Journal of Computer-Aided Molecular Design, **15:** 195–202, 2001.
KLUWER/ESCOM

A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors

Antti Poso[a], Jukka Gynther[a] & Risto Juvonen[b]
[a]*Department of Pharmaceutical Chemistry,* [b]*Department of Pharmacology and Toxicology, University of Kuopio, P.O. Box 1627, FIN 70211 Kuopio, Finland*

Received 7 February 2000; accepted 9 October 2000

Key words: Comparative molecular field analysis, Cytochrome P450, GOLPE, GRID, smart region definition

Summary

Structure-activity relationships of 23 P450 2A5 and 2A6 inhibitors were analysed using the CoMFA [1] and GOLPE/GRID with smart region definition (SRD) [2]. The predictive power of the resulting models was validated using five compounds not belonging to the model set. All models have high internal and external predictive power and resulting 3D-QSAR models are supporting each other. Both Sybyl and GOLPE highlight properties near lactone moiety to be important for 2A5 and 2A6 inhibition. Another important feature for pIC_{50} was the size of the substituent in the 7-positon of coumarin. The models suggest that the 2A5 binding site is larger that that of 2A6 due to larger steric regions in the CoMFA coefficient maps and corresponding GOLPE maps. In addition, the maps reveal that 2A6 disfavours negative charge near the lactone moiety of coumarin.

Introduction

Cytochrome P450 (CYPs) form a large superfamily of heme enzymes, which catalyse the oxygenation of many endogenous and exogenous compounds [3]. Coumarin is specifically hydroxylated to 7-hydroxycoumarin in mice by CYP2A5 and in humans by CYP2A6 [4]. These enzymes share a 82% similarity in their amino acid sequences. Most of the known CYP2A5 ligands include a lactone moiety (Figure 1). According to our earlier CoMFA study, a negative charge near this lactone moiety is favourable for the inhibition of CYP2A5 [5]. At that time, there was no information available concerning the structure-activity relationships of CYP2A6 inhibitors. Our group has recently measured the IC_{50} values of 28 compounds, using both CYP2A5 and CYP2A6 as target systems [6]. CoMFA is typically used to study only one type of binding activity (either ligands or inhibitors), but it has also proved to be useful in comparing the structure-activity relationships (SAR) of receptor subtypes [7, 8]. The initial aim of this study was to construct QSAR models for CYP2A5 and 2A6 to be used in future work. Another aim was to reveal the SAR differences between CYP2A5 and 2A6 ligands.

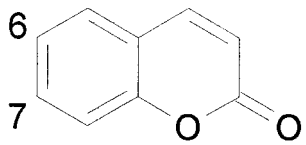


Figure 1. Structure of coumarin.

As stated above, we have earlier found lactone moiety to be important for 2A5 inhibition. Since GRID [11] includes directional hydrogen bonding function [12, 13], we decided to use it with GOLPE/SRD [10,2] to reveal the importance of lactone moiety. In addition, CoMFA [1], as implemented in Sybyl [9], was used. In SRD, normal CoMFA regions are replaced by regions that are defined by the type of information that is included, and not by the 3D geometry. The grouping algorithm uses either PLS/PCA weights or loadings to choose starting points, 'seeds', for the regions. Every X-variable is then assigned to the nearest seed, if the distance is within a user-defined value. Variables, which are not assigned to any seed, are discarded. Generated re-

Table 1. Actual and predicted pIC_{50}-values for mouse (CYP-2A5) and human (CYP-2A6) coumarin 7-hydroxylation. The predicted pIC_{50}-values were obtained from the SRD GOLPE models presented in Table 2. The asterisks shown with non-lactone compounds denote the atoms used for superimposition. Test set compounds are marked with °. The *in vitro* pIC_{50} values were obtained from Juvonen et al. [6]

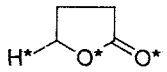
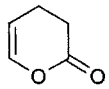
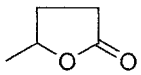
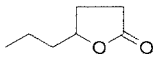
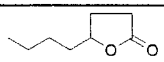
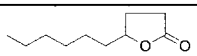
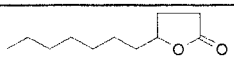
		pIC_{50} CYP2A5		pIC_{50} CYP2A6	
Inhibitor	Structure	Actual	Predicted	Actual	Predicted
γ-butyro-lactone		1.73	2.33	0.46	1.69
5,6-dihydro-2H-pyran-2-one		2.22	2.38	1.48	1.68
2H-pyran-2-one		2.28	2.40	1.67	1.74
ε-capro-lactone		3.05	2.53	1.79	1.68
γ-valero-lactone*		2.23	2.84	2.08	2.08
γ-hepta-lactone		4.54	4.81	3.25	3.07
γ-octanoic-lactone		5.24	4.99	3.61	3.09
γ-decano-lactone		5.68	5.06	2.98	2.85
γ-undecano-lactone		5.41	5.25	2.67	2.85

Table 1. Continued

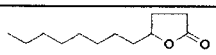
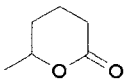
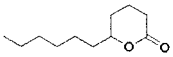
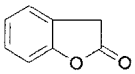
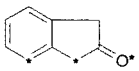
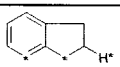
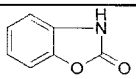
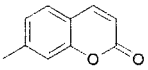
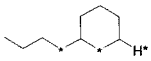
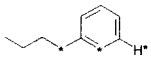
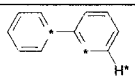
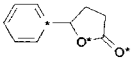
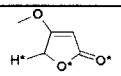
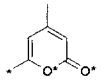
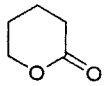
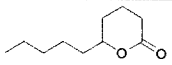
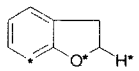
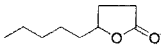
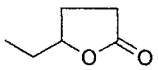
γ-dodecano-lactone		5.41	5.26	2.94	2.65
δ-hexa-lactone		2.66	2.96	1.91	2.50
undecanoic-δ-lactone		4.93	5.22	2.66	2.98
2-couma-rone		4.33	4.07	3.52	3.83
2-indanone		4.29	4.11	4.25	3.73
indan		4.08	4.04	4.30	4.38
2-benzoxa-zolinone		2.98	4.27	3.06	3.95
7-methyl-coumarin		5.40	4.88	4.52	3.65
butylcyclo-hexane		5.24	4.99	4.37	3.85
butyl-benzene		4.79	5.22	4.12	4.08
biphenyl		4.54	4.71	4.09	4.17
γ-phenyl-γ-butyro-		5.62	4.77	3.80	3.04

Table 1. Continued

lactone					
4-methoxy-2(5H)-furanone		2.14	2.11	1.75	1.31
4,5-dimethyl-α-pyrone		3.38	2.80	2.80	2.55
δ-valero-lactone°		2.01	2.32	1.54	1.38
δ-decano-lactone°		5.42	5.23	3.13	3.13
2,3-dihydro-benzofuran°		4.10	4.09	3.75	4.00
γ-nonanoic-lactone°		5.72	5.13	3.62	3.07
γ-capro-lactone°		3.14	3.86	2.42	3.08

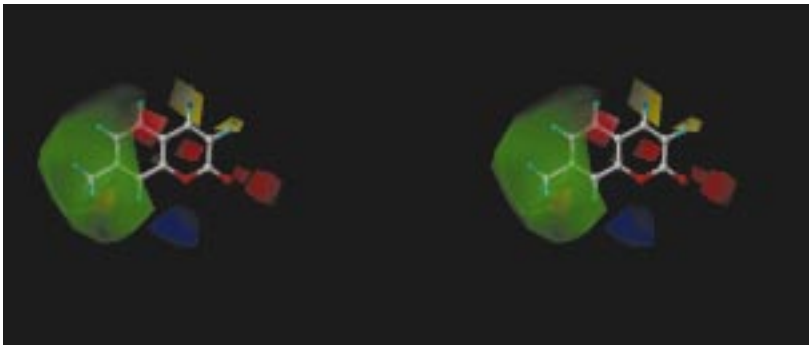


Figure 2. Stereo figure of CoMFA (CYP2A5) coefficient maps (actual values) for steric (yellow −0.001; green 0.003) and electrostatic properties (red −0.003; blue 0.003). Reference structure is 7-methylcoumarin.

Table 2. Statistics from the CoMFA and GOLPE analyses; q^2 = the crossvalidated correlation coefficient; S_{PRESS} (Sybyl) and SDEP (GOLPE) = standard deviations for the error of predictions; N = number of PLS components; $r^2_{pred} = r^2$ of prediction.

Models	q^2	r^2	S_{PRESS}/SDEP	N	r^2_{pred}
CYP2A5CoMFA	0.79	0.94	0.66	3	0.83
CYP2A5 GOLPE	0.86	0.94	0.48	2	0.90
CYP2A6 CoMFA	0.81	0.97	0.54	4	0.77
CYP2A6 GOLPE	0.78	0.93	0.51	2	0.76

gions are further collapsed if they meet the criteria of similarity and minimum distance. After region selection, F-factorial design is used to choose those regions, which contribute to biological activity. The benefit of using SRD over traditional geometric region definition is structural differences in the series are reflected in the groups of variables. These groups are less prone to produce artifacts than individual variables, as the group formation requires similarity inside the group.

Materials and methods

All molecular modelling and QSAR studies were performed on an SGI O2 workstation. Construction of the molecules, superimposition and CoMFA was done by using Sybyl 6.6 [9]. The molecular structures were created using the sketch option in Sybyl, and structures were further minimised with MMFF94 force field [14]. Atomic point charges (scaled ESP) were calculated using MNDO Hamiltonian within MOPAC 6.0 [15]. The conformation of any side chain was fully extended. Minimised molecules were superimposed by using a rigid fit with three matching atoms, namely the oxygens of the lactone moiety and the 7-position of coumarin, or corresponding moieties (marked with an asterisk in Table 1). The SRD PLS-analysis was made by using GOLPE 4.5 [10] and GRID (version 18) [11] using phenolic hydroxyl probe. The selection of the GRID probe was based on the fact that phenolic hydroxyl probe is capable to act as a hydrogen bond donor and acceptor and to interact with aromatic ring systems. In case of classical CoMFA steric (sp^3-carbon probe) and electrostatic fields (charge +1) were included in the analysis. The grid spacing was 2 Å (Sybyl) or 0.5 Å (GOLPE). Prior to PLS analysis PCA (based on GRID phenolic OH-probe) was carried out to check structural variation of compounds (no outliers found). PCA was also used to select five diverse compounds (δ-valerolactone, δ-decanolactone, 2,3-dihydro-benzofuran, γ-nonanoiclactone and γ-capro-lactone) to be used as a test set.

The standard deviation threshold for exclusion of data columns from the PLS analysis was set at 2 kcal/mol (in Sybyl) and 0.1 kcal/mol (GOLPE) and applied for both cross-validated and nonvalidated analyses. Because the SRD method includes effective noise filtering, the value for the standard deviation threshold can be lower than in Sybyl. In the GOLPE

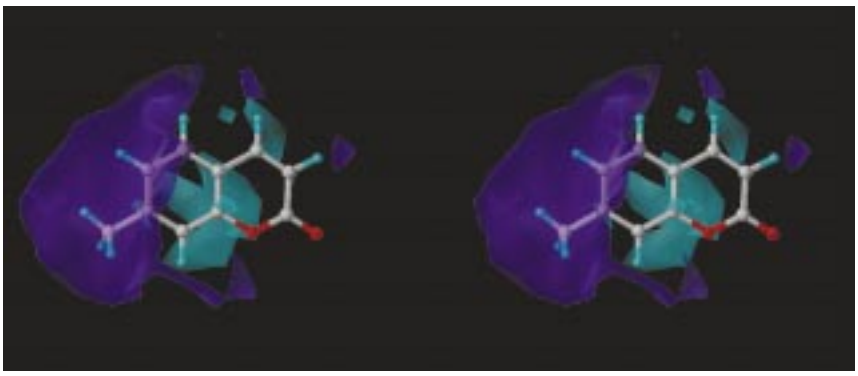


Figure 3. Stereo figure of GOLPE (CYP2A5) coefficient map (actual values) for phenolic OH-probe (cyan −0.001; purple 0.003). Reference structure is 7-methylcoumarin.

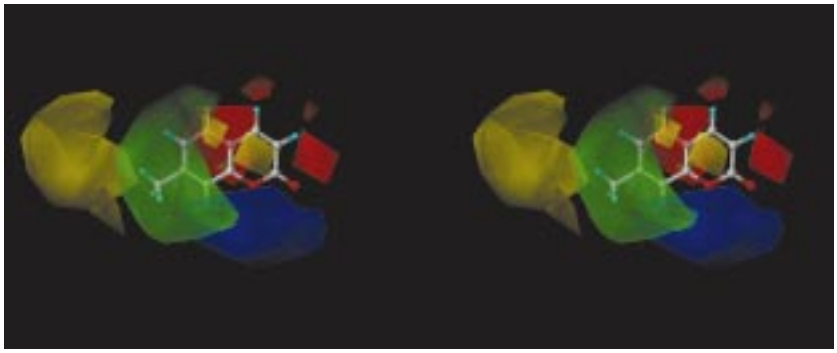


Figure 4. Stereo figure of CoMFA (CYP2A6) coefficient maps (actual values) for steric (yellow −0.001; green 0.003) and electrostatic properties (red −0.004; blue 0.003). Reference structure is 7-methylcoumarin.

analysis, all two and three level variables were removed and positive values (repulsive term) lower than 0.1 equated with zero. The optimum number of PLS components was selected by using a cross-validation with 5 groups to derive the lowest S_{PRESS} in Sybyl and SDEP in GOLPE. In GOLPE, this calculation was repeated 20 times to verify the stability of q^2-values. The number of PLS components was not allowed to be higher than five to ensure simplicity of the models. SRD, with F-factorial selection, was used with two PLS-components in the selection phase, together with 5 random group PLS validations (20 SDEP). Test set was used to calculate the predictive r^2 (r^2_{pred}).

The biological data (IC_{50} values) were obtained from Juvonen et al. [6] and transformed to pIC_{50}.

Results

Statistics from the CoMFA and GOLPE/SRD analyses for CYP2A5 and 2A6 are shown in Table 2. Both data sets gave q^2 values over 0.7, with 2–4 PLS components and good external predictivity. The coefficient fields for CYP2A5 and CYP2A6 are shown in Figures 2–5.

Discussion

The 3D-QSAR models in both cases, and with both methods, were statistically valid and the statistical differences between CoMFA and GOLPE were not of significance. All models had both internal and external q^2 values over 0.7. The only clear statistical difference between the two methods was in the number of selected PLS components (2 with GOLPE, 3 and 4 with Sybyl). 2-Benzoxazolinone showed high residual values (1.29/2A5 and 0.89/2A6 in GOLPE models) but this compound was not dropped from the final analysis because it was not an outlier in the PLS plot (plot not shown).

The CoMFA/GOLPE coefficient contour maps were clear with large continous areas in both data sets (Figures 2–5). In the case of 2A5, one can find two

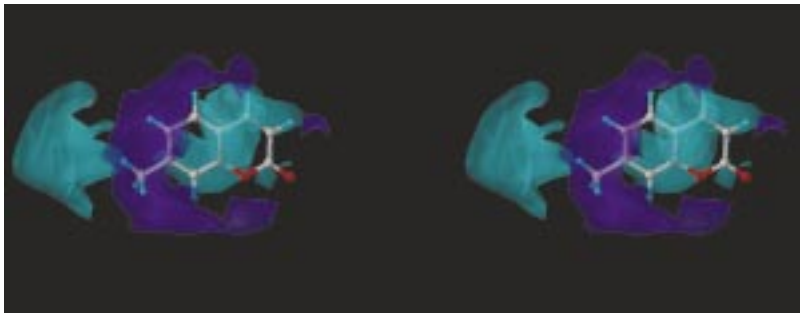


Figure 5. Stereo figure of GOLPE (CYP2A6) coefficient map (actual values) for phenolic OH-probe (cyan −0.001; purple 0.003). Reference structure is 7-methylcoumarin.

areas of interest, one near the lactone oxygen and another near the position para to the lactone oxygen, where a positive charge was correlated with the pIC_{50} values (Figure 2). A negative charge is favoured near the carbonyl oxygen and over both sides of coumarin ring plane. In the steric map, a negative interaction (attractive) near the position para to the lactone oxygen correlated with a high pIC_{50} value. A large steric allowed region was located around the 7-position of coumarin. This area coincided with a large area in the GOLPE/GRID map (Figure 3) where a repulsive interaction between a ligand and the phenolic hydroxyl probe gave a positive correlation with pIC_{50}. In addition, an OH-attractive area was found over coumarin ring plane and it coincided with area where negative charge is favourable. This OH-attractive area also had the highest coefficient values in the GOLPE/GRID maps.

In 2A6 models, the findings were about the same as for 2A5. The main difference was a large area in both CoMFA (Figure 4) and GOLPE/GRID maps (Figure 5), where attractive (negative) interactions had a positive correlation with pIC_{50}. This area was located near the region where steric interaction (repulsive) was favourable. In electrostatic the main difference also was near the lactone moiety. In this area, positive charge is again favourable, but the size of the area was much larger than in the case of 2A5.

In conclusion, substituents near position 7 of coumarin in the 2A5 model seems to be important for the SARs in Figures 2 and 3. If the 'negative charge favourable' areas and GOLPE/GRID information are combined then the properties of the ring system are also prominent. Probably the ring system should be planar, with negative charge and capable to attractive interaction with OH-probe. This would correspond to aromatic system, or at least to system, where – electrons are available. On the other hand, lactone moiety do not seem to be the best one, since negative charge near ring oxygen is not favourable while opposite is true for carbonyl oxygen. It is remarkable, however, that the GRID fields near this area are not correlated with pIC_{50} values. This would indicate that hydrogen bonding between the lactone oxygen and enzyme residues are not of major importance in the inhibition process. This phenomenon is also seen in case of 2A6 (Figure 4). The area near the carbonyl oxygen, where a positive charge is favourable, is even larger than with 2A5. The meaning of this area's significance is not clear, as there is also a small but strong area near the carbonyl oxygen, where a negative charge is favourable. In addition, GOLPE/GRID (Figure 5) indicates that a repulsion near the lactone moiety, between the OH-probe and a ligand, is correlated with pIC_{50}. All of the ligands within the data set were lacking substituents that occupy this area. Because the OH-probe had a strong attractive interaction with lactone, the results from GOLPE has at least partly explain the increased activity of non-lactone compounds by a decreased attraction between ligand and probe. For example, 2-coumarone had a 2A6 pIC_{50} value of 3.52, and for the two corresponding non-lactone compounds 2-indanone 4.25, and indan 4.3. Correlations between steric interaction and pIC_{50} (2A6) appeared almost identically in the steric CoMFA and in GOLPE/GRID maps (Figures 4–5). Small substituents at the 7-position were favourable. However, the large steric unfavourable region indicated that the optimal size of the ring substituent was limited, and substituents that were too large eventually lead to a less-potent inhibitor.

When we compare the QSAR-data, it can easily be recognize that the size of the ring substituent was the major discriminating factor between the 2A5 and

2A6 models. Inhibitors of 2A5 are allowed to include larger substituents on the coumarin ring system, while the binding site of 2A6 inhibitors was, in this sense, smaller. Another main factor was the difference in electrostatic requirements near the lactone moiety. It seems clear that the most potent 2A6 inhibitors should not include lactone moiety, or maybe any heteroatoms, near the corresponding position. The same is probably true for 2A5, but only partially, because the 'positive charge favourable' area did not cover as large area as in the case of 2A6, and because of the small but strong area near carbonyl oxygen where negative charge was favourable.

References

1. Cramer, R.D., III, Patterson, D.E. and Bunce J.D., J. Am. Chem. Soc., 110 (1988) 5959.
2. Pastor, M., Cruciani, G. and Clementi, S., J. Med. Chem., 40 (1997) 1455.
3. Honkakoski, P. and Negishi, M., Drug Metab Rev., 29 (1977) 977.
4. Pelkonen, O., Raunio, H., Rautio, A., Mäenpää, J., and Lang, M.A., J. Irsh. Coll. Phys. Surg., 22 (1993) 24.
5. Poso, A., Juvonen, R. and Gynther, J., Quant. Struct.-Act. Relat. 14 (1995) 507.
6. Juvonen, R., Gynther, J., Pasanen, M., Alhava, E. and Poso, A., Xenobiotica, 30 (2000) 81.
7. Huang, Q., Liu, R., Zhang, P., He, X., McKernan, R., Gan, T., Bennett, D.W. and Cook, J.M., J. Med. Chem., 41 (1998) 4130.
8. Agarwal, A., Pearson, P.P., Taylor, E.W., Li, H.B., Dahlgren, T., Herslof, M., Yang, Y., Lambert, G., Nelson, D.L. and Regan, J.W., J. Med. Chem., 36 (1993) 4006.
9. Tripos Inc. St. Louis, USA, 1999.
10. GOLPE, Multivariate Informetric Analysis, Perugia, Italy, 1999.
11. GRID, Molecular Discovery Limited, Oxford, England, 2000.
12. Wade, R.C., Clark, K.J. and Goodford, P.J., J. Med. Chem., 36 (1993) 140.
13. Wade, R.C. and Goodford, P.J., J. Med. Chem., 36 (1993) 148.
14. Halgren, T.A., J. Am. Chem. Soc., 114 (1992) 7827.
15. QCPE No. 455, version 6.0.