Journal of Molecular Graphics and Modelling 30 (2011) 110–119

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Prediction of probable mutations in influenza virus hemagglutinin protein based on large-scale ab initio fragment molecular orbital calculations

Akio Yoshioka[a], Kaori Fukuzawa[b], Yuji Mochizuki[c], Katsumi Yamashita[d], Tatsuya Nakano[e], Yoshio Okiyama[f], Eri Nobusawa[g], Katsuhisa Nakajima[h], Shigenori Tanaka[a,*]

[a] Graduate School of System Infrormatics, Kobe University, 1-1, Rokkodai, Nada-ku, Kobe 657-8501, Japan
[b] Mizuho Information and Research Institute, Inc., 2-3, Kanda Nishi-cho, Chiyoda-ku, Tokyo 101-8443, Japan
[c] Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1, Nishi-Ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
[d] 1st Manufacturing Industries Solutions Division, NEC Soft Ltd., 1-18-6, Shinkiba, Koto-ku, Tokyo 136-8608, Japan
[e] Division of Safety Information on Drug, Food and Chemicals, National Institute of Health Sciences, 1-18-1, Kamiyoga, Setagaya-ku, Tokyo 158-8501, Japan
[f] Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan
[g] National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku-ku, Tokyo 162-8640, Japan
[h] Department of Virology, Medical School, Nagoya City University, 1, Kawasumi, Mizuho-cho, Mizuho-ku, Nagoya 467-8601, Japan

ARTICLE INFO

Article history:
Received 22 April 2011
Received in revised form 24 June 2011
Accepted 27 June 2011
Available online 6 July 2011

Keywords:
Influenza
Hemagglutinin (HA)
Antibody
Mutation
Fragment molecular orbital (FMO) method

ABSTRACT

Ab initio electronic-state calculations for influenza virus hemagglutinin (HA) trimer complexed with Fab antibody were performed on the basis of the fragment molecular orbital (FMO) method at the second and third-order Møller–Plesset (MP2 and MP3) perturbation levels. For the protein complex containing 2351 residues and 36,160 atoms, the inter-fragment interaction energies (IFIEs) were evaluated to illustrate the effective interactions between all the pairs of amino acid residues. By analyzing the calculated data on the IFIEs, we first discussed the interactions and their fluctuations between multiple domains contained in the trimer complex. Next, by combining the IFIE data between the Fab antibody and each residue in the HA antigen with experimental data on the hemadsorption activity of HA mutants, we proposed a protocol to predict probable mutations in HA. The proposed protocol based on the FMO-MP2.5 calculation can explain the historical facts concerning the actual mutations after the emergence of A/Hong Kong/1/68 influenza virus with subtype H3N2, and thus provides a useful methodology to enumerate those residue sites likely to mutate in the future.

1. Introduction

Hemagglutinin (HA), a major antigenic protein of the influenza virus, is a homo-trimeric glycoprotein. HA plays an important role in the early stage of infection; binding to the receptors (sialic acids) on the host cells and trigger for the fusion between virus and endosome membranes. The receptor binding site (RBS) locates on the membrane-distal globular domain of HA. The neutralizing antibody targeting antigenic regions located around RBS prohibits the virus from binding to the receptors [1,2]. Antigenic variants are generally selected during circulation of the viruses among human population. In these variants, amino acid differences are observed in the antigenic region compared to the original viruses [3]. The structural analyses of HA–antibody complexes have provided the information about the amino acid residues on HA directly interacting with the antibody in the complexes. Amino acid substitutions at these sites allow the virus to escape from the neutralizing antibody [4–6].

In this paper, we first specify the sites of the antigenic region under the antibody pressure based on the X-ray structure of complex of the HA A/Hong Kong/1/68 (H3N2) with the antibody HC63 [7]. This structure is a complex of HA trimer and fragment of antigen binding (Fab) dimer. The pandemic of human influenza in 1968 was caused by the viruses that were human/avian virus reassortants [8]. Moreover, this virus often has undergone antigenic drifts after this outbreak [9]. We may further be faced with influenza pandemics by the causes of antigenic shift with the highly lethal viruses and antigenic drift in the future.

To predict the probable mutation sites of the antigenic region in HA, in an earlier work [10], we developed a practical scheme by combining the inter-fragment interaction energy (IFIE) analysis of fragment molecular orbital (FMO) method with experimental information about the hemadsorption activity of mutants [11,12], in which an HA monomer structure [6] was used. However, it was indicated [7] that the Fab monomer can effectively recognize not only directly linked (e.g., Fab(I) versus HA(I) or Fab(II) versus HA(II))

* Corresponding author. Tel.: +81 78 803 6620; fax: +81 78 803 6621.
E-mail address: tanaka2@kobe-u.ac.jp (S. Tanaka).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.06.011

domains but also cross-linked (e.g., Fab(I) versus HA(II) or Fab(II) versus HA(III)) domains in the complex structure of HA trimer and Fab dimer (see Fig. 1). Thus, we here investigate the interactions of the HA trimer with the Fab dimer by ab initio FMO calculations to quantify the interaction energies at antigenic sites, which would provide a more quantitative measure of how each amino-acid residue is recognized by the Fab antibody.

In the following, we first explain the computational procedure of FMO method in Section 2. Next we show in Section 3 the results for the assessment of the accuracy of calculation, the correlation between hydrogen bonds and IFIE values on the surface of epitopes, and the prediction of amino acid mutations in HA. Concluding remarks are given in Section 4, where we address a possible contribution of the present scheme to the development of influenza vaccine.

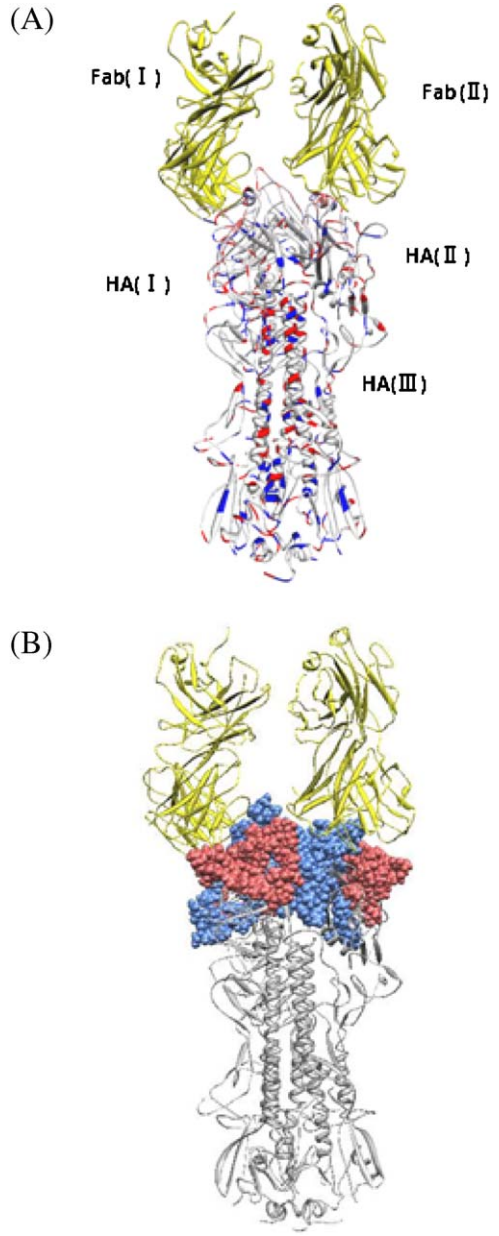

Fig. 1. (A) Visualization of IFIEs between HA trimer (I, II, III) and Fab dimer (I, II) calculated at the FMO-MP2.5/6-31G level. The color represents the sign and strength of the interactions between each residue in the HA trimer and the Fab dimer. For the Fab domain indicated in yellow, the red and blue fragments refer to stabilized and destabilized interactions, respectively, and the deepness of the hue indicates the strength of the interaction. (B) Visualization of antigenic regions A (pink) and B (light blue) by sphere representation. The illustration was created with BioStation Viewer [16]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2. Materials and computational methods

We consider a complex of HA trimer and Fab dimer that is registered in Protein Data Bank (PDB) with ID code of 1KEN [7]. Each HA monomer has the antigenic regions A–E [1,3], although they are further subdivided and sometimes overlaps of these areas have been noted [13,14]. In particular, the Fab has been observed to be bound with the antigenic regions of A, B and E in electron microscope experiments [15]. The neutralization by antibodies is effective for these regions because they are considered to bind to glycoprotein before it binds to sugar moieties of host cell. In the case of 1KEN, it is observed that the Fab dimer is located over both the antigenic regions of A and B (see Fig. 1 [16]). The antibody HC63 [17] also recognizes the multiple HA monomers primarily through the hydrogen bonds, then suggesting the locations of epitopes containing the residues of numbers 136, 137, 153, 158, 159, 186, 187, 189, 190, 192, 193, 225, 226 (direct-linked), and 126, 128, 162, 163, 165 (cross-linked) [7].

The FMO calculations [18,19] have been performed even for enormous proteins like the complex of HA trimer and Fab dimer [20]. We here employ the Hartree–Fock (HF), the second and third-order Møller–Plesset (MP2 and MP3) perturbation methods for comparison. While the HF method is a rough mean-field approximation, the MP2 and MP3 perturbation methods are employed for the description of electron correlations in the present study, where the latter is expected to correct the tendency of overestimation of stabilization energy in the former [21], but too much. Thus, we can accurately evaluate the binding energy in terms of MP2.5 method which provides a half-and-half mixture of MP2 and MP3 energies [20,21].

It is remarked that the effective fragment–fragment interactions in the FMO scheme [22] are obtained in terms of the inter-fragment interaction energy (IFIE) that is defined as

$$\Delta E_{ij} = (E'_{ij} - E'_i - E'_j) + Tr(\Delta P_{ij} V_{ij}), \tag{1}$$

where ΔP_{ij} is a difference density matrix, V_{ij} is an environmental electrostatic potential for fragment dimer ij from other fragments, and E'_i and E'_{ij} are energies of fragment monomer i and dimer ij without environmental electrostatic potential, respectively. These values ΔE_{ij} then represent the interaction energies of an amino acid residue with a ligand or between amino acid residues because each amino acid is assigned as a single fragment in the present analysis [23–25]. The IFIEs were calculated in this study to analyze the interaction pattern and to estimate the contribution of each residue to binding, as seen in Fig. 1.

It is also convenient to introduce [10,26]

$$\Delta E_{IJ}^{\text{total}} = \Sigma_j \Delta E_{ij}, \tag{2}$$

which refers to the contribution of each fragment i to the binding affinity with the domain J containing the grouped residues j. It is noted here that

$$\Delta E_{IJ}^{\text{total}} = \Sigma_i \Delta E_{ij}^{\text{total}} \tag{3}$$

represents the inter–domain interaction between the domain I containing the residues i and the domain J containing the residues j.

The structure of the complex of HA antigen and Fab antibody for the FMO calculations was prepared as follows. Starting with the PDB structure (1KEN), the missing hydrogen atoms were complemented by the MOE (Molecular Operating Environment) software (Chemical Computing Group Inc.). The locations of the hydrogen atoms

were then optimized with the MMFF94x force field. The total numbers of residues and atoms were 2351 and 36,160, respectively. We carried out the FMO-HF (Hartree–Fock), MP2 and MP3 calculations [20] with the basis set 6-31G on the Earth Simulator in Yokohama, as in the previous work [27].

3. Results and discussion

3.1. Analysis of inter–domain interaction energies

We first study the inter–domain interactions in the complex of HA trimer and Fab dimer. As seen in Table 1, the calculated value of the interaction energy between the HA trimer and the Fab dimer is +38.1 kcal/mol at the HF level, while the MP2 and MP3 results have been found to be −163.5 kcal/mol and −127.0 kcal/mol, respectively. The reason why the interaction energy in the HF approximation is positive (repulsive) even though the HA trimer should be attractive to the Fab is that the dispersion energies are not appropriately described. In earlier studies [24,28] with the MP2 method for receptor–ligand systems, it has been demonstrated that the dispersion energies play an important role for describing the interactions of biomacromolecules. We evaluate the interaction energy as −145.3 kcal/mol by the MP2.5 method, which would provide a dependable value. The interaction energies between the HA monomers included in the complex are −1445.5 kcal/mol to −1223.1 kcal/mol in the MP2.5, showing the strong binding. The interaction energies between the Fab monomer and the associated HA monomer for Fab(I)–HA(I) and Fab(II)–HA(II) are −359.9 kcal/mol and −372.0 kcal/mol in the MP2.5, respectively. However, the interactions between the Fab monomer and the other HA monomer, and those between the Fab domains were found to be repulsive. These results suggest that the Fab monomers interact attractively only with the bonding HA monomers. It is then noted that, because of the disulfide bridge at the Fab connection site (not seen explicitly in the structure), there is some affinity instead of steric hindrance in this structure in spite of the repulsive interaction [7].

Table 1
IFIE results (in units of kcal/mol) for inter–domain interactions in the complex of HA-trimer (I, II and III) and Fab-dimer (I and II). The calculated values by the HF, MP2, MP3 and MP2.5 methods with the basis set of 6-31G are shown.

Inter–domain	HF	MP2	MP3	MP2.5
Fab dimer–HA trimer	38.1	−163.5	−127.0	−145.3
Fab(I)–HA(I)	−288.8	−367.0	−352.8	−359.9
Fab(I)–HA(II)	177.5	155.5	144.5	150.0
Fab(I)–HA(III)	134.3	134.3	134.3	134.3
Fab(II)–HA(I)	137.0	137.0	137.0	137.0
Fab(II)–HA(II)	−292.7	−380.4	−363.7	−372.0
Fab(II)–HA(III)	170.8	157.0	159.5	158.2
HA(I)–HA(II)	−1022.4	−1280.4	−1237.1	−1258.7
HA(II)–HA(III)	−981.7	−1245.7	−1200.6	−1223.1
HA(I)–HA(III)	−1189.0	−1469.7	−1421.3	−1445.5
Fab(I)–Fab(II)	210.8	197.7	199.5	198.6
Fab dimer–HA(I)	−151.8	−230.0	−215.8	−222.9
Fab dimer–HA(II)	−115.3	−224.9	−205.0	−214.9
Fab dimer–HA(III)	305.1	291.3	293.8	292.6

IFIE value (kcal/mol).

Table 2
Interaction energies (in units of kcal/mol) between residues calculated with FMO-MP2.5/6-31G method and the donor–acceptor distance (Å) of hydrogen bonds in epitopes. The signatures (A)–(J) are associated with those in Fig. 2. As an exceptional case in the calculation, LEU164 is listed instead of VAL163 in (J) because of the fragmentation at C_α employed in the FMO calculation.

Fab	HA (kcal/mol)	Distance (Å)	Fab	HA (kcal/mol)	Distance (Å)
(A)	ASN137		(B)	ASN137	
ASP104	−22.7	1.39	ASP104	−34.54	1.54
			SER32	−8.70	1.98
	TRP153			TRP153	
ASP104	−4.07		ASP104	−1.86	
(C)	GLY158		(D)	GLY158	
PRO60	−0.42		GLY58	0.88	
ALA61	−0.54		VAL59	0.56	
	SER159			SER159	
PRO60	1.36		GLY58	−5.70	1.86
ALA61	−0.43		VAL59	−11.66	1.80
(E)	SER186		(F)	SER186	
TYR100	−8.27	1.80	TYR100	−5.16	2.51
	THR187			THR187	
TYR100	−2.48		TYR100	−3.02	
	GLN189			GLN189	
TYR109	−9.71	1.52	TYR109	−11.12	1.76
	THR192			THR192	
SER57	−11.11	1.76	SER57	0.73	
	SER193			SER193	
SER57	−0.35		SER57	−15.31	1.60
(G)	THR126		(H)	THR126	
SER31	−1.26		SER31	−0.75	
	THR128			THR128	
THR74	0.14		SER28	−2.89	2.96
SER75	−12.70	1.76			
(I)	PR0162		(J)	PR0162	
GLY26	0.23		SER28	−2.27	
	VAL163			LEU 164	
GLY26	0.63		SER31	−11.44	1.54
TYR27	−11.06	2.03	GLY32	−2.08	
	ASN165			ASN165	
SER31	−0.60		SER31	−6.47	
GLY32	−12.55	1.64	GLY32	−12.05	1.52

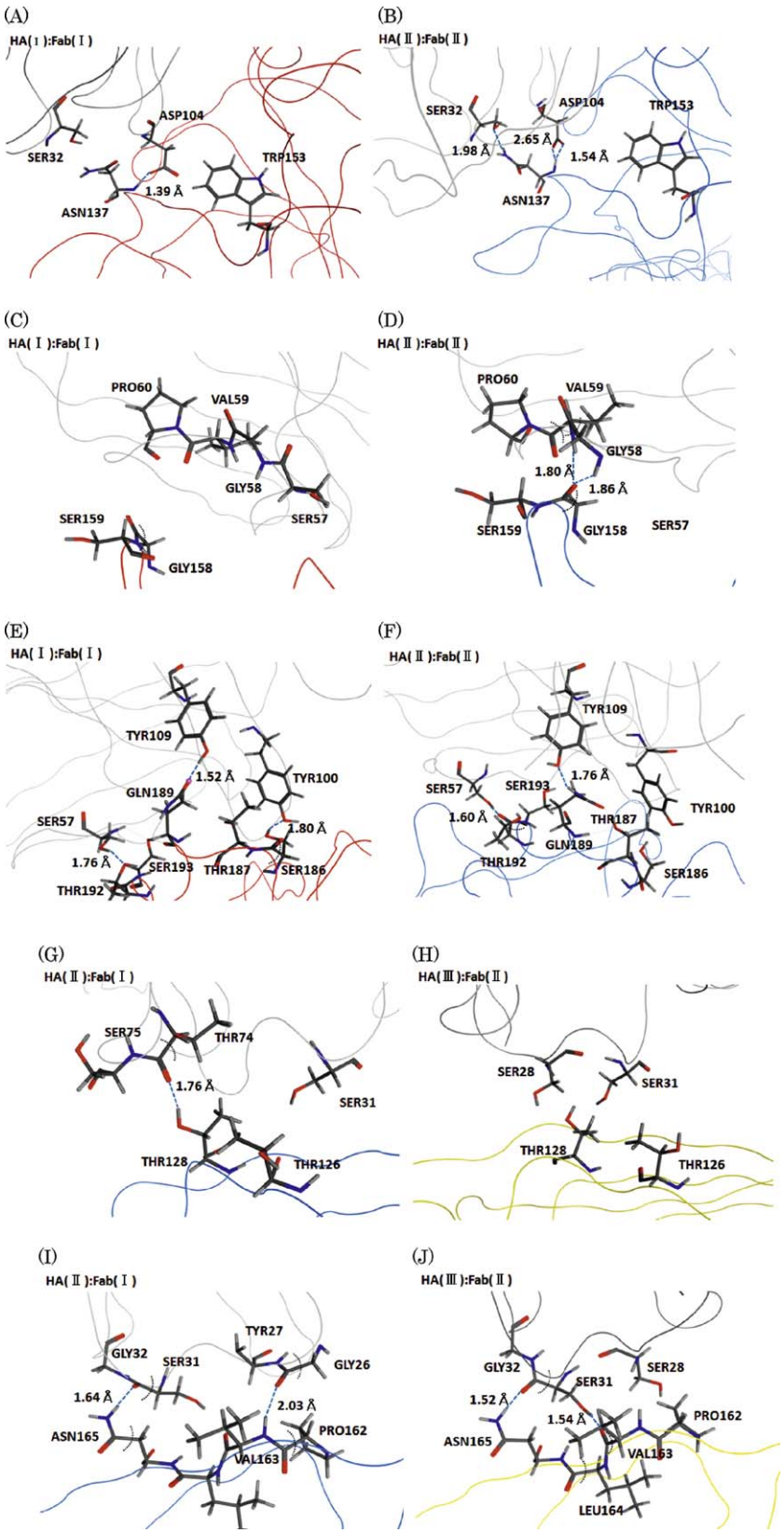

Fig. 2. Details of hydrogen bonds between the Fab and the epitopes in HA. (A)–(F) and (G)–(J) refer to direct-linked and cross-linked domains, respectively; (A), (C), (E) and (B), (D), (F) refer to the interactions between HA(I) and Fab(I) or HA(II) and Fab(II), respectively; Fab in gray, HA(I) in red, HA(II) in blue and HA(III) in yellow. Hydrogen bonds are represented by dotted lines with the distance between donor and acceptor atoms.

Table 3
The hemadsorption activity (p: positive/n: negative; – means no experimental data) and IFIE results (kcal/mol) for the Fab dimer calculated with FMO-MP2.5/6-31G method at the residue sites in antigenic regions A and B of HA. Trimer is the complex of HA(I)–(III), and the mutation years and the types of amino acid after mutation are also listed. The meanings of attached symbols for p sites are as follows. *: Attractive and already mutated site, **: attractive and yet to-be-mutated site, †: repulsive but already mutated site, ††: repulsive and not mutated site.

Region A	p/n	HA(I)	HA(II)	HA(III)	Trimer		Mutation year	Mutation
PHE120	n	0.05	0.39	−0.19	0.24			
ILE121	p	−0.14	−0.76	−0.01	−0.91	*	1995	I-N
THR122	p	0.43	1.24	0.31	1.98	†	1971	T-N
GLU123	–	−46.44	−62.48	−62.62	−171.54			
GLY124	p	0.15	0.24	−0.21	0.19	†	1986, 1996	G-D-S
PHE125	p	−0.72	−1.29	−0.51	−2.52	**		
THR126	p	0.70	−2.74	−0.89	−2.93	*	1973	T-N
TRP127	n	−0.87	−1.15	0.82	−1.20			
THR128	p	0.46	−15.66	−10.86	−26.05	**		
GLY129	n	−0.49	−2.91	−0.39	−3.80			
VAL130	n	0.23	−1.03	0.74	−0.06			
THR131	p	−1.92	−4.16	−1.27	−7.35	*	1987, 2001	T-A-T
GLN132	–	3.02	6.54	4.49	14.05			
ASN133	p	0.86	0.72	1.24	2.81	†	1977, 1991	N-S-D
GLY134	n	−0.13	0.89	−0.20	0.55			
GLY135	p	1.49	−0.03	1.35	2.81	†	1993	G-T
SER136	n	−15.88	−22.56	−0.13	−38.57			
ASN137	p	−24.48	−39.49	−0.54	−64.51	*	1976, 1997	N-Y-S
ALA138	–	−30.89	−31.00	0.89	−61.00			
CYS139	n	−23.09	−19.69	0.11	−42.67			
LYS140	n	21.08	29.32	55.39	105.79		2006	G-I
ARG141	–	33.60	52.72	59.95	146.28			
GLY142	p	−1.65	−3.12	−0.01	−4.78	*	1995	G-R
PR0143	p	−0.43	1.67	0.28	1.52	†	1977	P-S
GLY144	p	−4.71	−3.79	−0.48	−8.97	*	1970, 1982, 1985, 1995	G-D-N-V-R
SER145	p	1.52	−0.65	0.51	1.38	†	1975, 1993, 2004	S-N-K-N
GLY146	p	−6.34	−7.00	−0.48	−13.82	*	1977	G-S
PHE147	p	9.40	7.89	−0.34	16.94	††		
PHE148	n	−3.34	−1.86	0.36	−4.83			
SER149	–	−0.97	−1.42	−1.25	−3.64			
ARG150	–	40.23	55.59	56.94	152.76			
LEU151	n	−0.97	−0.64	−0.46	−2.07			
ASN152	n	3.06	1.05	−0.80	3.31			
TRP153	n	−3.57	−0.17	0.89	−2.84			

Region B	p/n	HA(I)	HA(II)	HA(III)	Trimer		Mutation year	Mutation
LEU154	–	−0.18	−1.72	−1.18	−3.09			
THR155	p	−1.78	−0.29	−0.55	−2.62	*	1971, 1986, 2001	T-Y-H-T
LYS156	p	−24.52	−33.47	82.77	24.78	†	1992, 1995, 2001	E-K-Q-H
SER157	p	0.97	1.02	1.15	3.14	†	1993	S-L
GLY158	p	−1.68	−9.35	−1.47	−12.50	*	1976, 1995	G-E-K
SER159	p	−6.54	−28.09	2.12	−32.52	*	1985, 2002	S-Y-F
THR160	p	3.90	−3.96	0.32	0.25	†	1977	T-K
TYR161	–	1.89	6.83	3.24	11.96			
PR0162	–	−2.61	−3.64	−2.38	−8.64			
VAL163	p	0.93	−4.04	−0.13	−3.24	*	1983	V-A
LEU164	–	0.01	−11.25	−16.43	−27.66			
ASN165	p	−1.16	−2.44	−8.93	−12.53	**		
VAL166	–	0.05	−8.92	−2.13	−11.00			
THR167	–	−0.04	5.16	2.79	7.91			
MET168	–	0.05	−3.88	−3.47	−7.30			
PR0169	–	−0.03	2.79	2.24	5.00			
ASN170	n	0.46	−2.03	−1.61	−3.18			
ASN171	–	−0.29	2.19	1.56	3.46			
ASP172	p	−45.49	−60.16	−59.56	−165.21	*	1976, 1993, 1998	D-G-D-E
ASN173	p	−0.24	0.27	0.04	0.07	†	1982	N-K
PHE174	p	0.32	0.75	0.78	1.85	††		
ASP175	–	−46.03	−49.56	−51.96	−147.55			
LYS176	–	46.24	58.73	59.82	164.80			
LEU177	–	−0.21	1.15	0.72	1.67			
TYR178	n	0.02	−0.76	−0.63	−1.37			
ILE179	n	−0.23	1.01	0.89	1.67			
TRP180	–	0.30	−1.31	−1.86	−2.87			
GLY181	n	0.23	1.07	1.24	2.54			
ILE182	p	−1.04	−1.24	−1.65	−3.94	*	1968	I-V
HIS183	n	2.45	35.21	64.98	102.64			
HIS184	–	33.53	43.09	55.72	132.34			
PR0185	n	6.73	4.00	0.58	11.31			
SER186	–	−15.94	−11.42	0.45	−26.91		1999, 2009	S-G
THR187	–	−12.59	−6.99	−0.02	−19.61			
ASN188	p	5.92	0.53	−1.49	4.96	†	1970	N-D
GLN189	p	−20.65	−28.95	−1.02	−50.61	*	1973, 1987, 1992, 2002	Q-K-R-S-N

Table 3 (*Continued*)

Region B	p/n	HA(I)	HA(II)	HA(III)	Trimer		Mutation year	Mutation
GLU190	–	−11.47	−25.83	−64.32	−101.62		1991	E-D
GLN191	n	5.85	3.99	0.86	10.70			
THR192	p	−12.54	−0.15	−1.35	−14.03	*	1998	T-I
SER193	p	−13.02	−33.20	−3.15	−49.37	*	1972, 1989, 2004	S-N-S-F
LEU194	–	11.86	13.37	−2.04	23.19			
TYR195	–	−0.48	−2.94	−0.50	−3.93			
VAL196	p	−3.64	−6.97	−2.05	−12.66	**		
GLN197	p	−3.37	0.21	1.97	−1.18	*	1977, 1993	Q-R-Q
ALA198	p	1.47	3.85	2.70	8.02	††		
SER199	–	0.67	2.44	1.72	4.82			
GLY200	–	−0.12	−1.41	−0.96	−2.49			
ARG201	–	55.63	55.57	64.96	176.16			
VAL202	n	0.25	−1.09	−1.57	−2.42		2001	V-I
THR203	–	−0.10	1.56	1.55	3.01			
VAL204	n	0.18	−1.48	−1.02	−2.33			

It is known that the MP2 method significantly overestimates the stabilization energy, while the MP3 method underestimates it. To include the higher-order perturbation effects needs the considerable computational costs even by one order. Therefore, we rely on the MP2.5 method [20,21] in the following analysis to represent the interaction energies of amino acid residues, which would provide a quantitative measure.

3.2. Fluctuations in monoclonal antibody recognition

In the analysis of inter–domain interaction energies, we have found that the effect of cross-linked interaction energies should be assessed as well. To gain the information on the molecular recognition by hydrogen bonds in epitopes, we investigate the structures (Fig. 2) and the interaction energy values (Table 2) of the HA antigen–antibody complex. (The signatures (A)–(J) correspond with each other between Fig. 2 and Table 2.)

As seen in Fig. 2(A) and (B) with Table 2(A) and (B), the residue ASP104 in the Fab(I) is bonded by the oxygen and hydrogen atoms to the residue ASN137 in HA(I) with the distance of 1.39 Å, where the calculation result for IFIE is −22.70 kcal/mol. (Hereafter, the distance represents that between hydrogen-bonded donor and acceptor atoms.) On the other hand, the residues ASP104 and SER32 in the Fab(II) are bonded to the residue ASN137 in HA(II) with 1.54 Å and 1.98 Å, respectively; the calculation results for the corresponding IFIEs are −34.54 kcal/mol and −8.70 kcal/mol, respectively. As observed in Fig. 2(C) and (D) with Table 2(C) and (D), the residue ALA61 in Fab(I) interacts attractively with GLY158 and SER159 in HA(I), while both the residues VAL59 and GLY58 in Fab(II) interact with GLY158 and SER159 in HA(II). We also see in Fig. 2(E) and (F) with Table 2(E) and (F) that the interactions between THR192 and SER57 are attractive and repulsive in HA(I)–Fab(I) and HA(II)–Fab(II) complexes, respectively. As an exceptional case in the calculation, as seen in Fig. 2(J) and Table 2(J), LEU164 instead of VAL163 interacts strongly with the Fab by the carbonyl oxygen because of the fragmentation made at C_α, which is employed in the standard FMO recipe [18,19]. In this way, as observed in Fig. 2 and Table 2, the energy values are calculated to be about −5 to −15 kcal/mol with forming the hydrogen bonds shorter than 2 Å, while other electrostatic interaction energy values like the interaction with ASP are larger in magnitude than them. Although the monoclonal antibodies recognize the identical region in HA, the Fab monomers interact with HA trimer with fluctuations. In addition, it is remarked that the present analysis is based on a stable snapshot structure obtained in terms of X-ray crystallographic experiment. Thermal effects at physiological temperatures may thus cause considerable fluctuations in molecular structures and associated interactions of complexes. We should take into account these circumstances in specifying the important amino-acid residues in the HA antigen.

3.3. Hypothetical scheme for predicting the mutations in HA

For the probable mutations of amino acid residues in HA, the following two conditions should be satisfied [10]: that is, the mutant HA should preserve its viral function and also be able to escape the antibody pressure. The former condition is associated with the experimental work carried out by Nakajima et al. [11,12], in which they have extensively introduced single-point mutations in HA and measured the hemadsorption activity of the mutants to assess whether the mutated sites are allowed (positive) or prohibited (negative). The latter condition is associated with the theoretical work in which attractive or repulsive interaction energies with the Fab dimer are evaluated in terms of the values of IFIE sum of the residues in the HA antigenic regions A and B (Table 3). Our hypothesis [10] is that the residues satisfying these two conditions above (i.e., allowed site and attractive interaction) will have a high probability of mutation, which will be examined below through comparison with the historical facts concerning the actual mutations in HA.

Table 3 shows the data regarding the positive(p)/negative(n) sites and the interaction energies with the Fab dimer at the MP2.5/6-31G level for all the amino acid residues in the HA antigenic regions A and B. At first, there are 21 residues of allowed and attractive sites which may be predicted to lead to mutations in our scheme. We see that 17 residues (represented by the symbol *) of them have already been mutated. The other four residues (represented by the symbol **) may be expected to be mutated in future. Next, there are 14 residues at allowed and repulsive sites. Three residues (represented by the symbol ††) of them have not mutated corresponding to our criterion for mutation. However, there are eleven residues (represented by the symbol †) seemingly against our prediction, which are located at the repulsive sites but have experienced the mutations. We will discuss and explain these cases below, where we will categorize them into the charged, polar and hydrophobic residues for quantitative characterization of amino acid residues. In addition, there are a few exceptions concerning the criterion above by the hemadsorption experiment; LYS140 and VAL 202 experienced the mutations in spite of prohibited (n) sites, which may be explained in terms of the differences of mutations between those introduced in the hemadsorption experiment [11,12] and observed actually. Fig. 3 shows the IFIE sums between Fab dimer and each HA monomer in antigenic regions A and B. It is noted that the red bars represent the sites that are allowed (p: positive) and show an attractive interaction

with Fab dimer, which will have a high probability of forthcoming mutation.

As shown in Fig. 3(A), the charged residues LYS156 and ASP172 have possibility to mutate because of positive (allowed) sites and attractive interactions with Fab dimer. In fact, these sites have mutated many times until now. Interaction of LYS156 in HA(III) is very repulsive against Fab dimer because of its non-bonding to Fab dimer with positive charge. However, it shows the large attractive interaction values in HA(I) and (II) so that LYS156 experiences the antibody pressure by the Fab monomers. It is also noted that K156E is favorable for receptor binding and was actually selected under the pressure of antibodies [29]. Interaction energies of charged residues were quantitatively too large due to the neglect of screening effect in the present FMO calculation in vacuo,

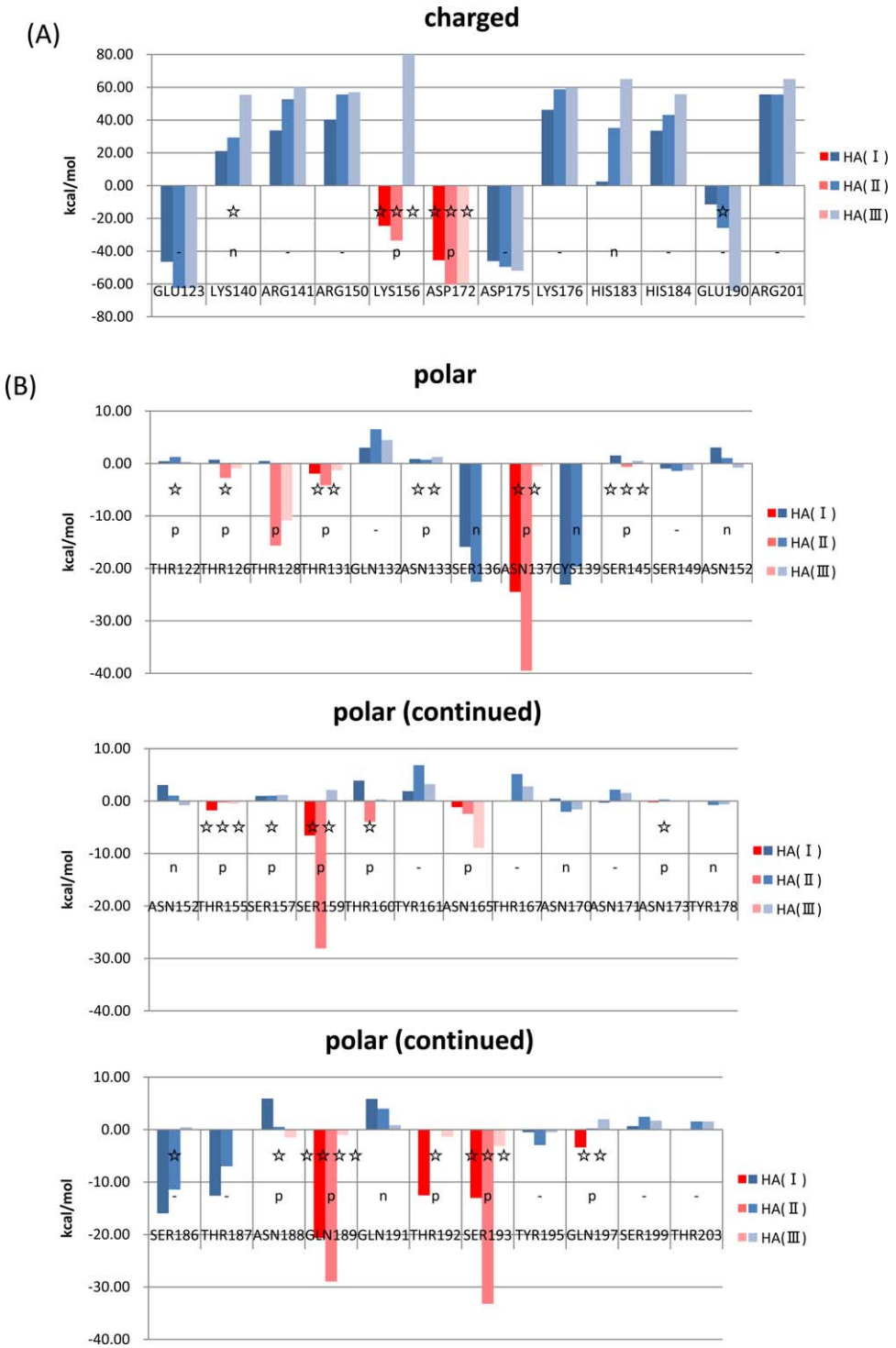

Fig. 3. IFIE sums between Fabs and the residues in the antigenic regions A and B calculated with FMO-MP2.5/6-31G method: (A) charged residues; (B) polar residues; (C) hydrophobic residues. The red bars represent the sites that are allowed (p: positive) and show an attractive interaction with Fab dimer, which will have a high probability of forthcoming mutation; other cases are depicted by the blue bars. The number of stars represents the times of mutations observed already. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

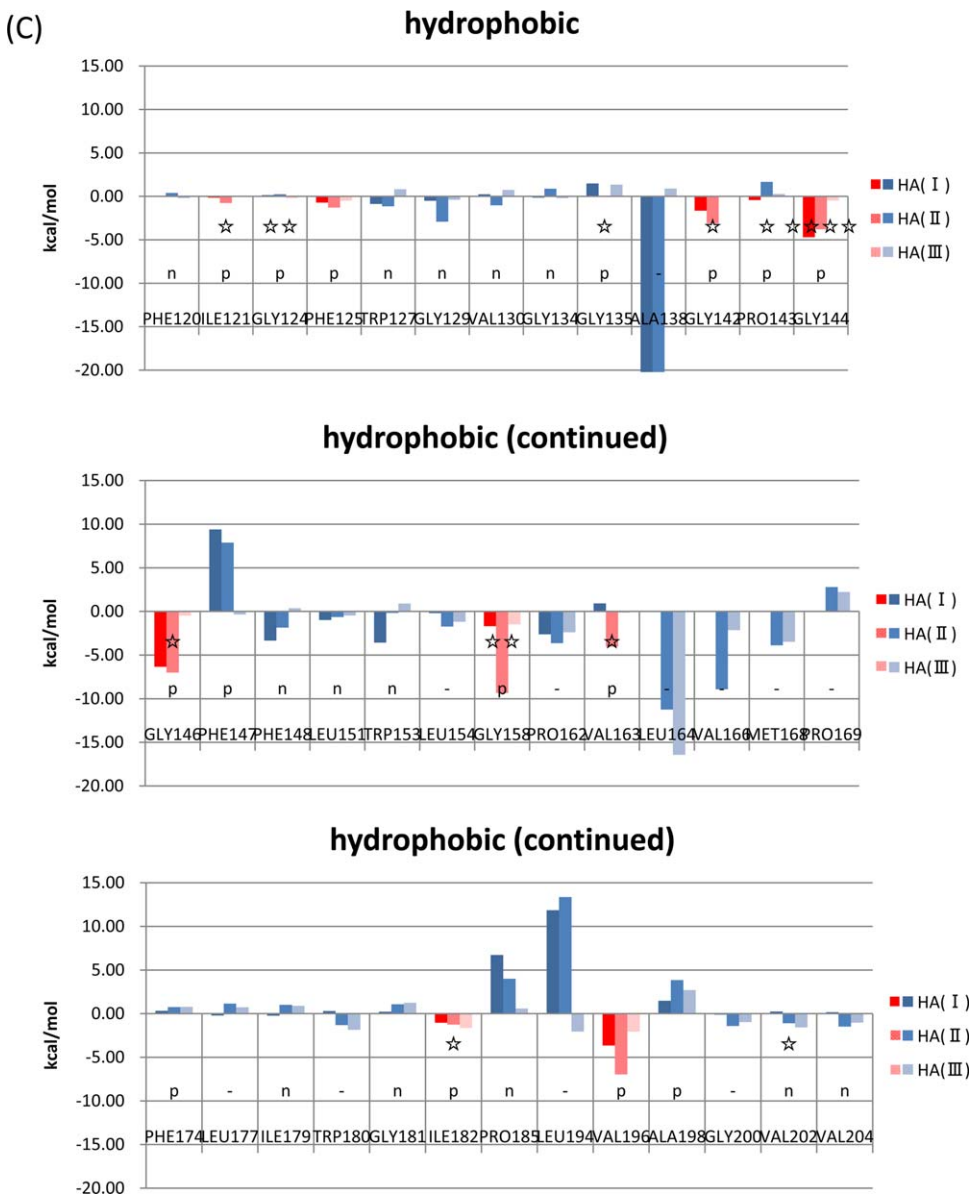

Fig. 3. (*Continued*).

while the comparative features (relative importance and attractive or repulsive interaction) of individual residues are supposed to be described relevantly.

As seen in Fig. 3(B), there are 18 polar residues located at positive (allowed) sites. The eleven sites (126, 128, 131, 137, 155, 159, 165, 173, 189, 192 and 193) are under the antibody pressure apparently, 10 of which have already mutated. The residue 128 may be expected to be mutated in future. Although seven residues (122, 133, 145, 157, 160, 173 and 188) show repulsive interactions with Fab dimer, they have been mutated. The mutations at the 122 and 133 sites have obtained the ability of the oligosaccharide attachment by mutations [30,31]. The residue 133 was actually recognized by monoclonal antibodies until recently [32]. The substitutions N145K and N188D have caused structural changes which allow for the escape from the neutralization by antibodies [33,34]. The substitution S157L allowed for the escape from the antibody HC19 [15]. THR160 receives opposite interactions from multiple Fab monomers. In this case, there are some bonding fluctuations concerning the HA(I) and HA(II) interactions with Fab monomers. We thus suggest that THR160 is under the pressure of an antibody. Because of very small interaction with Fab dimer, the residue ASN173 could not be treated by our prediction scheme.

As observed in Fig. 3(C), the hydrophobic residues show smaller interaction energies than the charged and polar residues. In the present electron-correlated FMO calculations, we can consider the dispersion interactions quantitatively. There are 15 hydrophobic residues located at the allowed (positive) sites. Nine residues (121, 125, 142, 144, 146, 158, 163, 182 and 196) are under the antibody pressure apparently, and seven of them have already been mutated. The residues PHE125 and VAL196 may be expected to be mutated in future. Although three residues (124, 135 and 143) show repulsive interactions with Fab dimer, they have been mutated. The substitution G135R enhances the attractive interaction with glycoprotein [29], and G135T enhances the attractive interaction with sialic acid [35]. The remaining residues GLY124 and PRO143 interact with Fab monomers by 0–1 kcal/mol, which are very weak interactions.

Thus, we have obtained satisfactory results in fair agreement with the historical mutation events, as well as in the earlier study [10]. These results above which were calculated in vacuo, however, ignore the screening effects by solvents [36,37]. Especially,

the IFIE values between charged fragments are substantially overestimated. To cope with this difficulty, although it costs much more, it would be desirable to incorporate the solvation effects into the FMO calculation with, e.g., the Poisson–Boltzmann equation [38]. If this method would be combined with the IFIE calculations, we can evaluate the interaction energies more quantitatively, while careful treatments for the protonation state and the counterions would be required. (Standard aqueous solution pK_a values are assumed in the present analysis.) Further issues to be investigated are the contributions associated with entropy and electronic polarization effects [36,37] in the complex. On the other hand, concerning the calculation levels, it has been observed [28,39] that the relative ordering of the binding energies can be obtained even at such low level as HF/STO-3G. In addition, we can further assess the relative importance of energy values by categorizing the charged, polar and hydrophobic residues, while the interaction energies associated with the hydrophobic residues should be addressed by taking into account the electron correlation effects appropriately.

4. Concluding remarks

The influenza vaccines have been developed mainly against target HA proteins, while it has been difficult to predict forthcoming mutations in HA. We note that only the current calculations cannot detect probable mutations, which may also be associated with the maintenance of HA functions. However, in our combination of computational and experimental methods, to overcome this difficulty, we attempted the detection of possible amino-acid mutations to escape from the antibodies with the aid of the measurement of hemadsorption activities in antigenic regions. As a result, our prediction scheme was found to be consistent with the historical facts of mutations by 83%. (We have picked up the 21 residues of allowed and attractive sites, and two more ones, LYS156 and THR160, showing the attractive interactions at HA monomer level, of which 19 sites have actually mutated.) On the other hand, some mutations unexplained in this analysis were found: however, for example, the substitutions N145K and N188D seemed to induce the structural changes between the HAs of antigenic mutant and the wild type, and G135R and K156E enhanced interactions with glycoprotein under the monoclonal antibodies. Almost all exceptional mutations against our prediction could be explained by considering these previous studies in the literature. Thus the challenge will now be focused on the elucidation of the possible functions of glycoproteins to identify these exceptions. In addition, amino acid changes at more than one residue on the same antigenic site are observed [40,41] for antigenic drift, whose elucidation associated with some cooperative effects may be another challenge. Nonetheless, we expect that this novel computational approach addressed in the present work would be useful for influenza vaccine developments as well.

Acknowledgments

This work was partially supported by the Health and Labour Sciences Research Grants on Emerging and Re-emerging Infectious Diseases (to E.N., No. H22-//Shinko-Ippan-006//) from the Ministry of Health, Labour and Welfare of Japan.

References

[1] J.J. Skehel, D.J. Stevens, R.S. Daniels, A.R. Douglas, M. Knossow, I.A. Wilson, D.C. Wiley, A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 1779–1783.
[2] R.S. Daniels, A.R. Douglas, J.J. Skehel, D.C. Wiley, Analyses of the antigenicity of influenza haemagglutinin at the pH optimum for virus-mediated membrane fusion, J. Gen. Virol. 64 (1983) 1657–1662.
[3] D.C. Wiley, I.A. Wilson, Structural identification of the antibody-binding site of Hong Kong influenza haemagglutinin and their involvement antigenic variation, Nature 289 (1981) 373–378.
[4] T. Bizebard, B. Gigant, P. Rigolet, B. Rasmussen, O. Diat, S. Bosecke, S. Wharton, J.J. Skehel, M. Knossow, Structure of influenza virus hemagglutinin complexed with a neutralizing antibody, Nature 376 (1995) 92–94.
[5] D. Fleury, B. Barrere, T. Bizebard, R.S. Daniels, J. Skehel, M. Knossow, A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site, Nat. Struct. Biol. 6 (1999) 530–534.
[6] D. Fleury, R.S. Daniels, J.J. Skehel, M. Knossow, T. Bizebard, Structural evidence for recognition of a single epitope by two distinct antibodies, Struct. Funct. Gen. 40 (2000) 572–578.
[7] C. Barbey-Martin, B. Gigant, T. Bizebard, L.J. Calder, S.A. Wharton, J.J. Skehel, M. Knossow, An antibody that prevents the hemagglutinin low pH fusogenic transition, Virology 294 (2002) 70–74.
[8] W.D. Kundin, Hong Kong A-2 influenza virus infection among swine during a human epidemic in Taiwan, Nature 228 (1970) 857.
[9] C.A. Russell, The global circulation of seasonal influenza A (H3N2) viruses, Science 320 (2008) 340–346.
[10] K. Takematsu, K. Fukuzawa, K. Omagari, S. Nakajima, K. Nakajima, Y. Mochizuki, T. Nakano, H. Watanabe, S. Tanaka, Possibility of mutation prediction of influenza hemagglutinin by combination of hemadsorption experiment and quantum chemical calculation for antibody binding, J. Phys. Chem. B 113 (2009) 4991–4994.
[11] K. Nakajima, E. Nobusawa, K. Tonegawa, S. Nakajima, Restriction of amino acid change in influenza a virus H3HA: comparison of amino acid changes observed in nature and in vitro, J. Virol. 77 (2003) 10088–10098.
[12] K. Nakajima, E. Nobusawa, A. Nagy, S. Nakajima, Accumulation of amino acid substitutions promotes irreversible structural changes in the hemagglutinin of human influenza AH3 virus during evolution, J. Virol. 79 (2005) 6472–6477.
[13] A.J. Caton, G.G. Brownlee, The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (HI Subtype) , Cell 31 (1982) 417–427.
[14] D.C. Wiley, The structure and function of the hemagglutinin membrane glycoprotein of inflenza virus, Ann. Rev. Biochem. 56 (1987) 365–394.
[15] N.G. Wrigley, E.B. Brown, R.S. Daniels, A.R. Douglas, J.J. Skehel, D.C. Wiley, Electron microscopy of influenza hemagglutinin-monoclonal antibody complexes, Virology 131 (1983) 308–314.
[16] BioStation Viewer, Available at http://www.fsis.iis.u-tokyo.ac.jp/en/result/software/.
[17] R.S. Daniels, S. Jeffries, P. Yates, G.C. Schild, G.N. Rogers, J.C. Paulson, S.A. Wharton, A.R. Douglas, J.J. Skehel, D.C. Wiley, The receptor-binding and membrane-fusion properties of influenza virus variants selected using anti-hemagglutinin monoclonal antibodies, EMBO J. 6 (1987) 1459–1465.
[18] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, Fragment molecular orbital method: an approximate computational method for large molecules, Chem. Phys. Lett. 313 (1999) 701–706.
[19] D.G. Fedorov, K. Kitaura, Extending the power of quantum chemistry to large systems with the fragment molecular orbital method, J. Phys. Chem. A 111 (2007) 6904–6914.
[20] Y. Mochizuki, K. Yamashita, K. Fukuzawa, K. Takematsu, H. Watanabe, N. Taguchi, Y. Okiyama, M. Tsuboi, T. Nakano, S. Tanaka, Large-scale FMO-MP3 calculations on the surface protein of influenza virus, hemagglutinin (HA) and neuraminidase (NA), Chem. Phys. Lett. 493 (2010) 346–352.
[21] M. Pitoňák, P. Neogrády, J. Černý, S. Grimme, P. Hobza, Scaled MP3 non-covalent interaction energies agree closely with accurate CCSD(T) benchmark data, Chem. Phys. Chem. 10 (2009) 282–289.
[22] K. Kitaura, T. Sawai, T. Asada, T. Nakano, M. Uebayasi, Pair interaction molecular orbital method: an approximate computational method for molecular interactions, Chem. Phys. Lett. 312 (1999) 319–324.
[23] K. Fukuzawa, Y. Komeiji, Y. Mochizuki, A. Kato, T. Nakano, S. Tanaka, Intra- and intermolecular interaction between cyclic-AMP receptor protein and DNA: ab initio fragment molecular orbital study, J. Comput. Chem. 27 (2006) 948–960.
[24] M. Ito, K. Fukuzawa, Y. Mochizuki, T. Nakano, S. Tanaka, Ab initio fragment molecular orbital study of molecular interactions between liganded retinoid X receptor and its coactivator: roles of helix 12 in the coactivator binding mechanism , J. Phys. Chem. B 111 (2007) 3525–3533.
[25] I. Kurisaki, K. Fukuzawa, Y. Komeiji, Y. Mochizuki, T. Nakano, J. Imada, A. Chmielewski, S.M. Rothstein, H. Watanabe, S. Tanaka, Visualization analysis of inter-fragment interaction energies of CRP-cAMP-DNA complex based on the fragment molecular orbital method, Biophys. Chem. 130 (2007) 1–9.
[26] T. Iwata, K. Fukuzawa, K. Nakajima, S. Aida-Hyugaji, Y. Mochizuki, H. Watanabe, S. Tanaka, Theoretical analysis of binding specificity of influenza viral hemagglutinin to avian and human receptors based on the fragment molecular orbital method, Comput. Biol. Chem. 32 (2008) 198–211.
[27] Y. Mochizuki, K. Yamashita, T. Murase, T. Nakano, K. Fukuzawa, K. Takematsu, H. Watanabe, S. Tanaka, Large scale FMO-MP2 calculations on a massively parallel-vector computer , Chem. Phys. Lett. 457 (2008) 396–403.
[28] K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura, T. Nakano, Molecular interactions between estrogen receptor and its ligand studied by ab initio fragment molecular orbital method , J. Phys. Chem. B 110 (2006) 16102–16110.
[29] P.A. Underwood, J.J. Skehel, D.C. Willy, Receptor-binding characteristic of monoclonal antibody-selected antigenic variants of influenza virus, J. Virol. 61 (1987) 206–208.
[30] M. Knossow, J.J. Skehel, Variation and infectivity neutralization in influenza, Immunology 119 (2006) 1–7.

[31] Y.P. Lin, V. Gregory, M. Bennett, A. Hay, Recent changes among human influenza viruses, Virus Res. 103 (2004) 47–52.
[32] J. Okada, N. Ohshima, R. Kubota-Koketsu, Y. Iba, S. Ota, W. Takase, T. Yoshikawa, T. Ishikawa, Y. Asano, Y. Okuno, Y. Kurosawa, Localization of epitopes recognized by monoclonal antibodies that neutralized the H3N2 influenza viruses in man, J. Gen. Virol. 92 (2011) 326–335.
[33] D.J. Smith, A.S. Lapedes, J.C. de Jong, T.M. Bestebroer, G.F. Rimmelzwaan, A.D.M.E. Osterhaus, R.A.M. Fouchier, Mapping the antigenic and genetic evolution of influenza virus, Science 305 (2004) 371–376.
[34] M. Knossow, R.S. Daniels, A.R. Douglas, J.J. Skehel, D.C. Willey, Three-dimentional structure of an antigenic mutant of the influenza virus haemagglutinin, Nature 311 (1984) 678–680.
[35] A.I. Karasin, M.M. Schutten, L.A. Cooper, C.B. Smith, K. Subbarao, G.A. Anderson, S. Carman, C.W. Olsen, Genetic characterization of H3N2 influenza viruses isolated from pigs in North America, 1977–1999: evidence for wholly human and reassortant virus genotypes, Virus Res. 68 (2000) 71–85.
[36] T. Sawada, D.G. Fedorov, K. Kitaura, Role of the key mutation in the selective binding of avian and human influenza hemagglutinin to sialosides revealed by quantum-mechanical calculations, J. Am. Chem. Soc. 132 (2010) 16862–16872.
[37] T. Sawada, D.G. Fedorov, K. Kitaura, Binding of influenza A virus hemagglutinin to the sialoside receptor is not controlled by the homotropic allosteric effect, J. Phys. Chem. B 114 (2010) 15700–15705.
[38] H. Watanabe, Y. Okiyama, T. Nakano, S. Tanaka, Incorpolation of solvation effects into fragment molecular orbital calculation with the Poisson–Boltzmann equation, Chem. Phys. Lett. 500 (2010) 116–119.
[39] K. Fukuzawa, K. Kitaura, M. Uebayashi, K. Nakata, T. Kaminuma, T. Nakano, Ab initio quantum mechanical study of the binding energies of human estrogen receptor α with its ligands: an application of fragment molecular orbital method, J. Comput. Chem. 26 (2005) 1–10.
[40] I.A. Wilson, N.J. Cox, Structural basis of immune recognition of influenza virus hemagglutinin, Annu. Rev. Immunol. 8 (1990) 737–771.
[41] J.B. Poltkin, J. Dushoff, S.A. Levin, Hemagglutinin sequence clusters and the antigenic evolution of influenza A virus, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 6236–6268.