Supplemental information

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

Chiduru Watanabe[1, 2*], Hirofumi Watanabe[3], Yoshio Okiyama[1,4], Daisuke Takaya[1], Kaori Fukuzawa[1, 5], Shigenori Tanaka[6], Teruki Honma[1**]

[1]*Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan*
[2]*JST, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan*
[3]*Education Center on Computational Science and Engineering, Kobe University 7-1-48, Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan*
[4]*National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan*
[5]*Department of Physical Chemistry, School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa, Tokyo 142-8501, Japan*
[6]*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

A. Structure preparation of the automated FMO calculation protocol.

Flowchart of the structure preparation for the automated FMO calculation protocol (Auto-FMO protocol) shown in Figure 1. The details are described as follows.

(1)   Input data.

An MDB file including fields of receptor structures, ligand structures, and unique IDs as the protocol's input file is prepared. The data of each PDB file should be registered as one entry of the MDB file.

(2)   Complementation of missing atoms.

Each entry of the above MDB file is loaded, and the heavy atoms of the complex are fixed at the original coordinates of the input data. Next, the missing atoms of the PDB data are complemented by one of the following two procedures. First, both missing atoms and missing residues are complemented by homology modeling, using the sequence file given in the FASTA format. Second, the residues next to the missing residues are capped by NME and ACE groups without complementation, and the partial missing atoms of the side chains are corrected using the "StructurePreparation" function. In the complementation of missing atoms step, the addition of hydrogen atoms is simultaneously conducted.

(2.1) Homology modeling and structure preparation.

The SVL function on MOE and its options used in this paper for homology modeling are as follows. Although the options are mainly based on the default settings, several options, `nSideModels`, `intermediate_refine`, `protonate3D`, and `final_refine`, can be changed according to the user's preference.

```
pro_HomologyModel [
    Chains[],
    [ ignore_outgaps     : 1,
    selected_res         : 1,
    env_atoms            : 1,
    auto_disulfide       : 1,
    nMainModels          : 10,
    nSideModels          : 1,
    sample_T             : 300,
    rot_Ecutoff          : 1.5,
```

```
        rot_maxD            : 1.2,
        intermediate_refine : 'Middle',
        intermediate_gtest  : 1,
        protonate3D         : 1,
        final_refine        : 'Middle',
        final_gtest         : 0.5,
        model_score         :  'GB/VI'
        ]
]
```

(2.2) Only structure preparation.

The SVL function and its options for structure correction used in this paper are shown, as follows.

```
local SPREP_DEF_MOEBATCH = [
        cmd                 : 'cli',
        autoIsolate         : 1,
        batch_protonate3d   : 1,
        disableTypes        : 'HName'
    ];


StructurePreparation SPREP_DEF_MOEBATCH
```

(3)   Local minimization of complemented atoms.

Structural minimization of only the complemented atoms in procedure (2) is performed using the force field selected by the option. The SVL function and its options of structural minimization used in this paper are shown, as follows. Although the options are mainly based on the default settings, rigidHOH and verbose were set to active in this protocol.

```
MM[
        pot_finalize  : 1,
        pot_charge    : 1,
        keep_chirality : 0,
        sd_maxit      : 100,
        sd_gtest      : 1000,
```

```
        cg_maxit        : 100,
        cg_gtest        : 100,
        maxit           : 500,
        gtest           : 0.1,
        maxstep         : 1,
        maxstep_x       : 5,
        tetherWeight    : 0,
        rigidHOH        : 1,
        verbose         : 1
];
```

(4)   Constraint setup of user specifications.

In the optimizing step, the following four options can be selected. The other atoms are fixed and remain at their original PDB coordinates. The default setting was used in this paper.

Selection of optimizing atoms on the protocol's option.
>   **Option 1.** All hydrogen atoms of the complex structure (default).
>   **Option 2.** All hydrogen atoms of the complex structure; heavy atoms of the ligand molecule.
>   **Option 3.** All hydrogen atoms of the complex structure; heavy atoms of the ligand molecule; heavy atoms of the receptor and solvent molecules within threshold distance from the ligand. The threshold distance can be specified by the protocol's option.
>   **Option 4.** All atoms of the complex structure.

(5)   Final minimization of unrestrained atoms.

Final minimization of unrestrained atoms is performed by the following procedure. The complex structure is treated by the selected force field, and the partial charges of the complex are subsequently calculated. Structure minimization is performed by the MM function under the same options as in procedure (3).

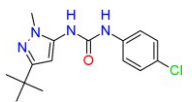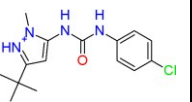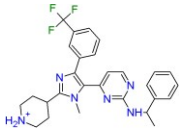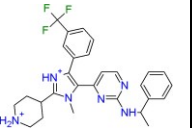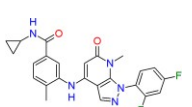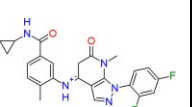**(6)   Save structures used for FMO calculation.**

The obtained optimized structure of each MDB entry is stored in the PDB file format. Subsequently, the structures are sent to the FMO calculation step.
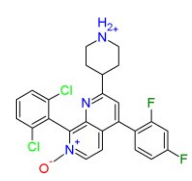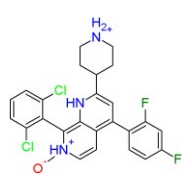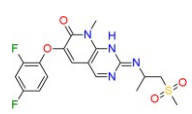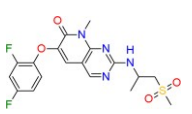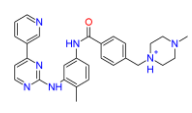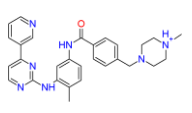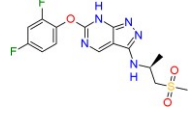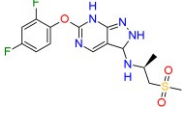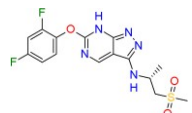
B. Difference of IFIE

Table S1 summarizes the entries with large differences between the protocol data and the manually prepared data for p38α. In the three complexes (PDB IDs: 3O8P, 1OUK, and 3GFE) with different ligand charges between the protocol data and the manual data, the differences of the I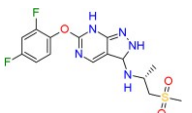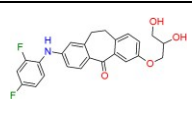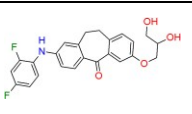FIE values were more than 100 kcal/mol. In the next entry (PDB ID: 3MW1), the ligand charges of both structures are equal; however the protonation state is different. The ligand structure (ligand name: MIH) by the protocol has two additional hydrogen atoms due to the treatment of the rare N-oxide moiety. As a result, the difference in the ligand binding energy recorded was 80 kcal/mol. There are different tautomerization states in the three complexes (PDB IDs: 3FLS, 3FML, 3FMM). The IFIE difference of the complex (PDB ID: 3HEC) resulted from the different protonated nitrogen atoms of the piperazine ring (ligand name: STI).

In the case of the 3QUE entry, the charge and tautomerization state of the ligand obtained by the protocol are the same as those in the manually prepared data. Here, Figure S1 shows the three-dimensional structures of the ligand binding pocket for both data sets. For the complementation of the missing residues by homology modeling, the complemented structures around the ligand showed significant differences. Glu173 formed a hydrogen bond with the ligand in the manually prepared data, while no hydrogen bond between the ligand and Glu173 was detected in the protocol data (Figure S1). The drastic conformational change of the Glu173 side chain caused a 20 kcal/mol difference in the ligand binding energies.

Table S1. Data set of the FMO protocol data and the manual data.

PDB ID (Ligand name)	Chain ID	Manual			Protocol		
		Structural Formula	Ligand charge	ΔE_{ligand}	Structural Formula	Ligand charge	ΔE_{ligand}
3O8P (BMU)	A		0	−131.16		1	−313.04
1OUK (084)	A		1	−247.85		2	−399.92
3GFE (P37)	A		0	−119.51		1	−252.79

3MW1 (MIH)	A		1	−234.81	1	−320.28
3FLS (FLS)	A		0	−86.03	0	−119.62
3HEC (STI)	A		1	−270.74	1	−301.61
3FML (FML)	A		0	−77.01	0	−105.67
3FMM (XI2)	A		0	−109.06	0	−131.13
3QUE (3FF)	A		0	−122.39	0	−99.19
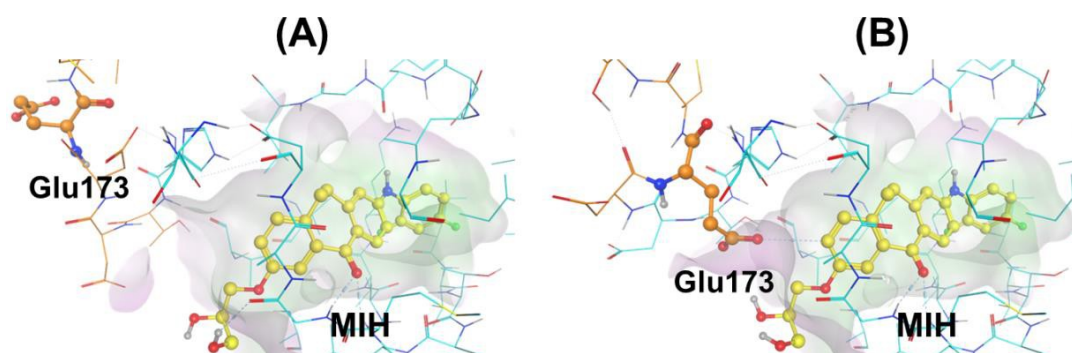
Figure S1. Three-dimensional structures of the p38α and ligand (MIH) complex (PDB ID: 3QUE, Chain: A) in the Auto-FMO protocoll data (A) and the manually prepared data (B)

The ligand and amino acid residues including the X-ray crystal structure are shown by yellow ball and stick models and the cyan line, respectively. Complemented structures regarding missing atoms are highlighted in orange. Glu173 is shown by an orange ball and stick model, because a very large conformational difference between the protocol data and the manual data was observed.

Figure S2 shows the correlation between the Auto-FMO protocol data and the manual data, excluding the obviously different data listed in Table S1. As a result, the correlations of all data (black), the neutral ligand charge data (red), and the positively charged data (blue) between the protocol and manual data were dramatically improved.
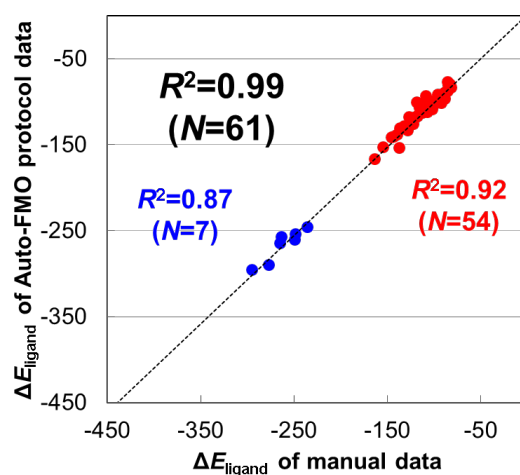
Figure S2. Validation of the binding energies compared with the manual data and the Auto-FMO protocol data without the obvious nine outliers in the p38α dataset

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