

Bericht 1 zum 4. Übungsblatt

Aufgabe 2:

Die ersten 100 Basenpaare der Genom-Sequenz von „Human T-cell leukemia virus type I“:

```
ggctcgcacatctctccttcacgcgccccgccgccttacctgaggccgccatccacgccggt  
tgagtcgcggttctgcgcctcccgcctgtggtgcctcctga
```

Die ersten 100 Basenpaare der CDS (coding sequence) von „Human T-cell leukemia virus type I“:

```
ATGGGCCAAATCTTTTCCCGTAGCGCTAGCCCTATTCCGCGGCCGCCCGGGGGCTGGC  
CGCTCATCACTGGCTTAACTTCCTCCAGGCGGCATATCGCC
```

Aufgabe 3:

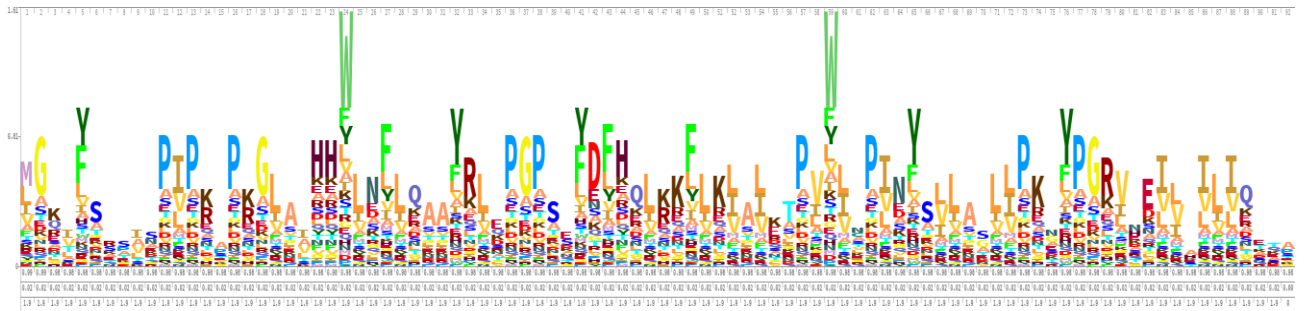
Die ersten 30 Aminosäuren des 1. 5'3' Frames (aus der CDS generiert):

```
MGQIFSRASPIPRPPRGLAAHHWLNFLQA
```

- Da der genetische Code redundant ist und damit eine Aminosäure durch mehrere Codons codiert wird, können verschiedene Organismen über die Aminosäuresequenz der codierenden DNA-Sequenzen verglichen werden. Auf diese Weise können kleinere Abweichungen in der DNA, die dennoch zur gleichen Aminosäure resultieren, nicht berücksichtigt und somit größere Fehler verhindert werden.
- Dies dient zur Überprüfung des Anfangs der Transkription für die mRNA des späteren Proteins. Es gibt innerhalb eines Genabschnitts mehrere Initiationsstellen für die Transkription, weshalb es sinnvoll ist, alle möglichen Übersetzungsframes zu durchsuchen.

Aufgabe 4:

HMM-Logo des Major core protein p19 (EMBL-EBI: Gag_p19 (PF02228),
<http://pfam.xfam.org/family/PF02228#tabview=tab4>)



Wie dem HMM-Logo entnommen werden kann, stellen die Aminosäuren der Suchsequenz zum größten Teil die am häufigsten vorkommenden Aminosäuren im Sequenz-Logo dar. Dies zeigt, dass diese an fast jeder Position im Logo konserviert vorliegen.

Bericht 2 zum 4. Übungsblatt

Aufgabe 5: für das Human papillomavirus type 85 (NCBI: NC_034616.1)

5.2)

Die ersten 100 Basenpaare der Genom-Sequenz von „Human papillomavirus type 85“:

```
cttataactttttttataattttattactgaaaaggggtgtaaccgaaaacgggtcatgaccg
aaatcgggtgca tatatataaccagcttacagtacatgtaa
```

Die ersten 100 Basenpaare der CDS (coding sequence) von „Human papillomavirus type 85“:

```
ATGATGGATGCCCTATCCCAACGTTTAAATGTGTTACAGGAAAAGATTCTAGAACATTA
CGAGACTGATAGTACAGACATATGTGATCAAATAGACTATT
```

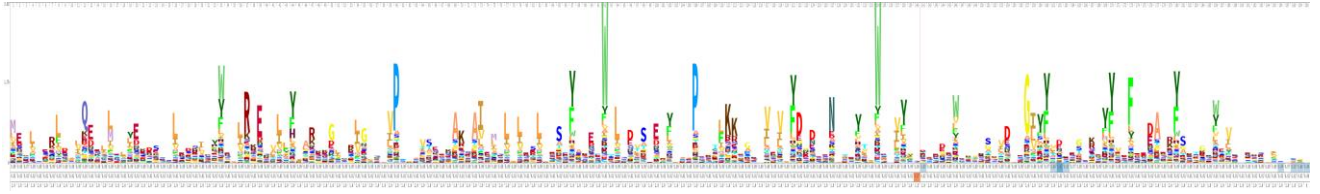
5.3)

Die ersten 30 Aminosäuren des 1. 5'3' Frames (aus der CDS generiert):

```
MMDALSQRLNVLQEKILEHYETDSTDICDQ
```

5.4)

HMM-Logo des E2 (early) protein (EMBL-EBI: PPV_E2_N (PF00508),
<http://pfam.xfam.org/family/PF00508.16#tabview=tab4>):



Hier stimmen die Aminosäuren der Suchsequenz nur gering mit der Logo-Sequenz überein. Generell gibt es hier keine eindeutige, am häufigsten vorkommende Aminosäure für mehrere Positionen. Dies zeigt, dass an vielen Positionen die Entropie höher ist (da die relative Höhe der einzelnen Aminosäuren kleiner ist) und dass es keine eindeutige, konservierte Aminosäuresequenz gibt (starke Abweichungen).