Journal of Virological Methods 297 (2021) 114266

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

# Design and structural bioinformatic analysis of polypeptide antigens useful for the SRLV serodiagnosis

Angela Ostuni [a,*], Magnus Monné [a], Maria Antonietta Crudele [a], Pier Luigi Cristinziano [a], Stefano Cecchini [a], Mario Amati [a], Jolanda De Vendel [b], Paolo Raimondi [b], Taxiarchis Chassalevris [c], Chrysostomos I. Dovas [c], Alfonso Bavoso [a]

[a] Department of Sciences, University of Basilicata, viale Ateneo Lucano 10, 85100, Potenza, Italy
[b] OneHEco APS, 84047, Capaccio Paestum, SA, Italy
[c] Diagnostic Laboratory, School of Veterinary Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, 11 Stavrou Voutyra Str., 54627, Thessaloniki, Greece

ARTICLE INFO

*Keywords:*
Protein clustering
Antigen design
Modelling and structural analysis
Molecular dynamics
Circular dichroism
ELISA test
SRLV serodiagnosis

ABSTRACT

Due to their intrinsic genetic, structural and phenotypic variability the Lentiviruses, and specifically small ruminant lentiviruses (SRLV), are considered viral quasispecies with a population structure that consists of extremely large numbers of variant genomes, termed mutant spectra or mutant cloud. Immunoenzymatic tests for SRLVs are available but the dynamic heterogeneity of the virus makes the development of a diagnostic "golden standard" extremely difficult. The ELISA reported in the literature have been obtained using proteins derived from a single strain or they are multi-strain based assay that may increase the sensitivity of the serological diagnosis. Hundreds of SRLV protein sequences derived from different viral strains are deposited in GenBank. The aim of this study is to verify if the database can be exploited with the help of bioinformatics in order to have a more systematic approach in the design of a set of representative protein antigens useful in the SRLV serodiagnosis. Clustering, molecular modelling, molecular dynamics, epitope predictions and aggregative/solubility predictions were the main bioinformatic tools used. This approach led to the design of SRLV antigenic proteins that were expressed by recombinant DNA technology using synthetic genes, analyzed by CD spectroscopy, tested by ELISA and preliminarily compared to currently commercially available detection kits.

## 1. Introduction

Infections by ovine and caprine lentiviruses, referred to as small ruminant lentiviruses (SRLV), cause slow-progressive, persistent and debilitating diseases that can lead to the death of the animal due to the failure of many organs. Herd productivity and international animal trade are also adversely affected. Two genetically, structurally and antigenically related viruses, members of the genus Lentivirus of the family Retroviridae, are in the SRLV group: Maedi Visna Virus (MVV) and Caprine Arthritis Encephalitis virus (CAEV) (Minguijón et al., 2015; Ramírez et al., 2013). In the past, sheep and goat lentivirus infections have been considered species-specific, with the MVV infecting the sheep and CAEV infecting the goats. More recently, sequence analyses and phylogenetic studies have concluded that the two apparently different viruses are part of a genetic continuum of lentiviral species (Leroux et al., 2010) and cross-species transmission has been demonstrated (Shah et al., 2004a,b; Minardi da Cruz et al., 2013; Fras et al., 2013).

SRLV have been classified in 5 genotypes or clades with 28 reported subtypes (Michiels et al., 2020; Molaee et al., 2020). Genotype A (subtypes A1-A22) corresponds to the original MMV, whereas genotype B (subtypes B1-B5) corresponds to the original CAEV; these two genotypes are widely distributed and comprise most of the subtypes. In addition, geographically restricted genotypes C, D and E have been reported (Reina et al., 2010; Shah et al., 2004a; Gjerset et al., 2006; Grego et al., 2007).

Diagnosis of SRLV infections can be made clinically, though only a small proportion of animals develop clinical signs. Serology represents a reliable and cost-effective method of diagnosing persistently SRLV infected animals (OIE, 2021, https://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.07.02-03_CAE_MV.pdf/). In the past

* Corresponding author.
*E-mail address:* angela.ostuni@unibas.it (A. Ostuni).

https://doi.org/10.1016/j.jviromet.2021.114266
Received 9 February 2021; Received in revised form 30 June 2021; Accepted 18 August 2021
Available online 26 August 2021

decades a number of methodologies have been developed for this purpose. These include the agar gel immunodiffusion (AGID), enzyme-linked immunosorbent assay (ELISA), radio-immunoprecipitation (RIPA), radioimmunoassay (RIA), and western blotting (WB) (de Andrés et al., 2005). Since no "gold standard'' method of diagnosis exists, it is quite usual that estimates of performance of a certain test, in particular sensitivity and specificity, are made relative to some other measure. The quality of the SRLV serological diagnosis generally depends on several factors: the format of the assay, the sequences and structural homologies between the strain of virus used in the assay and the strains of virus present in the testing populations, and the viral antigen used in the assay (OIE). Regarding the ELISA both indirect and competitive ELISA format have been developed and mainly three different recombinant protein antigens have been selected and used: the capsid protein p25 (CA), the transmembrane protein gp46 (TM), and the gp135 envelope protein. Synthetic peptides derived from p25 or TM as well whole virus antigens have also been used (OIE). The combination of the CA and TM antigens has shown to be quite effective in indirect ELISA and has been used in certain tests (Brinkhof et van Maanen, 2007; Saman et al., 1999).

However, due to the "quasispecies" nature of these viruses the high structural and immunogenic variability of SRLV represents an important limitation in the use of the ELISA. In fact, as stated above, the homology between the strain of virus used in the assay and the strains of virus present in the testing populations should always be taken into account. The ELISA reported in the literature have been obtained using proteins derived from a single strain in certain cases (Saman et al., 1999; Konishi et al., 2010); in another case it has been reported that a double-strain-based immunoassay may increase the sensitivity of serological diagnosis of SRLV infections (Grego et al., 2002).

Certain commercial kits use proteins derived from more than one strain apparently chosen on empirical bases since they represent the main genotypes but no detailed information is available, to the best of our knowledge. On the other hand, hundreds of SRLV protein sequences derived from different viral strains are deposited in GenBank. The scope of this investigation is to verify if the database can be exploited with the help of bioinformatics in order to have a more systematic approach in the design of a set of representative protein antigens useful in the SRLV serodiagnosis, since our ultimate goal is the development of an efficient and highly cost effective in-house indirect ELISA test. The main bioinformatic tools used in this investigation were: protein clustering, molecular modelling, molecular dynamics, epitope predictions and aggregative/solubility predictions. The conclusions from these analyses led to the design of SRLV antigenic proteins that were expressed recombinantly, analyzed by CD spectroscopy, tested by ELISA and preliminarily compared to currently commercially available detection kits.

## 2. Materials and methods

### 2.1. Sequence clustering analysis

The proteins investigated were the p25 capsid protein and the gp46 transmembrane (TM) protein that were chosen as suitable according to literature data (Saman et al., 1999). The exodomain of the TM protein was identified by using the program TMpred (Hofmann and Stoffel, 1993). Cd-Hit, a sequence-based clustering software (Huang et al., 2010), was used to analyze the protein sequences retrieved with BlastP from the Non-Redundant database.

Clustering can help to organize protein sequences into homologous and functionally similar groups. In our approach we make the reasonable assumption that a high immunological and epitope similarity could correspond to high protein sequence identity. A cluster is a set of sequences associated with one sequence, known as the centroid or representative sequence. Every sequence in the cluster must have an identity above a given threshold with the centroid. Therefore, clustering can reduce a large data set of sequences to a limited number of representative sequences.

In Cd-Hit sequences are first sorted in order of decreasing length. The longest one becomes the representative of the first cluster. Then each remaining sequence is compared to the existing representatives. If the identity with any representative is above a given threshold, it is included into that cluster. Otherwise a new cluster starts with it as representative. It is worth to remark that the sequences are processed in the order they appear in the input file. In our data set sequences of prototypic SRLV isolates were placed at the beginning of the file in order to verify if they were suitable as representative sequences. Three identity thresholds (80 %, 85 % and 90 %) and default input parameters were used in CD-Hit.

### 2.2. Homology modelling and structural analyses

The structural homology models of the three variants 1, 2 and 3 of the P25 domain of SRLV virus gag protein were built by Modeller (Fiser and Sali, 2003) with P26 of equine infectious anemia virus capsid (1EIA) (Jin et al., 1999) and the Cα-trace of hexameric P24 in HIV-1 (3GV2) (Pornillos et al., 2009) as templates. The structural homology models of trimeric TM_1 and TM_2 of the transmembrane protein of SRLV were built by Modeller using the closest related structures, i.e. the ectodomain of SIV gp41 (1QCE) and Visna TM core structure (1JEK) (Malashkevich et al., 2001), as templates. The structural restraints were used for intramolecular disulfide bond formation between C42 and C49 in each protomer. The TM_1_2 polypeptide contains the sequences of TM_1 and TM_2 joined by a short flexible linker. Two models of the multistrain TM_1_2 polypeptide were generated starting from the model of the trimeric TM_1. Chains A and B were joined (TM_1_2ab model) in one case and chains A and C (TM_1_2ac model) in the other case.

The top evaluated models were energy minimized with the YASARA energy minimization server (Krieger et al., 2009) and subsequently the protein geometry was evaluated by MolProbity (Williams et al., 2018). Structures were displayed and root mean square deviation (RMSD) calculated by PyMOL (Schrödinger, 2010).

Multiple sequence alignments of the protein sequences of the structural models and their homologous variants were generated by ClustalW in Seaview (Gouy et al., 2010) and sequence conservation was displayed on the models by ConSurf (Ashkenazy et al., 2016).

Linear B-cell epitopes of the sequences of P25-variants 1, 2 and 3 and as well as the two TM_1 and TM_2 were done with ABCpred (Saha and Raghava, 2006), BcePred (Saha and Raghava, 2004), BepiPred 2.0 (Larsen et al., 2006), COBEpro (Sweredoski and Baldi, 2009) and ElliPro linear (Ponomarenko et al., 2008). Discontinuous epitopes were predicted for the structural models of the same five proteins by using BEpro (Sweredoski and Baldi, 2008), DiscoTope 2.0 (Kringelum et al., 2012), ElliPro discontinuous (Ponomarenko et al., 2008), EPCES (Liang et al., 2009) and SEPPA 3.0 (Zhou et al., 2019). The epitope consensus prediction for each position was calculated on the basis of the results from the ten above-mentioned programs.

### 2.3. Molecular dynamics

The MD simulation was performed using Amber 12 software in the AMBER ff03 force field (Case et al., 2012). The systems were subjected to 100−200 ns simulations at 300 K temperature and 1 bar pressure, without restrictions, in a suitable explicit water box. The trajectory analysis was performed using a software developed in house. The RMSD is a measure of the spatial difference between two static structures, and in the MD simulation, the calculation was performed against the centroid (the mean structure trajectory) and all increasing time trajectory structures. The RMSF profile calculates the flexibility of a residue based on the fluctuation around an average position among all stuctures of an MD simulation. VMD software (http://www.ks.uiuc.edu/Research/vmd/) (Humphrey et al., 1996) was used to visualize protein behaviour throughout the simulation. Graphics were generated with Gnuplot 5.0 (http://www.gnuplot.info/).

### 2.4. Computational aggregation study

An in silico method, that can predict the aggregative properties of proteins starting from the modelled 3D structures, was used. Calculations were done with the Aggrescan3D software available at: http://bioc omp.chem.uw.edu.pl/A3D2/ web site (Zambrano et al., 2015). Structures were used with no further energy minimization and default input parameters were applied. Results were analyzed with the tools provided by the software.

### 2.5. Synthesis and purification of the polypeptides

All polypeptides were obtained by recombinant DNA technology in *Escherichia coli*. The related synthetic genes were inserted into cloning vectors (recombinant cloning vectors) purchased from Eurofins Genomics (Ebersberg, Germany) and subsequently subcloned into the HindIII- BamHI sites of plasmid pQE30 to obtained a protein with N-terminal His tag (MRGSHHHHHHGS).

All the proteins from lysed cells, were analyzed by 12 %-4% SDS-PAGE gel and polypeptides expression was confirmed by western blot analysis (Lee et al., 2014) with Monoclonal Anti-poly-Histidine-Peroxidase antibody (Sigma). Where indicated polypeptide was purified by HPLC (SHIMATZU-HPLC) on a reverse phase C4 column (Jupiter 5 µm, 300A,250 × 4.60 mm, Phenomenex). Purity analysis was made by SDS-PAGE gel and by HPLC on a reverse phase C18 column.

Polypeptides concentration were determined spectrophotometrically using extinction coefficients at 280 nm (https://web.expasy.org/protparam/).

### 2.6. CD spectroscopy analysis and CD spectra calculations

CD spectra were acquired at 25 °C with a Jasco J-815 CD Spectrometer with a light source constituted by a Xenon 250 W lamp and equipped with a thermoelectric temperature controller. The acquisition was carried out in a cylindrical quartz thermostated cell with an optical path of 0.1 cm in the spectral range 190–250 nm, a scanning speed of 100 nm/min, 1 nm bandwidth, a time-constant of 0.5 s, 20 mdeg of sensitivity and a total number of 16 accumulations for each spectrum. Then, the baseline spectra of the solvents were subtracted and spectra were smoothed using the Fourier transform. Data were expressed in terms of the molar ellipticity per residues in units of deg × $cm^2$×$dmol^{-1}$. Analysis of CD spectra for the evaluation of secondary structure content was performed with DICHROWEB using the SELCON 3 algorithm (Whitmore and Wallace, 2004; Micsonaia et al., 2015).

Circular dichroism spectra of the polypeptides were calculated with the software PDBMD2CD (Drew and Janes, 2020) at the web server https://pdbmd2cd.cryst.bbk.ac.uk/, starting from the molecular dynamic simulation data (e.g. MD trajectories) and generating spectra for each of the structures provided. 100 structures were sampled from each MD simulation and used for the CD spectra calculations. Calculated spectra were compared to the experimental ones using the same program PDBMD2CD.

### 2.7. Indirect ELISA

Sera used were obtained from sheep and goats of southern Italy (Basilicata and Campania regions) and Central Macedonia - Greece. All procedures were conducted in strict accordance with European legislation, regarding the protection of animals used for scientific purposes (European Directive 2010/63). Veterinary officers of the Italian National Health System collected blood samples during the compulsory, official eradication and surveillance programs on brucellosis of small ruminants, for which sheep and goats must be periodically sampled. Greek samples were collected as previously described (Chassalevris et al., 2020)

All the protein antigens were preliminary tested in Western Blotting and ELISA. A checkerboard titration was conducted to optimize the indirect ELISA reaction. Maxi Binding Immuno plates SPL (Life Sciences) were coated with polypeptides using a fixed mass ratio of 1:1:2 for P25_1, P25_2frag and TM_1_2 respectively at different total concentrations from 8 to 0.5 µg/mL in the coating buffer; 0.5 M carbonate buffer (pH 9.6), and then incubated at 4 °C overnight. Non-specific bindings were inhibited with blocking buffer 2% ovalbumin (Sigma) in PBS. Plates were washed twice with phosphate-buffered saline (PBS,pH 7.4). Sera were diluted with PBS Tween 0,05 % (PBS-T) at different dilutions from 1:25 to 1:1600. Protein G Peroxidase (Sigma) was diluted in (PBS-T) at different dilutions; 1:10000, 1:30000, 1:60000 and 1:120000 and added as secondary antibody. Then, tetramethylbenzidine (TMB, Sigma) was used as a substrate and the optical density (OD) read at 450 nm.

The checkerboard titration method was conducted using positive and negative samples. The diagnostic accuracy and the optimal cut-point value were evaluated by ROC analysis performed using the MedRoc software (https://stenstat.com/MedRoc/MedRoc.htm).

## 3. Results

### 3.1. Clustering analysis of the variable protein sequences of the SRLV P25 and TM proteins

Blastp retrieved 967 and 458 sequences for the SRLV P25 protein and for the 115-residue-polypeptide fragment of the SRLV TM (gp46) protein (sequence 680–794 envelope glycoprotein strain CAEV-cork accession P31626.1), respectively. In the Blastp analysis the lowest sequence identity was 70.7 % with respect to the query sequence (KV1772 strain) in the case of P25 and 67.8 % with respect to the query sequence (EV1 strain) in the case of the TM peptide. The 115-residue polypeptide represents the exodomain of the transmembrane protein and was chosen with the aid of a TMpred analysis. Sequences shorter than the full length one was also retrieved in the case of P25. Clustering results are summarized in Table 1.

In the case of the P25 at 90 % of sequence identity three main clusters were found. They are formed by 512, 390 and 34 sequences, respectively, and together represent 96 % of the retrieved P25 sequences. The representative sequences of the clusters are related to the accession numbers P03352.1 (Braun et al., 1987; Sonigo et al., 1985), ABU25359.1 (Grego et al., 2007) and ACV53612.1 (Reina et al., 2010), and belong to an A (strain 1514), B (strain It-007.5s03) and E (strain Seui) SRLV genotype, respectively. The clustering of the TM peptide at 85 % of sequence identity shows that two main clusters are present with 421 and 20 sequences, respectively, which represent 96 % of the retrieved TM polypeptide sequences. The representative sequences of the clusters are related to the accession numbers ABO32371.1 (Fraisier et al., 2007) and P31626.1 (Knowles et al., 1991; Saltarelli et al., 1990) and belong to an A (strain EV1) and B (strain Cork) genotypes, respectively. Clustering with a threshold value of 90 % resulted in a high number of clusters with the five most numerous ones representing 90 % of the retrieved sequences. Sequences to be expressed were chosen accordingly.

In conclusion this analysis reduced a large data set of sequences to a limited number of representative sequences that were used for the antigen design and development.

In particular three sequences were selected for P25 (named P25_1, P25_2 and P25_3) and two for the TM protein (named TM_1 and TM_2) with an identity level within each cluster of 90 % and 85 % respectively. Clusters corresponding to genotype C and D can also be identified in the analysis of the P25 sequence data set but contain only one sequence.

### 3.2. Structural bioinformatics analysis of the SRLV P25 and TM proteins

To simplify the set of polypeptides to be biosynthesized, in order to lower the development efforts and the production costs of a potential

**Table 1**
Clustering results for P25 and TM proteins retrieved with BLASTP.

**(a)**

| I.T. | Cluster1 | Cluster2 | Cluster3 | Cluster4 | Cluster5 | Cluster6 | Cluster7 | Cluster8 | Cluster9 |
|---|---|---|---|---|---|---|---|---|---|
| 80% | ABU25359.1<br>933 | ACV53612.1<br>34 | | | | | | | |
| 85% | ABU25359.1<br>702 | AGG09050.1<br>231 | ACV53612.1<br>34 | | | | | | |
| 90% | P03352.1<br>512 | ABU25359.1<br>390 | ACV53612.1<br>34 | AGH08192.1 *<br>14 | | | | | |

**(b)**

| I.T. | Cluster1 | Cluster2 | Cluster3 | Cluster4 | Cluster5 | Cluster6 | Cluster7 | Cluster8 | Cluster9 |
|---|---|---|---|---|---|---|---|---|---|
| 80% | ABO32371.1<br>428 | P31626.1<br>26 | AEF12562.1<br>2 | ACA81613.1<br>2 | AAG48632.1<br>1 | | | | |
| 85% | ABO32371.1<br>421 | P31626.1<br>20 | AEY84768.1<br>6 | AAS18424.1<br>4 | AEF12562.1<br>2 | ACA81613.1<br>2 | AYG99177.1<br>2 | AAG48632.1<br>1 | AMR71156.1<br>1 |
| 90% | AGS55105.1<br>254 | AGS55047.1<br>105 | ABO32371.1<br>40 | AYG99202.1<br>9 | P31626.1<br>7 | NP_040843.1<br>7 | AEY84760.1<br>7 | ALP75967.1<br>6 | AEY84767.1**<br>6 |

I.T. Identity Threshold. In green are highlighted the accession numbers related to the biosynthesized sequences. Under the accession numbers, the number of sequences in each cluster is indicated.
**(a)** Clustering results for the 967 sequences of the P25 protein. The Query sequence was the P25 (KV1772 strain). *There are other 6 clusters with all sequences that are shorter than 220 aa.
**(b)** Clustering results for the 459 sequences of the TM peptide. The Query sequence was TM_1 (EV1 strain). **There is another cluster with 4 sequences and other 12 clusters each with a number of sequences equal to or less than 2.

indirect ELISA test, a preliminary biostructural analysis and an epitope prediction was conducted on P25 and all the three complete sequences (P25_1, P25_2 and P25_3) were considered in the calculations. A similar analysis was performed for the transmembrane protein on the polypeptides TM_1, TM_2 and a multistrain TM_1_2 polypeptide. Molecular modelling and Molecular dynamics were the main tools used for the preliminary structural bioinformatics analysis.

### *3.2.1. Homology models of P25 (variants 1, 2 and 3), TM polypeptide (variants 1 and 2) and multistrain polypeptide TM_1_2*

Structural homology models of P25_1, P25_2 and P25_3 were generated to evaluate potential antibody-accessible epitopes. The structural templates for these homology models were P26 of EIAV (1EIA, 47, 50 and 47 % similar to P25_1, P25_2 and P25_3 respectively) and the hexameric form of the corresponding protein P24 of HIV (3GV2, 45, 47 and 44 % similar to P25_1, P25_2 and P25_3, respectively). The homology P25 models were analyzed in different aspects to evaluate if they were sound in terms of resemblance to the templates, protein geometry, location of polar, non-polar and conserved residues. The structural models of the P25-variants 1, 2 and 3 were superimposed on the templates and the RMSD were calculated showing values in the range 0.8–2.6 Å (Supplementary Fig. 1). These homology models display protein geometry comparable to the template 1EIA (Supplementary Table 1) and have mainly hydrophobic residues towards the core of the protein (Supplementary Fig. 2), where also the majority of evolutionary conserved residues are found (Supplementary Fig. 3). SRLV P25 models adopt the same domain folds reported for HIV-1 p24 (Momany et al., 1996), and EIAV p26 (Jin et al., 1999) with the first seven α-helices forming the larger N-terminal domain and the remaining four α-helices forming the C-terminal domain. A flexible linker joins the two domains. The major differences in the P25 models and the template structures are found in the longer loops, i.e. between helix 4 and 5 as well as in the linker between the domains, and towards the C-terminal end of the protein.

Regarding the gp46 (TM) exodomain as structural templates for the homology models the closest related 3D-structures (based on sequence similarity) were chosen: the trimeric helix-loop-helix structures of gp41 of HIV (1QCE, where only the helices are present, with 45 and 44 % sequence similarity with TM_1 and TM_2, respectively) and the TM of Visna Maedi (1JEK, where the loop connecting the two helices is missing, with 55 and 46 % sequence similarity with TM_1 and TM_2, respectively). Thus, also the obtained homology models consist of trimeric helix-loop-helix structures with a pseudo threefold symmetry axis. The TM homology models were analyzed in a similar way as the P25 models. The structural models of TM_1 and TM_2 were superimposed on the templates and RMSDs were measured and showed values in the range 0.4–2.2 Å with the major differences in the loop between the helices (Supplementary Fig. 4). However, it is noteworthy that also the two template structures differ with respect to each other, displaying an RMSD of 1.8 Å. The TM_1 and TM_2 models differ in the first few residues of the loop just after the N-terminal helix. This region as well as in the tip of the loop are also somewhat different in the homology models with respect to the template structure of HIV gp41. The homology models do not exhibit deteriorated protein geometry compared to the same statistics of the two templates (Supplementary Table 2). Moreover, the distribution of solvent-accessible and buried residues co-variates with the charged/polar and hydrophobic residues, respectively (Supplementary Fig. 5). The evolutionary conservation of residues within TM sequences in their positions in the structural homology models were analyzed and the results show that the residues involved in intramolecular helix-helix packing and intermolecular trimeric interaction are highly conserved whereas the solvent-exposed residues and the residues in the loop are variable (Supplementary Fig. 6). As above reported two different models were built for the multistrain polypeptide TM_1_2 using as template the TM_1 trimeric structure, due the pseudo-three-fold symmetry axis of the trimer. Both models were energy minimized and retain most of the structural features of the parent trimer model (Supplementary Fig. 7).

### *3.2.2. Molecular Dynamics (MD) of P25 (variants 1, 2 and 3), TM polypeptide (variants 1 and 2) and multistrain polypeptide TM_1_2*

Molecular dynamics simulations were performed to further validate

the homology models obtained with Modeller and to analyze their structural stability in different oligomeric states and in a fusion protein construct. Calculations were done for P25_1, TM_1 as monomer and as trimer and for the two models of the TM_1_2 multistrain polypeptide.

MD simulations of 200 nanoseconds (ns), were performed for P25_1, and TM_1 (both monomer and trimer) whereas 100 ns was the simulation time for TM_1_2. The dynamic behaviours were assessed by trajectory analysis. RMSD was used to evaluate the deviation of the predicted models from the original states during the simulation (Supplementary Fig. 8). High-magnitude RMSD fluctuations throughout the simulation can be an indication of a flexible and mobile natural protein / polypeptide. All MD simulations were visually inspected as well. According to the results of the analysis, the RMSD tends to stabilize at approximately 20−30 ns of simulation time for all the structures with the exception of the TM_1 monomer were larger fluctuations are observed during the entire simulation. Apart from TM_1 all the other

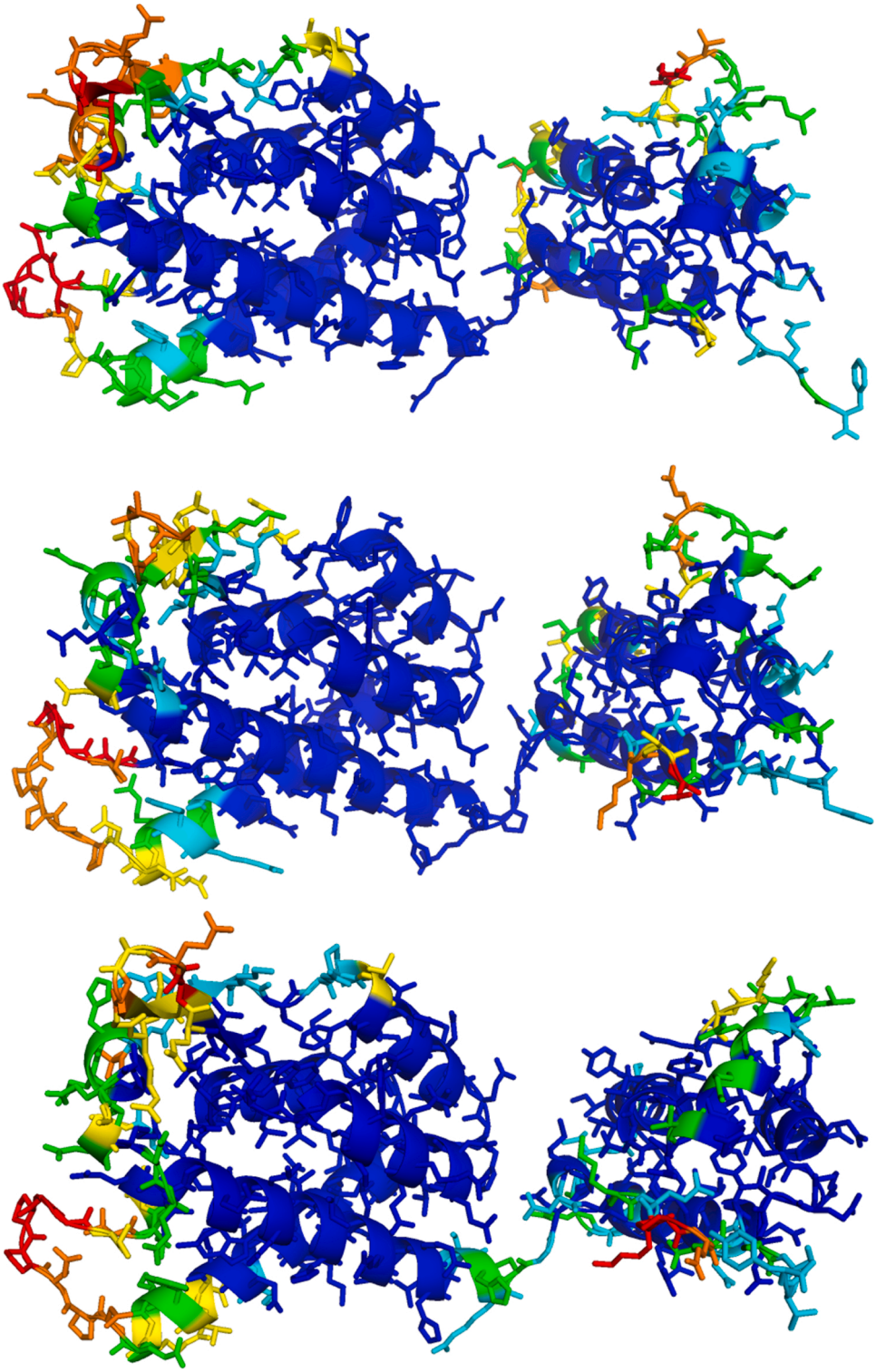

**Fig. 1.** The consensus predictions for linear and discontinuous epitopes are displayed on the homology models of P25_1 (top), P25_2 (middle) and P25_3 (bottom) in blue with the consensus intervals ≥90 % (red), 80-89 % (orange), 70-79 % (yellow), 60-69 % (green) and 50-59 % (cyan).

structures show, after the stabilization time, approximately a 2 nm deviation from the initial structure and a periodic fluctuation of low magnitude on the basis of this value.

Moreover, we evaluated the fluctuations resulting from movements of each of the residues in the protein, which highlights the most flexible chain segments (Supplementary Fig. 9). The root means square fluctuation (RMSF) analysis of each residue was performed to determine which residues may have given the main contributions to the RMSD values. In P25_1 all the residues show low mobility. TM_1 monomer has a high flexibility for many residues. For the other TM structures a significant flexibility is observed in the loop region and at the end of the C-terminal helices of the TM_1 trimer whereas TM_1_2ab and TM_1_2ac are quite flexible in the loop and in the linker.

As overall result, molecular dynamics simulation conducted on P25_1 confirms that homology models of P25 are quite stable as structures; in fact, no relevant variation from the initial structure are observed during the 200 ns of MD simulation of P25_1. Therefore, at least in terms of protein geometry as well as the positions of polar, non-polar and conserved residues, and dynamic behaviour the homology models of the P25-variants 1, 2 and 3 are reasonably acceptable. The 200 ns molecular dynamics simulation of the homology model of TM_1 in the trimeric structure demonstrate that it is quite stable whereas the 200 ns MD simulation of the TM_1 monomer shows that significant conformational distortions with respect to the starting structure appear at residues 16 and 27–28 of N-terminal α-helix and 75–77 in the C terminal α-helix. This may be due to the lack of favourable stabilizing intermolecular interactions as compared to the trimeric structure.

On the contrary MD simulations conducted starting from the TM_1_2 models do not show significant conformational variations for 100 ns, in this case the additional hydrophobic interactions between the two helix-loop-helix motifs may contribute to an effective stabilization of the structure with respect to the monomer. Model TM_1_2ac appears to be less flexible during the simulation than model TM_1_2ab. In conclusion our results indicate that the TM_1 polypeptide is quite stable in the trimer structure (that should correspond to the native structure) but flexible in the monomeric form (a similar result should be expected for the TM_2). The multistrain TM_1_2 polypeptide behaves like the trimer form in terms of flexibility in the MD simulation.

### *3.2.3. Antibody-accessible epitopes on the homology models of P25 (variants 1, 2 and 3) and TM polypeptide (variant 1 and 2)*

Potential B-cell epitopes in the sequences and on the homology models of the P25-variants 1, 2 and 3 were predicted using the sequence for the linear epitope and the modelled structures for the prediction of the discontinuous (conformational) epitopes (Supplementary Fig.10). The consensus epitope predictions mapped onto the models of the P25-variants 1, 2 and 3 show that the major epitope hotspots in the proteins are found in the more extended loops on the distal edges of the larger N-terminal and the smaller C-terminal domains and very few proximal to the domain interface area (Fig. 1).

The consensus of discontinuous epitope predictions displayed similar results (Supplementary Fig. 11) as those in using all programs. Thus, the P25-variants 1, 2 and 3 contain numerous potential epitopes including an immunodominant epitope experimentally found and previously

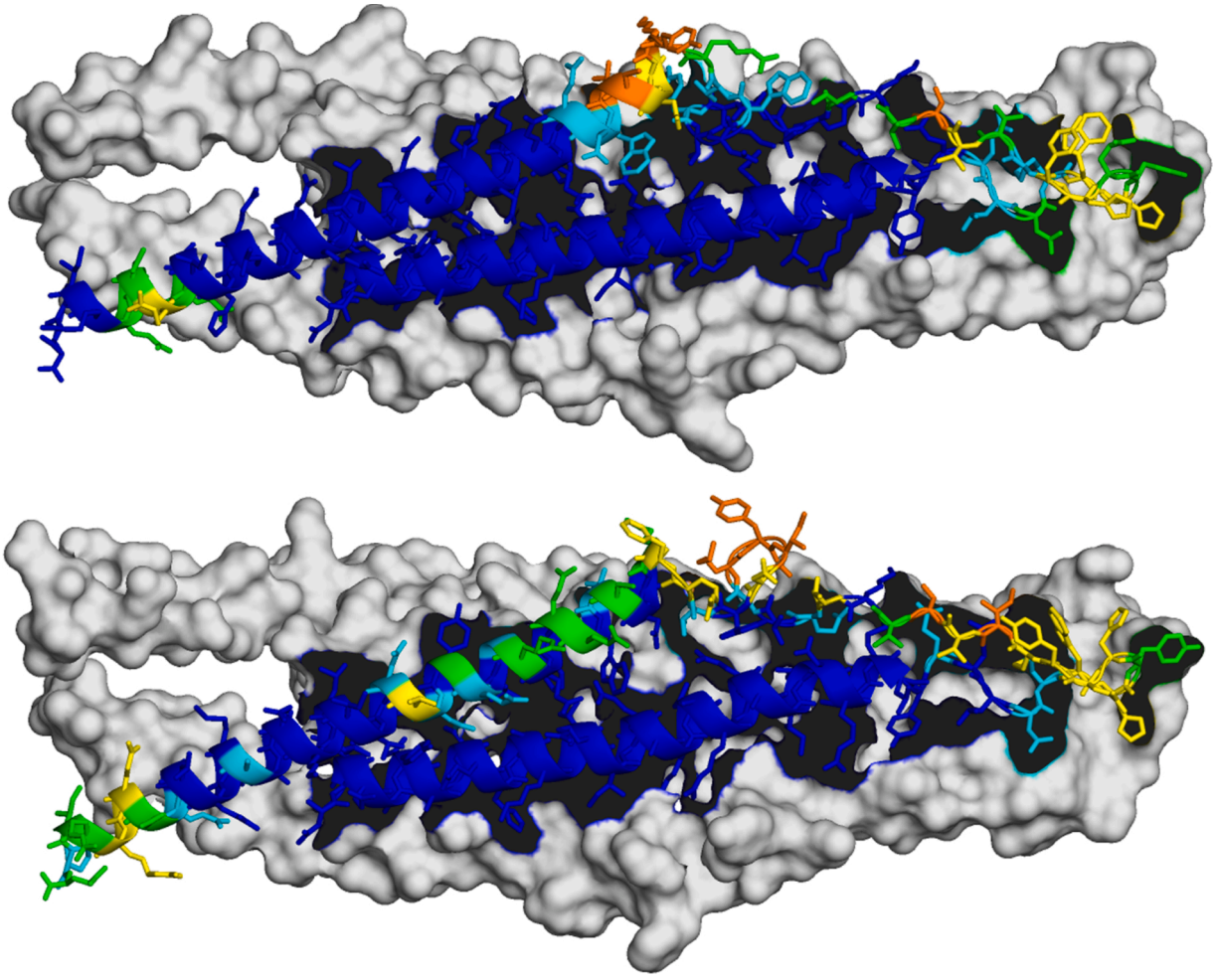

**Fig. 2.** The consensus predictions for linear and discontinuous epitopes are displayed on the homology models of TM_1 (top) and TM_2 (bottom) in blue with the consensus intervals ≥90 % (red), 80-89 % (orange), 70-79 % (yellow), 60-69 % (green) and 50-59 % (cyan).

reported in the literature (LNEEAERWVRQ) (Rosati et al., 1999). It is noteworthy that only one significant epitope is found in the N-terminal region of the protein.

Potential B-cell epitopes in the sequences and the homology models of TM_1 and TM_2 were predicted as in the case of P25 (Supplementary Fig. 12). The consensus epitope predictions were mapped onto the models of TM_1 and TM_2, showing that the major epitope hotspots in both proteins are located in the loop between the two α-helices and to some extent towards the C-terminal end of the second α-helix (Fig. 2).

The predicted epitopes consist of mainly exposed residues that would be accessible both in the monomer and dimer (TM_1_2) forms of the proteins. If considering the epitope consensus predictions of only the programs using the structural models, i.e. the predictions for the discontinuous epitopes, the results were similar to those of all epitope prediction programs (Supplementary Fig. 13). The differences in these results are mainly due to that the discontinuous epitope predictions also included exposed residues of the C-terminal a-helix as potential epitopes. In conclusion, these results demonstrate that several potential epitopes are found in the in TM_1 and TM_2. Results are partly in agreement with experimental findings by Bertoni et al. (Bertoni et al., 1994, 2000), however, our analysis did not find the first two epitopes in the N-terminal α-helix that were reported in their paper.

### 3.3. Design and production of the SRLV P25 and TM antigens

Based on the considerations from the sequence clustering and structural modelling analyses that are described in the previous sections, antigen constructs of the SRLV P25 and TM proteins were designed and expressed recombinantly and purified. P25_1 was wholly synthesized in order to have the complete set of epitopes of this protein whereas a multistrain polypeptide (P25_ 2_3frag) was designed and initially biosynthesized. In this case a suitable fragment was chosen from both proteins with amino acids lacking at the N-terminus (P25_2 and P25_3) and a short linker was used to join the two fragments. The specific sequence was chosen considering that, as noted above, most of the epitopes are in the central and C-terminal region of the protein. However, the chimeric P25_2_3frag was poorly expressed and difficult to purify, thus P25_2frag and P25_3frag were produced. Regarding the transmembrane protein three polypeptides were biosynthesized TM_1, TM_2 and the multistrain TM_1_2.

### 3.4. Secondary structure and aggregation analysis of the SRLV P25 and TM antigens

In order to evaluate if the secondary structure implied by the structural modelling of Section 3.2 is consistent with that of the recombinantly obtained antigens, we performed a spectroscopic study of the polypeptides and theoretical spectra were calculated for the structural homology models. CD spectra were recorded in TFE/water solutions at various concentrations and phosphate buffer (PB) solutions at room temperature. P25_1 is easily soluble in PB and its CD spectrum was qualitatively compared with the CD spectrum calculated from the MD trajectories using 100 structures sampled during the simulation. The calculated and experimental spectra appear qualitatively quite similar (Fig. 3), thus suggesting that the secondary structure of the structural models of P25 are coherent with the corresponding proteins.

Spectra were obtained for TM_1, TM_2 and the multistrain TM_1_2 in PB and TFE. Spectra of TM_1 and TM_2 in PB differ qualitatively from those measured in TFE (Supplementary Fig. 14). In particular the minimum at about 222 nm appears to be less deep in PB solution than in TFE for both TM_1 and TM_2, which most likely indicates a lower extent of α-helix structure. Moreover, the CD spectrum calculated from the MD simulation of TM_1 compares better with the spectrum in TFE (Supplementary Fig. 15).

On the contrary, spectra of the chimeric TM_1_2 taken in PB and TFE appear qualitatively quite similar to each other. Comparison of the CD spectrum calculated starting from the MD simulation of TM_1_2ac structure shows a reasonable qualitative agreement with both the spectrum in PB and TFE (Fig. 4).

However, in both cases the maximum at approximately 195 nm of the spectra appears to be lower than the corresponding maximum in the calculated spectrum. This may be due to random coil conformations that are not evidenced in the MD simulation since an analysis performed using Dichroweb indicates a contribution of 16 % and 24 % unordered conformation in the spectrum of TM_1_2 in TFE 70 % and PB, respectively. Since we observed a low solubility in PB especially for the TM_1 and the TM_2 polypeptides, we were prompted to investigate the aggregation propensity of TM_1 (monomer), and the multistrain TM_1_2 as compared to the TM_1 as trimer using a computational approach. In this respect, with the aid of Aggrescan3D, we qualitatively predicted the solubility of the proteins and polypeptides by using an in silico method that can anticipate the aggregative properties of protein and eventually

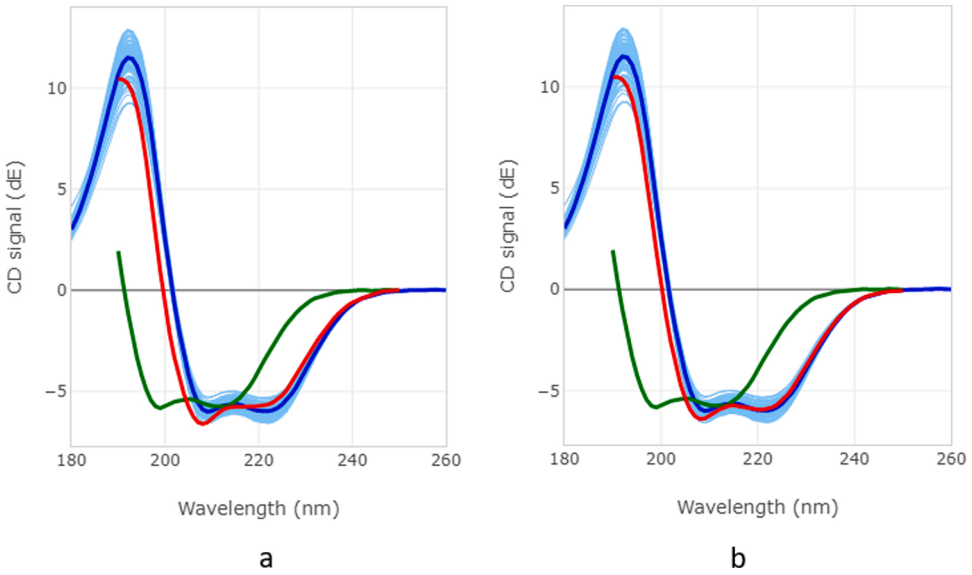


**Fig. 3.** Superimposition of the P25_1 CD spectrum (red) in water (a) and TFE 70 % in water (b) with the average spectrum calculated from the MD simulation (bold blue). The envelope of the spectra corresponding to 100 sampled structures is in light blue. In green is reported (shifted) the calculated spectrum closest to the experimental.

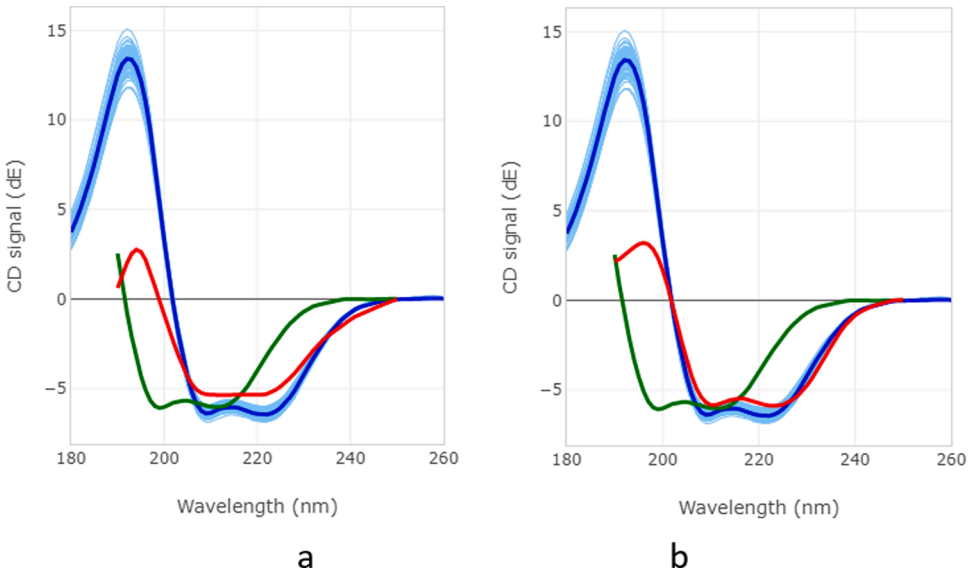


**Fig. 4.** Superimposition of the TM_1_2 CD spectrum (red in phosphate buffer (a) and TFE 70 % in water (b) with the average spectrum calculated from the MD simulation (bold blue). The envelope of the spectra corresponding to 100 sampled structures is in light blue. In green is reported (shifted) the calculated spectrum closest to the experimental.

assist the engineering of soluble protein. The aggregative properties of the molecular surface of the monomer and trimer TM_1, are shown in Fig. 5(a) and (b) respectively and that of the TM_1_2 polypeptide in Fig. 5 (c).

The trimer structure apart from the loop region, is lined with non-aggregative residues whereas in the case of the monomer the aggregative residues encompassing the loop and the N terminal α-helix are exposed to the solvent. In the multistrain TM_1_2 a number of

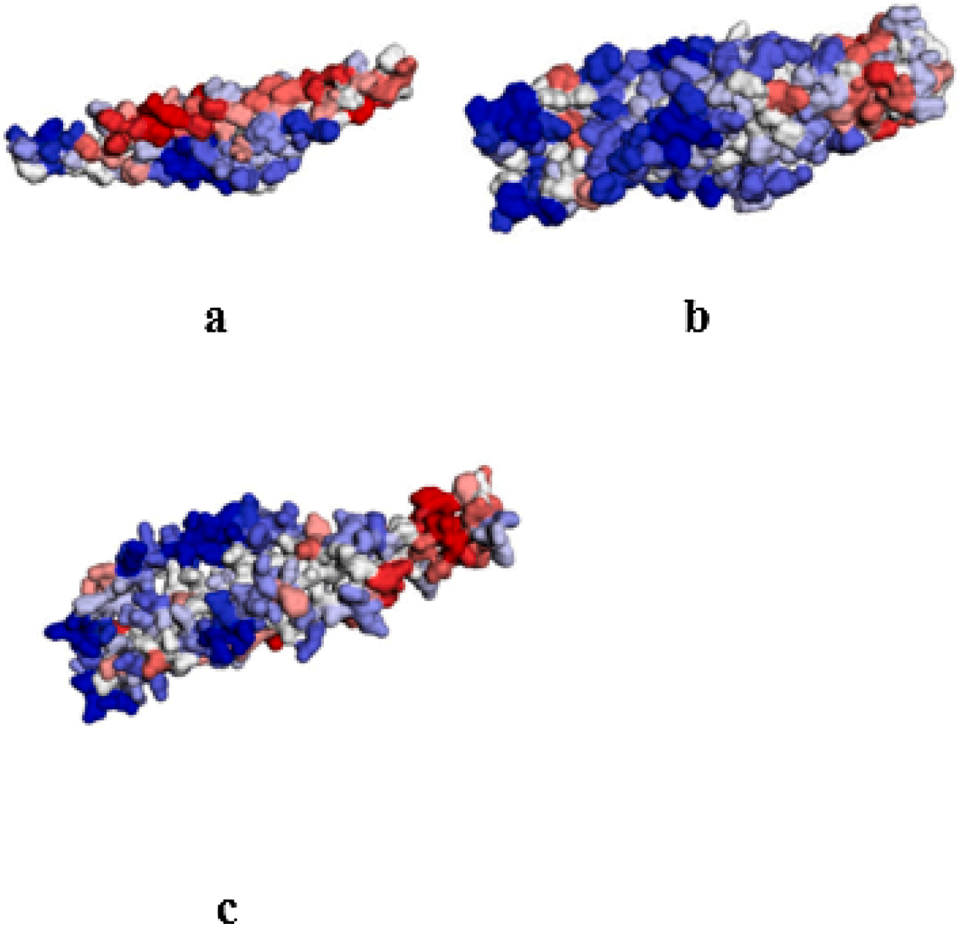


**Fig. 5.** Aggregative tendency of the molecular surface of TM_1 monomer (a), TM_1 trimer (b) and TM_1_2 (c) as evaluated with Aggrescan3d. High aggregative is red low aggregative is blue.

aggregative residues of the N-terminal α-helix are buried within the inner part the molecule. In conclusion this analysis qualitatively predicts a higher solubility for the trimer structure, a poor one for the monomer, with an intermediate value for the chimeric TM_1_2. This result could reasonable explain the findings derived from the CD analysis as compared to the MD simulations. This approach can be used in the future to try to avoid the biosynthesis of poorly soluble peptides.

### 3.5. ELISA to test the SRLV P25 and TM antigens

A preliminary ELISA test was set up using P25_1, P25_2frag and TM_1_2; P25_3frag (corresponding to genotype E) will be eventually introduced in future developments (data not shown). A checkerboard analysis was performed using a fixed mass ratio of 1:1:2 for P25_1, P25_2frag and TM_1_2 respectively, and the ELISA test conditions were optimized.

A panel of 47 samples of Greek origin previously tested with a commercial ELISA for SRLV and PCR, and giving identical results (Chassalevris et al., 2020), were retested with our in-house ELISA. A preliminary Roc analysis, showed a 96 % sensitivity and a 95 % specificity for our test (Supplementary Fig. 16). An additional comparison of 37 sera from southern Italy was done with another commercial indirect ELISA and a 100 % relative sensitivity and 100 % relative specificity were found.

## 4. Discussion and conclusions

RNA viruses generally exhibit heterogeneous and complex populations with similar but nonidentical genomes, since they evolve rapidly mainly due to the large population size, the high replication rate, the defective proofreading ability of their RNA-dependent RNA polymerase or their RNA-dependent DNA polymerase (Reverse Transcriptase) in the case of the retroviruses. The intrinsic genetic, structural and phenotypic variability of the Lentiviruses and specifically SRLV, has led numerous authors to classify them as viral quasispecies (Holland et al., 1992) with a population structure that consists of extremely large numbers of variant genomes, termed mutant spectra, or mutant clouds (each specific viral mutant is also sometimes referred as strain or clone). It is noteworthy that according to Holland and colleagues: “It is important to remember that every quasispecies genome swarm in an infected individual is unique and “new” in the sense that no identical population of genomes has ever existed before and none such will ever exist again”, i.e. the virus is constantly exploring the evolutionary sequence space.

Immunoenzymatic tests for SRLVs are available but the dynamic heterogeneity of the virus makes the development of a diagnostic “golden standard" extremely difficult. However, one can envisage strategies to adapt continuously and improve the performances of the diagnostic tests, as sequence data accumulate due to the intrinsic nature of a “viral quasispecies “. This implies that a sort of dynamic and flexible approach should, in principle, be considered in the design of optimized viral antigens used in the tests.

The starting point should be a reasonable description of the “mutant spectra” i.e. the sequence space explored by the viral quasispecies. The DNA sequences deposited in the data banks are the only available source to derive the protein sequence variability and for extracting consensus sequences to identify major viral sequences, to be exploited for DNA amplification both for diagnostic purposes and for traditional molecular cloning with the aim of expressing the proteins. Starting data could be biased due a number of factors including the very limited number of deposited sequences with respect to the heterogeneity of the viral population, the uneven geographic distribution with respect to the livestock geography and, taking into account the relatively fast evolution of the viral quasispecies, the potentially not updated nature of the data set with respect to the current viral strains distribution. Therefore, it is clear that improving the representativeness and coverage of SRLV sequences in the existing data bases is an auspicable future goal. A promising approach to tackle the problem with high sequence variability is presented in the work of Colitti et al. (2019) with Next Generation Sequencing (NGS) technology introduced to characterize SRLV isolates completely.

In any case, it is expected that relatively large sets of sequences have to be analyzed, therefore, algorithms to reduce the data set to a reasonable number of representative protein sequences are very useful. Sequence clustering with an acceptable threshold identity seems quite appropriate and appears to be the simplest approach. However, other more sophisticated approaches could be explored, for example a combination of a preliminary sequence clustering with a very high identity threshold (for example greater than 95–98 %), that results in high number of selected sequences, followed by structural clustering of the corresponding modelled structures. This would require the automation of the modelling step and by far a greater computation effort. Since the clustering analysis usually provides a number of sequences of the same protein corresponding to different strains, it is important to consider the possibility to express and use multistrain polypeptide antigens obtained by the fusion of multiepitope regions of the protein derived from the selected representative sequences, in order to have a reasonable low number of individual polypeptides in the test. In this respect, a preliminary consensus prediction based on linear and conformational (discontinuous) epitopes is essential. In order to perform a proper conformational epitope prediction, the knowledge of the protein 3D structure is necessary; when no experimentally-determined structure is available, the alternative is protein homology modelling. As shown in this study, the structural modelling, MD and analysis of the antigens can facilitate the choice of the boundaries of each antigen to be inserted into the fusion protein construct by selecting stable domains with structural integrity, considering oligomeric states of the protein, avoiding aggregation-prone fragments and retaining the majority of the epitopes of the parent proteins. Furthermore, the bioinformatic methods used can help in assessing structural stability and solubility also of the planned multistrain polypeptides, which are important not only for the final antigen product but also for successful production of the recombinant protein throughout the expression, folding and purification steps. Although "artificial" fusion proteins are created through linkers by this approach, probably each antigen incorporated in the multistrain polypeptide very closely mimics its native state, against which most likely the antibodies in the infected animals have been raised. The consequence of this assumption is that it might also contribute to higher affinity in the immunoassay with respect to other tests based on antigens of randomly-chosen fragments and small peptides.

It is notable that the TM exodomain contains an immunodominant epitope in the loop region. This epitope is common to all lentiviruses (Saman et al., 1999) and comprises a sequence of few amino acids between two conserved cysteine residues that are engaged in a disulfide bond. In fact, the propensity of this region to be a good epitope much depends on the conformational restriction imposed by the disulfide bond as shown by Saman et al., who used a cyclized oxidized peptide bound to the streptavidin in their indirect ELISA test as TM derived antigen. Our goal was to have as many epitopes as possible in the TM antigen and a folded conformation for the immunodominant epitope described above. The results show that very likely the TM_1_2 polypeptide has a folded structure in the loop regions due to the favourable intramolecular interactions between the four α-helices (two helix-loop-helix motifs) predicted in the model, thus retaining the immunogenic properties in the region between the two cysteines without any need of chemical oxidation of the polypeptide.

In this study we have applied the rational bioinformatic approach described above to define a set of antigens useful for SRLV serodiagnostics. We expressed a multistrain polypeptide related to the P25 (P25_2_3frag) with two partial sequences fused, derived from two different strains. A second multistrain polypeptide related to the transmembrane protein (gp46) (TM_1_2) was also expressed. In this case the sequences of the exodomain region of two TM proteins of two different

strains were used. The P25_2_3frag was obtained in scarce quantity due to difficulties with purification and not used in further applications. On the contrary, TM_1_2 was useful to develop an ELISA test together with the P25_1 protein and the P25_2frag polypeptide. This ELISA test gave promising results when compared with two commercial ELISA kits and PCR for SRLV detection.

Due to the quasispecies nature of the SRLV it has been reported in the more recent literature that ELISA tests based on proteins/(poly)peptides derived from a single strain are less effective in the diagnosis (Grego et al., 2002) even if previous reports on single strain-based tests claimed high sensitivity and specificity (Saman et al., 1999). It is now widely accepted that multistrain- multiprotein tests result in better diagnostic performances, and thus proteins/ (poly) peptides derived from different genotypes have been used. However, it is not known how the set of antigens used so far in each specific test is representative with respect to the whole mutant spectra of the SRLV quasispecies. The combination of three different ELISAs and a PCR-based molecular test in a multi-platform approach has been recently used in order to increase the diagnostic efficiency (Ramírez et al., 2021; Echeverría et al., 2020). Using a threshold higher than 90 % would result in an increased number of representative sequences that, when mixed, might compete for absorption leading to a non-optimal amount of each protein attached to the wells. Therefore, it would be better to split them between different single tests. We believe that having too many proteins in a single test would be problematic even using chimeras. Therefore, as future development the design of multiple sets of antigens and their application in a number of different indirect ELISA to be performed in parallel, could be advantageously used in a multi-platform approach to better cover the antigenic heterogeneity of the SRLV.

We believe that our approach, based on multistrain polypeptides of representative sequences derived using clustering procedures combined with biostructural studies and gene synthesis, is at least a systematic procedure to select a set of better representative antigens of SRLV viral quasispecies.

Over the past two decades Immunoinformatics (a science that helps to create significant immunological information using bioinformatics softwares and applications) (Tomar and De, 2014; De Groot et al., 2020) has been applied to design suitable antigens for human and veterinary vaccine development. However, application of bioinformatic and biostructural tools to design antigens for veterinary immunoassays are quite a few. In our opinion the veterinary diagnostic sector could benefit from these in silico analyses.

## Author contributions

**Angela Ostuni:** Investigation, Visualization, Writing - review & editing; **Magnus Monné:** Investigation, Writing - original draft,Visualization; **Maria Antonietta Crudele:** Data curation, Formal analysis; **Pierluigi Cristinziano:** Investigation; **Stefano Cecchini:** Investigation, Resources; **Mario Amati:** Investigation; **Jolanda De Vendel:** Investigation; **Paolo Raimondi:** Resources; **Taxiarchis Chassalevris:** Investigation; **Chrysostomos I Dovas:** Investigation; **Alfonso Bavoso:** Conceptualization, Project administration, Methodology, Writing - original draft, Writing - review & editing, Funding Acquisition.

## Declaration of Competing Interest

The authors report no declarations of interest.

## Acknowledgments

This Investigation was partially funded by Basilicata Innovazione project. Thanks are due to Dr. Carmine Del Galdo for providing serum samples from Campania region.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jviromet.2021.114266.

## References

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 44, W344–350.

Bertoni, G., Zahno, M.L., Zanoni, R., Vogt, H.R., Peterhans, E., Ruff, G., Cheevers, W.P., Sonigo, P., Pancino, G., 1994. Antibody reactivity to the immunodominant epitopes of the caprine arthritis-encephalitis virus gp38 transmembrane protein associates with the development of arthritis. J. Virol. 68 (11), 7139–7147. https://doi.org/10.1128/JVI.68.11.7139-7147.1994.

Bertoni, G., Hertig, C., Zahno, M.L., Vogt, H.R., Dufour, S., Cordano, P., Peterhans, E., Cheevers, W.P., Sonigo, P., Pancino, G., 2000. B-cell epitopes of the envelope glycoprotein of caprine arthritis-encephalitis virus and antibody response in infected goats. J. Gen. Virol. 81 (Pt 12), 2929–2940. https://doi.org/10.1099/0022-1317-81-12-2929.

Braun, M.J., Clements, J.E., Gonda, M.A., 1987. The visna virus genome: evidence for a hypervariable site in the env gene and sequence homology among lentivirus envelope proteins. Virology 61 (12), 4046–4054.

Brinkhof, J., van Maanen, C., 2007. Evaluation of five enzyme-linked immunosorbent assays and an agar gel immunodiffusion test for detection of antibodies to small ruminant lentiviruses. Clin. Vac. Immunol. CVI 14 (9), 1210–1214. https://doi.org/10.1128/CVI.00282-06.

Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Götz, A.W., Kolossváry, I., Wong, K.F., Paesani, F., Vanicek, J., Wolf, R.M., Liu, J., Wu, X., Brozell, S.R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D.R., Mathews, D.H., Seetin, M.G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman, P.A., 2012. AMBER 12. University of California, San Francisco.

Chassalevris, T., Chaintoutis, S.C., Apostolidi, E.D., Giadinis, N.D., Vlemmas, I., Brellou, G.D., Dovas, C.I., 2020. A highly sensitive semi-nested real-time PCR utilizing oligospermine-conjugated degenerate primers for the detection of diverse strains of small ruminant lentiviruses. Mol. Cell. Probes 51, 101528. https://doi.org/10.1016/j.mcp.2020.101528.

Colitti, B., Coradduzza, E., Puggioni, G., Capucchio, M.T., Reina, R., Bertolotti, L., Rosati, S., 2019. A new approach for Small Ruminant Lentivirus full genome characterization revealed the circulation of divergent strains. PLoS One 14 (2), e0212585. https://doi.org/10.1371/journal.pone.0212585.

de Andrés, D., Klein, D., Watt, N.J., Berriatua, E., Torsteinsdottir, S., Blacklaws, B.A., Harkiss, G.D., 2005. Diagnostic tests for small ruminant lentiviruses. Vet. Microbiol. 107, 49–62.

De Groot, A.S., Moise, L., Terry, F., Gutierrez, A.H., Hindocha, P., Richard, G., Hoft, D.F., Ross, T.M., Noe, A.R., Takahashi, Y., Kotraiah, V., Silk, S.E., Nielsen, C.M., Minassian, A.M., Ashfield, R., Ardito, M., Draper, S.J., Martin, W.D., 2020. Better epitope discovery, precision immune engineering, and accelerated vaccine design using immunoinformatics tools. Front. Immunol. 11, 442. https://doi.org/10.3389/fimmu.2020.00442.

Drew, Elliot D., Janes, Robert W., 2020. PDBMD2CD: providing predicted protein circular dichroism spectra from multiple molecular dynamics-generated protein structures. Nucleic Acids Res. 48 (W1) https://doi.org/10.1093/nar/gkaa296, 17–W24.

Echeverría, I., De Miguel, R., De Pablo-Maiso, L., Glaria, I., Benito, A.A., De Blas, I., De Andrés, D., Luján, L., Reina, R., 2020. Multi-platform detection of small ruminant lentivirus antibodies and provirus as biomarkers of production losses. Front. Vet. Sci. 7, 182. https://doi.org/10.3389/fvets.2020.00182.

Fiser, A., Sali, A., 2003. Modeller: generation and refinement of homology-based protein structure models. Meth. Enzymol. 374, 461–491. https://doi.org/10.1016/S0076-6879(03)74020-8.

Fraisier, C., Arnarson, H., Barbezange, C., Andresdottir, V., Carrozza, M.L., De Andres, D., Tolari, F., Rosati, S., Lujan, L., Pepin, M., Amorena, B., Harkiss, G., Blacklaws, B., Suzan-Monti, M., 2007. Expression of the gp150 maedi visna virus envelope precursor protein by mammalian expression vectors. Virol. Methods 146 (1-2), 363–367.

Fras, M., Leboeuf, A., Labrie, F.M., Laurin, M.A., Singh Sohal, J., L'Homme, Y., 2013. Phylogenetic analysis of small ruminant lentiviruses in mixed flocks: multiple evidence of dual infection and natural transmission of types A2 and B1 between sheep and goats. Infect. Genet. Evol. 19, 97–104. https://doi.org/10.1016/j.meegid.2013.06.019.

Gjerset, B., Storset, A.K., Rimstad, E., 2006. Genetic diversity of small-ruminant lentiviruses: characterization of Norwegian isolates of Caprine arthritis encephalitis virus. J. Gen. Virol. 87 (Pt 3), 573–580. https://doi.org/10.1099/vir.0.81201-0.

Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224.

Grego, E., Profiti, M., Giammarioli, M., Giannino, L., Rutili, D., Woodall, C., Rosati, S., 2002. Genetic heterogeneity of small ruminant lentiviruses involves immunodominant epitope of capsid antigen and affects sensitivity of single-strain-based immunoassay. Clin. Diagn. Lab. Immunol. 9 (4), 828–832. https://doi.org/10.1128/cdli.9.4.828-832.2002.

Grego, E., Bertolotti, L., Quasso, A., Profiti, M., Lacerenza, D., Muz, D., Rosati, S., 2007. Genetic characterization of small ruminant lentivirus in Italian mixed flocks: evidence for a novel genotype circulating in a local goat population. J. Gen. Virol. 88 (Pt 12), 3423–3427. https://doi.org/10.1099/vir.0.83292-0.

Hofmann, K., Stoffel, W., 1993. TMbase - A database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler 374, 166.

Holland, J.J., de La Torre, J.C., Steinhauer, D.A., 1992. RNA virus populations as quasispecies. Curr. Top. Microbiol. Immunol. 176, 1–20.

Huang, Y., Niu, B., Gao, Y., Fu, L., Li, W., 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics (Oxford, England) 26 (5), 680–682. https://doi.org/10.1093/bioinformatics/btq003.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD - visual molecular dynamics. J. Mol. Graph. Model. 14, 33–38.

Jin, Z., Jin, L., Peterson, D.L., Lawson, C.L., 1999. Model for lentivirus capsid core assembly based on crystal dimers of EIAV p26. J. Mol. Biol. 286 (1), 83–93. https://doi.org/10.1006/jmbi.1998.2443.

Knowles Jr., D.P., Cheevers, W.P., McGuire, T.C., Brassfield, A.L., Harwood, W.G., Stem, T.A.E., 1991. Structure and genetic variability of envelope glycoproteins of two antigenic variants of caprine arthritis-encephalitis lentivirus. Virology 65 (11), 5744–5750.

Konishi, M., Yamamoto, T., Shimada, T., Shirafuji, H., Kameyama, K., Sentsui, H., Murakami, K., 2010. Development of enzyme-linked immunosorbent assay for detection of antibody against Caprine arthritis encephalitis virus using recombinant protein of the precursor of the major core protein, p55gag. J. Vet. Diagn. Invest. 22 (3), 415–419. https://doi.org/10.1177/104063871002200312.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., Karplus, K., 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8. Proteins 77 (Suppl 9), 114–122.

Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M., 2012. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol. 8, e1002829.

Larsen, J.E.P., Lund, O., Nielsen, M., 2006. Improved method for predicting linear B-cell epitopes. Immunome Res. 2, 2.

Lee, H., Lara, P., Ostuni, A., Presto, J., Johansson, J., Nilsson, I., Kim, H., 2014. Live-cell topology assessment of URG7, MRP6$_{102}$ and SP-C using glycosylatable green fluorescent protein in mammalian cells. Biochem. Biophys. Res. Commun. 450 (4), 1587–1592. https://doi.org/10.1016/j.bbrc.2014.07.046.

Leroux, C., Cruz, J.C., Mornex, J.F., 2010. SRLVs: a genetic continuum of lentiviral species in sheep and goats with cumulative evidence of cross species transmission. Curr. HIV Res. 8 (1), 94–100. https://doi.org/10.2174/157016210790416415.

Liang, S., Zheng, D., Zhang, C., Zacharias, M., 2009. Prediction of antigenic epitopes on protein surfaces by consensus scoring. BMC Bioinformatics 10, 302.

Malashkevich, V.N., Singh, M., Kim, P.S., 2001. The trimer-of-hairpins motif in membrane fusion: visna virus. Proc. Natl. Acad. Sci. U. S. A. 98 (15), 8502–8506. https://doi.org/10.1073/pnas.151254798.

Michiels, R., Adjadj, N.R., De Regge, N., 2020. Phylogenetic analysis of belgian small ruminant lentiviruses supports cross species virus transmission and identifies new subtype B5 strains. Pathogens 9 (3), 183. https://doi.org/10.3390/pathogens9030183.

Micsonaia, A., Wienb, F., Kernyaa, L., Leec, Y.-H., Gotoc, Y., Réfrégiersb, M., Kardos, J., 2015. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. PNAS Online. https://doi.org/10.1073/pnas.1500851112.

Minardi da Cruz, J.C., Singh, D.K., Lamara, A., Chebloune, Y., 2013. Small ruminant lentiviruses (SRLVs) break the species barrier to acquire new host range. Viruses 5 (7), 1867–1884. https://doi.org/10.3390/v5071867.

Minguijón, E., Reina, R., Pérez, M., Polledo, L., Villoria, M., Ramírez, H., Leginagoikoa, I., Badiola, J.J., García-Marín, J.F., de Andrés, D., Luján, L., Amorena, B., Juste, R.A., 2015. Small ruminant lentivirus infections and diseases. Vet. Microbiol. 181 (1–2), 75–89. https://doi.org/10.1016/j.vetmic.2015.08.007.

Molaee, V., Bazzucchi, M., De Mia, G.M., Otarod, V., Abdollahi, D., Rosati, S., Lühken, G., 2020. Phylogenetic analysis of small ruminant lentiviruses in Germany and Iran suggests their expansion with domestic sheep. Sci. Rep. 10 (1), 2243. https://doi.org/10.1038/s41598-020-58990-9.

Momany, C., Kovari, L.C.A., Prongay, Keller, W., Gitti, R.K., Lee, B.M., Gorbalenya, A.E., Tong, L., McClure, J., Ehrlich, L.S., Summers, M.F., Carter, C., Rossmann, M.G., 1996. Crystal structure of dimeric HIV-1 capsid protein. Nat. Struct. Biol. 3 (9), 763–770. https://doi.org/10.1038/nsb0996-763.

OIE, https://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.07.02-03_CAE_MV.pdf/.

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P.E., Sette, A., Peters, B., 2008. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC Bioinformatics 9, 514.

Pornillos, O., Ganser-Pornillos, B.K., Kelly, B.N., Hua, Y., Whitby, F.G., Stout, C.D., Sundquist, W.I., Hill, C.P., Yeager, M., 2009. X-ray structures of the hexameric building block of the HIV capsid. Cell 137 (7), 1282–1292. https://doi.org/10.1016/j.cell.2009.04.063.

Ramírez, H., Reina, R., Amorena, B., de Andrés, D., Martínez, H.A., 2013. Small ruminant lentiviruses: genetic variability, tropism and diagnosis. Viruses 5 (4), 1175–1207.

Ramírez, H., Echeverría, I., Benito, A.A., Glaria, I., Benavides, J., Pérez, V., de Andrés, D., Reina, R., 2021. Accurate diagnosis of small ruminant lentivirus infection is needed for selection of resistant sheep through TMEM154 E35K genotyping. Pathogens 10 (1), 83. https://doi.org/10.3390/pathogens10010083.

Reina, R., Bertolotti, L., Dei Giudici, S., Puggioni, G., Ponti, N., Profiti, M., Patta, C., Rosati, S., 2010. Small ruminant lentivirus genotype E is widespread in Sarda goat. Vet. Microbiol. 144 (1–2), 24–31. https://doi.org/10.1016/j.vetmic.2009.12.020.

Rosati, S., Mannelli, A., Merlo, T., Ponti, N., 1999. Characterization of the immunodominant cross-reacting epitope of visna maedi virus and caprine arthritis-encephalitis virus capsid antigen. Virus Res. 61 (2), 177–183. https://doi.org/10.1016/S0168-1702(99)00031-3. ISSN 0168-1702.

Saha, S., Raghava, G.P.S., 2004. BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: Nicosia, G., Cutello, V., Bentley, P.J., Timmis, J. (Eds.), Artificial Immune Systems. Springer, Berlin, Heidelberg, pp. 197–204.

Saha, S., Raghava, G.P.S., 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65, 40–48.

Saltarelli, M., Querat, G., Konings, D.A., Vigne, R., Clements, J.E., 1990. Nucleotide sequence and transcriptional analysis of molecular clones of CAEV which generate infectious virus. Virology 179 (1), 347–364.

Saman, E., Van Eynde, G., Lujan, L., Extramiana, B., Harkiss, G., Tolari, F., Gonzàlez, L., Amorena, B., Watt, N., Badiola, J., 1999. A new sensitive serological assay for detection of lentivirus infections in small ruminants. Clin. Diagn. Lab. Immunol. 6 (5), 734–740.

Schrödinger, L.L.C., 2010. The PyMOL Molecular Graphics System, 1.6 ed. http://www.pymol.org/.

Shah, C., Böni, J., Huder, J.B., Vogt, H.-R., Mühlherr, J., Zanoni, R., et al., 2004a. Phylogenetic analysis and reclassification of caprine and ovine lentiviruses based on 104 new isolates: evidence for regular sheep-to-goat transmission and worldwide propagation through livestock trade. Virology 319 (1), 12–26. https://doi.org/10.1016/j.virol.2003.09.047. PMID: 14967484.

Shah, C., Huder, J.B., Böni, J., Schönmann, M., Mühlherr, J., Lutz, H., Schüpbach, J., 2004b. Direct evidence for natural transmission of small-ruminant lentiviruses of subtype A4 from goats to sheep and vice versa. J. Virol. 78 (14), 7518–7522. https://doi.org/10.1128/JVI.78.14.7518-7522.2004.

Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A., Wain-Hobson, S., 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell 42 (1), 369–382.

Sweredoski, M.J., Baldi, P., 2008. PEPITO: improved discontinuous B-cell epitope prediction using multiple distance thresholds and half sphere exposure. Bioinformatics 24, 1459–1460.

Sweredoski, M.J., Baldi, P., 2009. COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng. Des. Sel. 22, 113–120. https://doi.org/10.1007/978-1-4939-1115-8_3.

Tomar, N., De, R.K., 2014. Immunoinformatics: a brief review. In: De, R., Tomar, N. (Eds.), Immunoinformatics. Methods in Molecular Biology (Methods and Protocols), vol 1184. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-1115-8_3.

Whitmore, L., Wallace, B.A., 2004. DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 32, 668–673.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315.

Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., Ventura, S., 2015. AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. Nucleic Acids Res. 43, W306–W313. https://doi.org/10.1093/nar/gkv359.

Zhou, C., Chen, Z., Zhang, L., Yan, D., Mao, T., Tang, K., Qiu, T., Cao, Z., 2019. SEPPA 3.0-enhanced spatial epitope prediction enabling glycoprotein antigens. Nucleic Acids Res. 47, W388–W394.