

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2010-501074

(P2010-501074A)

(43) 公表日 平成22年1月14日(2010.1.14)

(51) Int.Cl.		F I		テーマコード (参考)
<b>GO 1 N 33/53</b>	<b>(2006.01)</b>	GO 1 N 33/53	S	
<b>GO 1 N 37/00</b>	<b>(2006.01)</b>	GO 1 N 37/00	1 O 2	

審査請求 未請求 予備審査請求 未請求 (全 55 頁)

(21) 出願番号 特願2009-524675 (P2009-524675) (86) (22) 出願日 平成19年8月14日 (2007. 8. 14) (85) 翻訳文提出日 平成21年3月13日 (2009. 3. 13) (86) 国際出願番号 PCT/US2007/018103 (87) 国際公開番号 W02008/021415 (87) 国際公開日 平成20年2月21日 (2008. 2. 21) (31) 優先権主張番号 60/837, 868 (32) 優先日 平成18年8月14日 (2006. 8. 14) (33) 優先権主張国 米国 (US) (31) 優先権主張番号 60/837, 869 (32) 優先日 平成18年8月14日 (2006. 8. 14) (33) 優先権主張国 米国 (US)	(71) 出願人 596060697 マサチューセッツ・インスティテュート・ オブ・テクノロジー アメリカ合衆国マサチューセッツ州021 39ケンブリッジ, マサチューセッツ・ア ヴェニュー・77 (74) 代理人 100078282 弁理士 山本 秀策 (74) 代理人 100062409 弁理士 安村 高明 (74) 代理人 100113413 弁理士 森下 夏樹
---	--

最終頁に続く

(54) 【発明の名称】 グリカンのデータマイニングシステム

## (57) 【要約】

本発明は、グリカンおよびそれらの相互作用パートナーを解析するためのシステムを提供する。本発明のシステムは、糖タンパク質結合相互作用の特定および解析に特に有用である。一実施形態において、本発明によって提供される方法は、グリカン構造物の特徴を決定する工程と、複数のグリカンに対するグリカン結合タンパク質の結合を、該グリカンに存在する該決定された特徴と関連させる工程とを含む。一局面において、上記の関連させる工程は、上記特徴を含む複数のグリカンに結合する上記グリカン結合タンパク質の結合データを比較することと、結合の度合いを特徴の存在または非存在と関連させることとを含む。

**【特許請求の範囲】****【請求項 1】**

グリカン構造物の特徴を決定する工程と、  
複数のグリカンに対するグリカン結合タンパク質の結合を、該グリカンに存在する該決定された特徴と関連させる工程とを含む、方法。

**【請求項 2】**

前記関連させる工程が、前記特徴を含む複数のグリカンに結合する前記グリカン結合タンパク質の結合データを比較することと、結合の度合いを特徴の存在または非存在と関連させることとを含む、請求項 1 に記載の方法。

**【請求項 3】**

前記特徴が、単糖レベルの特徴、より高次の特徴、G B P 結合特徴、およびそれらの組合せからなる群より選択される、請求項 1 または請求項 2 に記載の方法。

**【請求項 4】**

前記単糖レベルの特徴が、組成、明示的な組成、およびそれらの組合せからなる群より選択される、請求項 3 に記載の方法。

**【請求項 5】**

前記単糖レベルの特徴が末端組成を含む、請求項 3 に記載の方法。

**【請求項 6】**

前記より高次の特徴が、ペア、トリプレット、クワドルプレット、クラスター、平均リーフ深度、リーフ数、およびそれらの組合せからなる群より選択される、請求項 3 に記載の方法。

**【請求項 7】**

前記ペアが、通常ペア、末端ペア、およびそれらの組合せからなる群より選択される、請求項 6 に記載の方法。

**【請求項 8】**

前記トリプレットが、通常、末端、表面、およびそれらの組合せからなる群より選択される、請求項 6 に記載の方法。

**【請求項 9】**

前記クワドルプレットが、通常、末端、表面、およびそれらの組合せからなる群より選択される、請求項 6 に記載の方法。

**【請求項 10】**

前記 G B P 結合特徴が、グリカン毎の平均信号、信号対雑音比、およびそれらの組合せからなる群より選択される、請求項 3 に記載の方法。

**【請求項 11】**

グリカン構造物の特徴を決定する工程と、  
複数のグリカンに対するグリカン結合タンパク質の結合を、該グリカンに存在する該決定された特徴と関連させる工程と、  
該関連に基づき、該グリカン結合タンパク質により結合された一連のグリカンを決する工程とを含む、方法。

**【請求項 12】**

前記決定された一連のグリカンを含むグリカン結合タンパク質特異的グリカンアレイを調製する工程をさらに含む、請求項 11 に記載の方法。

**【発明の詳細な説明】****【技術分野】****【0001】**

(優先権主張)

この出願は、2006年8月14日に出願された同時係属中の米国仮特許出願第60/837,868号および2006年8月14日に出願された同時係属中の米国仮特許出願第60/837,869号への優先権を主張する。これら各々の先の出願の全体の内容は、参考として本明細書に援用される。

10

20

30

40

50

## 【 0 0 0 2 】

( 政府支援 )

本発明は、契約番号 U 5 4 G M 6 2 1 1 6 で米国総合医科学研究所から、および契約番号 G M 5 7 0 7 3 で国立衛生研究所から与えられた米国政府の支援で行われた。米国政府は、本発明において一定の権利を有する。

## 【 背景技術 】

## 【 0 0 0 3 】

複合糖質またはグリカンの構造機能相関に対する統合的手法であるグリコミクスは、ポストゲノミクスの細胞生物学および分子生物学における重要なパラダイムとして出現しつつある。過去数年間で、細胞増殖および発生、腫瘍の増殖および転移、抗凝血、免疫認識 / 応答、細胞間情報交換、ならびに微生物病原性などの基本的な生物学的プロセスにおけるグリカンの公知の生物学的役割は劇的に増加した。グリカンは、細胞表面、ならびに細胞およびその細胞外環境間の界面の主要な構成要素である。その結果、グリカンは、増殖因子、サイトカイン、免疫受容体、および酵素など多数のタンパク質と相互作用してそれらの活動を調節し、したがって上記の生物学的プロセスに影響を及ぼす。

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 4 】

したがって、グリカン結合能力を特定および / または特徴付ける必要がある。

## 【 課題を解決するための手段 】

## 【 0 0 0 5 】

要旨

本発明は、グリカンおよびそれらの相互作用パートナーを解析するためのシステムを提供する。本発明のシステムは、糖タンパク質結合相互作用の特定および解析に特に有用である。本明細書において、本発明のシステムは、幾つかの異なる糖タンパク質解析に適用され、全ての場合で相互作用の特徴を特定することに成功した。したがって、本発明のシステムの原理は、グリカン相互作用の全域にわたって広く適用可能である。

## 【 図面の簡単な説明 】

## 【 0 0 0 6 】

【 図 1 】本明細書で利用されるデータマイニングプラットフォームを示す図である。パネル A で示されるのは、データマイニングプラットフォームの主要構成要素である。その特徴は、データベースから抽出されるデータオブジェクトから導出される。特徴は、パターンまたは規則を導出するための分類法により使用されるデータセット中に調製される。パネル B は、ユーザがグリカンアレイデータにデータマイニングプロセスを適用することを可能にする、ある種のソフトウェアモジュールを示す。

【 図 2 】特徴の模式的記述を示す図である。パネル A で示されるのは、ペア ( pair )、トリプレット ( triplet )、およびクワドルプレット ( quadruplet ) の定義を例証する代表的な高マンノースモチーフである。パネル B で示されるのは、異なる種類のトリプレットを例証する代表的な O 結合グリカン [ コア 2 型 ] モチーフである。単糖を表わすために、以下の記号命名法：

## 【 数 1 】

●Man ●Gal ■GlcNAc ▲Fuc ◆Neu5Ac ◆Neu5Gc ◆KDN

を使用した。

【 図 3 】高親和性結合の分類を図示する。図中で示されるのは、グリカンアレイのグリカン [ x 軸に連続的に付番されている ] に対してスクリーニングされるガレクチン 3 の信号対雑音比 [ y 軸 ] である。赤色の点線は、高親和性結合体であるグリカンを分類するために自由裁量で定義される閾値を示す。これらの閾値は、解析に使用された G B P の各々について定義された。

【 図 4 】 D C S I G N の C R D ( 灰色のリボントレース ) に対する、 L e w i s <sup>x</sup> モチ

10

20

30

40

50

ーフを含有するグリカン構造物  $\text{Gal} \quad 4 (\text{Fuc} \quad 3) \text{GlcNAc} \quad 3 \text{Gal}$  の結合を図示する。単糖および結合は、グリカン構造物上に標識化している。グリカン結合部位に近接する  $\text{Gal}$  の 3-OH は、赤色の円で示している。したがって、この水酸基における任意の置換は結合に負の効果を示し、データマイニングから得られる分類指標規則と一致する。

【図 5】野生型 HA の典型的な配列アラインメント。配列は、NCBI インフルエンザウイルス配列データベース (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) から取得した。

【図 5 A】野生型 HA の典型的な配列アラインメント。配列は、NCBI インフルエンザウイルス配列データベース (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) から取得した。

【図 5 B】野生型 HA の典型的な配列アラインメント。配列は、NCBI インフルエンザウイルス配列データベース (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) から取得した。

【図 5 C】野生型 HA の典型的な配列アラインメント。配列は、NCBI インフルエンザウイルス配列データベース (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>) から取得した。

【図 6】グリカン受容体の特異性を理解するための枠組み。2-3 および / または 2-6 結合グリカンは、異なるトポロジをとることができる。本発明によると、HA ポリペプチドがこれらのトポロジのある種に結合する能力は、異なる宿主、例えばヒトの感染を媒介する能力を HA ポリペプチドに授与する。この図で例証されているように、本発明は、2 つの特に関連性のあるトポロジ、すなわち「円錐形」トポロジおよび「傘形」トポロジを定義する。円錐形トポロジをとることができるのは、2-3 および / または 2-6 結合グリカンであり、コアに結合した短鎖オリゴ糖または分岐オリゴ糖に典型的である（しかし、ある種の長鎖オリゴ糖がこのトポロジをとることができる）。傘形トポロジをとることができるのは、2-6 結合グリカンであり（おそらくは、2-6 結合に存在する追加的な C5-C6 結合により許容されるコンフォメーションの多数性の増加による）、支配的に傘形トポロジをとるのは、 $\text{Neu5Ac} \quad 2-6 \text{Gal} \quad 1-3 / 4 \text{GlcNAc}$  のモチーフを特に含有する、長鎖オリゴ糖または長鎖オリゴ糖分岐部を有する分岐グリカンである。本明細書において、HA ポリペプチドが傘形グリカントポロジに結合する能力は、ヒト受容体への結合および / またはヒト感染を媒介する能力を授与する。

【図 7】HA 残基と円錐形または傘形グリカントポロジとの相互作用。HA グリカン共結晶を解析すると、HA 結合部位に対する  $\text{Neu5Ac}$  の位置は、ほとんど不変であることが明らかになる。 $\text{Neu5Ac}$  との接触には、F98、S/T136、W153、H183、および L/I194 などの高度に保存された残基が関与する。他の糖との接触には、糖結合が 2-3 かまたは 2-6 かどうか、およびグリカントポロジが円錐形かまたは傘形かどうかに依存して、異なる残基が関与する。例えば、円錐形トポロジでは、 $\text{Neu5Ac}$  および  $\text{Gal}$  糖と主に接触する。E190 および Q226 は、この結合に特に重要な役割を果たす。この図は、円錐形構造への結合に参加し得る他の位置（例えば、137、145、186、187、193、222）も例証する。ある場合において、異なる残基が異なるグリカン構造物と異なる接触することがある。これらの位置のアミノ酸の種類は、HA ポリペプチドが、グリカン構造物における異なる修飾および / または分岐パターンを有する受容体に結合する能力に影響を与えることがある。傘形トポロジでは、 $\text{Neu5Ac}$  および  $\text{Gal}$  を越えた糖と接触する。この図は、傘形構造への結合に参加し得る残基（例えば、137、145、156、159、186、187、189、190、192、193、196、222、225、226）を例証する。ある場合において、異なる残基が異なるグリカン構造物と異なる接触することがある。これらの位置のアミノ酸の種類は、HA ポリペプチドが、グリカン構造物における異なる修飾および / または分岐パターンを有する受容体に結合する能力に影響を与えることがある。一部実施形

10

20

30

40

50

態において、１９０位のＤ型残基および／または２２５位のＤ型残基が、傘形トポロジーに対する結合に寄与する。

【図８】典型的な円錐形トポロジー。この図は、円錐形トポロジーをとる、ある典型的な（しかし網羅的ではない）グリカン構造物を例証する。

【図９】典型的な傘形トポロジー。この図は、傘形トポロジーをとる、ある典型的な（しかし網羅的ではない）グリカン構造物を例証する。

【図１０－１】ＨＡグリカン結合ドメインの配列アラインメント。灰色：シアル酸に対する結合に関与する保存されたアミノ酸。赤色：Neu5Ac<sub>2</sub>3/6Galモチーフに対する結合に関与する特定のアミノ酸。黄色：Q226（１３７、１３８）およびE190（１８６、２２８）の位置決めに影響を及ぼすアミノ酸。緑色：Neu5Ac<sub>2</sub>3/6Galモチーフに結合した他の単糖（または修飾）に対する結合に関与するアミノ酸。ASI30、APR34、ADU63、ADS97、およびViet04についての配列は、それぞれの結晶構造から取得した。他の配列は、SwissProt（<http://us.expasy.org>）から取得した。略語：ADA76、A/アヒル/Alberta/35/76（H1N1）；ASI30、A/ブタ/Iowa/30（H1N1）；APR34、A/Puerto Rico/8/34（H1N1）；ASC18、A/South Carolina/1/18（H1N1）、AT91、A/Texas/36/91（H1N1）；ANY18、A/New York/1/18（H1N1）；ADU63、A/アヒル/Ukraine/1/63（H3N8）；AAI68、A/Aichi/2/68（H3N2）；AM99、A/Moscow/10/99（H3N2）；ADS97、A/アヒル/Singapore/3/97（H5N3）；Viet04、A/Vietnam/1203/2004（H5N1）。

【図１０－２】ＨＡグリカン結合ドメインの配列アラインメント。灰色：シアル酸に対する結合に関与する保存されたアミノ酸。赤色：Neu5Ac<sub>2</sub>3/6Galモチーフに対する結合に関与する特定のアミノ酸。黄色：Q226（１３７、１３８）およびE190（１８６、２２８）の位置決めに影響を及ぼすアミノ酸。緑色：Neu5Ac<sub>2</sub>3/6Galモチーフに結合した他の単糖（または修飾）に対する結合に関与するアミノ酸。ASI30、APR34、ADU63、ADS97、およびViet04についての配列は、それぞれの結晶構造から取得した。他の配列は、SwissProt（<http://us.expasy.org>）から取得した。略語：ADA76、A/アヒル/Alberta/35/76（H1N1）；ASI30、A/ブタ/Iowa/30（H1N1）；APR34、A/Puerto Rico/8/34（H1N1）；ASC18、A/South Carolina/1/18（H1N1）、AT91、A/Texas/36/91（H1N1）；ANY18、A/New York/1/18（H1N1）；ADU63、A/アヒル/Ukraine/1/63（H3N8）；AAI68、A/Aichi/2/68（H3N2）；AM99、A/Moscow/10/99（H3N2）；ADS97、A/アヒル/Singapore/3/97（H5N3）；Viet04、A/Vietnam/1203/2004（H5N1）。

【図１１－１】Ｈ１型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１１－２】Ｈ１型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１２－１】Ｈ３型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１２－２】Ｈ３型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－１】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－２】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－３】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－４】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－５】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－６】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－７】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図１３－８】Ｈ５型ＨＡの保存される亜配列特徴を例証する配列アラインメント。

【図 13 - 9】H5 型 HA の保存される亜配列特徴を例証する配列アラインメント。

【図 14】Neu5Ac<sub>2-3</sub>Gal および Neu5Ac<sub>2-6</sub>Gal モチーフのコンフォメーション地図および溶媒露出度。パネル A は、Neu5Ac<sub>2-3</sub>Gal 結合のコンフォメーション地図を示す。丸で囲まれた領域 2 は、APR34\_\_H1\_\_23、ADU63\_\_H3\_\_23、および ADS97\_\_H5\_\_23 の共結晶構造に観察されたトランス型コンフォメーションである。丸で囲まれた領域 1 は、AAI68\_\_H3\_\_23 の共結晶構造に観察されたコンフォメーションである。パネル B は、Neu5Ac<sub>2-6</sub>Gal のコンフォメーション地図を示し、シス型コンフォメーション（丸で囲まれた領域 3）が全ての HA<sub>2-6</sub>シアル酸含有グリカン共結晶構造に観察された。パネル C は、それぞれの HA<sub>2-6</sub>グリカン共結晶構造における、Neu5Ac<sub>2-3</sub>と<sub>2-6</sub>シアル酸含有オリゴ糖との溶媒露出面積（SASA）の差を示す。赤色および青色バーは、それぞれ<sub>2-6</sub>（正の値）または<sub>2-3</sub>（負の値）シアル酸含有グリカンの Neu5Ac が、グリカン結合部位とより多く接触することを示す。パネル D は、ブタおよびヒト H1 に結合した<sub>2-3</sub>シアル酸含有グリカンの NeuAc の SASA の差（H1<sub>2-3</sub>）、トリおよびヒト H3 に結合した<sub>2-3</sub>シアル酸含有グリカンの NeuAc の SASA の差（H3<sub>2-3</sub>）、ならびにブタおよびヒト H1 に結合した<sub>2-6</sub>シアル酸含有グリカンの NeuAc の SASA の差（H1<sub>2-6</sub>）を示す。H3<sub>2-3</sub>の青色バーが負であるのは、トリ H3 の接触と比較して、ヒト H3 型 HA の Neu5Ac<sub>2-3</sub>Gal との接触がより少ないことを示す。ねじれ角：C2-C1-O-C3（Neu5Ac<sub>2-3</sub>/6 結合について）；C1-O-C3-H3（Neu5Ac<sub>2-3</sub>Gal について）または C1-O-C6-C5（Neu5Ac<sub>2-6</sub>Gal について）；O-C6-C5-H5（Neu5Ac<sub>2-6</sub>Gal について）結合。地図は、Martin Frank 博士および Claus Wilhelm von der Leth 博士（German Cancer Research Institute、Heidelberg、Germany）により開発された GlycoMaps DB（<http://www.glycosciences.de/modeling/glycomapsdb/>）から取得した。高エネルギーから低エネルギーへの着色体スキームは、それぞれ明るい赤から明るい緑へである。

【図 15 A】<sub>2-3</sub>/6 シアル酸含有グリカンに対する、H1 型、H3 型、および H5 型 HA の結合に關与する残基。パネル A ~ D は、それぞれ ASI30\_\_H1、APR34\_\_H1、ADU63\_\_H3、および ADS97\_\_H5 共結晶構造における、<sub>2-3</sub>および<sub>2-6</sub>シアル酸含有グリカンと相互作用する残基の溶媒露出面積（SASA）の差（横座標における）を示す。緑色バーは、グリカンと直接相互作用する残基に対応し、明橙色バーは、Glu / Asp190 および Glu / Leu226 に近位する残基に対応する。緑色バーの値が正であることは、その残基が<sub>2-6</sub>シアル酸含有グリカンとより多く接触することを示すが、値が負であることは、<sub>2-3</sub>シアル酸含有グリカンとより多く接触することを示す。パネル E は、<sub>2-3</sub>/6 シアル酸含有グリカンへの結合に關与する、H1 型、H3 型、および H5 型 HA の残基を表形式で要約する。<sub>2-3</sub>シアル酸含有グリカンに対する結合に關与するある種の重要な残基は青色に着色され、<sub>2-6</sub>シアル酸含有グリカンに対する結合に關与するある種の重要な残基は赤色に着色されている。

【図 15 B】<sub>2-3</sub>/6 シアル酸含有グリカンに対する、H1 型、H3 型、および H5 型 HA の結合に關与する残基。パネル A ~ D は、それぞれ ASI30\_\_H1、APR34\_\_H1、ADU63\_\_H3、および ADS97\_\_H5 共結晶構造における、<sub>2-3</sub>および<sub>2-6</sub>シアル酸含有グリカンと相互作用する残基の溶媒露出面積（SASA）の差（横座標における）を示す。緑色バーは、グリカンと直接相互作用する残基に対応し、明橙色バーは、Glu / Asp190 および Glu / Leu226 に近位する残基に対応する。緑色バーの値が正であることは、その残基が<sub>2-6</sub>シアル酸含有グリカンとより多く接触することを示すが、値が負であることは、<sub>2-3</sub>シアル酸含有グリカンとより多く接触することを示す。パネル E は、<sub>2-3</sub>/6 シアル酸含有グリカンへの結合に

関与する、H 1 型、H 3 型、および H 5 型 H A の残基を表形式で要約する。2 3 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は青色に着色され、2 6 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は赤色に着色されている。

【図 1 5 C】 2 3 / 6 シアル酸含有グリカンに対する、H 1 型、H 3 型、および H 5 型 H A の結合に関与する残基。パネル A ~ D は、それぞれ A S I 3 0 \_\_ H 1、A P R 3 4 \_\_ H 1、A D U 6 3 \_\_ H 3、および A D S 9 7 \_\_ H 5 共結晶構造における、2 3 および 2 6 シアル酸含有グリカンと相互作用する残基の溶媒露出面積 ( S A S A ) の差 ( 横座標における ) を示す。緑色バーは、グリカンと直接相互作用する残基に対応し、明橙色バーは、G l u / A s p 1 9 0 および G l n / L e u 2 2 6 に近位する残基に対応する。緑色バーの値が正であることは、その残基が 2 6 シアル酸含有グリカンとより多く接触することを示すが、値が負であることは、2 3 シアル酸含有グリカンとより多く接触することを示す。パネル E は、2 3 / 6 シアル酸含有グリカンへの結合に関与する、H 1 型、H 3 型、および H 5 型 H A の残基を表形式で要約する。2 3 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は青色に着色され、2 6 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は赤色に着色されている。

10

【図 1 5 D】 2 3 / 6 シアル酸含有グリカンに対する、H 1 型、H 3 型、および H 5 型 H A の結合に関与する残基。パネル A ~ D は、それぞれ A S I 3 0 \_\_ H 1、A P R 3 4 \_\_ H 1、A D U 6 3 \_\_ H 3、および A D S 9 7 \_\_ H 5 共結晶構造における、2 3 および 2 6 シアル酸含有グリカンと相互作用する残基の溶媒露出面積 ( S A S A ) の差 ( 横座標における ) を示す。緑色バーは、グリカンと直接相互作用する残基に対応し、明橙色バーは、G l u / A s p 1 9 0 および G l n / L e u 2 2 6 に近位する残基に対応する。緑色バーの値が正であることは、その残基が 2 6 シアル酸含有グリカンとより多く接触することを示すが、値が負であることは、2 3 シアル酸含有グリカンとより多く接触することを示す。パネル E は、2 3 / 6 シアル酸含有グリカンへの結合に関与する、H 1 型、H 3 型、および H 5 型 H A の残基を表形式で要約する。2 3 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は青色に着色され、2 6 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は赤色に着色されている。

20

30

【図 1 5 E】 2 3 / 6 シアル酸含有グリカンに対する、H 1 型、H 3 型、および H 5 型 H A の結合に関与する残基。パネル A ~ D は、それぞれ A S I 3 0 \_\_ H 1、A P R 3 4 \_\_ H 1、A D U 6 3 \_\_ H 3、および A D S 9 7 \_\_ H 5 共結晶構造における、2 3 および 2 6 シアル酸含有グリカンと相互作用する残基の溶媒露出面積 ( S A S A ) の差 ( 横座標における ) を示す。緑色バーは、グリカンと直接相互作用する残基に対応し、明橙色バーは、G l u / A s p 1 9 0 および G l n / L e u 2 2 6 に近位する残基に対応する。緑色バーの値が正であることは、その残基が 2 6 シアル酸含有グリカンとより多く接触することを示すが、値が負であることは、2 3 シアル酸含有グリカンとより多く接触することを示す。パネル E は、2 3 / 6 シアル酸含有グリカンへの結合に関与する、H 1 型、H 3 型、および H 5 型 H A の残基を表形式で要約する。2 3 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は青色に着色され、2 6 シアル酸含有グリカンに対する結合に関与するある種の重要な残基は赤色に着色されている。

40

【図 1 6】二分岐 2 6 シアル酸含有グリカン ( 円錐形トポロジー ) に対する、V i e t 0 4 \_\_ H 5 型 H A の結合。伸長型コンフォメーションにある N e u 5 A c 2 6 G a l 結合を伴った V i e t 0 4 \_\_ H 5 のグリカン結合部位を表面レンダリングした立体像である ( 百日咳トキシン共結晶構造から取得した ; P D B I D : 1 P T O ) 。 L y s 1 9 3 ( 橙色 ) は、このコンフォメーションにあるグリカンとの任意の接触も有しない。このコンフォメーションにあるグリカンに対する結合に潜在的に関与する追加的なアミノ酸は、A s n 1 8 6、L y s 2 2 2、および S e r 2 2 7 である。しかしながら、シス型コン

50

フォメーションにある 2 6 シアル酸含有オリゴ糖に対する H A の結合に観察されるある種の接触は、伸長型コンフォメーションにおいては存在しない。任意の特定の理論により束縛されることは望まないが、これは、伸長型コンフォメーションがシス型コンフォメーションほど至適には H A に結合できないことを示唆することに、発明者らは留意する。Neu5Ac 2 6 Gal 1 4 GlcNAc 分岐部が Man 1 3 Man (PDB ID: 1LGC) および Man 1 6 Man (PDB ID: 1ZAG) に結合した分岐 N 結合グリカンの構造を、この結合のシス型および伸長型コンフォメーションの両方について、Viet 04 \_\_ H5 型 H A 結合部位の Neu5Ac 2 6 Gal 結合にスーパーインポーズした。このスーパーインポーズにより、Neu5Ac 2 6 Gal 1 4 GlcNAc 分岐部がコアの Man 1 6 Man に結合した構造は、結合部位との好ましくない立体重複を有することが示される (両コンフォメーションにおいて)。一方では、この分岐部がコアの Man 1 3 Man に結合した構造 (図中、トリマンノースコアは紫色で示されている) は、シス型コンフォメーションでは Lys 193 との立体重複を有するが、伸長型コンフォメーションでは Lys 193 と任意の接触もせず、至適というほどではないが結合が可能である。

【図 17】野生型の H1 型、H3 型、および H5 型 H A の産生。パネル A は、H1N1 (A/South Carolina/1/1918)、H3N2 (A/Moscow/10/1999)、および H5N1 (A/Vietnam/1203/2004) に由来する可溶性形態の H A タンパク質を、4 ~ 12 % の SDS ポリアクリルアミドゲル上で流し、ニトロセルロース膜にプロットしたことを示す。H1N1 型 H A は、ヤギ抗 A 型インフルエンザ抗体および抗ヤギ IgG HRP を使用して探索した。H3N2 は、フェレット抗 H3N2 型 H A 抗血清および抗フェレット HRP を使用して探索した。H5N1 は、抗トリ H5N1 型 H A 抗体および抗ウサギ IgG HRP を使用して探索した。H1N1 型 H A および H3N2 型 H A は H A 0 として存在するが、H5N1 型 H A は、H A 0 および H A 1 の両方として存在する。パネル B は、全長 H5N1 型 H A および 2 つの変異体 (Glu190 Asp、Lys193 Ser、Gly225 Asp、Gln226 Leu、すなわち「DSDL」、および Glu190 Asp、Lys193 Ser、Gln223 Leu、Gly228 Ser、すなわち「DSL S」) を、SDS ポリアクリルアミドゲル上で流し、ニトロセルロース膜にプロットしたことを示す。H A は、抗トリ H5N1 型抗体および抗ウサギ IgG HRP を使用して探索した。

【図 18】上気道組織切片のレクチン染色。ジャカリン (緑色) および Con A (赤色) で気管組織を共染色すると、ジャカリンが気管の頂端面上の杯状細胞と優先的に結合する (O 結合グリカンに特異的に結合すること、および con A が纖毛性気管上皮細胞と優先的に結合する (N 結合グリカンに特異的に結合すること) が明らかになる。任意の特定の理論により束縛されることは望まないが、発明者らは、この知見が、杯状細胞は O 結合グリカンを支配的に発現し、纖毛性上皮細胞は N 結合グリカンを支配的に発現することを示唆することに留意する。ジャカリンおよび SNA (赤色; 2 6 と特異的に結合する) で気管を共染色すると、SNA が杯状細胞および纖毛細胞の両方と結合することが示される。一方では、ジャカリン (緑色) および 2 3 シアル酸含有グリカンに特異的に結合する MAL (赤色) で共染色すると、MAL は多列気管上皮と極弱く結合するか無結合であるが、組織の下部領域とは広範に結合することが示される。まとめると、レクチン染色データは、N 結合および O 結合の両グリカンの一部としての 2 6 シアル酸含有グリカンが、それぞれ気管上皮頂端側の纖毛細胞および杯状細胞において支配的に発現され、広範に分布することを示した。

【図 19】組換え野生型および突然変異型 H A の組織切片に対する結合。気管組織切片、気管支組織切片、および肺胞組織切片に結合する野生型 (WT)、DSL S、および DSDL が示されている。WT の場合、白色矢印は肺胞組織切片に結合する H A (緑色) を示す。DSL S 突然変異体の場合、気管および気管支組織切片の白色矢印は、組織の頂端側に結合するこの突然変異体 H A (緑色) を示す。DSDL 突然変異体は任意の組織切片とも結合しないことに留意されたい。

10

20

30

40

50



## 【発明を実施するための形態】

## 【0007】

## 定義

親和性：当該技術分野で公知のように、「親和性」とは、特定のリガンド（例えば、H A ポリペプチド）がそのパートナー（例えば、H A 受容体）に結合する緊密さの尺度である。親和性は様々な方法で測定できる。

## 【0008】

生物学的に活性がある：本明細書において、「生物学的に活性のある」という語句は、生体システム、および特に生物において活性を示す任意の薬剤の特徴を指す。例えば、生物に投与された際にその生物に生物学的効果を示す薬剤は、生物学的に活性があると見なされる。タンパク質またはポリペプチドが生物学的に活性である特定の実施形態において、タンパク質またはポリペプチドの少なくとも1つの生物学的活性を共有するそのタンパク質またはポリペプチドの部分は、典型的には「生物学的に活性のある」部分と呼ばれる。

10

## 【0009】

広域性ヒト結合（B S H B）H 5 型 H A ポリペプチド：本明細書において、「広域性ヒト結合 H 5 型 H A」という語句は、ヒト上皮組織に見出される H A 受容体、および特に 2 6 シアル酸含有グリカンを含むヒト H A 受容体に結合する H 5 型 H A ポリペプチドの型を指す。さらに、発明の B S H B 性 H 5 型 H A は、複数の異なる 2 6 シアル酸含有グリカンに結合する。一部実施形態において、B S H B 性 H 5 型 H A は、ヒト試料に見出される十分な数の 2 6 シアル酸含有グリカンに結合し、そのためそれらを含むウイルスはヒト集団に感染する広範な能力、および特にそれらの集団の上気道受容体に結合する広範な能力を有する。一部実施形態において、B S H B 性 H 5 型 H A は、本明細書において、傘形グリカン（例えば、長鎖 2 6 シアル酸含有グリカン）に結合する。

20

## 【0010】

特徴部分：本明細書において、タンパク質またはポリペプチドの「特徴部分」という語句は、共にタンパク質またはポリペプチドの特徴であるアミノ酸の連続的な伸長、またはアミノ酸の連続的な伸長の集まりを含む部分である。そのような連続的な伸長の各々は、一般的に少なくとも2個のアミノ酸を含む。さらに、当業者であれば、タンパク質の特徴であるためには、典型的には少なくとも5個、10個、15個、20個、またはそれより多くのアミノ酸が必要とされることを認識しよう。一般的に、特徴部分は、上記で特定した配列同一性に加えて、少なくとも1つの機能的な特徴を、関連する完全なタンパク質と共有する部分である。

30

## 【0011】

特徴配列：「特徴配列」とは、ポリペプチドまたは核酸ファミリーの全メンバーに見出され、したがって当業者がファミリーのメンバーを定義するために使用できる配列である。

## 【0012】

円錐形トポロジー：「円錐形トポロジー」という語句は、ある種のグリカンおよび特に H A 受容体上のグリカンがとる3次元配置を指すために本明細書で使用する。図6に例証したように、円錐形トポロジーをとることができるのは、2 3 シアル酸含有グリカンまたは 2 6 シアル酸含有グリカンであり、幾つかの長鎖オリゴヌクレオチドもこのコンフォメーションをとることができるが、短鎖オリゴヌクレオチド鎖に典型的である。円錐形トポロジーは、約 - 60、60、または180の値（ $C1 - C2 - O - C3 / C6$ ）により決定される最小エネルギーコンフォメーションの3つ領域をサンプリングする N e u 5 A c 2 3 G a l 結合のグリコシドのねじれ角によって特徴付けられ、（ $C2 - O - C3 / C6 - H3 / C5$ ）は - 60 から 60 をサンプリングする（図14）。図8は、円錐形トポロジーをとるグリカンのある代表的な（しかし網羅的でない）例を表す。

40

## 【0013】

対応する：本明細書において、「対応する」という用語は、H A ポリペプチドのアミノ

50

酸残基の位置 / 同一性を指定するために使用されることが多い。当業者であれば、標準付番方式（野生型の H 3 型 H A に基づく）が簡潔さのために本明細書で使用され（例えば図 5 および図 10 ~ 13 に例証されているように）、そのため 190 位の残基に「対応する」のアミノ酸は、例えば実際に特定のアミノ酸鎖の 190 番目のアミノ酸である必要はなく、むしろ野生型の H 3 型 H A の 190 位に見出される残基に対応することを理解し、当業者であれば、対応するアミノ酸を特定する方法を容易に理解する。

#### 【0014】

離隔度離れた：本明細書において、「離隔度離れた」アミノ酸とは、グリカン結合に間接的な影響を及ぼす H A アミノ酸である。例えば、1 離隔度離れたアミノ酸とは、（１）直接結合するアミノ酸と相互作用する；および / または（２）そうでなければ、直接結合するアミノ酸が宿主細胞 H A 受容体に会合するグリカンと相互作用する能力に影響を及ぼすかのいずれかであってよく、そのような 1 離隔度離れたアミノ酸は、グリカン自体に直接結合してもよく、結合しなくてもよい。2 離隔度離れたアミノ酸は、（１）1 離隔度離れたアミノ酸と相互作用する；および / または（２）そうでなければ、1 離隔度離れたアミノ酸が、直接結合するアミノ酸などと相互作用する能力に影響を及ぼすかのいずれかである。

10

#### 【0015】

直接結合するアミノ酸：本明細書において、「直接結合するアミノ酸」という語句は、宿主細胞 H A 受容体と会合する 1 つ以上のグリカンと直接的に相互作用する H A ポリペプチドアミノ酸を指す。

20

#### 【0016】

操作された：「操作された」という用語は、本明細書において、そのアミノ酸配列が人為的に選択されたポリペプチドを表す。例えば、操作された H A ポリペプチドは、天然のインフルエンザ分離株に見出される H A ポリペプチドのアミノ酸配列と異なるアミノ酸配列を有する。一部実施形態において、操作された H A ポリペプチドは、NCBI データベースに含まれている H A ポリペプチドのアミノ酸配列と異なるアミノ酸配列を有する。

#### 【0017】

H 1 ポリペプチド：「H 1 ポリペプチド」とは、その用語が本明細書で使用されているように、そのアミノ酸配列が、H 1 の特徴であり H 1 を他の H A 亜型と区別する少なくとも 1 つの配列エレメントを含む H A ポリペプチドである。代表的なそのような配列エレメントは、例えば図 5 および図 10 ~ 11 に例証されるものなどのようにアラインメントにより決定でき、例えば H A 配列エレメントの H 1 特異的な実施形態に関して本明細書に記載したものなどを含むことができる。

30

#### 【0018】

H 3 ポリペプチド：「H 3 ポリペプチド」とは、その用語が本明細書で使用されているように、そのアミノ酸配列が、H 3 の特徴であり H 3 を他の H A 亜型と区別する少なくとも 1 つの配列エレメントを含む H A ポリペプチドである。代表的なそのような配列エレメントは、例えば図 5、図 10、および図 12 に例証されたものなどのようにアラインメントにより決定でき、例えば H A 配列エレメントの H 3 特異的な実施形態に関して本明細書に記載したものなどを含むことができる。

40

#### 【0019】

H 5 ポリペプチド：「H 5 ポリペプチド」とは、その用語が本明細書で使用されているように、そのアミノ酸配列が、H 5 の特徴であり H 5 を他の H A 亜型と区別する少なくとも 1 つの配列エレメントを含む H A ポリペプチドである。代表的なそのような配列エレメントは、例えば図 5、図 10、および図 13 に例証されたものなどのようにアラインメントにより決定でき、例えば H A 配列エレメントの H 5 特異的な実施形態に関して本明細書に記載したものなどを含むことができる。

#### 【0020】

赤血球凝集素（H A）ポリペプチド：本明細書において、「赤血球凝集素ポリペプチド」（「または H A ポリペプチド」）という用語は、そのアミノ酸配列が H A の特徴配列を

50

少なくとも1つ含むポリペプチドを指す。インフルエンザ分離株に由来する多種多様なHA配列は、当該技術分野において公知であり、実際に、米国バイオテクノロジー情報センター(NCBI)は、データベース([www.ncbi.nlm.nih.gov/genomes/FLU/flu.html](http://www.ncbi.nlm.nih.gov/genomes/FLU/flu.html))を維持しており、本出願の出願時点で9796個のHA配列が含まれている。このデータベースを参照すると、当業者であれば、一般的にHAポリペプチドの特徴である配列、および/または特定のHAポリペプチド(例えば、H1、H2、H3、H4、H5、H6、H7、H8、H9、H10、H11、H12、H13、H14、H15、またはH16ポリペプチド)の特徴である配列;または特定の宿主、例えばトリ、ラクダ、イヌ、ネコ、ジャコウネコ、環境、ウマ、ヒト、ヒョウ、ミンク、マウス、アザラシ、イワツバメ、ブタ、トラ、クジラなどの感染を媒介するHAの特徴である配列を容易に特定できる。例えば、一部実施形態において、HAポリペプチドは、インフルエンザウイルスの天然分離株に見出されるHAタンパク質のおよそ残基97および185、324および340、96および100、ならびに/あるいは130~230間に見出される1つ以上の特徴配列を含む。一部実施形態において、HAポリペプチドは、本明細書で定義したように、少なくとも1つのHA配列エレメント1および2を含むアミノ酸配列を有する。一部実施形態において、HAポリペプチドは、一部実施形態において、約100~200、または約125~175、または約125~160、または約125~150、または129~139、または約129、130、131、132、133、134、135、136、137、138、もしくは139個のアミノ酸分だけ互いに隔てられているHA配列エレメント1および2を含むアミノ酸配列を有する。一部実施形態において、HAポリペプチドは、グリカン結合に参加する領域96~100および/または130~230内に位置する残基を含むアミノ酸配列を有する。例えば、多くのHAポリペプチドは、以下の残基: Tyr98、Ser/Thr136、Trp153、His183、およびLeu/Ile194を1つ以上含む。一部実施形態において、HAポリペプチドは、これらの残基の少なくとも2個、3個、4個、または5個全てを含む。

10

20

30

40

50

#### 【0021】

単離された:「単離された」という用語は、本明細書において、(i)(自然界であろうと、または実験的設定であろうとも)最初に産生された際に会合していた構成要素の少なくとも一部から分離された;または(ii)人為的に産生されたかのいずれかである薬剤または実体を指す。単離された薬剤または実体は、それらが最初に会合していた他の構成要素の少なくとも約10%、20%、30%、40%、50%、60%、70%、80%、90%、またはそれより多くから分離されていてよい。一部実施形態において、単離された薬剤は、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%を超えて純粋である。

#### 【0022】

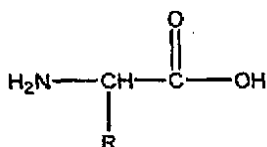
長鎖オリゴ糖:本開示の目的では、オリゴ糖は、少なくとも4つの単糖残基を有する少なくとも1つの直鎖を含む場合、典型的には「長鎖」と見なされる。

#### 【0023】

非天然アミノ酸:「非天然アミノ酸」という語句は、アミノ酸の化学構造(すなわち、

#### 【0024】

#### 【化1】



)を有し、したがって少なくとも2つのペプチド結合に参加できるが、自然界において見出されるものとは異なるR基を有する実体を指す。一部実施形態において、非天然アミノ酸は、水素以外の第2のR基を有してもよく、かつ/またはアミノ部分もしくはカルボン

酸部分に1つ以上の他の置換を有してもよい。

【0025】

ポリペプチド：「ポリペプチド」とは、一般的に言えば、ペプチド結合により互いに結合された少なくとも2つのアミノ酸の鎖である。一部実施形態において、ポリペプチドは、各々が少なくとも1つのペプチド結合により他と結合された少なくとも3～5個のアミノ酸を含んでいてよい。当業者であれば、ポリペプチドは、ともかくもポリペプチド鎖に組み込み可能な「非天然」アミノ酸または他の実体を場合により含むことが時にはあることを認識する。

【0026】

純粋である：本明細書において、薬剤または実体は、他の構成要素が実質的になければ「純粋である」。例えば、約90%を超えた特定の薬剤または実体を含有する調製物は、典型的には純粋な調製物であると見なされる。一部実施形態において、薬剤または実体は、少なくとも91%、92%、93%、94%、95%、96%、97%、98%、または99%純粋である。

【0027】

短鎖オリゴ糖：本開示の目的では、オリゴ糖は、任意の直鎖において4個未満の、またはもちろん3個未満の残基を有する場合、典型的には「短鎖」と見なされる。

【0028】

特異性：当該技術分野で公知のように、「特異性」とは、特定のリガンド（例えばHAポリペプチド）が、その結合パートナー（例えばヒトHA受容体、および特にヒト上気道HA受容体）を他の潜在的な結合パートナー（例えば、トリHA受容体）から区別する能力の尺度である。

【0029】

治療剤：本明細書において、「治療剤」という語句は、所望の生物学的または薬理学的効果を誘発する任意の薬剤を指す。

【0030】

治療：本明細書において、「治療」という用語は、疾病、障害、または状態の1つ以上の徴候もしくは態様の緩和、発症遅延、重症度もしくは発生率の低下、または予防をもたらすために使用される任意の方法を指す。本発明の目的では、治療は、徴候の発症前、発症中、および/または発症後に施すことができる。

【0031】

傘形トポロジー：「傘形トポロジー」という語句は、ある種のグリカンおよび特にHA受容体上のグリカンがとる3次元配置を指すために本明細書で使用する。本発明は、傘形トポロジーのグリカンに対する結合が、ヒト宿主の感染を媒介するHAタンパク質の特徴であるという認識を包含する。図6で例証したように、典型的には2-6シアル酸含有グリカンだけが傘形トポロジーをとり、長鎖（例えば、四糖を超えた）オリゴ糖に典型的である。傘形トポロジーの一例は、約-60のNeu5Ac-2-6Gal結合の角度により与えられる（例えば、図14を参照）。図9は、傘型トポロジーをとるグリカンのある代表的な（しかし網羅的ではない）例を表す。

【0032】

ワクチン接種：本明細書において、「ワクチン接種」という用語は、例えば疾病を引き起こす薬剤について免疫応答を生じさせるための組成物の投与を指す。本発明の目的では、ワクチン接種は、疾病を引き起こす薬剤への曝露前、曝露中、および/または曝露後に、ある実施形態において、薬剤への曝露前、曝露中、および/または曝露直後に投与できる。一部実施形態において、ワクチン接種は、適切に時間間隔をおいた、ワクチン接種組成物の多回投与を含む。

【0033】

変異体：本明細書において、「変異体」という用語は、目的とする特定のHAポリペプチドと、その配列が比較される「親」HAポリペプチドとの関係性を記述する相対的な用語である。目的とするHAポリペプチドが、特定の位置での少数の配列変更がなければ親

10

20

30

40

50

のアミノ酸配列と同一のアミノ酸配列を有する場合、目的とするHAポリペプチドは、親HAポリペプチドの「変異体」と見なされる。典型的には、親と比較して、20%、15%、10%、9%、8%、7%、6%、5%、4%、3%、2%未満の変異体の残基が置換される。一部実施形態において、変異体は、親と比較して10、9、8、7、6、5、4、3、2または1個の置換された残基を有する。変異体は、非常に少数（例えば、5、4、3、2、または1個未満）の置換された機能的残基（すなわち、特定の生物学的活性に参加する残基）を有することが多い。さらに変異体は、典型的には、親と比較して、5、4、3、2、または1つを超えない付加または欠失を有するが、付加または欠失を有しないことが多い。さらに、任意の付加または欠失は、典型的には、25、20、19、18、17、16、15、14、13、10、9、8、7、6個未満であり、一般には約5、4、3、または2個未満の残基である。一部実施形態において、親HAポリペプチドは、インフルエンザウイルスの天然分離株に見出されるポリペプチドである（例えば、野生型HA）。

10

#### 【0034】

ベクター：本明細書において、「ベクター」とは、それが連結されている別の核酸を輸送可能な核酸分子のことを指す。一部実施形態において、ベクターは、真核生物または原核細胞などの宿主細胞において、追加的な染色体複製および/またはそれが連結されている核酸の発現を可能にする。作用可能な連結された遺伝子の発現を指令できるベクターは、本明細書において「発現ベクター」と呼ばれる。

20

#### 【0035】

野生型：当該技術分野において理解されるように、「野生型」という語句は、一般的に、自然界に見出されるようなタンパク質または核酸の通常形態を指す。例えば、野生型HAポリペプチドは、インフルエンザウイルスの天然分離株に見出される。多様な異なる野生型HA配列は、NCBIインフルエンザウイルス配列データベース（<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>）に見出すことができる。

#### 【0036】

発明のある特定の実施形態の詳細な説明

グリカン G B P相互作用の定義および特徴付け

グリカン結合タンパク質（G B P）と呼ばれることが多いタンパク質の重要なファミリーは、種々の糖タンパク質のN結合およびO結合グリカンに結合し、免疫応答における細胞間接着、シグナリング、および輸送事象を媒介する。G B Pの主要な種類には、C型レクチン、ガレクチン、およびS I G L E Cが含まれる。G B Pは、典型的には、単量体または多量体の形態で、複数のグリカン結合部位を有する可溶性または膜結合性タンパク質のいずれかとして発現する。また、G B Pは、細胞表面上に分散、または微小環境に局所化され得る。

30

#### 【0037】

G B Pのグリカン結合部位は、糖質認識ドメイン（C R D）としても知られている。G B PのC R Dは、典型的には、単糖～四糖グリカンリガンドモチーフを収容する。単一C R Dおよびグリカンモチーフ間の相互作用は、典型的には、 $\mu$ M範囲の値を有する低親和性である。しかしながら、ほとんどの生理学的なグリカン G B P相互作用は多価性であり、グリカンモチーフの集団が、G B Pの会合により形成される多量体C R Dに結合することを伴う。したがって、タンパク質の機能を活性化するかまたは阻害するかのいずれかであるタンパク質間相互作用（デジタル制御）と異なり、グリカン G B P相互作用は、結合活性、段階的親和性、および多価性を介して、タンパク質の機能を微調整（アナログ調節）する。

40

#### 【0038】

生物学的機能に結びつく生化学的経路との関連で、グリカン タンパク質相互作用の構造機能相関を解釈することは、特有の課題を提示する。これらの課題の一態様は、その幾つかが追加的な組織特異的イソ型を有する複数のグリコシルトランスフェラーゼの協調的

50

な発現に伴うグリカンの非テンプレート性生合成による、グリカンの不均質性および化学的多様性から生じる。さらに、タンパク質上に複数ある糖鎖付加部位などのグリカンの生合成および細胞位置を考慮すると、グリカンは通常、細胞および組織から単離された際に、異なる化学構造の不均質な混合物と見なされるべきである。グリカン生合成が非テンプレート性であるという性質により、生体供給源由来の特定のグリカン構造物を増幅することも困難となる。

#### 【0039】

上記の課題を解決するために、数多くの進展があった。化学合成戦略に重要な進歩があり、細胞表面に存在するグリカンの多様性を捕捉する数百のグリカン構造物の合成に結びついた。これらの戦略を使用して、グリカンモチーフの様々な形態、すなわちクラスター、デンドロマー、ポリマーなどが、タンパク質上のグリカン結合部位の異なる型の多価性会合と一致するように構築される。これらの多価性複合糖質は、様々なGBPの相対結合親和性を評価する競合アッセイ、および生理学的なグリカン GBP 相互作用に対する阻害剤を設計するための競合アッセイにおいて、主として使用されている。これらの進展にもかかわらず、様々なGBPによる個々の生理学的グリカンモチーフの特異性または認識性、およびこれらの相互作用により調節される生物学的機能の選択性については、あまりよく知られていない。

10

#### 【0040】

公知である特異的なグリカン GBP 相互作用に関する現在の知識を急速に拡張するために、国際共同研究構想である Consortium for Functional Glycomics (CFG; [www.functionalglycomics.org](http://www.functionalglycomics.org)) は、新規グリカンリガンド特異性についてのGBPのハイスループットスクリーニングを可能にした幾つかのグリカン構造物を含むグリカンアレイを開発した。これらのグリカンアレイは、グリカンモチーフの多様性を増加させるために絶え間なく拡張されており、グリカンの生理学的多様性を最も良好に模倣する。CFGアレイ上のグリカンのほとんどは、化学的および化学酵素的合成に由来した。

20

#### 【0041】

また、CFGグリカンアレイは、一価性および多価性のグリカンモチーフ（すなわち、ポリアクリルアミド主鎖に結合した）の両方を含み、糖質生物学者が目的のGBPの新しいグリカンリガンドを発見するために広く使用されている資源として出現しつつある。グリカンアレイデータに加えて、CFGは、グリカン生合成酵素およびGBPの遺伝子発現から全生物グリコムおよびフェノーム解析に及ぶ多様なデータセットを生成するための、最先端の資源に進歩しつつある。

30

#### 【0042】

使い易いインターフェースを介するCFGデータセットの公開普及は、これらの複合データセットの解析により興味深いパターンを見出すための、または有意義な予測をするためのデータマイニングツールの開発に意欲を与え始めた。データマイニングツールは、ゲノミクスおよびプロテオミクスの領域で一般的になりつつある。多数の構成要素（遺伝子およびタンパク質）および生化学経路を代表する複雑なネットワークにおけるそれらの相互作用を扱うハイスループットデータは、統計的に有意な相関および予測を行うために解析される。グリコミクスの場合には、グリカン GBP 相互作用のアナログ的な性質を考慮すると、特定のGBPに対する結合を決定する構造集団の共通特徴を理解するためには、単一グリカンと単一GBPの相互作用の先を行くことが必要である。

40

#### 【0043】

ハイスループットグリコミクスデータ解析用のデータマイニングツール構築への第一歩として、発明者らは、本研究において、規則帰納法に基づくデータマイニング方法論を用いて、CFGグリカンアレイデータを解析するための新規手法を使用した。柔軟なソフトウェアアーキテクチャおよびCFGのリレーショナルデータベースを駆使して、発明者らは自らの手法を使用して、グリカンの集団が特定のGBPに結合する能力を決定するパターンを特定した。3つの異なるファミリーのGBPの具体例：(1) DC SIGN およ

50

びSIGNR；(2)ガラクチン；および(3)赤血球凝集素を使用して、発明者らは、これらのタンパク質との相互作用を決定する、アレイ上のグリカンの特異的なパターンを特定する。発明者らは、結晶構造を使用することにより特定され、かつグリカンアレイに見られないGBPとグリカンとの結合レベルを予測することにより特定されたパターンを確認する。これらのパターンは、グリカン構造物の集団（共通した一連の特徴を含む）と所定のGBPとの相互作用についての理解を初めて可能にし、それによりグリカン GBP相互作用の構造機能相関の解析および定義が可能になる。

#### 【0044】

したがって、本発明は、グリカン GBP相互作用の構造機能相関を理解するためのシステムを提供する。特に本発明は、グリカン構造物の集団とGBPの多価性CRDとの相互作用が、どのようにして基本的生物学的プロセスを調節するのかを理解するためのシステムを提供する。本発明は、所定の相互作用の特異性を決定する、グリカンまたはそれらの結合パートナーの特徴を特定する。本発明は、例えば解析情報（例えば、X線結晶解析、NMRなどから）に基づく特徴により提供される制約も定義した。そのような制約は単独で使用でき、または場合により機能情報もしくは他の情報と結び付けることができる。適切な機能情報は、例えば、グリカン結合の研究から取得できる。

10

#### 【0045】

本発明は、様々なGBPに対する新規グリカンリガンド候補を特定するためにますます利用されているCFGにより開発されたものなどのグリカンアレイから得られたデータセットを解析するための計算方法を提供する。これらのグリカンアレイが拡張し続けると共に、これらのグリカンアレイから得られたデータセットを解析し、グリカン GBP相互作用における特異性の基本を理解するためのそのような計算方法の価値は増加する一方である。

20

#### 【0046】

例えば、グリカンアレイデータ全体（高親和性、中親和性、低親和性、および非結合を含む）を解析するための規則に基づくデータマイニング方法論を使用して、本発明は、GBPに対する結合に正および負の効果を有するグリカンのパターンを特定するための新規手法を提供する。そのような規則に基づく手法の1つの利点は、これらの規則を満たす他の潜在的なグリカンを特定するために容易に適用できる一連の単純明快な規則として、最終パターンを提供することである。

30

#### 【0047】

本明細書において、本発明の原理は3つの多様なGBPファミリーに適用され、それらの有効性の原理証明が確立された。最初の例、すなわちDC SIGNおよびSIGNRシステムにおいて、規則は、DC SIGNについては3つの広範な特徴、すなわち高マンノース、ルイスx (Lewis x) [Gal b 4 (Fuc a 3) Glc NA c]、およびFuc a 4 Glc NA cを含有するモチーフを示し、DC SIGNRについては高マンノース特徴のみを示した。高親和性結合を決定する共通特徴の捕捉に加えて、規則は、ルイスx含有モチーフについてはGalに任意の3 O置換も存在しないことなど、結合に有害であった特徴も捕捉した。これらの負の結果はDC SIGNRの結晶構造の解析と一致し、したがって発明者らの手法の価値が強調される。

40

#### 【0048】

ガレクチンの場合には、規則はより複雑であった。ガレクチン1およびガレクチン3による高親和性リガンド結合に必要な主要な特徴 (Gal b 4 Glc NA c) の特定に加えて、発明者らは、鎖長に関連して、グリカンリガンドとの相互作用決定におけるこの単位への置換の役割も決定した。DC - SIGNの例と同様に、発明者らの知見は、ガレクチンの結晶構造の解析と一致した。上記の特徴に基づいて、ガレクチン1およびガレクチン3のグリカン結合間の主要な相異は、ガレクチン3が、N結合グリカンの様々な分岐部に存在するこれらの単位ではなく、Gal b 4 Glc NA cの直鎖反復単位を好んだということだった。ガレクチン1は、典型的には非共有結合的に会合されたCRDsを有するホモ二量体として生じるため、異なる分岐部上のGal b 4 Glc NA cの存在は、高

50

親和性の多価性結合を強化する。その一方でガレクチン3は、N末端リンカー領域を有する単量体で、これらの単位の分岐出現と比較して、ラクトサミン単位の直鎖反復への優先性を有する可能性が最も高い。

#### 【0049】

発明者らの規則に基づく帰納法的手法の全体的な精度は、上記規則が80%の高結合体(DC S I G Nの場合)から100%の高結合体(ガレクチン3およびDC S I G N R)を正確に特定し、全ての場合で偽陽性がなかったことを考慮すると、良好である。グリカンアレイは多様な一連のグリカンを含むが、まだグリカンの全体的な多様性を系統的に捕捉していない。その結果、スクリーニングデータに単集合データポイントがあり、すなわち、共通した一連の特徴により定義された任意の特定の群に当てはまらない高親和性グリカン構造物がある。そのような単集合データポイントは、発明者らの予測結果では偽陰性に結びつく。各規則は、高親和性結合を示す一連のグリカンにより共有されている主要なグリカンモチーフを含むことが、表2および表3から認められるべきである。さらに、主要なモチーフは、他のモチーフまたは鎖長の必要条件の非存在などの他の制約と併せて特定される。

10

#### 【0050】

データマイニングの全体的プロセスの一部として、これらの主要なパターンに基づく追加的な特徴が定義でき、グリカン結合におけるこれらの特徴の役割をさらに研究できる。例えば、還元末端または非還元末端からの距離、および直鎖の一部として出現するのかわたは分岐鎖の一部としての出現するのかわという点におけるGal b 4 Gl c N A cの位置は、結合におけるそれらの影響を評価するための追加的な特徴として定義できる。また、Gal N A c、Gal [ 3 O S O <sub>3</sub> ]、Gal [ 6 O S O <sub>3</sub> ]などの各単糖に対する修飾を全て組み合わせる追加的なグリカン特徴は、結合に対するこれらの修飾の各々の重要性を評価するために、単一の特徴に統合できる。

20

#### 【0051】

要約すると、CFGグリカンアレイデータをモデルシステムとして使用して、発明者らは、有意義な解釈を容易にする、複雑なデータセットにおける規則またはパターンを特定するための手法を概説した。数多くの大規模グリコミクス構想は、グリカン生合成酵素、GBPの遺伝子発現から、特定の細胞型および様々な供給源から単離された組織に由来するグリカンのレパートリーの特定に及ぶ多様なデータセットを取得するために、自らの資源を振り向けている。これらのデータセットが拡張する場合、本明細書で概説した規則に基づく帰納法的方法を使用して、グリカン GBP相互作用および生物学的機能に対する遺伝子発現を決定するパターンの組合せが取得できる。

30

#### 【0052】

##### 応用

本発明は、グリカン GBP結合相互作用の詳細な特徴付けを可能にする。したがって、本発明は、所定のGBPと相互作用する(またはしない)一連のグリカンの定義を提供する。したがって、本発明は、GBP特異的グリカンアレイ、すなわち特定のGBPの存在または同一性を確立または定義するのに十分な一連のグリカンを含むアレイの調製を可能にする。

40

#### 【0053】

例えば、本明細書で提示したように特定のGBPのグリカン結合特徴が定義されれば、結合したグリカン、結合しないグリカン、および/またはその組合せを含有するアレイが組立てられ、例えば、試料中のその特定のGBPの検出、および/またはそのGBPの誘導体の特徴付けに使用できる。

#### 【0054】

特定の一例を挙げると、GBPの1つは、赤血球凝集素(HA)H5タンパク質であり、その結合解析は下記で例証される。一般的に言えば、HAは、糖タンパク質受容体に結合することにより細胞表面と相互作用する。HA受容体に対するHAの結合は、HA受容体上のN結合グリカンが支配的に媒介する。具体的には、インフルエンザウイルス粒子表

50



面上の H A は、細胞宿主表面上の H A 受容体と会合しているシアル酸含有グリカンを経験する。認識および結合の後、宿主細胞はウイルス細胞を貪食し、ウイルスは自己複製してさらに数多くのウイルス粒子を産生でき、近隣の細胞に分散する。

【 0 0 5 5 】

H A 受容体は、受容体の H A 結合部位付近で 2 3 または 2 6 シアル酸含有グリカンのいずれかにより修飾され、受容体に結合したグリカンの結合型は、受容体の H A 結合部位のコンフォメーションに影響を与え、したがって様々な H A 亜型に対する受容体の特異性に影響を及ぼす。さらに、本発明者は、結合されたグリカンのトポロジー（傘形様または円錐形様）が、様々な H A に対する受容体の特異性に影響を及ぼすことを特定した。

10

【 0 0 5 6 】

例えば、トリ H A のグリカン結合ポケットは狭い。本発明によれば、このポケットは、2 3 シアル酸含有グリカンのトランス型コンフォメーション、および / または 2 3 結合かまたは 2 6 結合かに関わらず、円錐形トポロジーのグリカンに結合する。

【 0 0 5 7 】

トリ組織、ならびにさらにヒトの肺深部組織および胃腸管（ G I ）組織の H A 受容体は、2 3 シアル酸含有グリカン結合により特徴付けられ、さらに（本発明により）、支配的に円錐形トポロジーをとる 2 3 シアル酸含有グリカンおよび / または 2 6 シアル酸含有グリカンを含むグリカンにより特徴付けられる。

20

【 0 0 5 8 】

対照的に、上気道の気管支および気管のヒト H A 受容体は、2 6 シアル酸含有グリカンにより修飾されている。2 3 モチーフと異なり、2 6 モチーフは、C 6 C 5 結合によるコンフォメーションの追加的な自由度を有する（ R u s s e l l e t a l , G l y c o c o n j J 2 3 : 8 5 , 2 0 0 6 ）。そのような 2 6 シアル酸含有グリカンに結合する H A は、このコンフォメーションの自由度から生じる構造の多様性を収容するために、より開放的な結合ポケットを有する。さらに本発明によると、H A は、ヒト上気道組織の感染を効果的に媒介するためには、傘形トポロジーのグリカン（例えば、2 6 シアル酸含有グリカン）に結合する必要がある得、強い親和性および / または特異性でそのような傘形トポロジーのグリカンに結合する必要がある得。

30

【 0 0 5 9 】

これらの空間に制限されたグリコシル化特性の結果、ヒトは、数多くの野生型トリ H A （例えば、トリ H 5 ）を含有するウイルスにより通常は感染しない。具体的には、ウイルスに遭遇する可能性が最も高いヒトの気道部分（すなわち、気管と気管支）が、円錐形グリカン（例えば、2 3 シアル酸含有グリカン、および / または短鎖グリカン）に対する受容体を欠如し、野生型トリ H A は、円錐形グリカン（例えば、2 3 シアル酸含有グリカン、および / または短鎖グリカン）と会合した受容体に、主としてまたは排他的に結合するため、ヒトがトリウイルスに感染するのは稀有になる。ウイルスと十分に密接な接触をして、ウイルスが、傘形グリカン（例えば、長鎖 2 6 シアル酸含有グリカン）を有する肺深部および / または胃腸管の受容体に接近できる場合に限り、ヒトは感染する。

40

【 0 0 6 0 】

本明細書において、本発明は、H 5 型 H A タンパク質の検出、および / または変化した結合特異性で出現する可能性のあるタンパク質の変異体を検出するために使用できる一連のグリカンの特定を可能にする。特に、そのような発明のアレイは、上気道のヒト H A 受容体に結合する能力および / または傘形トポロジーのグリカンに結合する能力（場合によりは高親和性および / または特異性で、好ましくは高親和性で）を有する、任意の H 5 変異体または実際に任意の H A タンパク質もしくはその変異体を検出するために使用できる。

【 0 0 6 1 】

本明細書で実証したように、そのようなアレイは、様々な H A タンパク質およびそれら

50

のグリカン結合特徴の特定および／または特徴付けのために有用である。ある実施形態において、発明のH5型HA変異体タンパク質をそのようなアレイ上で検査して、それらが傘形トポロジー（例えば、2 6グリカン、および特に長鎖 2 6グリカン）に結合する能力を評価し、そのような複数のグリカンに結合する能力を特に評価する。

#### 【0062】

実際、本発明は、HA結合能力の特徴付けおよび／または例えばヒト結合性HAを検出するための特徴として使用できる、傘形グリカン（例えば、2 6グリカン、および特に長鎖 2 6グリカン）および場合によりは円錐形トポロジーのグリカン（例えば、2 3シアル酸含有グリカン）のアレイを提供する。当業者には明白であるが、そのようなアレイは、H5型HAを特徴付けるまたは検出するためだけでなく、実際は、2 6グリカンに結合する能力が望ましく評価されるH7および／またはH9を例えば含む任意のHAを検出するためにも有用である。

10

#### 【実施例】

#### 【0063】

（実施例1）データマイニング方法論

グリカンアレイの説明およびグリカンアレイデータ供給源

CFGは、2種類のグリカンアレイ：（1）ウェルに基づくマイクロアレイおよび（2）固相に印刷したアレイを開発した。印刷されたアレイはより最近になって開発されたため、初期のリガンドスクリーニングのほとんどはウェルに基づくマイクロアレイを使用し行なわれた。CFGにより開発されたウェルに基づくマイクロアレイの最初の型は、各グリカンの3重反復提示で約60個の異なるグリカンを含んでいた。各後続型のアレイは追加的なグリカンを組み込み、現行型は、各グリカンの4重反復提示で195個のグリカンを含む（<http://www.functionalglycomics.org/static/consortium/resources/resourcecoreh5.shtml>を参照）。アレイは、N結合およびO結合グリカンの生理学的多様性を捕捉する合成グリカンを支配的に含む。アレイは、ポリアクリルアミド骨格に結合した多価性グリカンリガンドも含む。合成グリカンに加えて、様々な哺乳類糖タンパク質に由来するN結合グリカン混合物も、アレイ上に提示されている。

20

#### 【0064】

本研究の解析用に選ばれたデータセットは、CFGウェブサイト：<http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp>から取得した。現在、40個の哺乳類GBPが、様々な型のグリカンアレイに対してスクリーニングされている。スクリーニングデータは、アレイ上の各グリカンリガンドについてのGBPの平均信号および信号対雑音比だけでなく、所定のウェルの所定のGBPの強度信号を含む生の書式でも両方で利用可能である。グリカンアレイが現行型へと進化したため、旧型のアレイを使用してスクリーニングされたGBPが、最新型を使用して再スクリーニングされることは一般的になったことを指摘しておくことは重要である。これらのデータポイントの非存在は、あるGBPのグリカンリガンドの結合を別のGBPのグリカンリガンドの結合と区別した特徴の特定に関する含意を有している（本明細書で考察したように）。DC SIGN、S I G N R、ヒトガレクチン1およびガレクチン3（およびその個々の糖質認識ドメイン）、ならびに赤血球凝集素H5のスクリーニングに対応するデータセットが取得された。これらのデータセットは、下記に記載したデータマイニングプラットフォームを使用して解析した。

30

40

#### 【0065】

データマイニングプラットフォーム

データマイニングプロセスに含まれる主要な工程を図1に例証する。これらの工程は、データオブジェクト、特徴、および分類指標という3つのエレメントに対する操作を含む。「データオブジェクト」とは、データベースに保存されている生データである。グリカンアレイデータの場合において、単糖および結合に関するグリカン構造物の化学的記述な

50

らびにスクリーニングされた様々なGBPとの結合信号が、データオブジェクトを構成する。データオブジェクトの重要な特性が「特徴」である。データオブジェクトを記述する特徴の選択は、規則またはパターンの取得を可能にする。「分類指標」は、データオブジェクトを特定の種類にクラスター化させるか、または特徴間の関係性を決定するかいずれかのために使用される規則またはパターンである。発明者らの下記の実施例で考察したように、分類指標は、高親和性でGBPに結合するグリカンにより満たされる特異的な特徴を提供する。これらの規則は：(1)結合を強化すると考えられる一連の高親和性グリカンリガンドに存在する特徴、および(2)結合に好ましくないと考えられる高親和性グリカンリガンドに存在すべきでない特徴の2種類である。

【0066】

データマイニングプラットフォームは、上述の操作を実施するために互いに相互作用するソフトウェアモジュール(図1)を含む。1つの構成要素は、特徴を抽出するためにCFGデータベースに接続する特徴抽出部である。CFGにより使用される、オブジェクトに基づくリレーショナルデータベースは、特徴の柔軟な定義を容易にする。

【0067】

特徴抽出およびデータ調製：

上記のように、特徴は、グリカンおよび/またはそれらの結合パートナーから抽出できる。本明細書で例証された特定の応用において、表1に列挙したように、ある種の特徴がグリカンアレイ上のグリカンから抽出された。

【0068】

表1．グリカンアレイ上のグリカンから抽出された特徴

この表に記載した特徴は、規則に基づく分類アルゴリズムによって使用され、特定のGBPに対する結合を特徴付けたパターンを特定した。

【0069】

10

20

【表 1】

抽出された特徴	特徴の説明	
単糖レベル		
組成	hex、hexNAc（複数）、dHex、シアル酸などの数 [図 1 では、組成物は Hex=5；HexNAc=4 である]。末端の組成は明確に記録される [図 1 では、末端の組成は Hex=2；HexNAc=2 である]。	
明示的な組成	Glc、Gal、GlcNAc、Fuc、GalNAc、Neu5Ac、Neu5Gc などの数 [図 1 では、明示的な組成は、Man=5；GlcNAc=4 である]。末端の明示的な組成は明示的に記録される [図 1 では、末端の明示的な組成は、Man=2；GlcNAc=2 である]。	10
より高次の特徴		
ペア	ペアとは、結合により共有結合で接続している 1 対の単糖を指す。ペアは、非還元末端で終了する単糖を 1 つ有するペアと区別するために、通常 [B] および末端 [T] の 2 つの範囲に分類される (図 2)。ペアの頻度は特徴として抽出された。	
トリプレット	トリプレットは、2 つの連結により共有結合で接続している一連の 3 つの単糖を指す。発明者らは、3 つの範囲、すなわち通常 [B]、末端 [T]、表面 [S] にそれらを分類する [図 2]。各範囲のトリプレットの組成物は特徴として抽出された。	20
クワドルプレット	トリプレットの特徴と同様に、クワドルプレットの特徴も、4 つの単糖およびそれらの結合 [図 2] と共に抽出される。クワドルプレットは、通常 [B] および表面 [S] の 2 つの種類に分類される。異なるクワドルプレットの頻度は特徴として抽出された。	
クラスター	上記の表面トリプレットおよびクワドルプレットの場合、結合情報を無視すれば、発明者らは一連の単糖クラスターを取得し、それらの出現頻度 (組成物) が作表される。これらの特徴は、単糖間の結合型の重要性を解析するために選ばれた。	30
平均リーフ深度 (Average Leaf Depth)	ブローブの効果的な長さの指標として、ツリーの還元末端の平均深度がグリカンの特徴として抽出される。図 2 B では、リーフ深度は 3、4、および 3 であり、平均は 3.34 である。	
リーフ数 (Number of Leaves)	グリカンツリーの広がり の尺度として、非還元単糖の数が特徴として抽出される。図 2 B の場合、リーフ数は 3 である。図 1 の場合は、リーフ数は 4 である。	
GBP 結合特徴	これらの特徴は、アレイを使用してスクリーニングされる全ての GBP について取得される。	
グリカン毎の平均信号	生の信号値は、同一グリカンの 3 重反復または 4 重反復 [アレイの型に依存する] を平均した。	40
信号対雑音比	陰性対照に基づいて平均雑音を計算し [CFG により開発された標準化法]、信号対雑音比 [S/N] が算出された。	

示された特徴を選ぶことの背後にある論理的根拠は、GBP のグリカン結合部位が典型的には二糖～四糖を収容するという事実だった。ツリーに基づく表示は、グリカン構造物における単糖および結合に関する情報を捕捉するために使用された (還元末端がツリーの根)。この表示は、接続する一連の単糖トリプレットなど、より高次の特徴を含む種々の特徴の抽出を容易にした (図 2)。データ調製は、各グリカンについて抽出した特徴 (表 1) と共に、最新型のグリカンアレイの全グリカンを列方向に列挙したリストの生成を含む。特定の GBP または一連の GBP に対する結合を決定する特定のパターンを決定する

ため、グリカンおよびそれらの特徴に関するこの基本表から、規則に基づく分類（下記を参照）のためのサブセットを選んだ。

#### 【0070】

分類指標：

様々な種類の分類指標が開発され、多くの応用で使用されてきた。それらは、主として3つの主要な範疇：数学法、間隔法、および論理法に分類される。これらの異なる方法ならびにそれらの利点および欠点は、Weiss & Indruchya (Predictive data mining - A practical guide. Morgan Kaufmann, San Francisco, 1998) で詳細に考察されている。この特定の応用の場合、発明者らは、論理的方法に該当する規則帰納法と呼ばれる方法を選ぶ。規則帰納法の分類指標は、IF - THEN 規則の形態のパターンを生成する。

10

#### 【0071】

IF - THEN (もし～ならば、～である) 規則を生成する規則帰納法などの論理的方法および特に分類指標の主要な利点の1つは、他の統計的または数学的方法と比較した際、分類指標の結果がより容易に説明できることである。これにより、発見された規則またはパターンの構造的および生物学的意義を探索することが可能になる。前述した特徴（表1を参照）を使用して生成される規則の例は、IF グリカンが「Gal b 4 Gl c NA c b 3 Gal [B]」を含有し、かつ「Fuc a 3 Gl c NA c [B]」を含有しないならば、THEN グリカンはより高い親和性でガレクチン3に結合しよう、というものである。

20

#### 【0072】

この場合に使用された特定の規則帰納アルゴリズムは、Weiss & Indruchya (Predictive data mining - A practical guide. Morgan Kaufmann, San Francisco, 1998) により開発されたアルゴリズムである。

#### 【0073】

結合レベル

グリカンアレイスクリーニングデータセットの各々について、低親和性結合と高親和性結合とを区別する閾値を定義した（図3）。

#### 【0074】

ハイスループットCFGグリカンアレイデータにデータマイニング法を適用することにより、発明者らは、様々なGBPに結合するグリカンの一連の特徴を特定した。3つの特定のシステム：(1) DC SIGNおよびSIGNR、(2) ガレクチン；ならびに(3) 赤血球凝集素H5を例として選んだ。これらのGBPファミリーの各々は、グリカンリガンドの優先性の点で合理的に十分定義される。最近の研究により、DC SIGNおよびSIGNRの特徴的なリガンド特異性の構造的基盤がグリカンアレイデータに基づいて概説されたため、第1の例は、発明者らの方法論の追加的な実証を提供する。過去の研究により、様々なガレクチンのリガンド特異性が系統的に評価されている。しかしながら、CFGグリカンアレイは、様々なガレクチンのリガンド特異性をスクリーニングするために使用されてきた、はるかに大きなドメインのグリカン構造物を提供する。したがって、発明者らの方法論をガレクチンデータセットに応用すると、様々なガレクチンのグリカンリガンドに対する結合を決定する追加的な規則が提供される。

30

40

#### 【0075】

(実施例2) DC SIGNおよびDC SIGNRへの方法論の応用

DC SIGNおよびDC SIGNRは、 $Ca^{2+}$  に依存する様式でグリカンリガンドを認識し結合する、C型レクチンのII型膜貫通受容体サブファミリーに属する。DC SIGNは、樹状細胞において豊富に発現され、ICAM 3分子を介して樹状細胞を提示する抗原へのT細胞の接着に重要な役割を果たし、それによって免疫応答を惹起する。加えて、DC SIGNは、樹状細胞によるHIVなどの病原体の認識に重要な役割を果たすことも示されている。実際、樹状細胞のDC SIGNに対するHIVの結合は、

50

T細胞の感染を強化することが実証されている。他方では、DC SIGNと77%の配列同一性を共有するDC-SIGNRは、肝臓、リンパ節、および胎盤の内皮細胞に見出される。

【0076】

これらのタンパク質の各々は、C末端に単一の糖質認識ドメイン(CRD)を含有する。両タンパク質の細胞外アルファヘリックスドメイン(CRDに隣接している)は、CRDの四量体化を促進し、したがってグリカンリガンドとの多価性相互作用を可能にする。DC SIGNおよびSIGNRについては、様々なグリカンリガンドとの結晶構造を含む豊富な結晶構造情報がある。より最近では、これらのタンパク質は、CFGグリカンアレイを使用してスクリーニングされ、特徴的なリガンド特異性および信号伝達特性を有することが実証された。したがって、これらのタンパク質のグリカンアレイデータは、データマイニング方法論を実証にするための良好な枠組みを提供した。

10

【0077】

上記の実施例1で概説したように、DC SIGNおよびDC SIGNRのグリカンスクリーニング解析に対応するグリカン特徴(表1)は、CFGデータベースから要約した。主要な目的関数が、2つのタンパク質の各々に対する各グリカンの結合の平均信号対雑音比であったこれらの特徴を使用して、規則に基づく分類法を実施した。前記分類法による結果を表2に集約する。

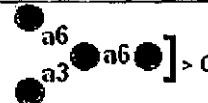
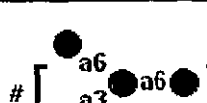

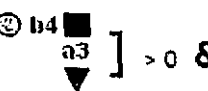









【0078】

表2. DC SIGNおよびDC SIGNRに対するグリカン結合を決定する規則特徴[表1]に基づいて規則が導出され、#[ ]は出現数を特定するために使用される。DC SIGN結合を決定する最後の規則は、個々の規則の各々が満たされなければならないことを示唆する、複数の特徴のより複雑な「&」を含む組合せである。DC SIGNについては、規則にクラスター化できなかったグリカン[偽陰性]も示す。

20

【0079】

【表 2】

DC - SIGN		DC-SIGNR		
グリカン特徴	真の陽性	グリカン特徴	真の陽性	
#  > 0	6	#  > 0	6	
#  > 0	4			
#  > 0 & !  & ! 	6			
概要 : 閾値 S / N 比 = 2.3 高結合体の総数 = 20 真の陽性 = 16 偽陽性 = 0 偽陰性 = 4 真の陰性 = 124		概要 : 閾値 S / N 比 = 3.43 高結合体の総数 = 6 真の陽性 = 6 偽陽性 = 0 偽陰性 = 0 真の陰性 = 42		
偽陰性 : 1.   2.   3.  4.  5. 				

規則に基づく分類法の全体的な成績は、DC SIGNRについては高マンノース構造候補の100%を予測し、DC SIGNについては20個の高結合体のうち16個を予測したことを考慮すると、良好であった。偽陽性がなかったこと、つまり結合すると予測されたが結合しなかったグリカンの実例がなかったこと留意することが重要である。結果を見て最初に明白になる含意は、DC SIGNおよびDC SIGNRの両方が、高マンノース構造に対する共通の高親和性結合を共有することである。Mana3 (Mana6) Mana6 Manの存在は、グリカンアレイデータに基づき、高親和性で結合する6つの異なる高マンノースグリカンリガンドを捕捉する強力な規則である。この観察は、過去の結晶構造研究と一致する。

【0080】

DC SIGNは、高マンノースリガンドに加えて、明確な特徴により特徴付けられた追加的な一連のフコシル化リガンドに結合した。これらのフコシル化リガンドは、DC SIGNRには結合しなかった。Fuca4GlcNAcは、ルイス<sup>a</sup> [Fuca4 (Galb3) GlcNAc] を含有するグリカン構造物に一般に観察されるモチーフである。Fuca3 (Galb4) GlcNAcは、N結合およびO結合グリカンの非還元末端に存在する、別の一般的に観察されるルイス<sup>x</sup>モチーフである。これらの両特徴は、DC SIGNに対する高親和性結合体に特徴的だった。この観察は、ルイス<sup>x</sup>を含有するグ

リカン構造物を伴ったDC SIGNの初期の結晶構造において観察された、フコシル化リガンドに対するDC SIGNの特徴的な結合と一致する。高マンノースおよびフコシル化リガンドを伴ったDC SIGNおよびSIGNRの結晶構造の詳細な研究に基づき、これらの両タンパク質が、高マンノースリガンドに対する同様の結合様式を共有する一方で、フコシル化リガンドに対する結合は完全に異なり、DC SIGNのCRDにあるアミノ酸によってのみ達成され得ることが示された。

#### 【0081】

発明者らの解析により提供された別の興味深い観察は、特定の特徴の非存在が、高親和性で結合するために必要とされることである。つまり、ルイス<sup>x</sup>モチーフと共にNeuA3Galb4GlcNAcおよびGala3Galが存在すると、DC SIGNに対するこれらのリガンドの結合に有害である。発明者らのデータマイニング手法の価値は、ルイス<sup>x</sup>を含有するグリカンリガンドを伴ったDC SIGNの結晶構造の研究によりこの規則が確認されることによって強調される。FucA3(Galb4)GlcNAcのGalにおける3-OHの位置は、DC SIGNのCRDに近接しており(図4)、硫酸化、シアル酸付加などを含むこの位置への任意のかさ高い置換は、タンパク質との好ましくない立体接触に結びつき、したがって結合を破壊する。

10

#### 【0082】

DC SIGNおよびSIGNRを伴ったルイス<sup>x</sup>含有グリカンの結晶構造に基づく、フコシル化リガンドの主要な結合には、Ca<sup>2+</sup>イオンとの配位を形成するFucのエクアトリアル酸素3-OHおよび4-OHが関与する。したがって、ルイス<sup>a</sup>抗原FucA4(Galb3)GlcNAcの場合でさえ、主要な結合には、Fucの3-OH、および4-OHが関与する(図4)。興味深いことには、このモチーフを含む規則である#[FucA4GlcNAc]>0は、Galb3GlcNAc結合の存在を明示的に含まなかった。したがって、この解析は、Galが存在すると、ルイス<sup>a</sup>含有モチーフを含有するものよりもルイス<sup>x</sup>含有モチーフの場合において、フコシル化リガンドに対するDC SIGNの結合により良好な影響を及ぼすことを示す。

20

#### 【0083】

##### (実施例3) ガレクチンへの方法論の応用

ガレクチンは、活性に還元性チオールを必要とするため以前はS型レクチンと定義されていたガラクトシドに結合することが知られている可溶性GBPのファミリーに属する。C型レクチン(DC SIGNおよびSIGNRなど)と異なり、ガレクチンはリガンド結合にCa<sup>2+</sup>を必要としない。ガレクチンは、多数の生物学的役割、すなわち細胞発生、アポトーシス、癌、および免疫応答に関わるとされている。ガレクチンは、一般的にI型(Galb3GlcNAc)およびII型(Galb4GlcNAc)ラクトサミン単位に結合することが知られているが、多数の生物学的役割についてのより細かい基質特異性およびその含意はあまり理解されていない。ヒトガレクチン1および3のデータセットを、規則に基づくデータマイニング手法を使用して解析した。これらの2つのガレクチンは、CRDの組織の点で基本的に異なる。ガレクチン1および3の両方は、類似したC末端F3型CRDを共有する。ガレクチン1は典型的にはCRDのホモ二量体であるが、ガレクチン3はN末端リンカードメインを有する単一CRDから構成される。ガレクチン3のN末端ドメインは、グリカンリガンドに対する親和性の強化に関わるとされている。

30

40

#### 【0084】

上記の例と同様に、グリカンアレイ上のグリカンリガンドの、ガレクチン1および3に対する結合を強化および減退させる特徴を、規則に基づくデータマイニング手法を使用して同定した(表3)。

#### 【0085】

##### 表3. ガレクチン1および3に対するグリカン結合を決定する規則

特徴[表1]に基づいて規則を導出し、#[ ]は出現数を特定するために使用され、&は各々が満たされなければならない規則の組合せを意味する。![ ]はパターンの非存在

50



を意味する。a 2 / 3 および a 3 / 6 における / は、それぞれ a 2 または a 3、および a 3 または a 6 を表わす。

【 0 0 8 6 】

【 表 3 】

ガレクチン - 1		
グリカン特徴	真の陽性	要約
$\#[\bigcirc b4 \blacksquare] > 0 \ \& \ ![\blacksquare b6 \square] \ \& \$ $![\blacktriangleleft a2/3 \blacksquare] \ \& \ ![\blacktriangleleft a2 \bigcirc] \ \& \$ $![\blacklozenge a3/6 \bigcirc] \ \& \ ![\blacklozenge a3/6 \bigcirc] \ \& \$ $![\bigcirc a3/4 \bigcirc] \ \& \ ![\blacklozenge a3 \bigcirc] \ \& \$ $\& \ [ \text{鎖長} > \text{ラクトサミンを含む 2 単位}]$	8	閾値 S / N 比 = 1.8 高結合体の総数 = 10 真の陽性 = 9 偽陽性 = 0 偽陰性 = 1 真の陰性 = 131 偽陰性 : $\bigcirc b3 \square$ $3-O-SO_3$
$\#[\bigcirc b4 \blacksquare] > 1 \ \& \ [\blacklozenge a3 \bigcirc] > 0$	1	
ガレクチン - 3		
グリカン特徴	真の陽性	要約
$\#[\bigcirc b4 \blacksquare] > 0 \ \& \ ![\blacksquare b6 \square] \ \& \$ $![\blacktriangleleft a2/3 \blacksquare] \ \& \ ![\bigcirc a4 \bigcirc] \ \& \$ $![\blacklozenge a3/6 \bigcirc] \ \& \ ![\blacklozenge a3/6 \bigcirc] \ \& \$ $\& \ ![\blacklozenge a3 \bigcirc] \ \& \ ![\text{a6} \bullet \text{a3} \bullet]$ $\& \ [ \text{ラクトサミン鎖の鎖長} > 2]$	6	閾値 S / N 比 = 2.7 高結合体の総数 = 8 真の陽性 = 8 偽陽性 = 0 偽陰性 = 0 真の陰性 = 139
$\#[\bigcirc b4 \blacksquare] > 2 \ \& \ [\blacklozenge a3 \bigcirc] > 0$	1	
$\#[\bigcirc b4 \blacksquare] > 1 \ \& \ [\blacklozenge a6 \bigcirc] > 0$	1	

ガレクチン 1 および 3 の高親和性リガンド結合を決定する規則は、D C S I G N および S I G N R について導出されたものより複雑である。ガレクチン 1 および 3 は、I I 型および I 型の両ラクトサミン単位に同様の親和性で結合することが知られているが、グリカンアレイからのデータでは、高結合体を区別するために使用された閾値強度に基づいて、任意の I 型 ( G a l b 3 G l c N A c ) 結合体も明らかにならなかった。

【 0 0 8 7 】

ガレクチン 1 の場合、9 つの高結合体のうち 8 つを捕捉した第 1 の規則 ( 表 3 ) は、鎖長が少なくとも単糖 3 つである少なくとも 1 つのラクトサミン単位の存在を含んでいた。かさねて、偽陽性がなかったことに留意することが重要である。低親和性および高親和性結合体の解析に基づき、規則 1 の幾つかのパターンは、結合に負の効果をも有すると意味づけられた。G l c N A c のフコシル化、G a l の末端フコシル化、G a l のシアル酸付加

、およびⅡⅡ型ラクトサミン単位と併せてGal a 3 GalまたはGal a 4 Galが存在することも、結合に負の効果をもたした。さらに、コア2型（またはコア4型）O結合コア上のⅡⅡ型ラクトサミンを含む Gal b 4 Gl c NAc b 6 Gal NAc 単位は、結合に負の効果をもたした。

#### 【0088】

第2の規則は、グリカンモチーフが少なくとも2つのGal b 4 Gl c NAc単位を有するⅡⅡ型ポリラクトサミン反復を含む場合、シアル酸付加は高親和性結合に影響を与えないことを示す興味深いパターンを示した。過去の研究では、末端にシアル酸が付加されたグリカンは、ガレクチン1に対するリガンド候補であるとされている。本研究で使用されたシアル酸含有グリカンは少なくとも2つのGal b 4 Gl c NAc単位を含んでいたため、これらの結果は発明者らの規則と一致する。さらに発明者らの規則は、ガレクチン1が内部Gal b 4 Gl c NAc単位に結合し、非還元末端に対して鎖のはるか遠方にある他の任意のパターンも高親和性結合に影響を及ぼさないことも示す。Gal [ 3 O S O<sub>3</sub> ] b 3 Gal NAcを含む偽陰性が1つだけがあった。

#### 【0089】

ガレクチン3結合の規則はガレクチン1に類似していたが、幾つかの違いがあった。これらの違いは表4に捕捉されている。

#### 【0090】

表4．ガレクチン1およびガレクチン3の結合の比較

#### 【0091】

#### 【表4】

	ヒトガレクチン -1	ヒトガレクチン -3
GlcNAc上のFuc ◀a2/3 ■	結合を阻害	結合を阻害
Gal上のFuc ◀a2 ⊙	結合を阻害	阻害しない
Gal上の NeuAcまたは NeuGc ◆a3/6 ⊙ / ◆a3/6 ⊙	結合を阻害	結合を阻害
ラクトースアミン単位の配置	直鎖配置より分岐配置を好む	分岐配置より直鎖配置を好む
最少鎖長	必要	必要

主な違いは、他のパターンと併用した、Man a 3 (Man a 6) Man単位の非存在の組合せを有する第1の規則にあった。この規則が全てのN結合グリカンを排除するのではないことを指摘することは重要である。その代わりに、ガレクチン3は、コアのMan a 3 (Man a 6) Manに結合した様々な分岐部に出現するGal b 4 Gl c NAcと比較して、Gal b 4 Gl c NAc（ポリラクトサミン）の直鎖反復を好むことを意味する。別の違いは、ガレクチン3に対する結合は、ラクトサミンにおけるGalのフコシル化により阻害されなかったが、ガレクチン1に対する結合はそれにより阻害されたことであった。

#### 【0092】

DC SIGNおよびSIGNRの例と同様に、ガレクチンデータの発明者らの解析結果を、リガンド結合の構造的態様と比較した。Gal b 4 Gl c NAc、Neu5Ac a 3 Gal b 4 Gl c NAc、Neu5Ac a 3 Gal b 4 (Fuc a 3) Gl c NAc、Neu5Ac a 6 Gal b 4 Gl c NAcなどのような様々なリガンドと、ガレクチン1および3との構造複合体を解析した。Gal b 4 Gl c NAcリガンドを伴ったガレクチン1および3の結晶構造は、他のリガンドの構造をスーパーインポーズし、様々な構造複合体を構築するための枠組みとして、それぞれ使用した。Gal b 4 Gl c NAc単位におけるGalの4 OH基および6 OH基ならびにGl c NAcの3 OHは、ガレクチン1および3のCRDのアミノ酸との相互作用に関与していた。したがって、これら



／ 6シアル酸含有グリカンと、H Aグリカン結合部位との相互作用についての重要な洞察を提供する。第2番目は、様々なH 1型、H 3型、およびH 5型H Aに関するグリカンアレイデータを解析するためのデータマイニング手法を含む。このデータマイニング解析は、様々な野生型および突然変異型H Aの強結合体、弱結合体、および非結合体を、マイクロアレイのグリカンの構造的特徴とを相関させる(表7)。

#### 【0099】

重要なことには、これらの相関(分類指標)は、2 3／6シアル酸含有結合および／または様々なH Aに対する結合の際の様々なトポロジーの、微妙な構造変化の影響を捕捉する。下記で考察するように、データマイニング解析から得られるグリカン特徴の相関は、H Aグリカン結合部位上にマッピングされ、異なるトポロジーのグリカンを含む2 3および2 6シアル酸含有グリカンに対する、H 1型、H 3型、およびH 5型H Aの結合を系統的に研究するための枠組みを提供する。

#### 【0100】

一例を挙げると、本発明によってこの枠組みをH 5型H Aに応用すると、特に分岐度との関連で、2 6オリゴ糖鎖の長さが、いかにしてグリカン周辺の構造変化の微妙な違いより重要になるかが例証される。例えば、より長鎖の2 6モチーフを有する二分岐構造と比較して、単一の2 6モチーフを有する三分岐構造は、個々の2 6モチーフ周辺での構造変化に比べて、H Aグリカン結合に影響を及ぼす。これは、本明細書でデータマイニングから得られた、2 6モチーフについての特徴的な鎖長依存性分類指標により確認される(表7)。

#### 【0101】

2 3および2 6シアル酸含有グリカンに対する、H 1型、H 3型、およびH 5型H Aの結合特異性のための枠組み

H 1(PDB ID: 1RD8、1RU7、1RUY、1RV0、1RV T、1RV X、1RV Z)、H 3(PDB ID: 1MQL、1MQM、1MQN)、およびH 5(1JSN、1JSO、2FK0)に由来するH Aの結晶構造、ならびに2 3および／または2 6シアル酸含有オリゴ糖とH Aとの複合体の結晶構造は、特定のH Aグリカン相互作用に関与する残基についての分子的洞察を提供した。より最近では、トリおよびヒトのH 1亜型およびH 3亜型のグリカン受容体特異性は、多様な2 3および2 6シアル酸含有グリカンを含むグリカンアレイ上で、野生型および突然変異体をスクリーニングすることにより詳説された。

#### 【0102】

1918年のヒト汎流行性ウイルスのH AにおけるAsp190Glu突然変異は、2 6シアル酸含有グリカンから2 3シアル酸含有グリカンへとその特異性を逆転させた(Stevensら、J. Mol. Biol., 355: 1143, 2006; Glaserら、J. Virol., 79: 11533, 2005)。他方では、トリH 1(A/アヒル/Alberta/35/1976)の二重突然変異Glu190AspおよびGly225Aspは、2 3シアル酸含有グリカンから2 6シアル酸含有グリカンへとその特異性を逆転させた。H 3亜型の場合、1963年のトリH 3N8型株および1967～68年の汎流行性ヒトH 3N2型株間での、Gln226からLeuへのアミノ酸変化およびGly228からSerへのアミノ酸変化は、2 3シアル酸含有グリカンから2 6シアル酸含有グリカンへの優先性の変化と相関する(Rogersら、Nature, 304: 76, 1983)。H Aグリカン結合の特異性と伝染効率との関係は、高病原性で伝染力の強い1918年のH 1N1型ウイルスを使用して、フェレットモデルで実証された(Tumpey, T. M. et al. Science 315: 655, 2007)。受容体結合特異性が、親のヒト2, 6シアル酸含有グリカン(SC18)受容体優先性からトリ2, 3シアル酸含有受容体優先性(AV18)に切り替わると、その結果として伝染不能であるウイルスがもたらされた。他方では、2, 3／2, 6混合シアル酸含有グリカン特異的ウイルス(A/New York/1/18(NY18))は伝染性を示さず、驚くべきことに、2, 3／2, 6混合シアル酸含

10

20

30

40

50

有グリカン特異性でもある A / T e x a s / 3 6 / 9 1 ( T x 9 1 ) ウイルスは、効率的に伝染可能であった。さらに、上述のように、高病原性 H 5 N 1 型ウイルスの種々の株は、2, 3 / 2, 6 混合シアル酸含有グリカン特異性も示すが ( Y a m a d a , S . e t a l . N a t u r e 4 4 4 : 3 7 8 , 2 0 0 6 )、ヒトからヒトに伝染可能である。H A のシアル酸含有グリカン特異性および伝染性に関する混乱した結果は、以下の疑問を提起した。まず、ヒトの上気道に見出されるシアル酸含有グリカンに多様性があるのか、それはウイルスの特異性および組織指向性を説明できるのか。次に、2 3 および / または 2 6 シアル酸含有グリカンの両方が、いかにして H A グリカン結合ポケットに結合するのかに関して役割を果たし得るグリカンのコンフォメーションに微妙な違いがあるのか。まとめると、ヒトに適応するための、インフルエンザ A 型ウイルス H A のグリカン結合の必要条件は何か。

10

#### 【 0 1 0 3 】

グリカントポロジと、2 3 シアル酸含有グリカンに結合する H 1 型、H 3 型、および H 5 型 H A における置換とにより課された構造的制約。

#### 【 0 1 0 4 】

全ての H A グリカン共結晶構造を解析すると、N e u 5 A c 糖 ( S A ) の配向性が H A グリカン結合部位に対して固定されていることが示される。様々な H A 亜型にわたって高度に保存されている一連のアミノ酸、P h e 9 5、S e r / T h r 1 3 6、T r p 1 5 3、H i s 1 8 3、L e u / I l e 1 9 4 は、S A の係留に関与する。したがって、2 3 または 2 6 に対する H A の特異性は、S A を越えたグリコシドの酸素原子および糖と H A グリカン結合部位との相互作用により決定される。

20

#### 【 0 1 0 5 】

N e u 5 A c 2 3 G a l 結合のコンフォメーションは、2 3 の G a l および G a l を越えた糖の位置決めが、この結合におけるグリコシドのねじれ角により決定される円錐形領域内に当てはまる状態である ( 図 6 )。最小エネルギーコンフォメーションの典型的な領域は、 $\phi$  が - 6 0 から 6 0 をサンプリングする約 - 6 0 または 6 0 または 1 8 0 の値により示される ( 図 1 4 )。これらの最小エネルギー領域では、2 3 の G a l を越えた糖が、H A グリカン結合部位から突き出ている。これは、 $\phi$  値が典型的には約 1 8 0 ( トランス型コンフォメーションと呼ぶ ) である 2 3 モチーフ ( N e u 5 A c 2 3 G a l 1 3 / 4 G l c N A c ) と H A との共結晶構造からも明らかである。トランス型コンフォメーションは、2 3 モチーフがポケットから突き出る原因となる。これは、3 つの糖 ( または三糖 ) の 2 3 モチーフの中央にある G a l および / または G l c N A c ( もしくは G a l N A c ) 糖で分岐する構造変化 ( 硫酸化およびフコシル化 ) が、H A 結合に最も影響を与えることを示唆する ( 図 7 )。この構造的含意は、データマイニング解析から得られた、2 3 シアル酸含有グリカンに対する H A 結合の 3 つの特徴的な分類指標と一致している ( 表 7 )。これらの 3 種類の全てに共通する特徴は、N e u 5 A c 2 3 G a l が G a l N A c / 1 4 G a l ) と共に存在すべきでないということである。結晶構造の解析は、N e u 5 A c 2 3 G a l の G a l に結合した G a l N A c がタンパク質との好ましくない立体接触をしたことを示し、それは分類指標と一致した。

30

40

#### 【 0 1 0 6 】

シアル酸結合のための保存された係留ポイントに加えて、2 つの重要な残基、すなわち G l n 2 2 6 および G l u 1 9 0 が、N e u 5 A c 2 3 G a l モチーフに対する結合に関与する。結合部位の基部に位置する G l n 2 2 6 は、N e u 5 A c 2 3 G a l 結合のグリコシド酸素原子と相互作用する ( 図 1 5、パネル C、D )。G l n 2 2 6 の反対側に位置する G l u 1 9 0 は、N e u 5 A c および G a l 単糖と相互作用する ( 図 1 5、パネル C、D )。さらに、トリ H A において高度に保存されている残基 A l a 1 3 8 ( G l n 2 2 6 の近位 ) および G l y 2 2 8 ( G l u 1 9 0 の近位 ) は、2 3 シアル酸含有グリカンと至適に相互作用するための G l n 2 2 6 および G l u 1 9 0 の正しいコンフォメーションの促進に関与し得る ( 図 1 5 )。ヒト H 1 亜型である A P R 3 4 は、A l a

50

138、Glu190、Gln226、およびGly228の4つのアミノ酸を全て含み、その結晶構造で観察されるように 2 3シアル酸含有グリカンに結合する(図14、パネルB)。

#### 【0107】

AAI68\_\_H3\_\_23、ADU67\_\_H3\_\_23、およびAPR34\_\_H1\_\_23の結晶構造中のグリカン結合部位をスーパーインポーズすると、Glu190側鎖の位置決め、および 2 3シアル酸含有グリカンへのHA結合に対するその影響についての洞察がさらに得られる。H1型HAのGlu190の側鎖は、H3型HAのGlu190側鎖と比較して、さらに(およそ1 )結合部位寄りである。これは、Glu190残基の近位にある、H3型HAのSer186と比べたH1型HAのアミノ酸の違いであるPro186に起因し得る。Glu190の側鎖コンフォメーションにおけるこの変化は、グリカンマイクロアレイデータのデータマイニング解析により示されたように(表7)、トリH1型(および非トリH3型)の、中程度の親和性で幾つかの 2 6シアル酸含有グリカンへの結合と相関し得る。さらに、トリおよびヒトH3亜型間の特徴的な変化である、Gly228のSerへの置換は、Glu190のコンフォメーションを変化させ、トランス型コンフォメーションのNeu5Ac 2 3GalとヒトH3型HAとの相互作用を妨害する。これは、ヒトAAI68\_\_H3\_\_23共結晶構造に観察されるNeu5Ac 2 3Galモチーフの特徴的なコンフォメーション(トランス型ではない)により、さらに詳説される。このコンフォメーションのNeu5Ac 2 3Galモチーフは、トランス型コンフォメーションのこのモチーフによりトリH3型HAにもたらされたものと比較して、あまり至適ではない接触をヒトH3型HA結合部位にもたらす。(図14)。この接触の喪失の結果として、ヒトH3型HAのGly228Ser突然変異は、そのグリカン結合部位を、 2 3シアル酸含有グリカンとの相互作用にそれほど好ましくないものにする。この構造的観察は、ヒトH3型HAが幾つかの 2 3シアル酸含有グリカンに対して中程度の親和性のみを有することを示すデータマイニング解析の結果と一致する(表7)。

#### 【0108】

Neu5Ac 2 3Gal周辺の構造変化は、どのようにHA グリカン相互作用に影響を及ぼすか。トリH5において高度に保存されているLys193(図5)は、Neu5Ac 2 3Gal 1 4GlcNAcの6 O硫酸化Galおよび/または6 O硫酸化GlcNAcと相互作用するように配置されている。この観察は、トリH5だけがGalまたはGlcNAcで硫酸化された 2 3シアル酸含有グリカンに高親和性で結合するというデータマイニング解析により確認される(表7)。同様に、222位の塩基性アミノ酸は、Neu5Ac 2 3Gal 1 3GlcNAcモチーフの4 O硫酸化GlcNAc、またはNeu5Ac 2 3Gal 1 4GlcNAcモチーフの6 O硫酸化GlcNAcと相互作用し得る。他方では、H1およびH5のLys222ならびにH3のTrp222などのかさ高い側鎖は、Neu5Ac 2 3Gal 1 4(Fuc 1 3)GlcNAcモチーフのフコシル化GlcNAcを潜在的に妨害する。この構造的観察は、トリH3型株およびH5型株に観察される分類指標規則 2 3タイプC(表7)を実証し、GlcNAcにおけるフコシル化が結合に有害であることを示す。 2 3シアル酸含有グリカンに対するViet04\_\_H5型HAの結合は、それぞれのグリカン結合部位のアミノ酸がほとんど同一であることを考慮すると、ADS97\_\_H5型HAの結合に類似している(表7)。

#### 【0109】

したがって、 2 3シアル酸含有グリカンに対する結合に関しては、Neu5Acを係留する残基とは別に、トリH1亜型、H3亜型、およびH5亜型の全てにおいて高度に保存されているGlu190およびGln226が、Neu5Ac 2 3Galモチーフに対する結合に重要である。GlcNAcまたはGalNAcとの接触、ならびに 2 3モチーフにおける硫酸化およびフコシル化などの置換は、137位、186位、187位、193位、および222位のアミノ酸を含む。H1、H3、およびH5に由来する

10

20

30

40

50

H A は、グリカンマイクロアレイに存在する多様な 2 3 シアル酸含有グリカンに対する異なる結合特異性を示す。これらの位置のアミノ酸残基は、様々な H A にわたって保存されているわけではなく、それが異なる結合特異性を説明する。

#### 【0110】

グリカントポロジーと、 2 6 シアル酸含有グリカンに結合する H 1 型、H 3 型および H 4 型 H A における置換とにより課された構造的制約

Neu5Ac 2 6 Gal 結合の場合には、追加的な C 6 C 5 結合がコンフォメーションの追加的な柔軟性を提供する。 2 6 の Gal およびそれに続く糖の位置は、 2 3 の場合の円錐形様領域と比較して、はるかに大きな傘形様領域にわたる (図 6)。

2 6 に対する結合は、そのような傘形トポロジーの基部にある Neu5Ac および Gal 糖との至適な接触を伴い、オリゴ糖の長さに依存してそれに続く糖とも至適な接触を伴う。Neu5Ac 2 6 Gal 1 3 / 4 Glc などの短鎖 2 6 オリゴ糖は、潜在的に円錐形様のトポロジーをとる。他方では、 2 6 モチーフである Neu5Ac 2 6 Gal 1 4 GlcNAc に Glc の代わりに GlcNAc が存在すると、GlcNAc および Neu5Ac の両方のアセチル炭素間の至適なファンデルワールス接触により安定する傘形トポロジーが潜在的に好まれる。しかしながら、 2 6 モチーフは、分岐および H A 結合などの追加要因が、傘形トポロジーにより提供される安定性を補償できるような円錐形トポロジーをとることもできる。 2 6 モチーフの円錐形トポロジーは、糖内の相互作用により安定され得る N 結合グリカンにおける複数の短鎖オリゴ糖分岐部の一部として存在する。他方では、傘形トポロジーは、長鎖オリゴ糖分岐部 (少なくとも 1 つの四糖) における 2 6 モチーフにより好まれる。H 1 型および H 3 型 H A と 2 6 モチーフ (Neu5Ac 2 6 Gal 1 4 GlcNAc) との共結晶構造は、 ~ - 60 (シス型コンフォメーションと呼ぶ) が、Neu5Ac 2 6 Gal を越える糖を H A タンパク質に向かって屈曲させて、結合部位と至適に接触させるという上記の概念を支援する (図 7)。

#### 【0111】

H 1 型 H A においては、ヒト H 1 N 1 (A / South Carolina / 1 / 19 18) 亜型に由来する H A のグリカン結合ドメインと、A S I 30 \_\_ H 1 \_\_ 26 および A P R 34 \_\_ H 1 \_\_ 26 の結合ドメインとのスーパーインポーズが、 2 6 シアル酸含有グリカンへの特異性をもたらすことに関与するアミノ酸についての洞察を提供した。L y s 222 および A s p 225 は、Neu5Ac 2 6 Gal モチーフにおける Gal の酸素原子と相互作用するように配置されている。A s p 190 および S e r / A s n 193 は、Neu5Ac 2 6 Gal 1 4 GlcNAc 1 3 Gal モチーフの追加的な単糖である GlcNAc 1 3 Gal と相互作用するように配置されている (図 15、パネル A、B)。

#### 【0112】

A s p 190、L y s 222、および A s p 225 は、1918 年のヒト汎流行性株に由来する H 1 型 H A の中で高度に保存されている。アミノ酸 G l n 226 は、トリおよびヒト H 1 亜型の全てにおいて高度に保存されているが、 2 3 シアル酸含有グリカン (トリ H 1 亜型の) に対する結合における役割と比較して、 2 6 シアル酸含有グリカン (ヒト H 1 亜型の) に対する結合に関与しないと考えられる。トリおよびヒト H 1 型 H A の野生型および突然変異型についてのグリカンアレイ結果のデータマイニング解析は、 2 6 シアル酸含有グリカンに対する結合における上記のアミノ酸の役割をさらに実証する (表 7)。トリ H 1 型 H A の G l u 190 A s p / G l y 225 A s p 二重突然変異体は、 2 6 シアル酸含有グリカンに対する結合を逆転させる (表 7)。さらに、ヒト A N Y 18 \_\_ H 1 の L y s 222 L e u 突然変異体は、アレイ上のシアル酸含有グリカンに対するその結合を全て取り除き、グリカン結合における L y s 222 の必須な役割と一致する。

#### 【0113】

2 6 シアル酸含有グリカンに結合する H 3 N2 型 H A に特異性をもたらすアミノ酸

を特定するため、ヒトH3N2型(AA168\_\_H3)、ADU63\_\_H3\_\_26、およびASI30\_\_H1\_\_26に由来するHAのグリカン結合ドメインをスーパーインポーズした。これらのスーパーインポーズされた構造を解析すると、Leu226は、Neu5  
2 6GalモチーフのC6原子との至適なファンデルワールス接触をもたらすように配置され、Ser228は、シアル酸のO9と相互作用するように配置されていることが示された。ヒトH3のSer228は、Glu190とも相互作用し(相互作用しないトリADU63\_\_H3のGly228とは異なる)、それによりその側鎖コンフォメーションに影響を与える。ヒトH3型HAのGlu190の側鎖は、トリH3型HAのGlu190の側鎖と比較して、約0.7 だけわずかに結合部位に向かって移動されている。これらの違いは、ヒトH3型HAが 2 3シアル酸含有グリカンに結合し、 2 6シアル酸含有グリカンに対するその優先的結合と相関する能力を制限する。したがって、Gln226LeuおよびGly228Ser突然変異は、1967年の大流行中に、トリH3亜型のグリカン受容体特異性をヒトH3亜型に逆転させる原因となった。

#### 【0114】

1967~68年の汎流行性H3N2型に由来するHAと、より最近のH3亜型(1990年以降)に由来するHAとを比較すると、最近の亜型では、Glu190がAspに突然変異していることが示される。ヒトH3のAsp190がこれらのグリカンと好ましく相互作用するように位置しているため、この突然変異は、ヒトH3の 2 6シアル酸含有グリカンに対する結合をさらに強化する。この構造的含意は、ヒトH3亜型(A/Moscow/10/1999)に関するグリカンアレイデータのデータマイニング解析によりさらに確認される。このHAは、Asp190、Leu226、およびSer228を含み(図10)、 2 6シアル酸含有グリカンに強い優先性を示す(表7)。

#### 【0115】

上記の観察は、 2 6シアル酸含有グリカンに結合するH1型およびH3型HA間の違いだけでなく類似性も両方とも強調する。H1型およびH3型HAの両方において、Asp190およびSer/Asn193は、Neu5Ac 2 6Galモチーフを越えた単糖と好ましい接触をするように位置する(図15、パネルA、B)。アミノ酸およびそれらと 2 6シアル酸含有グリカンとの接触における、H1型HAとH3型HAとの違いは、性をこれらのグリカンに結合するための特徴的な表面補完性およびイオン補完を提供する。Neu5Ac 2 6Gal結合は、Neu5Ac 2 3Galに比べて追加的なコンフォメーションの自由度を有する。したがって、 2 6シアル酸含有グリカンに結合するHAは、このコンフォメーション自由度を収容するためにより開放的な結合ポケットを有する。ヒトH3型HAのLeu226は、Neu5Ac 2 6Galとの至適なファンデルワールス接触をもたらすように位置しているが、H1型HAのGln226によりこのモチーフにもたらされるイオン性接触は、それほど至適ではない。その一方で、H1においては、アミノ酸Lys222およびAsp225は、H3のTrp222およびGly225と比較して、より至適なイオン性接触を 2 6シアル酸含有グリカンにもたらす。

#### 【0116】

2 6シアル酸含有グリカンに対する、野生型および突然変異体H5型HAの結合についての構造的制約

H1型およびH3型HAの異なるアミノ酸によりもたらされる 2 6シアル酸含有グリカンとの相互作用は、現在のトリH5N1型HAが、そのグリカン受容体特異性を逆転するために、H1様またはH3様グリカン結合部位に変異し得ることを示唆した。上記の枠組みに基づき、H5型HAについての仮想的なH1様およびH3様突然変異を、下記で考察したようにさらに詳述し、試験した。

#### 【0117】

スーパーインポーズされたASI30\_\_H1\_\_26、APR34\_\_H1\_\_26、ADS97\_\_H5\_\_26、およびViet04\_\_H5構造の解析は、 2 6シアル酸含有グリカンに対する、H5型HAのH1様結合に関する洞察をもたらした。H1型およびH5型



H Aは同一の構造クレードに属するため、それらのグリカン結合部位は、類似のトポロジーおよびアミノ酸分布を共有する(Russellら、Virology, 325: 287, 2004)。トリH5型H Aに高度に保存されているLys 222は、H1型H Aの類似性Lysと同様に、Neu5Ac 2 6GalモチーフのGalに至適な接触をもたらすように位置する。Viet04\_\_H5のGlu190およびGly225(H1のAsp190およびAsp225の代わり)は、H1と同様に、Neu5Ac 2 6Gal 1 4GlcNAcモチーフに必要な接触をもたらさない。したがって、H5型H AのGlu190AspおよびGly225Asp突然変異は、2 6シアル酸含有グリカンとの接触を、潜在的に改善し得る。

#### 【0118】

Neu5Ac 2 6Gal 1 4GlcNAc 1 3Gal 1 4Glcオリゴ糖のGlcNAcを越える相互作用、ならびにH1型およびH5型H Aのグリカン結合ポケットを解析すると、H1型H AのSer/Asn193は最後から二番目のGalとの好ましい接触をもたらすが、H5の類似性Lys193は、GlcNAc 1 3Galモチーフとの好ましくない立体重複を有することが示された。したがって、Lys193Ser突然変異は、2 6シアル酸含有グリカンとの追加的な好ましい接触を(Glu190AspおよびGly225Asp突然変異と共に)もたらしすることができる。

#### 【0119】

H1型H Aに高度に保存されているGln226は、トリH5型H Aにも保存されている。Gln226が2 6シアル酸含有グリカンに対するH1型H Aの結合に積極的な役割をあまり果たさないとすれば(上記で考察したように)、このアミノ酸がLeuなどの疎水性アミノ酸に突然変異すると、Neu5Ac 2 6GalモチーフにおけるGalのC6原子とのファンデルワールス接触が潜在的に強化され得る。

#### 【0120】

ADU63\_\_H3\_\_26、AAI68\_\_H3、ADS97\_\_H5\_\_26、およびViet04\_\_H5をスーパーインポーズすると、2 6シアル酸含有グリカンに対するH5型H AのH3様結合についての洞察がもたらされる。このスーパーインポーズは、H5型およびH3型H Aのグリカン結合部位を構造的にアラインメントしたが、H5およびH1間の構造的なアラインメントほど良好ではなかった。H3型H AのLeu226およびSer228によりそれぞれもたらされるNeu5 2 6Galモチーフとの好ましいファンデルワールス接触およびイオン性接触は、H5型H Aには存在しなかった(Gln226およびGly228とは)。ヒトH3型H Aにおいては、Leu226およびSer228が2 6シアル酸含有グリカンに対する結合に重要であるとすれば、H5型H AのGln226LeuおよびGly228Ser突然変異は、潜在的に2 6シアル酸含有グリカンとの至適な接触をもたらすことができる。さらに、H3とH5との比較においてさえ、Lys193は、好ましい接触をもたらすように位置しているヒトH3型H AのSer193に比べて、Neu5Ac 2 6Galモチーフを越えた単糖との好ましくない立体接触を有するように位置している。1967~68年の汎流行性H3N2型に由来するH AはGlu190を含むが、H5型H AのAsp190は、より長いオリゴ糖のNeu5Ac 2 6Galモチーフとのより良好なイオン性接触をもたらすように位置する。

#### 【0121】

上述の残基の役割は、野生型および突然変異型のViet04\_\_H5についてのグリカンアレイデータのデータマイニング解析により、さらに確認された(表7)。二重突然変異体、Glu190Asp/Gly225Aspは、2 3シアル酸含有グリカンに結合するためのアミノ酸Glu190を欠如し、2 6シアル酸含有グリカンに結合するためのLys193に由来する立体障害を有するため、任意のグリカン構造物とも結合しない。同様に、二重突然変異体、Gln226Leu/Gly228Serは、幾つかの2 3シアル酸含有グリカンに結合するが(2 3タイプB分類指標)、たった1つの二分岐2 6シアル酸含有グリカンにしか結合しない(2 6タイプA分類指標)

10

20

30

40

50

。

## 【0122】

二分岐 2 6シアル酸含有グリカンに対するこの結合を解析すると、このグリカンの Neu5Ac 2 6Gal 結合は、接触がより少ないにもかかわらず、伸長型コンフォメーションで二重突然変異体に潜在的に結合できることが示された(図16)。さらに、Mal 1 3Man 分岐部の Neu5Ac 2 6Gal は、H5型HAのグリカン結合部位との好ましくない立体接触を有する Man 1 6Man 分岐部の同一モチーフと比較して、より好ましく結合する(図16)。Gln226Leu/Gly228Ser 二重突然変異体の、2 6シアル酸含有グリカンに対する特異性が限られていることは、Lys193が結合を妨害することと一致する。

10

## 【0123】

任意の特定の理論により束縛されることは望まないが、本発明者らは、インフルエンザA型ウイルスHAのヒト適応に必要な条件は、長鎖 2 6(支配的にヒト上気道で発現される)に高親和性で結合する能力を獲得することであると提唱する。例えば、グリカン多様性の一態様は、シアル酸でキャッピングされるラクトサミン分岐部の長さである。これは、データマイニング解析から導出された 2 6シアル酸含有グリカンの2つの明確な特徴により捕捉される(表7)。1つの特徴は、N結合コアのManに結合する Neu5Ac 2 6Gal 1 4GlcNAcにより特徴付けられ、もう1つは、より長い分岐部(典型的には傘形トポロジーをとる)を形成する別のラクトースアミン単位に結合するこのモチーフにより特徴付けられる。したがって、上気道に対する突然変異体H5型HAの広範な結合は、これらの突然変異体が、傘形トポロジーをとる長鎖 2 6を有するグリカンに高親和性で結合する場合のみ可能であり得る。例えば、本発明によれば、望ましい結合パターンは、図9に示す傘形グリカンに対する結合を含む。

20

## 【0124】

対照的に、発明者らは、修飾H5型HAタンパク質(Gly228SerおよびGln226Leu/Gly228Ser置換を含む)が、グリカンアレイ上のたった1つの二分岐 2 6スアリルラクトサミングリカン構造物に対する結合を示したという最近の報告に留意する(Stevensら、Science 312:404, 2006)。したがって、本明細書において、そのような修飾H5型HAタンパク質は、BSHB性H5型HAではない。

30

## 【0125】

2 6シアル酸含有グリカンに対する、野生型および突然変異体H5型HAの結合したがって、本発明は、これらのグリカンとヒトH1型またはH3型HAとの相互作用に基づき、現在のトリH5N1型HAが、2 6グリカンに対するその特異性を変更する突然変異を起こし得ることを実証する。Glu190Asp、Lys193Ser、Gly225Asp、およびGln226Leu突然変異(「DSDL突然変異体」)は、H5型HA結合部位をヒトH1型HAの結合部位に類似させることが潜在的に可能である一方、Glu190Asp、Lys193Ser、Gln226Leu、およびGly228Ser(「DSL突然変異体」)は、2 6シアル酸含有グリカンとの至適な相互作用のために、H5型HA結合部位をヒトH3型HAの結合部位に類似させることが潜在的に可能である。DSDL性およびDSL性H5型HA突然変異体を設計し、上記の枠組みに基づき試験した。野生型および突然変異体BSHB性H5型HAは、過去に報告されているようにバキュロウイルスで発現させ、精製した(図10XXX Y)。

40

## 【0126】

発明者らは、2 3シアル酸含有グリカンに対するトリH5型HAの結合と一致して、組換え野生型H5型HAのみが肺胞領域に広範に結合し、気管または気管支には結合するとしてもほんのわずかしが結合しないことを見出した。対照的に、DSL突然変異体(H3様)だけが上気道の気管および気管支組織に結合し、さらにこの突然変異体は、肺深部の肺胞組織に結合しない。

## 【0127】

50

組織結合実験の場合、組織切片を脱パラフィンし、再水和し、WTおよび突然変異体H Aタンパク質（PBSで希釈）と共に3時間にわたってインキュベートした。精製後の所定ロットのタンパク質濃度に基づき、1：10～1：100の範囲の適切な連続希釈が試験された。PBSで入念に洗浄した後、切片を2%BSA PBSで30分間ブロッキングし、次にウサギ抗トリH5N1型赤血球凝集素抗体（Pro Sci Inc、2%BSA PBS中で1：1000）で3時間インキュベートした。PBSで切片を洗浄し、次にヤギ抗ウサギ2次抗体（Invitrogen；2%BSA PBS中で1：500）で90分間インキュベートした。ヨウ化プロピジウムで切片を対比染色（赤色；Invitrogen；PBS中で1：200）した後、共焦点顕微鏡（Zeiss LSM 510レーザー走査型共焦点顕微鏡）で観察した。インキュベーションはすべて室温であった。

10

#### 【0128】

発明者らの枠組みに基づき、DSDLおよびDSL S突然変異の両方が、上気道の26シアル酸含有グリカンに広範に結合すると予想されたことを考慮すると、DSDL型ではなくDSL S型のH5型HAが、気管および気管支切片に結合した（しかし肺胞には結合しなかった）という観察は興味深い。これらの両突然変異体においてLys193がSer193に換わると、Lys193（野生型H5型HAの）により課された立体障害が取り除かれ、両突然変異体に26シアル酸含有グリカンへの幅広い特異性をもたらす。さらに、H5およびH1が同一の構造クレードに属することを考慮すると、H5型HAがH1様グリカン結合部位に変異する可能性はより高い。

20

#### 【0129】

DSDL突然変異体が26シアル酸含有グリカンに結合不能であることをさらに理解するために、ASI30\_\_H1\_\_26およびAPR34\_\_H1\_\_26結晶構造とさらにスーパーインポーズされたVie04\_\_H5結晶構造上にこの突然変異をマッピングした。このマッピングは、26シアル酸含有オリゴ糖との接触が全て、H1型HAおよびDSDL突然変異体間で保存されていることを示した。しかしながら、トリH5型HAに高度に保存されているAsp187は、DSDL突然変異体のAsp190に緊密に近接していた。3つのアスパラギン酸（Asp187、Asp190、およびAsp225）の存在は、DSDL突然変異体のpIが（WTおよびDSL S突然変異体の7.3と比較して）6.8であることをさらに説明した。Asp187およびAsp190間の相互作用は、H3型HAにおけるSer228のGlu190に対する影響と同様に、Asp190のコンフォメーションを潜在的に変化させ得る。187位のアミノ酸がAsp190に近接することの影響は、26シアル酸含有グリカンと比べて23シアル酸含有グリカンと相互作用するASI30\_\_H1のThr187のSASAの差からも明白である。Asp190が、H1型HAの26シアル酸含有グリカンとの至適な接触の形成に関与するとすれば、Asp187のAsp190に対する影響は、潜在的にこの相互作用を破壊できる。恐らくは、H1型HAに高度に保存されているGln226がDSDL突然変異体のLeuに突然変異すると、この突然変異体のHA結合部位の環境が他のH1様突然変異との関連で影響を受け、26シアル酸含有グリカンに対する結合にはそれほど至適ではなくなった。

30

40

#### 【0130】

Gln226を保持するGlu190Asp/Lys193SerまたはDS突然変異体を使用して、H5型HAのH1様結合におけるGln226の役割をさらに試験した。DS突然変異体の肺深部組織に対する結合の欠如は、（Glu190Asp突然変異による）23シアル酸含有グリカンに対する結合の喪失と一致している。同様に、この突然変異体の上気道組織に対する結合の欠如は、この突然変異の26シアル酸含有グリカンに対する結合を低下させ得る、Asp187のAsp190に対する破壊的な影響をさらに支援する。したがって、現在のトリH5N1型HAの突然変異は、26シアル酸含有グリカンに幅広い特異性を有するH3様グリカン結合部位をもたらすことを好む（H1様と比較して）。

50

## 【0131】

D S L S 突然変異体の肺上部に対する結合は、上気道の 2 6 シアル酸含有グリカンの多様性に関する疑問を提起する。ヒト気管支上皮 (H B E) 細胞のレクチン染色は、これらの細胞が、N 結合、O 結合、および糖脂質などの様々な 2 6 シアル酸含有グリカンに豊富であることを示す (図 18)。これらの 2 6 シアル酸含有グリカンの多様性は、H B E 細胞の細胞表面から N 結合グリカンを単離し、M A L D I M S 解析を使用してそれらを特徴付けることにより、さらに詳説した。

## 【0132】

具体的には、100 mM のクエン酸緩衝生理食塩水で > 90 % 集密的になった際に、約  $70 \times 10^6$  の 16 H B E 14 o 細胞 (D . C . Gruener t 博士提供; University of California, San Francisco) を回収し、プロテアーゼ阻害剤 (Calbiochem) で処理し均質化した後で、細胞膜を単離した。P N G a s e F (New England Biolabs) で細胞膜画分を処理し、反応混合液を 37 で終夜インキュベートした。酵素を失活させるために反応混合液を 10 分間煮沸し、Sep Pak C18 SPE カートリッジ (Waters) を使用して、脱グリコシル化されたペプチドおよびタンパク質を取り除いた。さらにグリカンを脱塩し、黒鉛化炭素の固相抽出カラム (Supelco) を使用して、中性 (25 % アセトニトリル画分) および酸性 (0.05 % のトリフルオロ酢酸を含む 50 % アセトニトリル) 画分へと精製した。酸性画分 (シアル酸含有グリカンを含む) は、ソフトイオン化条件で陰イオンモードの M A L D I T O F M S で解析した (加速電圧 22 kV、グリッド電圧 93 %、ガイドワイヤー 0.3 %、および 150 ns の抽出遅延時間)。代表的な質量ピークのこの M A L D I T O F T O F 断片解析は、N 結合グリカンの分岐パターンおよび分岐鎖長の増加の点での多様性を例証した。グリカン特性で観察されたより高次の分岐に比べてより長い分岐鎖長は、H5 型 H A のこれらのグリカンに対する結合に影響を及ぼすことができる。


## 【0133】


例えば、グリカン多様性の一態様は、シアル酸でキャッピングされるラクトサミン分岐部の長さである。これは、データマイニング解析から導出された 2 6 シアル酸含有グリカンの 2 つの明確な特徴により捕捉される (表 7)。1 つの特徴は、N 結合コアの M a n に結合する N e u 5 A c 2 6 G a l 1 4 G l c N A c により特徴付けられ、もう 1 つは、より長い分岐部を形成する別のラクトースアミン単位に結合するこのモチーフにより特徴付けられる。したがって、突然変異体 H5 型 H A の上気道に対する広範な結合は、これらの突然変異体が 2 6 シアル酸含有グリカンに対する幅広い結合特異性を有する場合のみ可能である。例えば、本発明によれば、望ましい結合パターンには、図 9 に示すものならびに / あるいは :




## 【0134】

## 【化 2】

, ,  およびその組合せ:

,

, ならびに / あるいは

, , , およびその組合せ

が含まれる。

## 【0135】

対照的に、発明者らは、修飾 H5 型 H A タンパク質 (G l y 2 2 8 S e r および G l n 2 2 6 L e u / G l y 2 2 8 S e r 置換を含む) が、グリカンアレイ上のたった 1 つの二分岐 2 6 スアリルラクトサミングリカン構造物に対する結合を示したという最近の報

告に留意する (S t e v e n s ら、S c i e n c e 3 1 2 : 4 0 4 , 2 0 0 6 )。したがって、本明細書において、そのような修飾 H 5 型 H A タンパク質は、B S H B 性 H 5 型 H A ではない。

【 0 1 3 6 】

等価物

当業者であれば、単なる定常的な実験法を使用して、本明細書に記載した本発明の特定の実施形態に対する多数の等価物を認識または確認できよう。本発明の範囲は、上記の明細書に限定されることを意図するものではなく、むしろ以下の特許請求の範囲に記載した通りである。

【 0 1 3 7 】

【表 6】

表 6. HA - グリカン複合体の結晶構造

略語	(PDB ID)	ウイルス株	グリカン (座標が割り当てられている)
ASI30_H1_23 (1RV0)		A/ブタ /Iowa/30 (H1N1)	Neu5Ac
ASI30_H1_26 (1RVT)		A/ブタ /Iowa/30 (H1N1)	Neu5Aca6Galβ4GlcNAcβ3Galβ4Glc
APR34_H1_23 (1RVX)		A/Puerto Rico/8/34 (H1N1)	Neu5Aca3Galβ4GlcNAc
APR34_H1_26 (1RVZ)		A/Puerto Rico/8/34 (H1N1)	Neu5Aca6Galβ4GlcNAc
ADU63_H3_23 (1MQM)		A/7ヒム /Ukraine/1/63 (H3N8)	Neu5Aca3Gal
ADU63_H3_26 (1MQN)		A/7ヒム /Ukraine/1/63 (H3N8)	Neu5Aca6Gal
AAI68_H3_23 (1HGG)		A/Aichi/2/68 (H3N2)	Neu5Aca3Galβ4Glc
ADS97_H5_23 (1JSN)		A/7ヒム /Singapore/3/97 (H5N3)	Neu5Aca3Galβ3GlcNAc
ADS97_H5_26(1JSO)		A/7ヒム /Singapore/3/97 (H5N3)	Neu5Ac
Viet04_H5 (2FK0)		A/Vietnam/1203/2004 (H5N1)	

HA - α2 - 6シリアル糖含有グリカン複合体は、ASI30\_H1\_26およびAPR34\_H1\_26 (H1) 上に、ADU63\_H3およびADS97\_H5ならびにViet04\_H5のHA1サブユニットの CA1ドメインをスピンオフすることにより生成された。α2 - 6シリアル糖含有グリカンとHA/Aichi/2/68 (H3N2) との構造複合体は発表されているが<sup>11)</sup>、それらの座標はProtein Data Bankで入手可能ではなかった。SARF2 (<http://123d.ncicrf.gov/sarf2.html>) フォントを使用して、スピンオフ用の様々なHA1サブユニットの構造的アライメントを取得した。

10

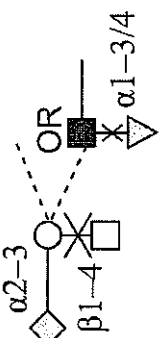



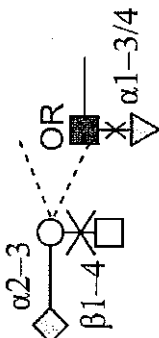

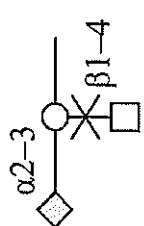

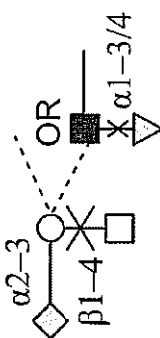

20

30

40

【表 7 - 1】

表 7. 分類指標規則に基づくHAのグリカン受容体特異性

インフルエンザ株	$\alpha 2-3$ $\text{HA}^{\text{a}}$	$\alpha 2-6$ $\text{HA}^{\text{b}}$
A/7 $\text{H}1$ /Alberta/35/76 (トリ <i>H1N1</i> )	 ( $\text{HA}^{\circ}$ C)	 ( $\text{HA}^{\circ}$ A <sup>1</sup> )
A/7 $\text{H}1$ /Alberta/35/76 (トリ <i>H1N1</i> ) Glu190Asp/Gly225 Asp 二重突然変異体	なし	 ( $\text{HA}^{\circ}$ B)
A/South Carolina/1/18 (ヒト <i>H1N1</i> )	なし	 (AまたはB)
A/New York/1/18 (ヒト <i>H1N1</i> )	 ( $\text{HA}^{\circ}$ C <sup>2</sup> )	 ( $\text{HA}^{\circ}$ B <sup>3</sup> )
A/Texas/36/91 (ヒト <i>H1N1</i> )	 ( $\text{HA}^{\circ}$ A <sup>5</sup> )	 (AまたはB)
A/New York/1/18 (ヒト <i>H1N1</i> ) Asp 190Glu突然変異体 <sup>4</sup>	 ( $\text{HA}^{\circ}$ C <sup>5</sup> )	 (AまたはB)

【 0 1 3 9 】

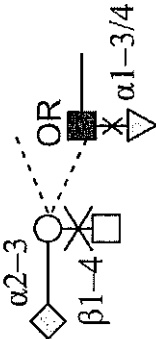

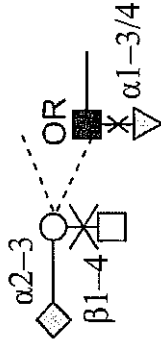
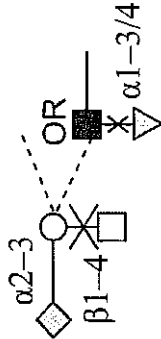
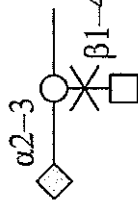
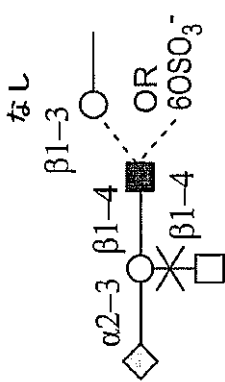
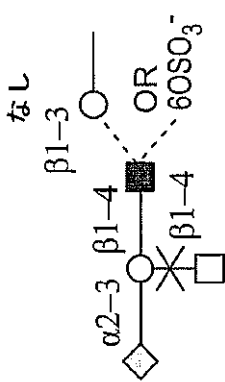

10

20

30

40

【表 7 - 2】

表 7 ( 続き )			
A/New York/1/18 ( ヒト H1N1 ) Lys222 Leu 突然変異体	なし 	なし	
A/ 7 <del>ヒ</del> /Ukraine/1/63 ( トリ H3N8 )	なし 6		( 417° B <sup>7</sup> )
A/Moscow/10/99 ( ヒト H3N2 )		なし	
A/ 7 <del>ヒ</del> /Singapore/3/97 ( トリ H5N3 )		なし	( 417° C <sup>2</sup> )
A/Vietnam/1203/04 ( トリ H5N1 )		なし	( 417° A )
A/Vietnam/1203/04 ( トリ H5N1 ) Glu 190Asp/ Gly225 Asp 二重突然変異体	なし 	なし	
A/Vietnam/1203/04 ( トリ H5N1 ) Gln226Leu/ Gly228Ser 二重突然変異体			( 417° A )

10

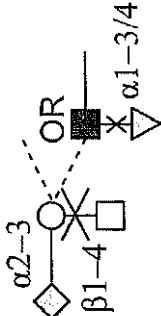
20

30

40



表 7 ( 続き )

<p>A/Vietnam/1203/04 ( トリ H5N1 ) Arg216Glu, Ser221 Pro 二重突然変異体</p>	
--	---

1 境界線上の高結合体 ; 2 硫酸化 GlcNAc[6S]/Gal[6S] 高結合体 ; <sup>3</sup> α 2 - 6 タイプ B に対する境界線上の高結合体 硫酸化 GlcNAc[6S]/Gal[6S] だけが高結合体である ; <sup>4</sup> 幾つかの非シアル酸含有グリカンに結合する ; <sup>5</sup> α 2 - 3 シアル酸含有グリカンに対して境界線上の高結合体 ; <sup>6</sup> Neu5Acα3Galβ3/4GlcNAc 上の硫酸化GlcNAcに対する境界線上の極少数の高結合体 ; <sup>7</sup> 高結合体は Neu5Acα6Galβ4GlcNAcβ3Gal & !GlcNAcα6Man である ; 他は境界線上の高結合体。

記号 : ■ GlcNAc; □ Gal; ● Man; ▽ Fuc; ◆ Neu5Ac;

H1 亜型、H3 亜型、および H5 亜型のグリカンマイクログリスクリングからのデータは、

Glycomics (CFG) のウェブサイト - <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp> から取得した。特徴および分類指標の記述を含むデータマイニング解析の詳細は、添付の図5で提供する。規則帰納分類法を使用して、α 2 - 3 / 6 シアル酸含有グリカンに対する H A の結合を決定する、以下の分類指標 (または規則) を生成した。

α 2 - 3 シアル酸含有グリカン結合の分類指標 - タイプ A: Neu5Acα3Gal & !GalNAcβ4Gal, Neu5Acα3Galβ4GlcNAc & !GalNAcβ4Gal & {GlcNAcβ3Gal または GlcNAc[6S]} , タイプ C: Neu5Acα3Galβ & !GalNAcβ4Gal & !Fucα3/4GlcNAc. α 2 - 6 シアル酸含有グリカン結合の分類指標 - タイプ A: Neu5Acα6Galβ4GlcNAcb?Man, タイプ B: Neu5Acα6Galβ4GlcNAc & !GlcNAcb?Man. これらの複雑な規則は、明瞭さのために表中では図表で表わされている。

規則は、結合を強化する高親和性結合体の特徴ならびに結合に有害である弱結合体および非結合体の特徴の論理的な組み合わせとして提供される (文中の記述では「!」という記号の後に示され、図表では「x」という記号を有する赤色の結合として示されている)。

α 2 - 6 分類指標におけるマンノースの存在は、グリカンアレイ上の二分岐 N 結合グリカンを含む単一の 6' - シアリルラクトサミンから生じる。

Consortium for Functional Glycomics (CFG) のウェブサイト - <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp> から取得した。特徴および分類指標の記述を含むデータマイニング解析の詳細は、添付の図5で提供する。規則帰納分類法を使用して、α 2 - 3 / 6 シアル酸含有グリカンに対する H A の結合を決定する、以下の分類指標 (または規則) を生成した。

α 2 - 3 シアル酸含有グリカン結合の分類指標 - タイプ A: Neu5Acα3Gal & !GalNAcβ4Gal, Neu5Acα3Galβ4GlcNAc & !GalNAcβ4Gal & {GlcNAcβ3Gal または GlcNAc[6S]} , タイプ C: Neu5Acα3Galβ & !GalNAcβ4Gal & !Fucα3/4GlcNAc. α 2 - 6 シアル酸含有グリカン結合の分類指標 - タイプ A: Neu5Acα6Galβ4GlcNAcb?Man, タイプ B: Neu5Acα6Galβ4GlcNAc & !GlcNAcb?Man. これらの複雑な規則は、明瞭さのために表中では図表で表わされている。

規則は、結合を強化する高親和性結合体の特徴ならびに結合に有害である弱結合体および非結合体の特徴の論理的な組み合わせとして提供される (文中の記述では「!」という記号の後に示され、図表では「x」という記号を有する赤色の結合として示されている)。

α 2 - 6 分類指標におけるマンノースの存在は、グリカンアレイ上の二分岐 N 結合グリカンを含む単一の 6' - シアリルラクトサミンから生じる。

【図 1】

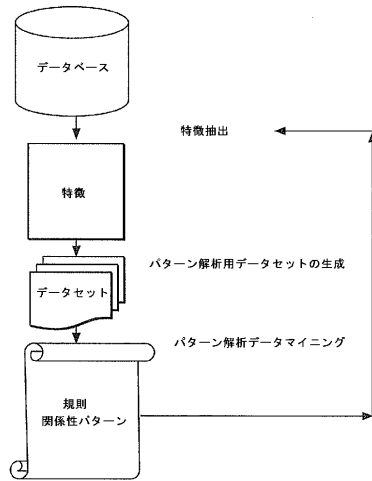


FIG. 1A

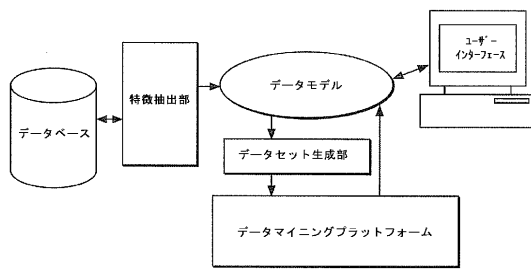


FIG. 1B

【図 3】

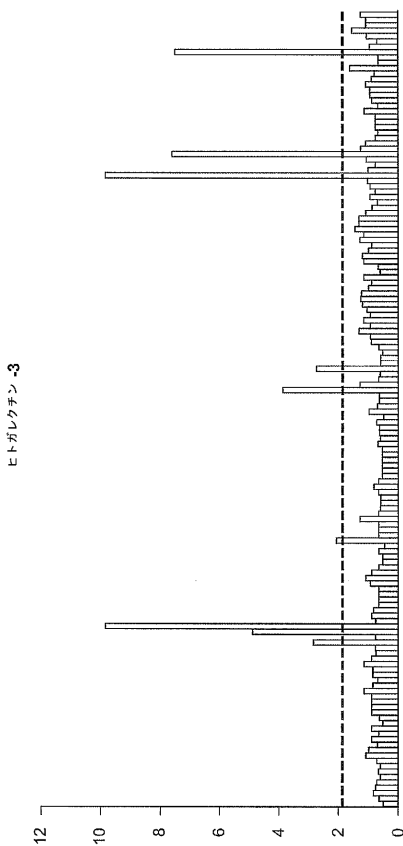


FIG. 3

【図 2】

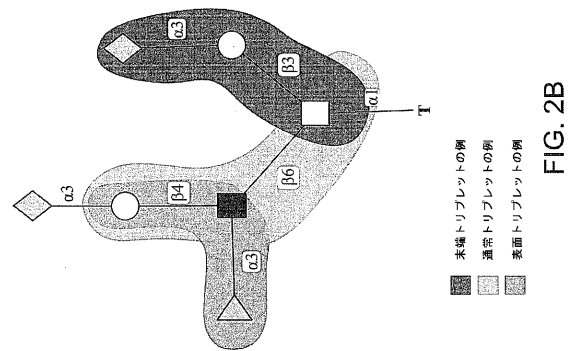


FIG. 2B

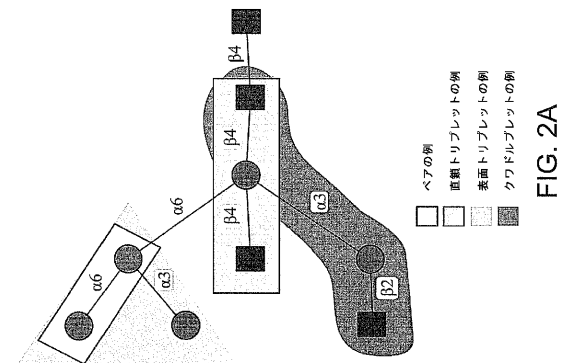


FIG. 2A

【図 4】

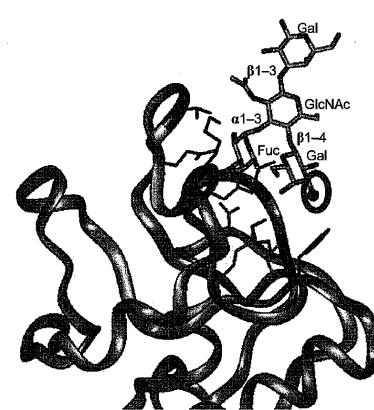


FIG. 4



【 図 6 】

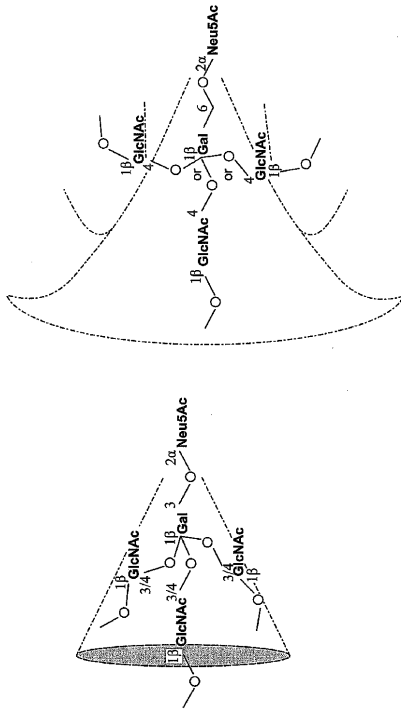


FIG. 6

【 図 7 】

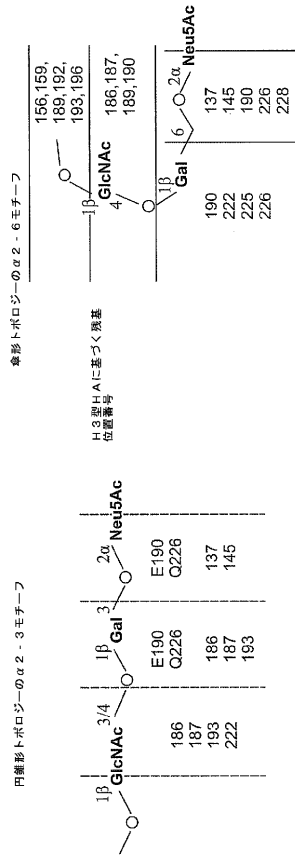
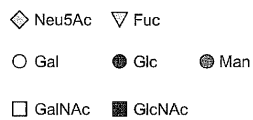
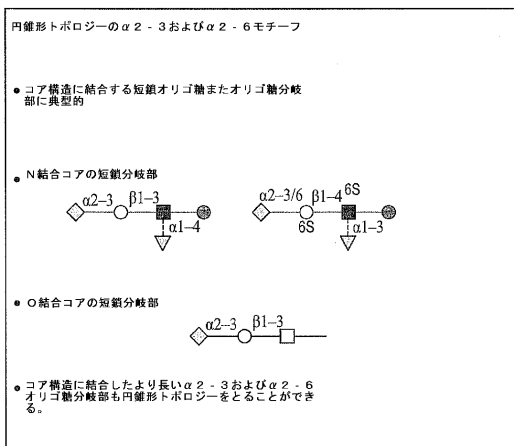


FIG. 7

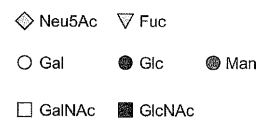
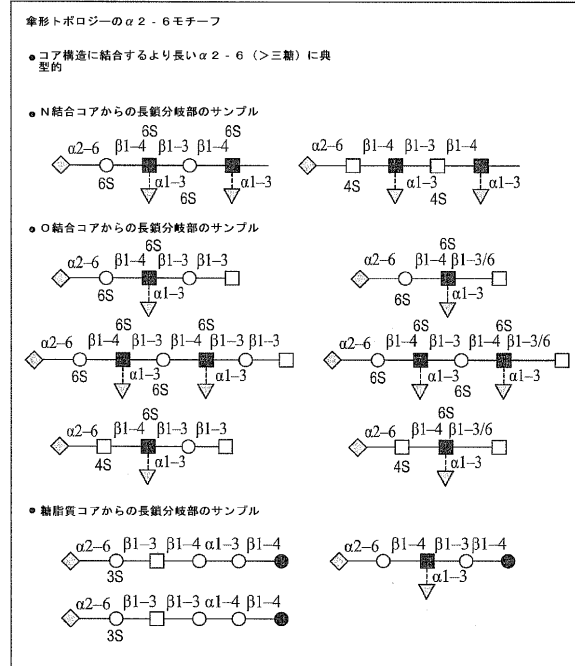
【 図 8 】



灰色の点線、4S、および6Sは、フコシル化および硫酸化修飾の潜在的な部位を示す。

FIG. 8

【 図 9 】



灰色の点線、4S、および6Sは、フコシル化および硫酸化修飾の潜在的な部位を示す。

FIG. 9













【 図 1 5 C 】

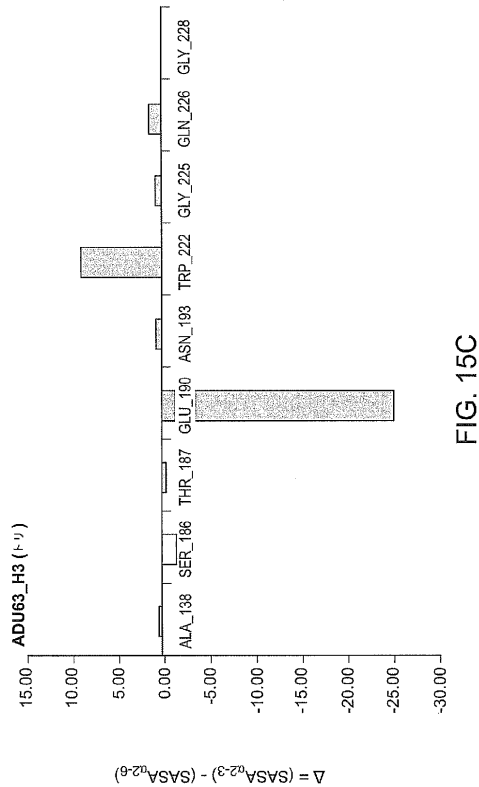


FIG. 15C

【 図 1 5 D 】

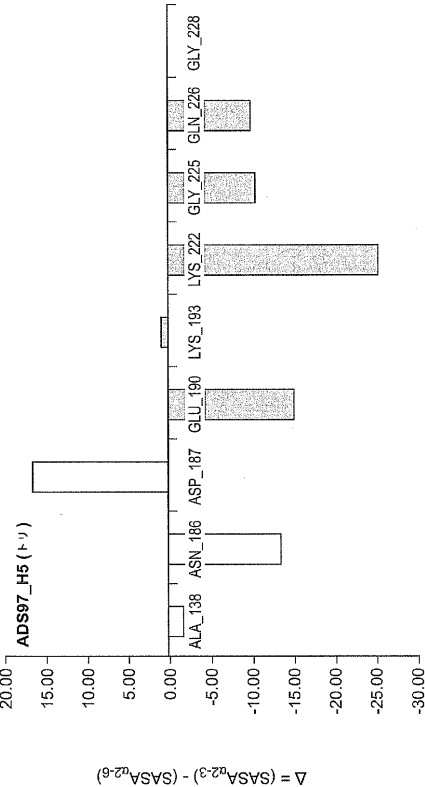


FIG. 15D

【 図 1 5 E 】

		138	186	187	190	193	222	225	226	228
H1	$\alpha 2-3$	A	P	N/T	E	S/N	K	G	Q	G
	$\alpha 2-6$	A	P	N/T	D	S/N	K	D	Q	G
H3	$\alpha 2-3$	A	S/N	N/T	E	S/N	W	G	Q	G
	$\alpha 2-6$	A	S/N	N/T	E/D	S/N	W	G	L	S
H5	$\alpha 2-3$	A	S/N	D	E	K	K	G	Q	G

FIG. 15E

【 図 1 7 】

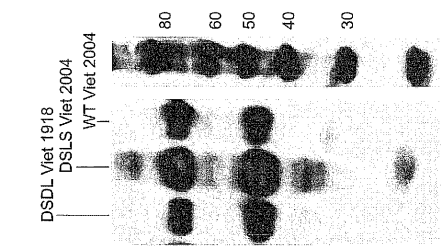


FIG. 17B

【 図 1 6 】

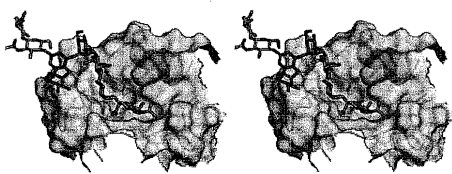


FIG. 16

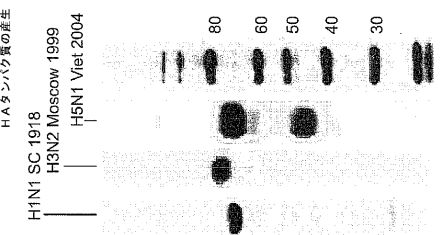


FIG. 17A

## 【図 18】

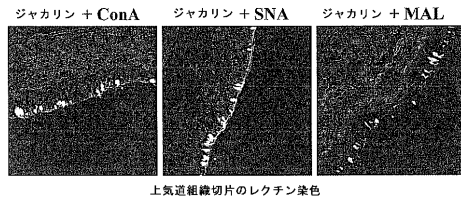


FIG. 18

## 【図 19】

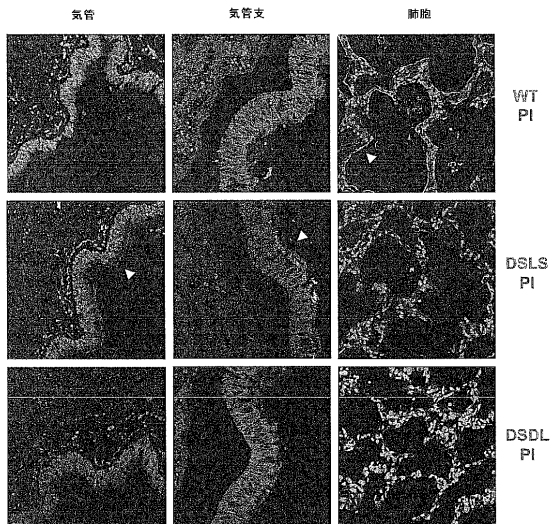


FIG. 19

## 【配列表】

2010501074000001.app

## 【 国際調査報告 】

<b>INTERNATIONAL SEARCH REPORT</b>		International application No. PCT/US07/18103		
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC: G06F 19/00( 2006.01);C07H 1/00( 2006.01);A61K 38/16( 2006.01);A61B 5/055( 2006.01)  USPC: 702/19;536/1.11;514/8;424/9.35 According to International Patent Classification (IPC) or to both national classification and IPC				
<b>B. FIELDS SEARCHED</b>				
Minimum documentation searched (classification system followed by classification symbols) U.S. : 702/19; 536/1.11; 514/8; 424/9.35				
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched				
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet				
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>				
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
A	FUKUI et al. Oligosaccharide Microarrays for High-Throughput Detection and Specificity Assignments of Carbohydrate-Protein Interactions. Nature Biotechnology. 1999, Vol. 20, pages 1011-1017, see entire document.	1-12		
A	HIRABAYASHI. Lectin-Based Structural Glycomics: Glycoproteomics and Glycan Profiling. Glycoconjugate Journal. 2004, Vol. 21, pages 35-40, see entire document.	1-12		
A	KIKUCHI et al. The Carbohydrate Sequence Markup Language (CabosML): an XML Description of Carbohydrate Structures. Bioinformatics. 2005, Vol. 21, pages 1717-1718, see entire document.	1-12		
A	PETRESCU et al. A Statistical Analysis of N- and O-Glycan Linkage Conformations from Crystallographic Data. Glycobiology. 1999, Vol. 9, pages 343-352, see entire document.	1-12		
X	HASHIMOTO et al. KEGG as a Glycome Informatics Resource. Glycobiology. 2006 (published online on July 13, 2005), Vol. 16, pages 63r-70r, see entire document.	1-12		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.				
* Special categories of cited documents: <table border="0" style="width: 100%;"> <tr> <td style="vertical-align: top;">           "A" document defining the general state of the art which is not considered to be of particular relevance            "E" earlier application or patent published on or after the international filing date            "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)            "O" document referring to an oral disclosure, use, exhibition or other means            "P" document published prior to the international filing date but later than the priority date claimed         </td> <td style="vertical-align: top;">           "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention            "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone            "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art            "Z" document member of the same patent family         </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "Z" document member of the same patent family			
Date of the actual completion of the international search 16 August 2008 (16.08.2008)		Date of mailing of the international search report <b>22 SEP 2008</b>		
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201		Authorized officer  /LORI A. CLOW/  Telephone No. 571-272-5000		

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US07/18103

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RAMAN et al. Glycomics: an Integrated Systems Approach to Structure-Function Relationships of Glycans. Nature Methods. 2005, Vol. 2, pages 817-824, see entire document.	1-12

**INTERNATIONAL SEARCH REPORT****International application No.**  
**PCT/US07/18103**

Continuation of B. FIELDS SEARCHED Item 3:  
WEST; STN-Medline, Caplus, Biosis; scholar.google.com  
search terms: glycomics database; glycan; glycoprotein; glycolipid; binding

## フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LS,MW,MZ,NA,SD,SL,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HU,IE,IS,IT,LT,LU,LV,MC,MT,NL,PL,PT,RO,SE,SI,SK,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BH,BR,BW,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IS,JP,KE,KG,KM,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LT,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PG,PH,PL,PT,RO,RS,RU,SC,SD,SE,SG,SK,SL,SM,SV,SY,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,ZA,ZM,ZW

(72)発明者 サシセカラン, ラム

アメリカ合衆国 マサチューセッツ 02139, ケンブリッジ, マサチューセッツ アベニュー 77, 16-561

(72)発明者 ラグラム, エス.

アメリカ合衆国 ニュージャージー 08844, ヒルズバラ, フリート ドライブ 34

(72)発明者 ベンカタラマン, マーデバン

アメリカ合衆国 マサチューセッツ 02474, アーリントン, マサチューセッツ アベニュー 285, アpartment 37

(72)発明者 カウンディンヤ, サブラマニアン

アメリカ合衆国 マサチューセッツ 02474, アーリントン, マサチューセッツ アベニュー 155, アpartment 5

(72)発明者 ラマン, ラウル

アメリカ合衆国 マサチューセッツ 02139, ケンブリッジ, メモリアル ドライブ 540, アpartment 1610