Mosaic HIV-1 Vaccines Expand the Breadth and Depth of Cellular Immune Responses in Rhesus Monkeys

Dan H. Barouch[1,2], Kara L. O'Brien[1], Nathaniel L. Simmons[1], Sharon L. King[1], Peter Abbink[1], Lori F. Maxfield[1], Ying-Hua Sun[1], Annalena La Porte[1], Ambryice M. Riggs[1], Diana M. Lynch[1], Sarah L. Clark[1], Katherine Backus[1], James R. Perry[1], Michael S. Seaman[1], Angela Carville[3], Keith G. Mansfield[3], James J. Szinger[4], Will Fischer[4], Mark Muldoon[5,6], Bette Korber[4,6]

[1]Division of Vaccine Research, Beth Israel Deaconess Medical Center, Boston, MA; [2]Ragon Institute of MGH, MIT, and Harvard, Boston, MA; [3]New England Primate Research Center, Southborough, MA; [4]Los Alamos National Laboratory, Los Alamos, NM; [5]School of Mathematics, University of Manchester, Manchester, UK; [6]Santa Fe Institute, Santa Fe, NM

Correspondence should be addressed to D.H.B. (dbarouch@bidmc.harvard.edu)

SUPPLEMENTARY FIGURES

Supplementary Figure 1. CD8+ T lymphocyte responses to PTE peptides at week 4 following immunization mapped onto HIV-1 Gag, Pol, and Env protein sequences. Colors denote monkeys that received the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. For each epitope, the monkey number, antigen (G, Gag; P, Pol; E, Env), subpool number, and individual PTE peptide number are indicated.

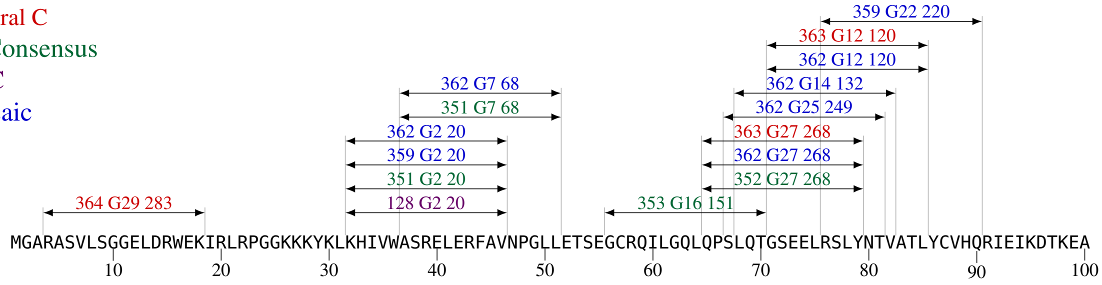
Supplementary Figure 2. CD4+ T lymphocyte responses to PTE peptides at week 4 following immunization mapped onto HIV-1 Gag, Pol, and Env protein sequences. Colors denote monkeys that received the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens. For each epitope, the monkey number, antigen (G, Gag; P, Pol; E, Env), subpool number, and individual PTE peptide number are indicated.

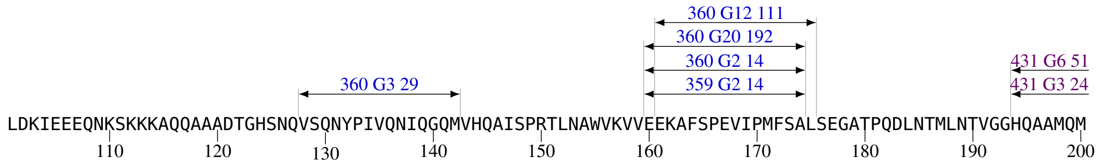
Supplementary Figure 3. Alignment of vaccine sequences with reactive PTE peptides in all monkeys at week 4 following immunization with rAd26 vectors expressing mosaic, M consensus, clade B + clade C, or optimal natural clade C HIV-1 Gag, Pol, and Env antigens. For each monkey, vaccine sequences are shown on the top, and reactive PTE peptides are shown beneath the vaccine sequences denoted by the antigen (G, Gag; P, Pol; E, Env) and PTE peptide numbers. The minimal overlap region is shown in bold. Sequence polymorphisms between the two mosaic or the two clade B + clade C antigens are shown in blue. Differences between the vaccine sequences and the reactive PTE peptides are shown in red.

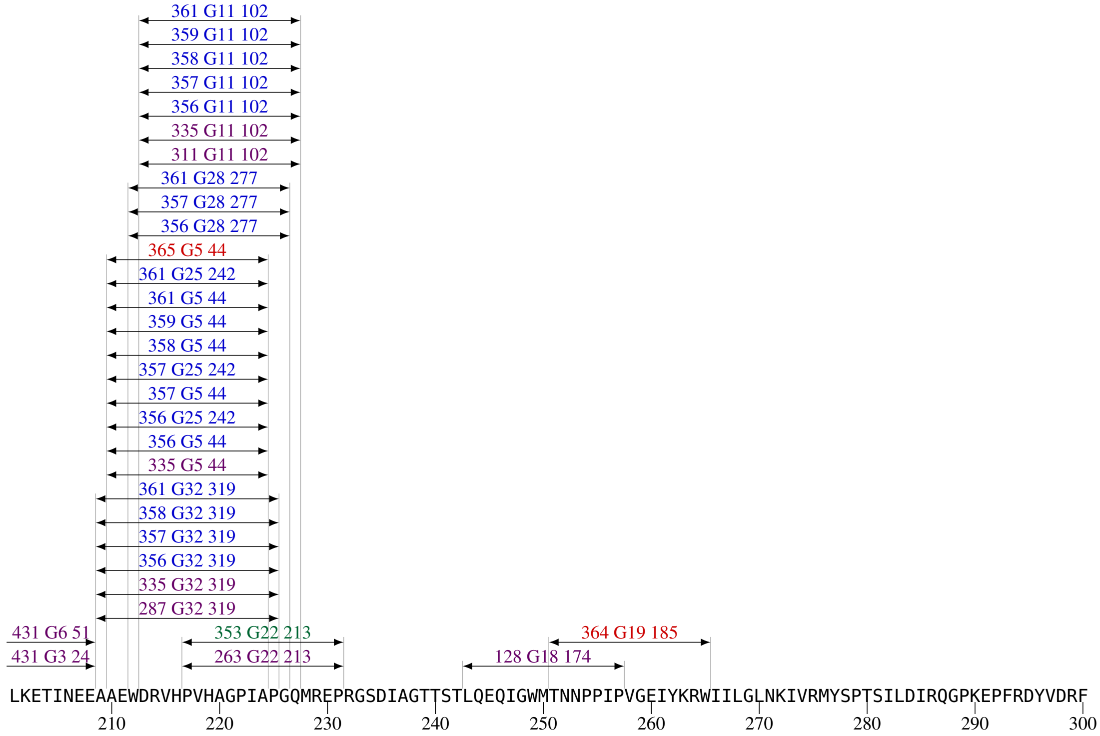
Supplementary Figure 4. Theoretical coverage of PTE peptides by the various vaccine antigens. Percentage of 9 amino acid PTE peptides that are covered by the mosaic (blue), M consensus (green), clade B + clade C (purple), or optimal natural clade C (red) HIV-1 Gag, Pol, and Env antigens are shown.

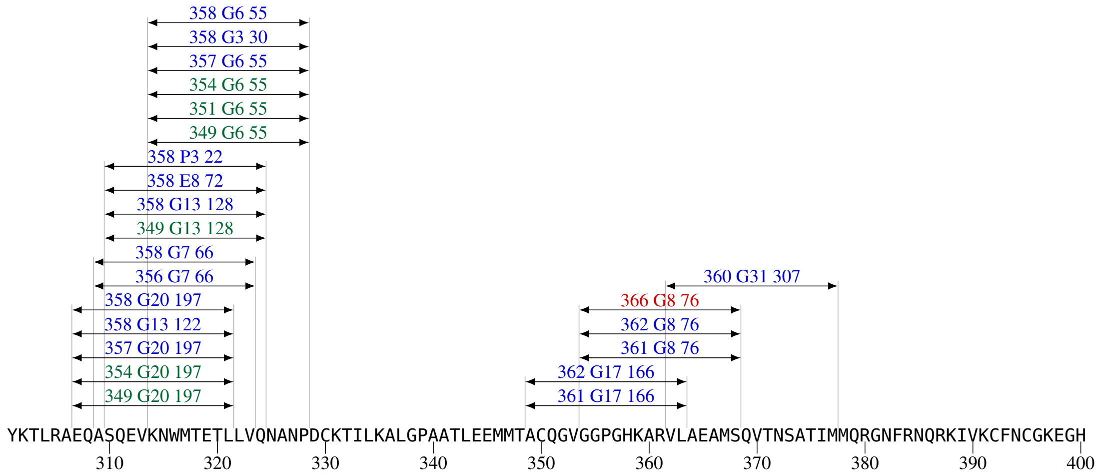
CD8 Gag

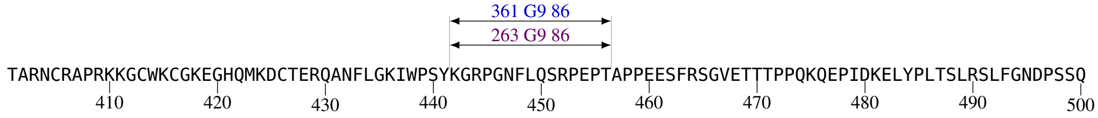
Natural C
M Consensus
B+C
Mosaic



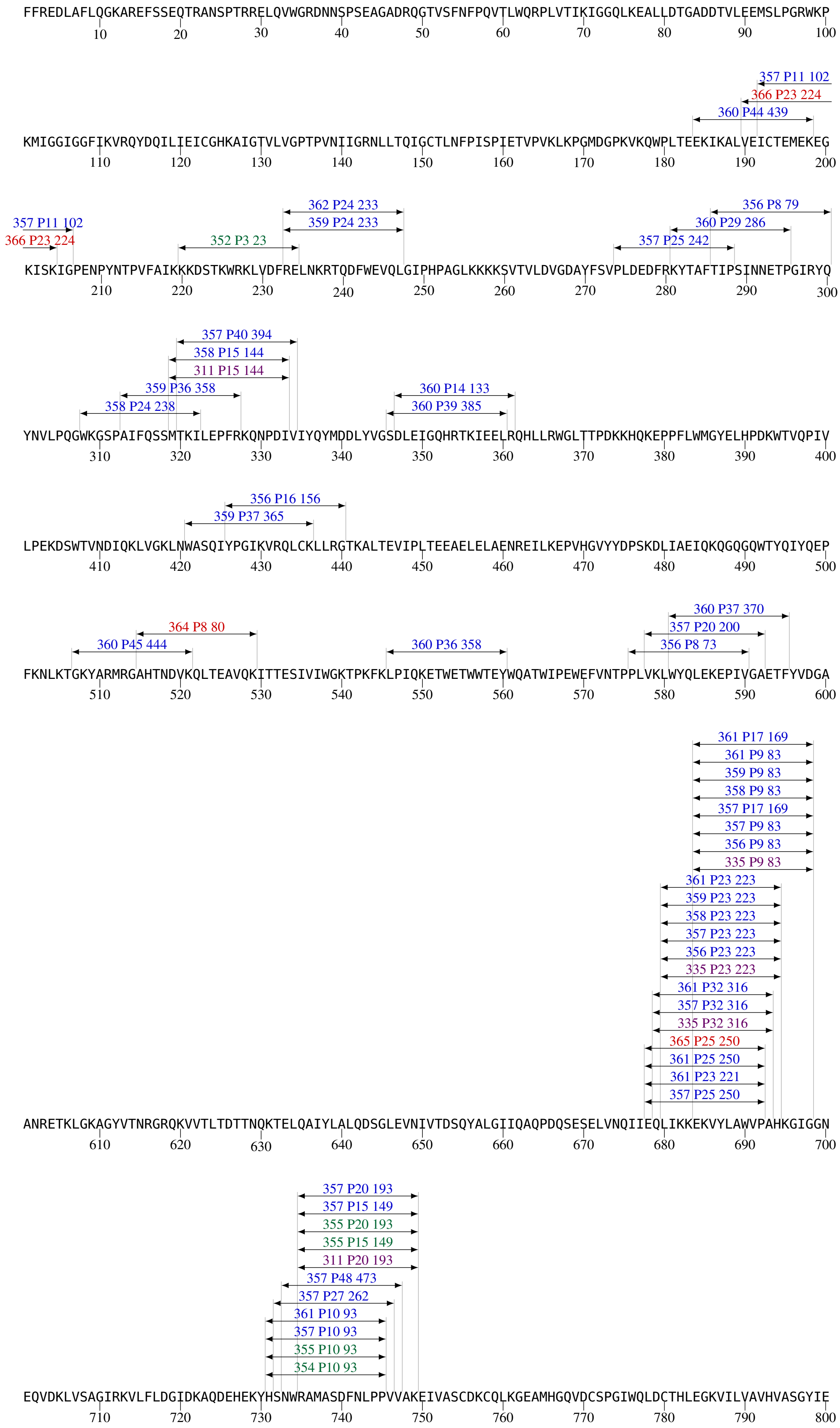


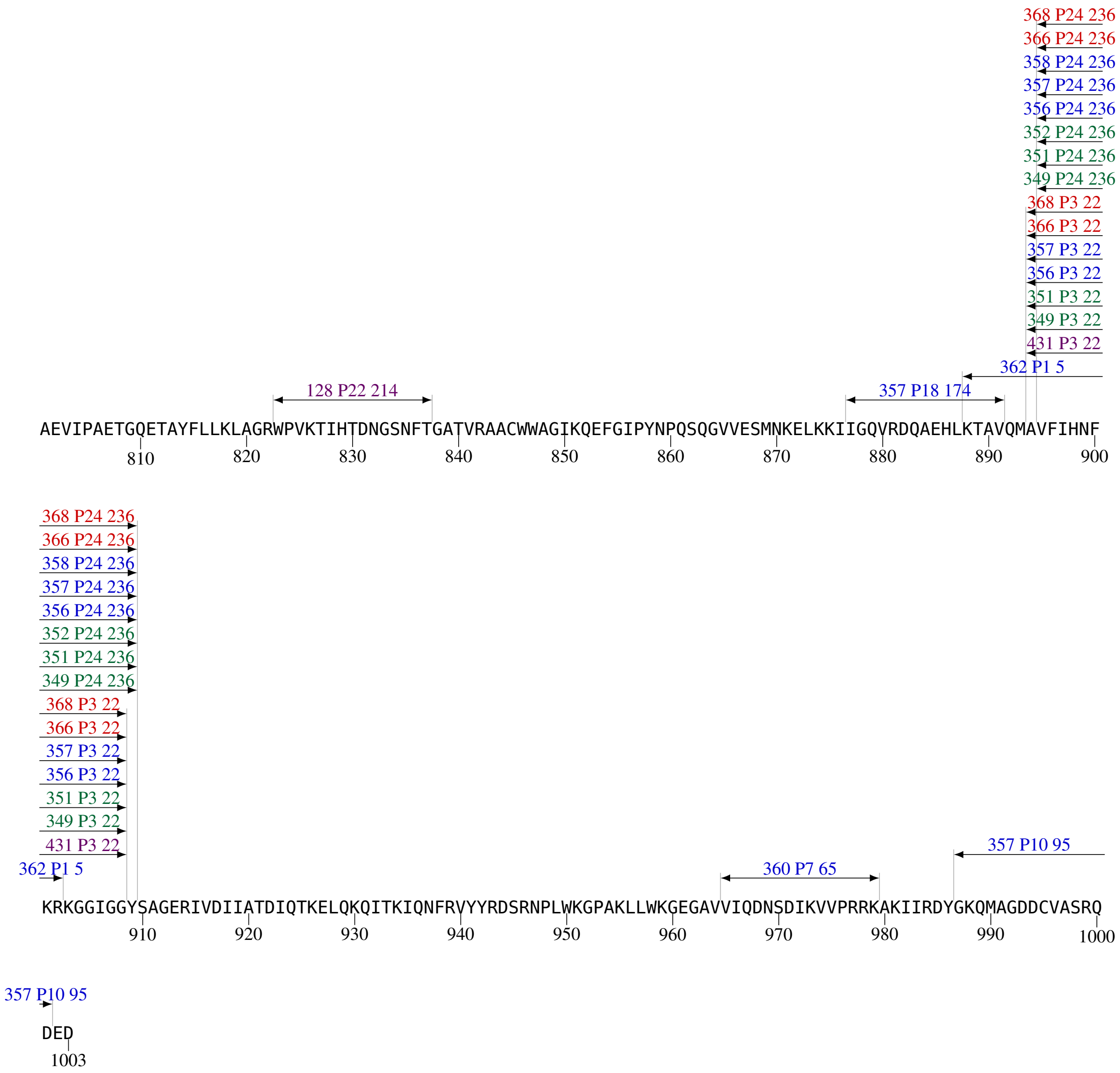




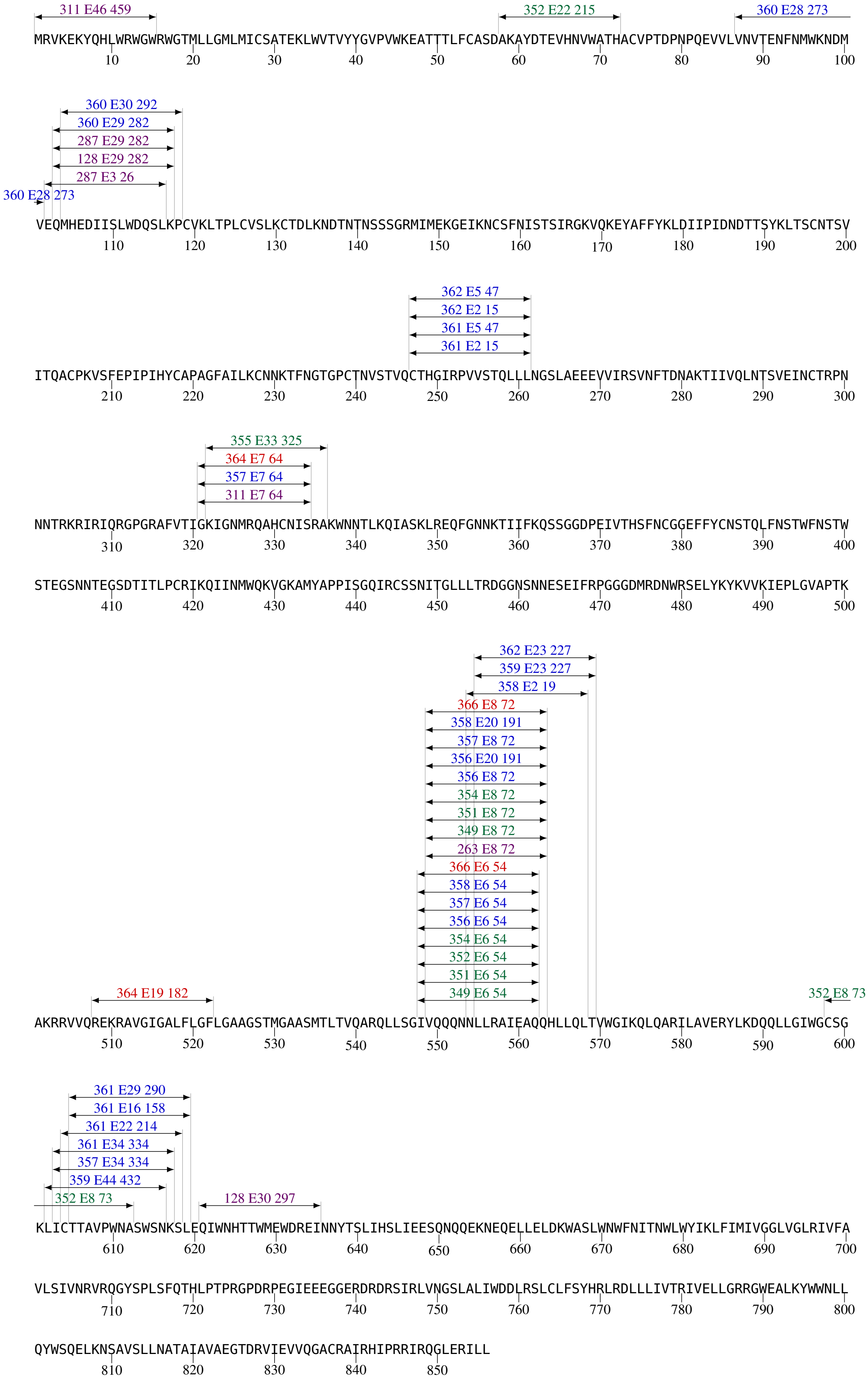


CD8 Pol


FFREDLAFLQGKAREFSSEQTRANSPTRRELQVWGRDNNSPSEAGADRQGTVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKP
10 20 30 40 50 60 70 80 90 100
357 P11 102
366 P23 224
360 P44 439
KMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEG
110 120 130 140 150 160 170 180 190 200
362 P24 233
359 P24 233
356 P8 79
360 P29 286
357 P11 102
366 P23 224
352 P3 23
357 P25 242
KISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQ
210 220 230 240 250 260 270 280 290 300
357 P40 394
358 P15 144
311 P15 144
359 P36 358
360 P14 133
358 P24 238
360 P39 385
YNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIV
310 320 330 340 350 360 370 380 390 400
356 P16 156
359 P37 365
LPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP
410 420 430 440 450 460 470 480 490 500
360 P37 370
364 P8 80
357 P20 200
360 P45 444
360 P36 358
356 P8 73
FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGA
510 520 530 540 550 560 570 580 590 600
361 P17 169
361 P9 83
359 P9 83
358 P9 83
357 P17 169
357 P9 83
356 P9 83
335 P9 83
361 P23 223
359 P23 223
358 P23 223
357 P23 223
356 P23 223
335 P23 223
361 P32 316
357 P32 316
335 P32 316
365 P25 250
361 P25 250
361 P23 221
357 P25 250
ANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGN
610 620 630 640 650 660 670 680 690 700
357 P20 193
357 P15 149
355 P20 193
355 P15 149
311 P20 193
357 P48 473
357 P27 262
361 P10 93
357 P10 93
355 P10 93
354 P10 93
EQVDKLVSAGIRKVLFLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIE
710 720 730 740 750 760 770 780 790 800


368 P24 236
366 P24 236
358 P24 236
357 P24 236
356 P24 236
352 P24 236
351 P24 236
349 P24 236
368 P3 22
366 P3 22
357 P3 22
356 P3 22
351 P3 22
349 P3 22
431 P3 22
362 P1 5
128 P22 214
357 P18 174
AEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNF
810
820
830
840
850
860
870
880
890
900
368 P24 236
366 P24 236
358 P24 236
357 P24 236
356 P24 236
352 P24 236
351 P24 236
349 P24 236
368 P3 22
366 P3 22
357 P3 22
356 P3 22
351 P3 22
349 P3 22
431 P3 22
362 P1 5
360 P7 65
357 P10 95
KRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQ
910
920
930
940
950
960
970
980
990
1000
357 P10 95
DED
1003

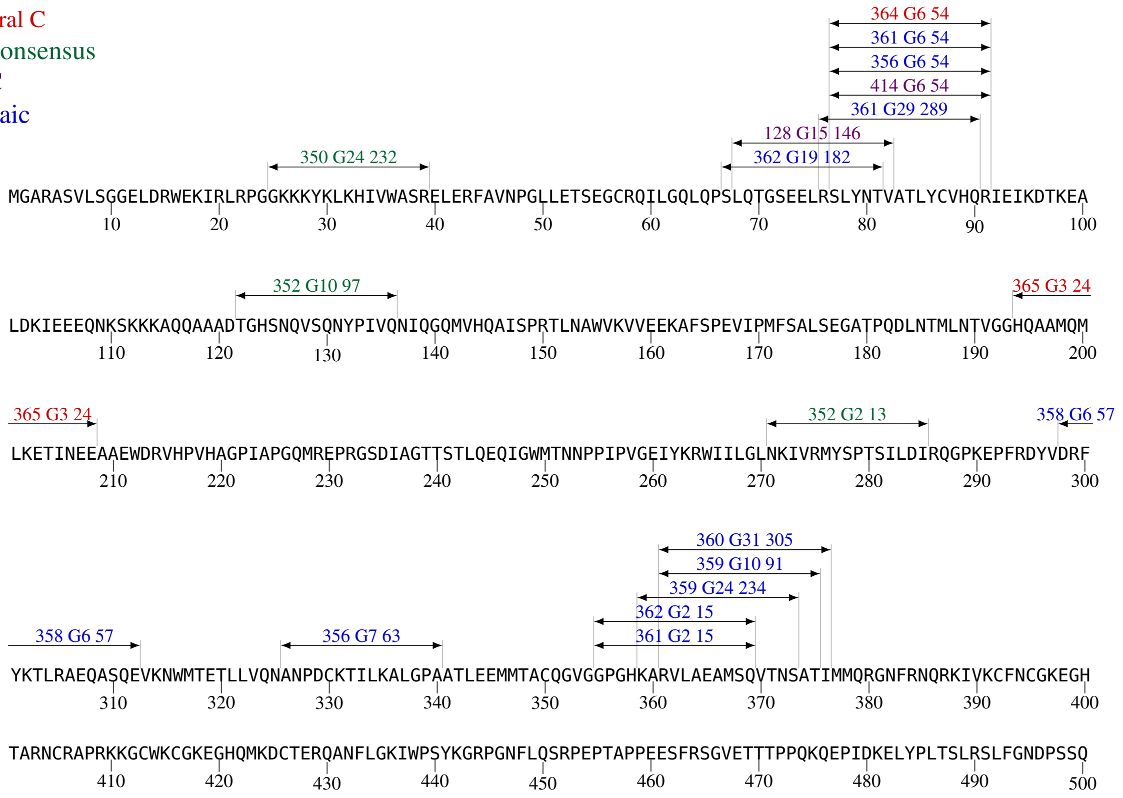
CD8 Env



```
MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDM
         10        20        30        40        50        60        70        80        90       100

VEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSV
         110       120       130       140       150       160       170       180       190       200

ITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDNAKTIIVQLNTSVEINCTRPN
         210       220       230       240       250       260       270       280       290       300

NNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTW
         310       320       330       340       350       360       370       380       390       400

STEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTK
         410       420       430       440       450       460       470       480       490       500

AKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSG
         510       520       530       540       550       560       570       580       590       600

KLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFA
         610       620       630       640       650       660       670       680       690       700

VLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
         710       720       730       740       750       760       770       780       790       800

QYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
         810       820       830       840       850
```

CD4 Gag

Natural C
M Consensus
B+C
Mosaic



MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEA

LDKIEEEQNKSKKKAQQAAADTGHSNQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKAFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQM

LKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF

YKTLRAEQASQEVKNWMTETLLVQNANPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGH

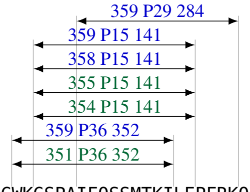
TARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLQSRPEPTAPPEESFRSGVETTTPPQKQEPIDKELYPLTSLRSLFGNDPSSQ

CD4 Pol

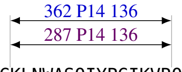
```
FFREDLAFLQGKAREFSSEQTRANSPTRRELQVWGRDNNSPSEAGADRQGTVSFNFPQVTLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKP
        10        20        30        40        50        60        70        80        90       100
```

```
KMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEG
       110       120       130       140       150       160       170       180       190       200
```

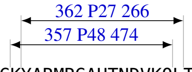
```
KISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQ
       210       220       230       240       250       260       270       280       290       300
```



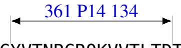
```
YNVLPQGWKGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIV
       310       320       330       340       350       360       370       380       390       400
```



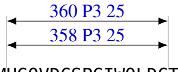
```
LPEKDSWTVNDIQKLVGKLNWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQKQGQGQWTYQIYQEP
       410       420       430       440       450       460       470       480       490       500
```



```
FKNLKTGKYARMRGAHTNDVKQLTEAVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIVGAETFYVDGA
       510       520       530       540       550       560       570       580       590       600
```

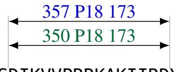


```
ANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYALGIIQAQPDQSESELVNQIIEQLIKKEKVYLAWVPAHKGIGGN
       610       620       630       640       650       660       670       680       690       700
```



```
EQVDKLVSAGIRKVLFLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKCQLKGEAMHGQVDCSPGIWQLDCTHLEGKVILVAVHVASGYIE
       710       720       730       740       750       760       770       780       790       800
```

```
AEVIPAETGQETAYFLLKLAGRWPVKTIHTDNGSNFTGATVRAACWWAGIKQEFGIPYNPQSQGVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNF
       810       820       830       840       850       860       870       880       890       900
```



```
KRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYYRDSRNPLWKGPAKLLWKGEGAVVIQDNSDIKVVPRRKAKIIRDYGKQMAGDDCVASRQ
       910       920       930       940       950       960       970       980       990      1000
```

```
DED
 1003
```

CD4 Env

```
                                                                                            349 E5 44
                                                                                           <---------
                                                                                            431 E5 44
                                                                                           <---------
                                                                                            128 E5 44
                                                                                           <---------
MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWKNDM
        10        20        30        40        50        60        70        80        90       100

349 E5 44
----->
431 E5 44
----->
128 E5 44
----->
VEQMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFNISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSV
       110       120       130       140       150       160       170       180       190       200

ITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEEVVIRSVNFTDNAKTIIVQLNTSVEINCTRPN
       210       220       230       240       250       260       270       280       290       300

                      357 E42 412
                   <------------>
NNTRKRIRIQRGPGRAFVTIGKIGNMRQAHCNISRAKWNNTLKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFNSTW
       310       320       330       340       350       360       370       380       390       400

                                              364 E32 316                           349 E5 42
                                            <-------------->                    <-------------->
STEGSNNTEGSDTITLPCRIKQIINMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTK
       410       420       430       440       450       460       470       480       490       500

AKRRVVQREKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQLLGIWGCSG
       510       520       530       540       550       560       570       580       590       600

                                                     357 E7 68
                                                  <-------------->
       361 E6 53                                263 E5 49
   <-------------->                         <-------------->
KLICTTAVPWNASWSNKSLEQIWNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVGGLVGLRIVFA
       610       620       630       640       650       660       670       680       690       700

VLSIVNRVRQGYSPLSFQTHLPTPRGPDRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCLFSYHRLRDLLLIVTRIVELLGRRGWEALKYWWNLL
       710       720       730       740       750       760       770       780       790       800

QYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQGACRAIRHIPRRIRQGLERILL
       810       820       830       840       850
```

Supplementary Figure 3. Alignments of vaccine sequences with reactive PTE peptides

Minimal regions within the peptides that are likely to contain the immune response epitope, based on overlap between reactive peptides when it occurs, are in bold in the vaccines. If there is no overlapping peptide, we assume the epitope can be anywhere in the peptide, so the whole region is bold. We cannot differentiate between different T cell responses targeting epitopes with different boundaries within a peptide, or more promiscuous clonal T cell responses that can tolerate variation when variants are present; either scenario could be advantageous in a vaccine immune response. The number of targeted regions corresponds to the minimum number of T cell responses required to account for the data.

Amino acids where the vaccine and the peptides don't match are written in red; if they fall within the region likely to carry the epitope, they are bold red. Amino acid differences outside of the overlapping regions when multiple peptides overlap are marked in red, but not bold.

The vaccines are always at the top. The letter for each protein (Gag is G, Pol is P, Envelope is E) and the peptide number are used to label for each reactive PTE peptide. The protein and HXB2 numbers follow each peptide.

For the mosaic and clade B+C vaccines, there are 2 antigens each and both are included in the alignment; amino acid differences in the vaccines are noted in blue, and if the reactive peptide carries the variant amino acid in the second mosaic, it is also in blue. In each of the positions where the two vaccine antigens differ, the reactive peptides are also marked in bold to indicate the positions where including two variants may have impacted the vaccine immune response and allowed greater breadth and depth.

For example, the first vaccine summarized is the clade B+C vaccine, and animal 287-95 is the first animal for which responses are listed. There were 3 CD8 responses to PTE peptides, 1 to CD4. Two of the CD8 peptides show substantial overlap, E26 and E282, so both may be targets for the same CTL response; thus we also note there are only 2 CD8 responsive regions, and 1 CD4 responsive regions. For each responsive region, we write out the number of overlapping peptides per region (e.g. CD8: 1 2 CD4: 1) to assess depth of responses; the two is red to indicate that the region of overlap is variable in the reactive peptides. If the vaccine differs, like the D/E in the second reactive region, it is marked in blue. Only the region of overlap is bold. The H in E282 was not found in either vaccine so it is marked with red; it is within the region of overlap so it is bold. Each reactive peptide has its protein and corresponding HXB2 numbering noted on the right.

Reactive peptides and breadth and depth of responses: Clade B+C Vaccine

Clade B+C 287-95:
3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 1 2 CD4: 1

3 CD8 responses:
C   **AAEWDRLHPVHAGPIA**
B   **AAEWDRLHPVHAGPIA**
G319 AAE**-**DRLHPVHAGPI**P** Gag 209 225

C   D**QMHEDIISLWDQSL**K
B   E**QMHEDIISLWDQSL**K
E26 EQMHEDIISLWDQSL Env 102 116
E282 **H**MHEDIISLWDESLK Env 103 117

1 CD4 response:

```
C     KLNWASQIYSGIKVR
B     KLNWASQIYPGIKVR
P136  KLNWASQIYAGIKVK  Pol   418   432
```

Clade B+C 128-92:
5 CD8 PTE+, 2 CD4 PTE+
5 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 1 1

5 CD8 responses:

```
C     KHLVWASRELERFAL
B     KHIVWASRELERFAV
G20   KHLVWASRELERFAL Gag   32    46

C     LQEQIAWMTNNPPVP
B     LQEQIGWMTNNPPIP
G174  LAEQIAWMTSNPPIP Gag   243   257

C     QMHEDIISLWDQSLK
B     QMHEDIISLWDQSLK
E282  HMHEDIISLWDESLK  Env   103   117

C     DIWDNMTWMQWDREI
B     EIWDNMTWMEWEREI
E297  EIWNNMTWMEWEKEI  Env   621   635

C     WPVKVIHTANGSNFT
B     WPVKTIHTANGSNFT
P214  WPVKVVHTDNGSNFT Pol   823   837
```

2 CD4 responses:

```
C     AQTGTEELRSLYNTV
B     LQTGSEELRSLYNTV
G146  LQTGSEELKSLFNTV  Gag   68    82

C     YFNMWKNDMVDQMHE
B     NFNMWKNNMVEQMHE
E544  NFNMWKNNMVEQMHE  Env   92    106
```

Clade B+C 263-00:
3 CD8 PTE+, 1 CD4 PTE+
3 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 1 1 1 CD4: 1

3 CD8 responses:

C **GRPGNFLQSRPEPT**
B **GRPGNFLQSRPEPT**
G86 GRPGNFLQ**N**RPEPT Gag 442 456

C **PVHAGPIAPGQMREP**
B **PVHAGPIAPGQMREP**
G213 PV**Q**AGPIAPGQMREP Gag 217 231

C **VQQQSNLLRAIEAQQ**
B **VQQQNNLLRAIEAQQ**
E72 VQQQ**N**NLLRAIEAQ**H** Env 549 563

1 CD4 response:
C **LLEESQNQQEKNEKD**
B **LIEESQNQQEKNEQE**
E49 L**I**EESQNQQEKNE**QD** Env 645 659

Clade B+C 311-00:
5 CD8 PTE+, 0 CD4 PTE+
5 CD8+ regions, 0 CD4+ region
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 0

5 CD8 responses:
C **DRLHPVHAGPIAPGQ**
B **DRLHPVHAGPIAPGQ**
G102 DR**V**HPVHAGPI**P**PGQ Gag 213 227

C **GDIIGDIRQAHCNIS**
B **GEIIGDIRQAHCNIS**
E64 G**D**IIGDIRQAHCNIS Env 321 334

C **TGMLRNCQPWWIWGI**
B **KGIRKNYQHLWRWGT**
E459 **KEIRK**N**YQHL**W**R**WG**T** Env 4 18

C **MTKILEPFRAKNPEI**
B **MTKILEPFRKQNPDI**
P144 MTKILEPFR**KQ**NP**D**I Pol 319 333

C **RAMASEFNLPPVVAK**
B **RAMASDFNLPPVVAK**
P193 **K**AMAS**D**FNLPPIVAK Pol 735 749

Clade B+C 335-96:
6 CD8 PTE+, 0 CD4 PTE+
2 CD8+ regions, 0 CD4+ region

Number of overlapping peptides per region: CD8: 3 3 CD4: 0

6 CD8 responses:

```
C    AAEWDRLHPVHAGPIAPGQ
B    AAEWDRLHPVHAGPIAPGQ
G319 AAE-DRLHPVHAGPIP       Gag   209   225
G44   AEWDRLHPVHAGPIA      Gag   210   224
G102     DRVHPVHAGPIPPGQ   Gag   213   227
```

```
C    QLINKERVYLSWVPAHKGIG
B    QLIKKEKVYLAWVPAHKGIG
P316 KLIEKDKVYLSWVPA       Pol   679   693
P223  LIKKERVYLSWVPAH      Pol   680   694
P83       EKVYLSWVPAHKGIG  Pol   684   698
```

Clade B+C 414-95:

0 CD8 PTE+, 1 CD4 PTE+
0 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 0 CD4: 1

1 CD4 responses:

```
C    SLYNTVATLYCVHAG
B    SLYNTVATLYCVHQK
G54  SLYNTVATLYCVHQR Gag   77    91
```

Clade B+C 431-01:

3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ region
Number of overlapping peptides per region: CD8: 2 1 CD4: 1

3 CD8 responses:

```
C    HQAAMQMLKDTINEE
B    HQAAMQMLKETINEE
G24  HQAAMQMLKDTINEE Gag   194   208
G51  HQAAMQMLKETINEE Gag   194   208
```

```
C    AVFIHNFKRKGGIGG
B    AVFIHNFKRKGGIGG
P22  AVFIHNFKRKGGIGG Pol   894   908
```

1 CD4 response:

```
C    YFNMWKNDMVDQMHE
B    NFNMWKNNMVEQMHE
E44  NFNMWKNNMVEQMHE Env   92    106
```

Reactive peptides and breadth and depth of responses: M Consensus Vaccine

M consensus 349-07:
7 CD8 PTE+, 2 CD4 PTE+
3 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 3 2 2 CD4: 1 1

3 CD8 responses:

```
Mcon  EQATQDVKNWMTDTLLVQNANP
G197  EQSTQEVKNWMTDTL            Gag   307   321
G128  ---TQDVKNWMTDTLLIQ         Gag   310   324
G55   -------KNWMTDTLLVQNANP     Gag   314   328

Mcon  IVQQQSNLLRAIEAQQ
E54   VVQQQSNLLRAIEAQ            Env   548   562
E72   -VQQQNNLLRAIEAQH           Env   549   563

Mcon  AVFIHNFKRKGGIGGY
P22   AVFIHNFKRKGGIGG            Pol   894   908
P236  -VLIHNFKRKGGIGGY           Pol   895   909
```

2 CD4 responses:

```
Mcon  SELYKYKVVKIEPLG
E42   SELYKYKVVKIEPLG            Env   481   495

Mcon  NFNMWKNNMVEQMHE
E44   NFNMWKNNMVEQMHE            Env   92    106
```

M consensus 350-07:
0 CD8 PTE+, 2 CD4 PTE+
0 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: none CD4: 1 1

2 CD4 responses:

```
Mcon  GKKKYRLKHLVWASR
G232  GRKKYRLKHIVWASR            Gag   25    39

Mcon  DIKVVPRRKAKIIRD
P173  DIKVVPRRKVKIIRD            Pol   971   985
```

M consensus 351-07:
7 CD8 PTE+, 1 CD4 PTE+
4 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 2 1 CD4: 1

4 CD8 responses:

Mcon	KHLVW**ASRELERFAL**NPGLL			
G20	KHLVWASRELERFAL	Gag	32	46
G68	-----ASRELERFA**V**NPGLL	Gag	37	5

Mcon	**IVQQQSNLLRAIEAQ**Q			
E6	VVQQQSNLLRAIEAQ	Env	548	562
E8	-VQQQ**N**NLLRAIEAQH	Env	549	563

Mcon	A**VFIHNFKRKGGIGG**Y			
P22	AVFIHNFKRKGGIGG	Pol	894	908
P236	-V**L**IHNFKRKGGIGGY	Pol	895	909

Mcon	**KNWMTDTLLVQNANP**			
G55	KNWMTDTLLVQNANP	Gag	314	328

1 CD4 response:

Mcon	**WKGSPAIFQSSMTKI**			
P352	WKGSPAIFQ**A**SMTKI	Pol	308	322

M consensus 352-07:
6 CD8 PTE+, 2 CD4 PTE+
6 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 1 CD4: 1

6 CD8 responses:

Mcon	**QPALQTGSEELRSLY**			
G268	**L**PAL**K**TGSEELRSLY	Gag	65	79

Mcon	**IVQQQSNLLRAIEAQ**			
E54	**V**VQQQSNLLRAIEAQ	Env	548	562

Mcon	**CSGKLICTTTVPWNS**			
E73	CSGKLICTT**A**VPWNS	Env	598	612

Mcon	**AKAYDTEVHNVWATH**			
E215	AKAY**EK**EVHNVWATH	Env	58	72

Mcon	**KKDSTKWRKLVDFRE**			
P23	KKDSTKWRKLVDFRE	Pol	220	234

Mcon	**VFIHNFKRKGGIGGY**			
P236	V**L**IHNFKRKGGIGGY	Pol	895	909

2 CD4 responses:

```
Mcon  NKIVRMYSPVSILDI
G13   NKIVRMYSPVSILDI       Gag   271   285

Mcon  KGNSSKVSQNYPIVQ
G97   TGNSSQVSQNYPIVQ       Gag   122   136
________________________________________
M consensus 353-07:
2 CD8 PTE+, 0 CD4 PTE+
2 CD8+ regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1  CD4: none

2 CD8 responses:
Mcon  GCKQIIGQLQPALQT
G151  GCRQILGQLQPSLQT       Gag   56    70

Mcon  PVHAGPIPPGQMREP
G213  PVQAGPIAPGQMREP       Gag   217   231
________________________________________
M consensus 354-07:
5 CD8 PTE+, 1 CD4 PTE+
3 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1  CD4: 1

3 CD8 responses:
Mcon  EQATQDVKNWMTDTLLVQNANP
G197  EQSTQEVKNWMTDTL            Gag   307      321
G55   -------KNWMTDTLLVQNANP     Gag   314      328

Mcon  IVQQQSNLLRAIEAQQ
E54   VVQQQSNLLRAIEAQ            Env   548      562
E72   -VQQQNNLLRAIEAQH           Env   549      563

Mcon  HSNWRAMASDFNLPP
P93   HSNWRAMASDFNLPP            Pol   731      745

1 CD4 response:
Mcon  GSPAIFQSSMTKILE
P141  GSPAIFQSSMTKILD            Pol   310      324
________________________________________
M consensus 355-07:
3 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 1 3 CD4: 1

2 CD8 responses:
Mcon  IIGDIRQAHCNISGT
```

E325	I**T**GDIRQAHCN**V**S**RS**	Env	322	336

Mcon	SNW**RAMASDFNLPP**IVAK			
P93	SNWRAMASDFNLPP	Pol	731	745
P149	---R**T**MASDFNLPPVVAK	Pol	735	749
P193	---**K**AMASDFNLPPIVAK	Pol	735	749

1 CD4 response:

Mcon	**GSPAIFQSSMTKILE**			
P141	GSPAIFQSSMTKIL**D**	Pol	310	324

Reactive peptides and breadth and depth of responses: Mosaic Vaccine

Mosaic 356-07:
15 CD8 PTE+, 2 CD4 PTE+
8 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 5 3 1 2 1 1 1 2 CD4: 1 1

8 CD8 responses:

Mos1	AAEW**DRVHPVHAGPIA**PGQ			
Mos2	AAEW**DR**L**HPVHAGP**V**A**PGQ			
G319	AAE-DRLHPVHAGP**IP**	Gag	209	225
G44	-AEWDRLHPVHAGP**IA**	Gag	210	224
G242	-ADWDRLHPVHAGP**VA**	Gag	210	224
G277	---WDR**V**HPVHAGP**NP**PG	Gag	212	226
G102	----DR**V**HPVHAGP**IP**PGQ	Gag	213	227

Mos1	I**VQQQNNLLRAIEAQ**Q			
Mos2	I**VQQQSNLLRAIEAQ**Q			
E54	VVQQQ**S**NLL**R**AIEAQ	Env	548	562
E72	-VQQQ**N**NLL**R**AIEAQH	Env	549	563
E191	-VQQQ**S**NLL**K**AIEAQQ	Env	549	563

Mos1	**ASQDVKNWMTETLLV**			
Mos2	**A**T**QDVKNWMT**D**TLLV**			
G66	A**S**Q**E**VKNWMT**E**TLL**I**	Gag	309	323

Mos1	A**VFIHNFKRKGGIGG**Y			
Mos2	A**VFIHNFKRKGGIG**EY			
P22	AV**F**IHNFKRKGGIG**G**	Pol	894	908
P236	-V**L**IHNFKRKGGIG**G**Y	Pol	895	909

Mos1	**PLVKLWYQLEKDPIA**			
Mos2	**PLVKLWYQLEKEPI**V			

P73	PLVKLWYQLEKEPIV	Pol	576	590

Mos1	**TIPSTNNETPGIRYQ**			
Mos2	**TIPSINNETPGIRYQ**			
P79	TIPS**I**NNETPGIRYQ	Pol	286	300

Mos1	**YPGIKVRQLCKLLRG**			
Mos2	**YAGIKVKQLCKLLRG**			
P156	Y**P**GIKV**R**QLCKLLRG	Pol	426	440

Mos1	LIKK**ERVYLSWVPAH**KGIG			
Mos2	LIKK**EKVYLAWVPAH**KGIG			
P223	LIKKE**R**VYL**S**WVPAH	Pol	680	694
P83	----E**K**VYL**S**WVPAHKGIG	Pol	684	698

2 CD4 responses:

Mos1	**SLYNTVATLYCVHQR**			
Mos2	**SLFNTVATLYCVHAE**			
G54	SL**Y**NTVATLYCVH**QR**	Gag	77	91

Mos1	**ANPDCKTILKALGPA**			
Mos2	**ANPDCKTILRALGPG**			
G63	ANPDCKTIL**R**ALGP**G**	Gag	326	340

Mosaic 357-07:

29 CD8 PTE+, 4 CD4 PTE+

14 CD8+ regions, 4 CD4+ regions

Number of overlapping peptides per region: CD8: 5 2 2 5 5 2 1 1 1 1 1 1 1 1 CD4: 1 1 1 1

14 CD8 responses:

Mos1	EQLIKK**ERVYLSWVP**AHKGIG			
Mos2	EQLIKK**EKVYLAWVP**AHKGIG			
P250	EPLIKKE**K**VYL**S**WVP	Pol	678	692
P316	-KLIEK**DK**VYL**S**WVPA	Pol	679	693
P223	--LIKKE**R**VYL**S**WVPAH	Pol	680	694
P169	------E**K**VYL**A**WVPAHKGIG	Pol	684	698
P83	------E**K**VYL**S**WVPAHKGIG	Pol	684	698

Mos1	I**VQQQNNLLRAIEAQ**Q			
Mos2	I**VQQQSNLLRAIEAQ**Q			
E54	VVQQQ**S**NLLRAIEAQ	Env	548	562
E72	-VQQQ**N**NLLRAIEAQH	Env	549	563

Mos1	EQASQDV**KNWMTETL**LVQNANP			
Mos2	EQATQDV**KNWMTDTL**LVQNANP			
G197	EQSTQEVKNWMT**D**TL	Gag	307	321

G55	-------KNWMT**D**TLLVQNANP	Gag	314	328

Mos1	HSNW**RAMASDFNLPP**VVAK			
Mos2	HSNW**RAMASEFNLPP**IVAK			
P93	HSNWRAMAS**D**FNLPP	Pol	731	745
P262	-SNWRAMAS**E**FNLPPI	Pol	732	746
P473	--NWR**T**MAS**D**FNLPPVI	Pol	733	747
P149	----R**T**MAS**D**FNLPPVVAK	Pol	735	749
P193	----**K**AMAS**D**FNLPPIVAK	Pol	735	749

Mos1	AEW**DRVHPVHAGPIA**PGQ			
Mos2	AEW**DR**L**HPVHAGP**V**A**PGQ			
G319	AAEDRLHPVHAGP**IP**	Gag	209	225
G242	ADWDRLHPVHAGP**V**A	Gag	210	224
G44	AEWDRLHPVHAGP**I**A	Gag	210	224
G277	--WDR**V**HPVHAGP**NP**PG	Gag	212	226
G102	---DR**V**HPVHAGP**IP**PGQ	Gag	213	227

Mos1	A**VFIHNFKRKGGIGG**Y			
Mos2	A**VFIHNFKRKGGIG**EY			
P22	AVFIHNFKRKGGIG**G**	Pol	894	908
P236	-V**L**IHNFKRKGGIG**G**Y	Pol	895	909

Mos1	**PLDEGFRKYTAFTIP**			
Mos2	**PLDE**D**FRKYTAFTIP**			
P242	PLDE**S**FRKYTAFTIP	Pol	274	288

Mos1	**TRILEPFRAKNPEIV**			
Mos2	**T**K**ILEPFR**KQ**NP**D**IV**			
P394	T**K**ILEPFR**AQ**NP**EL**V	Pol	320	334

Mos1	**ICEEMEKEGKITKIG**			
Mos2	**IC**T**EMEKEGKI**S**KIG**			
P102	IC**T**EMEKEGKI**S**KIG	Pol	192	206

Mos1	**VKLWYQLEKDPIAGV**			
Mos2	**VKLWYQLEK**E**PI**V**GA**			
P200	VKLWYQLEK**D**PI**V**G**A**	Pol	578	592

Mos1	**GDIIGDIRQAHCNIS**			
Mos2	**GDIIGDIRQAHCN**L**S**			
E764	GDIIGDIRQAHCN**I**S	Env	321	334

Mos1	**ICTTTVPWNASWSNK**			
Mos2	**ICTT**A**VPWN**T**SWSNK**			
E334	ICTT**T**VPWN**A**SWSN**R**	Env	603	617

Mos1 **IGQVRDQAEHLKTAV**
Mos2 **IGQVRDQAEHLKTAV**
P174 IGQVR**E**QAEHLKTAV Pol 877 891

Mos1 **GKQMAGADCVAGRQD**
Mos2 **GKQMAGDDCVASRQD**
P95 GKQMAG**D**DCVA**G**RQD Pol 987 1001

4 CD4 responses:
Mos1 **KYAKMRTAHTNDVKQ**
Mos2 **KYARMRGAHTNDVKQ**
P474 KYA**K**MR**T**AHTNDV**R**Q Pol 508 522

Mos1 **AGDIIGDIRQAHCNI**
Mos2 **TGDIIGDIRQAHCNL**
E412 **T**G**E**IIGDIRQAHCN**L** Env 320 333

Mos1 **NQQEKNEKDLLALDS**
Mos2 **NQQEKNEQELLELDK**
E68 NQQEKNE**QD**LL**A**LD**K** Env 651 665

Mos1 **DIKVVPRRKVKIIKD**
Mos2 **DIKVVPRRKAKIIRD**
P173 DIKVVPRRK**V**KII**R**D Pol 971 985

Mosaic 358-07:
19 CD8 PTE+, 3 CD4 PTE+
7 CD8+ regions, 3 CD4+ regions
Number of overlapping peptides per region: CD8: 2 4 2 6 3 1 1 CD4: 1 1 1

7 CD8 responses:
Mos1 LIKK**ERVYLSWVPAH**KGIG
Mos2 LIKK**EKVYLAWVPAH**KGIG
P223 LIKKE**R**VYL**S**WVPAH Pol 680 694
P83 ----E**K**VYL**S**WVPAHKGIG Pol 684 698

Mos1 IVQQQN**NLLRAIEAQ**QHLLQL
Mos2 IVQQQS**NLLRAIEAQ**QHMLQL
E54 VVQQQSNLL**R**AIEAQ Env 548 562
E191 -VQQQSNLL**K**AIEAQQ Env 549 563
E72 -VQQQNNLL**R**AIEAQH Env 549 563
E19 ------NLL**R**AIEAQQHLLQL Env 554 568

Mos1 A**VFIHNFKRKGGIGG**Y
Mos2 A**VFIHNFKRKGGIGE**Y

P222	AV**F**IHNFKRKGGIG**G**	Pol	894	908
P236	-V**L**IHNFKRKGGIG**G**Y	Pol	895	909

Mos1	EQASQDV**KNWMTETL**LVQNANP			
Mos2	EQATQDV**KNWMTDTL**LVQNANP			
G122	EQASQDVKNWMT**E**TL	Gag	307	321
G197	EQSTQEVKNWMT**D**TL	Gag	307	321
G66	--ASQEVKNWMT**E**TLLI	Gag	309	323
G128	---TQDVKNWMT**D**TLLIQ	Gag	310	324
G30	-------KNWMT**E**TLLVQNANP	Gag	314	328
G55	-------KNWMT**D**TLLVQNANP	Gag	314	328

Mos1	AEW**DRVHPVHAGPIA**PGQ			
Mos2	AEW**DR**L**HPVHAGP**V**A**PGQ			
G319	AAEDR**L**HPVHAGP**IP**	Gag	209	225
G44	AEWDR**L**HPVHAGP**I**A	Gag	210	224
G11	---DR**V**HPVHAGP**IP**PGQ	Gag	213	227

Mos1	**WKGSPAIFQCSMTRI**			
Mos2	**WKGSPAIFQ**S**SMT**K**I**			
P238	WKGSPAIFQ**C**SMT**K**I	Pol	308	322

Mos1	**MTRILEPFRAKNPEI**			
Mos2	**MTKILEPFRKQNP**D**I**			
P144	MT**K**ILEPFR**KQ**NP**D**I	Pol	319	333

3 CD4 responses:

Mos1	**GSPAIFQCSMTRILE**			
Mos2	**GSPAIFQ**S**SMT**K**ILE**			
P141	GSPAIFQ**S**SMT**K**IL**D**	Pol	310	324

Mos1	**HGQVDCSPGIWQLAC**			
Mos2	**HGQVDCSPGIWQLAC**			
P25	HGQVDCSPGIWQL**D**C	Pol	766	780

Mos1	**DRFYKTLRAEQASQD**			
Mos2	**DRF**F**KTLRAEQA**T**QD**			
G57	DRF**Y**KTLRAEQA**S**Q**E**	Gag	298	312

Mosaic 359-07:
12 CD8 PTE+, 5 CD4 PTE+
10 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1 1 1 1 1 1 1 1 CD4: 3 2

10 CD8 responses:

Mos1 LIKK**ERVYLSWVPAH**KGIG

Mos2	LIKK**EKVYLAWVPAH**KGIG			
P223	LIKKE**R**VYL**S**WVPAH	Pol	680	694
P83	----E**K**VYL**S**WVPAHKGIG	Pol	684	698

Mos1	AEW**DRVHPVHAGPIA**PGQ			
Mos2	AEW**DRLHPVHAGPVA**PGQ			
G44	AEWDR**L**HPVHAGP**I**A	Gag	210	224
G102	---DR**V**HPVHAGP**I**PPGQ	Gag	213	227

Mos1	**AIFQCSMTRILEPFR**			
Mos2	**AIFQSSMTKILEPFR**			
P358	AIFQ**S**SMT**R**ILEPFR	Pol	313	327

Mos1	**ASQIYPGIKVRQLCK**			
Mos2	**ASQIYAGIKVKQLCK**			
P365	**W**SQIY**A**GIKV**R**QLCK	Pol	421	436

Mos1	**RELNKRTQDFWEVQL**			
Mos2	**RELNKRTQDFWEVQL**			
P233	RELN**R**RTQDFWEVQL	Pol	233	247

Mos1	**LLRAIEAQQHLLQLT**			
Mos2	**LLRAIEAQQHMLQLT**			
E227	LL**M**AIEAQQH**L**LQLT	Env	555	569

Mos1	**LICTTTVPWNASWSN**			
Mos2	**LICTTAVPWNTSWSN**			
E432	**H**ICTT**N**VPWN**A**SWSN	Env	602	616

Mos1	**KHIVWASRELERFAV**			
Mos2	**KHLVWASRELERFAL**			
G20	KH**L**VWASRELERFA**L**	Gag	32	46

Mos1	**RSLYNTVATLYCVHQ**			
Mos2	**RSLFNTVATLYCVHA**			
G220	**K**SL**Y**NTVAVLYCVH**Q**	Gag	76	90

Mos1	**EEKAFSPEVIPMFSA**			
Mos2	**EEKAFSPEVIPMFTA**			
G14	EEKAFSPEVIPMF**S**A	Gag	160	174

2 CD4 responses:

Mos1	WKGSPA**IFQCSMTRI**LEPFRA			
Mos2	WKGSPA**IFQSSMTKI**LEPFRK			
P352	WKGSPAIFQ**A**SMTKI	Pol	308	322
P141	--GSPAIFQ**S**SMTKILD	Pol	310	324

```
P284  ------IFQCSMTKILEPFRA      Pol     314

Mos1  KARVLAEAMSQVTNSAT
Mos2  KARVLAEAMSQTNSTIL
G234  KAKVLAEAMSQVQQT             Gag     359     373
G91   --RVLAEAMSQVTNSAT           Gag     361     375
```

Mosaic 360-07:
16 CD8 PTE+, 2 CD4 PTE+
12 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 3 1 1 1 1 1 1 1 1 1 CD4: 1 1

12 CD8 responses:

```
Mos1  SDLEIGQHRAKIEELR
Mos2  SDLEIGQHRTKIEELR
P385  SDLEIGQHRIKIEEL      Pol     346     360
P133  -DLEIGQHRAKIEELR     Pol     347     361

Mos1  QMHEDIISLWDQSLKP
Mos2  QMHEDIIRLWDQSLKP
E282  HMHEDIISLWDESLK      Env   103     117
E292  -MHEDVISLWDQSLKP     Env   104     118

Mos1  EEKAFSPEVIPMFSAL
Mos2  EEKAFSPEVIPMFTAL
G14   EEKAFSPEVIPMFSA      Gag     160     174
G192  EEKGFNPEVIPMFSA      Gag     160     174
G111  -EKGFSPEVIPMFTAL     Gag     161     175

Mos1  KYTAFTIPSTNNETP
Mos2  KYTAFTIPSINNETP
P286  KYTAFTIPSTNNETP      Pol  281  295

Mos1  EKIKALTAICEEMEK
Mos2  EKIKALVEICTEMEK
P439  EKIEALTAICEEMEK      Pol  184  198

Mos1  GKYAKMRTAHTNDVK
Mos2  GKYARMRGAHTNDVK
P444  GKYARMRGAHTNDVR      Pol  507  521

Mos1  LPIQKETWETWWTDY
Mos2  LPIQKETWEAWWTEY
P358  LPIQKETWEAWWTEY      Pol  546  560

Mos1  WYQLEKDPIAGVETF
```

```
Mos2  WYQLEKEPIVGAETF
P370  WYQLEKDPIAGAETF       Pol   581   595

Mos1  ENVTENFNMWKNNMV
Mos2  ENVTENFNMWKNDMV
E273  ENVTEEFNMWKNNMV       Env   87    101

Mos1  VIQDNSDIKVVPRRK
Mos2  VIQDNSDIKVVPRRK
P765  VIQDNSDIKVVPRRK       Pol   965   979

Mos1  VSQNYPIVQNIQGQM
Mos2  VSQNYPIVQNLQGQM
G29   VSQNYPIVQNLQGQM       Gag   128   142

Mos1  VLAEAMSQVTNSATIM
Mos2  VLAEAMSQ-TNSTILM
G307  VLAEAMSQ-AQQTNIM      Gag   362   377
```

2 CD4 responses:

```
Mos1  HGQVDCSPGIWQLAC
Mos2  HGQVDCSPGIWQLAC
P25   HGQVDCSPGIWQLDC       Pol   766   780

Mos1  RVLAEAMSQVTNSAT
Mos2  RVLAEAMSQTNSTIL
G305  RVLAEAMSQTNSAIL       Gag   361   376
```

Mosaic 361-07:

21 CD8 PTE+, 6 CD4 PTE+

7 CD8+ regions, 5 CD4+ regions

Number of overlapping peptides per region: CD8: 6 2 4 2 5 1 1 CD4: 2 1 1 1 1

7 CD8 responses:

```
Mos1  EQLIKKERVYLSWVPAHKGIG
Mos2  EQLIKKEKVYLAWVPAHKGIG
P221  EELIKKEKVYLAWVP           Pol     678     692
P250  EPLIKKEKVYLSWVP           Pol     678     692
P316  -KLIEKDKVYLSWVPA          Pol     679     693
P223  --LIKKERVYLSWVPAH         Pol     680     694
P169  ------EKVYLAWVPAHKGIG     Pol     684     698
P83   ------EKVYLSWVPAHKGIG     Pol     684     698

Mos1  CTHGIRPVVSTQLLL
Mos2  CTHGIKPVVSTQLLL
E15   CTHGIKPVVSTQLLL            Env   247     261
```

E47 CTHGI**R**PVVSTQLLL Env 247 261

Mos1 IC**TTTVPWNASWSNK**SL
Mos2 IC**TTAVPWNTSWSNK**SQ
E334 ICTT**T**VPWN**A**SWSN**R** Env 603 617
E214 -CTT**T**VPWN**S**SWSNKT Env 604 618
E158 --TT**A**VPWN**A**SWSNKSL Env 605 619
E290 --TT**A**VPWN**T**SWSNKSL Env 605 619

Mos1 ACQGV**GGPGHKARVL**AEAMS
Mos2 ACQGV**GGPSHKARVL**AEAMS
G166 ACQEVGGP**G**HKARVL Gag 349 363
G76 -----GGP**S**HKARVLAEAMG Gag 354 368

Mos1 AAEW**DRVHPVHAGPIA**PGQ
Mos2 AAEW**DRLHPVHAGPVA**PGQ
G319 AAE-DRLHPVHAGP**IP** Gag 209 225
G242 -ADWDR**L**HPVHAGP**V**A Gag 210 224
G44 -AEWDR**L**HPVHAGP**I**A Gag 210 224
G277 ---WDR**V**HPVHAGP**NP**PG Gag 212 226
G102 ----DR**V**HPVHAGP**IP**PGQ Gag 212 226

Mos1 **HSNWRAMASDFNLPP**
Mos2 **HSNWRAMASEFNLPP**
P93 HSNWRAMAS**D**FNLPP Pol 731 745

Mos1 **KGRPGNFLQNRPEPT**
Mos2 **KGRPGNFLQSRPEPT**
G86 KGRPGNFLQ**N**RPEPT Gag 442 456

5 CD4 responses:
Mos1 R**SLYNTVATLYCVHQ**R
Mos2 R**SLFNTVATLYCVH**AE
G289 KSL**F**NTVATLYCVH**A** Gag 76 90
G54 -SL**Y**NTVATLYCVH**Q**R Gag 77 91

Mos1 **YVTDRGRQKIVSLTE**
Mos2 **YVTDRGRQKVVSLTD**
P134 YVTDRGRQK**V**VSLT**E** Pol 612 626

Mos1 **CTTTVPWNASWSNKS**
Mos2 **CTTAVPWNTSWSNKS**
E53 CTT**N**VPWN**S**SWSNKS Env 604 618

Mos1 **WWAGIQQEFGIPYNP**
Mos2 **WWAGIKQEFGIPYNP**

P135	WWAGI**Q**QEFGIPYNP	Pol	846	860

Mos1	**GPGHKARVLAEAMSQ**			
Mos2	**GPSHKARVLAEAMSQ**			
G15	GP**G**HKARVLAEAMSQ	Gag	355	369

Mosaic 362-07:
13 CD8 PTE+, 5 CD4 PTE+
7 CD8+ regions, 5 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 4 2 1 1 1 CD4: 1 1 1 1 1

7 CD8 responses:

Mos1	**CTHGIRPVVSTQLLL**			
Mos2	**CTHGIKPVVSTQLLL**			
E15	CTHGI**K**PVVSTQLLL	Env	247	261
E47	CTHGI**R**PVVSTQLLL	Env	247	261

Mos1	KHIVW**ASRELERFAV**NPGLL			
Mos2	KHLVW**ASRELERFAL**NPGLL			
G20	KHLVWASRELERFA**L**	Gag	32	46
G68	-----ASRELERFA**V**NPGLL	Gag	37	51

Mos1	QPSLQT**GSEELRSLY**NTVATL			
Mos2	QPALQT**GTEELRSLF**NTVATL			
G268	LPALKTG**S**EELRSL**Y**	Gag	65	79
G249	--ALQTG**S**EELRSL**F**NT	Gag	67	81
G132	---LQTG**T**EELRSL**F**NTV	Gag	68	82
G120	------G**S**EELRSL**Y**NTVATL	Gag	71	85

Mos1	ACQGV**GGPGHKARVL**AEAMS			
Mos2	ACQGV**GGPSHKARVL**AEAMS			
G166	ACQEVGGP**G**HKARVL	Gag	349	363
G76	-----GGP**S**HKARVLAEAMG	Gag	354	368

Mos1	**RELNKRTQDFWEVQL**			
Mos2	**RELNKRTQDFWEVQL**			
P233	RELN**R**RTQDFWEVQL	Pol	233	247

Mos1	**LLRAIEAQQHLLQLT**			
Mos2	**LLRAIEAQQHMLQLT**			
E227	LL**M**AIEAQQH**L**LQLT	Env	555	569

Mos1	**KTAVQMAVFIHNFKR**			
Mos2	**KTAVQMAVFIHNFKR**			
P5	KTAVQMAVFIHNFKR	Pol	888	902

5 CD4 responses:
Mos1 **KLNWASQIYPGIKVR**
Mos2 **KLNWASQIYAGIKVK**
P136 KLNWASQIY**A**GIKV**K** Pol 418 432

Mos1 **YAKMRTAHTNDVKQL**
Mos2 **YARMRGAHTNDVKQL**
P266 YA**RK**R**S**AHTNDVKQL Pol 509 523

Mos1 **DQAEHLKTAVQMAVF**
Mos2 **DQAEHLKTAVQMAVF**
P28 DQAEHLKTAVQMAVF Pol 882 896

Mos1 **SLQTGSEELRSLYNT**
Mos2 **ALQTGTEELRSLFNT**
G182 **A**L**K**TG**T**EELRSL**Y**NT Gag 67 81

Mos1 **GPGHKARVLAEAMSQ**
Mos2 **GPSHKARVLAEAMSQ**
G15 GP**G**HKARVLAEAMSQ Gag 355 369

Reactive peptides and breadth and depth of responses: Optimal Natural Clade C Vaccine

Natural Clade C 363-07:
2 CD8 PTE+, 0 CD4 PTE+
1 CD8+ regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 2 CD4: none

1 CD8 responses:
C QPALQT**GTEELRSLY**NTVATL
G268 LPALKTG**S**EELRSLY
G120 ------G**S**EELRSLYNTVATL

Natural Clade C 364-07:
5 CD8 PTE+, 2 CD4 PTE+
5 CD8+ regions, 2 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 1 1 1 CD4: 1 1

5 CD8 responses:
C **AHTNDVKQLTEAVQK**
P80 AHTNDVKQLTEAVQK Pol 515 529

C **GDIIGDIRQAHCNIS**
E64 GDIIGDIRQAHCNIS Env 321 334

```
C     SEKSAVGIGAVFLGF
E182  REKRAVGLGAVFLGF  Env   508   522

C     RASILRGGKLDKWEK
G283  SASILRGGKLDKWEK  Gag   4     18

C     TNNPPVPVGDIYKRW
G185  TSNPPVPVGDIYKRW  Gag   251   265

2 CD4 responses:
C     CKSNITGLLLTRDGG
E316  CTSNITGLLLVRDGG  Env   445   459

C     SLYNTVATLYCVHAG
G54   SLYNTVATLYCVHQR  Gag   77    91
```

```
Natural Clade C 365-07:
2 CD8 PTE+, 1 CD4 PTE+
2 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 1 1 CD4: 1

2 CD8 responses:
C     EQLINKERVYLSWVP
P250  EPLIKKEKVYLSWVP  Pol   678   692

C     AEWDRLHPVHAGPIA
G44   AEWDRLHPVHAGPIA  Gag   210   224

1 CD4 response:
C     HQAAMQMLKDTINEE
G24   HQAAMQMLKDTINEE  Gag   194   208
```

```
Natural Clade C 366-07:
6 CD8 PTE+, 1 CD4 PTE+
4 CD8+ regions, 1 CD4+ regions
Number of overlapping peptides per region: CD8: 2 2 1 1 CD4: 1

4 CD8 responses:
C     IVQQQSNLLRAIEAQQ
E54   VVQQQSNLLRAIEAQ        Env    548      562
E72   -VQQQNNLLRAIEAQH       Env    549      563

C     AVFIHNFKRKGGIGGY
P22   AVFIHNFKRKGGIGG        Pol    894      908
P236  -VLIHNFKRKGGIGGY       Pol    895      909
```

C **MAICEEMEKEGKIT**K
P224 **T**AICEEMEKEGKITK Pol 190 204

C **GGPSHKARVLAEAMS**
G76 GGPSHKARVLAEAM**G** Gag 354 368

1 CD4 response:
C **IIGQVRDQAEHLKTA**
P86 **L**IGQVRDQAEHLKTA Pol 876 890

Natural Clade C 367-07:
0 CD8 PTE+, 0 CD4 PTE+
0 CD8+ regions, 0 CD4+ regions
Number of overlapping peptides per region: none

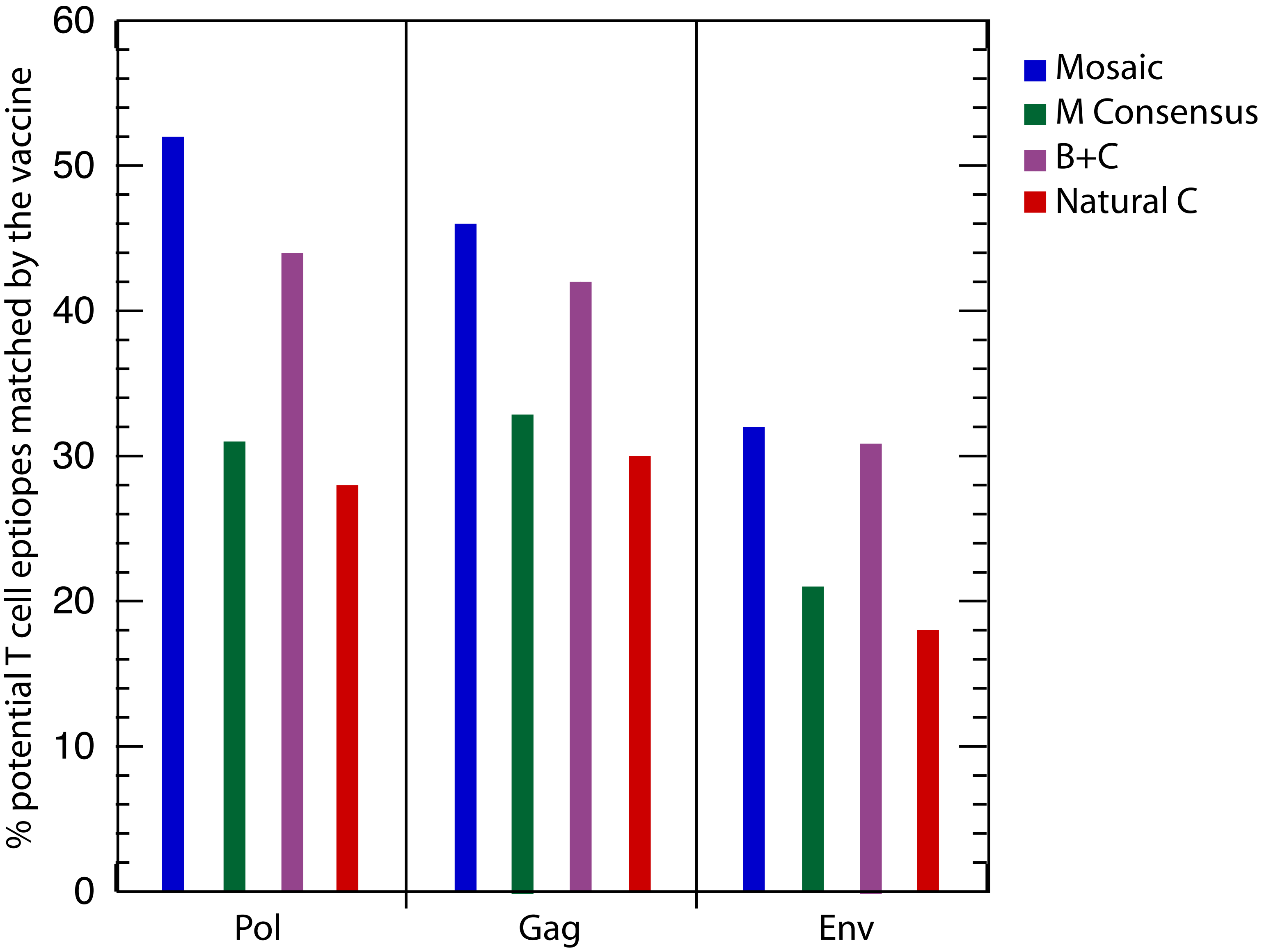
0 CD8 responses

0 CD4 responses

Natural Clade C 368-07:
2 CD8 PTE+, 0 CD4 PTE+
1 CD8+ regions, 0 CD4+ regions
Number of overlapping peptides per region: CD8: 2 CD4: 0

1 CD8 response
C A**VFIHNFKRKGGIGG**Y
P22 AVFIHNFKRKGGIGG Pol 894 908
P236 -V**L**IHNFKRKGGIGGY Pol 895 909

Fig. S4



% potential T cell eptiopes matched by the vaccine
0
10
20
30
40
50
60
Pol
Gag
Env
Mosaic
M Consensus
B+C
Natural C

Mosaic 1 Env

MRVTGIRKNYQHLWRWGTMLLGILMICSAAGKLWVTVYYGVPVWKEATTTLF
CASDAKAYDTEVHNVWATHACVPTDPNPQEVVLENVTENFNMWKNNMVEQM
HEDIISLWDQSLKPCVKLTPLCVTLNCTDDVRNVTNNATNTNSSWGEPMEKGEI
KNCSFNITTSIRNKVQKQYALFYKLDVVPIDNDSNNTNYRLISCNTSVITQACPKV
SFEPIPIHYCAPAGFAILKCNDKKFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGS
LAEEEVVIRSENFTNNAKTIMVQLNVSVEINCTRPNNNTRKSIHIGPGRAFYTAGD
IIGDIRQAHCNISRANWNNTLRQIVEKLGKQFGNNKTIVFNHSSGGDPEIVMHSFN
CGGEFFYCNSTKLFNSTWTWNNSTWNNTKRSNDTEEHITLPCRIKQIINMWQEV
GKAMYAPPIRGQIRCSSNITGLLLTRDGGNDTSGTEIFRPGGGDMRDNWRSELYK
YKVVKIEPLGVAPTKAKRRVVQSEKSAVGIGAVFLGFLGAAGSTMGAASMTLT
VQARLLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQL
LGIWGCSGKLICTTTVPWNASWSNKSLDKIWNNMTWMEWEREINNYTSLIYTLI
EESQNQQEKNEQELLELDKWASLWNWFDISNWLW

Mosaic 2 Env

MRVRGIQRNWPQWWIWGILGFWMIIICRVMGNLWVTVYYGVPVWKEAKTTLF
CASDAKAYEKEVHNVWATHACVPTDPNPQEMVLENVTENFNMWKNDMVDQM
HEDIIRLWDQSLKPCVKLTPLCVTLECRNVRNVSSNGTYNIIHNETYKEMKNCSF
NATTVVEDRKQKVHALFYRLDIVPLDENNSSEKSSENSSEYYRLINCNTSAITQA
CPKVSFDPIPIHYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLL
LNGSLAEEEIIIRSENLTNNAKTIIVHLNETVNITCTRPNNNTRKSIRIGPGQTFYAT
GDIIGDIRQAHCNLSRDGWNKTLQGVKKKLAEHFPNKTINFTSSSGGDLEITTHSF
NCRGEFFYCNTSGLFNGTYMPNGTNSNSSSNITLPCRIKQIINMWQEVGRAMYAP
PIAGNITCRSNITGLLLTRDGGSNNGVPNDTETFRPGGGDMRNNWRSELYKYKV
VEVKPLGVAPTEAKRRVVESEKSAVGIGAVFLGILGAAGSTMGAASITLTVQAR
QLLSGIVQQQSNLLRAIEAQQHMLQLTVWGIKQLQTRVLAIERYLQDQQLLGLW
GCSGKLICTTAVPWNTSWSNKSQTDIWDNMTWMQWDKEIGNYTGEIYRLLEES
QNQQEKNEKDLLALDSWKNLWNWFDITNWLW

M Consensus Env

MRVRGIQRNCQHLWRWGTLILGMLMICSAAENLWVTVYYGVPVWKEANTTLF
CASDAKAYDTEVHNVWATHACVPTDPNPQEIVLENVTENFNMWKNNMVEQM
HEDIISLWDQSLKPCVKLTPLCVTLNCTNVNVTNTTNNTEEKGEIKNCSFNITTEI
RDKKQKVYALFYRLDVVPIDDNNNNSSNYRLINCNTSAITQACPKVSFEPIPIHYC
APAGFAILKCNDKKFNGTGPCKNVSTVQCTHGIKPVVSTQLLLNGSLAEEEIIIRS
ENITNNAKTIIVQLNESVEINCTRPNNNTRKSIRIGPGQAFYATGDIIGDIRQAHCNI
SGTKWNKTLQQVAKKLREHFNNKTIIFKPSSGGDLEITTHSFNCRGEFFYCNTSG
LFNSTWIGNGTKNNNNTNDTITLPCRIKQIINMWQGVGQAMYAPPIEGKITCKSNI
TGLLLTRDGGNNNTNETEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAK
RRVVESEKSAVGIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLR
AIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWN
SSWSNKSQDEIWDNMTWMEWEREINNYTDIIYSLIEESQNQQEKNEQELLALDK
WASLWNWFDITNWLW

Clade C Env

MRVTGMLRNCQPWWIWGILGFWMLLIYNVGGNLWVTVYYGVPVWKEAKTTL
FCASDAKAYEKEVHNVWATHACVPTDPNPQEMVLENVTEYFNMWKNDMVDQ
MHEDIISLWDQSLKPCVKLTPLCVTLNCRNVTTSNNATSNDNPNGEIKNCSFNITT
ELRDKRRNEYALFYRLDIVPLSGSKNSSNSSEYRLINCNTSAITQACPKVSFDPIPI
HYCAPAGYAILKCNNKTFNGTGPCNNVSTVQCTHGIKPVVSTQLLLNGSLAEGEI
IIRSENLTNNAKTIIVHLNESIEIVCARPNNNTRKSMRIGPGQTFYATGDIIGDIRQA
HCNISGNWNATLEKVKGKLQEHFPGKNISFEPSSGGDLEITTHSFNCRGEFFYCDT
SKLFNGTTHTANSSITIQCRIKQIINMWQGVGRAIYAPPIAGNITCKSNITGLLLTR
DGGTLNNDTEKFRPGGGDMRDNWRSELYKYKVVEIKPLGIAPTKAKRRVVESE
KSAVGIGAVFLGFLGAAGSTMGAASITLTVQARQLLSGIVQQQSNLLRAIEAQQH
MLQLTVWGIKQLQTRVLAIERYLKDQQLLGIWGCSGKIICTTAVPWNTSWSNKS
LEDIWDNMTWMQWDREINNYTSIIYSLLEESQNQQEKNEKDLLALDSWNNLWN
WFNITKWLW

Mosaic 1 Gag

MGARASVLSGGELDRWEKIRLRPGGKKKYRLKHIVWASRELERFAVNPGLLETS
EGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDTKEALEKIEEEQNK
SKKKAQQAAADTGNSSQVSQNYPIVQNIQGQMVHQAISPRTLNAWVKVVEEKA
FSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEWDRVH
PVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWIILGLNK
IVRMYSPVSILDIRQGPKEPFRDYVDRFYKTLRAEQASQDVKNWMTETLLVQNA
NPDCKTILKALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQ
RGNFRNQRKTVKCFNCGKEGHIAKNCRAPRKKGCWKCGKEGHQMKDCTERQA
NFLGKIWPSNKGRPGNFLQNRPEPTAPPEESFRFGEETTTPSQKQEPIDKEMYPLA
SLKSLFGNDPSSQ

Mosaic 2 Gag

MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLET
SEGCKQIIKQLQPALQTGTEELRSLFNTVATLYCVHAEIEVRDTKEALDKIEEEQN
KSQQKTQQAKEADGKVSQNYPIVQNLQGQMVHQPISPRTLNAWVKVIEEKAFSP
EVIPMFTALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHPVH
AGPVAPGQMREPRGSDIAGTTSNLQEQIAWMTSNPPIPVGDIYKRWIILGLNKIVR
MYSPTSILDIKQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNANPD
CKTILRALGPGATLEEMMTACQGVGGPSHKARVLAEAMSQTNSTILMQRSNFK
GSKRIVKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKI
WPSHKGRPGNFLQSRPEPTAPPAESFRFEETTPAPKQEPKDREPLTSLRSLFGSDPL
SQ

M Consensus Gag

MGARASVLSGGKLDAWEKIRLRPGGKKKYRLKHLVWASRELERFALNPGLLET
SEGCKQIIGQLQPALQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQN
KSQQKTQQAAADKGNSSKVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEK
AFSPEVIPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRL
HPVHAGPIPPGQMREPRGSDIAGTTSTLQEQIAWMTSNPPIPVGEIYKRWIILGLN
KIVRMYSPVSILDIRQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQN

ANPDCKTILKALGPGATLEEMMTACQGVGGPGHKARVLAEAMSQVTNAAIMM
QRGNFKGQRRIIKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMKDCTERQA
NFLGKIWPSNKGRPGNFLQSRPEPTAPPAESFGFGEEITPSPKQEPKDKEPPLTSLK
SLFGNDPLSQ

Clade C Gag
MGARASILRGGKLDKWEKIRLRPGGKKHYMLKHLVWASRELERFALNPGLLET
SEGCKQILKQLQPALQTGTEELRSLYNTVATLYCVHAGIEVRDTKEALDKIEEEQ
NKGQQKTQQAKGADGKVSQNYPIVQNLQGQMVHQAISPRTLNAWVKVIEEKAF
SPEVIPMFTALSEGATPQDLNTMLNTVGGHQAAMQMLKDTINEEAAEWDRLHP
VHAGPIAPGQMREPRGSDIAGTTSTLQEQIAWMTNNPPVPVGDIYKRWIILGLNKI
VRMYSPVSILDIKQGPKEPFRDYVDRFFKTLRAEQATQDVKNWMTDTLLVQNA
NPDCKTILRALGPGATLEEMMTACQGVGGPSHKARVLAEAMSQTGSTIMMQRS
NFKGSKRIVKCFNCGKEGHIARNCRAPRKKGCWKCGKEGHQMKDCTERQANFL
GKIWPSHKGRPGNFLQSRPEPTAPPAESFRFEETTPAPKQELKDREPLTSLKSLFGS
DPLSQ

Mosaic 1 Pol
MAPISPIETVPVKLKPGMDGPRVKQWPLTEEKIKALTAICEEMEKEGKITKIGPEN
PYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
LAVGDAYFSVPLDEGFRKYTAFTIPSTNNETPGIRYQYNVLPQGWKGSPAIFQCS
MTRILEPFRAKNPEIVIYQYMAALYVGSDLEIGQHRAKIEELREHLLKWGFTTPD
KKHQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIY
PGIKVRQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIA
EIQKQGHDQWTYQIYQEPFKNLKTGKYAKMRTAHTNDVKQLTEAVQKIAMESI
VIWGKTPKFRLPIQKETWETWWTDYWQATWIPEWEFVNTPPLVKLWYQLEKDP
IAGVETFYVAGAANRETKLGKAGYVTDRGRQKIVSLTETTNQKTALQAIYLALQ
DSGSEVNIVTASQYALGIIQAQPDKSESELVNQIIEQLIKKERVYLSWVPAHKGIG
GNEQVDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPPVVAKEIV
ASCDQCQLKGEAMHGQVDCSPGIWQLACTHLEGKIILVAVHVASGYIEAEVIPAE
TGQETAYFILKLAGRWPVKVIHTANGSNFTSAAVKAACWWAGIQQEFGIPYNPQ
SQGVVASMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERI
IDIIATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDI
KVVPRRKVKIIKDYGKQMAGADCVAGRQDED

Mosaic 2 Pol
MAPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGPEN
PYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
LAVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSM
TKILEPFRKQNPDIVIYQYMAALYVGSDLEIGQHRTKIEELRQHLLRWGFTTPDK
KHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQIYA
GIKVKQLCKLLRGTKALTEVVPLTEEAELELAENREILKEPVHGVYYDPSKDLIA
EIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTEAVQKIATESIV
IWGKTPKFKLPIQKETWEAWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPI
VGAETFYVAGAANRETKLGKAGYVTDRGRQKVVSLTDTTNQKTALQAIHLALQ
DSGLEVNIVTASQYALGIIQAQPDKSESELVSQIIEQLIKKEKVYLAWVPAHKGIG

GNEQVDKLVSRGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPPIVAKEIV
ASCDKCQLKGEAIHGQVDCSPGIWQLACTHLEGKVILVAVHVASGYIEAEVIPAE
TGQETAYFLLKLAGRWPVKTIHTANGSNFTSATVKAACWWAGIKQEFGIPYNPQ
SQGVVASINKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGEYSAGERIV
DIIASDIQTKELQKQITKIQNFRVYYRDSRDPLWKGPAKLLWKGEGAVVIQDNSD
IKVVPRRKAKIIRDYGKQMAGDDCVASRQDED

M Consensus Pol
MAPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALTEICTEMEKEGKISKIGPEN
PYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
LDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSM
TKILEPFRTQNPEIVIYQYMDHLYVGSDLEIGQHRAKIEELREHLLRWGFTTPDKK
HQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYPGI
KVKQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQ
KQGQDQWTYQIYQEPFKNLKTGKYAKMRSAHTNDVKQLTEAVQKIATESIVIW
GKTPKFRLPIQKETWETWWTEYWQATWIPEWEFVNTPPLVKLWYQLEKEPIAG
AETFYVDGAANRETKLGKAGYVTDRGRQKVVSLTETTNQKTELQAIHLALQDS
GSEVNIVTDSQYALGIIQAQPDKSESELVNQIIEQLIKKEKVYLSWVPAHKGIGGN
EQVDKLVSTGIRKVLFLDGIDKAQEEHEKYHSNWRAMASDFNLPPIVAKEIVASC
DKCQLKGEAMHGQVDCSPGIWQLACTHLEGKIILVAVHVASGYIEAEVIPAETG
QETAYFILKLAGRWPVKVIHTDNGSNFTSAAVKAACWWAGIQQEFGIPYNPQSQ
GVVESMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDI
IATDIQTKELQKQITKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIK
VVPRRKAKIIRDYGKQMAGDDCVAGRQDED

Clade C Pol
MAPISPIETVPVKLKPGMDGPKIKQWPLTEEKIKALMAICEEMEKEGKITKIGPEN
PYNTPIFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKSVTV
LAVGDAYFSVPLDESFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSM
TKILEPFRAKNPEIVIYQYMAALYVGSDLEIGQHRAKIEELREHLLRWGFTTPDKK
HQKEPPFLWMGYELHPDKWTVQPIQLPEKDSWTVNDIQKLVGKLNWASQIYSGI
KVRQLCKLLRGAKALTDIVPLTEEAELELAENREILKEPVHGVYYDPSKDLIAEIQ
KQGYDQWTYQIYQEPFKNLKTGKYAKMRTAHTNDVKQLTEAVQKIALESIVIW
GKTPKFRLPIQKETWEIWWTDYWQATWIPEWEFVNTPPLVKLWYQLEKEPIAGA
ETFYVAGAANRETKIGKAGYVTDKGRQKIVTLTETTNQKTALQAIQLALQDSGS
EVNIVTASQYALGIIQAQPDKSESELVNQIIEQLINKERVYLSWVPAHKGIGGNEQ
VDKLVSSGIRKVLFLDGIDKAQEEHEKYHSNWRAMASEFNLPPVVAKEIVASCD
KCQLKGEAIHGQVDCSPGIWQLACTHLEGKVILVAVHVASGYMEAEVIPAETGQ
ETAYYILKLAGRWPVKVIHTANGSNFTSAAVKAACWWAGIQQEFGIPYNPQSQG
VVASMNKELKKIIGQVRDQAEHLKTAVQMAVFIHNFKRKGGIGGYSAGERIIDII
ATDIQTKELQKQIIKIQNFRVYYRDSRDPIWKGPAKLLWKGEGAVVIQDNSDIKV
VPRRKVKIIKDYGKQMAGADCVAGRQDED