3D-QSAR of histone deacetylase inhibitors: hydroxamate analogues†

Dhanshri C. Juvale,[a] Vishal V. Kulkarni,[b] Hemantkumar S. Deokar,[a] Nilesh K. Wagh,[a] Subhash B. Padhye[b] and Vithal M. Kulkarni*[a]

The histone deacetylase enzyme has increasingly become an attractive target for developing novel anticancer drugs. Hydroxamates are a new class of anticancer agents reported to act by selective inhibition of the histone deacetylase (HDAC) enzyme. Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were employed to study three-dimensional quantitative structure–activity relationships (3D-QSARs). QSAR models were derived from a training set of 40 molecules. An external test set consisting of 17 molecules was used to validate the CoMFA and CoMSIA models. All molecules were superimposed on the template structure by atom-based, multifit and the SYBYL QSAR rigid body field fit alignments. The statistical quality of the QSAR models was assessed using the parameters r^2_{conv}, r^2_{cv} and r^2_{pred}. In addition to steric and electronic fields, ClogP was also taken as descriptor to account for lipophilicity. The resulting models exhibited a good conventional r^2_{conv} and cross-validated r^2_{cv} values up to 0.910 and 0.502 for CoMFA and 0.987 and 0.534 for CoMSIA. Robust cross-validation by 2 groups was performed 25 times to eliminate chance correlation. The CoMFA models exhibited good external predictivity as compared to that of CoMSIA models. These 3D-QSAR models are very useful for design of novel HDAC inhibitors.

Introduction

Histone deacetylase (HDAC) inhibitors have gained considerable interest due to their ability to modulate transcriptional activity.[1] HDAC-mediated transcriptional activity represents a common molecular mechanism of alteration in chromin structure and blockage of normal cell differentiation. As a result, this class of inhibitors can block angiogenesis and cell cycling, and promote apoptosis and cell differentiation. HDAC has become a novel target for the discovery of drugs for the treatment of cancer and other diseases.[2–7] The number of HDAC enzyme subtypes has expanded considerably over the past few years, offering opportunities for the development of HDAC inhibitors with improved specificity.

A number of natural inhibitors such as trichostatin A (TSA),[8] cyclic tetrapeptide trapoxin (TPX),[9] HC toxin[10] and apicidin[11] have been so far reported. Among them, TSA has been identified as a potent and specific HDAC inhibitor. Synthetic inhibitors like sodium phenyl butyrate,[12] sodium valproate,[13] suberanilo hydroxamic acid (SAHA),[14] straight chain TSA and SAHA like analogues[15–17] and oxamflatin[18] have been reported. TSA and its analogues are considered to be mimics of the histone acetyl lysine side chain. Crystal structures of histone deacetylase-like protein (HDLP) with TSA and SAHA revealed that hydroxamic acid-based HDAC inhibitors bind to the deacetylase core by inserting their aliphatic chains into the HDLP pocket. Their hydroxamic acid group reaches the polar bottom of the pocket, where it coordinates with the zinc ion.[19]

HDAC inhibitors typically possess a metal-binding group, a hydrophobic cap functionality that interacts with the amino acid residues at the entrance of the *N*-acetyl lysine binding channel and an aliphatic spacer connecting the cap and the metal binding group. The factors contributing to the biological activity can be analyzed through the use of different physicochemical descriptors in the generation of quantitative structure–activity relationship (QSAR) models. Due to the flexibility of the spacer group between the metal binding and cap groups, it is difficult to choose a suitable conformation to achieve a meaningful superimposition. Only a few QSAR studies have been reported until now.[20–23] Wang *et al.*[20] reported QSAR models on TSA- and SAHA-like hydroxamic acid and suggested that the shape and area of molecules are important for their biological activity. Similarly Xie *et al.*[23] reported a QSAR study on a data set of 124 compounds which showed that the van der Waals surface area and hydrophobicity are important parameters required for the biological activity.

In order to gain further insight into the structural requirements of HDAC inhibitors, we have performed a three-dimensional quantitative structure–activity relationship (3D-QSAR) study using comparative molecular field analysis (CoMFA)[24] and comparative molecular similarity indices analysis (CoMSIA).[25] In CoMFA, it is assumed that the interaction between an inhibitor and its molecular target is preliminarily noncovalent and shape-dependent in nature. A QSAR can be derived correlating the differences in steric and electrostatic fields surrounding a set of molecules to the biological activity. This method can be used to develop a 3D pharmacophore model[26] that describes the structure–activity relationship (SAR). The CoMSIA method of 3D-QSAR was introduced by Klebe[25] in 1994, in which a common probe atom and similarity indices are calculated at regularly spaced grid points for prealigned molecules. CoMSIA considers five different fields: electronic, steric, hydrophobic field and hydrogen

[a]Poona College of Pharmacy, Bharati Vidyapeeth Deemed University, Erandwane, Pune, 411038, India. E-mail: vivivips5@gmail.com; Fax: +91-20-25439383; Tel: +91-20-25437237
[b]Department of Chemistry, University of Pune, Pune, 411008, India
† VMK dedicates this manuscript to Shridevi V. Kulkarni.

bond donor and acceptor fields, and is less alignment-sensitive than CoMFA. CoMFA and CoMSIA have been widely applied in drug design.[27–29]

Results and discussion

The 3D-QSAR models for hydroxamic acid analogues were derived using CoMFA and CoMSIA techniques. The negative logarithm of IC_{50} (pIC_{50}) was used as the biological activity in the 3D-QSAR study (Table 1). Conformation of the molecules used in the study was obtained by a systematic search and the lowest energy conformer was selected and minimized using Powell method to rms 0.001 kcal mol^{-1} Å^{-1}.

Alignment of the molecules was carried out using three techniques, namely RMS fitting (atom-based), multifit (flexible fitting) and SYBYL QSAR rigid body field fitting. The most active molecule 17 was used as a template molecule for alignment (Fig. 1).
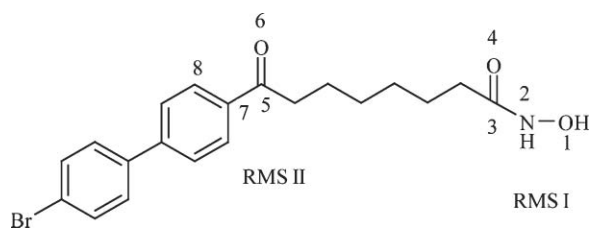
Fig. 1 Molecule 17 with atoms used for superimposition are marked. Atoms 1–4 were used for RMS I alignment and atoms 5–8 were used for RMS II.

CoMFA

CoMFA models were generated using a training set of 40 molecules (Table 1), with a column filtering value (σ min) of 2.0. A test set of 17 molecules (Table 1) was used to check the external predictivity of the models.

A preliminary study was performed on the atom-based alignment to study importance of each field individually. The cross-validated r_{cv}^2 value from the electrostatic field only was higher than that of the steric field only analysis. All further analyses were performed with steric and electrostatic fields calculated at each grid point simultaneously.

The atom-based alignment RMS I gave 0.405 with four components, a conventional r^2 (r_{conv}^2) of 0.934, a predictive r^2 (r_{pred}^2) of 0.210 and an F value of 122.995. The alignment of the molecules using atom-based selection RMS II shows good internal predictivity with an r_{cv}^2 of 0.514. However, the model exhibited rather a poor external predictivity with an r_{cv}^2 of 0.132. This is because the alignment did not have –CONHOH as template for superimposition.

CoMFA models generated for multifit alignment, MF I showed r_{cv}^2 of 0.243 with one component, r_{conv}^2 of 0.571, r_{pred}^2 of −0.018, F value of 50.610. Realignment of the molecules by field fit (FF I) with respect to the fields of template molecule (molecule 17) yielded r_{cv}^2 of 0.478 with seven components, r_{conv}^2 of 0.987, r_{pred}^2 of 0.327, F value of 174.687. Models generated for multifit (MF II) and field fit (FF II) alignments using RMS II data set had poor external predictivity (Table 2).

The activity data used in this study may have contributions from other factors than just steric and electrostatic interactions. The ClogP, the calculated logarithm of partition coefficient, was calculated and added to the CoMFA table. Inclusion of ClogP in the CoMFA model with field fit alignment increased internal ($r_{cv}^2 = 0.502$) as well as external predictivity ($r_{pred}^2 = 0.633$). We performed all further studies with ClogP in addition to CoMFA fields.

The model generated with FF I alignment (Fig. 2) with a good internal predictive ability ($r_{cv}^2 = 0.502$) and a small standard error of estimatation (SEE = 0.260) was selected as the best model to explain SAR and to carry out further analysis. Results obtained from the RMS I, MF I and FF I alignments with ClogP as additional descriptors are shown in Table 3. Observed and predicted biological activities of the training and test sets are plotted in Fig. 3 and 4, respectively. To further assess the robustness and statistical confidence of the derived 3D-QSAR model, bootstrapping analysis was performed and average of 100 runs is 0.919 (r_{bs}^2). Cross-validation with 2 groups was performed to ascertain the true predictivity of the model and repeated 25 times; the mean r_{cv}^2 was 0.299. A negative value of r^2 in a randomized biological activity test revealed that the results were not based on chance correlations. The results of these cross-validation tests are shown in Table 5.
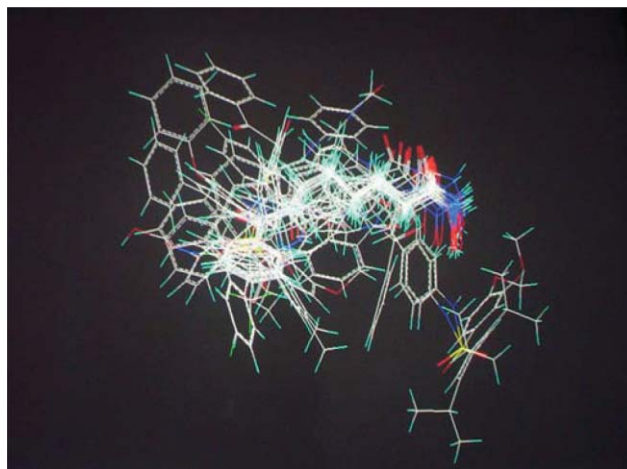

Fig. 2 Superimposition of all molecules using FF I.
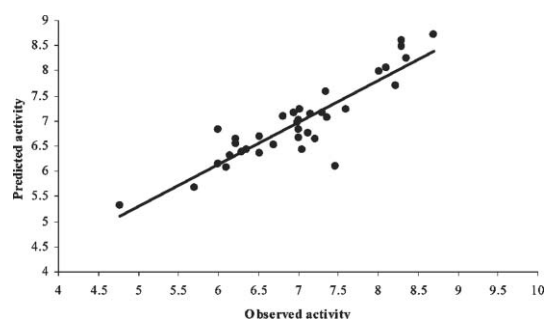

Fig. 3 Graph of observed versus predicted activities of the training set from FF I CoMFA analysis.

The results of 3D-QSAR using CoMFA, are represented as a "coefficient contour" map. The contour maps obtained from the field fit model are used to explain the SAR of molecules in the

Table 1 Structures and biological activities of molecules used in the present study
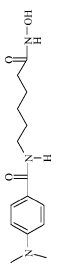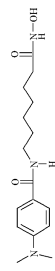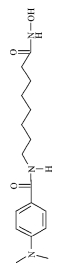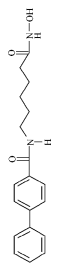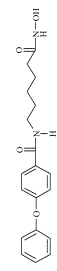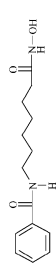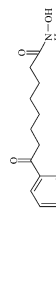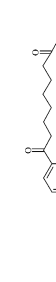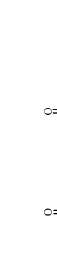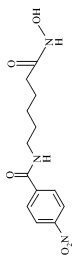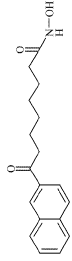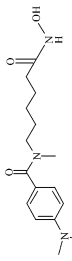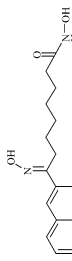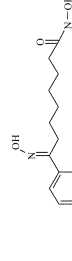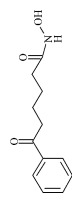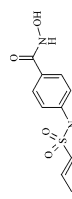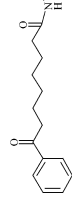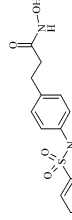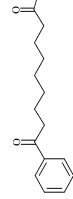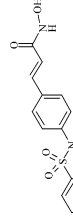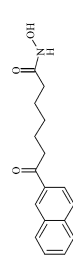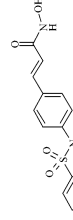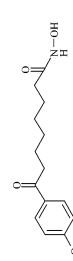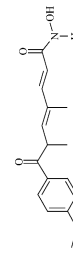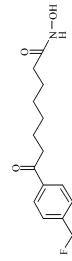
Structure number	Structure	$pIC_{50}{}^a$	Predicted activity by CoMFA[b]	Predicted activity by CoMSIA[b]	Structure number	Structure	$pIC_{50}{}^a$	Predicted activity by CoMFA[b]	Predicted activity by CoMSIA[b]
01		7.30	7.37	7.26	12		7.30	7.16	7.33
02		8.00	8.08	8.00	13		7.02	7.23	6.96
03		7.54	7.07	7.51	14		7.35	7.58	7.34
04		7.28	7.18	7.31	15		8.30	8.49	8.42
05		7.36	7.06	7.26	16		8.10	8.06	8.09
06		6.95	7.15	6.95	17		8.70	8.71	8.55
07		6.99	6.98	6.99	18		8.02	7.99	8.01
08		7.15	7.14	7.13	19		8.35	8.24	8.34
09		6.30	6.39	6.34	20		8.22	7.70	8.18
10		6.35	6.421	6.32	21		7.6	7.23	7.7
11		6.82	7.088	6.79	22		7.46	6.09	7.62

Table 1 (*Contd.*)

Structure number	Structure	pIC_{50} [a]	Predicted activity by CoMFA [b]	Predicted activity by CoMSIA [b]
23		6.00	6.84	5.99
24		6.00	6.14	5.99
25		6.70	6.53	6.82
26		5.70	5.67	5.62
27		4.77	5.31	4.83
28		7.12	6.76	6.89
29		7.00	6.84	7.00
30		6.52	6.70	6.66
31		6.15	6.30	6.01

Structure number	Structure	pIC_{50} [a]	Predicted activity by CoMFA [b]	Predicted activity by CoMSIA [b]
32		6.10	6.07	6.09
33		6.22	6.54	6.32
34		6.52	6.35	6.77
35		7.00	6.67	6.76
36		7.00	7.03	7.07
37		6.22	6.64	6.19
38		7.22	6.64	7.13
39		7.05	6.44	7.15
40		7.13	7.172	7.17

This journal is © The Royal Society of Chemistry 2006

Org. Biomol. Chem., 2006, **4**, 2858–2868 | 2861

Table 1 (Contd.)

Structure number	Structure	pIC$_{50}$[a]	Predicted activity by CoMFA[b]	Predicted activity by CoMSIA[b]
41		6.52	6.731	6.49
42		7.05	6.907	7.24
43		7.52	6.66	7.21
44		5.82	6.8	6.52
45		7.19	7.53	7.97
46		6.87	7.26	7.54
47		7.46	7.38	7.48
48		7.82	7.66	8.00
49		7.35	7.94	8.09
50		8.30	7.94	8.10
51		8.07	7.70	7.39
52		8.40	7.71	7.23
53		6.05	6.40	6.66
54		7.00	6.80	6.94
55		6.22	6.78	6.05
56		7.00	6.72	6.45
57	TSA	8.30	8.60	8.33

[a] Observed activity, normalized pIC$_{50}$ values. [b] Predicted activities from field fit alignment, molecules 40–56 used as test set.

Table 2 Summary of CoMFA results with steric and electrostatic fields

	Alignments					
	RMS I[a]	FF I[b]	MF I[c]	RMS II[a]	FF II[b]	MF II[c]
r^2_{cv} [a]	0.405	0.478	0.243	0.514	0.484	0.326
Components	4	7	1	4	5	5
SEP	0.675	0.602	0.756	0.613	0.628	0.731
r^2_{conv}	0.934	0.987	0.571	0.961	0.988	0.904
SEE	0.227	0.155	0.553	0.176	0.099	0.123
F value	122.995	174.687	50.610	168.022	551.325	123.24
Steric contribution	0.440	0.408	0.449	0.415	0.374	0.409
Electrostatic	0.560	0.592	0.501	0.585	0.626	0.591
r^2_{pred}	0.210	0.327	−0.018	0.132	0.035	−0.120
r^2_{bs}	0.959	0.990	0.676	0.975	0.994	0.945

[a] Alignment by RMS fit. [b] Alignment by field fit. [c] Alignment by multifit. [d] A cross-validated r^2 value was obtained from the 'leave-one-out' method.

Table 3 Summary of CoMFA results with the additional descriptor ClogP

	RMS I[a]	FF I[b]	MF I[c]
r^2_{cv} [d]	0.464	0.502	0.252
Components	3	3	2
SEP	0.641	0.609	0.752
r^2_{conv}	0.874	0.910	0.649
SEE	0.308	0.260	0507
F value	83.441	121.227	34.152
Steric contribution	0.409	0.372	0.440
Electrostatic	0.481	0.533	0.456
ClogP	0.110	0.096	0.104
r^2_{pred}	0.521	0.633	0.258
r^2_{bs}	0.890	0.919	0.732

[a] Alignment by RMS fit. [b] Alignment by field fit. [c] Alignment by multifit. [d] A cross-validated r^2 value was obtained from the 'leave-one-out' method.

Fig. 4 Graph of observed *versus* predicted activities of the test set from FF I CoMFA analysis.

Fig. 6 displays the electrostatic contour plot. The blue contours describe regions where positively charged groups enhance activity (80% contribution), and red contours describe regions where negatively charged groups enhance the activity (20% contribution).

CoMSIA

The CoMSIA analysis was performed using steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor descriptors. Only a few combinations of descriptors were used, which are complimentary to previously generated CoMFA models. Only FF I alignment was used for CoMSIA analysis. All the results of CoMSIA are shown in Table 4. CoMSIA models show lower

present study. The CoMFA contour maps are shown in Fig. 5 and 6. The field values were calculated at each grid point as the scalar product of the associated QSAR coefficient and the standard deviation of all the values in the corresponding column of the data table (STDDEV*COEFF) and are plotted as a percentage contribution to the QSAR equation.

Fig. 5 displays the steric contour plot. The green contours describe regions where sterically favorable groups enhance activity (80% contribution), and yellow contours describe regions of unfavorable steric effects (20% contribution).

Table 4 Results of CoMSIA analysis

	SE[a]	SEH[a]	SEA[a]	SEHA[a]	SEHAD[a]
r^2_{cv} [b]	0.297	0.343	0.419	0.441	0.534
Components	7	5	8	5	7
r^2_{conv}	0.984	0.976	0.991	0.975	0.987
SEE	0.118	0.138	0.091	0.140	0.109
F value	274.093	276.292	404.253	270.407	350.063
Steric contribution	0.406	0.280	0.316	0.218	0.179
Electrostatic	0.594	0.415	0.430	0.311	0.272
Hydrophobic	0.000	0.304	0.000	0.224	0.196
Hydrogen bond acceptor	0.000	0.000	0.254	0.246	0.197
Hydrogen bond donor	0.000	0.000	0.000	0.000	0.155
r^2_{pred}	0.054	0.286	0.304	0.389	0.464
r^2_{bs}	0.990	0.981	0.993	0.989	0.989

[a] S = steric; E = electroststic; H = hydrophobic, A = hHydrogen bond acceptor, D = hydrogen bond donor. [b] The same as in Table 2.
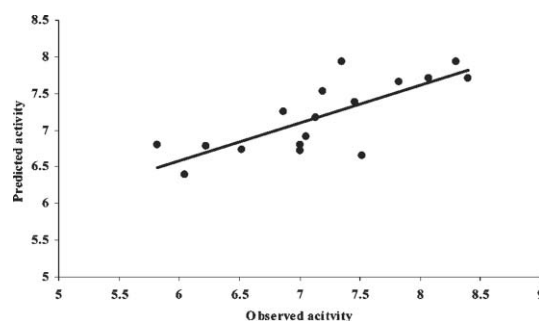
Table 5 Results of analysis with cross validation for 2 groups and randomized biological activities

| | r_{cv}^2 for 2 groups[a] | | Randomized r^2 [b] | |
	CoMFA[c]	CoMSIA[d]	CoMFA[c]	CoMSIA[d]
Mean	0.299	0.275	−0.359	−0.240
SD	0.117	0.085	0.023	0.064
High	0.438	0.424	−0.082	−0.063
Low	0.064	0.116	−0.463	−0.381

[a] Cross-validated r^2 for 2 groups with optimum number of components, average of 25 runs. [b] Cross-validated r^2 with randomized biological activity, average of 25 runs [c] CoMFA model generated by field fit. [d] CoMSIA analysis by combined steric, electronic, hydrophobic and hydrogen bond donor and acceptor fields.
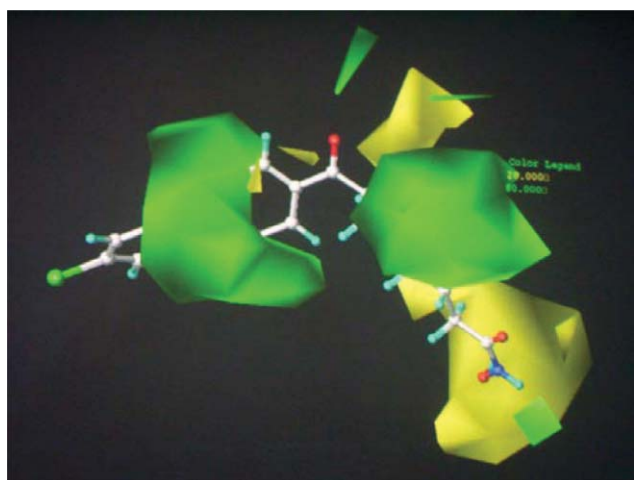

Fig. 5 CoMFA steric STDEV*COEFF contour plots from the field fit. Sterically favored areas are represented by green polyhedra. Sterically disfavorable are represented by yellow polyhedra. The active molecule 17 is shown in ball-and-stick representation.
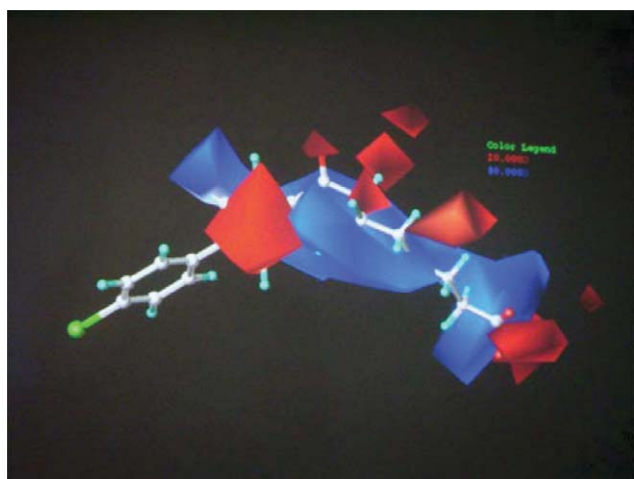
Fig. 6 CoMFA electrostatic STDEV*COEFF contour plots from the field fit. Positive-charge favored areas are represented by blue polyhedra. Negative-charge favored areas are represented by red polyhedra. The active molecule 17 is shown in ball-and-stick representation.

predictive properties than those of CoMFA models. In all models the electronic field is the common factor indicating the importance

of electrostatic interactions for the present series of molecules. The model with steric, electronic and hydrogen bond acceptor descriptors has a good predictivity (r_{pred}^2) of 0.304. Addition of a hydrophobic descriptor to this model caused an increase in the r_{cv}^2 (0.441). Combination of all fields gave a CoMSIA model with proper balance of all statistical terms. The models showed higher r_{cv}^2 (0.534) and a considerable r_{pred}^2 (0.464) values. The model, characterized by a good r_{conv}^2 (0.987) and a lower SEE (0.109), was selected as the best model to generate contour maps and explain the SAR. To further assess and validate the derived 3D-QSAR model, bootstrapping analysis was performed and average result of 100 runs is 0.989 (r_{bs}^2). To ascertain the true predictivity of the model, cross-validation with 2 groups was performed 25 times and the mean r_{cv}^2 was 0.275. A negative value of r^2 in a randomized biological activity test revealed that the results were not based on chance correlations. The results of CoMSIA are summarized in Table 4 and the observed *versus* predicted biological activities of the training and test sets are plotted in Fig. 7 and 8.
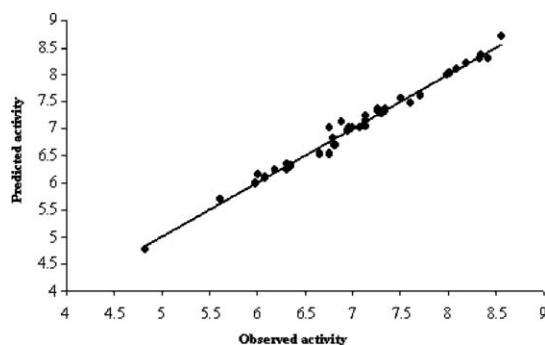
Fig. 7 Graph of observed *versus* predicted activities of the training set from FF I CoMSIA analysis.
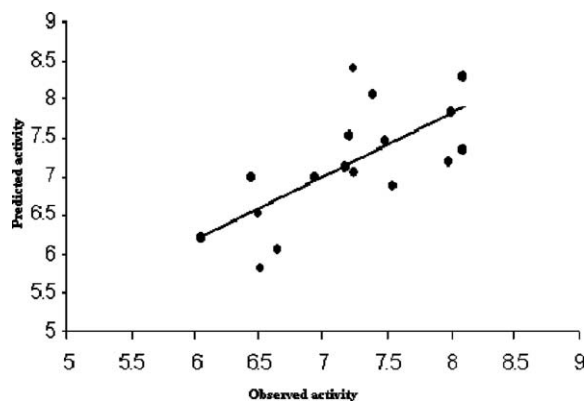
Fig. 8 Graph of observed *versus* predicted activities of the test set from FF I CoMSIA analysis.

The steric, electrostatic, hydrophobic and hydrogen bond donor and acceptor contours of CoMSIA are shown in Fig. 9–13, respectively. The steric fields (green, more steric bulk favored; yellow, steric bulk disfavored), electrostatic fields (blue, positive charge favored; red, negative charge favored), hydrophobic fields (yellow, hydrophobic favored; white, hydrophobic disfavored), hydrogen bond acceptor fields (magenta, favored; red,

2864 | *Org. Biomol. Chem.*, 2006, **4**, 2858–2868

This journal is © The Royal Society of Chemistry 2006

Fig. 9 CoMSIA steric fields. Yellow indicates sterically unfavorable region; green indicates a sterically favorable region. The active molecule 17 is shown in a ball-and-stick representation.

Fig. 11 CoMSIA hydrophobic fields. Yellow indicates regions where hydrophobic substituents enhance activity; white indicates hydrophobic substituents reduce activity. The active molecule 17 is shown in a ball-and-stick representation.

Fig. 10 CoMSIA electrostatic fields. Positive-charge favored areas are represented by blue polyhedra. Negative-charge favored areas are represented by red polyhedra. The active molecule 17 is shown in a ball-and-stick representation.


Fig. 12 CoMSIA hydrogen bond acceptor fields. Magenta indicates regions where hydrogen bond acceptor substituents enhance activity; red indicates hydrogen bond acceptor substituents reduce activity. The active molecule 17 is shown in a ball-and-stick representation.

disfavored) and hydrogen bond donor fields (cyan, favored; purple, disfavored).

Interpretation of QSAR models

The CoMFA and CoMSIA analyses were performed on a series of HDAC inhibitors. The conformations of the molecules were generated from a systematic search of all the rotatable bonds with a uniform increment. The rotation of the spacer produces many low energy conformations. The lowest energy conformation of the all the molecules was used in the study.

CoMFA

Alignment of the molecules is important for CoMFA studies. In the present study we have aligned these ligands onto the template structure (molecule 17) using two different strategies.

The PLS analysis on RMS I model was performed using a threshold column filtering value of 2.0 kcal mol^{-1}. The results of different alignments are reported in Table 2. The analysis showed that electrostatic fields play a major role in binding to the HDAC active site. The CoMFA models were validated by predicting the activity of the external test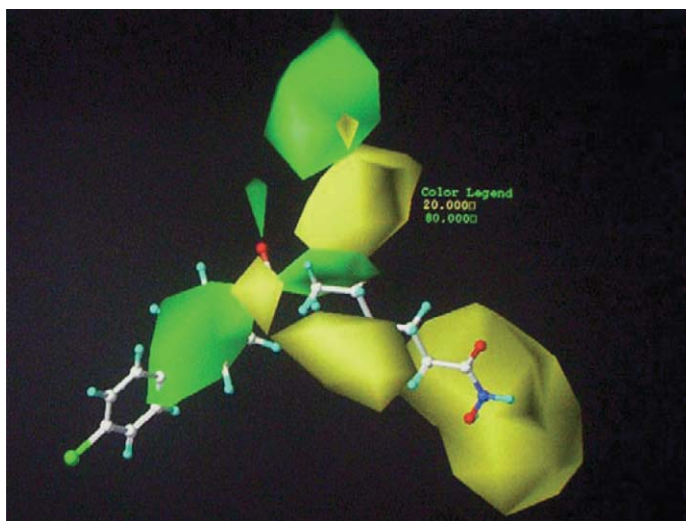 set. The results obtained show that RMS I alignment produces a statistically significant model. RMS II has improved internal predictivity but the external predictivity of the model was reduced. These analyses indicate that the hydroxamic acid group is very important for alignment. The hydroxamic acid is important for enzyme inhibitory activity, as it coordinates the zinc ion through CO and OH groups. It also forms hydrogen bonds between its NH and OH groups and the two charge relay systems His131/Asp166 and His132/Asp173 and between its CO and the Tyr297 hydroxyl group.[19]

Multifit alignment decreased the predictivity whereas field fit alignment showed improved predictions in both cases. Slight

This journal is © The Royal Society of Chemistry 2006

Org. Biomol. Chem., 2006, **4**, 2858–2868 | 2865

Fig. 13 CoMSIA hydrogen bond donor fields. Cyan indicates regions where hydrogen bond donor substituents enhance activity; purple indicates hydrogen bond donor substituents reduce activity. The active molecule 17, is shown in a ball-and-stick representation.

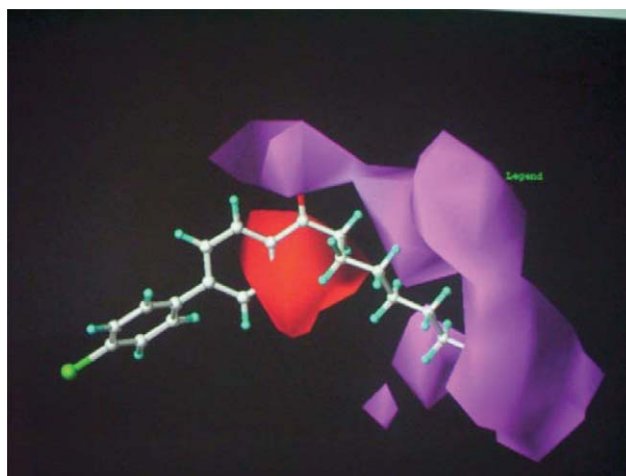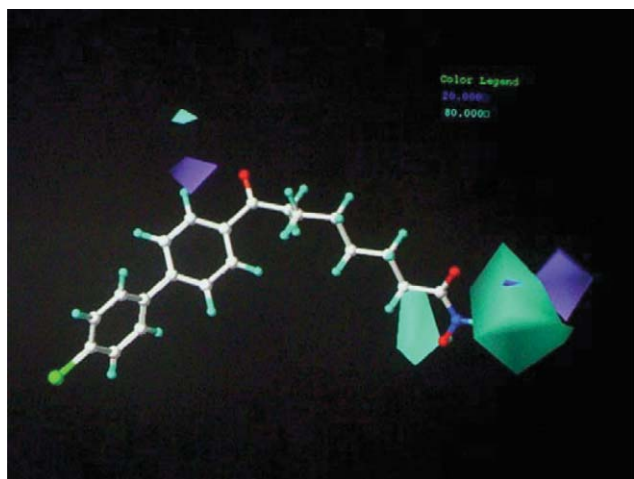variation in the alignment rules leads to dramatic differences in the external predictions. The good internal and external predictions with FF I alignment support the use of these atoms for superimposition.

An additional descriptor was added to the CoMFA table to study the influence of other factors on the CoMFA results. Inclusion of ClogP in the CoMFA table improved the statistical results of the FF I model. Hydrophobicity is important for HDAC inhibition.

PLS analysis on FF I was also performed using the 2 group cross-validation procedure (Table 5). The r_{cv}^2 values observed during 25 cross-validation runs were less than those observed from 'leave one out'. In no cases were these values were negative, and they showed good internal consistency. The field fit alignment FF I was used to analyze the CoMFA contour maps as this model exhibited good internal as well as external predictivity. The CoMFA steric and electrostatic maps are shown in Fig. 5 and 6, respectively.

Fig. 5 depicts the steric contour plot. The lipophilic fragment of the molecule was surrounded by the sterically favorable green contours. The most-active molecule 17 has a phenyl ring embedded in this green region. Other sterically-favorable green contours are observed near the spacer chain. This green contour is surrounded by the unfavorable yellow region. A substitution on the spacer atoms adjacent to the phenyl ring (as that of TSA) is favored, but any substitution on the carbon next to the hydroxamic acid functional group is unfavorable. This also suggests that the orientation of the bulky groups at these positions is important for activity.

Fig. 6 displays the electrostatic contour plots. A red contour was found near the phenyl ring of compound 17, suggesting that a high electron density in this region increases the activity. A large negative-charge unfavorable blue contour was found to surround the spacer chain. This indicates that substitutions in this region with high electron density reduce activity and emphasizes the necessity of positively charged groups; hydroxamate is essential for HDAC activity.

CoMSIA

CoMSIA analysis was performed for field fit alignment as it gave the best model in CoMFA. All the CoMSIA fields were considered for analysis. A combination of steric, electrostatic, hydrophobic and hydrogen bond donor and acceptor fields gave the best QSAR model with good internal as well as external predictivity. The model was further validated using 2 groups method and randomization tests. Fig. 9–13 show the CoMSIA contour maps with the most active molecule.

The steric contour maps of CoMSIA (Fig. 9) are also similar to CoMFA steric maps. The electrostatic contour maps are shown in Fig. 10. CoMSIA shows blue contours over the hydroxamic acid group, which signifies the importance of this group. The red polyhedra over the phenyl ring indicate that the presence of electron-rich functional groups at this position increase the activity. Analysis of CoMSIA hydrophobic maps (Fig. 11) indicates that a lipophilic-favorable yellow region is found near the aromatic ring. This indicates that lipophilicity of the cap portion of the molecule is important for activity. A hydrophilic-favorable white contour is observed surrounding the hydroxamate functional group. In the present study, hydrogen bond acceptor fields (Fig. 12) provide further support for the role of the positively-charged hydroxamate group. Hydrogen bond acceptor-unfavorable red polyhedra observed at the phenyl ring indicate that the presence of electron-rich groups improves activity. Hydrogen bond donor maps (Fig. 13) show hydrogen bond donor-favorable cyan contours near the hydroxyl group of hydroxamic acids.

The role of both the steric and electrostatic contribution can be clearly explained by analyzing the molecules 25, 26 and 28. Molecule 25, due to the double bond in its spacer chain, loses its orientation towards the sterically favorable region. Molecule 26 shows lesser biological activity, due to an additional methyl group adjacent to the hydroxamic acid group, which is also oriented towards the sterically unfavourable region. Molecule 28, with an electron-withdrawing chloro group substituted on the *para* position of the side chain, enhances the potency of the molecule, due to the proper orientation towards steric- and electrostatic-favorable regions. Molecules 32, 33 and 37 are less active. The bulky phenyl ring substituents of these molecules are oriented towards the sterically unfavourable yellow region. Molecule 15 has comparable biological activity to that of the most active molecule 17, because of the proper steric and electrostatic interactions of its hydrophobic phenyl groups.

Experimental

Biological data

A data set consisting of 57 hydroxamic acid analogues was taken from the literature.[15–17] The structure of the compounds and their biological data are given in Table 1. In this QSAR study, the biological activity of each compound has been expressed as the negative logarithm of normalized IC_{50} (pIC_{50}). These normalized IC_{50} values, and the negative logarithm of normalized IC_{50} (pIC_{50}) of compounds were taken from the literature and used in the present study.[23] Thus, the data correlated linearly to the free energy change. A training set of 40 molecules (Table 1) was used for

the generation of QSAR models. The training set molecules were selected in such a way that they contains information in terms of both their structural features and biological activity ranges. The most active molecules were included, so that they can provide critical information on pharmacophore requirements. Several moderately active and inactive molecules were also included, to spread out the range of activities. A test set of 17 molecules (40–56, Table 1) was used to access the predictive ability of the generated models. The test molecules represent a range of biological activity similar to the training set.

Computational details

All computational studies were performed using SYBYL 6.9.1[30] with a standard Tripos force field.[31] The compounds were constructed from the fragments in the SYBYL database with standard bond lengths and bond angles. Geometry optimization was carried out using the standard Tripos forcefield with distance dependent-dielectric function and energy gradient of 0.001 kcal mol^{-1} Å^{-1}. The initial conformations were obtained from a systematic search. The lowest energy conformers were selected and minimized using the Powell method till root-mean-square (rms) deviation 0.001 kcal mol^{-1} Å^{-1} was achieved. Partial atomic charges required for calculation of the electrostatic interaction were computed by a semiempirical molecular orbital method using AM1 in the MOPAC program.

Alignment rules.

The "alignment rule", *i.e.*, the positioning of a molecular model within the fixed lattice, is by far the most important input variable in CoMFA, since the relative interaction energies depend strongly on relative molecular positions. The most active molecule 17 was used as a template for aligning the other molecules.

In the present study, we have superimposed molecules by three alignment rules: (1) atom-based alignment, (2) multifit alignment, (3) field fit alignment.

Alignment (1) was done by atom-based fitting of the atoms to the most active molecule, 17. (a) The hydroxamic acid group of the molecules was used for rms fitting (RMS I) and (b) carbonyl group and phenyl ring atoms for RMS II, both as shown in Fig. 1.

Alignment (2) of the molecules was carried out by flexible fitting (multifit) of atoms, of the molecules to the template molecule 17. This involved energy calculations and fitting onto the template molecule by applying force (force constant 20 kcal mol^{-1}) and subsequent energy minimization.

Alignment (3) was carried out using the SYBYL QSAR rigid body field fit command within SYBYL and using compound 17 as template molecule. The superimposition of all the molecules is shown in Fig. 2.

CoMFA and CoMSIA interaction energies.

For each alignment, the steric and electrostatic potential fields for CoMFA were calculated at each lattice intersection of a regularly spaced grid box. The lattice spacing was set to a value of 2.0 Å in all X, Y and Z directions. The van der Waals potential (Lennard-Jones, 6–12) and the columbic term, which represent, respectively, steric and electrostatic fields, were calculated using the Tripos force field. A distance-dependent dielectric constant of 1.0 was used. A

sp^3 carbon atom with van der Waals radius of 1.52 Å and + 1.0 charge was served as the probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to ±30 kcal mol^{-1}. The electrostatic contributions were ignored at lattice intersections with maximum steric interactions.

CoMSIA calculates similarity indices at the intersections of the surrounding lattice. Five physicochemical properties steric, electrostatic, hydrophobic, hydrogen bond donor and acceptor were evaluated, using a common probe atom with 1 Å radius and charge, hydrophobicity and hydrogen bond property of +1.0. The attenuation factor was set to default vale, 0.3.

Calculation of ClogP

The ClogP values for 57 molecules were calculated using ClogP/CMR application within Sybyl 6.9.1. These methods are developed by the Biobyte Corporation.

Partial least square (PLS) analysis

The CoMFA and CoMSIA descriptors were used as independent variables and pIC$_{50}$ as dependent variables in the PLS[32] regression analysis to derive the 3D-QSAR models using the standard implementation in the Sybyl package. Initially, PLS was carried out in conjugation with the 'leave-one-out' (LOO)[33] option to determine the optimum number of components.

The results from cross-validation analysis were expressed as the cross-validated r^2 value (r^2_{cv}), which is defined as:

$$r^2_{cv} = 1 - \frac{\text{PRESS}}{\sum (Y - Y_{\text{mean}})^2},$$

where $\text{PRESS} = \Sigma (Y - Y_{\text{pred}})^2$.

The number of components that result in the highest r^2_{cv} and lowest standard error of predictions (SEP) were taken as the optimum. Equal weights were assigned to steric and electrostatic fields using CoMFA_STD scaling option. To speed up the analysis and reduce the noise, a minimum filter value "σ" of 2.0 kcal mol^{-1} was used. The LOO method of cross-validation is rather obsolete and it generally gives high r^2 value. Final analysis was performed to calculate the r^2_{conv} with a number of cross-validation groups set to zero using the optimum number of components. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis (100 runs) was performed. The statistical results obtained for CoMFA analysis are shown in Table 2 and 3.

To perform a more rigorous statistical test, cross-validation using 2 groups was carried out for the field fit analysis of CoMFA and CoMSIA. In this case, the data set is randomly divided into two groups, and the activity of the compounds from one group is predicted using the model from the other group. The process of group cross-validation was performed 25 times. The final r^2_{cv} value was calculated by taking the mean of 25 runs. These r^2_{cv} values were compared with r^2_{cv} obtained from LOO for each PLS analysis. The statistical results obtained from cross-validation with 2 groups for CoMFA and CoMSIA analyses are shown in Table 5.

To check the probability of chance correlation, PLS analysis was performed by randomization of the biological activity. This was done by randomly changing the biological activity data and

This journal is © The Royal Society of Chemistry 2006

Org. Biomol. Chem., 2006, **4**, 2858–2868 | 2867

performing PLS analysis to calculate the r^2_{cv} value for field fit of CoMFA and CoMSIA. The process was repeated 100 times.

Predictive r^2 value (r^2_{pred})

To test the predictive power of the derived CoMFA and CoMSIA models, biological activities of the test set molecules were predicted using models derived from training set. The plot of predicted *versus* observed activity of test set are shown in Fig. 3–4 and 7–8 based on CoMFA and CoMSIA, respectively.

The predictive r^2 value was calculated using the following formula:

$$r^2_{pred} = \frac{SD - PRESS}{SD},$$

where SD is the sum of squared deviation between the biological activities of the test set molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the observed and the predicted activities of the test molecules.

Conclusions

The CoMFA and CoMSIA methods have been applied to derive 3D-QSAR models for hydroxamic acid HDAC inhibitors. The models obtained using these methods showed high correlative and predictive abilities. A high bootstrapped r^2 value indicates that a similar relationship exits in all molecules. Inclusion of ClogP as an additional descriptor increased the statistical significance of the model, indicating that lipophilicity enhances the HDAC inhibitory activity. Different alignments were considered for the study. The atom-based alignment with the hydroxamic acid functional group gives a better result than other atom-based alignment. It emphasizes the importance of the interaction of hydroxamic acid with the enzyme residues. Out of all these alignments, the field fit alignment (along with ClogP) resulted in the best CoMFA model. The same alignment was also considered for CoMSIA where all five fields were considered in different combinations. The model generated using all five fields gave a statistically-significant model and explained the observed biological activities. The contour maps from both the models are similar in explaining influence of substitution on activity. The substitution by electron rich functional groups on the phenyl ring may improve activity. Hydrogen bond acceptor and donor groups also enhance the activity. Overall, electrostatic interactions play a major role in binding to the HDAC active site.

Acknowledgements

References

1   M. J. Pazin and J. T. Kadonaga, *Cell*, 1997, **89**, 325–328.
2   P. A. Marks, R. A. Rifkind, V. M. Richon, R. Breslow, T. Miller and W. K. Kelly, *Nat. Rev. Cancer*, 2001, **1**, 194–202.
3   W. K. Kelly, O. A. O'Connor and P. A. Marks, *Expert Opin. Invest. Drugs.*, 2002, **11**, 1695–1713.
4   P. A. Marks, V. M. Richon, R. Breslow and R. A. Rifkind, *Curr. Opin. Oncol.*, 2001, **13**, 477–483.
5   P. A. Marks, V. M. Richon, R. Breslow and R. A. Rifkind, *Clin. Cancer Res.*, 2001, **7**, 759–760.
6   C. J. Phiel, F. Zhang, E. Y. Huang, M. G. Guenther, M. A. Lazar and P. S. Klein, *J. Biol. Chem.*, 2001, **76**, 36734–36741.
7   P. T. Meinke and P. Liberator, *Curr. Med. Chem.*, 2001, **8**, 211–235.
8   M. Yoshida, M. Kijima, M. Akita and T. Beppu, *J. Biol. Chem.*, 1990, **265**, 17174–17179.
9   M. Kijima, M. Yoshida, K. Suguta, S. Horinouchi and T. Beppu, *J. Biol. Chem.*, 1993, **268**, 22429–22435.
10 R. E. Shute, B. Dunlap and D. H. Rich, *J. Med. Chem.*, 1987, **30**, 71–78.
11 J. W. Han, S. H. Ahn, S. Y. Wang and G. U. Seo, *Cancer Res.*, 2000, **60**, 6068–6074.
12 S. D. Gore and M. A. Carducci, *Expert Opin. Invest. Drugs.*, 2000, **9**, 2923–2934.
13 M. Göttlicher, S. Minucci, P. Zhu, O. H. Kramer, A. Schimpf, S. Giavara, J. P. Sleeman, F. Lo Coco, C. Nervi, P. G. Pelicci and T. Hainzel, *EMBO J.*, 2001, **20**, 6969–6978.
14 L. M. Butler, D. B. Agus, H. I. Scher, B. Higgins, A. Rose, C. Cordon-Cardo, H. T. Thaler, R. A. Rifkind, P. Marks and V. M. Richon, *Cancer Res.*, 2000, **60**, 5165–5170.
15 M. Jung, G. Brosch, D. Kölle, H. Scherf, C. Gerhäuser and P. Loidl, *J. Med. Chem.*, 1999, **42**, 4669–4679.
16 S. W. Remiszewski, L. C. Sambucetti, P. Atadja, K. W. Bair, W. D. Cornell, M. A. Green, K. L. Howell, M. Jung, P. Kwon, N. Trogani and H. Walker, *J. Med. Chem.*, 2002, **45**, 753–757.
17 S. H. Woo, S. Frechette, E. A. Khalil, G. Bouchain, A. Vaisburg, N. Bernstein, O. Moradei, S. Leit, M. Allan, M. Fournel, M.-C. Trachy-Bourget, Z. Li, J. M. Besterman and D. Delorme, *J. Med. Chem.*, 2002, **45**, 2877–2885.
18 Y. B. Kim, K. H. Lee, K. Sugita, M. Yoshida and S. Horinouchi, *Oncogene*, 1999, **18**, 2461–2470.
19 M. S. Finnin, J. R. Dongian, A. Cohen, V. M. Richon, R. A. Rifkina, P. A. Marks, R. Breslow and N. P. Pavletich, *Nature*, 1999, **401**, 188–193.
20 D.-F. Wang, O. G. Wiest, P. Helquist, H.-Y. Lan-Hargest and N. L. Wiech, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 707–711.
21 Y. Guo, J. Xiao, Z. Guo, F. Chu, Y. Cheng and S. Wu, *Bioorg. Med. Chem.*, 2005, **13**(18), 5424–5434.
22 S. Massa, A. Mai, G. Sbardella, M. Esposito, R. Ragno, P. Loidl and G. Brosch, *J. Med. Chem.*, 2003, **46**, 512–524.
23 A. Xie, C. Liao, Z. Li, Z. Ning, W. Hu, X. Lu, L. Shi and J. Zhou, *Curr. Med. Chem.: Anti-Cancer Agents*, 2004, **4**, 273–299.
24 R. D. Cramer, III, J. D. Bunce and D. E. Patterson, *Quant. Struct.–Act. Relat.*, 1988, **7**, 18–25.
25 G. Klebe and U. Abraham, *J. Comput. Aided Mol. Des.*, 1999, **13**, 1–10.
26 V. Hariprasad and V. M. Kulkarni, *J. Comput. Aided Mol. Des.*, 1996, **10**, 284–292.
27 S. S. Kulkarni and V. M. Kulkarni, *J. Med. Chem.*, 1999, **42**, 373–380.
28 S. V. Murthy and V. M. Kulkarni, *Bioorg. Med. Chem.*, 2002, **10**, 2267–2282.
29 M. T. Mahindra and V. M. Kulkarni, *J. Comput. Aided Mol. Des.*, 2002, **16**, 181–200.
30 *SYBYL 6.9.1*, Tripos Associates Inc., St Hanley Road, St Louis, MO, USA, 2003.
31 M. Clark, R. D. Cramer, III and N. V. Opdenbosh, *J. Comput. Chem.*, 1989, **10**, 982–1012.
32 W. J. Dunn, U. Wolds, S. Hellberg and J. Gasteiger, *Quant. Struct.–Act. Relat.*, 1984, **3**, 131–137.
33 S. Wold, *Technometrics*, 1978, **4**, 397–405.