Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 3093–3105
Bioorganic & Medicinal Chemistry

Connecting traditional QSAR and molecular simulations of papain hydrolysis—importance of charge transfer

Zsolt Lepp and Hiroshi Chuman*

Department of Pharmaceutical Sciences, The University of Tokushima Shomachi, Tokushima 770-8505, Japan

Received 1 February 2005; revised 28 February 2005; accepted 28 February 2005

Abstract—Molecular dynamics and full structure LocalSCF semi-empirical quantum mechanics calculations of receptor–ligand complexes were carried out to investigate structure–activity relationship for the papain hydrolysis of a series of *N*-benzoylglycine esters and reinterpret traditional QSAR descriptors using detailed structural information. A correlation of $r^2 = 0.694$ was obtained and it was shown that the pattern of charge distribution on the ester group is different if charges of free or complex ligands or are analyzed. The results can help to understand how traditional QSAR descriptors, such as F or σ, interacts with other electronic effects during complex formation.

1. Introduction

Predicting enzymatic reactions is a crucial problem of biochemistry thus, various methods to estimate biding properties for protein–ligand complexes have been emerged. These methods apply two main approaches; one is the QSAR type analysis,[1] which is based on the use of numerical descriptors of some properties of ligands and relates these properties to enzymic kinetic parameters, and the other approach directly calculates the binding properties from various molecular level simulations of the 3D structure of receptor–ligand complexes. Both approaches have their own advantages and drawbacks but both have been successfully applied to structure based drug design. However a 'final' method that could work out all the problems in the way of highly accurate prediction of enzymic activities has not been developed yet.

The purposes of this work was to search for links between traditional QSAR and molecular level simulations in order to help to understand how the traditional QSAR descriptors could be interpreted in a detailed 3D ligand–receptor complex.

To quantify the strength of protein–ligand interaction a score of properties calculated in different ways were applied. For rapid estimation of binding affinity various scoring functions were developed.[2,3] However highly accurate evaluation of protein–ligand interaction requires costly energy calculations on a large ensemble of conformations, such as free energy perturbation[4,5] or linear response approaches,[6] which use molecular dynamics or Monte Carlo calculations. Traditional QSAR uses three kinds of properties of unbound molecules, the electronic, steric, and hydrophobic effects of substituents of ligands. The latter two parameters are also employed by 3D structural simulations, however, the precise determination and usage of electronic properties to estimate binding affinities have not been implemented yet.

To find a bridge between the two different approaches it is necessary to take the protein–ligand electronic interactions into consideration. Our intention was to show the importance of inclusion of electronic parameters into structure–activity studies to clarify the significance of electronic factors in enzymatic activity. The usage of energy-type descriptors were avoided in order to correspond to traditional QSAR approach.

The work was based on previous analyses of the hydrolysis of *N*-benzoylglycine esters by a cysteine protease,

Keywords: Charge transfer; Field inductive effect; Hydrophobic effect; Linear regression; LocalSCF; Molecular dynamics; Papain; *N*-Benzoylglycine esters; Principal components analysis; QSAR; Semi-empirical quantum mechanics.
* Corresponding author. Tel./fax: +81 88 633 9508; e-mail: hchuman@ph.tokushima-u.ac.jp

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.02.057

papain.[7–10] It was selected because of the thorough and careful works that has already been carried out. It was a concern to select one reference with consistent measured enzymic kinetics to reduce the errors that can arise when combining different data sets.

As a result in the original study the following equations were derived:

$$\log 1/K_m = 0.57^*\sigma + 1.03^*\Pi_{3'} + 0.61^*MR_4 + 3.80 \quad \text{(Ref. 7)} \tag{1}$$

$$n = 25,\ r = 0.907,\ s = 0.202$$

$$\log 1/K_m = 8.13(\pm 1.8)^*F + 0.33(\pm 0.13)^*Z + 1.27(\pm 0.37)^*\Pi_{3'} + 1.95(\pm 0.29) \quad \text{(Ref. 8)} \tag{2}$$

$$n = 37,\ r = 0.939,\ s = 0.366$$

An aim of this study was to reinterpret the traditional QSAR descriptors by the lights of detailed molecular level structural information from the 3D protein–ligand structures. Such, physicochemically meaningful, parameters were searched that correspond to the classical descriptors.

The second equation has a worse correlation with the molecules that were selected than the first one (see later). However it would be difficult to find corresponding parameters in a 3D complex structure for the Hammet constant. Thus, the second equation was used as a base to search for links between the classical QSAR and molecular modeling approaches.

The field inductive effect (F),[11] the hydrophobic substituent constant (Π),[12] and the distance from oxygen through the end of the molecule (Z)[8] were used to quantify the electronic, hydrophobic, and steric effects from substituents, respectively.

To generate an ensemble of conformations molecular dynamics calculations were carried out. Corresponding parameters to the steric and hydrophobic QSAR descriptors were derived by measuring various geometric parameters in some snapshots of the molecular dynamics simulations. To accurately determine the electronic features of complexation, semi-empirical molecular calculations were on the entire structure of complexes carried out. Activities were estimated by the linear combination of the most significant ones of these descriptors.

Structural properties of complexes, the descriptors used for the regression equations of structure–activity analyses and the alteration of electronic structure of ligands due to complexation will be explained in the following.

2. Methods

2.1. Data set

Traditional QSAR parameters are all determined from the ligand alone, but in a protein–ligand complex there are many additional parameters, which might correspond to these descriptors and/or give additional information. Thus, there is no obvious answer for how to reinterpret the traditional QSAR descriptors. Because of this fact it was necessary to use as few variables as possible. Thus, it was decided to use a relatively homogeneous set of ligands. From the 37 ligands of the referred study 23 esters containing *para* or *meta* substituted phenol moiety were selected and are shown in Table 1. By removing the remaining molecules, which were aliphatic esters, the studied set of compounds differed only at the substituents of phenyl group. This simplification of the data set makes both the generation of consistent 3D structures and statistical modeling easier. Further four molecules (three nitrophenyl esters and the *p*-iodophenyl ester) from the original set were removed because of the erroneous treatment of these molecules by the LocalSCF software.

Although the diversity of the dataset is less than that of the original study, it was beneficial to use a more homologous set for the initial study. This makes it easier to find significant parameters and correlations and build a model. Furthermore models developed for a diverse set usually have worse predicting ability for a more homologues set. Thus, it seemed reasonable to develop a model for a less diverse set at first, than later expand it into a more general one.

2.2. Protein structure

The crystallographic structure of the cystein protease papain was obtained from the Protein Data Bank (PDB code 1KHQ).[13] This crystal structure contains a covalently bonded diazomethylketone inhibitor connected to the sulfur atom of residue cystein 25, which was removed before building complexes.

2.3. Describing electronic properties

It was found that for a variety of hydrolyses the coefficient with σ in QSAR equations is very much the same for enzymes acting on esters such as the ones of this study,[14] indicating similar electronic properties of these type of enzymes (Fig. 1). The original work used the field inductive parameter (F) to describe the electronic effects of the substituent. However, the advance of molecular orbital calculations made it possible to determine the atomic charges of the complexed ligands and describe the electronic features of those in a more direct way. This study makes a first attempt to use full structure molecular orbital calculations for structure–activity relationship analysis. To determine atomic charges single point semi-empirical calculations were carried out, which included all components of the complex; the receptor–ligand, water, and chlorine molecules. To characterize the differences between free and bound

Table 1. List of *N*-benzoylglycine ester compounds and the value of parameters used to derive regression Eqs. 2–6[8]

Compound	Obsd[a]	Pred[b]	Res[c]	F[d]	Z[d]	$\Pi_{3'}$[d]	C2[e]	O1[f]	WS_O1[g]	$D2$[h]
1 H	3.86	4.092	−0.232	0.12	1.48	0	−0.113	−0.205	2.200	−6.022
2 4-NH_2	3.59	3.925	−0.335	0.13	2.21	0	−0.200	−0.183	1.320	−73.541
3 4-F	3.78	3.837	−0.057	0.17	1.72	0	−0.158	−0.196	1.987	−3.401
4 4-CH_3	3.99	3.910	0.080	0.14	2.29	0	−0.130	−0.197	1.615	41.016
5 4-OCH_3	4.05	3.828	0.222	0.13	3.63	0	−0.152	−0.176	1.685	−59.376
6 4-Cl	4.20	4.309	−0.109	0.18	2.11	0	−0.107	−0.180	0.098	0.260
7 4-CN	4.23	4.377	−0.147	0.2	2.93	0	−0.081	−0.205	2.039	−37.082
8 4-SO_2NH_2	4.28	4.245	0.035	0.2	2.91	0	−0.003	−0.210	1.399	149.640
9 4-$CONH_2$	4.60	4.451	0.149	0.16	3.52	0	−0.065	−0.189	0.716	−3.717
10 4-$COCH_3$	4.95	4.985	−0.035	0.17	3.7	0	−0.065	−0.204	1.389	−136.766
11 3-NH_2	3.62	3.955	−0.335	0.13	1.47	0	−0.091	−0.181	1.785	−18.745
12 3-$NHCOCH_3$	3.83	4.221	−0.391	0.17	3.14	0	−0.107	−0.217	2.322	11.064
13 3-F	4.03	4.116	−0.086	0.17	1.47	0.14	−0.083	−0.165	0.727	−38.775
14 3-$CONH_2$	4.12	3.972	0.148	0.16	2.23	0	−0.128	−0.192	1.985	−27.156
15 3-CN	4.32	3.896	0.424	0.2	1.49	0	−0.129	−0.168	2.078	−122.723
16 3-SO_2NH_2	4.55	4.273	0.277	0.2	2.67	0	−0.132	−0.207	1.657	−18.997
17 3-CH_3	4.58	4.280	0.300	0.14	1.45	0.56	−0.113	−0.212	0.943	76.546
18 3-Cl	4.73	4.811	−0.081	0.18	1.47	0.71	−0.111	−0.238	1.688	−6.013
19 3-CF_3	4.28	4.498	−0.218	0.18	1.66	0.88	−0.105	−0.204	0.765	6.187
20 3,5-$(OCH_3)_2$	4.49	4.587	−0.097	0.21	1.43	0	0.144	−0.172	2.382	−48.733
21 3,5-$(CH_3)_2$	4.69	4.596	0.094	0.12	1.46	0.56	−0.058	−0.195	0.784	−13.618
22 3,5-$(Cl)_2$	4.91	4.913	−0.003	0.25	1.45	0.71	−0.040	−0.224	1.913	−44.279
23 3-CH_3, 5-C_2H_5	4.97	4.573	0.397	0.12	1.45	1.02	−0.056	−0.222	2.229	−2.254

[a] Observed $\log 1/K_m$.
[b] Predicted $\log 1/K_m$.
[c] Residual.
[d] From Ref. 7.
[e] Charge on atom C2.
[f] Charge on atom O1.
[g] Average watershell around atom O1.
[h] Dihedral angle $D2$ in deg.

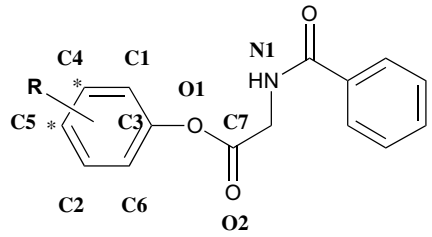

Figure 1. The general structure of *N*-benzoylglycine esters and the naming of atoms that were the subjects of the analyses. Atoms where substitution took place are marked with asterisk. In case of compounds 20–23 the C2 atom was also substituted.

molecules the ligands of complexes were extracted and similar calculations were carried out on those in vacuum. It was also a concern to understand how the field inductive parameter can influence the electronic properties of complexed ligands.

It was necessary to decide, which atoms were to be included in the structure–activity analysis. The ligands vary only in the substituent of phenolate group and are consisted of two conjugated aromatic systems; the benzamide and the phenolate groups. These two subparts are connected together by a methylene group at the middle of the molecule blocking the further conjugation of the two phenyl groups. Thus, these two aromatic systems can be considered independent regarding the electronic properties. Because the amide part is the same for all compounds, only the ester group of each ligands and for further simplification only its heavy atoms were studied. It means nine atoms altogether as shown in Figure 1. These are the six carbons of phenyl named C1–C6, and the carbon (C7) and two oxygen atoms (O1, O2) of the carboxyl group. As an additional descriptor the total charge transfer to the whole molecule were also used.

2.4. Describing hydrophobic and conformational properties

Another fundamental parameter in classical QSAR is the hydrophobic substitute constant. Molecular level simulations provide very strong tools to estimate the effect of hydration/dehydration during complex formation. To determine this effect, the original study used the hydrophobic substituent constant $\Pi_{3'}$, which was determined by the following way. It was used only for ligands with hydrophobic substituents and was zero for the other compounds. Furthermore it was hypothesized that only one of the two *meta* substituents can contact the hydrophobic surface, the other is perforce placed into the aqueous phase. Thus, to calculate $\Pi_{3'}$ values for these molecules only the effect of one substituent was considered. In this study a water cap of 20 Å radii around the Cys25 residue of the active site was added and the watershell values of each of the mentioned nine atoms of ligands, that is, the number of water molecules around those, were used as descriptors. The usage of

watershell is a direct and consistent way to describe dehydration process.

To take conformational effects into consideration various measures were also determined from the complex structures. These included key torsional angles along the carboxylic group and neighboring atoms. Atomic distances inside the ligands and some ligand–receptor residue distances were also measured.

2.5. Generating initial structures by docking

Each of the protein–ligand complexes used for molecular dynamics simulations were generated by docking the ligand into the active site by the FlexX module of Sybyl 6.91.[15] The active site to be extracted was defined by a radius of 6.5 Å around the inhibitor of the crystallographic structure. The covalently bonded inhibitor and the crystallographic water molecules were removed. The *N*-benzoylglycine ligands were built by PC Spartan'04.[16] The initial conformers of ligands were the predicted global minimum structures made by the built-in conformation search algorithm of PC Spartan'04 by using MMFF force field[17] and default setup values. For docking were all default parameters as implemented in FlexX of Sybyl 6.91 used. Cscore[18] calculations were used for ranking and the conformer of best score was used for the following simulations. Hydrogen atoms and those of the original crystallographic water molecules, which had a crystallographic B value of more than 15 were added to the protein of each complex. The coordinates of hydrogens were optimized by molecular mechanics minimization using the Tripos force field[19] and Sybyl 6.91. Non-hydrogen atoms were fixed and default force filed parameters and 2000 steps of minimization were used. Arginines and lysines were positive and aspartates and glutamates were negative whilst other amino acids were neutral.

2.6. Molecular dynamics simulations

Molecular dynamics calculation was used to obtain optimal conformation for complexes and statistical ensemble for structure–activity relationship analyses.

Amber 7[20] was used for the molecular dynamics simulations. The same procedures were applied for all the complexes by using Unix scripts. The Amber *ff99*[21] and GAFF[22] force fields were used for describing the protein and ligands, respectively. Input files were made by the tleap,[20] MM and MD calculations were carried out by the Sander,[20] and various distance related information from the trajectories were obtained by the ptraj[20] module of Amber 7. The atomic charges for ligands were calculated by carrying out single point calculation on the HF/6-31G* level using Gaussian 98[23] and were fitted using the standard RESP method[24] by the antechamber[20] module of Amber 7. The protein was reported to be positively charged and neutralized by tleap by adding five chlorine anions. A TIP3 water cap of 20 Å radius around the Cys25 residue was added.

First, the water molecules of the complex were minimized whilst keeping the other atoms constrained using a harmonic potential with a force constant of 200 kcal/mol/Å^2. The non-bonded electrostatic cutoff was 15 Å and all the other parameters were the default values of Amber 7. The maximum number of minimization steps was 3000 and the convergence criterion for the energy was 1e-4. One final minimization was done with all the restrains removed. All the parameters were the same as for the restrained minimization but with the maximum minimization cycles being 9000. The final minimized structure was used for molecular dynamics simulations.

The general setup for all phases of molecular dynamics calculations was the following. The method of temperature scaling applied was constant temperature using the weak-coupling algorithm with default values as implemented in Amber 7. All the force field parameters were the same as for the minimization processes but cutoff value was reduced to 10 Å. All bonds involving hydrogens were constrained by SHAKE algorithm as implemented in Amber 7.

The simulations were carried out as follows. First the water molecules were equilibrated in three steps. (1) The temperature was raised from 0 to 100 K during 20 ps; (2) The temperature was raised from 100 K to 300 K during 20 ps, and (3) Final 20 ps constrained simulation at 300 K was performed. The timestep was 0.5 fs and the constraining force constant was 10 kcal/mol/Å^2 for each case. The constrains were removed and the full complex was equilibrated at 300 K for 30 ps with 0.5 fs then for 100 ps with 2 fs timestep. The production run took 1500 ps and for all kinds of analysis the trajectory of only the last 1000 ps were used. Along the 1 ns trajectory 2000 snapshots were collected from which various descriptors were calculated. These descriptors were (1) structural parameters of ligands (Fig. 2) such as torsional angles (D1–D3) and Van der Waals distance between the phenolic oxygen and the end of the molecule in the direction of the axis connecting the oxygen with the adjacent atom (*Z*). (2) Watershells, which were defined by counting the number of water molecules around

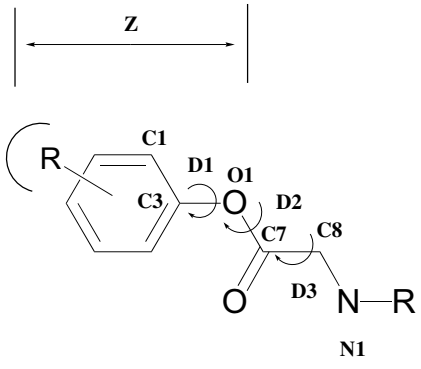

Figure 2. Generation of geometrical descriptors for regression analyses. D1–D3: dihedral (torsion) angles between atoms C1–C3–O1–C7, C3–O1–C7–C8, and O1–C7–C8–N1, respectively. *Z*: the Van der Waals distance from O1 to the end of the molecule through the axis of O1–C3.

the given atom inside a sphere of 5 Å radius, around various heavy atoms were determined. (3) Distances between the center of mass of carbon atoms of the substituted phenyl ring and the center of mass of protein residues or combination of residues forming the wall of active site. These residues (or sets) were as follows: Gln19-Gly20-Ser21; Asn64; Gly65-Gly66; Tyr67-Pro68; Ala136-Ala137; Gly138-Lys139; Gln142; Asp158-His159; Ala160; Ser176-Trp177. Finally the distance between the sulfur atom of Cys25 and the carboxylic oxygen of both the ester and amide groups (O2 and O3, respectively), were calculated.

2.7. LocalSCF calculations

Semi-empirical quantum mechanics calculations were performed on the full complexes by the LocalSCF method[25] and LocalSCF2003 program.[26] Ten snapshots during the 1 ns of MD production run were collected (one per every 100 ps) and single point calculations were carried out in vacuum with all the water molecules included. For all cases AM1[27] was used as the Hamiltonian with the initial local molecular orbital (LMO) involving the Lewis core plus four chemical bonds around it. The threshold of derivatives for the expansion of LMO's was set to 0.04 and the maximum allowed non-orthogonality of orbitals were set to 0.001. The ligands were exctracted from the complex and single point calculations with identical parameters to those of the complexes were performed in vacuum. The atomic charges from the output file of all the 10 snapshots of ligands and complexes were extracted and were used for analysis.

2.8. Statistical analyses

Multiple linear regression analyses were performed using the statistical software package Xlstat pro7.5.[28] Measured enzyme activities were obtained from the literature.[8] First the descriptors were standardized with the mean and standard deviation being 0 and 1, respectively. Descriptor sets were chosen so that the correlation coefficient, r^2 was maximized whilst the number of descriptors was fixed. The number of descriptors to be used varied between 2 and 5 and the leave-one-out procedure was applied to determine the maximum number of descriptor in order to avoid overfitting. The statistical significance of the descriptors was confirmed from the variance analysis using the Fisher's F ratios, requiring that the probability of a greater F value occurring by chance (Prob > F) is less than 0.01.

Principal coordinates (PCO)[29] and components (PCA)[30] analyses were performed using the software MVSP 3.12.[31] Axes were extracted by the Kaiser's rule that states that the minimum eigenvalue should be the average of all eigenvalues (or 1 if the correlation matrix is used), in all of the cases. For PCA the data were centered and standardized. In case of PCO analyses the measure of distance between the descriptors of compounds were measured by the Euclidean rule. The accuracies were 1e-7 for all the cases of PCA and PCO analyses.

3. Results and discussion

3.1. Binding modes

The structural features of enzyme Papain have already been extensively studied.[13,32] The active site of the enzyme has a hydrophobic subsite, which includes Tyr61, Tyr67, Pro68, Trp69, Phe207, and Ala160. It was assumed[33] that this subsite can firmly bind the unsubstituted phenyl ring of the *N*-benzoylglycine esters. Both the docking and molecular dynamics simulations have confirmed this statement. In Figure 3 some characteristic

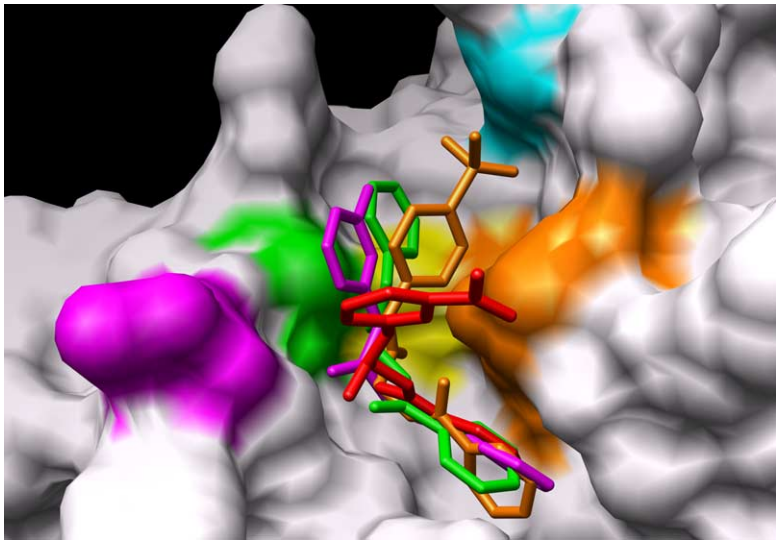

Figure 3. Characteristic structures of molecular dynamics simulations. Compounds **1**, **8**, **11**, and **14** (in green, orange, magenta, and red color, respectively; water molecules were removed). The graph was generated by aligning the protein structures of complexes. Only the receptor of compound **8** is shown, the conformations of those of the others may be slightly but not significantly different. The closest residues to the ligands are tinted in corresponding colors: Gly23-Ser24 (green), Glu158-His159 (orange), Asn64 (magenta). Two other residues are highlighted: Gln142 (cyan), and Cys25 (yellow).

structures at the end of 1 ns MD production runs of various complexes are shown. It is clearly visible that the unsubstituted phenyl rings of the four ligands (**1**, **8**, **11**, **14**, colored green, orange, magenta, and red, respectively), are firmly anchored in this pocket and well overlapping, whilst the other part of the molecules are diverging. Figure 3 depicts the difference between the spatial orientations of *para* and (di)-*meta* substituted ligands, too. Whilst the former ligands are closing toward residues Glu158 and His159 (colored orange), the latter ones are rather moving toward Asn64 (magenta).

In Ref. 33 it was also stated that a major difference between the binding modes of ligands is that *para* substituted ligands are capable to contact the amide moiety of Gln142. The validity of this assumption could be also confirmed and is shown in Figure 3 that the substituent of **8** is in the proximity of Gln142 (cyan) and in Figure 4, which shows a space filled model of **8** and **17**. In Figure 3 two other regions are also emphasized by color; (1) Cys25 (yellow), which is a part of 'oxyanion hole' formed by Gln19, Ser24, and Cys25, and it is the residue to which the inhibitor of the used crystallographic structure, 1KHQ was covalently bonded. (2) Gly23-Ser24 (green), which are partly forming the bottom of the active site. The tendencies in the orientation of substituted phenyl rings were numerically proved from the MD trajectories. Distances between the center of phenyl ring of the ester group and various residues of the protein were calculated, as described in the Methods section. The averages of these values (using 2000 snapshots for each complex) were calculated. It was found that the closest contacts between the protein and ligands for the case of practically all ligands are the ones belonging to the aforementioned residues, such as Gly23-Ser24, Glu158-His159, and Asn64.

Table 2. Average distance (Å) between the center of the phenyl of ester group and the center of mass of some significant residues of papain

	158–159[a]	64[a]	23–24[a]	142[a]
para	6.13	7.58	7.12	10.28
meta, di-*meta*	7.63	6.66	8.38	12.10

[a] Residue number; more than one number means that the center of mass of multiple residues were calculated.

The geometrical means of these average distances for each compounds are summarized in Table 2. It can be seen that on average the *para* substituted ligands are closer to residues Glu158-His159 and are farther from Asn64 than the di-*meta* substituted ones. Also, the Gln142 residue is closer to the *para* compounds, showing the interaction of *para* substituent with this residue. To residues Gly23-Ser24 the shortest distance (less than 6 Å) can be found for the non-substituted **1** whilst for the other compounds the interaction with residues 64 or 158–159 hinders the phenyl group to move further to the bottom of the active site. The difference between the orientations of the two series of ligands has a consequence on the electronic structure of the complexed molecules as it will be discussed later.

3.2. Effect of solvation/desolvation

Watershell around skeletal atoms of the ester groups were calculated as discussed in the methods chapter. In Table 3 the average watershell values of 2000 snapshots around the atoms of interest are summarized. C4

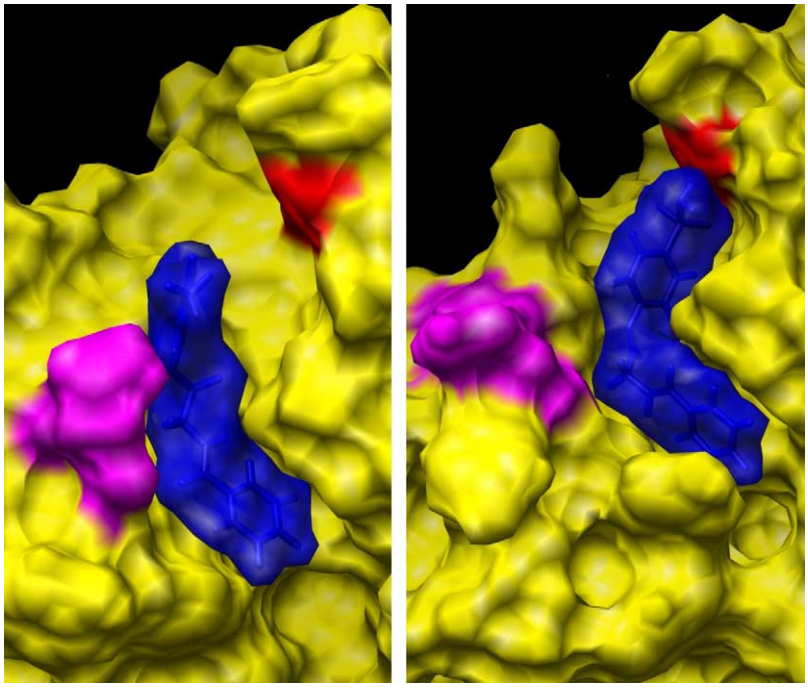

Figure 4. Difference between characteristic MD structures of a *meta* (**17**, left) and a *para* (**8**, right) substituted ligand. Water molecules were removed. Red and magenta colors shows residue 142 and 64, respectively.

Table 3. Average number of water molecules in the 5 Å water sphere around the positions of substitution

Compound	C2	C4	C5
1	2.54	1.06	2.29
2	3.05	3.31	2.68
3	2.16	2.09	1.75
4	1.80	1.81	1.13
5	2.82	3.04	2.72
6	2.68	0.77	1.31
7	2.39	2.29	1.73
8	2.49	0.41	0.91
9	1.24	2.68	1.59
10	0.78	1.96	1.08
11	1.87	2.26	2.50
12	0.00	0.31	0.03
13	2.42	2.05	2.57
14	2.52	1.74	2.22
15	2.67	2.03	2.79
16	0.62	0.05	0.04
17	2.42	1.70	2.52
18	2.67	1.84	2.61
19	2.57	0.86	1.87
20	2.41	2.39	2.72
21	0.65	0.40	0.93
22	0.88	1.12	1.73
23	2.01	1.91	2.39

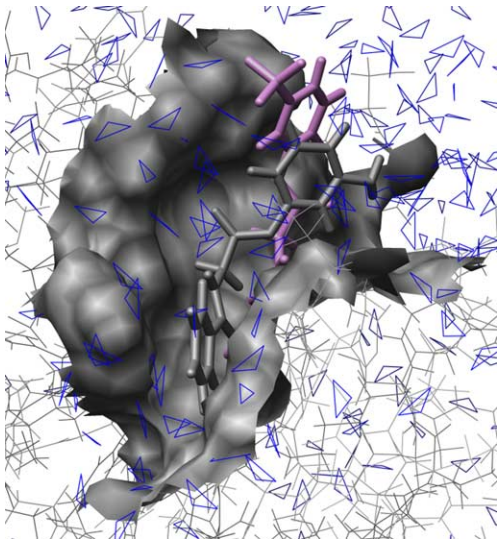

Figure 5. Average molecular dynamics structures of compounds **11** (gray) and **19** (pink). The phenyls of ester groups of both compounds are in close contact to residue Asn64. The substituent of **11**, contrary to other ligands with *meta* substituent such as **19**, is facing toward the water phase. Water molecules are depicted as triangles. The receptor of **19** is not shown.

and C5 are the atoms on which *meta* and *para* substitution, respectively, were done and C2 is substituted only in the case of di-*meta* compounds. First of all, we should see that there is a constraint on the rotation of phenyl groups inside the complex in the case of *para* compounds. C2 and C4 are equivalent atoms in the *para* ligand, thus the watershell values around these atoms would be equal. The reason for this hindered rotation is possibly the mentioned interaction of these ligands with the Gln142 residue, which limits the freedom of movement for this ligands.

However, the watershell values of these atoms are not equal, which shows the lack of rotational freedom. But, whilst either the value for C4 or C2 is the larger in a random fashion for these compounds, in the case of all but two (**11** and **12**) of the *meta* compounds, the side of unsubstituted C2 were facing the water face. Figure 5 shows the orientation of a typical (**19**) and of the irregular (**11**) *meta* compounds. The possible reason for the irregular behavior of **11** is that it is the only compound with basic character (aniline). The amine is forced into the water phase by repulsive forces between it and the probably basic side chain of protein around the Asn64 residue. The other compounds have either acidic (**16**) or neutral substituents. From the molecular dynamics trajectory it was clear that the orientation of the phenyl group of 11 relative to the protein is similar to those of the other *meta* compounds. The difference is that the substitute of this ligand faces the water phase rather than the protein phase contrary to the other molecules.

It should be noted that the watershell values around compounds **12** and **16** is relatively low. The reason for this phenomenon lies in the large substituent groups of these molecules. These shield the phenyl group from the water phase and push it downward the protein. Because of the small value, the watershell around C2 and C4 is not a very good indicator for these two compounds. However, considering the other *meta* ligands it is very possible that the difference between the watershell around the two halves of the molecules is not caused by the effect of random placement as in the case of *para* compounds. It is most likely caused by the attractive or repulsive electrostatic forces between the substituent and the protein side chain.

By examining the binding modes and the watershell values some assumptions of Ref. 8 can be validated. It was presumed that the di-substituted ligands are placed somewhat half between the protein and water phases. The previously discussed difference in watershell around C2 and C4 verifies this hypothesis. On the other hand, no evidence was found that the compounds with hydrophobic ligands bind differently than the others. Because this assumption gave the bases of the assignment $\Pi_{3'}$ values[8] it might be concluded that this treatment cannot be confirmed.

The different relative orientation of *para* and *meta* ligands, and the later discussed difference in charge transfer between the two groups of compounds show the difficulties of including all the compounds in one QSAR model.

3.3. Regression analyses of $\log 1/K_m$ with the use of electronic and structural descriptors

After generating the descriptors assumed to be significant and in correspondence with traditional QSAR descriptors, regression equations to describe the biological

activities of ligands using these parameters were developed. Various electronic and steric parameters were calculated from snapshots of the 1 ns molecular dynamics trajectory of each complex. (1) The descriptors for the electronic features of the complexed ligands were the atomic charges of nine atoms as mentioned in the Methods section; the heavy atoms of the substituted phenyl ring and the carboxylic group (Fig. 1) and the total charge transfer to the ligands. To calculate these charges 10 snapshots were extracted in a regular fashion, one structure were selected at every 100 ps and the charges for the entire complex, including all the water molecules and chlorine atoms were calculated by single point semi-empirical quantum mechanics calculation. The final values of atomic charges to be used for the analysis were the average (mean) of those from the 10 snapshot structures. The ligands were removed from the complexes and the same single point calculation was carried out in vacuum. Charge transfer was calculated by subtracting the atomic charges of the free ligands from those bonded inside the corresponding complexes. (2) To characterize the solvation effects watershell values were determined for each skeletal heavy atoms (see Methods). (3) In order to take the conformational effects into consideration dihedral angles and Z values were measured from the same 2000 snapshots as seen in Figure 2. Table 4 summarizes the descriptors used.

Among these descriptors a set was selected by finding the best linear regression model. All but two of the protein–ligand complexes were used. The two main criteria for the selection process were to get a linear combination of the parameters that gives the best correlation with the experimental Michaelis-Menten constants ($\log 1/K_m$) but avoiding any possible overfitting by using the least number of descriptors. It was found that the maximum number of descriptors that can be used was four. Using more parameters results in overfitting as it was proved by the leave-one-out procedure. The best correlation obtained was ($r = 0.833$, $r^2 = 0.694$ Eq. 6) better than the correlation that can be achieved using the original descriptors (F, Z, $\Pi_{3'}$) ($r = 0.734$, $r^2 = 0.524$, Eq. 3). It should be noted that the study referred is using 37, more diverse ligands and obtaining a reasonably good r^2 of 0.87. However, if we calculate the r^2 using the original parameters and only those compounds, which were the subjects of the current study, we get the mentioned correlation coefficient (Eq. 3). It means that the descriptors used in this study are more sensitive to subtle changes in the molecular structure of less diverse compounds than the traditional parameters. The values of the parameters of the best model for each molecule are listed in Table 1.

Table 4. Principal descriptors considered in the linear regression modeling

Symbol	Description
C1–C7	Atomic charge on C atoms no 1–7
O1–O3	Atomic charge on O atoms no 1–3
ChTrans	Total charge transferred to the ligand
*D*1–*D*3	Dihedral angles 1–3, see Figure 2
WS_Xn	Watershell around X atom no *n*
VdW_vol	Van der Waals volume
Z	Distance as depicted in Figure 2

Table 5. Correlation coefficients (R) between the regression model parameters

Correlations	O1	C2	WS_O1	D2
O1	1.00			
C2	0.01	1.00		
WS_O1[a]	0.15	−0.22	1.00	
D2[b]	0.11	−0.38	−0.19	1.00

[a] Average of watershell values around atom O1.
[b] Dihedral angle D2.

The correlation matrix between the $\log(1/K_m)$ values and the descriptors can be seen in Table 5. It is clear from the values of the matrix that neither any of the parameters can determine the activity alone, nor there is significant correlation between these parameters.

$$\log 1/K_m = 3.976(\pm 1.881)^* F + 0.168(\pm 0.093)^* Z + 0.884(\pm 0.218)^* \Pi_{3'} + 3.104 \quad (3)$$

$$n = 23,\ r = 0.734,\ F = 7.385,\ s = 0.303$$

$$\log 1/K_m = 0.188(\pm 0.071)^* \mathrm{C2} - 0.190(\pm 0.071)^* \mathrm{O1} + 4.289(\pm 0.069) \quad (4)$$

$$n = 23,\ r = 0.642,\ F = 7.000,\ s = 0.333$$

$$\log 1/K_m = 0.206(\pm 0.066)^* \mathrm{C2} - 0.249(\pm 0.071)^* \mathrm{O1} + 0.153(\pm 0.071)^* \mathrm{D2} + 4.289(\pm 0.064) \quad (5)$$

$$n = 23,\ r = 0.774,\ F = 8.474,\ s = 0.307$$

$$\log 1/K_m = 0.242(\pm 0.056)^* \mathrm{C2} - 0.316(\pm 0.062)^* \mathrm{O1} - 0.186(\pm 0.059)^* \mathrm{WS\_O1} - 0.218(\pm 0.062)^* \mathrm{D2} + 4.289(\pm 0.053) \quad (6)$$

$$n = 23,\ r = 0.833,\ F = 10.228,\ s = 0.253$$

The descriptors and the sign of the coefficients are physicochemically reasonable. Regarding charges, O1 is at the reaction center and C2 is the closest non-substituted atom to both *para* and *meta* substituents, that is, its state is identical in all but di-*meta* substituted molecules and the effect of substitution on the neighboring atoms is the strongest on it. Furthermore, C2 is practically independent from O1. These two parameters are the most important and can be found in all of the equations. The sign of the correlation coefficient (Eqs. 4–6) is

negative for O1 whilst that of the C2 is positive. It can be explained as charge polarization promotes the reaction, which corresponds with the chemical sense. The negative sign of the watershell descriptors is also reasonable; a larger dehydration (thus, smaller watershell value) around the reaction center is a sign of a stronger ligand–receptor interaction. The D2 dihedral angle determines the relative orientation of substituted phenyl group between the protein and water phase.

All of these descriptors are based on various properties of the ligands. The usage of any specific parameters related to the enzyme (e.g. distances between the ligand and residues of protein) could be avoided. The advantage of using only ligand based descriptors is the possibility that the method might be used for other targets too.

3.4. Effect of charge redistribution

From the regression equations it is clear that the atomic charge distribution plays a decisive role in the structure–activity relationship. Thus, it is important to examine the effect of complexation on the electronic structure of the molecules. Two methods were used to describe the behavior of atomic charges. (1) The tendency of internal charge distribution in free and bound molecules comparing all ligands to each other were studied by principal coordinates analyses. (2) The correlation of the field inductive effect (F) and atomic charges of ligands were also examined and the effect of F on the total charge transfer has been clarified.

The values of average total charge transfer to a ligand were calculated by summing the atomic charges of all of the atoms of the complexed ligand. Although, for a non-complexed neutral molecule the sum of charges of its atoms is, obviously zero, it is not zero for the complexed ligand, because of the charge transfer between the ligand and the protein. The results, summarized in Table 6, show that the hydrolysis of *N*-benzoylglycine esters by papain is mediated by a positive charge transfer to the ligand.

The charges on the atoms of a ligand after complex formation is determined by two forces; the electrostatic effects between the ligand and the protein and the relocalization of transferred charge on each atom by internal effects, mean the resonance and field/inductive effects. Although these forces cannot be independently studied it is of great interest to understand how the charges of each atoms of a ligand altered relative to the same atoms of the other ligands during complex formation. The atomic charges on the nine examined atoms (C1–C7 and O1–O2) were subjected to principal coordinates (PCO) and principal components (PCA) analysis.

These analyses describe the correlation between the aforementioned nine atoms regarding the 23 compounds. It can be imagined as there were 23 compounds with nine descriptors (these are the atomic charges). The PCO and PCA analyses are used to make groups of these compounds based on the values of descriptors. As it is shown in Figure 6.1.a, no tendency, no grouping can be observed in the scatter graph of the two principal coordinates axes in the case of the free ligands. However, a strikingly clear difference can be observed in the similar graph for the complexed ligands: the ligands with *para* and *meta* or di-*meta* substituents are heavily separated and located to the opposite sides of the first principal coordinate axis (Fig. 6.1.b). This difference shows that the process of charge transfer/redistribution take place to these atoms in a different manner in molecules substituted at *meta* or *para* position. A similar PCO analysis of the values of transferred charges (obtained by subtracting the corresponding values of free and bonded ligands) reveals that these values show the same tendency (Fig. 6.1.c) thus, the difference in charge transfer to the atoms is behind the difference of correlations between atomic charges considering bonded ligands versus free ones. If this phenomenon was caused by the internal electronic properties of ligand molecules than similar separation of the two groups of ligands could be observed on the PCO plot of free ligands (Fig. 6.1.a). The lack of such kind of tendency shows that the complexation alters the electronic structures of the two groups of molecules in a different manner.

Because PCO analysis does not give information about the role of each descriptor, PCA analysis was also carried out. As it can be seen in the scatter graph (Fig. 6.2.a), the *para* versus di-*meta* substituted molecules are separated along the second PCA axis.

Table 7 lists the loadings of each descriptor in the first 3 PCA axes. Regarding the second axis, parameters with the highest loadings are the charges on the O2 and C7

Table 6. The amount of total charge transferred to each ligands obtained by averaging 10 snapshots per compound

Compound	Total transferred charge
1	0.0284
2	0.0290
3	0.0414
4	0.0321
5	0.0386
6	0.0251
7	0.0193
8	0.0218
9	0.0411
10	0.0306
11	0.0179
12	0.0253
13	0.0153
14	0.0432
15	0.0414
16	0.0611
17	0.0294
18	0.0330
19	0.0361
20	0.0361
21	0.0405
22	0.0266
23	0.0179

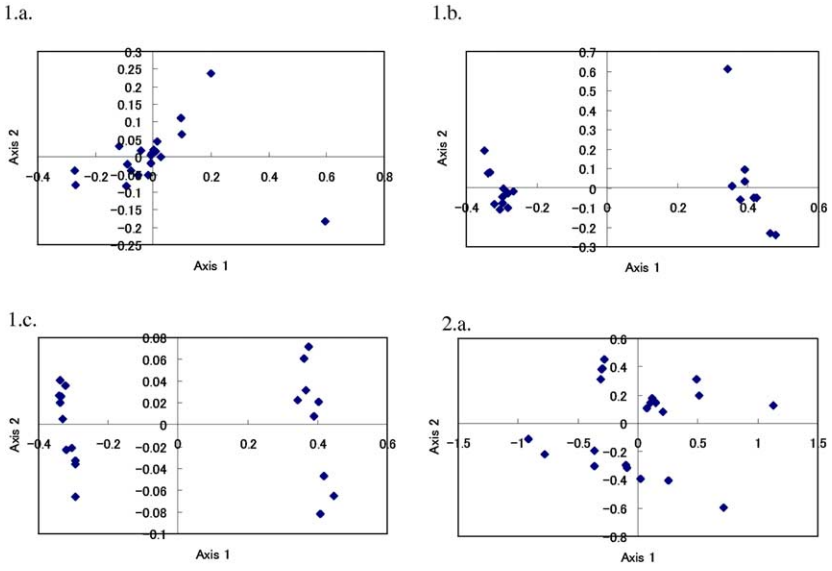

Figure 6. Scatter plots of the first two principal axises of principal coordinates (1.a.–c.) and principal components (2.a.) analyses of the charges on the skeletal atoms (C1–C7 and O1–O2) of ester group. 1.a. free ligands; 1.b. complexed ligands; 1.c. charge difference between complexed and free ligands; 2.a. complexed ligands. In 1.b. and 1.c. the points of *para* and di-*meta* substituted ligands are separated by axis 1 (located at the right and left side, respectively), and in 2.a. by axis 2 (below: *para*; above: *meta* and di-*meta* substituted compounds). Free ligands do not show the same tendency (1.a.).

Table 7. PCA variable loadings of the first three principal axis for the atomic charges on complexed ligands

	Axis 1	Axis 2	Axis 3
Cum. %[a]	29.230	46.768	62.912
C1	−0.402	0.097	−0.137
C2	0.41	−0.064	0.091
C3	0.395	−0.267	−0.194
C4	0.422	0.154	0.232
C5	−0.328	0.214	0.276
C6	−0.385	0.156	−0.225
O1	−0.051	0.158	0.832
O2	−0.226	−0.603	0.162
C7	0.163	0.658	−0.192

[a] Cumulative contribution of the PCA variables to the variance.

atoms. These highest loadings show the parameters with the largest variations. It means that the largest difference between differently substituted molecules was at the atoms of reaction center. Thus the complex formation has a significant effect on these atoms.

Correlation studies between charge transfer and field inductive effect (F) of substituents has also shown discrepancy between *para* and *meta* substituted ligands. The correlation between the total charge transferred to ligands and the F value shows the opposite tendency in case of *para* and *meta* ligands, and no correlation in the case of di-*meta* substituted ones (Fig. 7). Though the correlation coefficients are not large ($r = -0.720$, +0.597 and +0.026 for *para*, *meta*, and di-*meta* compounds, respectively), it might be stated that F is in inverse proportion to the total amount of charge transferred to the ligand in the case of *para* compounds whilst it is directly proportional to that in case of *meta* substituted ligands. The reason of this phenomenon is that the substituents interact with different residues of the protein as it was shown in Figure 4. This effect of substitution on the total charge transferred to the compounds also validates the use of receptor–ligand complexes instead of single molecules for the structure–activity studies of this example.

It is also interesting to examine how the F value correlates with atomic charges in free and complexed ligands. It was found out that the *para* and *meta* substituted compounds had behaved differently, again. Of course, due to the definition of F there should be a difference between the two types of compounds regarding correlation with atoms, but it seems that complexation has also a different effect. In Table 8 the correlation coefficients for the linear regressions are shown. The same atoms in free and bound *para* ligands correlates with F (note that C1 and C6 are chemically identical atoms) and the correlation coefficient has increased a bit after complex formation. In the case of *meta* substituted ones not only C1 was replaced by C5 (a chemically different atom) in the regression equation but also the correlation was a bit worse.

It is also worth noting that the atoms of which atomic charges correlate best with the Michaelis-Menten constants ($\log 1/K_m$) and with the F values are not identical.

As a side note: it was found that the charge transfer to the methylene group separating the two aromatic systems were on average about an order of magnitude smaller than those to the heavy atoms of ester group meaning that this group electronically separates the two rings as were hypothesized.

Finally, it can be concluded that the papain enzyme–substrate interaction amplifies the differences in

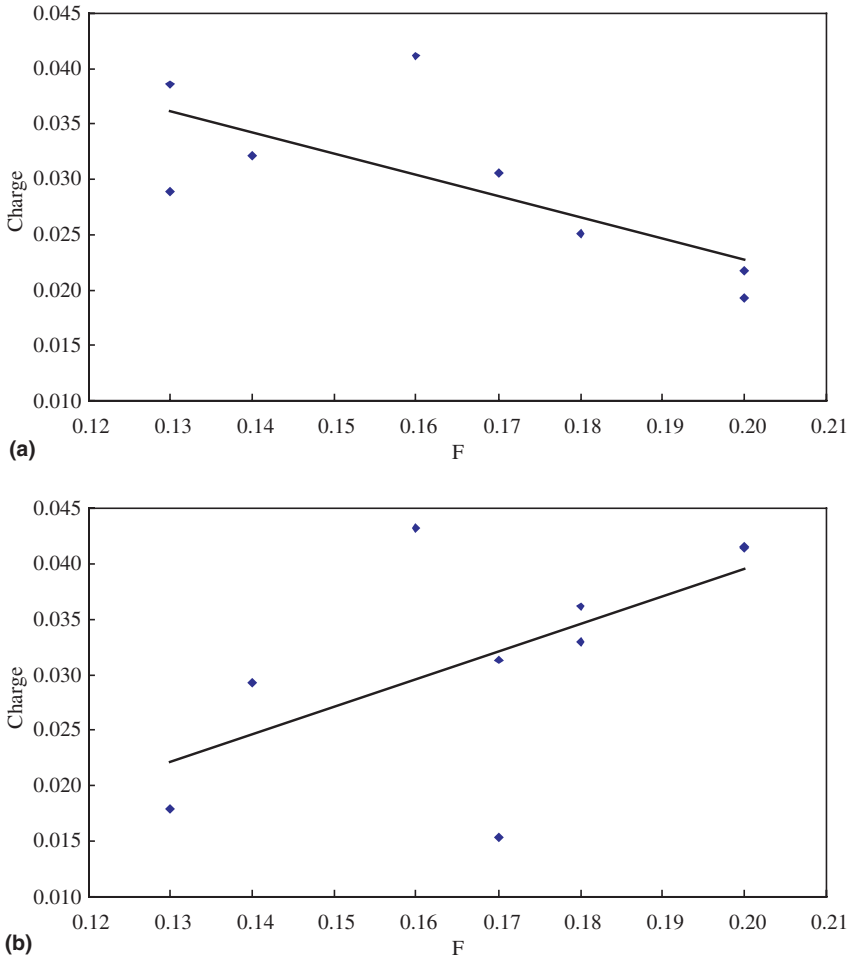

Figure 7. Correlation between the total transferred charge to the ligands and the *F* values of ligands: (a) *para*, (b) *meta* substituted compounds. The correlation coefficients for the regression equations are $r = -0.720$, $s = 0.021$; $r = 0.597$, $s = 0.008$, respectively.

Table 8. The variables and correlation coefficients for the regression equations of field inductive parameter (*F*) versus atomic charges of ester group of complexed ligands

Compounds	Non-complexed			Complexed		
	para	*meta*	All	*para*	*meta*	All
Variable (s)	C1, C4	C1, C3	C1, C3	C4, C6	C3, C5	C2, C6
r (coefficient of correlation)	0.907	0.791	0.603	0.938	0.759	0.612
r^2 (coefficient of determination)	0.822	0.625	0.364	0.880	0.576	0.375
SSR	0.001	0.006	0.016	0.001	0.007	0.015

electronic structure between molecules with differently substituted phenyl groups. Complex formation has some effects on the electronic structure of ligands, which makes tendencies of charge distribution and the regression model different from what could be obtained by considering only the electronic properties of free ligands.

The internal electronic properties of a molecule have the most important influence on enzyme reactions. But the interactions between the ligand and receptor could also alter the electronic properties of the ligands. As it was mentioned in the chapters discussing binding modes and hydration, the relative orientation of ligands with *para* and *meta* substituents relative to the side chains of the protein is different. Because of the different enzymic environment the charge transfer/redistribution was different as it was shown by the PCO analyses.

It might be true that the significance (mean, absolute value of regression coefficient) of *F* (and σ) values in a traditional QSAR equation is also influenced by this effect; complexation can amplify or weaken the contribution of *F* to the regression model. However, it is necessary to examine various other examples in order to prove the general validity of this conclusion.

4. Conclusions

Regression analyses were carried out to predict Michaelis-Menten constants for the papain hydrolysis of 23 homologous *N*-benzoylglycine esters by means of molecular dynamics and semi-empirical quantum mechanical calculations. Descriptors for the analysis were various atomic charge and structural parameters of complexes. This study, by our knowledge, was the first attempt to use charges obtained by full structure semi-empirical quantum mechanical calculations of entire complexes for structure–activity relationship calculations. The best correlation ($r^2 = 0.694$) was obtained using the charges of two significant atoms, namely the O1, which is at the reaction center and the C2, which is the closest independent non-substituted atom to the substituent, watershell and torsional angle values as descriptors. The effects of complexation on the electronic structure of ligands were also studied by multivariant analyses (PCA and PCO). A tendency was found for the change of inter-ligand correlation of atomic charges by complexation. The correlation among these charges after complex formation was different for the case of molecules substituted at *para* or at di-*meta* positions. Because no such tendency could be observed for free ligands this behavior of atomic charges in complexed compounds can be put down to ligand–receptor interactions. Same applies to the observation that there is opposite correlation between the field inductive parameter (F) and total charge transfer to the ligand in the case of *para* versus *meta* compounds.

It has been demonstrated that by following the way of thinking of traditional QSAR studies for the case of papain hydrolysis it was possible to predict enzymic reaction properties by 3D structural simulations without using any calculated energy terms. The atomic charge distribution was thus shown to play a crucial role thus; when carrying out structure–activity analysis of a receptor–ligand complex it might be a valid approach to use accurate charge values in combination with other descriptors. However, tendencies regarding atomic charges among a set of ligands can be different when free and complexed ligands are subjected to calculations. In the presented case the atoms of the reaction center (COO group) was differently affected by the complex formation in the case of highly similar compounds. The work tried to highlight the importance of further studying the charge transfer phenomenon in order to get a more thorough view of ligand–receptor interactions.

Acknowledgements

This work was supported by the 21st Century COE Program, Human Nutritional Science on Stress Control, Tokushima, and by CREST-JST.

References and notes

1. Hansch, C.; Leo, A. *Exploring QSAR*; American Chemical Society: Washington, 1995.
2. Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
3. Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., III. Assessing scoring functions for protein–ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–3034.
4. Kollman, P. A. Free energy calculations: application to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
5. Masukawa, K. M.; Kollman, P. A.; Kuntz, I. D. Investigation of neuraminidase—substrate recognition using molecular dynamics and free energy calculations. *J. Med. Chem.* **2003**, *46*, 5628–5637.
6. Tominaga, Y.; Jorgensen, W. L. Genaral model for estimation of the inhibition of protein kinases using Monte Carlo simulations. *J. Med. Chem.* **2004**, *47*, 2534–2549.
7. Smith, R. N.; Hansch, C.; Kim, K. H.; Omiya, B.; Fukumura, G.; Selassie, C. D.; Jow, P. Y. C.; Blaney, J. M.; Langridge, R. *Arch. Biochem. Biophys.* **1982**, *215*, 319.
8. Compadre, C. M.; Hansch, C.; Klein, T. E.; Langridge, R. The structure–activity relationship of the papain hydrolysis of *N*-benzoylglycine esters. *Biochim. Biophys. Acta* **1990**, *1038*, 158–163.
9. Hansch, C.; Klein, T. E. Molecular graphics and QSAR in the study of enzyme–ligand interactions. On the definition of bioreceptors. *Acc. Chem. Res.* **1986**, *19*, 392–400.
10. Selassie, D. S.; Klein, T. E. *Building Bridges: QSAR and Molecular Graphics. 3D QSAR in Drug Design*; Escom Science Publishers B.V.: Leiden, 1993.
11. Swain, C. G.; Lupton, E. C., Jr. Field and resonance components of substituent effects. *J. Am. Chem. Soc.* **1968**, *90*, 4328.
12. Hansch, C.; Leo, A. *Substituent constant for correlation analysis in chemistry and biology*; Wiley-Interscience: New York, 1979.
13. Janowski, R.; Kozak, M.; Jankowska, E.; Grzonka, Z.; Jaskolski, M. Two polymorphic forms of papain in covalent complex With ZLFG-DAM, to be published.
14. Morgenstern, L.; Recanatini, M.; Klein, T. E.; Steinmetz, W.; Yang, C. Z.; Langridge, R.; Hansch, C. Chymotrypsin hydrolysis of X-phenyl hippurates. A quantitative structure–activity relationship and molecular graphics analysis. *J. Biol. Chem.* **1987**, *262*, 10767–10772.
15. SYBYL® 6.9.2 Tripos Inc., 1699 South Hanley Road, St. Louis, MI 63144, USA.
16. Spartan'04 Wavefunction, Inc. Irvine, CA.
17. Halgren, T. Maximally diagonal force constants in dependent angle-bending coordinates. II. Implications for the design of empirical force fields. *J. Am. Chem. Soc.* **1990**, *112*, 4710–4723.
18. Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281.
19. Clark, M.; Cramer, R. D., III; Opdenbosch, N. Van *J. Comput. Chem.* **1989**, *10*, 982–1012.
20. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. AMBER 7, University of California, San Francisco, 2002.
21. Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

22. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollamn, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
23. Gaussian 98. Gaussian, Inc., Pittsburgh PA, 1998.
24. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for determining atom-centered charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269.
25. Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. LocalSCF method for semiempirical quantum-chemical calculation of ultra-large biomolecules. *J. Chem. Phys.* **2004**, *3*, 1266–1270.
26. LocalSCF2003. Fujitsu Limited, Tokyo, Japan, 2003.
27. Dewar, M. J. S.; Zoebisch, M. J. S.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
28. XLSTAT-Pro 7.5 Addinsoft, Brooklyn, NY, USA.
29. Gower, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **1966**, *53*, 325–338.
30. Dillon, W. R.; Goldstein, M. *Multivariate analysis. Methods and applications*; John Wiley & Sons: New York, 1984; 23–52.
31. MVSP 3.12. Kovach Computing Services, Wales, UK.
32. Walsh, C. *Enzymatic Reaction Mechanisms*; Freeman: San Francisco, 1979; p 98.
33. Smith, R. N.; Hansch, C.; Kim, K. H.; Omiya, B.; Fukumura, G.; Selassie, C. D.; Jow, P. Y.; Blaney, J. M.; Langridge, R. The use of crystallography, graphics, and quantitative structure–activity relationships in the analysis of the papain hydrolysis of X-phenyl hippurates. *Arch. Biochem. Biophys.* **1981**, *215*, 319–328.