Journal of Computer-Aided Molecular Design, **14:** 265–275, 2000.
KLUWER/ESCOM

A CoMFA analysis with conformational propensity: An attempt to analyze the SAR of a set of molecules with different conformational flexibility using a 3D-QSAR method

Keigo Gohda[a,1,*], Ichiro Mori[b,2], Daisaku Ohta[c,3] & Takeshi Kikuchi[a,4]
[a]*Computer Chemistry Unit,* [b]*Chemistry Department and* [c]*Bio-organic Research Department, International Research Laboratories, CIBA-GEIGY Japan Ltd., P.O. Box 1, Takarazuka, Hyogo 665, Japan*

Received 26 February 1999; Accepted 15 September 1999

Key words: CoMFA analysis, conformational flexibility, conformational propensity, imidazoleglycerol phosphate dehydratase, 3D-QSAR

Summary

CoMFA analysis, a widely used 3D-QSAR method, has limitations to handle a set of SAR data containing diverse conformational flexibility since it does not explicitly include the conformational entropic effects into the analysis. Here, we present an attempt to incorporate the conformational entropy effects of a molecule into a 3D-QSAR analysis. Our attempt is based on the assumption that the conformational entropic loss of a ligand upon making a ligand-receptor complex is small if the ligand in an unbound state has a conformational propensity to adopt an active conformation in a complex state. For a QSAR analysis, this assumption was interpreted as follows: a potent ligand should have a higher conformational propensity to adopt an 'active-conformation'-like structure in an unbound state than an inactive one. The conformational propensity value was defined as the populational ratio, N_{active}/N_{stable}, of the number of energetically stable conformers, N_{stable}, to the number of 'active-conformation'-like structures, N_{active}. The latter number was calculated by counting the number of conformers that satisfied the structural parameters deduced from the active conformation. A set of SAR data of imidazoleglycerol phosphate dehydratase inhibitors containing 20 molecules with different conformational flexibility was used as a training set for developing a 3D structure-activity relationship by a CoMFA analysis with the conformational propensity value. This resulted in a cross-validated squared correlation coefficient of the CoMFA model with the conformational propensity value ($R^2_{cross} = 0.640$) higher than that of the standard CoMFA model ($R^2_{cross} = 0.431$). Then we evaluated the quality of the CoMFA models by predicting the inhibitory activity for a new molecule.

Introduction

Recent advances in computer-aided molecular modeling are leading to the ultimate goals of predicting the potency of new bio-active molecules prior to synthesis and designing new therapeutic agents by computer. In the cases where the 3D structure of a target protein is determined by X-ray crystallography, spectroscopic measurements, or protein homology modeling methods, the approach of structure-based drug design can be used to design novel leads [1]. On the other hand, one must take another approach when structural information on a target protein is not available. One of the most promising methods in this case is the 3D-QSAR approach. 3D-QSAR is a new method that uses pharmacophoric or molecular shape comparison among ligands in three-dimensional space. It has been used to predict the biological activity of new com-

*To whom correspondence should be addressed. E-mail: keigo.gohda@ pharma.novartis.com.
Present addresses: [1]Research Department, Novartis Pharma K. K., P.O. Box 1, Takarazuka, Hyogo 665-8666, Japan. [2]Research Div., GlaxoWellcome K. K., 43 Wadai, Tsukuba, 300-4247, Japan. [3]College of Agriculture, Osaka Prefectural Univ., Gakuen-cho, Sakai, Osaka 599-8531, Japan. [4]Department of Chemical Technology, College of Science and Industrial Technology, Kurashiki University of Science and the Arts, 2640 Nishinoura, Tsurajima-cho, Kurashiki, Okayama 712-8505, Japan.

pounds and design novel chemical structures [2–3]. The comparative molecular field analysis (CoMFA) method is a widely used 3D-QSAR technique [4]. A standard CoMFA approach maps the steric and electrostatic potentials of ligands on a 3D grid lattice, and develops a quantitative structure-activity relationship by statistical analysis using the partial least squares (PLS) method [5]. Although it opens a new era in the QSAR research field, the methodology possesses certain limitations, as described in Cramer's first report [6].

The potential fields mapped by a CoMFA analysis principally describe the 3D shape of a pseudo-receptor derived from the steric and electrostatic complementarities of ligands. Thus, an appropriate or consistent structural matching among ligands is required to generate a reasonable QSAR model. Without knowing the binding geometry of a ligand, the alignments of pharmacophoric atoms or groups in 3D space are somewhat arbitrary. Especially, when ligands are a non-cognate series or contain very different conformational flexibility, finding an unique alignment is difficult. On the assumption that all potent ligands share a common geometry in a complex with a receptor, several ways to overcome the alignment issue have been developed. The active analogue approach is based on the systematic torsional search of molecules to find out their best matching on the pharmacophores [7]. To match with the whole molecule, not only pharmacophores, the flexible multiple-fit procedure or the field fit method have been developed [8]. DISCO is a method that focuses on hydrogen-bonding in ligand-receptor interactions by adding dummy atoms to hetero atoms as hydrogen-bond acceptor or donor sites [9]. CATALYST is used for searching 3D structural databases of compounds based on the hypothetical receptor model derived from the information of active molecules [10]. On making the hypothesis, a training set of molecules are aligned to each other according to the five relevant types of the interactions: hydrogen-bond donors and acceptors, hydrophobic groups, and groups with positive and negative charges. Multiple conformers considering the flexibility of each molecule are used in the alignments. A recently developed alignment method is FLEXS [11, 12]. In this method, a molecule to be fitted is fragmented and superimposed onto a reference molecule from an anchor fragment in a stepwise manner, considering their putative binding site geometry. During the incremental construction of the fragmented molecule, the reference molecule is assumed rigid, but the growing fragments are flexible.

When the binding mode of a ligand is known in the complex with a receptor, the experimental geometry, i.e., the bound conformation of a ligand, can be used for the alignments to directly develop a 3D-QSAR model. However, the experimental geometry does not always provide the sufficient quality of a 3D-QSAR model [13, 14]. This unsatisfactory result might be due to the lack of the entropic effects in the 3D-QSAR model. Although the CoMFA steric and electrostatic fields can represent the enthalpic nature of the interaction between ligands and receptors, the entropic nature of the interaction is not explicitly included in a CoMFA analysis. Apparently, the entropic contributions in a complex state, such as the entropy gain due to the desolvation effects in a binding site, or the entropy loss due to the decrease of rotational freedom have significant effects on ligand binding. To cope with the desolvation effects in a CoMFA analysis, several methods have been suggested, for example, by combining CLOGP values with a QSAR model, incorporating HINT parameters, or using the MLP force field as a potential field, or using a water molecule as a probe atom [15–19]. On the loss of the conformational entropy in a complex, the number of torsional degrees of freedom in a ligand is simply and widely used to estimate the conformational entropy effects. A method using the conformational entropy term defined to each torsion of a molecule was developed as torsion angle unit (TAU) theory by Hopfinger's group [20, 21]. In the TAU theory, the sum of the TAU entropy terms approximates the maximum first-nearest group neighbor conformational entropy available to the entire molecule. This term had success in a 3D-QSAR model of a renin-inhibitor system in the free energy force field (FEFF) 3D-QSAR analysis [21]. However, this method cannot be directly applied when there are no parameters for the TAU terms such as cyclic molecules. Relative to the conformational property, the strained conformational energy, which is the conformational energy difference between that in a bound state and that in a locally minimized unbound state, is used to estimate changes in conformational enthalpy of a ligand [14]. In addition, Hopfinger's group reported a novel and sophisticated method based on the atomic occupancy of each molecule in 3D space using MD conformational simulation combined with statistical analysis (i.e., 4D-QSAR analysis) [22]. In this method, no alignment rule is required to develop structure-activity relationships by use of a genetic algorithm. Using this method, they constructed 3D-QSAR models for three applications. Although this

method can handle the conformational entropy effects quantitatively, it would be difficult to interpret and use its results to design a novel molecule since the standard field analyses using steric, electrostatic, or hydrophobic potentials are not included as is done in a CoMFA analysis.

We here present an attempt to incorporate the conformational entropy effects into a CoMFA model through conformational propensity to adopt an 'active-conformation'-like structure. A set of SAR data of imidazoleglycerol phosphate dehydratase (IGPD) inhibitors comprising 20 molecules with different conformational flexibility were used as a training set to develop a 3D structure-activity relationship by a CoMFA analysis with the conformational propensity. To obtain the conformational propensity of each inhibitor, a conformational propensity (CP) value was defined by the selection using the structural parameters of an active conformation.

IGPD (E.C. 4.2.1.19) is a key enzyme in the histidine biosynthesis pathway in plants or microbes [23]. It catalyzes the dehydration of an imidazoleglycerol phosphate (IGP) to an imidazoleacetol phosphate (IAP). This enzyme is subjected to a promising target for plants or microbes growth inhibition [24–26]. Triazole phosphonate compounds with a three carbon length chain strongly inhibit IGPD enzyme [27]. However, the structure of this enzyme is unknown.

Computational procedures

Initial structures of inhibitors were constructed by the QUANTA molecular modeling package [28] with the standard geometry parameters of the CHARMm force-field [29]. The modeled structures were treated in their neutral state, i.e., the total molecular charge was zero. The full geometry optimization was performed on all the modeled structures by the PM3 method of the MOPAC 5.0 program package with the keyword PRECISE [30]. The atomic charge of the molecule was computed for the optimized structure by ab initio MO calculations, using a minimum basis-set STO-3G, by SPARTAN [31]. The charge population analysis was done by the natural atomic orbital population method. Monte Carlo Energy Minimization (MC/EM) sampling was conducted by the Conformational Search function of the QUANTA package [32, 33]. The sampling temperature was set to 300 K. The torsion angle window for conformational perturbation was 15 degrees, and the maximum root-mean-square (RMS) difference in torsion space at the Boltzmann jump step was set to 5.0 deg. The energy minimization was done by the conjugated gradient method with 0.01 kcal/Å gradient convergent criterion. Electrostatic energy was calculated by distance-dependent dielectric electrostatics with dielectric constant $\varepsilon = 2$. Non-bonding interaction was computed with 15 Å cut-off distance. CoMFA analyses were carried out by the QSAR module of the SYBYL modeling package [34]. Superimposition of all of the molecules in the CoMFA analyses was done by the Torsional Flexible Fit command in the Molecular Similarity function of the QUANTA package. The grid spacing and the size used for the CoMFA lattice were 2.0 Å and 648 points dimensions (16.188 Å $\times$ 15.942 Å $\times$ 17.104 Å), respectively. An sp3 carbon atom with $+1.0$ atomic charge was used for the calculations of the steric and electrostatic potentials with a distance-dependent dielectric constant $1/r$. The cutoff values for both steric and electrostatic interactions were $+30$ and -30 kcal/mol, respectively. For variable scaling, the CoMFA standard scaling was used [35]. The PLS analysis, and the leave-one-out cross-validation procedure were done by the SYBYL package [36]. When an extra numeric parameter was added in the PLS analysis with the CoMFA potential fields, the parameter was treated as an independent variable to extract the latent variables so that each latent variable was orthogonal in the PLS model [37]. The scaling of the extra parameters was done by the autoscaling procedure in the SYBYL package [35]. CLOGP values were calculated by Leo's method [38]. All computations were done on an INDIGO^2 workstation (Silicon Graphics Inc., U.S.A.). Synthesis of the inhibitors will be described elsewhere. The inhibitory activity was examined by the previously reported assay method [25].

Results and discussion

Concepts and computational methodology for conformational propensity

On the binding process of a ligand to a receptor, a ligand in an unbound state with ensemble conformations in equilibrium, first approaches the receptor. Among ligand-receptor complexes formed with various conformers of the ligand, the complex with the lowest free energy becomes a productive complex. The conformation of the ligand in the productive complex is called an active conformation. The binding

free energy ($\Delta G = G_{\text{complex}} - G_{\text{ligand}} - G_{\text{receptor}}$) of a ligand can be broken down into the enthalpic and entropic contributions ($\Delta G = \Delta H - T\Delta S$). The binding enthalpy ($\Delta H = H_{\text{complex}} - H_{\text{ligand}} - H_{\text{receptor}}$) is primarily as the binding energies of the steric (ΔE_{ster}) and electrostatic (ΔE_{elec}) interactions. To estimate the entropic contributions in the binding free energy, the solvating and conformational entropy effects have to be included. In a 3D-QSAR analysis, the conformation of a ligand should be superimposed as much as possible with the active conformation of a template-ligand to maximize the binding enthalpic contribution in a ligand-receptor complex. If the ligand in an unbound state has a conformational propensity to adopt a similar conformation to the active conformation, it can be assumed that the conformational entropic loss in ligand binding is small. Our attempt is based on the idea that a potent ligand should have a higher conformational propensity in an unbound state to adopt an 'active-conformation'-like structure than an inactive one. The entropic gain by the desolvation effects may not differ so much among the ligands if they are a congeneric series. In this study the conformational propensity value was defined as the populational ratio, $N_{\text{active}}/N_{\text{stable}}$, of the number of energetically stable conformers, N_{stable}, to the number of 'active-conformation'-like structures, N_{active}. This value became larger when the population of 'active-conformation'-like structures increased. Only energetically stable conformers were used in the analysis since a highly distorted conformer would not be suitable for binding a receptor protein [39]. The 'active-conformation'-like structure for each ligand was evaluated by geometric selection using the structural parameters of a defined active conformation. Experimental information and/or computationally derived models can define an active conformation. Based on the geometry of the defined active conformation, the structural parameters relevant to biological activity should be identified. This includes distances, angles, or planes among pharmacophoric atoms or groups, as are usually used in a query of the 3D database search method in a lead finding process. If the structural parameters of each stable conformer meet the structural criteria within tolerant deviations, it is counted as an 'active-conformation'-like conformer.

The flow of the CP value calculation is shown in Figure 1. First, a set of conformers was generated by the MC/EM sampling method [32, 33]. To search the conformational space as widely as possible within a reasonable CPU time, we used this sampling method which probably picks up conformations close to the lowest-energy conformation. Second, an energetically stable conformer was selected among the sampled conformers according to an energy cutoff value. This conformer was designated as a stable conformer (N_{stable}). Third, an 'active-conformation'-like structure (N_{active}) was selected from the N_{stable} conformers based on the structural parameters of an active conformation. In this study, the chair-form structure of compound **1** was used as an active conformation (Figure 2), and two inter-atomic distances among a triazole ring, a hydroxyl group and a phosphonate in the structure of compound **1** were used as the structural parameters. Justifications of the active conformation and the structural parameters used in this study are described in the following section. Finally, the CP value of each molecule was calculated by dividing the number of stable conformers (N_{stable}) into the number of 'active-conformation'-like conformers (N_{active}).

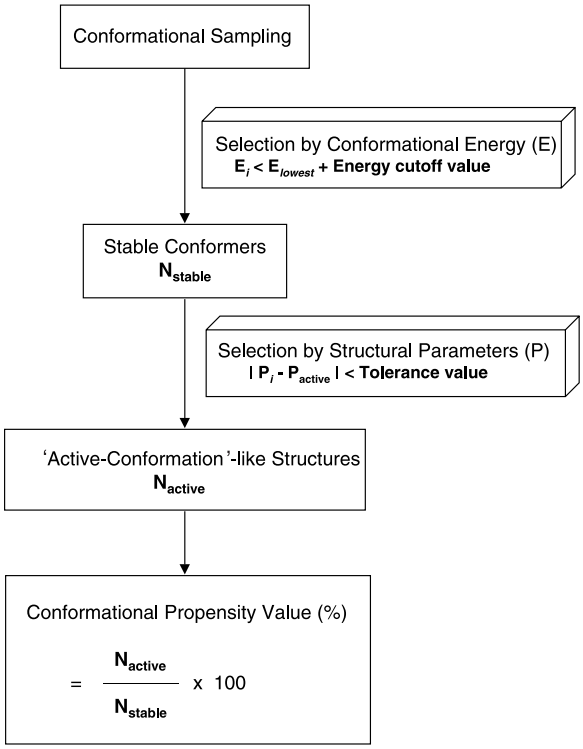

Figure 1. The flow-chart for computing the conformational propensity value.

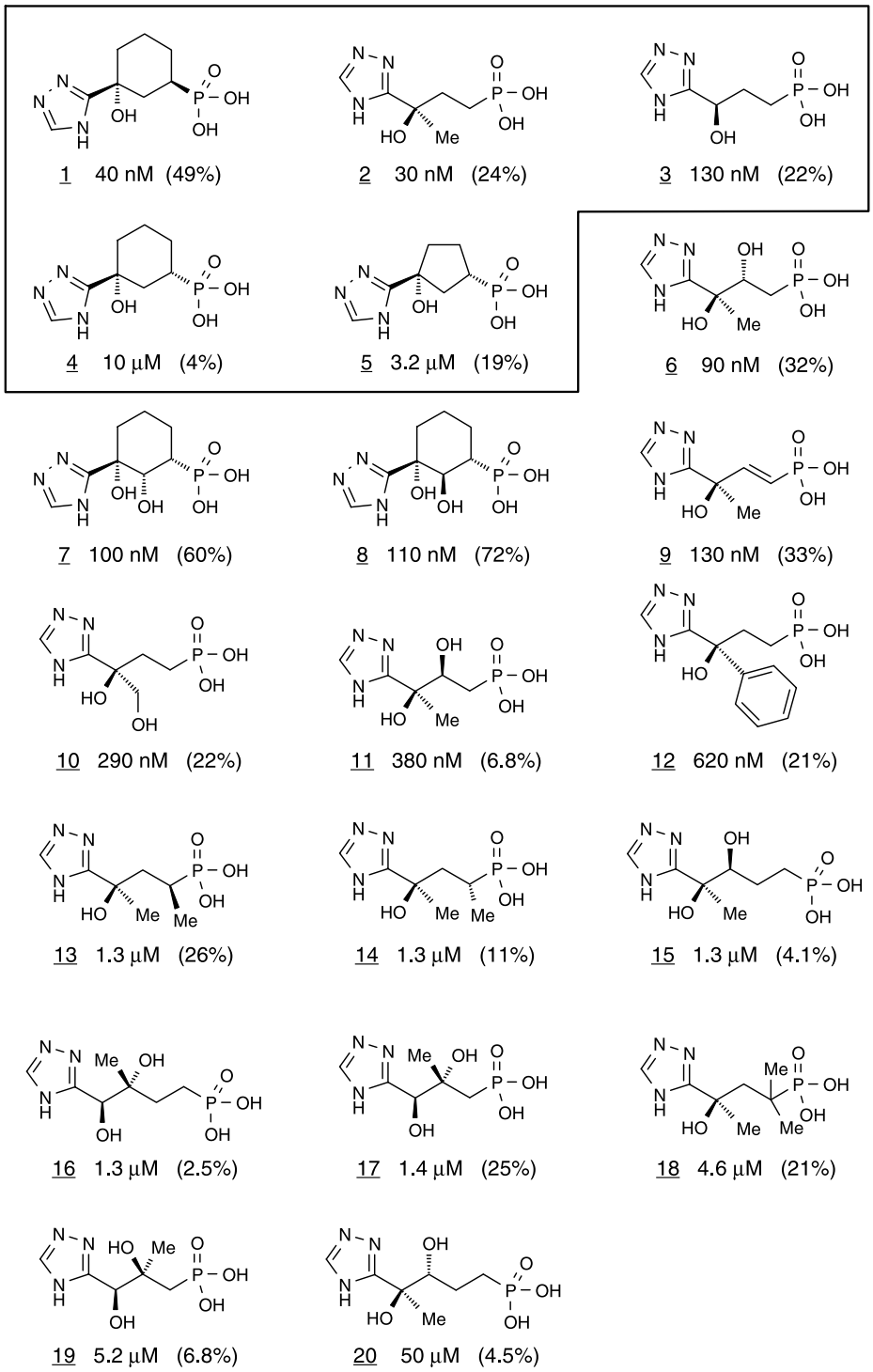

Figure 2. IGPD inhibitors with their IC_{50} values. The representative IGPD inhibitors are in the box. The conformational propensity values are in parentheses.

Selections of active conformation and structural parameters

Triazole phosphonate compounds are known as potent IGPD inhibitors [27]. Among the compounds synthesized in our laboratory, 20 inhibitors were chosen for a SAR study (Figure 2). They share a common chemical entity but their inhibitory activities vary from 30 nM to 50 μM. Among the inhibitors, the conformational entropic effects as well as the steric and electronic complementarity of the inhibitors may contribute to the differences in activities since the inhibitors possess diverse conformational flexibility. The desolvation effects are presumably similar in each inhibitor-enzyme complex since the number of functional groups and the molecular volumes of the inhibitors are almost identical. We previously determined the single crystal structure of the derivative of compound **1**, showing a chair-form of the cyclohexyl moiety in which the triazole ring and the phosphonate group are in equatorial orientations [27]. A solution structure of compound **1** observed by NMR measurements also showed the chair conformation. In addition, compound **1** is one of the most potent inhibitors that has restricted conformation. Thus, the X-ray structure of compound **1** was reasonably assumed to be a bound conformation in a complex with the enzyme, i.e., an active conformation. Three functional groups, a triazole ring, a hydroxyl group and a phosphonate, were defined as the pharmacophoric groups among the inhibitors. The importance of these functional groups in the IGPD inhibitors was reported in the SAR studies and our previous analysis of the enzyme's reaction mechanism [40]. In the chair-form of compound **1**, inter-pharmacophore distances among the triazole ring, the hydroxyl group, and the phosphonate group were used as the structural parameters for the selection of the 'active-conformation'-like structure (Figure 3). The positions of the three pharmacophoric groups were represented by three atoms in the groups; an oxygen atom in the hydroxyl group (O-atom), a carbon atom in the triazole ring (C-atom), and a phosphorus atom in the phosphonate group (P-atom). Among the inhibitors, the distance between the triazole ring and the hydroxyl group was most likely unchanged since both were the substituents on a common atom. Therefore, the structural parameters used for the selection were two interatomic distances in compound **1**, i.e., the distance between the C-atom and the P-atom [$d(C–P)_{active} = 5.2$ Å], and the distance between the O-atom and the P-atom [$d(O\text{-}P)_{active} = 4.6$ Å].

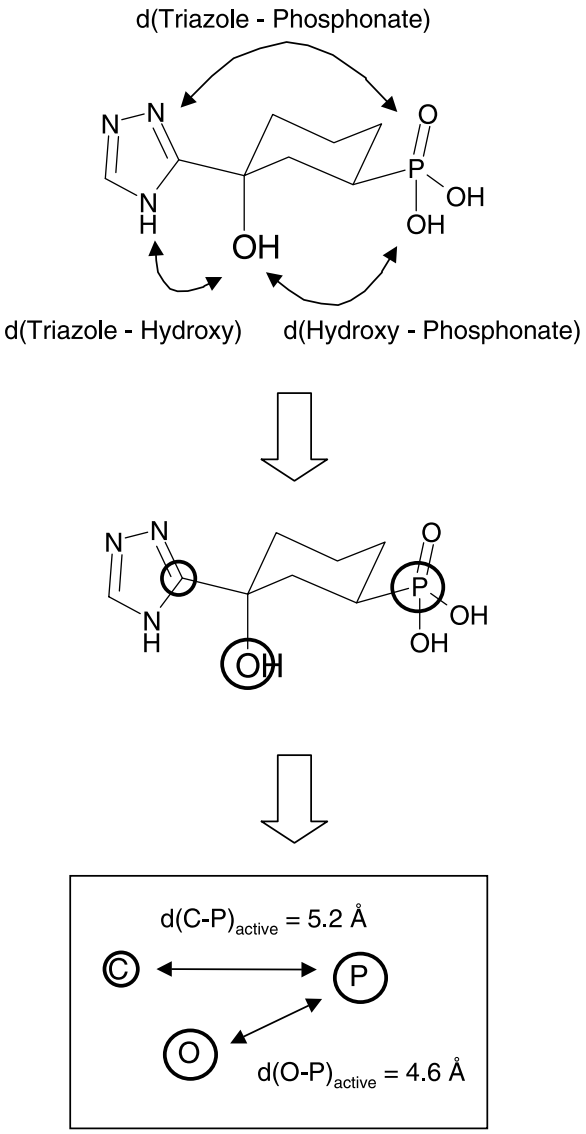

Figure 3. The structural parameters, $d(C–P)_{active}$ and $d(O–P)_{active}$, deduced from the geometric properties of compound **1**.

For the computational analysis, all the inhibitors were handled as neutral ligands (no formal charges). It is probable that the phosphonate group in the inhibitors is deprotonated and has negative charge(s) of -1 or -2 under the assay conditions [25]. However, the local pH in the binding site of the enzyme would be different from that of the solution, and non-deprotonated species might co-exist in the system. Thus, to simplify computation and also to avoid overestimation of electrostatic interaction which is sometimes seen in studies of molecular mechanics calculations with charged

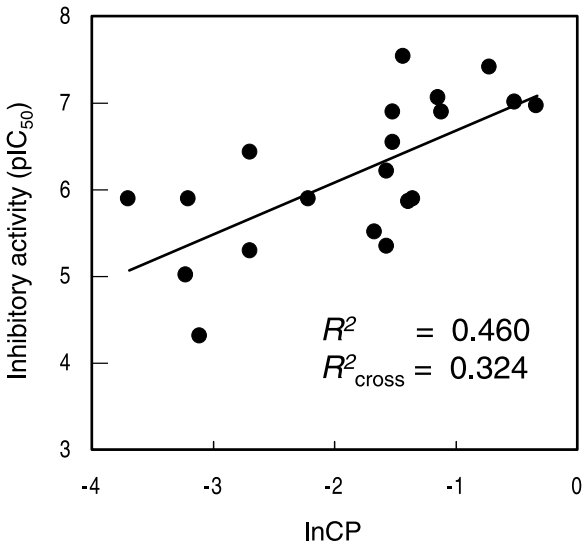

Figure 4. The correlation plot between the CP values and inhibitory activities with ±0.3 Å distance tolerance and +8 kcal/mol energy cutoff values.

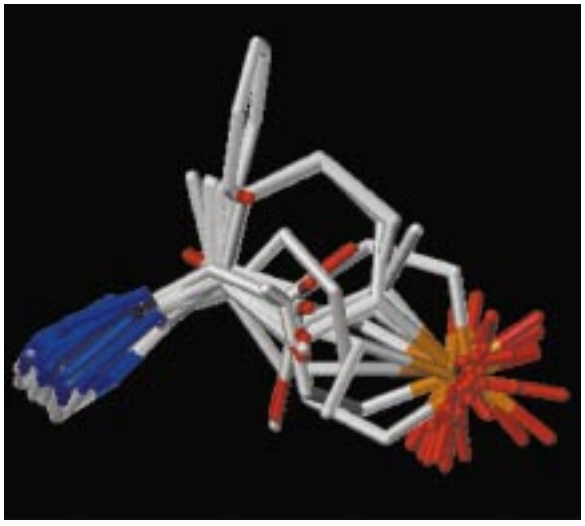

Figure 5. Superposition of 20 IGPD inhibitors.

species, only the neutral species of the inhibitors were used in this study.

Parameterization of energy cutoff value and distance tolerance value

To calculate the CP values for each ligand, the two parameter values of energy cutoff and distance tolerance, which are necessary to calculate the numbers N_{stable} and N_{active}, should be determined. Before analyzing the whole set of inhibitors, five compounds (**1**–**5**) having the simplest structure and different conformational flexibility were chosen as representatives. Three different energy cutoff values, and distance tolerance values were examined to identify an optimum combination of the two parameters, namely, +6, +8, and +10 kcal/mol energies from the lowest energy, and ±0.3 Å, ±0.5 Å, and ±0.7 Å distance tolerances for the structural parameters. A set of 400 MC/EM sampled conformations for the five inhibitors was used in the analyses. In total, nine CP values for each inhibitor were calculated, and then correlated with the inhibitory activity. For the correlation analysis, the logarithmic form of the CP value, i.e., ln CP, was used to adjust itself to an energy-like unit. Table 1 summarizes the CP values and the squared correlation coefficients (R^2) of QSARs. The combinations of ±0.3 Å distance tolerance value with the three energy cutoff values gave good correlations. To avoid chance correlations with little data, further parameterization to the energy cutoff value was done using all the inhibitors. The CP values of the other 15 inhibitors were calculated with ±0.3 Å distance tolerance value. Three compounds (**4**, **11** and **19**) were excluded from the analysis of the +6 kcal/mol cutoff correlation because their CP values were zero. As a result, the +8 kcal/mol cutoff correlation showed the highest R^2 value (= 0.461). Thus, we decided to use the combination of ±0.3 Å distance tolerance value with +8 kcal/mol cutoff value for further QSAR analyses.

CoMFA analysis with conformational propensity value

Figure 4 shows the correlation plot of the combination of '8–0.3'. This linear correlation with a moderate cross-validated coefficient (R^2_{cross}) suggested that the conformational entropic effects contributed to the inhibitory activity. Then, to develop full correlation with the enthalpic contributions, the 3D-QSAR analyses were implemented by the CoMFA method with the CP value calculated using the combination of '8–0.3', CoMFA + ln CP. In parallel, three CoMFA analyses, (1) with only CoMFA steric potential, (2) with only CoMFA electrostatic potential, and (3) by the standard CoMFA method were done as references. The inhibitors in the CoMFA analyses were aligned to the chair-form of compound **1**. The three functional groups were used for the alignments as represented by the four atoms, i.e., the three atoms used in the calculation of the CP value and the carbon atom next to the triazole ring. The alignment was done as follows [41]. The X-ray structure of compound **1** was

Table 1. Correlation analyses of conformational propensity values with inhibitory activities using nine combinations of parameters

CMPDS	IC_{50}	6–0.3[a]	6–0.5	6–0.7	8–0.3	8–0.5	8–0.7	10–0.3	10–0.5	10–0.7
1	40 nM	59%	59%	59%	49%	53%	66%	41%	51%	64%
2	30 nM	26	35	68	24	34	68	25	36	68
3	130 nM	24	33	55	22	32	55	21	32	55
4	10 μM	0	8	51	4	17	48	12	40	66
5	3.2 μM	19	69	74	19	68	74	19	68	74
R^2 (5)[b]		0.411[c]	0.254	0.027	0.678	0.090	0.147	0.714	0.127	0.138
R^2 (20)[d]		0.383[e]	–	–	0.461	–	–	0.394	–	–

[a]First digit is the energy cutoff value and the second number is the distance tolerance value. For example with '6–0.3', the energy cutoff and distance tolerance values were set to +6 kcal/mol and ±0.3 Å, respectively.
[b]Squared correlation coefficients of the correlation with the five inhibitors.
[c]Compound **4** was excluded from the correlation.
[d]Squared correlation coefficients of the correlation with all the inhibitors.
[e]Compounds **4**, **11**, and **19** were excluded from the correlation.

fixed as a template. Each inhibitor was fitted to the template structure by the rigid-body fit to maximize the overlapping on the four atoms, and the RMS deviation of the fitting was calculated. The RMS deviation value was then minimized by adjusting the rotatable bonds. Figure 5 shows the aligned structures of all the inhibitors. The standard CoMFA analyses with and without the CP value showed good correlations (Table 2). Although in the standard CoMFA analysis a good correlation $R^2 = 0.999$ was found, the cross-validated R^2 value was relatively small to be predictive enough ($R^2_{\text{cross}} = 0.431$). However, the cross-validated R^2 value increased to 0.640 with the CoMFA + ln CP analysis. It is widely accepted that a correlation with a cross-validated R^2 value greater than 0.5–0.6 would possess enough power to predict the biological activity of a new molecule [42].

To examine the stability of the CoMFA models, the number of rotatable bonds (Ntor), and the CLOGP value were used as supplemental parameters. These are the simple flexibility measure, and simple hydrophobic descriptor, respectively. The acyclic bonds between sp3-sp3 or sp3-sp2 atoms were included with the rotatable bonds, but the rotatable bonds of OH and CH_3 were not. The results in Table 2 show that no improvement in the CoMFA models with both parameters was obtained. In addition, to test the robustness of the CoMFA + ln CP model, the data set of the activity was randomly sorted in three different ways. The randomized activity data set was then used for developing a CoMFA + ln CP model. As a result, these sets provided the R^2 values in the range of 0.076 to 0.339 and the cross-validated R^2 values in the range of −0.623 to −0.002. These results suggest that the improvement of the CoMFA model by the addition of the CP value is not due to chance correlation, and that the conformational entropic effects likely influence the inhibitor-IGPD enzyme complex.

Prediction of inhibitory activity for a new molecule

To test the predictive power of the CoMFA + ln CP model, a new compound (**21**) excluded from the 3D-QSAR analyses was subjected to the prediction of its inhibitory activity (Figure 6). This compound is a derivative of compound **1** with the replacement of CH_2 at position C-4 to an oxygen atom, and showed the most potent inhibitory activity among the IGPD inhibitors used in this study ($IC_{50}^{\text{exp}} = 20$ nM). The CP value of the molecule was calculated by the same method used in the training set. Using both the CoMFA + ln CP and the standard CoMFA models, the inhibitory activities of compound **21** were predicted and compared. The predicted activity from the CoMFA + ln CP model ($IC_{50}^{\text{pred}} = 20$ nM) agreed with the experimental activity better than that of the standard CoMFA model ($IC_{50}^{\text{pred}} = 41$ nM). Figure 7 shows the steric and electrostatic potential field maps deduced from the CoMFA + ln CP model. There were no favorable or unfavorable areas relevant to the steric and electrostatic interactions around position C-4 where an oxygen atom was placed in compound **21**, suggesting that the replacement at that position might not significantly affect the inhibitory activity based on the standard CoMFA model. In fact, the prediction by the standard CoMFA model forecasted the activity of

Table 2. Statistical parameters from QSAR analyses

QSAR	SE[a]	R^2_{cross} [b]	*F* value	*p* value	R^2 [c]	SD	Param[d]	Ster[e]	Elec[e]
CoMFA + ln CP	0.619	0.640(6)[f]	467.5	0.000	0.995	0.070	0.164	0.384	0.452
CoMFA + Ntor	0.824	0.314(5)	65.4	0.000	0.959	0.202	0.073	0.374	0.553
CoMFA + CLOGP	0.819	0.317(6)	174.1	0.000	0.988	0.114	0.044	0.409	0.547
CoMFA	0.811	0.431(7)	1291.2	0.000	0.999	0.039	–	0.559	0.441
CoMFA(ster)	0.854	0.211(4)	70.0	0.000	0.949	0.217	–	1.000	–
CoMFA(elec)	0.805	0.345(5)	50.4	0.000	0.947	0.228	–	–	1.000
ln CP	0.721	0.324(1)	15.4	0.001	0.460	0.644	1.000	–	–
Ntor	0.958	−0.193(1)	1.3	0.278	0.065	0.848	1.000	–	–
CLOGP	0.967	−0.216(1)	0.5	0.497	0.026	0.866	1.000	–	–

[a]Cross-validated standard error of estimate.
[b]Cross-validated R^2.
[c]Conventional R^2.
[d]Contributions of additional parameters in QSAR equations.
[e]Contributions of potential fields in QSAR equations.
[f]Optimum numbers of components in parentheses.

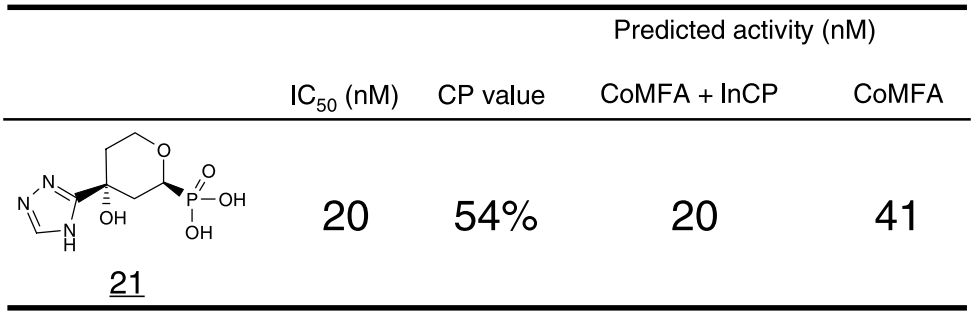

Figure 6. An IGPD inhibitor for the prediction.

compound **21** (IC_{50}^{pred} = 41 nM) to be nearly that of compound **1** (IC_{50} = 40 nM), while in reality the activity of compound **21** increased 2 times more than that of compound **1**. Compound **21** showed a bigger CP value (= 54%) than compound **1** (= 49%). The success of the prediction by the CoMFA + ln CP model might be due to the CP value. This suggests that the replacement to the oxygen atom at position C-4 made the molecules more flexible and hence more likely to take the chair conformation in an unbound state. This would lead to increased inhibition.

Based on the steric potential field map shown in Figure 7, a further activity improvement could occur by attaching larger substituents to the cyclohexyl carbon atom that faces the sterically favored region in green. In fact, there is enough space to accommodate a large substituent since the phenyl moiety of compound **12** did not completely diminish its activity. The favorable or unfavorable areas of the electrostatic interaction were mainly localized around the phosphonate group. Changing the electronic characteristics of the phosphonate group by the modification to phosphoramide or replacement with other acidic groups, such as sulfonic, carboxyl, or malonic acid groups could lead to further derivatization of the inhibitors.

Conclusions

Supplementing the CoMFA method with conformational propensity improved the quality of the 3D-QSAR model. Although the test compound was only one in this study, the CoMFA model with the CP value predicted the experimental activity more accurately than the standard CoMFA model. This result suggests that the degree of conformational propensity of each IGPD inhibitor in an unbound state contributes to the differences of the binding free energy in the complex with the enzyme. A similar observation was reported in the thermodynamic studies of a renin-inhibitor system [43]. That report concluded

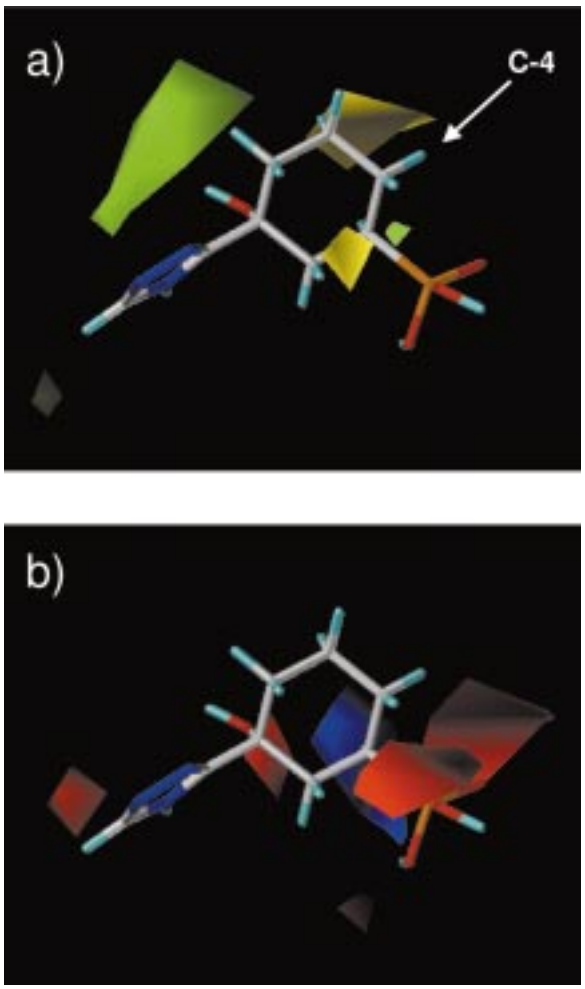

Figure 7. (a) The CoMFA steric stdev*coeff contour plot from the CoMFA + ln CP model. Sterically favored areas (contribution level of 80%) are represented by green polyhedra. Sterically disfavored areas (contribution level of 20%) are represented by yellow polyhedra. Compound **1** is represented by a wire structure. (b) The CoMFA electrostatic stdev*coeff contour plot from the CoMFA + ln CP model. Negative charge favored areas (contribution level of 80%) are represented by blue polyhedra. Negative charge disfavored areas (contribution level of 20%) are represented by red polyhedra. Compound **1** is represented by a wire structure.

that the conformation of the peptide inhibitor in solution is probably an important determinant of binding. By use of the FEFF 3D-QSAR analysis for the same renin system, it was shown that structural differences between the bound and unbound conformations of each of the ligands could affect their binding affinity and the greater similarity between the bound and unbound conformations infers smaller entropic and/or enthalpic energy binding penalties [21]. These reports support our assumption that the conformational propensity can represent the degree of entropy loss in a ligand-receptor complex.

However, the CoMFA method with conformational propensity implemented in this study has certain limitations. To obtain a set of stable conformers, we used the MC/EM sampling method. This method was originally developed to explore the conformational space of a various molecule, i.e., cyclic or acyclic molecule, effectively. Although it was successfully used to reproduce experimental binding modes of ligands in a complex with enzymes, the energetic and conformational distributions of the sampled conformers do not theoretically correspond to the Boltzmann distribution. The CP value, therefore, cannot be transformed to the entropic energetic values directly, and then the estimation of ΔG itself is not possible. Nevertheless, the MC/EM method should be quite useful for applying the calculation of the CP value since the conformers are sampled as a stable one near the lowest energy, and the CPU cost for their sampling is not expensive. In addition to the issues of the conformational sampling method, the selection of distance tolerance and energy cutoff values should be more rationalized to reduce the arbitrariness. Another limitation is that the structure of the 'active conformation' is a prerequisite for the CP calculation. For further improvement of the methodology, the entropic effects of a receptor and solvating effects in a system should be treated properly.

We are presently addressing these points and increasing the number of data sets to apply this method to other SAR results in our laboratories.

Acknowledgements

We gratefully thank professors H. Fujiwara and T. Takagi for many helpful discussions.

References

1. Bohacek, R.S., McMartin, C. and Guida, W.C., Med. Res. Rev., 16 (1996) 3.
2. Kubinyi, H., Drug Discov. Today, 2 (1997) 457.
3. Kubinyi, H., Drug Discov. Today, 2 (1997) 538.
4. Kubinyi, H. (Ed.) 3D-QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993.
5. Stahle, L. and Wold, S., Prog. Med. Chem., 25 (1988) 292.
6. Cramer, R.D., III, J. Am. Chem. Soc., 110 (1988) 5959.
7. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., in Olsen, E.C. and Christoffersen, E. (Eds.) Computer Assisted Drug Design, American Chemical Society, Washington, DC, 1979, pp. 87–112.

8. Klebe, G., In Kubinyi, H. (Ed.) 3D-QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 173–199.
9. Martin, Y.C., Bures, M.G., Danaher, E.A., DeLazzer, J., Lico, I. and Pavlik, P.A., J. Comput.-Aided Mol. Design, 7 (1993) 83.
10. Sprague, P. W., Perspect. Drug Discov. Des., 3 (1995) 1.
11. Lemmen, C. and Lengauer, T., J. Comput.-Aided Mol. Design, 11 (1997) 357.
12. Lemmen, C., Lengauer, T. and Klebe, G., J. Med. Chem., 41 (1998) 4502.
13. Klebe, G. and Abraham, U., J. Med. Chem., 36 (1993) 70.
14. DePriest, S.A., Mayer, D., Naylor, C.B. and Marshall, G.R., J. Am. Chem. Soc., 115 (1993) 5372.
15. Kellogg, G.E., Semus, S.F. and Abraham, D.J., J. Comput.-Aided Mol. Design, 5 (1991) 545.
16. Gaillard, P., Carrupt, P.-A., Testa, B. and Boudon, A., J. Comput.-Aided Mol. Design, 8 (1994) 83.
17. Masuda, T., Nakamura, K., Jikihara, T., Kasuya, F., Igarashi, K., Fukui, M., Takagi, T. and Fujiwara, H., Quant. Struct.-Act. Relat., 15 (1996) 194.
18. Ghose, A.K., Viswanadhan, V.N. and Wendoloski, J.J., J. Phys. Chem. A, 102 (1998) 3762.
19. Kim, K.H., Quant. Struct.-Act. Relat., 12 (1993) 232.
20. Koehler, M.G. and Hopfinger, A.J., Polymer, 30 (1989) 116.
21. Tokarski, J.S. and Hopfinger, A.J., J. Chem. Inf. Comput. Sci., 37 (1997) 792.
22. Hopfinger, A.J., Wang, S., Tokarski, J.S., Jin, B., Albuquerque, M., Madhav, P.J. and Duraiswami, C., J. Am. Chem. Soc., 119 (1997) 10509.
23. Ames, B.N., J. Biol. Chem., 228 (1957) 131.
24. Hawkes, T.R., Cox, J.M., Barnes, N.J., Beautement, K., Edwards, L.S., Kipps, M.R., Langford, M.P., Lewis, T., Ridley, S.M. and Thomas, P.G., Brighton Crop Protection Conference (Weeds), 1993, pp. 739–744.
25. Mori, I., Fonne-Pfister, R., Matsunaga, S., Tada, S., Kimura, Y., Iwasaki, G., Mano, J., Hatano, M., Nakano, T., Koizumi, S., Scheidegger, A., Hayakawa, K. and Ohta, D., Plant Physiol., 107 (1995) 719.
26. Lindell, S.D., Earnshaw, C.G., Wright, B.J., Carver, D.S., O'Mahony, M.J. and Saville-Stones, E.A., Bioorg. Med. Chem. Lett., 6 (1996) 547.
27. Mori, I., Iwasaki, G., Kimura, Y., Matsunaga, S., Ogawa, A., Nakano, T., Buser, H.-P., Hatano, M., Tada, S. and Hayakawa, K., J. Am. Chem. Soc., 117 (1995) 4411.
28. QUANTA v. 3.3.1., Molecular Simulation Inc., Burlington, MA, 1994.
29. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
30. Stewart, J.J.P., J. Comput. Chem., 10 (1989) 209.
31. SPARTAN v. 4.0.2., Wavefunction Co., Irvine, CA, 1995.
32. Chang, G., Guida, W.C. and Still, W.C., J. Am. Chem. Soc., 111 (1989) 4379.
33. Guida, W.C., Bohacek, R.S. and Erion, M.D., J. Comput. Chem., 13 (1992) 214.
34. SYBYL v. 6.0.2., Tripos Inc., St. Louis, MO, 1994.
35. Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., in Kubinyi, H. (Ed.) 3D-QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 443–485.
36. Wold, S., Technometrics, 20 (1978) 397.
37. SYBYL 6.0 manual: Ligand-Based Design, Tripos Inc., St. Louis, MO, 1994, pp. 195–262.
38. Leo, A.J., CLOGP version 3.63, Daylight Chemical Information Systems, Irvine, CA, 1991.
39. Bostrom, J., Norrby, P.-O. and Liljefors, T., J. Comput.-Aided Mol. Design, 12 (1998) 383.
40. Gohda, K., Kimura, Y., Mori, I., Ohta, D. and Kikuchi, T., Biochim. Biophys. Acta, 1385 (1998) 107.
41. QUANTA 3.3 manual: Simulation, search, and analysis, Molecular Simulation Inc., Burlington, MA, 1994, pp. 185–228.
42. Clark, M. and Cramer, R.D., III, Quant. Struct.-Act. Relat., 12 (1993) 137.
43. Epps, D.E., Cheney, J., Schostarez, H., Sawyer, T.K., Prairie, M., Krueger, W.C. and Mandel, F., J. Med. Chem., 33 (1990) 2080.