US 20110236969A1

(19) **United States**

(12) **Patent Application Publication**

(10) **Pub. No.: US 2011/0236969 A1**

(43) **Pub. Date: Sep. 29, 2011**

Lazar et al.

(54) **METHODS OF GENERATING VARIANT PROTEINS WITH INCREASED HOST STRING CONTENT AND COMPOSITIONS THEREOF**

(75) Inventors: **Gregory Alan Lazar**, Alhambra, CA (US); **John R. Desjarlais**, Pasadena, CA (US); **Philip W. Hammond**, Sierra Madre, CA (US)

(73) Assignee: **Xencor, Inc.**, Monrovia, CA (US)

(21) Appl. No.: **13/053,087**

(22) Filed: **Mar. 21, 2011**

Related U.S. Application Data

(63) Continuation of application No. 11/981,794, filed on Oct. 31, 2007, now Pat. No. 7,930,107, which is a continuation of application No. 11/004,590, filed on Dec. 3, 2004, now Pat. No. 7,657,380.

(60) Provisional application No. 60/619,483, filed on Oct. 16, 2004, provisional application No. 60/601,665, filed on Aug. 13, 2004, provisional application No. 60/581,613, filed on Jun. 21, 2004, provisional application No. 60/527,167, filed on Dec. 4, 2003.

Publication Classification

(51) **Int. Cl.**

C12N 5/10	(2006.01)
C12N 15/63	(2006.01)
C07K 1/107	(2006.01)
C07K 16/28	(2006.01)
C07H 21/00	(2006.01)

(52) **U.S. Cl.** **435/334**; 435/320.1; 530/409; 530/389.1; 530/387.3; 536/23.53

(57) **ABSTRACT**

The present invention relates to novel methods for generating variant proteins with increased host string content, and proteins that are engineered using these methods.



FIG. 1a

Human Germline VH Sequences

	Chain	FR1	CDR1	FR2
		1 2	3	4
	Kabat	12345678901234567890123456789012345ab67890123456789		
SEQ ID NO.:1	VH_1-2	QVQLVQSGAEVKKPGASVKVSCKASGYTFTGYYMH		WVRQAPGQGLEWMG
SEQ ID NO.:2	VH_1-3	QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYAMH		WVRQAPGQRLEWMG
SEQ ID NO.:3	VH_1-8	QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYDIN		WVRQATGQGLEWMG
SEQ ID NO.:4	VH_1-18	QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYGIS		WVRQAPGQGLEWMG
SEQ ID NO.:5	VH_1-24	QVQLVQSGAEVKKPGASVKVSCKVSGYTLTELSMH		WVRQAPGKGLEWMG
SEQ ID NO.:6	VH_1-45	QMQLVQSGAEVKKTGSSVKVSCKASGYTFTYRYLH		WVRQAPGQALEWMG
SEQ ID NO.:7	VH_1-46	QVQLVQSGAEVKKPGASVKVSCKASGYTFTSYYMH		WVRQAPGQGLEWMG
SEQ ID NO.:8	VH_1-58	QMQLVQSGPEVKKPGTSVKVSCKASGFTFTSSAVQ		WVRQARGQRLEWIG
SEQ ID NO.:9	VH_1-69	QVQLVQSGAEVKKPGSSVKVSCKASGGTFSSYAIS		WVRQAPGQGLEWMG
SEQ ID NO.:10	VH_1-f	EVQLVQSGAEVKKPGATVKISCKVSGYTFTDYYMH		WVQQAPGKGLEWMG
SEQ ID NO.:11	VH_2-5	QITLKESGPTLVKPTQTLTLTCTFSGFSLSTSGVGVG		WIRQPPGKALEWLA
SEQ ID NO.:12	VH_2-26	QVTLKESGPVLVKPTETLTLTCTVSGFSLSNARMGVS		WIRQPPGKALEWLA
SEQ ID NO.:13	VH_2-70	QVTLRESGPALVKPTQTLTLTCTFSGFSLSTSGMCVS		WIRQPPGKALEWLA
SEQ ID NO.:14	VH_3-7	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYWMS		WVRQAPGKGLEWVA
SEQ ID NO.:15	VH_3-9	EVQLVESGGGLVQPGRSLRLSCAASGFTFDDYAMH		WVRQAPGKGLEWVS
SEQ ID NO.:16	VH_3-11	QVQLVESGGGLVKPGGSLRLSCAASGFTFSDYYMS		WIRQAPGKGLEWVS
SEQ ID NO.:17	VH_3-13	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYDMH		WVRQATGKGLEWVS
SEQ ID NO.:18	VH_3-15	EVQLVESGGGLVKPGGSLRLSCAASGFTFSNAWMS		WVRQAPGKGLEWVG
SEQ ID NO.:19	VH_3-16	EVQLVESGGGLVQPGGSLRLSCAASGFTFSNSDMN		WARKAPGKGLEWVS
SEQ ID NO.:20	VH_3-19	TVQLVESGGGLVEPGGSLRLSCAASGFTFSNSDMN		WVRQAPGKGLEWVS
SEQ ID NO.:21	VH_3-20	EVQLVESGGGVVRPGGSLRLSCAASGFTFDDYGMS		WVRQAPGKGLEWVS
SEQ ID NO.:22	VH_3-21	EVQLVESGGGLVKPGGSLRLSCAASGFTFSSYSMN		WVRQAPGKGLEWVS
SEQ ID NO.:23	VH_3-23	EVQLLESGGGLVQPGGSLRLSCAASGFTFSSYAMS		WVRQAPGKGLEWVS

FIG. 1a-1

```
   CDR2                          FR3                                CDR3
5              6          7          8              9          10
012abc34567890123456789012345678901234567890123456789012abc34567890123456789012abcdefghijk12

RINP  NSGGTNYAQKFQGRVTSTRDTSISTAYMELSRLRSDDTVVYYCAR
WINA  GNGNTKYSQKFQGRVTITRDTSASTAYMELSSLRSEDTAVYYCAR
WMNP  NSGNTGYAQKFQGRVTMTRNTSISTAYMELSSLRSEDTAVYYCAR
WISA  YNGNTNYAQKLQGRVTMTTDTSTSTAYMELRSLRSDDTAVYYCAR
GFDP  EDGETIYAQKFQGRVTMTEDTSTDTAYMELSSLRSEDTAVYYCAT
WITP  FNGNTNYAQKFQDRVTITRDRSMSTAYMELSSLRSEDTAMYYCAR
IINP  SGGSTSYAQKFQGRVTMTRDTSTSTVYMELSSLRSEDTAVYYCAR
WIVV  GSGNTNYAQKFQERVTITRDMSTSTAYMELSSLRSEDTAVYYCAA
GIIP  IFGTANYAQKFQGRVTITADESTSTAYMELSSLRSEDTAVYYCAR
LVDP  EDGETIYAEKFQGRVTITADTSTDTAYMELSSLRSEDTAVYYCAT

LIY   WNDDKRYSPSLKSRLTITKDTSKNQVVLTMTNMDPVDTATYYCAHR
HIF   SNDEKSYSTSLKSRLTISKDTSKSQVVLTMTNMDPVDTATYYCARI
LID   WDDDKYYSTSLKTRLTISKDTSKNQVVLTMTNMDPVDTATYYCARI

NIKQ  DGSEKYYVDSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCAR
GISW  NSGSIGYADSVKGRFTISRDNAKNSLYLQMNSLRAEDTALYYCAKD
YISS  SGSTIYYADSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCAR
AIGT  AGDT YYPGSVKGRFTISRENAKNSLYLQMNSLRAGDTAVYYCAR
RIKSKTDGGTTDYAAPVKGRFTISRDDSKNTLYLQMNSLKTEDTAVYYCTT
GVSW  NGSRTHYVDSVKRRFIISRDNSRNSLYLQKNRRRAEDMAVYYCVR
GVSW  NGSRTHYADSVKGRFIISRDNSRNFLYQQMNSLRPEDMAVYYCVR
GINW  NGGSTGYADSVKGRFTISRDNAKNSLYLQMNSLRAEDTALYHCAR
SISS  SSSYIYYADSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCAR
AISG  SGGSTYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAK
```

FIG. 1a-2

SEQ ID NO.:24	VH_3-30	QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYAMH WVRQAPGKGLEWVA
SEQ ID NO.:25	VH_3-30-3	QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYAMH WVRQAPGKGLEWVA
SEQ ID NO.:26	VH_3-33	QVQLVESGGGVVQPGRSLRLSCAASGFTFSSYGMH WVRQAPGKGLEWVA
SEQ ID NO.:27	VH_3-35	EVQLVESGGGLVQPGGSLRLSCAASGFTFSNSDMN WVHQAPGKGLEWVS
SEQ ID NO.:28	VH_3-38	EVQLVESGGGLVQPRGSLRLSCAASGFTVSSNEMS WIRQAPGKGLEWVS
SEQ ID NO.:29	VH_3-43	EVQLVESGGVVVQPGGSLRLSCAASGFTFDDYTMH WVRQAPGKGLEWVS
SEQ ID NO.:30	VH_3-48	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYSMN WVRQAPGKGLEWVS
SEQ ID NO.:31	VH_3-49	EVQLVESGGGLVQPGRSLRLSCTASGFTFGDYAMS WFRQAPGKGLEWVG
SEQ ID NO.:32	VH_3-53	EVQLVESGGGLIQPGGSLRLSCAASGFTVSSNYMS WVRQAPGKGLEWVS
SEQ ID NO.:33	VH_3-64	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYAMH WVRQAPGKGLEYVS
SEQ ID NO.:34	VH_3-66	EVQLVESGGGLVQPGGSLRLSCAASGFTVSSNYMS WVRQAPGKGLEWVS
SEQ ID NO.:35	VH_3-72	EVQLVESGGGLVQPGGSLRLSCAASGFTFSDHYMD WVRQAPGKGLEWVG
SEQ ID NO.:36	VH_3-73	EVQLVESGGGLVQPGGSLKLSCAASGFTFSGSAMH WVRQASGKGLEWVG
SEQ ID NO.:37	VH_3-74	EVQLVESGGGLVQPGGSLRLSCAASGFTFSSYWMH WVRQAPGKGLVWVS
SEQ ID NO.:38	VH_3-d	EVQLVESRGVLVQPGGSLRLSCAASGFTVSSNEMS WVRQAPGKGLEWVS
SEQ ID NO.:39	VH_4-4	QVQLQESGPGLVKPPGTLSLTCAVSGGSISSSNWWS WVRQPPGKGLEWIG
SEQ ID NO.:40	VH_4-28	QVQLQESGPGLVKPSDTLSLTCAVSGYSISSSNWWG WIRQPPGKGLEWIG
SEQ ID NO.:41	VH_4-30-2	QLQLQESGSGLVKPSQTLSLTCAVSGGSISSGGYSWSWIRQPPGKGLEWIG
SEQ ID NO.:42	VH_4-30-4	QVQLQESGPGLVKPSQTLSLTCTVSGGSISSGDYYWSWIRQPPGKGLEWIG
SEQ ID NO.:43	VH_4-31	QVQLQESGPGLVKPSQTLSLTCTVSGGSISSGGYYWSWIRQHPGKGLEWIG
SEQ ID NO.:44	VH_4-34	QVQLQQWGAGLLKPSETLSLTCAVYGGSFSGYYWS WIRQPPGKGLEWIG
SEQ ID NO.:45	VH_4-39	QLQLQESGPGLVKPSETLSLTCTVSGGSISSSSYYWGWIRQPPGKGLEWIG
SEQ ID NO.:46	VH_4-59	QVQLQESGPGLVKPSETLSLTCTVSGGSISSYYWS WIRQPPGKGLEWIG
SEQ ID NO.:47	VH_4-61	QVQLQESGPGLVKPSETLSLTCTVSGGSVSSGSYYWSWIRQPPGKGLEWIG
SEQ ID NO.:48	VH_4-b	QVQLQESGPGLVKPSETLSLTCAVSGYSISSGYYWG WIRQPPGKGLEWIG
SEQ ID NO.:49	VH_5-51	EVQLVQSGAEVKKPGESLKISCKGSGYSFTSYWIG WVRQMPGKGLEWMG
SEQ ID NO.:50	VH_5-a	EVQLVQSGAEVKKPGESLRISCKGSGYSFTSYWIS WVRQMPGKGLEWMG
SEQ ID NO.:51	VH_6-1	QVQLQQSGPGLVKPSQTLSLTCAISGDSVSSNSAAWNWIRQSPSRGLEWLG
SEQ ID NO.:52	VH_7-4-1	QVQLVQSGSELKKPGASVKVSCKASGYTFTSYAMN WVRQAPGQGLEWMG
SEQ ID NO.:53	VH_7-81	QVQLVQSGHEVKQPGASVKVSCKASGYSFTTYGMN WVFQAPGQGLEWMG

FIG. 1a-3

VISY DGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR
VISY DGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR
VIWY DGSNKYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR
GVSW NGSRTHYADSVKGRFIISRDNSRNTLYLQTNSLRAEDTAVYYCVR
SIS GGSTYYADSRKGRFTISRDNSKNTLYLQMNNLRAEGTAAYYCARY
LISW DGGSTYYADSVKGRFTISRDNSKNSLYLQMNSLRTEDTALYYCAKD
YISS SSSTIYYADSVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCAR
FIRSKAYGGTTEYTASVKGRFTISRDGSKSIAYLQMNSLKTEDTAVYYCTR
VIY SGGSTYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR
AISS NGGSTYYANSVKGRFTISRDNSKNTLYLQMGSLRAEDMAVYYCAR
VIY SGGSTYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAR
RTRNKANSYTTEYAASVKGRFTISRDDSKNSLYLQMNSLKTEDTAVYYCAR
RIRSKANSYATAYAASVKGRFTISRDDSKNTAYLQMNSLKTEDTAVYYCTR
RINS DGSSTSYADSVKGRFTISRDNAKNTLYLQMNSLRAEDTAVYYCAR
SIS GGSTYYADSRKGRFTISRDNSKNTLHLQMNSLRAEDTAVYYCKK
EIY HSGSTNYNPSLKSRVTISVDKSKNQFSLKLSSVTAADTAVYCCAR
YIY YSGSTYYNPSLKSRVTMSVDTSKNQFSLKLSSVTAVDTAVYYCAR
YIY HSGSTYYNPSLKSRVTISVDRSKNQFSLKLSSVTAADTAVYYCAR
YIY YSGSTYYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
YIY YSGSTYYNPSLKSLVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
EIN HSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
SIY YSGSTYYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
YIY YSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
YIY YSGSTNYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR
SIY HSGSTYYNPSLKSRVTISVDTSKNQFSLKLSSVTAADTAVYYCAR

IIYP GDSDTRYSPSFQGQVTISADKSISTAYLQWSSLKASDTAMYYCAR
RIDP SDSYTNYSPSFQGHVTISADKSISTAYLQWSSLKASDTAMYYCAR

RTYYR SKWYNDYAVSVKSRITINPDTSKNQFSLQLNSVTPEDTAVYYCAR

WINT NTGNPTYAQGFTGRFVFSLDTSVSTAYLQICSLKAEDTAVYYCAR
WFNT YTGNPTYAQGFTGRFVFSMDTSASTAYLQISSLKAEDMAMYYCAR

FIG. 1a-4

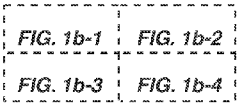


FIG. 1b

FIG. 1b-1

Human Germline VLk Sequences

	Chain	FR1		CDR1		FR2
		1	2	3		4
	Kabat	12345678901234567890123456789012345678901234567890123456789				
SEQ ID NO.:54	VLk_1-5	DIQMTQSPSTLSASVGDRVTITCRASQ		SISSWLAWYQQKPGKAPKLLIY		
SEQ ID NO.:55	VLk_1-6	AIQMTQSPSSLSASVGDRVTITCRASQ		GIRNDLGWYQQKPGKAPKLLIY		
SEQ ID NO.:56	VLk_1-8	AIRMTQSPSSFSASTGDRVTITCRASQ		GISSYLAWYQQKPGKAPKLLIY		
SEQ ID NO.:57	VLk_1D-8	VIWMTQSPSLLSASTGDRVTISCRMSQ		GISSYLAWYQQKPGKAPELLIY		
SEQ ID NO.:58	VLk_1-9	DIQLTQSPSFLSASVGDRVTITCRASQ		GISSYLAWYQQKPGKAPKLLIY		
SEQ ID NO.:59	VLk_1-12	DTQMTQSPSSVSASVGDRVTITCRASQ		GISSWLAWYQQKPGKAPKLLIY		
SEQ ID NO.:60	VLk_1D-12	DIQMTQSPSSVSASVGDRVTITCRASQ		GISSWLAWYQQKPGKAPKLLIY		
SEQ ID NO.:61	VLk_1-13_02	AIQLTQSPSSLSASVGDRVTITCRASQ		GISSALAWYQQKPGKAPKLLIY		
SEQ ID NO.:62	VLk_1D-13	AIQLTQSPSSLSASVGDRVTITCRASQ		GISSALAWYQQKPGKAPKLLIY		
SEQ ID NO.:63	VLk_1-16	DIQMTQSPSSLSASVGDRVTITCRASQ		GISNYLAWFQQKPGKAPKSLIY		
SEQ ID NO.:64	VLk_1D-16	DIQMTQSPSSLSASVGDRVTITCRASQ		GISSWLAWYQQKPEKAPKSLIY		
SEQ ID NO.:65	VLk_1-17	DIQMTQSPSSLSASVGDRVTITCRASQ		GIRNDLGWYQQKPGKAPKRLIY		
SEQ ID NO.:66	VLk_1D-17	NIQMTQSPSAMSASVGDRVTITCRARQ		GISNYLAWFQQKPGKVPKHLIY		
SEQ ID NO.:67	VLk_1-27	DIQMTQSPSSLSASVGDRVTITCRASQ		GISNYLAWYQQKPGKVPKLLIY		
SEQ ID NO.:68	VLk_1-33	DIQMTQSPSSLSASVGDRVTITCQASQ		DISNYLNWYQQKPGKAPKLLIY		
SEQ ID NO.:69	VLk_1D-33	DIQMTQSPSSLSASVGDRVTITCQASQ		DISNYLNWYQQKPGKAPKLLIY		
SEQ ID NO.:70	VLk_1-37	DIQLTQSPSSLSASVGDRVTITCRVSQ		GISSYLNWYRQKPGKVPKLLIY		
SEQ ID NO.:71	VLk_1D-37	DIQLTQSPSSLSASVGDRVTITCRVSQ		GISSYLNWYRQKPGKVPKLLIY		
SEQ ID NO.:72	VLk_1-39	DIQMTQSPSSLSASVGDRVTITCRASQ		SISSYLNWYQQKPGKAPKLLIY		
SEQ ID NO.:73	VLk_1D-39	DIQMTQSPSSLSASVGDRVTITCRASQ		SISSYLNWYQQKPGKAPKLLIY		
SEQ ID NO.:74	VLk_1D-42	DIQMIQSPSFLSASVGDRVSIICWASE		GISSNLAWYLQKPGKSPKLFLY		
SEQ ID NO.:75	VLk_1D-43	AIRMTQSPFSLSASVGDRVTITCWASQ		GISSYLAWYQQKPAKAPKLFIY		

FIG. 1b-2

```
 CDR2                FR3                          CDR3
5          6          7          8          9
01234567890123456789012345678901234567890123456789012345abcdef67

DASSLESGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCQQYNSYS
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCLQDYNYP
AASTLQSGVPSRFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSYP
AASTLQSGVPSRFSGSGSGTDFTLTISCLQSEDFATYYCQQYYSFP
AASTLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCQQLNSYP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQANSFP
DASSLESGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQFNSYP
DASSLESGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQFNNYP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQYNSYP
AASSLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP
AASSLQSGVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSYP
AASTLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYCQKYNSAP
DASNLETGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCQQYDNLP
DASNLETGVPSRFSGSGSGTDFTFTISSLQPEDIATYYCQQYDNLP
SASNLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYGQRTYNAP
SASNLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYGQRTYNAP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSYSTP
AASSLQSGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCQQSYSTP
DAKDLHPGVSSRFSGRGSGTDFTLTIISLKPEDFAAYYCKQDFSYP
YASSLQSGVPSRFSGSGSGTDYTLTISSLQPEDFATYYCQQYYSTP
```

SEQ ID NO.:76 VLk_2-24 DIVMTQTPLSSPVTLGQPASISCRSSQSLVHS DGNTYLSWLQQRPGQPPRLLIY
SEQ ID NO.:77 VLk_2D-24 DIVMTQTPLSSPVTLGQPASISFRSSQSLVHS DGNTYLSWLQQRPGQPPRLLIY
SEQ ID NO.:78 VLk_2-28 DIVMTQSPLSLPVTPGEPASISCRSSQSLLHS NGYNYLDWYLQKPGQSPQLLIY
SEQ ID NO.:79 VLk_2D-28 DIVMTQSPLSLPVTPGEPASISCRSSQSLLHS NGYNYLDWYLQKPGQSPQLLIY
SEQ ID NO.:80 VLk_2D-29 DIVMTQTPLSLSVTPGQPASISCKSSQSLLHS DGKTYLYWYLQKPGQPPQLLIY
SEQ ID NO.:81 VLk_2-29_2 DIVMTQTPLSLSVTPGQPASISCKSSQSLLHS DGKTYLYWYLQKPGQSPQLLIY
SEQ ID NO.:82 VLk_2-30 DVVMTQSPLSLPVTLGQPASISCRSSQSLVYS DGNTYLNWFQQRPGQSPRRLIY
SEQ ID NO.:83 VLk_2D-30 DVVMTQSPLSLPVTLGQPASISCRSSQSLVYS DGNTYLNWFQQRPGQSPRRLIY
SEQ ID NO.:84 VLk_2-40 DIVMTQTPLSLPVTPGEPASISCRSSQSLLDSDDGNTYLDWYLQKPGQSPQLLIY
SEQ ID NO.:85 VLk_2D-40 DIVMTQTPLSLPVTPGEPASISCRSSQSLLDSDDGNTYLDWYLQKPGQSPQLLIY
SEQ ID NO.:86 VLk_3-7 EIVMTQSPPTLSLSPGERVTLSCRASQS VSSSYLTWYQQKPGQAPRLLIY
SEQ ID NO.:87 VLk_3D-7 EIVMTQSPATLSLSPGERATLSCRASQS VSSSYLSWYQQKPGQAPRLLIY
SEQ ID NO.:88 VLk_3-11 EIVLTQSPATLSLSPGERATLSCRASQ SVSSYLAWYQQKPGQAPRLLIY
SEQ ID NO.:89 VLk_3D-11 EIVLTQSPATLSLSPGERATLSCRASQ GVSSYLAWYQQKPGQAPRLLIY
SEQ ID NO.:90 VLk_3-15 EIVMTQSPATLSVSPGERATLSCRASQ SVSSNLAWYQQKPGQAPRLLIY
SEQ ID NO.:91 VLk_3D-15 EIVMTQSPATLSVSPGERATLSCRASQ SVSSNLAWYQQKPGQAPRLLIY
SEQ ID NO.:92 VLk_3-20 EIVLTQSPGTLSLSPGERATLSCRASQS VSSSYLAWYQQKPGQAPRLLIY
SEQ ID NO.:93 VLk_3D-20 EIVLTQSPATLSLSPGERATLSCGASQS VSSSYLAWYQQKPGLAPRLLIY

SEQ ID NO.:94 VLk_4-1 DIVMTQSPDSLAVSLGERATINCKSSQSVLYSSNNKNYLAWYQQKPGQPPKLLIY

SEQ ID NO.:95 VLk_5-2 ETTLTQSPAFMSATPGDKVNISCKASQ DIDDDMNWYQQKPGEAAIFIIQ

SEQ ID NO.:96 VLk_6-21 EIVLTQSPDFQSVTPKEKVTITCRASQ SIGSSLHWYQQKPDQSPKLLIK
SEQ ID NO.:97 VLk_6D-21 EIVLTQSPDFQSVTPKEKVTITCRASQ SIGSSLHWYQQKPDQSPKLLIK
SEQ ID NO.:98 VLk_6D-41 DVVMTQSPAFLSVTPGEKVTITCQASE GIGNYLYWYQQKPDQAPKLLIK

FIG. 1b-3

KISNRFSGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCMQATQFP
KVSNRFSGVPDRFSGSGAGTDFTLKISRVEAEDVGVYYCTQATQFP
LGSNRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQALQTP
LGSNRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQALQTP
EVSNRFSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQSIQLP
EVSSRFSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGIHLP
KVSNRDSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP
KVSNWDSGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQGTHWP
TLSYRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQRIEFP
TLSYRASGVPDRFSGSGSGTDFTLKISRVEAEDVGVYYCMQRIEFP
GASTRATSIPARFSGSGSGTDFTLTISSLQPEDFAVYYCQQDHNLP
GASTRATGIPARFSGSGSGTDFTLTISSLQPEDFAVYYCQQDYNLP
DASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYYCQQRSNWP
DASNRATGIPARFSGSGPGTDFTLTISSLEPEDFAVYYCQQRSNWH
GASTRATGIPARFSGSGSGTEFTLTISSLQSEDFAVYYCQQYNNWP
GASTRATGIPARFSGSGSGTEFTLTISSLQSEDFAVYYCQQYNNWP
GASSRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP
DASSRATGIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSP

WASTRESGVPDRFSGSGSGTDFTLTISSLQAEDVAVYYCQQYYSTP

EATTLVPGIPPRFSGSGYGTDFTLTINNIESEDAAYYFCLQHDNFP

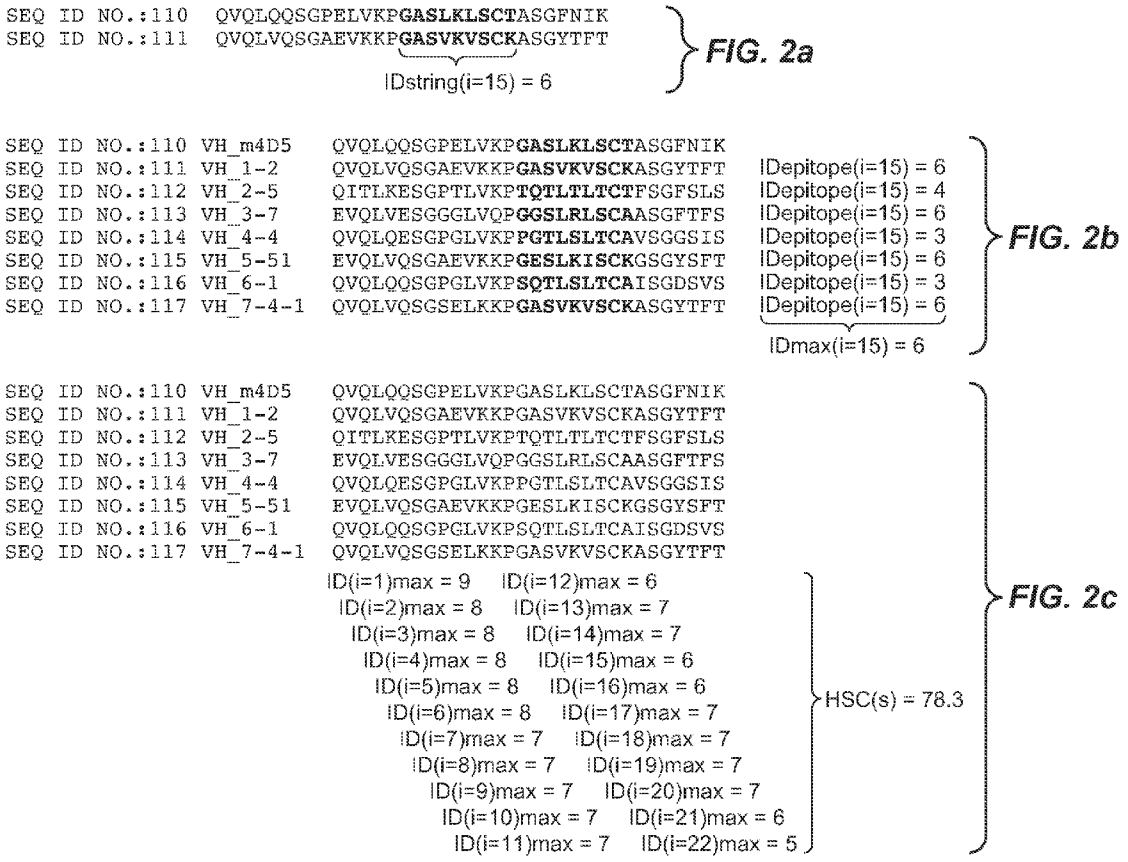
YASQSFSGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCHQSSSLP
YASQSFSGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCHQSSSLP
YASQSISGVPSRFSGSGSGTDFTFTISSLEAEDAATYYCQQGNKHP

FIG. 1b-4

		Kabat	9 10 678901234567
SEQ ID NO.:99	IGKJ1		WTFGQGTKVEIK
SEQ ID NO.:100	IGKJ2		YTFGQGTKLEIK
SEQ ID NO.:101	IGKJ3		FTFGPGTKVDIK
SEQ ID NO.:102	IGKJ4		LTFGGGTKVEIK
SEQ ID NO.:103	IGKJ5		ITFGQGTRLEIK

		Kabat	10 11 1234567890123
SEQ ID NO.:104	IGHJ1		AEYFQHWGQGTLVTVSS
SEQ ID NO.:105	IGHJ2		YWYFDLWGRGTLVTVSS
SEQ ID NO.:106	IGHJ3		AFDVWGQGTMVTVSS
SEQ ID NO.:107	IGHJ4		YFDYWGQGTLVTVSS
SEQ ID NO.:108	IGHJ5		NWFDSWGQGTLVTVSS
SEQ ID NO.:109	IGHJ6		YYYYYGMDVWGQGTTVTVSS

FIG. 1c


SEQ ID NO.:110 QVQLQQSGPELVKPGASLKLSCTASGFNIK
SEQ ID NO.:111 QVQLVQSGAEVKKPGASVKVSCKASGYTFT
IDstring(i=15) = 6
FIG. 2a
SEQ ID NO.:110 VH_m4D5 QVQLQQSGPELVKPGASLKLSCTASGFNIK
SEQ ID NO.:111 VH_1-2 QVQLVQSGAEVKKPGASVKVSCKASGYTFT IDepitope(i=15) = 6
SEQ ID NO.:112 VH_2-5 QITLKESGPTLVKPTQTLTLTCTFSGFSLS IDepitope(i=15) = 4
SEQ ID NO.:113 VH_3-7 EVQLVESGGGLVQPGGSLRLSCAASGFTFS IDepitope(i=15) = 6
SEQ ID NO.:114 VH_4-4 QVQLQESGPGLVKPPGTLSLTCAVSGGSIS IDepitope(i=15) = 3
SEQ ID NO.:115 VH_5-51 EVQLVQSGAEVKKPGESLKISCKGSGYSFT IDepitope(i=15) = 6
SEQ ID NO.:116 VH_6-1 QVQLQQSGPGLVKPSQTLSLTCAISGDSVS IDepitope(i=15) = 3
SEQ ID NO.:117 VH_7-4-1 QVQLVQSGSELKKPGASVKVSCKASGYTFT IDepitope(i=15) = 6
IDmax(i=15) = 6
FIG. 2b
SEQ ID NO.:110 VH_m4D5 QVQLQQSGPELVKPGASLKLSCTASGFNIK
SEQ ID NO.:111 VH_1-2 QVQLVQSGAEVKKPGASVKVSCKASGYTFT
SEQ ID NO.:112 VH_2-5 QITLKESGPTLVKPTQTLTLTCTFSGFSLS
SEQ ID NO.:113 VH_3-7 EVQLVESGGGLVQPGGSLRLSCAASGFTFS
SEQ ID NO.:114 VH_4-4 QVQLQESGPGLVKPPGTLSLTCAVSGGSIS
SEQ ID NO.:115 VH_5-51 EVQLVQSGAEVKKPGESLKISCKGSGYSFT
SEQ ID NO.:116 VH_6-1 QVQLQQSGPGLVKPSQTLSLTCAISGDSVS
SEQ ID NO.:117 VH_7-4-1 QVQLVQSGSELKKPGASVKVSCKASGYTFT
ID(i=1)max = 9 ID(i=12)max = 6
ID(i=2)max = 8 ID(i=13)max = 7
ID(i=3)max = 8 ID(i=14)max = 7
ID(i=4)max = 8 ID(i=15)max = 6
ID(i=5)max = 8 ID(i=16)max = 6
ID(i=6)max = 8 ID(i=17)max = 7
ID(i=7)max = 7 ID(i=18)max = 7
ID(i=8)max = 7 ID(i=19)max = 7
ID(i=9)max = 7 ID(i=20)max = 7
ID(i=10)max = 7 ID(i=21)max = 6
ID(i=11)max = 7 ID(i=22)max = 5
HSC(s) = 78.3
FIG. 2c

SEQ ID NO.: 118
DIVLTQSPASLAVSLGQRATISCKASQSVDFDGDSYMNWYQQKPGQPPKVLIYAAS
NLESGIPARFSGSGSGTDFTLNIHPVEEEDAATYYCQQSNEDPWTFGGGTKLEIK

FIG. 3a

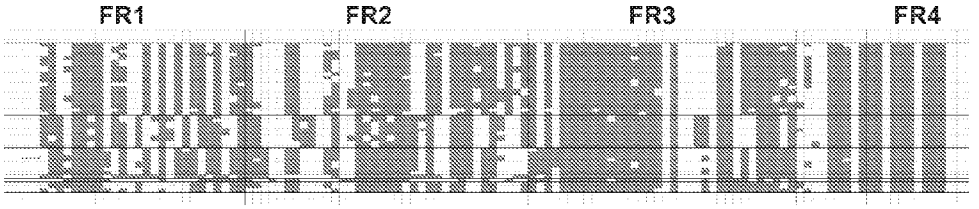


FIG. 3b

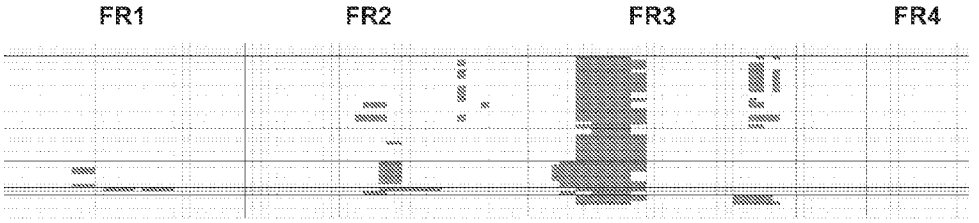


FIG. 3c

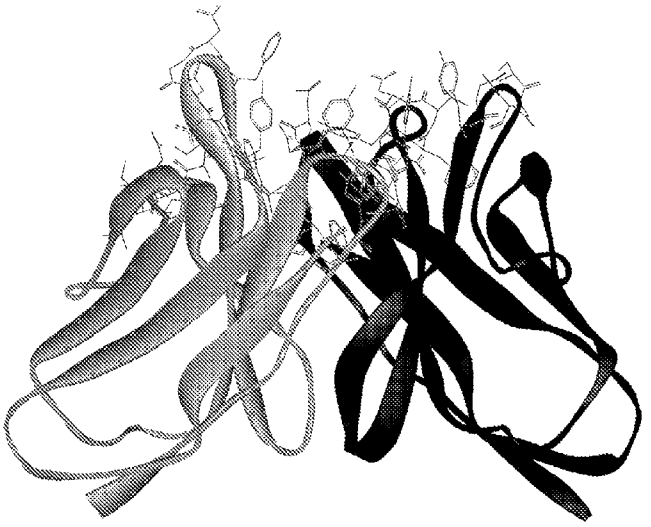


FIG. 3d

SEQ ID NO.: 119
QIQLQQSGPEVVKPGASVKISCKASGYTFTDYYITWVKQKPGQGLEWIGWIYPGSG
NTKYNEKFKGKATLTVDTSSSTAFMQLSSLTSEDTAVYFCANYGNYWFAYWGQGT
QVTVSA

FIG. 4a

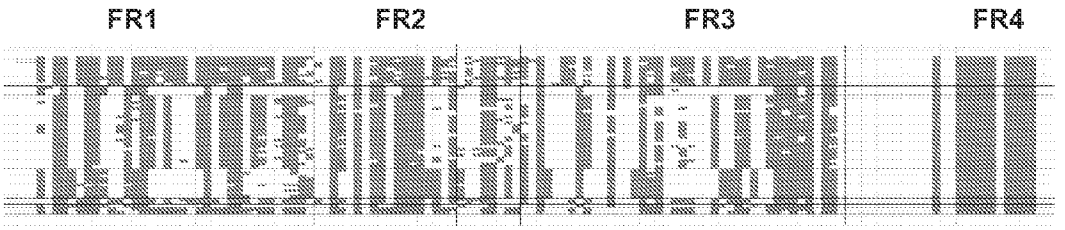


FIG. 4b

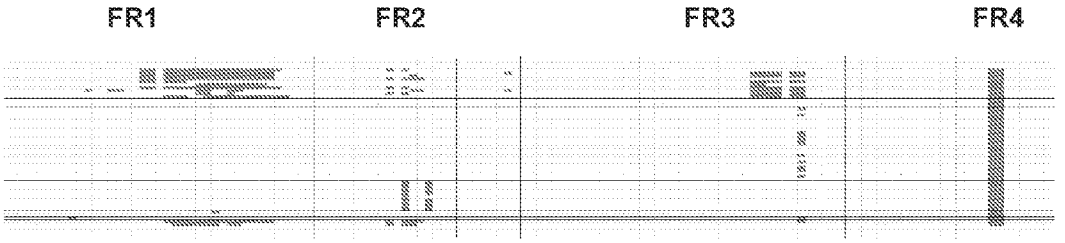


FIG. 4c



FIG. 4d

SEQ ID NO.: 120
DIVMTQSPDSLAVSLGERATINCKSSQSVDFDGDSYLAWYQQKPGQPPKLLIYAAS
NLESGVPDRFSGSGSGTDFTLTISSLQAEDVAVYYCQQSNEDPWTFGGGTKLEIK

FIG. 5a

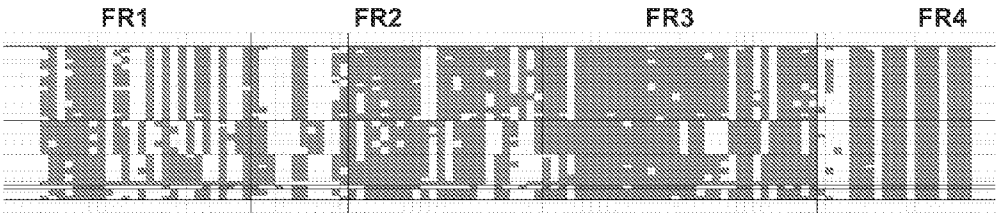


FIG. 5b

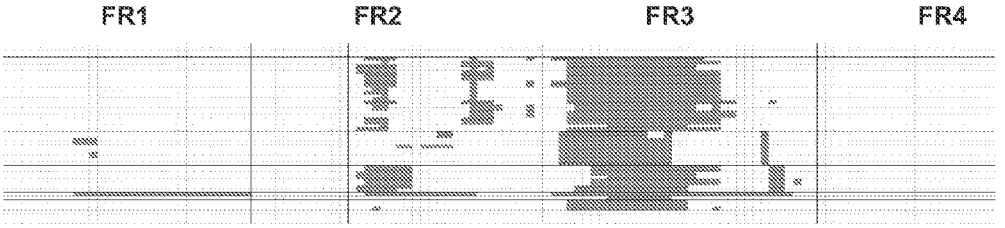


FIG. 5c



FIG. 5d

SEQ ID NO.: 121
QVQLVQSGAEVKKPGASVKVSCKASGYTFTDYYMHWVRQAPGQRLEWMGWIYP
GSGNTKYSQKFQGRVTITRDTSASTAYMELSSLRSEDTAVYYCARYGNYWFAYWG
QGTLVTVSS

FIG. 6a

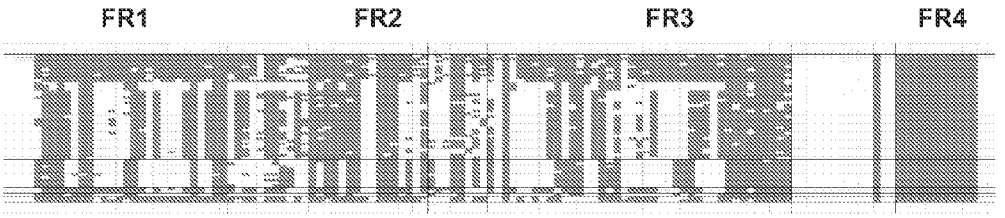


FIG. 6b

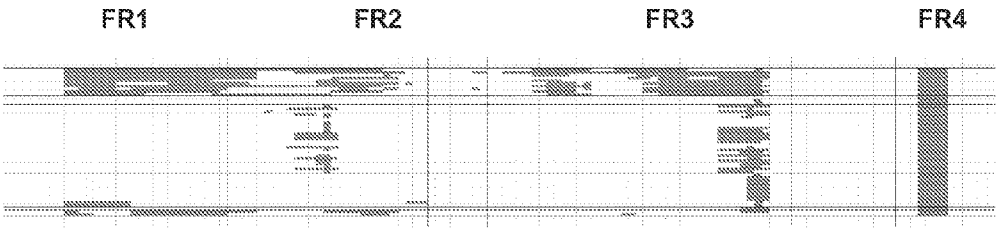


FIG. 6c



FIG. 6d

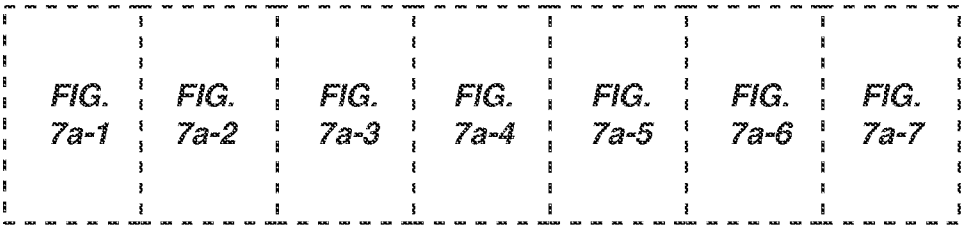

FIG. 7a-1
FIG. 7a-2
FIG. 7a-3
FIG. 7a-4
FIG. 7a-5
FIG. 7a-6
FIG. 7a-7

FIG. 7a

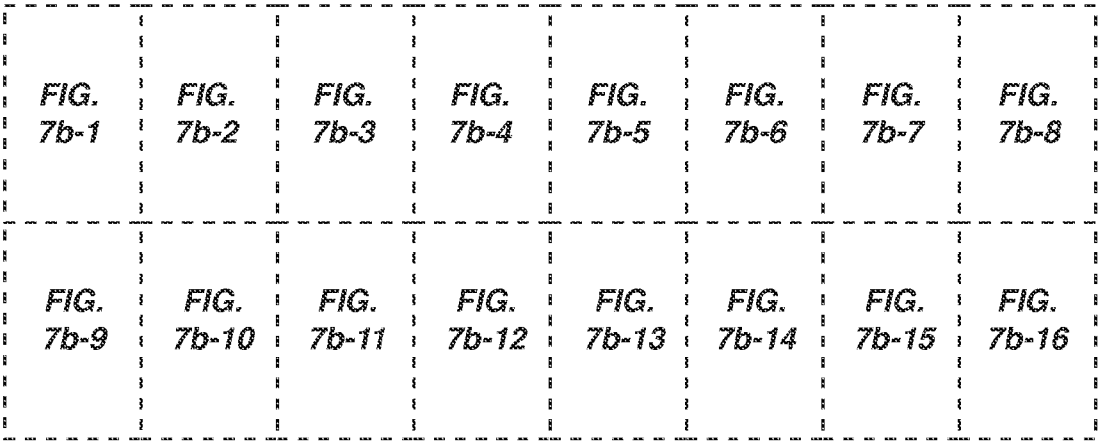

FIG. 7b-1
FIG. 7b-2
FIG. 7b-3
FIG. 7b-4
FIG. 7b-5
FIG. 7b-6
FIG. 7b-7
FIG. 7b-8
FIG. 7b-9
FIG. 7b-10
FIG. 7b-11
FIG. 7b-12
FIG. 7b-13
FIG. 7b-14
FIG. 7b-15
FIG. 7b-16

FIG. 7b

AC10 VL HEC Calculation 1

(WT) SEQ ID NO.:118	Iter	Structural Consensus	Structural Precedence	HSC	HSS	N_9max	FRH	Muts	Kabat WT Cluster
SEQ ID NO.:122	1	0.55	0.60	0.86	0.42	44	0.45	10	6.79
SEQ ID NO.:123	3	0.55	0.61	0.86	0.42	44	0.45	10	6.55
SEQ ID NO.:124	4	0.56	0.57	0.86	0.43	45	0.38	13	6.21
SEQ ID NO.:125	5	0.57	0.61	0.86	0.43	45	0.58	12	5.55
SEQ ID NO.:126	6	0.55	0.56	0.86	0.43	45	0.33	13	6.45
SEQ ID NO.:127	16	0.56	0.60	0.86	0.43	45	0.51	12	5.79
SEQ ID NO.:128	17	0.55	0.59	0.86	0.43	45	0.49	13	6.59
SEQ ID NO.:129	33	0.56	0.60	0.86	0.43	45	0.49	13	6.34
SEQ ID NO.:130	9	0.55	0.58	0.86	0.43	45	0.49	12	6.48
SEQ ID NO.:131	13	0.54	0.57	0.86	0.43	45	0.49	12	6.72
SEQ ID NO.:132	14	0.55	0.61	0.86	0.42	44	0.73	12	6.31
SEQ ID NO.:133	35	0.55	0.58	0.86	0.42	44	0.64	12	7.14
SEQ ID NO.:134	11	0.57	0.62	0.87	0.48	51	0.43	15	6.24
SEQ ID NO.:135	25	0.57	0.63	0.87	0.48	51	0.55	16	6.55
SEQ ID NO.:136	31	0.56	0.62	0.87	0.48	51	0.31	15	6.48
SEQ ID NO.:137	19	0.57	0.64	0.87	0.48	51	0.51	16	6.31
SEQ ID NO.:138	36	0.58	0.66	0.87	0.48	51	0.61	15	6.14
SEQ ID NO.:139	12	0.56	0.63	0.87	0.48	51	0.45	14	6.52
SEQ ID NO.:140	27	0.56	0.63	0.87	0.48	51	0.63	15	6.83
SEQ ID NO.:141	39	0.57	0.63	0.87	0.48	51	0.53	18	7.66
SEQ ID NO.:142	15	0.56	0.64	0.86	0.47	50	0.62	13	6.90
SEQ ID NO.:143	24	0.56	0.61	0.86	0.47	50	0.44	15	7.86
SEQ ID NO.:144	28	0.56	0.62	0.86	0.47	50	0.54	14	7.07
SEQ ID NO.:145	29	0.56	0.64	0.86	0.47	50	0.44	12	6.59
SEQ ID NO.:146	32	0.55	0.63	0.86	0.47	50	0.40	12	6.83
SEQ ID NO.:147	38	0.56	0.63	0.86	0.47	50	0.44	14	7.69
SEQ ID NO.:148	7	0.54	0.54	0.85	0.40	42	0.48	12	10.14
SEQ ID NO.:149	8	0.55	0.55	0.85	0.40	42	0.48	12	9.90
SEQ ID NO.:150	23	0.54	0.55	0.86	0.45	48	0.52	16	10.41
SEQ ID NO.:151	2	0.55	0.52	0.86	0.42	43	0.51	14	4.75
SEQ ID NO.:152	10	0.56	0.57	0.86	0.41	42	0.62	16	4.75
SEQ ID NO.:153	18	0.56	0.51	0.86	0.42	43	0.37	15	4.00
SEQ ID NO.:154	22	0.56	0.54	0.86	0.42	43	0.40	15	4.25
SEQ ID NO.:155	20	0.56	0.57	0.86	0.42	43	0.53	16	3.75
SEQ ID NO.:156	21	0.56	0.53	0.86	0.42	43	0.35	17	4.25
SEQ ID NO.:157	30	0.55	0.56	0.86	0.41	42	0.48	14	5.38
SEQ ID NO.:158	34	0.55	0.56	0.86	0.41	42	0.48	14	5.38
SEQ ID NO.:159	26	0.62	0.67	0.86	0.46	48	0.69	17	0.50
SEQ ID NO.:160	37	0.63	0.68	0.86	0.46	48	0.81	16	0.50

FIG. 7a-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D	I	V	L	T	Q	S	P	A	S	L	A	V	S	L	G	Q	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	V	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	V	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	D	S	L	A	V	S	L	G	E	R	A	T
D	I	V	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	V	L	T	Q	S	P	L	S	L	P	V	T	L	G	Q	P	A	S
D	I	V	L	T	Q	T	P	L	S	L	S	V	T	P	G	Q	P	A	S
D	I	V	L	T	Q	S	P	L	S	L	P	V	T	L	G	Q	P	A	S
D	I	V	L	T	Q	S	P	L	S	L	P	V	T	L	G	Q	P	A	S
D	I	V	L	T	Q	S	P	L	S	L	P	V	T	P	G	E	P	A	S
D	I	V	L	T	Q	T	P	L	S	L	S	V	T	P	G	Q	P	A	S
D	I	V	L	T	Q	S	P	L	S	L	P	V	T	P	G	E	P	A	S
D	I	V	L	T	Q	T	P	L	S	L	S	V	T	P	G	Q	P	A	S
D	I	V	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	V	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T

FIG. 7a-2

21	22	23	24	25	26	27	27a	27b	27c	27d			28	29	30	31	32	33	34
I	S	C	K	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	G	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
L	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	S	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	T	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N
I	T	C	R	A	S	Q	S	V	D	F	_	_	D	G	D	S	Y	M	N

FIG. 7a-3

35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
W	Y	Q	Q	K	P	G	Q	P	P	K	V	L	I	Y	A	A	S	N	L

55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	N
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	K
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	K
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	K
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	F	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T
E	S	G	I	P	A	R	F	S	G	S	G	S	G	T	D	F	T	L	T

FIG. 7a-5

75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
I	H	P	V	E	E	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	D	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	D	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	C	L	Q	S	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	C	L	Q	S	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	S	N	E	D
I	S	R	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	C	L	Q	S	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	C	L	Q	S	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	D	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	R	V	E	A	E	D	V	G	V	Y	Y	C	Q	Q	S	N	E	D
I	S	R	V	E	A	E	D	V	G	V	Y	Y	C	Q	Q	S	N	E	D
I	S	R	V	E	A	E	D	V	G	V	Y	Y	C	Q	Q	S	N	E	D
I	S	C	L	Q	S	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	D	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	S	N	E	D

FIG. 7a-6

	95	96	97	98	99	100	101	102	103	104	105	106	107
_ _	P	W	T	F	G	G	G	T	K	L	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K
_ _	P	W	T	F	G	G	G	T	K	X	E	I	K

FIG. 7a-7

FIG. 7b-1

AC10 VH HEC Calculation 1

(WT) SEQ ID NO.:119	Iter	Structural Consensus	Structural Precedence	HSC	HSS	$N_g$max	FRH	Muts	Kabat WT Cluster
SEQ ID NO.:161	1	0.47	0.62	0.81	0.42	49	0.88	18	5.41
SEQ ID NO.:162	17	0.48	0.62	0.81	0.42	49	0.76	17	5.23
SEQ ID NO.:163	42	0.47	0.62	0.81	0.42	49	0.71	18	5.74
SEQ ID NO.:164	2	0.46	0.60	0.81	0.42	49	0.65	19	6.38
SEQ ID NO.:165	6	0.47	0.62	0.81	0.42	50	0.58	18	5.33
SEQ ID NO.:166	14	0.47	0.62	0.81	0.42	50	0.58	17	5.15
SEQ ID NO.:167	7	0.47	0.64	0.81	0.43	50	0.80	18	4.92
SEQ ID NO.:168	15	0.48	0.64	0.81	0.43	50	0.76	17	4.74
SEQ ID NO.:169	33	0.48	0.64	0.81	0.43	50	0.72	18	5.26
SEQ ID NO.:170	10	0.48	0.64	0.81	0.42	49	0.76	17	5.33
SEQ ID NO.:171	12	0.47	0.61	0.81	0.42	49	0.63	19	6.77
SEQ ID NO.:172	36	0.47	0.61	0.81	0.42	49	0.63	18	5.92
SEQ ID NO.:173	16	0.47	0.62	0.81	0.43	50	0.54	18	5.72
SEQ ID NO.:174	20	0.45	0.61	0.81	0.42	49	0.51	20	7.41
SEQ ID NO.:175	23	0.46	0.59	0.81	0.42	49	0.55	19	6.67
SEQ ID NO.:176	37	0.46	0.60	0.81	0.42	49	0.59	18	5.82
SEQ ID NO.:177	26	0.46	0.62	0.81	0.42	49	0.61	19	6.49
SEQ ID NO.:178	28	0.46	0.61	0.81	0.42	49	0.59	18	6.46
SEQ ID NO.:179	41	0.46	0.61	0.81	0.42	49	0.55	19	7.31
SEQ ID NO.:180	30	0.46	0.62	0.81	0.42	49	0.63	18	5.92
SEQ ID NO.:181	40	0.47	0.62	0.81	0.42	49	0.63	17	5.74
SEQ ID NO.:182	32	0.46	0.62	0.81	0.42	49	0.80	20	7.00
SEQ ID NO.:183	50	0.47	0.63	0.81	0.43	50	0.62	19	5.62
SEQ ID NO.:184	51	0.48	0.63	0.81	0.43	50	0.62	18	5.44
SEQ ID NO.:185	35	0.46	0.63	0.81	0.43	50	0.62	21	7.67
SEQ ID NO.:186	39	0.46	0.61	0.81	0.42	49	0.69	20	7.79
SEQ ID NO.:187	44	0.45	0.62	0.81	0.43	50	0.56	21	7.95
SEQ ID NO.:188	49	0.46	0.62	0.81	0.42	49	0.73	20	6.92
SEQ ID NO.:189	53	0.45	0.62	0.81	0.42	50	0.50	20	7.38
SEQ ID NO.:190	4	0.48	0.61	0.81	0.46	53	0.66	21	7.95
SEQ ID NO.:191	13	0.47	0.59	0.81	0.46	53	0.66	21	8.69
SEQ ID NO.:192	22	0.47	0.59	0.81	0.46	53	0.75	21	8.36
SEQ ID NO.:193	25	0.48	0.61	0.81	0.46	53	0.70	20	7.44
SEQ ID NO.:194	38	0.48	0.61	0.81	0.46	53	0.75	21	7.62
SEQ ID NO.:195	45	0.47	0.58	0.81	0.45	53	0.55	20	8.59
SEQ ID NO.:196	48	0.47	0.59	0.81	0.46	53	0.58	22	9.05
SEQ ID NO.:197	52	0.48	0.61	0.81	0.46	53	0.58	21	8.13
SEQ ID NO.:198	43	0.46	0.59	0.81	0.46	53	0.74	23	10.41
SEQ ID NO.:199	11	0.48	0.58	0.81	0.38	44	0.75	20	9.13

FIG. 7b-2

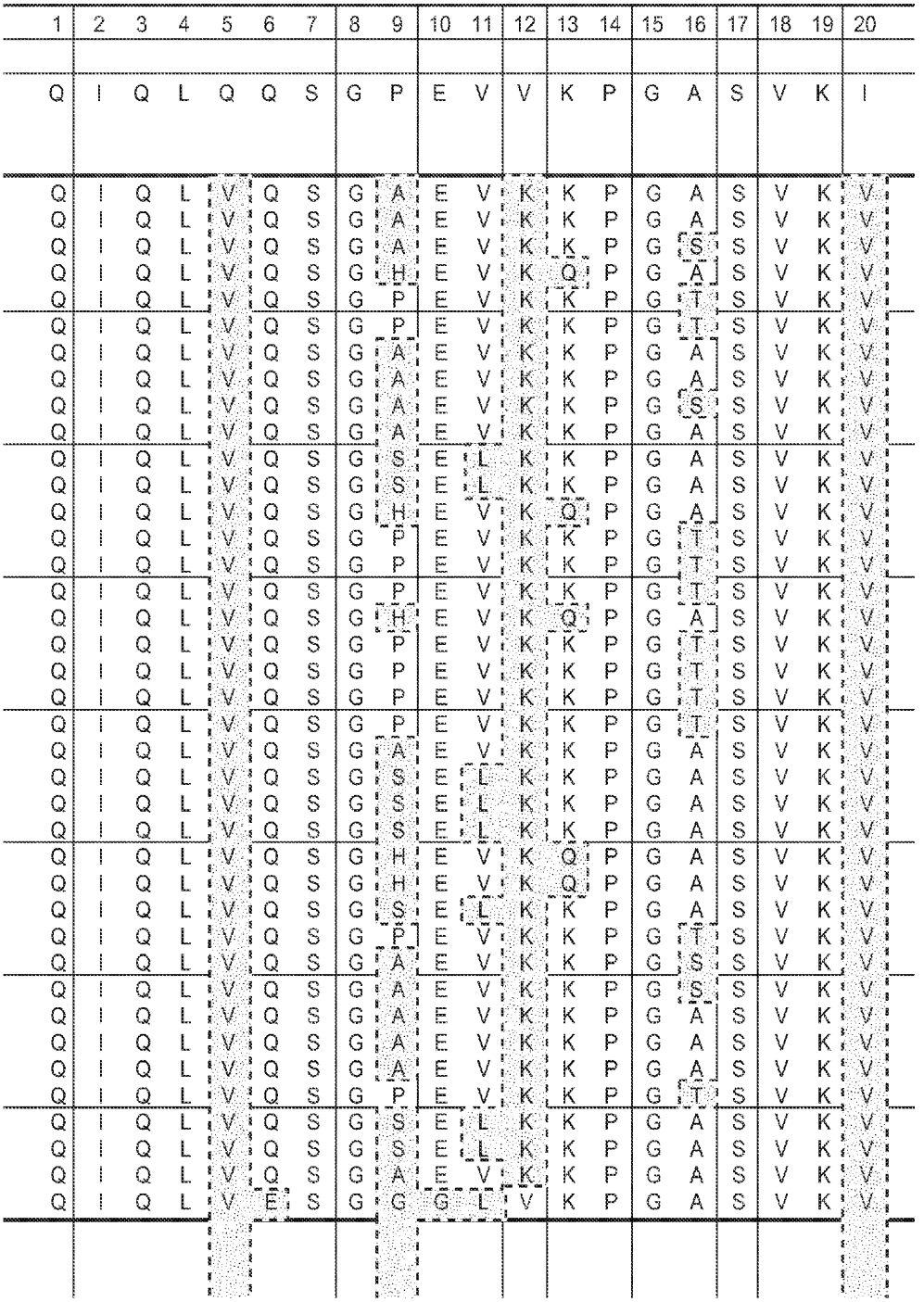

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Q I Q L Q Q S G P E V V K P G A S V K I

FIG. 7b-3

21	22	23	24	25	26	27	28	29	30	31			32	33	34	35	36	37	38
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	K
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D			Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R
S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R

FIG. 7b-4

39	40	41	42	43	44	45	46	47	48	49	50	51	52	52a	53			54	55	
Q	K	P	G	Q	G	L	E	W	I	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	P	G	Q	G	L	E	W	M	G	W	I	Y	P	G	_	_	S	G	
Q	A	T	G	Q	G	L	E	W	M	G	W	I	Y	P	G			S	G	

FIG. 7b-5

56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
N	T	K	Y	N	E	K	F	K	G	K	A	T	L	T	V	D	T	S	S
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	N	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	N	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	D	T	S	T
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	M	T	V	N	T	S	I
N	T	K	Y	N	E	K	F	Q	G	R	V	T	I	T	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	R	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	L	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	L	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	R	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	R	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	L	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	L	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	R	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	S	L	K	S	L	V	T	I	S	V	D	T	S	A
N	T	K	Y	N	E	K	F	Q	G	R	V	T	S	T	V	D	T	S	I

FIG. 7b-6

76	77	78	79	80	81	82	82a	82b	82c	83	84	85	86	87	88	89	90	91	92
S	T	A	F	M	Q	L	S	S	L	T	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	M	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	R	S	L	R	S	D	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	M	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	M	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	R	S	L	R	S	D	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	R	L	R	S	D	D	T	V	V	Y	F	C
S	T	A	Y	M	E	L	R	S	L	R	S	D	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	R	L	R	S	D	D	T	V	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	R	L	R	S	D	D	T	V	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C
S	T	A	Y	M	E	L	S	R	L	R	S	D	D	T	V	V	Y	F	C
S	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	F	C

FIG. 7b-7

93	94	95	96	97							98	99	100	101	102	103	104	105	106	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N							Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G	

FIG. 7b-8

	107	108	109	110	111	112	113
	T	Q	V	T	V	S	A
	T	L	V	T	V	S	S
	T	L	V	T	V	S	S
	T	L	V	T	V	S	S

SEQ ID NO.:200	3	0.46	0.57	0.81	0.40	45	0.53	22	5.18
SEQ ID NO.:201	5	0.46	0.58	0.81	0.40	45	0.67	21	4.09
SEQ ID NO.:202	56	0.46	0.56	0.81	0.39	45	0.47	21	3.91
SEQ ID NO.:203	8	0.47	0.59	0.81	0.42	49	0.71	22	4.73
SEQ ID NO.:204	18	0.48	0.60	0.81	0.42	49	0.65	20	3.73
SEQ ID NO.:205	29	0.47	0.59	0.81	0.42	49	0.55	21	3.91
SEQ ID NO.:206	19	0.47	0.58	0.81	0.42	49	0.55	20	3.55
SEQ ID NO.:207	31	0.47	0.58	0.81	0.38	45	0.60	20	4.36
SEQ ID NO.:208	34	0.45	0.56	0.81	0.41	48	0.40	22	6.36
SEQ ID NO.:209	54	0.46	0.57	0.81	0.42	50	0.42	22	5.55
SEQ ID NO.:210	46	0.45	0.58	0.81	0.40	45	0.60	23	6.09
SEQ ID NO.:211	9	0.48	0.58	0.81	0.46	53	0.68	23	2.83
SEQ ID NO.:212	24	0.47	0.57	0.81	0.46	53	0.68	23	3.50
SEQ ID NO.:213	21	0.47	0.58	0.81	0.46	53	0.57	24	3.83
SEQ ID NO.:214	27	0.47	0.57	0.81	0.46	53	0.47	25	3.83
SEQ ID NO.:215	47	0.49	0.55	0.81	0.45	52	0.60	24	4.50
SEQ ID NO.:216	57	0.48	0.55	0.81	0.45	52	0.46	25	5.83
SEQ ID NO.:217	55	0.50	0.56	0.80	0.36	41	0.51	26	

FIG. 7b-9

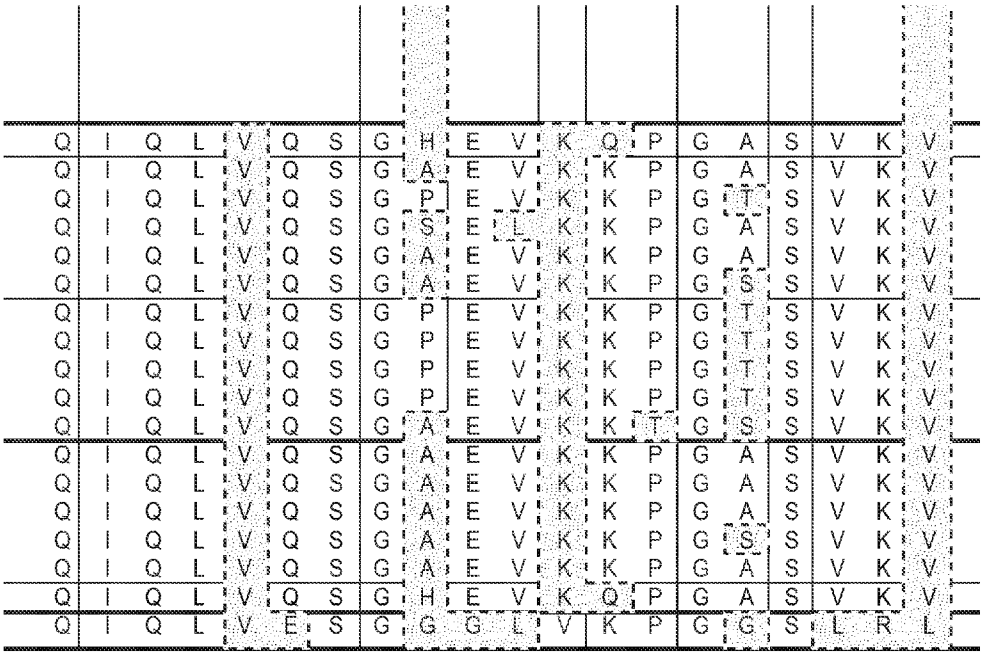


FIG. 7b-10

S	C	K	A	S	G	Y	T	F	T	D	_	_	Y	Y	I	T	W	V	R

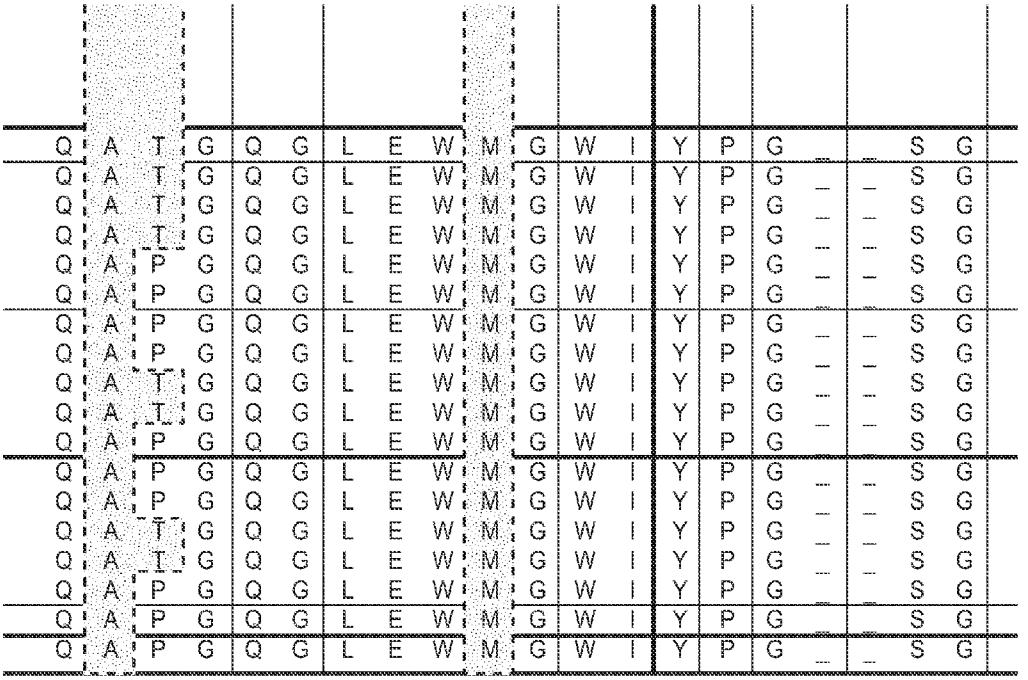

Q A T G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A T G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G
Q A P G Q G L E W M G W I Y P G _ _ S G

FIG. 7b-12

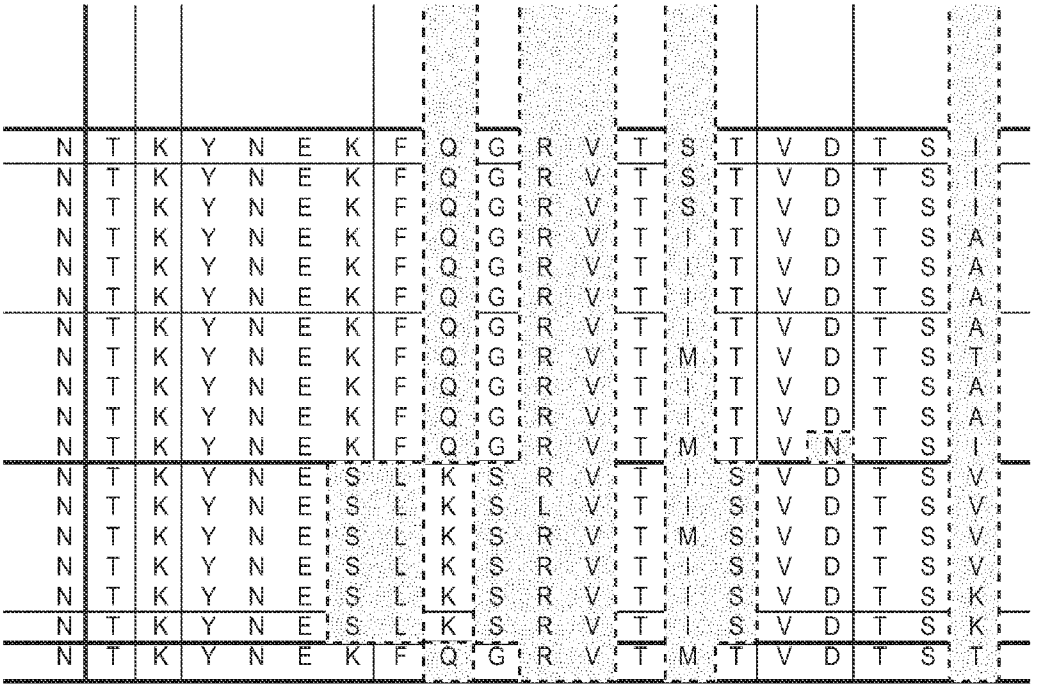

N T K Y N E K F Q G R V T S T V D T S I
N T K Y N E K F Q G R V T S T V D T S I
N T K Y N E K F Q G R V T S T V D T S I
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T M T V D T S T
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T I T V D T S A
N T K Y N E K F Q G R V T M T V N T S I
N T K Y N E S L K S R V T I S V D T S V
N T K Y N E S L K S L V T I S V D T S V
N T K Y N E S L K S R V T M S V D T S V
N T K Y N E S L K S R V T I S V D T S V
N T K Y N E S L K S R V T I S V D T S K
N T K Y N E S L K S R V T I S V D T S K
N T K Y N E K F Q G R V T M T V D T S T

FIG. 7b-13

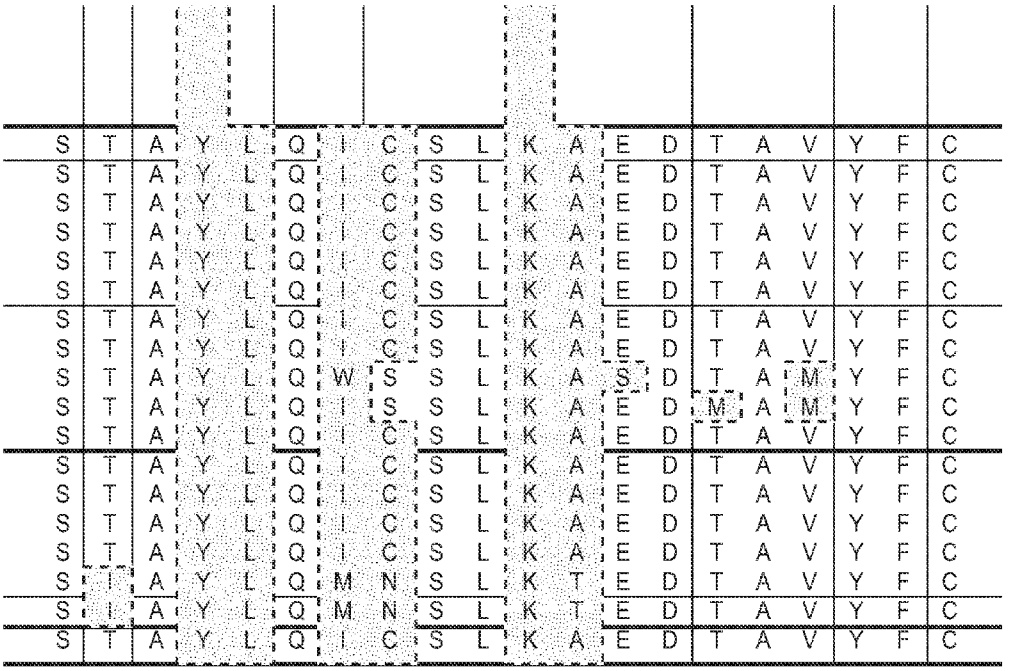

S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q W S S L K A S D T A M Y F C
S T A Y L Q I S S L K A E D M A M Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C
S I A Y L Q M N S L K T E D T A V Y F C
S I A Y L Q M N S L K T E D T A V Y F C
S T A Y L Q I C S L K A E D T A V Y F C

FIG. 7b-14

A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G
A	N	Y	G	N	_	_	_	_	_	_	Y	W	F	A	Y	W	G	Q	G

FIG. 7b-15

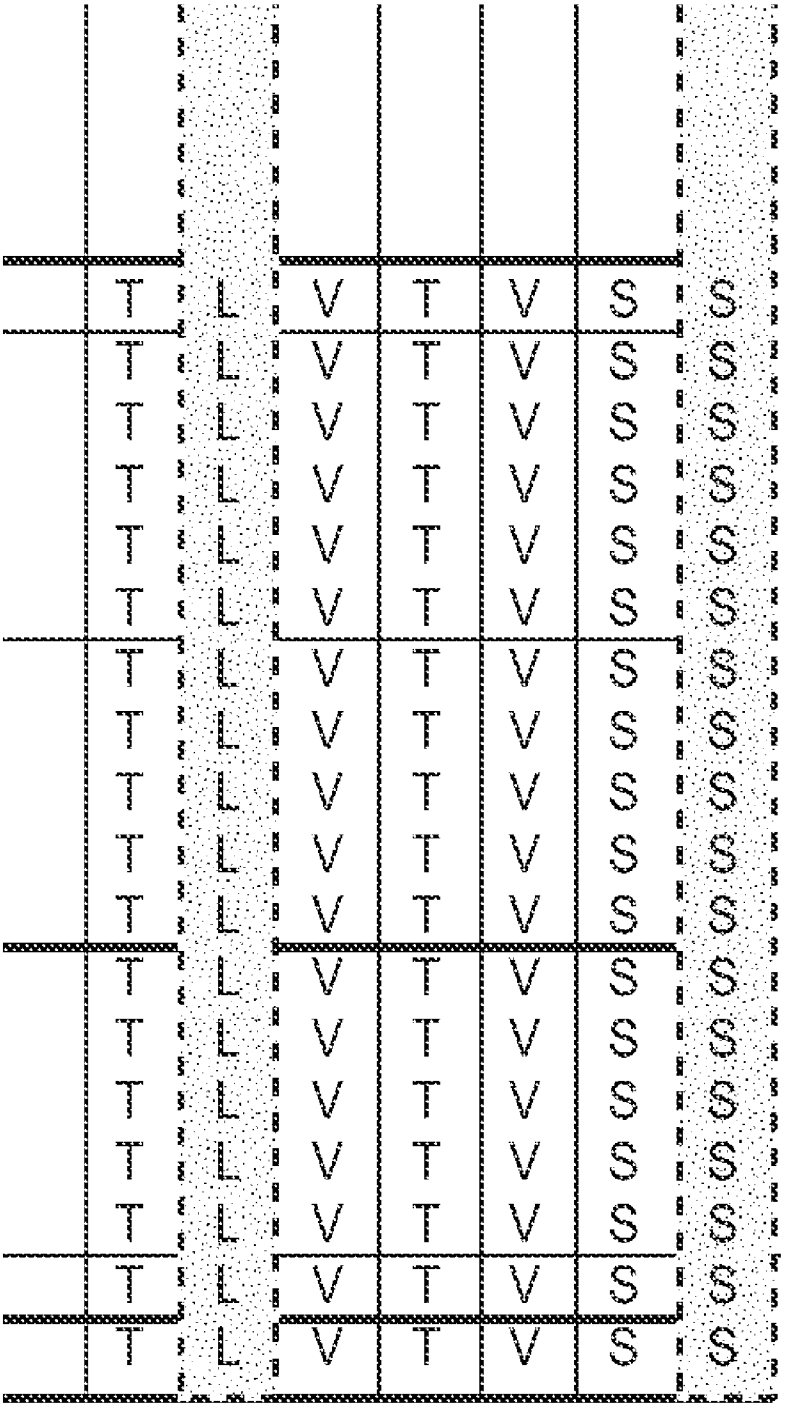


FIG. 7b-16

SEQ ID NO.: 218
DIVLTQSPATLSLSPGERATLSCRASQSVDFDGDSYMNWYQQKPGQPPKVLIYAAS
NLESGIPARFSGSGSGTDFTLTISSLQPEDFATYYCQQSNEDPWTFGGGTKVEIK

FIG. 8a

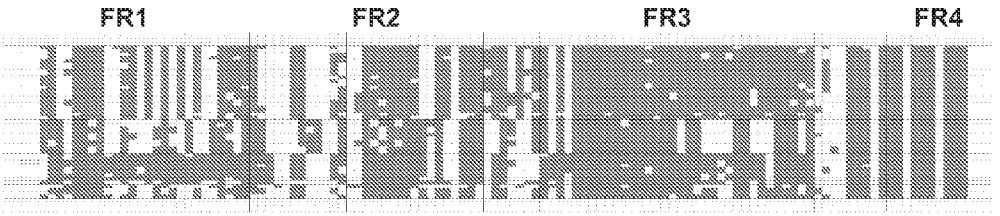


FIG. 8b

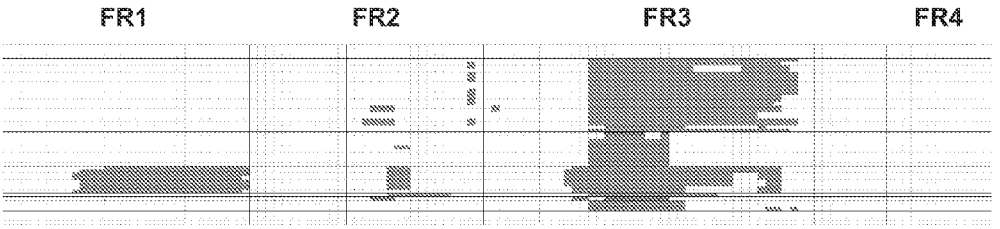


FIG. 8c



FIG. 8d

SEQ ID NO.: 219
DIVLTQSPSSLSASVGDRVTITCRASQSVDFDGDSYMNWYQQKPGQPPKVLIYAAS
NLESGIPARFSGSGSGTDFTLTISSLQPEDFATYYCQQSNEDPWTFGGGTKVEIK

FIG. 9a

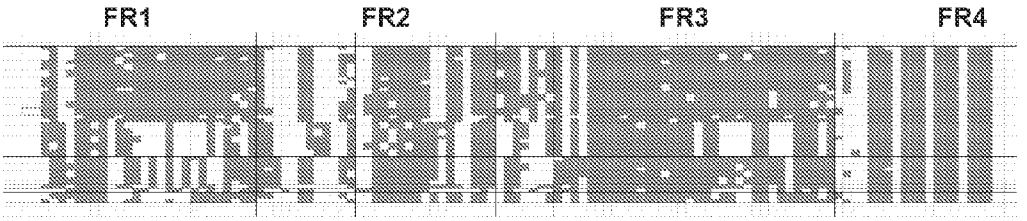


FIG. 9b

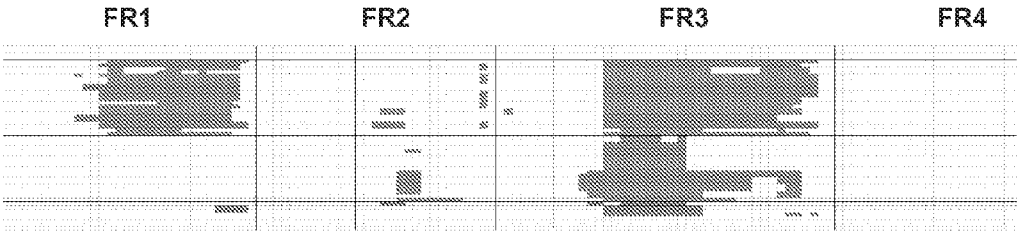


FIG. 9c

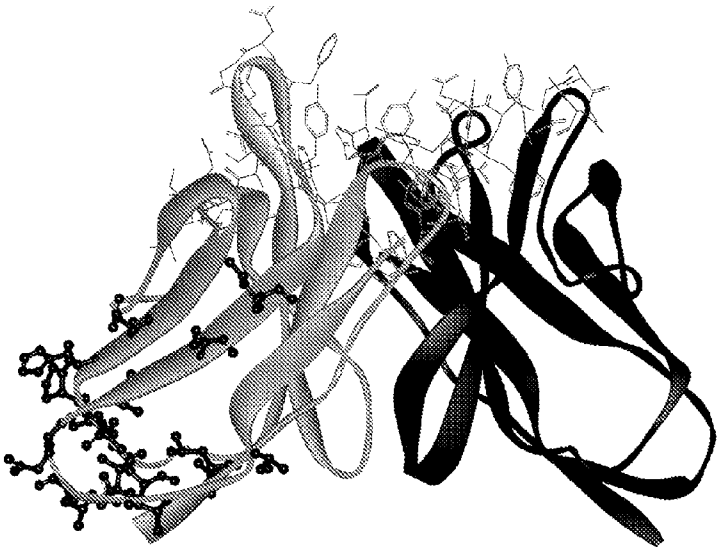


FIG. 9d

SEQ ID NO.: 220
DIVLTQSPDSLAVSLGERATINCKASQSVDFDGDSYMNWYQQKPGQPPKVLIYAAS
NLESGIPARFSGSGSGTDFTLTINSLEAEDAATYYCQQSNEDPWTFGGGTKVEIK

FIG. 10a

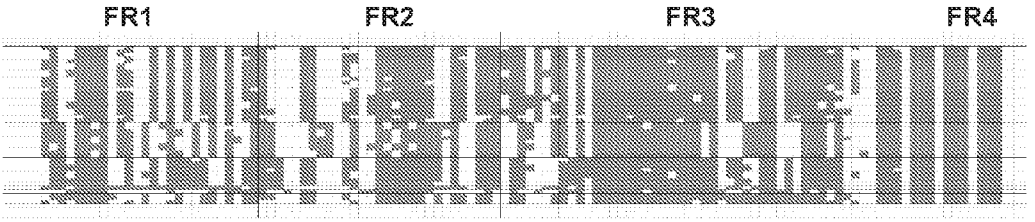


FIG. 10b

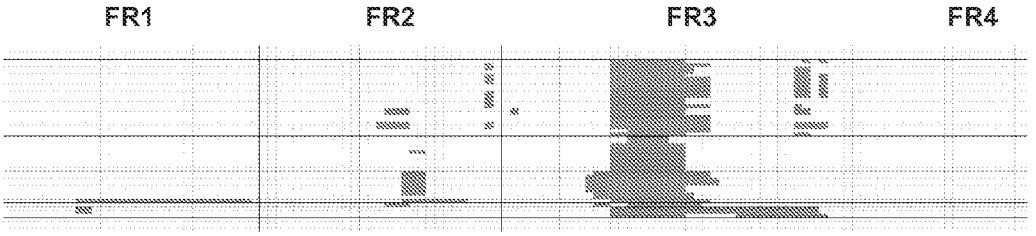


FIG. 10c



FIG. 10d

SEQ ID NO.: 221
QIQLVQSGPEVKKPGASVKVSCKASGYTFTDYYITWVRQAPGQGLEWMGWIYPG
SGNTKYNEKFQGRVTITVDTSASTAYMELSSLRSEDTAVYFCANYGNYWFAYWGQ
GTLVTVSS

FIG. 11a

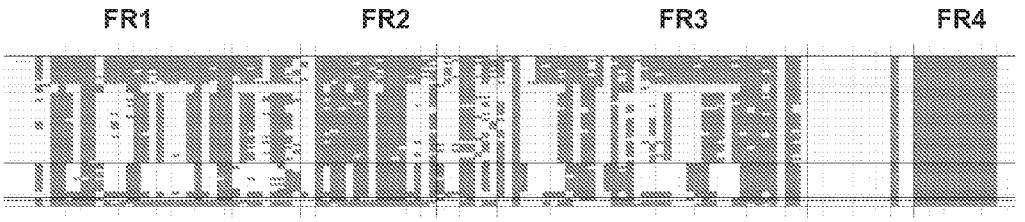


FIG. 11b

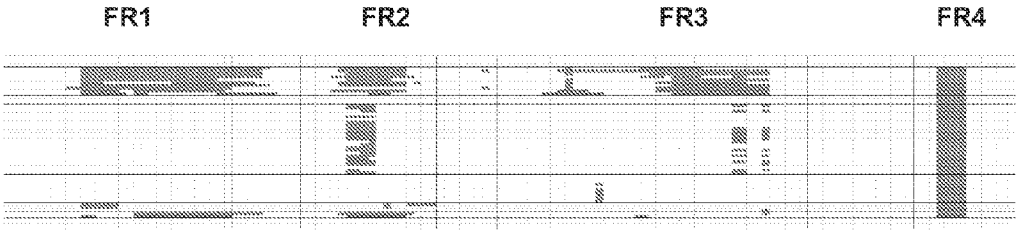


FIG. 11c

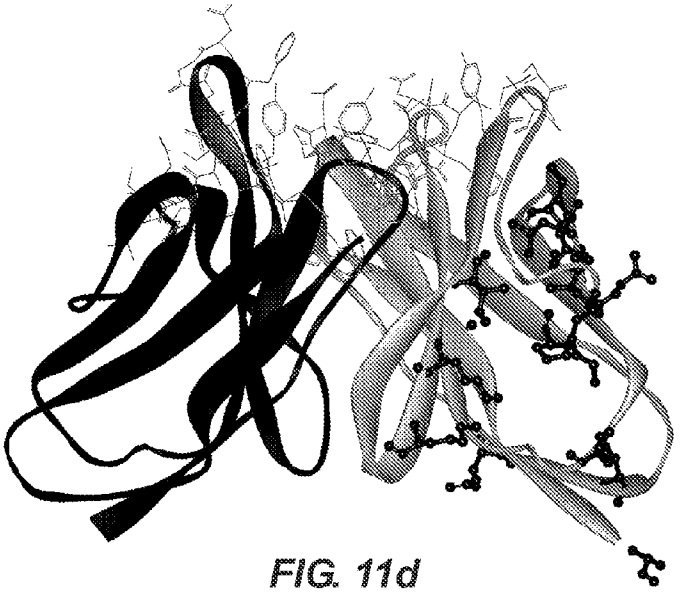


FIG. 11d

SEQ ID NO.: 222
QIQL**V**ESG**GGL**VKPG**GSLRL**SC**A**ASGYTFTDYYITWV**RQA**PGQGLEW**M**GWIYPG
SGNTKYNEKF**QG**R**VTM**TVDTST**S**TA**Y**M**E**LSSL**R**SEDTAVYFCANYGNYWFAYWG
QGTLVTVS**S**

FIG. 12a

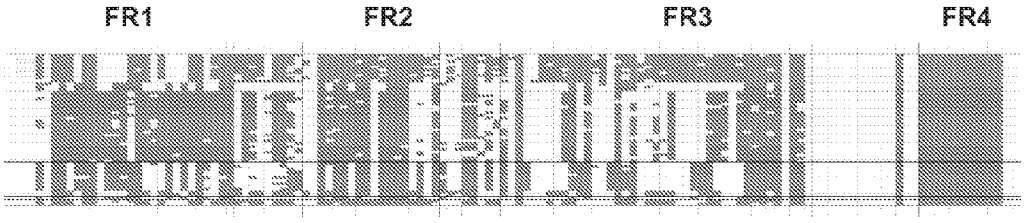


FIG. 12b

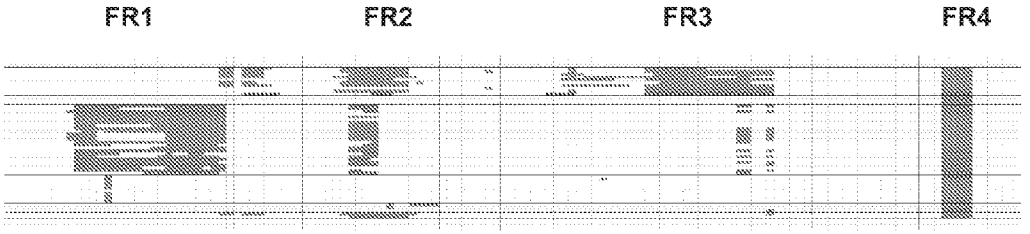


FIG. 12c

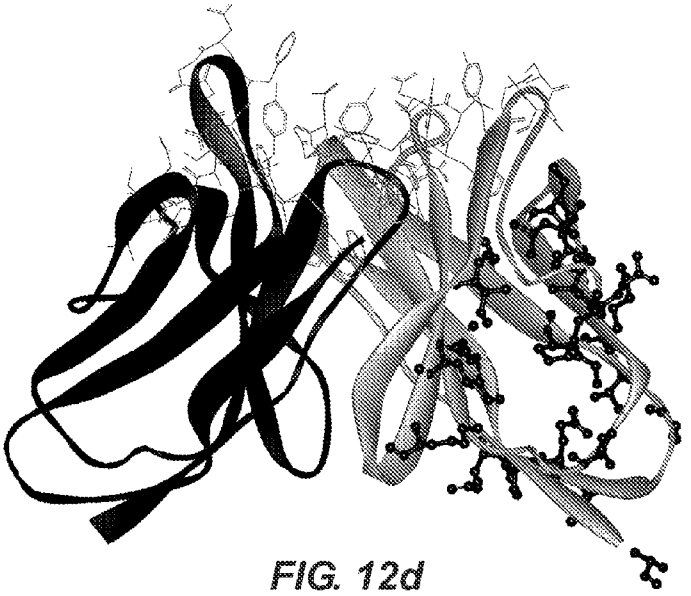


FIG. 12d

SEQ ID NO.: 223
QIQLVQSGPEVKKPGASVKVSCKASGYTFTDYYITWVRQAPGQGLEWMGWIYPG
SGNTKYNEKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYFCANYGNYWFAYWGQ
GTLVTVSS

FIG. 13a

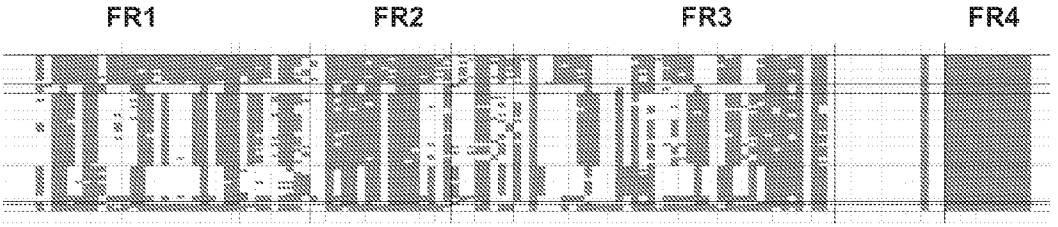


FIG. 13b

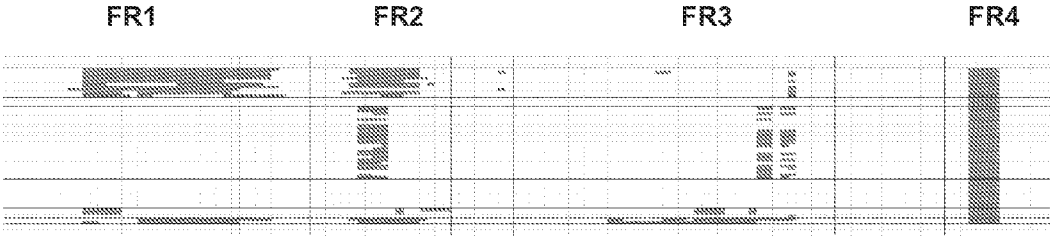


FIG. 13c

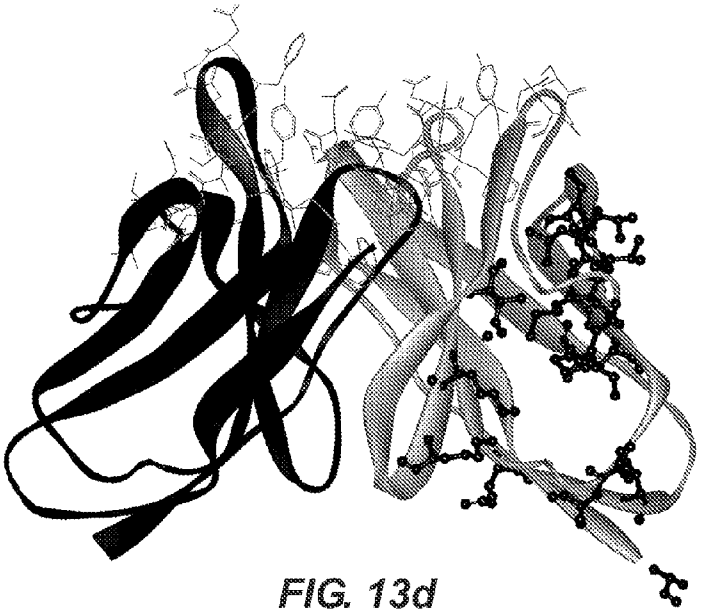


FIG. 13d

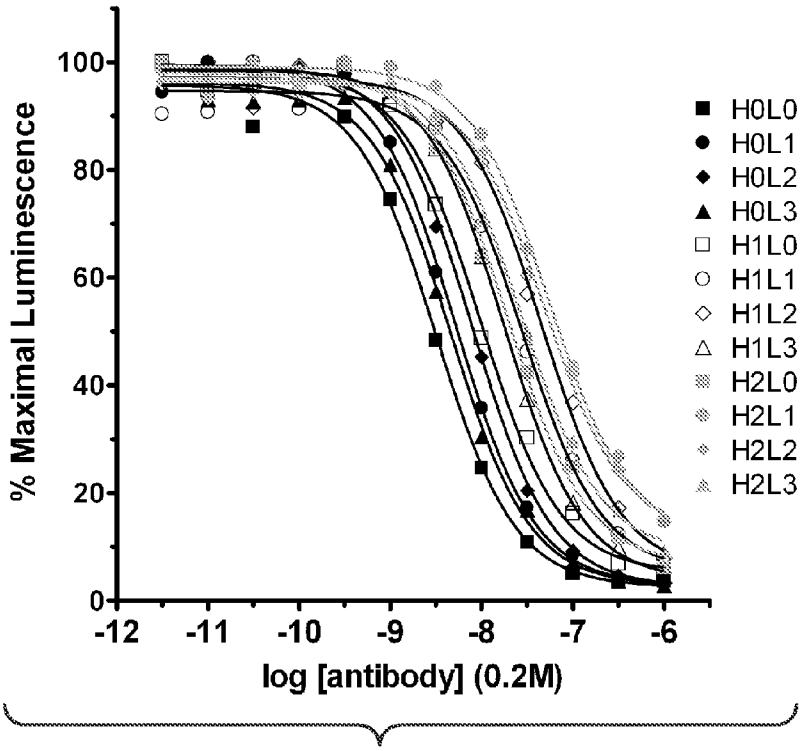


FIG. 14a

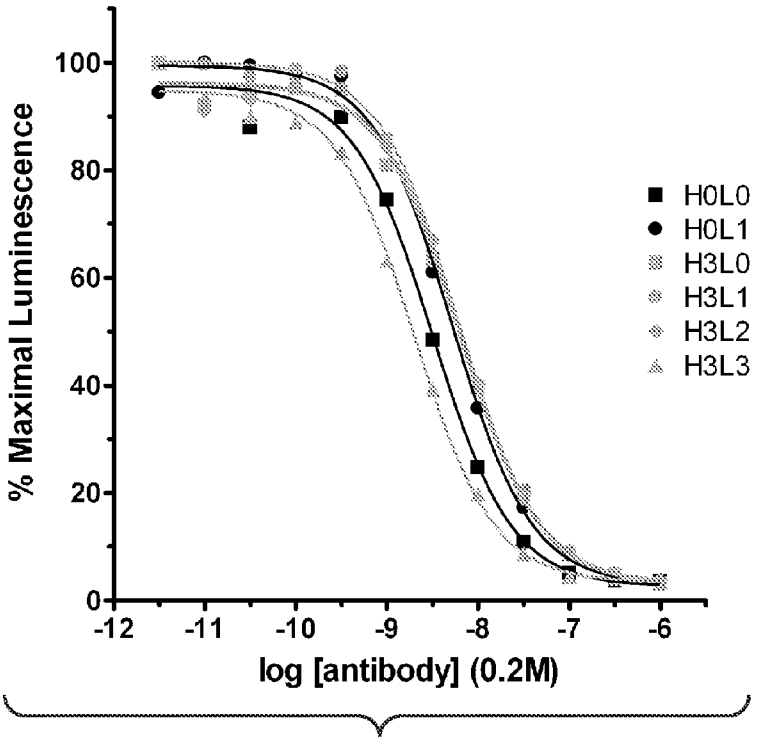


FIG. 14b

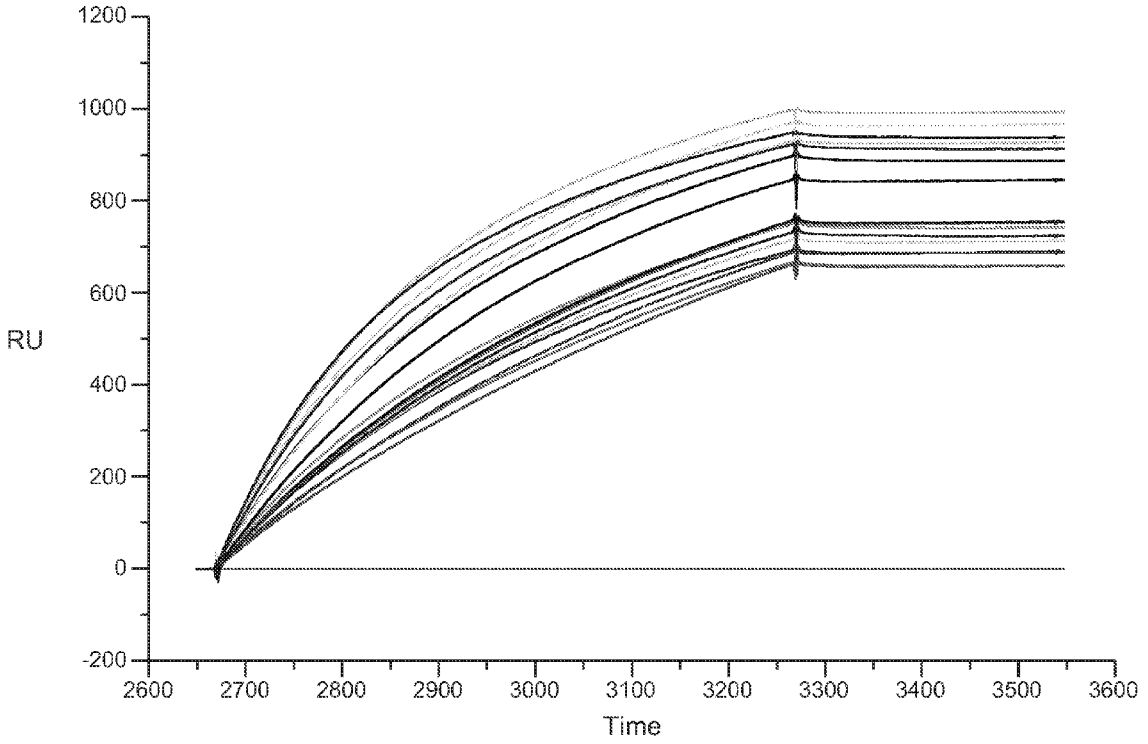

1200
1000
800
600
400
200
0
-200
RU
2600
2700
2800
2900
3000
3100
3200
3300
3400
3500
3600
Time

FIG. 15

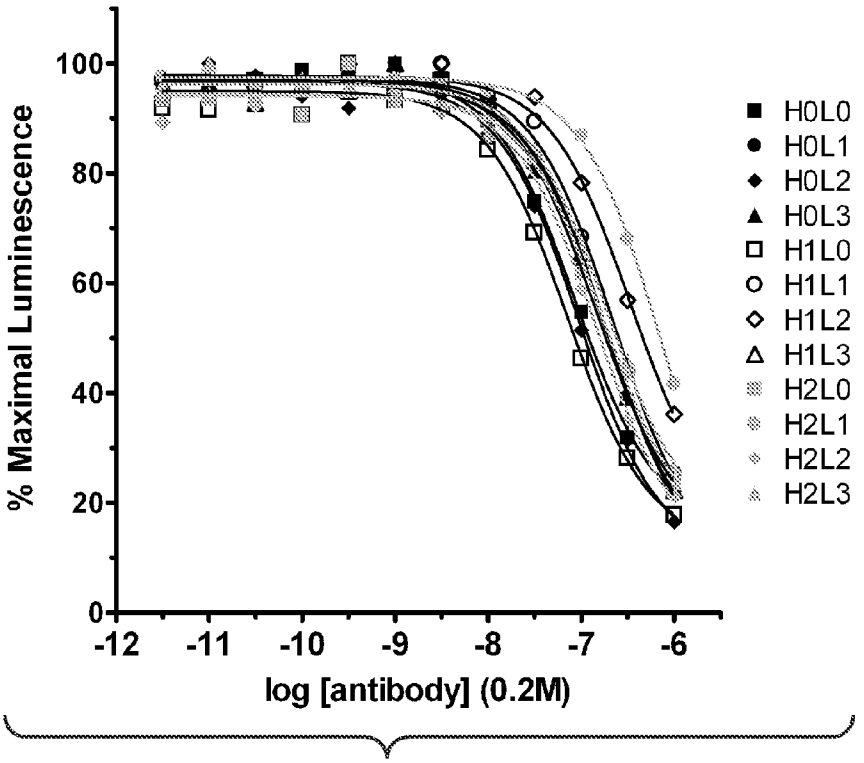


FIG. 16a

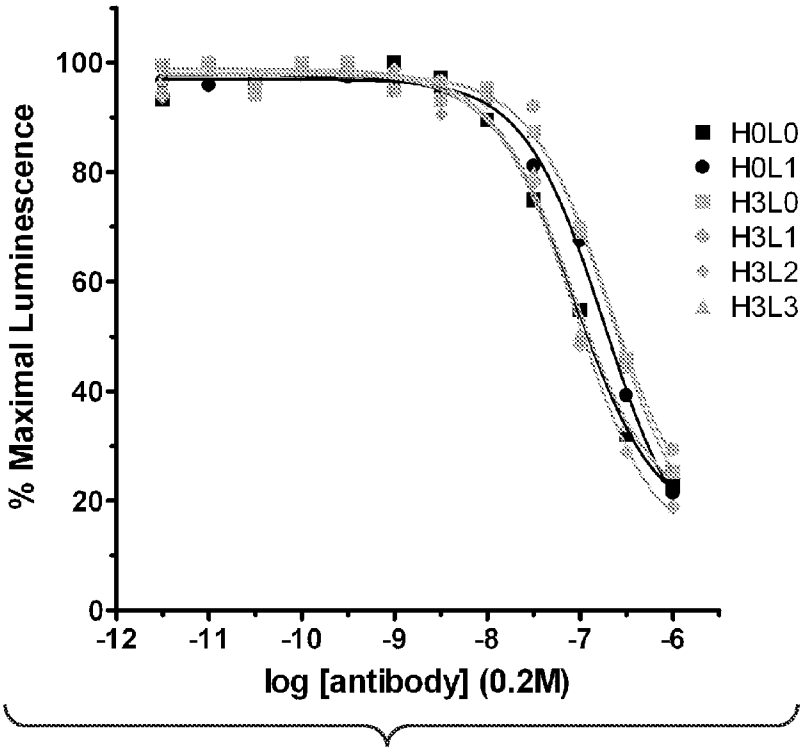


FIG. 16b

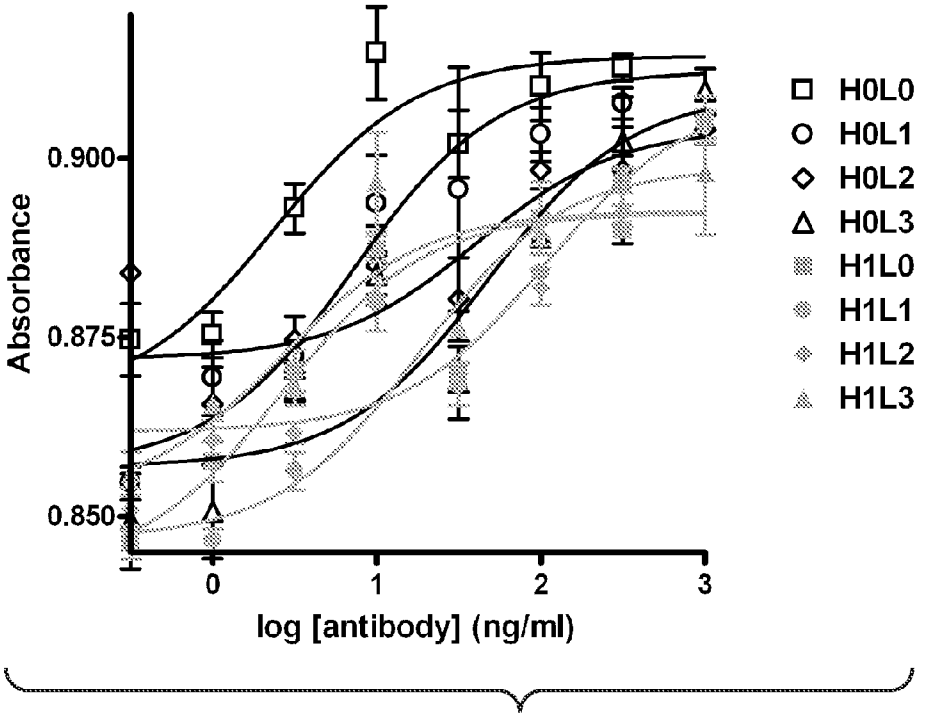


FIG. 17a

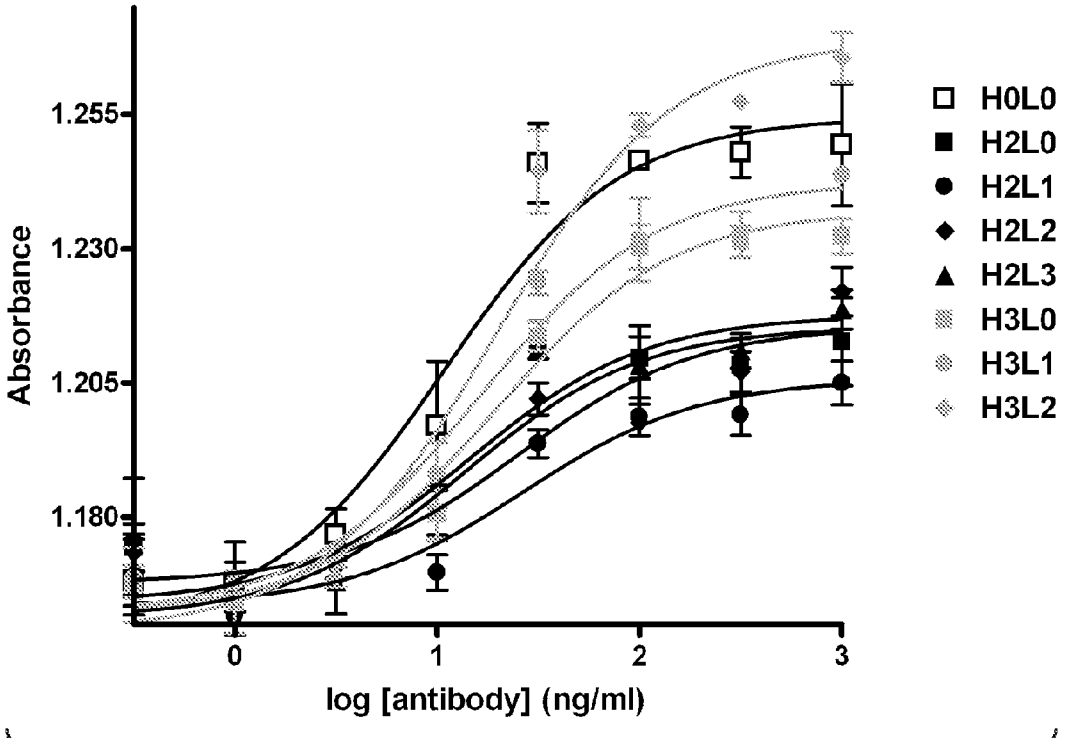


FIG. 17b

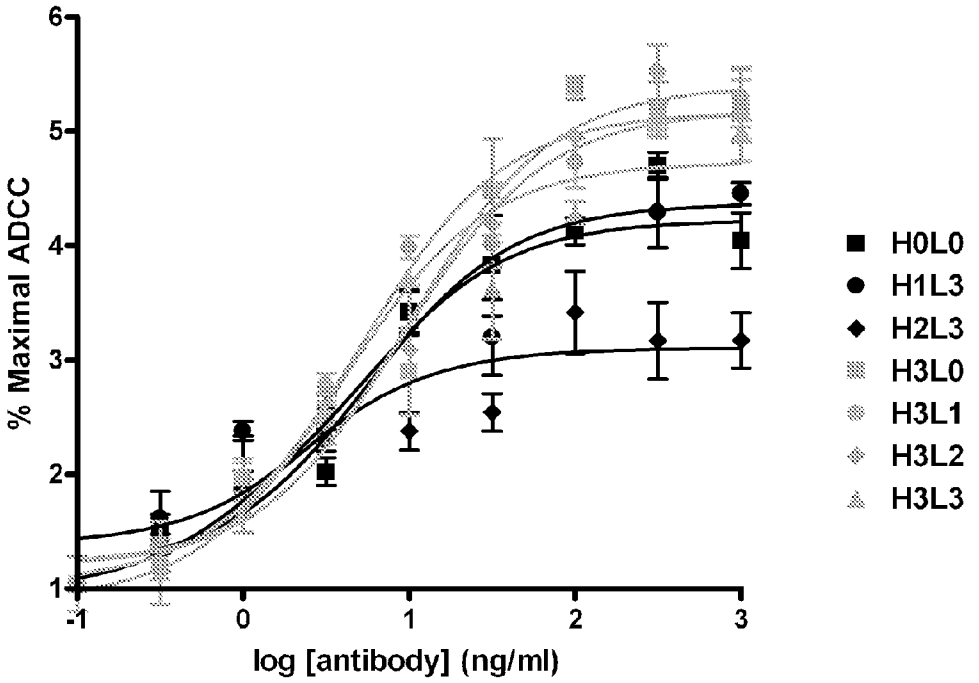


FIG. 17c

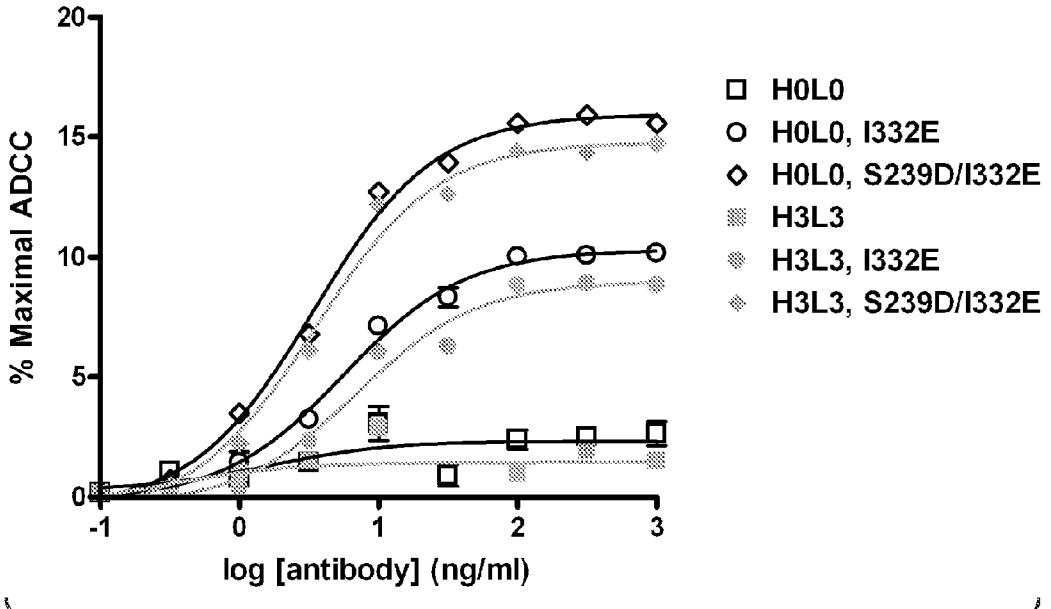


FIG. 18a

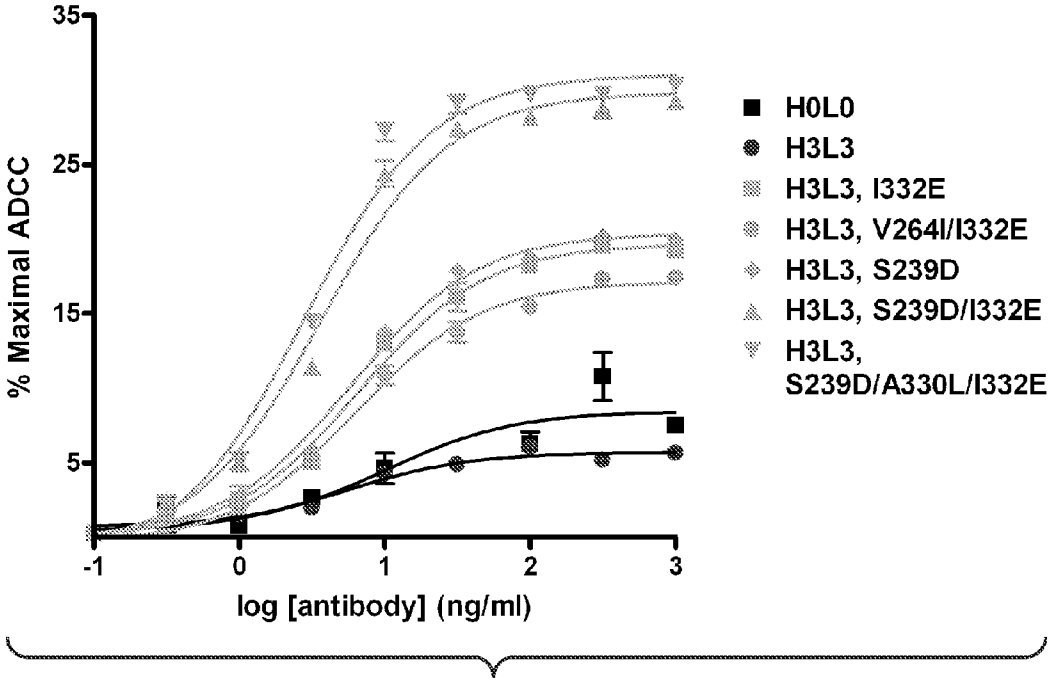


FIG. 18b

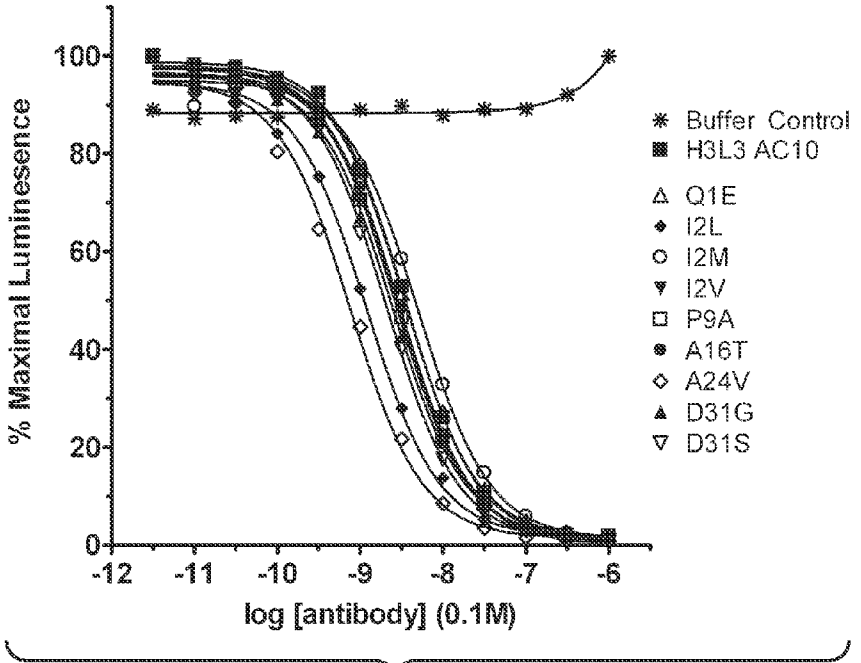


FIG. 19

SEQ ID NO.: 224
EIVLTQSPDSLAVSLGERATINCKASQSVDFDGDSYLNWYQQKPGQPPKVLIYAAST
LQSGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCQQSNEDPWTFGGGTKVEIK

FIG. 20a

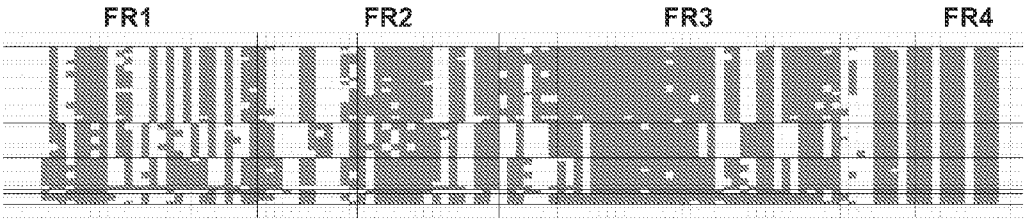


FIG. 20b

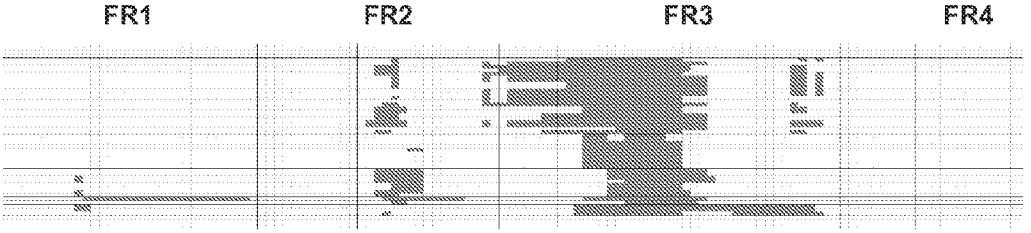


FIG. 20c

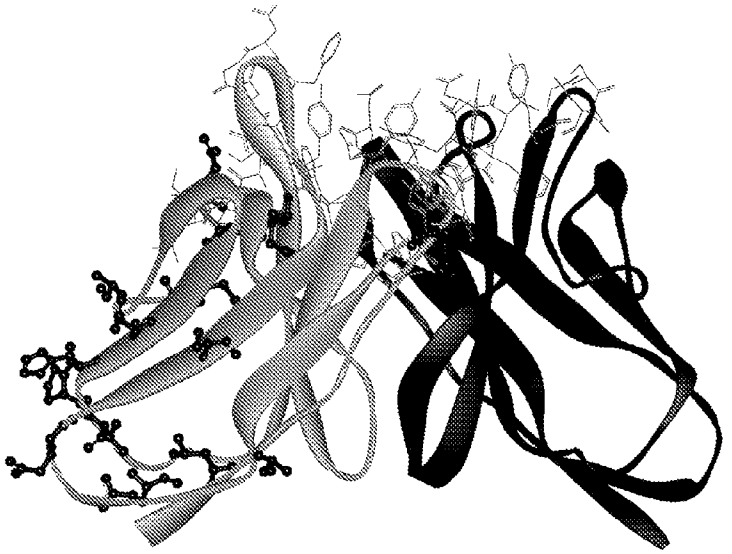


FIG. 20d

SEQ ID NO.: 225
AIVLTQSPDSLAVSLGERATINCKASQSVDFDGDSYLNWYQQKPGQPPKVLIYAAST
LETGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCQQSNEDPWTFGGGTKVEIK

FIG. 21a

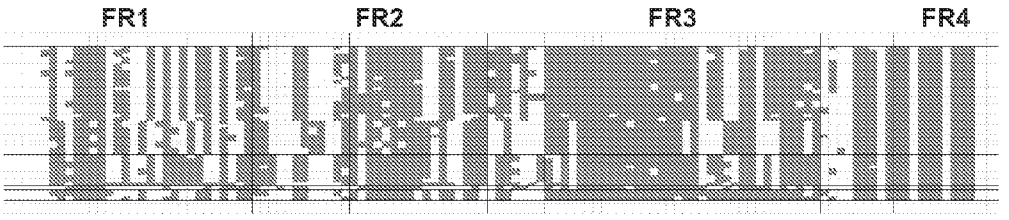


FIG. 21b

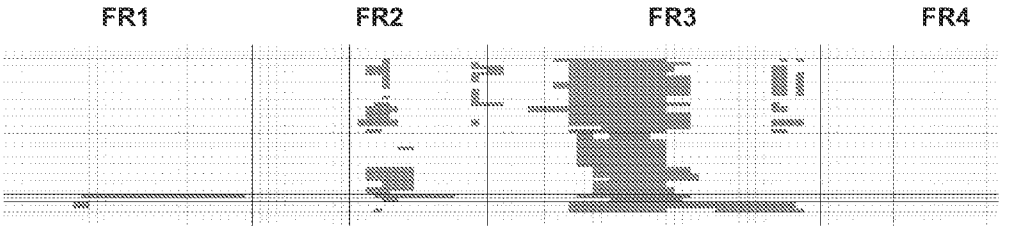


FIG. 21c



FIG. 21d

SEQ ID NO.: 226
QLQLVQSGPEVKKPGASVKVSCKVSGYTFTDYYITWVRQAPGQALEWMGWIYPG
SGNTKYNEKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYFCANYGNYWFAYWGQ
GTLVTVSS

FIG. 22a

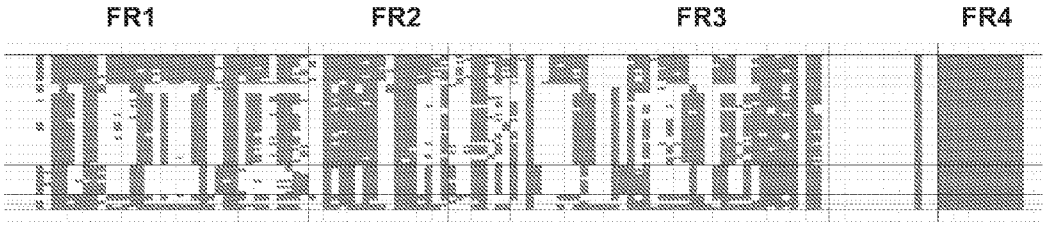


FIG. 22b

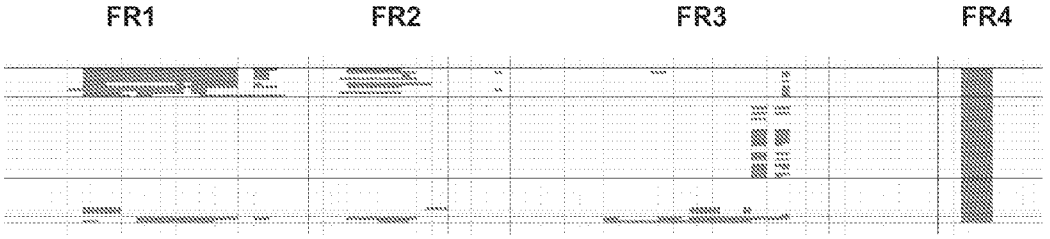


FIG. 22c

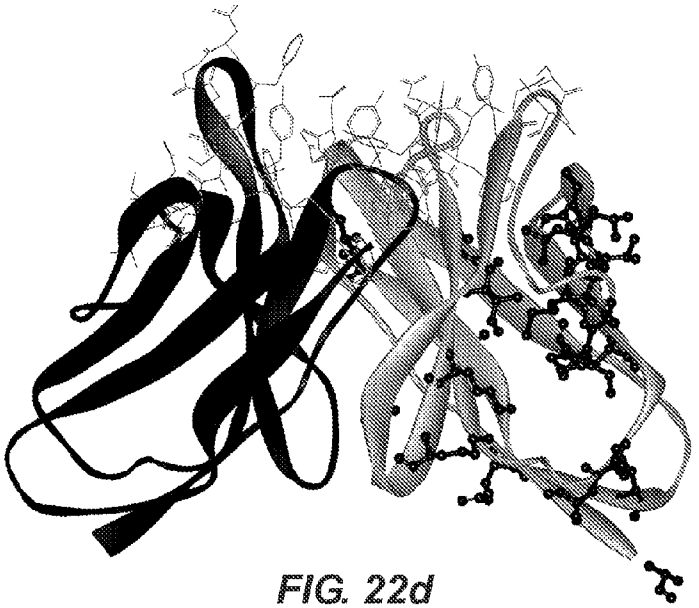


FIG. 22d

SEQ ID NO.: 227
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTDYYITWVRQAPGQALEWMGWIYPG
SGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFAYWGQ
GTLVTVSS

FIG. 23a

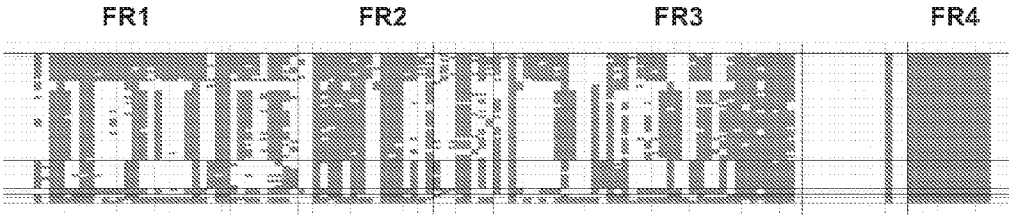


FIG. 23b

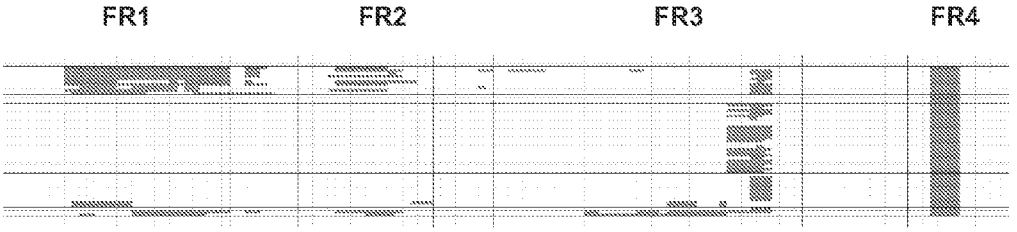


FIG. 23c

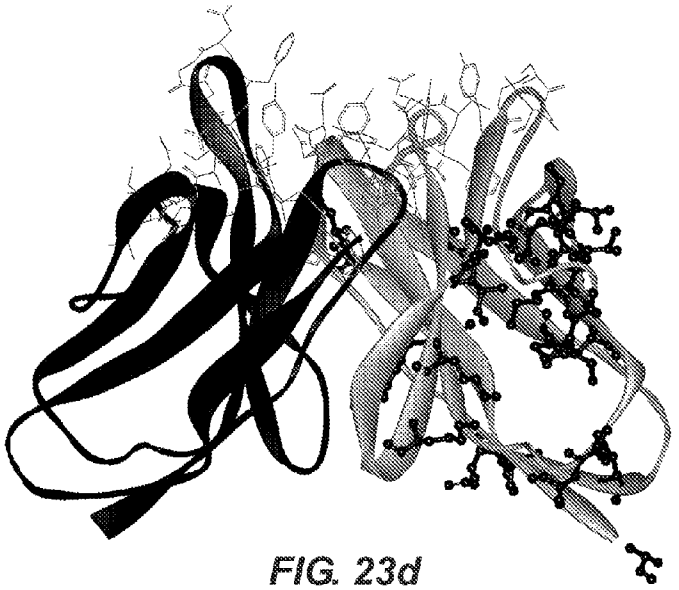


FIG. 23d

SEQ ID NO.: 228
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTSYYISWVRQAPGQALEWMGWIYAG
SGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFAYWGQ
GTLVTVSS

FIG. 24a

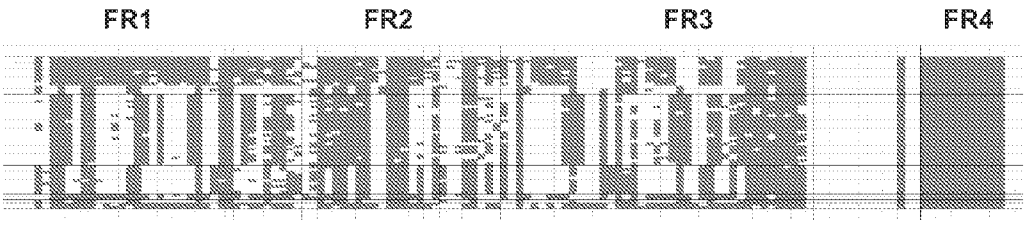


FIG. 24b

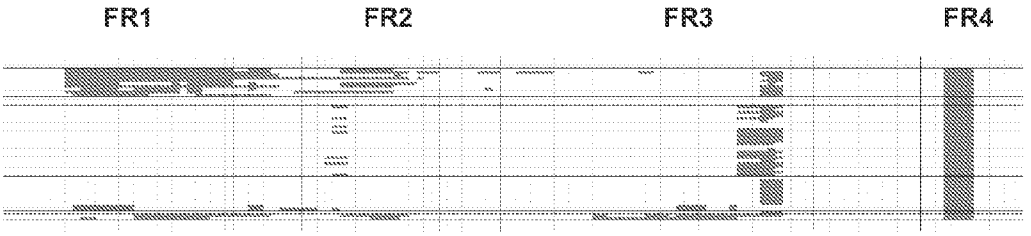


FIG. 24c

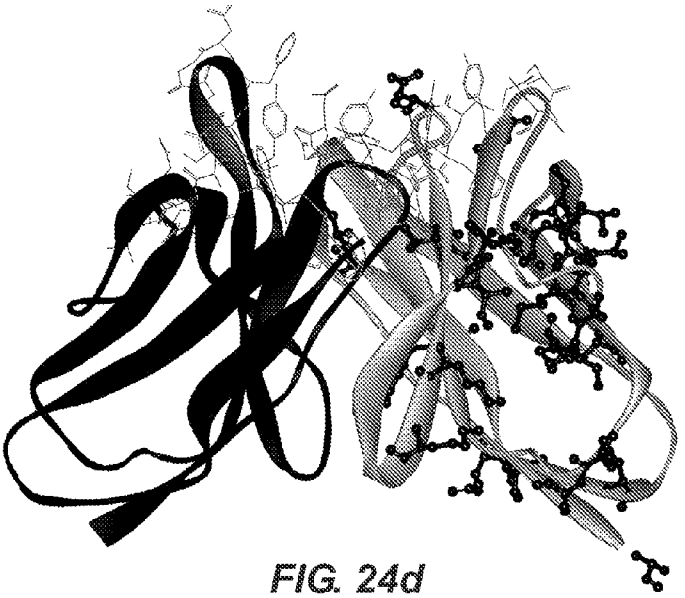


FIG. 24d

Anti-CD30 Light Chain

SEQ ID NO.:229
EIVLTQSPDSLAVSLGERATINCKASQSVDFDGDSYLNWYQQKPGQPPKVLIYA
ASTLQSGVPSRFSGSGSGTDFTLTINSLEAEDAATYYCQQSNEDPWTFGGGTK
VEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSG
NSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFN
RGEC

FIG. 25a

Anti-CD30 IgG1 Heavy Chain Comprising Possible Fc Variants

SEQ ID NO.:230
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTSYYISWVRQAPGQALEWMGWIY
AGSGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFA
YWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSW
NSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVD
KKAEPKSCDKTHTCPPCPAPELLGGP**X$_1$**VFLFPPKPKDTLMISRTPEVTCVV**X$_2$**
DVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNG
KEYKCKVSNKALP**X$_3$**P**X$_4$**EKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL
VKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQG
NVFSCSVMHEALHNHYTQKSLSLSPGK

Position	EU Index Position	WT	Possible Substitutions
X_1	239	S	D,E,N,Q,T
X_2	264	V	I,T,Y
X_3	330	A	Y,L,I
X_4	332	I	D,E,N,Q

FIG. 25b

Anti-CD30 Fc Variant IgG1 Heavy Chain

SEQ ID NO.:231
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTSYYISWVRQAPGQALEWMGWIY
AGSGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFA
YWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSW
NSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVD
KKAEPKSCDKTHTCPPCPAPELLGGP**D**VFLFPPKPKDTLMISRTPEVTCVV**V**DV
SHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKE
YKCKVSNKALP**L**P**E**EKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGF
YPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFS
CSVMHEALHNHYTQKSLSLSPGK

FIG. 25c

Anti-CD30 IgG2 Heavy Chain Comprising Possible Fc Variants

SEQ ID NO.:232
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTSYYISWVRQAPGQALEWMGWIY
AGSGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFA
YWGQGTLVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSW
NSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKVD
KTVERKCCVECPPCPAP$Z_1Z_2Z_3Z_4$GPX_1VFLFPPKPKDTLMISRTPEVTCVVX_2
DVSHEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNG
KEYKCKVSNKZ_5LPX_3PX_4EKTISKTKGQPREPQVYTLPPSREEMTKNQVSLTCL
VKGFYPSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQG
NVFSCSVMHEALHNHYTQKSLSLSPGK

Position	EU Index Position	WT	Possible Substitutions
X_1	239	S	D,E,N,Q,T
X_2	264	V	I,T,Y
X_3	330	A	Y,L,I
X_4	332	I	D,E,N,Q
Z_1	233	P	E
Z_2	234	V	L
Z_3	235	A	L
Z_4	236	-	G
Z_5	327	G	A

FIG. 25d

Anti-CD30 Fc Variant IgG2 Heavy Chain

SEQ ID NO.:233
QLQLVQSGAEVKKPGASVKVSCKVSGYTFTSYYISWVRQAPGQALEWMGWIY
AGSGNTKYSQKFQGRFVFSVDTSASTAYLQISSLKAEDTAVYYCANYGNYWFA
YWGQGTLVTVSSASTKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSW
NSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSNFGTQTYTCNVDHKPSNTKV
DKTVERKCCVECPPCPAP**ELLG**GP**D**VFLFPPKPKDTLMISRTPEVTCVV**V**DVS
HEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTFRVVSVLTVVHQDWLNGKEY
KCKVSNK**G**LP**L**P**E**EKTISK**T**KGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFY
PSDIAVEWESNGQPENNYKTTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFSC
SVMHEALHNHYTQKSLSLSPGK

FIG. 25e

SEQ ID NO.: 234
DILLTQSPVILSVSPGERVSFSCRASQSIGTNIHWYQQRTNGSPRLLIKYASESISGIP
SRFSGSGSGTDFTLSINSVESEDIADYYCQQNNNWPTTFGAGTKLELK

FIG. 26a

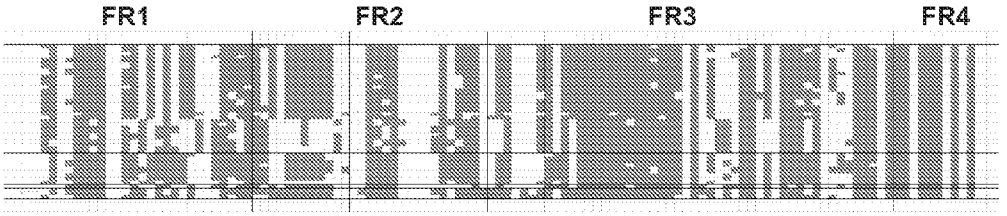


FIG. 26b

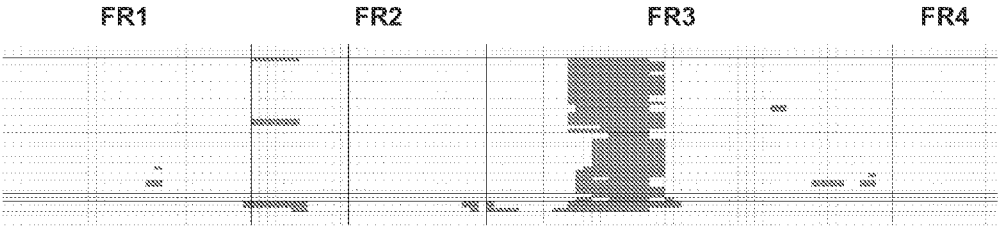


FIG. 26c

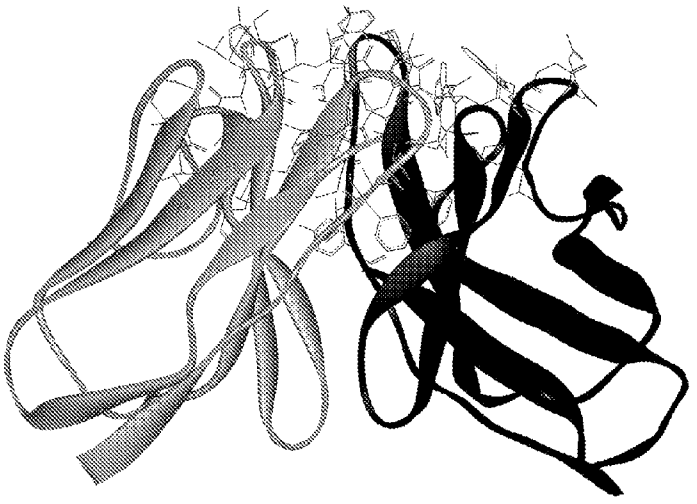


FIG. 26d

SEQ ID NO.: 235
QVQLKQSGPGLVQPSQSLSITCTVSGFSLTNYGVHWVRQSPGKGLEWLGVIWSG
GNTDYNTPFTSRLSINKDNSKSQVFFKMNSLQSNDTAIYYCARALTYYDYEFAYWG
QGTLVTVSA

FIG. 27a

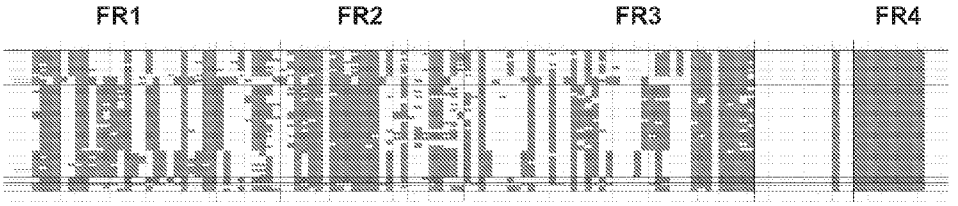


FIG. 27b

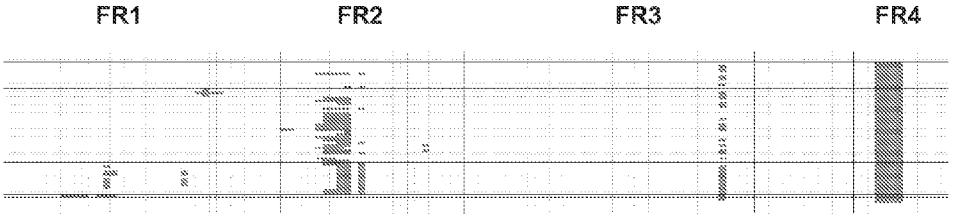


FIG. 27c

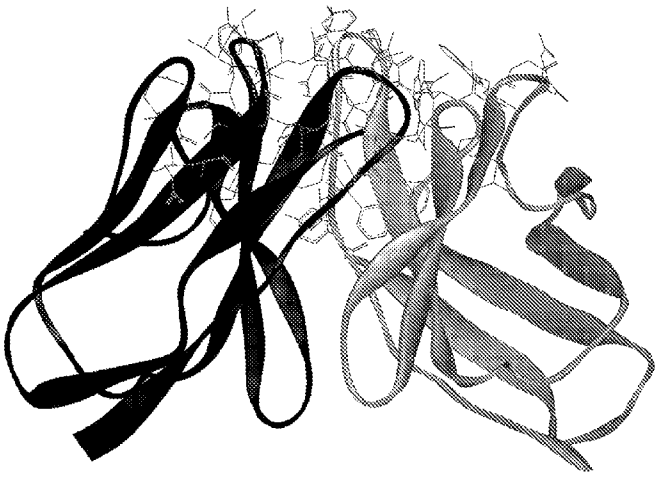


FIG. 27d

SEQ ID NO.: 236
EI**V**LTQSP**DFQ**SVTP**K**EK**VTIT**CRASQSIGTN**L**HWYQQ**KPDQ**SP**K**LLIKYASESISG
VPSRFSGSGSGTDFTL**T**INSLE**A**ED**A**ATYYC**H**QNNNWPTTFGAGTKLE**I**K

FIG. 28a

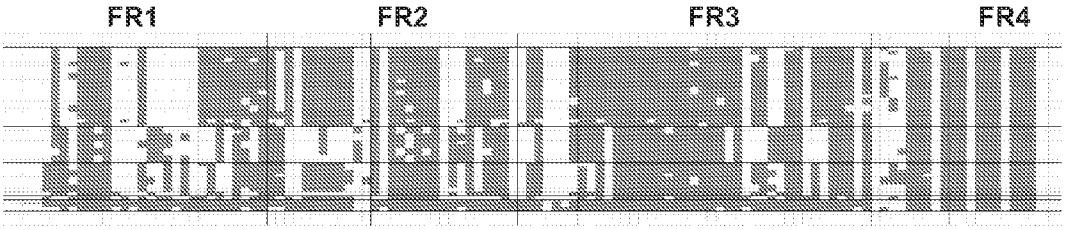


FIG. 28b

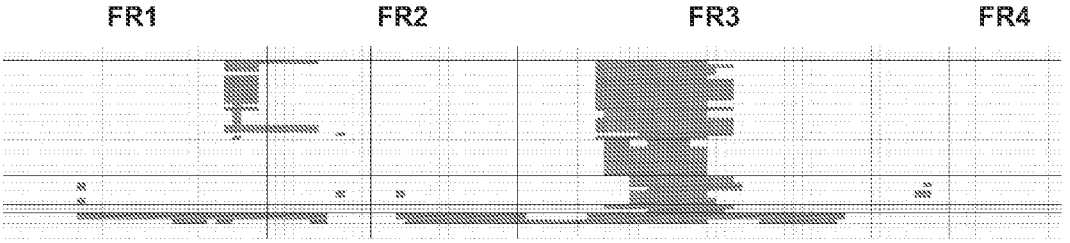


FIG. 28c

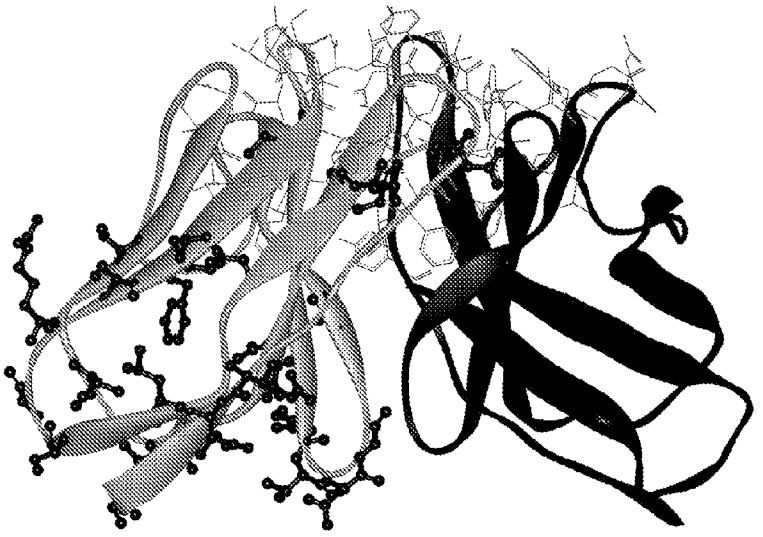


FIG. 28d

SEQ ID NO.: 237
QVQL**Q**ESGPGLV**K**PSQ**T**LSLTCTVSGFSLTNYG**WS**WIRQ**P**PGKGLEWIG**Y**IWSGG
NT**Y**YN**PSLK**SR**VTISV**DTSK**N**Q**FSL**KL**SS**VTAA**DTA**V**YYCARALTYYDYEFAYWGQ
GTLVTVS**S**

FIG. 29a

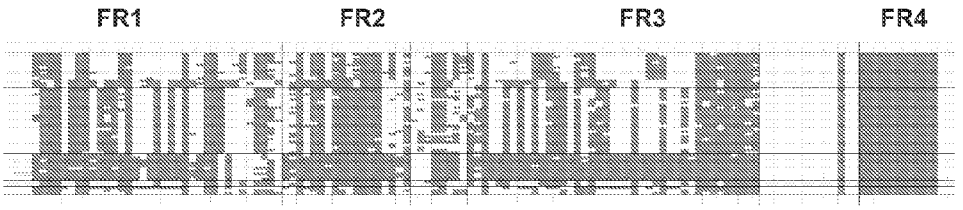


FIG. 29b

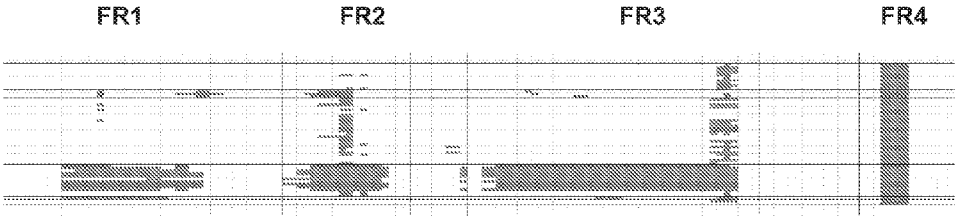


FIG. 29c

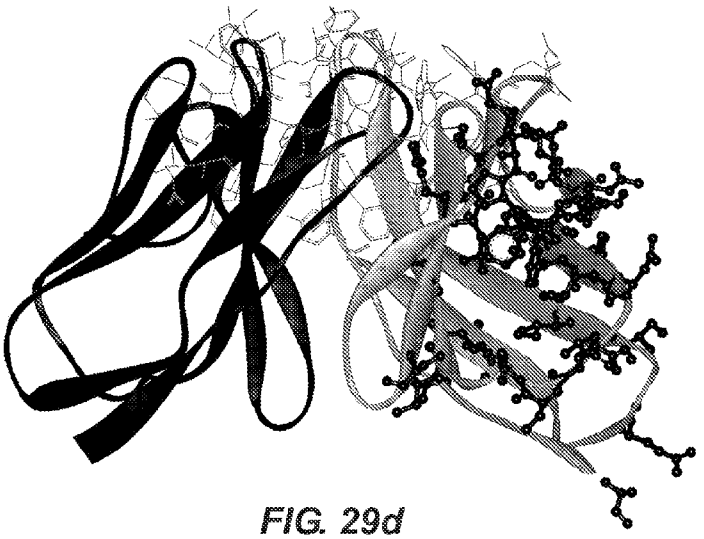


FIG. 29d

C225 VL HEC Calculation 1

(WT)									Kabat WT Muts
	SEQ ID NO.:234	Iter	Structural Consensus	Structural Precedence	HSC	HSS	N_9max	FRH	
	SEQ ID NO.:238	68	0.55	0.62	0.92	0.57	58	0.45	17
	SEQ ID NO.:239	94	0.56	0.60	0.92	0.57	58	0.47	18
	SEQ ID NO.:240	21	0.55	0.61	0.92	0.57	58	0.52	17
	SEQ ID NO.:241	16	0.54	0.61	0.92	0.57	58	0.52	17
	SEQ ID NO.:242	59	0.58	0.60	0.92	0.58	60	0.50	20
	SEQ ID NO.:243	86	0.54	0.58	0.92	0.57	58	0.67	17
	SEQ ID NO.:244	31	0.57	0.58	0.92	0.58	60	0.72	19
	SEQ ID NO.:245	25	0.54	0.58	0.92	0.57	58	0.67	16
	SEQ ID NO.:246	3	0.53	0.58	0.92	0.57	58	0.67	16
	SEQ ID NO.:247	41	0.55	0.58	0.92	0.57	58	0.45	17
	SEQ ID NO.:248	2	0.54	0.58	0.92	0.57	58	0.67	16
	SEQ ID NO.:249	54	0.53	0.57	0.91	0.51	52	0.75	17
	SEQ ID NO.:250	28	0.53	0.58	0.92	0.52	53	0.62	16
	SEQ ID NO.:251	78	0.53	0.58	0.92	0.52	53	0.62	16
	SEQ ID NO.:252	5	0.55	0.56	0.92	0.55	56	0.55	19
	SEQ ID NO.:253	8	0.54	0.56	0.92	0.55	56	0.55	19
	SEQ ID NO.:254	45	0.52	0.55	0.91	0.52	53	0.51	16
	SEQ ID NO.:255	4	0.52	0.55	0.91	0.52	53	0.51	16
	SEQ ID NO.:256	34	0.52	0.54	0.92	0.53	54	0.50	17
	SEQ ID NO.:257	6	0.52	0.55	0.92	0.53	54	0.50	16
	SEQ ID NO.:258	19	0.52	0.55	0.92	0.53	54	0.50	16
	SEQ ID NO.:259	17	0.52	0.55	0.92	0.53	54	0.50	16
	SEQ ID NO.:260	1	0.52	0.55	0.91	0.52	53	0.51	16
	SEQ ID NO.:261	76	0.53	0.54	0.91	0.52	53	0.57	17
	SEQ ID NO.:262	55	0.54	0.54	0.91	0.52	53	0.57	17
	SEQ ID NO.:263	48	0.53	0.53	0.91	0.52	53	0.53	17

FIG. 30a-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D	I	L	L	T	Q	S	P	V	I	L	S	V	S	P	G	E	R	V	S
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T

FIG. 30a-2

21	22	23	24	25	26	27	28	29	30							31	32	33	34
F	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
I	T	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
I	T	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
I	T	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
I	T	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H
L	S	C	R	A	S	Q	S	I	G	_	_	_	_	_	_	T	N	I	H

FIG. 30a-3

35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
W	Y	Q	Q	R	T	N	G	S	P	R	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	L	Q	K	P	G	K	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	L	Q	K	P	G	K	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	L	Q	K	P	G	Q	S	P	Q	L	L	I	K	Y	A	S	E	S

FIG. 30a-4

55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	S
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T

FIG. 30a-5

75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
I	N	S	V	E	S	E	D	I	A	D	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	Q	P	E	D	F	A	T	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	E	P	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	Q	S	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	S	S	L	Q	P	E	D	I	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	V	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	V	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	N	I	E	S	E	D	F	A	V	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W
I	N	S	L	E	A	E	D	A	A	T	Y	Y	C	Q	Q	N	N	N	W

FIG. 30a-6

		95	96	97	98	99	100	101	102	103	104	105	106	107
_	_	P	T	T	F	G	Q	G	T	K	V	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K
_	_	P	T	T	F	G	Q	G	T	K	X	E	I	K

FIG. 30a-7

FIG. 30a-1	FIG. 30a-2	FIG. 30a-3	FIG. 30a-4	FIG. 30a-5	FIG. 30a-6	FIG. 30a-7

FIG. 30a

C225 VH HEC Calculation 2

(WT) SEQ ID NO.:235	Iter	Structural Consensus	Structural Precedence	HSC	HSS	$N_g$max	FRH	Kabat WT Muts
SEQ ID NO.:264	23	0.48	0.51	0.79	0.42	47	0.40	20
SEQ ID NO.:265	5	0.49	0.51	0.79	0.42	47	0.49	20
SEQ ID NO.:266	7	0.47	0.50	0.79	0.42	46	0.41	20
SEQ ID NO.:267	3	0.48	0.50	0.79	0.42	46	0.46	20
SEQ ID NO.:268	4	0.48	0.50	0.79	0.42	47	0.38	20
SEQ ID NO.:269	27	0.47	0.50	0.79	0.42	46	0.39	20
SEQ ID NO.:270	9	0.47	0.49	0.79	0.43	48	0.40	21
SEQ ID NO.:271	2	0.48	0.49	0.79	0.43	48	0.50	21
SEQ ID NO.:272	49	0.47	0.49	0.78	0.41	45	0.40	20
SEQ ID NO.:273	100	0.47	0.49	0.79	0.43	48	0.38	21
SEQ ID NO.:274	1	0.46	0.43	0.79	0.43	48	0.35	23

FIG. 30b-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Q	V	Q	L	K	Q	S	G	P	G	L	V	Q	P	S	Q	S	L	S	I
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L

FIG. 30b-2

21	22	23	24	25	26	27	28	29	30	31		32	33	34	35	36	37	38	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	
T	C	T	V	S	G	F	S	L	T	N	_	_ Y	G	V	H	W	V	R	

FIG. 30b-3

39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54				55	
Q	S	P	G	K	G	L	E	W	L	G	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G				G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G				G	
Q	A	P	G	K	G	L	E	W	V	A	V	I	W	S	G	_	_	_	G	

FIG. 30b-4

56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
N	T	D	Y	N	T	P	F	T	S	R	L	S	I	N	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K

FIG. 30b-5

76	77	78	79	80	81	82	82a	82b	82c	83	84	85	86	87	88	89	90	91	92
S	Q	V	F	F	K	M	N	S	L	Q	S	N	D	T	A	I	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	T	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	T	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	Q	V	V	L	T	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	T	M	R	S	L	R	S	D	D	T	A	V	Y	Y	C
S	T	V	Y	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	T	V	Y	L	Q	W	S	S	L	K	A	S	D	T	A	M	Y	Y	C

FIG. 30b-6

93	94	95	96	97				98	99	100	100a	100b	100c	101	102	103	104	105	106
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G

FIG. 30b-7

107	108	109	110	111	112	113
T	L	V	T	V	S	A
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S

FIG. 30b-8

FIG. 30b-1	FIG. 30b-2	FIG. 30b-3	FIG. 30b-4	FIG. 30b-5	FIG. 30b-6	FIG. 30b-7	FIG. 30b-8

FIG. 30b

FIG. 31a-1

C225 VL HEC Calculation 2

(WT) SEQ ID NO.:234	Iter	Structural Consensus	Structural Precedence	HSC	HSS	N_9max	FRH	Muts	Kabat WT Cluster
SEQ ID NO.:275	1	0.54	0.58	0.92	0.57	58	0.55	21	3.13
SEQ ID NO.:276	12	0.54	0.57	0.92	0.57	58	0.55	22	3.63
SEQ ID NO.:277	51	0.54	0.57	0.92	0.57	58	0.50	22	3.63
SEQ ID NO.:278	48	0.53	0.55	0.92	0.57	58	0.55	21	4.13
SEQ ID NO.:279	28	0.52	0.55	0.92	0.58	60	0.75	22	4.13
SEQ ID NO.:280	35	0.52	0.52	0.92	0.58	60	0.75	23	4.75
SEQ ID NO.:281	49	0.53	0.55	0.92	0.58	60	0.75	23	3.63
SEQ ID NO.:282	38	0.53	0.60	0.92	0.57	58	0.78	19	5.25
SEQ ID NO.:283	2	0.56	0.59	0.92	0.58	60	0.50	22	3.50
SEQ ID NO.:284	17	0.57	0.60	0.92	0.58	60	0.50	22	3.25
SEQ ID NO.:285	47	0.58	0.61	0.92	0.58	60	0.50	22	4.25
SEQ ID NO.:286	15	0.57	0.57	0.92	0.58	60	0.72	20	3.75
SEQ ID NO.:287	32	0.56	0.57	0.92	0.58	60	0.72	20	4.00
SEQ ID NO.:288	40	0.57	0.58	0.92	0.58	60	0.52	22	3.75
SEQ ID NO.:289	55	0.59	0.61	0.92	0.58	60	0.58	22	4.50
SEQ ID NO.:290	21	0.54	0.50	0.92	0.56	58	0.48	22	6.25
SEQ ID NO.:291	3	0.54	0.57	0.92	0.57	58	0.45	19	4.78
SEQ ID NO.:292	4	0.53	0.59	0.92	0.57	58	0.45	19	4.65
SEQ ID NO.:293	5	0.54	0.58	0.92	0.57	58	0.45	18	4.10
SEQ ID NO.:294	7	0.55	0.60	0.92	0.57	58	0.47	20	3.75
SEQ ID NO.:295	14	0.55	0.61	0.92	0.57	58	0.45	19	3.68
SEQ ID NO.:296	20	0.54	0.58	0.92	0.57	58	0.50	18	4.30
SEQ ID NO.:297	22	0.54	0.58	0.92	0.57	58	0.45	18	4.38
SEQ ID NO.:298	23	0.54	0.57	0.92	0.57	58	0.40	20	4.38
SEQ ID NO.:299	25	0.54	0.57	0.92	0.57	58	0.40	21	4.45
SEQ ID NO.:300	27	0.55	0.61	0.92	0.57	58	0.50	19	3.88
SEQ ID NO.:301	30	0.54	0.61	0.92	0.57	58	0.47	19	4.38
SEQ ID NO.:302	33	0.54	0.55	0.92	0.57	58	0.41	19	4.80
SEQ ID NO.:303	34	0.54	0.61	0.92	0.57	58	0.45	18	4.20
SEQ ID NO.:304	36	0.55	0.59	0.92	0.57	58	0.57	19	4.15
SEQ ID NO.:305	37	0.55	0.59	0.92	0.57	58	0.47	19	3.95
SEQ ID NO.:306	39	0.55	0.59	0.92	0.57	58	0.52	19	3.80
SEQ ID NO.:307	41	0.55	0.59	0.92	0.57	58	0.41	20	4.43
SEQ ID NO.:308	43	0.54	0.58	0.92	0.57	58	0.34	19	4.85
SEQ ID NO.:309	45	0.54	0.59	0.92	0.57	58	0.45	20	4.58
SEQ ID NO.:310	46	0.54	0.59	0.92	0.57	58	0.45	21	4.65
SEQ ID NO.:311	53	0.56	0.60	0.92	0.57	58	0.47	19	3.58
SEQ ID NO.:312	54	0.55	0.59	0.92	0.57	58	0.45	18	4.00
SEQ ID NO.:313	56	0.54	0.60	0.92	0.57	58	0.45	18	4.58

FIG. 31a-2

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D	I	L	L	T	Q	S	P	V	I	L	S	V	S	P	G	E	R	V	S
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	K	V	T
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	K	V	T
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	K	V	T
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	K	V	T
D	I	L	L	T	Q	S	P	D	F	Q	S	V	T	P	K	E	K	V	T
D	I	L	L	T	Q	S	P	D	F	Q	S	V	T	P	K	E	K	V	T
D	I	L	L	T	Q	S	P	D	F	Q	S	V	T	P	K	E	K	V	T
D	I	L	L	T	Q	S	P	A	F	L	S	V	T	P	G	E	K	V	T
D	I	L	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	V	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	A	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	A	T	L	S	L	S	P	G	E	R	V	T
D	I	L	L	T	Q	S	P	G	T	L	S	L	S	P	G	E	R	V	T

FIG. 31a-3

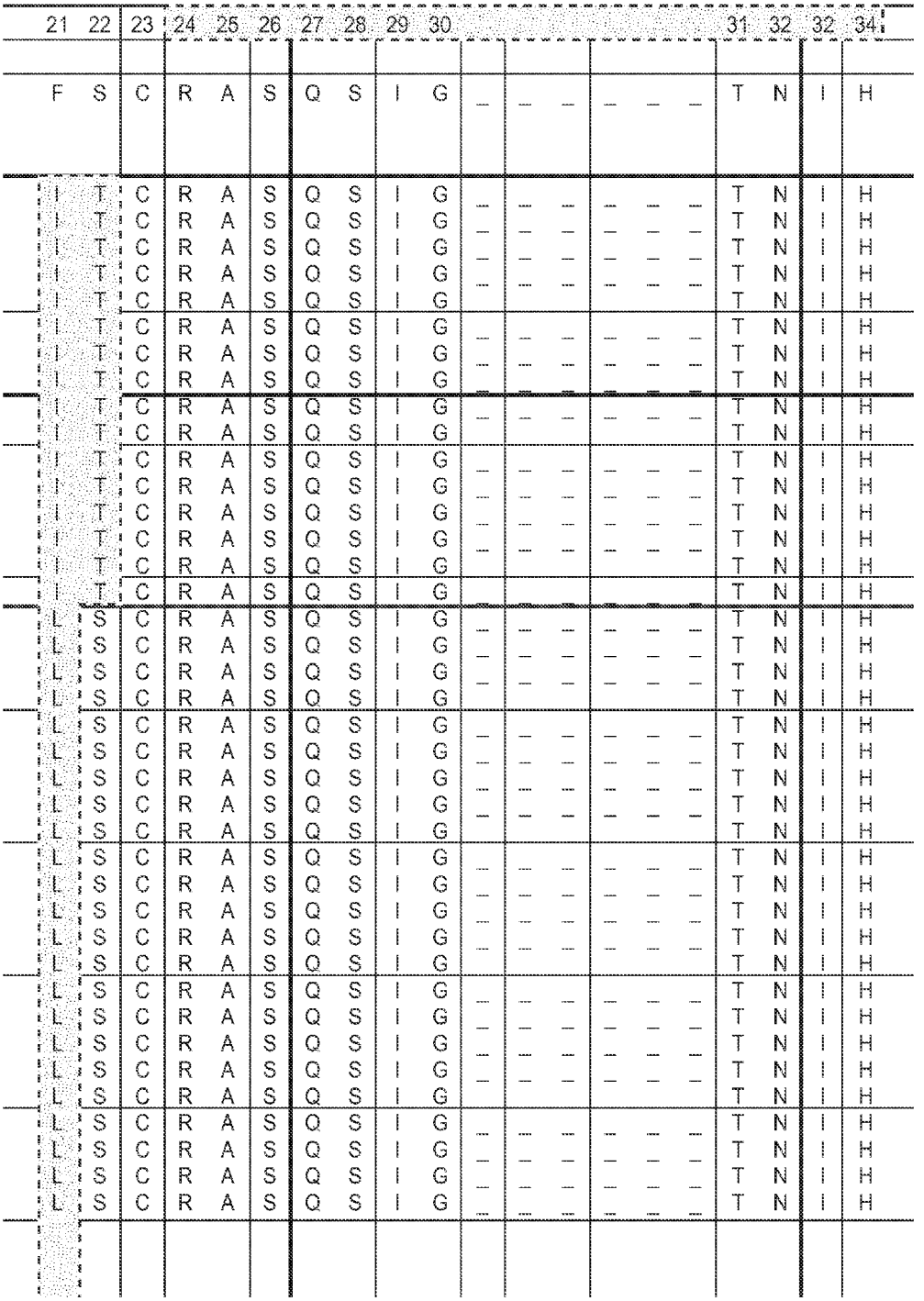


FIG. 31a-4

35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
W	Y	Q	Q	R	T	N	G	S	P	R	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S

FIG. 31a-5

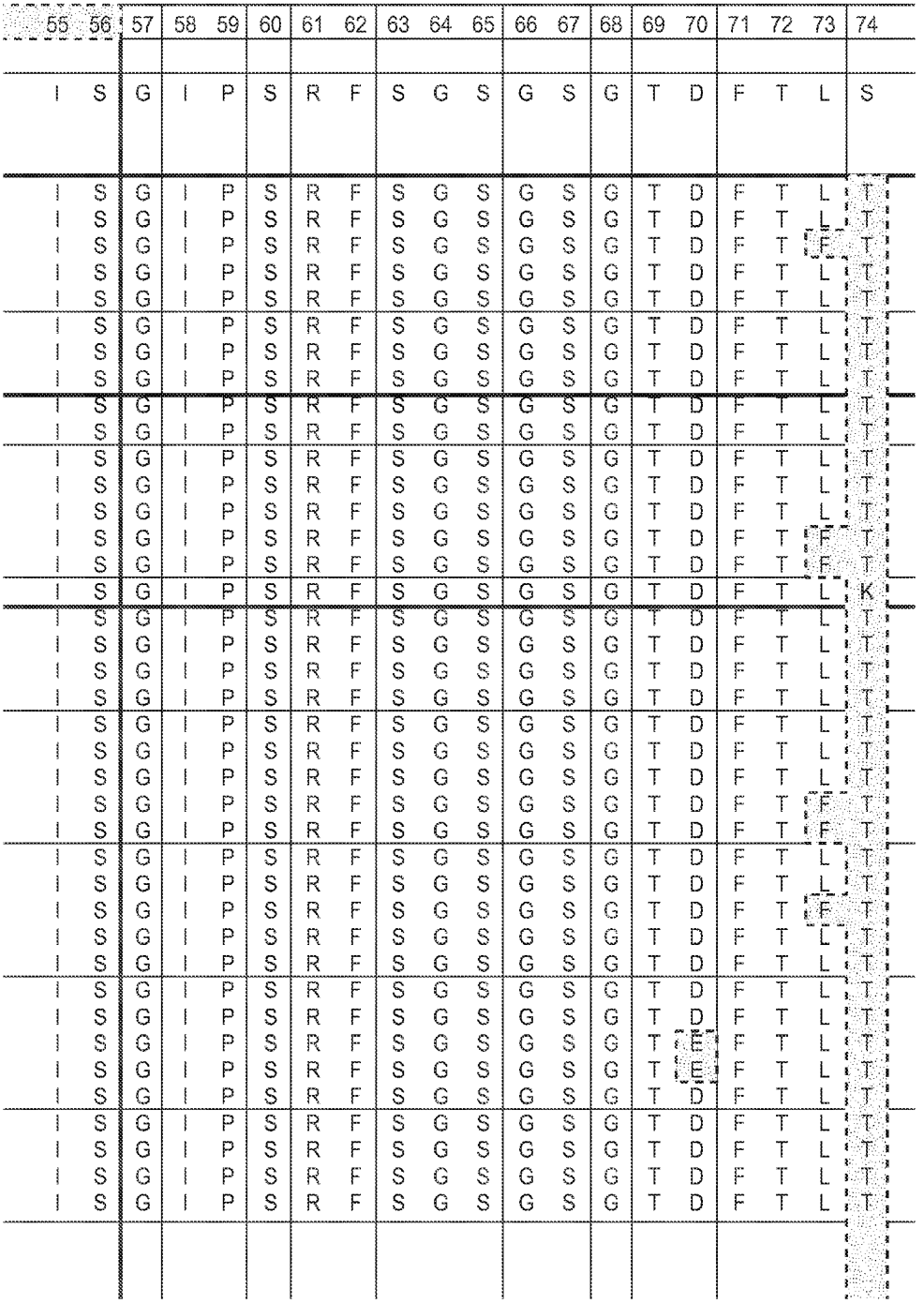


FIG. 31a-6

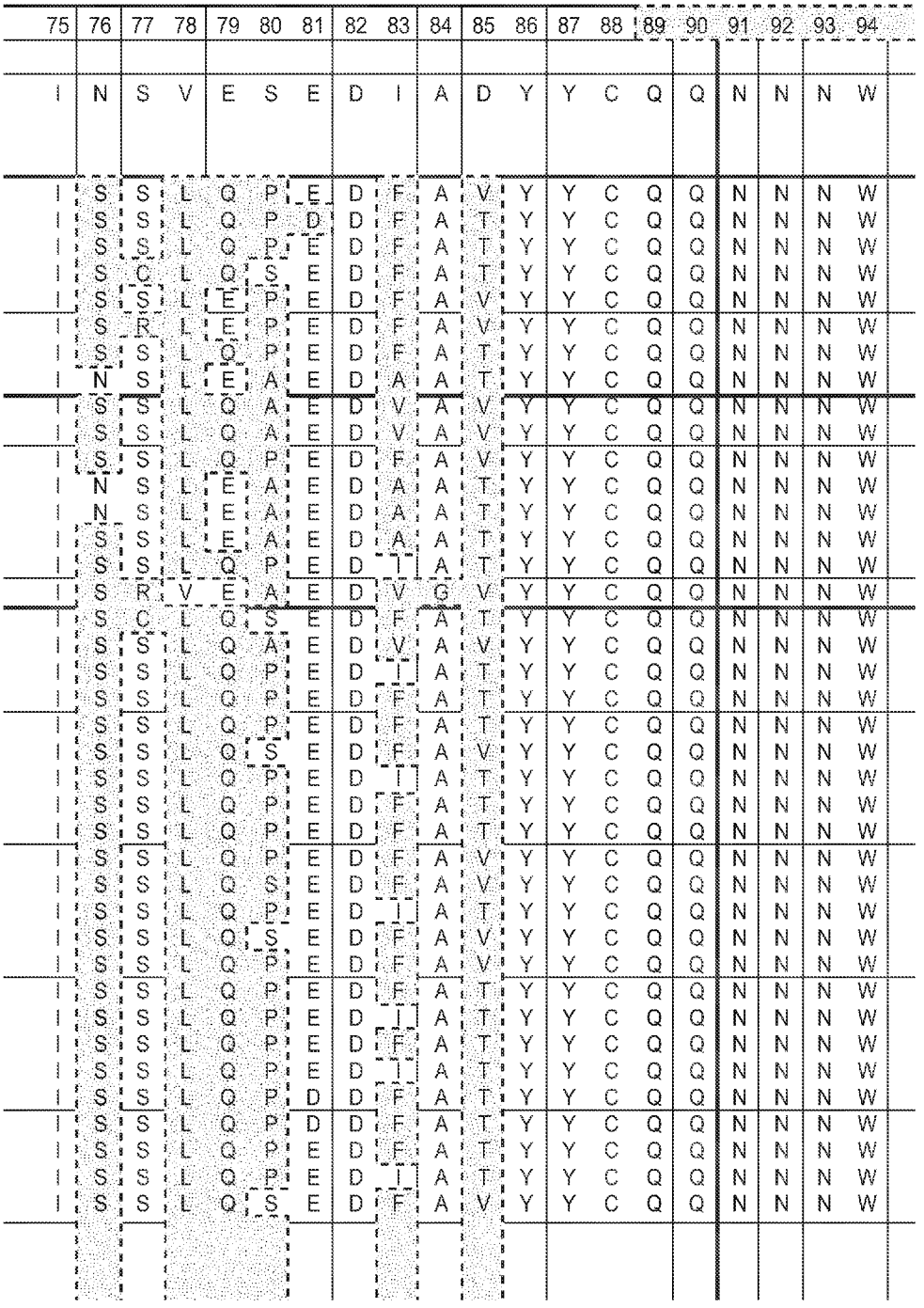
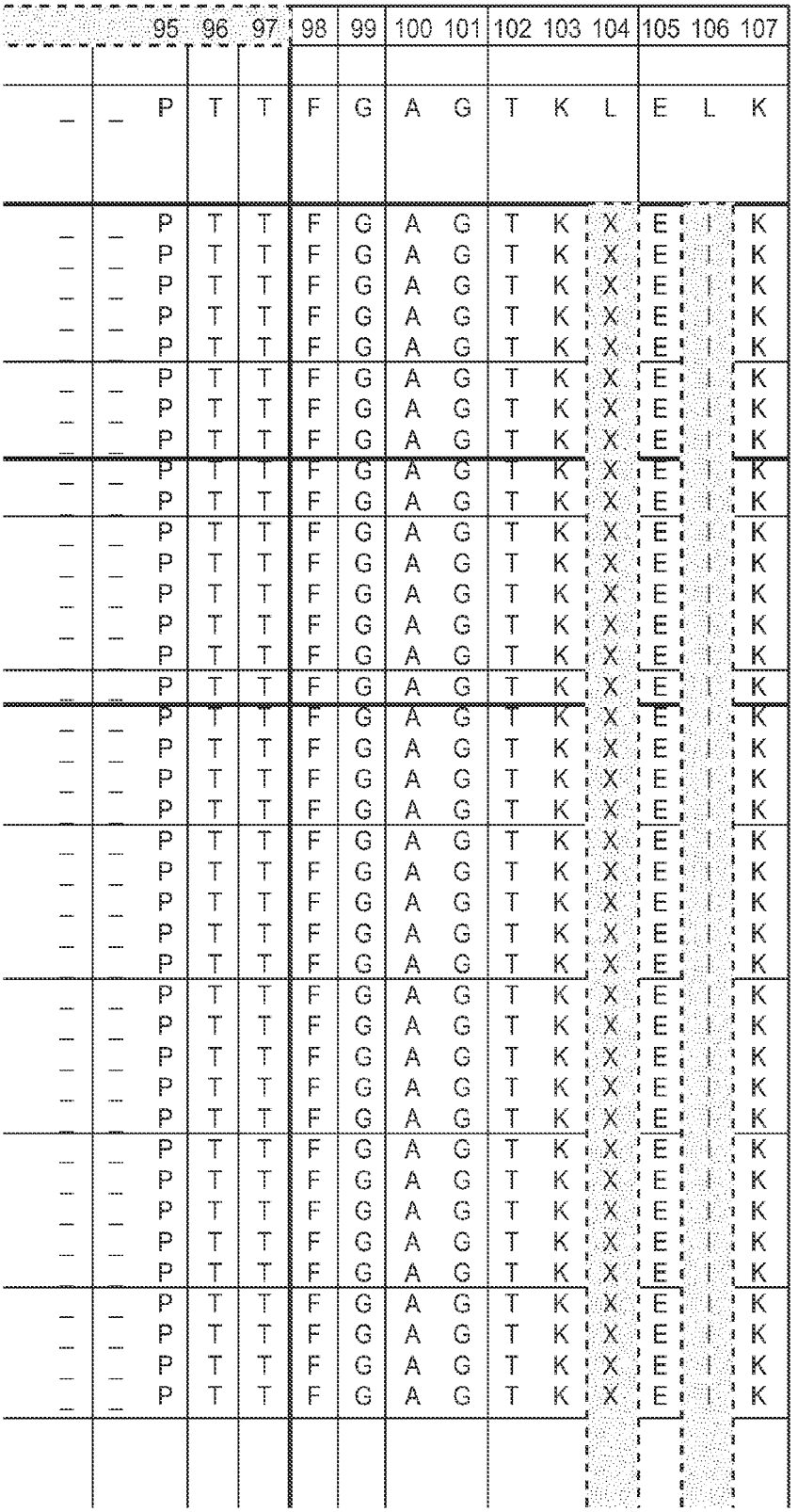

75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94
I N S V E S E D I A D Y Y C Q Q N N N W

FIG. 31a-7



SEQ ID NO.:314	6	0.53	0.57	0.92	0.57	58	0.45	19	4.93
SEQ ID NO.:315	31	0.54	0.58	0.92	0.57	58	0.45	19	4.55
SEQ ID NO.:316	9	0.55	0.61	0.92	0.57	58	0.59	20	4.58
SEQ ID NO.:317	10	0.54	0.61	0.92	0.57	58	0.50	19	4.88
SEQ ID NO.:318	11	0.55	0.61	0.92	0.57	58	0.50	19	4.78
SEQ ID NO.:319	26	0.55	0.62	0.92	0.57	58	0.52	18	4.08
SEQ ID NO.:320	42	0.55	0.60	0.92	0.57	58	0.67	20	4.95
SEQ ID NO.:321	16	0.53	0.58	0.92	0.57	58	0.45	19	4.88
SEQ ID NO.:322	18	0.53	0.60	0.92	0.57	58	0.48	19	5.03
SEQ ID NO.:323	8	0.53	0.58	0.92	0.57	58	0.67	17	5.48
SEQ ID NO.:324	13	0.54	0.58	0.92	0.57	58	0.67	17	5.38
SEQ ID NO.:325	19	0.54	0.58	0.92	0.57	58	0.67	18	5.18
SEQ ID NO.:326	50	0.53	0.57	0.92	0.57	58	0.67	18	5.55
SEQ ID NO.:327	24	0.54	0.57	0.92	0.57	58	0.47	19	5.28
SEQ ID NO.:328	29	0.53	0.56	0.92	0.57	58	0.47	19	5.38
SEQ ID NO.:329	52	0.53	0.54	0.91	0.53	54	0.50	21	8.88
SEQ ID NO.:330	44	0.51	0.51	0.92	0.55	56	0.43	20	8.58

FIG. 31a-8

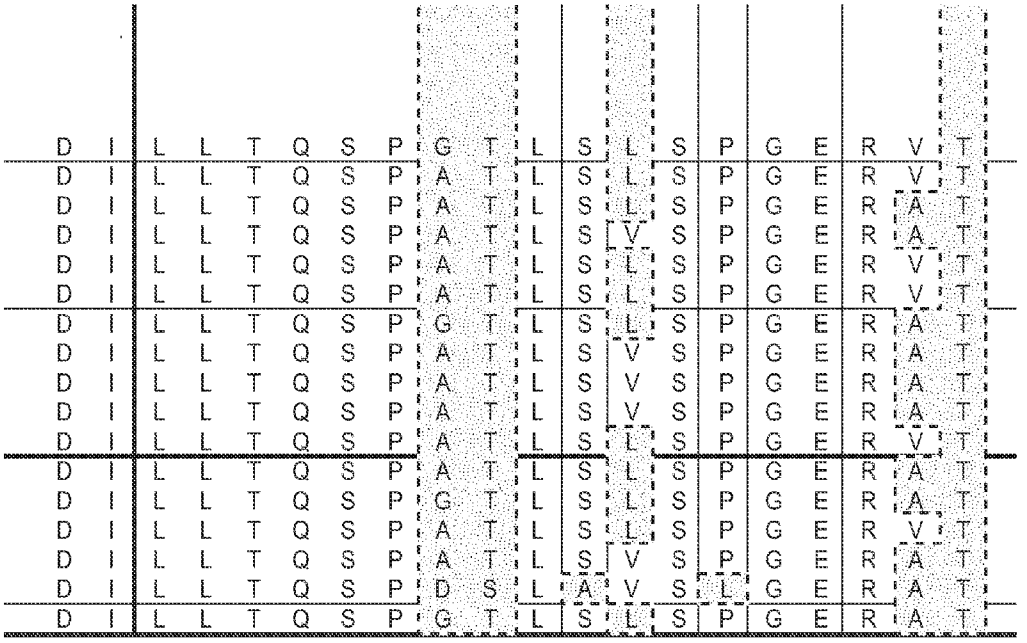

D I L L T Q S P G T L S L S P G E R V T
D I L L T Q S P A T L S L S P G E R V T
D I L L T Q S P A T L S L S P G E R A T
D I L L T Q S P A T L S V S P G E R A T
D I L L T Q S P A T L S L S P G E R V T
D I L L T Q S P A T L S L S P G E R V T
D I L L T Q S P G T L S L S P G E R A T
D I L L T Q S P A T L S V S P G E R A T
D I L L T Q S P A T L S V S P G E R A T
D I L L T Q S P A T L S V S P G E R A T
D I L L T Q S P A T L S L S P G E R V T
D I L L T Q S P A T L S L S P G E R A T
D I L L T Q S P G T L S L S P G E R A T
D I L L T Q S P A T L S L S P G E R V T
D I L L T Q S P A T L S V S P G E R A T
D I L L T Q S P D S L A V S L G E R A T
D I L L T Q S P G T L S L S P G E R A T

FIG. 31a-9

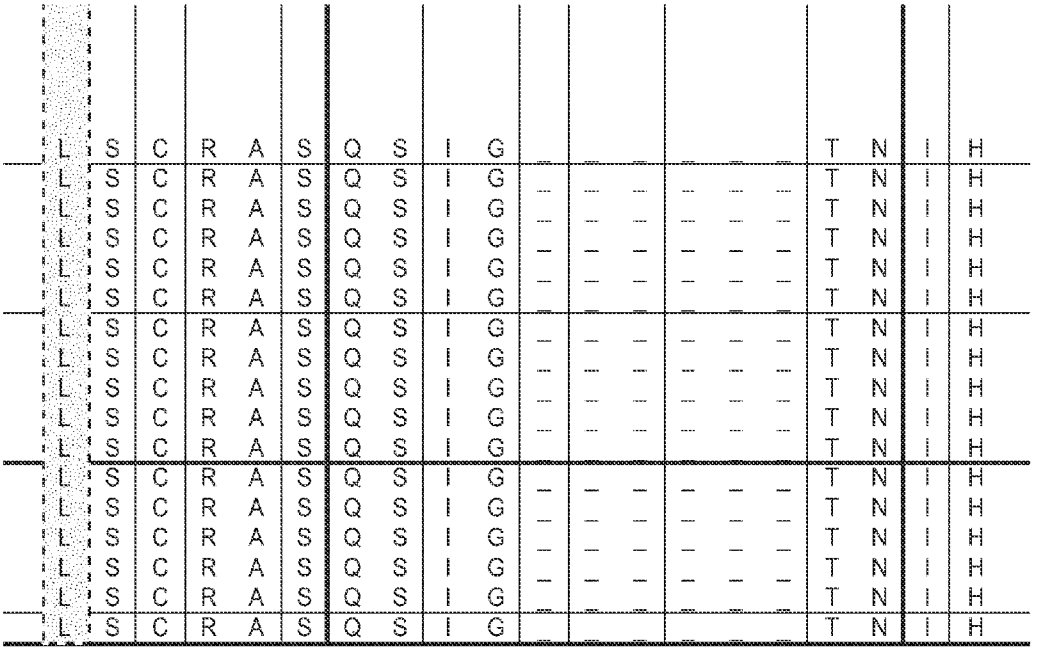


FIG. 31a-10

W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	
W	Y	Q	Q	K	P	D	Q	S	P	K	L	L	I	K	Y	A	S	E	S	

FIG. 31a-11

I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	E	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	F	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	F	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	F	T
I	S	G	I	P	S	R	F	S	G	S	G	S	G	T	D	F	T	L	K

FIG. 31a-12

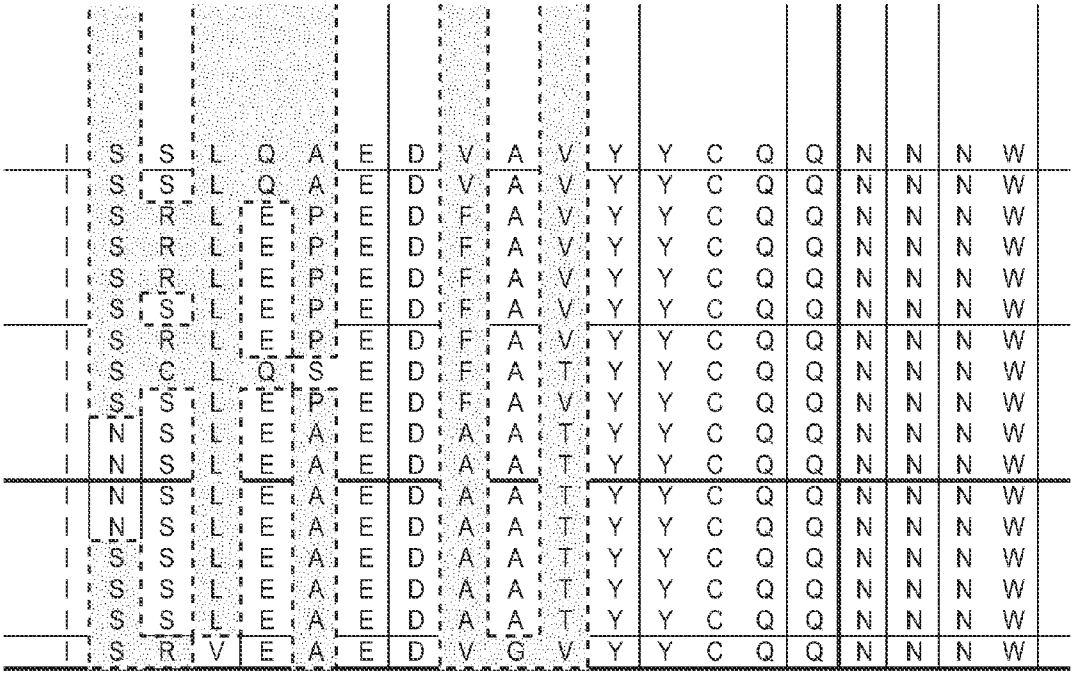


FIG. 31a-13

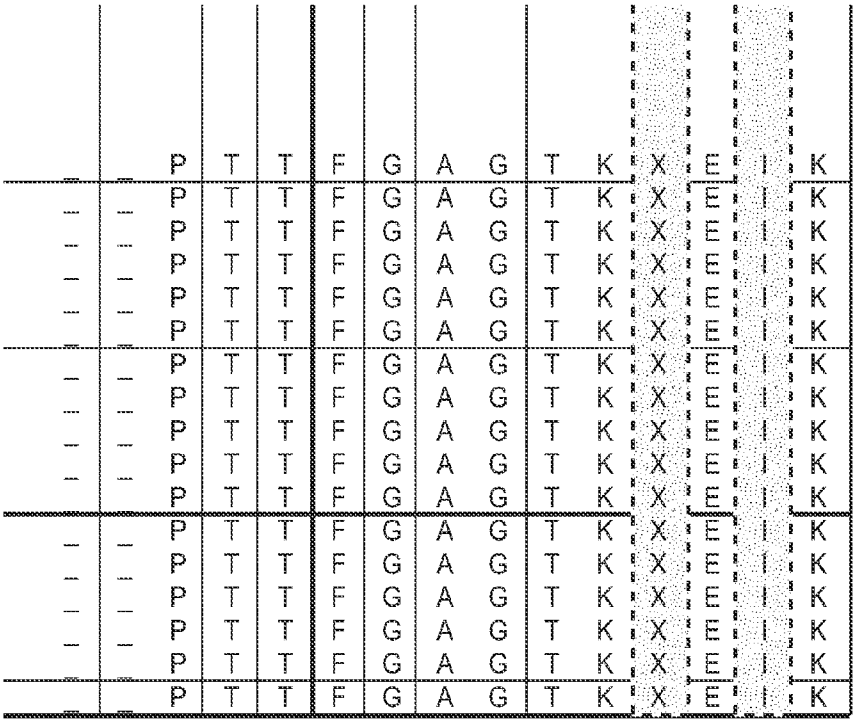


FIG. 31a-14

FIG. 31a-1	FIG. 31a-2	FIG. 31a-3	FIG. 31a-4	FIG. 31a-5	FIG. 31a-6	FIG. 31a-7
FIG. 31a-8	FIG. 31a-9	FIG. 31a-10	FIG. 31a-11	FIG. 31a-12	FIG. 31a-13	FIG. 31a-14

FIG. 31a

FIG. 31b-1

C225 VH HEC Calculation 2

(WT) SEQ ID NO.:235	Iter	Structural Consensus	Structural Precedence	HSC	HSS	N_9max	Kabat FRH	Kabat Muts	Kabat WT Cluster
SEQ ID NO.:331	1	0.45	0.52	0.82	0.44	52	0.48	24	1.80
SEQ ID NO.:332	2	0.45	0.51	0.82	0.44	52	0.48	23	2.00
SEQ ID NO.:333	15	0.45	0.51	0.82	0.44	52	0.48	24	2.20
SEQ ID NO.:334	22	0.45	0.52	0.82	0.44	52	0.48	25	2.00
SEQ ID NO.:335	40	0.43	0.52	0.82	0.44	52	0.60	28	4.80
SEQ ID NO.:336	3	0.45	0.51	0.81	0.41	48	0.58	27	
SEQ ID NO.:337	4	0.49	0.51	0.81	0.40	47	0.47	23	5.19
SEQ ID NO.:338	6	0.49	0.51	0.81	0.40	47	0.47	21	4.81
SEQ ID NO.:339	17	0.48	0.50	0.81	0.40	47	0.34	22	5.42
SEQ ID NO.:340	21	0.49	0.51	0.81	0.40	47	0.47	22	5.74
SEQ ID NO.:341	30	0.49	0.51	0.81	0.40	47	0.47	22	4.84
SEQ ID NO.:342	43	0.49	0.51	0.81	0.40	47	0.47	22	5.16
SEQ ID NO.:343	5	0.50	0.53	0.81	0.39	46	0.46	20	4.45
SEQ ID NO.:344	7	0.50	0.53	0.81	0.39	46	0.48	22	4.84
SEQ ID NO.:345	8	0.49	0.52	0.81	0.39	46	0.46	21	4.48
SEQ ID NO.:346	25	0.48	0.52	0.81	0.39	46	0.35	21	5.06
SEQ ID NO.:347	45	0.49	0.53	0.81	0.39	46	0.39	20	5.03
SEQ ID NO.:348	20	0.49	0.53	0.81	0.39	46	0.48	21	5.68
SEQ ID NO.:349	24	0.49	0.53	0.81	0.39	46	0.41	20	5.32
SEQ ID NO.:350	37	0.49	0.52	0.81	0.39	46	0.39	21	5.35
SEQ ID NO.:351	33	0.49	0.51	0.81	0.40	47	0.40	21	5.68
SEQ ID NO.:352	32	0.49	0.51	0.81	0.39	46	0.46	23	6.26
SEQ ID NO.:353	11	0.47	0.45	0.81	0.40	47	0.36	23	7.32
SEQ ID NO.:354	16	0.46	0.45	0.81	0.40	47	0.32	24	7.35
SEQ ID NO.:355	29	0.46	0.47	0.81	0.40	47	0.32	25	7.45
SEQ ID NO.:356	39	0.47	0.47	0.81	0.40	47	0.45	26	7.81
SEQ ID NO.:357	34	0.48	0.47	0.81	0.40	47	0.40	23	6.16
SEQ ID NO.:358	46	0.47	0.47	0.81	0.40	47	0.34	24	6.19
SEQ ID NO.:359	12	0.47	0.49	0.81	0.38	45	0.47	23	7.55
SEQ ID NO.:360	26	0.47	0.49	0.81	0.38	45	0.47	24	7.58
SEQ ID NO.:361	28	0.47	0.49	0.81	0.38	45	0.38	22	7.19
SEQ ID NO.:362	47	0.47	0.49	0.81	0.38	45	0.33	23	7.23
SEQ ID NO.:363	9	0.49	0.54	0.81	0.40	47	0.60	22	6.97
SEQ ID NO.:364	27	0.49	0.53	0.81	0.40	47	0.47	24	7.35
SEQ ID NO.:365	18	0.48	0.54	0.81	0.40	47	0.38	22	8.00
SEQ ID NO.:366	23	0.48	0.54	0.81	0.40	47	0.36	23	8.03
SEQ ID NO.:367	41	0.48	0.54	0.81	0.40	47	0.55	22	8.03
SEQ ID NO.:368	10	0.47	0.47	0.82	0.44	52	0.37	24	1.50
SEQ ID NO.:369	35	0.47	0.48	0.82	0.44	52	0.42	24	1.83

FIG. 31b-2

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Q	V	Q	L	K	Q	S	G	P	G	L	V	Q	P	S	Q	S	L	S	I
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	T	L	R	E	S	G	P	A	L	V	K	P	T	Q	T	L	T	L
Q	V	T	L	K	E	S	G	P	V	L	V	K	P	T	E	T	L	T	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	Q	T	L	S	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	S	E	T	L	S	L
Q	V	Q	L	Q	E	S	G	P	G	L	V	K	P	S	Q	T	L	S	L

FIG. 31b-3

21	22	23	24	25	26	27	28	29	30	31		32	33	34	35	36	37	38
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_ _	Y	G	V	H	W	V	R

FIG. 31b-4

39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54			55
Q	S	P	G	K	G	L	E	W	L	G	V	I	W	S	G	_	_	_ G
Q	A	P	G	K	G	L	E	W	V	S

FIG. 31b-5

56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
N	T	D	Y	N	T	P	F	T	S	R	L	S	I	N	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	S	K	D	N	S	K
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	T	K	D	N	S	T
N	T	D	Y	N	T	S	L	K	S	R	L	T	I	T	K	D	N	S	T

FIG. 31b-6

76	77	78	79	80	81	82	82a	82b	82c	83	84	85	86	87	88	89	90	91	92
S	Q	V	F	F	K	M	N	S	L	Q	S	N	D	T	A	I	Y	Y	C
S	Q	V	V	L	T	M	T	N	M	D	P	V	D	T	A	T	Y	Y	C
S	Q	V	V	L	T	M	T	N	M	D	P	V	D	T	A	T	Y	Y	C
S	Q	V	V	L	T	M	T	N	M	D	P	V	D	T	A	T	Y	Y	C
S	Q	V	V	L	T	M	T	N	M	D	P	V	D	T	A	T	Y	Y	C
S	Q	V	V	L	T	M	T	N	M	D	P	V	D	T	A	T	Y	Y	C
S	Q	V	S	L	K	L	S	S	V	T	A	A	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	R	A	G	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	I	V	Y	L	Q	M	N	S	L	K	T	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	Q	M	G	S	L	R	A	E	D	M	A	V	Y	Y	C
S	T	V	Y	L	Q	W	S	S	L	K	A	S	D	T	A	M	Y	Y	C
S	T	V	Y	L	Q	W	S	S	L	K	A	S	D	T	A	M	Y	Y	C
S	T	V	Y	L	Q	I	S	S	L	K	A	E	D	M	A	M	Y	Y	C
S	T	V	Y	L	Q	I	S	S	L	K	A	E	D	M	A	M	Y	Y	C
S	T	V	Y	L	Q	I	C	S	L	K	A	E	D	T	A	V	Y	Y	C
S	T	V	Y	L	Q	I	C	S	L	K	A	E	D	T	A	V	Y	Y	C
S	Q	V	V	L	T	M	N	N	L	R	A	E	G	T	A	A	Y	Y	C
S	Q	V	V	L	T	M	N	N	L	R	A	E	G	T	A	A	Y	Y	C
S	Q	V	V	L	T	M	N	N	L	R	A	E	G	T	A	A	Y	Y	C
S	Q	V	V	L	T	M	N	N	L	R	A	E	G	T	A	A	Y	Y	C
S	Q	V	S	L	Q	L	N	S	V	T	P	E	D	T	A	V	Y	Y	C
S	Q	V	S	L	Q	L	N	S	V	T	P	E	D	T	A	V	Y	Y	C
S	Q	V	S	L	K	L	S	S	V	T	A	V	D	T	A	V	Y	Y	C
S	Q	V	S	L	K	L	S	S	V	T	A	V	D	T	A	V	Y	Y	C
S	Q	V	S	L	K	L	S	S	V	T	A	A	D	T	A	V	Y	Y	C
S	T	V	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
S	T	V	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C

FIG. 31b-7

93	94	95	96	97				98	99	100	100a	100b	100c	101	102	103	104	105	106
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T				Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G

FIG. 31b-8

107	108	109	110	111	112	113
T	L	V	T	V	S	A
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S

SEQ ID NO.:370	36	0.47	0.48	0.82	0.44	52	0.37	23	1.50
SEQ ID NO.:371	38	0.47	0.47	0.82	0.44	52	0.33	23	2.17
SEQ ID NO.:372	48	0.47	0.48	0.82	0.44	52	0.42	25	1.83
SEQ ID NO.:373	31	0.44	0.45	0.81	0.42	49	0.31	26	3.50
SEQ ID NO.:374	13	0.46	0.48	0.82	0.40	47	0.49	27	
SEQ ID NO.:375	14	0.47	0.51	0.81	0.39	46	0.46	25	
SEQ ID NO.:376	19	0.49	0.49	0.81	0.40	47	0.49	27	
SEQ ID NO.:377	42	0.48	0.49	0.80	0.36	42	0.55	26	
SEQ ID NO.:378	44	0.51	0.53	0.80	0.35	41	0.71	25	

FIG. 31b-9

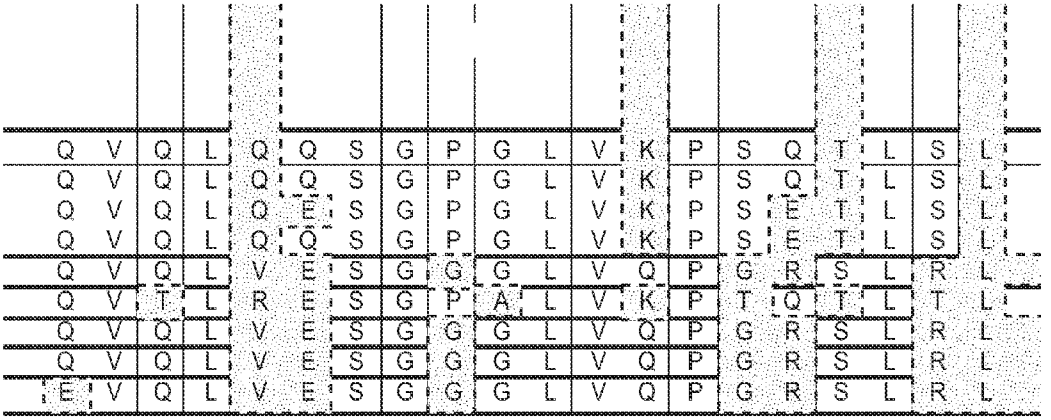


FIG. 31b-10

T	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
S	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
T	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
S	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
S	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R
S	C	T	V	S	G	F	S	L	T	N	_	_	Y	G	V	H	W	V	R

FIG. 31b-11

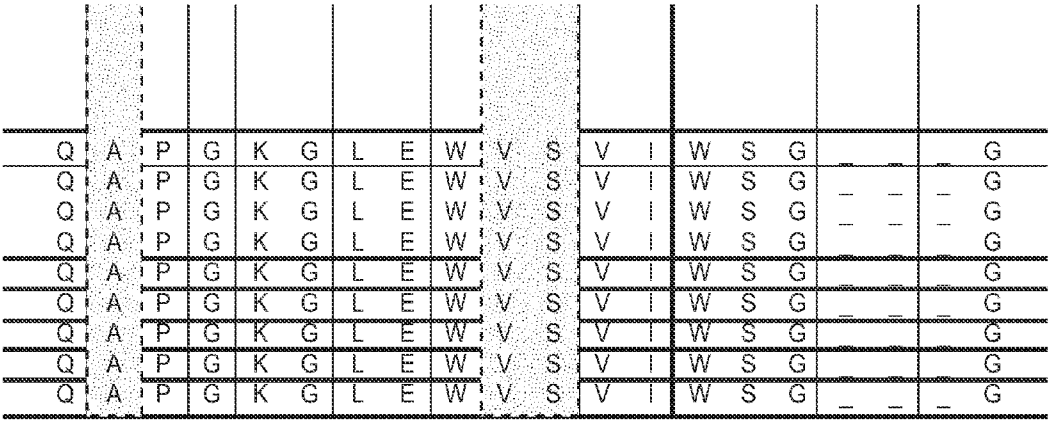

Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G
Q A P G K G L E W V S V I W S G _ _ _ G

FIG. 31b-12

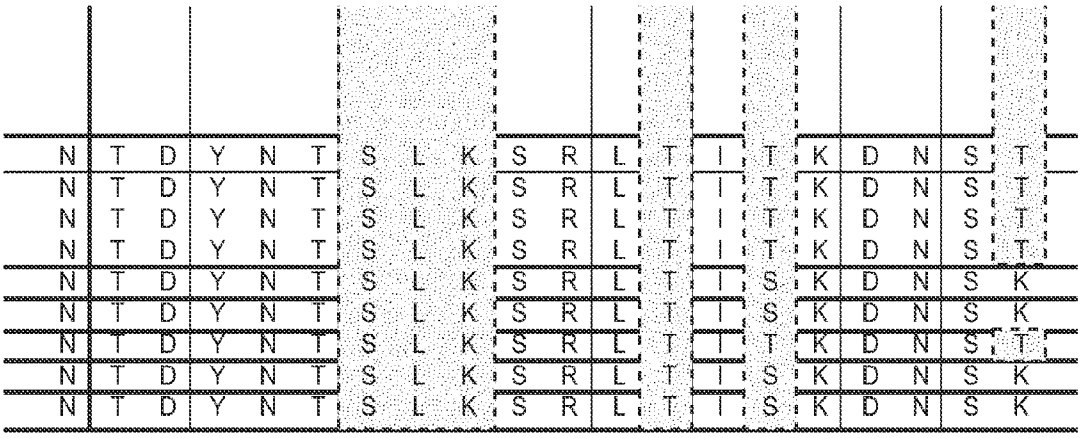

N T D Y N T S L K S R L T I T K D N S T
N T D Y N T S L K S R L T I T K D N S T
N T D Y N T S L K S R L T I T K D N S T
N T D Y N T S L K S R L T I T K D N S T
N T D Y N T S L K S R L T I S K D N S K
N T D Y N T S L K S R L T I S K D N S K
N T D Y N T S L K S R L T I T K D N S T
N T D Y N T S L K S R L T I S K D N S K
N T D Y N T S L K S R L T I S K D N S K

FIG. 31b-13

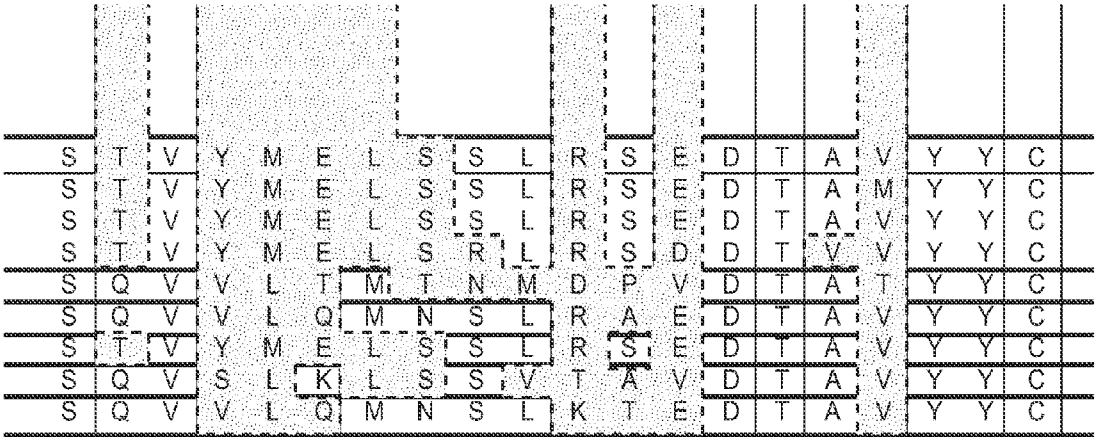

S T V Y M E L S S L R S E D T A V Y Y C
S T V Y M E L S S L R S E D T A M Y Y C
S T V Y M E L S S L R S E D T A V Y Y C
S T V Y M E L S R L R S D D T V V Y Y C
S Q V V L T M T N M D P V D T A T Y Y C
S Q V V L Q M N S L R A E D T A V Y Y C
S T V Y M E L S S L R S E D T A V Y Y C
S Q V S L K L S S V T A V D T A V Y Y C
S Q V V L Q M N S L K T E D T A V Y Y C

FIG. 31b-14

A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G
A	R	A	L	T	_	_	_	Y	Y	D	Y	E	F	A	Y	W	G	Q	G

FIG. 31b-15

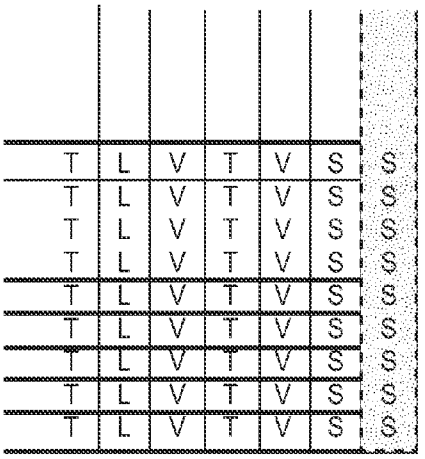


FIG. 31b-16

FIG. 31b-1	FIG. 31b-2	FIG. 31b-3	FIG. 31b-4	FIG. 31b-5	FIG. 31b-6	FIG. 31b-7	FIG. 31b-8
FIG. 31b-9	FIG. 31b-10	FIG. 31b-11	FIG. 31b-12	FIG. 31b-13	FIG. 31b-14	FIG. 31b-15	FIG. 31b-16

FIG. 31b

SEQ ID NO.: 379
DILLTQSP**AT**LS**L**SPGERV**TL**SCRASQSIGTNIHWYQQ**KPDQ**SPK**L**LIKYASESISGI
PSRFSGSGSGTDFTLTI**SSLQP**EDFADYYCQQNNNWPTTFGAGTKLE**I**K

FIG. 32a

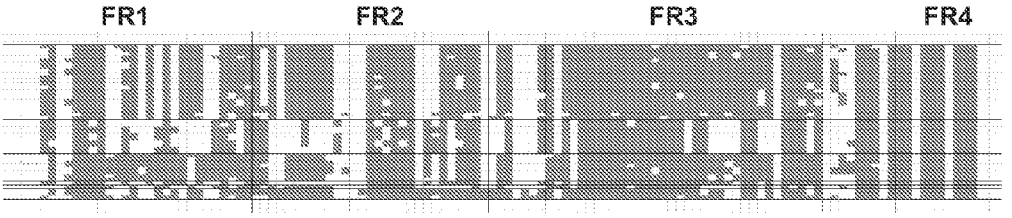


FIG. 32b

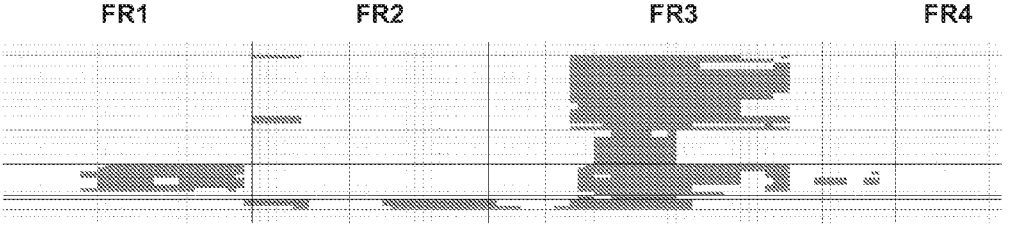


FIG. 32c



FIG. 32d

SEQ ID NO.: 380
DILLTQSPSSLSASVGDRVTITCRASQSIGTNIHWYQQKPDQSPKLLIKYASESISGIP
SRFSGSGSGTDFTLTISSLQAEDVAVYYCQQNNNWPTTFGAGTKLEIK

FIG. 33a

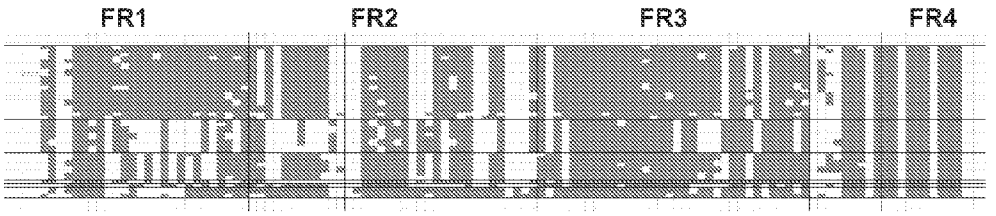


FIG. 33b

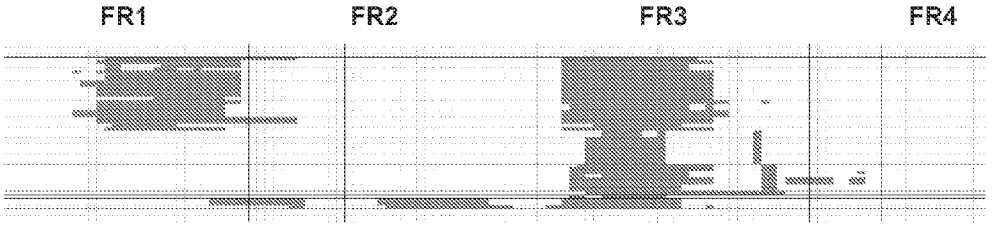


FIG. 33c

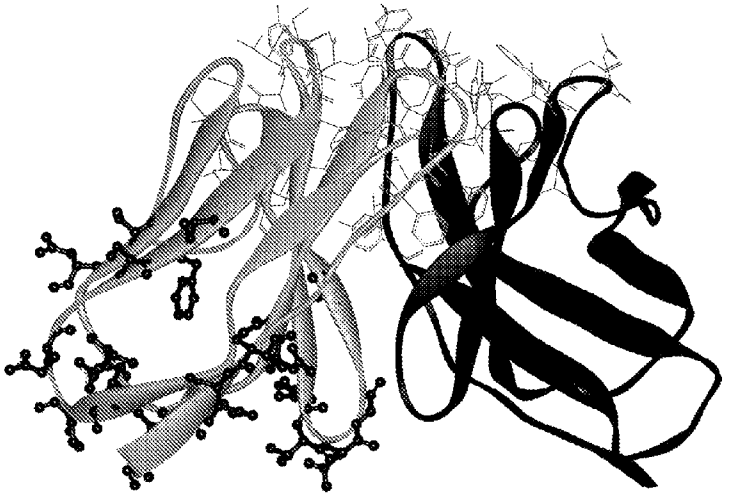


FIG. 33d

SEQ ID NO.: 381
DILLTQSPAFLSVTPGEKVTITCRASQSIGTNIHWYQQKPDQSPKLLIKYASESISGIP
SRFSGSGSGTDFTLTINSLEAEDAATYYCQQNNNWPTTFGAGTKLEIK

FIG. 34a

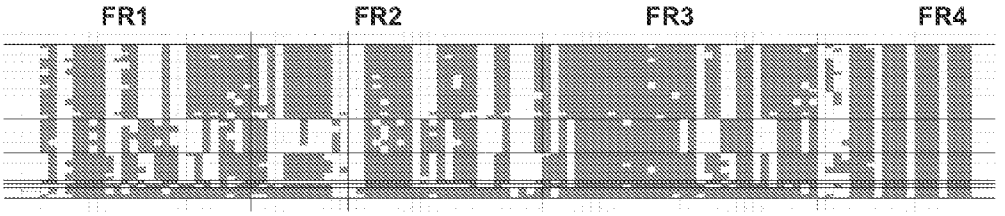


FIG. 34b

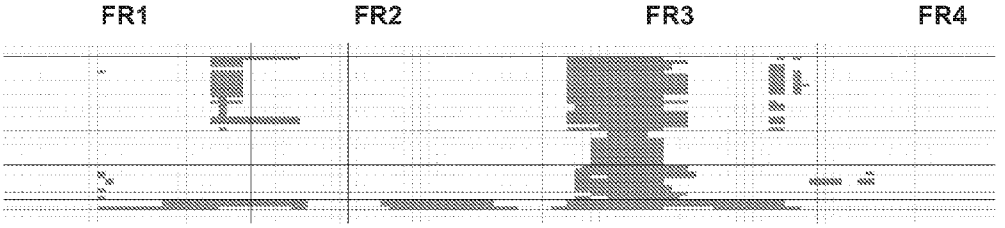


FIG. 34c



FIG. 34d

SEQ ID NO.: 382
QVQL**Q**QSGPGLV**K**PSQ**T**LS**L**TCTVSGFSLTNYGVHWVRQ**A**PGKGLEWLGVIWSG
GNTDYN**T**S**LK**SRLT**I**SKDNSKSQV**VLQ**MNSL**RAE**DTA**V**YYCARALTYYDYEFAYW
GQGTLVTVS**S**

FIG. 35a

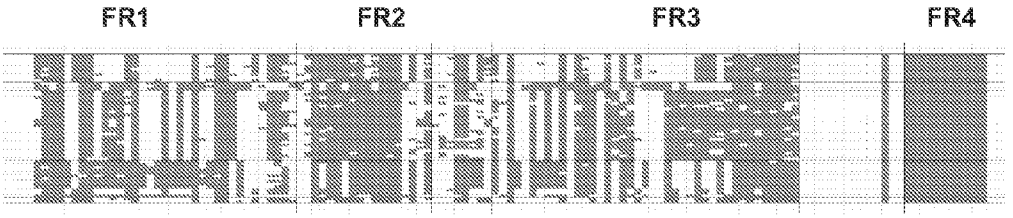


FIG. 35b

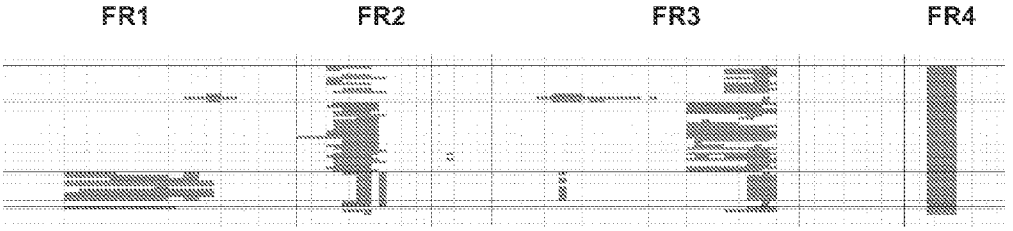


FIG. 35c

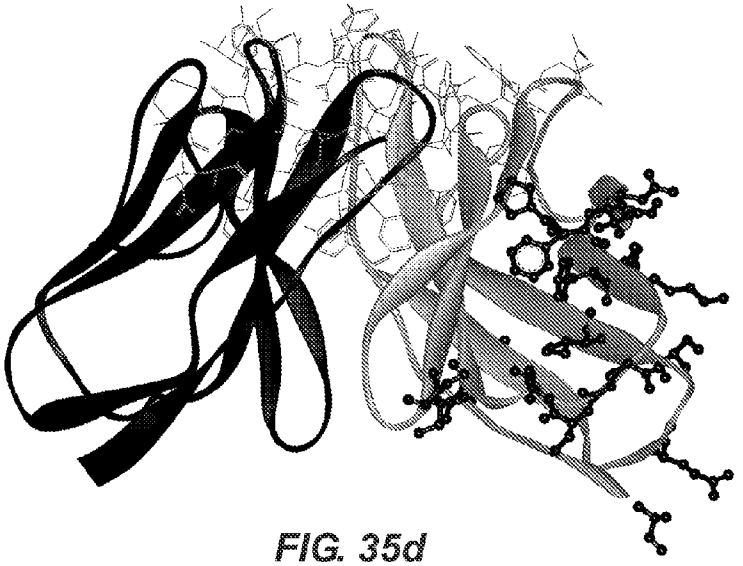


FIG. 35d

SEQ ID NO.: 383
QVQL**Q**QSGPGLV**K**PSQ**T**LS**L**TCTVSGFSLTNYGVHWVRQ**A**PGKGLEWLGVIWSG
GNTDYNT**SLK**SRL**T**I**S**KDNSKSQV**VLT**MT**NMDPV**DTATYYCARALTYYDYEFAYWG
QGTLVTVS**S**

FIG. 36a

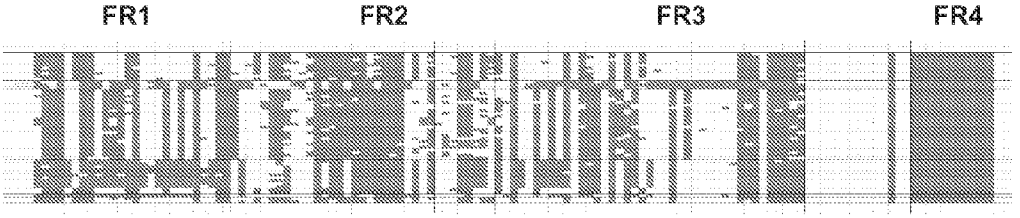


FIG. 36b

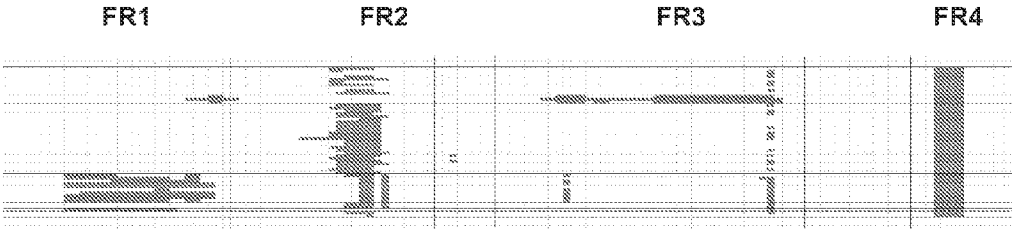


FIG. 36c

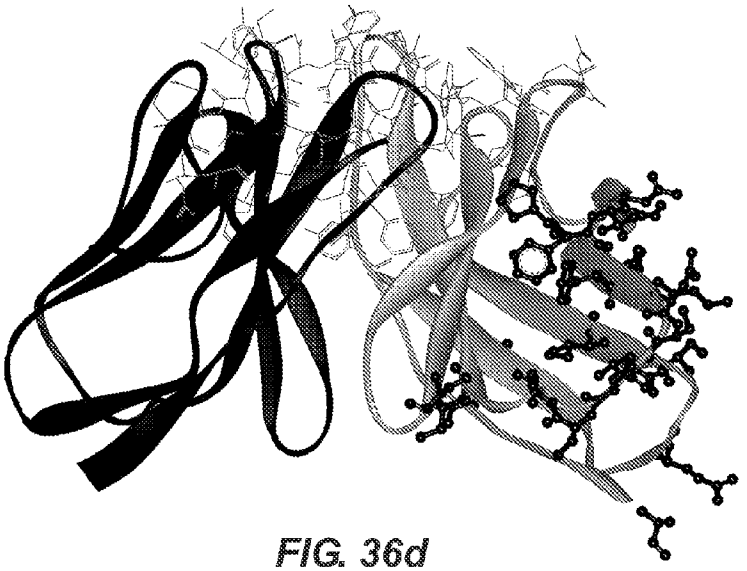


FIG. 36d

SEQ ID NO.: 384
QVQL**Q**QSGPGLV**K**PSQTLSLTCTVSGFSLTNYGVHWVRQ**A**PGKGLEWLGVIWSG
GNTDYNTP**L**TSRLTINKDNSKSQV**VLQ**MNSL**RAE**DTA**V**YYCARALTYYDYEFAYW
GQGTLVTVS**S**

FIG. 37a

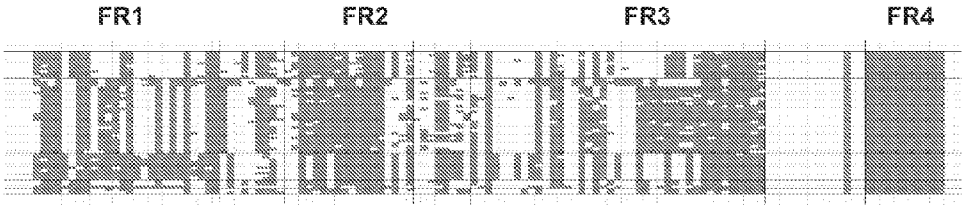


FIG. 37b

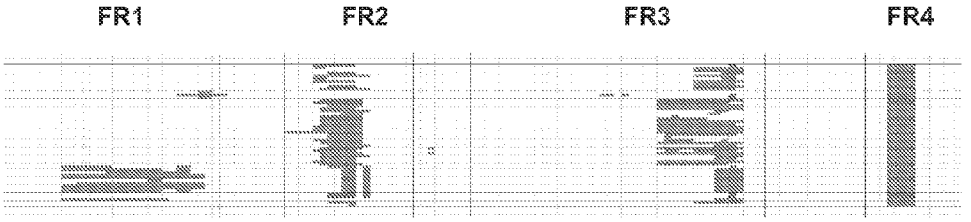


FIG. 37c

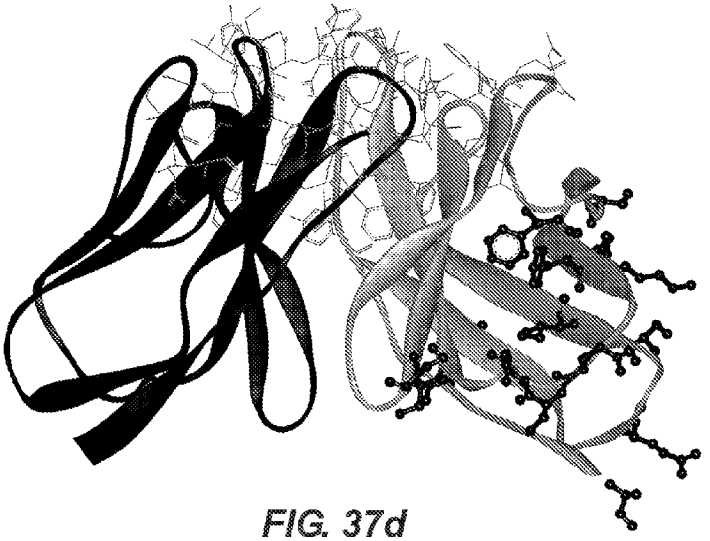


FIG. 37d

SEQ ID NO.: 385
QVQL**Q**QSGPGLV**K**PSQ**T**LSLTCTVSGFSLTNYGVHWVRQ**A**PGKGLEW**V**GVIWSG
GNTDYNT**SVKG**RFT**I**SKDNSKSQV**YLQ**MNSLR**AE**DTA**V**YYCARALTYYDYEFAYW
GQGTLVTVS**S**

FIG. 38a

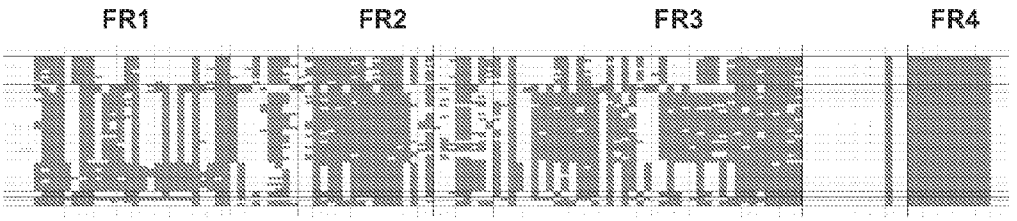


FIG. 38b

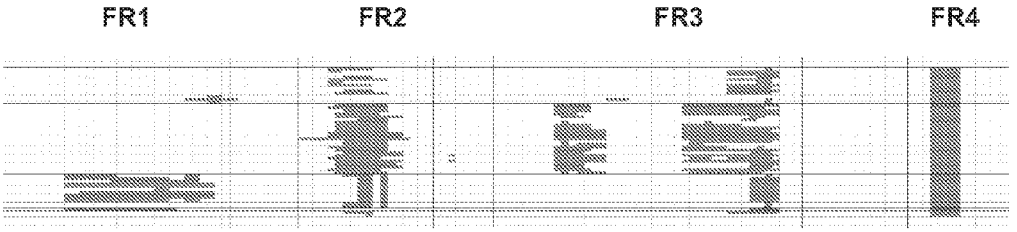


FIG. 38c

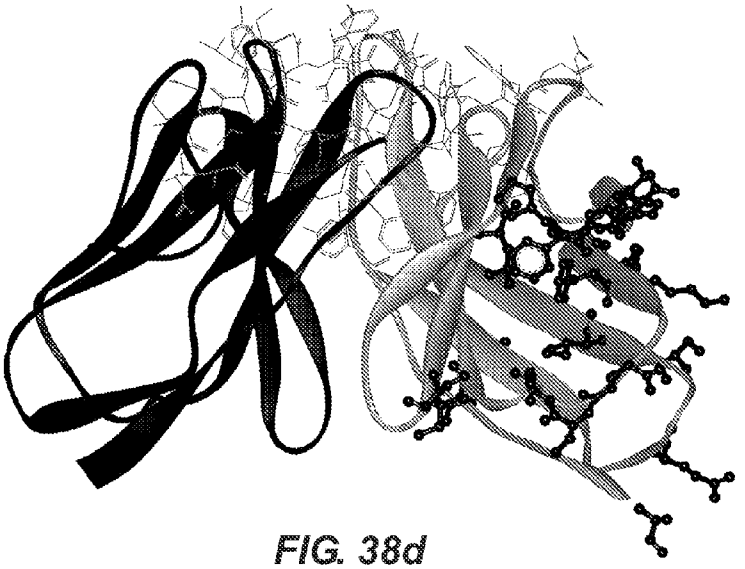


FIG. 38d

SEQ ID NO.: 386
QVQLQQSGPGLVKPSQTLSLTCTVSGFSLTNYGVHWVRQPPGKGLEWIGVIWSG
GNTDYNTSLKSRVTISKDNSKSQVSLKLSSVTAADTAVYYCARALTYYDYEFAYWG
QGTLVTVSS

FIG. 39a

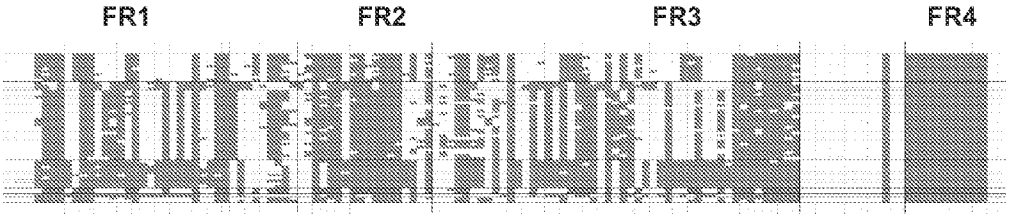


FIG. 39b

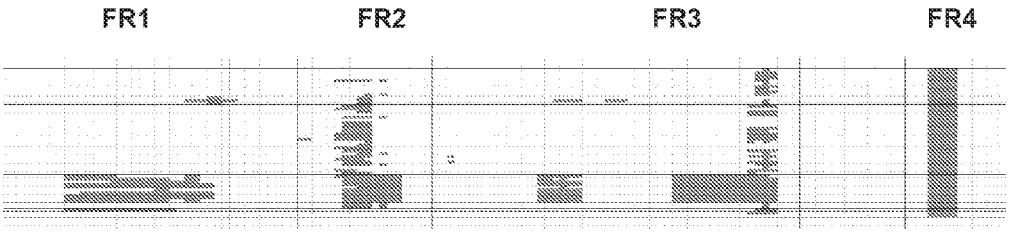


FIG. 39c

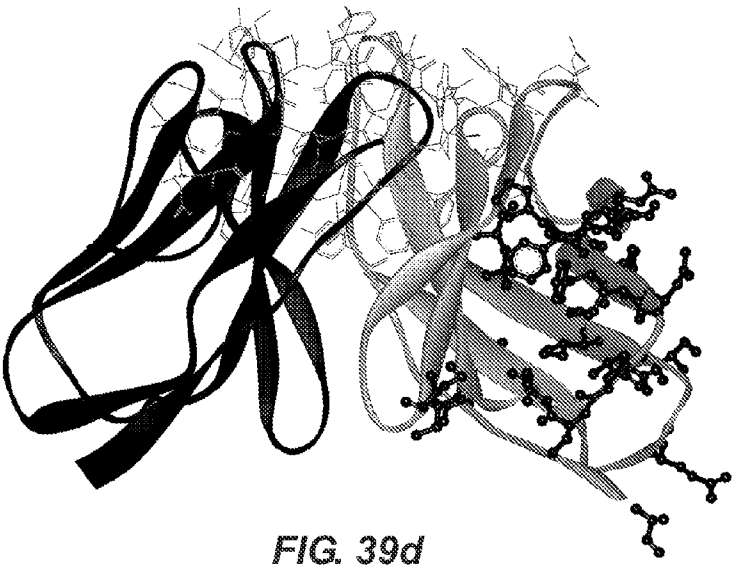


FIG. 39d

SEQ ID NO.: 387
QVQL**VE**SG**G**GLVQP**GR**SL**RLS**C**A**VSGFSLTNYGVHW**V**RQ**A**PGKGLEW**VS**VIWSG
GNTDYNT**SVKG**RFT**I**SKDNSKSTV**YLQ**MNSL**RAE**DTA**V**YYCARALTYYDYEFAYW
GQGTLVTVS**S**

FIG. 40a

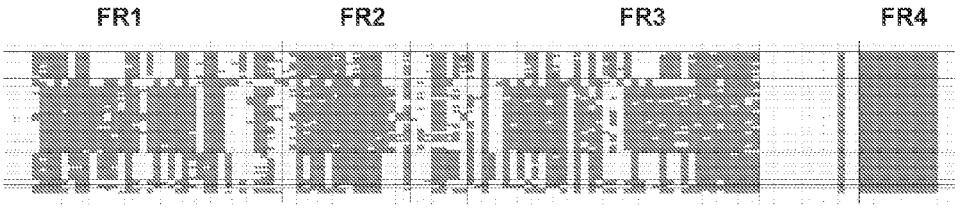


FIG. 40b

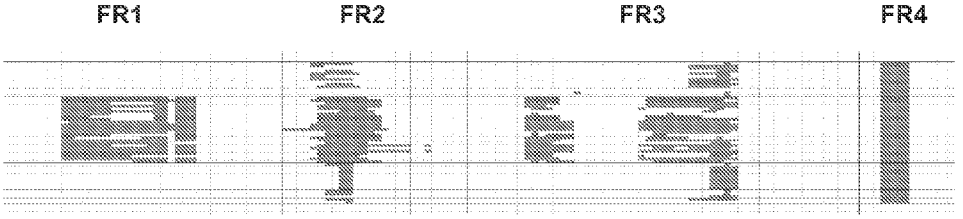


FIG. 40c

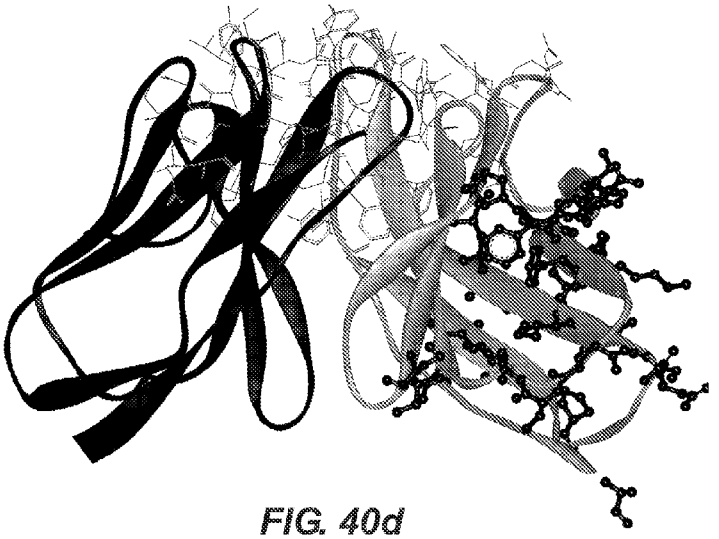


FIG. 40d

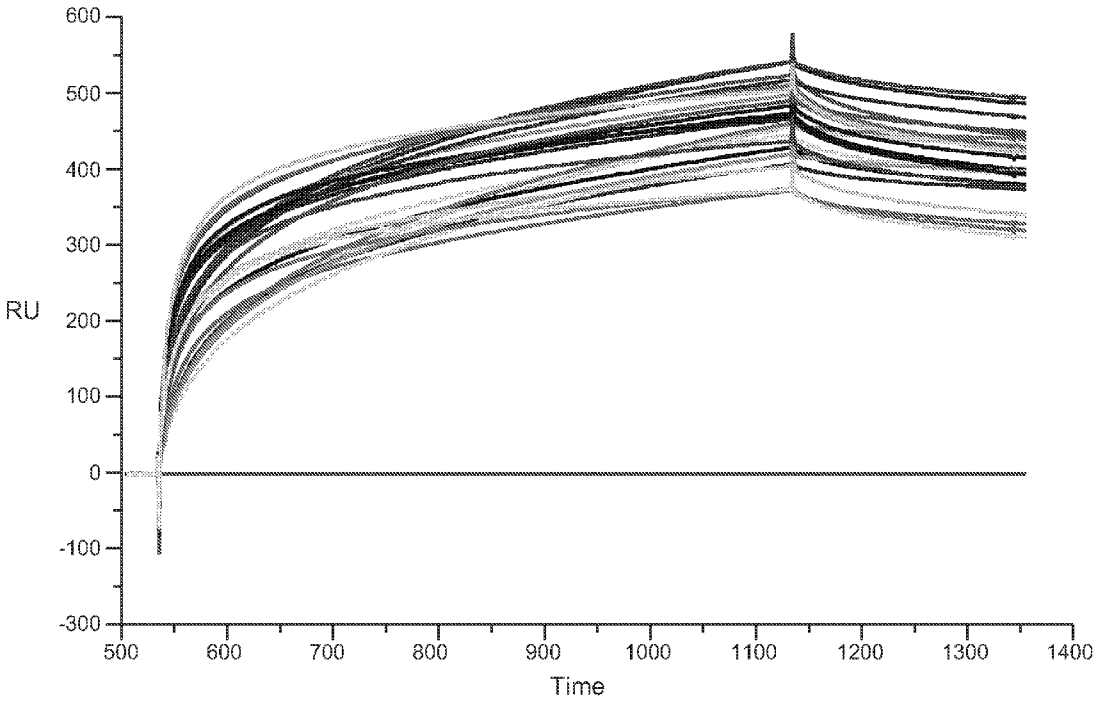

600
500
400
300
200
100
0
-100
-300
RU
500
600
700
800
900
1000
1100
1200
1300
1400
Time

FIG. 41

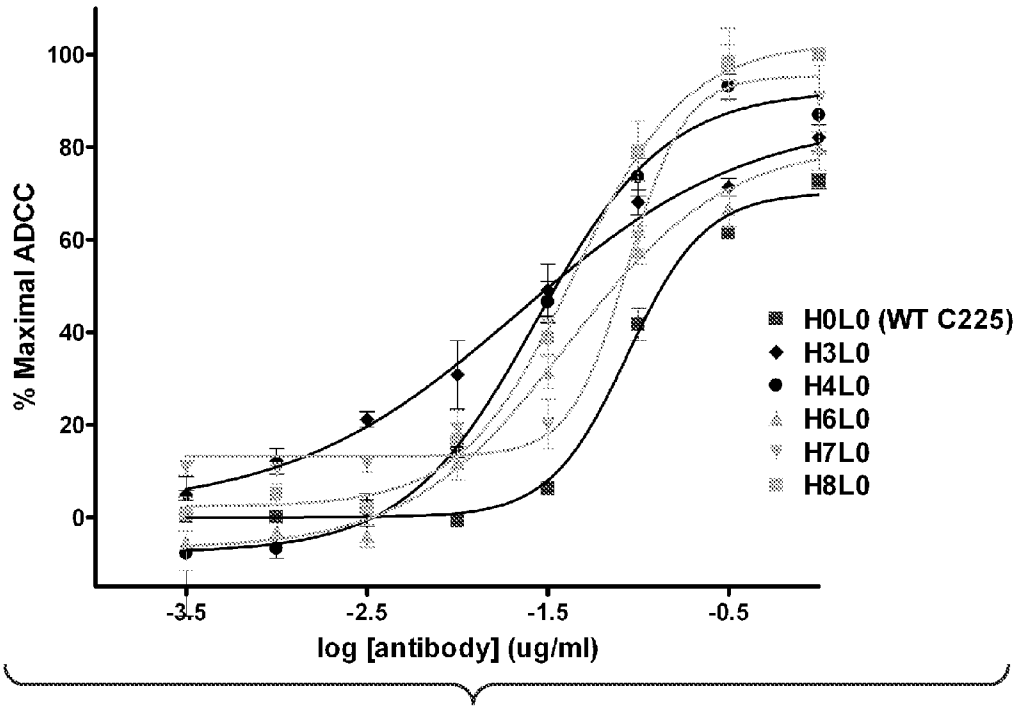


FIG. 42a

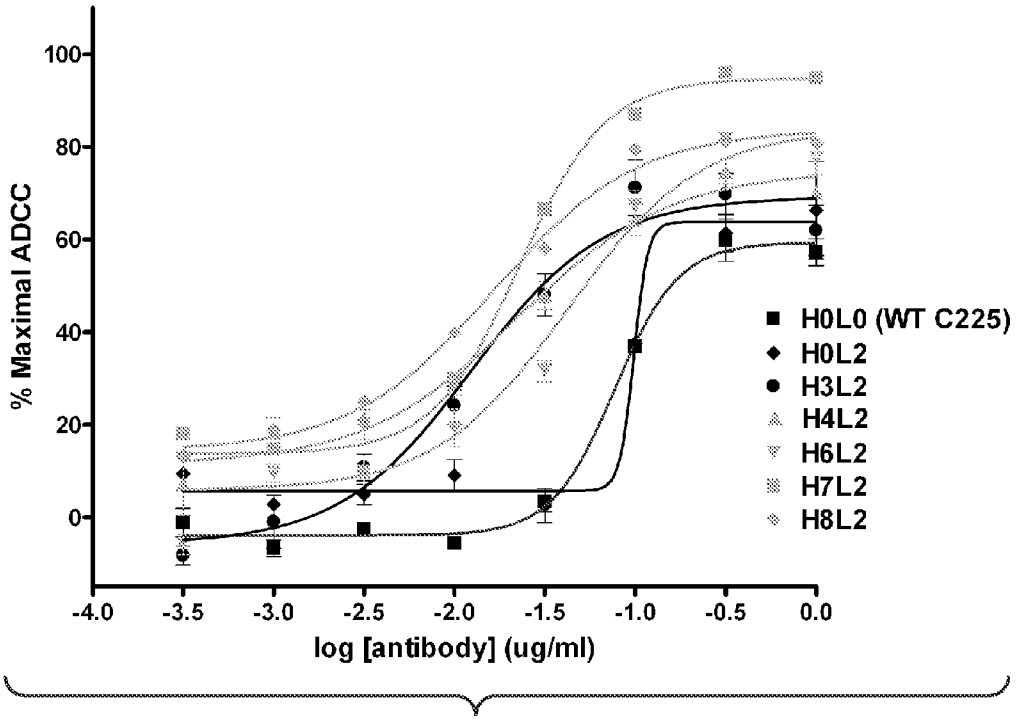


FIG. 42b

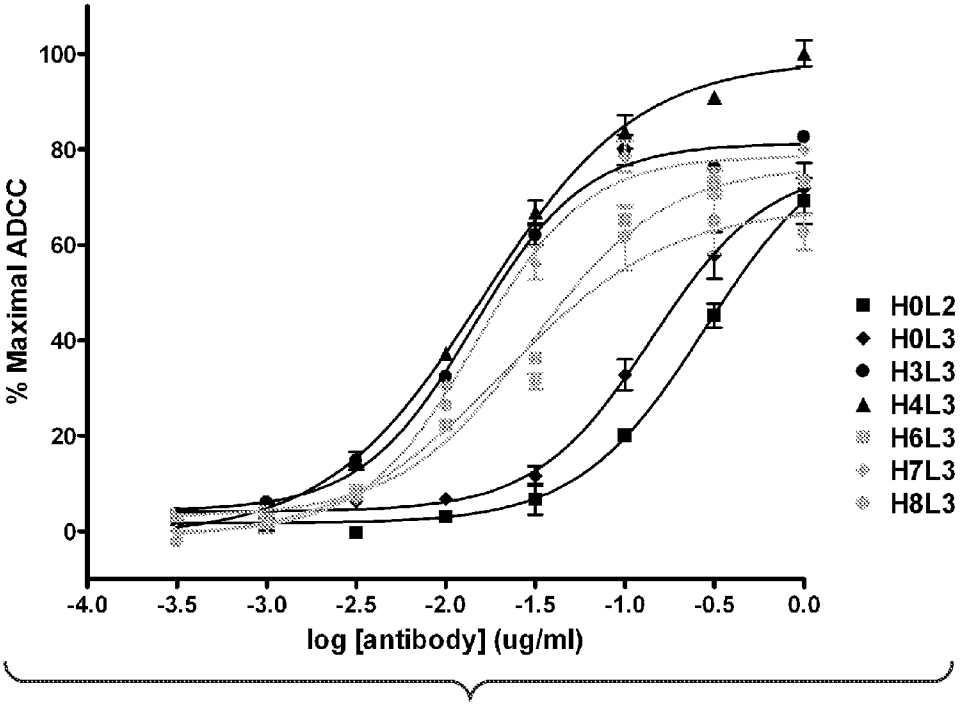


FIG. 42c

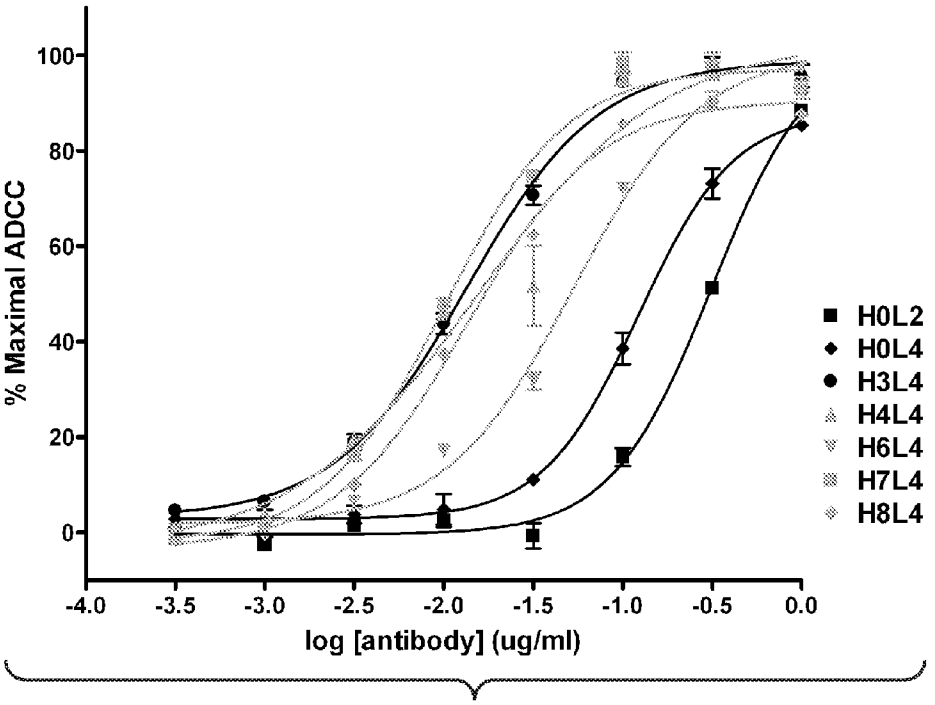


FIG. 42d

SEQ ID NO.: 388
DIQMTQSPSFLSASVGDRVTINCKASQNINNYLNWYQQKLGEAPKRLIYNTNNLQT
GIPSRFSGSGSGTDYTLTISSLQPEDFATYFCLQHNSFPTFGAGTKLELK

FIG. 43a

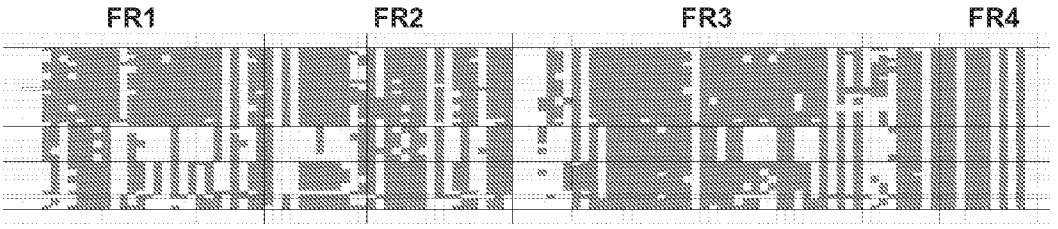


FIG. 43b

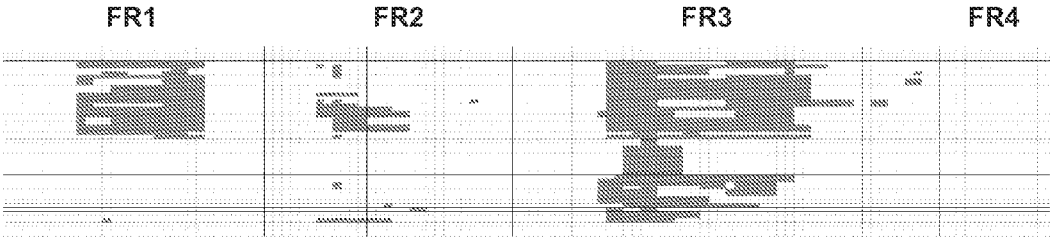


FIG. 43c

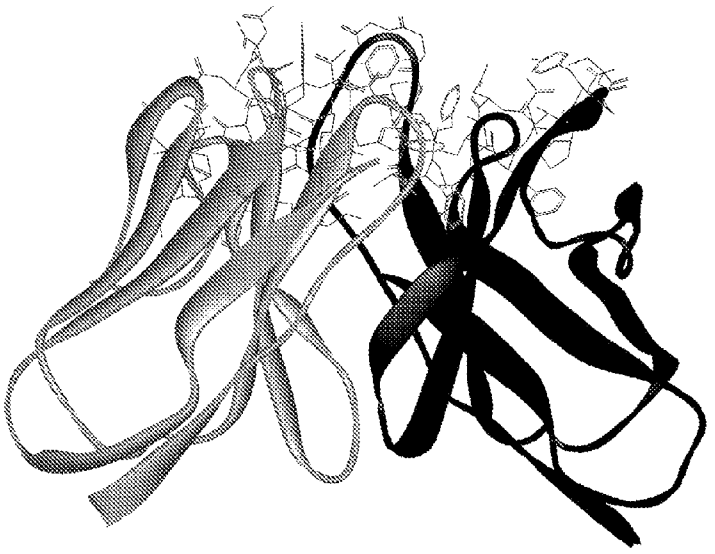


FIG. 43d

SEQ ID NO.: 389
QVNLLQSGAALVKPGASVKLSCKGSGFTFTDYKIHWVKQSHGKSLEWIGYFNPNS
GYSTYNEKFKSKATLTADKSTDTAYMELTSLTSEDSATYYCTRLSPGGYYVMDAWG
QGASVTVSS

FIG. 44a

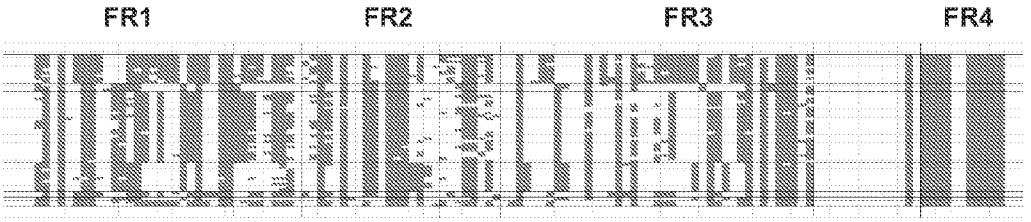


FIG. 44b

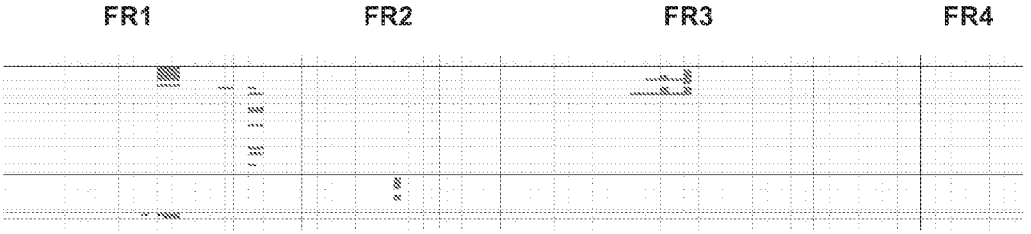


FIG. 44c

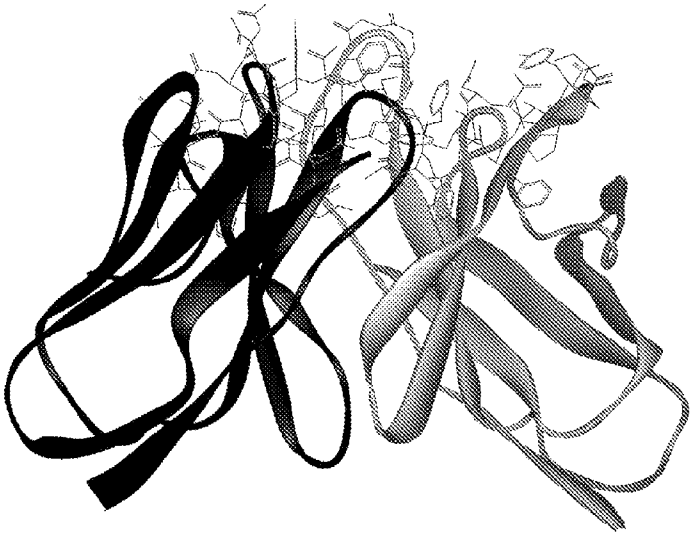


FIG. 44d

SEQ ID NO.: 390
DIQMTQSPSSLSASVGDRVTITCRASQNINNYLGWYQQKPGKAPKRLIYNTNNLQT
GVPSRFSGSGSGTEFTLTISSLQPEDFATYYCLQHNSFPTFGAGTKLEIK

FIG. 45a

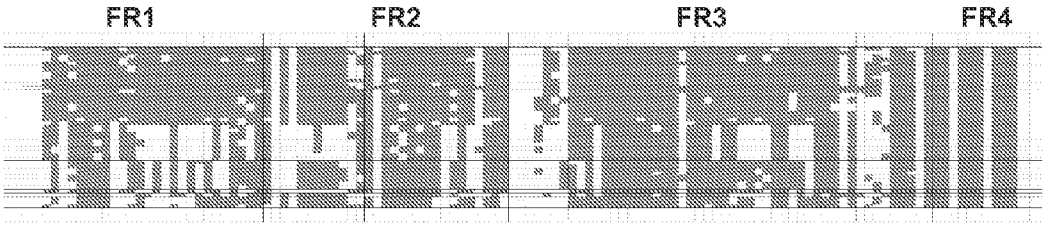


FIG. 45b

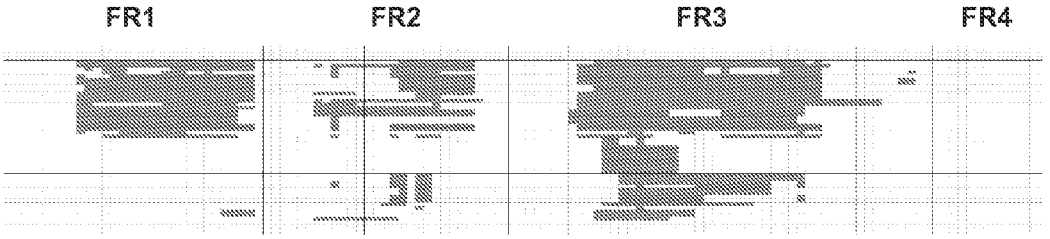


FIG. 45c

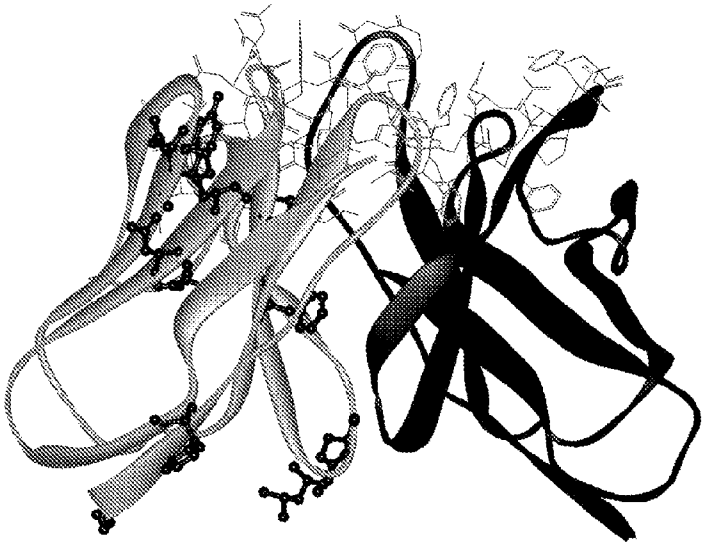


FIG. 45d

SEQ ID NO.: 391
EV**QL**VQSGA**EVK**KPGA**T**VK**I**SCK**V**SGFTFTDYK**M**HWV**QQAP**GKG**L**EW**M**G**LV**NPN
SGY**TIYA**EKF**QGRVTI**TADTSTDTAYMEL**S**SL**R**SEDTA**V**YYC**AT**LSPGGYYVMDAW
GQG**TL**VTVSS

FIG. 46a

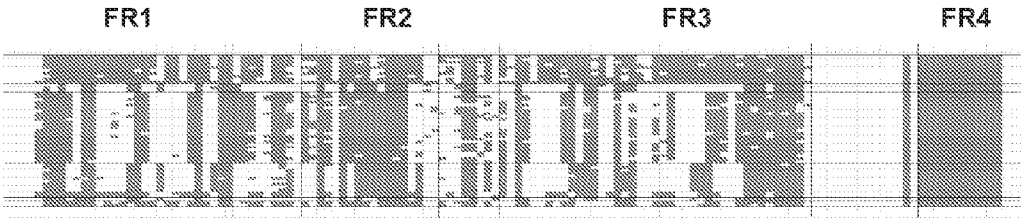


FIG. 46b

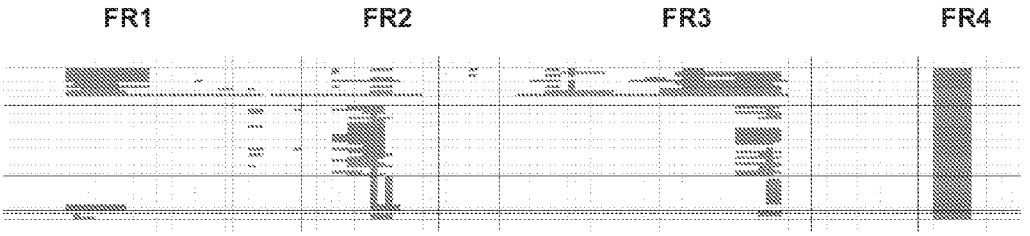


FIG. 46c



FIG. 46d

ICR62 VL HEC Calculation 1

(WT) SEQ ID NO.:388	Iter	Structural Consensus	Structural Precedence	HSC	HSS	Kabat N_9max	Kabat FRH	Kabat WT Muts
SEQ ID NO.:392	1	0.60	0.59	0.90	0.56	56	0.48	7
SEQ ID NO.:393	2	0.60	0.59	0.90	0.56	56	0.48	7
SEQ ID NO.:394	3	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:395	4	0.60	0.57	0.90	0.56	56	0.59	7
SEQ ID NO.:396	5	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:397	6	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:398	7	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:399	8	0.58	0.56	0.90	0.56	56	0.48	7
SEQ ID NO.:400	9	0.60	0.58	0.90	0.56	56	0.66	7
SEQ ID NO.:401	10	0.61	0.59	0.90	0.56	56	0.64	7
SEQ ID NO.:402	11	0.58	0.55	0.90	0.56	56	0.52	7
SEQ ID NO.:403	12	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:404	13	0.59	0.57	0.90	0.56	56	0.48	7
SEQ ID NO.:405	14	0.58	0.56	0.90	0.56	56	0.48	7
SEQ ID NO.:406	15	0.58	0.55	0.90	0.56	56	0.52	7
SEQ ID NO.:407	16	0.59	0.58	0.90	0.56	56	0.50	7
SEQ ID NO.:408	17	0.58	0.56	0.90	0.56	56	0.48	7
SEQ ID NO.:409	18	0.59	0.56	0.90	0.56	56	0.52	7
SEQ ID NO.:410	19	0.60	0.58	0.90	0.56	56	0.66	7
SEQ ID NO.:411	20	0.61	0.59	0.90	0.56	56	0.64	7

FIG. 47a-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D	I	Q	M	T	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	I	Q	S	P	S	F	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	T	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T
D	I	Q	M	T	Q	S	P	S	S	L	S	A	S	V	G	D	R	V	T

FIG. 47a-2

21	22	23	24	25	26	27	28	29	30							31	32	33	34	
I	N	C	K	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	R	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	R	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	W	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	R	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	W	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	R	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	W	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	W	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	Q	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	
I	T	C	R	A	S	Q	N	I	N	_	_	_	_	_	_	N	Y	L	N	

FIG. 47a-3

35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54
W	Y	Q	Q	K	L	G	E	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L
W	Y	Q	Q	K	P	G	K	A	P	K	R	L	I	Y	N	T	N	N	L

FIG. 47a-4

55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T
Q	T	G	I	P	S	R	F	S	G	S	G	S	G	T	D	Y	T	L	T

FIG. 47a-5

75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F
I	S	S	L	Q	P	E	D	F	A	T	Y	F	C	L	Q	H	N	S	F

FIG. 47a-6

		95	96	97	98	99	100	101	102	103	104	105	106	107
_	_	P	_	T	F	G	A	G	T	K	L	E	L	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K
_	_	P	_	T	F	G	A	G	T	K	X	E	I	K

FIG. 47a-7

FIG. 47a-1	FIG. 47a-2	FIG. 47a-3	FIG. 47a-4	FIG. 47a-5	FIG. 47a-6	FIG. 47a-7

FIG. 47a

ICR62 VH HEC Calculation 1

(WT) SEQ ID NO.:389	Iter	Structural Consensus	Structural Precedence	HSC	HSS	$N_g$max	FRH	Kabat WT Muts
SEQ ID NO.:412	1	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:413	9	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:414	11	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:415	12	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:416	16	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:417	19	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:418	20	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:419	3	0.44	0.44	0.77	0.28	33	0.67	20
SEQ ID NO.:420	4	0.44	0.44	0.77	0.28	33	0.67	20
SEQ ID NO.:421	10	0.44	0.44	0.77	0.28	33	0.67	20
SEQ ID NO.:422	5	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:423	6	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:424	8	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:425	15	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:426	18	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:427	2	0.45	0.44	0.77	0.29	34	0.65	21
SEQ ID NO.:428	7	0.45	0.44	0.77	0.29	34	0.65	21
SEQ ID NO.:429	13	0.45	0.44	0.77	0.29	34	0.65	21
SEQ ID NO.:430	14	0.45	0.44	0.77	0.29	34	0.65	21
SEQ ID NO.:431	17	0.45	0.44	0.77	0.29	34	0.65	21

FIG. 47b-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Q	V	N	L	L	Q	S	G	A	A	L	V	K	P	G	A	S	V	K	L
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V

FIG. 47b-2

21	22	23	24	25	26	27	28	29	30	31		32	33	34	35	36	37	38
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	K
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_ _	Y	K	I	H	W	V	Q

FIG. 47b-3

39	40	41	42	43	44	45	46	47	48	49	50	51	52	52a	53			54	55
Q	S	H	G	K	S	L	E	W	I	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	_	S	G

FIG. 47b-4

56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Y	S	T	Y	N	E	K	F	K	S	K	A	T	L	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T

FIG. 47b-5

76	77	78	79	80	81	82	82a	82b	82c	83	84	85	86	87	88	89	90	91	92
D	T	A	Y	M	E	L	T	S	L	T	S	E	D	S	A	T	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C

FIG. 47b-6

93	94	95	96	97				98	99	100	100a	100b	100c	101	102	103	104	105	106	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	
T	R	L	S	P	_	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G	

FIG. 47b-7

107	108	109	110	111	112	113
A	S	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S

FIG. 47b-8

FIG. 47b-1	FIG. 47b-2	FIG. 47b-3	FIG. 47b-4	FIG. 47b-5	FIG. 47b-6	FIG. 47b-7	FIG. 47b-8

FIG. 47b

ICR62 VH HEC Calculation 2

(WT) SEQ ID NO.:389	Iter	Structural Consensus	Structural Precedence	HSC	HSS	N_9max	Kabat FRH	Kabat WT Muts
SEQ ID NO.:432	5	0.44	0.47	0.78	0.33	39	0.85	21
SEQ ID NO.:433	78	0.43	0.47	0.78	0.33	39	0.82	21
SEQ ID NO.:434	2	0.44	0.47	0.78	0.33	39	0.56	21
SEQ ID NO.:435	77	0.43	0.47	0.78	0.33	39	0.54	22
SEQ ID NO.:436	51	0.43	0.46	0.78	0.33	39	0.56	21
SEQ ID NO.:437	30	0.44	0.45	0.77	0.31	36	0.92	22
SEQ ID NO.:438	90	0.44	0.45	0.77	0.31	36	0.89	23
SEQ ID NO.:439	9	0.44	0.45	0.77	0.31	36	0.61	22
SEQ ID NO.:440	24	0.44	0.44	0.77	0.31	36	0.89	22
SEQ ID NO.:441	58	0.43	0.44	0.77	0.31	36	0.58	22
SEQ ID NO.:442	6	0.45	0.44	0.77	0.29	34	0.62	22
SEQ ID NO.:443	25	0.45	0.44	0.77	0.29	34	0.62	21
SEQ ID NO.:444	98	0.45	0.44	0.77	0.29	34	0.65	22
SEQ ID NO.:445	4	0.45	0.45	0.77	0.28	33	0.64	20
SEQ ID NO.:446	75	0.44	0.45	0.77	0.28	33	0.64	21
SEQ ID NO.:447	3	0.44	0.44	0.77	0.28	33	0.67	20
SEQ ID NO.:448	8	0.43	0.43	0.77	0.28	33	0.64	20
SEQ ID NO.:449	1	0.46	0.43	0.77	0.27	31	0.68	23
SEQ ID NO.:450	74	0.45	0.42	0.77	0.27	31	0.68	22
SEQ ID NO.:451	22	0.45	0.42	0.77	0.26	31	0.65	23
SEQ ID NO.:452	7	0.45	0.42	0.77	0.26	31	0.65	22
SEQ ID NO.:453	16	0.45	0.43	0.76	0.26	30	0.70	21
SEQ ID NO.:454	11	0.44	0.42	0.76	0.26	30	0.67	21
SEQ ID NO.:455	87	0.44	0.41	0.76	0.26	30	0.70	21

FIG. 47c-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Q	V	N	L	L	Q	S	G	A	A	L	V	K	P	G	A	S	V	K	L
Q	V	Q	L	V	Q	S	G	A	E	V	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	A	E	V	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	S	E	L	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	S	E	L	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	S	E	L	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	A	E	V	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	A	E	V	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	S	E	L	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	A	E	V	K	K	P	G	A	S	V	K	V
Q	V	Q	L	V	Q	S	G	S	E	L	K	K	P	G	A	S	V	K	V
E	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
E	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	T	L	R	E	S	G	P	A	L	V	K	P	G	A	S	V	K	V
E	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
E	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	V	E	S	G	G	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	Q	L	Q	Q	S	G	P	G	L	V	K	P	G	A	S	V	K	V
Q	V	T	L	R	E	S	G	P	A	L	V	K	P	G	A	S	V	K	V

FIG. 47c-2

21	22	23	24	25	26	27	28	29	30	31			32	33	34	35	36	37	38
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	K
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	Q
S	C	K	G	S	G	F	T	F	T	D	_	_	Y	K	I	H	W	V	R

FIG. 47c-3

39	40	41	42	43	44	45	46	47	48	49	50	51	52	52a	53		54	55
Q	S	H	G	K	S	L	E	W	I	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	Q	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	Q	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	Q	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G
Q	A	P	G	K	S	L	E	W	M	G	Y	F	N	P	N	_	S	G

FIG. 47c-4

56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
Y	S	T	Y	N	E	K	F	K	S	K	A	T	L	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	K	F	Q	G	R	V	T	I	T	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T
Y	S	T	Y	N	E	S	L	K	S	R	V	T	I	S	A	D	K	S	T

FIG. 47c-5

76	77	78	79	80	81	82	82a	82b	82c	83	84	85	86	87	88	89	90	91	92
D	T	A	Y	M	E	L	T	S	L	T	S	E	D	S	A	T	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C
D	T	A	Y	M	E	L	S	S	L	R	S	E	D	T	A	V	Y	Y	C

FIG. 47c-6

93	94	95	96	97			98	99	100	100a	100b	100c	101	102	103	104	105	106
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G
T	R	L	S	P	_	_	G	G	Y	Y	V	M	D	A	W	G	Q	G

FIG. 47c-7

107	108	109	110	111	112	113
A	S	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S
T	L	V	T	V	S	S

FIG. 47c-8

FIG. 47c-1 | FIG. 47c-2 | FIG. 47c-3 | FIG. 47c-4 | FIG. 47c-5 | FIG. 47c-6 | FIG. 47c-7 | FIG. 47c-8

FIG. 47c

SEQ ID NO.: 456
DIQMTQSPSSLSASVGDRVTITCRASQNINNYLNWYQQKPGKAPKRLIYNTNNLQT
GIPSRFSGSGSGTDYTLTISSLQPEDFATYFCLQHNSFPTFGAGTKLEIK

FIG. 48a

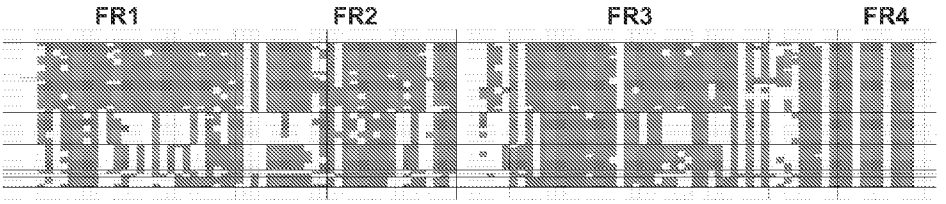


FIG. 48b

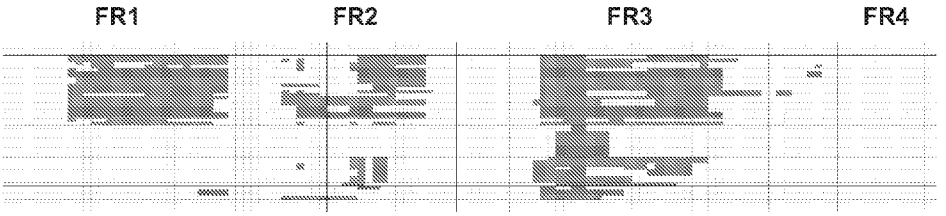


FIG. 48c

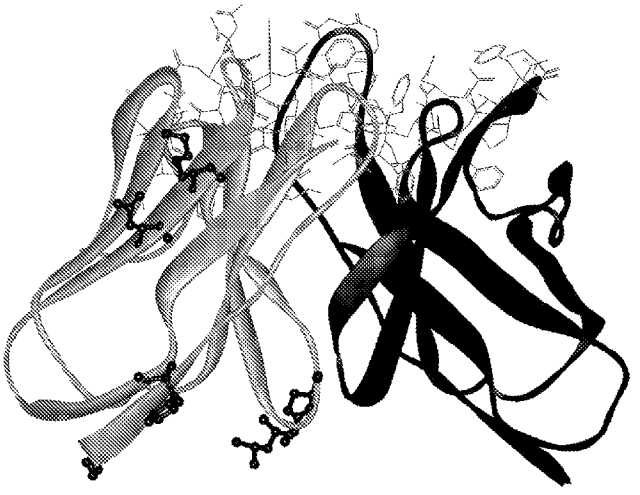


FIG. 48d

SEQ ID NO.: 457
QV**Q**L**Q**QSG**P**G**L**VKPGASVK**V**SCKGSGFTFTDYKIHWV**R**Q**A**PGKSLEW**M**GYFNPN
SGYSTYNEKF**Q**G**R**V**T**ITADKSTDTAYMEL**S**SL**R**SED**T**A**V**YYCTRLSPGGYYVMDA
WGQG**T**L**V**TVSS

FIG. 49a

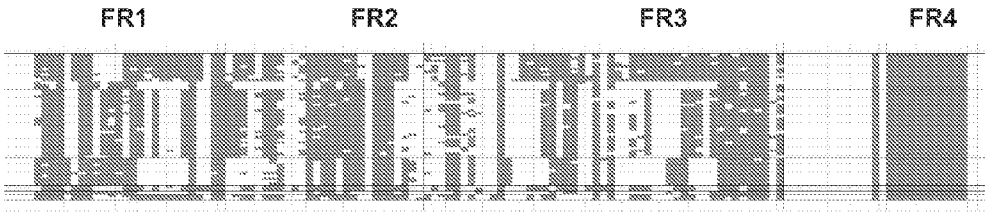


FIG. 49b

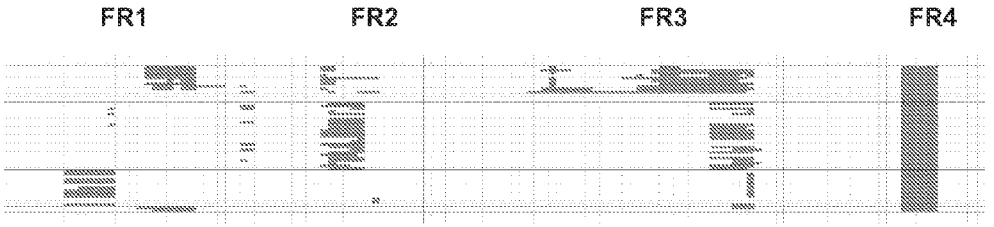


FIG. 49c

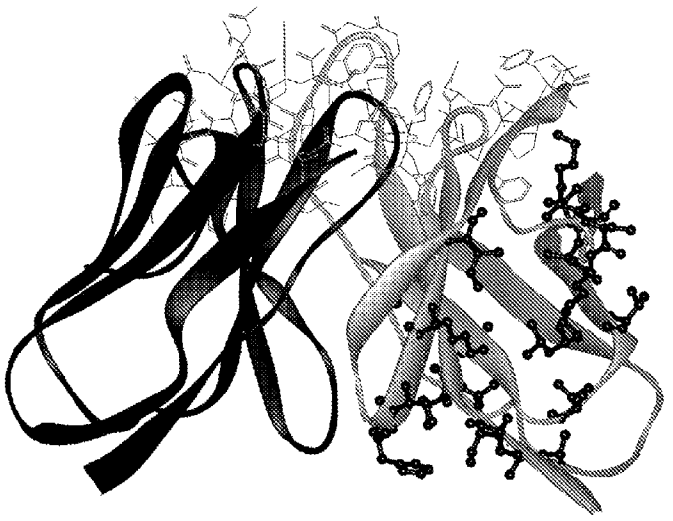


FIG. 49d

SEQ ID NO.: 458
QV**Q**L**V**QSGA**EVK**KPGASVK**V**SCKGSGFTFTDYKIHWV**R**Q**AP**GKSLEW**M**GYFNPN
SGYSTYNEKF**QGRV**TITADKSTDTAYMEL**S**SL**R**SED**TAV**YYCTRLSPGGYYVMDA
WGQG**TL**VTVSS

FIG. 50a

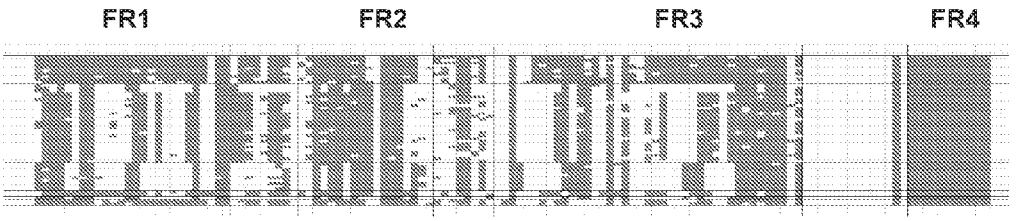


FIG. 50b

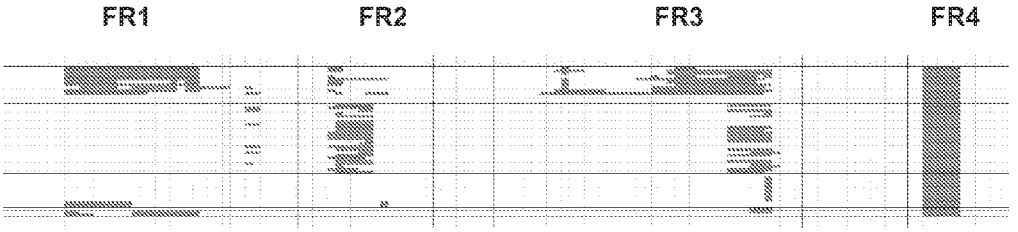


FIG. 50c



FIG. 50d

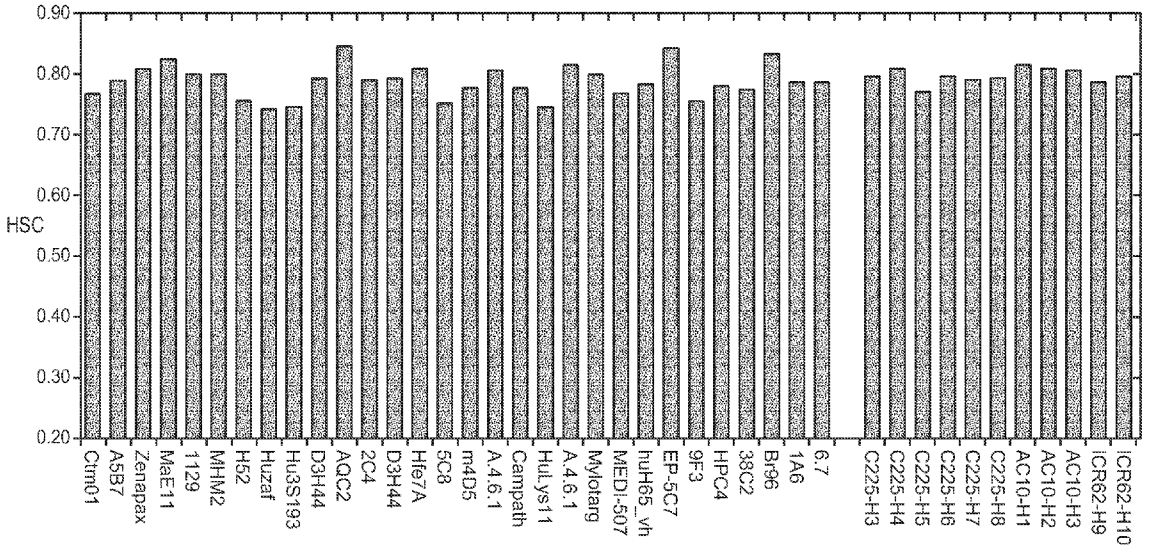

HSC
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
Ctm01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu3S193
D3H44
AQC2
2C4
D3H44
Hfe7A
5C8
m4D5
A.4.6.1
Campath
Hul,ys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225-H3
C225-H4
C225-H5
C225-H6
C225-H7
C225-H8
AC10-H1
AC10-H2
AC10-H3
iCR62-H9
iCR62-H10

FIG. 51a

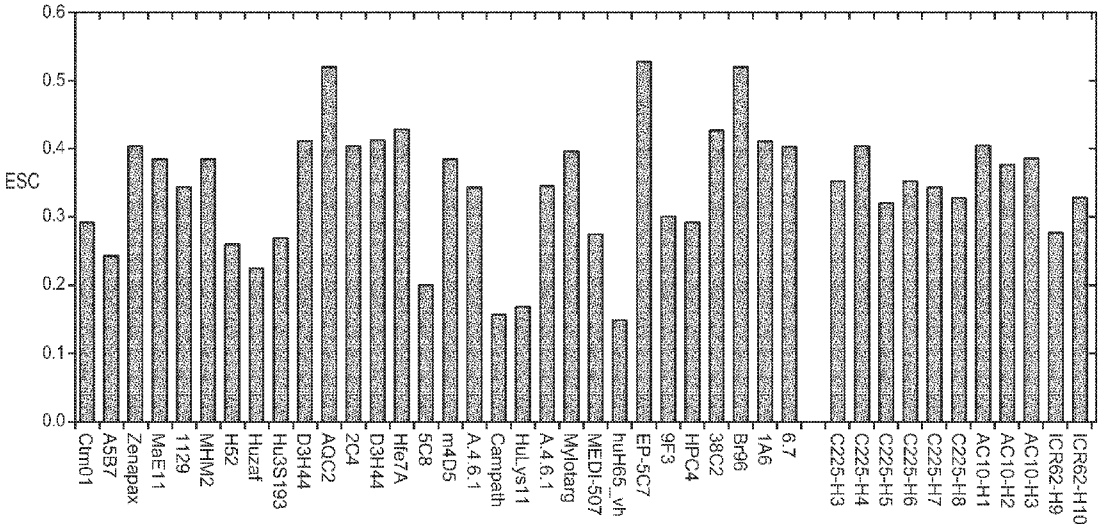

ESC
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Ctm01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu3S193
D3H44
AQC2
2C4
D3H44
Hfe7A
5C8
m4D5
A.4.6.1
Campath
HuLys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225-H3
C225-H4
C225-H5
C225-H6
C225-H7
C225-H8
AC10-H1
AC10-H2
AC10-H3
ICR62-H9
ICR62-H10

FIG. 51b

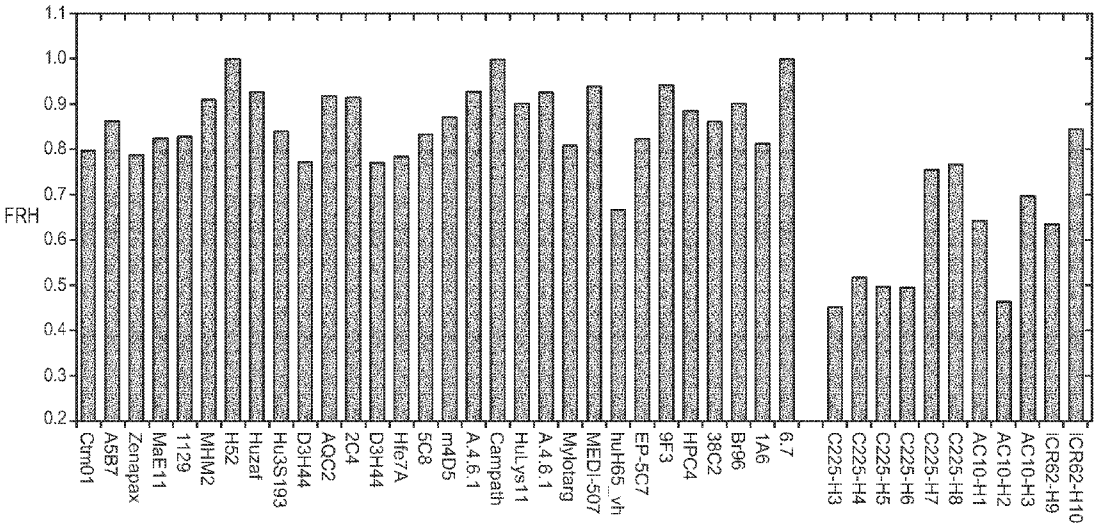

FRH
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Ctm01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu3S193
D3H44
AQC2
2C4
D3H44
Hfe7A
5C8
m4D5
A.4.6.1
Campath
HuLys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225-H3
C225-H4
C225-H5
C225-H6
C225-H7
C225-H8
AC10-H1
AC10-H2
AC10-H3
iCR62-H9
iCR62-H10

FIG. 51c

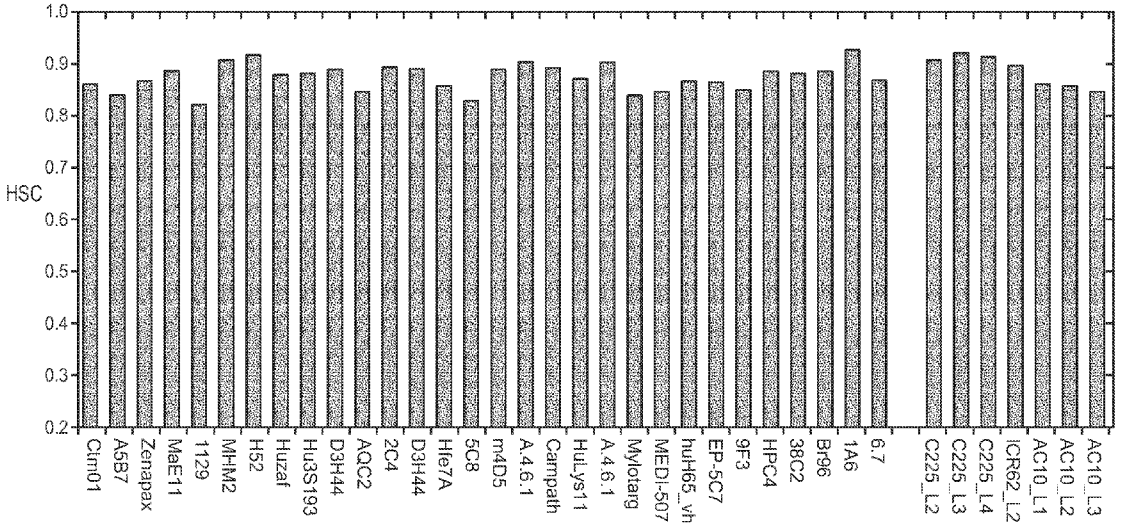

HSC
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Cim01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu3S193
D3H44
AQC2
2C4
D3H44
Hle7A
5C8
m4D5
A.4.6.1
Campath
Hul.ys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225_L2
C225_L3
C225_L4
ICR62_L2
AC10_L1
AC10_L2
AC10_L3

FIG. 52a

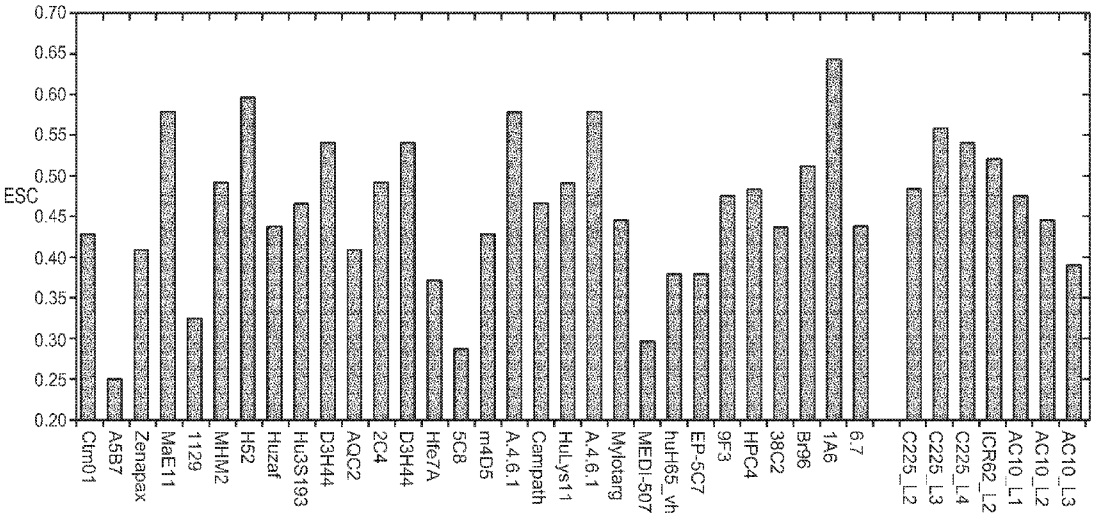

ESC
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
Ctm01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu33193
D3H44
AQC2
2C4
D3H44
Hfe7A
5C8
m4D5
A.4.6.1
Campath
HuLys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225_L2
C225_L3
C225_L4
ICR62_L2
AC10_L1
AC10_L2
AC10_L3

FIG. 52b

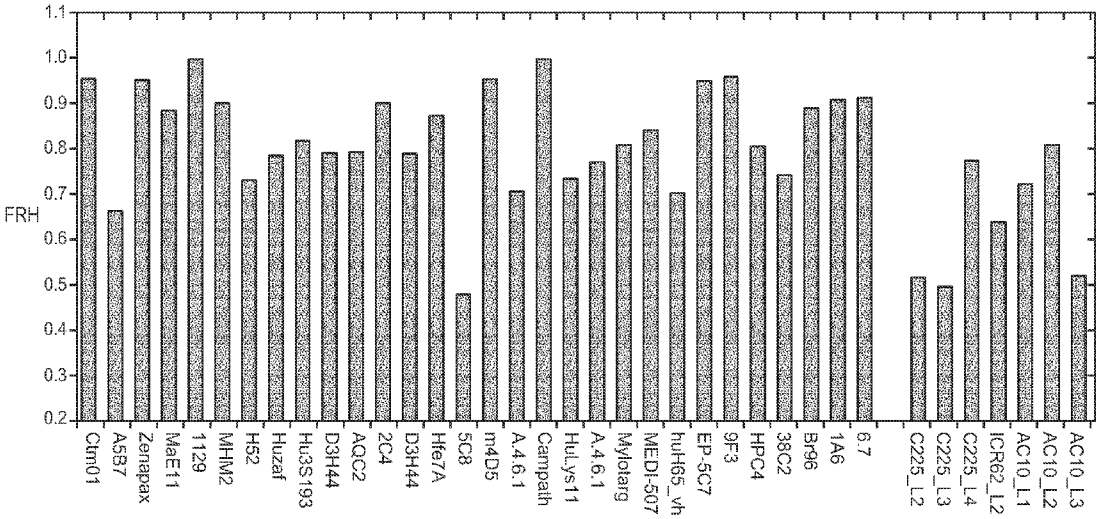

FRH
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Ctm01
A5B7
Zenapax
MaE11
1129
MHM2
H52
Huzaf
Hu3S193
D3H44
AQC2
2C4
D3H44
Hfe7A
5C8
m4D5
A.4.6.1
Campath
HuLys11
A.4.6.1
Mylotarg
MEDI-507
huH65_vh
EP-5C7
9F3
HPC4
38C2
Br96
1A6
6.7
C225_L2
C225_L3
C225_L4
ICR62_L2
AC10_L1
AC10_L2
AC10_L3

FIG. 52c

METHODS OF GENERATING VARIANT PROTEINS WITH INCREASED HOST STRING CONTENT AND COMPOSITIONS THEREOF

[0001] This application is a Continuation of U.S. application Ser. No. 11/981,794, filed Oct. 31, 2007; U.S. application Ser. No. 11/981,794 is a continuation of U.S. application Ser. No. 11/004,590, filed Dec. 3, 2004, now U.S. Pat. No. 7,657,380, issued Feb. 2, 2010, and U.S. application Ser. No. 11/004,590 claims the benefit of under 35 U.S.C. §119(e) to U.S. Ser. Nos. 60/527,167, filed Dec. 4, 2003; 60/581,613, filed Jun. 21, 2004; 60/601,665, filed Aug. 13, 2004; and, 60/619,483, filed Oct. 16, 2004; all of which are expressly incorporated by reference in their entirety.

FIELD OF THE INVENTION

[0002] The present invention relates to novel methods for generating variant proteins with increased host string content, and proteins that are engineered using these methods.

BACKGROUND OF THE INVENTION

[0003] Many proteins that have the potential to be useful human therapeutics have a xenogeneic origin. The use of xenogeneic proteins for therapeutic purposes may be advantageous for a variety of reasons, including, for example, the established success of hybridoma technology for raising antibodies in rodents, and the possibility of higher efficacy with a xenogeneic protein than with a human counterpart. Although xenogeneic proteins are a rich source of potential therapeutic molecules, they remain a relatively untapped one. One reason for this is that nonhuman proteins are often immunogenic when administered to humans, thereby greatly reducing their therapeutic utility. Additionally, even engineered proteins of human origin may become immunogenic due to changes in the protein sequence.

[0004] Immunogenicity is the result of a complex series of responses to a substance that is perceived as foreign, and may include production of neutralizing and non-neutralizing antibodies, formation of immune complexes, complement activation, mast cell activation, inflammation, hypersensitivity responses, and anaphylaxis. Several factors can contribute to protein immunogenicity, including but not limited to protein sequence, route and frequency of administration, and patient population. Immunogenicity may limit the efficacy and safety of a protein therapeutic in multiple ways. Efficacy can be reduced directly by the formation of neutralizing antibodies. Efficacy may also be reduced indirectly, as binding to either neutralizing or non-neutralizing antibodies typically leads to rapid clearance from serum. Severe side effects and even death may occur when an immune reaction is raised. One special class of side effects results when neutralizing antibodies cross-react with an endogenous protein and block its function.

[0005] Because of the clinical success of monoclonal antibodies, immunogenicity reduction of these proteins has been an intense area of investigation. Antibodies are a unique system for the development of immunogenicity reduction methods because of the large number of highly conserved antibody sequences and the wealth of high-resolution structural information. A number of strategies for reducing antibody immunogenicity have been developed. The central aim of all of these approaches has been the reduction of nonhuman, and correspondingly immunogenic content, while maintaining affinity for the antigen.

[0006] The dominant method in use for antibody immunogenicity reduction, referred to as "humanization", relies principally on the grafting of "donor" (typically mouse or rat) complementarity determining regions (CDRs) onto "acceptor" (human) variable light chain (VL) and variable heavy chain (VH) frameworks (FRs) (Tsurushita & Vasquez, 2004, Humanization of Monoclonal Antibodies, Molecular Biology of B Cells, 533-545, Elsevier Science (USA)). This strategy is referred to as "CDR grafting" (Winter U.S. Pat. No. 5,225,539). "*Backmutation*" of selected acceptor framework residues to the corresponding donor residues is often required to regain affinity that is lost in the initial grafted construct (U.S. Pat. No. 5,530,101; U.S. Pat. No. 5,585,089; U.S. Pat. No. 5,693,761; U.S. Pat. No. 5,693,762; U.S. Pat. No. 6,180,370; U.S. Pat. No. 5,859,205; U.S. Pat. No. 5,821,337; U.S. Pat. No. 6,054,297; U.S. Pat. No. 6,407,213). Despite the significant clinical application of antibodies engineered using these methods, these methods remain nonrobust with regard to their ability to reduce immunogenicity. A number of humanized antibodies have elicited substantial immune reaction in clinical studies, with incidences of immune response as high as 63% of patients (Ritter et al., 2001, *Cancer Research* 61: 6851-6859).

[0007] The incomplete capacity of current humanization methods for immunogenicity reduction are due to significant limitations imposed by the donor-acceptor approach. Historically, the use of a single donor has been part of methods aimed at engineering a single xenogeneic antibody to be suitable as a human biotherapeutic. However, the use of a single acceptor is not required. On the contrary, the use of an acceptor antibody, and the use of global homology to select it, place substantial restrictions on the immunogenicity reduction process. A principal problem is that the use of overall sequence similarity between nonhuman and human sequences as a metric for human immunogenicity is fundamentally flawed. This means of measuring the degree of humanness does not accurately account for the underlying molecular mechanisms of immune response. The immune system does not recognize antigens on the basis of global sequence similarity to human proteins. Rather, immune cells, including antigen presenting cells (APCs), T cells, and B cells, recognize linear or conformational motifs comprising only a handful of residues. A key step in antigen recognition is the formation of peptide-MHC-T cell receptor complexes. APCs express MHC molecules that recognize short (approximately nine residue) linear peptide sequences, referred to as MHC agretopes. T cells express T cell receptors that recognize T cell epitopes in the context of peptide-MHC complexes. T cells that recognize MHC agretopes that are present in human proteins typically undergo apoptosis or become anergic, while T cells that recognize foreign agretopes bound to MHC molecules may participate in an immune response. Thus the relevant quantity for the immunogenicity of a protein is not its global sequence similarity to a human sequence, but rather its sequence content of individual human epitopes.

[0008] The donor-acceptor model and the use of global sequence homology that it imposes fails in practice. Because CDRs are treated as inviolable, structural incompatibilities are introduced at the CDR-FR boundaries. Grafting of foreign donor CDRs onto a human acceptor framework creates a substantial number of nonhuman epitopes in each variable

chain, including not only the epitopes in the foreign CDRs, but also the large number of epitopes at the FR-CDR boundaries. This FR-CDR incompatibility is evident when one backs away from global homology and looks at more local sequence homologies. CDR grafting generally maximizes the donor-acceptor homology of the frameworks at the expense of the CDRs (Clark, 2000, *Immunology Today* 21: 397-402). Ironically this frequently results in lower global homology to human antibodies. In reality, the "cut and paste" approach to imparting the functional determinants of a nonhuman antibody onto the framework of a human one is unnecessary, as careful analysis of the antigen binding determinants of antibodies shows that, in fact, the majority of CDR residues are not involved in binding antigen (MacCallum et al., 1996, *J. Mol. Biol.* 262: 732-745). FR-CDR incompatibility causes not only immunological problems at the sequence level, but also causes conformational problems at the structural level. As a result, humanization methods based on CDR grafting often result in antigen affinity losses of 10-100-fold, necessitating backmutation to donor residues within the framework. This process of backmutation is a hallmark of essentially all current humanization efforts, and because it introduces yet additional nonhuman epitopes, highlights the inefficiency of these methods.

[0009] Methods that take an immune epitope approach to reducing antibody immunogenicity have been explored (U.S. Pat. No. 5,712,120; US 2003/0153043). Central to these methods is the determination of sequences within a xenogeneic antibody that are in fact immunogenic epitopes. Different methods for determination of immunogenicity both theoretical and experimental have been described and include determination of potential for amphipathic helix formation, binding to MHC, reactivity in a T-cell activation assay. A distinguishing feature between these strategies and the present invention is that the present invention makes no presumption as to the immunogenicity of specific epitopes. Rather, the primary goal is to maximize the content of human linear sequence strings in the xenogeneic antibody as determined by comparison to an alignment of human sequences. The relevant sequence dataset comprises strings that are non-immunogenic for all relevant reasons, including lack of interaction with MHC, lack of interaction with T cell receptor, lack of proper processing necessary for presentation, and tolerance.

[0010] It is noted that the methods described in U.S. Pat. No. 5,712,120 and US 2003/0153043 suffer additionally in that they fail to address a significant concern for local level sequence engineering, namely the requirement for maintaining protein structure, stability, solubility, and function. Thus, although the sequence string approach to immunogenicity reduction is more accurate than CDR grafting, it will be optimal when coupled with protein design methodology that takes into account both local sequence content and conformational compatibility at the local and global structural level. In addition to providing scoring functions for assessing host string content, the present invention also describes scoring functions that evaluate other relevant properties of a protein that may be employed for the simultaneous immunogenicity reduction and structural and functional optimization of proteins.

[0011] In summary, the donor-acceptor model imposes significant restrictions on the immunogenicity reduction process. With regard to sequence, global sequence homology is an inappropriate metric for immunogenicity. With regard to structure, backmutations are needed to repair conformational incompatibilities, thereby creating or reintroducing nonhuman epitopes. The present invention describes a novel method for antibody immunogenicity reduction that steps outside of the donor-acceptor model, and thus the sequence and structural restrictions it imposes. The central strategy of the described method is that it maximizes the content of human linear sequence strings. In this way immunogenicity is addressed at the local sequence level, typically by utilizing the local sequence information contained in an alignment of human sequences. This strategy not only provides a more accurate measure of the immunogenicity, it enables substitutions to be designed in a forward rather than backward manner to repair problems introduced by the graft. In effect, by addressing immunogenicity at the local sequence string level, the optimal balance between binding determinants and humanness can be designed.

[0012] The present invention describes a novel method for reducing the immunogenicity of proteins that leverages the nonimmunogenic information contained in natural human sequences to score protein sequences for immunogenic content at the sequence string level. Furthermore, the described method capitalizes on recent advances in computational sequence and structure-based protein engineering methods to quantitatively and systematically determine the optimal balance between human sequence content and protein functionality. Because of the wealth of human sequence information available for the immunoglobulin protein family, application to human antibodies is emphasized. Applications to other proteins are also possible.

SUMMARY OF THE INVENTION

[0013] The invention disclosed herein provides a novel method for reducing the immunogenicity of a protein, wherein the method maximizes the content of sequence strings. In a preferred embodiment, the method of the present invention maximizes the content of human sequence strings.

[0014] It is an object of the present invention to provide scoring functions that may be used to evaluate the human sequence string content of a protein. In a preferred embodiment, the scoring function compares the similarity of strings in a protein sequence to the strings that compose a set of natural protein sequences. In another preferred embodiment, the set of sequences is an aligned set of germline sequences. In additional preferred embodiments, the set of sequences contains mature sequences. In the most preferred embodiments, the sequences are human sequences.

[0015] It is an object of the present invention to provide scoring functions that may be used to evaluate the structural and/or functional fitness of a protein.

[0016] It is an object of the present invention to provide protein variants of a parent protein that are engineered using the methods described herein. In a preferred embodiment, the parent protein is an immunoglobulin.

[0017] It is an object of the present invention to provide experimental methods for screening and testing the protein variants of the present invention.

[0018] The present invention provides isolated nucleic acids encoding the protein variants described herein. The present invention provides vectors comprising the nucleic acids, optionally, operably linked to control sequences. The present invention provides host cells containing the vectors, and methods for producing and optionally recovering the protein variants.

[0019] The present invention provides compositions comprising the protein variants described herein, and a physiologically or pharmaceutically acceptable carrier or diluent.

[0020] The present invention provides novel antibodies and Fc fusions that comprise the protein variants disclosed herein. The novel antibodies and Fc fusions may find use in a therapeutic product.

[0021] The present invention provides therapeutic treatment and diagnostic uses for the protein variants disclosed herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] FIG. **1**. Human germ line sequences and diversity. The sequences that are known to encode the human VH chains (FIG. **1***a*) (SEQ ID NOS:1-53), human VL kappa chains (FIG. **1***b*) (SEQ ID NOS:54-98), and VH and VL kappa J chains (FIG. **1***c*) (SEQ ID NOS:99-109) are shown. The VL lambda germline sequences are not provided. The germline sequences are numbered according to the numbering scheme of Kabat (Kabat et al., 1991, *Sequences of Proteins of Immunological Interest,* 5th Ed., United States Public Health Service, National Institutes of Health, Bethesda). The regions of the variable region are indicated above the numbering in FIGS. **1***a* and **1***b,* and these include framework regions 1 through 3 and the CDRs 1 through 3. Positions that make up the Kabat CDRs are underlined. The germline chains are grouped into 7 subfamilies for V_H and and 6 subfamilies for V_L, as is known in the art, and these subfamilies are grouped together and separated by a blank line. The sequences of the five germlines that make up the IgG light kappa J chains (IGKJ1-IGKJ5), and the six germlines that make up the IgG heavy J chains (IGHJ1-IGHJ5) are shown in FIG. **1***c*. The kappa and lambda light J chains combine with the VLκ and VLλ germlines respectively to form the light chain variable region, and the heavy J chains combine with the VH germlines and heavy diversity (D) germlines (not shown) to form the heavy chain variable region. The V_H CDR3 is not part of the V_H germ line, and is encoded by the D and J genes.

[0023] FIG. **2**. The quantities described by equations 1, 2, and 3 are illustrated. In FIG. **2***a*, IDstring (Equation 1) is illustrated for the string beginning at position i=15, comparing a region of the murine antibody m4D5 VH sequence (VH_m4D5) (SEQ ID NO:110) as parent sequences with the homologous region from the VH human germline sequence (VH_1-2) (SEQ ID NOS:111) as human sequence h. Only 30 residues from each sequence are shown, and the residues that compose the relevant string are bolded. In FIG. **2***b*, IDmax (Equation 2) is illustrated for the parent sequence s string that begins at position i=15 (shown in bold) and the homologous regions from an aligned set of 7 VH human germline sequences (SEQ ID NOS:111-117). In FIG. **2***c*, HSC(s) (Equation 3) is illustrated for all strings (i=1 to i=22) in the parent sequence s and the homologous regions from an aligned set of 7 VH human germline sequences (SEQ ID NOS:111-117).

[0024] FIG. **3**. Sequence, host string content, and structure of WT AC10 VL. FIG. **3***a* shows the sequence of the WT AC10 VL (SEQ ID NO:117). FIG. **3***b* shows the identity of each residue in WT AC10 VL as compared to the corresponding residue in each sequence of the human VLκ germline. The black horizontal lines delineate the 7 different subfamilies as presented in FIG. **1**, and the black vertical lines delineate the different framework and CDR regions of the domain (in the order FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4). A grey square indicates that the germline sequence has the same amino acid identity to the residue at the corresponding position in the WT AC10 VL sequence. A white square indicates that the two sequences differ at that position. FIG. **3***c* shows the continuous 8- and 9-mer strings between WT AC10 VL and each sequence of the human VLκ germline. The black horizontal and vertical lines are as described in FIG. **3***b*. A grey square indicates that the germline sequence comprises an 9-mer string centered on that position that is an 8 out of 9 or 9 out of 9 identical match to the corresponding string (centered on the corresponding residue) in the WT AC10 VL sequence. FIG. **3***d* shows the structure of the modeled WT AC10 variable region. The light chain is shown as grey ribbon, the heavy chain is shown as black ribbon, and the CDR residues are indicated as black lines.

[0025] FIG. **4**. Sequence (SEQ ID NO:119), host string content, and structure of WT AC10 VH. The figure is as described in the figure legend for FIG. **3**, except that here the light chain is shown as black ribbon and the heavy chain is shown as grey ribbon.

[0026] FIG. **5**. Sequence (SEQ ID NO:120), host string content, and structure of CDR grafted AC10 VL. CDR grafted AC10 VL was derived from the CDRs of WT AC10 and the frameworks of the human germline sequence vlk_4-1. Differences between CDR grafted AC10 VL and WT AC10 VL are shown as bolded residues in the sequence in FIG. **5***a*, and as black ball and sticks in FIG. **5***d*.

[0027] FIG. **6**. Sequence (SEQ ID NO:121), host string content, and structure of CDR grafted AC10 VH. CDR grafted AC10 VH was derived from the CDRs of WT AC10 and the frameworks of the human germline sequence vh_1-3 and substitutions Q108L and A113S (Kabat numbering) in FR4

[0028] FIG. **7**. AC10 VL (SEQ ID NO:118) and VH variants with optimized HSC (SEQ ID NOS:122-160). AC10 VL (SEQ ID NO:119) and VH variants with optimized HSC (SEQ ID NOS:161-217). The nonredundant set of output sequences from the calculations described in Example 1 are shown. For each iteration (Iter) the following are provided: the Structural Consensus; Structural Precedence; Human String Content (HSC); Human String Similarity (HSS); N_9max; the Framework Region Homogeneity (FRH); and, the number of mutations from WT (Muts). The output sequences were clustered based on their mutational distance from the other sequences in the set. These clusters are delineated by the horizontal black lines. The "Cluster" column provides this mutational distance quantitatively. Differences between the parent WT AC10 sequence are shown in grey. Positions are numbered according to the Kabat numbering scheme, provided at the top. The light grey regions bracketed by the black horizontal lines indicate residues in or proximal to the Kabat defined CDRs.

[0029] FIG. **8**. Sequence (SEQ ID NO:218), host string content, and structure of L1 AC10 VL.

[0030] FIG. **9**. Sequence (SEQ ID NO:219), host string content, and structure of L2 AC10 VL

[0031] FIG. **10**. Sequence (SEQ ID NO:220), host string content, and structure of L3 AC10 VL.

[0032] FIG. **11**. Sequence (SEQ ID NO:221), host string content, and structure of H1 AC10 VH.

[0033] FIG. **12**. Sequence (SEQ ID NO:222), host string content, and structure of H2 AC10 VH.

[0034] FIG. **13**. Sequence (SEQ ID NO:223), host string content, and structure of H3 AC10 VH.

[0035] FIGS. **14**. AlphaScreen™ assay measuring binding between AC10 variants and the target antigen CD30. In the presence of competitor variant antibody, a characteristic inhibition curve is observed as a decrease in luminescence signal. The binding data were normalized to the maximum and minimum luminescence signal for each particular curve, provided by the baselines at low and high antibody concentrations respectively. The curves represent the fits of the data to a one site competition model using nonlinear regression, and the fits provide IC50s for each antibody.

[0036] FIG. **15**. FIG. **11**. SPR sensorgrams showing binding of AC10 WT and variant full length antibodies to the CD30 target antigen. The curves consist of an association phase and dissociation phase, the separation being marked by a little spike on each curve.

[0037] FIG. **16**. AlphaScreen™ assay measuring binding between AC10 variants and human V158 FcγRIIIa.

[0038] FIG. **17**. Cell-based ADCC assay of WT and AC10 variants. Purified human peripheral blood monocytes (PBMCs) were used as effector cells, L540 Hodgkin's lymphoma cells were used as target cells, and lysis was monitored by measuring LDH activity using the Cytotoxicity Detection Kit (LDH, Roche Diagnostic Corporation, Indianapolis, Ind.). Samples were run in triplicate to provide error estimates (n=3, +/−S.D.). FIG. **17** shows the dose dependence of ADCC at various antibody concentrations, and the curves represent the fits of the data to a sigmoidal dose-response model using nonlinear regression. Raw data are presented in FIGS. **17***a* and **17***b*, whereas in FIG. **17***c* the data were normalized to a percentage scale of maximal cytotoxicity determined by Triton-X100 lysis of target cells.

[0039] FIG. **18**. Cell-based assay measuring ADCC capacity of WT (H0/L0) and H3/L3 AC10 antibodies comprising Fc variants that provide enhanced effector function. Raw data were normalized to a percentage scale of maximal cytotoxicity determined by Triton-X100 lysis of target cells.

[0040] FIG. **19**. AlphaScreen™ assay measuring binding between select H3L3 secondary AC10 variants and the target antigen CD30.

[0041] FIG. **20**. Sequence (SEQ ID NO:224), host string content, and structure of L3.71 AC10 VL.

[0042] FIG. **21**. Sequence (SEQ ID NO:225), host string content, and structure of L3.72 AC10 VL.

[0043] FIG. **22**. Sequence (SEQ ID NO:226), host string content, and structure of H3.68 AC10 VH.

[0044] FIG. **23**. Sequence (SEQ ID NO:227), host string content, and structure of H3.69 AC10 VH.

[0045] FIG. **24**. Sequence (SEQ ID NO:228), host string content, and structure of H3.70 AC10 VH.

[0046] FIG. **25**. Amino acid sequences of a AC10 variant antibodies comprising the L3.71 AC10 variant VL with the CLK constant light chain (FIG. **25***a*) (SEQ ID NO:229) and the H3.70 AC10 variant VH with IgG constant chains (FIGS. **25***b*-**25***e*) (SEQ ID NOS:230-233) that may comprise amino acid modifications in the Fc region. FIG. **25***b* (SEQ ID NO:230) provides an IgG1 heavy chain with positions that may be mutated designated in bold as X_1, X_2, X_3, and X_4, referring to residues S239, V264, A330, and I332. FIG. **25***c* (SEQ ID NO:231) provides one example of a heavy chain described in FIG. **25***b*, here comprising the H3.70 AC10 variant VH region with the S239D/A330L/I332E IgG1 constant region. FIG. **25***d* (SEQ ID NO:232) provides an IgG2 heavy chain with positions that may be mutated and designated in bold as X_1, X_2, X_3, X_4, Z_1, Z_2, Z_3, Z_4, and Z_5 referring to residues S239, V264, A330, I332, P233, V234, A235, -236, and G237 (here -236 refers to a deletion at EU index position 236). FIG. **25***e* (SEQ ID NO:233) provides one example of a heavy chain described in FIG. **25***d*, here comprising the H3.70 AC10 variant VH region with the S239D/A330L/I332E/P233E/V234L/A235L/-236G IgG2 constant region.

[0047] FIG. **26**. Sequence (SEQ ID NO:234), host string content, and structure of WT C225 VL.

[0048] FIG. **27**. Sequence (SEQ ID NO:235), host string content, and structure of WT C225 VH.

[0049] FIG. **28**. Sequence (SEQ ID NO:236), host string content, and structure of CDR grafted C225 VL, which was derived from the CDRs of WT C225 and the frameworks of the human germline sequence vlk_6D-21 and an L106I (Kabat numbering) substitution in FR4.

[0050] FIG. **29**. Sequence (SEQ ID NO:237), host string content, and structure of CDR grafted C225 VH, which was derived from the CDRs of WT C225 and the frameworks of the human germline sequence vh_4-30-4 and an A113S (Kabat numbering) substitution in FR4.

[0051] FIG. **30**. C225 VL and VH variants with optimized HSC_(SEQ ID NOS:238-274). The nonredundant set of output sequences from the calculations described in Example 2 are shown.

[0052] FIG. **31**. C225 VL and VH variants with optimized HSC (SEQ ID NOS:275-378). The nonredundant set of output sequences from the calculations described in Example 2 are shown.

[0053] FIG. **32**. Sequence (SEQ ID NO:379), host string content, and structure of L2 C225 VL.

[0054] FIG. **33**. Sequence (SEQ ID NO:380), host string content, and structure of L3 C225 VL.

[0055] FIG. **34**. Sequence (SEQ ID NO:381), host string content, and structure of L4 C225 VL.

[0056] FIG. **35**. Sequence (SEQ ID NO:382), host string content, and structure of H3 C225 VH.

[0057] FIG. **36**. Sequence (SEQ ID NO:383), host string content, and structure of H4 C225 VH.

[0058] FIG. **37**. Sequence (SEQ ID NO:384), host string content, and structure of H5 C225 VH.

[0059] FIG. **38**. Sequence_(SEQ ID NO:385), host string content, and structure of H6 C225 VH.

[0060] FIG. **39**. Sequence (SEQ ID NO:386), host string content, and structure of H7 C225 VH.

[0061] FIG. **40**. Sequence (SEQ ID NO:387, host string content, and structure of H8 C225 VH.

[0062] FIG. **41**. SPR sensorgrams showing binding of full length antibody C225 variants to the EGFR target antigen. The sensorgrams show binding of C225 WT (L0/H0) and variant (L0/H3, L0/H4, L0/H5, L0/H6, L0/H7, L0/H8, L2/H3, L2/H4, L2/H5, L2/H6, L2/H7, L2/H8, L3/H3, L3/H4, L3/H5, L3/H6, L3/H7, L3/H8, L4/H3, L4/H4, L4/H5, L4/H6, L4/H7, and L4/H8) full length antibodies to the EGFR sensor chip. The curves consist of an association phase and dissociation phase, the separation being marked by a little spike on each curve.

[0063] FIG. **42**. Cell-based ADCC assay of C225 WT (L0/H0) and variant (L0/H3, L0/H4, L0/H5, L0/H6, L0/H7, L0/H8, L2/H3, L2/H4, L2/H5, L2/H6, L2/H7, L2/H8, L3/H3, L3/H4, L3/H5, L3/H6, L3/H7, L3/H8, L4/H3, L4/H4, L4/H5, L4/H6, L4/H7, and L4/H8) full length antibodies. Purified human peripheral blood monocytes (PBMCs) were used as

effector cells, A431 epidermoid carcinoma cells were used as target cells at a 10:1 effector:target cell ratio, and lysis was monitored by measuring LDH activity using the Cytotoxicity Detection Kit (LDH, Roche Diagnostic Corporation, Indianapolis, Ind.). Samples were run in triplicate to provide error estimates (n=3, +/−S.D.). FIG. **42** shows the dose dependence of ADCC at various antibody concentrations, normalized to the minimum and maximum levels of lysis for the assay. The curves represent the fits of the data to a sigmoidal dose-response model using nonlinear regression.

[0064] FIG. **43**. Sequence_(SEQ ID NO:388), host string content, and structure of WT ICR62 VL.

[0065] FIG. **44**. Sequence (SEQ ID NO:389), host string content, and structure of WT ICR62 VH.

[0066] FIG. **45**. Sequence (SEQ ID NO:390), host string content, and structure of CDR grafted ICR62 VL. CDR grafted ICR62 VL was derived from the CDRs of WT ICR62 and the frameworks of the human germline sequence vlk_1-17 and an L106I (Kabat numbering) substitution in FR4.

[0067] FIG. **46**. Sequence (SEQ ID NO:391), host string content, and structure of CDR grafted ICR62 VH. CDR grafted ICR62 VH was derived from the CDRs of WT ICR62 and the frameworks of the human germline sequence vh_14 and substitutions A107T and S108L (Kabat numbering) in FR4.

[0068] FIG. **47**. ICR62 VL and VH variants with optimized HSC (SEQ ID NOS:392-455).

[0069] FIG. **48**. Sequence (SEQ ID NO:456), host string content, and structure of L3 ICR62 VL.

[0070] FIG. **49**. Sequence (SEQ ID NO:457), host string content, and structure of H9 ICR62 VH.

[0071] FIG. **50**. Sequence (SEQ ID NO:458), host string content, and structure of H10 ICR62 VH.

[0072] FIG. **51**. Comparison of VH sequences humanized by the methods in the prior art versus the present method. Prior art antibodies include Ctm01, A5B7, Zenapax, MaE11, 1129, MHM2, H52, Huzaf, Hu3S193, D3H44, AQC2, 2C4, D3H44, Hfe7A, 5C8, m4D5, A.4.6.1, Campath, HuLys11, A.4.6.1, Mylotarg, MEDI-507, huH65_vh, EP-5C7, 9F3, HPC4, 38C2, Br96, 1A6, and 6.7. Sequences designed using the present invention, including AC10 H1, H2, and H3, C225 H3, H4, H5, H6, H7, and H8, and ICR62 H9 and H10, are offset to the right. FIG. **51***a* provides the host string content (HSC) as defined by equation 3, FIG. **51***b* provides the exact string content (ESC) as defined by equation 3a, and FIG. **51***c* provides the framework region homogeneity (FRH) as defined by equation 10. Window size w was 9 for all calculations.

[0073] FIG. **52**. Comparison of VL sequences humanized by the methods in the prior art versus the present method. Prior art antibodies include Ctm01, A5B7, Zenapax, MaE11, 1129, MHM2, H52, Huzaf, Hu3S193, D3H44, AQC2, 2C4, D3H44, Hfe7A, 5C8, m4D5, A.4.6.1, Campath, HuLys11, A.4.6.1, Mylotarg, MEDI-507, huH65_vh, EP-5C7, 9F3, HPC4, 38C2, Br96, 1A6, and 6.7. Sequences designed using the present invention, including AC10 L1, L2, and L3, C225 L2, L3, L4, and ICR62 L2, are offset to the right. FIG. **52***a* provides the host string content (HSC) as defined by equation 3, FIG. **52***b* provides the exact string content (ESC) as defined by equation 3a, and FIG. **52***c* provides the framework region homogeneity (FRH) as defined by equation 10. Window size w was 9 for all calculations.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[0074] In order that the invention may be more completely understood, several definitions are set forth below. Such definitions are meant to encompass grammatical equivalents.

[0075] By "amino acid" as used herein is meant one of the 20 naturally occurring amino acids or any non-natural analogues that may be present at a specific, defined position.

[0076] By "amino acid modification" herein is meant an amino acid substitution, insertion, and/or deletion in a polypeptide sequence. The preferred amino acid modification herein is a substitution.

[0077] By "amino acid substitution" or "substitution" herein is meant the replacement of an amino acid at a given position in a protein sequence with another amino acid.

[0078] By "antibody" herein is meant a protein consisting of one or more proteins substantially encoded by all or part of the recognized immunoglobulin genes. The recognized immunoglobulin genes, for example in humans, include the kappa (κ), lambda (λ), and heavy chain genetic loci, which together comprise the myriad variable region genes, and the constant region genes mu (μ), delta (δ), gamma (γ), sigma (σ), and alpha (α) which encode the IgM, IgD, IgG, IgE, and IgA isotypes respectively. Antibody herein is meant to include full length antibodies and antibody fragments, and may refer to a natural antibody from any organism, an engineered antibody, or an antibody generated recombinantly for experimental, therapeutic, or other purposes. By "IgG" as used herein is meant a protein belonging to the class of antibodies that are substantially encoded by a recognized immunoglobulin gamma gene. In humans this class comprises IgG1, IgG2, IgG3, and IgG4.

[0079] By "corresponding" or "equivalent" residues as meant herein are residues that represent similar or homologous sequence and/or structural environments between a first and second protein, or between a first protein and set of multiple proteins. In order to establish homology, the amino acid sequence of a first protein is directly compared to the sequence of a second protein. After aligning the sequences, using one or more of the homology alignment programs known in the art (for example using conserved residues as between species), allowing for necessary insertions and deletions in order to maintain alignment (i.e., avoiding the elimination of conserved residues through arbitrary deletion and insertion), the residues equivalent to particular amino acids in the primary sequence of the first protein are defined. Alignment of conserved residues preferably should conserve 100% of such residues. However, alignment of greater than 75% or as little as 50% of conserved residues is also adequate to define equivalent residues. Corresponding residues may also be defined by determining structural homology between a first and second protein that is at the level of tertiary structure for proteins whose structures have been determined. In this case, equivalent residues are defined as those for which the atomic coordinates of two or more of the main chain atoms of a particular amino acid residue of the proteins (N on N, CA on CA, C on C and O on O) are within 0.13 nm and preferably 0.1 nm of each other after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the proteins.

[0080] By "CDR" as used herein is meant a Complementarity Determining Region of an antibody variable domain. Systematic identification of residues included in the CDRs have been developed by Kabat (Kabat et al., 1991, Sequences of Proteins of Immunological Interest, 5th Ed., United States Public Health Service, National Institutes of Health, Bethesda) and alternately by Chothia (Chothia & Lesk, 1987, J. Mol. Biol. 196: 901-917; Chothia et al., 1989, Nature 342:

877-883; Al-Lazikani et al., 1997, J. Mol. Biol. 273: 927-948). For the purposes of the present invention, CDRs are defined as a slightly smaller set of residues than the CDRs defined by Chothia. VL CDRs are herein defined to include residues at positions 27-32 (CDR1), 50-56 (CDR2), and 91-97 (CDR3), wherein the numbering is according to Chothia. Because the VL CDRs as defined by Chothia and Kabat are identical, the numbering of these VL CDR positions is also according to Kabat. VH CDRs are herein defined to include residues at positions 27-33 (CDR1), 52-56 (CDR2), and 95-102 (CDR3), wherein the numbering is according to Chothia. These VH CDR positions correspond to Kabat positions 27-35 (CDR1), 52-56 (CDR2), and 95-102 (CDR3).

[0081] By "framework" as used herein is meant the region of an antibody variable domain exclusive of those regions defined as CDR's. Each antibody variable domain framework can be further subdivided into the contiguous regions separated by the CDR's (FR1, FR2, FR3 and FR4).

[0082] By "germline" as used herein is meant the set of sequences that compose the natural genetic repertoire of a protein, and its associated alleles.

[0083] By "host" as used herein is meant a family, genus, species or subspecies, group of individuals or even a single individual. A host group of individuals can be selected for based upon a variety of criteria, such as MHC allele composition, etc. In a preferred embodiment, a host is canine, murine, primate, or human. In the most preferred embodiment, a host is human.

[0084] By "host string" or "host sequence" as used herein is meant a string or sequence that encodes any part of a naturally occurring host protein.

[0085] By "humanized" antibody as used herein is meant an antibody comprising a human framework region and one or more CDR's from a non-human (usually mouse or rat) antibody. The non-human antibody providing the CDR's is called the "donor" and the human immunoglobulin providing the framework is called the "acceptor". One says that the donor antibody has been "humanized", by the process of "humanization".

[0086] By "identity" as used herein is meant the number of residues in a first sequence that are identical to the residues in a second sequence after alignment of the sequences to achieve the maximum identity.

[0087] By "immune epitope" or "epitope" herein is meant a linear sequence of amino acids that is located in a protein of interest. Epitopes may be analyzed for their potential for immunogenicity. Epitopes may be any length, preferably 9-mers.

[0088] By "immunogenicity" herein is meant the ability of a protein to elicit an immune response, including but not limited to production of neutralizing and non-neutralizing antibodies, formation of immune complexes, complement activation, mast cell activation, inflammation, and anaphylaxis.

[0089] By "immunoglobulin (Ig)" herein is meant a protein consisting of one or more proteins substantially encoded by immunoglobulin genes. Immunoglobulins include but are not limited to antibodies. Immunoglobulins may have a number of structural forms, including but not limited to full length antibodies, antibody fragments, and individual immunoglobulin domains. By "immunoglobulin (Ig) domain" herein is meant a region of an immunoglobulin that exists as a distinct structural entity as ascertained by one skilled in the art of protein structure. Ig domains typically have a characteristic β-sandwich folding topology. The known Ig domains in the IgG class of antibodies are V_H, Cγ1, Cγ2, Cγ3, V_L, and C_L.

[0090] By "natural sequence" or "natural protein" as used herein is meant a protein that has been determined to exist absent any experimental modifications. Also included are sequences that can be predicted to exist in nature based on experimentally determined sequences. An example of such a predicted sequence is an antibody that can be predicted to exist based on the established patterns of germline recombination. In this case the large size of the predicted antibody repertoire makes the actual experimental determination of all mature recombined antibodies not practical.

[0091] By "parent" or "parent protein" as used herein is meant a protein that is subsequently modified to generate a variant. The parent protein may be a naturally occurring protein, or a variant or engineered version of a naturally occurring protein. Parent protein may refer to the protein itself, compositions that comprise the parent protein, or the amino acid sequence that encodes it. Accordingly, by "parent antibody" as used herein is meant an antibody that is subsequently modified to generate a variant antibody. Accordingly, by "parent sequence" as used herein is meant the sequence that encodes the parent protein or parent antibody.

[0092] By "position" as used herein is meant a location in the sequence of a protein. Positions may be numbered sequentially, or according to an established format, for example Kabat, Chothia, and/or the EU index as in Kabat.

[0093] By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures.

[0094] By "reduced immunogenicity" herein is meant a decreased ability to activate the immune system, when compared to the parent protein. For example, a protein variant can be said to have "reduced immunogenicity" if it elicits neutralizing or non-neutralizing antibodies in lower titer or in fewer patients than the parent protein. A protein variant also can be said to have "reduced immunogenicity" if it shows decreased binding to one or more MHC alleles or if it induces T cell activation in a decreased fraction of patients relative to the parent protein.

[0095] By "residue" as used herein is meant a position in a protein and its associated amino acid identity. For example, proline 9 (also referred to as Pro9, also referred to as P9) is a residue in the WT AC10 VH region.

[0096] By "scoring function" herein is meant any equation or method for evaluating the fitness of one or more amino acid modifications in a protein. The scoring function may involve a physical or chemical energy term, or may involve knowledge-, statistical-, sequence-based energy terms, and the like.

[0097] By "string" as used herein is meant a contiguous sequence that encodes any part of a protein. Strings may comprise any 2 or more linear residues, with the number of contiguous residues being defined by the "window" or "window size". Window sizes of 2-20 are preferred, with 7-13 more preferred, with 9 most preferred.

[0098] By "target" as used herein is meant the molecule that is bound specifically by a protein. A target may be a protein, carbohydrate, lipid, or other chemical compound. The target of an antibody is its antigen, also referred to as its target antigen.

[0099] By "variable region" as used herein is meant the region of an immunoglobulin that comprises one or more Ig domains substantially encoded by any of the VL (including Vκ and Vλ) and/or V_H genes that make up the light chain (including kappa and lambda) and heavy chain immunoglobulin genetic loci respectively. A light or heavy chain variable region (VL and VH) consists of a "framework" or "FR" region interrupted by three hypervariable regions referred to as "complementarity determining regions" or "CDRs". The extent of the framework region and CDRs have been precisely defined, for example as in Kabat (see "Sequences of Proteins of Immunological Interest," E. Kabat et al., U.S. Department of Health and Human Services, (1983)), and as in Chothia. The framework regions of an antibody, that is the combined framework regions of the constituent light and heavy chains, serves to position and align the CDRs, which are primarily responsible for binding to an antigen.

[0100] By "variant protein" or "protein variant", or "variant" as used herein is meant a protein that differs from a parent protein by virtue of at least one amino acid modification. Protein variant may refer to the protein itself, a composition comprising the protein, or the amino sequence that encodes it. Preferably, the protein variant has at least one amino acid modification compared to the parent protein, e.g. from about one to about ten amino acid modifications, and preferably from about one to about five amino acid modifications compared to the parent. The protein variant sequence herein will preferably possess at least about 80% homology with a parent protein sequence, and most preferably at least about 90% homology, more preferably at least about 95% homology. Accordingly, by "immunoglobulin variant" as used herein is meant an immunoglobulin that differs from a parent immunoglobulin by virtue of at least one amino acid modification.

[0101] By "wild type or WT" herein is meant an amino acid sequence or a nucleotide sequence that is found in nature and includes allelic variations. A WT protein has an amino acid sequence or a nucleotide sequence that has not been intentionally modified.

[0102] The protein variants of the present invention may be derived from parent proteins that are themselves from a wide range of sources. The parent protein may be substantially encoded by one or more genes from any organism, including but not limited to humans, mice, rats, rabbits, camels, llamas, dromedaries, monkeys, preferably mammals and most preferably humans and mice and rats. Although in a preferred embodiment the parent protein is nonhuman, in some embodiments of the present invention the parent protein may be human or similar to human. The parent protein may comprise more than one protein chain, and thus may be a monomer or an oligomer, including a homo- or hetero-oligomer. In a preferred embodiment, the parent protein is an antibody, referred to as the parent antibody. The parent antibody need not be naturally occurring. For example, the parent antibody may be an engineered antibody, including but not limited to nonhuman and chimeric antibodies. The parent antibody may be fully human, obtained for example using transgenic mice (Bruggemann et al., 1997, *Curr Opin Biotechnol* 8:455-458) or human antibody libraries coupled with selection methods (Griffiths et al., 1998, *Curr Opin Biotechnol* 9:102-108). The parent antibody need not be naturally occurring. For example, the parent antibody may be an engineered antibody, including but not limited to chimeric antibodies and humanized antibodies (Clark, 2000, *Immunol Today* 21:397-402). The parent antibody may be an engineered variant of an antibody that is substantially encoded by one or more natural antibody genes. In one embodiment, the parent antibody has been affinity matured, as is known in the art, or engineered in some other way. The parent antibodies of the present invention may be substantially encoded by immunoglobulin genes belonging to any of the antibody classes, and may comprise sequences belonging to the IgG class of antibodies, including IgG1, IgG2, IgG3, or IgG4, or alternatively the IgA (including subclasses IgA1 and IgA2), IgD, IgE, IgG, or IgM classes of antibodies.

[0103] Virtually any binding partner or antigen may be targeted by the proteins of the present invention. A number biotherapeutic proteins and antibodies that are approved for use, in clinical trials, or in development may thus benefit from immunogenicity reduction methods of the present invention. In a preferred embodiment, the less immunogenic protein of the present invention is an antibody. The less immunogenic antibody may comprise sequences belonging to the IgG (including IgG1, IgG2, IgG3, or IgG4), IgA (including subclasses IgA1 and IgA2), IgD, IgE, IgG, or IgM classes of antibodies, with the IgG class being preferred. The less immunogenic antibodies of the present invention may be full length antibodies, or antibody fragments. Constant regions need not be present, but if they are, they will likely be substantially identical to human immunoglobulin constant regions.

[0104] The constant region of the antibody may be modified in some way to make it more effective therapeutically. For example, the constant region may comprise substitutions that enhance therapeutic properties. Most preferred substitutions and optimized effector function properties are described in U.S. Ser. No. 10/672,280, PCT US03/30249, and U.S. Ser. No. 10/822,231, and U.S. Ser. No. 60/627,774, filed Nov. 12, 2004 and entitled "Optimized Fc Variants". Other known Fc variants that may find use in the present invention include but are not limited to those described in U.S. Pat. No. 6,737,056; PCT US2004/000643; U.S. Ser. No. 10/370,749; PCT/US2004/005112; US 2004/0132101; U.S. Ser. No. 10/672,280; PCT/US03/30249; U.S. Pat. No. 6,737,056, US 2004/0002587; WO 2004/063351; Idusogie et al., 2001, J. Immunology 166:2571-2572; Hinton et al., 2004, J. Biol. Chem. 279(8): 6213-6216. In alternate embodiments, the constant region may comprise one or more engineered glycoforms, as is known in the art (Umaña et al., 1999, *Nat Biotechnol* 17:176-180; Davies et al., 2001, *Biotechnol Bioeng* 74:288-294; Shields et al., 2002, *J Biol Chem* 277: 26733-26740; Shinkawa et al., 2003, *J Biol Chem* 278:3466-3473); (U.S. Pat. No. 6,602,684; U.S. Ser. No. 10/277,370; U.S. Ser. No. 10/113,929; PCT WO 00/61739A1; PCT WO 01/29246A1; PCT WO 02/31140A1; PCT WO 02/30954A1); (Potelligent™ technology [Biowa, Inc., Princeton, N.J.]; GlycoMAb™ glycosylation engineering technology [GLYCART biotechnology AG, Zürich, Switzerland]).

[0105] The protein variants of the present invention may find use in a wide range of protein products. In one embodiment the protein is a therapeutic, a diagnostic, or a research reagent, preferably a therapeutic. Alternatively, the protein of the present invention may be used for agricultural or industrial uses. In a preferred embodiment, the protein is a therapeutic that is used to treat a disease. By "disease" herein is meant a disorder that may be ameliorated by the administration of a pharmaceutical composition comprising a protein of the present invention. Diseases include but are not limited to autoimmune diseases, immunological diseases, infectious

diseases, inflammatory diseases, neurological diseases, and oncological and neoplastic diseases including cancer. In one embodiment, a protein of the present invention is the only therapeutically active agent administered to a patient. Alternatively, the protein of the present invention is administered in combination with one or more other therapeutic agents, including but not limited to cytotoxic agents, chemotherapeutic agents, cytokines, growth inhibitory agents, anti-hormonal agents, kinase inhibitors, anti-angiogenic agents, cardioprotectants, or other therapeutic agents. The proteins of the present invention may be combined with other therapeutic regimens. For example, in one embodiment, the patient to be treated with the protein may also receive radiation therapy and/or undergo surgery. In an alternate embodiment, the protein of the present invention is conjugated or operably linked to another therapeutic compound. The therapeutic compound may be a cytotoxic agent, a chemotherapeutic agent, a toxin, a radioisotope, a cytokine, or other therapeutically active agent. In yet another embodiment, a protein of the present invention may be conjugated to a protein or molecule for utilization in tumor pretargeting or prodrug therapy. Other modifications of the proteins of the present invention are contemplated herein. For example, the protein may be linked to one of a variety of nonproteinaceous polymers, for example e.g., polyethylene glycol (PEG).

[0106] Pharmaceutical compositions are contemplated wherein a protein of the present invention and one or more therapeutically active agents are formulated. Formulations of the proteins of the present invention are prepared for storage by mixing the protein having the desired degree of purity with optional pharmaceutically acceptable carriers, excipients or stabilizers (Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed.,1980), in the form of lyophilized formulations or aqueous solutions. The formulations to be used for in vivo administration are preferably sterile. The proteins disclosed herein may also be formulated as immunoliposomes, or entrapped in microcapsules. The concentration of the protein of the present invention in the formulation may vary from about 0.1 to 100 weight %. In a preferred embodiment, the concentration of the protein is in the range of 0.003 to 1.0 molar. In order to treat a patient, a therapeutically effective dose of the protein of the present invention may be administered. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques. Dosages may range from 0.01 to 100 mg/kg of body weight or greater, for example 0.1, 1, 10, or 50 mg/kg of body weight, with 1 to 10 mg/kg being preferred. Administration of the pharmaceutical composition comprising a protein of the present invention, preferably in the form of a sterile aqueous solution, may be done in a variety of ways, including, but not limited to, orally, subcutaneously, intravenously, intranasally, intraotically, transdermally, topically, intraperitoneally, intramuscularly, intrapulmonary, inhalably, vaginally, parenterally, rectally, or intraocularly. As is known in the art, the pharmaceutical composition may be formulated accordingly depending upon the manner of introduction.

Description of the Methodology

[0107] The present invention provides a novel method for reducing the immunogenicity of a protein. A central principle of the described method is that substitutions are designed to maximize the content of human linear sequence strings using an alignment of human sequences. For application to antibodies, this approach to immunogenicity reduction excludes the use of the single donor-acceptor model employed in humanization methods. By stepping outside of the limitations imposed by the need to choose a human acceptor sequence a priori, a more immunologically relevant approach to immunogenicity reduction is enabled. Sequence information and structural information may be used to score potential amino acid substitutions. The scoring results are used to design protein variant libraries, which are subsequently screened experimentally to determine favorable substitutions. Feedback from experimental data may guide subsequent iterations of design and experimental screening, ultimately enabling protein variants to be engineered with the optimal balance between biophysical and immunological constraints.

Sequences

[0108] Central to the method described herein is that a set of host sequences provides information as to the degree to which linear sequence strings have the potential to be immunogenic. Thus the set of sequences employed is an important parameter. In the most common embodiment, the sequences are a set of human sequences that are homologous in sequence and/or structure to the parent sequence. As is known in the art, some proteins share a common structural scaffold and are homologous in sequence. This information may be used to gain insight into particular positions in the protein family. Sequence alignments are often carried out to determine which protein residues are conserved and which are not conserved. That is to say, by comparing and contrasting alignments of protein sequences, the degree of variability at a position may be observed, and the types of amino acids that occur naturally at positions may be observed. Thus for the present invention, typically the sequences are aligned such that the conserved or similar residues that exist between the parent sequence and the set of human sequences and among the set of human sequences can be identified. Methods for sequence alignment are well known in the art, and include alignments based on sequence and structural homology.

[0109] Protein sequence information can be obtained, compiled, and/or generated from sequence alignments of naturally occurring proteins from any organism, including but not limited to mammals. Because a preferred embodiment of present invention is directed towards immunogenicity reduction for biotherapeutics, the sequences that compose the set are most preferably human. The source of the sequences may vary widely, may be a database that is compiled publicly or privately, and may be may include one or more of the known general protein and nucleic acid sequences databases, including but not limited to SwissProt, Gen Bank® and Entrez®, and EMBL Nucleotide Sequence Database. Because a preferred embodiment of the present invention is its application to the immunogenicity reduction of immunoglobulins, a number of immunoglobulin databases may be useful for obtaining sequences, including but not limited to the Kabat database (Johnson & Wu, 2001, *Nucleic Acids Res* 29:205-206; Johnson & Wu, 2000, *Nucleic Acids Res* 28:214-218), the IMGT database (IMGT, the international ImMunoGeneTics information system®; Lefranc et al., 1999, *Nucleic Acids Res* 27:209-212; Ruiz et al., 2000 *Nucleic Acids Re.* 28:219-221; Lefranc et al., 2001, *Nucleic Acids Res* 29:207-209; Lefranc et al., 2003, *Nucleic Acids Res* 31:307-310), and VBASE.

[0110] As is well known in the art, immunoglobulins possess a high degree of sequence and structural homology, and

therefore alignment of sequences provides a wealth of information. Due to the existence of deletions and insertions in these alignments, numbering conventions have been adopted to enable a normalized reference to conserved positions in immunoglobulin families or subfamilies. Those skilled in the art will appreciate that these conventions consist of nonsequential numbering in specific regions of an immunoglobulin sequence, and thus accordingly the positions of any given immunoglobulin as defined by any given numbering scheme will not necessarily correspond to its sequential sequence or to those in an alternate numbering scheme. For all variable regions discussed in the present invention, numbering is according to the numbering scheme of Kabat (Kabat et al., 1991, *Sequences of Proteins of Immunological Interest,* 5th Ed., United States Public Health Service, National Institutes of Health, Bethesda). For all constant region positions discussed in the present invention, number is according to the EU index as in Kabat. Alternate numbering schemes may find use in the present invention, including but not limited that of Chothia (Chothia & Lesk, 1987, *J. Mol. Biol.* 196: 901-917; Chothia et al., 1989, *Nature* 342: 877-883; Al-Lazikani et al., 1997, *J. Mol. Biol.* 273: 927-948).

[0111] In a most preferred embodiment, the set of human sequences used is an aligned set of human germline immunoglobulin sequences. For example, FIGS. **1***a*-**1***c* (SEQ ID NOS: 1-109) provide the set of sequences that compose the human antibody variable region germline (VH, VL, and J chains), along with the corresponding diversity at each position. The human germline repertoire for immunoglobulin heavy chain variable regions and immunoglobulin light chain kappa variable regions have been reported (Matsuda et al., 1998, *J Exp Med* 188: 2151-2162; Zachau, 2000, Biol Chem 381:951-954; Pallares et al., 1999, *Exp Clin Immunogenet* 16(1): 36-60; Barbie & Lefranc, 1998, *Exp Clin Immunogenet* 15(3): 171-83). The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Barbie V, Lefranc M P). The rationale for use of this type of sequence information as a metric for humanness is that the strings that compose the human germline should be minimally immunogenic. Sequences need not be human genomic or germline sequences. In other preferred embodiments, human antibody variable region sequences are derived not from germline information, but rather from matured antibodies obtained for example from hybridoma technology or cDNA libraries.

[0112] For many of the genes in the human immunoglobulin germline, several different alleles have been identified. Although the polymorphisms detected in many of the alleles do not change the amino acid sequence of the gene, in a great number of cases the sequence is changed. In choosing a set of sequences to use in the method described herein, different sets of sequences may be chosen. When choosing a single allele as representative of a specific gene the most cautious approach is to choose that sequence which is closest to the consensus of the entire germline. This subset of sequences would thereby be most likely to be represented within the population as a whole. Alternatively, a much greater sequence diversity could be sampled by choosing representative sequences that are furthest from the consensus. Another approach yielding greater diversity would be to use multiple alleles where they exist for each germline. At this time, there is little or no quantitative data on allele frequency within the population. When allele frequency becomes available, a more informed decision can be made regarding the likelihood of tolerance for a specific non-consensus allele within the target patient population.

[0113] When two or more possible substitutions are being evaluated for use at a specific position when both are found in the human germline, the decision may become subjective. In such a case additional information can be incorporated that may reflect different levels of expression of particular genes (Cox et al. Eur J Immunol. 1994 April; 24(4):827-36). One underlying assumption of such a strategy would be that relative expression level of a particular germline (or corresponding sequence strings) correlates with the relative immunogenicity.

[0114] The sequences used for the method disclosed herein are those of homologous proteins with sufficient homology to allow their alignment with the protein whose immunogenicity is being reduced. One might argue that if a particular protein sequence is found anywhere within the expressed human genome that there is innate tolerance to that peptide. Such a proposition greatly increases the number of possible sequences that could be used to reduce the immunogenicity of a protein. In such a case however, alignment of proteins that are not structurally homologous would likely be prohibitive. In addition, the processing of a protein to produce the strings to which tolerance is developed may be structurally determined. Therefore, a specific strings may be nonimmunogenic in its native context but immunogenic in an altered structural context.

Scoring Functions—String Content

[0115] In order to evaluate the fitness of protein variants, amino acid modifications in the parent protein may be scored using a variety of scoring functions. Central to preferred embodiment of immunogenicity reduction method described herein is that at least one scoring function is aimed at maximizing the content of host linear sequence strings that are present in a set of host sequences. Typically, but not always, a computer is used to score potential amino acid substitutions.

[0116] In one embodiment, substitutions may be scored according to their occupancy in the set of host sequences, i.e., whether or not a given amino acid is part of the diversity at a given position. The use of position-specific alignment information to generate a list of considered amino acids at a variable position is well known in the art; see for example Lehmann & Wyss, 2001, *Curr Opin Biotechnol* 12(4): 371-5; Lehmann et al., 2000, *Biochim Biophys Acta* 1543(2):408-415; Rath & Davidson, 2000, *Protein Sci,* 9(12):2457-69; Lehmann et al., 2000, *Protein Eng* 13(1):49-57; Desjarlais & Berg, 1993, *Proc Natl Aced Sci USA* 90(6):2256-60; Desjarlais & Berg, 1992, *Proteins* 12(2):101-4; Henikoff & Henikoff, 2000, *Adv Protein Chem* 54:73-97; Henikoff & Henikoff, 1994, *J Mol Biol* 243(4):574-8. Thus, for example, for the parent nonhuman VLκ sequence aligned to the human sequences in FIG. **1***b* (SEQ ID NOS: 54-85), substitutions to be considered at position 1 would be Ala, Asp, Glu, Asn, and Val. In a more preferred embodiment, substitutions are scored based on their frequency in the set of human sequences listed. For example, in the previous example, Asp and Glu occur most frequently at position 1, and thus may be more preferable substitutions that Ala, Asn, or Val. The basis for this scoring function is that the frequency of a given amino acid at a given position in the alignment is proportional to its potential for being in a host string.

[0117] Occupancy and frequency provide relatively straightforward approximations for designing substitutions that have the potential for reduced immunogenicity. Their use, however, does not take into account the context of the parent sequence. Although frequency is proportional to the potential for a substitution to increase host content of a string, it is not a direct measure. In order to more accurately incorporate the information present in an aligned set of host sequences into a measure of immunogenicity, an approach can be taken wherein the linearity or contiguity of a given position in the context of the strings that comprise it is considered. In this most preferred embodiment, substitutions in a parent sequence are scored based on the probability of removing a nonhost string and replacing it with a less immunogenic string, namely one present in the set of host sequences. This method of scoring may employ the calculation of identity or percent identity of a parent string to a host string within a window of equivalent positions. In one embodiment, the identity of a string in sequence s to a host string in sequence h, (IDstring), can be presented as the sum of amino acid sequence identities in a given window size, according to equation 1:

$$IDstring(i, w) = \sum_{j=i}^{i+w-1} \delta_{aa^s_j, aa^h_j} \qquad \text{Equation 1}$$

where w is the string window size, i is the first position in the string, aa^s_j is the amino acid at position j of sequence s, aa^h_j is the amino acid at position j of the host sequence h, and the Kronecker delta function is used to return a value of 1 for a match (for example if the parent and host amino acids at position j are both serine) and 0 if there is no match (for example if the parent amino acid at position j is a serine but the host amino acid is a leucine). FIG. **2***a* illustrates equation 1 using a region of the VH of murine anti-Her2 antibody m4D5 (VH_m4D5) (SEQ ID NO: 110} as the parent sequence s and the homologous region from the VH human germline (VH_1-2) as human sequence (SEQ ID NO:111)

[0118] In a further embodiment, it is assumed that the most immunologically appropriate measure of host string content at position i is the maximal identity between a string of sequence s and any host sequence in the alignment, as calculated in equation 2:

$$IDmax = \max_{h \in HS} \left(\sum_{j=i}^{i+w-1} \delta_{aa^s_j, aa^h_j} \right) \qquad \text{Equation 2}$$

where HS is the set of host sequences. In other words, if IDstring at position i is equal to w for any one of the host sequences, IDmax=w as well, and the i^{th} string is assumed to be minimally immunogenic. The concept of the IDmax quantity represented by Equation 2 is illustrated in FIG. **2***b*.

[0119] Finally, these equations can be combined to calculate a single numerical metric for total host string content (HSC) of a sequence s by summing the IDmax values over all pertinent sequence positions, as in equation 3:

$$HSC(s) = 100 \cdot \frac{1}{(L-w+1) \cdot w} \sum_{i=1}^{L-w+1} \max_{h \in HS} \left(\sum_{j=i}^{i+w-1} \delta_{aa^s_j, aa^h_j} \right) \qquad \text{Equation 3}$$

where L is the length of the sequence and HS is the set of host sequences in the alignment. A perfectly host sequence would have an HSC of 100. One might alternatively say that such a sequence is 100% host. The concept of the HSC quantity represented by Equation 3 is illustrated in FIG. **2***c*. In alternative embodiments, Equation 3 can be modified further such that the final score is dependent on the relative usage of each host sequence in the alignment. Strings from sequences that are more frequently expressed by hosts are expected to be more tolerized, and therefore may be given correspondingly higher influence in a scoring system.

[0120] In an alternative embodiment, one can measure the exact string content (ESC) as in Equation 3a:

$$ESC(s) = 100 \cdot \frac{1}{(L-w+1) \cdot w} \sum_{i=1}^{L-w+1} \max_{h \in HS} \delta_{aa^s_{i \ldots i+w-1}, aa^h_{i \ldots i+w-1}}, \qquad \text{Equation 3a}$$

where the notation $aa^s_{i,i+w-1}$ refers the contiguous sequence string in protein s from position i to position i+w−1. In this embodiment, only perfect matches of size w are counted in the score.

[0121] It is worth noting that, since the scoring systems in Equations 3 and 3a are based on local sequence identity and/or similarity evaluated over windows of defined size, a sequence with high HSC can be constructed of sequence segments that are maximally similar to different members of the set of host sequences at different positions.

[0122] The above measure of hostness is likely to be more immunologically relevant than the more commonly used global identity measure of equation 4:

$$globalID = 100 \cdot \max_{h \in HS} \frac{1}{L} \sum_{i=1}^{L} \delta_{aa^s_i, aa^h_i} \qquad \text{Equation 4}$$

Equation 4 disregards the extent of contiguous sequence identity, which is particularly relevant for capturing the molecular behavior of the immune system.

[0123] Additional scoring functions similar to equations 3 are also possible. For example, as will be appreciated by those skilled in the art, there is some uncertainty regarding the hostness of a string wherein IDmax=w−1, w−2, etc. In one alternative embodiment, sequence similarity is compared instead of identity, using any of a variety of amino acid substitution matrices (e.g. PAM, BLOSUM62, etc.), providing a host string similarity (HSS) score as in equation 5:

$$HSS(s) = 100 \cdot \frac{1}{L-w+1} \sum_{i=1}^{L-w+1} \max_{h \in HS} \left(e^{\sum_{j=i}^{i+w-1} \left(S_{aa^s_j, aa^h_j} - S_{aa^s_j, aa^s_j} \right)} \right) \qquad \text{Equation 5}$$

where S is a substitution score comparing any two amino acids. In yet another alternative, sequence identities are weighted according to the extent of identity, as in equation 6:

$$HSC(s) = 100 \cdot \frac{1}{(L-w+1)} \sum_{i=1}^{L-w+1} f\left(\max_{h \in HS} \left(\sum_{j=i}^{i+w-1} \delta_{aa_j^s, aa_j^h} \right) \right) \quad \text{Equation 6}$$

where f is a continuous or noncontinuous function dependent on IDmax. For example, perfect matches can be weighted greater than near perfect matches (e.g. f(w)=1, f(w−1)=0.5, etc.), and poor matches can be discarded (e.g. f(w−3)=f(w−2)=0).

String Window Size

[0124] The fundamental binding units of class I and class II MHC proteins are both 9 amino acids. In a preferred embodiment, the window size w used to create and score parent sequences is 9. However, it is also known that additional peptide flanking residues (PFRs) can influence T-cell recognition (via the TCR) of class II MHC-peptide complexes (see for example Arnold et al., 2002, *J Immunology* 169(2): 739-49), with the residues at positions P-1 (one position before the 1st MHC binding position) and P11 being most influential. Because these effects might influence immune tolerance, a desirable goal of the invention, larger window sizes (e.g. 12) can be used. It should be noted however, that sequences optimized with similar window sizes are highly correlated.

Optimization of HSC

[0125] Although a definition of string scoring systems is useful, an efficient process for discovering sequences with high HSC is also desirable. It is therefore a further aspect of the invention to provide methods for dynamic optimization of HSC given the described scoring systems.

[0126] Desirable features of an optimization method include but are not limited to the following: 1) the output sequences are optimal or near-optimal (subject to design constraints) in their host string content; 2) structural constraints can be used to modulate the nature of the optimized sequences; and 3) multiple near-optimal solutions can be generated. Additionally, in some preferred embodiments, host string content may be maximized using a minimal number of substitutions.

[0127] In a preferred embodiment, an iterative algorithm for optimization of HSC works as follows. 1) a parent sequence and set of host sequences are defined; 2) mutational constraints are defined at functionally or structurally important positions, referred to herein as masking—in a preferred embodiment, for antibody applications, positions within or structurally proximal to CDR residue (as defined by herein, or alternatively as defined by Kabat or Chothia) and/or interface are masked, locked, or fixed so that mutations are not possible (in some embodiments this constraint can be relaxed if the potential mutation is a conservative substitution of the parent amino acid). In a preferred embodiment, positions within 5 angstroms of a CDR residue or interface are masked. In other preferred embodiments, positions within 6.5 angstroms of a CDR residue or interface are masked; 3) host sequence segments (up to a defined length: lengths from 1-6 are typical) are collected from the alignment and stored for each position: segments that violate the mutational constraints are not collected; 4) each segment is analyzed for its potential impact on HSC, in the context of the current parent sequence, defined as String Impact (SI) in equation 7:

$$SI(x_m(z) \rightarrow y_m(z)) = HSC(s(y_m)) - HSC(\text{parent}), \quad \text{Equation 7}$$

where $y_m(z)$ is a host segment of length z replacing segment x at position m, and $s(y_m)$ and parent are versions of the parent sequence that include these segments (the parent sequence contains $x_m(z)$). 5) a single string is randomly selected from all stored host strings; the probability of selection is biased and proportional to the impact on HSC and inversely proportional to the number of mutations relative to the current parent sequence, as in Equation 8:

$$P \propto \frac{SI(x_m(z) \rightarrow y_m(z))}{\sum_{i=m}^{m+z-1} \delta_{aa_i^{parent}, aa_i^{s(y_m)}}} \quad \text{Equation 8}$$

This selected string is substituted into the current parent sequence for its corresponding parent amino acids on an amino acid string by amino acid string basis. This kind of selection bias tends to optimize host string content with minimal perturbation of the original sequence. 6) steps 4 and 5 are repeated until no further optimization is possible (no segment substitutions have a favorable impact on host string content).

[0128] Such an algorithm is inherently non-deterministic, so independent runs of the algorithm will tend to generate different solutions (this is a favorable feature). In a preferred embodiment, such an algorithm is applied numerous times to generate a diverse array of unique solutions. These solutions can be further clustered such that representative sequences can be prioritized for further analysis. For example, in one embodiment, the solution sequences are clustered into groups of similar sequences according to mutational distance, using a nearest neighbor single linkage hierarchical clustering algorithm to assign sequences to related groups based on similarity scores. Clustering algorithms may be useful for classifying sequences into representative groups. Representative groups may be defined, for example, by similarity. Measures of similarity include, but are not limited to sequence similarity and energetic similarity. Thus the output sequences from computational screening may be clustered around local minima, referred to herein as clustered sets of sequences. Sets of sequences that are close in sequence space may be distinguished from other sets. In one embodiment, diversity across clustered sets of sequences may be sampled by experimentally testing only a subset of sequences within each clustered set. For example, all or most of the clustered sets could be broadly sampled by including the lowest energy sequence from each clustered set of sequences to be experimentally tested. Because the sequence space of solutions with optimized HSC can be large, additional methods can be applied to ensure that a broad set of sequences is created. In a preferred embodiment, individual framework sequences generated by the procedure are clustered separately to generate a list of nonredundant basis framework regions (FRs) with high HSC. These basis FRs are then computationally assembled in all combinations along with the CDRs to generate a secondary list of solution sequences (which will usually have some

overlap with the primary set). Alternatively, the basis FRs may be combined into an experimental library, for example a combinatorial library.

Framework Diversity

[0129] Application of this algorithm will generate variant protein solutions for which HSC is higher than the original parent sequence. It will also frequently generated solutions in which substituted strings are derived from different members of the alignment. The variant sequences derived using the present invention generally have unique properties relative to sequences generated using other methodologies. For example, in the context of an antibody, the protein variants of the invention frequently derive their host string content from a combination of different host germline sequences. This may be true even within a single FR. Quantification of these properties is useful for defining the nature of sequences derived using the present invention. A clear distinction emerges from a comparison of exact string content (meaning a perfect match over window w) in any single germline sequence versus exact w-mer string content within the set of all germline sequences (content of strings for which IDmax=w). Single germline exact string content (SGESC) of a variant sequence v may thus be defined as:

$$SGESC(v) = 100 \cdot \frac{1}{L-w+1} \cdot \max_{h \in HS} \sum_{i=1}^{L-w+1} \delta_{aa_i^v \ldots i+w-1, aa_i^h \ldots i+w-1} \quad \text{Equation 9}$$

[0130] This quantity provides the extent to which a string-optimized sequence has string identity with the closest single germline sequence. Using this definition, it is also possible to assess the extent to which the high host string content of a given variant sequence v is derived from a single germline as opposed to multiple germline sequences. Framework region homogeneity (FRH) is defined as follows:

$$FRH(v) = \frac{SGESC(v)}{ESC(v)} = \frac{\max_{h \in HS} \sum_{i=1}^{L-w+1} \delta_{aa_i^v \ldots i+w-1, aa_i^h \ldots i+w-1}}{\sum_{i=1}^{L-w+1} \max_{h \in HS} \delta_{aa_i^s \ldots i+w-1, aa_i^h \ldots i+w-1}} \quad \text{Equation 10}$$

[0131] In other words, if a variant sequence's exact string content is derived solely from a single germline sequence, the FRH would be close to 1.0. It should be noted that a similar or identical quantity can be defined for non-antibody proteins. Alternatively, as is the case with many of the variant sequences created by the present invention, FRH values can be significantly less than 1, with values ranging from 0.4 to 1.0, indicating, as expected, that sequences with high exact string content can be discovered with contributions from multiple germline subfamilies and sequences. As described more fully in Example 5 below, variant sequences generated using the present invention have high HSC values yet many have low FRH values, indicating their HSC is derived from multiple germline frameworks.

Additional Scoring

[0132] The above methods of scoring use the information present in an aligned set of host sequences as a metric of immunogenicity to maximize the content of host linear sequence strings in a parent sequence. In addition to such scoring functions, other scoring functions and methods may be employed. Such additional scoring functions may be aimed at the same goal as the aforementioned linear string scoring function, namely immunogenicity reduction of the parent protein. Alternatively, such additional scoring functions may be used to achieve other goals, for example optimization of protein stability, solubility, expression, pharmacokinetics, and/or aspects of protein function such as affinity of the parent protein for a target ligand, specificity, effector function, and/or enzymatic activity. For example an additional scoring function may be employed to enhance the affinity of an antibody variable domain for its target antigen. Such additional scoring functions may be employed statically or dynamically for the generation of optimized protein variants. A number of embodiments are described below as preferred additional scoring functions that may be used with the aforementioned linear string scoring method of the present invention. However, these are not meant to constrain the invention to these embodiments, and it should be clear that any method of scoring the fitness of an amino acid modification in a parent protein may be coupled with the novel linear string scoring method of the present invention so that optimal protein variants may be designed.

[0133] In a preferred embodiment, substitutions are scored based on their structural compatibility with the structure of the parent protein. Such methods of scoring may require the structural coordinates that describe the three-dimensional structure of the protein, for example as obtained by X-ray crystallographic and nuclear magnetic resonance (NMR) techniques. Suitable proteins structures may also be obtained from structural models, which may be generated by methods that are known in the art of structural biology, including but not limited to de novo and homology modeling. Structure-based scoring functions may include any number of potentials that describe or approximate physical or chemical energy terms, including but not limited to a van der Waals potential, a hydrogen bond potential, an atomic solvation potential or other solvation models, a secondary structure propensity potential, an electrostatic potential, a torsional potential, an entropy potential, and/or additional energy terms. In other preferred embodiments, scoring methods may also be derived from sequence information, including but not limited to knowledge-based potentials derived from protein sequence and/or structure statistics, threading potentials, reference energies, pseudo energies, homology-based energies, and sequence biases derived from sequence alignments. In alternately preferred embodiments, both structural and sequence-based potentials are used to generate one or more scoring functions that may be coupled with the linear string scoring method of the present invention.

[0134] In a most preferred embodiment, a scoring method is used wherein the structural and functional integrity of substitutions are evaluated using a sequence and structure-based scoring function described in U.S. Ser. No. 60/528,229, filed Dec. 8, 2003, entitled Protein Engineering with Analogous Contact Environments; and U.S. Ser. No. 60/602,566, filed Aug. 17, 2004, entitled Protein Engineering with Analogous Contact Environments. This method combines sequence alignment information and structural information to predict the structural compatibility of one or more substitutions with a protein structure template. Nearest neighbor structure-based scores generated by this method include Structural

Consensus and Structural Precedence as provided in the Examples. This method is particularly well suited for application to evaluating the structural fitness of immunoglobulins due to their substantial sequence and structural homology.

[0135] In a preferred embodiment, substitutions are scored using a scoring function or computational design program that is substantially similar to Protein Design Automation® (PDA®) technology, as is described in U.S. Pat. No. 6,188,965; U.S. Pat. No. 6,269,312; U.S. Pat. No. 6,403,312; U.S. Pat. No. 6,708,120; U.S. Pat. No. 6,804,611; U.S. Pat. No. 6,792,356; U.S. Ser. No. 09/782,004; U.S. Ser. No. 09/812,034; U.S. Ser. No. 09/927,790; U.S. Ser. No. 10/218,102; U.S. Ser. No. 10/101,499; U.S. Ser. No. 10/218,102; U.S. Ser. No. 10/666,311; U.S. Ser. No. 10/665,307; U.S. Ser. No. 10/888,748; PCT WO 98/07254; PCT WO 99/24229; PCT WO 01/40091; and PCT WO 02/25588. In another preferred embodiment, a computational design method substantially similar to Sequence Prediction Algorithm™ (SPA™) technology is used, as is described in (Raha et al., 2000, *Protein Sci.* 9: 1106-1119), U.S. Ser. No. 09/877,695, and U.S. Ser. No. 10/071,859. In another preferred embodiment, the computational methods described in U.S. Ser. No. 10/339,788, are used.

[0136] In another preferred embodiment, optimized sequences are also assessed for surface similarity with host antibodies. Ensuring similarity may be important for reducing the probability of introducing novel 3D epitopes, which are potentially recognized by B-cell receptors. In a preferred embodiment, surface similarity at position i is quantified as follows:

$$surfscore(i) = \max_{k \in HS} e^{f_i^{exp}\left(\frac{\left(\sum_{j=1}^{L} proximity(i,j) * S(aa_j^S, aa_j^k)\right) -}{\left(\sum_{j=1}^{L} proximity(i,j) * S(aa_j^S, aa_j^S)\right)}\right)/T} \quad \text{Equation 11}$$

where f_i^{exp} is the fraction accessibility of position i to solvent, proximity(i,j) is the spatial proximity of positions i and j in the three-dimensional structure of the protein, S is a measure of amino acid similarity, and T is a temperature factor used to tune the stringency of the similarity comparison. It will be appreciated from the equation that if sequences are identical in the region of position i, a surfscore of 1.0 will be approached. Alternatively, a score of 1.0 can also be achieved if a position is completely buried (i.e. $f_i^{exp}=0$), since the position would not be accessible to B-cell receptors. Lower scores represent surface positions for which there are significant differences between the variant sequence and the most similar host sequence. In a preferred embodiment, the proximity between two positions is inversely related to their distance (e.g. a Gaussian or exponential function of the distance), and the proximity of a position to itself is 1.0. In a preferred embodiment, the decay of the proximity function is tuned such that patches of positions correspond to the size of a typical antibody epitope.

[0137] Other surface properties may also be desirable. For example, optimized variant sequences may be assessed for the exposure of nonpolar amino acids, which is generally expected to decrease solubility. In such cases, variant sequences with lower nonpolar exposure can be prioritized over alternatives. Surface electrostatic properties may also be assessed for variant sequences. In a preferred embodiment, surface properties are assessed for multiple variants with optimized HSC, in order to limit the set of variants that will be experimentally screened.

[0138] In another embodiment, substitution matrices or other knowledge-based scoring methods are used to identify alternate sequences that are likely to retain the structure and function of the protein. Such scoring methods can be used to quantify how conservative a given substitution or set of substitutions is. In most cases, conservative mutations do not significantly disrupt the structure and function of proteins (see for example, Bowie et. al., 1990, *Science* 247: 1306-1310, Bowie & Sauer, 1989, *Proc. Nat. Acad. Sci. USA* 86: 2152-2156, and Reidhaar-Olson & Sauer, 1990, *Proteins* 7: 306-316). However, non-conservative mutations can destabilize protein structure and reduce activity (see for example, Lim et al., 1992, *Biochem.* 31: 4324-4333). Substitution matrices including but not limited to BLOSUM62 provide a quantitative measure of the compatibility between a sequence and a target structure, which can be used to predict non-disruptive substitution mutations (Topham et al., 1997, *Prot. Eng.* 10: 7-21). The use of substitution matrices to design peptides with improved properties has been disclosed (Adenot et al., 1999, *J. Mol. Graph. Model.* 17: 292-309). Substitution matrices include, but are not limited to, the BLOSUM matrices (Henikoff & Henikoff, 1992, *Proc. Nat. Acad. Sci. USA* 89: 10917, the PAM matrices, the Dayhoff matrix, and the like. For a review of substitution matrices, see for example Henikoff, 1996, *Curr. Opin. Struct. Biol.* 6: 353-360. It is also possible to construct a substitution matrix based on an alignment of a given protein of interest and its homologs; see for example Henikoff & Henikoff, 1996, *Comput. Appl. Biosci.* 12: 135-143.

[0139] In a preferred embodiment, other methods for scoring immunogenicity may additionally be used. Most preferably, immunogenicity may be scored using a function that considers peptide binding to one or more MHC molecules. For example, substitutions would be scored such that there are no or a minimal number of immune epitopes that are predicted to bind, with high affinity, to any prevalent MHC alleles. These methods of scoring may be useful, for example, for designing substitutions in VH CDR3, for which scoring using human germline strings may be less straightforward. Several methods of identifying MHC-binding epitopes in protein sequences are known in the art and may be used to score epitopes in an antibody. See for example WO 98/52976; WO 02/079232; WO 00/3317; U.S. Ser. No. 09/903,378; U.S. Ser. No. 10/039,170; U.S. Ser. No. 60/222,697; U.S. Ser. No. 10/339,788; PCT WO 01/21823; and PCT WO 02/00165; Mallios, 1999, *Bioinformatics* 15: 432-439; Mallios, 2001, *Bioinformatics* 17: 942-948; Sturniolo et al., 1999, *Nature Biotech.* 17: 555-561; WO 98/59244; WO 02/069232; WO 02/77187; Marshall et al., 1995, *J. Immunol.* 154: 5927-5933; and Hammer et al., 1994, *J. Exp. Med.* 180: 2353-2358. Sequence-based information can be used to determine a binding score for a given peptide-MHC interaction (see for example Mallios, 1999, *Bioinformatics* 15: 432-439; Mallios, 2001, *Bioinformatics* 17: p942-948; Sturniolo et. al., 1999, *Nature Biotech.* 17: 555-561). It is possible to use structure-based methods in which a given peptide is computationally placed in the peptide-binding groove of a given MHC molecule and the interaction energy is determined (for example, see WO 98/59244 and WO 02/069232). Such methods may be referred to as "threading" methods. Alternatively, purely experimental methods can be used; for example a set of

overlapping peptides derived from the protein of interest can be experimentally tested for the ability to induce T-cell activation and/or other aspects of an immune response. (see for example WO 02/77187). In a preferred embodiment, MHC-binding propensity scores are calculated for each 9-residue frame along the protein sequence using a matrix method (see Sturniolo et. al., supra; Marshall et. al., 1995, *J. Immunol.* 154: 5927-5933, and Hammer et. al., 1994, *J. Exp. Med.* 180: 2353-2358). It is also possible to consider scores for only a subset of these residues, or to consider also the identities of the peptide residues before and after the 9-residue frame of interest. The matrix comprises binding scores for specific amino acids interacting with the peptide binding pockets in different human class II MHC molecule. In the most preferred embodiment, the scores in the matrix are obtained from experimental peptide binding studies. In an alternate preferred embodiment, scores for a given amino acid binding to a given pocket are extrapolated from experimentally characterized alleles to additional alleles with identical or similar residues lining that pocket. Matrices that are produced by extrapolation are referred to as "virtual matrices".

[0140] In alternate embodiments, additional scoring functions are employed that predict reactive sites within a protein, such as deamidation sites, glycosylation sites, oxidation sites, protealytic cleavage sites, and the like.

[0141] It will be appreciated by one of skill in the art that the use of combinations of any of the aforementioned scoring functions and/or other scoring functions is contemplated. In one embodiment, this could be accomplished by evaluating the outputs of the results from separate calculations. Alternatively, scoring functions may be combined into one scoring term. This latter strategy enables different scoring terms to be weighted separately, thus providing more control over the relative contributions of the scoring terms, and a greater capacity to tune the scoring function for a desired engineering strategy.

Additional Optimization of Sequences

[0142] In a preferred embodiment, after optimized protein variants have been engineered using the aforementioned scoring functions, additional optimization of protein variants may be carried out. In this way, an optimized protein variant can be thought of as primary variant or template for further optimization, and variants of this primary variant can be thought of as secondary variants. Because variant sequences of the invention are preferably derived from a HSC-increasing procedure in which substitution of structurally important positions is disallowed or discouraged (for example masking), it is likely that additional optimization of HSC is possible if those positions are allowed to vary in a secondary analysis. Optimization of other properties is possible, including but are not limited to protein affinity, expression, specificity, solubility, activity, and effector function. Thus the variant sequences derived in the primary analysis can represent variants for further optimization. In a preferred embodiment, the secondary analysis comprises the steps of 1) string analysis of the template sequence to identify secondary amino acid diversity that will have neutral, positive, or minimal impact on HSC; 2) experimental production of secondary variant sequences using the diversity derived in step 1; and 3) experimental screening of secondary variant sequences.

[0143] For these purposes, the string impact (SI) of a single substitution at position m from amino acid x to amino acid y can be quantified as in Equation 7 (where segment length z=1). As will be appreciated, the maximum possible string increase (for a single substitution at an internal position) is w and the maximum possible decrease is w.

[0144] In a most preferred embodiment, substitutions in a primary variant are chosen as those substitutions that will result in zero or positive string impact, referred to herein as string neutral and string positive substitutions respectively. In other embodiments, substitutions that result in negative string impact, i.e. string negative substitutions, may also be considered for engineering secondary variants. Secondary variants then may be constructed, expressed, and tested experimentally. Secondary substitutions that show favorable properties with respect to antigen affinity, effector function, stability, solubility, expression, and the like, may be combined in subsequent variants to generate a more optimized therapeutic candidate.

Optimization of Non-Xenogeneic Proteins

[0145] It can be appreciated that the optimization described above is not restricted to immunogenicity reduction of xenogeneic proteins. Evaluation of potential substitutions for string impact provides an excellent strategy for generating substitution diversity for engineering protein variants with optimized properties. A clear advantage of this approach is that it generates protein variants with minimal immunogenicity risk, the importance of which has been discussed extensively and is a primary goal of the present invention. An additional advantage of this approach is that because the sequences being used to evaluate string impact are typically derived from a set of naturally evolved host sequences, variants designed are effectively enriched for stability, solubility, and other favorable properties. The utility of this capability lies in the fact that there are innumerable amino acid modifications that are detrimental or deleterious to proteins. By screening a quality set of variant diversity, the chances are increased that a protein variant of the desired property will be obtained. The capacity of the string impact approach to generate a quality set of variant diversity derives from the greater tolerance to mutation of positions which sample greater diversity, and the greater propensity of amino acids in a set of naturally evolved sequences to be compatible with a homologous protein's structure, stability, solubility, function, and the like.

[0146] This string impact approach to variant design may be applied not only to the generation of secondary variants as described above, but may also be used to engineer amino acid modifications in proteins that are presumably already minimally immunogenic. This may include, for example, natural host proteins. Alternatively, and in a preferred embodiment, the string impact strategy may be applied to engineer modifications in an antibody variable region (VH or VL) that is humanized (Clark, 2000, *Immunol Today* 21:397-402), or "fully human" as obtained for example using transgenic mice (Bruggemann et al., 1997, *Curr Opin Biotechnol* 8:455-458) or human antibody libraries coupled with selection methods (Griffiths et al., 1998, *Curr Opin Biotechnol* 9:102-108). As with optimization of primary variant sequences described above, the string impact analysis described here can be used to identify secondary diversity that will have neutral, positive, or minimal impact on HSC, as well as potentially other favorable properties. Such diversity can then be used to screen for optimized versions of these sequences without increasing the risk of immunogenicity.

Experimental Production, Screening, and Testing

[0147] Methods for production and screening of protein variants are well known in the art. General methods for anti-

body molecular biology, expression, purification, and screening are described in Antibody Engineering, edited by Duebel & Kontermann, Springer-Verlag, Heidelberg, 2001; and Hayhurst & Georgiou, 2001, *Curr Opin Chem Biol* 5:683-689; Maynard & Georgiou, 2000, *Annu Rev Biomed Eng* 2:339-76. Also see the methods described in U.S. Ser. No. 10/339,788, filed on Mar. 3, 2003, U.S. Ser. No. 10/672,280, filed Sep. 29, 2003, and U.S. Ser. No. 10/822,231, filed Mar. 26, 2004.

[0148] In one embodiment of the present invention, the library sequences are used to create nucleic acids that encode the member sequences, and that may then be cloned into host cells, expressed and assayed, if desired. These practices are carried out using well-known procedures, and a variety of methods that may find use in the present invention are described in Molecular Cloning—A Laboratory Manual, 3rd Ed. (Maniatis, Cold Spring Harbor Laboratory Press, New York, 2001), and Current Protocols in Molecular Biology (John Wiley & Sons). The nucleic acids that encode the protein variants of the present invention may be incorporated into an expression vector in order to express the protein. Expression vectors typically comprise a protein operably linked, that is placed in a functional relationship, with control or regulatory sequences, selectable markers, any fusion partners, and/or additional elements. The protein variants of the present invention may be produced by culturing a host cell transformed with nucleic acid, preferably an expression vector, containing nucleic acid encoding the protein variants, under the appropriate conditions to induce or cause expression of the protein. A wide variety of appropriate host cells may be used, including but not limited to mammalian cells, bacteria, insect cells, and yeast. For example, a variety of cell lines that may find use in the present invention are described in the ATCC cell line catalog, available from the American Type Culture Collection. The methods of introducing exogenous nucleic acid into host cells are well known in the art, and will vary with the host cell used.

[0149] In a preferred embodiment, protein variants are purified or isolated after expression. Proteins may be isolated or purified in a variety of ways known to those skilled in the art. Standard purification methods include chromatographic techniques, electrophoretic, immunological, precipitation, dialysis, filtration, concentration, and chromatofocusing techniques. As is well known in the art, a variety of natural proteins bind antibodies, for example bacterial proteins A, G, and L, and these proteins may find use in the present invention for purification. Purification can often be enabled by a particular fusion partner. For example, proteins may be purified using glutathione resin if a GST fusion is employed, Ni^{+2} affinity chromatography if a His-tag is employed, or immobilized anti-flag antibody if a flag-tag is used. For general guidance in suitable purification techniques, see Protein Purification: Principles and Practice, 3r^{d} Ed., Scopes, Springer-Verlag, NY, 1994.

[0150] Protein variants may be screened using a variety of methods, including but not limited to those that use in vitro assays, in vivo and cell-based assays, and selection technologies. Automation and high-throughput screening technologies may be utilized in the screening procedures. Screening may employ the use of a fusion partner or label, for example an immune label, isotopic label, or small molecule label such as a fluorescent or colorimetric dye.

[0151] In a preferred embodiment, the functional and/or biophysical properties of protein variants are screened in an in vitro assay. In a preferred embodiment, the protein is screened for functionality, for example its ability to catalyze a reaction or its binding affinity to its target. Binding assays can be carried out using a variety of methods known in the art, including but not limited to FRET (Fluorescence Resonance Energy Transfer) and BRET (Bioluminescence Resonance Energy Transfer)-based assays, AlphaScreen™ (Amplified Luminescent Proximity Homogeneous Assay), Scintillation Proximity Assay, ELISA (Enzyme-Linked Immunosorbent Assay), SPR (Surface Plasmon Resonance, also known as BIACORE®), isothermal titration calorimetry, differential scanning calorimetry, gel electrophoresis, and chromatography including gel filtration. These and other methods may take advantage of some fusion partner or label. Assays may employ a variety of detection methods including but not limited to chromogenic, fluorescent, luminescent, or isotopic labels. The biophysical properties of proteins, for example stability and solubility, may be screened using a variety of methods known in the art. Protein stability may be determined by measuring the thermodynamic equilibrium between folded and unfolded states. For example, protein variants of the present invention may be unfolded using chemical denaturant, heat, or pH, and this transition may be monitored using methods including but not limited to circular dichroism spectroscopy, fluorescence spectroscopy, absorbance spectroscopy, NMR spectroscopy, calorimetry, and proteolysis. As will be appreciated by those skilled in the art, the kinetic parameters of the folding and unfolding transitions may also be monitored using these and other techniques. The solubility and overall structural integrity of a protein variant may be quantitatively or qualitatively determined using a wide range of methods that are known in the art. Methods which may find use in the present invention for characterizing the biophysical properties of protein variants include gel electrophoresis, chromatography such as size exclusion chromatography and reversed-phase high performance liquid chromatography, mass spectrometry, ultraviolet absorbance spectroscopy, fluorescence spectroscopy, circular dichroism spectroscopy, isothermal titration calorimetry, differential scanning calorimetry, analytical ultra-centrifugation, dynamic light scattering, proteolysis, and cross-linking, turbidity measurement, filter retardation assays, immunological assays, fluorescent dye binding assays, protein-staining assays, microscopy, and detection of aggregates via ELISA or other binding assay. Structural analysis employing X-ray crystallographic techniques and NMR spectroscopy may also find use.

[0152] In a preferred embodiment, protein variants are screened using one or more cell-based or in vivo assays. For such assays, purified or unpurified proteins are typically added exogenously such that cells are exposed to individual variants or pools of variants belonging to a library. These assays are typically, but not always, based on the function of the protein; that is, the ability of the protein to bind to its target and mediate some biochemical event, for example effector function, ligand/receptor binding inhibition, apoptosis, and the like. Such assays often involve monitoring the response of cells to the protein, for example cell survival, cell death, change in cellular morphology, or transcriptional activation such as cellular expression of a natural gene or reporter gene. For example, such assays may measure the ability of antibody variants to elicit ADCC, ADCP, or CDC. For some assays additional cells or components, that is in addition to the target cells, may need to be added, for example serum complement, or effector cells such as peripheral blood monocytes (PB-

MCs), NK cells, macrophages, and the like. Such additional cells may be from any organism, preferably humans, mice, rat, rabbit, and monkey. Proteins may cause apoptosis of certain cell lines expressing the target, or they may mediate attack on target cells by immune cells which have been added to the assay. Methods for monitoring cell death or viability are known in the art, and include the use of dyes, immunochemical, cytochemical, and radioactive reagents. For example, caspase staining assays may enable apoptosis to be measured, and uptake or release of radioactive substrates or fluorescent dyes such as alamar blue may enable cell growth or activation to be monitored. In a preferred embodiment, the DELFIA® EuTDA-based cytotoxicity assay (Perkin Elmer, Mass.) is used. Alternatively, dead or damaged target cells may be monitored by measuring the release of one or more natural intracellular proteins, for example lactate dehydrogenase. Transcriptional activation may also serve as a method for assaying function in cell-based assays. In this case, response may be monitored by assaying for natural genes or proteins which may be upregulated, for example the release of certain interleukins may be measured, or alternatively readout may be via a reporter construct. Cell-based assays may also involve the measure of morphological changes of cells as a response to the presence of a protein. Cell types for such assays may be prokaryotic or eukaryotic, and a variety of cell lines that are known in the art may be employed. Alternatively, cell-based screens are performed using cells that have been transformed or transfected with nucleic acids encoding the variant proteins. That is, protein variants are not added exogenously to the cells. For example, in one embodiment, the cell-based screen utilizes cell surface display. A fusion partner can be employed that enables display of variants on the surface of cells (Witrrup, 2001, *Curr Opin Biotechnol,* 12:395-399).

[0153] As is known in the art, a subset of screening methods are those that select for favorable members of a library. The methods are herein referred to as "selection methods", and these methods find use in the present invention for screening protein variants. When protein libraries are screened using a selection method, only those members of a library that are favorable, that is which meet some selection criteria, are propagated, isolated, and/or observed. As will be appreciated, because only the most fit variants are observed, such methods enable the screening of libraries that are larger than those screenable by methods that assay the fitness of library members individually. Selection is enabled by any method, technique, or fusion partner that links, covalently or noncovalently, the phenotype of a protein with its genotype, i.e., the function of a protein with the nucleic acid that encodes it. For example the use of phage display as a selection method is enabled by the fusion of library members to the gene III protein. In this way, selection or isolation of protein variants that meet some criteria, for example binding affinity to the protein's target, also selects for or isolates the nucleic acid that encodes it. Once isolated, the gene or genes encoding variants may then be amplified. This process of isolation and amplification, referred to as panning, may be repeated, allowing favorable protein variants in the library to be enriched. Nucleic acid sequencing of the attached nucleic acid ultimately allows for gene identification.

[0154] A variety of selection methods are known in the art that may find use in the present invention for screening protein libraries. These include but are not limited to phage display (Phage display of peptides and proteins: a laboratory manual, Kay et al., 1996, Academic Press, San Diego, Calif., 1996; Lowman et al., 1991, *Biochemistry* 30:10832-10838; Smith, 1985, *Science* 228:1315-1317) and its derivatives such as selective phage infection (Malmborg et al., 1997, J Mol Biol 273:544-551), selectively infective phage (Krebber et al., 1997, *J Mol Biol* 268:619-630), and delayed infectivity panning (Benhar et al., 2000, *J Mol Biol* 301:893-904), cell surface display (Witrrup, 2001, *Curr Opin Biotechnol,* 12:395-399) such as display on bacteria (Georgiou et al., 1997, *Nat Biotechnol* 15:29-34; Georgiou et al., 1993, *Trends Biotechnol* 11:6-10; Lee et al., 2000, *Nat Biotechnol* 18:645-648; Jun et al., 1998, *Nat Biotechnol* 16:576-80), yeast (Boder & Wittrup, 2000, *Methods Enzymol* 328:430-44; Boder & Wittrup, 1997, *Nat Biotechnol* 15:553-557), and mammalian cells (Whitehorn et al., 1995, *Bio/technology* 13:1215-1219), as well as in vitro display technologies (Amstutz et al., 2001, *Curr Opin Biotechnol* 12:400-405) such as polysome display (Mattheakis et al., 1994, *Proc Natl Acad Sci USA* 91:9022-9026), ribosome display (Hanes et al., 1997, *Proc Natl Acad Sci USA* 94:4937-4942), mRNA display (Roberts & Szostak, 1997, *Proc Natl Acad Sci USA* 94:12297-12302; Nemoto et al., 1997, *FEBS Lett* 414:405-408), and ribosome-inactivation display system (Zhou et al., 2002, *J Am Chem Soc* 124, 538-543).

[0155] Other selection methods that may find use in the present invention include methods that do not rely on display, such as in vivo methods including but not limited to periplasmic expression and cytometric screening (Chen et al., 2001, *Nat Biotechnol* 19:537-542), the protein fragment complementation assay (Johnsson & Varshaysky, 1994, *Proc Natl Acad Sci USA* 91:10340-10344; Pelletier et al., 1998, *Proc Natl Acad Sci USA* 95:12141-12146), and the yeast two hybrid screen (Fields & Song, 1989, *Nature* 340:245-246) used in selection mode (Visintin et al., 1999, *Proc Natl Acad Sci USA* 96:11723-11728). In an alternate embodiment, selection is enabled by a fusion partner that binds to a specific sequence on the expression vector, thus linking covalently or noncovalently the fusion partner and associated variant library member with the nucleic acid that encodes them. For example, U.S. Ser. No. 09/642,574; U.S. Ser. No. 10/080,376; U.S. Ser. No. 09/792,630; U.S. Ser. No. 10/023,208; U.S. Ser. No. 09/792,626; U.S. Ser. No. 10/082,671; U.S. Ser. No. 09/953,351; U.S. Ser. No. 10/097,100; U.S. Ser. No. 60/366,658; PCT WO 00/22906; PCT WO 01/49058; PCT WO 02/04852; PCT WO 02/04853; PCT WO 02/08023; PCT WO 01/28702; and PCT WO 02/07466 describe such a fusion partner and technique that may find use in the present invention. In an alternative embodiment, in vivo selection can occur if expression of the protein imparts some growth, reproduction, or survival advantage to the cell.

[0156] A subset of selection methods referred to as "directed evolution" methods are those that include the mating or breading of favorable sequences during selection, sometimes with the incorporation of new mutations. As will be appreciated by those skilled in the art, directed evolution methods can facilitate identification of the most favorable sequences in a library, and can increase the diversity of sequences that are screened. A variety of directed evolution methods are known in the art that may find use in the present invention for screening protein variants, including but not limited to DNA shuffling (PCT WO 00/42561 A3; PCT WO 01/70947 A3), exon shuffling (U.S. Pat. No. 6,365,377; Kolkman & Stemmer, 2001, *Nat Biotechnol* 19:423-428), family shuffling (Crameri et al., 1998, *Nature* 391:288-291; U.S. Pat.

No. 6,376,246), RACHIT™ (Coco et al., 2001, *Nat Biotechnol* 19:354-359; PCT WO 02/06469), STEP and random priming of in vitro recombination (Zhao et al., 1998, *Nat Biotechnol* 16:258-261; Shao et al., 1998, *Nucleic Acids Res* 26:681-683), exonuclease mediated gene assembly (U.S. Pat. No. 6,352,842; U.S. Pat. No. 6,361,974), Gene Site Saturation Mutagenesis™ (U.S. Pat. No. 6,358,709), Gene Reassembly™ (U.S. Pat. No. 6,358,709), SCRATCHY (Lutz et al., 2001, *Proc Natl Acad Sci USA* 98:11248-11253), DNA fragmentation methods (Kikuchi et al., *Gene* 236:159-167), single-stranded DNA shuffling (Kikuchi et al., 2000, *Gene* 243:133-137), and AMEsystem™ directed evolution protein engineering technology (Applied Molecular Evolution) (U.S. Pat. No. 5,824,514; U.S. Pat. No. 5,817,483; U.S. Pat. No. 5,814,476; U.S. Pat. No. 5,763,192; U.S. Pat. No. 5,723,323).

[0157] In a preferred embodiment, the immunogenicity of the protein variants is determined experimentally to confirm that the variants do have reduced or eliminated immunogenicity relative to the parent protein. Several methods can be used for experimental confirmation of epitopes. In a preferred embodiment, ex vivo T-cell activation assays are used to experimentally quantitate immunogenicity. In this method, antigen presenting cells and naïve T cells from matched donors are challenged with a peptide or whole protein of interest one or more times. Then, T cell activation can be detected using a number of methods, for example by monitoring production of cytokines or measuring uptake of tritiated thymidine. In the most preferred embodiment, interferon gamma production is monitored using Elispot assays (Schmittel et. al., 2000, *J. Immunol. Meth.*, 24: 17-24). If sera are available from patients who have raised an immune response to protein, it is possible to detect mature T cells that respond to specific epitopes. In a preferred embodiment, interferon gamma or IL-5 production by activated T-cells is monitored using Elispot assays, although it is also possible to use other indicators of T cell activation or proliferation such as tritiated thymidine incorporation or production of other cytokines. Other suitable T cell assays include those disclosed in Meidenbauer et al., 2000, Prostate 43, 88-100; Schultes & Whiteside, 2003, *J. Immunol. Methods* 279, 1-15; and Stickler et al., 200, *J. Immunotherapy,* 23, 654-660. In a preferred embodiment, the PBMC donors used for the above-described T cell activation assays will comprise class II MHC alleles that are common in patients requiring treatment for protein responsive disorders. For example, for most diseases and disorders, it is desirable to test donors comprising all of the alleles that are prevalent in the population. However, for diseases or disorders that are linked with specific MHC alleles, it may be more appropriate to focus screening on alleles that confer susceptibility to protein responsive disorders. In a preferred embodiment, the MHC haplotype of PBMC donors or patients that raise an immune response to the wild type or protein variant are compared with the MHC haplotype of patients who do not raise a response. This data may be used to guide preclinical and clinical studies as well as aiding in identification of patients who will be especially likely to respond favorably or unfavorably to the protein therapeutic.

[0158] In an alternate preferred embodiment, immunogenicity is measured in transgenic mouse systems. For example, mice expressing fully or partially human class II MHC molecules may be used. In an alternate embodiment, immunogenicity is tested by administering the protein variants to one or more animals, including rodents and primates, and monitoring for antibody formation. Nonhuman primates with defined MHC haplotypes may be especially useful, as the sequences and hence peptide binding specificities of the MHC molecules in nonhuman primates may be very similar to the sequences and peptide binding specificities of humans. Similarly, genetically engineered mouse models expressing human MHC peptide-binding domains may be used (see for example Sonderstrup et al., 1999, *Immunol. Rev.* 172: 335-343; and Forsthuber et al., 2001, J. *Immunol.* 167: 119-125).

[0159] The biological properties of the proteins of the present invention may be characterized in cell, tissue, and whole organism experiments. As is known in the art, drugs are often tested in animals, including but not limited to mice, rats, rabbits, dogs, cats, pigs, and monkeys, in order to measure a drug's efficacy for treatment against a disease or disease model, or to measure a drug's pharmacokinetics, toxicity, and other properties. The animals may be referred to as disease models. Therapeutics are often tested in mice, including but not limited to nude mice, SCID mice, xenograft mice, and transgenic mice (including knockins and knockouts). Such experimentation may provide meaningful data for determination of the potential of the protein to be used as a therapeutic. Any organism, preferably mammals, may be used for testing. For example because of their genetic similarity to humans, monkeys can be suitable therapeutic models, and thus may be used to test the efficacy, toxicity, pharmacokinetics, or other property of the proteins of the present invention. Tests of the in humans are ultimately required for approval as drugs, and thus of course these experiments are contemplated. Thus the proteins of the present invention may be tested in humans to determine their therapeutic efficacy, toxicity, immunogenicity, pharmacokinetics, and/or other clinical properties.

[0160] In one embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes two or more amino acid substitutions derived from two or more natural antibodies. In this embodiment, a first resultant variant string in the variant antibody is rendered most homologous to a first natural antibody, a second resultant variant string in the variant antibody is rendered most homologous to the corresponding string in an second natural antibody, the substitutions are not in a CDR, and at least one resultant string is not a consensus of homologous natural sequences. In a preferred embodiment, the variant strings in the variant antibody do not include CDR residues. In a further preferred embodiment, the first and second natural antibodies are from different subfamilies.

[0161] In another embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes two or more amino acid substitutions derived from three or more natural antibodies. In this embodiment, a first resultant variant string in the variant antibody is rendered most homologous to a first natural antibody, a second resultant variant string in the variant antibody is rendered most homologous to the corresponding string in an second natural antibody, a third resultant variant string in the variant antibody is rendered most homologous to the corresponding string in an third natural antibody, the substitutions are not in a CDR, and at least one resultant string is not a consensus of homologous natural sequences. In an additional embodiment, the first, second and third natural antibodies are from different subfamilies. If a further additional embodiment, the variant antibody further comprises a fourth resultant variant string that is rendered most homologous to the corresponding string in a fourth natural antibody. If an additional embodiment, the variant antibody includes at least one substitution that is made

at a position that is not surface exposed, the first, second and third natural antibodies are from different antibody groups, one of the substitutions is made at a position that is part of the VH/VL interface, and at least one amino acid substitution is not a back mutation.

[0162] In other embodiments, the first, second and third natural antibodies are from different antibody groups.

[0163] In other embodiments, the variant antibody includes at least one substitution that is not surface exposed.

[0164] In other embodiments, at least one of the substitutions is made at a position that is part of the VH/VL interface.

[0165] In other embodiments, the variant antibody includes at least one amino acid substitution is not a back mutation.

[0166] In one embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes a variant VH antibody region with host string content (HSC) greater than about 75%, and a framework region homogeneity (FRH) less than about 60%, wherein the HSC and FRH are calculated with a window size of 9.

[0167] In one embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes a variant VH antibody region with exact string content greater than about 20%; and a framework region homogeneity less than about 60%, wherein the HSC and FRH are calculated with a window size of 9.

[0168] In one embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes a variant VL antibody region with exact string content greater than about 35%, and a framework region homogeneity less than about 60%, wherein HSC and FRH are calculated with a window size of 9.

[0169] In one embodiment of the present invention, a variant antibody for a host as compared to a parent antibody includes a first set of one or more amino acid substitutions from a first natural antibody and a second set of one or more amino acid substitutions from a second natural antibody, wherein the identity of said substituted amino acids from said second antibody differ from the corresponding amino acids of said first natural antibody, the substitutions are not in a CDR, and at least one substitution is not a consensus of homologous natural sequences. In an additional embodiment, the variant antibody further includes a third set of one or more amino acids substitutions from a third natural antibody wherein the identity of the substituted amino acids of said third set differ from the identity the corresponding amino acids from said first and second sets of amino acid substitutions. In a further embodiment, the variant antibody includes a fourth set of one or more amino acids from a fourth natural antibody wherein the identity of the substituted amino acids of said fourth set differ from the identity the corresponding amino acids from said first, second and third sets of amino acid substitutions. In other embodiments, the variant antibody includes multiple sets of one or more amino acids from multiple natural antibodies, wherein the identity of the substituted amino acids of any set differ from the identity the corresponding amino acids from the other sets of amino acid substitutions.

EXAMPLES

[0170] Examples are provided below to illustrate the present invention. These examples are not meant to constrain the present invention to any particular application or theory of operation.

[0171] For reference to immunoglobulin variable regions, positions are numbered according to the Kabat numbering scheme. For reference to immunoglobulin constant regions, positions are numbered according to the EU index as in Kabat (Kabat et al., 1991, *Sequences of Proteins of Immunological Interest,* 5th Ed., United States Public Health Service, National Institutes of Health, Bethesda).

Example 1

Immunogenicity Reduction of AC10

[0172] To illustrate application of the method described in the present invention, and to validate its broad applicability to immunogenicity reduction of proteins, a xenogeneic antibody example is provided using as the parent sequence the anti-CD30 antibody AC10 (Bowen et al. Journal of Immunology, 1993, 151: 5896). A structural model of the mouse AC10 variable region was constructed using standard antibody modeling methods known in the art. FIGS. **3** and **4** show the sequences (SEQ ID NOS: 118 and 1190, host string content, and structures of the AC10 VL and VH domains (referred to as L0 AC10 VL and H0 AC10 VH respectively). A CDR graft of this antibody was constructed by placing the AC10 CDRs into the context of the frameworks of the most homologous host germlines, determined to be vlk_4-1 for VL and vh_1-3 for VH using the sequence alignment program BLAST. The sequences (SEQ ID NOS: 120 and 121) and string content of these CDR grafts are shown in FIGS. **5** and **6**, along with structures of modeled AC10 highlighting the mutational differences between the CDR grafted AC10 variable chains and WT.

[0173] AC10 variants with reduced immunogenicity were generated by applying a string optimization algorithm on the WT AC10 VL and VH sequences. This algorithm heuristically samples multiple amino acid mutations that exist in the diversity of the human VLκ and VH germline sequences, and calculates the host string content (HSC) of each sequence according to Equation 3 described above, using a window size w=9. In this set of calculations, residues in the CDRs and close to a CDR or to the VL/VH interface were masked, that is were not allowed to mutate. CDRs were defined as a slightly smaller set of residues than the CDRs defined by Chothia (Chothia & Lesk, 1987, *J. Mol. Biol.* 196: 901-917; Chothia et al., 1989, *Nature* 342: 877-883; Al-Lazikani et al., 1997, *J. Mol. Biol.* 273: 927-948). For the purposes of the present invention, VL CDRs are herein defined to include residues at positions 27-32 (CDR1), 50-56 (CDR2), and 91-97 (CDR3), wherein the numbering is according to Chothia. Because the VL CDRs as defined by Chothia and Kabat are identical, the numbering of these VL CDR positions is also according to Kabat. For the purposes of the present invention, VH CDRs are herein defined to include residues at positions 27-33 (CDR1), 52-56 (CDR2), and 95-102 (CDR3), wherein the numbering is according to Chothia. These VH CDR positions correspond to Kabat positions 27-35 (CDR1), 52-56 (CDR2), and 95-102 (CDR3). Masked residues in these calculations were set at positions 1-4, 25-34, 36, 38, 43, 44, 46, 48-58, 60, 63-69, 71, 87, and 89-98 for VL, and 2, 4, 24, 26-35, 37, 39, 44, 45, 47, 50-58, 60, 61, 71, 73, 76, 78, 91, and 93-106 for VH, wherein the numbering is according to Kabat. Masking of potentially critical residues is a conservative approach to generating more host antibody variants, however it is but one embodiment of the present invention, and calculations wherein positions are not masked are also contemplated. This calculation was run for AC10 VL and VH in 100 separate interactions, generating a

set of diverse AC10 variants with more host string content than WT. FIG. **7** shows the clustered nonredundant set of output sequences from these calculations for the AC10 VL and VH region (SEQ ID NOS:122-217), referred to as AC10 VL HSC Calculation 1 and AC10 VH HSC Calculation 1 respecitvely. For each iteration (Iter), the HSC (Equation 3), HSS (Equation 5), and number (Mut) and identity (shaded residues) of mutations from WT are presented. In addition to the HSC score, each sequence was evaluated for its structural and functional integrity using a nearest neighbor structure-based scoring method (U.S. Ser. No. 60/528,229, filed Dec. 8, 2003, entitled Protein Engineering with Analogous Contact Environments). Two measures of structural fitness, referred to as "Structural Consensus" and "Structural Precedence", are also provided in the FIG. **7**. Although the Analogous Contact Environments method is particularly well-suited for antibodies because of the wealth of sequence and structure information, any structure-based and/or sequence-based scoring method may be used to evaluate the structural and functional fitness of the variant sequences. The output sequences were clustered based on their mutational distance from the other sequences in the set, and these clusters are delineated by the horizontal black lines in the Figure. The "Cluster" column provides the quantitative mutational distance between each sequence and the rest of the sequence in its cluster; sequences with a lower cluster value are more representative of that particular sequence cluster.

[0174] These calculations were used to generate a set of AC10 VL and VH variants. In some cases, further substitutions were made to output sequences, using string and structural scores, as well as visual inspection of the modeled AC10 structure, to evaluate fitness. FIGS. **8-13** present the sequences (SEQ ID NOS: 218-223), host string content, and mapped mutational differences on the modeled AC10 structure for each of the AC10 VL and VH variants. Iteration 36 from AC10 VL HSC calculation 1 served as the precursor for L1 AC10 VL, iteration 37 served as the precursor for L2 AC10 VL, and iteration 3 served as the precursor for L3 AC10 VL. Iteration 15 from AC10 VH HSC calculation 1 served as the precursor for H1 AC10 HL, iteration 55 served as the precursor for H2 AC10 VH, and iteration 18 served as the precursor for H3 AC10 VH.

[0175] Tables 1 and 2 present the number of mutations from the parent sequence, structural fitness scores, and host string scores for the AC10 VL and VH variants as compared to the WT and CDR grafted AC10 sequences. In addition to the aforementioned structural and host string analysis, each sequence was analyzed for its global homology to the host germline. The maximum identity match to the germline for each string in the sequences was also determined, referred to as $N_o$max. This represents the total number of strings in each sequence whose maximum identity to the corresponding strings in the host germline is the indicated value. For w=9, Tables 1 and 2 list N_9max, N_8max, N_7max, and $N_{\leqq 6}$max for each sequence. N_9max represents the number of strings in the sequence for which 9 of 9 residues match at least one string in the host germline, N_8max represents the number of strings for which 8 of 9 residues match at least one string in the host germline, N_7max represents the number of strings for which 7 of 9 residues match at least one string in the host germline, and $N_{\leqq 6}$max represents the number of strings for which 6 or less residues of 9 residues match at least one string in the host germline. This last category (ID≦6) could, for example, be regarded as the number of poorly scoring strings. In addition to the aforementioned structural and host string analysis, each sequence was analyzed for its global homology to the host germline; Tables 1 and 2 present the most homologous human germline sequence for each sequence (Closest Germline) and corresponding identity to that germline (ID to Closest Germline), determined using the sequence alignment program BLAST. Finally, the Framework region homogeneity (FRH) of each variant was evaluated for w=9, and is presented in Tables 1 and 2, providing the extent to which the host string content of each variant is derived from a single germline as opposed to multiple germline sequences.

TABLE 1

AC10 VL Variants

	WT	CDR Graft	L1	L2	L3
Mutations		18	15	16	9
Structural Consensus	0.57	0.57	0.59	0.64	0.56
Structural Precedence	0.68	0.57	0.66	0.67	0.58
Human String Content	0.78	0.88	0.86	0.86	0.85
Human String Similarity	0.15	0.57	0.48	0.46	0.41
Framework Region Homogeneity	0.60	0.97	0.73	0.81	0.52
N_9max	15	61	51	48	42
N_8max	27	11	13	15	21
N_7max	31	15	21	22	20
$N_{\leqq 6}$max	34	20	22	22	24
Closest Germline	4-1	4-1	3-11	1-39	4-1
ID to Closest Germline	68/101 67%	86/101 85%	78/99 79%	80/99 81%	75/101 74%

TABLE 2

AC10 VH Variants

	WT	CDR Graft	H1	H2	H3
Mutations		26	16	23	20
Structural Consensus	0.49	0.48	0.48	0.50	0.47
Structural Precedence	0.63	0.67	0.63	0.59	0.59
Human String Content	0.69	0.87	0.81	0.81	0.80
Human String Similarity	0.07	0.68	0.41	0.39	0.38
Framework Region Homogeneity	0.60	0.86	0.65	0.47	0.55
N_9max	5	81	48	45	44
N_8max	31	13	30	32	28
N_7max	34	8	20	21	24
$N_{\leqq 6}$max	49	17	21	21	23
Closest Germline	1-3	1-3	1-3	1-3	7-4-1
ID to Closest Germline	69/98 70%	93/98 95%	83/98 85%	72/98 73%	76/98 78%

[0176] An important observation is that, whereas the CDR grafted antibodies are most homologous to a single human germline sequence (the "acceptor" sequence in humanization terminology), the present invention describes variants that are homologous to different host germline sequences in different regions of the sequence. This is evident from the significant differences in Framework region homogeneity (FRH) scores for the AC10 variants of the present invention and CDR grafted AC10 variants. Furthermore, whereas CDR grafted AC10 VL and VH are most homologous to human germline subfamilies 4 (VL) and 1 (VH) respectively across their entire sequences, a number of the AC10 variants are most homologous to different subfamilies in different frameworks. Addi-

tionally, whereas the CDR grafted antibodies are most homologous to a single germline sequence that is also the most homologous sequence to the parent sequence, the present invention presents a set of antibodies for a given antibody that are most homologous to different human germline sequences, which need not be the most homologous germline sequence to WT. For example, Table 1 shows that CDR grafted AC10 VL is most homologous to 4-1, which is also the most homologous human germline to the WT AC10 parent. However L1, L2, and L3 are most homologous to three different human germlines—3-11, 1-39, and 4-1 respectively. Thus the variants of the present invention explore a substantially greater amount of diversity than CDR grafted antibodies. One obvious advantage of this is that the method of the present invention provides a greater chance of success with respect to antigen affinity. The choice of an "acceptor" in humanization methods places a single bet; if the donor CDRs are in fact incompatible with the acceptor FRs, a set of back-mutations that regain WT affinity may not exist. In contrast, the method of the present invention enables a greater diversity of sequence and structure space to be sampled in the immunogenicity reduction process, increasing the chances of obtaining a final less immunogenic version with WT affinity or better. An additional advantage of sampling greater sequence diversity is that some sequences may have more optimal properties than others, for example with regard to stability, solubility, and effector function. For example, as disclosed in U.S. Ser. No. 60/614,944, and U.S. Ser. No. 60/619,409, filed Oct. 14, 2004, entitled "Immunoglobulin Variants Outside the Fc Region with Optimized Effector Function", the variable region of an antibody may impact effector functions such as antibody dependent cell-mediated cytotoxicity (ADCC), antibody dependent cell-mediated phagocytosis (ADCP), and complement dependent cytotoxicity (CDC).

[0177] The genes for the variable regions of AC10 WT (L0 and H0) and variants (L1, L2, L3, H1, H2, and H3) were constructed using recursive PCR, and subcloned into a the mammalian expression vector pcDNA3.1Zeo (Invitrogen) comprising the full length light kappa (CDκ) and heavy chain IgG1 constant regions. All sequences were sequenced to confirm the fidelity of the sequence. Plasmids containing heavy chain gene (VH-CH1-CH2-CH3) (wild-type or variants) were co-transfected with plasmid containing light chain gene (VL-CLκ) in all combinations (L0/H0, L0/H1, L0/H2, L0/H3, L1/H0, L1/H1, L1/H2, L1/H3, L2/H0, L2/H1, L2/H2, L2/H3, L3/H0, L3/H1, L3/H2, L3/H3) into 293T cells. Here, for example, L2/H3 refers to the L2 AC10 VL paired with H3 AC10 VH. Media were harvested 5 days after transfection, and antibodies were purified from the supernatant using protein A affinity chromatography (Pierce, Catalog #20334).

[0178] WT and variant antibodies were experimentally tested for their capacity to bind CD30 antigen. Binding affinity to human CD30 by the AC10 WT and variant antibodies was measured using a quantitative and extremely sensitive method, AlphaScreen™ assay. The AlphaScreen™ assay is a bead-based non-radioactive luminescent proximity assay. Laser excitation of a donor bead excites oxygen, which if sufficiently close to the acceptor bead will generate a cascade of chemiluminescent events, ultimately leading to fluorescence emission at 520-620 nm. The AlphaScreen™ assay was applied as a competition assay for screening the antibodies. WT AC10 antibody was biotinylated by standard methods for attachment to streptavidin donor beads (Perkin Elmer). Commericial CD30 was conjugated to digoxigenin (DIG) (Roche Diagnostics) for attachment to anti-DIG acceptor beads (Perkin Elmer). In the absence of competing AC10 variants, WT antibody and CD30 interact and produce a signal at 520-620 nm. Addition of untagged AC10 variant competes with the WT AC10/CD30 interaction, reducing fluorescence quantitatively to enable determination of relative binding affinities. FIGS. **14***a* and **14***b* show binding of WT (H0L0) and AC10 variant antibodies to CD30 using the AlphaScreen™ assay. The data were fit to a one site competition model using non-linear regression, and these fits are represented by the curves in the figure. These fits provide the inhibitory concentration 50% (IC50) (i.e. the concentration required for 50% inhibition) for each antibody, thus enabling the relative binding affinities relative to WT to be determined. Table 3 provides the IC50's and Fold IC50's relative to WT for fits to these binding curves. The AC10 variants display an array of CD30 binding affinities, with a number of variants binding CD30 with affinity comparable to or better affinity than WT AC10.

[0179] Antigen affinity of the AC10 variants was also measured using Surface Plasmon Resonance (SPR) (Biacore, Uppsala, Sweden). SPR allows for the measurement of direct binding rates and affinities of protein-protein interactions, and thus provides an excellent complementary binding assay to the AlphaScreen™ assay. CD30 fused to the Fc region of IgG1 (R&D Systems) was immobilized on a Protein A SPR chip, the surface was blocked with Fc, and WT and variant AC10 antibodies were flowed over the chip at a range of concentrations. The resulting sensorgrams are shown in FIG. **15**. Global Langmuir fits were carried out for the concentrations series using the BiaEvaluation curve fitting software, providing the on-rate constant (ka), off-rate constant (kd), and equilibrium binding constant (KD=kd/ka) for the curves. Table 3 provides the KDs and Fold KDs relative to WT for the SPR data. The excellent agreement between the rank ordering of the variants as determined by SPR and AlphaScreen™ assay support the accuracy of the data.

TABLE 3

CD30 Binding of AC10 Variants

AC10 Variant	SPR KD (nM)	SPR Fold KD	AlphaScreen IC50 (nM)	AlphaScreen Fold IC50
H2L1	9.49	0.36	55.1	0.06
H2L2	5.95	0.57	49.2	0.06
H1L2	7.55	0.45	45.3	0.07
H1L1	5.63	0.60	27.7	0.11
H2L3	6.75	0.50	27.2	0.12
H2L0	8.00	0.42	19.4	0.16
H1L3	5.09	0.67	17.4	0.18
H1L0	6.39	0.53	9.77	0.32
H0L2	3.48	0.97	7.81	0.41
H3L2	2.86	1.19	6.57	0.48
H3L0	3.08	1.10	6.18	0.51
H3L1	2.44	1.39	6.09	0.52
H0L1	3.29	1.03	5.19	0.61
H0L3	3.00	1.13	4.61	0.69
H0L0	3.39	1.00	3.18	1.00
H3L3	2.33	1.45	1.99	1.59

[0180] In addition to assessing the antigen affinity and biophysical properties of the variants of the present invention, they may also be tested for effector functions in the context of a full length antibody. One advantage of generating multiple reduced immunogenicity variants of a parent immunoglobulin is that it enables a greater degree of sequence diversity to

be sampled, diversity which may provide optimal properties. Some sequences may have more optimal properties than others, for example with regard to effector function. For example, as disclosed in U.S. Ser. No. 60/614,944, and U.S. Ser. No. 60/619,409, filed Oct. 14, 2004, entitled "Immunoglobulin Variants Outside the Fc Region with Optimized Effector Function", the variable region of an antibody may impact effector functions such as antibody dependent cell-mediated cytotoxicity (ADCC), antibody dependent cell-mediated phagocytosis (ADCP), and complement dependent cytotoxicity (CDC).

[0181] In order to explore any differences in capacity to mediate effector function, the affinities of the AC10 variants for FcγRIIIa were measured using the AlphaScreen™ assay. The extracellular region of human V158 FcγRIIIa was obtained by PCR from a clone obtained from the Mammalian Gene Collection (MGC:22630), and the receptor was fused with glutathione S-Transferase (GST) to enable screening. Tagged FcγRIIIa was transfected in 293T cells, and media containing secreted FcγRIIIa were harvested and purified. The AlphaScreen™ assay was applied as a competition assay for screening AC10 variants for binding to FcγRIIIa. Biotinylated WT AC10 antibody was bound to streptavidin donor beads (Perkin Elmer), and GST-fused human V158 FcγRIIIa was bound to anti-GST acceptor beads (Perkin Elmer). The binding data are shown in FIGS. **16***a* and **16***b*, and the resulting IC50's and Fold IC50's relative to WT are provided in Table 4. FcγRIIIa affinity of the AC10 variants was also measured using SPR. GST-fused human FcγRIIIa (V158 isoform) was immobilized on a chip, and WT and variant AC10 antibodies were flowed over the chip at a range of concentrations. Binding constants were obtained from fitting the data using standard curve-fitting methods. The equilibrium dissociation constants (KDs) obtained from the fits to these binding curves, and the calculated fold improvement or reduction relative to WT (Fold KD) are shown in Table 4.

TABLE 4

FcγRIIIa Binding of AC10 Variants

AC10 Variant	SPR KD (nM)	SPR Fold KD	AlphaScreen IC50 (nM)	AlphaScreen Fold IC50
H2L1	14.9	1.25	751	0.12
H2L2	4.01	4.64	146	0.60
H1L2	1.6.6	1.12	340	0.26
H1L1	11.2	1.66	221	0.39
H2L3	3.52	5.28	183	0.48
H2L0	12.9	1.44	175	0.50
H1L3	11.2	1.66	178	0.49
H1L0	22.0	0.85	71.6	1.22
H0L2	9.09	2.05	93.8	0.93
H3L2	3.57	5.21	88.7	0.98
H3L0	20.0	0.93	216	0.40
H3L1	17.4	1.07	209	0.42
H0L1	11.6	1.60	183	0.48
H0L3	12.7	1.46	146	0.60
H0L0	18.6	1.00	87.2	1.00
H3L3	6.13	3.03	83.5	1.04

[0182] To assess the capacity of the AC10 variants to mediate effector function against CD30 expressing cells, the AC10 variants were tested in a cell-based ADCC assay. Human peripheral blood monocytes (PBMCs) were isolated from buffy-coat and used as effector cells, and CD30 positive L540 Hodgkin's lymphoma cells were used as target cells. L540 target cells were seeded at 20,000 per well in 96-well plates and treated with designated antibodies in triplicates starting at 1 μg/ml and in reduced concentrations in ½ log steps. PBMCs isolated using a Ficoll gradient and allotyped as FcγRIIIa 158 V/F were added at 25-fold excess of L540 cells and co-cultured for 4 hrs before processing for LDH activity using the Cytotoxicity Detection Kit (LDH, Roche Diagnostic Corporation, Indianapolis, Ind.) according to the manufacturer's instructions. The plates were read using a Wallac 1420 Victor^{2}™. FIGS. **17***a*-**17***c* show the results. The graphs show that the antibodies differ not only in their EC50, reflecting their relative potency, but also in the maximal level of ADCC attainable by the antibodies at saturating concentrations, reflecting their relative efficacy. These two terms, potency and efficacy, are sometimes used loosely to refer to desired clinical properties. In the current experimental context, however, they are denoted as specific quantities, and therefore are here explicitly defined. By "potency" as used in the current experimental context is meant the EC50 of an EGFR targeting protein. By "efficacy" as used in the current experimental context is meant the maximal possible effector function of an antibody at saturating levels. Differences in capacity to mediate ADCC may be due to differences in antigen affinity, different capacities of the variant variable regions to effect FcγR binding, or both. Regardless, the contribution of an antibody variable region to FcγR binding and effector function may be an important parameter for selecting a clinical candidate. The choice of an antibody clinical candidate based in whole or in part on the impact on effector function of the variable region represents a novel dimension in antibody therapeutics.

[0183] Based on the CD30 binding, FcγRIIIa binding, and ADCC results, the H3/L3 AC10 variant was chosen as a potential biotherapeutic candidate. Because this antibody is intended for clinical use as an anti-cancer therapeutic, it may be advantageous to optimize its effector function. As previously described, substitutions can be engineered in the constant region of an antibody to provide favorable clinical properties. In a most preferred embodiment, one or more amino acid modifications that provide optimized binding to FcγRs and/or enhanced effector function described in U.S. Ser. No. 10/672,280, PCT US03/30249, and U.S. Ser. No. 10/822,231, and U.S. Ser. No. 60/627,774, filed Nov. 12, 2004 and entitled "Optimized Fc Variants", are combined with the AC10 variants of the present invention. A number of optimized Fc variants obtained from these studies, including I332E, S239D, V264I/I332E, S239D/I332E, and S239D/A330L/I332E, were constructed in the H0/L0 and H3/L0 AC10 antibodies using quick change mutagenesis (Stratagene). Antibodies were expressed and purified as described above. FIGS. **18***a* and **18***b* show the results of the ADCC assay, carried out as described above, comparing WT (H0/L0) and H3/L3 AC10 in combination with the optimized Fc variants. Considerable enhancements in potency and efficacy are observed for the Fc variant antibodies as compared to H0/L0 and H3/L3 AC10.

[0184] As described above, because variant sequences of the invention are preferably derived from a HSC-increasing procedure in which substitution of structurally important positions is disallowed (or discouraged), it is likely that additional optimization of HSC is possible if those positions are allowed to vary in a secondary analysis. It is noted that, due to residue masking, mutations in the variants occur distal to the CDRs and VL/VH interface. This is in contrast to CDR grafted antibodies, which have mutations in the parent that are at or near these critical regions and thus have a significantly

greater potential for perturbing antigen affinity. This is corroborated by the fact that CDR grafted antibodies typically require backmutations to the donor sequence to regain WT affinity for antigen. Such backmutations are usually made out of structural and immunogenic context with respect to host sequences, and cause dramatic reductions in the host string content of the final variant. In contrast, the variants presented herein are simultaneously optimized for host string and structural fitness within the same context, and no backmutations need be made. Nonetheless, one or more subsequent substitutions may be explored to increase antigen affinity or further improve HSC, for example by mutating residues that were masked in the calculations and/or residues in or close to the CDRs or VL/VH interface. Thus the H3/L3 variant can be thought of as a primary variant or template for further optimization, and variants of H3/L3 can be thought of as secondary variants. In contrast to backmutating as with CDR grafted antibodies, secondary substitutions in the variants of the present invention will comprise forward or neutral mutations with respect to the host germline, and thus are expected to only improve or unaffected HSC. An additional benefit of generating secondary variants is that, by exploring quality structural and string diversity, it is also possible that other properties can be optimized, for example affinity, activity, specificity, solubility, expression level, and effector function.

[0185] String analysis was carried out on the H3/L3 sequence to design a set of secondary substitutions that have neutral, positive, or minimal impact on HSC, and/or that have significant potential for optimization of antigen affinity and/or effector function. Table 5 provides this set of 70 VL (Table 5a) and 64 VH (Table 5b) single mutations. The H3 column provides the WT H3 amino acid, and the Sub column provides the designed substitution. Positions are numbered according to the Kabat numbering format, with Kabat CDR positions bolded. The provided string impact, defined according to Equation 7, describes the difference in HSC between the primary variant sequence, here H3/L3, and the secondary variant sequence.

TABLE 5a

L3 AC10 Secondary Variants

Variant	Pos (Kabat)	L3	Sub	String Impact	Fold Prot A	Fold CD30	Fold FcγRIIIa
L3.1	1	D	A	0	0.89	1.30	0.96
L3.2	1	D	E	1	1.09	1.24	1.46
L3.3	1	D	N	0	1.24	1.66	1.35
L3.4	1	D	S	0	0.97	1.04	1.18
L3.5	3	V	Q	0	1.13	1.32	1.32
L3.6	4	L	M	4	1.65	1.64	1.21
L3.7	25	A	S	6			
L3.8	27a	S	D	0	1.02	0.94	0.61
L3.9	27b	V	I	1	1.04	0.58	0.81
L3.10	27c	D	L	5			
L3.11	27c	D	S	8			
L3.12	27c	D	V	3	1.24	1.15	1.19
L3.13	27d	F	D	0	1.07	0.32	0.98
L3.14	27d	F	H	4	0.86	0.08	0.93
L3.15	27d	F	Y	5	1.04	0.63	1.19
L3.16	28	D	N	−1	1.10	1.61	1.18
L3.17	30	D	K	4	1.01	1.21	1.24
L3.18	30	D	N	6	1.09	1.45	0.94
L3.19	30	D	S	4	1.07	1.13	0.82
L3.20	30	D	Y	0	0.82	0.78	0.73
L3.21	31	S	D	0	1.01	0.81	0.95
L3.22	31	S	T	3	1.03	0.46	0.97
L3.23	31	S	N	−1	1.03	0.71	1.00

TABLE 5a-continued

L3 AC10 Secondary Variants

Variant	Pos (Kabat)	L3	Sub	String Impact	Fold Prot A	Fold CD30	Fold FcγRIIIa
L3.24	32	Y	D	0	1.31	0.46	1.33
L3.25	33	M	L	8	1.38	1.36	1.37
L3.26	34	N	S	0			
L3.27	34	N	A	−1	1.39	0.38	1.36
L3.28	34	N	D	−6	1.19	0.41	1.76
L3.29	46	V	H	4	0.06	0.03	0.11
L3.30	46	V	L	9	0.86	0.39	0.75
L3.31	46	V	R	4			
L3.32	46	V	S	4	1.05	0.32	0.90
L3.33	50	A	D	6	0.98	0.26	0.66
L3.34	50	A	S	5	1.01	0.47	1.20
L3.35	50	A	W	2			
L3.36	53	N	S	8			
L3.37	53	N	T	5	0.99	1.26	1.01
L3.38	54	L	R	1	1.19	1.46	1.55
L3.39	55	E	A	2	1.01	0.85	1.32
L3.40	55	E	Q	6	0.99	0.87	1.07
L3.41	56	S	T	8	1.50	1.80	1.23
L3.42	58	I	V	4	1.44	1.55	0.95
L3.43	60	A	D	1	1.11	1.16	1.08
L3.44	60	A	S	2	0.82	1.08	0.85
L3.45	67	S	P	0			
L3.46	89	Q	H	1	1.37	0.08	1.64
L3.47	91	S	A	8			
L3.48	91	S	G	9	0.85	0.29	0.80
L3.49	91	S	H	2	1.20	0.01	1.32
L3.50	91	S	L	8	1.10	0.02	1.59
L3.51	91	S	Y	8	1.00	0.02	1.50
L3.52	92	N	I	3			
L3.53	92	N	S	2	3.02	0.48	1.34
L3.54	92	N	Y	8	1.39	0.96	1.05
L3.55	93	E	K	8	0.62	0.27	0.49
L3.56	93	E	N	8	1.06	0.64	0.84
L3.57	93	E	Q	2			
L3.58	93	E	S	8	0.90	0.49	0.87
L3.59	94	D	A	3	1.16	0.09	1.14
L3.60	94	D	F	9	1.22	0.02	1.19
L3.61	94	D	H	8			
L3.62	94	D	L	3	0.87	0.46	0.79
L3.63	94	D	S	1	1.74	0.57	1.42
L3.64	94	D	T	7	1.24	0.14	1.16
L3.65	96	W	F		0.33	0.34	0.29
L3.66	96	W	I		0.75	0.00	0.57
L3.67	96	W	L				
L3.68	96	W	Y				
L3.69	100	G	P				
L3.70	100	G	Q				

TABLE 5b

H3 AC10 Secondary Variants

Variant	Position (Kabat)	H3	Sub	String Impact	Fold Prot A	Fold CD30
H3.1	1	Q	E	−1	0.83	1.00
H3.2	2	I	L	0	1.60	2.76
H3.3	2	I	M	2	0.88	0.68
H3.4	2	I	V	0	0.98	1.28
H3.5	9	P	A	2	0.95	1.29
H3.6	16	A	T	2	0.89	1.13
H3.7	24	A	V	2	1.54	4.45
H3.8	31	D	G	2	0.80	1.40
H3.9	31	D	S	2	0.82	1.65
H3.10	33	Y	D	2	0.68	0.07
H3.11	33	Y	G	3	0.96	0.73
H3.12	33	Y	W	0	0.84	0.00
H3.13	34	I	L	1	0.96	1.52

TABLE 5b-continued

H3 AC10 Secondary Variants

Variant	Position (Kabat)	H3	Sub	String Impact	Fold Prot A	Fold CD30
H3.14	34	I	M	8	1.05	1.62
H3.15	35	T	D	1	1.55	0.05
H3.16	35	T	G	2	1.03	0.15
H3.17	35	T	H	8	0.86	0.04
H3.18	35	T	N	4	1.07	0.13
H3.19	35	T	S	6	0.88	1.11
H3.20	44	G	A	0	1.20	2.04
H3.21	44	G	R	0	1.36	2.60
H3.22	50	W	I	6	1.25	0.01
H3.23	50	W	R	0	0.99	0.16
H3.24	52	Y	N	2	1.03	0.03
H3.25	52	Y	T	1	1.11	0.06
H3.26	52	Y	V	1	1.33	0.06
H3.27	52a	P	A	1	1.02	2.00
H3.28	52a	P	V	1	1.44	1.34
H3.29	54	S	D	1	1.45	1.81
H3.30	54	S	N	5	1.13	1.45
H3.31	58	K	G	4		
H3.32	58	K	I	2	1.22	1.09
H3.33	58	K	N	5	1.26	0.50
H3.34	60	N	A	7	0.87	1.30
H3.35	60	N	P	0		
H3.36	60	N	S	7	1.02	1.24
H3.37	60	N	T	0	1.12	1.01
H3.38	60	N	V	0	1.16	1.14
H3.39	60	N	D	0	1.09	1.00
H3.40	61	E	Q	7	1.51	1.83
H3.41	64	Q	T	0	0.98	1.38
H3.42	71	V	L	4	1.10	0.66
H3.43	71	V	M	9	1.17	0.88
H3.44	71	V	R	1	1.25	1.76
H3.45	87	T	M	−1	0.99	1.14
H3.46	89	V	M	−4	1.41	1.39
H3.47	91	F	H	2		
H3.48	91	F	Y	9	1.32	1.60
H3.49	93	A	T	0	1.47	0.38
H3.50	93	A	V	0	1.01	1.40
H3.51	94	N	A	9	1.51	0.08
H3.52	94	N	H	5	1.23	1.24
H3.53	94	N	K	9	1.67	0.02
H3.54	94	N	R	9	1.26	0.00
H3.55	94	N	T	9	1.24	0.91
H3.56	99	W	Y		1.26	0.07
H3.57	101	A	D		1.31	0.53
H3.58	101	A	Q		1.17	0.16
H3.59	102	Y	H		1.69	1.05
H3.60	102	Y	S		1.04	0.82
H3.61	102	Y	V		1.33	1.21
H3.62	102	Y	L		1.34	1.22
H3.63	102	Y	F		1.18	1.24
H3.64	105	Q	R		1.15	1.28

[0186] The secondary H3/L3 variants were constructed using quick change mutagenesis, and the full length antibodies were expressed and purified as described above. H3 variants comprised H3 variant VH chains (H3.1-H3.64) in combination with L3 VL, and L3 variants comprised L3 variant VL chains (L3.1-L3.70) in combination with H3 VH. The AlphaScreen™ assay was used to measure binding of the H3/L3 secondary variants to CD30 and FcγRIIIa (as described earlier), as well as to protein A using biotinylated AC10 bound directly to protein A acceptor beads and streptavidin donor beads. FIG. **19** provides AlphaScreen™ binding curves for binding of select AC10 variants to CD30. The Fold IC50's relative to WT H3/L3 for binding to CD30, FcγRIIIa, and protein A are provided in Table 5. A number of H3/L3 secondary variants provide comparable or improved binding to CD30 antigen relative to the H3/L3 parent, enabling the engineering of additional variants that comprise combinations of these substitutions, which may provide further enhancements in HSC and/or antigen affinity.

[0187] Secondary substitutions that show favorable properties with respect to antigen affinity, effector function, stability, solubility, expression, and the like, may be combined in subsequent variants to generate a more optimized therapeutic candidate. Two new VL and three new VH variants were designed that comprise combinations of the described secondary substitutions, referred to as L3.71, L3.72, H3.68, H3.69, and H3.70. FIGS. **20-24** present the sequences_(SEQ ID NOS: 224-228), host string content, and mapped mutational differences on the modeled AC10 structure for each of these new AC10 VL and VH variants. Table 6 presents the number of mutations from the parent sequence, structural fitness scores, host string scores, and homology scores for these AC10 VL and VH variants.

TABLE 6

AC10 Variants

	L3.71	L3.72	H3.68	H3.69	H3.70
Mutations	15	15	23	27	30
Structural Consensus	0.56	0.55	0.46	0.46	0.45
Structural Precedence	0.54	0.52	0.55	0.57	0.56
Human String Content	0.88	0.87	0.80	0.83	0.84
Human String Similarity	0.52	0.45	0.39	0.47	0.47
Framework Region Homogeneity	0.47	0.51	0.33	0.40	0.42
N_9max	55	47	46	55	55
N_8max	19	24	26	26	34
N_7max	14	17	24	20	12
$N_{\leq 6}$max	19	19	23	18	18
Closest Germline	4-1	4-1	7-4-1	1-3	1-3
ID to Closest Germline	76/101 75%	76/101 75%	74/98 76%	77/98 79%	79/98 81%

[0188] Because the provided AC10 variants antibodies are clinical candidates for anti-cancer therapeutics, it may be advantageous to optimize their effector function. As previously described, substitutions can be engineered in the constant region of an antibody to provide favorable clinical properties. Combinations of the variants of the present invention with Fc modifications that alter effector function are anticipated. In a most preferred embodiment, one or more amino acid modifications that provide optimized binding to FcγRs and/or enhanced effector function described in U.S. Ser. No. 10/672,280, PCT US03/30249, and U.S. Ser. No. 10/822,231, and U.S. Ser. No. 60/627,774, filed Nov. 12, 2004 and entitled "Optimized Fc Variants", are combined with the AC10 variants of the present invention. The optimal anti-CD30 clinical candidate may comprise amino acid modifications that reduce immunogenicity and enhance effector function relative to a parent anti-CD30 antibody. FIGS. **25***a*-**25***c* (SEQ ID NOS:229-231) provide the light and heavy chain sequences of AC10 variants that comprise L3.71/H3.70 AC10 as described above, combined with a number of possible variant IgG1 constant regions, comprising one or more modifications at S239, V264, A330, and I332, that provide enhanced effector function.

[0189] Although human IgG1 is the most commonly used constant region for therapeutic antibodies, other embodiments may utilize constant regions or variants thereof of other IgG immunoglobulin chains. Effector functions such as

ADCC, ADCP, CDC, and serum half-life differ significantly between the different classes of antibodies, including for example human IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, IgD, IgE, IgG, and IgM (Michaelsen et al., 1992, Molecular Immunology, 29(3): 319-326). A number of studies have explored IgG1, IgG2, IgG3, and IgG4 variants in order to investigate the determinants of the effector function differences between them. See for example Canfield & Morrison, 1991, J. Exp. Med. 173: 1483-1491; Chappel et al., 1991, Proc. Natl. Acad. Sci. USA 88(20): 9036-9040; Chappel et al., 1993, Journal of Biological Chemistry 268:25124-25131; Tao et al., 1991, J. Exp. Med. 173: 1025-1028; Tao et al., 1993, J. Exp. Med. 178: 661-667; Redpath et al., 1998, Human Immunology, 59, 720-727. Using methods known in the art, it is possible to determine corresponding or equivalent residues in proteins that have significant sequence or structural homology with each other. By the same token, it is possible to use such methods to engineer amino acid modifications in an antibody or Fc fusion that comprise constant regions from other immunoglobulin classes, for example as described in U.S. Ser. No. 60/621,387 and 60/629,068, to provide optimal properties. As an example, the relatively poor effector function of IgG2 may be improved by replacing key FcγR binding residues with the corresponding amino acids in an IgG with better effector function, for example IgG1. For example, key residue differences between IgG2 and IgG1 with respect to FcγR binding may include P233, V234, A235, -236 (referring to a deletion in IgG2 relative to IgG1), and G327. Thus one or more amino acid modifications in the parent IgG2 wherein one or more of these residues is replaced with the corresponding IgG1 amino acids, P233E, V234L, A235L, -236G (referring to an insertion of a glycine at position 236), and G327A, may provide enhanced effector function. Furthermore, one or more additional amino acid modifications, for example the S239D, V264I, A330L, I332E, or combinations thereof as described above, may provide enhanced FcγR binding and effector function relative to the parent IgG2. FIGS. **25***a* (SEQ ID NO:229), **25***d* (SEQ ID NO:232), and **25***e* (SEQ ID NO:233) illustrate this embodiment, providing the light and heavy chain sequences of AC10 variants that comprise L3.71/H3.70 AC10 combined with a number of possible variant IgG2 constant regions.

[0190] The Fc modifications defined in FIG. **25** that provide enhanced effector function are not meant to constrain the invention to only these modifications for effector function optimization. For example, as described in U.S. Pat. No. 6,737,056, PCT US2004/000643, U.S. Ser. No. 10/370,749, and PCT/US2004/005112, the substitutions S298A, S298D, K326E, K326D, E333A, K334A, and P396L provide optimized FcγR binding and/or enhanced ADCC. Furthermore, as disclosed in Idusogie et al., 2001, J. Immunology 166: 2571-2572, substitutions K326W, K326Y, and E333S provide enhanced binding to the complement protein C1q and enhanced CDC. As described in Hinton et al., 2004, J. Biol. Chem. 279(8): 6213-6216, substitutions T250Q, T250E, M428L, and M428F provide enhanced binding to FcRn and improved pharmacokinetics. Modifications need not be restricted to the Fc region. It is also possible that the mutational differences in the Fab and hinge regions may provide optimized FcγR and/or C1q binding and/or effector function. For example, as disclosed in U.S. Ser. No. 60/614,944, and U.S. Ser. No. 60/619,409, filed Oct. 14, 2004, entitled "Immunoglobulin Variants Outside the Fc Region with Optimized Effector Function", the Fab and hinge regions of an antibody may impact effector functions such as antibody dependent cell-mediated cytotoxicity (ADCC), antibody dependent cell-mediated phagocytosis (ADCP), and complement dependent cytotoxicity (CDC). Thus immunoglobulin variants comprising substitutions in the Fc, Fab, and/or hinge regions are contemplated. For example, the antibodies may be combined with one or more substitutions in the VL, CL, VH, CH1, and/or hinge regions. Furthermore, further modifications may be made in non IgG1 immunoglobulins to corresponding amino acids in other immunoglobulin classes to provide more optimal properties, as described in U.S. Ser. No. 60/621,387, filed Oct. 21, 2004, entitled "IgG Immunoglobulin Variants with Optimized Effector Function". For example, in one embodiment, an IgG2 antibody, similar to the antibody presented in FIG. **25**, may comprise one or more modifications to corresponding amino acids in IgG1 or IgG3 CH1, hinge, CH2, and/or CH3. In another embodiment, an IgG2 antibody, similar to the antibody presented in FIG. **25**, may comprise all of the IgG1 CH1 and hinge substitutions,. i.e. said IgG2 variant comprises the entire CH1 domain and hinge of IgG1.

Example 2

Immunogenicity Reduction of C225

[0191] To illustrate further application of the method described in the present invention, and to validate its broad applicability to immunogenicity reduction of proteins, a second xenogeneic antibody example is provided using the variable region of C225 as the parent sequence (cetuximab, Erbitux®, Imclone®) (U.S. Pat. No. 4,943,533; PCT WO 96/40210). C225 is a murine anti-EGFR antibody, a chimeric version of which is currently approved for the treatment of cancer. A structural model of the murine C225 variable region was constructed using standard antibody modeling methods known in the art. FIGS. **26** and **27** show the sequences (SEQ ID NOS:234- and 235), host string content, and structures of the C225 VL and VH domains. A CDR graft of this antibody was constructed by placing the C225 CDRs into the context of the frameworks of the most homologous human germlines, determined to be vlk_6D-21 for VL and vh_4-30-4 for VH using the sequence alignment program BLAST. The sequences (SEQ ID NOS:236 and 237) and string content of these CDR grafts are shown in FIGS. **28** and **29**, along with structures of modeled C225 highlighting the mutational differences between the CDR grafted C225 variable chains and WT.

[0192] Variants with reduced immunogenicity were generated by applying a string optimization algorithm on the WT C225 VL and VH sequences, similar to as described above for AC10 except that single instead of multiple amino acid substitutions were sampled. HSC of each sequence was optimized using a window size w=9, and the same set of CDR and VL/VH interface proximal residues were masked. The calculation was run for C225 VL and VH in 100 separate iterations, generating a set of diverse C225 variants with more host string content than WT. FIG. **30** shows the nonredundant set of output sequences from these calculations for the C225 VL and VH regions (SEQ ID NOS:238-274), referred to as C225 VL HSC Calculation 1 and C225 VH HSC Calculation 1, respectively. In addition to the HSC score, the structural consensus and structural precedence of each sequence was evaluated (U.S. Ser. No. 60/528,229, filed Dec. 8, 2003, entitled

Protein Engineering with Analogous Contact Environments) in order to evaluate its structural integrity.

[0193] A second set of similar calculations were run on the C225 VL and VH sequences, except that the algorithm was allowed to sample multiple amino acid substitutions, rather than only single substitutions, in order to optimize HSC. FIG. **31** shows the nonredundant set of output sequences from these calculations for the C225 VL and VH regions (SEQ ID NOS:275-378), referred to as C225 VL HSC Calculation 2 and C225 VH HSC Calculation 2, respectively. Here, two measure of structural fitness, referred to as "Structural Consensus" and "Structural Precedence" (U.S. Ser. No. 60/528,229 and U.S. Ser. No. 60/602,566), are used to evaluate the structural and functional integrity of the sequences, in addition to HSC score. The output sequences were clustered based on their mutational distance from the other sequences in the set, and these clusters are delineated by the horizontal black lines in the Figure.

[0194] The calculations described above and presented in FIGS. **30** and **31** were used to generate a set of C225 VL and VH variants (SEQ ID NOS:238-378). In some cases, further substitutions were made to output sequences, using string and structural scores, as well as visual inspection of the modeled C225 structure, to evaluate fitness. FIGS. **32-40** present the sequences (SEQ ID NOS:379-387), structural scores, string scores, and mapped mutational differences on the modeled C225 structure for each of the C225 VL and VH variants. Iteration 21 from C225 VL HSC calculation 1 served as the precursor for L2 C225 VL, iteration 17 from C225 VL HSC calculation 2 served as the precursor for L3 C225 VL, and iteration 38 from C225 VL HSC calculation 2 served as the precursor for L4 C225 VL. Iteration 23 from C225 VH HSC calculation 1 served as the precursor for H3, H4, and H5 C225 VH, iteration 5 from C225 VH HSC calculation 2 served as the precursor for H6 C225 VH, iteration 41 from C225 VH HSC calculation 2 served as the precursor for H7 C225 VH, and iteration 44 from C225 VH HSC calculation 2 served as the precursor for H8 C225 VH.

[0195] Tables 7 and 8 present the mutational, structural fitness, and host string content scores for the C225 VL and VH variants as compared to the WT and CDR grafted C225 sequences. In addition, the maximum identity match to the germline for each string in the sequences was also determined, referred to as N_{ID}max. This represents the total number of strings in each sequence whose maximum identity to the corresponding strings in the human germline is the indicated value. For w=9, Tables 7 and 8 list N_9max, N_8max, N_7max, and $N_{\leq 6}$max for each sequence. Also provided is the framework region homogeneity. In addition to the aforementioned structural and host string analysis, each sequence was analyzed for its global homology to the human germline; tables 7 and 8 present the most homologous human germline sequence for each sequence (Closest Germline) and corresponding identity to that germline (ID to Closest Germline), determined using the sequence alignment program BLAST.

TABLE 7

C225 VL Variants

	WT	CDR Graft	L2	L3	L4
Mutations		25	17	21	18
Structural Consensus	0.49	0.52	0.56	0.58	0.54
Structural Precedence	0.53	0.56	0.57	0.59	0.57
Human String Content	0.79	0.94	0.91	0.92	0.91
Human String Similarity	0.15	0.65	0.51	0.58	0.57
Framework Region Homogeneity		0.97	0.52	0.50	0.78
N_9max	13	69	52	60	58
N_8max	27	15	28	24	23
N_7max	37	22	20	16	19
$N_{\leq 6}$max	30	1	7	7	7
Closest Germline	6D-21	6D-21	3-11	1D-13	6D-21
ID to Closest Germline	63/95 66%	87/95 91%	72/95 75%	73/94 77%	79/95 83%

TABLE 8

C225 VH Variants

	WT	CDR Graft	H3	H4	H5	H6	H7	H8
Mutations		33	18	21	15	21	22	28
Structural Consensus	0.44	0.48	0.51	0.46	0.49	0.52	0.49	0.53
Structural Precedence	0.55	0.54	0.55	0.54	0.51	0.55	0.58	0.55
Human String Content	0.67	0.84	0.79	0.81	0.77	0.79	0.79	0.79
Human String Similarity	0.04	0.56	0.36	0.41	0.33	0.36	0.35	0.33
Framework Region Homogeneity		0.97	0.45	0.52	0.50	0.50	0.76	0.77
N_9max	3	66	42	48	38	42	41	39
N_8max	23	17	25	24	17	24	21	27
N_7max	32	10	19	16	26	23	25	23
$N_{\leq 6}$max	61	26	33	31	38	30	32	30
Closest Germline	4-30-4	4-30-4	4-30-4	2-26	4-30-4	3-33	4-30-4	3-33
ID to Closest Germline	56/99 56%	88/99 88%	67/99 66%	74/99 74%	64/99 64%	69/99 70%	67/99 67%	80/99 81%

[0196] Again, whereas the CDR grafted C225 antibodies are most homologous to a single human germline sequence, the C225 variants of the present invention are homologous to different human germline sequences in different regions of the sequence. Whereas CDR grafted C225 VH is most homologous to human germline subfamily 4 across its entire sequence, H4 C225 VH is most homologous to subfamily 4 in FR1, subfamily 3 in FR2, and subfamily 2 in FR3. Additionally, whereas the CDR grafted antibodies are most homologous to a single germline sequence that is also the most homologous sequence to the parent sequence, the present invention presents a set of antibodies for a given antibody that are most homologous to different human germline sequences, which need not be the most homologous germline sequence to WT. For example, Table 7 shows that CDR grafted C225 VL is most homologous to vlk_6D-21, which is also the most homologous human germline to WT C225. However L2, L3, and L4 are most homologous to three different human germlines—ylk_3-11, ylk_1D-13, and vlk_6D-21 respectively. Thus the variants of the present invention explore a substantially greater amount of diversity than CDR grafted antibodies.

[0197] The genes for the C225 variable regions were constructed as described above, and subcloned into a modified pASK84 vector (Skerra, 1994, Gene 141: 79-84) comprising mouse constant regions for expression as Fabs. Select C225 variants were experimentally tested for their capacity to bind EGFR antigen. L2/H3 and L2/H4 C225 Fabs were expressed from the pASK84 vector in *E. Coli* with a His-tag, and purified using Nickel-affinity chromatography. Antigen affinity of the C225 variants was tested using SPR similar to as described above. EGFR extracellular domain (purchased commercially from R&D Systems) was covalently coupled to the dextrane matrix of a CM5 chip using NHS-linkage chemistry. C225 Fabs were reacted with the EGFR sensor chip surface at varying concentrations. Global Langmuir fits were been carried out for the concentrations series using the BiaEvaluation curve fitting software. The on-rate constant (ka), off-rate constant (kd), equilibrium binding constant (KD=kd/ka), and predicted saturation binding signal (Rmax) derived from these fits are presented in Table 9, along with the Chi2 which quantifies the average deviation of the fit curve from the actual data curve. The data indicate that both the L2/H3 and L2/H4 C225 variants bind EGFR antigen.

TABLE 9

SPR data on C225 Variants

C225	ka (1/Ms)	kd (1/s)	KD (M)	Rmax (RU)	Chi2
L2/H3	2.79×10^{4}	5.35×10^{-3}	1.92×10^{-7}	174	8.83
L2/H4	1.79×10^{4}	4.73×10^{-3}	2.64×10^{-7}	153	2.69

[0198] In order to investigate the anti-EGFR variants in the context of a full length antibody, the C225 WT (L0 and H0) and variant (L2, L3, L4, H3, H4, H5, H6, H7, and H8) regions were subcloned into the mammalian expression vector pcDNA3.1Zeo (Invitrogen) as described above. All combinations of the light and heavy chain plasmids were co-transfected into 293T cells, and antibodies were expressed, harvested, and purified as described above. Binding of the C225 WT (L0/H0) and variant (L0/H3, L0/H4, L0/H5, L0/H6, L0/H7, L0/H8, L2/H3, L2/H4, L2/H5, L2/H6, L2/H7, L2/H8, L3/H3, L3/H4, L3/H5, L3/H6, L3/H7, L3/H8, L4/H3, L4/H4, L4/H5, L4/H6, L4/H7, and L4/H8) antibodies was determined using SPR similar to as described above. Full length antibodies were flowed over the EGFR sensor chip described above. FIG. **41** shows the SPR sensorgrams obtained from the experiments. The curves consist of a association phase and dissociation phase, the separation being marked by a little spike on each curve. As a very rough approximation the signal level reached near the end of the association phase can be used as an indicator for relative binding. For all the curves this signal level is within 25% of the average level indicating that none of the antibody variants have significantly lost their ability to bind to EGFR.

[0199] To assess the capacity of the anti-EGFR antibodies to mediate effector function against EGFR expressing cells, the C225 variants were tested in a cell-based ADCC assay. Human peripheral blood monocytes (PBMCs) were used as effector cells, A431 epidermoid carcinoma cells were used as target cells, and lysis was monitored by measuring LDH activity using the Cytotoxicity Detection Kit as described above. FIG. **42** shows the dose dependence of ADCC at various antibody concentrations for WT and variant C225 antibodies. The results show that a number of the C225 variants have comparable or better ADCC than WT C225 with respect to potency and efficacy. These data may be weighed together with the antigen affinity data and other data to choose the optimal anti-EGFR clinical candidate. As exemplified above with AC10 variants, combinations of the C225 variants of the present invention with amino acid modifications that alter effector function are contemplated.

Example 3

Immunogenicity Reduction of ICR62

[0200] To further illustrate application of the method described in the present invention, and to validate its broad applicability to immunogenicity reduction of proteins, an example is provided using as the parent sequence the anti-EGFR antibody ICR62 (Institute of Cancer Research) (PCT WO 95/20045; Modjtahedi et al., 1993, *J. Cell Biophys.* 1993, 22(1-3):129-46; Modjtahedi et al., 1993, *Br J Cancer.* 1993, 67(2):247-53; Modjtahedi et al, 1996, *Br J Cancer,* 73(2): 228-35; Modjtahedi et al, 2003, *Int J Cancer,* 105(2):273-80). A structural model of the rat ICR62 variable region was constructed using standard antibody modeling methods known in the art. FIGS. **43** and **44** show the sequences (SEQ ID NOS:388 and 389), host string content, and structures of the ICR62 VL and VH domains. A CDR graft of this antibody was constructed by placing the ICR62 CDRs into the context of the frameworks of the most homologous human germlines, determined to be vlk_1-17 for VL and vh_1-f for VH using the sequence alignment program BLAST. The sequences and string content of these CDR grafts are shown in FIGS. **45** and **46** (SEQ ID NOS:390 and 391), along with structures of modeled ICR62 highlighting the mutational differences between the CDR grafted ICR62 variable chains and WT.

[0201] Variants with reduced immunogenicity were generated by applying a string optimization algorithm on the WT ICR62 VL and VH sequences, similar to as described above for AC10 except that single instead of multiple amino acid substitutions were sampled. HSC of each sequence was optimized using a window size w=9, and the same set of CDR and VL/VH interface proximal residues were masked. The calculation was run for ICR62 VL and VH in 100 separate interactions, generating a set of diverse ICR62 variants with more host string content than WT. FIG. **47** shows the nonredundant

set of output sequences (SEQ ID NOS:392 and 431) from these calculations for the ICR62 VL and VH regions, referred to as ICR62 VL HSC Calculation 1 and ICR62 VH HSC Calculation 1 respectively. In addition to the HSC score, the structural consensus and structural precedence of each sequence was evaluated (U.S. Ser. No. 60/528,229, filed Dec. 8, 2003, entitled Protein Engineering with Analogous Contact Environments) in order to evaluate its structural integrity.

[0202] The calculations described above and presented in FIG. **47** were used to generate a set of ICR62 VL and VH variants (SEQ ID NOS:392 and 455). In some cases, further substitutions were made to output sequences, using HSC and Structural Precedence scores, as well as visual inspection of the modeled ICR62 structure, to evaluate fitness. FIGS. **48-50** present the sequences (SEQ ID NOS:456 and 458), host string content, and mapped mutational differences on the modeled ICR62 structure for each of the ICR62 VL and VH variants. Iteration 20 from ICR62 VL HSC calculation 1 served as the precursor for L2 ICR62 VL. Iteration 1 from ICR62 VH HSC calculation 1 served as the precursor for H9, and iteration 5 from ICR62 VH HSC calculation 2 served as the precursor for H10 ICR62 VH.

[0203] Tables 10 and 11 present the mutational, structural fitness, and host string content scores for the ICR62 VL and VH variants as compared to the WT and CDR grafted ICR62 sequences. In addition, the maximum identity match to the germline for each string in the sequences, referred to as N_{ID}-max, is also provided, as well as the framework region homogeneity. In addition to the aforementioned structural and host string analysis, each sequence was analyzed for its global homology to the host germline; tables 10 and 11 present the most homologous host germline sequence for each sequence (Closest Germline) and corresponding identity to that germline (ID to Closest Germline), determined using the sequence alignment program BLAST.

TABLE 10

ICR62 VL Variants

	WT	CDR Graft	L2
Mutations	0	11	6
Structural Consensus	0.56	0.60	0.61
Structural Precedence	0.52	0.58	0.57
Human String Content	0.86	0.91	0.90
Human String Similarity	0.38	0.58	0.56
Framework Region Homogeneity	0.62	0.97	0.64
N_9max	37	59	56
N_8max	26	21	19
N_7max	31	18	22
$N_{\leq 6}$max	13	9	10
Closest Germline	1-17	1-17	1-17
ID to Closest Germline	76/95	86/95	81/95
	80%	90%	85%

TABLE 11

ICR62 VH Variants

	WT	CDR Graft	H9	H10
Mutations	0	34	20	21
Structural Consensus	0.43	0.44	0.46	0.45

TABLE 11-continued

ICR62 VH Variants

	WT	CDR Graft	H9	H10
Structural Precedence	0.42	0.52	0.47	0.49
Human String Content	0.64	0.85	0.79	0.79
Human String Similarity	0.01	0.54	0.28	0.33
Framework Region Homogeneity		1.00	0.64	0.85
N_9max	1	64	33	39
N_8max	16	24	33	30
N_7max	35	14	28	25
$N_{\leq 6}$max	67	17	25	25
Closest Germline	1-f	1-f	1-f	1-f
ID to Closest Germline	60/98	92/98	72/98	77/98
	61%	93%	73%	79%

[0204] Again, as observed from the significant differences in FRH and closest germlines, the ICR62 variants are homologous to different host germline sequences in different regions of the sequence. The genes for the ICR62 WT and L2/H9 variable regions were constructed as described above, and subcloned into a modified pASK84 vector (Skerra, 1994, Gene 141: 79-84). The ICR62 Fabs experimentally tested for their capacity to bind EGFR antigen. WT and L2/H9 ICR62 Fabs were expressed from the pASK84 vector in *E. Coli* with a His-tag, and purified using Nickel-affinity chromatography. Antigen affinity of the ICR62 antibodies was tested using SPR similar to as described above, with EGFR covalently coupled to the CM5 chip reacted with ICR62 antibodies at varying concentrations. The fits to the data, as described above, are provided in Table 12. fits were been carried out for the concentrations series using the BiaEvaluation curve fitting software. As can be seen, L2/H9 ICR62 binds with comparable affinity as WT to the EGFR antigen.

TABLE 12

SPR data on ICR62 Variants

ICR62	ka (1/Ms)	kd (1/s)	KD (M)	Rmax (RU)	Chi2
WT	9.86×10^4	2.53×10^{-5}	2.57×10^{-10}	402	1.86
L2/H9	2.35×10^5	1.06×10^{-4}	4.50×10^{-10}	508	4.91

Example 4

String Diversity Exploration of Immunoglobulins

[0205] The generation of mutational diversity based on HSC is much broader than the primary variant—secondary variant strategy described above for H3/L3 AC10. Indeed substitutions can be designed for any parent protein wherein the substitutions result in positive or neutral impact on the host string content of the parent sequence. Again, the advantage of such a strategy is that it generates a diverse set of minimally immunogenic variants that have the potential for optimized properties, including but not limited to antigen affinity, activity, specificity, solubility, expression level, and effector function. Such a set of variants may be designed, for example, to explore diversity for other parent immunoglobulins, including but not limited to nonhuman antibodies,

humanized or otherwise engineered antibodies (Clark, 2000, *Immunol Today* 21:397-402), (Tsurushita & Vasquez, 2004, Humanization of Monoclonal Antibodies, Molecular Biology of B Cells, 533-545, Elsevier Science (USA)), and "fully human" antibodies, obtained for example using transgenic mice (Bruggemann et al., 1997, *Curr Opin Biotechnol* 8:455-458) or human antibody libraries coupled with selection methods (Griffiths et al., 1998, *Curr Opin Biotechnol* 9:102-108).

Example 5

Unique Properties of Variant Proteins Generated by the Methods of the Present Invention

[0206] The methods described in the present invention generate variant proteins that possess a number of unique properties relative to variant proteins generated by other methods that attempt to achieve the same or similar goal. FIGS. **51** and **52** provide the host string content (HSC, Equation 3), exact string content (ESC, Equation 3a), and framework region homogeneity (FRH, Equation 10) of the AC10, C225, and ICR62 VH (FIG. **51**) and VL (FIG. **52**) variants of the present invention, compared with a number of antibody variable regions "humanized" by methods in the prior art. If a variant sequence's exact string content is derived solely from a single germline sequence, the FRH would be close to 1.0. Alternatively, as is the case with many of the variant sequences created by the present invention, FRH values can be significantly less than 1, with values ranging from 0.4 to 1.0, indicating, as expected, that sequences with high exact string content can be discovered with contributions from multiple germline subfamilies and sequences. At the same time, the variant sequences engineered using the present invention have high host string content, and thus are predicted to have low potential for immunogenicity in humans. For example, as shown in FIG. **51** variant VH sequences generated using the present invention have HSC values generally higher than 75%, and many of them have FRH values lower than 0.6, indicating their HSC is derived from multiple germline frameworks. As shown in FIG. **52**, similar trends apply for variant VL sequences generated using the present invention

[0207] Whereas particular embodiments of the invention have been described above for purposes of illustration, it will be appreciated by those skilled in the art that numerous variations of the details may be made without departing from the invention as described in the appended claims. All references are herein expressly incorporated by reference.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 1

<210> SEQ ID NO 1
<211> LENGTH: 98
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

```
Gln Val Gln Leu Val Gln Ser Gly Ala Glu Val Lys Lys Pro Gly Ala
1               5                   10                  15

Ser Val Lys Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Gly Tyr
            20                  25                  30

Tyr Met His Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp Met
        35                  40                  45

Gly Arg Ile Asn Pro Asn Ser Gly Gly Thr Asn Tyr Ala Gln Lys Phe
    50                  55                  60

Gln Gly Arg Val Thr Ser Thr Arg Asp Thr Ser Ile Ser Thr Ala Tyr
65                  70                  75                  80

Met Glu Leu Ser Arg Leu Arg Ser Asp Asp Thr Val Val Tyr Tyr Cys
                85                  90                  95

Ala Arg
```

We claim:

1. An antibody that binds the CD30 antigen, wherein said antibody comprises:

a) a heavy chain comprising a heavy chain variable region and a constant heavy region, wherein said variable heavy region is selected from the group consisting of:

i) the sequence of SEQ ID NO:223;

ii) the sequence of SEQ ID NO:226;

iii) the sequence of SEQ ID NO:227;

iv) the sequence of SEQ ID NO:228; and

b) a light chain comprising a light variable region and a light constant region, wherein said light variable region is selected from the group consisting of:

i) the sequence of SEQ ID NO:220;

ii) the sequence of SEQ ID NO:224; and

iii) the sequence of SEQ ID NO:225;

wherein the numbering of the variable region is according to the Kabat numbering system.

2. The antibody according to claim **1**, wherein said variable light chain is the sequence of SEQ ID NO:224.

3. The antibody according to claim **1** or **2**, wherein said variable heavy chain is the sequence of SEQ ID NO:227.

4. The antibody according to claim **3**, wherein said heavy chain constant region comprises a substitution at position selected from the group consisting of 239, and 332, wherein number of the constant region is according to the EU index.

5. The antibody according to claim **4**, wherein said substitution comprises 332E.

6. The antibody according to claim **4**, wherein said heavy chain constant region comprises 239D and 332E.

7. An antibody that binds the CD30 antigen, wherein said antibody comprises:

a) a heavy chain variable region comprising:
   i) a first CDR comprising amino acids 31-35 of SEQ ID NO:227;
   ii) a second CDR comprising amino acids 50-66 of SEQ ID NO:227;
   iii) a third CDR comprising amino acids 99-106 of SEQ ID NO:227; and

b) a light chain variable region comprising:
   i) a first CDR comprising amino acids 24-38 of SEQ ID NO:224;
   ii) a second CDR comprising amino acids 54-60 of SEQ ID NO:224;
   iii) a third CDR comprising amino acids 93-101 of SEQ ID NO:224;

wherein the numbering of the variable region is according to the Kabat numbering system.

8. A polynucleotide encoding the heavy and light chains of a human monoclonal antibody that competes for binding to human CD30 with an antibody comprising the heavy chain FR1 through FR4 amino acids sequences of SEQ ID NO:227 and the light chain FR1 through FR4 amino acid sequence of SEQ ID NO:224.

9. An expression vector comprising the polynucleotide of claim **7**.

10. A mammalian host cell line comprising the polynucleotide of claim **7**.

11. A method of generating a variant protein for a host as compared to a parent protein comprising:

a) comparing said parent protein sequence with two or more natural protein sequences from said host species;
b) analyzing one or more amino acid strings of said parent protein sequence with a corresponding amino acid string of each natural protein sequence;
c) substituting one or more amino acids of said parent protein sequence with a corresponding amino acid string of a natural protein sequence on an amino acid string by amino acid string basis;
d) wherein said variant protein has increased host string content as compared to said parent protein; and
e) wherein said substituted amino acids include a first substitution in a first string from a first natural protein, and a second substitution in a second string from a second natural protein.

* * * * *