SINGAPORE PATENTS ACT(CHAPTER 221)
PATENTS RULES

PATENTS FORM 1	Request for the Grant of a Patent
2001	

Pre-requisites:

a. *Among other requirements, a Date of Filing will be issued only if a description of the invention is filed or if a complete and valid statement is made under Part 9 of this form.*

Estimated Time:

This form may take approximately15 - 20minutes to complete.

General:

a. * *denotes mandatory field.*

PART 1 Reference

Applicant/Agent Reference	2200840/MIT/MK/ANG/ely

PART 2 Title of Invention*

Title of Invention*	HEMAGGLUTININ POLYPEPTIDES, AND REAGENTS AND METHODS RELATING THERETO

PART 3 Details of Applicant*

Note: If there is insufficient space, please use the continuation sheet CS 1.

Applicants 1

Type of Applicant	Corporate
UEN/ Company Code	E0169511A
Name	MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Address Type	Foreign
Address	77 MASSACHUSETTS AVENUE, CAMBRIDGE, MASSACHUSETTS 02139-4307, UNITED STATES OF AMERICA Massachusetts-United States of America
Is SME	
Country of Incorporation	United States of America
State of Incorporation	Massachusetts
Sole Proprietor or Partners' Names	

PART 4 Declaration of Priority under Section 17(2)

Declaration of Priority(s) under Section 17(2)

Application No.	Country	Filing Date
61/389,639	United States of America	04/10/2010

PART 5 Inventor*

A. Are all applicants named above also inventors? Yes ☐ No ☒

B. If yes, are there any other inventors?	Yes ☐ No ☐

PART 6 Claiming the Filing Date of an Earlier Singapore Application

This application is	☒ A divisional of an earlier Singapore application.
	☐ An application filed in response to an order by the Registrar after determination of a question regarding the entitlement of an earlier Singapore application. Application No. 10201709806V Filing Date *(DD/MM/YYYY)* 04/10/2011

PART 7 Disclosure of Invention Prior to Filing

☐ The invention was displayed at an international exhibition on a date later than the start of the period of 12 months before the date of filing this application.

PART 8 Micro-Organism Deposited under the Budapest Treaty

☐ The invention requires the use of a micro-organism and for disclosure purposes, a culture of the micro-organism has been deposited for with an international depository authority under the Budapest Treaty.

PART 9 Description Incorporated by Reference to an Earlier Relevant Application

☐ No description is filed with this request. Instead the description of the invention for which the patent is sought is incorporated in the application by reference to, and is completely contained in, the following earlier relevant application, as filed:

Country

Application No.

Filing Date *(DD/MM/YYYY)*

PART 10 Checklist*

Checklist A — The application consists of the following no. of sheets

Description	513	sheet(s)
Claim	5	sheet(s)
Drawing	54	sheet(s)
Abstract	1	sheet(s)

(Note:Please indicate on the abstract the figure of the drawing, if any, which should accompany the abstract when published)

Total number of sheets	573	sheet(s)

Checklist B — The application as filed is accompanied by:

	☒ Statement of inventorship & right to grant *(Patents Form 8)*
	☐ International exhibition certificate

PART 11 Contact Details*

Note:
a. This part need not be filled up if the person's address in Part 3 is to be used as an address for service in Singapore.
b. The email address and telephone number provided are for contact purposes in case of query.
c. The official correspondence will be sent electronically via IP2SG.

Singapore registered patent agent, or an advocate and solicitor, who has in force a practising certificate. Please refer to Singapore Patents Act Section 105 for details.

Agent UEN/Company Code	T14LL0024B
Agent Name	DONALDSON & BURKINSHAW LLP
Representative or C/O Name	
	Address for Service in Singapore
	Block/ House No.
	Street Name
	Level - Unit 3667
	Building Name ROBINSON ROAD POST OFFICE
	Postal Code 905667
Contact Person Direct Telephone No. Email Address	

PART 12 Declaration*

Declaration	**By Person Filing the Application** I do hereby declare that the information furnished above is true to the best of my knowledge. **By Agent** I do hereby declare that : i. I have been duly authorised to act as an agent, for the purposes of this application, on behalf of the person(s) filing this form. ii. The information furnished above on behalf of the person(s) filing this form is true to the best of the person(s)' knowledge.
Name	EVELINE ANGSTETRA

Signature		
Date *(DD/MM/YYYY)*	27/5/2022	
No. of Extra Sheets Attached to this Form	0	sheet(s)

GST Information

The tax invoice will be issued to:

Name: MASSACHUSETTS INSTITUTE OF TECHNOLOGY

HEMAGGLUTININ POLYPEPTIDES, AND REAGENTS AND METHODS RELATING THERETO

Cross-Reference to Related Applications

[0001] The present application claims the benefit of U.S. Provisional Application No. 61/389,639 filed October 4, 2010, the complete contents of which is incorporated herein by reference.

Government Support

[0002] This invention was made with government support under grant numbers GM57073 and U54 GM62116 awarded by National Institutes of Health. The United States government has certain rights in this invention.

Background

[0003] Influenza has a long history of pandemics, epidemics, resurgences and outbreaks. Avian influenza, including the H5N1 strain, is a highly contagious and potentially fatal pathogen, but it currently has only a limited ability to infect humans. However, avian flu viruses have historically observed to accumulate mutations that alter its host specificity and allow it to readily infect humans. In fact, two of the major flu pandemics of the last century originated from avian flu viruses that changed their genetic makeup to allow for human infection.

[0004] There is a significant concern that the current H5N1, H7N7, H9N2 and H2N2 avian influenza strains might accumulate mutations that alter their host specificity and allow them to readily infect humans. Therefore, there is a need to assess whether the HA protein in these strains can, in fact, convert to a form that can readily infect humans, and a further need to identify HA variants with such ability. There is a further need to understand the characteristics of HA proteins generally that allow or prohibit infection of different subjects, particularly humans. There is also a need for vaccines and therapeutic strategies for effective treatment or delay of onset of disease caused by influenza virus.

Summary

[0005] The present invention binding agents with particular glycan binding characteristics. In particular, the present invention provides binding agents that bind to sialylated glycans having an umbrella-like topology. In some embodiments, binding agents in accordance with the invention bind to umbrella-topology glycans with high affinity and/or specificity. In some embodiments, binding agents in accordance with the invention show a binding preference for umbrella-topology glycans as compared with cone-topology glycans. In some embodiments, binding agents in accordance with the invention compete with hemagglutinin for binding to glycans on hemagglutinin receptors. In some embodiments, binding agents in accordance with the invention compete with hemagglutinin for binding to umbrella-topology glycans.

[0006] The present invention also provides diagnostic and therapeutic reagents and methods associated with provided binding agents, including vaccines.

[0007] The present invention particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with altered glycosylation can show increased (or decreased) binding to human HA receptors as compared with a reference HA polypeptide. In some embodiments, the reference polypeptide is an HA polypeptide of any of the following:

[0008] **A/South Carolina/1/18 (H1):**

MEARLLVLLCAFAATNADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCKLKGI
APLQLGKCNIAGWLLGNPECDLLLTASSWSYIVETSNSENGTCYPGDFIDYEELREQLSSVSSF
EKFEIFPKTSSWPNHETTKGVTAACSYAGASSFYRNLLWLTKKGSSYPKLSKSYVNNKGKEVLV
LWGVHHPPTGTDQQSLYQNADAYVSVGSSKYNRRFTPEIAARPKVRDQAGRMNYYWTLLEPGDT
ITFEATGNLIAPWYAFALNRGSGSGIITSDAPVHDCNTKCQTPHGAINSSLPFQNIHPVTIGEC
PKYVRSTKLRMATGLRNIPSIQSRGLFGAIAGFIEGGWTGMIDGWYGYHHQNEQGSGYAADQKS
TQNAIDGITNKVNSVIEKMNTQFTAVGKEFNNLERRIENLNKKVDDGFLDIWTYNAELLVLLEN
ERTLDFHDSNVRNLYEKVKSQLKNNAKEIGNGCFEFYHKCDDACMESVRNGTYDYPKYSEESKL
NREEIDGVKLESMGVYQILAIYSTVASSLVLLVSLGAISFWMCSNGSLQCRICI (SEQ ID
NO: 43)

[0009] **A/Brisbane/59/07 (H1):**

MKVKLLVLLCTFTATYADTICIGYHANNSTDTVDTVLEKNVTVTHSVNLLENSHNGKLCLLKGI
APLQLGNCSVAGWILGNPECELLISKESWSYIVEKPNPENGTCYPGHFADYEELREQLSSVSSF

ERFEIFPKESSWPNHTVTGVSASCSHNGESSFYRNLLWLTGKNGLYPNLSKSYANNKEKEVLVL
WGVHHPPNIGDQKALYHTENAYVSVVSSHYSRKFTPEIAKRPKVRDQEGRINYYWTLLEPGDTI
IFEANGNLIAPRYAFALSRGFGSGIINSNAPMDKCDAKCQTPQGAINSSLPFQNVHPVTIGECP
KYVRSAKLRMVTGLRNIPSIQSRGLFGAIAGFIEGGWTGMVDGWYGYHHQNEQGSGYAADQKST
QNAINGITNKVNSVIEKMNTQFTAVGKEFNKLERRMENLNKKVDDGFIDIWTYNAELLVLLENE
RTLDFHDSNVKNLYEKVKSQLKNNAKEIGNGCFEFYHKCNDECMESVKNGTYDYPKYSEESKLN
REKIDGVKLESMGVYQILAIYSTVASSLVLLVSLGAISFWMCSNGSLQCRICI (SEQ ID
NO: 44)

[0010] A/California/04/09 (H1):

MKAILVVLLYTFATANADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCK
LRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELRE
QLSSVSSFERFEIFPKTSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSK
SYINDKGKEVLVLWGIHHPSTSADQQSLYQNADTYVFVGSSRYSKKFKPEIAIRPKVRDQ
EGRMNYYWTLVEPGDKITFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDCNTTCQTPK
GAINTSLPFQNIHPITIGKCPKYVKSTKLRLATGLRNIPSIQSRGLFGAIAGFIEGGWTG
MVDGWYGYHHQNEQGSGYAADLKSTQNAIDEITNKVNSVIEKMNTQFTAVGKEFNHLEKR
IENLNKKVDDGFLDIWTYNAELLVLLENERTLDYHDSNVKNLYEKVRSQLKNNAKEIGNG
CFEFYHKCDNTCMESVKNGTYDYPKYSEEAKLNREEIDGVKLESTRIYQILAIYSTVASS
LVLVVSLGAISFWMCSNGSLQCRICI (SEQ ID NO: 45)

[0011] A/Albany/6/58 (H2):

MAIIYLILLFTAVRGDQICIGYHANNSTEKVDTILERNVTVTHAKDILEKTHNGKLCKLNGIPP
LELGDCSIAGWLLGNPECDRLLSVPEWSYIMEKENPRDGLCYPGSFNDYEELKHLLSSVKHFEK
VKILPKDRWTQHTTTGGSRACAVSGNPSFFRNMVWLTKKGSNYPVAKGSYNNTSGEQMLIIWGV
HHPNDETEQRTLYQNVGTYVSVGTSTLNKRSTPDIATRPKVNGLGSRMEFSWTLLDMWDTINFE
STGNLIAPEYGFKISKRGSSGIMKTEGTLGNCETKCQTPLGAINTTLPFHNVHPLTIGECPKYV
KSEKLVLATGLRNVPQIESRGLFGAIAGFIEGGWQGMVDGWYGYHHSNDQGSGYAADKESTQKA
FDGITNRVNSVIEKMNTQFEAVGKEFSNLERRLENLNKKMEDGFLDVWTYNAELLVLMENERTL
DFHDSNVKNLYDKVKMQLRDNVKELGNGCFEFYPKCDDECMNSVKNGTYDYPKYEEESKLNRNE
IKGVKLSSMGVYQILAIYATVAGSLSLAIMMAGISFWMCSNGSLQCRICI (SEQ ID NO:
46)

[0012] A/Aichi/1/68 (H3):

MKTIIALSYIFCLALGQDLPGNDNSTATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQSSST

GKICNNPHRILDGIDCTLIDALLGDPHCDVFQNETWDLFVERSKAFSNCYPYDVPDYASLRSLV
ASSGTLEFITEGFTWTGVTQNGGSNACKRGPGSGFFSRLNWLTKSGSTYPVLNVTMPNNDNFDK
LYIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRPWVRGLSSRISIYWTIVKPG
DVLVINSNGNLIAPRGYFKMRTGKSSIMRSDAPIDTCISECITPNGSIPNDKPFQNVNKITYGA
CPKYVKQNTLKLATGMRNVPEKQTRGLFGAIAGFIENGWEGMIDGWYGFRHQNSEGTGQAADLK
STQAAIDQINGKLNRVIEKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALE
NQHTIDLTDSEMNKLFEKTRRQLRENAEEMGNGCFKIYHKCDNACIESIRNGTYDHDVYRDEAL
NNRFQIKGVELKSGYKDWILWISFAISCFLLCVVLLGFIMWACQRGNIRCNICI (SEQ ID
NO: 47)

[0013] A/Moscow/10/99 (H3):

MKTIIALSYILCLVFAQKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQ
SSSTGRICDSPHQILDGENCTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPD
YASLRSLVASSGTLEFNNESFNWTGVAQNGTSSSCKRRSIKSFFSRLNWLHQLKYRYPAL
NVTMPNNDKFDKLYIWGVHHPSTDSDQTSLYTQASGRVTVSTKRSQQTVIPNIGSRPWVR
GISSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITP
NGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWE
GMMDGWYGFRHQNSEGTGQAADLKSTQAAINQINGKLNRLIEKTNEKFHQIEKEFSEVEG
RIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLFERTRKQLRENAEDMGN
GCFKIYHKCDNACIGSIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWILWISFAISC
FLLCVVLLGFIMWACQKGNIRCNICI (SEQ ID NO: 48)

[0014] A/Perth/16/09 (H3):

MKTIIALSYILCLVFAQKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSST
GEICDSPHQILDGKNCTLIDALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDYASLRSLV
ASSGTLEFNNESFNWTGVTQNGTSSACIRRSKNSFFSRLNWLTHLNFKYPALNVTMPNNEQFDK
LYIWGVHHPGTDKDQIFLYAQASGRITVSTKRSQQTVSPNIGSRPRVRNIPSRISIYWTIVKPG
DILLINSTGNLIAPRGYFKIRSGKSSIMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGA
CPRYVKQNTLKLATGMRNVPEKQTRGIFGAIAGFIENGWEGMVDGWYGFRHQNSEGRGQAADLK
STQAAIDQINGKLNRLIGKTNEKFHQIEKEFSEVEGRIQDLEKYVEDTKIDLWSYNAELLVALE
NQHTIDLTDSEMNKLFEKTKKQLRENAEDMGNGCFKIYHKCDNACIGSIRNGTYDHDVYRDEAL
NNRFQIKGVELKSGYKDWILWISFAISCFLLCVALLGFIMWACQKGNIRCNICI (SEQ ID
NO: 49)

[0015] A/Vietnam/1203/04 (H5):

MEKIVLLFAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFE
KIQIIPKSSWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVNGQSGRMEFFWTILKPNDAIN
FESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESVRNGTYDYPQYSEEA
RLKREEISGVKLESIGIYQILSIYSTVASSLALAIMVAGLSLWMCSNGSLQCRICI (SEQ ID
NO: 50)

[0016] A/Egypt/2786-NAMRU3/06 (H5):

MEKIVLLLAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFE
KIQIIPKSSWSDHEASSGVSSACPYQGRSSFFRNVVWLIKKDNAYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKSNDAIN
FESNGNFIAPENAYKIVKKGDSTIMKSELEYGNCNTKCQTPIGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHRCDNECMESVRNGTYDYPQYSEEA
RLKREEISGVKLESIGTYQILSIYSTVASSLALAIMVAGLFLWMCSNGSLQCRICI (SEQ ID
NO: 51)

[0017] A/New York/107/03 (H7):

MNTQILAFIACVLTGVKGDKICLGHHAVANGTKVNTLTERGIEVVNATETVETTNIKKICTQGK
RPTDLGQCGLLGTLIGPPQCDQFLEFSSDLIIERREGTDICYPGRFTNEESLRQILRRSGGIGK
ESMGFTYSGIRTNGATSACTRSGSSFYAEMKWLLSNSDNAAFPQMTKAYRNPRNKPALIIWGVH
HSESVSEQTKLYGSGNKLITVRSSKYQQSFTPNPGARRIDFHWLLLDPNDTVTFTFNGAFIAPD
RTSFFRGESLGVQSDAPLDSSCRGDCFHSGGTIVSSLPFQNINSRTVGKCPRYVKQKSLLLATG
MRNVPEKPKPRGLFGAIAGFIENGWEGLINGWYGFRHQNAQGEGTAADYKSTQSAIDQITGKLN
RLIGKTNQQFELIDNEFNEIEQQIGNVINWTRDAMTEIWSYNAELLVAMENQHTIDLADSEMSK
LYERVKKQLRENAEEDGTGCFEIFHKCDDQCMESIRNNTYDHTQYRTESLQNRIQIDPVKLSSG
YKDIILWFSFGASCFLLLAIAMGLVFICIKNGNMQCTICI (SEQ ID NO: 52)

[0018] The present invention also particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with alterations in the HA loop region, can show increased (or decreased) binding to human HA receptors as compared with a reference HA polypeptide (including, for example, an HA polypeptide of any of SEQ ID NOs: 43-52).

[0019] In some embodiments, the present invention encompasses the recognition that H5 HA polypeptide variants with altered glycosylation can show increased (or decreased) binding to human HA receptors as compared with a reference HA polypeptide. In some embodiments, the reference polypeptide is an HA polypeptide of any of the following:

[0020] **A/Hongkong/486/97**

MEKIVLLLATVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILERTHNGKLCDLNGVK
PLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKASPANDLCYPGNFNDYEELKHLLSRINHFE
KIQIIPKSSWSNHDASSGVSSACPYLGRSSFFRNVVWLIKKNSAYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPEIATRPKVNGQSGRMEFFWTILKPNDAIN
FESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNTPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIINKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESVKNGTYDYPQYSEEA
RLNREEISGVKLESMGTYQILSIYSTVASSLALAIMVAGLSLWMCSNGSLQCRICI (SEQ ID
NO: 53)

[0021] **A/Hongkong/213/03**

MEKIVLLFAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPANDLCYPGDFNDYEELKHLLSRINHFE
KIQIIPKNSWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNNAYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQNGRMEFFWTILKPNDAIN
FESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESVRNGTYDYPQYSEEA
RLKREEISGVKLESIGTYQILSIYSTVASSLALAIMVAGLSLWMCSNGSLQCRICI (SEQ ID
NO: 54)

[0022] A/Vietnam/1203/04

MEKIVLLFAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKKHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPVNDLCYPGDFNDYEELKHLLSRINHFE
KIQIIPKSSWSSHEASLGVSSACPYQGKSSFFRNVVWLIKKNSTYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPRIATRSKVNGQSGRMEFFWTILKPNDAIN
FESNGNFIAPEYAYKIVKKGDSTIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQRERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESVRNGTYDYPQYSEEA
RLKREEISGVKLESIGIYQILSIYSTVASSLALAIMVAGLSLWMCSNGSLQCRICI (SEQ ID
NO: 50)

[0023] A/Indonesia/5/05

MEKIVLLLAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFINVPEWSYIVEKANPTNDLCYPGSFNDYEELKHLLSRINHFE
KIQIIPKSSWSDHEASSGVSSACPYLGSPSFFRNVVWLIKKNSTYPTIKKSYNNTNQEDLLVLW
GIHHPNDAAEQTMLYQNPTTYISIGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKPNDAIN
FESNGNFIAPEYAYKIVKKGDSAIMKSELEYGNCNTKCQTPMGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQRESRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHKCDNECMESIRNGTYNYPQYSEEA
RLKREEISGVKLESIGTYQILSIYSTVASSLALAIMMAGLSLWMCSNGSLQCRICI (SEQ ID
NO: 55)

[0024] A/Egypt/2786-NAMRU3/06

MEKIVLLLAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDLDGVK
PLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKINPANDLCYPGNFNDYEELKHLLSRINHFE
KIQIIPKSSWSDHEASSGVSSACPYQGRSSFFRNVVWLIKKDNAYPTIKRSYNNTNQEDLLVLW
GIHHPNDAAEQTRLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKSNDAIN
FESNGNFIAPENAYKIVKKGDSTIMKSELEYGNCNTKCQTPIGAINSSMPFHNIHPLTIGECPK
YVKSNRLVLATGLRNSPQGERRRKKRGLFGAIAGFIEGGWQGMVDGWYGYHHSNEQGSGYAADK
ESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFNNLERRIENLNKKMEDGFLDVWTYNAELLVLM
ENERTLDFHDSNVKNLYDKVRLQLRDNAKELGNGCFEFYHRCDNECMESVRNGTYDYPQYSEEA
RLKREEISGVKLESIGTYQILSIYSTVASSLALAIMVAGLFLWMCSNGSLQCRICI (SEQ ID

NO: 51)

[0025] The present invention also particularly encompasses the recognition that H5 HA polypeptide variants with alterations in the HA loop region, can show increased (or decreased) binding to human HA receptors as compared with a reference HA polypeptide (including, for example, an HA polypeptide of any of SEQ ID NO: 50, 51, and 53-55).

Brief Description of the Drawings

[0026] *Figure 1A-1C. Alignment of exemplary sequences of wild type HA.* Sequences were obtained from the NCBI influenza virus sequence database (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html). H1_Av (SEQ ID NO: 1). H1_Hu1 (SEQ ID NO: 2). H1_Hu2 (SEQ ID NO: 3). H2_Av (SEQ ID NO: 4). H2_Hu (SEQ ID NO: 5). H3_Av (SEQ ID NO: 6). H3_Hu1 (SEQ ID NO: 7). H3_Hu2 (SEQ ID NO: 8). H4_Av (SEQ ID NO: 9). H5_Av1 (SEQ ID NO: 10). H5_Av2 (SEQ ID NO: 11). H6_Av (SEQ ID NO: 12). H7_Av (SEQ ID NO: 13). H8_Av (SEQ ID NO: 14). H9_Av (SEQ ID NO: 15). H10_Av (SEQ ID NO: 16). H11_Av (SEQ ID NO: 17). H12_Av (SEQ ID NO: 18). H13_Av (SEQ ID NO: 19). H14_Av (SEQ ID NO: 20). H15_Av (SEQ ID NO: 21). H16_Av (SEQ ID NO: 22).

[0027] *Figure 2A-B. Sequence alignment of HA glycan binding domain.* Gray: conserved amino acids involved in binding to sialic acid. Red: particular amino acids involved in binding to Neu5Acα2-3/6Gal motifs. Yellow: amino acids that influence positioning of Q226 (137, 138) and E190 (186, 228). Green: amino acids involved in binding to other monosaccharides (or modifications) attached to Neu5Acα2-3/6Gal motif. The sequence for ASI30, APR34, ADU63, ADS97 and Viet04 were obtained from their respective crystal structures. The other sequences were obtained from SwissProt (http://us.expasy.org). Abbreviations: ADA76, A/duck/Alberta/35/76 (H1N1) (SEQ ID NO: 23); ASI30, A/Swine/Iowa/30 (H1N1) (SEQ ID NO: 24); APR34, A/Puerto Rico/8/34 (H1N1) (SEQ ID NO: 25); ASC18, A/South Carolina/1/18 (H1N1) (SEQ ID NO: 26); AT91, A/Texas/36/91 (H1N1) (SEQ ID NO: 27); ANY18, A/New York/1/18 (H1N1) (SEQ ID NO: 28); ADU63, A/Duck/Ukraine/1/63 (H3N8) (SEQ ID NO: 29); AAI68, A/Aichi/2/68 (H3N2) (SEQ ID NO: 30); AM99, A/Moscow/10/99 (H3N2) (SEQ ID NO: 31); ADS97, A/Duck/Singapore/3/97 (H5N3) (SEQ ID NO: 32); Viet04, A/Vietnam/1203/2004 (H5N1) (SEQ ID NO: 33).

[0028] *Figure 3. Sequence alignment illustrating conserved subsequences characteristic of H1 HA.* Figure 3A presents the same alignment that was presented in Figure 1A, except that Figure 3A indicates the presence of an additional conserved subsequence. Figure 3B presents the same alignment that was presented in Figure 1C, except that Figure 3A indicates the presence of an additional conserved subsequence.

[0029] *Figure 4. Sequence alignment illustrating conserved subsequences characteristic of H3 HA.* Figure 4A presents the same alignment that was presented in Figure 1A, except that Figure 4A indicates the presence of an additional conserved subsequence. Figure 4B presents the same alignment that was presented in Figure 1C, except that Figure 4A indicates the presence of an additional conserved subsequence.

[0030] *Figure 5A-1. Sequence alignment illustrating conserved subsequences characteristic of H5 HA.* Figure 5A-1 presents the same alignment that was presented in Figure 1A, except that Figure 5A-1 indicates the presence of an additional conserved subsequence.

[0031] Figure 5A-2 presents the same alignment that was presented in Figure 1C, except that Figure 5A-2 indicates the presence of an additional conserved subsequence.

[0032] *Figure 5B1-B5* presents additional H5 HA sequence alignments. Consensus (SEQ ID NO: 34); AAL59142 (SEQ ID NO: 35); AAZ29963 (SEQ ID NO: 36); ABA70758 (SEQ ID NO: 37); ABB87042 (SEQ ID NO: 38); ABD14810 (SEQ ID NO: 39); ABD46740 (SEQ ID NO: 40); ABD85144 (SEQ ID NO: 41); and ABE97569 (SEQ ID NO: 42).

[0033] *Figure 5B-6* shows additional avian 4 (HA) H5N1 viral strains.

[0034] *Figure 6. Framework for understanding glycan receptor specificity.* α2-3- and/or α2-6-linked glycans can adopt different topologies. According to the present invention, the ability of an HA polypeptide to bind to certain of these topologies confers upon it the ability to mediate infection of different hosts, for example, humans. As illustrated in Panel A of this figure, the present invention defines two particularly relevant topologies, a "cone" topology and an "umbrella" topology. The cone topology can be adopted by α2-3- and/or α2-6-linked glycans, and is typical of short oligosaccharides or branched oligosaccharides attached to a core (although this topology can be adopted by certain long oligosaccharides). The umbrella topology can only be adopted by α2-6-linked glycans (presumably due to the increased conformational plurality afforded by the extra C5-C6 bond that is present in the α2-6 linkage), and is predominantly adopted by long oligosaccharides or branched glycans with long oligosaccharide

branches, particularly containing the motif Neu5Acα2-6Galβ1-3/4GlcNAc-. As described herein, ability of HA polypeptides to bind the umbrella glycan topology, confers binding to human receptors and/or ability to mediate infection of humans. Panel B of this Figure specifically shows the topology of α2-3 and α2-6 as governed by the glycosidic torsion angles of the trisaccharide motifs – Neu5Acα2-3Galβ1-3/4GlcNAc and Neu5Acα2-6Galβ1-4GlcNAc respectively. A parameter (θ) – angle between C2 atom of Neu5Ac and C1 atoms of the subsequent Gal and GlcNAc sugars in these trisaccharide motifs was defined to characterize the topology. Superimposition of the θ contour and the conformational maps of the α2-3 and α2-6 motifs shows that α2-3 motifs adopt 100% *cone*-like topology and α2-6 motifs sampled both *cone*-like and *umbrella*-like topologies (Panel C). In the *cone*-like topology sampled by α2-3 and α2-6, GlcNAc and subsequent sugars are positioned along a region spanning a cone. Interactions of HA with *cone*-like topology primarily involve contacts of amino acids at the numbered positions (based on H3 HA numbering) with Neu5Ac and Gal sugars. On the other hand, in *umbrella*-like topology, which is unique to α2-6, \ GlcNAc and subsequent sugars bend towards the HA binding site (as observed in HA- α2-6 co-crystal structures). Longer α2-6 oligosaccharides (*e.g.* at least a tetrasaccharide) would favor this conformation since it is stabilized by intra-sugar van der Waals contact between acetyl groups of GlcNAc and Neu5Ac. HA interactions with *umbrella*-like topology involve contacts of amino acids at the numbered positions (based on H3 HA numbering) with GlcNAc and subsequent sugars in addition to contacts with Neu5Ac and Gal sugars. Panel C of this Figure depicts conformational sampling of cone- and umbrella-like topology by α2-3 and α2-6. Sections (A) – (D) show the conformational (ϕ, ψ) maps of Neu5Acα2-3Gal, Neu5Acα2-6Gal, Galβ1-3GlcNAc, and Galβ1-4GlcNAc linkages, respectively. These maps obtained from GlycoMaps DB (http://www.glycosciences.de/modeling/glycomapsdb/) were generated using *ab initio* MD simulations using MM3 force field. Energy distribution is color coded starting from *dark* (representing highest energy) to *light* representing lowest energy. Encircled regions 1 – 5 represent (ϕ,ψ) values observed for the α2-3 and α2-6 oligosaccharides in the HA-glycan co-crystal structures. The *trans* conformation (encircled region 1) of Neu5Acα2-3Gal predominates in HA binding pocket with the exception of the co-crystal structure of A/Aichi/2/68 H3N2 HA with α2-3 where this conformation is gauche (encircled region 2). On the other hand, the *cis* conformation of Neu5Acα2-6Gal (encircled region 3) predominates in HA binding pocket. The

cone-like topology is sampled by encircled regions 1 and 2 and the *umbrella*-like topology is sampled by encircled region 3. Sections (E) – (F) show sampling of *cone*-like and *umbrella-like* topologies by α2-3 and α2-6 motifs, respectively. The darker regions in the conformational maps were used as the outer boundaries to calculate the θ parameter (angle between C2 atom of Neu5Ac and C1 atoms of subsequent Gal and GlcNAc sugars) for a given set of (ϕ,ψ) values. Based on the energy cutoff, the value of $\theta > 110°$ was used to characterize *cone*-like topology and $\theta < 100°$ was used to characterize *umbrella*-like topology. Superimposition of the θ contour with the conformational energy map indicated that α2-3 motif adopts 100% *cone*-like topology since it was energetically unfavorable to adopt *umbrella*-like topology. On the other hand, the α2-6 motif sampled both the *cone*-like and *umbrella*-like topologies and this sampling was classified based on the ω angle (O-C6-C5-H5) of Neu5Acα2-6Gal linkage.

[0035] *Figure 7. Interactions of HA residues with cone vs. umbrella glycan topologies.* Analysis of HA-glycan co-crystals reveals that the position of Neu5Ac relative to the HA binding site is almost invariant. Contacts with Neu5Ac involve highly conserved residues such as F98, S/T136, W153, H183 and L/I194. Contacts with other sugars involve different residues, depending on whether the sugar linkage is α2-3 or α2-6 and whether the glycan topology is cone or umbrella. For example, in the cone topology, the primary contacts are with Neu5Ac and with Gal sugars. E190 and Q226 play particularly important roles in this binding. This Figure also illustrates other positions (*e.g.*, 137, 145, 186, 187, 193, 222) that can participate in binding to cone structures. In some cases, different residues can make different contacts with different glycan structures. The type of amino acid in these positions can influence ability of an HA polypeptide to bind to receptors with different modification and/or branching patterns in the glycan structures. In the umbrella topology, contacts are made with sugars beyond Neu5Ac and Gal. This Figure illustrates residues (*e.g.*, 137, 145, 156, 159, 186, 187, 189, 190, 192, 193, 196, 222, 225, 226) that can participate in binding to umbrella structures. In some cases, different residues can make different contacts with different glycan structures. The type of amino acid in these positions can influence ability of an HA polypeptide to bind to receptors with different modification and/or branching patterns in the glycan structures. In some embodiments, a D residue at position 190 and/or a D residue at position 225 contribute(s) to binding to umbrella topologies.

[0036] *Figure 8. Exemplary cone topologies.* This Figure illustrates certain exemplary (but not exhaustive) glycan structures that adopt cone topologies.

[0037] *Figure 9. Exemplary umbrella topologies.* (A) Certain exemplary (but not exhaustive) N- and O-linked glycan structures that can adopt *umbrella* topologies. (B) Certain exemplary (but not exhaustive) O-linked glycan structures that can adopt umbrella topologies.

[0038] *Figure 10A-B. Glycan profile of human bronchial epithelial cells and human colonic epithelial cells.* To further investigate the glycan diversity in the upper respiratory tissues, N-linked glycans were isolated from HBEs (a representative upper respiratory cell line) and analyzed using MALDI-MS. The predominant expression of α2-6 in HBEs was confirmed by pre-treating the sample with Sialidase S (α2-3 specific) and Sialidase A (cleaves and SA). The predominant expression of glycans with long branch topology is supported by TOF-TOF fragmentation analysis of representative mass peaks. To provide a reference for glycan diversity in the upper respiratory tissues, the N-linked glycan profile of human colonic epithelial cells (HT29; a representative gut cell line) was obtained. This cell line was chosen because the current H5N1 viruses have been shown to infect gut cells. Sialidase A and S pre-treatment controls showed predominant expression of α2-3 glycans in the HT-29 cells. Moreover, the long branch glycan topology is not as prevalent as observed for HBEs. Therefore, human adaptation of the H5N1 HA would involve HA mutations that would enable high affinity binding to the diverse glycans expressed in the human upper respiratory tissues (*e.g.*, umbrella glycans).

[0039] *Figure 11. Conformational map and solvent accessibility of Neu5Acα2-3Gal and Neu5Acα2-6Gal motifs.* **Panel A** shows the conformational map of Neu5Acα2-3Gal linkage. The encircled region 2 is the *trans* conformation observed in the APR34_H1_23, ADU63_H3_23 and ADS97_H5_23 co-crystal structures. The encircled region 1 is the conformation observed in the AAI68_H3_23 co-crystal structure. **Panel B** shows the conformational map of Neu5Acα2-6Gal where the *cis*-conformation (encircled region 3) is observed in all the HA–α2-6 sialylated glycan co-crystal structures. **Panel C** shows difference between solvent accessible surface area (SASA) of Neu5Ac α2-3 and α2-6 sialylated oligosaccharides in the respective HA-glycan co-crystal structures. The red and cyan bars respectively indicate that Neu5Ac in α2-6 (positive value) or α2-3 (negative value) sialylated glycans makes more contact with glycan binding site. **Panel D** shows difference between SASA of NeuAc in α2-3 sialylated glycans bound to swine and human H1 ($H1_{\alpha 2\text{-}3}$), avian and human

H3 ($H3_{\alpha 2\text{-}3}$), and of NeuAc in α2-6 sialylated glycans bound to swine and human H1 ($H1_{\alpha 2\text{-}6}$). The negative bar for $H3_{\alpha 2\text{-}3}$ indicates lesser contact of the human H3 HA with Neu5Acα2-3Gal compared to that of avian H3. Torsion angles – ϕ: C2-C1-O-C3 (for Neu5Acα2-3/6 linkage); ψ: C1-O-C3-H3 (for Neu5Acα2-3Gal) or C1-O-C6-C5 (for Neu5Acα2-6Gal); ω: O-C6-C5-H5 (for Neu5Acα2-6Gal) linkages. The ϕ, ψ maps were obtained from GlycoMaps DB (http://www.glycosciences.de/modeling/glycomapsdb/) which was developed by Dr. Martin Frank and Dr. Claus-Wilhelm von der Lieth (German Cancer Research Institute, Heidelberg, Germany). The coloring scheme from high energy to low energy is from bright red to bright green, respectively.

[0040] *Figure 12A-B. Lectin staining of upper respiratory tissue sections.* A co-stain of the tracheal tissue with Jacalin (*lighter*) and ConA (*darker*) reveals a preferential binding of Jacalin (binds specifically to O-linked glycans) to goblet cells on the apical surface of the trachea and ConA (binds specifically to N-linked glycans) to the ciliated tracheal epithelial cells. Without wishing to be bound by any particular theory, we note that this finding suggests that goblet cells predominantly express O-linked glycans while ciliated epithelial cells predominantly express N-linked glycans. Co-staining of trachea with Jacalin and SNA (*dark*; binds specifically to α2-6) shows binding of SNA to both goblet and ciliated cells. On the other hand, co-staining of Jacalin (*lighter)* and MAL (*darker*), which specifically binds to α2-3 sialylated glycans, shows weak minimal to no binding of MAL to the pseudostratified tracheal epithelium but extensive binding to the underlying regions of the tissue. Together, the lectin staining data indicated predominant expression and extensive distribution of α2-6 sialylated glycans as a part of both N-linked and O-linked glycans respectively in ciliated and goblet cells on the apical side of the tracheal epithelium.

[0041] *Figure 13*: *Comparison of RBS of H5N1 HAs across different genetic clades.* A/Hong Kong/486/97, Genetic Clade 0 (HK_486_97_c0) (SEQ ID NO: 76); A/Duck/Hunan/795/02, Genetic Clade 2.1.1 (DK_Hunan_795_02_c.2.1.1) (SEQ ID NO: 77); A/Hong Kong/213/03, Genetic Clade 1 (Hk_213_03_c1) (SEQ ID NO: 78); A/Vietnam/1194/04, Genetic Clade 1 (Viet_1194_04_c1) (SEQ ID NO: 79); A/Vietnam/1203/04, Genetic Clade 1 (Viet_1203_04_c1) (SEQ ID NO: 80); A/Indonesia/5/05, Genetic Clade 2.1.3 (Ind_5_05_c2.1.3) (SEQ ID NO: 81); A/Anhui/1/05, Genetic Clade 2.3.4 (Anhui_1_05_c2.3.4) (SEQ ID NO: 82); A/Egypt/2786-NAMRU3/06, Genetic Clade 2.2 (Egypt_2876-N3_06_c2.2) (SEQ ID NO: 83);

A/goose/Guiyang/337/06, Genetic Clade 4 (Go_Guiy_337_06_c4) (SEQ ID NO: 84); A/Egypt/2321-NAMRU3/07, Genetic Clade 2.2.1 (Egypt_2321-N3_07_c2.2.1) (SEQ ID NO: 85); A/Egypt/3300-NAMRU3/08, Genetic Clade 2.2.1 (Egypt_3300-N3_08_c.2.2.1) (SEQ ID NO: 86); A/common magpie/Hong Kong/5052/07, Genetic Clade 2.3.2 (Mag_HK_5052_07_c.2.3.2) (SEQ ID NO: 87); A/Chicken/Vietnam/NCVD-016/2008, Genetic Clade 7 (Ck_Viet_NCVD-016_08_c7) (SEQ ID NO: 88).

[0042] *Figure 14*: Adapted from Stevens *et al.*, 2008, *J. Mol. Biol.*, 381:1382-94. *Viet0304*: A/Vietnam/1203/04. *Ind505*: A/Indonesia/5/05. *LS*: *Q226L and G228S* mutations RBS of HA.

[0043] *Figure 15*: *Viet0304-LS* and *Viet0304-RLS* (equivalent to *Ind505-LS*) HAs were analyzed in a dose-dependent fashion on our glycan array comprising of representative avian and human receptors. Both mutants showed minimal α2-6 binding (binding signals observed only at high HA concentration), which contrasts with the high affinity α2-6 binding shared by human adapted HAs.

[0044] *Figure 16: Dose dependent analysis of H5N1 HAs that lack glycosylation at N158 in the context of* LS *mutation.* Removal of glycosylation at N158 in the context of *LS* mutation increased human receptor binding affinity to the same range (Kd' ~ 20pM) as observed for human adapted H1N1 and H2N2 HAs. Removal of glycosylation at N158 on H5 HA template with Q226L mutation alone (*i.e.*, without G228S of the *LS* mutation) showed improved preference to human receptors (over avian receptors), but substantially lowered human receptor-binding affinity in comparison with N158-deglycosylated *LS* mutant.

[0045] *Figure 17: Comparing salient features of RBS between human adapted H2N2 HA (Alb6_58) and H5N1 (Viet1203_04) HA.* Four differences in H2N2 HA as compared with H5N1 HA include: (1) Composition of 130 loop in H2N2 HA which includes a deletion, (2) lack of glycosylation at 158 position in H2N2 HA, (3) amino acid composition of the base of the RBS involving positions 137, 221, 226 and 228, and (4) amino acid composition at top of the RBS involving positions 188, 192 and 193.

[0046] *Figure 18*: *Comparison of the RBS of the H5 HA mutants with that of H2N2 HA.* A1b6_58_H2N2 (SEQ ID NO: 89): A/Albany/6/58 H2N2 HA; Viet1203_04_D (SEQ ID NO: 90): modified version of A/Vietnam/1203/04 HA; Viet1203_04_D_H2RBS (SEQ ID NO: 91): mutant of Viet1203_04_D with deletion at 130 and 13 amino acid substitutions; Viet1203_04_D_H2RBSmin (SEQ ID NO: 92): mutant of Viet1203_04_D with deletion at 130

and 7 substitutions; ckEgy_07 (SEQ ID NO: 93): A/chicken/Egypt/R2/2007 H5N1 HA that already has deletion at 130; ckEgy_07_H2RBS (SEQ ID NO: 94): mutant of ckEgy_07 with 8 substitutions; ckViet_08 (SEQ ID NO: 95): A/chicken/Vietnam/NCVD-093/2008 H5N1 HA that already has switch in charge at 192 and 193 positions; ckViet_08_H2RBS (SEQ ID NO: 96): mutant of ckViet_08 with deletion at 130 and 6 substitutions; ckViet_08H2RBSmin (SEQ ID NO: 97): mutant of ckViet_08 with deletion at 130 and 6 substitutions; ckViet_08_H2RBSmin (SEQ ID NO: 98): mutant of ckViet_08 with deletion at 130 and 4 substitutions. The residue positions that are substituted are in bold and highlighted. The deletion in 130 loop is shown in bold highlighted. Glycosylation at 158 is highlighted.

[0047] *Figure 19: Dose dependent analysis of mutant of H5N1 HA* (Viet1203_04_D_H2RBS) *designed such that the molecular composition of its RBS mimics that of human-adapted H2N2 HA.* This mutant shows highly specific high affinity binding to human receptors that is characteristic of other human adapted HAs. The binding affinity of this mutant to human receptor (6'SLN-LN) is quantified by *Kd'* ~ 3 pM that is in the same range as that of human adapted H1N1 and H2N2 HAs.

[0048] *Figure 20: Characteristic glycan-receptor binding properties of pandemic HAs* The HAs from prototypic human-adapted pandemic 1918 H1N1 (*S1*) and 1958 H2N2 (*S2*) and 2009 H1N1 (*S3*) strains show specific high affinity binding to human receptors (6'SLN-LN) with minimal to substantially lower affinity binding (relative to human receptor affinity) to avian receptors (3'SLN-LN). On the other hand introducing the hallmark *LS* mutation in *Viet1203_04* H5N1 HA sequence does not switch its receptor preference to the human receptor binding seen with the pandemic HAs.

[0049] *Figure 21: Phylogeny tree of representative sequences from HA subtypes* Branches leading to clade 1 & 2 HAs are labeled and colored in red and blue, respectively. Closely related subtypes are located on branches close to one another.

[0050] *Figure 22: Key structural features within the RBS of H5 HA* Shown in the figure is the cartoon rendering of RBS of *Alb6_58* HA (*gray*) and *Viet0304* (*green*) with side chains of amino acids that are different between these HAs. The four features that distinguish the H2 and H5 RBS as described in the text are highlighted in dotted *red* circles.

[0051] *Figure 23: Sequence alignment of RBS of representative H2 and H5 HAs*

HA sequences from the pandemic H2N2 strain (A/Albany/6/58 or Alb58) (SEQ ID NO: 98), representative human isolates from 1997-2006 (A/Hong Kong/486/97 or HK_486_97 (SEQ ID NO: 105), A/Hong Kong/213/03 or HK_213_03 (SEQ ID NO: 104), A/Vietnam/1203/04 or Viet1203_04 (SEQ ID NO: 103), A/Indonesia/5/05 or Ind_5_05 (SEQ ID NO: 102), A/Egypt/2786-NAMRU3 or Egy_2786-N3_06 (SEQ ID NO: 101)) along with the chosen H5 HA template (A/chicken/Egypt/R2/07 or ckEgy_07 (SEQ ID NO: 100)) for introducing *LS* mutation (ckEgy_07mutv5.3 (SEQ ID NO: 99)) are aligned.

[0052] *Figure 24: Glycan receptor-binding properties of ckEgy_07 and ckEgy_07 harboring LS amino acid changes*

Dose-dependent direct glycan-binding of HAs were performed on a glycan array platform comprising of representative human and avian receptors. **A**, the wild-type *ckEgy_07* HA shows the typical specific and high affinity avian-receptor binding characteristic of other wild-type H5N1 HAs. **B**, Introduction of the *LS* mutations on this HA quantitatively shifts its specificity to human receptor (6'SLN-LN) and substantially reducing its avian receptor binding to a minimal level. **C**, binding of *ckEgy_07 LS* mutant to physiological human receptors expressed predominantly on apical surface of human tracheal tissue section. The human receptor specificity and affinity from the dose-dependent binding profile together with the human tracheal tissue staining of the mutant HA is such that it may be sufficient to confer aerosol transmission of H5N1 virus harboring this mutant HA in the context of other required changes (such as PB2).

[0053] *Figure 25: Analysis of effect of loss of glycosylation at the 158 position in the context of LS mutations on glycan receptor binding of H5 HA*

A, the *ckEgy_07_LS* mutant shows a quantitative human receptor switch resembling other pandemic HAs. **B,** the binding curve for A/California/04/2009 H1N1 HA adapted from previous study (*S3*) is shown for comparison. **C**, the T160A mutation in *Viet03_04_ALS* mutant removes glycosylation sequon at N158 leading to loss of glycosylation at this site. Although this mutant shows dramatic improvement in human receptor-binding, it retains most of its avian receptor-binding that is not characteristic of pandemic HAs and the *ckEgy_07_LS* mutant (in the *top* panel). **D**, *LS* amino acid mutations introduced in *Egy_06* HA sequence that naturally lacks glycosylation at 158 also shows the same binding profile as *Viet03_04_ALS* mutant.

[0054] *Figure 26: Human-adaptive amino acid changes on H5N1 HA sequence that naturally acquired feature 2*

A/chicken/Vietnam/NCVD-093/2008 avian H5N1 HA already acquired amino acid changes in 190-helix where 192 position typically comprising of Thr has mutated to Lys and 193 position typically comprising of Lys/Arg has mutated to Met. Introducing 6 amino acid changes and a deletion to match feature 1 (deletion in 130-loop + A130T) and feature 3 (S137R/S221P/Q226L/S227G/G228S) resulted in a mutant HA that quantitatively switched its preference to human receptors even in the presence of glycosylation at 158. However, introduction of the T160A loss of glycosylation change along with *LS*, without the 130-loop deletion results in dramatic reduction in binding to human and avian receptors (data not shown). Therefore, for these H5N1 HAs, the deletion in 130-loop is a more critical change than the loss of glycosylation in the context of *LS* mutation.

[0055] *Figure 27: Emergence of key features in recent avian H5 and human H5N1 isolates* **A,** Percentage of avian and human H5N1 isolates whose HA has acquired amino acid changes to match features 1 and 4 of H2 HA RBS is plotted as function of year when the strain was isolated. There has been a dramatic increase in percentage of isolates having these key features since their initial emergence in 2007. Phylogenetic analysis of the sequences of these isolates showed that they belonged to clade 2.2.1. **B**, Percentage of avian and human isolates whose HA has acquired amino acid changes to match feature 2 of H2 HA RBS. Only a small percentage of H5N1 isolates have acquired this feature. Full-length non-redundant HA sequences from NCBI Influenza Virus Resource were aligned, and number of occurrences of each of the features was calculated in a given year and expressed as percentage. A total of 2277 full-length non-redundant H5N1 sequences were employed for the analysis.

[0056] *Figure 28: Expanded nomenclature of glycans used in the glycan array* Neu5Ac: N-acetyl D-neuraminic acid; Gal: D-galactose; GlcNAc: N-acetyl D-glucosamine. α / β: anomeric configuration of the pyranose sugars. All the sugars are linked via a spacer to biotin (-Sp-LC-LC-Biotin as described in http://www.functionalglycomics.org/static/consortium/resources/resourcecored5.shtml)

Description of HA Sequence Elements

HA Sequence Element 1

[0057] *HA Sequence Element 1* is a sequence element corresponding approximately to residues 97-185 (where residue positions are assigned using H3 HA as reference) of many HA proteins found in natural influenza isolates. This sequence element has the basic structure:

C (Y/F) P X_1 C X_2 W X_3 W X_4 H H P (SEQ ID NO.106), wherein:

X_1 is approximately 30-45 amino acids long;

X_2 is approximately 5-20 amino acids long;

X_3 is approximately 25-30 amino acids long; and

X_4 is approximately 2 amino acids long.

[0058] In some embodiments, X_1 is about 35-45, or about 35-43, or about 35, 36, 37, 38, 38, 40, 41, 42, or 43 amino acids long. In some embodiments, X_2 is about 9-15, or about 9-14, or about 9, 10, 11, 12, 13, or 14 amino acids long. In some embodiments, X_3 is about 26-28, or about 26, 27, or 28 amino acids long. In some embodiments, X_4 has the sequence (G/A) (I/V). In some embodiments, X_4 has the sequence GI; in some embodiments, X_4 has the sequence GV; in some embodiments, X_4 has the sequence AI; in some embodiments, X_4 has the sequence AV. In some embodiments, HA Sequence Element 1 comprises a disulfide bond. In some embodiments, this disulfide bond bridges residues corresponding to positions 97 and 139 (based on the canonical H3 numbering system utilized herein).

[0059] In some embodiments, and particularly in H1 polypeptides, X_1 is about 43 amino acids long, and/or X_2 is about 13 amino acids long, and/or X_3 is about 26 amino acids long. In some embodiments, and particularly in H1 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} T (A/T) (A/S) C X_2 W X_3 W X_4 H H P (SEQ ID NO.107), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 26-41, or approximately 31-41, or approximately 31-39, or approximately 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long, and X_2-X_4 are as above.

[0060] In some embodiments, and particularly in H1 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} T (A/T) (A/S) C X_2 W (I/L) (T/V) X_{3A} W X_4 H H P (SEQ ID NO.108), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long,

X_{3A} is approximately 23-28, or approximately 24-26, or approximately 24, 25, or 26 amino acids long, and X_2 and X_4 are as above.

[0061] In some embodiments, and particularly in H1 polypeptides, HA Sequence Element 1 includes the sequence:

Q L S S I S S F E K (SEQ ID NO.109),

typically within X_1, (including within X_{1A}) and especially beginning about residue 12 of X_1 (as illustrated, for example, in **Figures 1-3**).

[0062] In some embodiments, and particularly in H3 polypeptides, X_1 is about 39 amino acids long, and/or X_2 is about 13 amino acids long, and/or X_3 is about 26 amino acids long.

[0063] In some embodiments, and particularly in H3 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} S (S/N) (A/S) C X_2 W X_3 W X_4 H H P (SEQ ID NO.110), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 23-38, or approximately 28-38, or approximately 28-36, or approximately 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long, and X_2-X_4 are as above.

[0064] In some embodiments, and particularly in H3 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} S (S/N) (A/S) C X_2 W L (T/H) X_{3A} W X_4 H H P (SEQ ID NO.111), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long,

X_{3A} is approximately 23-28, or approximately 24-26, or approximately 24, 25, or 26 amino acids long, and X_2 and X_4 are as above.

[0065] In some embodiments, and particularly in H3 polypeptides, HA Sequence Element 1 includes the sequence:

(L/I) (V/I) A S S G T L E F (SEQ ID NO.112),

typically within X_1 (including within X_{1A}), and especially beginning about residue 12 of X_1 (as illustrated, for example, in **Figures 1, 2** and **4)**.

[0066] In some embodiments, and particularly in H5 polypeptides, X_1 is about 42 amino acids long, and/or X_2 is about 13 amino acids long, and/or X_3 is about 26 amino acids long.

[0067] In some embodiments, and particularly in H5 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} S S A C X_2 W X_3 W X_4 H H P (SEQ ID NO.113), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 23-38, or approximately 28-38, or approximately 28-36, or approximately 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long, and X_2-X_4 are as.

[0068] In some embodiments, and particularly in H5 polypeptides, HA Sequence Element 1 has the structure:

C Y P X_{1A} S S A C X_2 W L I X_{3A} W X_4 H H P (SEQ ID NO.114), wherein:

X_{1A} is approximately 27-42, or approximately 32-42, or approximately 32-40, or approximately 32, 33, 34, 35, 36, 37, 38, 39, or 40 amino acids long, and

X_{3A} is approximately 23-28, or approximately 24-26, or approximately 24, 25, or 26 amino acids long, and X_2 and X_4 are as above.

[0069] In some embodiments, and particularly in H5 polypeptides, HA Sequence Element 1 is extended (*i.e.*, at a position corresponding to residues 186-193) by the sequence:

N D A A E X X (K/R) (SEQ ID NO.115)

[0070] In some embodiments, and particularly in H5 polypeptides, HA Sequence Element 1 includes the sequence:

Y E E L K H L X S X X N H F E K (SEQ ID NO.116),

typically within X_1, and especially beginning about residue 6 of X_1 (as illustrated, for example, in **Figures 1, 2,** and **5)**.

HA Sequence Element 2

[0071] *HA Sequence Element 2* is a sequence element corresponding approximately to residues 324-340 (again using a numbering system based on H3 HA) of many HA proteins found in natural influenza isolates. This sequence element has the basic structure:

G A I A G F I E (SEQ ID NO.117)

In some embodiments, HA Sequence Element 2 has the sequence:

P X_1G A I A G F I E (SEQ ID NO.118), wherein:

X_1 is approximately 4-14 amino acids long, or about 8-12 amino acids long, or about 12, 11, 10, 9 or 8 amino acids long. In some embodiments, this sequence element provides the HA0 cleavage site, allowing production of HA1 and HA2.

[0072] In some embodiments, and particularly in H1 polypeptides, HA Sequence Element 2 has the structure:

P S (I/V) Q S R X_{1A} G A I A G F I E (SEQ ID NO.119), wherein:

X_{1A} is approximately 3 amino acids long; in some embodiments, X_{1A} is G (L/I) F.

[0073] In some embodiments, and particularly in H3 polypeptides, HA Sequence Element 2 has the structure:

P X K X T R X_{1A} G A I A G F I E (SEQ ID NO.120), wherein:

X_{1A} is approximately 3 amino acids long; in some embodiments, X_{1A} is G (L/I) F.

[0074] In some embodiments, and particularly in H5 polypeptides, HA Sequence Element 2 has the structure:

P Q R X X X R X X R X_{1A} G A I A G F I E (SEQ ID NO.121), wherein:

X_{1A} is approximately 3 amino acids long; in some embodiments, X_{1A} is G (L/I) F.

Definitions

[0075] *Affinity*: As is known in the art, "affinity" is a measure of the tightness with a particular ligand (*e.g.*, an HA polypeptide) binds to its partner (*e.g.*, and HA receptor). Affinities can be measured in different ways.

[0076] *Binding*: It will be understood that the term "binding", as used herein, typically refers to a non-covalent association between or among agents. In many embodiments herein, binding is addressed with respect to particular glycans (*e.g.*, umbrella topology glycans or cone topology glycans). It will be appreciated by those of ordinary skill in the art that such binding may be assessed in any of a variety of contexts. In some embodiments, binding is assessed with respect to free glycans. In some embodiments, binding is assessed with respect to glycans attached (*e.g.*, covalently linked to) a carrier. In some such embodiments, the carrier is a polypeptide. In some embodiments, binding is assessed with respect to glycans attached to an HA receptor. In such embodiments, reference may be made to receptor binding or to glycan binding.

[0077] *Binding agent*: In general, the term "binding agent" is used herein to refer to any entity that binds to glycans (*e.g.,* to umbrella-topology glycans) as described herein. Binding agents may be of any chemical type. In some embodiments, binding agents are polypeptides (including, *e.g.,* antibodies or antibody fragments); in some such embodiments, binding agents are HA polypeptides and/or variants thereof and/or characteristic portions thereof; in some embodiments, binding agents are polypeptides whose amino acid sequence does not include an HA characteristic sequence (*i.e.,* "Non-HA polypeptides"). In some embodiments, binding agents are small molecules. In some embodiments, binding agents are nucleic acids. In some embodiments, binding agents are aptamers. In some embodiments, binding agents are polymers; in some embodiments, binding agents are non-polymeric. In some embodiments, binding agents are carbohydrates. In some embodiments, binding agents are lectins. In some embodiments, binding agents as described herein bind to sialylated glycans having an umbrella-like topology. In some embodiments, binding agents bind to umbrella-topology glycans with high affinity and/or specificity. In some embodiments, binding agents show a binding preference for umbrella-topology glycans as compared with cone-topology glycans. In some embodiments, binding agents compete with hemagglutinin for binding to glycans on hemagglutinin receptors. In some embodiments, binding agents compete with hemagglutinin for binding to umbrella-topology glycans. In some embodiments, a binding agent provided herein is an umbrella

topology blocking agent. In some embodiments, a binding agent provided herein is an umbrella topology specific blocking agent. In some embodiments, binding agents bind to umbrella topology glycan mimics.

[0078] *Biologically active*: As used herein, the phrase "biologically active" refers to a characteristic of any agent that has activity in a biological system, and particularly in an organism. For instance, an agent that, when administered to an organism, has a biological effect on that organism, is considered to be biologically active. In some embodiments, where a protein or polypeptide is biologically active, a portion of that protein or polypeptide that shares at least one biological activity of the protein or polypeptide is typically referred to as a "biologically active" portion.

[0079] *Characteristic portion*: As used herein, the phrase a "characteristic portion" of a protein or polypeptide is one that contains a continuous stretch of amino acids, or a collection of continuous stretches of amino acids, that together are characteristic of a protein or polypeptide. Each such continuous stretch generally will contain at least two amino acids. Furthermore, those of ordinary skill in the art will appreciate that typically at least 5, at least 10, at least 15, at least 20 or more amino acids are required to be characteristic of a protein. In general, a characteristic portion is one that, in addition to the sequence identity specified above, shares at least one functional characteristic with the relevant intact protein.

[0080] *Characteristic sequence*: A "characteristic sequence" is a sequence that is found in all members of a family of polypeptides or nucleic acids, and therefore can be used by those of ordinary skill in the art to define members of the family.

[0081] *Cone topology:* The phrase "cone topology" is used herein to refer to a 3-dimensional arrangement adopted by certain glycans and in particular by glycans on HA receptors. As illustrated in **Figure 6**, the cone topology can be adopted by α2-3 sialylated glycans or by α2-6 sialylated glycans, and is typical of short oligonucleotide chains, though some long oligonucleotides can also adopt this conformation. The cone topology is characterized by the glycosidic torsion angles of Neu5Acα2-3Gal linkage which samples three regions of minimum energy conformations given by ϕ (C1-C2-O-C3/C6) value of about -60, about 60, or about 180 and ψ (C2-O-C3/C6-H3/C5) samples -60 to 60 (**Figure 11**). **Figure 8** presents certain representative (though not exhaustive) examples of glycans that adopt a cone topology.

[0082] *Corresponding to*: As used herein, the term "corresponding to" is often used to designate the position/identity of an amino acid residue in an HA polypeptide. Those of ordinary skill will appreciate that, for purposes of simplicity, a canonical numbering system (based on wild type H3 HA) is utilized herein (as illustrated, for example, in **Figures 1-5**), so that an amino acid "corresponding to" a residue at position 190, for example, need not actually be the 190th amino acid in a particular amino acid chain but rather corresponds to the residue found at 190 in wild type H3 HA; those of ordinary skill in the art readily appreciate how to identify corresponding amino acids.

[0083] *Degree of separation removed*: As used herein, amino acids that are a "degree of separation removed" are HA amino acids that have indirect effects on glycan binding. For example, one-degree-of-separation-removed amino acids may either: (1) interact with the direct-binding amino acids; and/or (2) otherwise affect the ability of direct-binding amino acids to interact with glycan that is associated with host cell HA receptors; such one-degree-of-separation-removed amino acids may or may not directly bind to glycan themselves. Two-degree-of-separation-removed amino acids either (1) interact with one-degree-of-separation-removed amino acids; and/or (2) otherwise affect the ability of the one-degree-of-separation-removed amino acids to interact with direct-binding amino acids, *etc.*

[0084] *Direct-binding amino acids*: As used herein, the phrase "direct-binding amino acids" refers to HA polypeptide amino acids which interact directly with one or more glycans that is associated with host cell HA receptors.

[0085] *Engineered*: The term "engineered", as used herein, describes a polypeptide whose amino acid sequence has been selected by man. For example, an engineered HA polypeptide has an amino acid sequence that differs from the amino acid sequences of HA polypeptides found in natural influenza isolates. In some embodiments, an engineered HA polypeptide has an amino acid sequence that differs from the amino acid sequence of HA polypeptides included in the NCBI database.

[0086] *H1 polypeptide*: An "H1 polypeptide", as that term is used herein, is an HA polypeptide whose amino acid sequence includes at least one sequence element that is characteristic of H1 and distinguishes H1 from other HA subtypes. Representative such sequence elements can be determined by alignments such as, for example, those illustrated in

Figures 1-3 and include, for example, those described herein with regard to H1-specific embodiments of HA Sequence Elements.

[0087] *H3 polypeptide*: An "H3 polypeptide", as that term is used herein, is an HA polypeptide whose amino acid sequence includes at least one sequence element that is characteristic of H3 and distinguishes H3 from other HA subtypes. Representative such sequence elements can be determined by alignments such as, for example, those illustrated in **Figures 1, 2, and 4** and include, for example, those described herein with regard to H3-specific embodiments of HA Sequence Elements.

[0088] *H5 polypeptide*: An "H5 polypeptide", as that term is used herein, is an HA polypeptide whose amino acid sequence includes at least one sequence element that is characteristic of H5 and distinguishes H5 from other HA subtypes. Representative such sequence elements can be determined by alignments such as, for example, those illustrated in **Figures 1, 2, and 5** and include, for example, those described herein with regard to H5-specific embodiments of HA Sequence Elements.

[0089] *HX polypeptide*: An "HX polypeptide", as that term is used herein, is an HA polypeptide whose amino acid sequence includes at least one sequence element that is characteristic of HX and distinguishes HX from other HA subtypes, wherein "X" refers to the numbering of the HA subtype (*e.g.*, wherein when "X" = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16, "HX" = H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16, respectively).

[0090] *Hemagglutinin (HA) polypeptide*: As used herein, the term "hemagglutinin polypeptide" (or "HA polypeptide') refers to a polypeptide whose amino acid sequence includes at least one characteristic sequence of HA. A wide variety of HA sequences from influenza isolates are known in the art; indeed, the National Center for Biotechnology Information (NCBI) maintains a database (www.ncbi.nlm.nih.gov/genomes/FLU/flu.html) that, as of the filing of the present application included 9796 HA sequences. Those of ordinary skill in the art, referring to this database, can readily identify sequences that are characteristic of HA polypeptides generally, and/or of particular HA polypeptides (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 polypeptides); or of HAs that mediate infection of particular hosts, *e.g.*, avian, camel, canine, cat, civet, environment, equine, human, leopard, mink, mouse, seal, stone martin, swine, tiger, whale, *etc.* For example, in some embodiments, an HA polypeptide

includes one or more characteristic sequence elements found between about residues 97 and about 185, about 324 and about 340, about 96 and about 100, and/or about 130 and about 230 of an HA protein found in a natural isolate of an influenza virus. In some embodiments, an HA polypeptide has an amino acid sequence comprising at least one of HA Sequence Elements 1 and 2, as defined herein. In some embodiments, an HA polypeptide has an amino acid sequence comprising HA Sequence Elements 1 and 2, in some embodiments separated from one another by about 100 to about 200, or by about 125 to about 175, or about 125 to about 160, or about 125 to about 150, or about 129 to about 139, or about 129, about 130, about 131, about 132, about 133, about 134, about 135, about 136, about 137, about 138, or about 139 amino acids. In some embodiments, an HA polypeptide has an amino acid sequence that includes residues at positions within the regions 96-100 and/or 130-230 that participate in glycan binding. For example, many HA polypeptides include one or more of the following residues: Tyr98, Ser/Thr136, Trp153, His183, and Leu/Ile194. In some embodiments, an HA polypeptide includes at least 2, 3, 4, or all 5 of these residues.

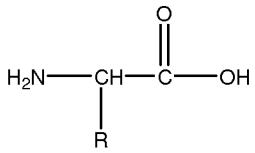
[0091] *Isolated*: The term "isolated", as used herein, refers to an agent or entity that has either (i) been separated from at least some of the components with which it was associated when initially produced (whether in nature or in an experimental setting); or (ii) produced by the hand of man. Isolated agents or entities may be separated from at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, or more of the other components with which they were initially associated. In some embodiments, isolated agents are more than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% pure.

[0092] *Linkage Specific Blocking Agent (LSBA):* As used herein, the term "linkage specific blocking agent" refers to an agent which binds to an HA receptor having an α2-6 sialylated glycan. In some embodiments, an LSBA selectively binds to an HA receptor having an α2-6 sialylated glycan with at least about 40, 50, or 75% of the affinity of that for an HA receptor having an α2-3 sialylated glycan. In some embodiments, an LSBA selectively binds to an HA receptor having an α2-6 sialylated glycan with at least about 2, 4, 5, or 10 times greater affinity than that for an HA receptor having an α2-3 sialylated glycan. In some embodiments, an LSBA has an affinity for an α2-6 sialylated glycan that is at least 50, 100, 150, or 200 % of its affinity for an α2-3 sialylated glycan. In some embodiments, an LSBA may compete with hemagglutinin

for binding to an HA receptor. For example, an LSBA may selectively inhibit the binding of an influenza virus particle (*e.g.*, human or avian influenza virus) to an HA receptor based on the linkage characteristics (*e.g.*, α2-6 sialylated glycan or α2-3 sialylated glycan). In some embodiments, an LSBA is a polypeptide. In some such embodiments, an LSBA polypeptide has an amino acid sequence that is substantially identical or substantially homologous to that of a naturally-occurring polypeptide. In some embodiments, an LSBA polypeptide is an HA polypeptide. In some embodiments, an LSBA polypeptide is a naturally-occurring HA polypeptide, or a fragment thereof. In some embodiments, an LSBA polypeptide has an amino acid sequence that is not related to that of an HA polypeptide. In some embodiments, an LSBA polypeptide is an antibody or fragment thereof. In some embodiments, an LSBA polypeptide is a lectin (*e.g.,* SNA-1). In some embodiments, an LSBA is not a polypeptide. In some embodiments, an LSBA is a small molecule. In some embodiments, an LSBA is a nucleic acid.

[0093] *Long oligosaccharide*: For purposes of the present disclosure, an oligosaccharide is typically considered to be "long" if it includes at least one linear chain that has at least four saccharide residues.

[0094] *Non-natural amino acid*: The phrase "non-natural amino acid" refers to an entity having the chemical structure of an amino acid (*i.e.*,:



and therefore being capable of participating in at least two peptide bonds, but having an R group that differs from those found in nature. In some embodiments, non-natural amino acids may also have a second R group rather than a hydrogen, and/or may have one or more other substitutions on the amino or carboxylic acid moieties.

[0095] *Polypeptide*: A "polypeptide", generally speaking, is a string of at least two amino acids attached to one another by a peptide bond. In some embodiments, a polypeptide may include at least 3-5 amino acids, each of which is attached to others by way of at least one peptide bond. Those of ordinary skill in the art will appreciate that polypeptides sometimes include "non-natural" amino acids or other entities that nonetheless are capable of integrating into a polypeptide chain, optionally.

[0096] *Pure*: As used herein, an agent or entity is "pure" if it is substantially free of other components. For example, a preparation that contains more than about 90% of a particular agent

or entity is typically considered to be a pure preparation. In some embodiments, an agent or entity is at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% pure.

[0097] *Short oligosaccharide*: For purposes of the present disclosure, an oligosaccharide is typically considered to be "short" if it has fewer than 4, or certainly fewer than 3, residues in any linear chain.

[0098] *Specificity*: As is known in the art, "specificity" is a measure of the ability of a particular ligand (*e.g.*, an HA polypeptide) to distinguish its binding partner (*e.g.*, a human HA receptor, and particularly a human upper respiratory tract HA receptor) from other potential binding partners (*e.g.*, an avian HA receptor).

[0099] *Substantial homology*: The phrase "substantial homology" is used herein to refer to a comparison between amino acid or nucleic acid sequences. As will be appreciated by those of ordinary skill in the art, two sequences are generally considered to be "substantially homologous" if they contain homologous residues in corresponding positions. Homologous residues may be identical residues. Alternatively, homologous residues may be non-identical residues will appropriately similar structural and/or functional characteristics. For example, as is well known by those of ordinary skill in the art, certain amino acids are typically classified as "hydrophobic" or "hydrophilic" amino acids, and/or as having "polar" or "non-polar" side chains. Substitution of one amino acid for another of the same type may often be considered a "homologous" substitution. Typical amino acid categorizations are summarized below:

Alanine	Ala	A	nonpolar	neutral	1.8
Arginine	Arg	R	polar	positive	-4.5
Asparagine	Asn	N	polar	neutral	-3.5
Aspartic acid	Asp	D	polar	negative	-3.5
Cysteine	Cys	C	nonpolar	neutral	2.5
Glutamic acid	Glu	E	polar	negative	-3.5
Glutamine	Gln	Q	polar	neutral	-3.5
Glycine	Gly	G	nonpolar	neutral	-0.4

Histidine	His	H	polar	positive	-3.2
Isoleucine	Ile	I	nonpolar	neutral	4.5
Leucine	Leu	L	nonpolar	neutral	3.8
Lysine	Lys	K	polar	positive	-3.9
Methionine	Met	M	nonpolar	neutral	1.9
Phenylalanine	Phe	F	nonpolar	neutral	2.8
Proline	Pro	P	nonpolar	neutral	-1.6
Serine	Ser	S	polar	neutral	-0.8
Threonine	Thr	T	polar	neutral	-0.7
Tryptophan	Trp	W	nonpolar	neutral	-0.9
Tyrosine	Tyr	Y	polar	neutral	-1.3
Valine	Val	V	nonpolar	neutral	4.2

Ambiguous Amino Acids	**3-Letter**	**1-Letter**
Asparagine or aspartic acid	Asx	B
Glutamine or glutamic acid	Glx	Z
Leucine or Isoleucine	Xle	J
Unspecified or unknown amino acid	Xaa	X

As is well known in this art, amino acid or nucleic acid sequences may be compared using any of a variety of algorithms, including those available in commercial computer programs such as BLASTN for nucleotide sequences and BLASTP, gapped BLAST, and PSI-BLAST for amino acid sequences. Exemplary such programs are described in Altschul, *et al.*, Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, *et al.*, *Methods in Enzymology*; Altschul, *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis, *et al.*,

Bioinformatics : A Practical Guide to the Analysis of Genes and Proteins, Wiley, 1998; and Misener, *et al.*, (eds.), *Bioinformatics Methods and Protocols* (Methods in Molecular Biology, Vol. 132), Humana Press, 1999; all of the foregoing of which are incorporated herein by reference. In addition to identifying homologous sequences, the programs mentioned above typically provide an indication of the degree of homology. In some embodiments, two sequences are considered to be substantially homologous if at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more of their corresponding residues are homologous over a relevant stretch of residues. In some embodiments, the relevant stretch is a complete sequence. In some embodiments, the relevant stretch is at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 125, at least 150, at least 175, at least 200, at least 225, at least 250, at least 275, at least 300, at least 325, at least 350, at least 375, at least 400, at least 425, at least 450, at least 475, at least 500 or more residues.

[00100] *Substantial identity*: The phrase "substantial identity" is used herein to refer to a comparison between amino acid or nucleic acid sequences. As will be appreciated by those of ordinary skill in the art, two sequences are generally considered to be "substantially identical" if they contain identical residues in corresponding positions. As is well known in this art, amino acid or nucleic acid sequences may be compared using any of a variety of algorithms, including those available in commercial computer programs such as BLASTN for nucleotide sequences and BLASTP, gapped BLAST, and PSI-BLAST for amino acid sequences. Exemplary such programs are described in Altschul, *et al.*, Basic local alignment search tool, *J. Mol. Biol.*, 215(3): 403-410, 1990; Altschul, *et al.*, *Methods in Enzymology*; Altschul, *et al.*, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402, 1997; Baxevanis, *et al.*, *Bioinformatics : A Practical Guide to the Analysis of Genes and Proteins*, Wiley, 1998; and Misener, *et al.*, (eds.), *Bioinformatics Methods and Protocols* (Methods in Molecular Biology, Vol. 132), Humana Press, 1999; all of the foregoing of which are incorporated herein by reference. In addition to identifying identical sequences, the programs mentioned above typically provide an indication of the degree of identity. In some embodiments, two sequences are considered to be substantially identical if at least 50%, at least

55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more of their corresponding residues are identical over a relevant stretch of residues. In some embodiments, the relevant stretch is a complete sequence. In some embodiments, the relevant stretch is at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 125, at least 150, at least 175, at least 200, at least 225, at least 250, at least 275, at least 300, at least 325, at least 350, at least 375, at least 400, at least 425, at least 450, at least 475, at least 500 or more residues.

[00101] *Therapeutic agent*: As used herein, the phrase "therapeutic agent" refers to any agent that elicits a desired biological or pharmacological effect.

[00102] *Treatment*: As used herein, the term "treatment" refers to any method used to alleviate, delay onset, reduce severity or incidence, or yield prophylaxis of one or more symptoms or aspects of a disease, disorder, or condition. For the purposes of the present invention, treatment can be administered before, during, and/or after the onset of symptoms.

[00103] *Umbrella topology:* The phrase "umbrella topology" is used herein to refer to a 3-dimensional arrangement adopted by certain glycans and in particular by glycans on HA receptors. The present invention encompasses the recognition that binding to umbrella topology glycans is characteristic of HA proteins that mediate infection of human hosts. As illustrated in **Figure 6**, the umbrella topology is typically adopted only by α2-6 sialylated glycans, and is typical of long (*e.g.*, greater than tetrasaccharide) oligosaccharides. In some embodiments, *umbrella*-topology glycans are glycans exhibiting a three-dimensional structure substantially similar to the structure presented in **Figure 6** (right panel). In some embodiments, *umbrella*-topology glycans are glycans which contact HA polypeptides via the amino acid residues shown in **Figure 6** (right panel). In some embodiments, *umbrella*-topology glycans are glycans which are able to contact and/or specifically bind to the amino acid binding pocket shown in **Figure 6** (right panel). In some embodiments, glycan structural topology is classified based on parameter θ defined as angle between C_2 of Sia, C_1 of Gal, and C_1 of GlcNAc. Values of $\theta < 100°$ represent *cone*-like topology adopted by α2-3 and short α2-6 glycans. Values of $\theta > 110°$ represent *umbrella*-like topology, such as topology adopted by long α2-6 glycans (**Figure 6**). An

example of umbrella topology is given by ϕ angle of Neu5Acα2-6Gal linkage of around -60 (see, for example, **Figure 11**). **Figure 9** presents certain representative (though not exhaustive) examples of glycans that can adopt an umbrella topology. The long α2-6 motifs presented in **Figure 9** includes Neu5Acα2-6 linked at the non-reducing end to a long chain (*e.g.*, at least a trisaccharide) found as a part of biological N-linked glycans, O-linked glycans, and glycolipids. The boxed inset shows examples of the *umbrella*-topology long α2-6 glycan moieties that are found as a part of biological glycans that bind to high affinity with HA. In some embodiments, *umbrella*-topology glycans (*e.g.*, at a site) comprise a greater proportion of long (*e.g.* multiple lactosamine units) α2-6 oligosaccharide branches than short α2-6 (*e.g.* single lactosamine) branches. In some embodiments, *umbrella*-topology glycans (*e.g.*, at a site) comprise about 2-fold, about 3-fold, about 4-fold, about 5-fold, about 10-fold, about 20-fold, about 50-fold, or greater than about 50-fold more long α2-6 oligosaccharide branches than short α2-6 (*e.g.* single lactosamine) branches. In some embodiments, the unique characteristic of HA interactions with umbrella-topology glycans and/or glycan decoys is the HA contact with a glycan comprising sialic acid (SA) and/or SA analogs at the non-reducing end. In some embodiments, chain length of the oligosaccharide is at least a trisaccharide (excluding the SA or SA analog). In some embodiments, a combination of the numbered residues shown in the right-hand panel of **Figure 6** is involved in contacts with umbrella-like topology. In some embodiments, umbrella topology glycans are oligosaccharides of the following form:

Neu5Acα2-6Sug1-Sug2-Sug3

where:

(a) Neu5Ac α2-6 is typically (but not essentially) at the non-reducing end;

(b) Sug1:

(i) is a hexose (frequently Gal or Glc) or hexosamine (GlcNAc or GalNAc) in α or β configuration (frequently β- for N- and O-linked extension and α- in the case of GalNAcα- that is O-linked to glycoprotein);

(ii) no sugars other than Neu5Acα2-6 are attached to any of the non-reducing positions of Sug1 (except when Sug1 is GalNAcα- that is O-linked to the glycoprotein); and/or

(iii) non-sugar moieties such as sulfate, phosphate, guanidium, amine, N-acetyl, *etc.* can be attached to non-reducing positions (typically 6 position) of Sug1 (*e.g.*, to improve contacts with HA);

(c) Sug2 and/or Sug3 is/are:

(i) hexose (frequently Gal or Glc) or hexosamine (GlcNAc or GalNAc) in α or β configuration (frequently β); and/or

(ii) sugars (such as Fuc) or non-sugar moieties such as sulfate, phosphate, guanidium, amine, N-acetyl, *etc.* can be attached to non-reducing positions of Sug2, Sug3, and/or Sug4;

(d) Linkage between any two sugars in the oligosaccharide apart from Neu5Acα2-6 linkage can be 1-2, 1-3, 1-4, and/or 1-6 (typically 1-3 or 1-4); and/or

(e) Structure where Neu5Acα2-6 is linked GalNAcα that is O-linked to the glycoprotein and additional sugars are linked to the non-reducing end of GalNAcα for example

(i) Neu5Acα2-6(Neu5Acα2-3Galβ1-3)GalNAcα-

(ii) Neu5Acα2-6(Galβ1-3)GalNAcα-

[00104] *Umbrella topology blocking agent (UTBA)*: As used herein, the term "umbrella topology blocking agent" refers to an agent which binds to an HA receptor having an umbrella topology glycan. In some embodiments, a UTBA binds to an HA receptor having an umbrella topology glycan found in human upper airways. A UBTA can bind to either an umbrella topology glycan and/or to a cone topology glycan. In some embodiments, a UTBA selectively binds to an umbrella topology glycan with 50, 100, 150, or 200% of its affinity for a cone topology glycan. In some embodiments a UTBA selectively binds to an umbrella topology glycan with 50-150% of its affinity for a cone topology glycan. In some embodiments, a UTBA binds to an umbrella topology glycan with about the same affinity as for a cone topology glycan. For example, in some embodiments, a UTBA binds an umbrella topology glycan (*e.g.*, 6'SLN-LN) with about 50-200%, 50-150%, or about the same affinity to which it binds a cone topology glycan (*e.g.*, 3'SLN-LN). In some embodiments, a UTBA selectively inhibits the binding of an influenza virus particle (*e.g.*, a human or avian influenza virus) to the HA receptor based on the glycan topology of the receptor (*e.g.*, umbrella or cone). In some embodiments, a UTBA is a polypeptide. In some such embodiments, a UTBA polypeptide has an amino acid sequence that is substantially identical or substantially homologous to that of a naturally-occurring polypeptide. In some embodiments, a UTBA polypeptide is an HA polypeptide. In some embodiments, a UTBA polypeptide is a naturally-occurring HA polypeptide, or a fragment thereof. In some

embodiments, a UTBA polypeptide has an amino acid sequence that is not related to that of an HA polypeptide. In some embodiments, a UTBA polypeptide is an antibody or fragment thereof. In some embodiments, a UTBA polypeptide is a lectin (*e.g.,* SNA-1). In some embodiments, a UTBA is not a polypeptide. In some embodiments, a UTBA is a small molecule. In some embodiments, a UTBA is a nucleic acid.

[00105] *Umbrella topology glycan mimic:* An "umbrella topology glycan mimic" is an agent, other than an umbrella topology glycan, that binds to binding agents as described herein. In some embodiments, umbrella topology glycan mimics are agents that bind to HA polypeptides. In some such embodiments, umbrella topology glycan mimics are agents that interact with HA polypeptide residues selected from the group consisting of residues 95, 98, 128, 130, 131, 132, 133, 135, 136, 137, 138, 145, 153, 155, 156, 158, 159, 160, 183, 186, 187, 188, 189, 190, 192, 193, 194, 195, 196, 219, 221, 222, 224, 225, 226, 227, 228, and combinations thereof. In some such embodiments, umbrella topology glycan mimics are agents that interact with HA polypeptide residues selected from the group consisting of residues 130, 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, 228, and combinations thereof. In some such embodiments, umbrella topology glycan mimics are agents that interact with HA polypeptide residues selected from the group consisting of residues 160, 192, 193, and combinations thereof. Note that amino acid positions stated above are based on H3 HA numbering. In some embodiments, an HA topology glycan mimic is an agent that competes with umbrella topology glycans for interaction with an HA polypeptide.

[00106] *Umbrella topology specific blocking agent (UTSBA):* As used herein, the term "umbrella topology specific blocking agent" refers to an agent which binds to an HA receptor having an umbrella topology glycan found in human upper airways. A UTSBA selectively binds an umbrella topology glycan HA. For example, a UTSBA binds an umbrella topology glycan (*e.g.*, 6'SLN-LN) with about at least 2, at least 4, at least 5, or at least 10 times greater affinity than it binds to a cone topology glycan (*e.g.*, 3'SLN-LN). Typically, the affinity of a UTSBA for an umbrella topology glycan is greater than 1 nM. Typically the affinity of a UTSBA for a cone topology glycan is less is at least within 2 to 3 orders of magnitude of the binding affinity of umbrella topology glycans to human adapted HAs such as SC18, Mos99, Tx91, *etc.* and α2-6 binding plant lectins such as SNA-I. The binding affinity of UTSBA as measured by the dose-dependent direct binding assay (**Figs 19 and 20**) would typically be at least 1 nM. Typically the

affinity of a UTSBA for a cone topology glycan is at most 1 to 3 orders of magnitude less than the binding affinity of cone topology glycans to avian HAs such as Viet0405, Av18, *etc.* In some embodiments, a UTSBA selectively inhibits binding of an influenza virus particle (*e.g.*, a human or avian influenza virus) to the HA receptor (*e.g.*, an H1, H2 or H3 or a human-adapted H5, H7 or H9) based on glycan topology (*e.g.*, umbrella or cone). In some embodiments, a UTSBA is a polypeptide. In some such embodiments, a UTSBA polypeptide has an amino acid sequence that is that is substantially identical or substantially homologous to that of a naturally-occurring polypeptide. In some embodiments, a UTSBA polypeptide is an HA polypeptide. In some embodiments, a UTSBA polypeptide is a naturally-occurring HA polypeptide, or a fragment thereof. In some embodiments, a UTSBA polypeptide has an amino acid sequence that is not related to that of an HA polypeptide. In some embodiments, a UTSBA polypeptide is an antibody or fragment thereof. In some embodiments, a UTSBA polypeptide is a lectin (*e.g.*, SNA-1). In some embodiments, a UTSBA is not a polypeptide. In some embodiments, a UTSBA is a small molecule. In some embodiments, a UTSBA is a nucleic acid.

[00107] *Vaccination*: As used herein, the term "vaccination" refers to the administration of a composition intended to generate an immune response, for example to a disease-causing agent. For the purposes of the present invention, vaccination can be administered before, during, and/or after exposure to a disease-causing agent, and in some embodiments, before, during, and/or shortly after exposure to the agent. In some embodiments, vaccination includes multiple administrations, appropriately spaced in time, of a vaccinating composition.

[00108] *Variant*: As used herein, the term "variant" is a relative term that describes the relationship between a particular polypeptide (*e.g.*, HA polypeptide) of interest and a "parent" polypeptide to which its sequence is being compared. A polypeptide of interest is considered to be a "variant" of a parent polypeptide if the polypeptide of interest has an amino acid sequence that is identical to that of the parent but for a small number of sequence alterations at particular positions. Typically, fewer than 20%, 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% of the residues in the variant are substituted as compared with the parent. In some embodiments, a variant has 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 substituted residue as compared with a parent. Often, a variant has a very small number (*e.g.*, fewer than 5, 4, 3, 2, or 1) number of substituted functional residues (*i.e.*, residues that participate in a particular biological activity). Furthermore, a variant typically has not more than 5, 4, 3, 2, or 1 additions or deletions, and

often has no additions or deletions, as compared with the parent. Moreover, any additions or deletions are typically fewer than about 25, about 20, about 19, about 18, about 17, about 16, about 15, about 14, about 13, about 10, about 9, about 8, about 7, about 6, and commonly are fewer than about 5, about 4, about 3, or about 2 residues. In some embodiments, the parent polypeptide is one found in nature. For example, a parent HA polypeptide may be one found in a natural (*e.g.*, wild type) isolate of an influenza virus (*e.g.*, a wild type HA).

[00109] *Vector*: As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. In some embodiment, vectors are capable of extra-chromosomal replication and/or expression of nucleic acids to which they are linked in a host cell such as a eukaryotic or prokaryotic cell. Vectors capable of directing the expression of operatively linked genes are referred to herein as "expression vectors."

[00110] *Wild type*: As is understood in the art, the phrase "wild type" generally refers to a normal form of a protein or nucleic acid, as is found in nature. For example, wild type HA polypeptides are found in natural isolates of influenza virus. A variety of different wild type HA sequences can be found in the NCBI influenza virus sequence database, http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html.

Detailed Description of Certain Embodiments

[00111] The present invention provides binding agents (*e.g.,* HA polypeptides, HA polypeptide variants, LSBAs, UTBAs, UTSBAs, *etc.*) that bind to umbrella topology glycans. In some embodiments, the present invention provides binding agents that bind to umbrella topology glycans found on HA receptors of a particular target species. For example, in some embodiments, the present invention provides binding agents that bind to umbrella topology glycans found on human HA receptors, *e.g.*, HA receptors found on human epithelial cells, and particularly binding agents that bind to umbrella topology glycans found on human HA receptors in the upper respiratory tract.

[00112] The present invention provides binding agents that bind to HA receptors found on cells in the human upper respiratory tract, and in particular provides binding agents that binds to such receptors (and/or to their glycans, particularly to their umbrella glycans) with a designated affinity and/or specificity.

[00113] In some embodiments, binding agents in accordance with the present invention are or comprise HA polypeptide sequences. In some embodiments, binding agents in accordance with the present invention comprise an HA polypeptide sequence which differs from a parent naturally-occurring HA polypeptide sequence at one or more of residues selected from the group consisting of residues 130, 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, 228, and combinations thereof. In some embodiments, binding agents in accordance with the present invention comprise an HA polypeptide sequence which differs from a parent naturally-occurring HA polypeptide sequence at one or more of residues selected from the group consisting of residues 160, 192, 193, and combinations thereof.

[00114] The present invention encompasses the recognition that gaining an ability to bind umbrella topology glycans (*e.g.*, long α2-6 sialylated glycans), and particularly an ability to bind with high affinity, may confer upon an HA polypeptide variant the ability to infect humans (where its parent HA polypeptide cannot). Without wishing to be bound by any particular theory, the present inventors propose that binding to umbrella topology glycans may be paramount, and in particular that loss of binding to other glycan types may not be required.

[00115] The present invention further provides various reagents and methods associated with binding agents in accordance with the invention (*e.g.,* HA polypeptides, HA polypeptide variants, UTBAs, UTSBAs, *etc.*) including, for example, systems for identifying them, strategies for preparing them, antibodies that bind to them, and various diagnostic and therapeutic methods relating to them. Further description of certain embodiments of these aspects, and others, of the present invention, is presented below.

Hemagglutinin (HA)

[00116] Influenza viruses are RNA viruses which are characterized by a lipid membrane envelope containing two glycoproteins, hemagglutinin (HA) and neuraminidase (NA), embedded in the membrane of the virus particular. There are 16 known HA subtypes and 9 NA subtypes, and different influenza strains are named based on the number of the strain's HA and NA subtypes. Based on comparisons of amino acid sequence identity and of crystal structures, the HA subtypes have been divided into two main groups and four smaller clades. The different HA subtypes do not necessarily share strong amino acid sequence identity, but the overall 3D structures of the different HA subtypes are similar to one another, with several subtle differences

that can be used for classification purposes. For example, the particular orientation of the membrane-distal subdomains in relation to a central α-helix is one structural characteristic commonly used to determine HA subtype (Russell *et al.*, 2004, *Virology*, 325:287; incorporated herein by reference).

[00117] HA exists in the membrane as a homotrimer of one of 16 subtypes, termed H1-H16. Only three of these subtypes (H1, H2, and H3) have thus far become adapted for human infection. One reported characteristic of HAs that have adapted to infect humans (*e.g.*, of HAs from the pandemic H1N1 (1918) and H3N2 (1967-68) influenza subtypes) is their ability to preferentially bind to α2-6 sialylated glycans in comparison with their avian progenitors that preferentially bind to α2-3 sialylated glycans (Skehel & Wiley, 2000, *Annu Rev Biochem*, 69:531; Rogers, & Paulson, 1983, *Virology*, 127:361; Rogers *et al.*, 1983, *Nature*, 304:76; Sauter *et al.*, 1992, *Biochemistry*, 31:9609; Connor *et al.*, 1994, *Virology*, 205:17; Tumpey *et al.*, 2005, *Science*, 310:77; all of which are incorporated herein by reference). The present invention, however, encompasses the recognition that ability to infect human hosts correlates less with binding to glycans of a particular linkage, and more with binding to glycans of a particular topology. Thus, the present invention demonstrates that HAs that mediate infection of humans bind to umbrella topology glycans, often showing preference for umbrella topology glycans over cone topology glycans (even though cone-topology glycans may be α2-6 sialylated glycans).

[00118] Several crystal structures of HAs from H1 (human and swine), H3 (avian) and H5 (avian) subtypes bound to sialylated oligosaccharides (of both α2-3 and α2-6 linkages) are available and provide molecular insights into the specific amino acids that are involved in distinct interactions of the HAs with these glycans (Eisen *et al.*, 1997, *Virology*, 232:19; Ha *et al.*, 2001, *Proc Natl Acad Sci USA*, 98:11181; Ha *et al.*, 2003, *Virology*, 309:209; Gamblin *et al.*, 2004, *Science,* 303:1838; Stevens *et al.*, 2004, *Science*, 303:1866; Russell *et al.*, 2006, *Glycoconj J* 23:85; Stevens *et al.*, 2006, *Science,* 312:404; all of which are incorporated herein by reference).

[00119] For example, the crystal structures of H5 (A/duck/Singapore/3/97) alone or bound to an α2-3 or an α2-6 sialylated oligosaccharide identifies certain amino acids that interact directly with bound glycans, and also amino acids that are one or more degree of separation removed (Stevens *et al.*, 2001, *Proc Natl Acad Sci USA* 98:11181; incorporated herein by reference). In some cases, conformation of these residues is different in bound versus unbound states. For

instance, Glu190, Lys193, and Gln226 all participate in direct-binding interactions and have different conformations in the bound versus the unbound state. The conformation of Asn186, which is proximal to Glu190, is also significantly different in the bound versus the unbound state.

Binding Agents

[00120] As noted above, the present invention encompasses the finding that binding to umbrella topology glycans correlates with ability to mediate infection of particular hosts, including for example, humans. Accordingly, the present invention provides binding agents (*e.g.*, HA polypeptides, HA polypeptide variants, LSBAs, UTBAs, UTSBAs, *etc.*) that bind to umbrella glycans (and/or to umbrella topology glycan mimics). In some embodiments, binding agents in accordance with the invention bind to umbrella glycans (and/or to umbrella topology glycan mimics) with high affinity. In some embodiments, binding agents in accordance with the invention bind to a plurality of different umbrella topology glycans, often with high affinity and/or specificity.

[00121] In some embodiments, binding agents in accordance with the invention bind to umbrella topology glycans (*e.g.*, long α2-6 sialylated glycans such as, for example, Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc-) with high affinity. For example, in some embodiments, binding agents in accordance with the invention bind to umbrella topology glycans with an affinity comparable to that observed for a wild type HA that mediates infection of a humans (*e.g.*, H1N1 HA or H3N2 HA). In some embodiments, binding agents in accordance with the invention bind to umbrella glycans with an affinity that is at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% of that observed under comparable conditions for a wild type HA that mediates infection of humans. In some embodiments, binding agents in accordance with the invention bind to umbrella glycans with an affinity that is greater than that observed under comparable conditions for a wild type HA that mediates infection of humans.

[00122] In some embodiments, binding affinity of binding agents in accordance with the invention is assessed over a range of concentrations. Such a strategy provides significantly more information, particularly in multivalent binding assays, than do single-concentration analyses. In

some embodiments, for example, binding affinities of binding agents in accordance with the invention are assessed over concentrations ranging over at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10 or more fold.

[00123] In some embodiments, binding agents in accordance with the invention show high affinity if they show a saturating signal in a multivalent glycan array binding assay such as those described herein. In some embodiments, binding agents in accordance with the invention show high affinity if they show a signal above about 400000 or more (*e.g.*, above about 500000, about 600000, about 700000, about 800000, *etc.*) in such studies. In some embodiments, binding agents as described herein show saturating binding to umbrella glycans over a concentration range of at least 2 fold, at least 3 fold, at least 4 fold, at least 5 fold or more, and in some embodiments over a concentration range as large as 10 fold or more.

[00124] Furthermore, in some embodiments, binding agents in accordance with the invention bind to umbrella topology glycans (and/or to umbrella topology glycan mimics) more strongly than they bind to cone topology glycans. In some embodiments, binding agents in accordance with the invention show a relative affinity for umbrella glycans vs. cone glycans that is about 10, about 9, about 8, about 7, about 6, about 5, about 4, about 3, or about 2.

[00125] In some embodiments, binding agents in accordance with the invention bind to α2-6 sialylated glycans; in some embodiments, binding agents in accordance with the invention bind preferentially to α2-6 sialylated glycans. In some embodiments, binding agents in accordance with the invention bind to a plurality of different α2-6 sialylated glycans. In some embodiments, binding agents in accordance with the invention are not able to bind to α2-3 sialylated glycans, and in some embodiments binding agents in accordance with the invention are able to bind to α2-3 sialylated glycans.

[00126] In some embodiments, binding agents in accordance with the invention bind to receptors found on human upper respiratory epithelial cells. In some embodiments, binding agents in accordance with the invention bind to HA receptors in the bronchus and/or trachea. In some embodiments, binding agents in accordance with the invention are not able to bind receptors in the deep lung, and in some embodiments, binding agents in accordance with the invention are able to bind receptors in the deep lung.

[00127] In some embodiments, binding agents in accordance with the invention bind to at least about 10%, about 15%, about 20%, about 25%, about 30% about 35%, about 40%, about

45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, or more of the glycans found on HA receptors in human upper respiratory tract tissues (*e.g.*, epithelial cells).

[00128] In some embodiments, binding agents in accordance with the invention bind to one or more of the glycans illustrated in **Figure 9**. In some embodiments, binding agents in accordance with the invention bind to multiple glycans illustrated in **Figure 9**. In some embodiments, binding agents in accordance with the invention bind with high affinity and/or specificity to glycans illustrated in **Figure 9**. In some embodiments, binding agents in accordance with the invention bind to glycans illustrated in **Figure 9** preferentially as compared with their binding to glycans illustrated in **Figure 8**. In some embodiments, binding agents in accordance with the invention bind to an oligosaccharide of the following form:

Neu5Acα2-6Sug1-Sug2-Sug3

where:

1. Neu5Ac α2-6 is always or almost always at the non-reducing end;
2. Sug1:
    a. is a hexose (frequently Gal or Glc) or hexosamine (GlcNAc or GalNAc) in α or β configuration (frequently β- for N- and O-linked extension and α- in the case of GalNAcα- that is O-linked to glycoprotein);
    b. no sugars other than Neu5Acα2-6 should be attached to any of the non-reducing positions of Sug1 (except when Sug1 is GalNAcα- that is O-linked to the glycoprotein); and/or
    c. non-sugar moieties such as sulfate, phosphate, guanidium, amine, N-acetyl, *etc.* can be attached to non-reducing positions (typically 6 position) of Sug1 to improve contacts with HA;
3. Sug2 and/or Sug3:
    a. hexose (frequently Gal or Glc) or hexosamine (GlcNAc or GalNAc) in α or β configuration (frequently β); and/or
    b. sugars (such as Fuc) or non-sugar moieties such as sulfate, phosphate, guanidium, amine, N-acetyl, *etc.* can be attached to non-reducing positions of Sug2, Sug3, and/or Sug4;

4. Linkage between any two sugars in the oligosaccharide apart from Neu5Acα2-6 linkage can be 1-2, 1-3, 1-4, and/or 1-6 (typically 1-3 or 1-4); and/or
5. Structure where Neu5Acα2-6 is linked GalNAcα that is O-linked to the glycoprotein and additional sugars are linked to the non-reducing end of GalNAcα for example
    i. Neu5Acα2-6(Neu5Acα2-3Galβ1-3)GalNAcα-
    ii. Neu5Acα2-6(Galβ1-3)GalNAcα-

[00129] The present invention provides binding agents with designated binding specificity, and also provides binding agents with designated binding characteristics with respect to umbrella glycans.

[00130] Certain particular binding agents provided by the present invention are described in more detail below.

HA polypeptides

[00131] In some embodiments, binding agents in accordance with the invention are HA polypeptides. For example, the present invention provides isolated HA polypeptides with designated binding specificity, and also provides engineered HA polypeptides with designated binding characteristics with respect to umbrella glycans.

[00132] In some embodiments, provided HA polypeptides with designated binding characteristics are H1 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H2 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H3 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H4 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H5 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H6 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H7 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H8 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding

characteristics are H9 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H10 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H11 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H12 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H13 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H14 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H15 polypeptides. In some embodiments, HA polypeptides in accordance with the invention with designated binding characteristics are H16 polypeptides.

[00133] In some embodiments, HA polypeptides in accordance with the invention do not include the H1 protein from any of the strains: A/South Carolina/1/1918; A/Puerto Rico/8/1934; A/Taiwan/1/1986; A/Texas/36/1991; A/Beijing/262/1995; A/Johannesburg/92/1996; A/New Caledonia/20/1999; A/Solomon Islands/3/2006.

[00134] In some embodiments, HA polypeptides in accordance with the invention are not the H2 protein from any of the strains of the Asian flu epidemic of 1957-58). In some embodiments, HA polypeptides in accordance with the invention do not include the H2 protein from any of the strains: A/Japan/305+/1957; A/Singapore/1/1957; A/Taiwan/1/1964; A/Taiwan/1/1967.

[00135] In some embodiments, HA polypeptides in accordance with the invention do not include the H3 protein from any of the strains: A/Aichi/2/1968; A/Philippines/2/1982; A/Mississippi/1/1985; A/Leningrad/360/1986; A/Sichuan/2/1987; A/Shanghai/11/1987; A/Beijing/353/1989; A/Shandong/9/1993; A/Johannesburg/33/1994; A/Nanchang/813/1995; A/Sydney/5/1997; A/Moscow/10/1999; A/Panama/2007/1999; A/Wyoming/3/2003; A/Oklahoma/323/2003; A/California/7/2004; A/Wisconsin/65/2005.

Variant HA polypeptides

[00136] In some embodiments, a provided HA polypeptide is a variant of a parent HA polypeptide in that its amino acid sequence is identical to that of the parent HA but for a small number of particular sequence alterations. In some embodiments, the parent HA is an HA polypeptide found in a natural isolate of an influenza virus (*e.g.*, a wild type HA polypeptide).

[00137] In some embodiments, HA polypeptide variants in accordance with the invention have different glycan binding characteristics than their corresponding parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for umbrella glycans (*e.g.*, as compared with for cone glycans) than do their cognate parent HA polypeptides. In some embodiments, such HA polypeptide variants are engineered variants.

[00138] In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for umbrella glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have reduced affinity and/or specificity for cone topology glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for umbrella glycans and reduced affinity and/or specificity for cone topology glycans as compared with their cognate parent HA polypeptides.

[00139] In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for α2-6 glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have reduced affinity and/or specificity for α2-3 glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for α2-6 and α2-3 glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for α2-6 glycans and reduced affinity and/or specificity for α2-3 glycans as compared with their cognate parent HA polypeptides.

[00140] In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for umbrella topology glycans and α2-3 glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for umbrella topology glycans and reduced affinity and/or specificity for α2-3 glycans as compared with their cognate parent HA polypeptides.

[00141] In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for α2-6 glycans and cone topology glycans as compared with their cognate parent HA polypeptides. In some embodiments, HA variant polypeptides in accordance with the invention have greater affinity and/or specificity for α2-6 glycans and reduced affinity and/or specificity for cone topology glycans as compared with their cognate parent HA polypeptides.

[00142] In some embodiments, HA polypeptide variants in accordance with the invention contain one or more sequence alterations that are consistent with HA sequences found in a different HA subtype. To give but one particular example, in some embodiments, an H5 HA polypeptide variant in accordance with the invention contains one or more sequence alterations which make the H5 HA polypeptide variant more closely resemble an H2 HA polypeptide. To give another particular example, in some embodiments, an H5 HA polypeptide variant in accordance with the invention contains one or more sequence alterations which make the H5 HA polypeptide variant more closely resemble an H1 HA polypeptide.

[00143] The present invention particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with altered glycosylation can show 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 1000-fold, 10,000-fold, or greater affinity to human HA receptors as compared with a reference HA polypeptide. The present invention also particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with alterations in the HA loop region, can show 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 1000-fold, 10,000-fold, or greater affinity to human HA receptors as compared with a reference HA polypeptide (*e.g.*, an HA polypeptide of any of SEQ ID NOs: 43-55).

[00144] The present invention particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with altered glycosylation can show 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 1000-fold, 10,000-fold, or greater specificity for human HA receptors as compared with a reference HA polypeptide. The present invention also particularly encompasses the recognition that HA polypeptide variants (*e.g.*, H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, or H16 HA polypeptide variants) with alterations in the

HA loop region, can show 2-fold, 3-fold, 4-fold, 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 1000-fold, 10,000-fold, or greater specificity for human HA receptors as compared with a reference HA polypeptide (*e.g.*, an HA polypeptide of any of SEQ ID NOs: 43-55).

[00145] In some embodiments, the reference HA polypeptide is the HA from A/South Carolina/1/18 (H1N1) (SEQ ID NO: 43). In some embodiments, the reference HA polypeptide is the HA from A/Brisbane/59/07 (SEQ ID NO: 44). In some embodiments, the reference HA polypeptide is the HA from A/California/04/09 (SEQ ID NO: 45). In some embodiments, the reference HA polypeptide is the HA from A/Albany/6/58 (H2N2) (SEQ ID NO: 46). In some embodiments, the reference HA polypeptide is the HA from A/Aichi/1/68 (SEQ ID NO: 47). In some embodiments, the reference HA polypeptide is the HA from A/Moscow/10/99 (SEQ ID NO: 48). In some embodiments, the reference HA polypeptide is the HA from A/Perth/16/09 (SEQ ID NO: 49). In some embodiments, the reference HA polypeptide is the HA from A/Vietnam/1203/04 (SEQ ID NO: 50). In some embodiments, the reference HA polypeptide is the HA from A/Egypt/2786-NAMRU3/06 (SEQ ID NO: 51). In some embodiments, the reference HA polypeptide is the HA from A/New York/107/03 (SEQ ID NO: 52). In some embodiments, the reference HA polypeptide is the HA from A/Hongkong/486/97 (SEQ ID NO: 53). In some embodiments, the reference HA polypeptide is the HA from A/Hongkong/213/03 (SEQ ID NO: 54). In some embodiments, the reference HA polypeptide is the HA from A/Indonesia/5/05 (SEQ ID NO: 55).

[00146] In some embodiments, HA polypeptide variants with altered glycan binding characteristics have one or more sequence alternations in residues within or affecting the glycan binding site. In some embodiments, such substitutions are of amino acids that interact directly with bound glycan; in some embodiments, such substitutions are of amino acids that are one degree of separation removed from those that interact with bound glycan, in that the one degree of separation removed–amino acids either (1) interact with the direct-binding amino acids; (2) otherwise affect the ability of the direct-binding amino acids to interact with glycan, but do not interact directly with glycan themselves; or (3) otherwise affect the ability of the direct-binding amino acids to interact with glycan, and also interact directly with glycan themselves. HA polypeptide variants in accordance with the invention contain substitutions of one or more direct-binding amino acids, one or more first degree of separation–amino acids, one or more second degree of separation–amino acids, or any combination of these. In some embodiments, HA

polypeptide variants in accordance with the invention may contain substitutions of one or more amino acids with even higher degrees of separation.

[00147] In some embodiments, HA polypeptide variants with altered glycan binding characteristics have sequence alterations in residues that make contact with sugars beyond Neu5Ac and Gal (see, for example, **Figure 7**).

[00148] In some embodiments, HA polypeptide variants have at least one amino acid substitution, as compared with a wild type parent HA. In some embodiments, HA polypeptide variants in accordance with the invention have at least two, three, four, five or more amino acid substitutions as compared with a cognate wild type parent HA; in some embodiments HA polypeptide variants in accordance with the invention have two, three, or four amino acid substitutions. In some embodiments, all such amino acid substitutions are located within the glycan binding site.

[00149] In some embodiments, HA polypeptide variants in accordance with the invention contain one or more amino acid substitutions as described in any of U.S. Patent Publication Number 2009/0269342 and 2010/0004195, and in U.S. Patent Application Serial Number 12/829931, filed July 2, 2010, entitled "COMPOSITIONS AND METHODS FOR DIAGNOSING AND/OR TREATING INFLUENZA INFECTION" (all of which are incorporated herein by reference).

[00150] In some embodiments, HA polypeptide variants have sequence substitutions at positions corresponding to one or more of residues 95, 98, 128, 130, 131, 132, 133, 135, 136, 137, 138, 145, 153, 155, 156, 158, 159, 160, 183, 186, 187, 188, 189, 190, 192, 193, 194, 195, 196, 219, 221, 222, 224, 225, 226, 227, and 228. In some embodiments, HA polypeptide variants, particularly H5 polypeptide variants, have one or more amino acid substitutions relative to a wild type parent HA (*e.g.*, H5) at residues selected from the group consisting of residues 95, 98, 128, 130, 131, 132, 133, 135, 136, 137, 138, 145, 153, 155, 156, 158, 159, 160, 183, 186, 187, 188, 189, 190, 192, 193, 194, 195, 196, 219, 221, 222, 224, 225, 226, 227, and 228. In some embodiments, HA polypeptide variants, particularly H5 polypeptide variants, have one or more amino acid substitutions relative to a wild type parent HA (*e.g.*, H5) at any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, or 37 residues selected from the group consisting of residues 95, 98, 128, 130, 131,

132, 133, 135, 136, 137, 138, 145, 153, 155, 156, 158, 159, 160, 183, 186, 187, 188, 189, 190, 192, 193, 194, 195, 196, 219, 221, 222, 224, 225, 226, 227, and 228.

[00151] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that reduce or abolish glycosylation at a site corresponding to amino acid position 158. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that affect and/or alter the identity and/or structure of the glycan linked to a site corresponding to amino acid position 158. In some embodiments, such a sequence substitution is a mutation at a site corresponding to position 158, *e.g.*, Asn158Xaa, wherein Xaa is any amino acid other than Asn. In some embodiments, such a sequence substitution is a mutation at a site corresponding to position 160, *e.g.*, Thr160Xaa, wherein Xaa is any amino acid other than Asn. In some embodiments, such a sequence substitution comprises the mutation Thr160Ala. In some embodiments, a sequence substitution that reduces, abolishes, affects, or alters glycosylation at a site corresponding to amino acid position 158 can make a non-H2 HA polypeptide (*e.g.*, an H5 HA polypeptide) more closely resemble (*e.g.*, both structurally and functionally) an H2 HA polypeptide. In some embodiments, a mutation at a site corresponding to position 160 (*e.g.*, Thr160Xaa, such as Thr160Ala) can make a non-H2 HA polypeptide (*e.g.*, an H5 HA polypeptide) more closely resemble (*e.g.*, both structurally and functionally) an H2 HA polypeptide.

[00152] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 226, 228, and 160. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226, 228, and 160. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226 and 160. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 228 and 160. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226 and 228.

[00153] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 226, 228, and 158. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226, 228, and 158. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226 and 158. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 228 and 158. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to residues 226 and 228.

[00154] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion in one or more of the loop regions of an HA polypeptide. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion at a site corresponding to the 128-137 loop region of an HA polypeptide. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion at one or more of amino acid positions corresponding to residues 128, 129, 130, 131, 132, 133, 134, 135, 136, and/or 137 of an HA polypeptide. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion at a site corresponding to the 128-134 loop region of an HA polypeptide. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion at one or more of amino acid positions corresponding to residues 128, 129, 130, 131, 132, 133, and/or 134 of an HA polypeptide. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions that include a deletion of an amino acid corresponding to residue 130. In some embodiments, such loop region substitutions can make a non-H2 HA polypeptide (*e.g.*, an H5 HA polypeptide) more closely resemble (*e.g.*, both structurally and functionally) an H2 HA polypeptide. In some embodiments, a deletion of an amino acid corresponding to residue 130 can make a non-H2 HA polypeptide (*e.g.*, an H5 HA

polypeptide) more closely resemble (*e.g.*, both structurally and functionally) an H2 HA polypeptide.

[00155] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, or 14 of residues 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 of residues 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, and 228.

[00156] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 132, 135, 188, 192, 221, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, or 7 of residues 131, 132, 135, 188, 192, 221, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 132, 135, 188, 192, and 221. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, or 6 of residues 131, 132, 135, 188, 192, and 221.

[00157] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 133, 137, 155, 193, 226, 227, 228, and 130. In some

embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, or of residues 133, 137, 155, 193, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 133, 137, 155, 193, 226, 227, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, or 7 of residues 133, 137, 155, 193, 226, 227, and 228.

[00158] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 130, 192, and 193. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, or 3 of residues 130, 192, 193. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or both of residues 192 and 193.

[00159] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16 of residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 of residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, and 228.

[00160] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 137, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, or 7 of residues 137, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 137, 188, 192, 193, 226, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, or 6 of residues 137, 188, 192, 193, 226, and 228.

[00161] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 137, 188, 192, 193, 226, 227, 228, 131, 132, 133, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 of residues 137, 188, 192, 193, 226, 227, 228, 131, 132, 133, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 137, 188, 192, 193, 226, 227, 228, 131, 132, and 133. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 of residues 137, 188, 192, 193, 226, 227, 228, 131, 132, and 133.

[00162] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 227, 131, 132, 133, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, or 5 of residues 227, 131, 132, 133, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5

HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 227, 131, 132, and 133. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, or 4 of residues 227, 131, 132, and 133.

[00163] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 133, 137, 155, 188, 192, 193, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 of residues 131, 133, 137, 155, 188, 192, 193, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 133, 137, 155, 188, 192, 193, 226, 227, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 of residues 131, 133, 137, 155, 188, 192, 193, 226, 227, and 228.

[00164] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 133, 137, 155, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 of residues 131, 133, 137, 155, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 133, 137, 155, 188, 192, 193, 226, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, or 9 of residues 131, 133, 137, 155, 188, 192, 193, 226, and 228.

[00165] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13 of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 227, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 227, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 227, and 228.

[00166] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 228, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, and 228. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 of residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, and 228.

[00167] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 137, 188, 192, 193, 226, 228, 131, 132, 133, 221, 227, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions

corresponding to any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 of residues 137, 188, 192, 193, 226, 228, 131, 132, 133, 221, 227, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 137, 188, 192, 193, 226, 228, 131, 132, 133, 221, and 227. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11 of residues 137, 188, 192, 193, 226, 228, 131, 132, 133, 221, and 227.

[00168] In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to one or more of residues 131, 132, 133, 221, 227, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to any 1, 2, 3, 4, 5, or 6 of residues 131, 132, 133, 221, 227, and 130. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) one or more of residues 131, 132, 133, 221, and 227. In some embodiments, HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) have sequence substitutions relative to a wild type parent HA (*e.g.*, H5 HA) at positions corresponding to (1) 130, and (2) any 1, 2, 3, 4, or 5 of residues 131, 132, 133, 221, and 227.

[00169] In some embodiments, an HA polypeptide variant, and particularly an H5 polypeptide variant, has one or more amino acid substitutions relative to a wild type parent HA at residues selected from amino acids located in the region of the receptor that directly binds to the glycan, including but not limited to residues 98, 136, 153, 155, 183, and 194. In some embodiments, an HA polypeptide variant in accordance with the invention, and particularly an H5 polypeptide variant, has one or more amino acid substitutions relative to a wild type parent HA at residues selected from amino acids located adjacent to the region of the receptor that directly binds the glycan, including but not limited to (a) residues 98 and 195, (b) residues 98, 138, 186, 187, 195, and 228), or (c) residues 138, 186, 187, and 228.

[00170] In some embodiments, an HA polypeptide variant in accordance with the invention, and particularly an H5 variant has one or more of the following amino acid substitutions:

Ser132Thr, Ala133Thr, Ser133Thr, Ser137Ala, Ser137Arg, Ile155Thr, Lys156Glu, Asn158Xaa (wherein Xaa = any amino acid besides Asn), Thr160Ala, Asn186Pro, Asp187Ser, Asp187Thr, Ala188Glu, Ala188Asp, Ala189Gln, Ala189Lys, Ala189Thr, Glu190Asp, Glu190Thr, Thr192Arg/Lys, Lys193Arg, Lys193Asn, Lys193His, Lys193Ser, Lys/Arg193Thr/Ala/Met/Val, Ser221Pro, Gly225Asp, Gln226Ile, Gln226Leu, Gln226Val, Ser227Ala, Gly228Ser.

[00171] In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) in accordance with the invention has an amino acid substitution at a position corresponding to residue 192, which switches the charge at that position. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) in accordance with the invention has an amino acid substitution at a position corresponding to residue 193, which switches the charge at that position. For example, in some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) in accordance with the invention has a Thr or a hydrophobic residue (*e.g.*, Val or Ile) at a position corresponding to residue 192, and an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a hydrophilic residue at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a hydrophilic residue at a position corresponding to residue 192. To give another example, in some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) in accordance with the invention has a Thr or a hydrophobic residue (*e.g.*, Val or Ile) at a position corresponding to residue 192, and an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a basic residue (*e.g.*, Lys or Arg) at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a basic residue (*e.g.*, Lys or Arg) at a position corresponding to residue 192. To give yet another example, in some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) in accordance with the invention has a basic residue (*e.g.*, Lys or Arg) at a position corresponding to residue 193, and an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a neutral or acidic residue at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) (*e.g.*, a human-adapted variant) has a neutral or acidic residue at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA

polypeptide variant) (*e.g.*, a human-adapted variant) has a Thr, Ala, Met, or Val at a position corresponding to residue 193.

[00172] In some embodiments, human adaptation of HA polypeptides (*e.g.*, H5 HA polypeptides) is associated with the propert(ies) of the residue at position 188. In H5 HA, residue 188 is frequently Ala, which makes contacts with Thr or a hydrophobic residue at 192. In contrast, in H2 HA, residue 188 is frequently Glu or Asp, which makes contacts with Arg or Lys at 192. Hence, in some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at position 188. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at position 188. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Ala188Glu substitution. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Ala188Asp substitution.

[00173] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 188 and a Thr at a position corresponding to residue 192. In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 188 and a hydrophobic residue at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188 and an Arg at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188 and an Arg at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188 and a Lys at a position corresponding to residue 192. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188 and a Lys at a position corresponding to residue 192.

[00174] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 188, a Thr at a position corresponding to residue 192, and a Lys at a position corresponding to residue 193. In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 188, a hydrophobic residue at a position corresponding to residue 192, and a Lys at a position corresponding to residue 193. In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala

at a position corresponding to residue 188, a Thr at a position corresponding to residue 192, and an Arg at a position corresponding to residue 193. In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 188, a hydrophobic residue at a position corresponding to residue 192, and an Arg at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188, an Arg at a position corresponding to residue 192, and a Thr at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188, an Arg at a position corresponding to residue 192, and a Thr at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188, a Lys at a position corresponding to residue 192, and a Thr at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188, a Lys at a position corresponding to residue 192, and a Thr at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188, an Arg at a position corresponding to residue 192, and a Thr, Ala, Met, or Val at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188, an Arg at a position corresponding to residue 192, and a Thr, Ala, Met, or Val at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Glu at a position corresponding to residue 188, a Lys at a position corresponding to residue 192, and a Thr, Ala, Met, or Val at a position corresponding to residue 193. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Asp at a position corresponding to residue 188, a Lys at a position corresponding to residue 192, and a Thr, Ala, Met, or Val at a position corresponding to residue 193.

[00175] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ala at a position corresponding to residue 131. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Thr at a position corresponding to residue 131.

[00176] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Ser at a position corresponding to residue 132. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Thr at a position corresponding to residue 132.

[00177] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Ser at a position corresponding to residue 133. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Thr at a position corresponding to residue 133.

[00178] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) includes Ala, Thr, and/or Ser at any position corresponding to residues 131, 132, and/or 133. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) includes Ala, Thr, and/or Ser at any position corresponding to residues 131, 132, and/or 133. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) includes Thr at all of positions corresopnding to 131, 132, and 133.

[00179] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Val at a position corresponding to residue 135. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has any amino acid other than Val at a position corresponding to residue 135.

[00180] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Ser at a position corresponding to residue 137. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has an Arg at a position corresponding to residue 137.

[00181] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has an Ile at a position corresponding to residue 155. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Thr at a position corresponding to residue 155. In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) includes a Thr at a position corresponding to residue 155. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) includes a Thr at a position corresponding to residue 155.

[00182] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Ser at a position corresponding to residue 221. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Pro at a position corresponding to residue 221.

[00183] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) includes a Ser at a position corresponding to residue 221. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) includes a Pro at a position corresponding to residue

221. Without wishing to be bound by any one particular theory, Pro221 might influence conformation of 220 loop which is involved with the RBS of H2 HA.

[00184] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Gln at a position corresponding to residue 226. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Leu at a position corresponding to residue 226.

[00185] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Ser at a position corresponding to residue 227. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Gly at a position corresponding to residue 227.

[00186] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) has a Gly at a position corresponding to residue 228. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) has a Ser at a position corresponding to residue 228.

[00187] In some embodiments, an HA polypeptide (*e.g.*, an H5 HA polypeptide) includes Gln, Ser, and Gly residues at positions corresponding to residues 226, 227, and 228, respectively. In some embodiments, an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) includes a Leu, Gly, and Ser at positions corresponding to residues 226, 227, and 228, respectively.

[00188] In some embodiments, an HA polypeptide variant in accordance with the invention, and particularly an H5 variant has one or more of the following amino acids at the indicated positions (the numbering of these positions corresponds to the numbering of H3 HA):

- Glu190Asp, Lys193Ser, Gly225Asp, Gln226Leu
- Glu190Asp, Lys193Ser, Gln226Leu, Gly228Ser
- Ala189Gln, Lys193Ser, Thr160Ala
- Ala189Gln, Lys193Ser, Gln226Leu, Gly228Ser
- Asp187Ser/Thr, Ala189Gln, Lys193Ser, Gln226Leu, Gly228Ser
- Ala189Lys, Lys193Asn, Gln226Leu, Gly228Ser
- Asp187Ser/Thr, Ala189Lys, Lys193Asn, Gln226Leu, Gly228Ser
- Lys156Glu, Ala189Lys, Lys193Asn, Gln226Leu, Gly228Ser
- Lys193His, Gln226Leu/Ile/Val, Gly228Ser
- Lys193Arg, Gln226Leu/Ile/Val, Gly228Ser
- Ala189Lys, Lys193Asn, Gly225Asp
- Lys156Glu, Ala189Lys, Lys193Asn, Gly225Asp
- Ser137Ala, Lys156Glu, Ala189Lys, Lys193Asn, Gly225Asp

- Glu190Thr, Lys193Ser, Gly225Asp
- Asp187Thr, Ala189Thr, Glu190Asp, Lys193Ser, Gly225Asp
- Asn186Pro, Asp187Thr, Ala189Thr, Glu190Asp, Lys193Ser, Gly225Asp
- Asn186Pro, Asp187Thr, Ala189Thr, Glu190Asp, Lys193Ser, Gly225Asp, Ser227Ala
- Gln226Leu, Gly228Ser, Thr160Ala
- Gln226Leu, Gly228Ser, Thr160Ala
- Gly228Ser, Thr160Ala
- Gln226Leu, Thr160Ala
- Gln226Leu, Gly228Ser
- Thr160Ala
- Gln226Leu, Gly228Ser, Asn158Xaa (wherein Xaa = any amino acid besides Asn)
- Gly228Ser, Asn158Xaa
- Gln226Leu, Asn158Xaa
- Gln226Leu, Gly228Ser
- Asn158Xaa
- Δ130 (wherein "Δ130" indicates a deletion at an amino acid corresponding to position 130) plus any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 188, 192, 193, 221, 226, 227, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 132, 135, 188, 192, and 221
- Δ130 plus any possible combination of mutations at positions corresponding to residues 133, 137, 155, 193, 226, 227, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 133, 137, 155, 188, 192, 193, 226, 227, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 133, 137, 155, 188, 192, 193, 226, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, 227, and 228

- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 133, 137, 155, 159, 160, 188, 192, 193, 226, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 137, 188, 192, 193, 226, 228, 131, 132, 133, 221, and 227
- Δ130 plus any possible combination of mutations at positions corresponding to residues 131, 132, 133, 221, and 227
- Δ130 plus any possible combination of mutations at positions corresponding to residues 137, 188, 192, 193, 226, and 228
- Δ130 plus any possible combination of mutations at positions corresponding to residues 137, 188, 192, 193, 226, 227, 228, 131, 132, and 133
- Δ130 plus any possible combination of mutations at positions corresponding to residues 227, 131, 132, and 133
- Gln226Leu, Gly228Ser, Thr160Ala, Δ130
- Gln226Leu, Gly228Ser, Δ130
- Gln226Leu, Thr160Ala, Δ130
- Gly228Ser, Thr160Ala, Δ130
- Gln226Leu, Δ130
- Gly228Ser, Δ130
- Thr160Ala, Δ130
- Δ130
- Δ130, Ala131Thr, Leu133Thr, Ser137Arg, Ile155Thr, Ala188Glu, Thr/Ile192Arg/Lys, Arg/Lys193Thr/Ala, Gln226Leu, Ser227Gly, Gly228Ser
- Δ130, Ala131Thr, Leu133Thr, Ser137Arg, Ile155Thr, Ala188Glu, Thr/Ile192Arg/Lys, Arg/Lys193Thr/Ala, Gln226Leu, Gly228Ser
- Δ130, Ala131Thr, Leu133Thr, Ser137Arg, Ile155Thr, Asn159Asp (or Thr160Ala or both), Ala188Glu, Thr/Ile192Arg/Lys, Arg/Lys193Thr/Ala, Gln226Leu, Ser227Gly, Gly228Ser
- Δ130, Ala131Thr, Leu133Thr, Ser137Arg, Ile155Thr, Asn159Asp (or Thr160Ala or both), Ala188Glu, Thr/Ile192Arg/Lys, Arg/Lys193Thr/Ala, Gln226Leu, Gly228Ser

- Δ130, Ser137Arg, Ala188Glu, Thr192Arg/Lys, Arg/Lys193Thr/Met/Ala/Val, Gln226Leu, Gly228Ser
- Δ130, Ser137Arg, Ala188Glu, Thr192Arg/Lys, Arg/Lys193Thr/Met/Ala/Val, Gln226Leu, Gly228Ser, Xaa131Ser/Thr, Xaa132Ser/Thr, Xaa133Ser/Thr, Ser221Pro, Ser227Gly (wherein Xaa = any amino acid)
- Δ130, Xaa131Ser/Thr, Xaa132Ser/Thr, Xaa133Ser/Thr, Ser221Pro, Ser227Gly (wherein Xaa = any amino acid)
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid)
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue)
- Δ130, Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue)
- Δ130, Lys/Arg193Thr/Ala/Met/Val
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue)
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue)
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Lys/Arg193Thr/Ala/Met/Val
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Lys/Arg193Thr/Ala/Met/Val
- Δ130, Ala188Glu
- Δ130, Ala188Asp
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Ala188Glu
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Ala188Glu
- Δ130, Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Glu
- Δ130, Lys/Arg193Thr/Ala/Met/Val, Ala188Glu

- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Ala188Asp
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Ala188Asp
- Δ130, Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Asp
- Δ130, Lys/Arg193Thr/Ala/Met/Val, Ala188Asp
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Glu
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Glu
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Lys/Arg193Thr/Ala/Met/Val, Ala188Glu
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Lys/Arg193Thr/Ala/Met/Val, Ala188Glu
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Asp
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Xaa193Xaa' (wherein Xaa = a basic residue, *e.g.*, Lys or Arg, and Xaa' = a neutral or acidic residue), Ala188Asp
- Δ130, Xaa192Xaa' (wherein Xaa = any hydrophobic amino acid, and Xaa' = any hydrophilic amino acid), Lys/Arg193Thr/Ala/Met/Val, Ala188Asp
- Δ130, Xaa192Lys/Arg (wherein Xaa = any hydrophobic residue), Lys/Arg193Thr/Ala/Met/Val, Ala188Asp

[00189] In some embodiments, the present invention provides HA polypeptides (*e.g.*, HA polypeptide variants, engineered HA polypeptides, and/or engineered HA polypeptide variants) whose amino acid sequence includes an element as set forth below:

- X190, X193, X225 and X226
- X190, X193, X226 and X228

- X189, X193, X160
- X189, X193, X226, X228
- X187, X189, X193, X226, X228
- X189, X193, X226, X228
- X187, X189, X193, X226, X228
- X156, X189, X193, X226, X228
- X193, X226, X228
- X193, X226, X228
- X189, X193, X225
- X156, X189, X193, X225
- X137, X156, X189, X193, X225
- X190, X193, X225
- X187, X189, X190, X193, X225
- X186, X187, X189, X190, X193, X225
- X186, X187, X189, X190, X193, X225, X227
- X226, X228, X160
- X226, X228, X160
- X228, X160
- X226, X160
- X226, X228
- X160
- X226, X228, Xaa158 (wherein Xaa = any amino acid besides Asn)
- X228, Xaa158 (wherein Xaa = any amino acid besides Asn)
- X226, Xaa158 (wherein Xaa = any amino acid besides Asn)
- X226, X228
- X158 (wherein Xaa = any amino acid besides Asn)
- X130 plus any possible combination of X131, X132, X133, X135, X137, X155, X188, X192, X193, X221, X226, X227, and X228
- X130 plus any possible combination of X131, X132, X135, X188, X192, and X221

- X130 plus any possible combination of X133, X137, X155, X193, X226, X227, and X228
- X130 plus any possible combination of X131, X132, X133, X135, X137, X155, Xaa158 (wherein Xaa = any amino acid besides Asn), X160, X188, X192, X193, X221, X226, X227, and X228
- X130 plus any possible combination of X131, X133, X137, X155, X188, X192, X193, X226, X227, and X228
- X130 plus any possible combination of X131, X133, X137, X155, X188, X192, X193, X226, and X228
- X130 plus any possible combination of X131, X133, X137, X155, X159, X160, X188, X192, X193, X226, X227, and X228
- X130 plus any possible combination of X131, X133, X137, X155, X159, X160, X188, X192, X193, X226, and X228
- X130 plus any possible combination of X137, X188, X192, X193, X226, X228, X131, X132, X133, X221, and X227
- X130 plus any possible combination of X131, X132, X133, X221, and X227
- X130 plus any possible combination of X137, X188, X192, X193, X226, and X228
- X130 plus any possible combination of X137, X188, X192, X193, X226, X227, X228, X131, X132, and X133
- X130 plus any possible combination of X227, X131, X132, and X133
- X226, X228, X160, X130
- X226, X228, X130
- X226, X160, X130
- X228, X160, X130
- X226, X130
- X228, X130
- X160, X130
- X130
- X130, X131, X133, X137, X155, X188, X192, X193, X226, X227, X228
- X130, X131, X133, X137, X155, X188, X192, X193, X226, X228

- X130, X131, X133, X137, X155, X159, X160, X188, X192, X193, X226, X227, X228
- X130, X131, X133, X137, X155, X159, X160, X188, X192, X193, X226, X228
- X130, X137, X188, X192, X193, X226, X228
- X130, X137, X188, X192, X193, X226, X228, X131, X132, X133, X221, X227
- X130, X131, X132, X133, X221, X227
- X130, X192
- X130, X193
- X130, X192, X193
- X130, X188
- X130, X192, X188
- X130, X193, X188
- X130, X192, X193, X188

wherein X = any amino acid (unless otherwise specified above), and/or X = a missing amino acid. The numbering of these positions corresponds to the numbering of H3 HA.

[00190] In some embodiments X130 is a deletion at position 130. In some embodiments, X160 is an Ala. In some embodiments, X158 is any amino acid other than Asn.

[00191] In some such embodiments, the H5 HA polypeptide variant has at least one further substitution as compared with a wild type H5 HA, such that affinity and/or specificity of the variant for umbrella glycans is increased.

[00192] In some such embodiments, the HA polypeptide has at least one further substitution as compared with a wild type HA, such that affinity and/or specificity of the variant for umbrella glycans is increased.

[00193] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include L226, S228, and A160. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include L226 and A160. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include S228 and A160. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include A160.

[00194] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include L226, S228, and X158 (wherein X =

any amino acid besides Asn). In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include L226 and X158. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include S228 and X158. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include X158.

[00195] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130 and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, 227, and 228.

[00196] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, S228, A160 and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 188, 192, 193, 221, and 227. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, A160, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 188, 192, 193, 221, 227, and 228. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, S228, A160, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 188, 192, 193, 221, and 227. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, S228, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, and 227.

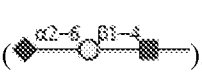
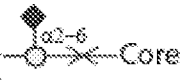
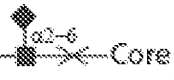
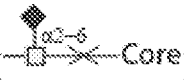
[00197] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, A160, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 188, 192, 193, 221, 226, 227, and 228. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, and 227, and 228. In some

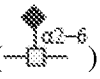
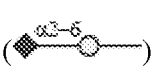
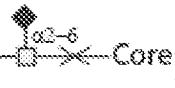
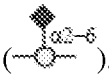
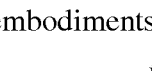
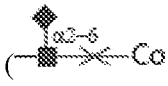
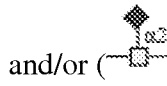
embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, S228, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, 226, and 227.

[00198] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, S228, X158 (wherein X = any amino acid besides Asn) and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 160, 188, 192, 193, 221, and 227. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, X158, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 160, 188, 192, 193, 221, 227, and 228. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, S228, X158, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 160, 188, 192, 193, 221, and 227. In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, L226, S228, and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 158, 160, 188, 192, 193, 221, and 227.

[00199] In some embodiments, HA polypeptides in accordance with the invention (including H5 HA polypeptide variants) have sequences that include Δ130, X158 (wherein X = any amino acid besides Asn), and any possible combination of mutations at positions corresponding to residues 131, 132, 133, 135, 137, 155, 160, 188, 192, 193, 221, 226, 227, and 228.

[00200] In some embodiments, H5 HA polypeptide variants (*e.g.*, H5 HA polypeptide variants) in accordance with the invention have an open binding site as compared with a parent HA, and particularly with a parent wild type HAs.

[00201] In some embodiments, HA polypeptides (*e.g.*, H5 HA polypeptides) in accordance with the invention bind to the following α2-6 sialylated glycans: (), (), (), (), (), (), (), (), and combinations thereof. In some embodiments, H5 HA polypeptides in accordance with the

invention bind to glycans of the structure: (), (), () and combinations thereof; and/or (); and/or (); and/or (), (), (), and combinations thereof. In some embodiments, HA polypeptides (*e.g.*, H5 HA polypeptides) in accordance with the invention bind to (), (), and/or (); in some embodiments to (); in some embodiments to (); and in some embodiments to (), (), and/or ().

[00202] In some embodiments, HA polypeptides (*e.g.*, H5 HA polypeptides) in accordance with the invention bind to umbrella topology glycans. In some embodiments, H5 HA polypeptides in accordance with the invention bind to at least some of the glycans (*e.g.*, α2-6 sialylated glycans) depicted in **Figure 9**. In some embodiments, HA polypeptides (*e.g.*, H5 HA polypeptides) in accordance with the invention bind to multiple glycans depicted in **Figure 9**.

[00203] In some embodiments, HA polypeptides (*e.g.*, H5 HA polypeptides) in accordance with the invention bind to at least about 10%, about 15%, about 20%, about 25%, about 30% about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90% about 95%, or more of the glycans found on HA receptors in human upper respiratory tract tissues (*e.g.*, epithelial cells).

[00204] In one aspect, the present invention provides the particular recognition that high affinity binding to umbrella-topology glycans alone may not be sufficient to confer effective transmission to/infectivity of humans. The present invention provides the insight that reduced binding to cone-topology glycans may also be important. In some embodiments, high affinity binding to umbrella-topology glycans and reduced affinity binding to cone-topology glycans may be involved in conferring effective transmission to/infectivity of humans. In some embodiments, high affinity binding to umbrella-topology glycans is sufficient to confer effective transmission to/infectivity of humans. In some embodiments, high affinity binding to umbrella-topology glycans is sufficient to confer effective transmission to/infectivity of humans, even if the affinity of binding to cone-topology glycans is not reduced (*e.g.*, unchanged, increased, *etc.*).

[00205] In some embodiments, increased affinity and/or specificity of binding of an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) to umbrella-topology glycans and reduced affinity and/or specificity binding to cone-topology glycans may be involved in increasing or enhancing transmission to/infectivity of humans relative to a reference polypeptide (*e.g.*, the HA polypeptide variant's cognate parent HA polypeptide). In some embodiments, increased affinity and/or specificity of binding of an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) to umbrella-topology glycans is sufficient to increase or enhance transmission to/infectivity of humans relative to a reference polypeptide (*e.g.*, the HA polypeptide variant's cognate parent HA polypeptide). In some embodiments, increased affinity and/or specificity of binding of an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) to umbrella-topology glycans is sufficient to increase or enhance transmission to/infectivity of humans relative to a reference polypeptide (*e.g.*, the HA polypeptide variant's cognate parent HA polypeptide), even if the affinity and/or specificity of binding to cone-topology glycans is not reduced (*e.g.*, unchanged, increased, *etc.*). In some embodiments, increased affinity and/or specificity of binding of an HA polypeptide variant (*e.g.*, an H5 HA polypeptide variant) to umbrella-topology glycans is sufficient to increase or enhance transmission to/infectivity of humans relative to a reference polypeptide (*e.g.*, the HA polypeptide variant's cognate parent HA polypeptide), even if the affinity and/or specificity of binding to cone-topology glycans is equal to and/or greater than that of the affinity and/or specificity of binding to umbrella-topology glycans.

Portions or fragments of HA polypeptides

[00206] The present invention further provides characteristic portions (which may or may not be binding agents) of HA polypeptides in accordance with the invention and nucleic acids that encode them. In general, a characteristic portion is one that contains a continuous stretch of amino acids, or a collection of continuous stretches of amino acids, that together are characteristic of the HA polypeptide. Each such continuous stretch generally will contain at least two amino acids. Furthermore, those of ordinary skill in the art will appreciate that typically at least 5, at least 10, at least 15, at least 20 or more amino acids are required to be characteristic of a H5 HA polypeptide. In general, a characteristic portion is one that, in addition to the sequence identity specified above, shares at least one functional characteristic with the relevant intact HA

polypeptide. In some embodiments, characteristic portions of HA polypeptides in accordance with the invention share glycan binding characteristics with the relevant full-length HA polypeptides.

Non-HA polypeptides

[00207] In some embodiments, binding agents provided in accordance with the present invention are polypeptides whose amino acid sequence does not include a characteristic HA sequence. Such polypeptides are referred to herein as "Non-HA polypeptides". In some embodiments, a Non-HA polypeptide has an amino acid sequence selected in advance (*e.g., via* rational design, including for example, introduction of strategic amino acid alterations [*e.g.*, additions, deletions, and/or substitutions] as compared with a reference sequence). In some embodiments, a Non-HA polypeptide has an amino acid sequence that is determined stochastically and, for example, identified on the basis of the desirable binding characteristics defined herein.

Antibodies

[00208] In some embodiments, binding agents provided in accordance with the present invention are antibodies (*e.g.*, that bind to umbrella topology glycans and/or to umbrella topology glycan mimics). Antibodies suitable for the invention include antibodies or fragments of antibodies that bind immunospecifically to any umbrella topology glycan epitope. As used herein, the term "antibodies" is intended to include immunoglobulins and fragments thereof which are specifically reactive to the designated protein or peptide, or fragments thereof. Suitable antibodies include, but are not limited to, human antibodies, primatized antibodies, chimeric antibodies, bi-specific antibodies, humanized antibodies, conjugated antibodies (*i.e.*, antibodies conjugated or fused to other proteins, radiolabels, cytotoxins), Small Modular ImmunoPharmaceuticals ("SMIPs™"), single chain antibodies, cameloid antibodies, and antibody fragments. As used herein, the term "antibodies" also includes intact monoclonal antibodies, polyclonal antibodies, single domain antibodies (*e.g.*, shark single domain antibodies (*e.g.*, IgNAR or fragments thereof)), multispecific antibodies (*e.g.* bi-specific antibodies) formed from at least two intact antibodies, and antibody fragments so long as they exhibit the desired

biological activity. Antibody polypeptides for use herein may be of any type (*e.g.*, IgA, IgD, IgE, IgG, IgM).

[00209] As used herein, an "antibody fragment" includes a portion of an intact antibody, such as, for example, the antigen-binding or variable region of an antibody. Examples of antibody fragments include Fab, Fab', F(ab')2, and Fv fragments; triabodies; tetrabodies; linear antibodies; single-chain antibody molecules; and multi specific antibodies formed from antibody fragments. The term "antibody fragment" also includes any synthetic or genetically engineered protein that acts like an antibody by binding to a specific antigen to form a complex. For example, antibody fragments include isolated fragments, "Fv" fragments, consisting of the variable regions of the heavy and light chains, recombinant single chain polypeptide molecules in which light and heavy chain variable regions are connected by a peptide linker ("ScFv proteins"), and minimal recognition units consisting of the amino acid residues that mimic the hypervariable region.

[00210] Antibodies can be generated using methods well known in the art. For example, protocols for antibody production are described by Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*; incorporated herein by reference. Typically, antibodies can be generated in mouse, rat, guinea pig, hamster, camel, llama, shark, or other appropriate host. Alternatively, antibodies may be made in chickens, producing IgY molecules (Schade *et al.*, 1996, *ALTEX* 13(5):80-85; incorporated herein by reference). In some embodiments, antibodies suitable for the present invention are subhuman primate antibodies. For example, general techniques for raising therapeutically useful antibodies in baboons may be found, for example, in Goldenberg *et al.*, international patent publication number WO 91/11465, 1991; and in Losman *et al.*, 1990, *Int. J. Cancer* 46:310; both of which are incorporated herein by reference). In some embodiments, monoclonal antibodies may be prepared using hybridoma methods (Milstein and Cuello, 1983, *Nature* 305(5934):537-40; incorporated herein by reference). In some embodiments, monoclonal antibodies may also be made by recombinant methods (U.S. Patent Number 4,166,452, 1979; incorporated herein by reference).

[00211] In some embodiments, antibodies suitable for the invention may include humanized or human antibodies. Humanized forms of non-human antibodies are chimeric Igs, Ig chains or fragments (such as Fv, Fab, Fab', F(ab')2 or other antigen-binding subsequences of Abs) that contain minimal sequence derived from non-human Ig. Generally, a humanized antibody has

one or more amino acid residues introduced from a non-human source. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization is accomplished by substituting rodent complementarity determining regions (CDRs) or CDR sequences for the corresponding sequences of a human antibody (Riechmann *et al.*, 1988, *Nature* 332(6162):323-7; Verhoeyen *et al.*, 1988, *Science*. 239(4847):1534-6; both of which are incorporated herein by reference). Such "humanized" antibodies are chimeric Abs (U.S. Patent Number 4,816,567, 1989; incorporated herein by reference), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In some embodiments, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent Abs. Humanized antibodies include human Igs (recipient antibody) in which residues from a CDR of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit, having the desired specificity, affinity and capacity. In some instances, corresponding non-human residues replace Fv framework residues of the human Ig. Humanized antibodies may comprise residues that are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody comprises substantially all of at least one, and typically two, variable domains, in which most if not all of the CDR regions correspond to those of a non-human Ig and most if not all of the FR regions are those of a human Ig consensus sequence. The humanized antibody optimally also comprises at least a portion of an Ig constant region (Fc), typically that of a human Ig (Riechmann *et al.*, 1988, *Nature* 332(6162):323-7; Verhoeyen *et al.*, 1988, *Science*. 239(4847):1534-6; both of which are incorporated herein by reference).

[00212] Human antibodies can also be produced using various techniques, including phage display libraries (Hoogenboom *et al.*, 1991, *Mol Immunol.* 28(9):1027-37; Marks *et al.*, 1991, *J Mol Biol.* 222(3):581-97; both of which are incorporated herein by reference) and the preparation of human monoclonal antibodies (Reisfeld and Sell, 1985, *Cancer Surv.* 4(1):271-90; incorporated herein by reference). Similarly, introducing human Ig genes into transgenic animals in which the endogenous Ig genes have been partially or completely inactivated can be exploited to synthesize human antibodies. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene

rearrangement, assembly, and antibody repertoire (Fishwild *et al.*, 1996, *Nat Biotechnol.* 14(7):845-51; Lonberg *et al.*, 1994, *Nature* 368(6474):856-9; Lonberg and Huszar, 1995, *Int. Rev. Immunol.* 13(1):65-93; Marks *et al.*, 1992, *Biotechnology* (N Y). 10(7):779-83; all of which are incorporated herein by reference).

Lectins

[00213] In some embodiments, binding agents provided in accordance with the present invention are lectins. Lectins are sugar-binding proteins which may bind to a soluble carbohydrate or to a carbohydrate moiety which is a part of a glycoconjugate (*e.g.*, a glycopeptide or glycolipid). Lectins typically agglutinate certain animal cells and/or precipitate glycoconjugates by recognizing a particular sugar moiety. For example, SNA-1 is a lectin that has a high affinity for α2-6 sialic acids. As yet another example, polyporus squamosus lectins (PSL1a and PSL1b) have high affinity for binding sialylated glycoconjugates containing Neu5Acα2,6Galβ1,4Glc/GlcNAc trisaccharide sequences of asparagine-linked glycoproteins. Non-limiting exemplary lectins that may act as binding agents include SNA-1, SNA-1', PSL1a, PSL1b, and polypeptides derived therefrom.

[00214] Amino acid sequences of exemplary lectins are provided below:

[00215] ***Sambucus Nigra Lectin 1 (Genbank Accession No. U27122):***

MRLVAKLLYLAVLAICGLGIHGALTHPRVTPPVYPSVSFNLTGADTYEPFLRALQEKVILGNHTA
FDLPVLNPESQVSDSNRFVLVPLTNPSGDTVTLAIDVVNLYVVAFSSNGKSYFFSGSTAVQRDNL
FVDTTQEELNFTGNYTSLERQVGFGRVYIPLGPKSLDQAISSLRTYTLTAGDTKPLARGLLVVIQ
MVSEAARFRYIELRIRTSITDASEFTPDLLMLSMENNWSSMSSEIQQAQPGGIFAGVVQLRDERN
NSIEVTNFRRLFELTYIAVLLYGCAPVTSSSYSNNAIDAQIIKMPVFRGGEYEKVCSVVEVTRRI
SGWDGLCVDVRYGHYIDGNPVQLRPCGNECNQLWTFRTDGTIRWLGKCLTASSSVMIYDCNTVPP
EATKWVVSIDGTITNPHSGLVLTAPQAAEGTALSLENNIHAARQGWTVGDVEPLVTFIVGYKQMC
LRENGENNFVWLEDCVLNRVQQEWALYGDGTIRVNSNRSLCVTSEDHEPSDLIVILKCEGSGNQR
WVFNTNGTISNPNAKLLMDVAQRDVSLRKIILYRPTGNPNQQWITTTHPA (SEQ ID NO: 56)

[00216] ***Sambucus Nigra Lectin 1' (Genbank Accession No. U66191):***

MKVVATILYLVVLAICGLGIHGAHPTHSAPPTVYPSVSFNLTEANSNEYRHFLQELRGKVILGSH
RAFDLPVLNPESKVSDSDRFVLVRLTNPSRKKVTLAIDVVTFYVVAFAQNDRSYFFSGSSEVQRE
NLFVDTTQEDLNFKGDYTSLEHQVGFGRVYIPLGPKSLAQSISSLSTYKSSAGDNKRLARSLLVV
IQMVSEAARFRYIQLRIQASITDAKEFTPDLLMLSMENKWSSMSSEIQQAQPGGAFAQVVKLLDQ
RNHPIDVTNFRRLFQLTSVAVLLHGCPTVTKMPAYIIKMPVFNGGEDEERCSVVEEVTRRIGGRD
GFCAEVKNGDEKDGTPVQLSSCGEQSNQQWTFSTDGTIQSLGKCLTTSSSVMIYNCKVVPPESTK
WVVSIDGTITNPRSGLVLTAPKAAEGTLVSLEKNVHAARQGWIVGNVEPLVTFIVGYEQMCLETN
PGNNDVSLGDCSVKSASKVDQKWALYGDGTIRVNNDRSLCVTSEGKSSNEPIIILKCLGWANQRW
VFNTDGTISNPDSKLVMHVDQNDVPLRKIILSHPSGTSNQQWIASTHPA (SEQ ID NO: 57)

[00217] ***Polyporous squamosus lectin 1a (UniProt Q75WT9)***

MSFQGHGIYYIASAYVANTRLALSEDSSANKSPDVIISSDAVDPLNNLWLIEPVGEADTYTVRNA
FAGSYMDLAGHAATDGTAIIGYRPTGGDNQKWIISQINDVWKIKSKETGTFVTLLNGDGGGTGTV
VGWQNITNNTSQNWTFQKLSQTGANVHATLLACPALRQDFKSYLSDGLYLVLTRDQISSIWQASG
LGSTPWRSEIFDCDDFATVFKGAVAKWGNENFKANGFALLCGLMFGSKSSGAHAYNWFVERGNFS
TVTFFEPQNGTYSANAWDYKAYFGLF (SEQ ID NO: 58)

[00218] ***Polyporous squamosus lectin 1b (UniProt Q75WT8)***

MSFEGHGIYHIPHAHVANIRMALANRGSGQNGTPVIAWDSNNDAFDHMWLVEPTGEADTYTIHNV
STGTYMDVTASAVADNTPIIGYQRTGNDNQKWIIRQVQTDGGDRPWKIQCKATGTFATLYSGGGS
GTAIVGWRLVNSNGNQDWVFQKLSQTSVNVHATLLACGATVGQDFKNYLYDGLYLVLPRDRISAI
WKASGLGETARRDGIYDSDEFAMTFKSAAATWGKENFKADGFAILCGMMFGTKASTNRHAYNWVV
ERGSFSTVTFFEPQNGTYSDDAWGYKAYFGLF (SEQ ID NO: 59)

Aptamers

[00219] In some embodiments, binding agents provided in accordance with the present invention are aptamers. Aptamers are macromolecules composed of nucleic acid (*e.g.*, RNA, DNA) that bind tightly to a specific molecular target (*e.g.*, an umbrella topology glycan). A particular aptamer may be described by a linear nucleotide sequence and is typically about 15 to about 60 nucleotides in length. Without wishing to be bound by any theory, it is contemplated that the chain of nucleotides in an aptamer form intramolecular interactions that fold the molecule into a complex three-dimensional shape, and this three-dimensional shape allows the aptamer to bind tightly to the surface of its target molecule. Given the extraordinary diversity of molecular shapes that exist within the universe of all possible nucleotide sequences, aptamers

may be obtained for a wide array of molecular targets, including proteins and small molecules. In addition to high specificity, aptamers have very high affinities for their targets (*e.g.*, affinities in the picomolar to low nanomolar range for proteins). Aptamers are chemically stable and can be boiled or frozen without loss of activity. Because they are synthetic molecules, they are amenable to a variety of modifications, which can optimize their function for particular applications. For example, aptamers can be modified to dramatically reduce their sensitivity to degradation by enzymes in the blood for use in *in vivo* applications. In addition, aptamers can be modified to alter their biodistribution or plasma residence time.

[00220] Selection of aptamers that can bind umbrella topology glycans (and/or to umbrella topology glycan mimics) can be achieved through methods known in the art. For example, aptamers can be selected using the SELEX (Systematic Evolution of Ligands by Exponential Enrichment) method (Tuerk and Gold, 1990, *Science* 249:505-510; incorporated herein by reference). In the SELEX method, a large library of nucleic acid molecules (*e.g.*, 10^{15} different molecules) is produced and/or screened with the target molecule (*e.g.*, an umbrella topology glycan of umbrella topology glycan epitope). The target molecule is allowed to incubate with the library of nucleotide sequences for a period of time. Several methods, known in the art, can then be used to physically isolate the aptamer target molecules from the unbound molecules in the mixture, which can be discarded. The aptamers with the highest affinity for the target molecule can then be purified away from the target molecule and amplified enzymatically to produce a new library of molecules that is substantially enriched for aptamers that can bind the target molecule. The enriched library can then be used to initiate a new cycle of selection, partitioning, and amplification. After 5-15 cycles of this iterative selection, partitioning and amplification process, the library is reduced to a small number of aptamers that bind tightly to the target molecule. Individual molecules in the mixture can then be isolated, their nucleotide sequences determined, and their properties with respect to binding affinity and specificity measured and compared. Isolated aptamers can then be further refined to eliminate any nucleotides that do not contribute to target binding and/or aptamer structure, thereby producing aptamers truncated to their core binding domain. See Jayasena, 1999, *Clin. Chem.* 45:1628-50 for review of aptamer technology; the entire teachings of which are incorporated herein by reference).

Production of Polypeptides

[00221] Polypeptides in accordance with the invention (*e.g.*, HA polypeptides and/or Non-HA polypeptides), and/or characteristic portions thereof, or nucleic acids encoding them, may be produced by any available means.

[00222] Polypeptides in accordance with the invention (or characteristic portions thereof) may be produced, for example, by utilizing a host cell system engineered to express a polypeptide-encoding nucleic acid in accordance with the invention.

[00223] Any system can be used to produce polypeptides (or characteristic portions), such as egg, baculovirus, plant, yeast, Madin-Darby Canine Kidney cells (MDCK), or Vero (African green monkey kidney) cells. Alternatively or additionally, polypeptides (or characteristic portions) can be expressed in cells using recombinant techniques, such as through the use of an expression vector (Sambrook *et al., Molecular Cloning: A Laboratory Manual*, CSHL Press, 1989; incorporated herein by reference).

[00224] Alternatively or additionally, polypeptides in accordance with the invention (or characteristic portions thereof) can be produced by synthetic means.

[00225] Alternatively or additionally, polypeptides in accordance with the invention (or characteristic portions thereof), and particularly HA polypeptides, may be produced in the context of intact virus, whether otherwise wild type, attenuated, killed, *etc.* Polypeptides in accordance with the invention (*e.g.*, HA polypeptides), or characteristic portions thereof, may also be produced in the context of virus like particles.

[00226] In some embodiments, HA polypeptides (or characteristic portions thereof) can be isolated and/or purified from influenza virus. For example, virus may be grown in eggs, such as embryonated hen eggs, in which case the harvested material is typically allantoic fluid. Alternatively or additionally, influenza virus may be derived from any method using tissue culture to grow the virus. Suitable cell substrates for growing the virus include, for example, dog kidney cells such as MDCK or cells from a clone of MDCK, MDCK-like cells, monkey kidney cells such as AGMK cells including Vero cells, cultured epithelial cells as continuous cell lines, 293T cells, BK-21 cells, CV-1 cells, or any other mammalian cell type suitable for the production of influenza virus for vaccine purposes, readily available from commercial sources (*e.g.*, ATCC, Rockville, MD). Suitable cell substrates also include human cells such as MRC-5

cells. Suitable cell substrates are not limited to cell lines; for example primary cells such as chicken embryo fibroblasts are also included.

[00227] Also, it will be appreciated by those of ordinary skill in the art that polypeptides, and particularly variant HA polypeptides as described herein, may be generated, identified, isolated, and/or produced by culturing cells or organisms that produce the polypeptide (whether alone or as part of a complex, including as part of a virus particle or virus), under conditions that allow ready screening and/or selection of polypeptides capable of binding to umbrella-topology glycans. To give but one example, in some embodiments, it may be useful to produce and/or study a collection of polypeptides (*e.g.,* HA variant polypeptides) under conditions that reveal and/or favor those variants that bind to umbrella topology glycans (*e.g.*, with particular specificity and/or affinity). In some embodiments, such a collection of polypeptides (*e.g.,* HA variant polypeptides) results from evolution in nature. In some embodiments, such a collection of polypeptides (*e.g.,* HA variant polypeptides) results from engineering. In some embodiments, such a collection of polypeptides (*e.g.,* HA variant polypeptides) results from a combination of engineering and natural evolution.

HA Receptors

[00228] HA interacts with the surface of cells by binding to a glycoprotein receptor. Binding of HA to HA receptors is predominantly mediated by N-linked glycans on the HA receptors. Specifically, HA on the surface of flu virus particles recognizes sialylated glycans that are associated with HA receptors on the surface of the cellular host. After recognition and binding, the host cell engulfs the viral cell and the virus is able to replicate and produce many more virus particles to be distributed to neighboring cells. Some crystal structures of exemplary HA-glycan interactions have been identified and are presented in Table 1:

Table 1. Crystal Structures of HA-Glycan Complexes

Abbreviation (PDB ID)	Virus Strain	Glycan (with assigned coordinates)
ADkALB76_H1_26 (2WRH)	A/duck/Alberta/76 (H1N1)	Neu5Ac
ASI30_H1_23 (1RV0)	A/Swine/Iowa/30 (H1N1)	Neu5Ac
ASI30_H1_26 (1RVT)	A/Swine/Iowa/30 (H1N1)	Neu5Acα6Galβ4GlcNAcβ3Galβ4Glc
ASC18_H1_26 (2WRG)	A/South Carolina/1/18 (H1N1)	Neu5Acα6Galβ4GlcNAcβ3Gal

Abbreviation (PDB ID)	Virus Strain	Glycan (with assigned coordinates)
APR34_H1_23 (1RVX)	A/Puerto Rico/8/34 (H1N1)	Neu5Acα3Galβ4GlcNAc
APR34_H1_26 (1RVZ)	A/Puerto Rico/8/34 (H1N1)	Neu5Acα6Galβ4GlcNAc
ACkNY91_H2_23 (2WR2)	A/chicken/NY/29878/91 (H2N2)	Neu5Acα3Galβ3GlcNAc
AckNY91_H2_26 (2WR1)	A/chicken/NY/29878/91 (H2N2)	Neu5Acα6Galβ4GlcNAc
AdkON77_H2_23 (2WR3)	A/duck/Ontario/77 (H2N2)	Neu5Acα3Galβ4GlcNAc
AdkON77_H2_26 (2WR4)	A/duck/Ontario/77 (H2N2)	Neu5Acα6Galβ4GlcNAc
AckPD84_H2_26 (2WRF)	A/chicken/Potsdam/475/84 (H2N2)	Neu5Acα6Gal
ASING57_H2_23 (2WRB)	A/Singapore/1/57 (H2N2)	Neu5Ac
ASING57_H2_26 (2WR7)	A/Singapore/1/57 (H2N2)	Neu5Acα6Galβ4GlcNAcβ3Gal
AJAP57_H2_26(2WRE)	A/Japan/305/57 (H2N2)	Neu5Acα6Gal
ADU63_H3_23 (1MQM)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Acα3Gal
ADU63_H3_26 (1MQN)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Acα6Gal
AAI68_H3_23 (1HGG)	A/Aichi/2/68 (H3N2)	Neu5Acα3Galβ4Glc
ADS97_H5_23 (1JSN)	A/Duck/Singapore/3/97 (H5N3)	Neu5Acα3Galβ3GlcNAc
ADS97_H5_26(1JSO)	A/Duck/Singapore/3/97 (H5N3)	Neu5Ac
Viet1203_04_H5 (2FK0)	A/Vietnam/1203/2004 (H5N1)	
Viet1194_04_H5 (2IBX)	A/Vietnam/1194/2004 (H5N1)	
ASI30_H1_23 (1RV0)	A/Swine/Iowa/30 (H1N1)	Neu5Ac
ASI30_H1_26 (1RVT)	A/Swine/Iowa/30 (H1N1)	Neu5Acα6Galβ4GlcNAcβ3Galβ4Glc
APR34_H1_23 (1RVX)	A/Puerto Rico/8/34 (H1N1)	Neu5Acα3Galβ4GlcNAc
APR34_H1_26 (1RVZ)	A/Puerto Rico/8/34 (H1N1)	Neu5Acα6Galβ4GlcNAc
ADU63_H3_23 (1MQM)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Acα3Gal
ADU63_H3_26 (1MQN)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Acα6Gal
AAI68_H3_23 (1HGG)	A/Aichi/2/68 (H3N2)	Neu5Acα3Galβ4Glc
ADS97_H5_23 (1JSN)	A/Duck/Singapore/3/97 (H5N3)	Neu5Acα3Galβ3GlcNAc
ADS97_H5_26(1JSO)	A/Duck/Singapore/3/97 (H5N3)	Neu5Ac
Viet04_H5 (2FK0)	A/Vietnam/1203/2004 (H5N1)	

HA - α2-6 sialylated glycan complexes were generated by superimposition of the CA trace of the HA1 subunit of ADU63_H3 and ADS97_H5 and Viet04_H5 on ASI30_H1_26 and APR34_H1_26 (H1). Although the structural complexes of the human A/Aichi/2/68 (H3N2) with α2-6 sialylated glycans are published (Eisen *et al.*, 1997, *Virology*, 232:19; incorporated herein by reference), their coordinates were not available in the Protein Data Bank. The SARF2