

Novel Quantitative Structure-Activity Studies of HIV-1 Protease Inhibitors of the Cyclic Urea Type Using Descriptors Derived from Molecular Dynamics and Molecular Orbital Calculations

Tatsusada Yoshida¹, Toshio Fujita² and Hiroshi Chuman^{*,1}

¹Institute of Health Biosciences, The University of Tokushima Graduate School, 1-78 Shomachi, Tokushima 770-8505, Japan

²Kyoto University, Office #305 Heights Kyogosho, Fuyacho-Nishikikoji-agaru, Nakagyoku, Kyoto, 604-8057, Japan

Abstract: Human immunodeficiency virus type 1 protease (HIV-1 PR) is an essential enzyme for the replication cycle of HIV-1. HIV-1 PR inhibitors have been extensively investigated as anti-AIDS drugs. For developments of HIV-1 PR inhibitors more promising than those utilized at the moment, the construction of reliable QSAR models that can elucidate the inhibitory mechanism as consistently as possible should be one of the most significant issues. Garg, Kurup, and their groups published comprehensive QSAR studies using past structure-activity data for HIV-1 PR inhibitors, and summarized some physicochemical structural factors of inhibitors that govern variations in the inhibitory activity for various structural types. There seem to exist much to be clarified further, especially for effects of electronic structure of inhibitors. It is also expected to incorporate structural and physicochemical information about the enzyme protein into the QSAR model. In this article, we reviewed our own QSAR study on a series of cyclic urea inhibitors with newly proposed QSAR descriptors. We performed molecular dynamics simulations of HIV-1 PR-inhibitor complexes to provide the accurate geometry to the fragment molecular orbital (FMO) calculations as well as to the estimation of an accessible surface area descriptor for inhibitors and amino acid residues. With the FMO procedure to cover full electronic feature of three-dimensional structure of protease-inhibitor complexes, we derived electronic descriptors for inhibitors and amino acid residues. The successful results are believed to provide a new insight into QSAR and understanding of binding mechanism of inhibitors with HIV-1 PR at atomic and electronic levels.

Keywords: *Ab initio* fragment molecular orbital (FMO), accessible surface area (ASA), charge transfer, cyclic urea type inhibitor (CUI), HIV-1 protease (HIV-1 PR), Inter-fragment interaction energy (IFIE), molecular dynamics (MD), quantitative structure-activity relationship (QSAR).

1. INTRODUCTION

1.1. AIDS, HIV, and HIV-1 Protease

More than twenty million people have died of the acquired immunodeficiency syndrome (AIDS) caused by Human Immunodeficiency Virus Type 1 (HIV-1) and many are presently suffering from this disease [1, 2]. HIV-1 protease (HIV-1 PR) is an essential enzyme for the replication cycle of this virus [3]. Inhibition of this enzyme leads to the production of immature, noninfectious viral particles. Thus, HIV-1 PR inhibitors have been extensively investigated as anti-AIDS agents and numerous synthetic and structure-activity studies for possible candidates have been practiced and published [4-6]. HIV-1 PR is expressed as a part of the gag and gag-pol polyproteins, and the target of inhibitors is certain amino acid sequences including the active site of the protease. There are other mechanisms to interfere with the replication cycle of HIV. Inhibition of HIV reverse transcriptase and viral entry to target cells are such mechanisms, which also have been investigated with much effort. Most drugs approved and clinically used to date are due to either inhibitory mechanism of the HIV-1 PR, HIV reverse transcriptase, or the viral target-cell entry [7-9].

1.2. Structure of the HIV-1 Protease

HIV-1 PR is a C₂ symmetrical homodimeric aspartic protease, each monomer containing 99 amino acid residues (1 ~ 99 and 1' ~ 99' in chains A and B, respectively), as shown in Fig. (1a) [10]. Since the first report of the X-ray crystallographic structure of this enzyme was published in 1989 [11-13], a number of crystal structures of its mutants and their complexes with various inhibitors have been accumulated and provided much information about the mechanism of the protease reaction as well as the complex formation with inhibitors [14, 15]. Thus, the active site of HIV PR has been well recognized to be constructed by a pair of catalytic triads, Asp25 – Thr26 – Gly27 and Asp25' – Thr26' – Gly27', located at the interface of the two symmetric core domains (two triads together function as a single active site) as shown in Fig. (1b) [16]. The active site/pocket is surrounded by subsites S1, S2, S1', and S2', which consist of hydrophobic amino acid residues, except for Asp29 (29') and Asp30 (30') in S2/2', as shown in Fig. (2) with the inhibitor molecule (DMP323 analogs) incorporated [16]. The active site is also covered by two symmetrical flap domains including the β hairpin "flap tips" from Trp42 to Gln58 and Trp42' to Gln58' on the dimer interface (not shown in figures). The flexible flaps change their conformation between free and complexed (with substrates and inhibitors) states, and possibly contribute to the molecular motion responsible for substrate entry and release.

*Address correspondence to this author at the Institute of Health Biosciences, The University of Tokushima Graduate School, 1-78 Shomachi, Tokushima 770-8505, Japan; Tel: +81-88-633-7257, Fax: +81-88-633-9508; E-mail: hchuman@ph.tokushima-u.ac.jp

Fig. (1). Three-dimensional structure of HIV-1 PR–DMP323 complex (PDB code; 1QBS). (a) Catalytic residues Asp25 (25') of the active site are represented in a sphere model. (b) Magnified view of DMP323 (represented in a stick model) and the active site with a pair of catalytic triads, Asp25 (25') – Thr26 (26') – Gly27 (27').

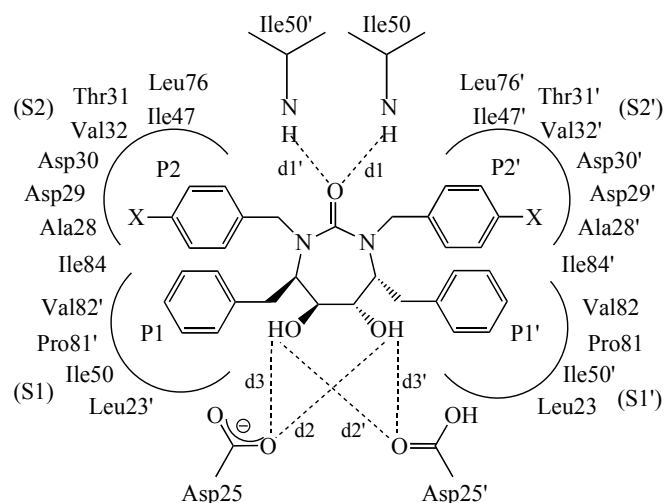


Fig. (2). Schematic representation of the binding sites in the HIV-1 PR–cyclic urea type inhibitor (CUI) complex. According to the standard notation, P1/P1' and P2/P2' are for CUI analogs, and S1/S1' and S2/S2' represent corresponding subsites in HIV-1 PR.

1.3. Inhibitors of HIV-1 Protease

HIV-1 PR inhibitors can be either peptidomimetic or non-peptidomimetic. Representative inhibitors are shown in Fig. (3). Although such peptidomimetic inhibitors as saquinavir (1), ritonavir (2), and indinavir (3) have been observed to increase the lifespans of HIV-infected patients [17], they exhibit limited oral bioavailability owing to their high molecular weight/volume, poor solubility, and substantial peptide character vulnerable to hydrolysis. To overcome these drawbacks, a large variety of non-peptidomimetic inhibitors have been developed. Among them, such “cyclic urea-type inhibitors” (CUIs) [18, 19] as DMP450 (4), DMP323 (5), and their cyanoimine analogs (6) are unique in their structure with the C_2 symmetry and four benzyl and two hydroxyl groups on the seven-membered ring designed to fit well with the active site of HIV-1 PR. They were also intended to maintain the efficient hydrogen bond network commonly observed in the enzyme complexed with

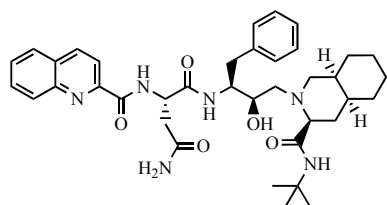
peptidomimetic inhibitors [18, 19]. Thus, the two hydroxyl groups build up hydrogen-bonds with the catalytic Asp25 (25'), and the urea carbonyl-oxygen (and the imino-nitrogen) forms hydrogen-bonds with the flap-tip Ile50 (50') NH hydrogen atoms, the latter simulating those achieved by structural water molecule in the complexes with peptidomimetic inhibitors [18, 19].

The HIV-1 PR is a member of aspartic protease superfamily, and the mechanism of the enzyme reaction is considered as a type of general-base catalytic hydrolysis [14, 15]. The inhibitors have been accepted to form transition-state analogs with Asp25 and 25' to interrupt the otherwise ongoing hydrolysis and to create multiple hydrogen bond networks to stabilize the “nonproductive” complex [24]. There should be electronic mechanisms for enzyme reaction as well as the enzyme inhibition rationalized in terms of QSAR.

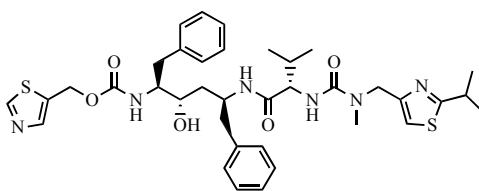
1.4. Past QSAR Studies for HIV-1 Protease Inhibitors

Corresponding to a number of synthetic and structure-activity publications [10, 25-33], a number of QSAR studies have been pursued to understand the molecular mechanism of HIV-1 PR inhibition as well as to obtain clues for designing more active inhibitors. These studies have been compiled, analyzed, and reviewed by Garg [34], Kurup [35], and respective coworkers. In their studies, structure-activity data were collected from various sources, in which original authors had not necessarily intended for the QSAR analyses. In each series of inhibitors with a fixed scaffold, structure modifications are performed mostly as substituent variations at certain defined positions. Substituent variations are not restricted in a narrow range at least for hydrophobic and steric points of view to make some lateral and comparative QSAR examinations. Thus, Garg *et al.* were able to present a predictable generalization for the molecular hydrophobicity required for the better complex formation of this type of inhibitors [34, 36]. However, the range of variations is not enough in many cases in terms of electronic properties to obtain information about the electronic mechanism of interaction including hydrogen-bond formation occurring between the enzyme and inhibitors.

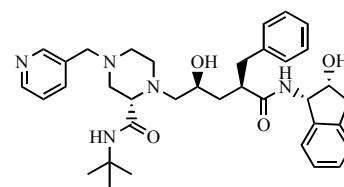
(a) Peptidomimetic type



1
Saquinavir (Invirase)
Hoffmann-La Roche [Ref. 20] (4.73)

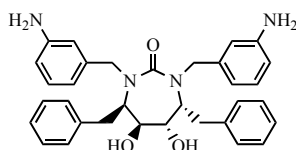


2
Ritonavir (Norvir)
Abbott, Glaxco Smith Kline [Ref. 21] (4.94)

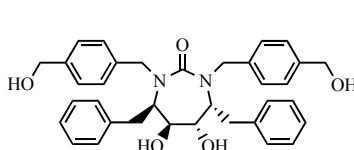


3
Indinavir (Crixivan)
Merck [Ref. 22] (3.68)

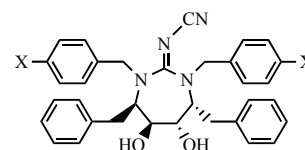
(b) Cyclic urea type



4
DMP450 [Ref. 19] (4.78)



5
DMP323 [Refs. 18,19] (5.16)



6
Cyanoimine analogs of
DMP450 and DMP323 [Ref. 23]

Fig. (3). Structure of representative HIV-1 PR inhibitors. (a) FDA-approved peptidomimetic inhibitors. (b) Non-peptidomimetic cyclic urea type inhibitors. Values in parentheses show *Clog P* (see below).

1.5. Novel Descriptors Derived from Molecular Dynamics and Molecular Orbital Methods and their Application to QSAR for Protein–Ligand Complexes

Although classical QSAR studies mentioned above and others [37-45] have indicated various degrees of success, information provided is only based on the structure of inhibitors, and not on the structure of the complexes with the target enzyme. Recently, structures of crystalline HIV-1 PR complexes with inhibitors have been accumulated rapidly as mentioned above. Likewise, the computational algorithms and machine power have been advancing enormously. Thus, molecular simulation and calculation techniques such as molecular dynamics (MD) and the *ab initio* molecular orbital (MO) method for ligand–protein complexes have become applicable to accurate evaluation of ligand–protein interactions. Molecular dynamics simulation could be useful to understand dynamic behaviors of the target protein molecule during the complex formation. Molecular orbital calculation provided for macromolecular compounds, such as the fragment MO (FMO) method [46], has made it possible to calculate the full electronic feature of proteins as well as to describe the changes in the atomic charges induced by the complex formation. The overlooking of possible significance among essential terms for electronic, steric, and hydrophobic effects and others in classical QSAR equations could hopefully be overcome with a novel procedure using physico-chemical information about the ligand–protein complex.

In this article, we would like to present our newly developed QSAR procedure using descriptors calculated for the

ligand–protein complex and show how these new descriptors are able to interpret the mechanism of the protein–ligand complex formation [47]. We also discuss possible links between this “MD-MO” QSAR and the classical procedure in terms of relationships of descriptors between the two. As the structure-activity data for the analysis, we selected those for the CUI analogs of DMP450 (**4**) and DMP323 (**5**), for which Garg *et al.* formulated QSARs showing significance in the negative electronic effect along with the negative hydrophobic effect of substituents on the benzene ring of the P2/P2' benzyl groups. In fact, we used only *para*-substituted benzyl compounds at the moment as will be mentioned later. Other reasons for this selection are because the cyclic urea structure is compact without much conformational freedom to make the computation easier and the X-ray crystallographic data for their complex with HIV-1 PR are available for setting the initial conditions for calculations.

2. MODELING OF DYNAMIC 3D-STRUCTURE OF HIV-1 PR–CUI COMPLEX

2.1. Significance of Dynamic Structure of the Protein–Ligand Complex

With the recent progress in fine experimental techniques, substantial dynamic fluctuations have been shown to occur in conformation of proteins. Internal dynamics plays a key role in protein function. In fact, the biological activity of proteins requires flexibility because interactions with substrates and inhibitors usually involve at least small rearrangements of atoms, and these fluctuations are essential

to biological activity [48]. A number of MD studies have demonstrated the importance of the conformational transition in HIV-1 PR induced by ligand binding [49-54]. Other MD studies have also revealed that certain amino acid mutations dramatically perturb the conformation at the active site and the flap region in HIV-1 PR [55]. In the present QSAR study, where MD-MO derived descriptors are to be used, we defined first 3D atomic coordinates of protein-ligand complex under "equilibrated fluctuation" conditions with the MD procedure and provided the coordinates to calculations by the *ab initio* fragment molecular orbital (FMO) method.

2.2. Preparation for Molecular Dynamics Calculations

We started with the static conformation of the initial geometry of the HIV-1 PR complex of DMP323 (**5**) obtained from the X-ray crystallographic coordinates (PDB code: 1QBS [10]) as the reference. For complexes of CUI analogs with other *para*-substituted benzyl groups, compounds **5** and **7-18**, the coordinates were derived by replacing the CH₂OH group of DMP323 with corresponding substituents. Each of the complexes was then surrounded by a shell of about 8000 TIP3P water molecules with a thickness of 20 Å. The dissociation states of the two catalytic aspartic acids, Asp25 (25'), have been discussed from both experimental [56-58] and theoretical [59-68] aspects, but their exact states are still ambiguous. However, according to the catalytic mechanism of the peptide bond hydrolysis by the aspartic protease family [69, 70], one of the Asp25 (25') residues was taken to work as the ionized (deprotonated) form while the other was regarded to be neutral (protonated). We assumed that other residues with ionizable side-chains are ionized. To make the electrostatic charge of the entire system neutral, five chloride anions were introduced into the surrounding shell.

2.3. Molecular Dynamics Calculations and Conformational Sampling of the Structure of Complexes

Using the AMBER program package [71], we carried out the calculations for the entire complex under TIP3P-aqueous conditions. First the static conformation of the complex with the initial geometry was energy-minimized. With this energy-minimized structure, MD calculations were carried out at 298 K. The total energy of the system was calculated as the summation of (a) potential energies for stretching, bending, and torsional freedoms, (b) Lennard-Jones and Coulomb potential energies for the non-bonded atom pairs with the 1-4 atomic relation and greater in each of monomers of the total protein and inhibitor, and (c) the intermolecular interaction energies among monomers of the protein, inhibitor and TIP3P water molecules. The hydrogen-bonding interaction is taken into account effectively within the summation of these potential functions. The basic force field parameters are pre-defined for peptide and water molecules empirically and implemented in the program. The partial atomic charges for CUIs were calculated *via* the RESP (restrained electrostatic potential) fitting algorithm [72] so as to be consistent with the AMBER force field used for the MD calculation.

After several preliminary MD runs for heating and equilibration at 298 K taking 60 and 250 ps, respectively, the final productive dynamics were examined at 298 K [73]. Ten snapshot structures during the 500 ps productive run were extracted (one snapshot per every 50 ps). Each snapshot structure was energy-minimized using the conjugate

gradient method. With these energy-minimized coordinates, we calculated inter-atomic distance values defined below as the average \pm fluctuation for each of the HIV-1 PR-CUI complexes.

2.4. Fluctuation of the Structure of the Protein-Inhibitor Complex

The fluctuation in terms of root-mean-square deviations (RMSD) of all atoms in the HIV-1 PR-DMP323 complex during the productive MD run was within 0.80 Å. This indicates that the complex was taking on essentially an identical conformation with small amplitudes of fluctuation. In particular, the seven-membered ring of the inhibitor was tightly fixed through interactions with the Ile50 (50') and Asp25 (25') residues. Fig. (4) shows that the inter-atomic distance values for d1 and d1' (defined in Fig. (2)) between urea oxygen of DMP323 (**5**) and the amide hydrogen atoms of Ile50 (50') are nearly equivalent to each other and the fluctuation (± 0.17 Å) with time occurs around 2.30 Å, which is the distance for the effective hydrogen bond formation. Fig. (4) also shows that the two hydroxyl groups on the seven-membered ring are located closely to the side-chain atoms of Asp25 (25'), with considerably varying distances, d2, d2', d3, and d3' (also defined in Fig. (2)), 2.87 ± 0.13 , 3.15 ± 0.16 , 1.78 ± 0.09 , and 3.44 ± 0.18 Å, respectively. The two hydroxyl groups interact with the negatively ionized side-chain of Asp25 owing to a higher electrostatic attraction more strongly than with the neutral Asp25' side-chain. The interactions of DMP323 with chain A and B are not symmetric in this model.

Fig. (4). Fluctuations in inter-atomic distance between hydrogen-bonding partners of DMP323 and the active site residues in HIV-1 PR (see Fig. (2) for d1, d1', d2, d2', d3, and d3').

2.5. Superimposition of Structures of Protein-Inhibitor Complexes

Every energy-minimized and averaged snapshot structure of the HIV-1 PR-DMP323 complex was superimposed so as to make the RMSD of distances between corresponding C _{α} atoms in each of residues among snapshot structures minimum. The average RMSD for heavy atoms of DMP323 was 0.167 Å among averaged snapshot structures, again indicating that DMP323 in the complex takes an almost identical conformation with very small amplitudes of fluctuation. The structure of complexes of the HIV-1 PR with CUI other than DMP323 (**5**) was composed according to similar procedures. Fig. (5) shows the superimposition of such structures for DMP323 (**5**) and CUI analogs used for the present correla-

tion study. The seven-membered ring and the P1/P1' moieties (Fig. 2) are tightly incorporated into a similar conformational space inside the active site of HIV-1 PR. The P2/P2' moieties (Fig. 2), in which various substituents are introduced at the *para* position, are shown to occupy somewhat varying conformations depending on the substituents.

Fig. (5). Superimposition of Compounds 5 and 7-18. The background protein structure is the one bound with DMP323 (5). The average RMSD value for superimposition of all compounds was 0.421 (± 0.158) Å (for heavy atoms).

3. *AB INITIO* FRAGMENT MOLECULAR ORBITAL (FMO) CALCULATION

3.1. Features of the FMO Method

Recently, Kitaura *et al.* developed an *ab initio* molecular orbital (FMO) method enabling us to make the full electronic calculation of proteins [46]. In this method, a large biomolecular system is formally “divided” into fragments and the MO calculation covering the full molecular system was avoided to reduce computational time dramatically without appreciable loss of accuracy. The total energy of the system can be evaluated by summation of fragment energies and Inter-Fragment Interaction Energies (IFIEs). IFIE is one of the most advantageous outcomes directly obtained from the FMO procedure [74-87], and detailed analyses of IFIEs between ligand and amino acid residues in the complex can provide significant insights into the interaction mechanism. FMO can also give atomic charges that are difficult to be estimated properly with classical approaches, so it makes possible to evaluate differences in the atomic charge induced by the complex formation. We have introduced the IFIE and the charge difference values as a new type of “QSAR descriptors” in the present series of HIV-1 PR-CUI complexes to identify significant electronic features responsible for variations in the binding potency as will be described in the later sections.

3.2. Outline of the FMO Theory and Calculation

In the FMO method, MO calculations are performed on single fragments (monomers) and their pairs (dimers) among “*N* fragments” to obtain the total energy and the properties of the entire system. The ligand molecule is treated as a single

“fragment” in this study. The fragmentation of a polypeptide is usually carried out as shown in Fig. (6). It should be noted that the definition of FMO fragment “slightly” differs from the standard notation of the amino acid residues. Both the peptide carbonyl of the (*i* - 1)-th residue and NH of the *i*-th residue belong to the same fragment *i*.

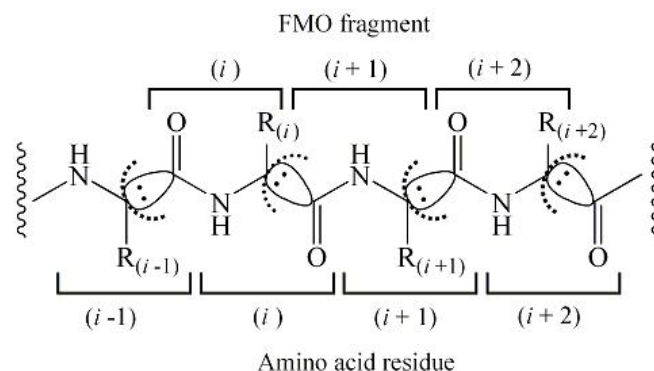


Fig. (6). Fragmentation of polypeptides in the FMO method. For fragments to which no electron pair is allocated from the bond when detached in the fragmentation, the MO is built from usual atomic basis functions of the constituent atoms according to the conventional LCAO-MO framework. For fragments in which bonding electron pair is left, the atomic valence basis function of the partner atom, with which the fragment is connected originally, is used additionally in the LCAO-MO model.

With the Hartree-Fock (HF) approximation on the closed shell systems, the Fock equations for monomer *I* and dimer *IJ* are solved repeatedly under the environmental electrostatic potential due to surrounding fragments, until they converge self-consistently. The total energy of the system *E* in the FMO theory is expressed by Eq. 1 with the energies of monomer *I*, *J*, and dimer *IJ*, denoted as E_I , E_J , and E_{IJ} , respectively. The first term is the sum of monomer energies and the second term accumulates energy variations due to interaction between monomers *I* and *J* in dimer *IJ*.

$$E = \sum_I^N E_I + \sum_{I>J}^N (E_{IJ} - E_I - E_J) \quad (1)$$

We can transform Eq. 1 to Eq. 2 by introducing ΔE_{IJ} defined by Eq. 3.

$$E = \sum_{I>J}^N E'_I + \sum_I^N \Delta E_{IJ} \quad (2)$$

$$\Delta E_{IJ} = (E'_{IJ} - E'_I - E'_J) + \text{Tr}(\Delta \mathbf{P}^{IJ} \mathbf{V}^{IJ}) \quad (3)$$

All the quantities in the right hand side of Eqs. 2 and 3 are now computable. E'_I , E'_J , and E'_{IJ} are, respectively, the energy of monomers *I* and *J*, and the dimer *IJ* in their interaction-free state under conditions ignoring the environmental electrostatic potential. They are defined as the counterparts of elements in the second term of Eq. 1. In Eq. 3, ΔE_{IJ} is the interaction energy on the basis of E'_I , E'_J , and E'_{IJ} ignoring environmental electrostatic potential (the first term) and the additional term to account for interactions of monomers *I* and *J*, and dimer *IJ* with surrounding fragments (the second term). $\Delta \mathbf{P}^{IJ}$ is the matrix of density difference between dimer *IJ* and the sum of monomers, *I* and *J*, \mathbf{V}^{IJ} is the environmental electrostatic potential for the dimer *IJ* governed by

electrons in surrounding fragments and all nuclei in the molecule, and Tr (trace) denotes the sum of diagonal matrix elements. In this situation, ΔE_{IJ} can be defined as the inter-fragment interaction energy (IFIE) between components of dimer IJ . In following sections, ΔE_{IJ} is denoted as IFIE(i) when J is the inhibitor (as a single fragment) and I is the i -th fragment in the protein.

Introduction of three approximations mentioned below can reduce computational time without discernible loss of accuracy. The Coulomb interaction approximation for well-separated dimers makes it unnecessary to solve the HF equation (approximation 1). The other two approximations are related to the environmental electrostatic potentials \mathbf{V}^{IJ} ; namely, the fractional point charge approximation (approximation 2) and the Mulliken orbital charge approximation (approximation 3). In this study, the approximations 1, 2, and 3 were applied to fragments whose separations were greater than 2.0, 2.0, and 0.0, respectively (these distances are given in a unit of the sum of the van der Waals radii of the closest contact atoms in each fragment). All FMO calculations were performed at the HF/6-31G level using the ABINIT-MP program [88].

4. QSAR ANALYSIS OF CYCLIC UREA TYPE HIV-1 PR INHIBITORS USING MOLECULAR DYNAMICS AND *AB INITIO* FMO CALCULATIONS

4.1. Set of Compounds

The structure-activity data of cyclic urea type inhibitors (CUIs) for the present QSAR analysis were taken from publications of the DuPont-Merck group [10, 23]. Every analog ($n = 13$) of DMP323 (Compound 5) with *para*-substituted benzyl group at the P2/P2' positions in the original publications was used for the analyses as briefly mentioned in Section 1.5 unless otherwise noted. The *meta* and *ortho* substituted compounds were excluded because of their multiplicity of the benzene ring conformation causing the prolonged computational time. The structure, the inhibition constant in terms of pK_i (K_i in M) against HIV-1 PR, and QSAR descriptors for thirteen compounds are listed in Tables below. The number of compounds included in the analysis was limited because of constraints in computational time, cost, and resources. CPU time for the FMO calculation is about 30 hours for a single HIV-1 PR-CUI complex with 4 dual Intel Xeon 2.33 GHz mini-clusters (8 CPUs).

4.2. Accessible Surface Area Parameter

The Accessible Surface Area (ASA) is a significant index in analyzing free energy difference for the phase transfer of organic molecules from polar to non-polar environment [89]. For (non)branched and cyclic aliphatic hydrocarbons and amino acids with the alkyl side-chain, the free energy difference has been shown to be linearly related with variations in the ASA value [89, 90]. The ASA difference, ΔASA value, has been considered to reflect the difference in the entropy and dispersion energy accompanied with the phase transfer of the molecule [90-92].

In the present study, we used the water accessible surface area. The structure of each of protein and ligands was taken as being equivalent to that in the complex, and the value of ASA for each of the complex and its components was calculated separately with a water probe of 1.4 Å radius and

Bondi's van der Waals atomic radius [93] of each element. For the difference between before and after complex formation, ΔASA was defined as in Eq. 4. For ligand molecules, ΔASA values thus calculated for ten snapshot structures were averaged to use in the QSAR analyses.

$$\Delta ASA = ASA(\text{complex}) - [ASA(\text{protein}) + ASA(\text{ligand})] \quad (4)$$

According to this definition, the greater the common contact area between components of the complex (the greater the total disappearing surface area of components), the lower is the ASA value of the complex and the more negative is the ΔASA value. This definition is reflecting the difference in the entropy and dispersion energy mentioned above, and the common contact area for components is half the $-\Delta ASA$ value. Although the ΔASA value is calculated on the basis of surface area on which water molecules are accessible, it does not mean the participation of water molecules in the complex formation. It is the parameter for the difference in the "accessible surface" defined just with the "(de)hydration" model. The ΔASA value for each of ligands is listed in Table 1.

It should be mentioned that the ΔASA value can be calculated and assigned for each of the amino acid residues/side chains because the structural coordinates of complexes as well as each of the protein and ligands are available in this study. Which residues of the protein are brought contacted after the complex formation can be detected and how intense is each of the contacts can be estimated. The analyses using these types of ΔASA information will be described in Section 5.

4.3. New QSAR Descriptors for Electronic Contribution

The ligand binding energy is defined by Eq. 5. $E(\text{complex})$ and $E(\text{protein})$ in Eq. 5 are calculated on the basis of Eq. 2.

$$\Delta E_{\text{ligand}} = E(\text{complex}) - [E(\text{protein}) + E(\text{ligand})] \quad (5)$$

The inter-fragment interaction energy, IFIE defined by Eq. 3, was calculated among 199 fragments in total, i.e., 198 (99×2) amino acid residues and the inhibitor molecule treated as a single "fragment". In the current case, IFIE(i), is defined in Eq. 6.

$$\text{IFIE}(i) = \text{IFIE}(i)_A + \text{IFIE}(i)_B \quad (6)$$

IFIE(i)_A and IFIE(i)_B express the interaction energies between CUI and the i -th FMO fragment in A and B chains of HIV-1 PR, respectively. As observed in X-ray crystallographic structures of some complexes including 1QBS (for DMP323), no water molecule participates in the interaction of CUI with the active site residues of HIV-1 PR in the complex [10, 18]. Therefore, we ignored the water molecules in the IFIE calculation. ΔE_{ligand} and IFIE values for important ligands are indicated in Table 1.

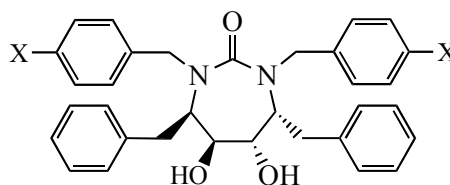
The charge differences for the i -th amino acid residues, Δq [residue(i)], and the ligand, $[\Delta q$ (ligand)], between before and after the complex formation are defined in Eqs. 7 and 8, respectively.

$$\Delta q [\text{residue}(i)] = q [\text{residue}(i)]^{\text{complex}} - q [\text{residue}(i)]^{\text{free}} \quad (7)$$

$$\Delta q [\text{ligand}] = -\sum \Delta q [\text{residue}(i)] \quad (8)$$

In these equations, q [residue(i)]^{free} and q [residue(i)]^{complex} are the total Mulliken net atomic charges [94] inside the i -th amino acid residue in the free protein and the CUI-complexed state, respectively. Similar to IFIE (i) in Eq. 6, Δq

Table 1. Structure and Descriptors of CUIs Used in Eqs. 9-13



Compound		p <i>K</i> _i			Δ <i>ASA</i> ^b	Δ <i>E</i> _{ligand} ^c	IFIE ^c				
		Obsd.	Calcd. (Eq. 9)	Calcd. (Eq. 13)			Σ	(25)	(30)	(31)	(50)
No.	X										
5 ^a	CH ₂ OH	9.47	9.22	9.40	-1202	-80.5	-118.1	-65.6	-34.0	-17.5	-6.72
7	H	8.52	8.62	8.90	-1102	-47.3	-77.7	-65.2	-3.07	-6.24	-11.1
8	F	8.85	8.42	8.46	-1123	-47.1	-76.5	-68.0	-2.42	0.346	-9.73
9	Cl	8.28	7.65	7.75	-1173	-39.6	-70.7	-71.6	0.03	0.606	-9.09
10	Br	7.57	7.69	7.57	-1201	-45.9	-76.0	-72.4	1.33	2.09	-8.05
11	CF ₃	7.29	7.67	7.30	-1242	-53.5	-81.4	-76.8	3.44	-2.85	-4.79
12	CN	7.28	7.81	7.50	-1203	-49.2	-76.3	-74.3	9.59	-4.77	-3.66
13	CH ₃	8.24	7.72	7.79	-1195	-45.5	-75.9	-65.6	-0.25	-1.36	-11.6
14	OCH ₃	6.80	7.72	7.87	-1229	-52.1	-83.0	-64.7	-14.7	1.75	-9.93
15	OCH ₂ Ph	6.27	5.80	6.01	-1430	-48.3	-85.5	-73.0	-10.4	-1.24	-8.07
16	OH	9.92	9.60	9.62	-1122	-73.3	-106.6	-67.1	-5.29	-31.3	-10.4
17	NO ₂	7.50	8.21	7.85	-1189	-55.4	-86.2	-79.3	4.99	-6.52	-4.02
18	NH ₂	8.96	8.84	8.94	-1159	-63.3	-96.6	-63.6	-11.2	-14.9	-11.4
Variance (kcal ² /mol ²)						128	176	23.9	120	85.8	7.12

a. DMP323.

b. in Å².

c. in kcal/mol (FMO/HF/6-31G).

[residue(*i*)] expresses the sum of the values of the *i*-th amino acid residues in the A and B chains.

The electronic descriptors for the individual Asp25 and Asp25' residues should be different from each other because of the use of the "unsymmetrical" model for their dissociation state as mentioned in Section 2.4. However, no constraint is required, because such descriptors as IFIE(*i*) and Δ*q* [residue(*i*)] are in fact defined as the sum of values for residues *i* and *i*'. The same token should apply also to Δ*ASA* values for residues (see below).

4.4. QSAR Analyses with Δ*ASA* and New Electronic Descriptors

We performed regression analyses using new parameters. First, Eq. 9 was formulated using Δ*E*_{ligand} and Δ*ASA* values listed in Table 1.

$$pK_i = -0.0445 (\pm 0.0307) \Delta E_{\text{ligand}} + 8.75 \cdot 10^{-3} (\pm 4.30 \cdot 10^{-3}) \Delta ASA + 16.1 (\pm 5.79)$$

$$n = 13, r = 0.881, s = 0.555, S_{cv} = 0.655, F = 17.3 \quad (9)$$

In this and the following correlation equations, *s* is the standard deviation of the correlation, *S*_{cv} is the standard deviation of the leave-one-out (cross-validated) prediction, *F* is the ratio of regression and residual variances, and the figures in parentheses are the 95% confidence intervals. The coeffi-

cient of Δ*E*_{ligand} in Eq. 9 is negative as expected. It indicates that stabilization of the binding energy increases the inhibitory potency. The positive coefficient of Δ*ASA* suggests that the inhibitory potency increases with decreasing the intermolecular contact area. This observation is not necessarily as we originally anticipated, and will be discussed in detail in Section 5.

Fig. (7) [95] patterns the IFIE value for each of 99 fragments of 13 CUI analogs. Some fragments are observed to show appreciable electronic interaction but others are not. Among them, fragments 25, 30, 31, and 50 are most significant in governing variations of the interaction with the ligand. In Table 1, IFIE values only for these fragments are indicated. As shown in Fig. (8), variance of IFIE(25), IFIE(30), IFIE(31), and IFIE(50) over 13 analogs are the broadest four among IFIE values of 99 fragment pairs.

Because inter-fragment interaction energy (IFIE) corresponds with the contribution of each fragment to the total protein–ligand interaction energy, it is reasonable to expect the good quality of correlation for Eq. 10 in which the slope of ΣIFIE term is close to unity (Σ: summation over 99 fragments).

$$\Delta E_{\text{ligand}} = 0.838 (\pm 0.106) \sum \text{IFIE} + 17.7 (\pm 9.16)$$

$$n = 13, r = 0.982, s = 2.30, S_{cv} = 2.40, F = 303 \quad (10)$$

Fig. (7). Interaction pattern of Compounds **5** and **7-18** with HIV-1 PR. IFIEs colored in red and blue represent stabilized and destabilized interactions, respectively, and the deepness of the hue indicates the strength of the interaction [IFIE(*i*; *i* = 1~99) is plotted as abscissa]. The color coding of the compound name column shows the hue for \sum IFIE. The illustration was created with BioStation Viewer [96]. For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.

Fig. (8). Variance, $\Sigma(\text{IFIE}(i) - \langle \text{IFIE}(i) \rangle)^2$, of IFIE values among thirteen CUI analogs for some important FMO fragments.

The “slight” deviation of the slope from unity arises from a “slight” difference of environmental electrostatic potential (Eq. 3) of the protein between complexed and free states. The conformation of the complexed protein is defined by crystallography and further energy-minimized, but that of the free protein is not. Conditions for electrostatic interaction of the free protein could differ from those for the energy-minimized state.

The contribution of four most widely varied fragments (in terms of variance) together to \sum IFIE is expressed as Eq. 11 with its correlation quality almost equivalent to that of Eq. 10.

$$\begin{aligned} \sum \text{IFIE} &= 0.872(\pm 0.123) [\text{IFIE}(25) + \text{IFIE}(30) + \text{IFIE}(31) \\ &+ \text{IFIE}(50)] - 8.01(\pm 11.1) \\ n &= 13, r = 0.978, s = 2.99, S_{\text{cv}} = 3.25, F = 244 \end{aligned} \quad (11)$$

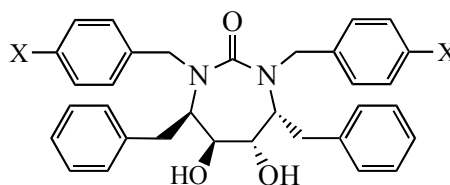
The value of IFIE(30) and IFIE(31) is more positive in average than that of IFIE(25), while their variance is much greater than that of IFIE(25) as well as IFIE(50) as shown in Fig. (8). Accordingly, the sum of four IFIE values in Eq. 11 is further replaceable by the sum of IFIE(30) and IFIE(31) to give Eq. 12 without losing much in the correlation quality.

$$\begin{aligned} \sum \text{IFIE} &= 0.760(\pm 0.145) [\text{IFIE}(30) + \text{IFIE}(31)] - 77.3(\pm 2.89) \\ n &= 13, r = 0.961, s = 4.00, S_{\text{cv}} = 4.17, F = 132 \end{aligned} \quad (12)$$

Eqs. 10 and 12 together indicate that variations in the binding energy are governed mostly by variations in the IFIE value of two (four) fragments, 30 (30'), and 31 (31'). Replacement of ΔE_{ligand} in Eq. 9 with \sum IFIE using Eqs. 10 and 12 gives Eq. 13.

$$\begin{aligned} \text{p}K_i &= -0.0334(\pm 0.0176) [\text{IFIE}(30) + \text{IFIE}(31)] \\ &+ 9.06 \cdot 10^{-3} (\pm 3.77 \cdot 10^{-3}) \Delta \text{ASA} + 18.6(\pm 4.54) \\ n &= 13, r = 0.914, s = 0.475, S_{\text{cv}} = 0.564, F = 25.4 \end{aligned} \quad (13)$$

The quality of correlation for $\text{p}K_i$ expressed in Eq. 9 is somewhat improved in Eq. 13 in spite of the “reduction” for the electronic interaction term. Eq. 13 clearly indicates that variations in the electronic interaction between ligand and protein is primarily governed by those between ligand and fragments 30 and 31. By the definition of the fragmentation shown in Fig. (6), IFIE(30) consists of the hydrogen-bonding and electrostatic interaction energies working with the side-chain and amide NH of Asp30 (30') and the amide C=O of Asp29 (29') with the P2/P2' moieties of CUIs. IFIE(31) con-

Table 2. Charge Redistribution Due to the Binding for the Amino Acid Residues and Ligand

Compound		Δq^b								
No.	X	(Ligand)	(Asp25)	(Ala28)	(Asp29)	(Asp30)	(Thr31)	(Gly48)	(Gly49)	(Ile50)
5 ^a	CH ₂ OH	-0.1437	0.0881	-0.0322	0.0328	0.0908	0.0285	-0.0131	-0.0351	0.0291
7	H	-0.0464	0.0941	0.0038	-0.0051	0.0086	0.0086	-0.0202	-0.0284	0.0247
8	F	-0.0171	0.1025	-0.0175	-0.0158	0.0246	-0.0015	-0.0248	-0.0379	0.0164
9	Cl	-0.0662	0.1043	-0.0151	-0.0023	0.0240	0.0004	-0.0182	-0.0379	0.0174
10	Br	-0.0769	0.0989	-0.0038	0.0044	0.0351	0.0002	-0.0186	-0.0427	0.0243
11	CF ₃	-0.0103	0.1032	-0.0255	-0.0041	0.0090	0.0008	-0.0182	-0.0402	0.0181
12	CN	-0.0356	0.1046	-0.0260	-0.0221	0.0393	-0.0005	-0.0271	-0.0307	0.0163
13	CH ₃	-0.0561	0.0851	0.0038	0.0046	0.0117	0.0082	-0.0157	-0.0393	0.0261
14	OCH ₃	-0.0546	0.0837	-0.0168	0.0246	0.0449	0.0016	-0.0144	-0.0408	0.0235
15	OCH ₂ Ph	-0.0430	0.0853	-0.0165	0.0061	0.0418	0.0044	-0.0082	-0.0419	0.0210
16	OH	-0.0891	0.1029	-0.0496	0.0109	0.0479	0.0321	-0.0200	-0.0335	0.0135
17	NO ₂	-0.0261	0.0842	-0.0418	0.0114	0.0411	0.0016	-0.0238	-0.0318	0.0089
18	NH ₂	-0.0396	0.0811	-0.0162	-0.0011	0.0286	0.0214	-0.0209	-0.0336	0.0193
Variance (10 ⁻⁵ esu ²)		113	7.96	23.3	20.3	43.6	12.4	2.40	1.94	2.90

a. DMP323.

b. in esu (FMO/HF/6-31G).

tains energies of corresponding interactions with respective amino acid residues Thr31(31') and Asp30(30').

It should be mentioned here that we also ran the conventional regression procedure in which significance of individual IFIE terms is analyzed stepwise. Except for IFIE(30) and IFIE(31), however, the variance is not enough as to participate in explaining the residuals as indicated above for IFIE(25) and IFIE(50), and some degrees of (multiple) colinearity was revealed to exist among variables. Whereas IFIE(30) and IFIE(31) variables are almost independent (in terms of $r = 0.373$) from each other as well as from others, the addition of the terms of other IFIE variables was statistically insignificant. The "combined" variable is used in Eq. 13, because it is originated from Eqs. 10, 11, and 12, where the component IFIE terms are not weighted. In the counterpart of Eq. 13 with two individual terms, the two terms are nearly equivalently weighted (not shown). Moreover, the "combined" variable does not reduce the number of freedom, so as not to increase the statistically insignificant terms in the regression analyses, when the number of compound is limited like in the present study.

4.5. Correlation of Inhibitory Potency with Charge Transfer and Redistribution

Table 2 shows that a considerable amount of charge redistribution occurs among amino acid residues of the protease active site on the complex formation with CUI analogs.

The negative value of $\Delta q(\text{ligand})$ for every CUI analog indicates a charge transfer from the protease molecule to ligands so that the ligands in the complex are charged more negatively than those in the free state. The $\Delta q(\text{residue})$ value varies depending upon the structure (substituent) of the CUI ligand. As shown in Table 2 and Fig. (9), the variance of the $\Delta q(\text{residue})$ value for Ala28 (28'), Asp29 (29'), Asp30 (30'), and Thr31 (31') are higher than others or among the highest. The charge redistribution occurring among these residues could be decisive in governing the variation in the interaction energy between protein and each of the CUI ligands.

Among above residues, Asp30 (30') exhibits the highest variance in the $\Delta q(\text{residue})$ value. The interaction energy of CUIs with Asp30 (30') is divided and included in IFIE(30) and IFIE(31) by definition in Fig. (6), which are the highest two in terms of the variance of IFIE(fragment) as shown in Table 1 and Fig. (8). As expected, the sum of interaction energy [IFIE(30) + IFIE(31)] is fairly nicely represented by the charge transfer and redistribution occurring at the Asp30 (30') and its neighboring residues, Asp29 (29') and Thr31 (31'), as shown in Eq. 14.

$$\text{IFIE}(30) + \text{IFIE}(31) = -370(\pm 139) [\Delta q(\text{Asp29}) + \Delta q(\text{Asp30}) + \Delta q(\text{Thr31})] + 6.33(\pm 8.40)$$

$$n = 13, r = 0.871, s = 8.96, S_{cv} = 9.29, F = 34.5 \quad (14)$$

The $\Delta q(\text{Asp30})$ value alone as the independent variable does not work well, because Fragment(30) and Fragment(31)

together cover not only Asp30 (30') but also a part of Asp29 (29') and Thr31 (31'). The use of $\Delta q(\text{C=O}$ of Asp29) and $\Delta q(\text{NH}$ of Thr31) instead of $\Delta q(\text{Asp29})$ and $\Delta q(\text{Thr31})$ in Eq. 14, however, lowered the correlation quality ($r = 0.825$, $s = 10.3$). By introducing Eq. 14 into Eq. 13, Eq. 15 can be derived.

$$\begin{aligned} \text{p}K_i &= 11.4(\pm 9.59) [\Delta q(\text{Asp29}) + \Delta q(\text{Asp30}) + \Delta q(\text{Thr31})] \\ &+ 10.2 \cdot 10^{-3} (\pm 4.84 \cdot 10^{-3}) \Delta ASA + 19.80 (\pm 5.78) \\ n &= 13, r = 0.854, s = 0.609, S_{\text{cv}} = 0.834, F = 13.5 \end{aligned} \quad (15)$$

Fig. (9). Variance of charge redistribution values, $\Sigma(\Delta q_i - \langle \Delta q_i \rangle)^2$.

The correlation quality of Eq. 15 is slightly lower than that of Eq. 9. Eqs. 14 and 15 indicate that the more positive the $\Delta q(\text{residue})$ values, i.e. the higher the electron flow from residues toward ligands, the more negative is the IFIE sum as well as the interaction energy leading to the more potent inhibitory activity, under conditions where the ΔASA value is not changed much in the series of 13 compounds.

4.6. CORRELATION OF INHIBITORY POTENCY WITH CHARGE REDISTRIBUTION PATTERNS

The charge redistribution and the interaction energy for certain residue(s)/fragment(s) seem to be, however, not always "simply" related as observed for Asp30 (30') and its neighbors. Depending upon local situations, electron flow does not necessarily occur from individual residues toward the ligand. As understood from Δq values in Table 2, electron flow occurs toward Gly48 (48') and Gly49 (49') with every CUI analog and toward Ala28 (28') with most CUIs after complex formation. Even toward Asp29, the next neighbor residue of Asp30, electron flow is observed after the complex formation with six CUI analogs.

To understand what types of pattern of charge redistribution are accompanied with the complex formation, we performed the principal component analysis (PCA) [97] of the Δq variables of each CUI ligand and 19 active site residues for 13 members of CUIs. PCA involves a mathematical procedure that transforms a number of (possibly) correlated variable sets into a smaller number of uncorrelated variable sets called principal components. In the present case, the PCA procedure corresponds to treat 20 data points defined by Δq variables of 20 constituents, the ligand and 19 residues, which are scattered within a 13-dimensional hyperspace with various degrees of possible correlations mutually. The mathematical procedure starts with the formulation of

variance-covariance matrix of the original 13 sets of Δq variables.

The principal components (PCs) can be estimated from the eigenvalue and eigenvector of this matrix. The first PC axis is to explain the largest variance in the original set of multi-dimensional data points, being oriented in the direction along which data points show the largest variance. In this study, the direction showing the largest variance is supposed to pass through the $\Delta q(\text{ligand})$ data point because individual $\Delta q(\text{ligand})$ values for 13 CUIs are generally more negative than and somewhat isolated from $\Delta q(\text{residue})$ values so that the $\Delta q(\text{ligand})$ data point is located at the edge of "approximately ellipsoid shaped" batch of data points in the multi-dimensional hyperspace. The first PC axis toward the ligand data point is probably directed (as the major axis of the ellipsoid) from that for Asp25 (25') of which Δq values are most positive among other residues throughout 13 sets without broad variances. The second PC, which can explain variances of data points not explained by the first principal component, can be perpendicular to the first PC axis (ensuring independence), crossing the first axis at the point of the greatest concentration of data. The second PC is directed along the line that runs through the minor axis of the ellipsoidal batch of data points.

The score of the first PC for the k -th complex formation, $Z_1(k)$ (k : from 1 to 13), is defined as a weighted linear combination of the original Δq_i (ligand and residues, i : from 1 to 20) variables for 13 CUIs as $\sum a_i [\Delta q_i(k) - \langle \Delta q_i(k) \rangle]$, the second term in parentheses meaning the average of 13 $\Delta q_i(k)$ values. The set of weighting factor (a_i : for ligand and residues) corresponds to the eigenvector for the largest eigenvalue of the variance-covariance matrix. The score of the second PC, $Z_2(k)$, is defined similar to $Z_1(k)$ using the eigenvector belonging to the second largest eigenvalue. The Z_1 and Z_2 values, which are calculated as a linear combination of Δq , are listed with other parameters in Table 3 and shown in Fig. (10).

Fig. (10). The coefficient values of the first eigenvector obtained from principal component analysis based on a variance-covariance matrix of charge redistribution values.

The Z_1 value is a descriptor supposed to represent a component of the total electron flow directing toward the ligand during the complex formation. It explains about 67 % of the total variance in Δq values. There is a very high correlation

between “original” $\Delta q(\text{ligand})$ and “synthesized” Z_1 variables, as shown in Eq. 16.

$$Z_1 = 1.26(\pm 0.214) \Delta q(\text{ligand}) + 0.0681(\pm 0.0137) \\ n = 13, r = 0.969, s = 0.012, S_{cv} = 0.012, F = 167 \quad (16)$$

Eq. 16 is reasonable in terms of the definition of the first PC in this study. Note that a differently weighted $\Delta q(\text{ligand})$ term is included as a component in the left-hand-side Z_1 value.

The Z_2 value represents perhaps a pattern including electron efflux from the ligand and redistribution among residues upon the complex formation. The first and second PCs together are able to explain about 83% of the total variance in Δq values. Up to a level corresponding to this, two types of charge redistribution patterns are expected to explain variations in the energy difference during the complex formation. In fact, there is a considerably good correlation between PC scores and the ligand binding energy, ΔE_{ligand} , as indicated in Eq. 17.

$$\Delta E_{\text{ligand}} = 159(\pm 96.9) Z_1 + 307(\pm 198) Z_2 - 53.9(\pm 4.23) \\ n = 13, r = 0.847, s = 6.84, S_{cv} = 7.48, F = 12.7 \quad (17)$$

Replacing ΔE_{ligand} in the right hand side of Eq. 9 with Z_1 and Z_2 variables together, however, does not give significant

correlation. The use of Z_1 parameter singularly as in Eq. 18 shows an almost equivalent quality of the correlation with Eqs. 9 and 15.

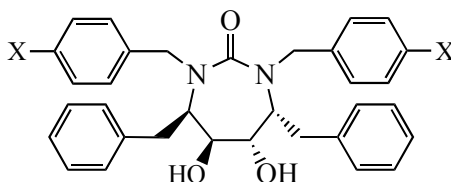
$$pK_i = -10.8(\pm 8.27) Z_1 + 9.65 \cdot 10^{-3}(\pm 4.62 \cdot 10^{-3}) \Delta ASA \\ + 19.6(\pm 5.55) \\ n = 13, r = 0.867, s = 0.584, S_{cv} = 0.781, F = 15.1 \quad (18)$$

Eqs. 15, 17 and 18 seem to show that the variations in the inhibitory potency are also governed by the electron redistribution patterns in the complex formation. Among charge transfer patterns, the one directing toward the CUI ligand is most important. It should be noted that the variations in the Z_1 variable are controlled to a large extent by the weighted sum of Δq values for Asp29, Asp30, and Thr31 for 13 CUI analogs. The reason, why the addition of Z_2 (the second PC score) term to the right hand side of Eq. 18 was insignificant ($|t| > 5\%$), is not certain at the moment.

5. PROPERTY OF ACCESSIBLE SURFACE AREA PARAMETER

We have thus far presented our new QSAR results without describing the detail of characteristics of the ΔASA variable. This is because this variable is highly thought-provoking in nature and worthy of being discussed in a separate section.

Table 3. Structure and Descriptors of CUIs Used in Eqs. 15, 17, and 18



Compound		pK _i			ΔASA ^b	ΔE _{ligand} ^c	[Δq(Asp29) + Δq(Asp30) + Δq(Thr31)] ^d	Z ₁ ^e	Z ₂ ^e
No.	X	Obsd.	Calcd. (Eq. 15)	Calcd. (Eq. 18)					
5 ^a	CH ₂ OH	9.47	9.24	9.27	-1202	-80.5	0.152053	-0.114612	-0.019614
7	H	8.52	8.67	8.84	-1102	-47.3	0.012029	0.014432	0.031339
8	F	8.85	8.40	8.36	-1123	-47.1	0.007346	0.039885	0.001971
9	Cl	8.28	8.05	8.39	-1173	-39.6	0.022057	-0.007421	0.017274
10	Br	7.57	7.96	8.28	-1201	-45.9	0.039755	-0.022059	0.021029
11	CF ₃	7.29	7.16	7.13	-1242	-53.5	0.005683	0.047424	0.001090
12	CN	7.28	7.68	7.80	-1203	-49.2	0.016619	0.019896	-0.005712
13	CH ₃	8.24	7.85	8.07	-1195	-45.5	0.024503	0.002229	0.030908
14	OCH ₃	6.80	8.04	7.90	-1229	-52.1	0.071083	-0.011546	-0.006564
15	OCH ₂ Ph	6.27	5.77	5.81	-1430	-48.3	0.052270	0.002235	-0.003327
16	OH	9.92	9.36	9.29	-1122	-73.3	0.090878	-0.045699	-0.020937
17	NO ₂	7.50	8.26	7.45	-1189	-55.4	0.054156	0.065734	-0.048261
18	NH ₂	8.96	8.51	8.35	-1159	-63.3	0.048808	0.009503	0.000805

a. DMP323.

b. in Å².

c. in kcal/mol (FMO/HF/6-31G).

d. in esu (FMO/HF/6-31G).

e. PCA score obtained from a variance-covariance matrix of charge differences.

5.1. Composition of the Accessible Surface Area Parameter

In Table 1, the ΔASA values are listed for the complex formation of 13 CUI inhibitors. The conformation of each of ligand and protein was taken to be unchanged before and after the complex formation. Although the area defined by $-\Delta ASA$ is shared by ligand and protein as two times the contacting surface area as mentioned with the definition of Eq. 4, the ΔASA value calculated for the complex with a certain ligand can be “fragmented” and assigned to each of amino acid residues accessible to the inhibitor molecule (residue ΔASA).

Fig. (11) shows the situation of ΔASA for the complex formation with Compound 5 (DMP323, X = *p*-HOCH₂-). The fragments of ΔASA value are allocated to each of residues located in the active site. The unsubstituted benzyl group (of the P1/P1' moieties) is in contact with the side chain of such hydrophobic residues as Leu23 (23'), Ile50 (50'), Pro81 (81'), Val82 (82'), and Ile84 (84') of the S1/S1' subsites of the protein (see also Fig. (2)). The *p*-HOCH₂-benzyl group (of the P2/P2' moieties) is in a close contact with such hydrophobic residues as Ile47 (47') and Ile50 (50') in addition to polar residues as Asp29 (29') and Asp30 (30') of S2/S2' subsites. The urea carbonyl oxygen and two hydroxyl groups on the seven membered ring are close to amide hydrogen atom of Ile50 (50') and the carboxyl group of Asp25 (25'), respectively, with hydrogen-bond formation. In Fig. (11), a part of certain “residue ΔASA ” values attributable to contacts brought about by hydrogen-bond formation is indicated as the “polar ΔASA ”. The contribution of energies with these hydrogen-bonds to the complex formation is, however, included within energies evaluated with the MO calculation and not necessary to consider additionally beyond the residue ΔASA values in the present study.

Fig. (11). ΔASA value of amino acid residues in the active site for DMP323.

Table 4 shows that variations of the ΔASA value for S2/S2' residues are generally greater (in terms of variance) than those for the S1/S1' residues throughout the 13 inhibitors. This feature is more apparent in Fig. (12). Variations of the ΔASA values of S2/S2' residues to the negative side are most conspicuous for Compound 15 with the OCH₂Ph substituent, the bulkiest among others. The higher variability of S2/S2' residues is expected because they are the residues to primarily fit the P2/P2' moieties of the ligand molecule where structure (substituent) variations occur. The variability

of some S2/S2' residues is lower than that of certain S1/S1' residues, reflecting the fact that ligand substituents locating at the *para*-position are able to touch subsites other than S2/S2' subsites depending upon their structure and geometry.

Fig. (12). Variations in ΔASA value of some amino acid residues (shown in the right side) among thirteen CUI analogs. ΔASA values are for the sum of correspondent residues in chains A and B.

5.2. Dual Feature of the Accessible Surface Area Parameter

As presented above, the ΔASA value is an additive-constitutive parameter but the composition is not necessarily “uniform”, since individual residue ΔASA values can be composed from components of different locational (subsite) origins. Despite the fact that the ΔASA is composite in nature, this parameter was revealed to be well correlated with some conventional steric/volume parameter of the *para*-substituents of the P2/P2' benzyl moiety as indicated in Eq. 19.

$$\Delta ASA = -396.4(\pm 63.1) MgVol - 1117.5(\pm 17.6) \\ n = 13, r = 0.972, s = 19.8, S_{cv} = 47.6, F = 191 \quad (19)$$

MgVol is a “molecular” volume parameter of substituents, shown in Table 5, calculated by Abraham and McGowan's procedure [98] similar to Bondi's van der Waals volume [93]. Eq. 19 indicates that the greater the substituent volume, the more negative the ΔASA value so that the wider the contacting area. Deleting the most isolated compound 15, the quality of correlation was slightly changed ($r = 0.956, s = 13.5$) with a steeper slope of the *MgVol* term (-600.4). The use of other volume parameters does not change the situation much. We examined the relationship of pK_i with ΔE_{ligand} and *MgVol* similar to that represented by Eq. 9. The correlation was poorer (not shown; $r = 0.828, s = 0.656$). The significance of Eq. 19 means, however, that the ΔASA value has a dual facet. It represents the difference in the accessible surface by definition, but, in addition, it is linearly related to the “molecular” volume of substituents.

Table 4. Variations in ΔASA Value for Amino Acid Residues Among Complexes with 13 CUI Analogs

Contact Residue	Max. Value ^a	Min. Value ^a	Max. - Min. ^a	Average ^a	Variance ^b	Subsite
Arg8 (8')	-11.85	-24.31	12.46	-17.61	11.40	S3/S3'
Leu23 (23')	-6.83	-12.88	6.05	-10.16	3.52	S1/S1'
Asp25 (25')	-23.03	-26.95	3.92	-25.08	1.39	Catalytic triads
Gly27 (27')	-18.82	-27.40	8.58	-23.55	8.21	Catalytic triads
Ala28 (28')	-42.68	-56.58	13.89	-49.54	10.75	S2/S2'
Asp29 (29')	-18.81	-56.53	37.72	-33.68	86.89	S2/S2'
Asp30 (30')	-14.27	-45.69	31.42	-21.21	56.39	S2/S2'
Val32 (32')	-4.98	-7.64	2.65	-6.63	0.68	S2/S2'
Leu47 (47')	-27.15	-74.75	47.60	-37.96	144.53	S2/S2'
Gly48 (48')	-38.47	-58.17	19.71	-42.57	23.57	-
Gly49 (49')	-33.68	-42.47	8.79	-37.72	7.79	-
Ile50 (50')	-45.31	-58.26	12.95	-51.22	14.22	S1/S1'
Leu76 (76')	-0.03	-2.43	2.40	-1.09	0.47	S2/S2'
Pro81 (81')	-15.21	-21.03	5.82	-17.72	2.92	S1/S1'
Val82 (82')	-35.13	-42.77	7.64	-39.66	6.10	S1/S1'

a. in \AA^2 .b. in \AA^4 .Table 5. ΔASA and $MgVol$ Values for Eq. 19

Compound		ΔASA ^b		$MgVol$
No.	X	Caled.		
		(Eq. 4)	(Eq. 19)	
5 ^a	CH ₂ OH	-1202	-1205	0.2210
7	H	-1102	-1127	0.0222
8	F	-1123	-1134	0.0400
9	Cl	-1173	-1175	0.1440
10	Br	-1201	-1195	0.1970
11	CF ₃	-1242	-1203	0.2160
12	CN	-1203	-1188	0.1770
13	CH ₃	-1195	-1182	0.1630
14	OCH ₃	-1229	-1205	0.2210
15	OCH ₂ Ph	-1430	-1445	0.8290
16	OH	-1122	-1149	0.0810
17	NO ₂	-1189	-1195	0.1960
18	NH ₂	-1159	-1166	0.1220

a. DMP323.

b. in \AA^2 .

We assume that this observation is originated in the neglect of relaxation process of the protein and inhibitor in the complex formation in defining Eq. 4. Whereas the $ASA(\text{complex})$ value is calculated at the geometry of the complex optimized (energy-minimized) by MD, each of the $ASA(\text{protein})$ and $ASA(\text{ligand})$ is not at their own individually optimized structure, but on the basis of optimized structure under conditions existing as the

complex. The relaxation energy of the protein and inhibitor is probably nearly proportional to the common contact area between the two in the complex. It is interesting to observe such a liner relationship as Eq. 19 between complicatedly puckered contact surface and a bulk volume.

In general, two factors could be included in the ΔASA parameter term. The one represents the destabilization due to the geometry distortion of the protein and inhibitor induced by the "virtual" complex formation without considering the relaxation process. The other is the stabilization due to a change of hydrophobic energy. Because the participation of water molecules in the complex formation can be neglected in this study, the above mentioned stabilization effect need not be considered. The positive ΔASA term appearing in Eqs. 9, 13, 15 and 18 indicates that the destabilization effect due to the steric distortion is at work to reduce the pK_i value for the complex formation.

6. QSARs OBTAINED WITH NEW PARAMETERS AND THE COMPARISON WITH CLASSICAL CORRELATION

6.1. Comparison Among QSARs with MO and ASA Parameters

From above sections, QSAR correlation equations for pK_i were picked up and lined up below.

$$pK_i = -0.0445(\pm 0.0307) \Delta E_{\text{ligand}} + 8.75 \cdot 10^{-3} (\pm 4.30 \cdot 10^{-3}) \Delta ASA + 16.1 (\pm 5.79) \\ n = 13, r = 0.881, s = 0.555, S_{cv} = 0.655, F = 17.3 \quad (9)$$

$$pK_i = -0.0334(\pm 0.0176) [\text{IFIE}(30) + \text{IFIE}(31)] \\ + 9.06 \cdot 10^{-3} (\pm 3.77 \cdot 10^{-3}) \Delta ASA + 18.6 (\pm 4.54) \\ n = 13, r = 0.914, s = 0.475, S_{cv} = 0.564, F = 25.4 \quad (13)$$

$$pK_i = 11.4(\pm 9.59) [\Delta q(\text{Asp29}) + \Delta q(\text{Asp30}) + \Delta q(\text{Thr31})] + 10.2 \cdot 10^{-3} (\pm 4.84 \cdot 10^{-3}) \Delta ASA + 19.80(\pm 5.78) \\ n = 13, r = 0.854, s = 0.609, S_{cv} = 0.834, F = 13.5 \quad (15)$$

$$pK_i = -10.8(\pm 8.27) Z_1 + 9.65 \cdot 10^{-3} (\pm 4.62 \cdot 10^{-3}) \Delta ASA + 19.6(\pm 5.55) \\ n = 13, r = 0.867, s = 0.584, S_{cv} = 0.781, F = 15.1 \quad (18)$$

Although the quality of the correlation is not excellent, it does not change much among equations obtained with different MO electronic parameters. The enzyme inhibition in terms of pK_i is correlated with use of the total binding energy, ΔE_{ligand} , and the residue–ligand interaction energy, [IFIE(30) + IFIE(31)]. The lower the energy of the complex, the higher is the enzyme inhibition. It is also related to the electron transfer occurring at residue sites, [$\Delta q(\text{Asp29}) + \Delta q(\text{Asp30}) + \Delta q(\text{Thr31})$], and the electron transfer pattern corresponding to that toward the ligand, Z_1 . The higher the electron efflux from the residues (the more positive the charge at the residue sites) and the higher the electron influx into the ligand (the more negative the charge at the ligand), the more severely is inhibited the enzyme. The first term of these equations is mutually correlated by Eqs. 10, 11, 14, 16, and 17. The size of the ΔASA term is almost unchanged among Eqs. 9, 13, 15, and 18. This could be reasonable, because the contribution of the “steric distortion” effect of the ligand (inhibitor) to the enzyme inhibition should not vary significantly depending on the change of “definition” of electronically interacting factors between enzyme and inhibitor. The higher the ΔASA value, i.e., the lower the common contact area and the steric distortion effect, the more easily occurs the enzyme inhibition. The size of the intercept is also nearly stable among four equations.

The terms in the right hand side of these equations could represent mostly the effective enthalpy change for the complex formation. Because ligands used in this study consist of a series of compounds in which a single substituent at a fixed (*para*) position is varied, they are distributed only within a limited region in the multidimensional physicochemical parameter space. Under such conditions, other effects probably effective on the complex formation such as the solvation and conformational entropic energies could be nearly linearly related to the enthalpy term(s) according to the entropy–enthalpy compensation [99]. The dependent variable, pK_i , is kinetically measured value for the enzyme inhibition. The enzyme inhibition is not totally equivalent with the binding but probably includes various physicochemical processes occurring with and without participation of substrate protein. The effect of these processes could also be subject to the entropy–enthalpy compensation. Thus, the pK_i value should require adjusting factors for being connected to the theoretically calculated enthalpy term, and the factors are probably included as in(de)crements in the regression coefficient and intercept of the correlation equations.

6.2. Comparison with the Classical QSAR

We also examined to analyze the complex formation with classical procedure using such empirical parameters as the Hammett σ and its modifications, the substituent hydrophobicity parameter π , and various types of steric parameters [100]. After trials with various combinations of parameters,

our best correlation equation capable of explaining variations in the binding is shown as Eq. 20.

$$pK_i = -1.11(\pm 0.52) \sigma^o - 3.28(\pm 0.88) MgVol + 1.35(\pm 0.49) HB + 8.80(\pm 0.30) \\ n = 12, r = 0.977, s = 0.259, S_{cv} = 0.341, F = 56.6, \\ \text{outlier} = p\text{-OCH}_3 \quad (20)$$

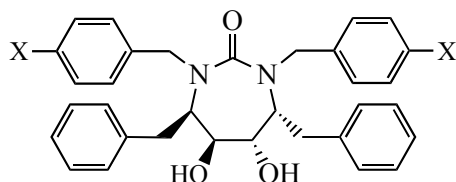
The σ^o parameter, shown in Table 6, is one of the Hammett-type constant from which the through resonance effect was supposed to be eliminated [101]. The use of this parameter is justified because the substituent location is on the benzyl group and a resonance effect of substituents does not work through the phenylene-methylene linker up to the cyclic urea nitrogen. In particular, the σ^o value should be used for the *p*-hydroxy and *p*-amino groups in compounds **16** and **18**, because σ and σ^o values of these substituents (-0.37 vs -0.13 for *p*-OH and -0.66 vs -0.38 for *p*-NH₂) are more conspicuously different than those of other substituents.

HB is an indicator variable that takes the value of unity for such hydrogen bond donor groups as *p*-CH₂OH and -OH but otherwise zero [102]. The *p*-NH₂ group was not assigned as the hydrogen bond donor. This is reasonable because (1) the “hydrogen bonding acidity” of aromatic NH₂ group is much lower than that of phenolic OH [103] and (2) the electron efflux from Asp30, $\Delta q(\text{Asp30})$, which varies most susceptibly depending upon the substituent variations, is lowest toward the *p*-NH₂ and highest toward the *p*-CH₂OH among hydrogen donors (Table 2). Perhaps the side chain COO⁻ of Asp 30 coordinates the hydrogen donor groups of ligands to make the hydrogen bond resulting in the electron efflux from the residue toward the P2 benzyl. In correlation equations in the preceding section, the *HB* indicator as an independent variable was not necessary because the hydrogen bonding effect was included in descriptors calculated by MO. In Eq. 20, the electronic effect of substituents is expressed by the through resonance-free σ^o and the additional hydrogen bonding effect, *HB*, for OH and CH₂OH substituents. *MgVol* is the McGowan molar volume used in Eq. 19 and the negative term represents that the lower the steric distortion, the higher is the enzyme inhibition. Overall, the correspondence of classical parameters with new QSAR descriptors is regarded to be very good. It is also to be mentioned that neither use nor addition of the hydrophobicity π term was significant in Eq. 20 as expected by the fact that the participation of water molecules can be neglected in the process of complex formation.

Unfortunately, the methoxy compound **14** was not included in Eq. 20. The reason is not clear but this compound always takes the most deviant behavior although it is included in Eqs. 9, 13, 15, and 18. Although the number of compounds relative to the number of descriptors is low and no definite mechanistic interpretation is plausible, Eq. 20 is helpful to understand the electron flow inside the ligand/inhibitor molecule that is subject to the substituent effect. As Eq. 15 shows, the higher the electron efflux from the residue sites and the higher the electron incorporation in the ligand molecule, the higher is the enzyme inhibition. The size and sign of the σ^o term in Eq. 20 indicates that the effect of substituent is to push the incorporated charge toward the cyclic urea region. The electron flow is highest at the site of catalytic triad as shown by the value of $\Delta q(\text{Asp25})$, but it

does not undergo so well the substituent effect of P2/P2' benzyl groups (Table 2).

Table 6. Structure and Parameters of CUI Analogs Used in the Classical QSAR Eq. 20



Compound		p <i>K</i> _i		σ ^o	<i>MgVol</i>	<i>HB</i>
No.	X	Obsd.	Calcd. (Eq. 20)			
5 ^a	CH ₂ OH	9.47	9.37	0.05	0.2210	1
7	H	8.52	8.73	0.00	0.0222	0
8	F	8.85	8.48	0.17	0.0400	0
9	Cl	8.28	8.03	0.27	0.1440	0
10	Br	7.57	7.87	0.26	0.1970	0
11	CF ₃	7.29	7.51	0.53	0.2160	0
12	CN	7.28	7.46	0.69	0.1770	0
13	CH ₃	8.24	8.40	-0.12	0.1630	0
14	OCH ₃ ^b	6.80	(8.25)	-0.16	0.2210	0
15	OCH ₂ Ph	6.27	6.23	-0.13	0.8290	0
16	OH	9.92	10.02	-0.13	0.0810	1
17	NO ₂	7.50	7.25	0.82	0.1960	0
18	NH ₂	8.96	8.82	-0.38	0.1220	0

a. DMP323.

b. Outlier in Eq. 20.

6.3. Comparison with Classical QSARs Previously Published

As mentioned earlier in this article, Garg and her co-workers have been examining the QSAR of CUI series of compounds comprehensively [34-36, 104-107]. Their study covers structure-activity data not only for variously substituted P2/P2' benzyl but also for P2/P2' substituted alkyl and aryl compounds. For the DMP323 analogs including *m*- and *p*-substituted benzyl compounds [23], they formulated Eq. 21.

$$pK_i = -1.29(\pm 0.99) \sigma - 0.61(\pm 0.20) \text{Clog } P + 12.79(\pm 1.44) \\ n = 12, r = 0.922, s = 0.571, S_{cv} = 0.743, \text{outlier} = p\text{-NH}_2 \quad (21)$$

Clog P is the molecular hydrophobicity parameter, *log P*, calculated by the BioByte system [108]. In this series of compounds, structure variations occur only in the benzyl group so that variations in the *log P* (and also *Clog P*) are equivalent to those in substituent π value. In Eq. 21, H, *m*-NH₂, and *m*- and *p*-substituted NO₂, CN, OH, OCH₂Ph, and CH₂OH compounds are included, but *p*-NH₂ compound is deleted as outlier.

The same group from DuPont-Merck also studied the cyanoimine analogs of CUI (6). From 12 compounds, the substituents in which are identical to those included in Eq. 21, Garg *et al.* deleted the unsubstituted compound further as outlier and derived Eq. 22 for 11 compounds.

$$pK_i = -1.26(\pm 0.67) \sigma - 1.77(\pm 0.65) \text{MgVol} + 16.21(\pm 3.14) \\ n = 11, r = 0.919, s = 0.473, S_{cv} = 0.514, \\ \text{outliers} = \text{H}, p\text{-NH}_2 \quad (22)$$

In Eqs. 21 and 22, the negative electronic σ term is similar to each other, but the higher hydrophobicity (*log P*) is detrimental for CUI compounds, whereas the greater steric bulk of substituents is unfavorable to their cyanoimine analogs. In fact, Eq. 22 shows similar features to those of Eq. 20. Although an additional parameter term for hydrogen bond donor groups was required and the *p*-OMe compound was not included, the *p*-NH₂ compound behaves well in Eq. 20. It could be certain if the σ^o value is used in Eqs. 21 and 22, the *p*-NH₂ compound would behave regularly.

The second term of Eqs. 21 and 22 could correspond with that of the Δ*ASA* parameter in our new correlation equations, which has the dual character as being steric and/or hydrophobic. Although no participation of water molecules was assumed for the complex formation process in our QSAR model, either steric or hydrophobic parameter term could be significant depending upon how to select compounds with varying substituents under conditions without using MO derived electronic descriptors.

7. CONCLUDING REMARKS

To conclude, the present QSAR study is reasonably considered to show that the novel descriptors defined by MD-MO procedures are able to represent physicochemical features of molecules and substituents clearly. In particular, the fragmental MO parameters are very useful for QSAR of protein-ligand interactions. It is indeed noteworthy that physicochemical parameters attributable to amino acid residues, instead of those of ligands, are able to work as QSAR descriptors in analyzing the complex formation. The IFIE (Inter Fragmental Interaction Energy) and Δ*ASA* values can be used as residue/fragment parameters similar to substituent parameters in the classical QSAR.

Although we performed detailed geometry calculation with MM-MD procedures, structure fluctuations of the complex among ten snapshots can be negligible as indicated above in Section 2.4. It is true that, for the FMO calculation, we need the structure of the complex geometrically as accurate as possible. It should be admitted, however, that the effect of fluctuation on the physicochemical descriptors such as Δ*ASA*, Δ*E*_{ligand}, IFIE, and Δ*q* is not very significant.

The outcome of the present MD-MO QSAR can be regarded to conform to that of classical QSAR formulations good enough and, in fact, reinforce the mechanistic interpretation of the enzyme-ligand interaction in terms of the classical QSAR. In the present study, the number of compounds utilized for the QSAR analyses is not large and the correlation quality is not sufficiently good. Therefore, definite conclusion about the mechanism of inhibition of CUI analogs against HIV-1 PR should be carried over in future with analyses using additional compounds suitably selected and MO parameters further improved. Along with our recent

successful QSAR results for carbonic anhydrase inhibitors (work in progress), however, the new procedure that was used in this article is believed as being a very powerful tool for the advanced QSAR, especially for the protein–ligand interactions that are the very fundamentals for the biological activity. The QSAR results and discussion about inhibition of CUI analogs against HIV-1 PR in the present article supersede those about similar topics from this laboratory.

ACKNOWLEDGEMENTS

This work was supported by the Japan Science and Technology Corporation (JST-CREST), Grants-in-Aid for Scientific Research (No.18590034 and 20590036) from the Ministry of Education, Culture, Sports, Science and Technology, and the 21st Century COE Program, Human Nutritional Science on Stress Control, Tokushima, Japan. We wish to thank Professor Shigenori Tanaka (Kobe University, Japan), Dr. Kaori Fukuzawa (Mizuho Information & Research Institute, Inc, Japan), and Professor Yuji Mochizuki (Rikkyo University, Japan) for their instructive suggestions.

REFERENCES

- [1] Gallo, R.C.; Sarin, P.S.; Gelmann, E.P.; Robert-Guroff, M.; Richardson, E.; Kalyanaraman, V.S.; Mann, D.; Sidhu, G.D.; Stahl, R.E.; Zolla-Pazner, S.; Leibowitch, J.; Popovic, M. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science*, **1983**, *220*, 865-876.
- [2] Barre-Sinoussi, F.; Chermann, J.C.; Rey, F.; Nugeyre, M.T.; Chamaret, S.; Gruest, J.; Dautet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, **1983**, *220*, 868-871.
- [3] Kohl, N.E.; Emini, E.A.; Schleif, W.A.; Davis, L.J.; Heimbach, J.C.; Dixon, R.A.F.; Scolnick, E.M.; Sigal, I.S. Active human immunodeficiency virus protease is required for viral infectivity immunodeficiency. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 4686-4690.
- [4] Applet, K. Crystal structures of HIV-1 protease-inhibitor complexes. *Perspect. Drug Discov. Des.*, **1993**, *1*, 23-48.
- [5] Erickson, J.W. Design and structure of symmetry-based inhibitors of HIV-1 protease. *Perspect. Drug Discov. Des.*, **1993**, *1*, 109-128.
- [6] Huff, J.R. Discovery and clinical development of HIV-1 protease inhibitors. *Adv. Protein Chem.*, **2001**, *56*, 213-251.
- [7] Flexner, C. HIV drug development: the next 25 years. *Nat. Rev. Drug Discov.*, **2007**, *6*, 959-964.
- [8] Emmelkamp J.M.; Rockstroh, J.K. CCR5 antagonists: comparison of efficacy, side effects, pharmacokinetics and interactions—review of the literature. *Eur. J. Med. Res.*, **2007**, *12*, 409-417.
- [9] Deeks, S.G.; Kar, S.; Guberman, S.I.; Kirkpatrick, P. News and analysis; Raltegravir. *Nat. Rev. Drug Discov.*, **2008**, *7*, 117-118.
- [10] Lam, P.Y.S.; Ru, Y.; Jadhav, P.K.; Aldrich, P.E.; DeLucca, G.V.; Eyermann, C.J.; Chang, C.H.; Emmett, G.; Holler, E.R.; Daneker, W.F.; Li, L.; Confalone, P.N.; McHugh, R.J.; Han, Q.; Li, R.; Markwalder, J.A.; Seitz, S.P.; Sharpe, T.R.; Bachelier, L.T.; Rayner, M.M.; Klabe, R.M.; Shum, L.; Winslow, D.L.; Kornhauser, D.M.; Jackson, D.A.; Viitanen, S.E.; Hodge C.N. Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas. *J. Med. Chem.*, **1996**, *39*, 3514-3525.
- [11] Navia, M.A.; Fitzgerald, P.M.; McKeever, B.M.; Leu, C.T.; Heimbach, J.C.; Herber, W.K.; Sigal, I.S.; Darke, P.L.; Springer, J.P. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature*, **1989**, *337*, 615-620.
- [12] Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B.K.; Baldwin, E.; Weber, I.T.; Selk, L.M.; Clawson, L.; Schneider, J.; Kent, S.B. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science*, **1989**, *245*, 616-621.
- [13] Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J.R.; Whittle, P.J.; Danley, D.E.; Geoghegan, K.F.; Hawrylyk, S.J.; Lee, S.E.; Scheld, K.G.; Hobart, P.M. X-ray analysis of HIV-1 proteinase at 2.7 Å resolution confirms structural homology among retroviral enzymes. *Nature*, **1989**, *342*, 299-302.
- [14] Brik, A.; Wong, C.-H. HIV-1 protease: mechanism and drug discovery. *Org. Biomol. Chem.*, **2003**, *1*, 5-14.
- [15] Das, A.; Prashar, V.; Mahale, S.; Serre, L.; Ferrer, J.L.; Hosur, M.V. Crystal structure of HIV-1 protease *in situ* product complex and observation of a low-barrier hydrogen bond between catalytic aspartates. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 18464-18469.
- [16] Babine, R.E.; Bender, S.L. Molecular recognition of protein–ligand complexes: Applications to drug design. *Chem. Rev.*, **1997**, *97*, 1359-1472.
- [17] Morse, G.D.; Catanzaro, L.M.; Acosta, E.P. Clinical pharmacodynamics of HIV-1 protease inhibitors: use of inhibitory quotients to optimise pharmacotherapy. *Lancet Infect. Dis.*, **2006**, *6*, 215-225.
- [18] Lam, P.Y.; Jadhav, P.K.; Eyermann, C.J.; Hodge, C.N.; Ru, Y.; Bachelier, L.T.; Meek, J.L.; Otto, M.J.; Rayner, M.M.; Wong, Y.N.; Chang, C.H.; Weber, P.C.; Jackson, D.A.; Sharpe, T.R.; Viitanen, S.E. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science*, **1994**, *263*, 380-384.
- [19] De Lucca, G.V.; Viitanen, S.E.; Lam, P.Y.S. Cyclic HIV protease inhibitors capable of displacing the active site structural water molecule. *Drug Discov. Today*, **1997**, *2*, 6-18.
- [20] Roberts, N.A.; Martin, J.A.; Kinchington, D.; Broadhurst, A.V.; Craig, J.C.; Duncan, I.B.; Galpin, S.A.; Handa, B.K.; Kay, J.; Kröhn, A.; Lambert, R.W.; Merrett, J.H.; Mills, J.S.; Parkes, K.E.B.; Redshaw, S.; Ritchie, A.J.; Taylor, D.L.; Thomas, G.J.; Machin, P.J. Rational design of peptide-based HIV proteinase inhibitors. *Science*, **1990**, *248*, 358-361.
- [21] Dorsey, B.D.; Levin, R.B.; McDaniel, S.L.; Vacca, J.P.; Guare, J.P.; Darke, P.L.; Zugay, J.A.; Emini, E.A.; Schleif, W.A.; Quintero, J.C.; Lin, J.H.; Chen, I.W.; Holloway, M.K.; Fitzgerald, P.M.D.; Axel, M.G.; Ostovic, D.; Anderson, P.S.; Huf, J.R. L-735, 524: the design of a potent and orally bioavailable HIV protease inhibitor. *J. Med. Chem.*, **1994**, *37*, 3443-3451.
- [22] Kempf, D.J.; Sham, H.L.; Marsh, K.C.; Flentge, C.A.; Betebenner, D.; Green, B.E.; McDonald, E.; Vasavanonda, S.; Saldívar, A.; Wideburg, N.E.; Kati, W.M.; Ruiz, L.; Zhao, C.; Fino, L.; Patterson, J.; Molla, A.; Plattner, J.J.; Norbeck, D.W. Discovery of ritonavir, a potent inhibitor of HIV protease with high oral bioavailability and clinical efficacy. *J. Med. Chem.*, **1998**, *41*, 602-617.
- [23] Jadhav, P.K.; Woerner, F.J.; Lam, P.Y.S.; Hodge, C.N.; Eyermann, C.J.; Man, H.-W.; Daneker, W.F.; Bachelier, L.T.; Rayner, M.M.; Meek, J.L.; Erickson-Viitanen, S.; Jackson, D.A.; Calabrese, J.C.; Schadt, M.; Chang, C.-H. Nonpeptide cyclic cyanoguanidines as HIV-1 protease inhibitors: synthesis, structure-activity relationships, and X-ray crystal structure studies. *J. Med. Chem.*, **1998**, *41*, 1446-1455.
- [24] Tucker, T.J.; Lumma, W.C., Jr.; Payne, L.S.; Wai, J.M.; De Solms, S.J.; Giuliani, E.A.; Darke, P.L.; Heimbach, J.C.; Zugay, J.A.; Schleif, W.A.; Quintero, J.C.; Emini, E.A.; Huff, J.R.; Anderson, P.S. A series of potent HIV-1 protease inhibitors containing a hydroxyethyl secondary amine transition state isostere: synthesis, enzyme inhibition, and antiviral activity. *J. Med. Chem.*, **1992**, *35*, 2525-2533.
- [25] Nugiel, D.A.; Jacobs, K.; Kaltenbach, R.F.; Worley, T.; Patel, M.; Meyer, D.T.; Jadhav, P.K.; De Lucca, G.V.; Smyser, T.E.; Klabe, R.M.; Bachelier, L.T.; Rayner, M.M.; Seitz, S.P. Preparation and structure-activity relationship of novel P1/P1'-substituted cyclic urea-based human immunodeficiency virus type-1 protease inhibitors. *J. Med. Chem.*, **1996**, *39*, 2156-2169.
- [26] Romines, K.R.; Morris, J.K.; Howe, W.J.; Tomich, P.K.; Horng, M.M.; Chong, K.T.; Hinshaw, R.R.; Anderson, D.J.; Strohbach, J.W.; Turner, S.R.; Mizzak, S.A. Cycloalkylpyranones and cycloalkyldihydropyrones as HIV protease inhibitors: exploring the impact of ring size on structure-activity relationships. *J. Med. Chem.*, **1996**, *39*, 4125-4130.
- [27] De Lucca, G.V.; Liang, J.; De Lucca, I. Stereospecific synthesis, structure-activity relationship, and oral bioavailability of tetrahydropyrimidin-2-one HIV protease inhibitors. *J. Med. Chem.*, **1999**, *42*, 135-152.
- [28] Boyer, F.E.; Vara Prasad, J.V.; Domagala, J.M.; Ellsworth, E.L.; Gajda, C.; Hagen, S.E.; Markoski, L.J.; Tait, B.D.; Lunney, E.A.; Palovsky, A.; Ferguson, D.; Graham, N.; Holler, T.; Hupe, D.; Nouhan, C.; Tummino, P.J.; Urumov, A.; Zeikus, E.; Zeikus, G.; Gracheck, S.J.; Sanders, J.M.; VanderRoest S.; Brodfuehrer, J.

- Iyer, K.; Sinz, M.; Gulnik, S.V. 5,6-Dihydropyran-2-ones possessing various sulfonyl functionalities: potent nonpeptidic inhibitors of HIV protease. *J. Med. Chem.*, **2000**, *43*, 843-858.
- [29] Myers, A.G.; Barbay, J.K.; Zhong, B. Asymmetric synthesis of chiral organofluorine compounds: use of nonracemic fluoroiodoacetic acid as a practical electrophile and its application to the synthesis of monofluoro hydroxyethylene dipeptide isosteres within a novel series of HIV protease inhibitors. *J. Am. Chem. Soc.*, **2001**, *123*, 7207-7219.
- [30] Mühlman, A.; Classon, B.; Hallberg, A.; Samuelsson, B. Synthesis of potent C(2)-symmetric, diol-based hiv-1 protease inhibitors. Investigation of thioalkyl and thioaryl P1/P1' substituents. *J. Med. Chem.*, **2001**, *44*, 3402-3406.
- [31] Chen, X.; Kempf, D.J.; Li, L.; Sham, H.L.; Vasavanonda, S.; Wideburg, N.E.; Saldivar, A.; Marsh, K.C.; McDonald, E.; Norbeck, D.W. Synthesis and SAR studies of potent HIV protease inhibitors containing novel dimethylphenoxyl acetates as P2 ligands. *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 3657-3660.
- [32] Surleraux, D.L.; Tahri, A.; Verschuere, W.G.; Pille, G.M.; de Kock, H.A.; Jonckers, T.H.; Peeters, A.; De Meyer S, Azijn, H.; Pauwels, R.; de Bethune, M.P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C.A.; Wigerinck, P.B. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.*, **2005**, *48*, 1813-1822.
- [33] Blum, A.; Böttcher, J.; Heine, A.; Klebe, G.; Diederich, W.E. Structure-guided design of C2-symmetric HIV-1 protease inhibitors based on a pyrrolidine scaffold. *J. Med. Chem.*, **2008**, *51*, 2078-2087.
- [34] Garg, R.; Gupta, S.P.; Gao, H.; Babu, M.S.; Debnath, A.K.; Hansch, C. Comparative quantitative structure-activity relationship studies on anti-HIV drugs. *Chem. Rev.*, **1999**, *99*, 3525-3601.
- [35] Kurup, A.; Babu, M.S.; Garg, R.; Hansch, C. HIV-1 protease inhibitors: a comparative QSAR analysis. *Curr. Med. Chem.*, **2003**, *10*, 1679-1688.
- [36] Garg, R.; Bhatarai, B. A mechanistic study of 3-aminoindazole cyclic urea HIV-1 protease inhibitors using comparative QSAR. *Bioorg. Med. Chem.*, **2004**, *12*, 5819-5831.
- [37] Kellog, G.E.; Semus, S.; Abraham, D. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Des.*, **1991**, *5*, 545-552.
- [38] Debnath, A.K. Three-dimensional quantitative structure-activity relationship study on cyclic urea derivatives as HIV-1 protease inhibitors: application of comparative molecular field analysis. *J. Med. Chem.*, **1999**, *42*, 249-259.
- [39] Jayatilleke, P.N.; Nair, A.C.; Zauhar, R.; Welsh, W.J. Computational studies on HIV-1 protease inhibitors: influence of calculated inhibitor-enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models. *J. Med. Chem.*, **2000**, *43*, 4446-4451.
- [40] Schaal, W.; Karlsson, A.; Ahlsén, G.; Lindberg, J.; Andersson, H.O.; Danielson, U.H.; Classon, B.; Unge, T.; Samuelsson, B.; Hultén, J.; Hallberg, A.; Karlén, A. Synthesis and comparative molecular field analysis (CoMFA) of symmetric and nonsymmetric cyclic sulfamide HIV-1 protease inhibitors. *J. Med. Chem.*, **2001**, *44*, 155-169.
- [41] Nair, A.C.; Jayatilleke, P.; Wang, X.; Miertus, S.; Welsh, W.J. Computational studies on tetrahydropyrimidine-2-one HIV-1 protease inhibitors: improving three-dimensional quantitative structure-activity relationship comparative molecular field analysis models by inclusion of calculated inhibitor- and receptor-based properties. *J. Med. Chem.*, **2002**, *45*, 973-983.
- [42] Avram, S.; Svab, L.; Bologna, C.; Flonta, M.-L. Correlation between the predicted and the observed biological activity of the symmetric and nonsymmetric cyclic urea derivatives used as HIV-1 protease inhibitors. A 3D-QSAR-CoMFA method for new antiviral drug design. *J. Cell. Mol. Med.*, **2003**, *7*, 287-296.
- [43] Tervo, A.J.; Nyrönen, T.H.; Rönkkö, T.; Poso, A. Comparing the quality and productiveness between 3D QSAR models obtained from manual and automated alignment. *J. Chem. Inf. Sci.*, **2004**, *44*, 807-816.
- [44] Fernández, M.; Caballero, J. Modeling of activity of cyclic urea HIV-1 protease inhibitors using regularized-artificial neural networks. *Bioorg. Med. Chem.*, **2006**, *14*, 280-294.
- [45] Milac, A.-L.; Avram, S.; Petrescu, A.-J. Evaluation of a neural networks QSAR method based on ligand representation using substituent descriptors: Application to HIV-1 protease inhibitors. *J. Mol. Graph. Model.*, **2006**, *25*, 37-45.
- [46] Fedorov, D.G.; Kitaura, K. In *Modern Methods for Theoretical Physical Chemistry of Biopolymers*; Starikov, E.B., Lewis, J.P., Tanaka, S., Eds.; Elsevier: Amsterdam, **2006**, pp. 3-38.
- [47] Wilkerson, W.W.; Akamike, E.; Cheatham, W.W.; Hollis, A.Y.; Collins, R.D.; DeLuca, I.; Lam, P.Y.S.; Ru, Y. HIV protease inhibitory bis-benzamide cyclic ureas: a quantitative structure-activity relationship analysis. *J. Med. Chem.*, **1996**, *39*, 4299-4312.
- [48] McCammon, J.A.; Harvey, S.C. In *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, **1987**, pp. 1-34.
- [49] Scott, W.R.P.; Schiffer, C.A. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure*, **2000**, *8*, 1259-1265.
- [50] Meagher, K.L.; Carlson, H.A. Solvation influences flap collapse in HIV-1 protease. *Proteins*, **2005**, *58*, 119-125.
- [51] Hamelberg, D.; McCammon, J.A. Fast peptidyl cis-trans isomerization within the flexible Gly-rich flaps of HIV-1 protease. *J. Am. Chem. Soc.*, **2005**, *127*, 13778-13779.
- [52] Hornak, V.; Okur, A.; Rizzo, R.C.; Simmerling, C. HIV-1 protease flaps spontaneously close to the correct structure in simulations following manual placement of an inhibitor into the open state. *J. Am. Chem. Soc.*, **2006**, *128*, 2812-2813.
- [53] Toth, G.; Borics, A. Closing of the flaps of HIV-1 protease induced by substrate binding: a model of a flap closing mechanism in retroviral aspartic proteases. *Biochemistry*, **2006**, *45*, 6606-6614.
- [54] Hornak, V.; Simmerling, C. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov. Today*, **2007**, *12*, 132-138.
- [55] Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational simulations of HIV-1 proteases-multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.*, **2006**, *128*, 7887-7895.
- [56] Hyland, L.J.; Tomaszek, T.A.; Meek, T.D. Human immunodeficiency virus-1 protease. 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism. *Biochemistry*, **1991**, *30*, 8454-8463.
- [57] Yamazaki, T.; Nicholson, L.K.; Torchia, D.A.; Wingfield, P.; Stahl, S.J.; Kaufman, J.D.; Eyermann, C.J.; Hodge, C.N.; Lam, P.Y.S.; Ru, Y.; Jadhav, P.K.; Chang, C.H.; Weber, P.C. NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.*, **1994**, *116*, 10791-10792.
- [58] Smith, R.; Brereton, I.M.; Chai, R.Y.; Kent, S.B. Ionization states of the catalytic residues in HIV-1 protease. *Nat. Struct. Biol.*, **1996**, *3*, 946-950.
- [59] Tropsha, A.; Hermans, J. Application of free energy simulations to the binding of a transition-state-analogue inhibitor to HTV protease. *Protein Eng.*, **1992**, *5*, 29-33.
- [60] Wang, L.; Duan, Y.; Stouten, P.; Lucca, G.V.D.; Klabe, R.M.; Kollman, P.A. Does a diol cyclic urea inhibitor of HIV-1 protease bind tighter than its corresponding alcohol form? A study by free energy perturbation and continuum electrostatics calculations. *J. Comput.-Aided Mol. Des.*, **2001**, *15*, 145-156.
- [61] Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *Ab initio* molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site. *J. Am. Chem. Soc.*, **2001**, *123*, 8730-8737.
- [62] Mardis, K.L.; Luo, R.; Gilson, M.K. Interpreting trends in the binding of cyclic ureas to HIV-1 protease. *J. Mol. Biol.*, **2001**, *309*, 507-517.
- [63] Piana, S.; Carloni, P.; Parrinello, M. Role of conformational fluctuations in the enzymatic reaction of HIV-1 protease. *J. Mol. Biol.*, **2002**, *319*, 567-583.
- [64] Carloni, P. Density functional theory-based molecular dynamics of biological systems. *Quant. Struct.-Act. Relat.*, **2002**, *21*, 166-172.
- [65] Sirois, S.; Proynov, E.I.; Truchon, J.F.; Tsoukas, C.M.; Salahub, D.R. A density functional study of the hydrogen-bond network within the HIV-1 protease catalytic site cleft. *J. Comput. Chem.*, **2003**, *24*, 1110-1119.
- [66] Piana, S.; Bucher, D.; Carloni, P.; Rothlisberger, U. Reaction mechanism of HIV-1 protease by hybrid Car-Parrinello/classical MD simulations. *J. Phys. Chem. B*, **2004**, *108*, 11139-11149.
- [67] Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational simulations of HIV-1 proteases-multi-drug resistance due to

- nonactive site mutation L90M. *J. Am. Chem. Soc.*, **2006**, *128*, 7887-7895.
- [68] Aruksakunwong, O.; Wittayanarakul, K.; Sompornpisut, P.; Sanghiran, V.; Parasuk, V.; Hannongbua, S. Structural and dynamical properties of different protonated states of mutant HIV-1 protease complexed with the saquinavir inhibitor studied by molecular dynamics simulations. *J. Mol. Graphics Modell.*, **2006**, *25*, 324-332.
- [69] Suguna, K.; Padlan, E.A.; Smith, C.W.; Carlson, W.D.; Davies, D.R. Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: implications for a mechanism of action. *Proc. Natl. Acad. Sci. USA*, **1987**, *84*, 7009-7013.
- [70] Northrop, D.B. Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.*, **2001**, *34*, 790-797.
- [71] Case, D.A.; Cheatham, T.E. III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M. Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.
- [72] Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.*, **1993**, *97*, 10269-10280.
- [73] Yoshida, T.; Yamagishi, K.; Chuman, H. QSAR study of cyclic urea type HIV-1 PR inhibitors using *ab initio* MO calculation of their complex structures with HIV-1 PR. *QSAR Comb. Sci.*, **2008**, *27*, 694-703.
- [74] Fukuzawa, K.; Kitaura, K.; Nakata, K.; Nakano, T. Fragment molecular orbital study of the binding energy of ligands to the estrogen receptor. *Pure Appl. Chem.*, **2003**, *75*, 2405-2410.
- [75] Fukuzawa, K.; Kitaura, K.; Uebayashi, M.; Nakata, K.; Kaminuma, T.; Nakano, T. *Ab initio* quantum mechanical study of the binding energies of human estrogen receptor α with its ligands: an application of fragment molecular orbital method. *J. Comp. Chem.*, **2005**, *26*, 1-10.
- [76] Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Kato, A.; Nakano, T.; Tanaka, S. Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: *Ab initio* fragment molecular orbital study. *J. Comput. Chem.*, **2006**, *27*, 948-960.
- [77] Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. Molecular interactions between estrogen receptor and its ligand studied by the *ab initio* fragment molecular orbital method. *J. Phys. Chem. B*, **2006**, *110*, 16102-16110.
- [78] Yamagishi, K.; Yamamoto, K.; Yamada, S.; Tokiwa, H. Functions of key residues in the ligand-binding pocket of vitamin D receptor: Fragment molecular orbital-interfragment interaction energy analysis. *Chem. Phys. Lett.*, **2006**, *420*, 465-468.
- [79] Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Ishida, H.; Kiso, M. Why does avian influenza A virus hemagglutinin bind to avian receptor stronger than to human receptor? *Ab initio* fragment molecular orbital studies. *Biochem. Biophys. Res. Commun.*, **2006**, *351*, 40-43.
- [80] Ito, M.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. *Ab initio* fragment molecular orbital study of molecular interactions between liganded retinoid X receptor and its coactivator: roles of helix 12 in the coactivator binding mechanism. *J. Phys. Chem. B*, **2007**, *111*, 3525-3533.
- [81] Sawada, T.; Hashimoto, T.; Nakano, H.; Suzuki, T.; Suzuki, Y.; Kawaoka, Y.; Ishida, H.; Kiso, M. Why does avian influenza A virus hemagglutinin bind to avian receptor stronger than to human receptor? *Ab initio* fragment molecular orbital studies. *Biochem. Biophys. Res. Commun.*, **2006**, *355*, 6-9.
- [82] Watanabe, T.; Inadomi, Y.; Fukuzawa, K.; Nakano, T.; Tanaka, S.; Nilsson, L.; Nagashima, U. DNA and estrogen receptor interaction revealed by fragment molecular orbital calculations. *J. Phys. Chem. B*, **2007**, *111*, 9621-9627.
- [83] Watanabe, H.; Enomoto, T.; Tanaka, S. *Ab initio* study of molecular interactions in higher plant and *Galdieria partita* Rubiscos with the fragment molecular orbital method. *Biochem. Biophys. Res. Commun.*, **2007**, *361*, 367-372.
- [84] Ito, M.; Fukuzawa, K.; Mochizuki, Y.; Nakano, T.; Tanaka, S. *Ab initio* fragment molecular orbital study of molecular interactions between liganded retinoid X receptor and its coactivator; part II: influence of mutations in transcriptional activation function 2 activating domain core on the molecular interactions. *J. Phys. Chem. A*, **2008**, *112*, 1986-1998.
- [85] Iwata, T.; Fukuzawa, K.; Nakajima, K.; Aida-Hyugaji, S.; Mochizuki, Y.; Watanabe, H.; Tanaka, S. Theoretical analysis of binding specificity of influenza viral hemagglutinin to avian and human receptors based on the fragment molecular orbital method. *Comput. Biol. Chem.*, **2008**, *32*, 198-221.
- [86] Harada, T.; Yamagishi, K.; Nakano, T.; Kitaura, K.; Tokiwa, H. *Ab initio* fragment molecular orbital study of ligand binding to human progesterone receptor ligand-binding domain. *Naunyn-Schmiedeberg's Arch. Pharmacol.*, **2008**, *377*, 607-615.
- [87] Ito, M.; Fukuzawa, K.; Ishikawa, T.; Mochizuki, Y.; Nakano, T.; Tanaka, S. *Ab initio* fragment molecular orbital study of molecular interactions in liganded retinoid X receptor: specification of residues associated with ligand inducible information transmission. *J. Phys. Chem. B*, **2008**, *112*, 12081-12094.
- [88] ABINIT-MP ver 4.1 is available from the website of RSS21 project: <http://www.ciss.iis.u-tokyo.ac.jp/rss21/>
- [89] Tanford, C. In *The Hydrophobic Effect*, 2nd ed.; Wiley-Interscience: New York, **1980**, pp. 5-20.
- [90] Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature*, **1974**, *248*, 338-339.
- [91] Eisenberg, D.; McLachlan, A.D. Solvation energy in protein folding and binding. *Nature*, **1986**, *319*, 199-203.
- [92] Chuman, H.; Mori, A.; Tanaka, H.; Yamagami, C.; Fujita, T. Analyses of the partition coefficient, log *P*, using *ab initio* MO parameter and accessible surface area of solute molecules. *J. Pharm. Sci.*, **2004**, *93*, 2681-2697.
- [93] Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.*, **1964**, *68*, 441-451.
- [94] Mulliken, R.S. Electronic population analysis on LCAO-MO molecular wave functions. I. *J. Chem. Phys.*, **1955**, *23*, 1833-1840.
- [95] Amari, S.; Aizawa, M.; Zhang, J.; Fukuzawa, K.; Mochizuki, Y.; Iwasawa, Y.; Nakata, K.; Chuman, H.; Nakano, T. VISCANA: visualized cluster analysis of protein-ligand interaction based on the *ab initio* fragment molecular orbital method for virtual ligand screening. *J. Chem. Inf. Model.*, **2006**, *46*, 221-230.
- [96] BioStation Viewer ver 8.0 is available from the website of RSS21 project: <http://www.ciss.iis.u-tokyo.ac.jp/rss21/>
- [97] Leach, A.W. In *Molecular Modelling: Principles and Applications*; Longman: Essex, **1996**, pp. 451-453.
- [98] Abraham, M.H.; McGowan, J.C. The use of characteristic volumes to measure cavity terms in reversed phase liquid chromatography. *Chromatographia*, **1987**, *23*, 243-246.
- [99] Sharp, K. Entropy-enthalpy compensation: fact or artifact? *Protein Sci.*, **2001**, *10*, 661-667.
- [100] Hansch, C.; Leo. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society: Washington, D.C., **1995**.
- [101] Exner, O. In *Correlation Analysis in Chemistry*; Chapman, N. B.; Shorter, J. Eds.; Plenum Press: New York, **1978**, pp. 439-540.
- [102] Fujita, T.; Nishioka, T.; Nakajima, M. Hydrogen-bonding parameter and its significance in quantitative structure-activity studies. *J. Med. Chem.*, **1977**, *20*, 1071-1081.
- [103] Joesten, M.D.; Schaad, L.J. In *Hydrogen Bonding*; Marcel Dekker, Inc.: New York, **1974**, pp. 291-381.
- [104] Gupta, S.P.; Babu, M.S.; Garg, R.; Sowmya, S. Quantitative structure-activity relationship studies on cyclic urea-based HIV protease inhibitors. *J. Enz. Inhib.*, **1998**, *13*, 399-407.
- [105] Gupta, S.P.; Babu, M.S. Quantitative structure-activity relationship studies on cyclic cyanoguanidines acting as HIV-1 protease inhibitors. *Bioorg. Med. Chem.*, **1999**, *7*, 2549-2553.
- [106] Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S.P. A quantitative structure-activity relationship study on some HIV-1 protease inhibitors using molecular connectivity index. *Bioorg. Med. Chem.*, **2001**, *9*, 3059-3063.
- [107] Garg, R.; Patel, D. Hydrophobicity in the design of P2/P2' tetrahydropyrimidinone HIV protease inhibitors. *Bioorg. Med. Chem. Lett.*, **2005**, *15*, 3767-3770.
- [108] ClogP for windows ver. 4.0, BioByte Corp., 201 W. Fourth St., Suite 204, Claremont, CA 91711 USA.