Application of singular value decomposition to the inter-fragment interaction energy analysis for ligand screening

Maruyama, Keiya
Sheng, Yinglei
Watanabe, Hirofumi
Fukuzawa, Kaori
Tanaka, Shigenori

(Citation)
Computational and Theoretical Chemistry,1132:23-34

(Issue Date)
2018-05-15

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90004957

Application of singular value decomposition to the inter-fragment interaction energy analysis for ligand screening

Keiya Maruyama[1], Yinglei Sheng[1], Hirofumi Watanabe[2], Kaori Fukuzawa[3], Shigenori Tanaka[1*]

1 Department of Computational Science, Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Hyogo 657-8501, Japan

2 Education Center on Computational Science and Engineering, Kobe University, 7-1-48 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan

3 Department of Physical Chemistry, School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-8501, Japan

**email: tanaka2@kobe-u.ac.jp*

Keywords: p38 MAP kinase; Ligand binding; Fragment Molecular Orbital (FMO) method; Inter-Fragment Interaction Energy (IFIE); Singular Value Decomposition (SVD)

Abstract

We evaluated the binding affinity between p38 MAP kinase and various inhibitors through use of the fragment molecular orbital (FMO) method at MP2/6-31G* level in comparison to experimental values of half maximal inhibitory concentration (IC_{50}). Initially, the calculated results of the FMO-IFIE (inter-fragment interaction energy) sums for 60 complex structures registered in the Protein Data Bank were not well correlated with the IC_{50} activity data. Therefore, we performed the singular value decomposition (SVD) for the calculated results of the IFIE matrix (amino acid residues × various ligands) to improve the correlation and determine the cause of the initial poor results. In SVD, the original matrix is divided into multiple vectors that are orthogonal to each other. Through this method, we improved the correlation by removing some particular vectors that involved noise components and impaired the correlation. In addition, the correlation between the IC_{50} and FMO-IFIE for 22 complex structures of estrogen receptor α (ERα) was also improved in this way. We analyzed the amino acid residues of receptors that were mainly involved in the removed vectors and found an overestimation of the strength of the hydrogen bond between glutamic acid and the ligand.

1. Introduction

P38 mitogen activated protein (MAP) kinase is activated by environmental stressors such as ultraviolet light, proinflammatory cytokines, and pathogens [1-4]. It plays a role in immune responses, inflammation, and the signal transduction pathways leading to cell division and apoptosis [5-7]. Dysregulation of the MAP kinase pathway leads to the development of cancer [8,9] and autoimmune diseases such as rheumatism and asthma.

Kinases have an N-lobe, C-lobe, and hinge region (THR106-ASP112) as well as a DFG-loop structure (ASP168-THR175), as shown in Fig. 1. A ligand binding pocket exists around the hinge region and DFG-loop structure [3,10-12]. The DFG-loop structure changes depending on the state of the protein (active or inert); the DFG-in structure is associated with an active state and the DFG-out structure with an inert state [10].

In the present study, we perform fragment molecular orbital (FMO) calculations for p38 MAP kinase [13-17] with ligand binding for the application to virtual screening. The FMO method [18-20] has been successfully used for drug discovery [21,22] as it is able to describe the electron state of the entire protein-ligand complex by quantum chemical calculation [23,24]. It is thus a useful tool for the accurate evaluation of binding affinity between a receptor protein and its ligand molecules. However, when we compare the evaluated binding energies of ligands with experimental affinity data (half maximal inhibitory concentration, IC_{50}), we often do not find fair correlations between them; some causes for the discrepancy may be attributed to the inaccuracies in employed calculation methods and molecular structures. It would thus be essential for obtaining the desirable correlations to remove the noises associated with these structural and computational inaccuracies through some statistical techniques.

Here, we employ the singular value decomposition (SVD) technique [25-27] to reduce the undesirable effects caused by involved noises. By processing the inter-fragment interaction energy (IFIE) data obtained by FMO calculations for multiple ligands in terms of the SVD, we can improve the agreement with experimental affinity data and thereby the prediction accuracy for ligand binding activity. After the introduction of computational methods, we will illustrate successful examples for p38 MAP kinase and also for estrogen receptor α (ERα) in Appendix A, thus examining the usefulness of novel approach.

2. Computational methods and structure preparation

2.1. Fragment molecular orbital (FMO) method

The FMO method [18-20,23,24] is a computational method that divides large molecules such as proteins into relatively small units called fragments, and then calculates the energy of the whole molecule and the electron density quantum-chemically by molecular orbital (MO) calculations of fragments-alone (monomers) and fragment-pairs (dimers) (sometimes, trimers

and tetramers are also considered). The FMO method is an *ab initio* molecular orbital method [28] and does not use experimental (or empirical) values except for physical constants. By using this method, we can apply an *ab initio* MO method that has been shown to succeed for small compounds to macromolecules such as proteins without a significant loss in accuracy.

There are three main benefits of the FMO method. First, it can greatly reduce computation time because each calculation is independent and the high parallelism effect is realized. Second, the inter-fragment interaction energy (IFIE) can be obtained [29-31]. Third, because it is a quantum chemical calculation, it is possible to accurately verify the force other than the electrostatic interaction, which cannot be accurately described by classical mechanical calculations. In particular, the second benefit is important, thus making the IFIEs essential in this study.

2.2. Inter-fragment interaction energy (IFIE)

When dividing a large molecule into N_f fragments and letting the total electron energies of a fragment and its pair be E_I, E_{IJ}, the total electron energy of a molecule can be approximated (FMO2 approximation) as [18-20,23,24] :

$$\mathrm{E} = \sum_{I>J} E_{IJ} - \left(N_f - 2\right) \sum_I E_I \ . \qquad (1)$$

If ΔP is the difference matrix of the electron density of monomer and dimer, Eq. (1) can be transformed into the following equation:

$$\mathrm{E} = \sum_{I>J} (E'_{IJ} - E'_I - E'_J) + \sum_{I>J} Tr(\Delta \boldsymbol{P}^{IJ} \boldsymbol{V}^{IJ}) + \sum_I E'_I \ , \qquad (2)$$

where $\mathrm{E}'_I = E_I - V_I$, $\mathrm{E}'_{IJ} = E_{IJ} - V_{IJ}$, $\mathrm{V}_I = Tr(\boldsymbol{P}^I \boldsymbol{V}^I)$ and $\mathrm{V}_{IJ} = Tr(\boldsymbol{P}^{IJ} \boldsymbol{V}^{IJ})$; V_I and V_{IJ} are the electrostatic potentials that fragment I and fragment pair IJ receive from surrounding fragments, respectively. Since this formula contains only the electrostatic potential for the dimer, different approximate electrostatic potentials can be used for monomer and dimer. Then,

$$\Delta \mathrm{E}_{IJ} = \left(E'_{IJ} - E'_I - E'_J\right) + Tr(\Delta \boldsymbol{P}^{\boldsymbol{IJ}} \boldsymbol{V}^{\boldsymbol{IJ}}) \ , \qquad (3)$$

where $\Delta \mathrm{E}_{IJ}$ can be interpreted as the interaction energy between the fragment pair IJ. This $\Delta \mathrm{E}_{IJ}$ is referred to as inter-fragment interaction energy (IFIE) [29-31].

In the FMO method, the interaction between a ligand and each amino acid residue of a receptor can be identified as an IFIE. Then, the total value of IFIEs (namely, IFIE-sum) is an index representing the strength of the binding between the ligand and the receptor as a whole.

2.3. Structure preparation

In this study, we used 60 three-dimensional structures of p38 MAP kinase and ligand complexes (see Appendix B) obtained from X-ray crystal structure analysis which are available in the Protein Data Bank (PDB). If there were missing residues in the downloaded PDB data, we complemented these defects by using the “Complement Main Chain” or “Overlay

Molecules" function of BioStation Viewer (MIZUHO / BioStation) [24]. For a template structure with relatively good resolution and without missing residues, we used PDB ID: 3GC7 for DFG-in and PDB ID: 3D83 for DFG-out. We then performed hydrogenation and structure optimization using molecular calculation software MOE [32]. For the structural optimization, we employed Amber10+EHT force field with fixed heavy atoms. We set N-terminal and C-terminal to NH_3^+ and COO^- respectively. Concerning the crystal water molecules, we removed all except those that interact with ASP168, LYS53 or the ligand (Fig. 2). Finally, for the prepared structures, we performed FMO calculations using ABINIT-MP software [24], in which we adopted the MP2 method with the basis function 6-31G* to take into account the electron correlations via the (Moeller-Plesset) second-order perturbation from the Hartree-Fock solution. The use of the 6-31G* basis set gives a good compromise between accuracy and computational cost to perform FMO calculations for many protein-ligand complexes.

2.4. Singular value decomposition (SVD)

An m × n matrix A with m and n-dimensional vectors as rows and columns can be related to an m × n diagonal matrix Σ that satisfies the following equation:

$$\Sigma = U^T A V . \quad (4)$$

Here, U is an m × m orthogonal matrix, and V is an n × n orthogonal matrix. If U and V are chosen appropriately, a pair of matrices can be made with Σ satisfying the condition described later. When it is rewritten, the following equation is satisfied:

$$A = U \Sigma V^T . \quad (5)$$

This type of decomposition is called Singular Value Decomposition (SVD) [25-27, 33]. For simplicity of description, assume $m \geqq n$. Otherwise, we can think of the transposed matrix A^T of A.

If σ_{ij} is an element of Σ, in the case of $i \neq j$, $\sigma_{ij}=0$; in the case of $i = j$, for $1 \leqq i \leqq n$, $\sigma_{ij}=\sigma_i \geqq 0$. At this time, $\sigma_1 \geqq \sigma_2 \geqq \sigma_3 \geqq \ldots$, where σ_i is called a singular value of A, a column vector of U is a left singular vector, and a row of V^T is a right singular vector [33].

For example, in Fig. 3, the left-hand side is the matrix A before decomposition; the right-hand side is U, Σ, and V^T in order from the left side matrix. When the original matrix A is m × n, the shape of each matrix is the following:

A: m × n, U: m × m, Σ: m × n, and V^T: n × n.

The column vector of U is an orthonormal basis for each amino acid residue and the row vector of V^T is an orthonormal basis for each PDB structure with its respective ligands; each singular vector has an independent meaning.

In this paper, we analyze left singular vectors (U) and call them singular vector 1, singular vector 2, ... and so forth in order from the vector of the first column of the left singular vectors.

To execute singular value decompositions, we use numpy.linalg.svd in NumPy (www.numpy.org).

3. Results and discussion

3.1. FMO calculation results for p38 MAP kinase

The average of the calculated IFIEs between each amino acid residue and ligand molecules of 60 structures (see Appendix B) of p38 is shown in Fig. 4. Most of the residues involved in strong binding were charged residues, including GLU71 (-30.74 kcal/mol), ASP112 (-11.29 kcal/mol), and ASP168 (-18.95 kcal/mol).

Here, we examined the statistical correlation between IFIE-sum that is a total IFIE of the contributions from each residue and experimental pIC_{50} (median value in multiple experiments). As shown in Fig. 5, the result is almost uncorrelated (correlation coefficient R^2=0.0129). Moreover, we see that charged ligands have a lower IFIE-sum (i.e., they are more stable) and have a tendency to gather in the lower part indicated by a red circle in the figure. Therefore, we decided to consider the correlation separately depending on the presence or absence of electric charge in the ligand. However, as seen in Fig. 6, we still could not obtain good correlations for both charged ligands and uncharged ligands. Here, the correlation (R^2=0.2399) observed in Fig. 6(a) is irrelevant, because the IFIE-sum should be more negative with the increase of pIC_{50}.

We further separated the data for the uncharged ligands in Fig. 6(b) based on the structure of the DFG-loop (in or out). As a result, although we could not obtain good correlation with the DFG-out structure, the DFG-in structure paired with uncharged ligands showed a correlation, R^2=0.2438 (Fig. 7).

From these results and excluding a limited number of cases, we find that the IFIE-sum obtained by the present FMO calculation and pIC_{50} do not have a reasonable correlation. However, as shown in Fig. 7(a), the IFIE-sum for the DFG-in structure paired with uncharged ligands is slightly correlated with experimental IC_{50} data. In order to improve the overall correlation, it is necessary to correct the IFIE values for the DFG-out structure and/or the charged ligands, thus successfully extracting the positive correlation such as in Fig. 7(a). Therefore, we decided to perform the singular value decomposition for the calculated results of the IFIE matrix (amino acid residues × various ligands) to improve the correlation.

In the following, we will describe in detail the correction of the IFIE values by singular value decomposition (SVD) and the improvement of the correlation. Furthermore, we will discuss in subsection 3.2.3 the difference in IFIE due to the DFG-loop structure. Hereafter, we will analyze left singular vectors [33] and call them singular vector 1, singular vector 2, etc. in order from the vector of the first column of left singular vectors. Likewise, singular values

having a high contribution rate will be called singular value 1, singular value 2, etc. in order. (Concerning the contribution rate of each singular vector, see Appendix C.)

3.2. Singular value decomposition (SVD) for IFIEs of p38 MAP kinase

3.2.1. SVD for 60 complex structures

As described in Sec. 2.4, in singular value decomposition (SVD) the matrix is decomposed into a combination of relevant matrices. Here, when one vector component contains substantial noise, it is possible to eliminate noise components by removing these corresponding vectors. By using this approach, we expect to improve the correlation between calculated IFIE values and experimental activity values. As a method of vector removal, after performing the SVD, we transformed the singular values that we wanted to remove into zero and then reverted the matrix before the decomposition.

We applied SVD to all 60 complex structures and considered the resulting left singular vectors. Peak residues for each singular vector are shown in Fig. 8. Singular vector 1 (Fig. 8(a)) is very similar to the IFIE average (Fig. 4), which can thus be considered to represent the IFIE average itself. However, singular vector 2 (Fig. 8(b)) was significantly different from the IFIE average. The reason why GLU71 and PHE169 were peaked in the singular vector 2 is discussed in the next subsection. Singular vector 3 (Fig. 8(c)) was also different from the IFIE average.

In order to identify noisy vectors to remove, we examined the correlation between the IFIE-sum that depends on each one singular vector and the pIC_{50}. However, since several top singular vectors have no characteristic correlation with pIC_{50} (Fig. 9), we could not identify the noisy vectors at this stage.

3.2.2. SVD for 45 complex structures with uncharged (neutral) ligand

Since the IFIE varies greatly depending on the presence or absence of ligand charge and it is difficult to handle all the data for charged and uncharged ligands on the equal footing, we decided to perform SVD for only the 45 complex structures that contained a neutral ligand.

The correlation of IFIE-sum with pIC_{50} is shown in Fig. 6. By selecting only the uncharged ligand data, the distribution of IFIE changed as shown in Table 1. In contrast to the IFIE values of ASP168 and ASP112, which showed much less attractive interactions, we did not observe significant changes in the IFIE data for GLU71 or PHE169, which we had previously identified as peak residues of singular vector 2 in subsection 3.2.1 (Fig. 8(b)).

Next, we examined each left singular vector. As well as SVD for all 60 structures, singular vector 1 was again similar to the average value of IFIE itself, while singular vector 2 had completely different peak residues (Fig. 10). Further, we can see that singular vector 1 from the uncharged data (Fig. 10(a)) has the same peak residues as singular vector 2 from all of the data

(Fig. 8(b)). From these results, we thought that singular vector 2 from all of the data included a significantly important factor for binding that is not substantially affected by ligand charge. Interestingly, we show in Table 1 that GLU71 and PHE169 are not affected by the presence or absence of ligand charge.

Here, as in the case of the total data, we examined the correlation of each IFIE-sum with each singular vector and pIC_{50} to look for specific vectors containing noise components (Fig. 11). This revealed a correlation coefficient of R^2=0.3176 for singular vector 2. On the contrary, singular vector 1 showed an inverse correlation, R^2=0.3298. We then thought that we may be able to take out the correlation included in FMO calculation such as singular vector 2 by removing singular vector 1.

As seen in Fig. 12, while the original R^2 was 0.0448 (Fig. 6(b)), we were able to improve it to 0.3715 by removing the contribution from singular vector 1. From this result, we can say that the correlation was improved by reducing the contribution of GLU71 because the maximum peak residue of removed singular vector 1 is GLU71. To confirm this issue, we should consider the interaction between GLU71 and the ligand in more detail. Furthermore, in the next subsection, we also investigate the relation between this result and Fig. 7(a).

3.2.3. Hydrogen bonds between GLU71 and the ligand

In Fig. 13, we can see that GLU71 and the ligand are strongly connected through hydrogen bonds [35]. However, there is a great difference in the strength of the hydrogen bonds depending on the kind of ligands, as seen in Tables 2 and 3.

From Table 2 and 3, we can see that IFIE between GLU71 and the ligand tends to be inversely correlated with the pIC_{50} and this is also reflected in the singular vector 1 (Fig. 11(a)), which has GLU71 as a peak residue. This tendency is due to the fact that the magnitude of IFIE between GLU 71 and the ligand is larger in the DFG-out structures than in the DFG-in. The IFIE between GLU71 and the ligand naturally affected IFIE-sum, and we think that this is the reason why we could see no correlation in the DFG-out structures (Fig. 7(b)).

From these results, especially for the DFG-out structures, we supposed that some data may overestimate the IFIE value of the hydrogen bonds between GLU71 and the ligand in the FMO calculation, partially because of the inaccuracies in employed molecular structures. Thus, by removing the singular vector 1, the IFIE between GLU71 and the ligand could be corrected and we could extract the correlation as in Fig. 7(a).

3.2.4. SVD for DFG-out structures with uncharged ligand

Following on from the identification of GLU71 as a strong contributor to the correlation between pIC_{50} and IFIE-sum, we tried to improve the correlation for the data set comprised of

only the DFG-out structures with uncharged ligands, which was originally nearly uncorrelated (Fig. 7(b)). As shown in Fig. 14, the peak residues of singular vectors were almost the same as those (Fig. 10) from the data set of 45 structures described in subsection 3.2.2. Table 4 shows top 10 strongly attractive residues of average IFIE values in the case of DFG-out structure and uncharged ligands.

As shown in Fig. 15(a), we can see that singular vector 1 has a stronger inverse correlation with pIC_{50} than in Fig. 11(a). Figure 15(b) then shows the result of removing the singular vector 1, which succeeds in improving the correlation from R^2 values of 0.0171 to 0.5746. This result supports our consideration in subsection 3.2.3 above, thus illustrating that the inaccuracies associated with GLU71 in ligand-binding calculations significantly impaired the correlation.

In addition, in order to confirm the usefulness of SVD, we also applied it to estrogen receptor (ER) α [36] and succeeded in improving its correlation through use of the same procedure as for p38 MAP kinase (see Appendix A for details).

4. Concluding remarks

We performed FMO calculations for p38 MAP kinase and attempted to reproduce or predict the ligand-binding activity, but were unable to obtain expected correlation between the calculated IFIE-sum and experimental pIC_{50}. However, by removing the noise components using the singular value decomposition (SVD), we succeeded in improving the correlation between the calculated values for p38 and the experimental activity values. The same method was also effective for estrogen receptor (ER) α. Empirically, we found that singular vector 1 in the SVD substantially reflects the average value of IFIE itself. If we use data from charged ligands, singular vector 2 extracts the key residues that are less susceptible to ligand charge. Glutamic acid is representative of such residues and we found that the binding energies of hydrogen bonds between glutamic acid and the ligand were sometimes overestimated in the present FMO calculations, thus identifying one of the sources of noise. One of the reasons that hydrogen bonds are reflected strongly in the calculated results is that we performed the energy calculations *in vacuo*. To reduce this, we could perform FMO calculations that take into account solvent effects [37,38] or scale down evaluation of the hydrogen bond, which remain to be investigated in the future study. The present approach in which the FMO and SVD techniques are combined is expected to provide a useful tool for ligand screening, which may be regarded as a complement to database-driven molecular fingerprint approaches such as structural interaction fingerprint [39] and substructure profiles [40].

Acknowledgements

This study was conducted as one of the activities of the “FMO Drug Design Consortium”, under the task of K computer project “hp150160” and “hp160103”. We thank the members of the FMO Drug Design Consortium, Satoshi Anzaki (Kobe University), Atsushi Inoue (Eisai Co., Ltd.), Yoshitaka Maeda (Mochida Pharmaceutical Co., Ltd.), Fumio Nakajima (Carna Biosciences, Inc.), and Chiduru Watanabe (RIKEN) for providing the IFIE data. S.T. would like to acknowledge the Grants-in-Aid for Scientific Research (Nos. 26460035 and 17H06353) from the Ministry of Education, Cultute, Sports, Science and Technology (MEXT). K.F. acknowledges JSPS KAKENHI Grant Number JP15K05397. The computations were partially performed using the π-computer at Kobe University.

Appendix

A. SVD for ERα

The female hormone, estrogen, is recognized by the estrogen receptor (ER), of which there are two types, α and β [41,42]. For breast cancer cells to proliferate, it is necessary that the ER binds with estrogen and is activated [36]. The structure of ERα around a ligand is shown in Fig. A1. There are two types of ligands for ERα, agonists and antagonists [41-43], as shown in Fig. A2. There is great difference of the ligand structure between these two groups, with antagonistic ligands usually carrying +1 charge.

Structure preparation for the FMO calculation was performed as described earlier for p38 (Sec. 2.3 in the main text). For the template structure, we used PDB ID: 2YJA for agonists and PDB ID: 3ERT for antagonists. We left only a crystal water interacting with GLU353 and ARG394 (Fig. A3).

When we performed SVD for the IFIEs of p38 in the main text, we used data procured from the use of uncharged ligands. In this time for ERα, we neutralized the antagonists and performed the SVD on all 22 data (see Appendix B) employed for the FMO calculations, comprising 11 agonists and 11 neutralized antagonists.

The average IFIE values and the correlation between the IFIE-sum and pIC_{50} are shown in Fig. A4 and Fig. A5, respectively. The IFIE-sum and pIC_{50} were virtually uncorrelated (R^2=0.0057), as seen in Fig. A5.

The peak residues of singular vectors 1 and 2 following the SVD are depicted in Fig. A6. Similar to the case of p38, the peak figure of singular vector 1 (Fig. A6(a)) corresponded closely with the figure of the average value of IFIE (Fig. A4), while singular vector 2 did not.

The correlations between the IFIE-sum for each of singular vector 1, 2, and the pIC_{50} are shown in Fig. A7. Whereas the IFIE for singular vector 1 is correlated inversely with pIC_{50}

(R^2=0.112), that for singular vector 2 and pIC_{50} showed a relevant correlation with R^2=0.1962, again similar to the case of p38.

To improve these correlations via the same procedure as for p38, we removed the contribution from singular vector 1 and confirmed the correlation with pIC_{50}. As seen in Fig. A8, the correlation was improved to a certain extent, with R^2 improving from 0.0057 to 0.2973.

Figure A6(a) indicates that the most important residue in singular vector 1 is GLU353. Interestingly, as in the case of p38, when we use only uncharged ligand data, the peak residue is glutamic acid. Glutamic acid is likely not affected by ligand charge, as compared to other charged residues such as aspartic acid.

GLU353 is hydrogen bonded with the ligand as shown in Fig. A9. That we could improve the correlation by mitigating the effects of GLU353 was similar to what we saw with the SVD for p38. Therefore, there is a possibility that the strength of the hydrogen bonds with the ligand was overestimated, as addressed in the main text.

B. PDB codes employed for the calculations

The Protein Data Bank (PDB) codes for the complexes of p38 and ligand compounds used in the FMO calculations are as follows:
1bmk, 1m7q, 1zz2, 1zz2 (fragmented ligand), 1zzl, 2baj, 2bak, 2gfs, 2rg5, 2zb0, 2zb1, 3c5u, 3d7z, 3d83, 3ds6, 3flq, 3fls, 3fly, 3fmh, 3fmk, 3fml, 3fmm, 3gc7, 3gcq, 3gcu, 3gcv, 3gfe, 3hec, 3heg, 3hll, 3hp5, 3hv3, 3hv4, 3hv6, 3hv7, 3iph, 3itz, 3mpt, 3mvl, 3mw1, 3nnv, 3nww, 3o8p, 3o8t, 3obj, 3oc1, 3ocg, 3pg3, 3que, 3roc, 3s4q, 3u8w, 3zsg, 3zsh, 3zsi, 3zya, 4aa5, 4aac, 4kin, 4kip.

The PDB codes for the complexes of ERα and ligand compounds used in the FMO calculations are as follows:
1ere, 1err, 1gwr, 1sj0, 1x7e, 1x7r, 1xp1, 1xp6, 1xp9, 1xpc, 1yim, 1yin, 2iog, 2ouz, 2qa6, 2qzo, 2yja, 3erd, 3ert, 3hm1, 3q95, 3uu7.

C. Contribution rate of each singular vector in SVD

In SVD, the significance of each singular vector is determined by the magnitude of the corresponding singular value. The ratio of each singular value to the sum of all the singular values is called the contribution ratio or rate. The contribution ratio of the top 10 singular vectors for each of 4 data sets is shown in Table A1.

From Table A1, we can see that in the case of p38 (all data) and p38 (uncharged ligand data), the contribution ratio of singular vector 1 is about 30% and that of singular vector 2 is about 13%, whereas in p38 (DFG-out and uncharged ligand data) and ERα, the singular vector 1

displays a contribution ratio of 50 % or more. Table A2 shows the average ratio of the IFIE-sum of the top 3 residues with a large negative IFIE value against the IFIE-sum of all the residues.

From Table A2, we can see that in the case of p38 (DFG-out and uncharged ligand data) and ERα, the ratio of the top 3 residues to the IFIE-sum is larger than others. We can consider this as one of the reasons why the contribution ratio of singular vector 1 is large.

References

[1] J. Saklatvala, The p38 MAP kinase pathway as a therapeutic target inflammatory disease, *Curr. Opin. Pharmacol.* 4 (2004) 372-377.

[2] A. Cuenda, S. Rousseau, P38 MAP-kinase pathway regulation, function and role in human diseases, *Biochim. Biophys. Acta – Mol. Cell Res.* 1773 (2007) 1358-1375.

[3] K.P. Wilson, P.G. McCaffrey, K. Hsiao, S. Pazhanisamy, V. Galullo, G.W. Bemis, M.J. Fitzgibbon, P.R. Caron, M.A. Murcko, M.S.S. Su (1997). The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase, *Chem. Biol.* 4 (1997) 423-431.

[4] R. Winzen, M. Kracht, B. Ritter, A. Wilhelm, C.Y.A. Chen, A.B. Shyu, M. Muller, M. Gaestel, K. Resch, H. Holtman (1999). The p38 MAP kinase pathway signals for cytokine induced mRNA stabilization via MAP kinase activated protein kinase 2 and an AU-rich region targeted mechanism, *EMBO J.* 18 (1999) 4969-4980.

[5] M. Qi, E.A. Elion, MAP kinase pathways, *J. Cell Sci.* 118 (2005) 3569-3572.

[6] Z. Xia, M. Dickens, J. Raingeaud, R.J. Davis, M.E. Greenberg, Opposing effects of ERK and JNK-p38 MAP kinase on apoptosis, *Science* 270 (1995) 1326.

[7] B. Pucci, M. Kasten, A. Giordano, Cell cycle and apoptosis, *Neoplasia* 2 (2000) 291-299.

[8] A.S. Dhillon, S. Hagan, O. Rath, W. Kolch, MAP kinase signaling pathways in cancer, *Oncogene* 26 (2007) 3279-3290.

[9] J.M. Olson, A.R. Hallahan, P38 MAP kinase: a convergence point in cancer therapy, *Trends Mol. Med.* 10 (2004) 125-129.

[10] Y. Liu, N.S. Gray, Rational design of inhibitors that bind to inactive kinase conformations, *Nature Chem. Biol.* 2 (2006) 358-364.

[11] M. Sawa, Strategies for the design of selective protein kinase inhibitors, *Mini Rev. Med. Chem.* 8 (2008) 1291-1297.

[12] L. Tong, S. Pav, D.M. White, S. Rogers, K.M. Crane, C.L. Cywin, M.L. Brown, C.A. Pargellis, A highly specific inhibitor of human p38 MAP kinase binds in the ATP pocket, *Nature Struct. Biol.* 4 (1997) 311-316.

[13] J. Raingeaud, S. Gupta, J.S. Rogers, M. Dickens, J. Han, R.J. Ulevitch, R.J. Davis, Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine, *J. Biol. Chem.* 270 (1995) 7420-7426.

[14] T. Zarubin, J. Han, Activation and signaling of the p38 MAP kinase pathway, *Cell Res.* 15 (2005) 11-18.

[15] B. Derijard, J. Raingeaud, T. Barrett, I.H. Wu, J. Han, R.J. Ulevitch, R.J. Davis, Independent human MAP-kinase signal transduction pathways define by MEK and MKK isoforms, *Science* 267 (1995) 682-685.
[16] C. Pargellis, L. Tong, L. Churchill, P.F. Cirillo, T. Gilmore, A.G. Graham, P.M. Grob, E.R. Hickey, N. Moss, S. Pav, J. Regan, Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site, *Nature Struct. Biol.* 9 (2002) 268-272.
[17] J. Guay, H. Lambert, G. Gingras-Breton, J.N. Lavoie, J. Huot, J. Landry (1997). Regulation of actin filament dynamics by p38 map kinase mediated phosphorylation of heat shock protein 27, *J. Cell Sci.* 110 (1997) 357-368.
[18] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayashi, Fragment molecular orbital method: An approximate computational method for large molecules, *Chem. Phys. Lett.* 313 (1999) 701-706.
[19] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayashi, K. Kitaura, Fragment molecular orbital method: Application to polypeptides, *Chem. Phys. Lett.* 318 (2000) 614-618.
[20] T. Nakano, T. Kaminuma, T. Sato, K. Fukuzawa, Y. Akiyama, M. Uebayashi, K. Kitaura, Fragment molecular orbital method: Use of approximate electrostatic potential, *Chem. Phys. Lett.* 351 (2002) 475-480.
[21] S. Amari, M. Aizawa, J. Zhang, K. Fukuzawa, Y. Mochizuki, Y. Iwasawa, K. Nakata, H. Chuman, T. Nakano, 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.* 46 (2006) 221–230.
[22] R. Kurauchi, C. Watanabe, K. Fukuzawa, S. Tanaka, Novel type of virtual ligand screening on the basis of quantum-chemical calculations for protein-ligand complexes and extended clustering techniques, *Comput. Theor. Chem.* 1061 (2015) 12-22.
[23] D.G. Fedorov, T. Nagata, K. Kitaura, Exploring chemistry with the fragment molecular orbital method, *Phys. Chem. Chem. Phys.* 14 (2012) 7562-7577.
[24] S. Tanaka, Y. Mochizuki, Y. Komeji, Y. Okiyama, K. Fukuzawa, Electron-correlated fragment-molecular-orbital calculations for biomolecular and nano systems, *Phys. Chem. Chem. Phys.* 16 (2014) 10310-10344.
[25] V. Klema, A.Laud, The singular value decomposition: Its computation and some applications, *IEEE Trans. Automat. Contr.* 25 (1980) 164-176.
[26] F. Kleibergen, R. Paap, Generalized reduced rank tests using the singular value decomposition, *J. Economet.* 133 (2006) 97-126.
[27] M. Brand, Fast low-rank modification of the thin singular value decomposition, *Linear Algebra Appl.* 415 (2006) 20-30.

[28] A. Szabo, N.S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (McGraw-Hill, New York, 1989).
[29] K. Fukuzawa, Y. Mochizuki, S. Tanaka, K. Kitaura, T. Nakano, Molecular interactions between estrogen receptor and its ligand studied by the ab initio fragment molecular orbital method, *J. Phys. Chem. B* 110 (2006) 16102-16110.
[30] I. Kurisaki, K. Fukuzawa, Y. Komeji, Y. Mochizuki, T. Nakano, J. Imada, A. Chmielewski, S.M. Rothstein, H. Watanabe, S. Tanaka, Visualization analysis of inter-fragment interaction energies of CRP-cAMP-DNA complex based on the fragment molecular orbital method, *Biophys. Chem.* 130 (2007) 1-9.
[31] Y. Okiyama, K. Fukuzawa, H. Yamada, Y. Mochizuki, T. Nakano, S. Tanaka, Counterpoise-corrected interaction energy analysis based on the fragment molecular orbital scheme, *Chem. Phys. Lett.* 509 (2011) 67-71.
[32] N. Nirmal, G.O. Praba, D. Velmurugan, Modeling studies on phospholipase A_2-inhibitor complexes, *Ind. J. Biochem. Biophys.* 45 (2008) 256-262.
[33] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, *SIAM J. Mat. Anal.* 21 (2000) 1253-1278.
[34] M.J. Greenacre, *Theory and Applications of Correspondence Analysis* (Academic Press, London, 1984).
[35] T. Steiner, The hydrogen bond in the solid state, *Angew. Chem. Intern. Ed.* 41 (2002) 48-76.
[36] N. Kondo, T. Toyama, H. Sugiura, Y. Fujii, H. Yamashita, MiR-206 expression is down-regulated in estrogen receptor α-positive human breast cancer, *Cancer Res.* 68 (2008) 5004-5008.
[37] S. Miertus, E. Scrocco, J. Tomasi, Electrostatic interaction of solute with a continuum: A utilization of ab initio molecular potentials for the prevision of solvent effects, *Chem. Phys.* 55 (1981) 117-129.
[38] H. Watanabe, Y. Okiyama, T. Nakano, S. Tanaka, Incorporation of solvation effects into the fragment molecular orbital calculations with the Poisson-Boltzmann equation, *Chem. Phys. Lett.* 500 (2010) 116-119.
[39] Z. Deng, C. Chuaqui, J. Singh, Structural interaction fingerprint (SIFt): A novel method for analyzing three-dimensional protein-ligand binding interactions, *J. Med. Chem.* 47 (2004) 337-344.
[40] W.K. Yeo, M.L. Go, S. Nilar, Extraction and validation of substructure profiles for enriching compound libraries, *J. Comput. Aided Mol. Des.* 26 (2012) 1127-1141.

[41] T. Barkhem, B. Carlsson, Y. Nilsson, E. Enmark, J.-A. Gustafsson, S. Nilsson, Differential response of estrogen receptor α and estrogen receptor β to partial estrogen agonists / antagonists, *Mol. Pharmacol.* 54 (1998) 105-112.
[42] G.G.J.M. Kuiper, B. Carlsson, K. Grandien, E. Enmark, J. Haggblad, S. Nilsson, J.-A. Gustafsson, Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors α and β, *Endocrinol.* 138 (1997) 863-870.
[43] A.L. Wijayaratne, D.P. McDonnell, The human estrogen receptor-α is a ubiquitinated protein whose stability is affected differentially by agonists, antagonists, and selective estrogen receptor modulators, *J. Biol. Chem.* 276 (2001) 35684-35692.

Figure legends

Figure 1. Structure of p38 MAP kinase around a ligand (B45; $C_{28}H_{27}Cl_3FN_3O$) shown by sphere representation. Two structures of activation loop, DFG-in and DFG-out, are denoted by pink and cyan lines, respectively, where only the loop part of DFG-out (PDB ID: 3D83) is superimposed on the whole structure of the DFG-in (PDB ID: 3GC7). The hinge region (green) and the gate keeper (yellow) are also depicted, where the gate keeper is a part (THR106) of the hinge region.

Figure 2. Crystal water near a ligand ($C_{25}H_{26}N_4O_2$; sphere representation) and two residues, Lys53 and Asp168, of p38 MAP kinase (PDB ID: 3NNV).

Figure 3. Example of singular value decomposition (SVD) [34].

Figure 4. Average IFIEs between each amino acid residue and ligand molecules of 60 calculated structures of p38. Several important residues that show strong interactions with ligands are highlighted.

Figure 5. Correlation between pIC_{50} (median value obtained from multiple experiments) and calculated IFIE-sum for all 60 data. The results for charged ligands are highlighted.

Figure 6. Correlation between pIC_{50} (Median) and IFIE-sum for (a) charged ligands and (b) uncharged (neutral) ligands.

Figure 7. Correlation between pIC_{50} (Median) and IFIE-sum for the uncharged ligands. (a) DFG-in structure. (b) DFG-out structure. (c) All uncharged ligand data (that is, Fig. 6(b)). Orange points represent DFG-in data and blue points represent DFG-out data.

Figure 8. Peak residues (highlighted) of each singular vector for all 60 complexes. (a) Singular vector 1. (b) Singular vector 2. (c) Singular vector 3.

Figure 9. Correlation between the IFIE-sum for each singular vector and pIC_{50} (Median). (a) Singular vector 1. (b) Singular vector 2. (c) Singular vector 3.

Figure 10. Peak residues (highlighted) of singular vectors in the case of uncharged ligands. (a) Singular vector 1. (b) Singular vector 2.

Figure 11. Correlation between the IFIE-sum for each singular vector and pIC_{50} (Median) in the case of uncharged ligands. (a) Singular vector 1. (b) Singular vector 2.

Figure 12. Correlation between IFIE-sum and pIC_{50} (Median) after removing the contribution from singular vector 1 in the case of uncharged ligands.

Figure 13. Various examples for hydrogen bonds (green line) between GLU71 and the ligand (PDB ID: 3O8T, 3HEG, 3GFE, and 3GCQ) along with the bond length in units of Å. Corresponding values of experimental pIC_{50} and calculated GLU71-ligand IFIE are also shown.

Figure 14. Peak residues (highlighted) of singular vectors in the case of DFG-out structure with uncharged ligands. (a) Singular vector 1. (b) Singular vector 2.

Figure 15. Correlation between the IFIE-sum for each singular vector and pIC_{50} (Median) in the case of DFG-out structure with uncharged ligands. (a) Singular vector 1. (b) All vectors except singular vector 1.

Figure A1. Amino acid residues of ERα around a ligand ($C_{18}H_{24}O_2$). The structure was retrieved from PDB entry of 2YJA.

Figure A2. Example of ligand structures of agonist and antagonist for ERα with their PDB entries.

Figure A3. Crystal water near ligand ($C_{18}H_{24}O_2$) binding pocket of ERα (PDB ID: 2YJA).

Figure A4. Average IFIE values between residues and ligands for ERα.

Figure A5. Correlation between IFIE-sum and pIC_{50} (Median) for ligand binding of ERα.

Figure A6. Peak residues (highlighted) of singular vectors in the case of ERα. (a) Singular vector 1. (b) Singular vector 2.

Figure A7. Correlation between the IFIE-sum for each singular vector and pIC_{50} (Median) for ERα. (a) Singular vector 1. (b) Singular vector 2.

Figure A8. Correlation between the IFIE-sum and pIC_{50} (Median) obtained by removing the contribution from singular vector 1 in the case of ERα.

Figure A9. Hydrogen bonds between GLU353 and the ligand of ERα (PDB ID: 1X7E) along with the bond length in units of Å.

Tables

Table 1. Top 10 attractive IFIE (residue-ligand) average values for both the total and uncharged-ligand datasets.

(a) All 60 data.

Residue	IFIE average	Rank
GLU71	−30.74	1
ASP168	−18.95	2
ASP112	−11.29	3
ASP150	−7.69	4
LEU108	−7.20	5
GLY110	−7.06	6
PHE169	−6.88	7
MET109	−5.46	8
GLU178	−5.26	9
THR106	−4.92	10

(b) 45 data with uncharged ligands.

Residue	IFIE average	Rank
GLU71	−24.46	1
GLY110	−7.40	2
PHE169	−7.35	3
MET109	−6.99	4
LEU108	−5.96	5
ASP168	−5.10	6
THR106	−4.74	7
VAL52	−4.01	8
ALA111	−3.80	9
LEU167	−3.76	10

Table 2. Comparison between the IFIE of GLU71-ligand and pIC_{50} (Median) along with the type of structure of DFG loop.

PDB ID	3O8T	3GCQ	3HEG	3GFE
IFIE[kcal/mol]	−71.31	−63.39	−66.38	−24.21
pIC50(Median)[M]	4.7	6.0	6.7	8.5
DFG-loop	out	out	out	in

Table 3. Average values of pIC_{50} (Median), IFIE-sum (uncharged ligand), and GLU71-ligand IFIE (charged / uncharged) for DFG-in and DFG-out structures.

DFG-loop	In	Out
pIC50(Median)[M]	7.69	6.28
IFIE-sum(uncharged ligands)[kcal/mol]	−101.05	−127.79
GLU71(charged ligands)[kcal/mol]	−41.66	−71.39
GLU71(uncharged ligands)[kcal/mol]	−10.77	−50.32

Table 4. Top 10 residues of average IFIE (residue-ligand) values in the case of DFG-out structure and uncharged ligands.

Residue	IFIE average	Rank
GLU71	−50.32	1
PHE169	−14.73	2
ASP168	−11.71	3
ILE84	−4.65	4
MET109	−4.51	5
LEU108	−4.36	6
LEU167	−3.79	7
GLY110	−3.52	8
THR106	−3.11	9
VAL38	−2.69	10

Table A1. Contribution ratio in each SVD.

(a) p38 MAP kinase (all data).

Vector number	Contribution rate
1	0.32
2	0.13
3	0.08
4	0.06
5	0.05
6	0.04
7	0.03
8	0.03
9	0.03
10	0.02

(b) p38 MAP kinase (uncharged ligand data).

Vector number	Contribution rate
1	0.29
2	0.13
3	0.08
4	0.05
5	0.05
6	0.04
7	0.03
8	0.03
9	0.03
10	0.03

(c) p38 MAP kinase (DFG-out and uncharged ligand data).

Vector number	Contribution rate
1	0.52
2	0.10
3	0.08
4	0.06
5	0.04
6	0.04
7	0.03
8	0.02
9	0.02
10	0.02

(d) ERα.

Vector number	Contribution rate
1	0.54
2	0.10
3	0.08
4	0.05
5	0.03
6	0.03
7	0.03
8	0.02
9	0.02
10	0.02

Table A2. Average ratio of the IFIE-sum of the top 3 residues against the total IFIE-sum for the 4 cases shown in Table A1.

p38 (all)	p38 (uncharged)	p38 (DFG-out)	ER α
0.42	0.35	0.60	0.55

Figures

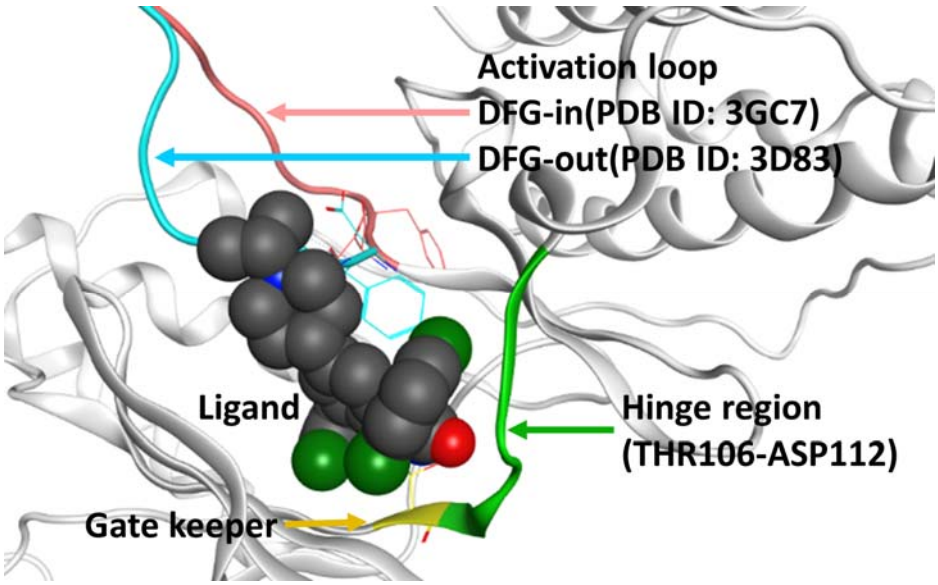

Figure 1

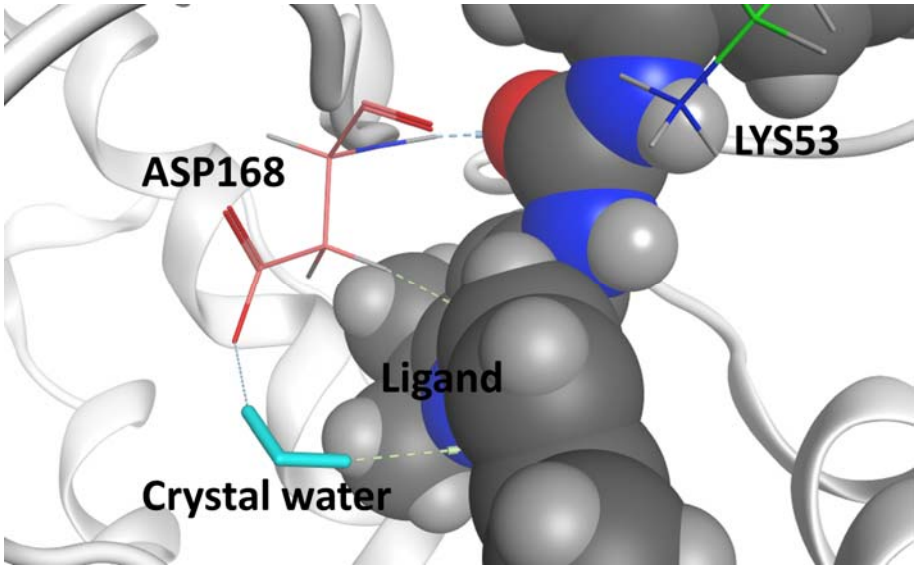

Figure 2

$$\begin{bmatrix} 1 & 2 & 3 \\ 8 & 9 & 7 \\ 10 & 11 & 12 \\ 6 & 4 & 5 \end{bmatrix} = \begin{bmatrix} -0.14 & 0.62 & 0.05 & * \\ -0.55 & -0.54 & 0.58 & * \\ -0.75 & 0.44 & -0.06 & * \\ -0.34 & -0.37 & -0.81 & * \end{bmatrix} \begin{bmatrix} 25.35 & 0 & 0 \\ 0 & 2.15 & 0 \\ 0 & 0 & 1.71 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.56 & -0.59 & -0.59 \\ -0.68 & -0.09 & 0.73 \\ -0.48 & 0.81 & -0.35 \end{bmatrix}$$

Figure 3

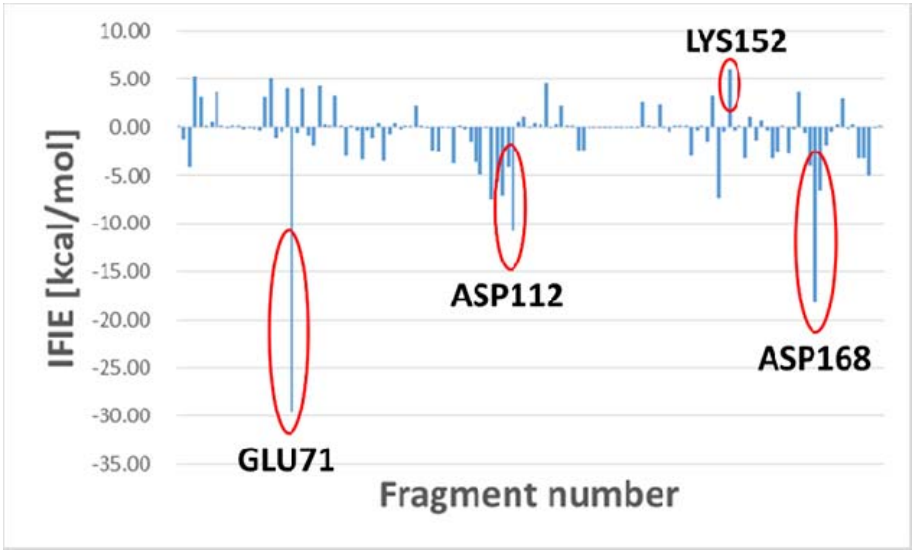

Figure 4

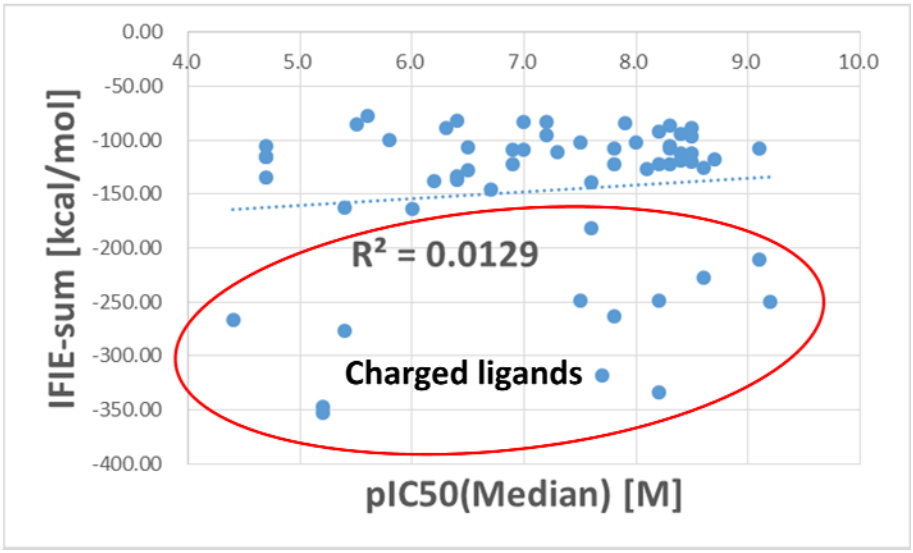

Figure 5

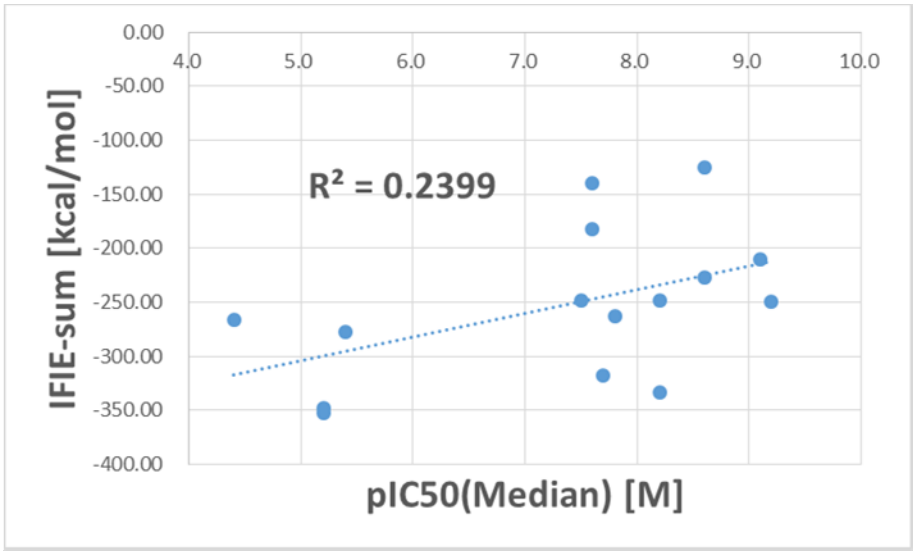

Figure 6 (a)

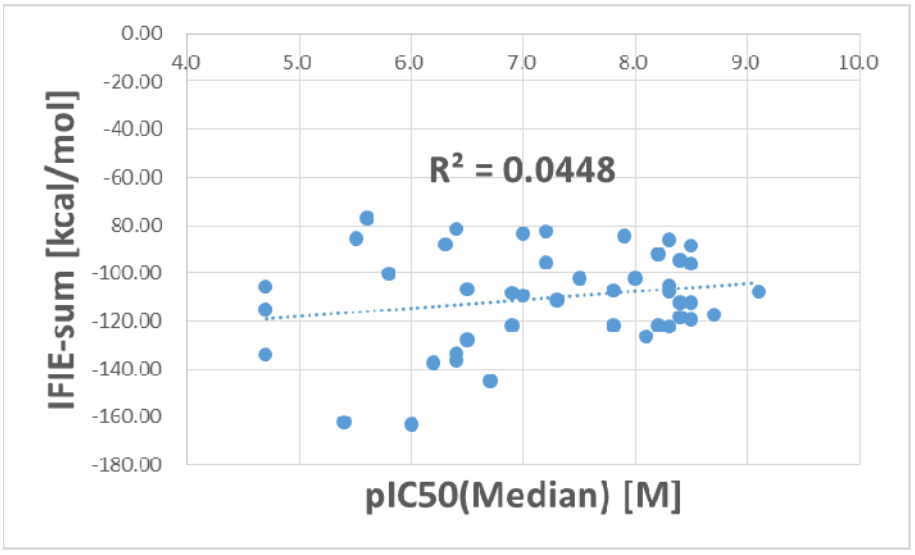

Figure 6 (b)

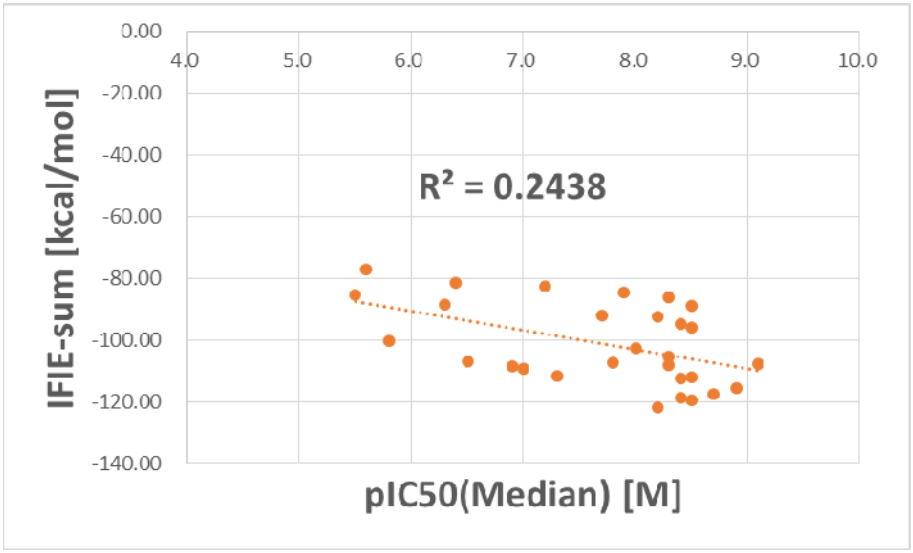

Figure 7 (a)

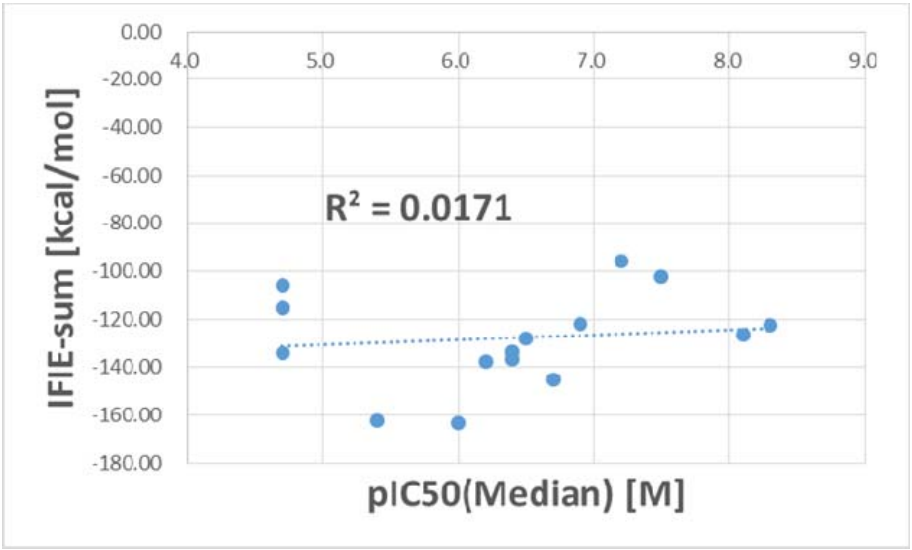

Figure 7 (b)

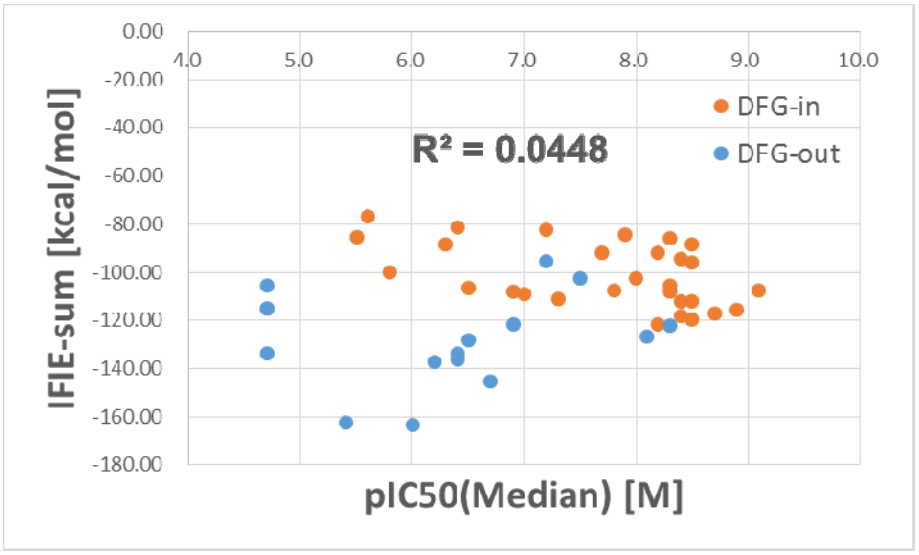

Figure 7(c)

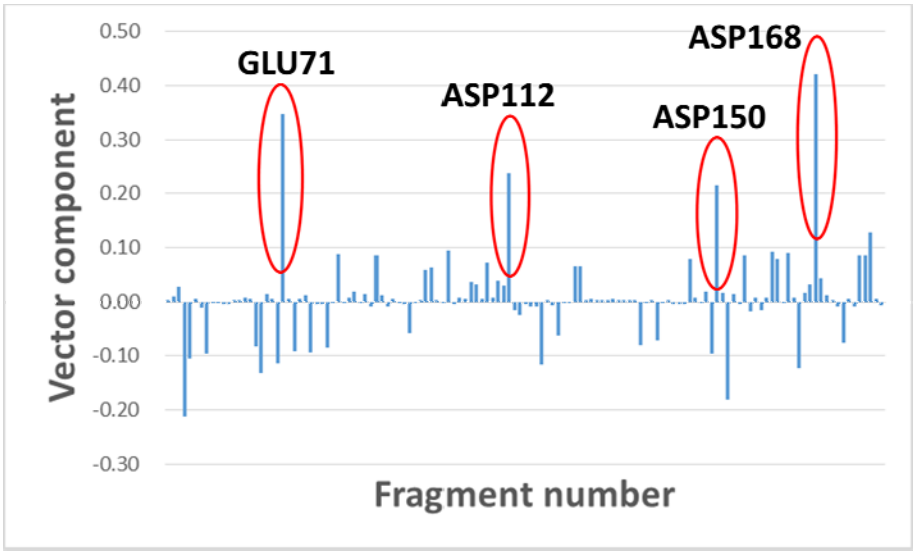

Figure 8 (a)

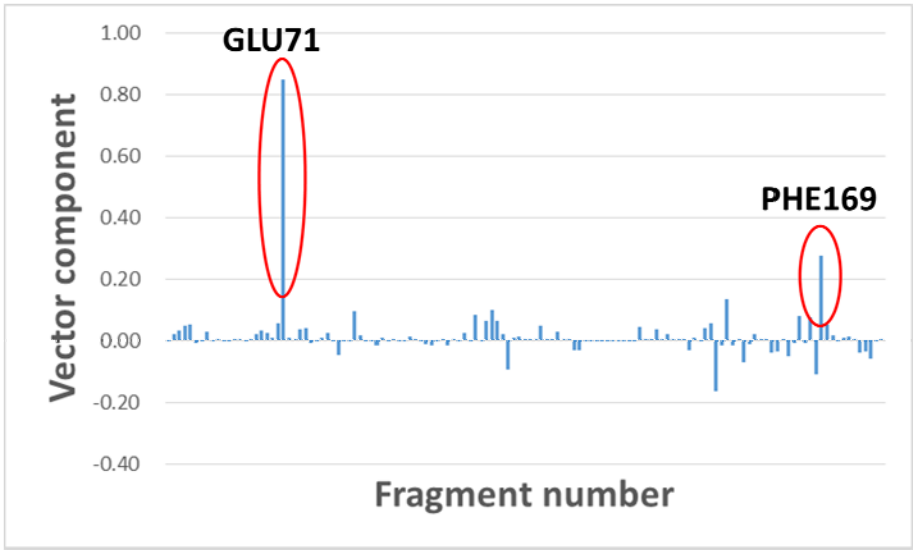

Figure 8 (b)

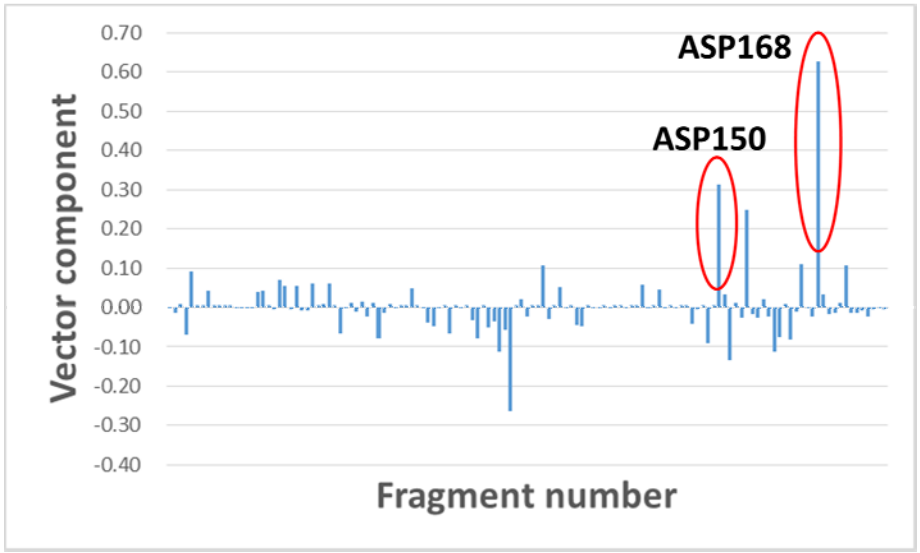

Figure 8 (c)

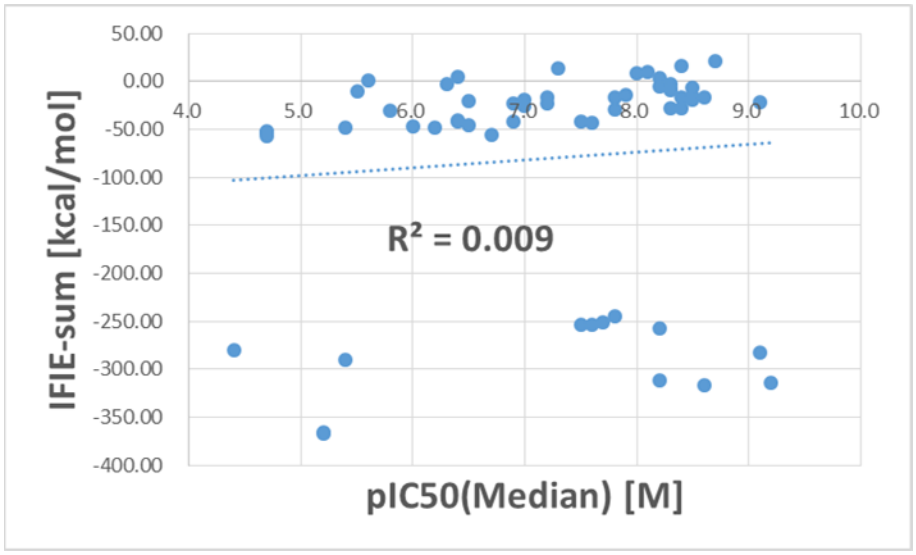

Figure 9 (a)

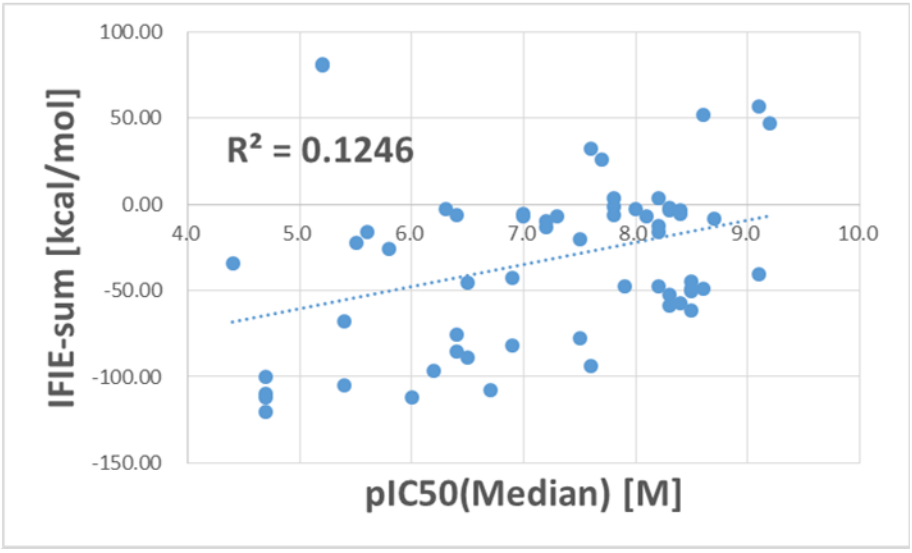

Figure 9 (b)

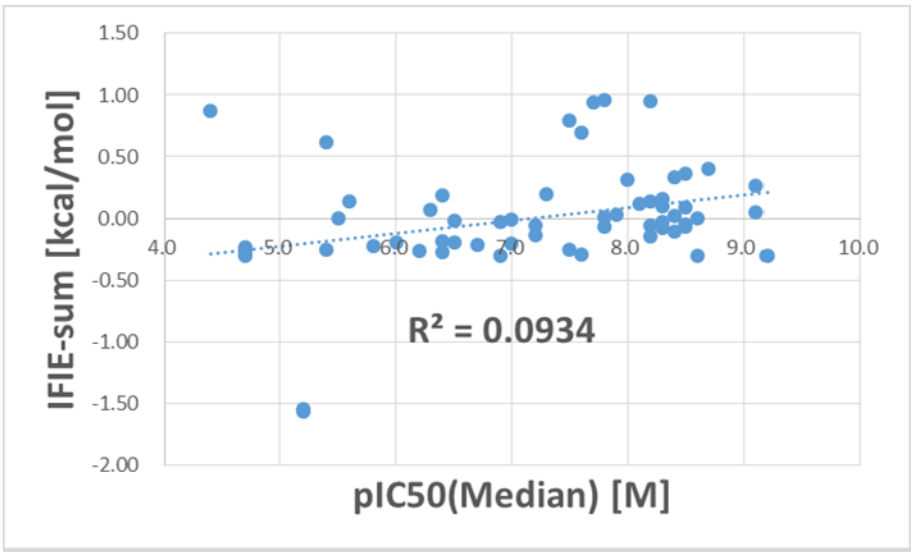

Figure 9 (c)

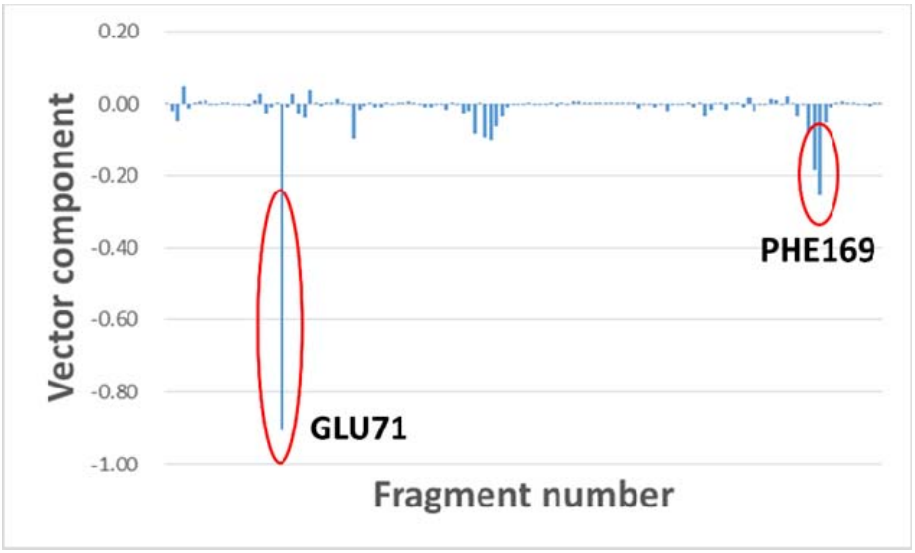

Figure 10 (a)

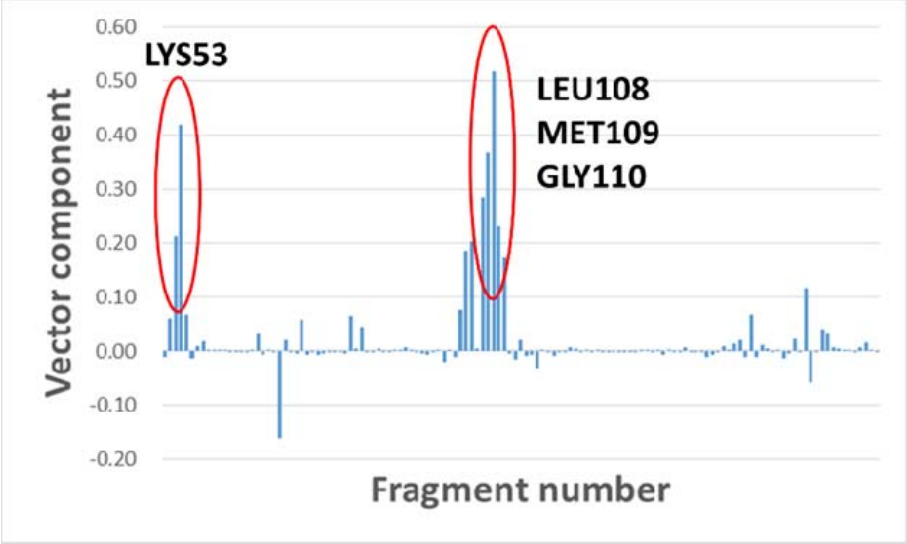

Figure 10 (b)

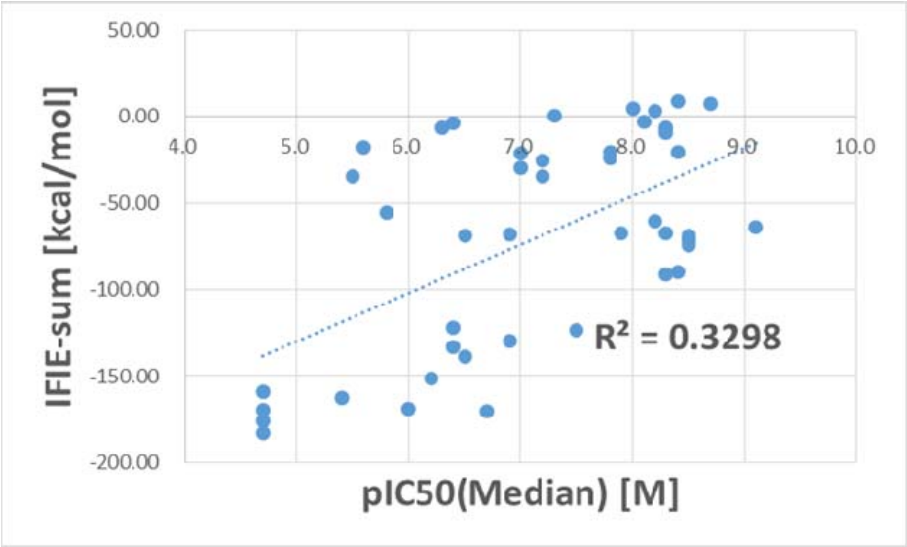

Figure 11 (a)

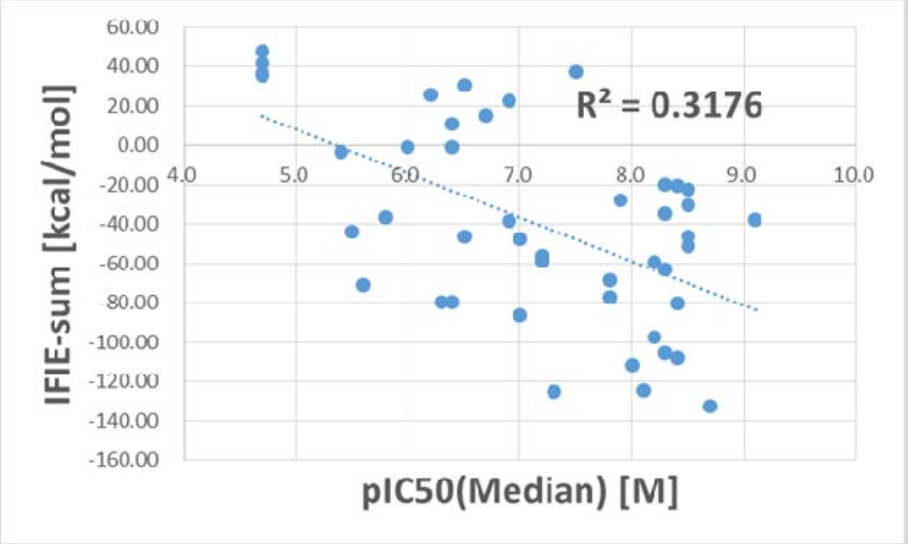

Figure 11 (b)

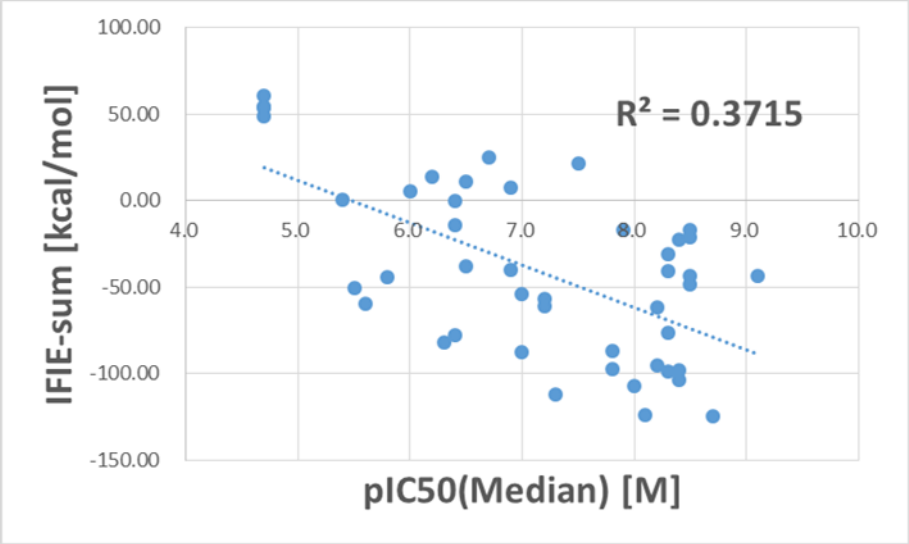

Figure 12

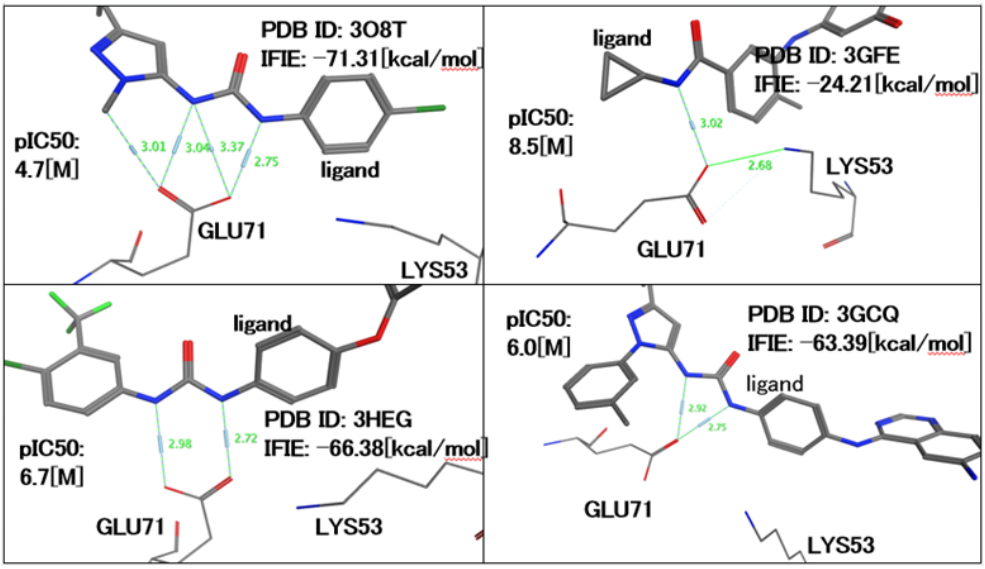

Figure 13

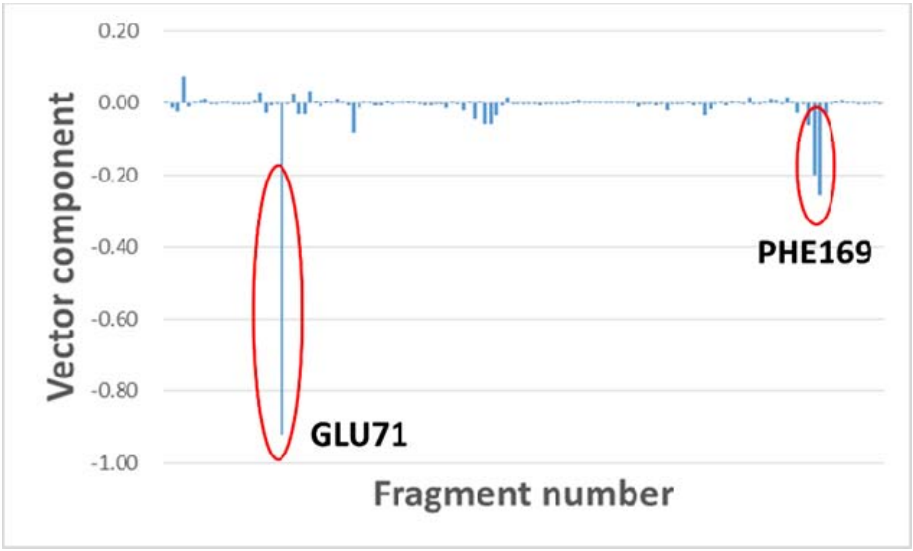

Figure 14 (a)

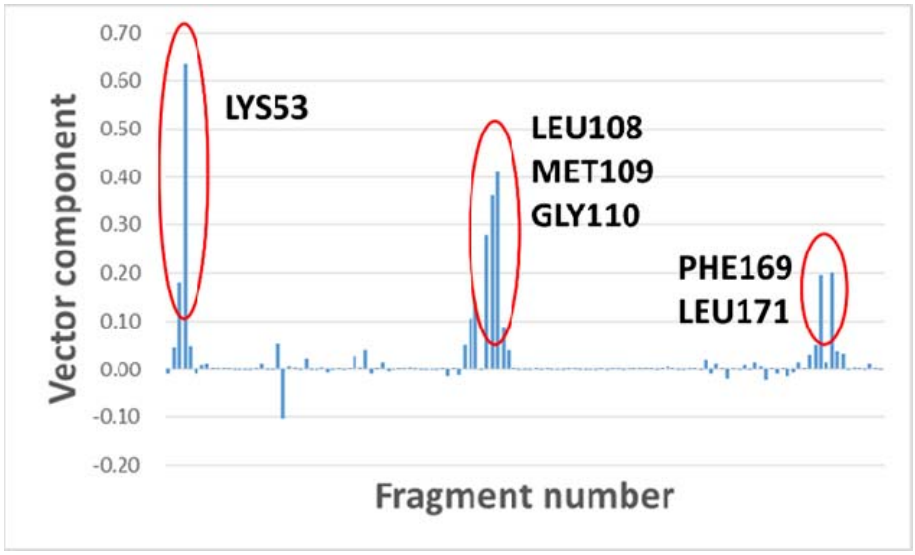

Figure 14 (b)

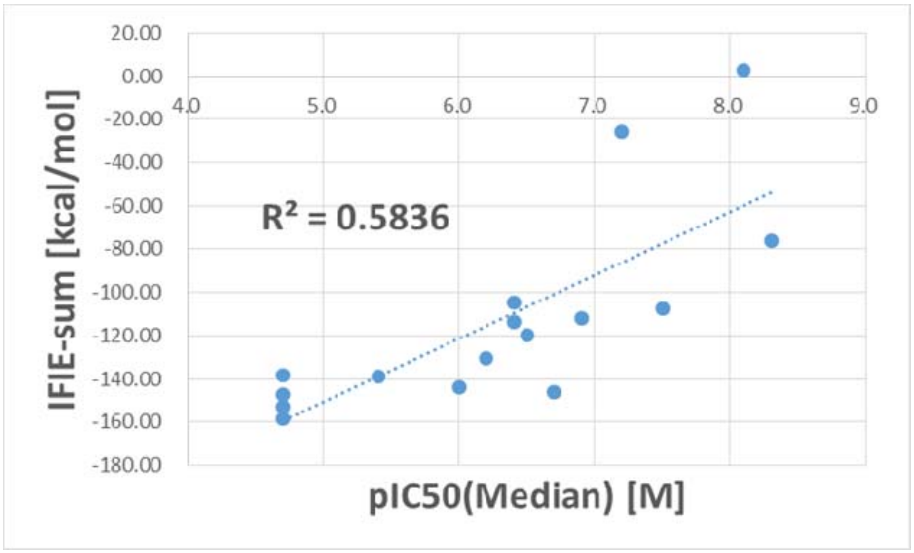

Figure 15 (a)

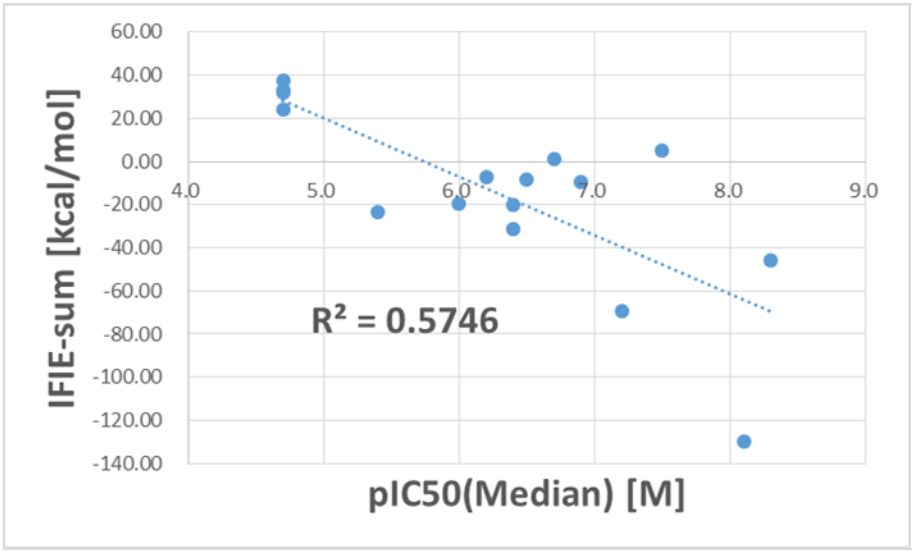

Figure 15 (b)

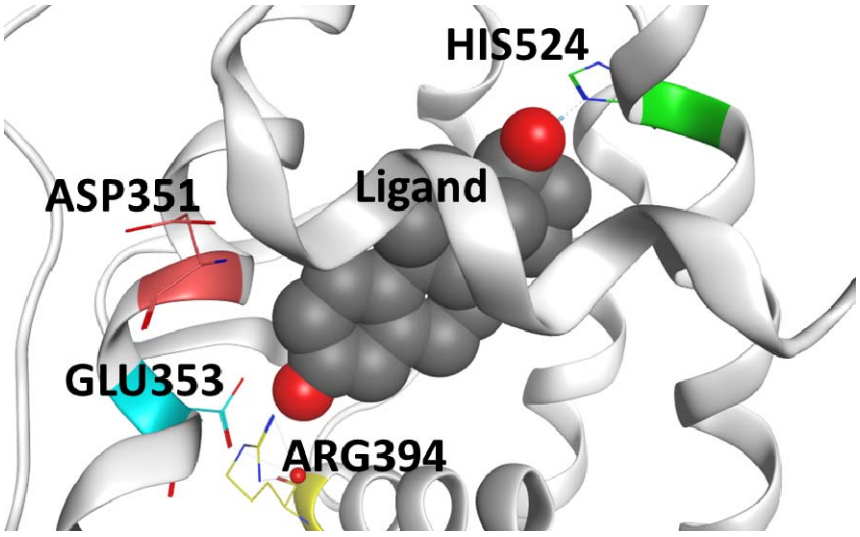

Figure A1

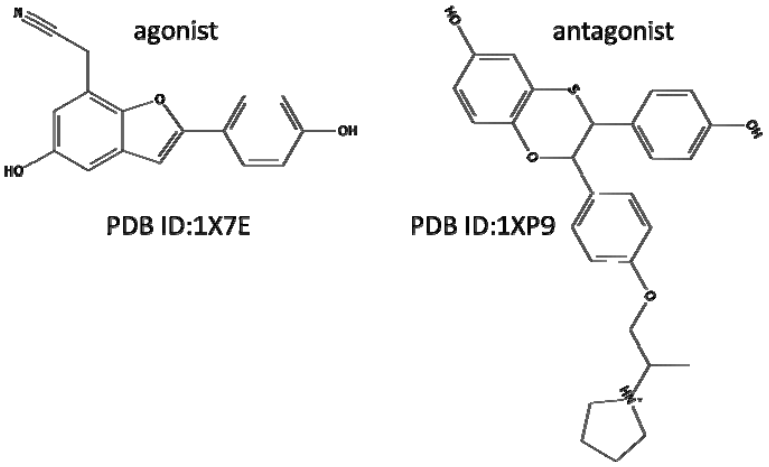

Figure A2

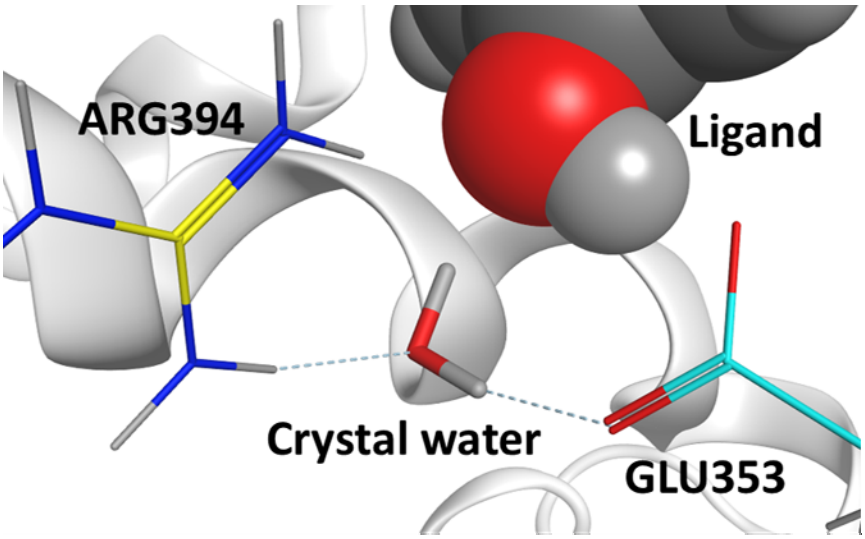

Figure A3

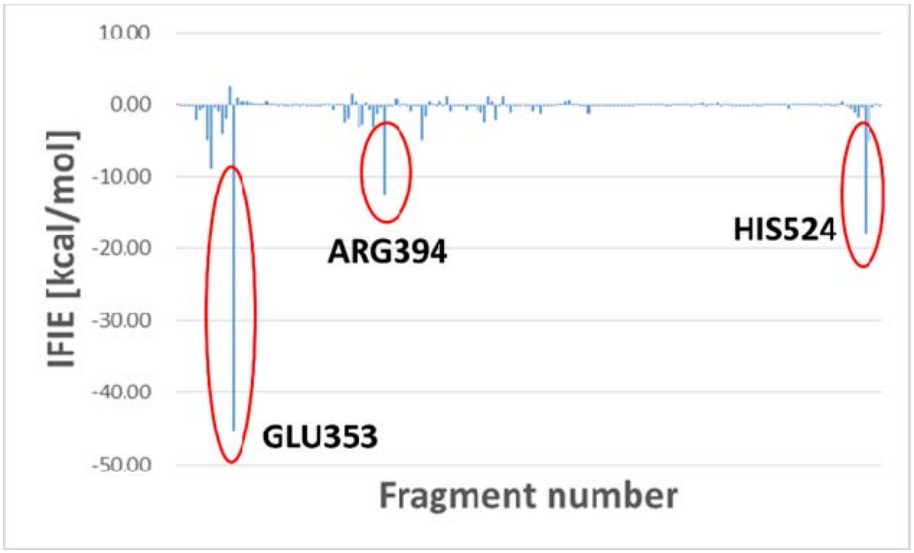

Figure A4

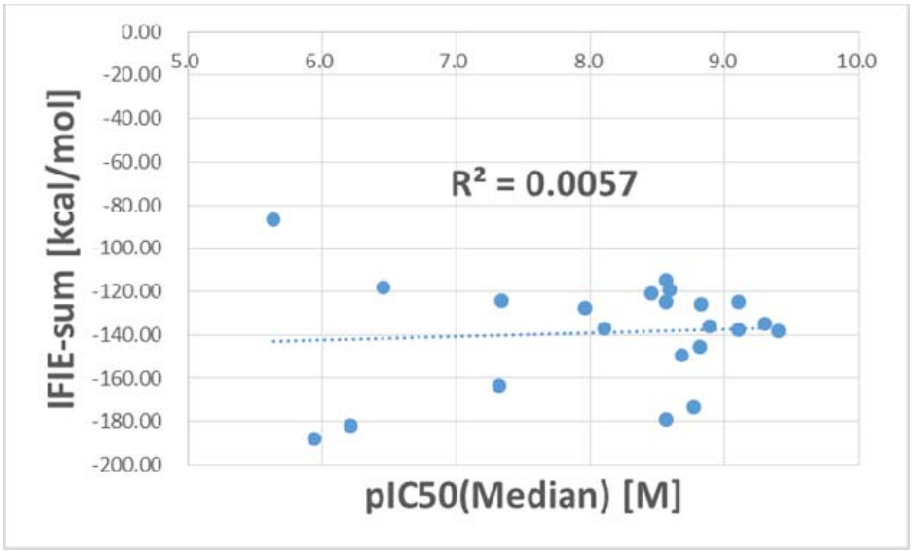

Figure A5

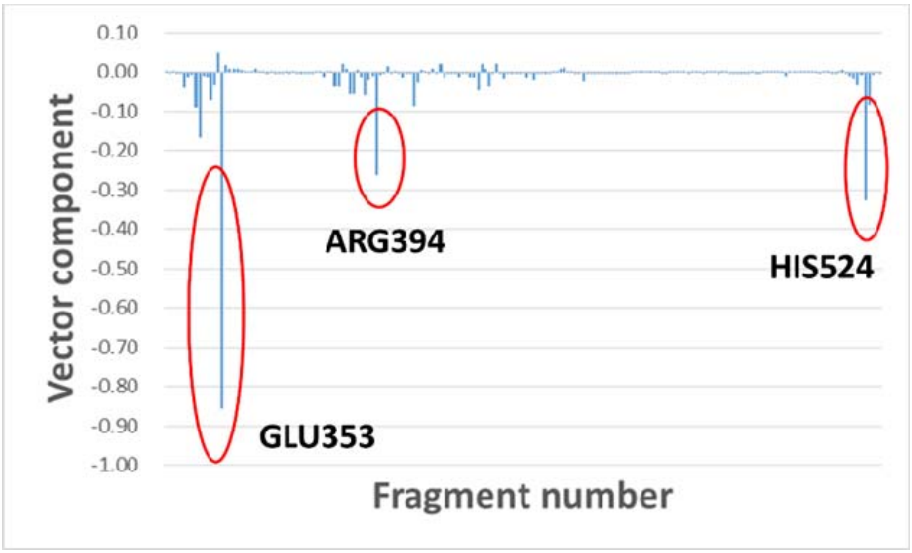

Figure A6 (a)

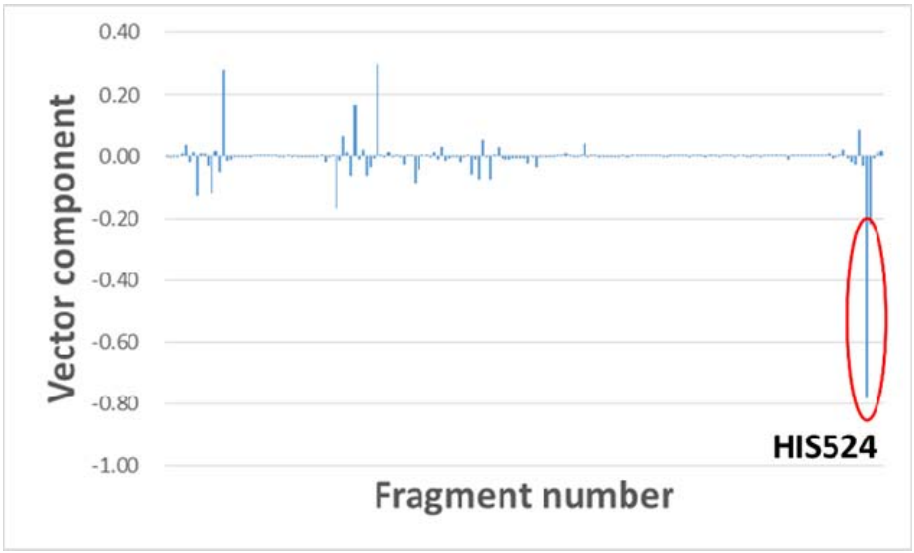

Figure A6 (b)

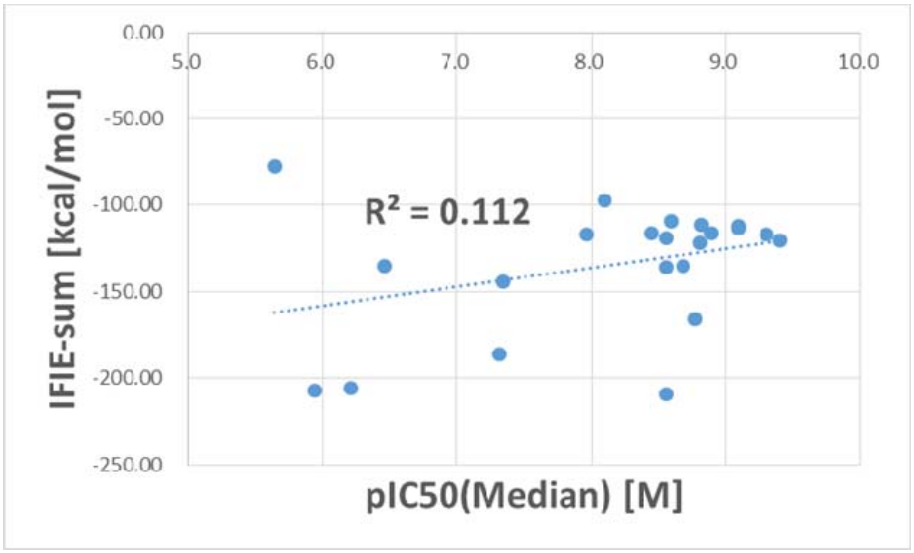

Figure A7 (a)

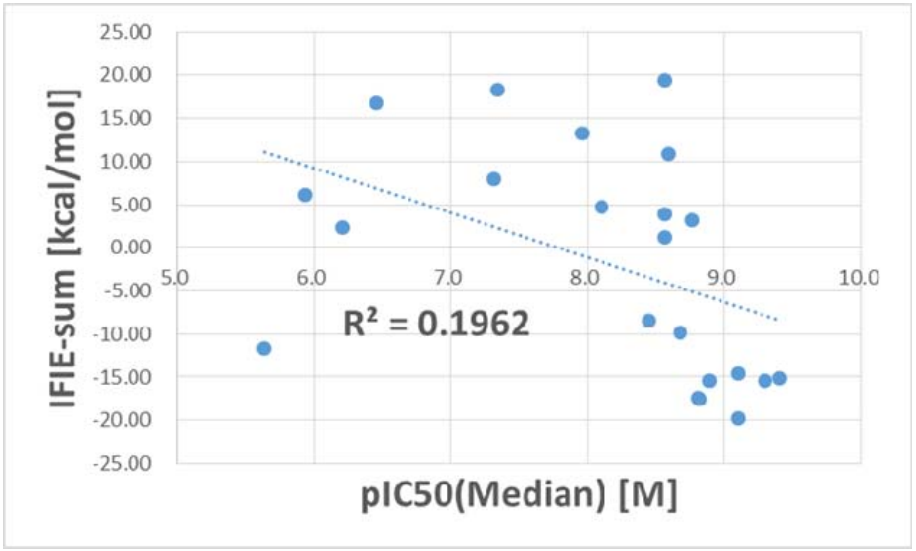

Figure A7 (b)

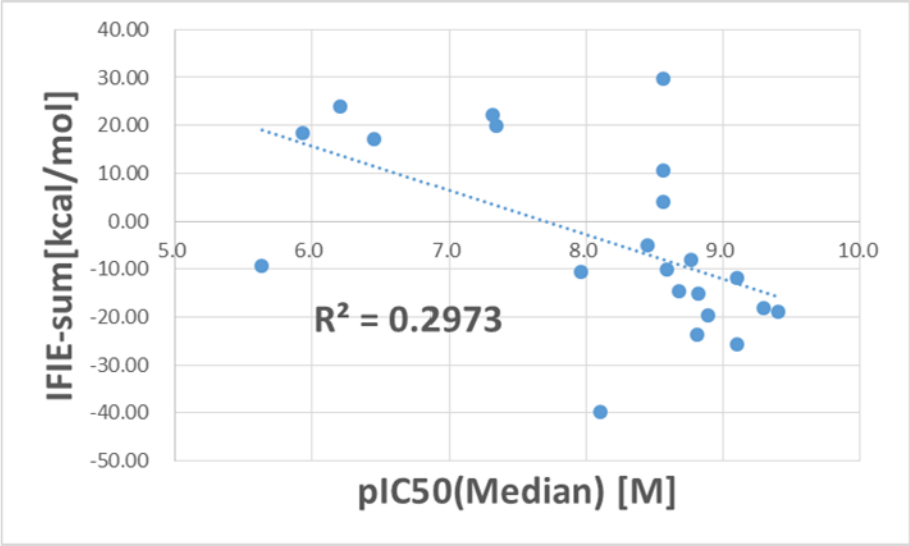

Figure A8

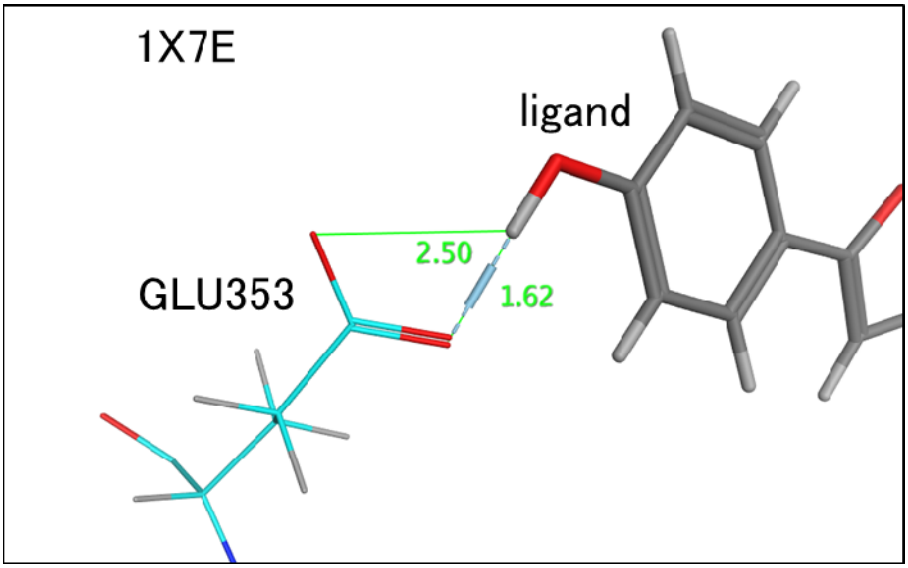

Figure A9