Meta Gene 25 (2020) 100754

Contents lists available at ScienceDirect

Meta Gene

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

A disclosure of hidden secrets in human cytomegalovirus: An in-silico study of identification of novel genes and their analysis for vaccine development

Shalja Verma[a,b], Anand Kumar Pandey[a,*]

[a] Department of Biotechnology Engineering, Institute of Engineering and Technology, Bundelkhand University, Jhansi 284128, India
[b] School of Biochemical Engineering, Indian Institute of Technology, Hauz Khas, New Delhi 110016, India

ARTICLE INFO

Keywords:
Human cytomegalovirus
Novel gene
Functional annotation
T and B cell epitopes prediction
Vaccine development
Structure prediction

ABSTRACT

Human cytomegalovirus (CMV), has a largest, linear double stranded DNA genome of 236Kbps. It causes adverse congenital infections, resulting in morbid, irreversible CNS associated pathologies. Present study deals with in-silico human CMV genome analysis for novel genes identification and their analysis, for expressed protein structure, function and epitope predictions for vaccine development. Our analysis resulted in identification of 7 novel gene sequences and exposed their functions in essential viral processes. All the identified sequences were having conserved sequences for the different MHC class I and class II alleles. Sequence 2, 4 and 5 can absolutely act as B-cell epitope whereas sequence 7 and 3 contained some B-cell receptor responsive residues. Sequence 7(ATGCAGCACCTAGATATCCAGTTTAACCCCGTATATCACAAGTCTCTGTGTCACTTTTTTTTTGTCTGTTTTTTTTTT CTTCTCCTGGTTCAGACGTTCTCTTCTTCGTCAGAGTCTTTCAAGTGTCGGTAG) was most effective for vaccine development containing B-cell and T-cell receptors epitopes with high 3D-structure prediction score and function annotated as lymphocyte proliferation retardation, thus can also be a highly potent target to develop therapies against the infection.

1. Introduction

Human cytomegalovirus (CMV), a herpes DNA virus, is a major contributor for congenital central nervous system abnormalities via intrauterine infections including pathologies like mental retardation, microcephaly, cerebral palsy and cognitive impairment. (Muller et al., 2010). The other pathological outcomes of CMV congenital infection are hearing and vision loss along with intellectual defects (Pass and Anderson, 2014). It was first discovered in the 1950s and in present scenario every year around 30,000 infants in the US suffer from severe developmental problems associated with CMV congenital infection (Riley, 1997; Patro, 2019). Out of them approximately 10% suffers from severe damage of the brain and other 10% have a congenital subclinical infection and later suffer from epilepsy, mental retardation and seizures. The infection rate of CMV is reported to be higher than Down's syndrome. In adults, CMV infection is quite prevalent in immunocompromised individuals especially those suffering from Acquired Immune Deficiency Syndrome (AIDS). Its prevalence in Asia and Africa is greater than 90% and in USA and Europe it is around 80%. It is mainly transmitted through blood, saliva, urine or semen of infected person (Bialas et al., 2015; Machala et al., 2019). Latent infections, in the spleen, salivary glands, kidney, lungs and bone marrow are quite common in immunocompetent individuals and reactivation at these sites as well as in brain have been studied by many researchers.

Moreover, uncontrolled replication of virus and dissemination in immunocompromised individuals, the elderly and new born babies having congenital infection, often leads to high mortality (La Rosa and Diamond, 2012). Congenital infections within the 16–18 weeks of gestation period, when the neuronal migration and organ development is occurring, results in severe neurological damages. It is well established that cell mediated immune responses of neonates is delayed, compared to adults (Cheeran et al., 2009; La Rosa and Diamond, 2012). Reduced neonate T cells cytolysis activity, low primary antigen responding capacity, low monoclonal stimulatory antibody levels, low toxins stimulating toxic shocks and deficient IL-2 production became evident in vitro studies in neonates against CMV. In addition, response of the mature CD8+ CMV specific T cells resulted in low IFN-γ and high IL-8 levels. Thus, low responsiveness of immune system, neonatal bias for

Abbreviations: CMV, Cytomegalovirus; DNA, Deoxyribo Nucleic Acid; AIDS, Acquired Immune Deficiency Syndrome; ORF, Open Reading Frame; EST, Expressed Sequence Tags; MHC, Major Histocompatibility Complex; HLA, Human Leucocyte Antigen; MUSCA, Multiple Sequence Alignment

* Corresponding author at: Department of Biotechnology Engineering, Institute of Engineering and Technology, Bundelkhand University, Jhansi 284128, Uttar Pradesh, India.
E-mail address: pandayanandkumar@gmail.com (A.K. Pandey).

https://doi.org/10.1016/j.mgene.2020.100754
Received 10 January 2020; Received in revised form 22 May 2020; Accepted 17 June 2020
Available online 23 June 2020

Th2 cell response, where as lack of Th1 cell response against CMV, and constant IL-8 levels that directly enhance viral replication are key characteristic features of neonate immune system and prevent high degree attack by the host against CMV. Such deficient host response causes viremia leading to severe damage in new borns with congenital infection. During pregnancy, antibodies against CMV from seropositive mother prevents congenital foetus infection and also vertical transfer can occur from seropositive mothers (Hassan et al., 2007; Schleiss, 2013a, 2013b; Pass and Anderson, 2014).

In individuals with allogenic transplantations, the immune system is depressed and is unable to inhibit viral replication. Infection or reactivation of CMV is highly frequent in such immunocompromised individuals. Doses of suppressive agents given to avoid T cell mediated transplant rejections, by targeting CD4+ and CD8+ T cells result in highly adverse symptoms during CMV infection as antiviral efficiency of immune system becomes limited. Lytic infection becomes prevalent in such conditions and often result in uncontrolled conditions of viral replication and hence morbidity and mortality (Azevedo et al., 2015; La Rosa and Diamond, 2012).

The complex immunobiology associated to CMV includes a wide range of cells being affected by the virus including epithelial, endothelial, fibroblast, dendritic, leukocytes and smooth muscle cells. In immunocompetent host, primary infection of CMV commence with replication of viral dsDNA in mucosal epithelium, later distributing it to myeloid lineage monocytic cells like CD34+ cells and monocytes finally leading to establishment of latency. This latency stage is achieved after subjection of virus to various high potential immunogenic responses posed by host but can revert back to active productive or lytic stage if immune system is compromised at any stage of life resulting in adverse diseased state (Gredmark-Russ and Söderberg-Nauclér, 2012; Forte et al., 2020).

The immune responses subsequent to primary infection in immunocompetent host against virus bodies and particles are elicited upon encounter with antigen presenting cells (APCs). APCs process these virus particles and activates antigen specific response by immune system (Gredmark-Russ and Söderberg-Nauclér, 2012). Adaptive response in humans is the strongest against CMV and initiates both cellular and humoral immunity. Antibodies are elicited having high specificities to many proteins of CMV including structural proteins like pp150 and pp65, glycoproteins of envelope like gH and gB and multimeric complexes gH/gL and IE1 non-structural proteins thus contribute to circumvent clinical manifestations by limiting viral distribution to different organs of host (McVoy, 2013). Further influential studies have revealed that multimeric complex gH/gL/pUL is essential for infection in both epithelial and endothelial cells whereas gH/gL/gO complex is required for fibroblast cells. Antibodies against gH/gL/pUL complex depicted neutralizing and distribution limiting or inhibiting activities against CMV (Fouts et al., 2012; Xia et al., 2018).

Cellular immunity accounts for the highly striking T-cell response, and frequencies of T-cells specific to CMV are extremely higher compared to other viruses which infect humans except HIV which shows similar levels. However, characteristics associated to such strikingly high and long-lasting response by T cells even when viremia is absent have not been clarified yet but the extended low viral replication may recruit specific CMV T memory cells. T cells recruitment can restrict replication of virus but do not abolish virus from host or prevent its transmission (van den Berg et al., 2019). Broad spectrum T cells against CMV accounts for around 10% of CD8+ and CD4+ memory compartments in blood peripheral stream. Such frequencies of T cell continue to increase lifelong hence a very large response may get displayed ex vivo even for one single epitope which may lead to considerable fraction of T memory cells in seropositive healthy candidate. Past researches, on CMV seropositive candidates having different types of MHC class I molecules, have detected responses against CMV from both CD8+ and CD-4+ T cells to 213 CMV predicted ORFs (Open Reading Frames) out of 13,687 peptides. Significant differences have been observed in cellular responses among individuals, some having T cells specific for only one ORF and others having for around 39 ORFs (Sylwester et al., 2005). Several studies have recognised T cell responses for long term as well as for primary infection against pp65 and IE proteins (Bao et al., 2008; Jackson et al., 2019). Another study reported that pp65 (UL83) and IE (UL123) have been identified in greater than 50% of affected population by CD8+ and CD8+ T cells along with other ORFs of UL122, UL99, UL82, UL55, UL48 and UL32 (Sylwester et al., 2005). Several encoded proteins by CMV disrupts MHC class I and class II pathways of antigen presentation to evade recognition by T cells. Also, IL-10 cytokine homolog, cmvLA, which is expressed in latent stage reduces the expression of MHC I and II and inhibit cytokine production and proliferation of mononuclear cells in blood. The balance between the viral immune system evading mechanisms and the immune responses posed by host leads to achievement of latency (Jenkins et al., 2008; Miller-Kittrell and Sparer, 2009).

Extensive involvement of immune system not only protects from CMV but also poses harmful effects on immunocompetent individuals. Continuous increased expression of IFNγ and other cytokines specific for Th1 cells, by T effector cells specific for CMV stimulate acute state of proinflammation which causes damaging effects on vasculature and finally results in atherosclerosis (Clement and Humphreys, 2019). Further several studies have revealed that chronic immune system activation can potentially lead to immune senescence which finally leads to morbidity and mortality in individuals with age greater than 65. Along with this, infection from CMV, changes the pool composition of T cells by stimulating, a reduction of T naïve cells but a prominent rise in exceedingly differentiated T cells that accumulate in body with the increment in age. Such shift in T cell levels leads to fast and constant decrement in telomere length of lymphocytes and can promote cancers (Jackson et al., 2017; Saunderson and McLellan, 2017). Also, the essential increase of oligoclonal or monoclonal terminally differentiated T cells pool specific for CMV can dominate the whole T cell population, constricting space or damaging the T cells developed against other antigens (La Rosa and Diamond, 2012).

Wide range of dreaded effects of CMV on healthy, unborn or immunocompromised population pose a great challenge to the scientific society and demands exhaustive research for the development of vaccine or therapies against CMV. The capability of CMV to develop resistance against antiviral drugs in turn provokes the need for vaccine development which can target the virus with mechanisms different form antiviral agents (Strasfeld and Chou, 2010). Unfortunately, due to species-restricted viral replication of CMV, no natural model for vaccine evaluation have been developed which pose great difficulty. Despite of that, variety of efforts have been made by numerous researchers to develop vaccine against CMV but most of them are in clinical phase trial stages (Schleiss, 2013a, 2013b).

An effective strategy to develop vaccine should aim to evoke both adaptive and innate immunity at desired time point. For congenital infections a vaccine targeting both cell mediated and humoral immunity would be of great benefit where as for CMV infected transplant patients, vaccine targeting only cell mediated response would be sufficient (Anderholm et al., 2016). Various strategies to develop vaccines against CMV include chimeric attenuated virus vaccine, dense viral bodies, nucleic acid vaccines, peptides, recombinant proteins and viral vector vaccines. These vaccines could be of great benefit for child bearing age women who is seronegative, child bearing age women who is seropositive, receivers of solid organs from CMV seropositive donators and receivers of seropositive stem cells which are hematogenous (Plotkin and Boppana, 2019). The CMV vaccine development dates late back in the 1970s when two attenuated vaccines were developed from two strains Towne and AD169. AD169 strain was soon rejected but the other continued to clinical trials in recipients of solid organ transplants and healthy female and male volunteers. Although better results were reported in transplant receiver's against severe disease forms and rejections but defence against higher virus dose infection was

Table 1
Detected novel sequences their start and end site in genome, sequence length and translated peptide sequence.

S. no.	Novel sequence	Start site	End site	length	Strand	Translated sequence
1	ATGATCAACGTGTCTACGAACGTCA TTGTGAAAGTGACGTCTCAGGCT TTCCGAAACCGCGTCAAGTTCAACG TTGGTTTCGGTTTAGCCTGCGTCA CCGAGGCGGAGGTGGAAATG AGCCGTCCTGTGGGGGAGTGTA CGACCCTGTAG	186,368	186,517	150	+	MINVSTNVIVKVTSQA FRNRVKFNVGFGLAC VTEAEVEMSRPVGECTTL
2	TTACGCTGGGGACAGGGA CGGGGGTTGCGCCGGGGA CGGGGGGTGTGCGGGG ACGGGGGGGTGTGCGGGGACG GGGGGTGTGCGGGGACGGG GGGTGTGCGGGGACGGGGCGT CGCGGGATGGCGGGCTGTTGCGTG CCGGGGACGGGGGACTCTT GCGGCGGGGACGGTGGTGAGGAC GGGGACAGGGGCAT	995	1186	192	−	MPLSPSSPPSPPQESPVPGT QQPAIPRRPVPAHPPS PHTPRPRTPPVPAHPP SPAQPPSL
3	ATGGTCTCCGATGATGATGTTG TTATTGATCGAATCATGGTGC AGAACGGCGACGGAGAGGAG CGTGTCCGCCGCCGGGAAGG TGGTCTCTTTCTCTTTTCT TTTTTCAAGAAATCT TCCATGTGTTTATCGTAG	3400	3534	135	+	MVSDDDVVIDRIMVQ NGDGEERVRRRE GGLFLFSFFKKSSMCLS
4	CTACGAGGAACGGATAACG CGGTGGCGACGGCAGCGGTGGT GGCGCTGGGGGTGGCGGTAG TGGTACTGCTGATGGTAGTC GGGACGGAGGAGAGACGATGC ATACATACACGCGTGCAT	4564	4683	120	−	MHACMYASSLLRPDYHQ QYHYRHPQRHHR CRRHRVIRSS
5	ATGCCCCTGTCCCCGTCCTCA CCACCGTCCCCGCCGCAAGAGTC CCCCGTCCCCGGCACGCAACAG CCCGCCATCCCGCGACGCCC CGTCCCCGCACACCCCCCG TCCCCGCACACCCCCCGTCCCCG CACACCCCCCGTCCCCGCA CACCCCCCGTCCCCGGCGCA ACCCCCGTCCCTG TCCCC AGCGTAA	194,482	194,673	192	+	MPLSPSSPPSPPQES PVPGTQQPAIPRR PVPAHPPSPHTPR PRTPPVPAHPPSP AQPPSLSPA
6	TCACGTGACCATCAGTGCAGG AAGCGAGGCCCGTAGAACCGCCC AAGAGGCGGTGCCAGATGCC AACGTCATAATCACAAGGTGAT TTGTTACGTCACGCGTGTGC GCACACGCGCACGCGCGCAC ACGCGCGCGGTAGAATACAG CGATCCCTAGTGAAGCCAC ACCCAT	92,935	93,105	171	−	MGVASLGIAVFYRA RVRACACAHTRD VTNHLVIMTLASG TASWAVLRASLPALMVT
7	ATGCAGCACCTAGATATCCA GTTTAACCCCGTATATCACAA GTCTCTGTGTCACTTTTTTTTTG TCTGTTTTTTTTTTCTTCTCCT GGTTCAGACGTTCTCTTC TTCGTCAGAGTCTTTC AAGTGTCGGTAG	81,977	82,108	132	+	MQHLDIQFNPVYHK SLCHFFFVCFFFLLLV QTFSSSSESFKCR

insignificant (Plotkin et al., 1975; Neff et al., 1979; Plotkin and Boppana, 2019). The next approach used CMV surface glycoprotein B in conjugation with the oil-in-water adjuvant MF-59. Significantly good outcomes were noticed upon three viral exposures during a period of 6 months in humans, which exhibited high neutralizing antibody levels. But antibody levels reduced quickly with this adjuvant and increased and became consistent when booster dose combined with AS01 adjuvant was given subsequently, this was due to stimulation of tool like receptors 4. As glycoprotein B is a trimeric protein thus there is possibility that other highly immunogenic form may exist which restricted the use of this vaccine (Schleiss, 2018; Anderholm et al., 2016). The development of pentameric vaccine which consisted a complex of glycoprotein L, H and protein expressed by gene UL128, UL130, and UL131 gave better results in eliciting antibodies compared to glycoprotein B vaccine. This was also effective in providing protection to foetus from an infected mother who was vaccinated during pregnancy (Lehmann et al., 2019. Further, attempt to enhance attenuated Towne virus vaccine immunogenicity by making Toledo wild CMV recombinant was made. Four new recombinants entered the clinical trial stage and one of them showed high immunogenicity. Another approach for the development of vaccine is replication defective virus in which 2 proteins were chemically destabilized but they get stabilized by Shld 1. In human host this defective strain does not cause infection but expression of proteins causing immunogenicity remains consistent. This candidate vaccine gave improved neutralizing antibody yield in phase I trial, thus showed significant results (Plotkin and Boppana, 2019).

Table 2
Physical and chemical parameters predicted by Protpram tool.

Seq No.	No. of amino acids	Molecular wt.	pI	No. of negatively charged residues	No. of positively charged residues	Estimated half-life in human reticulocytes (in vitro) (hr)	Instability index (in test tube)	Aliphatic index	Grand average of hydropathicity
1	51	5577.45	6.10	5	5	30	37.37 (stable)	87.65	0.222
2	60	6233.12	12	1	4	30	154.97 (unstable)	40.67	−1.003
3	44	5070.81	5.09	8	7	30	44.52 (unstable)	77.27	−0.230
4	39	4971.67	10.85	1	8	30	135.25 (unstable)	42.56	−1.418
5	63	6488.39	12	1	4	30	158.10 (unstable)	40.32	−0.965
6	56	5902.03	10.66	1	5	30	23.83 (stable)	111.61	0.838
7	43	5210.09	7.78	2	3	30	43.63 (Unstable)	74.65	0.451

Many vector-based vaccines which carry specific CMV genes have been developed. Proteins like phosphoprotein 65 of tegument and glycoprotein B were leading contributors for development of vector-based vaccines and gave significant results in generating immunogenicity along with safety. Moreover, inactivated virus, DNA, peptide and mRNA vaccines gave significant outcome in various clinical studies. Although the efforts were significant but still commercialization of these vaccines is in half way, further a particular formulation which can be effective for congenital infection, immunosuppressed as well as transplant patients is yet to be discovered (Sung and Schleiss, 2010).

Genome sequencing of CMV provided better understanding of pathology for the development of potential treatments. The whole genetic makeup of a wild type human CMV contained 235,645 bp genome from a lowly passaged strain and 235,476 bp with 16 ORF from JHC strain procured from a Korean bone marrow transplant patient, the later study of 11 HCMV strains further depicted inter strain gentic differences found in the considered virus (Dolan et al., 2004; Jung et al., 2011). Such sequencing studies lead the pathway for exploring the actively participating genes in the pathology of infection. In-silico gene and protein prediction provided great assistance in predicting and annotating the structurally and functionally significant genes. Novotny et al., in their in-silico structural and functional study of human CMV genome analysed 213 proteins with unknown functions for their structure and functional attributes by using Pro Ceryon program (Novotny et al., 2001). Rigoutsos et al., in their pattern based in-silico human CMV genome analysis, analysed 200 ORFs having the potential to code for proteins. Bio-Dictionary based annotation method and algorithm based on sequence patterns for multiple sequence alignment (MUSCA) were applied and combinations of amino acids which were unique to herpes group of viruses or to human CMV were evaluated. Results revealed that a large number of ORFs were membrane proteins and some proteins which were uncharacterized were found to be homologs of G protein receptors (Rigoutsos et al., 2003). Today many other in-silico gene prediction online tools like Augustus, Gene Mark, ChemGenome, FgeneSV and EasyGene are frequently being used for the analysis of potential open reading frames from the genome in several studies on different organisms (Stanke and Morgenstern, 2005; Besemer and Borodovsky, 2005; Díaz-Cruz et al., 2017; Nielsen and Krogh, 2005; Singhal et al., 2008).

Discovery of novel viral ORFs coding for proteins responsible for either evading immune response by supressing it or eliciting immune response can be of great importance in development of treatments against CMV. Further, detection of conserved regions in such proteins which can act as epitopes for B cell receptors, dendritic cells or APCs to develop B or T cell mediated immune response can lead to high potency vaccine to specifically prepare host's immune system to target the virus and combat the adverse effects of the disease upon infection.

Hence, this study focuses on in-silico identification of novel genes and structure prediction of proteins encoded by them, their functional attributes and relevance in causing disease, as well as the identification of B-cell and T cell epitopes for the development of highly promising vaccine candidate against the human CMV.

2. Method and tools

The complete genome sequence of Human cytomegalovirus (NC_006273.2 Human herpesvirus 5 strain Merlin) was procured from the GeneBank database of National Centre for Biotechnology

Table 3
B-cell epitope prediction by Disco Tope 2.0 server.

S. no.	Amino acid sequence	Sequence of B-cell epitope highlighted in red
1	MINVSTNVIVKVTSQA FRNRVKFNVGFGLAC VTEAEVEMSRPVGECTTL	Null
2	MPLSPSSPPSPPQESPVPGTQ QPAIPRRPVPAHPPSPHTPRP RTPPVPAHPPSPAQPPSL	MPLSPSSPPSPPQESPVPGTQQPAIPRRPV PAHPPSPHTPRPRTPPVPAHPPSPAQPPSL
3	MVSDDDVVIDRIMVQNGD GEERVRRREGGLFLFSF FKKSSMCLS	MVSDDDVVIDRIMVQNGDGEERVRRR EGGLFLFSFFKKSSMCLS
4	MHACMYASSLLRPDYHQQY HYRHPQRHHRCRRHRVIRSS	MHACMYASSLLRPDYHQQYHYRHPQ RHHRCRRHRVIRSS
5	MPLSPSSPPSPPQESPVPG TQQPAIPRRPVPAHPPSPHTP RPRTPPVPAHPPSPAQPPSLSPA	MPLSPSSPPSPPQESPVPGTQQPAIPRRPV PAHPPSPHTPRPRTPPVPAHPPSPAQPPSL SPA
6	MGVASLGIAVFYRARVRACA CAHTRDVTNHLVIMTLASGT ASWAVLRASLPALMVT	Null
7	MQHLDIQFNPVYHKSLCHFFF VCFFFLLLVQTFSSSSESFKCR	MQHLDIQFNPVYHKSLCHFFFVCF FFLLLVQTFSSSSESFKCR

Information (NCBI). The possible ORFs were identified by using web-servers for viral gene prediction fgenesV (http://www.softberry.com) and GeneMarkS (http://exon.gatech.edu/GeneMark/genemarks.cgi) (Besemer and Borodovsky, 2005). The uncommon ORFs in results of above two tools were found by using VENNY 2.1 (Oliveros, 2007–2015). These uncommon ORFs were then mapped against EST dataset by using BLAST N and the sequences having no significant similarity with the sequences of EST dataset were considered for further analysis. Then these sequences were mapped with non-redundant protein sequence dataset by using BLAST X and the sequences which were having insignificant similarity with this respective dataset were finally considered as novel gene sequences (Adams et al., 1991). These gene sequences were translated into protein primary sequences by using EMBOSS Transeq (Rice et al., 2000). Then the prediction of B-cell and T-cell epitopes in each sequence was done by using Disco-Tope 2.0 (Kringelum et al., 2012) and Propred I (MHC I) and II (MHC II) (Singh and Raghava, 2001) webservers respectively. Also, the sequences were analysed for the presence of protective antigen by using VaxiJen v2.0 server to test their effectiveness for vaccine development (Doytchinova and Flower, 2007). Further, to analyse the affinity of peptide epitopes for MHC I and MHC II alleles (present in potential sequences which resulted from VAxiJen analysis), IC50 values were calculated by using TepiTool developed by IEDB analysis resource by utilizing NetMHCpan method (Paul et al., 2016; Karosiene et al., 2013; Nielsen et al., 2008). The upper limit for IC50 was set to 100 nM to analyse most effective peptide epitopes having very high affinity for respective receptors. The secondary structures were then predicted from the primary sequences by using PSIPRED webserver (Jones, 1999) and the physio-chemical properties were analysed by using ProtParam (Gasteiger et al., 2005). Also, the predictions of tertiary structures and functions were done by using I-TASSER webserver (Roy et al., 2010; Yang et al., 2015) for each novel gene sequence to detect their potentiality to be employed as targets for treatment. Refinement of predicted tertiary structures was done by using 3D refine online web server (Bhattacharya et al., 2016). Ramachandran plots were generated to depict residues lying in favourable regions by using Discovery studio 2019 software. The tertiary structures of proteins were analysed by estimating the *Z*-score by using ProsaWeb webserver to look for the predicted structures similarity with experimentally analysed structures in the database (Wiederstein and Sippl, 2007).

3. Results and discussion

Human cytomegalovirus is responsible for intrauterine infections to infants causing congenital central nervous system disorders (Muller et al., 2010). Presence of unknown novel gene sequences in its genome opens new scope for understanding the viral disease pathology and development of new treatments as these sequences will be unique to the organism and will enhance specificity of treatment if considered. Though viral genome is known to mutate at a significant rate, identification of specific conserved sequences can benefit in development of effective vaccines against them (Sanjuán and Domingo-Calap, 2016). Epitopic regions in novel sequences which can be expressed by MHC class I, II and B cell receptor can be of great significance in this regard. Also, the information about the 3D structure and function of the protein encoded by these sequences can reveal new targets for therapies.

Present study deals with in-silico analysis of Human herpesvirus 5 strain Merlin virus genome for the identification of novel gene sequences. Our results revealed 7 novel sequences out of which two sequences (sequence 2 present on negative strand and sequence 5 present on positive strand) code for same protein sequence thus increasing the significance of the sequence. The identified sequences and their start and end positions along with translated amino acid sequence are mentioned in Table 1. Various studies have proposed different pipelines and protocols for identification of novel gene sequences. Klasberg et al., in their review presented various protocols relevant for novel gene prediction. In-silico study on *Bombyx mori* identified novel gene responsible for expressing chitinase like protein, another study identified 28 novel genes expressed in spermatogenic cells having primary role in development of male germ cell using bioinformatic tools (Klasberg et al., 2016; Pan et al., 2012; Hong et al., 2004).

As the physiochemical properties are essential to find the stability of

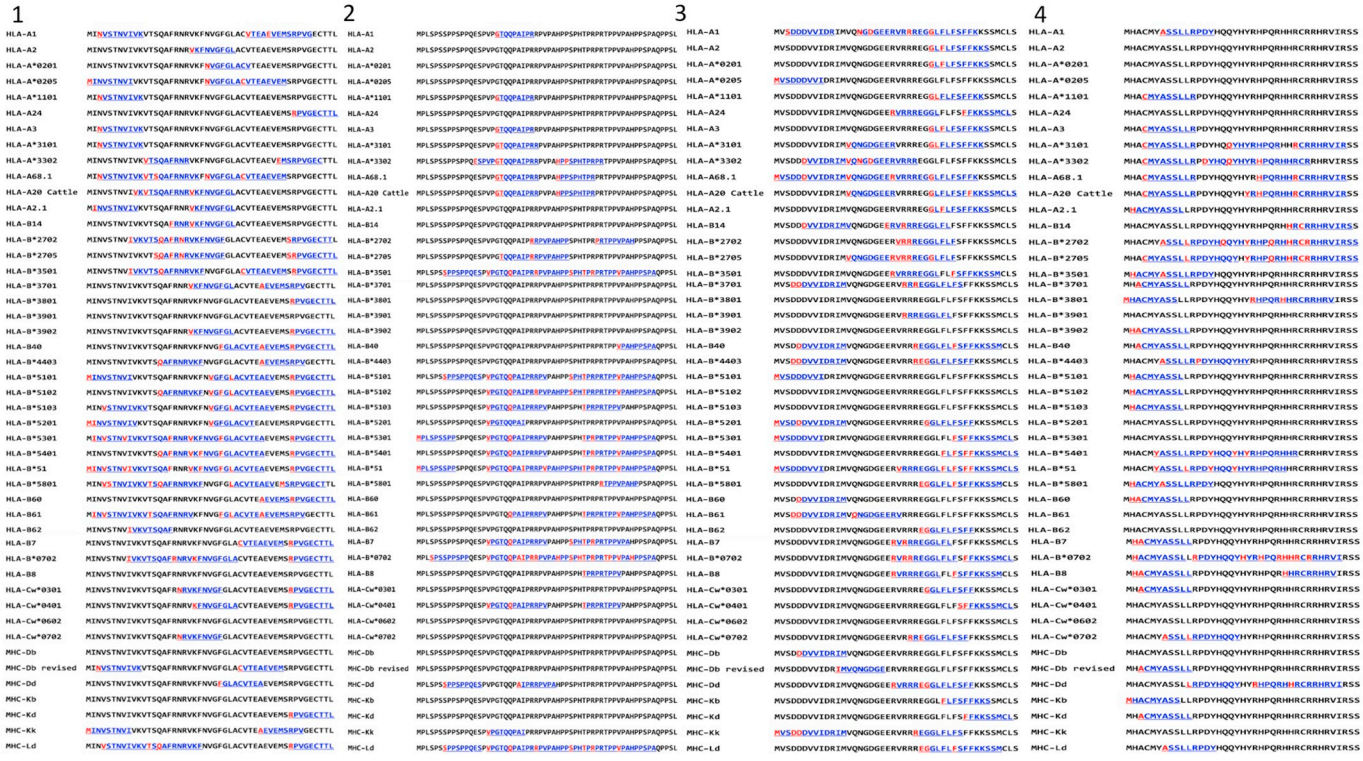

Fig. 1. MHC class I epitope prediction by Propred I for a) 1–4 and b) 5–7 novel sequence peptides.

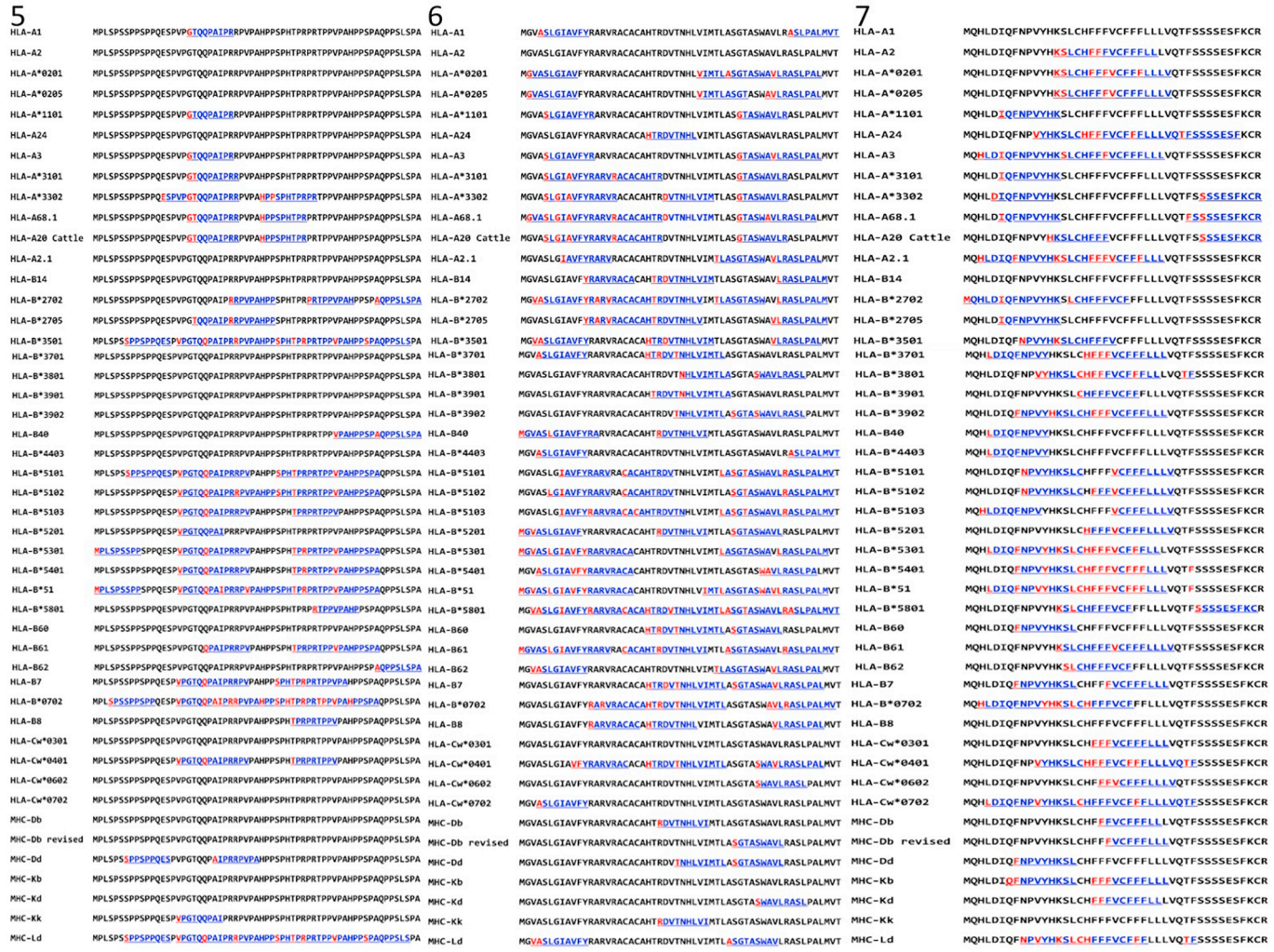

Fig. 1. (*continued*)

expressed protein, following the prediction, we then analysed the resulting novel genes for physiochemical properties by ProtParam tool and the results are presented in Table 2. Thereafter analysis of the concerned primary protein sequences for the presence of B-cell epitopes as well as MHC class I and II epitopes, was performed which can be beneficial for the development of highly specific peptide vaccine against the disease. Our result demonstrated that all the residues of sequence 2, 4 and 5 possesses potential to act as B-cell epitopes proving high significance of the sequences for use in vaccine to evoke humoral immunity, along with this sequence 7 and 3 also had some residue long stretches which can be employed for this purpose (Table 3). In MHC class I and MHC class II epitope prediction we found that all the sequences had epitope sequences corresponding to most of the allele types of MHC I. Sequence 1, 6 and 7 contained epitopes for majority of alleles hence have high probability to be identified by the immune system cells to evoke cell mediate response and memory thus can be highly effective if used as peptide vaccines (Fig. 1). In case of MHC class II, sequences 1, 3, 6 and 7 showed specific region which possess potential to act as effective epitope corresponding to different allelic forms of MHC class II but sequences 2, 4 and 5 had small sequences corresponding to only few of the allelic forms (Fig. 2). Thus sequence 6 and 7 having epitope sequence for both MHC class I and II can be effectively used to evoke strong cell mediated immunity. We further analysed our sequences for their ability to act as protective antigenic sequence capable of eliciting antigenic response when used as potential vaccine. Our results showed that sequence 1, 6 and 7 had great potential to be used for the purpose where sequence 7 showed the highest score of 0.9132 and have potential to act as highly protective antigen among all considered sequences proving its high immunogenicity thus can evoke significant immune response when used as epitope for B and T cell receptors to provide protection against the infection (Table 4). Moreover, the IC50 values of sequences 1,6 and 7 were evaluated for the most common MHC I and MHC II alleles to predict the affinity of considered peptide sequences for MHC receptors and to get an estimate of the efficiency of present epitopes to elicit T cell response. The resulting 9 mers and 15 mers epitopic sequences for MHC I and MHC II respectively gave significantly low IC50 values of less than 100 nM thus proving high affinity of epitopes for T cell receptors and high potential of sequences to be used as vaccine candidates (Table 5) (Paul et al., 2016).

Many studies have reported role of proteins in eliciting immune responses, monomeric glycoprotein B in case of CMV showed 50% efficiency, in avoiding infection in reproductive age women who was seronegative and in lowering viremia in recipients of transplants of solid organs. Further, the discovery of novel recombinant glycoprotein B which was trimeric, showed 11 times higher antibody titers compared to monomeric protein titers capable of neutralizing the infection in mice model. Furthermore, 50- and 20-times higher complement independent and dependent neutralizing titers were reported respectively in fibroblast cells (Cui et al., 2018). A recent study by Gu et al., on

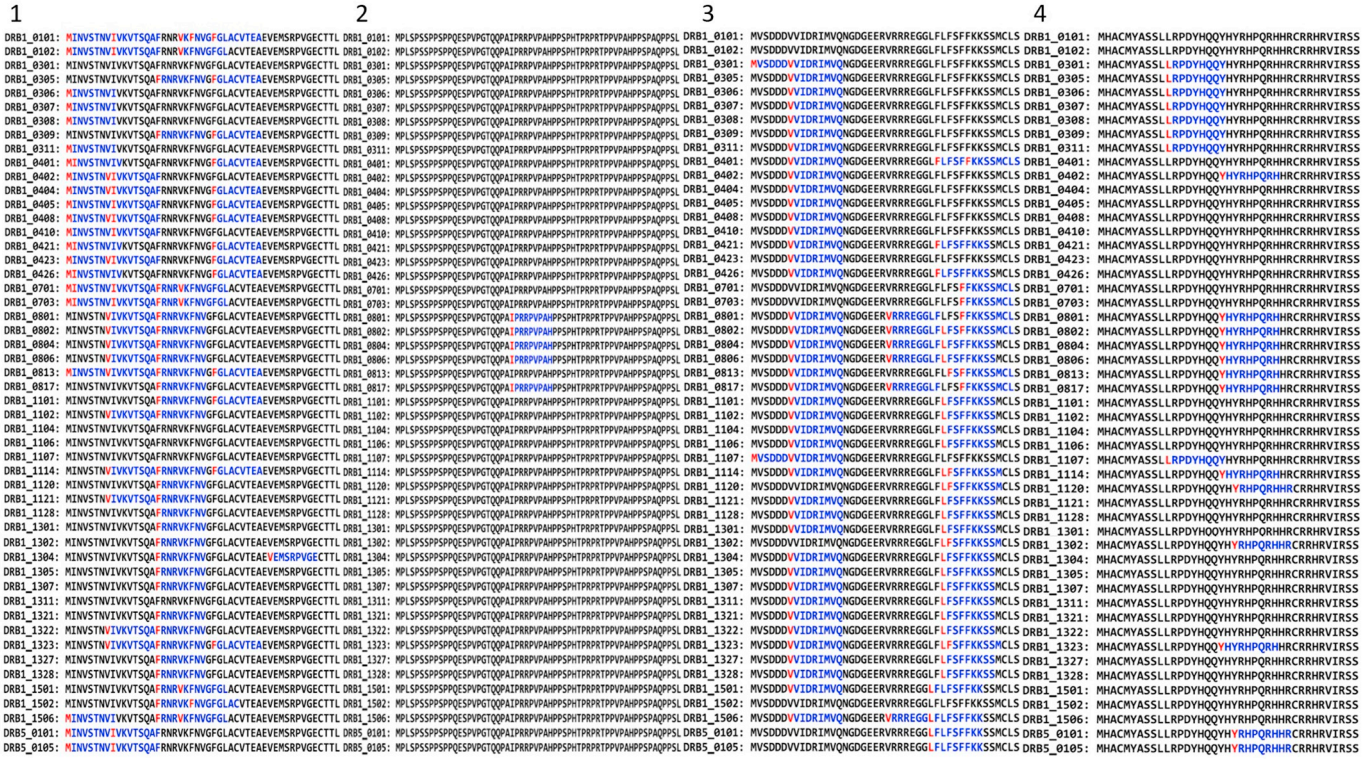

Fig. 2. MHC class II epitope prediction by Propred II for a) 1–4 and b) 5–7 novel sequence peptides.

infection of *Trichinella spiralis*, proved that peptide vaccine containing multiple epitopes both of T and B cells showed significantly high titers of IgG2a and IgG1 along with production of mixed cytokines secreted by immunized mice splenocytes (Gu et al., 2013). Another peptide vaccine developed by Herrera-Rodriguez et al., against influenza virus contained mixture of epitopes for both T and B cells induced both cell mediated and humoral immune reactions and displayed highly significant results against severe infection in mice model (Herrera-Rodriguez et al., 2018). Firbas et al., 2006, reported controlled human trial on 128 individuals of a novel peptide-based vaccine, against hepatitis C virus infection, called IC41 which consists of 5 peptides having T cell epitopes and proved the efficacy of peptide vaccine to elicit T cell response and increased levels of IFN gamma (Firbas et al., 2006).

Many studies have used in-silico approaches for epitope predictions for development of vaccines. In-silico study on Nipah Virus predicted MKLQFSLGS as effective epitope present in matrix protein and showed high binding affinity of this epitope with MHC class II allele (DRBI*0421) (Kamthania and Sharma, 2016). Sequence and structure based in-silico study of flagellin protein of *Burkholderia pseudomallei* predicted 3 peptides which can act as both B cell as well as T cell epitope and immunoreactivity analysis confirmed that all the three peptides were reactive against IgG antibodies and evoked production of cytokines. Further antibodies generated by 2 peptides also enhanced the bactericidal activities of purified neutrophils from human thus proving high efficiency of in-silico analysis to be employed for vaccine development (Nithichanon et al., 2015). Epitope prediction study on Oropouche virus identified, 18 high antigenic and immunogenic epitopes which can bind to CD8 + Tcells and 5 non-toxic and conserved B-cell epitopes, out of the 18 epitopes of T-cell 5 were found to be highly conserved. Another study on Hepatitis B virus polymerase protein, predicted 4 epitopes having high affinity for HLA-A0201 by molecular docking and 2 regions having very high probability to act as B cell epitope were detected (Adhikari et al., 2018; Zheng et al., 2017). Shey et al., in their study of in-silico designing of multi-epitope vaccine for onchocerciasis and related diseases identified novel vaccine candidates having capability to work as both T-cell and B-cell epitopes. Immune simulation study reported significant increase in IgG, T-cytotoxic, T-helper, IL-2 and INF-γ levels hence validating the in-silico analysis (Shey et al., 2019).

To get deep insight about the role of the novel sequences identified in the present study, we predicted the secondary and tertiary structures to reach the functional significance of these sequences inside the virus. The secondary structures of these sequences were predicted by using PSIPRED as shown in Fig. 3. Along with the secondary structure, the confidence of prediction at each residue is displayed which conveys the quality of prediction. The tertiary structure predictions and functional annotations done by using I-TASSER webserver were based on template homology and the similarity with the analog proteins in the database respectively (Fig. 4). The quality of prediction was evaluated by C-score, which signifies effective template threading. The TM-score for alignment of predicted structure with the analog structure display structural alignment quality and denotes the probability of the modelled structure to have similar function as that of the analog. The results and evaluated parameters for tertiary structure and function predictions are present in Supplementary Table 1. Our result showed good alignment of all the sequences with their template with normalized *Z*-score of greater than 1 except the 6th sequence whose Z-score was 0.75. the highest scoring was achieved by the 7th sequence signifying high confidence of structure prediction. Further validation by ProSAweb tool gave significant scores for the predicted structures which compares with the structures generated through experimental techniques like X-ray diffiraction and NMR present in databases (Table 6). Functional annotation of these predicted structures based on structural analogy is of great importance as it will provide the role of these viral proteins in relation to disease pathology and will also help to reveal new targets for treatment or identification of effective peptide vaccine candidates. The more the functional effectiveness of the vaccine peptide sequence in the disease pathology the more effective will be the immune response developed against it when used in vaccine for the prevention of disease (Li et al., 2014). According to gene ontology terms annotated to our

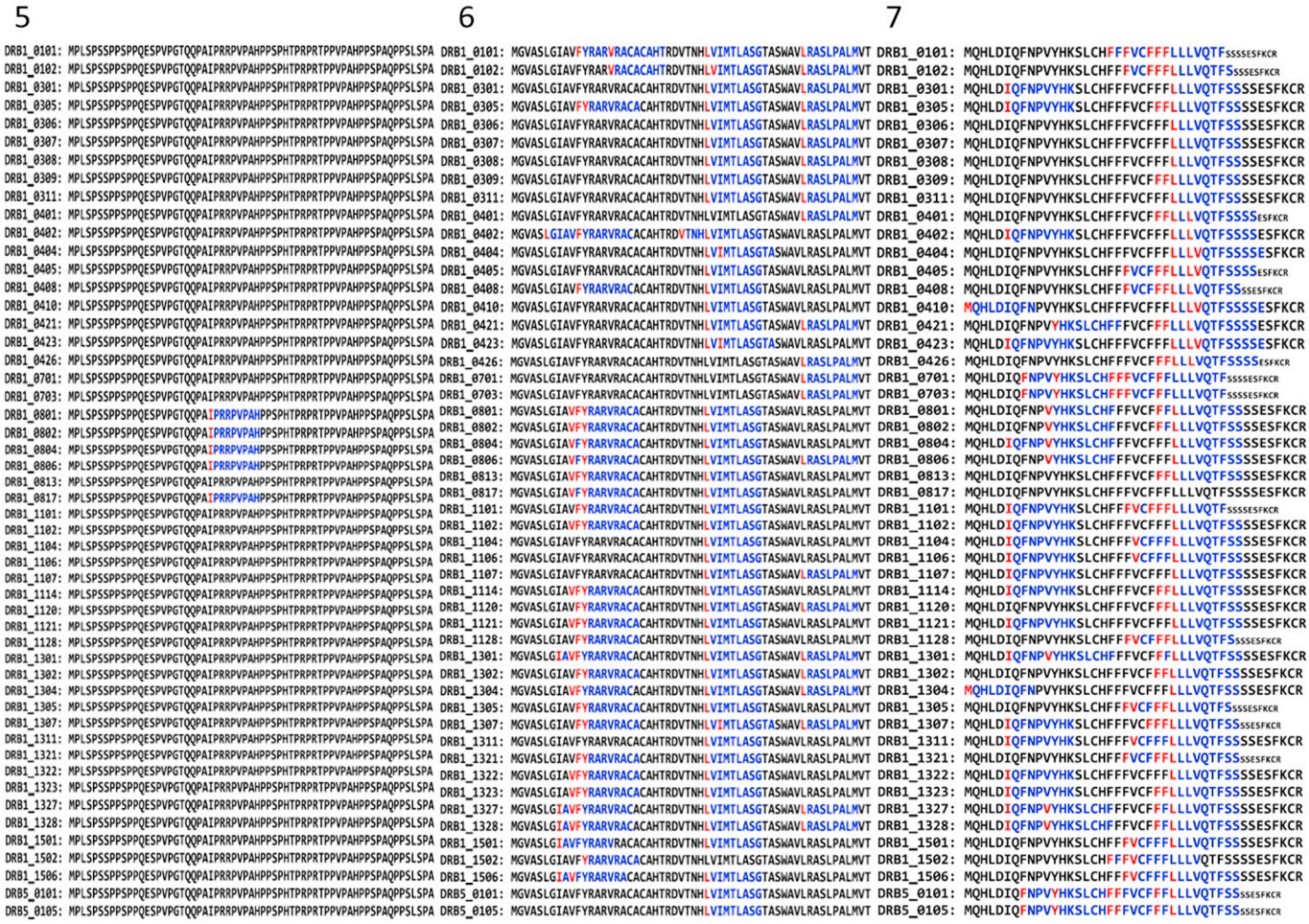

Fig. 2. (*continued*)

sequences, sequence 1 possess ion binding efficiency at molecular level and have biological cell adhesion capability including intracellular attachment in the regions between the membranes which can impart human cell adhesion efficiency to the virus to transfer it genetic material into it. It can also work for cellular communication and can take part in providing interaction with the host cell environment after getting expressed. Sequence 2 and 5 being different in genetic sequence and location on strand encodes for same amino acid sequence. Our analysis showed very low confidence score of structure prediction as well as for functional annotation but further analysis for such novel, difficult to predict sequence can reveal new facts in concerns to disease pathology. Our functional annotation result for sequence 2 and 5 showed that they can have significant role in regulation of DNA transcription which can effectively be associated with the multiplication of viruses in host body leading to infection. Gene ontology terms related to sequence 3 revealed their role as hydro-lyase enzyme for the utilization of carbon, hence the protein encoded can have potential role in providing raw material for the growth and multiplication of virus particles. The ontology terms associated with sequence 4 revealed its ability to bind ATP and DNA and thus function as DNA-dependant ATPase in DNA repair. Its cellular localization predicted was in cytoplasm which prove high potential of this sequence to be a leading target to treat the disease or for vaccine development. The ontology related to the protein structure of sequence 6 proved its affinity to bind to adenosine ribonucleotide and role in nucleic acid metabolism thus displaying its part in viral nucleic acid metabolism. Having highest TM score for similarity with the analog structure among the identified sequences, functional annotation of sequence 7 disclosed its activity as 3′,5′-cyclic-nucleotide phosphodiesterase at molecular level and as negative lymphocyte proliferation regulator at biological level hence, proving its efficient role in retarding the host immune system. Thus, development of immune response against this sequence as peptide vaccine can help to fight the virus ability to elude the host immune system. Considering the functional annotation analysis, we can conclude that the novel sequences identified in the present study can disclose deep secrets behind the pathologies caused by human cytomegalovirus if analysed exhaustively. Several computational functional annotation studies have been taken by many researchers in different context, mulberry transcriptome functional annotation identified the stress responsiveness of 3 protein with unknown functions by using automated pipeline of bioinformatics tools, function prediction of 765 conserved domain sequences of hypothetical proteins in *Exiguobacterium antarcticum* strain B7 by

Table 4
Protective antigen prediction by VaxiJen v.2.0.

Sequence no.	Overall prediction for the protective antigen (Threshold = 0.4)	Antigen/non-antigen
1	0.6542	Antigen
2	0.222	Non-antigen
3	0.2990	Non-antigen
4	−0.0061	Non-antigen
5	0.2896	Non-antigen
6	0.4769	Antigen
7	0.9132	Antigen

Table 5
IC50 values of predicted epitopes of sequences 1,6 and 7 for various MHC I and MHC II alleles.

Seq. no.	Peptide	IC50	MHC I allele	Peptide	IC50	MHC II allele
1	RVKFNVGFG	15.9	HLA-A*30:01	MINVSTNVIVKVTSQ	22.78	HLA-DRB1*13:02
	IVKVTSQAF	17.9	HLA-B*15:01	AFRNRVKFNVGFGLA	36.93	HLA-DRB1*13:02
	AEVEMSRPV	35.6	HLA-B*40:01	VIVKVTSQAFRNRVK	73.95	HLA-DRB5*01:01
	RPVGECTTL	37.7	HLA-B*07:02	VIVKVTSQAFRNRVK	95.3	HLA-DRB1*07:01
	NVSTNVIVK	58.1	HLA-A*68:01	VIVKVTSQAFRNRVK	64.36	HLA-DRB1*01:01
	VTSQAFRNR	59.4	HLA-A*31:01	AFRNRVKFNVGFGLA	89.52	HLA-DRB1*01:01
	AEVEMSRPV	63	HLA-B*44:03			
	AEVEMSRPV	66.6	HLA-B*44:02			
	NVGFGLACV	82.9	HLA-A*68:02			
6	GTASWAVLR	10.5	HLA-A*68:01	ASWAVLRASLPALMV	25.59	HLA-DRB1*09:01
	RACACAHTR	14.8	HLA-A*31:01	VIMTLASGTASWAVL	25.65	HLA-DQA1*05:01/DQB1*03:01
	SLGIAVFYR	22.8	HLA-A*31:01	ASWAVLRASLPALMV	9.44	HLA-DRB1*01:01
	GTASWAVLR	23.1	HLA-A*31:01	ASWAVLRASLPALMV	23.11	HLA-DRB1*07:01
	AVFYRARVR	25.7	HLA-A*31:01	ASWAVLRASLPALMV	21.59	HLA-DRB1*13:02
	GTASWAVLR	27.4	HLA-A*11:01	LGIAVFYRARVRACA	27.1	HLA-DRB1*11:01
	VLRASLPAL	27.4	HLA-A*02:03	LGIAVFYRARVRACA	51.26	HLA-DRB1*15:01
	GIAVFYRAR	35.5	HLA-A*31:01	ASWAVLRASLPALMV	60.32	HLA-DRB4*01:01
	ASLGIAVFY	47.5	HLA-A*11:01	LGIAVFYRARVRACA	29.31	HLA-DRB5*01:01
	GVASLGIAV	47.6	HLA-A*02:03	ASWAVLRASLPALMV	83.56	HLA-DRB3*02:02
	ASLGIAVFY	54.3	HLA-A*30:02	ASWAVLRASLPALMV	65.91	HLA-DRB1*15:01
	HTRDVTNHL	67	HLA-A*30:01	ASWAVLRASLPALMV	46.39	HLA-DRB5*01:01
	GVASLGIAV	69.4	HLA-A*02:06	VIMTLASGTASWAVL	87.29	HLA-DRB1*09:01
	VASLGIAVF	71	HLA-B*35:01	VIMTLASGTASWAVL	24.87	HLA-DRB1*01:01
	SLGIAVFYR	83	HLA-A*68:01	ASWAVLRASLPALMV	98.36	HLA-DQA1*05:01/DQB1*03:01
	IMTLASGTA	87.9	HLA-A*02:03	LGIAVFYRARVRACA	88.65	HLA-DRB1*07:01
	RVRACACAH	93.7	HLA-A*30:01	LGIAVFYRARVRACA	31.93	HLA-DRB1*01:01
	SLGIAVFYR	95.7	HLA-A*11:01			
7	KSLCHFFFV	5.8	HLA-A*02:06	FFLLLVQTFSSSSES	73.57	HLA-DRB1*04:05
	FSSSSESFK	9.7	HLA-A*68:01	FFLLLVQTFSSSSES	78.46	HLA-DRB1*04:01
	KSLCHFFFV	15.7	HLA-A*30:01	QHLDIQFNPVYHKSL	90.35	HLA-DRB1*13:02
	KSLCHFFFV	24	HLA-A*02:01	FFLLLVQTFSSSSES	40.72	HLA-DRB1*01:01
	VYHKSLCHF	25.7	HLA-A*23:01			
	IQFNPVYHK	26.4	HLA-A*11:01			
	VYHKSLCHF	29.1	HLA-A*24:02			
	FSSSSESFK	47	HLA-A*11:01			
	KSLCHFFFV	58.8	HLA-A*31:01			
	FFLLLVQTF	60.9	HLA-A*23:01			
	HLDIQFNPV	63.5	HLA-A*02:06			
	HLDIQFNPV	64.3	HLA-A*02:01			
	HLDIQFNPV	69.5	HLA-A*02:03			
	IQFNPVYHK	70	HLA-A*03:01			
	HFFFVCFFF	82.7	HLA-A*23:01			

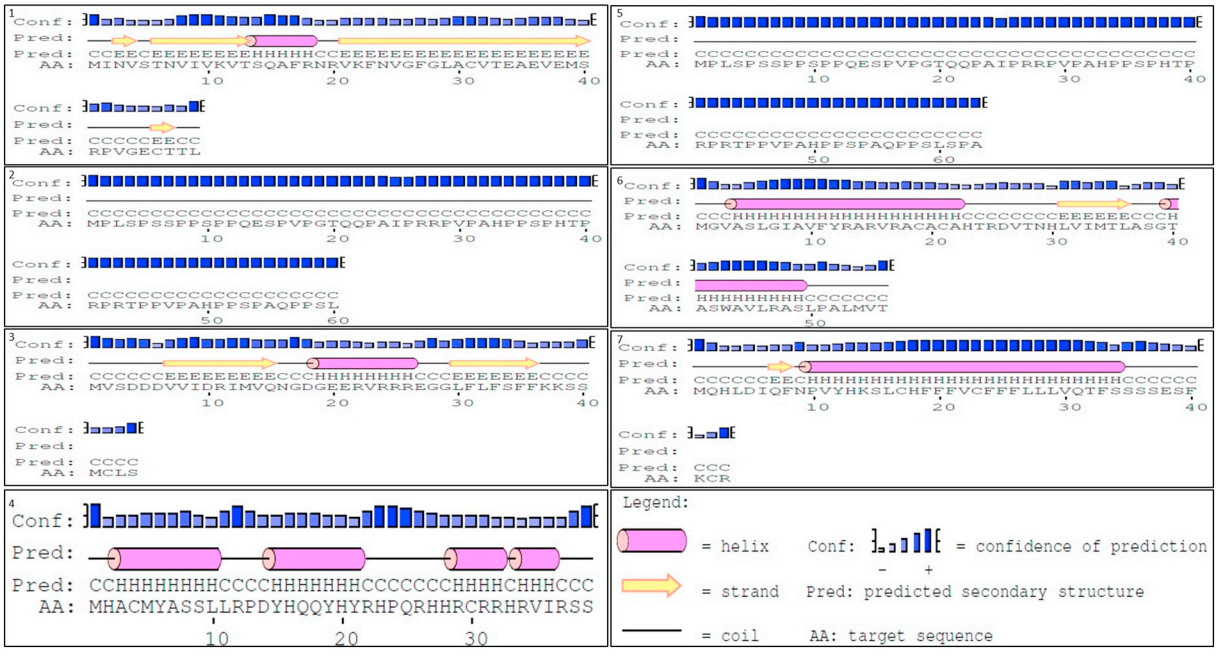

Fig. 3. Secondary structure prediction by PSIPRED for 7 novel sequence peptides.

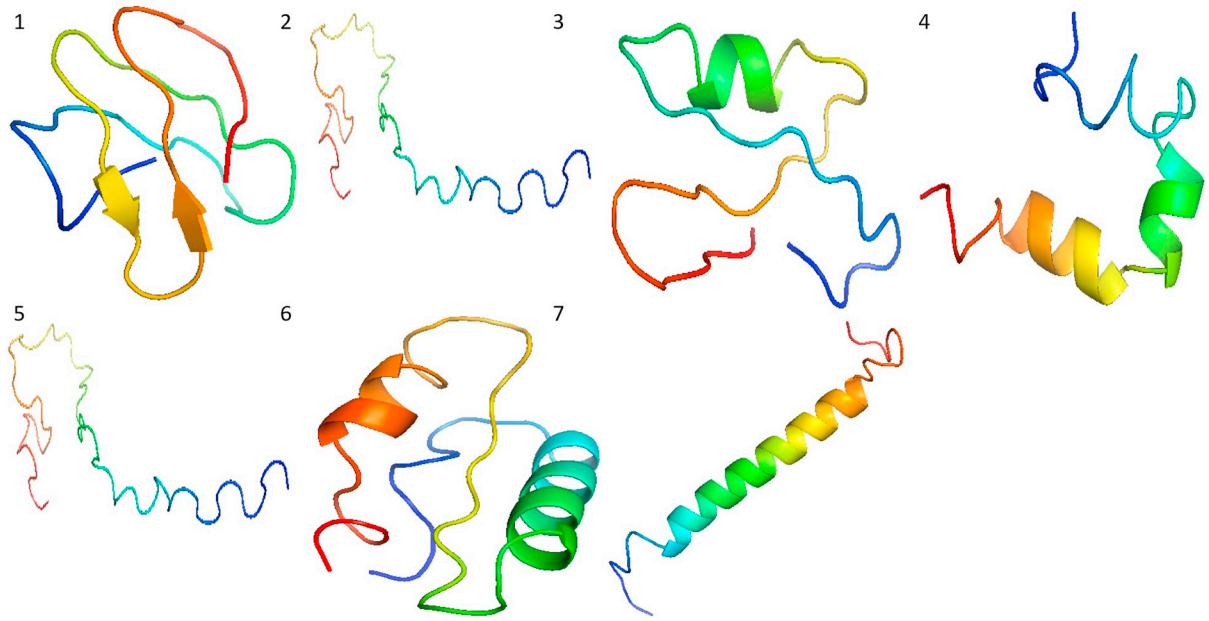

Fig. 4. a) Tertiary structures and b) Ramachandran plots of refined models predicted for 7 novel sequence peptides.

combining several in-silico tools, revealed 132 hypothetical proteins having role in adaptation mechanism to adverse conditions (Dhanyalakshmi et al., 2016; da Costa et al., 2018). All such researches provide great support to our approach of using in-silico bioinformatic tools for the prediction of function.

Hence, our study put forward novel gene sequences, having vital role in viral pathways that can be used as specific and highly effective peptide vaccines to generate immune response against virus without hampering host system. Also, the probable functions of the proteins encoded, in virus growth and multiplication displays their potential to be targeted for therapy development against the virus. Henceforth further analysis by molecular docking and molecular dynamic simulation can be performed to confirm the interaction of proteins translated from novel sequences with the respective immune receptors or with potential therapeutic compounds, which can be highly beneficial for the development of treatment against dreaded congenital CNS abnormalities which occur through intrauterine CMV infections.

4. Conclusion

Human cytomegalovirus (CMV) is responsible for widely diverse congenital infections including adverse CNS developmental abnormalities in new borns. It also adversely infect the immunocompromised adults mainly those who had gone through organ transplantations. Vaccine development against CMV had been a top priority for several scientific organizations for past two decades but effective candidates are still under clinical trials. Several researches have displayed high effectiveness of novel peptide vaccines candidates in eliciting both

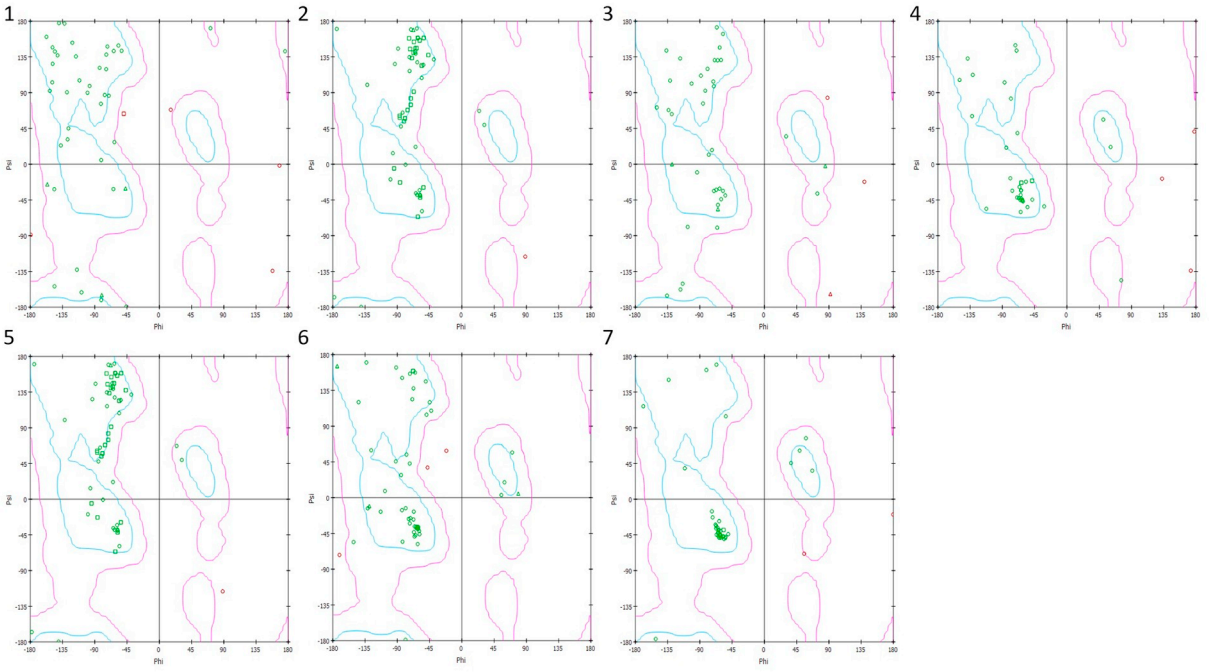

Fig. 4. (*continued*)

Table 6
ProSA-web prediction based on similarity with native structures.

Seq no.	*Z*-score	Similarity of model protein with NMR/XRD analysed structures based on Z-score
1	−5.11	Score in range of NMR structures
2	−1.8	Score in range of NMR structures
3	−3.88	Score in range of NMR structures
4	−1.65	Score in range of NMR structures
5	−1.8	Score in range of NMR structures
6	−3.93	Score in range of NMR structures
7	1.31	Score in range of NMR structures

humoral and cell mediated immune response. Discovery of novel gene sequences and thus the proteins encoded by them can provide new scope in this context. Present study deals with disclosure of 7 novel gene sequences and their analysis for development of CMV treatment as well as peptide vaccine which can evoke significant immunogenic response. The effective functions of the discovered sequences in virus multiplication and mechanisms to evade host immune response, along with their ability to act as epitope for B-cell and MHC class I and II receptors prove them highly potent tools to fight CMV infection. Out of seven sequences, three (Muller et al., 2010; La Rosa and Diamond, 2012; Cheeran et al., 2009) were considered effective for vaccine development. The IC50 values of the epitopes (for MHC I and MHC II) analysed form these three sequences were less than 100 nM thus exhibiting them as highly efficient candidates for vaccine development. Sequence 7 was considered the most potent candidate as it was predicted to contain both B cell and T cell receptor epitopes along with highest protective antigen score and lowest IC50 (less than 10) value for MHC I allele HLA-A*02:06. It also secured good score in tertiary structure prediction and has function related to virus ability to evade host immune system mechanism. Henceforth, extensive research in highly demanded to further analyse the potential of the newly discovered sequences which seems to be highly promising for development of vaccine against CMV infection.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mgene.2020.100754.

Credit author statement

Both authors have equal contribution in designing and writing of manuscript.

Declaration of Competing Interest

There is no conflict of interest among authors.

References

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., Mccombie, W.R., Venter, J.C., 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252 (5013), 1651–1656.

Adhikari, U.K., Tayebi, M., Rahman, M.M., 2018. Immunoinformatics approach for epitope-based peptide vaccine design and active site prediction against polyprotein of emerging Oropouche virus. J. Immunol. Res. 2018, 1–22.

Anderholm, K.M., Bierle, C.J., Schleiss, M.R., 2016. Cytomegalovirus vaccines: current status and future prospects. Drugs 76 (17), 1625–1645.

Azevedo, L.S., Pierrotti, L.C., Abdala, E., Costa, S.F., Strabelli, T.M., Campos, S.V., Ramos, J.F., Latif, A.Z., Litvinov, N., Maluf, N.Z., Caiaffa Filho, H.H., Pannuti, C.S., Lopes, M.H., Santos, V.A., Linardi Cda, C., Yasuda, M.A., Marques, H.H., 2015. Cytomegalovirus infection in transplant recipients. Clinics (Sao Paulo) 70 (7), 515–523.

Bao, L., Dunham, K., Stamer, M., Mulieri, K.M., Lucas, K.G., 2008. Expansion of cytomegalovirus pp65 and IE-1 specific cytotoxic T lymphocytes for cytomegalovirus-specific immunotherapy following allogeneic stem cell transplantation. Biol. Blood Marrow Transpl. 14 (10), 1156–1162.

Besemer, J., Borodovsky, M., 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33 (Web server), W451–W454.

Bhattacharya, D., Nowotny, J., Cao, R., Cheng, J., 2016. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic Acids Res. 44, W406–W409 (Web server issue.

Bialas, K.M., Tanaka, T., Tran, D., Varner, V., Cisneros De La Rosa, E., Chiuppesi, F., Wussow, F., Kattenhorn, L., Macri, S., Kunz, E.L., Estroff, J.A., Kirchherr, J., Yue, Y., Fan, Q., Lauck, M., O'Connor, D.H., Hall, A.H., Xavier, A., Diamond, D.J., Barry, P.A., Kaur, A., Permar, S.R., 2015. Maternal CD4+ T cells protect against severe congenital cytomegalovirus disease in a novel nonhuman primate model of placental cytomegalovirus transmission. Proc. Natl. Acad. Sci. 112 (44), 13645–13650.

Cheeran, M.C., Lokensgard, J.R., Schleiss, M.R., 2009. Neuropathogenesis of congenital cytomegalovirus infection: disease mechanisms and prospects for intervention. Clin. Microbiol. Rev. 22 (1), 99–126.

Clement, M., Humphreys, I.R., 2019. Cytokine-mediated induction and regulation of tissue damage during cytomegalovirus infection. Front. Immunol. 10 (78), 1–9.

Cui, X., Cao, Z., Wang, S., Lee, R.B., Wang, X., Murata, H., Adler, S.P., McVoy, M.A., Snapper, C.M., 2018. Novel trimeric human cytomegalovirus glycoprotein B elicits a high-titer neutralizing antibody response. Vaccine 36 (37), 5580–5590.

da Costa, W.L.O., de Aragão Arau'jo, C.L., Dias, L.M., de Sousa Pereira, L.C., Alves, J.T.C., Arau'jo, F.A., Folador, E.L., Henriques, I., Silva, A., Ribeiro, A., Folador, C., 2018. Functional annotation of hypothetical proteins from the Exiguobacterium antarcticum strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLoS One 13 (6), 1–28.

Dhanyalakshmi, K.H., Naika, M.B.N., Sajeevan, R.S., Mathew, O.K., Shafi, K.M., Sowdhamini, R., Natraja, K.N., 2016. An approach to function annotation for proteins of unknown function (PUFs) in the transcriptome of Indian mulberry. PLoS One 11 (3), 1–18.

Díaz-Cruz, G.A., Smith, C.M., Wiebe, K.F., Cassone, B.J., 2017. First complete genome sequence of tobacco necrosis virus D isolated from soybean and from North America. Genome Announc. 5 (32), 1–2.

Dolan, A., Cunningham, C., Hector, R.D., Hassan-Walker, A.F., Lee, L., Addison, C., Dargan, D.J., McGeoch, D.J., Gatherer, D., Emery, V.C., Griffiths, P.D., Sinzger, C., McSharry, B.P., Wilkinson, G.W., Davison, A.J., 2004. Genetic content of wild-type human cytomegalovirus. J. Gen. Virol. 85 (5), 1301–1312.

Doytchinova, I.A., Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8 (2007), 1–7.

Firbas, C., Jilma, B., Tauber, E., Buerger, V., Jelovcan, S., Lingnau, K., Buschle, M., Frisch, J., Klade, C.S., 2006. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. Vaccine 24 (20), 4343–4353.

Forte, E., Zhang, Z., Thorp, E.B., Hummel, M., 2020. Cytomegalovirus latency and reactivation: an intricate interplay with the host immune response. Front. Cell. Infect. Microbiol. 10, 1–18.

Fouts, A.E., Chan, P., Stephan, J.P., Vandlen, R., Feierbach, B., 2012. Antibodies against the gH/gL/UL128/UL130/UL131 complex comprise the majority of the anti-cytomegalovirus (anti-CMV) neutralizing antibody response in CMV hyperimmune globulin. J. Virol. 86 (13), 7444–7447.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein identification and analysis tools on the ExPASy Server. In: Walker, John M. (Ed.), The Proteomics Protocols Handbook. Humana Press, pp. 571–607.

Gredmark-Russ, S., Söderberg-Nauclér, C., 2012. Dendritic cell biology in human cytomegalovirus infection and the clinical consequences for host immunity and pathology. Virulence 3 (7), 621–634.

Gu, Y., Wei, J., Yang, J., Huang, J., Yang, X., Zhu, X., 2013. Protective immunity against Trichinella spiralis infection induced by a multi-epitope vaccine in a murine model. PLoS One 8 (10), 1–12.

Hassan, J., Dooley, S., Hall, W., 2007. Immunological response to cytomegalovirus in congenitally infected neonates. Clin. Exp. Immunol. 147 (3), 465–471.

Herrera-Rodriguez, J., Meijerhof, T., Niesters, H.G., Stjernholm, G., Hovden, A.O., Sørensen, B., Ökvist, M., Sommerfelt, M.A., Huckriede, A., 2018. A novel peptide-based vaccine candidate with protective efficacy against influenza A in a mouse model. Virology 515, 21–28.

Hong, S., Choi, I., Woo, J.M., Oh, J., Kim, T., Choi, E., Kim, T.W., Jung, Y.K., Kim, D.H., Sun, C.H., Yi, G.S., Eddy, E.M., Cho, C., 2004. Identification and integrative analysis of 28 novel genes specifically expressed and developmentally regulated in murine spermatogenic cells. J. Biol. Chem. 280 (9), 7685–7693.

Jackson, S.E., Redeker, A., Arens, R., van Baarle, D., van den Berg, S.P.H., Benedict, C.A., Čičin-Šain, L., Hill, A.B., Wills, M.R., 2017. CMV immune evasion and manipulation of the immune system with aging. GeroScience 39 (3), 273–291.

Jackson, S.E., Sedikides, G.X., Okecha, G., Wills, M.R., 2019. Generation, maintenance and tissue distribution of T cell responses to human cytomegalovirus in lytic and latent infection. Med. Microbiol. Immunol. 208 (3–4), 375–389.

Jenkins, C., Garcia, W., Godwin, M.J., Spencer, J.V., Stern, J.L., Abendroth, A., Slobedman, B., 2008. Immunomodulatory properties of a viral homolog of human interleukin-10 expressed by human cytomegalovirus during the latent phase of infection. J. Virol. 82 (7), 3736–3750.

Jones, D.T., 1999. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292 (2), 195–202.

Jung, G.S., Kim, Y.Y., Kim, J.I., Ji, G.Y., Jeon, J.S., Yoon, H.W., Lee, G.C., Ahn, J.H., Lee, K.M., Lee, C.H., 2011. Full genome sequencing and analysis of human cytomegalovirus strain JHC isolated from a Korean patient. Virus Res. 156 (1–2), 113–120.

Kamthania, M., Sharma, D.K., 2016. Epitope-based peptides prediction from proteome of Nipah virus. Int. J. Pept. Res. Ther. 22 (4), 465–470.

Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., Nielsen, M., 2013. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including

all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics 65 (10), 711–724.
Klasberg, S., Bitard-Feildel, T., Mallet, L., 2016. Computational identification of novel genes: current and future perspectives. Bioinform. Biol. Insights 10, 121–131.
Kringelum, J.V., Lundegaard, C., Lund, O., Nielsen, M., 2012. Reliable B Cell epitope predictions: impacts of method development and improved benchmarking. PLoS Comput. Biol. 8 (12), 1–11.
La Rosa, C., Diamond, D.J., 2012. The immune response to human CMV. Future Virol. 7 (3), 279–293.
Lehmann, C., Falk, J.J., Büscher, N., Penner, I., Zimmermann, C., Gogesch, P., Sinzger, C., Plachter, B., 2019. Dense bodies of a gH/gL/UL128/UL130/UL131 pentamer-repaired towne strain of human cytomegalovirus induce an enhanced neutralizing antibody response. J. Virol. 93 (17), 1–15.
Li, W., Joshi, M.D., Singhania, S., Ramsey, K.H., Murthy, A.K., 2014. Peptide Vaccine: Progress and Challenges. Vaccines (Basel) 2, 515–536.
Machala, E.A., Avdic, S., Stern, L., Zajonc, D.M., Benedict, C.A., Blyth, E., Gottlieb, D.J., Abendroth, A., McSharry, B.P., Slobedman, B., 2019. Restriction of human cytomegalovirus infection by galectin-9. J. Virol. 93 (3), 1–18.
McVoy, M.A., 2013. Cytomegalovirus vaccines. Clin. Infect. Dis. 4 (Suppl. 4), S196–S199.
Miller-Kittrell, M., Sparer, T.E., 2009. Feeling manipulated: cytomegalovirus immune manipulation. Virol. J. 6 (4), 1–20.
Muller, W.J., Jones, C.A., Koelle, D.M., 2010. Immunobiology of herpes simplex virus and cytomegalovirus infections of the fetus and newborn. Curr. Immunol. Rev. 6 (1), 38–55.
Neff, B.J., Weibel, R.E., Buynak, E.B., McLean, A.A., Hilleman, M.R., 1979. Clinical and laboratory studies of live cytomegalovirus vaccine Ad-169. Proc. Soc. Exp. Biol. Med. 160 (1), 32–37.
Nielsen, P., Krogh, A., 2005. Large-scale prokaryotic gene prediction and comparison to genome annotation. Bioinformatics 21 (24), 4322–4329.
Nielsen, M., Lundegaard, C., Blicher, T., Peters, B., Sette, A., Justesen, S., Buus, S., Lund, O., 2008. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. PLoS Comput. Biol. 4 (7), 1–10.
Nithichanon, A., Rinchai, D., Gori, A., Lassaux, P., Peri, C., Conchillio-Solé, O., Ferrer-Navarro, M., Gourlay, L.J., Nardini, M., Vila, J., Daura, X., Colombo, G., Bolognesi, M., Lertmemonkolchai, G., 2015. Sequence- and structure-based immunoreactive epitope discovery for *Burkholderia pseudomallei* Flagellin. PLoS Negl. Trop. Dis. 9 (7), 1–20.
Novotny, J., Rigoutsos, I., Coleman, D., Shenk, T., 2001. In silico structural and functional analysis of the human cytomegalovirus (HHV5) genome. J. Mol. Biol. 310 (5), 1151–1166.
Oliveros, J.C., 2007–2015. Venny. An Interactive Tool for Comparing Lists With Venn's Diagrams. http://bioinfogp.cnb.csic.es/tools/venny/index.html.
Pan, Y., Lü, P., Wang, Y., Yin, L., Ma, H., Ma, G., Chen, K., He, Y., 2012. In silico identification of novel chitinase-like proteins in the silkworm, *Bombyx mori*, Genome. J. Insect Sci. 12 (150), 1–14.
Pass, R.F., Anderson, B., 2014. Mother-to-child transmission of cytomegalovirus and prevention of congenital infection. J. Pediatr. Infect. Dis. Soc. (Suppl. 1), S2–S6.
Patro, A.R.K., 2019. Subversion of immune response by human cytomegalovirus. Front. Immunol. 10 (1155), 1–7.
Paul, S., Sidney, J., Sette, A., Peters, B., 2016. TepiTool: a pipeline for computational prediction of T cell epitope candidates. Curr. Protoc. Immunol. 114, 1–35.
Plotkin, S.A., Boppana, S.B., 2019. Vaccination against the human cytomegalovirus. Vaccine. 37 (50), 7437–7442.
Plotkin, S.A., Furukawa, T., Zygraich, N., Huygelen, C., 1975. Candidate cytomegalovirus strain for human vaccination. Infect. Immun. 12 (3), 521–527.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16 (6), 276–277.
Rigoutsos, I., Novotny, J., Huynh, T., Chin-Bow, S.T., Parida, L., Platt, D., Coleman, D., Shenk, T., 2003. In silico pattern-based analysis of the human cytomegalovirus genome. J. Virol. 77 (7), 4326–4344.
Riley, H.D., 1997. History of the cytomegalovirus. South. Med. J. 90 (2), 184–190.
Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5 (4), 725–738.
Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. Cell. Mol. Life Sci. 73 (23), 4433–4448.
Saunderson, S.C., McLellan, A.D., 2017. Role of lymphocyte subsets in the immune response to primary B cell–derived exosomes. J. Immunol. 1–12 ji1601537.
Schleiss, M.R., 2013a. Cytomegalovirus in the neonate: immune correlates of infection and protection. Clin. Dev. Immunol. 2013, 1–14.
Schleiss, M.R., 2013b. Developing a vaccine against congenital cytomegalovirus (cmv) infection: what have we learned from animal models? Where should we go next? Future Virol. 8 (12), 1161–1182.
Schleiss, M.R., 2018. Recombinant cytomegalovirus glycoprotein B vaccine: rethinking the immunological basis of protection. Proc. Natl. Acad. Sci. U. S. A. 115 (24), 6110–6112.
Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L., Souopgui, J., 2019. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci. Rep. 9 (1), 1–18.
Singh, H., Raghava, G.P., 2001. ProPred: prediction of HLA-DR binding sites. Bioinformatics 17 (12), 1236–1237.
Singhal, P., Jayaram, B., Dixit, S.B., Beveridge, D.L., 2008. Prokaryotic gene finding based on physicochemical characteristics of codons calculated from molecular dynamics simulations. Biophys. J. 94 (11), 4173–4183.
Stanke, M., Morgenstern, B., 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33 (Web server), W465–W467.
Strasfeld, L., Chou, S., 2010. Antiviral drug resistance: mechanisms and clinical implications. Infect. Dis. Clin. N. Am. 24 (3), 809–833.
Sung, H., Schleiss, M.R., 2010. Update on the current status of cytomegalovirus vaccines. Expert Rev. Vaccines 9 (11), 1303–1314.
Sylwester, A.W., Mitchell, B.L., Edgar, J.B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P.R., Grabstein, K.H., Hosken, N.A., Kern, F., Nelson, J.A., Picker, L.J., 2005. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. J. Exp. Med. 202 (5), 673–685.
van den Berg, S.P.H., Pardieck, I.N., Lanfermeijer, J., Sauce, D., Klenerman, P., van Baarle, D., Arens, R., 2019. The hallmarks of CMV-specific CD8 T-cell differentiation. Med. Microbiol. Immunol. 208 (3–4), 365–373.
Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35 (Web server), W407–W410.
Xia, L., Su, R., An, Z., Fu, T.M., Luo, W., 2018. Human cytomegalovirus vaccine development: immune responses to look into vaccine strategy. Hum. Vaccin. Immunother. 14 (2), 292–303.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER suite: protein structure and function prediction. Nat. Methods 12 (1), 7–8.
Zheng, J., Lin, X., Wang, X., Zheng, L., Lan, S., Jin, S., Ou, Z., Wu, J., 2017. In silico analysis of epitope-based vaccine candidates against hepatitis B virus polymerase protein. Viruses 9 (5), 1–18.