# Übung 5

# Human T-cell leukemia virus type I OATL9B proviral DNA

**Aufgabe 2:**

Komplettes Genom:

https://www.ncbi.nlm.nih.gov/nuccore/LC378575.1 (02.07.18, 20:39 Uhr)

ORIGIN

1 tgacaatgac catgagcccc aaatatcccc cgggggctta gagcctccca gtgaaaaaca

61 tttccgcgaa acagaagtct gaaaaggtca gggcccagac taaggctctg acgtctcccc

121 ccggagggac agctcagcac cggctcaggc taggccctga cgtgtccccc tgaagacaaa

181 tcataagctc agacctccgg gaagccaccg gaaccaccca tttcctcccc atgtttgtca

241 agccgccctc aggcgttgac gacaacccct cacctcaaaa aacttttcat ggcacgcata

301 tggctgaata aactaacagg agtctataaa agcgtggaga cagttcagga gggggctcgc

361 atctctcctt cacgcgcccg ccgccctacc tgaggccgcc atccacgccg gttgagtcgc

421 gttctgccgc ctcccgcctg tggtgcctcc tgaactgcgt ccgccgtcta ggtaagttta

481 gagctcaggt cgagaccggg cctttgtccg gcgctccctt ggagcctacc tagactcagc

541 cggttctcca cgctttgcct gaccctgctt gctcaactct gcgtctttgt ttcgttttct

601 gttctgcgcc gctacagatc gaaagttccg cccctttccc tttcattcac gactgactgc

661 cggcttggcc cacggccaag taccggcgac tccgttggct cggagccagc gacagcccat

721 tctatagcac tctccaggag agaaacttag tacacagttg ggggctcgtc cgggatacga

781 gcgccccttt attccctagg caatgggcca aatcttttcc cgtagcgcta gccctattcc

841 gcggccgccc cgggggctgg ccgctcatca ctggcttaac ttcctccaag cggcatatcg

901 cctagaaccc ggtccctcca gttacgattt ccaccagttg aaaaaatttc ttaaaatagc

961 tttagaaaca ccggtctgga tctgtcccat taactactcc ctcctagcca gcctactccc

1021 aaaaggatac cccggccggg tgaatgaaat tttacacata ctcatccaaa cccaagccca

1081 gatcccgtcc cgtcccgcgc caccgccgcc gtcatccccc acccacgacc ccccggattc

1141 tgatccacaa atcccccctc cctatgttga gcctacggcc ccccaagtcc ttccagtcat

1201 gcacccacat ggtgcccctc ccaaccatcg cccatggcaa atgaaggacc tacaggccat

1261 taagcaagaa gtctcccaag cagcccctgg gagcccccag tttatgcaga ccatccggct

1321 tgcggtgcag cagtttgacc ccactgccaa agacctccaa gacctcctgc agtacctttg

1381 ctcctccctc gtggcttccc tccatcacca gcagctagat agccttatat cagaggccga

1441 aacccgaggt attacaggtt ataacccctt agccggtccc ctccgtgtcc aagccaacaa

1501 tccacaacaa caaggattaa ggcgagaata ccagcaactc tggctcgccg ccttcgccgc

1561 cctgccaggg agtgccaaag acccttcctg ggcctctatc ctccaaggcc tggaggagcc

1621 ttaccacgcc ttcgtagaac gcctcaacat agctcttgac aatgggctgc cagaaggcac

1681 gcccaaagac cccatcttac gttccttagc ctactccaat gcaaacaaag aatgccaaaa

1741 attactacag gcccgaggac acactaatag ccctctagga gatatgttgc gggcttgtca

1801 gacctggacc cccaaagaca aaaccaaagt gttagttgtc cagcctaaaa aacccccccc

1861 aaatcagccg tgcttccggt gcgggaaagc aggccactgg agtcgggact gcactcagcc

1921 tcgtcccccc cccgggccat gccccctatg tcaagaccca actcactgga agcgagactg

1981 cccccgccta aagcccacta tcccagaacc agagccagag gaagatgccc tcctattaga

2041 cctccccgct gacatcccac acccaaaaaa ctccataggg ggggaggttt aacctccccc

2101 cccacattac agcaagtcct tcttaaccaa gacccagcat ctattctgcc agttataccg

2161 ttagatcccg cccgtcggcc cgtaattaaa gcccaggttg acacccagac cagccaccca

2221 aagactatcg aagctttact agatacagga gcagacatga cagtccttcc gatagccttg

2281 ttctcaagta atactcccct caaaaataca tccgtattag gggcaggggg ccaaacccaa

2341 gatcacttta agctcacctc ccttcctgtg ctaatacgcc tccctttccg gacaacgcct

2401 attgttttaa catcttgcct agttgatatc aaaaacaact gggccatcat aggtcgcgat

2461 gccttacaac aatgccaggg cgtcctgtac ctccctgagg caaaaaggcc gcctgtaatc

2521 ttgccaatac aggcgccagc cgtccttggg ctagaacacc tcccaaggcc ccccgaaatc

2581 agccagttcc ctttaaacca gaacgcctcc aggccttgca acacttggtc cggaaggccc

2641 tggaggcagg ccatatcgaa ccctacaccg gaccaggaaa taacccagta ttcccagtta

2701 aaaaggccaa tggaacctgg cgattcatcc acgacctgcg ggccactaac tctctaacca

2761 tagatctctc atcatcttcc cccgggcccc ctgacttgtc cagcctgcca actacactag

2821 cccacttgca aactatagac cttaaagacg cctttttcca aatcccctta cctaaacagt

2881 tccagcccta ctttgctttc actgtcccac agcagtgtaa ctacggcccc ggcactagat

2941 acgcctggaa ggtactaccc caagggttta aaaatagtcc caccctgttc gaaatgcagc

3001 tggcccatat cctgcagccc attcggcaag ctttccccca atgcactatt cttcagtaca

3061 tggatgacat tctcctggca agcccctccc atgaggacct actactactc tcagaggcca

3121 caatggcttc cctaatctcc catgggttgc ctgtgtccga aaacaaaacc cagcaaaccc

3181 ctggaacaat taagttccta gggcaaataa tttcacccaa ccacctcact tatgatgcag

3241 tccccacggt acctatacgg tcccgctggg cgctacctga acttcaagcc ctacttggcg

3301 agattcagtg ggtctccaag ggaactccta ccttacgcca gccccttcac agtctctact

3361 gtgccttaca aaggcatact gatccccgag accaaatata tttaaatcct tctcaagttc

3421 aatcattagt gcagctgcgg caggccctgt cacagaactg ccgcagtaga ctagtccaaa

3481 ccctgcccct cctaggggct attatgctga ccctcactgg caccactact gtagtgttcc

3541 agtccaagca gcagtggcca cttgtctggc tacatgcccc cctaccccac actagccagt

3601 gcccctgggg gcagctactt gcctcagctg tgttattact cgacaaatac accttgcaat

3661 cctatgggct actctgccaa accatacatc ataacatctc cacccaaacc ttcaaccaat

3721 tcattcaaac atctgaccac cccagtgttc ctatcttact ccaccacagt caccgattca

3781 aaaatttagg tgcccaaact ggagaacttt ggaacacttt tcttaaaaca gctgccccat

3841 tggctcctgt aaaagccctc atgccagtgt ttactctttc cccggtgatc ataaacaccg

3901 ccccctgcct gttttcagac ggatctacct cccgggcagc ctatattctc tgggacaagc

3961 atacattgtc acaaagatca ttcccccttc cgccaccgca caagtcggcc caacgggccg

4021 aacttctcgg acttttgcat ggcctttcca gcgcccgttc gtggcgctgt ctcaacatat

4081 ttctagactc caagtatctt tatcattacc ttcggaccct tgccctgggc accttccaag

4141 gcaggtcctc tcaggccccc tttcaggccc tcctgccccg cttactatcg cgtaaggtcg

4201 tctatttgca ccacgttcgc agccatacca atctacctga tcccatctcc aggctcaacg

4261 ctctcacaga tgccctacta atcacccctg tcctgcagct ctctcctgca gaactacaca

4321 gtttcaccca ttgcggacag acggccctca cattgcaagg ggcaaccaca actgaggctt

4381 ccaatatcct gcgctcttgc cacgcctgcc gcaaaaataa cccacaacat cagatgcctc

4441 ggggacacat ccgccgtggc ctacttccta accacatctg gcaaggcgac attacccatt

4501 tcaaatataa aaatacgctg taccgccttc atgtatgggt agacaccttt tcaggagcca

4561 tctcagctac ccaaaagaga aaagaaacaa gctcagaagc tatttcctct ttgcttcagg

4621 ccattgccta tctaggcaag cctagctaca taaacacaga caacggccct gcctatattt

4681 cccaagactt cctcaatatg tgtacctccc ttgctattcg acatactacc catgtcccct

4741 acaatccaac cagctcagga cttgtagaac gctctaatgg cattcttaaa accctattat

4801 ataagtactt tactgacaaa cccgacctac ccatggataa tgctctatcc atagccctat

4861 ggacaatcaa ccacctgaat gtgttaacca actgccacaa aacccgatgg cagcttcacc

4921 actccccccg actccagccg atcccagaga cacattccct cagcaataaa caaacccatt

4981 ggtattattt caagcttcct ggtcttaata gccgccagtg gaaaggacca caggaggctc

5041 tccaagaagc tgccggcgct gctctcatcc cggtaagcgc tagttctgcc cagtggatcc

5101 cgtggagact cctcaagcga gctgcatgcc caagacccgt cggaggcccc gccgatccca

5161 aagaaaaaga ccaccaacac catgggtaag tttctcgcca ctttgatttt attcttccag

5221 ttctgccccc tcatcctcgg tgattacagc cccagctgct gtactctcac aattggagtc

5281 tcctcatacc actctaaacc ctgcaatcct gcccagccag tttgttcgtg gaccctcgac

5341 ctgccggccc tttcagcaga tcaggcccta cagcccccct gccctaatct agtaagttac

5401 tccagctacc atgccaccta ttccctatat ctattccctc attggattaa aaagccaaac

5461 cgaaatggcg gaggctatta ttcagcctct tattcagacc cttgttcctt aaagtgccca

5521 tacctggggt gccaatcatg gacctgcccc tatacaggag ccgtctccag cccctactgg

5581 aagtttcagc aagatgtcaa ttttactcaa gaagtttcac gcctcaatat taatctccat

5641 ttttcgaaat gcggttttcc cttctccctt ctagtcgacg ctccaggata tgaccccatc

5701 tggttcctta ataccgaacc cagccaactg cctcccaccg cccctcctct actcccccac

5761 tctaacctag accacatcct cgagccctct ataccatgga aatcaaaact cctgaccctt

5821 gtccagttaa ccctacaaag cactaattat acttgcattg tctgtatcga tcgtgccagc

5881 ctatccactt ggcacgtcct atactctccc aacgtctctg ttccatcctc ttcttctacc

5941 cccctccttt acccatcgtt agcgcttcca gccccccacc tgacgttacc atttaactgg

6001 acccactgct ttgaccccca gattcaagct atagtctcct ccccctgtca taactccctc

6061 atcctgcccc ccttttcctt gtcacctgtt cccaccctag gatcccgctc ccgccgagcg

6121 gtaccggtgg cggtctggct tgtctccgcc ctggccatgg gagccggggt ggctggcggg

6181 attaccggct ccatgtccct cgcctcagga aagagcctct tacatgaggt ggacaaagat

6241 atttcccaat taactcaagc aatagtcaaa aaccacaaaa atctactcaa aattgcgcag

6301 tatgctgccc agaacagacg aggccttgat ctcctgttct gggagcaagg aggattatgc

6361 aaagcattac aagaacagtg ctgttttctg aatattacta attcccatgt ctcaatacta

6421 caagagagac cccccctgga gaatcgagtc ctgactggct ggggccttaa ctgggacctt

6481 ggcctctcac agtgggctcg agaggcctta caaactggaa tcacccttgt cgcgctactc

6541 cttcttgtta tccttgcagg accatgcatc ctccgtcagc tacgacacct cccctcgcgc

6601 gtcagatacc cccattactc tcttataaac cctgagtcat ccctgtaaac caagcacata

6661 attattgcaa ccacatcgcc tccagcctcc cctgccaata attaacctct cccattaaat

6721 cctccttctc ctgcagcaac ttcctccgtt cagcctccaa ggactccacc tcgccttcca

6781 actgtctagt atagccatca atccccaact cctgcatttt ttctttccta gcactatgct

6841 gtttcgcctt ctcagcccct tgtctccact tgcgctcacg gcgctcctgc tcttcctgct

6901 ttctccgggc gacgtcagcg gccttcttct ccgcccgcct cctgcgccgt gccttctcct

6961 cttccttcct tttcaaatac tcagcaatct gcttttcctc ctctttctcc cgctcttttt

7021 ttcgcttcct cttctcctca gcccgtcgct gccgatcacg atgcgtttcc ccgcgaggtg

7081 gcgctttctc ccctggaggg ccccgtcgca gccggccgcg gctttcctct tctaaggata

7141 gcaaaccgtc aagcacagct tcctcctcct ccttgtcctt taactcttcc tccaaggata

7201 atagcccgtc caccaattcc tccaccagca ggtcctccgg gcatggcaca ggcaagcatc

7261 gaaacagccc tacagataca aagttaacca tgcttattat cagcccactt cccagggttt

7321 ggacagagcc ttcttttcgg atacccagtc tacgtgtttg gagactgtgt acaaggcgac

7381 tggtgcccca tctctggggg actatgttcg gcccgcctac atcgtcacgc cctactggcc

7441 acctgtccag agcatcagat cacctgggac cccatcgatg gacgcgttat cggctcagct

7501 ctacagttcc ttatccctcg actcccctcc ttccccaccc agagaacctc taagaccctc

7561 aaggtcctta ccccgccaat cactcataca acccccaaca ttccaccctc cttcctccag

7621 gccatgcgca aatactcccc cttccgaaat ggatacatgg aacccaccct tgggcagcac

7681 ctcccaaccc tgtcttttcc agaccccgga ctccggcccc aaaacctgta caccctctgg

7741 ggaggctccg ttgtctgcat gtacctctac cagctttccc cccccatcac ctggcccctc

7801 ctgccccacg tgattttttg ccaccccggc cagctcgggg ccttcctcac caatgttccc

7861 tacaagcgaa tagaagaact cctctataaa atttccctta ccacaggggc cctaataatt

7921 ctacccgaag actgtttgcc caccaccctt ttccagcctg ttagggcacc cgtcacgcta

7981 acagcctggc aaaacggcct ccttccgttc cactcaaccc tcaccactcc aggccttatt

8041 tggacattta ccgatggcac gcctatgatt tccgggccct gccctaaaga tggccagcca

8101 tctttagtac tacagtcctc ctcctttata tttcacaaat ttcaaaccaa ggcctaccac

8161 ccctcattcc tactctcaca cggcctcata cagtactctt cctttcataa tttacatctc

8221 ctgtttgaag aatacaccaa catccccatt tctctacttt ttaacgaaaa agaggcagat

8281 gacaatgacc atgagcccca aatatccccc gggggcttag agcctcccag tgaaaaacat

8341 ttccgcgaaa cagaagtctg aaaaggtcag ggcccagact aaggctctga cgtctccccc

8401 cggagggaca gctcagcacc ggctcaggct aggccctgac gtgtccccct gaagacaaat

8461 cataagctca gacctccggg aagccaccgg aaccacccat ttcctcccca tgtttgtcaa

8521 gccgccctca ggcgttgacg acaacccctc acctcaaaaa acttttcatg gcacgcatat

8581 ggctgaataa actaacagga gtctataaaa gcgtggagac agttcaggag ggggctcgca

8641 tctctccttc acgcgcccgc cgccctacct gaggccgcca tccacgccgg ttgagtcgcg

8701 ttctgccgcc tcccgcctgt ggtgcctcct gaactgcgtc cgccgtctag gtaagtttag

8761 agctcaggtc gagaccgggc ctttgtccgg cgctcccttg gagcctacct agactcagcc

8821 ggttctccac gctttgcctg accctgcttg ctcaactctg cgtctttgtt tcgttttctg

8881 ttctgcgccg ctacagatcg aaagttccgc ccctttccct ttcattcacg actgactgcc

8941 ggcttggccc acggccaagt accggcgact ccgttggctc

In III bzw III sind die codierenden Sequenzabschnitte gekennzeichnet. Es befinden sich insgesamt 7 Gene auf der DNA-Sequenz

**First 100 bp of genome sequence (part of gene 1):**

atgggcca aatcttttcc cgtagcgcta gccctattcc

gcggccgccc cgggggctgg ccgctcatca ctggcttaac ttcctccaag cggcatatcg

cc

*(803-902 bp)*

**Aufgabe 3:**

Codierende Sequenzen in Aminosäuren übersetzt (1. 5´3`Frame):

1.Gen (803-2092 bp):

MGQIFSRSASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFH

QLKKFLKIALETPVWICPINYSLLASLLPKGYPGRVNEILHILIQTQAQIPSRPAPPP

PSSPTHDPPDSDPQIPPPYVEPTAPQVLPVMHPHGAPPNHRPWQMKDLQAIKQEVSQA

APGSPQFMQTIRLAVQQFDPTAKDLQDLLQYLCSSLVASLHHQQLDSLISEAETRGIT

GYNPLAGPLRVQANNPQQQGLRREYQQLWLAAFAALPGSAKDPSWASILQGLEEPYHA

FVERLNIALDNGLPEGTPKDPILRSLAYSNANKECQKLLQARGHTNSPLGDMLRACQT

WTPKDKTKVLVVQPKKPPPNQPCFRCGKAGHWSRDCTQPRPPPGPCPLCQDPTHWKRD

CPRLKPTIPEPEPEEDALLLDLPADIPHPKNSIGGEV

2.Gen (2994-5189 bp):

MQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSEATM

ASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLG

EIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRL

VQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDK

YTLQSYGLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTF

LKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKHTLSQRSFPLPP

PHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQA

LLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQLSPAELHSFTHCGQT

ALTLQGATTTEASNILRSCHACRKNNPQHQMPRGHIRRGLLPNHIWQGDITHFKYKNT

LYRLHVWVDTFSGAISATQKRKETSSEAISSLLQAIAYLGKPSYINTDNGPAYISQDF

LNMCTSLAIRHTTHVPYNPTSSGLVERSNGILKTLLYKYFTDKPDLPMDNALSIALWT

INHLNVLTNCHKTRWQLHHSPRLQPIPETHSLSNKQTHWYYFKLPGLNSRQWKGPQEA

LQEAAGAALIPVSASSAQWIPWRLLKRAACPRPVGGPADPKEKDHQHHG

3.Gen (5126-5185 bp und 7304-7813 bp):

MPKTRRRPRRSQRKRPPTPWPTSQGLDRAFFSDTQSTCLETVYK

ATGAPSLGDYVRPAYIVTPYWPPVQSIRSPGTPSMDALSAQLYSSLSLDSPPSPPREP

LRPSRSLPRQSLIQPPTFHPPSSRPCANTPPSEMDTWNPPLGSTSQPCLFQTPDSGPK

TCTPSGEAPLSACTSTSFPPPSPGPSCPT

4.Gen (5182-5185 bp und 7304-8361 bp):

MAHFPGFGQSLLFGYPVYVFGDCVQGDWCPISGGLCSARLHRHA

LLATCPEHQITWDPIDGRVIGSALQFLIPRLPSFPTQRTSKTLKVLTPPITHTTPNIP

PSFLQAMRKYSPFRNGYMEPTLGQHLPTLSFPDPGLRPQNLYTLWGGSVVCMYLYQLS

PPITWPLLPHVIFCHPGQLGAFLTNVPYKRIEELLYKISLTTGALIILPEDCLPTTLF

QPVRAPVTLTAWQNGLLPFHSTLTTPGLIWTFTDGTPMISGPCPKDGQPSLVLQSSSF

IFHKFQTKAYHPSFLLSHGLIQYSSFHNLHLLFEEYTNIPISLLFNEKEADDNDHEPQ

ISPGGLEPPSEKHFRETEV

5.Gen (5182-5185 bp und 6832-7553 bp):

MALCCFAFSAPCLHLRSRRSCSSCFLRATSAAFFSARLLRRAFS

SSFLFKYSAICFSSSFSRSFFRFLFSSARRCRSRCVSPRGGAFSPGGPRRSRPRLSSS

KDSKPSSTASSSSLSFNSSSKDNSPSTNSSTSRSSGHGTGKHRNSPTDTKLTMLIISP

LPRVWTEPSFRIPSLRVWRLCTRRLVPHLWGTMFGPPTSSRPTGHLSRASDHLGPHRW

TRYRLSSTVPYPSTPLLPHPENL

6.Gen (5182-6648 bp):

MGKFLATLILFFQFCPLILGDYSPSCCTLTIGVSSYHSKPCNPA

QPVCSWTLDLPALSADQALQPPCPNLVSYSSYHATYSLYLFPHWIKKPNRNGGGYYSA

SYSDPCSLKCPYLGCQSWTCPYTGAVSSPYWKFQQDVNFTQEVSRLNINLHFSKCGFP

FSLLVDAPGYDPIWFLNTEPSQLPPTAPPLLPHSNLDHILEPSIPWKSKLLTLVQLTL

QSTNYTCIVCIDRASLSTWHVLYSPNVSVPSSSSTPLLYPSLALPAPHLTLPFNWTHC

FDPQIQAIVSSPCHNSLILPPFSLSPVPTLGSRSRRAVPVAVWLVSALAMGAGVAGGI

TGSMSLASGKSLLHEVDKDISQLTQAIVKNHKNLLKIAQYAAQNRRGLDLLFWEQGGL

CKALQEQCCFLNITNSHVSILQERPPLENRVLTGWGLNWDLGLSQWAREALQTGITLV

ALLLLVILAGPCILRQLRHLPSRVRYPHYSLINPESSL

7.Gen (6836-7135 bp):

MLFRLLSPLSPLALTALLLFLLSPGDVSGLLLRPPPAPCLLLFL

PFQILSNLLFLLFLPLFFSLPLLLSPSLPITMRFPARWRFLPWRAPSQPAAAFLF

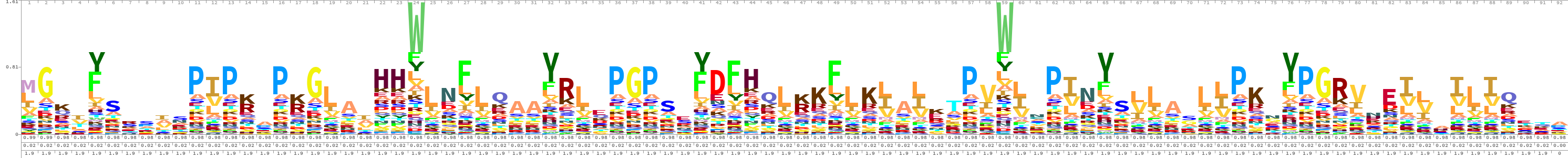
**Die ersten 30 AS des 5´3´-Frames des 1. Gens lauten:**

MGQIFSRSASPIPRPPRGLAAHHWLNFLQA

1. In der Genomsequenz liegen Basentripletts vor, die für bestimmte Aminosäuren codieren. Jedoch ist es möglich, dass mehrere unterschiedliche Basentripletts für ein und dieselbe Aminosäure codieren. Es gibt 64 mögliche Codon-Variationen, jedoch nur 21 Aminosäuren, die codiert werden. Der genetische Code ist also „degeneriert“. Sucht man nun in der Aminosäuresequenz, hat man zum einen eine viel kürzere Sequenz (nur 1/3 so lang) und zum anderen ist die Charakterisierung deutlich einfacher (eine eindeutig zugeordnete Aminosäure). Des Weiteren kann mithilfe der Aminosäure-Sequenz direkt gezielter nach Genen gesucht werden.
2. Es ist sinnvoll alle 6 möglichen Übersetzungs-Frames zu durchsuchen, da zum Einen nicht immer klar ist, ab welcher Base translatiert wird und ob man sich auf dem codierenden- oder nicht-codierenden-Strang befindet. Es ergeben sich 6 „reading frames“, da die DNA in aus 3 Basen bestehenden Codons eingeteilt wird, was 3 Leserahmen ergibt. Außerdem ist die DNA eine Doppelhelix, also gibt es 3 weitere Leserahmen auf dem komplementären Strang.

**Aufgabe 4:**

*http://pfam.xfam.org/family/PF02228.15#tabview=tab4*



Abweichungen von eingegebener Suchsequenz in III markiert:

Eingegebene Suchsequenz:

MGQIFSRSASPIPRPPRGLAAHHWLNFLQA

Erhaltenes HMM Logo:

MGKIYSRSISPIPKAPKGLAIHHWLNFLQA(AYRLFPGPSEYDFHQLKKFLKLALKTPVWLNPINYSLLASLLPKNYPGRVNEILAILIQETA)

# In Klammern ist die weitere Aminosäure-Sequenz des Major core protein p19, welches über die HMM-Suche, aufgrund des eigegebenen 30 AS – langen Abschnittes, erhalten wurde.

# Die eingegebene Sequenz weicht von der erhaltenen Sequenz um 23,3 % ab. Die Abweichung beträgt ungefähr ¼ und kann beispielsweise durch vereinzelte Mutationen zustande gekommen sein.

# Hätte man nicht nur die ersten 30 Aminosäuren des 1. Gens, sondern die gesamte Aminosäure-Sequenz mit der Pfam-Datenbank verglichen, hätte man 3 HMM-Logos erhalten:

# https://www.ebi.ac.uk/Tools/hmmer/results/9D7DA4D2-7E28-11E8-8778-339353F04F9B/score (02.07.18, 20:49 Uhr)

# Aufgabe 5:

# Ausgewähles Genom: *Blueberry fruit drop associated virus*

# Wiederholung Aufgabe 2:

# Komplettes Genom:

# *(https://www.ncbi.nlm.nih.gov/nuccore/NC\_028462.1 (01.07.18, 20:15 Uhr))*

ORIGIN

1 tggtatcaga gccctgtaac tctcctgtta ctcggcattt attattagca tatgtttgat

61 tggacatttt gtagatctcc cacagcatca cagatcgctt ggaatagaca aaattgccct

121 aatcacctat ccataactag tgtgcatcgt agtcacattt cgtcaggctt agaagccaaa

181 cctagtgtct tagctgaagt cctcaattat cagtcagatc agttatttga cataaaagga

241 aaatctcaaa aaatagcaga agtcgtagac tggactgcag agtttttaga aaatcaaaaa

301 caatcagatc tgcctaacaa caagcttgac ttcgtcaaga cttccatcga cttgttagca

361 acagaatttg cagggttcaa agaagccaga acagagttca ttgaacaaat tatcaggcgt

421 gttcttgacg ctgataaaaa gaatcttacc caactcaccg gggtaattga agaacaagtc

481 tctcaactca aaggagtaat tgaagaactt aggtttgacc cagataaacc taagccagct

541 gaaatacgaa cacaaacact tgcagaagaa atagcagctt cggcggttgc tatccagaag

601 gtccttagca aactagagga gctagtggct aaggttgact ctacggtatc agaagttttc

661 agagtcacca agagaatcga ggctacctca acatgagtgg cttcaaacta tcttcaccat

721 ccgacgaaga tttgaaccgt tggctaactg atgagttaga tcgggttcaa aataaaatta

781 ttggaaaggt caaagatgtt gacccagatc atgtcgccat ggcagaacca ggcgaaaatc

841 atttacgaaa tggtaaagtt ctgattcacc aaatgaatcg aatgtttgcc accatcaata

901 ccatgttcga aaagcaactt tggttcgaac aaagacagaa ggtgctattc caaaatcttt

961 gggattttct caacagctat gatgagaacc agaaggcaca acagaaagct ttaaaagaca

1021 gtcaagacaa tattgctctt gccacaagaa acattgctcg catgttggag caagtcaaag

1081 gaattgaaga agtcatcaat actgtctttc aaaccattga aggaattctc aatgtcaaca

1141 aacagtttat gacagcagag aaacaaaggg agcagatctt gcaaagtcaa cttaagcaaa

1201 ttaaagacca attggcagaa ccagttgtgt taagcaaacc tgtcagtgat ccagctcttg

1261 aaggtcacct caaaagcatc aacagaggag taaaaaacct cgttgaacgt gcagaactcc

1321 taatcgacta tgaagaactt gctgcaaaaa ttggcaagaa tatttctact caaggaattg

1381 cagtcagtga aaggagcaaa tcatttatga gcttcgatca cccaaagcca gtagaatcag

1441 ttctacccaa ggcaagaatg ttcagtatgt ttgcaggagg caaacccaaa agcagcattg

1501 tcagatccct ggagttagaa gttgaagaac aacctaaagt cctcaaagac tataaggttc

1561 caaaattatc tgtcaagcag ctctatgaac aaggaagcat gatgcagcgc aacgcttatc

1621 gcatcaagca gatcgagaga aagattccct acgatcccaa cactaaagtt atccatcttg

1681 tggataaagg agaaatggat cgcatgaagg caaagggata cgagtatgtg cacttcggag

1741 cagtccaagt tgttttcaga ctattggcca gaagagatct caactgctca gttcttagca

1801 tttgcagaga caaccgttta aaatctcttc aaaagggatt actcggggct ttgcaagcaa

1861 gcctcaacaa tcaggttgca tattttgtat gcatccccaa tttctcgtgc acaattgaag

1921 atgcccatga agctttggtt ttatgtgttc agacacatgg aacagatttc aaagaaggct

1981 acactgatct tgccattgaa tacattgtca catacaaatt tatgacaagt gatatggaag

2041 ccatgtcaaa actcaaagcc actgaaggat tcaactctat ggtaatttct gatcgaaata

2101 accataccat ttctcatact aagaagatag actggaagga actcaaccaa agtgagtttc

2161 ccaaagaatg gagtatcaac agtgaagaac ctccagaagc caagtcttcg tcagtcatca

2221 ctggcattta tcagcatgaa aatggcaata taagtatgcg ttttcaccaa gctggagaag

2281 gaagtagtca cagcaaagct caaacgcatc ctgttcacat cttcgatgat gaacccattc

2341 aactggatgt atctaaatct ggcagaccta aggccagaat gatcagagct agcacaagaa

2401 aggaagttgt ggtggaatta catgacatca acagcatcag agaaatctcc gacgaaaaac

2461 tcgacgagat ttatgctacc acttcagatc cacagtactt caactataca gtactgccag

2521 agatgcaaag aagaatcaac caagggatat cggtcagagc tatcagtcaa tcggtcaaca

2581 gacccagtac ctctggagat agcaaagagg tgcccattac ctctgaagtc cctagcagca

2641 gtacactcga aggcaaagga aaaactaaga agttcactcc agctggatca agtatcaagc

2701 taccactcat caccggagta ccaactggaa agcctaacga gttcaacact gaaagataca

2761 aaccaaggat tgaacccaat ccggctgtat caccaaatgg agtctggctc gacatcgaca

2821 acaaccccga tccaagagga aatattgacg aatgggtcaa tactatgaac atgttgattg

2881 tctcagaaac cggaaaagca tggagcccag atgaagcaag aacctactat gcctgtacat

2941 tcaaaggcat agccagaaaa tggttcgaca aatacaagga gacagatgaa tacgtcttat

3001 ggttgagcac catagaaagg tcaaagtcac cctctgactt tgccacatgt ctctacaagc

3061 agttctgcag agaacaagat cccagcaaga aggagaagga tcttgagatt ctaaaaggaa

3121 aacttgtgaa tctcaccatt tcagatctaa agtacttcct cagttatgtc aacgagtaca

3181 acttggcata ctacaacctt ggcatattca acaataacac atacagagat atgttcattc

3241 aaaagcttcc agatccatgg aaggaatatc ttaccaatga aagagccaaa gaagcaggaa

3301 atgaagaatg gacgcttgga aatcttatcg agtttgtcaa gacaaaactt gcagatctcg

3361 ccattcaagc ctggagagcc aagaggatgc aaaaacttta cagcaaaagt ggtctcagtt

3421 tttgccagca tctcattgac atccctactc aatggggtgg aagaaacaaa gagctcaagc

3481 ataagaagaa aggaaagaag atcaaatctt tcaagagata caagttcaga aaatccaaga

3541 agaaaggatt tgaaaaatca aagatcaatc cccagaagaa tttcaaaaga agattcaaac

3601 agaagtctgg aaaggagatt gaggatcgaa agaagcagag caaaagcaaa cctacatcct

3661 ctaaaggagt agccaacgtt tgctggaact gtaatcagca gggtcatttc tcaagagaat

3721 gcccaagaaa gaatgctcgt cttattgagc tacaagaagc agcagataca tgggaaattc

3781 ctctcatgag aattgaagaa gacgacgact acagtgatac ttcaagtatc tattcacttg

3841 aaacagaaag cgattcagag tcatctgatt ctgatcaaga aagtgaagac gaagacactg

3901 tcaggatgat caagacccaa tttggtatca agacaatacg cttgaacaat tggccacatg

3961 acacccagga gattggggac gtcgtgtctt ttgaagctat actggagtat gaaaagttcg

4021 agtgtgattt ctgcaaaaca tctgcagacc aatacacacc agtcatacaa cgaagcgatc

4081 ttagcaaaga aggtaccatc agctatcaca catggtgttt tacttctctt gtcatgagag

4141 aaaccaagaa cgctgaggtc ccacaactgg tacaaagaga agttgatttg gcaaaactca

4201 gagatgccaa aaagaagttt gaacaagctc agctcaaaga agaacaagaa agacttagag

4261 caacaagaag acctgtcata cttgaggtct atgatgatgc agaggagctt gaagaaatgt

4321 caaaaattgg aggccatttg gaagaactat ccatcaaact aagaaggcct gttgctgcta

4381 gtccaggtgc tacaagcagt aatccacttg aagaaaacgt ggatactccc aaagacctaa

4441 aaggtcaaat acttctctgc cccatattca aaggtccaat ccctaagaga actgaagctg

4501 attggaaaag agaagaagaa gagagaaaca gaaagcttgt tgaaagctac agaaaattga

4561 agctgtttga aagatataag gaagttagga atgaagtcct acagcaaatt ccacagcatc

4621 aaattcccat caacacaatc atcaccgtaa agagttctcc gacggagatg cttgctgata

4681 ttgaaaaact gatggaagaa gtcaagaagc ttcaacagaa ggaaagcacg ccagaagata

4741 agccagatgc aacccctgaa ttcaggcata aagacaacag catctacatc tctgccgctt

4801 tcagattcaa aggatatgaa gaatatcatc tccactgtct agtcgacacc ggagccagca

4861 tcacacttgc tacccaatgg gccattccag aagaaaaatg gatcactatt ccccacggaa

4921 atgaaatcaa aatcacagtg gcaaatggag aaacttgtaa aatcgacaag tatgccacca

4981 atgtgcccat atacatttca ggtatcaagt ttactattga aaaagtattg caatttaata

5041 agcaaggttg cgatttcctc atcggaaatg atttctttat gaaatacgca ccatatacgc

5101 aatacccaga tcgtatttca atagccaaaa acggaatttc gtatatcaca aacgtggtag

5161 aacgaccagt ttccgtagca tacagaaaag aatttctgga aagacgaaaa attccagacc

5221 atcagcagtc acacaaggtg aagtgtggtg aaaattacaa agtccaaaaa caagcccaag

5281 aaaatgactt cactattgaa gactatggta ctacggatag cgaagctgaa gagcccatat

5341 atgatgccat atatgacaaa ggatttcagc ccatcgaatc tgacgatgag cccaatgacg

5401 aaatccagcc catcaatgat gagcccaatg aagaagtcag gcccggattc cagcccatcg

5461 aatctaacga tgagcccaat aatgaaatcc agcccagcaa tgacgagccc aataaagagg

5521 tcaggcccaa gccaacaaaa gatatcaggc ccaagcccat tactgaagaa gacttccaag

5581 cagagcttcg cctcgtcgaa gcccactatg aaaagcaggc taggcccaaa caacggccca

5641 acctaaatgt gctcgaagaa caaatcagca gatgcagcat tgaagagctc accaaacagc

5701 ttgagcccaa catcagtgac aatccaatcc acaagtggga tgtcacccaa acctatgctg

5761 acatacaact caaggatccg tatgcactca tacgtgtaaa gcccatgcca tacagtgctg

5821 aagatgaagc tgagtttgaa agacaactcg aagagcaact caaactggaa ctcattcagc

5881 ccagtaagag tccacattcc agcccagcat tctgtgtcag aaatcatgct gagatcaaga

5941 gacagaaagc ccgcattgtc atcaactaca aagaactcaa tcgagtcacc aaagacgatg

6001 gctactttct acccaatttg gagaccttag tatatcaagt tgctgaggcc agagttttct

6061 ccaagttcga ctgtaagagt ggatattggc aaatcaagct cacacctgag agcatcccac

6121 tcactgcatt cagcaccccc aagggtcagt atgaatggaa ggtccttcca tttggattaa

6181 agaatgcacc aaacatattc cagcgaagga tggacaaaat cttcaaagat tgttatcagt

6241 tctgtggtgt atacgtcgac gatatactcg tcttctccaa agacaaagaa gaacacctac

6301 tgcatctgga tatcatcatc aacaaaattg tccagaatgg aatcatcatc ggaagaacca

6361 agtacgaatt ggtcagagaa agcattgact ttttaggagt caccatatgc aaaggaaaga

6421 tcaagttgca gcctcacatc cttactcacc tcatggagtt tcctgacaaa ctcaaggatc

6481 gacttcaggt ccagcagttt ctcggatgtg tcaactacgc cggaaagttc atacaaaggc

6541 ttgcagagtt atgcagtgtg ttgcatacca aactcaagaa agatcacaag tggaagttca

6601 cagaggacga tgtcaaaaga gtccaactca ttaagaaagt ctgtgccaat cttccagagc

6661 ttaccttgcc cccaccaaca tggaagatcg ttttgcaaac tgatgcaagt gatcaccact

6721 gggcaggaac aattggagct atcagccctg acaagaagga agaagtaatt tgtgcttaca

6781 gaagcggaac gttcaaacct gctgagcaaa attactccac ccaagaaaaa gaagtcttag

6841 ccatcatccg tactattcaa aaggccaaga tcttcgtcct aaaacccttc cttgtccgaa

6901 ccgattcaaa gtttgcagcc ctattccttg ataaaaagat caacgagtct ttagctaggg

6961 gtcgtctcat tagatggcaa atcatgctta ggcagtacta cttagacatt gaacacattc

7021 ctggtgtttc taattatttg cctaatgccc ttactcgtga aatggcaggt tgcaatatca

7081 tttaccagat caaaatggta aaccatggaa aagccccaag gatggatctt agagatgtca

7141 tcacttccaa ggctagagct agagaaaatc agcagctcga aaatgaagac atccattgga

7201 aacagcatct tgcttcaaca tcaagaagct ttgctaaaag aaaagatgtt ccaatgccaa

7261 gtccaattcc tccaccgttc gaatacaaga cttatcgacg gctaactgag gaacagaaaa

7321 gtcagcttga ctggcttacc agctcatggc acacagttga tcaagacaca ttcttcgatg

7381 tgttaaaagc catggccaaa gaatttgaag gactcaacaa attgaagaat ggcaatccag

7441 ccaaaagagt caaaacaagc agagaggcca aactggccag atggaaagct caacccaaga

7501 agttcacctt ctttgaagaa tggcgtacaa tcctactacg gaggctacgc aactggacca

7561 acggagctac cataatggac gaagaattgt atcaacggat catcaacact ccttgcagag

7621 cccaattgca atggtggtcc caactcaaca agttatggca gtatcagccg cctgagcaaa

7681 gtttcatctg caaagctact tgctttgaca acaatcatac catgctcaaa ctcaaatgtc

7741 ctcccatcaa caaactcacc cctgagttca tggactggtt tgaaaatgga ttcattcaaa

7801 gctttgccct taagagaacc agtgattaca gtttattacc tactgatttt atcaaggtaa

7861 ttgaacagat ttgggctatt aaagatattg gtaatacatt gtttattatg atgtattcag

7921 ccccaggaga atggaactac aaagattgtt cctattatta tccttatcat ctggtgagga

7981 tttatgatag agttccatct gtcactagag atacaatcct tgaaggagaa atcttccccc

8041 atctcagaag ccaggagcaa attcaaagag atgaagaaat gggtatgtac gaaccactgt

8101 actgtcccat cgacgattgg ctcagccgtg acccgagaca tgtcagcaaa tggagagcct

8161 ttgctctcat ggagatacaa gacctcatca atgaagacaa gggatttttg ccacgatggc

8221 attacatcca cagcaaccag aggttcatct ttgaagttga aggaatcagc atgacgcgtg

8281 cttgtgaaag ttacagcctt ttccaaagct tatccttaaa ggtacccatg aattatgata

8341 cccagctgga aattgactgt attcagacat tggatgctgc tcaaagtgaa gacccatgtg

8401 acgagccata tgatgaatca gattcgtact ggcaggacat ggatgatgag acgtggcaca

8461 atattgaaaa catgatggaa ggataaggcc tggcagactc caccatccca tgtgctactt

8521 ttgctttcgg tgtgctttat tgctttcggt ttgtcggttt gctttcggca gagcatcttg

8581 cttttgcttt tgttagtgcg tgtacgaaaa gcatcttttg cttttcattt gtcggaaatt

8641 tcatttgaat ttccatttca tttgcttttg taaactccta tataaggaga tgtgtgttta

8701 gttgtaaggc attcgaatcc caacctagaa atcactacat attgtaagaa gtgtggaagc

8761 tctcaagatc ttcaagattt ctcttgtaag ttctttttat gaataaaagt ttgatgtttt

8821 attctaacat ggagtagtcc ttagggcgcc tctctggagg aggcctctgt tatgcatgtt

8881 tgtggttatg ctttaactat atgaatgttg tattagagta tgatataact atggaggttc

8941 gaagcactag agtctgttaa cataactcta tgataaacca gaagagaggt tatggtttca

9001 ttacttcact catgtctact gcaaaatttt ggaaaacaaa tgttctttgt caccagacct

9061 tatcaaatat acttttttgg ctaatgggaa aaatggagtt ttctgtttgg aagaacgaag

9121 gaggaatggc agtgaggtgt aggcctagat cccggtgttt gccaagaggg cgttgtgttg

9181 ggtgtggttg aagccgaagg caagtttcct ctggttgttt gtattcttaa aacaaagttt

9241 ggcccttcta gaggaaaagg tttgataaac gaatggtatt ttggtaacaa aagattttta

9301 gtatctgaaa tattgcagat gtatgagcgc tcttgtattt tttggccaca agtcttggta

9361 tggtttatct atgatgagtt tgaaactacg ttgcttcaga gtccttgtat atcagagtct

9421 tttataacgg aagttggttg gcccataccc acaggtcttg cataataagc cgtccgaaag

9481 agaatagcca tttacattat attcctgttt agagcatcac ttttatgtcc tgttatcatt

9541 ggagaaatac agatctacag catataagtc ttcttgcatc tattcgtagt tatcagagca

9601 tttataatta taaacttgca tctttaggta ttaactatcc gcatagtctt gctttagaag

9661 ctagatctag aattagtcag ttatttagta tttactgcag taaaagtgtg cttacagtag

9721 atgagttgtt atattgctta gaaagtcata atttctttat ccgcattcaa gcactaccag

9781 atcgtgttta ctttaggttt tataggtctg tttgcacttg ttttgcatta taacatcatc

9841 ttgtttttat

In III ist der codierenden Sequenzabschnitt gekennzeichnet. Es befindet sich ein Gen auf der DNA-Sequenz

**First 100 bp of genome sequence:**

atgagtgg cttcaaacta tcttcaccat

ccgacgaaga tttgaaccgt tggctaactg atgagttaga tcgggttcaa aataaaatta

ttggaaaggt ca

*(693-792 bp)*

# Wiederholung Aufgabe 3:

Codierende Sequenz in Aminosäure-Sequenz übersetzt (1. 5´3`Frame):

1.Gen (693-8486 bp):

MSGFKLSSPSDEDLNRWLTDELDRVQNKIIGKVKDVDPDHVAMA

EPGENHLRNGKVLIHQMNRMFATINTMFEKQLWFEQRQKVLFQNLWDFLNSYDENQKA

QQKALKDSQDNIALATRNIARMLEQVKGIEEVINTVFQTIEGILNVNKQFMTAEKQRE

QILQSQLKQIKDQLAEPVVLSKPVSDPALEGHLKSINRGVKNLVERAELLIDYEELAA

KIGKNISTQGIAVSERSKSFMSFDHPKPVESVLPKARMFSMFAGGKPKSSIVRSLELE

VEEQPKVLKDYKVPKLSVKQLYEQGSMMQRNAYRIKQIERKIPYDPNTKVIHLVDKGE

MDRMKAKGYEYVHFGAVQVVFRLLARRDLNCSVLSICRDNRLKSLQKGLLGALQASLN

NQVAYFVCIPNFSCTIEDAHEALVLCVQTHGTDFKEGYTDLAIEYIVTYKFMTSDMEA

MSKLKATEGFNSMVISDRNNHTISHTKKIDWKELNQSEFPKEWSINSEEPPEAKSSSV

ITGIYQHENGNISMRFHQAGEGSSHSKAQTHPVHIFDDEPIQLDVSKSGRPKARMIRA

STRKEVVVELHDINSIREISDEKLDEIYATTSDPQYFNYTVLPEMQRRINQGISVRAI

SQSVNRPSTSGDSKEVPITSEVPSSSTLEGKGKTKKFTPAGSSIKLPLITGVPTGKPN

EFNTERYKPRIEPNPAVSPNGVWLDIDNNPDPRGNIDEWVNTMNMLIVSETGKAWSPD

EARTYYACTFKGIARKWFDKYKETDEYVLWLSTIERSKSPSDFATCLYKQFCREQDPS

KKEKDLEILKGKLVNLTISDLKYFLSYVNEYNLAYYNLGIFNNNTYRDMFIQKLPDPW

KEYLTNERAKEAGNEEWTLGNLIEFVKTKLADLAIQAWRAKRMQKLYSKSGLSFCQHL

IDIPTQWGGRNKELKHKKKGKKIKSFKRYKFRKSKKKGFEKSKINPQKNFKRRFKQKS

GKEIEDRKKQSKSKPTSSKGVANVCWNCNQQGHFSRECPRKNARLIELQEAADTWEIP

LMRIEEDDDYSDTSSIYSLETESDSESSDSDQESEDEDTVRMIKTQFGIKTIRLNNWP

HDTQEIGDVVSFEAILEYEKFECDFCKTSADQYTPVIQRSDLSKEGTISYHTWCFTSL

VMRETKNAEVPQLVQREVDLAKLRDAKKKFEQAQLKEEQERLRATRRPVILEVYDDAE

ELEEMSKIGGHLEELSIKLRRPVAASPGATSSNPLEENVDTPKDLKGQILLCPIFKGP

IPKRTEADWKREEEERNRKLVESYRKLKLFERYKEVRNEVLQQIPQHQIPINTIITVK

SSPTEMLADIEKLMEEVKKLQQKESTPEDKPDATPEFRHKDNSIYISAAFRFKGYEEY

HLHCLVDTGASITLATQWAIPEEKWITIPHGNEIKITVANGETCKIDKYATNVPIYIS

GIKFTIEKVLQFNKQGCDFLIGNDFFMKYAPYTQYPDRISIAKNGISYITNVVERPVS

VAYRKEFLERRKIPDHQQSHKVKCGENYKVQKQAQENDFTIEDYGTTDSEAEEPIYDA

IYDKGFQPIESDDEPNDEIQPINDEPNEEVRPGFQPIESNDEPNNEIQPSNDEPNKEV

RPKPTKDIRPKPITEEDFQAELRLVEAHYEKQARPKQRPNLNVLEEQISRCSIEELTK

QLEPNISDNPIHKWDVTQTYADIQLKDPYALIRVKPMPYSAEDEAEFERQLEEQLKLE

LIQPSKSPHSSPAFCVRNHAEIKRQKARIVINYKELNRVTKDDGYFLPNLETLVYQVA

EARVFSKFDCKSGYWQIKLTPESIPLTAFSTPKGQYEWKVLPFGLKNAPNIFQRRMDK

IFKDCYQFCGVYVDDILVFSKDKEEHLLHLDIIINKIVQNGIIIGRTKYELVRESIDF

LGVTICKGKIKLQPHILTHLMEFPDKLKDRLQVQQFLGCVNYAGKFIQRLAELCSVLH

TKLKKDHKWKFTEDDVKRVQLIKKVCANLPELTLPPPTWKIVLQTDASDHHWAGTIGA

ISPDKKEEVICAYRSGTFKPAEQNYSTQEKEVLAIIRTIQKAKIFVLKPFLVRTDSKF

AALFLDKKINESLARGRLIRWQIMLRQYYLDIEHIPGVSNYLPNALTREMAGCNIIYQ

IKMVNHGKAPRMDLRDVITSKARARENQQLENEDIHWKQHLASTSRSFAKRKDVPMPS

PIPPPFEYKTYRRLTEEQKSQLDWLTSSWHTVDQDTFFDVLKAMAKEFEGLNKLKNGN

PAKRVKTSREAKLARWKAQPKKFTFFEEWRTILLRRLRNWTNGATIMDEELYQRIINT

PCRAQLQWWSQLNKLWQYQPPEQSFICKATCFDNNHTMLKLKCPPINKLTPEFMDWFE

NGFIQSFALKRTSDYSLLPTDFIKVIEQIWAIKDIGNTLFIMMYSAPGEWNYKDCSYY

YPYHLVRIYDRVPSVTRDTILEGEIFPHLRSQEQIQRDEEMGMYEPLYCPIDDWLSRD

PRHVSKWRAFALMEIQDLINEDKGFLPRWHYIHSNQRFIFEVEGISMTRACESYSLFQ

SLSLKVPMNYDTQLEIDCIQTLDAAQSEDPCDEPYDESDSYWQDMDDETWHNIENMME

G

**Die erste AS-Sequenz (79 bp lang) des 1. 5´3´-Frames, die ein HMM-Profil ergibt, liegt im Bereich 303-381 bp des 1. Gens und lautet:**

MMQRNAYRIKQIERKIPYDPNTKVIHLVDKGEMDRMKAKGYEYVHFGAVQVVFRLLARRDLNCSVLSICRDNRLK

SLQK

*(303-381 bp)*

# Wiederholung Aufgabe 4:

# *http://pfam.xfam.org/family/PF01107.17#tabview=tab4 (02.07.18, 20:43 Uhr)*

# HMM logo

# Das abgebildete HMM-Logo gehört zum Viral movement protein (MP).

Eingegebene Suchsequenz:

MMQRNAYRIKQIERKIPYDPNTKVIHLVDKGEMDRMKAKGYEYVHFGAVQVVFRLLARRDLNCSVLSICRDNRLK

SLQK

Erhaltenes HMM Logo:

Siehe Bild.

Es konnten keine übereinstimmenden Sequenzabschnitte zwischen dem erhaltenen HMM-Logo und der eingegebenen Sequenz erkannt werden.

Somit kann nicht festgelegt werden, wie ähnlich sich die Sequenzen bzw. kann gesagt werden, dass sie überhaupt nicht miteinander einstimmen.

Der Grund hierfür konnte nicht ermittelt werden, da mangelnde Kenntnisse zum HMM-Suchvorgang vorliegen.

Meines Erachtens hätte zumindest eine kurze Sequenz (einzelne Aminosäure-Abweichungen möglich) mit der eingegebenen Sequenz übereinstimmen müssen, um einen Treffer für ein charakteristisches Protein zu erlangen (hier: Viral movement protein).