

Assignment 5

Aufgabe 2: Genom-Sequenz von „Human T-cell leukemia virus type I“

<https://www.ncbi.nlm.nih.gov/nuccore/LC183873.1>

Die ersten 100 Basenpaare:

```
tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctccca  
gtgaaaaaca tttccgcgaa acagaagtct gaaaagggtca gggcccagac
```

Aufgabe 3:

5'3' Frame 1																			
tga	caa	tga	cca	tga	gcc	cca	aat	atc	ccc	ggg	ggg	ctt	aga	gcc	tcg	cag	tga	aaa	aca
-	Q	-	P	-	A	P	N	I	P	R	G	L	R	A	S	Q	-	K	T
ttt	ccg	cga	aac	aga	agt	ctg	aaa	agg	toa	ggg	ccc	aga	ota	agg	ctc	tga	cgt	ctc	ccc
F	P	R	N	R	S	L	K	R	S	G	P	R	L	R	L	-	R	L	P

5'-Q-P-APNIPRGLRASQ-KTFPRNRSLKRS3'

bzw: 5'QPAPNIPRGLRASQKTFPRNRSLKRSGPRL3'

a) AS-Sequenz vs Genomsequenz

- benutzt man den Einbuchstabencode für die AS-Sequenz verkürzt sich der string auf 33%, da nun nicht mehr mit Tripletts (Codons) gearbeitet wird
- die Betrachtung des genetisches Codes entfällt bei der AS-Sequenz
- Bei dem Virus hier egal, bei Eukaryoten aber muss Segmentierung der Gene in Exons und Introns beachtet werden (alternatives Splicen)

b) 6 mögliche Übersetzungsframes

- Überprüfung ob die Zuordnung des codogenen Stranges korrekt ist
- Identifikation verschiedener Leseraster, um den ORF zu identifizieren

Aufgabe 4:

Tools:

NCBI Genom-Datenbank:

<https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239>

Übersetzungs-Tools „DNA -> Proteinsequenz“:

<https://web.expasy.org/cgi-bin/translate/dna2aa.cgi>

Profil-HMM Such-Tool:

<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>

Genomsequenz (1000):

```
tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctccca
gtgaaaaaca tttccgcgaa acagaagtct gaaaagggtca gggcccagac
taaggctctg acgtctcccc ccggagggac agctcagcac cggctcaggc
taggccctga cgtgtccccc tgaagacaaa tcataagctc agacctccgg
gaagccaccg gaaccaccca tttcctcccc atgtttgtca agccgccctc
aggcgttgac gacaaccctt cacctcaaaa aacttttcat ggcacgcata
tggctgaata aactaacagg agtctataaa agcgtggaga cagttcagga
gggggctcgc atctctcctt cacgcgcccg ccgccctacc tgaggccgcc
atccacgccg gttgagtcgc gttctgccgc ctcccgcctg tgggtgcctcc
tgaactgcgt ccgccgtcta ggtaagttta gagctcaggt cgagaccggg
cctttgtccg gcgctccctt ggagcctacc tagactcagc cggttctcca
cgctttgcct gaccctgctt gctcaactct gcgtctttgt ttcgttttct
gttctgcgcc gctacagatc gaaagttccg cccctttccc tttcattcac
gactgactgc cggcttggcc cacggccaag taccggcgac tccgttggct
cggagccagc gacagcccat tctatagcac tctccaggag agaaacttag
tacacagttg ggggctcgtc cgggatacga gcgccccttt attccctagg
caatgggcca aatcttttcc cgtagcgcta gccctattcc gcggccgccc
cgggggctgg ccgctcatca ctggcttaac ttcctccaag cggcatatcg
cctagaacct ggtccctcca gttacgattt ccaccagttg aaaaaatttc
ttaaaatagc tttagaaaca ccggtctgga tctgtcccat taactactcc
```

AS-Sequenz:

```
QPAPNIPRGLRASQKTFPRNRSLKRSGPRLRLRLPPEGQLSTGSGALTCP
PEDKSAQTSGKPPEPPISSPCLSSRPQALTTTPHLKKLFMARIWLNKLTG
VYKSVETVQEGARISPSRARRPTGRHPRRLSRVLPPPACGASTASAVVSL
ELRSRPGLCPALPWSLPRLSRFSTLCLTLLAQLCVFVSFSVLRRYRSKVP
PLSLSFTTDCRLGPRPSTGDSVGSEPATAHSIALSRRETYTVGGGSSGIRA
PLYSLGNGPNLFPRPYSAAPGAGRSSLALPPSGISPRTSLQLRFPPVE
KISNSFRNTGLDLSHLL
```