Journal of Theoretical and Computational Chemistry
Vol. 17, No. 2 (2018) 1850017 (17 pages)

DOI: 10.1142/S0219633618500177

Computational identification of antibody epitopes of human papillomavirus 16 (HPV16) L1 proteins

Xin Tan[*,†], Na Liu[‡,§], Fiona S. Legge[¶], Minghui Yang[‡] and Jun Zeng[¶,‖]

[*]*Department of Obstetrics and Gynecology*
West China Second University Hospital
Chengdu 610041, P. R. China

[†]*Key Laboratory of Birth Defects and Related*
Diseases of Women and Children
Sichuan University Ministry of Education
Chengdu 610041, P. R. China

[‡]*Key Laboratory of Magnetic Resonance*
in Biological Systems
National Center for Magnetic Resonance
and Atomic and Molecular Physics
Wuhan Institute of Physics and Mathematics
Chinese Academy of Sciences
Wuhan 430071, P. R. China

[§]*University of Chinese Academy of Sciences*
Beijing 100049, P. R. China

[¶]*MedChemSoft Solutions Wheelers Hill*
VIC 3150, Australia
[‖]*Jun.Zeng@medchemsoft.com*

Received 8 January 2018
Accepted 26 February 2018
Published 22 March 2018

Previously, we developed a method to predict epitopes on a protein recognized by specific antibodies. In this study, we have applied this method to identify the epitopes of the human papillomavirus 16 (HPV16) L1 capsomer that is bound by monoclonal antibodies U4, AE3 and AG7. Initially, the method was validated by the identification of epitopes of HPV16 L1 capsomer that bind to antibody U4. Our predicted epitopes were in agreement with the cryto-electron microscopy (cryto-EM) structure of the complex. The method was then used to predict the epitopes of HPV16 L1 binding of antibodies AE3 and AG7. Our calculations indicated that antibody AE3 binds to the HPV16 L1 capsomer at two different regions. Firstly, the region recognized by antibody U4 and secondly, the region recognized by antibody V5, which have been shown in the cryto-EM structure of the V5 and HPV16 L1 complex. In comparison, the antibody AG7 binds to the capsomer only at the epitopes bound by antibody U4. Therefore, antibody AE3 is predicted to have higher affinity than antibody AG7 and could be used for

‖Corresponding author.

developing highly efficient anti-HPV monoclonal antibodies in the clinical treatment of HPV infections.

Keywords: Computational epitope design; multiple copy simultaneous search; vaccines; antibody-antigen recognition.

1. Introduction

Human papillomavirus (HPV) infections have been a significant health challenge worldwide for many years.[1,2] Although commercial vaccines targeting the viral capsid proteins have been developed and used successfully to protect against high-risk HPVs, their efficacy is genotype specific and provides little benefit against existing infections.[3] Identification of epitopes on the capsid protein structure is thus important to develop improved recombinant vaccines that maximize effective and long-term antibody-mediated protection against multiple HPV types.

Papillomavirus capsids are composed of 360 copies of L1 structural protein and up to 72 copies of L2 minor structural protein.[4,5] Each capsomer is formed by five L1 proteins, and 72 capsomers make up one capsid. The C-terminus of each L1 protein extends along the capsid floor. A disulfide bond between Cys428 and Cys175 stabilizes the capsid structure and is important in virus maturation.[6,7] The core of the capsomer is composed of the common viral motif, the anti-parallel β-strands which are connected by surface loops (as shown in Fig. S1 in supplementary materials). Nearly all the conformational epitopes are located on one or more of these outwardly facing surface-exposed loops, having been studied by antibody binding and neutralizing assays, hybrid virus loop exchange studies, and structural analysis from cryto-electron microscopy (cryto-EM) studies.[8,9] However, a more recent cryto-EM study on the complex of HPV16 L1 capsomer and antibody U4 revealed a different recognition region at the canyons between capsomer knobs, including an epitope involving the L1 C-terminal invading arm.[9] The surface loops are poorly conserved caused by selective pressure for HPV to accumulate mutations that prevent the binding of neutralizing antibodies raised by prior infection. Therefore, epitopes that bind to antibody U4 are potentially important for generation and selection of highly efficient anti-HPV L1 monoclonal antibodies crucial for the clinical diagnosis and treatment of HPV infection.

Understanding how these antibodies bind to their antigens is essential for developing therapeutic antibodies and vaccines for viral infections. Previously, we developed a new approach to predict linear antibody-binding epitopes of antigens.[10–12] This approach has been successfully used for the prediction of the epitopes on a bunyavirus glycoprotein ectodomain recognition of severe fever with thrombocytopenia syndrome (SFTS) virus to its human antibody Mab 4–5,[10] Shiga Toxin 2 (Stx2) subunit A to its specific antibodies 11E10 and S2C2,[11] and the Dengue virus NS1 protein to its specific antibodies.[12] Briefly, it consists of three steps. (1) Identification of the locations of chemical functional groups on key regions of the antibody using the "multiple copy simultaneous search" (MCSS)

approach.[13–19] (2) MCSS clusters of a specific functional group with favorable interaction energies with the protein selected to derive the key pattern of functional groups on the surface of the antigen. These patterns are subsequently converted into the amino acid sequence pattern. (3) The sequence pattern is used to search the antigen protein sequence, and the peptides with the highest scores matching the key pattern are considered to be binder peptides.[10–12] This method is considered an extension of our computational combinatorial inhibitor design (CCLD) approach,[13,16–18] which has been used previously to design peptide inhibitors of Ras–Raf interactions.[18]

Two new monoclonal antibodies (AE3 and AG7) have been recently developed against human papillomavirus type 16 (HPV16) L1 protein.[20] However, details of their binding to HPV16 are unknown. In this work, we use the computational method developed previously to predict the binding epitopes of HPV16 L1 for these two antibodies. Firstly, we verified our method by identifying the epitopes of HPV16 L1 protein that are recognized by antibody U4.[9] The backbone structure of the complex between U4 and HPV16 L1 capsomer has been solved by cryto-EM, so that the binding epitope is known.[8,9] Secondly, we applied our approach to predict epitopes of HPV16 L1 for the two antibodies AE3 and AG7.

2. Methods

2.1. *Homology modeling of the antibodies*

The sequences of the three antibodies U4, AE3 and AG7 were aligned and shown in Fig. S1 of Supplementary Material. The sequence similarities of these antibodies are 68.5% between U4 and AE3, 69.4% between U4 and AG7, and 74.3% between AE3 and AG7, respectively. The sequences were also used to search for the closest related antibody with known 3D structure using a BLAST search against sequences of proteins deposited in the protein data bank (http://blast.ncbi.nlm.nih.gov). The complex structure between U4 antibody and HPV16 L1 capsomer has been solved by cryto-EM study. However, at the low resolution of 12 Å, only backbone structures were revealed (PDB entry: 3JBA).[9] To obtain the full structure of antibody U4, a homology model was built using the single-chain Fv (scFv) fragment of an anti-ErbB2 antibody chA21 as the template (PDB entry: 3H3B).[21] The VL and VH domains of U4 show sequence identities of 84.3% and 44.0% to 3H3B, respectively. For the AE3, the best matching antibody sequence identified was the crystal structure of anti-polysialic acid antibody single chain Fv fragment (mAb735) (PDB entry: 3WBD).[22] The amino acid identities of the VL and VH domains of the AE3 were 76.8% and 63.6% to 3WBD, respectively. As for the AG7, the template used was the monoclonal antibody 2D10 in complex with DBP (PDB entry: 5F3J).[23] The amino acid identities of the VL and VH domains of the AG7 were 83.1% and 75.0% to 5F3J, respectively. All the model constructions were carried out using the Swissport modeling service (http://swissmodel.expasy.org).[24–27]

2.2. *MCSS of functional groups*

The "MCSS" method has been described previously.[13–19] Using the homology models of the antibodies, our Qu-Cbit software (http://www.medchemsoft.com) implementing our MCSS approach[10–12] was used to identify the minima locations of functional groups on the binding surfaces, i.e. the "Complementary Determining Regions" (CDRs). Eleven functional groups were used, each of which corresponds to the side chains of different amino acids (Table 1). We use CHARMM22 all-hydrogen atom force field[28] for both protein and functional groups.

Three hundred replicas of each functional group were randomly placed inside a sphere with a 12 Å radius around the CDRs of the antibodies. Details of the CDR loop conformations have been demonstrated to be insignificant for the distribution of MCSS minima and on the sequence pattern derived from the minima.[10] We use only single conformation of the CDRs.[10] The CDRs are defined by (Ile2, Leu29, Tyr38, Trp56, Ser73, Tyr97, Pro101, Leu102) of L chain and (Met33, Ser49, Ser52, Tyr96, Phe105) of H chain for the U4; (Val2, Leu29, Tyr37, Leu55, Ser72, Gly96, Pro100, Arg101) of L chain and (Tyr33, Arg50, Asp52, Gly100) of H chain for AE3; and (Ile2, Tyr37, Arg55, Ser72, His96, Leu97, Pro100, Phe101) of L chain and (Asn50, Tyr54, Asn100) of H chain for AG7. A 500-step MCSS was performed. During all the MCSS calculations, each replica only interacts with a target protein, and not with the other replicas. The nonbonded interaction was truncated at 20 Å. The dielectric constant was set to 10 to mimic solvent screening effects.[29]

2.3. *Identification of sequence pattern*

Interaction energy of −10.00 kcal/mol was used as the threshold for the minima of polar and apolar functional groups for the antibodies U4, AE3 and AG7. For the charged groups ACET, MAMM and MGUA, and bulky group INDO, a threshold of −15.00 kcal/mol was used due to large electrostatic interactions arising from charged groups and strong hydrophobic packing from INDO group. The spatial patterns of the locations of the MCSS minima on the surface of the antibody were converted into a sequence pattern according to the relationship between the functional groups and amino acids as given in Table 1. This sequence pattern will serve as the fingerprint to identify the epitopes of antigens.

2.4. *Search for epitopes based on the sequence pattern*

The sequence pattern obtained using the method described in Sec. 2.3 was used to identify candidate peptides from the sequence of HPV16 L1 protein. The details of the method have been described previously.[12] Briefly, L1 protein sequence was sliced into 7-mer peptide libraries several times, starting from residue 1 up to 7. Therefore, seven libraries of 7-mer peptides were obtained and each peptide library was checked for sequence matches with the key pattern. The peptides with a sequence that matches the key pattern were deemed to be binder peptides and residues occurring in

Table 1. The relationship between functional groups and amino acids.

	Functional group	Abbreviation	Amino acids
Charged (−)	Acetate ion	ACET	ASP, GLU
Charged (+)	Methylguanidinium	MGUA	ARG
Charged (+)	Methylammonium	MAMM	LYS
Polar	Acetamide	ACEM	ASN, GLN
Polar	Methanol	MEOH	SER, THR
Hydrophobic	Methanethiol	MESH	CYS, MET
Aromatic Polar	Phenol	PHEN	TYR
Aromatic Polar	Indole	INDO	TRP
Aromatic Polar	Imidazole	IMIA	HIS
Aromatic Hydrophobic	Benzene	BENZ	PHE
Hydrophobic	Ibutane	IBUT	VAL, ILE, LEU

binder peptides from more than three libraries were considered part of the epitope. Therefore, the epitopes predicted vary in their length.

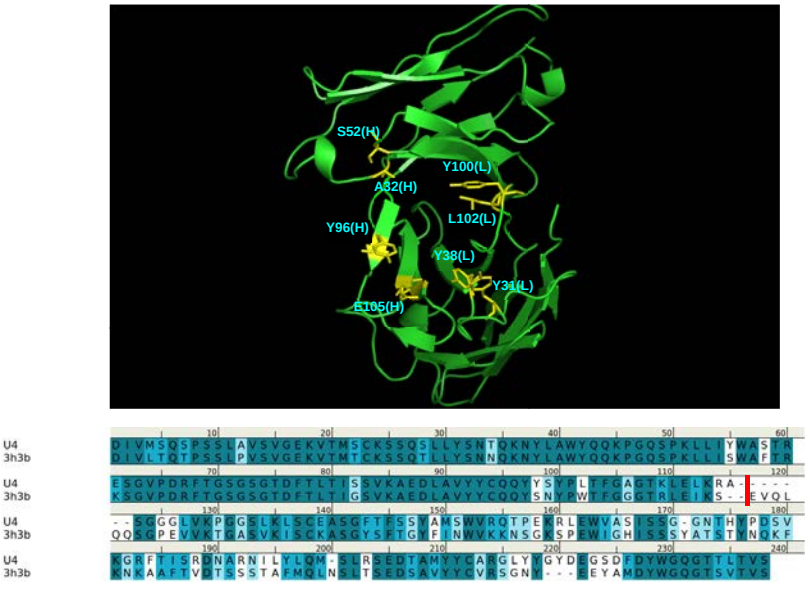
Searching the sequence of the antigen only represents linear epitopes on the antigen. For HPV L1, it is observed that the L1 protein forms capsomers on the cell surface[30]; some predicted epitopes could be either buried inside the L1 protein or exposed to the capsomer pore, so that these peptides would be unlikely to bind antibodies. Using the cryto-EM structure of HPV16 L1 capsomer with antibody U4,[9] we only selected the peptides with significant surface accessibility as the final epitopes.

3. Results

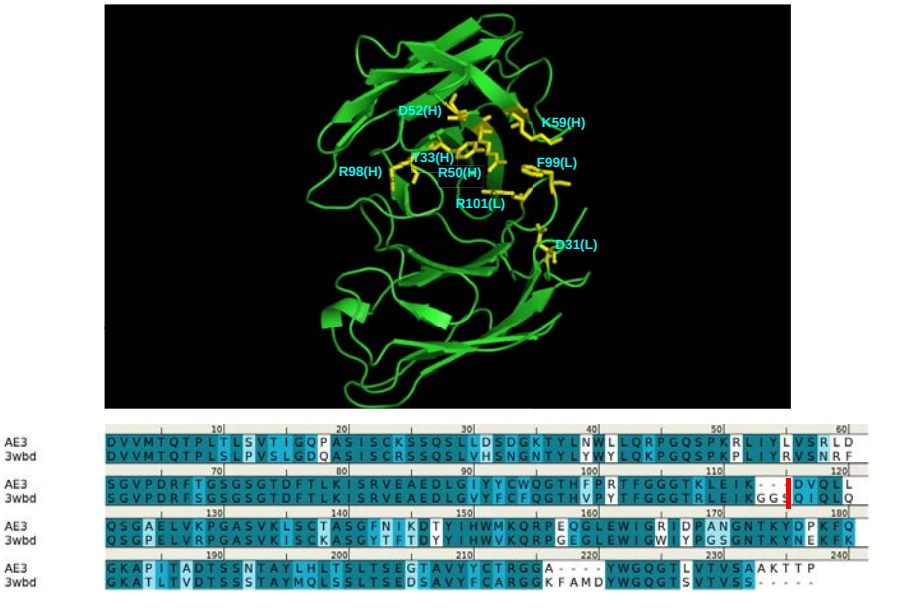
3.1. *Recognition of HPV16 L1 protein by antibody U4*

The antibody structure of U4 was built using the crystal structure of homologous antibody PDB entry 3H3B[21] as template. The model structure and the sequence alignment used for the homology modeling are shown in Fig. 1(a). Two distinct regions S1 and S2 are identified around the CDR3 loop, with a large groove between. Region S1 is formed by H chain residues (Ala32, Tyr96), while region S2 is formed by L chain residues (Tyr38, Tyr100, Leu102). The groove is formed by residues (Gln95, Tyr97) of L chain (not shown in Fig. 1(a)) and (Ser40, Glu105) of H chain. The bottom of the groove (Tyr99 of H chain) is ca. 8 Å inside protein from the surface; the groove is thus too deep to be accessed by the side chains of the antibody. Therefore, the minima inside the groove are discarded.

Figure 2(a) shows the distribution of MCSS minima of functional groups on the surface of U4. Overall, the distribution of the MCSS minima closely corresponds to the physical properties of the surface with the minima forming specific interactions with the residues of U4. The groups identified at regions S1 and S2 were ACET, MGUA, MAMM, and PHEN minima. For the negatively charged ACET group, only two minima were found at S1 region, interacting with Ser52 of H chain with interaction energies of (−12.80, −10.00) kcal/mol, respectively. As the S1 surface is mainly hydrophobic, MGUA and MAMM minima were only observed at region S2.

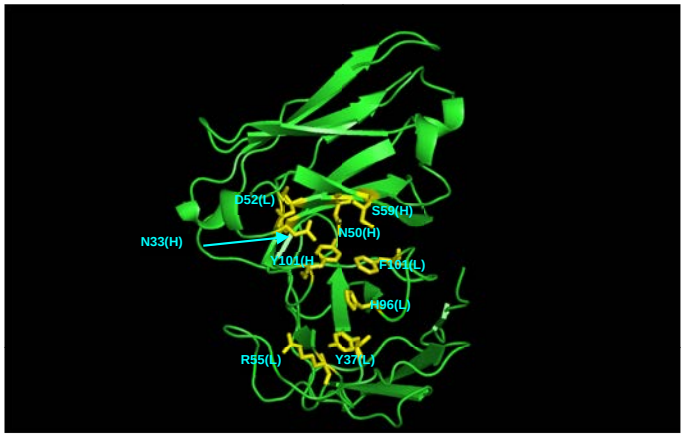


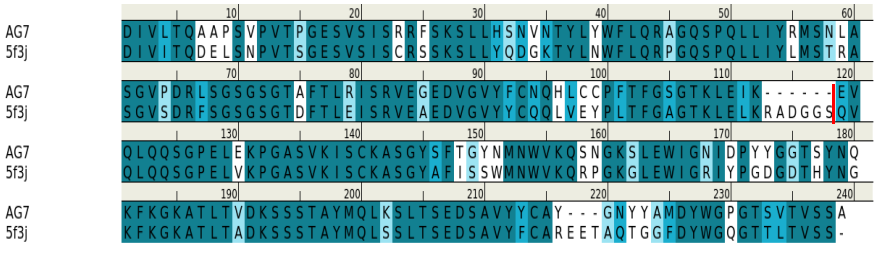
(a)



(b)

Fig. 1. (a) Model structure and sequence alignment used for homology modeling between antibody U4 and its template 3H3B[21]; (b) model structure and sequence alignment used for homology modeling between antibody AE3 and its template 3WBD[22]; (c) model structure and sequence alignment used for homology modeling between antibody AG7 and its template 5F3J.[23] L ad H chains of antibodies in the alignment are separated by red bar. The contributing residues around CDR3 loop are highlighted in stick, colored in yellow, and labeled in cyan. Figures were prepared using PyMOL.[34]





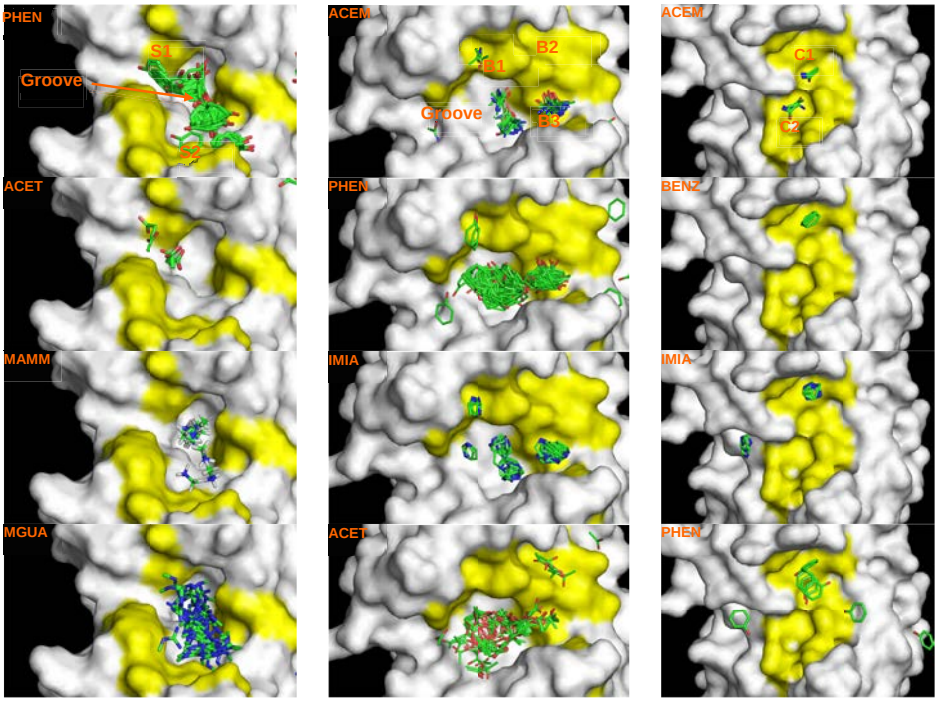
(c)

Fig. 1. (*Continued*)

Thirteen MGUA minima were identified interacting with the negatively charged surface given by L chain residues (Tyr31, Tyr38) and H chain residues (Tyr96, Glu101). The interaction energies were calculated to be (−15.00, −20.40) kcal/mol, respectively. Similarly, eight MAMM minima were located at S2 interacting with the carbonyl oxygen of Tyr97 of L chain with interaction energies of (−15.00, −15.40) kcal/mol, respectively. For the PHEN group, three minima were found at S1, located between Ala32 and Tyr96 of H chain with interaction energies of (−10.80, −11.00, −10.60) kcal/mol, respectively, while five minima were found at S2 forming hydrophobic interactions with Tyr31 and Tyr38 of L chain with interaction energies of (−10.30, −11.10) kcal/mol, respectively.

Using the MCSS minima on the two surfaces S1 and S2, we constructed a sequence pattern for the peptides that could potentially bind to the antibody. As the maximum distance between the two binding sites is approximately 13.5 Å, a gap of two amino acids is introduced. Therefore, the key sequence pattern for the binding epitopes is defined as X–Z, in which X = (R/K,Y), and Z = (D/E,Y), as given in Table 2.

The sequence pattern was then used to search for peptide binders from the peptide libraries derived from the sequence of HPV16 L1. There were four peptides



(a) (b) (c)

Fig. 2. Selected MCSS minima of functional groups on the surface of antibodies U4 (a), AE3 (b) and AG7 (c). The binding surfaces are colored in yellow and labeled in orange. Figures were prepared using PyMOL.[34]

obtained, i.e. **E1**-"DTSFYNPDT", **E2**-"PLLNKLDDT", **E3**-"TSICKYPDYI", and **E4**-"KEYLRHGEE". By examining the peptides inside the cryto-EM of L1 Capsomers, **E1** is found to be exposed to solvent while **E2** and **E4** are located to be interacting with another monomer and thus are disregarded. In fact, epitope **E4** is near the surface loop FG which antibody V5 binds to.[31,32] Figure 3 shows the HPV16 L1 sequence with these epitopes highlighted in lower case and colored in orange, and their positions in the L1 protein structure. Overall, epitope **E3** is located in the interface of the experimental structure of the complex between U4 and HPV16 L1 protein.

3.2. *Prediction of epitopes of HPV16 L1 to antibody AE3*

The antibody structure of AE3 was built using the crystal structure of homologous antibody from PDB entry 3WBD[22] as template. Figure 1(b) shows the model structure and the sequence alignment used for homology modeling. Three regions, B1, B2 and B3, are identified around the CDR3 loop, with a groove located between B1 and B3. Region B1 is formed by H chain residues (Tyr33, Arg50, Asp52)

Table 2. Distribution of key minima and the derived sequence pattern for the binding epitope peptides to antibodies U4, AE3 and AG7. Sequence patterns of "X–Z" [X = R/K,Y and Z = D/E,Y], "X−Z" [X = Y,H,Q/N,D/E and Z = Y,H,Q/N,D/E], and "X−Z" [X = F, Y,H,Q/N and Z = Q/N] were obtained for antibodies U4, AE3 and AG7, respectively.

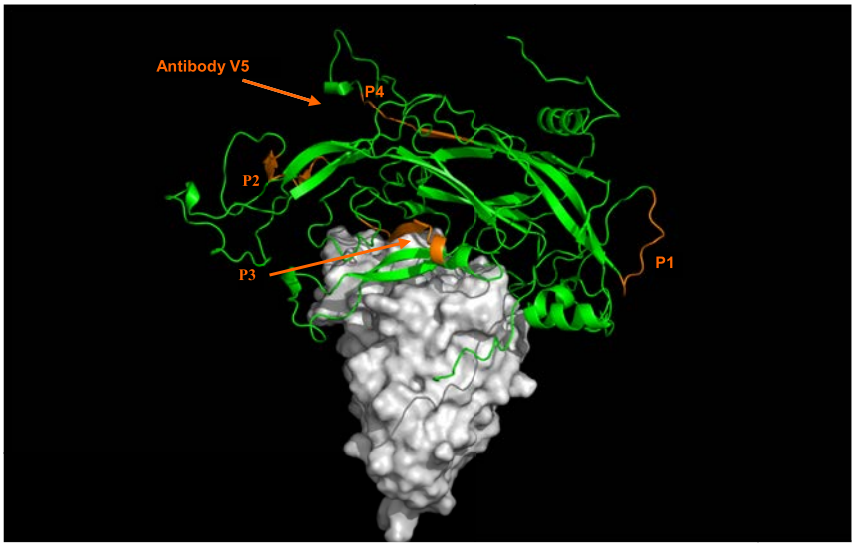
	Antibody U4		
Binding surface	S2		S1
MCSS minima (Sequence) Pattern	MGUA(R) MAMM(K) PHEN (Y)		ACET (D/E) PHEN(Y)
	Antibody AE3		
Binding surfaces	B1	B2	B3
MCSS minima (Sequence) Pattern	PHEN(Y) IMIA (H) ACEM (Q/N)	ACET(D/E)	PHEN(Y) IMIA(H) ACEM(Q/E) ACET (D/E)
	Antibody AG7		
Surfaces	C1		C2
MCSS minima (Sequence) Pattern	PHEN(Y) BENZ(F) IMIA(H) ACEM(Q/E)		ACEM(Q/E)

and L chain residue Arg101; region B2 is formed by H chain residues (Arg50, Lys59) and L chain residues (Phe99, Arg101); and region B3 is formed by H chain residue (Arg50, Asp52) and L chain residues (Asp31, Phe99, Arg101). A deep groove is formed by residues (His35, Gly100) of H chain and residues (Asn39, Leu41, Trp94) of L chain. The bottom of the groove (Trp94 of L chain) is ca. 9.0 Å inside protein from the surface, so that the minima inside the groove are discarded.

ATVYLPPVPVSKVVSTDEYVARTNIYYHAGTSRLLAVGHPYFPIKKPNNNKILVPKVSGLQYRVFRIHLPDPNKFGFPdtsfyn
pdtQRLVWACVGVEVGRGQPLGVGISGHpllnklddtENASAYAANAGVDNRECISMDYKQTQLCLIGCKPPIGEHWGKGSP
CTQVAVQPGDCPPLELINTVIQDGDMVDTGFGAMDFTTLQANKSEVPLDICtsickypdyiKMVSEPYGDSLFFYLRREQMFV
RHLFNRAGTVGENVPDDLYIKGSGSTANLASSNYFPTPSGSMVTSDAQIFNKPYWLQRAQGHNNGICWGNQLFVTVVDT
TRSTNMSLCAAISTSETTYKNTNFkeylrhgeeYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRF
VTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKEKFSADLDQFPLGRKFLLQLGLKAKPKFTLGKRK

(a)

Fig. 3. (a) The predicted epitopes of HPV16 L1 protein to antibody U4 are highlighted in lower case and colored orange in the protein sequence. The peptides identified as binders using the sequence search only are shown in lower case. (b) Backbone presentation of HPV16 L1 complexed with antibody U4 as revealed by cryto-EM (PDB 3JBA).[9] The antibody is displayed as surface in white and predicted epitopes are colored in orange. Figures were prepared using PyMOL.[34]



(b)

Fig. 3. (*Continued*)

Figure 2(b) shows the distribution of MCSS minima of functional groups on the surface of antibody AE3. The surface around the CDR3 loop is highly positively charged due to charged amino acids (Arg30, Arg50, Lys59, Arg98, Arg101) of H chain. Therefore, no MAMM and MGUA minima were found around CDR3 loop. However, the negatively charged ACET group has seven minima at B2 by interacting to residues (Lys59, Arg50) of H chain with interaction energies of (−15.30, −17.90) kcal/mol and has 18 minima at B3 by interacting to Arg50 of H chain with interaction energies (−15.60, −17.80) kcal/mol, respectively. The two clusters of ACET minima at B2 and B3 are separated by residues Arg50 of H chain and Phe99 of L chain. For the ACEM group, 2 and 52 minima were found at B1 and B3, respectively, by interacting to residue Asp52 of H chain in B1 and to residues Arg50 of H chain and Asp31 of L chain in B3. For the PHEN group, 1 minimum was found at B1 with electrostatic interaction to Asp52 of H chain and hydrophobic interaction to Tyr45 of L chain, and 27 minima at B3 by forming electrostatic interactions to Asp31 of L chain and (Tyr25, Arg50) of H chain, respectively. Similar distribution was also found for the IMIA group with 3 and 25 minima located at B1 and B2, respectively.

Due to their closeness to CDR3 loop, B1 and B2 minima were grouped into the same cluster as a binding site. Maximum distance from the center of B1 and B2 to B3 is approximately 7.5 Å, indicating a gap of one amino acid. Based on the MCSS minima distribution, the key sequence pattern was defined as X–Z, in which X = (Y, H, Q/N, D/E) and Z = (Y, H, Q/N, D/E), as given in Table 2. The sequence pattern

was subsequently used to search HPV16 L1 protein sequence to identify the peptide binders. Altogether, 16 peptides were identified as potential binders (Fig. 4). Figure 4(a) shows the HPV16 L1 sequence with these epitopes highlighted in lower case and colored in orange. These binders were further analyzed based on their locations in the HPV16 L1 capsomer from cryto-EM studies and the results are summarized in Fig. 4(b). Overall, two peptides (AE3-P5, AE3-P10) are located at the inner pore of capsomer (PDB entry: 3JBA[9]), three peptides (AE3-P3, AE3-P12, AE3-P13) are buried inside L1 protein, and peptide AE3-P16 is interacting with another monomer; therefore, these six peptides are discarded as potential epitopes. The remaining 10 peptides could be classified into two groups: the first group (listed as EA) consists of five peptides exposed to solvent, four of which (AE3-P2: "IKKPNNNKIL", AE3-P7: "PPIGEHWGK", AE3-P8: "NTVIQDGDM", and AE3-P9: "CKYPDYIKM") are located at the region where the antibody U4 binds[9]; the second group (listed as EB) consists of peptides located at the surface loops from two L1 monomers (Fig. S1 supplementary material). These loops (labeled BC, DE, FG, and H1[31,32]) are shown to bind to antibody V5 and other antibodies (e.g. 1A, 14J and H263.A2) from cryto-EM structures.[9] Figures 4(c) and 4(d) showed the position of these two groups of epitopes in protein structures (cryto-EM structure of HPV16 L1 complexed with antibody U4[9] and complexed with antibody V5[31,32]), respectively.

3.3. *Prediction of epitopes of HPVs to antibody AG7*

The antibody structure of antibody AG7 was built using the crystal structure of homologous antibody from PDB entry 5F3J[23] as template. Figure 1(c) shows the model structure and the sequence alignment used for homology modeling. Around the CDR3 loop, two distinct regions, C1 and C2, are identified by H chain residues (Asn33, Asn50, Ser59, Tyr101) and L chain residues (His96, Phe101), and by H chain residue Asp52 and L chain residues (Tyr37, Arg55, His96), respectively.

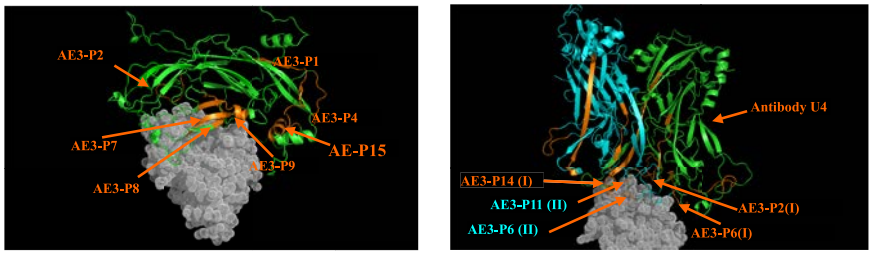
Figure 2(c) shows the distribution of MCSS minima of functional groups on the surface of AG7. A comparison with the U4 and AE3 antibodies shows that the surface around the CDR3 loop is relatively flat. For the ACEM group, three minima are located at C1 region with hydrogen bonding to the amide of Tyr101(H) and interaction energies of −10.40, −10.00 and −10.30 kcal/mol, respectively, while four minima are located at C2 region with hydrogen bonding to residues Arg55(L) and the amide group of Tyr101(H) with interaction energies of (−10.20, −11.50) kcal/mol, respectively. For aromatic groups such as PHEN, IMIA and BENZ, four PHEN minima were found at C1 region with specific interactions to residues (Asn50, Asp52, Ser59, and amide of Tyr101) of H chain. Five IMIA minima and three BENZ minima are located on the surface between residues Asp52 and Ser59 of H chain in C1. The interaction energies were calculated to be (−10.10, −11.20) kcal/mol, (−10.10, −10.80) kcal/mol) and (−10.10, −10.50) kcal/mol for the PHEN, IMIA and BENZ minima, respectively.

ATVYLPPVPVSKvvstdeyvartniyyhagtSRLLAVGHPYFPikkpnnnkilVPKvsglqyrvfRIHLPDPNKFGFPdtsfynpdtQRLVWA
CVGVEVGRGQPLGVGISGHPLlnklddtenasaYAAnagvdnrecismdykqtQLCLIGCKppigehwgkGSPCTQVAVQPGDCPPL
ELIntviqdgdmVDTGFGAMDFTTLQANKSEVPLDICTSIckypdyikmVSEPYGDSLFFylrreqmfvRHLFNRAgtvgenvpddlyi
KGSGSTANlassnyfptPSGSMVTSDAQIFNKPYWLQraqghnngicwgnqlfvTVVDTTRSTNMSLCAAISTSETTYKNtnfkeylr
hgeeydlqfIFQLCKITLTADVMTYIHSMNstiledwnfGLQPPPggtledtyrFVTSQAIACQKHTPPAPKEDPLKKYTFWEVNLKE
KFSADLDQFPLGRKFLLQLGLKAKPKFTLGKRK

(a)

	sequences	location	Surface Loop	Epitopes
AE3-P1	vvstdeyvartniyyhag	Exposed		EA
AE3-P2	ikkpnnnkil	Exposed	BC	EA
AE3-P3	vsglqyrvf	Buried		-
AE3-P4	dtsfynpdt	Exposed		EA
AE3-P5	lnklddtenasa	Inner Pore		-
AE3-P6	nagvdnrecismdykqt	Partially exposed	DE	EB
AE3-P7	ppigehwgk	Exposed		EA
AE3-P8	ntviqdgdm	Exposed		EA
AE3-P9	ckypdyikm	Exposed		EA
AE3-P10	ylrreqmfv	Inner pore		-
AE3-P11	gtvgenvpddlyi	Partially exposed	FG	EB
AE3-P12	lassnyfpt	Buried		-
AE3-P13	raqghnngicwgnqlfv	Buried		-
AE3-P14	tnfkeylrhgeeydlqf	Partially exposed	HI	EB
AE3-P15	stiledwnf	Exposed		EA
AE3-P16	ggtledtyr	Interacting other monomer		-

(b)



(c) (d)

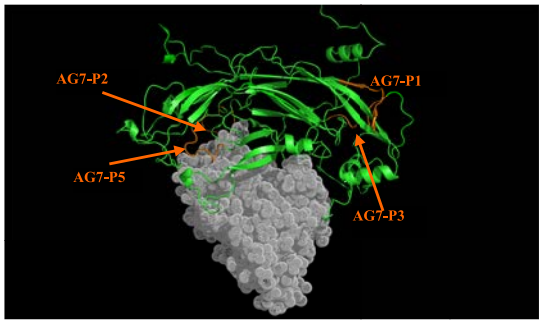
Fig. 4. (a) The predicted epitopes of HPV16 L1 protein to antibody AE3 are highlighted in lower case and colored orange in the protein sequence. The peptides identified as binders using the sequence search only are shown in lower case. (b) The 16 predicted peptides, their location in the structure, and any interaction with the surface loops of two L1 monomers are listed (Fig. S1). The surface loop labeled BC, DE, FG, and HI were defined previously.[31,32] See the context for details. (c) Backbone presentation of the HPV L1 with the group 1 epitopes (as listed as EA in B) shown in orange. Antibody AE3 is displayed as mesh surface based on cryto-EM structure of HPV16 L1 and antibody U4 (PDB 3JBA).[9] (d) Backbone presentation of the HPV L1 with the group 2 epitopes (listed as EB in B) shown in orange. Antibody AE3 is displayed as mesh surface in white based on the cryto-EM structure of HPV16 L1 and antibody V5 (PDB 3J7E).[32] The epitope peptides from different L1 proteins in capsomer were labelled as I and II in parenthesis and colored in orange (I) and labeled in cyan (II). Figures were prepared using PyMOL.[34]

ATVYLPPVPVSKVVSTDEYvartniyyhAGTSRLLAVGHPYFPikkpnnnkiLVPKVSGLQYRVFRIHlpdpnkfgfPDTSFYNPDTQ
RLVWACVGVEVGRGQPLGVGISGHPLLNKLDDTENASAYAANAGVDNRECISmdykqtqlcLIGCKPPIGEHWGKGSPCTQ
VAVQPGDCPPLELINTVIQDGDMVDTGFGAMDfttlqanksEVPLDICTSICKYPDYIKMVSEPYGDSLFFYlrreqmfvrHLFNR
AGTVGENVPDDLYIKGSGSTANlassnyfptPSGSMVtsdaqifnkPYWlqraqghnngicwgnqlfvtVVDTTRSTNMSLCAAISTSEt
tykntnfkEYLRHGEEYDLQFIFQLCKITLTADVMTYIHSMNSTILEDWNFGLQPPPGGTLEDTYRFVTSQAIACQKHTPPAPK
EDPLKKYTFWEVNLKEKFSADLDQFPLGRKFLLQLGLKAKPKFTLGKRK

(a)

	sequences	location	Epitope
AG7-P1	**vartniyyh**	**exposed**	
AG7-P2	**ikkpnnnki**	**exposed (BC)**	**Yes**
AG7-P3	**lpdpnkfgf**	**exposed**	
AG7-P4	**mdykqtqlc**	**buried**	
AG7-P5	**fttlqanks**	**exposed**	**Yes**
AG7-P6	**lrreqmfvr**	**Inner pore**	
AG7-P7	**lassnyfpt**	**Inner pore**	
AG7-P8	**tsdaqifnk**	**Inner pore**	
AG7-P9	**lqraqghnngicwgnqlfvt**	**Buried**	
AG7-P10	**ttykntnfk**	**Interacting other monomer**	

(b)



(c)

Fig. 5. (a) The predicted epitopes of HPV16 L1 protein to antibody AG7 are highlighted in lower case and colored orange in the protein sequence. The peptides identified as binders using the sequence search only are shown in lower case. (b) The selection based on their surface exposure. See the context for details. (c) Backbone presentation of the HPV L1 with predicted epitopes shown in orange. Antibody AG7 is displayed as mesh surface based on cryto-EM structure of HPV16 L1 and U4.[9] Figures were prepared using PyMOL.[34]

The maximum distance between the two binding sites is approximately 7.5 Å, indicating a gap of one amino acid. The key sequence pattern for the binding epitopes can thus be defined as X–Z, in which X = (F, Y, H, Q/N), and Z = Q/N, as given in Table 2. After searching using the protocol described in Method 2.3, 10 peptide binders were predicted (Fig. 5); however, only four of them are exposed in the capsomer. Figure 5(a) shows the HPV16 L1 sequence with these epitopes highlighted

in lower case and colored in orange, and Fig. 5(b) shows their positions in the protein structure. Two peptides (AG7-P2-"IKKPNNKI" and AG7-P5-"FTTLQANKS") are identified as binders to the antibody based on comparison with the cryto-EM structure of the L1 capsomer and U4 antibody.[9] However, the other two peptides (AG7-P1 and AG7-P3) are located at the region opposite to the interface between L1 capsomer and U4, and unlikely to be epitopes of an antibody. Figure 5(c) shows the position of these binders in the cryto-EM protein structure.

4. Discussions and Conclusions

In this work, we have applied a new method[10–12] to identify binding epitopes of HPV16 to three antibodies (U4, AE3 and AG7). Overall, our results identified a binding peptide which is located at the interface of the complex between U4 and HPV16 L1 capsomer identified by cryto-EM[9] and predicted a set of peptides that form epitopes or recognition surface for antibodies AE3 and AG7. This will significantly reduce the experimental work needed to identify antibody-binding epitopes for the newly developed antibodies AE3 and AG7.

Our calculations on U4 antibody have identified four peptides as potential binding epitopes. One of the peptides (residues 226–235, seq-"TSICKYPDYI") is part of the observed interface from the cryto-EM crystal, while the other three peptides are located at different loop positions on the L1 surface. One of these binding peptides **E4** (residues 361–369, seq-"KEYLRHGEE") is close to the FG surface loop, where antibody H16.V5 is shown to bind.[31,32] This binder overlays the epitope RHGEEYDLQFIFQLCKITLT that was predicted previously for vaccine production against HPV16 and HPV18.[33] A comparison with previous neutralization and structural studies[8,9] shows that the epitope at L1 C-terminal invading arm (427–445) is missing from our prediction. This is probably due to the fact that our method only searches epitopes over the monomer of L1 protein, while the experiments were performed using the L1 capsid; the epitope at the C-terminal region revealed from the cryto-EM structure could be caused by the formation of the L1 capsomer.

Our model structure shows that antibody AE3 is significantly charged around the CDR3 loop (Fig. 1(b)). Our calculations predicted a set of binding peptides based on the sequence pattern derived from the MCSS minima around the CDR3 loop. The binder peptides can be clustered into two groups of 10 potential epitopes. The first group consists of five peptides: AE3-P2 (resides 52–61, seq-"IKKPNNNKIL"), AE3-P7 (residues 163–171, seq-"PPIGEHWGK"), AE3-P8 (residues 192–200, seq-"NTVIQDGDM"), AE3-P9 (residues 229–237, seq-"CKYPDYIKM"), and AE3-P15 (residues 396–404, seq-"STILEDWNF"). Out of these peptides, three peptides (AE3-P7, AE3-P8 and AE3-P9) are close to each other and form a recognition surface similar to that found for antibody U4. The second group of four peptides includes AE3-P2 (residues 52–61, seq-"IKKPNNNKIL"), AE3-P6 (residues 138–154, seq-"NAGVDNRESCISMDYKQT"), AE3-P11 (residues 265–277, seq-"GTVGEN-

VPDDLYI"), and AE3-P14 (residues 358–374, seq-"TNFKEYLRHEEYDLQF"). These peptides are located at surface loops BC, DE, FG, and HI from different monomers, respectively, representing the binding modes of several antibodies such as V5, 1A, 14J, and 263A2 revealed experimentally (Fig. 4(d)).[9] Therefore, dual binding and neutralization mechanisms could occur for the antibody AE3 to HPV16 L1 protein, indicating a strong binding affinity for AE3 to HPV16 L1 capsids. In fact, antibodies have been shown previously to be capable of binding two types of epitopes of antigen, therefore, neutralizing the virus via two mechanisms.[35] Since the predicted binders of AE3 are consistent with the epitopes of both antibodies V5 and U4, AE3 will potentially neutralize the HPV infections very effectively. In contrast, the AG7 antibody has an apolar flat surface around CDR3 loop, resulting in different MCSS minima distribution and binding epitopes (Figs. 2(c) and 5). Out of the four epitope peptides obtained, only peptides AG7-P2 (residues 52–60, seq-"IKKPNNNKI") and AG7-P5 (residues 210–218, seq-"FTTLQANKS") could interact with the antibody. Therefore, this key difference between the L1 protein epitopes of the binding antibodies AG7 and AE3 suggests that the binding affinity of AG7 to L1 capsids is lower than that of AE3. This is in agreement with experimental results in which the affinity of AG7 is 10-fold weaker than AE3.[20]

Overall, our calculations have predicted several binding peptides of HPV16 L1 protein where the newly developed antibodies AE3 and AG7 could bind. These peptides will be tested using *in vitro* binding assay, and the verified epitopes will be used in vaccine design to elicit the neutralizing antibody response. Moreover, the recognition surface formed by three epitopes predicted for AE3 (located at regions of residues 163–171, residues 192–200 and residues 229–237) and the epitope for AG7 (located at region of residues 210–218) are highly conserved among the high-risk HPVs (Fig. S2 in supplementary materials). These epitopes, upon experimental verification, could be used for the development of effective vaccines against cervical cancer.

Acknowledgment

Na Liu and Minghui Yang were supported by the National Natural Science Foundation of China (Grant Nos. 21773297 and 21773298).

References

1. Gilmer LS, Human papillomavirus vaccine update, *Prim Care* **42**:17–32, 2015.
2. Dickson EL *et al.*, Recent trends in type-specific HPV infection rates in the United States, *Epidemiol Infect* **143**:1042–1047, 2015.
3. McKee SJ *et al.*, Recent progress in vaccination against human papillomavirus-mediated cervical cancer, *Rev Med Virol* **25**:54–71, 2015.
4. Baker TS *et al.*, Structures of bovine and human papillomaviruses-analysis by cryoelectron microscopy and 3-dimensional image-reconstruction, *Biophys J* **60**: 1445–1456, 1991.
5. Buck CB *et al.*, Generation of HPV pseudovirions using transfection and their use in neutralization assays, *Methods Mol Med* **119**:445–462, 2005.

6. Rossmann MG, Johnson JE, Icosahedral RNA virus structure, *Annu Rev Biochem* **58**:533–573, 1989.
7. Buck CB *et al.*, Maturation of papillomavirus capsids, *J Virol* **79**:2839–2846, 2005.
8. Carter JJ *et al.*, Identification of a human papillomavirus type 16-specific epitope on the C-terminal arm of the major capsid protein L1, *J Virol* **77**:11625–11632, 2003.
9. Guan J *et al.*, The U4 antibody epitope on human papillomavirus 16 identified by cryo-electron microscopy, *J Virol* **89**:12108–12117, 2015.
10. Zhang WS *et al.*, Computational identification of epitopes in the glycoproteins of novel bunyavirus (SFTS virus) recognized by a human monoclonal antibody (MAb 4-5), *J Comput Aided Mol Des* **27**:539–550, 2013.
11. Jiao YJ *et al.*, Antibody recognition of Shiga toxins (Stxs): Computational identification of the epitopes of Stx2 subunit A to the antibodies 11E10 and S2C4, *Plos One* **9**:e88191, 2014.
12. Jones ML *et al.*, Computational identification of antibody epitopes on the dengue virus NS1 protein, *Molecules* **22**:607–628, 2017.
13. Caflisch A, Computational combinatorial ligand design: Application to human alpha-thrombin, *J Comput Aided Mol Des* **10**:372–396, 1996.
14. Caflisch A, Karplus M, Acid and thermal-denaturation of barnase investigated by molecular-dynamics simulations, *J Mol Biol* **252**:672–708, 1995.
15. Caflisch A *et al.*, Multiple copy simultaneous search and construction of ligands in binding-sites application to inhibitors of HIV-1 aspartic proteinase, *J Med Chem* **36**:2142–2167, 1993.
16. Zeng J, Mini-review: Computational structure-based design of inhibitors that target protein surfaces, *Comb Chem High Throughput Screen* **3**:355–362, 2000.
17. Zeng J, Treutlein HR, A method for computational combinatorial peptide design of inhibitors of Ras protein, *Protein Eng* **12**:457–468, 1999.
18. Zeng J *et al.*, Design of inhibitors of Ras-Raf interaction using a computational combinatorial algorithm, *Protein Eng* **14**:39–45, 2001.
19. Zeng J *et al.*, Predicting sequences and structures of MHC-binding peptides: A computational combinatorial approach, *J Comput Aided Mol Des* **15**:573–586, 2001.
20. Wang Y *et al.*, Characterization of two new monoclonal antibodies against human papillomavirus type 16 L1 protein, *Diagn Pathol* **9**:101–107, 2014.
21. Zhou HH *et al.*, Structural insights into the down-regulation of overexpressed p185(her2/neu) protein of transformed cells by the antibody chA21, *J Biol Chem* **286**:31676–31683, 2011.
22. Nagae M *et al.*, Crystal structure of anti-polysialic acid antibody single chain Fv fragment complexed with octasialic acid insight into the binding preference for polysialic acid, *J Biol Chem* **288**:33784–33796, 2013.
23. Chen E *et al.*, Broadly neutralizing epitopes in the Plasmodium vivax vaccine candidate Duffy Binding Protein, *Proc Natl Acad Sci USA* **113**:6277–6282, 2016.
24. Biasini M *et al.*, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Res* **42**:W252–W258, 2014.
25. Kiefer F *et al.*, The SWISS-MODEL repository and associated resources, *Nucleic Acids Res* **37**:D387–D392, 2009.
26. Arnold K *et al.*, The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling, *Bioinformatics* **22**:195–201, 2006.
27. Guex N *et al.*, Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective, *Electrophoresis* **30**:S162–S173, 2009.
28. MacKerell AD *et al.*, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B* **102**:3586–3616, 1998.

29. Simonson T, Brunger AT, Solvation free-energies estimated from macroscopic continuum theory — An accuracy assessment, *J Phys Chem* **98**:4683–4694, 1994.
30. Schiller JT *et al.*, A review of clinical trials of human papillomavirus prophylactic vaccines, *Vaccine* **30**:F123–F138, 2012.
31. Guan J *et al.*, Structural comparison of four different antibodies interacting with human papillomavirus 16 and mechanisms of neutralization, *Virology* **483**:253–263, 2015.
32. Lee H *et al.*, A cryo-electron microscopy study identifies the complete H16.V5 epitope and reveals global conformational changes initiated by binding of the neutralizing antibody fragment, *J Virol* **89**:1428–1438, 2015.
33. Baidya S *et al.*, Epitope design of L1 protein for vaccine production against human papilloma virus types 16 and 18, *Bioinformation* **13**(3):86–93, 2017.
34. DeLano WL, The PyMOL Molecular Graphics System: DeLano Scientific, San Carlos, CA, 2009.
35. Tan Y *et al.*, A novel humanized antibody neutralizes H5N1 influenza virus via two different mechanisms, *J Virol* **89**:3712–3722, 2015.