

Bericht zur Analyse des Virus-Genoms des „Human T-cell leukemia virus type I“ mittels HMM-Profilen aus der Pfam Datenbank

Das humane T-lymphotrope Virus 1 gehört zur Gruppe der Retroviren und löst beim Menschen Leukämie aus. Um Domänen innerhalb eines Genoms oder einer Proteinsequenz zu finden, wird das Hidden Markov Modell verwendet.

2) Aus der NCBI-Datenbank (Accession: NC_001436) wird die Basensequenz ermittelt und exemplarisch die ersten 100 Basen hier dargestellt:

```
ggctcgcatc tctccttcac ggcggcgccg ccttacctga ggccgccatc  
cacgccggtt gagtcgcgtt ctgccgcctc ccgcctgttg tgccctctga
```

3) Mithilfe des ExPASy-Übersetzungstools werden die ersten 30 Aminosäuren ermittelt.

```
GSHLSTFRPPPYLPPSTPVESRSAASRLWC
```

- Da die selbe Aminosäure durch unterschiedliche Basentriplets codiert werden kann, ist es günstiger die Analyse von der Aminosäuresequenz durchzuführen. Stille Mutationen innerhalb Spezies und zwischen verschiedenen Spezies erleichtern somit den Vergleich der Aminosäuresequenz.
- Da bei einem unbekannten gegebenen Strang nicht klar ist, ob es sich um den codierenden oder nicht codierenden Strang handelt, müssen stets alle Frames durchgetestet werden. Desweiteren ist der OpenReadingFrame nicht bekannt, sodass auch dieser erst bestimmt werden muss.

4) Innerhalb des ersten Reading Frames wurde kein HMM-Logo gefunden, deshalb wurde die Suche bis zum 3. Frame erweitert

5'3' Frame 1
GSHLSFTRPPPYL-PPSTPVESRSAASRLWCLNLYVRLGKF-AQVETGPLSGAPLEPT-TQPALHALPDPACSTLRCLFVFCAPLQIESSTPFFFIHDWLPAPWPTAKYRQLCWLASDS
PFYSTLQE-NLVHSGWGLVRDSSAPLFP-QWAKSFPVALALFRGRPGGWPLITGLTSS-RHIA-NPVPVTISTS-KNFLK-L-KHRSGSAPLTPS-PAYSQKDTAGW-MKFYTYSSKPKP
-SRPAPRRRRRHPPPTPRILTHKSPLF-MLSLQPPKSFQSCTHMVPLETTAHGKWKTY-PLSKKSPKRPLEAPSLC-PSGLRCSSTPLPK

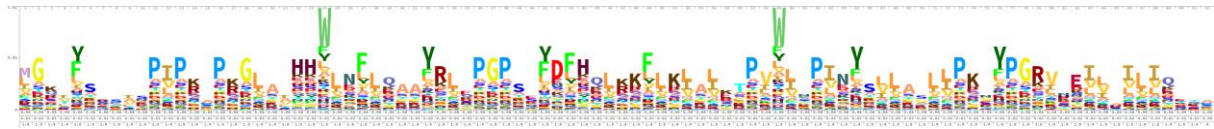
5'3' Frame 2
ARISPSRARRLTWGRHPRRLSRVLPACGASWTTSV-VSLEL-S-PGLCPALPWSLP-LSRLSTLCLTLAQLYVVFVSFVLRRY-SKVPPLSLSFTTDCRLGPRPSTGNSAGSEPATA
HSMALS--EI-YTVGGSSGIRAPLYSLGNGPNLFP-R-PYSAAAPGAGRSLA-LPPGGMSP-TRSLQLRFPVVKIS-NSF-NTGLDLPH-LLPSPQTPK-MPRPGEWNFTHTHPNPS
DVPFPRAAAVILHPRPPGFWPTNPPSLCWAYSPSPSSHAPTWCPSQPPPMANE-PTGH-A-SLPSGPKPPVYADHPACGAAVPHCQ-

5'3' Frame 3
LASLLHAPALPEAAIHAGWVAFCLRPVVPPELRPPS--V-SSGRDAFVRRSLGAYLDSAGSPRFAWPCLLNSTSLFRFLCAVTDKRFHFFPFSRLTAGLAHQVQVATLLARSQRP
IL-HSPGEKFSTQLGARPGFERPFIIP-AMGQIFSRASAPIPRPRGLAAHHWLNFLQAAAYRLEPGPSSYDFHQLKKFLKMALETVPVWICPINYSLASLLPKGYPRVNEILHMLIQTQAO
IPSPAPPPSSSTHDPDSDFQIPPPYVEPTAPQVLFVMPHGPAPPNHRPWQMKDLQAIKQVVSQAAPGSPQFMQTIRLAVQQFDPTAK

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

Die Suche ergab das HMM-Logo des Major core Proteins p19.

Id	Accession	Clan	Description	Cross-references	Start	End	Ind.	Cond.
> Gag_p19	PF02228.15	CL0074	Major core protein p19		1	30	5.2e-15	3.1e-19



http://pfam.xfam.org/family/Gag_p19#tabview=tab4

Die Größe des Buchstabens teilt die Höhe der Säule gemäß der Häufigkeit der Buchstaben. Diese Buchstaben stimmen mit dem HMM-Logo anderer Retroviren überein. Es handelt sich also um hochkonservierte Bereiche.

5) Aus der NCBI-Datenbank wird die Basensequenz Human immunodeficiency virus 1 (Accession: NC_001802) ermittelt und exemplarisch die ersten 100 Basen hier dargestellt:

```
gggtctctctg gttagaccag atctgagcct gggagctctc tggctaacta
gggaaccacac tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt
```

Mithilfe des ExPASy-Übersetzungstools werden die ersten 30 Aminosäuren ermittelt.

```
GLSGTSEPGSSLANGTHCLSLNKACLECFK
```

Innerhalb des ersten Reading Frames wurde kein HMM-Logo gefunden, deshalb wurde die Suche bis zum 3. Frame erweitert

5'3' Frame 1

GLSG-T-SEPGSSLAN-GTHCLSLNKACLECFK-CVPVCCVTILT-DPSDFSQCGKSLAFAPEQGPES-ET-GALSTQDSACWSAHGK-RGAATGEYAKNFD-R-LEGE-WVRERQY-A
GEN-IDGKFKF-GQGE-KNMN-NM-YGQAGS-NDSQLILAC-KHQKAVDKYWDSYNHPF-QDQNLDDHYMMQ-QPSIVCIKG---KTP-KL-T---KSKTKV-KKHSKQQLTQDTAI-SAK
ITL-C-TS-GKWIY-PYHLEL-MHG-K-K--LSAQKWYPCFQHYQKEPPHKI-TPC-TQWGDIKQPCCK-K-PSM-KLQNGMECTQCMQG

5'3' Frame 2

VSLV-PDLSLALWLT-EPTA-ASMKLALSASSVCPVSVVNLW-LEIPQTLVSVENL-QWRPN-DLKAKGKPELSRR-TRLAEARTA-GEGRRLVSTPKILTSGG-KE-DGCESVSIKR
G-I-SMGKNSVKA-GKEKM-IKTYSMGKQGA-TIRS-SWPV-NI--L-TNTGTATTIPSD-I--T-IIM-YSSNPLLCASKD-DK-HQGSF-QD-G-AKQK-EKSTASSSWH-TQQSGQPK
LPYSAEHFGANGTSGHIT-NFKCMGKSS--EGFQP-SDTHVFSII--SHPT-FKHAKHSGGTSSSHANVK-DHQWGC-MG-SASSAC-A

5'3' Frame 3

SLWLDQIWAWEISG-LGNPLLKPQ-SLPWVLQVVCARLLCDSGN--SL-PF-SVWKKISSSGARTGTWKRKGNQ-SSLDAGLGLLKRARQEA-GGDWWVRQKFWLAE--EMGA-ASVLSG
GELDRWEKIRL-PGGKKYKYLKHMVWAS-ELERFAVNPGLLETSEGC-QMLGQLQPSLQTGSEEL-SLYNTVATLYCVHQ-MEMKDTKEALDKMEEQNKSKKKAQQAADTGHNSQVSQN
YPMVQNIQGGQMVHQAMSB-TLNAWVKVVEEKAFSPV-MFMFSALSEGATPDQDNTMLNTVGGHQAAMQMLKETINEEAEDW-VHPVHAG

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

Id	Accession	Clan	Description	Cross-references	Start	End	Ind.	Cond.
Gag_p24	PF00607.19	CL0148	gag gene protein p24 (core nucleocapsid protein)		1	30	9.3e-06	5.5e-10

<https://www.ebi.ac.uk/Tools/hmmer/results/BBED452A-7F00-11E8-9314-60E0DBC3747A/score>

Die Suche ergab das HMM-Logo des gag gene protein p24.



http://pfam.xfam.org/family/Gag_p24#tabview=tab4

Die Größe des Buchstabens teilt die Höhe der Säule gemäß der Häufigkeit der Buchstaben. Diese Buchstaben stimmen mit dem HMM-Logo anderer Retroviren überein. Es handelt sich also um hochkonservierte Bereiche.