Partial Charge Calculation Method Affects CoMFA QSAR Prediction Accuracy

Ruchi R. Mittal,† Lisa Harris,† Ross A. McKinnon, and Michael J. Sorich*

Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA 5000, Australia

Received August 19, 2008

The 3D-QSAR method comparative molecular field analysis (CoMFA) involves the estimation of atomic partial charges as part of the process of calculating molecular electrostatic fields. Using 30 data sets from the literature the effect of using different common partial charge calculation methods on the predictivity (cross-validated R^2) of CoMFA was studied. The partial charge methods ranged from the popular Gasteiger and the newer MMFF94 electronegativity equalization methods, to the more complex and computationally expensive semiempirical charges AM1, MNDO, and PM3. The MMFF94 and semiempirical MNDO, AM1, and PM3 methods for computing charges were found to result in statistically significantly more predictive CoMFA models than the Gasteiger charges. Although there was a trend toward the semiempirical charges performing better than the MMFF94 charges, the difference was not statistically significant. Thus, semiempirical partial charge calculation methods are suggested for the most predictive CoMFA models, but the MMFF94 charge calculation method is a very good alternative if semiempirical methods are not available or faster calculation speed is important.

INTRODUCTION

Predicting the biological activity of a candidate drug, as well as its pharmacokinetic properties and toxicity, early in the drug discovery process has the ability to significantly reduce the cost of developing drugs.[1] An important computational approach in predicting properties in drug discovery is the quantitative structure−activity relationship (QSAR) method, which correlates biological activity to chemical properties/descriptors in a series of structures.[2] Comparative molecular field analysis (CoMFA) is a popular QSAR method based on the analysis of molecular field data.

One of the two standard molecular field types used in CoMFA is the electrostatic molecular field. Although the electrostatic molecular field can be computed directly from chemical structure using ab initio quantum chemical methods, a much faster and common approach is to calculate the electrostatic molecular field on the basis of the atomic partial charge values. There are a number of methods available to calculate the atom partial charge values, and a handful of these are commonly used as part of the CoMFA process.

One of the simplest and quickest methods of assigning partial atomic charges uses the concept of equalization of electronegativities. The idea behind this method is that highly electronegative atoms draw electrons away from less electronegative atoms. These empirical approaches include the Gasteiger−Marsili charge calculation, which is assigned using a two stage algorithm. First the seed charges are assigned to each atom in the molecule, and then these initial charges are shared across bonds with a certain amount of charge moved from one atom to another. Gasteiger−Hückel is a combination of the Gasteiger and Hückel methods, which utilizes the Gasteiger−Marsili method for σ charges and an extension thereof for distributing charges across π bonds.[3] The Gasteiger−Marsili and Gasteiger−Hückel partial charge calculation methods are by far the most commonly used methods in published CoMFA studies to date. The more recently developed MMFF94 partial charge calculation method is also based on electronegativity equalization. It was primarily developed as part of the popular MMFF94 molecular force field but has been given increasing attention as a potentially more accurate general-use empirical partial charge method.

Semiempirical quantum methods are a tradeoff between the empirical and ab initio quantum chemical approaches in terms of accuracy and computational time. The semiempirical methods most frequently used in the CoMFA analysis include the modified neglect of differential overlap (MNDO), Austin model 1 (AM1), and parametric model 3 (PM3).

Because partial charges are used to calculate the electrostatic fields in CoMFA studies the quality of the partial charges calculated should influence the quality of the electrostatic field and hence the ability of CoMFA models to predict the activity of new chemicals. There have been a number of studies that directly or indirectly have assessed the influence of different partial charge calculation methods on CoMFA prediction accuracy.[4−7] Much of the available literature is conflicting, thus making it very difficult to give general guidelines for the partial charge calculation method of choice for CoMFA studies. Most studies that assess the affect of partial charge calculation method on CoMFA predictivity are based on a single data set. There are no studies that attempt to assess the statistical significance of the differences found in model predictivity and hence generalization of the findings to other data sets is very difficult.

* To whom correspondence should be addressed. Phone: +61 8 83022034. Fax: +61 8 83021087. E-mail: michael.sorich@unisa.edu.au.
† Contributed equally to the paper.

10.1021/ci800390m CCC: $40.75
Published on Web 02/25/2009

Table 1. Prealigned Molecular Datasets Sourced from the Literature for the CoMFA and CoMSIA Analysis

data set	description	*N*
ACE[8]	inhibitors of angiotensin converting enzyme	114
ACHE[8]	inhibitors of acetyl-cholinesterase	111
AI[9,10]	steroid aromatase inhibitors	78
ARB[4]	nonpeptide angiotensin II receptor antagonists	28
ATA[11]	antituberculosis agents	94
BZR[8]	inhibitors of benzodiazepine receptor	163
CBRA[12]	cannabinoid CB1 receptor agonists	32
COMT[13]	inhibitors of catechol-*O*-methyltransferase	92
COX2[8]	inhibitors of cyclooxygenase-2	322
DAT[14]	piperidine analogues for dopamine transporter	42
DHFR[8]	inhibitors of rat dihydrofolate reductase	397
DR[15,16]	antagonists of dopamine receptor	38
ECR[17]	binding of diacylhydrazine to ecdysone receptor	50
EDC[18]	estrogen disrupting chemicals	123
GHS[19]	growth hormone secretagogue mimics	31
GPB[8]	inhibitors of glycogen phosporylase b	66
GSK3B[20]	inhibition of glycogen synthase kinase 3	42
HIVPR[21]	inhibitors of human immunodeficiency virus protease	113
HIVRT[22,23]	inhibition of HIV-1 reverse transcriptase	101
KOA[24]	κ-opioid antagonists	39
MX[25]	mutagenicity of mutagen X analogues	29
PDE[26]	inhibition of phosphodiesterase-IV	29
PTC[27]	phase-transfer asymmetric catalysts	40
RYR[10,28]	binding of ryanoids to the ryanodine receptor	18
steroids	binding of steroids to carrier proteins	21
TCHK[30]	inhibition of *Trypanosoma cruzi* hexokinase	42
THERM[8]	inhibitors of thermolysin	76
THR[8]	inhibitors of thrombin	88
TP2A[31]	inhibition of topoisomerase-IIa	25
YOPH[32]	inhibitors of *Yersinia* protein tyrosine phosphatase	39

The present paper addresses the need to develop better predictive 3D-QSAR models by evaluation of the partial charge calculation methods that are readily accessible to the molecular modeler. The primary aim of this study was to compare the standard empirical Gasteiger methods with the newer MMFF94 empirical method and the commonly used semiempirical methods (AM1, MNDO, and PM3).

METHODS

Data Set. Thirty prealigned molecular data sets were sourced from the literature. The aligned 3D molecular structures and values of their experimentally derived activity were either requested from the relevant author or were extracted from the publication's Supporting Information. QSAR data sets and their respective references are listed in Table 1.

3D-QSAR Studies. CoMFA and CoMSIA modeling and analyses were carried out on Red Hat Linux enterprise edition WS4 using the QSAR and Advanced COMFA modules in Sybyl 7.3.[33] Default settings were used for the analyses. The transform (CoMFA field class) was set to Tripos standard field. The field type was set to both steric and electrostatic. Volume average type was set to none. The transition was set to smooth. In the region settings, the fields were sampled at a density (grid spacing) of 2 Å with an extension of 4 Å in all directions beyond the aligned molecules. Analyses were repeated at lower grid spacing of 1 and 1.5 Å as well to confirm the results are unchanged. The maximum field values were truncated to ±30.0 kcal/mol for both steric and coulomb-derived electrostatic interaction energies. The maximum number of partial least-squares regression (PLSR)[29,34] components was set to six. The column filter was set to 2.0 kcal/mol. The same grids constructed for the calculation of CoMFA fields were used for the CoMSIA fields' calculation. The steric and electrostatic similarity indices descriptors were calculated using a sp3 carbon probe atom with a charge of +1.0 and a radius of 2.0 Å. The attenuation factor for the Gaussian-type distance was set to 0.3. All of the above procedures (QSAR table generation, molecular field calculations, and PLSR) were automated with the use of in-house SYBYL Programming Language (SPL) scripts.

Sybyl software was used to calculate partial charges by the Gasteiger−Marsili, Gasteiger−Hückel, and MMFF94 methods. QUACPAC[35] software was used to calculate partial charges using the AM1 method, and MOPAC[36] software was used to calculate partial charges by the MNDO and PM3 methods.

CoMFA model predictivity was assessed primarily by leave-several-out (LSO) cross-validation R^2 (R_{cv}^2), but also with leave-one out cross validation. LSO was performed by dividing the data set into training and test sets in the ratio 70% and 30%, respectively. The R_{cv}^2 values for each data set and partial charge calculation methods were analyzed using SPSS[37] and *R*[38] statistical software. Paired Wilcoxon signed ranks test[39] were carried out to determine any statistical significance in regards to the difference in R_{cv}^2 values between models generated with each partial charge calculation method. Differences with $p < 0.05$ were considered statistically significant.

RESULTS

Data on the LSO R_{cv}^2 values for CoMFA and CoMSIA models of individual data sets with each partial charge calculation method can be found in Table 2 and in the Supporting Information, respectively.

Comparison of Gasteiger−Marsili to Other Partial Charge Calculation Methods. The results of the comparisons are detailed in Tables 3 and 4. In summary, the Gasteiger−Marsili and Gasteiger−Hückel partial charge calculation methods result in CoMFA models of equivalent predictivity, but the CoMSIA models were more predictive with Gasteiger−Hückel partial charges. The use of the MMFF94 or the semiempirical partial charge calculation methods, however, results in a statistically significant increase in R_{cv}^2 over Gasteiger−Marsili (or Gasteiger−Hückel) for both CoMFA and CoMSIA.

Comparison of MMFF94 and Semiempirical Partial Charge Calculation Methods. All of the semiempirical charges used in this study resulted in CoMFA/CoMSIA models of equivalent prediction ability. No statistically significant difference was observed between the semiempirical charges from the Wilcoxon signed ranks test of CoMFA/CoMSIA R_{cv}^2.There were also no statistically significant differences in the predictive ability between MMFF94 and any of the semiempirical partial charges (AM1, MNDO, or PM3).

Practical Significance of Changes in Predictive Ability. Tables 5 and 6 display the percent of data sets that have a noteworthy change in predictive ability following substitution of the Gasteiger−Marsilli partial charge calculation method. Both a typical (>10%) and conservative (>20%) definition of "noteworthy improvement" have been used in these tables to help readers judge the practical significance of the results.

Table 2. CoMFA Leave Several Out Cross-Validated R^2 Values for All Data Sets

data sets	Gasteiger	GAST_HUCK	MMFF94	AM1	MNDO	PM3
ACE	0.644	0.649	0.653	0.646	0.651	0.663
ACHE	0.544	0.545	0.535	0.529	0.542	0.533
AI	0.610	0.658	0.628	0.651	0.607	0.601
ARB	0.406	0.262	0.430	0.294	0.284	0.325
ATA	0.115	0.178	0.117	0.144	0.127	0.138
BZR	0.375	0.412	0.461	0.468	0.456	0.465
CBRA	0.604	0.582	0.639	0.631	0.606	0.652
COMT	0.362	0.457	0.592	0.657	0.610	0.644
COX2	0.358	0.360	0.395	0.360	0.388	0.357
DAT	0.773	0.823	0.828	0.845	0.850	0.855
DHFR	0.682	0.681	0.690	0.682	0.681	0.675
DR	0.757	0.787	0.791	0.790	0.781	0.770
ECR	0.492	0.425	0.514	0.526	0.491	0.545
ERB	0.349	0.349	0.360	0.367	0.353	0.374
GHS	0.306	0.407	0.352	0.360	0.390	0.368
GPB	0.141	0.262	0.444	0.336	0.322	0.321
GSK3B	0.698	0.690	0.632	0.668	0.671	0.618
HIVPR	0.597	0.493	0.591	0.600	0.582	0.598
HIVRT	0.808	0.827	0.821	0.821	0.823	0.821
KOA	0.695	0.691	0.694	0.668	0.669	0.679
MX	0.723	0.726	0.730	0.723	0.650	0.668
PDE	0.381	0.533	0.514	0.462	0.445	0.394
PTC	0.768	0.786	0.819	0.804	0.788	0.774
RYR	0.466	0.378	0.448	0.503	0.516	0.494
STEROIDS	0.760	0.766	0.776	0.814	0.800	0.810
TCHK	0.381	0.438	0.343	0.405	0.382	0.431
THERM	0.458	0.497	0.445	0.434	0.442	0.443
THR	0.595	0.586	0.599	0.611	0.596	0.617
TP2A	0.519	0.486	0.595	0.639	0.651	0.640
YOPH	0.771	0.770	0.771	0.766	0.772	0.774

Table 3. Results of Wilcoxon Signed Ranks Test to Compare the Gasteiger−Marsili Partial Charge Calculation Method to the Other Common Partial Charge Calculation Methods in Terms of CoMFA R_{cv}^2

Gasteiger−Marsili vs	P value	95% CI of R_{cv}^2 improvement over Gasteiger−Marsili
Gasteiger−Hückel	>0.05	−0.005−0.034
MMFF94	0.003	0.007−0.042
AM1	0.002	0.010−0.049
MNDO	0.039	0.0004−0.041
PM3	0.014	0.003−0.042

Table 4. Results of Wilcoxon Signed Ranks Test to Compare the Gasteiger−Marsili Partial Charge Calculation Method to the Other Common Partial Charge Calculation Methods in Terms of CoMSIA R_{cv}^2

Gasteiger−Marsili vs	P value	95% CI of R_{cv}^2 improvement over Gasteiger−Marsili
Gasteiger−Hückel	0.013	0.006−0.072
MMFF94	0.0004	0.023−0.086
AM1	0.004	0.014−0.108
MNDO	0.0005	0.019−0.112
PM3	0.002	0.024−0.114

Further Validation. To further validate the results presented above, the analyses were repeated in slightly different forms. The CoMFA models for each partial charge calculation method were assessed using leave-one-out cross validation (c.f., leave−30%-out cross validation in the earlier results) and by varying the grid spacing to 1 and 1.5 Å. The statistical significance of each test was equivalent.

DISCUSSION

On the basis of a comparison of the Gasteiger−Marsili partial charge calculation method with the MMFF94 method using 30 lead optimization projects, this study has demonstrated that the MMFF94 method results in statistically significantly higher R_{cv}^2 values than the Gasteiger−Marsili partial charge method for both CoMFA and CoMSIA. A previous study by Hou et al. found, when comparing the Gasteiger−Marsili and MMFF94 partial charges for one large set of quinazoline type epidermal growth factor receptor (EGF-R) inhibitors using CoMFA and CoMSIA, that the Gasteiger−Marsili charge resulted in a model of slightly greater predictivity than the MMFF94 charge calculation method.[40] Another study on one data set containing 44 compounds of diaryloxymethano-phenanthrene derivatives suggested that the Gasteiger−Marsili and MMFF94 charges resulted in CoMFA models with similar predictive ability.[7] The discrepancy between the results presented here with these previous studies is most likely the result of the previous comparisons being limited to a single data and hence lacking tests of statistically significance to assess the probability of the result occurring by chance.

The results presented here indicate statistically significant greater R_{cv}^2 values of models generated with the semiempirical (AM1, MNDO, and PM3) partial charge calculation methods compared to the Gasteiger−Marsili method. As an estimate, nearly 25−40% of the data sets will have a noteworthy improvement in predictive ability simply by changing the charge method (as summarized in Tables 4and 5). These results were observed in both the CoMFA and CoMSIA analysis.

Kim et al. reviewed the literature on the effect of partial charge calculation method on CoMFA quality.[5] The results

Table 5. Percentage of Data Sets That Have a Noteworthy (Either >10% or >20%) Increase and Decrease in CoMFA Model Predictive Ability Following Switching from Gasteiger−Marsilli Partial Charge Calculation to the Charge Calculation Method Indicated in the Column Header

	GH	MMFF94	AM1	MNDO	PM3
			increase		
>10%	23	23	23	29	30
>20%	17	13	20	17	20
			decrease		
>10%	13	0	3	7	7
>20%	3	0	3	3	0

Table 6. Percentage of Data Sets That Have a Noteworthy (Either >10% or >20%) Increase and Decrease in CoMSIA Model Predictive Ability Following Switching from Gasteiger−Marsilli Partial Charge Calculation to the Charge Calculation Method Indicated in the Column Header

	GH	MMFF94	AM1	MNDO	PM3
			increase		
>10%	30	40	44	40	50
>20%	27	27	27	33	33
			decrease		
>10%	10	3	7	3	10
>20%	7	0	0	0	3

from 8 studies,[4,6,41−46] each involving a single data set, were explored. The review indicated that there was no clear consensus on the question of how important the method of partial charge calculation and commented that the Gasteiger methods commonly yield CoMFA results of comparable or only slightly inferior quality compared to the semiempirical methods.[5] In a study of antitumor activity of 32 aryl-sulfonyl-imidazolidinone derivatives, Choo et al. compared the effect on CoMFA of using charges computed using AM1 and Gasteiger-Huckel methods. The AM1 charges yielded a model with higher cross-validated R^2 value of 0.642 than the Gasteiger−Huckel method yielding R_{cv}^2 value of 0.246.[47] In a study of human catechol *O*-methyltransferase enzyme kinetics, Sipila compared the CoMFA results obtained with the partial charges computed using MOPAC AM1 electrostatic potential (ESP), MOPAC AM1 Coulson charges, and Gasteiger−Huckel charges. AM1 ESP charges yielded in a model with the R_{cv}^2 and S_{press} values of 0.77 and 0.76, respectively. In contrast, the Gasteiger−Hückel model yielded an inferior model with the R_{cv}^2 and S_{press} values of 0.59 and 0.99, respectively.[48] On the basis of these previous studies, it is not clear whether the semiempirical partial charge calculation methods result in more predictive CoMFA models. The results presented here, however, are able to confirm the hypothesis that using a semiempirical method instead of the Gasteiger methods does indeed result in a statistically significant improvement in the predictivity of the CoMFA model produced.

There were no statistically significant differences in CoMFA and CoMSIA predictivity between any of the semiempirical charges used in this study. Previously the effect of MNDO, AM1, and PM3 partial charge calculation methods on CoMFA performance was studied in a single data set with AM1 and PM3 producing better CoMFA models than MNDO.[49] The Gasteiger−Marsili and semiempirical MNDO, AM1, and PM3 methods were also examined using a single data set of 37 benzodiazepine receptor ligands. The study suggested that the AM1 charge method results in higher predictivity when compared to the other charges.[6] With the larger number of data sets analyzed in the current study, however, it has been demonstrated that all the semiempirical charges result in models of comparable predictive ability.

In general, the prior literature comparing the effect of partial charge calculation methods on CoMFA predictivity are very difficult to interpret because of the conflicting results and lack of assessment of statistical significance of any differences observed. An interesting result of this study was that there were statistically significant differences in the predictivity of CoMFA models based on different electronegativity equalization methods. The more recent MMFF94 charge calculation method yielded CoMFA models that were statistically significantly more predictive than the standard Gasteiger−Hückel and Gasteiger−Marsili partial charge calculation methods. Although the semiempirical calculations provide a more detailed description of the molecule and are considered more accurate, this is at the cost of greater computational time. A simpler calculation provided by the MMFF94 method may be accurate enough to provide satisfactory predictions with the advantage of shorter run times. It should be noted, however, that with current computer hardware the additional computation time required for semiempirical partial charge calculation is unlikely to be a limiting factor for most molecular modelers. The biggest factor influencing the choice between the semiempirical and MMFF94 methods may be the relative availability of software implementing these methods to the individual modeler.

There are many factors that can affect the CoMFA analysis; therefore, there is the possibility that the results presented in the current study are dependent on other parameter settings. One study examined the effect of optimization of various parameter settings on CoMFA results using 9 different data sets. In total, 6120 models were generated using various combinations of parameter values for each data set. The study concluded that predictivity of CoMFA models can be increased by optimizing various parameters.[50] However, there was no clear consensus about the general model or optimal parameter values to be used. Most published CoMFA studies use the default settings; therefore, by using the default settings in this study, we believe the results presented are most generally applicable. Caution, however, should be applied in using these results to guide nonstandard CoMFA analysis. Alignment is a process that can affect the outcome of the CoMFA results yet it is not easily standardized between studies. The significance of alignment on partial charge calculation is unknown; therefore using one alignment method may potentially have an effect on the comparison. To partially overcome this influence we have used alignments from the original studies. The broad range of alignment methods has the advantage of reducing this bias. However, the large collection of data sets presented here could be used further to study the possible effect of these parameters (alignment and other issues) in future.

Furthermore, the present paper limits to the comparisons to partial charge calculation methods that have been commonly used in CoMFA studies to date, namely, mechanical

and semiempirical approaches. Now that differences in CoMFA predictive ability have been demonstrated to be effected by partial charge quality an obvious direction for future work is to determine whether ab initio method for calculating the electrostatic field can further improve CoMFA predictive ability.

CONCLUSIONS

The current study suggests that the partial charges calculation method can influence the predictivity of the resulting CoMFA and CoMSIA models in a statistically significant manner. The semiempirical charges (AM1, MNDO, and PM3) give superior predictive CoMFA and CoMSIA results than using Gasteiger−Hückel and Gasteiger−Marsili partial charge calculation methods. There was no significant difference between the predictivity of models generated with the different semiempirical charges.

The MMFF94 partial charge calculation method was found to result in statistically significantly more predictive CoMFA and CoMSIA models than the Gasteiger methods. No statistically significant difference, however, was observed between MMFF94 and the semiempirical charges used, although there was a trend toward slightly better predictivity with the semiempirical approaches.

On the basis of these results, semiempirical partial charges are recommended when possible to give the most predictive CoMFA/CoMSIA models. When not easily accessible or if there are concern about the computational time, MMFF94 is a very good alternative.

ACKNOWLEDGMENT

We thank the authors of the data sets used for analysis in this study.

Supporting Information Available: Supporting Table 1 details the LSO R^2_{CV} values for CoMSIA models, Supporting Table 2 provides information regarding the molecular data set files, which are also available as Supporting Information, and related links for other data sets used in the present study. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Roberts, S. A. High-throughput screening approaches for investigating drug metabolism and pharmacokinetics. *Xenobiotica* **2001**, *31* (8−9), 557–589.
(2) Dias, M.; Mittal, R.; McKinnon, R.; Sorich, M. Systematic statistical comparison of comparative molecular similarity indices analysis molecular fields for computer-aided lead optimization. *J. Chem Inf. Model.* **2006**, *46*, 2015–2021.
(3) Clark, R. D. Boosted leave-many-out cross-validation: The effect of training and test set diversity on PLS statistics. *J. Comput.-Aided Mol. Des.* **2003**, *17* (2−4), 265–275.
(4) Belvisi, L.; Bravi, G.; Catalano, G.; Mabilia, M.; Salimbeni, A.; Scolastico, C. A 3D QSAR CoMFA study of non-peptide angiotensin II receptor antagonists. *J. Comput.-Aided Mol. Des.* **1996**, *10* (6), 567–582.
(5) Kim, K. H.; Greco, G.; Novellino, E. A critical review of recent CoMFA applications. *Perspect. Drug Discovery Des.* **1998**, *12*, 257–315.
(6) Kroemer, R. T.; Hecht, P.; Liedl, K. R. Different electrostatic descriptors in comparative molecular field analysis: A comparison of molecular electrostatic and Coulomb potentials. *J. Comput. Chem.* **1996**, *17* (11), 1296–1308.
(7) Shagufta;Kumar, A.; Panda, G.; Siddiqi, M. I. CoMFA and CoMSIA 3D-QSAR analysis of diaryloxy-methano-phenanthrene derivatives as anti-tubercular agents. *J. Mol. Model.* **2007**, *13*, 99–109.
(8) Sutherland, J. J.; O'Brien, L. A.; Weaver, D. F. A comparison of methods for modeling quantitative structure−activity relationships. *J. Med. Chem.* **2004**, *47* (22), 5541–5554.
(9) Sulea, T.; Oprea, T. I.; Muresan, S.; Chan, S. L. A different method for steric field evaluation in CoMFA improves model robustness. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (6), 1162–1170.
(10) *Tripos Bookshelf*, version 7.3; Tripos Inc.: St.Louis, MO, 2006.
(11) Nayyar, A.; Malde, A.; Jain, R.; Coutinho, E. 3D-QSAR study of ring-substituted quinoline class of anti-tuberculosis agents. *Bioorg. Med. Chem.* **2006**, *14* (3), 847–856.
(12) Salo, O. M. H.; Savinainen, J. R.; Parkkari, T.; Nevalainen, T.; Lahtela-Kakkonen, M.; Gynther, J.; Laitinen, J. T.; Jarvinen, T.; Poso, A. 3D-QSAR studies on cannabinoid CB1 receptor agonists: G-protein activation as biological data. *J. Med. Chem.* **2006**, *49* (2), 554–566.
(13) Tervo, A. J.; Nyronen, T. H.; Ronkko, T.; Poso, A. A structure−activity relationship study of catechol-*O*-methyltransferase inhibitors combining molecular docking and 3D QSAR methods. *J. Comput.-Aided Mol. Des.* **2003**, *17* (12), 797–810.
(14) Yuan, H. B.; Kozikowski, A. P.; Petukhov, P. A. CoMFA study of piperidine analogues of cocaine at the dopamine transporter: Exploring the binding mode of the 3 α-substituent of the piperidine ring using pharmacophore-based flexible alignment. *J. Med. Chem.* **2004**, *47* (25), 6137–6143.
(15) Bostrom, J.; Bohm, M.; Gundertofte, K.; Klebe, G. A 3D QSAR study on a set of dopamine D-4 receptor antagonists. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (3), 1020–1027.
(16) Melville, J. L.; Hirst, J. D. On the stability of CoMFA models. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (4), 1294–1300.
(17) Nakagawa, Y.; Takahashi, K.; Kishikawa, H.; Ogura, T.; Minakuchi, C.; Miyagawa, H. Classical and three-dimensional QSAR for the inhibition of H-3 ponasterone A binding by diacylhydrazine-type ecdysone agonists to insect Sf-9 cells. *Bioorg. Med. Chem.* **2005**, *13* (4), 1333–1340.
(18) Marini, F.; Roncaglioni, A.; Novic, M. Variable selection and interpretation in structure−affinity correlation modeling of estrogen receptor binders. *J. Chem Inf. Model.* **2005**, *45* (6), 1507–1519.
(19) Wang, R. X.; Gao, Y.; Liu, L.; Lai, L. H. All-orientation search and all-placement search in comparative molecular field analysis. *J. Mol. Model.* **1998**, *4* (8), 276–283.
(20) Zhang, N.; Jiang, Y. J.; Zou, J. W.; Zhang, B.; Jin, H. X.; Wang, Y. H.; Yu, Q. S. 3D QSAR for GSK-3 β inhibition by indirubin analogues. *Eur. J. Med. Chem.* **2006**, *41* (3), 373–378.
(21) Tervo, A. J.; Nyronen, T. H.; Ronkko, T.; Poso, A. Comparing the quality and predictiveness between 3D QSAR models obtained from manual and automated alignment. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (3), 807–816.
(22) Hannongbua, S.; Lawtrakul, L.; Sotriffer, C.; Rode, B. Comparative molecular field analysis of HIV-1 reverse transcriptase inhibitors in the class of 1 (2-hydroxyethoxy)-methyl-6(phenylthio)thymine. *Quant. Struct.−Act. Relat.* **1996**, *15* (5), 389–394.
(23) Luco, J. M.; Feretti, F. H. QSAR based on multiple linear regression and PLS methods for the anti-HIV activity of a large group of HEPT derivatives. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 392–401.
(24) Li, W.; Tang, Y.; Zheng, Y. L.; Qiu, Z. B. Molecular modeling and 3D-QSAR studies of indolomorphinan derivatives as kappa opioid antagonists. *Bioorg. Med. Chem.* **2006**, *14* (3), 601–610.
(25) Bang, S. J.; Cho, S. J. Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) study of mutagen X. *Bull. Korean Chem. Soc.* **2004**, *25* (10), 1525–1530.
(26) Srivani, P.; Kiran, K.; Sastry, G. N. Understanding the structural requirements of triarylethane analogues towards PDE-IV inhibitors: A molecular modeling study. *Indian J. Chem., Sect. A: Inorg., Bioinorg., Phys., Theor. Anal. Chem.* **2006**, *45* (1), 68–76.
(27) Melville, J. L.; Lovelock, K. R. J.; Wilson, C.; Allbutt, B.; Burke, E. K.; Lygo, B.; Hirst, J. D. Exploring phase-transfer catalysis with molecular dynamics and 3D/4D quantitative structure-selectivity relationships. *J. Chem Inf. Model.* **2005**, *45* (4), 971–981.
(28) Welch, W.; Ahmad, S.; Airey, J. A.; Gerzon, K.; Humerickhouse, R. A.; Besch, H. R.; Ruest, L.; Deslongchamps, P.; Sutko, J. L. Structural determinants of high-affinity binding of ryanoids to the vertebrate skeletal-muscle ryanodine receptor—A comparative molecular-field analysis. *Biochemistry* **1994**, *33* (20), 6074–6085.
(29) Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (COMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
(30) Hudock, M. P.; Sanz-Rodriguez, C. E.; Song, Y. C.; Chan, J. M. W.; Zhang, Y. H.; Odeh, S.; Kosztowski, T.; Leon-Rossell, A.; Concepcion, J. L.; Yardley, V.; Croft, S. L.; Urbina, J. A.; Oldfield, E. Inhibition of *Trypanosoma cruzi* hexokinase by bisphosphonates. *J. Med. Chem.* **2006**, *49* (1), 215–223.

(31) Jensen, L. H.; Liang, H.; Shoemaker, R.; Grauslund, M.; Sehested, M.; Hasinoff, B. B. A 3D-QSAR study of the inhibition of the ATPase activity and the strand passing catalytic activity of topoisomerase IIα by substituted purine analogs. *Mol. Pharmacol.* **2006**, *73* (3), 686–696.

(32) Hu, X.; Stebbins, C. E. Molecular docking and 3D-QSAR studies of *Yersinia* protein tyrosine phosphatase YopH inhibitors. *Bioorg. Med. Chem.* **2005**, *13* (4), 1101–1109.

(33) *SYBYL*, version 7.3; Tripos International: St. Louis MO, 2006.

(34) Wold, S.; Johansson, E.; Cocchi, M. PLS-partial least squares projections to latent structuresIn *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed.; ESCOM: Leiden, The Netherlands, 1993; Vol. *1*, pp 523−550.

(35) *QUACPAC*, version 1.1; OpenEye Scientific Software Inc.: Santa Fe, NM, 2005.

(36) *MOPAC*, version MOPAC2007; Stewart Computational Chemistry: Colorado Springs, CO, 2007.

(37) *SPSS for Windows*, version 15.0.0; SPSS, Inc.: Chicago, IL, 2005.

(38) *R:A Language and Environment for Statistical Computing*, version 2.5.1; R Foundation for Statistical Computing: Vienna, Austria, 2005.

(39) Downing, D.; Clark, J., *Statistics The Easy Way*, 3rd ed.; Barron's: Hauppauge, NY, 1997; pp 240−247.

(40) Hou, T. J.; Zhu, L. L.; Chen, L. R.; Xu, X. J. Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (1), 273–287.

(41) Bureau, R.; Lancelot, J. C.; Prunier, H.; Rault, S. Conformational analysis and 3D QSAR study on novel partial agonists of 5-HT3 receptors. *Quant. Struct.−Act. Relat.* **1996**, *15* (5), 373–381.

(42) Folkers, G.; Merz, A.; Rognan, D. CoMFA: Scope and limitationsIn *3D QSAR in Drug Design: Theory, Methods and Applications*; Kubinyi, H., Ed. ESCOM: Leiden, The Netherlands, 1996; Vol *I*, pp 583−618.

(43) Krystek, S. R.; Hunt, J. T.; Stein, P. D.; Stouch, T. R. 3-Dimensional quantitative structure−activity-relationships of sulfonamide endothelin inhibitors. *J. Med. Chem.* **1995**, *38* (4), 659–668.

(44) Navajas, C.; Kokkola, T.; Poso, A.; Honka, N.; Gynther, J.; Laitinen, J. T. A rhodopsin-based model for melatonin recognition at its G protein-coupled receptor. *Eur. J. Pharmacol.* **1996**, *304* (1−3), 173–183.

(45) Recanatini, M. Comparative molecular field analysis of non-steroidal aromatase inhibitors related to fadrozole. *J. Comput.-Aided Mol. Des.* **1996**, *10* (1), 74–82.

(46) Waller, C. L.; Marshall, G. R. 3-Dimensional quantitative structure−activity relationship of angiotesin-converting enzyme and thermolysin inhibitors 2. A comparison of CoMFA models incorporating molecular-orbital fields and desolvation free-energies based on active-analog and complementary-receptor-field alignment rules. *J. Med. Chem.* **1993**, *36* (16), 2390–2403.

(47) Choo, H. Y. P.; Choi, S.; Jung, S. H.; Koh, H. Y.; Pae, A. N. The 3D-QSAR study of antitumor arylsulfonylimidazolidinone derivatives by CoMFA and CoMSIA. *Bioorg. Med. Chem.* **2003**, *11* (21), 4585–4589.

(48) Sipila, J.; Taskinen, J. CoMFA modeling of human catechol *O*-methyltransferase enzyme kinetics. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (1), 97–104.

(49) Navajas, C.; Poso, A.; Tuppurainen, K.; Gynther, J. Comparative molecular field analysis (CoMFA) of MX compounds using different semi-empirical methods: LUMO field and its correlation with mutagenic activity. *Quant. Struct−Act. Rel.* **1996**, *15* (3), 189–193.

(50) Peterson, S. D.; Schaal, W.; Karlen, A. Improved CoMFA modeling by optimization of settings. *J. Chem Inf. Model.* **2006**, *46* (1), 355–364.

CI800390M