A 3D-QSAR Analysis of CDK2 Inhibitors Using FMO Calculations and PLS Regression

Tomoki Yoshida* and Shuichi Hirono*

School of Pharmacy, Kitasato University; 5–9–1 Shirokane, Minato-ku, Tokyo 108–8641, Japan.

Received December 15, 2018; accepted February 23, 2019

We report a three-dimensional quantitative structure–activity relationship (3D-QSAR) analysis of CDK2 inhibitors using fragment molecular orbital (FMO) calculations and partial least squares (PLS) regression. In our analysis, fragment binding energies of individual amino acids and fragment binding energy of a single ligand in a protein–ligand complex are evaluated by FMO calculations and used as descriptors in PLS regression to estimate biological activities of the ligands. The analysis was applied to the system of CDK2 protein and its inhibitors and the effectiveness of the method was tested. Application of the 3D-QSAR model demonstrated that it offered good predictive ability and was able to predict not only biological activity of ligands but also identify important amino acid residues which could be targeted in order to improve ligand activity.

Key words fragment molecular orbital; partial least squares; protein–ligand interaction

Introduction

The fragment molecular orbital (FMO) method1,2) is one of the fragment-based calculation methods that make it possible to calculate the electronic state of an entire macromolecular system within a feasible time. In FMO methods, highly accurate quantum chemical calculations can be executed efficiently by means of dividing the whole protein into amino acid residue fragments. When the FMO method is applied to a protein–ligand complex, precise energetic information on the interaction between individual amino acid residue fragments–ligand pairs can be evaluated quantitatively.3,4) For this reason, the method has drawn a good deal of attention especially from the field of drug discovery. Fragment level energy information can also be obtained by using force field based energy decomposition approaches, however, in current force fields important physical effects, for example electronic polarizability of the electron clouds is not explicitly included and in general not accurate enough. The number of publications using FMO calculations to study the interaction between proteins and drug molecules is increasing.5,6) However, how to use the large amount of fragment level energy information in drug design has been a problem.

In the field of drug discovery, three-dimensional quantitative structure–activity relationship (3D-QSAR) is an important computational approach used for discriminating and predicting active compounds. Here, objective variables, such as the inhibitory activity of compounds are predicted using descriptors that depend on the 3D structural information of the compounds as explanatory variables. Among the linear regression methods which can be used in 3D-QSAR, partial least squares (PLS) regression is a well-known and often successfully applied technique.6) PLS regression is effective when the number of explanatory variables is larger than that of objective variables to be predicted and/or multi-collinearity between explanatory variables are high. For example, in the Comparative Molecular Field Analysis (CoMFA) method, which is a typical 3D-QSAR analysis method, an electrostatic field and a steric field surrounding each structurally superimposed series of compounds are used as descriptors. In this method, the number of explanatory variables exceeds the number of objective variables and the PLS regression is used. Important electrostatic and steric features are extracted from the obtained PLS regression coefficient and are used for designing more active compounds.7) PLS regression is suitable for processing a large amount of fragment level energy information obtained by FMO calculation such that multiple regression cannot be applied because the number of fragments (explanatory variables) is larger than the number of compounds (objective variables).

Some previous studies combining FMO calculations and regression analysis have been reported, including a Linear Expression by Representative Energy Terms (LERE)-QSAR analysis by Chuman and colleagues,8) a PLS regression of CDK2 inhibitors by Mazanetz et al.9) and a PLS regression of FK506-binding protein (FKBP) ligands by Otsuka et al.10) In all cases, the usefulness of the FMO method and regression analysis was demonstrated. However, in these studies, the interaction/binding energy between the whole protein and a single ligand was used as a descriptor, and individual energy information about the amino acid residue fragments was not used. It is predicted that by using the large amount of fragment level energy information more information can be extracted which can then be utilized for drug design.

In this study, we report a 3D-QSAR analysis method utilizing PLS regression with fragment binding energy11) which can be calculated using the FMO method as descriptors. Selecting CDK2 protein and its inhibitors as a test case, we examined whether it is possible to identify amino acid residues which would give ligand activity improvement using the PLS regression coefficient like CoMFA analysis. Protein–ligand complex structures required to calculate fragment binding energies were generated by docking calculations with CDK2 inhibitors. In order to obtain reliable complex structures, for each of the CDK2 inhibitors, shape and chemical similarity to the Pro-
tein Data Bank (PDB) ligand conformation in X-ray complex structures was evaluated, and the protein structure in complex with the most similar PDB ligand was used for the ligand docking calculations.

**Experimental**

**Preparation of ChEMBL Ligands** Known ChEMBL ligands with inhibitory activity ($K_i$) against CDK2 were extracted in SMILES format from the ChEMBL database according to the procedure shown in Fig. 1. They were converted to a three-dimensional structure, and possible ionization/tautomerization forms at pH 7 were generated by the program LigPrep. Subsequently, multiple conformations were generated using the intermediate mode of the program ConfGen.

**Preparation of Protein Structures** CDK2 protein is one of the cyclin dependent kinases that regulates the cell cycle and has been actively studied as a target for the development of anti-cancer drugs. We selected this protein as a research target because there are many known ligands whose inhibitory activity has been measured and for which there are X-ray structures of protein–ligand complexes.

Initially, X-ray complex structures containing CDK2 (UniProtKB Accession Number: P24941) were obtained from the PDB (379 entries), Fig. 1. From these entries, we extracted 96 high-quality X-ray structures of CDK2 in complex with a ligand. The following conditions were used for quality assessment: resolution $\leq 2.5$ Å, $R_{work} \leq 0.40$, $R_{free} \leq 0.45$, Single ligand bound, Ligand occupancy $= 1$, and Ligand EDIAm $> 0.8$. Here, the $R$-values and the DPI are the measures of the quality of the atomic model obtained from the crystallographic data. A good atomic model has a value close to zero. The ligand EDIAm is an index for evaluating the goodness of fit of the ligand model with the electron density around the model. A perfect fit would have a value of 1.2. The program DPICalc was used for the DPI calculation and ProteinsPlus Server was used for the EDIAm calculation. For each of these 96 PDB ligand–protein pairs, we performed re-docking calculations with the ligand and 83 PDB ligand–protein pairs were selected for which the ligand was successfully re-docked. Prior to the re-docking calculation, addition of hydrogen atoms, prediction of missing residues/loops, removal of molecules that are not ligands containing water molecules, optimization of hydrogen bonding networks and structural relaxation were performed for all complex structures using Protein Preparation Wizard. X-Ray ligand conformations were prepared using LigPrep and ConfGen under the same conditions as for ChEMBL ligands and the docking calculation was performed in standard accuracy SP mode docking using the program Glide. The success of the re-docking calculation was judged by whether or not the top docking score ligand pose reproduced its X-ray ligand pose (heavy atom root-mean-square deviation (RMSD) $\leq 2.0$ Å).

**Determination of Protein Structures for ChEMBL Ligand Docking** For each of the 152 ChEMBL ligands, the most similar X-ray ligand was identified from the 83 PDB ligands (Fig. 1). The TanimotoCombo score from the ROCs program was adopted for similarity evaluation. The score is calculated as the sum of the ShapeTanimoto score for evaluating the shape similarity and the ColorTanimoto score for evaluating the chemical similarity. Ligand conformations were generated by LigPrep and ConfGen for each of the 152 ChEMBL ligands, and the CDK2 protein conformation bound to the X-ray ligand with the highest ChEMBL ligand TanimotoCombo score was used as the target protein structure for the ChEMBL ligand docking calculation. Of the 83 PDB ligand–
CDK2 pairs, 5 CDK2 X-ray crystal structures were used as target protein structures. Of the 152 ChEMBL ligands, 34 ChEMBL ligands were selected which had TanimotoCombo scores greater than 1.2 (ShapeTanimoto >0.6 and ColorTanimoto >0.6). The correspondence between the 5 target protein structures and 34 ChEMBL ligands is shown in Table 1. For example, ChEMBL ligand 10 has the highest similarity to the protein structure from the complex (PDB ID: 1H1P) was used for the docking calculation of ligand 10. The 2D structures and experimental pK values of the 34 ChEMBL ligands are summarized in Tables 2 to 4. Table 2 shows the training set for model generation, and Table 3 shows the validation set for model selection and Table 4 shows the test set to check the predictive ability of the generated model.

**ChEMBL Ligand Docking and Partial Structural Optimization**

Docking calculations with the target protein structure for each of the prepared 34 ChEMBL ligands were performed in the SP mode of the program Glide. Docking score top poses for each of the ChEMBL ligands were extracted. Structural optimization of the docked ligand and amino acid residues within 5 Å around the ligand was performed using a VSGB aqueous solvent model and OPLS2 force field using the MM-GBSA method. Structural optimizations at the same level of theory were also performed on the ligand structures extracted from the complex structures. The program used for the MM-GBSA method was Prime. Obtained docking poses for the ligands against their target protein structures are shown in Fig. 2. It can be seen that all ChEMBL ligands have very good overlap with the PDB ligand. We used only ChEMBL ligands with high similarity to the PDB ligands in the X-ray crystal structure for the docking calculation, and as a result, the docking poses were very similar. The reliability of the obtained complex structures can be considered sufficiently high.

**Calculation of Fragment Binding Energy**

The fragment binding energy is not the interaction energy between fragments but rather the measure of the stability of the fragment itself upon ligand binding. By performing FMO calculations on each of the structures of the complex (PL), protein alone (P), and ligand alone (L), the fragment binding energy for each amino acid residue fragment \( \Delta E_{L}^{bind} (I = 1, 2, 3, \ldots) \) and the fragment binding energy for the ligand fragment \( L, \Delta E_{L}^{bind} \) were calculated (Eq. 1).

\[
\Delta E_{PL}^{bind} = E_{PL} - (E_{P} + E_{L}) = \sum_{I \in P} \Delta E_{I}^{bind} + \Delta E_{L}^{bind} \tag{1}
\]

Here, \( \Delta E_{PL}^{bind} \) is the FMO total binding energy of the system. \( E_{PL}, E_{P} \) and \( E_{L} \) represent the total energy of PL, P and L, respectively. According to the subsystem analysis of FMO, \( \Delta E_{L}^{bind} \) is the FMO total binding energy of the ligand fragment \( \Delta E_{L}^{bind} \) were obtained (Eqs. 2 and 3).

\[
\Delta E_{L}^{bind} = (E_{L}^{PL} - E_{L}^{P}) + \frac{1}{2} \sum_{(j \neq e) \in P} (\Delta E_{inter,PL} - \Delta E_{inter,P}) + \Delta E_{inter}^{PL} \tag{2}
\]

\[
\Delta E_{L}^{bind} = E_{L}^{PL} - E_{L}^{P} \tag{3}
\]

We adopted the formula that the interaction energy between the ligand fragment and the amino acid residue fragments \( \Delta E_{inter}^{PL} \) is entirely owned by the amino acid residue fragments (Eq. 2). Thus, \( \Delta E_{L}^{bind} \) includes deformation energy and desolvation energy of the ligand upon binding and does not include interaction energies with the amino acid residues in the protein. The program used was GAMESS (Apr 20, 2017) and the computational level used was FMO2-RHF/D/PCM<1>/STO-3G. Fragmentation was 1 residue/1 fragment and 1 ligand/1 fragment. We used the program Faciocom to create GAMESS input files. Figure 3 shows the complex structure with ligand 12 used for the FMO calculation as an example. The FMO calculation was performed on all complex structures.

**PLS Regression**

PLS regression is a powerful linear regression method when explanatory variables \( \{x_{i}\}_{i=1, \ldots, m} \) are highly correlated and/or the number of explanatory variables is greater than the number of objective variables \( \{y_{i}\}_{i=1, \ldots, n} \) \((m > n)\).

\[
y \sim Xe \tag{4}
\]

Here, \( y \) is the \( n \times 1 \) matrix of the objective variables, \( X \) is the \( n \times m \) matrix of the explanatory variables and \( e \) is the \( m \times 1 \) matrix of the PLS regression coefficients that we would like to know. In the PLS regression, a set of new linearly independent variables (PLS scores) \( \{t_{i}\}_{i=1, \ldots, d} \) are sequentially determined up to \( d \) components on condition that they have the largest covariance with \( y \) (or its remaining part that was not explained by the previous PLS scores). As a result, it corresponds to the linear regression of the objective variables using the new variables of the PLS scores.

\[
T = XR \tag{5}
\]

\[
y \sim Tq^{T} \tag{6}
\]
Here, $T$ is the $n \times A$ PLS score matrix, $R$ is the $m \times A$ weight matrix and $q^T$ is the $A \times 1$ y-loading matrix. The PLS regression coefficient $c$ for the original explanatory variables can be obtained using the following equation.

$$c = Rq^T = \sum_{j=1}^{A} r_j q_j$$  \hspace{1cm} (8)

PLS regression for predicting the inhibitory activity of the ChEMBL ligands was performed. The objective variables were $pK_i (= -\log K_i)$ and the explanatory variables were the fragment binding energies of ligand fragments and the fragment binding energies of the 45 amino acid residue fragments surrounding the ligands. We used those residues instead of all in order to reduce noise because the residues away from the ligands are not so much important for ligand stability and the structures are less reliable. The 45 amino acid residue frag-  

| ID | Structure | pKi(exp) |
|----|-----------|----------|
| 1  | ![Structure 1](image1.png) | 8.52     |
| 2  | ![Structure 2](image2.png) | 8.70     |
| 3  | ![Structure 3](image3.png) | 6.54     |
| 4  | ![Structure 4](image4.png) | 7.62     |
| 5  | ![Structure 5](image5.png) | 7.00     |
| 6  | ![Structure 6](image6.png) | 7.29     |
| 7  | ![Structure 7](image7.png) | 7.08     |
| 8  | ![Structure 8](image8.png) | 6.10     |
| 9  | ![Structure 9](image9.png) | 7.8      |
| 10 | ![Structure 10](image10.png) | 4.92     |
| 11 | ![Structure 11](image11.png) | 6.77     |
| 12 | ![Structure 12](image12.png) | 8.70     |
| 13 | ![Structure 13](image13.png) | 5.70     |
| 14 | ![Structure 14](image14.png) | 8.15     |
| 15 | ![Structure 15](image15.png) | 5.76     |
| 16 | ![Structure 16](image16.png) | 6.06     |
| 17 | ![Structure 17](image17.png) | 6.22     |
| 18 | ![Structure 18](image18.png) | 6.36     |
| 19 | ![Structure 19](image19.png) | 6.54     |
| 20 | ![Structure 20](image20.png) | 6.81     |

Table 2. The List of IDs, 2D Structures and Experimental pKi(exp) for the Training Set Ligands
ments were defined by the union of amino acid residue fragments that have an inter-fragment distance between a ligand is less than twice the sum of vDW radii. An example of the complex structure used for the FMO calculations and the 45 amino acid residue fragments are shown in Fig. 3.

Of the 34 ChEMBL ligands that had TanimotoCombo scores greater than 1.2 (ShapeTanimoto >0.6 and ColorTanimoto >0.6), 20 were randomly selected as a training set (Table 2) for the PLS model creation, 7 were used as a validation set for choosing the best model (Table 3) and the other 7 were used as a test set (Table 4). The number of PLS components was determined from the point at which the $R^2$ value reached a maximum when the model created from the training set was applied to the validation set. The SIMPLS algorithm27) was used for the PLS regression and the pls package 28) of the program R 29) was used. Standardization of explanatory variables (descriptors) was not performed because differences in the magnitude of fragment binding energy values are meaningful and units are common.

### Results and Discussion

FMO calculations were performed on the 34 reliable complex structures generated from the docking calculations, and fragment binding energies (FBEs) for all fragments were obtained. Figure 4 is a heat map showing the FBEs obtained from the FMO calculations for the 34 ChEMBL ligands. Although FBEs have been obtained for all amino acid residues of CDK2, only those for the 45 amino acid residue fragments surrounding the ligand fragment are shown. FBE value is a measure of the stability of the fragment upon ligand binding. Stabilized fragments upon ligand binding have negative FBE values and destabilized fragments have positive FBE values (refer to Table 5). Some amino acid residue fragments (PHE-82, HIS-84, LEU-134, ASP-135) tended to have large negative FBEs and we would predict that the stabilization of these amino acid residue fragments is essential for the binding of the CDK2 inhibitors. The FBE of the ligand fragment (LIG) is a large positive value. This means that the ligand is destabilized by structural change and dehydration upon binding to the protein. Note that, although the FBE of the ligand fragment is positive, the ligand binding itself is stabilized by the interaction with the surrounding amino acid residues. The vertical axis of the heat map is arranged in order of ligand activity; however, it is difficult to judge which amino acid residues are important for improving the ligand activity from

| ID | Structure | pKᵢ(exp)  |
|----|-----------|-----------|
| 21 |          | 6.60      |
| 22 |          | 7.89      |
| 23 |          | 8.22      |
| 24 |          | 6.49      |
| 25 |          | 5.50      |
| 26 |          | 5.67      |
| 27 |          | 7.00      |

| ID | Structure | pKᵢ(exp)  |
|----|-----------|-----------|
| 28 |          | 7.40      |
| 29 |          | 7.10      |
| 30 |          | 7.70      |
| 31 |          | 5.98      |
| 32 |          | 6.44      |
| 33 |          | 5.20      |
| 34 |          | 6.41      |

Table 3. The List of IDs, 2D Structures and Experimental $pKᵢ$ (exp) for the Validation Set Ligands

Table 4. The List of IDs, 2D Structures and Experimental $pKᵢ$ (exp) for the Test Set Ligands
In Fig. 5, the correlation between FMO total binding energy and experimentally measured $pK_a^{\text{exp}}$ for all 34 ligands is shown. The total binding energy is just the sum of FBEs as shown in Eq. 1. There is the correct trend that highly active ligands with large $pK_a$ values have large negative total binding energies. However, the trend is weak as is often the case when congeneric ligands are included. The simple linear regression using total binding energy was not adequate to predict the experimental activity for our ligands. In order to obtain a good predictive model, we then performed PLS regression using FBEs. The PLS regression also provides information on the important amino acid fragments from its regression coefficients. If the relative trend in the FBEs of a given fragment is correctly calculated among a series of ligands and correlated with the activity of the ligands, the contribution of the fragment in the regression coefficient will be significant.

Figure 6 shows the PLS regression results for the 27 ChEMBL ligands (20 from the training set in Table 2, 7 from the validation set in Table 3). The correlation between the predicted $pK_a$ and the experimental $pK_a$ for the training set and validation set were $R^2_{\text{training}} = 0.84$ and $R^2_{\text{validation}} = 0.86$, respectively. The number of PLS components determined from the validation set was four. Figure 7 shows the results of applying the created PLS model to the 7 test set ligands in Table 4 that were not used in the PLS model creation. As a result, the external prediction performance for the test set was $R^2_{\text{test}} = 0.69$. Although the prediction performance decreased from 0.86 to 0.69; the value is sufficiently high to use the PLS model for further analysis because 70% of the ligand activity is explained.

Since a sufficiently good PLS model was obtained, PLS regression coefficients were analyzed. Figure 8 shows the PLS regression coefficients of the created PLS model (Eq. 8). In the PLS regression, it is expected that the contribution of the explanatory variables correlated with the objective variable will be large, and the contribution of the uncorrelated explanatory variables will be small. Ligand activity equals the product sum of the FBE value and the PLS regression coefficient of each fragment. Thus, the larger the absolute value of the PLS regression coefficient, the greater the contribution of the FBE of the fragment to the predicted $pK_a$ value. That is, the fragments with larger absolute values of the PLS regression coefficient are expected to be important for improving the activity of the CDK2 inhibitors.

Using PROSITE database\textsuperscript{30} to identify important residues we observe that the residues of the ATP binding region (Ile10 to Val18, Lys33, Glu81 to Leu83, Asp86, Lys129 to Leu132, Asp145) tend to have large PLS regression coefficients. Furthermore, seven fragments (GLU-12, GLY-13, PHE-80, ASP-86, LYS-89, LYS-133, ARG-297, LIG) are identified as important fragments with absolute fragment binding energies greater than 0.05. Here, we used different notation for amino acid residues and FMO residue fragments because FMO fragments contain the C=O of the preceding residue and do not include own C=O. The FMO fragments are denoted by three capital letters followed by residue numbers with a hyphen. Four of the seven important fragments are within the ATP binding region. The remaining three are fragments outside the

---

**Fig. 2. Docking Poses for the ChEMBL Ligands Bound to Their Target Protein Structure**

PDB ligands are shown as green sticks. (Color figure can be accessed in the online version.)

**Fig. 3. Example of a Complex Structure Used for FMO Calculations**

Amino acid residues for which fragment binding energies are used in the subsequent PLS regression are shown in red. Ligand 12 is shown as green spheres. (Color figure can be accessed in the online version.)
ATP binding region. These are the important fragments expected to be important for improvement of the activity of the CDK2 inhibitors. The FBE value of the amino acid fragments that are common to all the ligands and not correlated with ligand activity leads to a small PLS regression coefficient. Therefore, in the heat map in Fig. 4, FBEs for several residue fragments (such as PHE-82, HIS-84, LEU-134, ASP-145) were large and negative; however, as expected the PLS regression coefficients of these residues are small. Ligands that bind to CDK2 tend to stabilize these amino acid residue fragments; however, the stabilization of these residues is important just for ligand binding and not important for improving the ligand activity.

The sign of the PLS regression coefficient is also important information. Table 5 shows the relationship of the sign of the FBE value and PLS regression coefficient of a fragment to the ligand activity. When a fragment with negative PLS regression coefficient is stabilized upon ligand binding (FBE value is negative), the ligand activity increases. Also, when a fragment with positive PLS regression coefficient is destabilized upon ligand binding (FBE value is positive), the ligand activity also increases. The other combinations of FBE value and PLS regression coefficient lead to decreased ligand activity. The PLS regression coefficient of the ligand fragment LIG is large and negative. The FBE value of the ligand fragment LIG is always positive because it includes only ligand deformation and desolvation contributions upon binding to the protein. Thus, the larger the ligand destabilization due to the deformation and desolvation upon binding to the protein, the more inactive the ligand is.

There are four important amino acid residue fragments with large negative PLS regression coefficients. GLY-13 has the

![Fig. 4. Heat Map of the Fragment Binding Energies for the 45 Amino Acid Residue Fragments and the Ligand Fragment](image)

The ligands are sorted according to their inhibitory activity: the most active ligand 12 is at the top. The fragment binding energies are in kcal/mol. The members of the validation set ligands are marked by single asterisks. The members of the test set ligands are marked by double asterisks. (Color figure can be accessed in the online version.)

| PLS regression coefficient | Stabilized fragment | Destabilized fragment |
|----------------------------|---------------------|-----------------------|
| -                          | Activity increase   | Activity decrease     |
| +                          | Activity decrease   | Activity increase     |

Ligand activity equals the product sum of FBE value and PLS regression coefficient of each fragment. The combination of stabilized fragment with negative FBE value and negative PLS regression coefficient leads to activity increase.
largest value, followed by LYS-89, ASP-86 and PHE-80. It can be considered that the more stabilized these fragments are (the more negative FBEs are), the more active the ligand is (refer to Table 5). In contrast, there are three important residue fragments which have large positive PLS regression coefficients. Glu-12 has the largest positive value, followed by ARG-297 and LEU-133. It is expected that the more stabilized these fragments are (the more negative FBEs are), the less active the ligand is (Table 5). In order to increase the activity of the ligand, it is better to destabilize these fragments.

Figure 9 shows the seven important amino acid residue fragments for improving ligand activity. The structure is the complex structure of the most active ligand (ligand 12) and CDK2. The four important amino acid residue fragments with large negative PLS regression coefficients are colored blue (GLY-13, LYS-89, ASP-86, PHE-80) and the three important amino acid residue fragments with large positive PLS regression coefficients are colored red (GLU-12, ARG-297, LEU-133). The FBE value of the amino acid residue fragment includes not only the interaction between the ligand but also fragment deformation, fragment desolvation and half of the interaction with the other amino acid fragments; However, roughly speaking, favorable interaction of a ligand with an amino acid residue fragment would stabilize the amino acid residue fragment and lead to a negative FBE value. Thus, a favorable interaction of a ligand with the blue-colored amino acid residue fragments would contribute to increased activity of the ligand (Table 5). Favorable interactions are, for example, electrostatic interactions with ASP-86 and LYS-89 side chains, hydrophobic interactions with GLY-13 and PHE-80 side chains, and hydrogen bonding with main chain carbonyl groups (note that FMO fragments include a carbonyl group from the previous residue). For ligand 12 in Fig. 9, electrostatic interaction with LYS-89 seems to contribute to increased activity of the ligand. Unfavorable interaction of a ligand with an amino acid residue fragment would destabilize the amino acid residue fragment and lead to a positive FBE value. Thus, the unfavorable interaction of a ligand with the red-colored amino acid residue fragments could also contribute to increasing the activity of the ligand (Table 5). The unfavorable interaction is, for example, electrostatic repulsion with GLU-12 and ARG-297 side chains. Paying attention to the decrease in activity, another expression is also possible: the favorable interaction of a ligand with the red-colored amino acid residue fragments could also contribute to a decrease in activity of the ligand (Table 5). Ligand 12 in Fig. 9 has no contacts with the red-colored amino acid residue fragments having neither favorable or unfavorable interaction with these residue fragments. Taking into account the magnitude and the sign of the obtained PLS regression coefficient and the stability of the amino acid residues upon ligand binding, it is possible to identify amino acid residues important for improving the activity of the ligand. As a result, it is possible to study how to optimize the ligand structure in order to improve the stability of these residues by interaction with these residues.

Conclusion
In this study, we present a 3D-QSAR analysis of CDK2...
inhibitors by PLS regression using fragment binding energy obtained by FMO calculations as a descriptor. In addition to the quantitative prediction of the biological activity of the inhibitors, our 3D-QSAR analysis is able to identify amino acid residues important for improving activity. The protein–ligand complex structures for the FMO calculations were prepared by docking calculations, where the protein structures used were carefully determined according to the similarity between the ligand to be docked and the ligand in X-ray complex structures. Residue fragments with large PLS regression coefficients correspond to regions containing important residues. The magnitude and the sign of the obtained PLS regression coefficient are useful information that can be used to improve the ligand activity by designing the ligand while considering the stability of amino acid residues. In order to ascertain whether the knowledge obtained from this study is applicable for improving activity, careful experimental verification and application to other protein–ligand systems are necessary.

In this work, FMO2–RHF/D/PCM<1/>/STO-3 was used as the computational level for the FMO method. Although this is an insufficiently accurate calculation level, it has successfully identified important amino acid residues. By increasing the computational level, it should be possible to improve the accuracy of this method. In this study, we reported a 3D-QSAR analysis of CDK2 inhibitors using FMO calculation and PLS regression. The combination of FMO and the other chemometrics approach is also possible. Such analysis method has already been reported by other groups: FMO and clustering (VISCANA), FMO and self-organizing map, or FMO and multidimensional scale construction methods. Computational drug discovery research using the FMO method is expected to develop more with the improvement of computer speed and development of the theory.

Conflict of Interest The authors declare no conflict of interest.

References
1) Kitaura K., Sawai T., Asada T., Nakano T., Uebayashi M., Chem. Phys. Lett., 312, 319–324 (1999).
2) Kitaura K., Ikeo E., Asada T., Nakano T., Uebayasi M., Chem. Phys. Lett., 313, 701–706 (1999).
3) Fedorov D. G., Kitaura K., J. Phys. Chem. A, 111, 6904–6914 (2007).
4) Fedorov D. G., Kitaura K., J. Phys. Chem. A, 116, 704–719 (2012).
5) Tanaka S., Mochizuki Y., Komeiji Y., Okiyama Y., Fukuzawa K., Phys. Chem. Chem. Phys., 16, 10310–10344 (2014).
6) Wold S., Sjöström M., Eriksson L., Chemom. Intell. Lab. Syst., 58, 109–130 (2001).
7) Cramer R. D. III, Patterson D. E., Bunce J. D., J. Med. Chem., 110, 5959–5967 (1988).
8) Hitaoka S., Matoba H., Harada M., Yoshida T., Tsuji D., Hirokawa T., Itoh K., Chuman H., J. Chem. Inf. Model., 51, 2706–2716 (2011).
9) Mazanetz M. P., Ichihara O., Law R. J., Whittaker M., J. Cheminf., 3, 2 (2011).
10) Otsuka T., Okimoto N., Taiji M., J. Comput. Chem., 36, 2209–2218 (2015).
11) Fedorov D. G., Kitaura K., J. Phys. Chem. A, 120, 2218–2231 (2016).
12) Gaulton A., Bellis L. J., Bento A. P., Chambers J., Davies M., Hersey A., Light V., McLinchey S., Michalovich D., Al-Lazikani B., Overington J. P., Nucleic Acids Res., 40 (D1), D100–D110.
13) Schrödinger Release, 2017-4: LigPrep, Schrödinger, LLC, New York, NY, 2017.
14) Schrödinger Release, 2017-4: ConfGen, Schrödinger, LLC, New York, NY, 2017.
15) Law M. E., Corsino P. E., Narayan S., Law B. K., Mol. Pharmacol., 88, 846–852 (2015).
16) RCSB Protein Data Bank.: <http://www.rcsb.org/pdb/home/home.do>, accessed Aug 3, 2017.
17) Goto J., Kataoka R., Hirayama N., J. Med. Chem., 47, 6804–6811 (2004).
18) Meyder A., Nittinger E., Lange G., Klein R., Rarey M., J. Chem. Inf. Model., 57, 2437–2447 (2017).
19) Vainio M. J. DPICalc, <http://users.abo.fi/mivaino/shaep/download.php>.
20) The Proteins Plus Server: <http://proteinsplus.zbh.uni-hamburg.de>, accessed Sep. 8, 2017.
21) Schrödinger Release, 2017-4: Schrödinger Suite 2017-4 Protein Preparation Wizard, Epik, Schrödinger, LLC, New York, NY, 2017, Impact, Schrödinger, LLC, New York, NY, 2017, Prime, Schrödinger, LLC, New York, NY, 2017.
22) Schrödinger Release, 2017-4: Glide, Schrödinger, LLC, New York, NY, 2017.
23) Hawkins P. C. D., Skillman A. G., Nicholls A., J. Med. Chem., 50, 74–82 (2007).
24) Schrödinger Release, 2017-4: Prime, Schrödinger, LLC, New York, NY, 2017.
25) Schmidt M. W., Baldrige K. K., Boatz J. A., Elbert S. T., Gordon M. S., Jensen J. H., Koseki S., Matsunaga N., Nguyen K. A., Su S., Winds & M., Dupuis M., Montgomery J. A. Jr., J. Comput. Chem., 14, 1347–1363 (1993).
26) Suenaga M., J. Comput. Chem. Jpn., 7, 33–54 (2008).
27) de Jong S., Chemom. Intell. Lab. Syst., 18, 251–263 (1993).
28) Mevik B.-H., Wehrens R., J. Stat. Softw., 18, 1–24 (2007).
29) R Core Team, R: A Language and Environment for Statistical Computing, version 3.2.3, R Foundation for Statistical Computing: Vienna, Austria (2015).
30) Sigrist C. J. A., de Castro E., Cerutti L., Cuche B. A., Hulo N., Bridge A., Bouguieret L., Xenarios I., Nucleic Acids Res., 41, D344–D347 (2013).
31) Amari S., Aizawa M., Zhang J., Fukuzawa K., Mochizuki Y., Iwasawa Y., Nakata K., Chuman H., Nakano T., J. Chem. Inf. Model., 46, 221–230 (2006).
32) Kurauchi R., Watanabe C., Fukuzawa K., Tanaka S., Comput. Theor. Chem., 1061, 12–22 (2015).