

# Virus Report

19-01-2015 10:04:53

## Secondary structure results

Virus name:  
HI-V

Sequence:

```
TGGTATCAGAGCAAGGTTAAACATGGTCATGTCCGGCTAAAAACCTAGTGTTAGGTTTCGG
GGAAAGAGGTCTACTGTGTGTTTCAAATATCCATTTGTTATTATCCAAACAGTGTCTGTTT
CTTGTGTTTAGGATGGAAACCTGAGGTTGCGAACTACGTGTTCAAAGGAGTGATCATTTT
AAAAGGAATGTCATTCTAACTATCGTATGCTCTAGACTTGCCCTGTTATAGCAACAGAGGA
TCCTGTCTATCAGAAAATAGTTTGAACAATTCCGATGATCTCTCATACTTTAGGTATAACAA
CGCCACGTATAAAGGAGCTAATTCTTTTGTAAGACGTAACCCAGAAGAGAGTACCCAGA
GAGGGGAAGACTTGGGGAGAAGGTGAACAGAAGAGAGAGCTTACTGTACCTGGTACTGT
TATAGGAAGATTTTGTATGTCCAGTCGCTGGGAGAATAGTATACAAGAGTGGTATGAGA
AGTCACACACAGCAAACCTTGAGTACCTTGACCTAGCATCTACTAGTAAAGTGACCAACA
ACCAGTTAGCACATAACCTTGACGTAACCTTTGATAGAGTAAATTTAGGTAACCGAGTTTT
TATTA AAAACCTGAAGCAGATTCAAGAGTCTATTCTTGAATTAACACCCCGTGTTGATACTA
TAGAAGTAGCTCTAAGGAGGTTAACCAAGACCTTCAGAGAAAGCAAACCACTTACCGAAA
GTGAAGTCCGTAAGTTGGTTGAAGAAATTGCTCAACAACCTAAGATCGTCGAGAAACAAG
CACTCGAGATCTCTCAACAACCTTGAGCAGAAACTAGAGAAGGTGGAAAAGCTCTTGACACA
AACTTGACCAGTGGGTTGGTCAATGACTGATAGCCCAGCCTATCAAGAAGCTTTAAAGGA
AGCAGAAAAGGTTGACCCACCAGCAGTTGGGATAACAACCTCCACCGGAGTAACAGCGG
TTCAAGGGTTTTAAACCGTTATTAAGCAGAACAACGTTGAGATCTGCTTACTCGCTGTCAT
AGCAGATAAACTAGAAGAGCTGATACAGGATCAGAAGAAGGCCAGGAAGGATAAGGCAA
AGGAAGTTGCCATTCCAGAAGACTTAATCACAAAACCTCCAAGGATTATCTATTCAGGAGAA
AGGAGAGGCCAAAAGTCACAAGAAAACCCGAGCCAAAAGGAACGCTGTTTGGATTCAAAG
ATCCTTACAAGATCCTAGCAGCAGAAAAGGCTAAGATCACACTTAAGCCTGTGAAAGAGT
AGAAAGATGAGTCGAGCACGACCGCAACACCCAGTTCCTAGTGTGACTACAACCACTAGT
GAGCAAAAACAGGGAAGGACCTCTATACGAGGATCAGATCAGAGACTACCGAAGAGGACA
AAGAAGGATCTTTAACCTTCGAAGAAGAGCCAGAAGGTTAAGAAGATCAATGATGGGGTC
TAGATATCAAGAGACCCTAGAACAAGAAATAGATCCACAGACAACACTGAGGTTGTCCAT
GCAAGAACGAGCAAGACTAGTACCAGCTGAGGTACTGTACAGATCCAGGCGAGACACTG
TTCATCACAGAGTCTACACACACCGATCTGAAGAGTCAGTCCTATGTGTTGGCGGAAGCC
AAGTGGATAGGGCGTTTTATACAGCCTGAAAGTCTGGAACAACCTCAAAGGACAGGAATGT
CCTTTATTCATATCGGAATTCTGCAAGTGAGGATTCAGATTCTGCACCGACAAGAAGAAG
GCACCATGGCCTTGGTAGTATTCAGGGATAACAGATGGTTCGGGAGATCAGTCTATTTTCG
CTCAAATGGAAATAGACCTAACAAAAGGAAGCCAATTGGTGTTTCGTTATACCAGACACCAT
GATGACCATCGGAGACTTTGCCCGGAACGTACAACCTATCCATCCTCACAAGAGGGGTATGA
AAATTGGCAGAACGGGAGAAGCCAATCTCCTGATAACACGTGGCATGACAGGGGCGACTGT
CCAACACTCCCAATGTCGCCTTTGCTTACCAAATTGCCAGCGCAACAGATTACTTGGCAA
GTCACGGTGTA AAAAGCTATCGCAGGAAAGAAAATGAATCTGCAACACCTGCGAAACCAAC
AGTGGATACTACGTCCACCACAAACGGACATCACTCCAATGCAACCAAGATCGGTTGAAA
CAAGAAATCTCGTAGATGGAAGTATCTCCATCAGATTCCATGATTATGAGGCAGCTACTTC
AGCTTCAAGACCTCACTACAATGAGGAAGATGAAGAAGTGGAATCAGAAACGGAGTCAGA
AATAAGGGGAGCATACTATTGCAGTCTGGATAGGAGAAGAAGAAATTCAGACCAGACAGG
GAGAAAGAAGGTATGGGAAGAATCTAGTAATGGAAATGGAAGATTCTTCCGGTATTACAC
TCCTCCACCAACATTTGAAGGGGCAAATCATTGCTACAGGATGGGGAAAGTGATGATGATAA
TGAAAAAACTCCTCCAAAATGGGATGAAAGCCCAGATGAAGAAGGACCCACAGAACCCAT
ATGGGATCAAGAAGAAGAAGAAGATGAATATGATCCCAATGTCTATAGGGGCATACTTACA
AAAGGAGGAAGATGAGTGGCAAGAAATCACAGCTAGTCTCAGGGAAGAAATGGAGTACC
CAAAAAGACGACCACAAACAGAGATGGCGTTCTCTGAAACAGTCGACTATACTCCACCTG
GTGACACTATGATGACACCTGTCGGATACCCCCCGGCCTCGTCATCAAGATCAACAGTCA
CAACGCCAAGTAGACCGCCACTTTTTGAAGGAAGGACCACACACGTGCCACGATTCTTAA
AAAGGGATGAGTACACAGAATGGTGGCAATTACCATCATCGCAAGGCACAACAGGGGGCG
TTATTTGTAATGCCTAAACAAATGGGCCTATTCCATGAAGTCTTCTCGAGATGGGAATCCA
TCACCAAAAACCTACGTTGCGGCCCAAGGTTTTACGGACCCAACAGAAAAGATGGAGTTCA
TGGA AAATCTGTTAGGAGAAACAGAAAAGCTAACCTGGATCCAATGGAGAATGAATTACG
AGGCTGAATACCAGCAGCTGCTAACCCAAGCAGATGGGCGACAAGGAACCCAAAATATC
TTGTCCCAGATCAAGAGAATCTTTTCTTTAGAAGACCCAGCCTCTGGATCCACGAGAATAC
AAGATGCTGCATACAGAGACCTGGAGAGATTAACCTGCCACAACATAAAAGACATAGTTC
```

AATTTCTAAATGACTATGGAAGGCTTGCAGCAAAGAGTGGGCGACTGTTCTTAGGAACTG  
AGCTCAGTGAAAAGTTATGGATGAAGATGCCACCAGAAGTGGGCATCGAATGAAAGAAG  
CATTTCAAAAGGAGTATTCAGGCAATGAAGTAGGAGTCTTCCCGCGCATCTTGTTTGCCT  
ACAGATATCTGGAGCAAGAGTGTAAAGATGCAGCTTTTAAGAGAAGCCTGAAGTCATTAA  
GCTTCTGCAAAGACATGCCATTAACAGGATACTATGATAAAACCTCCAAGTATGGCATGA  
GAAAGTCAAGAACCTATAAAGGAAAACCATGCATCACATGCAAGGGTGGAGAAGAGG  
AAGCACTTGATCAGAAATAAAAAGTGCAAATGTTATCTCTGCGGGGATGAAGGACACTTT  
GCCCCGGAATGCCCCAATCAGAGGAGAGATGTCAAGAGAGTAGCCATTTTTGAAGGAAT  
TGATCTTCCTGAAGGCTTTGATATCGTCTCAGTAGAAGAAGGAGAAGAAGAGTCAGATGC  
TATTTATAGCATATCTGAAAATGAAGACGGAGAAGTGGACACTGAAGTAGTCCATGAGAAA  
GTCTTCATGATGAGAGAAGAAGACCAGTCCTATTGGTTAGGAAAAGACAAACCATTGGACG  
GCTATGGTGCGAGTCAGCAGTCAGCAATATCACTGCATGCACCAATGGGAGCACAAATAAA  
GAAATATTAGTGGTGGCCACATCAACTGCCACTTCTGTAAGCAACCCACTCAGTTAAGG  
AGTCGAATACACTGTCCACAGTGTCAACTCACCAGTTGCTTCATGTGTGCCCCAATATACT  
GCAATATGATAGTCCAGCAGCAGCCTAAACCACCAGTACCGTTTAATACTCACACACTGC  
TCCAACAACAAGCGGCTTATATCCAGTGGTTGGAAAAAGAAAATCAGCGGTAACTGAAG  
CCGTTGAATTCTATAAAAAGGAGGCTGAAGAATTAAGGCTCGAAAGAGACTTAGAGCAAG  
ATAGAAGGAGTCTGGAACCTACGTTGTTAGACAAAGGAAAGAAGGTTTCAGATTCTTGATC  
CAGATGAAGATCAGCACACAGCGTATCTTGAAGAAGATACCATCAGCCGTGTTATCGGCC  
ATACTGTGGAACAACAAGAGGTTAGAAAGCCCGTTAAAAAGGGAAACATGCTCTATAACC  
TCGATGTGGTGTACATATCCAGAGGTTAGGAAGACCTATCAAAGTCAAAGCAATTCTAG  
ACACTGGAGCAACCACATGCTGTATAAACATCAACTCTGTACCACAAACAGCAATTGAACA  
GAACACTTTTTCTGGTACAATTCCGAGGCATAAATTCCACGCAATCTGTGGATAAGAAACTC  
AAATATGGGAGGATGACTATCAGCAATCACCAGTTCAGAATCCCGTACTGTTATGCCTTTC  
CTCTATCCCTTGGTGATGGAATAGAGATGATCCTAGGGTGTAATTTTCATCCGTGGGATGT  
ATGGCGGTTTGCCTATTGAAGGTACACAAATCACCTTCTACAAAAATGTCACTACAATCCA  
AACCCGCTTGTCTGCCGTAATGGTTGGTGGTACAACCGCTTCTGAGTTAGGGGGAGGGG  
AGGAGTCCAAATCCGATTCTGAATCCATGTTTGACCTCTCAGAAACAGAAGAATTTGACTC  
AGAAACCCACCAGCAGATTGTGAGTCATGTTGCAGCCCAAGCCCAACAACAAAAATTGGA  
TCCAAAACCTCCAACAATAATGGTCCAACTTCAGGATCAGGGCTTTATTGGGGAAAATCCT  
ATGCAACATTGGGCTAAAAACAAGATCCTATGTGCTAGATATCAAGAATCCTGATCTTA  
TAATAGAAGACAAGCCCATCAAGCATCTAACACCGGCCATGGAGAAACAGTTCCAGAAGC  
ACATCAAAGCACTCCTGGACATTGGTGTATCAGGCCTAGTAAGTCAAAACACAGGACTA  
CGGCCTTCATTGTGAATCAGGCACTGTTATTGATCCAGTCACGAAAAAGACTATACACG  
GTAAAGAACGTTTGGTCTTCAACTACAAACGCCTGAACGACAACACTGAAAAGGACCAGT  
ACTCGCTACCCGGTATACAGACCATCCTAAAGCGGGTGGGCAACAAGAAGGTCTTCAGC  
AAGTTTCGATCTAAAATCGGGCTTCCATCAAGTTGCCATGGCGGAAGAGTCCATCCCTTGG  
ACTGCTTTCTGGGTACCGCAGGGCCTTTATGAATGGTTAGTGATGCCCTTTGGGCTCAAA  
AACGCTCCTGCAGTATTTCAAAGAAAAATGGACCAATGCTTCAAAGGTACAGAAGAATTCA  
TTGCAGTGTACATTGATGACATTTTGGTCTTCAGCGAAAATATGGCAGAGCATACCAAGCA  
CATTGGAATCATGCTCAAGATCTGCCAAGAAAATGGGCTGGTACTAAGCCCAAGTAAGAT  
ATGTCTTGCCCAACGCGAGATTGAATTTTTGGGCACAGTCATCTCACAAGGACAAATGAA  
GCTTCAGGCCCATGTAATCAAAAAGATAGTCAACAAAGCCAACATAGAGCTAGAAACAAC  
AAAAGGCCTGAGATCCTTTTTGGGCCTCCTCAACTATGCCCGTATCTACATACCCAACCT  
GGGTAGAAAGCTAAGTCCACTATATGCCAAGACTAGTCCCACCGGAGAGAAAAGGTTTAA  
TCGACAAGATTGGCATCTGATAAAGGAAATCAAAGATATGGTCCAAAAGCTCCCAAACCT  
CGCTATCCCACCAGCAAGATGCTACATTATCATTGAAAGCGATGGCTGCATGGAAGGATG  
GGGGGCCGTATGTAAGTGGAATTAGCAAAAGAAGATTCCCGCACTACTGAAAAGATTG  
TGCGTACGCTAGTGGGAAATTGCGGTGTTGTCAAGTCTACCATCGACGCCGAAATTTACGC  
ACTCATAAAAGCATTGGAATCTTTTAAAATCTTCTATCTGGACAAAAAACATTTGGTGGTGC  
GAACAGACTGTCAAGCGATAGTGACATTTTACAACAAGACAAGTACTCATAAACCCAGTA  
GAATACGTTGGATCACCTTTTCCGACTATATAACGGGGTTAGGAGTTCCGGTTACTATCG  
AACACATAGACGGAAAAGAGAACCAGTTGGCTGATACTAAGCAGACTCGTGTATACAA  
CATGGAACCAGTCCCAAACCTACCAACCGGAGGAAGAAGAGCTGGAGAAGTCCCAACAT  
CTCAGCTTCGCGGGGCTAGCTATCCCTATAGCTTGGCCTATGATGGGCTCCTACAACAAA  
AGAAGGACGCCATTACTCACGGGACAATCACTCTGGCAACGGAACAAGCCATCTCAGCA  
CAGCTCTACCGCATCGAAGAGCAGGCAGCCAAGAAAGCACTATTGGCCCTACGTGACCT  
ACAGGGCGTACTCCACTTCAAGAGAGACTATTTGGCCGCTACTGCCACTAGAGACAACCTG  
GGCTAGCGACAGACTGCCAGCTGCCCAACAAGACTCAGCCGCCCTAGACCAACATGCTG  
GCGTGATTAACGCCATTATTGAAAGGGCTGTCCAACCCTAGTTTGGACGGTAGTAGTAGG  
TGTAATAATAATAGTTAGGTGTGCTTTACTTTTCCAAGCTGTCACTCATTATAGAGTAGACA  
TGATGATCGACGATGGGGCCAATGAGCACCCGGATCACCATTCTCCCATCTATAAATGA  
GAGTTTGTAAAGGCTTAGCCATCAGAGAGTGAAAACCTCAACTGATCCTAAGTGTTAGA  
GTTTGTATTTTCTAAGAGTTTGTAAAGATTTTATGAAATAAAGAGTCTACTTTGTGTTTATC

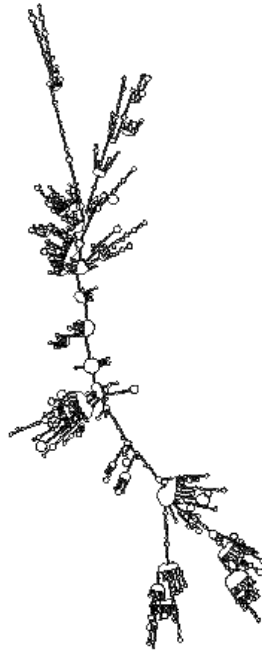
TCTTTGTTTCACCTGGGATTTAAACAGTTTTTTGTTTTTCCGCACCATCGGTTTGCGCCCGA  
TCGATGT

Structure in dot-bracket format:

[illegible]

Energy:  
-2056.9

### RNA-Structure:



## Virus Domain(s)

Number of proteins:  
3

### Protein 1

Number of domains:  
1

Aminoacidsequence:  
VTNNQLAHNLA VTFDRVNLGNRVFIKNLKQIQESILELNTRVDTIEVALRRLTKTFRESKPLTES  
EVRKLV EEEIAQQPKIVEKQALEISQQLEQKLEKVEKLLHKLDQWVGQ

Starting nucleotide position:  
533

Ending nucleotide position:  
872



### Domain 1

Starting aminoacid position:  
1

Ending aminoacid position:  
99

Domain description:  
Protein of unknown function (DUF1319)

Identifier:  
DUF1319

### Protein 2

Number of domains:  
0

Aminoacidsequence:

MTDSPAYQEALKEAEKVDPPAVGITTSTGVTAVQGFKTVIKQNNVQICLLAVIADKLEELIQDQ  
KKARKDKAKEVAIPEDLITKLQGLSIQEKGEAKVTRKPEPKGTLFGFKDPYKILAAEKAKITLKP  
VKE

Starting nucleotide position:

868

Ending nucleotide position:

1267



### Protein 3

Number of domains:

4

Aminoacidsequence:

MSRARPQHVPVSVTTTTSEQNREGPLYEDQIRDYRRGQRRIFNLRRRARRLRRSMMGSRYSQ  
ETLEQEIDPQTTLRLSMQERARLVPAEVLYRSRRDTVHHRVYTHRSEESVLCVGGGSQVDRAF  
IQPESLEQLQRTGMSFIHIGILQVRIQILHRQEEGTALVVFRDNRWSGDQSIFAQMEIDLTGK  
SQLVFVIPDTMMTIGDFARNVQLSILTRGYENWQNGEANLLITRGMTGRLSNTPNVAFAYQIA  
SATDYLASHGVKAIAGKKMNLQHLRNQQWILRPPQTDITPMQPRSVETRNLVDGSISIRFHDY  
EAATSASRPHYNEEDEEVESETESEIREHTIAVWIGEEEEIPDQTGRKKVWEESSNGNGRFFRY  
YTPPPTFEGQIIATGWGSDDDNEKTPPKWDESPDEEGPTEPIWDQEEEEDEYDPNVYRAYL  
QKEEDEWQEITASLREEMEYPKRRPQTEMAFSETVDYTPPGDTMMTPVGYPASSSRSTVT  
TPSRPPLFEGRTTHVPRFLKRDEYTEWWQLPSSQGTGALFVMPKQMGLFHEVFSRWESIT  
KNYVAAQGFTDPTEKMEFMENLLGETEKLTIQWRMNYEAEYQQLLTQADGRQGTQNILSQI  
KRIFSLEDPASGSTRIQDAAYRDLERLTCHNIKDIVQFLNDYGRLAAKSGRLFLGTELSEKLWM  
KMPPPELGHRMKEAFQKEYSGNEVGVPFPRILFAYRYLEQECKDAAFKRSLKSLSFCKDMPLTG  
YYDKTSKYGMRKSRTYKKGPHASHARVEKRKHLIRNKKCKCYLCGDEGHFARECPNQRRDV  
KRVAIFEGIDLPEGFDIVSVEEGEEESDAIYSISENEDGELDTEVVHEKVFMREEDQSYWLG  
KTNHWTAMVRVSSQQYHCMHQWEHNKEILVVAHINCHFCQPTQLRSRIHCPTCQLTSCFM  
CAPIYCNMIVQQQPKPPVPFNTHLLQQQAAYIQWLEKENQRLTEAVEFYKKEAEELRLERDL  
EQDRRSLEPTLLDKGKKVQILDPEDEDQHTAYLEEDTISRVIIGHTVEQQEVRKPKVKKGNMLYNL  
DVVLHIPEVGRPIKVKAILDTGATTCCININSVPQTAIEQNTFLVQFRGINSTQSVDKKLKYGRM  
TISNHQFRIPYCYAFPLSLGDGIEMILGCNFIRGMYGGLRIEGHTITFYKNVTTIQTRLAAVMVG  
GTTASELGGGEESKSDSESMFDLSETEEFDSETHQQIVSHVAAQAQQQKLDPKLQQLMVQL  
QDQGFIGENPMQHWAKNKILCRLDIKNPDLIIEDKPIKHLTPAMEKQFQKHIALLDIGVIRPSK  
SKHRTTAFIVESGTVIDPVTKKTIHGKERLVFNKYRLNDNTEKDQYSLPGIQTILKRVGNKKVFS  
KFDLKSGFHQVAMAEESIPWTAFWVPQGLYEWLVMPFGLKNAPAVFQRKMDQCFKGTEEFI  
AVYIDDILVFSENMAEHTKHIGIMLKICQENGLVLSPSKICLAQREIEFLGTVISQGMKLQAHVI  
KKIVNKANIELETTKGLRSFLGLLNYARIYIPNLGRKLSPLYAKTSPTGEKRFNRQDWHLIKEIKD  
MVQKLPNLAIPPARCYIIIESDGCMEGWGAVCKWKLAKEDSRTTEKICAYASGKFGVVKSTIDA  
EIYALIKALESFKIFYLDKKHLVVRTDCQAIVTFYNKTSTHKPSRIRWITFSYITGLGVPVTIEHI  
DGKENQLADTL SRLVYTTWNQSQTHQPEEEEELEKSQHLSFAGLAIPAWPMMGSYNKRRTPL  
LTGQSLWQRNKPSQHSSTASKSRQPRKHYPYVVTYRAYSTSRETIWPLLPLETTGLATDCQL  
PNKTQPP

Starting nucleotide position:

1271

Ending nucleotide position:

6776



### Domain 1

Starting aminoacid position:  
86

Ending aminoacid position:  
241

Domain description:  
Viral movement protein (MP)

Identifier:  
MP

## **Domain 2**

Starting aminoacid position:  
786

Ending aminoacid position:  
803

Domain description:  
Zinc knuckle

Identifier:  
zf-CCHC

## **Domain 3**

Starting aminoacid position:  
1046

Ending aminoacid position:  
1163

Domain description:  
Retroviral aspartyl protease

Identifier:  
RVP

## **Domain 4**

Starting aminoacid position:  
1333

Ending aminoacid position:  
1494

Domain description:  
Reverse transcriptase (RNA-dependent DNA polymerase)

Identifier:  
RVT\_1