

# PeptGen: Designing Peptides for Immunological Studies and Application to HIV Consensus Sequences

Charles Calef,<sup>1</sup> Rama Thakallapally,<sup>1</sup> Dorothy Lang,<sup>1</sup> Christian Brander,<sup>2</sup> Philip Goulder,<sup>3</sup> Otto Yang,<sup>4</sup> Bette Korber<sup>1</sup>

<sup>1</sup> *Theoretical Biology and Biophysics (T10), Los Alamos National Laboratory, Los Alamos, NM 87545*

<sup>2</sup> *Partners AIDS Research Center, Massachusetts General Hospital, Boston, MA*

<sup>3</sup> *John Radcliffe Hospital, Oxford OX3 9DS, U.K.*

<sup>4</sup> *AIDS Institute, UCLA Medical Center, Los Angeles, CA 90095*

## Abstract

PeptGen is a Web-based tool that enables users to subdivide protein sequences into shorter overlapping peptides according to user-specified criteria. Immunologists can use it to generate peptide sequences for immunological analysis according to a consistent set of rules. As an example, this tool was applied to the HIV-1 B subtype consensus sequences of Pol, Gag, Env, and the HIV regulatory proteins. A stringent set of rules was followed that prohibited certain amino acids from occurring at the N- and C-terminal positions. These “forbidden” amino acids were problematic either in terms of peptide chemistry or in terms of being unlikely to serve as anchor residues for binding to class I MHC proteins. This peptide set is meant to serve as an example and users can select their own parameters and protein sequences (HIV or other proteins) as input to the interactive Web version of the program. The use of consensus sequences for generation of testing reagents, versus laboratory strains, is discussed. The consensus sequence more closely resembles currently circulating forms than any one isolate resembles other circulating forms, so would be more likely to elicit a cross-reactive CTL response. An alignment for comparison of all major subtype consensus sequences, the global consensus, and two B subtype isolates is provided. Consensus sequences for all HIV-1 subtypes and for all HIV-1 proteins are available at the PeptGen Web site:

<http://hiv-web.lanl.gov/PEPTGEN/PeptGenSubmitForm.html>

## Introduction

Immunologists frequently need to generate overlapping peptides to determine immunogenic regions and define T cell and contiguous antibody epitopes within a protein. We have designed a simple tool with a Web interface to facilitate this process. The user inputs a protein sequence, and specifies the length of peptide and the extent of the overlap. In the past a typical HIV study might involve a single protein serially broken down into peptides 20 amino acids in length with an overlap of 10 amino acids. The initial determination of reactive peptides would be followed by subsequent fine mapping within the reactive peptides to more precisely define the optimal epitope. Because of recent experimental advances [Kaul & Rowland-Jones(1999)] it has become possible to use smaller peptides that more closely identify the reactive epitope in the first screening, as well as to perform a more comprehensive examination of many HIV proteins in the context of a single study.

It occurred to us that it would be helpful to provide a Web-based tool that could neatly map out overlapping peptides that follow a set of pre-specified rules. In particular, for CTL studies, it may be advantageous to modify peptide lengths when amino acids that are unlikely to serve as C-terminal anchor residues fall in the C-terminal position in a peptide. Additionally, certain amino acids could be forbidden based on the pragmatic difficulties related to peptide synthesis. The variables that the user may specify are

1. peptide length
2. peptide overlap
3. forbidden amino acids at the C- and N-terminal ends of a peptide
4. how much to shorten or lengthen the peptide in search of allowed terminal amino acids.

The output yields a list of peptides generated according to the users intentions. The Kyte-Doolittle hydrophobicity score [Kyte & Doolittle(1982)] for the peptide can be provided, and the amino acids in the peptides can be colored according to a hydrophobicity scale or according to amino acid side-chain chemistry if desired.

## What does PeptGen do?

Given an amino acid sequence, PeptGen splits the sequence into shorter overlapping peptide fragments and displays them. Figure 1 shows what part of the output might look like. The first line is the submitted amino acid sequence, while the peptides generated by the program form the stair-step pattern below. The submitted protein sequence may have spaces and newline characters within

Word length: 18  
 Represent all n-mers of length: 11  
 Forbidden C-term characters: **GPEDQNTSC**  
 Forbidden N-term characters: **Q**  
 Shorten by: 3  
 Lengthen by: 2

```

MGARASvLsggeLDrWEkiRLRPGgKkKykLKHivWASRELERFavnpgLLEtseGCrQILgLqpsLqTGSElrSlyNtvATLLyCvHgrievkDTKeAI
MGARASvLSggeLDrWEkiRLRPGgK (18) [-0.54]
  SgeLDrWEkiRLRPGgK (18) [-1.37]
    EkiRLRPgKkKykLKHi (18) [-1.46]
      GkkKykLKHivWASREL (17) [-0.94]
        KHivWASRELERFav (15) [-0.11]
          ASRELERFavnpgLL (15) [-0.02]
            ERFavnpgLLEtseGCr (17) [-0.47]
              gLLEtseGCrQILgL (16) [0.06]
                eGCrQILgLqpsLqTGS (18) [-0.48] All C-term amino acids forbidden
                  gLqpsLqTGSElrSly (18) [-0.76]
                    TGSElrSlyNtvATLLy (17) [-0.25]
                      SlyNtvATLLyCvHgriev (18) [0.33]
                        LyCvHgrevkDTKeAI (17) [-0.39]
  
```

Figure 1. Sample PeptGen output for part of the HIV-1 Gag Consensus B protein. User-specified input parameters are shown at the top. Peptide length is shown in parentheses; Kyte-Doolittle hydropathy score in square brackets. Forbidden amino acids are underlined and emboldened.

it, but these (and a few other “funny characters”) will be removed. The number in parentheses following each peptide records the length of the peptide. In this particular example, the user had specified that the program first attempt to construct peptides of 18 amino acids, which we will refer to as the “ideal” length.

#### Forbidden Amino Acids

Why then are some of the peptides shorter than 18 amino acids? Because the user also specified a list of amino acids that are “forbidden” to occur at the C-terminus because of their inimical effect on binding to class I molecules. Forbidden amino acids are underlined and printed in bold. In this run the forbidden amino acids were declared to be “**GPEdQNTSc**”. Note, forbidden amino acids can be entered as either capital letters or lowercase letters so GpeDqnTSc would give the same results.

As can be seen the fourth peptide in Figure 1 is only 17 AAs-long because the 18th amino acid is an “E”, a member of the forbidden set. The fifth peptide is only 15 amino acids long because amino acids 16, 17, and 18 are N, P, and G, all forbidden. The first position at which non-forbidden amino acid is found determines the length of the peptide. How much the program will shorten the peptide in its search for an allowed amino acid is specified by the user. In this example the “shorten by” parameter was set to 3. If all the amino acids from 18 down to 15 are forbidden, then the program begins to lengthen the peptide beyond the ideal of 18, one amino acid at a time until it finds an allowed amino acid. It will add amino acids to the peptide up to the limit set by the “lengthen by” parameter (2 in this case). If all amino acids between between 15 and 20 are forbidden, then the 18-mer is used even though it ends in a forbidden amino acid. Such peptides—one of which is shown in Figure 1—are marked with the words “All C-term amino acids forbidden”. The number of times such “forbidden” peptides are generated is reported at the end of your output. In the peptides only the forbidden amino acids near the C-terminus are printed in bold and underlined in the peptide, but all forbidden amino acids are underlined in the reference strain.

The rarely-invoked “Proline rule” says that a peptide may never end in a proline residue. In this sense proline may be thought of as ultra-forbidden. If all the AAs near the C-terminus are forbidden, the program selects the peptide of “ideal length” (18-mer in the examples above) even though it ends with a forbidden AA. But if the proline rule is being observed and the eighteenth AA is proline then the 17-mer will be chosen as the peptide. Under the proline rule the algorithm will shorten the ideal peptide and then lengthen it searching for non-proline AAs. If all AAs at the C-term are prolines this is reported and the

18-mer would be used.

The user may also specify forbidden character(s) at the N terminus. Because of difficulties of synthesizing peptides that initiate with an N-terminal Glutamine (Q), that is the default forbidden character. If after shifting right the proper amount to begin the next peptide the program detects an N-terminal Q it will back up one position and try to generate a peptide from that starting position, and so on until it finds a non-Q start. It is possible that given severe constraints on the N- and C-terminal amino acids that a peptide cannot be generated. In this case the program terminates and suggests you adjust your constraints.

#### Peptide Overlap

The overlap determines how many amino acids are shared between adjacent peptides. In the example, the word length was set to 18, the overlap was set to 10. With an overlap of 10, every 11-mer will be represented exactly once as illustrated in Figure 2. The reason for implementing this parameter is that in immunological studies it is usually the shorter peptides (9- to 11-mers) that are of most interest, but the cost and labor of generating every single one of these is prohibitive. As a compromise, longer peptides are synthesized in such a way to insure that shorter peptides of a given length (e.g. 11-mers) will be represented at least once in the longer peptides. The algorithm this program uses insures that this will be the case.

```

MGARASVLSGGELDRWEK (18)  first peptide
MGARASVLSGG
GARASVLSGGE
ARASVLSGGEL
RASVLSGGELD
ASVLSGGELDR
SVLSGGELDRW
VLSGGELDRWE
LSGGELDRWEK
SGGELDRWEKIRLRPFGK (18)  second peptide
SGGELDRWEKI
GGELDRWEKIR
GELDRWEKIRL
ELDRWEKIRLRL
LDRWEKIRLRP
DRWEKIRLRPFG
RWEKIRLRPFGG
WEKIRLRPFGGK
- 11-mers within 1st peptide
- 11-mers within 2nd peptide
EKIRLRPFGKKYKLRKHI (18)  third peptide etc. . .

```

Figure 2. A peptide overlap of ten insures each 11-mer will be represented exactly once.

### Hydropathy

The user may exercise options to color the amino acids of the peptides according to Kyte-Doolittle hydropathy classes and to compute the average hydropathy score for each peptide. If these options are selected, the most hydrophobic will be colored red and the most hydrophilic blue, moving through the spectrum accordingly. The hydropathy score is printed in square brackets after each peptide.

### How to use this site

Figure 3 shows the submission form for the PeptGen Web site. To use, paste your raw amino acid sequence into the large text box beneath the “Sample Input” button. Below that box is a table which contains the adjustable parameters and their default values described above in this article. Change any of these parameters to your specifications and click on the “FIRE” button to submit your PeptGen request.

### Applying PeptGen to the B consensus sequence

Epitopes can be missed due to differences between an infecting strain and the strain used to probe the immune response [Jin(2000)]. In circumstances where it is not possible to use autologous strains to explore individual immune responses, there may be advantages in using HIV consensus sequences or modeled ancestral sequences as a basis for scanning for immunologically reactive peptides. In phylogenetic analysis, sequences from the same subtype tend to be approximately equidistant, with a tendency to have longer branch lengths relative to the ancestral node of the subtype as time goes on. There is a star-like pattern to these subtype groupings, such that there is often little clear phylogenetic structure within a subtype. Under these circumstances, a consensus sequence, comprising the most common amino acid found in a particular position, closely resembles the most recent common ancestor of the subtype. Because the consensus is “central”, there are on average approximately half the number of amino acid differences between the consensus and a contemporary strains as there are between two contemporary strains (Figure 2). For example, there are 220 substitutions between B subtype sequences HXB2 and DH123 in the 856 amino acids of Envelope. But there are only 98 or 128 changes between HXB2 or DH123 and the B subtype consensus, respectively. A subtype consensus is also more similar to other subtypes than contemporary strain inter-subtype sequence relationships, which may be a consideration for vaccine design (*i.e.*, a B subtype consensus would be more similar than would a contemporary B strain to a contemporary C strain).

### What does this site do?

Paste a plain amino acid sequence into the window, then change the default values for the other parameters if you like and hit FIRE!

Paste your sequence here:

### Sample Input

```

MENRWQVHLVWQYDRIRIRITWKSLSVKHHHTYSGKARGWFYRHHYSEPHRISSEYHIPLG
DARLYITTYWGLHTGERDWHLGGQVSIWPKKPYSTQYDPELADQLIHLYYFDCESDSAI
RKALLGHIVSFRCEYQAGHNKYGSLQYLALAAALITPEKKIKPFLPSVTKLTEDRWKPKQKT
KCHRGSHHTINGH

```

Make peptides of this length:  C-term forbidden amino acids:

Overlap peptides by:  N-term forbidden amino acids:

Shorten by:  Apply proline rule?  Yes  No

Lengthen by:  Calculate hydrophathy?  Yes  No

Color amino acids?  Yes  No

If you have problems with this site please contact the

Figure 3. The submission form of the PeptGen Web site with a protein pasted into the text box beneath the “Sample Input” button. This site can be found at <http://hiv-web.lanl.gov/PEPTGEN/PeptGenSubmitForm.html>

The consensus of the subtype consensus sequences is not weighted for any one subtype, and through maximum likelihood phylogenetic analysis essentially mimics the M group ancestral sequence [Korber(2000)]. This sequence is most central in the tree, and may be most appropriate for consideration of a single sequence with the greatest potential for global breadth of cross-reactivity to M group strains. The consensus sequence from all major HIV subtypes, and HIV-1 M group “consensus of the consensus” is provided at the web site for user input.

For the sake of comparison, an alignment of all HIV proteins, including two subtype B reference strains, a consensus for each subtype, and the consensus of the consensus is provided. The B consensus is used as the input to PeptGen, with the peptide length of 18, and an overlap of 10, with the settings as in Figure 1.

### References

- [Kaul & Rowland-Jones(1999)] Rupert Kaul and Sarah L. Rowland-Jones, Methods of Detection of HIV-specific CTL and Their Role in Protection Against HIV Infection. In: *Human Retroviruses and AIDS 1999: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences*. Kuiken CL, Foley B, Hahn B, Korber B, McCutchan F, Marx PA, Mellors JW, Mullins JJ, Sodroski J, and Wolinsky S, Eds. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM.
- [Kyte & Doolittle(1982)] Jack Kyte and Russell F. Doolittle, A Simple Method for Displaying the Hydrophobic Character of a Protein, *J. Mol. Biol.* **157**:105–132.
- [Jin(2000)] Jin X, Roberts CG, Nixon DF, Safrin JT, Zhang LQ, Huang YX, Bhardwaj N, Jesdale B, DeGroot AS, Koup RA Identification of subdominant cytotoxic T lymphocyte epitopes encoded by autologous HIV type 1 sequences, using dendritic cell stimulation and computer-driven algorithm, *AIDS Res Hum Retroviruses* **16**(1):67–76.
- [Korber(2000)] Korber B, Muldoon, J Theiler, F Gao, R Gupta, A Lapedes, B Hahn, S Wolinsky, T Bhattacharya, Timing the Ancestor of the HIV-1 Pandemic Strains, *Science* **288**:1789–96.

Word length: 18  
 Overlap peptide by: 10  
 Forbidden C-term characters: GPEDQNTSC  
 Forbidden N-term characters: Q  
 Shorten by: 3  
 Lengthen by: 2

# Gag

```
[Number of peptides generated: 67]
MGARASVLSggeLDrWEKiRLRPgGkKkYkLkHiVWASRELERFavmpgLLLEtseGCrQIIgQLqpsLqTGSSElrSlyNtVAtLYCvHgrIevkDTKeAl
-----K-S-----E-----I-----Gag Con-B
-----k-a-----r-l-----L-s-t--Q-me--A-k--T-----Gag HXB2
-----I-R-K-k-----h-m--l-----l-----k-mk--A--T-----Gag DH123
-----K-a-----r-l-----L-----k-i--ai-----E-----Gag Con-A
-----K-a-----rm-L-----r-R-----k-L-s--T--KK-I-----K--aiVV--Y--d-R-----Gag Con-C
-----K-a-----r-l-----L-----L-d--A--q-Ma--a--T--ik-f-----k-----Gag Con-D
-----I-----K-D-----q-RI--L-----D-L-----SAK--Q--V--A--Tq-ik-----I-----Gag Con-F1
-----k-a-----r-l-----l-----l-----a--q-me--a-k-----k-f-----Gag Con-F2
-----k-a-----r-l-----l-----l-----l-----q-i-----a-----Gag Con-G
MGARASVLSggeLDrWEk (18) [-0.54]
SggelDrWEKiRLRPgGk (18) [-1.37]
EKiRLRPgGkKkYkLkHi (18) [-1.46]
GkKkYkLkHiVWASREL (17) [-0.94]
KHiVWASRELERFav (15) [-0.11]
ASRELERFavmpgLL (15) [-0.02]
ERFavmpgLLLEtseGCr (17) [-0.47]
gLLLEtseGCrQIIgQL (16) [0.06]
eGCrQIIgQLqpsLqTGS (18) [-0.48] All C-term AAs forbidden
gQLqpsLqTGSSElrSly (18) [-0.76]
TGSSElrSlyNtVAtLY (17) [-0.25]
SlyNtVAtLYCvHgrIev (18) [0.33]
LYCvHgrIevkDTKeAl (17) [-0.39]
```

```

ievkDTKeAlekiEEEQnkskkkaggaaadtgnssgvsqNYPiVQNIqGQMVHQaiSPRTLnAWVKvEEKAFSPFIPMFSAALSEGATPQDInTnMLNTVGGHQAAAMgMLK Gag Con-B
--I-----D-----H-N-----I-----I-----L-----A-----sl-----i-----m-----i----- Gag HXB2
-d-----d-----i-----Q-t-----k-----A-----s-----i-----m-----i----- Gag DH123
--I-----D-----m-----...kgk-----i-----I-----T-----I----- Gag Con-A
--m-----d-----L-----qQ-T-----...KG-----l-----i-----L-----s-----I----- Gag Con-C
-d-R-----D-LQ-----D-hQQ-t-----...K-ke-----l-----i-----L-----s-----I----- Gag Con-D
--EV-KI-K-Qg-i-----...e-n-----A-----A-----A-----I----- Gag Con-F1
-g-----I-----I-----rQ-T-----KEkDnK-----A-----A-----A-----I----- Gag Con-F2
-I-----d-----I-----N-QQ-K-E--KKDN-----P-----L-----I-----I-----I----- Gag Con-G
-gir-----D-l-----i-----qg-t-----...kg-----k-----p-----L-----I-----m-----i----- Gag Con-H
--d-----d-----i-----qg-t-----...k-----k-----a-----a-----I-----m-----i----- Gag Con-J
--d-----d-----i-----qg-t-----...k-----k-----a-----a-----I-----m-----i----- Gag Con-K
ievkDTKeAlekiEEEQnkskkkagg (19) [-1.53]
lekIEEEQnkskkkagg (18) [-2.02]
nkskkkaggaaadtgnssgv (20) [-1.43]
aadtgnssgvsqNYPiV (17) [-0.44]
sqvsqNYPiVQNIqGQMV (18) [-0.37]
IVQNIqGQMVHQaiSPR (17) [-0.18]
GQMVHQaiSPRTLnAWVK (18) [-0.24]
SpRTLnAWVKvEEKAF (17) [-0.24]
WVKVvEEKAFSPFIPMF (18) [0.41]
SpRTLnAWVKvEEKAF (17) [-0.24]
AFSPEVIPMFSAALSEGATPQDInTnMLNTVGGHQAAAMgMLK (18) [0.73]
PMFSAALSEGATPQDInTnMLNTVGGHQAAAMgMLK (18) [-0.16]
GATPQDI nTnMLNTVGGHQAAAMgMLK (17) [-0.39]
nTnMLNTVGGHQAAAMgMLK (18) [-0.12]

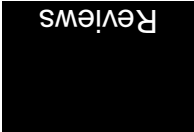
```

GHQAAMgMLK**ETT**NeE**EAa**EWDRlHPvA>Gp*ia*FQMr**e**PRGSDIAGT**rs**TL**g**egIqWMTnnP*pi*PVGeIYKRWIIIGLNkIVRMYS**Pts**II**L**DIr**Q**GPk**EP**FRDYVDRFYK**T**LR Gag Con-B  
 -----V-----M----- Gag HXB2  
 -----D-----p-----g-----D-----V-----k-----F----- Gag DHL23  
 -----D-----p-----a-----s-----v-----d-----V-----k-----F----- Gag Con-A  
 -----v-----a-----s-----V-----V-----F-----F----- Gag Con-C  
 -----D-----p-----Q-----s-----V-----d-----V-----K-----F----- Gag Con-D  
 -----D-----p-----a-----s-----V-----V-----K-----F----- Gag Con-F1  
 -----d-----m-----q-----p-----I-----r-----s-----V-----V-----F-----F----- Gag Con-F2  
 -----D-----p-----A-----G-----D-----D-----V-----K-----F----- Gag Con-G  
 -----D-----V-----V-----G-----G-----V-----V-----F-----F----- Gag Con-H  
 -----d-----v-----p-----a-----s-----v-----v-----f-----f----- Gag Con-J  
 -----d-----v-----p-----a-----s-----v-----v-----f-----f----- Gag Con-K  
 -----d-----p-----a-----s-----v-----v-----f-----f----- Gag Con-O  
 -----p-----a-----s-----v-----v-----f-----f----- Gag Con-Con  
 GHQAAMgMLKETIN**EE**Aa (18) [-0.55]  
 IKETTIN**EE**AaEWDRlHPv (18) [-0.88]  
 AaEWDRlHPvA>Gp*ia* (16) [-0.17]  
 IHPvA>Gp*ia*FGMr**e**PR (18) [-0.64]  
 iaPGQMr**e**PRGSDIA (15) [-0.65]  
 MrePRGSDIAGT**rs**TL (16) [-0.63]  
 SDIAGT**rs**TLgqI**g**WMM (17) [-0.17]  
 STLgqI**g**WMTnnP*pi*V (18) [-0.38]  
 WMTnnP*pi*PVGeIYKRWI (18) [-0.46]  
 PVGeIYKRWIIIGLNkIV (18) [0.56]  
 WIIIGLNkIVRMYS**Pts**I (18) [0.74]  
 IYRMYS**Pts**II**L**Dir**Q**GPk (18) [-0.21]  
 SILDir**Q**GPk**EP**FRDYV (17) [-0.75]  
 GPk**EP**FRDYVDRFYK**T**LR (18) [-1.39]



YVDRFYkTLraEQAsQeVKNwmTETLLVQNaNPDCKTLLKALGPaATLleMMTACQGVGGPqHKARVLAEMSGvtnsatIMmQrGhfNgrktvkGfncGKeGHIaknCrAPR Gag Con-B  
 -----S-----G-----I-----G-----I-----G-----I-----T-R-----Gag HXB2  
 -----t-----g-----s-----r-----aG-----s-----r-----I-----G-----,-----r-----I-----L-----r-----S-----Gag DH123  
 -----t-----d-----d-----r-----g-----s-----r-----g-----s-----Kgp-----r-----i-----L-----r-----Gag Con-A  
 -----d-----d-----g-----g-----G-----I-----K-----S-----k-----G-----R-----I-----R-----Gag Con-C  
 -----F-----T-----g-----D-----r-----g-----S-----I-----a-----n-----a-----K-----K-----G-----P-----R-----I-----R-----Gag Con-F1  
 -----T-----D-----D-----D-----T-----N-----S-----G-----A-----K-----K-----G-----P-----R-----I-----L-----R-----Gag Con-F2  
 -----f-----t-----t-----d-----g-----s-----i-----I-----a-----v-----a-----v-----a-----v-----k-----g-----P-----i-----R-----Gag Con-H  
 -----f-----t-----t-----d-----g-----s-----i-----I-----a-----v-----a-----v-----k-----g-----P-----i-----R-----Gag Con-K  
 -----f-----t-----t-----d-----g-----s-----i-----I-----a-----v-----a-----v-----k-----g-----P-----i-----R-----Gag Con-O  
 -----f-----t-----t-----d-----g-----s-----i-----I-----a-----v-----a-----v-----k-----g-----P-----i-----R-----Gag Con-Con

YVDRFYKTLraEQAsQeV (18) [-0.88]  
 LraEQAsQeVKNwmTETL (18) [-0.84]  
 eVKNwmTETLLVQNa (15) [-0.27]  
 mTETLLVQNaNPDCKTLL (18) [0.07]  
 NAnPDCKTLLKALGPaA (17) [-0.05]  
 TLLKALGPaATLleMMTA (18) [0.66]  
 aATLleMMTACQGVGGPqH (19) [0.07]  
 ACQGVGGPqHKARVLAEA (18) [0.01]  
 gHKARVLAEMSGvtnsa (18) [-0.19]  
 EAMSGvtnsatIMmQr (16) [-0.22]  
 mQrGhfNgrktvkGf (16) [-1.37]  
 tnsatIMmQrGhfNgrk (18) [-1.39]  
 mQrGhfNgrktvkGf (16) [-1.39]  
 rngrktvkGfncGKeGHI (18) [-1.27]  
 CfnCGKeGHIaknCrAPR (18) [-0.81]



```

HiaknCrAPRkkGCWkcGkGkGkMkdcteRQAnFLGkIwpsShkGRpgnFLQERPEPLAPpeesfrfgeettbpsgkqepidkelyPlasLrsLFgndPssg Gag Con-B
-T-R-----Y-----S-V-----P-----T-----Gag HXB2
--S-----E-----S-----A-----A-----K-----OR Gag DH123
-L-R-----s-----P-----A-i-gm---ia.Spp--qk-r-qn--v--K-----LL-- Gag Con-A
--Y-----a-----T...ap---k-r-..-t--k--s--L-- Gag Con-C
--R-----r-----q-----N-----N-----A-G---I-..-OK---t--K-----l-- Gag Con-D
--L-R-----r-----q-----N-----N-----A-G---Ia.-P--QK--Qa-i--K--S--l-- Gag Con-F1
--R-----r-----r-----N-----N-----A-G---IA.-P--Lk--P-----l-- Gag Con-F2
--L-R-----r-----s-----S-----S-----A-LG1--I..-P--K-----T-----S--L-- Gag Con-G
--Y-----s-----s-----n-----n-----a--gm--i-.-pr--tk--gg--T--k-----l-- Gag Con-H
--R-----r-----s-----s-----a--g---i-.-p--qk-----k-----l-- Gag Con-K
--Y-----s-----s-----a--g---i-.-p--qk-----k-----l-- Gag Con-O
HiaknCrAPRkkGCWk (16) [-1.32]
  rAPRkkGCWkcGkeGkGM (18) [-1.66]
    WkcGkeGkMkdcteRQA (18) [-1.70]
      HqMkdcteRQAnFLGkIw (18) [-0.98]
        RQAnFLGkIwpsShkGR (16) [-1.14]
          GkIwpsShkGRpgnFLQER (18) [-1.27]
            GRpgnFLQSRPEPTA (15) [-1.21]
              FLQSRPEPTAPpeesSfrf (18) [-0.98]
                tAPpeesSfrfgeettbpsgk (20) [-1.41]
                  geettbpsqkqepidkelyPlasLr (18) [-1.69]
                    sqkqepidkelyPlasLr (18) [-1.01]
                      kelyPlasLrsLFgndPss (18) [-0.36] All C-term AAs forbidden
                        LrsLFgndPssg (12) [-0.75]

```

# Pol

[Number of peptides generated: 136]  
 FFRRedlafggkArefseQtRansptrrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 ---L-----N-----V-----S-----  
 ---N-----G-S-----E-I-L-L-----K-----I-----K-----  
 ---N-q-E-k-----g-----S-d-wdg-----s-----e-----p-----t-----V-----DI-----K-----  
 ---n-----e-----p-----I-----r-----p-----e-----l-----n-----s-----v-----i-----k-----  
 ---n-----g-----l-----S-----R-----g-----p-----T-----e-----n-----l-----K-----  
 ---N-Q-E-K-p-----a-----S-----r-----g-----r-----E-----R-----p-----L-----r-----DI-----K-----  
 ---N-Q-E-K-----AS-----R-----RrG-----s-----p-----P-----E-----S-----L-----a-----r-----V-----R-----DI-----K-----  
 ---N-Q-E-K-----AS-----R-----RrG-----s-----p-----P-----E-----S-----L-----a-----r-----V-----R-----DI-----K-----  
 ---Q-re-k-p-A-----S-----R-----RrG-----d-----p-----e-----l-----c-----V-----E-----r-----I-----k-----  
 ---Q-RE-l-P-----S-----P-----R-----A-----R-----G-----S-----N-----R-----R-----d-----I-----D-----r-----K-----  
 ---e-s-rE-k-----s-----w-----r-----g-----p-----p-----d-----q-----p-----N-----v-----v-----v-----I-----K-----  
 ---n-q-e-k-----s-----r-----d-----g-----e-----s-----e-----s-----l-----di-----k-----  
 ---n-q-e-k-----s-----r-----r-----g-----p-----e-----s-----l-----di-----k-----  
 FFRRedlafggkArefseQtRansptrrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 afpgkArefseQtRansptrrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 refsseQtRansptrrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 RansptrrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 trrlgVwgrdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 grdnnslsEAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 EAGadrgatvsfPqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 PqiTLwQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 wQRPIVtikigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 kigQLKeALLDTGADDTVleemnLPgrWKPkMiGGI  
 lkeALLDTGADDTVleemnLPgrWKPkMiGGI  
 GADDTVleemnLPgrWKPkMiGGI  
 eemnLPgrWKPkMiGGI  
 [-0.68]

```

rWKPkMiGGIGGFkVrQydgilileICGHKAIgFVLVgPTPvNIIGrLLlTQIGCTLhFPISPIeTVPVkLkPgmdGPKvKQWPLtEEKIKaLvEITcEMeKKGKiskki Pol Con-B
-----V-----R-----S-----T-----T----- Pol HXB2
K-----K-----K-----M-----M-----T-----T----- Pol DH123
k-----p-----k-----m-----l-----ta-----e-----t----- Pol Con-A
K-----K-----h-----M-----M-----T-----d-----r----- Pol Con-C
K-----p-----Q-----Q-----Q-----I-----I-----TQ-----A-----l-----E-----R----- Pol Con-D
K-----e-----VA-----K-----K-----I-----I-----T-----l----- Pol Con-F1
K-----NEVP-----E-----K-----I-----M-----L-----I-----I-----r----- Pol Con-F2
k-----vc-----q-----k-----k-----m-----m-----t-----t----- Pol Con-G
k-----k-----k-----k-----k-----m-----m-----t-----t----- Pol Con-H
rWKPkMiGGIGGFkVr (17) [-0.14] LVGPTPVNIIGrLLlTQI (18) [0.72] nFPISPIeTVPVkLkPgmm (18) [0.16] gmdGPKvKQWPLtEEKIK (18) [-1.16]
GGIGGFkVrQydgilile (17) [0.41] IIGrLLlTQIGCTLhFPISPI (18) [0.88] TQIGCTLhFPISPIeTV (17) [0.58] TVPVkLkPgmdGPKvKQW (18) [-0.64] KQWPLtEEKIKaLvE (16) [-0.39]
kVrQydgilileICGHKAI (18) [-0.05] kVrQydgilileICGHKAI (18) [0.85] GHKAIgFVLVgPTPvNII (18) [0.85] IIGrLLlTQIGCTLhFPISPI (18) [0.88] TQIGCTLhFPISPIeTV (17) [0.58] TVPVkLkPgmdGPKvKQW (18) [-0.64] EEKIKaLvEITcEMeK (16) [-0.42]
lieICGHKAIgFVLV (15) [1.45] lieICGHKAIgFVLV (15) [1.45] GHKAIgFVLVgPTPvNII (18) [0.85] LVGPTPVNIIGrLLlTQI (18) [0.72] IIGrLLlTQIGCTLhFPISPI (18) [0.88] TQIGCTLhFPISPIeTV (17) [0.58] TVPVkLkPgmdGPKvKQW (18) [-0.64] gmdGPKvKQWPLtEEKIK (18) [-1.16]
KHKAIgFVLVgPTPvNII (18) [0.85] GHKAIgFVLVgPTPvNII (18) [0.85] LVGPTPVNIIGrLLlTQI (18) [0.72] IIGrLLlTQIGCTLhFPISPI (18) [0.88] TQIGCTLhFPISPIeTV (17) [0.58] TVPVkLkPgmdGPKvKQW (18) [-0.64] KQWPLtEEKIKaLvE (16) [-0.39]
-----k-----k-----k-----m-----m-----t-----t----- LVEITcEMeKKGKiskki (17) [-0.10]

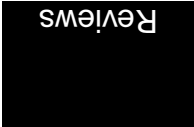
```

```

MeKEGKiskikiGPeNPYNTPvFAlkKKGdETkWRKLVDFRRLNkrTQDFWEVQLGIPHPeAGLkKkkkSVtVLDVGDDayFsvpLdkdfRKYTaFTIPsInNetPgiRYQYNVLPQGWK
-----E-----
-----I--N--R-----V--AA-----
-----i-----es-----t-----
-----t-----E-----
-----r-----e-----v-----
-----I-----r-----S-----
-----I-----S-----
-----E-----YE-----
-----i-----e-----
-----i-----
MeKEGKiskikiGPeNPY (16) [-1.31]
iskikiGPeNPYNTPvFAlk (18) [-0.22]
PYNTPvFAlkKKKdskWR (18) [-1.19]
IKKKdskWRKLVDFRRL (18) [-1.24]
WRKLVDFRRLNkrTQDFW (18) [-1.33]
ELNkrTQDFWEVQLGIPH (18) [-0.95]
FWEVQLGIPHPaGLKkkk (18) [-0.54]
PHPaGLKkkkSVtVLDV (17) [-0.32]
kkkSVtVLDVGDDayFsv (17) [0.15]
LDVGDDayFsvpLdkdfRk (18) [-0.39]
LDVGDDayFsvpLdkdfRk (18) [-0.28]
SvpLdkdfRKYTaFTIPsInNetPgi (20) [-0.46]
dFRKYTaFTIPsInNetPgi (20) [-0.75]
PsinNetPgiRYQYNV (17) [-0.93]
PgiRYQYNVLPQGWK (15) [-0.93]

```

PeptGen

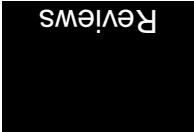


YQYNVLPQGWKGS	PaIFqsSMTkILEP	FkqNPdiViYQYMD	DLiYVGSdLEI	gQHRtKiEeLRghL	LrWgfETP	DkkHQKePP	FLLMNGYELH	HpdKkWTVP	IVLPEK	Pol Con-B
-----	-----	-----	-----	-----	-L-	-----	-----	-----	-----	Pol HXB2
-----	-----	-----	-----	-----	-LF--	Q-	-----	-----	-----	Pol DH123
-----	sk-e-i	-----	a-s	-----	-----	-----	-----	e-	-----	Pol Con-A
-----	a-e	-----	A-e-k	-----	-----	-----	-----	q-	-----	Pol Con-C
-----	E	-----	e	-----	-----	s-k-	-----	-----	-----	Pol Con-D
-----	c	-----	ak	-----	E-k	-----	-----	Q-d	-----	Pol Con-F1
-----	c-i	-----	ak-E	-----	E	-----	-----	Q-d	-----	Pol Con-F2
-----	c-i	-----	ak-E	-----	a	-----	-----	Q-d	-----	Pol Con-G
-----	-----	E-i	-----	a	A	-----	-----	Vk-	-----	Pol Con-H
-----	C	-----	ER-E	-----	E-R-K-E	-----	K-y	-----	-----	Pol Con-J
-----	h	-----	iK-em	-----	a	-----	e-k	-----	q-d	Pol Con-K
-----	-----	tk-e	-----	a	e-k	-----	-----	q-	-----	Pol Con-O
-----	-----	ak-e	-----	a	e-k	-----	-----	q-	-----	Pol Con-Con
YQYNVLPQGWKGS	PaIF (17) [-0.33]									
PQGWKGS	PaIFqsSMTkI (18) [-0.41]									
aIFqsSMTkILEP	Fk (16) [-0.07]									
MTkILEP	FkqNPdiViY (18) [-0.29]									
rkqNP	diViYQYMDLYV (18) [-0.66]									
IYQYMD	LYVGSdLEI (16) [0.01]									
DLiYVGSdLEI	gQHRtKi (17) [-0.49]									
LEI	gQHRtKiEeLRghL (18) [-0.76]									
KiEeLRghL	LrWgfETP	DkkHQKePP	FLLMNGYELH (18) [-1.56]							
LrWgfETP	DkkHQKePP	FLLMNGYELH (18) [-1.42]								
DkkHQKePP	FLLMNGYELH (18) [-1.42]									
PFLMW	GYELHpdKkWTVP	(16) [-0.31]								
YELHpdKkWTVP	IVLPEK (18) [-0.68]									

```

TVQP I V L P E K d S W T V N D I Q K L V G K L m W a S Q I Y a G I K V k g L C K L L R G E K A L T E v i p l T e E A E L E L A e N R E I L L k e p v H g Y y D P s k d l i a E i Q K G G G W T Y Q I y Q e p f k n L k T g K Y Pol Con-B
-----P-----R-----A-----I-----
-----e-----e-----a-----Divt-----d-----D-----
-----q-----p-----r-----a-----Div-----hd-----
-----s-----k-----E-----r-----r-----r-----h-----
-----Q-----d-----S-----P-----R-----i-----H-----r-----a-----D-----V-----a-----i-----i-----i-----h-----
-----Q-----d-----S-----P-----R-----i-----H-----r-----a-----D-----V-----a-----i-----i-----i-----h-----
-----V-----K-----Q-----d-----S-----P-----R-----i-----i-----A-----div-----K-----r-----r-----r-----p-----d-----
-----q-----d-----ED-----P-----i-----i-----A-----D-----I-----V-----R-----K-----A-----r-----E-----V-----L-----D-----h-----
-----q-----q-----d-----P-----v-----v-----div-----a-----a-----div-----d-----
-----q-----q-----P-----a-----a-----div-----hd-----
TVQP I V L P E K d S W T V N D I (18) [-0.13]
      EKdSWTVNDIQKLVGKLL (17) [-0.68]
            NDIQKLVGKLnWaSQIYa (18) [-0.24]
                  KLnWaSQIYaGIKvkgL (17) [-0.07]
                          IYaGIKvkgLCKLLRGEK (18) [0.14]
                                  kgLCKLLRGEKALTEvi (17) [0.19]
                                          RGTKALTEvipLTeEAELELAE (18) [-0.16]
                                                  vipLTeEAELELAE (18) [-0.17]
                                                          ELELAE (18) [-0.47]
                                                                  EILKEPVHGYYDFSKDL (18) [-0.64]
                                                                          gvYYDFSKDLIAEIQK (16) [-0.53]
                                                                              skDLIAEIQKGGGWTY (18) [-1.04]
                                                                                  iQKGGGWTYQIYQEPF (18) [-1.29]
                                                                                                                                 WTYQIYQEPfknLkTgKY (18) [-1.27]

```



```

pFknlkTgKYArmrGaHTNDvkQLtEaVQKiaTeesIvIwKtEPkfkLPIQkETWeawWteYvgATWIPeWEfvNTIpEIVKLVWQLLePePIvgAETIFVYVdGAAAr
-----T-----
-----V--VT--C-----R-----T-----S-----
-----K-S-----a-v--Vvm-----t--md-----D-a-----
-----K-t-----m-----R-----t--d-----a-----
-----K-S-----l-----R-----T--i-----i-----
-----RKS-----V-----l-----R-----i-----d-----t-----T--a-----S-----
-----RKS-----V-----R-----i-----R-----I-----I-----T--a-----T--a-----
-----KR-S-----aqV-----L-A-----R-----R-----T--h-----0-----H-----t--a-----Y-----
-----k-s-----v-----v-----R-----R-----R-----T--D-----D-----M-----
-----s-----Y-----Y-----t-----t-----T-----T-----a-----
pFknlkTgKYArmrGaH (-1.16)
gKYArmrGaHTNDvkQL (17) [-1.16]
GaHTNDvkQLtEaVQKia (18) [-0.46]
kQLtEaVQKiaTeesIvIw (18) [0.24]
KiatesIvIwKtEPkfkL (18) [0.11]
IwKtEPkfkLPIQkETW (17) [-0.81]
fkLPIQkETWeawWteYw (18) [-0.91]
TWeawWteYvgATWIPeW (18) [-0.91]
YvgATWIPeWEfvNTIpL (18) [-0.39]
eWEfvNTIpEIVKLVWQL (17) [-0.14]
pPLVKLVWQLLekePIvgA (18) [0.02]
YQLekePIvgAETfYV (16) [-0.12]
PIvgAETfYVdGAAAr (16) [0.11]

```



```

TFYVdGAAAnreTklgKAGYVtdrGRQkVvslltDFTTNOkTELgAihLALQDSGLFNIVtDSQYALGIgAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV Pol Con-B
-----N-----T-----Y-----Y-----e-----h-----S-----S-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol HXB2
-----S-----R-----N-----N-----e-----h-----S-----S-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol DH123
-----i-----i-----E-----E-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol Con-A
-----i-----i-----E-----E-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol Con-C
-----S-----K-----i-----i-----E-----E-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol Con-D
-----S-----K-----i-----i-----E-----E-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol Con-F1
-----S-----K-----i-----i-----E-----E-----i-----E-----p-----E-----A-----i-----p-----E-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e----- Pol Con-F2
Y-----i-----i-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e-----e----- Pol Con-G
-----S-----t-----K-----K-----i-----T-----I-----e-----a-----h-----e-----e-----e----- Pol Con-H
-----q-----q-----I-----e-----a-----h-----e-----e-----e-----e----- Pol Con-J
-----q-----q-----I-----e-----a-----h-----e-----e-----e-----e----- Pol Con-K
-----q-----q-----I-----e-----a-----h-----e-----e-----e-----e----- Pol Con-O
-----q-----q-----I-----e-----a-----h-----e-----e-----e-----e----- Pol Con-Con

TFYVdGAAAnreTklgKA (17) [-0.59]
AnreTklgKAGYVtdrGR (18) [-1.34]
KAGYVtdrGRQkVvslltDFTTNOkTELgAihLALQDSGLFNIVtDSQYALGIgAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (16) [-0.58]
drGRQkVvslltDFTTNOkTELgAihLALQDSGLFNIVtDSQYALGIgAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (17) [-1.49]
vsltDFTTNOkTELgAihLALQDSGLFNIVtDSQYALGIgAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (18) [-0.35]
NOKTELgAihLALQDSGLFNIVtDSQYALGIgAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (18) [-0.37]
IhLALQDSGLFNIV (15) [0.81]
LQDSGLFNIVtDSQYAL (18) [0.06]
NIVtDSQYALGIQA (15) [0.53]
SQYALGIQAQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (19) [-0.55]
AQPDkSESElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (18) [-0.54]
ElVsgIIEglIkKekVYLaWVPAHKIGIGNEgV (18) [0.11]
EglIkKekVYLaWVPAHKIGIGNEgV (18) [-0.50]
VYLaWVPAHKIGIGNEgV (18) [0.11]

```

```

HKGIGNEqvDkLVsAGiRkVLFldgIdkAQeeHEkYHsNWrAMASdFNlPpVvAKeIVAScdKQlKGEAmHGQVDesPGIwGLDCTHLeGkiILVAVHVASGYIeAEVITPa
-----D-----V-----
T--R--E--A--C--V--
s--s--r--t--i--
s--s--e--i--
q--i--n--N--
s--s--
S--s--q--r--n--I--
s--s--d--
s--s--r--n--i--i--
s--s--
HKGIGNEqvDkLVsAGi (18) [-0.24]
EqvDkLVsAGiRkVLF1 (17) [0.54]
SAGiRkVLFldgIdkA (16) [0.39]
VLFldgIdkAQeeHEkYH (18) [-0.89]
kAQeeHEkYHsNWrAMA (17) [-1.88]
kYHsNWrAMASdFNlPpV (18) [-0.71]
MASdFNlPpVvAKeIVA (17) [0.74]
PpVvAKeIVAScdKQlK (18) [0.18]
VAScdKQlKGEAmHGQV (18) [-0.22]
IKGEAmHGQVDesPGIw (17) [-0.20]
GQVDesPGIwGLDCTHL (17) [-0.04]
GIWqLDCTHLeGkiILVA (18) [0.74]
HLeGkiILVAVHVASGYI (18) [1.14]
VAVHVASGYIeAEVITPa (17) [1.10]

```

```

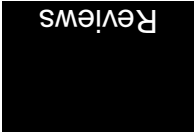
GYIeAEVTPaeTgQeTAYVFLlLkLAgRWPVktIHTDNGsNFtsttVkaAcWwagGikQeFGiPYNPQSGVveSmmkeLkkIIgQVRDQAEHLKtAVgMAVFIHNfkRK
-----GA-R-----
-----I-----
-----vv-----AA-----nvq-----
-----i-----v-----aa-----Q-----E-----
-----vv-----Aa-----
-----i-----a-----Q-----
-----a-----I-----aV-----Q-----i-----
-----a-----I-----aV-----Q-----i-----
-----i-----m-----aA-----D-q-----
-----A-F-I-----a-----GA-----D-----E-----
-----I-----V-----av-----d-----
-----i-----v-----aa-----n-q-----
-----i-----v-----aa-----q-----
GYIeAEVIPaeTgQeTAY (18) [-0.29]
  PaeTgQeTAYFLLkLAgR (18) [-0.24]
    AYFLLkLAgRWPVktIh (17) [0.36]
      AgrWPVktIHTDNGsNF (17) [-0.84]
        tIHTDNGsNFtsttVkaA (18) [-0.44]
          NftsttVkaAcWwagGik (17) [0.18]
            kAAcWwagGikQeFGiPY (17) [-0.04]
              GikQeFGiPYNPQSGVv (18) [-0.43]
                PYNPQSGVveSmmkeLk (18) [-1.18]
                  VveSmmkeLkkIIgQVR (17) [-0.24]
                    eLkkIIgQVRDQAEHLK (17) [-0.86]
                      gQVRDQAEHLKtAVgMAV (18) [-0.36]
                        HLKtAVgMAVFIHNfkRK (18) [-0.14]
  
```

AVFIHNEKRRKGGIGYSaGerIvdiIAEDiQeLeLQkQiTkQnfrVYYRdsrDPLWKgpAKLLWKGEGAVVIQDnsdIKVVPRRkakiIRDYgKQmagDDDCVAsRQDED Pol Con-B  
 -----N-----K----- Pol HXB2  
 -----S-----K----- Pol DH123  
 -----I-----i-----i-----g----- Pol Con-A  
 -----i-----i-----i-----k-----A-----g----- Pol Con-C  
 -----I-----i-----i-----v----- Pol Con-D  
 -----I-----r-----V-----E-----G----- Pol Con-F1  
 -----I-----s-----F-----NE-----i-----G----- Pol Con-F2  
 -----I-----s-----F-----NE-----i-----G----- Pol Con-G  
 -----I-----r-----I-----E-----p-----g----- Pol Con-H  
 -----I-----l-----e-----I-----e-----g----- Pol Con-J  
 -----i-----i-----i-----e-----g----- Pol Con-K  
 -----i-----i-----i-----e-----g----- Pol Con-O  
 -----i-----i-----i-----e-----g----- Pol Con-Con

AVFIHNEKRRKGGIGYSA (18) [-0.02]  
 RKGIGGYSaGerIvdiI (18) [-0.00]  
 sAGeRivdiIAEDiQtK (17) [0.05]  
 diIAEDiQeLeLQkQiTk (18) [-0.62]  
 tkeLQkQiTkIqnfrVYY (18) [-0.99]  
 TkiQnfrVYYRdsrDPLW (18) [-1.26]  
 YRdsrDPLWKgpAKLLW (18) [-1.08]  
 LWKgpAKLLWKGEGAVVI (18) [0.45]  
 LWKGEgAVVIQDnsdIKV (18) [-0.03]  
 VIQDnsdIKVVPRRkaki (18) [-0.51]  
 KVVPRRkakiIRDYgKQm (18) [-1.02]  
 KiIRDYgKQmagDDCVA (17) [-0.42]  
 KQmagDDCVAsRQDED (16) [-1.37]

**Vif**

[Number of peptides generated: 25]  
 menrwqvmiVWQVDRmrIrtwksLVkhhmyisgkagwfyrrhHye**st**hPr**is**sevhiPlgdarlvitI**Y**wG**l**htger**d**Wh**l**gg**g**vs**i**ewrkk**r**Y**st**q**v**dp**d**l Vif Con-B  
 -----V-----R-----P-----S-V-----R-----R-----h-----l-----i----- Vif HXB2  
 -----V-K-----V-k-----R-kv-----vr-----k-----h-----l-----r-----i----- Vif DH123  
 -----n-----n-----V-rr-n-----R-kv-----e-----k-----q-----H-----l-----r-----g----- Vif Con-A  
 -----L-----k-----n-----v-k-q-l-----dcp-K-----e-----vk-----e-----r-----r-----g----- Vif Con-C  
 -----n-----n-----Y-hv-K-----R-R-a-----R-V-----V-----Fq-R-k-----Ee-k-----p-----Q**G**-R-I-g----- Vif Con-D  
 -----n-----n-----v-K-----R-V-----R-V-----R-V-----k-----Vr-----K-----q-----H-----Q**r**-----i----- Vif Con-F1  
 -----n-----n-----Y-nv-k-rk-l-----d-----k-----e-----v-----v-----e-----v-----L-k-r----- Vif Con-F2  
 -----n-----n-----Y-v-K-nr-y-----h-K-----I-----v-----v-----e-----v-----m-----h-----Y-k-----i----- Vif Con-H  
 -----n-----n-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k----- Vif Con-G  
 -----n-----n-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k----- Vif Con-K  
 -----n-----n-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k----- Vif Con-O  
 -----n-----n-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k-----v-k----- Vif Con-Con  
 menrwqvmiVWQVDRmrI (18) [-0.31]  
 iVWQVDRmrIrtwksLVk (18) [-0.24]  
 rIrtwksLVkhhmyisgk (18) [-0.73]  
 VkhHmyisgkagwfyrrh (18) [-0.87]  
 gkagwfyrrhHye**st**hPr (18) [-1.82]  
 rhHye**st**hPrissevhi (17) [-1.27]  
 hPrissevhiPlgdarlv (18) [-0.04]  
 hiPlgdarlvitI**Y**wG**l**h (18) [0.33]  
 lvitI**Y**wG**l**htger**d**Wh**l** (18) [-0.21]  
 lhtgerdWhl**g**ggvs**i**ew (18) [-0.73]  
 hlgggvsiewrkk**r**Y (15) [-1.22]  
 vsiewrkk**r**Y**st**q**v**dp**d**l (18) [-1.12]



```

rYstqvbdladgLIHlyyfdCfsesAirnaIlghivsprCeygaGhkvgsLgYLAalaaltpkkkikPPiPsvtkltEDRWnkpgtkghrgshmtmngH Vif Con-B
-----E-----D-----K-L-----R-----V-R-----A-----SH-----Vif HXB2
--i-----h-----d-----k-----qv-----s-----k-----v-----r-----I-----R-----Vif DH123
--G-----mh-----ad-----k-----i-----d-----T-----K-----r-----V-----R-----Y-----Vif Con-A
--G-----m-----a-----k-----t-----n-----t-----a-----r-----r-----e-----R-----Y-----Vif Con-C
--R-----I-----I-----K-----r-----I-----n-----T-----a-----t-----q-----V-----e-----R-----Vif Con-D
--R-----h-----h-----d-----k-----R-----N-----T-----V-----R-----V-----N-----Y-----Vif Con-F1
--i-----h-----h-----d-----k-----q-----r-----p-----K-----V-----tr-----r-----r-----R-----emp-----Vif Con-F2
--r-----g-----mh-----d-----k-----q-----d-----T-----a-----rr-----q-----V-----r-----e-----Vif Con-G
k-r-----i-----h-----h-----d-----k-----r-----y-----k-----q-----T-----va-----p-----v-----q-----Vif Con-H
k-----i-----h-----h-----d-----k-----qr-----r-----vk-----rt-----q-----r-----Vif Con-K
-----h-----h-----d-----k-----k-----t-----a-----t-----q-----v-----r-----Vif Con-O
rYstqvbdladgLIHly (18) [-0.50]
dladgLIHlyyfdCf (15) [0.40]
LIHlyyfdCfsesAirna (18) [0.33]
CfsesAirnaIlghiv (16) [0.86]
irnaIlghivsprCeyga (18) [0.04]
ivsprCeygaGhkvgsL (18) [-0.36]
yqaGhkvgsLgYLAala (18) [0.06]
gslgYLAalaaltpkkkik (18) [0.53]
aaltpkkkikPPiPsvtk (18) [0.01]
ikPPiPsvtkltEDRWnk (18) [-1.02]
tkltEDRWnkpgtkghrgshmtmngH (18) [-2.39]
NkpgtkghrgshmtmngH (18) [-2.19]
hrghshmtmngH (10) [-1.8]

```

# Vpr

[Number of peptides generated: 12]

```

MegaPedaGpqrpepynewtLelLeeLknEAvrhFpriWlhslgghiyetygdTWaGveaiRiLQQLLfiHFriGcGHSRigigtgrrarmgasrs Vpr Con-B
-----H-----G-----R-----V-R----- Vpr HXB2
---V-Q-----P-G-Y-----L-----IR-T----- Vpr DH123
-----m---d-h---p-g---n---E-----t---v-----irg-v-d-sg-- Vpr Con-A
-----q-----p-g-y-----t---l-----lr----- Vpr Con-C
-----S-----y-----r-----s----- Vpr Con-D
-----a-----p-g---N---E-----r-----ir---V----- Vpr Con-F1
---P-----h-----EY-G---N---r-----H-----IR---i---S--- Vpr Con-F2
-----a-----p-g---N---E-----pr---v-d-pg-- Vpr Con-G
---H-0-----s-----p---q---N---e---l---T---R---V---P--- Vpr Con-H
-----h-----i---r-----p---d---n---E-----h-----ip---g----- Vpr Con-K
-----a-----p---g---y-----e-----v-----nrr---g----- Vpr Con-O
-----p---g---n---e-----ir---v----- Vpr Con-Con

MegaPedaGpqrpepynew (18) [-2.21]
GpqrpepynewtLelLeeL (18) [-1.13]
  eWtLelLeeLknEAvrhF (18) [-0.57]
    eLknEAvrhFpriWlhsl (18) [-0.47]
      hFpriWlhslgghiyety (18) [-0.48]
        slgghiyetygdTWaGv (17) [-0.37]
          etYgdTWaGveaiRiL (17) [0.34]
            aGveaiRiLQQLLfiHF (18) [1.34]
              iLQQLLfiHFriGcGHSR (18) [0.33]
                hFRigCqHSRigigtgrr (18) [-1.02]
                  SRigigtgrrarmgasr (17) [-1.38]
                    grrarmgasrs (11) [-2.13]

```

## Vpu

[Number of peptides generated: 9]  
 mgsLqilaivalvvaaiiaIvVwsiivfleyrkiilrQrkidrlidrireraedsGNeSegdqeelsalvemGhhapwvdvddI Vpu Con-B  
 T-PIP-V-----I-----L-----L-----L-----K-----D-----D-----R--L-----I--- Vpu HXB2  
 --P-V-----L--V-----L--g--k--l--k-----e-----d--t-----i---nYDLG-dNn- Vpu DH123  
 -SP-e-w--g-i-l-l-l-----T-g--k--l--k-----e-----d--t-----i---nYDLG-dNn- Vpu Con-A  
 akvdyr-gvg--l-l-----t-aY-----l-----w--k-----t-----tm-d--LRLl--n-- Vpu Con-C  
 --p-v-----l-----l-----t-----c-rlk-----w-----Y-----m Vpu Con-D  
 -Sh-laig-a-I-L-----T-Y--K-LV-----N--Yk-----A--A--G--pFI-G-iNN- Vpu Con-F1  
 -p-sLaVv-a-YI-L-----I--T--Y--K-----KR-N--YE--I-----A--A--G--V-PLI-G-INN- Vpu Con-F2  
 --p-e-s--g-I--f-A--i--t-----e-Rk-k-ek-l-----t-----at-m--dFD--VG-N- Vpu Con-G  
 -YI-G.G-G-----F-----T--Y-----LvK-k-----e--i-----D--t--k-----LnLGY-A-- Vpu Con-H  
 -ip--a-----I--iFL--cm-T--Y--K-L-----K-----D--Td--Ad--r-P-DL-N-N-- Vpu Con-J  
 -v--isg.-i--a-l-l-I--t-aY-----v--kr-nw-----a--a-ig-t--liLgSisn- Vpu Con-K  
 --p-e-----g-i-l-----t-l-----kk-----l-----t-----d-----n Vpu Con-O  
 --p-----i--l-----t--Y-----k-----t-----t-----ld-----n Vpu Con-Con  
 mgsLqilaivalvvaaii (18) [2.41]  
 ivalvvaaiiaIvVwsiiv (18) [3.17]  
 iiaIvVwsiivfleyrkiil (18) [1.84]  
 ivfleyrkiilrQrkidrl (18) [-0.28]  
 ilrQrkidrlidrirera (18) [-0.97]  
 rliidrireraedsGNe (18) [-1.66] All C-term AAs forbidden  
 raedsGNeSegdqeelsa (18) [-1.77]  
 SegdqeelsalvemGhha (18) [-0.69]  
 salvemGhhapwvdvddI (17) [-0.15]



# Tat

[Number of peptides generated: 13]

```

MepVdprLEPWkHPGSGPktactnCyCKKCCfHCqVCFttkgLgIsyGrKKRRrRRrapqdsqthQvsLskQpasgprDDptgppkeskkkVerEtEtDpvdg Tat Con-B
-----I--A-----H-N--A-----T-----F-----0-----H Tat HXB2
-----L-----K--S-H--D-A-----T-----H Tat DH123
-d--n--n--t-p-nk-----y-----ln-----gt--s-kd--npip--ip-tq-is--e-----ska--rf- Tat Con-A
-----n--n--s-y-l--q-----s--ps-ed--nli--lprtq-----se-----sK-----f- Tat Con-C
-d--n--n--r-p-Nk-----y-----i-----p--gg-a--dpiP--S----- Tat Con-D
--l--n-D--N-----pL--r-----yw--a-----h-T--S--i--dpvP--l--a-----E--SKak--c- Tat Con-F1
--V--kiD-----E-P-Nk-----r-----pL--r-----T--s-ki--DPVP--L--T--k--E-----SQ----- Tat Con-F2
--D--k--n--p-Nk-----v--w-----LN-----kh--gt--s-Kd--nVPV--lptT--N-----e-aSK-----f- Tat Con-G
--D--ng--N-----q--N-----Y--l--Lk-----s--aT-aSl-D--nhIp--L-RTh----- Tat Con-H
-----nr--N-----Nq--r--Y--i--lk-----S--PG-K--DLIP--L--TQRK-----E-----E--SKA-P-RF- Tat Con-J
-----n--n--t--nk-----y-----lk-----tT-ya-kn--Dpip--L-----g-----a--SK-K-- Tat Con-K
-----n--n--nk-----y-----l-----t--s-kd--dpip--l--t-----e--ska-----f-w Tat Con-O
MepVdprLEPWkHPGSGPk (19) [-1.56]
PWkHPGSGPktactnCy (17) [-1.22]
SQPKtactnCyCKKCCfH (18) [-0.55]
nCyCKKCCfHCqVCFttk (18) [0.01]
fHCqVCFttkgLgIsyGr (18) [0.18]
tkgLgIsyGrKKRRrRRr (18) [-2.11]
GrKKRRrRRrapqdsqth (18) [-2.98]
RrapqdsqthQvsLsk (16) [-1.56]
sqthQvsLskQpasgpr (17) [-1.35]
LskQpasgprGDptgpk (17) [-1.57]
sqprGDptgppkeskkkV (17) [-2.02]
tgpkeskkkVerEtEtDpvdg (19) [-1.88]
VerEtEtDpvdg (12) [-1.68]

```

# Rev

[Number of peptides generated: 15]

MAGRsgdsEeLlktvrlkflYgSNppPspeTrgARRNRRRRWRreFQRqirsisiSgwiLstYlGrpaepVpLgLPPlerLtLdcneDCGtsGtgGvGspgilv	Rev Con-B
-----IR-----L-----L-----L-----L-----NH-----S-----DA-----T-----	Rev HXB2
-----E-----D-----N-----L-----L-----L-----L-----L-----C-----C-----C-----C-----	Rev DH123
-----RAi-i-i-----y-k-k-s-----K-----D-l-eR-----C-----h-----s-----r-----vs-----	Rev Con-A
-----a--qA--i-i-----Y-e-----k-----A-----h-----eR-----c-----i-----hi-----S--sg-----	Rev Con-C
-----d-----A-----i-----i-----A-----h-----gER-----se-----r-----N--s-----	Rev Con-D
-----t-----A--y--I-----Y-k-----Y-k-----A-----a-----QR-----sC-----e-----HIN--S-----	Rev Con-F1
-----r-----A--Y--I-----Y-p-----Y-p-----A-----h-----ER-----ac-----q-----r-----h-----S--g-----	Rev Con-F2
-----st-----ra--i-i-----y-E-a-----y-E-a-----A-----a-----ER-----tsc-----p-----	Rev Con-G
-----a--q--gVckI--i-----Y-K-N-S-----Y-K-N-S-----A-----N--D-----ER--p--SC-----I-----R-----S-----N-----	Rev Con-H
-----DQ--LA--i-i-----Y-k-----s-----Y-k-----s-----A-----A-----qRl--ac-----S-----i-----s--N--D--p--kg-----	Rev Con-J
-----R--p--qQ--Tp--i-i-----y-k-----k-----a-----h-----er-----c-----s-----h-----s-----	Rev Con-K
-----ra--i-i-----y-k-----k-----a-----h-----er-----c-----s-----h-----s-----	Rev Con-O
-----a--i--i-----y-k-----k-----a-----h-----er-----c-----s-----h-----s-----	Rev Con-Con
MAGRsgdsEeLlktvrl (18) [-0.59]	
dEeLlktvrlkflY (15) [0.13]	
ktvrlkflYgSNppPs (17) [-0.51] All C-term AAs forbidden and proline at position 18	
fLYgSNppPspeTrgAR (18) [-1.39]	
PspeTrgARRNRRRRWr (18) [-2.82]	
ARRNRRRRWRreFQRqir (17) [-3.15]	
RWreFQRqirsisiSgwiL (17) [-1.15]	
RqirsisiSgwiLgr (17) [-0.12]	
gwiLstYlGrpaepVpL (17) [0.27]	
LGrpaepVpLqLPPlerL (18) [-0.16]	
pLqLPPlerLtLdcneDC (18) [-0.39] All C-term AAs forbidden	
rLtLdcneDCGtsGtgGv (18) [-0.52]	
DCGtsGtgGvGspgilv (17) [0.15]	

```

tgGvGspqilvESpavIesGtKHe Rev Con-B
-----T----- Rev HXB2
-----T----- Rev DH123
--r--vs---v--g---n Rev Con-A
-----0sGk-c--g--a-K- Rev Con-C
-----s-----d---e-g Rev Con-D
..-----Sg--ht----- Rev Con-F1
..-----S---h-a-G-i-- Rev Con-F2
-----r---f---sV--G--- Rev Con-G
..-k-----sl--st--GT--- Rev Con-H
--D---SG-FCM--GA--- Rev Con-J
-.eL-----p--PcT-----d-- Rev Con-K
-.-----s-----i-g----- Rev Con-O
-.-----s-----g----- Rev Con-Con
tgGvGspqilvESpavL (17) [0.57]
      qilvESpavIesGtKHe (17) [-0.20]

```

# Env

[Number of peptides generated: 114]

```

MrvkgrknyghlwrwGmlGmlmicsaaeglwVvYVYGVVwKEatTTLFCASdAkaydeVHNvWATHAcVPTdPnpggevvlenVtenfnmwKNnmveqMh
---EKYQHWRWG-----T-K-----N-----IL-----D-----
---M-----KG--L--I-----i--ii-----n-----d--e-----ih-----e-----d--d---
---m--qr--c--qW--i--iLGFW-----nmgn-----k-----e-----s--e--a--l-----i--e-----a--k-----d---
---r--qr-----k--i--i--t--dn-----s--e--a--l-----s--Er--a-----p--L--G--s--k-----D---
---R-MQR-W--GK--LLF--i--I--n--dn-----p-----s--k-----i-----s--p-----d---
---REMQR-W--G--LLF--i--I--as--ad-----Li--lvi-----snn-----ed--d-----s-----s--p-----d---
---.retQr-W--t--g--ilf--i--i-----n-----d-----s-----l-----s--p-----d---
---r-mQr-W--gkr--ilf--i--i-----n-----d-----s-----l-----s--p-----d---
---retQr-W--t--g--ilf--i--i-----n-----d-----s-----l-----s--p-----d---
---r--qr--w-----ilf--i--i-----n-----d-----s-----l-----s--p-----d---
Mrvkgirknyghlwrw (16) [-1.29]
rknyghlwrwGtmllGml (18) [-0.48]
rwGtmllGmlmicsaa (16) [1.25]
lGmlmicsaaeglwVvY (18) [1.28]
aaeglwVvYVYGVVwK (17) [0.36]
TvYVYGVVwKEatTTLF (17) [0.31]
VwKEatTTLFCASdAkay (18) [0.03]
LFCASdAkaydeVHNvWATHAcV (18) [-0.11]
aydeVHNvWATHAcV (16) [0.00]
HNvWATHAcVPTdPnpggev (19) [-0.65]
vPTdPnpggevlenV (15) [-0.39]
npggevlenVtenfnmwK (18) [-0.78]
nVtenfnmwKNnmveqMh (18) [-1.10]

```

```

wkNmvmveqWnhedIsLWdgsLKPeVkiLPLCvtLnctdlnmtnnttssgkmeiknCSfnittsirdkvqkeyalfyklvvpidndntsyrlisCntsviEQACPkv Env Con-B
---D---S-K---K-D--N---RM---S---G---F---II---RH---RN.IT---TL--- Env HXB2
---H---K-G-LKGTKIIG-.R---V-KN-I--K---k-v-s-r--q-e-nssq--N---a--- Env DH123
---t---snv-tnt-n-...-.re---m-El--k-v--r--i-le-nse--n---a--- Env Con-A
---d-d---nv-tts-n...mg...-e-p-am---ev-k-qvh--r-i--ns-n--n---a--- Env Con-C
---t---natstsga-i-tl-E-p-A-Q---m-Ev--kl-vh--r-I--n-ssse--n---T--- Env Con-D
----D--l---kaiItmsstlapl-tiqe---E---qr-qq--r-e-.kNSst---Tl--- Env Con-F1
---e---nv-nt-nt...t-ia---e---kk---r---dn.sn--n-v-t-k--- Env Con-F2
---d--t---vt..s..nsdtnsnatvispd---E-k-kKq---r---.sesnke.-N--- Env Con-G
---t---yygtrrnatniratsp.a---Elk--KKqvs--r---nssn--N---tv--- Env Con-J
---d--t---vt..s..dtnsnaispd---E-k-kKq---r---.sesn--N---t--- Env Con-K
---t---nv-ttst-n---l---p---e---k-v---r---nssn--n---t--- Env Con-Con
wkNmvmveqWnhedIsLW (17) [-0.58]
eqWnhedIsLWdgsLK (16) [-0.76]
iIsLWdgsLKPeVkiLPL (18) [0.54]
LKPcVkiLPLCvtLnctdl (19) [0.78]
LCvtLnctdlnmtnntts (18) [-0.26] All C-term AAs forbidden
dlnmtnnttssgkmeik (18) [-1.75]
tsssgkmeiknCSf (18) [-1.04]
ekgeiknCSfnittsir (17) [-0.64]
CSfnittsirdkvqkeya (18) [-0.57]
irdkvqkeyalfyklvvpidndntsy (19) [-0.19]
yalfyklvvpidndntsy (19) [-0.17]
vpidndntsyrlisCntsv (19) [-0.26]
yrlisCntsviEQACPkv (18) [0.37]

```

```

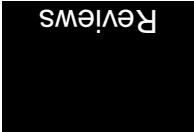
svitQACPKvsfepIpihycapagfaillkCndkkFngLppCtnVstvgCtHgirPvvvStgllLlngSLaeeevvirsenftdnaktIivqLnesveinctrpnmntksihhi Env Con-B
-----N-T-----K-----e-----n-t-----y-----r-----Y-----D-----d-----t-----r-----t-----r-----r-----r-----r-----r-----r-----R-R- Env HXB2
-TL-----K-----e-----n-t-----y-----r-----Y-----D-----d-----t-----r-----t-----r-----r-----r-----r-----G-TL Env DH123
-a-----d-----r-----WD-----D-----t-----k-----k-----k-----k-----k-----k-----k-----k-----k-----k-----k-----k-----k-----k-----r- Env Con-A
-a-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r- Env Con-C
-T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T-----T- Env Con-D
-t-----k-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r- Env Con-F1
-t-----k-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r-----r- Env Con-F2
-e-----N-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t- Env Con-G
-t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t- Env Con-J
-t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t- Env Con-K
-t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t- Env Con-K
-t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t-----t- Env Con-Con
SvitQACPKvsfepIpih (18) [0.43]
  kvsfepIpihycapagfa (18) [0.49]
    ihycapagfaillkCndkk (18) [0.04]
      failkCndkkFngLppCtnV (20) [-0.23]
        FngtgpCtnVstvgCtH (17) [-0.21]
          tnVstvgCtHGirPvv (16) [0.26]
            vqCtHGirPvvSTqLL (17) [0.71]
              rpvvSTqLLngSLa (15) [0.64]
                TqLLngSLaeeevvir (17) [0.35]
                  Slaeevvirsenftdn (18) [-0.46]
                    irsenftdnaktIivqL (17) [-0.12]
                      dnaktIivqLnesvei (16) [0.07]
                        ivqLnesveinctrpnmnt (20) [-0.86]
                          nctrpnmntksihhi (15) [-1.49]

```

```

nmnrksihigpgrafytlgtdiirgAhCnisrakwnntLkqiviklregfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti Env Con-B
-R-R-QR-GRAFV-IGK-NM-----AS-----KV-H-----RV-E-----K-E-----K-----I-K-----T-----S-----KK-----GTEGSYN----- Env HXB2
-G-TL-----V-----K-----V-----se-k-qkvakg-ty-n-----i-tn-----l-TT-----sg-----dqe-s----- Env DH123
-r-----q--a-d-----gt--e--qrVsk-a-h-p--k-ap-----l--tt-----r-----sk-----y-----sng----- Env Con-A
Y---grp-l-q-l-rs-----K-----gtQ--k--ayVka--ksh-p--k-----L-T-----r-----d-sg--d-----sng----- Env Con-C
-l-----q--a-----K-----kq-yd--ik-atEFkkyh.--sVa-qp-a--l-TT-----R-----y--si-----h-r--Vadit-htN----- Env Con-F1
-r-----q--a-d-----v-gt--em-gkvkaq-gki-n-----t-s-a--l--tt-----r-----sg--s-t-----sng----- Env Con-F2
-t-----m--q--a-d-----ged-----rrvkk-----h--h--i-tp-----r-----sk-----sk----- Env Con-G
S-----a-d-----ged--p--nrVkee-kkh-k-i-i-qp-----vth-m--r-----k-----dt----- Env Con-J
-m--q--a-d-----ged-----rrvkk--h--i-tp-----r-----sk-----sk-----nt----- Env Con-K
nmnrksihigpgrafy (17) [-1.07]
ihigpgrafytlgtdiirgAhCnisrakwnntLkqiviklregfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.52] All C-term AAs forbidden
rafytlgtdiirgAhCnisrakwnntLkqiviklregfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-1.03]
rafytlgtdiirgAhCnisrakwnntLkqiviklregf (17) [-0.37]
  rafytlgtdiirgAhCnisrakwnntLkqiviklregf (17) [-0.45]
    AhCnisrakwnntLkqiviklregf (17) [-0.49]
      kqiviklregfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.69]
        kqiviklregfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.74]
          egfgmkIvFngsGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.57]
            ngssGGDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.10]
              GDpeivmhsfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.03]
                sfnCgGEffYCnttgLFnstmntegsnntdti (17) [-0.10]
                  fYCnttgLFnstmntegsnntdti (17) [-0.52] All C-term AAs forbidden
                    LFnstmntegsnntdti (17) [-1.03]

```



```

tegsnntdtitlpCrIkqiNmwoVgkamYAppPirGqirCsSnIGllLlrdggnnntteIFrpggdmrdnwrseLykYKvkiEplgVAPITkakarrrvvgrEkrav Env Con-B
STEGS-S-----K-----S-----W-----g-v-e-----s-n-t-----s-n-t-----RV--I----- Env HXB2
GTEGSYN-----S-----K.-SS-----s-n-t-----ttt--t-----e-k-----r----- Env DH123
dqe--s-----ra-q-----a-n-t-k-----i-s-n-t-----i-----r-----e----- Env Con-A
.....ns-----g-r-----e-n-n-----a-n-t-N-----i-s-n-t-----E-----q-----a----- Env Con-C
.....sng-----v-----r-----a-A-n-t-N-----i-skidek-L-i-----ts-n-t-----nrsetg--t-----r-R-----e----- Env Con-D
dit-htN-----r-V-----Ri-q-----a-k-Q-t-----a-n-t-r-----r-g-I-----a-n-t-t-----i-----dthe-t-----q-----i-----r----- Env Con-F1
.....N-----k-----Vr-r-q-----a-n-t-r-----i-----r-----a-n-t-t-----mi-----rsetg--t-----g-----l-----r----- Env Con-F2
sk-----.....ent-----k-----I-----a-s-n-r-----i-----rsetg--t-----g-----l-----r----- Env Con-G
.....dt-----r-----q-I-----a-n-t-t-----i-----rsetg--t-----g-----l-----r----- Env Con-J
.....nt-----i-----r-----q-I-----a-n-t-t-----i-----rsetg--t-----g-----l-----r----- Env Con-K
.....n-----r-----a-n-t-----s-n-t----- Env Con-Con
tegsnntdtitlpCrIk (17) [-0.75]
dtitlpCrIkqiNmw (16) [0.21]
    CrIkqiNmwoVgkamY (18) [-0.17]
        mwQeVgkamYAppPirGqi (18) [-0.25]
            mYAppPirGqirCsSnI (16) [-0.17]
                rGqirCsSnITGllLlLr (17) [-0.08]
                    SnITGllLlrdggnnnt (18) [-0.76] All C-term AAs forbidden
                        trdggnnntteIFr (15) [-1.71]
                            nnnntteIFrpggdmr (17) [-1.47]
                                iFrpggdmrdnwrseLy (18) [-1.13]
                                    mrdnwrseLykYKvki (17) [-0.99]
                                        eLykYKvkieplgVA (16) [0.20]
                                            vvkieplgVAPITkakarrrv (18) [0.01]
                                                VAPITkakarrrvvgrEkrav (18) [-0.93]

```



```

RvvqrEkravgiqamfIGFLgaAGSTMGaasmLLTvQaRgllsGIvGQmmLLrAieaQQhllqLTwGikQLQarvLaverYlLkdQgLLgiwGCSgkLlctTavPwnasw Env Con-B
-----L-----I-----M-----Q-----T-----T-----
--v--v--i--I-----s-----k-----n-----s-----
--e--v--v--I-----s-----m-----t-----i-----h-----n-----s-----
--e--i--l--v-----aI--a-----S--K-----I-----p--t-----s-----
q--a--M--l-----i-----i-----v-----I-----s-----N-----s-----
q--a--M--l-----i-----i-----v-----I-----s-----p--t-----s-----
--e--l--v-----I-----v-----I-----s-----r-----r-----n-----t-----
--l--l--I-----I-----s-----I-----r-----r-----n-----s-----
-----v-----i-----s-----s-----n-----s-----
RvvqrEkravgiqamfI (17) [0.24]
ravgiqamfIGFLgaA (16) [1.56]
  amfIGFLgaAGSTMGaa (17) [1.35]
    gaAGSTMGaasmLLTvQa (18) [0.69]
      vQaRqllsGIvGQmmLL (18) [0.93]
        AasmtLLTvQaRqllsGIv (18) [0.02]
          GIVGQmmLLrAieaQQh (18) [-0.65]
            vQaRqllsGIvGQmmLL (18) [0.47]
              LLrAieaQQhllqLTw (17) [0.05]
                aQQhllqLTwGikQLQa (18) [0.09]
                  TVwGikQLQarvLaverY (18) [0.04]
                    IQarvLaverYlLkdQgLL (18) [-0.04]
                      erYlLkdQgLLgiwGCSgk (18) [-0.67]
                        LlgwGCSgkLlctTav (17) [1.36]
                          sgkLlctTavPwnasw (16) [0.28]

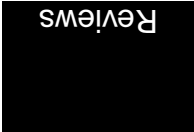
```

```

tTavpwnaswsnksl dgiwnnmTmwewereidnytsliytLlieesgqQekNegeLleLdkwaslwnwf dItEnMLwYikiFimivgGLvGlrivfavLsiivnrVRggy Env Con-B
-----E---HT---D---N---N---HS-----N-----L-----I-----S-----                               Env HXB2
--T---T---G---Q---K---N---G---N---A-----N---S-----I-----S-----                               Env DH123
--n---s---qse-d---Lq-dk-s---qi-q-----d-a-n-----s-----I-----i-----                               Env Con-A
--n---s---qee-d---q-d---s---nt-r-l-d---kd-a-s-qn-----I-----i-----                               Env Con-C
--n---s---eee-q-----g-s-i-----s-q-----I-----I-----                               Env Con-D
--N---s---q-E-----q-K-s-Skt-r-----a-----s-----I-----K-----                               Env Con-F1
p-t---s---y-d-d---q-K-----kT-r-sA---r-d-A--D--s-S-r-----i-s-I-----i-----                               Env Con-F2
--n---t---yne-d---i-d---s---gg-s-----D-A-----I-----I-----                               Env Con-G
--n---s---qse-----q-k-r-r-s-t-r-----i-----q-k-n-gt-s-a-----d-a-----s-----i-----i-----                               Env Con-J
--n---s---yee-e-----g-k-n-gt-s-a-----d-a-----k-----i-----a-----v-----                               Env Con-K
--n---s---yee-e-----g-k-n-gt-s-a-----d-a-----s-----i-----i-----                               Env Con-J
--n---s---q-e-d---q-k-s---qt-s-----d-a-----i-----i-----                               Env Con-Con
tTavpwnaswsnksl dgi (18) [-0.50]
swnksldgiwnnmTwm (17) [-0.89]
dgiwnnmTmwewerei (16) [-1.23]
mTmwewereidnytsliy (18) [-0.67]
eidnytsliyTLieesgq (18) [-0.50] All C-term AAs forbidden
iyTLieesgqQekNegeL (19) [-1.51]
nqQekNegeLleLdkwa (17) [-1.79]
eqeLleLdkwaslwnwf (17) [-0.51]
dkwaslwnwfdItEnMLwY (18) [-0.47]
wfdItEnMLwYikiFimiv (18) [1.24]
wYikiFimivgGLvGLrivi (18) [1.74]
ivgGLvGLrivfavLsiiv (18) [2.44]
rivfavLsiivnrVRggy (17) [0.66]

```





```

elknSavlllnatAiaavaegtDrvievvgracrailhihprirrgleFral1 Env Con-B
-----G-----R-----I-- Env HXB2
--G--I--IL--G--N--T----- Env DH123
--i--in--DtI--gW--ig--ig--n--f----- Env Con-A
--k--i--dtI--i--li--i--rn--f-a-q----- Env Con-C
--i--dt-----i--v--n----- Env Con-D
--I--t--v-----aL--g--v--N----- Env Con-F1
--I--Dri-----I--iL--G--V-----a----- Env Con-F2
--In--dtI--nw-----a-----n----- Env Con-G
.....gl.....L--vw--g-----I--T----- Env Con-J
--I--T-----I--ia--f--l----- Env Con-K
--I--T-----I--ia--f----- Env Con-J
--i--dt-----ia-----n----- Env Con-Con
elknSavlllnatAiaava (18) [0.92]
llnatAiaavaegtDrvi (17) [0.91]
avaegtDrvievvgragr (18) [0.03]
vievvgracrailhihprir (18) [0.34]
craihprirrgleFera (18) [-0.42]
rrirglerall (12) [-0.64]

```

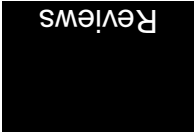
# Nef

[Number of peptides generated: 27]

```

mggkSkrsrvvgtvrrmrrraepaadgvgAvsrdlekhhgaitsntaamadcawleageeeeevgfpvrpgvplrPmttykaavDlshFlkekgglegliysgk Nef Con-B
-----S-I-----R--A-----T--A-----T--A-----T--A-----T--A-----H--R Nef HXB2
-----L-CGG--S-----R--A--G-----T--A-----Q-----L-----L-----L-----Nef DH123
-----s-i---q---i-qtp---e---Q-d---v---.i---v---.i---v---.i---v---.i---v---.i---v---.i---v---.i---v---d---k Nef Con-A
-----c-i---ai---t---e---a-q-d-y-l-----dtt-----r-----f-g-f--f-----k Nef Con-C
-----S-i---Ai---i--Td-----sT--a-----s-----w-k Nef Con-D
-----S-I---AI---pP---E---Q---RR-----r-t-p-L-----dd-----xx-----x-Kr Nef Con-F
-----s-i---a---i-qtP---E---Q---ar-----t-p-----dS-----f--F-----D---K Nef Con-G
-----sris-a-i-----e-----DRR--V-IN--st-p-a-----a-----g-f-----d---K Nef Con-H
--n---S...-q-----p--PaA---Q-A-----t-----t-----x-----g-f--F-----d---k Nef Con-J
--na---skfa-e-r-----rsppCA-q--e-aar-g-p-h-Pq--aL-f--shgd-----F--f-----h Nef Con-O
-----s-i---a---p-----q-----t-----t-----t-----g-f-----k Nef Con-Con

mggkSkrsrvvgtvrrWptvrr (17) [-0.54]
rsvvgtvrrermrra (16) [-0.88]
    ptvrrermrraepaadgv (17) [-1.01]
        rraepaadgvgAvsrdl (17) [-0.46]
            dgvgAvsrdlekhhgai (16) [-0.24]
                srdlekhhgaitsntaa (17) [-0.74]
                    gaitsntaamadcawl (18) [0.08]
                        aamadcawleageeeeev (18) [-0.72]
                            wleageeeeevgfpvrpgv (18) [-0.69]
                                evgfpvrpgvplrPmttyk (18) [-0.42]
                                    pgvplrPmttykaavDlsh (18) [-0.26]
                                        tykaavDlshFlkekggl (18) [-0.19]
                                            shFlkekgglegliysgk (18) [-0.60]
```



```

glegliysqkrqgILdlwvynhTQgyfPdwanyTppgiryPItfgWcfkLVPvepekeveeanegeemusllhPmslHgmdDperlevIewkfdPsr.lafhmarelhPeYyykdc Nef Con-B
-----H-R-----I-----N-----N-----D-I-----K-----T-----V-----C-----I-----E-----K-----L-----R-----V-----F-N-----Nef HXB2
--d--k--e--n--f--c--d--re--d--k--c--d--re--d--k--c--Q--ie--e--m--a--rk--i--f--Nef DH123
--w--k--e--i--i--e--d--ge--T--d--c--icQ--E--e--q--v--R--N--E--K--k--f--Nef Con-A
--x--Kr--E--l--n--f--m--d--aE--k--k--C--Q--E--ED--K--Lr--I--R--q--Nef Con-D
--D--K--e--n--f--t--f--E--E--Q--e--q--d--Qe--q--d--se--a--q--ttcdaa--pcq--ae--ahk--i--k--q--rs--gl--t--i--iq--ss--s--Nef Con-F
--d--k--e--hn--N-----E-----d--se-----l-----t--F-----l-----d-----e-----c-----cg-----e-----m-----rr--i-----Nef Con-G
--d--k--e--hn-----F-----t--F-----l-----d-----e-----c-----cg-----e-----m-----rr--i-----Nef Con-H
--h--aE-----F-----t--F-----l-----d-----e-----c-----cg-----e-----m-----rr--i-----Nef Con-J
--k--e-----F-----t--F-----l-----d-----e-----c-----cg-----e-----m-----rr--i-----Nef Con-O
glegliysqkrqgILdlw (18) [-0.31]
sqkrqgILdlwvynhTQgy (18) [-1.01]
dlwvynhTQgyfPdwany (17) [-1.08]
TQgyfPdwanyTppgiry (18) [-1.23]
wqnyTppgiryPIIfgW (18) [-0.68]
giryPIItfgWcfkLVPv (17) [0.78]
fgWcfkLVPvepekevee (18) [0.00]
Pvepekeveeanegeemusl (18) [-1.43]
eanegeemusllhPmslH (17) [-0.94]
nslhPmslHgmdDper (17) [-0.88]
slHgmdDper (17) [-0.96]
perevIewkfdsrIafhh (18) [-0.97]
kfdsrIafhmarelh (16) [-0.72]
afhmarelhPeYyykdc (17) [-1.06]

```