

(21) Application No: 0912754.9  
(22) Date of Filing: 22.07.2009  
(30) Priority Data:  
(31) 61179850 (32) 20.05.2009 (33) US

(51) INT CL:  
C07K 16/46 (2006.01) C07K 16/00 (2006.01)  
C40B 4/08 (2006.01)

(56) Documents Cited:  
US 20060222654 A1  
Journal of Molecular Biology 376 (2008) Rothe et al  
'The Human Combinatorial Antibody Library HuCAL  
GOLD Combines Diversification of All Six CDRs  
According to the Natural Immune System with a Novel  
Display Method for Efficient Selection of High-Affinity  
Antibodies' 1182 - 1200  
Trends in Biotechnology 23 (2005) Lu 'Seamless  
cloning and gene fusion' 199 - 207

(71) Applicant(s):  
Novimmune S.A  
14 Ch Des Aulx, Plan-Les-Ouates, Geneva 1228,  
Switzerland

(72) Inventor(s):  
Ulla Ravn  
Franck Gueneau  
Nicolas Fischer  
Marie Kosco-Vilbois

(58) Field of Search:  
INT CL C07K, C40B  
Other: BIOSIS, MEDLINE, XPESP, EPODOC, WPI,  
TXTJP, TXTWOT

(74) Agent and/or Address for Service:  
Mintz Levin Cohn Ferris Glovsky and Popeo  
Intellectual Property LLP, Alder Castle,  
10 Noble Street, LONDON, EC2V 7JX,  
United Kingdom

(54) Title of the Invention: **Synthetic polypeptide libraries and methods for generating naturally diversified polypeptide variants**

Abstract Title: **Immunoglobulin variable domain production using type IIs restriction enzymes**

(57) Methods (claims 1 to 3) for producing a library of nucleic acids encoding immunoglobulin variable domains for custom recombinant antibody and intrabody production comprise i) providing a plurality of acceptor antibody framework regions (FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4) wherein one of the CDR regions is replaced by a stuffer DNA fragment comprising at least two Type IIs restriction sites interspersed by random sequence; ii) providing a plurality of diversified sequences that encode replacements for the missing CDR region flanked by type IIs restriction sites; iii) inserting the replacement CDR regions in place of the stuffer fragment by restriction digestion and ligation to produce complete immunoglobulin variable domains that do not contain the type IIs restriction sites used in the method. Independent method claims 26 to 28 further concern the expression and screening of the resultant immunoglobulin variable domains for affinity to a target antigen.

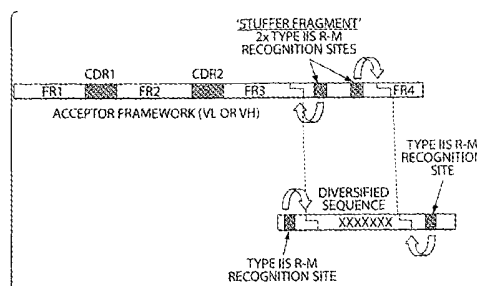


Fig. 2

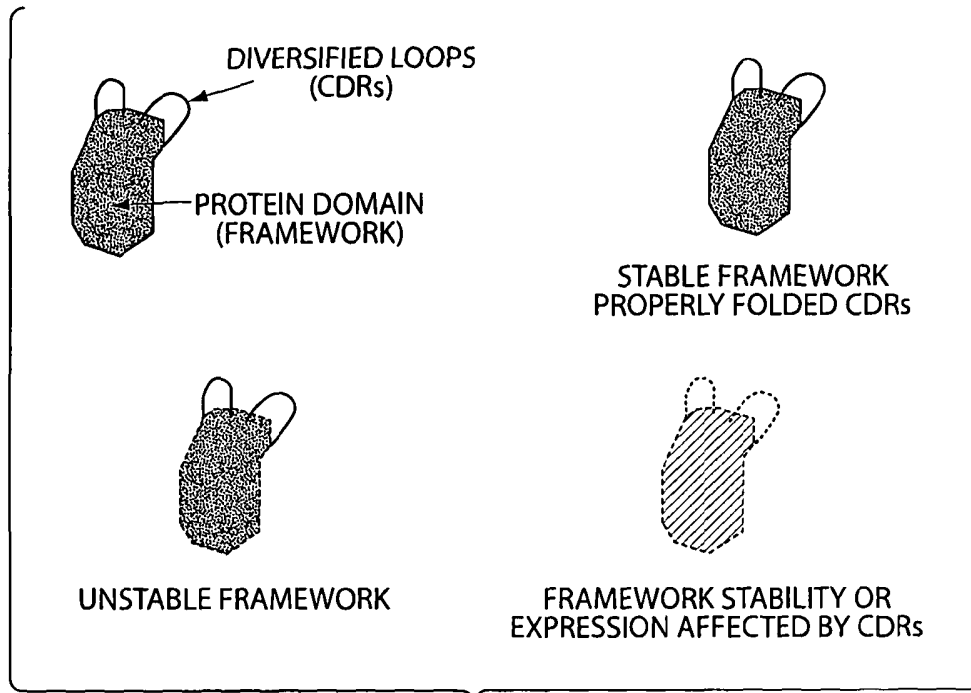


Fig. 1A

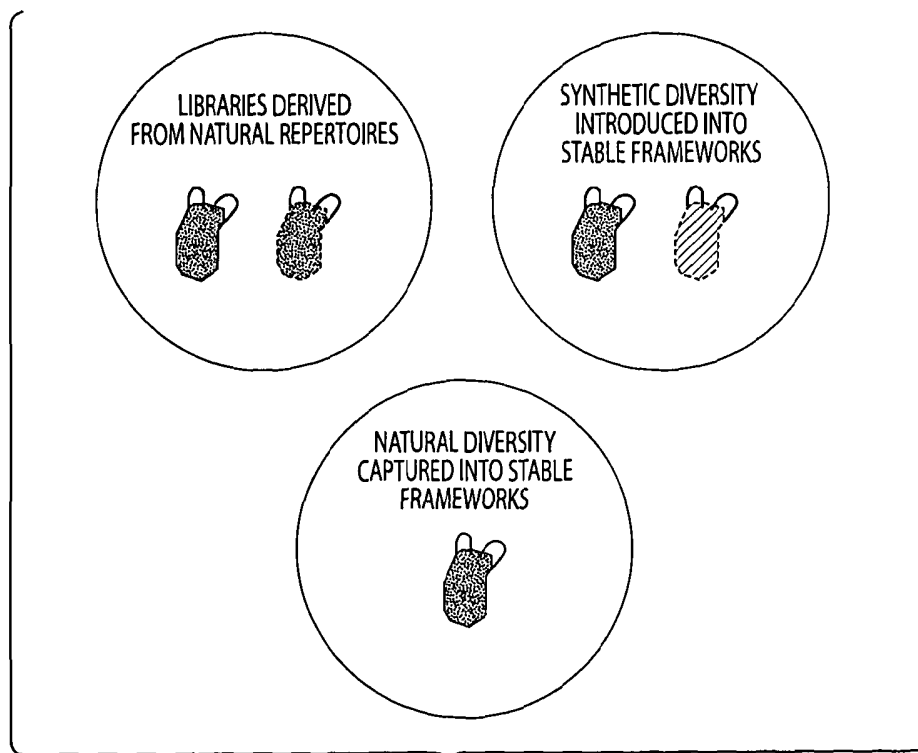
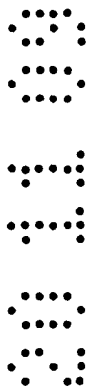


Fig. 1B

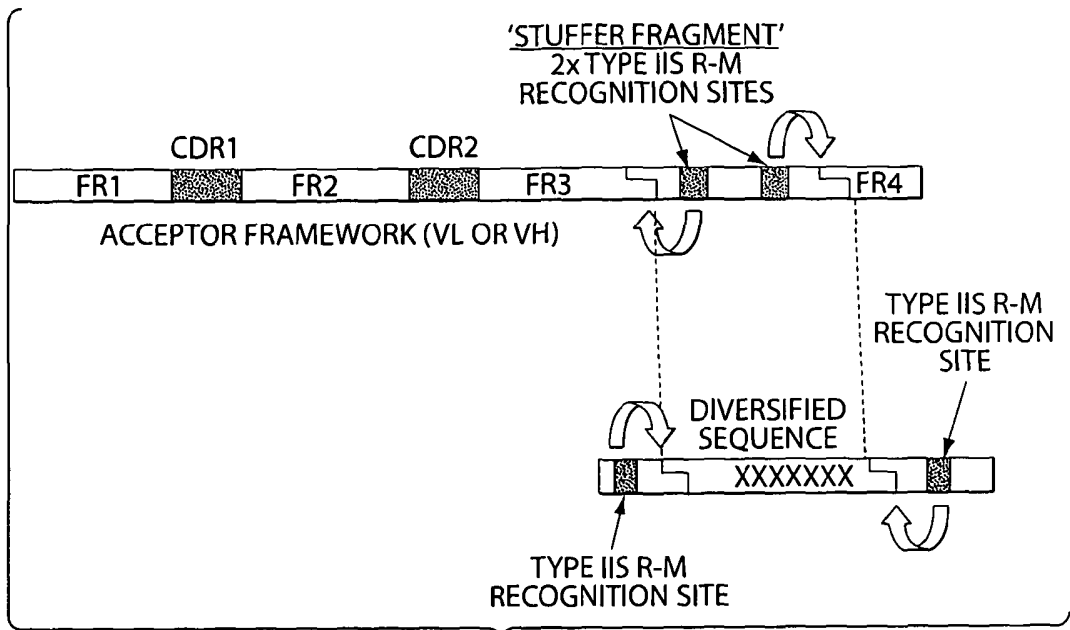


Fig. 2

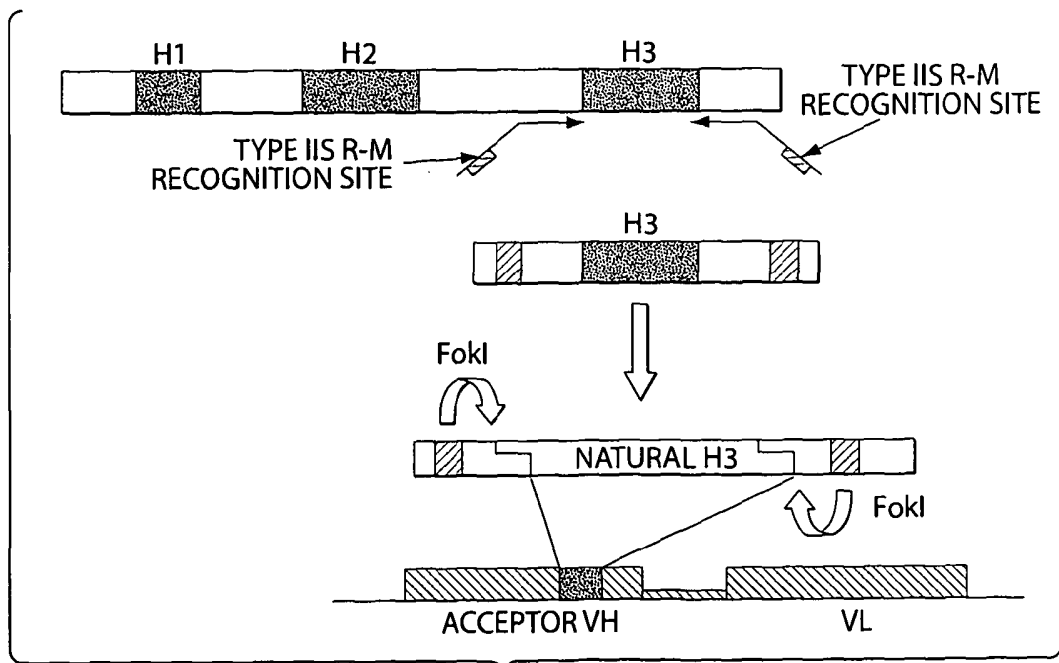
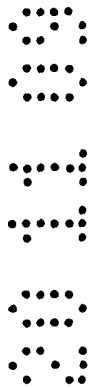


Fig. 3

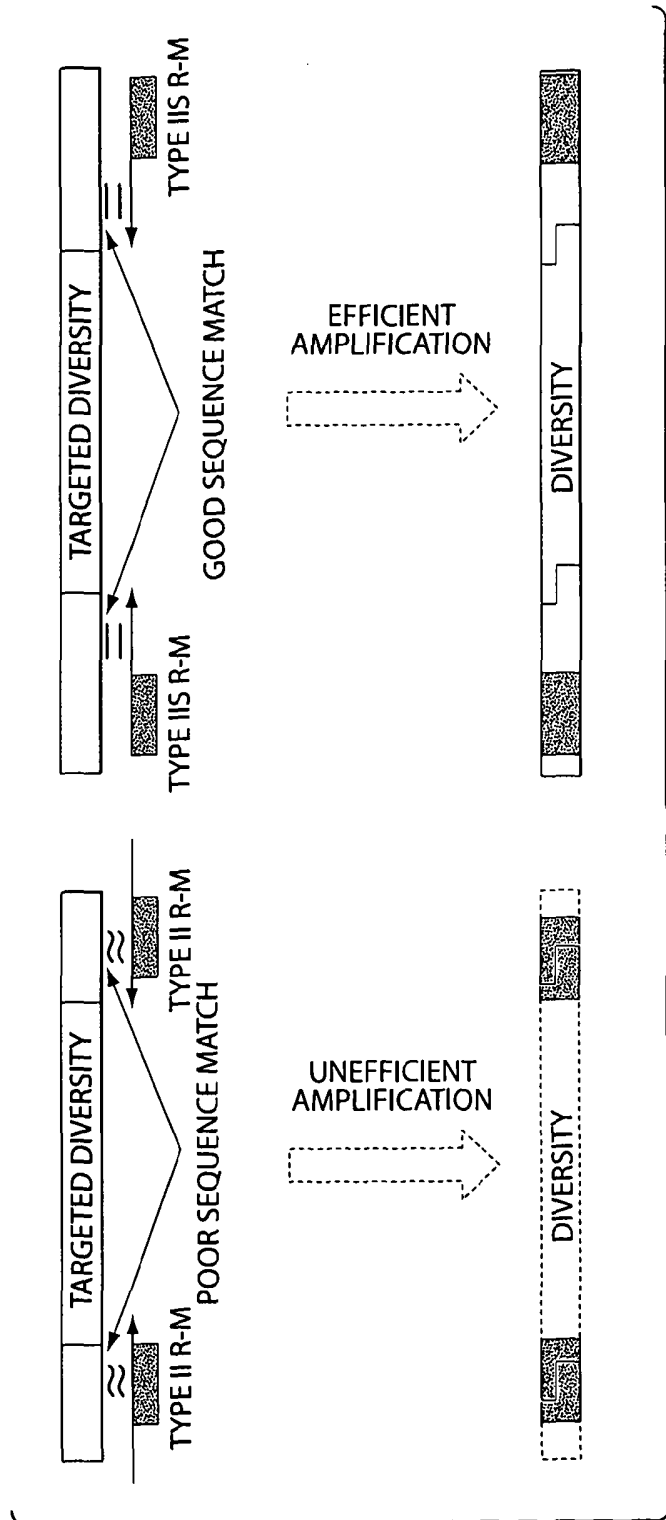
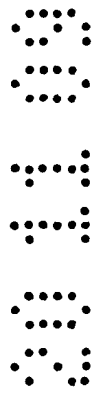
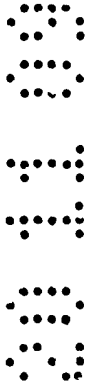


Fig. 4



VH1-2 QVOLVQSGAEVKKPGASVKVCKASGYTFTGYMHVWRQAPGGLEWMMGWIINPNSGNTVAOKFOGRVTMTRDTISITAYMELSLRSLRSDDTAVYYCAR SEQ ID NO: 1  
 VH1-18 QVOLVQSGAEVKKPGASVKVCKASGYTFTSYGISWVRQAPGGLEWMMGWIINPNSGNTVAOKFOGRVTMTRDTISITAYMELSLRSLRSDDTAVYYCAR SEQ ID NO: 2  
 VH1-69 QVOLVQSGAEVKKPGASVKVCKASGGTFSVAISWVRQAPGGLEWMMGWIIFGTANVAOKFOGRVTITADESTSTAYMELSSLRSEDTAVYYCAR SEQ ID NO: 3  
 VH3-30 QVOLVESGGGVQGRSLRLSCLAAAGTFFSSYAMHWVRQAPGKGLWVAVISYDGSNKYYADSVKGRFTISRDNKNTLYLQMNLSRAEDTAVYYCAR SEQ ID NO: 4  
 VH3-48 EVQLVESGGGLVQPGGSLRLSCLAAAGTFFSSYAMHWVRQAPGKGLWVAVISYDGSNKYYADSVKGRFTISRDNKNTLYLQMNLSRAEDTAVYYCAR SEQ ID NO: 5  
 VH3-23 EVQLLESGGGLVQPGGSLRLSCLAAAGTFFSSYAMSWVRQAPGKGLWVAVISYDGSNKYYADSVKGRFTISRDNKNTLYLQMNLSRAEDTAVYYCAK SEQ ID NO: 6  
 VH5-51 EVQLVQSGAEVKKPGESLKISCKGGYGFSTYIWGWVRQMPGKGLWMMGIIYPGDSDFRISPSFOGQVTISADKKSISTAYLQWSSLKASDTAMYYCAR SEQ ID NO: 7  
 VK1-33 DIQMTQSPSSLSASVGRVTITCAQASODIS NYLNWYQQKPKGAPKLLIYDASNLETVPSRFRSGSGGTDFTFISSLQPEDIATYYCQQYDNLPPITV SEQ ID NO: 8  
 VK1-39 DIQMTQSPSSLSASVGRVTITCRASQIS SYLNWYQQKPKGAPKLLIYAASSLQSGVPSRFRSGSGGTDFTFISSLQPEDFATYYCQSYSTPPTV SEQ ID NO: 9  
 VK3-11 EIVLTQSPATLSLSPGERATLSCRASQSVS SYLAWYQQKPKGAPRLLIYDASNRAITGIPARFSGSGGTDFTFISSLQPEDFAYVYCCQQRNWPPTV SEQ ID NO: 10  
 VK3-15 EIVMTQSPATLSVSPGERATLSCRASQSVS SNLAWYQQKPKGAPRLLIYGASTRATGIPARFSGSGGTEFTFISSLQSEDFAYVYCCQYNNWPPTV SEQ ID NO: 11  
 VK3-20 EIVLTQSPGTLSPGERATLSCRASQSVSSYLAWYQQKPKGAPRLLIYGASSRATGIPDRFSGSGGTDFTFISSLRLEPDEFAYVYCCQYGGSSPPTV SEQ ID NO: 12  
 VL1-44 QSVLTQPPSASGTPGQRVTIISCSGSSSNIISNTVNWYQQLPGTAPKLLIYSNNQRPSGVPDRFSGSKSGTSASLAIISGLQSEDEADYYCAAWDDSLNG SEQ ID NO: 13  
 VL1-51 QSVLTQPPPSVSAAPGQKVTIISCSGSSSNIISNTVNWYQQLPGTAPKLLIYENNRKPSGIIPDRFSGSKSGTSATLGTGLQTGDEADYYCGTWDSSLISA SEQ ID NO: 14

Fig. 5

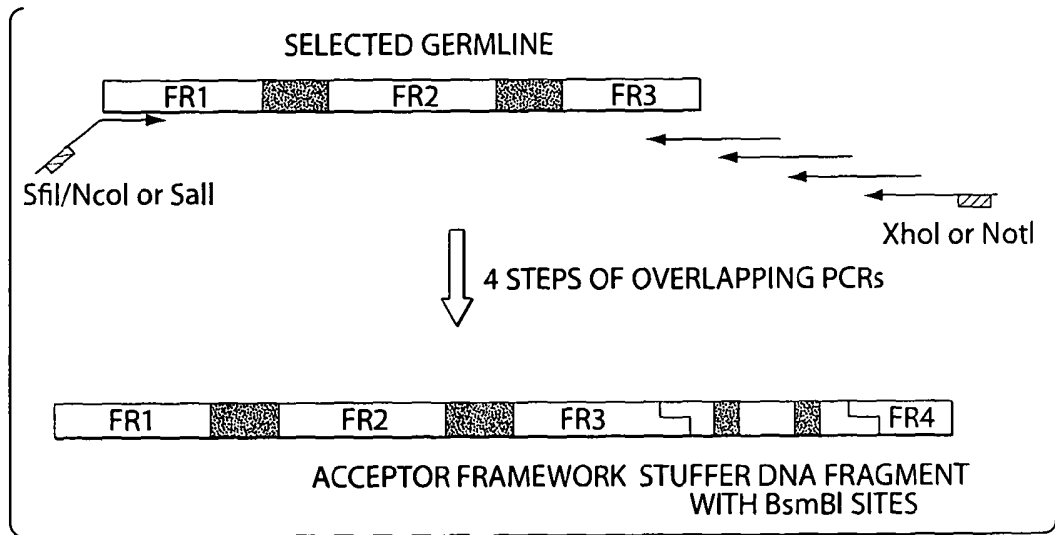
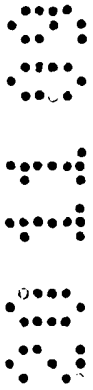


Fig. 6





550	2560	2570	2580	2590	2600	2610	2620	2630	2640	2650		
TATTA	CTGTCGAG	AGAGACCA	TAAACGGTAA	GGGTTTACC	AGGTTTAA	ACCGTATGG	AAGCGG	TCTCTG	TTGGGGC	AGGGAAC	CCCTGTCACAGTCTCG	SEQ ID NO: 15
VLLC	EMRR	ITVRI	TRFTR	FKRVL	GRVLR	GRVLR	GRVLR	GRVLR	GRVLR	GRVLR	GRVLR	SEQ ID NO: 16
YYCAR	.DER	.R	.GGLP	GLNAY	WEGAS	LSL	GLG	ARE	PWS	QSS	RL	SEQ ID NO: 17
ITV	RDE	TNNG	KAVY	QV	.TR	IGK	ARLL	GLG	PNP	PGH	SSL	SEQ ID NO: 18

5' ... CGTCTC|NNNNNNN...3' SEQ ID NO: 19  
 3' ... CGAGAG|NNNNNNN...5' SEQ ID NO: 20

Fig. 7



VH1-2 AF

```

cagtgacgctc tgggctgag gtgaagaagc ctggggcctc agtgaaggtc tcttgaagg cttctggata caccitcacc ggctactata tgcactgggt ggcagagccc cctggacaag
p g a a g a v w g - g e e a w g l s e g l l q g f w i h i h r l l y a l l g a t g p w t
q v q l v q s g a e v k k p g a s v k v s c k a s g y t f t g y y m h w v r q a p g q
p r c s w c s l g l r - r s l g p p q - r s p a r l l d t p s p a t i c t g c d r p l d k
>>.....VH1-2.....>
q v q l v q s g a e v k k p g a s v k v s c k a s g y t f t g y y m h w v r q a p g q

ggcttgagtg atcaacccta acagtgatgg cacaaactat gcacagaagt ttcaggcag ggtcaccatg accagggaca cgtccatcag cacagcctac atggagctga gcaggtgag
r a - v d g m d q p - q w w h k l c t e v s g q g h h d q g h v h g h s l h g a e q a e
g l e w m g w i n p n s g g t n y a q k f q g r v t m t r d t s i s t a y m e l s r l
g l s g w d g s t l t v v a q t m h r s f r a g s p - p g t r p s a q p t w s - a g -
>.....VH1-2.....>
g l e w m g w i n p n s g g t n y a q k f q g r v t m t r d t s i s t a y m e l s r l

atctgaqcac acggcctgtt attactgtgc gagatgagac gaataacggt aagggcgttt accagttta aacgctatt ggaagcgc gtctcttgg gccagggaac cctggtcaca gtctcg SEQ ID NO: 21
i - r h g r v l l c e m r i t v r i f t r i f k r v l g r i r v s w g q g t l v t v s SEQ ID NO: 22
r s d d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 23
d l t t r p c i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 24
>.....VH1-2.....>
r s d d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 23

```

Fig. 8



001103

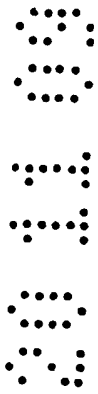
VH1-18 AF

caegtgcaqc tgggtgagtc tggagctgag gtgaagaagc ctggggcctc agtgaagtc tccgtcaagg cttctggtta cacctttacc agctatggtta tcagctgggt ggagacagcc cctggacaag  
p g a a g a v w s - g e e a w g l s e g l i l q g f w l h l y q l w y q l g a t g p w t  
q v q l v q s g a e v k k p g a s v k v s c k a s g y t f t s y g i s w v r g a p g q  
p r c s w c s l e l r - r s l g p q - r s p a r l l v t p l p a m v s a g c d r p l d k  
>>.....VH1-18.....>  
q v q l v q s g a e v k k p g a s v k v s c k a s g y t f t g y g i s w v r g a p g q

ggcttgatg gatgggatg atcagcgctt acaatggtta cacaaactat gcacagaagc tccagggcag agtcaccatg accacagaca catccaccag cacagcctac atggagctga ggagcctgag  
r a - v d g m d q r l q w - h k l c t e a p g q s h h d h r h i h e h s l h g a e e p e  
g l e w m g w i s a y n g n t n y a q k l q g r v t m t d t s t s t a y m e l r s l  
g l s g w d g s a l t m v t q t m h r s s r a e s p - p q t h p r a q p t w s - g a -  
>.....VH1-18.....>  
g l e w m g w i s a y n g n t n y a q k l q g r v t m t d t s t s t a y m e l r s l

atctgacac atcgccgtgt attactgtgc gagatgagac gaataacggt aagggcgttt accaggtta aacgcgtatt gggaaagcgc gtctcttggg gccaggaac cctggtcaca gtctcg SEQ ID NO: 25  
i - r h g r v l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 26  
r s d d t a v y y c a r - d e - r - g g l p p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 27  
d l t t r p c i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 28  
>.....VH1-18.....>  
r s d d t a v y y c a r - d e - r - g g l p p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 27

Fig. 8 Cont.



VH1-69 AF  
caggtgcagc tgggtcagtc tggggctgag gtgaagaagc ctgggcctc ggtgaagtc tctcaagg cttctgagg caccttcagc agctatgta tcagctgggt gcgacagccc cctggacaag  
p g a a g a v w g - g e e a w v l g e g l l q g f w r h l q q l c y q l g a t g p w t  
q v q l v q s g a e v k k p g s s v k v s c k a s g g t f s s y a i s w v r q a p g q  
p r c s w c s l g l r - r s l g p r - r s p a r l l e a p s a a m l s a g c d r p l d k  
>>.....VH1-69.....>  
q v q l v q s g a e v k k p g s s v k v s c k a s g g t f s s y a i s w v r q a p g q  
g g t t g a g t g a t g g g a g g a t c a t c c c t a t c t t t g g t a c a g c a a a c t a c g c a c a g a a g t t c c a g g c a g a g t c a c a g a g t a c c g g a g c a g a t c a g g a g c t a g g a g c t a g a g c c t g a g  
r a - v d g r d h p y l w y s k l r t e v p g q s h d y r g r i h e h s l h g a e q p e  
g l e w m g g i i p i f g t a n y a q k f q g r v t i t a d e s t s t a y m e l s s l  
g l s g w e g s s l s l v q q t t h r s s r a e s r l p r t n p r a q p t w s - a a -  
>.....VH1-69.....>  
g l e w m g g i i p i f g t a n y a q k f q g r v t i t a d e s t s t a y m e l s s l  
a t c t g a g a c a c g c c g t g t a t t a c t g t c g a t a g a c g a t a a c g g t t t a a c c g g t t a a a c c g t a t t g g g a a g c g c g t c t t t g g g c c c a g g a a c c c t g g t c a c a g t c t c g  
i - g h g r v l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s  
r s e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s  
d l r t r p c i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l  
>.....VH1-69.....>  
r s e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s  
SEQ ID NO: 29  
SEQ ID NO: 30  
SEQ ID NO: 31  
SEQ ID NO: 32  
SEQ ID NO: 31

Fig. 8 Cont.

001103

VH3-23R AF

```
gaggtgcagc tgttgagtc tggggagcc ttggtaacg ctaggggtc cctctggag cctctggatt cacctttagc agctatgcca tgagctgggt ccgccaggct ccagggaaag  
r g a a v g v w g r l g t a w g v p e t l l c s l w i h l l - q l c h e l g p p g s r e  
e v q l l e s g g g l v q p g g s l r l s c a a a g f t f s y a m s w v r q a p g k  
p r c s c w s l g e a w y s l g g p - d s p v q p l d s p l a a m p - a g s a r l q g r  
>>.....VH3-23.....>  
e v q l l e s g g g l v q p g g s l r l s c a a a g f t f s s y a m s w v r q a p g k  
  
ggctggagtg ggtctcagct attagtgta gtggtggtag cacatactac gcagactccg tgaaggccg gttcaccatc tccagagaca attccaagaa cacgctgtat ctgcaaatga acagctgag  
g a g v g l s y - w - w w - h i l r r l r e g p v h h l q r q f q e h a v s a n e q p e  
g l e w v s a i s g s g g s t y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
g w s g s g l l v v v v v a h t t q t p - I a g s p s p e t i p r t r c i c k - t a -  
>.....VH3-23.....>  
g l e w v s a i s g s g g s t y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
  
agccggagac acggccgtat attactgtc gagatgagac gaataacggt aagggcgttt accaggttta aacgcgtatt ggggaaggcc gtctcttggg gccagggaac cctggtcaca gtctcg SEQ ID NO: 33  
s r g h g r i l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 34  
r a e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 35  
e p r t r p y i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 36  
>.....VH3-23.....>  
r s e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 35
```

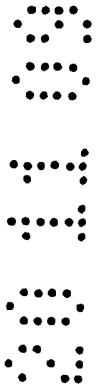
Fig. 8 Cont.

000103

VH3-23K AF

```
gggtgcagc tgttgagtc tgggggggc ttggtacagc ctggggggtc cctgagactc tctctgag cctctggatt cacctttagc agctatgcca tggagctgggt ccgccaggct ccaggaag  
r g a a v g v w g r l g t a w g v p e t l l c s l w i h l - g l c h e l g p p g s r e  
e v q l l e s g g g l v q p g g s l r l s c a a s g f t f s s y a m s w v r q a p g k  
p r c s c w s l g e a w y s l g g p - d s p v q p l d s p l a a m p - a g s a r l q g r  
>>.....VH3-23.....>>  
e v q l l e s g g g l v q p g g s l r l s c a a s g f t f s s y a m s w v r q a p g k  
ggctgagtg ggtctcagct attagtgtta gtggtggtag cacatactac gcagactccg tgaaggccg gttcaccatc tccagagaca attccaagaa cacctgtgtat ctgcaaatga acagctgag  
g a g v g l s y - w - w w - h i l r r l r e g p v h h l q r q f q e h a v s a n e q p e  
g l e w v s a i s g s g g s t y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
g w s g s q l l v v v v v a h t t q t p - r a g s p s p e t i p r t r c i c k - t a -  
>.....VH3-23.....>  
g l e w v s a i s g s g g s t y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
agccaggac acggccgtat attactgtc gaaatgagac gaataacggt aagcggttt accagttta aacgcgtatt gggaaggccg gtctcttggg gccagggaac cctggtcaca gtctcg SEQ ID NO: 36  
s r g h g r i l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 37  
r a e d t a v y y c a k - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 38  
e p r t r p y i t v r n e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 39  
>.....VH3-23.....>  
r a e d t a v y y c a k - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 38
```

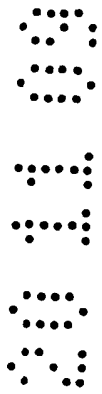
Fig. 8 Cont.



VH3-30 AF

```
cgggtgcagc tgggggagc gtggccagc ctgggagtc cctgagactc tctctgagc cctctggatt caccttcagt agctatgcta tgcactgggt ccgccaggct ccagcaagg  
p g a a g g v w g r r g p a w e v p e t l l c s l w i h l q - l c y a l g p p g s r q  
q v q l v e s g g g v v q p g r s l r l s c a a s g f t f s y a m h w v r q a p g k  
p r c s w w s l g e a w s s l g g p - d s p v q p l d s p s v a m l c t g s a r l q a r  
>>.....VH3-30.....>  
q v q l v e s g g g v v q p g r s l r l s c a a s g f t f s y a m h w v r q a p g k  
  
ggctggagt ggtggcagtt atatcatatg atggaagttaa taaatactac gcagactccg tgaaggccg attcaccatc tccagagaca attccaaga cacgctgtat ctgcaaatga acagctgag  
g a g v g g s y i i - w k - - i l r r l r e g p i h h l q r q f q e h a v s a n e q p e  
g l e w v a v i s y d g s n k y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
g w s g w q l y h m m e v i n t t q t p - r a d s p s p e t i p r t r c i c k - t a -  
>.....VH3-30.....>  
g l e w v a v i s y d g s n k y y a d s v k g r f t i s r d n s k n t l y l q m n s l  
  
agctgagac acgctgtgt attactgtc gagatgagac gaataacggt aagccggtt accagtttaaacgcgtatt gggaggccg gtctctggg gccagggaac cctgtcaca gtctcg SEQ ID NO: 40  
s - g h g c v l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 41  
r a e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 42  
e l r t r l c i t v i r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 43  
>.....VH3-30.....>  
r a e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 42
```

Fig. 8 Cont.



VH3-48 AF

```

gggtgcagc tgggtgagtc tgggggagc ttggtacagc ctggggggtc cctgagactc tctctgagc cctctggatt caccttcagt agctatagca tgaactgggt ccgccaggct ccagggaaagg
r g a a g g v w g r l g t a w g v p e t l l c s l w i h l q - l l - h e l g p p g s r e
e v q l v e s g g g l v q p g g s l r l s c a a s g f t f s s y s m n w v r q a p g k
p r c s w w s l g e a w y s l g g p - d s p v q p l d s p s v a i a - t g s a r l q g r
>>.....VH3-48.....>
e v q l v e s g g g l v q p g g s l r l s c a a s g f t f s s y s m n w v r q a p g k

```

```

ggctggagtg ggtttcatac attagtagta gtagtagtac catatactac gcagactctg tgaaggcgc attcaccatc tccagagaca atgccaagaa ctcaactgtat ctgcaaatga acagcctgag
g a g v g f i h - - - y h i l r l c e g p i h h l q r q c q e l t v s a n e q p e
g l e w v s y i s s s s s t i y y a d s v k g r f t i s r d n a k n s l y l q m n s l
g w s g f h t l v v v v v p y t t q t l - r a d s p s p e t m p r t h c i c k - t a -
>>.....VH3-48.....>
g l e w v s y i s s s s s t i y y a d s v k g r f t i s r d n a k n s l y l q m n s l

```

```

agccgagac accgctgtgt attactgtc gagatgagac gaataacggt aaggcggtt accaggttta aatcggtatt ggaaggcgc gtctctggg gccagggaac cctggtcaca gtctcg SEQ ID NO: 44
s r g h g c v l l c e m r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 45
r a e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 46
e p r t r l c i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 47
>>.....VH3-48.....>
r a e d t a v y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 46

```

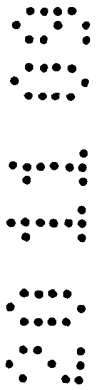
Fig. 8 Cont.



VH5-51 AF

gggtgcagtc tggagcagag gtgaaaaagc ccggggagtc tctgaagatc toctgaagg gtctcgata cagctttacc agctactgga tgggtcgggt gcgccagatg cccgggaaag  
r g a a g a v w s r g e k a r g v s e d l l l - g f w i q l y q l l d r l g a p d a r e  
e v q l v q s g a e v k k p g e s l k i s c k g s g y s f t s y w i g w v r q m p g k  
p r c s w c s l e q r - k s p g s l - r s p v r v l d t a l p a t g s a g c a r c p g k  
>>.....VH5-51.....>>  
e v q l v q s g a e v k k p g e s l k i s c k g s g y s f t s y w i g w v r q m p g k  
ggctggagtg gatggggtc atctatcctg gtgaactctga taacagatc agccctcct tccaaggcca ggtcaccatc tcagcgaca agtccatcag caccgcctac ctgcagtgga gcagctgaa  
r p g v d g d h l s w - l - Y q i q p v l p r p g h h l s r q v h q h r l p a v e q p e  
g l e w m g i i y p g d s d t r y s p s f q g q v t i s a d k s i s t a y l q w s s l  
a w s g w g s s i l v t l i p d t a r p s k a r s p s q p t s p s a p p t c s g a a -  
>.....VH5-51.....>  
g l e w m g i i y p g d s d t r y s p s f q g q v t i s a d k s i s t a y l q w s s l  
ggcctcgac accgccatgt attactgtc gagatgagac gaataacggt aaggcggtt accagttta aacgcgtatt gggaaaggcg gtctcttggg gccagggaac cctggtcaca gtctcg SEQ ID NO: 48  
g l g h r h v l l c e m r r i t v r r f t r f k r v l g r r v s w g q g t l v t v s SEQ ID NO: 49  
k a s d t a m y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 50  
r p r t p p c i t v r d e t n n g k a v y q v - t r i g k a r l l g p g n p g h s l SEQ ID NO: 51  
>.....VH5-51.....>>  
k a s d t a m y y c a r - d e - r - g g l p g l n a y w e g a s l g a r e p w s q s SEQ ID NO: 50

Fig. 8 Cont.

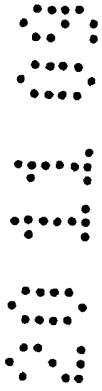


VK1-33G AF

```
gacatccaga tgaccagtc tcatactcc ctgtctgat ctgtaggaga cagagtcacc atcaattgcc aggcagtc gacattagc aactattaa attggtatca gcagaaacca gggaaagccc  
ghpddpvsillpvcicrirqshhhlpgesgh-qlfkllvsaretres  
diqumtqspsslsasvgsdirtvtitcqaasqdisnylnwyqqkpkp  
rtsr-pslhpplclhl-etespsla rrvirtlati-igisrnqgkpkp  
>>.....VK1-33G.....>  
diqumtqspsslsasvgsdirtvtitcqaasqdisnylnwyqqkpkp  
ctaagtccat gatctacgat gcatccaatt tggaaacagg ggtcccatca aggttcagtg gaagtggatc tggacagat ttactttca ccatacagc cctgcagcct gaagatattg caacatatta  
p-apidlrciqfgnrppikvqwkwiwdrfyfhhgqpaa-rycnill  
pklliydasnletgvpsrfsgsgtdftfthisslqpedit  
lss-stmhpniwkqgshqgsv evd lggqillspasacslkllqhl  
>>.....VK1-33G.....>  
pklliydasnletgvpsrfsgsgtdftfthisslqpedit  
cigtcagcag cgagacgaat aacggtaagg cggittacca ggittaaacy cgtattgga aggcgctct cattcgcca agggaccaag gtggaatca aagggcgc cgca SEQ ID NO: 52  
lsaarritvrrftxfkrvlgrrvsfgggtkveikgaaa SEQ ID NO: 53  
ycqqrde-rr-gglpplnawywegashsakgpprwxksgrrp SEQ ID NO: 54  
tvssetnngkavyqv-triqkarlirpirdqgnqirggr  
>>.....VK1-33G.....>  
ycqqrde-rr-gglpplnawywegashsakgpprwxksgrrp SEQ ID NO: 54
```

Fig. 8 Cont.

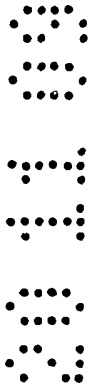




VH1-33R AF

gacatccaga tgaccagtc tccatcctcc ctgtctgat ctgtaggaga cagagtcacc atcactgcc aggcagta ggcattagc aactattaa attggtatca gcagaacca gggaaagccc  
ghpddpv silpvc icrir qsh hhl pges gh- qlfk lvs aet res  
d i q m t q s p s s l s a s v g d r v t i t c q a s q d i s n y l n w y q q k p g k a  
r t s r - p s l h p p c l h l - e t e s p s l a r r v r t l a t i - i g i s r n q g k p  
>>.....VK 1-33R.....>>  
d i q m t q s p s s l s a s v g d r v t i t c q a s q d i s n y l n w y q q k p g k a  
ctaagtcct gatctacgat gcatccaatt tggaaacagg ggtccatca aggttcagtg gaagtgatc tggacagat ttactttca ccatecagcag cctgcagcct gaagatattg caacatatta  
p - a p d l r c i q f g n r g p i k v q w k w i w d r f y f h h q q p a a - I Y c n i l  
p k l l i y d a s n l e t g v p s r f s g s g t d f t f t i s s l q p e d i a t y  
l s s - s t m h p i w k q g s h q g s v e v d l g q i l l s p s a a c s l k i l q h i  
>>.....VK 1-33R.....>>  
p k l l i y d a s n l e t g v p s r f s g s g t d f t f t i s s l q p e d i a t y  
ctgtcagcag cgagacgaat aacgtaagg cggttacca ggtttaacg cgtattgga aggcgctct cattcgcca agggaccaag gtggaatca aatcggcggc cyca SEQ ID NO: 56  
l s a a r r i t v r r f t r f k r v l g r r v s f g q g t k v e i k r a a a SEQ ID NO: 57  
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 58  
t v s s e t n n g k a v y q v - t r i g k a r l i r p r d q g g n q t g g r  
>.....VK 1-33R.....>>  
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 58

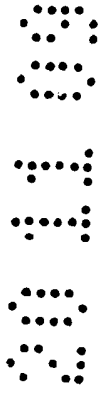
Fig. 8 Cont.



VK1-39G AF

gacatccaga tgaccacgct tccatcctcc ctgtctgcat ctgtaggaga cagagtcacc atcacttgcc gggcaagta ca gacattagc agctatttaa attggtatca gcagaaacca gggaaagccc  
ghpddpvsillpvcicrirqshhhlpgkseh-qllfkllvsaetreress  
diqmtqspsslsasvgsdrrvtitcraasqsis syl nwy qqkpkp  
rtsr- pslhpl-clhl-etespslaagqvraali-igisrnqqkpkp  
>>.....VK1-39G.....>>  
diqmtqspsslsasvgsdrrvtitcraasqsis syl nwy qqkpkpkp  
ctaagctect gatctatgct gcatccagtt tgcaagtgg ggtcccatca aggttcagtg ggaagatc tggacagat ttcactctca coactcagcag tctgcaacct gaagattttg caacttacta  
p- apdlcciqfakwgpikvqwwqwiwdrrfhs h h q q s a t - I f c n l l  
pklliyaaassllqs gvpssrfsgsgtdftltis slqp edf aty  
lss - smllhplvc khqgs vavdlggqislspsavcnllqlt  
>.....VK1-39G.....>  
pklliyaaassllqs gvpssrfsgsgtdftltis slqp edf aty  
ctgtcagcag cgagacgaat aacggtaagg cggtttaccga ggtttaaacg cgtattggga agggcgtct cattcgcca agggaccaag gtggaatca aagggcgcc cgca SEQ ID NO: 60  
lsaarritvrrftrfkrvlgrvrvsfgggtkveikgaaa SEQ ID NO: 61  
ycqqqrde-rr-gglp glnaywegas hsa kgp r w k s k g r p SEQ ID NO: 62  
tvssetnngkavyqv-t r i g k a r l i r p r d q g g n q r g g r  
>.....VK1-39G.....>  
ycqqqrde-rr-gglp glnaywegas hsa kgp r w k s k g r p SEQ ID NO: 63

Fig. 8 Cont.



VH1-39G AF

gacatccaaga tgaccacgctc tccatccctc ctgtctgcat ctgtaggaga cagagtcacc atcacctgccc gggcaagta ca gacattagc agctatitaa attggtatca gcagaaacca gggaaagccc  
ghpddpv silpvc icrir qshhhlpgkseh-qllfk lvs aet res  
d i q m t q s p s s l s a s v g d r v t i t c r a s q s i s s y l n w y q q k p g k a  
r t s r - p s l h p p c l h l - e t e s p s l a g q v r a l a a i - i g i s r n q g k p  
>>.....VK 1-39R.....>  
d i q m t q s p s s l s a s v g d r v t i t c r a s q s i s s y l n w y q q k p g k a  
ctaagctct gatctatgct gcatccagtt tgcaagtgg ggtccatca aggttcagtg gmagtgatc tggacagat tccactcca ccatcagcag tctgcaacct gaagattttg caacttacta  
p - a p d l c c i q f a k w g p i k v q w q w i w d r f h s h h q q s a t - I f c n l l  
p k l l i y a a s s l q s g v p s r f s g s g s g t d f t l t i s s l q p e d f a t y  
l s s - s m l h p v c k v g s h q g s v a v d l g q i s l s p s a v c n l k l l q l t  
>.....VK 1-39R.....>  
p k l l i y a a s s l q s g v p s r f s g s g s g t d f t l t i s s l q p e d f a t y  
ctgtcagcag cgagacgaat aacggttaagg cggtttacca ggtttaaacg cgtattggga aggcgctct cattcgcca agggaccaag gtggaatca aacgggcggc cgca SEQ ID NO: 64  
l s a a r r i t v r r f t r f k r v l g r r v s f g q g t k v e i k r a a a SEQ ID NO: 65  
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 66  
t v s s e t n n g k a v y q v - t r i g k a r l i r p r d q g g n q t g g r I  
>.....VK 1-39R.....>  
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 67

Fig. 8 Cont.



VK3-11G AF

```
gaaattggt tgacacagtc tccagccacc ctgtcttgtt ctccaggga aagagccacc ctctctgca gggccagtc gagtgtagc agctacttag cctggtacca acagaaacct gggccaggctc
gncvdtvssshpvsrfvshpshpllqgqsecc-qllslvptetwpgg
eivltqspatlsllspgerratpssqrasqsvslyllawyqqkplgqa
rklc-hsllqpplcclclqgkrepsspa gpvrvl aat-pgt n r n l a r l
>>.....VK 3-11G.....>>
eivltqspatlsllspgerratlsclrassqsvslyllawyqqkplgqa

ccagctct catctatgat gcatccaaca gggccactgg catccacc accgttcagtg gaagtggctc tgggacagac tteactctca ccatcagcag cctagagcct gaagatttg cagtttatta
sqaphl-Ciqqghwhpsqvqvwvdrlhshhqqqpra-irfcsl
prlliydasnratgiparfsqsgsgtdftltisplep edfavy
psssmhptgplasqpssvavq lggqtslspasa-sllqfi
>.....VK 3-11G.....>
prlliydasnratgiparfsqsgsgtdftltisplep edfavy

ctgtcagcag cgagacgaat aacgtaagg cggttacca ggtttaaacy cgtattgga aggcctct cattcgcca agggaccaag gtggaatca aagggcggc cgca
lsarritvrrftrfkrvlgrrvsgfggtkveikgaaa
ycqqrde-rgglplp glnaywegaa shsakgp r wks kgr p
tvssetnngkavyqv-t r i g k a r l i r p r d q g g n q r g g r
>.....VK 3-11G.....>
ycqqrde-r-gglplp glnaywegaa shsakgp r wks kgr p
```

SEQ ID NO: 68  
SEQ ID NO: 69  
SEQ ID NO: 70  
SEQ ID NO: 71  
SEQ ID NO: 70

Fig. 8 Cont.

VH3-11R AF

```

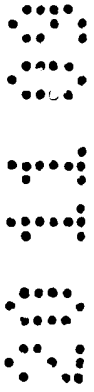
gaaattggtg tgacacagtc tccagccacc ctgtcttggc ctccagggga aagagccacc ctctctgca gggccagtc gagtgtagc agctacttag ccttggtacca acagaaacct gggcaggctc
gncvdtvsshpvfvsvrgkshplllqgqsec-qllslvptetwpgg
eivltqspatlsllspgerratlscrasqsvsylvawyqqkplgqa
rklc-hsllqpplcclclqqgkeepspagpvrvlaat-pgtnrnlarl
>>.....VK 3-11R.....>
eivltqspatlsllspgerratlscrasqsvsylvawyqqkplgqa
>>.....VK 3-11R.....>

ccagctcct catctatgat gcaccaaca gggccactgg catccagcc aggttcagtg gaagtggtc tggacagac ttcactctca ccatcagcag cctagagcct gaagatttg cagtttatta
sqaphl-ciqqghwhpsqvqwwvdrlhshhqqprra-irfcsl
prlliydasnratgiparfsqsgsgtdftltisslepedefavy
pgsssmhpttgplassqpgsvavqlgqtslspasasllkllqfi
>>.....VK 3-11R.....>
prlliydasnratgiparfsqsgsgtdftltisslepedefavy
>>.....VK 3-11R.....>

ctgtcagcag cgagacgaat aacggttaagg cggtttacca ggtttaaacg cgtattggga aggcgctct cattcgcca agggaccaag gtggaatca aacggggcgc cgca SEQ ID NO: 72
lsaarritvrrftrfkrvlgrrvsfgggtkveikraa SEQ ID NO: 73
ycqqrde-rr-gglpplnawyegaaashsakkgprrwksngrp SEQ ID NO: 74
tvssetnngkavyqv-tirigkarliprdqgnqtggr SEQ ID NO: 75
>>.....VK 3-11R.....>
ycqqrde-rr-gglpplnawyegaaashsakkgprrwksngrp SEQ ID NO: 74

```

Fig. 8 Cont.



VK3-15G AF

```
gaaatagtga tgacgcagtc tccagccacc ctgtctgtgt ctccaggga aagagccacc ctctctca gggccagta gagtgttagc agcaacttag ctgtgtacca gcagaaacct gcccaggctc
gnsddaavsshsvspat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat
eivmtqspat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat
>>.....VK 3-15G.....>
eivmtqspat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat lsvspggrat
>>.....VK 3-15G.....>
ccaggctct catctatggt gcattcaacca gggccactgg tatccagcc aggttcagtg gaagtgggtc tgggacagag ttcactctca ccatacagag ctgcagctt gaagattttg cagtttatta
sqaphlwcihgghwypsqvqwgwvwdrvvhs h h q q p a v - r f c s l l
p r l l i y g a s t r a t g i p a i f s g s g s g t e f t l t i s s l q s e d f a v y i
p g s s s m v h p p g p l v s q p g s v a v g l g q s s l s p s a a c s l k i l q f i
>>.....VK 3-15G.....>
p r l l i y g a s t r a t g i p a r f s g s g s g t e f t l t i s s l q s e d f a v y
ctgtcagcag cgagatcaag aacggttaag cggtttacca ggtttaaacg cgtattgga aggcgctct cattcgcca agggaccaag gtggaatca aagggggcgc cgca SEQ ID NO: 76
l s a a r i t v r r f t r f k r v l g r r v s f g q g t k v e i k g a a a SEQ ID NO: 77
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s k g r p SEQ ID NO: 78
t v s s e t n n g k a v y q v - t r i g k a r l i r p r d q g g n q r g g r
>>.....VK 3-15G.....>
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s k g r p SEQ ID NO: 79
```

Fig. 8 Cont.



VK3-15R AF

```

gaaatagtga tgacgcagtc tccagccacc ctgtctgtgt ctccaggga aagagccacc ctctctgca gggccagta gagtgtagc agcaacttag cctgggtacca gcagaaacct gggcaggctc
gns d d a v s s h p v c v s r g k s h p l l q g q s e c - g q l s l v p a e t w p g
e i v m t q s p a t l s v s p g e r a t l s c r a s q s v s n l a w y q q k p l a r l
r k - - r s l q p p c l c l q g k e p p s p a g p v r v l a a t - p g t s r n l a r l
>>.....VK 3-15R.....>
e i v m t q s p a t l s v s p g e r a t l s c r a s q s v s n l a w y q q k p g q a

ccaggctct catctatggt gcattcaaca gggccactgg tatcccagcc aggttcagtg ggaagggtc tgggaagag ttcactctca ccatacagag cctgcagctc gaagatttg cagtttatta
s q a p h l w c i h q g h w y p s q v q w q w v w d r v h s h h q q p a v - r f c s l l
p r l l i y g a s t r a t g i p a r f s g s g s g t e f t l t i s s l q s e d f a v y
p g s s s m v h p p g p l v s q p g s v a v q l g q s s l s p s a a c s l k i l q f i
>.....VK 3-15R.....>
p r l l i y g a s t r a t g i p a r f s g s g s g t e f t l t i s s l q s e d f a v y

ctgtcagcag cgagacgaat aacggttaagg cggtttacca ggtttaaacg cgtattgga aggcgctct cattcgcca agggaccaag gtggaatca aacgggcggc cgca
l s a a r r i t v r r f t r f k r v l g r r v s f g q g t k v e i k r a a a
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p
t v s s e t n n g k a v y q v - t r i g k a r l i r p r d q g g n q t g g r
>.....VK 3-15R.....>
y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s a g r p

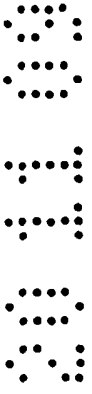
```

```

SEQ ID NO: 80
SEQ ID NO: 81
SEQ ID NO: 82
SEQ ID NO: 83
SEQ ID NO: 82

```

Fig. 8 Cont.



VK3-20G AF

```

gaaattgtgt tgaccagatc tccaggacc ctgtcttgt ctccaggga aagagcacc ctctctgca gggccagica gagtgtagc agcagctact tagcctgta ccagcagaaa cctggccagg
gncvdavsrhpvfvsrgkshpvlqgqsecc-qqlslvpaetwpe
eivltqspgtlslspgerratlscrasqsvsssy law Yqqkpgq
rklc-rsliqapclclqgkelpspagpvrvlaaat-ppgtstrnlair
>>.....VK 3-20G.....>
eivltqspgtlslspgerratlscrasqsvsssy law Yqqkpgq
ctcccaggct cctcatctat ggtgatcca gaggccac tggatcca gacagttca gtggagtg gtctggaca gacttactc tcaccatcag cagactggag cctgaagatt ttgcagttta
gsqaphlwciqqghwhp rqvqwvwdrllhshhggtg a-irfcsl
aprllygasstratgipdrfsgsgst dftlti srlepedfav
lpgsssmvhp agp lasgtgs vavglg gtsl sps adwslkidlqf
>.....VK 3-20G.....>
aprllygasstratgipdrfsgsgst dftlti srlepedfav
ttactgtag cagcagatc aataacgta agcggttta ccaggttta acggtattg gaaagccg tctcattcgg ccaaggacc aagtggaat tcaaaagggc ggccgca
llsaaarritvrrftrfkrvlgrrvsvf gqgtkveikgaaa
yycqqrde-ir-gglpgl n ay wega shs ak g prwk skg r p
itv sset nng kav yqv - triggkairlirprdqgnqirgr
>.....VK 3-20G.....>
yycqqrde-ir-gglpgl n ay wega shs ak g prwk skg r p

```

SEQ ID NO: 84  
 SEQ ID NO: 85  
 SEQ ID NO: 86  
 SEQ ID NO: 87  
 SEQ ID NO: 86

Fig. 8 Cont.



VK3-20R AF

gaaattggt tgaccgagtc tccaggacc ctgtctttgt ctccaggga aagagccacc ctctctqca gggccagta gagtggtagc agcagctact tagcctggta ccagcagaaa cctggccagg  
 gnc v d a v s r h p v f v s r g k s h p l l q g q s e c - q q l l s l v p a e t w p  
 e i v l t g s p g t l s l s p g e r a t l s c r a s q s v s s s y l a w y q q k p g q  
 r k l c - r s l q a p c l c l q g k e p p s p a g p v r v l a a a t - p g t s r n l a r  
 >>.....VK 3-20R.....>>  
 e i v l t q s p g t l s l s p g e r a t l s c r a s q s v s s s y l a w y q q k p g q

ctccaggct cctcatctat ggtgatcca gaggccac tggatcca gacaggttca gtggcagtg gctcggaca gacttcactc tcaccatcag cagactggag cctgaagatt ttgcagttta  
 g s q a p h l w c i q q g h w h p r q v q w q w v w d r l h s h h q q t g a - I f c s l  
 a p r l l i y g a s s r a t g i p d r f s g s g s g t d f t l t i s r l e p e d f a v  
 l p g s s s m v h p a g p l a s q t g s v a v g l g q t s l s p s a d w s l k i l q f  
 >.....VK 3-20R.....>  
 a p r l l i y g a s s r a t g i p d r f s g s g s g t d f t l t i s r l e p e d f a v

ttactgtcag cagcgagacg aataacggta aggcggttta ccaggtttaa acgcgtattg ggaagcgcg tctcattcgg ccaaggacc aaggtggaaa tcaaacgggc ggccgca SEQ ID NO: 88  
 l l s a a r r i t v r r f t r f k r v l g r r v s f g q g t k v e i k r a a a SEQ ID NO: 89  
 y y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 90  
 i t v s s e t n n g k a v y q v - t r i g k a r l i r p r d q g g n q t g g r SEQ ID NO: 91  
 >.....VK 3-20R.....>>  
 y y c q q r d e - r - g g l p g l n a y w e g a s h s a k g p r w k s n g r p SEQ ID NO: 90

Fig. 8 Cont.



V11-44 AF

cagtctgtg tgcagcagcc gccctcagtg tctgggccc caggacagaa ggtcaccatc tctctctctg gaagcagctc caacattggg aataattatg tatcctggta ccagcagctc ccaggaacag  
 avc vdaaalsvcg prte ghhlwklqlqhwe-lcivpaa prn  
 qsvltqppsvsaa pggqkvtkvtiscsgssni gnny vsw Yqqqlpgt  
 rslc-rs rppqclrp qdr rsvp s pal eaa p t l g i m y p g t s s s q e q  
 >>.....VL1-51.....>>  
 qsvltqppsvsaa pggqkvtkvtiscsgssni gnny vsw Yqqqlpgt

cccccaaact cctcatttat gacaataata agcgaccctc agggattctet gaccgattct ctggtcccaa gtctggcagc tcagccacc tgggcatcac cggactccag actggggacg aggccgatta  
 spqtp h l - g - - a t l r d s - p i l w l q v w h v s h p g h h r t p d w g r g r l  
 a p k l l i y d n n k r p s g i p d r i f s g s k s g t s a a t l g i t g l q t g d e a d  
 p p n s s f m t i i s d p q g f l t d s l a p s l a r q p p w a s p d s r l g t r p i  
 >.....VL1-51.....>  
 a p k l l i y d n n k r p s g i p d r i f s g s k s g t s a t l g i t g l q t g d e a d

ttactgcgga acaggagacg aataacgga aggcggttta ccaggttta acgcgtattg gaaagcgcg tctctgtgtt cggcgaggg accaagtga cgcctcagg ggcggcgca  
 l l r n r r i t v r r f t r f k r v l g r r v s v f g g g t k l t v l g a a a  
 Y Y c g t g d e - r - g g l p g l n a y w e g a s l c s a e g p s - p s - g r p  
 i t a e q e t n n g k a v y q v - t r i g k a r l c v r i r d q a d r p r g g r  
 >.....VL1-51.....>  
 Y Y c q t g d e - r - g g l p g l n a y w e g a s l c s a e g p s - p s - g r p

SEQ ID NO: 92  
 SEQ ID NO: 93  
 SEQ ID NO: 94  
 SEQ ID NO: 95

Fig. 8 Cont.



VL1-51 AF

```

cagctgtgc tgactcagcc accctcagcg tctgggacc ccgggcagag ggtcaccatc tcttttctg gaagcagctc caacatgga agtaatactg taaactggta ccagcagctc ccaggaaagc
avc adsa t l s v w d p r a e g h h l l f w k q l q h r k - y c k l v p a a p r n
q s v l t q p p s a s g t p g q r v t i s c s g s s s n i g s n t v n w y q q l p g t
r s l c - l s h p q r l g p p g r g s p s l v l e a a p t s e v i l - t g t s s s q e r
>>.....VL1-44.....>
q s v l t q p p s a s g t p g q r v t i s c s g s s s n i g s n t v n w y q q l p g t

cccccaact cctcatctat agtaataatc agggccctc aggggtccct gaccgattct ctggctcaa gtctgcacc tcagctccc tgggcacag tggcctccag tctgaggatg aggtgatta
g p q t p h l - - s a a l r g p - p i l w l q v w h l s l p g h q w a p v - g - l
a p k l l i y s n n q r p s g v p d r f s g s k s g t s a s l a i s g l q s e d e a d
p p n s s i v i i s g p q g s l t d s l a p s l a p q p p w p s v g s s l r m r l i
>.....VL1-44.....>
a p k l l i y s n n q r p s g v p d r f s g s k s g t s a s l a i s g l q s e d e a d

ttactgtca gcaggagacg aataacggta aggggttta ccaggttaa acggtattg ggaagcgcg tctctgtgtt cyggcgaggg accaagctga ccgtcctagg ggcggccgca
l l c s r r r i t v r r f t r f k r v l g r r v s v f g g g t k l t v l g a a a
y y c a a g d e - r - g g l p g l n a y w e g a s l c s a e g p s - p s - g r p
i t v q q e t n n g k a v y q v - t r i g k a r l c v r r r d q a d r p r g g r
>.....VL1-44.....>
y y c a a g d e - r - g g l p g l n a y w e g a s l c s a e g p s - p s - g r p

```

SEQ ID NO: 96  
 SEQ ID NO: 97  
 SEQ ID NO: 98  
 SEQ ID NO: 99  
 SEQ ID NO: 98

Fig. 8 Cont.

201109

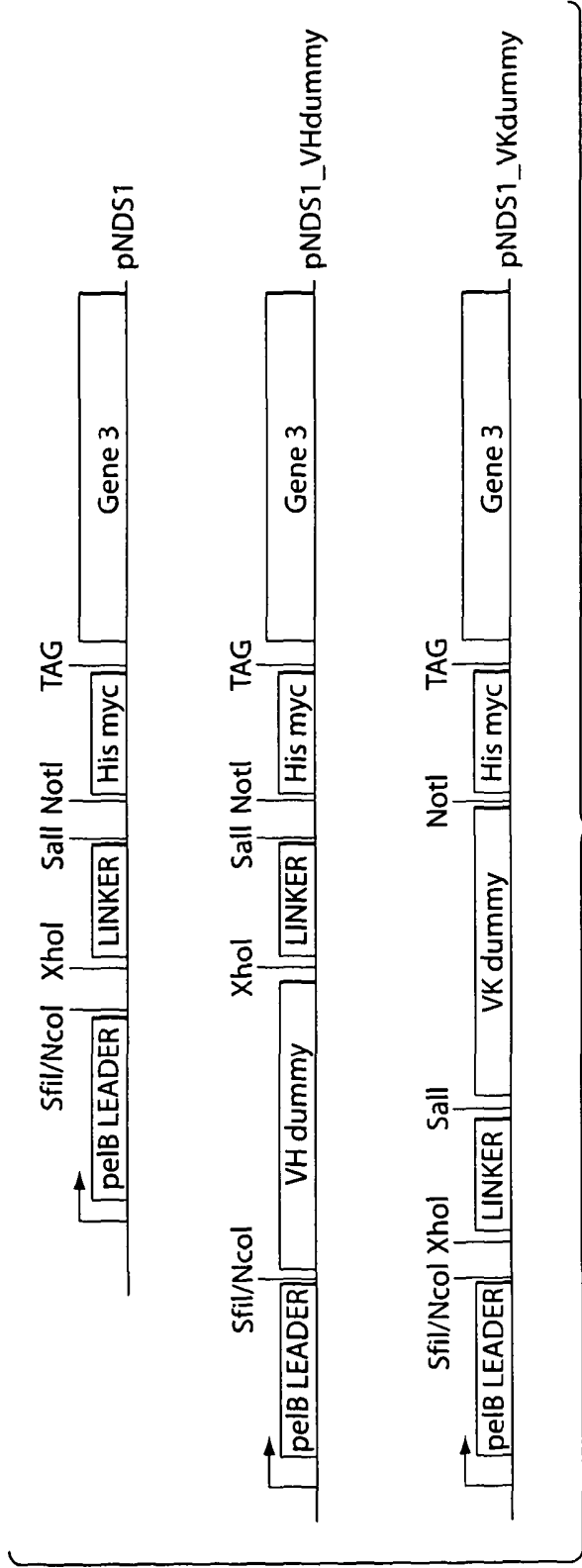
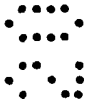
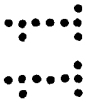
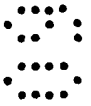


Fig. 9

SAMPLE	FW3	H3	H3/FW4	CDR3 LENGTH	
1a	CAK	AFRPNWGSRVLYF	DYW	15	SEQ ID NO: 100
1b	CAK	LNWGWFF	DPW	8	SEQ ID NO: 101
1c	CAK	GWEGEY	DYW	8	SEQ ID NO: 102
1d	CAK	DLWLDSSSNWF	DPW	13	SEQ ID NO: 103
1e	CAK	DLPG	DPHW	7	SEQ ID NO: 104
1f	CAK	APPTGYF	DYW	9	SEQ ID NO: 105
1g	CAK	VGLTGVHF	DYW	10	SEQ ID NO: 106
1h	CAK	DSYGSGSYYN	DYW	12	SEQ ID NO: 107
2a	CAK	VAGTWGRVAYYF	DYW	14	SEQ ID NO: 108
2b	CAK	VTGVFVGNF	DYW	11	SEQ ID NO: 109
2c	CAK	DARNSGSYF	DYW	11	SEQ ID NO: 110
2d	CAK	GVRTGVVF	DYW	10	SEQ ID NO: 111
2e	CAK	GVVNWGTRRKGWF	DPW	15	SEQ ID NO: 112
2f	CAK	WGGWF	DPW	7	SEQ ID NO: 113
2g	CAK	RRDNWGSV	DW	9	SEQ ID NO: 114
2h	CAK	GSGFSSGWF	DSW	11	SEQ ID NO: 115

Fig. 10





VH LENGTH FREQUENCY		CDR3																										THEORETICAL DIVERSITY							
		a b c d e f g h i j k l m n o p z 101 102 103 104 105 106																										NNS		DKV		NVT		DVT	
7	2.5	C	A	R	K	X	X	X	X	F	M	D	Y	W	G	Q	G	T	L	V	T	V	S	4	1.0E+06	1.0E+05	2.1E+04	6.6E+03							
8	4	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	5	3.4E+07	1.9E+06	2.5E+05	5.9E+04							
9	6	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	6	1.1E+09	3.4E+07	3.0E+06	5.3E+05								
10	8	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	7	3.4E+10	6.1E+08	3.6E+07	4.8E+06								
11	9	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	8	1.1E+12	1.1E+10	4.3E+08	4.3E+07								
12	11	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	9	3.5E+13	2.0E+11	5.2E+09	3.9E+08								
13	10	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	10	1.1E+15	6.4E+12	6.2E+10	3.5E+09								
14	9	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	11	3.6E+16	3.6E+13	7.4E+11	3.1E+10								
15	8	C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	12	1.2E+18	1.2E+15	8.9E+12	2.8E+11								
16		C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	13	3.7E+19	2.1E+16	1.1E+14	2.5E+12								
17		C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	14	1.2E+21	3.7E+17	1.3E+15	2.3E+13								
18		C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	15	3.8E+22	6.7E+18	1.5E+16	2.1E+14								
19		C	A	R	K	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	16	1.2E+24	1.2E+20	1.8E+17	1.9E+15								

VK LENGTH FREQUENCY	STRUCTURAL LOOP		CANONICAL RANDOMIZED		THEORETICAL DIVERSITY															
	89	90	91	92	93	94	95	96	96a	97	NNS	DKV	NVT	DVT						
4	C	Q	Q	X	X	X	X	-	-	S	F	G	Q	G	4	1.0E+06	1.0E+05	2.1E+04	6.6E+03	
5	+	C	Q	Q	X	X	X	P	-	T	F	G	Q	G	3	3.4E+07	1.9E+06	2.5E+05	5.9E+04	
6	+++	C	Q	Q	X	X	X	P	X	-	T	F	G	Q	G	1	1.1E+09	3.4E+07	3.0E+06	5.3E+05
7	+	C	Q	Q	X	X	X	P	X	X	T	F	G	Q	G	5				
6		C	Q	Q	X	X	X	P	X	X	T	F	G	Q	G	2				

VL LENGTH FREQUENCY	STRUCTURAL LOOP		CANONICAL RANDOMIZED		THEORETICAL DIVERSITY																	
	89	90	91	92	93	94	95	95a	95b	96	97	NNS	DKV	NVT	DVT							
VL1-44	6	C	A	W	Y	D	X	X	X	-	X	V	F	G	G	1	2.1E+06	2.1E+05	4.1E+04	1.3E+04		
7		C	A	W	Y	D	X	X	X	X	-	X	V	F	G	G	5	6.7E+07	3.8E+06	5.0E+05	1.2E+05	
8		C	A	W	Y	D	X	X	X	X	X	V	F	G	G	2	2.1E+09	6.8E+07	6.0E+06	1.1E+06		
VL1-51	6	C	G	T	W	Y	D	X	X	X	-	X	V	F	G	G	1	2.1E+06	2.1E+05	4.1E+04	1.3E+04	
7		C	G	T	W	Y	D	X	X	X	X	-	X	V	F	G	G	5	6.7E+07	3.8E+06	5.0E+05	1.2E+05
8		C	G	T	W	Y	D	X	X	X	X	X	V	F	G	G	2	2.1E+09	6.8E+07	6.0E+06	1.1E+06	

Fig. 11



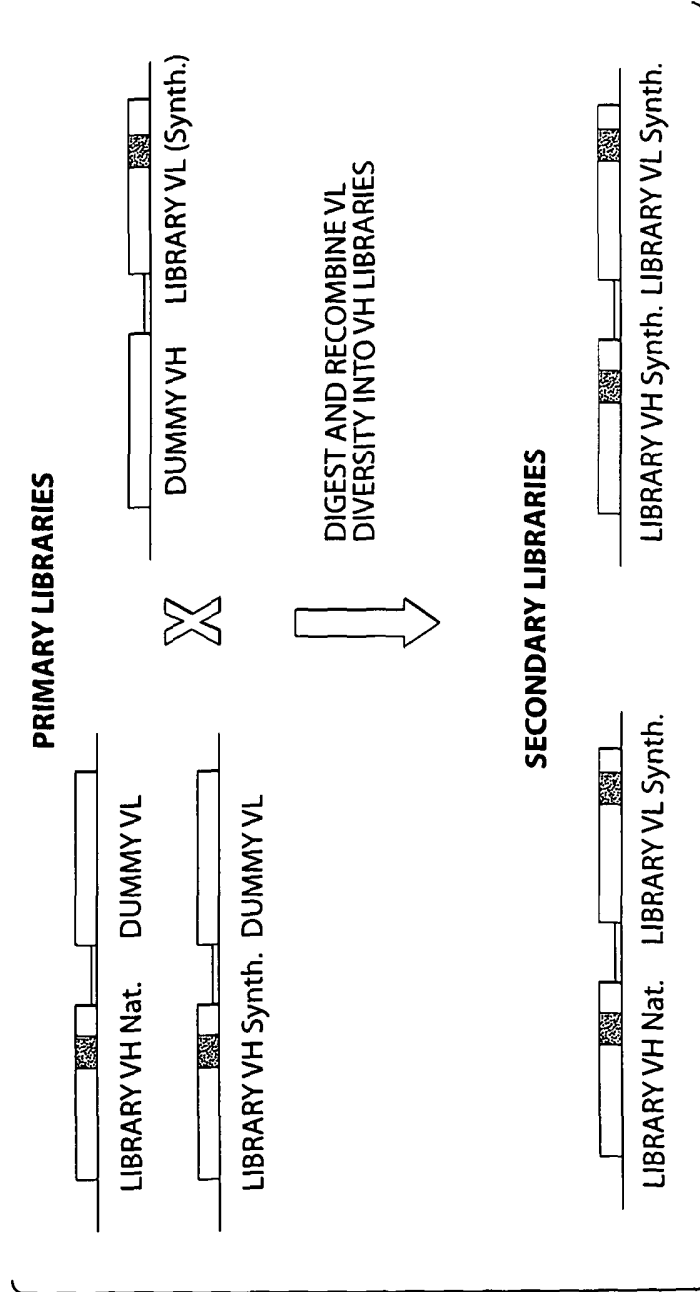


Fig. 13



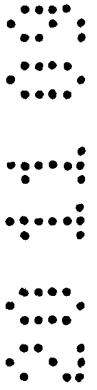
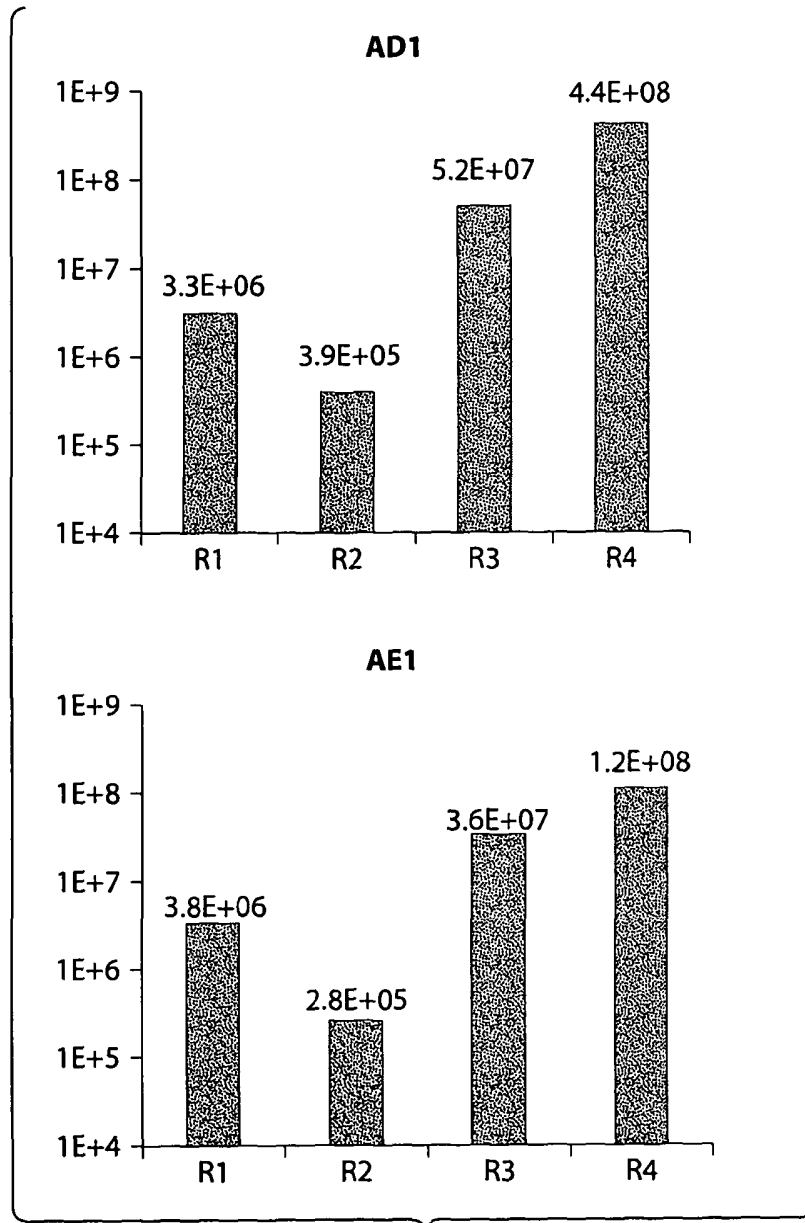


Fig. 14

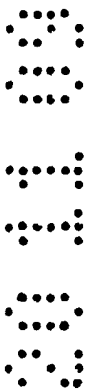
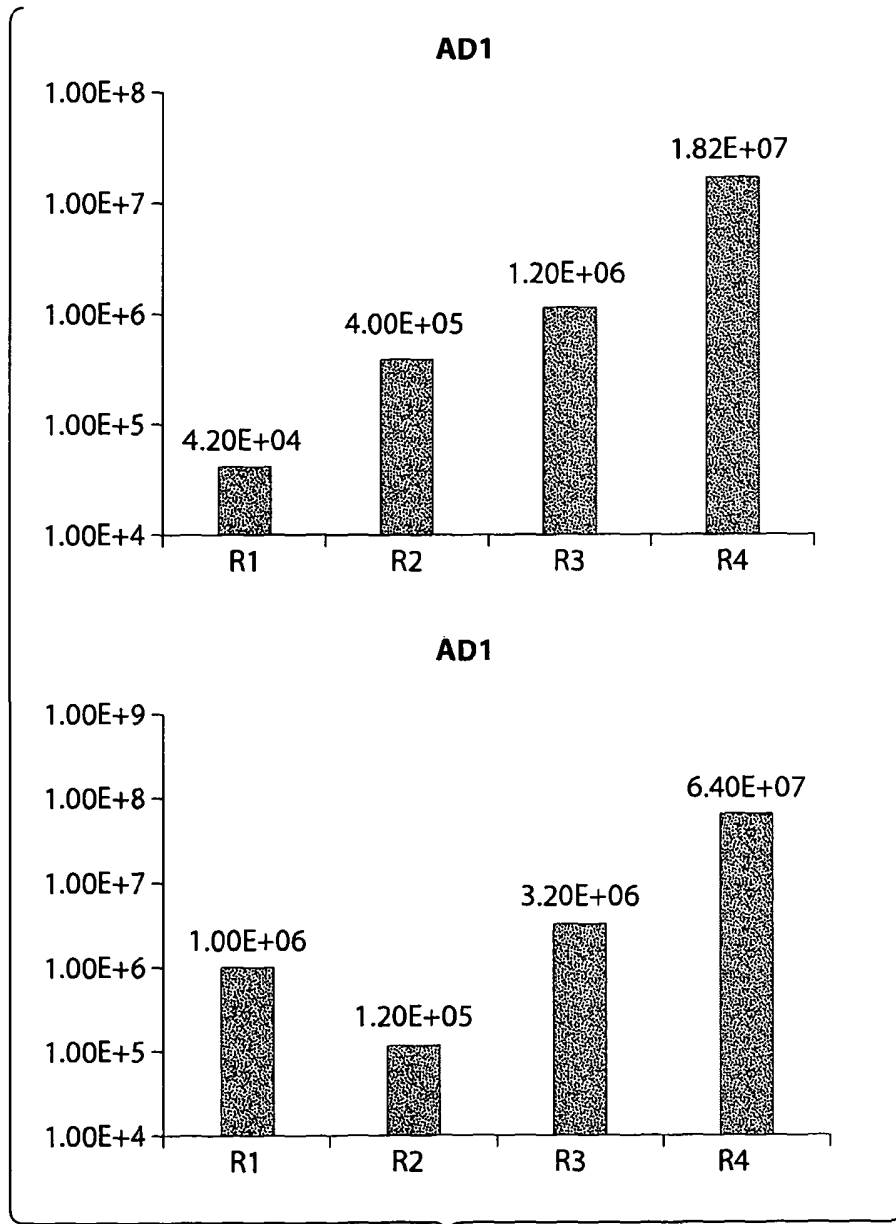


Fig. 15

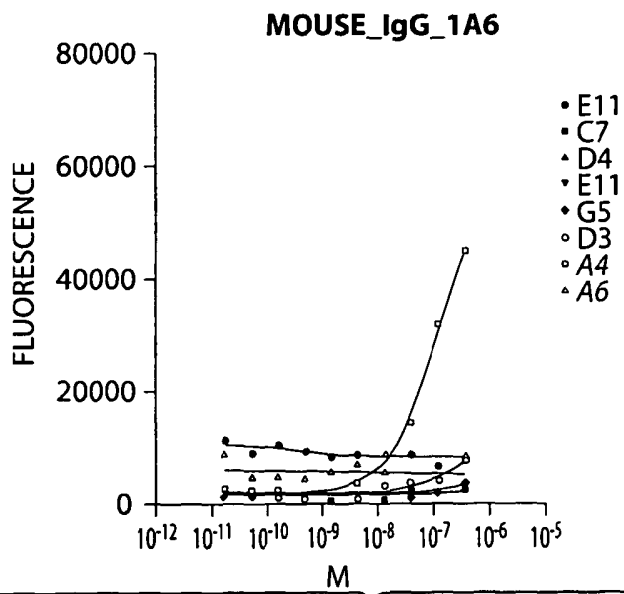
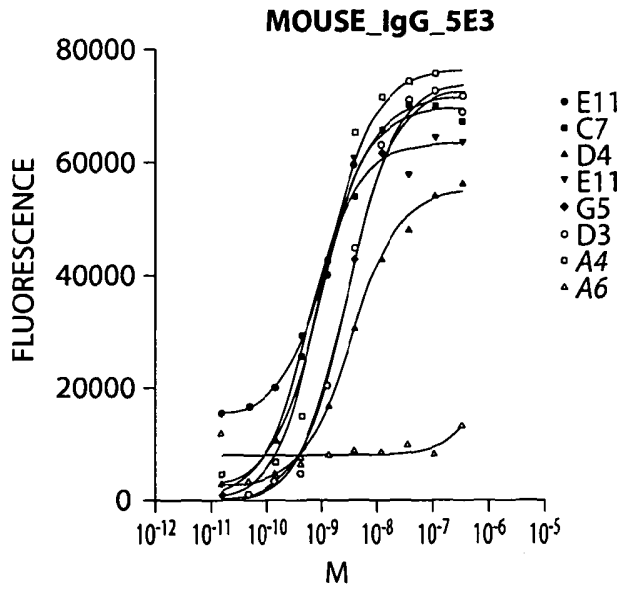
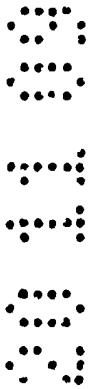


Fig. 16



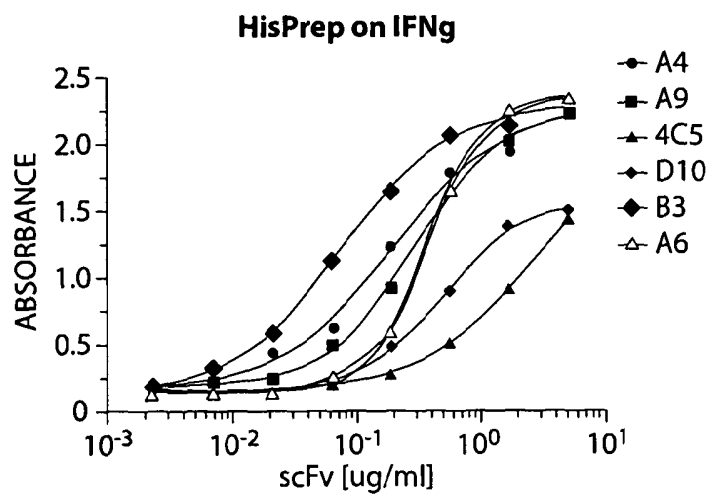
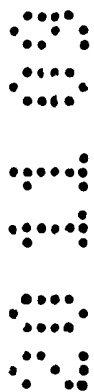


Fig. 17



## **SYNTHETIC POLYPEPTIDE LIBRARIES AND METHODS FOR GENERATING NATURALLY DIVERSIFIED POLYPEPTIDE VARIANTS**

### **Field of the Invention**

**[0001]** The invention relates to the generation of libraries of DNA sequences encoding homologous polypeptides and to the use of such libraries. This invention in particular relates to the generation of collections of synthetic antibody fragments in which one or several complementary determining regions (CDR) are replaced by a collection of the corresponding CDR captured from a natural source. The invention further relates to the diversification of a portion of a polypeptide by inserting a diversified sequence of synthetic or natural origin without the need for modification of the original polypeptide coding sequence.

### **Background of the Invention**

**[0002]** An antibody is composed of four polypeptides: two heavy chains and two light chains. The antigen binding portion of an antibody is formed by the light chain variable domain (VL) and the heavy chain variable domain (VH). At one extremity of these domains six loops form the antigen binding site and also referred to as the complementarity determining regions (CDR). Three CDRs are located on the VH domain (H1, H2 and H3) and the three others are on the VL domain (L1, L2 and L3). During B cell development a unique immunoglobulin region is formed by somatic recombination known as V(D)J recombination. The variable region of the immunoglobulin heavy or light chain is encoded by different gene segments. The heavy chain is encoded by three segments called variable (V), diversity (D) and joining (J) segments whereas the light chain variable is formed by the recombination of only two segments V and J. A large number of antibody paratopes can be generated by recombination between one of the multiple copies of the V, D and J segments that are present in the genome. The V segment encodes the CDR1 and CDR2 whereas the CDR3 is generated by the recombination events. During the course of the immune response further diversity is introduced into the antigen binding site by a process called somatic hypermutation (SHM). During this process point mutations are introduced in the variable genes of the heavy and light chains and in particular into the regions encoding the CDRs. This additional variability allows

for the selection and expansion of B cells expressing antibody variants with improved affinity for their cognate antigen.

**[0003]** In recent years several display technologies have emerged and allow for the screening of large collections of proteins or peptides. These include phage display, bacterial display, yeast display and ribosome display (Smith GP. *Science*. 1985 Jun 14;228(4705):1315-7; Hanes J and Plückthun A. *Proc Natl Acad Sci U S A*. 1997 May 13;94(10):4937-42.; Daugherty PS et al., *Protein Eng*. 1998 Sep; 11(9):825-32.; Boder ET and Wittrup KD. *Nat Biotechnol*. 1997 Jun; 15(6):553-7). In particular these methods have been applied extensively to antibodies and fragments thereof. A number of methods have been described to generate libraries of polypeptides and to screen for members with desired binding properties.

**[0004]** A first approach is to capture by gene amplification rearranged immunoglobulin genes from natural repertoires using either tissues or cells from humans or other mammals as a source of genetic diversity. These collections of rearranged heavy and light chains (VH and VL) are then combined to generate libraries of binding pairs that can be displayed on bacteriophage or on other display packages such as bacteria, yeast or mammalian cells. In this case a large fraction of the immunoglobulin repertoire found in the donor is captured. Thus all of the frameworks encoded by the donor germline genes can be found in such repertoires as well as diversity generated both by V(D)J recombination and by somatic hypermutation (Marks JD et al., *J Mol Biol*. 1991 Dec 5;222(3):581-97.; McCafferty US Patent No. 5,969,108). A limitation of natural repertoires is that naturally occurring antibodies can be based on frameworks with low intrinsic stability that limit their expression levels, shelf life and their usefulness as reagents or therapeutic molecules. In order to overcome these limitations a number of methods have been developed to generate synthetic antibody libraries. In these approaches, a unique or a limited number of selected antibody framework encoded by their corresponding germline genes are selected. The selection of these frameworks is commonly based on their biochemical stability and/or their frequency of expression in natural antibody repertoires. In order to generate a collection of binding proteins, synthetic diversity is then introduced in all or a subset of CDRs. Typically either the whole or part of the CDR is diversified using different strategies. In some cases diversity was introduced at selected positions within the CDRs (Knappik A et al., *J Mol Biol*. 2000 Feb 11;296(1):57-86). Targeted residues can be those frequently involved in antigen contact, those displaying maximal

diversity in natural antibody repertoires or even residues that would be preferentially targeted by the cellular machinery involved in generating somatic hypermutations during the natural affinity maturation process (Balint RF, Larrick JW. *Gene*. 1993 Dec 27;137(1):109-18.). Several methods have been used to diversify the antibody CDRs. Overlapping PCR using degenerate oligonucleotides have been extensively used to assemble framework and CDR elements to reconstitute antibody genes. In another approach, unique restriction enzyme sites have been engineered into the framework regions at the boundary of each CDR allowing for the introduction of diversified CDRs by restriction enzyme mediated cloning. In any case, as all the members of the library are based on frameworks with selected and preferred characteristics, it is anticipated that the antibodies derived from these repertoires are more stable and provide a better source of useful reagents. (Knappik, US 6696248; Sidhu SS, et al., *Methods Enzymol*. 2000;328:333-63; Lee CV et al., *J Mol Biol*. 2004 Jul 23;340(5):1073-93). However, an important limitation of these synthetic libraries is that a significant proportion of the library members are probably not expressed because the randomly diversified sequences do not allow for proper expression and/or folding of the protein. This problem is particularly significant for the CDR3 of the heavy chain. Indeed, this CDR often contributes to most of the binding energy to the antigen and is highly diverse in length and sequence. While the other CDR (H1, H2, L1, L2 and L3) can only adopt a limited number of three dimensional conformations, known as canonical folds, the number of conformations that can be adopted by the heavy chain CDR3 remains too diverse to be predicted (Al-Lazikani B et al., *J Mol Biol*. 1997 Nov 7;273(4):927-48). In addition, the use of long degenerate oligonucleotides used to cover long CDR H3 often introduces single base-pair deletions. These factors significantly reduce the functional size of synthetic repertoires. This serious limitations prompted many efforts to improve the process. First, in order to more effectively sample the huge diversity of possible sequence combination encoded by the CDRs, diversification strategies aiming at mimicking the amino acids usage found in Natural CDRs have been used (de Kruif J et al., *J Mol Biol*. 1995 Apr 21;248(1):97-105; Sidhu SS et al., *J Mol Biol*. 2004 Apr 23;338(2):299-310.). Another approach has been the pre-selection of synthetic repertoires by of binding the library to a generic ligand. This step allows for the enrichment of library members that are able to expressed and to fold properly (Winter and Tomlinson, US 6,696,245 B2).

**[0005]** In summary, both natural and synthetic repertoires have advantages and limitations. On one hand, strategies relying on the capture of naturally rearranged antibody variable genes are not optimal as they include potentially less favorable frameworks. A positive aspect is that these rearranged variable genes include CDRs which are compatible with proper domain folding as they have been expressed in context of a natural antibody. On the other hand, strategies based on selecting frameworks and inserting synthetic diversity benefit from the improved stability of the frameworks but are limited by the large number of CDR sequences that are not compatible with folding and/or expression and can destabilize the overall domain (Figure 1A).

**[0006]** The present invention improves over the methods described above by combining the benefits of stable framework selection and the insertion of naturally encoded CDRs that have been selected in a natural context of a functional antibody.

#### **Summary of the Invention**

**[0007]** The present invention provides methods of generating libraries of nucleic acid sequences that combine the benefits of stable framework selection and the insertion of naturally encoded complementarity determining regions (CDRs) or amino acid sequences that can fulfill the role of a CDR that have been selected in a natural context of a functional polypeptide such as an antibody. The method allows for the recovery of long CDRs or amino acid sequences that can fulfill the role of a CDR that are very difficult to encode using synthetic approaches. This invention, by combining stable frameworks and properly folded CDRs or amino acid sequences that can fulfill the role of a CDR, maximizes the proportion of functional antibodies in the library and therefore the performance of the selection process and the quality of selected clones. The present methods are also used to introduce CDRs of synthetic origin or amino acid sequences that can fulfill the role of a CDR with a higher success frequency than alternative methods. Libraries of variants generated according to this method are used for selection and screening with any described display, selection and screening technology.

**[0008]** The methods provided herein generate antibodies that contain a stable framework and correctly folded CDRs or amino acid sequences that can fulfill the role of a CDR. The methods capture the natural diversity of sequences in stable frameworks.



**[0009]** In the methods provided herein, the germline sequences for framework regions 1, 2 and 3 (FR1, FR2 and FR3) are selected from the desired organism, for example, from the human genome (see *e.g.*, Figures 2 and 6). In one embodiment of this method, selected antibody variable domains are modified by introducing a stuffer sequence that will serve as an integration site for diversified sequences. Diversity is introduced into the sequence outside of the immunoglobulin coding region by introducing restriction enzyme recognition sites, for example, Type IIs restriction sites, at a desired location such as the variable heavy chain complementarity determining region 3 (CDR H3), the variable light chain complementarity determining region 3 (CDR L3), the variable heavy chain complementarity determining region 1 (CDR H1), the variable light chain complementarity determining region 1 (CDR L1), the variable heavy chain complementarity determining region 2 (CDR H2) or the variable light chain complementarity determining region 2 (CDR L2). While the examples provided herein demonstrate diversity at the CDR3 region (in the variable heavy chain region and/or variable light chain region), it is understood that diversity can be achieved at any desired location, such as, but not limited to, the CDR1 region (in the variable heavy chain region and/or variable light chain region) or the CDR2 region (in the variable heavy chain region and/or variable light chain region). Diversified DNA sequences are generated with flanking sequences that include Type IIs restriction sites. In the methods provided herein, the cohesive ends generated by the restriction enzymes are compatible and the reading frame is maintained, thus allowing the diversified DNA fragments to be ligated into an acceptor framework.

**[0010]** The methods provided herein are also useful for generating amino acid sequences having diversified regions encoded therein. For example, in the methods provided herein, the sequences for the non-diversified portions of the encoded amino acid are selected from the desired organism, for example, from the human sequence. A portion of the encoded amino acid sequence is modified by introducing a stuffer sequence that will serve as an integration site for diversified sequences. Diversity is introduced into the sequence at the desired location(s) by introducing restriction enzyme recognition sites, for example, Type IIs restriction sites, at a desired location within the encoded amino acid sequence. Diversified DNA sequences are generated with flanking sequences that include Type IIs recognition sites. In the methods provided herein, the cohesive ends generated by the restriction enzymes are

compatible and the reading frame is maintained, thus allowing the diversified DNA fragments to be ligated into an acceptor framework.

**[0011]** In the methods provided herein, an "Acceptor Framework" is generated using a "stuffer fragment" of DNA that contain and are, preferably, bordered by two Type II restriction enzyme sites. (See *e.g.*, Figure 6). Preferably, these two Type II restriction enzyme sites digest sequences at the boundary of the site at which diversity is desired, such as, for example, the CDR H3 region, the CDR L3 region, the CDR H1 region, the CDR L1 region, the CDR H2 region or the CDR L2 region. As used herein, the term "Acceptor Framework" refers to a nucleic acid sequence that include the nucleic acid sequences encoding the FR1, FR2, FR3 and FR4 regions, the nucleic acid sequences encoding two CDRs or amino acid sequences that can fulfill the role of these CDRs, and a "stuffer fragment" that serves as the site of integration for diversified nucleic acid sequence. For example, in embodiments where diversity at the CDR3 region (in the variable heavy chain region and/or the variable light chain region) is desired, the Acceptor Framework includes the nucleic acid sequences encoding the FR1, FR2, FR3 and FR4 regions, the nucleic acid sequences encoding the CDR1 and CDR2 regions, and a "stuffer fragment" that serves as the site of integration for diversified nucleic acid sequence. For example, in embodiments where diversity at the CDR2 region (in the variable heavy chain region and/or the variable light chain region) is desired, the Acceptor Framework includes the nucleic acid sequences encoding the FR1, FR2, FR3 and FR4 regions, the nucleic acid sequences encoding the CDR1 and CDR3 regions, and a "stuffer fragment" that serves as the site of integration for diversified nucleic acid sequence. For example, in embodiments where diversity at the CDR1 region (in the variable heavy chain regions and/or the variable light chain regions) is desired, the Acceptor Framework includes the nucleic acid sequences encoding the FR1, FR2, FR3 and FR4 regions, the nucleic acid sequences encoding the CDR2 and CDR3 regions, and a "stuffer fragment" that serves as the site of integration for diversified nucleic acid sequence.

**[0012]** The terms "stuffer fragment", "stuffer DNA fragment" and "stuffer sequence" or any grammatical variation thereof are used interchangeably herein to refer to a nucleic acid sequence that includes at least two Type II recognition sites and a diversified sequence. The Acceptor Framework can be a variable heavy chain (VH) Acceptor Framework or a variable light chain (VL) Acceptor Framework. The use of the Acceptor Frameworks and the stuffer

fragments contained therein allow for the integration of a CDR sequence (natural or synthetic) or an amino acid sequence that can fulfill the role of the CDR into the acceptor framework with no donor framework nucleotides or residues contained therein or needed for integration. For example, the use of the Acceptor Frameworks and the stuffer fragments contained therein allow for the integration of a CDR sequence (natural or synthetic) selected from CDR H3, CDR L3, CDR H2, CDR L2, CDR H1 and CDR L1, or an amino acid sequence that can fulfill the role of a CDR selected from CDR H3, CDR L3, CDR H2, CDR L2, CDR H1 and CDR L1 into the acceptor framework with no donor framework nucleotides or residues contained therein or needed for integration. Thus, upon integration, the stuffer fragment is removed in full, and the coding region of the acceptor protein and the inserted proteins fragments (*i.e.*, the CDRs) are intact.

**[0013]** The methods provided herein use primers that are designed to contain cleavage sites for Type II restriction enzymes at the boundary of the site of at which diversity is desired, for example, the CDR H3 region, the CDR L3 region, the CDR H2 region, the CDR L2, the CDR H1 region or the CDR L1 region. Random, naturally occurring CDR clones (see *e.g.*, Figure 10) or synthetic CDR sequences (see *e.g.*, Example 6) or amino acid sequences that can fulfill the role of the CDR are captured in the Acceptor Frameworks used herein. For example, in embodiments where diversity at the CDR3 region (in the variable heavy chain region and/or the variable light chain region) is desired, random, naturally occurring CDR3 clones (see *e.g.*, Figure 10) or synthetic CDR3 sequences (see *e.g.*, Example 6) or amino acid sequences that can fulfill the role of a CDR3 are captured in the Acceptor Frameworks used herein. For example, in embodiments where diversity at the CDR2 region (in the variable heavy chain region and/or the variable light chain region) is desired, random, naturally occurring CDR2 clones (see *e.g.*, methods shown in Figure 10) or synthetic CDR2 sequences (see *e.g.*, methods shown in Example 6) or amino acid sequences that can fulfill the role of a CDR2 are captured in the Acceptor Frameworks used herein. For example, in embodiments where diversity at the CDR1 region (in the variable heavy chain region and/or the variable light chain region) is desired, random, naturally occurring CDR1 clones (see *e.g.*, methods shown in Figure 10) or synthetic CDR1 sequences (see *e.g.*, methods shown in Example 6) or amino acid sequences that can fulfill the role of a CDR1 are captured in the Acceptor Frameworks used herein. As an example, oligonucleotides primers specific for flanking regions of the DNA sequence encoding

the CDR H3 of immunoglobulins, *i.e.*, specific for the FR3 and FR4 of the variable region, were designed. Oligonucleotide primers specific for flanking regions of the DNA sequences encoding other regions, such as, for example, the CDR L3, CDR H1, CDR L1, CDR H2, or CDR L2, can also be designed. These oligonucleotides contain at their 5' end a site for a Type IIs restriction enzyme whereas their 3' portion matches the targeted DNA sequence.

**[0014]** In some embodiments, the primer is a nucleic acid selected from the group consisting of SEQ ID NOs: 120-254.

**[0015]** The methods provided herein use Type IIs restriction enzymes, such as, for example, FokI, to insert natural CDR sequences, such as, for example, natural CDR H3, CDR L3, CDR H1, CDR L1, CDR H2, or CDR L2 sequences into the acceptor frameworks described herein. The methods provided herein use Type IIs restriction enzymes, such as, for example, FokI, to insert synthetic CDR sequences, such as, for example, synthetic CDR H3, CDR L3, CDR H1, CDR L1, CDR H2, or CDR L2 sequences into the acceptor frameworks described herein. The methods provided herein use Type IIs restriction enzymes, such as, for example, FokI, to insert amino acid sequences that can fulfill the role of a desired CDR region, such as, for example, an amino acid sequence that can fulfill the role of a natural or synthetic CDR H3, CDR L3, CDR H1, CDR L1, CDR H2, or CDR L2 region into the acceptor frameworks described herein. The Type IIs restriction enzymes are enzymes that cleave outside of their recognition sequence to one side. These enzymes are intermediate in size, typically 400-650 amino acids in length, and they recognize sequences that are continuous and asymmetric. Suitable Type IIs restriction enzymes, also known as Type IIs restriction endonucleases, and the sequences they identify are described, for example, in Szybalski et al., "Class-IIS Restriction Enzymes - a Review." *Gene*, vol. 100: 13-26 (1991), the contents of which are hereby incorporated in their entirety by reference.

**[0016]** Primary Libraries include a VH Acceptor Framework and a fixed VL sequence (also referred to as a "dummy VL" sequence) or a VL Acceptor Framework and a fixed VH sequence (also referred to as a "dummy VH" sequence). Thus, Primary Libraries exhibit diversity in only one of the heavy or light chains. Secondary Libraries are generated by ligating a VH Acceptor Framework and a VL Acceptor Framework together (see *e.g.*, Example 7). Secondary Libraries have diversity in both the heavy and light chains.

**[0017]** The invention provides methods for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain by (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence; (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 3 (CDR3) regions or encoding amino acid sequences that can fulfill the role of a CDR3 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR3 regions or amino acid sequences that can fulfill the role of a CDR3 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); and (d) ligating the digested nucleic acid sequences encoding the CDR3 regions or the amino acid sequences that can fulfill the role of a CDR3 region of step (c) into the digested Acceptor Framework of step (c) such that the FR3 and FR4 regions are interspaced by the nucleic acid sequences encoding the CDR3 region or the amino acid sequence that can fulfill the role of a CDR3 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

**[0018]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0019]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0020]** In one embodiment, the plurality of diversified nucleic acids encode CDR3 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0021]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0022]** In one embodiment, the plurality of diversified nucleic acids encodes CDR3 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0023]** In one embodiment, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

**[0024]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0025]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0026]** In some embodiments, the methods provided include the additional step of (e) transforming the expression vector of step (d) into a host cell and culturing the host cell under conditions sufficient to express the plurality of Acceptor Framework sequences. For example, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0027]** The invention also provides methods for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain, by (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3); (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 1 (CDR1) regions or encoding amino acid sequences that can fulfill the role of a CDR1 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR1 regions or amino acid sequences that can fulfill the role of a CDR1 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); and (d) ligating the digested nucleic acid sequences encoding the CDR1 regions or the amino acid sequences that can fulfill the role of a CDR1 region of step (c) into the digested Acceptor Framework of step (c) such that the FR1 and FR2 regions are interspaced by the nucleic acid sequences encoding the CDR1 region or the amino

acid sequence that can fulfill the role of a CDR1 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

**[0028]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0029]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0030]** In one embodiment, the plurality of diversified nucleic acids encode CDR1 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0031]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0032]** In one embodiment, the plurality of diversified nucleic acids encodes CDR1 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from



immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0033]** In one embodiment, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

**[0034]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0035]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0036]** In some embodiments, the methods provided include the additional steps of (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector and (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domain encoded by the library. For example, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0037]** The invention also provides methods for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain, by (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type II restriction enzyme recognition sites interspaced by a random nucleic acid sequence, and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3); (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 2 (CDR2)

regions or encoding amino acid sequences that can fulfill the role of a CDR2 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR2 regions or amino acid sequences that can fulfill the role of a CDR2 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); and (d) ligating the digested nucleic acid sequences encoding the CDR2 regions or the amino acid sequences that can fulfill the role of a CDR2 region of step (c) into the digested Acceptor Framework of step (c) such that the FR2 and FR3 regions are interspaced by the nucleic acid sequences encoding the CDR2 region or the amino acid sequence that can fulfill the role of a CDR2 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

**[0038]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0039]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable

gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0040]** In one embodiment, the plurality of diversified nucleic acids encode CDR2 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0041]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0042]** In one embodiment, the plurality of diversified nucleic acids encode CDR2 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0043]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0044]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0045]** In some embodiments, the methods provided include the additional steps of (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector and (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domain encoded by the library. For example, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0046]** The invention also provides methods for making a target-specific antibody, antibody variable region or a portion thereof, by (a) providing a plurality of Acceptor

Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence; (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 3 (CDR3) regions or encoding amino acid sequences that can fulfill the role of a CDR3 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR3 regions or amino acid sequences that can fulfill the role of a CDR3 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); (d) cloning the digested nucleic acid sequences encoding the CDR3 regions or the amino acid sequences that can fulfill the role of a CDR3 region into an expression vector and ligating the digested nucleic acid sequences encoding the CDR3 regions or the amino acid sequences that can fulfill the role of a CDR3 region of step (c) into the Acceptor Framework such that the FR3 and FR4 regions are interspaced by the nucleic acid sequences encoding the CDR3 region or the amino acid sequence that can fulfill the role of a CDR3 region and a complete immunoglobulin variable gene encoding sequence is restored; (e) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express the plurality of Acceptor Framework sequences; (f) contacting the host cell with a target antigen; and (g) determining which expressed Acceptor Framework sequences bind to the target antigen.

**[0047]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme

recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0048]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0049]** In one embodiment, the plurality of diversified nucleic acids encode CDR3 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0050]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0051]** In one embodiment, the plurality of diversified nucleic acids encodes CDR3 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0052]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0053]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0054]** In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1. In some embodiments, the host cell is *E. coli*.

**[0055]** In some embodiments, the method includes the additional step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

**[0056]** The invention also provides methods for making a target-specific antibody, antibody variable region or a portion thereof, by (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3); (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 1 (CDR1) regions or encoding amino acid sequences that can fulfill the role of a CDR1 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR1 regions or amino acid sequences that can fulfill the role of a CDR1 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); (d) cloning the digested nucleic acid sequences encoding the CDR1 regions or the amino acid sequences that can fulfill the role of a CDR1 region into an expression vector and ligating the digested nucleic acid sequences encoding the CDR1 regions or the amino acid sequences that can fulfill the role of a CDR1 region of step (c) into the Acceptor Framework such that the FR1 and FR2 regions are interspaced by the nucleic acid sequences encoding the CDR1 region or the amino acid

sequence that can fulfill the role of a CDR1 region and a complete immunoglobulin variable gene encoding sequence is restored; (e) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express the plurality of Acceptor Framework sequences; (f) contacting the host cell with a target antigen; and (g) determining which expressed Acceptor Framework sequences bind to the target antigen.

**[0057]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites BsaI recognition sites, and/or BsmBI recognition sites.

**[0058]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0059]** In one embodiment, the plurality of diversified nucleic acids encode CDR1 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0060]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0061]** In one embodiment, the plurality of diversified nucleic acids encodes CDR1 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0062]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0063]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0064]** In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1. In some embodiments, the host cell is *E. coli*.

**[0065]** In some embodiments, the method includes the additional step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

**[0066]** The invention provides methods for making a target-specific antibody, antibody variable region or a portion thereof, by (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3); (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 2 (CDR2) regions or encoding amino acid sequences that can fulfill the role of a CDR2 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity; (c) digesting each of the plurality of nucleic acid sequences encoding the CDR2 regions or amino acid sequences that can fulfill the role of a CDR2 region using a Type IIs restriction enzyme



that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); (d) ligating the digested nucleic acid sequences encoding the CDR2 regions or the amino acid sequences that can fulfill the role of a CDR2 region of step (c) into the digested Acceptor Framework of step (c) such that the FR2 and FR3 regions are interspaced by the nucleic acid sequences encoding the CDR2 region or the amino acid sequence that can fulfill the role of a CDR2 region and complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored; (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector; (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domains encoded by the library; (g) contacting the plurality of immunoglobulin variable domains of step (f) with a target antigen; and (h) determining which expressed immunoglobulin variable domain encoding sequences bind to the target antigen.

**[0067]** In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0068]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable

gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0069]** In one embodiment, the plurality of diversified nucleic acids encode CDR2 regions, and wherein the plurality of diversified nucleic acids comprise naturally occurring sequences or sequences derived from immunized animals.

**[0070]** In one embodiment, the plurality of diversified nucleic acids includes or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0071]** In one embodiment, the plurality of diversified nucleic acids encodes CDR2 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0072]** In one embodiment, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

**[0073]** In another embodiment, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0074]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0075]** In some embodiments, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0076]** In some embodiments, the method includes the additional step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

**[0077]** The invention also provides methods for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain. These methods

include the steps of (a) providing a plurality of Ig Acceptor Framework nucleic acid sequences into which a source of diversity is introduced at a single complementarity determining region (CDR) selected from the group consisting of complementarity determining region 1 (CDR1), complementarity determining region 2 (CDR2), and complementarity determining region 3 (CDR3), wherein the Ig Acceptor Framework sequence comprises a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites, and wherein the source of diversity is a CDR selected from naturally occurring CDR sequences that contain Type IIs restriction enzyme recognition sites outside the CDR region, (b) introducing the source of diversity within each Ig Acceptor Framework by digesting both the source of diversity and the Ig Acceptor Frameworks with a Type IIs restriction enzyme; and (c) ligating the digested source of diversity into the Ig Acceptor Framework such that a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

**[0078]** The naturally occurring CDR region sequences are substantially unaltered from their wild-type, *i.e.*, natural state. These naturally occurring CDR region sequences naturally contain two Type IIs restriction enzyme recognition sites. The Type IIs restriction enzyme recognition sites are outside the CDR encoding region. The sequence of CDR regions are unaltered at the boundaries of the CDR encoding region -- the restriction enzymes recognize and splice at a region that is up to the boundary of the CDR encoding region, but does not splice within the CDR encoding region. Such Type IIs restriction enzyme recognition sites do not need to be engineered or otherwise artificially introduced into the CDR region sequence.

**[0079]** In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the naturally occurring CDR sequences are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the naturally occurring CDR sequences are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0080]** In some embodiments, the Ig Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene

sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0081]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0082]** In some embodiments, the set of naturally occurring nucleic acids encode CDR3 regions, and wherein the set of naturally occurring nucleic acids comprise immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0083]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0084]** In some embodiments, the set of naturally occurring nucleic acids encode CDR1 regions, and wherein the set of naturally occurring nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0085]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring

Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[0086]** In some embodiments, the set of naturally occurring nucleic acids encodes CDR2 regions, and wherein the set of naturally occurring nucleic acids comprises immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[0087]** In some embodiments, the plurality of Ig Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain (VL) Acceptor Framework nucleic acid sequence.

**[0088]** In some embodiments, the methods provided include the additional steps of (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector and (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domain encoded by the library. For example, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0089]** The invention also provides methods for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain. These methods include the steps of (a) providing a plurality of Ig Acceptor Framework nucleic acid sequences into which a source of diversity is introduced at a single complementarity determining region (CDR) selected from the group consisting of complementarity determining region 1 (CDR1), complementarity determining region 2 (CDR2), and complementarity determining region 3 (CDR3), where the Ig Acceptor Framework sequence comprises a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites, and wherein the source of diversity is a CDR selected from synthetically produced CDR sequences that contain Type IIs restriction enzyme recognition sites outside the CDR region, (b) introducing the source of diversity within each Ig Acceptor Framework by digesting both the source of diversity and the Ig Acceptor Framework with a Type IIs restriction enzyme; and (c) ligating the digested source

of diversity into the Ig Acceptor Framework such that a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

**[0090]** In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the synthetically produced CDR sequences are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the synthetically produced CDR sequences are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[0091]** In some embodiments, the Ig Acceptor Framework nucleic acid sequence is derived from a human sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[0092]** In some embodiments, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0093]** In some embodiments, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0094]** In some embodiments, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[0095]** In some embodiments, the plurality of Ig Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[0096]** In some embodiments, the methods provided include the additional steps of (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector and (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domain encoded by the library. For example, the host cell is *E. coli*. In some embodiments, the expression vector is a phagemid vector. For example, the phagemid vector is pNDS1.

**[0097]** The invention also provides methods for making an immunoglobulin polypeptide. These methods include the steps of (a) providing a plurality of Ig Acceptor Framework nucleic acid sequences into which a source of diversity is introduced at a single complementarity determining region (CDR) selected from the group consisting of complementarity determining region 1 (CDR1), complementarity determining region 2 (CDR2), and complementarity determining region 3 (CDR3), wherein the Ig Acceptor Framework sequence comprises a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites, and wherein the source of diversity is a CDR selected from naturally occurring CDR sequences that contain Type IIs restriction enzyme recognition sites outside the CDR region, (b) introducing the source of diversity within each Ig Acceptor Framework by digesting both the source of diversity and the Ig Acceptor Frameworks with a Type IIs restriction enzyme; (c) ligating the digested source of diversity into the Ig Acceptor Framework such that a complete immunoglobulin variable gene encoding sequence is restored; and (d) cloning the complete immunoglobulin variable gene encoding sequence from step (c) into an expression vector; and (e) transforming the expression vector of step (d) into a host cell and culturing the host cell under conditions sufficient to express the complete immunoglobulin gene encoding sequences that do not contain the Type IIs restriction enzyme recognition sites are restored.

**[0098]** In these embodiments, the naturally occurring CDR region sequences are substantially unaltered from their wild-type, *i.e.*, natural state. These naturally occurring CDR

region sequences naturally contain two Type IIs restriction enzyme recognition sites. Such Type IIs restriction enzyme recognition sites do not need to be engineered or otherwise artificially introduced into the CDR region sequence.

**[0099]** In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the naturally occurring CDR sequences are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the naturally occurring CDR sequences are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[00100]** In some embodiments, the Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[00101]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[00102]** In some embodiments, the set of naturally occurring nucleic acids encode CDR3 regions, and wherein the set of naturally occurring nucleic acids comprise immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or



sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[00103]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[00104]** In some embodiments, the set of naturally occurring nucleic acids encode CDR1 regions, and wherein the set of naturally occurring nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[00105]** In some embodiments, the set of naturally occurring nucleic acids includes or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

**[00106]** In some embodiments, the set of naturally occurring nucleic acids encodes CDR2 regions, and wherein the set of naturally occurring nucleic acids comprises immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

**[00107]** In some embodiments, the plurality of Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[00108]** In some embodiments, the expression vector is a phagemid vector. In some embodiments, the host cell is *E. coli*.

**[00109]** In some embodiments, the method also includes the steps of contacting the host cell with a target antigen, and determining which expressed complete Ig variable gene encoding sequences bind to the target antigen, thereby identifying target specific antibodies, antibody

variable regions or portions thereof. In some embodiments, the method includes the additional step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

**[00110]** The invention also provides methods for making an immunoglobulin polypeptide. These methods include the steps of (a) providing a plurality of Ig Acceptor Framework nucleic acid sequences into which a source of diversity is introduced at a single complementarity determining region (CDR) selected from the group consisting of complementarity determining region 1 (CDR1), complementarity determining region 2 (CDR2), and complementarity determining region 3 (CDR3), wherein the Ig Acceptor Framework sequence comprises a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites, and wherein the source of diversity is a CDR selected from synthetically produced CDR sequences that contain Type IIs restriction enzyme recognition sites outside the CDR region, (b) introducing the source of diversity within each Ig Acceptor Framework by digesting both the source of diversity and the Ig Acceptor Framework with a Type IIs restriction enzyme; (c) ligating the digested source of diversity into the Ig Acceptor Framework such that a complete immunoglobulin variable gene encoding sequence is restored; (d) cloning the ligated Ig Acceptor Framework from step (c) into an expression vector; and (e) transforming the expression vector of step (d) into a host cell and culturing the host cell under conditions sufficient to express the complete immunoglobulin gene encoding sequences that do not contain the Type IIs restriction enzyme recognition sites are restored.

**[00111]** In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the synthetically produced CDR sequences are recognized by the same Type IIs restriction enzyme. In some embodiments, the Type IIs restriction enzyme recognition sites within the stuffer nucleic acid sequences and the synthetically produced CDR sequences are recognized by different Type IIs restriction enzymes. For example, the Type IIs restriction enzyme recognition sites are FokI recognition sites, BsaI recognition sites, and/or BsmBI recognition sites.

**[00112]** In some embodiments, the Ig Acceptor Framework nucleic acid sequence is derived from a human gene sequence. For example, the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence. In some embodiments, the human heavy chain variable gene sequence is selected

from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51. In some embodiments, the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence. For example, the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20. In some embodiments, the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence. For example, the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.

**[00113]** In some embodiments, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[00114]** In some embodiments, the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[00115]** In some embodiments, the plurality of diversified nucleic acids encode amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprise synthetic sequences.

**[00116]** In some embodiments, the plurality of Ig Acceptor Framework nucleic acid sequences comprise a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

**[00117]** In some embodiments, the expression vector is a phagemid vector. In some embodiments, the host cell is *E. coli*.

**[00118]** In some embodiments, the method also includes the steps of contacting the host cell with a target antigen, and determining which expressed complete Ig variable gene encoding sequences bind to the target antigen, thereby identifying target specific antibodies, antibody variable regions or portions thereof. In some embodiments, the method includes the additional step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

### **Brief Description of the Drawings**

**[00119]** Figure 1A is a schematic representation of a protein domain with a framework and loops providing contact residues with another protein or molecule. Several situations are depicted: A stable protein domain with properly folded loop regions; properly folded loops inserted into a domain of limited intrinsic stability; an intrinsically stable protein domain which stability is affected by the loop regions.

**[00120]** Figure 1B is a schematic representation of different types of libraries of protein repertoires generated using different diversification strategies.

**[00121]** Figure 2 is a schematic representation of an antibody variable Acceptor Framework. Framework regions, CDRs and type IIS-RM restriction site are indicated.

**[00122]** Figure 3 is a schematic representation of a strategy used for capturing CDRH3 sequences from natural repertoires.

**[00123]** Figure 4 is a schematic representation of the benefit of using primers containing Type IIS-RM restriction enzymes for the amplification and insertion of natural CDR regions into Acceptor Frameworks.

**[00124]** Figure 5 is an illustration depicting the germline gene sequences of the variable heavy and light chain domain selected for the generation of Acceptor Frameworks.

**[00125]** Figure 6 is a schematic representation of an amplification strategy used for the generation of Acceptor Frameworks by addition to the germline sequences of a stuffer fragment and a FR4 region.

**[00126]** Figure 7, top panel, is an illustration depicting the sequence detail of Stuffer fragments of VH acceptor Framework. DNA sequences recognized and cleaved by the restriction enzyme BsmBI are boxed in red and black respectively and indicated in the lower panel of the figure. The reading frame corresponding to the antibody variable sequence is underlined.

**[00127]** Figure 8 is an illustration depicting the sequences of the 20 Acceptor Frameworks.

**[00128]** Figure 9 is a schematic representation of the pNDS1 vector alone or combined with a dummy heavy chain variable region or a dummy light variable region.

[00129] Figure 10 is a table depicting the sequences of CDRH3 sequences that were retrieved from a human cDNA source and inserted into human Acceptor Frameworks.

[00130] Figure 11 is a table representing the design of synthetic CDR sequences for VH, VK and Vλ. The positions are numbered according to the Kabat numbering scheme. The theoretical diversity of the CDR using a defined codon diversification strategy (NNS, DVK, NVT, DVT) is indicated. The strategies adopted for VH CDR synthesis are boxed.

[00131] Figure 12 is a schematic representation and sequence detail of synthetic CDR insertion into an Acceptor Framework.

[00132] Figure 13 is a schematic representation of Primary libraries and the chain recombination performed to generate Secondary libraries.

[00133] Figure 14 is a series of graphs depicting phage output titration during selection against hINFγ with the secondary libraries AD1 and AE1.

[00134] Figure 15 is a series of graphs depicting phage output titration during selection against monoclonal antibody 5E3 with the secondary libraries AD1 and AE1.

[00135] Figure 16 depicts dose response ELISA using purified 7 scFv preparations against mouse 5E3 or an irrelevant mouse antibody 1A6. The seven clones encode different scFvs. Clone A6 is a scFv specific for hINFγ and was used as a negative control.

[00136] Figure 17 depicts dose response ELISA using purified scFv preparations against mouse hINFγ and compared to a positive scFv specific for hINFγ (A6).

### **Detailed Description of the Invention**

[00137] Synthetic protein libraries and in particular synthetic antibody libraries are attractive as it is possible during the library generation process to select the building blocks composing these synthetic proteins and include desired characteristics. An important limitation, however, is that the randomization of portions of these synthetic proteins to generate a collection of variants often leads to non-functional proteins and thus can dramatically decrease the functional library size and its performance. Another limitation of synthetic diversity is that the library size needed to cover the theoretical diversity of randomized amino acid stretches cannot be covered because of practical limitations. Even with display systems such as ribosome display a diversity of  $10^{13}$  to  $10^{14}$  can be generated and sampled which can maximally cover the complete randomization of stretches of 9 amino acids. As the average size

of natural CDR H3 (also referred to herein as the heavy chain CDR3 or VH CDR3) is above 9 and can be over 20 amino acids in length, synthetic diversity is not a practicable approach to generate such CDRs.

**[00138]** The key advantage of the invention is that it combines selected acceptor antibody variable frameworks with CDR loops that have a high probability of correct folding. It allows for the capture of long CDRs that are difficult to cover with synthetic randomization approaches. Furthermore the methods described does not employ any modification within the coding region of acceptor antibody variable for cloning of the diversified sequences. Another advantage of this method is that several sources of diversity can be captured into the same set of acceptor antibody frameworks. These sources include but are not limited to: natural antibody CDRs of human or other mammal origin, CDR from chicken antibodies, CDRs of antibody-like molecules such as VHH from camelids, IgNARs from sharks, variable loops from T cell receptors. In addition, natural CDRs can be derived from naïve or immunized animals. In the latter case, the CDRs retrieved are enriched in sequences that were involved in recognition of the antigen used for immunization.

**[00139]** In this method selected protein domains, as exemplified by antibody variable domains, are modified by introducing a stuffer sequence that will serve as an integration site for diversified sequences. Upon integration, the stuffer fragment is removed in full, thus leaving intact the coding region of the acceptor protein and the inserted proteins fragments (*i.e.*, the CDRs). This integration event is mediated by a the use of Type IIs restriction enzyme that recognizes a defined site in the DNA sequence but cleave the DNA at a defined distance from this site. This approach has two major advantages: (1) it allows for the digestion of acceptors framework without affecting their coding sequences (no need to engineer silent restrictions sites); and (2) it allows for the digestion and cloning of naturally diversified sequences that by definition do not possess compatible restriction sites.

**[00140]** As described above, prior attempts to generate libraries and/or displays of antibody sequences differ from the methods provided herein. For example, some methods require the grafting of each CDR, as described for example by U.S. Patent No. 6,300,064, in which restriction enzyme sites are engineered at the boundary of each CDR, not just the CDR H3 region. In other methods, CDR sequences from natural sources are amplified and rearranged, as described in, *e.g.*, U.S. Patent No. 6,989,250. In some methods, such as those

described in US Patent Application Publication No. 20060134098, sequences from a mouse (or other mammal) is added to a human framework, such that the resulting antibody has CDR1 and CDR2 regions of murine origin and a CDR3 region of human origin. Other methods, such as those described in US Patent Application Publication No. 20030232333, generate antibodies that have synthetic CDR1 and/or CDR1/CDR2 regions along with a natural CDR3 region. However, these methods fail to provide libraries that contain stable framework regions and correctly folded CDRs.

#### Design of the antibody acceptor frameworks for diversity cloning.

**[00141]** A strategy was designed to introduce diversity into the CDR3 of selected human antibody domains that avoids the modification of the sequence of the original framework. The strategy relies on the introduction outside of the immunoglobulin coding region of Type IIs restriction sites. This class of restriction enzymes recognizes asymmetric and uninterrupted sequence of 4-7 base pairs but cleave DNA at a defined distance of up to 20 bases independently of the DNA sequence found at the cleavage site. In order to take advantage of this system for cloning of diversified sequences into selected frameworks, acceptor frameworks containing a stuffer DNA fragment, instead of the CDR3, that includes two Type IIs restriction sites were designed. Similarly, diversified DNA sequences are generated with flanking sequences that include Type IIs. Provided that the cohesive ends generated by the restriction enzymes are compatible and that reading frame is maintained, the DNA fragments can be ligated into the acceptor framework and restore the encoded CDR3 in the new context of the acceptor antibody framework (Figure 2).

#### Capture of natural CDR diversity

**[00142]** The strategy that was developed to capture naturally diversified protein fragments as a source of diversity also takes advantage of Type IIs restriction enzymes. As an example, oligonucleotide primers specific for flanking regions of the DNA sequence encoding the CDR H3 of immunoglobulins, *i.e.*, specific for the FR3 and FR4 of the variable region, were designed. These oligonucleotides contain at their 5' end a site for a Type IIs restriction enzyme whereas their 3' portion matches the targeted DNA sequence. The restriction enzyme site used is preferably an enzyme that cleaves DNA far away from the DNA recognition site

such as FokI. This is a key element of the method as it allows for the efficient amplification of natural DNA sequences as it maintains a good match between the 3' end of the primer and the DNA flanking the CDR H3 while allowing for excision of the CDRH3 coding sequence by DNA cleavage at the boundary between the CDR and framework regions (Figure 3). This precise excision of the CDR coding sequence is very difficult using Type II enzymes that cleave DNA at their recognition site as the corresponding restriction site is not present in the natural DNA sequences and that introduction of such sites during amplification would be difficult due poor primer annealing. Thus this method allows for the amplification of diversified protein sequences and their insertion into any the acceptor antibody framework regardless of origin of amplified diversity (Figure 4).

**[00143]** Unless otherwise defined, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures utilized in connection with, and techniques of, cell and tissue culture, molecular biology, and protein and oligo- or polynucleotide chemistry and hybridization described herein are those well known and commonly used in the art. Standard techniques are used for recombinant DNA, oligonucleotide synthesis, and tissue culture and transformation (*e.g.*, electroporation, lipofection). Enzymatic reactions and purification techniques are performed according to manufacturer's specifications or as commonly accomplished in the art or as described herein. The foregoing techniques and procedures are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification. *See e.g.*, Sambrook *et al.* *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). The nomenclatures utilized in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

**[00144]** As utilized in accordance with the present disclosure, the following terms, unless otherwise indicated, shall be understood to have the following meanings:



**[00145]** As used herein, the term "antibody" refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. By "specifically bind" or "immunoreacts with" or "immunospecifically bind" is meant that the antibody reacts with one or more antigenic determinants of the desired antigen and does not react with other polypeptides or binds at much lower affinity ( $K_d > 10^{-6}$ ). Antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, dAb (domain antibody), single chain,  $F_{ab}$ ,  $F_{ab}'$  and  $F_{(ab)2}$  fragments, scFvs, and an  $F_{ab}$  expression library.

**[00146]** The basic antibody structural unit is known to comprise a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The amino-terminal portion of each chain includes a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The carboxy-terminal portion of each chain defines a constant region primarily responsible for effector function. In general, antibody molecules obtained from humans relate to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG<sub>1</sub>, IgG<sub>2</sub>, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain.

**[00147]** The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

**[00148]** The term "antigen-binding site," or "binding portion" refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains, referred to as "hypervariable regions," are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus, the term "FR" refers to amino acid

sequences which are naturally found between, and adjacent to, hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to as "complementarity-determining regions," or "CDRs." The assignment of amino acids to each domain is in accordance with the definitions of Kabat Sequences of Proteins of Immunological Interest (National Institutes of Health, Bethesda, Md. (1987 and 1991)), or Chothia & Lesk J. Mol. Biol. 196:901-917 (1987), Chothia *et al.* Nature 342:878-883 (1989).

**[00149]** As used herein, the term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin, an scFv, or a T-cell receptor. The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. For example, antibodies may be raised against N-terminal or C-terminal peptides of a polypeptide. An antibody is said to specifically bind an antigen when the dissociation constant is  $\leq 1 \mu\text{M}$ ; *e.g.*,  $\leq 100 \text{ nM}$ , preferably  $\leq 10 \text{ nM}$  and more preferably  $\leq 1 \text{ nM}$ .

**[00150]** As used herein, the terms "immunological binding," and "immunological binding properties" refer to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant ( $K_d$ ) of the interaction, wherein a smaller  $K_d$  represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" ( $K_{on}$ ) and the "off rate constant" ( $K_{off}$ ) can be determined by calculation of the concentrations and the actual rates of association and dissociation. (*See* Nature 361:186-87 (1993)). The ratio of  $K_{off}$

$1/K_{on}$  enables the cancellation of all parameters not related to affinity, and is equal to the dissociation constant  $K_d$ . (See, generally, Davies et al. (1990) Annual Rev Biochem 59:439-473). An antibody of the present invention is said to specifically bind to its target, when the equilibrium binding constant ( $K_d$ ) is  $\leq 1 \mu\text{M}$ , *e.g.*,  $\leq 100 \text{ nM}$ , preferably  $\leq 10 \text{ nM}$ , and more preferably  $\leq 1 \text{ nM}$ , as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

**[00151]** The term "isolated polynucleotide" as used herein shall mean a polynucleotide of genomic, cDNA, or synthetic origin or some combination thereof, which by virtue of its origin the "isolated polynucleotide" (1) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (2) is operably linked to a polynucleotide which it is not linked to in nature, or (3) does not occur in nature as part of a larger sequence. Polynucleotides in accordance with the invention include the nucleic acid molecules encoding the heavy chain immunoglobulin molecules, and nucleic acid molecules encoding the light chain immunoglobulin molecules described herein.

**[00152]** The term "isolated protein" referred to herein means a protein of cDNA, recombinant RNA, or synthetic origin or some combination thereof, which by virtue of its origin, or source of derivation, the "isolated protein" (1) is not associated with proteins found in nature, (2) is free of other proteins from the same source, *e.g.*, free of marine proteins, (3) is expressed by a cell from a different species, or (4) does not occur in nature.

**[00153]** The term "polypeptide" is used herein as a generic term to refer to native protein, fragments, or analogs of a polypeptide sequence. Hence, native protein fragments, and analogs are species of the polypeptide genus. Polypeptides in accordance with the invention comprise the heavy chain immunoglobulin molecules, and the light chain immunoglobulin molecules described herein, as well as antibody molecules formed by combinations comprising the heavy chain immunoglobulin molecules with light chain immunoglobulin molecules, such as kappa light chain immunoglobulin molecules, and vice versa, as well as fragments and analogs thereof.

**[00154]** The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in

nature and which has not been intentionally modified by man in the laboratory or otherwise is naturally-occurring.

**[00155]** The term "operably linked" as used herein refers to positions of components so described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

**[00156]** The term "control sequence" as used herein refers to polynucleotide sequences which are necessary to effect the expression and processing of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence in eukaryotes, generally, such control sequences include promoters and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression and processing, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences. The term "polynucleotide" as referred to herein means a polymeric boron of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA.

**[00157]** As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See Immunology - A Synthesis* (2nd Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland Mass. (1991)). Stereoisomers (*e.g.*, D- amino acids) of the twenty conventional amino acids, unnatural amino acids such as  $\alpha$ -,  $\alpha$ -disubstituted amino acids, N-alkyl amino acids, lactic acid, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4 hydroxyproline,  $\gamma$ -carboxyglutamate,  $\epsilon$ -N,N,N-trimethyllysine,  $\epsilon$ -N-acetyllysine, O-phosphoserine, N- acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine,  $\sigma$ -N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4- hydroxyproline). In the polypeptide notation used herein, the left-hand direction is the amino terminal direction and the right-hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

**[00158]** As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity, and most preferably at least 99 percent sequence identity.

**[00159]** Preferably, residue positions which are not identical differ by conservative amino acid substitutions.

**[00160]** Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine valine, glutamic-aspartic, and asparagine-glutamine.

**[00161]** As discussed herein, minor variations in the amino acid sequences of antibodies or immunoglobulin molecules are contemplated as being encompassed by the present invention, providing that the variations in the amino acid sequence maintain at least 75%, more preferably at least 80%, 90%, 95%, and most preferably 99%. In particular, conservative amino acid replacements are contemplated. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into families: (1) acidic amino acids are aspartate, glutamate; (2) basic amino acids are lysine, arginine, histidine; (3) non-polar amino acids are alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan, and (4) uncharged polar amino acids are glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. The hydrophilic amino acids include arginine, asparagine, aspartate, glutamine, glutamate, histidine, lysine, serine, and threonine. The hydrophobic amino acids include alanine, cysteine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, tyrosine and valine. Other families of amino acids include (i) serine and threonine, which are the aliphatic-hydroxy

family; (ii) asparagine and glutamine, which are the amide containing family; (iii) alanine, valine, leucine and isoleucine, which are the aliphatic family; and (iv) phenylalanine, tryptophan, and tyrosine, which are the aromatic family. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid will not have a major effect on the binding or properties of the resulting molecule, especially if the replacement does not involve an amino acid within a framework site. Whether an amino acid change results in a functional peptide can readily be determined by assaying the specific activity of the polypeptide derivative. Assays are described in detail herein. Fragments or analogs of antibodies or immunoglobulin molecules can be readily prepared by those of ordinary skill in the art. Preferred amino- and carboxy-termini of fragments or analogs occur near boundaries of functional domains. Structural and functional domains can be identified by comparison of the nucleotide and/or amino acid sequence data to public or proprietary sequence databases. Preferably, computerized comparison methods are used to identify sequence motifs or predicted protein conformation domains that occur in other proteins of known structure and/or function. Methods to identify protein sequences that fold into a known three-dimensional structure are known. Bowie *et al.* Science 253:164 (1991). Thus, the foregoing examples demonstrate that those of skill in the art can recognize sequence motifs and structural conformations that may be used to define structural and functional domains in accordance with the invention.

**[00162]** Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinities, and (4) confer or modify other physicochemical or functional properties of such analogs. Analogs can include various muteins of a sequence other than the naturally-occurring peptide sequence. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to break a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized

polypeptide secondary and tertiary structures are described in *Proteins, Structures and Molecular Principles* (Creighton, Ed., W. H. Freeman and Company, New York (1984)); *Introduction to Protein Structure* (C. Branden and J. Tooze, eds., Garland Publishing, New York, N.Y. (1991)); and Thornton et al. *Nature* 354:105 (1991).

**[00163]** As used herein, the terms "label" or "labeled" refers to incorporation of a detectable marker, *e.g.*, by incorporation of a radiolabeled amino acid or attachment to a polypeptide of biotinyl moieties that can be detected by marked avidin (*e.g.*, streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods). In certain situations, the label or marker can also be therapeutic. Various methods of labeling polypeptides and glycoproteins are known in the art and may be used. Examples of labels for polypeptides include, but are not limited to, the following: radioisotopes or radionuclides (*e.g.*,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{15}\text{N}$ ,  $^{35}\text{S}$ ,  $^{90}\text{Y}$ ,  $^{99}\text{Tc}$ ,  $^{111}\text{In}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ), fluorescent labels (*e.g.*, FITC, rhodamine, lanthanide phosphors), enzymatic labels (*e.g.*, horseradish peroxidase, p-galactosidase, luciferase, alkaline phosphatase), chemiluminescent, biotinyl groups, predetermined polypeptide epitopes recognized by a secondary reporter (*e.g.*, leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance. The term "pharmaceutical agent or drug" as used herein refers to a chemical compound or composition capable of inducing a desired therapeutic effect when properly administered to a patient.

**[00164]** Other chemistry terms herein are used according to conventional usage in the art, as exemplified by *The McGraw-Hill Dictionary of Chemical Terms* (Parker, S., Ed., McGraw-Hill, San Francisco (1985)).

**[00165]** As used herein, "substantially pure" means an object species is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present.

**[00166]** Generally, a substantially pure composition will comprise more than about 80 percent of all macromolecular species present in the composition, more preferably more than about 85%, 90%, 95%, and 99%. Most preferably, the object species is purified to essential

homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species.

[00167] The term patient includes human and veterinary subjects.

[00168] Antibodies are purified by well-known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

[00169] The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

## EXAMPLES

### Example 1: Cloning of immunoglobulin variable germline genes

[00170] Seven human heavy chain variable germline genes (VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, VH5-51), five human kappa light chain variable germline genes (VK1-33, VK1-39, VK3-11, VK3-15, VK3-20) and two human lambda light chain variable germline genes (VL1-44, VL1-51) were selected to construct the libraries (Lefranc, M.-P. et al., 1999 Nucleic Acids Research, 27, 209-212). These genes were selected because they are often used in human expressed antibody repertoires and the frameworks they encode show favorable stability and expression profiles as individual domains or in the context of a VH/VL pair (Ewert S et al., J Mol Biol. 2003 Jan 17;325(3):531-53). Two sets of specific primers were used to amplify these genes from human genomic DNA by nested PCR. This approach was necessary as the 5' sequences of germline genes of the same family are identical or very similar. For each gene, a first pair of primers, called genomic locators, was designed to be specific to the 5' and 3' untranslated regions flanking the germline gene. The second pair was designed to be specific for the beginning of the framework 1 region (FR1) and the end of the FR2. The 14 independent PCR products were cloned into pGEMT-easy (Promega, Madison



WI) and their identity and integrity were verified by sequencing. The amino acid sequence of the selected germline genes is shown in Figure 5.

[00171] The primers and primer combination used are indicated below.

#### Genomic locators

5 K1-33 TGTTTCTAATCGCAGGTGCCAGATG (SEQ ID NO: 120)  
3 K1-33 ATTTATGTTATGACTTGTACTACTG (SEQ ID NO: 121)  
5 K1-39 TATTTGTTTTTATGTTTCCAATCTC (SEQ ID NO: 122)  
3 K1-39 CCTTGGAGGTTTATGTTATGACTTG (SEQ ID NO: 123)  
5 K3-11 TTATTTCCAATTTTCAGATACCACCG (SEQ ID NO: 124)  
3 K3-11 TTGTTGGGGTTTTTGTTCATGTGG (SEQ ID NO: 125)  
5 K3-15 TATTTCCAATTTTCAGATACCCTGG (SEQ ID NO: 126)  
3 K3-15 ATGTTGAATCACTGTGGGAGGCCAG (SEQ ID NO: 127)  
5 K3-20 TTATTTCCAATCTCAGATACCACCG (SEQ ID NO: 128)  
3 K3-20 TTTTGTTCAGCTGAATCACTGTG (SEQ ID NO: 129)  
5 L1-44 ATGCTGTGTCTCTCTCACTTCCAG (SEQ ID NO: 130)  
3 L1-44 TTCCCATTTGGCCTGGAGCACTGTG (SEQ ID NO: 131)  
5 L1-51 GTGTCTGTGTCTCTCCTGCTTCCAG (SEQ ID NO: 132)  
3 L1-51 CTTGTCTCAGTTCCCATTTGGGCTG (SEQ ID NO: 133)  
5 H1-2 ATCTCATCCACTTCTGTGTTCTCTC (SEQ ID NO: 134)  
3 H1-2 TTGGGTTTCTGACACCTCAGGATG (SEQ ID NO: 135)  
5 H1-18 CAGGCCAGTCATGTGAGACTTCACC (SEQ ID NO: 136)  
3 H1-18 CTGCCTCCTCCCTGGGGTTTCTGAA (SEQ ID NO: 137)  
5 H1-69 CCCCTGTGTCTCTCCACAGGTGTC (SEQ ID NO: 138)  
3 H1-69 CCGGCACAGCTGCCTTCTCCCTCAG (SEQ ID NO: 139)  
5 DP-47 GAGGTGCAGCTGTTGGAG (SEQ ID NO: 140)  
5 H3-23 TCTGACCAGGGTTTCTTTTGTGTTGC (SEQ ID NO: 141)  
3 H3-23 TTGTGTCTGGGCTCACAATGACTTC (SEQ ID NO: 142)  
5 H3-30 TGGCATTCTTGATAACGGTGTCC (SEQ ID NO: 143)  
3 H3-30 CTGCAGGGAGGTTTGTGTCTGGGCG (SEQ ID NO: 144)  
5 H3-48 ATATGTGTGGCAGTTTCTGACCTTG (SEQ ID NO: 145)  
3 H3-48 GGTTTGTGTCTGGTGTCACTGAC (SEQ ID NO: 146)  
5 H5-a GAGTCTGTGCCGAAGTGCAGCTGG (SEQ ID NO: 147)

#### Specific for coding sequence

5 VH1 TATCAGGTGCAGCTGGTGCAG (SEQ ID NO: 148)  
5 VH3 TATCAGGTGCAGCTGGTGGAG (SEQ ID NO: 149)  
5 VH5 TATGAGGTGCAGCTGGTGCAG (SEQ ID NO: 150)  
3 VH1/3 ATATCTCTCGCACAGTAATACAC (SEQ ID NO: 151)  
3 VH3 ATATCTCTCGCACAGTAATATAC (SEQ ID NO: 152)  
3 VH5 ATATGTCTCGCACAGTAATACAT (SEQ ID NO: 153)  
5 VK1 TATGACATCCAGATGACCCAGTCTCCATCCTC (SEQ ID NO: 154)  
3 DPK9 ATAGGAGGGTACTGTAAC (SEQ ID NO: 155)  
3 DPK1 ATAGGAGGGAGATTATCATA (SEQ ID NO: 156)  
5 DPK22\_L6 TATGAAATTGTGTTGACGCAGTCT (SEQ ID NO: 157)  
3 DPK22 ATAGGAGGTGAGCTACCATACTG (SEQ ID NO: 158)  
5 DPK21 TATGAAATAGTGATGACGCAGTCT (SEQ ID NO: 159)  
3 DPK21 ATAGGAGGCCAGTTATTATACTG (SEQ ID NO: 160)  
3 L6 CAGCGTAGCAACTGGCCTCCTAT (SEQ ID NO: 161)  
5 DPL2 TACAGTCTGTGCTGACTCAG (SEQ ID NO: 162)  
3 DPL2 ATAGGACCATTTCAGGCTGTCATC (SEQ ID NO: 163)  
5 DPL5 TATCAGTCTGTGTTGACGCAG (SEQ ID NO: 164)  
3 DPL5 ATAGGAGCACTCAGGCTGCTAT (SEQ ID NO: 165)

**Primer combinations used to amplify selected germline genes.**

Family	germline	1st PCR		2nd PCR	
		5'	3'	5'	3'
VH1	DP-8/75 HV 1-2	5 H1-2	3 H1-2	5 VH1	3 VH1/3
	DP-10 HV 1-69	5 H1-69	3 H1-69	5 VH1	3 VH1/3
	DP-14 HV 1-18	5 H1-18	3 H1-18	5 VH1	3 VH1/3
VH3	DP-49 HV 3-30	5 H3-30	3 H3-30	5 VH3	3 VH1/3
	DP-51 HV 3-48	5 H3-48	3 H3-48	5 VH3	3 VH1/3
	DP-47 HV 3-23	5 H3-23	3 H3-23	5 VH3	3 VH3
VH5	HV 5a	5 H5a	3 VH5	5 VH5	3 VH5
VK1	DPK-1 KV 1-33	5 K 1-33	3 K 1-33	5 VK1	3 DPK-1
	DPK-9 KV 1-39	5 K 1-39	3 K 1-39	5 VK1	3 DPK-9
VKIII	L6 KV 3-11	5 K3-11	3 K3-11	5 DPK22_L6	3 L6
	DPK-21 KV 3-15	5 K3-15	3 K3-15	5 DPK21	3 DPK21
	DPK-22 KV 3-20	5 K3-20	3 K3-20	5 DPK22_L6	3 DPK22
VL1	DPL-2 LV 1-44	5 L1-44	3 L1-44	5 DPL2	3 DPL2
	DPL-5 LV 1-51	5 L1-51	3 L1-51	5 DPL5	3 DPL5

**Example 2: Generation of Acceptor Frameworks**

[00172] The sequences of the selected germline genes were analyzed for the presence of Type IIs restriction sites. No BsmBI site was present in the selected antibody variable germline genes. Two BsmBI sites were found in the backbone of pNDS1, the phagemid vector in which the Acceptor Framework would be cloned. These two sites were removed by site-directed mutagenesis so that unique BsmBI sites could be introduced into the stuffer DNA sequences of the Acceptor Frameworks. Each germline gene was amplified by multiple nested PCR in order to add a stuffer DNA sequence at the 3' end of the FR3 sequence followed by a sequence encoding FR4 which is specific for each corresponding variable segment (VH, Vk, Vλ). The amino acid sequence of VH FR4 corresponds to the FR4 region encoded by the germline J genes JH1, JH3, JH4 and JH5. The amino acid sequence of VK FR4 corresponds to the FR4 region encoded by the germline J genes JK1. The amino acid sequence of Vλ FR4 corresponds to the FR4 region encoded by the germline J genes JL2 and JL3. Two variants of the Vk FR4

sequence were generated with a single amino acid substitution at position 106 (Arginine or Glycine). For the Acceptor Framework based on the germline gene VH3-23, two variants were also constructed differing by a single amino acid (Lysine to Arginine) at position 94, the last residue of FR3. During the final amplification step SfiI/NcoI and XhoI sites were introduced at the 5' and 3' end of the VH, respectively.

[00173] Similarly, Sall and NotI sites were introduced at the 5' and 3' end of the VL, respectively (Figure 6). The stuffer fragment was designed so that the translation reading frame was shifted thus preventing the expression of any functional protein from the Acceptor Frameworks (Figure 7). The primers used in this process are listed below.

### VH

5 VH1 CAGCCGGCCATGGCCAGGTGCAGCTGGTGCAG (SEQ ID NO: 166)  
 5 VH3-30 CAGCCGGCCATGGCCAGGTGCAGCTGGTGGAG (SEQ ID NO: 167)  
 5 VH3-23 CAGCCGGCCATGGCCAGGTGCAGCTGTTGGAG (SEQ ID NO: 168)  
 5 VH3-48 CAGCCGGCCATGGCCAGGTGCAGCTGGTGGAGTCTGGGGGAG (SEQ ID NO: 169)  
 5 VH5-51 CAGCCGGCCATGGCCAGGTGCAGCTGGTGCAG (SEQ ID NO: 170)  
 3 VH1/3 CTTACCGTTATTTCGTCTCATCTCGCACAGTAATACAC (SEQ ID NO: 171)  
 3 VH3-23 CTTACCGTTATTTCGTCTCATTTTCGCACAGTAATATAC (SEQ ID NO: 172)  
 3 VH3-48 CTCGCACAGTAATACACAGCCGTGTCCTCGGCTCTCAGGCTG (SEQ ID NO: 173)  
 3 VH5-51 CTTACCGTTATTTCGTCTCATCTCGCACAGTAATACAT (SEQ ID NO: 174)  
 3 VHex1 CAATACGCGTTTAAACCTGGTAAACCGCCTTACCGTTATTTCGTCTCA (SEQ ID NO: 175)  
 3 VHex2 GTTCCCTGGCCCCAAGAGACGCGCCTTCCCAATACGCGTTTAAACCTG (SEQ ID NO: 176)  
 3 VHex3 CCTCCACCGCTCGAGACTGTGACCAGGGTCCCTGGCCCCAAGAG (SEQ ID NO: 177)

### VK

5 VK1 CGGGTCGACGGACATCCAGATGACCCAGTC (SEQ ID NO: 178)  
 5 VK3-11 CGGGTCGACGGAAATTGTGTTGACACAGTCTCCAGC (SEQ ID NO: 179)  
 5 VK3-15 CGGGTCGACGGAAATAGTGATGACGCAGTCTCCAGC (SEQ ID NO: 180)  
 5 VK3-20 CGGGTCGACGGAAATTGTGTTGACGCAGTCTCCAGG (SEQ ID NO: 181)  
 3 VK1-33 CCTTACCGTTATTTCGTCTCGCTGCTGACAGTAATATGTTGCAATA (SEQ ID NO: 182)  
 3 VK1-39 CCTTACCGTTATTTCGTCTCGCTGCTGACAGTAGTAAGTTGCAAAA (SEQ ID NO: 183)  
 3 VK3 CCTTACCGTTATTTCGTCTCGCTGCTGACAGTAATAAACTGCAAAATC (SEQ ID NO: 184)  
 3 VKext1 CCAATACGCGTTTAAACCTGGTAAACCGCCTTACCGTTATTTCGTCTC (SEQ ID NO: 185)  
 3 VKext2 GGTCCCTTGGCCGAATGAGACGCGCCTTCCCAATACGCGTTTAAAC (SEQ ID NO: 186)  
 3 Vkext3R GTGCGGCCGCCCGTTTGAATTTCCACCTTGGTCCCTTGGCCGAATG (SEQ ID NO: 187)  
 3 VKext3G GTGCGGCCGCCCTTGAATTTCCACCTTGGTCCCTTGGCCGAATG (SEQ ID NO: 188)

### VL

5 VL1-44 CGGGTCGACGCAGTCTGTGCTGACTCAGCCAC (SEQ ID NO: 189)  
 5 VL1-51 CGGGTCGACGCAGTCTGTGTTGACGCAGCCGC (SEQ ID NO: 190)  
 3 VL1-44 CCTTACCGTTATTTCGTCTCCTGCTGCACAGTAATAATC (SEQ ID NO: 191)  
 3 VL1-51 CCTTACCGTTATTTCGTCTCCTGTTCCGCAGTAATAATC (SEQ ID NO: 192)  
 3 V1ext2 CCCTCCGCCGAACACAGAGACGCGCCTTCCCAATACGCGTTTAAAC (SEQ ID NO: 193)  
 3 V1ext3 GTGCGGCCGCCCTAGGACGGTCAGCTTGGTCCCTCCGCCGAACACAGA (SEQ ID NO: 194)

[00174] The sequences of the 20 final assembled Acceptor Frameworks are shown in Figure 8.

**EXAMPLE 3: Generation of phagemid Acceptor vectors containing an invariant variable domain**

[00175] The phagemid vector pNDS1 used for the expression of scFv was first modified to remove two BsmBI sites. A VH3-23 domain containing a defined CDR3 sequence was cloned into the modified pNDS1 using the SfiI and XhoI restriction sites to obtain the phagemid vector pNDS\_VHdummy. This domain contained a BsmBI site in the FR4 region, which was corrected by silent site directed mutagenesis. In parallel, a VK1-39 domain containing a defined CDR3 sequence was then cloned into the modified pNDS1 using the Sall and NotI restriction sites to obtain the phagemid vector pNDS\_VKdummy (Figure 9). The 8 VH Acceptor Frameworks were cloned into pNDS\_VKdummy using the Sall and NotI restrictions sites. The 12 VL Acceptor Frameworks were cloned into pNDS\_VHdummy using the SfiI and XhoI restrictions sites. The resulting 20 pNDS phagemid vectors that are listed below could at this stage be used for cloning of diversified CDR3 using the BsmBI sites present in the stuffer DNA fragments.

[00176] VH Acceptors: pNDS\_VH1-2\_VKd; pNDS\_VH1-18\_VKd; pNDS\_VH1-69\_VKd; pNDS\_VH3-23R\_VKd; pNDS\_VH3-23K\_VKd; pNDS\_VH3-30\_VKd; pNDS\_VH5-51\_VKd; pNDS\_VH3-48\_VKd.

[00177] VL Acceptors: pNDS\_VHd\_VK1-33G; pNDS\_VHd\_VK1-33R; pNDS\_VHd\_VK1-39G; pNDS\_VHd\_VK1-39R; pNDS\_VHd\_VK3-11G; pNDS\_VHd\_VK3-11R; pNDS\_VHd\_VK3-15G; pNDS\_VHd\_VK3-15R; pNDS\_VHd\_VK3-20G; pNDS\_VHd\_VK3-20R; pNDS\_VHd\_VL1-44; pNDS\_VHd\_VK1-51.

**EXAMPLE 4: Capturing natural CDR H3 diversity from human repertoires**

[00178] Multiple sources of human cDNA were used as a template for amplification of CDR H3 sequences. These sources included human fetal spleen as well as pools of male and female normal adult peripheral blood purified cells. Several strategies for amplification have been used in order to recover CDR H3 sequences originating from rearranged VH cDNA encoded by a specific germline gene or CDR H3 sequences originating from any VH cDNA.

**[00179]** First, mixtures of primers matching the 5' coding regions of the majority of human VH families were used in combination with primer mixtures matching all the human JH regions. This allowed for PCR amplification a majority of heavy chain immunoglobulin variable genes. The expected amplification products of approximately 400 base pairs (bp) were isolated by agarose gel electrophoresis and purified. This DNA served as template in a second PCR step using primers with a 13 bp and 14 bp match for the end FR3 region and the beginning of FR4, respectively. In most cases, the last residue of the FR3 is either an arginine or a lysine. As the last bp matches are critical for primer extension by the polymerase, two different 5' primers were used: 5 VHR\_FOK (SEQ ID NO: 205 shown below) and 5 VHK\_FOK (SEQ ID NO: 206 shown below). Importantly, these primers also contain a FokI restriction site for excision of the CDR H3 sequence (Figure 4). The primers used in the second PCR step were biotinylated at their 5' end to facilitate downstream purification steps (see example 5). This two step approach allows for an efficient amplification of the CDR H3 sequences despite the limited number of base pairs matches. Amplifications were performed at varying annealing temperatures (between 30°C and 70°C) and with several thermostable DNA polymerases to establish optimal conditions. An annealing temperature of 55-58°C in combination with GoTaq polymerase (Promega) was found to be optimal for this set of primers. The second amplification product was separated on a 2% agarose gel and resulted in a smear in the lower part of the gel corresponding to CDR H3 of different length. Either the complete DNA smear was extracted from the gel or a region corresponding to larger DNA fragments in order to enrich for long CDR H3.

**[00180]** Alternatively, the first amplification step was performed using the 5' primer 5 VH3-23H2 (SEQ ID NO: 201 shown below), which is specific for the sequence encoding the CDR H2 of the germline VH3-23. As the different germline genes are diverse in this CDR, VH cDNAs encoded by the selected germline gene can be preferentially amplified. The subsequent purification and amplification steps were identical. In this way, it is possible to retrieve CDRs originating from a specific framework environment and to re-introduce them into the same, a similar or different framework.

**[00181]** Below is a list of primers used for the amplification of natural human CDR H3 repertoires.

### 1st PCR step

5 VH1/5 CCGCACAGCCGGCCATGGCCCAGGTGCAGCTGGTGCAGTCTGG (SEQ ID NO: 195)  
5 VH3 CCGCACAGCCGGCCATGGCCCAGGTGCAGCTGGTGGAGTCTGG (SEQ ID NO: 196)  
5 VH2 CCGCACAGCCGGCCATGGCCCAGRTCACCTTGCTCGAGTCTGG (SEQ ID NO: 197)  
5 VH4 CCGCACAGCCGGCCATGGCCCAGGTGCAGCTGCAGGAGTCGGG (SEQ ID NO: 198)  
5 VH4DP64 CCGCACAGCCGGCCATGGCCCAGGTGCAGCTGCAGGAGTCCGG (SEQ ID NO: 199)  
5 VH4DP63 CCGCACAGCCGGCCATGGCCCAGGTGCAGCTACAGCAGTGGGG (SEQ ID NO: 200)  
5 VH3-23H2 TGGAGTGGGTCTCAGCTATTAGTGGTAGTGGT (SEQ ID NO: 201)  
3 HJ1/2 CGATGGGCCCTTGGTGGAGGCTGAGGAGACRGTGACCAGGGTGCC (SEQ ID NO: 202)  
3 HJ3/6 CGATGGGCCCTTGGTGGAGGCTGAAGAGACGGTGACCRKGTCCC (SEQ ID NO: 203)  
3 HJ4/5 CGATGGGCCCTTGGTGGAGGCTGAGGAGACGGTGACCAGGGTTCC (SEQ ID NO: 204)

### 2nd PCR step

5 VHR\_FOK GAGCCGAGGACACGGCCGGATGTTACTGTGCGAGA (SEQ ID NO: 205)  
5 VHK\_FOK GAGCCGAGGACACGGCCGGATGTTACTGTGCGAAA (SEQ ID NO: 206)  
3 JH1\_FOK GAGGAGACGGTGACGGATGTGCCCTGGCCCCA (SEQ ID NO: 207)  
3 JH2\_FOK GAGGAGACGGTGACGGATGTGCCACGGCCCCA (SEQ ID NO: 208)  
3 JH3456\_FOK GAGGAGACGGTGACGGATGTTCCTTGGCCCCA (SEQ ID NO: 209)

### EXAMPLE 5. Generation of primary libraries by cloning natural human CDR H3 into acceptor frameworks

[00182] The amplified CDR H3 were digested with FokI, and the cleaved extremities as well as undigested DNA was removed using streptavidin coated magnetic beads. In parallel, pNDS VH Acceptor vectors were digested using BsmBI. As the overhangs generated by these digestions are compatible, the collection of natural CDR H3 was able to be ligated into the VH Acceptor Framework restoring the appropriate reading frame. The ligated DNA was purified and concentrated for transformation into competent *E. coli* XL1 Blue cells, and random clones analyzed by sequencing in order to check that CDR H3 sequence had been reconstituted and that junctions between the CDR and the Framework region are correct (Figure 10). The results indicated that all the clones contained CDR H3 sequences and that the reading frame was restored, thus encoding an immunoglobulin variable heavy chain. In addition, all the CDRs were different, indicating that a large diversity of naturally occurring sequences had been captured by this approach. The length of the CDR H3 was also variable and relatively long CDRs of 10 to 15 residues were found, thus underscoring the advantage of this approach for sampling long CDR sequences that are difficult to cover using synthetic diversity.

[00183] Using this method, natural CDR H3 sequences, derived either from pooled human peripheral blood purified cells or human fetal spleen, were cloned into each of the pNDS VH Acceptor Frameworks and transformed into electrocompetent *E. coli* TG1 cells and

plated on 2xTYAG Bioassay plates (2xTY media containing 100 µg/ml ampicillin and 2% glucose). After overnight incubation at 30 °C, 10 ml of 2xTYAG liquid medium was added to the plates and the cells were scraped from the surface and transferred to a 50 ml polypropylene tube. 2xTYAG containing 50% glycerol was added to the cell suspension to obtain a final concentration of 17% glycerol. Aliquots of the libraries were stored at -80 °C. In this process, 14 primary libraries were generated representing a total of  $8.1 \times 10^9$  transformants. 180 randomly picked clones were sequenced to determine the quality and diversity of the libraries. All clones encoded different VH sequences and >89% were in frame. These primary libraries contain diversity in the CDR H3 only as they are combined with a dummy VL domain.

#### **Example 6. Generation of primary libraries by cloning synthetic CDR3 into acceptor frameworks**

**[00184]** Although the method is of particular interest for retrieving natural diversity, it can also be applied for the integration of synthetic diversity into Acceptor Frameworks. Synthetic CDR3 sequences were designed for both the VH and VL. The design took into account the frequency of CDRs with a given length and the diversification strategy (NNS, DVK, NVT or DVT codons) that would allow a complete coverage of the theoretical diversity within a reasonable number of transformants in a library ( $\sim 5 \times 10^9$  transformants) (Figure 11). Key residues to maintain the canonical structure of the CDR were kept constant in the design of CDR3 for VK and V $\lambda$  chains. For the heavy chain, only CDR3 with up to 10 diversified positions were generated as the number of clones required to cover the diversity encoded by longer CDRs is beyond practical limits of transformation efficiency.

**[00185]** Degenerate oligonucleotides of different length were synthesized using NNS, NVT, DVK or DVT randomized codons. For each CDR H3, two oligonucleotides were synthesized encoding either a methionine or a phenylalanine at position 100z (Figure 11). Each oligonucleotide was extended and amplified with two external biotinylated primers to generate double stranded DNA fragments encoding the designed CDRs. These external primers contain BsmBI restriction sites for subsequent excision of the CDR sequence and insertion into the Acceptor Frameworks (Figure 12). The assembled DNA fragments were processed without gel purification and digested with BsmBI. The cleaved extremities as well as undigested DNA was removed using streptavidin coated magnetic beads. The digested DNA fragments were

concentrated by ethanol precipitation and ligated into the corresponding pNDS VH, VK or VL Acceptor vectors. Ligation products were purified and concentrated for transformation into electrocompetent *E. coli* TG1 cells and plated on 2xTYAG Bioassay plates (2xTY media containing 100 µg/ml ampicillin and 2% glucose). After overnight incubation at 30 °C, 10 ml of 2xTYAG liquid medium was added to the plates and the cells were scraped from the surface and transferred to a 50 ml polypropylene tube. 2xTYAG containing 50% glycerol was added to the cell suspension to obtain a final concentration of 17% glycerol. Aliquots of the libraries were stored at -80 °C. A total of 24 primary heavy chain libraries were generated representing a total of  $1.6 \times 10^{10}$  transformants. Similarly, 13 primary light chain libraries were generated representing a total of  $6.9 \times 10^9$  transformants. These primary libraries contain diversity in the CDR H3 only as they are combined with a dummy VL domain. A total of 330 randomly picked clones were sequenced to determine the quality and diversity of the libraries. All clones encoded different variable domain sequences and >90% were in frame. This low frequency of sequences containing shifts in the reading frame is in sharp contrast with results traditionally obtained during the construction of synthetic antibody fragment libraries using overlapping PCR approaches which are more prone to the introduction of insertion, and significant loss of functional clones (20-40%) has frequently been reported.

**[00186]** The diversity in these primary libraries was restricted to the CDR H3 or CDR L3 as they are combined with a dummy VL or VH chain, respectively.

**[00187]** Primers used for synthetic CDR assembly are listed below.

5 H3\_R\_biot ATGATGCTGCTGGCAGTCTCCGAGA (SEQ ID NO: 210)  
 3 H3\_M\_biot CCACGTCATCCGATCCGCTCTCCCCAATAATCCAT (SEQ ID NO: 211)  
 3 H3\_F\_biot CCACGTCATCCGATCCGCTCTCCCCAATAATCAA (SEQ ID NO: 212)  
 H3\_4nnsF GCTGGCACGTCTCCGAGANNSNNSNNSNNSNTTGGATTATTGGGGGAGACG (SEQ ID NO: 213)  
 H3\_4nnsM GCTGGCACGTCTCCGAGANNSNNSNNSNNSNNSATGGATTATTGGGGGAGACG (SEQ ID NO: 214)  
 H3\_5nnsF GCTGGCACGTCTCCGAGANNSNNSNNSNNSNNSNTTGGATTATTGGGGGAGACG (SEQ ID NO: 215)  
 H3\_5nnsM GCTGGCACGTCTCCGAGANNSNNSNNSNNSNNSNNSATGGATTATTGGGGGAGACG (SEQ ID NO: 216)  
 H3\_6nnsF GCTGGCACGTCTCCGAGANNSNNSNNSNNSNNSNNSNTTGGATTATTGGGGGAGACG  
 (SEQ ID NO: 217)  
 H3\_6nnsM GCTGGCACGTCTCCGAGANNSNNSNNSNNSNNSNNSNNSATGGATTATTGGGGGAGACG  
 (SEQ ID NO: 218)  
 H3\_6dvkF GCTGGCACGTCTCCGAGADVVDVVDVVDVVDVVDVKTGGATTATTGGGGGAGACG  
 (SEQ ID NO: 219)  
 H3\_6dvkM GCTGGCACGTCTCCGAGADVVDVVDVVDVVDVVDVVDVKTGGATTATTGGGGGAGACG  
 (SEQ ID NO: 220)  
 H3\_7dvkF GCTGGCACGTCTCCGAGADVVDVVDVVDVVDVVDVVDVKTGGATTATTGGGGGAGACG  
 (SEQ ID NO: 221)  
 H3\_7dvkM GCTGGCACGTCTCCGAGADVVDVVDVVDVVDVVDVVDVVDVKTGGATTATTGGGGGAGACG  
 (SEQ ID NO: 222)





### **Example 7. Generation of secondary libraries**

**[00188]** In order to generate libraries of scFv carrying diversity in both the heavy and light chains, the Primary synthetic light chain libraries were combined with either the Primary synthetic heavy chain libraries or the Primary natural heavy chain libraries (Figure 13). Phagemid DNA was prepared from each primary library and digested with XhoI/NotI restriction enzymes. The DNA fragments corresponding to the linker and light chains from the Primary synthetic libraries were inserted by ligation into the digested Primary natural or synthetic heavy chain vectors. Alternatively the Linker-VL sequence was also amplified with specific primers before digestion with XhoI/NotI and ligation. The ligation products were purified by phenol/chloroform extraction and precipitation before transformation into electrocompetent *E. coli* TG1 cells and plating on 2xTYAG Bioassay plates (2xTY media containing 100 µg/ml ampicillin and 2% glucose). After overnight incubation at 30 °C, 10 ml of 2xTYAG liquid medium was added to the plates and the cells were scraped from the surface and transferred to a 50 ml polypropylene tube. 2xTYAG containing 50% glycerol was added to the cell suspension to obtain a final concentration of 17% glycerol. Aliquots of the libraries were stored at -80 °C. To limit the number of libraries to be recombined, they were pooled by chain subclasses (*i.e.*, VH1, VH3, VH5, VK1, VK3, Vλ1) and thus 9 library combination were performed for (*i.e.*, VH1xVK1, VH1xVK3, VH1xVλ1, VH3xVK1, VH3xVK3, VH3xVλ1, VH5xVK1, VH5xVK3, VH5xVλ1). The total size of the Secondary synthetic libraries (carrying synthetic diversity in both the VH and VL) was  $7.3 \times 10^9$  transformants. The total size of the Secondary natural libraries (carrying natural diversity in the VH and synthetic diversity in the VL) was  $1.5 \times 10^{10}$  transformants.

### **Example 8. Phage rescue of the libraries**

**[00189]** Each Primary and Secondary library was rescued independently according to standard phage display procedures briefly summarized hereafter. A volume of cell from the frozen library aliquots sufficient to cover at least 10 times the theoretical diversity of the library was added to 500 ml of 2xTYAG and grown at 37 °C with agitation (240 rpm) until an OD600 of 0.3 to 0.5 was reached. The culture was then super-infected with MK13K07 helper phage and incubated for one hour at 37 °C (150 rpm). The medium was then changed by centrifugating the cells at 2000 rpm for 10 minutes, removing the medium and resuspending the

pellet in 500 ml of 2xTY-AK (100µg/ml ampicilin; 50 µg/ml kanamycin). The culture was then grown overnight at 30 °C (240 rpm). The culture was centrifugated at 4000 rpm for 20 minutes to pellet the cells. The supernatant was collected and 30% (vol/vol) of PEG 8000 (20%)/2.5M NaCl was added to precipitate the phage particles by incubating the mixture 1 hour on ice. The phage particles were collected by centrifugation at 10,000 rpm for 30 minutes and resuspended in 10ml of TE buffer (10 mM tris-HCl pH 8.0; 1mM EDTA). The resuspended solution was centrifuged at 10,000 rpm to clear the bacterial debris and the precipitation procedure was repeated. After final resuspension, phage was titrated by infection of *E. coli* and absorption at 280 nm. The display level of scFv at the surface of phage was also evaluated by Western blot analysis using an anti-c-myc monoclonal antibody. Purified phage from different libraries was stored frozen at -80°C after addition of glycerol to a final concentration of 15% (w/v).

**[00190]** In order to use a manageable number of libraries during selection procedures, the purified phage was pooled into 4 working libraries: AA1 – Phage from all Primary synthetic VH libraries; AB1 - Phage from all Primary synthetic VL libraries; AC1 - Phage from all Primary natural VH libraries; AD1 - Phage from all Secondary natural libraries; AE1 - Phage from all Secondary synthetic libraries.

#### **Example 9. Phage display selections using Secondary Libraries**

**[00191]** *Liquid phase selections against human interferon gamma (hINF $\gamma$ ):* Aliquots of AD1 and AE1 phage libraries ( $10^{11}$ - $10^{12}$  Pfu) were blocked with PBS containing 3% (w/v) skimmed milk for one hour at room temperature on a rotary mixer. Blocked phage was then deselected on streptavidin magnetic beads (Dynal M-280) for one hour at room temperature on a rotary mixer. Deselected phage was then incubated with *in vivo* biotinylated hINF $\gamma$  (100 nM) for two hours at room temperature on a rotary mixer. Beads were captured using a magnetic stand followed by four washes with PBS/0.1% Tween 20 and 3 washes with PBS. Beads were then directly added to 10 ml of exponentially growing TG1 cells and incubated for one hour at 37 °C with slow shaking (100 rpm). An aliquot of the infected TG1 was serially diluted to titer the selection output. The remaining infected TG1 were spun at 3000 rpm for 15 minutes and re-suspended in 0.5 ml 2xTYAG (2xTY media containing 100 µg/ml ampicilin and 2% glucose) and spread on 2xTYAG agar Bioassay plates. After overnight incubation at 30 °C,

10 ml of 2xTYAG was added to the plates and the cells were scraped from the surface and transferred to a 50 ml polypropylene tube. 2xTYAG containing 50% glycerol was added to the cell suspension to obtain a final concentration of 17% glycerol. Aliquots of the selection round were kept at  $-80^{\circ}\text{C}$ . Phage outputs were titrated after each round and the progressive increase in outputs indicated that the enrichment of clones specific for the target was occurring (Figure 14).

**[00192]**      *Selections by panning against the rat monoclonal antibody 5E3:* Immunotubes were coated with 5E3 at  $10\mu\text{g/ml}$  in PBS over night at  $4^{\circ}\text{C}$  and immunotubes for phage deselection were coated with an irrelevant rat antibody under the same conditions. After washing immunotubes were blocked with PBS containing 3% (w/v) skimmed milk for one hour at room temperature. Aliquots of AD1 and AE1 phage libraries ( $10^{11}$ - $10^{12}$  Pfu) were blocked with PBS containing 3% (w/v) skimmed milk for one hour at room temperature on a rotary mixer. Blocked phage was then deselected in the immunotubes coated with an irrelevant rat antibody for one hour at room temperature on a rotary mixer. Deselected phage was then transferred to the immunotubes coated with 5E3 and incubated for two hours at room temperature on a rotary mixer. Tubes were washed five times with PBS/0.1% Tween 20 and 3 times with PBS. Phage was eluted with TEA 100mM for 10 minutes and neutralized with 1M Tris HCl pH 7.5. Phage was added to 10 ml of exponentially growing TG1 cells and incubated for one hour at  $37^{\circ}\text{C}$  with slow shaking (100 rpm). An aliquot of the infected TG1 was serially diluted to titer the selection output. The remaining infected TG1 were spun at 3000 rpm for 15 minutes and re-suspended in 0.5 ml 2xTYAG (2xTY media containing  $100\mu\text{g/ml}$  ampicillin and 2% glucose) and spread on 2xTYAG agar Bioassay plates. After overnight incubation at  $30^{\circ}\text{C}$ , 10 ml of 2xTYAG was added to the plates and the cells were scraped from the surface and transferred to a 50 ml polypropylene tube. 2xTYAG containing 50% glycerol was added to the cell suspension to obtain a final concentration of 17% glycerol. Aliquots of the selection round were kept at  $-80^{\circ}\text{C}$ . Rounds of selection were performed by alternating between rat 5E3 and a chimeric version of 5E3 in which the variable region were fused to mouse constant domains. This alternating rounds were performed in order to enrich for clones specific for the variable region of 5E3 and generate anti-idiotypic antibodies. Phage outputs were titrated after each round and the progressive increase in outputs indicated that the enrichment of clones specific for the target was occurring (Figure 15).

**[00193]** *Phage rescue:* 100 µl of cell suspension obtained from previous selection rounds were added to 20 ml of 2xTYAG and grown at 37 °C with agitation (240 rpm) until an OD600 of 0.3 to 0.5 was reached. The culture was then super-infected with  $3.3 \times 10^{10}$  MK13K07 helper phage and incubated for one hour at 37 °C (150 rpm). The medium was then changed by centrifugating the cells at 2000 rpm for 10 minutes, removing the medium and resuspending the pellet in 20 ml of 2xTY-AK (100 µg/ml ampicilin; 50 µg/ml kanamycin). The culture was then grown overnight at 30 °C (240 rpm).

**[00194]** *Monoclonal phage rescue for ELISA:* Single clones were picked into a microtiter plate containing 150 µl of 2xTYAG media (2% glucose) per well and grown at 37°C (100-120 rpm) for 5-6h. M13K07 helper phage was added to each well to obtain a multiplicity of infection (MOI) of 10 (*i.e.*, 10 phage for each cell in the culture) and incubated at 37°C (100 rpm) for 1h. Following growth, plates were centrifuged at 3,200 rpm for 10 min. Supernatant was carefully removed, cells resuspended in 150 µl 2xTYAK medium and grown overnight at 30 °C (120 rpm). For the ELISA, the phage are blocked by adding 150µl of 2x concentration PBS containing 5% skimmed milk powder followed by one hour incubation at room temperature. The plates were then centrifuged 10 minutes at 3000 rpm and the phage containing supernatant used for the ELISA.

**[00195]** *Phage ELISA:* ELISA plates (Maxisorb, NUNC) were coated overnight with 2 µg/ml hIFN $\gamma$  in PBS or 2 µg/ml rat 5E3 in PBs. Control plates were coated with 2µg/ml BSA or an irrelevant rat monoclonal antibody. Plates were then blocked with 3% skimmed milk / PBS at room temperature for 1h. Plates were washed 3 times with PBS 0.05% Tween 20 before transferring the pre-blocked phage supernatants and incubation for one hour at room temperature. Plates were then washed 3 times with PBS 0.05% Tween 20. 50µl of 3% skimmed milk / PBS containing (HRP)-conjugated anti-M13 antibody (Amersham, diluted 1:10,000) to each well. Following incubation at room temperature for 1 hr, the plates were washed 5 times with PBS 0.05% Tween 20. The ELISA was then revealed by adding 50µl of TMB (Sigma) and 50µl of 2N H<sub>2</sub>SO<sub>4</sub> to stop the reaction. Absorption intensity was read at 450nm. Clones specific for hIFN $\gamma$  could be identified and the hit rates ranged between 10% and 30% after the third round of selection. Clones specific for the variable region of 5E3 could also be identified and the hit rates ranged between 7 and 48% after the third round of selection.

**[00196]** *Phage clone sequencing:* Single clones were grown in 5 ml of 2xTYAG media (2% glucose) per well and grown at 37 °C (120 rpm) overnight. The next day phagemid DNA was purified and used for DNA sequencing using a primer specific for pNDS1: mycseq, 5'-CTCTTCTGAGATGAGTTTTTG. (SEQ ID NO: 255).

**[00197]** *Large scale scFv purification:* A starter culture of 1 ml of 2xTYAG was inoculated with a single colony from a freshly streaked 2xTYAG agar plate and incubated with shaking (240 rpm) at 37 °C for 5 hours. 0.9 ml of this culture was used to inoculate a 400 ml culture of the same media and was grown overnight at 30 °C with vigorous shaking (300 rpm).

**[00198]** The next day the culture was induced by adding 400 µl of 1M IPTG and incubation was continued for an additional 3 hours. The cells were collected by centrifugation at 5,000 rpm for 10 minutes at 4 °C. Pelleted cells were resuspended in 10 ml of ice-cold TES buffer complemented with protease inhibitors as described above. Osmotic shock was achieved by adding 15 ml of 1:5 diluted TES buffer and incubation for 1 hour on ice. Cells were centrifuged at 10,000 rpm for 20 minutes at 4 °C to pellet cell debris. The supernatant was carefully transferred to a fresh tube. Imidazole was added to the supernatant to a final concentration of 10 mM. 1 ml of Ni-NTA resin (Qiagen), equilibrated in PBS was added to each tube and incubated on a rotary mixer at 4 °C (20 rpm) for 1 hour. The tubes were centrifuged at 2,000 rpm for 5 minutes and the supernatant carefully removed. The pelleted resin was resuspended in 10 ml of cold (4 °C) Wash buffer 1 (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH to 8.0). The suspension was added to a polyprep column (Biorad). 8 ml of cold Wash Buffer 2 (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH to 8.0) were used to wash the column by gravity flow. The scFv were eluted from the column with 2 ml of Elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH to 8.0). Fractions were analyzed by absorption at 280 nm and protein containing fractions were pooled before buffer exchange on a PD10 desalting column (Amersham) equilibrated with PBS. The scFv in PBS were analyzed by SDS-PAGE and quantified by absorption at 280 nm. The purified scFv were aliquoted and stored at -20°C and at 4°C.

#### **Example 10. Evaluating and Testing Identified Protein Variants**

**[00199]** Purified scFvs preparations of clones having different sequences and that were identified positive against the variable region of 5E3 were tested for binding against chimeric

5E3 in a dose response ELISA. These preparations were also tested against an irrelevant mouse antibody (1A6). ELISA plates (Maxisorb, NUNC) were coated overnight with 2 µg/ml mouse 5E3 in PBS. Control plates were coated with 2µg/ml 1A6 monoclonal antibody. Plates were then blocked with 3% skimmed milk / PBS at room temperature for 1h. Plates were washed 3 times with PBS 0.05% Tween 20 before adding different concentrations of purified scFv and incubation for one hour at room temperature. Plates were then washed 3 times with PBS 0.05% Tween 20. 50µl of 3% skimmed milk / PBS containing (HRP)-conjugated anti-myc antibody to each well. Following incubation at room temperature for 1 hr, the plates were washed 5 times with PBS 0.05% Tween 20. The ELISA was then revealed by adding 50µl of Amplex Red fluorescent substrate and the signal was read on fluorescence spectrophotometer. The data shows that most of the clones are highly specific for 5E3 as they do not recognize 1A6 and that they are directed against the variable regions of 5E3 (Figure 16).

**[00200]** Similarly, purified scFvs preparations of clones having different sequences and that were identified in phage ELISA as binders against hIFN $\gamma$  were tested for binding against hIFN $\gamma$  in a dose response experiment . ELISA plates (Maxisorb, NUNC) were coated overnight with 2 µg/ml hIFN $\gamma$  in PBS and control plates were coated with 2µg/ml BSA in PBS. Plates were then blocked with 3% skimmed milk / PBS at room temperature for 1h. Plates were washed 3 times with PBS 0.05% Tween 20 before adding different concentration of purified scFv and incubation for one hour at room temperature. Plates were then washed 3 times with PBS 0.05% Tween 20. 50µl of 3% skimmed milk / PBS containing (HRP)-conjugated anti-myc antibody to each well. Following incubation at room temperature for 1 hr, the plates were washed 5 times with PBS 0.05% Tween 20. The ELISA was then revealed by adding 50µl TMB substrate and 50µl of 2N H<sub>2</sub>SO<sub>4</sub> to stop the reaction. The signal was read on an absorbance spectrophotometer at 450 nm. The data shows that the selected clones are binding to hIFN $\gamma$  in a dose dependent manner and gave a very good signal when compared to a positive control scFv A6 that has a high affinity for hIFN $\gamma$  (Figure 17).

#### **Other Embodiments**

**[00201]** While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

What is claimed is:

1. A method for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type II restriction enzyme recognition sites interspaced by a random nucleic acid sequence;
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 3 (CDR3) regions or encoding amino acid sequences that can fulfill the role of a CDR3 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type II restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR3 regions or amino acid sequences that can fulfill the role of a CDR3 region using a Type II restriction enzyme that binds to the Type II restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type II restriction enzyme that binds to the Type II restriction enzyme recognition site of step (a); and
- (d) ligating the digested nucleic acid sequences encoding the CDR3 regions or the amino acid sequences that can fulfill the role of a CDR3 region of step (c) into the digested Acceptor Framework of step (c) such that the FR3 and FR4 regions are interspaced by the nucleic acid sequences encoding the CDR3 region or the amino acid sequence that can fulfill the role of a CDR3 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type II restriction enzyme recognition sites of steps (a) and (b) are restored.



2. A method for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3);
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 1 (CDR1) regions or encoding amino acid sequences that can fulfill the role of a CDR1 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR1 regions or amino acid sequences that can fulfill the role of a CDR1 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); and
- (d) ligating the digested nucleic acid sequences encoding the CDR1 regions or the amino acid sequences that can fulfill the role of a CDR1 region of step (c) into the digested Acceptor Framework of step (c) such that the FR1 and FR2 regions are interspaced by the nucleic acid sequences encoding the CDR1 region or the amino acid sequence that can fulfill the role of a CDR1 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

3. A method for producing a library of nucleic acids, wherein each nucleic acid encodes an immunoglobulin variable domain, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3);
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 2 (CDR2) regions or encoding amino acid sequences that can fulfill the role of a CDR2 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR2 regions or amino acid sequences that can fulfill the role of a CDR2 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a); and
- (d) ligating the digested nucleic acid sequences encoding the CDR2 regions or the amino acid sequences that can fulfill the role of a CDR2 region of step (c) into the digested Acceptor Framework of step (c) such that the FR2 and FR3 regions are interspaced by the nucleic acid sequences encoding the CDR2 region or the amino acid sequence that can fulfill the role of a CDR2 region and a complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored.

4. The method according to any one of claims 1-3, wherein the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by a different Type IIs restriction enzyme.

5. The method of claim 4, wherein the Type II restriction enzyme recognition sites are BsmBI recognition sites, BsaI recognition sites, FokI recognition sites or a combination thereof.
6. The method according to any one of claims 1-3, wherein the Acceptor Framework nucleic acid sequence is derived from a human gene sequence.
7. The method of claim 6, wherein the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence.
8. The method of claim 7, wherein the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51.
9. The method of claim 6, wherein the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence.
10. The method of claim 9, wherein the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3-15, and VK3-20.
11. The method of claim 6, wherein the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence.
12. The method of claim 11, wherein the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.
13. The method of claim 1, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally

occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

14. The method of claim 1, wherein the plurality of diversified nucleic acids encodes CDR3 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

15. The method of claim 1, wherein the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

16. The method of claim 2, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

17. The method of claim 2, wherein the plurality of diversified nucleic acids encodes CDR1 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

18. The method of claim 2, wherein the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

19. The method of claim 3, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring

Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

20. The method of claim 3, wherein the plurality of diversified nucleic acids encodes CDR2 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

21. The method of claim 3, wherein the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR2 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

22. The method according to any one of claims 1-3, wherein the plurality of Acceptor Framework nucleic acid sequences comprises a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

23. The method according to any one of claims 1-3, further comprising the steps of (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector and (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domain encoded by the library.

24. The method of claim 23, wherein the host cell is *E. coli*.

25. The method according to claim 23, wherein the expression vector is a phagemid vector.

26. A method for making a target-specific antibody, antibody variable region or a portion thereof, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence;
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 3 (CDR3) regions or encoding amino acid sequences that can fulfill the role of a CDR3 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR3 regions or amino acid sequences that can fulfill the role of a CDR3 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a);
- (d) ligating the digested nucleic acid sequences encoding the CDR3 regions or the amino acid sequences that can fulfill the role of a CDR3 region of step (c) into the digested Acceptor Framework of step (c) such that the FR3 and FR4 regions are interspaced by the nucleic acid sequences encoding the CDR3 region or the amino acid sequence that can fulfill the role of a CDR3 region and complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored;
- (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector;
- (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domains encoded by the library;

- (g) contacting the plurality of immunoglobulin domains of step (f) with a target antigen;  
and
- (h) determining which expressed immunoglobulin variable domain encoding sequences bind to the target antigen.

27. A method for making a target-specific antibody, antibody variable region or a portion thereof, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type IIs restriction enzyme recognition sites interspaced by a random nucleic acid sequence, the FR2 and FR3 regions are interspaced by a complementarity determining region 2 (CDR2), and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3);
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 1 (CDR1) regions or encoding amino acid sequences that can fulfill the role of a CDR1 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type IIs restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR1 regions or amino acid sequences that can fulfill the role of a CDR1 region using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a);
- (d) ligating the digested nucleic acid sequences encoding the CDR1 regions or the amino acid sequences that can fulfill the role of a CDR1 region of step (c) into the digested Acceptor Framework of step (c) such that the FR1 and FR2 regions are interspaced by the nucleic acid sequences encoding the CDR1 region or the amino acid sequence that can fulfill the role of a CDR1 region and complete immunoglobulin variable domain

encoding sequences that do not contain the Type II's restriction enzyme recognition sites of steps (a) and (b) are restored;

- (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector;
- (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domains encoded by the library;
- (g) contacting the plurality of immunoglobulin domains of step (f) with a target antigen; and
- (g) determining which expressed immunoglobulin variable domain encoding sequences bind to the target antigen.

28. A method for making a target-specific antibody, antibody variable region or a portion thereof, the method comprising:

- (a) providing a plurality of Acceptor Framework nucleic acid sequences encoding distinct immunoglobulin variable domains, each Acceptor Framework nucleic acid sequence comprising a first framework region (FR1), a second framework region (FR2), a third framework region (FR3), and a fourth framework region (FR4), wherein the FR1 and FR2 regions are interspaced by a complementarity determining region 1 (CDR1), the FR2 and FR3 regions are interspaced by a stuffer nucleic acid sequence comprising at least two Type II's restriction enzyme recognition sites interspaced by a random nucleic acid sequence, and the FR3 and FR4 regions are interspaced by a complementarity determining region 3 (CDR3);
- (b) providing a plurality of diversified nucleic acid sequences encoding complementarity determining region 2 (CDR2) regions or encoding amino acid sequences that can fulfill the role of a CDR2 region, wherein each of the plurality of diversified nucleic acid sequences comprises a Type II's restriction enzyme recognition site at each extremity;
- (c) digesting each of the plurality of nucleic acid sequences encoding the CDR2 regions or amino acid sequences that can fulfill the role of a CDR2 region using a Type II's restriction enzyme that binds to the Type II's restriction enzyme recognition site of step (b) and digesting the stuffer nucleic acid sequence of step (a) from the Acceptor



- Framework using a Type IIs restriction enzyme that binds to the Type IIs restriction enzyme recognition site of step (a);
- (d) ligating the digested nucleic acid sequences encoding the CDR2 regions or the amino acid sequences that can fulfill the role of a CDR2 region of step (c) into the digested Acceptor Framework of step (c) such that the FR2 and FR3 regions are interspaced by the nucleic acid sequences encoding the CDR2 region or the amino acid sequence that can fulfill the role of a CDR2 region and complete immunoglobulin variable domain encoding sequences that do not contain the Type IIs restriction enzyme recognition sites of steps (a) and (b) are restored;
  - (e) cloning the library of nucleic acids encoding immunoglobulin variable domains of step (d) into an expression vector;
  - (f) transforming the expression vector of step (e) into a host cell and culturing the host cell under conditions sufficient to express a plurality of immunoglobulin variable domains encoded by the library;
  - (g) contacting the plurality of immunoglobulin variable domains of step (f) with a target antigen; and
  - (h) determining which expressed immunoglobulin variable domain encoding sequences bind to the target antigen.

29. The method according to any one of claims 26-28, wherein the method further comprises the step of (i) sequencing the immunoglobulin variable domain encoding sequences that bind the target antigen.

30. The method according to any one of claims 26-28, wherein the Type IIs restriction enzyme recognition sites of step (a) and step (b) are recognized by a different Type IIs restriction enzyme.

31. The method of claim 30, wherein the Type IIs restriction enzyme recognition sites are BsmBI recognition sites, BsaI recognition sites, FokI recognition sites or a combination thereof.

32. The method according to any one of claims 26-28, wherein the Acceptor Framework nucleic acid sequence is derived from a human gene sequence.
33. The method of claim 32, wherein the human sequence is a human heavy chain variable gene sequence or a sequence derived from a human heavy chain variable gene sequence.
34. The method of claim 33, wherein the human heavy chain variable gene sequence is selected from VH1-2, VH1-69, VH1-18, VH3-30, VH3-48, VH3-23, and VH5-51.
35. The method of claim 32, wherein the human sequence is a human kappa light chain variable gene sequence or a sequence derived from a human kappa light chain variable gene sequence.
36. The method of claim 35, wherein the human kappa light chain variable gene sequence is selected from VK1-33, VK1-39, VK3-11, VK3- 15, and VK3-20
37. The method of claim 32, wherein the human sequence is a human lambda light chain variable gene sequence or a sequence derived from a human lambda light chain variable gene sequence.
38. The method of claim 37, wherein the human lambda light chain variable gene sequence is selected from VL1-44 and VL1-51.
39. The method of claim 26, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR3 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.
40. The method of claim 26, wherein the plurality of diversified nucleic acids encodes CDR3 regions, and wherein the plurality of diversified nucleic acids comprises or is derived

from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

41. The method of claim 26, wherein the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR3 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

42. The method of claim 27, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR1 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

43. The method of claim 27, wherein the plurality of diversified nucleic acids encodes CDR1 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

44. The method of claim 27, wherein the plurality of diversified nucleic acids encodes amino acid sequences that can fulfill the role of a CDR1 region, and wherein the plurality of diversified nucleic acids comprises synthetic sequences.

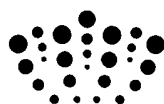
45. The method of claim 28, wherein the plurality of diversified nucleic acids comprises or is derived from sequences selected from naturally occurring CDR2 sequences, naturally occurring Ig sequences from humans, naturally occurring Ig sequences from a mammal, naturally occurring sequences from a loop region of a T cell receptor in a mammal, and other naturally diversified polypeptide collections.

46. The method of claim 28, wherein the plurality of diversified nucleic acids encodes CDR2 regions, and wherein the plurality of diversified nucleic acids comprises or is derived from immunoglobulin sequences that occur naturally in humans that have been exposed to a particular immunogen or sequences derived from animals that have been identified as having been exposed to a particular antigen.

47. The method according to any one of claims 26-28, wherein the plurality of Acceptor Framework nucleic acid sequences comprises a mixture of at least one variable heavy chain (VH) Acceptor Framework nucleic acid sequence and at least one variable light chain Acceptor Framework nucleic acid sequence.

48. The method according to any one of claims 26-28 wherein the expression vector is a phagemid vector.

49. The method according to any one of claims 26-28, wherein the host cell is *E. coli*.



**Application No:** GB0912754.9

**Examiner:** Dr Graham Feeny

**Claims searched:** 1 to 49

**Date of search:** 29 September 2009

**Patents Act 1977: Search Report under Section 17**

**Documents considered to be relevant:**

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
Y	1 to 49	Journal of Molecular Biology 376 (2008) Rothe et al 'The Human Combinatorial Antibody Library HuCAL GOLD Combines Diversification of All Six CDRs According to the Natural Immune System with a Novel Display Method for Efficient Selection of High-Affinity Antibodies' 1182 - 1200 The whole document, see especially the abstract, the introduction, the Results: 'Generation of HuCALs with six CDRs diversified' at p1185-p1188, figure 3 on p1188, 'Cloning strategies' p1192-1193.
Y	1 to 49	Trends in Biotechnology 23 (2005) Lu 'Seamless cloning and gene fusion' 199 - 207 The whole document, especially see p199 column 2 middle paragraph, table 1: 'Type IIS restriction....', figure 3 on p204.
Y	1 to 49	US2006/0222654 A1 (DELCAIRE) See [0185].

**Categories:**

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

**Field of Search:**

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC<sup>X</sup> :

--

Worldwide search of patent documents classified in the following areas of the IPC

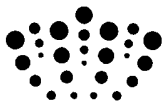
C07K; C40B
------------

The following online and other databases have been used in the preparation of this search report

BIOSIS, MEDLINE, XPESP , EPODOC, WPI, TXTJP, TXTWOT
---

**International Classification:**

Subclass	Subgroup	Valid From
C07K	0016/46	01/01/2006
C07K	0016/00	01/01/2006



<b>Subclass</b>	<b>Subgroup</b>	<b>Valid From</b>
C40B	0040/08	01/01/2006