

## Bericht für den 5. Übungszettel

### Verwendung der Sequenz des Human Lymphotropic Virus 1

1) sh. GitHub

2)

Angabe der ersten 100 bp:

Non-codogen:

```
5`GGCTCGCATCTCTCCTTCACGCGCCCGCGCCTTACCTGAGGCCGCCATCCACGCCGGTTGAGTCGCGTT
3`CCGAGCGTAGAGAGGAAGTGCGCGGGCGGCGGAATGGACTCCGGCGGTAGGTGCGGCCAACTCAGCGCAA
```

```
5`CTGCCGCCTCCCGCCTGTGGTGCCTCCTGA
3`GACGGCGGAGGGCGGACACCACGGAGGACT
```

codogen:

```
5`ATGGGCCAAATCTTTTCCCGTAGCGCTAGCCCTATTCCGCGGCCGCCCGGGGGCTGGCCGCTCATCACTG
3`TACCCGTTTTAGAAAAGGGCATCGCGATCGGGATAAGGCGCCGGCGGGGCCCCCGACCGGCGAGTAGTGAC
```

```
5`GCTTAACTTCCTCCAGGCGGCATATCGCC
3`CGAATTGAAGGAGGTCCGCCGTATAGCGG
```

3)

Verwendung der ersten 1000 bp der codogenen Sequenz für die Übersetzung in eine Proteinsequenz:

```
atgggccaaat cttttcccgtagcgtagcc
ctattccgcg gccgccccgg gggctggcgc ctcatcactg gcttaacttc ctccaggcgg
catatgcct agaaccgggt ccctccagtt acgatttcca ccagttaaaa aaatttctta
aaatagcttt agaaacaccg gtctggatct gccccattaa ctactccctc ctagccagcc
tactccaaa aggatacccc ggccgggtga atgaaatttt acacatactc atccaaaccc
aagcccagat cccgtcccgc cccgcgcgc gcgcgcgcgc atcctccacc caccagcccc
cggattctga cccacaaatc cccctccct atgttgagcc tacagcccc caagtccttc
cagtcatgca cccacatggt gcccctccca accaccgccc atggcaaatg aaagacctac
aggccattaa gcaagaagtc tcccaagcgg cccctggaag cccccagttt atgcagacca
tccggcttgc ggtgcagcag tttgaccca ctgccaaaga cctccaagac ctctgcagt
acctttgctc ctccctcgtg gcttccctcc atcaccagca gctagatagc cttatatcag
aggccgaaac tcgaggtatt acaggttata accccttagc cggccccctc cgtgtccaag
ccaacaatcc acaacaacaa ggattaaggc gagaatacca gcaactctgg ctgcgcgcct
tcgccgcct gccagggagt gccaaagacc ctctctgggc ctctatcctc caaggcctgg
aggagcctta ccacgccttc gtagaacgcc tcaacatagc tcttgacaat gggctgccag
aaggcacgcc caaagacccc attttacgtt ccttagccta ctctaagca aacaaagaat
gccaaaaatt actacaggcc cgagggcaca ctaatagccc tctaggagat atgttgcggg
cttgtcaggc ctggaccccc aaagacaaaa
```

## 1.5'-3' frame

MGQIFSRASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFHQKKFLKIALETPVWICPINYSLLASLLPKG  
YPGRVNEILHILIQTQAQIPSRPAPPPSSSTHDPDSDPQIPPPYVEPTAPQVLPVMHPHGAPPNHRPWQMKDL  
QAIKQEVSAAPGSPQFMQTIRLAVQQFDPTAKDLQDLLQYLCSSLVASLHHQQLDLSLISEAETRGITGYNPLAG  
PLRVQANNPQQQGLRREYQQLWLAFAALPGSAKDPASWASILQGLEEYPYHAFVERLNIALDNGLPEGTPKDPILR  
SLAYSNANKECQKLLQARGHTNSPLGDMRLACQAWTPKDK

Die ersten 30 Aminosäuren sind daher:

MGQIFSRASPIPRPPRGLAAHHWLNFLQA

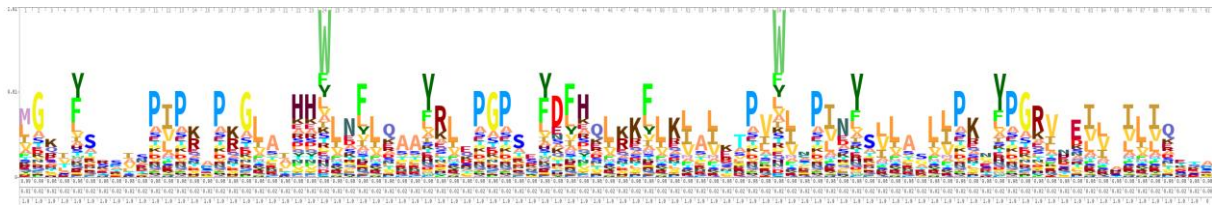
a) Aufgrund der Wobble-Hypothese ergibt sich, dass die dritte Position im Codon variabel ist. Hierdurch codieren verschiedene Codons für eine Aminosäure.

Durch die Verwendung von einer Aminosäuresequenz führen wir eine höhere Stufe der Variabilität ein: Da die dritte Position veränderlich ist, wäre die Verwendung von der DNA-Sequenz ungenauer. Mit der Proteinsequenz wird dies abgedeckt.

b) Es ist sinnvoll die anderen Frames zu durchsuchen, da unsere gesuchte Sequenz ebenso in den anderen Frames vorhanden sein kann.

4)

Verwendung der ersten 30 Aminosäuren für „protein sequence vs profile-HMM database“



Gag\_p19 (PF02228)

Aufgrund der Wahrscheinlichkeit der unterschiedlichen Aminosäuren im HMM-Logo ist von einer moderaten Ähnlichkeit auszugehen. Einige Bereiche sind sehr variabel. Viele Positionen werden durch eine höhere Wahrscheinlichkeit einer Aminosäure abgedeckt.

Höchstwahrscheinlich sind bessere Modelle möglich.

## Bericht Teil 2

### Verwendung der Sequenz des Human Lymphotropic Virus 2

#### 2) Angabe der ersten 100bp der codogenen Sequenz

```
5`ATGGGACAAATCCACGGGCTTTCCCCAACTCCAATACCCAAAGCCCCCAGGGGGCTATCAACCCACCACTGGC
3`TACCCTGTTTAGGTGCCCGAAAGGGGTTGAGGTTATGGGTTTCGGGGGTCCCCCGATAGTTGGGTGGTGACCG
```

```
5`TTAACTTTCTCCAGGCTGCTTACCGCT
3`AATTGAAAGAGGTCCGACGAATGGCGA
```

#### 3) Verwendung der ersten 1000 bp der codogenen Sequenz für die Übersetzung in eine Proteinsequenz:

```
atgg gacaaatcca cgggctttcc ccaactccaa
tacccaaagc cccagggggg ctatcaaccc accactgggt taactttctc caggetgctt
accgcttgca gcctaggccc tccgatttcg acttcagca gctacgacgc tttctaaaac
tagcccttaa aacgccatt tggctaaatc ctattgacta ctgcctttta gctagcctta
tccccaaagg atatccagga aggggtggtag agattataaa tatccttgtc aaaaatcaag
tctccccctag cgccccgcgc gccccagttc cgacacctat ctgccctact actactcctc
cgccacctcc ccccccttcc cgggaggccc atgttcccc cccttacgtg gaaccaccca
ccacgcaatg cttccctatc ttacatcccc caggagcccc ctcagctcat aggccctggc
agatgaaaga cttacaggcc atcaagcagg aggtcagctc ctctgctctt ggcagcccc
agttcatgca gaccctccgg ctggcgggtac aacagtttga cccaccgcc aaggacttac
aagatctcct ccagtaccta tgctcctccc tcgtagtctt cttacaccat cagcagctta
acacactaat taccgaggct gagaccgcgc ggatgacagg ctacaacccc atggcagggc
ccctaagaat gcaggcta atccccgcgc agcaaggctc tagacgggag taccagaatc
tttggtctgc tgctttctcc accctgccag gcaataaccg tgaccctctc tgggcagcta
tcctacaggg gctggaggaa cctattgctg cgttcgtaga gcgccttaac gtggcccttg
acaacggcct ccccgagggt acccccaaag agccatctt acgttcccta gcgtactcaa
acgccaacaa agaatgccaa aaaatcttac aagccgcgcg acacactaac agcccccttg
gggagatgct ccggacatgt caggcgtgga caccacagga caaaaccaag gtccttgagg
tccaaccacg gaggcccccc ccacacagc cctgctttcg ttgtggcaag gtaggacact
ggagtcggga ctgtaccag ccacgcccc ctctggccc ctgcccccta tgccaagatc
cttctcactg gaaaagggac tgcccacaac tcaaaccccc tcaggaggaa ggggaacccc
tcctgttgga tctcccttcc acctcaggca ctactgagga aaaaaactcc ttaagggggg
agatctaa
```

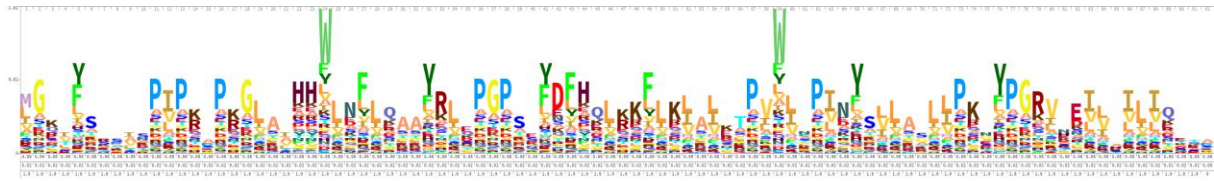
#### 1. 5'-3'Frame

```
MGQIHGLSPTPIPKAPRGLSTHHWLNFLQAAYRLQPRPSDFDFQQLRRFLKLALKTPIWLNPIDYSLLASLIPKG
YPGRVVEIINILVKNQVSPSAPAPVPTPICPTTTTPPPPPPSPEAHVPPPYVEPTTTQCFPI LHPPGAPSAHRP
WQMKDLQAIKQEVSSSALGSPQFMQTLRLAVQQFDPTAKDLQDLLQYLCSSLVVSLHHQQNLNLTAEATRGMTG
YNPMAGPLRMQANNPAQQGLRREYQNLWLA AFSTLPGNTRDPSWAAILQGLEEPYCAFVERL NVALDNGLPEGTP
KEPILRSLAYSNANKECQKILQARGHTNSPLGEMLR TCQAWTPKDKTKVLVVQPRRPPPTQPCFRCGKVGHWSRD
CTQPRPPPGPCPLCQDPSHWKRDCPQLKPPQEEGEPLLLDLPSTSGTTEEKNSLRGEI
```

#### Die ersten 30 Aminosäuren des Frames:

```
MGQIHGLSPTPIPKAPRGLSTHHWLNFLQA
```

#### 4) Verwendung der ersten 30 Aminosäuren für „protein sequence vs profile-HMM database“



Aufgrund der Ähnlichkeit des Genoms mit dem des Human Lymphotropic virus 1 ist das HMM-Logo das gleiche wie im ersten Teil des Berichts, wodurch die gleiche Aussage getätigt wird, dass das Logo eine moderate Analogie bzw. Wahrscheinlichkeit aufweist.

Im Vergleich zu einem Logo der **vollen** Aminosäuresequenz des BMPRIb und dem resultierenden Protein TGF- $\beta$  fällt auf, dass hier bspw. eine größere Wahrscheinlichkeit der Aminosäuren vorhanden ist.

