

Immunomics Datasets and Tools: To Identify Potential Epitope Segments for Designing Chimeric Vaccine Candidate to Cervix Papilloma

Satyavani Kaliasurthi ^{1,2}, Gurudeeban Selvaraj ^{1,2}, Sathishkumar Chinnasamy ³,
Qiankun Wang ³, Asma Sindhoo Nangraj ³, William C. Cho ⁴, Keren Gu ^{1,2}
and Dong-Qing Wei ^{1,3,*}

¹ Center of Interdisciplinary Science-Computational Life Sciences, College of Food Science and Engineering, Henan University of Technology, Zhengzhou High-tech Industrial Development Zone, 100 Lianhua Street, Zhengzhou 450001, China; satyavani.mkk@haut.edu.cn (S.K.); gurudeeb99@haut.edu.cn (G.S.); gkr@haut.edu.cn (K.G.)

² College of Chemistry, Chemical Engineering and Environment, Henan University of Technology, Zhengzhou High-tech Industrial Development Zone, 100 Lianhua Street, Zhengzhou 450001, China

³ The State Key Laboratory of Microbial Metabolism, College of Life Sciences and Biotechnology, Shanghai Jiao Tong University, No. 800 Dongchuan Road, Minhang, Shanghai 200240, China; sathishimb@sjtu.edu.cn (S.C.); wangqiankun@sjtu.edu.cn (Q.W.); sindhoo_sind@yahoo.com (A.S.N.)

⁴ Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong; chocs@ha.org.hk

* Correspondence: dqwei@sjtu.edu.cn; Tel.: +86-021-3420-4717; Fax: +86-021-3420-5709

Received: 6 December 2018; Accepted: 12 February 2019; Published: 15 February 2019



Abstract: Immunomics tools and databases play an important role in the designing of prophylactic or therapeutic vaccines against pathogenic bacteria and viruses. Therefore, we aimed to illustrate the different immunological databases and web servers used to design a chimeric vaccine candidate against human cervix papilloma. Initially, cellular immunity inducing major histocompatibility complex class I and II epitopes from L2 protein of papilloma 58 strain were predicted using the IEDB, NetMHC, and Tepi tools. Then, the overlapped segments from the above analysis were used to calculate efficiency on interferon-gamma and humoral immunity production. In addition, the allergenicity, antigenicity, cross-reactivity with human proteomes, and epitope conservancy of elite segments were determined. The chimeric vaccine candidate (SGD58) was constructed with two different overlapped peptide segments (23–36) and (29–42), adjuvants (flagellin and RS09), two Th epitopes, and amino acid linkers. The results of homology modeling demonstrated that SGD58 have 88.6% of favored regions based on Ramachandran plot. Protein–protein docking with Swarm Dock reveals SGD58 with receptor complex have -54.74 kcal/mol of binding energy with more than 20 interacting residues. Docked complex are stable in 100ns of molecular dynamic simulation. Further, coding sequences of SGD58 also show elevated gene expression in *E. coli*. In conclusion, SGD58 may prompt vaccine against cervix papilloma. This study provides insight of vaccine design against different pathogenic microbes as well.

Dataset: <http://doi.org/10.5281/zenodo.1997695>

Keywords: cellular immunity; codon frequency distribution; HPV58; minor capsid protein; TLR agonist; SGD58; prophylaxis

1. Summary

Currently, viral infection contributes to about 20% of the global burden of human cancer. Among the broad viral spectrum, human papillomavirus (HPV) is reported in about 5% of all human cancers,

specifically infection associated with the cervix with 250,000 mortalities every year [1]. HPV with its double-stranded DNA contains a nonenvelope small virus, which infects a region of the cutaneous epithelial membrane (skin or integumentary system), or the mucous membrane (i.e., coated as an internal line in hollow spaced organs like the mouth, reproductive organ, urinary tract, or rectum) in the host system [2]. The genomic relationship between different cancer types has demonstrated that more than 99% of cervical cancer patients are infected with 15 different types of α -clade HPV, defined as “high-risk” or “oncogenic” genital HPVs. The α -clade HPV (6 and 11) causes genital warts while the remaining strains of HPV are related to the risk of cervical cancer. HPV infection is attributed to more than 50% of oropharyngeal and anogenital cancers [3]. Generally, the human immune system can clear the pathogenic infection caused by HPV within two years, but this also depends on the efficiency of an individual’s immune system and the invading type of HPV. However, in the case of a very weak immune system, it fails to remove the invading high-risk HPVs (hrHPV) that may lead to the development of cervical cancer [4,5]. hrHPV infections are responsible for causing more than 99% of precancerous cervical intraepithelial neoplasias (CIN) and invasive cervical cancers (ICC) [6–8]. In China, HPV-mediated cervical cancer is a substantial public health issue, with 1 million new cervical cancer incidences and 30,000 mortalities registered every year [9,10]. In 2018, clinical, epidemiological, and clinicopathological studies reported HPV58 to be the second or third most predominant genotypes in precancerous CIN includes mild dysplasia (CIN I), mild to moderate dysplasia (CIN II), severe dysplasia to cancer (CIN III), and ICC lesions. Higher grades of squamous intraepithelial, or cell carcinoma, and adenocarcinoma of HPV positive patients were diagnosed in different geographical regions of China [11]. Seven provinces of China that have reported hrHPV-mediated cervical cancer incidences, include Guangdong, Liaocheng, Shanghai, Wenzhou, Wuhan, Southwestern China, and Western China [11–17]. Zhang et al. [18] reported that the HPV16 (6.4%) and HPV58 (5.3%) genotypes were predominantly found in males who had recently involved in sex, in Shanghai.

Cervarix[®], Gardasil[®], and Gardasil 9[®] are the three noninfectious prophylactic Food and Drug Administration (FDA)-approved HPV licensed subunit vaccines in active usage. These vaccines were developed from the major capsid L1 virus-like particles (VLPs) using recombinant DNA technology. Cervarix is a bivalent vaccine based on the Baculovirus fermentation and it provides ~70% protection to HPV (16 and 18)-mediated cervical cancer but not to genital warts [19]. Gardasil is a quadrivalent HPV (6, 11, 16, and 18) vaccine based on yeast fermentation technology. It is efficiently used for the prevention of genital warts and gives ~70% protection for cervical cancer [20]. In 2009, the FDA approved a nine-valent Gardasil 9[®] that provides protection to HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58. It has been used for both males and females in the age groups of 9–15 and 9–26 [21]. The new nine-valent vaccine exhibited a positive outcome in high-grade lesions in the absence of HPV (18 and 16) infections [22]. In October 2018, FDA extended the use of Gardasil 9 to the age group of 27–45 among both the sexes. In addition, the L1 VLP (absence of viral genomic materials)-mediated vaccine production in the eukaryotic (ex. Baculovirus) host system is a complex and tedious process [23,24]. The main limitations of currently available prophylactic vaccines are strain specific, not therapeutic for patients already infected with HPV22, require multiple dosages, and is expensive [25,26]. In addition, the effective straightforward delivery of HPV vaccines can enhance the immunogenic potential against HPVs.

The implementation of L2 minor capsid protein is a potential alternative in the HPV prophylactic vaccine production. Since the N-terminal region of the L2 protein is highly conserved in low-risk HPV (6 and 11) and 13 different hrHPVs, it is contrasted with the type-specific protection of L1 prophylactic VLPs [27]. The single copy of L2 protein (~473 amino acids (AA)) is present in each L1 capsomere, resulting in 72 copies per virion [28]. Incidentally, L2 protein plays a vital role in L1 assembly into the VLPs and enhances the encapsidation of double-stranded ~8kb circular viral genome [29]. Moreover, the full-length or polypeptides (1–8 or 11–200 AA in length) of L2 protein enhance the production of neutralizing antibodies in vaccinated experimental models including mice, cattle, and rabbit [30–32].

To date, no L2 VLP-derived prophylactic vaccine has been approved in clinical trials due to their limitation of weak immunogenicity, which imitates the incapability of multimerizing into the VLPs.

With this information, in this study, we aimed to design the novel chimeric vaccine from the N-terminal region of the L2 sequence of HPV58 targets to hrHPVs. Immunomics tools and databases play an important role in the designing of prophylactic or therapeutic vaccines against pathogenic bacteria and viruses [33–35]. Immunomic tools include immune epitope database (IEDB) and NetMHCv4.0, Tepitool, CTLPred, PAComplex, IFNepitope, ABCPred, AllerTOP, AllergenFPv 1.0, ANTIGENpro, program of protein information resource (PIR), and epitopes conservancy were implemented to discover the overlapped epitope segment to induce B cell and T cell immunity. Then, the chimeric vaccine (SGD58) was constructed using overlapped epitope segments, TLR adjuvants, Th epitopes, and amino acid linkers. The physiochemical and immunological properties of the chimeric vaccine was validated using ProtParam, SolPro, VaxiJen, and ANTIGENpro tools. In addition, homology modeling using iterative threading ASSEmblY refinement (I-TASSER), structural refinement (GalaxyRefine and 3DRefine), and structural validation (protein structure analysis (ProSA), Ramachandran plot, and ERRAT) were performed to obtain the best three-dimensional (3D) model of the chimeric vaccine and target TLR5 receptor. Then, the interaction of the chimeric vaccine with TLR5 and stability of this complex were determined through PP docking and molecular dynamic (MD) simulation. Moreover, the virtual cloning and gene expression of the chimeric vaccine in *E. coli* were analyzed to obtain a low-cost HPV vaccine. All the necessary supporting information for SGD58 design illustrated in the study and the original results was reported in our previous study [36].

2. Data Description

The specifications of the data description namely subject area, types of data, method of acquiring data, format, experimental factors, source and accessibility of data for selection of potential epitope segments and designing of chimeric vaccine was illustrated in Table 1.

Table 1. Specifications table.

Subject Area	Immuno-Informatics, Structural Vaccinology
More specific subject area	Chimeric vaccine for cervix papilloma
Type of data	Image, Excel, doc
How data was acquired	Online tools based on manual selective algorithms
Data format	Raw and Manual Annotations
Experimental factors	Epitopes, antigenicity, allergenicity, and modeled structures
Experimental features	The epitopes were identified from the proteome of papilloma virus. It has antigenic, non-allergenic and INF inducing properties. The elite epitopes with designed vaccine structure was modeled and validated.
Data source location	Public databases and online tools based on manual selective algorithms
Data accessibility	http://doi.org/10.5281/zenodo.1997695

The overlapped epitope segments obtained from the major histocompatibility complex class I (MHC-I) prediction were compared with the results of both CTLPred and PAComplex servers. Furthermore, the shared epitope segments obtained from the CTLPred and PAComplex were used for epitope selection, and vaccine design as shown in Table 2.

The lowest percentile rank with strong binding affinity epitope segments with human MHC-II alleles, such as the DQB1-, DRB1-, and DPB1-restricted epitopes, were obtained using IEDB consensus and Tepitool servers. The overlapping promiscuous epitope segments from the above prediction (Table 3) were selected and evaluated for their INF- γ production ability. The overlapped INF- γ producing CD4+ (MHC-II) epitope segments are as given in Table 3. Therefore, the shared MHC-II epitope segments could produce INF- γ against viral infection. Interestingly, the above-obtained overlapped CD4+ epitopes shared the CD8+ epitope segments. In addition, Table 3 illustrates overlapped B cell epitopes predicted through the ABCPred tool.

Table 2. Overlapped epitope segments of major histocompatibility complex class I (MHC-I), CTL, and TCR from the N-terminal region of HPV58 were predicted using different servers.

MHC-I		CTL	TCR–Peptide/Peptide–MHC Interfaces
IEDB ^a	NetMHC 4.0 ^b	CTLPred ^c	PAComplex ^d
23–36	23–35	16–24	19–27
30–43	29–42	40–48	38–46
10–23	9–22	16–244–12	4–12
29–42	28–41	38–46	38–46

MHC-I overlapped epitope segments prediction by using different tools as ^a IEDB consensus and ^b NetMHC4.0; CTL epitopes prediction by using ^c CTLPred; TCR–peptide and peptide–MHC interface predicted by ^d PAComplex.

Table 3. The overlapped epitope segments of MHC-II, INF-gamma-producing, and B cell epitopes N-terminal region of HPV58 by using different servers.

MHC-II		INF- γ Producing Epitopes	B Cell Epitopes
^a IEDB Consensus	^b Tepitool	^c INFepitope	^d ABCPred
23–36	23–36	0.51	
23–37	23–37	0.53	26–41
29–43	29–43	+1	
30–44	30–44	+1	26–41
7–21	7–21	+1	
6–20	6–20	+1	7–22
29–43	29–43	+1	
28–42	28–42	+1	33–48

MHC-II overlapped epitope segments prediction by using different tools as ^a IEDB consensus and ^b Tepitool; INF- γ production of the overlapped epitope segments by using ^c INFepitope; Overlapped B cell linear epitope segments prediction by using ^d ABCPred.

Table 4 gives the comprehensive analysis of overlapped epitope ($\geq 30\%$), positions, subsequences identity, and hrHPV. The conservation of selected epitopes has cross-protection to the 15 hrHPV as shown in Figure 1.

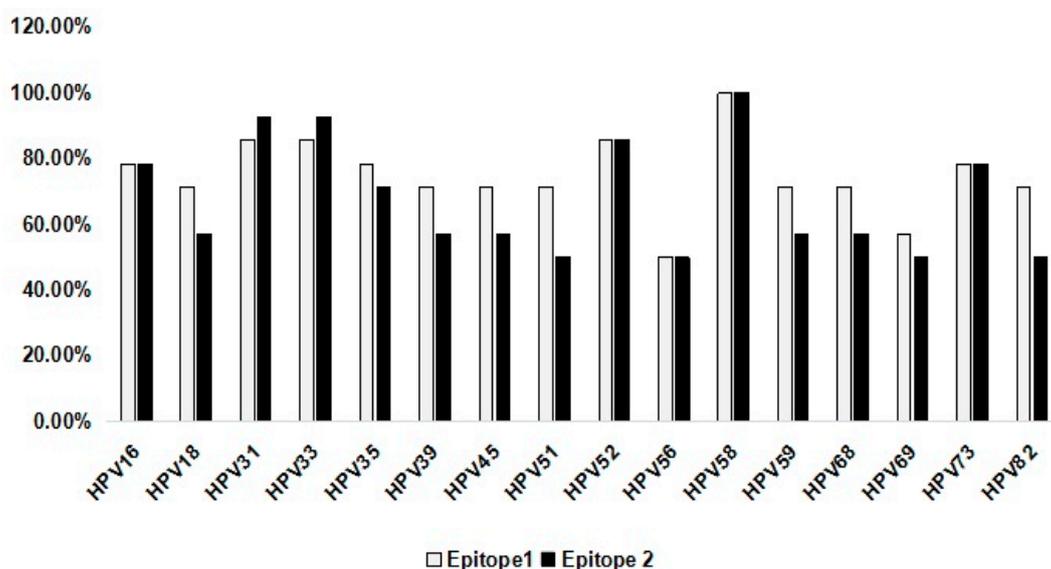
**Figure 1.** Conservation of two overlapped epitope segments (Epitope 1: KVEGTTIADQILRY; Epitope 2: IADQILRYGSLGVF) in fifteen hrHPV strains. Fifteen hrHPV strains conserved epitope segments were represented in X-axis. Percentage (%) of epitope conservancy among the hrHPV strains were showed in Y-axis.

Table 4. Conservation across- high-risk human papillomavirus (hrHPV) strains by the overlapped HPV58 epitope segments.

No.	Epitopes	Positions	Protein Sub Sequences	Identity (%)	Name of the Strain
1	CKASGTCPPDVIPK	21–34	CKASGTCPPDVIPK	100.00	HPV52
2		21–34	CKASGTCPPDVIPK	100.00	HPV58
3		21–34	CKATGTCPPDVIPK	92.86	HPV33
4		22–35	CKAAGTCPPDVIPK	92.86	HPV35
5		21–34	CKAAGTCPPDVIPK	92.86	HPV69
6		21–34	CKAAGTCPPDVIPK	92.86	HPV82
7		21–34	CKQSGTCPPDVVVK	85.71	HPV18
8		22–35	CKAAGTCPSDVIPK	85.71	HPV31
9		21–34	CKQSGTCPPDVINK	85.71	HPV45
10		23–36	CKQAGTCPPDVIPK	85.71	HPV73
11		22–35	CKQAGTCPPDIIPK	78.57	HPV16
12		21–34	CKQSGTCPPDVVDK	78.57	HPV39
13		21–34	CKAAGTCPPDVVVK	78.57	HPV51
14		21–34	CKQSGTCPSDVINK	78.57	HPV68
15		21–34	CKLSGTCPEDVVVK	71.43	HPV56
16		21–34	CKQAGTCPSDVINK	71.43	HPV59
1	KVEGTTIADQILRY	34–47	KVEGTTIADQILRY	100.00	HPV58
2		35–48	KIEHTTIADQILRY	85.71	HPV31
3		34–47	KVEGSTIADQILKY	85.71	HPV33
4		34–47	KVEGTTIADQLLY	85.71	HPV52
5		35–48	KVEGKTIAEQILQY	78.57	HPV16
6		35–48	KVEGNTVADQILKY	78.57	HPV35
7		36–49	KVEGSTIADNILKY	78.57	HPV73
8		34–47	KVEGTTLADKILQW	71.43	HPV18
9		34–47	KVEGTTLADKILQW	71.43	HPV39
10		34–47	KVEGTTLADKILQW	71.43	HPV45
11		34–47	KVEGTTLADKILQW	71.43	HPV51
12		34–47	KVEGTTLADKILQW	71.43	HPV59
13		34–47	KVEGTTLADKILQW	71.43	HPV68
14		34–47	KVEGTTLADKILQW	71.43	HPV82
15		34–47	KIEGSTLADKILQW	57.14	HPV69
16		34–47	KIEQKTWADRILQW	50.00	HPV56
1	IADQILRYGSLGVF	40–53	IADQILRYGSLGVF	100.00	HPV58
2		41–54	IADQILRYGSMGVF	92.86	HPV31
3		40–53	IADQILKYGSLGVF	92.86	HPV33
4		40–53	IADQLLYGSLGVF	85.71	HPV52
5		41–54	IAEQILQYGSMGVF	78.57	HPV16
6		42–55	IADNILKYGSIGVF	78.57	HPV73
7		41–54	VADQILKYGSMAVFF	71.43	HPV35
8		40–53	LADKILQWSSLGIF	57.14	HPV18
9		40–53	LADKILQWTSLGIF	57.14	HPV39
10		40–53	LADKILQWSSLGIF	57.14	HPV45
11		40–53	LADKILQWTSLGIF	57.14	HPV59
12		40–53	LADKILQWTSLGIF	57.14	HPV68
13		40–53	LADKILQWSSLGIF	50.00	HPV51
14		40–53	WADRILQWGSIFTY	50.00	HPV56
15		40–53	LADKILQWSSLGIF	50.00	HPV69
16		40–53	LADKILQWSSLGIF	50.00	HPV82
1	ADQILRYGSLGVFF	41–54	ADQILRYGSLGVFF	100.00	HPV58
2		42–55	ADQILRYGSMGVFF	92.86	HPV31
3		41–54	ADQILKYGSLGVFF	92.86	HPV33
4		41–54	ADQLLYGSLGVFF	85.71	HPV52
5		42–55	AEQILQYGSMGVFF	78.57	HPV16
6		42–55	ADQILKYGSMAVFF	78.57	HPV35
7		43–56	ADNILKYGSIGVFF	78.57	HPV73
8		41–54	ADKILQWSSLGIFL	57.14	HPV18
9		41–54	ADKILQWTSLGIFL	57.14	HPV39
10		41–54	ADKILQWSSLGIFL	57.14	HPV45
11		41–54	ADRILQWGSIFTYF	57.14	HPV56
12		41–54	ADKILQWTSLGIFL	57.14	HPV59
13		41–54	ADKILQWTSLGIFL	57.14	HPV68
14		41–54	ADKILQWSSLGIFL	50.00	HPV51
15		41–54	ADKILQWSSLGIFL	50.00	HPV69
16		41–54	ADKILQWSSLGIFL	50.00	HPV82

Residues that are different from their corresponding residues in the reference sequence are highlighted in bold with gray shadow. Identity indicates the number (%) of residues in the homologous sequences that are identical to the corresponding residues in the reference sequence.

The refined 3D structure obtained in the above section underwent quality improvement using three potential tools: ProSA-web, RAMPAGE, and the ERRAT. The z-score (ProSA), overall quality factor (ERRAT), and favored, allowed, and the outlier region (RAMPAGE) of the validated 3D structure of SGD58 are given in Table 5 and TLR5 in Table 6.

Table 5. Validation of 3D structures of the designed SGD58 obtained by the iterative threading ASSEmbly refinement (I-TASSER) and its refinement by the Galaxy Refine (named as I-T Gal) and 3Drefine (named as I-T 3DR).

Model	ProSA	ERRAT	RAMPAGE		
	z-Score	Overall Quality Factor	Favored Region	Allowed Region	Outlier Region
I-TASSER	−5.76	83.2258	249 (78.8%)	44 (13.9%)	23 (7.3%)
I-T Gal1	−5.54	75.6494	282 (89.2%)	23 (7.3%)	11 (3.5%)
I-T Gal2	−5.55	75.1613	281 (88.9%)	22 (7.0%)	13 (4.1%)
I-T Gal3	−5.77	88.889	280 (88.6%)	24 (7.6%)	12 (3.8%)
I-T Gal4	−5.63	79.8701	279 (88.3%)	24 (7.6%)	13 (4.1%)
I-T Gal5	−5.75	77.7419	280 (88.6%)	24 (7.6%)	12 (3.8%)
I-T 3DR1	−5.72	86.8056	261 (82.6%)	35 (11.1%)	20 (6.3%)
I-T 3DR2	−5.72	88.8114	259 (82.0%)	32 (10.1%)	25 (7.9%)
I-T 3DR3	−5.87	88.8112	258 (81.6%)	35 (11.1%)	23 (7.3%)
I-T 3DR4	−5.86	88.8112	259 (82.0%)	30 (9.5%)	27 (8.5%)
I-T 3DR5	−5.89	80.9677	259 (82.0%)	30 (9.5%)	27 (8.5%)

The I-T Gal.3 structure was chosen as the most appropriate model, which is shown in bold.

Table 6. Validation of 3D structures of the TLR5 obtained by the I-TASSER and its refinement by the GalaxyRefine (named as I-T Gal) and 3Drefine (named as I-T 3DR).

Model	ProSA	ERRAT	RAMPAGE		
	z-Score	Overall Quality Factor	Favored Region	Allowed Region	Outlier Region
I-TASSER	−5.93	79.7619	635 (74.2%)	169 (19.7%)	52 (6.1%)
I-T Gal1	−6.52	68.9781	779 (91.0%)	71 (8.3%)	6 (0.7%)
I-T Gal2	−6.35	73.7864	778 (90.9%)	70 (8.2%)	8 (0.9%)
I-T Gal3	−6.64	73.3414	783 (91.5%)	65 (7.6%)	8 (0.9%)
I-T Gal4	−6.65	70.3163	785 (91.7%)	63 (7.4%)	8 (0.9%)
I-T Gal5	−6.61	72.6176	782 (91.4%)	68 (7.9%)	6 (0.7%)
I-T 3DR1	−6.47	85.967	699 (81.7%)	118 (13.8%)	39 (4.6%)
I-T 3DR2	−6.52	86.3208	708 (82.7%)	109 (12.7%)	39 (4.6%)
I-T 3DR3	−6.53	86.6745	714 (83.4%)	102 (11.9%)	40 (4.7%)
I-T 3DR4	−6.63	86.4387	713 (83.3%)	102 (11.9%)	41 (4.8%)
I-T 3DR5	−6.77	87.6179	712 (83.2%)	103 (12.0%)	41 (4.8%)

The I-T 3DR5 structure was chosen as the most appropriate model, which is shown in bold.

From overall comparison of the results, model 3 of GalaxyRefine of SGD58 (Figure 2a) and model 5 of 3Drefine of TLR5 (Figure 2b) using UCSF Chimera were selected for further analysis.

Table 7 explains the respective amino acid, residue with contact number, propensity, and Discotope score of the predicted B cell epitopes.

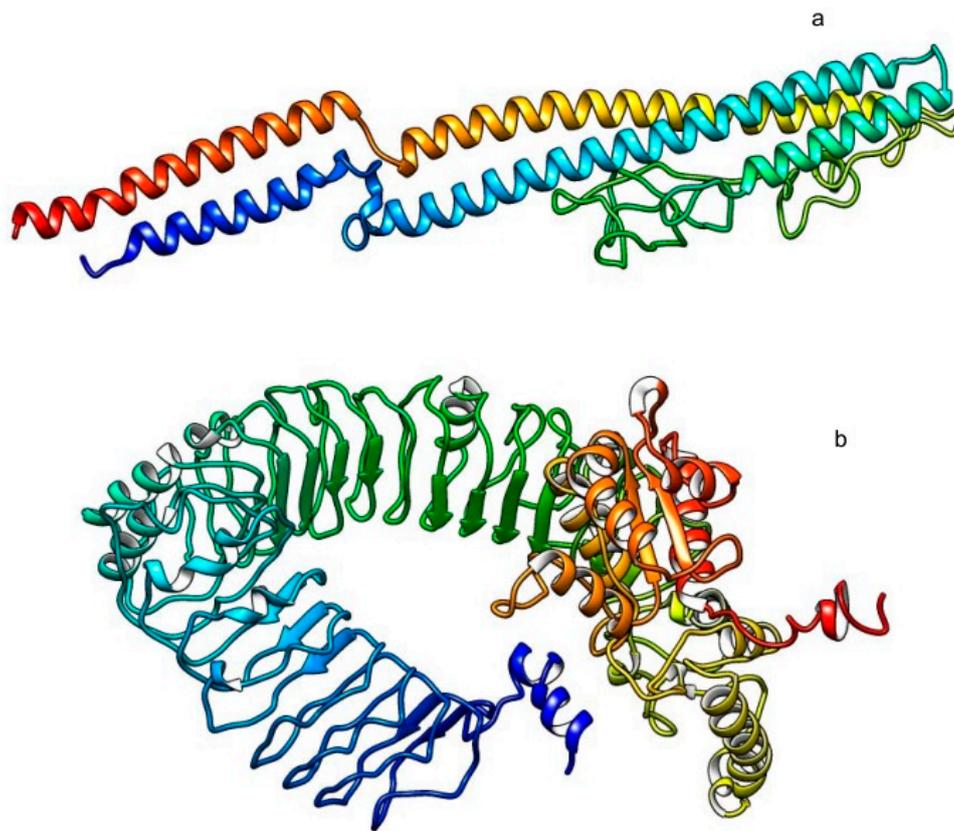


Figure 2. Refined 3D structure of the SGD58 and mTLR5 by using UCSF Chimera. (a) The 3D structure of the SGD58 was obtained through homology modeling by using i-TASSER, and then the best proposed model was refined by using GalaxyRefine. (b) The 3D structure of the mouse TLR5 was obtained through homology modeling by using i-TASSER, and then the best proposed model was refined by using 3Drefine.

Table 7. Dis-continuous B cell epitopes identified in the refined 3D structure of designed vaccine constructs of HPV58 by using DiscoTope 2.0.

S.No.	Residue Number	Amino Acid	Contact Number	Propensity Score	DiscoTope Score
1	12	ASN	5	−3.272	−3.471
2	25	ILE	7	−3.159	−3.601
3	37	ALA	0	−3.037	−2.688
4	38	LYS	5	−2.621	−2.895
5	41	ALA	0	−3.549	−3.141
6	42	ALA	3	−3.414	−3.366
7	55	LYS	6	−3.291	−3.602
8	99	THR	0	−1.665	−1.474
9	101	SER	3	−1.96	−2.079
10	103	SER	0	−2.842	−2.515
11	107	SER	6	−2.739	−3.114
12	130	GLY	5	−2.944	−3.181
13	265	GLY	8	−2.67	−3.283
14	266	ASN	5	−0.617	−1.121
15	269	THR	6	−2.481	−2.886
16	270	ASN	7	−2.764	−3.251
17	284	ALA	1	−3.567	−3.272
18	288	SER	5	−3.336	−3.528

SwarmDock modeling demonstrated the list of clusters with SGD58 and TLR5 complex. The clusters are ranked based on interacting residues between the complexes with binding energies. The input TLR5 receptor contains 858 amino acids and SGD58 contains 2923 amino acids. The human TLR5

sequence contains 21 different leucine-rich repeats (LRR) segments with 443 amino acids. Flagellin contains two D1/D0 TLR binding domains in the N and C terminals of the sequence. After restraining the residues in TLR5 (45–68, 71–93, 95–117, 120–143, 146–166, 171–192, 197–211, 214–229, 234–235, 260–284, 289–301, 313–334, 337–355, 385–401, 412–431, 449–470, 474–495, 503–524, 527–546, 549–567, and 579–631) and in SGD58 (5–143 and 419–504) were chosen for docking. The initial preprocessing of receptor and ligand took 20 h and 10 min, and restrained docking took 7 h and 40 min in SwarmDock web server. In total, 352 docked complexes were obtained from the docking results. The result demonstrated that each best model interacts with the LRR segments. The binding energies of the top 10 models are -54.74 , -49.12 , -49.07 , -46.84 , -42.57 , -41.31 , -40.59 , -40.53 , -38.78 , and -33.51 , respectively. The energy function is calculated based on van der Waals and a Coulombic term on individual interacting atoms in receptor and ligand. In addition, the highest percentage of interacting residues (88.87%) and the lowest percentage of interacting residues (57.71%) were observed in models 1 and 9 of the SGD58 and TLR5 complex, respectively. Table 8 provides the list of top ten clusters, binding residues, and interaction energies.

Figure 3 illustrates that potential energy (PE), temperature, total energy (TE), and pressure of SGD58 was stable during the simulation period. The average TE of SGD58 is $-7,206,525.282$ with a standard deviation of 4373.407. In addition, the average PE of SGD58 is $-8,984,582.127$ with a standard deviation of 3472.905. PE and TE attained equilibrium at a temperature of 300 K. The result of the radius of gyration (Rg) analysis is shown in Figure 4. The simultaneous changes in Rg plots of the SGD58 and (Figure 4a) and complex with TLR5 (Figure 4b) indicate that the substantial nature of the complex frequently increases. Rg plots compression of SGD58 with TLR5 and are similar to the RMSD parameter, which indicates the effort of SGD58 to reach internal configuration in TLR5

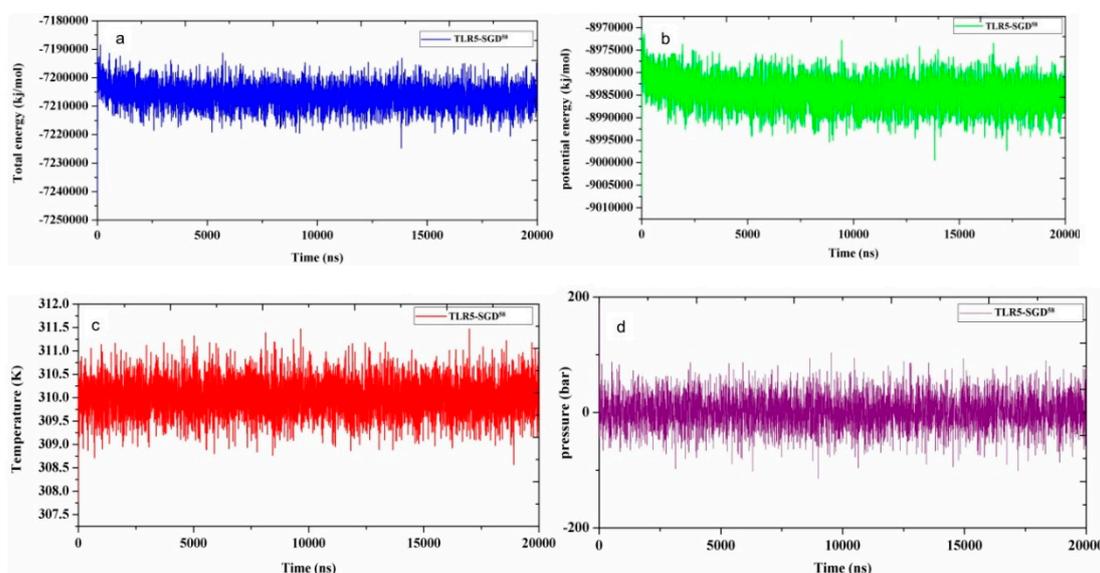


Figure 3. (a) Total energy, (b) potential energy, (c) temperature, and (d) pressure plots of molecular dynamic (MD) simulation for mTLR5- SGD58 complex in simulations of 20 ns.

The maximal protein expression of this optimized coding sequence in the host (*E. coli*) was analyzed by the GenScript's Optimum Gene™ codon optimization tool. Figure 5 illustrates the CAI, GC, and CFD of the gene transcript. The gene (reverse translated coding sequence of the vaccine construct) having ideal CAI value of 1.00 (>0.8) is more suitable for the above expression (*E. coli*) in the host organism. Moreover, 59.92% of ideal GC content is presented in the gene (between 30% and 70%). However, the outside of these peak ranges would severely inhibit the transcriptional and translational efficiency of the gene products. The CFD value of the gene is 100%, representing their highest codon frequency distribution in the preferred expression organism.

Table 8. List of top ten clusters, binding residues, and interaction energy of SGD58 and TLR5 complex.

Model Number	Reference Model Number	Starting Amino Acid in TLR5	Binding Energy	Number of Clusters	Total Number of Contacts between the SGD58 and TLR5 Complex	Number of Contact Residues in Receptor	Number of Contact Residues in Ligand	Percentage of Interacting Residues between the SGD58 and TLR5 Complex
Model 1	64a.pdb	112	−54.74	3	663	538	208	88.87
Model 2	63c.pdb	111	−49.12	1	1172	969	682	70.98
Model 3	57d.pdb	104	−49.07	1	731	615	361	74.89
Model 4	84d.pdb	183	−46.84	7	802	693	403	73.17
Model 5	56b.pdb	103	−42.57	7	687	459	386	81.30
Model 6	57c.pdb	104	−41.31	1	634	543	387	68.17
Model 7	72c.pdb	121	−40.59	1	668	603	506	60.23
Model 8	49c.pdb	96	−40.53	1	592	511	414	64.00
Model 9	46d.pdb	92	−38.78	2	793	604	770	57.71
Model 10	68b.pdb	115	−33.51	1	508	508	267	65.55

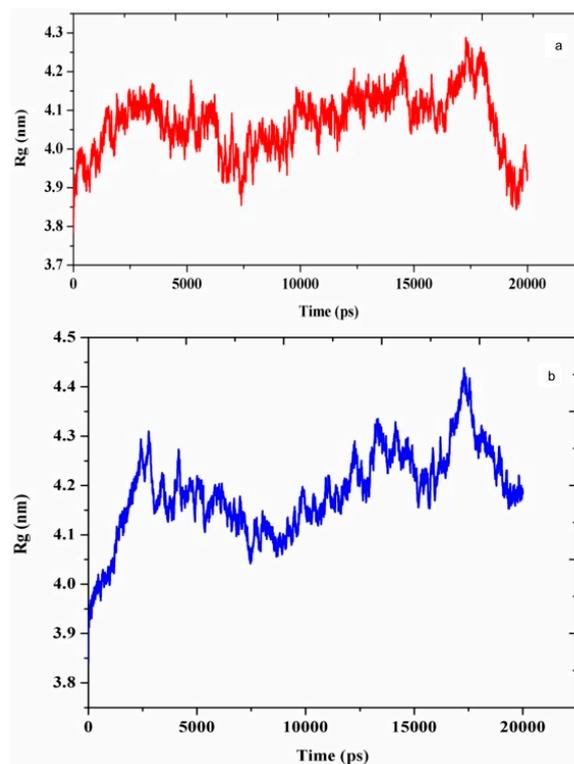


Figure 4. (a) Rg plot of vaccine molecule and (b) Rg plot of complex vaccine molecule.

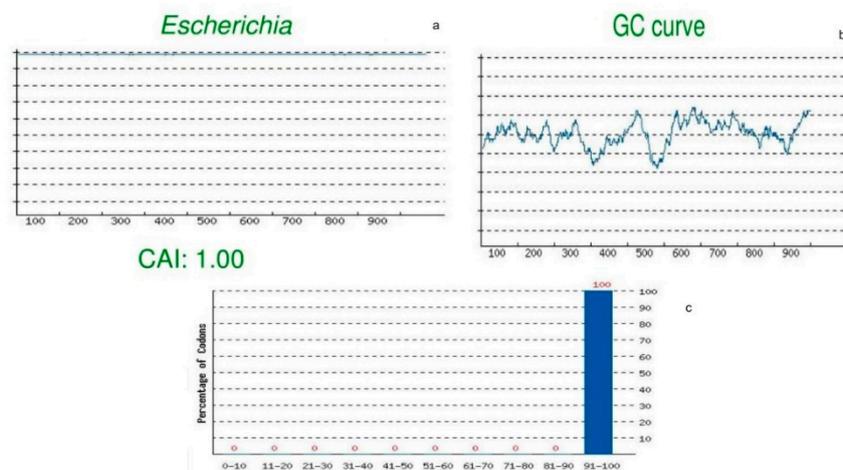


Figure 5. Codon optimization and *in silico* cloning of the gene. (a) The gene (reverse translated coding sequence of the vaccine construct) having ideal CAI value of 1.00 (>0.8), which is more suitable for higher expression in the *E. coli* host organism. (b) The percentage of GC content in the gene is 59.92%, which is in the ideal range of GC content (between 30 to 70%). (c) CFD value of the gene is 100%. The value of 100 is set for the codon with the highest usage frequency for a given amino acid in the desired expression organism.

We conclude SGD58 may prompt vaccine against cervix papilloma. This study provides insight of vaccine design against different pathogenic microbes as well.

3. Methods

The L2 protein of HPV58 (Accession No.: P26538), flagellin of *Salmonella enterica serovar Dublin* (Accession No.: Q06971), and human TLR5 (Accession No.: O60602) sequences were obtained from the

Swiss-Prot reviewed universal protein knowledgebase (UniProt) [37]. The designed chimeric vaccine was named SGD58, using the name of the first and principal authors along with the strain number. Two servers—IEDB and NetMHCv4.0—have been exploited for identification of major histocompatibility complex class I (MHC-I) binding epitopes from the N-terminal region of the L2 sequence. Specific human MHC-I alleles such as human leukocyte antigen (HLA)-A* (01:01, 02:01, 02:07, 11:01, 24:01), HLA-B* (46:01, 58:01), and HLA-C* (07:02, 12:03) were abundantly diagnosed in different regions of China, including Guizhou, Henan, Taihu River Basin, Tibetan, Yunnan, Wenzhou, and Wuhan. These 11 alleles were selected for epitope prediction [38–44]. IEDB [45] is a freely available analysis resource with specified algorithms for the identification and determination of immunogenic epitopes. A consensus method was implemented to predict the MHC-I binding epitopes and its production pathway [46]. In this consensus method, three algorithms, including neural network (artificial), matrix method (stabilized), and peptide libraries (combinatorial), were combined to predict the promising CTL epitope segments. The epitopes involve proteasomal cleavage (pCle), transporter associated with antigen processing (TAP), and MHC-I binding pathway. The lowest percentile rank (<10%) indicated the good binding efficiency of epitopes with the restricted alleles. NetMHCv4.0 is another potential tool implemented to find MHC-I binding peptides with the best Pearson’s correlation coefficient (PCC) of 0.895, based on the combined neural network. The strong and weak binding peptides were predicted based on the thresholds of <0.5 and <2, respectively [47]. The CTLPred tool is a direct method for the prediction of CTL epitope segments instead of MHC binders. The prominent combined approaches were implemented to find the epitopes, based on both the artificial neural networks (ANN) trained by the Stuttgart neural network simulator (SNNS) and support vector machine (SVM) methods. The combined methods demonstrate a higher level of accuracy (75.8 %) compared with other individual methods of prediction such as ANN (72.2%) and SVM (75.4%). The default cutoff scores of 0.51 of ANN and 0.36 of SVM were used to find the epitopes or nonpeptides at which the sensitivity and specificity of the predictions are almost similar [48]. A web server PACComplex provides access to examine and visualize the TCR–peptide and peptide–MHC interface (pMHC), respectively. For a given viral protein query sequence, the joint Z-value obtained with threshold 2.5. Moreover, it allows the selection of only limited allotypes of MHC class I such as HLA-A0201, HLA-B (0801, 3501, 3508, and 4405), and HLA-E, respectively. The Z-value was calculated using the following formula.

$$Jz = Z_{MHC} \times Z_{TCR} \quad (1)$$

where, Z_{MHC} and Z_{TCR} are the score of a TCR–pMHC complex, calculated by $(E-\mu)/\sigma$. E denotes interaction score, μ denotes mean, and σ denotes standard deviation from 10,000 random interfaces [49]. MHC-II alleles include DQB1*(03:01, 03:03, 06:01), DRB1*(07, 09, 14:01:01, 15:01, 15:07:01), and DPB1* (05:01,05:02:01), specific to Henan, Taihu River Basin, Tibetan, Yunnan, Wenzhou, and Wuhan provinces of China, which have been selected for epitope prediction [38–44]. The IEDB consensus approaches were used to predict MHC-II binding epitope segments using neural network-based alignment, stabilized matrix methods-based alignment, and combinatorial library-based algorithms [50]. The peptides with the lowest percentile rank were considered as the higher binding affinity. Tepitool is a tool from IEDB analysis resources, which provides accession to the prediction of both class I and II binders. The peptides which show the lowest percentile rank ($IC_{50} < \text{or} = 500 \text{ nM}$) are potentially considered as higher affinity binding peptides [51]. INFepitope is a potential server useful for the prediction and design of INF- γ inducing epitopes. INF- γ inducing epitopes were identified based on motif-based SVM or hybrid algorithms. The hybrid method using residue or dipeptide composition shows 81.39% accuracy [52]. ABCPred is used to predict linear B cell epitopes. It provides 65.93% of accuracy with the involvement of the recurrent neural network (RNN) algorithm. It consists of 700 B cell and non-B cell epitope segment datasets each with a length of 20 amino acids [53]. AllerTOP is the first proper alignment-free allergenicity server. In this, five machine learning methods such as partial least squares, logistic regression, decision tree, naive Bayes, or k nearest neighbors ($kNN = 1$) were implemented to find the allergen. It shows 88.7%, 90.7%, and 86.7% for accuracy, specificity, and

sensitivity, respectively [54]. AllergenFPv 1.0 is another essential tool for the allergenicity prediction based on novel descriptor fingerprint approaches. Twenty naturally existing amino acids in the protein sequences were classified into five descriptors (E) such as E1 (hydrophobicity), E2 (size), E3 (helix-forming propensity), E4 (relative abundance of amino acids), and E5 (β -strand forming propensity). Based on this, the strings were transformed into normal vectors by ACC transformation to find the allergen protein. It exhibits accuracy (87%), specificity (89%), and sensitivity (86%) [55]. ANTIGENpro is the potential alignment-free and sequence-based antigenicity prediction server with 79% accuracy and area under curve (AUC) of 0.89. It shows results based on amino acid composition and random-forest algorithm. The datasets were trained using 5-fold cross-validation. It consists of both protective antigen (193) and nonantigen (193) sequences. It predicts whether the given query epitope segments is antigenic or nonantigenic with their respective probability [56]. The presence or absence of similarity in predicted epitope segments with the human proteome was analyzed using the peptide matching program of PIR [57]. The EC tool [58] was employed to find the degree of conservancy of the epitope segments within the set of given hrHPV L2 protein sequences. The selected epitope segments of HPV58 with 14 hrHPV (16, 18, 31, 33, 35, 39, 45, 51, 56, 59, 68, 69, 73, and 82) strains performed the EC analysis.

The complete chimeric vaccine was designed by joining the optimized epitope segments (02), TLR adjuvants (02), and Th epitopes (02) with suitable amino acid linkers. Moreover, it is required to find the solubility of the designed chimeric vaccine on overexpression in *E. coli*. SOLpro is a useful tool to find the solubility of protein based on the two-stage SVM algorithm. It achieves an overall accuracy of 74%, which develops on standard evaluation metrics with 10-fold cross-validation. It predicts the query protein to be soluble or insoluble at $p \geq 0.5$ [59]. A range of physicochemical characteristics of the designed chimeric vaccine was also determined through ProtParam [60]. VaxiJen is the primary server used for prediction of antigenicity of the input sequence against different targets such as virus, bacteria, fungi, parasites, and tumors. Antigenicity was calculated based on the physicochemical properties of the protein sequences. Every target organism dataset contained 100 antigens and nonantigens. Moreover, the model organisms were validated using leave-one-out cross-validation (LOO-CV): providing 89% and 0.964 of accuracy and AUC at the threshold of 0.4 [61]. I-TASSER tool was employed to design the 3D structure of SGD58 and TLR5. It is a potential server that depends on the secondary-structure-mediated program of "Profile-Profile threading alignment (PPA) and iterative implementation of the TASSER". It has predicted a number of protein structures on request basis from 35 countries in the world. For the query inputs, the user obtains the confidence score, TM score (topology similarity assessments of the two various protein structures), root-mean-square deviation (RMSD), and cluster density values. Nevertheless, the higher C-score (ranging from -5 to $+2$) determines the best model with a higher confidence level [62]. Moreover, the 3D structure of the modeled protein was visualized using UCSF Chimera. Besides, unavailability of the crystal structure of TLR5, we have chosen TLR5 (PDB ID: 3J0A) as a template model to perform the homology modeling using I-TASSER. The high C-score model of the designed vaccine from the I-TASSER was further refined using the GalaxyRefine and 3DRefine tools. The GalaxyRefine is tool accessible in the GalaxyWeb server, is useful to refine the structure of a protein from the given query sequences based on template-based modeling, and undergoes loop and terminus portion refinement through the ab initio modeling method. The ninth critical assessment of techniques for protein structure prediction (CASP9) optimizes refinement and produces consistent core structures [63]. Another tool is 3Drefine, which prompts iteration analysis for ~ 300 amino acid residues efficiently in less than 5 min. It performs post-refinement model analysis with both or single MolProbity and random walk (RW) plus methods. The results are visualized using Javascript-based molecular viewer JSmol [64]. The top five models of each tool were used for further validation. The refined 3D models from the above steps were validated using the three interactive services such as ProSA, Ramachandran plot analysis, and ERRAT. ProSA-web is a potential tool for the refinement, validation, prediction, and modeling of protein structures. It indicates the difference in the protein structures through the respective score

and energy spot. It also facilitates the validation process of the protein structure that is acquired from X-ray, nuclear magnetic resonance (NMR) spectroscopy analysis, and theoretical calculations. As an output, the Z-score corresponds to the overall feature of the validated model [65]. RAMPAGE is used to validate the percentage (%) of favored, allowed, and the outlier region in the given query chimeric vaccine [66]. The statistics of noncovalent interactions between carbon, nitrogen, and oxygen atoms in the input sequence with best-resolution crystallographic structures were compared using the ERRAT tool. It implements empirical atom-based approach for verification of the protein structure and are more sensitive to errors (1.5Å) [67]. Similar steps were followed to validate the TLR5 model. DiscoTope 2.0 is a potential tool used to analyze the conformational (discontinuous) B cell epitopes from the input sequence. It showed a highly significant prediction performance with AUC at 0.824. The default -3.7 of threshold limit provides significant specificity (0.75) and sensitivity (0.47). It was selected for the present analysis, and the final score was evaluated by the combined calculation of the propensity score (PS) and contact numbers [68].

The PP interactions are the midpoint for all the biochemical pathways that involved in the biological functions. SwarmDock is a potential server for producing the 3D structure of the PP complexes. The validated best model of the vaccine construct (GalaxyRefine 3 model), and TLR5 (3DRefine 5 model) was chosen as a ligand and receptor respectively. The SwarmDock algorithm was implemented to perform the docking by restrained mode with the success rate of 71.6%. The following steps are majorly involved in this server such as input structures stepped into in preprocessing and minimization, docking by applying the hybrid algorithm of particle swarm optimization (PSO), minimizing (CHARMM), reranking and return the clustered structures into the users [69]. The molecular dynamics (MD) simulation determines the strength of the docked complex and designed vaccine SGD58. GROMACS 5.1.2 package with CHARMM force field was used to perform MD simulation. The details of MD simulation illustrated in our previous study [36]. EMBOSS Backtranseq v1.0 [70] is a suitable tool to uptake the query protein sequences, reverse translated and returned the optimizing coding sequences. The GenScript rare codon analysis [71] is a prominent tool for codon usage and its distribution analysis (codon adaptation index-CAI, glycine–cystine (GC) content, and codon frequency distribution (CFD) in the individual expression host organism based on the Optimum GeneTM algorithm. The details of virtual gene expression and cloning illustrated in our previous study [36]. All the supporting information deposited in Zenodo database [72].

Author Contributions: G.S., S.K., and D.-Q.W. conceived and designed the experiment. S.K and G.S. performed the immunoinformatics, vaccine design, and molecular docking studies. S.C., Q.W., and A.S.N. performed the molecular dynamics simulation. W.C.C., S.K., G.S., S.C., and D.-Q.W. wrote the main manuscript. S.K. and A.S.N. formatting the manuscript and figures according to the instructions. G.S., W.C.C., G.K., and D.-Q.W. critically reviewed the manuscript. All authors approved the final manuscript.

Funding: This paper was funded by the Ministry of Science and Technology of China (Grant No.: 2016YFA0501703), Henan Natural Science (Grant No.: 162300410060), to DQ.W. Henan University of Technology (Grant No.: 21450004) to S.K., Henan University of Technology (Grant No.: 21450003), and China Postdoctoral Science Foundation (Grant No.: 2018M632766) to G.S.

Acknowledgments: Authors are grateful to the Center for High Performance Computing, Shanghai Jiao Tong University, China for simulation studies.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Frazer, I.H. Development and implementation of papillomavirus prophylactic vaccines. *J. Immunol.* **2014**, *192*, 4007–4011. [[CrossRef](#)] [[PubMed](#)]
2. McLaughlin-Drubin, M.E.; Munger, K. Oncogenic activities of human papillomaviruses. *Virus Res.* **2009**, *143*, 195–208. [[CrossRef](#)] [[PubMed](#)]
3. Haedicke, J.; Iftner, T. Human papillomaviruses and cancer. *Radiother. Oncol.* **2013**, *108*, 397–402. [[CrossRef](#)] [[PubMed](#)]

4. Tao, G.; Yaling, G.; Zhan, G.; Pu, L.; Miao, H. Human papillomavirus genotype distribution among HPV-positive women in Sichuan province, Southwest China. *Arch. Virol.* **2018**, *163*, 65–72. [[CrossRef](#)] [[PubMed](#)]
5. Dunne, E.F.; Unger, E.R.; Sternberg, M.; McQuillan, G.; Swan, D.C.; Patel, S.S.; Markowitz, L.E. Prevalence of HPV infection among females in the United States. *JAMA* **2007**, *297*, 813–819. [[CrossRef](#)] [[PubMed](#)]
6. Kenter, G.G.; Welters, M.J.; Valentijn, A.R.; Lowik, M.J.; Berends-van der Meer, D.M.; Vloon, A.P.; Drijfhout, J.W.; Wafelman, A.R.; Oostendorp, J.; Fleuren, G.J.; et al. Phase I immunotherapeutic trial with long peptides spanning the E6 and E7 sequences of high-risk human papillomavirus 16 in end-stage cervical cancer patients shows low toxicity and robust immunogenicity. *Clin. Cancer Res.* **2008**, *14*, 169–177. [[CrossRef](#)] [[PubMed](#)]
7. De Vos van Steenwijk, P.J.; Ramwadhoebe, T.H.; Lowik, M.J.; van der Minne, C.E.; Berends-van der Meer, D.M.; Fathers, L.M.; Valentijn, A.R.; Oostendorp, J.; Fleuren, G.J.; Hellebrekers, B.W.; et al. A placebo-controlled randomized HPV16 synthetic long-peptide vaccination study in women with high-grade cervical squamous intraepithelial lesions. *Cancer Immunol. Immunother.* **2012**, *61*, 1485–1492. [[CrossRef](#)] [[PubMed](#)]
8. Rerucha, C.M.; Caro, R.J.; Wheeler, V.L. Cervical Cancer Screening. *Am. Fam. Phys.* **2018**, *97*, 441–448.
9. Hong, Y.; Zhang, C.; Li, X.; Lin, D.; Liu, Y. HPV and cervical cancer related knowledge, awareness and testing behaviors in a community sample of female sex workers in China. *BMC Public Health* **2013**, *13*, 696. [[CrossRef](#)]
10. Chen, W.; Zheng, R.; Baade, P.D.; Zhang, S.; Zeng, H.; Bray, F.; Jemal, A.; Yu, X.Q.; He, J. Cancer statistics in China, 2015. *CA Cancer J. Clin.* **2016**, *66*, 115–132. [[CrossRef](#)]
11. Zhou, H.L.; Zhang, W.; Zhang, C.J.; Wang, S.M.; Duan, Y.C.; Wang, J.X.; Yang, H.; Wang, X.Y. Prevalence and distribution of human papillomavirus genotypes in Chinese women between 1991 and 2016: A systematic review. *J. Infect.* **2018**, *76*, 522–528. [[CrossRef](#)] [[PubMed](#)]
12. You, W.; Li, S.; Du, R.; Zheng, J.; Shen, A. Epidemiological study of high-risk human papillomavirus infection in subjects with abnormal cytological findings in cervical cancer screening. *Exp. Ther. Med.* **2018**, *15*, 412–418. [[CrossRef](#)] [[PubMed](#)]
13. Long, W.; Yang, Z.; Li, X.; Chen, M.; Liu, J.; Zhang, Y.; Sun, X. HPV-16, HPV-58, and HPV-33 are the most carcinogenic HPV genotypes in Southwestern China and their viral loads are associated with severity of premalignant lesions in the cervix. *Virol. J.* **2018**, *15*, 94. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, C.; Huang, C.; Zheng, X.; Pan, D. Prevalence of human papillomavirus among Wenzhou women diagnosed with cervical intraepithelial neoplasia and cervical cancer. *Infect. Agent. Cancer* **2018**, *13*, 37. [[CrossRef](#)] [[PubMed](#)]
15. Dai, X.; Chen, L.; Li, J.; Wu, Y.; Hu, Y.; Xiang, F.; Guan, Q. Distribution characteristics of different human papillomavirus genotypes in women in Wuhan, China. *Cancer Med.* **2018**, *32*, e22581.
16. Zhao, P.; Liu, S.; Zhong, Z.; Hou, J.; Lin, L.; Weng, R.; Su, L.; Lei, N.; Hou, T.; Yang, H. Prevalence and genotype distribution of human papillomavirus infection among women in northeastern Guangdong Province of China. *J. Clin. Lab. Anal.* **2018**, *18*, 204. [[CrossRef](#)] [[PubMed](#)]
17. Liu, S.; Zhong, Z.; Hou, J.; Lin, L.; Weng, R.; Su, L.; Lei, N.; Hou, T.; Yang, H.; Li, K.; et al. Analysis of HPV distribution in patients with cervical precancerous lesions in Western China. *BMC Infect. Dis.* **2017**, *96*, e7304. [[CrossRef](#)]
18. Zhang, C.; Zhang, C.; Huang, J.; Shi, W. The Genotype of Human Papillomavirus and Associated Factors Among High Risk Males in Shanghai, China: A Molecular Epidemiology Study. *Med. Sci. Monit.* **2018**, *24*, 912–918. [[CrossRef](#)]
19. FDA licensure of bivalent human papillomavirus vaccine (HPV2, Cervarix) for use in females and updated HPV vaccination recommendations from the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb. Mortal. Wkly. Rep.* **2010**, *59*, 626–629.
20. Recommendations on the use of quadrivalent human papillomavirus vaccine in males—Advisory Committee on Immunization Practices (ACIP), 2011. *MMWR Morb. Mortal. Wkly. Rep.* **2011**, *60*, 1705–1708.
21. Petrosky, E.; Bocchini, J.A., Jr.; Hariri, S.; Chesson, H.; Curtis, C.R.; Saraiya, M.; Unger, E.R.; Markowitz, L.E. Use of 9-valent human papillomavirus (HPV) vaccine: Updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR Morb. Mortal. Wkly. Rep.* **2015**, *64*, 300–304. [[PubMed](#)]

22. Paz-Zulueta, M.; Alvarez-Paredes, L.; Rodriguez Diaz, J.C.; Paras-Bravo, P.; Andrada Becerra, M.E.; Rodriguez Ingelmo, J.M.; Ruiz Garcia, M.M.; Portilla, J.; Santibanez, M. Prevalence of high-risk HPV genotypes, categorised by their quadrivalent and nine-valent HPV vaccination coverage, and the genotype association with high-grade lesions. *BMC Cancer* **2018**, *18*, 112. [[CrossRef](#)] [[PubMed](#)]
23. Jiang, R.T.; Schellenbacher, C.; Chackerian, B.; Roden, R.B. Progress and prospects for L2-based human papillomavirus vaccines. *Expert Rev. Vaccines* **2016**, *15*, 853–862. [[CrossRef](#)] [[PubMed](#)]
24. Chroboczek, J.; Szurgot, I.; Szolajska, E. Virus-like particles as vaccine. *Acta Biochim. Pol.* **2014**, *61*, 531–539. [[PubMed](#)]
25. Pandhi, D.; Sonthalia, S. Human papilloma virus vaccines: Current scenario. *Indian J. Sex. Transm. Dis. AIDS* **2011**, *32*, 75–85. [[CrossRef](#)] [[PubMed](#)]
26. Monie, A.; Hung, C.F.; Roden, R.; Wu, T.C. Cervarix: A vaccine for the prevention of HPV 16, 18-associated cervical cancer. *Biologics* **2008**, *2*, 97–105. [[PubMed](#)]
27. Angioli, R.; Lopez, S.; Aloisi, A.; Terranova, C.; De Cicco, C.; Scaletta, G.; Capriglione, S.; Miranda, A.; Luvero, D.; Ricciardi, R.; et al. Ten years of HPV vaccines: State of art and controversies. *Crit. Rev. Oncol. Hematol.* **2016**, *102*, 65–72. [[CrossRef](#)] [[PubMed](#)]
28. Karanam, B.; Jagu, S.; Huh, W.K.; Roden, R.B. Developing vaccines against minor capsid antigen L2 to prevent papillomavirus infection. *Immunol. Cell Biol.* **2009**, *87*, 287–299. [[CrossRef](#)]
29. Schiller, J.T.; Muller, M. Next generation prophylactic human papillomavirus vaccines. *Lancet Oncol.* **2015**, *16*, e217–e225. [[CrossRef](#)]
30. Wang, J.W.; Roden, R.B. L2, the minor capsid protein of papillomavirus. *Virology* **2013**, *445*, 175–186. [[CrossRef](#)]
31. Chandrachud, L.M.; Grindlay, G.J.; McGarvie, G.M.; O’Neil, B.W.; Wagner, E.R.; Jarrett, W.F.; Campo, M.S. Vaccination of cattle with the N-terminus of L2 is necessary and sufficient for preventing infection by bovine papillomavirus-4. *Virology* **1995**, *211*, 204–208. [[CrossRef](#)] [[PubMed](#)]
32. Gaukroger, J.M.; Chandrachud, L.M.; O’Neil, B.W.; Grindlay, G.J.; Knowles, G.; Campo, M.S. Vaccination of cattle with bovine papillomavirus type 4 L2 elicits the production of virus-neutralizing antibodies. *J. Gen. Virol.* **1996**, *77*, 1577–1583. [[CrossRef](#)] [[PubMed](#)]
33. Wei, D.Q.; Selvaraj, G.; Kaushik, A.C. Computational Perspective on the Current State of the Methods and New Challenges in Cancer Drug Discovery. *Curr. Pharm. Des.* **2018**, *24*, 3725–3726. [[CrossRef](#)] [[PubMed](#)]
34. Kaliampurthi, S.; Selvaraj, G.; Junaid, M.; Khan, A.; Gu, K.; Wei, D.Q. Cancer Immunoinformatics: A Promising Era in the Development of Peptide Vaccines for Human Papillomavirus-induced Cervical Cancer. *Curr. Pharm. Des.* **2018**, *24*, 3791–3817. [[CrossRef](#)] [[PubMed](#)]
35. Kaliampurthi, S.; Selvaraj, G.; Kaushik, A.C.; Gu, K.R.; Wei, D.Q. Designing of CD8+ and CD8+-overlapped CD4+ epitope vaccine by targeting late and early proteins of human papillomavirus. *Biologics* **2018**, *12*, 107. [[PubMed](#)]
36. Kaliampurthi, S.; Selvaraj, G.; Chinnasamy, S.; Wang, Q.; Nangraj, A.S.; Cho, W.; Gu, K.; Wei, D.Q. Exploring the Papillomaviral Proteome to Identify Potential Candidates for a Chimeric Vaccine against Cervix Papilloma Using Immunomics and Computational Structural Vaccinology. *Viruses* **2019**, *11*, 63. [[CrossRef](#)] [[PubMed](#)]
37. UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [[CrossRef](#)]
38. Chen, S.; Hong, W.; Shao, H.; Fu, Y.; Liu, X.; Chen, D.; Xu, A. Allelic distribution of HLA class I genes in the Tibetan ethnic population of China. *Int. J. Immunogenet.* **2006**, *33*, 439–445. [[CrossRef](#)]
39. Chen, S.; Hu, Q.; Xie, Y.; Zhou, L.; Xiao, C.; Wu, Y.; Xu, A. Origin of Tibeto-Burman speakers: Evidence from HLA allele distribution in Lisu and Nu inhabiting Yunnan of China. *Hum. Immunol.* **2007**, *68*, 550–559. [[CrossRef](#)]
40. Chen, S.; Ren, X.; Liu, Y.; Hu, Q.; Hong, W.; Xu, A. Human leukocyte antigen class I polymorphism in Miao, Bouyei, and Shui ethnic minorities of Guizhou, China. *Hum. Immunol.* **2007**, *68*, 928–933. [[CrossRef](#)]
41. Wang, X.C.; Sun, L.Q.; Ma, L.; Li, H.X.; Wang, X.L.; Wang, X.; Yun, T.; Meng, N.L.; Lv, D.L. Prevalence and genotype distribution of human papillomavirus among women from Henan, China. *Asian Pac. J. Cancer Prev.* **2014**, *15*, 7333–7336. [[CrossRef](#)] [[PubMed](#)]
42. Zhang, H.Y.; Fei, M.D.; Jiang, Y.; Fei, Q.Y.; Qian, H.; Xu, L.; Jin, Y.N.; Jiang, C.Q.; Li, H.X.; Tiggelaar, S.M.; et al. The diversity of human papillomavirus infection among human immunodeficiency virus-infected women in Yunnan, China. *Virol. J.* **2014**, *11*, 202. [[CrossRef](#)] [[PubMed](#)]

43. Lu, J.F.; Shen, G.R.; Li, Q.; Chen, X.; Ma, C.F.; Zhu, T.H. Genotype distribution characteristics of multiple human papillomavirus in women from the Taihu River Basin, on the coast of eastern China. *BMC Infect. Dis.* **2017**, *17*, 226. [[CrossRef](#)] [[PubMed](#)]
44. Wang, Y.; Xue, J. Distribution and role of high-risk human papillomavirus genotypes in women with cervical intraepithelial neoplasia: A retrospective analysis from Wenzhou, southeast China. *Cancer Med.* **2018**. [[CrossRef](#)] [[PubMed](#)]
45. Dai, X.; Chen, L.; Li, J.; Wu, Y.; Hu, Y.; Vita, R.; Overton, J.A.; Greenbaum, J.A.; Ponomarenko, J.; Clark, J.D.; et al. The immune epitope database (IEDB) 3.0. *Cancer Med.* **2015**, *43*, D405–D412.
46. Moutaftsi, M.; Peters, B.; Pasquetto, V.; Tschärke, D.C.; Sidney, J.; Bui, H.H.; Grey, H.; Sette, A. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat. Biotechnol.* **2006**, *24*, 817–819. [[CrossRef](#)] [[PubMed](#)]
47. Nielsen, M.; Lundegaard, C.; Worning, P.; Lauemøller, S.L.; Lamberth, K.; Buus, S.; Brunak, S.; Lund, O. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **2003**, *12*, 1007–1017. [[CrossRef](#)]
48. Bhasin, M.; Raghava, G.P. Prediction of CTL epitopes using QM, SVM and ANN techniques. *Vaccine* **2004**, *22*, 3195–3204. [[CrossRef](#)]
49. Liu, I.H.; Lo, Y.S.; Yang, J.M. PAComplex: A web server to infer peptide antigen families and binding models from TCR-pMHC complexes. *Nucleic Acids Res.* **2011**, *39*, W254–W260. [[CrossRef](#)]
50. Wang, P.; Sidney, J.; Dow, C.; Mothe, B.; Sette, A.; Peters, B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput. Biol.* **2008**, *4*, e1000048. [[CrossRef](#)]
51. Paul, S.; Sidney, J.; Sette, A.; Peters, B. TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr. Protoc. Immunol.* **2016**, *114*, 18.19.1–18.19.24. [[CrossRef](#)] [[PubMed](#)]
52. Dhanda, S.K.; Vir, P.; Raghava, G.P. Designing of interferon-gamma inducing MHC class-II binders. *Biol. Direct.* **2013**, *8*, 30. [[CrossRef](#)]
53. Saha, S.; Raghava, G.P. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* **2006**, *65*, 40–48. [[CrossRef](#)] [[PubMed](#)]
54. Dimitrov, I.; Bangov, I.; Flower, D.R.; Doytchinova, I. AllerTOP v.2—A server for in silico prediction of allergens. *J. Mol. Model.* **2014**, *20*, 2278. [[CrossRef](#)] [[PubMed](#)]
55. Dimitrov, I.; Naneva, L.; Doytchinova, I.; Bangov, I. AllergenFP: Allergenicity prediction by descriptor fingerprints. *Bioinformatics* **2014**, *30*, 846–851. [[CrossRef](#)] [[PubMed](#)]
56. El-Manzalawy, Y.; Dobbs, D.; Honavar, V. Predicting protective bacterial antigens using random forest classifiers. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, Orlando, FL, USA, 7–10 October 2012; pp. 426–433.
57. Chen, C.; Li, Z.; Huang, H.; Suzek, B.E.; Wu, C.H. A fast Peptide Match service for UniProt Knowledgebase. *Bioinformatics* **2013**, *29*, 2808–2809. [[CrossRef](#)] [[PubMed](#)]
58. Bui, H.H.; Sidney, J.; Li, W.; Fusseder, N.; Sette, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinform.* **2007**, *8*, 361. [[CrossRef](#)]
59. Magnan, C.N.; Randall, A.; Baldi, P. SOLpro: Accurate sequence-based prediction of protein solubility. *Bioinformatics* **2009**, *25*, 2200–2207. [[CrossRef](#)]
60. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.E.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607.
61. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinform.* **2007**, *8*, 4. [[CrossRef](#)]
62. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinform.* **2008**, *9*, 40. [[CrossRef](#)]
63. Ko, J.; Park, H.; Heo, L.; Seok, C. GalaxyWEB server for protein structure prediction and refinement. *Nucleic Acids Res.* **2012**, *40*, W294–W297. [[CrossRef](#)] [[PubMed](#)]
64. Bhattacharya, D.; Cheng, J. i3Drefine software for protein 3D structure refinement and its assessment in CASP10. *PLoS ONE* **2013**, *8*, e69648. [[CrossRef](#)] [[PubMed](#)]
65. Wiederstein, M.; Sippl, M.J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–W410. [[CrossRef](#)] [[PubMed](#)]

66. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi,psi and Cbeta deviation. *Proteins* **2003**, *50*, 437–450. [[CrossRef](#)] [[PubMed](#)]
67. Colovos, C.; Yeates, T.O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* **1993**, *2*, 1511–1519. [[CrossRef](#)] [[PubMed](#)]
68. Kringelum, J.V.; Lundegaard, C.; Lund, O.; Nielsen, M. Reliable B cell epitope predictions: Impacts of method development and improved benchmarking. *PLoS Comput. Biol.* **2012**, *8*, e1002829. [[CrossRef](#)] [[PubMed](#)]
69. Torchala, M.; Moal, I.H.; Chaleil, R.A.; Fernandez-Recio, J.; Bates, P.A. SwarmDock: A server for flexible protein-protein docking. *Bioinformatics* **2013**, *29*, 807–809. [[CrossRef](#)]
70. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **2000**, *16*, 276–277. [[CrossRef](#)]
71. The GenScript Rare Codon Analysis. Available online: <https://www.genscript.com/tools/rare-codon-analysis> (accessed on 1 October 2018).
72. Kaliamurthi, S.; Selvaraj, G.; Wei, D.Q. Immunomics datasets: To identify potential candidates for chimeric vaccine design to cervix papilloma [Data set]. *Zenodo* **2018**. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).