Systematic Statistical Comparison of Comparative Molecular Similarity Indices Analysis Molecular Fields for Computer-Aided Lead Optimization

Mafalda M. Dias, Ruchi R. Mittal, Ross A. McKinnon, and Michael J. Sorich*

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

Received May 24, 2006

Comparative molecular similarity indices analysis (CoMSIA) is a 3D quantitative structure−activity relationship technique used to determine structural and electronic features influencing biological activity. This proves particularly useful for facilitating lead optimization projects. This study aimed to compare CoMSIA models produced using different subsets of the CoMSIA molecular fields (steric, electrostatic, hydrophobic, hydrogen-bond donor, and hydrogen-bond acceptor) in a systematic and statistically valid manner. A total of 23 data sets sourced from the literature were used to compare molecular field contribution and model predictivity using leave-one-out cross-validated R^2 values. Predictive ability varied in a highly statistically significant manner depending on the set of CoMSIA molecular fields used. In general, the greater the number of CoMSIA molecular fields included in the analysis, the better the model predictivity was. There is great redundancy in the information contained in the different CoMSIA molecular fields. When all five CoMSIA molecular fields are included, the hydrophobic and electrostatic fields had the largest and the steric field the smallest contribution. Data sets were clustered into four groups on the basis of the utility of molecular field sets to generate predictive models.

INTRODUCTION

The process of bringing a drug to the market is both expensive and time-consuming, primarily because of the low success rate of transiting chemicals from the laboratory to the market.[1,2] This can be attributed to, among other factors, problems in the lead optimization stage. Lead optimization is an important phase responsible for improving the efficacy and safety of compounds that have demonstrated some biological activity.

Certain computer technologies may improve the speed, cost, and efficiency of the lead optimization stage. Quantitative structure−activity relationship (QSAR) methods aim to decipher the relationship between the structural features of molecules and their biological activities.[3,4] Such ligand-based approaches are especially important when the 3D structure of the target protein is unavailable.[3,4] Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA)[5,6] are two of the most popular and interpretable QSAR methods. These methods employ interactive graphics and statistical techniques for correlating the shapes and other 3D properties of molecules (sampled as molecular fields) with their biological activity.[7] CoMSIA differs from CoMFA primarily in the way that the molecular fields are calculated. CoMSIA uses Gaussian-based similarity functions for molecular field calculations, while force-field-like potentials (e.g., Lennard-Jones and Coulomb) are predominantly used in CoMFA.[5,6] Because of their highly interpretable nature, they are particularly useful to medicinal chemists for rationally modifying a lead compound to improve its biological activity. By analyzing the structure−activity data, it is possible to shorten and maximally utilize the repetitive design−synthesis−test cycles of lead optimization.[8]

There are three main steps in CoMFA/CoMSIA analyses: molecular alignment, the calculation of molecular fields, and the analysis of molecular fields. In each of these steps, there are a great many possible methodological variations.[7] Despite the popularity of CoMFA/CoMSIA, there is a paucity of evidence for choosing one methodological option over another. Past comparisons of methodological variations have been limited primarily by the small number of data sets used and the lack of statistical tests to assess the validity of any differences.[9−12]

A CoMSIA model is based on the analysis of grid-sampled data from CoMSIA molecular fields. There are five different CoMSIA molecular fields, each of which is based on Gaussian functions. These are steric, electrostatic, hydrophobic, hydrogen-bond (H-bond) donor, and H-bond acceptor molecular similarity fields. Most CoMSIA models reported include steric and electrostatic molecular field data, and the hydrophobic field is also very common. However, there is little data to indicate which subset(s) of molecular fields results in the most predictive CoMSIA models. While the addition of more molecular fields will increase the chances of capturing the chemical features important for biological activity, it will also increase the difficulty of subsequent statistical analysis.

This study aimed to compare the five CoMSIA molecular fields (steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor) in order to derive insight into their relative importance and guidelines for the use of molecular fields in future CoMSIA studies. Specifically the study aimed to determine: (1) whether the predictive ability of CoMSIA

* Corresponding author tel.: +61 8 83022034; fax: +61 8 83021087; e-mail: michael.sorich@unisa.edu.au.

10.1021/ci600214b CCC: $33.50
Published on Web 08/31/2006

Table 1. Data Sets Used in Analyses

data set	description	*N*
ACE[12,14]	inhibitors of angiotensin converting enzyme	114
ACHE[14]	inhibitors of acetyl-cholinesterase	111
AI[15]	steroid aromatase inhibitors	78
ARB[16]	nonpeptide angiotensin II receptor antagonists	28
BZR[14,17]	inhibitors of benzodiazepine receptor	163
PTC[18]	phase-transfer asymmetric catalysts	40
COMT[19]	inhibitors of catechol-O-methyltransferase	92
COX2[14,20]	inhibitors of cyclooxygenase-2	322
DAT[21]	piperidine analogues for dopamine transporter	42
DHFR[14,22,23]	inhibitors of rat dihydrofolate reductase	397
DR[24,25]	antagonists of dopamine receptor	38
ECR[26]	binding of diacylhydrazine to ecdysone receptor	53
GHS[11]	growth hormone secretagogue mimics	31
GPB[14,27]	inhibitors of glycogen phosporylase b	66
HIVRT[28−30]	inhibition of HIV-1 reverse transcriptase	101
HIVPR[10]	inhibitors of HIV protease	113
MX[9]	mutagenicity of mutagen X analogues	29
PLA2[11,31]	indole-based inhibitors of phospholipase A2	11
RYR[30,32]	binding of ryanoids to the ryanodine receptor	18
STEROIDS[5,30]	binding of steroids to carrier proteins	21
THERM[6,14]	inhibitors of thermolysin	76
THR[14,33]	inhibitors of thrombin	88
YOPH[34]	inhibitors of Yersinia protein tyrosine phosphatase	39

models is significantly affected by the number and choice of CoMSIA molecular fields used, and if so, what subsets of CoMSIA molecular fields give the best average predictive ability, (2) the extent of redundancy/complementarity between pairs of CoMSIA molecular fields, (3) the relative contribution of each CoMSIA molecular field when all five CoMSIA molecular fields are included in model generation, and (4) whether the optimal choice of CoMSIA molecular fields is data-set-dependent.

DATA AND METHODS

A total of 23 data sets were sourced from the literature (Table 1). The 3D aligned compound structures of these data sets were either obtained by contacting the authors of the relevant publications or by accessing the publication's supporting information. The experimentally determined biological activity data of each data set was also either taken from the publication, its supporting information, or one of its cited references.

Once obtained, the molecules were imported into SYBYL 7.1 (Tripos Inc, MO) operating on a Silicon Graphics Octane 2 workstation. The QSAR module in SYBYL was used to calculate the CoMSIA steric, electrostatic, hydrophobic, H-bond donor, and H-bond acceptor molecular similarity fields.[6] The molecular alignment and partial charges used in this analysis were those sourced from the literature. The standard settings were used for the generation of all CoMSIA molecular fields for all data sets:

- The Gaussian function attenuation factor (α) was 0.3.
- The region file was automatically generated with grid spacing of 2 Å and extra spacing of 4 Å.
- The probe atom was C3 with a +1 charge, 1 Å radius, and +1 hydrophobicity.
- The steric contribution was reflected by the third power of the atomic radii, and hydrophobicity was assigned according to the method developed by Viswanadhan et al.[35]

For all subsets of the five CoMSIA molecular fields, a model was generated for each of the 23 data sets using partial-least-squares regression (PLSR). Up to six latent variables could be included in the model. The leave-one-out cross-validated R^2 (R_{cv}^2) was used to measure the predictive ability of each model.

$$R_{cv}^2 = \frac{\sum(Y_{pred} - Y_{mean})^2}{\sum(Y_{obs} - Y_{mean})^2}$$

In the leave-one-out method, one compound is removed from the data set and its activity is predicted using the model derived from the compounds remaining in the data set.[36] All of the above procedures (QSAR table generation, molecular field calculations, and PLSR) were automated with the use of SYBYL Programming Language scripts.

The R_{cv}^2 values for each data set and subset of CoMSIA molecular fields were statistically analyzed using SPSS 13.0 (SPSS Inc.) and R.[37] Box plots were produced to show the variation of R_{cv}^2 obtained with the different CoMSIA molecular field sets. Paired t-test and Friedman and Wilcoxon signed rank nonparametric tests were carried out to determine whether the difference between groups and pairs of CoMSIA molecular field sets was statistically significant. Statistical significance was defined here as $p < 0.05$. The tests had approximately 70% power to detect a change in R_{cv}^2 of 0.1 (based on the paired t-test and a standard deviation of 0.2).

Hierarchical clustering using the Euclidean distance function was undertaken using the R language.[37] Prior to clustering, the 31 molecular field sets were ranked for each data set on the basis of the R_{cv}^2. Thus, data sets that are most similar (i.e., clustered together) have the most predictive models generated from the same CoMSIA molecular field sets (and least predictive models generated from the same CoMSIA molecular field sets).

RESULTS

The distribution of R_{cv}^2 across the 23 data sets analyzed (Figure 1) was found to vary in a highly statistically significant manner between PLSR models generated using different subsets of CoMSIA molecular fields (Friedman test,

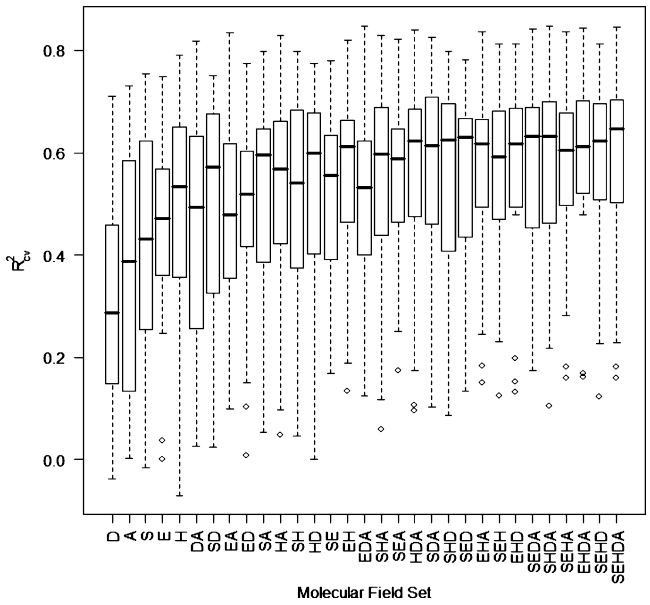

Figure 1. Distribution of the R_{cv}^2 across the 23 data sets analyzed for all subsets of CoMSIA molecular fields used, where S = steric, E = electrostatic, H = hydrophobic, D = H-bond donor, and A = H-bond acceptor.

Add

	S	E	H	D	A
S		**	**	ns	**
E	*		**	**	ns
H	ns	*		ns	ns
D	**	**	**		**
A	**	**	**	**	

ns $p > 0.05$
* $p < 0.05$ & > 0.005
** $p < 0.005$

Figure 2. The statistical significance of CoMSIA model predictivity improvement upon adding a second CoMSIA molecular field, where S = steric, E = electrostatic, H = hydrophobic, D = H-bond donor, and A = H-bond acceptor.

$p < 0.001$). In general, models built with more molecular fields were more predictive. Supporting Information Table 1 displays the statistical significance of all pairwise comparisons between CoMSIA molecular field sets.

Closer examination indicated that there were statistically significant differences between the R_{cv}^2 values of different models generated using a single CoMSIA molecular field (Friedman test, $p = 0.001$). Specifically, the CoMSIA hydrophobic field was found to be more predictive than steric and the H-bonding fields (Wilcoxon signed ranks test, $p < 0.05$). However, when comparing models generated with the same number of molecular fields (e.g., all models made with three molecular fields), there were relatively few statistically significant differences in R_{cv}^2.

Figure 2 demonstrates the statistical significance of improvement in model predictivity between models generated with a single CoMSIA molecular field and models generated with the same molecular field plus an additional CoMSIA molecular field. The areas labeled with '*' or '**' indicate instances of complementarity between molecular fields. For example, the first row indicates that adding a CoMSIA electrostatic, hydrophobic or H-bond acceptor (but not an H-bond donor) molecular field to the CoMSIA steric field results in a highly statistically significant improvement in model predictivity.

The best R_{cv}^2 values of models containing one (hydrophobic), two (hydrophobic and electrostatic), three (hydrophobic, electrostatic, and H-bond donor), four (hydrophobic, electrostatic, H-bond donor, and steric), and five CoMSIA molecular fields were compared, and a statistically significant difference was established (Friedman test, $p < 0.001$). When pairwise statistical tests (Wilcoxon signed ranks test) were used, the combination of CoMSIA hydrophobic and electrostatic molecular fields was found to result in a statistically significant larger R_{cv}^2 than that with the CoMSIA hydrophobic field alone. Similarly, the combination of hydrophobic, electrostatic, H-bond donor, and steric fields and the combination of all five CoMSIA fields both resulted in models with a statistically significant improvement in R_{cv}^2 over the combination of hydrophobic and electrostatic molecular fields.

The most commonly used CoMSIA molecular fields are steric and electrostatic followed by hydrophobic and H-bonding. Models were built progressively including molecular fields in this order to determine the advantage of adding the less commonly used CoMSIA molecular fields. Models built using steric and electrostatic CoMSIA fields have a statistically significant improvement in predictivity over using steric CoMSIA fields alone (Wilcoxon signed ranks test, $p = 0.001$). Similarly, models built using steric, electrostatic, and hydrophobic CoMSIA fields have a statistically significant improvement in predictivity over using only steric and electrostatic CoMSIA fields (Wilcoxon signed ranks test, $p = 0.02$). While not quite statistically significant, there was a strong trend indicating that models generated with all five molecular fields have superior predictive ability to those based on only the steric, electrostatic, and hydrophobic molecular fields (Wilcoxon signed ranks test, $p = 0.06$).

The relative importance of each molecular field in CoMSIA models generated using all five CoMSIA molecular fields is display in Figure 3. There was a highly statistically significant difference in relative contribution between the CoMSIA molecular fields (Friedman test, $p < 0.001$). Specifically, both the hydrophobic and electrostatic CoMSIA fields had a statistically significant greater relative contribution over the steric (Wilcoxon signed ranks test, $p < 0.005$) and H-bond acceptor (Wilcoxon signed ranks test, $p < 0.02$) CoMSIA fields. The H-bond donor and acceptor fields had a strong trend toward having a higher contribution than the steric field (Wilcoxon signed ranks test, $p \sim 0.05$).

There are 31 different subsets of the five CoMSIA molecular fields. For each data set, the CoMSIA molecular field sets can be ranked according to R_{cv}^2 (i.e., from worst to best). When these rankings of CoMSIA molecular field sets were used, data sets were clustered into four main groups (Figure 4). In the first two clusters, the worst-performing CoMSIA models were generated using only one or two molecular fields. Additionally, the CoMSIA electrostatic field

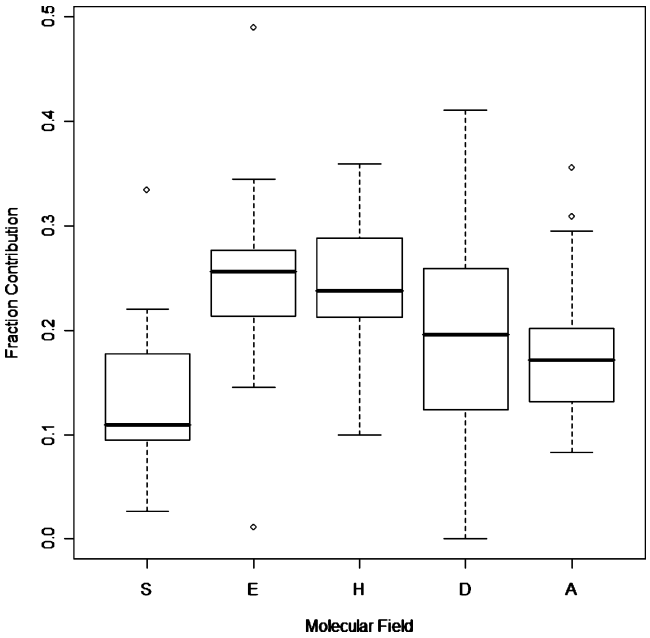

Figure 3. Distribution of relative molecular field contributions in models using all five CoMSIA molecular fields across all 23 data sets

seems to be associated with consistently better performance, particularly so in cluster two. The CoMSIA hydrophobic field appears to be associated with good performance in cluster one but not cluster two. In cluster three, the CoMSIA hydrophobic field seems to be present in the best performing models. In cluster four, the CoMSIA hydrogen-bond donor molecular field is most strongly associated with good performance.

DISCUSSION

It is quite common to compare the predictivity of models generated using different variations of the CoMFA/CoMSIA methodology. The variation may be in the alignment method, the method for calculating and sampling molecular fields, or the variable selection and statistical analysis. The major difference between this study and most if not all other previous comparisons is the number of data sets used to compare the methodological changes and the statistical validation of the comparisons. All data sets are different, and subsequently, the effect of different methodological variations will not be the same between different data sets. This has been demonstrated here by hierarchically clustering data sets on the basis of the utility of CoMSIA molecular field sets. Thus, interpreting the effect of methodological variations using a small number of data sets can easily lead to the wrong conclusions. A large sample of data sets is required to determine the general effect of the methodological variation. Tests for statistical significance are a foundation of the scientific interpretation of comparative data. However, tests for statistical significance are not possible using the small number of data sets compared in the past. Here, the predictive abilities of models generated for 23 data sets are used to compare the effects of using different subsets of molecular fields in CoMSIA. Only those differences that are statistically significant are given any weight.

There is some evidence from previous studies to indicate that including the hydrophobic and H-bond CoMSIA molecular fields in models can enhance CoMSIA model predictivity.[14,38,39] Other literature studies have also put forward the idea that hydrophobic and H-bond molecular fields together contribute most to a model's predictivity when compared to the steric and electrostatic fields.[6,39,40]

The results presented here indicate that the CoMSIA hydrophobic field plays an important role in CoMSIA models. It is the molecular field that results in the best model predictivity when used alone, and it has the equal-highest relative contribution (with the electrostatic molecular field) when all five molecular fields are included in the model. Even if the steric and electrostatic fields continue to serve as the standard molecular fields, it has been demonstrated here that adding a hydrophobic field still results in a statistically significant improvement in model predictivity. Thus, in general, it seems prudent to include a hydrophobic molecular field when undertaking CoMSIA in order to improve the predictivity of the model generated.

The importance of the CoMSIA H-bonding molecular field is less conclusive than that of the CoMSIA hydrophobic molecular field. In isolation, these two fields produced the least predictive models, although the difference from the steric field was not statistically significant. When all five CoMSIA molecular fields are included in the PLSR model, the H-bonding fields have a lesser contribution than the electrostatic or hydrophobic fields but a greater contribution than the steric field (borderline statistical significance). This theme plays out through most of the other results presented here; that is, CoMSIA H-bonding fields are of secondary utility compared to the CoMSIA electrostatic and hydrophobic fields. However, for a subset of data sets (cluster 4 in Figure 4), the CoMSIA hydrogen-bond donor was the most influential molecular field, indicating its importance in some cases.

The steric and electrostatic fields are currently the most commonly used fields in CoMFA and CoMSIA studies. As previously discussed, adding a CoMSIA hydrophobic field to the CoMSIA steric and electrostatic fields results in a statistically significant improvement in model predictivity over the combination of steric and electrostatic alone. While adding a CoMSIA H-bond donor or acceptor field to the CoMSIA steric and electrostatic fields improved model predictivity over the combination of steric and electrostatic alone, the difference does not reach statistical significance. The results presented here also indicate that models generated using all five CoMSIA molecular fields have a greater predictivity than using only steric, electrostatic, and hydrophobic CoMSIA molecular fields. However, once again, the difference falls short of statistical significance. Thus, there is a trend indicating that H-bond molecular fields are useful to include in CoMSIA, but the improvement across all data sets is relatively small and will require studies with a larger number of data sets to reach statistical significance.

On the basis of all of the results presented here, it is clear that there is great redundancy between the CoMSIA molecular fields. For example, there are few statistically significant differences in R_{cv}^2 between all of the different combinations of two CoMSIA molecular fields. Although there is a clear trend toward models built with a greater number of CoMSIA molecular fields having greater predictivity, the improvement in predictivity dwindles as more fields are added. While each CoMSIA molecular field is quite predictive alone, the benefit

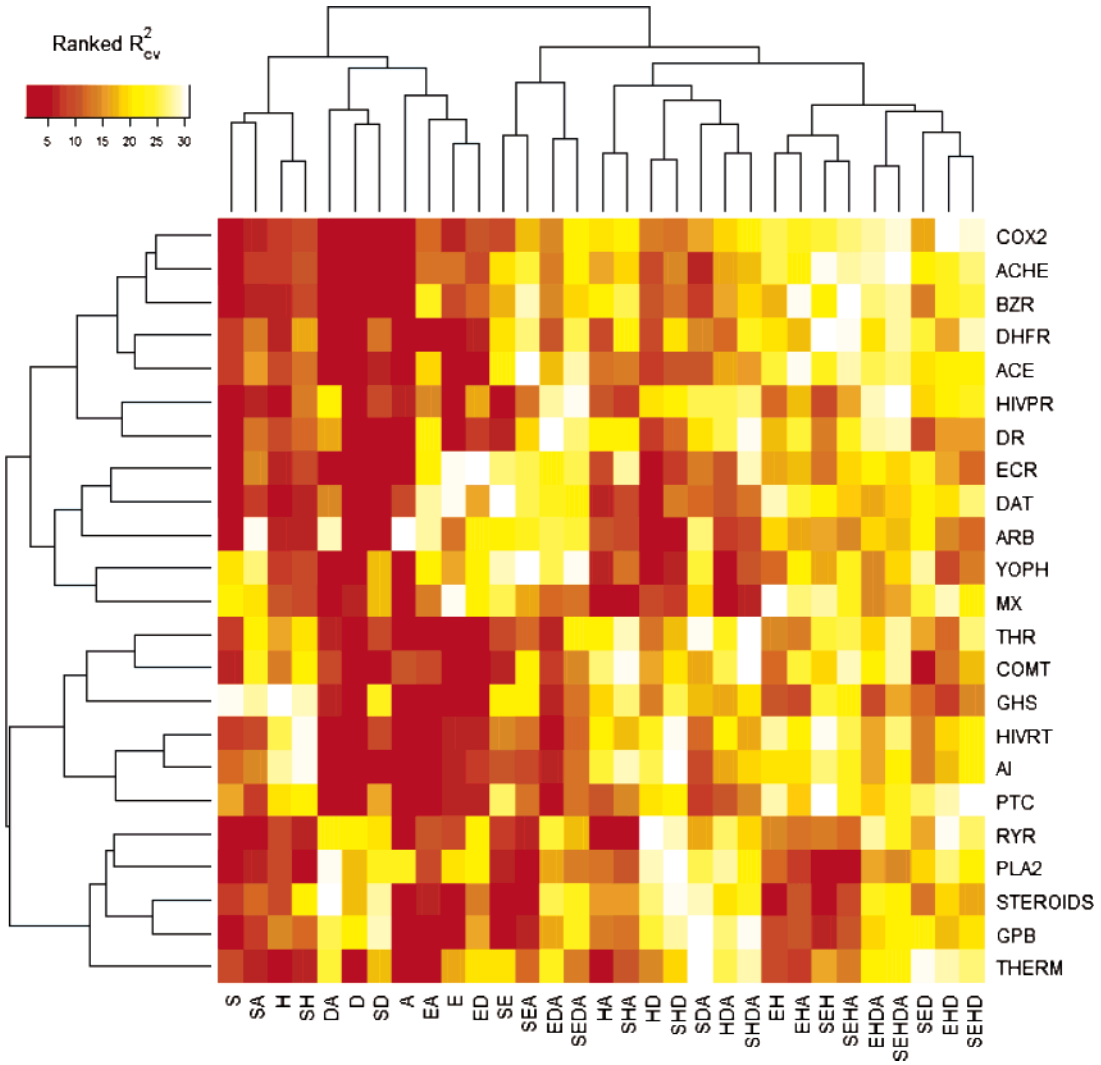

Figure 4. Heat map and hierarchical clustering of R_{cv}^2 rankings of each CoMSIA molecular field subset for each data set. The higher the R_{cv}^2 ranking (the brighter the color on the heat map), the better the CoMSIA molecular field subset is able to generate a predictive model for the respective data set.

of adding a molecular field to others is reduced because of a significant proportion of the information contained in a CoMSIA molecular field being redundant given the presence of other CoMSIA molecular fields (as demonstrated in Figure 2). Adding a molecular field greatly increases the number of explanatory variables in the model, and there was concern that this may reduce model predictivity through model overfitting. However, the results presented here indicate that in general the new information gained by including a different CoMSIA molecular field outweighs the model generation problems associated with increasing the number of explanatory variables.

The 23 data sets cluster on the basis of the utility of CoMSIA molecular field subsets to predict accurately (Figure 4). Here, the 23 data sets fell into four main clusters, each with a different CoMSIA molecular field preference profile. It will be interesting to see whether further clusters can be detected using a number of data sets. Additionally, it would be useful to be able to find molecular features of a data set which indicate which cluster a data set will fall into. This would allow the determination of the best set of CoMSIA molecular fields for the specific data set without having to try multiple molecular field subsets.

This study was undertaken in order to gain insight into how the predictivity of standard CoMSIA models can be maximized by the selection of CoMSIA molecular fields. The standard sampling grid and CoMSIA field parametrizations were used in this study, and hence, the results presented here may or may not be useful in nonstandard usage of CoMSIA. In addition to molecular field selection, the alignment of chemicals is a particularly important step in CoMSIA. To make the findings as generalizable as possible, bias toward a particular method for aligning chemicals was avoided by using the alignments of the original data set authors. It should also be noted that because of the great variation in how molecular fields can be calculated and sampled the extrapolation of the results presented here to other molecular field methodologies is not recommended without further study.

CONCLUSIONS

When CoMSIA models are based on a single CoMSIA molecular field, there are statistically significant differences in R_{cv}^2 between the models. The hydrophobic and electrostatic CoMSIA molecular fields resulted in the best prediction accuracy. Similarly, when CoMSIA models were developed

with all five CoMSIA molecular fields, the hydrophobic and electrostatic molecular fields had a statistically significant greater contribution than the H-bond and steric fields.

In general, the more CoMSIA molecular fields used, the better the R_{cv}^2. However, as more CoMSIA molecular fields are added, the increase in R_{cv}^2 diminishes because of redundancy between different CoMSIA molecular fields.

Data sets can be clustered into four groups on the basis of the utility of the CoMSIA molecular field sets, indicating that future studies focusing on identifying the differences between groups of data sets may be useful in choosing the best set of CoMSIA molecular fields to use.

ACKNOWLEDGMENT

We thank the authors who contributed their data sets for analysis in this study.

Supporting Information Available: Supporting Table 1 details the statistical significance of all pairwise comparisons between the R_{cv}^2 of models generated with different molecular field sets. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Spellmeyer, D. C.; Blaney, J. M.; Martin, E. In *Practical Application of Computer-Aided Drug Design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 165−193.

(2) Williams, M.; Gordon, E. M. In *A Textbook of Drug Design and Development*, 2nd ed.; Krogsgaard-Larsen, P., Liljefors, T., Madsen, U., Eds.; Harwood Academic Publishers: Sydney, Australia, 1996; pp 1−33.

(3) Giersiefen, H.; Hilgenfeld, R.; Hillisch, A. In *Modern Methods of Drug Discovery*; Hilgenfeld, R., Hillisch, A., Eds.; Birkhäuser Verlag: Basel, Switzerland, 2003; pp 1−18.

(4) Charifson, P. S.; Kuntz, I. D. In *Practical Application of Computer-Aided Drug Design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 1−37.

(5) Cramer, R. D., III; 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.

(6) Klebe, G.; Abraham, U.; Mietzner, T. Molecular Similarity Indexes in a Comparative-Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological-Activity. *J. Med. Chem.* **1994**, *37*, 4130−4146.

(7) Kim, K. H. In *Molecular Similarity in Drug Design*; Dean, P., Ed.; Blackie Academic & Professional: London, 1995; pp 291−331.

(8) Cramer, R. D. Topomer CoMFA: A Design Methodology for Rapid Lead Optimization. *J. Med. Chem.* **2003**, *46*, 374−388.

(9) 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*, 1525−1530.

(10) 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*, 807−816.

(11) 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*, 276−283.

(12) Depriest, S. A.; Mayer, D.; Naylor, C. B.; Marshall, G. R. 3D-QSAR of Angiotensin-Converting Enzyme and Thermolysin Inhibitors. A Comparison of CoMFA Models Based on Deduced and Experimentally Determined Active-Site Geometries. *J. Am. Chem. Soc.* **1993**, *115*, 5372−5384.

(13) Rahnasto, M.; Raunio, H.; Poso, A.; Wittekindt, C.; Juvonen, R. O. Quantitative Structure−Activity Relationship Analysis of Inhibitors of the Nicotine Metabolizing CYP2A6 Enzyme. *J. Med. Chem.* **2005**, *48*, 440−449.

(14) 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*, 5541−5554.

(15) 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*, 1162−1170.

(16) 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*, 567−582.

(17) Maddalena, D. J.; Johnston, G. A. Prediction of Receptor Properties and Binding Affinity of Ligands to Benzodiazepine/GABAA Receptors Using Artificial Neural Networks. *J. Med. Chem.* **1995**, *38*, 715−24.

(18) 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*, 971−981.

(19) 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*, 797−810.

(20) Chavatte, P.; Yous, S.; Marot, C.; Baurin, N.; Lesieur, D. Three-Dimensional Quantitative Structure−Activity Relationships of Cyclooxygenase-2 (COX-2) Inhibitors: A Comparative Molecular Field Analysis. *J. Med. Chem.* **2001**, *44*, 3223−30.

(21) 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 Alpha-Substituent of the Piperidine Ring Using Pharmacophore-Based Flexible Alignment. *J. Med. Chem.* **2004**, *47*, 6137−6143.

(22) Sutherland, J. J.; Weaver, D. F. Three-Dimensional Quantitative Structure−Activity and Structure−Selectivity Relationships of Dihydrofolate Reductase Inhibitors. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 309−331.

(23) Mattioni, B. E.; Jurs, P. C. Prediction of Dihydrofolate Reductase Inhibition and Selectivity Using Computational Neural Networks and Linear Discriminant Analysis. *J. Mol. Graphics Modell.* **2003**, *21*, 391−419.

(24) Melville, J. L.; Hirst, J. D. On the Stability of CoMFA Models. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1294−1300.

(25) Bostrom, J.; Bohm, M.; Gundertofte, K.; Klebe, G. A 3D QSAR Study on a Set of Dopamine D4 Receptor Antagonists. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1020−7.

(26) 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*, 1333−1340.

(27) Gohlke, H.; Klebe, G. DrugScore Meets CoMFA: Adaptation of Fields for Molecular Comparison (AFMoC) or How To Tailor Knowledge-Based Pair-Potentials to a Particular Protein. *J. Med. Chem.* **2002**, *45*, 4153−4170.

(28) Hannongbua, S.; Lawtrakul, L.; Sotriffer, C. A.; Rode, B. M. 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*, 389−394.

(29) Luco, J. M.; Ferretti, 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.

(30) *Tripos Bookshelf 7.1*; Tripos Inc.: St. Louis, Missouri, 2005.

(31) Schevitz, R. W.; Bach, N. J.; Carlson, D. G.; Chirgadze, N. Y.; Clawson, D. K.; Dillard, R. D.; Draheim, S. E.; Hartley, L. W.; Jones, N. D.; Mihelich, E. D.; Olkowski, J. L.; Snyder, D. W.; Sommers, C.; Wery, J. P. Structure-Based Design of the First Potent and Selective Inhibitor of Human Nonpancreatic Secretory Phospholipase-a(2). *Nat. Struct. Biol.* **1995**, *2*, 458−465.

(32) 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 (Moscow, Russ. Fed.)* **1994**, *33*, 6074−6085.

(33) Bohm, M.; Sturzebecher, J.; Klebe, G. Three-Dimensional Quantitative Structure−Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis to Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa. *J. Med. Chem.* **1999**, *42*, 458−477.

(34) Hu, X.; Stebbins, C. E. Molecular Docking and 3D-QSAR Studies of Yersinia Protein Tyrosine Phosphatase YopH Inhibitors. *Bioorg. Med. Chem.* **2005**, *13*, 1101−1109.

(35) Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three-Dimensional Structure Directed Quantitative Structure−Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally

Occurring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163−172.

(36) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579−586.

(37) *R Development Core Team, R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2005.

(38) Dixit, A.; Kashaw, S. K.; Gaur, S.; Saxena, A. K. Development of CoMFA, Advance CoMFA and CoMSIA Models in Pyrroloquinazolines as Thrombin Receptor Antagonist. *Bioorg. Med. Chem.* **2004**, *12*, 3591−3598.

(39) Prathipati, P.; Pandey, G.; Saxena, A. K. CoMFA and Docking Studies on Glycogen Phosphorylase A Inhibitors as Antidiabetic Agents. *J. Chem. Inf. Model.* **2005**, *45*, 136−145.

(40) Chakraborti, A. K.; Gopalakrishnan, B.; Sobhia, M. E.; Malde, A. 3D-QSAR Studies of Indole Derivatives as Phosphodiesterase IV Inhibitors. *Eur. J. Med. Chem.* **2003**, *38*, 975−982.

CI600214B