Analysis of Mutation Mechanisms of Influenza Virus Based on the Fragment Molecular Orbital Method

Project Representative
Shigenori Tanaka Graduate School of System Informatics, Kobe University

Authors
Shigenori Tanaka Graduate School of System Informatics, Kobe University
Yuji Mochizuki Department of Chemistry, Rikkyo University
Kaori Fukuzawa Mizuho Information and Research Institute, Inc.
Katsumi Yamashita NEC Soft, Ltd.
Tatsuya Nakano National Institute of Health Sciences

On the basis of the fragment molecular orbital (FMO) method, we performed the FMO-MP2 and FMO-MP3 electronic-state calculations on the Earth Simulator (ES2) for a protein complex consisting of HA (hemagglutinin) trimer and two Fab fragments and a complex of NA (neuraminidase) and oseltamivir inhibitor of influenza virus. The FMO-MP2/6-31G jobs were completed in 0.8 hours for the HA monomer and in 4.3 hours for the HA trimer with 128 nodes (1024 VPUs) of ES2. Comparison of these timings illuminates the low scaling nature of the present FMO calculations. In addition, the increase rate of computational cost by MP3 compared to MP2 is significantly low. In particular, the FMO-MP3 calculation for the HA trimer, which might be the world's largest target system (36160 atoms) for correlated all-electron calculations to date, was completed in only 5.8 hours with 128 nodes. In the case of NA-oseltamivir, a favorable performance of FMO-MP3 calculation was obtained as well. Thus, a realistic applicability of FMO-MP3 calculations to large-scale proteins has been demonstrated with the ES2 system. Molecular interaction analyses based on the FMO results were then carried out for the prediction of probable mutations in influenza proteins associated with the escape from antibody pressure and the resistance against drugs.

Keywords: FMO (fragment molecular orbital) method, MP (Moeller-Plesset) perturbation theory, influenza virus, HA (hemagglutinin), NA (neuraminidase)

1. Introduction

Influenza is one of the most important infectious diseases of humans. Recent concerns about the avian and swine influenza viruses highlight its threat and the need to understand its evolutionary dynamics. The influenza virus has a remarkable ability to escape host defense mechanisms by altering its binding characters through changes of amino acid residues in the pertinent proteins. This property is referred to as antigenic drift and has been thought to result from the accumulation of a series of amino acid changes in antigenically important regions of proteins. In addition, the viral resistance against some drugs is associated with analogous mutation properties as well. It is thus essential to elucidate the molecular mechanisms by which viruses alter their ligand binding characters in order to find an efficient way to suppress the pandemics and epidemics of influenza.

There are two types of well-known proteins on the surface of influenza virus. One is hemagglutinin (HA) associated with the infection into host cells. Another is neuraminidase (NA) involved in the escape from infected cells. To investigate the interactions between these proteins and their binding partners at the molecular level, we need to resort to some theoretical methods in computational chemistry. Considering the accuracy in molecular simulations, ab initio quantum-chemical approaches would be most dependable for the computational analysis on molecular interactions, whereas these kinds of simulations would demand huge amount of computer resources for biomolecular systems. Here, we employ the fragment molecular orbital (FMO) method [1], which has been developed for efficient and accurate ab initio calculations for biomolecules, for the detailed analysis of molecular interactions in HA and NA systems. In this context, we pay attention to the inclusion of electron correlation effects in terms of Moeller-Plesset (MP) type perturbative treatments in order to appropriately describe the weak interactions such as dispersion forces between hydrophobic residues.

In 2008, we performed [2] the FMO-MP2/6-31G calculation for an antigen-antibody system consisting of the HA monomer

and the Fab fragment (14086 atoms, 921 residues and 78390 AOs), where a total of 4096 vector processors (VPUs) of the Earth Simulator (ES) were utilized to complete the job within an hour. Later, the calculation with the extended 6-31G* basis set (121314 AOs) was carried out on cluster computers, and some specific residues associated with probable mutations were successfully identified through the IFIE (inter-fragment interaction energy) analysis, thus providing a method to predict the forthcoming mutations in HA [3].

Although the second-order MP2 calculations have become feasible even for large proteins in conjunction with the FMO scheme, there has been a potential demand for correlated methods better than MP2. The third-order MP (MP3) theory can be a straightforward option by the perturbative inclusion of electron pair-pair interactions. Here, we employ a parallelized integral-direct implementation of FMO-MP3 scheme in ABINIT-MPX software. The OpenMP shared-memory parallelization is also introduced for the intra-fragment calculations of monomers and dimers at the lower level processing. The Earth Simulator, which was renewed in 2009 as ES2, was used as a massively parallel-vector computational platform, in which some technical points for the vectorization were addressed [4].

Table 1 Timing data for HA monomer, HA trimer and NA systems. The timing shown here is the turn-around job time in hours. Each node with eight VPUs was assigned to the intra-fragment calculations with OpenMP. The 64 node jobs were processed during the usual production run hours, whereas the 128 node jobs were performed under a dedicated usage with special permission. The 6-31G basis set was used throughout, except for the cases of HA monomer and NA with the asterisk (*) meaning the use of the 6-31G* basis set.

(System) Calc. level	Nodes	Time (hour)	Rel.[a]	Acc.[b]	TFLOPS
(HA monomer)					
FMO-MP2	64	1.7			0.97
FMO-MP3	64	2.7	1.6		2.27
FMO-MP2*	64	4.4			1.19
FMO-MP3*	64	8.7	2.0		3.02
FMO-MP2	128	0.8		2.1	2.06
FMO-MP3	128	1.3	1.6	2.1	4.67
(HA trimer)					
FMO-MP2	64	9.4			0.83
FMO-MP3	64	11.9	1.3		1.66
FMO-MP2	128	4.3		2.2	1.83
FMO-MP3	128	5.8	1.3	2.1	3.44
(NA)					
FMO-MP3	64	1.0			3.04
FMO-MP3*	64	4.4			3.09

[a]Cost factor of MP3 job relative to MP2 job.

[b]Acceleration due to the increase of VPUs from 512 to 1024.

2. Results

In the present study, we performed [4] the FMO-MP2 and FMO-MP3 calculations with the 6-31G or 6-31G* basis set for a complex consisting of HA trimer and two Fab fragments (2351 residues and 201276 AOs; PDB-ID: 1KEN) and a complex of NA and oseltamivir ligand (386 residues; PDB-ID: 2HU4) on the ES2. The modeling of the complex structures was performed with the aid of MOE software, in which the addition and structural optimization of hydrogen atoms were carried out. Table 1 compiles the timing data of benchmark calculations of HA and NA systems by using 64 nodes (total 512 VPUs) and 128 nodes (1024 VPUs) of ES2. The FMO-MP2 jobs were processed in 0.8 hours (48.3 minutes) for the HA monomer and in 4.3 hours (260.6 minutes) for the HA trimer with 128 nodes. Comparison of these timings illuminates the low scaling nature of the present FMO calculations. The acceleration from 64 to 128 nodes was slightly over 2, presumably due to the difference in background conditions. Nevertheless, a value close to 2 was expected because of an inherent parallelism of the FMO calculations. It is remarkable that the increase rate of computational cost by MP3 compared to MP2 is quite low. In particular, the FMO-MP3 calculation for the HA trimer, which might be the world's largest target system (36160 atoms) for correlated all-electron calculations to date, was completed in only 5.8 hours with 128 nodes. In the case of NA-oseltamivir, a favorable performance of FMO-MP3 calculation was obtained as well. As a whole, a realistic applicability of FMO-MP3 calculations to large-scale proteins has just been demonstrated with the ES2 system, while further improvements in the ABINIT-MPX code would still be required for better efficiency.

Figure 1 (a) shows the structure of HA trimer complexed with two Fab fragments. From the top view (b) of Fig. 1, it is observed that a hollow structure is formed by the bundled monomers (labeled with roman numbers and colors) and

Table 2 IFIE results (in units of kcal/mol) for pairs of monomeric domains in HA trimer complex obtained by the Hartree-Fock (HF), MP2 and MP3 level FMO calculations with the 6-31G basis set.

	HF	MP2	MP3
HA(I)-HA(II)	-1022.4	-1280.3	-1237.1
HA(II)-HA(III)	-981.7	-1245.7	-1200.7
HA(I)-HA(III)	-1189.0	-1469.9	-1421.3
Fab(I)-HA(I)	-288.8	-367.0	-352.8
Fab(I)-HA(II)	177.5	155.6	158.7
Fab(I)-HA(III)	134.3	134.2	134.3
Fab(II)-HA(I)	137.0	137.0	137.0
Fab(II)-HA(II)	-292.7	-380.5	-363.7
Fab(II)-HA(III)	170.8	157.0	159.5
Fab(I)-Fab(II)	210.8	197.8	199.6
HA(sum)-Fab(sum)	38.1	-163.6	-127.0

also that the Fab fragment is situated in contact with plural monomers. Figure 2 then illustrates the results of IFIEs calculated at the FMO-MP3/6-31G level for the complex consisting of the HA trimer and two Fab fragments. The interactions between the yellow domain and each colored residue are depicted in the figures, where the red and blue correspond to the attractive and repulsive interactions, respectively. On the basis of the evaluated interactions with the Fab fragment antibody, as shown in Fig. 2 (a) and (b), it would be possible [3] to enumerate those residues (identified with deep-red) in HA that have a high probability of forthcoming mutations to escape from antibody pressure. This information about the probable mutations in HA would, in turn, facilitate the development of effective vaccines against influenza viruses. Figure 2 (c) represents the IFIEs between the HA monomer (II) and other amino acid residues in the complex. Such analyses should be useful for the comprehensive understanding of the specific roles played by various domains in the complex. For instance, we can discuss the inter-domain interactions by partially summing up the IFIE values. Table 2 then shows the interaction energies between various monomeric domains (see Fig. 1). Interestingly, there are clear differences among the Hartree-Fock (HF), MP2 and MP3 values for the pairs of attractively contacted domains, e.g. HA(I)−HA(II), being in contrast to the cases of repulsive pair. Pitonak et al. [5] pointed out that the MP3 correction could reduce the tendency of overstabilization at the MP2 level, and the numerical results in Table 2 are consistent with this indication. It is noteworthy that the inter-domain interactions break the symmetries and also that the attractive and repulsive interactions are balanced in a delicate way in the antigen-antibody complex. More detailed analyses concerning the biochemical functions based on interaction patterns have to be carried out for the HA trimer system.

The IFIE analysis between each residue of NA and the oseltamivir inhibitor was also performed [4] with the FMO-MP3/6-31G* calculation. The results are visualized in Fig. 3. The energetically stabilized residues would play key roles in recognizing oseltamivir, and therefore mutations at these residues could reduce the binding affinity. Such mutations have potential to yield the emergence of oseltamivir-resistant viruses. In contrast, the rational drug design to alter repulsive (blue) interactions to attractive (red) interactions could make up better inhibitors with high binding affinities, which would be more effective against the drug-resistant viruses. Figure 4 illustrates a set of attractive interactions between oseltamivir and surrounding residues in the pharmacophore. These stablizations

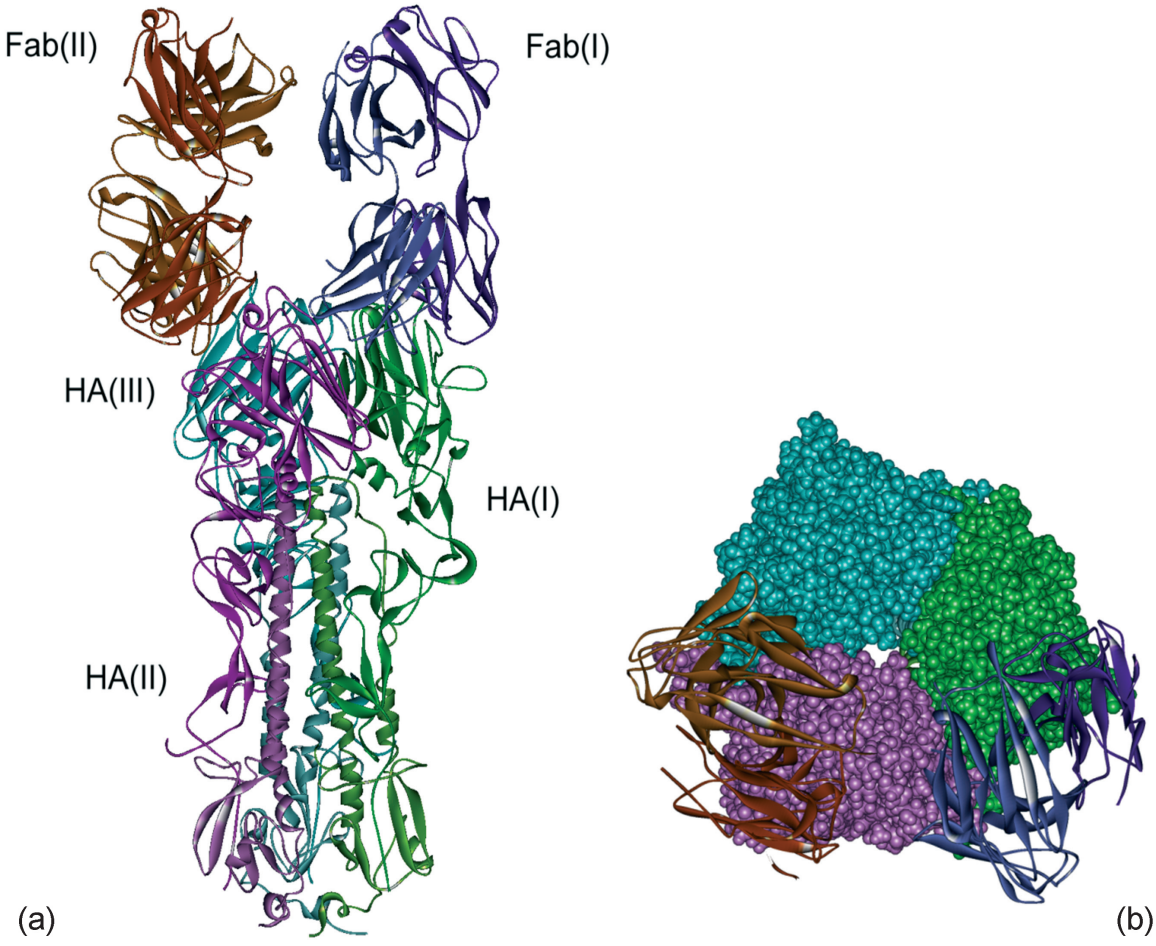
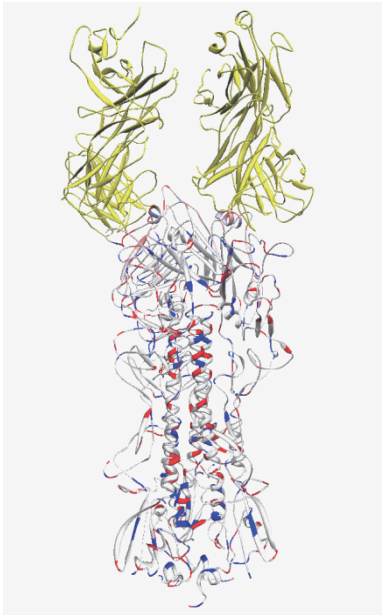
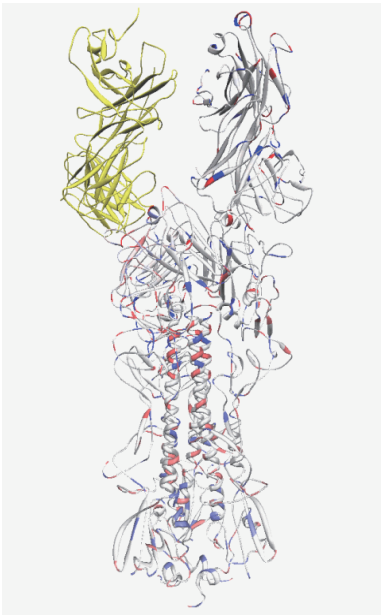


Fig. 1 Graphic representations of the influenza HA trimer with two Fab fragments: (a) side view, (b) top view. Each monomeric domain is identified with roman numbers and colors.

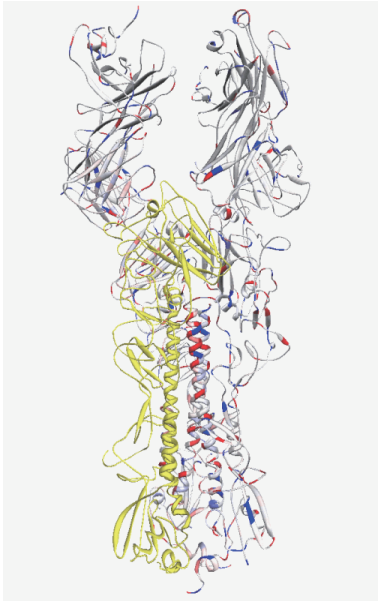
could be characterized by the electrostatic and hydrogen-bond interactions with Glu119, Asp151, Arg152, Arg292, Arg371 and Tyr347. These theoretical analyses could be helpful in revealing the oseltamivir-resistant mechanism and also in developing the effective inhibitors.



(a)



(b)



(c)

Fig. 2 Visualized IFIE results calculated at the FMO-MP3/6-31G level for the influenza HA trimer with two Fab fragments, by gathering all residues in yellow domains: (a) Fab (I)-fragment and Fab (II)-fragment, (b) Fab (II)-fragment, (c) HA (II)-monomer. Red and blue for residues refer to the interaction energies of stabilization (negative) and destabilization (positive), respectively.

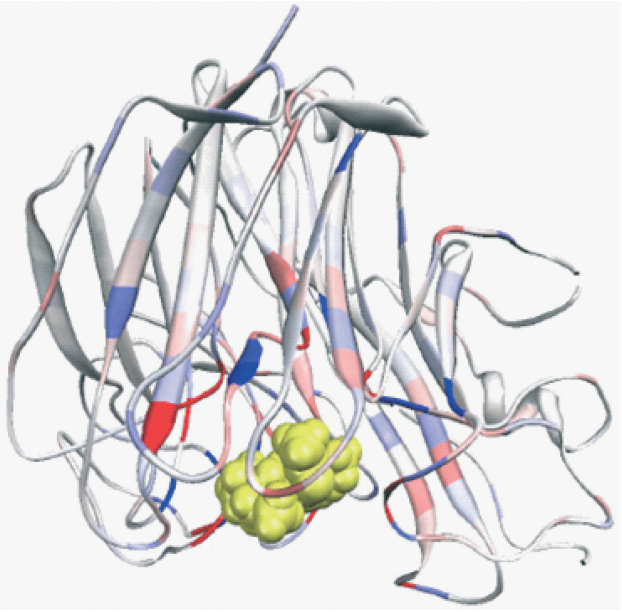


Fig. 3 Visualized IFIE results calculated at the FMO-MP3/6-31G* level for the influenza NA monomer with oseltamivir (yellow). Red and blue for residues refer to the interaction energies of stabilization (negative) and destabilization (positive), respectively.

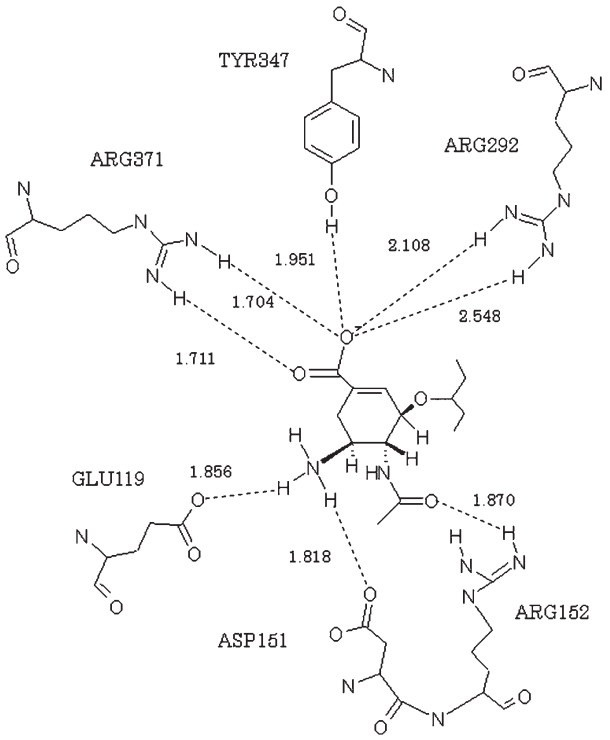


Fig. 4 Hydrogen bond network surrounding oseltamivir in pharmacophore of the influenza NA monomer.

Acknowledgement

This work was supported by the CREST project of the Japan Science and Technology Agency (JST) and by the CISS project at the Institute of Industrial Science (IIS) of the University of Tokyo.

References

[1] D.G. Fedorov and K. Kitaura, ed., *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems,* CRC Press, Boca Raton, 2009.

[2] Y. Mochizuki, K. Yamashita, T. Murase, T. Nakano, K. Fukuzawa, K. Takematsu, H. Watanabe, and S. Tanaka, Chem. Phys. Lett. 457, pp.396-403, 2008.

[3] K. Takematsu, K. Fukuzawa, K. Omagari, S. Nakajima, K. Nakajima, Y. Mochizuki, T. Nakano, H. Watanabe, and S. Tanaka, J. Phys. Chem. B 113, pp.4991-4994, 2009.

[4] Y. Mochizuki, K. Yamashita, K. Fukuzawa, K. Takematsu, H. Watanabe, N. Taguchi, Y. Okiyama, M. Tsuboi, T. Nakano, and S. Tanaka, Chem. Phys. Lett. 493, pp.346-352, 2010.

[5] M. Pitonak, P. Neogrady, J. Cerny, S. Grimme, and P. Hobza, Chem. Phys. Chem. 10, pp.282-289, 2009.

フラグメント分子軌道法によるインフルエンザウイルスの変異機構の解析

プロジェクト責任者
田中　成典　　神戸大学　大学院システム情報学研究科
著者
田中　成典　　神戸大学　大学院システム情報学研究科
望月　祐志　　立教大学　理学部化学科
福澤　　薫　　みずほ情報総研　サイエンスソリューション部
山下　勝美　　NEC ソフト　ニューソリューション事業部
中野　達也　　国立医薬品食品衛生研究所　医薬安全科学部

フラグメント分子軌道（Fragment Molecular Orbital; FMO）法に基づき、インフルエンザウイルスの表面タンパク質に対する FMO-MP2 および FMO-MP3 計算を地球シミュレータ（ES2）を用いて行った。計算の対象としたのは、ヘマグルチニン（HA）三量体と Fab 抗体二量体の複合体（2351 残基、36160 原子）ならびにノイラミニダーゼ（NA）とタミフル分子の複合体（386 残基）である。HA 単量体および三量体系に対して、電子相関を 2 次の Moeller-Plesset 摂動法で考慮した FMO-MP2/6-31G 計算は、ES2 の 128 ノード（1024 VPUs）を用いて、それぞれ 0.8 時間および 4.3 時間で完了した。このことは、本 FMO 計算がサイズ増加に対して良好なスケーリング関係を有することを示している。また、MP2 計算に対する 3 次の MP3 計算の相対的なコスト増加も極めて低く、例えば HA 三量体系に対する FMO-MP3/6-31G 計算は ES2 の 128 ノードを用いてわずか 5.8 時間で終了した。NA - タミフル系に対しても、同様に FMO-MP3 計算を 6-31G および 6-31G* 基底関数を用いて実行し、それぞれ 1.0 および 4.4 時間で完了した。このように、これらの計算を通じて、大規模タンパク質系に対する FMO-MP3 計算が ES2 上で効率的に実行可能であることが示された。また、これらの FMO 計算の結果を用いて系に含まれるフラグメント間の相互作用解析を行い、HA タンパク質や NA タンパク質内のアミノ酸が抗体圧や薬剤による阻害から逃れるためにどのように変異を起こすのかのメカニズムを理論的に考察した。これらの分析手法はインフルエンザウイルスの将来の変異予測や薬剤開発にも役立てることができる。

キーワード: フラグメント分子軌道法, メラー・プレセット摂動法, インフルエンザウイルス, ヘマグルチニン, ノイラミニダーゼ