Content available at: https://www.ipinnovative.com/open-access-journals

Indian Journal of Microbiology Research

Journal homepage: www.ijmronline.org

Original Research Article

Predictive analysis of binding affinity among human cytomegalovirus (HCMV) proteins and class-I major histocompatibility complex (MHC-I) molecules

Balaka Biswas[1], Ankita Bhaduri[1], Mousumi Saha[2], Dipanwita Das[3], Agniswar Sarkar[1]*

[1]Dept. of Biosciences, JIS University, Kolkata, West Bengal, India
[2]Dept. of Microbiology, Ballygunge Science College, University of Calcutta, Kolkata, West Bengal, India
[3]The Institute of Cancer Research, Chester Beatty Laboratories, London, England

Abstract

Background: Human Cytomegalovirus (HCMV) is a widespread herpesvirus that establishes persistent infections by evading host immune surveillance. A critical strategy involves the disruption of antigen presentation via Class I Major Histocompatibility Complex (MHC-I) molecules, thereby impairing cytotoxic T lymphocyte (CTL) recognition. This immune evasion is facilitated by a group of HCMV-encoded glycoproteins, US2, US3, US6, US10, and US11, which target distinct stages of the MHC-I processing and presentation pathway.
Materials and Methods: A comprehensive bioinformatics workflow was employed to characterize the structure and function of key HCMV proteins. Protein sequences were sourced from NCBI, and domain structures were analyzed using the Conserved Domain Database (CDD). Coding potential was assessed through reverse translation and ORF prediction. Structural modelling and homology were evaluated via Phyre2, PSI-BLAST, and Clustal Omega. Physicochemical properties were determined using ExPASy ProtParam, and transmembrane regions were predicted with TMHMM. Model validation involved RCSB-PDB, PDBsum, Ramachandran plots, and TM-align. Protein-MHC interactions were visualized using Discovery Studio and PyMOL.
Results: US2 and US3 mimic MHC-I structures to bind and retain them within the endoplasmic reticulum, while US6 inhibits TAP-mediated peptide translocation. US10, with its dual transmembrane topology, disrupts HLA-G trafficking, impacting both CTL and NK cell responses.
Conclusion: This study demonstrates how HCMV proteins interfere with MHC-I antigen presentation, emphasizing their roles in immune evasion. US10 emerges as a key therapeutic target. The findings offer novel insights into HCMV's molecular strategies, paving the way for the development of targeted antiviral treatments and vaccine design.

Keywords: Human cytomegalovirus (HCMV), Human herpesvirus 5, Major histocompatibility complex, Protein structure and function, Viral bioinformatics.

Received: 06-02-2025; **Accepted:** 10-06-2025; **Available Online**: 20-09-2025

1. Introduction

HCMV, a member of the *Herpesviridae* family, establishes lifelong infections and often remains asymptomatic or causes mild symptoms like fever and body aches. After primary infection, HCMV enters a latent phase within host cells, remaining dormant until reactivation triggers replication and symptoms.[1,2] Neonatal infections result from exposure during delivery, with congenital infections affecting 10 in 100,000 live births globally.[3,4] The risk of congenital CMV (cCMV) infections is highest through vertical transmission, occurring in 0.3–1.2% of births, often due to maternal primary infection.[5,6] HCMV can also severely impact immunocompromised individuals and is spread through bodily fluids such as saliva and blood.[7,8]

The HCMV genome is approximately 235 kb of linear double-stranded DNA, the largest among human herpesviruses, encoding over 200 open reading frames (ORFs) and containing unique long (UL) and unique short (US) gene families. A study indicated that 282 viral transcripts were actively translated into proteins, reflecting a significant protein-coding potential. HCMV encodes major long non-coding RNAs (RNA1.2, RNA2.7, RNA4.9,

*Corresponding author: Agniswar Sarkar
Email: ognish@gmail.com

https://doi.org/10.18231/j.ijmr.30210.1758356479

RNA5.0) and numerous miRNAs, which are vital for gene expression and viral pathogenesis.[9,10] The mature HCMV envelope features a lipid bilayer with eleven viral glycoproteins, alongside intrinsic membrane proteins and a capsid covered by a tegument that aids replication. Glycoproteins such as gB, gH, gL, gN, and gM are crucial for host cell penetration, acting like keys that fit specific receptors on host surfaces.[11,12] HCMV replicates in various cell types, including epithelial and renal cells. Class I Major Histocompatibility Complex (MHC-I) molecules are essential for activating cytotoxic cells like NK and CD8+ T cells. HCMV evades immune detection by downregulating MHC-I molecules and impairing antigen presentation, which is critical for the immune response.[13,14] HCMV proteins (US2, US3, US6, US10, US11) target MHC-I pathways for immune evasion. US2 directs MHC-I to degradation pathways, diminishing its presence on the cell surface.[15,16] US3 prevents MHC-I transport to the Golgi, blocking maturation and presentation. US6 binds MHC-I in the ER, hindering its transport. US10 internalizes mature MHC-I via endocytosis, and US11 inhibits MHC-I assembly with peptides, collectively diminishing MHC-I availability, allowing HCMV to evade immune detection [17-19]. This study employs bioinformatics to analyze the properties of key HCMV proteins, focusing on their affinities to MHC-I, providing insights relevant to vaccine design, particularly concerning US10. Multiple viral protein groups of HCMV are summarized in **Table 1**.

2. Materials and Methods

2.1. Sequence retrieval and analysis of biological roles

HCMV encodes US2, US3, US6, US10, and US11; and sequences from human class I MHC molecules were retrieved from the National Centre for Biotechnology Information (NCBI) protein database (https://www.ncbi.nlm.nih.gov/). These sequences were acquired in FASTA format, a standard format for nucleotide or protein sequences. These sequences were compared using the NCBI Conserved Domains Database (NCBI-CDD) to identify conserved domains- regions of proteins that are structurally and functionally conserved across species (https://www.ncbi.nlm.nih.gov/cdd/). The protein sequences were reverse-translated into nucleotide sequences using Reverse Translate, and potential protein-coding regions were identified using ORF Finder. This analysis also highlighted the genetic makeup of the sequences (https://www.ncbi.nlm.nih.gov/orffinder/).[20]

2.2. Protein secondary structure prediction and analysis

Protein sequences were analyzed using the Phyre2 Protein Fold Recognition Server, which predicts three-dimensional structures, domain architecture, and functional annotations based on homology modelling.[21] Phyre2 also identified intrinsically disordered regions and associated protein families and superfamilies. To investigate evolutionary relationships and conserved motifs, multiple sequence alignment (MSA) was performed using Clustal Omega version 1.2.4.[22] Homologous sequences were identified using PSI-BLAST, facilitating comparative analysis of structural similarities and variations. Physicochemical properties such as sequence length, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, and GRAVY (grand average of hydropathicity) were computed using the ExPASy ProtParam tool.[22,23] These parameters provided critical insights into each protein's solubility, stability, and hydrophobicity. A phylogenetic tree was generated from Clustal Omega alignment data to visualize evolutionary relationships. Transmembrane helices were predicted using TMHMM 2.0, which applies a hidden Markov model to identify membrane-spanning regions, essential for understanding epitope accessibility and membrane association.[24] Additionally, FASTA sequences of MHC Class I molecules obtained from the NCBI database were analyzed using GeneThreader. This analysis included pH profiles and amino acid composition, offering further insight into the structural stability and functional features of MHC Class I proteins.[25]

2.3. Protein structure validation and property analysis

Protein sequences were submitted to the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB-PDB) to retrieve detailed structural information, including three-dimensional atomic coordinates, ligand interactions, water molecules, and experimental conditions such as crystallization and X-ray diffraction data.[26] The resulting PDB files were further analyzed using PDBsum, which provides functional annotations and structural summaries, including information on secondary structure elements, active sites, ligand binding, and structural motifs such as pores and tunnels, key features for identifying potential epitope regions.[27] Ramachandran plots generated through PDBsum were used to evaluate the phi (ϕ) and psi (ψ) torsion angles of the backbone atoms, allowing assessment of stereochemical quality and identification of residues in disallowed regions, which could compromise protein stability.[28] Bond lengths and bond angles were also examined to ensure structural integrity. To assess structural similarity and evolutionary conservation, TM-align was applied to compare modelled HCMV proteins with homologous structures identified via PSI-BLAST. TM-align outputs included TM-score and RMSD (Root Mean Square Deviation), with higher TM-scores and lower RMSD values indicating a high degree of structural similarity and potential functional relevance.[29] Discovery Studio 2024 Client was then used to refine and visualize the structural alignments, including the interaction interfaces between HCMV proteins and human MHC molecules.[30] TM-align was also employed to predict binding affinities with MHC Class I proteins, supporting epitope mapping and immunological relevance. Final structural validation and visualization were performed using PyMOL, which confirmed the alignment results and

provided high-resolution images for accurate interpretation of epitope positioning and structural features.[31]

3. Results

HCMV employs a set of viral proteins, US2, US3, US6, US10, and US11, to evade immune detection by interfering with MHC Class I-mediated antigen presentation. These proteins act as immune modulators, enabling the virus to escape cytotoxic T cell responses. MHC Class I protein sequences were retrieved from NCBI and analyzed using Gene Threader, revealing that these immune molecules are highly acidic.[32] Alanine was the most abundant amino acid (9.39–10.84%), supporting structural stability and adaptability, while asparagine, cysteine, and methionine were least represented, yet likely play critical roles in function and binding.[33,34] Domain architecture analysis via NCBI-CDD showed that each HCMV protein possesses distinct structural features tailored for immune evasion. US2 and US3, part of the same family (pfam05963), are ER-resident proteins that bind MHC-I molecules directly, preventing their surface expression. Structural modelling based on US2-HLA-A2/tax crystal structures confirmed high-confidence interactions, with significant TM-align scores and e-values (US2: 294.76, e-value 1.32e-102; US3: 248.54, e-value 1.25e-84). These findings are consistent with earlier studies by Lee et al. and Gewurz et al., which showed US2 and US3 trap MHC-I in the ER, disrupting immune surveillance during early infection stages.[34,35]

3.1. Functional and structural diversification of HCMV immune evasion proteins:

HCMV proteins US6 and US10 exhibit unique mechanisms for evading the immune response and are categorized into separate PFAM families, with e-values of 3.35e-111 and 1.67e-125, respectively. While they are comparable in size, with US6 comprising residues 25–183 and US10 consisting of residues 25–185, their functional roles are markedly different. US6 is involved in inhibiting the transporter-associated protein (TAP), disrupting the translocation of peptides into the endoplasmic reticulum (ER) and preventing the loading of antigens onto MHC-I.[36] Conversely, US10 targets the tri-leucine motif within MHC-I molecules, impacting their surface trafficking and resulting in downregulation of HLA-G expression through its cytoplasmic tail, distinguishing it from other viral immune modulators.[37]

Additionally, US11, associated with a lower-confidence domain assignment (bit score 51.73, e-value 4.81e-08), also interferes with antigen presentation. These immune-modulating proteins are encoded by distinct ORFs and identified as US2: ORF 3, US3: ORF 4, US6: ORF 1, US10: ORF 2, and US11: ORF 2. The region extending from US2 to US11 is organized as an immune evasion module.[38,39]

Biochemical analyses performed via the ExPASy-Swiss Bioinformatics Resource Portal reveal that these transmembrane proteins, ranging from 183 to 215 amino acids and possessing molecular weights between 20,611.96 and 25,288.44 Da, display diverse physicochemical properties. Their isoelectric points (pI) range from 5.46 to 8.73, with US2 exhibiting the greatest stability (instability index of 28.71) and US11 the least stable (56.46). The aliphatic indices, varying between 92.73 and 100.97, indicate high thermal stability among these proteins. The GRAVY scores highlight their hydrophobicity profiles; for example, US2 (0.131) is associated with membranes, while US6 (−0.138) is more suited to aqueous environments (**Figure 1** C, D). These variances mirror their functional specialization; membrane-anchored proteins such as US2 disrupt host signalling pathways, whereas hydrophilic proteins like US6 affect intracellular processing. Such distinctions also offer critical insights into protein purification, solubility, and therapeutic targeting strategies.[18,40,41]

3.2. Template-based structural modelling and membrane integration of HCMV proteins

Template-based structural modelling with Phyre2 identified various fold types and secondary structure profiles in HCMV transmembrane proteins. US2 and US3 showed 100% confidence with immunoglobulin-like β-sandwich folds, covering 48% and 47% of their sequences, respectively, and had significant β-strand content (46% for US2 and 39% for US3) alongside α-helical regions (19% for US2 and 31% for US3). US6 was modelled on a phospholipase A2 template (15.2% confidence, 52% coverage), showing 46% α-helical content and 27% disordered regions. US10 used a PMMO template (45% coverage, 6.1% confidence) with 34% α-helix and 23% β-strand. US11 aligned with yeast autophagy protein Atg3 (40% coverage, 12.6% confidence), showing 28% α-helix and 32% β-strands. Structural templates from the RCSB Protein Data Bank provided insights into the architecture and therapeutic strategies for HCMV immune evasion proteins. US2 and US11 share a single transmembrane domain (residues 181–203), impacting MHC Class I trafficking.[42,43] US3 and US6 have single transmembrane helices at residues 161–183 and 147–169, respectively, while US10 has a dual transmembrane topology (residues 127–149 and 159–181), possibly enhancing membrane anchoring and interactions with host cell machinery.

Table 1: Lists HCMV proteins identified in virions and dense bodies via LC-MS/MS and FTICR mass spectrometry. Proteins with lower LC-MS/MS confidence but confirmed by FTICR are marked with asterisks

Viral protein group	HCMV ORF	LC-MS/MS		FTICR	Coverage (%)
		No. of peptides	Max X_{Corr}	No. of peptides	
Virion proteins					
Capsid	UL46	20	5.30	14	44.8
	UL48-49	8	6.52	5	54.7
	UL80	37	6.36	30	35.6
	UL85	21	6.73	22	63.1
	UL86	149	3.97	123	71.0
Tegument	UL24	8	5.06	9	38.3
	UL25	60	7.04	59	59.2
	UL26	9	4.77	10	53.7
	UL32	135	3.01	100	70.5
	UL43	7	5.50	10	28.1
	UL47	53	6.10	64	57.5
	UL48	111	4.29	109	56.8
	UL82	70	6.39	47	69.3
	UL83	123	5.44	86	92.0
	UL94	10	5.08	12	26.4
	UL99	8	5.87	9	64.7
	US22	2	3.16	2	5.4
	US23	1	2.61	1	4.6
	US24	1	4.83	2	7.0
Glycoproteins	RL10	5	2.36	4	22.8
	TRL14	*[a]		1	7.5
	UL5	*		1	5.4
	UL22A	1	5.04	1	19.4
	UL33	4	6.11	4	14.1
	UL38	*		1	5.7
	UL41A	2	5.72	2	25.6
	UL50	1	2.82	4	10.6
	UL55	21	6.16	23	24.8
	UL73	2	3.47	2	6.5
	UL74	4	5.07	4	13.5
	UL75	21	6.15	22	35.7
	UL77	14	5.65	12	31.2
	UL93	15	5.35	14	31.7
	UL100	13	5.24	7	15.9
	UL115	11	4.73	9	47.1
	UL119	2	2.23	1	4.6
	UL132	8	5.89	8	47.0
	US27	4	4.25	2	7.7
Transcription-replication machinery	IRS1	15	6.01	17	25.8
	TRS1	10	6.92	23	34.7
	UL44	1	4.32	9	31.0
	UL45	43	5.85	52	52.2
	UL54	*		1	1.6
	UL57	*		1	0.4
	UL69	6	4.17	7	19.0
	UL72	*		1	4.6

	UL84	1	2.50	3	12.8
	UL89	*		1	3.1
	UL97	13	5.95	9	32.1
	UL122	2	4.26	4	11.7
Uncharacterized	UL35	42	6.27	40	56.1
	UL51	*		1	3.2
	UL71	12	6.32	11	40.4
	UL79	*		1	10.9
	UL88	14	6.8	17	33.6
	UL96	1	4.46	1	19.7
	UL103	8	5.18	8	37.0
	UL104	9	4.68	9	23.0
	UL112	1	3.30	4	4.7
Dense body proteins					
Capsid	UL46	1	3.6	6	
	UL48-49	2	5.8	1	
	UL80	1	6.1	2	
	UL85	4	5.0	4	
	UL86	22	5.0	19	
Tegument	UL25	17	6.3	13	
	UL26	3	3.6	3	
	UL32	11	5.4	15	
	UL35	5	5.6	9	
	UL47	2	4.3	6	
	UL48	7	5.4	12	
	UL82	9	5.1	6	
	UL83	40	6.3	14	
	UL75	4	5.6	2	
Transcription-replication machinery	UL45	2	4.3	6	
	IRS1	3	5.6	2	
	TRS1	1	4.7	5	

Table 2: Shows physicochemical properties of MHC Class I proteins, highlighting their acidic nature (pI -11.08 to -4.96) and a high alanine content linked to structural stability; asparagine, cysteine, and methionine are less abundant

S. No.	Protein	pH	Highest	Value	Lowest	Value
1.	ACR55720.1	-6.06	Alanine	10.14	Asparagine	1.1
2.	CAL85437.2	-9.47	Alanine	10.84	Asparagine	10.84
3.	UPW16506.1	-6.63	Alanine	9.94	Cysteine	1.38
4.	BDU99860.1	-7.86	Alanine	10.66	Methionine Asparagine	1.37
5.	BDU99859.1	-5.64	Alanine	10.22	Asparagine Cysteine	1.38
6.	BDU99858.1	-4.96	Alanine	9.39	Methionine	1.1
7.	BDU99857.1	-7.64	Alanine	9.86	Asparagine	1.1
8.	BDU49430.1	-11.08	Alanine	10.11	Asparagine	1.37
9.	BDU18538.1	-5.68	Alanine	9.39	Methionine	1.1
10.	BDU18537.1	-5.64	Alanine	9.94	Cysteine	1.38

Table 3: Summarizes key properties of five HCMV proteins, noting US2 and US11's acidity, US2's high stability and hydrophobicity, and variability in extinction coefficients related to disulphide bonds

Protein (s)	NCBI Reference Sequence	No. of Amino Acids	Molecular Weight (Da)	Theoretical pI	Instability Index	Gravy	Total Atom Nos.	Extinction Coefficient (Cys-Cys)	Extinction Coefficient (Reduced Cys)
US2	QNT12687.1	199	23110.95	6.82	28.71	0.131	3236	68660	68410
US3	AAS49002.1	186	21514.98	8.59	44.20	0.076	3032	39545	39420
US6	AAS49004.1	183	20611.96	8.73	45.66	-0.138	2889	17710	16960
US10	YP_081595.1	185	20771.09	8.1	47.21	0.051	2918	40380	39880
US11	YP_081596.1	215	25288.44	5.46	56.46	-0.029	3558	70610	70360

Table 4: Presents structural alignment results of HCMV proteins using TM-align, with US3 and US10 showing the highest similarity (US10 perfectly aligned), moderate similarity for US6 and US2, and the lowest for US11

Proteins	Sample Acc. No.	Control Acc. No.	Length of Chain_1	Length of Chain_2	Aligned Length	RMSD	Seq. ID	TM-score (Chain_1)	TM-score (Chain_2)
US2	A398799	B398799	275	613	159	6.55	0.031	0.32891	0.18085
US3	A768951	B768951	275	275	275	0.00	1.000	1.00000	1.00000
US6	A627561	B627561	119	738	90	4.86	0.056	0.41385	0.09977
US10	A456694	B456694	362	362	362	0.00	1.000	1.00000	1.00000
US11	A342978	B342978	119	213	62	4.70	0.065	0.30271	0.19793

Table 5: Details superimposition of HCMV proteins on ten MHC Class I molecules, identifying closest structural matches (e.g., US2/US3 with 1m3), illustrating molecular mimicry and potential immune interactions

S. No.	MHC CLASS I molecules	HCMV proteins	Chain 1	Chain 2	Aligned Length	RMSD	TM-Score 1	TM-Score 2
1.	8rbu	1im3	A39262	B39262	275	1.09	0.97375	0.97375
	8rbu	2not	A43136	B43136	87	4.52	0.22385	0.41101
	8rbu	3chx	A775119	B775119	125	6.17	0.26921	0.22068
	8rbu	2dyt	A225357	B225357	93	5.54	0.21168	0.253
2.	6at5	1im3	A915955	B915955	275	0.93	0.97831	0.97831
	6at5	2not	A76528	B76528	62	4.12	0.16994	0.32446
	6at5	3chx	A375128	B375128	125	6.26	0.26725	0.21876
	6at5	2dyt	A938623	B938623	94	5.84	0.20983	0.25083
3.	8rh6	1im3	A830088	B830088	275	1.07	0.97431	0.97431
	8rh6	2not	A265052	B265052	84	4.97	0.20479	0.36051
	8rh6	3chx	A770980	B770980	128	6.7	0.25794	0.21251
	8rh6	2dyt	A614219	B614219	94	5.79	0.21694	0.26199
4.	6avf	1im3	A995997	B995997	91	1.98	0.80504	0.30499
	6avf	2not	A823717	B823717	49	4.38	0.27706	0.23798
	6avf	3chx	A162769	B162769	86	3.95	0.52829	0.1888
	6avf	2dyt	A849825	B849825	73	4.34	0.43199	0.23963
5.	8rcv	1im3	A567132	B567132	275	0.97	0.97318	0.97665
	8rcv	2not	A710917	B710917	84	5.11	0.20065	0.34949
	8rcv	3chx	A206273	B206273	125	6.16	0.26646	0.21923
	8rcv	2dyt	A261797	B261797	96	5.59	0.22061	0.26591
6.	8ref	1im3	A38683	B38683	275	0.87	0.97754	0.98104
	8ref	2not	A657214	B657214	83	5.02	0.19822	0.34421
	8ref	3chx	A651801	B651801	125	6.1	0.26846	0.22067
	8ref	2dyt	A845288	B845288	94	5.46	0.21787	0.26214

7.	7tlt	1im3	A790795	B790795	272	0.89	0.97686	0.96987
	7tlt	2not	A643556	B643556	77	5.31	0.18021	0.30753
	7tlt	3chx	A272931	B272931	132	6.67	0.26233	0.21609
	7tlt	2dyt	A736837	B736837	96	5.9	0.21743	0.25974
8.	6avg	1im3	A319539	B319539	94	2.16	0.79055	0.31365
	6avg	2not	A665036	B665036	52	4.81	0.25042	0.22573
	6avg	3chx	A103048	B103048	90	3.94	0.53554	0.19852
	6avg	2dyt	A7491	B7491	72	4.24	0.42602	0.24219
9.	7tlt	1im3	A893416	B893416	275	0.95	0.97705	0.97705
	7tlt	2not	A267335	B267335	82	4.91	0.20548	0.3755
	7tlt	3chx	A269864	B269864	126	6.24	0.27106	0.22162
	7tlt	2dyt	A521269	B521269	94	5.88	0.21576	0.26075
10.	7rtd	1im3	A549898	B549898	274	0.73	0.98646	0.98292
	7rtd	2not	A424143	B424143	84	5.03	0.20322	0.35317
	7rtd	3chx	A510367	B510367	122	6.1	0.26341	0.21552
	7rtd	2dyt	A172716	B172716	100	5.76	0.22927	0.27483

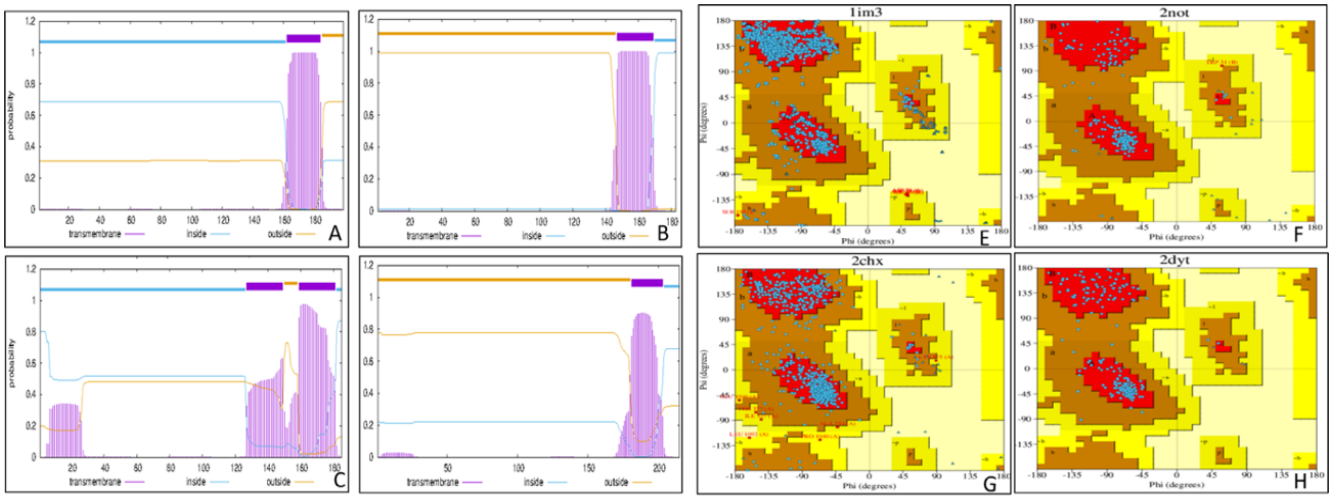

Figure 1: Sequence alignment and hydropathicity of HCMV proteins. (**A**): US2 shows a conserved core region modelled on 1im3.1, with colour-coded domains. (**B**): US3 has lower alignment and more gaps than US2. (**C**): US6 exhibits the lowest hydropathicity (−0.138), highlighting possible membrane interaction sites. (**D**): US2 has the highest hydropathicity (0.131), dominated by hydrophilic residues

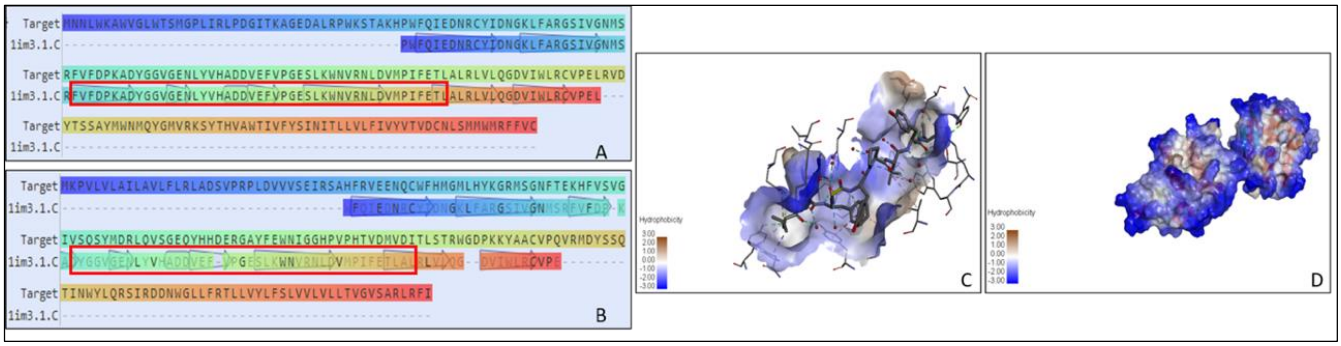

Figure 2: Domain annotations and 3D models of HCMV proteins. US2 & US3 belong to Cytomega_US3 family. US6 models to US6 superfamily; US10 to US10 domain; US11 to CMV_US superfamily. Homology modelling shows US2 complexed with MHC-I, US6 related to neurotoxic phospholipase A2, US10 to methane monooxygenase, and US11 to autophagy protein Atg3

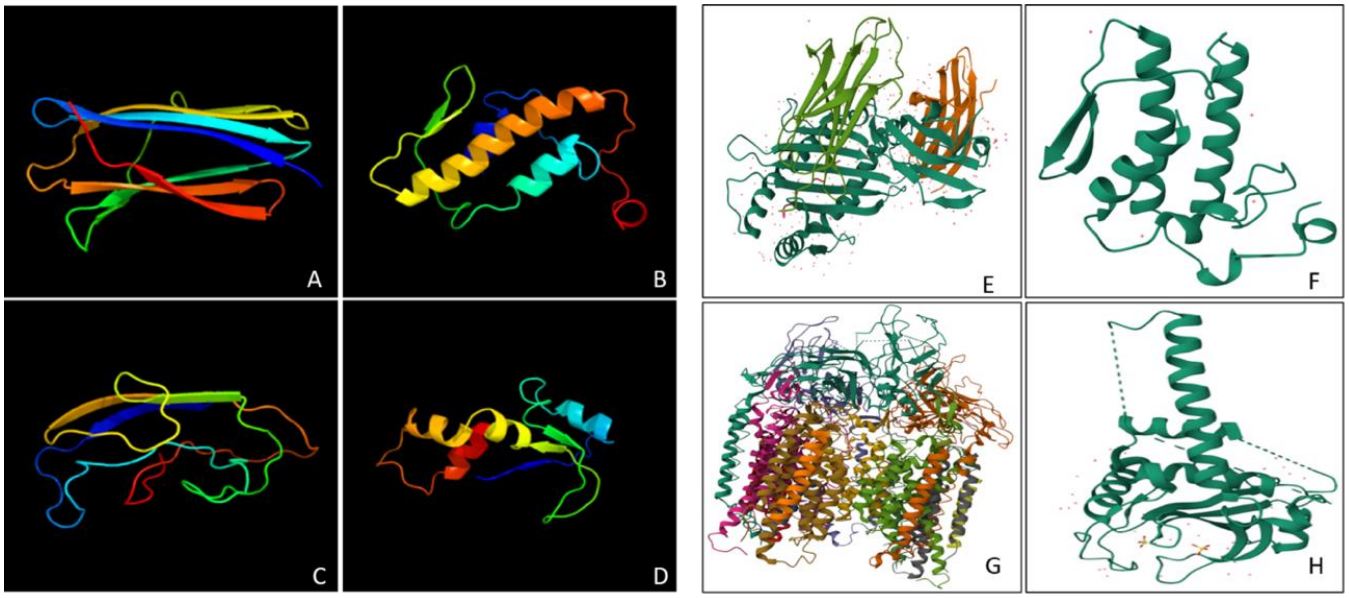

Figure 3: Transmembrane helix predictions and Ramachandran plots. US2 and US3: single TM helix (~161-184). US6 & US11: one TM helix each (147-169 and 181-203, respectively). US10: two TM helices (127-149 and 159-181). Ramachandran plots show >88% residues in favoured regions for all proteins, indicating stable folds

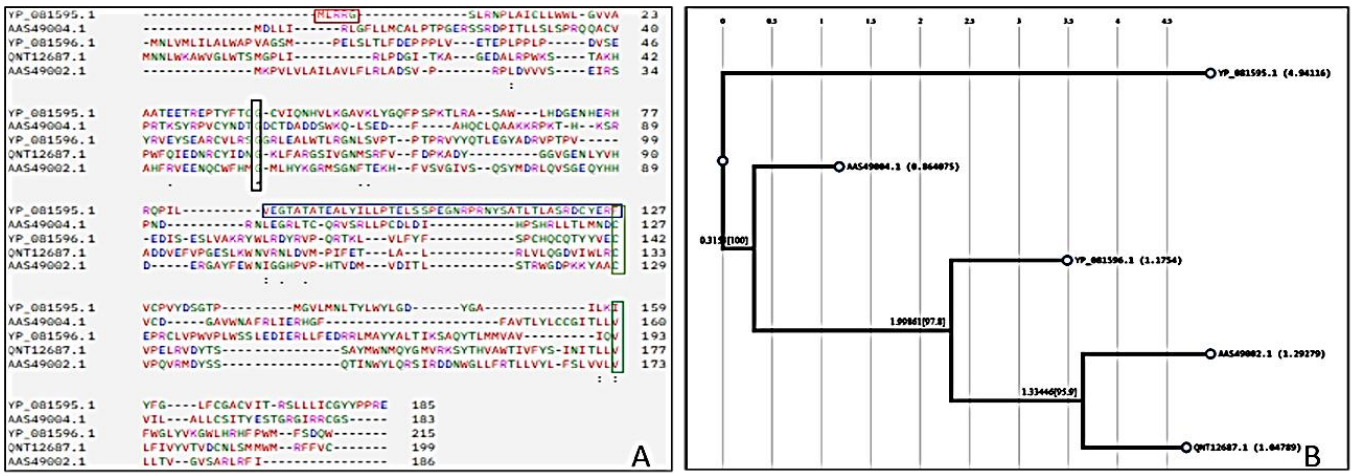

Figure 4: Multiple sequence alignment (MSA) and phylogeny of US2, US3, US6, US10, and US11. (**A**): US2, US3, and US6 share conserved regions; US10 has many indels and the shortest N-terminus. (**B**): Phylogenetic tree groups US2 and US3 closely; US6 and US11 cluster separately; US10 is a distant outlier

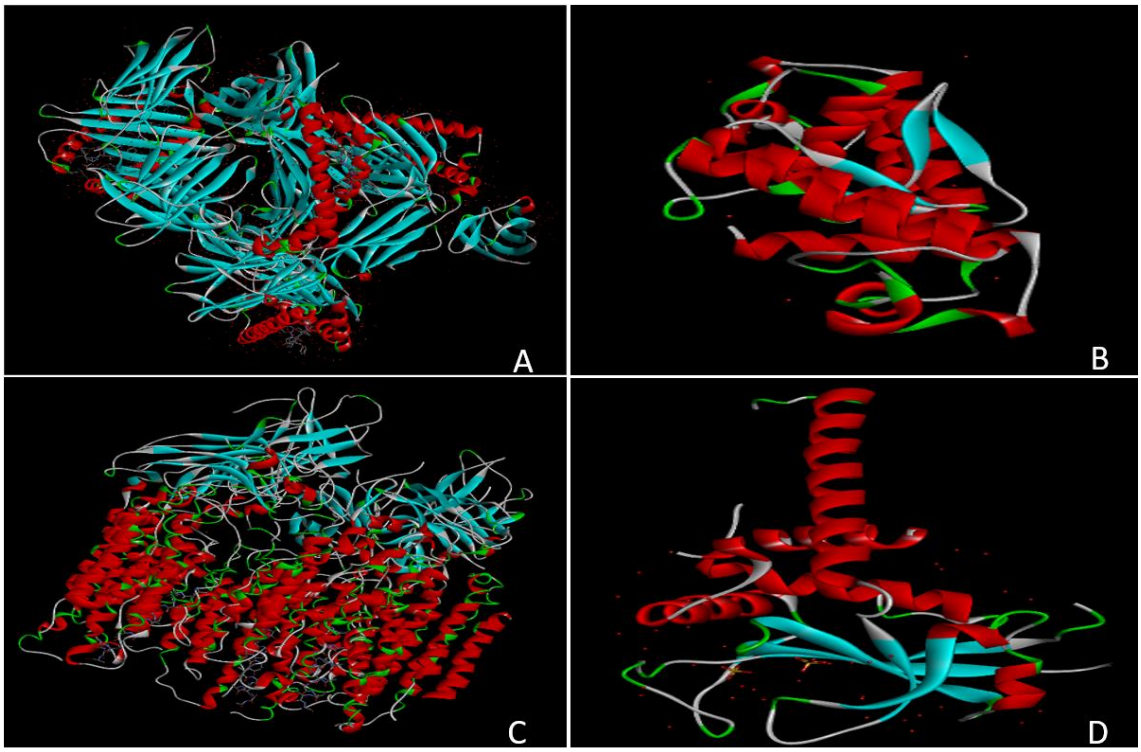

Figure 5: Structural comparison of HCMV protein models. US2 and US3: β-helix rich. US6: compact α-helical globular structure. US10: largest, complex with α-helices and β-sheets. US11: extended, flexible conformation

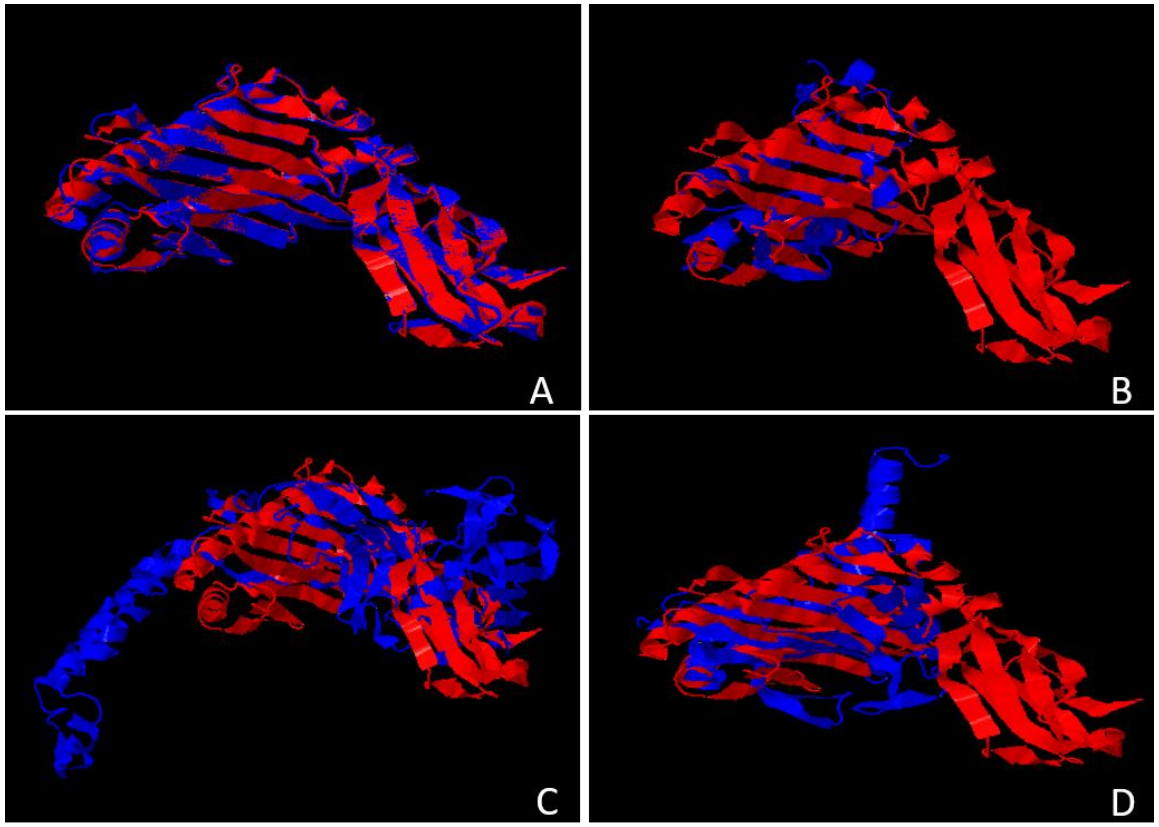

Figure 6: Superimposition of HCMV proteins with MHC Class I molecules. (**A**): US2 and US3 are closely aligned with MHC-I. (**B**): US6 fits tightly in the MHC core. (**C**): US10 shows poor alignment, weak interaction. **(D)**: US11 moderate alignment, less stable interaction

4. Discussion

4.1. Structural quality and evolutionary analysis of HCMV immune evasion proteins

Ramachandran plot analysis demonstrated that most HCMV proteins are well-folded and structurally stable. US6 and US11 exhibited exceptional structural quality, with over 91% of residues residing in favoured conformational regions. US2 and US3 also showed high-quality folding, with approximately 90% of residues in favourable geometries. These proteins benefit from the strategic distribution of glycine and proline residues, glycine conferring flexibility through hinge-like regions, and proline introducing rigidity that stabilizes protein folds.[44,45] In contrast, US10 displayed comparatively lower structural regularity, with only 88.4% of residues in favoured regions and a negative G-factor score, indicating atypical geometric features.

Multiple sequence alignment (MSA) revealed conserved domains primarily between residues 60–70 and 120–140 across the proteins, implicating these regions in essential functional roles. US2, US3, and US6 share notable sequence conservation and structural similarity, supporting a common evolutionary origin and coordinated functional mechanisms. Phylogenetic analysis further corroborated these relationships, US2 and US3 formed a sister clade, closely associated with US6, reflecting their cooperative roles in MHC Class I modulation. US11 clustered more distantly yet remained related, while US10 emerged as a distinct outlier, underscoring its evolutionary divergence and unique functional specialization. TM-align structural analysis provided quantitative measurements of structural similarity among HCMV proteins, with results summarized in **Table 4**. The most striking finding was that US2 and US10 exhibited perfect structural alignment with TM-scores of 1.000 and RMSD values of 0.00, indicating identical three-dimensional structures. This analysis suggested these proteins share high structural conservation despite their evolutionary distance, though this finding may reflect limitations in structural prediction models rather than true structural identity.

4.2. Structural comparison of HCMV immune evasion proteins

Complementing these findings, TM-align structural comparisons quantified protein similarities (**Table 4**). Notably, US2 and US10 exhibited perfect structural alignment with TM-scores of 1.000 and RMSD values of 0.00, suggesting identical predicted three-dimensional conformations. This high degree of structural similarity, despite their evolutionary distance, may reflect inherent limitations in predictive modelling rather than true structural identity. In contrast, US2, US6, and US11 displayed moderate to low structural similarity, consistent with their evolutionary divergence and distinct functional roles. **Figure 5** illustrates detailed structural comparisons: US2 and US3 share a combined fold dominated by beta-barrel elements, likely conferring stability essential for their trafficking and interaction with MHC Class I molecules. US6 adopts a compact, globular structure rich in alpha-helices, facilitating its role in molecular recognition and regulatory interactions. US10's structure is notably complex, integrating extensive alpha-helical and beta-sheet content, reflective of its dual transmembrane domains and specialized targeting of HLA-G molecules. Finally, US11 exhibits an extended alpha-helical conformation that provides conformational flexibility, enabling versatile binding to diverse host targets, a feature critical for dynamic immune evasion.

4.3. Comparative structural analysis of HCMV proteins and MHC class I molecules

The comparative structural analysis between HCMV proteins and MHC Class I molecules (**Table 5**) revealed variable degrees of similarity corresponding to their functional roles. MHC molecules exhibited high structural conservation both among themselves and with the US2/US3 complex (PDB ID: 1IM3), with TM-scores consistently above 0.9 and minimal RMSD values. This strong resemblance underscores a sophisticated molecular mimicry by US2 and US3, enabling their integration into the host MHC processing pathway. US6 displayed moderate similarity to MHC molecules, with TM-scores indicative of complementary but not identical structural features, reflecting an evolutionary strategy distinct from direct mimicry. In contrast, US10 demonstrated the lowest structural similarity, with RMSD values generally exceeding 6Å, highlighting its specialized targeting of alternative immune recognition components. US11 showed slightly better alignment than US10, but low TM-scores and RMSD values above 5 Å confirmed its divergence from MHC-like structures, consistent with its unique functional adaptations. Alignments were performed between ten MHC Class I variants and four HCMV proteins (US2, US6, US10, US11), yielding 40 interaction models. The TM and RMSD metrics were consistent across different MHC variants for each viral protein, indicating stable structural relationships independent of the MHC allele. Representative alignment examples for each protein-MHC pair are illustrated in **Figure 6**, effectively capturing characteristic interaction patterns while minimizing redundancy.US2 and US3 proteins demonstrate exceptional structural confidence and therapeutic potential. Both belong to the pfam05963 domain family and achieved 100% confidence in Phyre2 modelling with 48% and 47% coverage, respectively, using immunoglobulin-like beta-sandwich fold templates (**Figure 1** A, B). RCSB modelling with 1im3 revealed hetero-trimer (1-1-1-mer) configurations involving identical ligands: LEU-LEU-PHE-GLY-TYR-PRO-VAL-TYR-VAL.

Ramachandran plot analysis showed 90.3% of residues in favoured regions with single transmembrane helix architecture and high TM scores compared to 1im3 **(Figure 3** E**)**. These characteristics position US2 and US3 as prime candidates for epitope-based vaccine design and therapeutic applications.

4.4. Conserved domains and vaccine potential of HCMV immune evasion proteins

Multiple sequence alignment and phylogenetic analyses revealed conserved regions among US2, US3, and US6, particularly between residues 60-140, reflecting strong evolutionary conservation and functional importance in disrupting MHC-I antigen presentation (**Figure 4** A, B). These ER-resident glycoproteins' conserved domains make them promising targets for epitope-based vaccine development. The high structural similarity of US2 and US3 to MHC-I (PDB: 1IM3), coupled with stable modelling results, supports their potential use as subunit vaccine antigens to prime immune responses against early viral immune evasion. US10, although showing low structural similarity to MHC-I, possesses unique immunomodulatory functions via its conserved tri-leucine motif and complex dual transmembrane topology. It specifically downregulates HLA-G, modulating NK cell responses and expanding the immunogenic scope of potential vaccines to include both T cell and NK cell evasion mechanisms.[18] Incorporating proteins like US10 could enhance vaccine breadth by targeting multiple immune evasion pathways. US10 is a distinctive and less-characterised HCMV protein expressed early during infection.[13] Unlike other US proteins that degrade MHC-I, US10 delays HLA-G trafficking through its tri-leucine motif in the cytoplasmic tail, enabling NK cell evasion.[18] Structurally, US10 is a 185-residue protein containing a conserved Pfam17617 domain (residues 25-185). Phyre2 modelling yielded low confidence (6.1%) with partial coverage (45%), reflecting structural complexity and unusual conformations. Ramachandran analysis showed significant deviations in allowed regions, suggesting structural features challenging current prediction methods.

4.5. Clinical and therapeutic implications

Given their early expression and roles in immune evasion, these HCMV proteins offer valuable targets for diagnostics and therapeutics, especially in immunocompromised patients and congenital infections.[1,2] Assays targeting specific transcripts or epitopes (e.g., via ELISA or qPCR) could improve detection sensitivity. US10's modulation of HLA-G makes it a particularly important biomarker for assessing NK cell-related immune dysfunction in vulnerable populations.[37] Therapeutically, US6 and US10 represent promising targets to counteract immune evasion. US6 inhibits antigen processing by blocking TAP-mediated peptide transport, while US10 internalizes MHC-I molecules through its tri-leucine motif, impacting both T cell and NK cell responses.[37] Structural analyses characterize US6 as hydrophilic and thermally stable, whereas US10's dual transmembrane helices and disordered regions highlight its structural divergence and poor alignment with MHC-I molecules (**Figure 3**; **Table 3**, **Table 5**), suggesting novel mechanisms suitable for targeted intervention.

5. Conclusion

This study outlines the roles of HCMV immune evasion proteins US2, US3, US6, US10, and US11 about MHC-I molecules. US2 and US3 show strong alignment and stability with MHC-I, suggesting their potential as vaccine or therapeutic candidates. US6 and US11 effectively disrupt antigen presentation but vary in stability and interaction. US10 features dual transmembrane helices and targets the non-classical MHC-I molecule HLA-G, indicating a specialized mechanism against NK cell responses. This underscores the need for further validation. Overall, these findings highlight US10 and other viral proteins as targets for diagnostics, therapies, and vaccines.

6. Source of Funding

No funds were available for this research.

7. Conflicts of Interest

The authors reported no potential conflict of interest.

8. Author Contributions

Conceptualization, A.S., M.S. and B.B.; execution of methodology, B.B. and A.B.; investigation, B.B., A.B. and A.S.; writing-original draft preparation, B.B., A.B., D.D. and M.S.; writing-review and editing, A.S., D.D. and M.S. All authors have read and agreed to the published version of the manuscript.

9. Acknowledgements

The authors would like to thank Department of Biosciences, JIS University for providing high speed internet facility and additional lab support.

References

1. Griffiths P, Reeves M. Pathogenesis of human cytomegalovirus in the immunocompromised host. *Nat Rev Microbiol*. 2021;19(12):759–73.
2. Novelli M, Natale F, Di Norcia A, Boiani A, Temofonte S, Calandriello F, et al. Early neurodevelopmental outcomes in children with asymptomatic congenital CMV infection. *Ital J Pediatr*. 2022;48(1):203.
3. Turriziani Colonna A, Buonsenso D, Pata D, Salerno G, Chieffo DPR, Romeo DM, et al. Long-term clinical, audiological, visual, neurocognitive and behavioral outcome in children with symptomatic and asymptomatic congenital cytomegalovirus infection treated with valganciclovir. *Front Med (Lausanne)*. 2020;7:268.
4. Uematsu M, Haginoya K, Kikuchi A, Hino-Fukuyo N, Ishii K, Shiihara T, et al. Asymptomatic congenital cytomegalovirus infection with neurological sequelae: a retrospective study using umbilical cord. *Brain Dev*. 2016;38(9):819–26.
5. Corey L, Wald A. Maternal and neonatal herpes simplex virus infections. *N Engl J Med*. 2009;361(14):1376–85.
6. Kimberlin DW. Herpes simplex virus infections of the newborn. *Semin Perinatol*. 2007;31(1):19–25.

7. Adler A, Nigro G. Prevention of maternal-fetal transmission of cytomegalovirus. *Clin Infect Dis*. 2013;57(Suppl 4):S189–92.
8. Zheng QY, Huynh KT, van Zuylen WJ, Craig ME, Rawlinson WD. Cytomegalovirus infection in day care centres: a systematic review and meta-analysis of prevalence of infection in children. *Rev Med Virol*. 2019;29(1):e2011.
9. Tai-Schmiedel J, Karniely S, Lau B, Ezra A, Eliyahu E, Nachshon A, et al. Human cytomegalovirus long noncoding RNA4.9 regulates viral DNA replication. *PLoS Pathog*. 2020;16(4):e1008390.
10. Zhang L, Yu J, Liu Z. MicroRNAs expressed by human cytomegalovirus. *Virol J.* 2020;17(1):34.
11. Avitabile E, Lombardi G, Campadelli-Fiume G. Herpes simplex virus glycoprotein K, but not its syncytial allele, inhibits cell-cell fusion mediated by the four fusogenic glycoproteins, gD, gB, gH, and gL. *J Virol.* 2003;77(12):6836–44.
12. Heming JD, Conway JF, Homa FL. Herpesvirus capsid assembly and DNA packaging. *Adv Anat Embryol Cell Biol.* 2017;223:119–42.
13. Nguyen CC, Kamil JP. Pathogen at the gates: Human cytomegalovirus entry and cell tropism. *Viruses.* 2018;10(12):704.
14. Kinzler ER, Compton T. Characterization of human cytomegalovirus glycoprotein-induced cell-cell fusion. *J Virol.* 2005;79(12):7827–37.
15. Hegde NR, Johnson DC. Human cytomegalovirus US2 causes similar effects on both major histocompatibility complex class I and II proteins in epithelial and glial cells. *J Virol.* 2003;77(17):9287–94.
16. Liu W, Zhao Y, Biegalke BJ. Analysis of human cytomegalovirus US3 gene products. *Virology.* 2002;299(1):49–59.
17. van der Wal FJ, Kikkert M, Wiertz E. The HCMV gene products US2 and US11 target MHC class I molecules for degradation in the cytosol. *Curr Top Microbiol Immunol.* 2002;269:37–55.
18. Gerke C, Bauersfeld L, Schirmeister I, Mireisz CNM, Oberhardt V, Mery L, et al. Multimodal HLA-I genotype regulation by human cytomegalovirus US10 and resulting surface patterning. *Elife*. 2024;13:e85560.
19. Gabor F, Jahn G, Sedmak DD, Sinzger C. In vivo downregulation of MHC class I molecules by HCMV occurs during all phases of viral replication but is not always complete. *Front Cell Infect Microbiol*. 2020;10:283.
20. Rombel IT, Sykes KF, Rayner S, Johnston SA. ORF-FINDER: a vector for high-throughput gene identification. *Gene*. 2002;282(1–2):33–41.
21. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845–58.
22. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539.
23. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784–8.
24. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol.* 2001;305(3):567–80.
25. Lobley A, Sadowski MI, Jones DT. pGenTHREADER and pDomTHREADER: New methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics.* 2009;25(14):1761–7.
26. Bittrich S, Bhikadiya C, Bi C, Chao H, Duarte JM, Dutta S, et al. RCSB Protein Data Bank: efficient searching and simultaneous access to one million computed structure models alongside the PDB structures enabled by architectural advances. *J Mol Biol*. 2023;435(14):167994.
27. Laskowski RA, Jabłońska J, Pravda L, Svobodová Vařeková R, Thornton JM. PDBsum: structural summaries of PDB entries. *Protein Sci*. 2018;27(1):129–34.
28. Wiltgen M. Algorithms for structure comparison and analysis: Homology modelling of proteins. In: *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier; 2019. p. 38–61.
29. Zhang Y, Skolnick J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 2005;33(7):2302–9.
30. Pawar SS, Rohane SH. Review on Discovery Studio: An important tool for molecular docking. *Asian J Res Chem.* 2021;14(2):86–8.
31. Tomar NR, Singh V, Marla SS, Chandra R, Kumar R, Kumar A. Molecular docking studies with rabies virus glycoprotein to design viral therapeutics. *Indian J Pharm Sci.* 2010;72(4):486–90.
32. McGuffin LJ, Jones DT. Improvement of the GenTHREADER method for genomic fold recognition. *Bioinformatics.* 2003;19(7):874–81.
33. Whitcomb SJ, Rakpenthai A, Brückner F, Fischer A, Parmar S, Erban A, et al. Cysteine and methionine biosynthetic enzymes have distinct effects on seed nutritional quality and on molecular phenotypes associated with accumulation of a methionine-rich seed storage protein in rice. *Front Plant Sci*. 2020;11:1118.
34. Lee S, Yoon J, Park B, Jun Y, Jin M, Sung HC, et al. Structural and functional dissection of human cytomegalovirus US3 in binding major histocompatibility complex class I molecules. *J Virol*. 2000;74(23):11262–9.
35. Gewurz BE, Wang EW, Tortorella D, Schust DJ, Ploegh HL. Human cytomegalovirus US2 endoplasmic reticulum-lumenal domain dictates association with major histocompatibility complex class I in a locus-specific manner. *J Virol.* 2001;75(11):5197–204.
36. Dugan GE, Hewitt EW. Structural and functional dissection of the human cytomegalovirus immune evasion protein US6. *J Virol*. 2008;82..
37. Park B, Spooner E, Houser BL, Strominger JL, Ploegh HL. The HCMV membrane glycoprotein US10 selectively targets HLA-G for degradation. *J Exp Med.* 2010;207(9):2033–41.
38. Kim S, Lee S, Shin J, Kim Y, Evnouchidou I, Kim D, et al. Human cytomegalovirus microRNA miR-US4-1 inhibits CD8(+) T cell responses by targeting the aminopeptidase ERAP1. *Nat Immunol*. 2011;12(10):984–91.
39. Ye L, Qian Y, Yu W, Guo G, Wang H, Xue X. Functional profile of human cytomegalovirus genes and their associated diseases: a review. *Front Microbiol*. 2020;11:2104.
40. Tokmakov AA, Kurotani A, Sato KI. Protein pI and intracellular localization. *Front Mol Biosci.* 2021;8:775736.
41. Bhattacharya M, Chatterjee S, Nag S, Dhama K, Chakraborty C. Designing, characterization, and immune stimulation of a novel multi-epitopic peptide-based potential vaccine candidate against monkeypox virus through screening its whole genome encoded proteins: an immunoinformatics approach. *Travel Med Infect Dis*. 2022;50:102481.
42. Loureiro J, Ploegh HL. Antigen presentation and the ubiquitin-proteasome system in host-pathogen interactions. *Adv Immunol.* 2006;92:225–305.
43. Wiertz EJ, Jones TR, Sun L, Bogyo M, Geuze HJ, Ploegh HL. The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell.* 1996;84(5):769–79.
44. Trevino SR, Schaefer S, Scholtz JM, Pace CN. Increasing protein conformational stability by optimizing beta-turn sequence. *J Mol Biol*. 2007;373(1):211–8.
45. Krieger F, Möglich A, Kiefhaber T. Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *J Am Chem Soc.* 2005;127(10):3346–52.

Cite this article: Biswas B, Bhaduri A, Saha M, Das D, Sarkar A. Predictive analysis of binding affinity among human cytomegalovirus (HCMV) proteins and class-I major histocompatibility complex (MHC-I) molecules. *Indian J Microbiol Res.* 2025;12(3):334–345.