K08 Investigating the Links Between Traditional QSAR Approach and Molecular Simulations of Protein-Ligand binding.

○ Zsolt Lepp, Hiroshi Chuman. Dept. of Pharmaceutical Sciences, The University of Tokushima

Introduction

Predicting enzymatic reactions is a crucial problem of biochemistry thus, various methods to estimate biding properties for protein– ligand complexes has been emerged. These methods apply two main approaches; one is the QSAR type analysis 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.

The purposes of this work were to search for links between traditional QSAR and molecular level simulations and to reinterpret traditional QSAR descriptors using detailed 3D structural information.

Results and discussion

Docking and Molecular dynamics and semiempirical quantum mechanics simulations were carried out to create accurate 3D receptor- ligand complexes. Various steric, electronic and hydrophobic descriptors were derived from these structures and used to estimate activities by the linear combination of the most significants of these descriptors.

The work was based on previous analyses of the hydrolysis of *N*–benzoylglicine esters by a cisteine protease, papain.

In the original study the following descriptors were used: the field inductive effect (F)[1], the hydrophobic substituent constant (Π)[2] and the distance of from steric effects from substituents, respectively.

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

n=37, r=0.939, s=0.366

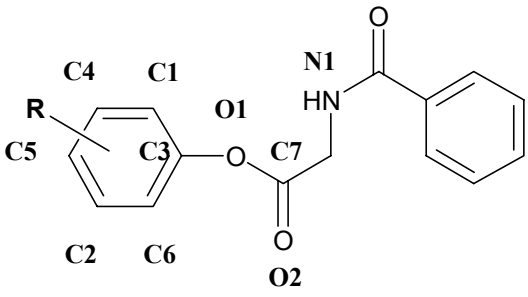

Figure 1. General structure of substrates

Compound	Obs[a].	Pred[b].
1 H	3.86	4.341
2 4–NH_2	3.59	3.581
3 4–CH_3	3.99	4.333
4 4–OCH_3	4.05	3.981
5 4–Cl	4.2	4.070
6 4–CN	4.23	4.262
7 4–SO_2NH_2	4.28	4.694
8 4–$CONH_2$	4.6	4.726
9 4–$COCH_3$	4.95	4.792
10 3–NH_2	3.62	3.957
11 3–$NHCOCH_3$	3.83	4.021
12 3–F	4.03	3.952
13 3–$CONH_2$	4.12	3.960
14 3–CN	4.32	4.006

zlepp@fc.ph.tokushima-u.c.jp

Compound	Obs[a].	Pred[b].
15 3–CH_3	4.58	4.447
16 3–Cl	4.73	4.660
17 3–CF_3	4.28	4.121
18 3,5–$(OCH_3)_2$	4.49	4.431
19 3,5–$(CH_3)_2$	4.69	4.513
20 3,5–$(Cl)_2$	4.91	4.785
21 3–CH_3,5–C_2H_5	4.97	4.686

Table1. The substituents, the (*a*) measured and the (*b*) predicted activities of ligands

The requirement for the descriptors determined from the 3D protein–ligand structures were to correspond to these classical parameters but give more accurate and physicochemically meaningful description for the properties of complexes. The homogenous set of ligands that were subjected to simulation are summarized in Table1.

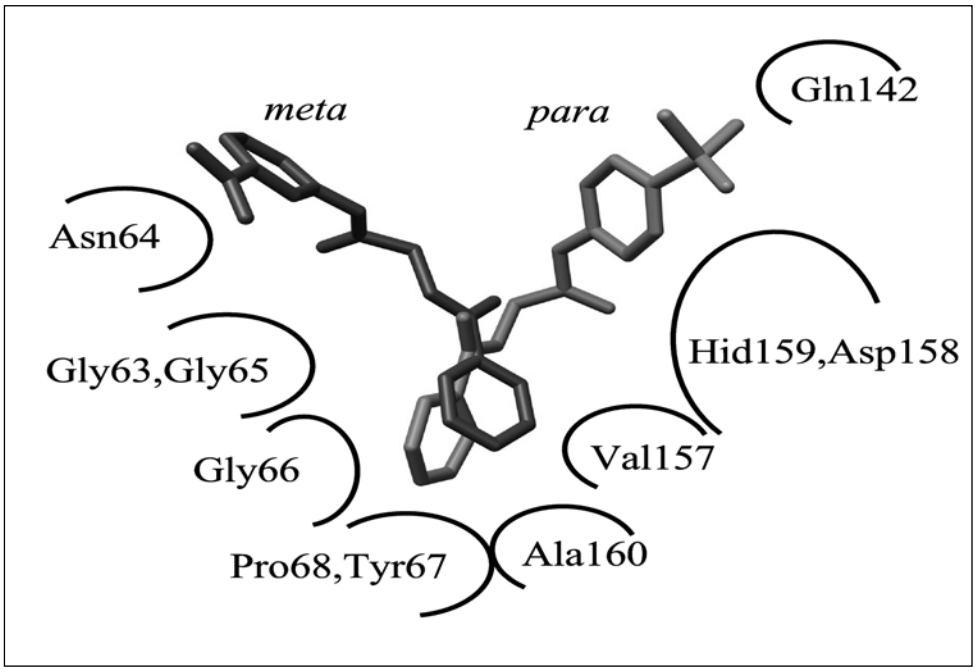

Figure 2. Difference between the orientation of *para* and *meta* substituted ligands after molecular dynamics simulation

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[4].

Molecular dynamics calculation was used to further optimize conformation for complexes at 300K and to create a statistical ensemble for the following structure– activity relationship analyses.

Amber 7[5] was used for the molecular dynamics simulations with Amber *ff99* (protein) and GAFF (ligand) force fields. The atomic charges for ligands were calculated by carrying out single point calculation on the HF/6–31G* level using Gaussian 98[6] and were fitted using the standard RESP[7] method by Amber 7. The protein was neutralized adding five chlorine anions. A TIP3 water cap of 20Å radius around the Cys25 residue was added to include solvent effects at the active center. The simulations were carried out at 300K. The timestep was 0.5fs. The equilibration run took 690 ps and the production run was1000 ps.

Semi empirical quantum mechanics calculations were performed on the full complexes by the LocalSCF[8] method and LocalSCF2003[9] program. Ten snapshots during the 1ns of MD production run were collected (one per every 100ps) and single point calculations were carried out in vacuo with all the water molecules included. For all cases the

semi-empirical AM1 Hamiltonian was used. The 10 snapshots of each ligands and complexes were averaged and the mean values were used for analysis. The ligands were separated from the complexes and single point calculations with identical parameters to those of the complexes were performed in vacuo.

The active site of the enzyme has a hydrophobic subsite, which includes Tyr61, Tyr67, Pro68, Trp69, Phe207, and Ala160. This subsite can firmly bind the unsubstituted phenyl ring of the *N*-benzoylglicine esters. The ester part of molecules are more diverging. There is an unambiguous difference between *para* and *meta* substituted ligands, because the former ones are closing towards residues Glu158 and His159, whilst the latter ones are rather moving towards Asn64, Figure 2.

Descriptors were determined from molecular dynamics trajectories and snapshots. Along the1ns trajectory of production run 2000 snapshots were collected from which various descriptors were calculated. These descriptors were (1) structural parameters of ligands (Figure 1) 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). Water shells, which were defined by counting the number of water molecules around the given atom inside a sphere of 5Å radius, around various heavy atoms were determined. (3). Atomic charges of nine atoms as depicted in Figure 1 were used as descriptors for electronic properties of ligands.

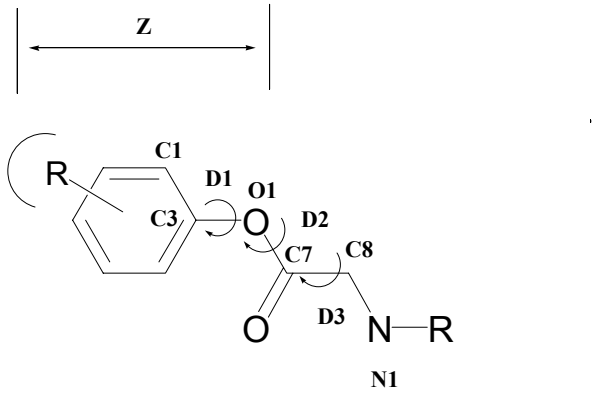

Figure 3. Steric descriptors for regression models

Multiple linear regression analyses were performed to establish structure-activity relationship models using the statistical software package Xlstat pro7.5[10]. Measured enzyme activities were obtained from the literature[3]. The correlation coefficient, r^2 was maximized whilst the number of descriptors 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.

The equations with best correlation using 2, 3 and 4 descriptors were as follows (Eq. 2-4);

$$\log 1/K_m = 2.24 + 2.65*C2 - 11.73*O1 \quad (2)$$

n=21, r=0.717, s=1.675, F=9.516,

$$\log 1/K_m = 1.67 + 2.64*C2 - 11.67*O + 4.04E-03*D1 \quad (3)$$

n=21, r=0.774, s=1.381, F=8.474,

$$\log 1/K_m = 2.93 + 2.56*C2 - 12.04*O1 - 0.22*WSC2C6 - 0.33*WSN1 \quad (4)$$

n = 21, r = 0.835, s = 1.043, F = 9.214.

Where O1 and C2 are atomic charges, WSC2C6 is the average watershell value around atoms C2 and C6 and WSN1 is the watershell of atom N1.

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. The sign of the correlation coefficient (Eqs. 3-5) is negative for O1 whilst that of the C2

meaning that charge polarization promotes the reaction which corresponds with the chemical sense. The negative sign of the watershell descriptors is also reasonable; it reveals that the more dehydration occurs the stronger the ligand– receptor interaction is. The watershell as a descriptor does not only characterizes dehydration effect, but it is probably also influenced by the relative orientation of the ligand between the water– receptor phases. For example the two watershell values can be replaced by the D1 dihedral angle which is a geometrical measure (Eq 3). The significance of watershell value around N1, which is the nitrogen atom of amide group, reveals that though the orientation of amide part is very similar for all the complexes this substructure should also be taken into consideration.

The best correlation obtained was (r=0.835, r^2=0.697 Eq. 5) better than the correlation that can be achieved using the original descriptors (r=0.724, r^2=0.524, Eq. 5).

$$\log 1/K_m = 3.976*F + 0.168*Z + 0.884*\Pi_{3'} + 3.104 \quad (5)$$

n=23, r=0.724, s=0.304, F=6.604.

Conclusion

It can be concluded, that by following the way of thinking of traditional QSAR studies it is possible to accurately predict enzymic reaction properties by 3D structural simulations without using any calculated energy terms. Especially the atomic charge distribution plays crucial role thus, when carrying out structure–activity analysis of a receptor–ligand complex it is recommended to use accurate charge values in combination with other descriptors.

ACKNOWLEDGEMENT

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

References

(1). Swain C. G.; Lupton, E. C., Jr. Field and resonance components of substituent effects *J. Am. Chem. Soc.* 1968, *90,* 4328.

(2). Hansch, C.; Leo, A.; Substituent Constant for Correlation Analysis in Chemistry and Biology. Wiley–Interscience, New York, 1979.

(3). Compadre C.M.; Hansch, C.; Klein, T. E.; Langridge, R. The structure–activity relationship of the papain hydrolysis of *N*–benzoylglycine esters. Biochimica et Biophysica Acta 1990, *1038,* 158-163

(4). SYBYL® 6.9.2 Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.

(5). AMBER 7, University of California, San Francisco, (2002).

(6). Gaussian 98, Gaussian, Inc., Pittsburgh PA 1998.

(7). C.I. Bayly, P. Cieplak, W.D. Cornell & P.A. Kollman. A Well–Behaved Electrostatic Potential Based Method Using Charge Restraints For Determining Atom–Centered Charges: The RESP Model. J. Phys. Chem. 1993, *97,* 10269.

(8). Anikin, N. A.; Anisimov, V. M.; Bugaenko, V. L.; Bobrikov, V. V.; Andreyev, A. M. LocalSCF method for semiempirical quantum–chemical calculation of ultralarge biomolecules *J. Chem. Phys.* 2004, *3* 1266–1270.

(9). LocalSCF2003, Fujitsu Limited, 2003, Tokyo, Japan.

(10). XLSTAT–Pro 7.5 Addinsoft, Brooklyn, NY, USA.