

# 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  ttcccgcgaa acagaagtct gaaaagggtca gggcccagac taaggctctg acgtctcccc
121 cgggagggac agctcagcac cggctcaggc taggccctga cgtgtccccc tgaagacaaa
181 tcataagctc agacctccgg gaagccaccg gaaccaccca ttctctcccc atgtttgtca
241 agccgccttc aggcgttgac gacaacccct cacctcaaaa aacttttcat ggcacgcata
301 tggttgaata aactaacagg agtctataaa agcgtggaga cagttcagga gggggctcgc
361 atctctcctt cagcgccccg ccgccttacc tgaggccgcc atccacgcgc gttgagtcgc
421 gttctgcgcg ctcccgcctg tgggtgcctcc tgaactgcgt ccgccgtcta ggtaagttta
481 gagctcaggt cgagaccggg cctttgtccg gcgtccctt ggagcctacc tagactcagc
541 cggttctcca cgctttgcct gacctgctt gctcaactct gcgtctttgt ttcgttttct
601 gttctgcgcg gctacagatc gaaagttccg cccctttccc ttctattcac gactgactgc
661 cggcttgccc cagggccaag tacggcgac tccgttggt cggagccagc gacagcccat
721 tctatagcac tctccaggag agaaacttag tacacagttg ggggctcgtc cgggatacga
781 gcgccccctt attccctagg caatgggcca aatcttttcc cgtagcgcta gccctattcc
841 gggcgccccc cgggggctgg ccgctcatca ctggcttaac ttctccaag cggcatatcg
901 cctagaaccc ggtccctcca gttacgattt ccaccagttg aaaaaatttc ttaaaatagc
961 tttagaaaca ccggtctgga tctgtcccat taactactcc ctctagcca gctactccc
1021 aaaaggatac cccggccggg tgaatgaaat ttacacata ctcatccaaa cccaagccca
1081 gatcccgctc cgtcccgccc caccgcgcc gtcaccccc acccacgacc ccccggatcc
1141 tgatccacaa atccccctc cctatgttga gcctacggcc cccaagtcc ttccagtcac
1201 gcacccacat ggtgccccct ccaaccatcg cccatggcaa atgaaggacc tacaggccat
1261 taagcaagaa gtctcccaag cagccccctg gagccccagc tttatgcaga ccatccggct
1321 tgcggtgcag cagtttgacc ccaactgcaa agacctccaa gacctcctgc agtacctttg
1381 ctctccctc gtggttccct tccatcacca gcagctagat agccttatat cagaggccga
1441 aacccgaggt attacaggtt ataaccctt agccggtccc ctccgtgtcc aagccaacaa
1501 tccacaacaa caaggattaa ggcgagaata ccagcaactc tggtcgcgcg ccttcgcgcg
1561 cctgccaggg agtgccaaag acccttctct ggctctatc ctccaaggcc tggaggagcc
1621 ttaccacgcc ttcgtagaac gcctcaacat agctcttgac aatgggctgc cagaaggcac
1681 gcccaagac cccatcttac gttccttagc ctactccaat gcaatccaaat aatgccaata
1741 attactacag ccccgaggac acactaatag cctctagga gatagtgtg gggcttgta
1801 gacctggacc cccaaagaca aaaccaaggt gttagtgtgc cagcctaaaa aacccccccc
1861 aaatcagccg tgcttcgggt gggggaaagc aggccactgg agtcgggact gcactcagcc
1921 tcgtccccc cccgggccat gccccctatg tcaagaccca actcactgga agcgagactg
1981 ccccgcccta aagccacta tcccagaacc agagccagag gaagatgccc tcctattaga
2041 cctccccgct gacatcccac acccaaaaaa ctccataggg ggggaggttt aaacctcccc
2101 cccacattac agcaagtcct tcttaaccaa gaccagcat ctattctgcc agttataccg
2161 ttagatcccg ccgctcgccc cgtaattaaa gccaggttg acaccagac cagccacca
2221 aagactatcg aagctttact agatacagga gcagacatga cagtccttcc gatagccttg
2281 ttctcaagta atactccct caaaaataca tccgtattag gggcaggggg ccaaacccaa
2341 gatcacttta agctcacctc ccttctctgt ctaatacgcc tccctttccg gacaacgcct
2401 attgttttaa catcttgctt agttgatata 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 taaccagta ttcccagtta
2701 aaaaggccaa tggaacctgg cgattcatcc acgacctgcg ggccactaac tctctaacca

```

Frei verwendbar

```

2761 tagatctctc atcatcttcc cccggggcccc ctgacttgtc cagcctgcca actacactag
2821 cccacttgca aactatagac cttaaagacg cttttttcca aatcccctta cctaaacagt
2881 tccagcccta ctttgctttc actgtcccac agcagtgtaa ctacggcccc ggcactagat
2941 acgcctggaa ggtactaccc caagggttta aaaatagtc caccctgttc gaaatgcagc
3001 tggcccatat cctgcagccc attcggcaag ctttccccc atgcactatt cttcagtaca
3061 tggatgacat tctcctggca agcccctccc atgaggacct actactactc tcagaggcca
3121 caatggcttc cctaattctc catgggttgc ctgtgtccga aaacaaaacc cagcaaaccc
3181 ctggaacaat taagtctcta gggcaaataa tttcacccaa ccacctcact tatgatgcag
3241 tccccacggg acctatacgg tcccgtctgg cgctacctga acttcaagcc ctacttggcg
3301 agattccagt ggtctccaag ggaactccca ccttacgcca gcccttccac agtctctact
3361 gtgccttaca aaggcatact gatccccgag accaaatata tttaaatcct tctcaagttc
3421 aatcattagt gcagctgcgg caggccctgt cacagaactg ccgcagtaga ctagtccaaa
3481 ccctgcccc cctaggggct attatgctga ccctcactgg caccactact gtagtgttcc
3541 agtccaagca gcagtggcca cttgtctggc tacatgcccc cctacccccac actagccagt
3601 gcccttgggg 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 caagtccggc caacggggcg
4021 aacttctcgg acttttgcct ggccctttcca gcgcccgttc gtggcgctgt ctcaacatat
4081 ttctagactc caagtatctt tatcattacc ttccggacct tgccctgggc accttccaag
4141 gcaggctctc tcaggcccc tttcaggccc tctgccccg cttactatcg cgtaaggctc
4201 tctatttgca ccacgttcgc agccatacca atctacctga tcccatctcc aggtcaacg
4261 ctctcacaga tgccctacta atcacccctg tctgcagct ctctcctgca gaactacaca
4321 gtttcaccca ttgcggacag acggccctca cattgcaagg ggcaaccaca actgaggctt
4381 ccaatatcct gcgctcttgc cactctgcc gcaaaaataa cccacaacat catagtctc
4441 ggggacacat ccgcctggc ctacttccca accacatctg gcaaggcgac attaccatt
4501 tcaaataata aaatacgtc taccgccttc atgtatgggt agacacctt tcaggagcca
4561 tctcagctac caaaagaga aaagaaacaa gtcagaagc tatttcctct ttgcttcagg
4621 ccattgcta tctaggcaag cctagctaca taaacacaga caacggccct gcctatat
4681 cccaagactt cctcaatatg tgtacctccc ttgctattcg acatactacc catgtcccct
4741 acaatccaac cagctcagga cttgtagaac gctctaattg cattcttaaa accctattat
4801 ataagtactt tactgacaaa cccgacctac ccatggataa tgctctatcc atagccctat
4861 ggacaatcaa ccacctgaat gtgttaacca actgccacaa aaccgatgg cagcttcacc
4921 actccccccg actccagccg atcccagaga cacattccct cagcaataaa caaacccatt
4981 ggtattatct caagcttctt ggtcttaata gccgccagtg gaaaggacca caggaggctc
5041 tccaagaagc tgccggcgct gctctcatcc cggtaagcgc tagttctgcc cagtggatcc
5101 cgtggagact cctcaagcga gctgcctgcc caagaccctg cggaggcccc gccgatccca
5161 aagaaaaaga ccaccaacac catgggtaag tttctcgcca ctttgatttt attcttccag
5221 ttctgcccc tcatectcgg tgattacagc cccagctgct gtactctcac aattggagtc
5281 tctcataacc actctaaacc ctgcaatcct gccagccag tttgttcgtg gacctcgac
5341 ctgccggccc tttcagcaga tcagcccta cagccccct gccctaattc agtaagttac
5401 tccagctacc atgccaccta tccctatat ctattccctc attggattaa aaagccaaac
5461 cgaaatggcg gaggtatta ttcagcctct tattcagacc cttgttccct aaagtgtcca
5521 tacctggggg gccaatcatg gacctgcccc tatacaggag ccgtctccag cccctactgg
5581 aagtttcagc aagatgtcaa ttttactcaa gaagtttcac gcctcaatat taatctccat
5641 ttttcgaaat ggggttttcc otttccctct ctagtcgacg ctccaggata tgacccccatc
5701 tggttcctta ataccgaacc cagccaactg cctcccaccg cccctcctct actccccac
5761 tctaacctag accacatcct cgagccctct ataccatgga aatcaaaact cctgaccctt
5821 gtccagttaa ccctacaaag cactaattat acttgcatcg tctgtatoga tcgtgccagc
5881 ctatccactt ggcacgtcct atactctccc aacgtctctg ttccatcctc ttcttctacc
5941 cccctccttt acccatcggt agcgttcca gccccccacc tgacgttacc atttaactgg
6001 acccactgct ttgaccccca gattcaagct atagtctcct cccctgtgca taactccctc
6061 atcctgcccc ctttttccct gtcacctgtt cccaccctag gatcccgcct ccgcgcagcg
6121 gtaccgggtg cggctctggc tgtctccgcc ctggccatgg gagccggggg ggctggcggg
6181 attaccggct ccatgtccct cgctcagga aagagcctct tacatgaggt ggacaaagat
6241 atttcccaat taactcaagc aatagtcaaa aaccacaaaa atctactcaa aattgcgcag
6301 tatgctgccc agaacagacg aggccttgat ctctgttctt 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	accatgcata	ctccgtcagc	tacgacacct	cccctcgcgc
6601	gtcagatacc	cccattactc	tcttataaac	cctgagtcac	ccctgtaa	ac caagcacata
6661	attattgcaa	ccacatcgcc	tccagcctcc	cctgcccaata	attaacctct	cccattaaat
6721	cctccttctc	ctgcagcaac	ttcctccgtt	cagcctccaa	ggactccacc	tgccttcca
6781	actgtctagt	atagccatca	atccccaa	cctgcatttt	ttctttccta	gcactatgct
6841	gttttcgctt	ctcagccctt	tgtctccact	tgcgtcacg	gcgtccctgc	tcttcctgct
6901	ttctccgggc	gacgtcagcg	gccttcttct	ccgcccgcct	cctgcgcctg	gccttctcct
6961	cttcttctct	tttcaaatac	tcagcaatct	gcttttctct	ctctttctcc	cgtctttttt
7021	ttcgttctct	cttctcctca	gcccgtcgtc	gccgatcacg	atgcttttcc	ccgcgagggtg
7081	gcgttttctc	ccctggaggg	ccccgtcgca	gccggccgcg	gcttttctct	tctaaggata
7141	gcaaaccgtc	aagcacagct	tctctctcct	ccttgtcctt	taactcttcc	tccaaggata
7201	atagcccgct	caccaattcc	tccaccagca	ggctctccgg	gcatggcaca	ggcaagcatc
7261	gaaacagccc	tacagatata	aagttaacca	tgcttattat	cagccactt	cccagggttt
7321	ggacagagcc	ttcttttctg	ataccagtc	tacgtgtttg	gagactgtgt	acaaggcgac
7381	tggtgcccc	tctctggggg	actatgttct	gcccgcctac	atcgtcacgc	cctactggcc
7441	acctgtccag	agcatcagat	cacctgggac	cccacgatg	gacgcgttat	cggctcagct
7501	ctacagtctc	ttatccctcg	actccctctc	ttccccaccc	agagaacctc	taagacctc
7561	aaggctctta	ccccgccaat	cactcataca	acccccaa	ttccacctc	cttctccag
7621	gccatgcgca	aatactcccc	cttcgaaat	ggatacatgg	aaccacctc	tgggcagcac
7681	ctcccaaccc	tgtcttttcc	agaccccgga	ctccggcccc	aaaacctgta	cacctctgg
7741	ggaggtctcg	ttgtctgcat	gtacctctac	cagctttccc	cccccatcac	ctggccccctc
7801	ctgccccacg	tgattttttg	ccacccccgg	cagctcgggg	ccttctctac	caatgttccc
7861	tacaagcgaa	tagaagaact	cctctataaa	atttccctta	ccacaggggc	cctaataatt
7921	ctaccggaag	actgtttgct	caccacctt	ttccagcctg	ttagggcacc	cgtcacgcta
7981	acagcctggc	aaaacggcct	ccttccgttc	cactcaaccc	tcaccaactcc	aggccttatt
8041	tggacattta	ccgatggcac	gcctatgatt	tccgggcccc	gccctaaaga	tggccagcca
8101	cttttagtac	tacagtcctc	ctcctttata	tttcacaaat	ttcaaaccac	ggcctaccac
8161	ccctcattcc	tactctcaca	cggcctcata	cagtactctt	cctttcataa	ttacatctc
8221	ctgtttgaag	aatacaccaa	catccccatt	tctctacttt	ttaacgaaaa	agaggcagat
8281	gacaatgacc	atgagcccca	aatatccccc	gggggcttag	agcctcccag	tgaaaaacat
8341	ttccgcgaaa	cagaagtctg	aaaaggctcag	ggcccagact	aaggctctga	cgtctcccc
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	cgcctacct	gaggccgcca	tccacgcggg	ttgagtcgcg
8701	ttctgcccgc	tcccgctgt	ggtgcctcct	gaactgcgtc	cgcctctag	gtaagttag
8761	agctcaggtc	gagaccgggc	ctttgtccgg	cgtcccttg	gagcctacct	agactcagcc
8821	ggttctccac	gctttgcctg	accctgcttg	ctcaactctg	cgtctttgtt	tcgttttctg
8881	ttctgcgcg	ctacagatcg	aaagtccgc	ccctttccct	ttcattcacg	actgactgcc
8941	ggcttgcccc	acggccaagt	accggcgact	ccgttggtc		

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

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

```

atgggccca aatcttttcc cgtagcgcta gccctattcc
gcgccgcccc cgggggctgg ccgctcatca ctggcttaac ttctccaag cggcataatcg
cc
(803-902 bp)

```

### Aufgabe 3:

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

##### 1.Gen (803-2092 bp):

MGQIFSRASPIPRPPRGLAAHHWLNFLQAAYRLEPGPSSYDFH  
QLKKFLKIALETVPWICPINYSLLASLLPKGYPGRVNEILHILIQTAQIPSRPAPPP  
PSSPTHDPDSDPQIPPPYVEPTAPQVLPVMHPHGAPPNHRPWQMKDLQAIKQEVSA  
APGSPQFMQTIRLAVQQFDPTAKDLQDLLQYLCSSLVASLHHQQLDLISEAETRGIT  
GYNPLAGPLRVQANNPQQGLRREYQQLWLAFAALPGSAKDPSWASILQGLEEYPYHA  
FVERLNIALDNGLPEGTPKDPILRSLAYSANKECQKLLQARGHTNSPLGDMLRACQT  
WTPKDKTKVLVVQPKPPPNQPCFRCGKAGHWSRDCTQPRPPGPCPLCQDPTHWKRD  
CPRLKPTIPEPEEEDALLLDLPADIPHPKNSIGGEV

##### 2.Gen (2994-5189 bp):

MQLAHILQPIRQAFQCTILQYMDDILLASPSHEDLLLLSEATM  
ASLISHGLPVSENKTQQTPTGIKFLGQIISP NHLTYDAVPTVPIRSRWALPELQALLG  
EIQWVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRL  
VQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDK  
YTLQSYGLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHSHRFKNLGAQTGELWNTF  
LKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYILWDKHTLSQRSFPLPP  
PHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQA  
LLPRLLSRKVVYLHHVRSHTNLPDPISRLNALTDALLITPVLQLSPAELHSFTHCGQT  
ALTQGATTTEASNILRSCHACRKNNPQHQMPPRGHIRRGLLPNHIWQGDITHFKYKNT  
LYRLHVWVDTFSGAISATQKRKETSSEAISSLLQAIAYLGKPSYINTDNGPAYISQDF  
LNMCTSLAIRHTTHVPYNPTSSGLVERSNGILKTLLYKYFTDKPDLPMDNALSIALWT  
INHNLVLTNCHKTRWQLHHSRPLQPIPETHSLSNKQTHWYYFKLPGLNSRQWKGPQEA  
LQEAAGAALIPVSASSAQWIPWRLKRAACPRPVGGPADPKEKDHQHHG

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

MPKTRRRPRRSQRKRPPTPWPTSQGLDRAFFSDTQSTCLETVYK  
ATGAPSLGDYVRPAYIVTPYWPPVQSIRSPGTSPMDALSAQLYSSLSLDSPPSPPREP  
LRPSRSLPRQSLIQPTTFHPPSSRPCANTPPSEMDTWNPPLGSTSQPCLFQTPDSGPK  
TCTPSGEAPLSACTSTSFPPSPGPGSCPT

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

MAHFPGFGQSLFLGYPVYVFGDCVQGDWCPISSGGLCSARLHRHA  
LLATCPEHQITWDPIDGRVIGSALQFLIPRLPSFPTQRTSKTLKVLTPPITHHTPNIP  
PSFLQAMRKYSFPRNGYMEPTLGQHLPTLSFPDPGLRPQONLYTLWGGSVVCMYLYQLS  
PPITWPLLPHVIFCHPGQLGAFLTNPYKRIEELLYKISLTTGALIILPEDCLPTTLF  
QPVRAPVTLTAWQNGLLPFHSTLTTPLGIWTFDGTMPISGPCPKDGQPSLVLQSSSF  
IFHKFQTKAYHPSFLLSHGLIQYSSFHNLHLLFEEYTNIPISLLFNEKEADDNDHEPQ  
ISPGGLEPPSEKHFRETEV

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

MALCCFAFSAPCLHLRSRRSCSSCFLRATSAAFFSARLLRRAFS  
SSFLFKYSAICFSSSFSSFRFLFSSARRCRSRCVSPRGGAFFSPGGPRRSRPRLSSS  
KDSKPSSTASSSSLSFNSSSKDNSTNSSTSRSSGHGTGKHRNSPTDTKLTMIIISP  
LPRVWTEPSFRIPSLRVWRLCTRRLVPHLWGTMTFGPPTSSRPTGHLSRASDHLGPHRW  
TRYRLSSTVPYPSTPLLPHPENL

6.Gen (5182 -6648 bp):

MKGFLATLILFFQFCPLILGDYSPSCCTLTIGVSSYHSKPCNPA  
QPVCSTWLDLPALSADQALQPPCPNLVSYSSYHATYSLYLFPHWIKKPNRNNGGGYSA  
SYSDPCSLKCPYLGCSWTCPYTGAUSSPYWKFQQDVNFTQEVSRNLNINLHFSKCGFP  
FSLLDVAPGYDPIWFLNTEPSQLPPTAPPLPHSNLDHILEPSIPWKSLLTLVQLTL  
QSTNYTCIVCIDRASLSTWHVLYSPNVSVSSSSTPLLYPSLALPAPHLTLFPNWTHC  
FDPQIQAIUSSPCHNSLILPPFSLSPVPTLGSRSRRAVPVAVWLVSALAMGAGVAGGI  
TGSMASLASGKSLLEVDKDISQLTQAIKVNHNLLKIAQYAAQNRRLDLLFWEQGGL  
CKALQECCFLNITNSHVSILQERPPLENRVLTGWGLNWDLGLSQWAREALQTGITLV  
ALLLLVILAGPCILRQLRHLPSRVRYPHYSLINPESSL

7.Gen (6836-7135 bp):

MLFRLLSPLSPLALTALLLFLSPGDVSGLLLRPPPAPCLLLFL  
PFQILSNLLFLLFLPLFFSLPLLLSPSLPITMRFPARWRFLPWRAPSQPAAAFLE

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

MGQIFSRASPIPRPPRGLAAHHWLNFLQA

- (a) In der Genomsequenz liegen Basentriplets vor, die für bestimmte Aminosäuren codieren. Jedoch ist es möglich, dass mehrere unterschiedliche Basentriplets 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.
- (b) 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. In der Summe erhalten wir also 6 Leseraster.

Bsp.: Basensequenz ACGTATG

Leserahmen 1: ACG

Leserahmen 2: CGT

Leserahmen 3: GTA

## Aufgabe 4:

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



Abweichungen von eingegebener Suchsequenz in      markiert:

Eingegebene Suchsequenz:

MGQIFSRASPIPRPPRLAAHHWLNFLQA

Erhaltenes HMM Logo:

MGKIYSRSISPIPKAPKGLAIHHWLNFLQA (AYRLFPGPSEYDFHQLKKFLKLALKTPVWLNPINYSLLASLLPK  
NYPGRVNEILAILIQETA)

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

- (a) Die eingegebene Sequenz weicht von der erhaltenen Sequenz um 23,3 % ab. Die Abweichung beträgt ungefähr  $\frac{1}{4}$  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:

Ausgewähltes Genom: *Blueberry fruit drop associated virus*

Wiederholung Aufgabe 2:

Komplettes Genom:

([https://www.ncbi.nlm.nih.gov/nuccore/NC\\_028462.1](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	tggacattht	gtagatctcc	cacagcatca	cagatcgctt	ggaatagaca	aaattgccct
121	aatcacctat	ccataactag	tgtgcatcgt	agtcacatth	cgtcaggctt	agaagccaaa
181	cctagtgtct	tagctgaagt	cctcaattat	cagtcagatc	agttatttga	cataaaagga
241	aaatctcaaa	aaatagcaga	agtcgtagac	tggactgcag	agtttttaga	aaatcaaaaa
301	caatcagatc	tgccatacaa	caagcttgac	ttcgtcaaga	cttccatcga	cttgtttagca
361	acagaatttg	caggggttcaa	agaagccaga	acagagttca	ttgaacaaat	tatcaggcgt
421	gttcttgacg	ctgataaaaa	gaatcttacc	caactcaccg	gggtaattga	agaacaagtc
481	tctcaactca	aaggagtaat	tgaagaactt	aggtttgacc	cagataaacc	taagccagct
541	gaaatacgaa	cacaaacact	tgcagaagaa	atagcagctt	cggcgggtgc	tatccagaag
601	gtccttagca	aactagagga	gctagtggct	aaggttgact	ctacgggtatc	agaagttttc
661	agagtcacca	agagaatcga	ggctaactca	acatgagtgg	cttcaaaacta	tcttcaccat
721	ccgacgaaga	tttgaaccgt	tggctaactg	atgagttaga	tcggggttcaa	aataaaaatta
781	ttggaaaggt	caaagatggt	gaccagatc	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	tattgtctct	gccacaagaa	acattgctcg	catgttgag	caagtcaaa
1081	gaattgaaga	agtcacatca	actgtctttc	aaaccattga	aggaattctc	aatgtcaaca



```

1141 aacagttttat gacagcagag aaacaaaggg agcagatcct gcaaagtcaa cttaagcaaa
1201 ttaaagacca attggcagaa ccagtttgtt taagcaaacc tgtcagtgat ccagctcctg
1261 aaggtcacct caaaagcatc aacagaggag taaaaaacct cgttgaacgt gcagaactcc
1321 taatcgacta tgaagaactt gctgcaaaaa ttggcaagaa tttttctact caaggaattg
1381 cagtcagtga aaggagcaaa tcatttatga gcttcgatca cccaaagcca gtagaatcag
1441 ttctacccaa ggcaagaatg ttcagtatgt ttgcaggagg caaacccaaa agcagcattg
1501 tcagatccct ggagtttaga gttgaagaac aacctaaagt cctcaaagac tataaggttc
1561 caaaattatc tgtcaagcag ctctatgaac aaggaagcat gatgcagcgc aacgcttata
1621 gcatcaagca gatcgagaga aagattccct acgatcccaa cactaaagtt atccatcttg
1681 tggataaagg agaaatggat cgcattgaagg caaagggata cgagtatgtg cacttcggag
1741 cagtccaagt tgttttcaga ctattggcca gaagagatct caactgctca gttcttagca
1801 tttgcagaga caaccgttta aaatctcttc aaaagggatt actcggggct ttgcaagcaa
1861 gcctcaacaa tcaggttgca ttttttgtat gcatccccaa tttctcgtgc acaattgaag
1921 atgcccatag agctttgggt ttatgtgttc agacacatgg aacagatttc aaagaaggct
1981 aactgatctt tgccattgaa tacattgtca catacaaatt tatgacaagt gatatggaag
2041 ccattgtcaa 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 caaacgcata ctgttcacat cttcgatgat gaaccattc
2341 aactggatgt atctaaatct ggcagacctt aggccagaat gatcagagct agcacaagaa
2401 aggaagttgt ggtggaatta catgacatca acagcatcag agaaatctcc gacgaaaaac
2461 tcgacgagat ttatgtctac acttcagatc cacagtactt caactataca gtactgccag
2521 agatgcaaag aagaatcaac caagggatat cggtcagagc tatcagtcaa tcggtcaaca
2581 gaccagtagc 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 agtgtgattg
2881 tctcagaaac cggaaaagca tggagcccag atgaagcaag aacctactat gcctgtacat
2941 tcaaaggcat agccagaaaa tggttcgaca aatacaagga gacagatgaa tacgtcttat
3001 ggttgagcac catagaaagg tcaaagtcac cctctgactt tgccacatgt ctctacaagc
3061 agttctgcag agaacaagat ccagcaaga aggagaagga tcttgagatt ctaaaaggaa
3121 aacttgtgaa tctcaccatt tcagatctaa agtacttcct cagttatgtc aacgagtaca
3181 acttggcata ctacaacctt ggcataattc 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 aatgggggtg aagaaacaaa gagctcaagc
3481 ataagaagaa aggaaagaag atcaaactct 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 ttcaagtata tattcacttg
3841 aaacagaag cgattcagag tcactgtatt ctgatcaaga aagtgaagac gaagacactg
3901 tcaggatgat caagacccaa tttggtatca agacaatacg cttgaacaat tggccacatg
3961 acaccagga gattggggac gtcgtgtctt ttgaagctat actggagtat gaaaagttcg
4021 agtgtgattt ctgcaaaaaca totgcagacc aatacacacc agtcatacaa cgaagcgatc
4081 ttagcaaaga aggtaccatc agctatcaca catggtgttt tacttctctt gtcagagag
4141 aaaccaagaa cgctgaggtc ccacaactgg tacaagaga agttgatatt gcaaaaactc
4201 gagatgccaa aaagaagttt gaacaagctc agctcaaaga agaacaagaa agacttagag
4261 caacaagaag acctgtcata cttgaggtct atgatgatgc agaggagctt gaagaaatgt
4321 caaaaatttg aggccatttg gaagaactat ccatcaaact aagaaggcct gttgctgcta
4381 gtccaggtgc tacaagcagt aatccacttg aagaaaacgt ggatactccc aaagacctaa
4441 aagggtcaaat acttctctgc cccatattca aagggtccaa ccctaagaga actgaagctg
4501 attggaaaag agaagaagaa gagagaaaca gaaagcttgt tgaaagctac agaaaaattg
4561 agctgtttga aagatataag gaagtttaga atgaagtcct acagcaaat ccacagcatc
4621 aaattcccat caacacaatc atcaccgtaa agagttctcc gacggagatg cttgctgata
4681 ttgaaaaact gatggaagaa gtcaagaagc ttcaacagaa ggaaagcacg ccagaagata
4741 agccagatgc aaccctgaa ttcaggcata aagacaacag catctacatc tctgccgctt
4801 tcagattcaa aggatatgaa gaatatcatc tccactgtct agtcgacacc ggagccagca
4861 tcacacttgc taccaatgg gccattccag aagaaaaatg gatcactatt cccacaggaa

```

```

4921 atgaaatcaa aatcacagtg gcaaattggag aaacttgttaa aatcgacaag tatgccacca
4981 atgtgccccat atacattttca ggtatcaagt ttactattga aaaagtattg caattttaata
5041 agcaagggttg cgattttcctc atcggaaatg atttctttat gaaatacgc ccatatacgc
5101 aatacccaga tcgtattttca atagccaaaa acggaatttc gtatatcaca aacgtggtag
5161 aacgaccagt ttccgtagca tacagaaaaa aatttctgga aagacgaaaa attccagacc
5221 atcagcagtc acacaagggtg aagtgtggtg aaaattacaa agtccaaaaa caagcccaag
5281 aaaatgactt cactattgaa gactatggta ctacggatag cgaagctgaa gagcccatat
5341 atgatgccat atatgacaaa ggatttcagc ccatcgaatc tgacgatgag cccaatgacg
5401 aaatccagcc catcaatgat gagcccaatg aagaagtcag gcccgattc cagcccatcg
5461 aatctaacga tgagcccaat aatgaaatcc agcccagcaa tgacgagccc aataaagagg
5521 tcaggcccaa gccaacaaaa gatatcaggc ccaagcccat tactgaagaa gacttccaag
5581 cagagcttcg cctcgtcgaa gccactatg aaaagcaggc taggcccaaa caacggccca
5641 acctaaatgt gtcgaagaa caaatcagca gatgcagcat tgaagagctc accaaacagc
5701 ttgagcccaa catcagtgac aatccaatcc acaagtggga tgtcacccaa acctatgctg
5761 acatacaact caaggatccg tatgcaacta tacgtgtaaa gcccatgcca tacagtgctg
5821 aagatgaagc tgagtttgaa agacaactcg aagagcaact caaactggaa ctcatcagc
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 agcatccac
6121 tcaactgcatt cagcaccccc aagggtcagt atgaatggaa ggtccttcca tttggattaa
6181 agaatgcacc aaacatatcc cagcgaagga tggacaaaat cttcaaagat tgttatcagt
6241 tctgtggtgt atacgtcgac gatatactcg tcttctccaa agacaaaagaa gaacacctac
6301 tgcattctgga tatcatcatc aacaaaattg tccagaatgg aatcatcatc ggaagaacca
6361 agtacgaatt ggtcagagaa agcattgact ttttaggagt caccatatgc aaaggaaaga
6421 tcaagttgca gcctcacatc cttactcacc tcatggagtt tctgacaaa ctcaaggatc
6481 gacttcagggt ccagcagttt ctcggatgtg tcaactacgc cggaaagtcc atacaaaggc
6541 ttgcagagtt atgcagtgtg ttgcatacca aactcaagaa agatcaaaag tggaaagtca
6601 cagaggacga tgtcaaaaga gtccaactca ttaagaaagt ctgtgccaat cttccagagc
6661 ttaccttgcc cccaccaaca tgggaagatcg ttttgcaaac tgatgcaagt gatcaccact
6721 gggcaggaac aattggagct atcagccctg acaagaagga agaagtaatt tgtgcttaca
6781 gaagcggaa cgttcaaact gctgagcaaa attactccac ccaagaaaaa gaagtcttag
6841 ccatcatccg tactattcaa aaggccaaga tcttgcctcct aaaacccttc cttgtccgaa
6901 ccgattcaaa gtttgcagcc ctattccttg ataaaaagat caacgagtct ttagctaggg
6961 gtcgtctcat tagatggcaa atcatgttta ggcagtacta cttagacatt gaacacattc
7021 ctggtgtttc taattatttg cctaattgcc ttactcgtga aatggcagggt tgcaatatca
7081 tttaccagat caaaatggta aaccatggaa aagccccaag gatggatctt agagatgtca
7141 tcaacttcaa ggctagagct agagaaaatc agcagctcga aaatgaagac atccattgga
7201 aacagcatct tgcttcaaca tcaagaagct ttgctaaaag aaaagatggt 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 caaccaaga
7501 agttcacctt ctttgaagaa tggcgtacaa tctactacg gaggctaocg aactggacca
7561 accgagctac cataatggac gaagaattgt atcaacggat catcaacact ccttgcagag
7621 cccaattgca atggtggtcc caactcaaca agttatggca gtatcagccg cctgagcaaa
7681 gtttcatctg caaagctact tgctttgaca acaatcatac catgctcaaa ctcaaatgtc
7741 ctcccatcaa caaactcacc cctgagttca tggactgggt tgaaaatgga ttcatcacia
7801 gctttgccct taagagaacc agtgattaca gtttattacc tactgatttt atcaaggtaa
7861 ttgaacagat ttgggctatt aaagatattg gtaatacatt gtttattatg atgtattcag
7921 ccccaggaga atggaactac aaagattggt cctattatta tccttatcat ctggtgagga
7981 tttatgatag agttccatct gtcactagag atacaatcct tgaaggagaa atcttcccc
8041 atctcagaag ccaggagcaa attcaaagag atgaagaaat gggatgttac gaaccactgt
8101 actgtcccat cgacgatttg 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 tcaaagtga gacccatgtg
8401 acgagccata tgatgaatca gattcgtact ggcaggacat ggatgatgag acgtggcaca
8461 atattgaaaa catgatggaa ggataaggcc tggcagactc caccatccca tgtgctactt
8521 ttgctttcgg tgtgctttat tgctttcggg ttgtcgggtt 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 ttcaagatct ctcttgtaag ttctttttat gaataaaagt ttgatgtttt
8821 attctaacat ggagtagtcc ttagggcgcc tctctggagg aggcctctgt tatgcatggt
8881 tgtggttatg ctttaactat atgaatgttg tattagagta tgatataact atggagggttc
8941 gaagcactag agtctgttaa cataactcta tgataaacca gaagagaggt tatggtttca
9001 ttacttcact catgtctact gcaaaatctt ggaaaacaaa tgttctttgt caccagacct
9061 tatcaaatat acttttttgg ctaatgggaa aaatggaggt ttctgttttg aagaacgaag
9121 gaggaatggc agtgagggtg aggcctagat cccggtgttt gccaaagagg cggtgtgttg
9181 ggtgtggttg aagccgaagg caagtttcct ctggtgtgtt gtattcttaa aacaaagttt
9241 ggcccttcta gaggaaaagg ttgataaac gaatgggtatt ttggtaacaa aagattttta
9301 gtatctgaaa tattgcagat gtatgagcgc tcttgatctt ttggccaca agtcttggtt
9361 tgggttatct atgatgagtt tgaaactacg ttgcttcaga gtccttggtt atcagagctt
9421 tttataacgg aagttgggtg gccataccc acagggtctg cataataagc cgtccgaaag
9481 agaataacca ttacattat 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 atgagttggt atattgctta gaaagtcata atttctttat ccgcattcaa gcactaccag
9781 atcgtgttta ctttaggttt tatagggtctg tttgcacttg ttttgcatta taacatcatc
9841 ttgtttttat
```

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

### First 100 bp of genome sequence:

```
atgagtggtg cttcaaaacta tcttcaccat
ccgacgaaga tttgaaccgt tggctaactg atgagttaga tggggttcaa aataaaatta
ttggaaaggt ca
(693-792 bp)
```

### Wiederholung Aufgabe 3:

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

#### 1.Gen (693-8486 bp):

```
MSGFKLSSPSDEDLNRWLTDELDRVQNKIIIGKVKDVPDPHVAMA
EPGENHLRNGKVLIIHQMNRMFATINTMFEKQLWFEQRQKVLQNLWDFLNSYDENQKA
QQKALKDSQDNIALATRNIAARMLEQVKGIEEVINTVFQTIEGILNVNKKQFMTAEKQRE
QILQSQLKQIKDQLAEPVVLSPVSDPALEGHLKSINRGVKNLVERAELLIDYEELAA
KIGKNISTQGIASERSKSFMSFDHPKPVESVLPKARMSMFAGGKPKSSIVRSLELE
VEEQPKVLKDYKVPKLSVKQLYEQGSMMQRNAYRIKQIERKIPYDPNTKVIHLVDKGE
MDRMKAKGYEYVHFGAVQVVFRLRLARRDLNCSVLSICRDNRLKSLQKGLLGALQASLN
NQVAYFVCIPNFSCTIEDAHEALVLCVQTHGTDKFGYTDLAIEYIVTYKFMTSDMEA
MSKLGKATEGFNSMVISDRNNHTISHTKKIDWKELNQSEFPKEWSINSEEPPEAKSSSV
ITGIYQHENGNI SMRFHQAGEGSSHKAQTHPVHIFDDEPIQLDVSKSGRPKARMIRA
STRKEVVVELHDINSIREISDEKLDEIYATTSDPQYFNYTVLPEMQRRINQGISVRAI
```

SQSVNRPSTSGDSKEVPITSEVPSSSTLEGKGKTKKFTPAGSSIKLPLITGVPTGKPN  
EFNTERYKPRIEPNPAVSPNGVWLDIDNNPDPRGNIDEWVNTMNMLIVSETGKAWSPD  
EARTYYACTFKGIARKWFDKYKETDEYVLWLSTIERSKSPSDFATCLYKQFCREQDPS  
KKEKDLEILKGKLVNLTISDLKYFLSYVNEYNLAYYNLGIFNNNTYRDMFIQKLPDPW  
KEYLTNERAKEAGNEEWTLGNLIEFVKTKLADLAIQAWRAKRMQKLYSKSGLSFCQHL  
IDIPTQWGGRNKELKHKKKGKKIKSFKRYKFRKSKKKGFEEKSKINPQKNFKRRFKQKS  
GKEIEDRKKQSKSKPTSSKGVANVCWNCNQGHFSRECPRK NARLIELQEAA DTWEIP  
LMRIEEDDDYSDTSSIIYSLETESDSESSDSQSEDEDTVRMIKTQFGIKTIRLNNWP  
HDTQEIGDVVSFEAILEYEKFECD FCKTSADQYTPVIQRS DLSKEGTISYHTWCFTSL  
VMRETKNAEVPQLVQREVDLAKLRDAKKKFEQAQLKEEQERLRATR RPVILEVYDDAE  
ELEEMSKIGGHLEELSIKLRRPVAASPGATSSNPLEENVDTPKDLKGQILLCP IFKGP  
IPKRTEADWKREEEERNRKLVESYRKLKLFERYKEVRNEVLQQIPQH QIPINTIITVK  
SSPTEMLADIEKLMEEVKKLQQKESTPEDKPDATPEFRHKDNSIYISAAFRFKGYEEY  
HLHCLVDTGASITLATQWAIPEEKWITIPHGNEIKITVANGETCKIDKYATNVPIYIS  
GIKFTIEKVLQFNKQGCD FLIGNDFFMKYAPYTQYPDRISIAKNGISYITNVVERPVS  
VAYRKEFLERRKIPDHQQSHKVKCGENYKVQKQAQENDFTIEDYGT TDSEAE EPIYDA  
IYDKGFQPIESDDEPNDEIQPINDEPNEEVRPGFQPIESNDEPNNEIQPSNDEPNKEV  
RPKPTKDIRPKPITEEDFQAELRLVEAHYEKQARPKQRPNLNVLEEQISRCSIEELTK  
QLEPNISDNPIHKWDVTQTYADIQLKDPYALIRVKPMPYSAEDEAEFERQLEEQLKLE  
LIQPSKSPHSSPAFCVRNHA EIKRQKARIVINYKELNRVTKDDGYFLPNLETLVYQVA  
EARVFSKFDCKSGYWQIKLTPESIPLTAFSTPKGQYEWKVL PFGLKNAPNIFQRRMDK  
IFKDCYQFCGVYVDDILVFSKDKEEHLHLLDIIINKIVQNGIIIGRTKYELVRESIDF  
LGVTICKGKIKLQPHILTHLMEFPDKLKDRLQVQQFLGCVNYAGKFIQRLAELCSVLH  
TKLKKDHKWKFTEDDVKRVQLIKKVCANLPELTLPPTWKIVLQTDASDHHWAGTIGA  
ISPDKKEEVICAYRSGTFKPAEQNYSTQEKEVLAIIRTIQKAKIFVLKPFLVRTDSKF  
AALFLDKKINESLARGRLIRWQIMLRQYYLDIEHIPGVSNYLPNAL TREMAGCNIIYQ  
IKMVNHGKAPRMDLRDVITSKARARENQQLENE DIHWKQH LASTSRSFAKRKDVPMP S  
PIPPPF EYKTYRRLTEEQKSQLDWLTSSWHTVDQDTFFDVLKAMAKEFEGLNKLKNGN  
PAKRVKTSREAKLARWKAQPKKFTFFEEWRTILLRRLRNWTNGATIMDEELYQRIINT  
PCRAQLQWWSQLNKLWQYQPPEQS FICKATCFDNNHTMLKLKCPPINKLTPEFMDWFE  
NGFIQS FALKRTSDYSL LPTDFIKVIEQIWA IKDIGNTLFIMMYSAPGEWNYKDCSYY

Amelie Faber

YPYHLVRIYDRVPSVTRDTILEGEIFPHLRSQEQIQRDEEMGMYEPLYCPIDDWLSRD  
PRHVSKWRAFALMEIQDLINEDKGFLPRWHYIHSNQRFIFEVEGISMTRACESYSLFQ  
SLSLKVPMNYDTQLEIDCIQTLDAAQSEDPCEPYDESDSYWQDMDETWHNIENMME  
G

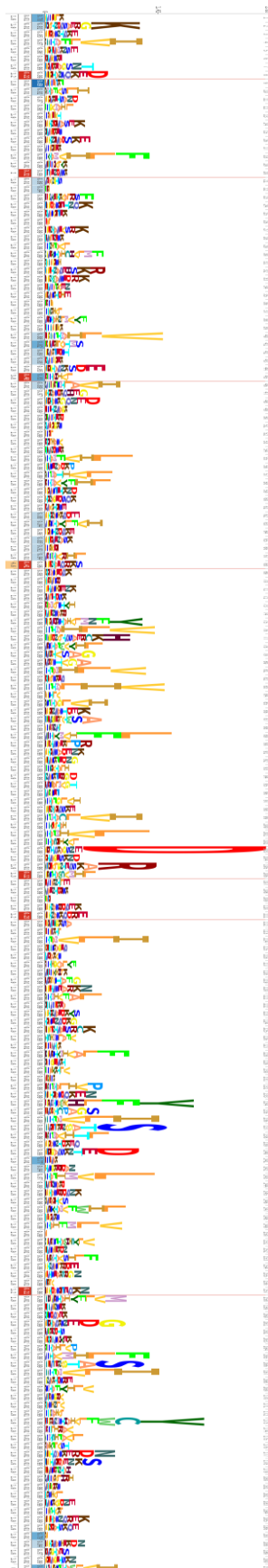
**Die erste AS-Sequenz (79 bp lang) des 1. 5'3'-Frames, die ein HMM-Profil ergibt, liegt im Bereich 303-381 bp des 1. Gens und lautet:**

MMQRNAYRIKQIERKIPYDPNTKVIHLVDKGEMDRMKAKGYEYVHFGAVQVVFRLRLARRDLNCSVLSICRDNRLK  
SLQK  
(303-381 bp)

Amelie Faber

## Wiederholung Aufgabe 4:

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



Frei verwendbar

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

Eingegebene Suchsequenz:

MMQRNAYRIKQIERKIPYDPNTKVIHLVDKGEMDRMKAKGYEYVHFHAVQVVFRLARLDNCSVLSICRDNRLK  
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).