

Assignment 5

Aufgabe 2:

Genom-Sequenz von „Human T-cell leukemia virus type I“ (Accession: NC_001436):

Basenpaare 1-100:

```
1   ggctcgcatc tctccttcac gcgcccgcgc ccttacctga ggccgccatc cacgccggtt
61  gagtcgcgtt ctgccgcctc ccgcctgtgg tgcctcctga
```

Aufgabe 3:

Übersetzung der oben gegebenen DNA-Sequenz in die Aminosäuresequenz:

AS 1-30:

5'3' Frame 1

G S H L S F T R P P P Y L R P P S T P V E S R S A A S R L W

- (a) die Suche nach HMM Logos in Aminosäuresequenzen ist der Suche in Genomsequenzen vorzuziehen, da die selbe Aminosäure grundsätzlich durch verschiedene Basentriplets codiert werden kann. Demnach kann die selbe AS-Sequenz durch verschiedene DNA-Sequenzen dargestellt werden. Die Funktionalität bestimmter Proteindomänen ist durch ihre 3D-Struktur gegeben, welche durch die spezifische AS-Abfolge bestimmt wird. Demnach entscheidet nicht die Genomsequenz, sondern die daraus resultierende AS-Sequenz über das auftretende Motiv, welches mit den gegebenen HMM Logos verglichen wird.
- (b) Es ist sinnvoll, alle 6 möglichen Übersetzungs-Frames zu durchsuchen, weil nicht klar ist, welcher der drei reading frames pro Einzelstrang bei der Translation verwendet wird. Die DNA bzw. RNA wird dabei in Basentriplets ausgelesen und in die AS übersetzt. Daher kann das Leseraster sowohl bei Base 1, 2 oder 3 Beginnen, was zu komplett verschiedenen AS-Sequenzen führt. Dies gilt sowohl für den forward, als auch für den reverse strand, sodass 6 verschiedene reading frames entstehen. Alle 6 stellen mögliche Grundlagen für die Translation dar.

Aufgabe 4:

Innerhalb der ersten rund 2000 Basenpaare/ 670 AS wurde kein HMM Logo gefunden, daher wurde die Suche mit den Basenpaaren 2041-3000 und der entsprechenden AS-Sequenz wiederholt. Dabei wurde das folgende HMM Logo gefunden:

Reverse transcriptase (RNA-dependent DNA polymerase)

AS-Sequenz aus gegebenem Genom (Suchsequenz), welche Analogien zum HMM Logo der RVT aufweist:

```

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

```

In rosa sind open reading frames hervorgehoben, wo eine Translation von Start- bis Stop-Codon möglich ist.

In gelb sind Positionen markiert, wo die Suchsequenz mit Basen hoher Wahrscheinlichkeit des HMM Logos übereinstimmt. Demnach weist die Suchsequenz punktuell größere Ähnlichkeit mit der RVT-Domäne auf. Grundsätzlich könnte die 3D-Struktur dieser Sequenz also eine RVT-ähnliche Domäne bilden.

Aufgabe 5:

Genom-Sequenz von „Avian leukemia virus isolate SCDY1, complete genome“ (Accession: NC_015116):

Basenpaare 1-100:

```

1  tgtagtcata  tgcaatactc  ttgtgtaacg  ataaaacagc  aatatgcctt  ataaggagtg
61  aaaaggcacc  gtacacgtcg  attggtggaa  ctaggatggt

```

Übersetzung der oben gegebenen DNA-Sequenz in die Aminosäuresequenz:

AS 1-30:

5'3' Frame 1

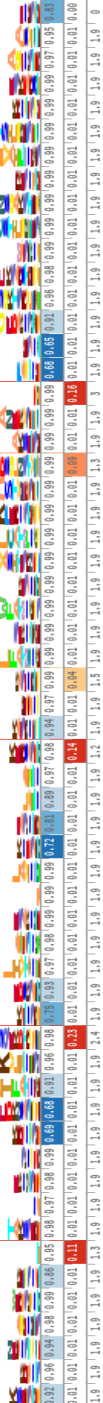
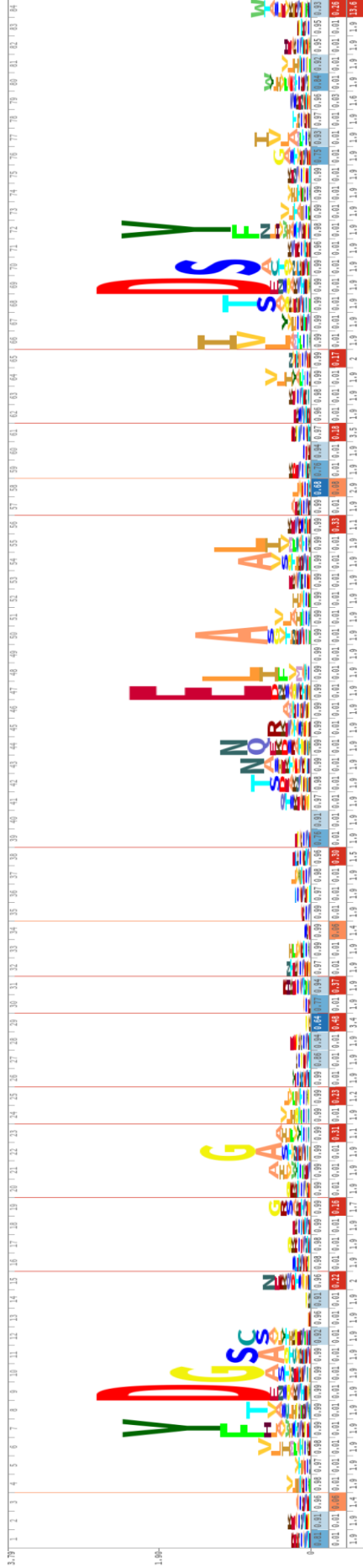
```

C S H Met Q Y S C V T I K Q Q Y A L Stop G V K R H R T R R L V E

```

Die Untersuchung der Basenpaaren 4021-5040 und der entsprechenden AS-Sequenz ergab Übereinstimmungen mit dem folgenden HMM Logo:

RNase_H



AS-Sequenz aus gegebenem Genom (Suchsequenz), welche Analogien zum HMM Logo der RVT aufweist:

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

In rosa sind open reading frames hervorgehoben, wo eine Translation von Start- bis Stop-Codon möglich ist.

In gelb sind Positionen markiert, wo die Suchsequenz mit Basen hoher Wahrscheinlichkeit des HMM Logos übereinstimmt. Demnach weist die Suchsequenz eher zu Beginn größere Ähnlichkeit mit der RNaseH-Domäne auf. Grundsätzlich könnte die 3D-Struktur dieser Sequenz also eine RNaseH-ähnliche Domäne bilden.

Außerdem könnten verschiedene Spleißisoformen des Proteins auftreten, sodass die Domäne anhand der Suchsequenz eventuell nicht als zusammenhängende AS-Abfolge gefunden werden kann. Evtl. entsteht die charakteristische 3D-Struktur erst durch die Kombination verschiedener Exons.