3-D-QSAR CoMFA and CoMSIA Studies on Tetrahydrofuroyl-L-phenylalanine Derivatives as VLA-4 Antagonists☆

Akash Khandelwal,[a] Ramamurthi Narayanan[b] and Bulusu Gopalakrishnan[a,]*

[a]Department of Molecular Modeling and Drug Design, Dr. Reddy's Laboratories Ltd, Discovery Research,
Bollaram Road, Miyapur, Hyderabad 500 050, India
[b]Discovery Chemistry, Dr. Reddy's Laboratories Ltd, Discovery Research, Bollaram Road,
Miyapur, Hyderabad 500 050, India

Abstract—Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) were performed on tetrahydrofuroyl-L-phenylalanine derivatives as VLA-4 antagonists. The best CoMFA and CoMSIA models that were generated using atom based alignment from a training set of twenty five tetrahydrofuroyl-L-phenylalanine derivatives, are six-component models with good statistics; CoMFA: $r^2_{cv} = 0.366$, $r^2 = 0.983$, $s = 0.099$, $F = 172.661$ and PRESS $= 4.435$; CoMSIA: $r^2_{cv} = 0.528$, $r^2 = 0.995$, $s = 0.054$, $F = 577.87$ and PRESS $= 3.563$. Both of these 3-D-QSAR models were validated using a test set of eleven compounds, whose predicted pIC_{50} values fall within one log unit of the actual pIC_{50}. The contour diagrams obtained for the various CoMFA and CoMSIA field contributions can be mapped back onto structural features to explain the activity trends of the molecules analysed. Based on the spatial arrangement of the various field contributions, novel molecules with improved activity can be designed.

Introduction

VLA-4 ($\alpha_4\beta_1$; 'Very Late Antigen-4') is a heterodimeric cell-surface protein expressed in high levels on all lymphocytes except platelets.[1] VLA-4 binds to an alternatively spliced segment (CS-1) of fibronectin (FN) on the extracellular matrix and to vascular cell adhesion molecule-1 (VCAM-1) on the endothelium. Both CS-1 and VCAM-1 are expressed in inflammed tissues and actively participate in physiologic and pathologic responses in inflammatory and autoimmune diseases. The binding of VLA-4 to VCAM-1 leads to lymphocyte infiltration to extra cellular tissues. Antibodies against VLA-4 have been demonstrated to block lymphocyte infiltration and prevent tissue damage in animal models of inflammatory diseases such as asthma,[2] multiple

sclerosis,[3] rheumatoid arthritis[4] and inflammatory bowel disease[5] and are in advanced stages of clinical trials.[6] The search for orally active VLA-4 antagonists has resulted in many potential new chemical entities (NCEs) and many of them are in advanced stages[7] of clinical trials; however it is surprising that the application of computer aided drug design (CADD) methods for the design of novel VLA-4 antagonists has received very little attention except for few scattered reports.[8] This paper describes the application of two three dimensional quantitative structure activity relationship (3-D-QSAR) methods, comparative molecular field analysis (CoMFA)[9] and comparative molecular similarity indices analysis (CoMSIA)[10,11] of tetrahydrofuroyl-L-phenylalanine derivatives as VLA-4 antagonists[12] and significantly it is first of this type for this class of VLA-4 antagonists. The present study is aimed to gain insights into steric, electrostatic, hydrophobic and hydrogen-bonding properties, their influence on activity and to derive 3-D-QSAR models for design and prediction of the activities of new derivatives for this class of antagonists.
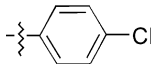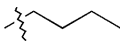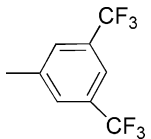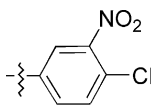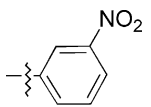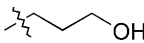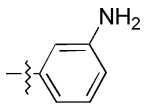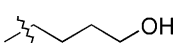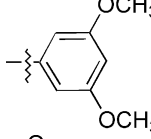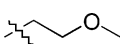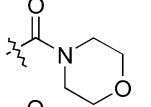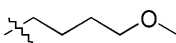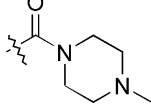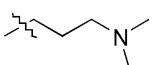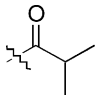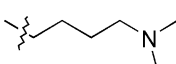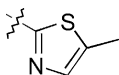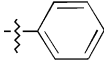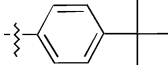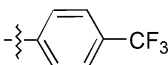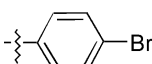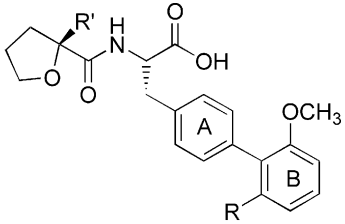
Results and Discussion
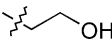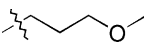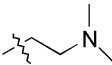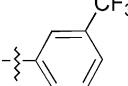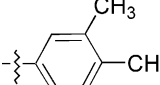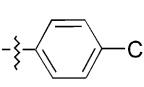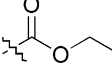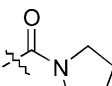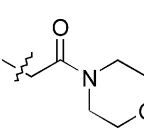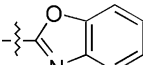
CoMFA and CoMSIA methods were employed for deriving 3-D-QSAR models for the training set of 25 Tetrahydrofuroyl-L-phenylalanine derivatives[12] (Table 1) keeping in vitro activity pIC_{50} as a dependent variable. The conformations of the molecules used in the study were generated by systematic search method. The relative alignments of the molecules were then carried out using three techniques, namely atom based fit (RMS fitting),[13] database alignment[13,14] and multifit[13] methods. The 3-D-QSAR models were validated using a test set (Table 2) of 11 compounds, which were not included in the development of the models. Results of the Partial

Table 1. Structures and biological activities of the molecules used in the training set

S. No.	Code	R	R′	IC_{50} (nM)	S. No.	Code	R	R′	IC_{50} (nM)
1	1	H	Me	3.9	**16**	20	H	(4-Cl-phenyl)	4.09
2	3	OCH_3	(butyl)	0.48					
3	4	OCH_3	(isopropyl)	0.7	**17**	23	H	(3,5-bis-CF_3-phenyl)	4.62
4	5	OCH_3	(isobutyl)	1.15					
5	7	OCH_3	(propyl-OH)	2.12	**18**	25	OCH_3	(2-Cl-3-NO_2-phenyl)	0.47
6	8	OCH_3	(butyl-OH)	255					
7	9	OCH_3	(ethyl-O-methyl)	0.7	**19**	26	OCH_3	(3-NO_2-phenyl)	0.23
8	11	OCH_3	(butyl-O-methyl)	0.62					
9	13	OCH_3	(propyl-N(Me)$_2$)	9.18	**20**	27	OCH_3	(3-NH_2-phenyl)	0.92
10	14	OCH_3	(butyl-N(Me)$_2$)	16.3	**21**	28	OCH_3	(3,5-bis-OCH_3-phenyl)	1.48
11	15	H	(phenyl)	2.2					
12	16	H	(4-tert-butyl-phenyl)	1.8	**22**	31	H	(morpholine carbonyl)	0.54
13	17	H	(4-F-phenyl)	4.3	**23**	32	H	(N-methyl-piperazine carbonyl)	0.59
14	18	H	(4-CF_3-phenyl)	5.31	**24**	35	OCH_3	(isopropyl carbonyl)	0.67
15	19	H	(4-Br-phenyl)	3.62	**25**	37	OCH_3	(5-methyl-thiazol-2-yl)	0.31

Table 2. Structures and biological activities of the molecules used in the test set

S. No.	Code	R	R'	IC$_{50}$ (nM)
1	2	OCH$_3$	CH$_3$	0.34
2	6	OCH$_3$	—CH$_2$CH$_2$—OH	1.56
3	10	OCH$_3$	—CH$_2$CH$_2$CH$_2$—O—	0.35
4	12	OCH$_3$	—CH$_2$CH$_2$—N(CH$_3$)$_2$	15.5
5	21	H	phenyl-CF$_3$	4.65
6	22	H	dimethylphenyl (CH$_3$, CH$_3$)	6.18
7	24	OCH$_3$	phenyl-Cl	0.55
8	30	H	—CH$_2$—C(O)—O—C$_2$H$_5$	4.06
9	33	H	—C(O)—pyrrolidine	1.33
10	34	H	—CH$_2$—C(O)—morpholine	23.15
11	36	OCH$_3$	benzoxazol-2-yl	0.42

Least Squares (PLS) analysis are shown in Table 3a and b.

CoMFA 3-D-QSAR models

Three CoMFA models using atom-based, database and multifit alignments were developed. The cross-validated r^2 for atom-based and multifit alignments are similar (0.366 and 0.340), while it is slightly higher (0.410) in case of database alignment. The non-cross-validated r^2 with six components in case of atom based and database alignments are 0.983 and 0.992, while it is 0.951 in case of multifit alignments with five components. The Predictive Residual Sum of Squares (PRESS) values of

atom-based and multifit alignments are 4.43 and 4.59, respectively, while in case of database alignment it is slightly higher 5.309. Though the cross validated r^2 is better in case of database alignment than the other two alignment methods, the PRESS value is higher in case of database alignment method. This further supports the finding of Golbraikh et al.[15] that cross validated r^2 is not the actual indicator of external predictivity of the 3-D-QSAR models. The internal consistency of the model obtained from database alignment appears better, compared to the models obtained from the other two methods. However, the external consistency is better in case of models obtained from atom-based and multifit alignment methods. The steric and electrostatic field contributions of all the three models are similar. Based on the above-mentioned observations it was difficult to choose the model for further analysis; so all the three models were analysed. The actual and predicted pIC$_{50}$ values of the training set are shown in Table 4.

The CoMFA steric and electrostatic field contour plots obtained from atom-based alignment is shown in Figures 1 and 2. The green regions indicate areas where steric bulk enhances biological activity, whereas the yellow contours indicate regions where steric bulk is detrimental to biological activity. Blue coloured regions indicate areas where electropositive groups enhance biological activity, while red regions represent areas where electronegative groups enhance activity. The higher activity of compound **3** as compared to **1**, **4**, and **5** is due to the fact the n-butyl group of **3** occupies space in green contours. The higher activity of **7** as compared to **8** is due to the fact that the OH group of **8** occupies yellow contours. The N,N-dimethyl group of compound **14** occupies position in yellow contour and hence exhibits lower activity than corresponding **13**. Red contours enclose part of phenyl ring of **15**, **16**, **17**, **18**, **19**, **20**, **23**, **25**, **26**, **27** and **28** the negative charge on phenyl ring of these compounds might have contributed to these contours. Compounds **17**, **19** and **20** with substituents F, Cl, Br have higher activity than **18** with CF$_3$ substitution because the former groups fit into the red contours while the later group protrudes out of the red contour and touches the blue contour (electropositive region). Because of electronegative groups present at the 3 position of the phenyl ring, compounds **25**, **26** and **27** have higher activity as compared to 3, 5 disubstituted compounds **23** and **28**; this can be attributed to the fact that the substituents at the 5th position occupy blue contours (electropositive region). The higher activity of **16** as compared to **15** is due to the t-butyl group of **16** that occupies blue contour. The difference in the activity of compounds **31**, **32** and **37** is due to the relative positioning of piperidine, morpholine and 2-methyl thiazole rings in green contours. The methyl group of **37** completely occupies green contours while the piperidine and morpholine rings of **31** and **32** partly cover the green contours.

The contours and the orientation of the molecules inside the contours in case of database and multifit alignments are similar to that of atom based alignment except that a small green contour is present in the 6-position of ring B in case of database alignment.

Table 3.
(a) Summary of CoMFA results

Fields[a]	Atom based alignment	Data base alignment	Multifit
r^{2}[b]	**0.983**	**0.992**	**0.951**
N[c]	6	6	5
r^{2}_{cv} [d]	**0.366**	**0.410**	**0.340**
SEE[e]	0.099	0.070	0.163
F-Value[f]	172.661	352.21	73.688
Field contribution[g]	0.577 (S), 0.423 (E)	0.518(S), 0.482 (E)	0.599 (S), 0.401(E)
PRESS[h]	4.435	5.309	4.595
r^{2}_{pred}	−0.00874	−0.2076	−0.0451

(b) Summary of CoMSIA results

Fields[a]	Atom based Alignment						
	S, E	S, E, H	S, E, A	S, E, D	S, E, H, A	S, E, H, D	ALL
r^{2} [b]	0.81	0.817	0.963	0.981	0.991	0.993	**0.995**
N [c]	2	3	4	5	6	6	6
r^{2}_{cv} [d]	0.38	0.446	0.472	0.475	0.474	0.518	**0.528**
SEE [e]	0.299	0.301	0.138	0.102	0.074	0.064	0.054
F-Value [f]	47.01	31.18	131.34	197.01	315.08	412.81	577.87
Field contribution [g]	0.29, 0.71	0.16, 0.47, 0.37	0.17, 0.43, 0.39	0.22, 0.57, 0.23	0.12, 0.33, 0.27, 0.29	0.15, 0.38, 0.31, 0.15	0.12, 0.27, 0.25, 0.24, 0.12
PRESS [h]	3.985	3.637	5.3247	4.236	4.992	4.03	3.563
r^{2}_{pred}	0.0936	0.173	−0.211	0.365	−0.135	0.0833	0.189
Database alignment							
r^{2} [b]	0.862	0.817	0.971	0.985	0.985	0.991	**0.996**
N [c]	3	3	4	5	5	6	6
r^{2}_{cv} [d]	0.345	0.422	0.418	0.468	0.432	0.461	**0.496**
SEE [e]	0.261	0.3	0.123	0.09	0.091	0.071	0.05
F-Value [f]	43.66	31.2	165.82	253.58	246.58	338.31	677.83
Field contribution [g]	0.28, 0.72	0.16, 0.44, 0.40	0.18, 0.41, 0.41	0.23, 0.55, 0.23	0.12, 0.29, 0.30, 0.29	0.16, 0.34, 0.34, 0.16	0.12, 0.25, 0.27, 0.23, 0.13
PRESS [h]	5.385	3.877	5.669	4.366	5.709	4.429	3.684
r^{2}_{pred}	−0.225	0.118	−0.289	0.007	−0.298	−0.007	0.162
Multifit							
r^{2} [b]	0.735	0.79	0.984	0.994	0.977	0.994	**0.992**
N [c]	2	3	5	6	5	6	6
r^{2}_{cv} [d]	0.374	0.404	0.505	0.542	0.459	0.516	**0.516**
SEE [e]	0.353	0.322	0.044	0.058	0.111	0.06	0.07
F-Value [f]	30.55	26.34	227.93	505.9	164.374	483.45	351.89
Field contribution [g]	0.27, 0.73	0.16, 0.48, 0.36	0.16, 0.45, 0.38	0.19, 0.56, 0.25	0.12, 0.32, 0.28, 0.28	0.14, 0.37, 0.30, 0.19	0.11, 0.29, 0.25, 0.22, 0.13
PRESS [h]	4.528	3.108	5.179	4.212	5.016	3.821	3.682
r^{2}_{pred}	−0.0299	0.2931	−0.178	0.0419	−0.1408	0.1309	0.1625

[a]Comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) with different field combinations like steric (S), electrostatic (E), hydrophobic (H), donor (D), and acceptor (A) fields.
[b]Non cross-validated correlation coefficient.
[c]Optimum number of components obtained form cross-validated PLS analysis and same used in final non-cross-validated analysis.
[d]Cross-validated correlation coefficient.
[e]Standard error of estimate.
[f]F-test value.
[g]Field contributions: Steric and electrostatic fields from CoMFA. Steric, electrostatic, hydrophobic, donor and acceptor fields from CoMSIA.
[h]Predictive residual sum of squares of the training set.

CoMSIA 3-D-QSAR models

CoMSIA is a relatively new alternative molecular field analysis method to CoMFA. It is believed to be less affected by changes in molecular alignment and provides more smooth and interpretable contour maps as a result of employing Gaussian type distance dependence with the molecular similarity indices it uses.[10] Furthermore, in addition to the steric and electrostatic fields of CoMFA, CoMSIA defines explicit hydrophobic and hydrogen bond donor and acceptor descriptor fields. In most instances CoMSIA performs similar[16,17] to

**Table 4.** Actual and predicted pIC_{50} of training set molecules from CoMFA

Code	Actual	Atom based alignment		Database alignment		Multifit	
		Predicted	Residual	Predicted	Residual	Predicted	Residual
1	8.409	8.492	−0.083	8.519	−0.110	8.434	−0.025
3	9.319	9.331	−0.012	9.229	0.090	9.123	0.196
4	9.155	9.153	0.002	9.117	0.038	9.166	−0.011
5	8.939	8.934	0.005	8.947	−0.008	9.064	−0.125
7	8.674	8.598	0.076	8.674	0.000	8.79	−0.116
8	6.594	6.634	−0.041	6.59	0.003	6.742	−0.149
9	9.155	8.947	0.208	9.174	−0.019	8.715	0.440
11	9.208	9.104	0.104	9.212	−0.004	9.117	0.091
13	8.037	8.03	0.007	8.013	0.024	7.823	0.214
14	7.788	7.879	−0.091	7.820	−0.032	7.633	0.155
15	8.658	8.597	0.061	8.615	0.043	8.662	−0.004
16	8.745	8.711	0.034	8.696	0.049	8.729	0.016
17	8.367	8.394	−0.027	8.440	−0.073	8.478	−0.111
18	8.275	8.276	−0.001	8.252	0.023	8.437	−0.162
19	8.441	8.453	−0.012	8.400	0.041	8.58	−0.139
20	8.388	8.442	−0.054	8.346	0.042	8.501	−0.113
23	8.335	8.16	0.175	8.348	−0.013	8.319	0.016
25	9.328	9.355	−0.027	9.345	−0.017	9.409	−0.081
26	9.638	9.562	0.076	9.583	0.055	9.549	0.089
27	9.036	9.129	−0.093	9.018	0.018	9.199	−0.163
28	8.830	8.891	−0.061	8.877	−0.047	8.785	0.045
31	9.268	9.256	0.012	9.276	−0.008	9.217	0.051
32	9.229	9.244	−0.015	9.260	−0.031	9.184	0.045
35	9.174	9.372	−0.198	9.147	0.027	9.337	−0.163
37	9.509	9.553	−0.044	9.599	−0.090	9.504	0.005

CoMFA with regard to predictive ability, sometimes slightly better,[10] or slightly worse[18] than CoMFA.

PLS analyses of CoMSIA models obtained from three different alignments are shown in Table 3b. There is little to choose between the different alignments and the
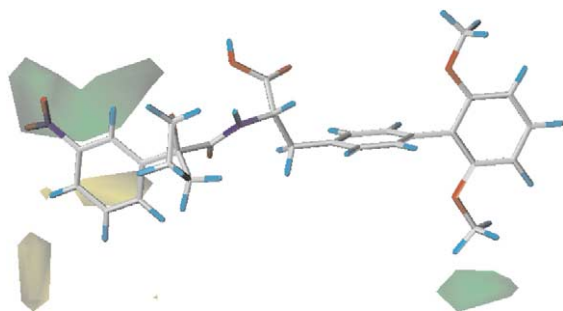
Figure 1. CoMFA stdev*coeff. steric contour plots; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky groups decrease activity.
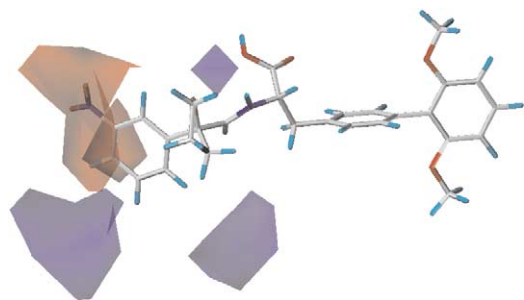
Figure 2. CoMFA stdev*coeff. electrostatic contour plot; blue contour indicate regions where positive groups increase activity, whereas red contours indicate regions where negative charge increases activity.

various models in each case, based on the combination of descriptors used (Table 3b). In the case of atom-based and database alignment methods, the model that included all the descriptors turns out to be the best. In contrast, for the multifit alignment method, models that did not include the hydrogen bond acceptor fields performed better. The steric, electrostatic, hydrophobic, H-bond donor and acceptor field contributions for all the three alignment methods are similar. The cross-validated r^2 for all the three models obtained from atom based, database and multifit alignments are 0.528, 0.496 and 0.516, respectively. The external predictivity of the models obtained from atom based alignment, database alignment and multifit are very similar, PRESS value of 3.56, 3.68 and 3.68, respectively. The actual and predicted pIC_{50} values of the training set are shown in Table 5.

The CoMSIA steric and electrostatic contour plots obtained from atom-based alignment are shown in Figures 3 and 4 respectively. The green regions indicate areas where steric bulk enhances biological activity, whereas the yellow contours indicate regions where steric bulk is detrimental to biological activity. Blue coloured regions indicate areas where electropositive groups enhance biological activity, while red regions represent areas where electronegative groups enhance activity. The hydrophobic, hydrogen bond donor and acceptor contours are displayed in Figures 5–7, respectively. The hydrophobic fields (Yellow, hydrophobic group favoured; White, hydrophobic disfavoured), H-bond donor (Cyan, favoured; purple, disfavoured) and the H-bond acceptor (Magenta, favoured; Red, disfavoured) fields indicate areas around the molecules where changes either increase or decrease activity.

Table 5. Actual and Predicted pIC_{50} of training set molecules in CoMSIA

| Code	Actual	Atom based alignment		Database alignment		Multifit	
		Predicted	Residual	Predicted	Residual	Predicted	Residual
1	8.409	8.519	−0.110	8.503	−0.094	8.55	−0.14
3	9.319	9.229	0.090	9.222	0.097	9.19	0.13
4	9.155	9.117	0.038	9.111	0.044	9.105	0.045
5	8.939	8.947	−0.008	8.956	−0.017	8.923	0.017
7	8.674	8.674	0.000	8.684	−0.010	8.69	−0.02
8	6.594	6.59	0.003	6.595	−0.002	6.607	−0.017
9	9.155	9.174	−0.019	9.172	−0.017	9.151	−0.001
11	9.208	9.212	−0.004	9.200	0.008	9.203	0.007
13	8.037	8.013	0.024	8.026	0.011	7.992	0.048
14	7.788	7.82	−0.032	7.807	−0.019	7.85	−0.06
15	8.658	8.615	0.043	8.636	0.022	8.653	0.007
16	8.745	8.696	0.049	8.716	0.029	8.735	0.005
17	8.367	8.44	−0.073	8.426	−0.059	8.383	−0.013
18	8.275	8.252	0.023	8.259	0.016	8.266	0.004
19	8.441	8.4	0.041	8.379	0.062	8.364	0.076
20	8.388	8.346	0.042	8.336	0.052	8.342	0.048
23	8.335	8.348	−0.013	8.355	−0.020	8.307	0.033
25	9.328	9.345	−0.017	9.324	0.004	9.33	0.000
26	9.638	9.583	0.055	9.607	0.031	9.57	0.07
27	9.036	9.018	0.018	9.018	0.018	8.956	0.084
28	8.830	8.877	−0.047	8.868	−0.038	8.894	−0.064
31	9.268	9.276	−0.008	9.272	−0.004	9.319	−0.049
32	9.229	9.26	−0.031	9.268	−0.039	9.28	−0.05
35	9.174	9.147	0.027	9.161	0.013	9.241	−0.071
37	9.509	9.599	−0.090	9.596	−0.087	9.594	−0.084

Small green and red contours are present near the 6′-position of the B ring; the methoxy group present in compounds **3**, **4**, **5**, **7**, **8**, **11**, **14**, **25**, **26**, **27**, **28**, **35** and **37** might have contributed to these contours. Higher activity of **3** is due to the fact that it occupies green contours while in **4** and **5**, it occupies space in yellow contours.
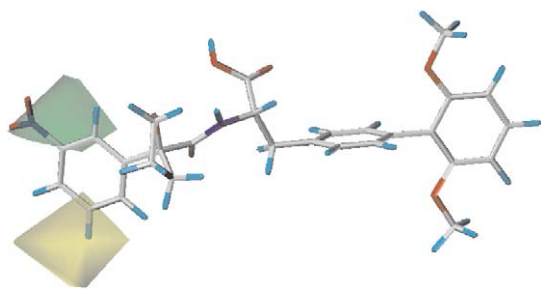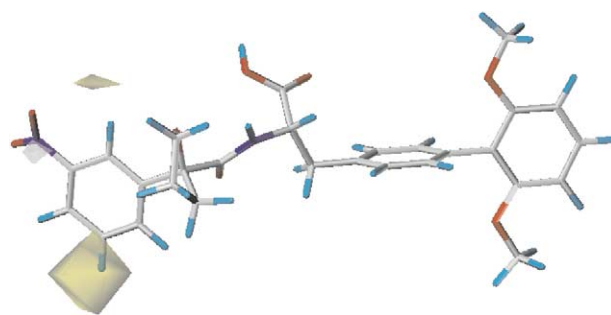

Figure 3. CoMSIA stdev*coeff. steric contour plots; green contours indicate regions where bulky groups increase activity, whereas yellow contours indicate regions where bulky group decrease activity.
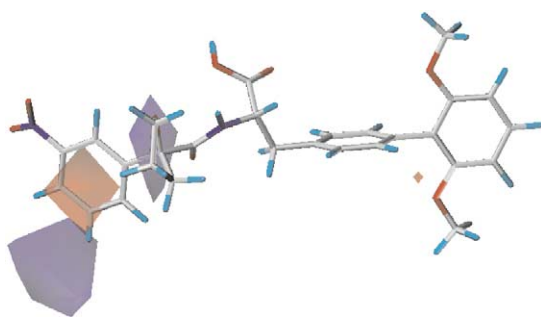

Figure 4. CoMSIA stdev*coeff. electrostatic contour plot; blue contour indicate regions where positive groups increase activity, whereas red contours indicate regions where negative charge increases activity.

The *N,N* dimethyl group of compound **14** occupies position in yellow contour and hence exhibits lower activity than corresponding **13**. The phenyl and *t*-butyl phenyl substituent of compound **15** and **16** are sandwitched between the green and yellow contours, in addition the *t*-butyl substituent of **16** occupies position in blue contour and hence exhibits better activity than **15**. Substituent at the 3 position of the phenyl ring of R^1 as in compound **26** and **27** occupies space in green contours and hence exhibits higher activity as compared to the unsubstituted R^1, compound **15**. The lower activity of compound **23** and **28** (disubstituted) as compared to monosubstituted **26** and **27** is because substitution at the 5-position occupies space in yellow contour (sterically unfavourable region) while substituents at 3 position occupies space in green contours. The higher activity of **31**, **32**, and **37** may be because their morpholine, piperidine and substitituted thiazole occupies space in green
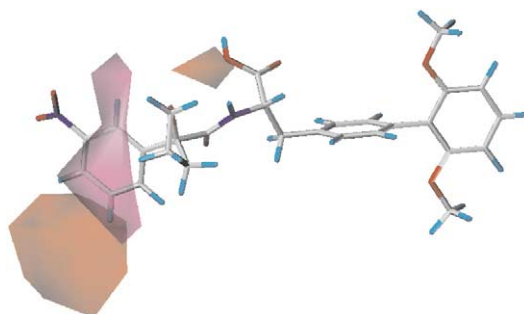

Figure 5. CoMSIA stdev*coeff. hydrophobic contour plots; yellow contours indicate regions where hydrophobic groups increases activity, whereas white contours indicates regions where hydrophobic group decreases activity.
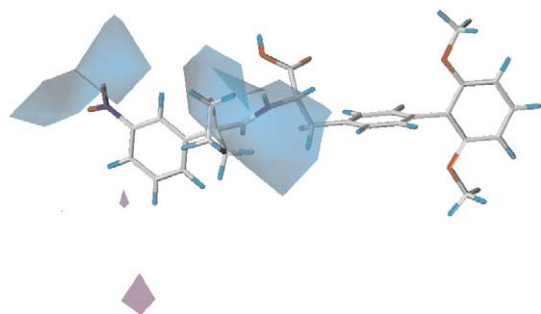
Figure 6. CoMSIA stdev*coeff H-bond donor plots: cyan contours indicate regions where H-bond donor group increases activity, whereas purple contours indicate regions where H-bond donor group decreases activity.

Figure 7. CoMSIA stdev*coeff H-bond acceptor plots: magenta contours indicate regions where H-bond acceptor group increases activity, whereas red contours indicate regions where H-bond acceptor group decreases activity.

contours. The OH group (R^1) of **8** occupies position near purple contour (donor, unfavourable) and hence exhibits lesser activity. The relatively lower activity of **9** as compared to **11** can be ascribed by the observation that the OCH_3 group of **9** touches the red contour (H-bond acceptor, unfavourable) while in case of **11** OCH_3 is near to red contour. The *N,N*-dimethyl group of **13** is near to red contour (H-bond acceptor, unfavourable) while it occupies red contour in case of **14** and hence **14** shows less activity than **13**. The CO substituent at the R^1 position in compounds **31**, **32** and **35** occupies space in magenta contours (H-bond acceptor favourable) and, in the case of **37**, the N of the thiazole ring occupies magenta contours and hence these contours exhibit higher activity.

The contours and the orientation of the molecules inside the contours in case of database and multifit alignments are similar to that of atom based alignment except that no small contour is present at the 6-position of ring B in case of database alignment.

The predictive power of each of the 3-D-QSAR models was evaluated by predicting the pIC_{50} of test set of 11 Tetrahydrofuroyl-L-phenylalanine derivatives. In almost all the cases, the predictive values fall close to the actual pIC_{50} values, deviating by not more than 1 logarithmic unit (Tables 6 and 7).

Based on the PLS statistics of both CoMFA (Table 3a) and CoMSIA (Table 3b) 3-D-QSAR models, it is clear

Table 6. Actual and predicted pIC_{50} of test set molecules in CoMFA

Code	Actual	Atom based alignment		Database alignment		Multifit	
		Predicted	Residual	Predicted	Residual	Predicted	Residual
2	9.469	8.696	0.772	8.793	0.675	8.642	0.826
6	8.807	9.072	−0.265	9.369	−0.562	8.847	−0.040
10	9.456	8.629	0.827	8.728	0.728	8.885	0.571
12	7.810	8.379	−0.569	8.749	−0.939	8.395	−0.586
21	8.333	9.050	−0.717	9.229	−0.897	9.265	−0.933
22	8.209	8.639	−0.430	8.897	−0.688	8.922	−0.713
24	9.260	8.747	0.513	8.629	0.630	8.794	0.466
30	8.392	9.246	−0.854	8.871	−0.480	9.112	−0.720
33	8.876	9.259	−0.383	9.283	−0.407	9.032	−0.156
34	7.635	8.635	−1.000	8.660	−1.024	8.603	−0.968
36	9.376	9.591	0.215	9.549	0.173	9.426	−0.05

Table 7. Actual and predicted pIC_{50} of test set molecules in CoMSIA

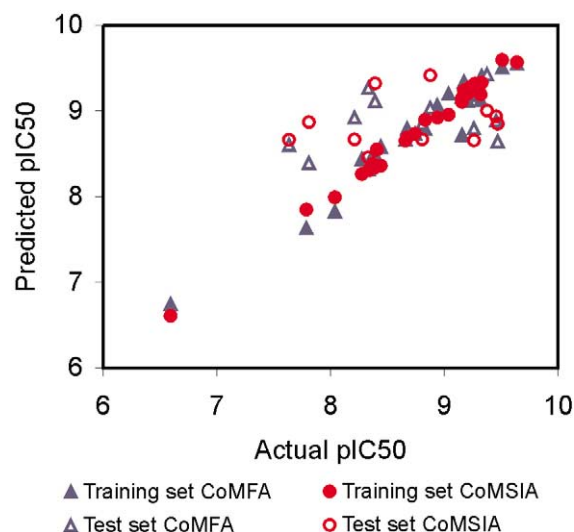| Code	Actual	Atom based alignment		Database alignment		Multifit	
		Predicted	Residual	Predicted	Residual	Predicted	Residual
2	9.469	8.857	0.611	8.825	0.644	8.850	0.619
6	8.807	8.739	0.068	8.752	0.055	8.666	0.141
10	9.456	8.874	0.582	8.898	0.558	8.936	0.520
12	7.810	8.830	−1.021	8.793	−0.983	8.869	−1.059
21	8.333	8.536	−0.204	8.588	−0.255	8.462	−0.129
22	8.209	8.652	−0.443	8.695	−0.486	8.665	−0.456
24	9.260	8.797	0.463	8.718	0.541	8.655	0.604
30	8.392	9.363	−0.971	9.382	−0.990	9.321	−0.930
33	8.876	9.395	−0.519	9.421	−0.545	9.417	−0.541
34	7.635	8.783	−1.147	8.924	−1.288	8.664	−1.028
36	9.376	9.564	−0.188	9.034	0.342	9.002	0.374
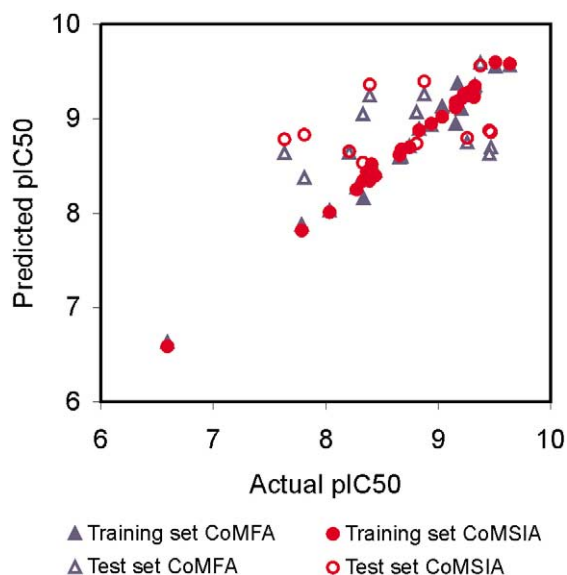
Figure 8. Graph of actual versus predicted pIC_{50} of training and test set molecules for CoMFA and CoMSIA 3-D-QSAR models (atom based alignment).
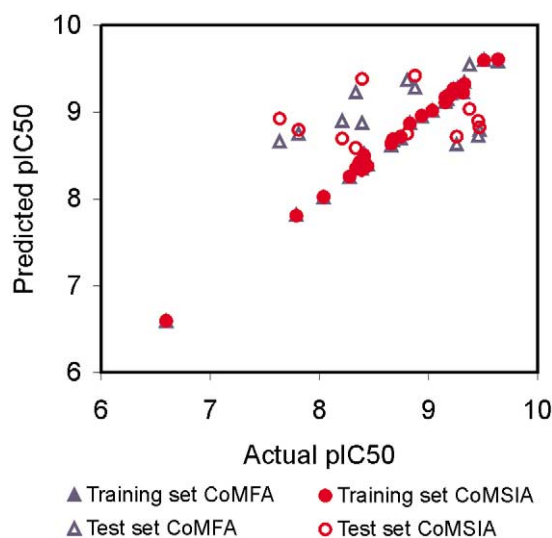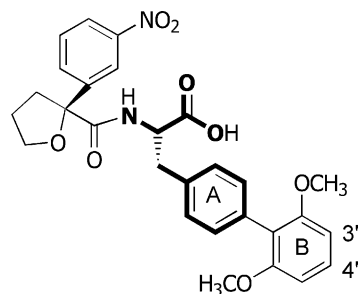

Figure 9. Graph of actual versis predicted pIC_{50} of training and test set molecules for CoMFA and CoMSIA 3-D-QSAR models (database alignment).

Figure 10. Graph of actual versus predicted pIC_{50} of training and test set molecules for CoMFA and CoMSIA 3-D-QSAR models (multifit).

Figure 11. Template used for molecular alignment. Reference atoms (in bold face) used in atom based, database and multifit alignment protocols.

Figure 12. Alignment of the tetrahydrofuroyl-L-phenylalanine derivatives.
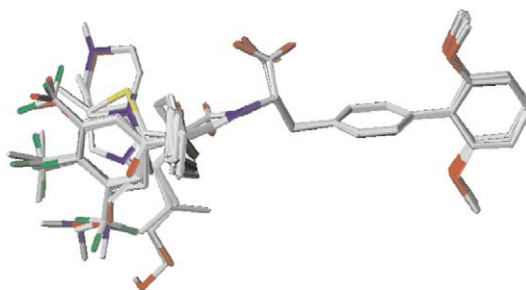
that CoMSIA is better than CoMFA; this is further validated by the residual values of the test set (Tables 6 and 7). Figures 8–10 represent graphs of the actual versus predicted pIC_{50} values of the training and test set molecules for CoMFA and CoMSIA 3-D-QSAR models using atom based, database and multifit alignment, respectively. The superior performance of CoMSIA relative to CoMFA may be ascribed mainly to the higher contributions from the hydrophobic, H-bond donor and acceptor fields to the CoMSIA 3-D-QSAR models (Table 3b). CoMFA unlike CoMSIA, does not have explicit hydrophobic and hydrogen bonding descriptors, which are assumed to be implicitly treated in the CoMFA steric and electrostatic fields, respec-

tively. Overall, the electrostatic field components made higher contributions to the CoMSIA 3-D-QSAR models (Table 3b).

Conclusion

We have developed predictive CoMFA and CoMSIA 3-D-QSAR models for Tetrahydrofuroyl-L-phenyl-alanine derivatives, and the CoMSIA 3-D-QSAR models appear better than CoMFA models. The contour diagrams obtained for the various CoMFA and CoM-SIA field contributions can be mapped back onto

structural features accounting for the activity trends among the molecules analysed. Based on the spatial arrangement of the various field contributions, novel molecules can be designed that may possess improved activity. The CoMFA and CoMSIA models obtained in this study are, to our knowledge, the first 3-D-QSAR studies reported for Tetrahydrofuroyl-l-phenylalanine derivatives as VLA-4 antagonists.

Methods

Data sets and biological activity

Recently Doherty et al.[12] have reported tetrahydrofuroyl-L-phenylalanine derivatives as potent and specific antagonists of VLA-4 and they constitute the training and test sets used in the present study. The IC_{50} values were converted to pIC_{50} according to the formula.

$$pIC_{50} = -\log IC_{50}$$

pIC_{50} values were used as dependent variables in the CoMFA and CoMSIA analysis. The training set of 25 molecules with structures and their activities are shown in Table 1. They are selected based on (a) structural diversity available at the 2-position of tetrahydrofuran and at the 4-position of phenylalanine moieties and (b) the biological activities (pIC_{50}) in the range of three log units. Molecules of the test set were selected similarly. Thus, the mean of the biological activity (pIC_{50}) of the chosen training set and test set molecules were 8.740 and 8.799, respectively. The predictive power of the models was assessed using a test set of 11 compounds whose structures and activities are shown in Table 2.

Molecular modelling and alignment

Three-dimensional structure building and molecular modelling studies were performed using SYBYL 6.8[13] installed on a Silicon Graphics workstation with IRIX 6.5 operating system. Energy minimisations were performed using MMFF94[19] force field and charges with distance dependent dielectric and conjugate gradient method with a convergence criterion of 0.01 kcal/mol.

The most important requirement for 3-D-QSAR techniques (CoMFA and CoMSIA) is that the 3-D structures of the molecules to be analysed be aligned according to a suitable conformational template, which is assumed to adopt a 'bioactive conformation'. In the present study, the structural information on these inhibitor-protein complexes is not available; therefore, the conformation of the molecules was obtained from systematic conformational search[13] procedures. Three different molecular alignment methods were employed in the present study.

Alignment 1. This method involves atom based fitting (RMS fitting)[13] of the ligands. The compounds were fitted to the template molecule 26, one of the most

active molecules (Fig. 11) and all the aligned molecules of the training set are shown in Figure 12.

Alignment 2. The molecules were aligned by database alignment[13,14] method.

Alignment 3. The alignment of the molecules was carried out by flexible fitting (multifit)[13] of the atoms of the ligand to the template molecule 26. This method involves fitting to the template molecule by applying force (the default SYBYL spring constant value of 20 kcal/mol was used), followed by minimisation.

CoMFA and CoMSIA 3-D QSAR models

In deriving the CoMFA and CoMSIA descriptor fields a 3-D cubic lattice with grid spacing of 2.0 Å in x, y and z directions was created to encompass the aligned molecules. CoMFA descriptors were calculated using an sp^3 carbon probe atom carrying $+1$ charge to generate steric (Lennard–Jones 6–12 potential) field energies and electrostatic (Coulomb potential) fields with a distance dependent dielectric at each lattice points. Values of steric and electrostatic energy were truncated at 30 kcal/mol. The CoMFA steric and electrostatic fields generated were scaled by the CoMFA-STD method in SYBYL.

CoMSIA similarity indices descriptors were derived according to Klebe et al.[10] in the same manner as for the CoMFA calculations. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q are calculated by eq 1 as follows:

$$A_{F,K}^{q}(j) = -\Sigma \omega_{probe,k} \omega_{ik} e_{iq^2}^{-\alpha,r} \tag{1}$$

Five physicochemical properties k (steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor) were evaluated using the probe atom. A Gaussian type dielectric dependence was used between the grid point q and each atom i of the molecule. The default value of 0.3 was used as the attenuation factor (α). In CoMSIA, the steric indices are related to the third power of the atomic radii, the electrostatic descriptors are derived from partial atomic charges, the hydrophobic field are derived from atom-based parameters,[20] and the hydrogen bond donor and acceptor indices are obtained by a rule-based method derived from experimental values.[21]

The CoMFA and CoMSIA descriptors were used as independent variables, and pIC_{50} values were used as dependent variables in partial least square analysis to derive 3-D-QSAR models using the standard implementation in the SYBYL package. The predictive value of the models was evaluated first by leave-one-out (LOO) cross-validation. The cross-validated r_{CV}^2 was calculated using eq 2.

$$r_{CV}^2 = 1 - [\Sigma (Y_{predicted} - Y_{observed})^2 /$$
$$\Sigma (Y_{observed} - Y_{mean})^2] \tag{2}$$

where, $Y_{predicted}$, $Y_{observed}$ and Y_{mean} are predicted, actual and mean values of the target property (pIC_{50}), respectively. $\Sigma(Y_{predicted} - Y_{observed})^2$ are the predictive residual sum of squares (PRESS). The optimum number of components was chosen which gave less standard error of prediction and more r_{CV}^2. In addition to the r_{CV}^2 and number of components, the conventional correlation coefficient r^2 and its standard error were also computed.

Acknowledgements

References and Notes

1. Helmer, M. E.; Elices, M. J.; Parker, C.; Takada, Y. *Immunol. Rev.* **1990**, *114*, 45.

2. (a) Abraham, W. M.; Sielczak, M. W.; Ahmed, A.; Cortes, A.; Lauredo, I. T.; Kim, J.; Pepinsky, B.; Benjamin, C. D.; Leone, D. R.; Lobb, R. R.; Weller, P. F. *J. Clin. Invest.* **1994**, *93*, 776. (b) Milne, A. A. Y.; Piper, P. J. *Eur. J. Pharmacol.* **1995**, *282*, 243.

3. (a) Yednock, T. A.; Cannon, D.; Fritz, L. C. *Nature* **1992**, *356*, 63. (b) Kesztheylyi, E.; Karlik, S.; Hyduk, S.; Rice, A.; Gordon, G.; Yednock, T.; Horner, H. *Neurology* **1996**, *47*, 1053.

4. Seifge, D. J. *Rheumatologia* **1996**, *23*, 2076.

5. Podolsky, D. K.; Lobb, R.; King, N.; Benjamin, C. D.; Pepinsky, B.; Sehgal, P.; DeBaumont, M. *J. Clin. Invest.* **1993**, *92*, 372.

6. Natalizumab, Biogen, Inc, Press Release 2002, April 29, Reference number RF449588.

7. (a) SB-683698, GlaxoSmithKline plc, Company Brochure 2001, February 22, Reference number RF399771. (b) GW-559090, GlaxoSmithKline plc, Company Publication 2001, October 23, Reference number RF426830. (c) IVL-745 and IVL-1031, Aventis Pharma AG, Company Presentation 2001, May 15, Reference number RF409257. (d) R-411, Hoffmann-La Roche Inc, Company Presentation 2002, April 16, Reference number RF448015.

8. (a) Chen, Q.; Li, j.; Maxwell, D.; Holland, G.; Dixon, R.; You, T. 220th National Meeting of the American Chemical Society, Boston, Aug 20–24, 2000; American Chemical Society: Washington, DC, 2000; MEDI 140. (b) Singh, J.; Zheng, Z.; Sprague, P.; Van Vilhemen, H.; Castro, A.C.; Adams, S.P. WO 9804913. (c) Singh, J.; Vlijmen, H. V.; Liao, Y.; Lee, W.; Mark, C.; Harris, M.; Shu, I.; Gill, A.; Cuervo, J.; Abraham, W. M.; Adams, S. *P. J. Med. Chem.* **2002**, *45*, 2988.

9. Cramer, R. D.III; Patterson, DeE.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959.

10. Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130.

11. Bohm, M.; Stuzerbercher, J.; Kliebe, G. *J. Med. Chem.* **1999**, *42*, 458.

12. Doherty, G. A.; Yang, G. X.; Chang, L. L.; MacCoss, M.; Tong, S.; Kidambi, U.; Egger, L. A.; McCauly, E.; Riper, G. V.; Mumford, R. A.; Schmidt, J. A.; Hagmann, W. K. *Bioorg. Med. Chem.* **2002**, *12*, 1501.

13. *SYBYL 6.8*, Tripos Inc.: 1669, S. Hanley Road, Suite 303, St. Louis, MO 63144-2913, USA.

14. Desiraju, G. R.; Gopalakrishnan, B.; Jetti, R. K. R.; Nagaraju, A.; Raveendra, D.; Sarma, J. A. R. P.; Sobhia, M. E.; Thilagavathi, R. *J. Med. Chem.* **2002**, *22*, 4847.

15. Golbraikh, A.; Tropsha, A. *J. Mol. Mod* **2002**, *20*, 269.

16. Suh, M.-E.; Kang, M-J.; Park, S.-O. *Bioorg. Med. Chem.* **2001**, *9*, 2987.

17. Kim, K. H. *Bioorg. Med. Chem.* **2001**, *9*, 1517.

18. Gopalakrishnan, B., Khandelwal, A., Rajjak, S.A., Selvakumar, N., Das, J., Trehan, S., Iqbal, J., Sitaram Kumar, M., Rajagopalan, R. 42nd Interscience Conference on Antimicrobial Agents and Chemotherapy, San Diego, CA, Sept 27–30, 2002; American Society for Microbiology: Washington, DC, 2002.

19. Halgren, T. *J. Comp. Chem.* **1999**, *20*, 720.

20. Viswanadhan, V. N.; Ghose, A. K.; Revenkar, G. R.; Robins, R. *J. Chem. Info. Comput. Sci.* **1989**, *29*, 163.

21. Klebe, G. . *J. Mol. Biol.* **1994**, *237*, 212.