TÜBİTAK

Turkish Journal of Biology

http://journals.tubitak.gov.tr/biology/

Research Article

Turk J Biol
(2014) 38: 628-632
© TÜBİTAK
doi:10.3906/biy-1401-21

Identification of HLA-A*0201-restricted CTL epitopes from the receptor-binding domain of MERS-CoV spike protein using a combinatorial in silico approach

Naghmeh POORINMOHAMMAD, Hassan MOHABATKAR*
Department of Biotechnology, Faculty of Advanced Sciences and Technologies, University of Isfahan, Isfahan, Iran

Received: 08.01.2014 • **Accepted:** 23.05.2014 • **Published Online:** 05.09.2014 • **Printed:** 30.09.2014

Abstract: A novel SARS-like illness called Middle East respiratory syndrome coronavirus (MERS-CoV), caused by an emerging coronavirus, has been a recent cause for concern due to its fatality and pandemic potential. Developing a peptide-based vaccine could be helpful in fighting against the virus. Since the experimental procedure is time-consuming and expensive, computational analysis can play an important role in accelerating the process. Therefore, the aim of this study was to computationally identify cytotoxic T-lymphocyte epitopes presented by the human leukocyte antigen (HLA)-A*0201, as this is the most frequent HLA class I allele among Middle Eastern populations. The receptor-binding domain of the spike glycoprotein of MERS-CoV, by which the virus binds to its entry receptor to further infect host cells, is a potential candidate used here for running our in silico epitope identification process. The results include predicted epitopes together with their interaction properties with major histocompatibility complex (MHC) molecules and also the binding behavior of MHC–epitope complexes to human T-cell receptor. Predicted epitopes with the most preferable binding properties are beneficial for vaccine development. Therefore, the huge experimental workload for epitope-based vaccine design will be minimized.

Key words: Middle East respiratory syndrome coronavirus, cytotoxic T-lymphocyte epitopes, HLA-A2, computational prediction, vaccine design

1. Introduction

Middle East respiratory syndrome coronavirus (MERS-CoV), which was called human coronavirus EMC (hCoV-EMC) before (Zaki et al., 2012), was first isolated from the sputum of a patient with acute pneumonia and renal failure in Jeddah, Saudi Arabia, in September 2012. By 17 July 2013, the World Health Organization had confirmed 82 cases of infection, including 45 deaths, mostly from the Arabian Peninsula (Saudi Arabia, Qatar, Jordan, and the United Arab Emirates).

Like severe acute respiratory syndrome coronavirus (SARAS-CoV), MERS-CoV is also a beta-coronavirus that originated from bats and causes pulmonary illness. The spike (S) protein of the virus is responsible for mediating the infection by binding to its entry receptor, which has been identified as CD26 (also known as dipeptidyl peptidase 4, DPP4) (Raj et al., 2013). The MERS-CoV S protein is a membrane glycoprotein with 1353 amino acids, including a domain at the N-terminal region (S1), a membrane-proximal domain (S2), a transmembrane domain, and an intracellular domain. The determinant part by which the virus targets its biologic receptor is located in the S1 domain (Wang et al., 2013), called the receptor-binding domain (RBD), which was determined to be a fragment of 231 amino acids (residues 358 to 588) (Mou et al., 2013).

Obviously, the RBD is a crucial part to be further studied for vaccine development and drug design in order to stimulate the immune system against the virus or to inhibit its interaction with its receptor, respectively. Prior experience in SARS-CoV vaccine development would suggest that vaccine candidates designed on the basis of the RBD subunit of the SARS-CoV S protein (located in the S2 subunit) are more effective in comparison with vaccines based on DNA or an inactivated virus (He et al., 2005; Du et al., 2009; Jiang et al., 2012; Lu et al., 2013). Therefore, in the case of MERS-CoV, searching for similar peptide vaccines seems to be effective. Moreover, in novel strategies for vaccine design process such as the use of epitope-loaded dendritic cells (Hatipoğlu et al., 2013), the exact identification of important antigenic sites of a protein is undoubtedly needed. Experimental screening of specific major histocompatibility complex (MHC)-binding peptides is expensive and time-consuming, as it requires a binding assay for each single peptide; by performing a reliable in silico analysis, we can summarize the steps for a MERS-CoV vaccine development procedure.

* Correspondence: h.mohabatkar@ast.ui.ac.ir

Cytotoxic T lymphocytes (CTL)-mediated cellular immunity is the most important mechanism in controlling viral infections (Wodarz and Nowak, 2000). This type of immunity is mediated by MHC class I CTLs: restricted CD8+ CTLs. MHC class I CTLs have short epitopes (usually about 9 residues) from proteolytically cleaved proteins to CTLs; however, statistics say that only 1 CTL epitope out of 200 nanomer epitopes will be able to bind to a specific MHC (Lundegaard et al., 2006). On the other hand, MHCs are very polymorphic, though only some of their alleles may become frequent in a given population.

What we present in this study is based on 3 steps. First, the CTL nanomer epitopes from the MERS-CoV receptor-binding domain of its S protein were obtained using the best 3 of all epitope prediction servers available on the World Wide Web. Once the specific epitopes were identified, the next step was to examine whether the predicted epitopes could bind efficiently to the most frequent MHC class I allele in Middle Eastern populations, which is HLA-A*0201 (Sheth et al., 1985; Valluei et al., 2005; Ferrante and Gorski, 2007). Finally, epitope-loaded MHC class I molecule interactions with human T-cell receptor (TCR) α and β were modeled. The latter 2 steps were done using an in silico molecular docking technique.

2. Materials and methods

2.1. Retrieval of target amino acid sequence

The amino acid sequence of the S glycoprotein of MERS-CoV was obtained from the NCBI protein database (accession number AGN52936.1). The receptor-binding domain (residues 358–588) was highlighted for further analysis.

2.2. CTL epitope prediction and modeling

CTL-specific epitope prediction was performed using NetCTL 1.2 (Larsen et al., 2007), EpiJen (Doytchinova et al., 2006), and NHLApred (Bhasin and Raghava, 2007). NHLApred and EpiJen expanded the MHC class I binding prediction to different MHC alleles, making it possible to specify a favored MHC allele from the existing list of HLA alleles. Here, the HLA allele selected for the prediction was HLA-A*0201. However, the NetCTL server predicts its MHC class I binding peptides for different supertypes; as a result, NetCTL selected the HLA-A2 supertype.

Having obtained the predicted epitopes from the servers, their 3D structures were modeled using the PEPstr server (Kaur et al., 2007), which models the tertiary structure of peptides (7 to 25 amino acids long) with high accuracy.

2.3. MHC and TCR model retrieval and molecular docking

The structure of a complex between HLA-A*0201 and TCRαβ was fetched from the Protein Database Bank (PDB ID: 1AO7) and used as our reference structure.

A molecular docking technique using ClusPro server (Comeau et al., 2004a, 2004b; Kozakov et al., 2010) was then performed to model the interactions of the 3D-modeled epitopes with MHC and also to model the interactions of epitope–MHC with the TCR molecule. We first docked the peptidic epitopes into the reference MHC molecule. After identifying the most appropriate peptide–MHC complex for each epitope, it was separately docked into TCRα and TCRβ.

3. Results and discussion

A protective high-affinity epitope binds the MHC molecule with enough strength and in a near-native binding orientation. The resulting epitope–MHC complex should bind the TCR molecule in the same manner. Therefore, since not every peptide with high affinity to MHC proteins is considered an epitope, it is better to make sure that the predicted epitopes can efficiently interact with MHC and TCR molecules.

3.1. CTL epitope prediction and modeling

In this study, 3 CTL epitope prediction servers were employed to construct an in silico approach to identify the HLA-A*0201-restricted T-cell epitopes for MERS-CoV. Each server uses a different algorithm, guaranteeing that nearly all possible epitopes would be predicted.

In NetCTL 1.2, the MHC peptide binding and the proteasome cleavage events are predicted using artificial neural networks (ANNs). NetCTL 1.2 also makes predictions about TAP transport efficiency via a weight matrix method proposed by Peters et al. (2003). In NHLApred, the CTL epitopes are predicted with a combinatorial approach consisting of ANNs and quantitative matrices (QMs) (Bhasin and Raghava, 2007), and the EpiJen server implements a QM method to predict CTL epitopes (Doytchinova et al., 2006). Six T-cell epitopes were identified by these prediction algorithms and are shown in Table 1. The number of predicted epitopes may seem low, but it is important to bear in mind that we studied the receptor-binding domain of the S glycoprotein, which is a fragment only 231 amino acids in length.

There is a further more subtle point that we must consider, which is that the epitope should not be hidden by glycosylation. Among all experimentally validated human CTLs, only a small number of them are shown to be glycosylated (whether N- or O-glycosylated) (Szabó et al., 2009); however, there is experimental evidence that shows that glycosylation might deactivate the peptidic T-cell epitope (Lisowska, 2002; Szabó et al., 2009). Therefore, predicted epitopes were checked to see if their residues were glycosylated. Using UniProt, it was shown that the only glycosylated residue in the receptor-binding domain of the MERS-CoV spike is an asparagine, which is not located in any of predicted epitopes (Magrane, 2011).

Table 1. Peptides predicted, via epitope prediction methods, as HLA-A2/A*0201-restricted T-cell epitopes of the receptor-binding domain of MERS-CoV spike protein.

Peptide	Sequence	Prediction methods
P1	LLSGTPPQV	NetCTL, EpiJen
P2	ILDYFSYPL	NetCTL, EpiJen
P3	ILATVPHNL	NetCTL, EpiJen
P4	NLTTITKPL	EpiJen
P5	LQMGFGITV	NetCTL, EpiJen
P6	FSNPTCLIL	NHLApred

Using the PEPster server, tertiary structures of the peptides were modeled. With this server, models can be generated under different modeling conditions, such as vacuum, hydrophobic, or hydrophilic environment simulations. Here, hydrophilic conditions were selected due to the fact that not only are epitopes in a hydrophilic environment while binding to MHC molecules, but there are also water molecules located in the MHC groove to stabilize the peptide–MHC interaction. The server gives one model for each short peptide in PDB format. The modeling algorithm is concerning with β-turns, which are the most important counterpart in short peptides, in addition to regular structures (Kaur et al., 2007).

3.2. Molecular docking

Docking of each predicted epitope into a HLA-A*0201 molecule, obtained from the PDB (PDB ID: 1AO7), was done using ClusPro. It provides different docking options and, if we have the knowledge of what forces dominate in our complex, it is possible to choose from these options. Since the major interaction between an immunogenic peptide and MHC is hydrophobic force (Ferrante and Gorski, 2007), models were created under hydrophobic conditions. The server's output contained at least 6 different results for each peptide. Among these docked models, conformations with high similarity to the native complex of 1AO7 were selected. The considerations included: 1) whether the peptide is docked into the peptide binding groove of MHC in a way such that the whole peptide is completely located into the groove, 2) whether the distance between the peptide and groove is comparable to that of the native structure, and 3) whether the C-terminal residue of the peptide has interaction with the groove. The latter consideration is because in MHC class I restricted CTL epitopes, the anchoring residue is often the residue located in the peptide's C-terminal. After choosing well-oriented peptide–MHC complexes (the closest ones to the native structure 1AO7) from the 6 models, their binding free energy scores (generated by the ClusPro server itself) were compared and the peptide–MHC complex with the lowest score was selected for further analysis. However, it was the orientation factor that played the most important role in our investigations. The binding free energy score only came into consideration when there were negligible differences between well-oriented models. This process was done for each previously predicted epitope.

The selected peptide–MHC complex was docked into TCRα and TCRβ separately. In this case, models created with the "balanced" mode of ClusPro were extracted and analyzed, as there is a combination of chemical forces participating in the interaction between the peptide–MHC and TCR molecule, not just one. Here, native TCR-binding orientation over peptide–MHC class I is a way that TCRβ binds from its N-terminal domain; for TCRα, it binds from where the first N-terminal residue is located in its single domain. The peptide–MHC/TCR complex with the lowest binding free energy was selected among the well-oriented coordinates. This is the same scenario previously done for selecting the best peptide–MHC complex.

Molecular docking results are prepared in Table 2.

Finally, the predicted epitopes' potential to be the true epitopes was evaluated on the basis of their docking

Table 2. Molecular docking results of predicted epitopes.

Peptide	Peptide/MHC		Peptide-MHC / TCRαβ				Real epitope potential
	Orientation score[a]	Binding free energy score	Orientation score[a]		Binding free energy score		
			TCRα	TCRβ	TCRα	TCRβ	
P1	+	–971	+	+	–601	–916	High
P2	+	–1286	+/-	+/-	–679	–971	Medium
P3	+	–1394	+/-	+/-	–635	–762	Medium
P4	+	–1009	-	-	–623	–778	Low
P5	+	–1338	+	+/-	–685	–855	Medium
P6	++	–1389	+/-	+	–636	–910	High

[a] ++: certifiable, +:acceptable, +/-: intermediate, -:false.

properties. Peptides with binding orientations closer to the native structure and lower binding free energy scores are ranked higher in having the potential to be real epitopes.

Among the 6 predicted epitopes, it was concluded that 2 of them (P1 and P6) have the potential to be real epitopes, as their binding orientations in interaction with either TCRαβ or HLA-A*0201 molecules were close to the reference coordinate and their binding free energies were the lowest scores among all binding scores. Among the predicted epitopes, P4 seemed to have low potential to be a real epitope, demonstrating that not all predicted epitopes have the potential to be real ones.

In conclusion, the combination of epitope prediction processes and a knowledge-based molecular docking technique can provide a more reliable identification of MHC class I CTL epitopes compared to merely using automated epitope-predicted servers. Therefore, this study was conducted to predict HLA-A*0201-restricted CTL epitopes of the receptor-binding domain of MERS-CoV S protein in a more reliable way. Results revealed possible CTL epitopes and further computational analysis revealed their real epitope potential.

Since bioinformatic tools are growing quickly in both technique and number, they have become a crucial part of different biological analyses, such as drug design, drug discovery, and the vaccine design process (Öztürk, 2013). Although computational studies are not as accurate as experimental assays, by making use of bioinformatics tools we can improve the output results of experimental assays by adding complementary steps based on our theoretical knowledge.

The combinatorial approach used here can improve the process of epitope prediction compared to the conventional methods. Such in silico approaches can dramatically reduce the number of peptides necessary for further experimental screenings. Therefore, this study was conducted to predict HLA-A*0201-restricted CTL epitopes of the receptor-binding domain of MERS-CoV S protein with a new, improved method. The results can suggest the best epitope candidates to be further tested experimentally in vaccine designing procedures.

Acknowledgment
Support of this study by the University of Isfahan is gratefully acknowledged.

References

Bhasin M, Raghava G (2007). A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes. J Biosciences 32: 31–42.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004a). ClusPro: a fully automated algorithm for protein–protein docking. Nucleic Acids Res 32: W96–W99.

Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004b). ClusPro: an automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20: 45–50.

Doytchinova IA, Guan P, Flower DR (2006). EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics 7: 131.

Du L, He Y, Zhou Y, Liu S, Zheng BJ, Jiang S (2009). The spike protein of SARS-CoV—a target for vaccine and therapeutic development. Nat Rev Microbiol 7: 226–236.

Ferrante A, Gorski J (2007). Cooperativity of hydrophobic anchor interactions: evidence for epitope selection by MHC class II as a folding process. J Immunol 178: 7181–7189.

Hatipoğlu İ, Söğüt İ, Ercan D, Aksu S, Sivas H, Başalp A (2013). Stimulation of dendritic cells with vaccine and vaccine-antibody complex and effect on immune response. Turk J Biol 37: 457–463.

He Y, Lu H, Siddiqui P, Zhou Y, Jiang S (2005). Receptor-binding domain of severe acute respiratory syndrome coronavirus spike protein contains multiple conformation-dependent epitopes that induce highly potent neutralizing antibodies. J Immunol 174: 4908–4915.

Jiang S, Bottazzi ME, Du L, Lustigman S, Tseng CTK, Curti E, Jones K, Zhan B, Hotez PJ (2012). Roadmap to developing a recombinant coronavirus S protein receptor-binding domain vaccine for severe acute respiratory syndrome. Expert Rev Vaccines 11: 1405–1413.

Kaur H, Garg A, Raghava G (2007). PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides. Protein Pept Lett 14: 626–631.

Kozakov D, Brenke R, Comeau SR, Vajda S (2006). PIPER: an FFT-based protein docking program with pairwise potentials. Proteins 65: 392–406.

Kozakov D, Hall DR, Beglov D, Brenke R, Comeau SR, Shen Y, Li K, Zheng J, Vakili P, Paschalidis IC (2010). Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13–19. Proteins 78: 3124–3130.

Larsen MV, Lundegaard C, Lamberth K, Buus S, Lund O, Nielsen M (2007). Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinformatics 8: 424.

Lisowska E (2002). The role of glycosylation in protein antigenic properties. Cell Mol Life Sci 59: 445–455.

Lu G, Hu Y, Wang Q, Qi J, Gao F, Li Y, Zhang Y, Zhang W, Yuan Y, Bao J (2013). Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. Nature 500: 227–231.

Lundegaard C, Nielsen M, Lund O (2006). The validity of predicted T-cell epitopes. Trends Biotechnol 24: 537–538.

Magrane M (2011). UniProt Knowledgebase: a hub of integrated protein data. Database 2011; 2011: bar009.

Mou H, Raj VS, van Kuppeveld FJ, Rottier PJ, Haagmans BL, Bosch BJ (2013). The receptor binding domain of the new MERS coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. J Virol 87: 9379–9383.

Öztürk MA (2013). Computational design of a pentapeptide inhibitor for fibroblast growth factor receptor 3b (FGFR3b). Turk J Biol 37: 675–682.

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.

Raj VS, Mou H, Smits SL, Dekkers DH, Müller MA, Dijkman R, Muth D, Demmers JA, Zaki A, Fouchier RA et al. (2013). Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495: 251–254.

Sheth KV, Edwards JA, Godwin JT (1985). Study of the HLA gene and antigen frequency from a Saudi Arabian hospital. Tissue Antigens 25: 156–162.

Szabó TG, Palotai R, Antal P, Tokatly I, Tóthfalusi L, Lund O, Nagy G, Falus A, Buzás EI (2009). Critical role of glycosylation in determining the length and structure of T cell epitopes. Immunome Res 5: 4.

Valluei V, Mustafa M, Santhosh A, Middleton D, Alvales M, El Haj E, Gumama O, Abdel-Waieth L (2005). Frequencies of HLA-A, HLA-B, HLA-DR, and HLA-DQ phenotypes in the United Arab Emirates population. Tissue Antigens 66: 107–113.

Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X et al. (2013). Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. Cell Res 23: 986–993.

Wodarz D, Nowak MA (2000). Correlates of cytotoxic T–lymphocyte–mediated virus control: implications for immunosuppressive infections and their treatment. Philos T R Soc Lon B 355: 1059–1070.

Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. New Engl J Med 367: 1814–1820.