Chapter 3

Reannotation of Foot-and-Mouth Disease Virus proteome

3.1. Introduction

Foot-and-Mouth Disease is a vesicular disease of cloven-hoofed animals and is caused by the Foot-and-Mouth Disease Virus (FMDV). It is a highly contagious and often fatal disease that infects economically important animals such as cattle and pigs. FMDV presents symptoms such as oral blisters and blistered hooves, which may result in lameness. In young animals infection can result in a miocarditis that can be fatal to the animal. Although most animals usually recover from FMDV infections, problems such as weight loss and swelling can continue for several months and this affects among others, milk production in cows, reduction in the availability of meat as well as affect working cattle used for ploughing in the African rural setting. FMDV is mostly transmitted via physical contact between animals kept in the same enclosure or via the clothes of the animal handlers.

FMDV occurs naturally throughout the world in wild populations but can cause economic problems when it infects domestic livestock populations (Fig. 3.1). FMDV infections can spread with great speed as seen in the outbreaks in the UK (Mason *et al.*, 2003*b*) in 2001. This outbreak resulted in an estimated loss of £4.1bn which illustrates the huge costs associated with FMD outbreaks.

FMDV is a small Aphthovirus that forms part of the Picornaviridae family (Levy *et al.*, 1994). It is non-enveloped and consists of an icosahedral capsid consisting of up to 60

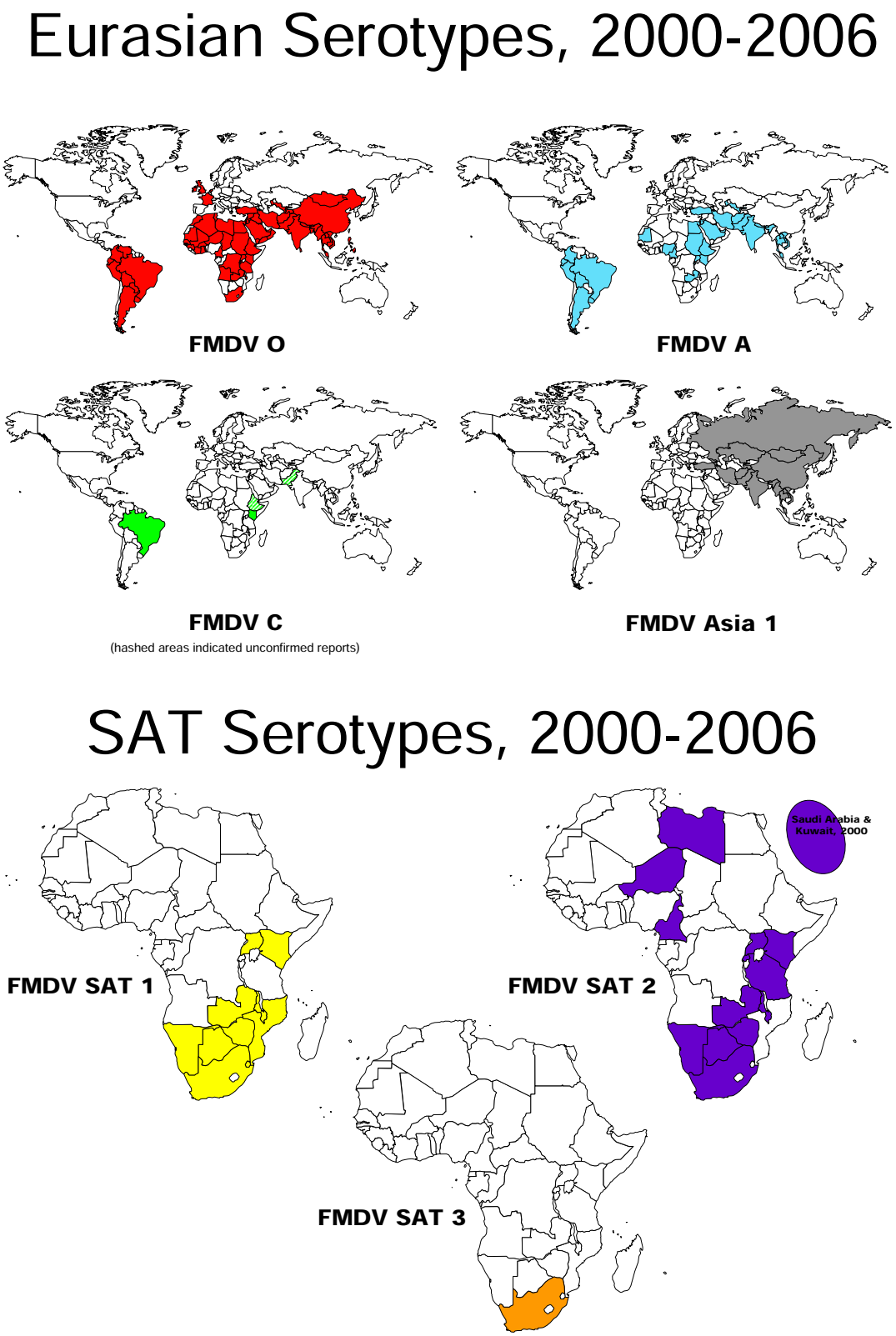

Figure 3.1: The distribution of FMDV outbreaks from 2000-2006 (FAO World Reference Laboratory for Foot-and-Mouth Disease, http://www.wrlfmd.org/maps/fmd_maps.htm). Top: Eurasian serotype outbreaks. Bottom: SAT serotype outbreaks.

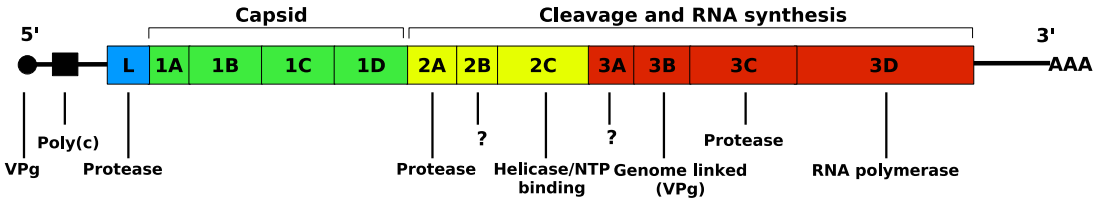

Figure 3.2: The genome organization of FMDV. It is divided into four basic sections. The 5' end is attached to the VPg protein and the 3' end is polyadenylated.

copies of four structural proteins. The structural aspects of FMDV will be discussed in more detail in chapters 4 and 5. The capsid contains a small 8.4 kb, single stranded RNA genome of positive polarity. In most cellular RNAs and some viral RNAs, a methylated G cap is usually found at the 5' terminus. In picornaviruses this is not the case and a VPg (3B) protein is bound to the 5' end (Fig. 3.2). This protein is 20-24 amino acids in length and is functionally, but not structurally, similar to several plant virus 5' terminal moieties. (Levy *et al.*, 1994). The virus also carries a polyadenylated tail at the 3' terminal. The length of this tail is encoded genetically and differs between the picornavirus members. This poly(A) tail is implicated in various roles related to genome replication.

The genome of FMDV is organized into a 5' untranslated region (5' UTR), an open reading frame (ORF) and a 3' UTR (Fig. 3.2). The ORF is divided into four basic regions: L, P1, P2, P3. The first section (L) encodes a protease that is responsible for early autocleavage of itself from the the polypeptide produced after translation. L^{pro} (Gradi *et al.*, 2003). In the L-coding region there are 2 AUG start codons. These code for proteins Lab and Lb. Both proteins appear to be present in the host but mutation studies have shown that Lb is vital to virus viability (Mason *et al.*, 2003*a*). Deletion studies have also shown that L^{pro} is needed for the virus to spread and infect its host. If L^{pro} is missing, the animal shows none of the symptoms typically associated with FMDV (Mason *et al.*, 2003*a*).

The second section produces four structural proteins (1A-D) and 2A. Post-translational cleavage by the 3C protease produces 1A-D that assembles into the icosahedral capsid. This capsid is unaffected by solvents such as ether and chloroform as there is no lipid membrane surrounding the virus (Levy *et al.*, 1994).

The third section produces three peptides after full cleavage, 2A-C. 2A seems to be an autoprotease that helps L^{pro} with early cleavage of cellular proteins and has some membrane binding ability. 2A is a short peptide consisting of only 18 residues. 2B enhances membrane permeability and blocks secretory pathways and seems to localize to sites of viral genome replication in vesicles derived from the ER (Carrillo *et al.*, 2005; Moffat *et al.*, 2005). It is also known to associate with the endoplasmic reticulum which is the site of virus genome replication. 2C appears to be associated with nucleotide binding (ATPase) and may have some helicase abilities (Mason *et al.*, 2003*a*). 2C has also been implicated in RNA synthesis initiation and localizes to virus replication vesicles. 2B and 2C are also implicated in virus-induced cytopathic effects.

The fourth section also produces 4 proteins after cleavage, namely 3A-D. The function of 3A is unknown but it seems to be involved in RNA replication (Mason *et al.*, 2003*a*) and may play a role in virus virulence (Carrillo *et al.*, 2005). Other studies have also shown that 3A directly associates with 3D and can function as a 3D co-factor (Hope *et al.*, 1997). In addition, previous studies have shown 3A to be the most invariable protein in FMDV (Carrillo *et al.*, 2005). 3A also forms a precursor with 3B i.e. 3AB, which has been implicated in RNA replication and supporting evidence comes from the fact that 3A fractionates with the ER membranes (Mason *et al.*, 2003*a*). FMDV contains 3 copies of 3B which is unique among the Picornaviridae. These 3 copies are referred to as 3B1 (23 aa), 3B2 (24 aa) and 3B3 (24 aa). The 3B becomes VPg after cleavage from a 3AB precursor. 3B appears to be associated with RNA replication, as the homologue in poliovirus helps to initiate genomic RNA synthesis (Carrillo *et al.*, 2005). Carillo and co-workers examined the variability in 3B and found that 3B1 and 3B2 are the most variable, and thus may play a role in host range and virulence. 3C is a protease of 213 amino acids, which helps to cleave the different precursor peptides from the main polypeptide produced during translation as well as cleaving host translation factors. The $3C^{pro}$ is responsible for ten of the thirteen cleavages of the polypeptide. Previous 3C studies have shown this protein to be conserved and thus have a limited tolerance for mutations (van Rensburg *et al.*, 2002). 3D is a virally encoded RNA dependant RNA polymerase (RdRp). It is the biggest protein encoded by the FMDV genome and is

comprised of 469 amino acids. It is also one of the most highly conserved sequences in the FMDV genome (Carrillo *et al.*, 2005). 3D is responsible for the elongation of nascent RNA strands during replication. 3C and 3D will be discussed in more detail in chapters 4 and 5.

FMDV exists as various subtypes even within a serotype, a likely consequence of the high mutation rate of the virus, and although some comparisons have been done between one or two viruses, there has been no detailed proteome comparison between the different serotypes. In this section various serotype proteomes were analyzed and compared to determine if there are any major protein differences or shifts in patterns in the sequences which may help to explain the phenotypic differences seen between the serotypes. These differences include effects such as host specificity, spreading and infection speed and virulence. By identifying the differences, it should be possible to map which areas are responsible for these effects. FMDV is a devastating disease and understanding how the proteins differ from serotype to serotype will help in unraveling the important regions in each protein. In this section four methods were used to characterize each protein. A Pfam family prediction was done to identify the family. This was followed by a Prosite pattern search. The absence or presence of certain patterns can help to explain differences seen between the various serotypes. It can also help to identify structurally important areas on a protein as these areas will be conserved throughout the various serotypes. A secondary structure prediction helped to identify areas that play a vital role on the structure of the protein. It has also assisted in identifying areas where variability has a possible effect on the structure, however small that might be. A final tool that was used were hydrophobic plots. As mentioned before, various of the FMDV proteins are membrane-associated and changes in hydrophobicity of a sequence may affect the association of these proteins with the various membranes.

3.2. Methods

Dr. F. Maree (ARC) supplied 3 proteomes for annotation (SAT1/SAR/09/81, SAT1/KNP /196/91, SAT2/ZIM/07/83) and 6 more were generated from genome sequences obtained

from Genbank (A24 (gi:46810792), A10 (gi:46810758), C3 (gi:46810870), O1/BFS/46 (gi:46810888), O/SAR/19/2000 (gi:30145780), SAT3/BEC/29 (gi:46810958)). Each proteome was split into its separate proteins: L, VP1, VP2, VP3, VP4, 2A, 2B, 2C, 3A, 3B1, 3B2, 3B3, 3C and 3D. All sequences are provided in the Appendix. Each protein was analyzed using the following programs: Pepwindow, garnier, Pfam and Prosite. Pepstats is part of the EMBOSS package (Rice *et al.* 2000) and calculates various protein statistics. Pepwindow is part of the EMBOSS package and was used to calculate protein hydropathy based on the Kyte-Doolittle parameters (Kyte and Doolittle, 1982). The hydrophobicity scale used is the same for every set of proteins and shows variation above and below 0, with 0 being neutral. Garnier is a secondary structure prediction tool incorporated into EMBOSS (Garnier *et al.*, 1978). Any secondary structure element longer than two residues was taken into consideration. Pfam (Finn *et al.* 2006) is a protein families database and contains Hidden Markov Models of each protein family. Hmmer (http://hmmer.janelia.org, as implemented in FunGIMS) was used to search protein sequences against the Pfam database (downloaded on 2008/05/8) with a 1e-03 cut-off value. Prosite (de Castro *et al.* 2006) is a database of patterns that identify proteins. The FunGIMS implementation of Prosite was used to scan each protein sequence.

3.3. Results and Discussion

Overall, the proteome annotation showed that the different subtypes within a serotype do not differ extensively yet local, protein specific or subtype-specific pattern changes were seen. Each set of protein sequences was submitted to the respective analysis methods. The results for each protein (L, VP1, etc.) were integrated to show any differences between the sequences (Figs. 3.3 - 3.13).

3.3.1. Pfam Results

The Pfam E-values of each protein is given in Table 3.1. The Pfam scan showed that all the proteins match the same Pfam family profile except in the case of the VP1 protein from SAT1/SAR/09/81. Upon closer inspection, it was seen that it matched the same

Pfam protein family as the other VP1 proteins but in this case it was above the cut-off of 1.0 e-03 (Table 3.1). Another interesting observation was that in VP3 (Fig. 3.6) the Pfam pattern had a far longer sequence length match in the SAT1/SAR/09/81, SAT1/KNP/196/91 and SAT3/BEC/29 subtypes. A similar situation was seen in VP1 (Fig. 3.4) where the SAT serotypes had the matching Pfam pattern split over two domains while the other serotypes had one domain match. A few proteins did not generate a match in the Pfam database. For protein 2A (Fig. 3.7) and 3B1-3 (Fig. 3.11) this is a result of their short length (about 20 amino acids in length) but for 2B (Fig. 3.8) and 3A (Fig. 3.10), each about 154 amino acids long, this is simply a matter of a lack of coverage in the Pfam database and a lack of general knowledge about the function of the protein in FMDV. The DUF1865 pattern match seen in VP4 (Fig. 3.7) is also a result of a lack of knowledge about the protein, but in this case it has already been assigned to a protein family of unknown function.

3.3.2. Prosite Results

As was to be expected, there were many Prosite hits due to certain amino acid patterns having a high probability of occurrence. Throughout most of the sequences the patterns appeared to be relatively conserved within serotypes e.g. the subtypes within SAT1 serotypes would have a certain pattern that differs slightly from the O subtypes (Figs. 3.3-3.6). It was decided not to exclude Prosite matches with a high probability of occurrence as these can provide clues to shifting patterns in the protein. There were a few interesting cases where patterns differed between proteins. The VP3 protein (Fig. 3.6) is an example of this. The VP3 protein varied from 221 to 222 amino acids in length for SAT1/3 and SAT2 isolates, respectively and with 58% overall variable aa positions. Most of the VP3 amino acid substitutions for SAT1, 2 and 3 were concentrated at four hypervariable regions, i.e. N-terminus (27-46), βB-βC loop (62-78), βE-βF loop (121-141) and βG-βH loop (165-183).

Certain matches are present in all the sequences (first two patterns) yet other patterns vary based on the genetic relatedness between the subtypes. In most of the proteins a definitive set of patterns was seen with small variations between the serotypes. An

Table 3.1: The Pfam pattern matches and E-values identified in each protein group. SAT1/KNP did not have a 3D sequence available.

		Pfam E-value								
Protein	Pfam Pattern	A24	A10	C3	O1/BFS	O/SAR	SAT1/SAR	SAT1/KNP	SAT2/ZIM	SAT3/BEC
L	Foot-and-mouth virus L-proteinase	2.2e-124	3.7e-128	8.7e-126	4.6e-130	2.2e-136	1.1e-129	1.1e-127	9.3e-128	7.7e-130
VP1	Picornavirus capsid protein	2.4e-26	4.1e-27	8.2e-25	3.7e-30	4.6e-23	Above cut-off	4.4e-05	2.9e-05	6.2e-08
VP2	Picornavirus capsid protein	4.2e-56	1.4e-56	6.2e-58	4.5e-56	1.6e-55	1.2e-42	1.3e-41	1.3e-43	8.4e-42
VP3	Picornavirus capsid protein	3.8e-41	8.9e-44	5.3e-33	3.5e-38	6.6e-38	3.3e-21	4.2e-21	1.7e-21	9.2e-25
VP4	Domain of unknown function (DUF1865)	8.8e-62	8.8e-62	3.6e-62	3.6e-62	3.6e-62	3.4e-61	3.4e-61	1.2e-60	8.5e-62
2A	None	-	-	-	-	-	-	-	-	-
2B	None	-	-	-	-	-	-	-	-	-
2C	RNA helicase	4.4e-23	4.4e-23	4.4e-23	4.4e-23	4.4e-23	7.3e-23	7.3e-23	4.4e-23	7.3e-23
3A	None	-	-	-	-	-	-	-	-	-
3B1	None	-	-	-	-	-	-	-	-	-
3B2	None	-	-	-	-	-	-	-	-	-
3B3	None	-	-	-	-	-	-	-	-	-
3C	3C cysteine protease (picornain 3C)	1.1e-80	4.8e-80	1.9e-79	2.8e-81	2.3e-79	1.5e-67	8e-69	8e-69	8.8e-68
3D	RNA dependent RNA polymerase	2.9e-162	1.4e-163	2.3e-162	1.4e-162	2.1e-161	9.2e-157	N/A	1.4e-155	2.4e-156

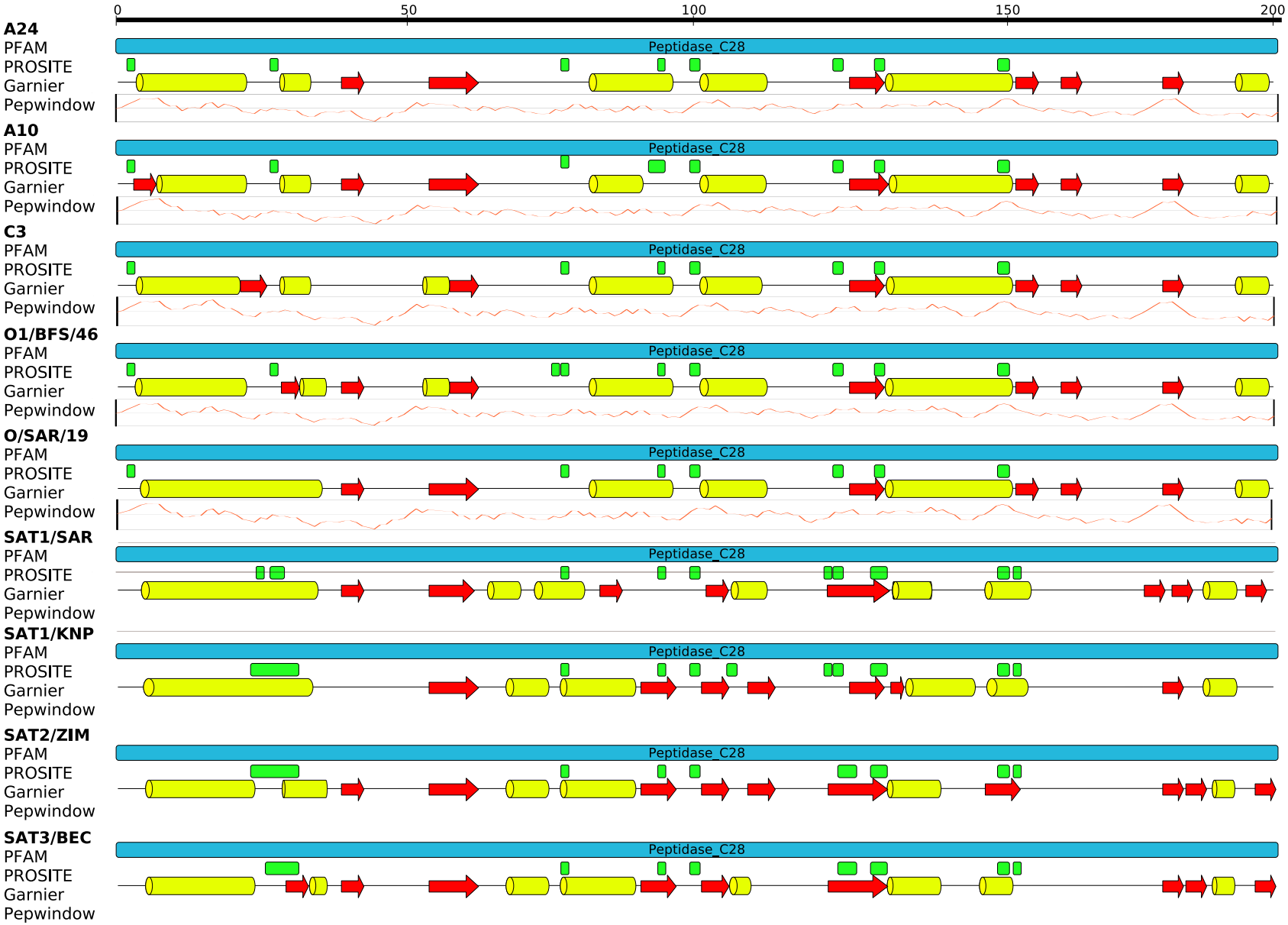

Figure 3.3: The annotation of protein L. α-helices are represented by cylinders and β-strands by red arrows.

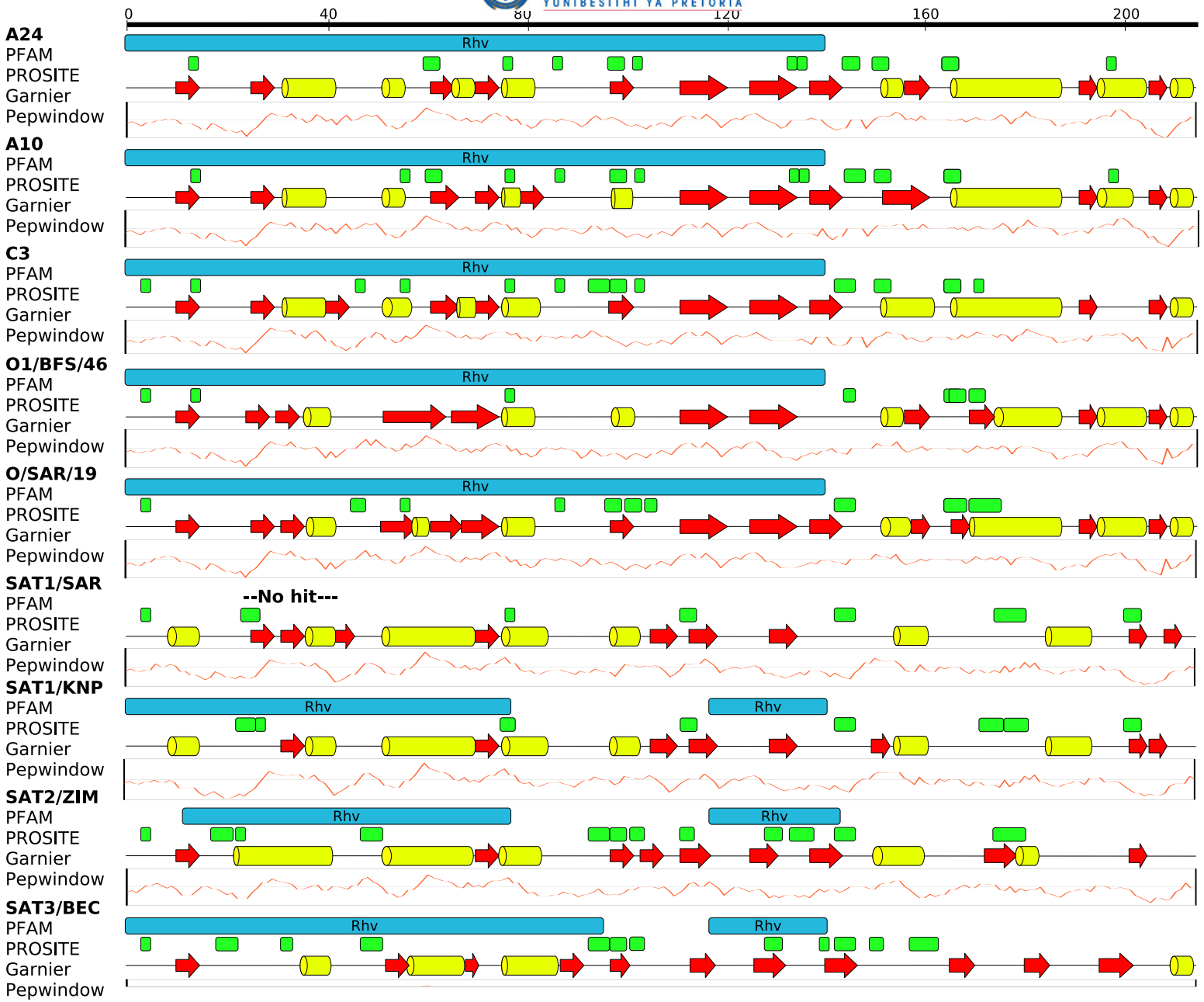

Figure 3.4: The annotation of protein VP1. α-helices are represented by cylinders and β-strands by red arrows.

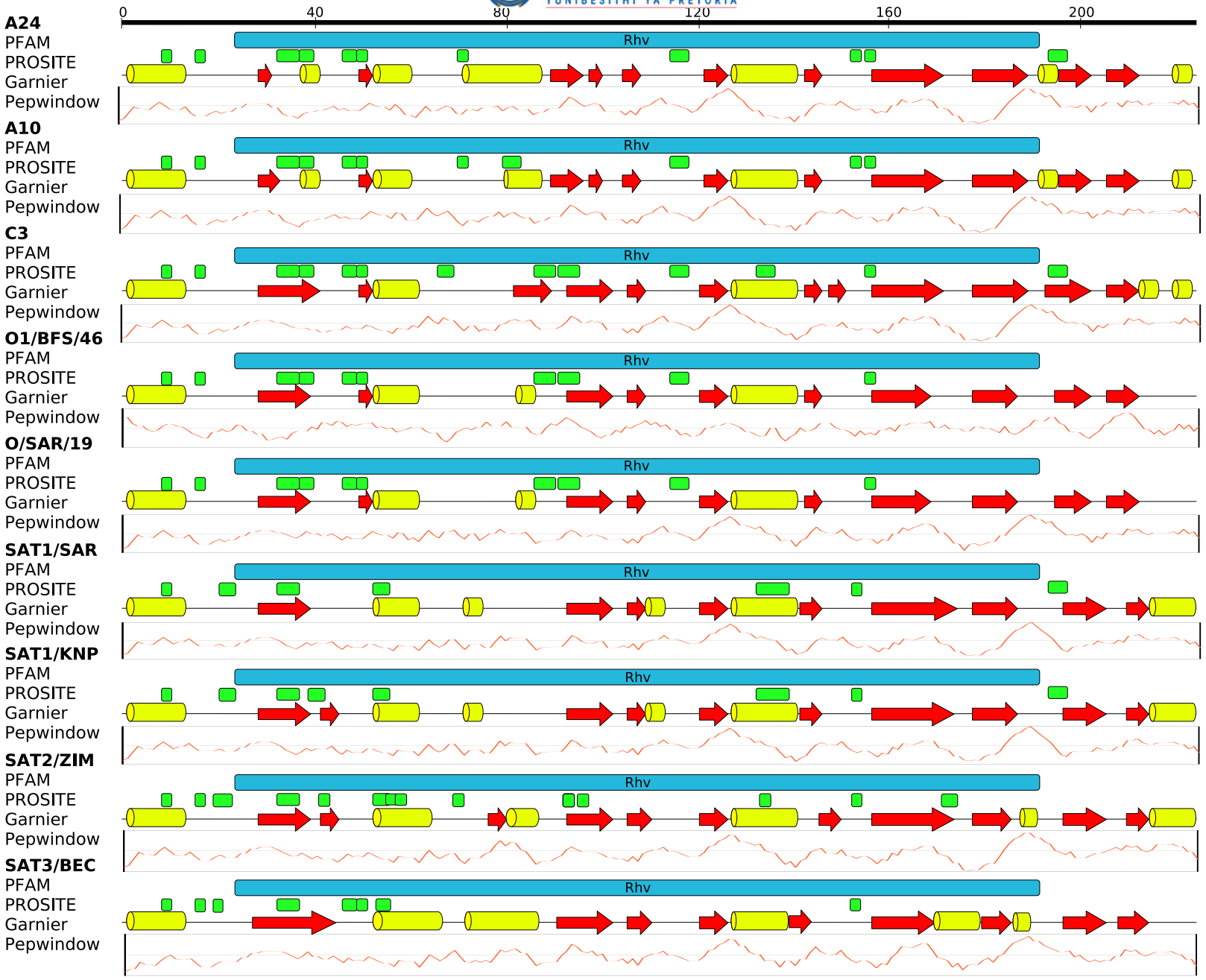

Figure 3.5: The annotation of protein VP2. α-helices are represented by cylinders and β-strands by red arrows.

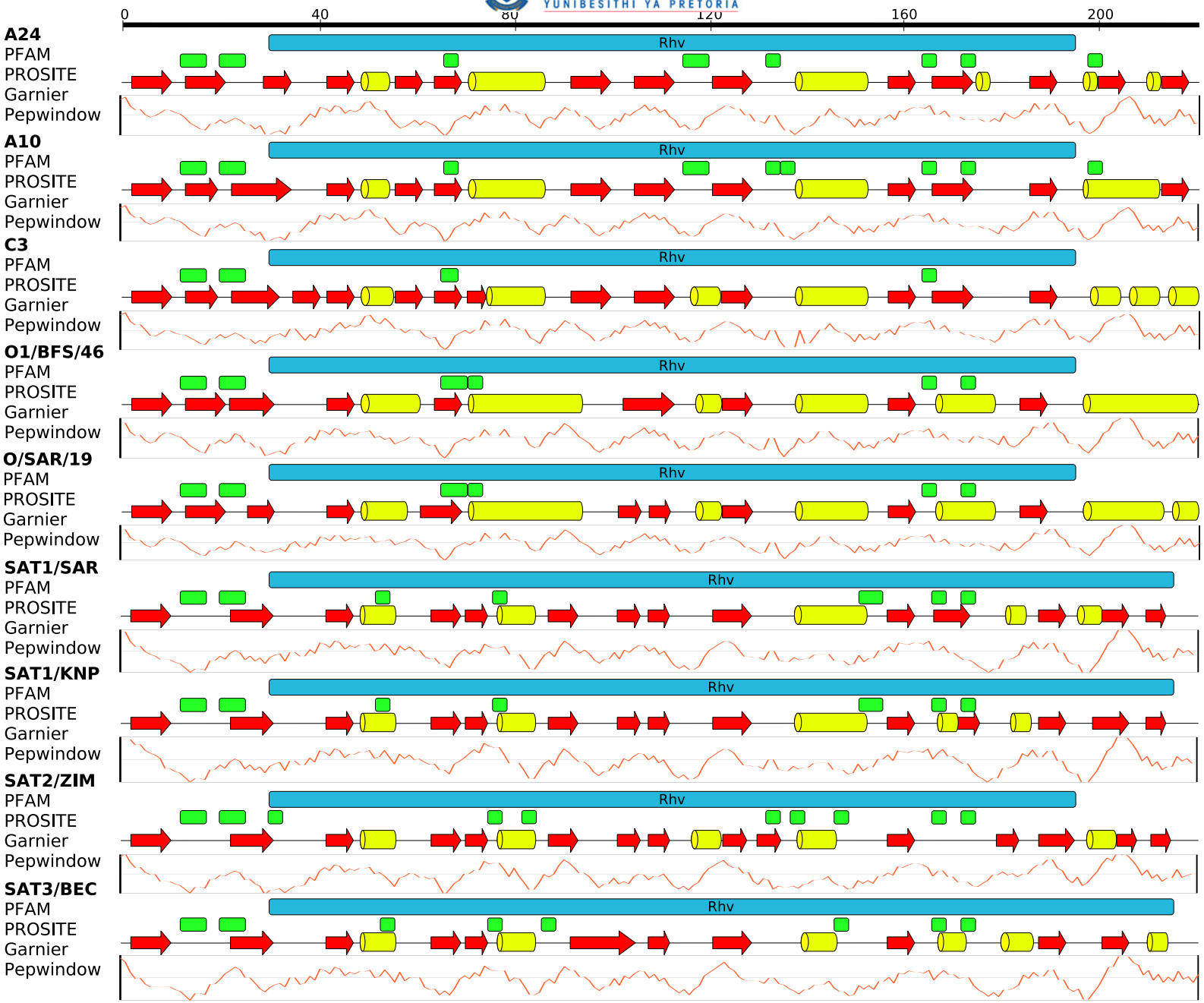

Figure 3.6: The annotation of protein VP3. α-helices are represented by cylinders and β-strands by red arrows.

example of this pattern conservation among subtypes can be seen in protein 2C (Fig. 3.9) where all the SAT serotypes share the same pattern. The SAT serotypes have an additional Prosite pattern match at the beginning and end of the sequence, which is not seen in the other serotypes analyzed. A clear pattern across all the proteins was seen for the SAT serotypes that confirms the close genetic relationship between the SAT1-3 non-structural protein coding regions. In most cases such as VP4 (Fig. 3.7) the SAT serotype displayed similar Prosite pattern hits that differ from the other serotypes. All the proteins showed a number of matches to many short patterns (3-6 residues in length) but in 2C a long pattern was found (Fig. 3.9). This pattern corresponds with the "Superfamily 3 helicase of positive ssRNA viruses domain profile". Another long pattern was found in the 3D protein (Fig. 3.13). A match to "RdRp of positive ssRNA viruses catalytic domain profile" was found, which is a RNA dependant RNA polymerase. A possible reason for these two long matches are the conserved nature of the proteins that are encoded by 2C and 3D. These proteins cannot accommodate many changes because of structural constraints and thus make it easier to construct a pattern match with a longer length.

3.3.3. Secondary Structure Results

The secondary structure prediction results showed that secondary structure is well conserved among the proteins but not as high as was expected. It was expected that the method would predict the same secondary structure for each sequence in a set, yet there were differences. This is possibly due to the method used, which is sequence-based. In most of the proteins the predicted secondary structure patterns stayed the same. In a few cases it was seen that an α-helix was split into two helices in another serotype as in the case of protein 2B (Fig. 3.8) or that an α-helix in one sequence is predicted to be a β-strand in another sequence (Fig. 3.3). Carillo and co-workers (Carrillo *et al.*, 2005) mention that a transmembrane region has been identified from position 120-140 but a transmembrane prediction using the Structural module showed no evidence of a transmembrane helix. However, hydrophobicity plots showed that the area from residue 120-140 is hydrophobic and may thus be associated with the membrane. A fact that

must be kept in mind is that secondary structure prediction is a sequence-based method and thus a one residue difference, such as a proline in the middle of a α- helix, may influence the algorithm and cause it to predict two separate helices instead of a longer, bent α-helix. This is also the possible cause of secondary structure being predicted as a α-helix in one serotype but in another serotype the same region is predicted to be a β-strand as seen in a comparison of 3B2 (Fig. 3.11). Carillo and co-workers reported on variation in three hypervariable regions in 3D (aa 1-12, 64-76 and 143-153, George *et al.*, 2001, Carrillo *et al.*, 2005). These areas were found to have a low variability in the proteomes examined here. This is reflected in the secondary structure predictions that predict the same structure for these areas in all the proteomes examined (Fig. 3.13). The Prosite patterns for the last two hypervariable regions are also the same, thus indicating low variation. The amino acid and Prosite pattern variation observed for the VP3 protein was also reflected in the secondary structure prediction. Similarly, VP1, the most variable of the outer capsid proteins, showed more variation in the secondary structure prediction. The VP1 protein varied in length from 213-214 aa for SAT2, 219 aa for SAT1 and 215-217 for SAT3 with 71% overall variable amino acid positions.

It must be kept in mind that the secondary structure predictions done here was to detect patterns in the sequences and not to get residue specific accurate predictions. There is currently no tool available which does such an accurate prediction of secondary structures. Moreover the sequences used here included local strains which have not been crystallized and thus no 3D data could be used to validate predictions. Main features such as a long α-helix or a sequence of helices or sheets seem to be conserved among the sequences, but short helices and strands seem to be conserved only among closely related serotypes. The results from the Garnier predictions showed that overall secondary structure patterns can be detected by the predictions, and predictions that differ across similar sequences must be investigated with further methods (either using structures or more advanced methods such HMMSTR (Bystroff *et al.*, 2000). Crystal structure data were not used in this section as the focus was on detecting pattern similarities/differences between the various strains.

3.3.4. Pepstat Hydrophobic Plot Results

The hydrophobicity plots for each set of sequences were kept on the same scale to allow comparison between plots. Each graph shown in Figures 3.3 to 3.13 have positive values indicating hydrophobicity and negative values indicating hydrophilicity below the line. The hydrophobicity plots showed, in contrast to the secondary structure predictions, that hydrophobicity remains mostly constant even though the sequence changes. Whereas the Garnier predictions made different predictions for a section based on the residues, the hydrophobicity plot was still the same indicating that there was some measure of structural integrity being maintained in spite of sequence differences. This was especially evident with the $3C^{pro}$ (Fig. 3.12). O1/BFS/46 VP2 (Fig. 3.5) showed one of the biggest shifts in hydrophobicity around residue 180. Whereas all the other sequences have a relatively hydrophilic stretch of residues, O1/BFS/46 appears to be very neutral in that region. This area was predicted to contain a β-strand by Garnier in all the sequences and may thus indicate a buried β-strand that can afford to be less hydrophilic. An interesting feature was also seen at the beginning (around residue 20) of the SAT VP3 sequences (Fig. 3.6). All the SAT serotypes are very hydrophilic at the start of the sequence, while the other serotypes show a slight increase in hydrophobicity in the same area. The SAT serotypes showed very similar hydrophobic plots as were seen for the secondary structure predictions and the Prosite pattern matches. This provides support for a possible ancestral sequence from which the SAT serotypes emerged.

An interesting feature was seen in VP2 (Fig. 3.5). Residues 30-40 were predicted to be a β-strand in C3, O1/BFS/46, O/SAR/19/2000 and SAT1-3 but in the A serotypes it was predicted to a be short β-strand and a short α-helix. Whereas the hydrophobic plots for the rest of the proteins in VP2 are the same, this area has a different plot for each serotype. A24 and A10 start out neutral from residues 30-35 and then turn fairly hydrophilic from residues 35-40. C3's plot is relatively neutral. O1/BFS/46 and O/SAR10/2000 differ. In O1/BFS/46 residues 30-40 is hydrophilic over most of the region whereas in O/SAR/19/2000 the region is far more neutral. The two SAT1 subtypes show the same pattern but the plots for SAT2/ZIM/07/83 and SAT3/BEC/29 appear more neutral for the area. The SAT serotypes all start out with a hydrophobic area from

residues 30-35 but then differ slightly from residues 35-40. Despite this difference the same Prosite pattern is conserved among all the sequences around position 35.

O1/BFS/46 shows another difference with the rest of the VP2 sequences. Overall VP2 from O1/BFS/46 is very neutral. If the hydrophobic plots are compared with the other sequences, it can be seen that O1/BFS/46 has none of the major hydrophobic plot spikes as seen at residues 120-130 and 185-195 in the other sequences. However O1/BFS/46 appears to have a unique hydrophobic area from residues 200-210 which is not seen in other subtypes.

The VP2 protein varied from 219 amino acids for SAT1 and SAT2 viruses and 218 amino acids for SAT3 viruses (52% overall variation within VP2) and the conserved N-terminal motif described by Carrillo *et al.* (2005) was supported in an alignment of SAT VP2 aa sequences, i.e. DKKTEETTLLEDRI(L/M/V)TT(S/R)H(G/N)TTT(S/T)TTQSSVG. In a structural model of the SAT type viruses this motif is located internally in the virion suggesting structural or functional constrains on this sequence and was recently mapped as a serotype-independent epitope (Filgueira *et al.*, 2000). Within the VP2 protein four hypervariable sites were identified, i.e. βA-βB loop (aa positions 31-44), βB-βC loop (aa 62-81), βC-βD loop (aa 91-101) and βE-βF loop (130-134/140 for SAT1 and 2, respectively).

3.4. Conclusion

Some authors have noted how variation in proteins such as L and 3A influence virulence and host range (Carrillo *et al.*, 2005; Mason *et al.*, 2003*a*). When looking at the annotation results, a clear picture emerges. There is variation, not only on a residue level, but also on a higher structural and potentially at a regulatory level, in almost all the proteins in the FMDV proteome. The main task now is to separate relevant and irrelevant variation. In this section global changes were looked at. Patterns such as Pfam only give a general idea of the function of the protein and thus are not as highly informative when looking at lower level differences. Lower level differences become obvious when Prosite patterns are looked at. As can be seen in the annotation results, some serotypes can be

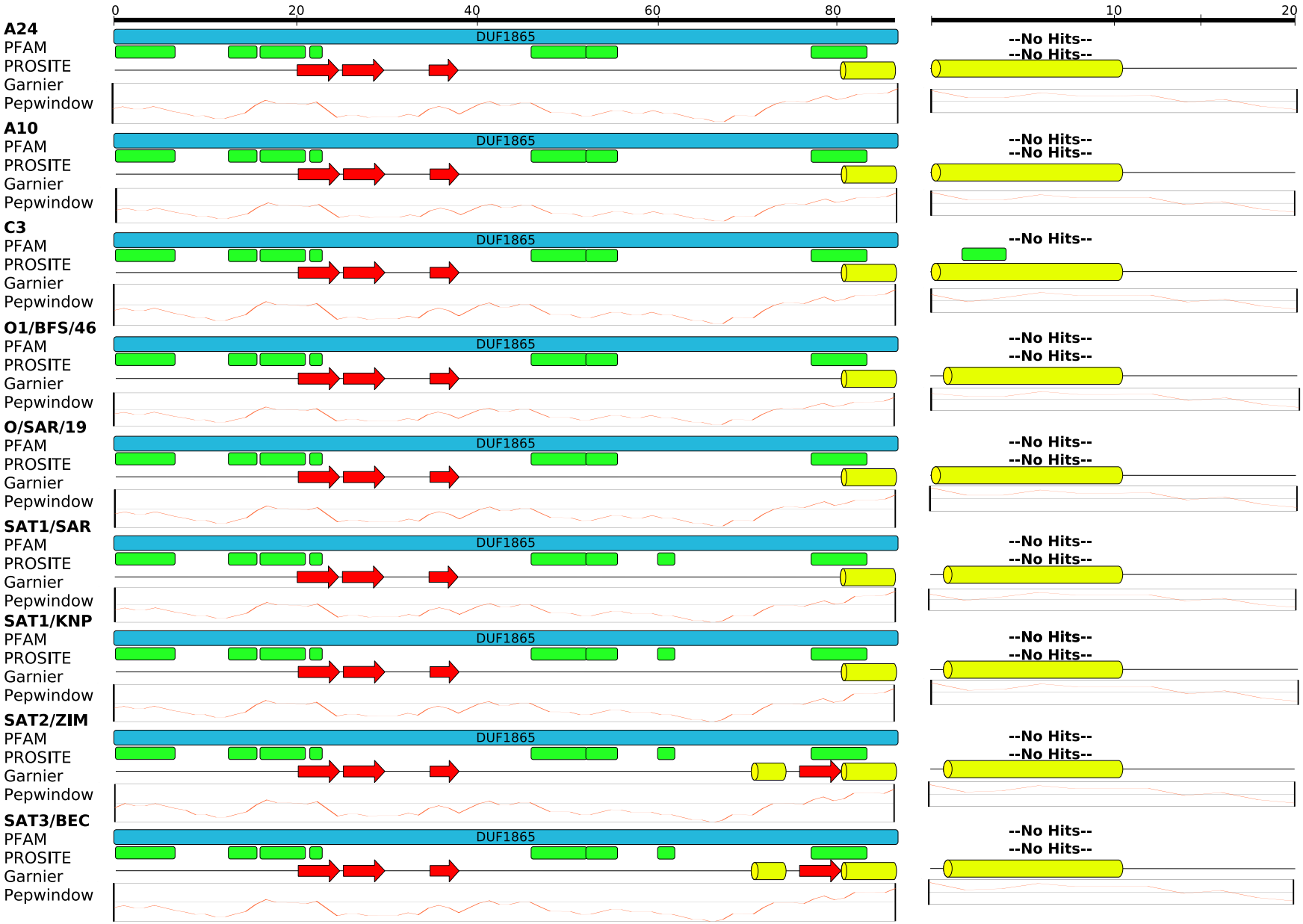

Figure 3.7: The annotation of protein VP4 (left) and 2A (right). α-helices are represented by cylinders and β-strands by red arrows.

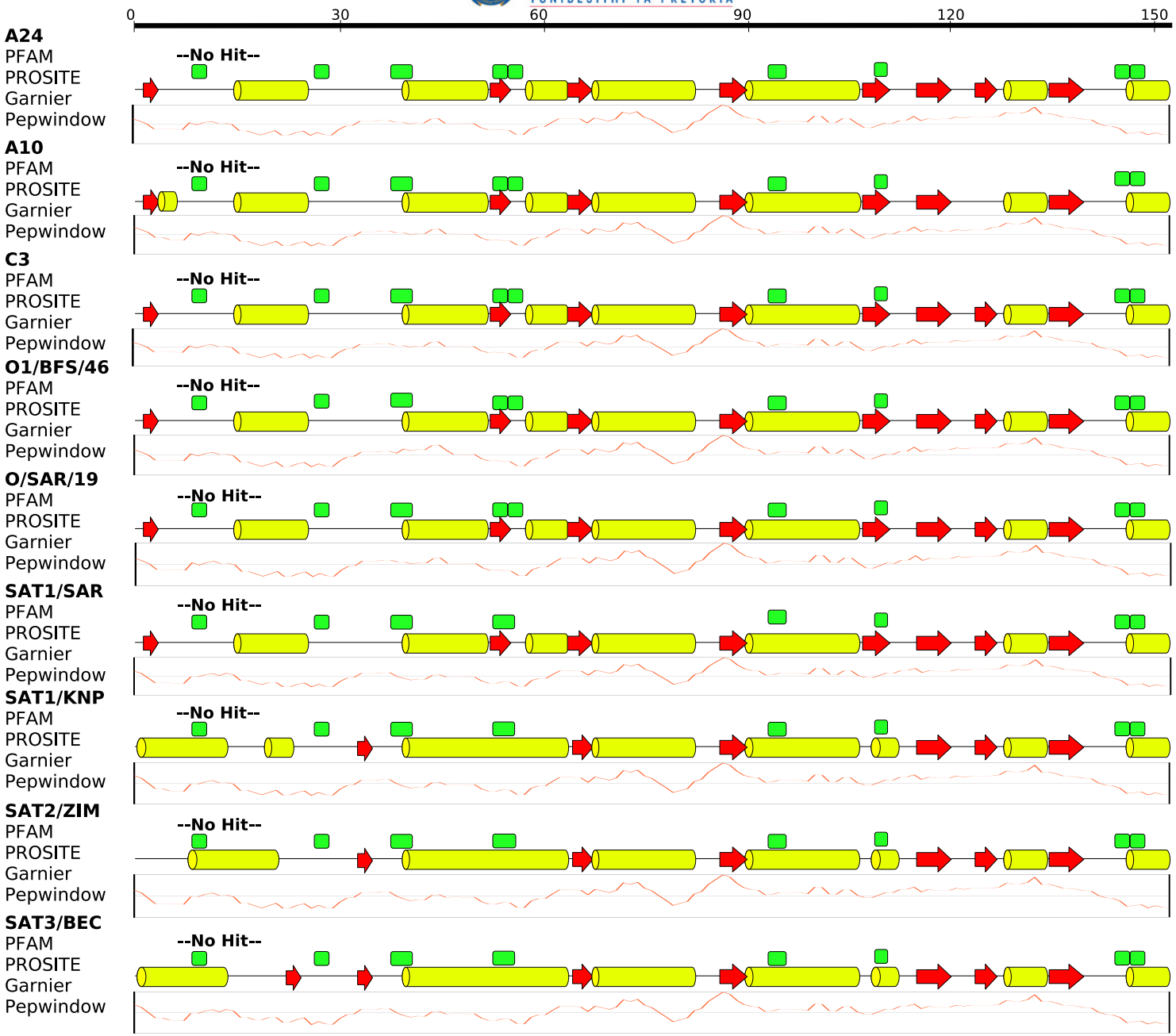

Figure 3.8: The annotation of protein 2B. α-helices are represented by cylinders and β-strands by red arrows.

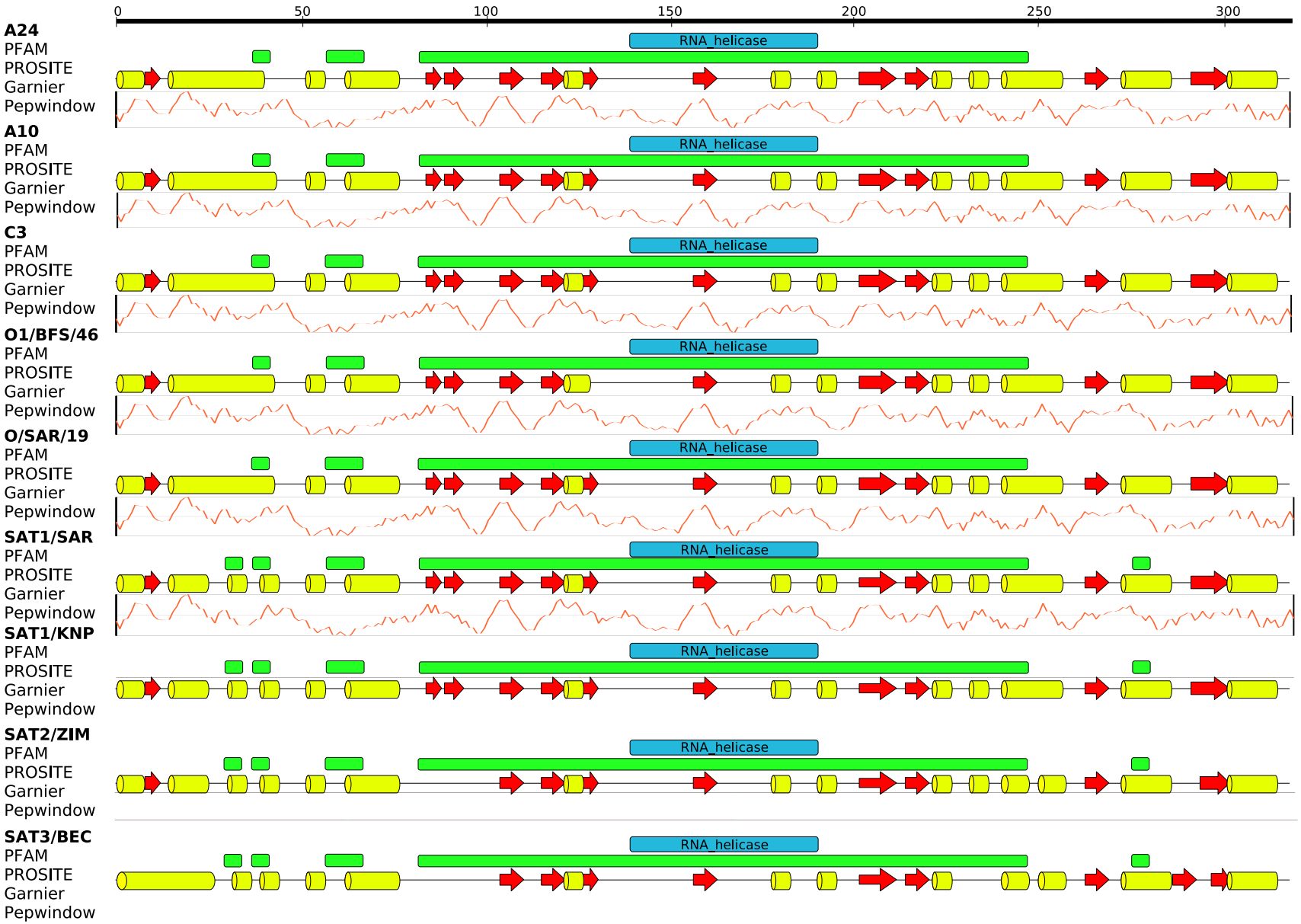

Figure 3.9: The annotation of protein 2C. α-helices are represented by cylinders and β-strands by red arrows.

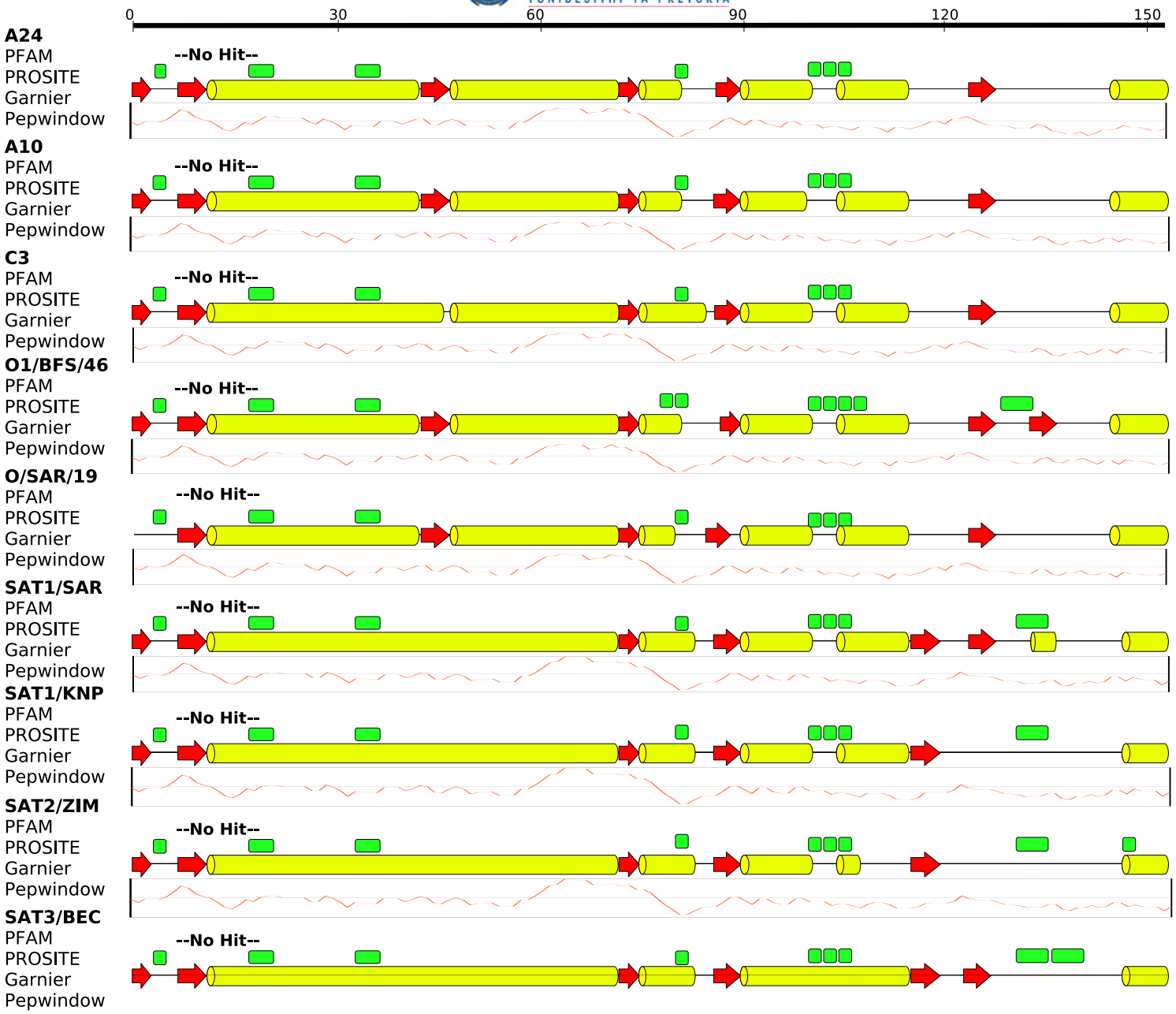

Figure 3.10: The annotation of protein 3A. α-helices are represented by cylinders and β-strands by red arrows.

Figure 3.11: The annotation of protein 3B. Left: 3B1; middle: 3B2; right: 3B3. α-helices are represented by cylinders and β-strands by red arrows.

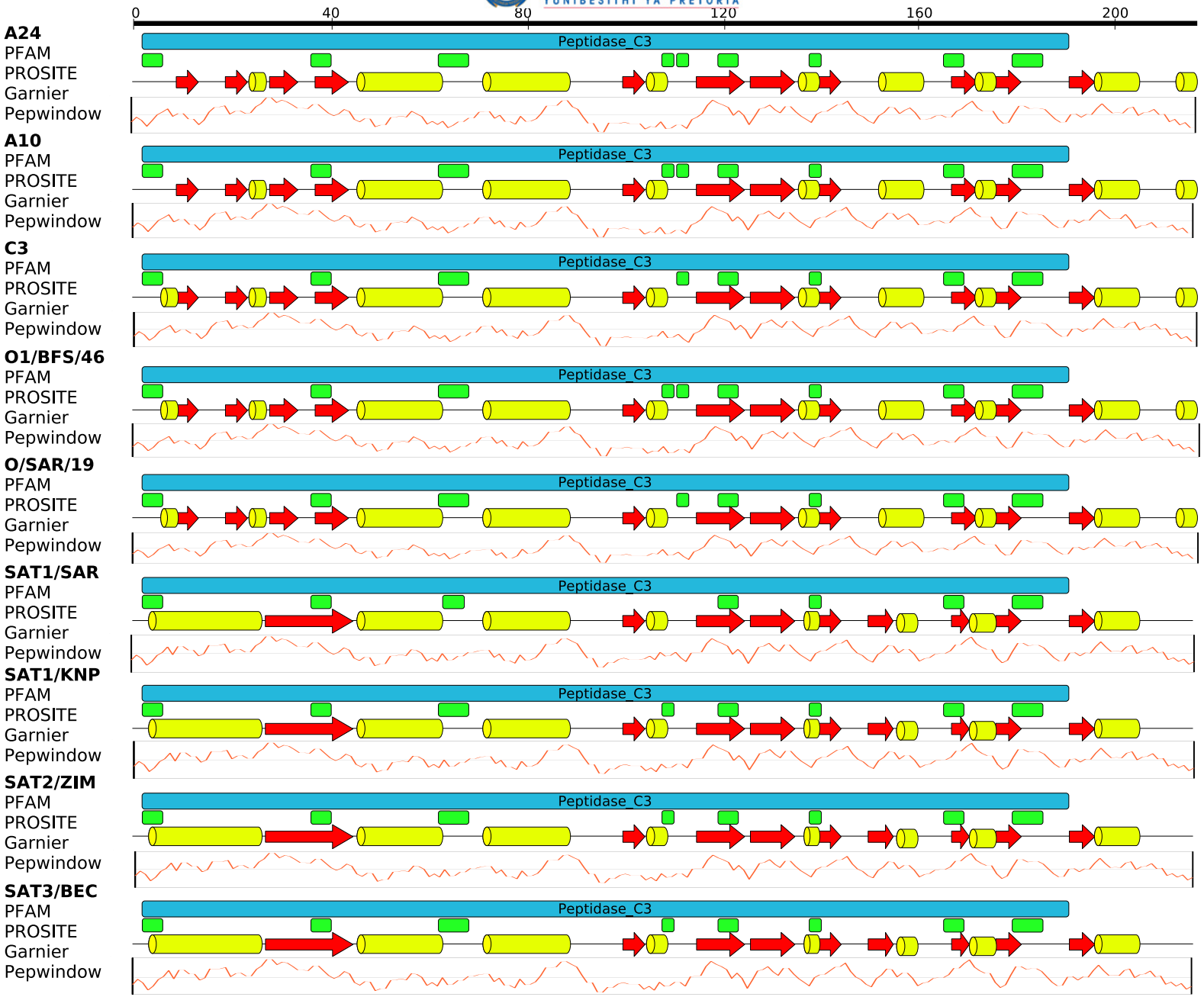

Figure 3.12: The annotation of protein 3C. α-helices are represented by cylinders and β-strands by red arrows.

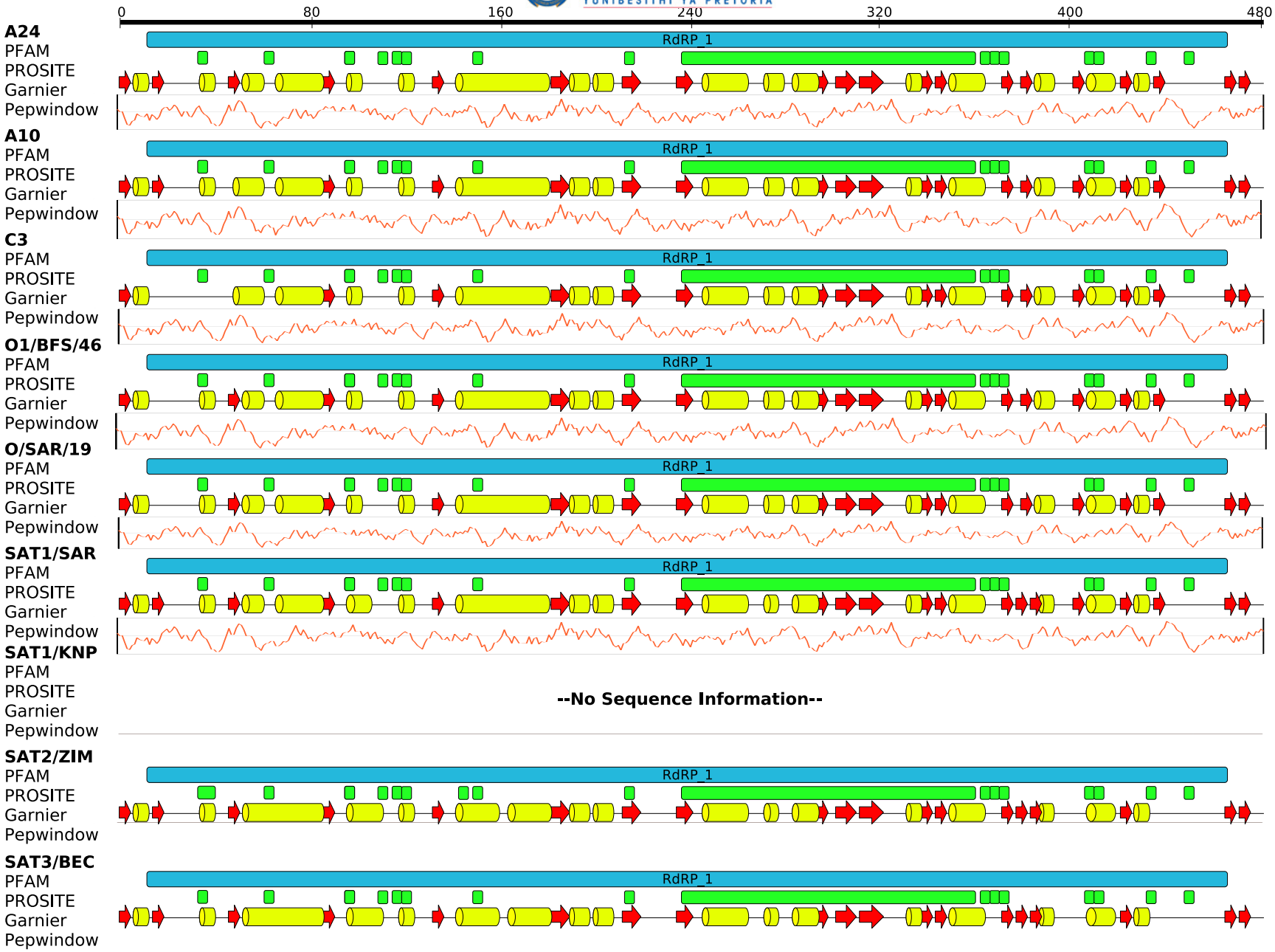

Figure 3.13: The annotation of protein 3D. α-helices are represented by cylinders and β-strands by red arrows.

grouped together on the basis of their distribution of Prosite patterns. The host uses some of these patterns for regulation and changes in the patterns may have an effect on the way the viral proteins function, on their activity or on protein-protein interactions. Changes in hydrophobicity patterns also affect the strength and affinity with which a protein, such as 2A and 3A, associates with the ER vesicle membranes and thus their duration of influence over RNA replication. A combination of all these factors may explain some of the differences seen in the host range, virulence and possibly even the spreading of the virus. The best approach to investigate these differences would be to make chimeras that contain conserved patterns found in every protein and thus determine which parts affects virus infection, translation and replication. A large-scale study involving all the sequences known for FMDV using the proteome annotation approach may yield valuable results, especially when coupled with epidemiology information such as virulence.

An important practical application of the proteome annotation is with regards to the substitution of structural proteins in the production of recombinant, chimeric viruses. The question becomes how much of the structural protein coding regions can be exchanged between serotypes in order to conserve structural constraints but be able to transfer the antigenic determinants to allow protection in the host animal. Previously it was shown that viable FMDV chimeras can be produced containing the complete or portions of the capsid coding sequence of different FMDV serotypes (Rieder *et al.*, 1994; Almeida *et al.*, 1998; van Rensburg and Mason, 2002). For example, the replacement of the pSAT2 (SAT2/ZIM/7/83) outer capsid sequences by those of A12 or SAT1/NAM/307/98 virus, rendered the resulting virus viable and stable during successive passages in BHK-21 cells (van Rensburg *et al.*, 2004; Storey *et al.*, 2007). The capsid and other sequences of the genome can be readily exchanged between serotypes and still rendered the chimeric viruses viable during successive passage *in vitro* (Almeida *et al.*, 1998; van Rensburg *et al.*, 2004; Storey *et al.*, 2007), implicating some pliability/versatility outside residues essential in the structural constraints of the virus particle. We have utilized the chimera technology in the development of recombinant FMDV vaccines specific for certain geographic locations. The virion stability, *in vitro* immunological profiles against a panel of reference sera and the receptor preferences were successfully transferred from the parental field viruses to

the chimeras with the substitution of the VP1, VP2 and VP3 coding regions (Blignaut *et al.*, unpublished; Maree *et al.*, unpublished). In addition, a chimera containing the outer capsid coding region of a SAT1 virus, KNP/196/91, in the genetic background of a SAT2 virus, ZIM/7/83, protected pigs against homologous KNP/196/91 challenge (Blignaut *et al.*, unpublished).

From the proteome analysis of the capsid-coding region it became clear that the structural proteins function as a unit, a fact that is supported by numerous recombinational studies. In these studies it was found that recombination rarely occured within the structural protein coding region, that breakpoint hotspots were detected at the 1A/1B and 1D/2AB boundaries and that hot spots on either side of the structural protein coding region function as a breakpoint pair (Jackson *et al.*, 2007; Heath *et al.*, 2006; Simmonds, 2006). Both the infrequency of recombination events within the structural protein coding region and the unique secondary structure prediction and hydrophobicity profiles in this study suggest that there are severe functional constraints limiting the exchange of structural protein coding regions between divergent parental viruses. This is mostly due to interaction patterns (hydrophobic as well as electrostatic) between the different proteins in the capsid. We predict that substitution of the VP2, VP3 and VP1-2A as a complete unit may allow the best success for recovery of viable viruses in the chimera vaccine technology. The work done here a starting point for the local researchers to start comparing phenotypic traits with patterns seen on the genomes of the various local SAT strains as well as assess how these strains compare with other serotypes.

Chapter 4 will deal with a more in-depth analysis of variation in FMDV 3C and 3D and their effect on the protein structure.