

## Development of an automated fragment molecular orbital (FMO) calculation protocol toward construction of quantum mechanical calculation database for large biomolecules

Chiduru Watanabe<sup>1,2\*</sup>, Hirofumi Watanabe<sup>3</sup>, Yoshio Okiyama<sup>1,4</sup>, Daisuke Takaya<sup>1</sup>,  
Kaori Fukuzawa<sup>1,5</sup>, Shigenori Tanaka<sup>6</sup>, Teruki Honma<sup>1\*\*</sup>

<sup>1</sup>Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

<sup>2</sup>JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

<sup>3</sup>Education Center on Computational Science and Engineering, Kobe University 7-1-48, Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

<sup>4</sup>National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan

<sup>5</sup>Department of Physical Chemistry, School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa, Tokyo 142-8501, Japan

<sup>6</sup>Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe 657-8501, Japan

E-mail: \*chiduru.watanabe@riken.jp, \*\*E-mail: honma.teruki@riken.jp

(Received ; November 13, 2018; accepted February 18,2019; published online March 23,2019)

### Abstract

We developed an automated FMO calculation protocol (Auto-FMO protocol) to calculate huge numbers of protein and ligand complexes, such as drug discovery targets, by an *ab initio* FMO method. The protocol performs not only FMO calculations but also pre-processing of input structures by homology modeling of missing atoms and subsequent MM-based optimization, as well as post-processing of calculation results. In addition, QM/MM optimization of complex structures, conformational searches of ligand structures in solvent, and MM-PBSA/GBSA calculations can be optionally carried out. In this paper, FMO calculations for 149 X-ray complex structures of estrogen receptor  $\alpha$  and p38 MAP kinase were performed at the K computer and in-house PC cluster server by using the Auto-FMO protocol. To demonstrate the usefulness of the Auto-FMO protocol, we compared the ligand binding interaction energies by the Auto-FMO protocol with those of manually prepared data. In most cases, the data calculated by the Auto-FMO protocol showed reasonable agreement with the manually prepared data. Further improvement of the protocol is necessary for the treatment of ionization and tautomerization at the structure preparation stage, because some outlier data were observed due to these issues. The Auto-FMO protocol provides a powerful tool to deal with huge numbers of complexes for drug design, as well as for the construction of the FMO database (<http://drugdesign.riken.jp/FMODB/>) released in 2019.

**Key Words:** Fragment molecular orbital (FMO), intermolecular interaction, ligand binding energy, estrogen receptor  $\alpha$  (ER $\alpha$ ), p38 mitogen-activated protein (MAP) kinase, FMO database

**Area of Interest:** Molecular recognition and molecular modeling

## 1. Introduction

*Ab initio* quantum mechanical calculations for whole large biomolecules can be efficiently performed by the fragment molecular orbital (FMO) method [1–3]. An inter-fragment interaction energy (IFIE) analysis based on FMO calculations can easily represent the detailed interactions in fragment units. The FMO method is already recognized as a useful drug design tool to analyze ligand binding interactions, incorporating electrostatic interactions such as hydrogen bonds and dispersion forces such as CH/ $\pi$  interactions, using the pair interaction energy decomposition analysis (PIEDA) [4, 5] and fine fragmentation by the functional group unit, rather than the amino acid residue unit and the whole ligand [6–8]. Recently, the IFIE analysis and its energy decomposition analysis have been applied to the prediction of binding affinity for rational drug design [9–18]. Using FMO calculations of tens of complexes for one target protein, the essential and characteristic interactions of the ligand binding mode can be abstracted from the IFIE and PIEDA data by clustering methods [19–20] and singular value decomposition [21]. In addition, the prediction of the activity cliff, which is very difficult using conventional molecular mechanics (MM)-based scoring functions, such as Glide score and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA), was successfully accomplished by the FMO method with molecular mechanics Poisson–Boltzmann surface area (FMO+MM-PBSA) approach incorporating MM-based desolvation effects, using protein–ligand complexes optimized by the quantum mechanics/molecular mechanics (QM/MM) method [22]. Moreover, the FMO based polarizable continuum model (FMO-PCM) [23] or FMO based Poisson–Boltzmann surface area (FMO-PBSA) [24–26] methods provide more reliable results in solution, by using a fully-polarizable medium for the solute.

Since 2014, we have performed FMO calculations for various drug discovery targets, such as kinases, nuclear receptors, proteases, and protein–protein interactions (PPIs), with experimental binding affinities ( $IC_{50}$ ,  $K_i$ ,  $K_d$  values) as the activities of the FMO drug design consortium (FMOOD) [27]. To calculate the huge number of different structures by a manual procedure, we must investigate and choose the various modeling conditions and FMO settings one by one. For example, appropriate structure preparation, which includes complementation of missing atoms or missing residues, addition of hydrogen atoms, and structure minimization, is critically important as the preprocessing before the FMO calculations. However, appropriate methods for preprocessing have not yet been established. We discussed the modeling conditions for the complementation of heavy atoms, with/without water molecules, and the restraint of heavy atoms on the minimization in the FMOOD consortium. As a result, some modeling case studies have been reported [28–34]. Another issue is the treatment of a large amount of structure data, including more than 145,000 protein data bank (PDB) entries, to construct an FMO database [35] in the future. There are limits to human power in preparing a huge number of structures by a manual operation. In addition, it is not easy to appropriately perform FMO calculations for inexperienced researchers, in terms of the structure preparation and FMO settings. Thus, we have started to develop “an automated FMO calculation protocol” (Auto-FMO protocol). In this paper, we constructed the Auto-FMO protocol consisting of structural preparation based on the MM method, structural optimization based on the QM/MM calculations with our own  $N$ -layered integrated molecular orbital and molecular

mechanics (ONIOM) method, MM-PBSA and GBSA calculations, FMO calculations, and ligand interaction analysis from the MM and FMO calculations. We validated the Auto-FMO protocol by comparison of the ligand binding energies between the protocol data and the manually prepared data for human estrogen receptor  $\alpha$  (ER $\alpha$ ) and human p38 mitogen-activated protein (MAP) kinase. Based on the results, we discussed the current accuracy and issues to solve in the protocol in future work.

## 2. Material and Methods

The structure preparations and the FMO calculations for the X-ray crystal structures of the ER $\alpha$  and p38 MAP kinase were performed using the Auto-FMO protocol described in Section 2.1, and compared with the manually prepared data shown in Section 2.2. We verified the difference in the ligand binding energy values between the data obtained by the Auto-FMO protocol and those generated by the manual procedure. The ligand binding energy is defined in Section 2.3.

### 2.1 Workflow of Auto-FMO protocol

The workflow of the Auto-FMO protocol is shown in Figure 1, and the method for each step is described in the following sections. The protocol was implemented in BIOVIA Pipeline Pilot [36]. Further details of the structure preparation step, such as the default parameters, are described in the supplementary information, Section A.

#### Input files

The three-dimensional structures of protein–ligand complexes or apo protein structures were prepared in PDB files and transformed into one MDB file [37]. This part should be done by the users, rather than by the Auto-FMO protocol. Water molecules far from each ligand can be removed if necessary.

#### Structure preparation

To complement the missing residues and atoms, two different types of functions were newly implemented with built-in functions in MOE: (1) Both missing residues and atoms were simultaneously complemented by the “pro\_HomologyModel” function to build a whole protein model, and (2) only the missing atoms of the side chains are complemented by the “StructurePreparation” function. In gap regions due to disorder, the residues next to the gaps are capped by an ACE or NME group to retain the original amide bonds. After the corrections, hydrogen atoms were added to each complex by using the “Protonate3D” function under pH 7.0 conditions. The residues at the N- and C-termini are treated as zwitterionic states with NH $_3^+$  and COO $^-$ , respectively.

Subsequently, structure optimization with the force field (Amber10:EHT) was performed, using the “MM” function under partial constraints. As the force field, other force fields such as MMFF94x can be selected. For example, only the complemented atoms and hydrogen atoms can be optimized. The constraint settings can be adjusted by the users.

#### QM/MM-based structure optimization (advanced option)

Structural optimization based on the QM/MM calculations with the ONIOM method at the HF/6-31G\*:MM/AMBER level can be automatically performed using the Gaussian09 program package [38] at the RIKEN super computer HOKUSAI, as an advanced option. A ligand and its

surrounding residues are generally selected as the high layer, and the rest of the biomolecule is assigned to the low layer region. Atoms of the high layer region are only optimized and the other atoms are fixed during the calculations. This option is not active as the default setting, because the QM/MM optimization requires high computational cost. Users should employ the QM/MM optimization appropriately, where there are the necessities of QM calculation to cope with the cases such as halogen bond. In the protocol, the QM/MM-based total energies were calculated by ONIOM method.

### MM-PBSA and GBSA calculations

The total MM energies and the solvation energies of each protein–ligand complex, apo-protein, and ligand were calculated by the MM-PBSA and MM-GBSA methods with the AMBER tools program [39]. The Amber99SB force field was used for the protein, and the general Amber force field (GAFF) [40] with the AM1-BCC [41] charge was used for the ligands by antechamber [42]. Conformational searches of ligands in solvent can be performed by MOE to estimate the deformation energies of the ligands in the bound state and the solvent for MM-PBSA, as an advanced option. Several candidates of stable structures in the solvent are used.

### FMO calculations

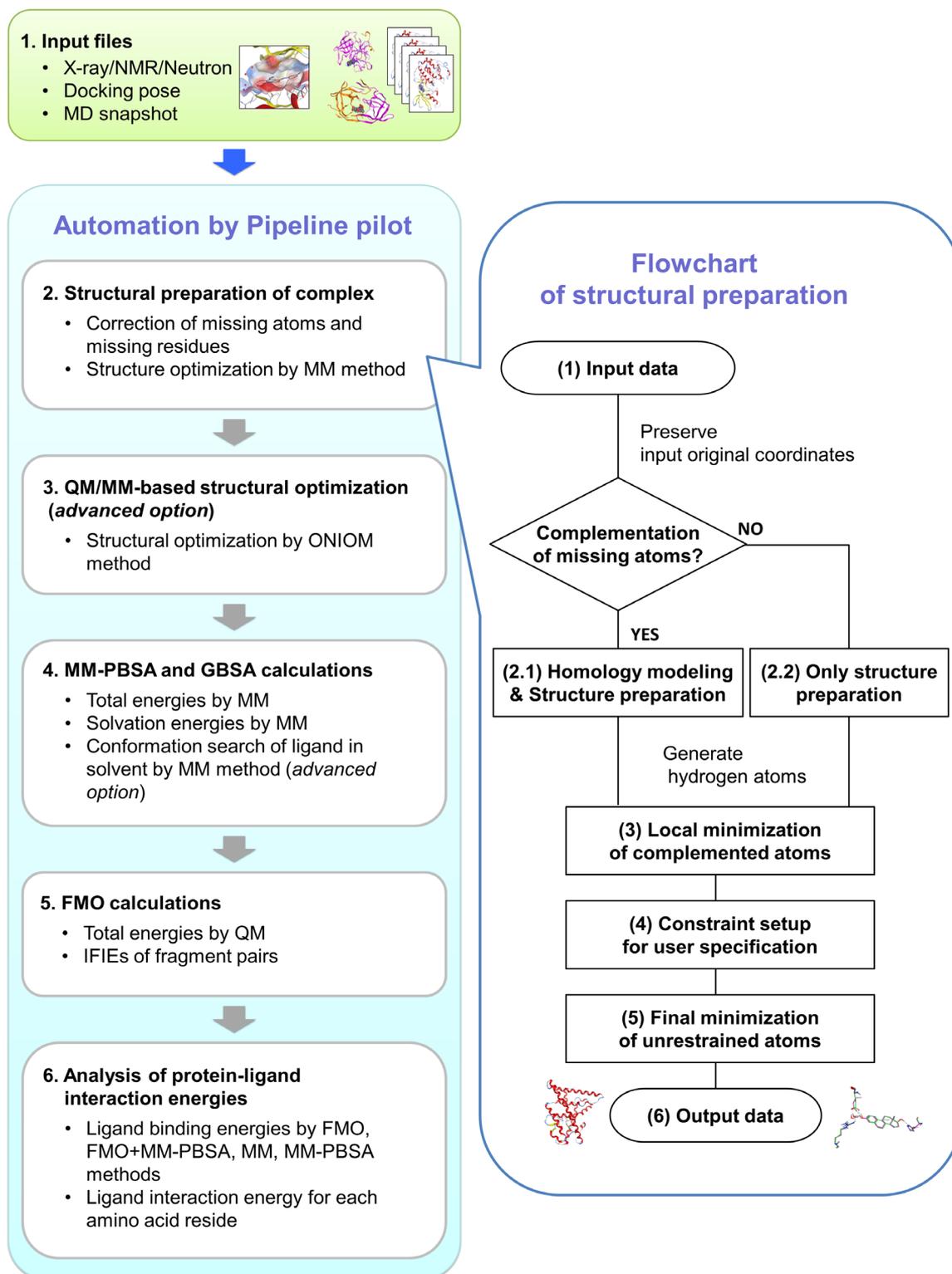
FMO calculations were performed with the ABINIT-MP program [43, 44] on the K computer or our in-house PC cluster server. Users can select various options for the FMO method: a computational level from the Hartree–Fock (HF) and the Møller–Plesset (MP2) level, and a basis set from 6-31G, 6-31G\*, and cc-pVDZ. In this paper, we used the MP2/6-31G\* level. The output log files of the FMO calculations include the IFIE and PIEDA values, as well as partial atomic charges by a Mulliken population analysis, a natural population analysis (NPA), and RESP fitting.

### Analysis of protein–ligand interaction energies

Ligand binding energies based on the MM, MM-PBSA, FMO, FMO+MM-PBSA methods can be extracted from the log files and summarized by the protocol. In addition, the IFIE values of the ligands are listed in an Excel file. The detailed definitions of these energies regarding the ligand are explained in ref. [22]

## 2.2 Validation data sets of ER $\alpha$ and p38 MAP kinase

For ER $\alpha$  and p38 MAP kinase (p38 $\alpha$ ), the complex structures with IC<sub>50</sub> values were selected from the PDB and ChEMBL databases. The datasets of ER $\alpha$  and p38 $\alpha$  have 37 and 96 PDB entries, respectively (Tables S2 and S3). Here, we executed the FMO calculations of 38 ER $\alpha$  and 111 p38 $\alpha$  complexes by the Auto-FMO protocol because the PDB data includes multiple chain complexes such as homo dimer. Since the coordinates of these residues are commonly given by X-ray crystallography, the amino acid residues of #309–544 (ER $\alpha$ ) and #6–351 (p38 $\alpha$ ) were employed. The results of the FMO calculations at the MP2/6-31G\* level using the automatically prepared structures were compared to those obtained using manually prepared structures. As the manually prepared data, we used already reported 70 p38 $\alpha$  complex data [34]. In terms of ER $\alpha$ , we selected 20 ER $\alpha$  complexes considering the crystal resolutions and the other reasons. These complex structures were included in the above mentioned 38 and 111 complex structures calculated by the Auto-FMO protocol. Thus, we used the 20 and 70 complex structures for comparison (Table 1). The PDB entries used for comparison are shown in “PDB and chain IDs used for analysis” column, Table S2 and S3.



**Figure 1.** The workflow of the Auto-FMO protocol  
The detailed procedures of the structural preparation are depicted on the right.

The numbers of data for each preparation step by the Auto-FMO protocol and the manual procedure are summarized in Table 1. The details are described below. Figure 2 shows the superimposed complexes of ER $\alpha$  and p38 $\alpha$  obtained from the manual preparation. In the case of ER $\alpha$ , the position of helix 12 (H12), composed of residues Leu536–Ala546, was different depending on whether the bound compound was an agonist or antagonist (Figure 2A). Hence, the loop between H11 and H12 near the ligand binding pocket is flexible, and there are missing residues on the loop in some complexes with antagonists (Table S2). In the case of p38 $\alpha$ , the DFG loop around Ile166–Ala184, including Asp168, Phe169, and Gly170, adopts three types of conformations, such as DFG-in, DFG-out, and DFG-intermediate conformations (Figure 2B). The DFG-loop is placed in the ATP binding pocket and is very flexible and disordered, including the missing residues (Table S2). In this paper, because these missing residues of ER $\alpha$  and p38 $\alpha$  are located near the bound ligands, they were complemented by both modeling methods (Auto-FMO protocol and manual operation).

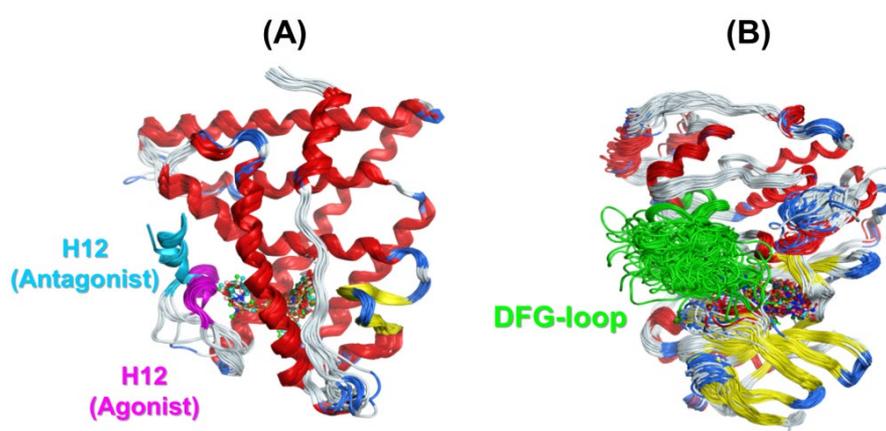
**Table 1.** Data set list for the automated FMO protocol and manually prepared data.

	Target protein	Estrogen receptor $\alpha$	p38-MAP kinase
<b>Manually prepared data</b>	# of completed FMO calculations	20	70
<b>Auto-FMO protocol data</b>	# of completed FMO calculations	20	70
	Average time per complex for structure preparation	3.2 minutes <sup>a</sup>	4.7 minutes <sup>a</sup>
	Average time per complex for FMO calculation	11 hours <sup>b</sup>	14 hours <sup>c</sup>

<sup>a</sup>Average time of structure preparation by MM optimization for each structure using 1 core at our in-house PC cluster server (Intel(R) Xeon(R) CPU E5-2695 v3 @ 2.30GHz).

<sup>b</sup>Average time of FMO calculation for each structure using 48 cores at our in-house PC cluster server (Intel(R) Xeon(R) CPU E5-2670 v3 @ 2.30GHz).

<sup>c</sup>Average time of FMO calculation for each structure using 960 cores at K computer.



**Figure 2.** Structures of ER $\alpha$  (A) and p38 $\alpha$  (B)

For ER $\alpha$  complexes, the H12 regions of the agonist-bound form and antagonist-bound form are represented by pink and light blue ribbons, respectively. For p38 $\alpha$ , the DFG-loop is shown by a green ribbon.

### Auto-FMO protocol data

In the cases of both ER $\alpha$  and p38 $\alpha$ , the bond orders of all of the ligands were corrected based on LigandExpo and used for the input structures of the Auto-FMO protocols. Subsequently, the Auto-FMO protocol was employed using the same options, as follows. All of the crystal water molecules were removed; two residues neighboring missing residues as a margin were removed and complemented by homology modeling using the wild type sequence; and the protonation states of the ligands and amino acid residues were determined by Protonate3D at pH 7.0. Finally, all of the hydrogen and missing heavy atoms were optimized by the Amber10:EHT force field.

### Manually prepared data

In the case of ER $\alpha$ , a few crystallographic water molecules forming bridging hydrogen bonds among Glu353, Arg394, and the ligand were retained, in addition to the ligands and proteins. By using the BioStation viewer, the missing residues of the complexes with agonists and antagonists were complemented according to the templates of the agonist form (PDB IDs: 1A52 and 2YJA) and the antagonist form (PDB ID: 1Y1M), which are high-resolution X-ray structures that have no missing residues in the H12 regions. The bond orders of all of the ligands were manually corrected. The tautomeric states of amino acid residues and ligands were consequently determined with Protonate3D in MOE. Finally, all of the hydrogens and missing heavy atoms of each ER $\alpha$  complex were optimized by an Amber10:EHT force field.

For p38 $\alpha$ , the structure coordinates prepared by Sheng *et al.* [34] were used in this study. First, a few crystallographic water molecules were kept in the following two cases. In the first case, the water molecule forming hydrogen bonds with Asp168 and Lys53 was retained in the DFG-in structures, and in the second case, the water molecule forming the hydrogen bond with Asp168 was kept in the DFG-out structures. Secondly, by using the BioStation viewer, the missing residues were complemented with the 3GC7 and 3D83 PDB entries as the template structures of the DFG-in and DFG-out structures, respectively. The templates were full sequences and high resolution structures, without missing residues. The bond orders of all of the ligands were corrected manually. The tautomeric states of the amino acid residues and ligands were consequently determined with Protonate3D in MOE. Only His312 was dealt with as a cationic protonation state (HIP), considering the surrounding hydrogen bonding network, and the other histidine residues were set as neutral states (HIE or HID). Moreover, the ligand charges were assigned as the corresponding protonation state of pH 7.0 in water. For aliphatic amines, we set the formal charges to +1e, while aromatic amines such as anilines were set to a formal charge of zero. Finally, all hydrogen and missing heavy atoms for each p38 $\alpha$  complex were optimized by an Amber 10:EHT force field.

## 2.3 Ligand binding interaction energy evaluated by FMO method

In this article, to demonstrate the utility of the Auto-FMO protocol, we compared the ligand binding energies  $\Delta E_{\text{ligand}}$  of the protocol with those of the manually prepared data. To obtain  $\Delta E_{\text{ligand}}$  based on FMO calculations, we summed up the IFIE values of all ligand–residue pairs. The FMO contribution of the ligand-binding energy is given by the following equation:

$$\Delta E_{\text{ligand}} = \sum_{\substack{I=\text{ligand} \\ J \neq I}} \Delta \tilde{E}_{IJ}. \quad (1)$$

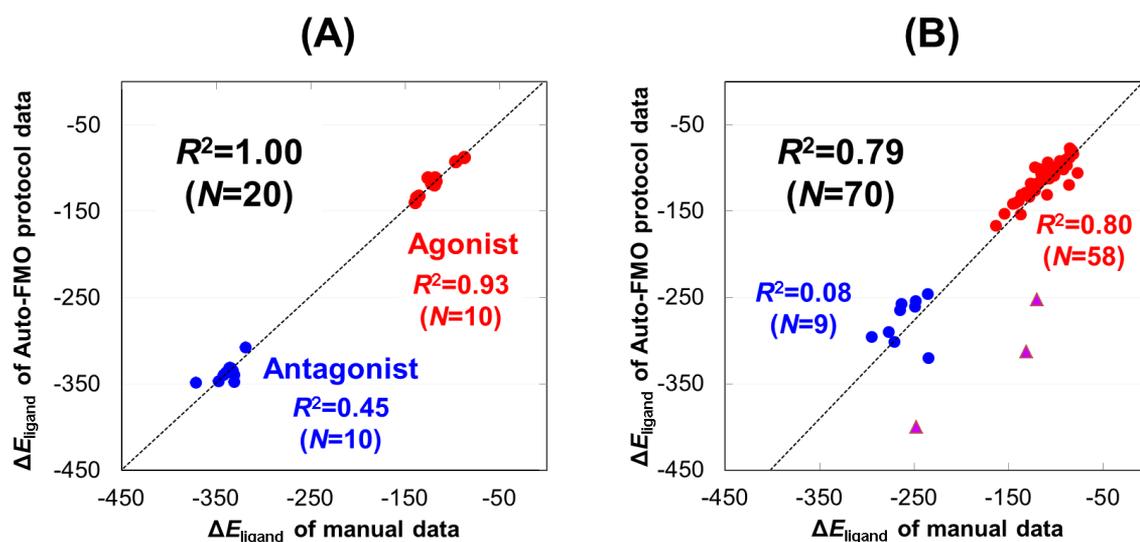
where  $\Delta \tilde{E}_{IJ}$  is the IFIE; and  $I$  and  $J$  are fragment indices.

### 3. Results and discussion

We constructed 38 and 111 complexes of ER $\alpha$  and p38 $\alpha$  by the MM-based structure preparation in the Auto-FMO protocol. The number of calculated structures includes multiple chains in each PDB entry. Using our in-house PC cluster server and the K computer, 38 and 110 FMO calculations were successfully completed for the ER $\alpha$  and p38 $\alpha$  complexes, respectively. There was one structure of p38 $\alpha$  (PDB ID: 4KIN, Chain: D) for which the FMO calculation was not completed. At the structure complement step by homology modeling, the complemented atoms of the complex severely clashed among Gly33, Ala34, and Tyr35. As a result, the FMO calculations were not completed for this structure. However, 99.3% of the FMO calculations using the structures created by the Auto-FMO protocol were completed, showing that this protocol is sufficiently robust for daily FMO research, including drug design. Next, we validated the Auto-FMO protocol by comparing the ligand binding energies  $\Delta E_{\text{ligand}}$  between the protocol data and the manually prepared data, and discussed the accuracy and the issues to improve in the protocol.

#### 3.1 Validation of the Auto-FMO protocol data

For accuracy validation of the protocol, the ligand binding interaction energies calculated by the Auto-FMO protocol were compared to those of the manually prepared data (Figure 3). Here, we compared 20 and 70 FMO calculation results of the structures in which the PDB ID and its chain ID matched in both methods for ER $\alpha$  and p38 $\alpha$ , respectively (Table 1).



**Figure 3.** Correlation of ligand binding energies between the manual data and the Auto-FMO protocol data for ER $\alpha$  (A) and p38 $\alpha$  (B)

The neutral and positively charged ligands are marked with red and blue circles, respectively.

Obvious outlier data (PDB IDs: 3O8P, 1OUK, 3GFE) in Figure 3B, for which the ligand charge of the protocol data differed from that of the manual data, are shown by purple triangles.

The coefficients of determination ( $R^2$ ) for all calculation data of ER $\alpha$  and p38 $\alpha$  were 1.00 and 0.79, respectively. Note that the p38 $\alpha$  data included different ligand charge data between the Auto-FMO protocol data and the manual data. In Figure 3B, the entries with different charges shown by purple triangles were located at significant outlier positions. Although the ionization state assignment using MOE should be improved, except for the three outliers (PDB IDs: 3O8P, 1OUK, 3GFE, also see Table S1),  $R^2$  improved to 0.95. In general, the differences in the formal charges seriously affect the calculation results, especially for a QM calculation. Therefore, careful double checking of the structure preparation by the Auto-FMO protocol is needed for practical use. The high correlations between the manual and the protocol data partially arise from the wide  $\Delta E_{\text{ligand}}$  range between the compound groups (agonist and antagonist groups in ER $\alpha$  and neutral ligand and charged ligand groups in p38 $\alpha$ ). To confirm the effects of the ligand charge, the compounds were classified by each ligand charge (neutral (red) and positively charged (blue) in Figure 3). For only neutral ligands, the  $R^2$  values for ER $\alpha$  and p38 $\alpha$  were sufficiently high (0.93 and 0.80, respectively), in spite of the narrow  $\Delta E_{\text{ligand}}$  range. In contrast, only positively charged ligands provided poor correlations ( $R^2 = 0.45$  and 0.08, respectively). In the cases of charged ligands, slight changes of the atomic coordinates (*e.g.*, 0.1 to 0.2 Å) can greatly affect the interaction energy values.

### 3.2 Issues to be solved in the Auto-FMO protocol

There was one data that the FMO calculation did not complete in Table 1. To reduce the computational cost, we plan to add an MM-based structure filter that removes incorrect structures before submission to the FMO calculation. The filter will check the bond lengths, bond angles, dihedral angles, and MM-based repulsion energies of residue in each complex according to their standard values observed in X-ray structures.

In the case of p38 $\alpha$ , we detected outliers for the comparison between the Auto-FMO protocol data and the manually prepared data in Figure 3B. The large differences in the ligand binding energies between the protocol data and the manually prepared data, which are more than 20 kcal/mol for p38 $\alpha$ , are listed in Table S1. The main reasons for the large differences in the ligand binding energies between the manual data and the protocol data were the differences in the ionization and tautomerization states of the ligands (Table S1), and the complemented structures regarding missing residues on flexible loops (Figures 2 and S1). The  $R^2$  values of the ligand binding energies between the protocol data and the manual data without the nine obviously different compounds were dramatically improved, from 0.79 to 0.99 in the p38 $\alpha$  data set (Figure S2). In this article, we handled the manually prepared data as representative data of ordinary computational researchers. However, with respect to the ligand structures, the manual preparation is not always effective, because many of the computational researchers are not familiar with the tautomerization and protonation states of a wide variety of chemical species. Indeed, the data with large  $\Delta E_{\text{ligand}}$  differences between manually prepared data and Auto-FMO protocol data (Table S1) contain inappropriate ligand structures on both sides. Therefore, it is necessary to manually check the ionization states and conformations of the complemented residues, using a list of the 2D structures of all ligands and the 3D structures of the complexes output from the protocol. Currently, the ionization assignment depends on the MOE protonate3D function, and it is not easy to enhance the function within MOE. In the future, we would like to develop our original ionization and tautomerization prediction model by combining MM or QM calculations and artificial intelligence (AI), based on the upcoming FMO database [35] including more than thousand FMO data.

The Auto-FMO protocol can be used for not only construction of the FMO database but also binding affinity prediction and interaction analysis. In this paper, we described the details of protocol and good agreements with the manual data. In general, for the binding affinity prediction

based on the FMO method, we have to consider many factors such as ligand properties including charges and functions (i.e. agonist and antagonist), entropy terms, and solvent effects. Indeed, the previous study of p38 $\alpha$  [34] showed good correlation between experimental values and estimated binding interaction energies after appropriate classification based on ligand charges, scaffolds, and types of DFG-loop structures. Considering the entropy and solvation effects [11, 22] is also very important to improve the correlation. These effects for the current datasets will be considered in our future works by using FMO+MM-PBSA approach. To enhance accuracy of the FMO+MM-PBSA method, we plan to use the QM-based atomic charges obtained by FMO calculations in addition to the QM/MM optimization. Because the above mentioned calculations are time consuming and the ligand clustering requires a deep understanding of the target and dataset, more advanced expert system is needed for practical applications to drug design. One of the future directions of development is an AI-based guidance for the binding affinity prediction.

## 4. Conclusion

We have developed the automated FMO calculation protocol (Auto-FMO protocol) to perform a large number of FMO calculations, together with pre- and post-processing. To validate the Auto-FMO protocol, the ligand binding interaction energies calculated by the Auto-FMO protocol were compared to those of the manually prepared data in the ER $\alpha$  and p38 $\alpha$  datasets. The automated FMO calculation data agreed reasonably well with the manually prepared data ( $R^2$ : 1.00 (ER $\alpha$ ) and 0.79 (p38 $\alpha$ )). One of the factors that reduced the correlation is the misassignment of the ionization and tautomerization; for only the neutral ligands or ligands with the same ionization and tautomerization assignments, the  $R^2$  values were significantly improved. Therefore, Auto-FMO protocol users need to check the ionization and tautomerization states of the ligands to avoid systematic misassignments. In the cases of complex structures with missing atoms in flexible regions, there is some room for improvement in the protein modeling to complement the missing atoms. With this protocol, not only computational researchers but also inexperienced researchers can easily perform FMO calculations without complicated procedures. The Auto-FMO protocol would reduce not only operative fluctuations but also artificial mistakes in such as bond-order setting for ligands. This makes us prepare the uniformly organized data set. In addition, the Auto-FMO protocol will enable us to construct the FMO database [35], which will provide quantum mechanical calculations of total energies, interaction energies of fragment pairs, and atomic charges for various protein–ligand complexes for more than a thousand complex structures in the future. This database will be useful for various research fields, such as drug design and structural biology, as well as molecular modeling [11, 45–47]. For example, using thousands of FMO calculation results, AI models that predict molecular force and atomic charges considering QM interactions are currently being developed. The new AI-based molecular force field will lead to more accurate molecular dynamics and docking simulations. Furthermore, we plan to establish an “on the fly” FMO calculation service for structural biologists. Once a structural biologist solves a new X-ray structure, the structure data can be uploaded to the service and the FMO calculation data can be obtained in one day.

## Supplemental information

The details of the structure preparation of complexes by the MM method for the automated FMO calculation protocol are shown in section A. The large differences in the ligand binding

energies between the protocol and the manually prepared data are discussed in section B. All PDB data of ER $\alpha$  and p38 $\alpha$  are available in Tables S2 and S3, respectively.

## Acknowledgements

The authors would like to thank Dr. Kazumi Tsuda and Mr. Daisuke Murayama at Science & Technology Systems, Inc., for supporting the development of the automated FMO calculation protocol, and Mr. Satoshi Anzaki and Ms. Yinglei Sheng at Kobe University for providing the manually prepared ER $\alpha$  and p38 $\alpha$  data, and Mr. Yuya Seki and Mr. Yuya Handa at Hoshi University for supporting of IFIE data analysis. This research was conducted as part of the activity of the FMO drug design consortium (FMOOD). The FMO calculations were performed on the K computer (project IDs: hp150160, hp160103, hp170183, and hp180147). For QM/MM optimizations, the supercomputer HOKUSAI (RIKEN Advanced Center for Computing and Communications) was used. PIEDA calculations were performed with the MIZUHO/BioStation software package. This research was partially supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP18am0101113. Finally, CW and DT acknowledge JSPS KAKENHI Grant Number JP18K06619.

## References

- [1] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: an Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.
- [2] Fedrov, D.G.; Kitaura, K. The Fragment Molecular Orbital Method: Practical Application to Large Molecular System; CRC Press, 2009.
- [3] Tanaka, S.; Mochizuki, Y.; Komeiji, Y.; Okiyama, Y.; Fukuzawa, K. Electron-correlated Fragment-molecular-orbital Calculations for Biomolecular and Nano Systems. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10310–10344.
- [4] Fedorov, D. G.; Kitaura, K. Pair interaction energy decomposition analysis. *J. Comp. Chem.* **2007**, *28*, 222–237.
- [5] Tsukamoto, T.; Kato, K.; Kato, A.; Nakano, T.; Mochizuki, Y. *et al.* Implementation of Pair Interaction Energy Decomposition Analysis and its Applications to Protein–Ligand Systems. *J. Comput. Chem. Jpn.* **2015**, *14*, 1–9.
- [6] Hitaoka, S.; Harada, M.; Yoshida, T.; Chuman, H. Correlation Analyses on Binding Affinity of Sialic Acid Analogues with Influenza Virus Neuraminidase-1 Using *ab Initio* MO Calculations on Their Complex Structures. *J. Chem. Inf. Model.* **2010**, *50*, 1796–1805.
- [7] Hitaoka, S.; Matoba, H.; Harada, M.; Yoshida, T.; Tsuji, D. *et al.* Correlation Analyses on Binding Affinity of Sialic Acid Analogues and Anti-Influenza Drugs with Human Neuraminidase Using *ab Initio* MO Calculations on Their Complex Structures-LERE-QSAR Analysis (IV). *J. Chem. Inf. Model.* **2011**, *51*, 2706–2716.
- [8] Watanabe, C.; Fukuzawa, K.; Okiyama, Y.; Tsukamoto, T.; Kato, A. *et al.* Three- and Four-body Corrected Fragment Molecular Orbital Calculations with a Novel Subdividing Fragmentation Method Applicable to Structure-based Drug Design. *J. Mol. Graph. Model.* **2013**, *41*, 31–42.
- [9] Fukuzawa, K.; Kitaura, K.; Uebayasi, M.; Nakata, K.; Kaminuma, T. *et al.* *Ab Initio* Quantum

- Mechanical Study of the Binding Energies of Human Estrogen Receptor  $\alpha$  with Its Ligands: An Application of Fragment Molecular Orbital Method. *J. Comput. Chem.* **2005**, *26*, 1–10.
- [10] Watanabe, H.; Tanaka, S.; Okimoto, N.; Hasegawa, A.; Taiji, M. *et al.* Comparison of Binding Affinity Evaluations for FKBP Ligands with State-of-the-Art Computational Methods: FMO, QM/MM, MM-PB/SA, and MP-CAFE Approaches. *CBIJ.* **2010**, *10*, 32–45.
- [11] Mazanetz, M. P.; Ichihara, O.; Law, R. J.; Whittaker, M. Prediction of cyclin-dependent kinase 2 inhibitor potency using the fragment molecular orbital method. *J. Cheminform.* **2011**, *3*, 2.
- [12] Heifetz, A.; Chudyk, E. I.; Gleave, L.; Aldeghi, M.; Cherezov, V. *et al.* The fragment molecular orbital method reveals new insight into the chemical nature of GPCR–ligand interactions. *J. Chem. Inf. Model.* **2016**, *56*, 159–172.
- [13] Heifetz, A.; Aldeghi, M.; Chudyk, E. I.; Fedorov, D. G.; Bodkin, M. J. *et al.* Using the fragment molecular orbital method to investigate agonist–orexin-2 receptor interactions. *Biochem. Soc. Trans.* **2016**, *44*, 574–581.
- [14] Heifetz, A.; Trani, G.; Aldeghi, M.; MacKinnon, C. H.; McEwan, P. A. *et al.* Fragment molecular orbital method applied to lead optimization of novel interleukin-2 inducible T-cell kinase (ITK) inhibitors. *J. Med. Chem.* **2016**, *59*, 4352–4363.
- [15] Morao, I.; Fedorov, D. G.; Robinson, R.; Southey, M.; Townsend-Nicholson, A. *et al.* Rapid and accurate assessment of GPCR–ligand interactions using the fragment molecular orbital-based density-functional tight-binding method. *J. Comp. Chem.* **2017**, *38*, 1987–1990.
- [16] Okimoto, N.; Otsuka, T.; Hirano, Y.; Taiji, M. Use of the Multilayer Fragment Molecular Orbital Method to Predict the Rank Order of Protein–Ligand Binding Affinities: A Case Study Using Tankyrase 2 Inhibitors. *ACS Omega*, **2018**, *3*, 4475–4485.
- [17] Takaya, D.; Inaka, K.; Omura, A.; Takenuki, K.; Kawanishi, M. *et al.* Characterization of crystal water molecules in a high-affinity inhibitor and hematopoietic prostaglandin D synthase complex by interaction energy studies. *Bioorg. Med. Chem.* **2018**, *26*, 4726–4734.
- [18] Xu, F.; Tanaka, S.; Watanabe, H.; Shimane, Y.; Iwasawa, M.; Ohishi, K. *et al.* Computational Analysis of the Interaction Energies between Amino Acid Residues of the Measles Virus Hemagglutinin and Its Receptors. *Viruses* **2018**, *10*, 236.
- [19] Amari, S.; Aizawa, M.; Zhang, J.; Fukuzawa, K.; Mochizuki, Y. *et al.* VISCANA: Visualized Cluster Analysis of Protein–Ligand Interaction Based on the ab Initio Fragment Molecular Orbital Method for Virtual Ligand Screening. *J. Chem. Inf. Model.* **2006**, *46*, 221–230.
- [20] Kurauchi, R.; Watanabe, C.; Fukuzawa, K.; Tanaka, S. Novel Type of Virtual Ligand Screening on the Basis of Quantum-Chemical Calculations for Protein–Ligand Complexes and Extended Clustering Techniques. *Comput. Theor. Chem.* **2015**, *1061*, 12–22.
- [21] Maruyama, K.; Sheng, Y.; Watanabe, H.; Fukuzawa, K.; Tanaka, S. Application of Singular Value Decomposition to the Inter-Fragment Interaction Energy Analysis for Ligand Screening. *Comput. Theor. Chem.* **2018**, *1132*, 23–34.
- [22] Watanabe, C.; Watanabe, H.; Fukuzawa, K.; Parker, L. J.; Okiyama, Y. *et al.* Theoretical Analysis of Activity Cliffs among Benzofuranone-Class Pim1 Inhibitors Using the Fragment Molecular Orbital Method with Molecular Mechanics Poisson–Boltzmann Surface Area (FMO+MM-PBSA) Approach. *J. Chem. Inf. Model.* **2017**, *57*, 2996–3010.
- [23] Alexeev, Y.; Mazanetz, M. P.; Ichihara, O.; Fedorov, D. G. GAMESS as a free quantum-mechanical platform for drug research. *Curr. Top. Med. Chem.* **2012**, *12*, 2013–2033.
- [24] Watanabe, H.; Okiyama, Y.; Nakano, T.; Tanaka, S. Incorporation of Solvation Effects into the Fragment Molecular Orbital Calculations with the Poisson–Boltzmann Equation. *Chem. Phys. Lett.* **2010**, *500*, 116–119.

- [25] Okiyama, Y.; Nakano, T.; Watanabe, C.; Fukuzawa, K.; Mochizuki, Y. *et al.* Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson–Boltzmann Equation: Implementation and DNA Study. *J. Phys. Chem. B*, **2018**, *122*, 4457–4471.
- [26] Okiyama, Y.; Watanabe, C.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T. *et al.* Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson–Boltzmann Equation: II. Protein and Its Ligand-Binding System Studies. *J. Phys. Chem. B*, **2019**, *123*, 957–973.
- [27] FMO drug design consortium (FMOODD), <http://eniac.scitec.kobe-u.ac.jp/fmodd/> (accessed on February 20, 2019).
- [28] Kobayashi, I.; Takeda, R.; Suzuki, R.; Shimamura, K.; Ishimura, H. *et al.* Specific interactions between androgen receptor and its ligand: *ab initio* molecular orbital calculations in water. *J. Mol. Graph. Model.* **2017**, *75*, 383–389.
- [29] Ozawa, M.; Ozawa, T.; Nishio, M.; Ueda, K. The role of CH/ $\pi$  interactions in the high affinity binding of streptavidin and biotin. *J. Mol. Graph. Model.* **2017**, *75*, 117–124.
- [30] Ozawa, M.; Ozawa, T.; Ueda, K. Application of the fragment molecular orbital method analysis to fragment-based drug discovery of BET (bromodomain and extra-terminal proteins) inhibitors. *J. Mol. Graph. Model.* **2017**, *74*, 73–82.
- [31] Takeda, R.; Kobayashi, I.; Shimamura, K.; Ishimura, H.; Kadoya, R. *et al.* Specific interactions between vitamin-D receptor and its ligands: *ab initio* molecular orbital calculations in water. *J. Steroid Biochem. Mol. Biol.* **2017**, *171*, 75–79.
- [32] Takeda, R.; Kobayashi, I.; Suzuki, R.; Kawai, K.; Kittaka, A. *et al.* Proposal of potent inhibitors for vitamin-D receptor based on *ab initio* fragment molecular orbital calculations. *J. Mol. Graph. Model.* **2018**, *80*, 320–326.
- [33] Takeda, R.; Suzuki, R.; Kobayashi, I.; Kawai, K.; Kittaka, A. *et al.* Specific interactions between vitamin D receptor and ligand depending on its chirality: *ab initio* fragment molecular orbital calculations. *CBIJ*, **2018**, *18*, 32–43.
- [34] Sheng, Y.; Watanabe, H.; Maruyama, K.; Watanabe, C.; Okiyama, Y. *et al.* Towards good correlation between fragment molecular orbital interaction energies and experimental IC<sub>50</sub> for ligand binding: A case study of p38 MAP kinase. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 421–434.
- [35] FMO database, <http://drugdesign.riken.jp/FMODB/> (accessed on February 20, 2019).
- [36] *BIOVIA Pipeline Pilot*, 17.2.0.1361; Dassault Systèmes BIOVIA: 5005 Wateridge Vista Drive, San Diego, CA 92121 USA, 2017.
- [37] *Molecular Operating Environment (MOE)*, 20151001; Chemical Computing Group Inc.: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.
- [38] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A. *et al.* *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2009.
- [39] Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J. *et al.* 2008, *AMBER 10*; University of California: San Francisco CA, 2008.
- [40] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- [41] Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- [42] Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- [43] *BioStation: ABINIT-MP and BioStation Viewer*; The program package is available at: <http://www.ciss.iis.u-tokyo.ac.jp/english/dl/> (accessed on February 20, 2019) and <http://www.ciss.iis.u-tokyo.ac.jp/riss/dl/download/index.php> (accessed on February 20, 2019).

- [44] *MIZUHO/BioStation Viewer 3.0*; Mizuho Information and Research Institute Inc., 2013.
- [45] Saitou, S.; Iijima, J.; Fujimoto, M.; Mochizuki, Y.; Okuwaki, K. *et al.* Application of TensorFlow to recognition of visualized results of fragment molecular orbital (FMO) calculations. *CBIJ*. **2018**, *18*, 58–69.
- [46] Doi, H.; Okuwaki, K.; Mochizuki, Y.; Ozawa, T.; Yasuoka, K. Dissipative particle dynamics (DPD) simulations with fragment molecular orbital (FMO) based effective parameters for 1-Palmitoyl-2-oleoyl phosphatidyl choline (POPC) membrane. *Chem. Phys. Lett.* **2017**, *684*, 427–432.
- [47] Yoshida, M.; Hirono, S. Interaction analyses of CDK2 with its inhibitors by FMO calculation and PLS regression. *SAR News* **2017**, *32*, 32–33.  
[http://bukai.pharm.or.jp/bukai\\_kozo/SARNews/SARNews\\_32.pdf](http://bukai.pharm.or.jp/bukai_kozo/SARNews/SARNews_32.pdf).