

50

配列：  
【化 2 2】

AAACCATGGATGAAAATTCTTATTAGCGG

【0 2 9 5】

配列番号:20ー制限部位には下線をつけた、合成オリゴヌクレオチドgne-RV(プライマー)のヌクレオチド配列

【0 2 9 6】

長さ:57

【0 2 9 7】

種類:DNA

【0 2 9 8】

生物:

【0 2 9 9】

配列:

【化 2 3】

AAATCTAGATTAAGCGTAATCTGGAAC

ATCGTATGGGTAGCACTGTTTTTCCCAATC

10

20

【0 3 0 0】

配列番号:21ーNheI制限酵素のための制限部位を含むオリゴヌクレオチドのヌクレオチド配列

【0 3 0 1】

長さ:11

【0 3 0 2】

種類:DNA

【0 3 0 3】

生物:

【0 3 0 4】

配列:

【化 2 4】

30

AAAAAGCTAGC

【0 3 0 5】

配列番号:22ーAscI制限酵素のための制限部位を含むオリゴヌクレオチドのヌクレオチド配列

【0 3 0 6】

長さ:8

【0 3 0 7】

種類:DNA

【0 3 0 8】

生物:

【0 3 0 9】

配列:

【化 2 5】

40

CCGCGCGG

【0 3 1 0】

配列番号:23ーC-末端赤血球凝集素タグを持つZ3206をコードしているプラスミドpMLBAD

50

:Z3206(プラスミド内の大腸菌0157の挿入断片)のヌクレオチド配列

【0311】

定義 生成物のZ3206-pMLBAD\*へのライゲーション

【0312】

特徴 位置/限定子

【0313】

CDS 2105..3098

/ラベル=Z3206

【0314】

CDS 3098..3127

/ラベル=HA

【0315】

長さ：7794 bp

【0316】

種類：DNA環状、未注釈型(UNA)

【0317】

配列：

【化 2 6】

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC  
CCAGTTGACA

61 TAAGCCTGTT CGGTTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG  
CAACTGGTCC

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGCGCAGTGG CGGTTTTTCAT  
GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC 10  
GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA  
GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG  
GTAGGCAGTC

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTTCGACA  
TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG CTTCCCCTGA 20  
GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC  
AAGTCGTCCG

541 TTGGTATTGC ACAAACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT  
CCCACCCAGG

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA  
TATCTTAGGA 30

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG  
CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT  
ATCATCGATG

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCTATGC  
TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA  
TCATTCACTT 40

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT  
TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC  
GACGGTGGCG ATAGGCATCC

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCCTCGCGC  
CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG  
CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG  
ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC  
GCTTCCATGC

10

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG  
CGCCCTTCCC

1321 CTTGCCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA ATGCGGCTGG  
TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA  
GTTAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC  
GCGAGCCTCC GGATGACGAC

20

1501 CGTAGTGATG AATCTCTCCT GCGGGGAACA GCAAAATATC ACCCGGTCGG  
CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA  
ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAT CGAGATAACC GTTGGCCTCA  
ATCGGCGTTA

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG  
CAGGGGATCA TTTTGCGCTT

30

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA  
TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAACCG  
GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT  
GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC  
ACACTTTGCT

40

1981 ATGCCATAGC ATTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT  
TTATCGCAAC

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA  
CCATGGTACC

2101 CGGGATGAAC GATAACGTTT TGCTCATAGG AGCTTCCGGA TTCGTAGGAA  
CCCGACTACT

2161 TGAAACGGCA ATTGCTGACT TTAATATCAA GAACCTGGAC  
AAACAGCAGA GCCACTTTTA

2221 TCCAGAAATC ACACAGATTG GCGATGTTTCG CGATCAACAG GCACTCGACC  
AGGCGTTAGT

10

2281 CGGTTTTGAC ACTGTTGTAC TACTGGCAGC GGAACACCGC GATGACGTCA  
GCCCTACTTC

2341 TCTCTATTAT GATGTCAACG TTCAGGGTAC CCGCAATGTG CTGGCGGCCA  
TGGAAAAAAA

2401 TGGCGTTAAA AATATCATCT TTACCAGTTC CGTTGCTGTT TATGGTTTGA  
ACAAACACAA

2461 CCCTGACGAA AACCATCCAC ACGACCCTTT CAACCACTAC GGCAAAAGTA  
AGTGGCAGGC

20

2521 AGAGGAAGTG CTGCGTGAAT GGTATAACAA AGCACCAACA  
GAACGTTTCAT TAACCATCAT

2581 CCGTCCTACC GTTATCTTCG GTGAACGCAA CCGCGGTAAC GTCTATAACT  
TGCTGAAACA

2641 GATCGCTGGC GGCAAGTTTA TGATGGTGGG CGCAGGGACT  
AACTATAAGT CCATGGCTTA

2701 TGTTGGAAAC ATTGTTGAGT TTATCAAGTA CAAACTGAAG AATGTTGCCG  
CAGGTTATGA

30

2761 GGTTTATAAC TACGTTGATA AGCCAGACCT GAACATGAAC CAGTTGGTTG  
CTGAAGTTGA

2821 ACAAAGCCTG AACAAAAAGA TCCCTTCTAT GCACTTGCCT TACCCACTAG  
GAATGCTGGG

2881 TGGATATTGC TTTGATATCC TGAGCAAAAT TACGGGCAAA AAATACGCTG  
TCAGCTCAGT

2941 GCGCGTGAAA AAATTCTGCG CAACAACACA GTTTGACGCA  
ACGAAAGTGC ATTCTTCAGG

40

3001 TTTTGTGGCA CCGTATACGC TGTCGCAAGG TCTGGATCGA AACTGCAGT  
ATGAATTCGT

3061 TCATGCCAAA AAAGACGACA TAACGTTTGT TTCTGAGTAC CCATACGATG  
TTCCAGATTA

3121 CGCTTAATCT AGAGTCGACC TGCAGGCATG CAAGCTTGGC TGTTTTGGCG  
GATGAGAGAA

3181 GATTTTCAGC CTGATACAGA TTAAATCAGA ACGCAGAAGC GGTCTGATAA  
AACAGAATTT

3241 GCCTGGCGGC AGTAGCGCGG TGGTCCCACC TGACCCCATG CCGAACTCAG  
AAGTGAAACG

10

3301 CCGTAGCGCC GATGGTAGTG TGGGGTCTCC CCATGCGAGA  
GTAGGGAATC GCCAGGCATC

3361 AAATAAAACG AAAGGCTCAG TCGAAAGACT GGGCCTTTCG TTTTATCTGT  
TGTTTGTCGG

3421 TGAACGCTCT CCTGAGTAGG ACAAATCCGC CGGGAGCGGA TTTGAACGTT  
GCCAAGCAAC

3481 GGCCCGGAGG GTGGCGGGCA GGACGCCCGC CATAAACTGC  
CAGGCATCAA ATTAAGCAGA

20

3541 AGGCCATCCT GACGGATGGC CTTTTGCGT TTCTACAAAC TCTTCCACTC  
ACTACAGCAG

3601 AGCCATTTAA ACAACATCCC CTCCCCCTTT CCACCGCGTC AGACGCCCGT  
AGCAGCCCGC

3661 TACGGGCTTT TTCATGCCCT GCCCTAGCGT CCAAGCCTCA CGGCCGCGCT  
CGGCCTCTCT

3721 GGCGGCCTTC TGGCGCTGAG GTCTGCCTCG TGAAGAAGGT GTTGCTGACT  
CATACCAGGC

30

3781 CTGAATCGCC CCATCATCCA GCCAGAAAGT GAGGGAGCCA  
CGGTTGATGA GAGCTTTGTT

3841 GTAGGTGGAC CAGTTGGTGA TTTTGAACCT TTGCTTTGCC ACGGAACGGT  
CTGCGTTGTC

3901 GGGAAGATGC GTGATCTGAT CCTTCAACTC AGCAAAAGTT CGATTTATTC  
AACAAAGCCG

3961 CCGTCCCGTC AAGTCAGCGT AATGCTCTGC CAGTGTTACA ACCAATTAAC  
CAATTCTGAT

40

4021 TAGAAAAACT CATCGAGCAT CAAATGAAAC TGCAATTTAT TCATATCAGG  
ATTATCAATA

4081 CCATATTTTT GAAAAAGCCG TTTCTGTAAT GAAGGAGAAA ACTCACCGAG  
GCAGTTCCAT

4141 AGGATGGCAA GATCCTGGTA TCGGTCTGCG ATTCCGACTC GTCCAACATC  
AATACAACCT

4201 ATTAATTTCC CCTCGTCAAA AATAAGGTTA TCAAGCGAGA AATCACCATG  
AGTGACGACT

4261 GAATCCGGTG AGAATGGCAA AAGCTAAAAA GGCCGTAATA  
TCCAGCTGAA CGGTCTGGTT

10

4321 ATAGGTACAT TGAGCAACTG ACTGAAATGC CTCAAAATGT TCTTTACGAT  
GCCATTGGGA

4381 TATATCAACG GTGGTATATC CAGTGATTTT TTTCTCCATT TTAGCTTCCT  
TAGCTCCTGA

4441 AAATCTCGAT AACTCAAAAA ATACGCCCGG TAGTGATCTT ATTCATTAT  
GGTGAAAGTT

4501 GGAACCTCTT ACGTGCCGAT CAACGTCTCA TTTTCGCCAA AAGTTGGCCC  
AGGGCTTCCC

20

4561 GGTATCAACA GGGACACCAG GATTTATTTA TTCTGCGAAG TGATCTTCCG  
TCACAGGTAT

4621 TTATTCGAAG ACGAAAGGGC CTCGTGATAC GCCTATTTTT ATAGGTTAAT  
GTCATGATAA

4681 TAATGGTTTC TTAGACGTCA GGTGGCACTT TTCGGGGAAA TGTGCGCGCC  
CGCGTTCCTG

4741 CTGGCGCTGG GCCTGTTTCT GCGCTGGAC TTCCCGCTGT TCCGTCAGCA  
GCTTTTCGCC

30

4801 CACGGCCTTG ATGATCGCGG CGGCCTTGGC CTGCATATCC CGATTCAACG  
GCCCCAGGGC

4861 GTCCAGAACG GGCTTCAGGC GCTCCCGAAG GTCTCGGGCC GTCTCTTGGG  
CTTGATCGGC

4921 CTTCTTGCGC ATCTCACGCG CTCCTGCGGC GGCCTGTAGG GCAGGCTCAT  
ACCCCTGCCG

4981 AACCGCTTTT GTCAGCCGGT CGGCCACGGC TTCCGGCGTC TCAACGCGCT  
TTGAGATTCC

40

5041 CAGCTTTTCG GCCAATCCCT GCGGTGCATA GGCGCGTGGC TCGACCGCTT  
GCGGGCTGAT

5101 GGTGACGTGG CCCACTGGTG GCCGCTCCAG GGCCTCGTAG  
AACGCCTGAA TGC GCGTGTG  
5161 ACGTGCCTTG CTGCCCTCGA TGCCCCGTTG CAGCCCTAGA TCGGCCACAG  
CGGCCGCAA  
5221 CGTGGTCTGG TCGCGGGTCA TCTGCGCTTT GTTGCCGATG AACTCCTTGG  
CCGACAGCCT  
5281 GCCGTCCTGC GTCAGCGGCA CCACGAACGC GGTCATGTGC GGGCTGGTTT 10  
CGTCACGGTG  
5341 GATGCTGGCC GTCACGATGC GATCCGCCCC GTACTTGTCC GCCAGCCACT  
TGTGCGCCTT  
5401 CTCGAAGAAC GCCGCCTGCT GTTCTTGGCT GGCCGACTTC CACCATTCCG  
GGCTGGCCGT  
5461 CATGACGTAC TCGACCGCCA ACACAGCGTC CTTGCGCCGC TTCTCTGGCA  
GCAACTCGCG  
5521 CAGTCGGCCC ATCGCTTCAT CGGTGCTGCT GGCCGCCCAG TGCTCGTTCT 20  
CTGGCGTCCT  
5581 GCTGGCGTCA GCGTTGGGCG TCTCGCGCTC GCGGTAGGCG TGCTTGAGAC  
TGGCCGCCAC  
5641 GTTGCCCATTTTCGCCAGCT TCTTGCATCG CATGATCGCG TATGCCGCCA  
TGCTTGCCCC  
5701 TCCCTTTTGG TGTCCAACCG GCTCGACGGG GGCAGCGCAA GGCGGTGCCT  
CCGGCGGGCC  
5761 ACTCAATGCT TGAGTATACT CACTAGACTT TGCTTCGCAA AGTCGTGACC 30  
GCCTACGGCG  
5821 GCTGCGGCGC CCTACGGGCT TGCTCTCCGG GCTTCGCCCT GCGCGGTCGC  
TGCGCTCCCT  
5881 TGCCAGCCCG TGGATATGTG GACGATGGCC GCGAGCGGCC  
ACCGGCTGGC TCGCTTCGCT  
5941 CGGCCCCGTGG ACAACCCTGC TGGACAAGCT GATGGACAGG  
CTGCGCCTGC CCACGAGCTT  
6001 GACCACAGGG ATTGCCCACC GGCTACCCAG CCTTCGACCA CATACCCACC 40  
GGCTCCAACCT  
6061 GCGCGGCCTG CGGCCTTGCC CCATCAATTT TTTTAATTTT CTCTGGGGAA  
AAGCCTCCGG

6121 CCTGCGGCCT GCGCGCTTCG CTTGCCGGTT GGACACCAAG TGGAAGGCGG  
GTCAAGGCTC

6181 GCGCAGCGAC CGCGCAGCGG CTTGGCCTTG ACGCGCCTGG  
AACGACCCAA GCCTATGCGA

6241 GTGGGGGCGAG TCGAAGGCGA AGCCCGCCCG CCTGCCCCC  
GAGCCTCACG GCGGCGAGTG

6301 CGGGGGTTC AAGGGGGCAG CGCCACCTTG GGCAAGGCCG  
AAGGCCGCGC AGTCGATCAA

10

6361 CAAGCCCCGG AGGGGCCACT TTTTGCCGGA GGGGGAGCCG  
CGCCGAAGGC GTGGGGGAAC

6421 CCCGCAGGGG TGCCCTTCTT TGGGCACCAA AGAACTAGAT  
ATAGGGCGAA ATGCGAAAGA

6481 CTAAAAATC AACAACTTAA AAAAGGGGGG TACGCAACAG  
CTCATTGCGG CACCCCCCGC

6541 AATAGCTCAT TGCGTAGGTT AAAGAAAATC TGTAATTGAC TGCCACTTTT  
ACGCAACGCA

20

6601 TAATTGTTGT CGCGCTGCCG AAAAGTTGCA GCTGATTGCG CATGGTGCCG  
CAACCGTGCG

6661 GCACCCTACC GCATGGAGAT AAGCATGGCC ACGCAGTCCA  
GAGAAATCGG CATTCAAGCC

6721 AAGAACAAGC CCGGTCCTG GGTGCAAACG GAACGCAAAG  
CGCATGAGGC GTGGGCCGGG

6781 CTTATTGCGA GGAAACCCAC GCGGCAATG CTGCTGCATC ACCTCGTGCC  
GCAGATGGGC

30

6841 CACCAGAACG CCGTGGTGGT CAGCCAGAAG ACACTTTCCA  
AGCTCATCGG ACGTTCTTTG

6901 CGGACGGTCC AATACGCAGT CAAGGACTTG GTGGCCGAGC  
GCTGGATCTC CGTCGTGAAG

6961 CTCAACGGCC CCGGCACCGT GTCGGCCTAC GTGGTCAATG ACCGCGTGCC  
GTGGGGCCAG

7021 CCCC GCGACC AGTTGCGCCT GTCGGTGTTT AGTGCCGCCG TGGTGGTTGA  
TCACGACGAC

40

7081 CAGGACGAAT CGCTGTTGGG GCATGGCGAC CTGCGCCGCA TCCCGACCCT  
GTATCCGGGC

7141 GAGCAGCAAC TACCGACCGG CCCC GGCGAG GAGCCGCCCA  
GCCAGCCCGG CATTCCGGGC

7201 ATGGAACCAG ACCTGCCAGC CTTGACCGAA ACGGAGGAAT  
GGGAACGGCG CGGGCAGCAG

7261 CGCCTGCCGA TGCCCGATGA GCCGTGTTTT CTGGACGATG GCGAGCCGTT  
GGAGCCGCCG

7321 ACACGGGTCA CGCTGCCGCG CCGGTAGCAC TTGGGTTCG  
CAGCAACCCG TAAGTGCGCT

10

7381 GTTCCAGACT ATCGGCTGTA GCCGCCTCGC CGCCCTATAC CTTGTCTGCC  
TCCCCGCGTT

7441 GCGTCGCGGT GCATGGAGCC GGGCCACCTC GACCTGAATG  
GAAGCCGGCG GCACCTCGCT

7501 AACGGATTCA CCGTTTTTAT CAGGCTCTGG GAGGCAGAAT AAATGATCAT  
ATCGTCAATT

7561 ATTACCTCCA CGGGGAGAGC CTGAGCAAAC TGGCCTCAGG  
CATTTGAGAA GCACACGGTC

20

7621 AACTGCTTC CGGTAGTCAA TAAACCGGTA AACCAGCAAT  
AGACATAAGC GGCTATTTAA

7681 CGACCCTGCC CTGAACCGAC GACCGGGTCG AATTTGCTTT CGAATTTCTG  
CCATTCATCC

7741 GCTTATTATC ACTTATTCAG GCGTAGCACC AGGCGTTTAA GTCGACCAAT  
AACC

30

# 【0318】

配列番号:24-C-末端赤血球凝集素タグを持つGneをコードしているpMLBAD:gne(プラスミド内の大腸菌0157の挿入断片)のヌクレオチド配列

# 【0319】

遺伝子座 gne-pMLBAD

# 【0320】

定義 dig galeEのpmlbad did(NcoI-XbaI)へのライゲーション

# 【0321】

特徴 位置/限定子

# 【0322】

CDS 2097..3080

/ラベル=galeE

# 【0323】

CDS 3081..3107

/ラベル=HA

# 【0324】

領域 3108..3110

/ラベル=停止

# 【0325】

長さ:7776 bp

40

50

【0 3 2 6】

種類：DNA環状、未注釈型(UNA)

【0 3 2 7】

配列：

【化 2 7】

1 TCTACGGGGT CTGACGCTCA GTGGAACGAA ATCGATGAGC TCGCACGAAC  
CCAGTTGACA

61 TAAGCCTGTT CGGTTCGTAA ACTGTAATGC AAGTAGCGTA TGCGCTCACG  
CAACTGGTCC

10

121 AGAACCTTGA CCGAACGCAG CGGTGGTAAC GGC GCAGTGG CGGTTTTCAT  
GGCTTGTTAT

181 GACTGTTTTT TTGTACAGTC TAGCCTCGGG CATCCAAGCT AGCTAAGCGC  
GTTACGCCGT

241 GGGTCGATGT TTGATGTTAT GGAACAGCAA CGATGTTACG CAGCAGGGTA  
GTCGCCCTAA

301 AACAAAGTTA GGCAGCCGTT GTGCTGGTGC TTTCTAGTAG TTGTTGTGGG  
GTAGGCAGTC

20

361 AGAGCTCGAT TTGCTTGTCG CCATAATAGA TTCACAAGAA GGATTCGACA  
TGGGTCAAAG

421 TAGCGATGAA GCCAACGCTC CCGTTGCAGG GCAGTTTGCG CTTCCCCTGA  
GTGCCACCTT

481 TGGCTTAGGG GATCGCGTAC GCAAGAAATC TGGTGCCGCT TGGCAGGGTC  
AAGTCGTCGG

541 TTGGTATTGC ACAA AACTCA CTCCTGAAGG CTATGCGGTC GAGTCCGAAT  
CCCACCCAGG

30

601 CTCAGTGCAA ATTTATCCTG TGGCTGCACT TGAACGTGTG GCCTAAGCGA  
TATCTTAGGA

661 TCTCCCATCG GTGATGTCGG CGATATAGGC GCCAGCAACC GCACCTGTGG  
CGCCGGTGAT

721 GCCGGCCACG ATGCGTCCGG CGTAGAGGAT CTGCTCATGT TTGACAGCTT  
ATCATCGATG

40

781 CATAATGTGC CTGTCAAATG GACGAAGCAG GGATTCTGCA AACCCCTATGC  
TACTCCGTCA

841 AGCCGTCAAT TGTCTGATTC GTTACCAATT ATGACAACTT GACGGCTACA  
TCATTCACTT

901 TTTCTTCACA ACCGGCACGG AACTCGCTCG GGCTGGCCCC GGTGCATTTT  
TTAAATACCC

961 GCGAGAAATA GAGTTGATCG TCAAAACCAA CATTGCGACC  
GACGGTGGCG ATAGGCATCC

10

1021 GGGTGGTGCT CAAAAGCAGC TTCGCCTGGC TGATACGTTG GTCCTCGCGC  
CAGCTTAAGA

1081 CGCTAATCCC TAACTGCTGG CGGAAAAGAT GTGACAGACG  
CGACGGCGAC AAGCAAACAT

1141 GCTGTGCGAC GCTGGCGATA TCAAAATTGC TGTCTGCCAG GTGATCGCTG  
ATGTACTGAC

1201 AAGCCTCGCG TACCCGATTA TCCATCGGTG GATGGAGCGA CTCGTTAATC  
GCTTCCATGC

20

1261 GCCGCAGTAA CAATTGCTCA AGCAGATTTA TCGCCAGCAG CTCCGAATAG  
CGCCCTTCCC

1321 CTTGCCCGGC GTTAATGATT TGCCCAAACA GGTCGCTGAA ATGCGGCTGG  
TGCGCTTCAT

1381 CCGGGCGAAA GAACCCCGTA TTGGCAAATA TTGACGGCCA  
GTAAAGCCAT TCATGCCAGT

1441 AGGCGCGCGG ACGAAAGTAA ACCCACTGGT GATACCATTC  
GCGAGCCTCC GGATGACGAC

30

1501 CGTAGTGATG AATCTCTCCT GGC GGGAACA GCAAAATATC ACCCGGTCGG  
CAAACAAATT

1561 CTCGTCCCTG ATTTTTCACC ACCCCCTGAC CGCGAATGGT GAGATTGAGA  
ATATAACCTT

1621 TCATTCCCAG CGGTCGGTCG ATAAAAAAT CGAGATAACC GTTGGCCTCA  
ATCGGCGTTA

40

1681 AACCCGCCAC CAGATGGGCA TTAAACGAGT ATCCCGGCAG  
CAGGGGATCA TTTTGCGCTT

1741 CAGCCATACT TTTCATACTC CCGCCATTCA GAGAAGAAAC CAATTGTCCA  
TATTGCATCA

1801 GACATTGCCG TCACTGCGTC TTTTACTGGC TCTTCTCGCT AACCAAACCG  
GTAACCCCGC

1861 TTATTAAAAG CATTCTGTAA CAAAGCGGGA CCAAAGCCAT  
GACAAAAACG CGTAACAAAA

1921 GTGTCTATAA TCACGGCAGA AAAGTCCACA TTGATTATTT GCACGGCGTC  
ACACTTTGCT

1981 ATGCCATAGC ATTTTATCC ATAAGATTAG CGGATCCTAC CTGACGCTTT  
TTATCGCAAC

10

2041 TCTCTACTGT TTCTCCATAC CCGTTTTTTT GGGCTAGCAG GAGGAATTCA  
CCATGGATGA

2101 AAATTCTTAT TAGCGGTGGT GCAGGTTATA TAGGTTCTCA TACTTTAAGA  
CAATTTTTAA

2161 AAACAGATCA TGAAATTTGT GTTTTAGATA ATCTTTCTAA GGGTTCTAAA  
ATCGCAATAG

2221 AAGATTTGCA AAAAATAAGA ACTTTTAAAT TTTTGAACA AGATTTAAGT  
GATTTTCAAG

20

2281 GCGTAAAAGC ATTGTTTGAG AGAGAAAAAT TTGACGCTAT TGTGCATTTT  
GCAGCGAGCA

2341 TTGAAGTTTT TGAAAGTATG CAAAACCTT TAAAGTATTA TATGAATAAC  
ACTGTTAATA

2401 CGACAAATCT CATCGAACT TGTTTGCAA CTGGAGTGAA TAAATTTATA  
TTTTCTTCAA

2461 CGGCAGCCAC TTATGGCGAA CCACAACTC CCGTTGTGAG  
CGAAACAAGT CCTTTAGCAC

30

2521 CTATTAATCC TTATGGGCGT AGTAAGCTTA TGAGCGAAGA GGTTTTGCCT  
GATGCAAGTA

2581 TGGCAAATCC TGAATTTAAG CATTGTATTT TAAGATATTT TAATGTTGCA  
GGTGCTTGCA

2641 TGGATTATAC TTTAGGACAA CGCTATCCAA AAGCGACTTT GCTTATAAAA  
GTTGCAGCTG

2701 AATGTGCCGC AGAAAAACGT AATAAACTTT TCATATTTGG CGATGATTAT  
GATACAAAAG

40

2761 ATGGCACTTG CATAAGAGAT TTTATCCATG TGGATGATAT TTCAAGTGCG  
CATTTATCGG

2821 CTTTGGATTA TTAAAAGAG AATGAAAGCA ATGTTTTTAA TGTAGGTTAT  
GGACATGGTT

2881 TTAGCGTAAA AGAAGTGATT GAAGCGATGA AAAAAGTTAG  
CGGAGTGGAT TTAAAGTAG

2941 AACTTGCCCC ACGCCGTGCG GGTGATCCTA GTGTATTGAT TTCTGATGCA  
AGTAAAATCA

3001 GAAATCTTAC TTCTTGGCAG CCTAAATATG ATGATTTAGG GCTTATTTGT  
AAATCTGCTT

10

3061 TTGATTGGGA AAAACAGTGC TACCCATACG ATGTTCCAGA TTACGCTTAA  
TCTAGAGTCG

3121 ACCTGCAGGC ATGCAAGCTT GGCTGTTTTG GCGGATGAGA GAAGATTTTC  
AGCCTGATAC

3181 AGATTAAATC AGAACGCAGA AGCGGTCTGA TAAACAGAA  
TTTGCCTGGC GGCAGTAGCG

3241 CGGTGGTCCC ACCTGACCCC ATGCCGAAC CAGAAGTGAA  
ACGCCGTAGC GCCGATGGTA

20

3301 GTGTGGGGTC TCCCCATGCG AGAGTAGGGA ACTGCCAGGC  
ATCAAATAAA ACGAAAGGCT

3361 CAGTCGAAAG ACTGGGCCTT TCGTTTTATC TGTTGTTTGT CGGTGAACGC  
TCTCCTGAGT

3421 AGGACAAATC CGCCGGGAGC GGATTTGAAC GTTGCGAAGC  
AACGGCCCCG AGGGTGGCGG

3481 GCAGGACGCC CGCCATAAAC TGCCAGGCAT CAAATTAAGC  
AGAAGGCCAT CCTGACGGAT

30

3541 GGCCTTTTTG CGTTTCTACA AACTCTTCCA CTCACTACAG CAGAGCCATT  
TAAACAACAT

3601 CCCCTCCCCC TTTCCACCGC GTCAGACGCC CGTAGCAGCC CGCTACGGGC  
TTTTTCATGC

3661 CCTGCCCTAG CGTCCAAGCC TCACGGCCGC GCTCGGCCTC TCTGGCGGCC  
TTCTGGCGCT

3721 GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG ACTCATACCA GGCCTGAATC  
GCCCCATCAT

40

3781 CCAGCCAGAA AGTGAGGGAG CCACGGTTGA TGAGAGCTTT  
GTTGTAGGTG GACCAGTTGG

3841 TGATTTTGAA CTTTGTCTT GCCACGGAAC GGTCTGCGTT GTCGGGAAGA  
TGC GTGATCT

3901 GATCCTTCAA CTCAGCAAAA GTTCGATTTA TTCAACAAAG CCGCCGTCCC  
GTCAAGTCAG

3961 CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA  
ACTCATCGAG

4021 CATCAAATGA AACTGCAATT TATTCATATC AGGATTATCA ATACCATATT  
TTTGAAAAAG

10

4081 CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC  
CATAGGATGG CAAGATCCTG

4141 GTATCGGTCT GCGATTCCGA CTCGTCCAAC ATCAATACAA CCTATTAATT  
TCCCCTCGTC

4201 AAAAATAAGG TTATCAAGCG AGAAATCACC ATGAGTGACG  
ACTGAATCCG GTGAGAATGG

4261 CAAAAGCTAA AAAGGCCGTA ATATCCAGCT GAACGGTCTG  
GTTATAGGTA CATTGAGCAA

20

4321 CTGACTGAAA TGCCTCAAAA TGTTCTTTAC GATGCCATTG GGATATATCA  
ACGGTGGTAT

4381 ATCCAGTGAT TTTTTCTCC ATTTAGCTT CCTTAGCTCC TGAAAATCTC  
GATAACTCAA

4441 AAAATACGCC CGGTAGTGAT CTTATTTTAT TATGGTGAAA GTTGGAACCT  
CTTACGTGCC

4501 GATCAACGTC TCATTTTCGC CAAAAGTTGG CCCAGGGCTT CCCGGTATCA  
ACAGGGACAC

30

4561 CAGGATTTAT TTATTCTGCG AAGTGATCTT CCGTCACAGG TATTTATTCG  
AAGACGAAAG

4621 GGCCTCGTGA TACGCCTATT TTTATAGGTT AATGTCATGA TAATAATGGT  
TTCTTAGACG

4681 TCAGGTGGCA CTTTTCGGGG AAATGTGCGC GCCCGCGTTC CTGCTGGCGC  
TGGGCCTGTT

4741 TCTGGCGCTG GACTTCCCGC TGTTCGTCA GCAGCTTTTC GCCCACGGCC  
TTGATGATCG

40

4801 CGGCGGCCTT GGCCTGCATA TCCCGATTCA ACGGCCCCAG GCGTCCAGA  
ACGGGCTTCA

4861 GGCCTCCCCG AAGGTCTCGG GCCGTCTCTT GGGCTTGATC GGCCTTCTTG  
CGCATCTCAC

4921 GCGCTCCTGC GCGGGCCTGT AGGGCAGGCT CATACCCCTG CCGAACCGCT  
TTTGTCAGCC

4981 GGTGCGCCAC GGCTTCCGGC GTCTCAACGC GCTTTGAGAT TCCCAGCTTT  
TCGGCCAATC

5041 CCTGCGGTGC ATAGGCGCGT GGCTCGACCG CTTGCGGGCT GATGGTGACG  
TGGCCCACTG 10

5101 GTGGCCGCTC CAGGGCCTCG TAGAACGCCT GAATGCGCGT GTGACGTGCC  
TTGCTGCCCT

5161 CGATGCCCCG TTGCAGCCCT AGATCGGCCA CAGCGGCCGC AAACGTGGTC  
TGGTCGCGGG

5221 TCATCTGCGC TTTGTTGCCG ATGAACTCCT TGGCCGACAG CCTGCCGTCC  
TGCCTCAGCG

5281 GCACCACGAA CGCGGTCATG TCGGGGCTGG TTTCGTACG GTGGATGCTG  
GCCGTCACGA 20

5341 TGCATCCGC CCCGTACTTG TCCGCCAGCC ACTTGTGCGC CTTCTCGAAG  
AACGCCGCCT

5401 GCTGTTCTTG GCTGGCCGAC TTCCACCATT CCGGGCTGGC CGTCATGACG  
TACTCGACCG

5461 CCAACACAGC GTCCTTGCGC CGCTTCTCTG GCAGCAACTC GCGCAGTCGG  
CCCATCGCTT

5521 CATCGGTGCT GCTGGCCGCC CAGTGCTCGT TCTCTGGCGT CCTGCTGGCG  
TCAGCGTTGG 30

5581 GCGTCTCGCG CTCGCGGTAG GCGTGCTTGA GACTGGCCGC CACGTTGCCC  
ATTTTCGCCA

5641 GCTTCTTGCA TCGCATGATC GCGTATGCCG CCATGCCTGC CCCTCCCTTT  
TGGTGTCCAA

5701 CCGGCTCGAC GGGGGCAGCG CAAGGCGGTG CCTCCGGCGG  
GCCACTCAAT GCTTGAGTAT

5761 ACTCACTAGA CTTTGCTTCG CAAAGTCGTG ACCGCCTACG GCGGCTGCGG  
CGCCCTACGG 40

5821 GCTTGCTCTC CGGGCTTCGC CCTGCGCGGT CGCTGCGCTC CCTTGCCAGC  
CCGTGGATAT

5881 GTGGACGATG GCCGCGAGCG GCCACCGGCT GGCTCGCTTC GCTCGGCCCC  
TGGACAACCC

5941 TGCTGGACAA GCTGATGGAC AGGCTGCGCC TGCCCACGAG  
CTTGACCACA GGGATTGCCC

6001 ACCGGCTACC CAGCCTTCGA CCACATACCC ACCGGCTCCA ACTGCGCGGC  
CTGCGGCCTT

6061 GCCCCATCAA TTTTTTTAAT TTTCTCTGGG GAAAAGCCTC CGGCCTGCGG  
CCTGCGCGCT

10

6121 TCGCTTGCCG GTTGGACACC AAGTGGAAGG CGGGTCAAGG  
CTCGCGCAGC GACCGCGCAG

6181 CGGCTTGGCC TTGACGCGCC TGGAACGACC CAAGCCTATG  
CGAGTGGGGG CAGTCGAAGG

6241 CGAAGCCCGC CCGCCTGCCC CCCGAGCCTC ACGGCGGCGA  
GTGCGGGGGT TCCAAGGGGG

6301 CAGCGCCACC TTGGGCAAGG CCGAAGGCCG CGCAGTCGAT  
CAACAAGCCC CGGAGGGGGC

20

6361 ACTTTTTGCC GGAGGGGGAG CCGCGCCGAA GGCGTGGGGG  
AACCCCGCAG GGGTGCCCTT

6421 CTTTGGGCAC CAAAGAACTA GATATAGGGC GAAATGCGAA  
AGACTTAAAA ATCAACAAC

6481 TAAAAAAGG GGGTACGCAA CAGCTCATTG CGGCACCCCC  
CGCAATAGCT CATTGCGTAG

6541 GTTAAAGAAA ATCTGTAATT GACTGCCACT TTTACGCAAC GCATAATTGT  
TGTCGCGCTG

30

6601 CCGAAAAGTT GCAGCTGATT GCGCATGGTG CCGCAACCGT GCGGCACCCT  
ACCGCATGGA

6661 GATAAGCATG GCCACGCAGT CCAGAGAAAT CGGCATTCAA  
GCCAAGAACA AGCCCGGTCA

6721 CTGGGTGCAA ACGGAACGCA AAGCGCATGA GGCGTGGGCC  
GGGCTTATTG CGAGGAAACC

6781 CACGGCGGCA ATGCTGCTGC ATCACCTCGT GGCGCAGATG  
GGCCACCAGA ACGCCGTGGT

40

6841 GGTCAGCCAG AAGACACTTT CCAAGCTCAT CGGACGTTCT TTGCGGACGG  
TCCAATACGC

6901 AGTCAAGGAC TTGGTGGCCG AGCGCTGGAT CTCCGTCGTG AAGCTCAACG  
GCCCCGGCAC

6961 CGTGTCTGGCC TACGTGGTCA ATGACCGCGT GGCCTGGGGC CAGCCCCGCG  
ACCAGTTGCG

7021 CCTGTCTGGTG TTCAGTGCCG CCGTGGTGGT TGATCACGAC GACCAGGACG  
AATCGCTGTT

7081 GGGGCATGGC GACCTGCGCC GCATCCCGAC CCTGTATCCG  
GGCGAGCAGC AACTACCGAC

10

7141 CGGCCCCGGC GAGGAGCCGC CCAGCCAGCC CGGCATTCCG  
GGCATGGAAC CAGACCTGCC

7201 AGCCTTGACC GAAACGGAGG AATGGGAACG GCGCGGGCAG  
CAGCGCCTGC CGATGCCCCA

7261 TGAGCCGTGT TTTCTGGACG ATGGCGAGCC GTTGGAGCCG CCGACACGGG  
TCACGCTGCC

7321 GCGCCGGTAG CACTTGGGTT GCGCAGCAAC CCGTAAGTGC GCTGTTCCAG  
ACTATCGGCT

20

7381 GTAGCCGCCT CGCCGCCCTA TACCTTGTCT GCCTCCCCGC GTTGCGTCGC  
GGTGCATGGA

7441 GCCGGGCCAC CTCGACCTGA ATGGAAGCCG GCGGCACCTC  
GCTAACGGAT TCACCGTTTT

7501 TATCAGGCTC TGGGAGGCAG AATAAATGAT CATATCGTCA ATTATTACCT  
CCACGGGGAG

7561 AGCCTGAGCA AACTGGCCTC AGGCATTTGA GAAGCACACG  
GTCACACTGC TTCCGGTAGT

30

7621 CAATAAACCG GTAAACCAGC AATAGACATA AGCGGCTATT  
TAACGACCCT GCCCTGAACC

7681 GACGACCGGG TCGAATTTGC TTTCGAATTT CTGCCATTCA TCCGCTTATT  
ATCACTTATT

7741 CAGGCGTAGC ACCAGGCGTT TAAGTCGACC AATAAC

【 0 3 2 8 】

40

配列番号:25-シグナル配列を持つ修飾されたEPAのアミノ酸配列

【 0 3 2 9 】

WO 2009/104074に開示(配列番号:6として)

【 0 3 3 0 】

種類:PRT

【 0 3 3 1 】

生物:人工

【 0 3 3 2 】

/ノート="人工配列の説明:合成ポリペプチド"

【 0 3 3 3 】

50

長さ : 643

【 0 3 3 4 】

配列 :

【 化 2 8 】

Met Lys Lys Ile Trp Leu Ala Leu Ala Gly Leu Val Leu Ala Phe Ser  
 1 5 10 15

Ala Ser Ala Ala Glu Glu Ala Phe Asp Leu Trp Asn Glu Cys Ala Lys  
 20 25 30

10

Ala Cys Val Leu Asp Leu Lys Asp Gly Val Arg Ser Ser Arg Met Ser  
 35 40 45

Val Asp Pro Ala Ile Ala Asp Thr Asn Gly Gln Gly Val Leu His Tyr  
 50 55 60

Ser Met Val Leu Glu Gly Gly Asn Asp Ala Leu Lys Leu Ala Ile Asp  
 65 70 75 80

20

Asn Ala Leu Ser Ile Thr Ser Asp Gly Leu Thr Ile Arg Leu Glu Gly  
 85 90 95

Gly Val Glu Pro Asn Lys Pro Val Arg Tyr Ser Tyr Thr Arg Gln Ala  
 100 105 110

Arg Gly Ser Trp Ser Leu Asn Trp Leu Val Pro Ile Gly His Glu Lys  
 115 120 125

30

Pro Ser Asn Ile Lys Val Phe Ile His Glu Leu Asn Ala Gly Asn Gln  
 130 135 140

Leu Ser His Met Ser Pro Ile Tyr Thr Ile Glu Met Gly Asp Glu Leu  
 145 150 155 160

40

Leu Ala Lys Leu Ala Arg Asp Ala Thr Phe Phe Val Arg Ala His Glu  
 165 170 175

Ser Asn Glu Met Gln Pro Thr Leu Ala Ile Ser His Ala Gly Val Ser  
 180 185 190

Val Val Met Ala Gln Ala Gln Pro Arg Arg Glu Lys Arg Trp Ser Glu  
 195 200 205

10

Trp Ala Ser Gly Lys Val Leu Cys Leu Leu Asp Pro Leu Asp Gly Val  
 210 215 220

Tyr Asn Tyr Leu Ala Gln Gln Arg Cys Asn Leu Asp Asp Thr Trp Glu  
 225 230 235 240

Gly Lys Ile Tyr Arg Val Leu Ala Gly Asn Pro Ala Lys His Asp Leu  
 245 250 255

20

Asp Ile Lys Asp Asn Asn Asn Ser Thr Pro Thr Val Ile Ser His Arg  
 260 265 270

Leu His Phe Pro Glu Gly Gly Ser Leu Ala Ala Leu Thr Ala His Gln  
 275 280 285

Ala Cys His Leu Pro Leu Glu Ala Phe Thr Arg His Arg Gln Pro Arg  
 290 295 300

30

Gly Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val  
 305 310 315 320

Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val  
 325 330 335

Ile Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu  
 340 345 350

40

Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala

|                                                                 |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|
| 355                                                             | 360 | 365 |     |
| Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu |     |     |     |
| 370                                                             | 375 | 380 |     |
| Ala Gly Ala Ala Ser Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala |     |     |     |
| 385                                                             | 390 | 395 | 400 |
| Lys Asp Gln Asn Arg Thr Lys Gly Glu Cys Ala Gly Pro Ala Asp Ser |     |     |     |
|                                                                 | 405 | 410 | 415 |
| Gly Asp Ala Leu Leu Glu Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu |     |     |     |
|                                                                 | 420 | 425 | 430 |
| Gly Asp Gly Gly Asp Val Ser Phe Ser Thr Arg Gly Thr Gln Asn Trp |     |     |     |
|                                                                 | 435 | 440 | 445 |
| Thr Val Glu Arg Leu Leu Gln Ala His Arg Gln Leu Glu Glu Arg Gly |     |     |     |
| 450                                                             | 455 | 460 |     |
| Tyr Val Phe Val Gly Tyr His Gly Thr Phe Leu Glu Ala Ala Gln Ser |     |     |     |
| 465                                                             | 470 | 475 | 480 |
| Ile Val Phe Gly Gly Val Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile |     |     |     |
|                                                                 | 485 | 490 | 495 |
| Trp Arg Gly Phe Tyr Ile Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr |     |     |     |
|                                                                 | 500 | 505 | 510 |
| Ala Gln Asp Gln Glu Pro Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala |     |     |     |
| 515                                                             | 520 | 525 |     |
| Leu Leu Arg Val Tyr Val Pro Arg Trp Ser Leu Pro Gly Phe Tyr Arg |     |     |     |
| 530                                                             | 535 | 540 |     |
| Thr Gly Leu Thr Leu Ala Ala Pro Glu Ala Ala Gly Glu Val Glu Arg |     |     |     |
| 545                                                             | 550 | 555 | 560 |

10

20

30

40

10

20

30

【化 2 9】

40

Met Ile Ile Ser Asn Asp Gly Tyr Ala Phe Ala Glu Gly Ala Arg Asp

40

Lys Arg Leu Asn Phe Met Ile Ile Gly Ile Leu Gly Ser Ala Thr Leu  
 260 265 270

Ile Phe Leu Ile Leu Ser Gly Gly Val Asp Pro Ile Leu Tyr Gln Leu  
 275 280 285

Lys Phe Tyr Ile Phe Arg Ser Asp Glu Ser Ala Asn Leu Thr Gln Gly  
 290 295 300

10

Phe Met Tyr Phe Asn Val Asn Gln Thr Ile Gln Glu Val Glu Asn Val  
 305 310 315 320

Asp Leu Ser Glu Phe Met Arg Arg Ile Ser Gly Ser Glu Ile Val Phe  
 325 330 335

Leu Phe Ser Leu Phe Gly Phe Val Trp Leu Leu Arg Lys His Lys Ser  
 340 345 350

20

Met Ile Met Ala Leu Pro Ile Leu Val Leu Gly Phe Leu Ala Leu Lys  
 355 360 365

Gly Gly Leu Arg Phe Thr Ile Tyr Ser Val Pro Val Met Ala Leu Gly  
 370 375 380

Phe Gly Phe Leu Leu Ser Glu Phe Lys Ala Ile Met Val Lys Lys Tyr  
 385 390 395 400

30

Ser Gln Leu Thr Ser Asn Val Cys Ile Val Phe Ala Thr Ile Leu Thr  
 405 410 415

Leu Ala Pro Val Phe Ile His Ile Tyr Asn Tyr Lys Ala Pro Thr Val  
 420 425 430

Phe Ser Gln Asn Glu Ala Ser Leu Leu Asn Gln Leu Lys Asn Ile Ala  
 435 440 445

40

Asn Arg Glu Asp Tyr Val Val Thr Trp Ala Ala Tyr Gly Tyr Pro Val  
 450 455 460

Arg Tyr Tyr Ser Asp Val Lys Thr Leu Val Asp Gly Gly Lys His Leu  
465 470 475 480

Gly Lys Asp Asn Phe Phe Pro Ser Phe Ala Leu Ser Lys Asp Glu Gln  
485 490 495

Ala Ala Ala Asn Met Ala Arg Leu Ser Val Glu Tyr Thr Glu Lys Ser  
500 505 510

10

Phe Tyr Ala Pro Gln Asn Asp Ile Leu Lys Thr Asp Ile Leu Gln Ala  
515 520 525

Met Met Lys Asp Tyr Asn Gln Ser Asn Val Asp Leu Phe Leu Ala Ser  
530 535 540

Leu Ser Lys Pro Asp Phe Lys Ile Asp Thr Pro Lys Thr Arg Asp Ile  
545 550 555 560

20

Tyr Leu Tyr Met Pro Ala Arg Met Ser Leu Ile Phe Ser Thr Val Ala  
565 570 575

Ser Phe Ser Phe Ile Asn Leu Asp Thr Gly Val Leu Asp Lys Pro Phe  
580 585 590

Thr Phe Ser Thr Ala Tyr Pro Leu Asp Val Lys Asn Gly Glu Ile Tyr  
595 600 605

30

Leu Ser Asn Gly Val Val Leu Ser Asp Asp Phe Arg Ser Phe Lys Ile  
610 615 620

Gly Asp Asn Val Val Ser Val Asn Ser Ile Val Glu Ile Asn Ser Ile  
625 630 635 640

40

Lys Gln Gly Glu Tyr Lys Ile Thr Pro Ile Asp Asp Lys Ala Gln Phe  
645 650 655

Tyr Ile Phe Tyr Leu Lys Asp Ser Ala Ile Pro Tyr Ala Gln Phe Ile  
                   660                                  665                                  670

Leu Met Asp Lys Thr Met Phe Asn Ser Ala Tyr Val Gln Met Phe Phe  
                   675                                  680                                  685

Leu Gly Asn Tyr Asp Lys Asn Leu Phe Asp Leu Val Ile Asn Ser Arg  
                   690                                  695                                  700

10

Asp Ala Lys Val Phe Lys Leu Lys Ile Tyr Pro Tyr Asp Val Pro Asp  
       705                                  710                                  715                                  720

Tyr Ala

【0 3 4 1】

配列番号:27-pCC1FOS空のプラスミドのヌクレオチド配列

【0 3 4 2】

20

遺伝子座 MCSカセットを持つpCC1FOS

【0 3 4 3】

特徴 位置/限定子

【0 3 4 4】

領域 230..256

【0 3 4 5】

/ラベル="pCC1/pEpiFOS fwd"

【0 3 4 6】

領域 311..330

【0 3 4 7】

30

/ラベル="T7プロモーター"

【0 3 4 8】

領域相補体(504..529)

【0 3 4 9】

/ラベル="pCC1pEpiFOS rv"

【0 3 5 0】

CDS相補体(805..1464)

【0 3 5 1】

/ラベル=cat

【0 3 5 2】

40

CDS 1683..2030

【0 3 5 3】

/ラベル=redF

【0 3 5 4】

CDS 3425..4180

【0 3 5 5】

/ラベル=repE

【0 3 5 6】

CDS 4759..5934

【0 3 5 7】

50

/ ラベル=parA

【0358】

CDS 5934..6905

【0359】

/ ラベル=parB

【0360】

起源

【0361】

長さ：8171 bp

【0362】

種類：DNA環状、未注釈型(UNA)

【0363】

生物：人工

【0364】

配列：

【化30】

1 GCGGCCGCAA GGGGTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG  
CTTAACATATG

61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA  
CCGCACAGAT

121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA  
ACTGTTGGGA

181 AGGGCGATCG GTGCGGGCCT CTTGCTATT ACGCCAGCTG GCGAAAGGGG  
GATGTGCTGC

241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA  
AAACGACGGC

301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTCGAG CTCGGTACCC  
GGGGATCCCA

361 CGTGGCGCGC CACTAGTGCT AGCGACGTCG TGGGATCCTC TAGAGTCGAC  
CTGCAGGCAT

421 GCAAGCTTGA GTATTCTATA GTCTCACCTA AATAGCTTGG CGTAATCATG  
GTCATAGCTG

481 TTTCCTGTGT GAAATTGTTA TCCGCTCACA ATTCCACACA ACATACGAGC  
CGGAAGCATA

541 AAGTGTAAG CCTGGGGTGC CTAATGAGTG AGCTAACTCA CATTAAATTGC  
GTTGCGCTCA

601 CTGCCCCTT TCCAGTCGGG AAACCTGTCG TGCCAGCTGC ATTAATGAAT  
CGGCCAACGC

661 GAACCCCTTG CGGCCGCCCG GGCCGTCGAC CAATTCTCAT GTTTGACAGC  
TTATCATCGA

721 ATTTCTGCCA TTCATCCGCT TATTATCACT TATTCAGGCG TAGCAACCAG  
GCGTTTAAGG

781 GCACCAATAA CTGCCTTAAA AAAATTACGC CCCGCCCTGC CACTCATCGC  
AGTACTGTTG

841 TAATTCATTA AGCATTCCTGC CGACATGGAA GCCATCACAA ACGGCATGAT  
GAACCTGAAT

901 CGCCAGCGGC ATCAGCACCT TGTCGCCTTG CGTATAATAT TTGCCCATGG  
TGAAAACGGG

961 GGCGAAGAAG TTGTCCATAT TGGCCACGTT TAAATCAAAA CTGGTGAAAC  
TCACCCAGGG

10

20

30

40

1021 ATTGGCTGAG ACGAAAAACA TATTCTCAAT AAACCCTTTA GGGAAATAGG  
 CCAGGTTTTTC  
 1081 ACCGTAACAC GCCACATCTT GCGAATATAT GTGTAGAAAC TGCCGGAAAT  
 CGTCGTGGTA  
 1141 TTCACTCCAG AGCGATGAAA ACGTTTCAGT TTGCTCATGG AAAACGGTGT  
 AACAAGGGTG  
 1201 AACACTATCC CATATCACCA GCTCACCGTC TTTCATTGCC ATACGAAATT  
 CCGGATGAGC  
 1261 ATTCATCAGG CGGGCAAGAA TGTGAATAAA GGCCGGATAA AACTTGTGCT  
 TATTTTTCTT  
 1321 TACGGTCTTT AAAAAGGCCG TAATATCCAG CTGAACGGTC TGGTTATAGG  
 TACATTGAGC  
 1381 AACTGACTGA AATGCCTCAA AATGTTCTTT ACGATGCCAT TGGGATATAT  
 CAACGGTGGT  
 1441 ATATCCAGTG ATTTTTTTCT CCATTTTAGC TTCCTTAGCT CCTGAAAATC  
 TCGATAACTC  
 1501 AAAAAATACG CCCGGTAGTG ATCTTATTTT ATTATGGTGA AAGTTGGAAC  
 CTCTTACGTG  
 1561 CCGATCAACG TCTCATTTTC GCCAAAAGTT GGCCCAGGGC TTCCCGGTAT  
 CAACAGGGAC  
 1621 ACCAGGATTT ATTTATTCTG CGAAGTGATC TTCCGTCACA GGTATTTATT  
 CGCGATAAGC  
 1681 TCATGGAGCG GCGTAACCGT CGCACAGGAA GGACAGAGAA AGCGCGGATC  
 TGGGAAGTGA  
 1741 CGGACAGAAC GGTCAGGACC TGGATTGGGG AGGCGGTTGC CGCCGCTGCT  
 GCTGACGGTG  
 1801 TGACGTTCTC TGTTCCGGTC ACACCACATA CGTTCCGCCA TTCCTATGCG  
 ATGCACATGC  
 1861 TGTATGCCGG TATACCGCTG AAAGTTCTGC AAAGCCTGAT GGGACATAAG  
 TCCATCAGTT  
 1921 CAACGGAAGT CTACACGAAG GTTTTTGCGC TGGATGTGGC TGCCCGGCAC  
 CGGGTGACGT  
 1981 TTGCGATGCC GGAGTCTGAT GCGGTTGCGA TGCTGAAACA ATTATCCTGA  
 GAATAAATGC  
 2041 CTTGGCCTTT ATATGGAAAT GTGGAAGTGA GTGGATATGC TGTTTTTGTC  
 TGTTAAACAG  
 2101 AGAAGCTGGC TGTTATCCAC TGAGAAGCGA ACGAAACAGT CGGGAAAATC  
 TCCATTATC  
 2161 GTAGAGATCC GCATTATTAA TCTCAGGAGC CTGTGTAGCG TTTATAGGAA  
 GTAGTGTCT  
 2221 GTCATGATGC CTGCAAGCGG TAACGAAAAC GATTTGAATA TGCCTTCAGG  
 AACAATAGAA  
 2281 ATCTTCGTGC GGTGTTACGT TGAAGTGGAG CGGATTATGT CAGCAATGGA  
 CAGAACAACC  
 2341 TAATGAACAC AGAACCATGA TGTGGTCTGT CCTTTTACAG CCAGTAGTGC  
 TCGCCGCAGT  
 2401 CGAGCGACAG GGCGAAGCCC TCGGCTGGTT GCCCTCGCCG CTGGGCTGGC  
 GGCCGTCTAT  
 2461 GGCCCTGCAA ACGCGCCAGA AACGCCGTCG AAGCCGTGTG CGAGACACCG  
 CGGCCGGCCG

10

20

30

40

2521 CCGGCGTTGT GGATACCTCG CGGAAACTT GGCCCTCACT GACAGATGAG  
 GGGCGGACGT  
 2581 TGACACTTGA GGGGCCGACT CACCCGGCGC GGC GTTGACA GATGAGGGGC  
 AGGCTCGATT  
 2641 TCGGCCGGCG ACGTGGAGCT GGCCAGCCTC GCAAATCGGC GAAAACGCCT  
 GATTTTACGC  
 2701 GAGTTTCCCA CAGATGATGT GGACAAGCCT GGGGATAAGT GCCCTGCGGT  
 ATTGACACTT  
 2761 GAGGGGCGCG ACTACTGACA GATGAGGGGC GCGATCCTTG ACACTTGAGG  
 GGCAGAGTGC  
 2821 TGACAGATGA GGGGCGCACC TATTGACATT TGAGGGGCTG TCCACAGGCA  
 GAAAATCCAG  
 2881 CATTTGCAAG GGTTCGCGC CGTTTTTCGG CCACCGCTAA CCTGTCTTTT  
 AACCTGCTTT  
 2941 TAAACCAATA TTTATAAACC TTGTTTTTAA CCAGGGCTGC GCCCTGTGCG  
 CGTGACCGCG  
 3001 CACGCCGAAG GGGGGTGCCC CCCCTTCTCG AACCTCCCG GTCGAGTGAG  
 CGAGGAAGCA  
 3061 CCAGGGAACA GCACTTATAT ATTCTGCTTA CACACGATGC CTGAAAAAAC  
 TTCCCTTGGG  
 3121 GTTATCCACT TATCCACGGG GATATTTTAA TAATTATTTT TTTTATAGTT  
 TTTAGATCTT  
 3181 CTTTTTTAGA GCGCCTTGTA GGCCTTTATC CATGCTGGTT CTAGAGAAGG  
 TGTTGTGACA  
 3241 AATTGCCCTT TCAGTGTGAC AAATCACCTT CAAATGACAG TCCTGTCTGT  
 GACAAATTGC  
 3301 CCTTAACCCT GTGACAAATT GCCCTCAGAA GAAGCTGTTT TTTCACAAAG  
 TTATCCCTGC  
 3361 TTATTGACTC TTTTTTATTT AGTGTGACAA TCTAAAACT TGTCACACTT  
 CACATGGATC  
 3421 TGTCATGGCG GAAACAGCGG TTATCAATCA CAAGAAACGT AAAAATAGCC  
 CGCGAATCGT  
 3481 CCAGTCAAAC GACCTCACTG AGGCGGCATA TAGTCTCTCC CGGGATCAAA  
 AACGTATGCT  
 3541 GTATCTGTTT GTTGACCAGA TCAGAAAATC TGATGGCACC CTACAGGAAC  
 ATGACGGTAT  
 3601 CTGCGAGATC CATGTTGCTA AATATGCTGA AATATTCGGA TTGACCTCTG  
 CGGAAGCCAG  
 3661 TAAGGATATA CGGCAGGCAT TGAAGAGTTT CGCGGGGAAG GAAGTGGTTT  
 TTTATCGCCC  
 3721 TGAAGAGGAT GCCGGCGATG AAAAAGGCTA TGAATCTTTT CCTTGGGTTA  
 TCAAACGTGC  
 3781 GCACAGTCCA TCCAGAGGGC TTTACAGTGT ACATATCAAC CCATATCTCA  
 TTCCCTTCTT  
 3841 TATCGGGTTA CAGAACCGGT TTACGCAGTT TCGGCTTAGT GAAACAAAAG  
 AAATCACCAA  
 3901 TCCGTATGCC ATGCGTTTAT ACGAATCCCT GTGTCAGTAT CGTAAGCCGG  
 ATGGCTCAGG  
 3961 CATCGTCTCT CTGAAAATCG ACTGGATCAT AGAGCGTTAC CAGCTGCCTC  
 AAAGTTACCA

10

20

30

40

4021 GCGTATGCCT GACTTCCGCC GCCGCTTCCT GCAGGTCTGT GTTAATGAGA  
 TCAACAGCAG  
 4081 AACTCCAATG CGCCTCTCAT ACATTGAGAA AAAGAAAGGC CGCCAGACGA  
 CTCATATCGT  
 4141 ATTTTCCTTC CGCGATATCA CTTCCATGAC GACAGGATAG TCTGAGGGTT  
 ATCTGTCACA  
 4201 GATTTGAGGG TGGTTCGTCA CATTTGTTCT GACCTACTGA GGGTAATTTG  
 TCACAGTTTT  
 4261 GCTGTTTCCT TCAGCCTGCA TGGATTTTCT CATACTTTTT GAACTGTAAT  
 TTTTAAGGAA  
 4321 GCCAAATTTG AGGGCAGTTT GTCACAGTTG ATTTCTTCT CTTTCCCTTC  
 GTCATGTGAC  
 4381 CTGATATCGG GGGTTAGTTC GTCATCATTG ATGAGGGTTG ATTATCACAG  
 TTTATTACTC  
 4441 TGAATTGGCT ATCCGCGTGT GTACCTCTAC CTGGAGTTTT TCCCACGGTG  
 GATATTTCTT  
 4501 CTTGCGCTGA GCGTAAGAGC TATCTGACAG AACAGTTCTT CTTTGCTTCC  
 TCGCCAGTTC  
 4561 GCTCGCTATG CTCGGTTACA CGGCTGCGGC GAGCGCTAGT GATAATAAGT  
 GACTGAGGTA  
 4621 TGTGCTCTTC TTATCTCCTT TTGTAGTGTT GCTCTTATTT TAAACAACCT  
 TGCGGTTTTT  
 4681 TGATGACTTT GCGATTTTGT TGTGCTTTG CAGTAAATTG CAAGATTTAA  
 TAAAAAACG  
 4741 CAAAGCAATG ATTAAAGGAT GTTCAGAATG AAATCATGG AAACACTTAA  
 CCAGTGCATA  
 4801 AACGCTGGTC ATGAAATGAC GAAGGCTATC GCCATTGCAC AGTTTAATGA  
 TGACAGCCCG  
 4861 GAAGCGAGGA AAATAACCCG GCGCTGGAGA ATAGGTGAAG CAGCGGATTT  
 AGTTGGGGTT  
 4921 TCTTCTCAGG CTATCAGAGA TGCCGAGAAA GCAGGGCGAC TACCGCACCC  
 GGATATGGAA  
 4981 ATTCGAGGAC GGGTTGAGCA ACGTGTTGGT TATACAATTG AACAAATTAA  
 TCATATGCGT  
 5041 GATGTGTTTG GTACGCGATT GCGACGTGCT GAAGACGTAT TTCCACCGGT  
 GATCGGGGTT  
 5101 GCTGCCATA AAGGTGGCGT TTACAAAACC TCAGTTTCTG TTCATCTTGC  
 TCAGGATCTG  
 5161 GCTCTGAAGG GGCTACGTGT TTTGCTCGTG GAAGGTAACG ACCCCCAGGG  
 AACAGCCTCA  
 5221 ATGTATCACG GATGGGTACC AGATCTTCAT ATTCATGCAG AAGACACTCT  
 CCTGCCTTTC  
 5281 TATCTTGGGG AAAAGGACGA TGTCATTAT GCAATAAAGC CCACTTGCTG  
 GCCGGGGCTT  
 5341 GACATTATTC CTTCTGTCT GGCTCTGCAC CGTATTGAAA CTGAGTTAAT  
 GGGCAAATTT  
 5401 GATGAAGGTA AACTGCCCAC CGATCCACAC CTGATGCTCC GACTGGCCAT  
 TGAAACTGTT  
 5461 GCTCATGACT ATGATGTCAT AGTTATTGAC AGCGCGCCTA ACCTGGGTAT  
 CGGCACGATT

10

20

30

40

5521 AATGTCGTAT GTGCTGCTGA TGTGCTGATT GTTCCCACGC CTGCTGAGTT  
 GTTTGACTAC  
 5581 ACCTCCGCAC TGCAGTTTTT CGATATGCTT CGTGATCTGC TCAAGAACGT  
 TGATCTTAAA  
 5641 GGGTTCGAGC CTGATGTACG TATTTTGCTT ACCAAATACA GCAATAGTAA  
 TGGCTCTCAG  
 5701 TCCCCGTGGA TGGAGGAGCA AATTCGGGAT GCCTGGGGAA GCATGGTTCT  
 AAAAAATGTT  
 5761 GTACGTGAAA CGGATGAAGT TGGTAAAGGT CAGATCCGGA TGAGAACTGT  
 TTTTGAACAG  
 5821 GCCATTGATC AACGCTCTTC AACTGGTGCC TGGAGAAATG CTCTTTCTAT  
 TTGGGAACCT  
 5881 GTCTGCAATG AAATTTTCGA TCGTCTGATT AAACCACGCT GGGAGATTAG  
 ATAATGAAGC  
 5941 GTGCGCCTGT TATTCCAAAA CACACGCTCA ATACTCAACC GGTGGAAGAT  
 ACTTCGTTAT  
 6001 CGACACCAGC TGCCCCGATG GTGGATTCGT TAATTGCGCG CGTAGGAGTA  
 ATGGCTCGCG  
 6061 GTAATGCCAT TACTTTGCCT GTATGTGGTC GGGATGTGAA GTTTACTCTT  
 GAAGTGCTCC  
 6121 GGGGTGATAG TGTTGAGAAG ACCTCTCGGG TATGGTCAGG TAATGAACGT  
 GACCAGGAGC  
 6181 TGCTTACTGA GGACGCACTG GATGATCTCA TCCCTTCTTT TCTACTGACT  
 GGTCAACAGA  
 6241 CACCGGCGTT CGGTCGAAGA GTATCTGGTG TCATAGAAAT TGCCGATGGG  
 AGTCGCCGTC  
 6301 GTAAAGCTGC TGCATTACC GAAAGTGATT ATCGTGTTCT GGTGCGGAG  
 CTGGATGATG  
 6361 AGCAGATGGC TGCATTATCC AGATTGGGTA ACGATTATCG CCCAACAAGT  
 GCTTATGAAC  
 6421 GTGGTCAGCG TTATGCAAGC CGATTGCAGA ATGAATTTGC TGGAAATATT  
 TCTGCGCTGG  
 6481 CTGATGCGGA AAATATTTCA CGTAAGATTA TTACCCGCTG TATCAACACC  
 GCCAAATTGC  
 6541 CTAAATCAGT TGTTGCTCTT TTTTCTCACC CCGGTGAACT ATCTGCCCCG  
 TCAGGTGATG  
 6601 CACTTCAAAA AGCCTTTACA GATAAAGAGG AATTACTTAA GCAGCAGGCA  
 TCTAACCTTC  
 6661 ATGAGCAGAA AAAAGCTGGG GTGATATTTG AAGCTGAAGA AGTTATCACT  
 CTTTTAACTT  
 6721 CTGTGCTTAA AACGTCATCT GCATCAAGAA CTAGTTTAAAG CTCACGACAT  
 CAGTTTGCTC  
 6781 CTGGAGCGAC AGTATTGTAT AAGGGCGATA AAATGGTGCT TAACCTGGAC  
 AGGTCTCGTG  
 6841 TTCCAACCTGA GTGTATAGAG AAAATTGAGG CCATTCTTAA GGAACCTGAA  
 AAGCCAGCAC  
 6901 CCTGATGCGA CCACGTTTTA GTCTACGTTT ATCTGTCTTT ACTTAATGTC  
 CTTTGTTACA  
 6961 GGCCAGAAAG CATAACTGGC CTGAATATTC TCTCTGGGCC CACTGTTCCA  
 CTTGTATCGT

10

20

30

40

7021 CGGTCTGATA ATCAGACTGG GACCACGGTC CCACTCGTAT CGTCGGTCTG  
 ATTATTAGTC  
 7081 TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA GTCTGGGACC  
 ACGGTCCCAC  
 7141 TCGTATCGTC GGTCTGATAA TCAGACTGGG ACCACGGTCC CACTCGTATC  
 GTCGGTCTGA  
 7201 TTATTAGTCT GGGACCATGG TCCCACTCGT ATCGTCGGTC TGATTATTAG  
 TCTGGGACCA  
 7261 CGGTCCCCTACT CGTATCGTCG GTCTGATTAT TAGTCTGGAA CCACGGTCCC  
 ACTCGTATCG  
 7321 TCGGTCTGAT TATTAGTCTG GGACCACGGT CCCACTCGTA TCGTCGGTCT 10  
 GATTATTAGT  
 7381 CTGGGACCAC GATCCCACTC GTGTTGTCGG TCTGATTATC GGTCTGGGAC  
 CACGGTCCCA  
 7441 CTTGTATTGT CGATCAGACT ATCAGCGTGA GACTACGATT CCATCAATGC  
 CTGTCAAGGG  
 7501 CAAGTATTGA CATGTCGTCG TAACCTGTAG AACGGAGTAA CCTCGGTGTG  
 CGGTTGTATG  
 7561 CCTGCTGTGG ATTGCTGCTG TGTCTGCTT ATCCACAACA TTTTGCGCAC  
 GGTTATGTGG  
 7621 ACAAATACC TGGTTACCCA GGCCGTGCCG GCACGTTAAC CGGGCTGCAT  
 CCGATGCAAG 20  
 7681 TGTGTCGCTG TCGACGAGCT CGCGAGCTCG GACATGAGGT TGCCCCGTAT  
 TCAGTGTGCG  
 7741 TGATTTGTAT TGTCTGAAGT TGTTTTTACG TTAAGTTGAT GCAGATCAAT  
 TAATACGATA  
 7801 CCTGCGTCAT AATTGATTAT TTGACGTGGT TTGATGGCCT CCACGCACGT  
 TGTGATATGT  
 7861 AGATGATAAT CATTATCACT TTACGGGTCC TTTCCGGTGA TCCGACAGGT  
 TACGGGGCGG  
 7921 CGACCTCGCG GGTTTTCGCT ATTTATGAAA ATTTTCCGGT TTAAGGCGTT  
 TCCGTTCTTC  
 7981 TTCGTCATAA CTTAATGTTT TTATTTAAAA TACCCTCTGA AAAGAAAGGA 30  
 AACGACAGGT  
 8041 GCTGAAAGCG AGCTTTTTTG CCTCTGTCGT TTCCTTTCTC TGTTTTTGTC  
 CGTGGAATGA  
 8101 ACAATGGAAG TCCGAGCTCA TCGCTAATAA CTTCGTATAG CATAATTAT  
 ACGAAGTTAT  
 8161 ATTCGATCCA C

## 【0365】

配列番号:28-pCC1FOSの切断(pFOS)及びZ3206を持たないシゲラ・フレックスネリ6 0-抗原のヌクレオチド配列

## 【0366】

遺伝子座 pFOSの切断及び0-抗原の切断(-Z3206)

## 【0367】

定義 NheI及びAscIで切断されたMCSカセットを持つ逆位pCC1FOSの、NheI及びAscIで切断されgalFNheI及びwzzAscIで増幅されたS.フレックスネリ6 0抗原クラスターへのライゲーション

## 【0368】

特徴 位置/限定子

## 【0369】

CDS 3..411

## 【0370】

10

20

30

40

50

/ラベル='galF  
【0 3 7 1】  
CDS 784..1869  
【0 3 7 2】  
/ラベル=rmlB  
【0 3 7 3】  
CDS 1869..2768  
【0 3 7 4】  
/ラベル=rmlD  
【0 3 7 5】 10  
CDS 2826..3704  
【0 3 7 6】  
/ラベル=rmlA  
【0 3 7 7】  
CDS 3709..4266  
【0 3 7 8】  
/ラベル=rmlC  
【0 3 7 9】  
CDS 4263..5495  
【0 3 8 0】 20  
/ラベル=wzx  
【0 3 8 1】  
CDS 5551..6738  
【0 3 8 2】  
/ラベル=wzy  
【0 3 8 3】  
CDS 6755..7624  
【0 3 8 4】  
/ラベル=wfbY  
【0 3 8 5】 30  
CDS 7621..8454  
【0 3 8 6】  
/ラベル=wfbZ  
【0 3 8 7】  
CDS 8559..9965  
【0 3 8 8】  
/ラベル=gnd  
【0 3 8 9】  
CDS 10187..11380  
【0 3 9 0】 40  
/ラベル=ugd  
【0 3 9 1】  
CDS相補体(11446..12450)  
【0 3 9 2】  
/ラベル=uge  
【0 3 9 3】  
CDS 12802..12828  
【0 3 9 4】  
/ラベル=wzz'  
【0 3 9 5】 50

領域相補体(12868..12887)  
【0396】  
/ラベル="T7プロモーター"  
【0397】  
領域相補体(12942..12968)  
【0398】  
/ラベル="pCC1/pEpiFOS fwd"  
【0399】  
CDS相補体(14460..15431)  
【0400】 10  
/ラベル=parB  
【0401】  
CDS相補体(15431..16606)  
【0402】  
/ラベル=parA  
【0403】  
CDS相補体(17185..17940)  
【0404】  
/ラベル=repE  
【0405】 20  
CDS相補体(19335..19682)  
【0406】  
/ラベル=redF  
【0407】  
CDS 19901..20560  
【0408】  
/ラベル=cat  
【0409】  
領域 20836..20861  
【0410】 30  
/ラベル="pCC1pEpiFOS rv"  
【0411】  
長さ：20982 bp  
【0412】  
種類：DNA環状、未注釈型(UNA)  
【0413】  
配列：  
【化31】  
1 CTAGCGGCAA AACGTATGCC GGGTGACCTC TCTGAATACT CCGTCATCCA  
GACCAAAGAA 40  
61 CCGCTGGATC GCGAAGGTAA AGTCAGCCGC ATTGTTGAAT TTATCGAAAA  
ACCGGATCAG  
121 CCGCAGACGC TGGACTCAGA CATCATGGCC GTTGGTCGCT ATGTGCTTTC  
TGCCGATATT  
181 TGGCCGGAAC TTGAACGTAC TCAGCCTGGT GCATGGGGAC GTATTCAGCT  
GACTGATGCC

241 ATTGCCGAGC TGGCGAAAAA ACAGTCCGTT GATGCAATGC TGATGACCGG  
CGACAGCTAC

301 GACTGCGGTA AAAAAATGGG CTATATGCAG GCGTTTGTGA AGTATGGGCT  
GCGCAACCTG

361 AAAGAAGGGG CGAAGTTCCG TAAAGGTATT GAGAAGCTGT  
TAAGCGAATA ATGAAAATCT

421 GACCGGATGT AACGGTTGAT AAGAAAATTA TAACGGCAGT GAAGATTCGT  
GGTGAAAGTA

10

481 ATTTGTTGCG AATATTCCTG CCGTTGTTTT ATATAAACAA TCAGAATAAC  
AACGAGTTAG

541 CAATAGGATT TTAGTCAAAG TTTTCCAGGA TTTTCCTTGT TTCCAGAGCG  
GATTGGTAAG

601 ACAATTAGCT TTTGAATTTT TCGGGTTTAG CGCGAGTGGG TAACGCTCGT  
CACATCGTAG

661 GCATGCATGC AGTGCTCTGG TAGCTGTAAA GCCAGGGGCG GTAGCGTGCA  
TTAATACTTC

20

721 TATTAATCAA ACTGAGAGCC GCTTATTTCA CAGCATGCTC TGAAGCAATA  
TGGAATAAAT

781 TAGGTGAAAA TACTTGTTAC TGGTGGCGCA GGATTTATTG GTTTTGCTGT  
AGTTCGTCAC

841 ATTATAAATA ATACGCAGGA TAGTGTTGTT AATGTCGATA AATTAACGTA  
CGCCGGAAAC

901 CTGGAATCAC TTGCTGATGT TTCTGATTCT GAACGCTATG TTTTGAACA  
TGCGGATATT

30

961 TGCGATGCAG CTGCAATGGC ACGGATTTTT GCTCAGCATC AGCCAGATGC  
AGTGATGCAC

1021 CTGGCTGCTG AAAGCCATGT TGACCGTTCA ATTACAGGTC CTGCGGCATT  
TATTGAAACC

1081 AATATTGTTG GTACATATGT CCTTTTGGAA GCCGCTCGCA ATTATTGGTC  
TGCTCTTGAT

40

1141 AGCGACAAGA AAAC TAGATT CCGTTTTTCAT CATATTTCTA CTGACGAAGT  
CTATGGTGAT

1201 TTGCCTCATC CTGACGAGGT AAATAATACA GAAGAATTAC CCTTATTAC  
AGAGACAACA

1261 GCTTACGCGC CAAGCAGCCC TTATTCCGCT TCAAAAGCAT CCAGCGATCA  
TTTAGTCCGC

1321 GCGTGGAAC GTACCTATGG TTTACCAACC ATTGTGACTA ATTGCTCTAA  
TAATTATGGT

1381 CCTTATCATT TCCCGGAAAA ATTGATTCCA TTGGTTATTC TGAATGCTCT  
GGAAGGTAAG

1441 GCATTACCTA TTTATGGCAA AGGGGATCAA ATTCGTGACT GGCTGTATGT  
TGAAGATCAT

10

1501 GCGCGTGCGT TATATACCGT CGTAACCGAA GGTAAAGCGG  
GTGAAACTTA TAACATTGGT

1561 GGACACAACG AAAAGAAAAA CATCGATGTA GTGCTCACTA TTTGTGATTT  
GCTGGATGAG

1621 ATTGTACCGA AAGAGAAATC TTACCGCGAG CAAATTACTT ATGTTGCCGA  
TCGCCCCGGA

1681 CACGATCGCC GTTATGCGAT TGATGCAGAG AAGATTAGCC  
GCGAATTGGG CTGGAAACCG

20

1741 CAGGAAACGT TTGAGAGCGG GATTCGGAAG ACATTGGAAT  
GGTACCTGTC CAATACAAAA

1801 TGGGTTGATA ATGTGAAAAG TGGTGCTTAT CAATCGTGGA TTGAACAGAA  
CTATGAGGGC

1861 CGCCAGTAAT GAATATCCTC CTTTTCGGCA AAACAGGGCA GGTAGGTTGG  
GAACTACAGC

1921 GTGCTCTGGC ACCTTTGGGT AATTTGATTG CTCTTGATGT TCACTCCACT  
GATTATTGTG

30

1981 GTGATTTTAG TAATCCTGAA GGTGTAGCTG AAACAGTCAA AAGAATTCGA  
CCTGATGTTA

2041 TTGTTAATGC TCGGGCTCAC ACCGCAGTAG ATAAGGCTGA GTCAGAACCC  
GAATTTGCAC

2101 AATTACTCAA TGCGACTAGT GTTGAATCAA TTGCAAAAGA GGCTAATGAA  
GTTGGGGCTT

40

2161 GGGTAATTCA TTA CTCAACT GACTACGTAT TCCCTGGAAA TGGCGACACG  
CCATGGCTGG

2221 AGACGGATGC AACCGCACCG CTAAATGTTT ACGGTGAAAC  
CAAGTTAGCC GGAGAAAAAG

2281 CGTTACAGGA ACATTGCGCG AAGCATCTTA TTTTCCGTAC CAGCTGGGTA  
TACGCAGCTA

2341 AAGGAAATAA CTTGCGCCAAA ACGATGTTGC GTCTGGCAAA  
AGAGCGCGAA GAACTGGCTG

2401 TGATAAATGA TCAATTTGGT GCGCCAACAG GTGCTGAGCT GCTGGCTGAT  
TGTACGGCAC

2461 ATGCTATTCG TGTGGCACTG AATAAACCGG AAGTCGCAGG TTTGTACCAT  
CTGGTAGCCA

10

2521 GTGGTACCAC AACCTGGCAC GATTATGCTG CGCTGGTTTT TGAAGAGGCG  
CGCAAAGCAG

2581 GTATTCCCCT TGCACTCAAC AAGCTCAACG CAGTACCAAC AACAGCCTAT  
CCTACACCAG

2641 CTCGTCGTCC ACATAACTCT CGCCTTAATA CAGAAAAATT TCAGCAGAAC  
TTTGCGCTTG

2701 TCTTGCCTGA CTGGCAGGTT GGTGTGAAAC GAATGCTCAA CGAATTAATT  
ACGACTACAG

20

2761 CAATTTAATA GTTTTTGCAT CTTGTTCGTG ATGGTGGAGC AAGATGAATT  
AAAAGGAATG

2821 ATGAAATGAA AACGCGTAAA GGTATTATTT TAGCGGGTGG TTCTGGTACA  
CGTCTTTATC

2881 CTGTGACTAT GGCTGTCAGT AAACAGCTAT TACCTATTTA TGATAAGCCG  
ATGATCTATT

2941 ACCCGCTCTC TACACTGATG TTGGCGGGTA TTCGCGATAT TCTGATTATT  
AGTACGCCAC

30

3001 AGGATACTCC TCGTTTTCAA CAACTGCTAG GTGACGGTAG CCAGTGGGGG  
CTAAATCTTC

3061 AGTACAAAGT GCAACCGACT CCAGATGGGC TTGCGCAGGC GTTTATTATC  
GGTGAAGAGT

3121 TTATCGGTGG TGATGATTGT GCTTTGGTTC TTGGTGATAA TATCTTCTAC  
GGTCATGATC

40

3181 TGCCGAAGTT AATGGATGTC GCTGTTAACA AAGAAAGTGG  
TGCAACGGTA TTTGCCTATC

3241 ACGTTAATGA TCCTGAACGC TACGGCGTCG TTGAGTTTGA TAAAAACGGT  
ACGGCAATAA

3301 GCCTGGAAGA AAAACCGCTA CAACCAAAAA GTAATTATGC  
GGTAACCGGG CTTTATTTCT

3361 ATGATAACGA CGTTGTCGAA ATGGCGAAAA ACCTTAAGCC TTCTGCCCCG  
GGTGAAGTGG

3421 AAATTACCGA TATTAACCGT ATTTATATGG AACAGGGGCG TTTATCCGTT  
GCCATGATGG

3481 GCGTGTTA TGCATGGCTG GATACGGGGA CACATCAGAG TCTTATTGAA  
GCAAGCAACT

10

3541 TCATTGCCAC CATTGAAGAG CGCCAGGGAC TAAAGGTTTC CTGCCCAGAA  
GAAATTGCTT

3601 ACCGTAAAGG GTTTATTGAT GCTGAACAGG TGAAAGCATT  
AGCGGAGCCG CTGAAAAAAA

3661 ATGCTTATGG ACAGTATCTG CTGAAAATGA TTAAAGGTTA TTAATAAAAT  
GAACGTAATT

3721 AAAACAGAAA TTCCTGATGT GTTAATTTTC GAGCCGAAAG TTTTGGTGA  
TGAGCGTGGT

20

3781 TTCTTTATGG AAAGCTTTAA TCAGAAAGTT TTCGAAGAAG CTGTAGGACG  
TAAGGTTGAA

3841 TTTGTTTCAAG ATAACCATTC GAAGTCTAGT AAAGGTGTTT TACGCGGGCT  
GCATTATCAG

3901 TTAGAACCTT ATGCGCAAGG GAAACTGGTA CGTTGCGTTG TTGGTGAGGT  
TTTTGATGTA

3961 GCTGTTGATA TTCGTAAATC GTCGCCTACC TTTGGTAAAT GGGTTGGGGT  
GAATTTATCT

30

4021 GCTGAGAATA AGCGGCAATT GTGGATCCCT GAGGGATTTG CACATGGTTT  
TTTGGTGCTG

4081 AGCGAGACTG CGGAATTTTT ATATAAAACG ACGAACTATT ATCATCCTGA  
TAGTGATAGA

4141 GGGATTGTAT GGAATGATCC TATTCTGAGC ATAAAATGGC CGACGATAGA  
ACATAATAAT

4201 TATATTTTAT CGATTAAAGA TGCAAGGGCT AAAGAATTGC ATAACATGAA  
GGAATTATTT

40

4261 TTGTGAGTAT TGTAAGAAT ACTTTATGGA ATATAAGTGG GTATATTATA  
CCATCATTA

4321 TAGCAATTCC TGC GTTAGGT ATACTGTCTA GAATTCTAGG GACCGAGCAA  
TTTGGCCTTT

4381 TTACGTTAGC TATTGCCTTA GTTGGATATG CAAGTATTTT TGATGCTGGA  
TTGACCAGAG

4441 CTGTTATAAG AGAAGTATCA ATATATAAAA ATGTTCATAA AGAATTAAGA  
GCGATCATTT

4501 CAACTTCAAC GGTAATTCTA ACTATATTGG GCTTGATTGG CGGTAGTGTA  
CTATTTTTGA

10

4561 GTAGCAATGT AATTGTTAAA TTATTAAACA TTAACGCGAA TCATGTTGTA  
GAATCTGTCA

4621 AAGCAATATA TATTATTTCA GCTACCATAC CCTTATACTT GTTAAACCAA  
GTCTGGTTGG

4681 GGATTTTTGA GGGGATGGAA AAGTTCAGAA AAGTAAATTT  
AATAAAATCA ATTAACAAC

4741 CTTTTGTGGC TGGATTACCA GTGATTTTCT GTTTTTTTCA TGGAGGATTA  
CTAAGTGCTA

20

4801 TATATGGTTT AGTTATGGCA AGAGTCTTAT CACTTATAGT GACCTTTATA  
TTTAGTCGAA

4861 AACTAATAAT ATCATCTGGG CTGTCTGTAA AAATTGTAAC AGTTAAAAGA  
TTAATCGGCT

4921 TTGGAAGCTG GATAACAGTT AGCAATATTA TTAGCCCTAT TATGACATAT  
ATGGATCGTT

4981 TTATTCTTTC ACACATTGTG GGGGCTGATA AAGTTTCTTT TTATACTGCT  
CCGTCTGAAG

30

5041 GTATACAACG CTTAACGATA TTACCAAGTG CGTTGTCCAG AGCTATTTTT  
CCAAGATTAA

5101 GTTCAGAATT GCAATCGGTA AAGCAAATA AAATATTATC ATATTTTATA  
ATGGTTATTG

5161 GTATACTTCC AATTGTAATG TTGATAATTA TTTTATCAGA TTTTATAATG  
TCCGCTTGGA

5221 TGGGACCTAC ATATCATGGG ACGCCAGGTA TAGTATTAAA AATTCTTGCA  
ATAGGTTTCT

40

5281 TTTTAAATTG CATTGCACAA ATCCCATTTG TTTCAGTTCA GGCTAGTGGA  
AGATCAAAAA

5341 TTACAGCTAT TATTCATTTG CTCGAAGTTA TCCCATATTT ATGCATATTA  
TATATTTTAA

5401 TTTATCATTG GGAATTGTT GGAGCCGCAA TAGCATGGTC TGTAAGAACA  
TCGTTAGATT

5461 TTTTGATATT ATTATTAATT GATACGAAAT ATTAATAGCG AATTGATTTT  
AGGGATTACT

5521 TCCTCAAGCC CATCTAATTA GAGTGCAAAC ATGACTTCTG ATTTTTATAA  
CTCAAAAGAC

10

5581 AAAAGTTTAA GTGTTCTTTT GTTTTTTGGG TTTATATTTT TCCTTACACG  
TAGCTTTCCA

5641 TTTATTCAAT ATAGTTGGAT TATGGAGGGG TTTTATGTC TTTGTATCAT  
GTCATTTACA

5701 AAGAAAATTG CAAACGGAAT ATATCACTAT CCTGTTATTT TAATATTTCT  
ATTAGCTCTT

5761 TTTATAAATT TTATTTATTC CTATATCAAG GGTAACGATA TAGCGATAAT  
AATTAGGTTT

20

5821 TATATTATCA TATTATTTAT ATTATGTGCT TATTTCTGCT CTTATGGAAC  
CATCTCGATT

5881 GTTAAAATAT TTTTATATTT AATGGTATTA CAGGCGGTTA TTATATCCAT  
CATTAGTATT

5941 TATATGACAA AAACATATGG TATTGGTGAT TATTCAGCAC TAAGACATTA  
TTTTTTGGAG

6001 AATGATTATG GTGATGTTTA TACATATGGA AGTGGTTTCT ATAGAGTTCA  
AATTAAAGGA

30

6061 AATGCTCTCA TTCCATTTGC CTTTATGTTG CATATAGTCA TAAAAGATTA  
TTTCTATTAT

6121 CGATTCAAAA ATACAATAAC CGTTATTCTG GCTATAGGTA CTATAGTGGC  
TGGTAATTTT

6181 GCATATTTTG TTTCGATATG CTGTGTTTTT ATGTATATTA TACTATGTTC  
TAAATCTAAC

6241 TCACGATACG CTAAATTAAG GAAAATTATT TTTGGGGTTT TTCTTACTGT  
GATTCTCCCT

40

6301 TTTTTTATTA CATATTCAAT TGAGTTGATA ATCATGAAAT CAAATGGAGC  
TGATTCTTCT

6361 TTAGGAGTTA GATGGGATCA GTTACTGTA TTAATTAATG ATCTTACAGA  
GTCTGTATCA

6421 AATTTTGTTA TAGGTTCTGG TTTGGGTAAT GTCATCAAAA TTCAAACCTCC  
TATCCGTGAT

6481 TATAGTGCAT ATATATATTA TGAATTGCAG TCAGTTTATT TTTTAAATCA  
ACTTGGCGTT

6541 ATTTTATTTA CTTTGTTTTT ATTAATTAAT CTCCTTCTCA CGATTAAAAT  
CATAAAATAC

10

6601 AGTGAGTTGT GTGTGCTATA TTTTCTATAT GTTCTTATG CAATTACTAA  
TCCTTATATT

6661 TTAGACTCTA ACCATGTTGC TGTAATAATT GTATTAGTGA CATTAAGTAA  
TGTCTAAAA

6721 AAGATGAAAG CTAAATGAAG GTTTTAAGGT GAAGATGGAC  
ACTGTATATG CCGTTTTGGT

6781 TGCTTACAAC CCAGAACATA ATGATTTAAA AAATGCGGTT GAATTATTGT  
TGAGACAAGT

20

6841 TACTAAAGTT GTCGTTTGCA ATAACCTCTAC AAATGGTTAT AAATATGCTG  
AAAATTCTTC

6901 AGGCGATGTA AAAATATTCA ATTTCAATGA TAATTTAGGC ATAGCAGAAG  
CCCAAAGTAT

6961 AGGAATGAAA TGGGCTTTTG AAAATGGCGC TGATTTTATA TTGCAAATGG  
ATCAGGATAG

7021 TATTCCTGAT CCTAAGATGG TAGAGCAGTT ACTTACTTGT TACAAAAAAT  
TGCTTAAACA

30

7081 AAATGTCAAT GTTGGTTTAG TTGGTTCACA AGATTTTGAT AAAGTAACTG  
GTGAATTAAA

7141 TAAAGCAAGG GTAAAAAAG GGAAACCACT TACAGAAGTT  
TATTATGAGG TAGATAGTAC

7201 ATTAAGTTCT GGCAGTCTAA TACCAAAAAA TAGTTGGTTG ATTGTTGGAG  
GAATGAAAGA

7261 TGAGCTTTTT ATCGATGCGG TAGACCATGA ATATTGTTGG AGATTAAGAG  
CTGCTGGGTT

40

7321 TAAAGTAATT AGGAATAAAA ATGCGTTACT TGCACATAGA CTTGGAGATG  
GGCGATTAA

7381 GATCTTAAAT ATTCTTTCTG TCGGTTTGCC AAGCCCATTT CGTCATTATT  
ATGCTACTCG

7441 AAATATCTTT CTTTTATTAA ATAAAAATTA TGTACCCATC TACTGGAAAA  
TTTCTAGTCT

7501 GGTAAATTA ATTGGAAAGG TTTTTTTATA TCCTATTTTC CTTCCAAATG  
GTAATAAAAG

7561 GTTATATTTT TTTTAAAAG GCATTAATGA CGGTTTAATG GGTCGAAGTG  
GTAAAATGAA

10

7621 ATGAATCATA GATTAGAAAA ATTCTCAGTT TTAATTAGCA TTTATAAAAA  
TGATCTACCG

7681 CAATTTTTTG AGGTGGCTCT ACGCTCTATT TTTCACGATC AAACACTTAA  
GCCAGATCAA

7741 ATAGTAATTG TTGCAGATGG AGAACTCCAT CAAACACACA TCGATATTAT  
AAATTCATTCT

7801 ATTGATGATG TTGGCAATAA AATAGTAACA TTTGTACCTT TACCTAGAAA  
TGTTGGATTG

20

7861 GCTAATGCCT TAAATGAAGG ATTAAAGGCT TGTAGGAATG AGTTAGTGGC  
AAGAATGGAT

7921 GCTGATGATA TTTCTTTGCC TCATCGGTTT GAGAAACAAA TTTCTTTTAT  
GATTAATAAT

7981 TCAGAAATAG ATGTATGTGG CAGTTTTATT GATGAAATTG AACTGTGTAC  
TGAGGAGTTT

8041 ATTTCAACAC GCAAAGTGCC TCTCGAACAT AGAGAAATAG TTAAATTTCG  
GAGGAAACGA

30

8101 AGCGCAGTTA GCCATCCTTC TGTAATTTTT AGAAAGAATA CAGTATTAGC  
TGTTGGTGGT

8161 TATCCTCCAT TCAGAAAATC TCAAGATTTT GCATTGTGGA GCCTATTAAT  
TGTACATAAT

8221 GCAAGATTTG CAAATCTTCC AGATATTTTA TTAAAAATGC GAACTGGTCG  
TAATCTTATG

8281 GCTCGACGTG GATTGTCATA TTTATTGTAC GAGTATAAAG TATTGTATTA  
TCAATATAAA

40

8341 ATTGGTTTTA TTCGAAAAAA TGAATTAATA AGTAATGCTA TGTTGAGAAC  
ATTTTTTCGT

8401 ATAATGCCAT CTAAATTAAA GGAGCTGATG TATTCAATCG TTAGGAATCG  
ATAATAATAA

8461 TTTTCTGATT AAGTGTTATG GATTTATTTT TATTAGGCAT ATTCTATAAT  
TAAGCATAAC

8521 CCGCATACCA CCCAGCGGTA TCCTGACAGG AGTAAACAAT  
GTCAAAGCAA CAGATCGGCG

8581 TCGTCGGTAT GGCAGTGATG GGGCGCAACC TTGCGCTCAA TATCGAAAGC  
CGTGGTTATA

10

8641 CCGTCTCTAT TTTCAACCGT TCCCGTGAAA AGACCGAAGA AGTGATTACC  
GAAAATCCAG

8701 GCAAGAAACT GGTTCTTAC TATACGGTGA AAGAATTTGT TGAATCTCTG  
GAAACGCCTC

8761 GTCGCATCCT GTTAATGGTG AAAGCAGGTG CTGGCACGGA TGCTGCTATT  
GATTCCCTCA

8821 AGCCATACCT CGATAAAGGT GACATCATCA TTGATGGTGG TAACACCTTC  
TTCCATGACA

20

8881 CCATTCGTCG TAACCGTGAG CTTTCTGCAG AAGGCTTTAA CTTTATCGGT  
ACCGGTGTTT

8941 CCGGTGGTGA AGAAGGTGCG CTGAAAGGTC CTTCCATTAT GCCTGGTGGG  
CAGAAAGAAG

9001 CTTATGAACT GATTGCGCCG ATCCTGACCA AAATCGCCGC TGTGGCTGAA  
GACGGCGAAC

9061 CGTGCGTTAC CTATATTGGT GCCGATGGTG CAGGTCATTA TGTGAAGATG  
GTTTACAACG

30

9121 GTATTGAATA CGGTGATATG CAGCTGATTG CTGAAGCCTA TTCTCTGCTT  
AAAGGTGGCT

9181 TGAACCTCAC CAACGAAGAA CTGGCGCAGA CCTTTACCGA  
GTGGAATAAC GGTGAACTGA

9241 GCAGCTACCT GATCGACATC ACCAAAGATA TCTTCACCAA  
AAAAGATGAA GAGGGTAACT

9301 ACCTGGTTGA TGTGATTCTG GATGAAGCAG CAAACAAAGG  
TACGGGCAAA TGGACCAGCC

40

9361 AGAGCGCGCT GGATCTCGGC GAACCGCTGT CGCTGATTAC CGAGTCTGTG  
TTTGACGTT

9421 ATATCTCTTC TCTGAAAGAG CAGCGTGTTG CCGCATCTAA AGTTCTCTCT  
GGCCCCGAAG

9481 CGCAGCCAGC TGGCGACAAT GCTGAGTTCA TCGAAAAAGT  
TCGCCGTGCG CTGTATCTGG

9541 GCAAAATCGT TTCTTACGCT CAGGGCTTCT CTCAGCTACG CGCTGCGTCT  
GAAGAGTACA

9601 ACTGGGATCT GAACTACGGT GAAATCGCGA AGATTTTCCG TGCTGGCTGC  
ATCATCCGTG

10

9661 CGCAGTTCCT GCAGAAAATC ACCGATGCTT ATGCCGAAAA TCCGCAGATC  
GCTAACCTGT

9721 TGCTGGCTCC TTA CTTC AAG CAAATTGCCG ATGACTACCA GCAGGCGCTG  
CGCGATGTCG

9781 TCGCTTACGC AGTACAGAAC GGTATCCCGG TGCCGACCTT CGCCGCTGCG  
GTTGCCTATT

9841 ACGACAGCTA CCGCGCCGCT GTTCTGCCTG CGAACCTGAT CCAGGCACAG  
CGTGACTATT

20

9901 TCGGTGCGCA TACTTATAAG CGCATTGATA AAGAAGGTGT GTTCCATACC  
GAATGGCTGG

9961 ATTAATCTGA TTAAATCAA TTAATCAAAG CAAGGCCCGG AGAAACCCTC  
CGGGCTTTTT

10021 TATTATACAA AGCGGCAGGT TAGGGCCTTT TTTTATAATT TATAGTTAAA  
AACGCGATAT

10081 AATACAGCGC CGCACAGCAG GATCGCTGCC TTGACAGTTC  
ATCTACATCA GCGTTAAAAA

30

10141 TCCCGCAGTA GATGAAGCTG TGGTGGTGGA TTAATGACCA CTCTAAATGT  
TTAACCGGAA

10201 GAAGTCAGAG CTAATGAAAA TAACAATTTC AGGAACAGGT  
TATGTTGGTC TTCAAATGG

10261 TATTCTGATT GCGCAAAACC ACGAAGTGGT TGC ACTGGAT ATCGTTCAGG  
CCAAAGTGGA

40

10321 CATGCTTAAC AAGAGGCAGT CACCGCTTGT TGATAAGGAG  
ATTGAAGAGT ATCTGGCGAC

10381 TAAAGATCTC AATTTCCGCG CTACGACAGA TAAGTATGAC  
GCGTATAAAA ATGCCGATTA

10441 CGTTATTATT GCCACACCTA CCGATTATGA TCCGAAAACA AATTATTTTA  
ATACCTCAAG

10501 CGTGGAAGCG GTCATTCGTG ATGTGACAGA AATTAATCCC  
AACGCGGTAA TGATTATAAA

10561 ATCAACTATC CCTGTTGGTT TTACAGAGTC CATTAAAGAA CGTTTTGGTA  
TTGAAAATGT

10621 GATCTTTTCG CCTGAGTTTT TCGTGAAGG TAAAGCACTT TATGATAACT  
TACACCCATC

10

10681 ACGCATTGTG ATTGGCGAGC AGTCTGAACG CGCTAAACGT TTTGCTGCGT  
TATTACAGGA

10741 AGGCGCCATT AAGCAAGACA TACCAACATT GTTTACTGAC  
TCAACCGAGG CTGAGGCGAT

10801 TAAACTTTTT GCGAACACTT ATCTGGCGAT GCGTGTAGCG TATTTCAATG  
AACTTGATAG

10861 TTATGCTGAA AGCCTGGGAC TTAATTCACG CCAGATTATT GAGGGCGTAT  
GCCTTGACCC

20

10921 GCGTATCGGT AATCACTACA ACAACCCGTC ATTCGGTTAT GGTGGTTATT  
GTCTGCCGAA

10981 AGATACTAAG CAGTTACTGG CAAATTACCA GTCTGTGCCG  
AATAACCTGA TCTCGGCAAT

11041 TGTTGACGCC AACCGCACGC GCAAAGATTT TATTGCCGAT TCTATCCTTG  
CACGTAAACC

11101 GAAAGTTGTT GGCGTCTATC GTTTGATTAT GAAGAATGGT TCAGACAATT  
TTCGTGCTTC

30

11161 CTCGATTCAG GGTATTATGA AGCGAATCAA GGCGAAAGGT  
GTGCCTGTAA TCGTTTATGA

11221 GCCAGCTATG AAAGAGGACG ATTTTTTCCG GTCGCGCGTG GTACGTGATC  
TGGATGCGTT

11281 CAAACAAGAA GCTGATGTTA TTATTTCTAA CCGTATGTCT GCCGATCTGG  
CTGATGTAGC

40

11341 AGATAAAGTT TATACGCGCG ACTTGTTTGG CAATGATTAA TTATTTTGTT  
TCATTCTAAG

11401 AAAAGGCCCT AATAAATTAG GGCCTTTTCT TATGGTTTTG TAAAATCAAA  
CTTTATAGAA

11461 GTTACGATAC CATTCTACAA AGTTCTTTAC CCCTTCTTTA ACTGACGTTT  
CAGGTTTGAA

11521 TCCTATTACG TCATACAGTG CTTTTGTATC AGCACTGGTT TCCAGTACAT  
CACCGGGTTG

11581 GAGAGGCATC ATATTTTTGT TGGCTTCAAT ACCCAGAGCC TCTTCTAACG  
CATTGATATA

11641 GTCCATCAAC TCCACAGGCG AACTATTACC AATGTTATAG ACACGATATG  
GTGCTGAACT

10

11701 TGTTGCAGGC GAGCCTGTTT CTACAGCCCA CTGTGGGTTT TTTTCTGGAA  
TAACATCCTG

11761 TAAGCGAATA ATAGCTTCGG CAATATCATC AATGTAAGTA  
AAGTCACGCT TCATTTTGCC

11821 GAAGTTGTAA ACATCAATGC TTTTACCTTC CAGCATGGCT TTAGTGAATT  
TAAATAATGC

11881 CATATCCGGA CGTCCCCATG GACCATAAAC CGTAAAGAAA  
CGCAGCCCTG TGGTCGGTAA

20

11941 GCCATACAAA TGAGAATATG TATGGGCCAT GAGTTCATTC GCTTTTTTAG  
TTGCTGCATA

12001 AAGCGAAACA GGATGATCTA CAGAGTCATC TGTAGAGAAA  
GGCATCTTGC GGTTTCATGCC

12061 ATAAACAGAA CTGGAGGAAG CGTAAAGTAG ATGCTGAACA  
TTATTATGGC GACATCCTTC

12121 TAGTATGTTC AGGAATCCAA TCAGGTTTGC ATCTGCATAT GCATTGGGAT  
TTTCAAGAGA

30

12181 GTAACGTACA CCGGCTTGCG CAGCGAGGTT TATTACGCGT TCGAACCGCT  
CGTCTGCAAA

12241 CAGTGCCGCC ATTTTCTCAC GATCGGCCAG GTCAATTTTA TAAAAACTGA  
AGTTGTCGTG

12301 CTTGAGTAAA TCAAGTCGTG CTTGTTTGAG GTTGACATCG TAATAATCAT  
TTAAGTTGTC

40

12361 AATGCCTACA ACCTGATGAC CAGCTGCAAG AAGCCGTTTA  
CTTAGATAGA AACCGATAAA

12421 GCCAGCAGCT CCCGTAACCA GAAATTTTCAT TTATAATCCT CGCTCAGGCT  
AGAATATAGC

12481 CAATCTTCAT CTGGCATAAC TGAAAGTTAA ATTATACCGT TAGACAAGAA  
AAAAAGATAA

12541 TCGGTATCAG TTCTAAACTT GGCTGTTTTT TCTGGTAACG TGCTCATTTT  
ACAATCAAAG

12601 CTGTTCTAAG CTGACTATAC AAGCCGACGT CATTATCTCC AACCGTATGG  
CAGAAGAGCT

12661 TAAGGATGTG GCAGACAAAG TCTACACCCG CGATCTCTTT  
GGCAGTGA CT AACATCCTGT

10

12721 TATCATGGCG ATTTTCGCCC TGATTCTCTT ATGTTCCCTT TGTAATAATT  
CATTATTTTT

12781 ATCATTTATC CTATAGCATT CATGGCGATT ATCGCTAAAC TATGGCGGCG  
CGCCACGTGG

12841 GATCCCCGGG TACCGAGCTC GAATTCGCCC TATAGTGAGT CGTATTACAA  
TTCACCTGGC

12901 GTCGTTTTAC AACGTCGTGA CTGGGAAAC CCTGGCGTTA CCCAACTTAA  
TCGCCTTGCA

20

12961 GCACATCCCC CTTTCGCCAG CTGGCGTAAT AGCGAAGAGG  
CCCGCACCGA TCGCCCTTCC

13021 CAACAGTTGC GCAGCTGAAT GGCGAATGGC GCCTGATGCG GTATTTTCTC  
CTTACGCATC

13081 TGTGCGGTAT TTCACACCGC ATATGGTGCA CTCTCAGTAC AATCTGCTCT  
GATGCCGCAT

13141 AGTTAAGCCA GCCCCGACAC CCGCCAACAC CCGCTGACGC  
GAACCCCTTG CGGCCGCATC

30

13201 GAATATAACT TCGTATAATG TATGCTATAC GAAGTTATTA GCGATGAGCT  
CGGACTTCCA

13261 TTGTTTCATC CACGGACAAA AACAGAGAAA GGAAACGACA  
GAGGCCAAAA AGCTCGCTTT

13321 CAGCACCTGT CGTTTCCTTT CTTTTCAGAG GGTATTTTAA ATAAAAACAT  
TAAGTTATGA

13381 CGAAGAAGAA CGGAAACGCC TTAAACCGGA AAATTTTCAT  
AAATAGCGAA AACCCGCGAG

40

13441 GTCGCCGCCC CGTAACCTGT CGGATCACCG GAAAGGACCC  
GTAAAGTGAT AATGATTATC

13501 ATCTACATAT CACAACGTGC GTGGAGGCCA TCAAACCACG  
TCAAATAATC AATTATGACG

13561 CAGGTATCGT ATTAATTGAT CTGCATCAAC TTAACGTAAA AACAACTTCA  
GACAATACAA

13621 ATCAGCGACA CTGAATACGG GGCAACCTCA TGTCCGAGCT  
CGCGAGCTCG TCGACAGCGA

13681 CACACTTGCA TCGGATGCAG CCCGGTTAAC GTGCCGGCAC  
GGCCTGGGTA ACCAGGTATT

10

13741 TTGTCCACAT AACCGTGCGC AAAATGTTGT GGATAAGCAG  
GACACAGCAG CAATCCACAG

13801 CAGGCATACA ACCGCACACC GAGGTTACTC CGTTCTACAG  
GTTACGACGA CATGTCAATA

13861 CTTGCCCTTG ACAGGCATTG ATGGAATCGT AGTCTCACGC TGATAGTCTG  
ATCGACAATA

13921 CAAGTGGGAC CGTGGTCCCA GACCGATAAT CAGACCGACA  
ACACGAGTGG GATCGTGGTC

20

13981 CCAGACTAAT AATCAGACCG ACGATACGAG TGGGACCGTG  
GTCCCAGACT AATAATCAGA

14041 CCGACGATAC GAGTGGGACC GTGGTTCCAG ACTAATAATC  
AGACCGACGA TACGAGTGGG

14101 ACCGTGGTCC CAGACTAATA ATCAGACCGA CGATACGAGT  
GGGACCATGG TCCCAGACTA

14161 ATAATCAGAC CGACGATACG AGTGGGACCG TGGTCCCAGT  
CTGATTATCA GACCGACGAT

30

14221 ACGAGTGGGA CCGTGGTCCC AGACTAATAA TCAGACCGAC  
GATACGAGTG GGACCGTGGT

14281 CCCAGACTAA TAATCAGACC GACGATACGA GTGGGACCGT  
GGTCCCAGTC TGATTATCAG

14341 ACCGACGATA CAAGTGGAAC AGTGGGCCCA GAGAGAATAT  
TCAGGCCAGT TATGCTTTCT

14401 GGCCTGTAAC AAAGGACATT AAGTAAAGAC AGATAAACGT  
AGACTAAAAC GTGGTTCGAT

40

14461 CAGGGTGCTG GCTTTTCAAG TTCCTTAAGA ATGGCCTCAA TTTTCTCTAT  
ACACTCAGTT

14521 GGAACACGAG ACCTGTCCAG GTTAAGCACC ATTTTATCGC CCTTATACAA  
TACTGTCGCT

14581 CCAGGAGCAA ACTGATGTCG TGAGCTTAAA CTAGTTCTTG  
ATGCAGATGA CGTTTTAAGC

14641 ACAGAAGTTA AAAGAGTGAT AACTTCTTCA GCTTCAAATA  
TCACCCCAGC TTTTTTCTGC

14701 TCATGAAGGT TAGATGCCTG CTGCTTAAGT AATTCCTCTT TATCTGTAAA  
GGCTTTTTGA

10

14761 AGTGCATCAC CTGACCGGGC AGATAGTTCA CCGGGGTGAG  
AAAAAAGAGC AACAACTGAT

14821 TTAGGCAATT TGGCGGTGTT GATACAGCGG GTAATAATCT TACGTGAAAT  
ATTTTCCGCA

14881 TCAGCCAGCG CAGAAATATT TCCAGCAAAT TCATTCTGCA ATCGGCTTGC  
ATAACGCTGA

14941 CCACGTTTAT AAGCACTTGT TGGGCGATAA TCGTTACCCA ATCTGGATAA  
TGCAGCCATC

20

15001 TGCTCATCAT CCAGCTCGCC AACCAGAACA CGATAATCAC TTTCGGTAAG  
TGCAGCAGCT

15061 TTACGACGGC GACTCCCATC GGCAATTTCT ATGACACCAG ATACTCTTCG  
ACCGAACGCC

15121 GGTGTCTGTT GACCAGTCAG TAGAAAAGAA GGGATGAGAT  
CATCCAGTGC GTCCTCAGTA

15181 AGCAGCTCCT GGTCACGTTT ATTACCTGAC CATACCCGAG AGGTCTTCTC  
AACACTATCA

30

15241 CCCCAGGAGCA CTTCAAGAGT AACTTCACA TCCCGACCAC  
ATACAGGCAA AGTAATGGCA

15301 TTACCGCGAG CCATTACTCC TACGCGCGCA ATTAACGAAT CCACCATCGG  
GGCAGCTGGT

15361 GTCGATAACG AAGTATCTTC AACCGGTTGA GTATTGAGCG TATGTTTTGG  
AATAACAGGC

40

15421 GCACGCTTCA TTATCTAATC TCCAGCGTG GTTTAATCAG ACGATCGAAA  
ATTTCAATTGC

15481 AGACAGGTTT CCAAATAGAA AGAGCATTTT TCCAGGCACC  
AGTTGAAGAG CGTTGATCAA

15541 TGGCCTGTTC AAAAACAGTT CTCATCCGGA TCTGACCTTT ACCAACTTCA  
TCCGTTTCAC

15601 GTACAACATT TTTTAGAACC ATGCTTCCCC AGGCATCCCG AATTTGCTCC  
TCCATCCACG

15661 GGGACTGAGA GCCATTACTA TTGCTGTATT TGGTAAGCAA  
AATACGTACA TCAGGCTCGA

15721 ACCCTTTAAG ATCAACGTTT TTGAGCAGAT CACGAAGCAT  
ATCGAAAAAC TGCAGTGCGG

10

15781 AGGTGTAGTC AAACAACTCA GCAGGCGTGG GAACAATCAG  
CACATCAGCA GCACATACGA

15841 CATTAATCGT GCCGATACCC AGGTTAGGCG CGCTGTCAAT AACTATGACA  
TCATAGTCAT

15901 GAGCAACAGT TTCAATGGCC AGTCGGAGCA TCAGGTGTGG  
ATCGGTGGGC AGTTTACCTT

15961 CATCAAATTT GCCCATTAAC TCAGTTTCAA TACGGTGCAG AGCCAGACAG  
GAAGGAATAA

20

16021 TGTCAGCCC CGGCCAGCAA GTGGGCTTTA TTGCATAAGT GACATCGTCC  
TTTTCCCCAA

16081 GATAGAAAGG CAGGAGAGTG TCTTCTGCAT GAATATGAAG  
ATCTGGTACC CATCCGTGAT

16141 ACATTGAGGC TGTTCCCTGG GGGTCGTTAC CTTCCACGAG CAAAACACGT  
AGCCCCTTCA

16201 GAGCCAGATC CTGAGCAAGA TGAACAGAAA CTGAGGTTTT  
GTAAACGCCA CCTTTATGGG

30

16261 CAGCAACCCC GATCACCGGT GGAAATACGT CTTCAGCACG  
TCGCAATCGC GTACCAAACA

16321 CATCACGCAT ATGATTAATT TGTTC AATTG TATAACCAAC ACGTTGCTCA  
ACCCGTCCTC

16381 GAATTTCCAT ATCCGGGTGC GGTAGTCGCC CTGCTTTCTC GGCATCTCTG  
ATAGCCTGAG

40

16441 AAGAAACCCC AACTAAATCC GCTGCTTCAC CTATTCTCCA GCGCCGGGTT  
ATTTTCCTCG

16501 CTTCCGGGCT GTCATCATTA AACTGTGCAA TGGCGATAGC CTTGTCATT  
TCATGACCAG

16561 CGTTTATGCA CTGGTTAAGT GTTCCATGA GTTTCATTCT GAACATCCTT  
TAATCATTGC

16621 TTTGCGTTTT TTTATTAAAT CTTGCAATTT ACTGCAAAGC AACAACAAAA  
TCGCAAAGTC

16681 ATCAAAAAC CGCAAAGTTG TTAAAATAA GAGCAACACT  
ACAAAAGGAG ATAAGAAGAG

16741 CACATACCTC AGTCACTTAT TATCACTAGC GCTCGCCGCA GCCGTGTAAC  
CGAGCATAGC

10

16801 GAGCGAACTG GCGAGGAAGC AAAGAAGAAC TGTTCGTCA  
GATAGCTCTT ACGCTCAGCG

16861 CAAGAAGAAA TATCCACCGT GGGAAAAACT CCAGGTAGAG  
GTACACACGC GGATAGCCAA

16921 TTCAGAGTAA TAACTGTGA TAATCAACCC TCATCAATGA  
TGACGAACTA ACCCCCGATA

16981 TCAGGTCACA TGACGAAGGG AAAGAGAAGG AAATCAACTG  
TGACAAACTG CCCTCAAATT

20

17041 TGGCTTCCTT AAAAATTACA GTTCAAAAAG TATGAGAAAA  
TCCATGCAGG CTGAAGGAAA

17101 CAGCAAACT GTGACAAATT ACCCTCAGTA GGTCAGAACA  
AATGTGACGA ACCACCCTCA

17161 AATCTGTGAC AGATAACCCT CAGACTATCC TGTCGTCATG GAAGTGATAT  
CGCGGAAGGA

17221 AAATACGATA TGAGTCGTCT GCGGCCTTT CTTTTTCTCA ATGTATGAGA  
GGCGCATTGG

30

17281 AGTTCGTCTG TTGATCTCAT TAACACAGAC CTGCAGGAAG  
CGGCGGCGGA AGTCAGGCAT

17341 ACGCTGGTAA CTTTGAGGCA GCTGGTAACG CTCTATGATC CAGTCGATTT  
TCAGAGAGAC

17401 GATGCCTGAG CCATCCGGCT TACGATACTG ACACAGGGAT  
TCGTATAAAC GCATGGCATA

17461 CGGATTGGTG ATTTCTTTTG TTTCATAAG CCGAACTGC GTAAACCGGT  
TCTGTAACCC

40

17521 GATAAAGAAG GGAATGAGAT ATGGGTTGAT ATGTACACTG  
TAAAGCCCTC TGGATGGACT

17581 GTGCGCACGT TTGATAAACC AAGGAAAAGA TTCATAGCCT TTTTCATCGC  
CGGCATCCTC

17641 TTCAGGGCGA TAAAAAACCA CTTCTTCCC CGCGAACTC TTCAATGCCT  
GCCGTATATC

17701 CTTACTGGCT TCCGCAGAGG TCAATCCGAA TATTTAGCA TATTTAGCAA  
CATGGATCTC

17761 GCAGATACCG TCATGTTCTT GTAGGGTGCC ATCAGATTTT CTGATCTGGT  
CAACGAACAG

10

17821 ATACAGCATA CGTTTTTGAT CCCGGGAGAG ACTATATGCC GCCTCAGTGA  
GGTCGTTTGA

17881 CTGGACGATT CGCGGGCTAT TTTACGTTT CTTGTGATTG ATAACCGCTG  
TTTCCGCCAT

17941 GACAGATCCA TGTGAAGTGT GACAAGTTTT TAGATTGTCA CACTAAATAA  
AAAAGAGTCA

18001 ATAAGCAGGG ATAACCTTGT GAAAAAACAG CTTCTTCTGA  
GGGCAATTTG TCACAGGGTT

20

18061 AAGGGCAATT TGTCACAGAC AGGACTGTCA TTTGAGGGTG  
ATTTGTCACA CTGAAAGGGC

18121 AATTTGTCAC AACACCTTCT CTAGAACCAG CATGGATAAA  
GGCCTACAAG GCGCTCTAAA

18181 AAAGAAGATC TAAAAACTAT AAAAAAATA ATTATAAAAA  
TATCCCCGTG GATAAGTGGA

18241 TAACCCCAAG GGAAGTTTTT TCAGGCATCG TGTGTAAGCA  
GAATATATAA GTGCTGTTCC

30

18301 CTGGTGCTTC CTCGCTCACT CGACCGGGAG GGTTCGAGAA  
GGGGGGGCAC CCCCTTCGG

18361 CGTGCGCGGT CACGCGCACA GGGCGCAGCC CTGGTTAAAA  
ACAAGGTTTA TAAATATTGG

18421 TTTAAAAGCA GGTAAAAGA CAGGTTAGCG GTGGCCGAAA  
AACGGGCGGA AACCTTGCA

18481 AATGCTGGAT TTTCTGCCTG TGGACAGCCC CTCAAATGTC AATAGGTGCG  
CCCCTCATCT

40

18541 GTCAGCACTC TGCCCCTCAA GTGTCAAGGA TCGCGCCCCT CATCTGTCAG  
TAGTCGCGCC

18601 CCTCAAGTGT CAATACCGCA GGGCACTTAT CCCCAGGCTT GTCCACATCA  
TCTGTGGGAA

18661 ACTCGCGTAA AATCAGGCGT TTTCGCCGAT TTGCGAGGCT GGCCAGCTCC  
ACGTCGCCGG

18721 CCGAAATCGA GCCTGCCCCT CATCTGTCAA CGCCGCGCCG  
GGTGAGTCGG CCCCTCAAGT

18781 GTCAACGTCC GCCCCTCATC TGTCAGTGAG GGCCAAGTTT TCCGCGAGGT  
ATCCACAACG

10

18841 CCGGCGGCCG GCCGCGGTGT CTCGCACACG GCTTCGACGG CGTTTCTGGC  
GCGTTTGCAG

18901 GGCCATAGAC GGCCGCCAGC CCAGCGGCGA GGGCAACCAG  
CCGAGGGCTT CGCCCTGTCG

18961 CTCGACTGCG GCGAGCACTA CTGGCTGTAA AAGGACAGAC  
CACATCATGG TTCTGTGTTC

19021 ATTAGGTTGT TCTGTCCATT GCTGACATAA TCCGCTCCAC TTCAACGTAA  
CACCGCACGA

20

19081 AGATTTCTAT TGTTCTGAA GGCATATTCA AATCGTTTTT GTTACCGCTT  
GCAGGCATCA

19141 TGACAGAACA CTACTTCCTA TAAACGCTAC ACAGGCTCCT GAGATTAATA  
ATGCGGATCT

19201 CTACGATAAT GGGAGATTTT CCCGACTGTT TCGTTCGCTT CTCAGTGGAT  
AACAGCCAGC

19261 TTCTCTGTTT AACAGACAAA AACAGCATAT CCACTCAGTT CCACATTTCC  
ATATAAAGGC

30

19321 CAAGGCATTT ATTCTCAGGA TAATTGTTTC AGCATCGCAA CCGCATCAGA  
CTCCGGCATC

19381 GCAAACCTGCA CCCGGTGCCG GGCAGCCACA TCCAGCGCAA  
AAACCTTCGT GTAGACTTCC

19441 GTTGAACCTGA TGGACTTATG TCCCATCAGG CTTTGCAGAA CTTTCAGCGG  
TATACCGGCA

40

19501 TACAGCATGT GCATCGCATA GGAATGGCGG AACGTATGTG  
GTGTGACCGG AACAGAGAAC

19561 GTCACACCGT CAGCAGCAGC GGCGGCAACC GCCTCCCCAA  
TCCAGGTCCT GACCGTTCTG

19621 TCCGTCACCTT CCCAGATCCG CGCTTTCTCT GTCCTTCCTG TGCGACGGTT  
ACGCCGCTCC

19681 ATGAGCTTAT CGCGAATAAA TACCTGTGAC GGAAGATCAC  
TTCGCAGAAT AAATAAATCC

19741 TGGTGTCCCT GTTGATACCG GGAAGCCCTG GGCCAACTTT TGGCGAAAAT  
GAGACGTTGA

19801 TCGGCACGTA AGAGGTTCCA ACTTTCACCA TAATGAAATA  
AGATCACTAC CGGGCGTATT

10

19861 TTTTGAGTTA TCGAGATTTT CAGGAGCTAA GGAAGCTAAA  
ATGGAGAAAA AAATCACTGG

19921 ATATACCACC GTTGATATAT CCCAATGGCA TCGTAAAGAA CATTTTGAGG  
CATTTCACTC

19981 AGTTGCTCAA TGTACCTATA ACCAGACCGT TCAGCTGGAT ATTACGGCCT  
TTTTAAAGAC

20041 CGTAAAGAAA AATAAGCACA AGTTTTATCC GGCTTTTATT CACATTCTTG  
CCCGCCTGAT

20

20101 GAATGCTCAT CCGGAATTTT GTATGGCAAT GAAAGACGGT  
GAGCTGGTGA TATGGGATAG

20161 TGTTACCCCT TGTTACACCG TTTTCCATGA GCAAACCTGAA ACGTTTTTCAT  
CGCTCTGGAG

20221 TGAATACCAC GACGATTTCG GGCAGTTTCT ACACATATAT TCGCAAGATG  
TGGCGTGTTA

20281 CGGTGAAAAC CTGGCCTATT TCCCTAAAGG GTTTATTGAG AATATGTTTT  
TCGTCTCAGC

30

20341 CAATCCCTGG GTGAGTTTCA CCAGTTTTGA TTAAACGTG GCCAATATGG  
ACAACTTCTT

20401 CGCCCCCGTT TTCACCATGG GCAAATATTA TACGCAAGGC  
GACAAGGTGC TGATGCCGCT

20461 GGCGATTCAAG GTTCATCATG CCGTTTGTGA TGGCTTCCAT GTCGGCAGAA  
TGCTTAATGA

20521 ATTACAACAG TACTGCGATG AGTGGCAGGG CGGGGCGTAA  
TTTTTTTAAG GCAGTTATTG

40

20581 GTGCCCTTAA ACGCCTGGTT GCTACGCCTG AATAAGTGAT  
AATAAGCGGA TGAATGGCAG

20641 AAATTCGATG ATAAGCTGTC AAACATGAGA ATTGGTCGAC  
GGCCCGGGCG GCCGCAAGGG

20701 GTTCGCGTTG GCCGATTCAT TAATGCAGCT GGCACGACAG GTTCCCGAC  
TGGAAAGCGG

20761 GCAGTGAGCG CAACGCAATT AATGTGAGTT AGCTCACTCA  
TTAGGCACCC CAGGCTTTAC

20821 ACTTTATGCT TCCGGCTCGT ATGTTGTGTG GAATTGTGAG CGGATAACAA  
TTTCACACAG

10

20881 GAAACAGCTA TGACCATGAT TACGCCAAGC TATTTAGGTG  
AGACTATAGA ATACTCAAGC

20941 TTGCATGCCT GCAGGTCGAC TCTAGAGGAT CCCACGACGT CG

【0414】

配列番号:29-pCC1FOSの切断(pFOS)及びZ3206を持つシゲラ・フレックスネリ60-抗原  
のヌクレオチド配列

【0415】

遺伝子座 pFOSの切断及び0-抗原の切断(Z3206+)

20

【0416】

定義 NheI及びAscIで切断されZ3206Nhe及びwzzAscIで増幅された逆位S. フレックスネ  
リ60-抗原クラスターの、NheI及びAscIで切断されたMCSカセットを持つpCC1FOSへのライ  
ゲーション

【0417】

特徴 位置/限定子

【0418】

CDS相補体 (370..396)

【0419】

/ラベル=wzz'

30

【0420】

CDS 748..1752

【0421】

/ラベル=uge

【0422】

CDS相補体(1818..3011)

【0423】

/ラベル=ugd

【0424】

CDS相補体(3233..4639)

40

【0425】

/ラベル=gnd

【0426】

CDS相補体(4744..5577)

【0427】

/ラベル=wfbZ

【0428】

CDS相補体(5574..6443)

【0429】

/ラベル=wfbY

50

|                      |    |
|----------------------|----|
| 【 0 4 3 0 】          |    |
| CDS相補体(6460..7647)   |    |
| 【 0 4 3 1 】          |    |
| /ラベル=wzy             |    |
| 【 0 4 3 2 】          |    |
| CDS相補体(7703..8935)   |    |
| 【 0 4 3 3 】          |    |
| /ラベル=wzx             |    |
| 【 0 4 3 4 】          |    |
| CDS相補体(8932..9489)   | 10 |
| 【 0 4 3 5 】          |    |
| /ラベル=rmlC            |    |
| 【 0 4 3 6 】          |    |
| CDS相補体(9494..10372)  |    |
| 【 0 4 3 7 】          |    |
| /ラベル=rmlA            |    |
| 【 0 4 3 8 】          |    |
| CDS相補体(10430..11329) |    |
| 【 0 4 3 9 】          |    |
| /ラベル=rmlD            | 20 |
| 【 0 4 4 0 】          |    |
| CDS相補体(11329..12414) |    |
| 【 0 4 4 1 】          |    |
| /ラベル=rmlB            |    |
| 【 0 4 4 2 】          |    |
| CDS相補体(12787..13680) |    |
| 【 0 4 4 3 】          |    |
| /ラベル=galF            |    |
| 【 0 4 4 4 】          |    |
| CDS相補体(13912..14907) | 30 |
| 【 0 4 4 5 】          |    |
| /ラベル=Z3206           |    |
| 【 0 4 4 6 】          |    |
| CDS相補体(15065..15097) |    |
| 【 0 4 4 7 】          |    |
| /ラベル='wcaM           |    |
| 【 0 4 4 8 】          |    |
| CDS相補体(15525..16184) |    |
| 【 0 4 4 9 】          |    |
| /ラベル=cat             | 40 |
| 【 0 4 5 0 】          |    |
| CDS 16403..16750     |    |
| 【 0 4 5 1 】          |    |
| /ラベル=redF            |    |
| 【 0 4 5 2 】          |    |
| CDS 18145..18900     |    |
| 【 0 4 5 3 】          |    |
| /ラベル=repE            |    |
| 【 0 4 5 4 】          |    |
| CDS 19479..20654     | 50 |

【 0 4 5 5 】

/ ラベル=parA

【 0 4 5 6 】

CDS 20654..21625

【 0 4 5 7 】

/ ラベル=parB

【 0 4 5 8 】

長さ : 22887 bp

【 0 4 5 9 】

種類 : DNA環状、未注釈型 (UNA)

10

【 0 4 6 0 】

配列 :

【 化 3 2 】

1 GCGGCCGCAA GGGGTTTCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG  
CTTAACTATG

61 CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA  
CCGCACAGAT

121 GCGTAAGGAG AAAATACCGC ATCAGGCGCC ATTCGCCATT CAGCTGCGCA  
ACTGTTGGGA

20

181 AGGGCGATCG GTGCGGGCCT CTTGCTATT ACGCCAGCTG GCGAAAGGGG  
GATGTGCTGC

241 AAGGCGATTA AGTTGGGTAA CGCCAGGGTT TTCCCAGTCA CGACGTTGTA  
AAACGACGGC

301 CAGTGAATTG TAATACGACT CACTATAGGG CGAATTTCGAG CTCGGTACCC  
GGGGATCCCA

361 CGTGGCGCGC CGCCATAGTT TAGCGATAAT CGCCATGAAT GCTATAGGAT  
AAATGATAAA

30

421 AATAATGAAT TATTACAAAG GGAACATAAG AGAATCAGGG  
CGAAAATCGC CATGATAACA

481 GGATGTTAGT CACTGCCAAA GAGATCGCGG GTGTAGACTT TGTCTGCCAC  
ATCCTTAAGC

541 TCTTCTGCCA TACGGTTGGA GATAATGACG TCGGCTTGTA TAGTCAGCTT  
AGAACAGCTT

601 TGATTGTAAA ATGAGCACGT TACCAGAAAA AACAGCCAAG TTTAGAACTG  
ATACCGATTA

40

661 TCTTTTTTTC TTGTCTAACG GTATAATTTA ACTTTCAGTT ATGCCAGATG  
AAGATTGGCT

721 ATATTCTAGC CTGAGCGAGG ATTATAAATG AAATTTCTGG TTACGGGAGC  
TGCTGGCTTT

781 ATCGGTTTCT ATCTAAGTAA ACGGCTTCTT GCAGCTGGTC ATCAGGTTGT  
AGGCATTGAC

841 AACTTAAATG ATTATTACGA TGTCAACCTC AAACAAGCAC GACTTGATTT  
ACTCAAGCAC

901 GACAACCTCA GTTTTTATAA AATTGACCTG GCCGATCGTG AGAAAATGGC  
GGCACTGTTT

10

961 GCAGACGAGC GGTTCGAACG CGTAATAAAC CTCGCTGCGC AAGCCGGTGT  
ACGTTACTCT

1021 CTTGAAAATC CCAATGCATA TGCAGATGCA AACCTGATTG GATTCCTGAA  
CATACTAGAA

1081 GGATGTCGCC ATAATAATGT TCAGCATCTA CTTTACGCTT CCTCCAGTTC  
TGTTTATGGC

1141 ATGAACCGCA AGATGCCTTT CTCTACAGAT GACTCTGTAG ATCATCCTGT  
TTCGCTTTAT

20

1201 GCAGCAACTA AAAAAGCGAA TGAAGTCATG GCCCATACAT ATTCTCATTT  
GTATGGCTTA

1261 CCGACCACAG GGCTGCGTTT CTTTACGGTT TATGGTCCAT GGGGACGTCC  
GGATATGGCA

1321 TTATTTAAAT TCACTAAAGC CATGCTGGAA GGTAAGCA TTGATGTTTA  
CAACTTCGGC

1381 AAAATGAAGC GTGACTTTAC TTACATTGAT GATATTGCCG AAGCTATTAT  
TCGCTTACAG

30

1441 GATGTTATTC CAGAAAAAAA CCCACAGTGG GCTGTAGAAA  
CAGGCTCGCC TGCAACAAGT

1501 TCAGCACCAT ATCGTGTCTA TAACATTGGT AATAGTTCGC CTGTGGAGTT  
GATGGACTAT

1561 ATCAATGCGT TAGAAGAGGC TCTGGGTATT GAAGCCAACA  
AAAATATGAT GCCTCTCCAA

1621 CCCGGTGATG TACTGGAAAC CAGTGCTGAT ACAAAGCAC TGTATGACGT  
AATAGGATTC

40

1681 AAACCTGAAA CGTCAGTTAA AGAAGGGGTA AAGAACTTTG  
TAGAATGGTA TCGTAACTTC

1741 TATAAAGTTT GATTTTACAA AACCATAAGA AAAGGCCCTA ATTTATTAGG  
GCCTTTTCTT

1801 AGAATGAAAC AAAATAATTA ATCATTGCCA AACAAGTCGC  
GCGTATAAAC TTTATCTGCT

1861 ACATCAGCCA GATCGGCAGA CATACGGTTA GAAATAATAA  
CATCAGCTTC TTGTTTGAAC

1921 GCATCCAGAT CACGTACCAC GCGCGACCGG AAAAAATCGT CCTCTTTCAT  
AGCTGGCTCA

10

1981 TAAACGATTA CAGGCACACC TTTCGCCTTG ATTCGCTTCA TAATACCCTG  
AATCGAGGAA

2041 GCACGAAAAT TGTCTGAACC ATTCTTCATA ATCAAACGAT AGACGCCAAC  
AACTTTCGGT

2101 TTACGTGCAA GGATAGAATC GGCAATAAAA TCTTTGCGCG TCGGGTTGGC  
GTCAACAATT

2161 GCCGAGATCA GGTTATTCGG CACAGACTGG TAATTTGCCA GTAACCTGCTT  
AGTATCTTTC

20

2221 GGCAGACAAT AACCACCATA ACCGAATGAC GGGTTGTTGT AGTGATTACC  
GATACGCGGG

2281 TCAAGGCATA CGCCCTCAAT AATCTGGCGT GAATTAAGTC CCAGGCTTTC  
AGCATAACTA

2341 TCAAGTTCAT TGAAATACGC TACACGCATC GCCAGATAAG TGTTCGCAAA  
AAGTTTAATC

2401 GCCTCAGCCT CGGTTGAGTC AGTAAACAAT GTTGGTATGT CTTGCTTAAT  
GGCGCCTTCC

30

2461 TGTAATAACG CAGCAAAACG TTTAGCGCGT TCAGACTGCT CGCCAATCAC  
AATGCGTGAT

2521 GGGTGTAAGT TATCATAAAG TGCTTTACCT TCACGCAAAA ACTCAGGCCA  
AAAGATCACA

2581 TTTTCAATAC CAAAACGTTC TTTAATGGAC TCTGTAAAAC CAACAGGGAT  
AGTTGATTTT

2641 ATAATCATT CCGCGTTGGG ATTAATTTCT GTCACATCAC GAATGACCGC  
TTCCACGCTT

40

2701 GAGGTATTAA AATAATTTGT TTTCGGATCA TAATCGGTAG GTGTGGCAAT  
AATAACGTAA

2761 TCGGCATTTT TATACGCGTC ATACTTATCT GTCGTAGCGC GGAAATTGAG  
ATCTTTAGTC

2821 GCCAGATACT CTTCAATCTC CTTATCAACA AGCGGTGACT GCCTCTTGTT  
AAGCATGTCC

2881 ACTTTGGCCT GAACGATATC CAGTGCAACC ACTTCGTGGT TTTGCGCAAT  
CAGAATACCA

2941 TTTGAAAGAC CAACATAACC TG TTCCTGAA ATTGTTATTT TCATTAGCTC  
TGACTTCTTC

10

3001 CGGTAAACA TTTAGAGTGG TCATTAATCC ACCACCACAG CTTCATCTAC  
TGCGGGATTT

3061 TTAACGCTGA TG TAGATGAA CTGTCAAGGC AGCGATCCTG CTGTGCGGCG  
CTGTATTATA

3121 TCGCGTTTTT AACTATAAAT TATAAAAAAA GGCCCTAACC TGCCGCTTTG  
TATAATAAAA

3181 AAGCCCGGAG GGTTTCTCCG GGCCTTGCTT TGATTAATTG ATTTAAATCA  
GATTAATCCA

20

3241 GCCATTCGGT ATGGAACACA CCTTCTTTAT CAATGCGCTT ATAAGTATGC  
GCACCGAAAT

3301 AGTCACGCTG TGCCTGGATC AGGTTTCGAG GCAGAACAGC  
GGCGCGGTAG CTGTCGTAAT

3361 AGGCAACCGC AGCGGCGAAG GTCGGCACCG GGATACCGTT  
CTGTACTGCG TAAGCGACGA

3421 CATCGCGCAG CGCCTGCTGG TAGTCATCGG CAATTTGCTT GAAGTAAGGA  
GCCAGCAACA

30

3481 GGTTAGCGAT CTGCGGATTT TCGGCATAAG CATCGGTGAT TTTCTGCAGG  
AACTGCGCAC

3541 GGATGATGCA GCCAGCACGG AAAATCTTCG CGATTTACAC GTAGTTCAGA  
TCCCAGTTGT

3601 ACTCTTCAGA CGCAGCGCGT AGCTGAGAGA AGCCCTGAGC  
GTAAGAAACG ATTTTGCCCA

40

3661 GATACAGCGC ACGGCGAACT TTTTCGATGA ACTCAGCATT GTCGCCAGCT  
GGCTGCGCTT

3721 GCGGGCCAGA GAGAACTTTA GATGCGGCAA CACGCTGCTC  
TTTCAGAGAA GAGATATAAC

3781 GTGCAAACAC AGACTCGGTA ATCAGCGACA GCGGTTGCGC  
GAGATCCAGC GCGCTCTGGC

3841 TGGTCCATTT GCCCGTACCT TTGTTTGCTG CTTTCATCCAG AATCACATCA  
ACCAGGTAGT

3901 TACCCTCTTC ATCTTTTTTG GTGAAGATAT CTTTGGTGAT GTCGATCAGG  
TAGCTGCTCA

3961 GTTCACCGTT ATTCCACTCG GTAAAGGTCT GCGCCAGTTC TTCGTTGGTG  
AGGTTCAAGC

10

4021 CACCTTTAAG CAGAGAATAG GCTTCAGCAA TCAGCTGCAT ATCACCGTAT  
TCAATACCGT

4081 TGTGAACCAT CTTACATAA TGACCTGCAC CATCGGCACC AATATAGGTA  
ACGCACGGTT

4141 CGCCGTCTTC AGCCACAGCG GCGATTTTGG TCAGGATCGG CGCAATCAGT  
TCATAAGCTT

4201 CTTTCTGCCC ACCAGGCATA ATGGAAGGAC CTTTCAGCGC ACCTTCTTCA  
CCACCGGAAA

20

4261 CACCGGTACC GATAAAGTTA AAGCCTTCTG CAGAAAGCTC ACGGTTACGA  
CGAATGGTGT

4321 CATGGAAGAA GGTGTTACCA CCATCAATGA TGATGTCACC TTTATCGAGG  
TATGGCTTGA

4381 GGGAATCAAT AGCAGCATCC GTGCCAGCAC CTGCTTTCAC CATTAACAGG  
ATGCGACGAG

4441 GCGTTTCCAG AGATTCAACA AATTCTTTCA CCGTATAGTA AGGAACCAGT  
TTCTTGCCTG

30

4501 GATTTTCGGT AATCACTTCT TCGGTCTTTT CACGGGAACG GTTGAAAATA  
GAGACGGTAT

4561 AACCACGGCT TTCGATATTG AGCGCAAGGT TCGCCCCAT CACTGCCATA  
CCGACGACGC

4621 CGATCTGTTG CTTTGACATT GTTACTCCT GTCAGGATAC CGCTGGGTGG  
TATGCGGGTT

4681 ATGCTTAATT ATAGAATATG CCTAATAAAA ATAAATCCAT AACACTTAAT  
CAGAAAATTA

40

4741 TTATTATCGA TTCCTAACGA TTGAATACAT CAGCTCCTTT AATTTAGATG  
GCATTATACG

4801 AAAAAATGTT CTCAACATAG CATTACTTAT TAATTCATTT TTTCGAATAA  
AACCAATTTT

4861 ATATTGATAA TACAATACTT TATACTCGTA CAATAAATAT GACAATCCAC  
GTCGAGCCAT

4921 AAGATTACGA CCAGTTCGCA TTTTAAATAA AATATCTGGA AGATTGCAA  
ATCTTGCAAT

4981 ATGTACAATT AATAGGCTCC ACAATGCAAA ATCTTGAGAT TTTCTGAATG  
GAGGATAACC

10

5041 ACCAACAGCT AATACTGTAT TCTTTCTAAA AATTACAGAA GGATGGCTAA  
CTGCGCTTCG

5101 TTTCCTCGCG AATTAACTA TTTCTCTATG TTCGAGAGGC ACTTTGCGTG  
TTGAAATAAA

5161 CTCCTCAGTA ACAGTTTCAA TTTCATCAAT AAAACTGCCA CATACTCTA  
TTTCTGAATT

5221 ATTAATCATA AAAGAAATTT GTTTCTCAA CCGATGAGGC AAAGAAATAT  
CATCAGCATC

20

5281 CATTCTTGCC ACTAACTCAT TCCTACAAGC CTTAATCCT TCATTAAAGG  
CATTAGCCAA

5341 TCCAACATTT CTAGGTAAAG GTACAAATGT TACTATTTTA TTGCCAACAT  
CATCAATGAA

5401 TGAATTTATA ATATCGATGT GTGTTTGATG GAGTTCTCCA TCTGCAACAA  
TTACTATTTG

5461 ATCTGGCTTA AGTGTTTGAT CGTGAAAAAT AGAGCGTAGA GCCACCTCAA  
AAAATTGCGG

30

5521 TAGATCATTT TTATAAATGC TAATTAAAAC TGAGAATTTT TCTAATCTAT  
GATTCATTTT

5581 ATTTTACCAC TTCGACCCAT TAAACCGTCA TTAATGCCTT TAAAAAATAA  
ATATAACCTT

5641 TTATTACCAT TTGGAAGGAA AATAGGATAT AAAAAACCT TTCCAATTAA  
TTTAACCAGA

5701 CTAGAAATTT TCCAGTAGAT GGGTACATAA TTTTATTTA ATAAAAGAAA  
GATATTTTGA

40

5761 GTAGCATAAT AATGACGAAA TGGGCTTGGC AAACCGACAG  
AAAGAATATT TAAGATCTTA

5821 AATCGCCCAT CTCCAAGTCT ATGTGCAAGT AACGCATTTT TATTCCTAAT  
TACTTTAAAC

5881 CCAGCAGCTC TTAATCTCCA ACAATATTCA TGGTCTACCG CATCGATAAA  
AAGCTCATCT

5941 TTCATTCTC CAACAATCAA CCAACTATTT TTTGGTATTA GACTGCCAGA  
ACTTAATGTA

6001 CTATCTACCT CATAATAAAC TTCTGTAAGT GGTTTCCCTT TTTTACCCT  
TGCTTTATTT

10

6061 AATTCACCAG TTACTTTATC AAAATCTTGT GAACCAACTA AACCAACATT  
GACATTTTGT

6121 TTAAGCAATT TTTTGTAACA AGTAAGTAAC TGCTCTACCA TCTTAGGATC  
AGGAATACTA

6181 TCCTGATCCA TTTGCAATAT AAAATCAGCG CCATTTTCAA AAGCCCATT  
CATTCCTATA

6241 CTTTGGGCTT CTGCTATGCC TAAATTATCA TTGAAATTGA ATATTTTAC  
ATCGCCTGAA

20

6301 GAATTTTCAG CATATTTATA ACCATTTGTA GAGTTATTGC AAACGACAAC  
TTTAGTAACT

6361 TGTCTCAACA ATAATTCAAC CGCATTTTTT AAATCATTAT GTTCTGGGT  
GTAAGCAACC

6421 AAAACGGCAT ATACAGTGTC CATCTTCACC TTAAAACCTT CATTTAGCTT  
TCATCTTTTT

6481 TAGAACATTA CTTAATGTCA CTAATACAAT TATTACAGCA ACATGGTTAG  
AGTCTAAAAT

30

6541 ATAAGGATTA GTAATTGCAT AAGAAACATA TAGAAAATAT  
AGCACACACA ACTCACTGTA

6601 TTTTATGATT TTAATCGTGA GAAGGAGATT AATTAATAAA AACAAAGTAA  
ATAAAATAAC

6661 GCCAAGTTGA TTAAAAAAT AAAGTACTG CAATTCATAA TATATATATG  
CACTATAATC

6721 ACGGATAGGA GTTTGAATTT TGATGACATT ACCCAAACCA GAACCTATAA  
CAAAATTTGA

40

6781 TACAGACTCT GTAAGATCAT TAATTAATAC AGTAACTGA TCCCATCTAA  
CTCCTAAAGA

6841 AGAATCAGCT CCATTTGATT TCATGATTAT CAACTCAATT GAATATGTAA  
TAAAAAAAGG

6901 GAGAATCACA GTAAGAAAAA CCCCAAAAT AATTTTCCTT AATTTAGCGT  
ATCGTGAGTT

6961 AGATTTAGAA CATAGTATAA TATACATAAA AAACAAGCAT  
ATCGAAACAA AATATGCAAA

7021 ATTACCAGCC ACTATAGTAC CTATAGCCAG AATAACGGTT ATTGTATTTT  
TGAATCGATA

10

7081 ATAGAAATAA TCTTTTATGA CTATATGCAA CATAAAGGCA AATGGAATGA  
GAGCATTTCC

7141 TTTAATTTGA ACTCTATAGA AACCACCTCC ATATGTATAA ACATCACCAT  
AATCATTCTC

7201 CAAAAAATAA TGTCTTAGTG CTGAATAATC ACCAATACCA TATGTTTTTG  
TCATATAAAT

7261 ACTAATGATG GATATAATAA CCGCCTGTAA TACCATTAAA TATAAAAATA  
TTTTAACAAT

20

7321 CGAGATGGTT CCATAAGAGC AGAAATAAGC ACATAATATA  
AATAATATGA TAATATAAAA

7381 CCTAATTATT ATCGCTATAT CGTTACCCTT GATATAGGAA TAAATAAAAT  
TTATAAAAAG

7441 AGCTAATAGA AATATTAAAA TAACAGGATA GTGATATATT CCGTTTGCAA  
TTTTCTTTGT

7501 AAATGACATG ATACAAAGAC ATAAAAACCC CTCCATAATC CAACTATATT  
GAATAAATGG

30

7561 AAAGCTACGT GTAAGGAAAA ATATAAACCC AAAAAACAAA  
AGAACACTTA AACTTTTGTC

7621 TTTTGAGTTA TAAAAATCAG AAGTCATGTT TGCACTCTAA TTAGATGGGC  
TTGAGGAAGT

7681 AATCCCTAAA ATCAATTCGC TATTAATATT TCGTATCAAT TAATAATAAT  
ATCAAAAAAT

7741 CTAACGATGT TCTTACAGAC CATGCTATTG CGGCTCCAAC AATTCCCCAA  
TGATAAATAA

40

7801 AAATATATAA TATGCATAAA TATGGGATAA CTTCGAGCAA  
ATGAATAATA GCTGTAATTT

7861 TTGATCTTCC ACTAGCCTGA ACTGAAACAA ATGGGATTTG TGCAATGCAA  
TTAAAAAAGA

7921 AACCTATTGC AAGAATTTTT AATACTATAC CTGGCGTCCC ATGATATGTA  
GGTCCCATCC

7981 AAGCGGACAT TATAAAATCT GATAAAATAA TTATCAACAT TACAATTGGA  
AGTATACCAA

8041 TAACCATTAT AAAATATGAT AATATTTTAG TTTGCTTTAC CGATTGCAAT  
TCTGAACTTA

10

8101 ATCTTGAAA AATAGCTCTG GACAACGCAC TTGGTAATAT CGTTAAGCGT  
TGTATACCTT

8161 CAGACGGAGC AGTATAAAAA GAACTTTAT CAGCCCCCAC  
AATGTGTGAA AGAATAAAAC

8221 GATCCATATA TGTCATAATA GGGCTAATAA TATTGCTAAC TGTTATCCAG  
CTTCCAAAGC

8281 CGATTAATCT TTAACTGTT ACAATTTTTA CAGACAGCCC AGATGATATT  
ATTAGTTTTC

20

8341 GACTAAATAT AAAGGTCACT ATAAGTGATA AGACTCTTGC CATAACTAAA  
CCATATATAG

8401 CACTTAGTAA TCCTCCATGA AAAAAACAGA AAATCACTGG TAATCCAGCC  
ACAAAAGAGT

8461 TGTTAATTGA TTTATTAAA TTTACTTTTC TGAACTTTTC CATCCCCTCA  
AAAATCCCCA

8521 ACCAGACTTG GTTTAACAAG TATAAGGGTA TGGTAGCTGA AATAATATAT  
ATTGCTTTGA

30

8581 CAGATTCTAC AACATGATTC GCGTTAATGT TTAATAATTT AACAATTACA  
TTGCTACTCA

8641 AAAATAGTAC ACTACCGCCA ATCAAGCCCA ATATAGTTAG AATTACCGTT  
GAAGTTGAAA

8701 TGATCGCTCT TAATTCTTTA TGAACATTTT TATATATTGA TACTTCTCTT  
ATAACAGCTC

8761 TGGTCAATCC AGCATCAAAA ATACTTGCAT ATCCAATAA GGCAATAGCT  
AACGTAAAAA

40

8821 GGCCAAATTG CTCGGTCCCT AGAATTCTAG ACAGTATACC TAACGCAGGA  
ATTGCTATTA

8881 ATGATGGTAT AATATACCCA CTTATATTCC ATAAAGTATT CTTTACAATA  
CTCACAAAAA

8941 TAATTCCTTC ATGTTATGCA ATTCTTTAGC CCTTGCATCT TTAATCGATA  
AAATATAATT

9001 ATTATGTTCT ATCGTCGGCC ATTTTATGCT CAGAATAGGA TCATTCCATA  
CAATCCCTCT

9061 ATCACTATCA GGATGATAAT AGTTCGTCGT TTTATATAAA AATTCCGCAG  
TCTCGCTCAG

9121 CACCAAAAAA CCATGTGCAA ATCCCTCAGG GATCCACAAT TGCCGCTTAT  
TCTCAGCAGA

9181 TAAATTCACC CCAACCCATT TACCAAAGGT AGGCGACGAT TTACGAATAT  
CAACAGCTAC

9241 ATCAAAAACC TCACCAACAA CGCAACGTAC CAGTTTCCCT TGCGCATAAG  
GTTCTAACTG

9301 ATAATGCAGC CCGCGTAAAA CACCTTTACT AGACTTCGAA TGGTTATCCT  
GAACAAATTC

9361 AACCTTACGT CCTACAGCTT CTTCGAAAAC TTTCTGATTA AAGCTTTCCA  
TAAAGAAACC

9421 ACGCTCATCA CCAAAAACCTT TCGGCTCGAA AATTAACACA TCAGGAATTT  
CTGTTTTAAT

9481 TACGTTCAAT TTATTAATAA CCTTTAATCA TTTTCAGCAG ATACTGTCCA  
TAAGCATTTT

9541 TTTTCAGCGG CTCCGCTAAT GCTTTCACCT GTTCAGCATC AATAAACCCCT  
TTACGGTAAG

9601 CAATTTCTTC TGGGCAGGAA ACCTTTAGTC CCTGGCGCTC TTCAATGGTG  
GCAATGAAGT

9661 TGCTTGCTTC AATAAGACTC TGATGTGTCC CCGTATCCAG CCATGCATAA  
CCACGCCCCA

9721 TCATGGCAAC GGATAAACGC CCCTGTTCCA TATAAATACG GTTAATATCG  
GTAATTTCCA

9781 GTTCACCACG GGCAGAAGGC TTAAGGTTTT TCGCCATTTT GACAACGTCG  
TTATCATAGA

9841 AATAAAGCCC GGTTACCGCA TAATTACTTT TTGGTTGTAG CGGTTTTTCT  
TCCAGGCTTA

10

20

30

40

9901 TTGCCGTACC GTTTTTATCA AACTCAACGA CGCCGTAGCG TTCAGGATCA  
TTAACGTGAT

9961 AGGCAAATAC CGTTGCACCA CTTTCTTTGT TAACAGCGAC ATCCATTAAC  
TTCGGCAGAT

10021 CATGACCGTA GAAGATATTA TCACCAAGAA CCAAAGCACA  
ATCATCACCA CCGATAAACT

10081 CTTACCCGAT AATAAACGCC TGC GCAAGCC CATCTGGAGT CGGTTGCACT  
TTGTA CTGAA

10

10141 GATTTAGCCC CCACTGGCTA CCGTCACCTA GCAGTTGTTG  
AAAACGAGGA GTATCCTGTG

10201 GCGTACTAAT AATCAGAATA TCGCGAATAC CCGCCAACAT  
CAGTGTAGAG AGCGGGTAAT

10261 AGATCATCGG CTTATCATAA ATAGGTAATA GCTGTTTACT GACAGCCATA  
GTCACAGGAT

10321 AAAGACGTGT ACCAGAACCA CCCGCTAAAA TAATACCTTT ACGCGTTTTTC  
ATTT CATCAT

20

10381 TCCTTTTAAT TCATCTTGCT CCACCATCAC GAACAAGATG CAAAACTAT  
TAAATTGCTG

10441 TAGTCGTAAT TAATTCGTTG AGCATTCGTT TCACACCAAC CTGCCAGTCA  
GGCAAGACAA

10501 GCGCAAAGTT CTGCTGAAAT TTTTCTGTAT TAAGGCGAGA GTTATGTGGA  
CGACGAGCTG

10561 GTGTAGGATA GGCTGTTGTT GGTACTGCGT TGAGCTTGTT GAGTGCAAGG  
GGAATACCTG

30

10621 CTTTGC GCGC CTCTTCAAAA ACCAGCGCAG CATAATCGTG CCAGGTTGTG  
GTACCACTGG

10681 CTACCAGATG GTACAAACCT GCGACTTCCG GTTTATTCAG TGCCACACGA  
ATAGCATGTG

10741 CCGTACAATC AGCCAGCAGC TCAGCACCTG TTGGCGCACC  
AAATTGATCA TTTATCACAG

40

10801 CCAGTTCTTC GCGCTCTTTT GCCAGACGCA ACATCGTTTT GGCGAAGTTA  
TTTCCTTTAG

10861 CTGCGTATAC CCAGCTGGTA CGGAAAATAA GATGCTTCGC GCAATGTTCC  
TGTAACGCTT

10921 TTTCTCCGGC TAACTTGGTT TCACCGTAAA CATTTAGCGG TCGGGTTGCA  
TCCGTCTCCA

10981 GCCATGGCGT GTCGCCATTT CCAGGGAATA CGTAGTCAGT  
TGAGTAATGA ATTACCCAAG

11041 CCCCAACTTC ATTAGCCTCT TTTGCAATTG ATTCAACACT AGTCGCATTG  
AGTAATTGTG

11101 CAAATTCGGG TTCTGACTCA GCCTTATCTA CTGCGGTGTG AGCCGCAGCA  
TTAACAATAA

10

11161 CATCAGGTCG AATTCTTTTG ACTGTTTCAG CTACACCTTC AGGATTACTA  
AAATCACCAC

11221 AATAATCAGT GGAGTGAACA TCAAGAGCAA TCAAATTACC  
CAAAGGTGCC AGAGCACGCT

11281 GTAGTTCCCA ACCTACCTGC CCTGTTTTGC CGAAAAGGAG GATATTCATT  
ACTGGCGGCC

11341 CTCATAGTTC TGTTC AATCC ACGATTGATA AGCACCCTT TTCACATTAT  
CAACCCATTT

20

11401 TGTATTGGAC AGGTACCATT CCAATGTCTT CCGAATCCCG CTCTCAAACG  
TTTCCTGCGG

11461 TTTCCAGCCC AATTCGCGGC TAATCTTCTC TGCATCAATC GCATAACGGC  
GATCGTGTCC

11521 CGGGCGATCG GCAACATAAG TAATTTGCTC GCGGTAAGAT TTCTCTTTTCG  
GTACAATCTC

11581 ATCCAGCAAA TCACAAATAG TGAGCACTAC ATCGATGTTT TTCTTTTCGT  
TGTGTCCACC

30

11641 AATGTTATAA GTTTCACCCG CTTTACCTTC GGTACGACG GTATATAACG  
CACGCGCATG

11701 ATCTTCAACA TACAGCCAGT CACGAATTG ATCCCCTTTG CCATAAATAG  
GTAATGCCTT

11761 ACCTTCCAGA GCATTCAGAA TAACCAATGG AATCAATTTT TCCGGGAAAT  
GATAAGGACC

11821 ATAATTATTA GAGCAATTAG TCACAATGGT TGGTAAACCA TAGGTACGTT  
TCCACGCGCG

40

11881 GACTAAATGA TCGCTGGATG CTTTTGAAGC GGAATAAGGG  
CTGCTTGGCG CGTAAGCTGT

11941 TGTCTCTGTA AATAAGGGTA ATTCTTCTGT ATTATTTACC TCGTCAGGAT  
GAGGCAAATC

12001 ACCATAGACT TCGTCAGTAG AAATATGATG AAAACGGAAT CTAGTTTTCT  
TGTCGCTATC

12061 AAGAGCAGAC CAATAATTGC GAGCGGCTTC CAAAAGGACA  
TATGTACCAA CAATATTGGT

12121 TTCAATAAAT GCCGCAGGAC CTGTAATTGA ACGGTCAACA TGGCTTTCAG  
CAGCCAGGTG 10

12181 CATCACTGCA TCTGGCTGAT GCTGAGCAAA AATCCGTGCC ATTGCAGCTG  
CATCGCAAAT

12241 ATCCGCATGT TCAAAAACAT AGCGTTCAGA ATCAGAAACA  
TCAGCAAGTG ATTCCAGGT

12301 TCCGGCGTAC GTTAATTTAT CGACATTAAC AACACTATCC TCGTATTAT  
TTATAATGTG

12361 ACGAACTACA GCAAAACCAA TAAATCCTGC GCCACCAGTA 20  
ACAAGTATTT TCACCTAATT

12421 TATTCCATAT TGCTTCAGAG CATGCTGTGA AATAAGCGGC TCTCAGTTTG  
ATTAATAGAA

12481 GTATTAATGC ACGCTACCGC CCCTGGCTTT ACAGCTACCA GAGCACTGCA  
TGCATGCCTA

12541 CGATGTGACG AGCGTTACCC ACTCGCGCTA AACCCGAAAA  
ATTCAAAAGC TAATTGTCTT

12601 ACCAATCCGC TCTGGAAACA AGGAAAATCC TGGAAAACCT 30  
TGACTAAAAT CCTATTGCTA

12661 ACTCGTTGTT ATTCTGATTG TTTATATAAA ACAACGGCAG GAATATTCGC  
AACAAATTAC

12721 TTTCACCACG AATCTTCACT GCCGTTATAA TTTTCTTATC AACCGTTACA  
TCCGGTCAGA

12781 TTTTCATTAT TCGCTTAACA GCTTCTCAAT ACCTTTACGG AACTTCGCCC  
CTTCTTTCAG

12841 GTTGCGCAGC CCATACTTCA CAAACGCCTG CATATAGCCC ATTTTTTTAC 40  
CGCAGTCGTA

12901 GCTGTCGCCG GTCATCAGCA TTGCATCAAC GGACTGTTTT TTCGCCAGCT  
CGGCAATGGC

12961 ATCAGTCAGC TGAATACGTC CCCATGCACC AGGCTGAGTA CGTTCAAGTT  
CCGGCCAAAT

13021 ATCGGCAGAA AGCACATAGC GACCAACGGC CATGATGTCT  
GAGTCCAGCG TCTGCGGCTG

13081 ATCCGGTTTT TCGATAAATT CAACAATGCG GCTGACTTTA CCTTCGCGAT  
CCAGCGGTTT

13141 TTTGGTCTGG ATGACGGAGT ATTCAGAGAG GTCACCCGGC ATACGTTTTG  
CCAGCACCTG

10

13201 GCTACGGCCC GTTTCATTGA AGCGCGCAAT CATGGCAGCA  
AGGTTGTAGC GTAGCGGGTC

13261 GGCGCTGGCG TCGTCGATCA CAACGTCTGG CAGCACCACG  
ACAAATGGAT TGTCACCAAT

13321 GGCGGGTCGT GCACACAAAA TGGAGTGACC TAAACCTAAA  
GGTTCGCCCT GACGCACGTT

13381 CATAATAGTC ACGCCCGGCG GGCAGATAGA TTGCACTTCC  
GCCAGTAGTT GACGCTTCAC

20

13441 GCGCTGCTCA AGGAGAGATT CTAATTCATA AGAGGTGTCG AAGTGGTTTT  
CGACCGCGTT

13501 CTTGGACGCA TGAGTTACCA GGAGGATTTT TTTGATCCCT GCAGCCACAA  
TCTCGTCAAC

13561 AATGTACTGA ATCATTGGCT TGTCGACGAT CGGTAGCATC TCTTTGGGTA  
TCGCCTTAGT

13621 GGCAGGCAAC ATATGCATCC CAAGACCCGC TACCGGTATA ACTGCTTTTA  
AATTCGTCAT

30

13681 TATTTTCCTA CCTCTAAGGG GCTGATAGTG CGTAAATTAT TGTCATAGGT  
TAGCCAAACG

13741 GTATGGCTAT ATACCAAGCA TAACTTTGAT TAAACCTTAC GATAACACTA  
CACACCATCA

13801 GCATCTGGGT TACTCGGATT ACTCGGAAAT CCACATACTG ATAATTTAAT  
CAGTACCTCT

40

13861 TTCCGAATAA TCGTAGTCCA ACCTGGTCCT TTTTCTCTG ACTCGTCTGC  
ATTACTCAGA

13921 AACAAACGTT ATGTCGTCTT TTTTGGCATG GACGAATTCA TACTGCAGAG  
TTCGATCCAG

13981 ACCTTGCGAC AGCGTATACG GTGCAACAAA ACCTGAAGAA  
TGCACTTTTCG TTGCGTCAAA

14041 CTGTGTTGTT GCGCAGAATT TTTTCACGCG CACAGAGCTG ACAGCGTATT  
TTTTGCCCCGT

14101 AATTTTGCTC AGGATATCAA AGCAATATCC ACCCAGCATT CCTAGTGGGT  
AAGGCAAGTG

14161 CATAGAAGGG ATCTTTTTGT TCAGGCTTTG TTCAACTTCA GCAACCAACT  
GGTTCATGTT

10

14221 CAGGTCTGGC TTATCAACAT AGTTATAAAC CTCATAACCT GCGGCAACAT  
TCTTCAGTTT

14281 GTACTTGATA AACTCAACAA TGTTTCCAAC ATAAGCCATG GACTTATAGT  
TAGTCCCTGC

14341 GCCCACCATC ATAAACTTGC CGCCAGCGAT CTGTTTCAGC AAGTTATAGA  
CGTTACCGCG

14401 GTTGCGTTCA CCGAAGATAA CGGTAGGACG GATGATGGTT  
AATGAACGTT CTGTTGGTGC

20

14461 TTTGTTATAC CATTACGCA GCACTTCCTC TGCCTGCCAC TTACTTTTGC  
CGTAGTGGTT

14521 GAAAGGGTTCG TGTGGATGGT TTTCGTCAGG GTTGTGTTTG TTCAAACCAT  
AAACAGCAAC

14581 GGAAGTGGTA AAGATGATAT TTTTAACGCC ATTTTTTTCC ATGGCCGCCA  
GCACATTGCG

14641 GGTACCCTGA ACGTTGACAT CATAATAGAG AGAAGTAGGG  
CTGACGTCAT CGCGGTGTTC

30

14701 CGCTGCCAGT AGTACAACAG TGTCAAAACC GGCTAACGCC  
TGGTCGAGTG CCTGTTGATC

14761 ACGAACATCA CCAATCTGTG TGATTTCTGG ATAAAAGTGG CTCTGCCGTT  
TGTCCAGGTT

14821 CTTGATATTA AAGTCAGCAA TTGCCGTTTC AAGTAGTCGG GTTCCTACGA  
ATCCGGAAGC

14881 TCCTATGAGC AAAACGTTAT TGTTCATAAA TCACTTTAGT CTGGTTGTTA  
CGTAAGAAAC

40

14941 ACAAGATAAA GATGAGTACC TTCCCTGAGT AGTCAATGCT  
GCCCAGCCCC AGCTTTAACA

15001 GTTAGTGTGA GGATTATAAT CTTT TAGAAC ATTATATCCA GTAAGTTTAT  
GAATGGTCGC

15061 AAATCTACTC TCTCCGTTCC GGCAATCTAA AGTTAATGCT AGCGACGTCG  
TGGGATCCTC

15121 TAGAGTCGAC CTGCAGGCAT GCAAGCTTGA GTATTCTATA GTCTCACCTA  
AATAGCTTGG

15181 CGTAATCATG GTCATAGCTG TTTCTGTGT GAAATTGTTA TCCGCTCACA  
ATTCCACACA

10

15241 ACATACGAGC CGGAAGCATA AAGTGTAAG CCTGGGGTGC  
CTAATGAGTG AGCTAACTCA

15301 CATTAATTGC GTTGCCTCA CTGCCCCTT TCCAGTCGGG AAACCTGTCTG  
TGCCAGCTGC

15361 ATTAATGAAT CGGCCAACGC GAACCCCTTG CGGCCGCCCG  
GGCCGTCGAC CAATTCTCAT

15421 GTTTGACAGC TTATCATCGA ATTTCTGCCA TTCATCCGCT TATTATCACT  
TATTCAGGCG

20

15481 TAGCAACCAG GCGTTTAAGG GCACCAATAA CTGCCTTAAA  
AAAATTACGC CCCGCCCTGC

15541 CACTCATCGC AGTACTGTTG TAATTCATTA AGCATTCTGC CGACATGGAA  
GCCATCACAA

15601 ACGGCATGAT GAACCTGAAT CGCCAGCGGC ATCAGCACCT  
TGTCGCCTTG CGTATAATAT

15661 TTGCCCATGG TGAAAACGGG GCGAAGAAG TTGTCCATAT  
TGGCCACGTT TAAATCAAAA

30

15721 CTGGTGAAAC TCACCCAGGG ATTGGCTGAG ACGAAAAACA  
TATTCTCAAT AAACCCTTTA

15781 GGGAAATAGG CCAGGTTTTT ACCGTAACAC GCCACATCTT  
GCGAATATAT GTGTAGAAAC

15841 TGCCGGAAAT CGTCGTGGTA TTCCTCCAG AGCGATGAAA ACGTTTCAGT  
TTGCTCATGG

40

15901 AAAACGGTGT AACAAGGGTG AACACTATCC CATATCACCA  
GCTCACCGTC TTTCATTGCC

15961 ATACGAAATT CCGGATGAGC ATTCATCAGG CGGGCAAGAA  
TGTGAATAAA GGCCGGATAA

16021 AACTTGTGCT TATTTTCTT TACGGTCTTT AAAAAGGCCG TAATATCCAG  
CTGAACGGTC

16081 TGGTTATAGG TACATTGAGC AACTGACTGA AATGCCTCAA AATGTTCTTT  
ACGATGCCAT

16141 TGGGATATAT CAACGGTGGT ATATCCAGTG ATTTTTTCT CCATTTTAGC  
TTCCTTAGCT

16201 CCTGAAAATC TCGATAACTC AAAAAATACG CCCGGTAGTG ATCTTATTTC  
ATTATGGTGA

10

16261 AAGTTGGAAC CTCTTACGTG CCGATCAACG TCTCATTTTC GCCAAAAGTT  
GGCCCAGGGC

16321 TTCCCGGTAT CAACAGGGAC ACCAGGATTT ATTTATTCTG CGAAGTGATC  
TTCCGTCACA

16381 GGTATTTATT CGCGATAAGC TCATGGAGCG GCGTAACCGT  
CGCACAGGAA GGACAGAGAA

16441 AGCGCGGATC TGGGAAGTGA CGGACAGAAC GGTGAGGACC  
TGGATTGGGG AGGCGGTTGC

20

16501 CGCCGCTGCT GCTGACGGTG TGACGTTCTC TGTTCCGGTC ACACCACATA  
CGTTCCGCCA

16561 TTCCTATGCG ATGCACATGC TGTATGCCGG TATACCGCTG AAAGTTCTGC  
AAAGCCTGAT

16621 GGGACATAAG TCCATCAGTT CAACGGAAGT CTACACGAAG  
GTTTTTGCGC TGGATGTGGC

16681 TGCCCGGCAC CGGGTGCACT TTGCGATGCC GGAGTCTGAT  
GCGGTTGCGA TGCTGAAACA

30

16741 ATTATCCTGA GAATAAATGC CTTGGCCTTT ATATGGAAAT GTGGAACTGA  
GTGGATATGC

16801 TGTTTTTGTC TGTTAAACAG AGAAGCTGGC TGTTATCCAC TGAGAAGCGA  
ACGAAACAGT

16861 CGGGAAAATC TCCCATTATC GTAGAGATCC GCATTATTAA TCTCAGGAGC  
CTGTGTAGCG

40

16921 TTTATAGGAA GTAGTGTTCT GTCATGATGC CTGCAAGCGG  
TAACGAAAAC GATTTGAATA

16981 TGCCTTCAGG AACAATAGAA ATCTTCGTGC GGTGTTACGT  
TGAAGTGGAG CGGATTATGT

17041 CAGCAATGGA CAGAACAACC TAATGAACAC AGAACCATGA  
TGTGGTCTGT CCTTTTACAG

17101 CCAGTAGTGC TCGCCGCAGT CGAGCGACAG GGCGAAGCCC  
TCGGCTGGTT GCCCTCGCCG

17161 CTGGGCTGGC GGCCGTCTAT GGCCCTGCAA ACGCGCCAGA  
AACGCCGTCG AAGCCGTGTG

17221 CGAGACACCG CGGCCGGCCG CCGGCGTTGT GGATACCTCG  
CGGAAAACCTT GGCCCTCACT

10

17281 GACAGATGAG GGGCGGACGT TGACACTTGA GGGGCCGACT  
CACCCGGCGC GGC GTTGACA

17341 GATGAGGGGC AGGCTCGATT TCGGCCGGCG ACGTGGAGCT  
GGCCAGCCTC GCAAATCGGC

17401 GAAACGCCT GATTTTACGC GAGTTTCCCA CAGATGATGT  
GGACAAGCCT GGGGATAAGT

17461 GCCCTGCGGT ATTGACACTT GAGGGGCGCG ACTACTGACA  
GATGAGGGGC GCGATCCTTG

20

17521 ACACTTGAGG GGCAGAGTGC TGACAGATGA GGGGCGCACC  
TATTGACATT TGAGGGGCTG

17581 TCCACAGGCA GAAAATCCAG CATTTGCAAG GGTTTCCGCC CGTTTTTCGG  
CCACCGCTAA

17641 CCTGTCTTTT AACCTGCTTT TAAACCAATA TTTATAAACC TTGTTTTTAA  
CCAGGGCTGC

17701 GCCCTGTGCG CGTGACCGCG CACGCCGAAG GGGGGTGCCC  
CCCCTTCTCG AACCTCCCCG

30

17761 GTCGAGTGAG CGAGGAAGCA CCAGGGAACA GCACTTATAT  
ATTCTGCTTA CACACGATGC

17821 CTGAAAAAAC TTCCCTTGGG GTTATCCACT TATCCACGGG GATATTTTAA  
TAATTATTTT

17881 TTTTATAGTT TTTAGATCTT CTTTTTTAGA GCGCCTTGTA GGCCTTTATC  
CATGCTGGTT

40

17941 CTAGAGAAGG TGTTGTGACA AATTGCCCTT TCAGTGTGAC AAATCACCCCT  
CAAATGACAG

18001 TCCTGTCTGT GACAAATTGC CCTTAACCCT GTGACAAATT GCCCTCAGAA  
GAAGCTGTTT

18061 TTTCACAAAG TTATCCCTGC TTATTGACTC TTTTATTTT AGTGTGACAA  
TCTAAAAACT

18121 TGTCACACTT CACATGGATC TGTCATGGCG GAAACAGCGG TTATCAATCA  
CAAGAAACGT

18181 AAAAATAGCC CGCGAATCGT CCAGTCAAAC GACCTCACTG  
AGGCGGCATA TAGTCTCTCC

18241 CGGGATCAAA AACGTATGCT GTATCTGTTC GTTGACCAGA  
TCAGAAAATC TGATGGCACC

10

18301 CTACAGGAAC ATGACGGTAT CTGCGAGATC CATGTTGCTA  
AATATGCTGA AATATTCGGA

18361 TTGACCTCTG CGGAAGCCAG TAAGGATATA CGGCAGGCAT  
TGAAGAGTTT CGCGGGGAAG

18421 GAAGTGGTTT TTTATCGCCC TGAAGAGGAT GCCGGCGATG  
AAAAAGGCTA TGAATCTTTT

18481 CCTTGGTTTA TCAAACGTGC GCACAGTCCA TCCAGAGGGC TTTACAGTGT  
ACATATCAAC

20

18541 CCATATCTCA TTCCCTTCTT TATCGGGTTA CAGAACCGGT TTACGCAGTT  
TCGGCTTAGT

18601 GAAACAAAAG AAATCACCAA TCCGTATGCC ATGCGTTTAT  
ACGAATCCCT GTGTCAGTAT

18661 CGTAAGCCGG ATGGCTCAGG CATCGTCTCT CTGAAAATCG ACTGGATCAT  
AGAGCGTTAC

18721 CAGCTGCCTC AAAGTTACCA GCGTATGCCT GACTTCCGCC GCCGCTTCCT  
GCAGGTCTGT

30

18781 GTTAATGAGA TCAACAGCAG AACTCCAATG CGCCTCTCAT  
ACATTGAGAA AAAGAAAGGC

18841 CGCCAGACGA CTCATATCGT ATTTTCCTTC CGCGATATCA CTTCCATGAC  
GACAGGATAG

18901 TCTGAGGGTT ATCTGTCACA GATTTGAGGG TGGTTCGTCA CATTTGTTCT  
GACCTACTGA

40

18961 GGGTAATTTG TCACAGTTTT GCTGTTTCCT TCAGCCTGCA TGGATTTTCT  
CATACTTTTT

19021 GAACTGTAAT TTTTAAGGAA GCCAAATTTG AGGGCAGTTT GTCACAGTTG  
ATTTCTTCT

19081 CTTTCCCTTC GTCATGTGAC CTGATATCGG GGGTTAGTTC GTCATCATTG  
ATGAGGGTTG

19141 ATTATCACAG TTTATTACTC TGAATTGGCT ATCCGCGTGT GTACCTCTAC  
CTGGAGTTTT

19201 TCCCACGGTG GATATTTCTT CTTGCGCTGA GCGTAAGAGC TATCTGACAG  
AACAGTTCTT

19261 CTTTGCTTCC TCGCCAGTTC GCTCGCTATG CTCGGTTACA CGGCTGCGGC  
GAGCGCTAGT

10

19321 GATAATAAGT GACTGAGGTA TGTGCTCTTC TTATCTCCTT TTGTAGTGTT  
GCTCTTATTT

19381 TAAACAACCTT TGCGGTTTTT TGATGACTTT GCGATTTTGT TGTGCTTTG  
CAGTAAATTG

19441 CAAGATTTAA TAAAAAACG CAAAGCAATG ATTAAGGAT  
GTTTCAAGATG AAATCATGG

19501 AAACACTTAA CCAGTGCATA AACGCTGGTC ATGAAATGAC  
GAAGGCTATC GCCATTGCAC

20

19561 AGTTTAATGA TGACAGCCCG GAAGCGAGGA AAATAACCCG  
GCGCTGGAGA ATAGGTGAAG

19621 CAGCGGATTT AGTTGGGGTT TCTTCTCAGG CTATCAGAGA TGCCGAGAAA  
GCAGGGCGAC

19681 TACCGCACCC GGATATGGAA ATTCGAGGAC GGGTTGAGCA  
ACGTGTTGGT TATACAATTG

30

19741 AACAAATTA TCATATGCGT GATGTGTTTG GTACGCGATT GCGACGTGCT  
GAAGACGTAT

19801 TTCCACCGGT GATCGGGGTT GCTGCCATA AAGGTGGCGT TTACAAAACC  
TCAGTTTCTG

19861 TTCATCTTGC TCAGGATCTG GCTCTGAAGG GGCTACGTGT TTTGCTCGTG  
GAAGGTAACG

19921 ACCCCCAGGG AACAGCCTCA ATGTATCACG GATGGGTACC  
AGATCTTCAT ATTCATGCAG

40

19981 AAGACACTCT CCTGCCTTTC TATCTTGGGG AAAAGGACGA TGTCATTAT  
GCAATAAAGC

20041 CCACTTGCTG GCCGGGGCTT GACATTATTC CTTCTGTCT GGCTCTGCAC  
CGTATTGAAA

20101 CTGAGTTAAT GGGCAAATTT GATGAAGGTA AACTGCCCCAC  
CGATCCACAC CTGATGCTCC

20161 GACTGGCCAT TGAAACTGTT GCTCATGACT ATGATGTCAT AGTTATTGAC  
AGCGCGCCTA

20221 ACCTGGGTAT CGGCACGATT AATGTCGTAT GTGCTGCTGA TGTGCTGATT  
GTTCCCACGC

20281 CTGCTGAGTT GTTTGACTAC ACCTCCGCAC TGCAGTTTTT CGATATGCTT  
CGTGATCTGC

10

20341 TCAAGAACGT TGATCTTAAA GGGTTCGAGC CTGATGTACG TATTTTGCTT  
ACCAAATACA

20401 GCAATAGTAA TGGCTCTCAG TCCCCGTGGA TGGAGGAGCA  
AATTCGGGAT GCCTGGGGAA

20461 GCATGGTTCT AAAAAATGTT GTACGTGAAA CGGATGAAGT  
TGGTAAAGGT CAGATCCGGA

20521 TGAGAACTGT TTTTGAACAG GCCATTGATC AACGCTCTTC AACTGGTGCC  
TGGAGAAATG

20

20581 CTCTTTCTAT TTGGGAACCT GTCTGCAATG AAATTTTCGA TCGTCTGATT  
AAACCACGCT

20641 GGGAGATTAG ATAATGAAGC GTGCGCCTGT TATTCCAAAA  
CATACGCTCA ATACTCAACC

20701 GGTTGAAGAT ACTTCGTTAT CGACACCAGC TGCCCCGATG GTGGATTCGT  
TAATTGCGCG

20761 CGTAGGAGTA ATGGCTCGCG GTAATGCCAT TACTTTGCCT GTATGTGGTC  
GGGATGTGAA

30

20821 GTTTACTCTT GAAGTGCTCC GGGGTGATAG TGTTGAGAAG ACCTCTCGGG  
TATGGTCAGG

20881 TAATGAACGT GACCAGGAGC TGCTTACTGA GGACGCACTG  
GATGATCTCA TCCCTTCTTT

20941 TCTACTGACT GGTCAACAGA CACCGGCGTT CGGTCGAAGA  
GTATCTGGTG TCATAGAAAT

40

21001 TGCCGATGGG AGTCGCCGTC GTAAAGCTGC TGCACTTACC  
GAAAGTGATT ATCGTGTCT

21061 GGTTGGCGAG CTGGATGATG AGCAGATGGC TGCATTATCC  
AGATTGGGTA ACGATTATCG

21121 CCCAACAAGT GCTTATGAAC GTGGTCAGCG TTATGCAAGC  
CGATTGCAGA ATGAATTTGC

21181 TGGAAATATT TCTGCGCTGG CTGATGCGGA AAATATTTC ACGTAAGATTA  
TTACCCGCTG

21241 TATCAACACC GCCAAATTGC CTAAATCAGT TGTGCTCTT TTTTCTCACC  
CCGGTGAAC

21301 ATCTGCCCCG TCAGGTGATG CACTTCAAAA AGCCTTTACA  
GATAAAGAGG AATTACTTAA

10

21361 GCAGCAGGCA TCTAACCTTC ATGAGCAGAA AAAAGCTGGG  
GTGATATTTG AAGCTGAAGA

21421 AGTTATCACT CTTTAACTT CTGTGCTTAA AACGTCATCT GCATCAAGAA  
CTAGTTTAA

21481 CTCACGACAT CAGTTTGCTC CTGGAGCGAC AGTATTGTAT AAGGGCGATA  
AAATGGTGCT

21541 TAACCTGGAC AGGTCTCGTG TTCCAACCTGA GTGTATAGAG  
AAAATTGAGG CCATTCTTAA

20

21601 GGAACCTGAA AAGCCAGCAC CCTGATGCGA CCACGTTTTA GTCTACGTTT  
ATCTGTCTTT

21661 ACTTAATGTC CTTTGTTACA GGCCAGAAAG CATAACTGGC CTGAATATTC  
TCTCTGGGCC

21721 CACTGTTCCA CTTGTATCGT CGGTCTGATA ATCAGACTGG GACCACGGTC  
CCACTCGTAT

21781 CGTCGGTCTG ATTATTAGTC TGGGACCACG GTCCCACTCG TATCGTCGGT  
CTGATTATTA

30

21841 GTCTGGGACC ACGGTCCAC TCGTATCGTC GGTCTGATAA TCAGACTGGG  
ACCACGGTCC

21901 CACTCGTATC GTCGGTCTGA TTATTAGTCT GGGACCATGG TCCCACTCGT  
ATCGTCGGTC

21961 TGATTATTAG TCTGGGACCA CGGTCCCACT CGTATCGTCG GTCTGATTAT  
TAGTCTGGAA

40

22021 CCACGGTCCC ACTCGTATCG TCGGTCTGAT TATTAGTCTG GGACCACGGT  
CCCACTCGTA

22081 TCGTCGGTCT GATTATTAGT CTGGGACCAC GATCCCACTC GTGTTGTCGG  
TCTGATTATC

22141 GGTCTGGGAC CACGGTCCCA CTTGTATTGT CGATCAGACT ATCAGCGTGA  
GACTACGATT

22201 CCATCAATGC CTGTCAAGGG CAAGTATTGA CATGTCGTCG TAACCTGTAG  
AACGGAGTAA

22261 CCTCGGTGTG CGGTTGTATG CCTGCTGTGG ATTGCTGCTG TGTCCTGCTT  
ATCCACAACA

22321 TTTTGCGCAC GGTTATGTGG ACAAATACC TGGTTACCCA GGCCGTGCCG  
GCACGTTAAC

10

22381 CGGGCTGCAT CCGATGCAAG TGTGTCGCTG TCGACGAGCT  
CGCGAGCTCG GACATGAGGT

22441 TGCCCCGTAT TCAGTGTCGC TGATTTGTAT TGTCTGAAGT TGTTTTTACG  
TTAAGTTGAT

22501 GCAGATCAAT TAATACGATA CCTGCGTCAT AATTGATTAT TTGACGTGGT  
TTGATGGCCT

22561 CCACGCACGT TGTGATATGT AGATGATAAT CATTATCACT TTACGGGTCC  
TTTCCGGTGA

20

22621 TCCGACAGGT TACGGGGCGG CGACCTCGCG GGTTTTCGCT ATTTATGAAA  
ATTTTCCGGT

22681 TTAAGGCGTT TCCGTTCTTC TTCGTCATAA CTTAATGTTT TTATTAAAA  
TACCCTCTGA

22741 AAAGAAAGGA AACGACAGGT GCTGAAAGCG AGCTTTTTTG  
CCTCTGTCGT TTCCTTTCTC

22801 TGTTTTTGTC CGTGGAATGA ACAATGGAAG TCCGAGCTCA TCGCTAATAA  
CTTCGTATAG

30

22861 CATAATTAT ACGAAGTTAT ATTCGAT

【図 1】

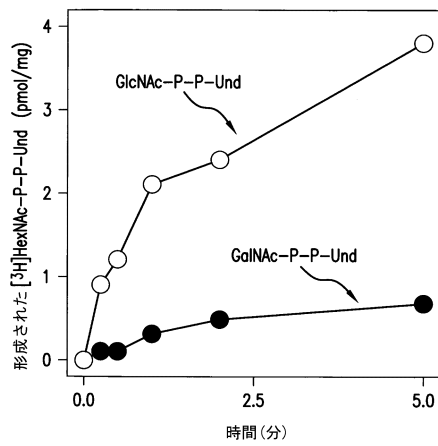


図 1

【図 2】

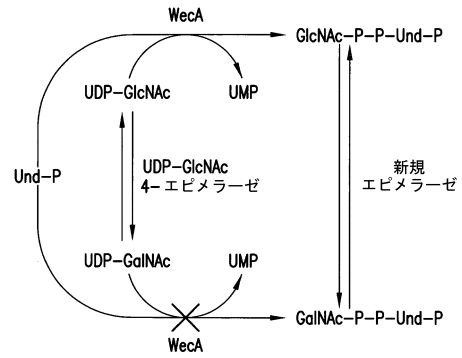


図 2

【図 3】

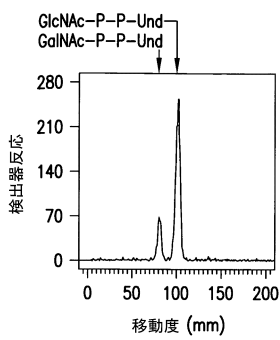


図 3A

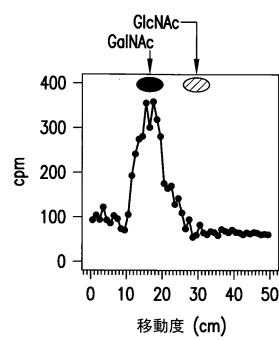


図 3C

【図 4】

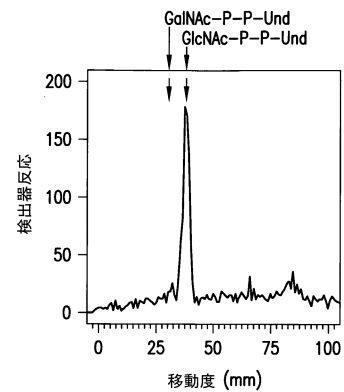


図 4A

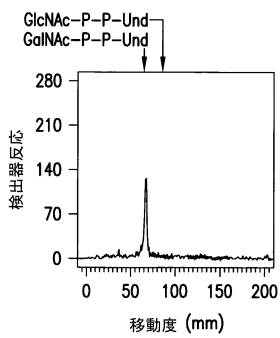


図 3B

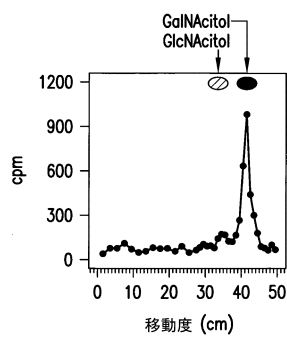


図 3D

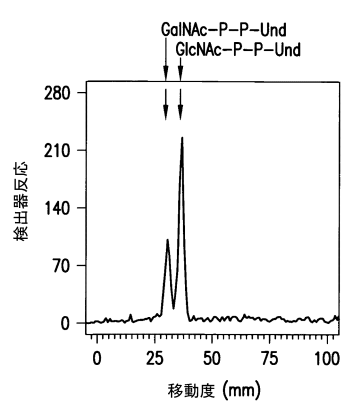


図 4B

【図 5】

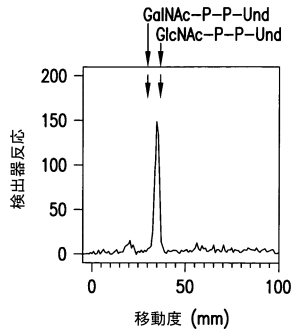


図 5A

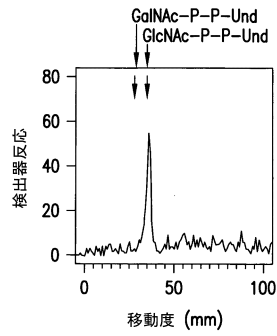


図 5C

【図 6】

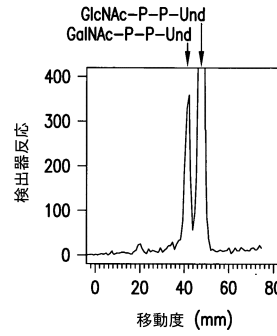


図 6A

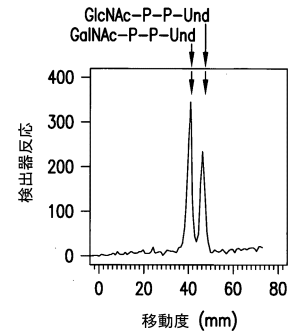


図 6B

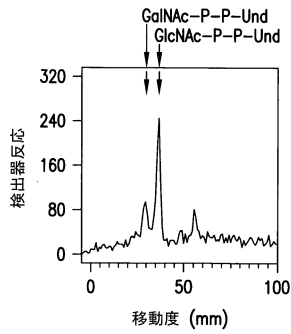


図 5B

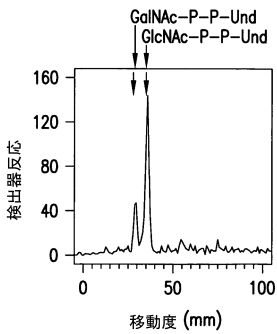


図 5D

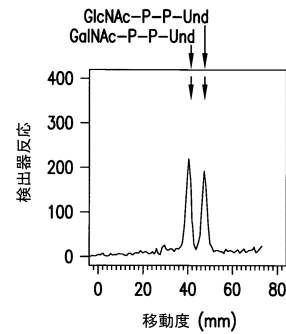


図 6C

【図 7 A - C】

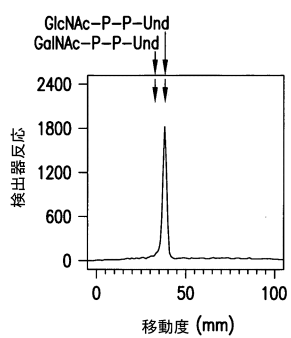


図 7A

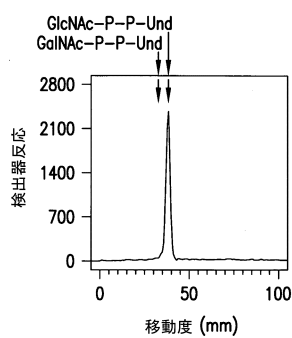


図 7B

【図 7 D - F】

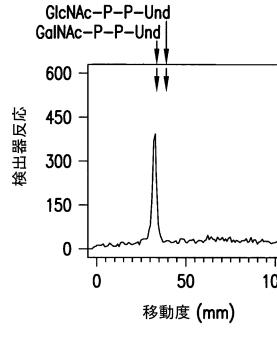


図 7D

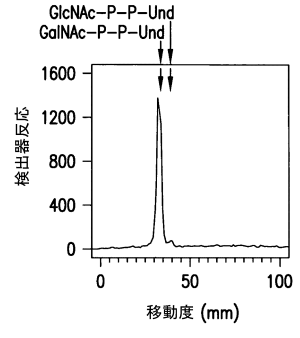


図 7E

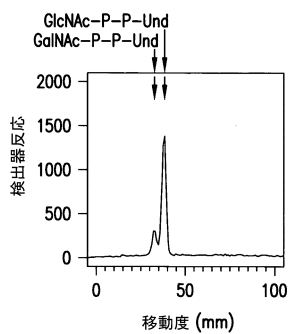


図 7C

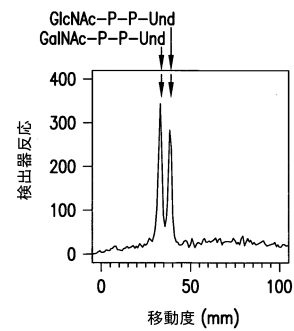
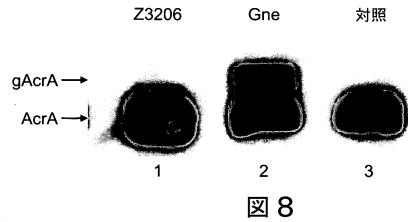
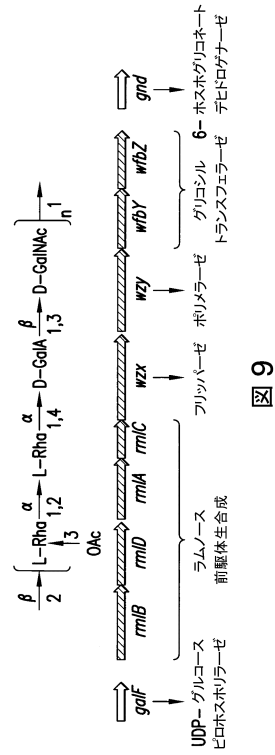


図 7F

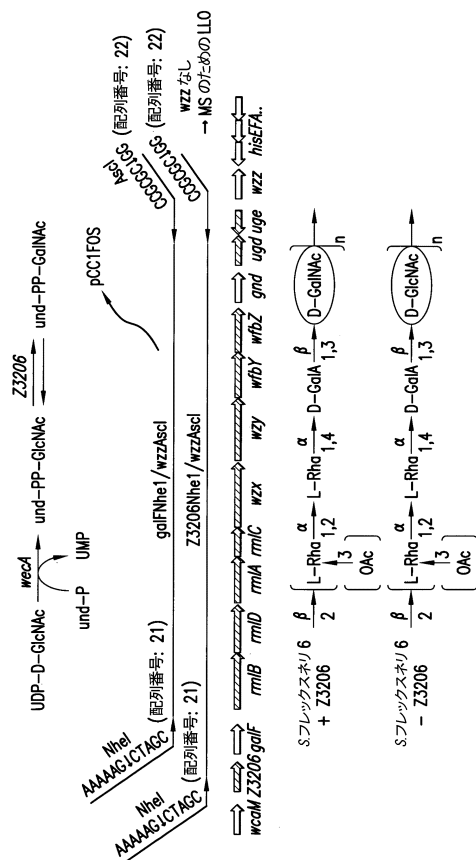
【図 8】



【図 9】

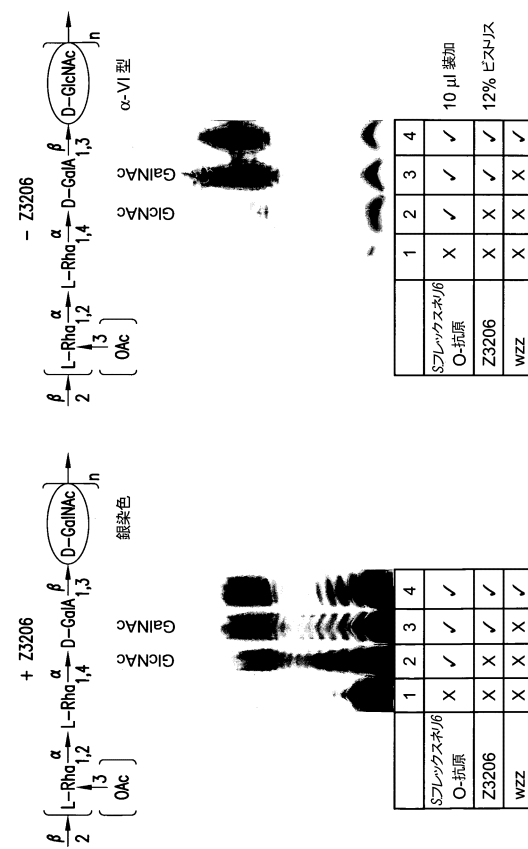


【図 10】



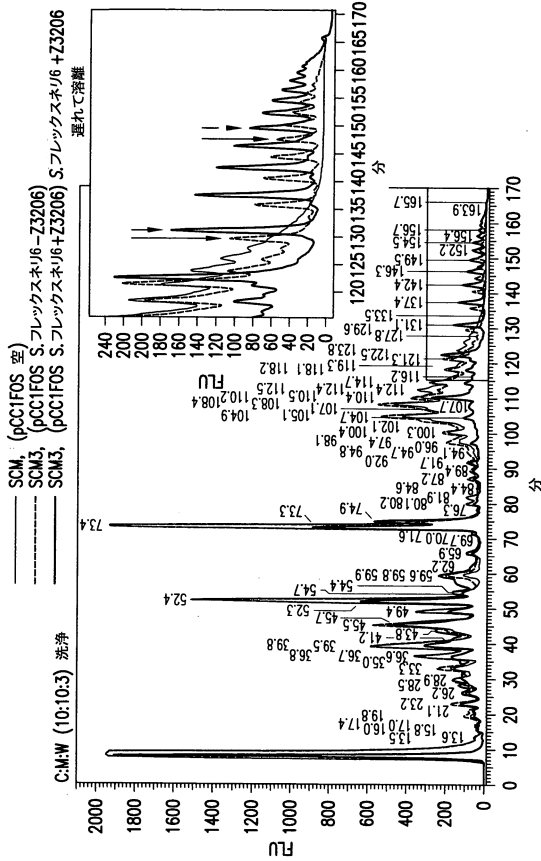
10

【图 1 1】



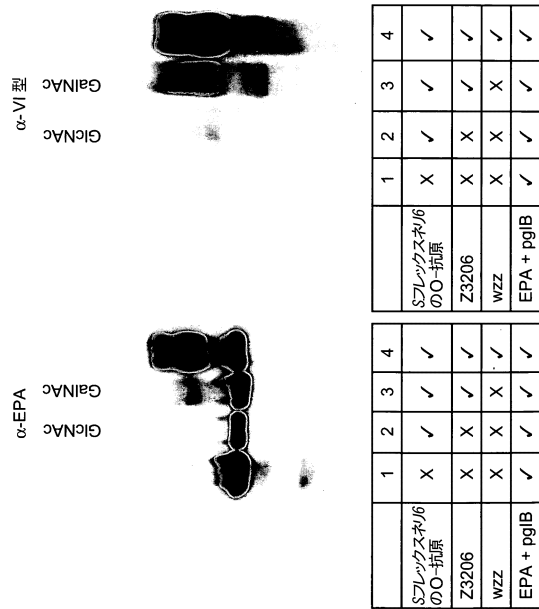
11图

【図 1 2】



12

【図 1 3】



13

【配列表】

0005968784000001.app

フロントページの続き

(51)Int.Cl. F I  
C O 7 K 14/195 (2006.01) C O 7 K 14/195

(74)代理人 100171505

弁理士 内藤 由美

(72)発明者 マイケル ワッカー

スイス国 シーエイチー 8 1 0 3 ウンテレングストリンゲン ダーリエンシュトラーセ 1 7

(72)発明者 チャールズ ワエクター

アメリカ合衆国 ケンタッキー州 4 0 5 0 2 レキシントン レインツリー ロード 6 0 5

審査官 松原 寛子

(56)参考文献 特表 2 0 0 8 - 5 3 9 7 4 3 (J P, A)

国際公開第 2 0 0 9 / 1 0 4 0 7 4 (W O, A 1)

国際公開第 2 0 0 9 / 0 8 9 3 9 6 (W O, A 1)

Feldman et al., Proc. Natl. Acad. Sci., Vol.102, No.8, p.3016-3021(2005)

Dmitriev et al., Eur. J. Biochem., Vol.98, p.309-316(1979)

Rush et al., The Journal of Biological Chemistry, Vol.285, No.3, p.1671-1680(2010), published online 2009-11-18

(58)調査した分野(Int.Cl., DB名)

C 1 2 N 1 / 2 1

C 1 2 P 2 1 / 0 2

J S T P l u s / J M E D P l u s / J S T 7 5 8 0 (J D r e a m I I I)

C A p l u s / R E G I S T R Y (S T N)

C A p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)