



In silico analysis and modeling of putative T cell epitopes for vaccine design of Toscana virus

Amisha Jain · Pranav Tripathi · Aniket Shrotriya ·
Ritu Chaudhary · Ajeet Singh

Received: 20 May 2014 / Accepted: 6 August 2014 / Published online: 26 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The sandfly fever Toscana virus is an important etiological agent known to cause human neurological infections in endemic Mediterranean countries during summer season. In the present study, prediction and modeling of T cell epitopes of Toscana virus (TOSV) antigenic proteins followed by the binding simulation studies of predicted highest binding scorers with their corresponding MHC class II alleles were done. Immunoinformatics was applied in computational vaccinology to analyze the viral proteins which generate possible outcomes to elicit vaccine for TOSV. Here, immunoinformatic tool ProPred was used to predict the promiscuous MHC class II epitopes of viral antigenic proteins. The molecular modeling of the selected epitopes as well as MHC alleles was done at CPH model 3.2 server. Molecular dynamics (MD) simulation studies were performed through the NAMD graphical user interface embedded in visual molecular dynamics. The epitope/peptide VKMMIVLNL of viral nucleoprotein as well as VMILGLLSS of viral glycoprotein has shown the highest binding score with the same DRB1*1104 MHC II allele. These two predicted peptides are highly potential to induce T cell-mediated immune response and are expected to be useful in designing epitope-based vaccines after further testing. The results signify that the nucleoprotein, glycoprotein or the combination of both could be useful for future development of a vaccine controlling the spread of this emerging virus that could pose a new threat for humans.

Keywords Toscana virus · ProPred · Structural modeling · Docking · Molecular dynamics · Vaccine design

Introduction

Toscana virus (TOSV) is an arbovirus (arthropod-borne virus), which belongs to the family *Bunyaviridae* and genus *Phlebovirus* (Bishop et al. 1990; Elliott et al. 1991). There are at least 45 members of the Phlebovirus genus, which are mostly transmitted to vertebrates by phlebotomine sandflies. Its genome is comprised of an enveloped negative-strand RNA that consists of three segments, designated as large (L), medium (M) and small (S), which encodes the viral RNA polymerase, the envelope glycoprotein and the nucleoprotein, respectively (Olal et al. 2014). The segment L is the largest about 6,400 nucleotides, M is about 4,200 nucleotides and S about 1,900 nucleotides in length. TOSV is among the three most prevalent viruses associated with meningitis during the warm seasons and, therefore, must be considered as an emerging pathogen. Currently, there is no effective antiviral treatment or vaccine for the infection caused by TOSV; the best prevention will likely to come by development of effective and safe vaccine.

Naples and *Sicilian* serocomplexes are the two main serocomplexes of sandfly fever viruses associated with human diseases. Sandfly fever Toscana virus (SFTV) belongs to *Naples* serocomplex and is endemic in Mediterranean countries like Italy, Portugal, Spain, and Cyprus. Within the group of sandfly-transmitted *phleboviruses*, TOSV is the only virus known to cause Central Nervous System (CNS) manifestations (Braitto et al. 1998; Cusi et al. 2010). Data from Italy, France and Spain indicate that it is among the three leading causes of aseptic meningitis during periods of sandfly activity (Charrel et al. 2005). The

A. Jain · P. Tripathi · R. Chaudhary · A. Singh (✉)
Department of Biotechnology, G B Pant Engineering College,
Pauri Garhwal, Uttarakhand, India
e-mail: ajeetsoniyal@gmail.com

A. Shrotriya
Department of Biotechnology, Delhi Technological University,
Delhi, India

other viruses are associated with systemic febrile illness, but not CNS infections. TOSV is the most important etiologic agent of meningitis, meningoencephalitis, or encephalitis associated with fever, severe frontal headache, vomiting, myalgia, ocular pain and, nuchal rigidity (Braito et al. 1998; Dionisio et al. 2001; Nicoletti et al. 1991; Valassina et al. 1998; 2000).

Vaccination is generally considered as a cost-effective weapon for preventing life-threatening infectious diseases. Awareness of differences in epitopes recognized by T cells and B cells has enabled immunologists to begin designing vaccine candidates to maximize activation of both arms of the immune system. Major histocompatibility complex (MHC) plays an important role in intercellular recognition of foreign antigenic peptides and thus participates in the development of both humoral and cell-mediated immune responses. T cells are present on the surface of antigen-presenting cells (APCs) and they recognize antigenic fragments only when they are combined with MHC molecules exposed on the surface of all vertebrate cells (Shekhar et al. 2012; Mohabatkar and Mohammadzadegan 2007). MHC molecules are heterodimeric glycoproteins which present a highly diverse set of peptides on the surface of a cell and thus induce T cell activation, hence play a profound role in immune response regulation (Viret and Janeway 1999; Tambunan and Parikesit 2011). As they are highly polymorphic in nature, the binding region for one allele may not induce an immune response in a population with different alleles. This raises the need for the identification of promiscuous viral peptides which can bind to multiple MHC alleles (Jiao et al. 2012).

Epitope- or peptide-based vaccines offer several advantages over conventional vaccines as they are more specific, easy to produce, cost-effective, less time-consuming and also safe. Apart from these, the main advantage of epitope- or peptide-based vaccine is the ability to deliver high doses of potential immunogen at low cost (Von Hoff et al. 2005; Tang et al. 2012). The viral protein that could act as a vaccine candidate must be antigenic and responsible for pathogenicity (Cerdino-Tarraga et al. 2003; Verma et al. 2011).

Materials and methods

Retrieval of Toscana virus antigenic proteins

The complete sequences of viral nucleoprotein (NP) and glycoprotein (GP) were retrieved from National Center for Biotechnology Information (NCBI), having accession code, ABY19521 and AGP51529, with length 253 and 1,339 amino acids, respectively (<http://www.ncbi.nlm.nih.gov/protein>).

T cell epitope prediction by ProPred

Several different prediction methods have been employed for the identification of promiscuous T cell epitopes. Identification of various T cell epitopes by more than one MHC allele and their recognition by more than one T cell clone are known as “promiscuous” epitopes. ProPred, the graphical web-based tool was used in this study to predict MHC class II binding regions in an antigenic sequence, using quantitative matrices that assist in locating promiscuous binding regions that are useful in selecting vaccine candidates (Sturniolo et al. 1999). In ProPred, the NP and GP sequences were analyzed, taken into consideration a total of 51 MHC class II alleles. Each nonamer comprises an anchor or starting residue that helps to identify those antigenic determinants in NP and GP which binds to several Human Leukocyte Antigen (HLA) class II molecules with good binding affinity. The predicted binders can be visualized either as peaks in graphical interface or as colored residues in HTML interface. Peptides displaying highest score were selected for immunoinformatic analysis.

Modeling and validation of favored MHC II alleles and epitopes

The NP epitope (VKMMIVLNL) and GP epitope (VMILGLLSS) of TOSV were modeled at CPH model 3.2 server. The 3-D structure of DRB1*1104 MHC allele was unavailable in the Protein Data Bank (PDB) server, therefore, CPH model 3.2 server was used to acquire its three-dimensional coordinate files employing its respective structural template (1A6A.B). The structure of MHC allele was then validated using various online structural analysis and verification servers, viz., Errat, ProQ, ProSA, RAMPAGE, etc. Errat (Colovos and Yeates 1993), a protein structure verification algorithm was especially used to analyze the statistics of non-bonded interactions between different atom types. The quality of a protein model was predicted by a protein quality predictor tool, ProQ (Wallner and Elofsson 2003) on the basis of two quality measures, LG score and MaxSub. A web-based Protein Structure Analysis tool ProSA (Wiederstein and Sippl 2007) calculates an overall quality score of specific protein structure by determining the overall model quality (Z-score) and local model quality. RAMPAGE (Lovell et al. 2002) was used for Ramachandran Plot Assessment of protein model.

Binding simulation of epitopes with the alleles

After structural modeling, rigid docking of selected peptides and alleles was accomplished with the help of Hex 8.0 (Ritchie et al. 2008) followed by semi-flexible docking using AutoDock Vina (Trott and Olson 2010) to determine

the energy minimization values. In Hex, the prediction of energy minimization is based on various parameters including correlation type of shape only, 3-D FFT mode, Grid dimension of 0.6, receptor range of 180 with step size 7.5 and ligand range of 180 with step size 7.5. Semi-flexible docking was achieved using AutoDock Vina which involves choosing of macromolecule, adding polar H-bonds, computing Gasteiger, and Kollman charges for the receptor. Then, for ligand all bonds, except amide bonds, were made rotatable. The docked complexes were finally visualized and analyzed using Autodock Tools.

Molecular dynamics simulation of epitope and HLA allele complex

Molecular dynamics simulation was done using NAMD graphical interface module (James et al. 2005) incorporated visual molecular dynamics (VMD) (Humphrey et al. 1996). Automatic PSF builder tool of VMD was used to make the PSF file and by accessing PSF and PDB files, NAMD generated the trajectory DCD file. Root Mean Square Deviation (RMSD) of the complex was completed using rmsd.tcl source file from the Tk console and finally rmsd.dat was saved and accessed in Microsoft office excel 2007.

Results

Prediction and analysis of MHC Class II binding peptides

The NP peptide VKMMIVLNL at position 58–66 showed maximum binding score of 62.65 with DRB1*1104 MHC II allele. The GP peptide VMILGLLSS at position 824–832 showed maximum binding score of 73.49 with the same DRB1*1104 allele. ProPred scores and the no. of MHC II alleles showing binding affinities for epitopes are shown below (Table 1).

Modeling and validation of target peptides and MHC allele

3-D coordinate files of preferred peptides and allele were obtained through CPH model server. The structure of

DRB1*1104 allele was validated using Errat, ProQ, ProSA, and RAMPAGE web-based tools. The overall quality factor of allele as determined by Errat was found to be 67.039. Quality of a model as predicted by ProQ was ‘fairly good’ with LG score of 2.096 and MaxSub of 0.131. Using ProSA, the calculated Z-score of above-mentioned allele was -5.52 (Fig. 1) and the local model quality was also good to consider. Through RAMPAGE analysis, the number of amino acid residues in favored region was 177 (95.7 %) out of 187 amino acid residues.

Binding energy determination by Hex 8.0 and AutoDock Vina

Binding simulation studies of NP epitope VKMMIVLNL with DRB1*1104 allele as well as GP epitope VMILGLLSS with DRB1*1104 allele formed stable HLA-peptide complexes with the energy minimization values of -5.5 and -6.0 kcal/mol, respectively (Table 2). After docking studies, we determined the number of H-bonds present in the stable complex formed. Using AutoDock Vina, it was found that two H-bonds were present in NP-DRB1*1104 complex at positions CYS44:HN1 and GLY154:HN1 (Fig. 2), whereas two H-bonds were present in GP-DRB1*1104 complex at positions CYS44:HN1 and PHE46:HN1 (Fig. 3). While using Hex 8.0, the predicted number of H-bond was 1 in both the complexes with good E-total values. Thus, it shows maximum probability of proper binding as there is possibility of formation of H-bonds.

Molecular dynamics simulation of peptide-allele complex through NAMD

The peptide-allele complexes formed by Hex 8.0 were subjected to molecular dynamics simulation and RMSD. NP peptide VKMMIVLNL-DRB1*1104 allele complex displayed the highest peak at RMSD value of 6.8 \AA (Fig. 4). On the other hand, GP peptide VMILGLLSS-DRB1*1104 allele complex resulted in highest peak at RMSD value of 11 \AA (Fig. 5). As displayed by the graph, the NP-allele complex runs till 10,500 picoseconds, while the GP-allele complex runs till 6,000 picoseconds which is lesser than the time value of NP-allele complex.

Table 1 Scores generated by ProPred

Protein	Position	Peptide/epitope	ProPred Score	HLA alleles	No. of reactivated MHC alleles
NP	58–66	VKMMIVLNL	62.65	DRB1*1104	51
GP	824–832	VMILGLLSS	73.49	DRB1*1104	49

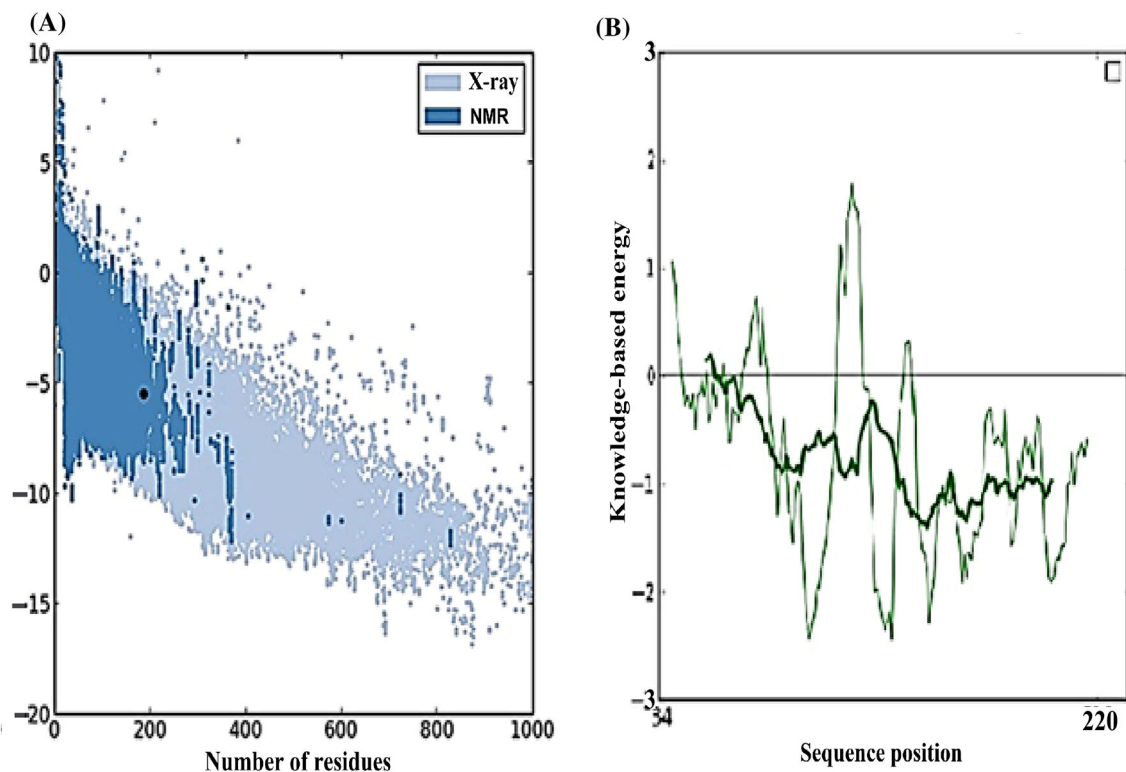


Fig. 1 ProSA analysis. **a** Z-score plot of DRBI*1104 allele. **b** Graphical plot of residue scores of allele structure

Table 2 Energy minimization results of docking studies with Hex 8.0 and Autodock Vina followed by the identification of H-bonds

Protein	Peptide/Epitope	Hex 8.0		Autodock Vina		HLA alleles
		E-total	H-bonds	Energy	H-bonds	
NP	VKMMIVLNL	-446.19	1	-5.5	2	DRB1*110
GP	VMILGLLSS	-301.04	1	-6.0	2	DRB1*1104

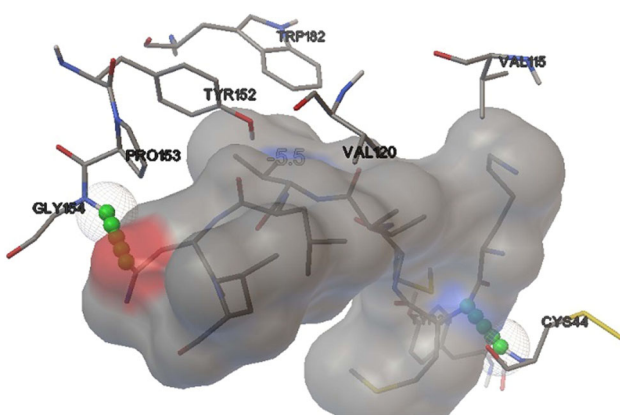


Fig. 2 Docked NP peptide VKMMIVLNL-DRB1*1104 allele complex by Autodock Vina depicting detailed position of amino acids along with formation of 2 H-bonds with CYS44 and GLY154

Discussion

The pathogenic activity of Toscana virus creates a challenge to develop novel vaccines. Addressing these challenges involves a time-consuming battery of in vivo and in vitro experiments. The growth of virus and isolation of peptides on a large array of experiments is also a huge and valid constraint. The current study incorporates immunoinformatics approach for reducing the time consumed in the long array of experiments to avoid hit and trial sets. Valassina et al. 2003 reported the antigenic properties of the viral proteins (NP and GP) to better understand their immunogenic role. It has been also reported that the viral NP is highly immunogenic for many members in the genus of *Phlebovirus* and is a major antigen that acts as a scaffold for packing of virus (Martin-Folgar et al. 2010; Magurano et al. 1999). A glycoprotein of length 1,339 amino acids

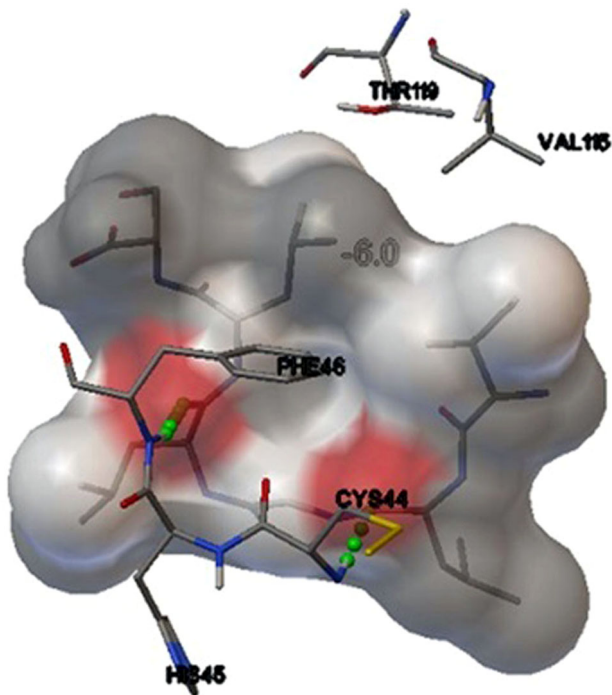


Fig. 3 Docked GP peptide VMILGLLSS-DRB1*1104 allele complex depicting position of amino acids along with formation of 2 H-bonds with CYS44 and PHE46

has possibility of holding large number of epitopes. So, two proteins, one 253 amino acids long (NP) and the other 1,339 amino acids long (GP), was taken into consideration, so that T cell bound MHC II epitopes can be predicted. Only one sequence variant for each of the two proteins was studied because NP is highly conserved and according to Collao et al. 2009 geographic differences in genotype distribution relate to differences in vector distribution only. For nucleoprotein, VKMMIVLNL peptide was predicted with the highest score of 62.65 % with DRB1*1104 allele. The homology modeling of the above-mentioned allele was satisfactorily validated with an Errat score of 67.039, LG score of 2.096 by ProQ, Z-score of -5.52 by ProSA. Ramachandran plot analysis through RAMPAGE resulted in 177 (95.7 %) amino acids in favored region. The rigid docking approach was carried out successfully by Hex 8.0 on a correlation type of shape in 3-D FFT mode. The E-total value of the predicted peptide was -446.19 using Hex 8.0. Semi-flexible approach used AutoDock Vina at default parameters which resulted in binding affinity of -5.5 kcal/mol along with two H-bonds. First H-bond was detected with CYS44:HN1 and second was predicted with GLY154:HN1. The molecular dynamics simulation showed that complex formed between a peptide and allele was attaining proper stability by creating a parallelism in

Fig. 4 Graph displaying molecular dynamic simulation of NP-allele complex, resulted in highest peak at 6.8 Å

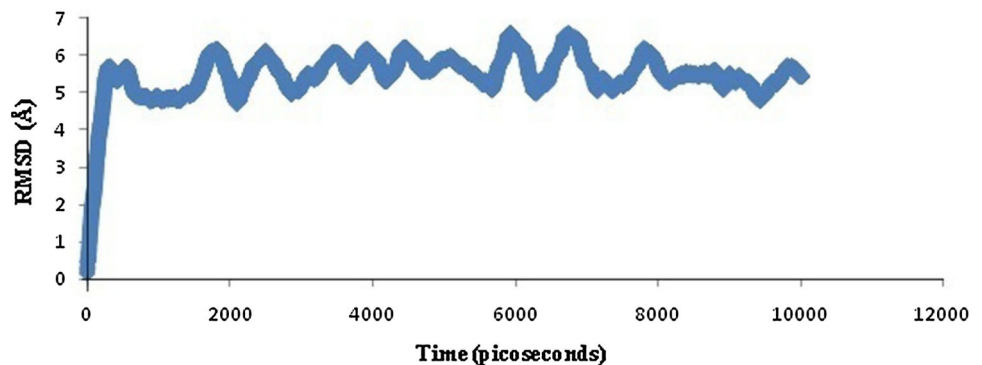
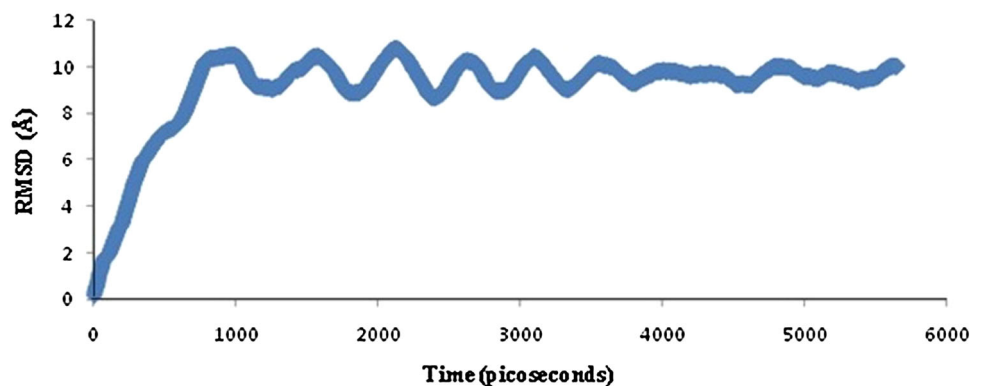


Fig. 5 Graph displaying molecular dynamic simulation of GP-allele complex, resulted in highest peak at 11 Å



RMSD over a time window of 10,500 picoseconds, with the highest peak at 6.8 Å.

For glycoprotein, VMILGLLSS peptide was predicted with highest score of 73.49 % and activating same allele as by the nucleoprotein. The homology modeling of above-mentioned allele was already completed satisfactorily, so rigid docking approach was initiated by Hex 8.0 on a correlation type of shape in 3-D FFT mode. E-total value of the predicted peptide was −301.04. Autodock Vina at default parameters resulted into binding affinity of −6.0 kcal/mol along with two H-bonds, first with CYS44:HN1 and second was predicted with PHE46:HN1. The molecular dynamics simulation here also showed proper stability over a time window of 6000 picoseconds in all the frames, with the highest peak at 11 Å.

It is observed that peptides from NP as well as GP activating the same allele were providing a solid base to the prediction for vaccine design. The mentioned peptides can be either isolated or formulated for further in vitro and in vivo testings.

Conclusion

The conclusion drawn from present study is that the two nonameric epitopes VKMMIVLNL and VMILGLLSS of NP and GP, respectively, have considerable binding with DRB1*1104 MHC class II allele and low-energy minimization values providing stability to the peptide–MHC complex. These peptide constructs may further be undergone wet lab studies for the development of targeted vaccine against Toscana virus. This would reduce the time and expenses of in vivo and in vitro experiments.

Acknowledgments This study was conducted in the Department of Biotechnology, G.B. Pant Engineering College (GBPEC), Pauri Garhwal (Uttarakhand). Authors gratefully acknowledge the necessary computational facilities and constant supervision provided by the Department of Biotechnology, GBPEC. Amisha Jain is thankful to AICTE (All India Council for Technical Education, New Delhi, India) for fellowship.

Conflict of interest Authors have no conflict of interest regarding the publication of paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

Bishop DHL, Knipe BN, Chanock RM, Hirsch MS, Melnick JL, Monath TP, Roizman B (1990) *Bunyaviridae* and their replication, Part 1: *Bunyaviridae*. In: Fields virology, 2nd edn, vol 1. Raven Press, New York, pp 1155–1173

- Braito A, Corbisiero R, Corradini S, Fiorentini C, Ciufolini MG (1998) Toscana virus infections of the central nervous system in children: a report of 14 cases. *J Pediatr* 132:144–148
- Cerdino-Tarraga AM, Efstratiou A, Dover LG, Holden MTG, Pallen M (2003) The complete genome sequence and analysis of *Corynebacterium diphtheria* NCTC13129. *Nucleic Acids Res* 31:6516–6523
- Charrel RN, Gallian P, Navarro-Mari JM, Nicoletti L, Papa A, Sanchez-Secco MP, Tenorio A, De Lamballerie X (2005) Emergence of Toscana virus in Europe. *Emerg Infect Dis* 11:1657–1663
- Collao X, Palacios G, Sanbonmatsu-Gómez S, Pérez-Ruiz M, Negredo AI, Navarro-Marí J, Grandadam M, Aransay AM, Lipkin WI, Tenorio A, Sánchez-Secco M (2009) Genetic diversity of Toscana virus. *Emerg Infect Dis* 15:574–577
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of non-bonded atomic interactions. *Protein Sci* 2:1511–1519
- Cusi MG, Savellini G, Zavelli G (2010) Toscana virus epidemiology from Italy to beyond. *Open virol J* 4:22–28
- Dionisio D, Valassina M, Ciufolini MG, Vivarelli A, Esperti F, Cusi MG, Mazzoli F, Lupi C (2001) Encephalitis without meningitis due to sandfly fever virus serotype Toscana. *Clin Infect Dis* 32:1241–1248
- Elliott RM, Schmaljohn CS, Collett MS (1991) *Bunyaviridae* genome structure and gene expression. *Curr Top Microbiol Immunol* 169:91–114
- Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. *J Mol Graphics* 14:33–38
- James CP, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802
- Jiao W, Yin W, Li X, Liu J (2012) Research advancement in RV novel vaccine. *Asian J Anim Vet Adv* 7:1234–1244
- Lovell SC, Davis IW, Arendall WB III, De Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC (2002) Structure validation by Calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50:437–450
- Magurano F, Nicoletti L (1999) Humoral response in Toscana virus acute neurologic disease investigated by viral-protein-specific immunoassays. *Clin Diagn Lab Immunol* 6:55–60
- Martin-Folgar R, Lorenzo G, Boshra H, Iglesias J, Mateos F (2010) Development and characterization of monoclonal antibodies against Rift Valley fever virus nucleocapsid protein generated by DNA immunization. *MAbs* 2:275–284
- Mohabatkar H, Mohammadzadegan R (2007) Computational comparison of T-cell epitopes of GP 120 of Iranian HIV-1 with different subtypes of the virus. *Pak J Biol Sci* 10:4295–4298
- Nicoletti L, Verani P, Cacioli S, Ciufolini MG, Renzi A, Bartolozzi D, Paci P, Leoncini F, Padovani P, Traini E (1991) Central nervous system involvement during infection by Phlebovirus toscana of residents in natural foci in central Italy (1977–1988). *Am J Trop Med Hyg* 45:429–434
- Olal D, Dick A, Woods VL, Liu T, Li S, Devignot S, Weber F, Saphire EO, Daumke O (2014) *Nucleic Acids Res* 42:6025–6037
- Ritchie DW, Kozakov D, Vajda S (2008) Accelerating Protein–Protein docking correlations using a six-dimensional analytic FFT generating function. *Bioinform* 24(17):1865–1873
- Shekhar C, Dev K, Verma SK, Kumar A (2012) In-silico: screening and modeling of CTL binding epitopes of Crimean congo hemorrhagic fever virus. *Trends Bioinform* 5:14–24
- Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J (1999) Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 17:555–561
- Tambunan USF, Parikesit AA (2011) In-silico design of drugs and vaccines for dengue disease. *Trends Bioinform* 4:1–9

- Tang H, Liu XS, Fang YZ, Pan L, Zhang ZW et al (2012) The epitopes of foot and mouth disease. *Asian J Anim Vet Adv* 7:1261–1265
- Trott O, Olson AJ (2010) Autodockvina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem* 31:455–461
- Valassina M, Cuppone AM, Bianchi S, Santini L, Cusi MG (1998) Evidence of Toscana virus variants circulating in Tuscany, Italy, during the summers of 1995–1997. *J Clin Microbiol* 36:2103–2104
- Valassina M, Meacci F, Valensin PE, Cusi MG (2000) Detection of neurotropic viruses circulating in Tuscany: the incisive role of Toscana virus. *J Med Virol* 60:86–90
- Valassina M, Cusi MG, Valensin PE (2003) A Mediterranean arbovirus: the Toscana virus. *J Neurovirol* 9:577–583
- Verma SK, Yadav SSP, Kumar A (2011) In silico T cell antigenic determinants from proteome of H1N2 swine influenza A virus. *Online J Bioinform* 12:371–378
- Viret C, Janeway CA (1999) MHC and T cell development. *Rev Immunogenet* 1:91–104
- Von Hoff DD, Evans DB, Hruban RH (2005) Pancreatic cancer. Jones and Bartlett, Sudbury
- Wallner B, Elofsson A (2003) Can correct protein models be identified. *Protein Sci* 12:1073–1086
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35:407–410