

Epitope Mapping of Tobacco Mosaic Virus Capsid Protein: Prediction and Experimental Data from Spot Synthesis

Daniel Ferreira de Lima Neto¹, Clarice Weis Arns¹, Dagmar Ruth Stach-Machado², Fernando Rosado Spilki³, Juliana Mattoso⁴ and Carlos Francisco Sampaio Bonafe^{**}

¹Departamento de Genética, Laboratório de Virologia Animal, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, Brazil

²Departamento de Biologia Estrutural e Funcional, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, Brazil

³Laboratório de Microbiologia Molecular, Instituto de Ciências da Saúde, Universidade Feevale, Novo Hamburgo, Brazil

⁴Departamento de Bioquímica, Laboratório de Termodinâmica de Proteínas, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, Brazil

Abstract

The immune system is a network of thousands of molecules, cells and regulatory factors that produce many interrelated responses. In this study, we used spot synthesis to map tobacco mosaic virus (TMV) epitopes in a mice animal model (Balb/c) and compared the results with those obtained using immunoinformatic prediction tools. Mice were inoculated with TMV and after immunization the sera were incubated with an array of overlapping pentadecapeptides that corresponded to the full sequence of the TMV capsid protein (TMVcp) that had been synthesized on a cellulose membrane for spot synthesis analysis. Six linear epitopes were identified experimentally, as shown by the IgG-elicited immune responses in mice. The data for epitope prediction based on epitope databases agreed with the results obtained by spot synthesis results. Comparison of the findings for spot intensities and those obtained with the prediction software allowed the identification of different responses according to the MHC class I alleles. The results of this work provide a detailed antigenic profile for TMV.

Keywords: Tobacco mosaic virus; Epitope mapping; Spot synthesis

Introduction

Tobacco mosaic virus (TMV) is an extensively studied virus model in numerous biological systems and, consequently, has become one of the best known viruses, alongside human immunodeficiency virus (HIV). The widespread use of TMV reflects the ease with which this virus can be manipulated, produced and purified. Studies of TMV have provided important insights into immunology and virology [1-3]. The TMV virion consists of a rod-shaped structure with 2130 identical protein subunits surrounding the genetic material (a single positive RNA strand). The capsid protein (TMVcp) subunit, which consists of 159 amino acids, is arranged as a right-handed antiparallel helical bundle composed of four α -helices, the left and right slanted (LS and RS) and left and right radial (LR and RR). A short inner loop connects the LS and RS helices while a longer one connects the LR and RR helices. The RNA binding site is formed by residues from the loops and the LR helix, and the N and C termini are located on the outer surface of the virion. The ability of TMVcp to aggregate is important for virus stability and assembly and is directly related to viral physiology [4,5]. The study of physical chemical properties of TMV is important to understand its physiology. We investigated the dissociation process by using of high pressure at low temperature ($\leq 0^\circ\text{C}$) [6], alkaline conditions [7] and presence of urea [8]. A combination of high pressure and such conditions revealed a potential tool for study of viral stability.

Several studies aimed at improving our understanding of structural factors and antigenicity have identified different antibody binding sites during the immune responses elicited by virus inoculation [4,9,10]. The immune system is a network of thousands of molecules, cells and regulatory factors that mediate many interrelated responses. Studies over the past few years have reported individual variation in the immune response as well as temporal changes within the host [11]. Advanced computational methods provide a valuable tool for analyzing complex experimental data. With such methods it is possible to identify specific patterns in the immune response based solely on protein sequences and to correlate these findings with experimental data; molecular predictions based on such analyses are widely used in

the literature [12,13]. In the present study, we used spot analysis to map the TMV epitopes and compared our findings with data from other studies in the literature. We also used immunoinformatics tools to map the potential B- and T-cell epitopes and compared these data with the results of epitope mapping.

Material and Methods

TMV purification and RNA sequencing

TMV was isolated from Turkish plants infected with the common strain of the virus. The purification method was described previously [14].

RNA from the reference virus strain was extracted with TRIzol[®] (Invitrogen), quantified by spectrophotometry and stored at -80°C until used. For reverse transcription, RNA (50 μg) was converted to cDNA using Superscript III reverse transcriptase, according to the manufacturer's instructions. The reactions were done using diethylpyrocarbonate (DEPC) water in sterile conditions. The polymerase chain reactions (PCR) were done by adding 10% of the RT reaction to a final volume of 50 μL . 10 % of the reverse transcriptase reaction was used as input for the PCR reaction. The PCR mixture consisted of 5 μL

***Corresponding author:** Dr. Carlos Francisco Sampaio Bonafe, Departamento de Bioquímica, Laboratório de Termodinâmica de Proteínas, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Rua Monteiro Lobato, 255, Campinas, 13083-862, SP, Brazil, Tel: +55 19 35216229; Fax: +55 19 35216129; E-mail: bonafe@unicamp.br

Received October 03, 2012; **Accepted** October 22, 2012; **Published** October 24, 2012

Citation: de Lima Neto DF, Arns CW, Stach-Machado DR, Spilki FR, Mattoso J, et al. (2013) Epitope Mapping of Tobacco Mosaic Virus Capsid Protein: Prediction and Experimental Data from Spot Synthesis. Mol Biol 2:108. doi:10.4172/2168-9547.1000108

Copyright: © 2013 de Lima Neto DF, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of each primer (5 pmol), 5 μ L of 10x buffer, 1 μ L of dNTPs (0.2 mM), 2 μ L of DTT, 0.5 μ L of *Taq* DNA polymerase and DEPC water to a final volume of 50 μ L. The PCR products were visualized in 2% agarose gels stained with ethidium bromide and then analyzed in a MEGABace 1000 sequencer following the manufacturer instructions. Flanking primers for the complete sequence of the TMVcp were used. The electropherogram of the sequences were aligned using the DNASTar software package (SeqMan) and the consensus sequence was exported to BLAST and to Protean (DNASTar). The amino acid sequence was deduced from the full nucleotide sequence using DNASTar software. Local alignment and basic protein structure properties were calculated using the CLC Main Workbench 5.5* (protein toolbox) and Lasergene DNASTar* software (EditSeq and SeqMan) packages.

Preparation of peptide arrays and epitope identification

Peptides based on the amino acid sequence determined above were synthesized by using Fmoc (9-fluorenylmethoxycarbonyl) chemistry [15,16] on PEG-derivatized cellulose membranes with an additional β Ala anchor for the C-terminal immobilized peptides. The membranes contained overlapping pentadecapeptides spanning the complete sequence of TMVcp (residues 1–159), with an offset of three amino acid residues.

Serum preparation and antibody-binding assay and quantification

Balb/c female mice (n=5) obtained from the Multidisciplinary Center for Biological Investigation (CEMIB/UNICAMP) were inoculated with native TMV and complete Freund adjuvant in the first inoculation and incomplete adjuvant in subsequent inoculations. The inoculations were done on the 1st, 15th, 28th and 35th day using 15, 10, 10 and 15 μ g of TMV, respectively. Blood was collected 14 days after the last inoculation, pooled and allowed to clot, after which the serum was obtained by centrifugation, inactivated by incubating for 1 h at 56 $^{\circ}$ C and stored at 4 $^{\circ}$ C. Control serum was obtained from five mice inoculated with porcine parvovirus using the same scheme as for TMV. The protocols involving animals were approved by the institutional Committee for Ethics in Animal Use (CEUA/UNICAMP, protocol no. 1717-2).

The antibody-binding assay used was described by Beutling and cols. [16]. A serum dilution of 1/100 was used in the membrane antibody-binding assays. An alkaline phosphatase-conjugated goat anti-mouse IgG (AP-IgG) (Jackson Immuno Research, CAT no. 115-055-071) was used as secondary antibody to detect bound antibodies. The color reaction was developed as described elsewhere [17], with an incubation time of up to 60 min. The membrane was subsequently scanned at 2400 d.p.i. with a table scanner and the software package Totallab Quant – Array Analysis was used to measure the intensity of each spot against the background intensity. The results were expressed as the mean for each spot. The membranes were regenerated as described by Beutling and cols. [16].

An epitope-based immunoinformatics study of TMVcp

The deduced protein sequence used was aligned with the protein sequence of TMV capsid protein retrieved from the National Centre for Biotechnology Institute (NCBI) database (BLAST reference accession no. ACY41215.1). Epitope flexibility was predicted with the Karplus and Schulz flexibility scale [18] and antigenicity was predicted with the Kolaskar and Tongaonkar antigenicity scale [19]. Hydrophobicity and hydrophilicity were analyzed with the Parker hydrophilicity scale. The Kyte-Doolittle hydrophobicity index [20] available from the Expsy Prot scale server was used to assess the distribution of polar and

apolar residues along the protein sequence. The Kyte-Doolittle [20] scale (window size of 7 and cutoff of 1.6) is widely used to examine the hydrophobic character of proteins and is useful in predicting membrane-spanning domains, potential antigenic sites and regions that are likely to be exposed on the protein surface. This approach was complemented with the Hopp method for a window size of 7-mers [20,21]. Ubiquitination sites were predicted as described elsewhere [22]. Proteasomal cleavage sites were predicted based on proteasomal degradation experiments with β -casein [23], enolase [24] and prion proteins [25]. The prediction of TAP affinity was based on a support vector machine (SVM) [22]. The SVMHC (support vector MHC) method was used to predict MHC binding [26]. This is an SVM-based method based on verified MHC binding peptides from the SYFPEITHI database [27]. These methods have been validated elsewhere [28]. Data available from the Expsy Prot scale server were used to counter-verify the *in silico* data [29].

Combining the prediction methods

The class II MHC algorithm and the linear and conformational B cell epitope prediction algorithms were used to simulate the processing pathways for which there are *in silico* prediction algorithms, particularly with regard to the virus and host used here.

The separate prediction methods were combined in order to model the entire processing pathway of MHC class I antigens. Predicted peptides should have a C terminus generated by the proteasome, a relatively high TAP affinity and some affinity for MHC molecules. The high accuracy of the final step of MHC binding prediction means that other methods can be used as filters to remove candidate peptides unlikely to be generated by proteasomal cleavage and/or transported by TAP. Peptides nine amino acids long (9-mer peptides) were extracted from the SYFPEITHI database [30]. We focused on 9-mer peptides because MHC class I binds these peptides more frequently than 8-, 10- or 11-mer peptides. The alleles Kd, Dd and Ld influence epitope display in Balb/C mice and were used to map the potential epitopes of TMV. The tentative epitopes were predicted with the PCM, SVMTAP and SVMHC methods [12,13,31-33]. In the first step of analysis, the protein sequences were screened individually for the best binding epitopes by using the Immune Epitope Database (IEDB) algorithm. The cutoff scores were adjusted by the algorithm itself.

Results

The PCR-amplified TMVcp sequence determined by sequencing is shown below. A BLASTn search revealed that the sequence shared 100% homology with a sequence retrieved from the National Center for Biotechnology Institute (NCBI) database (accession no. ACY41215.1). The corresponding amino acid sequence (determined with CLC Workbench software) is subsequently showed (Structure 1).

The TMV strain corresponded to the same as previously described (variant 1) [34], and deposited in the GenBank under the accession number V01408.1. Mice sera were titrated by ELISA, and the results for serial dilution testing are in Table 1.

Serum dilution	Optical density
0	0.000
2	1.906
1	1.592
0.5	1.119
0.25	1.034
0.125	0.476
0.0625	0.330
0.03125	0.148

Table 1: ELISA quantification of sera from mice immunized against TMV coat protein.

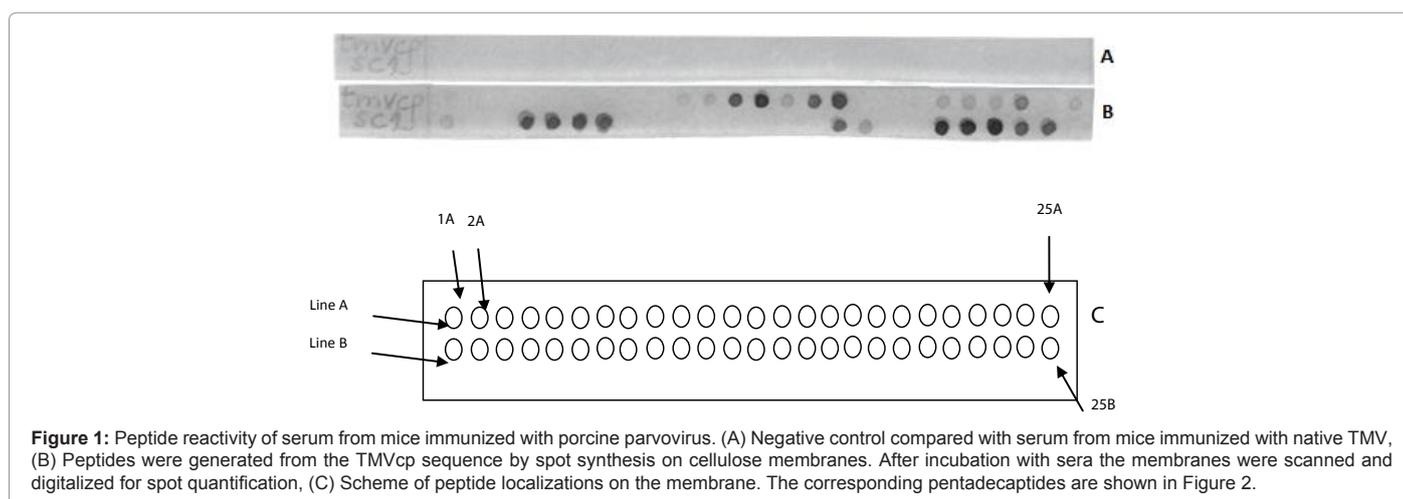
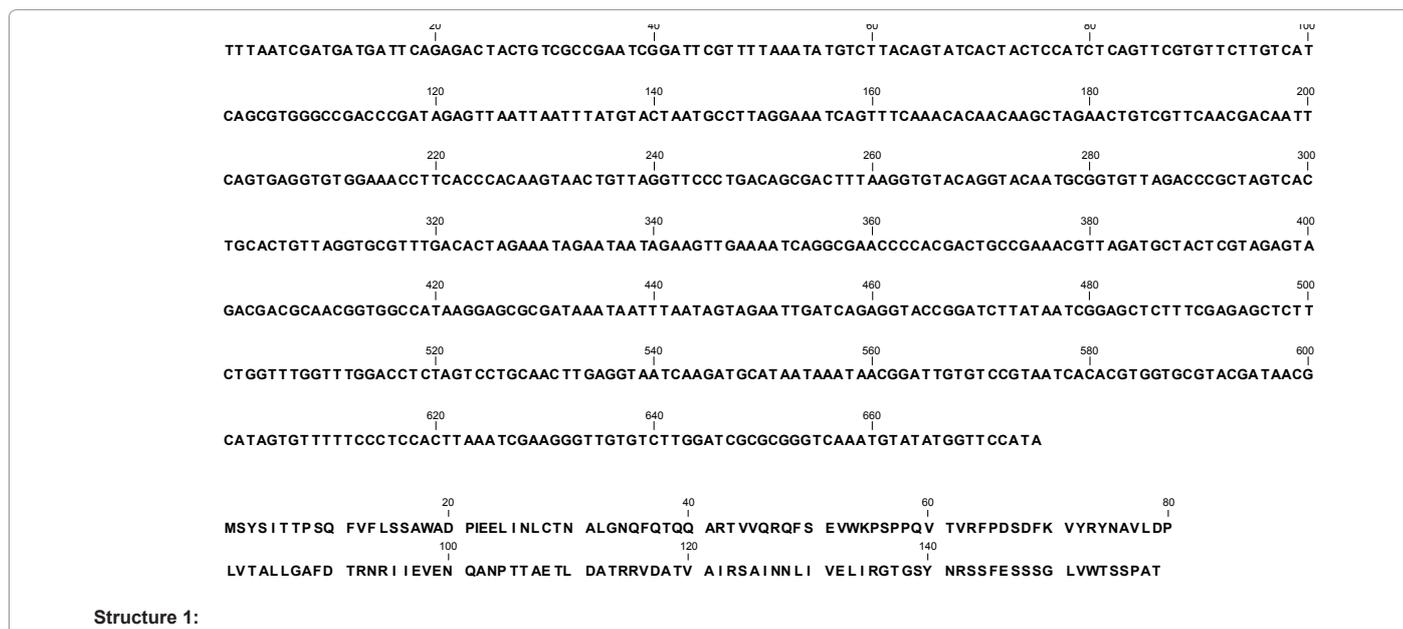


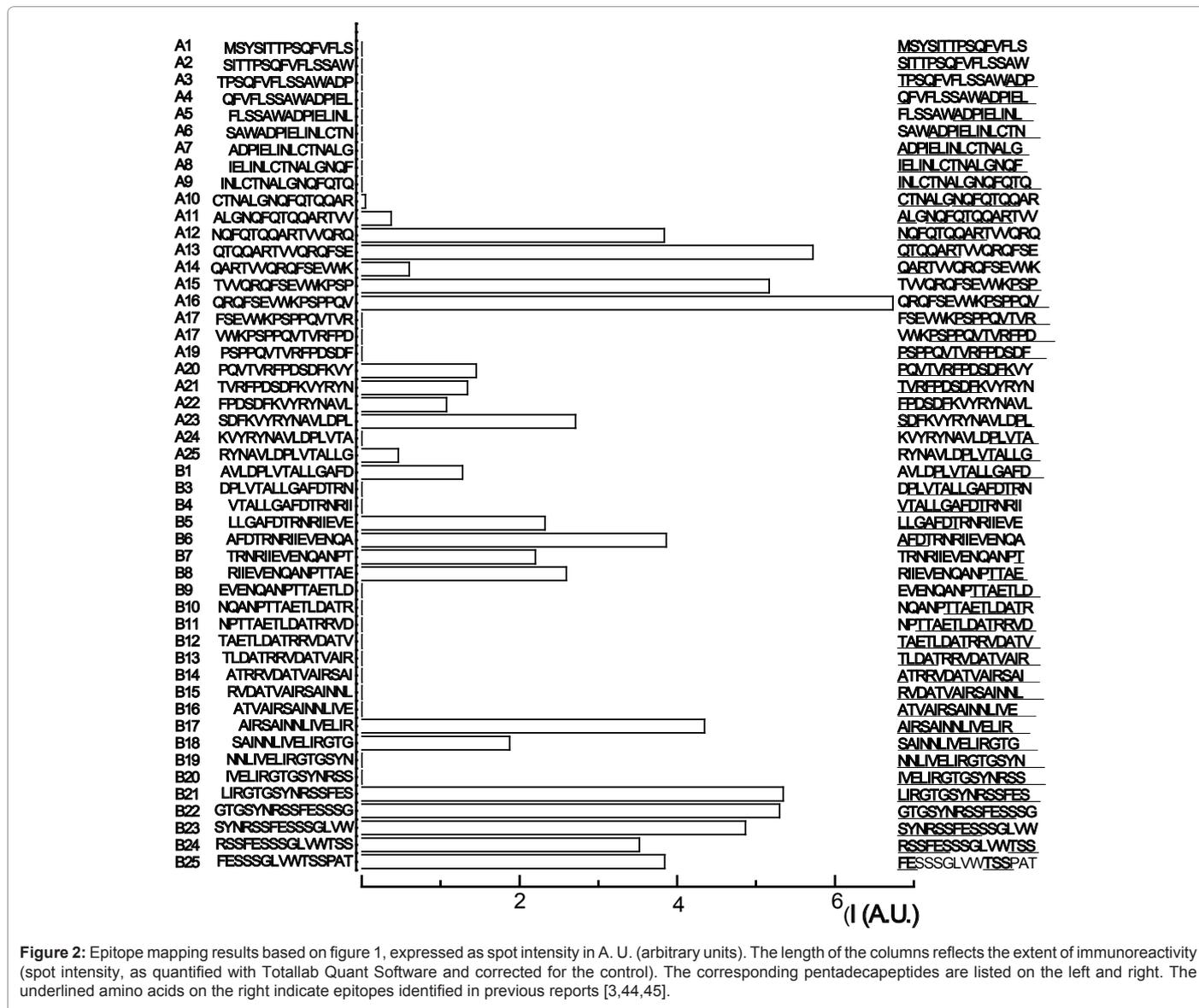
Figure 1: Peptide reactivity of serum from mice immunized with porcine parvovirus. (A) Negative control compared with serum from mice immunized with native TMV, (B) Peptides were generated from the TMVcp sequence by spot synthesis on cellulose membranes. After incubation with sera the membranes were scanned and digitalized for spot quantification, (C) Scheme of peptide localizations on the membrane. The corresponding pentadecapptides are shown in Figure 2.

To investigate the regions of TMVcp that were recognized by serum of immunized Balb/c mice, 49 overlapping decapentapeptides based on the deduced amino acid sequence of the protein were synthesized as an array on a cellulose membrane. Figure 1 shows the results for serum from mice inoculated with porcine parvovirus (A; negative control) and serum from TMV-immunized mice (B) when incubated with membrane-bound peptides. Significant antibody binding was observed with the latter serum compared to the control, which showed no detectable binding. Figure 2 shows the decapentapeptides sequences and the signal intensity of the positive samples (columns) after correction for the control. The underlined amino acids shown on the right side of Figure 2 correspond to previous reports of epitope mapping [3,9,35].

We found five strongly reactive regions: A10-A16 (residues 28-60), A20-A23 (residues 58-81), B4-B7 (residues 85-108), B16-B17 (residues 121-138) and B20-B24 (residues 133-159), and two weaker or less reactive regions: A4-A7 (residues 10-33) and A25-B1 (residues 73-90). The strong binding regions corresponded to α -helices that mainly projected away from the RNA binding site.

Initial protein analysis with CLC Workbench 5.5[®] and Lasergene DNASTar[®] showed that the TMVcp had a calculated molecular mass of 17,636 Da, an isoelectric point of 5.67 and an aliphatic index of 85.8 with 75 hydrophobic residues (A,F,G,I,L,M,P,V,W; frequency of 0.472) and 57 hydrophilic residues (C,N,Q,S,T,Y; frequency of 0.358). Figure 3 shows the antigenicity and hydrophobicity profiles predicted with the Hopp-Woods and Kyte-Doolittle algorithm, respectively.

Antigenicity and hydrophobicity predictions for the TMVcp identified a region between amino acid positions 28 and 60 that scored well for both parameters (Figure 3). Analysis of Figure 3 shows that the two plots were very symmetrical and appeared as a near mirror image of each other (in the Hopp-Woods plot apolar regions have negative values while in the Kyte-Doolittle [20] plot these regions have positive values). The Kolaskar and Tongaonkar "antigenicity plot" [19] also predicted antigenic regions at positions 16-39, 42-48, 68-77, 79-88 and 104-125 (not shown). Together, these findings identified four major peaks of antigenicity that corresponded to the regions where the epitopes were mapped. The peaks in these regions were polar in nature and were also accessible in the Kyte-Doolittle [20] plot. In contrast, the regions containing residues 30-70, 90-120 and 135-145 had negative



values, i.e., they were hydrophilic.

Prediction of solvent-accessible areas

The exposed fragments were studied by using surface accessible area software to assess which regions were available for further interactions such as antibody binding or capsid assembly. This software takes into account polarity, hydrophobicity and atomic radius to calculate which residues are located on the surface or buried within the protein. The results, depicted as scores given by the algorithm, are shown in Figure 4. The regions identified in this analysis are normally exposed on the surface of TMV and mediate the multiple interactions involved in capsid assembly in infected cells; they are also considered to be protective epitopes.

Based on the biochemical properties of each amino acid residue it is possible to predict the accessible area of a given protein and to determine which residues are exposed or buried within the structure of the protein. Other parameters taken into account by the algorithm include the solvent accessible surface area and atomic solvation.

Residues with values >60 (threshold) are considered to be located externally (on the outer surface) while those with values <60 are located internally (within the molecule). The residues identified on the surface of TMV were 3-15, 21-37, 57-60, 64-67, 99-105, 108-112, 126-135 and 142-154.

The susceptibility of TMVcp to proteasomal cleavage was assessed using the IEDB algorithm defined by Nielsen et al. [36] and the Prediction Algorithm for Proteasomal Cleavage, which uses two different prediction algorithms for proteasomal and immunoproteasomal cleavages [37]. Several regions yielded high scores with these methods, indicating presence of sites that were potentially available for cleavage. Figure 5 [38] shows the proteasomal cleavage predictions for the IEDB and PProC algorithms (NN and SVM, respectively).

The TAP binding affinity was assessed with the IEDB [39] and TAP Hunter [40] algorithms. As shown in Figure 6 [41], several regions were found to be suitable for binding. The regions that had high values but did not correlate with the predicted proteasomal cleavage sites and with our experimental results were not considered. The IEDB method

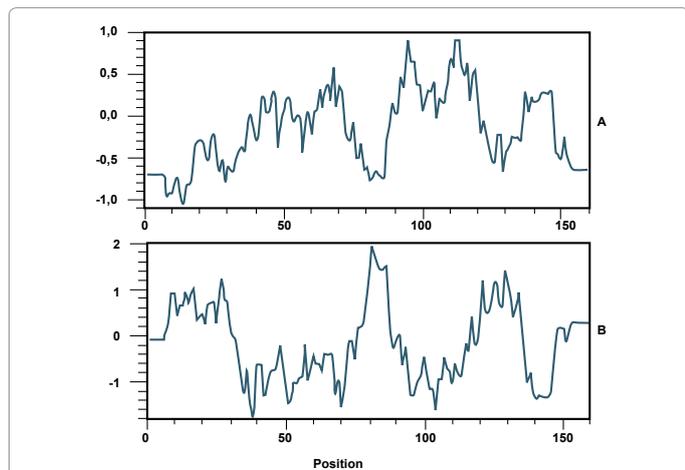


Figure 3: (A) Antigenicity plot generated with the Hopp-Woods algorithm. This plot identifies polar and apolar regions within the TMVcp sequence (apolar regions have negative values), (B) Hydrophobicity plot generated with the Kyte-Doolittle algorithm. Regions with positive values are hydrophobic.

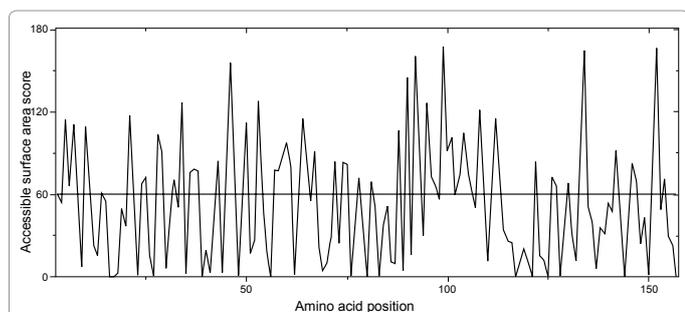


Figure 4: Solvent accessible surface area scores for TMVcp. Scores above the threshold of 60 indicates residues exposed on the surface of the protein.

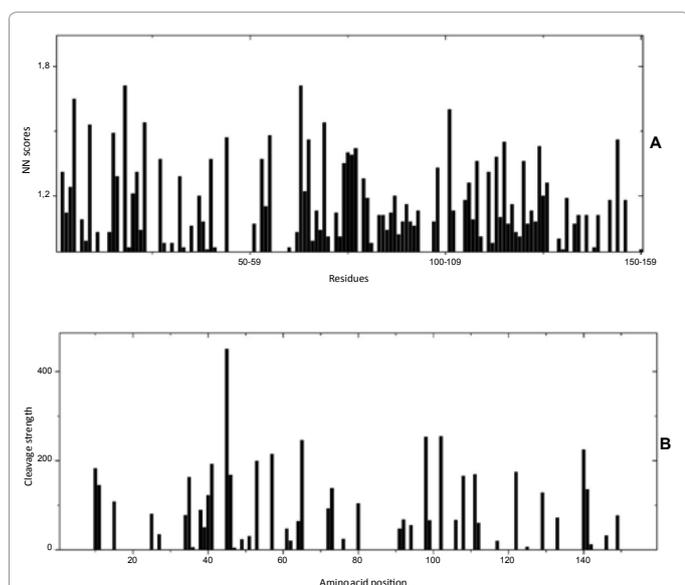


Figure 5: Predicted proteasomal cleavage of TMVcp based on the IEDB (A) and PProC (B) algorithms. In (A), the X axis represents the TMVcp sequence divided into decapeptides (segments of 10 peptides each) [35]. In (B), the X axis shows the amino acid position in TMVcp. In this case, greater cleavage strength indicates a higher probability of cleavage by the proteasome complex [34]. IEDB = Proteasome Cleavage Prediction by Neural Network. PProC = Proteasome Cleavage Prediction (human wild type I).

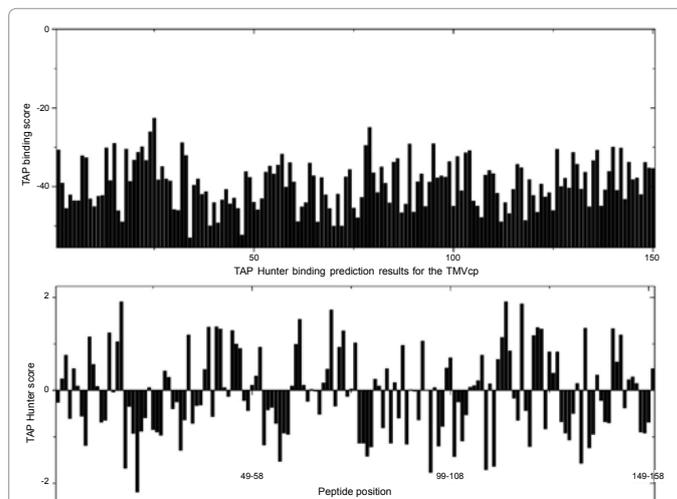


Figure 6: TAP-binding regions of TMVcp predicted using (A) Support Vector Machine TAP (IEDB algorithm) and (B) TAP Hunter. In (A), scores >40 indicate possible binding sequences [35]. In (B), positive values indicate possible TAP binding. The TAP Hunter algorithm identified several regions with low probability scores [38].

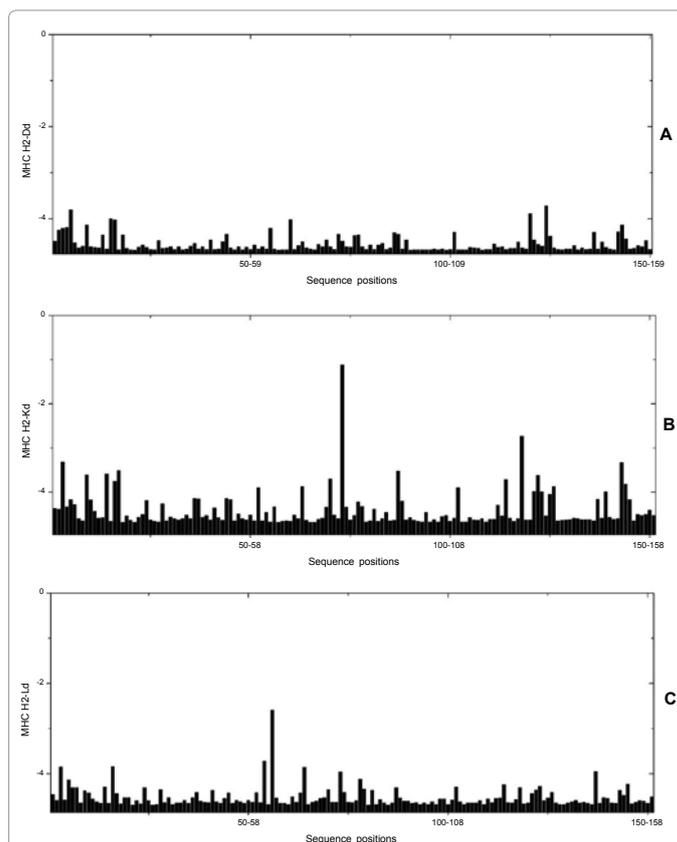


Figure 7: MHC class I predictions for TMVcp. A, B and C – MHC predictions for Balb/c mice alleles H2-Dd, H2-Kd and H2-Ld, respectively. Higher scores indicate sequence-specific affinity for the allele coded MHC.

revealed a similar landscape to that of quantification in the membrane and identified two peaks that correlated with our experimental findings. Amino acid residues 20-35, 47-58, 77-84 and 94-104 had the highest scores and were possibly associated with the TAP transporter; these regions were also identified in epitope mapping. The TAP Hunter

prediction algorithm also identified four broad regions of the sequence that were comparable to our experimental findings overlapped with the results from the IEDB prediction. The regions identified by these methods must also have high binding affinities for MHC class I receptors in order to be efficiently exposed outside normal cells.

Figure 7 shows the results for the analysis of alleles Dd, Kd and Ld using the IEDB database for MHC class I of Balb/c mice. These results reflect the peptide affinity for the MHC complex based on the database predictions. The affinity of a given peptide for the complex was proportional to the corresponding score. All three predictions identified a region spanning residues 3 to 16 as having high affinity. The affinity of the region that included residues 60 to 90 varied among the three predictions but was higher than for the fragment before it. The scores for alleles Dd and Kd in this region were similar whereas that for allele Ld was lower, with all three alleles showing different binding affinities in this region. The highest score for the H2Dd prediction (82-89) was -0.4 compared to -0.2 for the Kd region of residues 71-77 and -0.03 for the Ld prediction the region of residues 79-86. Of the three alleles only Ld was able to bind peptides in the region spanning residues 103-110. The terminal regions of TMVcp had similar binding regions although there was a progressive decrease in the binding affinities of alleles Dd, Kd and Ld. The greatest differences in the TMVcp peptide affinity scores were seen with alleles H2-Kd and H2-Ld. H2-Dd had low scores when compared to the other alleles and did not match the higher binding scores when the results from the three methods were combined with the overall scores (Figure 8) [42].

The MHC class II epitope predictions provided by the IEDB and CBS/ExPASy servers were used to analyze the possible responses of macrophages and dendritic cells to a challenge with TMV. Figure 9 [36] shows the results for the prediction of the two alleles IAd and IEd in Balb/c mice. The IAd allele recognized four long regions as possible binders to the MHC class II encoded by it; these regions were also detected in our experiments with TMV-immunized Balb/c mice, indicating that they were important in macrophage and T and B lymphocyte interactions and modulation. The MHC class II binding canyon allows the interaction of long peptides, even full proteins,

provided that the 9-15 amino acid size limit is respected [43,44]. Interaction between antigen processing cells (APC) and T lymphocytes leads to rapid cell proliferation and cell signaling as part of the immune response. The subsequent T CD4 and B cell communication that modulates antibody production and memory follows strict rules [45,46], and the end result is the binding of the endosome-lysosome-processed protein to the MHC class II complex, via CLIP. As shown here, the theoretical predictions and the experimental data yielded coherent results for the binding of IgG to epitopes on the membrane and high scores for the corresponding regions in IAd allele prediction. These findings suggest that there is a difference in the binding affinities of the IAd and IEd alleles in two regions (residues 30-70 and 105-115) that needs further investigation.

The BEpiPred online software identified six regions with positive values that corresponded to positive predictions (Figure 10). Some of these regions were also detected by the spot technique, e.g., regions formed by residues 50-75, 100-115 and 140-155. This method found linear epitopes based on the TMVcp protein sequence. Based on the scores provided by this method, the first sequence of eight residues (3-10 SITTPSQF) corresponded to the first and second spots of the first line; the second linear epitope had 24 residues (89-112; TRNRRIEVENQANPTTAETLDATR) that corresponded to spots 2-9 of the second line, although experimentally only spots 5-9 were detected (Figure 1B). The third linear epitope had 12 residues (146-157; SSGLVWTSSPA) and corresponded to spots 22-25 spots, of which only spots 22-24 were detected experimentally.

The TMVcp sequence was subsequently mapped for discontinuous epitopes using the EpiPred and IEDB algorithms, with the crystal structure of TMV being used as the input. Figure 11 shows front (left) and rear (right) views colored according to the predictions (BepiPred) (panel A) or shown in yellow on the structure predicted by IEDB (panel B). Both methods yielded similar results. Eight residues (7, 37, 39, 43, 47, 50, 134, 147) had single interaction points and three regions (63-66, 90-106, 140-143) were also predicted to be discontinuous.

Discussion

There is extensive structural information for TMV, particularly

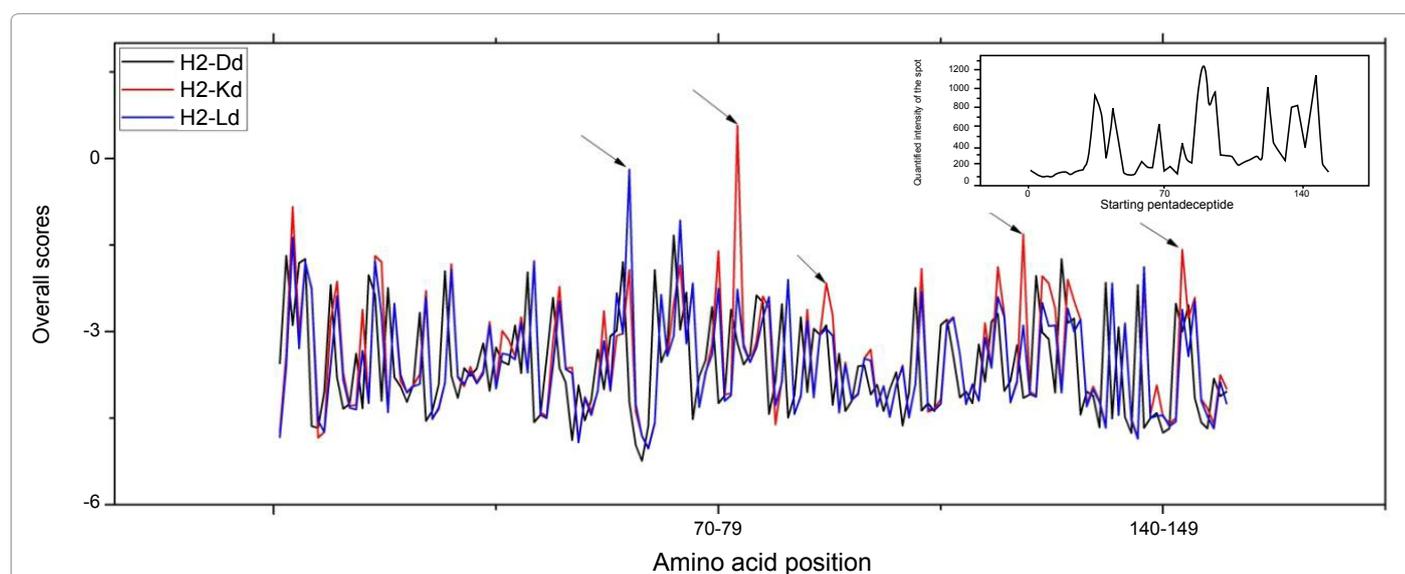


Figure 8: Combined overall results for the three Balb/c MHC class I alleles. Arrows indicate epitopes that were identified experimentally and predicted by bioinformatics tools [57]. Inset: Spot intensities quantified by densitometric analysis of the spot synthesis membrane.

the capsid protein, TMVcp, and the purified protein has been mapped to identify which epitopes are recognized in different animal models [3,47-50]. In this study, we used spot synthesis technology to map the epitopes of TMV that interact with B and T cells. The structural properties of TMV and TMVcp predicted by bioinformatics were also used to support our experimental results.

No significant binding was predicted for the first 42 residues of the TMVcp, in contrast to previous results for the first ten residues [47] and for residues 19-32 [48] using polyclonal sera. The positive regions that we identified agreed with those reported in the literature, i.e., regions corresponding to residues 34-39 [47], 55-61 [51] and 62-68 [49], with the strongest binding occurring in residues 34-51 and 43-60. The region involving residues 58-81 also reacted positively with polyclonal sera (specifically, regions 55-61 and 62-68). Studies with poly- and monoclonal antibodies (MAbs) have identified positive regions involving regions 80-90, 105-112, 115-134, 134-146 and 153-156 [52-54]. In agreement with this, we identified positive binding in the region 85-108, with the strongest signal in region 88-102 (spot 30). No binding was detected in the region containing residues 97-132, in agreement with other reports [52,53] for regions containing residues 121-144 (spots 41 and 42) and 133-159 (spots 45 to 49). These findings contrast with previous reports that have identified binding in this region [10,54].

With regard to the model used to investigate the antigenic profile of TMV, Anderer and Ströbel [50] used tryptic peptides and an ELISA-fixed complement detection system with polyclonal and monoclonal antibodies [53], whereas Al Moudallal et al. [52] used sera from immunized chickens and MAbs, and Altschuh and Van Regenmortel [51] used MAbs to map protein epitopes. Subsequent reports indicated that longer proteins should be analyzed in order to correctly address

the antigen presentation process [4,55,56]. The spot technique permits such an analysis and fast, accurate epitope mapping can be achieved by restricting the experiment to a mouse model, with epitope predictions specific for this model.

The antigenic determinants of proteins are generally located in exposed polar, hydrophilic regions. Consequently, the accessibility and flexibility of these segments are greater than those of other potential epitopes located deeper in the three-dimensional structure of the molecule. As shown here, a Kyte-Doolittle [20] plot of the TMVcp sequence revealed that the regions involving amino acid residues 30-78, 80-120 and 130-150 had negative values (Figure 3). These regions corresponded to the epitopes with higher binding intensities identified in the prediction analyses. Similar results have been described elsewhere using different, less accurate methods and animal models [3,4,57,58].

Several studies have demonstrated a correlation between experimental data and bioinformatics prediction results for complex processes such as proteasomal cleavage, TAP binding affinity and MHC binding [17,33,59-63]. A similar correlation was also observed here for the binding of TMVcp-derived peptides to MHC Class I and Class II in mice. Since TMV does not replicate in mice it is unlikely that mechanisms other than the usual intracellular antigen processing pathway were activated, i.e., there was probably no activation of antigen processing cells (APC) such as macrophages and dendritic cells that express only MHC class II epitope binding sites. In this context, the proteasome/TAP/MHC I predictions in the present study were done only for comparison since no mouse cells express receptors for TMV, although dendritic cells undergo autophagy and express both classes of MHC [64-66]. The Balb/c alleles studied here were chosen from the IEDB site. A number of studies have shown the effect of amino

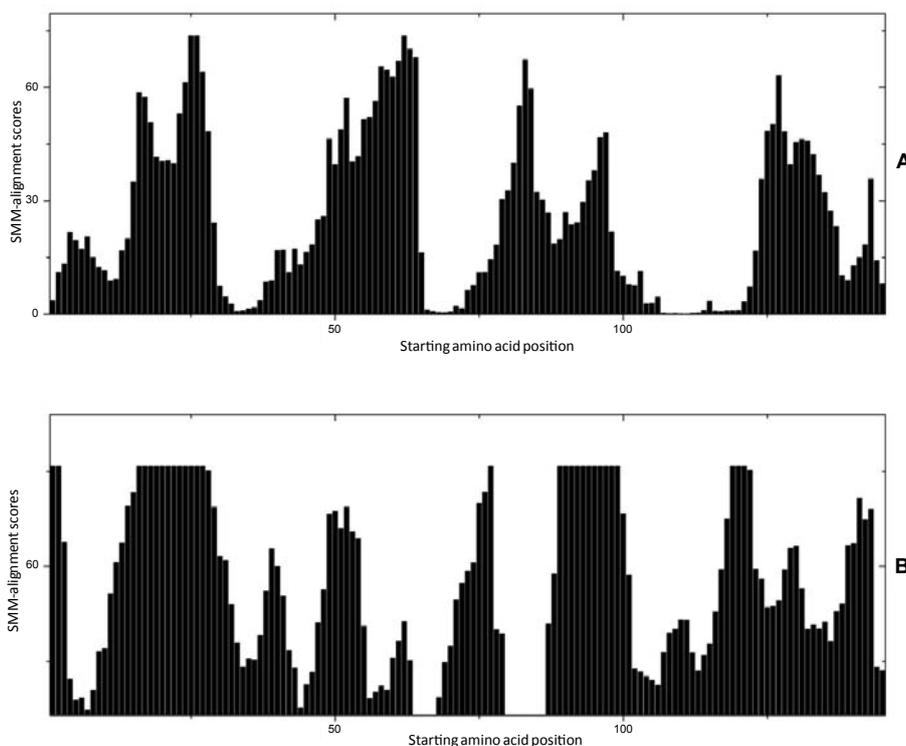


Figure 9: Identification of the TMVcp regions capable of binding to MHC class II epitopes [61]. The binding affinities of TMVcp pentadecapeptides to mouse MHC class II were predicted based on SMM alignments for alleles H2-IA_d (A) and H2-IE_d (B). In both panels, the higher the score the greater the affinity of the pentadecapeptide for MHC class II.

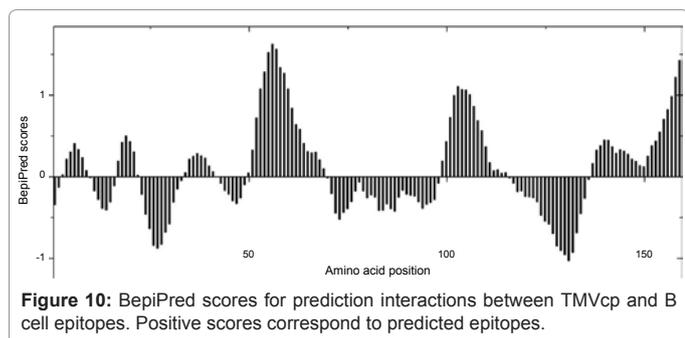


Figure 10: BepiPred scores for prediction interactions between TMVcp and B cell epitopes. Positive scores correspond to predicted epitopes.

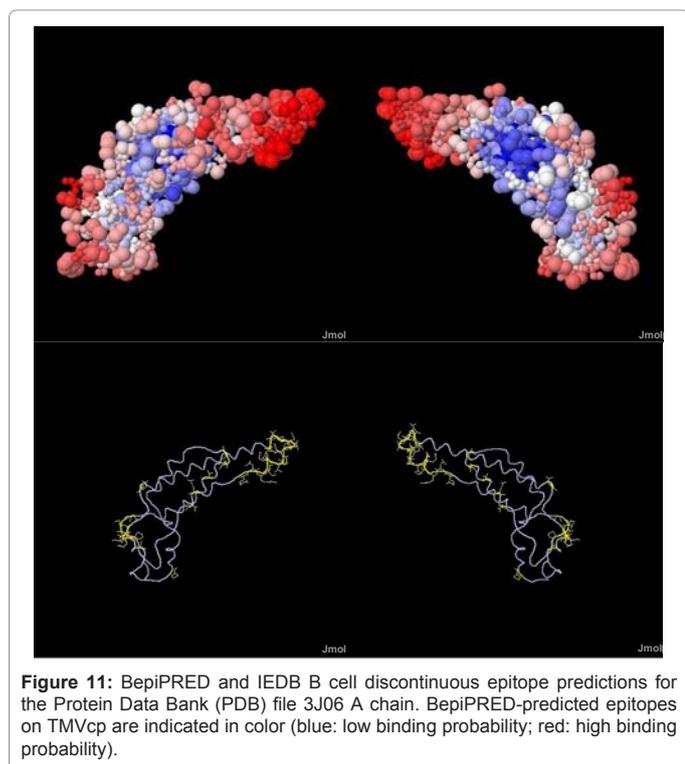


Figure 11: BepiPred and IEDB B cell discontinuous epitope predictions for the Protein Data Bank (PDB) file 3J06 A chain. BepiPred-predicted epitopes on TMVcp are indicated in color (blue: low binding probability; red: high binding probability).

acid replacements on protein function, e.g., the effect of mutations on the activity of influenza A polymerase [67,68]. Modification of capsid proteins may lead to a defective virus and consequently to an inability to assemble the virion particle, thereby hindering new infections [69,70]. We did not examine the functional domains of TMVcp since the ExPaSy functional domain prediction software did not find any additional functional domains in the TMV amino acid sequence used as the query sequence.

Hydrophobicity, flexibility, accessibility, the presence of turns, an exposed surface, polarity and antigenic propensity were parameters used to predict B cell epitopes. In general, turns occur on the surface of proteins and are generally accessible and hydrophilic. In contrast, the core is mostly devoid of water molecules [27]. In the case of herpes simplex virus type IVP5 capsid protein the secondary structure is important for antibody binding and even a minor modification can affect antigen identification by the immune system [71,72]. TMVcp has at least three major hydrophilic regions that span residues 30-70, 90-120 and 140-150; all three regions were also found to have higher binding capacities.

Based on the cutoff used for the binding of mouse IgG to the

membrane the TMV epitopes were found to be located in residues 31-60, 85 - 108 and 133 - 159. The IEDB tools showed that these regions had a higher probability of proteasomal cleavage. These results were complemented by the predictions of TAP-binding affinity. The predictions of binding to MHC class I took into account the combined results of the three prediction methods and the alleles used for this approach were based on the Balb/c immunological profile, which includes only three MHC alleles (Dd, Kd and Ld) available from the IEDB and CBS servers. The variation in the predicted binding observed with this method probably reflected structural differences in the MHCs that influenced MHC-peptide interactions [73]. In this context, *in silicio* analysis revealed different possible bindings for each allele tested and the detection of antibody binding to those were experimentally confirmed (Figures 1,2,7). In general, the CBS server analyzed residues individually whereas the IEDB used 9-mers in its analysis of each allele in the three major antigenic regions identified experimentally.

These three antigenic regions were also evaluated with the IDEB and CBS servers using MHC class II prediction software. The regions predicted to show high affinity binding were also found to be positive in the spot synthesis. Peptides derived from the segment containing residues 37-69 had lower IC_{50} scores by the SMM method in allele IAd whereas for allele IEd the lowest scores were observed with the longest segments. In the latter case, the peptides were derived from a segment containing residues 34-85. These findings suggest that these regions are not only immunogenic and must be further investigated. However, we could not confirm it experimentally at this time. When comparing the binding affinities of the restricted alleles for MHC class I and class II receptors one can find that the regions predicted to be of high affinity were different among the predictions.

The B cell prediction software found this region to be of great importance, with the highest values being found for residues 46-66, in agreement with our experimental results. The discontinuous B cell epitope prediction found this same region to be important and longer (stretching from residue 33 to 90), with two peaks being predicted within this region. These two peaks are separated only by a few residues (61-63), which markedly decreased the binding probability according to the algorithm used. The identification of these residues agreed with the theory governing their prediction.

B cells are of particular interest because their surface receptors are formed by rearrangement mediated by IgGs. Consequently, positive epitopes recognized by mouse IgGs and possibly presented via MHC II may be associated with B cell predictions through crosstalk with APCs. This possibility was examined by using several B epitope prediction engines, the results of which are shown in Figure 8. The parameters mentioned above were correlated to the location of continuous epitopes. As a result, five TMVcp regions were predicted to be B-cell epitopes. The shortest epitope was the first in the sequence and spanned residues 18-20 while the longest one was the second and spanned residues 52-64. The IAd and IEd alleles were chosen for this analysis. Peptides below the IC_{50} cutoff, i.e., those with the best predicted binding affinities, are shown in Figure 9.

The results described here agree with other reports for this virus and data from reference centers on TMV studies support our spot synthesis results; however, the epitopes that we identified differed from those reported in these other studies. In Balb/c mice, the inoculation of TMV the first ten overlapping pentadecapeptides of TMVcp did not produce marked signs of antibody response (Figures 1,2). Van Regenmortel [3] reported that these regions are important epitopes. However, the use of several different models to obtain the results described by these authors

could have led to different immune responses against the same antigen (as shown for MHC class I alleles in Balb/c mice) thereby making it difficult to interpret their findings.

In conclusion, the findings reported expand the epitope mapping of the TMV on mice models. We found discordant regions first regarded as well-defined epitopes that did not bind antibodies in the present study, suggesting a more careful exploration of different antigen processing machineries of the animal models chosen. Using prediction algorithms for comparison the results produced were in agreement with the experimental data. Epitope mapping is an effective tool for measuring and identifying which regions from a given protein binds effectively any antibody response and if used together with *in silico* predictions should provide a powerful instrument for deciding best options for epitope-based vaccine design, which can be confirmed thoroughly and further assessment of those decisions. Future studies should assess the extent to which the exposure of different protein surfaces in mutant TMV can elicit different immune responses.

Acknowledgements

The authors thank Ronald Frank and Susanne Daenicke (Helmholtz Centre for Infection Research in Braunschweig, Germany) for generous preparation of the peptide SPOT array, Ricardo Aparicio (Institute of Chemistry, UNICAMP) for helpful discussions about TMV structure, and Stephen Hyslop for editing the English of the manuscript. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil.

References

- Rappaport I, Siegel A, Haselkorn R (1965) Influence of the state of subunit aggregation on the antigenic specificity of TMV and TYMV. *Virology* 25: 325-328.
- Rappaport I (1965) The antigenic structure of tobacco mosaic virus. *Adv Virus Res* 11: 223-275.
- Van Regenmortel MH (1999) The antigenicity of tobacco mosaic virus. *Philos Trans R Soc Lond B Biol Sci* 354: 559-568.
- Van Regenmortel MH, Altschuh D, Klug A (1986) Influence of local structure on the location of antigenic determinants in tobacco mosaic virus protein. *Ciba Found Symp* 119: 76-92.
- Klug A (1999) The tobacco mosaic virus particle: structure and assembly. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 354: 531-535.
- Bonafe CF, Vital CM, Telles RC, Gonçalves MC, Matsuura MS, et al. (1998) Tobacco mosaic virus disassembly by high hydrostatic pressure in combination with urea and low temperature. *Biochemistry* 37: 11097-11105.
- Santos JL, Bispo JA, Landini GF, Bonafe CF (2004) Proton dependence of tobacco mosaic virus dissociation by pressure. *Biophys Chem* 111: 53-61.
- Santos JL, Aparicio R, Joekes I, Silva JL, Bispo JA, et al. (2008) Different urea stoichiometries between the dissociation and denaturation of tobacco mosaic virus as probed by hydrostatic pressure. *Biophys Chem* 134: 214-224.
- Morrow PR, Rennick DM, Leung CY, Benjamini E (1984) The antibody response to a single antigenic determinant of the tobacco mosaic virus protein: analysis using monoclonal antibodies, mutant proteins and synthetic peptides. *Mol Immunol* 21: 301-309.
- Altschuh D, Al Moudallal Z, Briand JP, Van Regenmortel MH (1985) Immunochemical studies of tobacco mosaic virus-VI. Attempts to localize viral epitopes with monoclonal antibodies. *Mol Immunol* 22: 329-337.
- Petrovsky N, Silva D, Brusica V (2003) The future for computational modelling and prediction systems in clinical immunology. *Novartis Found Symp* 254: 23-32.
- Kim Y, Sette A, Peters B (2011) Applications for T-cell epitope queries and tools in the Immune Epitope Database and Analysis Resource. *J Immunol Methods* 374: 62-69.
- Roopm K, Antes I, Lengauer T (2010) Predicting MHC class I epitopes in large datasets. *BMC Bioinformatics* 11: 90.
- Asselin A, Zaitlin M (1978) Characterization of a second protein associated with virions of tobacco mosaic virus. *Virology* 91: 173-181.
- Frank R, Overwin H (1996) SPOT synthesis. Epitope analysis with arrays of synthetic peptides prepared on cellulose membranes. *Methods Mol Biol* 66: 149-169.
- Beutling U, Stading K, Stradal T, Frank R (2008) Large-scale analysis of protein-protein interactions using cellulose-bound peptide arrays. *Adv Biochem Eng Biotechnol* 110: 115-152.
- Soutullo A, Santi MN, Perin JC, Beltramini LM, Borel IM, et al. (2007) Systematic epitope analysis of the p26 EIAV core protein. *J Mol Recognit* 20: 227-237.
- Karplus PA, Schulz, G.E. (1985) Prediction of Chain Flexibility in Proteins - A tool for the Selection of Peptide Antigens. *Naturwissenschaften* 72: 2.
- Kolaskar AS, Tongaonkar PC (1990) A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett* 276: 172-174.
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105-132.
- Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci U S A* 78: 3824-3828.
- Diez-Rivero CM, Chenlo B, Zuluaga P, Reche PA (2010) Quantitative modeling of peptide binding to TAP using support vector machine. *Proteins* 78: 63-72.
- Emmerich NP, Nussbaum AK, Stevanovic S, Priemer M, Toes RE, et al. (2000) The human 26 S and 20 S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. *J Biol Chem* 275: 21140-21148.
- Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, et al. (2001) Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med* 194: 1-12.
- Tenzer S, Stoltze L, Schönfisch B, Dengjel J, Müller M, et al. (2004) Quantitative analysis of prion-protein degradation by constitutive and immunoproteasomes indicates differences correlated with disease susceptibility. *J Immunol* 172: 1083-1091.
- Bhasin M, Lata S, Raghava GP (2007) TAPPred prediction of TAP-binding peptides in antigens. *Methods Mol Biol* 409: 381-386.
- Pellequer JL, Westhof E, Van Regenmortel MH (1991) Predicting location of continuous epitopes in proteins from their primary structures. *Methods Enzymol* 203: 176-201.
- Daniel S, Brusica V, Caillat-Zucman S, Petrovsky N, Harrison L, et al. (1998) Relationship between peptide selectivities of human transporters associated with antigen processing and HLA class I molecules. *J Immunol* 161: 617-624.
- Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA (2008) PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* 36: W35-41.
- Wingfield PT, Stahl SJ, Thomsen DR, Homa FL, Booy FP, et al. (1997) Hexon-only binding of VP26 reflects differences between the hexon and penton conformations of VP5, the major capsid protein of herpes simplex virus. *J Virol* 71: 8955-8961.
- Warren RL, Holt RA (2010) A census of predicted mutational epitopes suitable for immunologic cancer control. *Hum Immunol* 71: 245-254.
- Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, et al. (2008) Immune epitope database analysis resource (IEDB-AR). *Nucleic Acids Res* 36: W513-518.
- Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, et al. (2008) NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* 36: W509-512.
- Goelet P, Lomonosoff GP, Butler PJ, Akam ME, Gait MJ, et al. (1982) Nucleotide sequence of tobacco mosaic virus RNA. *Proc Natl Acad Sci U S A* 79: 5818-5822.
- Karpova O, Nikitin N, Chirkov S, Trifonova E, Sheveleva A, et al. (2012) Immunogenic compositions assembled from tobacco mosaic virus-generated spherical particle platforms and foreign antigens. *J Gen Virol* 93: 400-407.
- Nielsen M, Lundegaard C, Lund O (2007) Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* 8: 238.

37. Nussbaum AK, Kuttler C, Haderer KP, Rammensee HG, Schild H (2001) PAMProC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* 53: 87-94.
38. KeÄmir C, Nussbaum AK, Schild H, Detours V, Brunak S (2002) Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* 15: 287-296.
39. Peters B, Bulik S, Tampe R, Van Endert PM, Holzhütter HG (2003) Identifying MHC class I epitopes by predicting the TAP transport efficiency of epitope precursors. *J Immunol* 171: 1741-1749.
40. Zhang GL, Petrovsky N, Kwoh CK, August JT, Brusica V (2006) PRED(TAP): a system for prediction of peptide binding to the human transporter associated with antigen processing. *Immunome Res* 2: 3.
41. Lam TH, Mamitsuka H, Ren EC, Tong JC (2010) TAP Hunter: a SVM-based system for predicting TAP ligands using local description of amino acid sequence. *Immunome Res* 6 Suppl 1: S6.
42. Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, et al. (2005) Modeling the MHC class I pathway by combining predictions of proteasomal cleavage, TAP transport and MHC class I binding. *Cell Mol Life Sci* 62: 1025-1037.
43. Sette A, Adorini L, Colon SM, Buus S, Grey HM (1989) Capacity of intact proteins to bind to MHC class II molecules. *J Immunol* 143: 1265-1267.
44. Stern LJ, Wiley DC (1994) Antigenic peptide binding by class I and class II histocompatibility proteins. *Behring Institute Mitteilungen*: 1-10
45. Borghesi L, Milcarek C (2006) From B cell to plasma cell: regulation of V(D)J recombination and antibody secretion. *Immunol Res* 36: 27-32.
46. Santos P, Arumemi F, Park KS, Borghesi L, Milcarek C (2011) Transcriptional and epigenetic regulation of B cell development. *Immunol Res* 50: 105-112.
47. Altschuh D, Hartman D, Reinbolt J, Van Regenmortel MH (1983) Immunochemical studies of tobacco mosaic virus--V. Localization of four epitopes in the protein subunit by inhibition tests with synthetic peptides and cleavage peptides from three strains. *Mol Immunol* 20: 271-278.
48. Al Moudallal Z, Briand JP, Van Regenmortel MH (1982) Monoclonal antibodies as probes of the antigenic structure of tobacco mosaic virus. *EMBO J* 1: 1005-1010.
49. De L Milton RC, van Regenmortel MH (1979) Immunochemical studies of tobacco mosaic virus--III. Demonstration of five antigenic regions in the protein sub-unit. *Mol Immunol* 16: 179-184.
50. Anderer FA, Ströbel G (1972) Recognition of conjugated and native peptide determinants. II. Enhancement of antibodies with sequential specificity in anti-TMV sera by preimmunization with conjugated peptide antigens. *Eur J Immunol* 2: 278-282.
51. Altschuh D, Van Regenmortel MH (1982) Localization of antigenic determinants of a viral protein by inhibition of enzyme-linked immunosorbent assay (ELISA) with tryptic peptides. *J Immunol Methods* 50: 99-108.
52. Al Moudallal Z, Altschuh D, Briand JP, Van Regenmortel MH (1984) Comparative sensitivity of different ELISA methods for detecting monoclonal antibodies to viruses. *Dev Biol Stand* 57: 35-40.
53. ANDERER FA (1963) RECENT STUDIES ON THE STRUCTURE OF TOBACCO MOSAIC VIRUS. *Adv Protein Chem* 18: 1-35.
54. Benjamini E, Leung CY, Rennick DM (1978) Immunochemical studies on the tobacco mosaic virus protein. *Adv Exp Med Biol* 98: 165-179.
55. Dekker EL, Porta C, Van Regenmortel MH (1989) Limitations of different ELISA procedures for localizing epitopes in viral coat protein subunits. *Arch Virol* 105: 269-286.
56. Dore I, Altschuh D, Al Moudallal Z, Van Regenmortel MH (1987) Immunochemical studies of tobacco mosaic virus--VII. Use of comparative surface accessibility of residues in antigenically related viruses for delineating epitopes recognized by monoclonal antibodies. *Mol Immunol* 24: 1351-1358.
57. Van Regenmortel MH, Altschuh D, Zeder-Lutz G (1993) Tobacco mosaic virus: a model antigen to study virus-antibody interactions. *Biochimie* 75: 731-739.
58. Van Regenmortel MHV (1996) Mapping Epitope Structure and Activity: From One-Dimensional Prediction to Four-Dimensional Description of Antigenic Specificity Methods 9: 465-472.
59. Tsurui H, Takahashi T (2007) Prediction of T-cell epitope. *J Pharmacol Sci* 105: 299-316.
60. Mohabatkari H (2007) Prediction of epitopes and structural properties of Iranian HPV-16 E6 by bioinformatics methods. *Asian Pac J Cancer Prev* 8: 602-606.
61. Schuler MM, Nastke MD, Stevanović S (2007) SYFPEITHI: database for searching and T-cell epitope prediction. *Methods Mol Biol* 409: 75-93.
62. Lara J, Wohlhueter RM, Dimitrova Z, Khudyakov YE (2008) Artificial neural network for prediction of antigenic activity for a major conformational epitope in the hepatitis C virus NS3 protein. *Bioinformatics* 24: 1858-1864.
63. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M (2007) Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* 23: 3265-3275.
64. Gruber A, Brocker T (2005) MHC class I-positive dendritic cells (DC) control CD8 T cell homeostasis in vivo: T cell lymphopenia as a prerequisite for DC-mediated homeostatic proliferation of naive CD8 T cells. *J Immunol* 175: 201-206.
65. Chemali M, Radtke K, Desjardins M, English L (2011) Alternative pathways for MHC class I presentation: a new function for autophagy. *Cell Mol Life Sci* 68: 1533-1541.
66. Neeffjes J, Jongma ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* 11: 823-836.
67. Trifonov V, Racaniello V, Rabadan R () The Contribution of the PB1-F2 Protein to the Fitness of Influenza A Viruses and its Recent Evolution in the 2009 Influenza A (H1N1) Pandemic Virus. Version 3.
68. Ozawa M, Basnet S, Burley LM, Neumann G, Hatta M, et al. (2011) Impact of amino acid mutations in PB2, PB1-F2, and NS1 on the replication and pathogenicity of pandemic (H1N1) 2009 influenza viruses. *J Virol* 85: 4596-4601.
69. Grättinger M, Hohenberg H, Thomas D, Wilk T, Müller B, et al. (1999) In vitro assembly properties of wild-type and cyclophilin-binding defective human immunodeficiency virus capsid proteins in the presence and absence of cyclophilin A. *Virology* 257: 247-260.
70. Kuhlmann U, Sarkar S, Rohde W (1993) Capsid protein gene sequences of four tobacco mosaic virus strains defective for virus assembly. *Nucleic Acids Res* 21: 3325.
71. Schlosser G, Mezo G, Kiss R, Vass E, Majer Z, et al. (2003) Synthesis, solution structure analysis and antibody binding of cyclic epitope peptides from glycoprotein D of Herpes simplex virus type I. *Biophys Chem* 106: 155-171.
72. Homa FL, Brown JC (1997) Capsid assembly and DNA packaging in herpes simplex virus. *Rev Med Virol* 7: 107-122.
73. Grey HM, Ruppert J, Vitiello A, Sidney J, Kast WM, et al. (1995) Class I MHC-peptide interactions: structural requirements and functional implications. *Cancer Surv* 22: 37-49.