

1/44

Table 1. Features extracted from the glycans on the glycan array.

The features described in this table were used by the rule based classification algorithm to identify patterns that characterized binding to specific GBP.

Features extracted	Feature Description
<i>Monosaccharide level</i>	
Composition	Number of hex, hexNAcs, dHex, sialic acids, etc [In figure 1, the composition is Hex=5;HexNAc=4]. Terminal composition is distinctly recorded [In figure 1, the terminal composition is Hex=2;HexNAc=2].
Explicit Composition	Number of Glc, Gal, GlcNAc, Fuc, GalNAc, Neu5Ac, Neu5Gc, etc [In figure 1, the explicit composition is Man=5;GlcNAc=4]. Terminal explicit composition is explicitly recorded [In figure 1, the terminal explicit composition is Man=2;GlcNAc=2].
<i>Higher order features</i>	
Pairs	Pair refers to a pair of monosaccharide, connected covalently by a linkage. The pairs are classified into two categories, regular [B] and terminal [T] to distinguish between the pair with one monosaccharide that terminates in the non reducing end [Figure 2]. The frequency of the pairs were extracted as features
Triplets	Triplet refers to a set of three monosaccharides connected covalently by two linkages. We classify them into three categories namely regular [B], terminal [T] and surface [S] [Figure 2]. The compositions of each category of triplets were extracted as features
Quadruplets	Similar to the triplet features, quadruplets features are also extracted, with four monosaccharides and their linkages [Figure 2]. Quadruplets are classified into two varieties regular [B] and surface [S]. The frequencies of the different quadruplets were extracted as features

2/44

Table 1 (continued).

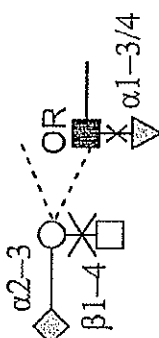
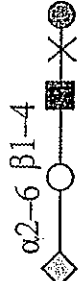
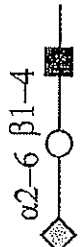
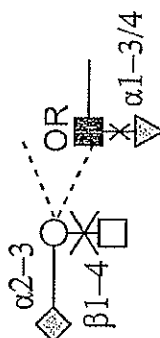
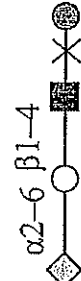
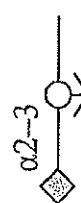
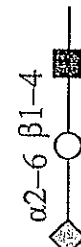
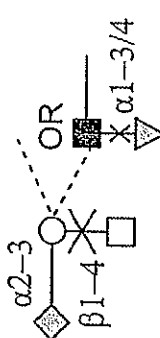
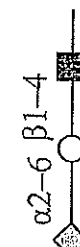
Clusters	In the case of surface triplets and quadruplets above, if the linkage information is ignored, we get a set of monosaccharide clusters, and their frequency of occurrence (composition) is tabulated. These features were chosen to analyze the importance of types of linkages between the monosaccharides.
Average Leaf Depth	As an indicator of the effective length of the probes, average depth of the reducing end of the tree is extracted as a glycan feature. In Figure 2B , the leaf depths are 3, 4 and 3, and the average is 3.34
Number of Leaves	As a measure of spread of the glycan tree, the number of non reducing monosaccharides is extracted as a feature. For Figure 2B , the number of leaves is 3. For figure 1 it is 4.
GBP binding features	<i>These features are obtained for all GBPs screened using the array</i>
Mean signal per glycan	Raw signal value averaged over triplicate or quadruplicate [depending on array version] representation of the same glycan
Signal to Noise Ratio	Mean noise computed based on negative control [standardized method developed by CFG] to calculate signal to noise ratio [S/N]

Table 2. Crystal structures of HA-glycan complexes

Abbreviation (PDB ID)	Virus strain	Glycan (with assigned coordinates)
ASI30_H1_23 (1RV0)	A/Swine/Iowa/30 (H1N1)	Neu5Ac
ASI30_H1_26 (1RVT)	A/Swine/Iowa/30 (H1N1)	Neu5Ac α 6Gal β 4GlcNAc β 3Gal β 4Glc
APR34_H1_23 (1RVX)	A/Puerto Rico/8/34 (H1N1)	Neu5Ac α 3Gal β 4GlcNAc
APR34_H1_26 (1RVZ)	A/Puerto Rico/8/34 (H1N1)	Neu5Ac α 6Gal β 4GlcNAc
ADU63_H3_23 (1MQM)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Ac α 3Gal
ADU63_H3_26 (1MQN)	A/Duck/Ukraine/1/63 (H3N8)	Neu5Ac α 6Gal
AAI68_H3_23 (1HGG)	A/Aichi/2/68 (H3N2)	Neu5Ac α 3Gal β 4Glc
ADS97_H5_23 (1JSN)	A/Duck/Singapore/3/97 (H5N3)	Neu5Ac α 3Gal β 3GlcNAc
ADS97_H5_26(1JSO)	A/Duck/Singapore/3/97 (H5N3)	Neu5Ac
Viet04_H5 (2FK0)	A/Vietnam/1203/2004 (H5N1)	

The HA- α 2-6 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 α 2-6 sialylated glycans are published¹⁷, their coordinates were not available in the Protein Data Bank. The SARF2 (<http://123d.ncifcrf.gov/sarf2.html>) program was used to obtain the structural alignment of the different HA1 subunits for superimposition.

Table 3. Glycan receptor specificity of HAs based on classifier rules

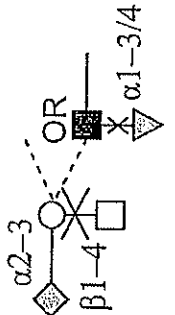
Influenza Strain	$\alpha 2-3$ Type ^a	$\alpha 2-6$ Type ^b	
A/Duck/Alberta/35/76 (<i>Avian H1N1</i>)	 (Type C)	 (Type A ¹)	
A/Duck/Alberta/35/76 (<i>Avian H1N1</i>) Glu190Asp/Gly225 Asp double mutant	No	 (Type B)	
A/South Carolina/1/18 (<i>Human H1N1</i>)	No	 (A or B)	4/44
A/New York/1/18 (<i>Human H1N1</i>)	 (Type C ²)	 (Type B ³)	
A/Texas/36/91 (<i>Human H1N1</i>)	 (Type A ⁵)	 (A or B)	
A/New York/1/18 (<i>Human H1N1</i>) Asp 190Glu mutant ⁴	 (Type C ⁵)	 (A or B)	

5/44

Table 3 (continued)

A/New York/1/18 (<i>Human H1N1</i>) Lys222 Leu mutant	No	No	No
A/Duck/Ukraine/1/63 (<i>Avian H3N8</i>)	No	No	No
A/Moscow/10/99 (<i>Human H3N2</i>)	No ⁶	No	(Type B ⁷)
A/Duck/Singapore/3/97 (<i>Avian H5N3</i>)	(Type C ²)	No	(Type A)
A/Vietnam/1203/04 (<i>Avian H5N1</i>)	(Type A)	No	(Type B)
A/Vietnam/1203/04 (<i>Avian H5N1</i>) Glu 190Asp/ Gly225 Asp double mutant	No	No	(Type A)
A/Vietnam/1203/04 (<i>Avian H5N1</i>) Gln226Leu/ Gly228Ser double mutant	No	No	(Type A)

Table 3 (continued)

A/Vietnam/1203/04 (<i>Avian H5N1</i>) Arg216Glu, Ser221 Pro double mutant		No (Type C)
--	---	----------------

¹ Border line high binder; ² Sulfated GlcNAc[6S]/Gal[6S] high binders ³; Border line high binders to α2-6 Type B. Only sulfated GlcNAc[6S]/Gal[6S] are high binders; ⁴ Binds to several non-sialylated glycans; ⁵ Border line high to α2-3 sialylated glycans; ⁶ Few border line high binders to sulfated GlcNAc on Neu5Acα3Galβ3/4GlcNAc; ⁷ High binders are Neu5Acα6Galβ4GlcNAcβ3Gal & !GlcNAcα6Man; Others are borderline high.

Keys: ■ GlcNAc; □ Gal; ● Man; ▼ Fuc; ◆ Neu5Ac;

The data from glycan microarray screening of H1, H3 and H5 subtypes were obtained from the Consortium for Functional Glycomics (CFG) web site – <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp>. The details of the data mining analysis including the description of features and classifiers are provided in **Suppl Figure 5**. The rule induction classification method was used to generate the following classifiers (or rules) that govern the binding of HA to α2-3/6 sialylated glycans. Classifiers for α2-3 sialylated glycan binding – Type A: Neu5Acα3Gal & !GalNAcβ4Gal, Type B: Neu5Acα3Galβ4GlcNAc & !GalNAcβ4Gal & {GlcNAcβ3Gal or GlcNAc[6S]}, Type C: Neu5Acα3Galβ & !GalNAcβ4Gal & !Fucα3/4GlcNAc. Classifiers for α2-6 sialylated glycan binding – Type A: Neu5Acα6Galβ4GlcNAcβ?Man, Type B: Neu5Acα6Galβ4GlcNAc & !GlcNAcβ?Man. These complex rules are graphically represented in the table for clarity. The rules are provided as a logical combination of features among high affinity binders that enhance binding and features among weak and non-binders that are detrimental to binding (shown after the '!' symbol in the text description and as a red linkage with a 'x' sign in the graphical representation). The presence of mannose in the α2-6 classifiers arises from the single 6'-sialyl lactosamine containing biantennary N-linked glycan on the glycan array.

7/44

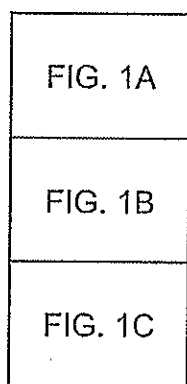


FIG. 1

8/44

97	139	153	183
H1_Av	ENGTCYPGEFIDYEELREQLSSISSEKFEIIPKASSWPNEHTTKGVTAA	CSYS-GASSFYRNLLMLTKKG-TS-YPKLSKSYTNKNGKEVLVIMGV	HHPPSVSEQOQSLYQVAD
H1_Hu1	ENGTCYPGDFIDYEELREQLSSVSSFEKFEIIPKTSWPNEHTTKGVTAA	CSYA-GASSFYRNLLMLTKKG-SS-YPKLSKSYTNKNGKEVLVIMGV	HHPPPTGTDQOQSLYQVAD
H1_Hu2	ENGTCYPGYADYEELREQLSSVSSFEKFEIIPKSSWPNEHTVT-GVSA	CSHN-GASSFYRNLLMLTKGN-CL-YPNLSKSYTNKNGKEVLVIMGV	HHPPNIGDQORALYHTEN
H2_Av	ANGLCYPSFNDYEELKHLTSVTHFEKVKILPR-DQWTOHTTGG-SRACA	VS-GNPSFFRNMMVWLTEKG-SN-YPIAKRSYNNNTSGQMLVIMGV	HHHENDDTEQRTLYQNVG
H2_Hu	RDGLCYPSFNDYEELKHLSSVVKHFEKVKILPK-DRWTOHTTGG-SRACA	VS-GNPSFFRNMMVWLTEKG-SN-YPVAKGSYNNNTSGEQLIIMGV	HHHENDDEKEQRTLYQNVG
H3_Av	FS-NCYPDIPDYASLSLVASSGTFEITEG---FTW-TGVTONGGSA	CKRG-PANGFFSRLLMLTKS--ESAYPVLNVTPMNNDFKLYIMGV	HHHESINQEQTDLIVQAS
H3_Hu1	FS-NCYPDVPDYASLSLVASSGTFEITEG---FTW-TGVTONGGSA	CKRG-PGSGFFSRLLMLTKS--GSTYPVLNVTPMNNDFKLYIMGV	HHHESINQEQTSLIVQAS
H3_Hu2	YS-NCYPDVPDYASLSLVASSGTFEFNNE---FNW-AGVTQNGTS	ACKRR-SNKSFFSRLLMLTHL--KYKYPALNVTPMNNKFDKLYIMGV	HHHVPVDSDDQISLYAQAS
H4_Av	VD-TCYPEDVPDYQSLRSILANNGKFEFIAEE---FQW-NTVKQNG	SGACKRA-NVNDFFNRLNMLTKSN-GDAYPLQNLTKVNNGDYARLYIMGV	HHHHPSTDTEQTDLYKNNP
H5_Av1	VNDLCYPGFNDYEELKHLLSRINHFEKIQIIPK-SSWSHEASLGV	SACPQY-GKSSFFRNVMILIKKN-ST-YPTIKRSYNNNTNQEDLLVIMGV	HHHENDAAEQTKLYQNPT
H5_Av2	ANDLCYPGFNDYEELKHLLSRINHFEKIQIIPK-NSWSHEASLGV	SACPQY-GKSSFFRNVMILIKKN-NA-YPTIKRSYNNNTNQEDLLVIMGV	HHHENDAAEQTKLYQNPT
H6_Av	QNGLCYPTLINEIEELKALIGSGRIERFEMFPK-STWSGVNTNNG	VTRACPDN-SGSSFYRNLLMLTKNSAA-YPVIKGTYNNTGNQPI	LYFIMGVHHHPPPTNAQNNLYGSGD
H7_Av	SD-VCYPGKFVNEELRQILRESGINKETMG---FTY-SGIRTING	ATSTCRRS-G-SSFYAEMKMLLSNTDNAAFPQWTKSYKNTRKDP	ALLIMGVHHHSGSTTEQTKLYGSGN
H8_Av	PEGMCYPSVENLEELRFVFSASAKYKRIRLFDY-SRWN--VTRSG	TSKACNASIGGQSFYRSINMLTKKAPDT-YDFNEGAYVNNEDG	DIIFIMGVHHHPPDTKEQTTLYKNAN
H9_Av	VNGTCYPGNVENLEELRTLFSASASYQRIQIFPD-TIWN--VTY	TGTSKACS-----GSFYRSMRMLTQK-SGS-YPVQDAQYTN	NREKSILFIMGVHHHPPPTDQAQTNLYINTD
H10_Av	IA-YCYPGATVNEELRQKIMESGGIDKISTG---FTYGSINS	AGTTRSCMRSGGNSFYAELKMLVSKNKGQNFQNTANTYRNTD	SAEHLIMGVHHHSSSTQEKNDLYGTQS
H11_Av	TNGTCYPTLENEELRLKFSGVLEFSKFPAFTS-NGWGA	VNSGAGVTAACKFG-SSNSFFRNMLLIH-QSGT-YPVIRRT	FNNTKGRDVLVIMGVHHHATLKEHQDLYKKDS
H12_Av	MEGVCPGSIENQELRSLFSSIKKYERVNMFDF-TKWN--VTY	TGTSRACNNTSNRSGSFYRSMRMLTIK-SGQ-FPVQTD	EYKNTFRSDILFTWALHHHPTSAEQVQLYKNPD
H13_Av	PHGLCYPELNNGELRLHLFSGIRSFSTRTELI PP-TSWGEVLD--	CATSACRDDKGTNSFYRNLMVFEVK-KNNR-YPVISKTYNN	TGTRDVLVIMGVHHHVSVEETKLYVNSD
H14_Av	VD-TCYPFDVPDYQSLRSILASSGLEFIAEQ---FTW-NGKV	DGSSSACLRG-GRNSFFSRLLMLTKET-NGNYGPINVT	KENTGSYRVLVIMGVHHHPSSDNEQTDLYKVAT
H15_Av	SD-ICYPGKFTNEELRQIIRRESGGIDKBPNG---FRY-SG	IKTDGATSACKPT-V-SSFYSEMMLLSKANQVFQNLQTYRNN	RKEPALIVIMGVHHHSSSLDEQNKLYGAGN
H16_Av	PNKLCTFGELDNNGELRLHLFSGVNSFSRTELI SP-NKWD	ILD--GVTASACRDN-GASSFYRNLMVIVKNKNGK-IPV	IKGDYNNNTGTRDVLVIMGVHHHPPDTETTAINLYASKN

FIG. 1A

9/44

H1_Av AYVSGSSKYNRRFAPEIARPEVRGOAGRMVYWTLLDQGDITTEATGNLIAPWYAFALNKGSD-----SGIITS-DAPVH-NCOTRCQTPHCAINSSLPFQNVHPIT
 H1_Hu1 AYVSGSSKYNRRFTEIARPKVRDQAGRMVYWTLLLEPGDITTEATGNLIAPWYAFALNRSG-----SGIITS-DAPVH-DCNTKCQTPHCAINSSLPFQNIHPVT
 H1_Hu2 AYVSVSSHYSRRFTEIARPKVRDQGRINYYWTLLLEPGDITTEANGNLIAPWYAFALSRGFG-----SGIITS-NAPMD-ECDAKCQTPQCAINSSLPFQNVHPVT
 H2_Av TYVSVGTSTLNKRISIPEIATRPKVNGQGRMEFSTLLETVDVINFEESTGNLIAPYGFKISKRG-----SGIMKT-EKTL-NCETKCQTPGLGAINITLTFHNHPLT
 H2_Hu TYVSVGTSTLNKRSTPDIAATRPKVNGLSRMEFSTLLDMWDINFEESTGNLIAPYGFKISKRG-----SGIMKT-EGTLE-NCETKCQTPGLGAINITLTFHNHPLT
 H3_Av GRVTVSTRRSQOTIIPNIGSRPWVRGQGRISIVYWTIVKPGDVIVNSNGNLIAPRGYFKMRTG-----KSSIMRS-DAPID-TCISECITPNGSIPNDKPFQNVNKIT
 H3_Hu1 GRVTVSTRRSQOTIIPNIGSRPWVRGLSSRISIVYWTIVKPGDVIVNSNGNLIAPRGYFKMRTG-----KSSIMRS-DAPID-TCISECITPNGSIPNDKPFQNVNKIT
 H3_Hu2 GRITVSTRRSQOTVIPNIGYRPRVRDISSRISIVYWTIVKPGDILLINSTGNLIAPRGYFKIRSG-----KSSIMRS-DAPIG-KCNSECITPNGSIPNDKPFQNVNRIT
 H4_Av GRVTVSTKTSQTSWPNIGSRPWVRGQGRISIVYWTIVDPGDILIVFNTIGNLIAPRGHYKLNQK-----KSTILNT-AVPIG-SCVSKCHTDGSIITTTKPFQNISRIS
 H5_Av1 TYISVGTSTLNQRLVPRIATRSKVNGQGRMEFFWTILKPDAINFESNGNLIAPYAYKIVKKGD-----STIMKS-ELEYG-NCWTKCQTPMGAINSSMPFHNHPLT
 H5_Av2 TYISVGTSTLNQRLVPKIAATRSKVNGQGRMEFFWTILKPDAINFESNGNLIAPYAYKIVKKGD-----SAIMKS-ELEYG-NCWTKCQTPMGAINSSMPFHNHPLT
 H6_Av RYVRMGTESMFAKGEISARPVVNGQGRIDYIWSVLKPGETLNVESNGNLIAPWYAYKFVSTNSK-----GAVFKS-NLPIE-NCDATCQTLAGVLRNKTQFQNVSPILW
 H7_Av KLITVSSSNYQOSFVPSGARPKVDGQGRIDFHWMLNPNNDITTESFNCAFIAPDRASFRL-----GKSMGIQS-GVQVDDNCEGDCYHSGGTIIISNLPFQNIINSRA
 H8_Av TLSSVTNTINRSFQPNIGRPLVRGQGRMDYYWGLKRGETILKIRTNGLIAPYFGLKGESY-----GRILIONEDIPIG-NCWTKCQTYAGAINSSLPFQNASRHY
 H9_Av TTTSVTTEDLNRIKPVIGRPLVNLQGRINYYWSVLKPGQTLRVRNGLIAPWYGHVLSGGSH-----GRILKT-DLNSG-NCVVQCQTEKGLNSTLTFHNISKYA
 H10_Av LSISVGSSTYQNNFVVGARPVNGQGRIDFHWMTWVQPDNITFESHNGLIAPSRVSKLK-----GRGLGIQS-GASVDNDCSKCFWKGGSIINTKLFPQNLSPRT
 H11_Av SYVAVDSESYNRRFTEISTRPKVNGQAGRMFTFWTIVKPGDAITFESNGAFIAPRYAFELVSLGN-----GKLFRS-DLNIE-SCSTKCQSEIGGINTRNRSFHNVRHT
 H12_Av TLSSVTTDEINRSFKPNIGRPLVRGQGRMDYYWAVLKPGQIVKIQTNGNLIAPYGHILITGKSH-----GRILKN-NLPVG-QCVTECOLNEGVMNTSKPFQNTSKHY
 H13_Av PYTLVSTKSWSEKYKLETGVVRPGYNGQSRWMMKIWSLLHPGEMITFESNGGLIAPRYGVIIIEYK-----GRIFQS-RIRMS-KCNTKCQTSVGGINTNRTFQNDKNA
 H14_Av GRVTVSTRSDQISIVPNIGSRPRVRNOSGRISIVYWTIVNPGDSIIFNSIGNLIAPRGHYKISKST-----KSTVLKS-DKRIG-SCITSPCLTDKGSIQSDKPFQNVSRIA
 H15_Av KLITVGSKYQOSFSPGDRPKVNGQAGRIDFHWMLLDPGDVTFTFNAGFIAPDRATFLRNPSPGVEYNGKSLGIQS-DAQIDESGECECFYSGGTINSPLPFQNDISWA
 H16_Av PYTLVSTKWSKRYELEIGTRIG-DGORSWMLVHLMRPGERIMFESNGGLIAPRYGVIIIEKYGT-----GRIFQS-GVRMA-KCNTKCQTSVGGINTNRTFQNIERNA

FIG. 1B

H1_	Av	IG	CP	YV	K	T	K	L	R	M	A	T	G	L	R	N	V	E	S	I	Q	----	S	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	Q	N	E	Q	G	S	Y	A	A	D	K	S	T	O	N	A	I	D	G	I	T	S	K	V	N	S	I	E	K	M	N	T	O	T	A	V	G	K	E	F	N	N	L	R	R	I	E	N	L	N											
H1_	Hu1	IG	CP	YV	R	T	K	L	R	M	A	T	G	L	R	N	I	P	S	I	Q	----	S	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	Q	N	E	Q	G	S	Y	A	A	D	K	S	T	O	N	A	I	D	G	I	T	N	K	V	N	S	I	E	K	M	N	T	O	T	A	V	G	K	E	F	N	N	L	R	R	I	E	N	L	N											
H1_	Hu2	IG	CP	YV	R	S	A	K	L	R	M	V	T	G	L	R	N	I	P	S	I	Q	----	S	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	Q	N	E	Q	G	S	Y	A	A	D	K	S	T	O	N	A	I	D	G	I	T	N	K	V	N	S	I	E	K	M	N	T	O	T	A	V	G	K	E	F	N	K	L	R	R	I	E	N	L	N										
H2_	Av	IG	CP	YV	K	S	D	R	L	V	L	A	T	G	L	R	N	V	E	Q	I	E	----	S	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	S	N	D	Q	G	S	Y	A	A	D	K	E	S	T	O	K	A	I	D	G	I	T	N	K	V	N	S	I	E	K	M	N	T	O	F	E	A	V	G	K	E	F	N	L	R	R	I	E	N	L	N									
H2_	Hu	IG	CP	YV	K	S	E	K	L	V	L	A	T	G	L	R	N	V	E	Q	I	E	----	S	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	S	N	D	Q	G	S	Y	A	A	D	K	E	S	T	O	K	A	I	D	G	I	T	N	K	V	N	S	I	E	K	M	N	T	O	F	E	A	V	G	K	E	F	N	L	R	R	I	E	N	L	N									
H3_	Av	Y	G	A	C	P	Y	V	K	O	N	T	L	K	A	T	G	M	R	N	V	E	K	----	Q	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	E	Q	G	T	O	A	A	I	D	Q	I	N	R	K	L	N	R	V	I	E	K	T	N	E	K	H	Q	I	E	K	E	F	S	E	V	E	G	R	I	O	D	L	E												
H3_	Hu1	Y	G	A	C	P	Y	V	K	O	N	T	L	K	A	T	G	M	R	N	V	E	K	----	Q	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	E	Q	G	T	O	A	A	I	D	Q	I	N	G	K	L	N	R	V	I	E	K	T	N	E	K	H	Q	I	E	K	E	F	S	E	V	E	G	R	I	O	D	L	E												
H3_	Hu2	Y	G	A	C	P	Y	V	K	O	N	T	L	K	A	T	G	M	R	N	V	E	K	----	Q	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	E	Q	G	T	O	A	A	I	D	Q	I	N	G	K	L	N	R	V	I	E	K	T	N	E	K	H	Q	I	E	K	E	F	S	E	V	E	G	R	I	O	D	L	E												
H4_	Av	I	G	C	P	Y	V	K	O	G	S	L	K	L	A	T	G	M	R	N	I	P	E	K	----	A	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	E	Q	G	T	A	A	I	D	Q	I	N	G	K	L	N	R	V	I	E	K	T	N	E	K	H	Q	I	E	K	E	F	S	E	V	E	G	R	I	O	D	L	E												
H5_	Av1	I	G	C	P	Y	V	K	S	N	R	L	V	L	A	T	G	L	R	N	S	P	O	R	R	R	K	K	R	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	S	N	E	Q	G	S	Y	A	A	D	K	E	S	T	O	K	A	I	D	G	V	T	N	K	V	N	S	I	D	K	M	N	T	O	F	E	A	V	G	R	E	F	N	N	L	R	R	I	E	N	L	N					
H5_	Av2	I	G	C	P	Y	V	K	S	N	R	L	V	L	A	T	G	L	R	N	S	P	O	R	R	R	K	K	R	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	S	N	E	Q	G	S	Y	A	A	D	K	E	S	T	O	K	A	I	D	G	V	T	N	K	V	N	S	I	D	K	M	N	T	O	F	E	A	V	G	R	E	F	N	N	L	R	R	I	E	N	L	N					
H6_	Av	I	G	C	P	Y	V	K	S	E	S	L	R	L	A	T	G	L	R	N	V	E	Q	I	A	----	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	H	H	S	N	E	Q	G	S	Y	A	A	D	R	E	S	T	O	K	A	I	D	G	I	T	N	K	V	N	S	I	D	K	M	N	T	O	F	E	A	V	D	H	E	F	S	N	L	R	R	I	D	N	N						
H7_	Av	V	G	C	P	Y	V	K	O	E	S	L	M	L	A	T	G	M	N	V	E	I	P	----	K	G	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	E	Q	G	T	A	A	I	D	Y	K	S	T	O	E	A	I	D	Q	I	T	G	K	L	N	R	V	I	E	K	T	N	O	Q	F	E	L	I	D	N	E	F	T	E	V	E	K	O	I	G	N	V	I			
H8_	Av	M	G	E	C	P	Y	V	K	A	S	L	R	L	A	V	G	L	R	N	T	P	S	V	E	----	P	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	H	H	S	N	E	Q	G	T	C	M	A	A	D	Q	K	S	T	O	E	A	I	D	K	I	T	N	K	V	N	I	V	D	K	M	N	R	E	F	V	N	H	E	F	S	E	V	E	K	R	I	N	N						
H9_	Av	F	G	I	C	P	Y	V	R	V	K	S	L	K	L	A	V	G	L	R	N	V	E	P	A	R	S	----	N	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	H	Q	N	D	Q	G	V	G	M	A	A	D	R	D	S	T	O	R	A	I	D	K	I	T	S	K	V	N	I	V	D	K	M	N	K	O	Y	E	L	I	D	H	E	F	S	E	V	E	T	R	L	N	N			
H10_	Av	V	G	Q	C	P	Y	V	N	K	S	L	L	L	A	T	G	M	R	N	V	E	V	----	Q	G	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	A	Q	G	T	O	A	A	I	D	Q	I	T	G	K	L	N	R	V	I	E	K	T	N	T	E	F	S	E	S	E	I	E	H	Q	I	G	N	V	I																
H11_	Av	I	G	D	C	P	Y	V	N	V	K	S	L	K	L	A	T	G	L	R	N	V	E	P	A	T	----	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	N	E	E	G	T	I	A	A	D	K	E	S	T	O	K	A	I	D	Q	I	T	S	K	V	N	I	V	D	R	M	T	N	F	S	V	Q	H	E	F	S	E	I	E	R	I	N	Q	I	S							
H12_	Av	I	G	C	P	Y	I	P	S	G	S	L	K	L	A	T	G	L	R	N	V	E	Q	I	----	N	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	H	Q	N	A	E	G	T	C	M	A	A	D	R	D	S	T	O	K	A	I	D	N	N	Q	N	K	L	N	V	I	D	K	M	N	K	O	F	E	V	N	H	E	F	S	E	S	R	I	N	N									
H13_	Av	I	G	D	C	P	Y	I	K	S	G	L	K	L	A	T	G	L	R	N	V	E	P	A	S	----	N	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	H	Q	N	E	Q	G	T	I	A	A	D	K	E	S	T	O	K	A	I	D	Q	I	T	K	I	N	I	I	D	K	M	G	N	V	D	S	I	R	G	E	F	N	O	V	E	K	R	I	N	M	L	A						
H14_	Av	I	G	N	C	P	Y	V	K	O	G	S	L	M	L	A	T	G	M	R	N	I	P	E	K	----	Q	A	K	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	A	E	G	T	A	A	I	D	K	S	T	O	A	A	I	D	Q	I	N	G	K	L	N	R	V	I	E	K	T	N	E	K	H	Q	I	E	K	E	F	S	E	V	E	G	R	I	O	D	L	E				
H15_	Av	V	G	R	C	P	Y	V	K	O	S	S	L	P	L	A	T	G	H	K	N	V	E	P	E	K	I	----	H	T	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	R	H	Q	N	A	Q	G	T	A	A	I	D	Y	K	S	T	O	A	A	I	D	Q	I	T	G	K	L	N	R	V	I	E	K	T	N	O	Q	F	E	L	I	D	N	E	F	T	E	V	E	Q	I	G	N	V	I
H16_	Av	I	G	D	C	P	Y	I	K	S	G	L	K	L	A	T	G	L	R	N	V	E	P	S	V	G	----	E	R	G	L	F	G	A	I	A	G	I	E	G	G	T	M	I	D	G	W	Y	G	F	H	Q	N	E	Q	G	T	I	A	A	D	K	E	S	T	O	K	A	I	D	E	I	T	K	I	N	I	I	E	K	M	G	N	V	D	S	I	R	G	E	F	N	O	V	E	K	R	I	N	M	L	A					

FIG. 1C

FIG. 2A
FIG. 2B

FIG. 2A

FIG. 2

98 136 153

H1 Subtype	:	:
ADA76	SYIETNSENGTCY	PGGEFIDYEEELREQLSSISSEKFEIFPKASSWPNHETTKGVTAACSYSGASSFYRNLLMLTKKGTSY
ASI30	SYIVETNSDNGTCY	PGDFIDYEEELREQLSSVSSFEKFEIFPKTSSWPNHETTRGVTAACPYAGASSFYRNLLMLVKKGNSY
APR34	SYIVETPNSENGICYP	PGDFIDYEEELREQLSSVSSFEKFEIFPKESSWPNHNTNG-VTAACSHGKSSFYRNLLMLTEKEGSY
ASC18	SYIVETNSENGTCY	PGDFIDYEEELREQLSSVSSFEKFEIFPKTSSWPNHETTKGVTAACSYAGASSFYRNLLMLTKKGSSY
AT91	SYIAETPNPENGTCY	PGCFADYEEELREQLSSVSSFEKFEIFPKESSWPNHTVTKGVTTSCSHNGKSSFYRNLLMLTKKNGLY
ANY18	SYIVETNSENGTCY	PGDFIDYEEELREQLSSVSSFEKFEIFPKTSSWPNHETTKGVTAACSYAGASSFYRNLLMLTKKGSSY
H3 Subtype		
ADU63	DLFVERNAES-NCY	PYDIPDYASRLSVASSGTLEFITEG----FTWTGVTQNGGSACKRGPANGFFSRLNMLTKSESAY
AAI68	DLFVERSKAFS-NCY	PYDVPDYASRLSVASSG---TLEFITEGFTWTG-VTQNGGSNACKRGPGSGFFSRLNMLTKSGSTY
AM99	DLFVERSKAYS-NCY	PYDVPDYASRLSVASSGTLEFNES----FNWTGVAONGTSSCKRRRSIKSFFSRLNMLHOLKYRY
H5 Subtype		
ADS97	SYIVEKDNFVNGLCY	PENFNNDYEEKHLLSSTNHFEKIRIIPR-SSWSNHDASSGVSSACPYNGRSSFFRNVMMLIKKNAY
Viet04	SYIVEKANPVNDLCY	PGDFNDYEEKHLLSRINHFEKIQIIPK-SSWSSHEASLGVSSACPYQKGSSFFRNVMMLIKKNSTY

	183	190	222	226
H1 Subtype	:	:	:	:
ADA76	PKLSKSYTNKKGKVLVWGVHHPSPVSE	QQSLYQNADAYVSVGSSKYNRRFAPEIAARPE	VRGQ	AGRMNYYWTLLDQGD
ASI30	PKLSKSYVNNKGKVLVWGVHHPPTSTD	QQSLYQNADAYVSVGSSKYDRRFTPEIAARPKVRSQ	AGRMNYYWTLLLEPGD	TI
APR34	PKLKSIVNNKGKVLVWGVHHPNSKE	QQNLVQENAYVSVTSNYYNRRFTPEIAERPVRDQ	AGRMNYYWTLLKPGD	TI
ASC18	PKLSKSYVNNKGKVLVWGVHHPPTGID	QQSLYQNADAYVSVGSSKYNRRFTPEIAARPKVRDQ	AGRMNYYWTLLLEPGD	TI
AT91	PNVSKSYVNNKEKVLVWGVHHPNSISD	QRAIYHTENAYVSVSSHYSRRFTPEIAKRPKVRDQ	EGRINYYWTLLLEPGD	TI
ANY18	PKLSKSYVNNKGKVLVWGVHHPPTGT	QQSLYQNADAYVSVGSSKYNRRFTPEIAARPKVRDQ	AGRMNYYWTLLLEPGD	TI
H3 Subtype				
ADU63	PVLNVTMPNNDNFDKLYIWGVHHPSTNQ	QTNLYVQASGRVTVSTRSQQTIPNIGSRPVRGQ	PCRISIIYWTIVKPGD	VL
AAI68	PVLNVTMPNNDNFDKLYIWGVHHPSTNQ	QTSLYVQASGRVTVSTRSQQTIPNIGSRPVRG	GLSSRISIIYWTIVKPGD	VL
AM99	PALNVTMPNNDKFDKLYIWGVHHPSTD	SDQTSLYTQASGRVTVSTKRSQQTVIPNIGSRPVRG	ISSRISIIYWTIVKPGD	IL
H5 Subtype				
ADS97	PTIKRSYNNTNQEDLLILWGVHHPNDAA	EQTKLYQNPTTYVSVGSTLNQRSVPEIATRPKNVC	QSGRMEFFWTILKPNDA	I
Viet04	PTIKRSYNNTNQEDLLYLWGVHHPNDAA	EQTKLYQNPTTYISVGSTLNQRLVPRIATRSKNVC	QSGRMEFFWTILKPNDA	I

FIG. 2B

13/44

FIG. 3

FIG. 3A
FIG. 3B

FIG. 3A

H1_Av	97	139	153	183
H1_Hu1	ENGICPGFDIYEELREQLSSISSEKFEIFPKASSWPNHETTKGTAACSYS-GASSFYRNLIWLTGKG-TS-YPKLSKSYTNKNGKEVLVWGVHHHPSVSEQQSIYQNAD			
H1_Hu2	ENGICPGFDIYEELREQLSSVSSHEKFEIFPKTSSWPNHETTKGTAACSya-GASSFYRNLIWLTGKG-SS-YPKLSKSYVNNKGEVLVWGVHHHPTGTDOQSIYQNAD			
H2_Av	ENGICPGFADYEELREQLSSVSSFERFEIFPKESSWPNHVT-GVSASCSHN-GKSSFYRNLIWLTGKN-CL-YPNLSKSYVNNKEKEVLVWGVHHHPPNIGDQRALYHTEN			
H2_Hu	ANGICPGSFNDYEELKHLITSVTHEKVKILPR-DQWTOHTTGG-SRACAVS-GNPSFFRNWVWLTGKG-SN-YPIAKRSYNNNTSGKQMLVWGVHHHNDDETEQRTLYQNVG			
H3_Av	RDGICPGSFNDYEELKHLSSVKHFEKVKILPK-DRWTQHTTGG-SRACAVS-GNPSFFRNWVWLTGKG-SN-YPIAKRSYNNNTSGEOMLIIWGVHHHNDDETEQRTLYQNVG			
H3_Hu1	FS-NCYPYDIPDYASLSRSLVASSGTFEITEG---FTW-TGVTQNGSSACKRG-PANGFFSRNLWLTGS--ESAYPVLNVTPMNNDFDKLYWGVHHHSTNQEQTDLVYQAS			
H3_Hu2	FS-NCYPYDIPDYASLSRSLVASSGTFEITEG---FTW-TGVTQNGSSACKRG-PCSGFFSRNLWLTGS--GSTYPVLNVTPMNNDFDKLYWGVHHHSTNQEQTDLVYQAS			
H4_Av	YS-NCYPYDIPDYASLSRSLVASSGTFEITENES---FNW-AGVTQNGTSSACKRR-SNKSFFSRNLWLTGHL--KYKYPALNVTPMNNKFDKLYWGVHHHPTDSDQISLYAQAS			
H5_Av1	VD-TCYPFDVPDYQSLRSILANNGKEFIAEE---FQW-NTVKQNGKSGACKRA-NVNDFFNRNLWLTGSN-GDAYPLQNLTKVNNGDYARLYWGVHHHPTDTEQTDLYKNNP			
H5_Av2	VNDICYPGDFNDYEELKHLISRLNHEKIQIIPK-SSWSSHEASLGVSSACPIQ-GKSSFFRNWVWLTGKG-ST-YPTIKRSYNNNTNQEDLLVWGVHHHNDAAEQTKLYQNPT			
H6_Av	ANDICYPGDFNDYEELKHLISRLNHEKIQIIPK-NSWSSHEASLGVSSACPIQ-GKSSFFRNWVWLTGKG-NA-YPTIKRSYNNNTNQEDLLVWGVHHHNDAAEQTKLYQNPT			
H7_Av	QNGICYPGTINEIEELKALIGSGERIERFEMFPK-STWGVNTNNGVTRACPDN-SGSSFYRNLIWLTGKNSAA-YPVIKGTYNNTGNQPILYWGVHHHPTDNTAQNLIYSGSD			
H8_Av	SD-VCYFGKFNVEEALRQILRESGGINKETMG---FTY-SGIRTNCASTCRRS-G-SSFYAEKMLLSNTDNAAFPQMTKSYKNTRKDPALIIWGVHHHSGSTTEQTKLYGSGN			
H9_Av	PEGICYPGSVENLEELRFVSSAASYKRIPLDY-SRWN--VTRSGTSCACNASTGGQSFYRSINWLTGKKPDT-YDFNEGAYVNNEDGDIIFWGVHHHPTDTEQTKLYKNAV			
H10_Av	VNGICYPGNVENLEELRTLFSSASSYQRIQIFPD-TIWN--VTTYGTSCACS-----GSFYRSFWLTQK-SGS-YPVQDAQYTNNREKSIIFWGVHHHPTDTEQTKLYINTD			
H11_Av	IA-ICYPGAFTVEEALRQIKIMESGGIDKISTG---FTYSSINSAGTFRSCMRSGNSFYAEKMLVSKNKGQNFQANTYRNTDQSAEHLIIWGVHHHPTDTEQTKLYGSGN			
H12_Av	TNGICYPGTLENEEELRLKFSGLVLEFSKFAFTS-NGWGAVNSGAGVTAACKEG-SSNSFFRNWVWLTGKG-SSGT-YPVIRTRFNNNTKGRDVLVWGVHHHPTDTEQTKLYKNAV			
H13_Av	MEGICYPGSVENLEELRSLFSSIKKYERVKMFD-TKWN--VTTYGTSCACNASTGGQSFYRSINWLTGKKPDT-YDFNEGAYVNNEDGDIIFWGVHHHPTDTEQTKLYKNAV			
H14_Av	PHGLICYPGELNNGELRHLFSGIRFSRTELIPP-TSWGEVLD--GATSAACRDKGTNSFYRNLIWLVK-KNNR-YPVLSKTYNNNTGRDVLVWGVHHHPTDTEQTKLYKNAV			
H15_Av	VD-TCYPFDVPDYQSLRSILANNGKEFIAEE---FTW-NGVKVDGSSACLRG-GRNSFFSRNLWLTGKG-SN-YPIAKRSYNNNTSGKQMLVWGVHHHNDDETEQRTLYQNVG			
H16_Av	SD-ICYPGKFTNEEALRQILRESGGIDKEPMG---FRY-SGKTGATSAACKRT-V-SSFYSEMMLLSKANQVFPQNLQTYRNNRKEPALIIVWGVHHHSSSDNEQTDLYKVA			
	PNKICFPGEELNNGELRHLFSGVNSFSRTELISP-NKWGDILD--GVTASCRDN-GASSFYRNLIWLVKKNCK-YPVIKGDYNNNTGRDVLVWGVHHHPTDTEQTKLYKNAV			

14/44

324	333
H1_Av	IGECPKYVKSITKLRLMATGLRNVPESIQ-----SRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H1_Hu1	IGECPKYVRSITKLRLMATGLRNVPESIQ-----SRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H1_Hu2	IGECPKYVRSITKLRLMATGLRNVPESIQ-----SRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H2_Av	IGECPKYVKSITKLRLMATGLRNVPESIQ-----SRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H2_Hu	IGECPKYVKSITKLRLMATGLRNVPESIQ-----SRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H3_Av	YGACPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H3_Hu1	YGACPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H3_Hu2	YGACPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H4_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H5_Av1	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H5_Av2	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H6_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H7_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H8_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H9_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H10_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H11_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H12_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H13_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H14_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H15_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN
H16_Av	IGDCPKYVKSITKLRLMATGLRNVPESIQ-----QTRGLFCAIAGFIIEGCGTGMIDGWYGVYHHONEQSGGYAADQKSTQNAIDGITSKVNSEIEKMTQFTAVGKFNFLERRIENLN

FIG. 3B

15/44

FIG. 4

FIG. 4A
FIG. 4B

FIG. 4A

97	H1_Av	ENGICYPGEFIDYEELREQLSSISSEKFEIPEKASSWNEHTTKGTAACSYS-GASSFYRNLLMTTKG-TS-YPKLSKSYTNKKGKVLVINGVHHBPSVSEQOSLYQNAD	139	153	183
	H1_Hu1	ENGICYPGDFIDYEELREQLSSVSSEKFEIPEKTSWNEHTTKGTAACSYA-GASSFYRNLLMTTKG-SS-YPKLSKSYVNNKGKVLVINGVHHBPTGTDOOSLYQNAD			
	H1_Hu2	ENGICYPGFADYEELREQLSSVSSEKFEIPEKSSWNEHTVT-GVSASCSHN-GKSSFYRNLLMTGKN-GL-YPNLSKSYVNNKEKVLVINGVHHBPNIGDQALYHTEN			
	H2_Av	ANGICYPGSFNDYEELKHLITSVTHFEKVKIIPR-DQWQHTTTGG-SRACAVS-GNPSFFRNVMVLTGK-SN-YPIAKRSYNNNTSGKQMLVINGVHHBNDDEQRTLYQNVG			
	H2_Hu	RDGICYPGSFNDYEELKHLISSVHKFEKVKIIPK-DRWQHTTTGG-SRACAVS-GNPSFFRNVMVLTGK-SN-YPIAKRSYNNNTSGKQMLVINGVHHBNDDEQRTLYQNVG			
	H3_Av	FS-NCYPYDIPDYASLSLVASSGTLIEFITEG---FTW-TGYTQNGGSSACKRG-PANGFFSRNLNMTKS--ESAYPVLNVTPNNDNFDKLVINGVHHBNDDEQRTLYQNVG			
	H3_Hu1	FS-NCYPYDIPDYASLSLVASSGTLIEFITEG---FTW-TGYTQNGGSSACKRG-PGSGFFSRNLNMTKS--GSTYPVLNVTPNNDNFDKLVINGVHHBNDDEQRTLYQNVG			
	H3_Hu2	YS-NCYPYDIPDYASLSLVASSGTLIEFNES---FNW-AGVTQNGTSSACKRR-SNKSFFSRNLNMTLTHL--KYKYPALNVTPNNEKFDKLVINGVHHBNDDEQRTLYQNVG			
	H4_Av	VD-TCYPDFDPDYQSLRSILANNGRFEFIAEE---FQW-NTVKQNGKSGACKRA-NVNDFFRNLNMTLTHL--KYKYPALNVTPNNEKFDKLVINGVHHBNDDEQRTLYQNVG			
	H5_Av1	VNDICYPGDNDEYELKHLISRINHFEKIQIIPK-SSWSHEASLGVSACPYQ-GKSSFFRNVMVLTGKN-GDAYPIQLNLTKNNGDYARLVINGVHHBNDDEQRTLYQNVG			
	H5_Av2	ANDICYPGDNDEYELKHLISRINHFEKIQIIPK-NSWSHEASLGVSACPYQ-GKSSFFRNVMVLTGKN-NA-YPIKRSYNNNTSGKQMLVINGVHHBNDDEQRTLYQNVG			
	H6_Av	QNGICYPGILNEIEELKALIGSGERIERFEMPFPK-STWGSVNTNNGVTRACPDN-SGSSFYRNLLMTTKNSAA-YPIKGTNNNTGQPILVINGVHHBNDDEQRTLYQNVG			
	H7_Av	SD-VCYPCKFVNEELRQILRESGGINKETMG---FTY-SGIRTINGATSTCRRS-G-SSFYAEMKMLLSWTDNAAFPQMTKSYNTRKDPALLINGVHHBNDDEQRTLYQNVG			
	H8_Av	PEGMCYPGSVENLEELRFVSSAASYKRIRLEDY-SRWNN--VTRSGTSKACNASIGGQSFYRSINWLTGKKPDT-YDFNEGAYVNNEDGDIIFINGVHHBNDDEQRTLYQNVG			
	H9_Av	VNGICYPGNVENLEELRFLSSASSYQRIQIPD-TIWN--VTYTCTSKACS-----GSFYASMRWLTQK-SGS-YPIVQDAQIYNNREKSIILFVINGVHHBNDDEQRTLYQNVG			
	H10_Av	IA-YCYPGATVNEELRQKIMESGIDKISG---FTYGSINSAGTTRSCMR-S-GGNSFYAELKWLVSNNKQNFQNTANTYRTDSEHLLINGVHHBNDDEQRTLYQNVG			
	H11_Av	TNGICYPGILNEEELRLKFSGLVLEPSKFEAFTS-NGWCAVNSGAGVTAACKFG-SSNSFFRNVMVLTGKN-QSGT-YPIVIRRTFNNTKGRDVLVINGVHHBNDDEQRTLYQNVG			
	H12_Av	MEGVICYPGSNIENQELRSLFSSIAKKYERVMNDF-TKWN--VTYTGTSRACNNNTSNRGSFYRSMRWLTGK-SGQ-FPVQTDYKNTNRDSDILFTWAIHHBPTSAEQVQLYKNPD			
	H13_Av	PHGICYPGLNNGELRHLFSGIRSFRTLEIPP-TSWGEVLD--GATSACRDDKGTNSFYRNLMVFK-KNNR-YPIVSKTYNNNTGRDVLVINGVHHBNDDEQRTLYQNVG			
	H14_Av	VD-TCYPDFDPDYQSLRSILASSGSLFIAEQ---FTW-NGVKVDCSSSACLRG-GRNSFFSRNLNMTLTHL--NGNIGPINVTKENTGSYRDLVINGVHHBNDDEQRTLYQNVG			
	H15_Av	SD-ICYPGFTNEELRQILRESGIDKEPMG---FRI-SGIKTDGATSACKRT-V-SSFYSEMKMLLSKANQVLPQNTQIRNNRKEPALIVINGVHHBNDDEQRTLYQNVG			
	H16_Av	PNKICFPGLDNNNGELRHLFSGVNSFSRTELISP-NKWDILD--GVTSACRDN-GASSFYRNLMVLTGKN-KNGK-YPIVKGDIYNNNTGRDVLVINGVHHBNDDEQRTLYQNVG			

16/44

324	333	
H1_Av	IGGCPKYVKSFKLRMTGLRNVP	PSIQ-----SRGLFCALAGFIEGGWTGMDGWYGVHHQNEQSGGYAADOKSTONADIGITSKVNSVIEKMNTOFTAVGKFNNLERRIENLN
H1_Hu1	IGGCPKYVRSFKLRMTGLRNVP	PSIQ-----SRGLFCALAGFIEGGWTGMDGWYGVHHQNEQSGGYAADOKSTONADIGITKNVNSVIEKMNTOFTAVGKFNNLERRIENLN
H1_Hu2	IGGCPKYVRSKLRMTGLRNVP	PSIQ-----SRGLFCALAGFIEGGWTGMDGWYGVHHQNEQSGGYAADOKSTONADIGITKNVNSVIEKMNTOFTAVGKFNNLERRIENLN
H2_Av	IGGCPKYVKSRLVLTGLRNVP	QIE-----SRGLFCALAGFIEGGWQMGMDGWYGVHHSNDQSGGYAADKSTQAKADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H2_Hu	IGGCPKYVKSRLVLTGLRNVP	QIE-----SRGLFCALAGFIEGGWQMGMDGWYGVHHSNDQSGGYAADKSTQAKADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H3_Av	YGACPKYVKNTLKLATGLRNVP	PGK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H3_Hu1	YGACPKYVKNTLKLATGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H3_Hu2	YGACPKYVKNTLKLATGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H4_Av	IGGCPKYVKSRLVLTGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H5_Av1	IGGCPKYVKSRLVLTGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H5_Av2	IGGCPKYVKSRLVLTGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H6_Av	IGGCPKYVKSRLVLTGLRNVP	PEK-----QTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H7_Av	VGKCPKYVKQESLMLATGMKNVP	PEIP-----KGRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H8_Av	MGFCPKYVKASRLVLTGLRNVP	PSVE-----PRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H9_Av	FGICPKYVRSKLRVLTGLRNVP	PARS-----NRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H10_Av	VGQCPKYVKNKSLLLATGMKNVP	PEVV---QGRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H11_Av	IGDCKPYVNVKSLKLATGLRNVP	PAIA-----TRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H12_Av	IGKCPKYIIPSGSLKLATGLRNVP	POVQ-----NRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H13_Av	LGDCPKYIKSGQLKLATGLRNVP	PAIS-----NRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H14_Av	IGNCPKYVKQSGSLMLATGMKNVP	IPCK-----QAKGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H15_Av	VGRCPRYVKQSSPLALGMKNVP	PEKI---HTRGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN
H16_Av	LGDCPKYIKSGQLKLATGLRNVP	PSVG-----ERGLFCALAGFIEGMEGMDGWYGFRRHNSQSGGYAADLKSQAAADIGITKNVNSVIEKMNTOFEAVGKFNNLERRIENLN

FIG. 4B

FIG. 5A

FIG. 5A-1
FIG. 5A-2

FIG. 5A-1

	97	139	153	183
H1_Av	ENGTCYPGDFIDYEEELREQLSSISSEKFEI	FPKASSWPNHETTKGVTAA	CSYS-CASSFYRNLI	WITKKG-TS-YPKLSKSYTNKKEVLV
H1_Hu1	ENGTCYPGDFIDYEEELREQLSSVSSFEKFEI	FPKTSWPNHETTKGVTAA	CSYA-CASSFYRNLI	WITKKG-SS-YPKLSKSYVNNKKEVLV
H1_Hu2	ENGTCYPGDFADYEEELREQLSSVSSFEKFEI	FPKESWPNHVT-VGSASCS	SHN-GKSSFYRNLI	WITGKN-GL-YPNLSKSYVNNKKEVLV
H2_Av	ANGLCYPSFENDYEEELKHLITSYTHFEKVIL	PR-DWTQHTT	TGG-SRACAVS-GNPSFFRN	WVLTGK-SN-YPVAKGSYNNNTSGKQ
H2_Hu	RDGLCYPSFENDYEEELKHLSSVKHFEKVIL	PK-DRWTQHTT	TGG-SRACAVS-GNPSFFRN	WVLTGK-SN-YPVAKGSYNNNTSGEOM
H3_Av	FS-NCYPVDIPDYASIRSLVASSGTLEFITEG	--FTW-TGVTQNGSSACKRG-PANGFFSR	LNWLTGS--GSTY	PVLNVTMPNNDNFDKLY
H3_Hu1	FS-NCYPVDIPDYASIRSLVASSGTLEFITEG	--FTW-TGVTQNGSSACKRG-PGSGFFSR	LNWLTGS--GSTY	PVLNVTMPNNDNFDKLY
H3_Hu2	YS-NCYPVDIPDYASIRSLVASSGTLEFNES	---FNW-AGVTQNGTSSACKRR-SNKSFFSR	LNWLTGHL--KYKY	PALNVTMPNNEKFDKLY
H4_Av	VD-TCYFDPDVPDYQSLRSILANNCKFE	FTAE---FQW-NTVKQNGKSGACKRA-NVND	FFNRLNWLTKSN-GDAY	PLQNLTKVNGDYARLY
H5_Av1	VNDICYPGDFNDYEEELKHLRSINHFEEK	LIQIIPK-SSWSSEASLGVSACPIQ-GKSS	FFRNWVWLTKN-ST-YPT	KRSYNNNTNQEDLLV
H5_Av2	ANDICYPGDFNDYEEELKHLRSINHFEEK	LIQIIPK-NSWSSSEASLGVSACPIQ-GKSS	FFRNWVWLTKN-NA-YPT	KRSYNNNTNQEDLLV
H6_Av	QNGICYPGTINEIEELKALIGSERIEREF	EMFPK-STWSGVNTNNGVTRACP	DN-SGSSFYRNLI	WITKNSAA-YPVI
H7_Av	SD-VCYPGKFVNEELRQILRESGGINK	FTMG---FTY-SGIRTWGATST	CRRS-G-SSEYAE	MKLLSNTDNA
H8_Av	PEGMCYPGSVENLEELRFVSSAASYKR	IRLFDY-SRWN--VTRSGTSKAC	NASTGGQSFYRS	INWLTGKKPDT-YD
H9_Av	VNGTCYPGNVENLEELRTLFSASSYQRI	QIPPD-TIWN--VITYGT	SKACS-----GSFY	SRMRLTQK-SGS-YPVQ
H10_Av	IA-YCYPGATVNEELRQIKIMESGGID	KISTG---FTYGSS	INSAGTTRSCVRS-CGNS	FVAELKWLVS
H11_Av	TNGICYPGTLENNEELRLKFGVLEFS	KEAFTS-NGGAVNSGAGVTA	ACKFG-SSNS	FFRNWVWLH-QSGT-YPV
H12_Av	MEGVCYPGSIENQELRSLFSSIKKYER	KMFDF-ITKN--VITYGTS	RACNNTSNRCSFY	SRMRLTLK-SGQ-FPV
H13_Av	PHGICYPGELNNGELRHLFGIRSF	SRFELIPP-TSGEVL	--GATSAC	DDKGTNSFYRN
H14_Av	VD-TCYFDPDVPDYQSLRSILASSG	LEFIAEQ---FTW-NGKV	DGSSACLRG-GRNS	FFSR
H15_Av	SD-ICYPCKFTNEELRQIIRRESGGID	KEPMG---FRY-SGIK	TDGATSA	CKRT-V-SSFY
H16_Av	PNKICFPGEELDNNGELRHLFGVNS	FORTLISP-NKWGD	ILD--GVTA	SRDN-GASS
				FYRN
				WLVK
				NGK-YPVI
				KGYNN
				TG
				RD
				VL
				MG
				IHH
				FP
				VS
				VE
				ET
				KT
				LY
				VNS
				D
				Y
				VE
				HP
				SS
				DN
				EC
				TD
				LY
				KV
				AT
				V
				AT
				KE
				QD
				LY
				K
				DS
				IL
				Y
				VE
				PT
				SA
				EQ
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				GS
				G
				SG
				SN
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA
				QN
				N
				LY
				GS
				G
				SG
				ST
				TE
				Q
				K
				LY
				KN
				P
				DP
				NA

18/44

324 333

H1_Av IGECPKYVKSRLMATGLRNVPISIQ-----SRGLEGAIAAGTIEGGWTGMIDGWYGYHHONEQSGGYAADQKSTQNAIDGITSKVN SVIEKMNTOFTAVGKEFNFLERRIENLN
 H1_Hu1 IGECPKYVRSKRLMATGLRNVPISIQ-----SRGLEGAIAAGTIEGGWTGMIDGWYGYHHONEQSGGYAADQKSTQNAIDGITSKVN SVIEKMNTOFTAVGKEFNFLERRIENLN
 H1_Hu2 IGECPKYVRSKRLMATGLRNVPISIQ-----SRGLEGAIAAGTIEGGWTGMIDGWYGYHHONEQSGGYAADQKSTQNAIDGITSKVN SVIEKMNTOFTAVGKEFNFLERRIENLN
 H2_Av IGECPKYVKSRLVATGLRNVPQIE-----SRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H2_Hu IGECPKYVKSRLVATGLRNVPQIE-----SRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H3_Av YGACPKYVKQNTLKLATGMNVPKG-----QTRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H3_Hu1 YGACPKYVKQNTLKLATGMNVPKG-----QTRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H3_Hu2 YGACPKYVKQNTLKLATGMNVPKG-----QTRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H4_Av IGDCKPYVKQSLKLATGMNVPKG-----ATRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H5_Av1 IGECPKYVKSRLVATGLRNVPQIERRRRRKKRLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H5_Av2 IGECPKYVKSRLVATGLRNVPQIERRRRRKKRLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H6_Av IGECPKYVKSRLVATGLRNVPQIA-----TRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H7_Av VGKCPRYVKQSLMLATGMNVPKIP-----KGRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H8_Av MGECPKYVKSRLVATGLRNVPKIP-----PRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H9_Av FGICPKYVVRKSLKLAVGLRNVPKIP-----NRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H10_Av VGQCPKYVNRKSLKLATGMNVPKIP-----QGRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H11_Av IGDCKPYVNRKSLKLATGMNVPKIP-----TRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H12_Av IGECPKYVPSGSLKLATGMNVPKIP-----NRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H13_Av LGDCKPYIKSGQLKLATGMNVPKIP-----NRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H14_Av IGECPKYVKSRLVATGLRNVPKIP-----QAKGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H15_Av VGRCPRYVKQSLPLALGMNVPKIP-----HTRGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN
 H16_Av LGDCKPYIKSGQLKLATGMNVPKIP-----ERGLEGAIAAGTIEGGWQGMVDGWYGYHHSNDQSGGYAADKESQKADIDGITSKVN SVIEKMNTOFFAVGKEFNFLERRIENLN

FIG. 5A-2

19/44

FIG. 5B-1
FIG. 5B-2
FIG. 5B-3
FIG. 5B-4
FIG. 5B-5
FIG. 5B-6

FIG. 5B

20/44

Positions from 1 till 60	
consensus	MEKIVLLLAIVSLVKSDQICIGYHANNSTEQVDTIMEKNVTVTHAQDILEKTHNGKLCDL
AAL59142	
AAZ29963	 F
ABA70758	 F
ABB87042	 I A I K E S .
ABD14810	
ABD46740	- - - -
ABD85144	- - - - - F
ABE97569	
Positions from 61 till 120	
consensus	DGVKPLILRDCSVAGWLLGNPNCDEFINVPESYIVEKANPANDLCYPGDFNDYEELKEEL
AAL59142	 S . D
AAZ29963	 V
ABA70758	 V
ABB87042	K K L D I G
ABD14810	 K I I
ABD46740	 L I I N
ABD85144	 L I I N
ABE97569	 N N

FIG. 5B-1

```

Positions from 121 till 180
consensus  LSRINHFKEIQIIPKSSWSDHEASSGVSSACPYQCKSSFFERNVWVLKKNSAYPTIKRSY
AAL59142      N . . . . . N . . . . . H . . . . .
AAZ29963      S . . . . . L . . . . . T . . . . .
ABA70758      S . . . . . L . . . . . P . . . . . T . . . . .
ABB87042      R . . . . . N D . . . . . N R . . . . . N . . . . . T .
ABD14810      . . . . . . . . . . . . . . . . . . . . . . I .
ABD46740      . . . . . . . . . . . . . . . . . . . . . . DN .
ABD85144      . . . . . . . . . . . . . . . . . . . . . . DN .
ABE97569      . . . . . . . . . . . . . . . . . . . . . . L .

```

Positions from 181 till 240

NNTNQEDLLVLWGIIHHHPNDAAEQTKKLYQNP	TTT	YISVGTSTLNQRLV	PKIATRSKVN	QSG
AAI59142				
AAZ29963			R	
ABA70758	V		R	
ABB87042	I	I	SN	V
ABD14810				
ABD46740			R	
ABD85144			R	
ABE97569			R	

FIG. 5B-2

22/44

Positions from 241 till 300	
consensus	RMEFFWTILKPNDAINFESNGNFIAP EYAYKI VKKGDS TIMKSBLEYGNCNTK CQT PMGA
AA159142	 A
AA729963	
ABA70758	
ABB87042	 S A D V .
ABD14810	 V D
ABD46740	 N I .
ABD85144	 S N I .
ABE97569	 A

Positions from 301 till 360

consensus	I N S S M P F H N I H P L T I G E C P K Y V K S N R L V L A T G L R N S P Q R E R R R K K R G L F G A I A C F I E G G W
AAL59142	 T G
AAZ29963	
ABA70758	 K
ABB87042	 V - - - T
ABD14810	 -
ABD46740	 G
ABD85144	 G
ABE97569	

FIG. 5B-3

23/44

Positions from 361 till 420

consensus	Q G M V D G W Y G Y H H S N E Q G S G Y A A D K E S T Q K A I D G V T N K V N S I I D K M N T Q F F E A V G R E F N N L E
AAI59142	 K
AAZ29963	
ABA70758	 K
ABB87042	 I T . K
ABD14810	
ABD46740	
ABD85144	
ABE97569	

Positions from 421 till 480

consensus	R R I E N L N K K M E D G F L D V W T Y N A E L L V L M E N E R T L D F H D S N V K N L Y D K V R L Q L R D N A K E L G
AAI59142	
AAZ29963	
ABA70758	
ABB87042	
ABD14810	
ABD46740	
ABD85144	
ABE97569	

FIG. 5B-4

```

Positions from 481 till 540
NGCFEFYHKCDNECMESVRNGTYDYPQYSEEARLKRFEISGVKLESIGTYQILSIYSTVA
. . . . . K . . . . . N . . . . . M . . . . .
AA159142
. . . . .
AA729963
. . . . .
ABA70758
. . . . . S . . . . . D . . . . . M . . . . .
ABB87042
. . . . .
ABD14810
. . . . . R . . . . .
ABD46740
. . . . . R . . . . .
ABD85144
. . . . . I . . . . . N . . . . . I . . . . .
ABE97569

```

```

Positions from 541 till 568
consensus      S S L A L A I M V A G L S L W M C S N G S L Q C R I C I
AAL59142      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
AAZ29963      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABA70758      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABB87042      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABD14810      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABD46740      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABD85144      . . . . . . . . . . . . . . . . . . . . . . . . . . . .
ABE97569      . . . . . . . . . . . . . . . . . . . . . . . . . . . .

```

25/44

AAL59142: 568 Avian 4 (HA) H5N1 Hong Kong 2000 Influenza A virus (A/Goose/Hong Kong/385.3.2000(H5N1))

AAZ29963: 556 Avian 4 (HA) H5N1 Thailand 2004 Influenza A virus (A/Ostrich/Samut Prakan/Thailand/CU-19/04(H5N1))

AB470758: 568 Avian 4 (HA) H5N1 Belgium 2004 Influenza A virus (A/crested eagle/Belgium/01/2004(H5N1))

ABB87042: 564 Avian 4 (HA) H5N2 Canada 1976/08/12 Influenza A virus (A/mallard duck/ALB/57/1976(H5N2))

ABD14810: 567 Avian 4 (HA) H5N1 China 2004 Influenza A virus (A/duck/Guangxi/13/2004(H5N1))

ABD46740: 556 Avian 4 (HA) H5N1 Nigeria 2006/01/17 Influenza A virus (A/chicken/Nigeria/641/2006(H5N1))

ABD85144: 557 Avian 4 (HA) H5N1 Egypt 2006 Influenza A virus (A/chicken/Egypt/960N3-004/2006(H5N1))

ABE97569: 553 Avian 4 (HA) H5N1 Indonesia 2004 Influenza A virus (A/turkey/Kedaton/BPPV3/2004(H5N1))

FIG. 5B-6

26/44

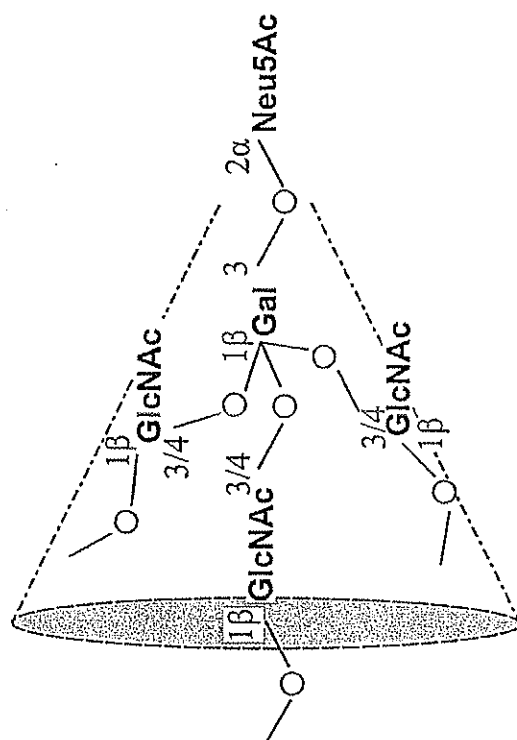
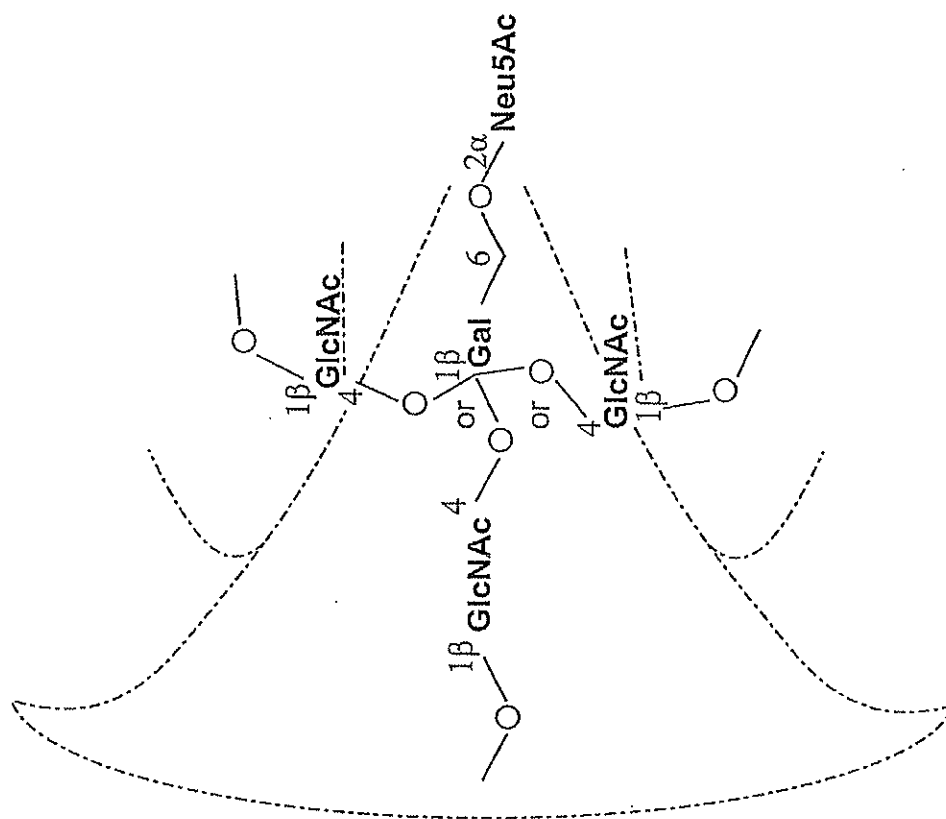


FIG. 6

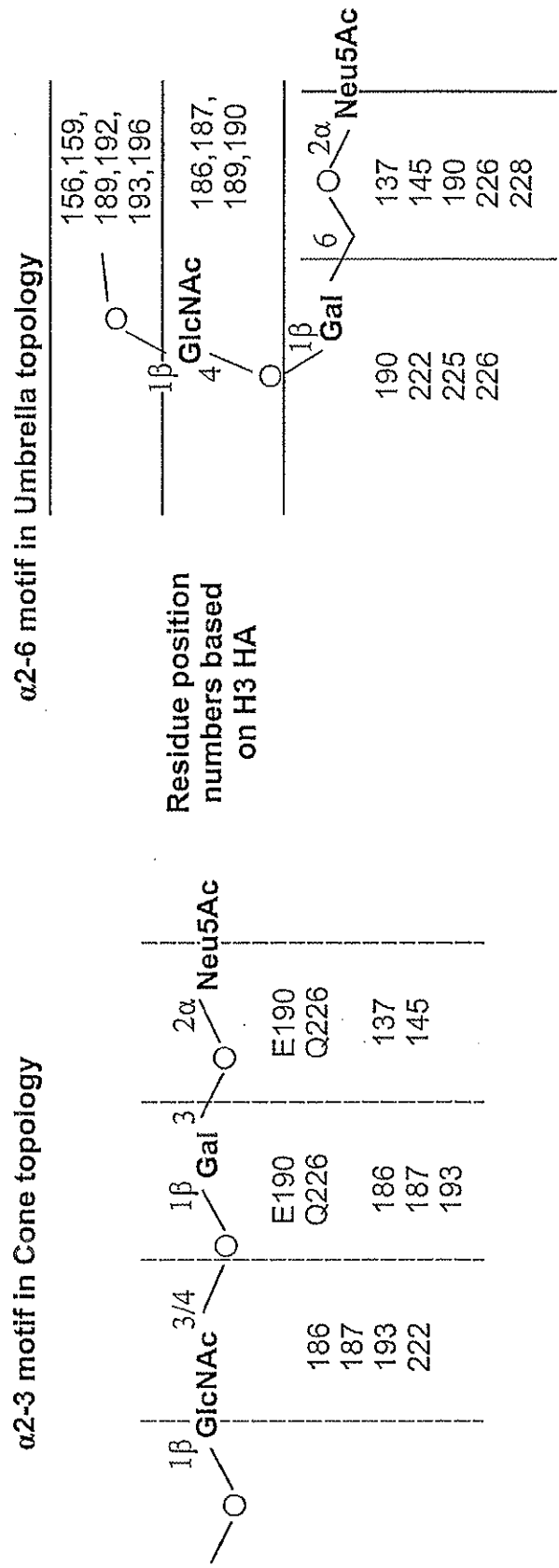


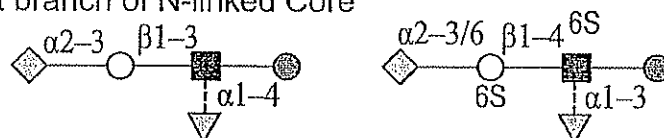
FIG. 7

28/44

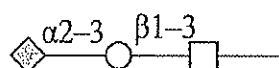
 $\alpha 2-3$ and $\alpha 2-6$ motif in Cone topology

- Typical of short oligosaccharide or oligosaccharide branch attached to a Core Structure

- Short branch of N-linked Core



- Short branch of O-linked Core



- The Cone topology can also be adopted by longer $\alpha 2-3$ and $\alpha 2-6$ oligosaccharide branch attached to Core Structure

◆ Neu5Ac ▼ Fuc

○ Gal ● Glc ● Man

□ GalNAc ■ GlcNAc

Dotted Gray lines, 4S and 6S indicate potential sites for fucosylation and sulfation modifications

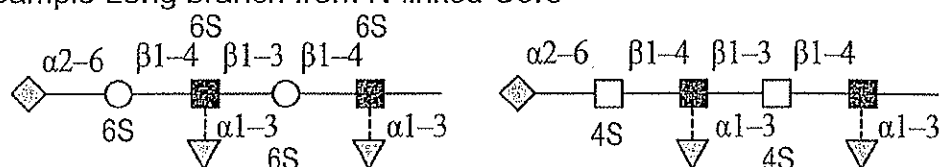
FIG. 8

29/44

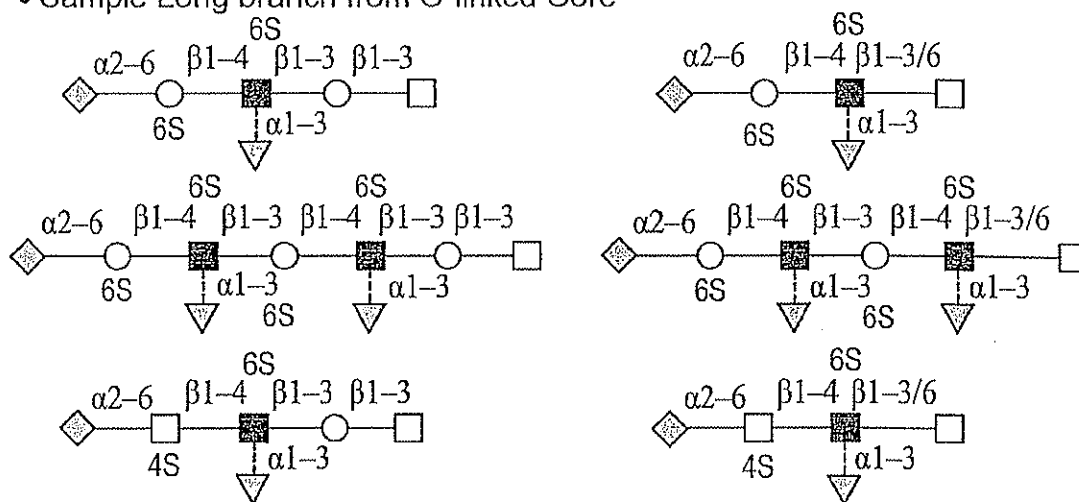
 α 2-6 motif in Umbrella topology

- Typical of longer α 2-6 (>trisaccharide) attached to Core Structure

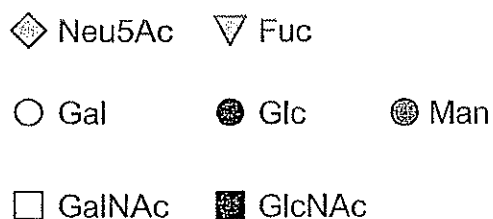
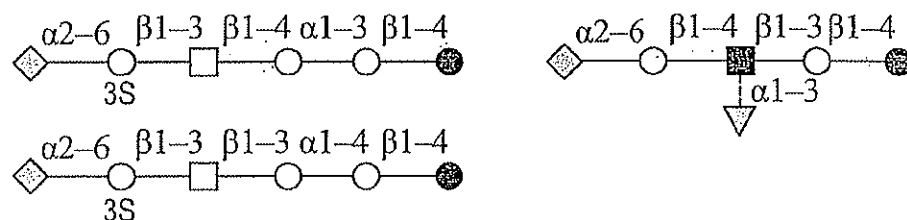
- Sample Long branch from N-linked Core



- Sample Long branch from O-linked Core



- Sample Long branch from Glycolipid Core



Dotted Gray lines, 4S and 6S indicate potential sites for fucosylation and sulfation modifications

FIG. 9

30/44

Human Bronchial Epithelial (HBE) Cells

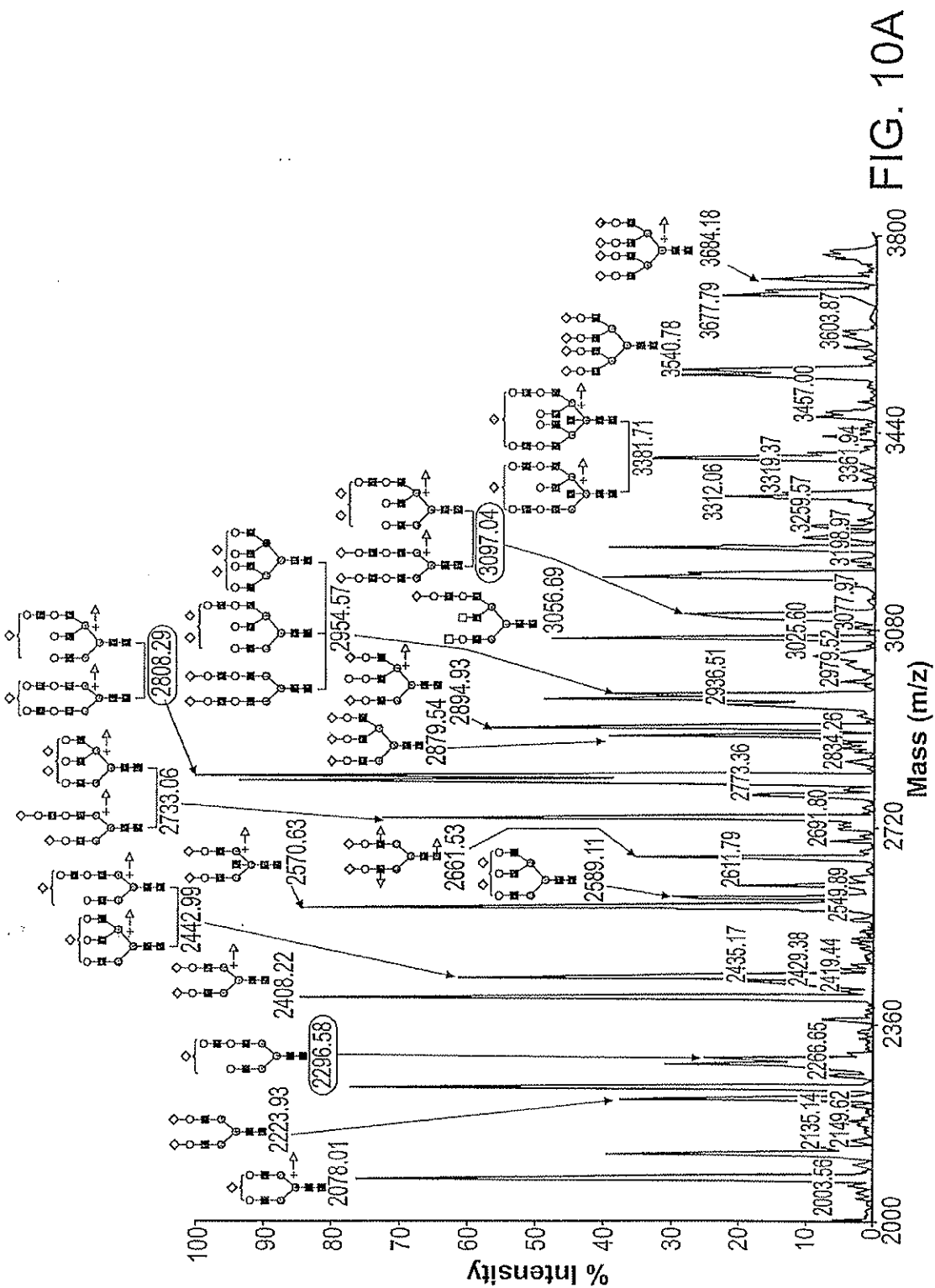


FIG. 10A

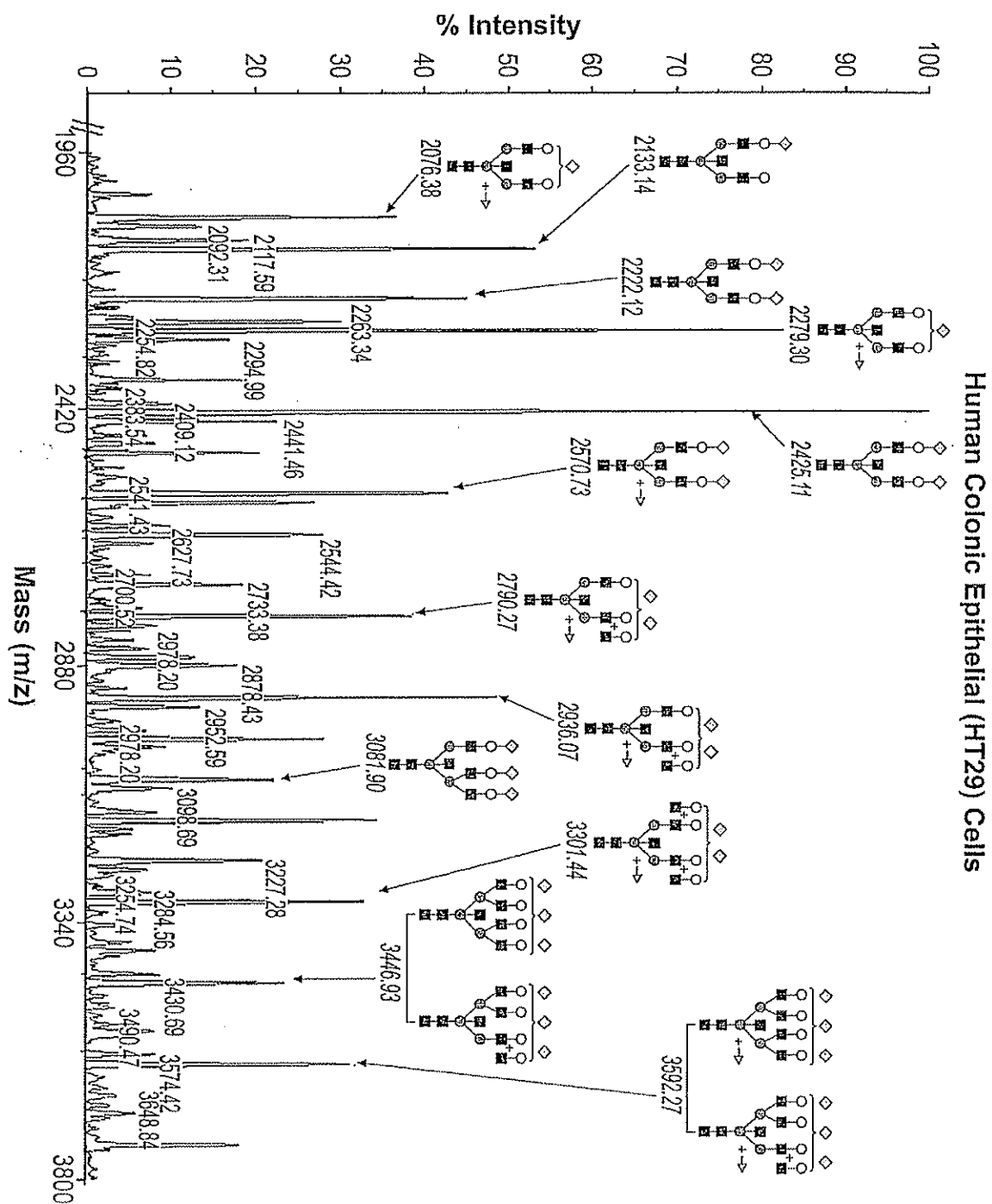


FIG. 10B

32/44

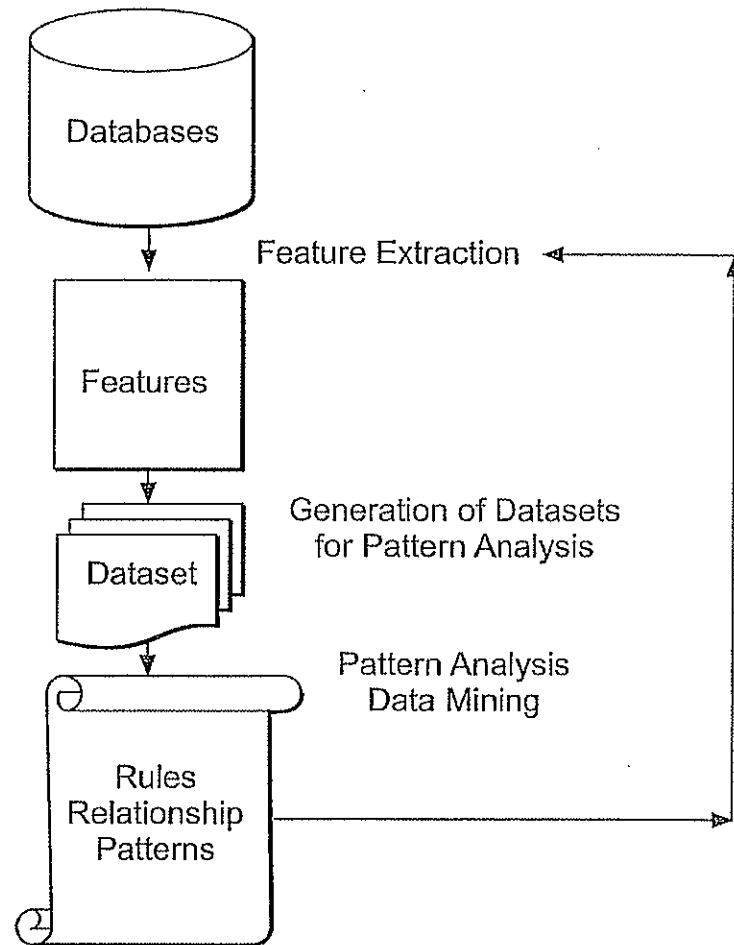


FIG. 11A

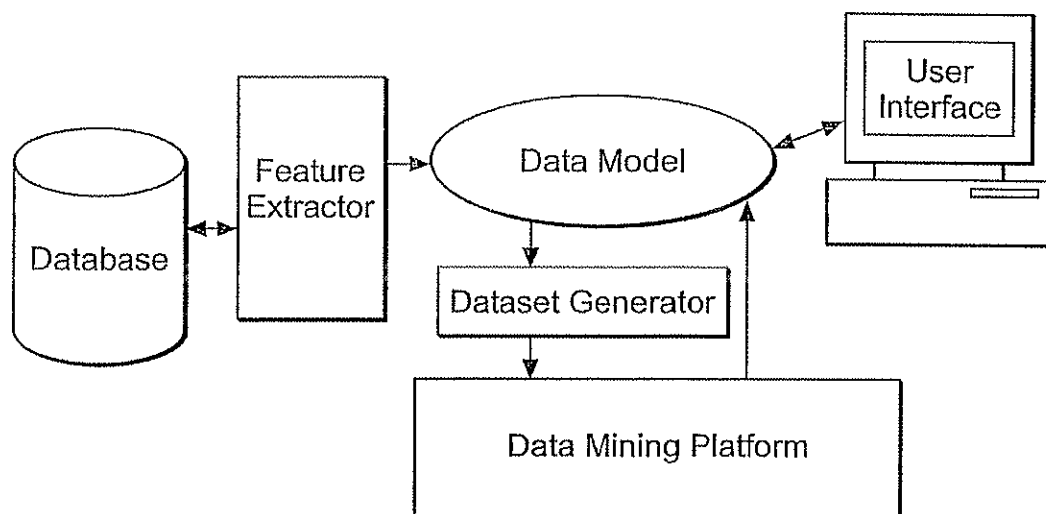


FIG. 11B

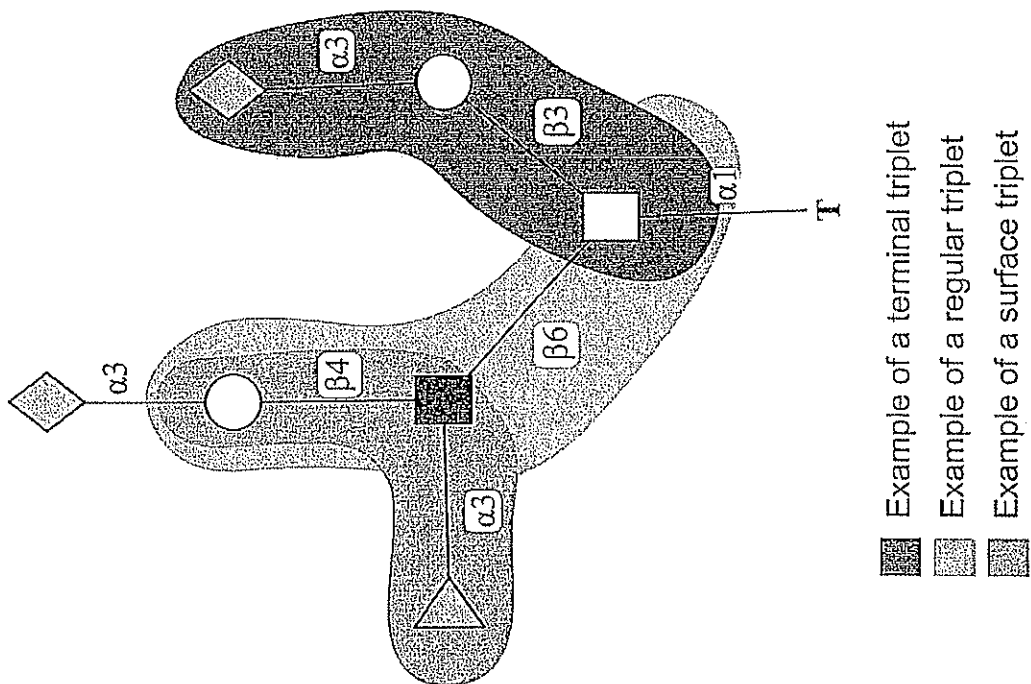


FIG. 12B

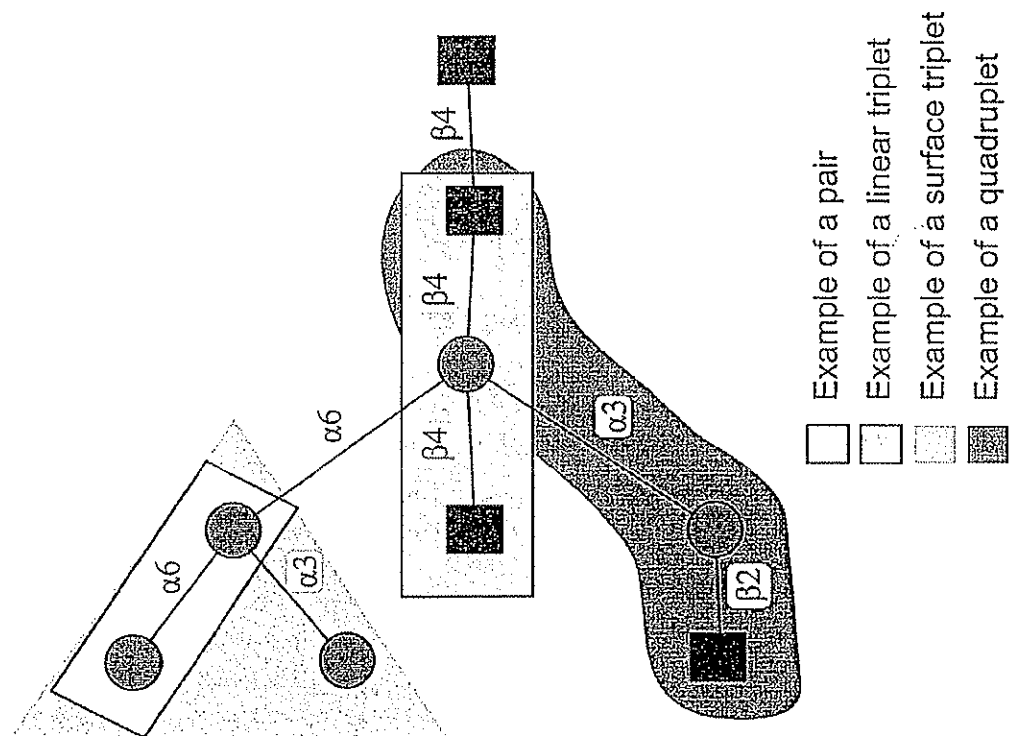


FIG. 12A

34/44

Classifier Id	Classifier rule
α 2-3 Type A	Neu5Aca α 3Gal & !GalNAc β 4Gal
α 2-3 Type B	Neu5Aca α 3Gal β 4GlcNAc & !GalNAc β 4Gal & {GlcNAc β 3Gal or GlcNAc[6s]}
α 2-3 Type C	Neu5Aca α 3Gal β & !GalNAc β 4Gal & !Fuc α 3/4GlcNAc
α 2-6 Type A	Neu5Aca α 6Gal β 4GlcNAcb2Man α 3Man
α 2-6 Type B	Neu5Aca α 6Gal β 4GlcNA & !GlcNAcb2Man α 3Man

FIG. 13

35/44

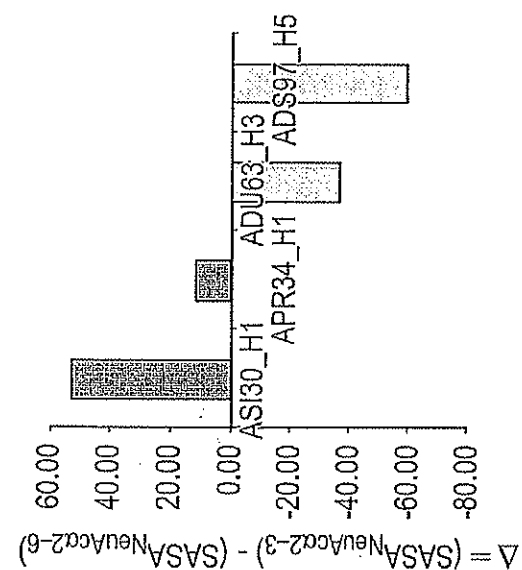


FIG. 14C

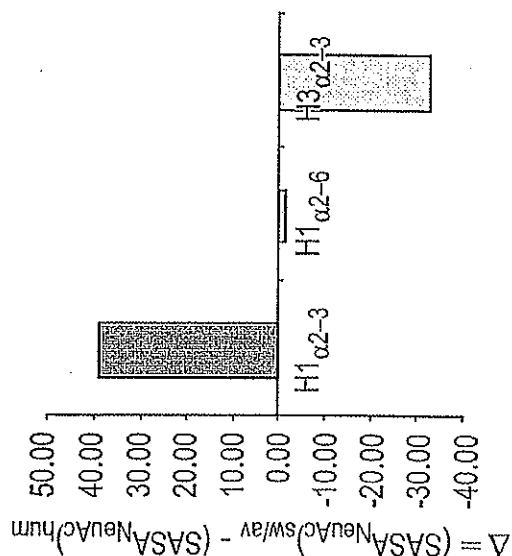


FIG. 14D

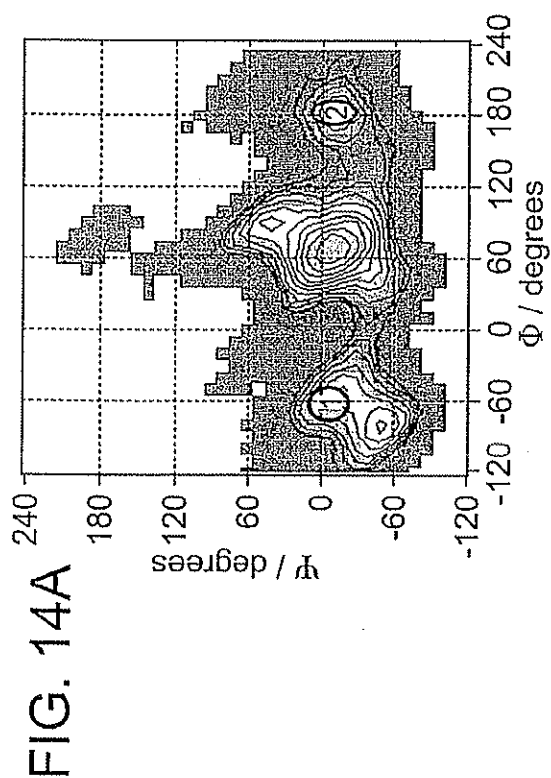


FIG. 14A

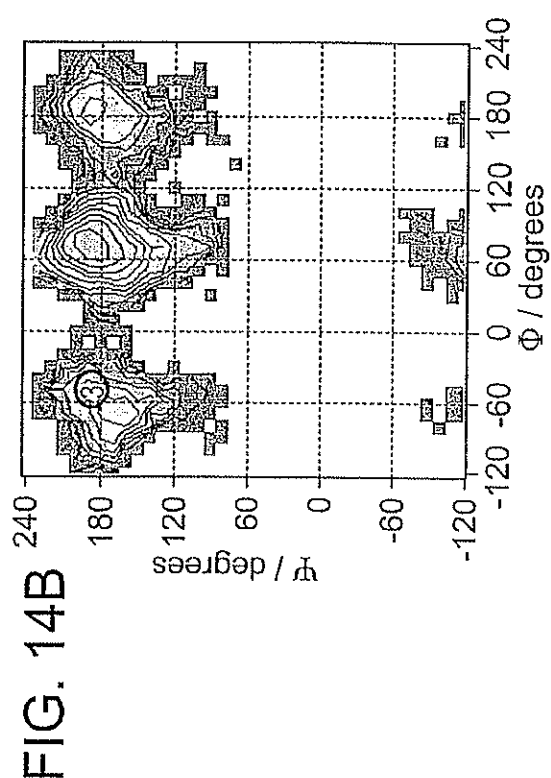


FIG. 14B

36/44

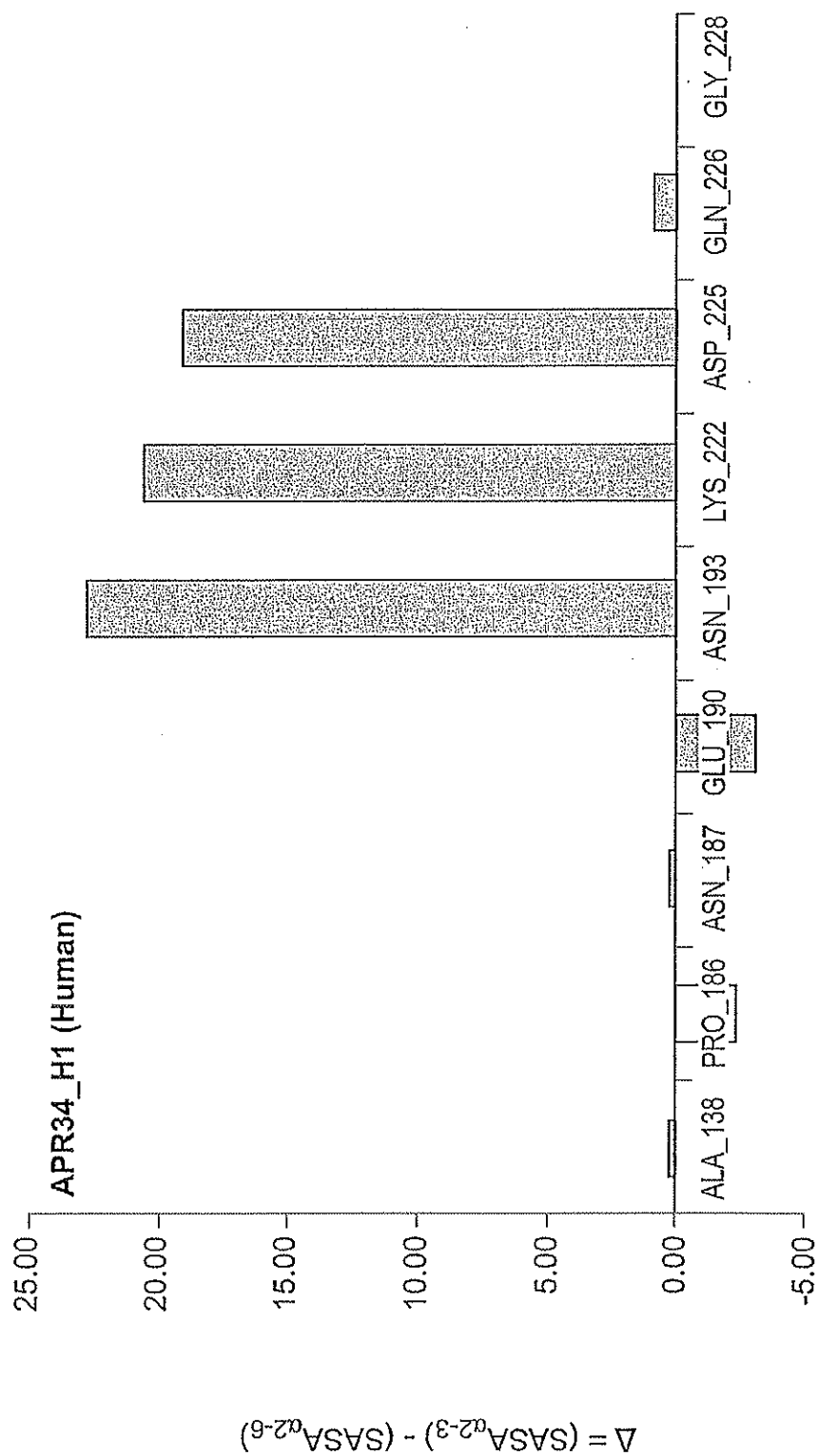


FIG. 15A

37/44

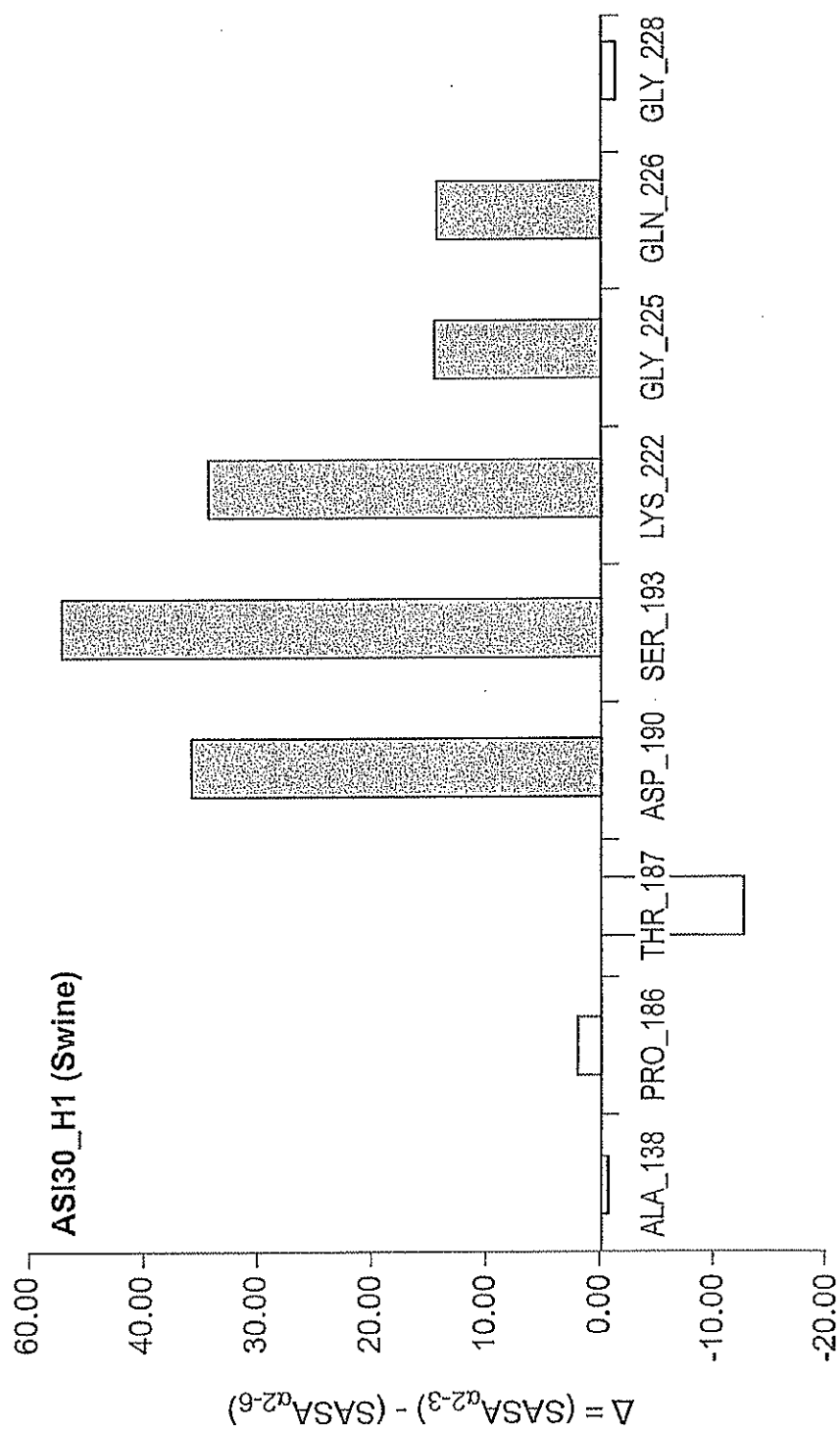


FIG. 15B

38/44

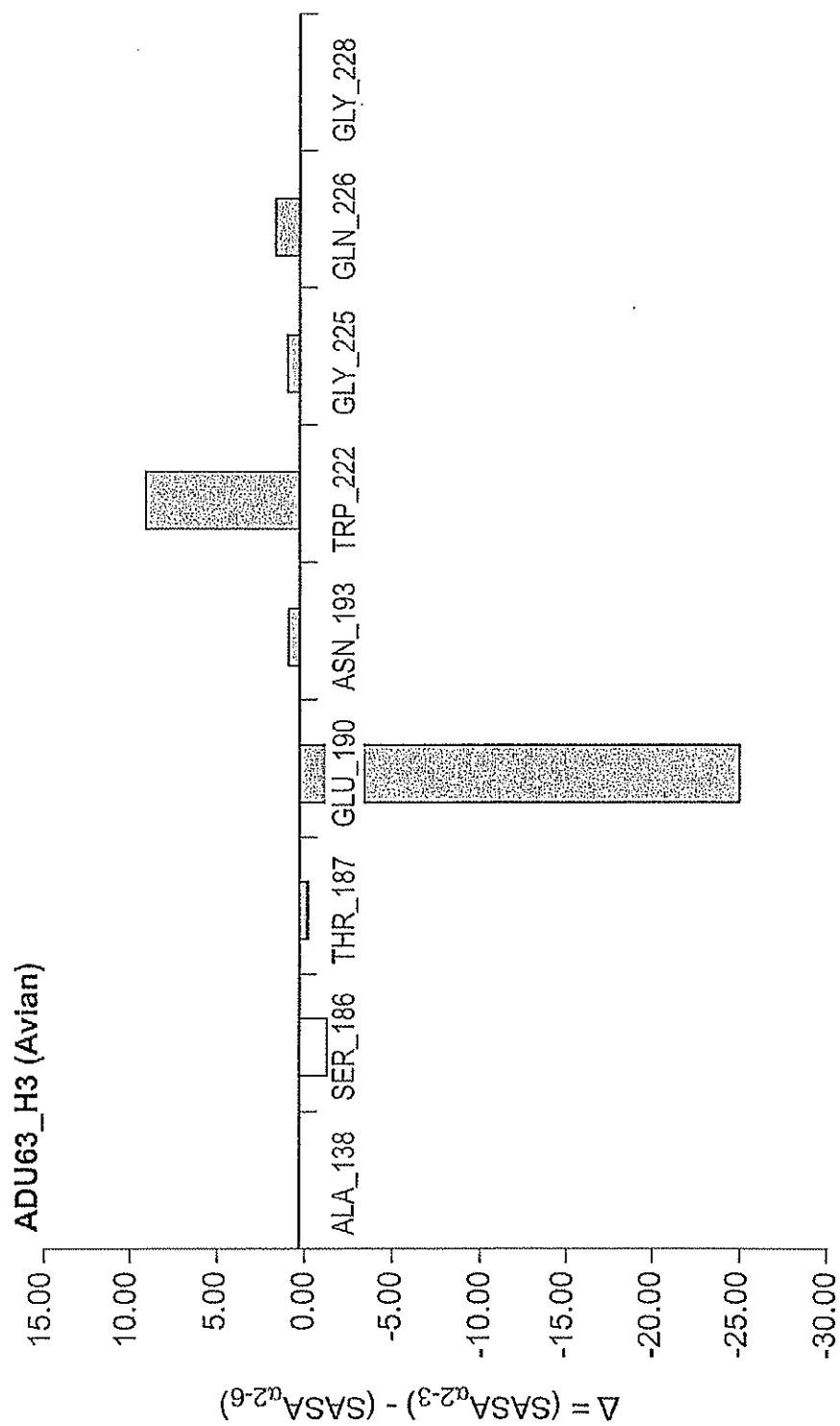


FIG. 15C

39/44

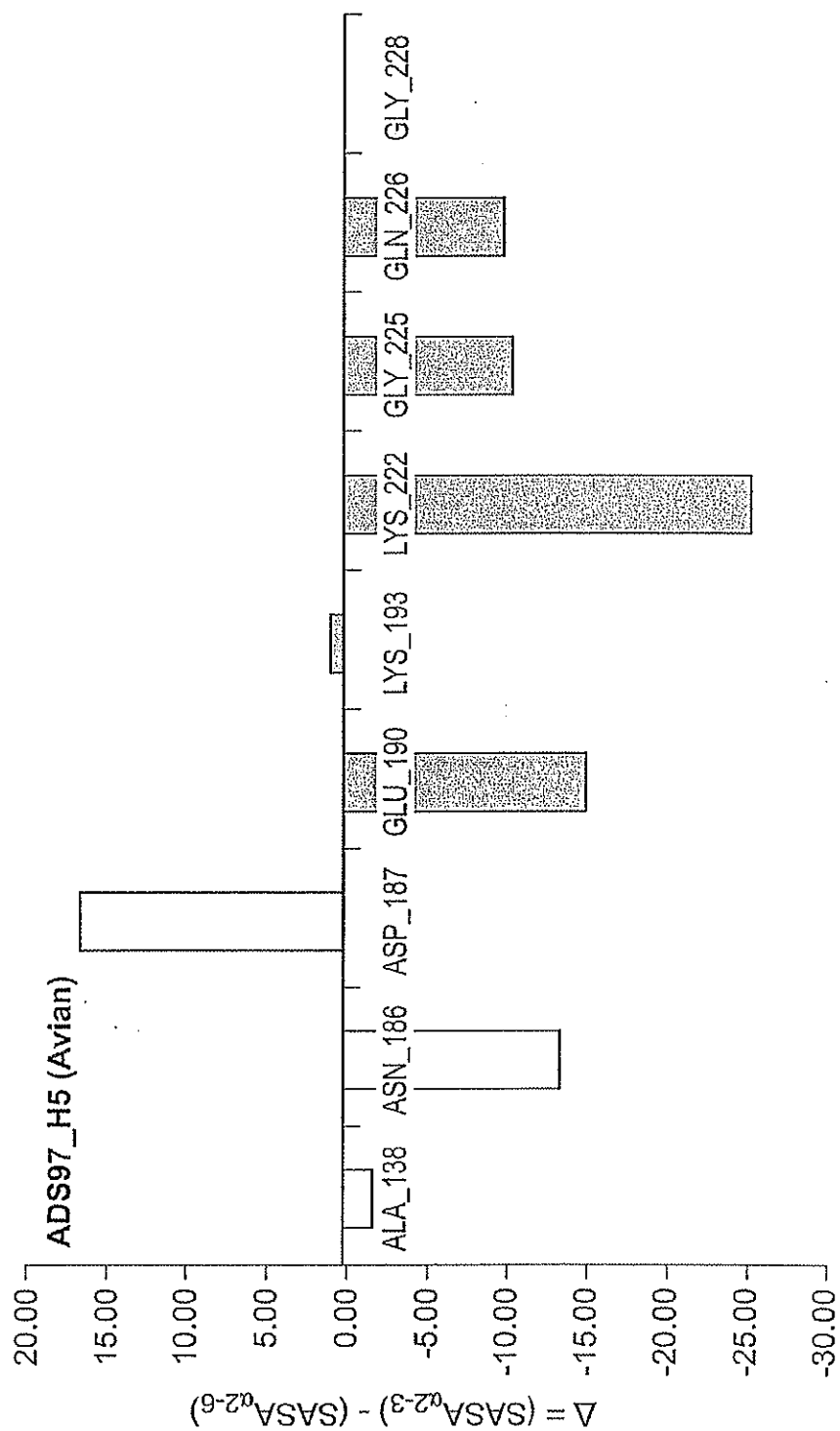


FIG. 15D

40/44

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

FIG. 15E

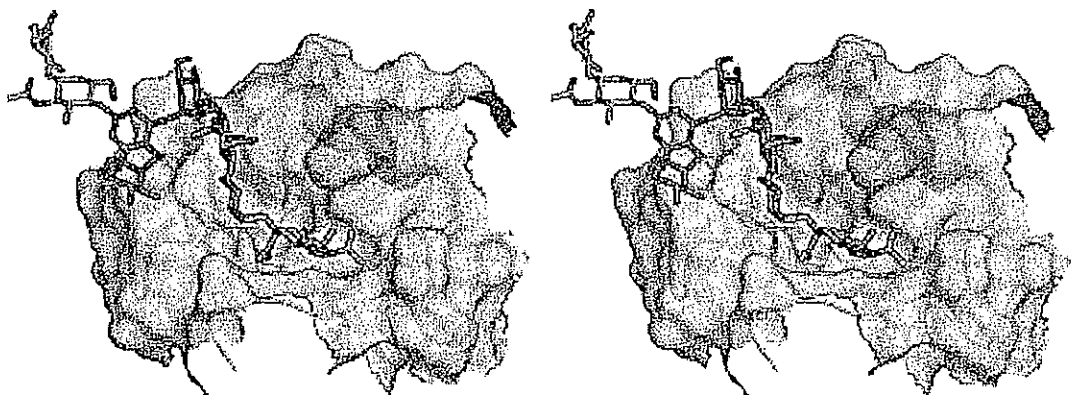


FIG. 16

41/44

Generation of HA protein.

H1N1 SC 1918
H3N2 Moscow 1999
H5N1 Viet 2004

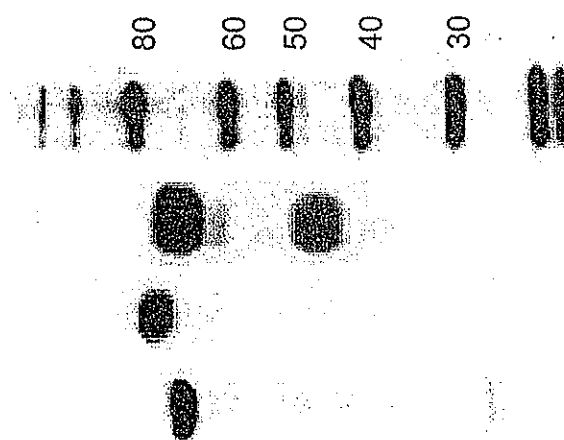


FIG. 17A

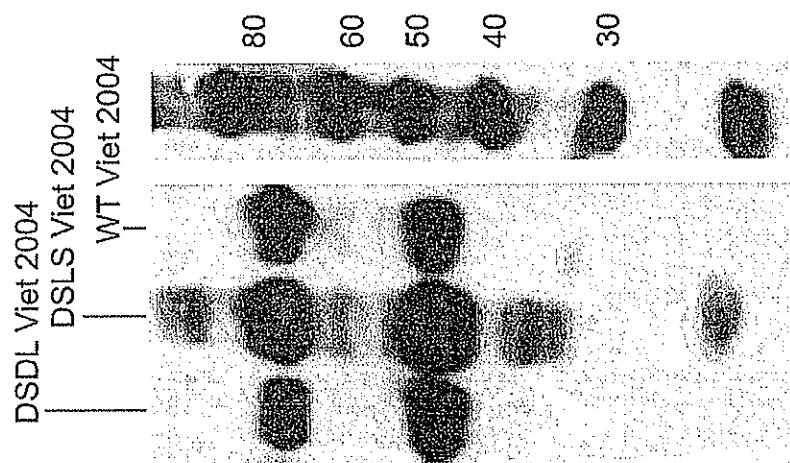


FIG. 17B

42/44

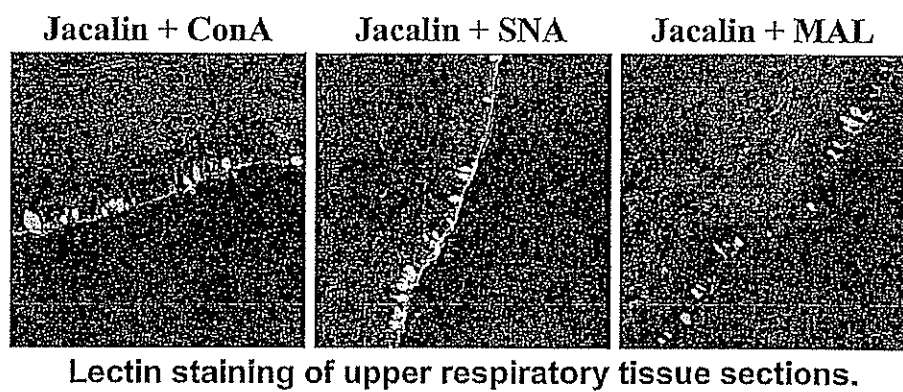


FIG. 18

43/44

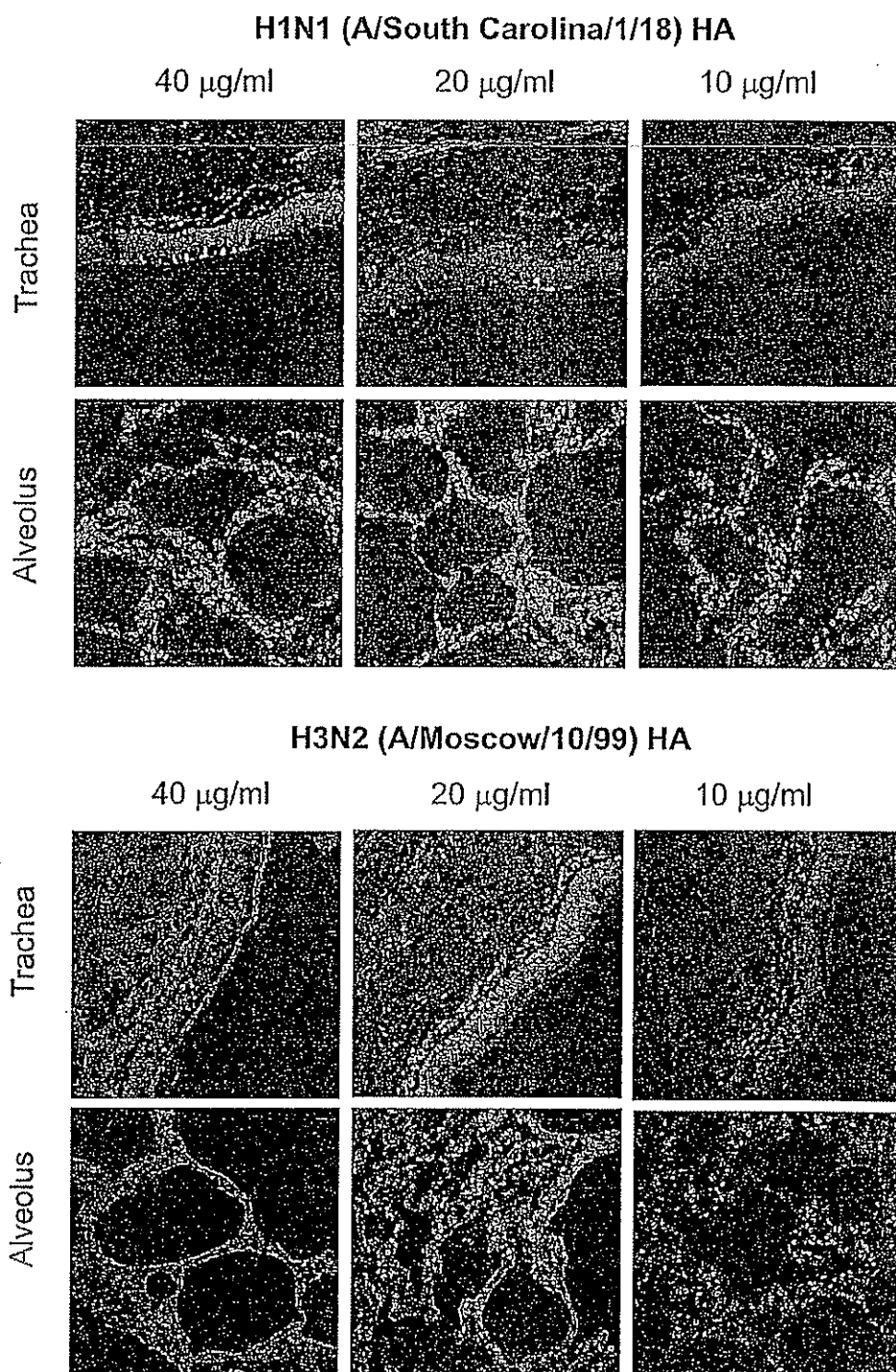


FIG. 19

44/44

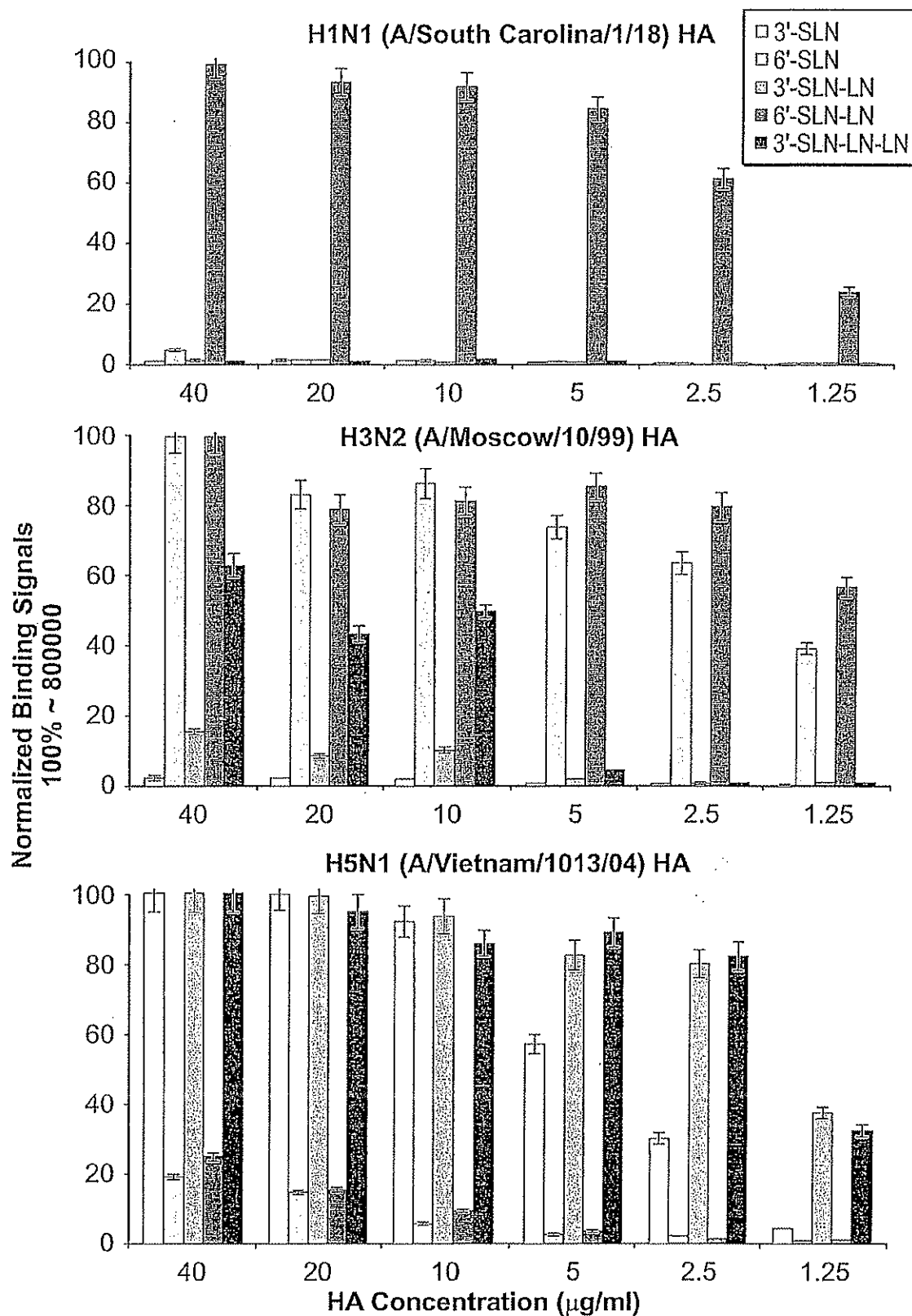


FIG. 20