

Übung 5

1.

Human T-cell leukemia virus type I NC_001436

Erste 100 bp:

```
TGACAAATGACCATGAGCCCCAAATATCCCCCGGGGGCTTAGAGCCTCCCAGTGAAAAACATTTCGCG  
AAACAGAAGTCTGAAAAGGTCAGGGCCCCAGAC [...]
```

FASTA: <https://www.ncbi.nlm.nih.gov/nuccore/LC378575.1?report=fasta>

2.

Die ersten 30 Aminosäuren des (+1) 5'3' Reading Frames (kein *Open Reading Frame*)

-Q-P-APNIPRGLRASQ-KTFPRNRSLKRS

a) Die Suche mit der Aminosäuresequenz ist der Suche mit DNA-Sequenz (z.B. bei der Analyse von Homologen Proteinen) vorzuziehen, da für die meisten Aminosäuren verschiedene Codons existieren (s.a. Wobble Base). Somit können theoretisch zwei vollkommen verschiedene (abgesehen vom Start-Codon AUG und dem Tryptophan-Codon UGG) DNA-Sequenzen existieren, die dennoch für die gleiche Aminosäuresequenz codieren.

b) Bei der Berücksichtigung aller 6 Reading Frames (0, +1, +2 für je 3'5' und 5'3') werden alle möglichen AA-Sequenzen und somit auch alle potentiellen Open Reading Frames berechnet.

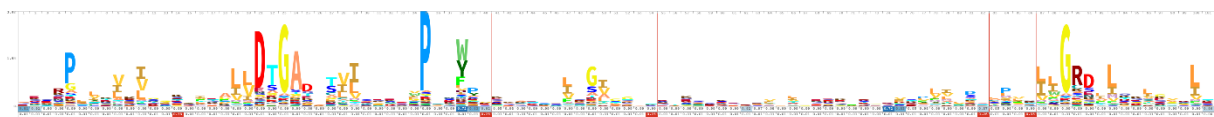
3.

Der 5. ORF (start_pos=752) erzielt einen Hit auf die **Retrovirale Aspartyl protease**

Sequenz:

```
MTVLPIALFSSNTPLKNTSVLGAGGQTQDHFKLTSPLVLIRLPFRTPPIVLTSCLVDIKNNWAIIGRD  
ALQQCQGVLYLPEAKRPPVILPIQAPAVLGLEHLPRPPEISQFPLNQNASRPCNTWSGRPWRQAISNP  
TPDQEITQYSQLKRPMEPGDSSTTCGPLTL
```

HMM-Logo:



Einschätzung:

Ein retrovirales Protein im Genom eines humanen Retrovirus' zu finden ist soweit plausibel.

Aus der Consensus Sequenz des HMM Logos gehen (z.B.) folgende hochkonservierte Reste hervor:

R5, D21, G23, A24, P35, aromatische AS bei 38, G89, Leucin/Isoleucin-reiche C-Term-Region

Aufgeführte Charakteristika konnten in der Sequenz identifiziert werden:

```
MTVLPIALFSSNTPLKNTSVLGAGGQTQDHFKLTSPLVLIRLPFRTPPIVLTSCLVDIKNNWAIIGRD  
ALQQCQGVLYLPEAKRPPVILPIQAPAVLGLEHLPRPPEISQFPLNQNASRPCNTWSGRPWRQAISNP  
TPDQEITQYSQLKRPMEPGDSSTTCGPLTL
```

Der für den Namen und die kat. Aktivität der **Aspartat**protease Rest (D23??) konnte nicht eindeutig auffindig gemacht werden.

Die Proteinsequenz zeigt im Vergleich zur HMM-Sequenz einen deutlich verlängerten C-Terminus von etwa 60 Aminosäuren, dieser könnte beispielsweise in dem Propeptid eine zymogenartige Funktion haben und zur Aktivierung des Enzyms abgespaltet werden (Spekulation!). Der ExPASy PeptideCutter zeigt hierfür einige endoproteolytische Schnittstellen im Bereich der AS Positionen 100-110 für Proteinase K oder Pepsin.

4.

Weiteres Beispiel: Abschnitt aus bakteriellem Genom: *bacillus subtilis* NC_000964.3:

Basen 2336830-2334581 (ATP-dependent Helicase)

ATGAAAAAGAAATCACTGACTGAACTCATTTCTGATTATAAGGAAATGAAAACGTTGTGAATTGGC [...]

(Sequenz beginnt gleich im ORF)

AS:

MKKKSLTELISDLKGNENVVNW [...]

Der HMMER Blast ergibt folgende Hits: **DEAD, DUF1998, Helicase_C**

DEAD:



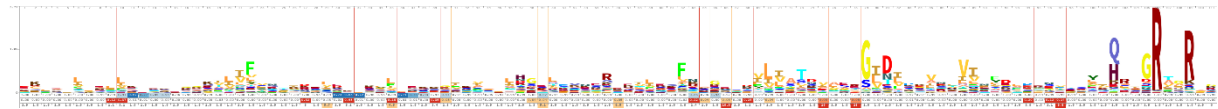
Das HMM-Logo zeigt einige (wenige) der typischen Helikase-Motive:

Q (Adenin-spezifisches NTP-Binding)

DEAD/H in diesem Fall DEAH, was die Helikase als translozierendes Enzym in die SF2 der Helikasen einordnet => RNA oder DNA bindend (noch nicht bekannt) *Motiv II*

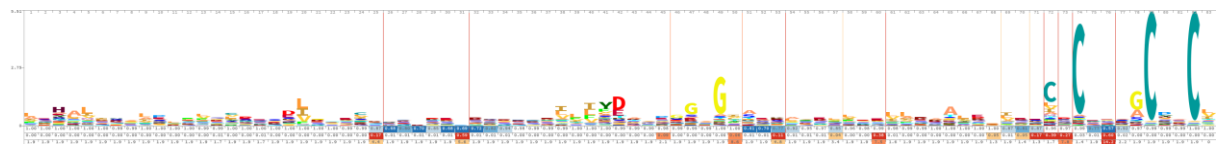
SAT ist ebenfalls ein organismenübergreifend hochkonserviertes Motiv in den Helikasen (*Motiv III*)

Helicase_C:



In dieser Sequenz sticht lediglich das Motiv VI (mit Q und R) hervor, möglicherweise auch Motiv Va mit G-I-D

DUF1998:



Die DUF (*Domain of unknown function*) 1998 zeigt im c-terminalen Bereich 4 ausgeprägte Cystein-Reste. Dieses Motiv lässt auf eine Ion-bindende Funktion (z.B. in Iron-Sulfur- oder Zinc-Sulfur-Clustern) schließen.