European Journal of Medicinal Chemistry 44 (2009) 2400–2407

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Development of CoMFA and CoMSIA models of cytotoxicity data of anti-HIV-1-phenylamino-1H-imidazole derivatives

Arijit Basu*, Kalsariya Jasu, Venkatesan Jayaprakash, Nibha Mishra, P. Ojha, S. Bhattacharya

Department of Pharmaceutical Sciences, Birla Institute of Technology, Mesra 835 215, Ranchi, India

ARTICLE INFO

Article history:
Received 15 February 2008
Received in revised form 5 September 2008
Accepted 8 September 2008
Available online 10 October 2008

Keywords:
Amino imidazoles
Anti-HIV
Cytotoxicity data
CoMFA and CoMSIA models

ABSTRACT

3D-QSAR models of 1-Phenylamino-1H-imidazole derivatives with cytotoxic potential have been developed using CoMFA and CoMSIA. Models were built keeping both 10% and 25% of molecules in test set. The Database and Field-Fit alignments were used for CoMFA model development. The best QSAR model was obtained from CoMFA analysis using Database alignment and employing 25% molecules in the test set (r^2_{pred} of 0.91 and r^2_m of 0.652). Database alignment with different fields such as Steric (S), Electrostatic (E), Acceptor (A), Donor (D) and Hydrophobic (H) was employed for CoMSIA model development. The best CoMSIA model was obtained by using the SHE fields and employing 25% molecules in the test set (r^2_{pred} of 0.789 and r^2_m of 0.410).

1. Introduction

Acquired Immuno Deficiency Syndrome (AIDS) is caused by the Human Immuno deficiency Virus type 1 (HIV-1). An estimated 38.6 million people live with HIV-1 worldwide, while about 25 million have died already [1,2]. The pandemic spread of this disease has prompted an unprecedented scientific and clinical effort to understand and combat it. Anti-AIDS drugs [3–5] which were earlier classified into three categories, the Nucleoside Reverse Transcriptase Inhibitors (NRTIs) that act as chain terminators to block the elongation of the HIV-1 viral DNA strand, the Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) that directly inhibit RT enzyme by binding to the alloSteric site near the polymerase active site and the Protease Inhibitors (PIs). Highly Active Anti Retroviral Therapy (HAART) regimens, which are based on triple or quadruple combinations of NRTIs, NNRTIs and PIs, reduce HIV to very low levels, but are unable to extricate the infection and long period therapies lead to the emergence of drug resistant mutant strains. Recently the first HIV-integrase inhibitor [6] has been introduced called ISENTRESS (raltegravir) by Merck Pharmaceutical Company. It works by inhibiting the insertion of HIV DNA into human DNA by the integrase enzyme. Inhibiting integrase from performing this essential function limits the ability of the virus to replicate and infect new cells [7]. Another class anti-HIV agents are the Entry Inhibitors they are presently represented only by one drug, Enfuvirtide, but many other compounds are in the process of clinical development [8].

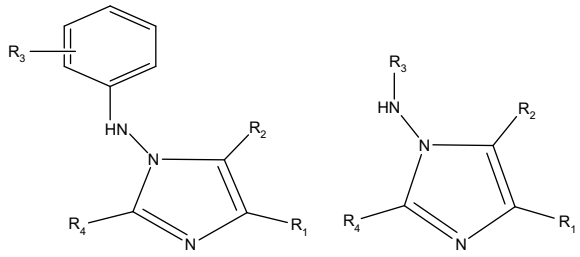
AIDS is one of the most fatal disorders for which no complete and successful chemotherapy has been developed so far. It is strongly desired to develop new anti-HIV-1 agents with superior efficacy and safety profiles. Under Anti-HIV screening program in Laboratory of Medicinal Chemistry and Laboratory of Virology and Chemotherapy, Rega Institute for Medical Research, Belgium, compounds with safety index over five were considered Anti-virally active [9] during preliminary screening and considered for further testing. Safety index is defined as the ratio of CC50/IC50, where CC50 is the Cytotoxic Concentration or concentration that reduced the MT-4 cell viability by 50% and IC50 is 50% inhibitory concentration or concentration required for viral cytopathic effect by 50%. Cytotoxic potential of any antiviral molecule thus bears valuable information about its specificity. Many groups [10,11] have performed a traditional 2D-QSAR analysis on this data set but 3D-QSAR analyses have never been performed. In this present work we have used CoMFA and CoMSIA methodologies to understand the influence of different physicochemical and structural parameters on cytotoxic activity.

Comparative Molecular Field Analysis (CoMFA) [12] and Comparative Molecular Similarity Indices Analysis (CoMSIA) [13] are two commonly used 3D-QSAR methodologies that have been employed in the current work to recognize pharmacophoric units that are responsible for cytotoxic activity. The CoMFA method calculates the energies of Steric and Electrostatic interactions between the compound and the probe atom kept at the various intersections of a regular 3D-lattice according to Lennard–Jones

* Corresponding author. Tel.: +91 9430353750.
E-mail address: arijit4uin@gmail.com (A. Basu).

0223-5234/$ – see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.09.043

Table 1
Molecular structure along with their respective Cytotoxic activity (pCC_{50}).



Structure No. - 37

Structure no.	R^1	R^2	R^3	R^4	Actual activity (pCC_{50} [μM])
1	CH_3	C_6H_5	3-Cl	SH	4.176
2	CH_3	C_6H_5	2-Cl	SH	3.791
3	CH_3	C_6H_5	4-F	SH	3.657
4	CH_3	C_6H_5	4-Cl	SH	3.871
5	CH_3	3-BrC_6H_4	3-Cl	SH	4.357
6	CH_3	4-BrC_6H_4	3-Cl	SH	4.479
7[a,b]	CH_3	3-ClC_6H_4	3-Cl	SH	4.509
8	CH_3	4-ClC_6H_4	3-Cl	SH	4.389
9[a,b]	CH_3	4-$CH_3OC_6H_5$	3-Cl	SH	4.412
10[b]	C_6H_5	CH_3	3-Cl	SH	4.301
11	$-(CH_2)_4-$		3-Cl	SH	3.898
12	CH_3	C_6H_5	H	SH	3.718
13[a,b]	CH_3	C_6H_5	3,4-CH_3	SH	4.255
14	CH_3	C_6H_5	3-Br	SH	4.354
15	CH_3	C_6H_5	3-Cl-4-CH_3	SH	4.354
16[a]	CH_3	C_6H_5	2,5-Cl	SH	3.832
17	CH_3	C_6H_5	3-NO_2	SH	4.316
18[a,b]	CH_3	C_6H_5	3-CH_3	SH	4.054
19	CH_3	CH_3	3-Cl	SH	3.586
20	CH_3	CH_3	H	SH	3.743
21[a,b]	CH_3	CH_3	3-CH_3	SH	3.444
22[a,b]	$CH(CH_3)_2$	C_6H_5	3-CH_3	SH	4.370
23	C_2H_5	C_6H_5	3-Cl	SH	4.409
24	C_2H_5	C_6H_5	3-CH_3	SH	4.271
25	C_6H_5	C_6H_5	3-Cl	SH	4.440
26	CH_3	CH_3OCO	3-Cl	SH	3.721
27	CH_3	COOH	3-Cl	SH	3.576
28	$OCOCH_3$	CH_3	3-Cl	SH	4.312
29[a,b]	CH_3	C_6H_5	3,5-CH_3	SH	3.712
30	CH_3	C_6H_5	3-CH_3O	SH	4.420
31	CH_3	3-CNC_6H_4	3-Cl	SH	4.003
32	CH_3	3-CNC_6H_4	3-CH_3	SH	4.341
33	CH_3	3-$CH_3OCOC_6H_4$	3-Cl	SH	4.172
34[a,b]	CH_3	3-$HOCOC_6H_4$	3-Cl	SH	3.921
35[a,b]	CH_3	3-$HOCOC_6H_4$	3-CH_3	SH	3.728
36	CH_3	3-$NH_2C_6H_4$	3-Cl	SH	4.375
37	CH_3	C_6H_5	α-naphthyl	SH	4.920
38[a,b]	CH_3	C_6H_5	4-C_2H_5	SH	4.463
39	CH_3	C_6H_5	4-CH_3S	SH	4.275
40	CH_3	C_6H_5	3-Cl	H	4.757
41	CH_3	3-BrC_6H_4	3-Cl	H	4.785
42[a,b]	CH_3	3-ClC_6H_4	3-Cl	H	4.923
43	$-(CH_2)-$		3-Cl	H	4.172
44	CH_3	C_6H_5	H	H	4.283
45	CH_3	CH_3	H	H	3.454
46	CH_3	CH_3	3-Cl	H	3.670
47	CH_3	CH_3	3-CH_3	H	3.506
48	CH_3	C_6H_5	3-CH_3	H	4.568
49	CH_3	C_6H_5	4-F	H	4.511
50	C_2H_5	C_6H_5	3-CH_3	H	4.555
51	CH_3	CH_3OCO	3-Cl	H	3.972
52[a]	CH_3	C_6H_5	3,5-CH_3	H	4.690
53	CH_3	C_6H_5	3-CH_3O	H	4.463
54	CH_3	3-CNC_6H_4	3-Cl	H	4.434
55[a,b]	CH_3	3-CNC_6H_4	3-CH_3	H	4.350
56[a,b]	CH_3	3-$NH_2C_6H_4$	3-Cl	H	3.584
57	CH_3	3-$NH_2C_6H_4$	3-CH_3	H	3.789
58	CH_3	3-$CH_3OCOC_6H_4$	3-Cl	H	4.480
59[a,b]	CH_3	3-$OCOHC_6H_4$	3-Cl	H	3.794
60	CH_3	C_6H_5	3-F	SH	3.964

Others – training set molecules.
[a] Test set molecules when 10% molecules are employed as test set.
[b] Test set molecules when 25% molecules are employed as test set.

Table 2
Statistical results of CoMFA and CoMSIA models.

	CoMFA				CoMSIA							
	Database alignment		Field-fit alignment		SEDA		SEA		SED		SHE	
	(10% test set)	(25% test set	(10% test set)	(25% test set	(10% test set)	(25% test set	(10% test set)	(25% test set	(10% test set)	(25% test set	(10% test set)	(25% test set)
q^2	0.735	0.702	0.620	0.640	0.520	0.392	0.685	0.485	0.413	0.413	0.598	0.698
N	10	5	10	5	10	5	10	5	10	5	10	5
R^2	0.984	0.975	0.990	0.972	0.848	0.863	0.932	0.844	0.782	0.913	0.963	0.983
SEE	0.052	0.065	0.041	0.617	0.151	0.141	0.107	0.151	0.181	0.123	0.078	0.058
F	247.02	231.686	412.59	225.65	51.51	37.886	56.33	32.583	30.93	55.815	108.14	179.992
r^2_{bs}	0.993	–	0.996	–	0.958	–	0.959	–	0.963	–	0.986	–
r^2_{cv}	0.743	–	0.620	–	0.620	–	0.677	–	0.488	–	0.589	–
r^2_{pred}	0.653	0.910	0.599	0.352	0.480	−0.101	0.537	0.237	0.439	0.434	0.161	0.789
r^2	0.649	0.898	0.619	0.350	0.477	−.131	0.571	0.247	0.441	0.424	0.172	0.791
r^2_m	−1.05	0.652	0.322	0.268	0.219	–	0.054	0.247	0.135	0.323	−0.080	0.410
Field contribution												
Steric	0.486	0.581	0.512	0.588	0.211	0.123	0.280	0.198	0.277	0.229	0.237	0.253
Electrostatic	0.514	0.489	0.488	0.412	0.289	0.378	0.399	0.462	0.388	0.457	0.399	0.367
Hydrophobic	–	–	–	–	–		–	–	–	–	0.364	0.380
Donor	–	–	–	–	0.241	0.256	–	–	0.336	0.315	–	–
Acceptor	–	–	–	–	0.259	0.243	0.321	0.340	–	–	–	–

q^2 = Cross validation correlation coefficient, N = Number of components, SEE = Standard error of estimate, R^2 = Correlation coefficient, F = F-ratio, r^2_{bs} = Bootstrapped correlation coefficient, r^2_{pred} = Predicted correlation coefficient. r^2 = squared correlation coefficient between observed and predicted values of test set molecules.

and Coulomb potentials. The resulting energies derived from these two potential functions can then be contoured to give a quantitative spatial description of the molecular properties. CoMSIA introduces the Gaussian function for the distance dependence between the molecular atoms and the probe atom in order to avoid some of the inherent deficiencies arising from the Lennard–Jones and Coulomb potential functional forms. CoMSIA is applied to gain insight into how Steric, Electrostatic, Hydrophobic and hydrogen bonding interactions influence the activity.

2. Material and methods

2.1. Data set

The data used in this QSAR study consisted of cytotoxicity data (CC50), the 50% Cytotoxic concentration to reduce MT-4 cell viability. 60 derivatives of 1-Phenylamino-1H-imidazole derivatives that have been reported by Lagoja et al. [9] were employed in the current study. The cytotoxic activity data [CC50 (μM)] for 1-Phenylamino-1H-imidazole derivatives shown in Table 1 were converted to the logarithmic scale [−log CC50 (μM)] and then used for subsequent 3D-QSAR (CoMFA and CoMSIA) analyses. The molecules were divided into two sets. In one set 10% of the molecules are chosen randomly and used as test set molecules. In another case a sorted approach is taken. In this approach we have sorted the compounds according to increasing order of their biological activity and then every fourth molecule is chosen as a test set molecule resulting in a 25% sorted test set.

2.2. Molecular modeling

All the molecular modeling studies, CoMFA and CoMSIA reported herein were performed using SYBYL7.1 molecular modeling software from Tripos, Inc., St. Louis, MO. All the compounds were built from fragments in the SYBYL database. Each structure was fully geometry-optimized using the standard Tripos force field with a distance-dependent dielectric function until a Root Mean Square (RMS) deviation of 0.001 kcal mol^{-1} $Å^{-1}$ was achieved. All the compounds were then subjected to simulated dynamic annealing. The least energy conformations were selected for each compound and further energy minimized using Powell (100 iterations) and Conjugation gradient (10,000 iterations) methods. Gasteiger–Huckel charges were computed for all molecules including the template after energy minimization.

2.3. Alignment procedure

A proper alignment of the structures is critical for obtaining valid 3D-QSAR models. Furthermore, it is vital that all compounds are aligned in a bioactive orientation since the 3D-QSAR model assumes that each structure exhibits activity at the same binding site of the receptor. To obtain a consistent alignment, the lowest energy conformation of compound 37 was used as the template for the alignments as because it has shown the highest cytotoxic activity. The atoms used for alignment were the common substructure of each molecule. An automatic alignment method was carried out by using Database alignment and Field-fit alignment. The molecules were aligned and put into a new database.

2.4. CoMFA and CoMSIA procedures

The standard CoMFA procedure, as implemented in SYBYL7.1, was performed. A 3D cubic lattice with a grid spacing of 2 Å was created automatically by the program to encompass all the aligned ligands. A default sp^3 C probe atom with a Van-Der Waals radius of 1.52 Å and a charge of +1.0 was used to generate Steric (Lennard–Jones 6–12 potential) field energies and Electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. The computed field energies were truncated to 30 kcal/mol for both Steric and Electrostatic fields. CoMSIA is a technique in which similarity indices are calculated at different points on a regularly spaced grid for pre-aligned molecules. In this approach, five different similarity fields are calculated: Steric, Electrostatic, Hydrophobic, Hydrogen bond Donor and Hydrogen bond Acceptor. These fields were selected to cover the major contributions to ligand binding. In CoMSIA fields, singularities were avoided at atomic positions because a Gaussian type distance dependence of each physicochemical property was adopted and thus no arbitrary cutoffs were necessary. The attenuation factor was set to the default value of 0.3.

2.5. Partial least squares

To form the basis for a statistically significant model, a Partial Least Squares Regression was used to generate a linear

Table 3
Comparison of actual and predicted values (pCC_{50}) of N-aminoimidazole derivatives (10%[a] and 25%[b] as test sets) for CoMFA models.

Structure no.	Actual activity (pCC_{50})	CoMFA			
		Database alignment		Field-fit alignment	
		Predicted (10% test set)	Predicted (25% test set)	Predicted (10% test set)	Predicted (25% test set)
1	4.176	4.205	4.215	4.156	4.158
2	3.791	3.789	3.796	3.774	3.785
3[a]	3.657	3.979	3.989	3.955	3.965
4[a]	3.871	4.074	3.869	4.004	3.941
5	4.357	4.298	4.298	4.289	4.29
6	4.479	4.471	4.470	4.514	4.414
7[a,b]	4.509	4.283	4.450	4.096	4.796
8	4.389	4.385	4.385	4.384	4.380
9[b]	4.412	4.410	4.561	4.372	4.797
10[b]	4.301	4.296	4.290	4.306	4.560
11	3.898	3.906	3.926	3.872	3.862
12	3.718	3.778	3.778	3.823	3.723
13[b]	4.255	4.241	4.111	4.252	4.521
14	4.354	4.268	4.258	4.326	4.360
15	4.354	4.438	4.43	4.361	4.461
16[b]	3.832	3.782	3.782	3.816	3.822
17	4.316	4.343	4.343	4.299	4.367
18[b]	4.054	4.011	3.939	4.022	3.753
19	3.586	4.057	4.057	4.179	3.579
20	3.743	3.692	3.752	3.724	3.724
21[b]	3.444	3.476	3.560	3.420	3.020
22[b]	4.370	4.309	4.272	4.359	4.590
23	4.409	4.322	4.422	4.379	4.410
24	4.271	4.318	4.278	4.299	4.302
25	4.440	4.496	4.561	4.421	4.410
26	3.721	3.701	3.717	3.736	3.716
27	3.576	3.617	3.579	3.571	3.575
28	4.312	4.333	4.431	4.341	4.311
29[a,b]	3.712	4.147	3.819	3.517	4.174
30	4.420	4.430	4.473	4.417	4.420
31	4.003	4.045	4.015	4.068	4.006
32	4.341	4.341	4.351	4.342	4.350
33	4.172	4.272	4.185	4.271	4.169
34[b]	3.921	4.004	4.009	3.946	3.432
35[b]	3.728	3.709	3.690	3.740	3.641
36	4.375	4.372	4.371	4.431	4.351
37	4.942	4.349	4.951	4.438	4.838
38[a,b]	4.463	4.354	4.780	4.154	4.254
39	4.275	4.351	4.251	4.268	4.258
40	4.757	4.746	4.760	4.870	4.800
41	4.785	4.796	4.794	4.751	4.767
42[b]	4.923	4.826	4.853	4.825	4.425
43	4.172	4.214	4.312	4.159	4.160
44	4.283	4.369	4.269	4.327	4.270
45	3.454	3.395	3.512	3.447	3.440
46	3.670	3.689	3.714	3.676	3.665
47	3.506	3.515	3.525	3.515	3.505
48	4.568	4.639	4.591	4.625	4.651
49	4.511	4.482	4.489	4.533	4.573
50	4.555	4.551	4.560	4.575	4.575
51	3.972	3.980	3.969	3.964	3.954
52[b]	4.690	4.710	4.799	4.672	4.871
53	4.463	4.418	4.458	4.439	4.525
54	4.434	4.471	4.439	4.418	4.428
55[b]	4.350	4.317	4.251	4.354	4.734
56	3.584	3.592	3.582	3.587	3.857
57	3.789	3.772	3.797	3.786	3.801
58	4.480	4.469	4.465	4.447	4.449
59[a,b]	3.794	3.786	3.721	3.989	4.289
60[a]	3.964	3.959	3.959	3.868	3.883

[a] Test set molecules when 10% molecules are employed as test set.
[b] Test set molecules when 25% molecules are employed as test set.

relationship that correlates changes in the computed fields with changes in the corresponding experimental values of the Cytotoxic activity [−log CC50 (μM)] for the data set of 1-Phenylamino-1H-imidazole derivatives. Cytotoxicity activity values of the data set were used as dependent variables in a PLS statistical analysis of all of the models. The column filtering value(s) was set to 2.0 kcal/mol to improve the signal-to-noise ratio by omitting those lattice points whose energy variations were below this threshold. Cross-validations were performed by the Leave-One-Out (LOO) procedure to determine the optimum number of components (*N*). The optimum number of components obtained (10 for models built using 10% molecules in the test set and 5 for models built using 25% molecules in the test set) then used to derive the final QSAR models.

Table 4
Comparison of actual and predicted values (pCC_{50}) of N-aminoimidazole derivatives (10%[a] and 25%[b] as test sets) for CoMSIA models.

Structure no.	Actual activity (pCC_{50})	CoMSIA							
		SEDA		SEA		SED		SHE	
		Predicted (10% test set)	Predicted (25% test set)	Predicted (10% test set)	Predicted (25% test set)	Predicted (10% test set)	Predicted (25% test set)	Predicted (10% test set)	Predicted (25% test set)
1	4.176	4.254	4.224	4.222	4.439	4.244	4.277	4.224	4.172
2	3.791	3.743	3.686	3.707	3.802	3.735	3.866	3.740	3.8
3[a]	3.657	4.126	3.874	4.083	3.826	4.131	3.957	4.091	3.672
4[a]	3.871	4.133	3.845	4.093	3.746	4.126	3.900	4.581	3.867
5	4.357	4.320	4.232	4.334	4.192	4.213	4.271	4.241	4.384
6	4.479	4.330	4.403	4.327	4.394	4.295	4.500	4.466	4.451
7[a,b]	4.509	4.269	4.197	4.281	4.321	4.165	4.143	4.440	4.441
8	4.389	4.290	4.378	4.296	4.348	4.257	4.440	4.373	4.391
9[b]	4.412	4.501	4.145	4.490	4.560	4.557	4.301	4.457	4.538
10[b]	4.301	4.311	4.567	4.320	3.842	4.317	3.972	4.318	4.431
11	3.898	3.940	3.774	3.953	3.830	3.961	3.751	3.902	3.851
12	3.718	4.012	4.076	3.966	4.162	3.977	4.158	3.875	3.728
13[b]	4.255	4.274	3.888	4.282	4.118	4.238	3.950	4.302	4.349
14	4.354	4.254	4.408	4.222	4.505	4.235	4.396	4.270	4.356
15	4.354	4.402	4.112	4.412	4.158	4.422	4.228	4.424	4.359
16	3.832	3.764	3.683	3.738	3.851	3.782	3.410	3.873	3.842
17	4.316	4.280	4.348	4.291	4.237	4.312	4.310	4.306	4.325
18[b]	4.054	4.134	4.603	4.100	4.436	4.068	3.871	4.101	4.114
19	3.586	3.487	4.206	3.574	3.891	3.562	3.956	3.612	3.579
20	3.743	3.697	3.769	3.693	3.726	3.706	3.684	3.629	3.751
21[b]	3.444	3.356	3.379	3.361	3.004	3.463	3.810	3.399	3.650
22[b]	4.370	4.475	4.295	4.444	4.672	4.441	4.512	4.411	4.622
23	4.409	4.303	4.483	4.267	4.347	4.322	4.406	4.355	4.406
24	4.271	4.397	4.358	4.397	4.336	4.335	4.180	4.350	4.269
25	4.440	4.513	4.255	4.463	4.190	4.434	4.445	4.493	4.434
26	3.721	3.681	4.402	3.666	4.461	3.802	3.590	3.739	3.731
27	3.576	3.605	3.733	3.571	3.613	3.654	3.593	3.535	3.611
28	4.312	4.250	3.585	4.266	4.461	4.272	4.442	4.350	4.294
29[a,b]	3.712	4.208	3.687	4.208	3.943	4.180	3.45	4.198	3.887
30	4.420	4.517	3.628	4.467	4.512	4.515	4.551	4.371	4.510
31	4.003	4.023	4.544	4.061	4.325	4.063	4.017	3.993	4.013
32	4.341	4.322	4.108	4.328	4.212	4.326	4.260	4.338	4.348
33	4.172	4.294	4.338	4.366	4.325	4.311	4.253	4.331	4.200
34[b]	3.921	3.959	3.819	3.993	4.281	3.967	3.812	3.922	4.110
35[b]	3.728	3.741	4491	3.779	3.561	3.685	4.100	3.741	3.819
36	4.375	4.435	3.649	4.416	4.398	4.357	4.517	4.316	4.381
37	4.942	4.342	4.323	4.376	4.779	4.366	4.438	4.297	4.933
38[a,b]	4.463	4.465	4.117	4.478	3.711	4.492	4.78	4.344	4.370
39	4.275	4.347	4.473	4.349	4.530	4.305	4.59	4.282	4.315
40	4.757	4.691	4.177	4.685	4.240	4.753	4.554	4.674	4.62
41	4.785	4.701	4.465	4.729	4.470	4.664	4.500	4.908	4.791
42[b]	4.923	4.701	4.275	4.719	4.784	4.675	4.61	4.801	4.822
43	4.172	4.016	3.892	4.121	3.901	4.096	4.168	4.174	4.202
44	4.283	4.464	4.308	4.445	4.285	4.479	4.261	4.331	4.257
45	3.454	3.518	3.600	3.570	3.648	3.441	3.676	3.471	3.514
46	3.670	3.713	3.710	3.768	3.805	3.696	3.792	3.731	3.7
47	3.506	3.517	3.551	3.491	3.449	3.478	3.542	3.534	3.514
48	4.568	4.513	4.383	4.510	4.380	4.508	4.469	4.531	4.569
49	4.511	4.531	4.646	4.517	4.654	4.590	4.704	4.496	4.521
50	4.555	4.466	4.587	4.408	4.537	4.430	4.690	4.468	4.557
51	3.972	3.961	4.119	3.959	4.018	3.917	3.661	3.909	3.972
52[b]	4.690	4.537	3.956	4.565	4.210	4.568	4.321	4.643	4.232
53	4.463	4.666	4.473	4.652	4.430	4.712	4.412	4.599	4.470
54	4.434	4.505	4.355	4.531	4.405	4.569	4.421	4.462	4.361
55[b]	4.350	4.322	4.217	4.345	4.289	4.321	3.852	4.308	4.556
56	3.584	4.654	3.929	4.685	4.121	3.654	3.681	3.778	3.564
57	3.789	3.718	3.903	3.618	3.765	3.722	3.723	3.674	3.789
58	4.480	4.445	4.477	4.462	4.480	4.493	4.398	4.448	4.483
59[a,b]	3.794	3.745	4.473	3.788	3.690	3.882	3.681	3.898	3.713
60[a]	3.964	3.955		3.978	4.018	3.889	3.958	4.005	3.962

SEDA = Steric + Electrostatic + Hydrogen bond Donor + Hydrogen bond Acceptor; SEA = Steric + Electrostatic + Hydrogen bond Acceptor; SED = Steric + Electrostatic + Hydrogen bond Donor; SHE = Steric + Electrostatic + Hydrophobic.

[a] Test set molecules when 10% molecules are employed as test set.

[b] Test set molecules when 25% molecules are employed as test set.

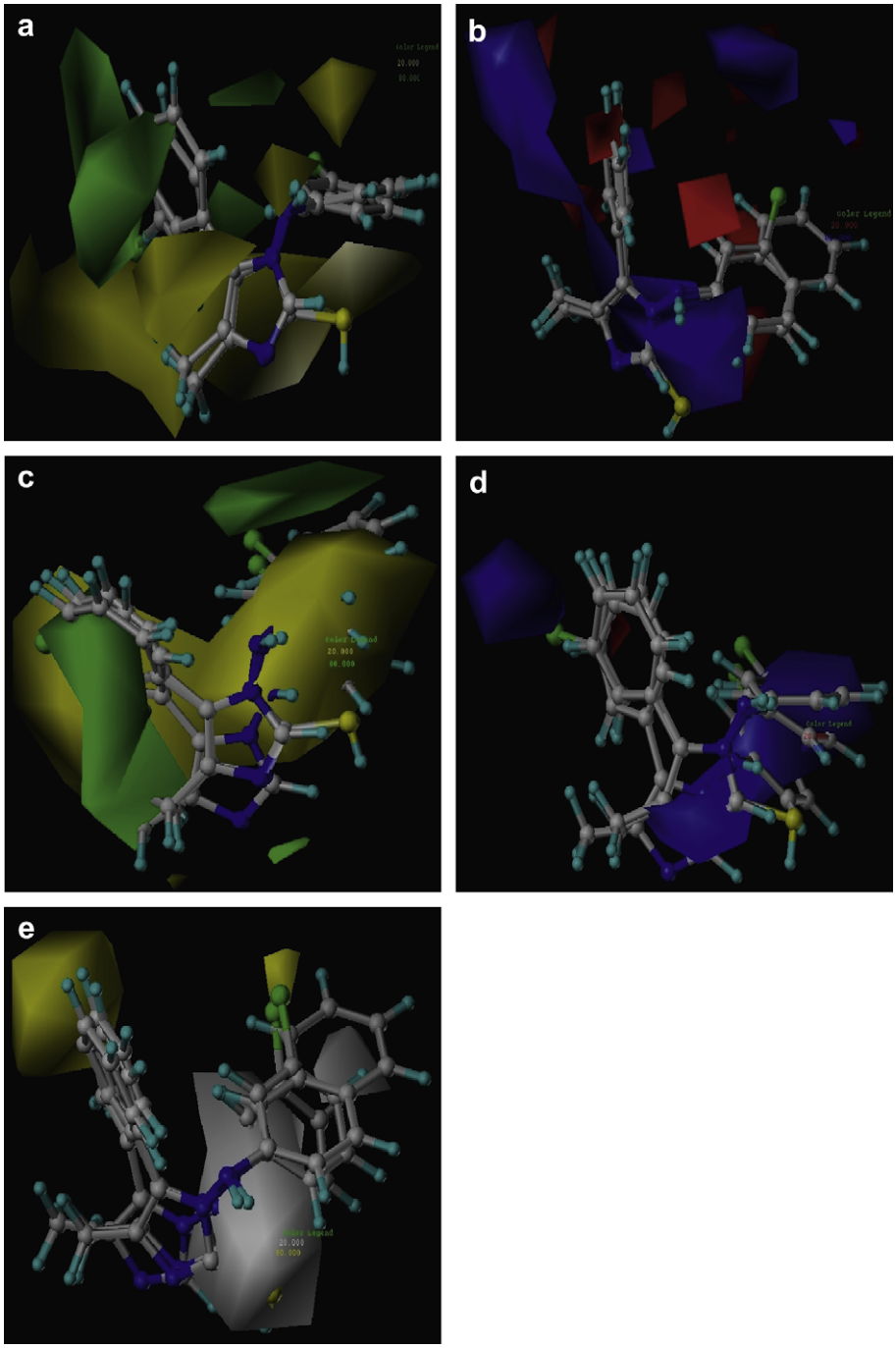


Fig. 1. Contour plots for the best CoMFA and CoMSIA models. (a) and (b) Steric and electrostatic field maps respectively for best CoMFA model obtained using Database alignment built by employing 25% of the molecules in the test set. (c)–(e) Steric, Electrostatic and Hydrophobic field maps for the best CoMSIA obtained using Database alignment built by employing 25% of the molecules in the test set and by considering the SHE fields. For steric field, sterically favored regions are in green, sterically disfavored regions are in yellow. For electrostatic field, red regions indicate that the negative potential is favored, blue region shows that positive potential is favored. For hydrophobic field, the yellow regions indicate favorable hydrophobic interactions and white regions indicate unfavorable hydrophobic interactions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.6. External validation

All the QSAR models are subjected to rigorous external validation process. Several statistics are employed [14] like r^2_{pred}, r^2 (squared correlation coefficient between observed and predicted values) and r^2_{m}.

r^2_{m} is obtained according to the following formula:

$$r^2_{\text{m}} = r^2\left(1 - \left|\sqrt{r^2 - r^2_0}\right|\right)$$

$r^2_0 = r^2$ with intercept set at zero, r^2 = squared correlation coefficient between observed and predicted values.

3. Results and discussion

3.1. CoMFA models

CoMFA Models were built by keeping both 10% and 25% of molecules in test set. The detail statistics are presented in Table 2. The models were validated based on both internal as well as

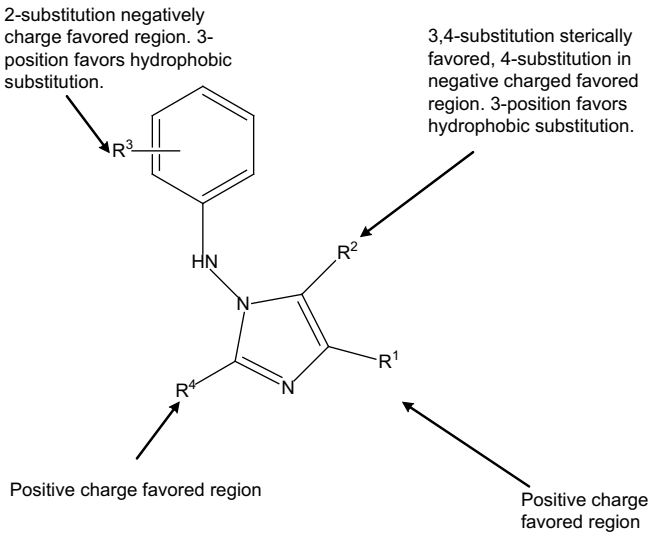


Fig. 2. Favorable and disfavorable interactions of several groups for Cytotoxic potential.

external predictions. CoMFA models built using Database alignment have shown better external predictivity with the model that was built using 25% of molecules in test set (r^2_{pred} of 0.91 and r^2_m of 0.652) than 10% of molecules in test set (r^2_{pred} of 0.653 and r^2_m of −1.05). Whereas the CoMFA models built using Field-Fit alignment have shown better external predictivity with the model that was built using 10% of molecules in test set (r^2_{pred} of 0.599 and r^2_m of 0.322) than 25% of molecules in test set (r^2_{pred} of 0.352 and r^2_m of 0.268). Therefore for this particular analysis it has been seen that changing or shuffling the population of test and training sets, has a significant influence on the external predictivity though internal prediction remains consistent in all the cases. In general models that were built using Database alignment were superior as compared to models built using Field-Fit alignment also increasing the population of the test set increases the external predictivity in case of Database alignment. The contributions of Steric and Electrostatic fields are almost similar in all the cases. Table 3 shows the actual versus predicted values of CoMFA models.

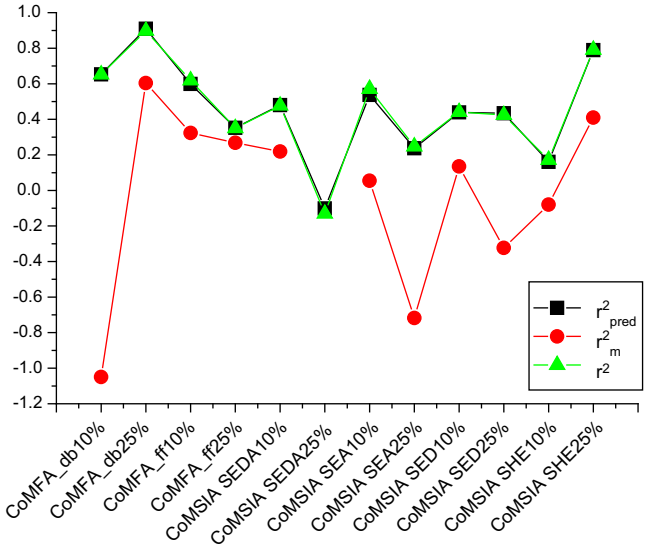


Fig. 3. Comparative study of different correlation coefficients used for external validation. r^2 (squared correlation coefficient between observed and predicted values), r^2_{pred} (predicted r^2). CoMFA_db10% (CoMFA with Database alignment using 10% molecules in test set), CoMFA_db25% (CoMFA with Database alignment using 25% molecules in test set), CoMFA_ff10% (CoMFA with Field-fit alignment using 10% molecules in test set), CoMFA_ff 25% (CoMFA with Database alignment using 25% molecules in test set), CoMSIA SEDA10% (CoMSIA with Database alignment using 10% molecules in test set and using SEDA fields), CoMSIA SEDA25% (CoMSIA with Database alignment using 25% molecules in test set and using SEDA fields), CoMSIA SEA10% (CoMSIA with Database alignment using 10% molecules in test set and using SEA fields), CoMSIA SEA25% (CoMSIA with Database alignment using 25% molecules in test set and using SEA fields), CoMSIA SED10% (CoMSIA with Database alignment using 10% molecules in test set and using SED fields), CoMSIA SED25% (CoMSIA with Database alignment using 25% molecules in test set and using SED fields), CoMSIA SHE10% (CoMSIA with Database alignment using 10% molecules in test set and using SHE fields), CoMSIA SHE25% (CoMSIA with Database alignment using 10% molecules in test set and using SHE fields).

3.2. CoMSIA models

Database alignment which gave the best CoMFA results was employed to build all the CoMSIA models. CoMSIA Models were also built by keeping both 10% and 25% of molecules in test set. PLS analyses using different fields, Steric (S), Electrostatic (E), Hydrophobic (H), Hydrogen bond Donor (D) and Hydrogen bond Acceptor (A) were employed. Using five different fields we developed four different CoMSIA models: SEDA, SEA, SED and SHE each for 10% and 25% of molecules in test set respectively. Amongst models that were built using 25% of the molecules in test set SHE produced the best model (r^2_{pred} of 0.789 and r^2_m of 0.410) whereas amongst models that were built using 10% of the molecules in test set SEA produced the best model (r^2_{pred} of 0.537 but poor r^2_m of 0.054). In SED model not much of a difference is observed by changing the population size of the test set (r^2_{pred} of 0.439 and 0.434 for 10% and 25% test sets respectively whereas their r^2_m was found to be 0.135 and 0.323 respectively). In general the SHE model with 25% molecule in the test set is superior amongst all the CoMSIA models for both internal as well as external predictions. However r^2_m value of 0.410 is unacceptable and therefore it may not be successfully extrapolated for predicting the activity of molecules which are not used for building the model. The field contributions for this model are Steric = 0.253, Electrostatic = 0.367, Hydrophobic = 0.380. This is suggesting that the Hydrophobicity of the molecules influences the cytotoxic potential of the molecules. Table 4 shows the actual versus predicted values of CoMSIA models for training and test sets. All the statistical parameters obtained from the CoMSIA models are presented in Table 2.

3.3. Contour maps

The contour maps of CoMFA denote the region in the space where the aligned molecules would favorably or unfavorably interact with the receptor while the CoMSIA contour maps denote these areas within the specified region where the presence of a group with a particular physicochemical property binds to the receptor [15]. The CoMFA/CoMSIA results were graphically interpreted by field contribution maps using the 'STDEV*COEFF' field type.

Fig. 1(a–e) shows the contour maps derived from the best CoMFA and CoMSIA PLS models. The best four analogues, compounds 37, 40, 41 and 42 were embedded in the maps. All of the contours represented the default 80 and 20% level contributions for favored and disfavored regions. From the contour plots the Steric and Electrostatic fields of both the CoMFA and CoMSIA models are similar and consistent. For both CoMFA and CoMSIA models the R^1, R^3 and R^4 substituents have no Steric contribution however R^2 substitution especially at 3 and 4 position of the phenyl ring strongly sterically favored for cytotoxicity. For Electrostatic maps of both CoMFA and CoMSIA models the R^4 substitution with –SH and the Imidazole ring is in positive charge favored region whereas the substitution at 4 position of R^2 and substitution at 2 position of R^3 are in negatively charge favored region. The Hydrophobic fields for the CoMSIA model suggest that the 3 position

substitution at R^3 and R^2 phenyl ring with a Hydrophobic substituent may be favorable for cytotoxicity. Favorable and disfavorable interactions of several groups for Cytotoxic potential are schematically depicted in Fig. 2.

4. Conclusions

CoMFA analyses using Database alignment employing 25% of the molecules in test set provided the best model whereas the model with Steric, Electrostatic and Hydrophobic field gave the best result in the case of CoMSIA analysis. In general the CoMFA model is statistically superior to CoMSIA model. The CoMSIA Steric and Electrostatic field maps are in accordance with field distribution of CoMFA maps and consistent with structure–activity relationships. Test sets that were built employing 25% of the molecules produced better models in most of the cases. Statistics like r^2, r^2_m employed for external validation of the models have helped in identifying the best model from amongst many similar looking models. A comparative depiction of the statistical quality of all the correlation coefficients used for external validation of the models is given in Fig. 3.

In the present study, we have successfully identified the significance of various structural elements responsible for Cytotoxic potential of this class of molecules.

Acknowledgements

The authors are grateful for the financial aid in the form of fellowship by All India Council for Technical Education, New Delhi, India.

References

[1] V. Simon, D.D. Ho, Q.A. Karim, Lancet 368 (2006) 489–504.
[2] http://www.unaids.org/Unaids/EN/Resources/Publications.
[3] E. De Clercq, J. Med. Chem. 38 (1995) 2491–2517.
[4] M. Artico, Farmaco 51 (1996) 305–331.
[5] E. De Clercq, Mini-Rev. Med. Chem. 2 (2002) 163–175.
[6] http://www.merck.com/newsroom/press_releases/product/2007_1012.html.
[7] S. Dubey, Y.D. Satyanarayana, H. Lavania, Eur. J. Med. Chem. 42 (2007) 1159–1168.
[8] V. Briz, E. Poveda, V. Soriano, J. Antimicrob. Chemother. 57 (2006) 619–627.
[9] I.M. Lagoja, C. Pannecouque, A.V. Aerschot, M. Witvrouw, Z. Debyser, J. Balzarini, P. Hardewijn, E. De Clercq, J. Med. Chem. 46 (2003) 1546.
[10] K. Roy, J.T. Leonard, Bioorg. Med. Chem. 14 (2006) 1039–1046.
[11] M. Arab Chamjangali, M. Beglari, G. Bagherian, J. Mol. Graphics Modell. 26 (2007) 360–367.
[12] Richard D. Cramer, David E. Patterson, Jeffrey D. Bunce, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[13] G. Klebe, U. Abraham, T. Mietzner, J. Med. Chem. 37 (1994) 4130–4146.
[14] P.P. Roy, K. Roy, QSAR Comb. Sci. 27 (2008) 302–313.
[15] Hongyu Cao, Huabei Zhang, Xuefangf Zheng, Dabin Gao, J. Mol. Graphics Modell. 26 (2007) 236–245.