T-CELL EPITOPE PREDICTION BY SEQUENCE-BASED METHODS AND MOLECULAR DOCKING OF PROTEINS FROM *BOOPHILUS MICROPLUS*

M. Atanasova, I. Dimitrov, I. Doytchinova

Department of Chemistry, School of Pharmacy, Medical University of Sofia, Sofia, Bulgaria

Abstract. Ticks are blood-sucking parasites that feed on variety of domestic animals. *Rhipicephalus (Boophilus) microplus* is a hard tick found in tropical and subtropical regions and feeds on cattle. The development of vaccine is very important in order to save the livestock. Threeglycosylphosphatidylinisotol(GPI)-anchored proteins (Contig2828, Contig7420, and CK181624) from the gut of *B. microplus* were selected by VaxiJen server as highly immunogenic. As the proteins will be tested further on BalbC mice (haplotype d), the T-cell epitope predictions were performed on mouse major histocompatibility complex (MHC) class II proteins IAd and IEd. The predictions were made by three sequence-based methods implemented in the servers MHCPred, NetMHCII and RANKPEP and by one structure-based method – molecular docking. Based on the consensus prediction by the sequence- and structure-based methods, the peptides 1-14, 31-47 and 49-67 from Contig2828, peptides 6-32 and 75-101 for Contig7420, peptides 1-26 and 45-67 for CK181624 were selected for further experimental studies.

Keywords: molecular docking, Rhipicephalus (Boophilus) microplus, MHCII

Introduction:

The tick *Rhipicephalus (Boophilus) microplus* is an obligate haematophagousectoparasite that feeds on variety of domestic animals and mainly on cattle. *B. microplus* is distributed worldwide principally in tropical and subtropical regions. This parasite causes blood loss and bodyweight reduction. It transmits numerous pathogen-causing diseases and death [1,2]. The protozoan parasites*Babesiabovis, Babesiabigemina* and the obligate intracellular bacteria *Anaplasmamarginale*, responsible for the infectious diseases babesiosis and anaplasmosis, respectively, are some of most prevalent worldwide disease-causing agents [1]. This impacts significantly the economy in cattle breeding worldwide [2-4]. Tick control is strongly desired in terms of adverse animal health effects and economic losses to cattle producers. One of the most common approaches for treatment ticks is to apply chemical acaricides in infected regions. Recently, it has been reported an increase in tick resistance against all classes acaricides [5-7]. Additionally, the intensive use of acaricides raises concerns of potential presence of chemical residues in food and environment [8,9]. Moreover, the toxic effect of organophosphorous class of acaricides in human is proven [10-12]. An alternative approach for tick control is vaccination. The first anti-tick vaccine was developed and commercialized by Willadsen et al. [13, 14]. It is based on a glycoprotein Bm86, which has been isolated from the gut epithelium of an Australian tick strain

(Yeerongpilly). The effectiveness of the vaccine varies at different regions in the world. Another protein, named Bm95, isolated from Argentinean tick strain A and differentiated from Bm86 in 21 amino acid residues, has been used for preparation of a vaccine [15]. An immune response after vaccination with Bm95-based vaccine has been achieved in some cases, but a variable effect has been found in others, where the infected animals were insensitive to protein Bm86 [16]. In 2002, a synthetic anti-*R.microplus* vaccine, named SBm7462, containing oligopeptides from Bm86, was developed by Patarroyo and co-workers [17]. It was shown that the synthesized epitopes used in SBm746 are highly conserved among the South American tick strains [18, 19]. Although in most cases the Bm86-based vaccines elicit protective immune responses against *R. microplus*, the effectiveness vary considerably due to genetic variability of tick and bovine populations [20-25]. Therefore, the discovery of new tick antigens covering the majority of *R. microplus* populations could improve efficacy and reduce the variation in protection level afforded by the Bm86-based vaccines.

T-cell recognition is a fundamental mechanism underlying the adaptive immune system through the action of which the host identifies and responds to foreign antigens [26]. The T cell is a specialized type of immune cell that mediates cellular immunity. The T-cell receptors (TCRs), found on the surface of T cells, bind major histocompatibility complex (MHC) proteins, which are presented on antigen-presenting cell (APCs) surfaces. T-cell immune responses are driven by recognition of peptide antigens (T-cell epitopes) bound to MHC molecules. MHC proteins are highly polymorphic glycoproteins, which bind small peptide fragments, or epitopes, derived from both pathogen and host protein and present them at the cell surface for interaction by T cells. The combination of two or more preferred peptide residues (anchors) is called a motif. The experimental determination of motifs for every allele is prohibitively expensive in terms of labor, time and resources. The only practical alternative is to make use of a bioinformatics approach.

In this study we applied sequence-based and structure-based approaches to identify probable T-cell epitopes bound to mouse MHC class II proteins IAd and IEd.

Methods

Immunogenicity prediction

VaxiJen is a server for prediction of protective antigens based on the physicochemical properties of proteins [27-29]. The alignment-free approach implemented in the server allows recognition between immunogenic and non-immunogenic proteins from different origin. The protein structure is presented by uniform vectors, generated by auto- and cross-covariance transformation of protein sequence described by three amino acid *z*-scores accounting for hydrophobicity, size and polarity [30].The accuracy of prediction of the implemented models into VaxiJen server varies between 70 and 90% for the internal and external test sets[27].

Sequence-based prediction

The sequence-based predictionfor binding to mouse MHC class II proteins IAd and IEd was performed by the servers MHCPred [31-34], NetMHCII [35] and RANKPEP [36-38]. In MHCPred server, the quantitative prediction of peptide binding affinity to MHC is based on the additive method. A combination of individual amino acid contribution at each position of the peptide and contributions from side-chain – side-chain interactions are taken into account. Several allele specific quantitative structure-activity relationship (QSAR) models, derived by partial least square (PLS), are implemented in the server. The quantitative prediction of peptide binding affinity to MHC in NetMHCII server uses NN-align, which is an artificial neural network-based alignment algorithm [35]. In NetMHCII, the binding core and binding affinity are estimated simultaneously. The method incorporates explicit encoding (composition and length) of peptide flanking residues and deals

with the data redundancy inherent in the peptide data due to multiple examples of identical binding cores. RANKPEP server uses position specific scoring matrices (PSSM) or profiles in prediction of peptide binding potential to MHC [36-38]. PSSM are derived from a set of aligned known binders to a given MHC molecule and contain observed sequence-weighted frequencies of all amino acids from the sequence alignments. Comparison between a peptide, binding to a given MHC and the corresponding PSSM for that MHC molecule gives the predicted binding potential of the peptide to the MHC molecule.

Structure-based prediction.Molecular docking

Structure-based methods derive predictions by analyzing the interactions of one single 3D structure of the complex between peptide and MHC protein. A widely used structure-based method, together with protein threading [39, 40] and homology modeling [41, 42], is molecular docking [43-45]. The molecular docking is a tool for prediction the binding mode(s) of a ligand within a protein with known 3D structure. Docking methodology consists of two steps, (1) generation the poses as sampling positional, conformational, and configurational space of the ligand within the binding site of the receptor and (2) selection the poses using scoring functions for ranking the ligand poses. GOLD v. 5.0.2.was used in the present study. GOLD is based on genetic algorithms for pose generation and several scoring functions are available for ranking the poses [46]. It has proven successful in virtual screening, lead optimization and identification of the correct binding mode of active molecules [47, 48]. GOLD takes into account the flexibility of the ligand as well as the flexibility of the residues within the binding site.

Results and Discussion

The selection of immunogenic proteins from *R. (B.) microplus* was performed by VaxiJen server. The proteins selected as highly immunogenic proteins were Contig2828 (VaxiJen score 1.2176), Contig7420 (VaxiJen score 1.0345) and CK181624 (VaxiJen score 0.9919). Each protein was presented as a set of overlapping nonamer peptides. Thus, for Contig2828 a set of 59 overlapping peptides was generated, for Contig7420 – a set of 93 peptides for CK181624 – a set of 61 peptides.

As the proteins are intended to be testedon BalbC mice (haplotype d), the T-cell epitope predictions were performed on mouse MHC class II proteins IAd and IEd. Prediction by MHCPredidentified high binders (IC_{50}< 50 nM) only to IAd, but not to IEd (Fig. 1).NetMHCII predicts binding peptides only to IAd, but no high binder was identified. RANKPEP, similarly to MHCPred, identified binders only to IAd, but not to IEd (Fig. 2).

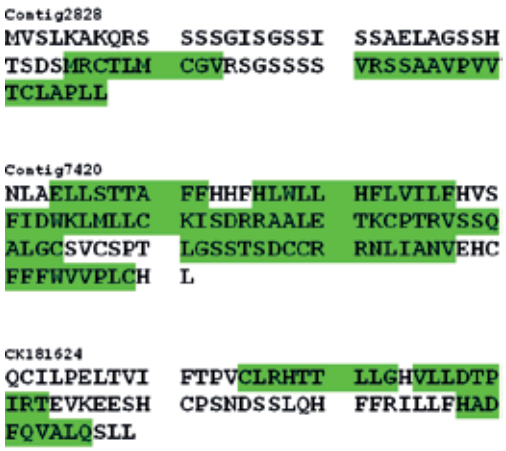

Fig. 1. Binding affinity prediction to IAd by MHCPred. Predicted binders with IC_{50}< 50 nM are highlighted in green.

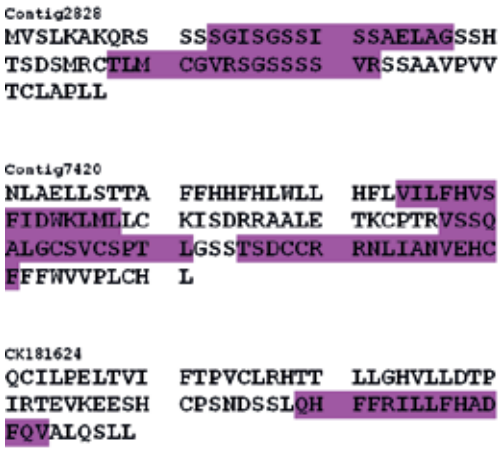

Fig. 2. Binding affinity prediction to IAd by RANKPEP. Predicted binders with binding threshold: 7.10 are highlighted in purple.

For molecular docking simulations, the X-ray structures of peptide – MHC protein complexes $OVA_{323-339}$/IAd (pdb code: 2iad) and HB/IEk (pdb code: 1iea) were selected as input structures. As MHC alleles are expressed by Balb/c strain, haplotype d was used as experimental mouse model. The structure of IEd was modeled from IEk applying single amino acid substitution method. Subsequently, a simulated annealing optimization of each peptide-protein complex was performed by AMBER force field [49]. The binding site was defined from the corresponding X-ray structure (pdb id: 2iad and 1iea) within 6 Å radius (Fig. 3). ChemScore scoring function was used to rank the solutions. Protein and peptide backbone were fixed. All peptides originating from one *B. microplus* proteins were ranked according to their docking-derived binding score and the top two clusters of peptides were selected for further experimental studies.

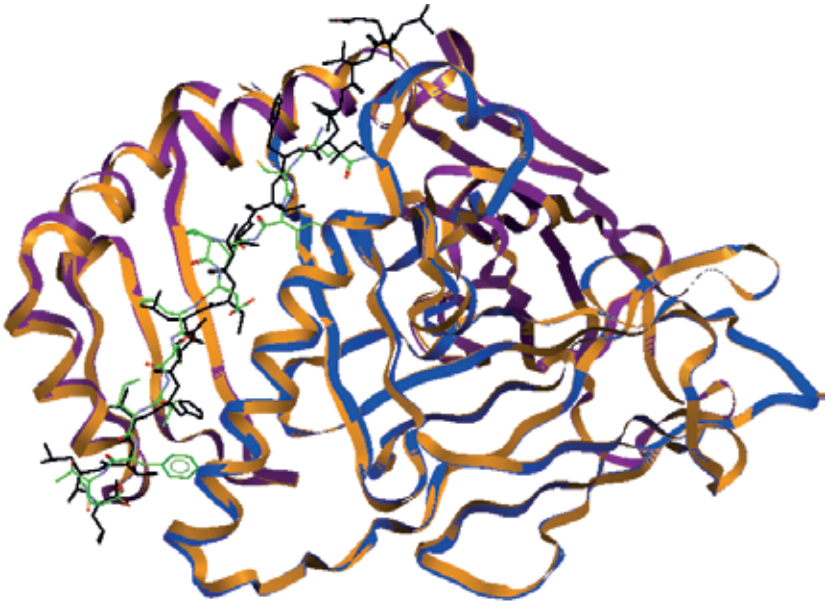

Fig. 3. Two docking solutions of peptides derived from CK181624 in complex with IEd.

The molecular docking was performed at the following settings:

- 6 Å radius of the binding site;
- fixed protein and peptide backbones;
- flexible side chains of the binding site residues and flexible peptide side chains;
- ChemScore function for ranking the poses.

The best binding peptides (ChemScore> 76.40) for IAd and IEd are given in Fig. 4 and 5, respectively.

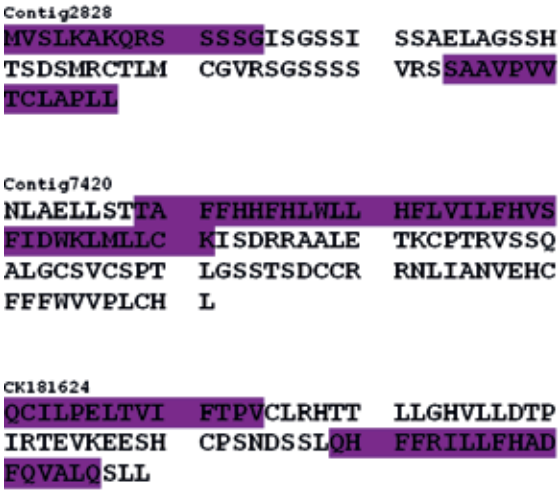

Fig. 4. Binding affinity prediction to IAd by molecular docking. The top two best predicted clusters of binders are highlighted in magenta.

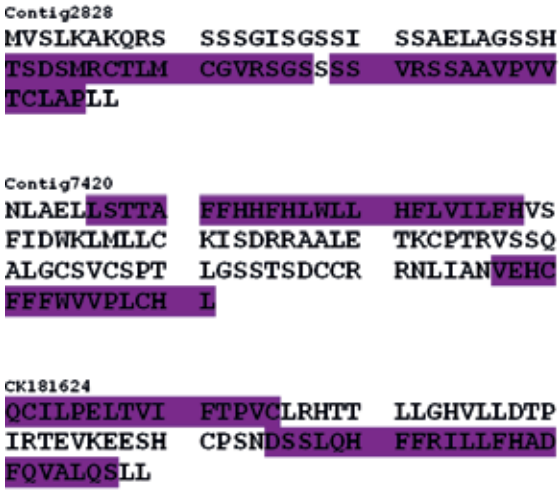

Fig. 5. Binding affinity prediction to IEd by molecular docking. The top two best predicted clusters of binders are highlighted in magenta.

Fig. 6. Peptides selected for further experimental studies based on consensus prediction to IAd and IEd by molecular docking, MHCPred and RANKPEP. Predicted binders are highlighted in blue.

Acknowledgments. The present research was supported by the National Research Fund, Bulgaria (Grant 02-1/2009).

Conclusion

Based on the consensus prediction by molecular docking, MHCPred and RANKPEP, we proposefor further experimental studies peptides 1-14, 31-47 and 49-67 from Contig2828, peptides 6-32 and 75-101 for Contig7420, andpeptides 1-26 and 45-67 for CK181624 (Fig. 6).

References

1. de Castro JJ. Sustainable tick and tickborne disease control in livestock improvement in developing countries. Vet Parasitol 1997; 71 (2-3): 77-97.

2. de la Fuente J, Estrada-Pena A, Venzal JM, Kocan KM, Sonenshine DE. Overview: Ticks as vectors of pathogens that cause disease in humans and animals. Front Biosci 2008; 13: 6938–6946.

3. Jonsson NN. The productivity effects of cattle tick (Boophilus microplus) infestation on cattle, with particular reference to Bos indicus cattle and their crosses. Vet Parasitol 2006; 137: 1–10.

4. Graf JF, Gogolewski R, Leach-Bing N, Sabatini GA, Molento MB, Bordin EL, Arantes GJ. Tick control: an industry point of view. Parasitology 2004; 129 (Suppl): S427-S442.

5. Guerrero FD, Pruett JH, Li AY. Molecular and biochemical diagnosis of esterasemediated pyrethroid resistance in a Mexican strain of Boophilus microplus (Acari: Ixodidae). Exp Appl Acarol 2002; 28 (1-4): 257-264.

6. Li AY, Davey RB, Miller RJ, George JE. Resistance to coumaphos and diazinon in Boophilus microplus (Acari: Ixodidae) and evidence for the involvement of an oxidative detoxification mechanism. J Med Entomol 2003; 40 (4): 482-490.

7. Cutulle C, Jonsson NN, Seddon J. Population structure of Australian isolates of the cattle tick Rhipicephalus (Boophilus) microplus. Vet Parasitol 2009; 161(3-4): 283-91.

8. Henrioud AN. Towards sustainable parasite control practices in livestock production with emphasis in Latin America. Vet Parasitol 2011; 180: 2–11.

9. Playford M. Review of research needs for cattle tick control, Phases I and II. Services SB. Ed.; Meat & Livestock Australia Ltd, Locked Bag 991, North Sydney NSW 2059: 2005.

10. Jamal GA. Neurological syndromes of organophosphorus compounds. Adverse Drug React Toxicol Rev. 1997; 16 (3): 133-70.

11. Singh S, Sharma N. Neurological syndromes following organophosphate poisoning. Neurol India. 2000; 48 (4): 308-313.

12. Costa LG. Current issues in organophosphate toxicology. Clin Chim Acta. 2006; 366 (1-2): 1-13.

13. Willadsen P, Mckenna RV, Riding GA. Isolation from the cattle tick, B. Microplus, of antigenic material capable of eliciting a protective immunological response in the bovine host. Int J Parasitol 1988; 18: 183–189.

14. Willadsen P, Riding GA, Mckenna RV, Kemp DH, Tellam RL, Nielsen JN, Lahnstein J, Cobon GS, Gough JM. Immunologic control of a parasitic arthropod Identification of a protective antigen from Boophilus microplus. J Immunol 1989; 143, 1346–1351.

15. García-García JC, González IL, González DM, Valdés M, Méndez L, Lamberti J, D'Agostino B, Citroni D, Fragoso H, Ortiz M, Rodríguez M, De La Fuente J. Sequence variations in the Boophilus microplus Bm86 locus and implications for immunoprotection in cattle vaccinated with this antigen. Exp Appl Acarol 1999; 23, 883–895.

16. García-García JC, Montero C, Redondo M, Vargas M, Canales M, Boué O, Rodríguez M, Joglar M, Machado H, González IL, Valdéz M, Mendéz L, De La Fuente J. Control of ticks resistant to immunization with Bm86 in cattle vaccinated with the recombinant antigen Bm95 isolated from the cattle tick Boophilus microplus. Vaccine 2000; 18, 2275–2287.

17. Patarroyo JH, Portela RW, De Castro RO, Couto-Pimentel J, Guzman F, Patarroyo ME, Vargas MI, Prates AA, Dias Mendes MA. Immunization of cattle with synthetic peptides derived from the Boophilus microplus gut protein (Bm86). Vet Immunol Immunop 2002; 88, 163–172.

18. Sossai S, Peconick AP, Sales-Junior PA, Marcelino FC, Vargas MI, Neves ES, Patarroyo JH. Polymorphism of the bm86 gene in South American strains of the cattle tick Boophilus microplus. Exp Appl Acarol 2005; 37, 199–214.

19. Peconick AP, Sossai S, Girão FA, Rodrigues MQRB, Souza e Silva CH, Guzman FQ, Patarroyo AMV, Vargas MI, Patarroyo JH. Synthetic vaccine (SBm7462) against the cattle tick Rhipicephalus (Boophilus) microplus: Preservation of immunogenic determinants in different strains from South America. Exp Parasitol 2008; 119, 37–43.

20. Andreotti R, Pedroso MS, Caetano AR, Martins NF. Comparison of predicted binders in Rhipicephalus (Boophilus) microplus intestine protein variants Bm86 Campo Grande strain, Bm86 and Bm95. Rev Bras Parasitol Vet 2008; 17: 93–98.

21. Ben Said M, Galai Y, Canales M, Nijhof AM, Mhadhbi M, Jedidi M, de la Fuente J, Darghouth MA. Hd86, the Bm86 tick protein ortholog in Hyalomma scupense (syn. H. detritum): expression in Pichia pastoris and analysis of nucleotides and amino acids sequences variations prior to vaccination trials. Vet Parasitol 2012; 183: 215–223.

22. Freeman JM, Davey RB, Kappmeyer LS, Kammlah DM, Olafson PU. Bm86 midgut protein sequence variation in South Texas cattle fever ticks. Parasit Vectors 2010; 3: 101-109.

23. Penichet M, Rodriguez M, Castellano O, Mandado S, Rojas Y, Rubiera R, Sánchez P, Lleonart R, de la Fuente J. Detection of Bm86 antigen in different strains of Boophilus microplus and effectiveness of immunization with recombinant Bm86. Parasite Immunol 1994; 16: 493–500.

24. Rodriguez M, Massard CL, da Fonseca AH, Ramos NF, Machado H, Labarta V, de la

Fuente J. Effect of vaccination with a recombinant Bm86 antigen preparation on natural infestations of Boophilus microplus in grazing dairy and beef pure and cross-bred cattle in Brazil. Vaccine 1995; 13: 1804–1808.

25. Rodríguez M, Penichet ML, Mouris AE, Labarta V, Luaces LL, Rubiera R,Cordobés C, Sánchez PA, Ramos E, Soto A, Canales M, Palenzuela D, Triguero A, Lleonart R, Herrera L, de la Fuente J. Control of Boophilus microplus populations in grazing cattle vaccinated with a recombinant Bm86 antigen preparation. Vet Parasitol 1995; 57: 339–349.

26. Flower DR. Epitopes: the immunological quantum. In: Bioinformatics for Vaccinology. Flower DR. Ed.; Wiley–Blackwell: Chichester UK 2008: pp. 94–95.

27. Doytchinova IA, FlowerDR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 2007; 8: 4-11.

28. Doytchinova IA, Flower DR. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 2007; 25: 856-866.

29. Doytchinova IA, Flower DR. Bioinformatic Approach for Identifying Parasite and Fungal Candidate Subunit Vaccines. Open Vaccine J 2008; 1: 22-26.

30. Nyström Å, Andersson PM, Lundstedt T. Multivariate data analysis of topographically modified á-melanotropin analoques using auto and cross auto covariances (ACC).Quant Struct-Act Relat 2000; 19: 264–269.

31. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: A server for quantitative prediction of peptide-MHC binding. Nucleic Acids Res 2003; 31(13): 3621-324.

32. Guan P, Doytchinova IA, Zygouri C, Flower DR. MHCPred: bringing a quantitative dimension to the online prediction of MHC binding. Appl Bioinformatics 2003; 2(1): 63-66.

33.Hattotuwagama CK, Guan P, Doytchinova IA, Zygouri C, Flower DR. J Mol Graph Model 2004; 22(3): 195-207.

34. Guan P, Hattotuwagama CK, Doytchinova IA, Flower DR. MHCPred 2.0: An Updated Quantitative T-Cell Epitope Prediction Server. Appl Bioinformatics 2006; 5(1): 55-61.

35. Nielsen M, Lund O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 2009; 18(10): 296-306.

36. Reche PA, Glutting JP, Reinherz EL. Prediction of MHC Class I Binding Peptides Using Profile Motifs. Hum Immunol 2002; 63: 701-709.

37.Reche PA,Glutting JP, Zhang H, Reinherz EL. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. Immunogenetics 2004; 56: 405-419.

38. Reche PA, Reinherz EL. Prediction of peptide-MHC binding using profiles. Methods Mol Biol 2007; 409: 185-200.

39. Singh SP, Mishra BN. Ranking of binding and nonbinding peptides to MHC class I molecules using inverse folding approach: Implications for vaccine design. Bioinformation 2008; 3(2): 72–82.

40. Knapp B, Omasits U, Frantal S, Schreiner W. A critical cross-validation of high throughput structural binding prediction methods for pMHC.J Comput Aided Mol Des 2009; 23 (5): 301–307.

41. Logean A, Rognan D, Recovery of known T-cell epitopes by computational scanning of a viral genome. J Comput Aided Mol Des 2002; 16 (4): 229–243.

42. Kosmopoulou A, Vlassi M, Stavrakoudis A, Sakarellos C, Sakarellos-Daitsiotis M. T-cell epitopes of the La/SSB autoantigen: Prediction based on the homology modeling of HLA-DQ2/DQ7 with the insulin-B peptide/HLA-DQ8 complex. J Comput Chem 2006; 27(9): 1033–1044.

43. Sauton N, Lagorce D, Villoutreix BO, Miteva MA. MS-DOCK: Accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. BMC Bioinformatics 2008; 9: 184.

44. Tong JC, Tan TW, Ranganathan S.

Modeling the structure of bound peptide ligands to major histocompatibility complex. Protein Sci 2004; 13 (9): 2523–2532.

45. Tong JC, Zhang GL, Tan TW, August JT, Brusic V, Ranganathan S. Prediction of HLA-DQ3.2 Ligands: evidence of multiple registers in class II binding peptides. Bioinformatics 2006; 22: 1232–1238.

46. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997: 267: 727–748.

47. Kellenberger E, Rodrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins: Struct Funct Bioinf, 2004; 5:, 225-242.

48. Perola E, Walters WP, Charifson PS. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. Proteins 200; 56: 235-249.

49. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003; 24(16): 1999-2012.

Corresponding author:
Mariyana Atanasova
Chemistry Department
Faculty of Pharmacy
Medical University of Sofia
2 Dunav st., Sofia 1000, Bulgaria
email: matanasova@ddg-pharmfac.net