BIOINFORMATICS ORIGINAL PAPER Vol. 27 no. 18 2011, pages 2529–2536
doi:10.1093/bioinformatics/btr437

Structural bioinformatics

Advance Access publication July 22, 2011

Structural analysis of the hot spots in the binding between H1N1 HA and the 2D1 antibody: do mutations of H1N1 from 1918 to 2009 affect much on this binding?

Qian Liu[1], Steven C. H. Hoi[1], Chinh T. T. Su[1], Zhenhua Li[1], Chee-Keong Kwoh[1], Limsoon Wong[2] and Jinyan Li[2,3,*]

[1]School of Computer Engineering, Nanyang Technological University, Singapore, [2]School of Computing, National University of Singapore, Singapore and [3]Advanced Analytics Institute, University of Technology Sydney, Australia

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Worldwide and substantial mortality caused by the 2009 H1N1 influenza A has stimulated a new surge of research on H1N1 viruses. An epitope conservation has been learned in the HA1 protein that allows antibodies to cross-neutralize both 1918 and 2009 H1N1. However, few works have thoroughly studied the binding hot spots in those two antigen–antibody interfaces which are responsible for the antibody cross-neutralization.

Results: We apply predictive methods to identify binding hot spots at the epitope sites of the HA1 proteins and at the paratope sites of the 2D1 antibody. We find that the six mutations at the HA1's epitope from 1918 to 2009 should not harm its binding to 2D1. Instead, the change of binding free energy on the whole exhibits an increased tendency after these mutations, making the binding stronger. This is consistent with the observation that the 1918 H1N1 neutralizing antibody can cross-react with 2009 H1N1. We identified three distinguished hot spot residues, including Lys^{166}, common between the two epitopes. These common hot spots again can explain why 2D1 cross-reacted. We believe that these hot spot residues are mutation candidates which may help H1N1 viruses to evade the immune system. We also identified eight residues at the paratope site of 2D1, five from its heavy chain and three from its light chain, that are predicted to be energetically important in the HA1 recognition. The identification of these hot spot residues and their structural analysis are potentially useful to fight against H1N1 viruses.

Contact: jinyan.li@uts.edu.au

Availability: *Z*-score is available at http://155.69.2.25/liuqian/indexz.py

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on April 10, 2011; revised on June 19, 2011; accepted on July 19, 2011

1 INTRODUCTION

The H1N1 influenza A caused two notable pandemics with substantial mortality in 1918 and 2009. Fortunately, it has been found that some antibodies can work against the Hemagglutinin (HA) proteins in these two pandemics (Xu *et al.*, 2010). HA is a homotrimeric glycoprotein. HA monomers are synthesized as precursors that are then cleaved into two proteins, HA1 and HA2, which form the major surface proteins of influenza A viruses. The infection is started by the binding of HA proteins to the sialic acid-containing receptors of target cells and by fusing the viral membrane with the endosomal membrane of the target cells. The viral genome enters and infects the target cells after the binding. So, inhibiting this binding by antibodies is an important way against flu. Previous works have learned that there is an epitope (binding site) conservation that exists between the 1918 and 2009 H1N1 HA proteins (Ekiert *et al.*, 2009; Xu *et al.*, 2010). Such epitope conservation enables the older population to avoid infection from 2009 H1N1 because their pre-existing immunity against 1918 H1N1 can neutralize the 2009 H1N1 HA proteins. Thus, studies on these antibody–HA binding interfaces are crucial to understand how the antibodies recognize the antigens. However, there are few studies on the energetic importance of the binding residues in the HA1 protein in complex with the 2D1 antibody.

We apply predictive and comparative methods to examine the interfaces between the 2D1 antibody and the HA1 proteins of 1918 and 2009 H1N1, and to investigate an assumed 2D1 binding to the seasonal influenza virus A/Brisbane/59/2007 to understand why 2D1 did not bind to the 2007 strain (Krause *et al.*, 2010; Xu *et al.*, 2010). This 2D1 antibody is a monoclonal antibody from a survivor of the 1918 Spanish influenza (Yu *et al.*, 2008), which is believed to bind to HA1s in both of 1918 and 2009 H1N1. Of particular interests, we identify *binding hot spot residues* from the above mentioned two antibody–antigen interfaces. A binding hot spot is a small fraction of interfacial residues that contribute most to binding free energy (Bogan and Thorn, 1998; Clackson and Wells, 1995). Their mutations—e.g. alanine mutations—can reduce binding affinity remarkably (Clackson and Wells, 1995).

We address the problems whether the interfacial mutations from 1918 H1N1's HA1 to 2009's are hot spot residues and whether these mutations make the binding stronger with 2D1. We explain how the computational methods find those antigenic residues that are energetically important in the antibody binding, such as Asn^{129}, Lys^{157} and Lys^{166}. These three hot spot residues are actually common between the 1918 and 2009 epitopes of HA1. Their mutations may make the 2D1-antibody binding ineffective.

*To whom correspondence should be addressed.

Therefore, they are mutation candidates which may help H1N1 viruses evade the immune system. We also describe and characterize hot spot residues at the paratope site of the 2D1 antibody—e.g. Asp^{52} and Arg^{97} from the heavy chain and Asn^{31}, Trp^{91} and Asp^{93} from the light chain. Knowledge gained from these binding hot spot studies can be useful to fight against H1N1 viruses in future.

2 METHODS

2.1 The HA-2D1 binding structures

We retrieve from the PDB entry 3LZF the crystal structure of the 2D1 antibody binding to the HA proteins of 1918 H1N1 (1918HA1), and from 3LZG the atomic coordinates of the HA proteins of the 2009 H1N1 A/California/04/2009 (Xu *et al.*, 2010) (2009HA1). The structure information of the HA1 proteins of A/Brisbane/59/2007 (2007HA1) is taken from Igarashi *et al.* (2010). Our comparative analysis is on the three interfaces: the interface between 2D1 and 1918HA1 (2D1-1918HA1), the interface between 2D1 and 2009HA1 (2D1-2009HA1) and an assumed artificial interface between 2D1 and 2007HA1 (2D1-2007HA1).

We use MAFFT to align the three HA1 sequences to examine the specific mutations among 1918HA1, 2009HA1 and 2007HA1. We also take the following steps to produce the structure of the 2D1-2009HA1 (and 2007HA1) binding. First, we use PyMOL to align the HA protein structure of 2009HA1 (or 2007HA1) onto the HA protein structure of the 1918 H1N1 in the PDB 3LZF with the antibody coordinates. PyMOL aligns the two HA protein structures to minimize root mean square deviation (RMSD) (Schrödinger, LLC, 2010). After that, we obtain the computational binding interface of 2009HA1 (or 2007HA1) with the 2D1 antibody by removing the HA coordinates of 1918 H1N1. 2009HA1 (or 2007HA1) in this computational binding interface remains in an unbound state with free side chains. So, we use FoldX (Schymkowitz *et al.*, 2005) to repair this interface when fixing the antibody binding site. The repaired interfaces are then used for our subsequent analysis.

2.2 Computational methods for predicting hot spots

Binding hot spot residues can be predicted by computational methods such as by Robetta (Kortemme and Baker, 2002), FoldX (Schymkowitz *et al.*, 2005), KFC (Darnell *et al.*, 2008), GCR (Li and Li, 2010) and a *Z*-score method.

Robetta is a simple physical model for estimating the binding energy of hot spots. This method uses all heavy atoms and polar hydrogens to represent proteins and proposes a free energy function for linearly combining such terms as Lennard–Jones potentials, an orientation-dependent hydrogen bond potential, Coulomb electrostatics and an implicit solvation model. Similarly, FoldX (Schymkowitz *et al.*, 2005) uses a linear combination of empirical terms to calculate free energy. The empirical terms are hydrophobic and polar solvation, hydrogen bonds (water-intermediate hydrogen bonds included), the Van der Waals terms, Coulomb electrostatics, and so on. Meanwhile, KFC (Darnell *et al.*, 2008) uses simple rules to identify binding hot spots. The following features are employed by KFC to represent a residue: physical and chemical features, shape specificity and biochemical contacts such as atomic contacts, hydrogen bonds and salt bridges. Then it uses a decision tree model to produce some rules for classifying hot spots. All these computational methods achieved good prediction performance based on experimental mutations. For example, the overall correlation between the observed and Robetta-calculated changes in binding free energy has an average unsigned error of 1.06 kcal/mol for interface mutations (Kortemme and Baker, 2002).

Recently, a novel descriptor of atoms and residues, called burial level by GCR (Li and Li, 2010), is also proposed to enhance hot spot prediction performance. By this method, an atomic contact graph is built for a protein complex, where vertices are atoms and edges are atom contacts. The burial level of an atom in this graph is defined as the length of the shortest path from this atom to its nearest exposed atom to the bulk solvent. The burial level of an atom or a residue indicates the extent it is buried inside the complex. As the hot spot residues are protected by O-rings (Bogan and Thorn, 1998; Li and Liu, 2009), hot spot residues always have low solvent accessible surface area (ASA) and high burial levels. But a high burial level is more sufficient than ASA: there are very few highly buried interfacial residues that are not hot spot residues. We have built a hot spot model (Li and Li, 2010) based on this concept; and the model has achieved good performance.

We have also proposed a *Z*-score biological significance for capturing the probability of residues occurring in or contributing to protein binding interfaces. This *Z*-score is actually intended to measure how far away certain properties of a putative contact residue at a binding interface are from those of crystal packing. So, we take crystal packing as the reference state to extract residue pairwise potentials. Then, the potential score of a residue is defined by using a knowledge-based potential function with ASA calculations. After that, a null distribution of this potential score is generated from artifact crystal packing contacts. Finally, the *Z*-score significance of a contact residue with a specific potential score is determined according to this null distribution. As binding hot spots contribute greatly to binding free energy, they should have big *Z*-score values. Here, a contact residue is considered as a hot spot residue if its *Z*-score is >1. Our evaluation on the ASEdb and BID datasets (Cho *et al.*, 2009) shows that *Z*-score is powerful for identifying protein binding hot spots. The details of how to calculate *Z*-score are given in the Supplementary Material.

2.3 A meta-learning approach to combine the computational methods for predicting hot spots

We use the computational methods above to predict whether contact residues are hot spot residues or not after alanine mutations in the three interfaces—2D1-1918HA1, 2D1-2007HA1 and 2D1-2009HA1. We use default parameters for the Robetta and KFC web servers and for the FoldX software. Since Robetta and FoldX estimate $\Delta\Delta G$, we are interested in those residues whose alanine mutation results in $\Delta\Delta G \geq 1$ kcal/mol. After that, we apply a meta-learning approach (Vilalta and Drissi, 2002) by combining the *Z*-score method with the other methods. The reason is that the *Z*-score method has a very high recall with low precision rate; however, the other methods generally have a low recall but a high precision rate. Therefore, in this work, we are interested in those hot spot residues that are predicted by *Z*-score and are also confirmed by at least one of the other methods (Robetta, FoldX or KFC). Meanwhile, we also trust with high confidence that non-hot spot residues predicted by *Z*-score generally have insignificant contribution to the binding. The hot spot residues which are predicted by a single method only are considered having intermediate contribution to binding.

3 RESULTS

The sequence alignment among 1918HA1, 2007HA1 and 2009HA1 is shown in Figure 1a. There are a total of six interfacial mutations between 1918HA1 and 2009HA1, namely E131D, T133N, S159N, V169I, N171D and T242K. The structural alignment of 1918HA1 and 2009HA1 between their interfacial segments is shown in Figure 1c where the structural match—based on the Cαs of these interfacial residues—has an RMSD of 0.725 Å. Previous works have reported that protein sequences with ~50% identity or above in crystallographic models can differ by ~1 Å RMSD, while proteins in NMR models can have even larger deviations (Chothia and Lesk, 1986; Schwede *et al.*, 2000). In some cases, sequences with >95% identity can also have an interface RMSD up to ~1.2 Å (Kinjo and Nakamura, 2010). Thus, the small 0.725 Å RMSD suggests that the interfacial segments of the two HA1 proteins have a very good match. This indicates that the mutations from 1918HA1 to 2009HA1 resulted in little structural change at the binding site, and that the

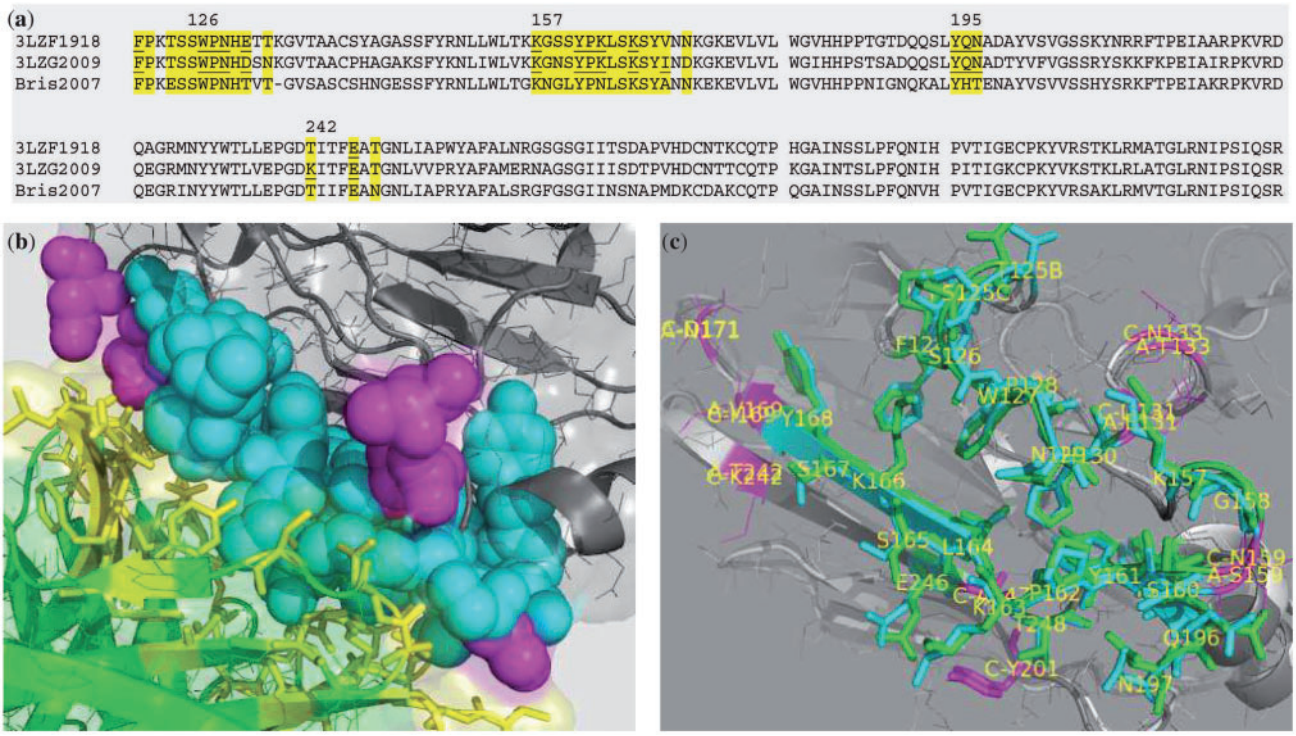

Fig. 1. The epitope of HA1 in 1918, 2009 and 2007, and their sequence and structural alignment (better viewed in color). (**a**) The sequence alignment between 1918HA1(3LZF1918), 2009HA1(3LZG2009) and 2007HA1(Bris2007). Interface residues are in yellow, and the positions are in accordance to the 1918HA1 numbering. (**b**) The binding interface of 2D1-1918HA1. The HA1 epitope is in cyan and in sphere view, and the antibody paratope is in yellow and in stick view. (**c**) The aligned structure of the epitopes between 1918HA1 (cyan) and 2009HA1 (green). In (b) and (c), residues in magenta are the mutations from 1918HA1 to 2009HA1.

computationally produced binding structure of 2009HA1 and the 2D1 antibody may not have any big deviation from the real one.

3.1 Energy change tendency of the six mutations

Using the *Z*-score method, three of the six mutations—T133N, S159N and N171D—are predicted as non-hot spot residues in both 1918HA1 and 2009HA1. Two mutations—V169I and T242K—are believed to contribute, though slightly, to the antibody binding in 2009HA1 only after the mutations. They may be newly formed hot spots in the epitope of 2009HA1 after the mutations. The remaining one of the six mutations—E131D—is predicted to contribute to the binding free energy both before and after the mutation. Robetta, FoldX and KFC predict all the mutations as non-hot spot residues in both 1918HA1 and 2009HA1. However, $\Delta\Delta G$ of V169I is predicted to increase from 0.4 to 0.76 kcal/mol by Robetta and from 0.07 to 0.83 kcal/mol by FoldX. Hence, the six mutations from 1918HA1 to 2009HA1 do not appear to adversely affect the binding between the 2D1 antibody and the two HA1s. Instead, on the whole, the change of binding free energy exhibits a possible increased tendency after the mutation, making the binding stronger. This is consistent with the result that the 1918HA1 neutralizing 2D1 antibody can cross-react with 2009HA1 (Xu *et al.*, 2010).

Geometrically, the six mutations are located at the rim of the binding interface (Fig. 1b and c), forming a part of an O-ring structure (Bogan and Thorn, 1998; Li and Liu, 2009). Most of them have a large exposed portion to water and are not deeply buried. Their absolute and relative ASA information and burial levels are presented in Table 1. Only the residue at position 131 is buried with little exposure to water. The influential O-ring theory (Bogan and Thorn, 1998; Li and Liu, 2009) stated that residues on an O-ring, though of structural importance, are usually not energetically important. Therefore, these six mutations can only slightly destruct the antibody binding of 2009HA1 energetically if they are adverse.

We closely examined the S159N mutation from 1918HA1 to 2009HA1. It was reported to have a high $\Delta\Delta G$ ~2 kcal/mol (Xu *et al.*, 2010). However, we believe that this residue does not contribute greatly to the antibody binding either before or after the mutation. First, Ser^{159} has no significant contact with the antibody; see Figure 3a. Second, the S159N mutation makes the backbone deviate far away—the Cα deviation is ~1.1 Å. This results from an increased flexibility for Asn^{159} side chain. At the side chain of Ser^{159}, OG has a hydrogen bond with its backbone O. The distance between O and OG is 2.2 Å, and the angle of OG-H···O is 151.0°. Therefore, it is this hydrogen bond that confines Ser^{159} side chain. But the mutation makes Asn^{159} side chain free to contact water molecules, which can drag and affect the backbone structure a lot. Third, the mutation breaks the hydrogen bond of OG and the backbone O at Ser^{159}, releasing some binding free energy. In fact, at least half of the 2 kcal/mol $\Delta\Delta G$ of the mutation should come from this hydrogen bond break. This can be seen from the mutation

Table 1. ASAs and burial levels of the six mutation residues from 1918HA1 to 2009HA1

Position	133		159		171		169		242		131	
Residue	Thr[a]	Asn[b]	Ser[a]	Asn[b]	Asn[a]	Asp[b]	Val[a]	Ile[b]	Thr[a]	Lys[b]	Glu[a]	Asp[b]
Absolute ASA (Å)	94.27	104.14	39.36	80.06	111.45	109.72	29.57	43.7	45.79	79.59	15.13	8.42
Relative ASA (%)	67.7	72.3	33.8	55.6	77.4	78.2	19.5	25.0	32.9	39.6	8.8	6.0
Burial level	0.43	0.5	0.83	0.63	0.5	0.63	0.86	0.75	0.57	0.78	1.22	1.25

[a]Residues in 1918HA1.
[b]Residues in 2009HA1.

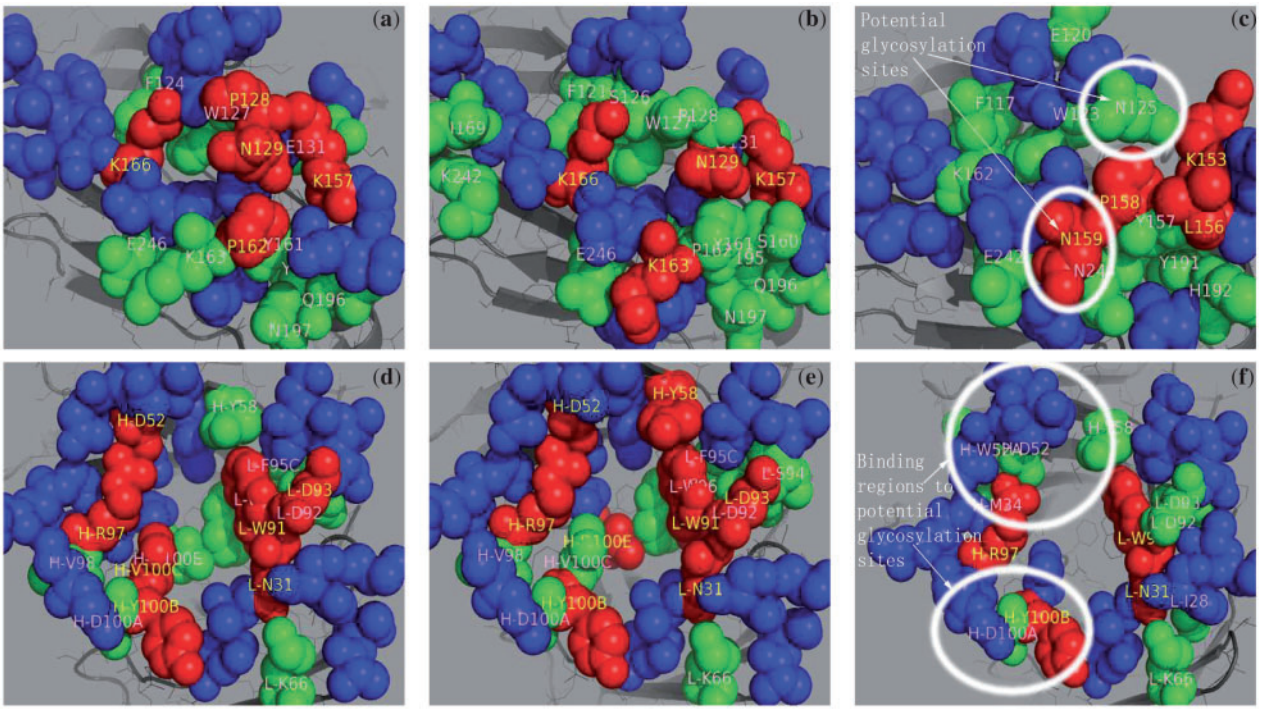

Fig. 2. The binding hot spots in the HA-2D1 binding interfaces (better viewed in color). (**a**) The hot spots in the 1918HA1 epitope. (**b**) The hot spots in the 2009HA1 epitope. (**c**) The artificial hot spots in the 2007HA1 epitope and their glycosylation sites. (**d**) The hot spots in the antibody paratope of 2D1-1918HA1. (**e**) The hot spots in the antibody paratope of 2D1-2009HA1. (**f**) The hot spots in the artificial paratope of 2D1-2007HA1. (a) and (d) form 2D1-1918HA1, (b) and (e) form 2D1-2009HA1, and (c) and (f) form artificial 2D1-2007HA1. In (a)–(f), binding sites are in sphere view; residues in red (with yellow labels) are predicted by *Z*-score to be hot spot residues and confirmed by at least one of other computational methods; residues in green (with light pink labels) are predicted only by one method; residues in blue are predicted as non-hot spots by all methods.

of Ser^{159} to Gly^{159} (no side chain in Gly), whose $\Delta\Delta G$ is bigger than 1 kcal/mol (Xu *et al.*, 2010). Therefore, the S159N mutation from 1918HA1 to 2009HA1 did not greatly destroy the binding of the 2D1 antibody to 2009HA1.

As the *Z*-score method has a high negative precision value for predicting non-hot spot residues, the above residues predicted as non-hot spot residues can be considered as energetically unimportant to the antibody binding with high confidence. The additional 12 non-hot spot predictions in the HA1 epitopes are Pro(^{125} in 1918HA1, ^{122} in 2009HA1), Thr(^{125B} in 1918HA1, ^{124} in 2009HA1), Ser(^{125C} in 1918HA1, ^{125} in 2009HA1), Ser^{126}, His^{130}, Gly^{158}, Ser^{160}, Leu^{164}, Ser^{165}, Ser^{167}, Tyr^{168} and Thr^{248}. Some of these non-hot spot predictions can be verified by past non-alanine mutation experiments (Xu *et al.*, 2010). For example, G158E/D, S160L or S165K cause only a small $\Delta\Delta G$ (<1 kcal/mol). These also happened between different residue-type groups, e.g. a mutation from a polar uncharged residue Ser^{160} to a hydrophobic residue Leu^{160}. These suggest that these non-hot spot residues have little contribution to the binding to either 1918HA1 or 2009HA1, just as *Z*-score predicts. So, mutating these predicted non-hot spot residues provides little chance for H1N1 to evade capture by the 2D1 antibody.

3.2 Hot spot residues at the epitopes of the two HA1s

The hot spot residues in 2D1-1918HA1 or 2D1-2009HA1 predicted by Robetta, FoldX, KFC and *Z*-score are shown in Figure 2. All of them are considered to have potential contribution to the antibody

Table 2. Binding hot spots in the two epitopes predicted by *Z*-score and confirmed by other previous computational methods

Residues	Robetta		FoldX		KFC			*P*-value*		Absolute ASA (Å)		Relative ASA (%)		Burial level	
	HA1[a]	HA1[b]	HA1[a]	HA1[b]	HA1[a]	HA1[b]	2007HA	HA1[a]	HA1[b]	HA1[a]	HA1[b]	HA1[a]	HA1[b]	HA1[a]	HA1[b]
Pro^{128}			~~0.83~~	~~0.9~~	√		No	0.02	0.0465	17.42	29.5	12.8	21.7	1.57	1.29
Asn^{129}				1.4	√		No	0.0312	0.0234	27.37	23.01	19.0	16.0	1.25	1.25
Lys^{157}	1.35			1.54	√	√	Yes_{153}	<1E-324	<1E-324	9.69	13.38	4.8	6.7	1.44	1.55
Pro^{162}			1.36	~~0.73~~			Yes_{158}	0.0027	0.0018	0.07	0.00	0.1	0.0	2.00	2.00
Lys^{163}	~~0.92~~	1.33		1.43			Yes^{c}_{159}	<1E-324	<1E-324	35.81	32.43	17.8	16.2	1.33	1.11
Lys^{166}	2.98	1.14	4.2	3.73			No	<1E-324	<1E-324	4.19	3.05	2.1	1.5	1.67	1.67

The subscript number is the position in 2007HA1.
[a]Residues in 1918HA1.
[b]Residues in 2009HA1.
[c]Asn mutation creates a potential glycosylation site in 2007HA1 (Xu *et al.*, 2010).
*P-values of Z-score.

binding. In particular, Pro^{128}, Asn^{129}, Lys^{157}, Pro^{162}, Lys^{163} and Lys^{166} are confirmed as hot spot residues at 2D1-1918HA1 or at 2D1-2009HA1 by *Z*-score and at least one of the other methods; see Table 2 and Figure 2a and b. Three of them are common in both 2D1-1918HA1 and 2D1-2009HA1. For the other three residues, two of them (Pro^{128} and Pro^{162}) are hot spot residues for 2D1-1918HA1 and also with not low $\Delta\Delta G$ in 2D1-2009HA1 (Table 2). Similar observation can be found for the remaining one (Lys^{163}) in 2D1-2009HA1. These residues all have a very small ASA, and are buried with a burial level up to 2.0 (Table 2). These double-confirmed hot spot residues are believed to contribute greatly to the antibody binding, as the combined prediction by *Z*-score and the other computational methods has a much higher precision. So, they are positions for mutations that can lead to H1N1's escape from 2D1's neutralization.

Some of these double-confirmed hot spot residues at 2D1-1918HA1 or 2D1-2009HA1 are also supported by wet-lab experiments. For example, the mutation of Asn to Lys at position 129 and the mutation of Lys to Asn at position 163 in wet-lab experiments resulted in >1 kcal/mol $\Delta\Delta G$ (Xu *et al.*, 2010). This fact indicates that Asn^{129} and Lys^{163} are truly energetically important although the mutations are non-alanine mutations.

Lys^{166} has been comprehensively studied in the past by wet-lab experiments. It was found that this residue contributes greatly to this antibody binding: its mutations to residue types such as its similar hydrophilic residues Glu and Gln, or Pro resulted in >3 kcal/mol $\Delta\Delta G$ (Xu *et al.*, 2010). As can be seen in Table 2, Lys^{166} is predicted by three computational methods (*Z*-score, Robetta and FoldX) as a hot spot residue in both 1918HA1 and 2009HA1 epitopes. To investigate why this residue is energetically so important, we examine its contacts using Figure 3b. First, this residue is deeply buried with a small ASA and high burial level. Second, Lys^{166} has several hydrogen bonds with its NZ atom as the donor: one hydrogen bond forms with the backbone O of the Ser^{126} from the same chain as Lys^{166}, and the other two form with the side chain O of Asp^{93} and of Asn^{31} from the light chain of the antibody. Third, CE of Lys^{166} has a π-involving contact with Trp^{127} from the same chain as Lys^{166}. These contacts suggest that this residue contributes greatly to the antibody binding and to the antigen folding by the π-involving contact with Trp^{127} and the hydrogen bond with the Ser^{126}. In fact, Ser^{126} is also at the epitope site, and it is predicted as a hot spot residue by FoldX and Robetta in the 2009HA1 epitope site but as a non-hot spot residue in the 1918HA1 by all the methods; so this hydrogen bond contributes to the antibody binding indirectly.

The residue Lys^{166} was also reported as a selected escape mutation at 2D1-antibody by several viruses including 2009 H1N1, 1918 H1N1 and the 1930 swine viruses (Krause *et al.*, 2010; Xu *et al.*, 2010; Yu *et al.*, 2008). What is more interesting is that Lys^{166}'s two hydrogen bond contact residues, Asp^{93} and Asn^{31}, are all predicted to be hot spot residues by *Z*-score and more than one other methods, instantiating the hot spot coupling property (Halperin *et al.*, 2004).

Although we have not found wet-lab evidence and report for Pro^{128}, Lys^{157} or Pro^{162}, we suggest that all of the six double-confirmed hot spot residues—Pro^{128}, Asn^{129}, Lys^{157}, Pro^{162}, Lys^{163} and Lys^{166}—are potential escape mutations for H1N1 to elude the 2D1 antibody.

3.3 Hot spot residues at 2D1's paratope

We also studied those residues in the paratope (the antigen binding site) of the 2D1 antibody that can contribute greatly to the binding. These antibody hot spot residues can uncover how the 2D1 antibody captures the H1N1 viruses.

Using the *Z*-score method, all and only six hot spot residues are predicted in the antibody light chain which are also common between 2D1-1918HA1 and 2D1-2009HA1. Meanwhile, eight hot spot residues in the antibody heavy chain are identified in 2D1-1918HA1, and seven in 2D1-2009HA1. These predicted paratope hot spots are depicted in Figure 2d and e. Among them, five from the heavy chain and three from the light chain are confirmed by more than one existing computational methods (Table 3). In the antibody light chain, the hot spot residues Asp^{93} and Asn^{31} have significant contacts with the antigen hot spot residue Lys^{166} as we discussed above. We believe that they are mainly responsible for the binding to the antigen.

In the antibody heavy chain, we are interested in the predicted hot spot residue Arg^{97}, as it is predicted to be energetically important by three methods in both 2D1-1918HA1 and 2D1-2009HA1 (Table 3).

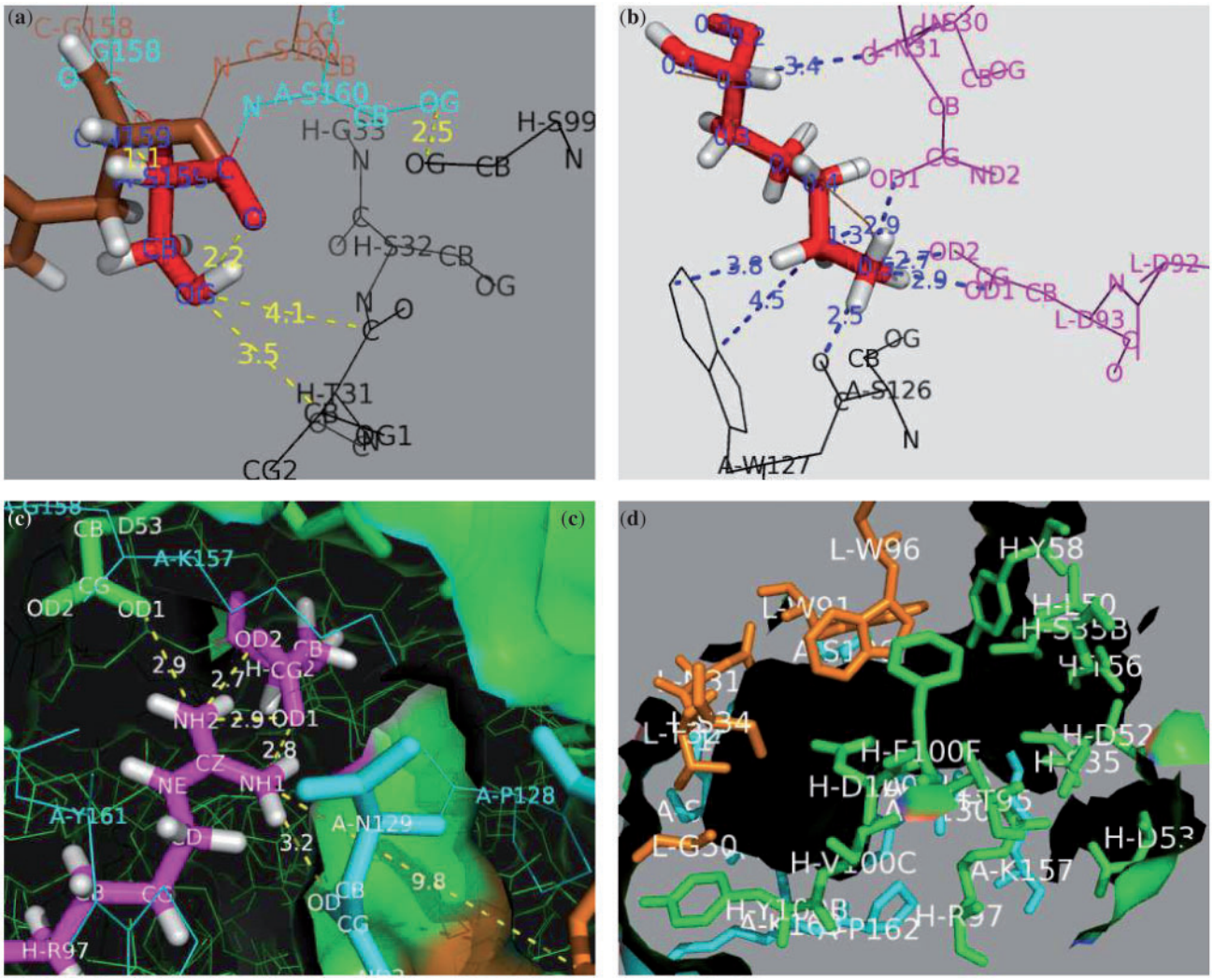

Fig. 3. Three examples of (non-)hot spot predictions in 2D1-1918HA1 and 2D1-2009HA1 (better viewed in color). (**a**) The mutation of Ser^{159} to Asn^{159} from 1918HA1 (cyan) to 2009HA1 (brown) when binding to 2D1 (the heavy chain in black); position 159 is in red and in stick view. (**b**) The hot spot residue Lys^{166} when both 1918HA1 (black) and 2009HA1 (brown) bind to 2D1 (the light chain in purple); the residues Lys^{166} are in red and in stick view. (**c**) The hot spot residues Asp^{52} and Arg^{97} in magenta in the paratope of the antibody heavy chain. (**d**) The cavity in the binding interface surrounded by the binding residues. In (c) and (d), the whole complex and the core interface are shown in surface view, and so the epitope and the paratope have no surface; 1918HA1, the antibody heavy chain and the antibody light chain are in cyan, green and brown, respectively.

Its close contact with Asp^{52} is shown in Figure 3c. Asp^{52} is also from the antibody heavy chain and confirmed as a hot spot residue by Robetta and FoldX. As seen in Figure 3c, Arg^{97} has two hydrogen bonds with Asp^{52} with NH1 and NH2 as donors and OD1 and OD2 as acceptors. Furthermore, Arg^{97}'s NH2 has a hydrogen bond with OD1 from Asp^{53}, and Arg^{97}'s NH1 has another hydrogen bond with OD1 from the antigen residue Asn^{129}. Asn^{129} is considered to contribute greatly to the antibody binding by the two existing methods (Table 2). These contacts form a hydrogen-bond network which is believed to generate a favorable electrostatic contribution to the protein binding that can strongly stabilize the protein complexes (Sheinerman and Honig, 2002). Another finding is that the side chain of Arg^{97} confines at least three side chains: the side chain of Arg^{97} which is positively electrically charged, and the side chains of Asp^{52} and Asp^{53} which are negatively charged. These side chains should prefer solvent water molecules if they are free. So, confining by the side chain of Arg^{97} can make remarkable contribution to the antibody binding by removing the freedom of these three charged side chains.

Our investigation also finds a large cavity at the core of the binding interface, as seen in Figure 3c and d. This cavity has a surface surrounded by the binding residues. Its narrowest part is >9 Å wide, which is equivalent to more than three water molecule diameters (2.75 Å). What is more important is that the side chains of Arg^{97} and Asp^{52} are in the rim of the cavity, and some side chain atoms of Asp^{53} contact the solvent as seen in Figure 3c and d. Thus, removing the side chains of Arg^{97} and/or Asp^{52} would increase the chance that this cavity is open to contact the solvent. In other words, the binding

Table 3. Binding hot spots in the antibody paratopes predicted by *Z*-score and confirmed by other previous computational methods

Residues	Robetta		FoldX		KFC			*P*-value*		Absolute ASA (Å)		Relative ASA (%)		Burial level	
	2D1[a]	2D1[b]	2D1[a]	2D1[b]	2D1[a]	2D1[b]	2007HA	2D1[a]	2D1[b]	2D1[a]	2D1[b]	2D1[a]	2D1[b]	2D1[a]	2D1[b]
Asp^{52}	3.93	2.85	1.06				No	0.2522	0.1735	2.44	2.06	1.7	1.5	1.88	1.88
Tyr^{58}	~~0.96~~	1.35					No	0.0585	0.1691	23.57	26.04	11.1	12.2	1.33	1.33
Arg^{97}	4.5	4.57	1.32	3.53			Yes[c]	0.0075	0.0106	0.71	0.69	0.3	0.3	2.00	2.00
Tyr^{100B}	2.28	1.14					Yes[c]	0.0178	0.015	26.35	33.65	12.4	15.8	1.17	1.08
[d]*Val*^{100C}			1.02	1.16			No[c]	0.2539	0.3271	16.97	16.5	11.2	10.9	1.57	1.57
[e]**Asn**^{31}	2.44	1.56	1.68	1.86			No	0.3236	0.3251	11.30	10.62	7.9	7.4	1.5	1.88
Trp^{91}	3.05	3.02	2.96	2.7		√	Yes	<1E-324	<1E-324	12.39	10.99	5.0	4.4	1.29	2.14
Asp^{93}	3.08	3.32		2.78		√	No	0.016	0.0065	13.34	11.13	9.5	7.9	1.25	1.25

Italic are for the antibody heavy chain, and bold for the light chain.
[a]Residues in 2D1-1918HA1.
[b]Residues in 2D1-2009HA1.
[c]Paratope residues which are close to potential glycosylation sites in 2D1-2007HA1.
[d]*Z*-score failed to identify this hot spot in 2D1-2009HA1.
[e] *Z*-score failed to identify this hot spot.
**P*-values of *Z*-score.

interface would have a larger open empty or solvent core. This makes it impossible to have stable binding. Hence, we believe that both Arg^{97} and Asp^{52} are energetically significant in the binding.

3.4 Analysis on the assumed 2D1-2007HA1 binding

The hot spot prediction results on the assumed artificial 2D1-2007HA1 interface are also presented in Tables 2, 3 and Figure 2c and f. We find that more than half of the predicted hot spots in the binding of 2D1-1918HA1 or 2D1-2009HA1 are not predicted to make contribution to the artificial binding 2D1-2007HA1, although Leu^{156} is newly predicted as a hot spot residue at the epitope of 2D1-2007HA1.

We have two interesting remarks about this assumed binding. One remark is about Lys^{162} in 2007HA1 (Lys^{166} in 1918HA1 and 2009HA1) which is predicted as a non-hot spot residue. First, this residue is conserved in the three HA1s. Second, it is predicted as a hot spot residue in 1918HA1 and 2009HA1 by all the computational methods here. Third, its binding importance in 2D1-1918HA1 and 2D1-2009HA1 has been demonstrated by Xu *et al.* (2010). So, the reduced contribution of this Lys in 2D1-2007HA1 indicates an escape of 2007HA1 from 2D1. This together with the less number of predicted hot spots in 2D1-2007HA1 suggests a very small occurrence probability of 2D1-2007HA1. The other remark is that although Asn^{159} in 2007HA1 (Lys^{163} in 1918HA1 and 2009HA1) is predicted to contribute to the artificial 2D1-2007HA1 binding, this mutation creates a potential *N*-glycosylation site (Xu *et al.*, 2010) as shown in Figure 2c. 2007HA1 contains another glycosylation site Asn^{125} (Xu *et al.*, 2010); see Figure 2c. To better understand the assumed binding, we use Figure 2f to depict the binding region of 2D1's paratope to the glycosylation sites, and this region covers two of the three predicted hot spot residues of the paratope. It can be observed that the glycosylation sites mask the surface of 2007HA1 to block the cross-neutralization by 2D1 (Xu *et al.*, 2010). The computational methods did make some predictions of hot spots in 2D1-2007HA1, because none of them considers the potential glycosylation sites but only the residue information. In fact, these hot spot predictions are not true if the glycosylation sites are considered. In summary, 2D1 cannot recognize 2007HA1 for neutralization.

4 CONCLUSION

We have done a structural analysis on the interfaces between the 2D1 antibody and the HA1 proteins of 2009 H1N1 and 1918 H1N1. The cross-neutralization of this antibody is clearly demonstrated by the hot spot residues common in the two binding interfaces. Our comprehensive investigation suggests that there are six outstanding epitope residues whose mutations will help H1N1 evade capture by this antibody. We further pinpointed the hot spot residues at the paratope site of the 2D1 antibody which are responsible for the antigen recognition. The understanding of these hot spot residues can potentially facilitate drug design to neutralize influenza viruses.

Funding: Singapore MOE Tier-2 funding grants (T208B2203 and MOE2009-T2-2-004 in part).

Conflict of Interest: none declared.

REFERENCES

Bogan,A.A. and Thorn,K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1–9.

Cho,K.-I. *et al.* (2009) A feature-based approach to modeling protein-protein interaction hot spots. *Nucleic Acids Res.*, **37**, 2672–2687.

Chothia,C. and Lesk,A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.*, **5**, 823–826.

Clackson,T. and Wells,J. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–386.

Darnell,S.J.J. *et al.* (2008) KFC server: interactive forecasting of protein interaction hot spots. *Nucleic Acids Res.*, **36**, W265–W269.

Ekiert,D.C. *et al.* (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science*, **324**, 246–251.

Halperin,I. *et al.* (2004) Protein-protein interactions: coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure*, **12**, 1027–1038.

Igarashi,M. *et al.* (2010) Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. *PLoS One*, **5**, e8553.

Kinjo,A.R. and Nakamura,H. (2010) Geometric similarities of protein-protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study. *J. Mol. Biol.*, **399**, 526–540.

Kortemme,T. and Baker,D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.

Krause,J.C. *et al.* (2010) Naturally occurring human monoclonal antibodies neutralize both 1918 and 2009 pandemic influenza a (H1N1) viruses. *J. Virol.*, **84**, 3127–3130.

Li,J. and Liu,Q. (2009) 'Double water exclusion': a hypothesis refining the O-ring theory for the hot spots at protein interfaces. *Bioinformatics*, **25**, 743–750.

Li,Z. and Li,J. (2010) Geometrically centered region: a "wet" model of protein binding hot spots not excluding water molecules. *Proteins*, **78**, 3304–3316.

Schrödinger,LLC (2010) The PyMOL molecular graphics system, version 1.3r1. Available at http://pymol.sourceforge.net/faq.html#CITE (last accessed date April 2011).

Schwede,T. *et al.* (2000) Protein structure computing in the genomic era. *Res. Microbiol.*, **151**, 107–112.

Schymkowitz,J. *et al.* (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.

Sheinerman,F.B. and Honig,B. (2002) On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.*, **318**, 161–177.

Vilalta,R. and Drissi,Y. (2002) A perspective view and survey of Meta-Learning. *Art. Intell. Rev.*, **18**, 77–95.

Xu,R. *et al.* (2010) Structural basis of preexisting immunity to the 2009 H1N1 Pandemic Influenza Virus. *Science*, **328**, 357–360.

Yu,X. *et al.* (2008) Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature*, **455**, 532–536.