

**Towards comprehensive bioinformatics analysis of viruses
on the presence of prion-like proteins.**

This dissertation is submitted in partial fulfilment of the requirements
for the degree of Master of Science in Bioinformatics
by

Konstantin V. Suslov

**New College
University of Oxford
2007**

Acknowledgements:

Many thanks to all our teachers and tutors on the course “MSc in Bioinformatics”. Special thanks to Dr.A.R.Dalby for useful discussions and criticism.

I am grateful to the Computational Biology Research Group (CBRG) at University of Oxford for the opportunity to use CBRG resources for this project.

My acknowledgements to Prof.A.J.McMichael for constant support. My acknowledgements to Mr.A.Smith, the Trustee of the Hill Foundation, for constant support. This work has been supported by The Hill Foundation.

Abstract:

Analysis of the literature data on sequence homologies, structural-functional similarities between viral proteins and prions led to the formulation of the “virus-prion ” theory according to which the virus is composed of the prion-like proteins and the nucleic acid which encodes them [1]. This theory opens a promising approach to create an anti-HIV-1 vaccine based on viral fibrillization [1, 2]. No comprehensive bioinformatics analysis has yet been done in the direction to support the “virus-prion” theory. In this dissertation possible approaches and the methodology to do such bioinformatics analysis are being developed. Different bioinformatics tools for structural alignment of proteins, pairwise/multiple sequence alignment of both DNA and proteins have been used. BLASTn searches for the sequence homology with a prion gene were successful and for some regions of HIV-1 genome sequence identity with a prion was 100%. For the first time multiple sequence alignment of different regions of HIV-1 genome and the prion gene has been performed. From more detailed analysis it follows that tRNA-like sequences in the genes might be potential hallmarks of the encoded prion-like proteins. Four Squares Code is expected to provide some clues to understanding of the nature of the prion-like proteins and prions.

Four squares code

		Second Position of Codon	
		W	S
First Position of Codon	W	Phe/Leu Asn/Lys Ile/Met Tyr/Ter	Thr Cys/Ter/Trp Ser Ser/Arg
	S	Leu Asp/Glu Val His/Gln	Ala Arg Pro Gly

The Table of Contents

Introduction	6
Methods.....	9
Results.....	12
1. Multiple sequence alignment of the 9 regions of HIV-1 genome and the prion gene.....	12
2. Structural alignment of viral proteins and prion.....	15
3. Aligning prion gene with genes/genomes of other viruses.....	15
4. Multiple protein sequence alignment of the HIV-1 proteins and prions.....	16
5. Analysis of the viral and prion sequences on the presence of the tRNA elements.....	16
6. BLAST searches.....	18
7. Four Squares Code.....	21
Discussion.....	22
References.....	24
Figures	27-60

Introduction

We live in the new era, “era of genomes” [3]. Although the human genome has been sequenced, many scientists still wonder what kind of information can be extracted from the data on sequence of the nucleotides in the genome. Bioinformatics is the field which uses computer-based tools to analyse both DNA and protein sequences for different biological tasks. Although bioinformatics nowadays is a very wide field which includes analysis and prediction of the protein structure, microarrays analysis etc, it’s no doubt that sequence alignment is the “gold of bioinformatics”. Many research fields started to use bioinformatics very intensively and bioinformatics can be considered as an applied science. It is nice to see the books like “Immunological bioinformatics” [4] in which bioinformatics is applied to immunology. In this dissertation bioinformatics will be applied to the analysis of the proteins, genes, genomes of different viruses and, in particularly, of the human immunodeficiency virus 1 (HIV-1). Such an analysis may prove to be useful for understanding what is the nature of HIV-1 with implications for an anti-HIV-1 vaccine.

Articles like [5, 6] about structural and functional similarities between prions and viral proteins inspired me to do more detailed search through the “PubMed” [7]. Then it was easy to find different articles about sequence homology and/or structural-functional similarities between prions and proteins of different viruses. Let us just mention several [8-15] of the articles, reviewed in [1]. Finally it led to the formulation of the so-called “Virus-Prion Theory” according to which the “Virus – is an aggregate of prion-like proteins with the nucleic acid encoding them” [1]. Prion-like proteins were defined as proteins which have “either sequence homology or structural-functional similarities, or both with prion”[1]. The main suggestion which follows from the virus-prion theory is the attractive possibility to induce the conformational changes in the prion-like proteins of HIV-1 to trigger the chain process of viral fibrillization [1, 2]. This approach might lead to the creation of the new type of anti-HIV-1 vaccine [1, 2]. It is expected that this vaccine will eradicate the virus in a way which is safe for the human organism [1, 2].

There was an independent study [16] which complements the work [1]. In both works [1, 16] we can find the data from the literature on sequence homologies, structural and functional similarities between viral proteins and prions. In the study [1] we can find: “Interestingly, how prions, these “slow viruses”, could have organized in a such complex aggregates as viruses.” This would automatically imply that viral proteins have their origin from prions, while in the study [16] we can find that prion has “its evolutionary origin as a horizontally transferred gene from an early RNA virus “. It should be noted that article [1] has been submitted to publication on the 19th of September 2004. The article [16] has been submitted for publication on the 16th of February 2005. Paradoxically, but two years after the publication of the articles [1] and [16] there appeared the article [17] with similar ideas as in the article [1]. Anyway, the trend is that now scientists are getting closer and closer to the understanding what is the virus in terms of prions and prion-like proteins [1]. The situation in the prion field became even more intriguing [18] after the publication of the article about detection of virus-like particles in the scrapie- or CJD agent- infected cells [19].

It should be noted that nobody yet has done a comprehensive bioinformatics analysis of the viral proteins to elucidate their inherent evolutionary link with prions. This dissertation does not pretend to do such a comprehensive bioinformatics analysis of viral proteins on their relation to prions, but rather has its small but important task to develop an approach for such an analysis. In other words the task of this dissertation would be to propose and try possible approaches to do such a bioinformatics analysis of viruses (particularly HIV-1), to detect some borders for such an analysis. There were several attempts to do sequence comparisons between prions and viral proteins. Pioneering studies were done by Haseltine and Patarka [20] but then these studies faced criticism [21-23]. Despite the fact that the strong support to the article [20] has been given by the

work [24], the discussion of the topic is not completed [23, 24]. The main topic was the discussion of the statistical significance of the alignments [20-24]. The main result from the article [20] was that four colinear subsequences in HIV-1(HTLV-III/LAV) *pol* gene have been found to have sequence similarity with hamster gene for PrP 27-30. It is of particular interest that the sequence identity with a prion gene was higher than 60 % for each of these 4 subsequences each with a total length over 50 nucleotides [20]. Also the sequence similarity was found on the protein level [20].

From the contemporary investigations, the article by Kuznetsov and Rackovsky [25] is an example of the article where the authors use bioinformatics analysis to identify sequence similarity between prion and viral protein, and show the statistical significance of such a comparison. As we can see from the articles [20, 25] it takes in average one article per one viral gene and/or protein to study deeply (with investigation of statistical significance) the sequence similarities between prions and viral proteins. The question arises how to manage the analysis, for example, of 1000 viral proteins? Is this the task for 1000 articles or rather it would be rational to develop the bioinformatics approach and then to do comprehensive analysis based on it? Intriguingly, but there is a silence in sequence comparisons of prions and viral proteins ... Maybe the task is too controversial based on the examples [20-24] or the statistical significance of the alignments is the significant obstacle, or that simple BLAST searches do not lead to significant hits? Maybe, as it can be seen from the review of the literature in [1], this is the task for the structural bioinformatics? Or maybe the clue is even more intriguing and includes both sequence analysis, structural analysis and requires to decipher some more sophisticated codes?

If the bioinformatics analysis would support the “Virus-Prion” theory [1], it would not only give some clues to the nature of HIV-1 but will also automatically support the vaccine based on viral fibrillization [1, 2].

In this dissertation different approaches have been chosen to develop the methodology of the bioinformatics analysis of the viruses. Taken altogether these approaches complement each other and allow to control the significance of findings. For example, the successful results from the BLASTn searches justify the multiple sequence alignment of the regions of the HIV-1 genome with the prion gene. Data from the structural alignment of proteins have also been used as a control for the results of the multiple alignment of protein and DNA sequences. Among the methods for pairwise and multiple sequence alignment the following bioinformatics tools have been used in the project: EMMA (ClustalW-based) [26], MAFFT [27], AMAP [28]. For the search for local sequence similarities different BLAST programs [29-31] have been used. For structural alignment of protein sequences CE (combinatorial extension) program [32, 33] and TM-align program [34, 35] have been used. Among the tools for visualization of the results the following tools proved to be very useful: PrettyPlot [36], JalView [37], Swiss-PdbViewer [38]. Among the websites with the collection of the bioinformatics tools and the collection of sequences/structures the most useful were: NCBI website [39, 40], CBRG website [41], MPI Bioinformatics Toolkit website [42, 43], PDB website [44, 45].

Human immunodeficiency virus 1 genome has been chosen among other viral genomes for the analysis on the presence of the prion-like proteins. The original algorithm has been developed for quick finding of the regions (in the HIV-1 genome) which potentially have sequence similarity/homology with a whole prion gene. Then multiple sequence alignment of these regions of HIV-1 genome and the prion gene has been obtained (Section 1 in Results: “Multiple sequence alignment of the 9 regions of HIV-1 genome and the prion gene”). In this global alignment there are regions of high local similarity. It is also important to note that the shuffled HIV-1 genome sequence served as a control in this bioinformatics experiment. The alignment was better in the case of the normal HIV-1 genome in comparison to the shuffled one. The significance of these findings has also been supported by the results from BLASTn searches (Section 6 in Results: “BLAST searches”) which led to the significant hits located in *nef*, *vpr* genes and the *vpu/env* region of the HIV-1. The lengths of the regions of hits were 18(100%

sequence identity), 17(100% sequence identity) and 43(79% sequence identity) nucleotides for *nef*, *vpr* genes and the *vpu/env* region, respectively.

It should be noted that prion gene has also been aligned with genomes/genes of the several other viruses/viroids (Section 3 in Results: “Aligning prion gene with genes/genomes of other viruses”). In the analysis of the structures (Section 2 in Results: “Structural alignment of viral proteins and prion”) as an example of the structural alignment, prion protein and the HIV-1 Vpr protein have been aligned in the CE program [32, 33]. Gp120 V3 loop of HIV-1 has been aligned with prion protein in TM-align program [34, 35] and the results have been compared to the literature data [6] on CE alignment of the same polypeptide chains. In all the cases the results of the structural alignment were used for comparison with findings from the pairwise/multiple sequence alignments.

Multiple sequence alignments of HIV-1 proteins with a prion protein were not as successful as MSA of HIV-1 regions on the DNA level. But it was interesting that in MSA (when using MAFFT program [27]) of the prion protein region of interest and HIV-1 proteins the results are consistent with structural alignment by CE program. Also analysis of MAFFT multiple sequence alignment led to the conclusion that the pattern of physico-chemical properties (volume; hydrophobicity etc) of amino acid residues along the polypeptide chain might explain better the structural similarity between the proteins rather than analysis of proteins on sequence homology/similarity (Section 4 in Results: “Multiple protein sequence alignment of the HIV-1 proteins and prions”).

In the case of prion-like proteins not only the structural similarity with prions might play a crucial role but also the patterns and types of the amino acid residues interacting with each other along the polypeptide chain upon the conformational changes (α/β -switch) in the prion-like proteins. Four Squares Code has been proposed to give some clues to the understanding of the these phenomena (Section 7 in Results: “Four Squares Code”).

Finally, after combining the results from the multiple sequence alignment (of the regions of HIV-1 genome with prion gene) (Section 1 in Results) with those from the structural alignment (Section 2 in Results) and taking into account the data on pairwise/ multiple sequence alignment of the tRNAs/sets of tRNAs with a prion gene and/or genomes of viruses (including HIV-1)/viroids (Section 5 in Results) the hypothesis has been put forward that tRNAs-like sequences in the genes might be the hallmarks of the prion-like properties of encoded proteins (Section 5 in Results: “Analysis of the viral and prion sequences on the presence of the tRNA elements”).

Further bioinformatics studies on the understanding what is a prion-like protein are required. All these approaches when combined and applied to the analysis of a wider set of viral genomes, genes and proteins might lead to the comprehensive analysis of viruses on the presence of the prion-like proteins. If automated, it is expected that in the future this analysis might take a relatively short time despite the huge number of viral genomes.

Methods:

1) Multiple sequence alignment of the 9 regions (sequences 1, 2, 3, 4, 7, 8, 9, 11, 12) of HIV-1 genome and prion gene (sequence 0).

Multiple sequence alignment has been done via the “Tutorials”[21] section in the CBRG group website [41], Oxford. MSA has been done by using EMMA program [26] and the results were visualized in PrettyPlot program [36] through the “tutorial” section at CBRG website [41]. The following sequences have been used:

Prion gene (>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds)

HIV-1 genome (>gi|4558520|gb|AF033819.3| HIV-1, complete genome). Sixteen sequences: Regions 1-16 (sequences 1-16) with the borders according to the Table 1 in the Results section.

In the control bioinformatics experiment to estimate the significance of the alignment ClustalW 1.82 program [26] through the MPI Bioinformatics ToolKit [42] has been used. “SHUFFLESEQ” program [61] has been used to randomize HIV-1 genome. Results were visualized by using PrettyPlot program [36]. Perl program has been written to get the subsequences of HIV-1 genome (results are in one (.txt) file) from the information (written in a separate (.txt) file) which includes a set of (start) and (end) positions of the corresponding regions.

2) Structural comparison of protein structures.

Structural alignments of protein sequences were done by using CE program [32] via CE website [33]. Protein Data Bank [44] codes for proteins used in multiple structural alignment: 1QLZ:A; 1IFD:.; 1DX0:A; 2CPS:.; 1AUM:.; 1LG7:A; 1G03:A; 1ARO:P; 1M8L:A and 1TKN:A. Protein Data Bank [44] codes for proteins used in pairwise structural alignment: 1QLX:A and 1VPC:.

Structural alignment of protein sequences was done by using TM-align program [35] through TM-align website [34]. Protein Data Bank [44] codes for proteins used in pairwise structural alignment: 1QLX:A and 1CE4:A.

3) Multiple sequence alignment of the region of HIV-1 *env* gene and the two regions of the prion gene was done by using EMMA program [26] through CBRG website [41]. Results were visualized in PrettyPlot [36] through the same website [22]. Sequences:

>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds
>gi|4558520|gb|AF033819.3| HIV-1, complete genome.

4) Pairwise sequence alignments of several viral genomes (or their genes) with a human prion gene were done using EMMA program [26] through CBRG website [41]. On the same website [22] the results were visualized through PrettyPlot [36]. The following genomes/genes of the viruses or viroids were investigated:

a) >gi|5805278|gb|AF144301.1|AF144301 Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) polymerase (PB1) gene, complete cds

b) >gi|335873|gb|J02428.1|VSVCG Vesicular stomatitis Indiana virus, complete genome

c) >gi|563600|emb|X78602.1|PCVRNA1 Peanut clump virus genomic RNA 1

d) >gi|58574|emb|X17101.1|AGVCCS Australian grapevine viroid central conserved sequence

e) >gi|59042|emb|X14638.1|CVDIV Citrus Viroid IV (CVd IV) mRNA

f) >gi|1304014|dbj|D84412.1|BNY25KP3 Beet necrotic yellow vein virus genomic RNA for 25k protein, complete cds

It should be noted that in all cases the following sequence for the prion gene has been taken:
>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds

Sequence for control alignment:

>gi|3249641|gb|AF070082.1|HSMKKIV03 Homo sapiens mitogen-activated protein kinase kinase 1 (MKK4) gene, exon 3

5) Multiple sequence alignment of the HIV-1 proteins (Pol, Env, Vif, Vpu) and prion proteins has been done via MAFFT [27] program at the MPI Bioinformatics Toolkit website [42]. Results were visualized in JalView Editor [37]. Also AMAP program [28] at AMAP website [28] has been used for multiple protein sequence alignments.

The following sequences have been picked up randomly from the NCBI website[7]:

```
>1_gi|17352345|gb|AAL01564.1| pol [Human immunodeficiency virus type 1]
>2_gi|114842142|dbj|BAF32553.1| Pol [Human immunodeficiency virus 1]
>3_gi|119959862|gb|ABM16847.1| pol polyprotein [Human immunodeficiency virus 1]
>4_gi|33518289|gb|AAQ20622.1| pol protein [Human immunodeficiency virus 1]
>5_gi|13604570|gb|AAK32307.1| pol protein [Human immunodeficiency virus type 1]
>6_gi|33235347|emb|CAE17440.1| pol polyprotein [Human immunodeficiency virus 1]
>7_gi|27466697|gb|AAO12804.1| pol protein [Human immunodeficiency virus 1]
>8_gi|111119126|gb|ABH05989.1| pol protein [Human immunodeficiency virus 1]
>9_gi|38231472|gb|AAR14670.1| pol protein [Human immunodeficiency virus 1]
>10_gi|21617331|gb|AAM66628.1| pol polyprotein [Human immunodeficiency virus type 1]
>11_gi|13736028|gb|AAK35310.1| pol polyprotein [Human immunodeficiency virus type 1]
>12_gi|116273601|gb|ABJ92132.1| pol protein [Human immunodeficiency virus 1]
>13_gi|73810441|gb|AAZ86182.1| pol protein [Human immunodeficiency virus 1]
>14_gi|32185958|gb|AAP72942.1| pol protein [Human immunodeficiency virus 1]
>15_gi|9939955|gb|AAF91064.1| pol polyprotein [Human immunodeficiency virus type 1]
>16_gi|10953650|gb|AAG25524.1| pol polyprotein [Human immunodeficiency virus type 1]
>17_gi|60616146|gb|AAX31194.1| pol protein [Human immunodeficiency virus 1]
>18_gi|9943619|gb|AAF90224.1| pol protein [Human immunodeficiency virus type 1]
>19_gi|14537670|gb|AAK66659.1| pol protein [Human immunodeficiency virus 1]
>20_gi|116273320|gb|ABJ91992.1| pol protein [Human immunodeficiency virus 1]
>21_gi|1575476|gb|AAB09538.1| env
>22_gi|4324895|gb|AAD17159.1| Env [Human immunodeficiency virus type 1]
>23_gi|530117|gb|AAB05009.1| env [Human immunodeficiency virus type 1]
>24_gi|8919904|emb|CAB96226.1| env polypeptide [Human immunodeficiency virus 1]
>25_gi|62904348|gb|AAY19907.1| envelope glycoprotein [Human immunodeficiency virus 1]
>26_gi|62906070|gb|AAY20701.1| envelope glycoprotein [Human immunodeficiency virus 1]
>27_gi|34328751|gb|AAQ63675.1| envelope glycoprotein [Human immunodeficiency virus 1]
>28_gi|80974730|gb|ABB54055.1| envelope glycoprotein [Human immunodeficiency virus 1]
>29_gi|57545250|gb|AAW51528.1| envelope glycoprotein [Human immunodeficiency virus 1]
>30_gi|424648|gb|AAA44789.1| envelope glycoprotein
>31_gi|87243404|gb|ABD34128.1| envelope glycoprotein [Human immunodeficiency virus 1]
>32_gi|3832007|gb|AAC70487.1| envelope glycoprotein [Human immunodeficiency virus type 1]
>33_gi|2190754|gb|AAC40360.1| envelope glycoprotein [Human immunodeficiency virus type 1]
>34_gi|15638437|gb|AAL04968.1|AF409618_1 envelope glycoprotein [Human immunodeficiency virus type 1]
>35_gi|62869171|gb|AAY17769.1| envelope glycoprotein [Human immunodeficiency virus 1]
>36_gi|37140501|gb|AAQ88383.1| gp120 [Human immunodeficiency virus 1]
>37_gi|55846192|gb|AAV67120.1| envelope glycoprotein [Human immunodeficiency virus 1]
>38_gi|31559621|dbj|BAC77450.1| Env glycoprotein [Human immunodeficiency virus 1]
>39_gi|14530252|gb|AAK65984.1|AF286238_8 env protein [Human immunodeficiency virus type 1]
>40_gi|1128989|gb|AAC37912.1| envelope glycoprotein gp120
>41_gi|17352346|gb|AAL01565.1| vif [Human immunodeficiency virus type 1]
>42_gi|70633576|gb|AAZ06067.1| vif protein [Human immunodeficiency virus 1]
```

>43_gi|46405171|gb|AAS93445.1| vif protein [Human immunodeficiency virus 1]
 >44_Vif_4
 MENRWQGLIVWQVDRMKIRTWNSLVKHHMYVSKRASGWFYRHHYESRHRVRSSEVHIPL--
 GEAKLVIITYWGLQTGEREWHLGHGVSIEWRLRRYSTQVDPGLADQLIHMYFYDFCFADS
 AIRKAILGHIVIPRCDYQAGHNKVGSLQYLALTALIKPKKRKPLPSVRKLVEDRWNKSQ
 KTRGRRGNHTMNGH
 >45_gi|9931092|emb|CAC05363.1| viral infectivity factor [Human immunodeficiency virus 1]
 >46_gi|122056625|ref|NP_001073591.1| prion protein preproprotein [Homo sapiens]
 >47_gi|34335270|ref|NP_898902.1| prion protein preproprotein [Homo sapiens]
 >48_gi|56204170|emb|CAI19053.1| prion protein (p27-30) (Creutzfeldt-Jakob disease, Gerstmann-
 Strausler-Scheinker syndrome, fatal familial insomnia) [Homo sapiens]
 >49_gi|32261460|gb|AAP76525.1| vpu protein; Vpu [Human immunodeficiency virus 1]
 >50_gi|122912584|gb|ABM67989.1| vpu protein [Human immunodeficiency virus 1]
 >51_gi|74315728|gb|ABA02456.1| vpu protein [Human immunodeficiency virus 1]
 >52_gi|25166826|gb|AAN73615.1|AF484497_6 vpu protein [Human immunodeficiency virus 1]
 >53_gi|61106261|gb|AAX38863.1| vpu protein [Human immunodeficiency virus 1]

6) Multiple sequence alignment of the HIV-1 proteins (Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef) and prion protein has been performed by using using EMMA program [26]. The results were visualized in PrettyPlot program [36] through the CBRG website [41]. Sequences:

>1_gi|28872819|ref|NP_057849.4| Gag-Pol [Human immunodeficiency virus 1]
 >2_gi|9629360|ref|NP_057850.1| Pr55(Gag) [Human immunodeficiency virus 1]
 >3_gi|9629361|ref|NP_057851.1| Vif [Human immunodeficiency virus 1]
 >4_gi|28872817|ref|NP_057852.2| Vpr [Human immunodeficiency virus 1]
 >5_gi|9629358|ref|NP_057853.1| Tat [Human immunodeficiency virus 1]
 >6_gi|9629359|ref|NP_057854.1| Rev [Human immunodeficiency virus 1]
 >7_gi|9629366|ref|NP_057855.1| Vpu [Human immunodeficiency virus 1]
 >8_gi|9629363|ref|NP_057856.1| Envelope surface glycoprotein gp160, precursor [Human immunodeficiency virus 1]
 >9_gi|28872818|ref|NP_057857.2| Nef [Human immunodeficiency virus 1]
 >10_gi|4506113|ref|NP_000302.1| prion protein preproprotein [Homo sapiens]

7) Analysis of the viral and prion sequences on the presence of the tRNA elements has been done by using EMMA program [26]. The results were visualized in PrettyPlot program [36] through the CBRG website [41]. Sequences:

>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds
 >gi|4558520|gb|AF033819.3| HIV-1, complete genome
 >Val_tRNA_gi|17981852:1604-1672 Homo sapiens mitochondrion, complete genome
 >Leu_tRNA_gi|17981852:12267-12337 Homo sapiens mitochondrion, complete genome
 >His_tRNA_gi|17981852:12139-12207 Homo sapiens mitochondrion, complete genome
 >Tyr_tRNA_gi|17981852:c5892-5827 Homo sapiens mitochondrion, complete genome
 >gi|323275|gb|J02049.1|CCC Cadang-cadang coconut viroid, short subspecies, complete genome
 >gi|4731630|gb|AF135121.1|HSM059JP2 Homo sapiens tumor suppressor protein p53 (p53) gene, exons 10 and 11 and complete cds

8) BLAST searchers have been done through the “BLAST” section [29] at the NCBI website [39] by using different BLAST programs [29-31]. The following sequences for the prion gene and corresponding protein have been used:

>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds
 >gi|5668953|gb|AAD46098.1| prion protein [Homo sapiens]

Results:

1) Multiple sequence alignment of the 9 regions of HIV-1 genome and the prion gene

The idea was to take the HIV-1 genome and to run the following algorithm for finding prion-like fragments in HIV-1:

The following algorithm has been developed: 1) to take the whole HIV-1 genome and to do pairwise sequence alignment between the prion gene and the HIV-1 genome. 2) Then to do the pairwise alignment between the rest of the HIV-1 genome and the prion gene etc. It should be noted that starting from this step of the algorithm and at every further step it does not matter what is the order in which parts of the HIV-1 genome for pairwise alignment with a prion gene are being taken, but what is important that only continuous parts of HIV-1 genome are being aligned. 3) Repeat the procedure until the whole HIV-1 genome is divided on the regions where the prion gene has been “annealed” according to pairwise alignment algorithm mentioned above. It should be noted that at the end of implementation of this algorithm some short patches of HIV-1 genome might not be covered due to their very short length in comparison to the length of the prion gene.

The algorithm proposed here in other words will first find the part of the HIV-1 genome which is most similar to the prion gene, then another which is less similar etc.

In the Table (1) regions are listed after the implementation of the algorithm mentioned above:

Analysis has been performed on:

HIV-1 genome (“>gi|4558520|gb|AF033819.3| HIV-1, complete genome”[7]) and *prion gene* (“>gi|5668952|gb|AF076976.1|AF076976 Homo sapiens prion protein (PRNP) gene, complete cds”[7]).

Table 1. Regions of HIV-1 after implementation of the “annealing” algorithm

Number of the region	Positions in the HIV-1 genome (9181 bp)
16	0-75
1	76-914
5	915-1178
4	1179-2059
6	2060-2584
3	2585-3387
8	3388-3883
7	3884-4779
10	4780-5006
9	5007-5910
12	5940-6789
13	6790-7088
11	7089-7975
14	7976-8110
2	8111-9031
15	9032-9181

In each step of the algorithm program Emma (which is based on ClustalW program)[26] was run to do pairwise sequence alignment. Alignments were visualized in PrettyPlot program. Both Emma and PrettyPlot were used as part of the tools available for medical students at CBRG [41], University of Oxford, Oxford.

Then the same program Emma was used to run multiple sequence alignment on the 16 regions of HIV-1 genome (Table 1) and prion gene (designated as 0 in MSA).

It should be noted that the regions 15 and 16 actually cover LTRs (long terminal repeats) of HIV-1 genome – regions which have regulatory sequences and do not encode any proteins, and therefore regions 15 and 16 can be simply excluded from the subsequent analysis.

But first, I made multiple sequence alignment of all 16 sequences including prion (sequence 0), using gap opening penalty (GOP) 40 and gap extension penalty (GEP) 10 as parameters for multiple alignment (the default parameters in Emma are gap opening penalty 10.0 and gap extension penalty 5.0). GOP was raised to 40 especially with a purpose to minimize the number of gaps in the alignment, to make multiple sequence alignment more informative in terms of sequence similarity especially if the alignment gets better if we raise GOP (and that was actually the case). Also multiple sequence alignment was done with the same sequences but without regions 10, 14, 15 and 16 (GOP:40, GEP: 10). And then the alignment was done without regions 5, 6, 13, 10, 14, 15 and 16 (GOP:40, GEP: 10). It should be noted that all these regions (5, 6, 13, 10, 14, 15 and 16; Table 1) are smaller in size than the prion gene (762 bp) and in a way mislead the alignment. The results are presented in the (Box 1). No doubt that in this alignment we can see the regions of high sequence similarity to the extent which was unexpected if we take into account that the sequences in MSA almost do not contain gaps.

Also different parameters for GOP and GEP have been varied to see how the alignment changes and especially how the region 12 of HIV-1 genome will be “moved” (and in which direction) compared to the prion gene. The following combinations were tested (results are not shown): GOP 50 and GEP 15; GOP 50 and GEP 1.0; GOP 1.0 and GEP 50.

To estimate the significance of the produced alignment (Box 1) the following bioinformatics experiment has been proposed (Dr.A.R.Dalby, personal communication). The idea was to shuffle (or in other words to randomize) the sequence of HIV-1 genome and to apply the same algorithm mentioned above, to construct multiple sequence alignment of the HIV-1 regions (of shuffled genome) with a prion gene. If the alignment will look like the same as in the case of the normal HIV-1 genome it will mean that the alignment (in the case of the normal HIV-1 genome) represents “noise” and nothing more. If the alignment in the case of the normal HIV-1 genome is significantly better than in the case of the shuffled genome, then this approach (algorithm and the alignment of the regions of the normal HIV-1 genome) is justified. HIV-1 genome has been shuffled with the use of the program “SHUFFLESEQ” [61]. After applying the same algorithm (mentioned above) for searching the regions of HIV-1 with homology to the prion gene and then doing multiple sequence alignment of the obtained regions with a prion gene there were the following results (Box 8; Figures 1 and 2). We can clearly see that in the case of the normal HIV-1 genome the alignment is significantly better than in the case of the shuffled HIV-1 genome. This justifies the approach and gives a chance that HIV-1 genome indeed might contain relatively long regions with sequence similarity/homology to the prion gene. It should be also noted that in both alignments there were almost no gaps because GOP and GEP values were set up to 40 and 10, respectively.

In the article [6] we can find the information about structural analysis (based on CE program [32] for structural alignment of protein structures) of the prion protein, HIV-1 gp120 protein and Alzheimer peptide. It would be of interest to figure out how the data on DNA on multiple sequence alignment (regions of HIV-1 genome and prion gene) that were presented above are consistent with the structural data. Below are listed DNA regions corresponding to the regions of the structural similarity between prion and HIV-1 gp120 according to analysis based on CE program [6]:

```

>1_(CVN...MMERVVE)_REGION_FROM_PRION_AF076976.1_REGION_535-633
TGCCTCAATATCACAATCAAGCAGCG
CACGGTCACCACAACCACCAAGGGGGGAGAACTTCACCGAGACCGACGTTAAGATGA
TGGAGCGCGTGGTTGAG
>2_(TRP...QAH)_REGION_6659-6760_HIV-1_AF033819.3
ACAAGACCCAACAACAATACAAGAAAAAGAATCCGTATCCAGAGAGGACCAGGGA
GAGCATT
TGTTACAATAGGAAAAATAGGAAATATGAGACAAGCACAT

```

If we find these sequences within the sequences of the multiple sequence alignment (please see Box 1) we can see that the start of the prion region of interest (encoding CVN...MMERVVE) is located on 102 nucleotides (this in turn corresponds to 34 aa) earlier than the start of the gp120 region of interest (encoding TRP...QAH).

Also when using TM-Align program [35] for protein structural alignment the results differ from CE program for the same chains (1QLX (for prion), 1CE4 (for V3 loop of gp120)): prion region of interest is located to the right-hand side to the region of interest (V3 loop) in gp120. The end of the prion region of interest is located on $55 \times 3 = 155$ bp to the right from the end of the V3 loop as region of interest:

TM-align Results:

```

Chain 1:A946439      Size= 35
Chain 2:B946439      Size= 104 (TM-score is normalized by 104)

Aligned length= 19, RMSD= 2.70, TM-score=0.14307, ID=0.077

----- rotation matrix to rotate Chain-1 to Chain-2 -----
i          t(i)          u(i,1)          u(i,2)          u(i,3)
1    -14.4114712472    0.0228588341    0.7207138554   -0.6928556938
2     -9.3241560802    0.9939210509   -0.0910398547   -0.0619087187
3     -1.9445106428   -0.1076959530   -0.6872286982   -0.7184137374

(":" denotes the residue pairs of distance < 5.0 Angstrom)

CTRPN-----NNTRK--SIHIGPGRAFYTTGEIIGDIRQAH-----
      .           .           ..: :::::::::::::::
-----LGGYMLGSAMSR----PII-----HFG-SDYEDRYRENMHRYPNQVYYRPMDEYSNQNNFVHD
-----

CVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCITQYERESQAYYQR

```

So different approaches to structural alignment give different results but what is nice is that the regions of interest in the MSA are shifted against each other (the start of the prion region of interest (CVN...MMERVVE) is located on 102 nucleotides to the left from the start of the (TRP...QAH) region in the V3 loop of gp120) just on 102 nucleotides which is actually just 12% of the total length (849 bp) of the region 12 of HIV-1. It should be noted that it is even better than 19.4% (155/849) for the shift which reflects the difference between the results of CE and TM-Align programs.

This means that in a way the algorithm which has been applied to figure out the regions of HIV-1 together with a global alignment approach give results which are consistent with structural alignment.

After the fact about 102 nucleotides shift (between CVN...MMERVVE - encoding part in prion gene and TRP...QAH- encoding region in the HIV-1 *env* gene) in the MSA alignment the idea was born that there should be the fragment that is much smaller in size than the whole prion gene and which could be the clue to the determination whether the gene encodes prion-like protein or not. Later on we will return to the discussion what this fragment (approximately 100 bp in size) could be. But at the moment let us to do the following alignment. If in the previous MSA the CVN...MMERVVE - encoding part in prion gene and TRP...QAH- encoding region in the HIV-1 *env* gene have not been aligned straight against each other in terms of positions in the alignment, let us align these regions together with the rest of the prion gene (which in fact in the previous MSA has been located straight against the TRP...QAH- encoding region in the HIV-1 *env* gene) altogether. The results are presented in the (Box 3). Parameters for the MSA alignment were the following: GOP 40; GEP 10. We can see that there are local regions of high similarity in this alignment.

2) Structural alignment of viral proteins and prion

It has been mentioned above about the results from the TM-Align program for pairwise structural alignment of the prion (1QLX:A) and V3 loop of HIV-1 gp120 (1CE4:A). As another example on the Figures (1-3) of the (Box 2) we can see the results from running CE program for pairwise sequence alignment of the prion protein (red; 1QLX:A) and the HIV-1 Vpr protein (green; 1VPC:_) . Protein structures are aligned and on the Figures (1-3) we can see the views of this alignment from the different sides (Swiss-PdbViewer [38] has been used for visualization of protein structures). Root mean square deviation of this alignment is 2.4Å which is good one because it has been mentioned in the CE website [33] that even proteins with the same fold and from the same protein family may differ from each other by 4.0Å or even more in RMSD (the default value of RMSD on the CE website is less than 5.0Å). The so-called Z-score (which is the measure of the statistical significance of the structural alignment) was 3.7 which is not the best, but a normal one because proteins with a similar fold may have Z-scores of 3.5 or higher [33].

Then in another experiment prion protein (1QLZ:A) as a representative of the prion proteins was run on the CE program against PDB structures and with the parameters (Z-score > 3.5, RMSD < 5.0Å) there were 455 chains as a result. From this chains several proteins (in total 9 structures; PDB codes are listed in the “Methods” section) including viral proteins were selected to see the alignment of the viral proteins and the prion. The results are presented in the Figure (4) in the Box (2) and it should be noted that it is very convenient for the quick view to use “Compare3D Java Applet” through the CE website [33] to figure out the relative positions of the backbone chains in multiple structural alignment of the proteins. It should be also mentioned that in multiple protein structural alignment among the selected viral proteins there were C-terminal domain of the HIV capsid protein (1AUM:_) and HIV-1 Vpr protein(1M8L:A).

3) Aligning prion gene with genes/genomes of other viruses

In another set of bioinformatics experiments several other viral genomes (or their genes) were aligned with a human prion gene. The following genomes/genes of the viruses or viroids were investigated: Influenza A virus (A/Goose/Guangdong/1/96(H5N1)) polymerase (PB1) gene, complete cds; Vesicular stomatitis Indiana virus, complete genome; Peanut clump virus genomic RNA 1; Australian grapevine viroid central conserved sequence; Citrus Viroid IV (CVd IV) mRNA; Beet necrotic yellow vein virus genomic RNA for 25k protein, complete cds (Please also see “Methods” section for more detailed information).

As an example in the (Box 4) we can see the pairwise sequence alignment of the prion gene and peanut clump virus. It is an amazing feature of this alignment that there are at least 10 regions of high local sequence similarity between the sequences in this global alignment.

4) Multiple protein sequence alignment of the HIV-1 proteins and prions

To complement the alignments on the genetic level, multiple sequence alignments of proteins have been done using MAFFT [27] program and EMMA(ClustalW -based) program[26]. Also it was of interest to see whether this alignment can be compatible with the results from the structural alignment.

The sequences for the HIV-1 proteins Pol, Env, Vif, Vpu and prion proteins have been taken randomly from the NCBI [7] collection of sequences. Then using MAFFT program [27] proteins Pol, Env, Vif and the selected region (covering the region of interest CVN...MMERVVE) of the prion protein have been aligned. It is worth while mentioning that (CVN...MMERVVE) region is located in MSA (Box 5; Figures 1-2) almost in front of the region (TRP...QAH) and this is exactly what we can see from the pairwise structural alignment in CE program [6]. So, the conclusion is that the data on structural alignment are compatible with the data on protein sequence alignment.

It is also interesting that the pattern of hydrophobicity in the HIV-1 proteins and prion region of interest in this MSA is similar to the extent that sometimes it looks like as if the sequences were aligned based on the hydrophobicity or polarity of the amino acid residues. Indeed, in the description of the MAFFT[27] program we can find: “Homologous regions are rapidly identified by the fast Fourier transform (FFT), in which an amino acid sequence is converted to a sequence composed of volume and polarity values of each amino acid residue”[27]. When doing MSA for all the same proteins but with addition of HIV-1 Vpu protein, the prion region of interest (CVN...MMERVVE) is more shifted to the left relatively to (TRP...QAH) region in HIV-1 gp120, but the pattern of hydrophobicity in the MSA (Box 5; Figure 3) remains similar to the previous alignment (without Vpu; Box 5; Figures 1-2). It should be noted that MAFFT default parameters were used when running MAFFT[27] through the MPI Bioinformatics Toolkit website [42]. Results were visualized in JalView Editor [37].

Proteins of HIV-1 and the prion protein were also aligned by using Emma (ClustalW-based) program [26], but unfortunately did not give any exciting results (Box 6) in comparison to those obtained on the DNA level. This actually can be explained in a way that the DNA level is more fundamental and the amount of information written on the DNA level is at least 3 times more than on the protein level. Also as we saw from the example of structural alignment between prion protein and V3 loop of gp120 of HIV-1 in [6], sequence identity between the polypeptide chains is low (9.4%), although selected chains have significant structural similarity. Taking into account the data from MSA (Box 5; Figures 1-3) when using MAFFT program [27] it can also be suggested that the pattern of physico-chemical properties (hydrophobicity; volume etc) of amino acid residues along the polypeptide chain would be much closer to the understanding of structural similarity between the proteins rather than simple analysis of protein sequence similarity/ homology.

5) Analysis of the viral and prion sequences on the presence of the tRNA elements

Let us return to the discussion of what kind of the element could be the feature or a hallmark that the particular sequence encodes a prion-like protein. If to return to the description of the 102 bp shift between CVN...MMERVVE - encoding part in prion gene and TRP...QAH-encoding region in the HIV-1 *env* gene (please see “Multiple sequence alignment of the 9 regions of HIV-1 genome and the prion gene” section) in MSA, we can see that this shift is close to the size of the tRNA (70-90 nucleotides [46]). In the article [46] we can find that “Due to their

short length of about 70 to 90 nucleotides, tRNAs have a high enough statistical likelihood (50-70%) for carrying full length orfs. Like in a children's construction kit, tRNAs then might have served as the building blocks for longer orfs". The genomic tag hypothesis [47, 48], the existence of the somatic hypermutation (SHM) signature motifs in Ig(immunoglobulin)-like, tRNA-like EBERs (EBV-encoded small non-translated RNAs) in EBV virus [46], sequence similarities between HIV-1 gp120 and immunoglobulins, reviewed in [49], SHM hypothesis for mutations in the HIV-1 gp120 glycoprotein [50] and the fact that immunoglobulins and gp120 are representatives from the proteins in the Prion-like World [1] give a clue that tRNAs-like sequences in the protein genes might potentially be the hallmarks of the prion-like proteins. tRNA has also a feature that in its structure it has loops and stems, i.e. the regions of self-complementarity. The hypothesis is that this self-complementarity of tRNA on the RNA level, re-written on the DNA level of protein-encoding genes then might have found its realization in the "complementarity" on the protein level (it might be the complementarity according to Root-Bernstein's principle of amino acid pairing [53-55]; this complementarity turned out to be automatically included in the four squares code, please see Section 7 below), in interaction between the groups of amino acid residues along the polypeptide chain upon conformational change (α/β -switch) in the structure of the prion or prion-like molecule. To support this hypothesis further bioinformatics analysis on the interacting amino acids residues in the prion structures is required. Respective structures of the prion molecules both in α -helix and β -sheet conformations should be taken and the interacting amino acids along the polypeptide chain in both cases might be determined. It should be also noted that in Root-Bernstein's principle of amino acid pairing [53-55], codons for respective pairs of interacting amino acids are in parallel orientation. The only problem is that complementarity within the single tRNA molecule is written in the anti-parallel orientation and the size of the tRNA molecule is relatively short. Therefore not the one tRNA-like element but rather several (if they exist) of the nearest tRNA-like elements in the prion-encoding gene should be considered for such a bioinformatics analysis on mapping the complementarity both on the genetic and protein levels.

Now let us take a "children's construction kit"[46] composed of tRNAs to find out whether the genes for viral and prion sequences have tRNA-like elements.

First of all, Val-tRNA and the prion gene have been aligned using EMMA program [26], and the results are presented in (Box 7; Figure 1). What is exciting about this alignment is that tRNA "sticks" exactly to the (CVN...MMERVVE) - encoding part in the prion gene. This can be clearly seen when the same alignment is produced when (CVN...MMERVVE) - encoding part in the prion gene is aligned with Val-tRNA (Box 7; Figure 2). Taking into account the fact that gp120 is a prion-like protein [6, 13, 14] which has a structural similarity [6] with a prion protein in the region (CVN...MMERVVE) of the prion sequence, it can be proposed that the region (CVN...MMERVVE) in the prion could potentially be the determinant of the prion-like properties. The fact that Val-tRNA "sticked" exactly to the (CVN...MMERVVE) - encoding region in the prion gene when two sequences (Val-tRNA and prion gene) were aligned allows to suggest that tRNA-like elements encoded in genes might serve as potential markers for prion-like properties of encoded proteins.

Then multiple sequence alignment (EMMA program) of the Val-tRNA, (TRP...QAH)-encoding region of the HIV-1 *env* gene, (CVN...MMERVVE)-encoding region of the prion gene, and the rest part of the prion gene has been performed. The results are shown in (Box 7; Figure 3).

By working with a prion gene by a similar approach as in the case of the obtaining 16 regions of HIV-1 (please see part "Multiple sequence alignment of the 9 regions of HIV-1 genome and the prion gene"), and then doing multiple sequence alignment from these sequences (parts of the prion gene) and the Val-tRNA there were the following results (Box 7; Figures 14-15). When using default parameters in EMMA [26] this alignment (Box 7; Figure 14) looks particularly interesting in terms of regions of sequence similarity.

Pairwise sequence alignment (EMMA) of the Val-tRNA and the region 12 of the HIV-1 genome has been done using: a) default parameters (GOP 10; GEP 5), results in (Box 7; Figure 4); b) GOP 40; GEP 10, results in (Box 7; Figure 5). We can see that the region, where Val-tRNA “sticks” to the HIV-1 region of interest (region 12), shifts significantly to the left upon an increase in gap penalties values.

It was of interest to align tRNA and the genome of the smallest viroid (Cadang-cadang coconut viroid). Results are presented in (Box 7; Figures 6-9). In (Figure 6) we can see that Val-tRNA “sticked” to the right part (or the “bottom” part) of the viroid genome. In (Figure 7) we can see the alignment of the “upper” part of the cadang-cadang coconut viroid with the Val-tRNA. In (Figure 8) it is interesting to see that when using two tRNAs (Val-tRNA and Leu-tRNA) for multiple sequence alignment with a viroid genome, the alignment becomes more informative because at particular regions viroid is more similar to Val-tRNA and at other regions to Leu-tRNA. In general we can predict that the more representatives from the tRNA family we will take for the alignment the more informative will be the alignment in terms of sequence conservation along the sequence. If we take the part of the cadang-cadang coconut viroid which has not been aligned in previous case (Figure 8) and align it with Val-tRNA and Leu-tRNA this leads to a nice alignment (Figure 9).

Protein p53 is a prion-like protein [51, 52], and it has been taken for alignment with the same couple of tRNAs: Val-tRNA and Leu-tRNA. Results are presented in the (Box 7; Figure 10). The same approach has been used to align prion protein with two tRNAs (Box 7; Figure 12). It is nice to see that in comparison to the case of the pairwise sequence alignment between prion and Val-tRNA (Box 7; Figure 1) the region where tRNAs “stick” to the prion gene shifts to the left approximately on the one size of the tRNA (Box 7; Figure 12).

Then the region 12 of the HIV-1 genome has been aligned with Val-tRNA and Leu-tRNA. Results (Box 7; Figure 11) show that two tRNAs “stick” to the HIV-1 just at the beginning of the region of interest (region 12). This gives a hint that probably it is worth while playing with tRNA in “children’s construction kit”[46] and just for the moment to consider HIV-1 as a set of the encoded tRNAs. For this bioinformatics experiment sequences comprising (Val-tRNA) eleven times in a row; (His-tRNA) thirteen times in a row and (Tyr-tRNA) thirteen times in a row have been prepared. Multiple sequence alignment results after the use of such constructs were beyond all the expectations (Box 7; Figure 13) and we can see numerous sequence similarities.

6) BLAST searches

BLASTp, PSI-BLAST, BLASTn and tBLASTx programs were used to do the BLAST searches [29-31] through the NCBI website [39] to find sequence similarity between prion (protein or gene) and viral protein or genes.

BLASTp (BLASTP 2.2.17) searches (query: prion protein; and in further searches there will be only prion protein or prion gene as a query) limited to “viruses; taxid:10239” did not give any significant matches when the parameters were “Default”(i.e. matrix BLOSUM 62; gap penalties: existence: 11, extension: 1). When using BLASTp with PAM 30 matrix (gap penalties: existence: 9, extension: 1), there was one significant match (from *Acanthamoeba polyphaga* mimivirus; Expect = 3.8).

With other gap penalties (existence: 7, extension: 2) when using the same matrix PAM 30 there were several more matches and what is particular interesting that in the case of the protein UL48 of the human herpesvirus 5 there was a match at very well known to us region (... MERVVE) of the prion protein:

```
>gi|44903274|gb|AAS48953.1| UL48 [Human herpesvirus 5]  
Length=2242
```

Score = 31.8 bits (66), Expect = 5.2, Method: Composition-based stats.
Identities = 9/11 (81%), Positives = 10/11 (90%), Gaps = 0/11 (0%)

```
Query 202 DVKMMERVVEQ 212
      D+K MERVVEQ
Sbjct 648 DLKQMERVVEQ 658
```

When doing BLASTp (default parameters; BLOSUM 62; gap penalties: existence: 11, extension: 1) search limited to Retro-transcribing viruses (taxid:35268) there was one match (Atlantic salmon swim bladder sarcoma virus; Expect =5.7).

Running PSI-BLAST (default parameters; BLOSUM 62; viruses; taxid:10239) with 4 iterations led to the 13 sequences, six of which had E-value less than 1.

When doing PSI-BLAST search limited to Retro-transcribing viruses (taxid:35268) after the third iteration there were just two matches (both had E-values higher than 1).

When using BLASTn program (default parameters: matrix: match:2; mismatch: -3; gap penalties: existence: 5, extension: 2; database nr/nt; viruses: taxid:10239; algorithm for “somewhat similar sequences”) there were 8 significant matches (with E-value less than 10).

When using the same parameters as in the previous case but running it against Retro-transcribing viruses (taxid:35268), there were 4 significant matches (with E-value less than 10). Among these sequences there was a nice match (100% identity) in the *nef* gene of HIV-1:

```
gb|DQ367330.1| HIV-1 isolate KSM4039 from Kenya nef protein (nef) gene,
partial cds
Length=615
```

Score = 33.7 bits (36), Expect = 8.4
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Minus

```
Query 731 TCTCTTTCCTCATCTTCC 748
      |||
Sbjct 536 TCTCTTTCCTCATCTTCC 519
```

Interestingly, neither “megablast” (which searches for highly similar sequences) nor “discontiguous megablast” (which searches for more dissimilar sequences) algorithms when selected as options did not give any significant hits for the same pool of sequences (Retro-transcribing viruses, taxid:35268).

It was very nice to reveal that when using BLASTn parameters (algorithm for “somewhat similar sequences”; Retro-transcribing viruses, taxid:35268; blastn matrix: match:1 and mismatch:-3; gap penalties: existence: 5, extension: 2) there were only two significant matches; and those matches were for the *nef* and *vpr* genes of HIV-1 with sequence identity 100% :

```
gb|DQ367330.1| HIV-1 isolate KSM4039 from Kenya nef protein (nef) gene,
partial cds
Length=615
```

Score = 36.2 bits (18), Expect = 1.6
Identities = 18/18 (100%), Gaps = 0/18 (0%)
Strand=Plus/Minus

```
Query 731 TCTCTTTCCTCATCTTCC 748
      |||
Sbjct 536 TCTCTTTCCTCATCTTCC 519
```

```
>gb|EF567320.1| HIV-1 isolate HOMER_HIV_VPR_1891 from Canada vpr protein
(vpr)
gene, partial cds
Length=288
```

```
Score = 34.2 bits (17), Expect = 6.3
Identities = 17/17 (100%), Gaps = 0/17 (0%)
Strand=Plus/Minus
```

```
Query 26 TGGTTCTCTTTGTGGCC 42
      |||
Sbjct 42 TGGTTCTCTTTGTGGCC 26
```

Interestingly, when increasing the “words size” either to 15 or decreasing it to 7, the program gives the same results as in previous case.

When using BLASTn parameters (algorithm for “somewhat similar sequences”; Retro-transcribing viruses, taxid:35268; blastn matrix: match:1 and mismatch:-2; gap penalties: existence: 1, extension: 1) there were three significant matches two of which refer to the integration site of HTLV-1 virus and one is located in the *vpu/env* region of the HIV-1:

```
>gb|DQ339443.1| HIV-1 isolate 04CMU2-1463 from South Korea vpu protein gene,
complete cds; and envelope glycoprotein gene, partial cds
Length=1548
```

```
Score = 34.8 bits (20), Expect = 3.6
Identities = 34/43 (79%), Gaps = 8/43 (18%)
Strand=Plus/Minus
```

```
Query 663 CCTCTCCACC--TG-TGA----TCCTCCTGATCTCTTTC-CTC 697
      |||
Sbjct 210 CCTCTCCACCAGTGCTGACAATTCCCTCCTGATCTCCTTCACTC 168
```

Interestingly, when using the same parameters but running the BLASTn program against “viruses, taxid:10239” : (algorithm for “somewhat similar sequences”; viruses, taxid:10239 ; blastn matrix: match:1 and mismatch:-2; gap penalties: existence: 1, extension: 1), there were 4 significant matches one of which had E-value 0.001, length 57 nucleotides and sequence identity 82%:

```
>gb|AF083424.1|AF083424 Ateline herpesvirus 3 complete genome
Length=108409
```

```
Sort alignments for this subject sequence by: E value Score Percent
identity
```

```
Query start position Subject start position
```

```
Score = 48.0 bits (28), Expect = 0.001
Identities = 47/57 (82%), Gaps = 7/57 (12%)
Strand=Plus/Plus
```

```
Query 654 TCCTCTTC-TCCTCT-CCACCTGTGATCCTCCTGATC-TC-T-TTCCTCATCTTCCT 705
      |||
Sbjct 63075 TCCTCTTCCTCCTCTTCCTCCTCT--TCCTCCTCATCCTCCTCTTCCTCATCTTCCT 63129
```

Score = 43.1 bits (25), Expect = 0.040
 Identities = 46/57 (80%), Gaps = 7/57 (12%)
 Strand=Plus/Plus

```
Query 654 TCCTCTTC-TCCTCT-CCACCTGTGATCCTCCTGATC-TC-T-TTCCTCATCTTCCT 705
      ||||| ||||| || ||| ||||| || || | |||||
Sbjct 63084 TCCTCTTCCTCCTCTTCCTCCTC--ATCCTCCTCTTCCTCATCTTCCTCATCTTCCT 63138
```

Also all the possible combinations of parameters (scores for matches and mismatches; gap penalties) were tested in the BLASTn run against Retro-transcribing viruses, taxid:35 under the use of the same algorithm “somewhat similar sequences”. There were no new matches for HIV-1 genes found (except for those mentioned above) when applying all these different combinations of parameters.

And finally when running tBLASTx program (query: prion protein; matrix: PAM 30) on the set of Retro-transcribing viruses (taxid:35268) there were 26 significant matches most of which refer to HIV-1.

To sum up the most exciting result from the BLAST searches with a prion gene as a query sequence was that for HIV-1 BLASTn revealed three significant hits located in *nef*, *vpr* genes and the *vpu/env* region of the HIV-1 with a length 18(100% sequence identity), 17(100% sequence identity) and 43(79% sequence identity) nucleotides respectively. This gives a clue that doing BLASTn (and trying different program parameters) on different human prion genes may potentially lead to more “100%” hits, - “footprints” of prions on the HIV-1 genome.

7) Four Squares Code

The genetic code was deciphered, but still there are some secrets. Interestingly, what would be the scenario on how codons were populated by amino acids during the evolution?

Considering the standard genetic code let us group codons according to the number of hydrogen bonds in the first and the second positions of the codon upon codon-anticodon pairing. A or U (designated as W) will form two hydrogen bonds, while G or C (designated as S) will form three ones. We obtained the table with four squares (Figure 16). If we look to the Figure 16 we can see some trends in which amino acids occupy each square but among the amino acids with similar chemical structures especially striking is the case of Serine, Threonine and Cysteine located in the same square. Intriguingly, the proposed table predicts the location of Selenocysteine and Pyrrolysine (encoded by stop codons) as the same squares as Cysteine and Lysine, respectively. Although it is not necessary, it is nonetheless a special property of this table that within each square amino acids can be paired according to Root-Bernstein’s principle of amino acid pairing [53-55]. It can be noted that comparative analysis of different genetic codes [7] supports the grouping of amino acids into four squares code.

Now let us consider pair GC with three hydrogen bonds. If we close the third hydrogen bond (exactly the same bond which Watson and Crick omitted in their classical article [56, 57]) with a hand, we will see that upon pairing the G will resemble T, and C will resemble A. In other words the feature of DNA that TA pair in a way is already “encoded” within the GC pair might have provided the ancient genetic code with some flexibility (the reader may track it by moving between the four squares keeping in mind the similarity between G and T; A and C). It should be noted that upon the change of G to T, T to G, A to C, C to A in the first and the second positions of the codon all the possible changes will divide amino acids into 4 groups: 1) Phe/Leu, Val, Gly, Cys/Trp 2) Ser, Ala, Asp/Glu, Tyr 3) Pro, Thr, Asn/Lys, His/Gln 4) Leu, Ile/Met, Ser/Arg, Arg. It should be noted that within each of the 4 groups the pattern of changes in the chemical

structures of the amino acids upon the changes of nucleotides in respective codons has been revealed (data not shown). But more analysis in this direction is required.

The information from the four squares code might be used in the construction of amino acid scoring matrixes for multiple protein sequence alignment. Hopefully these new amino acid scoring matrixes based on the four squares code might be useful for aligning protein sequences with low sequence similarity. Indeed, proteins can be evolutionary related to each other (this in turn can be reflected in their similar structures and / or similarities on the genetic level) and have simultaneously low sequence similarity, but the use of the standard amino scoring matrixes will never lead to such a conclusion about evolutionary relation between the proteins and, moreover, might mislead the alignment process. The significance of the results from the alignment based on the amino acid scoring matrixes constructed from the four squares code can be verified by the method of alignment of protein interaction networks [59, 60] (please see below in the discussion section).

While the genetic code provides information about which codon encodes particular amino acid, the four squares code might highlight the inherent relationship between amino acids and potentially provide some clues to protein 3D structure in terms of both intramolecular and intermolecular interactions between amino acid residues.

There are some interesting cases when one polypeptide chain may acquire two different structures (eg prion), or polypeptide chains from two different proteins with low sequence identity (eg HIV-1 gp120 protein and prion) may have similar structures [6]. Four squares code is expected to give some clues to these phenomena which are important for creation of an anti-HIV-1 vaccine based on viral fibrillization [1, 2].

Discussion

The results were summarized at the end of the “Introduction” part of the dissertation. Let us consider the future of the approaches listed above, some limitations and some open questions to be resolved with a help of bioinformatics.

It seems that in the future all of the methods (pairwise/multiple sequence alignment of DNA/protein sequences; structural alignment of proteins) will be used to do the comprehensive analysis of the viruses on the presence of the prion-like proteins. If to construct bioinformatics workflows and to try to do the process as automatic as possible it might be possible in the future to analyze the very big sets of viruses (even all of them) on the presence of the prion-like proteins by pressing just one button “start search” in the program on the computer. These programs and workflows is a challenging but an interesting task. But with every year it will be more and more simple to build workflows, and finally at some point the limiting stage might be the strategy/methodology behind such workflows. In this dissertation the first step towards the methodology of such a comprehensive analysis has been done. Indeed many of the approaches tested and mentioned above might in the future constitute small parts, tasks in such a bioinformatics workflow to find prion-like proteins in viruses.

Many questions remain unanswered about the prion-like proteins. According to the article [1] prion-like proteins are defined as proteins having “either sequence homology or structural-functional similarities, or both with prion”. Interestingly, what will be the exact criteria to say that particular protein is a prion-like protein? Is it possible to narrow the definition of the prion-like proteins given in the article [1]? Is it possible to develop purely bioinformatics criteria for the protein to be a prion-like? If the bioinformatics criteria have been developed what would be the experimental/biological tests to prove that the protein is a prion-like? Or the protein can be prion-like even in the case when its biological function is not obvious? As we can see it is not necessary for the protein to have sequence homology with a prion to be a prion-like. There are

examples when the prion-like proteins might have low sequence identity with a prion protein, but have structural similarity with it [6].

Let us introduce the concept of the “vertical” and “horizontal” bioinformatics. If we want to talk just about sequence similarity/homology it would be a task for “vertical” bioinformatics: in the vertical columns of the multiple sequence alignment we try to find the same or similar amino acid residues. “Vertical” bioinformatics cannot explain why proteins with low sequence similarity to the prion protein might have structural similarity with a prion and might undergo similar structural α/β -switch. If we want to talk about an interaction of amino acid residues along the polypeptide chain upon conformational changes in the prion-like proteins we can suppose that the information about such a conformational change (eg α/β -switch) might be written in the positions of the amino residues along the polypeptide chain. So we will be “aligning” amino acid residues which interact with each other in different conformations of the prion-like molecule and are located in different positions of the same polypeptide chain. To track all these amino acids interactions along the polypeptide chain and to determine the patterns of what kind of amino acids can interact with each other in the prion-like molecules would be the task for “horizontal” bioinformatics.

Four Squares Code is expected to help in understanding the nature of prions/prion-like proteins and might shed some light on the “horizontal” bioinformatics. The concept of the tRNA elements in genes as hallmarks of the prion-like proteins might have an interplay (please see Section 5 in “Results” part) with the concept of “horizontal” bioinformatics and the Four Squares Code. But further analysis of the prion sequences is required to support such a hypothesis.

It should be noted that bioinformatics analysis (please see Sections 5 and 7 in “Results” part) is required to support the grouping of amino acids into four squares code.

RepeatFinder [58] software might be used to find repeats of tRNA-like elements in the viral genomes (Dr.A.R.Dalby, personal communication). Such an analysis can support the hypothesis (about tRNA-like elements in genes as hallmarks of prion-like proteins) mentioned above. There should be strong criteria developed for such an analysis so that not to come to the conclusion that every protein is a prion-like or every protein is encoded by the repetitive tRNA-like sequences in DNA.

Hidden Markov Models (HMMs) for prion protein motifs might be constructed and used to find prion-like proteins (Dr.A.R.Dalby, personal communication).

At the ISMB/ECCB conference in Vienna this year (2007), there was a poster “Detecting Functional and Evolutionary Relationship by Aligning Protein Interaction Networks” by Michal Kolar, Michael Lassig and Johannes Berg. This poster represents a new approach to analyze evolutionary and functional relevance between the proteins which have low level of sequence identity. Indeed, very often it is difficult to make a conclusion about the evolutionary relation between the proteins based only on sequence comparison data. Authors’ idea is to align protein interaction networks to highlight functional similarity between the proteins, and this in turn will complement the data on sequence similarity. The application of the method [59, 60] to analysis of the proteins of herpesviruses was the topic of the poster.

The idea is that such an approach of aligning protein interaction networks can be applied to the analysis of the viruses on the presence of the prion-like proteins. Together with other approaches mentioned above this would provide a nice framework for the comprehensive bioinformatics analysis of viral genomes, genes and encoded proteins.

The final goal of such a bioinformatics analysis will be to support the “Virus-Prion” theory [1] for better understanding of the nature and the structure of the virus, and this in turn will help to create an anti-HIV-1 vaccine based on viral fibrillization [1, 2].

References:

- [1] Suslov KV. Towards an anti-(R5-X4 HIV-1 switching) vaccine based on prions. In: B.A. Veskler, Editor. *New Research on Immunology*. Nova Science Publishers. New York. pp.1-46. (2005).
- [2] Suslov KV. Fibrillizing HIV-1 naturally. *Med Hypotheses* (2007); 69:462-463.
- [3] O.Bosu, S.K. Thukral. *Bioinformatics. Databases, Tools, and Algorithms*. Oxford University Press, 2007.
- [4] O.Lund, M.Nielsen, C.Lundegaard, C.Kesmir, S.Brunak. *Immunological Bioinformatics*. MIT Press, 2005.
- [5] Sáez-Cirión A, Nieva JL, Gallaher WR. The hydrophobic internal region of bovine prion protein shares structural and functional properties with HIV type 1 fusion peptide. *AIDS Res Hum Retroviruses*.(2003);19(11):969-78.
- [6] Mahfoud R, Garmy N, Maresca M, Yahi N, Puigserver A, Fantini J. Identification of a common sphingolipid-binding domain in Alzheimer, prion, and HIV-1 proteins. *J Biol Chem*. (2002); 277(13):11292-6.
- [7] <http://www.ncbi.nlm.nih.gov>
- [8] Gabus C, Auxilien S, Péchoux C, Dormont D, Swietnicki W, Morillas M, Surewicz W, Nandi P, Darlix JL. The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. *J Mol Biol*. (2001);307(4):1011-21.
- [9] Gabus C, Derrington E, Leblanc P, Chnaiderman J, Dormont D, Swietnicki W, Morillas M, Surewicz WK, Marc D, Nandi P, Darlix JL. The prion protein has RNA binding and chaperoning properties characteristic of nucleocapsid protein NCP7 of HIV-1. *J Biol Chem*.(2001);276(22):19301-9.
- [10] Gordon LM, Mobley PW, Lee W, Eskandari S, Kaznessis YN, Sherman MA, Waring AJ. Conformational mapping of the N-terminal peptide of HIV-1 gp41 in lipid detergent and aqueous environments using ¹³C-enhanced Fourier transform infrared spectroscopy. *Protein Sci*. (2004);13(4):1012-30.
- [11] Crescenzi O, Tomaselli S, Guerrini R, Salvadori S, D'Ursi AM, Temussi PA, Picone D. Solution structure of the Alzheimer amyloid beta-peptide (1-42) in an apolar microenvironment. Similarity with a virus fusion domain. *Eur J Biochem*. (2002);269(22):5642-8.
- [12] Pécheur EI, Martin I, Bienvenüe A, Ruyschaert JM, Hoekstra D. Protein-induced fusion can be modulated by target membrane lipids through a structural switch at the level of the fusion peptide. *J Biol Chem*. (2000);275(6):3936-42.
- [13] Reed J, Kinzel V. A conformational switch is associated with receptor affinity in peptides derived from the CD4-binding domain of gp120 from HIV I. *Biochemistry*. (1991); 30(18):4521-8.
- [14] Reed J, Kinzel V. Primary structure elements responsible for the conformational switch in the envelope glycoprotein gp120 from human immunodeficiency virus type 1: LPCR is a motif governing folding. *Proc Natl Acad Sci U S A*. (1993);90(14):6761-5
- [15] Leblanc P, Baas D, Darlix JL. Analysis of the interactions between HIV-1 and the cellular prion protein in a human cell line. *J Mol Biol*. (2004); 337(4):1035-51.
- [16] McBride SM. Prion protein: a pattern recognition receptor for viral components and uric acid responsible for the induction of innate and adaptive immunity. *Med Hypotheses*.(2005); 65(3):570-7.
- [17] Lupi O, Dadalti P, Cruz E, Goodheart C. Did the first virus self-assemble from self-replicating prion proteins and RNA? *Med Hypotheses*. (2007); 69(4):724-30.
- [18] Ledford H. Virus paper reignites prion spat. *Nature* (2007);445(7128):575

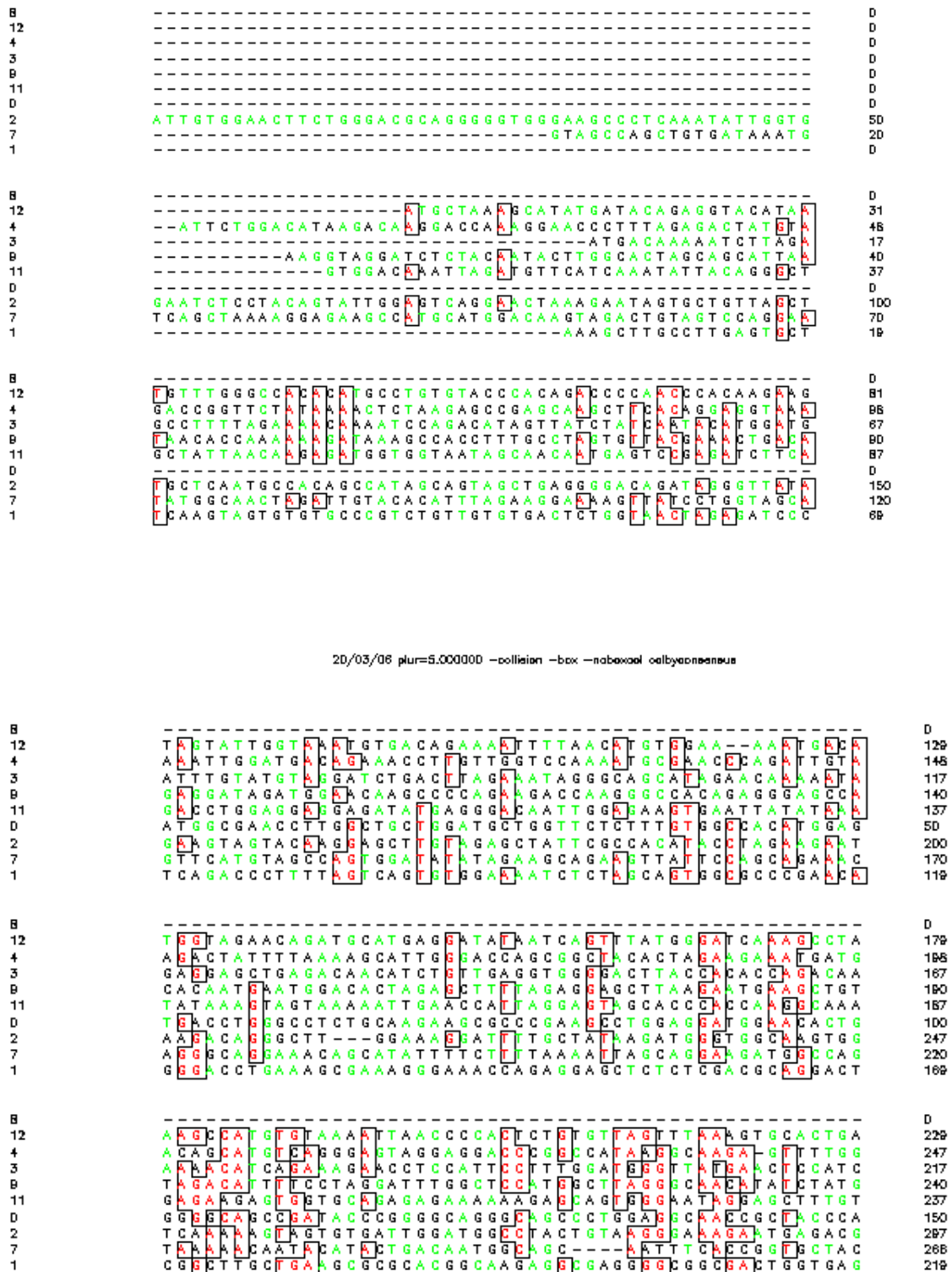
- [19] Manuelidis L, Yu ZX, Barquero N, Mullins B. Cells infected with scrapie and Creutzfeldt-Jakob disease agents produce intracellular 25-nm virus-like particles. *Proc Natl Acad Sci U S A.* (2007);104(6):1965-70.
- [20] Haseltine WA, Patarca R. AIDS virus and scrapie agent share protein. *Nature* (1986); 323: 115-6.
- [21] Braun MJ, Gonda MA. Is scrapie Prp 27-30 related to AIDS virus? *Nature* (1987); 325: 113-114.
- [22] Bazan JF, Fletterick RJ, Prusiner SB. AIDS virus and scrapie protein genes. *Nature* (1987); 325: 581.
- [23] Braun MJ, Gonda MA, George DG, Bazan JF, Fletterick RJ, Prusiner SB. The burden of proof in linking AIDS to scrapie. *Nature* (1987); 330: 525-526.
- [24] Patarca R, Haseltine WA, Webster T, Smith TF. Of how great significance? *Nature* (1987); 326: 749.
- [25] Kuznetsov IB, Rackovsky S. Similarity between the C-terminal domain of the prion protein and chimpanzee cytomegalovirus glycoprotein UL9. *Protein Eng.* (2003);16:861-3.
- [26] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* (1994); 22(22):4673-80.
<http://bioweb.pasteur.fr/docs/EMBOSS/emma.html>
- [27] Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* (2002);30(14):3059-66.
- [28] Schwartz AS, Pachter L. Multiple alignment by sequence annealing. *Bioinformatics.* (2007); 23(2):e24-9.
AMAP website: <http://baboon.math.berkeley.edu/amap/>
- [29] <http://www.ncbi.nlm.nih.gov/blast/>
- [30] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* (1990);215(3):403-10.
- [31] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* (1997); 25(17):3389-402.
- [32] Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Engineering* (1998); 11(9): 739-747.
- [33] CE website: <http://cl.sdsc.edu/ce.html>
- [34] TM-Align website: <http://zhang.bioinformatics.ku.edu/TM-align/>
- [35] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* (2005); 33(7):2302-9.
- [36] PrettyPlot: <http://bioweb.pasteur.fr/docs/EMBOSS/prettyplot.html>
<http://bioweb.pasteur.fr/seqanal/interfaces/prettyplot.html>
- [37] Clamp M, Cuff J, Searle SM, Barton GJ. The Jalview Java alignment editor. *Bioinformatics.* (2004);20(3):426-7.
- [38] Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* (1997); 18(15):2714-23.
- [39] NCBI website: <http://www.ncbi.nlm.nih.gov>
- [40] Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* (2007); 35(Database issue):D5-12

- [41] CBRG website: http://www.compbio.ox.ac.uk/CBRG_home.shtml
- [42] MPI Bioinformatics Toolkit website: <http://toolkit.tuebingen.mpg.de/>
- [43] Biegert A, Mayer C, Remmert M, Söding J, Lupas AN. The MPI Bioinformatics Toolkit for protein sequence analysis. *Nucleic Acids Res.* (2006);34(Web Server issue):W335-9.
- [44] PDB website: <http://www.rcsb.org/pdb/home/home.do>
- [45] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* (2000);28(1):235-42
- [46] Niller HH, Salamon D, Rahmann S, Ilg K, Koroknai A, Bánáti F, Schwarzmann F, Wolf H, Minárovits J. A 30 kb region of the Epstein-Barr virus genome is colinear with the rearranged human immunoglobulin gene loci: implications for a "ping-pong evolution" model for persisting viruses and their hosts. A review. *Acta Microbiol Immunol Hung.* (2004);51(4):469-84.
- [47] Weiner AM, Maizels N. The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication, and clues regarding the origin of protein synthesis. *Biol Bull.* (1999);196(3):327-8
- [48] Maizels N, Weiner AM, Yue D, Shi PY. New evidence for the genomic tag hypothesis: archaeal CCA-adding enzymes and tDNA substrates. *Biol Bull.* (1999);196(3):331-3
- [49] Metlas R, Veljkovic V. Does the HIV-1 manipulate immune network via gp120 immunoglobulin-like domain involving V3 loop? *Vaccine* (1995);13(4):355-9.
- [50] Suslov KV. Does AID aid AIDS? *Immunol Lett.* (2004); 91(1):1-2.
- [51] Ishimaru D, Andrade LR, Teixeira LS, Quesado PA, Maiolino LM, Lopez PM, Cordeiro Y, Costa LT, Heckl WM, Weissmüller G, Foguel D, Silva JL. Fibrillar aggregates of the tumor suppressor p53 core domain. *Biochemistry.* (2003) ; 42(30):9022-7.
- [52] Blagosklonny MV. p53 from complexity to simplicity: mutant p53 stabilization, gain-of-function, and dominant-negative effect. *FASEB J.* (2000); 14(13):1901-7.
- [53] Root-Bernstein RS. Amino acid pairing. *J Theor Biol.* (1982);94(4):885-94
- [54] Root-Bernstein RS. On the origin of the genetic code. *J Theor Biol.* (1982);94(4):895-904.
- [55] Root-Bernstein RS. Protein replication by amino acid pairing. *J Theor Biol.* (1983); 100(1):99-106.
- [56] Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* (1953); 171(4356):737-8.
- [57] Wain-Hobson S. The third Bond. *Nature* (2006); 439(7076):539.
- [58] <http://www.cbcb.umd.edu/software/RepeatFinder/>
- [59] Berg J, Lässig M. Local graph alignment and motif search in biological networks. *Proc Natl Acad Sci U S A.* (2004);101(41):14689-94.
- [60] Berg J, Lässig M. Cross-species analysis of biological networks by Bayesian alignment. *Proc Natl Acad Sci U S A.* (2006);103(29):10967-72.
- [61] <http://bioweb.pasteur.fr/seqanal/interfaces/shuffleseq.html>

Figures

Box 1. Multiple sequence alignment (EMMA) of the 9 regions (sequences 1, 2, 3, 4, 7, 8, 9, 11, 12) of HIV-1 genome and prion gene (sequence 0).

/proj/wabapp-data/production/1003887/.sequences



B		D
12	T T T G A A G A T S A T A C T A F A C A A T K S T A G T S C G G S A S K A T G A T A A T G S	279
4	C T G A A G C A A T S A G C C A A S F A A C A A A T T C A G C T A C C A T A T G A T F G C A G A G A	287
3	C T G A A A A T G S A C A G T A C A G C C T A T K G T G C T - G C C A G A A K A K A G A C A G C T S	286
9	A A A C T T A T G G S G A T A C T T G G S C A G G A K G T G G A S C C A F A A T A K A A A T T C T S	280
11	T C C T T G S T T C T T G G G K S C A G C A G G A K G C A C T A T G S C G C A G C C T C A A T S	287
0	C C T C A G S S C G S T G G T G G C T S S G G C K G C C T C T G G T S T G G C T G G G G G C A	200
2	A G C T G A S C C A S C A G C A G A T A G S G T G S A G C A G C A T C T C G S S C T G G A A A	347
7	G G T T A G S S C C S C C T G T T G T T S G C G G S A A T C A G C G A A T T T G G A A T T C	316
1	T A C G C C A A A A T T T T G K C F A G C S G A S S C T A G A G G A G A S K T F G G G T G C S	289

B		
12	-----A G A K C C K T A G F A G G A G C A G A A C C T T C T A F G T	33
4	A G A A A G G - - A G A G A T A A A A A C T S E T C T T C A A T A T C A S C A A A G C A T A A	327
3	G G C A A T T T T A G G A A C C A A A G A A A S A T T G T T A A S T G T T C A A T T S T G G C A A	347
9	G A C T G T C - - - A A T G A C A T A C A S A G T T A G T G G - - G S A A T T S A T T F G S	309
11	C A A C A A C - - - T G C T G T T T A T C K A T T T T C A G A A T T G S G T G T C A C A T F A G	335
0	A C G C T G A C - - - G G T A C K G S C A S A A A T A T T G T C G G T A T A S T G C A S C	333
2	G C C T C A T G G T G G T G G C T G S S G G C A G C C C A T G G T G G F G G C T G G S S A C S C	250
7	A A C A T G G A G C A A T C A C A K G T K S C A K T C A G C S C T A C C A A T G C T A T F G T	387
1	C C T C A A A T C C C C A A A G T C A A G G A S T A G T L G A A T C T A F G A A A A G A A T F A	366
	A G A G C G T C A G T A T T A G C S S G G G A G A A T K A G T C G A F G G S A A A A A T T C G	319

B		
12	A G A T S G S S C K S C T A A C A G G S K S A C T K A A - T T A S S A A A G C A G S A T A T G T T	82
4	G A G G T A A G G T G C A S A A G A A T A T S C A T T - T T T T A A A A C T T S A T A T A A T	376
3	A G A G G S C A C A C A S C C A G A A A T T B C A A S G C C C C T A G G A A A A S S G C T G T T	387
9	G C A A G T C A G S T T T A C C A S S G A T A A K G T A A G E A A T T A T G T A K C T C C T	359
11	C A G A T A G G C G T T A C T C S K C K S A G G K S K G C A A S A A T T G G A G C C K S T A S A T	365
0	A G C A G A C A C A T T T S C T G A S S G C T A T S A G S G C C A C A G C A T C T G T G C A A	363
2	C T C A C A G T S T G T G G C T - - - S S G A S G G T S C C C C A C A S T C A T C T T G	296
7	G C C T S G C T A G A A G C A C A S G A G G A G G S A G G T S G T T T C C A S T C A C C C	447
1	A A C A A A T T A T A A G S K C A G G T A A G A G K T C A G C T G A A C A T C T A K A C A C C	416
	G T T A A S C C K S G G S S A K A S A A A A T A T A A A T T A A A C A T A A G T A T G S G	389

B		
12	A C T A T A S A G G A L G A C A - - - - - K A K A B T T G T C S E C T A K C T G A C A C	123
4	A C C A T T A S A T A T S A T A E T A C A G E T A T A A G T T G A C A A G T T S A T A C C C T	426
3	G G A A T T G T G A A G S A G G A C A C C A A K T G A A G A T T G T A E T S A G A S A C A G	447
9	T A G A S G K C C A L A G C K C T A C A G A A G T A K T A C C A C T A C K G A K G A A G C A G	409
11	C C T A G - - A C T A S G C C C T G G A A G C A T C C K S S A A G T C A S C C T A K A K C T G C T	433
0	C T C A C A G T C T G G G S C K T C A K A S C A G C T C C K S S C A A G A A T C C T S G C T G T G G A	433
2	G A C A A G C C S K S T A A G C C A A A A A C C A C K A T S A - - A G C A C T G S C T G G T -	342
7	T C A G S T C C T T T A A G A C C A T G A C T K C A A G - - - C S E T G T A G A T C T T A	484
1	A G T C A A A T G G C A G T A T C A T C C A C A T T T T A - - A A A G A A A G S S G G G A	463
	C A A G C A G G A G C T A G K E S K T T C G C A G T K A T C C T G G C E T G T A G A A C A	419

B		
12	A A C K A T C A G A A G K C T G K S F T A C A A S C A A T T A T C T A S C T T T G C A G S A T T	173
4	C A G T C A T T A C A C A S G C T G T C C A A S S T A T C C T T T G A S C F A K T T C C A T A	476
3	G C T K A T T T T K G G G A A S A T C T G S C C C T T C C F A C A A G G S A A G G C C A G G S A A	487
9	A G C T A G A A C T T G G C A S A A A C A G A S A S A T C T A A A A G A A F C A S T A C A T S S A	459
11	T G T A C A A A T T S C T A T T G T A A A A A S T S T T G S C T T C A T T S C C K A - G T T T S T T	482
0	A A G A T C C T A A G G G T C A A C A G C T C C T G G G G A T T G S S S T T G C T C T G S A A	483
2	G C T S C A C A G C T S G G S C A S T G T S S S G G C C T G G C G S C T A C G T G C T S S G	382
7	G C C C T T T T A A A S A A A A S G G S A C F G G A S G G C T A A T T C A C T C C A A A	544
1	T T G G G G G T C K A G T S C A S G G G A A G A A T A G T A S A C A T A A T A S C A A C K S A C	513
	T C A G A S S G C T S T A S K C A A A T T C T S S A G C T A C A C C A T C C C T T C A S A C	469

B		
12	C G G S A T T A S A G T A A C A T K S T K A C A G K C T C A C A A T A T S C A T T A G G A A T C	223
4	C A T F A T T S T G C C C C G G C - T G G T T T T G C G A T T T A A A A T S T A A T A A T A A G A	525
3	T T T T C T T C A S A G C A S A C C A G A G C C A A C A C C C C A C C T A A G A S A G C T T C A	547
9	G T G A T T A T G A C C A T C A A K G K C T T A A T A G C A G A A T A C A S A - - - A G	504
11	T C A T A C K A A A G C C T A G G C A T C T C C T K T B G C A G G A S A A G C G G A G A C A G	532
0	A C T A A T T G A C C A C T G C T G T C C T T G G A A T G C T A S T T G G S T A A T A A A	533
2	A A G S C C A T G A G C A S G C C A T C K T A C A T T C G G C A S T G A C T A T G A G G A C C	442
7	A B A A G A C A K A A F A T C C T T G A T C T G T G S T C T A C C A C R B K C A A S G C T A C T T	584
1	A T A C A A C T A K A G A T T A C A A A C A A A T T A C A A A A T T C A A A T T T T C G	563
	A S G - A T C S A A G A K C T T A G A T C A T A T K T A A T A C A S T A G C A A C C C T C T A T	516

8 AT T C A G C C C A C C C A T C A K A G T G K T C A G A G T T K G T C A T C A A K T A A T 273
 12 C G T T C A A T G S A K C K G G C C A T G T A C K A K T G T C A G C K F G T A C A A T G T K C K 575
 4 G S T C T G G G T A G A G A C A C A C T C C C C T C A G A A G C A G S K G C C G A T A G A C 587
 3 C A G G G C A A G G C C A T T G G A C A T A T C K K K T T T A T C A A G A G C C A T T T K A A A 554
 9 C G A C G A G A G G C T C A T C A A K C K A T C K G A C C T C A T C A A G G T T C T C T A T C A A A 582
 11 T C T T G G A K C A T T T G G A A T C A C A C G A C C C T G G A T G A G T G G G A C A G A S K 583
 0 G T T A C T A T C S T G A A A A C A T G C A C C G T T A C C C A A C C A A G T G T A C T A C A S G 482
 2 C C C T G T T A G C A G A A C T A C A C C A G G G C C A G G G T C A G A T A T C C A C T G A 644
 7 G S T T A T T A G S G K C K S C A G A A T C C K C T T T G A K A G S K C C A G C K A K S C 613
 1 T S T S T G C A T C K A A G G A T A G S K A T A A K A G A C C A A G S A A G C T T T A G A C A A 566

8 A G A S C A G T F A K F A A A A S G A A K A G G T C T A F C T G G C A T S G G T A C C A S C A C 323
 12 C A T G G A A T T A G G C C A G T A S T A T C A A C T C K A C T G C T G T T A K A T G S C K S T C T 625
 4 A A G A A C T G T A T C C T T T A C T T C C C T C A G G F C A C T C T T G G C K A C G A C C C 647
 3 T C T G A A A C A G T A 583
 9 G C A T A A G T A G T A 632
 11 A A T T A C A A A T A C A C A C A C T T A T K C A C T G C T A A T F S A A K A T C S C A A 632
 0 C C A T G S A T A G T A C K C A A C C A G A A C A C K A C C T T T G T G C A C G A C T S C S T C A A 542
 2 C T T G A T S G T S C T K A K A G C T A K A C K A C C A G T G A C A G T A A G T A A G T A G A 684
 7 T C T C T S A A K G G T G A K G G G C A S T A S T K A K A C A A G - A T A A T A G T G A C A T 862
 1 G A T A G S A K G S A K G C A A A A S T A S T A A A A A A S C A C A G C A K G C A S C A S 816

8 A C - - - A A S G A A T T S S A G S A A A T G A C A K G T A G A T A A A T T A G F C A S T G C T 370
 12 A G C A - G A S A A A S A S A S T A G A T C T B T C A A T T T C A C G G A C A A T G C T 874
 4 C T C G T C A A T A 687
 3 C A T A A T G A T G - - - T A A A A K A C A T T A A C K G A G S A G T G C A A K A A T A C C 639
 9 C A T T A G T S T A S C A A T K A T A T A G C A A T K G T T S T G T G T C C A T T G T A T C 662
 11 A A C A G C A A G A A A S K T G A C A A G T A T A T S G A T T A G A T A A A T G G G C 662
 0 T A C A C A T C A A S C A G C S C A C G T C A C K C A C A C C A A G G G G G K A K A C T 582
 2 G A G G C C A A T A A S G A G A G A C C A G C T B T A C C C C T G T G A G C T G C A 744
 7 A A A G T A S T G C C C A K A G A A A S C A K A A K T C A T T A S G G A T T A T S G A A A A C 712
 1 C T G A C A C S G S C K A C A G C A A T C K S G T C S G C C A A A T T A C C T A T G T G C A G 666

8 G G A T C A G S A A S F A C F A T T T T S T S G K A T S A T A A S C C C A K S A T G A 420
 12 A K A C C A F A A T A S F A C A G C T T S A A C A C A T C T G T A S A A A F A K T F S T A C A A G 724
 4 A P G A G A G A T G A T A C A S T A T T G S A K G A A A A T G A G T T T G C C A G S A A G A T G G A 747
 3 A K S A A A G C A T S F A A T A T G G G S A - - - A A S A C T C C K A A K T F T A A C T G 684
 9 A K S A A A T A T A G G A A A A T A T T A A C A K A A A A A A T A G A C A G S T T A A T T G A 732
 11 A A G T T T G T G A A T S G T T A A C A T A K C A A A T T G S C T G S G T A T A T A A A T 732
 0 T C C C G A C C G A C G T T A A G A T S A T G G A G C G C G T G G T F G A G A G A T G T G T 642
 2 T G G A T G A T G C C C G S A G A G A G A G T G T T A S S T G A S S T T F S A C A G C C 784
 7 A S T G G C A G G T G A F S T T G F S T S S C A K G T S A C A G G A F G A G S A T T K S A A C 762
 1 A C A T C C A G G G C A K A T S G F A C A T C A S G C K T C A C C T A G A C T T T A A A 716

8 A S A T G S S A K A T C A C A S T A A T T G G K G A E C A A T S G C T A G T G A T T T F A A C C 470
 12 A K C C A A A A C A T A C K A A A A A G A F C C G A T C C A S A G K S S K C A G G G A 774
 4 A K C C A A A A T S A T S G S K A T T G G A G S T T F A T A A A G T A K A C A G A C A G T K T 787
 3 C C C A T A A A G G A K A C A T G G G A A A C A T S B T G G A C A S A G T A T T S C A A G C 734
 9 T A G A C T A A A G A K S C A G A G A C K G T S B C A A F A S A G T G A A G S A G A K A 762
 11 T A T T C T A A T S K T S A S A S A G C T T G S T A G G T T T A K A K A T G T T T T T G C 781
 0 K T C C C A S T C G L S S K A A T C T C K G S C C F L T T C C L G A S A S S A F C G A S 682
 2 G C T A G A T T C A T C A S T G G C C S A S A S C F S C A T C C G S A S F K T F C A K S 844
 7 T G G A K A S T T A G T A A A C K C A T A G S T A S T T C A G S S A A A G C T A G G G 812
 1 T G A T G G G T A A K A G T A S T A G A G A G A K A E C F T F C A G C C C A S A A G T G A T K C 766

8 T G C F A C C T S F A S T A G C A A A G A A A T A G - 486
 12 G F S K A T T T S T A C A K A S G A A A A A T A G S A A T A T G A G A C A K G C A C A T T S T 824
 4 G A T C G A T C T C A T A S A A T C T T G G A C A T A A A G T A T A G T A C A G T A T T 847
 3 C A C T G G A T F C T S A S T G G G A G T T S T T A T A C C C T C S C T T A G T G A A A T 764
 9 T A T C G C A C T F G T S A G A T G G G G G T F S S K A T T G G G C A C S F T G C F C T S G 852
 11 T G T G T T T C T A T A S T G A G T F A S S C A S G G A T A T C A C C A T A T A T C T G 831
 0 C A T G G T C C T C T C - - - T C C T C T C C A C C T S T G A T C C T C C T G A T C C T T T C C 739
 2 A K T G C T S A C T S A S C T T G C T A C A K S B A S T T T C C C T G G G A C T T T C C 884
 7 G A T G G T T A T A S A K T C A C T A T G K A A G C C C T C A T C C A S A A F A A G T T C 861
 1 C C A T G T T T C A S C A - - - T T A T C A G A A G B A S C A C C C A C A S A T T A A A C A 818

8	-----	486
12	AACATGAGTAAAGGCAAAATGSAATAA	850
4	AGTASGACCTACACTGTCAACATAATTGGAAAG	861
3	TATGSTACAGTTAGAA	803
9	GATGTTGATGATCTGTAGTCTCAGAAAAATTGTGGGTCAAGTCTATT	862
11	TTCAACCACTCCAAACCCGAGGGGACCCGACAGGCCCGAAGGAATA	861
0	TCACTCTGATAGTSGATGA	762
2	AGGGAGGCTGGCTGTGSGSGGACTGGGAGTGGCGAGCCCTCAGATCCT	844
7	AGAASTHACATCCACTAGSGGATGCTAGATTGG	866
1	CCATGCTAAACACAAGTSGSGG	839

8	-----	486
12	-----	850
4	-----	861
3	-----	803
9	ATGGGGTACCTGTGTGGAAGGA	804
11	GAGAA	867
0	-----	762
2	GCAATAAGCAGCTGCTTTTTGCTGTACTGGGTCTCTCTGGTTAGAACA	894
7	-----	866
1	-----	839

8	-----	486
12	-----	850
4	-----	861
3	-----	803
9	-----	804
11	-----	867
0	-----	762
2	GATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAAACCCACTGCTTAAGC	1044
7	-----	866
1	-----	839

8	-----	486
12	-----	850
4	-----	861
3	-----	803
9	-----	804
11	-----	867
0	-----	762
2	CTCAATAAAGCTTGCCTTGAAGTCTTC	1071
7	-----	866
1	-----	839

Box 2. Structural alignment

Figure 1. Structural alignment (CE program) of human prion protein (red) and HIV-1 Vpr protein (green).

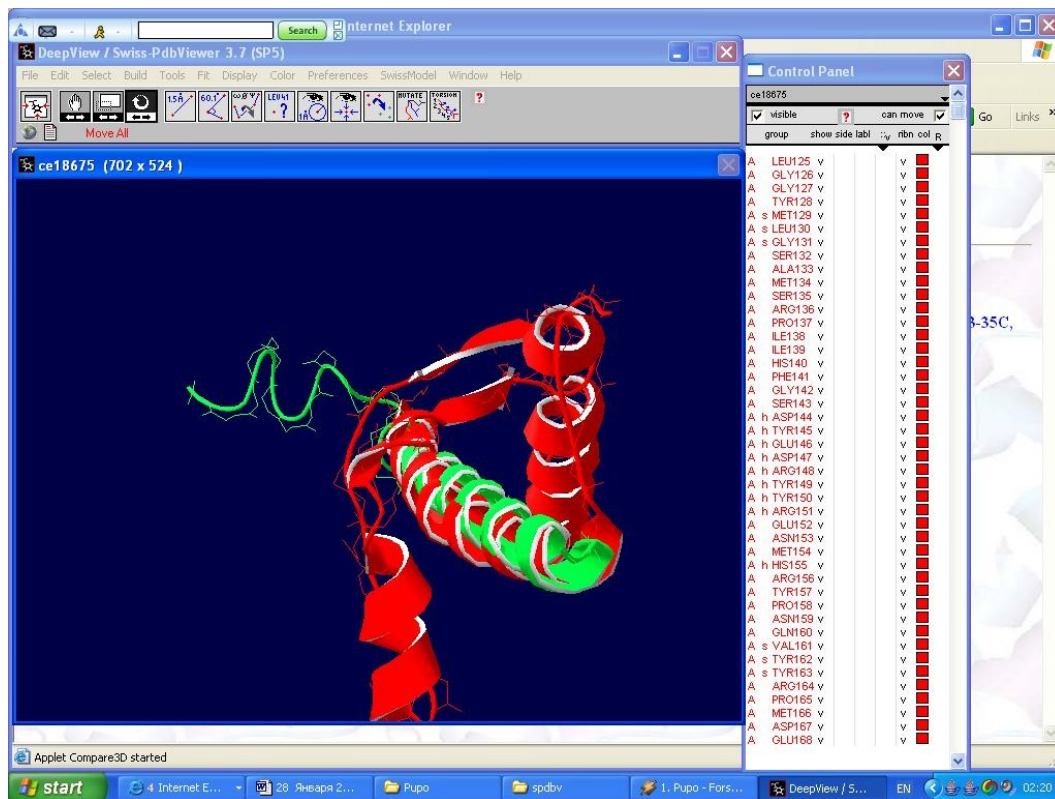


Figure 2. Structural alignment (CE program) of human prion protein (red) and HIV-1 Vpr protein (green).

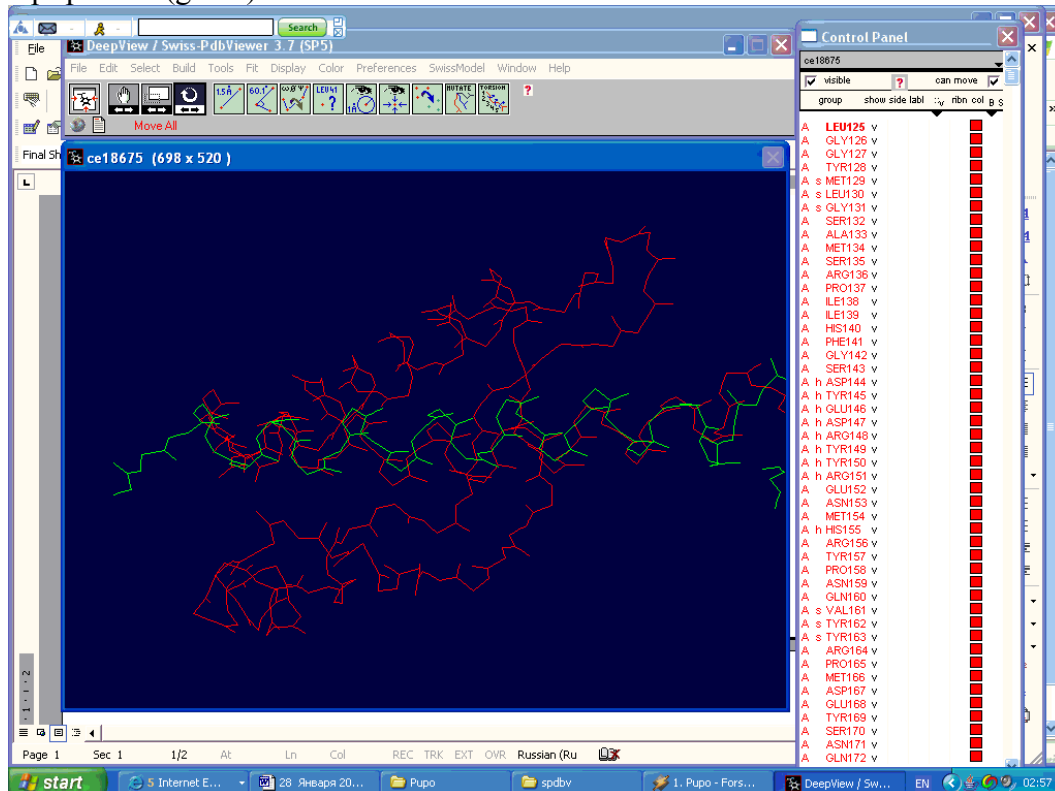


Figure 3. Structural alignment (CE program) of human prion protein (red) and HIV-1 Vpr protein (green).

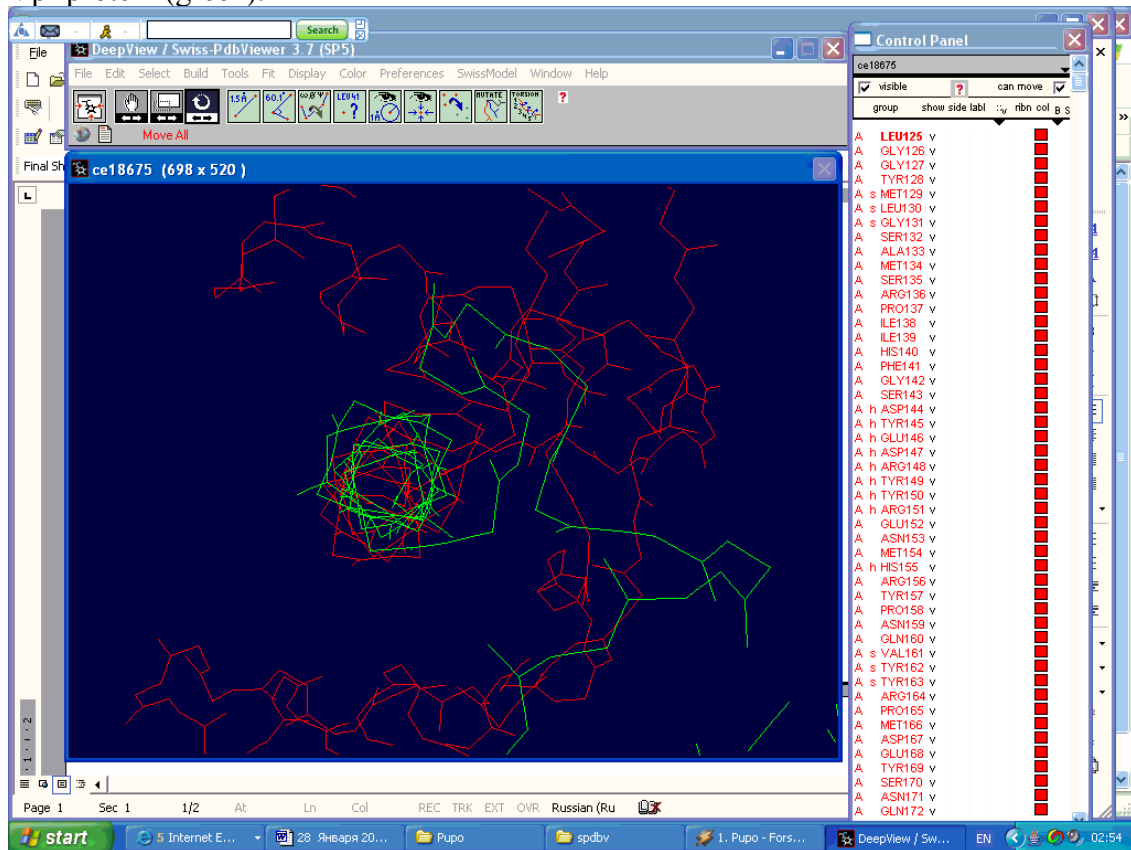
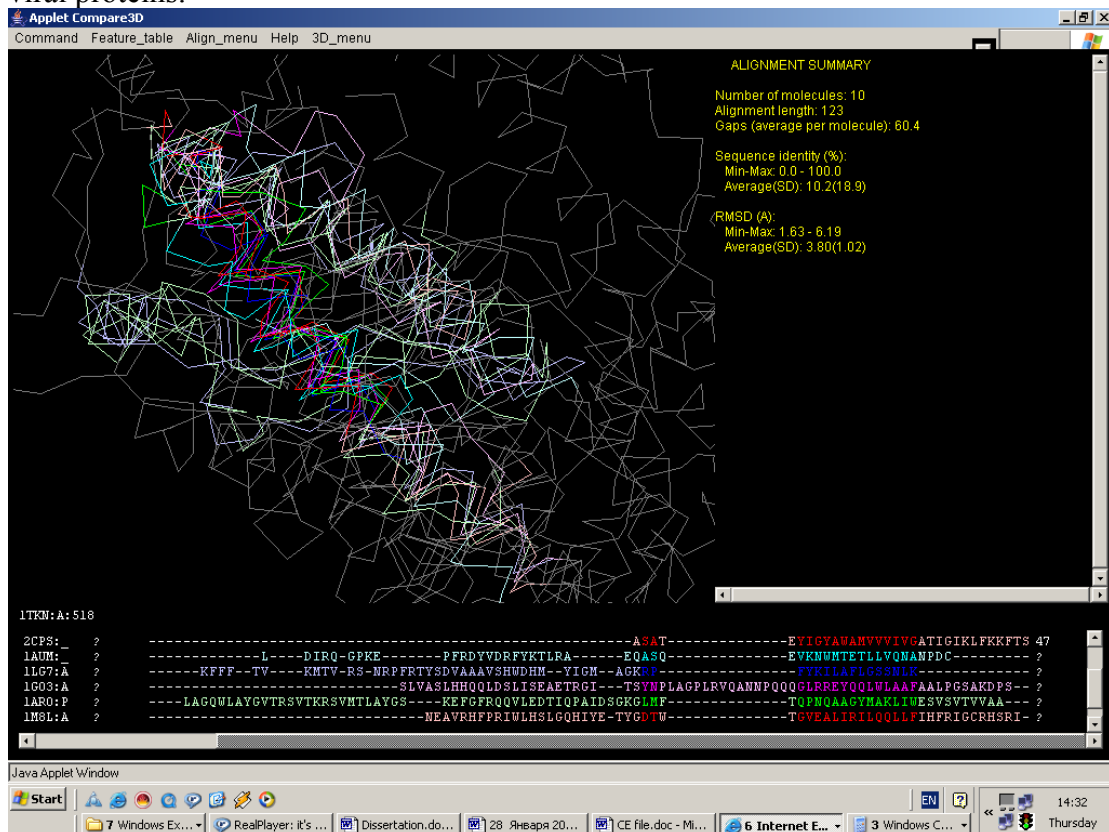
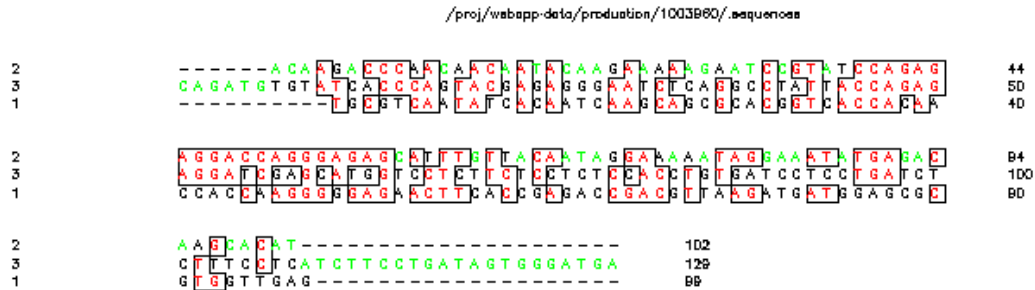


Figure 4. Multiple structural alignment (CE) of human prion protein and different viral proteins.



Box 3. Multiple sequence alignment (EMMA) of the region of HIV-1 env gene and the two regions of the prion gene. Sequences:

- 1: (CVN...MEERVE) region from prion (AF076976.1) region 535-633.
- 2: (TRP...QAH) region from HIV-1(AF033819.3) env fragment 6659-6760.
- 3: prion (AF076976.1) region 634-762.



Box 4. Pairwise sequence alignment (EMMA) of the prion gene and peanut clump virus. Sequences:

- 1: peanut clump virus
- 2: human prion gene

1	EACTACTCATGGCTATGAGCATGACCTAAATGATTATTTAGATTACATTA	1050
2	-----	D
1	AAAAATCTTTTGTGGTAGCTAAATGTGGTCATGTCTACATGCTTGAGCTT	1100
2	-----	D
1	TTTTCTGTAAGAGGGGACACCATCTTCTTTAAGCTGTATGATTACCGA	1150
2	-----	D
1	BTACAGTTTGTGAGAACTAGGCTTTTATCTTCCTTGTGCGGACTCGGG	1200
2	-----	D
1	CCATGTATTCAAAAGCTATGCCAATACCTAATACTGAGGAGGTAATAAGTG	1250
2	-----	D
1	CCGCTGTACTGGATTGAAAATGGAGAAAGTTGTGGTTGAAACGGCGTTACCT	1300
2	-----	D
1	FCCAAATCCCTCGTTTGTGAGGAAATGGAGTGGTTGATGCGTAATAAGG	1350
2	-----	D
1	CTAATGCTTTACAGTACGAGACATTGCTGAATTATTTGGTTTCAAATAAC	1400
2	-----	D
1	STTTCTGCCGTCTTTAATGGATGTCAGGTGCGAGAAAGTTTGAAGACGGA	1450
2	----- K TGGCGAA	B
1	FCCTTGGATATTGTGTAAAGTTGGCAATGACTTTGTAGTGCAGCAAGAAT	1500
2	-CCTTGGCTGCTGGATF-SCTGGTF-TCCTTTGTGGCCACATGGAGF--	5D
1	TTAATAGGGAAAAACAGAAAGAGATTCTGGAGATGTTGAGGATGAGCGGT	1550
2	-TGA CCTGGGCCTCTGC AAGAAAGCGCCGAAAGCCF-TGAGGATGGAACA	9B
1	FCAGAGCAGATAA-STTTAAAAG-AGTATCTTCCGGGAGTGTGTTGTAAA	1598
2	TGGGGGCAGCCGATA CCGGGGCGAGGGCAAGCCCTGGAAGCAACCCTACC	14B
1	TTTTTTGAAAGTGTCT-FCCTTGAAGGC--FCGAGTCTAAAACTGTAGC	1645
2	CA CCTCAAGGCGGTGTGTGGCTGGGGCGAAGCCCTCATGGTGGTGGCTGGGG	16B
1	FCGC-FGGTG-FGGTG-FTGAATTCGGAGCATGAAATACGATGTGGAA	1692
2	CA GGCCTCATGGTGGTGGCTGGGGCAAGCCCATGGTGGTGGCTGGGGACA	24B
1	FTCTTCCCTTATACGTTGAAATGAAAG--FAGCCTCTGTTTGTGGAGGA	1740
2	GCCTCATGGTGGTGGCTGGGGTCAAAGAGGTGGCAACCACAAGTCAAGTGA	29B
1	AGGTAAGTTGGAAAGGAATCGAGAGTTACGACATGAAAGATGAAGTCTG	1790
2	AF--EAAAGCCGF-AGTAAAGCCAAAAACCAF-KCATGAAGCACATGGCF-TGG	34B
1	AGGTACAATGAAAAGAAACAGGATGTTGAGCGTCTTACGAGATCCTTAGA	1840
2	TGCTGCAAGCF-AGCTGGGGCAGTGF-STGGGGGSCCTTGGCGGCTACGTGCT	38B
1	GCAGAAAGTAAATGATGGCTCGGCATCAGACACGTCTTCAAATTTGAAGA	1890
2	GGGAAGTGCCATGAGCAAGCCF-EATCATACATF--TTCGGCAGTGACTA	43A
1	AGTGGCTGCCCTTGCCTTGGCTAGTAATAAAGCACTCCTTTGGGG	1940
2	TGAGGACCGF--TACTATCGTGAAAACATGCAAGCGTTACCCCAACCAA	48B
1	AGGTTTATTGCTCT-FAGATCATTGTCCAGCGTCTCGAAACGGTCTGTCGT	1988
2	TGACTACAAGCCCATGGATGAF-STACAGCAACF-AGAACAA-FCTTGT	52B

1	GCAGTGAAGCGAGTCAATCCGGTCCGGATTGTCAAGCTAGTGAAGAGGTT	2038
2	GCATCGACGTCGTCGAATCTCACAATCAAAGCAGCCTCACCGGTCACC	570
1	GTTAGTTCCGAAGAGATTGCATCTCCGCAGATATCTGTTCTGAGAGTTG	2088
2	CAACCCACCAAGGGGGTGAAGCTTCAACCGAGACCGTAAAGATGATG	618
1	GTGGAAAGAGTCCACCTATGATTATACCTTACACTGTGAAGAAAAGT	2138
2	BAGCGCGTGGTTGAGCAGATGTGTATCACCCAGTACBAGAGGGAACTC	868
1	GGGCGGATTATTCTAGTGAAGTAGCGAAGACGTCGTAGCAGGGGAAGA	2188
2	CAGGCCTATTAACCTAGAGAGGAACGAGCATGGT	898
1	CTTTATGTACCGGTGTCAACTTTGGTCCTCCGAGACGTCTACCTGCC	2238
2	CTCTTCTCCTCTCTCAACTGTGATCCTCCTGATCTCTTCTCATC	744
1	CCCTGCATCCGGATGACGTTACAGGAGGCTGCTTGTCTGGAGTATCTG	2288
2	TTCCCTGTAGTGGATGAF	762
1	TGGTATCTGAAGTGAAGATTATTTGTGACTTCTCTGCTATGTATTCTGT	2338
2	-----	762
1	CGTGTGTGATTTCAAGGATCAGTTACTGCATGATGGGAGGTGCGAGTTTC	2388
2	-----	762
1	CGAAAAATGCTTTCTTTTTGAAAGTAGGGGAGGAAACAAAGTGGGCGATG	2438
2	-----	762
1	AGCGGCCTACATCTCAACAGGTCCGACATCAATATTGTGTAAAATTTTC	2488
2	-----	762

Box 5. Multiple sequence alignment (MAFFT) of the HIV-1 proteins and prion proteins.

Amino acids are coloured by hydrophobicity in JalView Editor.

Sequences: 1-20: HIV-1 Pol , 21-40: HIV-1 Env , 41-45: HIV-1 Vif, 46-48: Prion, 49-53: HIV-1 Vpu,

Figure 1. Multiple sequence alignment (MAFFT) of the HIV-1 proteins (Pol, Env, Vif) and prion proteins. Upper part of the alignment.

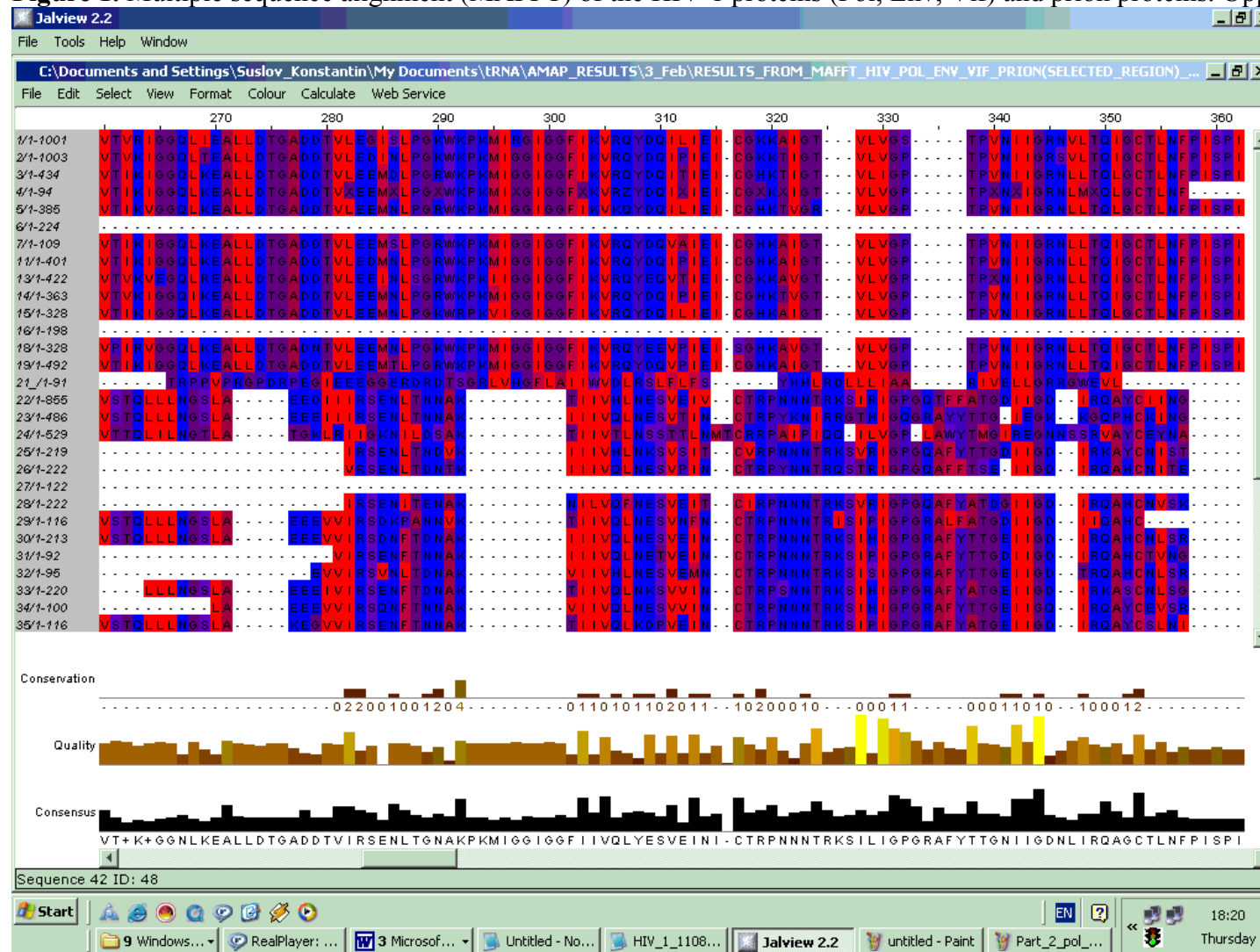


Figure 2. Multiple sequence alignment (MAFFT) of the HIV-1 proteins (Pol, Env, Vif) and prion proteins. Bottom part of the alignment.

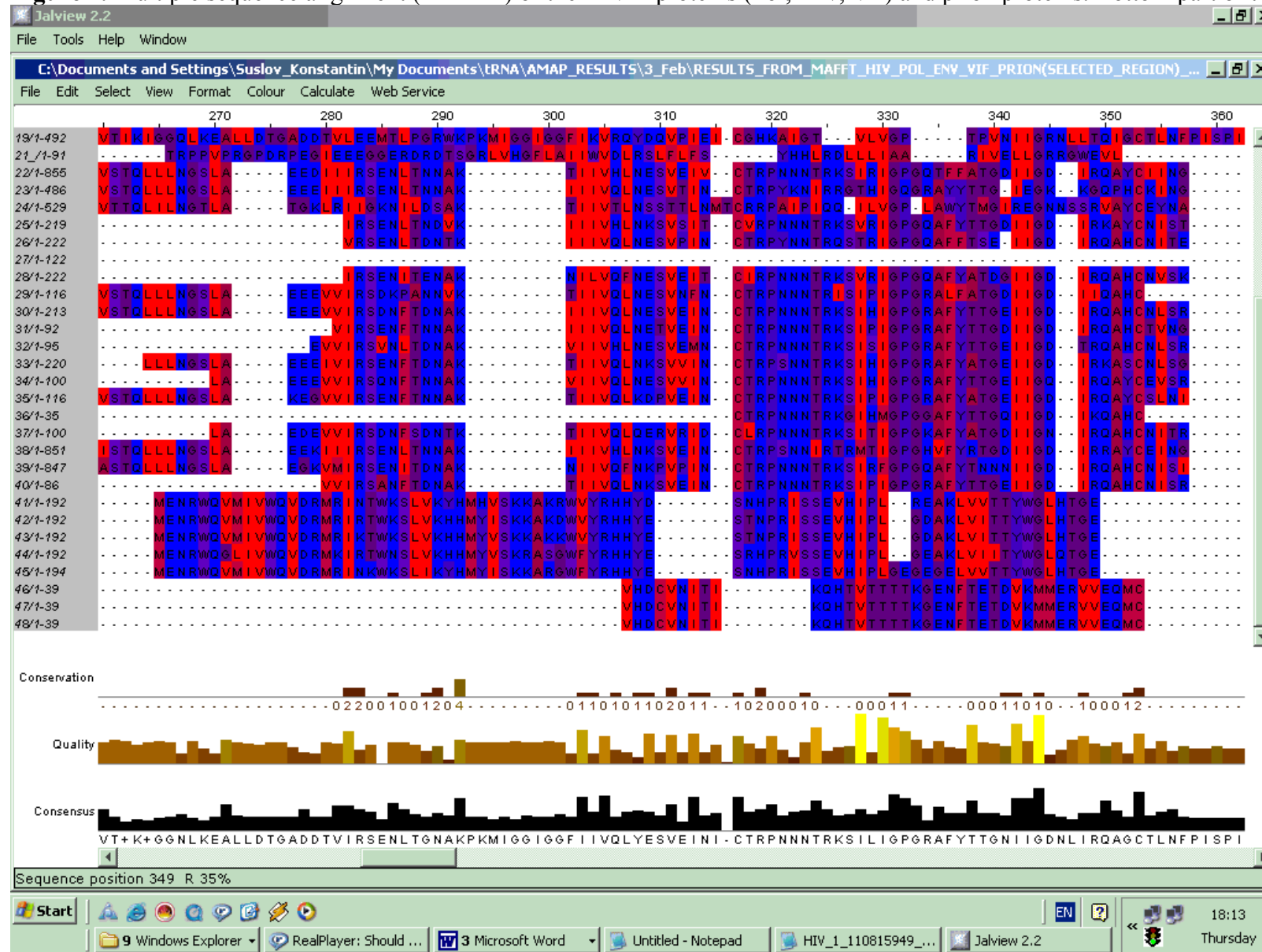
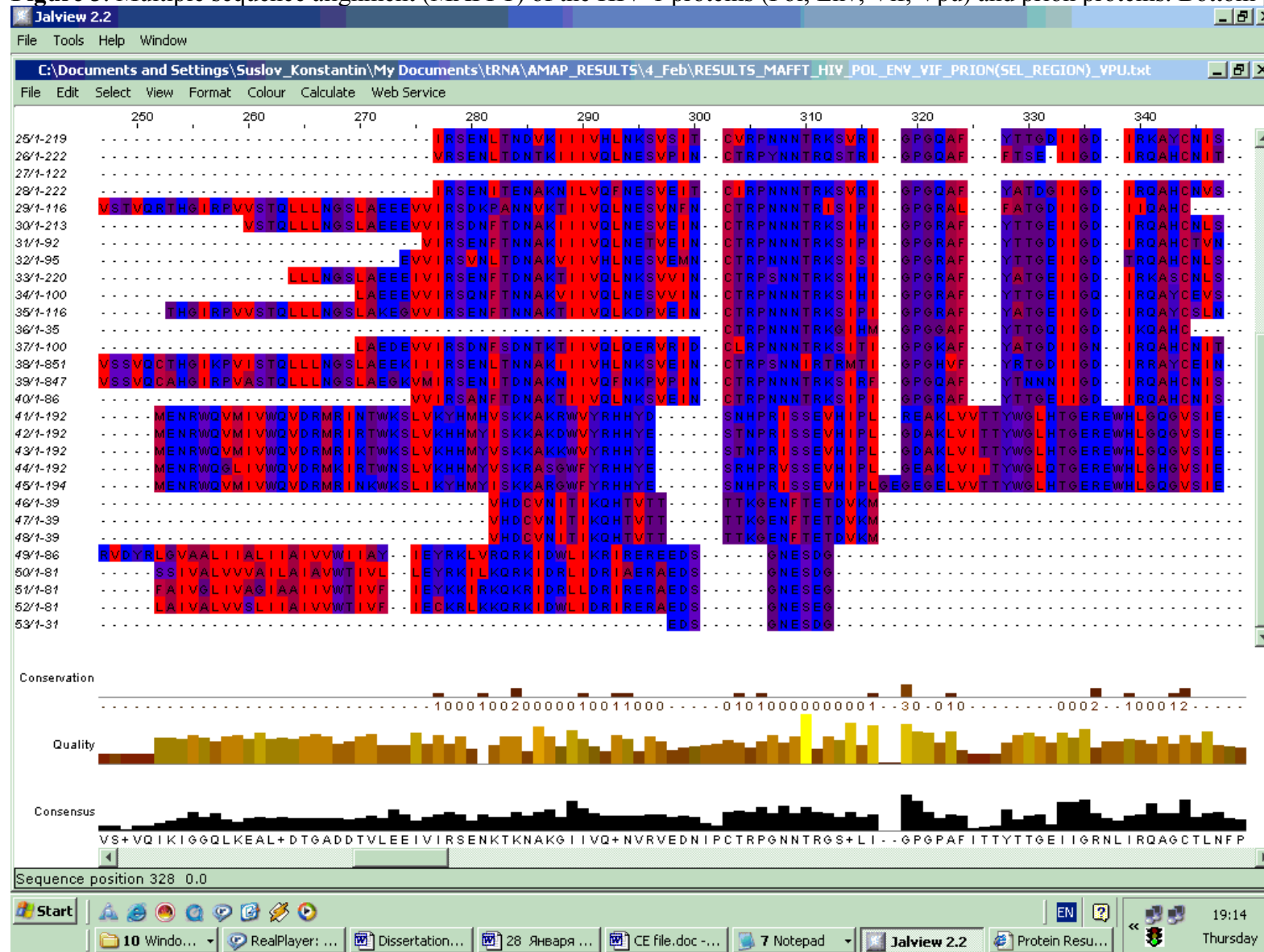


Figure 3. Multiple sequence alignment (MAFFT) of the HIV-1 proteins (Pol, Env, Vif, Vpu) and prion proteins. Bottom part of the alignment.



Box 6. Multiple sequence alignment (EMMA) of the HIV-1 proteins (Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env, Nef) and prion protein.

Sequences: 1: Gag-Pol, 2: Pr55(Gag), 3: Vif, 4: Vpr, 5: Tat, 6: Rev, 7: Vpu, 8: Env, 9: Nef, 10: human prion

/proj/wabapp-data/production/10025B2/.sequences

```

1      MGARA SVLSGGELDRWEKILRLRPGGKKKKYKLLKHIYWASRELERFAVNPGL      50
2      MGARA SVLSGGELDRWEKILRLRPGGKKKKYKLLKHIYWASRELERFAVNPGL      50
5      -----MAGRSGDSDEELIRTVR      17
3      -----MENRWQVMVQVYDRMRI RTWKS L V K H H M Y V S G K A R G W F Y R H H Y E      45
4      -----MEQAPL      8
7      -----
8      ---MRVKEKYYQHWRDQSWRWGTMLLGMLMICSATEKLVVTVYYGVVWKE      47
9      ---MGGKWSKSSVIQWPTVREMRRAEPAADRVGAAASRDLEKHGAIIS      45
10     ---MANEGCQMLVLFVATWSDLGLCKKRPKPGGWNTGGSRYPGK      41
    
```

```

1      LETSEGCRQILGQLQPSLQTFGSEELRSLYNTVATLYCVHQRLEIKDTKEA      100
2      LETSEGCRQILGQLQPSLQTFGSEELRSLYNTVATLYCVHQRLEIKDTKEA      100
5      ---MEPYDPRLEPWKHPGSSPKTACTNCYCKKCFHCQVCFETKALG      44
6      LIKLLYQSNPPPNPESTRQRNRNRFRWRERQRQIHSSERLELGTYLGRS      67
3      SPHPRISSVHIPLSDARLVIITVWGLHTGERDWHLGQGVSEWAKKRFYS      85
4      DQGPQREPHNEWTLELLLEKNEAVRHFPRIVLHGLGQHIYETVSDTWAQ      56
7      ---MQPIPIVAIVLVKIIIAIYVWSIVVEYRKILRQRK      39
8      ATTTLFCASDAKAYDFEVHNVWATHACVPTDPNPQEVVLVNVTENFNMWK      87
9      SNTAATNAACAWLEAQEEVGFPTPQVPLRPMTYKAAVDLSHFLKEKES      85
10     GSPGGNRYPPQGGGGWGGPHGGGGWGGPHGGGGWGGPHGGGGWGGPHGGGGWGG      81
    
```

```

1      LDKLEERQKRSKKKAQQAADTGHSNQVSNYPIVQNIQSQMVHQAISPR      150
2      LDKLEERQKRSKKKAQQAADTGHSNQVSNYPIVQNIQSQMVHQAISPR      150
5      ISYGRKKRRRRAKHQNSQTHQAQLSKQPTSQPRGDPTGPKE-----      86
6      AEPVPLQLPPLERLLDQNEDCGTSGTQGGVSPQILVESPTVLESSTKE-      118
3      TQVDPQLADQLLHLYYFDCFSDSALRKLGLGHIYSPRCVYQAQGHNKVGS      145
4      VEA RILQQLLHFHFGCRHSRQVTRQRRARNGASRS-----      98
7      DRLLDRLIERAEDSGNESEGEISLLVEMGVEMGHHPWVDVDDL-----      82
8      NDMVEGMHEDISLWQSSLKPCVKLTPLCVSLKCTDLKNDTNTNLSSSGRM      147
9      GLEGLIHSQRRLDLDLWIYHTQGYFPDWQNYTTPGGVRYPLTFGWICYKL      145
10     GGGTHSNWRKPSKPKTNNMKHMAQAQAAGAVYGGVGGVMLGSA MSAPIIH      141
    
```

27/02/06 plur=5.000000 -collision -box -nabaxaal oalbyaoneanus

```

1      TLNAWYKVVEEKAFSPEVIMPMSALSEGATPQDLNTMLNTVGGHQAAQM      200
2      TLNAWYKVVEEKAFSPEVIMPMSALSEGATPQDLNTMLNTVGGHQAAQM      200
5      -----
6      -----
3      QYLALAAALITPKKI KPLPSVTKLTEDRWNKPKQTKGHRGSHTMNGH---      182
4      -----
7      -----
8      I MEKGEIKNCSFNI STSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCN      187
9      VPVEPKIEEANKGENTSLLLHPVSLHGMDPEREVLEWRFDSSRLAFHHVA      185
10     GSDYEDRYYRENMHRYPNQVYYRPMDEYSNQNFVHDCVNI TIKQHTVTT      181
    
```

```

1      LKETINEEAAEWDRYHPYHAGPIAPGQMRPRGSDIAGT-TSTLQEQIGW      249
2      LKETINEEAAEWDRYHPYHAGPIAPGQMRPRGSDIAGT-TSTLQEQIGW      249
5      -----
6      -----
3      -----
4      -----
7      -----
8      TSVITQACPKVSEFPIPIHYCAPAGFAILKCNKTFNGTGPCTNYSTVQC      247
9      RELHPEYFKNC-----
10     TTKGENFTETDVKMMERVVEQMCITQYERESQAYYQRGSSMVLFSPPVI      241
    
```

```

1      MTNNPPIPVGEIYKRWII LGLNKIVRMYSPSTSLDI RQGPKEPFRDYVDR      289
2      MTNNPPIPVGEIYKRWII LGLNKIVRMYSPSTSLDI RQGPKEPFRDYVDR      289
5      -----
6      -----
3      -----
4      -----
7      -----
8      THGIRPVYSTQLLLNGSLAEEVYIRSVNFTDNAKTIIVQLNTSVEINCT      287
9      -----
10     LLI SFLIFLIVG-----
    
```

Box 7. Analysis of the viral and prion sequences on the presence of the tRNA elements.

Figure 1. Pairwise sequence alignment (EMMA) of the Val-tRNA and prion gene.

Sequences:

1: Val-tRNA, 2: Prion

/proj/webapp-data/production/1004067/.sequences



29/03/08 plus=1.000000 -collieian -box -nabaxool oelbyoensensus

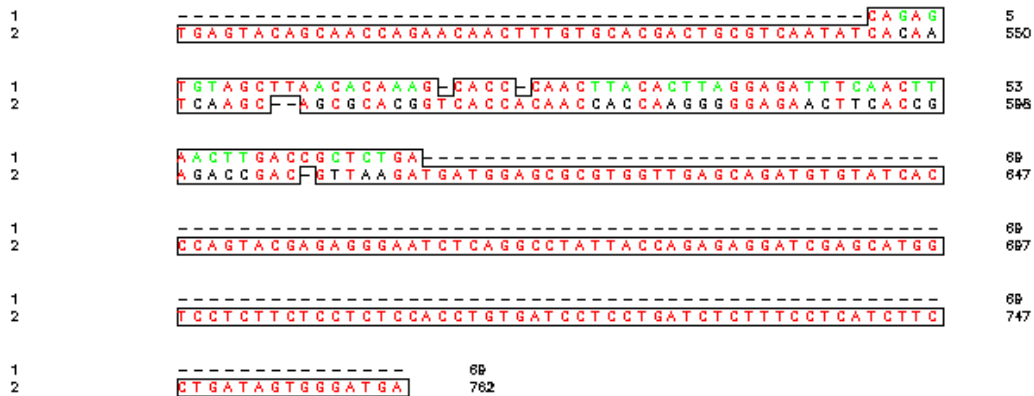
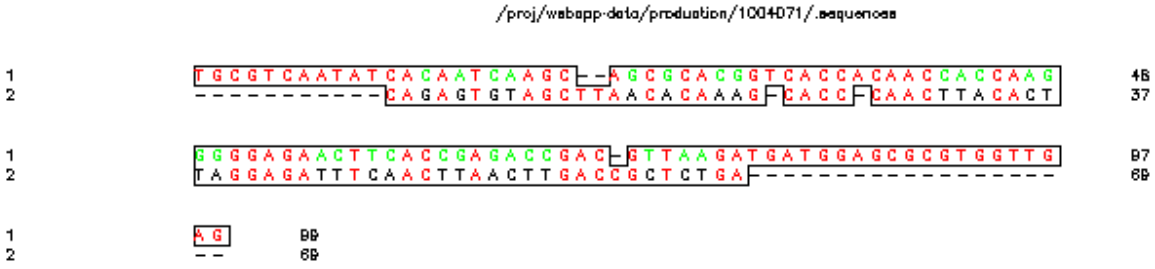


Figure 2. Pairwise sequence alignment (EMMA) of the Val-tRNA and (CVN...MEERVE)-encoding region of the prion gene.

Sequences: 1: (CVN...MEERVE) region from prion AF076976.1, region 535-633, 2: Val-tRNA



28/03/08 plus=1.000000 -collision -box -nabaxool oalbyconsensus

Figure 3. Multiple sequence alignment (EMMA) of the Val-tRNA, (CVN...MEERVE)-encoding region of the prion gene, (TRP...QAH)-encoding region of the HIV-1 env gene and the rest part of the prion gene.

Sequences:

- 1: (CVN...MEERVE) region from prion AF076976.1: region 535-633
- 2: (TRP...QAH) region from the HIV-1_AF033819.3: region 6659-6760
- 3: prion AF076976.1, region 634-762.
- 4: Val-tRNA

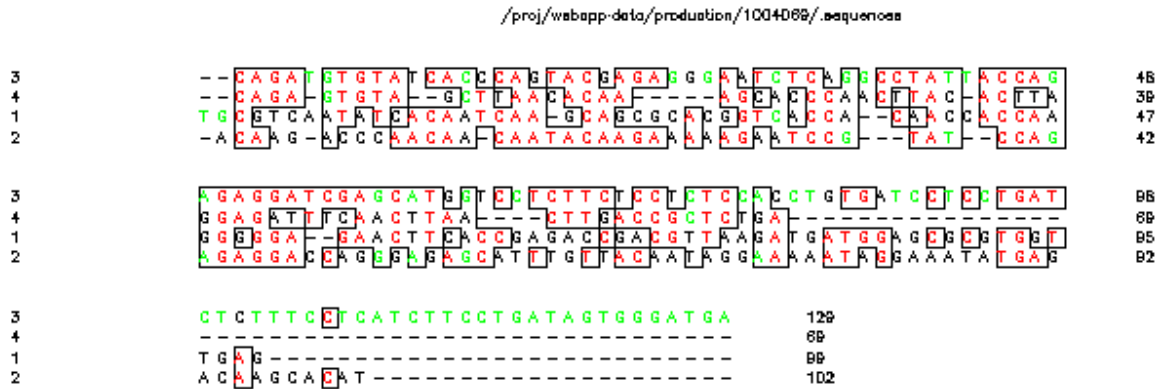


Figure 4. Pairwise sequence alignment (EMMA; default parameters) of the Val-tRNA and region 12 of the HIV-1 genome.

Sequences: 1: Region 12 of the HIV-1 genome; 2: Val-tRNA

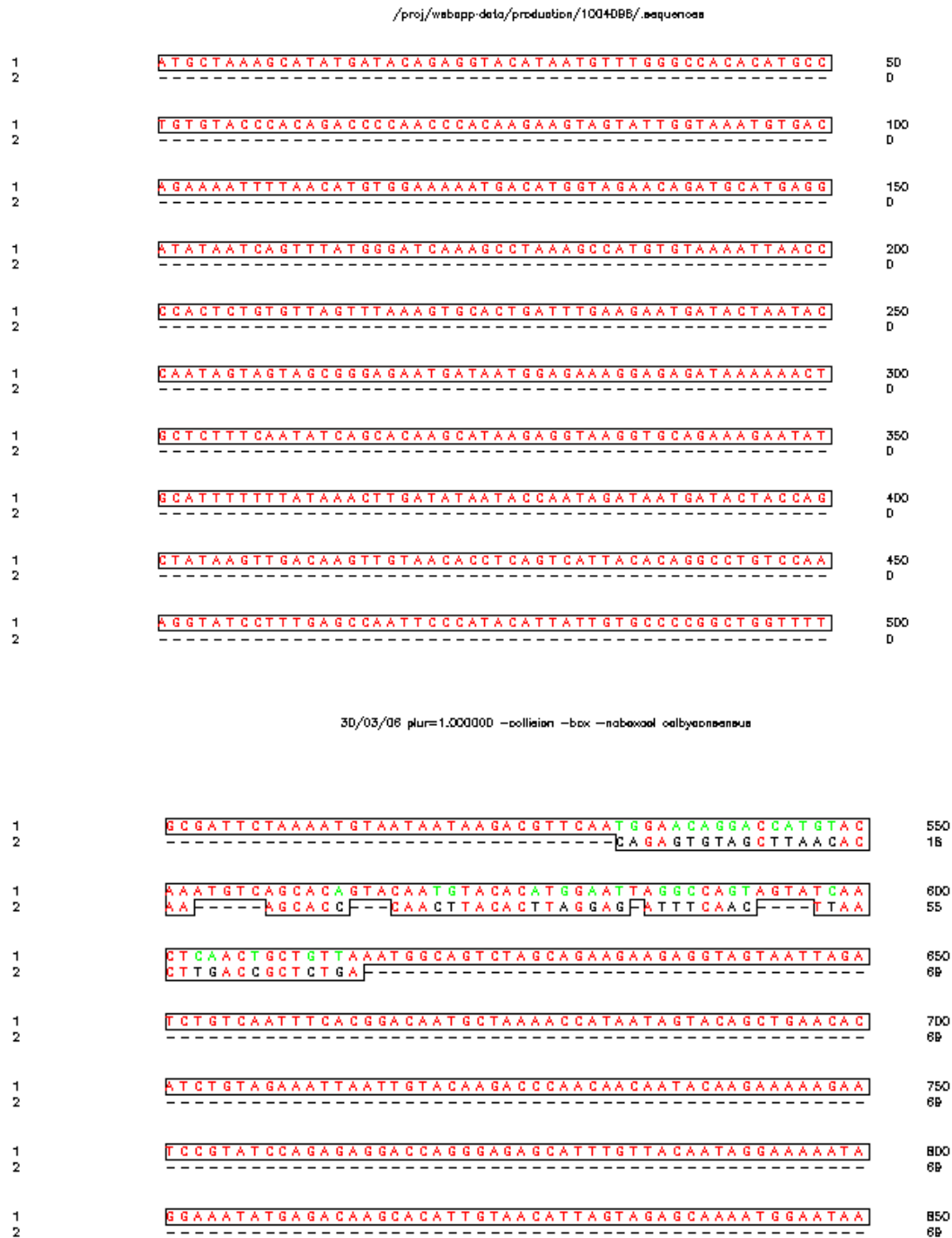


Figure 5. Pairwise sequence alignment (EMMA; GOP 40 and GEP 10) of the Val-tRNA and region 12 of the HIV-1 genome.
Sequences: 1: Region 12 of the HIV-1 genome; 2: Val-tRNA

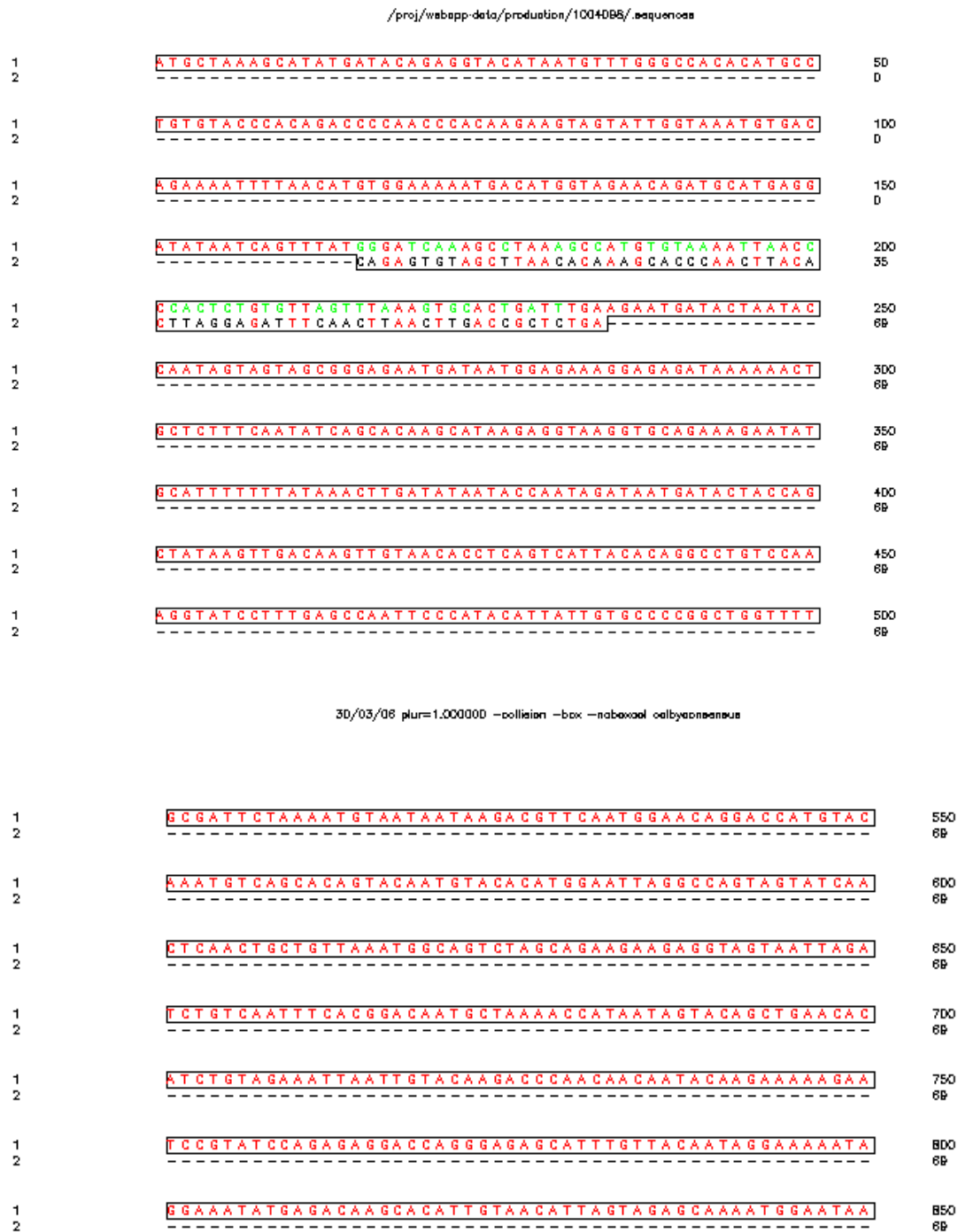


Figure 6. Pairwise sequence alignment (EMMA) of the Val-tRNA and Cadang-cadang coconut viroid.

Sequences: 1: Cadang-cadang coconut viroid; 2: Val-tRNA

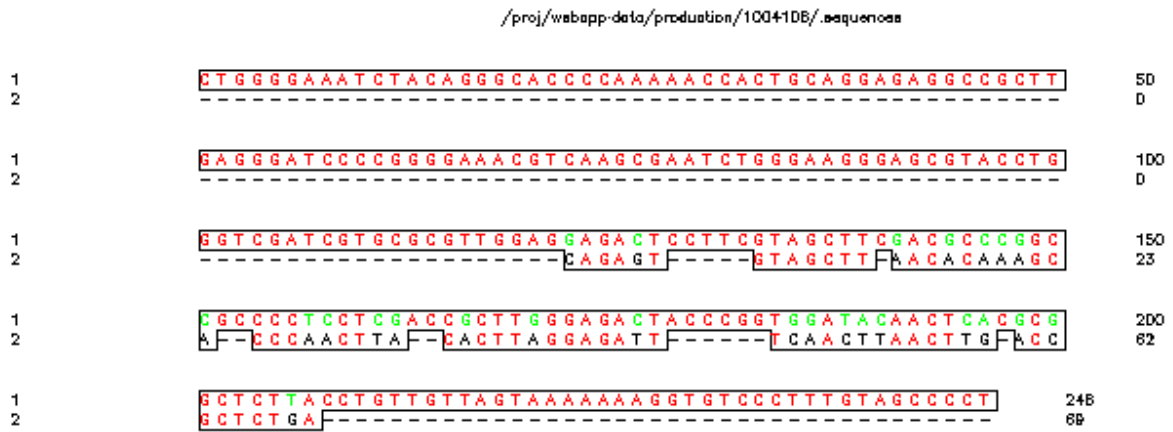


Figure 7. Pairwise sequence alignment (EMMA) of the Val-tRNA and the upper part of the Cadang-cadang coconut viroid.

Sequences: 1: Cadang-cadang coconut viroid, region 1-121; 2: Val-tRNA

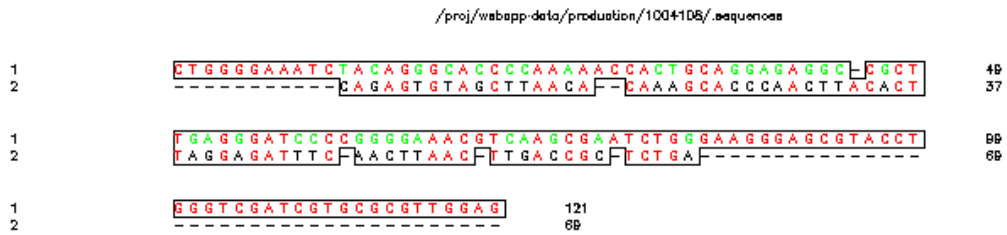


Figure 8. Multiple sequence alignment (EMMA) of the Val-tRNA, Leu-tRNA and Cadang-cadang coconut viroid.

Sequences: 0: Leu-tRNA;1: Cadang-cadang coconut viroid;2: Val-tRNA

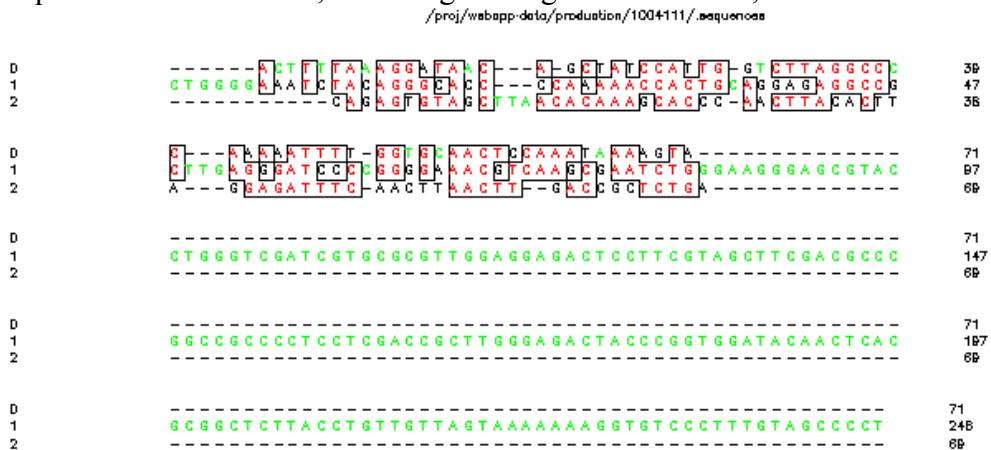
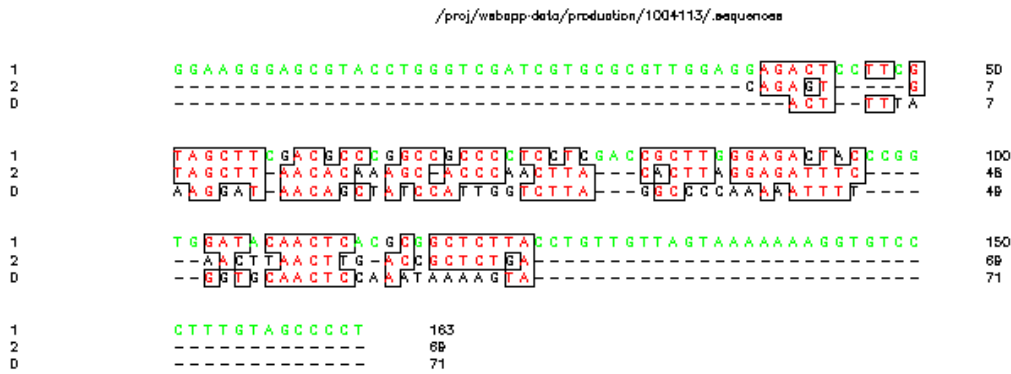


Figure 9. Multiple sequence alignment (EMMA) of the Val-tRNA, Leu-tRNA and the selected region of the Cadang-cadang coconut viroid.

Sequences: 0: Leu-tRNA;1: Cadang-cadang coconut viroid, region 84-246;2: Val-tRNA



30/03/08 plur=1.500000 -collision -box -nabowal oelbyaoneanus

Figure 10. Multiple sequence alignment (EMMA) of the Val-tRNA, Leu-tRNA and p53 gene.

Sequences: 0: Leu-tRNA;1: p53 ;2: Val-tRNA

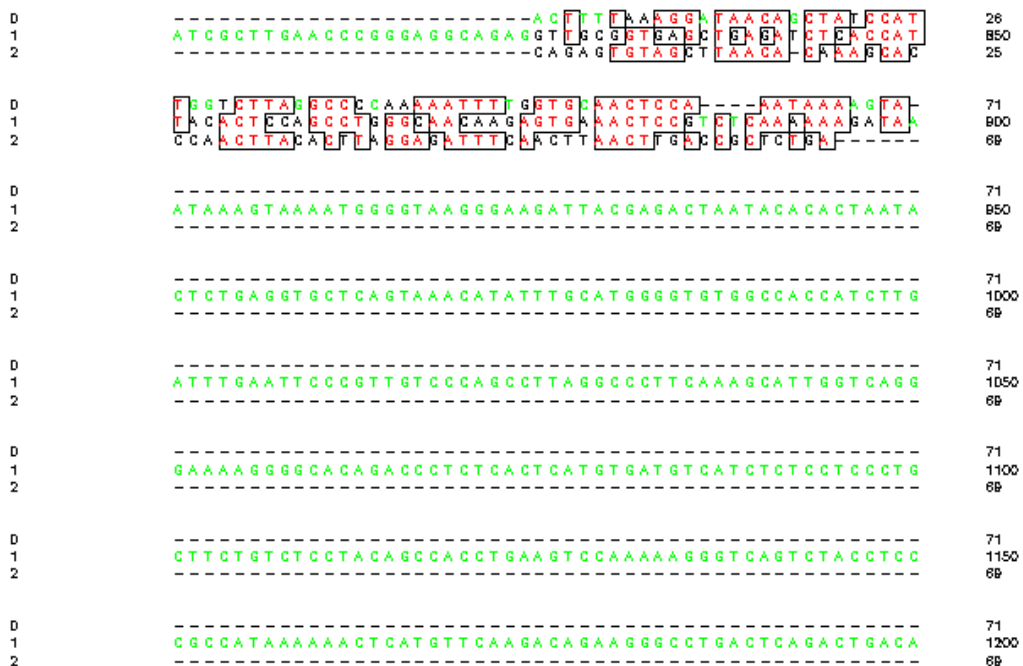
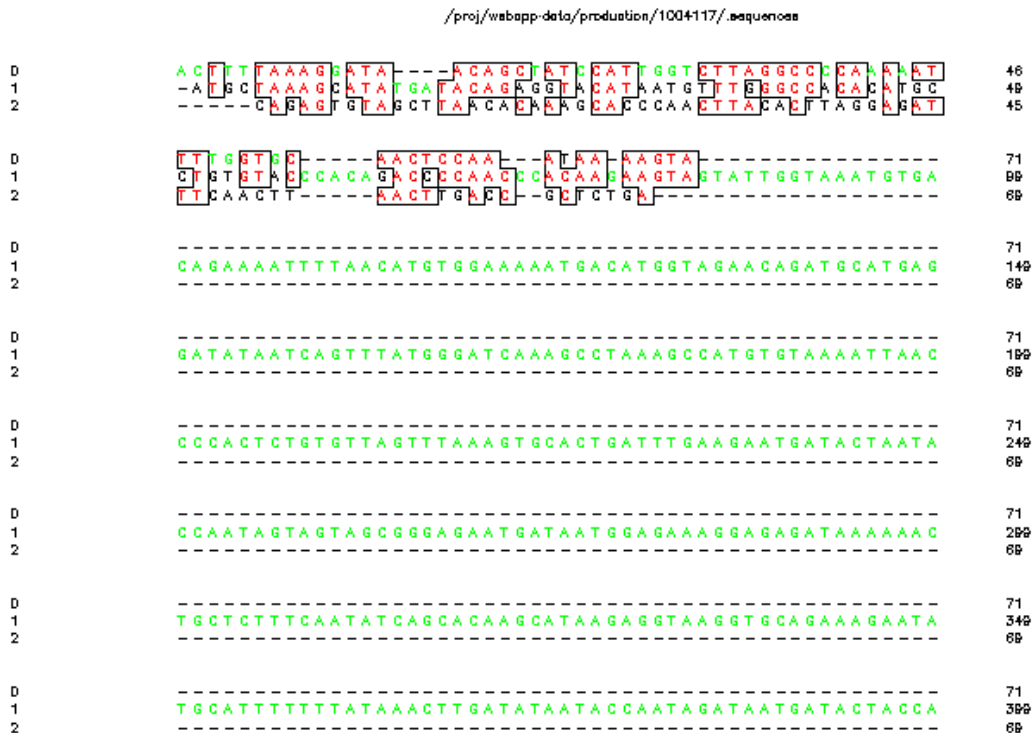


Figure 11. Multiple sequence alignment (EMMA) of the Val-tRNA, Leu-tRNA and region 12 of the HIV-1 genome.

Sequences: 0: Leu-tRNA;1: region 12 of the HIV-1 genome;2: Val-tRNA



30/03/08 plur=1.500000 -collision -box -nobaxool oalbyconsensus

Figure 12. Multiple sequence alignment (EMMA) of the Val-tRNA, Leu-tRNA and prion gene.

Sequences: 0: Leu-tRNA; 1: prion; 2: Val-tRNA

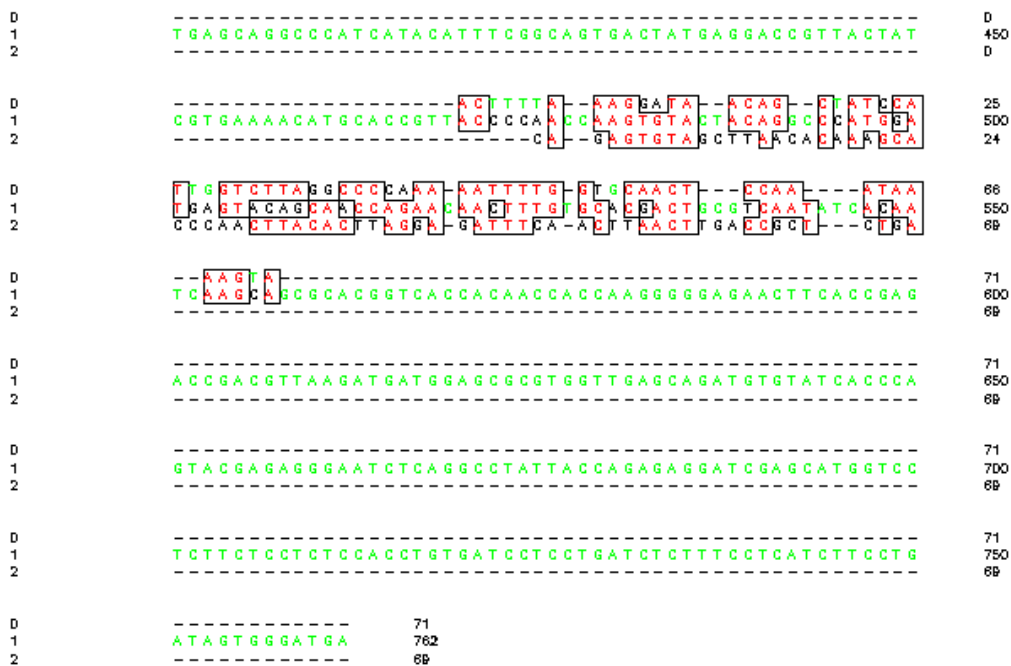
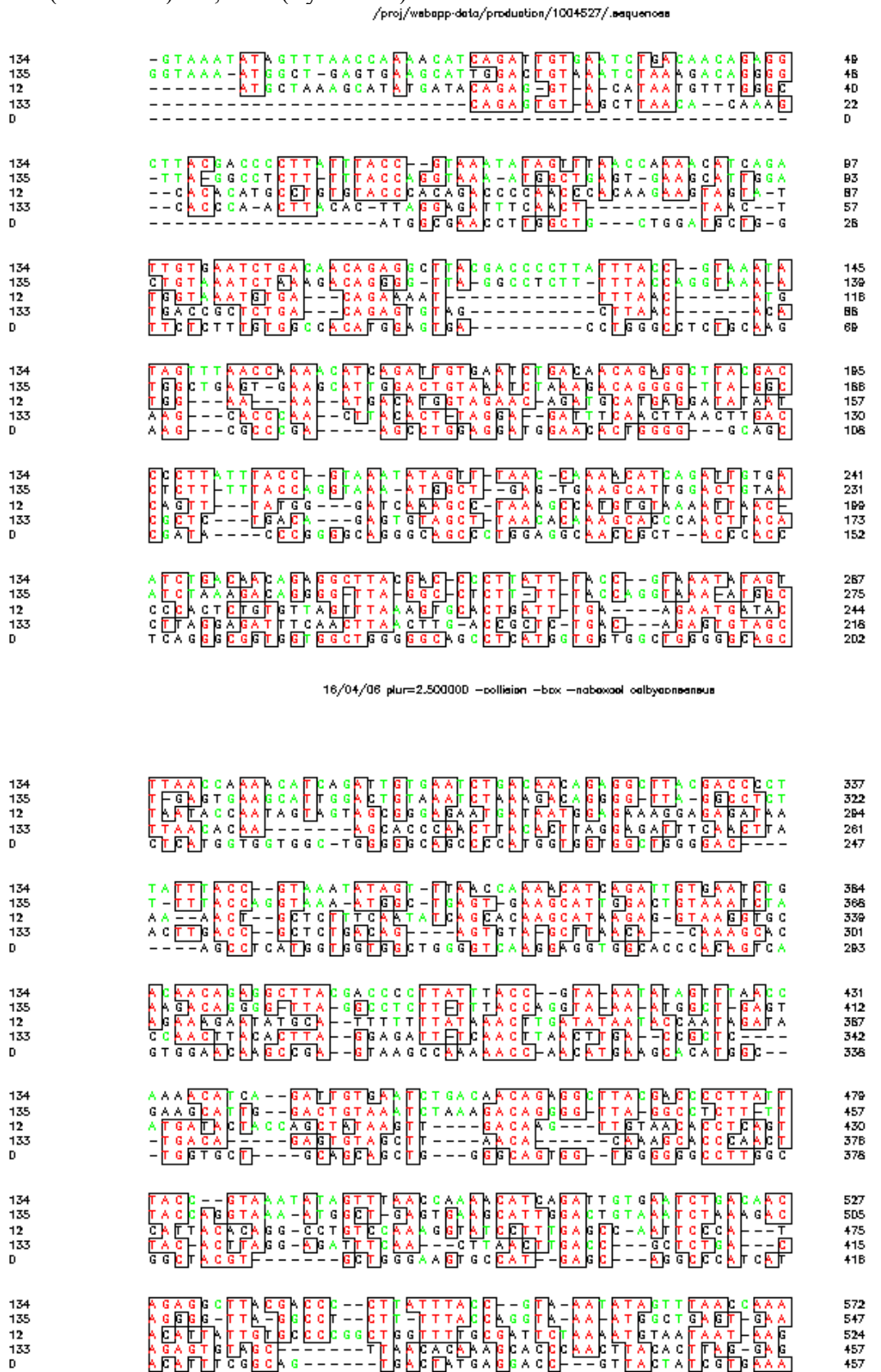


Figure 13. Multiple sequence alignment (EMMA) of the poly-tRNAs, prion gene and the region 12 of the HIV-1 genome.

Sequences: 0: prion; 12: region 12 of the HIV-1 genome; 133: (Val-tRNA)*11;
134: (His-tRNA)*13; 135: (Tyr-tRNA)*13



134 K C A T T C A G A T F G T G A A T C T G A C A C A G A G S C T T A C S K C C B C T - - - F A T T F A 819
 135 G C A T T G A C T G T A A T C T R A K G A C A G G S S - T F A S G C C T C T - - - T T T F A 591
 12 A C G T T C A F G B - - A A C A G S A C C A F G T - - A C A A K T B F C A G C K - - - C A G T A 568
 133 A T - T T C A C T T - - A A C T F G A C C G C T C T S A C A G A G T S T A G C T - - - F T A A C A 500
 0 A C A T G C - A C C B T T A C C C A A C C A A G T G T A C T A C A S G C C S A T S G A T S A S F A 508

134 C C - - S T A A A T A T A S T T T A A C C A A A C - A T C A G A T F G T S A A T B F S A C A A C A 868
 135 C C A S S T A A A A T S S C T G A G T - G A A G C C - A T T G B A C T G T A A A T C T A A A G A C A 838
 12 C A A T G T A C A C A T S S - - A A T T A S G C C - A S T A G - T A T C A A A C T C A A C T G C T 810
 133 C A A A G C A C C C A F - - - A E T T A C K C T T A S G A G A - T T T C A A C T F A A C T - - T 541
 0 C A F - S C A A C A S A A C - - A A C T T T G T G C A C G A C T G C S T C A A T A T C A C A K T C 552

134 G A G G C F T A C G A C C C C T T A F T T A C C - - G T A A A T A T A G T T T S A C C A A A A C A - 713
 135 G G G - - F T A - G G C C T C T T - T T A C C A G S T A A A - A T G G C T - G A B T G A A G C A - 862
 12 S - - - T T A - A A T G S C A G - - T C T A S C A G A K G A A S A G S T A S T A T T K S A T C T G 854
 133 S - - - - - A C C G C - - - T C T S A C A G A G T G T A G C T T A K C A F - - C A A A S C A C 577
 0 A - - - - - A S C A S C G C A C G G T C A C C - - K C A A C C A C A S - G G S G A S A A C T F - 582

134 T C A S A T T S T G A K T C T S A C A A C A S A S G C T F A C G A C C C T T A F T T A C C - - G T 761
 135 T T G G A C T S T A A A T C T A A A G C A S G G G - T F A - G G C C T C T T - T T T A C C A G S T 729
 12 T C A A T T T - C A C G G A C A A T B C T A A A C C A C A T A T A G T A C A S C T S - A A C A T 702
 133 C C A K C T T A C A C T F A G S A S A T T T C A A C T - T A A C T T G A C C S C T C T S A C A G A 825
 0 T C A C - - - - C S A S A C C S A C S T F A A G A T G A T S G A G C S G T G S T T S A S C A G A T 838

134 A A A T A T A G T T T A A C - - - C A A A C A T E A G A T F B T G A A T C T S A C A C A C A G A G S C 809
 135 A A A - A T S B C T - G A S - - T G A A S C A T T G A C T S T A A A T C T A A A G A C A G S S S - 774
 12 C T S T A S A A A T T A A T T G T A E A A A K C C C A A C A A C - A A T A C A A S A A - A A A G A A 750
 133 - - S T S T A G C T T A A C - A C A A A G A C C C A A C T - - - - T A C A C T T A - G S A G A T 868
 0 G T S T A T F A C C C A S T - - - - A C G A S A S G S G A A F C T C A G G C C T A T T A C C A G A G A 864

134 T T A C S A S C C C T T A T T T A C C - - S T A A A T A T S B T T F A A C C A S A A C A T C A S A T 857
 135 T T A - S S C C T C T T - T T T A C C A G S T A A A - A T S B C T G A S T - G A A B C A T T G G A C 820
 12 T C C G T A T C A G A G A G S A C C A G S A S A G C A T T S T T A C T - A A T A S G A A A A K A T 799
 133 T T C - - A A C T T A A C F T G A C C S C T C T G A C A S A B T S T A S C - T T A C A C A A A G C 713
 0 - - - G S A T C S A G C A T S S F C C - - - - - T C T T C T C C T C T C C S C C T G T S A T C C 724

134 T G T G A A T C T S A C A A C A G A G S C T T A C G A C C C C T T A T T T A C C - - - - - 887
 135 T G T A A A T C T A A G A C A G G S B - T T A - S S C C C C T T T - T T T A C C A - - - - - 858
 12 A S A A A T A T S A S A C A A G C A C A T T G T A C A C A T T A G - T A S A S A A A A T G G A A T 848
 133 A C C C A A C T T A C A C T T A G S A S A T T T C A A C T T A A C - T T G A C C S C T C T G A - 759
 0 T C C T G A T C T C T T T C T C A T C T T C C T S A F A G T G G G A T G A F - - - - - 762

134 - - 887
 135 - - 858
 12 A A 850
 133 - - 759
 0 - - 762

Figure 14. Multiple sequence alignment (EMMA; default parameters) of the Val-tRNA and different regions of the prion gene.

Sequences:

0: Val-tRNA

1: (CVN...MEERVE) region from prion AF076976.1, region 535-633

2: prion region 634-715

3: prion region 309-378

4: prion region 378-545

5: prion region 456-545

6: prion region 46-134

7: prion region 1-45

8: prion region 229-308

9: prion region 135-228

10: prion region 716-762

/proj/webapp-data/production/1004DB1/.sequences

```

2 -----
7 -----
0 -----
1 -----
4 C G G C T A C G T G C T G G G A A G T G C C A T G A G C A G G C C C A T C A T A C A T T T C G G C A
5 -----
10 -----
3 -----
9 -----
8 -----
6 -----
0
0
0
0
50
0
0
0
0
0
0
0

```

```

2 ----- C A G A T G T G T A T C A C C [E] A G T A [E] S 22
7 -----
0 -----
1 ----- T G C G T C A A T A T C A C A A T C A A G C T T A A [E] A 17
4 G T G A C T A T G A G G A C C G T T A C T A T C G T G A A A A C A T G C A C C S T T A C C C C A A C 26
5 ----- A A A C A T G C A C C S T T A C C C C A A C 22
10 -----
3 ----- T A A G C C A A A A C C A A C A 13
9 ----- A G G C A A C C S T T A [E] C C A C C T 17
8 -----
6 ----- T G G A G T G A C [E] T [E] S G C [E] T [E] G [E] A 0
0 22

```

```

2 A G A G G S A A T C T C A G S C C T A T T A C C A G A G A G S A T C G A S C A T S B T C [E] T C T T C 72
7 A T G S C G A A C C T T G S - C T S C T - - - - - S G A T - - - S C - T S B T T [E] T C T T T 36
0 C A A A S C A C C - C A - - C T T A C A C T T A G G A G A T T C A A C T T A A C T T G A C C G 63
1 C A C S G T C A C C A C A - - C A C C A A G G G G A G A A C T T C A C G A G A C [E] S A C - G 73
4 C A A G T G T A C T A C A G - G C C - C A T G G A T G A G T A C A G C A A C C A G A A C A A C T T 147
5 C A A G T G T A C T A C A G - G C C - C A T G G A T G A G T A C A G C A A C C A G A A C A A C T T 69
10 G A T C T C T T T [E] C T C A - T C T - T C T G A T - A G T G G G A T S A - - - - - 47
3 T G A A S A C - A T G S - C T S G T G C T - G C A G C A G C - T S G - G S C A G T G G T G G 59
9 C A S G G C G G T G G T G G - C T S G G G - - - G C A G C C T C A T S G T G S T G G C T G G G G 63
8 C A T S G T G - - G T G S - C T S G G G - - - A C A G C C T C A T S G T G S T G G C T G G G G 41
6 A G A A S C G C [E] S A A G - C T G G A S G - - A T G S A A C A C T S S G S C A G C C G A T A 66

```

30/03/08 plur=5.500000 -collision -box -nabaxal oalbyconsensus

```

2 T C C T [E] T C C A [E] C ----- 82
7 G T G S C C A C A [E] ----- 45
0 C T C T G A ----- 69
1 T T A A G A T G A T G S A G C G C G T G G T T G A G ----- 99
4 T G T S C A C G A C T S C G T C A A T A T ----- 168
5 T G T S C A C G A C T S C G T C A A T A T ----- 90
10 ----- 47
3 G G G S C C T - - T G S ----- 69
9 G C A S C C T C A T G S T G G T G G C T G G G G C A G C C C - - - - - 94
8 T C A A G G A G G T G S C A C C C A C A G T C A G T G G A A C A A G C C G A G ----- 80
6 C C C S S G G C A S G S C A G C C C T G G ----- 89

```

Figure 15. Multiple sequence alignment (EMMA; GOP 40 and GEP 10) of the Val-tRNA and different regions of the prion gene.

Sequences:

0: Val-tRNA

1: (CVN...MEERVE) region from prion AF076976.1, region 535-633

2: prion region 634-715

3: prion region 309-378

4: prion region 378-545

5: prion region 456-545

6: prion region 46-134

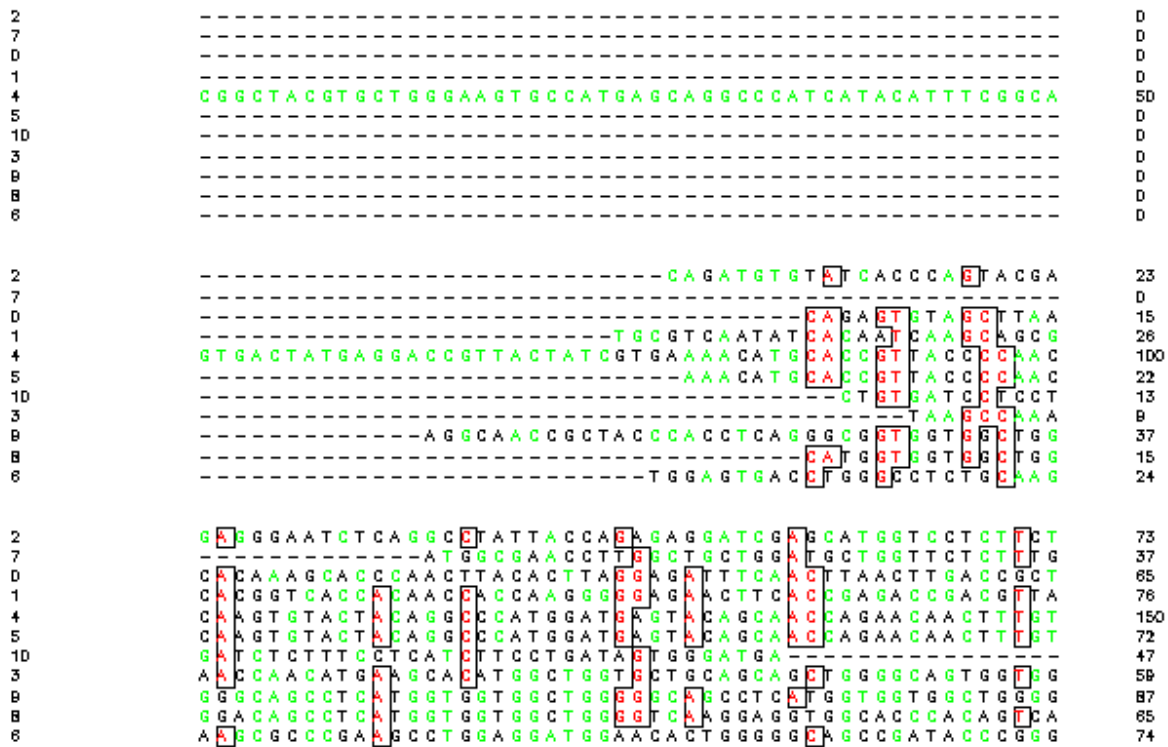
7: prion region 1-45

8: prion region 229-308

9: prion region 135-228

10: prion region 716-762

/proj/webapp-data/production/1004DB4/.sequences



30/03/08 plus=5.500000 -collision -box -nabaxaal oalbyaoneansus

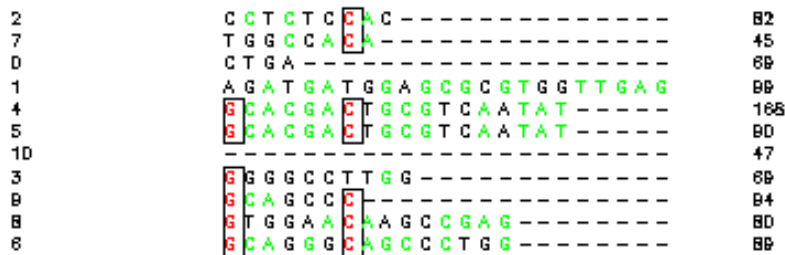
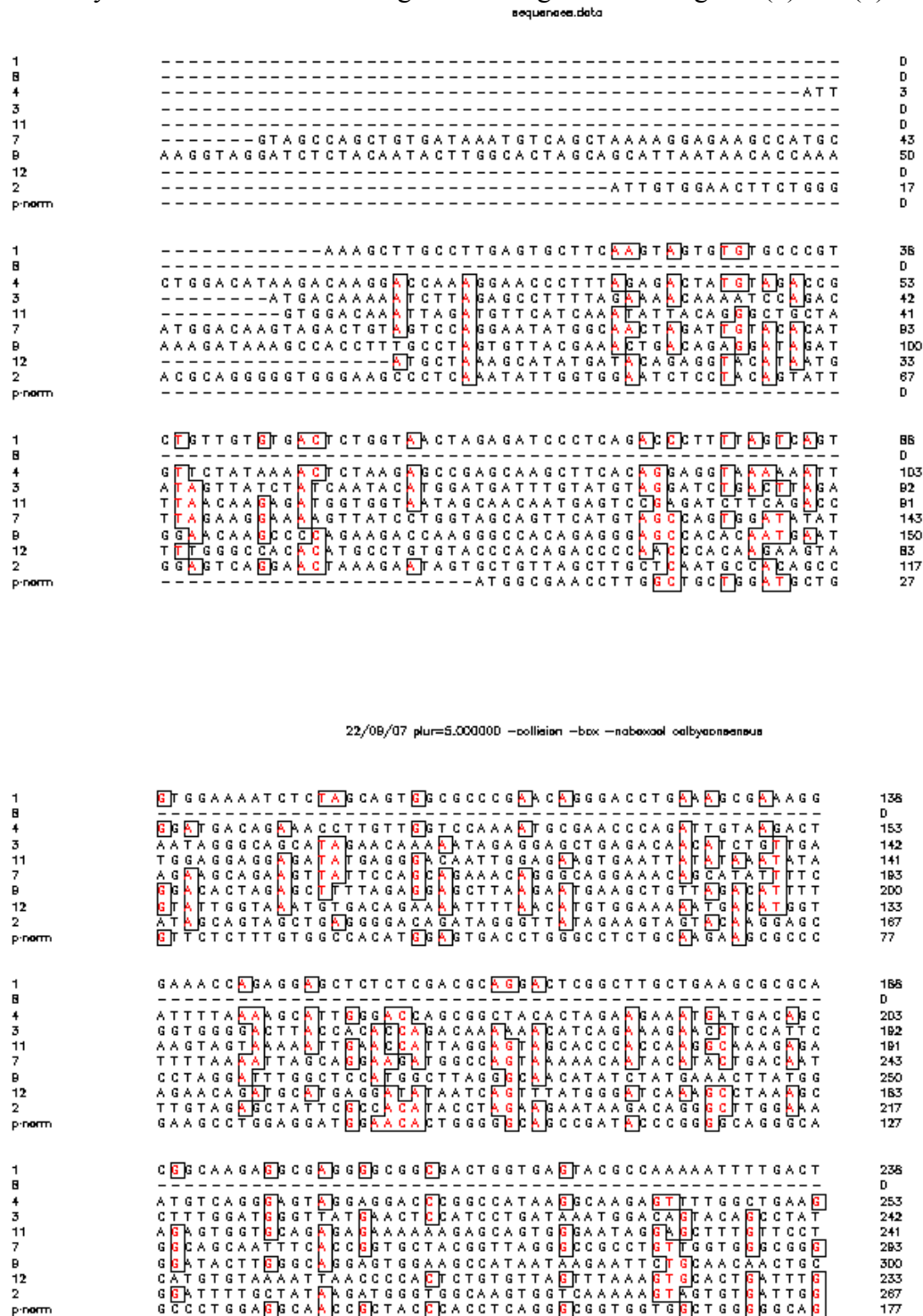


Figure 16. Four squares code

		Second Position of Codon	
		W	S
First Position of Codon	W	Phe/Leu Asn/Lys Ile/Met Tyr/Ter	Thr Cys/Ter/Trp Ser Ser/Arg
	S	Leu Asp/Glu Val His/Gln	Ala Arg Pro Gly

Box 8. Comparison of multiple sequence alignments of the prion gene with regions of the normal and shuffled HIV-1 genome.

Figure 1. Multiple sequence alignment (ClustalW 1.82) of the 9 regions (sequences 1, 2, 3, 4, 7, 8, 9, 11, 12) of the normal HIV-1 genome and prion gene (p-norm). Designation (p-norm) is used only for convenience to distinguish the alignments in Figures (1) and (2).



1 K G C S G A G B T A G K A S G A G A B A G A T G G G T G C G K B A S C G T S A G T K T T A A G C G 268
 B - A S A A F C C A T A G T A G S A S C A G A K A C C T T C T K T G T A G A T G S B S C A S C T K A 48
 4 C A A T G A G C C A K G T A C A C A A T T C K G C T A C C A T K A T S A T G C A G A G A G S C C A T 303
 3 K S T S C T S C C A K A A A A A G C T S G A C T G C A T K A T G A C A T A C A G S C T A S 282
 11 T G G S T T C T T G S G A S C A S C A S G A A G C A C T A T G G G C S C A G S C T C A K T S A C G C 281
 7 K A T C A A G C K G S A A T T T G S A A T T C C C T A C A A T C C C K A A G T C K A S G A G T K S 343
 9 T G T T T A T C C A T T T T C A G A A T T G G S T G T C G A C K T A S C A G A K T A G S C T T A C 350
 12 K A G A A T S A T K C T A A T A C C A A T A G T A G T A G C G G G A S A T G A T A T A T G S A G A K 263
 2 K T G S C C T A C T S T K A S G S A A A G A K T G A G A C G A G C T S A S C A G C A S C A G A T A 317
 p-norm C C T C A T S S T G S T G G C T S S G G C K G C C T C A T G G T G S T G G C T S S G C A G C C 227

1 G S G S A S K A T F A G A T C G K T G G G A A A K A K T F C G G T F A A G G C C A G S G G G K A A S 338
 B C A S S K G A C T A A A T T A G S A A A A G C A G S A T A T G T F A C T A A T A G A S G K A S A C 86
 4 T T T A G G A C C K A A S A A K G A T T G T T A A G - T G T T T C A A T T G T G S C A A A G A A S 352
 3 T G G S G A A T F S A A T T G G S C A A G T C A G A T F A C C C A G S G A T T A A G T K A A S 342
 11 T G A C S G T A C A G S C C A G S K A T T A T T S T C T S G T A F A G T G C A G C A S G C S A A C 341
 7 T K S A T C T A F S A T K A A G A A T T A A A G A A A A T T A F A G S A C A G S T A A G A S A T 383
 9 T S A C A G A S G K A G A S C A K A G A A A T G G A G E - C A G T A G A T C C T A G A C T A G A S C 388
 12 A S G A S K G A T A A A A A C T S T C T T T C A A T A T C A G C A C A A G C A T A G A G S T A 353
 2 G S G T S S S A S C A G C A T C T C G S B A C C T S S A A A A A C A T S S A G C A K A T C A C K A S T 367
 p-norm C C A T S G T G S T S C T S G G S A C A G C C T C K T G G T G G T S C T G G S T C A K S G S S 277

1 A A A A A T A T B A A T T A K A C A T A T A S T A T G G B C A A G C A G G G A B C F A G A C C G 368
 B A A A A G T T G T C A C C C T A A C T G A C A - - - C A R C A A T C A G A A S A C T G A S T T 144
 4 G G C A C A C A G C C K A A A K T T G C A S G G C C C C T A G S K A A A A G G G C T G T T G G A C A 402
 3 C A T T A T G T A A C T C C T T A G A S G K A C C A A A G C C T A A C A S A G B T A K A C C 382
 11 K A T T T G C T G A S G G C T A T G A G S C S C A A C A G C A T C T G T F S C A A C T C A C A S T 381
 7 C A S G C T G A A C T C T T A G A C A S C A S T A C A A T G G C A G T A T T C A T C C A A 443
 9 C C T G G A K G C T A C C A G A A G T C A G C C T A A A A C T G C T T G T A C C A A T T G C T A T 448
 12 K S G T G C A G A K A G A A T A T G C A T T T T T T A T A A C T T G A T A T A A T A C C A C A T A 383
 2 K S C A A T A C A G C A S C T A C C A A T S C T G C T T G T B C C T G G C T A S A B C A S C A 417
 p-norm G T G G C C C C K C A S T C A S T G G A C A A G C C G K G T A A G C C A A A A C C A K C A T 326

1 A T T C S C A S T A T C C T S C C T F S T T A G A A A C A T C A S A K G S C T G T A G K A C A A 438
 B A C A A S G C A A T T T A T C T A S C T T F S C A G S A T T C G - - - G S A T T A S B K S T A A A 189
 4 T G T G A A A S G A A G S A C A S C A A A T G A K A S A T T B T A C T G A G A G A C A G G C T K A 452
 3 A C T A C A S - - A S A S C A G A G C T A G A A A C T G S C A S K A C A G A S A S A T T C 439
 11 C T G G S C A T C A A S C A G C T C C A G S C K A G A A T C C T G S C T S T G G A A G A T A C C 441
 7 T T T A A A A S A A K A S G S G A T T S S G S G T A C A G T S C A G S G G A A A G A A T A S 483
 9 T G T A A A A S T G T T S C T T T C A T T S C C A S A T T T B T T C A T A A C A A A S C C T T 488
 12 G A T A T G A T A C T A C C A S C T A T A G T T G A C A A B T T S T A C A C C T C A S C A T 433
 2 G G A G S A S G A G G T G S T T T T C C A S T C A C A C C T C A G S T A C C T T T A G A C C K A 467
 p-norm G A A G C A C A T S G C T S S T S C T G C A S C A G C T G G S C A S T G S T G G G G S C T T G 378

1 F A C T G G G A C A G C F A C A - A C C A T C C T T C A S A C A G G A F C A S A A S A C T T A G 487
 B C A T S F A C A G A C T C A C A A T A T G C A T T A G S A T C A T T C A A S G C K A C C C A G 239
 4 F T T T T A S G G A A G A T C T S G C T T C T K C A A G G G A A S G C S G G A A T T T F C 502
 3 F A K A A G A C C A G F A C A T S G A G T S T A T T A T S A C C C A T C A A A S B C T T K A T A 489
 11 F A A S G G A T C A C A G C T C C T G G G A T T T G S B T T G C T C T G G A K A C T C A T F 491
 7 F A S C A T A A T A G C A C A S C A F A E A A C T A A G A A T T A C A A A A C A A T T 543
 9 A G G C A T C T C T A F S C A S G A A G A A G C G G A S A C A G C G A S A A G C F C A T C 548
 12 F A C C A A G C C T G C S A A S S T A T C C T T G A G C A K T T C C A T A K A T A T F 483
 2 F S A C T F A C A A G C A S C T S T A G A T C T F A G C C K C T T T T A A A S A A A G G G G 517
 p-norm G C G G C T A C G T G C F S G G A A S T G C C A T G A G C A G G C C A T C A F A C A T T T C G G C 428

1 K T C A T T A F A T K A T A G S F A G - - C A A C C T C F A T T G T G T G C A T C A A G G A T 535
 B A T C A A A G T G A K T C A G A G T F A G T C A A T C A A T A A S A S C A S T F A A T A A A A 289
 4 T T C A A S C A S A C C A G S G C K A - C A G C C C C A C C A G A A G A S A S C T F C A G G T C 551
 3 G C A G A A T C K A S A S C A G S G C A A G G C K A T S G C A T A T C A A T T T A T C A 539
 11 T G C A C C A C T S T G T G C C F T G S A A F S C T A G T S S A S T A A T A A A F C T C T G G A 541
 7 A C A A A A T T C K A A T T T C G S G T T A T T C A S S G A C A S C A S A A F C C A C T 583
 9 A G A C A S T C A G A C T C A C A G C T T C T A T T C A A A S C A S F A A S T A C A T 588
 12 G T - G C C C G S C T G G T T T F S C B A T T C T A A A A T S T A A T A A A S A C G T T C A A 532
 2 G G A C T G G A A S G C T A K T F C A C T C C A A A G A S B A C A A G A T A F C T G A C T C T 567
 p-norm K S T G A C T A T S A S G A C C G T F A C T A F C G T G S A A C K T G C A C C G T F A C C C A 478

1 A G G G A T A A A G K C K C C K A B B K A G C T T T A G A C A K B A T A G A G G K A B K G C A A A
 8 A A G G A K A A G G T C T A T C T G B C K T G G G T A C C A B C K C A C A A G G K A A T T G G A G G
 4 T G S G G T A G B B A C K A C A C K C T C C C C C T C A G A A B C K G G K G C C G A T A G K C A A G G
 3 A G A G C C A T T T A A A A T C T B K A A C A G G A A A T A T G C A K A G A A T G A G G G G T G
 11 A C K G A T T T G G A A T C A C A B B K C C T G G A T G G A B T G G G K C A G A G A A T T A K C A
 7 T T G A A K B G A C C K B C A A B C T C C T C T G G A A A G G T G A A G B B C K B T A G T A
 9 G T A A T G C A C C T A T A C C A A T K B T A G C A A T A B T K B C A T T A G T A B T A B C K A A T
 12 T G S A A K B B G A C C A T G T C A A K T G T C A G C A C A G T A C A A T G T A C A C T G G A K A
 2 G T G A T C T A C C K C A C A A G G C T A C T T C C C T G K T A K C A G A A C T A C A C K C
 p-norm C C K A G T B T A C T A C K A G C C A T G G A T G A G T K A G C A K C C A G A A C A C T T T G

585
339
801
589
591
843
848
582
817
526

1 K C A K A A G T A A B A A A K A B C A C A B C A K B C A G C K G C T G A C A C A B G K C A C K B C
 8 A A T G A A C A A B T A G A T A A T T A B T C K G T B C T G B A K T C A G G A A A B T A C T A T
 4 A A C T G T A T C C T T F A K C T T C C C T C A G G T C A C T C T T T G G C A A C G A C C C C T C G
 3 C C C A C A C T A A T G A T G T A A A C A A T F A C A G A G B C A G T B C A A A A A T A K C C
 11 K T T A C A C A B C T F A K C A C A C T C C T T A T T G A K G K A T C G C A A A A C K B C A
 7 T A C A A G A T A A T A G T G A C A T A A K A B T A G T B C C K G A K A A A A B C A K B K T C
 9 A A T A T A G C A T A G T T G T G T G G T C C A T A B T A K T C A T A G A A T A T A G G A A A
 12 T T A G G C C A G T A G T A T C A A C T C A A C T G C T B T T A A T G G C A B T C T A B C A G A A
 2 C K B G G C C A G G G G T C K B K A T A T C K A C T G A C C T T T B B K T G S T G C T A K B C T A
 p-norm T G C K C G K C T G C G T C K A T A T C A K A T C A A B C A G C G C A C S B T C K C A C A K C C

835
389
851
839
841
893
898
832
867
576

1 A A T C A G G T C A G C C K A A A A T F A C C C T A T A B T G C K B A A C A T C C A G B B S C K A K T
 8 T T T T A G T G B A A T A G A T A G B C C C A A S A T G A K C A F - - T B A G K A A T A T C A
 4 T C A C A T A A A G T A K S G G G G G C A A C T A A A G G A K B C T C T A T T A G A T K C A G G A
 3 A C A G A A G C A T A G T A A T B B G A A A G A C T C C T A A A T T A A A C T B C C C A T
 11 G A K A K A T B A A C A B A T T A T T G G A T A G A T A A - - A T G B B C A A G T T T
 7 A T T A G B G A T T A T G G A A A A C A B A T G G C K A G T G A T G T B T B G C A S T K B
 9 T A T T A A B A C A K A B K A A A A T A B A C A G G T T A A T T G K T A G A C T A T A G A A G A
 12 G A K A G G T A B T A A T T A A A T C T G T C A A T T C A C G G A C A A T B C T A A A C C A T
 2 G A C C A S T T B A G C C K B A T A B A T A G A A A G G C C A T A A G A B A G A A C A C
 p-norm A C C A K G G B G A B A K C T T C A C G A G K C G A C G T T A A S A T G A T B G A G C G C G T

885
435
701
889
888
743
748
882
717
826

1 - G G T A C A T C K B G C C K T A T C K C C T A G A A C T T A A A T G C A T G G G T A A A A G T A
 8 - C A G T A A T T B G A B A G C A A T G G C T A G T S A T T T F A C C T G C C A C C T G T A G T A
 4 G C A G A T B A T A C A B T A T A G A A G A A T T T G C C A B G A A G A T G G A A C C
 3 A C A K A G B A K C A T G G G A A K C A T G S T G G A C A G K B A T T T G G C A G C C C T
 11 G T G A A T G B T T T A K C A T A C A A A T T B G C T B T G G T A T A T A A A T T A T T C A
 7 A C A G B A T G A B G A T T A G A K C A T G G G A A A A T T A B A A A C A C A T A T G T A
 9 G C A G A A G A C K B T B C A A T G A G A T B A A A G A G A A T A T C A G C A C T T G T G A
 12 A A T A G T A C A B C T B K C K A C A T C T G T A G A A A T A A T T S T A C A A G A C C C K C A
 2 C A G C T T B T T K A C C C T G T B B C C T B C A T G G G K T G G A T B A C C C G G A G A G A
 p-norm G G T T G A G C A B A T B T G T K T C K C C C A G T A C G A G K B G G A A T C T C A B C C T A T T

734
484
751
739
738
793
798
732
767
878

1 S T A G A A G A G A A S B C T T T C A B C C C K S A A G T G A T A C C A T G T T T C K S C A T T
 8 B C A A A A G A A T A - - - - - T A - - - - - T T F A C C T G C C A C C T G T A G T A
 4 A A A A A T A A G G B G A A T F G G A G G T T T F A T C A A A G T A B A C A G T A T G A T C
 3 S B A T T C C T S A B T S G G A G T T B T T A T A C C C C T C C T T A B S B A A A T T A T G G
 11 T A A T G A T K G T A S G A G C T T G B T A G G T T F A G A A T A G T T T F T G C T G T A C T T
 7 S T T T C A G B A A A S C T A G G G B A T G G T T T A T A G A E A T C K C T A T G A K G C C C
 9 S A T G G G B T G G A B A T S G G G C A C C A T G C T C C T T G G G A T G T B A T A T C T G T
 12 A C K T A K A G K A A A A A A A T C C G T A T C C A G A G A G A C C C K B G G A G A C A T T
 2 A A G T T A G A B T B A A G T T B A C K S C C G C T A G C A T T T C A T C A C G T G G C C
 p-norm A C C A G A G A G B K T C G A S C A T G G T C C T C T F C T C C T C T C C K C C T G T K T C C F C

784
488
801
789
788
843
848
782
817
726

1 A T C A G A A G G A G C C K C C C C A C A A G A T T T A A C A C C A T G C T A A A C A C A G T G G
 8 A T C A A A G A A T A - - - - - T A - - - - - T T A C C A A G T A A B A C A G T A T G A T C
 4 A G A A C T C A T K B A A A T C T G T G G A C K T K A B C T A T A G G T A C A G T A T T A G T A
 3 T A C C A G T T A G K B A A - - - - - T T A C C A T T A T C G T T T C A G A C
 11 T C T A T A G T G A K T A S A G T T A G G C A G G G A T A T T C A C C A T T A T C G T T T C A G A C
 7 T C A T C C A A G A K T A A G T T C A G A A G T A C A C A T C C C A C T A G G G A T G C T A G A T
 9 A G T G C T A C A G A A A A T T B T G G G T C A C K A S T C T A T T A T G G G G T A C C T G T G
 12 G T T A C A A T A G A A A A T A G G A A A T A T G A G A C A A G C A C A T T G T A A C A T T A G
 2 C G A G A G C T G C K T C C G A G T A C T T C A G A C T G C T G A C A T C G A G C T T G C T A
 p-norm C T G A T C T C T T C C G C A F C T T C T G T K B T G G G A T G A - - - - - T G G G A T G A

834
488
851
803
838
893
898
832
867
762

1	G G G G A	-----	839
8	-----	-----	498
4	G G A C C T A C A C C T G T C A A C A T A A T T G G A A G A	-----	861
3	-----	-----	803
11	C C A C C T C C C A A C C C C G A G G G G A C C C G A C A G G C C C G A A G G A A T A G A A G A A	-----	867
7	T G G	-----	898
9	G A A G G A	-----	804
12	T A G A G C A A A A T G G A A T A A	-----	850
2	C A A G G G A C T T T C C G C T G G G G A C T T T C C A G G G A G G C G T G G C C T G G G C G G G A	-----	817
p-norm	-----	-----	762

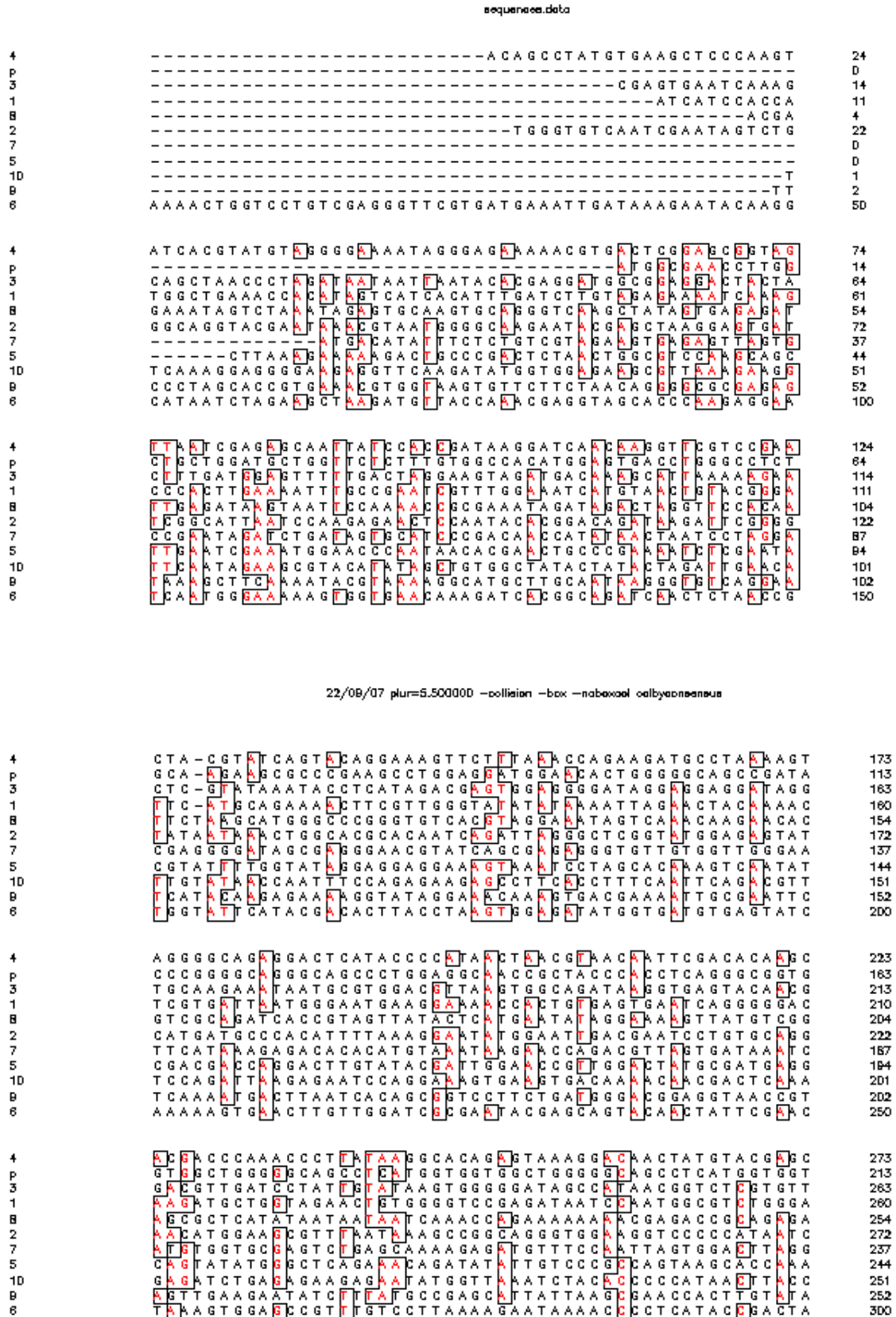
1	-----	-----	839
8	-----	-----	498
4	-----	-----	861
3	-----	-----	803
11	-----	-----	867
7	-----	-----	898
9	-----	-----	804
12	-----	-----	850
2	C T G G G G A G T G G C G A G C C C T C A G A T C C T G C A T A T A A G C A G C T G C T T T T T G C	-----	867
p-norm	-----	-----	762

1	-----	-----	839
8	-----	-----	498
4	-----	-----	861
3	-----	-----	803
11	-----	-----	867
7	-----	-----	898
9	-----	-----	804
12	-----	-----	850
2	C T G T A C T G G G T C T C T C T G G T T A G A C C A G A T C T G A G C C T G G G A G C T C T C T G	-----	1017
p-norm	-----	-----	762

1	-----	-----	839
8	-----	-----	498
4	-----	-----	861
3	-----	-----	803
11	-----	-----	867
7	-----	-----	898
9	-----	-----	804
12	-----	-----	850
2	G C T A A C T A G G G A A C C C A C T G C T T A A G C C T C A A T A A A G C T T G C C T T G A G T G	-----	1067
p-norm	-----	-----	762

1	-----	839
8	-----	498
4	-----	861
3	-----	803
11	-----	867
7	-----	898
9	-----	804
12	-----	850
2	C T T C	1071
p-norm	-----	762

Figure 2. Multiple sequence alignment (ClustalW 1.82) of the 10 regions (sequences 1-10) of the shuffled HIV-1 genome (“SHUFFLESEQ” program) and prion gene (p).



4 p 3 1 8 2 7 5 10 9 8

S G A - S A T G G C G A T A C A G A G A T K T A C K C G G S G T C C G A A T G G K A C G T A T C A T
S S C T A G S G G G C A G C C C A T G G T G S T G G C T S G G G A C A G C C T C A T G G T S T G T G G
A A C T A G C G C T A G A A G G T G A G A G S A A K G C C C A C T A T A G A T G A A G G T A T - A T
A A G G S A T A T T T A G A T T G A A G K A A G G C A S E A C A A G T A T G T A K A C K T C T
S S T A C C G T C T G T G A A T C A A A A K A G T G A C T C T T G T T C A A A A G A T T T S A T A A
S A A S G C G G A C A G A G T A T G T T K A G K C T G S T G G G C A C C G T C R A T A K A A G C
A C C A S A T A T T C A A G G T A G A T T C C A T A T C T A C G A C T A C A T G T G G A A K S G G G
S A C A C G G A T G A T T G C G A T C T T K A C T A C T C T A T T C C C G G G S A A A G A G A A
S A A T S T T A A A G A C A G T A G G A K A G T C T C T A A T T G G T C C G C A K A T T A C G A G G
T G A C C A A A G A T A G A A T T C T A G K A G C A K T S C A A A A A A A A A A S C G T S A A A G G
A C G C T A T C C A C A T C T G T T G G A K A G T A K A G S T C A T A G G T C T T A A S T A K A C A T C

322
263
311
310
304
322
267
284
301
302
350

4 p 3 1 8 2 7 5 10 9 8

A C K C G S A S C T G K A A A A G C K A A T A G T A G A C T A A S T C K T G K A A T A T C T G C C T A
C T G G G G T C A A G G K S G T G G C A C C C K C A G T C A G T G G A K A C A G C C G K A G T K A K G C
C C K C A T C A T G T K T A T A C T A C C K A S T A T G C A A S T G T T A K A C C G G A C K A T T G
T C C T G A S G C A A A T G K T C T T C A A C C A K A A G A C A C A C C G C T G G G - A A T G T C C A
A G K A G A T S C A T T T T C T A G C C A A G C T G A A C K A T A C C C A K A A K T C T C
A C A T C T A G S B C A T C A G G A T A G A A A T A G G T A G A A T A T A A C T G G C G T A A T
T A T A C A C G A G G C K A C T A A A C C G A K A G A C T A G T B T T A G T A A C G T C G A A G A A
T T K C A A T T B A T G K A A T C A A C A T A T K C C A T C C T A S C C A G K A A G A K T C T C C C
A C C A G C A G S C A C G T A T A T T A T C A T A G A C T A T A C A G G A G A G A K A G A K T A
A A G G C A C C S C A K A A A G G G C T C C A A G G A C A A B C A K A T G A A G A G A G G A
G T K A A G T A B T G C T K A A G A G T G T C C K C T A G G C T G C G T K A T T G T T T G G T G G A G

372
313
361
358
354
372
357
344
351
352
400

4 p 3 1 8 2 7 5 10 9 8

A G C K A A T A A T K T A A G G C T G G T A C A T A A G T A G K T G S T C A T S T C G T T G G C
C A A A A C C A A C K T G A A G C A C A T G G C T G G T G C T G C A S C A G C T G G G G C A G T G
T C C K T T A T A T C T A T K S G T C G T A G T A T A T T G A K T A T A C T A G T T A A T A A A G
A A T A S C T G A G C K S C A G T T T A C T A C T A C T T A G G G G S T G A C T G A A A G A C C G
T A T G A C A C A A A C C K C G T C A A C A G C C T G A A A T A A S G A G A B T A C T A G T A
A A A K S T C C A T A A A C A G T G G A G T G T C G T G C K A S S A T A A G T G T A C T C G A
A A A K T S B G G A K A S G C A T T G C T A A G C T A C T A C T A K T A S G A C C G C G C A C T T
T T G G A T A A A G C G G G T A T T C C T C C A G G A G A A C C A G S A G G G A C A A C A G G C
C C T A S T G A T G A T T A C A T A G T G A A A G G C T A K C C A G T A A A G A C G G A T T
T C C A T A C T T G T A C G T A C C G A G A G G T A C A T G A K A A C T G A T A C G T T A C A A
A G A K A C A T G T K A A A K C C C A C G T C T A C A C G A A T T C A A C G T A C A G T G G T A G

422
363
411
408
404
422
387
384
401
402
450

4 p 3 1 8 2 7 5 10 9 8

A T A A A C T C K A S T A G A T A T A A S A C T A A S G C C C T C A A S C K C C A T G
G T S G G G S C C T T G G C G G C T A C G T G C T G G G A K A S G C C A T G A S K A G C C C A T
A T A T A T A T A T A T A C A A S T A A T T S A A A A A C C C C G G A C A S B A A K T T A K
A A K T A G C T C G K A K A C A A G C A C A A C G G G T G K A T C G G G T A A A C G T T T A S
G A G C C C A G G G C A G S G S B T A C S T A T A A G T T A C A A A G G A A T S T T A S
C T A G A G A T G T A G A A C T S T B T A T G S T G T C T A A T G G C G A G A G K A T G A A K
T A G A K T G T A G A K A C T S T B T G C C G C A T T A A T G G G A A T T B A T K A A G G
T T A S A T S T A C C T T T C C T T A T C C G S C A T G G K A A T C T C C T A C C C T T C T T C
G T T C A G K C A A G K A A A G A G S T C G G A G G G A S T A T G T A A G A T A A T T A K C
C T T C A C A T B C K A A T G A S C T B A T S A T G C C K A S C G A T C T A G S A T T A K S
C G T T A C S A T A T G A A G A C G S C T A S S T T A C C T C A A A G G S S T T A A A S

472
413
461
458
454
472
437
444
451
452
500

4 p 3 1 8 2 7 5 10 9 8

A C T G T A C T C T T G A S G A G C C T C A C C C C T A G K T K G C A T - T C C G T A S T T A K S
C A T K C A T T T C G G C A S T G A C T A T G A S K C C G T T A C T A T - C G T G A A A C A T S
T A T G A G T G T T T G A C G A C T G A G A G C T A A C K A G S C S C A A G T A A C T A K T
C G T K G G C A T A G C G A C T G A C G A A A A S G S A T T A T C S A S C C C A T A T A S A T T C
G G T G C G A T T C C A A T T G A G T C G C S K A S G A S K T A S C S T C G T C A C G G T T S
A G A S G G A T A A C T A T G G G G A A C G G S A A A T K T T A A C T A T A A A T A A T T S A T
K T G C C C A C A G T A A C G C A G T A T T A G T C T C K G A A C A T G C T A A C A K A S
C T T A A T A C A G A S G T T G C A T A T G C C K T G C K T T S C A G C T A T A S G T C G S
T A T T T A G G C A G T A C T A A A T G C C S T A A C A T A G S C - - - A A G T C T T K G
C A A A T T A G A T C A C G T A T T T A A G A S A A C A T A T B T S C S C G A A A A A G A C T T
G A S S T T G G T C A A A T G T A A A G C G T A A S T C K A K G S T T T T G C A A A S A T K G

521
462
511
508
504
522
467
484
487
502
550

4 p 3 1 8 2 7 5 10 9 8

T G C K A A T A T - - - A A G G C K A K A A C A K A A A T A T B T A A G T A A G G C C - C G C C
C A C C S T T A C C C C A A C C A A S T G T A C T A C A G S C C A T G S K T G A G T K C A S C A
C G T T S T G A T G A T G C A G C A K A C A G A C A T T S T A S A C G T G T C A G T K C A S T
C A C A G C A G A A G G C C T C A T G C T G C C A T A T A G A S S A A G T G T C G G T A A A
T A G A G C T G T G A C C A T A C C A C G A A C T A G G C A C T B S C T A G C G A G A G K C C A C T T
A T T T S S T C C A C A A G T G T G C T G T T A C K A A T T C C A T A K S C A T A A K - G A A A G
G G C A G A T A G C A T C A A K T A G G G T C T A A C C T A G T K S C C A A T A S G T C G S
T A T A G A G A G - C T G G A T C A C T G A G G A C C C C T A T T C S S T C G T T G T A C G S G
G A T A C T A C G T C A G T A A T C A C G A G C C A C G T T G C B C K G C G A G A G K C C A G T
T G C C A A G C A G G T G C C A C A B A G C C K S T T B C A A A A G T A G A G G K C C K A A
G C G G S C G G A C A G A A C G K T A C A C G T A C T C A A K A T G C A G C K T C G A T

567
511
560
558
554
571
537
543
547
552
600

4 C T A G A A A C G K A G T F G G K G C K A T T A T T A K A G T G K C T A T C G G A K A G C G K C C 817
 p A C C K A A C C K A C T T F G T G C K A C G A C T G C G T C A A T - A T C K C A A T C K A G C G K G C G 850
 3 C T T G T T A T K A G C C A G T K A T C C T T T A G G C T T A A - A G T A G G A G K G C G G K T A 809
 1 A C C C A A A K A B B T C A G T A T T T K A A A A G A T B C A A A T T A C G A G G A G G G T T 808
 B T A C K S T A G K A G C G F G K K A K B T A A A G A T T A T C G T G G C K S T G C G K A C A K C C G 804
 2 G T G C T T A G K A C T T F T A C C C K A T T T G G T C A C G G K C C G C G T T G K A A K A G C K 821
 7 T T A K A G K A F T T T F G K A C A T G G A T G T A C C A G - C C G K T A T A G K C G C A A C T 568
 5 C C T A G T G T K A A G C T F A T A K A K A G G G T A G A T C A T A T C G T A A A K T T G A C C T 583
 10 A C A K A T F T A A G T C G A T A T A K C C A A A A A T T T G T G G G A A G G T T G C C K A A K T 587
 B C T A C C A C A C C B C A A G K A A K T K S T A A G G C A G T G - - T K A G A T T T K A G T C T G 589
 6 C C C K A G T A T G C G C C S T G T T T T G A A C T T T A C A T K A A A G T G G A C K C G G K G G 850

4 A T A K A C C C A C K A T G A K G A A G S T G G A K A A C A K C G A A T G T G G G A G G A F A 867
 p C A C G G T C A C C K A A C C A C C A K G G G G G A G K A C T T C A C C G A G A C C G A C G T F A 810
 3 C T A T A T C T G T C A C T A K C C G K A G A C A K T G G G A G G G C A G A C A A C C T G T C 859
 1 A T G K A S T T C C K C T A T A K T A A C T G T K A A C T G T A S T G G C G C T A T T A A A K G C K 858
 B T T T K A C T G C T K C C G A A C G K A S T A A A A G T C A A T A A T G T A C G A T T A K C F T 854
 2 G T G K A T A T G A G C A A A A T K A T K A G A A C K A T G A G K G C C A T A G G G T T T A G C K A 871
 7 A G C G K A S T T G G K C G A T A C G C G K C A A C G A A C G T T A G A T G A A G A G G C T A C C C 838
 5 A A C T C C A T G G C C T T G G C G T G K T C A G G G A F T G G G K A C A T T T G A A C C K C F A 843
 10 C A G T C A G C G C T C G A G A A A T K C T A G G A C T T G G K A T A T G G G C T G C A K A A C 847
 B C C A K A T T T G T K S T T A C T C T G K A A G G A K A S T G G K A C G A C C C T C A G G F A 849
 6 A C A K A T C A A T T C T T A T G A C A C K T G G C G C K A C C A G A C T T G G C C G T A A G A A T 700

4 K T C A C A C A A A A K S A A T T G A A G C T T T T T A G C G A A C A F A A A C K A T T C A A G G 717
 p K A G A T G A F B G A G C C G T G G T T G A G C A G A T G T G T K T C C C A G T A C G A G A G G 860
 3 K T G G K A T T G G T K S A G T C T G A T A G A C C G A A C G G G C C K A F A G T A K G C G A T K A 709
 1 G A T T G T C A T A G G A T A T T G A A A T A G T A T T T G G G G C K A F S T G A K T C T T A K A G 708
 B K A A G A T G G G G K S T C G G A G A T A A A T T G C C A A A G G T C T A T G G T T T A G A K G 704
 2 K A T A C K T A T G K A S T C C A A T G T B A T T G C A C A A A K A C A C A A A C T T A A T A K A 721
 7 K A T A C K T A T T T C K A A T T C A T T A G S T C A A C A A T G A T A K A F T G A G G C T T T G T 868
 5 C A A C A T F G A A G C T C A G A G G T A S A G A A T T C G T C K A A G T A G A G T T A T G G A K A 883
 10 K C C A C A T F G T A C C T C T G T T C T T A T G A A T A A G S T T A T A G T A G A A C T T G G 887
 B K A T A C K T C C C T T K A C A C A A A A G B T C A A G G G T A G K A C K C F A G C K A C T G C K T 889
 6 G G C G T G A A A G C G A T A A A T G C C S A G G T A C T T C G K A C C A T T A G K T C G T G T A 750

4 G A G T A T T T G K A A T A K T A A C A T A A G G A A C G G K T A K A K S C F C T A G A A G A T C T 767
 p G A A T C T C A G G C C T A T T A C C A G A G A S G A T C G K A G C A T G G T C C T - - - - C T T 704
 3 A T C A A G A G G G T C G A G G A A G A T C A K A C C G C A K C C A T K A A C G G B C C F A C G G A 759
 1 T A G G C T C T A T C A A A K C S C C A T T A C T C A A A G C G C S T K A C G C A S T C F A G G A 758
 B A A C A G G A T A K T T C T G F G S T C A T A K A A A T C G G K C C A K A A A F S A A T G T C A A T T 754
 2 C G G C G C A G A K T G A C C C C A A G C G K A A G A A G A C C C F C C C C F S K A A A F T C A A A 771
 7 A C A G G A T C G A T G C C A T G C G G K C T A G A C A G A G C G C A A A T A G F T C C A C A 738
 5 G C A T C C T C A K A A A G C A A A T T A A T G A T G A T K A G G G G T C G A K C T C A T A G A A 743
 10 T C T A T G T G G G T T A K T T A T T A G C G G T A G A A K C S G A K G F S C A G A F T A T T T 747
 B A C G A A T T A T K A A A G T G S C A A A T C T T T A T A C K A C A G K C G T C T C A A G C C T A C 749
 6 T G A A G C A G T T T A C G K T G C A G G G A C C C T A C A K T C A T K A C T A C A A C G T G C A C 800

4 K T A C G C C C C A C A T T G A A C C T A T C S T T T C A A F A C A A T F S C T G G A G T A G F C A 817
 p C T C C F C T C A C C T G T G A T C C T C T G A T C T C F T C C T C A T C T T C C T S K T A G 754
 3 K A A T F A T C C A G G T G T A T G A B C G A G G T A C C F A S T G T A T C T T A A C T S S F S C 809
 1 C T T C T C A A G A T G T T G T G G C T S S T A G C A A T T F A T C G T A A T T C A G G A C T G C G 808
 B K A G T A T A A C T G A T A G T C G A A A A S A C A G A C A K A C G C F C C C G T G G A A G C T T 804
 2 T A C C C A C A A C T C T G A T C T C T B S T S S T T T T A F S T - - - - - 804
 7 C G G G F C A A T T A A C G A G T A G A S A S A T G A T C F A C C A T F S C G A T C A G S C F S A 788
 5 K C G A F A C G F T C A G A A A T A T B C T A A A A T G T A K A C A A F S G G A G T A C S T F S A 783
 10 K C A C C G T A G A G A G A G G C C G A A A S C C C C G T F C T C G G A G G A A T G G C T F T C 787
 B K T A F S C G C T G A A T A A C C A A G A S C G T T T G A G T A G C F C C G A G A C T S T C T T 789
 6 G G G A C C T G T G G T G A C T T G A C S C C T A C C A A A G A T C G F F G A T C G A A S K C T A 850

4 T G C G G T S A - - - - - 825
 p T G G G A T S A - - - - - 762
 3 C T C A G A T A G T C C A A A T G A A - - - - - 828
 1 C C C G A T S A A T G G A G G C C A A T G A - - - - - 830
 B C G T A C T T G C G T T T C T A T A T A A T G A T C A G G A G A G A A A A G T T C - 845
 2 - - - - - 804
 7 T A A G C A S A G A C - - - - - 787
 5 C A A G C C A T A A A T A T - - - - - 807
 10 T G - - - - - 789
 B A T G C G C S A C G C T T T T C A A A T C T C A C C T A C A G G C T A G G A C A G 841
 6 G T C T T G S F T C T C T A A T G G A T G A G A G G G - - - - - 877