J Comput Aided Mol Des (2013) 27:539–550
DOI 10.1007/s10822-013-9661-7

Computational identification of epitopes in the glycoproteins of novel bunyavirus (SFTS virus) recognized by a human monoclonal antibody (MAb 4-5)

Wenshuai Zhang · Xiaoyan Zeng · Li Zhang · Haiyan Peng · Yongjun Jiao · Jun Zeng · Herbert R. Treutlein

Received: 28 November 2012 / Accepted: 24 June 2013 / Published online: 10 July 2013

Abstract In this work, we have developed a new approach to predict the epitopes of antigens that are recognized by a specific antibody. Our method is based on the "multiple copy simultaneous search" (MCSS) approach which identifies optimal locations of small chemical functional groups on the surfaces of the antibody, and identifying sequence patterns of peptides that can bind to the surface of the antibody. The identified sequence patterns are then used to search the amino-acid sequence of the antigen protein. The approach was validated by reproducing the binding epitope of HIV gp120 envelop glycoprotein for the human neutralizing antibody as revealed in the available crystal structure. Our method was then applied to predict the epitopes of two glycoproteins of a newly discovered bunyavirus recognized by an antibody named MAb 4-5. These predicted epitopes can be verified by experimental methods. We also discuss the involvement of different amino acids in the antigen–antibody recognition based on the distributions of MCSS minima of different functional groups.

Keywords Antibody–antigen interactions · Computational binding epitope prediction · Molecular modeling

Electronic supplementary material The online version of this article (doi:10.1007/s10822-013-9661-7) contains supplementary material, which is available to authorized users.

W. Zhang · X. Zeng · L. Zhang · H. Peng · Y. Jiao (✉)
Key Laboratory of Enteric Pathogenic Microbiology, Jiangsu Provincial Center for Disease Prevention and Control, Institute of Pathogenic Microbiology, Ministry Health, Nanjing 210009, China
e-mail: yongjunjiao@gmail.com

J. Zeng (✉) · H. R. Treutlein
Monash Institute of Pharmaceutical Sciences (MIPS), Monash University, 399 Royal Parade, Parkville, VIC 3050, Australia
e-mail: Jun.Zeng@computistresearch.com

J. Zeng · H. R. Treutlein (✉)
Computist Bio-Nanotech, 1 Dalmore Drive, Scoresby, VIC 3179, Australia
e-mail: herbert.treutlein@computistresearch.com

Introduction

Severe fever with thrombocytopenia syndrome (SFTS) is a fatal life-threatening infectious disease with a death rate of 30 % [1]. The disease is caused by a bunyavirus, SFTS virus (SFTSV) which was identified recently [1]. The symptoms include sudden onset of fever, respiratory tract or gastrointestinal symptoms, decline of the whole white blood cell count, hemorrhage and in the end stage, multi-organ failure [2]. The circulation of this disease covers a broad geographic area in the east and central China [1]. More recently, another bunyavirus (Heartland virus) has also been discovered in Missouri, USA [3]. It causes clinical symptoms in humans similar to SFTSV, possibly indicating a worldwide distribution of this virus [3]. However, no specific clinical treatment is available for the SFTSV infection besides the supportive care. The development of prophylactic and therapeutic measures such as therapeutic antibodies and vaccines to protect susceptible individuals and those at high risk of complications of infection is therefore of significant urgency.

SFTSV is a member of phlebovirus genus in the Bunyaviridae family [1]. Like all bunyavirues, SFTSV has a trisegmented, single stranded RNA genome with negative (L and M segments) or ambisense (S segment) polarity which encodes seven proteins [4]. The two glycoproteins,

Gn and Gc, which are obtained from cleavage of a precursor encoded by the M segment, are highly antigenic envelope proteins and responsible for receptor binding and membrane fusion [3]. These viral surface glycoproteins are thus ideal targets for neutralizing antibody responses.

Recently, a human monoclonal antibody named MAb 4-5, has been isolated from a phage antibody library using whole SFTSV virions [5]. Its neutralizing property was investigated and their binding domain mapping was performed [5]. MAb 4-5 was found to bind a functional domain at the N terminus of the Gn glycoprotein. This as yet uncharacterized epitope is well conserved among viruses in disparate geographic regions in China as the MAb 4-5 shows a cross-neutralizing activity. In this work, we performed a computational prediction of the epitope of the glycoproteins binding to the MAb 4-5 antibody. The identified epitope can be used for the design of antibody therapeutics in passive immunotherapy to the SFTSV viral infections.

The most common techniques currently used to determine an antibody-binding epitope are X-ray crystallography or NMR techniques, and the phage display technique [6, 7]. While the former methods are probably more accurate, they are usually not available for many cases due to technical difficulties. On the other hand, phage display has been frequently used for the determination and mapping of binding epitopes [7]. Phage display uses a mixture of phage clones each carrying a different peptide insertion in a surface protein, so that each phage displays a different peptide on its surface. From a set of random peptides, this technology can select those peptides which bind with high affinity to the antibody. The peptides selected mimic the epitope in terms of spatial organization and physicochemical properties [8, 9]. These peptides are named "mimotopes" [10]. However, it is very difficult to identify reliably the actual epitope mimicked by these peptides [10]. Although they show a partial functional equivalence, they often even do not share sequence similarities [10]. Therefore, it is useful to develop an algorithm that could use the information obtained in the set of peptides to correctly predict the corresponding epitope on the surface of the antigen. Not surprisingly many automated methods and procedures for this purpose have been developed [10]. Based on different approaches, these methods can be classified into three categories: firstly algorithms such as implemented in FindMap [10], where the software uses only sequence information from mimotopes and the antigen; secondly, several methods have been developed to use both sequences data and the antigen 3D structure, such as SiteLight [11], 3DEX [12], Mapitope [13–15], and Pepsurf [16]; and thirdly, web servers have been set up by integrating in various ways the two preceding approaches; these methods include MIMOX [17], MIMOP [18] and Pepitope [19], etc. More recently, a novel approach, which is implemented in the software program LocaPep [20] has been developed. It uses an algorithm that focuses on protein surface features for predicting potential epitopes from the set of mimotopes. It relies mainly on surface features with very little resort to sequence information, but shows advantages over the other algorithms which, for example, often predict several alternatives to the putative epitope [20].

In this work, we present a new approach for determining the antibody-binding epitope of an antigen. Our method consists of three steps: Firstly, we identify the locations of functional chemical groups on the key region of the antibody using an exhaustive "multiple copy simultaneous search" (MCSS) approach [21–26]. Each of these functional groups corresponds to an individual amino acid [26]. Secondly, the MCSS clusters of a specific functional group with favorite interaction energies with the protein, also referred to as "minima", are selected to identify a pattern of functional groups on the surface of the antigen. These functional group patterns are subsequently converted into a sequence pattern. Thirdly, the antigen protein sequence is sliced into short peptides of seven amino acids, and the set of peptide sequences are scored according to the number of matched amino acids with the sequence pattern identified. The peptides with high scores which match the key pattern are considered to be mimotopes. Our method presented here is an extension of our computational combinatorial inhibitor design (CCLD) approach, presented in refs. [22–26]. Previously, our CCLD approach has been successfully applied to design peptide inhibitors that could, e.g. block the Ras protein binding to its downstream target Raf [25, 26]. We had developed a novel scheme that allows the application of CCLD to identify several peptide inhibitors that target protein surface [24–26]. Several designed peptides were confirmed by in vitro Enzyme-Linked ImmunoSorbant Assay (ELISA), radioassay and Biosensor-based assays [25].

A validation study of our method for the prediction of antibody-binding epitopes is described in the first part of the "Results" section: As a test case, we choose the identification of the epitope in HIV gp120 envelope glycoprotein that is recognized by its human neutralizing antibody [27]. Since the crystal structure of the antibody-antigen complex has been solved [27], the binding epitope on the antigen is known. Using our new method to predict the epitope tested the accuracy of our method and could provide some insight into the antigen–antibody recognition as well.

In the second part of the "Results" section, we applied our method to predict the epitope on the ectodomains of the SFTSV glycoproteins recognized by the antibody MAb 4-5. The predicted epitopes will be verified experimentally

and will serve as the basis for the further development of a SFTSV vaccine.

In the next section we will describe details of the Materials and methods used, section will report the results, followed by the discussion and summary in sections "Discussions" and "Conclusions".

Methods

Figure 1S in the "Supplementary Information" presents the sequences of glycoproteins isolated from bunyavirus SFTSV. A sequence analysis was carried out using on-line protein structure analysis tools (http://www.cbs.dtu.dk/services) to classify different domains. In particular the TMHMM tool (http://www.cbs.dtu.dk/services/TMHMM/) was used to predict the extracellular, transmembrane, and intracellular domains of the proteins. Table 1 shows the classification of the glycoproteins Gn and Gc. As the ectodomains of both Gn and Gc share low sequence similarity (e.g. <10 %) with any protein with available experimental structure, there is no good template to construct homology models of the ectodomains. Therefore, we constructed a 3D structure of the antibody MAb 4-5, and used it to predict the epitope on the ectodomains of the glycoproteins.

Two ectodomains exist in the glycoprotein Gn and Gc. As previous experiments have demonstrated that the N-terminal ectodomains bind to the MAb 4-5 antibody [5], we, therefore, only considered the N-terminal domain for the Gn and Gc glycoproteins.

Homology modeling of the antibody MAb 4-5

The sequence of the antibody 4-5 was used to search for the closest related antibody structure with known 3D structure using the BLAST (http://blast.ncbi.nlm.nih.gov) database search method focused on sequences of proteins from the protein data bank (PDB) [28]. The best matching antibody sequence found was PDB entry "1GC1" which contains a complex of HIV-1 gp120 envelope glycoprotein with CD4 receptor and a human neutralizing antibody [27]. The VL and the VH domains show sequence similarity of 98 and 71 % with the corresponding domains of MAb 4-5, respectively. Figure 1 shows the sequence alignment of the antibody in 1GC1 with the MAb 4-5 VL and VH domains.

The domains VL and VH of the antibody in the 1GC1 PDB structure was used as the template to create a homology model the antibody MAb 4-5. The model was created using the Modeller software [36] including explicit optimization of the CDR loop regions as implemented in Modeller.

Table 1 Domain classification of glycoproteins Gn and Gc of SFTSV obtained using the CBS Prediction Servers on-line tool (http://cbs.dtu.dk)

Domains	Residue number	Length (aa)
Glycoprotein Gn		
Extracellular	1–452	452
Transmembrane	453–470	18
Intracellular	471–476	6
Transmembrane	477–495	19
Extracellular	496–562	67
Glycoprotein Gc		
Extracellular	1–473	473
Transmembrane	474–496	23
Intracellular	497–511	15

MCSS of functional groups

The MCSS method has been widely used to determine energetically favorable positions and orientations of functional groups in a target protein [22, 26]. Using our homology model of an antibody, our quCBit software (http://www.computistresearch.com) which implements our MCSS approach, was used to scan the preferred locations of functional chemical groups on the binding surfaces, i.e. the "Complementarity Determining Regions" (CDRs). 11 functional groups were used, each of which corresponds to the side chains of different amino acids. Table 2 lists the relationship between the functional groups and amino acids. The parameters for both protein and functional groups were taken from the CHARMM22 all-hydrogen atom force field [30].

300 replicas of each functional group were randomly distributed inside a sphere with a 12Å radius around selected side chains of the CDRs of the antibody MAb 4-5 and the human neutralizing antibody of HIV gp120 envelope glycoprotein. The residues selected from the CDRs for the HIV neutralizing antibody and MAb 4-5 are: Asp32, Tyr49, Ala55, and Trp94 of the L chain and Tyr32, Leu55, Val57, Tyr100, Asn112, Lys116 of the H chain, and Asn32, Tyr49, Ala55, and Trp94 of the VL domain and Tyr32, Phe52, Asp57, Lys210, Glu107, and Asp111 of the VH domain, respectively. A 500-step multiple copy simultaneous minimization was performed. During all the MCSS calculations, each replica only interacts with a target protein, and not with the other replicas. The interaction energy, U_{MCSS}, was defined as

$$U_{MCSS} = U_{protein-replica} + U_{protein} + U_{replica} \quad (1)$$

where $U_{protein\text{-}replica}$ represents non-bonded interactions (i.e. van der Waals and electrostatic interactions) between the target protein and the replica. $U_{protein}$ and $U_{replica}$ represent the internal energy of the protein and each replica,

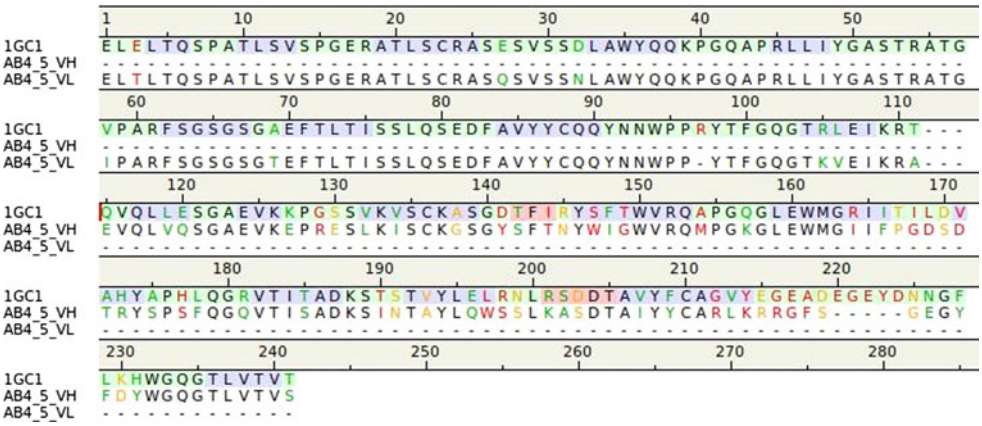

Fig. 1 Sequence alignment of the human neutralizing antibody for HIV gp120 (1GC1) with the VL and VH domains of antibody MAb4-5

Table 2 The relationship between the functional groups used 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, ALA

respectively. In the first protocol where the protein atoms were fixed and each replica treated as a single group, $U_{protein}$ and $U_{replica}$ were excluded.

The binding energy for a functional group in each minimum obtained from the MCSS calculations was defined as

$$U_{bind} = U_{protein-replica} + U_{replica} - U^{0}_{replica} \quad (2)$$

where $U^{0}_{replica}$ indicates the internal energy of each replica in vacuum. The nonbonded interaction was truncated at 20 Å. The dielectric constant was set to 10 to mimic solvent screening effects [31].

Identification of sequence pattern

MCSS functional groups showing strong interactions with the antibody were selected and are called "MCSS minima" in the following text. In the next step, the distribution and locations of the minima were identified. An interaction energy of −10.00 kcal/mol was used as the threshold for the minima of functional groups except for the larger or charged groups (i.e. INDO, MAMM and MGUA) where −15.00 kcal/mol was used. However, −10.00 kcal/mol was also used for the larger or charged groups on the HIV gp120 antibody. The spatial patterns of the locations of the MCSS minima on the surface of antibody were converted into a sequence pattern according to the relationship between the functional groups and amino acids as given in Table 2, and this sequence pattern served as the fingerprint to identify the epitopes of antigens.

Epitope search based on the sequence pattern

The sequence pattern obtained using the method described in section "Identification of sequence pattern" was used to identify the peptides derived from the antigens such as envelope glycoproteins of HIV gp120 and the ectodomains of glycoproteins Gn and Gc of SFTSV. In slight analogy to the experimental phage display technology described above, we decided to divide the whole protein sequence into overlapping peptides of length of seven amino acids. The typical peptide length in phage displays is around 12, however, we choose a length of seven because in our experience it allows a slightly more efficient scan of the MCSS minima distributions of the average sized binding epitopes. The peptides with a sequence that matches the key pattern derived from MCSS minima of functional groups were considered to be potentially part of the epitope and labeled as the "binders".

The sequence of each peptide derived from the protein depends on the starting point of the division into 7-m. To avoid artifacts by starting from a particular residue, the protein was sliced into 7-m peptide libraries several times, starting from the residue 1 up to 7. For example, the 1st set of peptides are the peptides of sequence number (1–7, 8–15…) and the 2nd set are the peptides of sequence number (2–8,9–16…), and the 7th set are the peptides of sequence number (7–13,14–20…). This results in seven

libraries of 7-m peptides. Each of the seven libraries was checked for sequence matches with the key pattern. Residues occurring in binder peptides from more than three libraries were considered part of the epitope. Therefore, the epitopes predicted from the seven sets of peptide libraries could vary in their length.

Results

Human neutralizing antibody binding to HIV gp120 envelope glycoprotein

We firstly applied our approach to predict the epitope of HIV gp120 envelope glycoprotein binding to a known human neutralizing antibody. The antibody structure was taken from the crystal structure 1GC1 [27]. Figures 2 a–h show all the minima of the functional groups on the binding surface of the antibody. Overall, the minima are grouped into three clusters, interacting with three binding grooves G1, G2, and G3: G1 is formed by residues Gln1, Tyr100, and Lys106 of the VH domain, G2 by residues Asn112, Phe114, and Tyr118 of the VH and Asp32 and Trp94 of the VL domain, and G3 by residues Ile54, Leu55, and Tyr109 of the VL domain. However, G1 is ca. 13 Å away from the standard antigen–antibody binding interface of the VL and VH domains, so that we only focused on the analysis of the MCSS minima at the G2 and G3. For the small functional groups, i.e. MEOH, MESH and IBUT, no well defined and clustered minima were found at G1 and G2, indicating that it is unlikely that small-size amino acids such as Ser, Thr, Cys, and Ala are significantly involved in the binding recognition of antibody-antigen. Previously, extensive analysis of protein–protein complexes has demonstrated that binding hot spots are abundant in charged residues Arg and Asp (via salt bridges), and Trp, Tyr and Asn in antigen–antibody complexes, as well as enriched in Ile, Gln and Asn in antibody proteins [29]. Our mapping results are consistent with these previous studies.

While 25 minima of ACEM are located in the binding groove G2, only one minimum is found at the G3. Therefore, amino acids such as ASN and GLN predominantly occur in positions of the epitope that bind to the G2 region. For the aromatic rings BENZ, IMIA, PHEN and INDO, the minima at the G2 are segregated into two clusters: One of them interacting with hydrophobic residues Tyr49 of the VL and Phe114 of the VH domain, and the other one with the side chains of Trp94 of the VL domain. The minima with stronger interaction energies are interacting with Trp94. For the positively charged groups MGUA and MAMM, the MCSS minima are spread along the CDR loops, with 62 MAMM minima and 67 MGUA minima found at the G2. Interestingly, no negatively charged minima of ACET are found at G2 and G3; probably due to highly hydrophobic nature of these grooves and the negatively charged Asp32 of VL near to G2. Overall, the majority of ACET minima are located in G1 and interacting with the charged residues Gln1 and Lys116 of VH chain of the antibody.

Figure 3 shows an overlay of a representative BENZ minima and the crystal structure of the complex 1GC1. The minima in the G3 groove overlay very well with the ILE423 of the antigen (HIV gp120 envelope glycoprotein). In the crystal structure, this residue interacts strongly with a hydrophobic patch formed by Ile54, Leu55, and Tyr109 of the VH chain and Trp94 of the VL chain of the antibody. Therefore, our MCSS minima reproduced the hot spot of antigen–antibody binding revealed experimentally by X-ray crystallography [29].

Using the minima at G2 and G3, we constructed sequence patterns for the peptides that could bind to the antibody. The maximum distance between the clusters of MCSS minima inside the binding grooves G2 and G3 is approximately 11.5 Å, corresponding to a separation of three amino acids. In the G2 groove, the best minima of MAMM and MGUA interact strongly with Asp32 of VL with interaction energies of −17.60 and −18.20 kcal/mol, respectively, while the best minima of aromatic rings ACEM, BENZ, INDO and IMIA interact with Trp94 showing energies of −10.40, −11.20, −15.00 and −9.60 kcal/mol, respectively. At the G3, the best minima of aromatic ring fragments BENZ, INDO and IMIA bind to residues Leu55166 and Tyr109 of VH domain with energies of −9.70, −12.00 and −9.90 kcal/mol, respectively. Therefore, the key sequence pattern for the binding epitope peptides can be defined as "XZ–J", in which X = R or K, Z = Q, N, Y, F, W or H and J = F, W or H. Table 3 lists the distribution of key MCSS minima and its derived sequence pattern. Please note that due to the short length (i.e. 7 residues) of peptides, it is sometimes necessary to loosen the search criteria for sequence pattern in order to find search hits: Here we used "XZ" as search criterion only.

The sequence pattern was subsequently used to search for "binders"(see "Methods" section above) from the peptide libraries derived from the sequence of the HIV gp120 envelope glycoprotein (given in Fig. 2S). The 7 libraries (see "Methods" section) were searched for peptides matching the calculated sequence pattern of the binding epitope. There are 302 residues in the HIV envelope glycoprotein which result in 43 peptides of a length of seven amino acids for each set of 7 libraries. Figure 3S in the "Supplementary Information" shows the identified epitopes of each set of peptide library with the epitopes highlighted in orange lower case characters. The predicted epitope of HIV gp120 envelope glycoprotein binding to the antibody is shown in Fig. 4a with the epitope highlighted in lower case and colored in orange. The five epitope regions

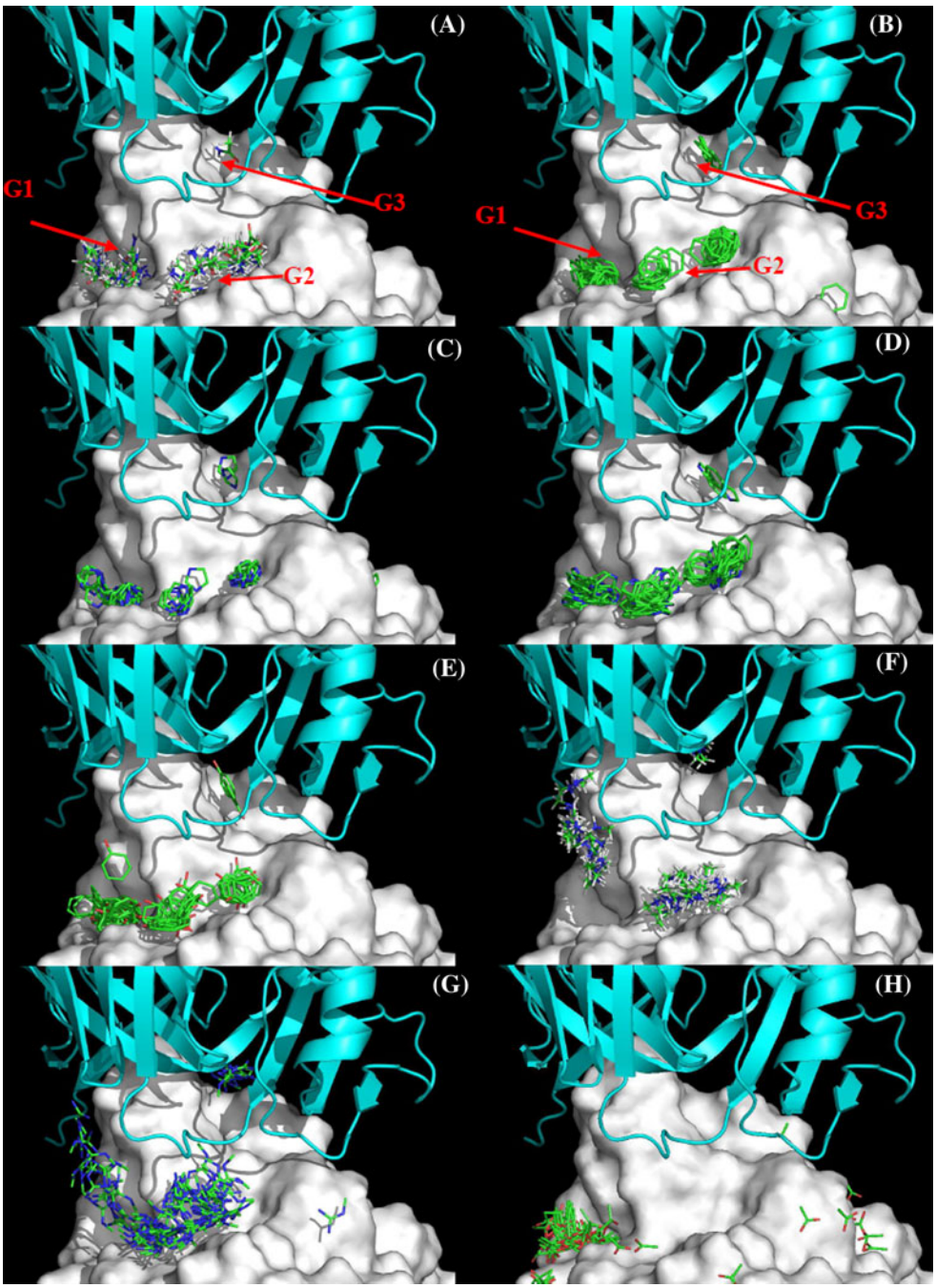

Fig. 2 Selected MCSS minima of functional groups on the surface of human neutralizing antibody of HIV gp120. **a** ACEM; **b** BENZ; **c** IMIA; **d** INDO; **e** PHEN; **f** MAMM; **g** MGUA; **h** ACET. Figures were prepared using PyMOL [32]

(FNMWKNDMW, ISRAKWNNTLKQIA, TIIFKQSSG, PCRIKQIIN and SELYKYKVVK) identified are colored in orange in the crystal structure of the complex (1GC1) (shown in Fig. 4b). One epitope (PCRIKQIIN) is in fact located in the interface between HIV gp120 envelope glycoprotein and its human neutralizing antibody.

Homology model of antibody MAb 4-5

The antigen of MAb 4-5 (glycoproteins Gc and Gn of SFTSV) does not currently have any homologous proteins in the PDB [28] that could be used as templates to build a reliable homology model. We, therefore, decided to only create a 3D model of the Mab 4-5 VL and VH domains of antibody which bind to the antigen. The VL and VH domains of the antibody in the crystal structure (PDB [28] 1GC1 [27]) were used as the template to create a model the antibody. Figure 5 shows the best model structure overlaid onto the template. The sequence similarities are 98 and 71 % for the corresponding L and H chains of the template and the MAb 4-5, respectively. Comparison of the template (in cyan) and model structure shows an excellent overlay,

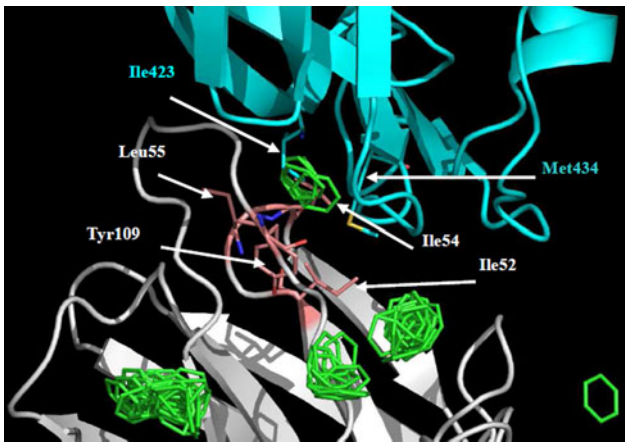

Fig. 3 Overlay of BENZ minima with side chains of residue Ile423 and Met434 of antigen in the crystal structure of the HIV gp120 antigen–antibody complex (1GC1). The antigen and antibody are colored *blue* and *white*, resp. The BENZ minima are colored *green*. *Hot spots* of the antigen–antibody interactions located in the protein–protein interface are highlighted using stick representations. Leu55 of VH chain and Ile52, Ile54 and Tyr109 of VL chain are colored *brown*, see text. Figures were prepared using PyMOL [32]

Table 3 Distribution of key minima and the derived sequence pattern for the binding epitope peptides to the HIV gp120 human neutralizing antibody

Binding groove	G2		11.5 Å	G3
MCSS minima pattern	MAMM	ACEM		BENZ
	MGUA	BENZ		INDO
		INDO		IMIA
		IMIA		
Sequence pattern	K	Q/N	Gap of 3 amino acid	F/Y
	R	F/Y		W
		W		H
		H		

A sequence pattern of "XZ–J"[X = (R, K), Z = (Q/N, Y/F, W, H) and J = (Y/F, W,H)] was obtained

however, the turn in the CDR3 loop is shifted by approx. 8.0 Å, resulting in two small cavities, one formed by residues Trp94 of VL domain, Trp33 and Leu99 of VH domain, and the other formed by residues Trp33, Phe52 and Arg101 of VH domain. These cavities differ from binding grooves seen in the crystal structure 1GC1 (Fig. 2), indicating that the binding epitopes for the glycoproteins of SFTSV are different to ones for HIV gp120.

Antibody MAb 4-5 binding to SFTSV glycoproteins

Using the homology model of the antibody MAb 4-5, the MCSS mapping of functional groups as described before was carried out. Figure 6 A-H shows the locations of the minima of functional groups on the surface of antibody MAb 4-5. Overall, three binding sites B1, B2 and B3 were identified from the surface of the antibody: Site B1 is formed by Asn32, Tyr91, and Tyr49 of the VL domain and backbone of Gly106 of the VH domain, B2 is defined by Trp94 of the VL and Gly105, Trp33, and Arg101 of the VH domain, and B3 consists of Tyr32, Asn31 and Lys100 of the VH domain. These binding sites are in similar locations as the sites G1 and G2 of the human neutralizing antibody described in section "Human neutralizing antibody binding to HIV gp120 envelope glycoprotein". However, no minima were found at a position corresponding to G3, indicating a different nature of epitopes in the SFTSV glycoproteins from the HIV gp120. As observed for the HIV antibody, only a very small number of minima of MEOH and IBUT were located on the surface of antibody, however, eighteen minima of MESH were found at the binding site B2 (Fig. 6b).

For the polar group ACEM, 4, 13 and 5 minima were found on the sites B1, and B2 and B3 sites, respectively. The most favorable interaction energy values are: −10.30 kcal/mol (B1), −12.10 kcal/mol (B2) and −11.70 kcal/mol (B3). For the aromatic rings BENZ, IMIA, PHEN and INDO, 93, 24, 90 and 57 minima are grouped into the three binding sites with the best minima at B2 showing favorable interactions with to Trp94 of VL and Trp33 of VH. The best minima have binding energies of (B1: −11.80, B2: −13.90, B3: −11.90) kcal/mol, (B1: −11.00, B2: −13.30, B3: −11.00) kcal/mol, (B1: −12.00, B2: −14.50, B3: −13.50) kcal/mol and (B1: −17.40, B2: −17.90, B3: −16.20) kcal/mol for the four groups BENZ, IMIA, PHEN and INDO, respectively. These minima are orientated perpendicularly to the Trp94, forming strong π–π interactions. For the charged groups ACET and MGUA, minima are spread along the CDR loops. While the majority of MGUA minima with strong interaction energies are located at the B1 site interacting with Asn92 and Trp94 of the VL domain, the best minima of ACET are located at the B2 and B3 sites, interacting with charged residues Arg101 and Arg102 of the VH chain, respectively. Due to the positive charged nature around the binding site, no minima were found for the MAMM which is also less hydrophobic than MGUA.

Based on the distribution of the important minima shown in Fig. 6a, sequence patterns for peptides that bind to MAb 4-5 were derived. The MCSS minima at the binding sites B1, B2 and B3 are separated by ca 9.0 Å, a distance that could accommodate two amino acids. However, all the functional groups have the best minima with the strong interaction energies at the B2 and B3 except the MGUA minima in which two minima are found at the B1 site and interact with Ser31 and Tyr9 of the VL domain. Therefore, no specific preference of amino acid is found to bind to B3. At the B2 site the best minima of ACEM, BENZ, INDO, IMIA and ACET are located at the same

Fig. 4 **a** The predicted epitopes of HIV gp120 envelope glycoprotein are highlighted in lower case and colored *orange* in the protein sequence. **b** Backbone presentation of the antigen–antibody complex showing the predicted epitopes in *orange*

(A)

TENfnmwkndmvEQMHEDIISLWDQSLKPCVKLTPLCVGAGSCNTSVITQACPKVSF
EPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAE
EEVVIRSVNFTDNAKTIIVQLNTSVEINCTGAGHCNisrakwnntlkqiasKLREQF
GNNKtiifkqssgGDPEIVTHSFNCGGEFFYCNSTQLFNSTWFGSDTITLpcrikqi
inMWQKVGKAMYAPPISGQIRCSSNITGLLLTRDGGNSNNESEIFRPGGGDMRDNWR
selykykvvKIE

(B)

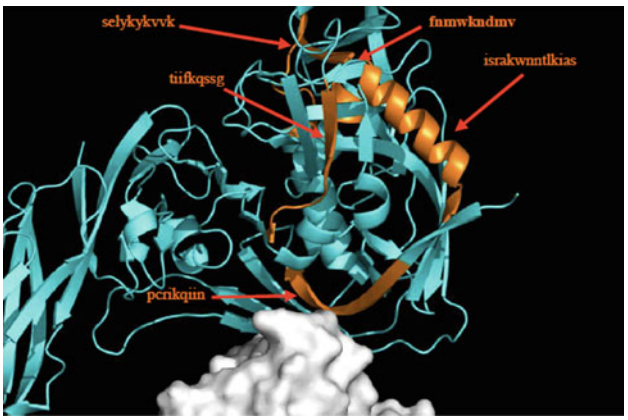

Fig. 5 The model structure of antibody MAb 4-5 (*green*) constructed using the antibody PDB id 1GC1 as template (in *cyan*). The figure was prepared using PyMOL [32]

position, while the MGUA minima are approx. 4.0 Å (position-1) away. The key sequence pattern for the binders was, therefore, defined as "R–ZJ", in which X = R, Z = R and J = Q or N, C, Y or F, H or W, D or E. Table 4 lists the distribution of key MCSS minima and its derived sequence pattern. Due to the short length of peptides (7mer), we loosened the search criterion to "ZJ" only.

Seven sets of peptide libraries of the ectodomains of glycoproteins Gn and Gc were generated using the protocol as described in the "Methods" section "Epitope search based on the sequence pattern". Figure 4S A and B in the "Supplementary Information" show the identified binders of each set of peptide libraries with the binder residues highlighted in lower case and colored orange. The final predicted epitopes of Gn and Gc using the protocol in section "Epitope search based on the sequence pattern" are shown in Fig. 7 with the epitope shown in orange lower case characters. Overall, the epitopes are located in the terminal regions, very likely to be located on the protein surface.

Influences of CDR flexibility on the MCSS mapping of MAb 4-5 antibody

To address the influences of errors in the loop conformations due to the homology modeling procedure, and the inherent flexibility of the CDRs on the distributions of MCSS minima of functional group, we built a second, different homology model of MAb 4-5 antibody using crystal structures of three different antibodies as templates (3RKD [33], 4HFU [34], and 2AJ3 [35]). As shown in Fig. 8a, the structural differences occur mainly around the CDR loops. Based on these different conformations, Fig. 8b shows the distribution of the benzene minima obtained using our two different models (cyan and orange), in comparison with the benzene minima (green) obtained from the model described in sections "Homology model of antibody MAb 4-5" and "Antibody MAb 4-5 binding to SFTSV glycoproteins" based on the 1GC1 template. While the benzene minima are located deeply inside a pocket at the binding site B2 of the model from 1GC1 (green), the distributions overlay very well when comparing the model based on 1GC1 and the model 2 based on the three templates specified above. For the model 1, the distributions overlay as well, but some minima can be found between the binding sites. Overall, the distributions of

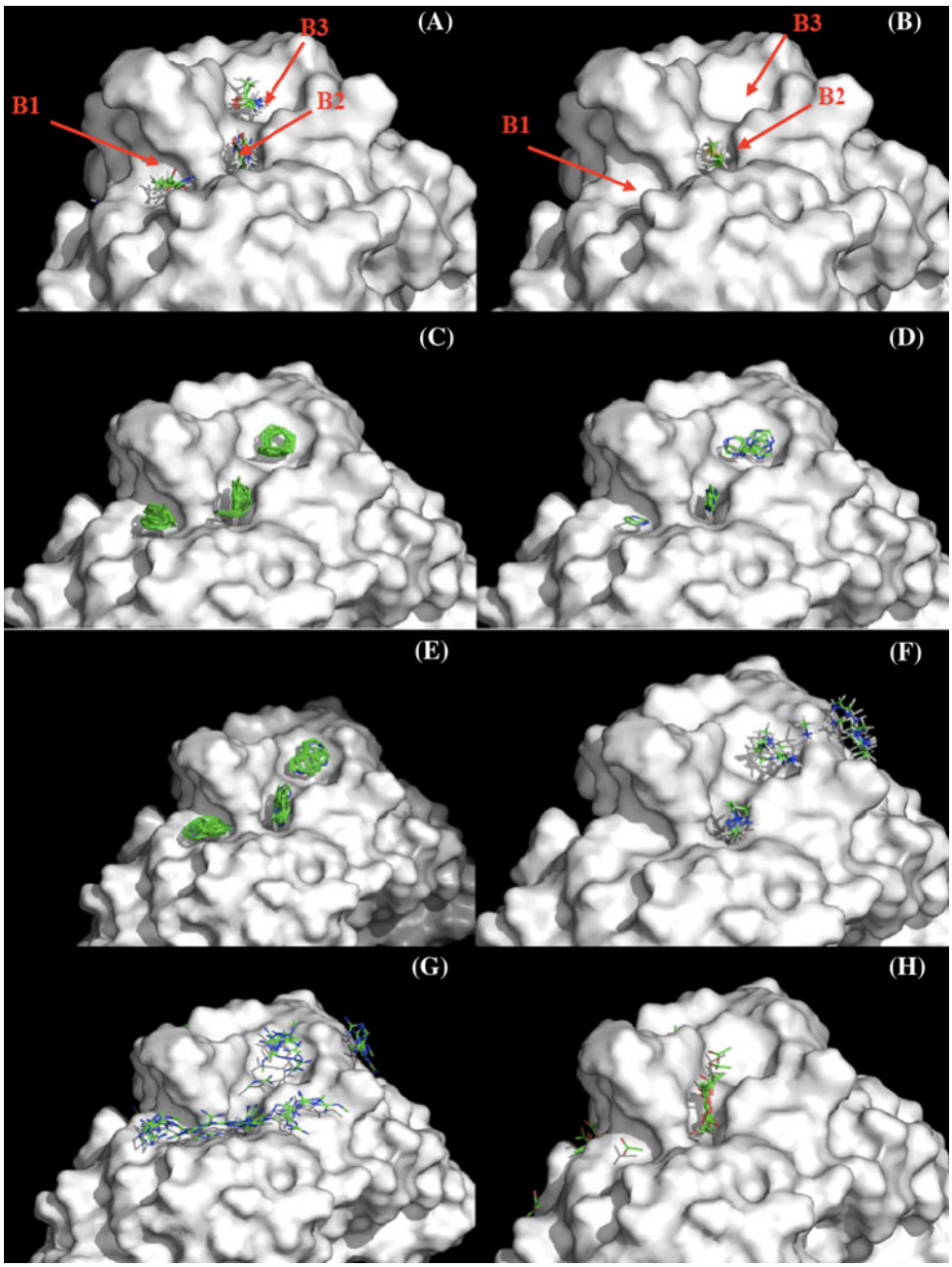

Fig. 6 Selected MCSS minima of functional groups on the surface of antibody Mab 4-5. **a** ACEM; **b** MESH; **c** BENZ; **d** IMIA; **e** INDO; **f** MAMM; **g** MGUA; **h** ACET. The PHEN minima are not shown due to their similarity to the distributions of BENZ, IMIA and INDO minima. Figures were prepared using PyMOL [32]

minima on the three binding sites B1, B2 and B3 are very close, indicating that the details of the CDR loop conformations only have an insignificant effect on the distribution of MCSS minima as calculated by our method and, therefore, also on the sequence pattern derived from the minima.

Discussions

We have presented a new method to predict the epitopes of an antigen that binds to an antibody. While our method cannot (yet) with certainty predict and identify the epitope but suggest a small number of peptides that could be part of the epitope, it already helps significantly to reduce the amount of experimental work needed to find an antibody binding epitope.

Our method involves three steps: (1) mapping of functional groups onto the surface of the antibody, (2) deriving a sequence pattern for potential binding peptides based on the distribution of important minima of functional groups, and (3) searching the binding peptides from the sequence of the antigen. The method was validated by application to the human neutralizing antibody of HIV gp120 envelope glycoprotein [27] and then used to predict the epitopes of ectodomains of the glycoproteins of SFTSV binding to the newly discovered antibody MAb 4-5.

Table 4 Distribution of key minima and the derived sequence pattern for the binding epitope peptides to the antibody MAb 4-5

Binding site	B1	9.0 Å	B2	
MCSS minima pattern	MGUA		MGUA	ACEM
				BENZ
				INDO
				IMIA
				ACET
Sequence pattern	R	Gap of 2 amino acid	R	Q/N
				F
				W
				H
				D/E

A sequence pattern of "R–XJ"(X = R and J = (Q/N, C, Y/F, H, W, D/E)) was obtained

Our MCSS calculations on the HIV gp120 antibody have identified 5 peptides as potential candidates for the binding epitope. One of them is in fact part of the observed binding epitope. The other 4 predicted peptides are not located close enough to make contact with the antibody, but the result nevertheless demonstrates a significant simplification of the experimental search for the binding epitope. Moreover, none of small functional groups was found on the surface of the antibody. This is consistent with the previous analysis of protein–protein interactions using the currently available 3D complex structures [29]. The published results in Ref. 29 demonstrated that the protein–protein interactions are controlled by salt bridges and aromatic ring hydrophobic/polar interactions, and that the occurrence of small amino acids (e.g. Ser and Thr etc.) is relatively rare [29].

It is noticeable that only one cluster of MCSS benzene minima at the surface of the antibody overlays with the side chain of residue Ile423 from the antigen (Fig. 3). This might be caused by the artifact that our method does not take induced fit into account: In the case of the HIV antibody, we used the 3D structure of the antibody-antigen complex to derive our MCSS minima. However, the structures of the antigen epitope and the antibody CDRs before forming the complex would likely be somewhat different to the conformation observed in the crystal structure of the complex. Nevertheless, the method does reproduce an important part of the epitope of HIV gp120 envelope glycoprotein to its neutralizing antibody.

For the antibody MAb 4-5, our calculation indicated several interesting features: Firstly, errors in the determination of the exact loop conformation as well as the flexibility of the CDRs have only a minor effect on the distribution of MCSS minima (Fig. 8). As the functional groups used here are small or flexible, they could accommodate the conformational changes of the CDRs. Secondly, the MCSS minima are different from those found in the neutralizing antibody of HIV gp120 (Fig. 6 vs. Fig. 2) due to changes in the amino acid sequences of the CDR of the two antibodies. Accordingly, the sequence pattern identified from the MCSS minima is "R-RX–", which is different from the pattern "XZ–J" derived from the MCSS minima on the HIV gp120 neutralizing antibody. The epitopes identified for the ectodomains of glycoproteins Gn and Gc are located in the terminal regions, which are very likely to be exposed to solvent and to interact with the antibody.

Conclusions

In our current report, we developed a simple qualitative method to predict the epitopes of the ectodomains of the Gn and Gc that bind to MAb 4-5. The best epitopes from the Gn and Gc are predicted to be within the first 100 amino acid of the proteins with sequences of "SSWLRNHSQ" and

Fig. 7 The sequences of ectodomains of glycoproteins Gn (**a**) and Gc (**b**) of SFTSV showing the predicted epitopes highlighted in *orange* lower case characters

(A)

MMKVIWFSSLICLVIQCGGDTGPIICAGPIHSNKSADIPHLLGYSEKICQIDRLIHVsswlrn
hsqFQGYVGQRGGRSQVSYYPAEnsysrwsglLSPCDADWLGMLVVKKAKGSDMIVPGPSYKG
KVFFERPTFDGYVGWGCGSGKSRTESGELCSSDSGTSSGLLPSDRVLWIGDVACQPMTPIPEE
TFLELKSFSQSEFPDICKIDGIVFNQCEGESPPQPFDVAWMDVGHSHkiimrehktKWVQESS
SKDFVCYKEGTGPCSESEEKTCKTSGSCRGDMQFCKVAGCEHGEEASeakcrcslvHKPGEVV
VSYGGMRVRPKCYGFSRMMATLEVNQPEQRIGQCTGCHLECINGGVRLITLTSELKSATVCAS
HFCSSATSGKKSTEIQFHSGSLVGKTAINVKGALVDGTEFTFEGSCMFPDGCDAVDctfcref
lkNPQCYPAK

(B)

CDEMVHADSKlvscrqgsgNMKECVTTGRALLPAVNPGQEACLHFTAPGSPDSKCLKIKVKRI
NLKCKKSSSYFVPdarsrctsvrrcrwagdCQSGCPPHFTSNSFSDDWAGKMDRAGLGFSGCS
DGCGGAACGCFNAAPSCIFWRKWVENPHGIIWKVSPCAAWVPSAVIELTMPSGEVRTFHPMSG
IPTQVFKGVSVTYLGSDMEVSGLTDLCEIEELKSKKLALAPCNQAGMGVVGKVGEIQCSSEES
ARTIKKDGCIWNADLVGIELRVDDAVCYSKITSVEAVANYSAIPTTigglrfersHDSQGKIS
GSPLDITAIRGSFSVNYRGLRLSLSEITATCTGEVTNVSGCYSCMTGAKVSIKLHSSKNStah
vrckgdETAFSVLEGVHSYSVSLSFDHAVVDEQCQLNCGGHESHVTLKGNLIFLDVPKFVDGS
YMQTYHSTVPTGANIPSPTDWLNALFGNGLS

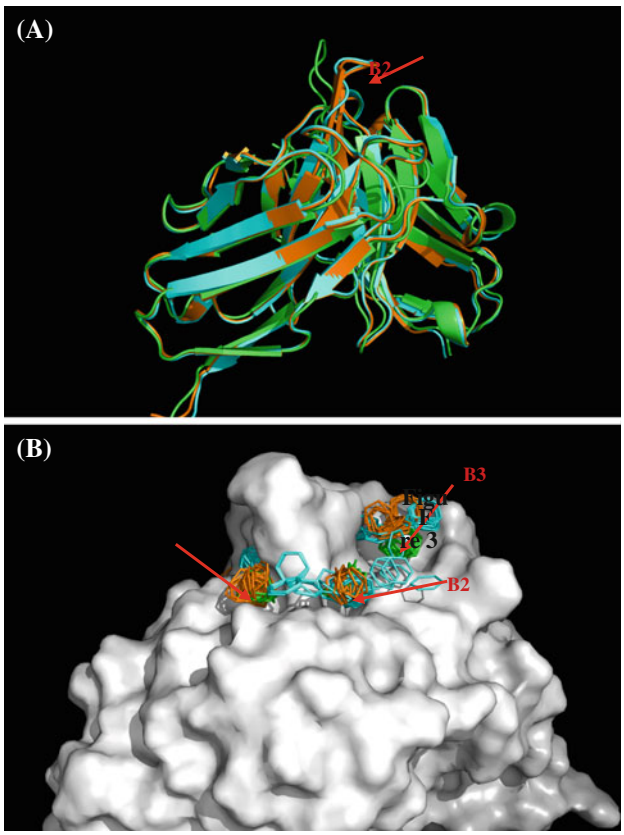

Fig. 8 **a** Three different conformations of MAb 4-5 antibody built from templates PDB id: 1GC1, and 2 models based on the templates with PDB ids 3RKD, 4HFU and 2AJ3. Three structures are colored in *green*, *cyan* and *orange*. **b** Distributions of MCSS minima of benzene groups on the surface of antibody. *Green* MCSS minima obtained using the homology model built from 1GC1; *cyan* MCSS minima obtained using the model 1 built from 3RKD, 4HFU, and 2AJ3; and *orange* MCSS minima obtained using the model 2 from 3RKD, 4HFU and 2AJ3. Three binding sites B1, B2 and B3 as shown in Fig. 6 are also labeled

"NSYSRWSGL" for Gn, and "LVSCRQGSG" and "DARSRCTSVRRCRWAGD" for Gc.

Our approach provides a qualitative assessment of peptide sequences from an antigen that could bind to an antibody which is useful to reduce the amount of experimental work necessary to identify an antibody binding epitope. Further improvement to our method could be implemented, such as a more rigorous algorithm for pattern recognition and a more quantitative scoring function to assess the binding affinities of the epitope peptides with the antibody. The work on an improved method is currently underway, together with the experimental verification of the predicted epitopes for Gn and Gc.

Acknowledgments This work was supported by grants from Jiangsu Province community development projects (BE2012768), Jiangsu Province Outstanding Medical Academic Leader Program (RC2011082), National Mega-Projects for Infectious Diseases from Ministry of Science and Technology, China and from the Ministry of Health, China (2013ZX09102029).

References

1. Yu XJ, Liang MF, Zhang SY, Liu Y, Li JD, Sun YL, Zhang L, Zhang QF, Popov VL, Li C, Qu J, Li Q, Zhang YP, Hai R, Wu W, Wang Q, Zhan FX, Wang XJ, Kan B, Wang SW, Wan KL, Jing HQ, Lu JX, Yin WW, Zhou H, Guan XH, Liu JF, Bi ZQ, Liu GH, Ren J, Wang H, Zhao Z, Song JD, He JR, Wan T, Zhang JS, Fu XP, Sun LN, Dong XP, Feng ZJ, Yang WZ, Hong T, Zhang Y, Walker DH, Wang Y, Li DX (2011) N Engl J Med 364:1523
2. Bao CJ, Qi X, Wang H (2011) N Engl J Med 365:862
3. McMullan LK, Folk SM, Kelly AJ, MacNeil A, Goldsmith CS, Metcalfe MG, Batten BC, Albarino CG, Zaki SR, Rollin PE, Nicholson WL, Nichol ST (2012) N Engl J Med 367:834
4. Li DX (2012) Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi 25:81
5. Yu L, Zhang L, Sun L, Lu J, Wu W, Li C, Zhang G, Zhang F, Jin C, Wang X, Bi Z, Li D, Liang M (2012) PLoS ONE 7:e38291
6. Scott JK, Smith GP (1990) Science 249:386
7. Smith GP, Petrenko VA (1997) Chem Rev 97:391
8. Cortese R, Felici F, Galfre G, Luzzago A, Monaci P, Nicosia A (1994) Trends Biotechnol 12:262
9. Cwirla SE, Peters EA, Barrett RW, Dower WJ (1990) Proc Natl Acad Sci U S A 87:6378
10. Geysen HM, Rodda SJ, Mason TJ (1986) Mol Immunol 23:709
11. Halperin I, Wolfson H, Nussinov R (2003) Protein Sci 12:1344
12. Schreiber A, Humbert M, Benz A, Dietrich U (2005) J Comput Chem 26:879
13. Bublil EM, Freund NT, Mayrose I, Penn O, Roitburd-Berman A, Rubinstein ND, Pupko T, Gershoni JM (2007) Proteins 68:294
14. Enshell-Seijffers D, Denisov D, Groisman B, Smelyanski L, Meyuhas R, Gross G, Denisova G, Gershoni JM (2003) J Mol Biol 334:87
15. Tarnovitski N, Matthews LJ, Sui J, Gershoni JM, Marasco WA (2006) J Mol Biol 359:190
16. Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, Sharan R, Pupko T (2007) Nucleic Acids Res 35:69
17. Huang J, Gutteridge A, Honda W, Kanehisa M (2006) BMC Bioinformatics 7:451
18. Moreau V, Granier C, Villard S, Laune D, Molina F (2006) Bioinformatics 22:1088
19. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, Freund NT, Bublil EM, Ruppin E, Sharan R, Gershoni JM, Martz E, Pupko T (2007) Bioinformatics 23:3244
20. Pacios LF, Tordesillas L, Palacin A, Sanchez-Monge R, Salcedo G and Diaz-Perales A (2011) J Chem Inf Model, 51 1465
21. Caflisch A, Karplus M (1995) J Mol Biol 252:672
22. Caflisch A (1996) J Comput Aided Mol Des 10:372
23. Zeng J, Treutlein HR, Rudy GB (2001) J Comput Aided Mol Des 15:573
24. Zeng J (2000) Comb Chem High Throughput Screen 3:355
25. Zeng J, Nheu T, Zorzet A, Catimel B, Nice E, Maruta H, Burgess AW, Treutlein HR (2001) Protein Eng 14:39
26. Zeng J, Treutlein HR (1999) Protein Eng 12:457
27. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA (1998) Nature 393:648
28. Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535
29. Chen R, Chen W, Yang S, Wu D, Wang Y, Tian Y and Shi Y (2011) BMC Bioinformatics, 12 311

30. Mackerell A et al (1998) J Phys Chem B 102:3586
31. Simonson T, Brunger AT (1994) J Phys Chem 98:4683
32. The PyMOL Molecular Graphics System (2012) Version 1.5.0.4 Schrödinger, LLC
33. Tang X, Yang C, Gu Y, Song C, Zhang X, Wang Y, Zhang J, Hew CL, Li S, Xia N, Sivaraman J (2011) Proc Natl Acad Sci U S A 108:10266–10271
34. Xu R, Karuse JC, McBride R, Paulson JC, Crowe JE Jr, Wilson IA (2013) Nat Struct Mol Biol 20:363–370
35. Prabakaran P, Gan J, Wu YQ, Zhang MY, Dimitrov DS, Ji X (2006) J Mol Biol 357:82–99
36. Săli A, Blundell TL (1993) J Mol Biol 234:779–815