Global variation in the HIV-1 V3 region

John C. Blouin, Esther A. Guzman and Brian T. Foley

MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

Introduction

Due to the immunogenicity and functional importance of the V3 loop, there has been a great deal of interest in the V3 region of the envelope protein, resulting in a large international effort to obtain V3 region sequences. This section, which includes sequences taken from 1651 individuals and complete references, provides an overview of the variation of sequences that span this region.

Sequences

To best summarize the spectrum of international HIV-1 variants, only one representative viral sequence was included per infected individual. A complete set of references accompanies the sequence alignments, and nomenclature was preserved from the original papers so individuals and isolates can be clearly identified. HIV1 was deleted from the sequence names in this section, as all sequences included here are HIV1. Included with the references when available are brief descriptions of critical features of the sequences. This includes the health status of the individual from whom the virus was derived, whether or not the virus was cultured, and the year the blood sample was taken.

All sequences are prefaced by a subtype association (see phylogenetic clustering below) and a two letter country code to identify the country that the individual resided in at the time that the blood sample was taken. If the person was a recent immigrant and this information was available, we included the country of origin in the references. The two letter code was developed for Internet (Copyright 1992, Lawrence H. Landwater and the Internet Society), and incorporated here based on a suggestion made by Dr. Francine McCutchan. The key to the country codes follows this introduction. Note that this key has been updated for 1996, with several country codes for eastern european nations added.

Sometimes only one viral sequence was available from a person: a clone from an isolate, or a direct sequence of PCR amplified peripheral blood DNA. For other individuals, up to 80 viral sequences from PCR amplified DNA or RNA from blood samples were available. Consequently, over 8000 sequences are represented by the 1651 included in this section. When two sequences were available from a person, one of the two was randomly selected. When three or more sequences were generated from a person, all available sequences were aligned (without regard to different time points of sampling) and either one representative sequence was chosen, or a consensus of the most common base found in each position in the alignment was generated. If there was a tie (e.g., 10 A's, 10 T's), the top base or amino acid in the alignment was used. If a set of sequences from two or more individuals was epidemiologically linked, and genetically very similar, only one sequence from the set was included, preferably the most recently infected. In the sequence description and references section, the short hand "PCR-direct, peripheral blood DNA" is used to signify that viral DNA was amplified from PBMCs, without culturing, and a single "direct" sequence was obtained from the amplification reaction products. The short hand "Consensus, PCR-clones, peripheral blood DNA" signifies that viral DNA was amplified from PBMCs and a set of clones was generated and sequenced from the PCR amplification products. The cloned sequences were aligned and a consensus was generated. In a handful of cases, a particular gp160 clone from an isolate was shown to be expressed and functional using a vaccinia virus T7 expression system. In these cases, the clone rather than the consensus of all sequences from a particular individual is included.

Phylogenetic clustering

Sequences have been organized according to the phylogenetic subtype association (A-J) of their envelope V3 regions only. The original sequence subtype (A-H) designations were defined based on the phylogenetic relationships determined by using both gag and env genes (when possible), are

approximately genetically equidistant in envelope, and have multiple members. The phylogenetic subtype designations and associations have generally been adopted by the HIV research community, and are now often presented with the publication of new sequences. We have either determined the subtype designations here, if not specified in the original manuscript, or else confirmed the subtype designations of the original manuscripts, and then used the subtypes to organize this section. Generally, confirmations were done by aligning a set HIV-1 V3 region sequences with longer env gene sequences (Part IIIC) that have clear subtype associations, and then using parsimony or neighbor joining trees to determine associations. Some of the shorter gene fragments from this region were given a subtype designation based on Hamming distances, using the similarity function of the MASE program (Faulkner DV, and Jurka J. TIBS 13:321-322 (1988)); these sequences have ".sh" appended to their name to indicate that they were too short for phylogenetic analysis. Parsimony trees were generated using PAUP (David Swofford, Illinois Natural History Survey), and neighbor-joining trees were generated with Kimura distances and a transition to transversion ratio of 1.3 using PHYLIP (Joseph Felsenstein, University of Washington). All available nucleotide sequence information was used for phylogenetic analysis; longer protein sequences were trimmed to be approximately the same length as the majority of the PCR fragments in this region, for the purposes of presentation. Some sequences were difficult to classify, and are included in the "U", or unclassified, section. In addition, it has recently been noted that recombination between HIV-1 occurs when an individual is infected with more than one strain. A meeting was held in Santa Fe, New Mexico in October, 1995 to discuss the implications of recombination and methods for detecting recombinant sequences. Because inter-subtype as well as intra-subtype recombination is known to occur, the subtype designations reported in this section should be interpretted only as pertaining to the V3 region of the envelope gene. For example HIV-1 MAL from Zaire, is known to be recombinant between subtypes A and D, with the V3 loop of env resembling subtype D. D_ZR-MAL is still listed with other subtype D sequences in this study, but may be moved to the U (uncertain) group in the future.

The set of sequences used to help resolve subtype associations included at least two sequences from each subtype (A-H), plus a simian immunodeficiency virus outgroup sequence. The sequences were selected based on being "typical" of the subtype they represent based on phylogenetic analysis. The set has changed as more sequences have accumulated. Thus not all subtype designations were based on the same reference set.

Limitations of phylogenetic analyses

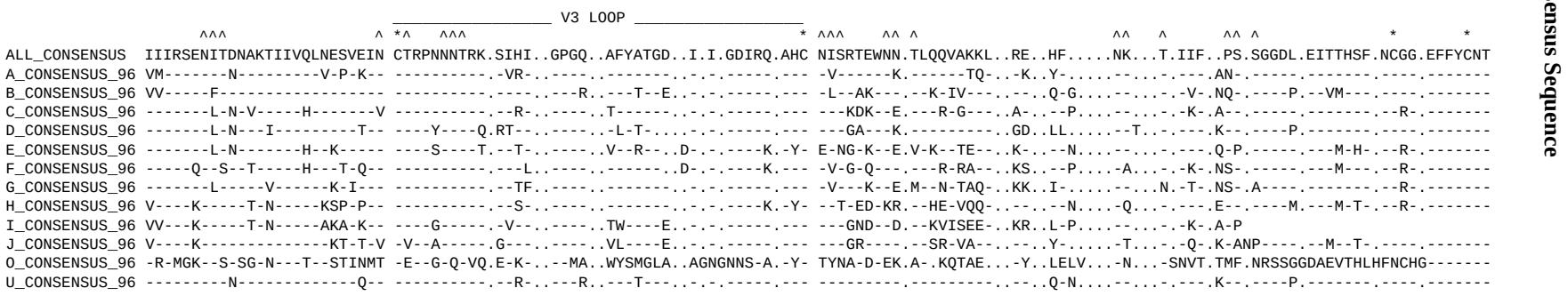
Most of the PCR derived sequences contain a sub-optimal length for phylogenetic analyses, given the level of variability in this region – typically on the order of 250 to 300 nucleotides. Due to this limitation, some of the classifications in this section are uncertain and are our best estimate given the available information. Control studies were performed to compare the phylogenetic clustering of the V3 region using available longer sequences, however, and these studies indicate that our subtype designations based on the V3 region are generally reliable. For 146 sequences, we had an approximately 700 base region of env available representing all of the subtypes A-H. (The limitation in length was due to including the H subtype sequences, which did not cover all of gp120.) After removing positions in the alignment which included gaps, 519 bases were left. When a 298 base V3 region fragment was excised from this set, and neighbor joining trees were constructed using both the 519 base and 298 base long sequences, the phylogenetic subtype designations were consistent in each case. Further, when a subset of longer gp120 sequences was analyzed (92 of the 146), including 935 bases after removing positions in the alignment which included gaps, the subtype designations were again clear in neighbor-joining trees. This indicates that the limited V3 region PCR fragments, which include more than the V3-loop, are generally able to serve as a reliable basis for subtype determination.

Without detailed analysis, genetic recombination between subtypes may obscure phylogenetic relationships between sequences. A characteristic of recombination is an indeterminate place in phylogenetic analyses, and some of the "Uncertain" category sequences may prove to be recombinant genomes upon further inspection. Also, while a subtype designation based on a gene or gene fragment may be correct, recombination events outside the region examined may have occurred. Therefore, care should be taken to not overinterpret the subtype designations. If one is to discuss the subtype

designations of viral isolates based on the data presented here, they should be refer to the designation as “B-like over V3 loop region,” rather than as “subtype B”.

Limitations of V3 amino acid consensus sequences

The V3 amino acid consensus sequences generated for each subtype have interesting features; however, one should be wary about assuming that any of the consensus sequences may broadly represent their subtype. Certainly many V3 loop variants in each of the subtypes are extremely divergent from the consensus sequences. These divergent forms may have very different biological and immunological characteristics from viruses which are similar to the consensus. Additionally, because of the relatively small sample size of most of the subtypes, consensus sequences can be dominated by a small group of highly similar sequences, which may in turn be a sampling artifact. Hence, these consensus sequences are “evolving” as new sequences from each subtype become available.



Subtype consensus sequences. This V3 region alignment shows a consensus sequence generated for each of the eleven subtypes. The subtype consensus sequences indicate the most common amino acid found in each position among the sequences associated with each subtype. The sequences are aligned to a consensus based on the most common amino acid in the subtype consensus sequences, which approximates a "global" consensus. It was generated in this way (rather than by using all 1651 sequences) to avoid over-representation of the B subtype, which has by far the largest number of available sequences. As is the convention in this compendium, a dash (-) indicates concurrence with the top sequence in the alignment; a period (.) indicates a deletion. The carets show where the N-linked glycosylation sites are found in the consensus. The V3 loop is set off from the surrounding sequence by a space on either side to facilitate viewing. Interesting features of the consensus alignment are: 1) Only in the B subtype is GPGR the most common tip of the V3 loop; globally, GPGQ is more prevalent. 2) A highly conserved N-linked glycosylation site is constitutively absent in the C subtype, proximal to the first cysteine (C) in loop. 3) The D subtype consensus has 34 amino acids from cysteine (C) to cysteine (C) rather than the more common 35; at the point where the deletion occurs, it is not uncommon to find insertions of 2 to 4 amino acids, as can be observed in the sequence alignments. 4) The D subtype has two arginine (R) residues in the V3 loop that are uncharacteristic relative to the other consensus sequences; positively charged amino acids in these positions may result in a syncytia inducing, non-monocytropic phenotype (Fouchier RAM, et al., *J. Virol*. **66**:3183–3187 (1992)). 5) A higher degree of variation is seen in the region just downstream of the V3 loop than within it. This difference is also observed internally among the sequences of the different subtypes. 6) The A, C, G and H consensus sequences have very similar V3 loop sequences.

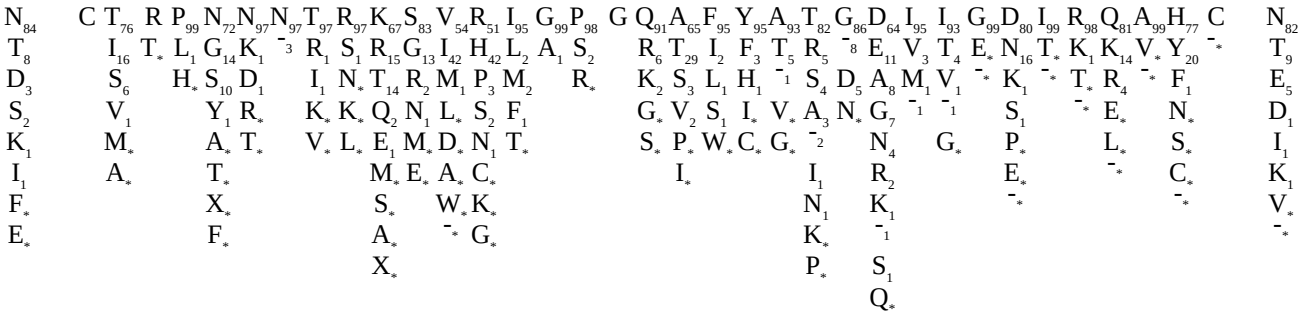
V3 Loop Amino Acids

The following pages present amino acid alignments of the V3 loop, arranged by phylogenetic subtype. For each subtype, the number of sequences used to construct the alignment is indicated. The top line in each alignment represents the consensus sequence for that subtype, where consensus simply means the most common amino acid found in each position among the sequences of the given subtype. The subscripts record the frequency with which that amino acid is observed at that location among members of the subtype. An amino acid which is conserved 100% is shown with no subscript. Directly beneath the most common amino acid in each position are the other amino acids observed in that position, listed from most common to least common. An asterisk (*) subscript means less than 0.5% of the sequences had the indicated amino acid at that location. A dash (-) indicates a gap inserted to maintain the alignment. Percentages were rounded to the nearest whole number.

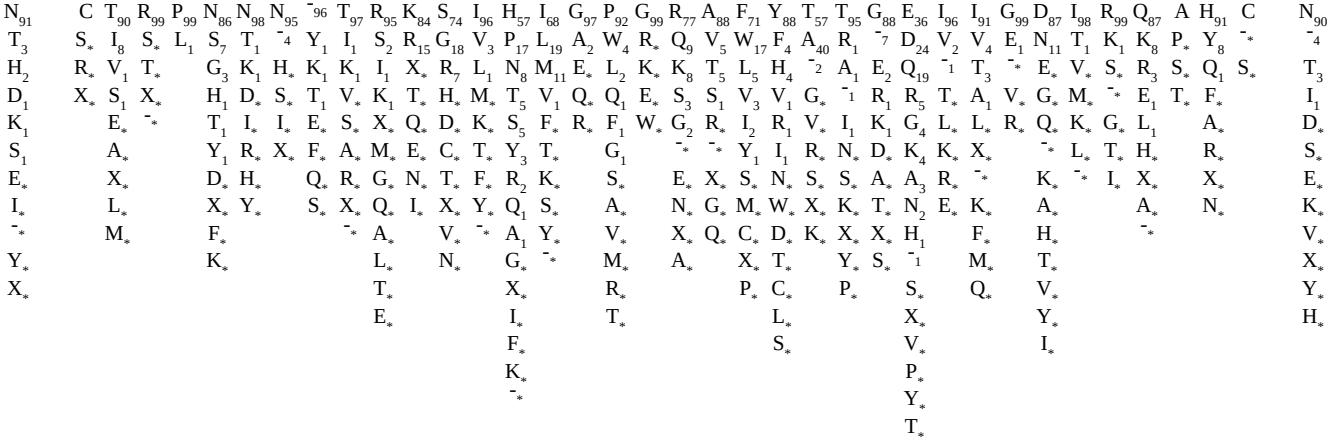
For this year's alignment, the HMMER (version 1.8) hidden Markov model software (Sean Eddy, Dept. of Genetics, Washington U. School of Medicine, St. Louis, MO 63110; eddy@genetics.wustl.edu) was used to objectively align all 1651 sequences. The frequency counts are derived from this alignment. Because each subtypes required different numbers and positions of gaps in order to create the full multiple sequence alignment, some sequences with unusual insertions were trimmed from the HMMER alignment, and a few positions were adjusted by hand, using MASE, prior to printing the full alignment which appears following the country codes description. The sequences which were culled from the alignment after counting frequencies, are appended.

Both the untouched HMMER alignment, and the edited version, will be available via ftp from the LANL HIV database (http://hiv-web.lanl.gov). Questions about these alignments should be directed to (btf@t10.lanl.gov) (505-665-1970).

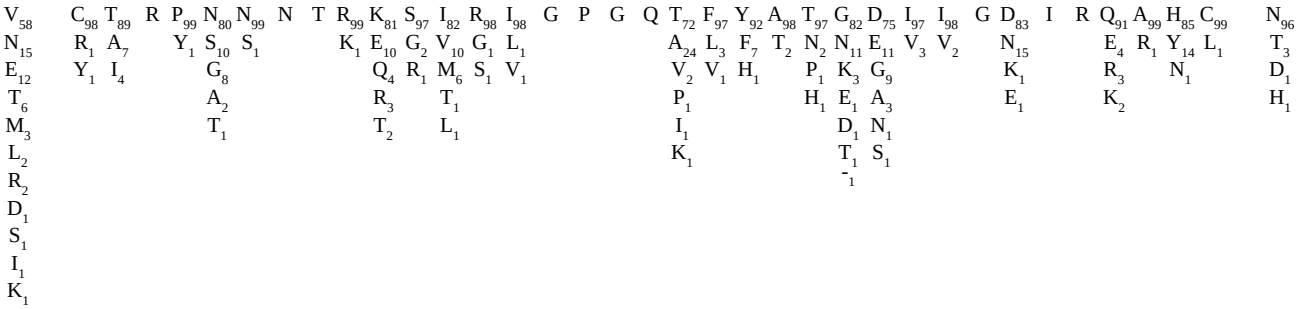
A subtype (207 sequences)



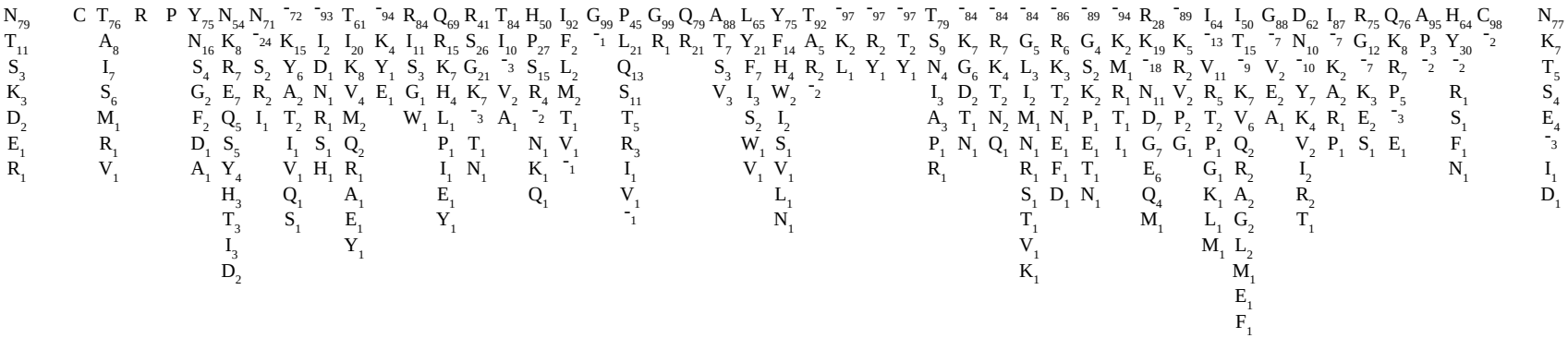
B subtype (975 sequences)



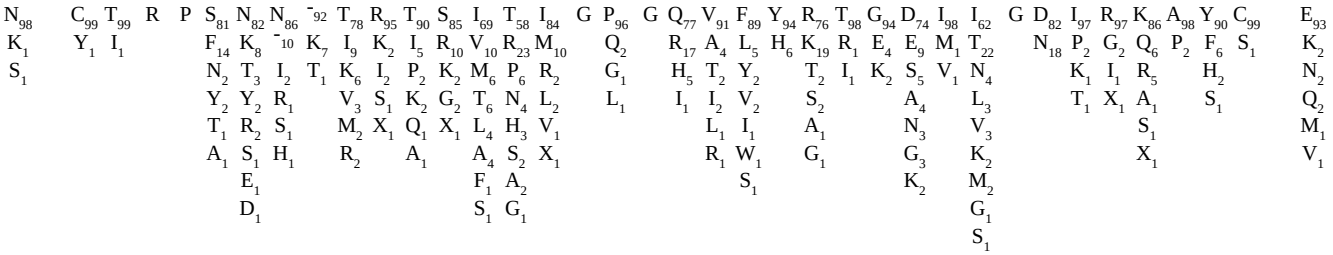
C subtype (119 sequences)



D subtype (107 sequences)



E subtype (124 sequences)



F subtype (59 sequences)

N_{90}	C_{98}	T_{97}	R_{98}	P_{95}	N_{95}	N_{97}	N_{95}	$-_{97}$	T_{95}	R_{98}	K_{92}	S_{86}	I_{98}	H_{71}	L_{58}	G_{98}	P_{97}	G	Q_{66}	A_{86}	F_{93}	Y_{95}	A_{66}	T_{97}	G_{90}	D_{66}	I_{97}	I_{90}	G	D_{93}	I	R	K_{90}	A	H_{83}	C	N_{83}
T_{5}	$-_{2}$	S_{2}	$-_{2}$	F_{2}	D_{2}	K_{2}	$-_{5}$	Y_{2}	I_{3}	$-_{2}$	R_{5}	G_{8}	$-_{2}$	R_{10}	I_{32}	$-_{2}$	L_{2}		R_{34}	V_{8}	I_{3}	F_{3}	T_{27}	A_{3}	S_{3}	E_{15}	V_{3}	T_{5}		N_{7}			Q_{8}		Y_{15}		K_{3}
D_{2}		$-_{2}$		X_{2}	S_{2}	$-_{2}$		T_{2}	$-_{2}$		X_{2}	R_{2}		Q_{7}	F_{3}		V_{2}			T_{2}	L_{2}	H_{2}	X_{3}		A_{2}	A_{5}		N_{2}					R_{2}		C_{2}		V_{3}
X_{2}				$-_{2}$	$-_{2}$						$-_{2}$	X_{2}		P_{5}	X_{3}					P_{2}	X_{2}		R_{2}		D_{2}	G_{3}		V_{2}									X_{3}
$-_{2}$												$-_{2}$		X_{3}	V_{2}					X_{2}			S_{2}		V_{2}	N_{3}		X_{2}									S_{2}
														Y_{2}	$-_{2}$										$-_{2}$	X_{3}											T_{2}
														$-_{2}$												Q_{2}											D_{2}
																										K_{2}											I_{2}

G subtype (23 sequences)

N_{52}	C	T_{87}	R	P	N_{91}	N	N	T	R_{96}	K_{96}	S_{96}	I_{96}	T_{35}	F_{57}	G_{91}	P_{91}	G	Q_{91}	A_{78}	F_{70}	Y	A_{96}	T	G_{78}	D_{35}	I_{96}	I_{96}	G	D_{70}	I	R_{96}	Q_{96}	A	H_{74}	C	N_{96}
T_{17}		I_{9}			S_{9}				I_{4}	R_{4}	R_{4}	K_{4}	R_{22}	I_{26}	A_{9}	T_{9}		R_{9}	T_{13}	L_{22}		T_{4}		D_{9}	E_{22}	V_{4}	T_{4}		N_{30}		K_{4}	K_{4}		Y_{26}		K_{4}
I_{13}		V_{4}											H_{13}	L_{17}					V_{9}	I_{9}				S_{9}	A_{17}											
V_{13}													K_{9}											T_{4}	N_{13}											
M_{4}													S_{9}												Q_{9}											
													N_{9}												G_{4}											
													P_{4}																							

H subtype (2 sequences)

N	C	T	R	P	N	N	N	T	R	K_{50}	S	I_{50}	R_{50}	I	G_{50}	I_{50}	G	P_{50}	G	Q	A_{50}	F	H_{50}	A	I_{50}	G	A_{50}	I	I	G	D	I	R	K_{50}	A	H_{50}	C	N
										R_{50}		M_{50}	S_{50}		$-_{50}$	$-_{50}$		R_{50}			T_{50}		Y_{50}		T_{50}		D_{50}							Q_{50}		Y_{50}		

I subtype (1 sequence)

N	C	T	R	P	G	N	N	T	R	K	S	V	H	I	G	P	G	Q	T	W	Y	A	T	G	E	I	I	G	D	I	R	Q	A	H	C	N

J subtype (5 sequences)

V_{40}	C	V_{60}	R	P	A_{60}	N	N	T	R	K_{80}	G_{80}	I	H	I_{60}	G	P	G	Q	V	L	Y	A	T	G	E_{60}	I_{60}	I	G	D_{60}	I	R	Q_{60}	A	H	C	N
K_{20}		T_{20}			N_{20}					E_{20}	S_{20}			M_{40}											G_{20}	V_{40}			N_{40}			E_{40}				
T_{20}		N_{20}			Y_{20}																				Q_{20}											
E_{20}																																				

O subtype (13 sequences)

T$_{92}$ A$_{8}$ C E$_{31}$ H$_{15}$ I$_{8}$ M$_{8}$ N$_{8}$ Q$_{8}$ S$_{8}$ T$_{8}$ V$_{8}$ R P$_{77}$ E$_{8}$ K$_{8}$ D$_{8}$ -$_{92}$ E$_{8}$ G$_{85}$ E$_{8}$ Q$_{8}$ N$_{46}$ I$_{15}$ R$_{15}$ M$_{8}$ H$_{8}$ D$_{8}$ -$_{77}$ H$_{8}$ N$_{8}$ D$_{8}$ L$_{23}$ Q$_{23}$ -$_{23}$ G$_{15}$ V$_{8}$ A$_{8}$ -$_{92}$ V$_{8}$ T$_{31}$ -$_{31}$ K$_{23}$ E$_{8}$ S$_{8}$ V$_{46}$ I$_{38}$ -$_{15}$ Q$_{85}$ K$_{8}$ -$_{8}$ E$_{62}$ D$_{8}$ K$_{8}$ Q$_{8}$ R$_{8}$ S$_{8}$ I$_{69}$ M$_{23}$ V$_{8}$ K$_{31}$ G$_{15}$ R$_{15}$ A$_{8}$ M$_{8}$ Q$_{8}$ H$_{8}$ Y$_{8}$ -$_{92}$ E$_{8}$ -$_{92}$ M$_{8}$ -$_{92}$ K$_{8}$ I$_{69}$ T$_{15}$ S$_{8}$ A$_{8}$ G P M$_{77}$ L$_{23}$ -$_{62}$ S$_{23}$ A$_{8}$ R$_{8}$ A$_{54}$ -$_{31}$ V$_{8}$ R$_{8}$ W$_{92}$ Y$_{8}$ Y$_{77}$ S$_{8}$ V$_{8}$ R$_{8}$ -$_{92}$ G$_{8}$ S M$_{85}$ L$_{15}$ G$_{46}$ A$_{23}$ Q$_{8}$ S$_{8}$ T$_{8}$ -$_{8}$ L$_{62}$ I$_{23}$ D$_{8}$ -$_{8}$ A$_{38}$ E$_{15}$ G$_{15}$ K$_{15}$ L$_{8}$ S$_{8}$ A$_{31}$ R$_{23}$ G$_{15}$ E$_{15}$ N$_{8}$ K$_{8}$ G$_{23}$ N$_{23}$ E$_{15}$ T$_{15}$ D$_{8}$ S$_{8}$ K$_{8}$ N$_{38}$ S$_{15}$ T$_{15}$ A$_{8}$ G$_{8}$ K$_{8}$ -$_{8}$ G$_{23}$ N$_{23}$ T$_{15}$ K$_{8}$ E$_{8}$ S$_{8}$ A$_{8}$ -$_{8}$ N$_{31}$ S$_{23}$ K$_{8}$ G$_{8}$ I$_{8}$ T$_{8}$ V$_{8}$ -$_{8}$ -$_{31}$ S$_{15}$ E$_{15}$ N$_{8}$ Q$_{8}$ R$_{8}$ A$_{8}$ K$_{8}$ -$_{77}$ N$_{15}$ P$_{8}$ -$_{85}$ S$_{8}$ R$_{8}$ S$_{62}$ T$_{23}$ K$_{8}$ -$_{8}$ -$_{92}$ G$_{8}$ -$_{85}$ G$_{8}$ D$_{8}$ -$_{85}$ T$_{8}$ S$_{8}$ R A$_{31}$ I$_{15}$ R$_{15}$ V$_{15}$ S$_{8}$ T$_{8}$ K$_{8}$ A Y$_{85}$ H$_{8}$ F$_{8}$ C N$_{23}$ K$_{15}$ T$_{15}$ M$_{8}$ E$_{8}$ R$_{8}$ D$_{8}$ X$_{8}$ -$_{8}$

U subtype (15 sequences)

N$_{80}$ K$_{7}$ S$_{7}$ T$_{7}$ C T$_{93}$ I$_{7}$ R P N$_{47}$ S$_{20}$ G$_{13}$ Y$_{13}$ A$_{7}$ N$_{80}$ K$_{7}$ S$_{7}$ T$_{7}$ N$_{73}$ -$_{27}$ -$_{73}$ I$_{7}$ K$_{7}$ T$_{7}$ D$_{7}$ -$_{87}$ R$_{7}$ K$_{7}$ -$_{87}$ N$_{7}$ K$_{7}$ T$_{60}$ I$_{33}$ K$_{7}$ -$_{93}$ K$_{7}$ R K$_{33}$ R$_{27}$ Q$_{20}$ E$_{7}$ T$_{7}$ X$_{7}$ S$_{60}$ G$_{13}$ R$_{7}$ I$_{7}$ T$_{7}$ X$_{7}$ I$_{53}$ T$_{27}$ M$_{13}$ V$_{7}$ R$_{47}$ H$_{40}$ P$_{7}$ N$_{7}$ I$_{80}$ F$_{13}$ L$_{7}$ G P$_{80}$ L$_{13}$ S$_{7}$ G Q$_{47}$ R$_{47}$ K$_{7}$ A$_{73}$ T$_{13}$ V$_{13}$ F$_{73}$ L$_{13}$ I$_{7}$ Y$_{7}$ Y$_{80}$ F$_{20}$ T$_{47}$ A$_{40}$ K$_{7}$ -$_{7}$ T$_{73}$ N$_{13}$ S$_{7}$ K$_{7}$ -$_{47}$ G$_{40}$ X$_{7}$ R$_{7}$ -$_{93}$ G$_{7}$ -$_{93}$ V$_{7}$ -$_{93}$ E$_{7}$ -$_{93}$ G$_{7}$ -$_{93}$ K$_{7}$ -$_{93}$ K$_{7}$ -$_{93}$ G$_{7}$ -$_{93}$ P$_{7}$ D$_{47}$ G$_{13}$ -$_{13}$ K$_{7}$ N$_{7}$ X$_{7}$ E$_{7}$ -$_{93}$ P$_{7}$ I$_{67}$ M$_{13}$ -$_{13}$ V$_{7}$ I$_{60}$ T$_{13}$ K$_{7}$ G$_{7}$ V$_{7}$ -$_{7}$ G$_{93}$ -$_{7}$ D$_{80}$ -$_{13}$ N$_{7}$ I$_{87}$ -$_{13}$ R$_{87}$ -$_{13}$ Q$_{73}$ K$_{20}$ -$_{7}$ A$_{93}$ -$_{7}$ H$_{67}$ Y$_{27}$ Q$_{7}$ C N$_{73}$ K$_{7}$ I$_{7}$ S$_{7}$ T$_{7}$

Summary of variations in the tetrameric tip of the V3 loop. This table is a tally of the different tetramers observed in the 1651 individuals analyzed. This tip is thought to form a turn, and is the focal point of the potent neutralizing antibody epitopes that have been mapped to the V3 loop, as well as of T cell epitopes. Each column shows the number of occurrences of a given tetramer in either the entire 976 sequences (combined), or in subsets consisting of subtypes A–O, and the unclassified sequences (U). Underneath the column heading is the number of sequences in each category. The most common form found in each subtype is highlighted in bold lettering. In the B subtype, GPGR is the predominant form, however globally GPGQ is more common.

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
GPGR	711	10	**643**		17	18	17						**6**
GPGQ	590	**184**	91	**119**	**31**	**95**	**39**	**19**	**1**	**1**	**5**		5
GPGK	81	4	76										1
GWGR	34		34										
GPGS	26	1	25										
GPGG	22	2	20										
GLGQ	21				20								1
GLGR	19		15		1	1	1						1
APGR	17	1	16										
GSGQ	16	3			12								1
GQGQ	13		1		12								
GQGR	9		6		1	2							
GPMA	8											**8**	
GPGH	6					6							
GFGR	6		6										
GTGQ	5				5								
GRGQ	5	1			3				1				
GVGR	4		2		1		1						
GSGR	4	1	3										
GPRR	3		3										
GPGA	3		3										
EPGR	3		3										
APGS	3		3										
APGQ	3	1						2					
GTGR	2							2					
GPKR	2		2										
GMGR	2		2										
GGGR	2		2										
GAGR	2		2										
AGGR	2		2										
GLGS	1		1										
GTGG	1		1										
GPMR	1											1	
GLRQ	1				1								
AQGR	1				1								
GPMS	1											1	
GPLR	1											1	
GARR	1		1										
GGGQ	1					1							
GPWG	1		1										
GIGQ	1				1								
RPGR	1		1										

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
GRGR	1		1										
GPGR	1		1										
GQGI	1					1							
GPLS	1											1	
GPWG	1		1										
GPGN	1		1										
GPGX	1		1										
GPGE	1		1										
GPEK	1		1										
GLGK	1		1										
GARR	1		1										
AWGR	1		1										
APGG	1		1										
AGGK	1		1										
AQGR	1				1								
GIGQ	1				1								
GGRA	1				1								
*PGR	1						1						

Summary of variations in the octameric tip of the V3 loop. This table is a tally of the different octamers observed in the 1651 individuals analyzed. This table is structured the same as the tetramer table on the previous pages. Amino acid changes proximal to the tip can influence the specificity of anti-V3 neutralizing antibodies. The forms that were found only once in the data set are not shown here, to save space, and are summarized in a row labeled "unique."

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
HIGPGRAF	279	3	**269**		2		3						**2**
RIGPGQTF	136	46		**82**	1	2	1	**3**					1
RIGPGQAF	75	45		27		1		1					1
PIGPGRAF	71	1	69				1						
HIGPGQAF	62	**56**	2				4						
NIGPGRAF	59		58										1
HLGPGQAW	44		44										
TIGPGQVF	39					**39**							
PLGPGQAW	31		31										
HLGPGQAF	31	2	1				**28**						
SIGPGRAF	26	1	25										
HIGPGKAF	25	2	23										
RIGPGQVF	23	3		2		16	1						1
TIGPGRAF	19		19										
HIGPGQAL	16		2		**13**			1					
YIGPGRAF	15		15										
PIGLGQAL	14				13								1
HMGWGRAF	14		14										
HIAPGRAF	14	1	13										
PLGPGRAW	11		11										
HLGPGRAW	10		10										
HIGPGSAF	10		10										
HIGPGRAY	10				10								
TMGPGQVF	9					9							
PMGPGRAF	9		9										
HMGPGRAF	9	1	8										
RIGPGRVF	8					8							
PIGPGQAF	7	3			1	2		1					
HMGWGRTF	7		7										
HIGPGRTF	7		7										
TIGPGRVF	6		1			5							
QIGPGRAF	6		5				1						
PMGPGKAF	6		6										
PLGPGKAW	6		6										
NIGPGRAW	6		6										
HLGWGRAF	6		6										
HIGPGRAV	6		6										
YLGPGRAF	5		5										
RFGPGQAF	5	1					1	1					2
PIGPGKAF	5		5										
NMGPGRAF	5		5										

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
NIGPGQVF	5					5							
HLGPGRAF	5	1	1				2						1
HLGPGGAF	5		5										
HIGSGQAL	5				5								
HIGPGRAW	5		5										
HIGPGRAL	5		2		1		1						1
HIGPGRAI	5		5										
HIGPGGAF	5		5										
SIGQGQAL	4				4								
SIGPGQAF	4	3							1				
PIGPGRAW	4		4										
PIGPGQVF	4					4							
KIGPMAWY	4											**4**	
HMGPGKAF	4		4										
HMGLGRAF	4		4										
HLGPGQAL	4		2		2								
HIGPGRVF	4		3				1						
HIGPGRSF	4	2	2										
HIGPGQVF	4	1				3							
HIGPGQTF	4	4											
HIGPGQAI	4	2			2								
YIGPGRAV	3		3										
YIGPGRAS	3		3										
TMGPGRVY	3		3										
TMGPGRVW	3		3										
TLGPGRVY	3		3										
TIGPGRVY	3		1			2							
TFGPGQAF	3							3					
SLGPGRAW	3		3										
SIGPGRAW	3		3										
RIGPGQTL	3			2	1								
RIGPGQSF	3	3											
PLGPGRAF	3		1				2						
NIGPGQAF	3	2			1								
HMGPGKTF	3	1	2										
HLGQGRAW	3		3										
HIGTGQAL	3				3								
HIGSGQAY	3				3								
HIGPGQVL	3										**3**		
HIGPGQAW	3		3										
GIGPGQTF	3	2		1									
AIGPGQVF	3					3							
XIGPGRAF	2		2										
TRGPGHVF	2					2							
TMGPGRVL	2		2										
TMGPGKVF	2		2										
TMGPGHVF	2					2							
TLGPGQAF	2							2					
TIGPGQVL	2					2							

V3 Loop Variation

	Combined	A	B	C	D	E	F	G	H	I	J	O	U
Totals	1651	208	975	119	107	124	59	23	2	1	5	13	15
TIGPGQIF	2					2							
SMGPGRAF	2		2										
SLGPGKAW	2		2										
SIGQGRVL	2					2							
SIGQGQTL	2				2								
SIGPGRVW	2		2										
SIGLGQAL	2				2								
SFGPGQAF	2							2					
RIGPMAWY	2											2	
RIGPGSAF	2		2										
RIGPGRVI	2						2						
RIGPGRTF	2		1										1
RIGPGRAV	2		2										
RIGPGRAF	2		2										
RIGPGQAL	2				2								
QLGPGRAW	2		2										
PLGPGRVW	2		2										
PIGSGQAL	2				2								
PIGRGQAL	2				2								
PIGLGQAY	2				2								
PIAPGSAW	2		2										
KIGPGQTF	2	1			1								
KFGTGRVL	2							2					
HVGPGQAF	2				1		1						
HMGPGRAL	2				2								
HMGPGQVL	2										2		
HMGPGGAF	2		2										
HLGPGKAW	2		2										
HLGPGKAF	2		2										
HLGLGRAF	2		2										
HLGFGRAL	2		2										
HIGSGRAF	2	1	1										
HIGSGQAI	2				1								1
HIGPGSAL	2		2										
HIGPGQAY	2				2								
HIGPGKVF	2		2										
HIGLGRAY	2				1								1
HIGGGRTL	2		2										
HIEPGRAF	2		2										
HFGPGQAL	2				1			1					
GIGPGRTV	2		2										
AIGPGRTV	2		2										
UNIQUE	189	26	93	3	26	8	10	11	2	3	7		