



# The possible regions to design Human Papilloma Viruses vaccine in Iranian L1 protein

Behzad Dehghani<sup>1</sup> · Zahra Hasanshahi<sup>1</sup> · Tayebeh Hashempour<sup>1</sup> · Mohamad Motamedifar<sup>1,2</sup>

Received: 16 June 2019 / Accepted: 5 November 2019 / Published online: 24 December 2019  
© Institute of Molecular Biology, Slovak Academy of Sciences 2019

## Abstract

Human Papilloma Virus (HPV) genome encodes several proteins, as L1 is major capsid protein and L2 is minor capsid protein. Among all HPV types HPV-16 and HPV-18 are the most common high-risk HPV (HR-HPV) types globally and the majority of cases are infected with these types. HPV entry and the initial interaction with the host cell are mainly related to the L1 protein which is the main component of HPV vaccines. The aim of this research was comparison analysis among all Iranian L1 protein sequences submitted in NCBI GenBank to find the major substitutions as well as structural and immune properties of this protein. All sequences HPV L1 protein from Iranian isolates from 2014 to 2016 were selected and obtained from NCBI data bank. “CLC Genomics Workbench” was used to translate alignment. To predict B cell epitopes, we employed several programs. Modification sites such as phosphorylation, glycosylation, and disulfide bonds were determined. Secondary and tertiary structures of all sequences were analyzed. Several mutations were found and major mutations were in amino acid residues 102, 202, 207, 292, 379, and 502. The mentioned mutations showed the minor effect on B cell and physicochemical properties of the L1 protein. Six disulfide bonds were determined in L1 protein and also in several N-link glycosylation and phosphorylation sites. Five L1 loops were determined, which had great potential to be B cell epitopes with high antigenic properties. All in all, this research as the first report from Iran described the tremendous potential of two L1 loops (BC and FG) to induce immune system which can be used as the descent candidate to design a new vaccine against HPV in the Iranian population. In addition, some differences between the reference sequence and Iranian patients’ sequences were determined. It is essential to consider these differences to monitor the effectiveness and efficacy of the vaccine for the Iranian population. Our results provide a vast understanding of L1 protein that can be useful for further studies on HPV infections and new vaccine generations.

**Keywords** HPV · L1 · Bioinformatics

## Abbreviations

HPV	Human Papilloma Virus
PVs	Papillomaviruses
VLPs	Self-assemble in virus-like particles
C	Cysteine
T	Threonine
NmAbs	Neutralizing monoclonal antibodies

MAbs	Monoclonal antibodies
HR-HPV	High-risk HPV

## Introduction

Globally, 2784 million women aged 15 years and older are at risk of developing cervical cancer; 569,847 women are diagnosed with cervical cancer, and 311,365 die from the disease annually (Bruni et al. 2019). After breast cancer, it is the second major cancer of women’s reproductive organs related to cancer deaths. (Beckmann et al. 2010; Rahimifar et al. 2010). Studies showed more than 69.4% of this cancer is related to HPV infection as HPV was found in pre-invasive lesions (Bruni et al. 2019).

The rate of Human Papillomavirus (HR HPV) among Iranian women has increased in the recent years; several

✉ Tayebeh Hashempour  
thashem@sums.ac.ir

<sup>1</sup> Shiraz HIV/AIDS Research Center, Shiraz University of Medical Sciences, Shiraz, Fars, Iran

<sup>2</sup> Department of Bacteriology and Virology, School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran

studies have estimated this rate among healthy women from 31% to 34% (Jamdar et al. 2018; Shafaghi et al. 2013; Yousefzadeh et al. 2014). This rate was around 38.6% in Iranian patients with cervical infections and 23% with breast cancer (Dadashi et al. 2017; Haghshenas et al. 2016).

HPV is a non-enveloped particle and like other Papillomaviruses (PVs) it is characterized by having an icosahedral capsid of about 55 nm diameter (Doorbar et al. 2015). HPV genome expresses six nonstructural viral regulatory proteins (E1, E2, E4, E5, E6, and E7) and the virus capsid contains two encoded proteins, the L1 major and L2 minor capsid proteins (Frazer 2009; Zheng and Baker 2006). L1 protein is highly conserved among different HPV types and only 10% of the L1 sequence is different among HPV types, and it has the ability to self-assemble in virus-like particles (VLPs) (Kim 2016; Xu et al. 2016); it seems that a surface cell protein was involved in VLP binding. Based on VLPs, VLP-based vaccines were introduced that offered highly effective protection against HPV infections (Buck et al. 2013). L1 has an indispensable role in HPV entry and the initial interaction is remarkably attributable to L1 interactions with proteoglycans (Buck et al. 2013). This virus is categorized into low-risk or high-risk. Prior studies showed that about 50–60% of human anogenital carcinomas belonged to HPV-16 and around 20% was HPV-18 (Vera-Bravo et al. 2003). Likewise, several studies showed that the majority of HPV infections in Iranian patients were related to HPV-16. Two recommended HPV vaccines, Gardasil and Cervarix contain VLPs assembled from the L1 proteins of different HPV types (Monie et al. 2008; Siddiqui and Perry 2006). Consequently, any change in HPV L1 can lead to a decrease in the effectiveness and efficacy of vaccines that are used commonly. The aim of this study, as the first report of the most prevalent substitutions of HPV L1 protein among sequences obtained from Iranian patients, was to find mutations and their effect on the structure, physicochemical properties, and antigenic features. We also attempted to provide a comprehensive view of L1 protein as a major candidate designing vaccines against HPV infections.

## Materials and methods

### L1 protein sequences availability

All 60 L1 sequences from 2014 to 2016 (KM058636-KM058666, KP160988-KP160999, KP161000-KP161014, and KX827590) and a reference sequence (K02718.1) were obtained from NCBI databank at (<http://www.ncbi.nlm.nih.gov>), all sequences were used in this study belonged to HPV-16. Table 1 shows all software which were used in this study.

### Substitutions

The CLC Genomics Workbench was used for the analysis of the mutations, alignment of the translated peptides and homology among sequences. Based on mutations, the selected sequenced were categorized into 3 groups (Table 2).

### Signal peptide prediction

To predict signal peptide, we used “Signal-BLAST” (Frank and Sippl 2008) and “predisi” (Hiller et al. 2004).

### Physicochemical properties

Prediction of instability index, aliphatic index, theoretical isoelectric point (pI), and grand average of hydrophathy (GRAVY) was done by “Expasy’sProtParam” (Gasteiger et al. 2005).

### Postmodification positions

“DiANNA” (Ferrè and Clote 2006) and “SCRATCH” (Cheng et al. 2005) were used to predict disulfide bonds prediction. Phosphorylation sites were computed by “NetPhosK” (Blom et al. 2004), “DISPHOS” (Iakoucheva et al. 2004), and “NetPhos” (Blom et al. 1999), and finally, N-glycosylation sites were found by “NetNGlyc” (Gupta et al. 2017) and “GlycoEP” (Chauhan et al. 2013).

### Secondary and tertiary structure

“SOPMA” (Geourjon and Deleage 1995) and “Phyre” (Kelley and Sternberg 2009) server were reliable programs we employed to predict the secondary structure. “I-TASSER” (Roy et al. 2010), “Phyre<sup>2</sup>” (Kelley and Sternberg 2009), and “(PS)2-v2” (Chen et al. 2006) were utilized to predict the tertiary structure of the selected sequences. All predicted 3D structures were evaluated for the stereochemistry, reliability, and quality by “Qmean” (Benkert et al. 2008) and “Rampage”. The obtained 3D structures were refined by “GalaxyRefine” (Heo et al. 2013) program. “Discovery Studio” was used to find the loops on HPV L1.

### Immunoinformatics analysis

“Immuneepitope” (Vita et al. 2018), “BcePred” (Saha and Raghava 2004), “ABCpred” (Saha and Raghava 2006b), and “Bepipred” (Jespersen et al. 2017) online programs were performed to find B Cell epitopes sites. Allergenic properties were predicted by “AlgPred” (Saha and Raghava 2006a) and “VaxiJen” (Doytchinova and Flower 2007) software computed protective antigens and subunit vaccines prediction. Hydrophilicity, flexibility, and surface accessibility of GP were anticipated using “IEDB” tools.

**Table 1** The software used in this study and related URLs

	Software	URL	Function
1	Signal-BLAST	<a href="http://sigpep.services.came.sbg.ac.at/signalblast.html">http://sigpep.services.came.sbg.ac.at/signalblast.html</a>	Signal peptide prediction
2	predisi	<a href="http://www.predisi.de/">http://www.predisi.de/</a>	Signal peptide prediction
3	SignalP	<a href="http://www.cbs.dtu.dk/services/SignalP/">http://www.cbs.dtu.dk/services/SignalP/</a>	Signal peptide prediction
4	ProtParam	<a href="http://expasy.org/tools/protparam.html">http://expasy.org/tools/protparam.html</a>	Physicochemical properties
5	DiANNA	<a href="http://clavius.bc.edu/~clotelab/DiANNA/">http://clavius.bc.edu/~clotelab/DiANNA/</a>	Disulfide bonds prediction
6	SCRATCH	<a href="http://scratch.proteomics.ics.uci.edu/">http://scratch.proteomics.ics.uci.edu/</a>	Disulfide bonds prediction
7	NetPhosK	<a href="http://www.cbs.dtu.dk/services/NetPhosK/">http://www.cbs.dtu.dk/services/NetPhosK/</a>	Phosphorylation sites prediction
8	DISPHOS	<a href="http://www.dabi.temple.edu/disphos/">http://www.dabi.temple.edu/disphos/</a>	Phosphorylation sites prediction
9	NetPhos	<a href="http://www.cbs.dtu.dk/services/NetPhos/">http://www.cbs.dtu.dk/services/NetPhos/</a>	Phosphorylation sites prediction
10	NetNGlyc	<a href="http://www.cbs.dtu.dk/services/NetNGlyc/">www.cbs.dtu.dk/services/NetNGlyc/</a>	N-glycosylation sites prediction
11	GlycoEP	<a href="http://www.imtech.res.in/raghava/glycoep/submit.html">www.imtech.res.in/raghava/glycoep/submit.html</a>	N-glycosylation sites prediction
12	SOPMA	<a href="https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html">https://npsa-prabi.ibcp.fr/NPSA/npsa_sopma.html</a>	Secondary structure prediction
13	Phyre2	<a href="http://www.sbg.bio.ic.ac.uk/~phyre2/">http://www.sbg.bio.ic.ac.uk/~phyre2/</a>	Secondary structure prediction
14	I-TASSER	<a href="https://zhanglab.cmb.med.umich.edu/I-TASSER/">https://zhanglab.cmb.med.umich.edu/I-TASSER/</a>	Tertiary structure prediction
15	Phyre <sup>2</sup>	<a href="http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index">http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index</a>	Tertiary structure prediction
16	(PS)2-v2	<a href="http://ps2.life.nctu.edu.tw/">http://ps2.life.nctu.edu.tw/</a>	Tertiary structure prediction
17	Qmean	<a href="https://swissmodel.expasy.org/qmean/">https://swissmodel.expasy.org/qmean/</a>	Tertiary structure qualification
18	Rammpage	<a href="http://mordred.bioc.cam.ac.uk/~rapper/rampage.php">http://mordred.bioc.cam.ac.uk/~rapper/rampage.php</a>	Ramachandran Plot Analysis
19	immuneepitope	<a href="https://www.iedb.org">https://www.iedb.org</a>	Immuno-informatic analysis
20	BcePred	<a href="http://crdd.osdd.net/raghava/bcepred/">crdd.osdd.net/raghava/bcepred/</a>	Immuno-informatic analysis
21	ABCpred	<a href="http://crdd.osdd.net/raghava/abcpred/">http://crdd.osdd.net/raghava/abcpred/</a>	Immuno-informatic analysis
22	Bepipred	<a href="http://www.cbs.dtu.dk/services/BepiPred/">http://www.cbs.dtu.dk/services/BepiPred/</a>	Immuno-informatic analysis
23	AlgPred	<a href="http://crdd.osdd.net/raghava/algpred/">http://crdd.osdd.net/raghava/algpred/</a>	Immuno-informatic analysis
24	VaxiJen	<a href="http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen">http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen</a>	Immuno-informatic analysis
25	IEDB	<a href="http://tools.iedb.org/bcell/">http://tools.iedb.org/bcell/</a>	Immuno-informatic analysis

## Results

### Mutation

All the selected sequences contained a mutation in amino acid 228, and the most prevalent mutations occurred in positions 102, 202, 207, 292, 379, 502 (Table 2).

### Signal peptide

Analysis by two selected software showed that signal-blast and predisi could not find any position for signal peptide.

### Protparam analysis

L1 protein with 531 amino acids length and molecular weight of 59,554.02 Da had pI of 8.27 indicated that it is a basic protein. Its' in vivo half-life was estimated to be 30 h in mammalian cells and more than 10 h in E.coli. Its instability index and aliphatic index were computed to be a stable protein (Table 3).

### Post-modification prediction

Table 4 illustrates the disulfide bonds results, determining 6 probable disulfide bonds in all sequences; however, because of mutations, different bonds were found in some positions. Prediction of glycosylation sites by several reliable tools showed 4 conserve positions for all sequences (157,242,367, and 421). The analysis showed that L1 is a highly phosphorylated protein and the software determined a large number of phosphorylation positions in such protein (Table 5). In comparison with the reference sequence, 4 new phosphorylation positions (229, 376, 507, and 514) were found in 3 groups of selected sequences. In considering the list of mutations we could not find any related mutation to these new positions. Therefore, other factors may have influenced one software to predict these positions.

### Structural analysis

According to the analysis, the majority of the secondary structures of L1 were random coil (42.5%); extended strand and alpha helix were 28% and 19%, respectively (Table 6).

**Table 2** All mutations which were summarized and all sequences harbored substitution in amino acid 60 and Selected sequences groups based on the most prevalent mutations

Mutations	Frequency
Q 2 E	1(1.6%)
L 39 S	1(1.6%)
H 102 Y	40(66.6%)
T 202 N	40(66.6%)
N 207 T	1(1.6%)
V 220 I	1(1.6%)
H 228 D	60 (100%)
T 292 A	53(88.3%)
A 310 V	1(1.6%)
T 379 P	40(66.6%)
T 424 S	1(1.6%)
L 502 F	40(66.6%)
K 514 N	1(1.6%)
Groups	Mutations for each group
1	102, 202, 207,228, 292, 379, and 502
2	228 and292
3	228

Figure 1 shows the tertiary structure of L1 predicted by reliable software. The Qmean and Ramachandran plot results are summarized in Table 7. The coverage of each predicted structure is shown in Fig. 2. By using “discovery Studio”, 5 loops were found on the L1 structure which was refined by “GalaxyRefine”; the results were summarized in Table 8 and illustrated in Fig. 3.

### Immune properties of L1

B cell epitope prediction by using software showed 6 highly potential regions (37–50, 79–86, 194–210, 303–313, 427–442, and 509–530). “AlgPred” could not find any IgE epitope, and “VexiJen” showed L1 is a Probable antigen. “IEDB”

**Table 4** Disulfide bonds were computed by DiANNA and SCRATCH with several similar positions among the sequences

disulfide bonds	Reference	1	2	3
13–13	+	+	+	+
13–211	+	+	+	+
13–371	+	+	+	+
128–211	+	+	+	+
172–405	+	+	+	+
201–453	+	–	–	–
211–453	+	–	–	–
211–454	–	+	+	+
255–371	+	+	+	+
405–453	+	–	–	–
350–454	–	+	+	+
187–454	–	+	+	+
350–371	–	+	–	–

+: Mentioned band was predicted in the sequence. -: Mentioned band was not found in the sequence

results are displayed in Fig. 3, with 3 distinct parameters which used to locate regions with the highest antigenic ability and all prediction calculations are based on propensity scales for each of the 20 amino acids. In each diagram, the regions with higher scale than the threshold were favorable regions consisting of higher scored residues. Combination of the results of all three diagrams could define the region with high potential to be a candidate as an antigen.

### Discussion

Several studies on Iranian patients with HPV infections illustrated that the majority of cases were infected with HPV-16 and after that HPV-18 is ranked second (Farjadian et al. 2003; Ghaffari et al. 2006; Salehi-Vaziri et al. 2016; Shahramian et al. 2011). As there were few sequences related to HPV-18, all the sequences used in this study belonged to HPV-16.

**Table 3** Physicochemical properties of L1 protein in selected sequences and reference sequence

Protparam	ref(K02718.1)	1	2	3
Molecular weight	59,554.02	59,542.99	59,473.93	59,503.95
pI	8.27	8.26	8.27	8.27
Estimated half-life	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h ( <i>Escherichia coli</i> , in vivo).	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h ( <i>Escherichia coli</i> , in vivo).	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h ( <i>Escherichia coli</i> , in vivo).	30 h (mammalian reticulocytes, in vitro). >20 h (yeast, in vivo). >10 h ( <i>Escherichia coli</i> , in vivo).
Instability index	36.11 stable.	35.87 stable	36.19 stable	36.19 stable
GRAVY	–0.298	–0.294	–0.289	–0.294
Aliphatic index	76.53	75.99	76.72	76.53

**Table 5** Phosphorylation sites found by NetPhosK<sup>+</sup>, “DISPHOS”, “NetPhos”, several serine, threonine, and tyrosine amino acids being predicted

	Serine	Threonine	Tyrosine
Reference	49–115–244–308–324–369–375–422–518–519–521	36–65–121–292–320–362–380–439–443–544–516–523–507–517–520–522	53–161–260–268–302–381–444
1	+	+ and 376, 514	+
2	+	+ and 507, 514	+
3	+	+ and 229, 507, 514	+

+: All mentioned positions were predicted in the sequence

**Table 6** Secondary structure prediction results for L1 protein; the majority of L1 protein structure consisting of random coil

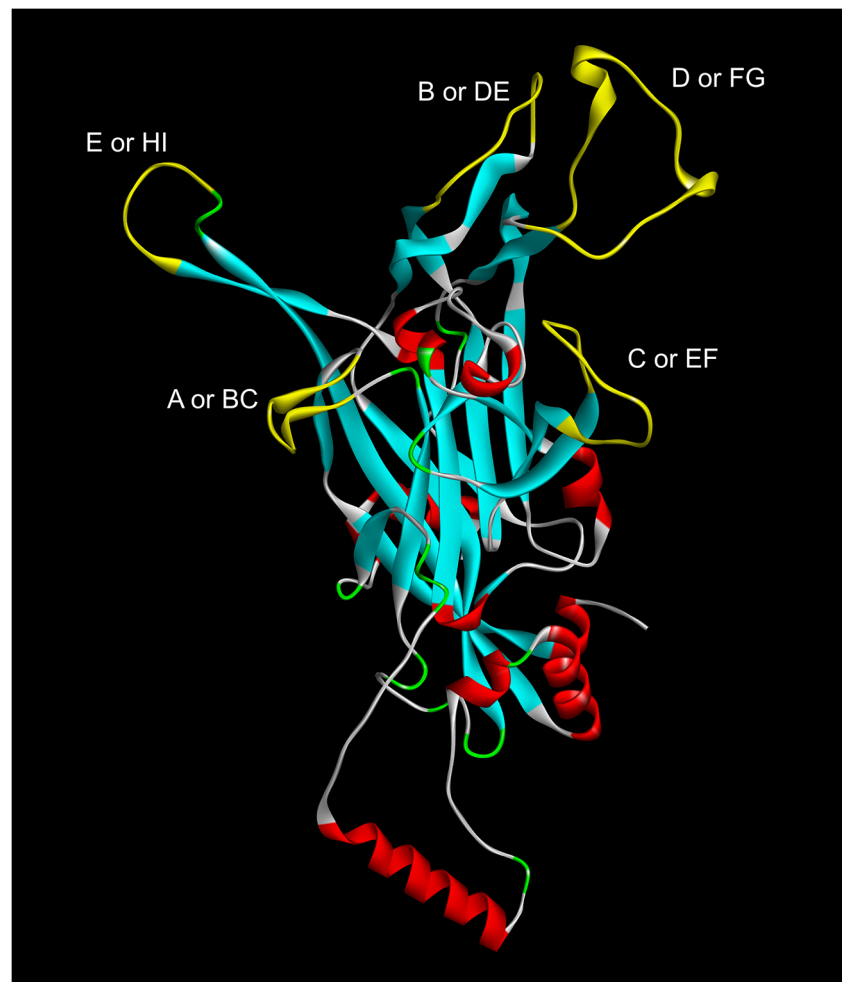
Secondary structure	Reference	1	2	3
Alpha helix	19.77%	19.59%	19.96%	19.40%
Extended strand	28.25%	28.81%	28.63%	28.63%
Beta turn	9.42%	8.85%	9.04%	9.42%
Random coil	42.56%	42.75%	42.37%	42.56%

436 Iranian women with different cervical lesions were studied by Salehi-Vaziri et al. in 2013 and the results showed 32.8% of the samples were HPV-16 (Salehi-Vaziri et al. 2016). A study on 100 Iranian patients with cervical carcinomas by Mortazavi et al. in 2002 showed that more than 73% of them were infected with HPV-16 and around 12% with HPV-18 and HPV-33 types (Mortazavi et al. 2002). In 2014, research on 851 Iranian women showed 7.3% of them were infected with HPV-16 (Yousefzadeh et al. 2014). While numerous investigations aimed to determine HPV and genotyping by using L1 sequences, there are no study reporting mutations in this gene. Therefore, this study was the first report on HPV L1 mutations in Iranian patients and the results are applicable to increase the efficiency of the current vaccine.

Antoine Touze et al. in 1998 showed substitution in the amino acid region from residues 83 to 97 seemed to affect the level of expression of the L1 protein that has an important effect on HPV vaccines development and serological tests (Touze et al. 1998). In 2014, Fleury et al. found that only a few mutations within the FG loop are sufficient to generate a new serotype escaping vaccination (Fleury et al. 2014). Shailja Pande in 2008 described that the most frequent variation in amino acid (T353P), threonine, was replaced by non-polar aliphatic amino acid and may have an effect on the structure or function of the L1 protein (Pande et al. 2008). It was suggested that it had a significant role in immune recognition and vaccine development strategies and led to conformational changes within epitopes relevant for viral neutralization. In 2003, Yoshiyuki Ishii found that three cysteines (175, 185, and 428) are required for the normal assembly of L1-capsids through trimerization and dimerization of L1; also, substitution in the mentioned regions affects the normal function of HPV (Ishii et al. 2003). The present study showed several mutations in L1 protein; however, we did not find any mutation similar to previous studies in other countries, probably due to different regions or different types of their studies.

Lee et al. in 2008 showed the expression rate of HPV L1 capsid protein in the uterine cervical specimens and determined that HPV low-risk group had a higher L1 capsid expression rate than HPV high-risk group; they suggested that L1 capsid expression might be related to favorable disease biology (Lee et al. 2008). Expression of L1 in the insect cells

**Fig. 1** HPV L1 3D model structure; yellow: 5 identified loops, and red:  $\alpha$ -helix



by using recombinant baculoviruses system by Antoine Touze et al. (1998) in 1998 was done, showing the stability of this protein in eukaryotic cells and the yield depend on mutations in a region (83 to 97) that seemed to affect the level of expression. In 2009, Bazan et al. expressed, purified recombinant HPV16 L1 in methylotrophic yeast (*Pichia pastoris*) that could be useful to produce low-cost vaccines (Bazan et al. 2009). Wang in 1999 and Zhang in 1998 produced a fused form in *Escherichia coli* using an inducible expression system and showed the stability of the L1 protein in this host (Zhang et al. 1998). Likewise, our results confirmed the stability of the L1 protein in prokaryotic and eukaryotic hosts and showed

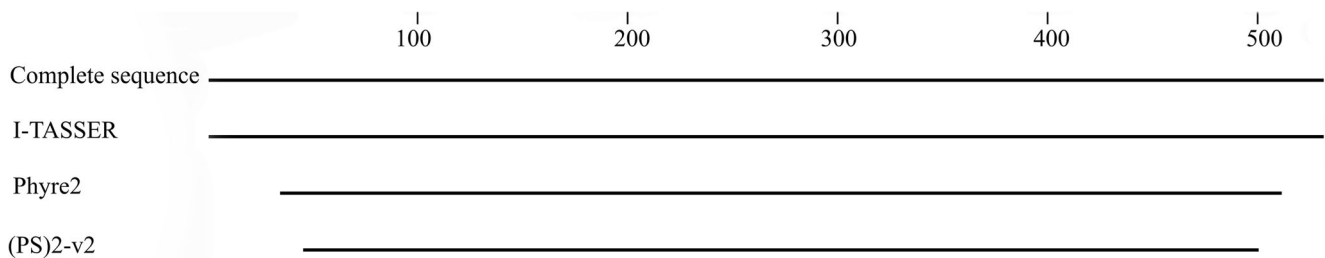
it was a thermostable and hydrophilic protein and two selected software did not predict any signal peptide for the L1 protein.

Li in 1998 determined the vital role of disulfide bonds in papillomavirus capsid assembly and suggested a conserved position (cysteine (C) 424) and also the mutant one could not assemble in vitro into capsid-like structures (Li et al. 1998). In 1998, Martin Sapp et al. described the conserved exclusive disulfide bond of C176 with C427 and confirmed the importance of this bond in the structure of the papillomavirus capsid and DNA packaging (Sapp et al. 1998). Conway et al. in 2011 indicated that capsids properly mature and become stabilized over time (10-day to 20-day) (Conway et al.

**Table 7** Ramachandran plot and Qmean results for the selected and reference sequences

	Phyre <sup>2</sup>		(PS)2-v2		I-TASSER	
	Ramachandran <sup>a</sup>	Qmean	Ramachandran <sup>a</sup>	Qmean	Ramachandran <sup>a</sup>	Qmean
Ref	86%,10.2%	-7.34	95.4%,4.2%	-3.73	79.2%, 14.4%	-8.79
1	85.5%,10.3%	-7.6	95.1%,4.4%	-3.51	80.3%, 14.0%	-7.71
2	85.7%,10.1%	-7.57	95.8%,4.0%	-3.98	81.1%, 14.0%	-6.77
3	85.7%,10.1%	-7.57	95.6%,4.0%	-3.72	81.9%, 14.0%	-6.12

<sup>a</sup> Percentage of favored, and allowed regions



**Fig. 2** Coverage of the predicted tertiary structures by 3 reliable software. The coverage of “I-TASSER” was 100%; it was 90% for “Phyre<sup>2</sup>”, and around 85% for “(PS)2-v2”

2011). In addition, several individual L1 cysteine residues (428, 185, and 175) have an indispensable role in this process. Similar to previous studies, the present results confirmed several cysteine bonds and found some bonds which were omitted as a result of some mutations occurred in such regions in different selected sequences. However, eight cysteine residues (13, 211, 371, 128, 405, 172, 255 and 371) were found in all selected and reference sequences which have the critical role in the structure of the L1 protein. The difference between the positions found in the present study and previous studies may be related to the different geographical regions, the different methods which used, and the different HPV genotypes.

It can be suggested that the disulfide bond 13–13, as well as all other disulfide bonds, may play critical roles in constructing L1 pentamer which has interaction with L2 proteins.

Although our results showed 4 conserved glycosylation positions in all sequences, based on previous studies we could not find any function for glycosylated regions. Zhou et al. in 1993 showed while the majority of L1 protein localizes in the cell nucleus, glycosylated L1 remains in the endoplasmic reticulum and it is not exported from the cell nor translocated to the cell membrane or the cell nucleus (Zhou et al. 1993). Therefore, they concluded that glycosylated L1 was not important in the construction of papillomavirus virion. While it was suggested that N-linked glycosylation played a significant role in VLP binding cells, in 1999 Joyce et al. could not find any concentration dose of tunicamycin to measure the effect of glycosylation (Joyce et al. 1999).

Two conserved amino acid residues (Threonine(T) 340 and T129) were introduced as phosphorylation positions on L1

protein by Buck et al. in 2013 (Buck et al. 2013). In addition, Xi et al. showed that L1 was phosphorylated only in the earlier weeks and this seems to be fairly unstable (Xi and Banks 1991). The present results indicated several phosphorylation sites that were conserved among all selected sequences and the reference sequence.

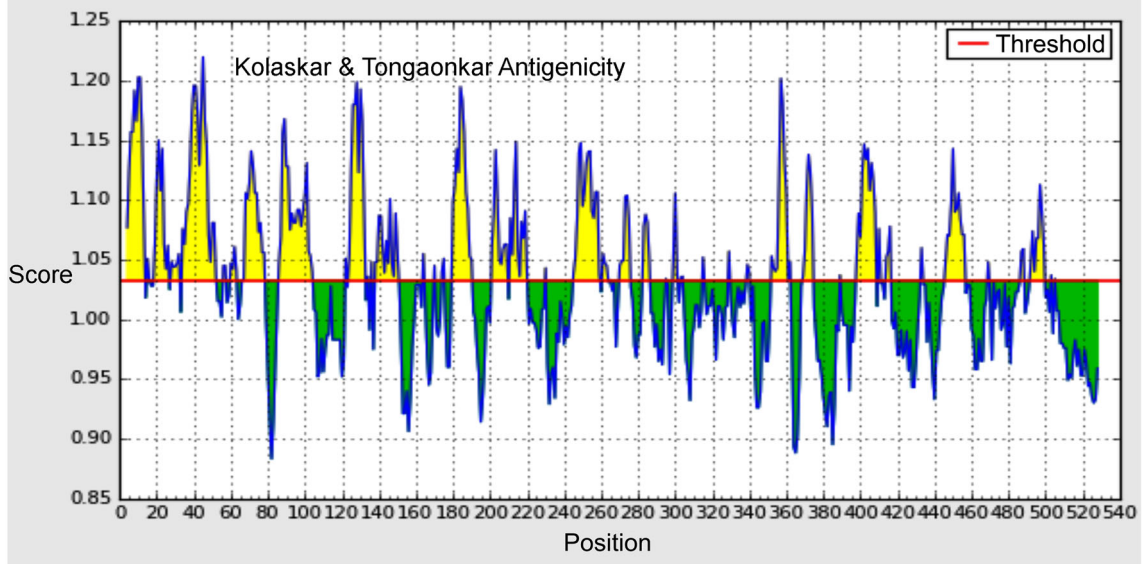
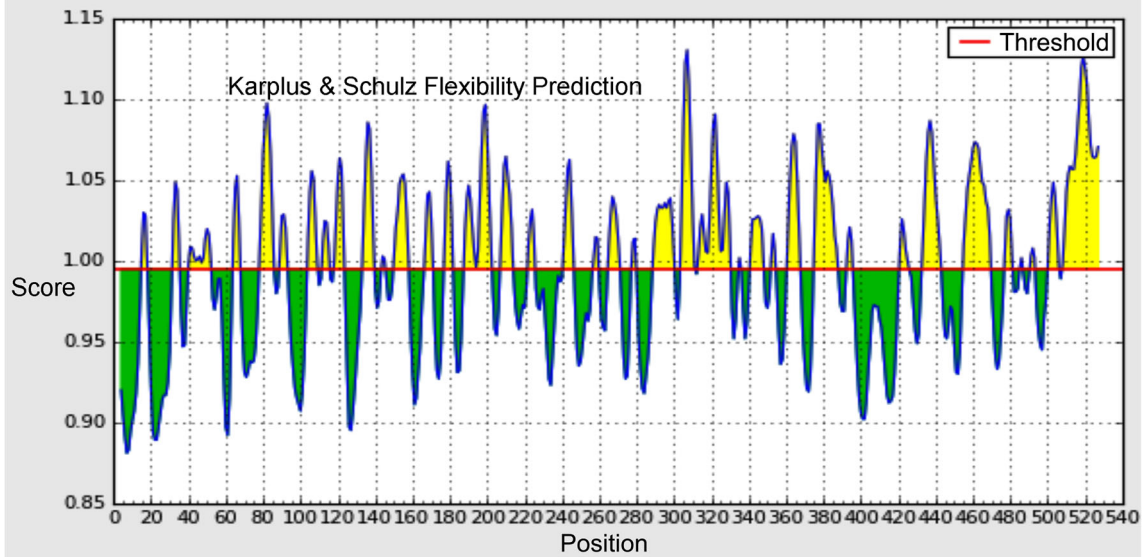
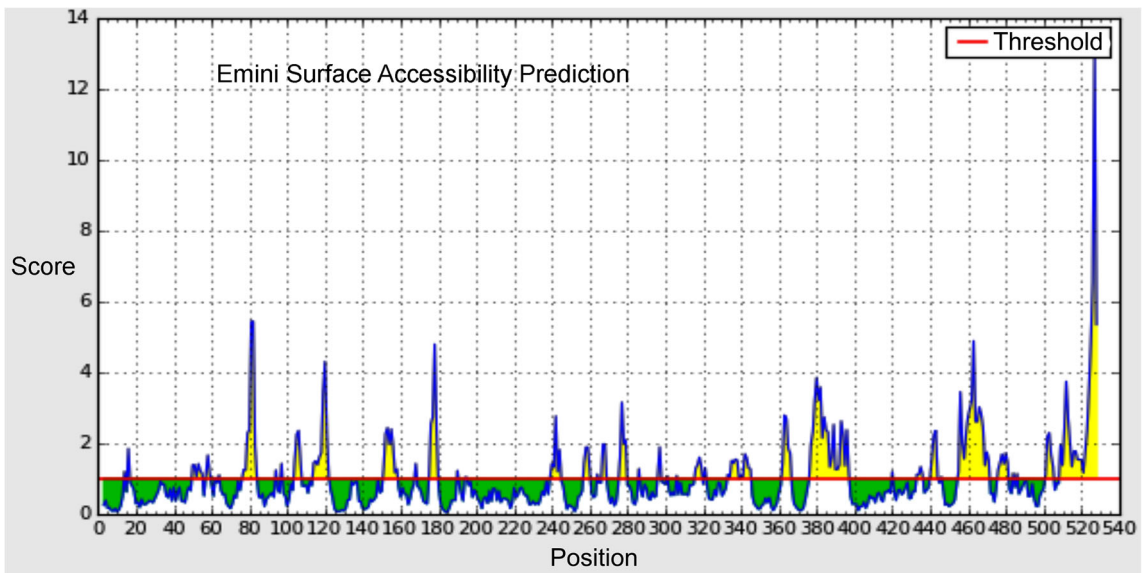
Bioinformatics has provided a great context to define the structure of virus proteins (Behzad et al. 2019; Dehghani et al. 2017, 2019a, b; Moattari et al. 2015; Zahra et al. 2020). Bioinformatics tools were employed widely in this research to determine the structure of L1 and illustrated that the majority of the secondary structures belonged to the random coil and extended strand. The tertiary structure prediction by using three reliable software showed (PS)2-v2 could be able to construct a more reliable structure with the highest percentage of amino acid residues in the favored region in Ramachandran plot and the highest Qmean score. However, the predicted structure by (PS)2-v2 just covered 85% of the L1 sequence and by considering this fact we can conclude that I-TASSER predicted a more useful structure while it covered the complete sequence. In addition, Phyre<sup>2</sup> showed the lower accuracy in comparison with (PS)2-v2 and its’ coverage was around 85%.

In our study, by using “discovery studio”, the structural analysis showed five distinct loops on L1 protein and the related sequences for each loop. Some mutations were located in these loops, i.e. T202 N and N207 T in EF, T292A in FG, and T379P in H1, which may have a significant effect on the loops structure and the antibodies response.

Carter et al. showed multiple neutralizing epitopes on the HPV virion surface, three loops of which (DE, FG, and HI) were most important and vital for binding by neutralizing antibodies (Carter et al. 2006). The crystal structures of four L1 pentamers were determined by Bishop et al. in 2007 (Bishop et al. 2007). They determined that the surface loops contain the known epitopes for neutralizing monoclonal antibodies (NmAbs). Bissett et al. found the role of the DE loop and the late region of the FG loop as neutralizing antibodies epitopes while their findings were critical to describe vaccine-induced cross-neutralizing antibodies which can play a vital role in vaccine-induced cross-protection (Bissett et al. 2012). Roth et al. by using monoclonal antibodies (mAbs)

**Table 8** Five identified loops on the L1 protein and the regions related to each loop. Codons which were mutated in selected sequences were bolded

Region	Loop
76-FPIKKPNNNKILVPA-89	BC or A
157-NASAYAANAGVDNR-170	DE or B
198-GSPCTNVAVNPGDCPP-213	EF or C
292-TVGENVPDDLYIKGSGSTANLASSN-316	FG or D
374-ISTSETTYK-382	H1 or E





**Fig. 3** Propensity scale plots of L1, Flexibility, Hydrophilicity, and Surface accessibility. The Horizontal red line is the threshold. Yellow colors, above the threshold, indicate favorable regions consisting of higher scored residues

recognized three loops (BC, DE, and FG) as the main epitopes and showed the majority of neutralizing epitopes are located on the tip of the capsomere that were related to these variable loops; this data is crucial for designing intertypic HPV vaccines (Roth et al. 2006).

Comparison between the identified loops, hydrophilicity, flexibility, and surface accessibility revealed that BC, DE, and H1 loops were located in regions with the highest surface accessibility scores. In addition, the highest flexibility scores belonged to DE, EF, and FG loops and 4 loops (BC, DE, FG, and H1) had high hydrophilicity scores. Based on three indices (hydrophilicity, flexibility, and surface accessibility), we found BC, DE, FG, and H1 were the highest antigenic ability among other regions of the L1 protein. The combination of B cell epitope prediction and antigenic properties clarified that BC and FG were the most capable regions of L1 with the highest ability to apply in recombinant vaccines against HPV infections (Table 9). By considering the list of mutations, the BC loop was completely conserved and just one codon (292) in FG loop showed substitution among selected sequences.

Humoral immunity has always the key role to control virus infections (Ayorloo et al. 2015; Alborzi et al. 2017; Negahdaripour et al. 2017). B cell epitope prediction showed 6 potential regions, 3 of which were completely consistent with 3 loops: BC, EF, and FG that illustrated the importance of these loops in humoral response against HPV L1 protein. In comparison with previous studies, it can be concluded that FG is the most reported loop which has tremendous potential as an epitope for neutralizing antibodies. It seemed that the mentioned mutations did not have a significant effect on B cell epitope regions and their scores; however, substitution in amino acid 202 had a positive impact on EF loop and increased the score of this region as a potential B cell epitope.

**Table 9** The comparison among the all identified loops by considering hydrophilicity, flexibility, surface accessibility, and B cell epitope prediction

Loops	Surface accessibility	Flexibility	Hydrophilicity	B-cell epitope	Final selection
BC	*		*	*	*
DE	*	*	*		
EF		*		*	
FG		*	*	*	*
H1	*		*		

Finally, two loops BC, and FG were selected as the most capable regions. \*: indicates the more capable loop in each parameter

To conclude, the present results suggested two regions in HPV L1 protein with high potential to induce immune system which can be used as new candidates for new HPV vaccines. In spite of several substitutions in the studied sequences, they had an insignificant impact on all features of L1 protein; it can be inferred that L1 has a conserved sequence and the existing vaccines are still functional for Iranian patients. However, monitoring mutations is crucial to examine the effectiveness and efficacy of the vaccines. Structural and immune properties of Iranian L1 showed the vital role of 5 loops which lead to a strong immune response against HPV infections.

**Acknowledgements** The authors would like to thank Shiraz University of Medical Sciences, Shiraz, Iran for support (Grant number: 97-01-59-18192), and also Center for Development of Clinical Research of Nemazee Hospital and Dr. Nasrin Shokrpour for editorial assistance and Jean Krugman, retired professor of General Studies, DeVry University, Decatur, Georgia, USA for help in revising the English of the manuscript. Special thanks to Nooshin Zare, Anahid khaanoomi Hasanshahi and Rozha Rezaei for their kind assistance and also Dr. Marjan Zare and Dr. Behnam Honarvar for statistical and design consultation, respectively.

**Compliance with ethical standards**

**Conflict of interest** Author Behzad Dehghani declares that he has no conflict of interest. Author Zahra Hasanshahi declares that she has no conflict of interest. Author Tayeb Hashempour declares that she has no conflict of interest. Author Mohamad Motamedifar declares that he has no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**References**

Ayorloo M et al. (2015) Detection of specific antibodies to HCV-ARF/CORE+ 1 protein in cirrhotic and non-cirrhotic patients with hepatitis C: a possible association with progressive fibrosis archives of Iranian Medicine (AIM) 18(5): 304-307

Alborzi A, Hashempour T, Moayedi J, Musavi Z, Pouladfar G, Merat S (2017) Role of serum level and genetic variation of IL-28B in interferon responsiveness and advanced liver disease in chronic hepatitis C patients. *Med Microbiol Immunol* 206:165–174. <https://doi.org/10.1007/s00430-017-0497-y>

Bazan SB, Chaves AAM, Aires KA, Cianciarullo AM, Garcea RL, Ho PL (2009) Expression and characterization of HPV-16 L1 capsid protein in *Pichia pastoris*. *Arch Virol* 154:1609. <https://doi.org/10.1007/s00705-009-0484-8>

Beckmann CR et al (2010) *Obstetricia y ginecología*. Wolters Kluwer,

Behzad D, Tayebeh H, Zahra H (2019) Interaction of human herpesvirus 8 viral interleukin-6 with human interleukin- 6 receptor using in silico approach: the potential role in HHV-8 pathogenesis. *Curr Proteomics* 16:1–10. <https://doi.org/10.2174/1570164616666190626151949>

Benkert P, Tosatto SC, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. *Proteins* 71:261–277. <https://doi.org/10.1002/prot.21715>

Bishop B, Dasgupta J, Klein M, Garcea RL, Christensen ND, Zhao R, Chen XS (2007) Crystal structures of four types of Human Papillomavirus L1 Capsid Proteins understanding the specificity of

- neutralizing monoclonal antibodies. *J Biol Chem* 282:31803–31811. <https://doi.org/10.1074/jbc.M706380200>
- Bissett SL et al (2012) Human papillomavirus antibody reference reagents for use in postvaccination surveillance serology. *Clin Vaccine Immunol* 19:449–451. <https://doi.org/10.1128/CVI.05641-11>
- Blom N, Gammeltoft S, Brunak S (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294:1351–1362. <https://doi.org/10.1006/jmbi.1999.3310>
- Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S, Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4:1633–1649. <https://doi.org/10.1002/pmic.200300771>
- Bruni L AG, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de Sanjosé S (2019) CO/IARC information Centre on HPV and Cancer (HPV information Centre). Human Papillomavirus and Related Diseases in the World Summary Report
- Buck CB, Day PM, Trus BL (2013) The papillomavirus major capsid protein L1. *Virology* 445:169–174
- Carter JJ, Wipf GC, Madeleine MM, Schwartz SM, Koutsky LA, Galloway DA (2006) Identification of human papillomavirus type 16 L1 surface loops required for neutralization by human sera. *J Virol* 80:4664–4672. <https://doi.org/10.1128/JVI.00537-17>
- Chauhan JS, Rao A, Raghava GP (2013) In silico platform for prediction of N-, O- and C-glycosites in eukaryotic protein sequences. *PLoS One* 8:e67008. <https://doi.org/10.1371/journal.pone.0067008>
- Chen C-C, Hwang J-K, Yang J-M (2006) 2: protein structure prediction server. *Nucleic Acids Res*. <https://doi.org/10.1093/nar/gkl187>
- Cheng J, Randall AZ, Sweredoski MJ, Baldi P (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Res* 33:W72–W76. <https://doi.org/10.1093/nar/gki396>
- Conway MJ, Cruz L, Alam S, Christensen ND, Meyers C (2011) Differentiation-dependent interprotameric disulfide bond stabilizes native human papillomavirus type 16. *PLoS One* 6:e22427. <https://doi.org/10.1371/journal.pone.0022427>
- Dadashi M et al (2017) Epidemiology of human papillomavirus (HPV) infection among Iranian women identified with cervical infections: a systematic review and meta-analysis of national data. *Infect Epidemiol Microbiol* 3:68–72. <https://doi.org/10.5812/ijm.40855>
- Dehghani B et al (2017) Functional and structural characterization of Ebola virus glycoprotein (1976–2015)—an in silico study. *Int J Biomath* 10:1750108. <https://doi.org/10.1142/S179352451750108X>
- Dehghani B, Hashempour T, Hasanshahi Z (2019a) Using immunoinformatics and structural approaches to design a novel HHV8 vaccine. *Int J Peptide Res Therap*: 1-11. <https://doi.org/10.1007/s10989-019-09839-x>
- Dehghani B, Hashempour T, Hasanshahi Z, Moayed J (2019b) Bioinformatics analysis of domain 1 of HCV-Core protein: Iran international journal of peptide research and therapeutics:1-18. <https://doi.org/10.1007/s10989-019-09838-y>
- Doorbar J, Egawa N, Griffin H, Kranjec C, Murakami I (2015) Human papillomavirus molecular biology and disease association. *Rev Med Virol* 25:2–23. <https://doi.org/10.1002/rmv.1822>
- Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines *BMC bioinformatics* 8:4. <https://doi.org/10.1186/1471-2105-8-4>
- Farjadian S, Asadi E, Doroudchi M, Dehaghani AS, Tabei S, Kumar V, Ghaderi A (2003) High risk HPV types in southern Iranian patients with cervical cancer. *Pathol Oncol Res* 9:121–125. <https://doi.org/10.1007/BF03033756>
- Ferré F, Clote P (2006) DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res* 34:W182–W185. <https://doi.org/10.1093/nar/gkl189>
- Fleury MJ, Touzé A, Coursaget P (2014) Human papillomavirus type 16 pseudovirions with few point mutations in L1 major capsid protein FG loop could escape actual or future vaccination for potential use in gene therapy. *Mol Biotechnol* 56:479–486. <https://doi.org/10.1007/s12033-014-9745-1>
- Frank K, Sippl MJ (2008) High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics* 24:2172–2176. <https://doi.org/10.1093/bioinformatics/btn422>
- Frazer IH (2009) Interaction of human papillomaviruses with the host immune system: a well evolved relationship *virology* 384:410–414. <https://doi.org/10.1016/j.virol.2008.10.004>
- Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExpASY server. *The proteomics protocols handbook*. Springer, In, pp 571–607
- Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11:681–684
- Ghaffari SR et al (2006) Prevalence of human papillomavirus genotypes in women with normal and abnormal cervical cytology in Iran. *Asian Pac J Cancer Prev* 7:529–532
- Gupta R, Jung E, Brunak S (2017) Prediction of N-glycosylation sites in human proteins. 2004 preparation available online: <http://www.cbs.dtu.dk/services/NetNGlyc>
- Haghshenas MR, Mousavi T, Moosazadeh M, Afshari M (2016) Human papillomavirus and breast cancer in Iran: a meta-analysis. *Iran J Basic Med Sci* 19:231. <https://doi.org/10.22038/IJBMS.2016.6640>
- Heo L, Park H, Seok C (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic Acids Res* 41:W384–W388. <https://doi.org/10.1093/nar/gkt458>
- Hiller K, Grote A, Scheer M, Münch R, Jahn D (2004) PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res* 32:W375–W379. <https://doi.org/10.1093/nar/gkh378>
- Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK (2004) The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32:1037–1049. <https://doi.org/10.1093/nar/gkh253>
- Ishii Y, Tanaka K, Kanda T (2003) Mutational analysis of human papillomavirus type 16 major capsid protein L1: the cysteines affecting the intermolecular bonding and structure of L1-capsids. *Virology* 308:128–136. [https://doi.org/10.1016/s0042-6822\(02\)00099-5](https://doi.org/10.1016/s0042-6822(02)00099-5)
- Jamdar F, Farzaneh F, Navidpour F, Younesi S, Balvayeh P, Hosseini M, Ghodssi-Ghasemabadi R (2018) Prevalence of human papillomavirus infection among Iranian women using COBAS HPV DNA testing. *Infect Agent Cancer* 13:6. <https://doi.org/10.1186/s13027-018-0178-5>
- Jespersen MC, Peters B, Nielsen M, Marcotili P (2017) BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 45:W24–W29
- Joyce JG et al (1999) The L1 major capsid protein of human papillomavirus type 11 recombinant virus-like particles interacts with heparin and cell-surface glycosaminoglycans on human keratinocytes. *J Biol Chem* 274:5810–5822. <https://doi.org/10.1074/jbc.274.9.5810>
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4:363. <https://doi.org/10.1038/nprot.2009.2>
- Kim SM (2016) Human papilloma virus in oral cancer. *Journal of the Korean Association of Oral and Maxillofacial Surgeons* 42:327–336. <https://doi.org/10.5125/jkaoms.2016.42.6.327>
- Lee H et al (2008) Expression of HPV L1 capsid protein in cervical specimens with HPV infection. *Diagn Cytopathol* 36:864–867. <https://doi.org/10.1002/dc.20922>
- Li M, Beard P, Estes PA, Lyon MK, Garcea RL (1998) Intercapsomeric disulfide bonds in papillomavirus assembly and disassembly. *J Virol* 72:2160–2167
- Moattari A, Dehghani B, Khodadad N, Tavakoli F (2015) In Silico functional and structural characterization of H1N1 influenza A viruses Hemagglutinin, 2010–2013. Shiraz, Iran *Acta biotheoretica* 63:183–202. <https://doi.org/10.1007/s10441-015-9260-1>

- Monie A, Hung C-F, Roden R, Wu TC (2008) Cervarix™: a vaccine for the prevention of HPV 16, 18-associated cervical cancer. *Biologics: targets & therapy* 2:107
- Mortazavi S, Zali M, Raoufi M, Nadji M, Kowsarian P, Nowroozi A (2002) The prevalence of human papillomavirus in cervical cancer in Iran. *Asian Pac J Cancer Prev* 3:69–72
- Negahdaripour M et al (2017) A novel HPV prophylactic peptide vaccine, designed by immunoinformatics and structural vaccinology approaches infection. *Genetics and Evolution* 54:402–416. <https://doi.org/10.1016/j.meegid.2017.08.002>
- Pande S et al (2008) Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in biopsy samples from cervical cancer patients in North India. *J Clin Microbiol* 46:1060–1066. <https://doi.org/10.1128/JCM.02202-07>
- Rahimifar S, Erfani N, Sarraf Z, Ghaderi A (2010) cta-4 gene variations may influence cervical cancer susceptibility. *Gynecol Oncol* 119:136–139. <https://doi.org/10.1016/j.ygyno.2010.06.006>
- Roth SD, Sapp M, Streeck RE, Selinka H-C (2006) Characterization of neutralizing epitopes within the major capsid protein of human papillomavirus type 33. *Virology* 33:83. doi: <https://doi.org/10.1186/1743-422X-3-83>
- Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5:725. <https://doi.org/10.1038/nprot.2010.5>
- Saha S, Raghava G (2004) BcePred: prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties. In: *International Conference on Artificial Immune Systems*, Springer, pp 197–204
- Saha S, Raghava G (2006a) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 34:W202–W209. <https://doi.org/10.1093/nar/gkl343>
- Saha S, Raghava G (2006b) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network proteins: structure, function, and Bioinformatics 65:40–48. <https://doi.org/10.1002/prot.21078>
- Salehi-Vaziri M, Sadeghi F, Hashemi FS, Haeri H, Bokharaci-Salim F, Monavari SH, Keyvani H (2016) Distribution of human papillomavirus genotypes in Iranian women according to the severity of the cervical lesion. *Iran Red Crescent Med J*:18. <https://doi.org/10.5812/ircmj.24458>
- Sapp M, Fligge C, Petzak I, Harris JR, Streeck RE (1998) Papillomavirus assembly requires trimerization of the major capsid protein by disulfides between two highly conserved cysteines. *J Virol* 72:6186–6189
- Shafaghi B, Jarollahi A, Yousefzadeh B, Ameri A, Moghadam S, Mostafavi M (2013) Human papilloma virus prevalence and types among Iranian women attending regular gynecological visits. *Reports of Radiotherapy and Oncology* 1
- Shahramian I, Heidari Z, Mahmoudzadeh-Sagheb H, Moradi A, Forghani F (2011) Prevalence of HPV infection and high risk HPV genotypes (16, 18), among monogamous and polygamous women, in Zabol, Iran. *Iranian J Public Health* 40:113
- Siddiqui MAA, Perry CM (2006) Human papillomavirus quadrivalent (types 6, 11, 16, 18) recombinant vaccine (Gardasil®). *Drugs* 66:1263–1271. <https://doi.org/10.2165/00003495-200666090-00008>
- Touze A, El Mehdaoui S, Sizaret P-Y, Mougins C, Muñoz N, Coursaget P (1998) The L1 major capsid protein of human papillomavirus type 16 variants affects yield of virus-like particles produced in an insect cell expression system. *J Clin Microbiol* 36:2046–2051
- Vera-Bravo R et al (2003) Human papillomavirus type 16 and 18 L1 protein peptide binding to VERO and HeLa cells inhibits their VLPs binding. *Int J Cancer* 107:416–424. <https://doi.org/10.1002/ijc.11433>
- Vita R et al (2018) The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* 47:D339–D343. <https://doi.org/10.1093/nar/gky1006>
- Xi S-Z, Banks LM (1991) Baculovirus expression of the human papillomavirus type 16 capsid proteins: detection of L1–L2 protein complexes. *J Gen Virol* 72:2981–2988. <https://doi.org/10.1099/0022-1317-72-12-2981>
- Xu W-X et al (2016) Epitomics: IgG-epitome decoding of E6, E7 and L1 proteins from oncogenic human papillomavirus type 58. *Sci Rep* 6. <https://doi.org/10.1038/srep34686>
- Yousefzadeh A et al (2014) Human papillomavirus (HPV) prevalence and types among women attending regular gynecological visit in Tehran. *Iran Clin Lab* 3:1
- Zahra M et al (2020) Antibody development to HCV alternate reading frame protein in liver transplant candidate and its computational analysis. *Current Proteomics* 17:1–1. <https://doi.org/10.2174/1570164617666190822103329>
- Zhang W, Carmichael J, Ferguson J, Inglis S, Ashrafi H, Stanley M (1998) Expression of human papillomavirus type 16 L1 protein in *Escherichia coli*: denaturation, renaturation, and self-assembly of virus-like particles in vitro. *Virology* 243:423–431. <https://doi.org/10.1006/viro.1998.9050>
- Zheng Z-M, Baker CC (2006) Papillomavirus genome structure, expression, and post-transcriptional regulation. *Frontiers in bioscience: a journal and virtual library* 11:2286
- Zhou J, Sun XY, Frazer IH (1993) Glycosylation of human papillomavirus type 16 L1 protein. *Virology* 194:210–218. <https://doi.org/10.1006/viro.1993.1251>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.