QSAR Analyses of 3-(4-Benzylpiperidin-1-yl)-*N*-phenylpropylamine Derivatives as Potent CCR5 Antagonists

Kunal Roy* and J. Thomas Leonard

Drug Theoretics & Cheminformatics Lab, Division of Medicinal & Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Received May 20, 2005

CCR5 receptor binding affinity of a series of 3-(4-benzylpiperidin-1-yl)propylamine congeners was subjected to QSAR study using the linear free energy related (LFER) model of Hansch. Appropriate indicator variables encoding different group contributions and different physicochemical variables such as hydrophobicity (π), electronic (Hammett σ), and steric (molar refractivity, STERIMOL values) parameters of phenyl ring substituents of the compounds were used as predictor variables. The Hansch analysis explores the importance of the lipophilicity and electron-donating substituents for the binding affinity. However, this method could not give more insight into the structure−activity relationships because of the diverse molecular features in the data set. 3D-QSAR analyses of the same data set using Molecular Shape Analysis (MSA), Receptor Surface Analysis (RSA), and Molecular Field Analysis (MFA) techniques were also performed. The best model with acceptable statistical quality was derived from the MSA, which showed the importance of the relative negative charge (RNCG): substituents with a high RNCG value have more binding affinity than the unsubstituted piperidine and phenyl (R_1 position) congeners. The relative negative charge surface area (RNCS) is detrimental (e.g. R_2 = 3,4-Cl_2) for the activity. An increase in the length of the molecule in the Z dimension (Lz) is conducive (e.g. R_3 = sulfonylmorpholino), while an increase in the area of the molecular shadow in the XZ plane (Sxz) is detrimental (e.g. R_1 = *N*-*c*-hexylmethyl-5-oxopyrrolidin-3-yl) for the binding affinity. The presence of a chiral center makes the molecule less active (e.g. R_1 = *N*-methyl-5-oxopyrrolidin-3-yl). An increase in the van der Waals area, the molecular volume, and the difference between the volume of the individual molecule and the shape reference compound are conducive (e.g. $R_3 = (CH_3)_2NSO_2-$) for the binding affinity. Substituents with higher JursFPSA_2 values (fractional charged partial surface area) like the *N*-methylsulfonylpiperidin-4-yl (R_1 position) group have better binding affinity than the substituents such as 4-chlorophenylamino (R_1 position). Unsubstituted piperidines (R_1 position) with less JursFNSA_1 values have lower binding affinity than the 4-chlorophenyl substituted compounds. The MFA derived equation shows interaction energies at different grid points, while the RSA model shows the importance of hydrophobicity and charge at different regions of the molecules. The models were validated through the leave-one-out, leave-15%-out, and leave-25%-out cross-validation techniques. The developed models were also subjected to a randomization test (99% confidence level). Although the MSA derived models had excellent statistical qualities both for the training as well as test sets, RSA and MFA results for the test sets are not comparable statistically with the MSA derived models.

1. INTRODUCTION

Acquired immunodeficiency syndrome (AIDS) is the fatal disorder for which no complete and successful chemotherapy has been developed so far. Human immunodeficiency virus subtype 1 (HIV-1), a retrovirus of the lentivirus family, has been found to be prevalent in causing this disease. HIV-1 produces a progressive immunosuppression by destruction of CD4+ T lymphocytes ("helper" cells, which lead attack against infections) and macrophages and results in opportunistic infections, neurological and neoplastic diseases, and death.[1] The global HIV/AIDS epidemic[2] has claimed more than 3 million lives in 2004, and an estimated 5 million people acquired HIV in the same year, bringing to 40 million (including 2.5 million children under 15 years of age) the number of people globally living with HIV/AIDS. The pandemic spread of this disease has prompted an unprecedented scientific and clinical effort to understand and combat it.

The replicative cycle of HIV can be divided into entry and post entry steps.[3,4] Entry of the HIV into a target cell consists of three vital steps: (1) The trimeric HIV-1 envelope glycoprotein complex mediated viral entry into susceptible target cells: the surface subunit (gp120) attaches to the receptor (CD4); (2) gp120-co-receptor (CXCR4 or CCR5) interaction, which results in the exposure of a co-receptor-binding domain in gp120 on the cell surface; (3) and subsequent conformational changes within the Env complex which lead to membrane fusion mediated by the transmembrane subunit (gp41). A number of potential sites for therapeutic intervention thus become accessible during the narrow window between virus attachment and the subsequent fusion of a viral envelope with the cell membrane.[5] Specific compounds that may be developed as novel types of antiretrovirals may target all steps in the process of HIV entry into the cell. The recent approval of fusion inhibitor

* Corresponding author phone: +91-33-28670786; e-mail: kunalroy_in@yahoo.com; URL: http://www.geocities.com/kunalroy_in.

10.1021/ci050205x CCC: $30.25
Published on Web 08/19/2005

enfuvirtide (FUZEON, formerly known as T-20) for the commercial market therefore provides a much-needed newer class of drugs for the treatment of HIV, the entry inhibitors. Some of these compounds demonstrated in vitro synergism with other classes of antivirals, offering thus the rationale for their combination in therapies for HIV-infected individuals. Entry inhibitors are expected to have different toxicity and resistance profiles than the existing reverse transcriptase and protease inhibitors.

Post entry steps[6] require the viral reverse transcriptase (RT), integrase, and protease (PR) to complete the viral replication cycle. The virally encoded RT enzyme mediates reverse transcription. RT is a heterodimeric (p51 and p66 subunits) and multifunctional enzyme presenting both RNA and DNA polymerase and RNaseH activities, being responsible for the conversion of the single stranded viral RNA into the double stranded proviral DNA.[1] Reverse transcriptase inhibitors were the first agents approved for the treatment of HIV-1. The viral integrase enzyme is required for the integration of proviral DNA into the host genome before replication. Integrase inhibitors are in clinical trials. When the infected cell synthesizes new protein, integrated proviral DNA is also translated into the protein building blocks of a new viral progeny. Subsequent expression of the virus by the host cells produces the gag and gag-pol proteins Pr44 and Pr160 of HIV-DNA that are processed by the HIV-encoded PR into functional proteins and enzymes. The viral components then assemble on the cell surface and bud out as immature viral particles. The final maturation of newly formed viruses requires the HIV-1 protease to make up an infectious virion. The inhibition of key enzymes, HIV-1 reverse transcriptase and HIV-1 protease, provides the most attractive target for the anti-HIV drug development. Appropriate combinations of these drugs (referred to as highly active antiretroviral therapy or HAART) markedly suppress viral replication in most treated persons, leading to significant restoration of the immune system function. HAART is responsible for dramatic reductions in HIV-associated morbidity and mortality.[7,8] However, the quest for improved therapies continues, because of problems that seriously limit the current HAART regimens, including toxic side effects, viral persistence, and difficulties in adhering to treatment, high cost, and the emergence of drug-resistant escape variants.[9] The compounds generally referred to as entry inhibitors (which include chemokine-receptor inhibitor, co-receptor inhibitor, membrane-fusion inhibitor, and other attachment inhibitors) are expected to have different toxicity and resistance profiles than the existing reverse transcriptase and protease inhibitors.[10] Among various methods of anti-HIV activity screening, some important methods are cyto-protection assay, integration enzyme assay, multinuclear-activation galactosidase indicator (MAGI) assay, RT inhibition assay, HIV attachment assay, fusion assay, cytotoxicity assay, time-of-addition assay, inhibition of HIV-1 transactivation, etc.[11,12]

Predictive pharmacophore models[13] have been developed for the CCR5 antagonists using piperidine and piperazine based compounds. Molecular docking[14] and 3D-QSAR[15] analyses of piperidine-based CCR5 receptor antagonists have also been reported. The present group of authors has developed a QSAR model for the CCR5 antagonist benzylpyrazole[16] derivatives. We have also developed a few quantitative structure−activity relationship (QSAR) models for anti-HIV activities of different groups of compounds, e.g., 2-amino-6-arylsulfonylbenzonitriles,[17,18] alkenyldiarylmethanes,[19] and imidazole derivatives.[20,21] In continuation of such efforts, the present paper deals with the QSAR modeling of CCR5 antagonist 3-(4-benzylpiperidin-1-yl)-*N*-phenylpropylamine derivatives.

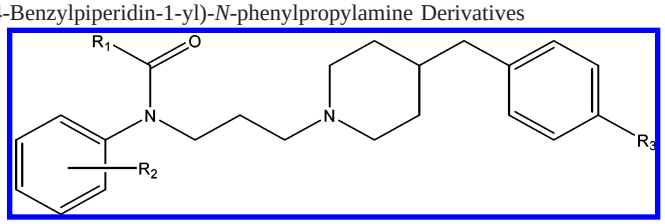
2. MATERIALS AND METHODS

CCR5 binding affinity data reported by Imamura et al.[22−24] have been used for the present QSAR study: the affinity data [K_i(nM)] and [K_i(μM)] of 3-(4-benzylpiperidin-1-yl)-propylamine congeners (Table 1) for ^{125}I-labeled RANTES to Chinese Hamster Ovary (CHO) cells expressing human CCR5 have been converted to the logarithmic scale [pK_i-(mM)] and then used for subsequent QSAR analyses as the response variable. Structural variations are present in three substitution sites of the compounds: the first one is the R_1 position (showing a diverse substitution pattern), the second one is the R_2 position (showing a limited substitution pattern), and the third one is the R_3 position (showing limited structural variations) (Table 1). Some of the compounds reported in the original papers were excluded in the present study because of their nongraded quantitative activity data or the presence of uncommon structural features. The data set was first divided into two subsets based on structural diversity: one training set composed of 70 compounds and one test set composed of 23 compounds (Table 1). All models were constructed based on the training set, and the generated models were then validated using the leave-one-out technique (internal validation) and predicting the activities of the test set (external validation).

The activity data were subjected to Classical QSAR analysis using a linear free energy related (LFER) model of Hansch[25,26] with lipophilicity (π), electronic (Hammett σ), and steric (molar refractivity MR and STERIMOL L, B1 and B5) parameters of the aryl ring substituents along with appropriate dummy parameters as predictor variables. Different indicator and physicochemical variables used in the study have been defined in Table 3. The values of the physiochemical substituent constants (Table 4) were taken from the literature.[27]

Though the classical approach of multiple linear regression (MLR) technique was used as the final statistical tool for developing classical QSAR relations, factor analysis (FA)[28,29] was used as the data-preprocessing step to identify the important predictor variables contributing to the response variable and to avoid collinearities among them. In a typical factor analysis procedure, the data matrix is first standardized, and the correlation matrix and the subsequently reduced correlation matrix are constructed. An eigenvalue problem is then solved, and the factor pattern can be obtained from the corresponding eigenvectors. The principal objectives of factor analysis are to display multidimensional data in a space of lower dimensionality with a minimum loss of information (explaining >95% of the variance of the data matrix) and to extract the basic features behind the data with ultimate goal of interpretation and/or prediction. Factor analysis was performed on the data set(s) containing biological activity and all descriptor variables, which were to be considered. The factors were extracted by the principal component

Table 1. Structural Features of 3-(4-Benzylpiperidin-1-yl)-*N*-phenylpropylamine Derivatives



compd no.	R_1	R_2	R_3	compd no.	R_1	R_2	R_3
				Training Set			
1	*N*-benzylcarbonylpiperidin-4-yl	H	H	**48**	Ph(4-COOH)-NH	H	F
2	piperidin-4-yl	H	H	**49**	Ph(4-$CONH_2$)-NH	H	F
3	*N*-acetylpiperidin-4-yl	H	H	**50**	Ph(4-OCH_3)-NH	H	F
4	*N*-isobutyrylpiperidin-4-yl	H	H	**51**	Ph(4-SCH_3)-NH	H	F
5	*N*-benzoylpiperidin-4-yl	H	H	**54**	Ph(4-Cl)-NH	3-Cl	F
6	*N*-methylsulfonylpiperidin-4-yl	H	H	**55**	Ph(4-Cl)-NH	3,4-Cl_2	F
9	*N*-acetylpiperidin-3-yl	H	H	**57**	Ph(4-Cl)-NH	H	SO_2-morpholino
10	*N*-methylsulfonylpiperidin-3-yl	H	H	**58**	*N*-methyl-5-oxopyrrolidin-3-yl	H	H
12	*N*-acetylpiperidin-4-yl	4-CH_3	F	**59**	*N*-methyl-5-oxopyrrolidin-3-yl	2-CH_3	H
13	*N*-acetylpiperidin-4-yl	3,4-Cl_2	H	**60**	*N*-methyl-5-oxopyrrolidin-3-yl	3-CH_3	H
14	*N*-acetylpiperidin-4-yl	3,4-Cl_2	F	**61**	*N*-methyl-5-oxopyrrolidin-3-yl	4-CH_3	H
15	*N*-acetylpiperidin-4-yl	3-Cl, 4-CH_3	H	**64**	*N*-methyl-5-oxopyrrolidin-3-yl	3-Cl	H
16	*N*-acetylpiperidin-4-yl	3-Cl, 4-CH_3	F	**65**	*N*-methyl-5-oxopyrrolidin-3-yl	4-Cl	H
17	*N*-acetylpiperidin-4-yl	3-Cl, 4-*i*-C_3H_7	F	**66**	*N*-methyl-5-oxopyrrolidin-3-yl	3,4-Cl_2	H
18	*N*-acetylpiperidin-4-yl	3-Cl, 4-OCH_3	F	**68**	*N*-methyl-5-oxopyrrolidin-3-yl	3,4-F_2	H
21	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	CF_3	**71**	*N*-methyl-5-oxopyrrolidin-3-yl	3-CN	H
24	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	$NHCOCH_3$	**72**	*N*-methyl-5-oxopyrrolidin-3-yl	3,4-Cl_2	F
25	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	$NHSO_2CH_3$	**73**	*N*-butyl-5-oxopyrrolidin-3-yl	H	H
28	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SCH_3	**74**	*N*-*c*-hexyl-5-oxopyrrolidin-3-yl	H	H
29	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SO_2CH_3	**75**	*N*-*c*-hexylmethyl-5-oxopyrrolidin-3-yl	H	H
32	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SO_2NH_2	**77**	*N*-benzyl-5-oxopyrrolidin-3-yl	H	H
33	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SO_2NHCH_3	**78**	*N*-(2-chlorobenzyl)-5-oxopyrrolidin-3-yl	H	H
34	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	$SO_2N(CH_3)_2$	**79**	*N*-(3-chlorobenzyl)-5-oxopyrrolidin-3-yl	H	H
35	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SO_2-morpholino	**80**	*N*-(4-chlorobenzyl)-5-oxopyrrolidin-3-yl	H	H
36[a]	*N*-acetylpiperidin-4-yl	3,4-Cl_2	F	**81**	*N*-(4-methylbenzyl)-5-oxopyrrolidin-3-yl	H	H
37[b]	*N*-acetylpiperidin-4-yl	3,4-Cl_2	F	**82**	*N*-benzylmethyl-5-oxopyrrolidin-3-yl	H	H
38	Ph-NH	H	H	**83**	*N*-(furan-2-yl-methyl)-5-oxopyrrolidin-3-yl	H	H
39	Ph(4-Cl)-NH	H	H	**85**	5-oxopyrrolidin-3-yl	H	H
40	Ph(4-Cl)-NH	H	F	**86**	*N*-(1,1,1-trifluoroethyl)-5-oxopyrrolidin-3-yl	H	H
41	Ph(4-F)-NH	H	F	**87**	*N*-(2-methylbenzyl)-5-oxopyrrolidin-3-yl	H	H
42	Ph(4-Br)-NH	H	F	**88**	*N*-benyl-5-oxopyrrolidin-3-yl	3-Cl	H
43	Ph(4-CH_3)-NH	H	F	**89**	*N*-benzyl-5-oxopyrrolidin-3-yl	3,4-Cl_2	H
44	Ph(4-*i*-Pr)-NH	H	F	**90**	*N*-methyl-5-oxopyrrolidin-3-yl	H	F
45	Ph(4-CF_3)-NH	H	F	**91**[d]	*N*-methyl-5-oxopyrrolidin-3-yl	H	F
47	Ph(4-$COOC_2H_5$)-NH	H	F	**92**[a,d]	*N*-methyl-5-oxopyrrolidin-3-yl	H	F
				Test Set			
7	*N*-benzylcarbonylpiperidin-3-yl	H	H	**52**	Ph(4-SO_2CH_3)-NH	H	F
8	piperidin-3-yl	H	H	**53**	Ph(4-Cl)-NH	H	F
11	*N*-acetylpiperidin-4-yl	3-Cl	F	**56**	Ph(4-Cl)-NH	H	SO_2CH_3
19	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	F	**62**	*N*-methyl-5-oxopyrrolidin-3-yl	4-*t*-Bu	H
20	*N*-methylsulfonylpiperidin-4-yl	3-Cl,4-CH_3	F	**63**	*N*-methyl-5-oxopyrrolidin-3-yl	4-OCH_3	H
22	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	NO_2	**67**	*N*-methyl-5-oxopyrrolidin-3-yl	3-Cl,4-F	H
23	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	NH_2	**69**	*N*-methyl-5-oxopyrrolidin-3-yl	3-CF_3	H
26	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	morpholino	**70**	*N*-methyl-5-oxopyrrolidin-3-yl	4-CF_3	H
27	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	OCH_3	**76**	*N*-phenyl-5-oxopyrrolidin-3-yl	H	H
30	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	$SO_2C_2H_5$	**84**	*N*-pyridini-4-yl-methyl-5-oxopyrrolidin-3-yl	H	H
31	*N*-methylsulfonylpiperidin-4-yl	3,4-Cl_2	SO_2-*i*-Pr	**93**[c,d]	*N*-methyl-5-oxopyrrolidin-3-yl	H	F
46	Ph(4-CN)-NH	H	F				

[a] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_4$. [b] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_2$. [c] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_5$. [d] Benzoyl group at the fourth position of the piperidine nucleus.

method and then rotated by VARIMAX rotation (a kind of rotation which is used in principle component analysis so that the axes are rotated to a position in which the sum of the *vari*ances of the loadings is the *max*imum possible) to obtain Thurston's simple structure. The simple structure is characterized by the property that as many variables as possible fall on the coordinate axes when presented in common factor space, so that the largest possible number of factor loadings becomes zero. This is done to obtain a numerically comprehensive picture of the relatedness of the variables. Only variables with nonzero loadings in such factors where biological activity also has nonzero loadings were considered important in explaining the variance of the activity. Further, variables with nonzero loadings in different factors were combined in a multivariate equation.

Along with FA-MLR, PCRA was also tried for the data set. In principle component regression analysis (PCRA),[29] factor scores are used as the predictor variables. PCRA has an advantage that collinearities among X variables are not a disturbing factor and that the number of variables included

Table 2. Observed, Calculated, Predicted, and Cross-Validated Predicted (LOO) Data of CCR5 Antagonists 3-(4-Benzylpiperidin-1-yl)-*N*-phenylpropylamine Derivatives

compd no.	obs[a]	calc[b]	pred[b]	calc[c]	pred[c]	calc[d]	pred[d]	calc[e]	pred[e]	calc [f]	pred [f]
					Training Set						
1	4.319	4.435	4.451	4.255	4.246	4.631	4.693	4.922	4.962	4.269	4.184
2	2.367	3.483	3.525	2.725	2.804	3.557	3.830	4.011	4.079	3.272	3.446
3	4.796	4.435	4.383	4.593	4.567	4.884	4.894	4.376	4.367	4.857	4.825
4	4.409	4.435	4.438	4.319	4.307	5.010	5.090	4.188	4.183	4.955	5.123
5	4.481	4.435	4.428	4.370	4.355	5.144	5.193	4.552	4.554	4.108	3.962
6	4.398	4.435	4.440	4.310	4.299	4.187	4.113	4.263	4.259	4.887	5.005
9	3.167	3.483	3.495	3.238	3.308	3.839	3.928	3.471	3.512	3.705	3.788
10	3.745	3.483	3.473	3.647	3.550	3.620	3.567	4.962	5.022	3.625	3.435
12	5.553	5.355	5.271	5.605	5.617	4.982	4.937	4.795	4.771	5.683	5.676
13	5.721	5.082	5.007	5.450	5.430	5.357	5.314	6.545	6.718	5.670	5.703
14	5.921	5.082	4.984	5.587	5.546	5.440	5.382	5.629	5.604	5.953	5.925
15	6.208	5.930	5.840	6.129	6.120	5.197	5.081	4.466	4.420	5.467	5.106
16	6.538	5.930	5.733	6.359	6.327	5.261	5.152	5.264	5.168	5.984	5.922
17	4.745	4.749	5.478	4.393	3.183	5.340	5.422	5.314	5.353	4.979	5.055
18	4.523	5.545	6.000	5.386	5.545	4.501	4.499	4.623	4.627	5.389	5.583
21	5.060	5.082	5.084	4.981	4.976	5.562	5.656	4.997	4.975	5.556	5.671
24	5.229	5.082	5.065	5.100	5.091	5.050	5.003	4.456	4.441	5.042	4.994
25	5.658	6.104	6.209	6.016	6.061	5.946	6.004	5.638	4.537	5.423	5.320
28	5.509	5.082	5.032	5.299	5.284	5.785	5.860	5.030	4.890	5.546	5.637
29	5.658	6.104	6.209	6.016	6.061	5.454	5.431	5.832	5.907	6.025	6.112
32	5.469	6.104	6.253	5.881	5.932	5.390	5.372	5.644	5.708	5.836	5.916
33	5.824	5.492	5.400	5.761	5.743	5.710	5.688	5.389	5.255	5.238	5.305
34	5.921	6.104	6.147	6.203	6.240	5.945	5.950	5.743	5.710	5.133	5.154
35	5.886	5.492	5.383	5.804	5.781	6.291	6.425	5.444	5.396	5.318	4.817
36[g]	4.284	4.469	4.536	4.053	3.985						
37[h]	3.319	4.469	4.887	3.369	3.388						
38	4.745	4.607	4.597	4.703	4.700	4.386	4.349	4.496	4.491	3.580	3.304
39	5.229	4.607	4.564	5.046	5.032	4.814	4.779	4.722	4.709	4.832	4.810
40	5.108	4.607	4.572	4.956	4.944	4.870	4.850	4.504	4.491	4.590	4.545
41	4.886	4.607	4.587	4.799	4.792	4.530	4.498	4.341	4.327	4.225	4.049
42	4.745	4.607	4.597	4.699	4.695	5.008	5.034	4.644	4.641	4.767	4.797
43	5.18.0	4.607	4.567	5.007	4.993	4.544	4.502	5.067	5.061	5.118	5.136
44	4.119	4.607	4.640	4.254	4.264	4.580	4.606	4.905	4.931	4.444	4.448
45	4.854	4.607	4.590	4.775	4.770	4.450	4.392	4.839	4.833	4.795	4.763
47	3.553	4.607	4.679	3.853	3.880	4.147	4.235	2.488	2.143	4.196	4.261
48	3.456	4.607	4.686	3.784	3.814	3.343	3.293	4.255	4.277	4.172	4.294
49	4.538	4.607	4.611	4.552	4.553	4.491	4.487	4.651	4.661	4.787	4.780
50	4.523	4.607	4.612	4.541	4.542	4.045	3.990	5.106	5.148	4.923	4.984
51	4.824	4.607	4.592	4.754	4.749	4.856	4.858	4.381	4.370	4.409	4.343
54	4.721	5.181	5.291	4.865	4.889	5.272	5.353	4.703	4.702	4.297	4.261
55	4.208	4.607	4.634	4.318	4.326	5.364	5.592	4.644	4.654	4.248	4.230
57	6.000	5.016	4.484	5.831	5.729	5.557	5.518	4.652	4.580	5.985	5.852
58	3.319	3.483	3.489	3.409	3.413	3.253	3.246	4.023	4.051	3.619	3.653
59	2.328	3.483	3.526	2.701	2.729	3.490	3.559	2.687	2.741	3.275	3.482
60	3.796	3.936	3.954	4.025	4.042	3.370	3.336	3.832	3.834	4.392	4.515
61	4.051	4.403	4.594	4.408	4.461	3.548	3.514	4.051	4.051	3.435	3.321
64	3.921	4.057	4.084	3.890	3.885	3.721	3.706	3.850	3.845	4.105	4.121
65	2.921	3.555	3.724	3.273	3.330	3.684	3.746	3.579	3.617	3.216	3.306
66	4.244	4.129	4.111	4.266	4.269	3.902	3.863	3.936	3.922	3.436	3.305
68	2.678	3.692	3.732	2.633	2.627	3.130	3.198	3.051	3.125	2.965	2.993
71	2.387	3.021	3.129	1.892	1.614	3.401	3.471	2.888	2.942	2.865	3.000
72	4.301	4.129	4.101	4.303	4.303	3.973	3.925	3.354	3.306	4.585	4.711
73	3.886	3.483	3.467	3.810	3.807	3.826	3.821	3.777	3.769	4.422	4.471
74	3.959	3.483	3.464	3.862	3.857	3.974	3.976	4.260	4.268	3.701	3.488
75	3.638	3.483	3.477	3.635	3.634	3.564	3.554	4.330	4.351	2.815	2.538
77	4.420	3.483	3.447	4.189	4.176	3.563	3.507	4.045	4.031	4.387	4.406
78	4.481	3.483	3.445	4.233	4.218	3.978	3.939	4.045	4.029	4.321	4.366
79	4.066	3.483	3.460	3.938	3.932	3.861	3.842	4.045	4.044	3.796	3.747
80	3.658	3.483	3.476	3.648	3.648	3.805	3.824	4.151	4.165	3.458	3.423
81	3.481	3.483	3.483	3.523	3.525	3.656	3.672	4.045	4.066	3.481	3.466
82	3.444	3.483	3.484	3.497	3.499	3.651	3.671	4.045	4.067	3.931	4.014
83	4.086	3.483	3.460	3.952	3.946	3.963	3.943	4.045	4.043	4.001	4.003
85	3.244	3.483	3.492	3.355	3.360	3.063	3.040	4.362	4.407	3.939	4.126
86	4.125	3.483	3.458	3.980	3.973	3.355	3.247	4.043	4.039	3.968	3.894
87	4.469	3.483	3.445	4.224	4.209	3.776	3.719	4.024	4.007	3.967	3.912
88	4.357	4.057	3.998	4.199	4.173	4.411	4.422	4.561	4.575	4.247	4.290
89	4.367	4.129	4.091	4.353	4.351	4.632	4.710	4.271	4.268	3.987	4.021
90	3.509	3.483	3.482	3.539	3.541	3.290	3.267	3.139	3.098	3.345	3.252
91[j]	2.721	3.483	3.511	2.980	2.997	3.471	3.534	3.268	3.366	2.705	2.705
92[g,j]	2.495	2.870	2.995	2.447	2.427						

Table 2. (Continued)

compd no.	obs[a]	calc[b]	pred[b]	calc[c]	pred[c]	calc[d]	pred[d]	calc[e]	pred[e]	calc [f]	pred [f]
					Test Set						
7	2.921	3.482				3.201		4.060		4.510	
8	3.108	3.482				2.391		4.239		4.383	
11	5.523	5.711				5.228		4.482		4.304	
19	5.481	4.621				4.772		4.353		4.943	
20	6.301	5.529				4.698		4.336		4.249	
22	5.620	4.621				4.884		12.110		5.163	
23	5.036	4.621				5.305		5.259		5.349	
26	5.553	4.009				5.838		5.013		5.616	
27	5.102	4.621				5.488		4.953		5.551	
30	5.721	5.031				5.591		4.539		5.556	
31	5.824	5.031				6.357		5.555		5.886	
46	4.824	4.606				4.648		4.743		4.371	
52	4.585	4.606				4.766		9.103		5.056	
53	4.959	4.424				4.875		4.025		3.751	
56	5.921	5.815				5.137		5.080		5.410	
62	3.796	4.262				3.805		3.925		4.707	
63	2.824	4.083				2.327		4.285		4.863	
67	3.222	4.449				3.491		3.825		4.156	
69	3.292	5.064				3.301		2.980		3.333	
70	2.456	1.713				3.461		3.121		3.754	
76	3.509	3.482				3.425		5.448		4.027	
84	3.620	3.482				3.442		4.045		3.903	
93[i,j]	2.301	2.870									

[a] Obs = observed (refs 22−24), calc = calculated, pred = predicted. [b] From eq 1. [c] From eq 2. [d] From eq 3. [e] From eq 6. [f] From eq 7. [g] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_4$. [h] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_2$. [i] Linkage between the amide nitrogen and the piperidine nitrogen $(CH_2)_5$. [j] Benzoyl group at the fourth position of the piperidine nucleus.

Table 3. Definitions of Indicator and Physicochemical Parameters

parameter	definition
π_{m_R2}	π value of *meta* substituent present at R_2
σ_{m_R2}	Hammett σ constant of *meta* substituent present at R_2
π_{p_R2}	π value of *para* substituent present at R_2
σ_{p_R2}	Hammett σ constant of *para* substituent present at R_2
MR_{m_R2}	molar refractivity value of *meta* substituent present at R_2
MR_{p_R2}	molar refractivity value of *para* substituent present at R_2
I_m	indicator variable having a value of 1 if any substituent is present at the *meta* position of the R_2, a value of 0 otherwise
I_{Pip_3}	indicator variable having a value of 1 if piperidine is attached at the third position at R_1, a value of 0 otherwise
I_{Pip_4}	indicator variable having a value of 1 if piperidine is attached at the fourth position at R_1, a value of 0 otherwise
I_{CS_3}	indicator variable having a value of 1 if the carbonyl/sulfonyl group is attached to 3-piperidinyl nitrogen at R_1, a value of 0 otherwise
I_{CS_4}	indicator variable having a value of 1 if the carbonyl/sulfonyl group is attached to 4-piperidinyl nitrogen at R_1, a value of 0 otherwise
I_{N3}	indicator variable having a value of 1 if the $-(CH_2)_3-$ linker is present between amide nitrogen and piperidine nitrogen, a value of 0 otherwise
I_{Pip_R1}	indicator variable having a value of 1 if piperidine is present at the R_1, a value of 0 otherwise
I_{Py_R1}	indicator variable having a value of 1 if pyrrolidine is present at the R_1, a value of 0 otherwise
I_{Ph_R1}	indicator variable having a value of 1 if phenyl is present at the R_1, a value of 0 otherwise
I_{SO2_R3}	indicator variable having a value of 1 if the sulfonyl group is present at R_3, a value of 0 otherwise
I_{F_R3}	indicator variable having a value of 1 if fluorine is present at R_3, a value of 0 otherwise

in the analysis may exceed the number of observations.[26] In PCRA, all descriptors are assumed to be important, while the aim of factor analysis is to identify relevant descriptors.

Table 4. Values of Physicochemical Parameters (Substituent Constants)[a]

substituents	substituent constants						
R	π	MR[b]	σ_m	σ_p	L	B1	B5
H	0	0.10	0	0	2.06	1.00	1.00
Cl	0.71	0.60	0.37	0.23	3.52	1.80	1.80
F	0.14	0.09	0.34	0.06	2.65	1.35	1.35
CH_3	0.56	0.56	−0.07	−0.17	2.87	1.52	2.04
OCH_3	−0.02	0.79	0.12	−0.27	3.98	1.35	3.07
CF_3	0.88	0.50	0.43	0.54	3.30	1.99	2.61
$CH(CH_3)_2$	1.53	1.50	−0.04	−0.15	4.11	1.90	3.17
CN	−0.57	0.63	0.56	0.66	4.23	1.60	1.60
$C(CH_3)_3$	1.98	1.96	−0.10	−0.20	4.11	2.60	3.17

[a] Taken from ref 27. [b] MR values scaled to a factor of 0.1 as usual.

Molecular shape analysis (MSA), receptor surface analysis (RSA), and molecular field analysis (MFA) were used as the 3D-QSAR techniques. All computational 3D-QSAR experiments were conducted with Cerius[2] 4.8 version[30] from Accelrys (San Diego, U.S.A.) under a QSAR+ environment on a Silicon Graphics O2 workstation running under the IRIX 6.5 operating system. Of the 93 compounds (training set n=70, test set n=23) used in the Classical QSAR analysis, only 89 compounds (training set n=67, test set n=22) were used for the 3D-QSAR techniques because of the presence of uncommon structural features.

Molecular shape analysis[31] is a formalism that deals with quantitative characterization, representation, and manipulation of molecular shape in the construction of a QSAR. The overall aim of molecular shape analysis is to identify the biologically relevant conformation without knowledge of the receptor geometry and in a quantitative fashion explain the activity of a series of congeners. The major steps of molecular shape analysis were (1) generation of conformers and energy minimization; (2) hypothesizing an active conformer (global minimum of the most active compound);

Table 5. Factor Loadings of the Variables (Indicator and Physicochemical) after VARIMAX Rotation

	F1	F2	F3	F4	F5	F6	F7	F8	F9	communality
pC	0.278	0.162	0.394	−0.115	0.286	−0.012	0.342	−0.074	0.689	0.951
I_{Pip_R1}	0.248	0.248	**0.890**	0.234	−0.105	0.058	0.066	0.063	0.051	0.996
I_{Py_R1}	−0.075	−0.075	−0.682	−0.162	−0.671	−0.032	−0.153	−0.003	−0.096	0.986
I_{SO2_R3}	0.191	0.121	0.182	−0.028	0.006	0.195	**0.792**	0.317	0.187	0.886
I_{F_R3}	0.012	0.143	−0.048	−0.097	**0.728**	−0.124	−0.519	0.226	0.056	0.901
I_{Ph_R1}	−0.199	−0.199	−0.229	−0.080	**0.904**	−0.030	0.102	−0.069	0.054	0.974
π_{m_R2}	**0.775**	0.347	0.223	−0.043	−0.046	0.169	0.027	0.079	0.283	0.890
π_{p_R2}	0.407	**0.778**	0.253	−0.050	−0.076	0.303	0.063	0.068	0.150	0.966
$\pi_{p^2\_R2}$	0.276	**0.923**	0.182	−0.033	−0.001	0.091	0.080	0.016	−0.021	0.977
MR_{m_R2}	**0.858**	0.302	0.319	−0.022	0.082	0.110	0.101	0.049	0.091	0.968
MR_{p_R2}	0.423	**0.818**	0.290	−0.046	−0.055	0.096	−0.013	0.066	0.173	0.980
$MR_{p^2\_R2}$	0.263	**0.927**	0.195	−0.029	0.016	−0.120	0.012	0.005	−0.017	0.984
σ_{m_R2}	**0.887**	0.296	0.182	−0.065	−0.059	0.118	0.089	0.041	−0.053	0.941
σ_{p_R2}	0.334	0.103	0.073	−0.023	−0.065	**0.890**	0.193	0.121	−0.015	0.977
$\sigma_{p^2\_R2}$	0.545	0.463	0.311	−0.053	−0.120	0.419	−0.023	0.138	0.287	0.902
I_m	**0.873**	0.318	0.184	−0.067	−0.151	0.131	0.082	0.055	0.014	0.951
I_{Pip_4}	0.274	0.268	**0.899**	−0.111	−0.096	0.064	0.069	0.069	0.062	0.995
I_{Pip_3}	−0.061	−0.045	0.020	**0.995**	−0.031	−0.015	−0.004	−0.017	−0.029	0.999
I_{CS_4}	0.301	0.288	**0.861**	−0.105	−0.089	0.061	0.069	0.082	0.149	0.971
I_{CS_3}	−0.061	−0.045	0.020	**0.995**	−0.031	−0.015	−0.004	−0.017	−0.029	0.999
I_{N3}	−0.103	−0.049	−0.103	0.026	−0.043	−0.104	−0.147	**−0.947**	0.026	0.957
% variance	0.458	0.109	0.106	0.081	0.068	0.050	0.041	0.029	0.017	0.960

(3) selecting a candidate shape reference compound (based on active conformation); (4) performing pairwise molecular superimposition using the maximum common subgroup [MCSG] method; (5) measuring molecular shape commonality using MSA descriptors; (6) determination of other molecular features by calculating spatial, electronic, and conformational parameters; (7) selection of conformers; and (8) generation of QSAR equations by genetic function algorithm (GFA) or stepwise regression. A complete list of descriptors used in MSA is given in Table 6. Multiple conformations of each molecule were generated using the optimal search as a conformational search method. The upper limit of the number of conformations per molecule was 50. Each conformer was subjected to an energy minimization procedure using a smart minimizer under open force field (OFF) to generate the lowest energy conformation for each structure. A conformer of the most active antagonist **16** for the CCR5 receptor was selected as a shape reference to which all the structures in the study compounds were aligned through pairwise superpositioning. The method used for performing the alignment was a maximum common subgroup (MCSG).[30] This method looks at molecules as points and lines and uses the techniques of graph theory to identify patterns. It finds the largest subset of atoms in the shape reference compound that is shared by all the structures in the study table and uses this subset for alignment. A rigid fit of atom pairings was performed to superimpose each structure so that it overlays the shape reference compound.

In the case of the receptor surface analysis, the major steps were (1) generating conformers and energy minimization; (2) aligning molecules using the MCSG method; (3) generating the receptor model; (4) evaluating the compounds in the generated receptor model; and (5) generation of equations by the genetic function approximation method. RSA is a useful tool in situations when the 3D structure of the receptor is unknown,[32] since one can build a hypothetical model of the receptor site. RSA differs from pharmacophore models in that the former tries to capture essential information about the receptor, while the latter captures information about the commonality of compounds that bind to a receptor. A receptor surface model embodies essential information about the hypothetical receptor site as a three-dimensional surface with associated properties such as hydrophobicity, partial charge, electrostatic (ELE) potential, van der Waals (VDW) potential, and hydrogen bonding propensity. The surface points that organize as triangle meshes in the construction of the RSA store these properties as associated scalar values. Receptor surface models provide compact, quantitative descriptors, which capture three-dimensional information of interaction energies in terms of steric and electrostatic fields at each surface point, which in other techniques such as CoMFA are calculated using probe interactions at various grid points. These descriptors can be used for 3D QSAR studies.

The major steps of molecular field analysis[31] were (1) generating conformers and energy minimization; (2) matching atoms using maximum common substructure (MCS) search and aligning molecules using default option; (3) setting MFA preferences (rectangular grid with 2 Å step size, charges by Gasteiger algorithm, H^+ and CH_3 as probes); (4) creating the field; and (5) analysis by G/PLS method. MFA models are predictive and sufficiently reliable to guide the chemist in the design of novel compounds. This approach is effective for the analysis of data sets where activity information is available but the structure of the receptor site is unknown. MFA attempts to postulate and represent the essential features of a receptor site from the aligned common features of the molecules that bind to it. This method generates multiple models that can be checked easily for validity. The MFA formalism calculates probe interaction energies on a rectangular grid around a bundle of active molecules. The surface is generated from a "Shape Field". The atomic coordinates of the contributing models are used to compute field values on each point of a 3D grid. Grid size was adjusted to default 2.00 Å. MFA evaluates the energy between a probe (H^+ and CH_3) and a molecular model at a series of points defined by a rectangular grid. Fields of molecules are represented using grids in MFA, and each energy associated with an MFA grid point can serve as input for the calculation of a QSAR. These energies were added to the study table to form new columns headed according to the probe type.

Table 6. List of Values of Selected Descriptors Used in Molecular Shape Analysis

compd no.	Vm	DIFFV	Jurs-FNSA-1	Jurs-FPSA-2	Jurs-RNCG	Jurs-RNCS	Shadow-XZfrac	Shadow-Zlength	area	chiral centers
					Training Set					
1	543.164	57.016	0.213	2.266	0.111	0.073	0.690	7.724	688.161	0
2	426.948	−59.200	0.163	1.711	0.155	2.135	0.673	8.628	546.334	0
3	463.210	−22.937	0.208	1.843	0.147	0.097	0.690	7.447	587.864	0
4	495.943	9.795	0.176	2.004	0.141	0.031	0.676	8.646	631.718	0
5	517.330	31.183	0.247	1.958	0.130	0.000	0.681	7.360	652.215	0
6	479.471	−6.676	0.246	2.554	0.167	6.460	0.731	7.451	605.668	0
9	462.569	−23.579	0.205	1.869	0.146	0.127	0.564	9.366	591.254	1
10	478.976	−7.171	0.250	2.549	0.167	5.831	0.566	9.888	612.552	1
12	483.675	−2.473	0.230	1.860	0.142	0.093	0.702	7.260	619.152	0
13	489.984	3.837	0.309	1.608	0.147	0.097	0.739	6.995	627.925	0
14	495.973	9.826	0.347	1.564	0.143	0.125	0.728	7.026	638.670	0
15	492.825	6.678	0.215	1.833	0.147	0.064	0.700	7.592	633.471	0
16	498.197	12.049	0.282	1.725	0.143	0.125	0.720	7.156	637.934	0
17	532.173	46.026	0.252	1.865	0.137	0.120	0.731	7.041	683.118	0
18	507.062	20.914	0.280	1.904	0.132	1.692	0.678	7.684	654.606	0
21	539.194	53.047	0.407	2.253	0.149	4.449	0.717	7.924	681.423	0
24	555.232	69.084	0.358	2.489	0.146	5.678	0.716	7.232	710.482	0
25	571.258	85.111	0.343	3.228	0.116	2.872	0.690	8.403	721.456	0
28	542.853	56.706	0.337	2.283	0.165	4.913	0.693	7.937	692.290	0
29	558.548	72.400	0.355	2.972	0.125	4.570	0.635	9.519	703.005	0
32	553.773	67.626	0.369	3.163	0.113	3.346	0.732	7.807	697.366	0
33	571.392	85.245	0.346	3.292	0.113	3.455	0.675	8.481	720.179	0
34	589.271	103.123	0.334	3.360	0.112	3.470	0.659	9.098	741.968	0
35	620.403	134.255	0.314	3.679	0.106	3.175	0.686	8.213	779.507	0
38	424.023	−62.124	0.259	1.619	0.141	0.216	0.617	8.109	544.614	0
39	438.250	−47.898	0.317	1.479	0.142	0.217	0.608	8.184	564.428	0
40	443.145	−43.002	0.352	1.447	0.138	0.211	0.603	8.200	567.271	0
41	434.064	−52.084	0.332	1.549	0.133	0.232	0.603	8.174	552.986	0
42	447.247	−38.900	0.360	1.413	0.139	0.213	0.598	8.234	573.181	0
43	445.728	−40.419	0.279	1.638	0.135	0.178	0.615	8.172	576.609	0
44	479.752	−6.395	0.253	1.767	0.130	0.199	0.638	8.237	621.593	0
45	460.356	−25.792	0.389	1.627	0.116	0.202	0.618	8.209	589.425	0
47	490.917	4.770	0.286	1.979	0.113	0.966	0.605	8.195	628.090	0
48	456.975	−29.172	0.337	1.822	0.119	3.068	0.588	8.249	581.953	0
49	460.364	−25.783	0.316	1.911	0.131	1.687	0.588	9.088	577.335	0
50	455.103	−31.044	0.280	1.780	0.141	2.666	0.604	8.233	580.889	0
51	464.894	−21.253	0.312	1.584	0.134	0.234	0.593	8.234	593.679	0
54	457.371	−28.777	0.410	1.307	0.139	0.212	0.599	8.281	589.451	0
55	471.427	−14.721	0.450	1.228	0.138	0.211	0.651	7.982	606.454	0
57	551.048	64.901	0.337	2.743	0.136	4.270	0.653	9.156	686.943	0
58	429.325	−56.823	0.196	1.833	0.151	0.165	0.679	8.066	542.646	1
59	445.344	−40.803	0.220	1.797	0.150	0.164	0.671	8.298	575.561	1
60	445.476	−40.671	0.181	1.887	0.150	0.164	0.667	8.781	574.386	1
61	445.870	−40.278	0.209	1.821	0.150	0.164	0.643	8.399	574.424	1
64	443.605	−42.542	0.291	1.607	0.152	0.233	0.684	8.250	569.116	1
65	443.324	−42.823	0.286	1.616	0.152	0.133	0.705	7.375	573.799	1
66	457.415	−28.732	0.329	1.531	0.151	0.132	0.718	7.356	588.639	1
68	438.361	−47.786	0.290	1.765	0.139	0.152	0.729	7.217	564.707	1
71	445.942	−40.206	0.260	1.758	0.145	0.159	0.699	7.694	565.451	1
72	461.765	−24.383	0.369	1.481	0.147	0.129	0.709	7.341	595.375	1
73	479.615	−6.532	0.203	1.940	0.141	0.061	0.585	8.904	616.457	1
74	502.067	15.920	0.201	1.994	0.136	0.030	0.570	10.207	638.191	1
75	519.242	33.095	0.148	2.178	0.134	0.000	0.695	8.416	657.811	1
77	500.615	14.468	0.234	1.993	0.130	0.085	0.662	8.901	629.247	1
78	514.200	28.052	0.280	1.864	0.130	0.086	0.628	9.742	648.580	1
79	514.756	28.609	0.284	1.850	0.131	0.086	0.672	8.862	649.204	1
80	514.646	28.498	0.299	1.810	0.131	0.086	0.704	8.445	647.195	1
81	517.295	31.148	0.219	2.050	0.129	0.113	0.662	9.037	651.587	1
82	517.438	31.290	0.221	2.056	0.129	0.085	0.686	9.173	659.525	1
83	482.281	−3.867	0.222	2.163	0.168	3.172	0.646	9.769	613.424	1
85	411.542	−74.605	0.252	1.714	0.155	0.882	0.698	7.072	531.642	1
86	460.592	−25.555	0.337	1.841	0.120	0.052	0.599	9.738	588.825	1
87	517.275	31.128	0.227	2.028	0.129	0.113	0.633	9.766	655.302	1
88	514.858	28.711	0.300	1.809	0.131	0.029	0.525	9.883	651.372	1
89	528.705	42.558	0.341	1.714	0.130	0.085	0.526	9.780	667.814	1
90	433.472	−52.675	0.266	1.722	0.147	0.161	0.715	7.253	559.219	1
91	436.281	−49.867	0.300	1.752	0.138	0.151	0.617	7.934	559.204	1

Table 6. (Continued)

compd no.	Vm	DIFFV	Jurs-FNSA-1	Jurs-FPSA-2	Jurs-RNCG	Jurs-RNCS	Shadow-XZfrac	Shadow-Zlength	area	chiral centers
					Test Set					
7	542.642	56.495	0.185	2.368	0.109	0.096	0.673	9.726	689.067	1
8	425.719	−60.429	0.154	1.744	0.153	1.476	0.669	8.603	550.082	1
11	481.080	−5.068	0.262	1.754	0.144	0.095	0.670	8.509	617.421	0
19	511.567	25.420	0.402	2.068	0.164	6.334	0.743	7.013	653.928	0
20	514.139	27.992	0.340	2.286	0.164	6.355	0.730	7.234	646.352	0
22	530.496	44.349	0.399	2.451	0.139	5.335	0.693	8.338	671.923	0
23	518.192	32.045	0.314	2.520	0.154	4.500	0.714	7.868	665.065	0
26	585.364	99.217	0.278	2.903	0.141	4.265	0.694	8.527	742.049	0
27	532.338	46.190	0.303	2.516	0.157	4.783	0.682	8.076	672.501	0
30	576.099	89.952	0.346	3.058	0.124	3.924	0.729	7.754	729.278	0
31	593.932	107.784	0.325	3.194	0.122	2.618	0.719	7.918	752.573	0
46	446.378	−39.769	0.337	1.553	0.131	0.172	0.608	8.173	566.240	0
52	481.245	−4.903	0.345	2.274	0.167	6.357	0.643	9.025	610.028	0
53	460.204	−25.943	0.337	1.494	0.136	0.119	0.642	7.986	578.968	0
56	490.969	4.822	0.362	2.160	0.171	6.260	0.622	7.886	633.394	0
62	497.032	10.885	0.189	1.974	0.142	0.155	0.641	8.508	638.158	1
63	455.493	−30.654	0.181	2.048	0.140	2.215	0.665	8.662	581.211	1
67	448.219	−37.928	0.318	1.625	0.145	0.158	0.735	7.209	575.007	1
69	460.707	−25.440	0.321	1.830	0.128	0.168	0.681	7.954	585.998	1
70	460.539	−25.608	0.322	1.825	0.128	0.168	0.632	8.228	592.847	1
76	484.338	−1.810	0.268	1.834	0.125	0.000	0.593	10.084	609.196	1
84	496.104	9.956	0.220	2.060	0.128	0.056	0.660	8.920	625.617	1

Model extraction from the data was done using genetic function approximation (GFA) for molecular shape analysis and receptor surface analysis and G/PLS for molecular field analysis using QSAR+ environment of Cerius^2 software.[30]

The genetic function approximation (GFA) technique[33,34] was used to generate a population of equations rather than one single equation for correlation between biological activity and physicochemical properties. GFA involves the combination of multivariate adaptive regression splines (MARS) algorithm with a genetic algorithm to evolve population of equations that best fit the training set data. It provides an error measure, called the lack of fit (LOF) score, that automatically penalizes models with too many features. It also inspires the use of splines as a powerful tool for nonlinear modeling. GFA is done as follows: (i) an initial population of equations is generated by a random choice of descriptors; (ii) pairs from the population of equations are chosen at random and "crossovers" are performed and progeny equations are generated; (iii) it is better at discovering combinations of features that take advantage of correlations between multiple features; (iv) the fitness of each progeny equation is assessed by the LOF measure; (v) it can use a larger variety of equation term types in construction of its models; and (vi) if the fitness of a new progeny equation is better, then it is preserved. The model with the proper balance of all statistical terms will be used to explain the variance of the biological activity. A distinctive feature of GFA is that it produces a population of models (e.g., 100), instead of generating a single model, as do most other statistical methods. The range of variations in this population gives added information on the quality fit and importance of the descriptors.

The genetic partial least squares (G/PLS) algorithm may be used as an alternative to a GFA calculation. G/PLS is derived from two QSAR calculation methods: GFA and partial least squares (PLS). The G/PLS algorithm uses GFA to select appropriate basis functions to be used in a model of the data and PLS regression as the fitting technique to weigh the basis functions relative contributions in the final model. PLS is a generalization of regression, which can handle data with strongly correlated and/or noisy or numerous X variables.[35] It gives a reduced solution, which is statistically more robust than multiple linear regression (MLR). The linear PLS model finds "new variables" (latent variables or X scores) which are linear combinations of the original variables. To avoid overfitting, a strict test for the significance of each consecutive PLS component is necessary and then stopping when the components are nonsignificant. Cross-validation is a practical and reliable method of testing this significance.[34] Application of G/PLS thus allows the construction of larger QSAR equations while still avoiding overfitting and eliminating most variables.

The factor analysis (FA) and principal component regression analysis were performed using the statistical software SPSS.[36] The statistical qualities of the multiple linear regression (MLR) equations[37] were judged by the parameters such as *explained variance* (R_a^2), *correlation coefficient* (R), *standard error of estimate* (s), and *variance ratio* (F) at specified *degrees of freedom* (df). All accepted MLR equations have regression coefficients and F ratios significant at 95% and 99% levels, respectively, if not stated otherwise. For the PLS equation R_a^2, R^2, and least-squares error (LSE) were taken as a statistical measure, while lack-of-fit (LOF) was noted for the GFA derived equations. The generated QSAR equations were validated by *PRESS* (leave-one-out)[38,39] and bootstrap statistics which were calculated using the QSAR+ module of Cerius^2 software,[30] and the reported parameters are *cross-validation* R^2 (Q^2), *predicted residual sum of squares* (*PRESS*), *standard deviation based on PRESS* (S_{PRESS}), *standard deviation of error of prediction* (*SDEP*), and *bootstrap* r^2 (bsr^2). The finally developed models were subjected to a randomization test for validation purpose. Additionally, the final models were subjected to leave-15%-out and leave-25%-out cross-validation with 7 and 4 cycles, respectively.

3. RESULTS AND DISCUSSION

Classical QSAR. FA-MLR (Multiple Linear Regression using Factor Analysis as the Data-Preprocessing Step). From the factor analysis of the data matrix consisting of the binding affinity data, physiochemical parameters, and indicator variables, it was observed that nine factors could explain the data matrix to the extent of 95.96%. The results of the factor analysis are presented in Table 5. Based on the factor analysis, the following equation was derived with eight variables.

$$\begin{aligned} p\mathit{K}_i = {} & 0.809(\pm 0.630)\pi_{m\_R_2} + 1.798(\pm 1.460)\pi_{p\_R_2} - \\ & 1.422(\pm 1.182)\pi_{p\_R_2}^{\ 2} - 2.119(\pm 1.646)\sigma_{p\_R_2} + \\ & 0.952(\pm 0.420)I_{CS\_4} + 0.612(\pm 0.552)I_{N3} + \\ & 1.124(\pm 0.370)I_{Ph\_R_1} + 1.022(\pm 0.572)I_{SO_2\_R_3} + \\ & 2.870(\pm 0.584) \end{aligned}$$

$$n = 70,\ R_a^{\ 2} = 0.657,\ R^2 = 0.697,\ R = 0.835,$$

$$F = 17.5(\text{df } 8,61),\ s = 0.584,\ \text{AVRES} = 0.430,\ \text{RMSE} = 0.545,$$

$$Q^2 = 0.585,\ \text{SDEP} = 0.638,\ S_{\text{PRESS}} = 0.683,$$

$$\text{PRESS} = 28.5,\ \text{bsr}^2 = 0.723(\pm 0.003),$$

$$n_{\text{Test}} = 23,\ r^2_{\text{Pred}} = 0.592,\ \text{AVRES}_{\text{Test}} = 0.642,\ \text{RMSE}_{\text{Test}} = 0.791 \quad (1)$$

The 95% confidence intervals of the regression coefficients are shown within parentheses. Equation 1 could explain 65.7% of the variance and predict 58.5% of the variance. The positive coefficient of the lipophilic substituent constant (π_{m_R2}) of the *meta* substituents of the R_2 position is conducive for the binding affinity. The lipophilic substituent constant (π_{p_R2}) of the *para* substituents of the R_2 position shows a parabolic relation with the binding affinity. This suggests that the binding affinity increases with an increase in lipophilicity of *para* substituents at the R_2 position until it reaches the critical value after which the affinity decreases. The negative coefficient of σ_{p_R2} indicates that the presence of electron-withdrawing substituents at the *para* position of the R_2 position is not favorable for the binding affinity. Again, the presence of either the sulfonyl or the carbonyl substituent at the 4-piperidinyl nitrogen is conducive for the binding affinity as evidenced from the positive coefficient of I_{CS_4}. The presence of the sulfonyl group at the R_3 position and phenyl at the R_1 position is conducive for the binding affinity as evidenced from the positive coefficients of I_{SO2_R3} and I_{Ph_R1}. Furthermore, the positive coefficient of I_{N3} indicates that presence of the $-(CH_2)_3-$ group as the linker between amide nitrogen and piperidine nitrogen increases the binding affinity. The calculated and leave-one-out cross-validation values of binding affinity according to eq 1 are given in Table 2. The predictive R^2 value for the test set was found to be 0.592. Figure 1(a),(b) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation tests, and the results are also presented in Table 8. The statistics of all the models developed are presented in Table 9.

PCRA. When factor scores were used as the predictor parameters in a multiple regression equation using forward selection method (PCRA), the following equation was obtained.

$$\begin{aligned} p\mathit{K}_i = {} & 0.278(\pm 0.056)\ \text{fs1} + 0.162(\pm 0.056)\ \text{fs2} + \\ & 0.393(\pm 0.056)\ \text{fs3} - 0.115(\pm 0.056)\ \text{fs4} + \\ & 0.285(\pm 0.056)\ \text{fs5} + 0.341(\pm 0.056)\ \text{fs7} - \\ & 0.074(\pm 0.056)\ \text{fs8} + 0.687(\pm 0.056)\ \text{fs9} + \\ & 4.368(\pm 0.056) \end{aligned}$$

$$n = 70,\ R_a^{\ 2} = 0.944,\ R^2 = 0.951,\ R = 0.975,$$

$$F = 146.6(\text{df } 8,61),\ s = 0.236,$$

$$Q^2 = 0.898,\ \text{SDEP} = 0.316,\ S_{\text{PRESS}} = 0.339,$$

$$\text{PRESS} = 7.0,\ \text{Pres}_{\text{av}} = 0.208 \quad (2)$$

Eq 2 shows excellent equation statistics (94.4% explained variance) and cross-validation parameters (89.8% predicted variance). The variables (factor scores) used in eq 2 are perfectly orthogonal to each other. As factor scores are used instead of selected descriptors in the MLR equation in PCRA and any one factor score contains information from different descriptors, loss of information is thus avoided and the quality of the PCRA equation is better than those derived from other techniques. From the factor scores used, significance of the original variables for modeling the binding affinity can be obtained. Factor scores 1 and 2 indicate the importance of lipophilicity of substituents at different positions (π_{m_R2} and π_{p_R2} respectively). Factor score 3 indicates the importance of the piperidine and 4-piperidinyl moieties at (R_1), while factor score 4 signifies the importance of the 3-piperidinyl moiety. Factor score 5 signifies the importance of fluorine (at R_3) and phenylamino (at R_1) substituents, factor score 7 signifies the importance of the sulfonyl group at the R_3 position. Factor score 8 signifies the importance of the $-(CH_2)_3-$ linker between the amide nitrogen and the piperidine nitrogen. The calculated and leave-one-out cross-validation values of binding affinity according to eq 2 are

Table 7. Results of a Randomization Test Applied on the Developed Models

eq. no.	3	4	5	6	7
3D QSAR method	MSA	MSA	MSA	RSA	MFA
modeling technique	GFA	GFA	GFA	GFA	G/PLS
R from nonrandom model	0.850	0.851	0.850	0.783	0.880
no. of random trials	99	99	99	99	99
no. of random *R*'s less than nonrandom *R*	99	99	99	99	99
no. of random *R*'s more than nonrandom *R*	0	0	0	0	0
confidence level	99%	99%	99%	99%	99%
mean value of *R* from random trials ± standard deviation	0.316 ± 0.083	0.326 ± 0.074	0.321 ± 0.083	0.262 ± 0.091	0.410 ± 0.077

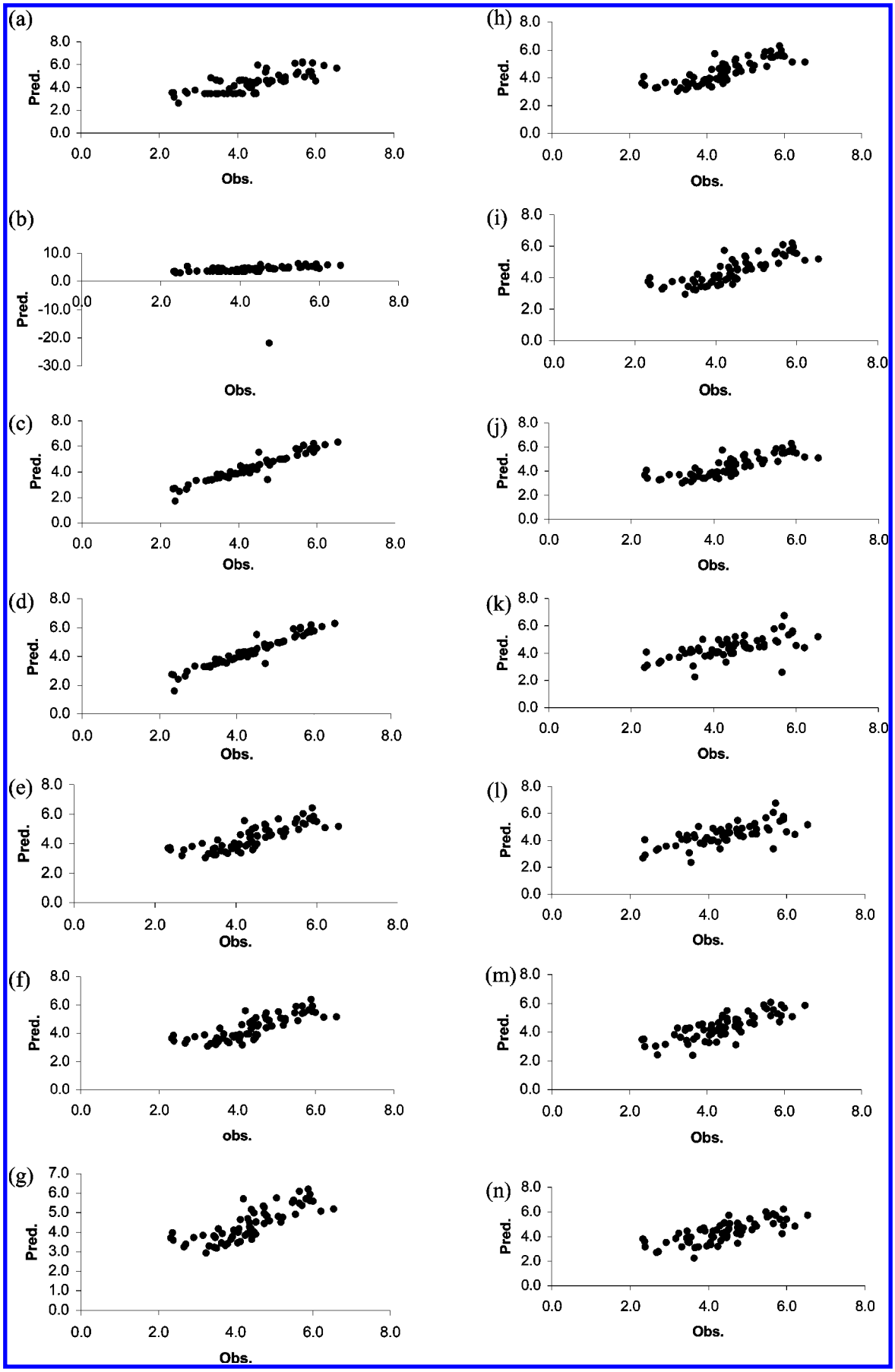


Figure 1. Scatter plots of observed vs leave-*n*%-out predicted CCR5 binding affinity values 3-(4-benzylpiperidin-1-yl)-*N*-phenylpropylamine derivatives for the models: (**a**) model (1) (15%-out), (**b**) model (1) (25%-out), (**c**) model (2) (15%-out), (**d**) model (2) (25%-out), (**e**) model (3) (15%-out), (**f**) model (3) (25%-out), (**g**) model (4) (15%-out), (**h**) model (4) (25%-out), (**i**) model (5) (15%-out), (**j**) model (5) (25%-out), (**k**) model (6) (15%-out), (**l**) model (6) (25%-out), (**m**) model (7) (15%-out), (**n**) model (7) (25%-out).

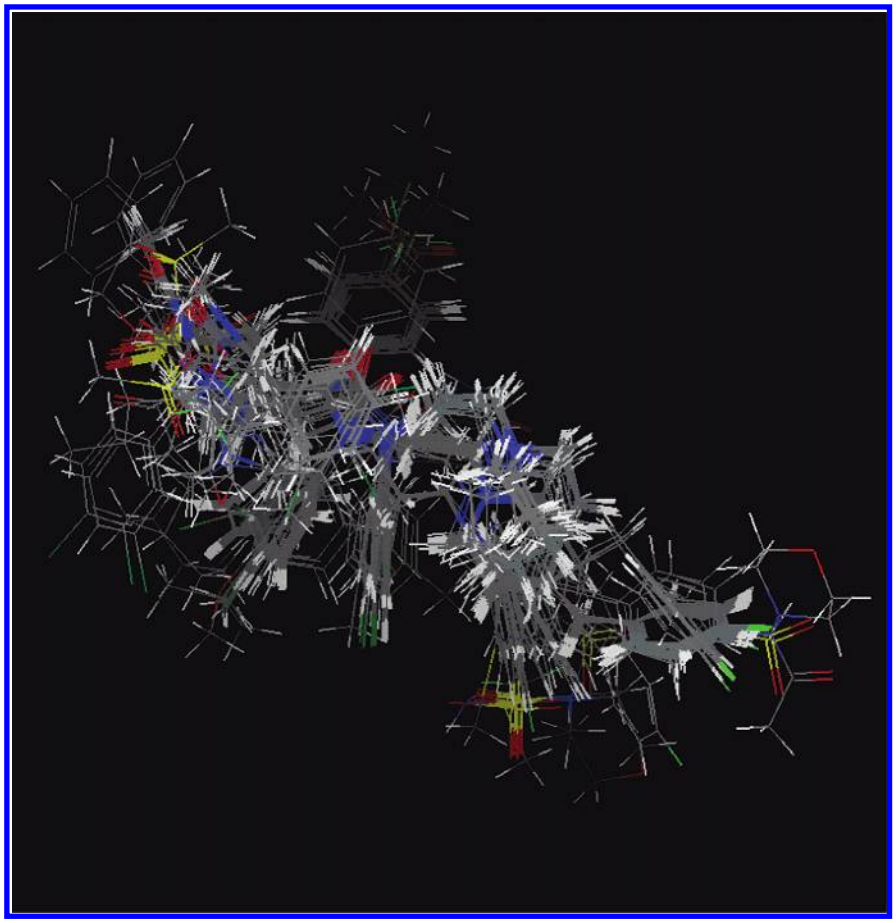


Figure 2. The view of aligned geometry of the training set molecules.

Table 8. Results of Leave-*n*%-out Cross-Validation

method	model no.	*n*	Q^2 (average Pres)	PRESS
PCA	(1)	15[a]	0.601(0.488)	27.340
		25[b]	−9.788(0.881)	740.643
PCRA	(2)	15[a]	0.912(0.198)	5.987
		25[b]	0.919(0.188)	5.533
MSA	(3)	15[a]	0.663(0.447)	21.533
		25[b]	0.659(0.444)	21.777
	(4)	15[a]	0.658(0.443)	21.862
		25[b]	0.665(0.436)	21.420
	(5)	15[a]	0.658(0.442)	21.847
		25[b]	0.666(0.433)	21.356
RSA	(6)	15[a]	0.397(0.567)	38.519
		25[b]	0.486(0.528)	32.836
MFA	(7)	15[a]	0.617(0.488)	24.488
		25[b]	0.543(0.525)	29.180

[a] Compounds were deleted in 7 cycles in the following manner: (**1**, **8**, **15**,....., **64**), (**2**, **9**, **16**,...., **65**),....., (**7**, **14**, **21**, ..., **70**). [b] Compounds were deleted in 4 cycles in the following manner: (**1**, **5**, **9**, ..., **69**), (**2**, **6**, **10**, ...,**70**),....., (**4**, **8**, **12**, ..., **68**). Average Pres is average of the absolute values of predicted residuals.

given in Table 2. Figure 1(c),(d) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation test, and the results are also presented in Table 8.

However, the Hansch approach could not give more insight into the structure−activity relationships because of the diverse molecular features in the data set. In cases where specific three-dimensional characteristics such as stereochemistry affect significantly the biological activity of the drug molecule, 3-D-QSAR scores over normal QSAR

Table 9. Comparison of Statistical Qualities of the Developed Models

method	model no.	R^2	R_a^2	F(df)	Q^2(LOO)	PRESS	R^2_{Pred}
PCA	(1)	0.697	0.657	17.5(8,61)	0.585	28.5	0.592
PCRA	(2)	0.951	0.944	146.6(8,61)	0.898	7.0	
MSA	(3)	0.722	0.689	21.9(7,59)	0.650	22.4	0.774
	(4)	0.724	0.691	22.1(7,59)	0.646	22.6	0.770
	(5)	0.723	0.690	22.0(7,59)	0.646	22.6	0.760
RSA	(6)	0.613	0.582	19.4(5,61)	0.535	29.7	−1.652
MFA	(7)	0.774	0.719		0.660	21.8	0.327

analysis. We also studied the three-dimensional structure−activity relationship of 3-(4-benzylpiperidin-1-yl)-*N*-phenylpropylamine derivatives reported by Imamura et al.[22−24] to gain more insight into steric, electrostatic, hydrophobic, and hydrogen bonding properties that influence the activity. The favorable and unfavorable interactions, usually represented graphically in 3D-QSAR contours around the representative molecule, are easier to visualize than a mathematical formula. In 3D-QSAR the properties are calculated individually, and there is no reliance on external or tabulated factors. As long as all the study compounds share the same pharmacophore and interact in the same way with the target, they can be analyzed.[40]

Molecular Shape Analysis. The view of the aligned training set molecules is shown in Figure 2. The values of the important descriptors used in MSA derived equations are given in Table 6. The following three equations (eqs 3−5) were among the best ones obtained from the genetic function

approximation (500 000 crossovers, linear terms, and other default settings):

$$\begin{aligned} pK_i = {} & 0.011(\pm 0.006)\mathrm{Vm} + \\ & 45.828(\pm 19.206)\mathrm{JursRNCG} - \\ & 0.438(\pm 0.192)\mathrm{JursRNCS} + \\ & 6.277(\pm 2.833)\mathrm{JursFNSA\_1} + \\ & 1.243(\pm 0.774)\mathrm{JursFPSA\_2} - \\ & 1.362(\pm 0.338)\mathrm{Chiralcenters} - \\ & 3.220(\pm 2.979)\mathrm{Shadow\_XZfrac} - 8.202(\pm 4.052) \end{aligned}$$

$$n = 67,\ R_a^2 = 0.689,\ R^2 = 0.722,\ R = 0.850,$$

$$F = 21.9(\mathrm{df}\ 7,59),\ \mathrm{LOF} = 0.424,\ \mathrm{AVRES} = 0.408,\ \mathrm{RMSE} = 0.515,$$

$$Q^2 = 0.650,\ \mathrm{SDEP} = 0.578,\ S_{\mathrm{PRESS}} = 0.616,$$

$$\mathrm{PRESS} = 22.4,\ \mathrm{bsr}^2 = 0.722(\pm 0.002),$$

$$n_{\mathrm{Test}} = 22,\ r^2_{\mathrm{Pred}} = 0.774,\ \mathrm{AVRES}_{\mathrm{Test}} = 0.419,\ \mathrm{RMSE}_{\mathrm{Test}} = 0.562 \quad (3)$$

The 95% confidence intervals of regression coefficients are given within parentheses. Equation 3 could explain 68.9% of the variance and predict 65.0% of the variance. The calculated values of the binding affinity according to eq 3, the leave-one-out predictions, and the predicted activity data of the test set are given in Table 2. Figure 1(e),(f) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation tests, and the results (Q^2) are presented in Table 8. The predictive R^2 value for the test set was found to be 0.774.

The positive coefficient of JursRNCG indicates that the relative negative charge (RNCG), the charge of the most negative atom divided by the total negative charge, is conducive for binding affinity. Substituents with a high RNCG value (e.g. R_1 = *N*-methylsulfonylpiperidin-4-yl) have more binding affinity than the unsubstituted piperidinyl and phenylamino (R_1 position) congeners. The relative negative charge surface area (RNCS), which is the ratio between the solvent-accessible surface area of most negative atoms and RNCG, is detrimental for the activity. Compounds with methyl and 3,4-dichloro substituents at the phenyl ring (R_2 position) have lower JursRNCS values and thus have better binding affinity than the congeners with an unsubstituted phenyl nucleus. An increase in the area of the molecular shadow in the XZ plane (Sxz) is detrimental (e.g. R_1 = *N*-*c*-hexylmethyl-5-oxopyrrolidin-3-yl) for the binding affinity. The presence of a chiral center makes the molecule less active (e.g. R_1 = 5-oxopyrrolidin-3-yl or piperidin-3-yl). An increase in the molecular volume is conducive (e.g. R_3 = $(CH_3)_2NSO_2-$) for the binding affinity. Compounds with the substituents of higher JursFPSA_2 values (fractional charged partial surface area) like the *N*-methylsulfonylpiperidin-4-yl (R_1 position) group have better binding affinity than the compounds with substituents such as 4-chlorophenylamino (R_1 position). Unsubstituted piperidines (at R_1 position) with less JursFNSA_1 values have lower binding affinity than the 4-chlorophenylamino substituted compounds.

$$\begin{aligned} pK_i = {} & 0.009(\pm 0.006)\mathrm{DIFFV} + \\ & 44.224(\pm 18.351)\mathrm{JursRNCG} - \\ & 0.440(\pm 0.192)\mathrm{JursRNCS} + \\ & 6.661(\pm 2.893)\mathrm{JursFNSA\_1} + \\ & 1.175(\pm 0.752)\mathrm{JursFPSA\_2} - \\ & 1.489(\pm 0.382)\mathrm{Chiralcenters} + \\ & 0.227(\pm 0.202)\mathrm{Shadow\_Zlength} - 6.647(\pm 4.806) \end{aligned}$$

$$n = 67,\ R_a^2 = 0.691,\ R^2 = 0.724,\ R = 0.851,$$

$$F = 22.1(\mathrm{df}\ 7,59),\ \mathrm{LOF} = 0.421,\ \mathrm{AVRES} = 0.403,\ \mathrm{RMSE} = 0.513,$$

$$Q^2 = 0.646,\ \mathrm{SDEP} = 0.581,\ S_{\mathrm{PRESS}} = 0.619,$$

$$\mathrm{PRESS} = 22.6,\ \mathrm{bsr}^2 = 0.724(\pm 0.003),$$

$$n_{\mathrm{Test}} = 22,\ r^2_{\mathrm{Pred}} = 0.770,\ \mathrm{AVRES}_{\mathrm{Test}} = 0.424,\ \mathrm{RMSE}_{\mathrm{Test}} = 0.568 \quad (4)$$

Eq 4 could explain 69.1% of the variance and predict 64.6% of the variance. Increase in the length of molecule in the Z dimension (Lz) is conducive (e.g. R_3 = sulfonylmorpholino) for the binding affinity. An increase in the difference between the volume of the individual molecule and the shape reference compound is conducive (e.g. $R_3 = (CH_3)_2NSO_2-$) for the binding affinity. The predictive R^2 value for the test set was found to be 0.770. Figure 1(g),(h) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation test, and the results (Q^2) are also presented in Table 8.

$$\begin{aligned} pK_i = {} & 0.007(\pm 0.004)\mathrm{Area} + \\ & 43.680(\pm 18.373)\mathrm{JursRNCG} - \\ & 0.438(\pm 0.192)\mathrm{JursRNCS} + \\ & 6.660(\pm 2.897)\mathrm{JursFNSA\_1} + \\ & 1.205(\pm 0.744)\mathrm{JursFPSA\_2} - \\ & 1.487(\pm 0.384)\mathrm{Chiralcenters} + \\ & 0.241(\pm 0.200)\mathrm{Shadow\_Zlength} - 11.277(\pm 4.936) \end{aligned}$$

$$n = 67,\ R_a^2 = 0.690,\ R^2 = 0.723,\ R = 0.850,$$

$$F = 22.0(\mathrm{df}\ 7,59),\ \mathrm{LOF} = 0.423,\ \mathrm{AVRES} = 0.402,\ \mathrm{RMSE} = 0.514,$$

$$Q^2 = 0.646,\ \mathrm{SDEP} = 0.581,\ S_{\mathrm{PRESS}} = 0.620,$$

$$\mathrm{PRESS} = 22.6,\ \mathrm{bsr}^2 = 0.723(\pm 0.003),$$

$$n_{\mathrm{Test}} = 22,\ r^2_{\mathrm{Pred}} = 0.760,\ \mathrm{AVRES}_{\mathrm{Test}} = 0.430,\ \mathrm{RMSE}_{\mathrm{Test}} = 0.579 \quad (5)$$

Eq 5 could explain 69.0% of the variance and predict 64.6% of the variance. Eqs 4 and 5 are close in statistical

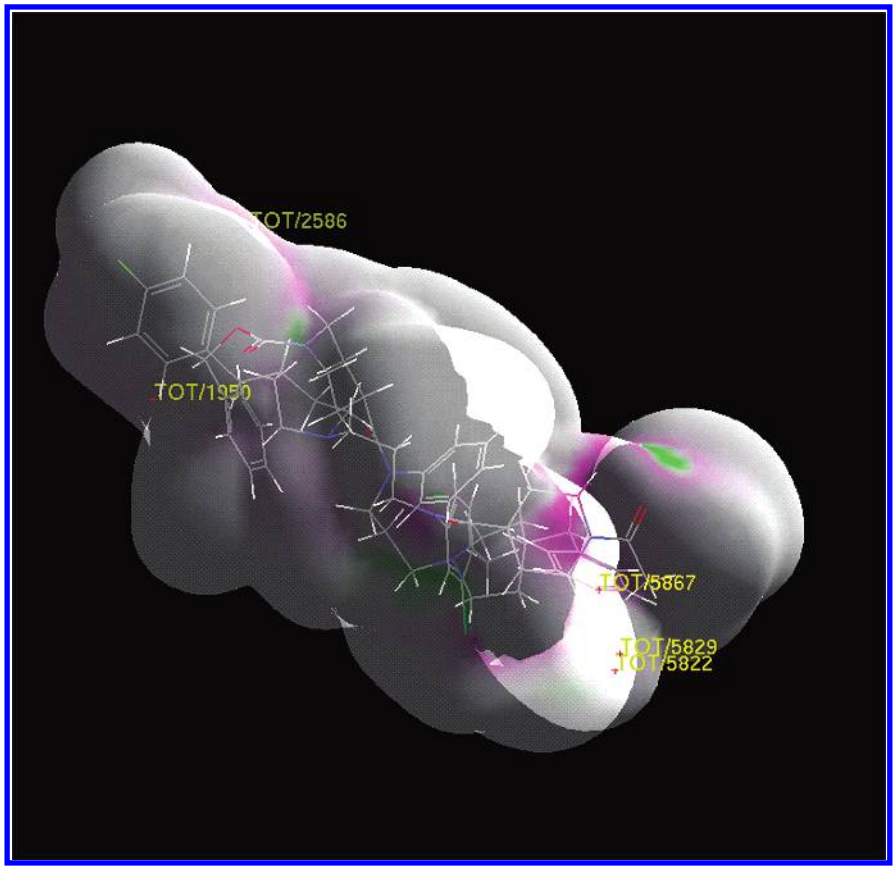


Figure 3. The view of receptor surface model and the interaction points with the most active compound **16** embedded into the model.

quality but not superior to eq 3. An increase in the van der Waals area is conducive (e.g. $R_3 = (CH_3)_2NSO_2-$) for the binding affinity. The predictive R^2 value for the test set was found to be 0.760. The final models (eq 3−5) were also subjected to a randomization test with 99% confidence intervals (Table 7). Figure 1(i),(j) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation test, and the results (Q^2) are also presented in Table 8.

Receptor Surface Analysis. The receptor surface model is normally generated from the most active compounds in the data set. The rationale is that the most active molecules tend to explore the best spatial and electronic interactions with the receptor, while the least active do not and tend to have unfavorable steric and electrostatic interactions. We initiated our study with receptor models generation using the 10 best active compounds (compound nos. **13, 14, 15, 16, 25, 29, 33, 34, 35, 57**) in the training set. After the receptor surface model has been generated, all the structures in the training and test sets can be evaluated against the model. The model can be used to calculate the energy associated with the binding of a molecule in the model. It can also be used to minimize a molecule by adjusting the geometry of the structure into a "best-fit configuration" based on the constraints imposed by the receptor model. The view of aligned geometry of the training set molecules is shown in Figure 2. The best model (GFA, 50 000 crossovers, linear terms, other default settings) obtained from the training set compounds is the following:

$$\begin{aligned} pK_i = &-1.192(\pm0.834)\mathrm{TOT/1950} - \\ &1.863(\pm0.814)\mathrm{TOT/2586} - 8.245(\pm2.042)\mathrm{TOT/5822} + \\ &4.312(\pm1.440)\mathrm{TOT/5829} - \\ &3.371(\pm1.220)\mathrm{TOT/5867} + 4.045(\pm0.240) \end{aligned}$$

$$n = 67,\ R_a^2 = 0.582,\ R^2 = 0.613,\ R = 0.783,$$

$$F = 19.4(\mathrm{df}\ 5,61),\ \mathrm{LOF} = 0.509,\ \mathrm{AVRES} = 0.464,\ \mathrm{RMSE} = 0.607,$$

$$Q^2 = 0.535,\ \mathrm{SDEP} = 0.666,\ S_{\mathrm{PRESS}} = 0.698,$$

$$\mathrm{PRESS} = 29.7,\ \mathrm{bsr}^2 = 0.614(\pm0.006),$$

$$n_{\mathrm{Test}} = 22,\ r^2_{\mathrm{Pred}} = -1.652,\ \mathrm{AVRES}_{\mathrm{Test}} = 1.235,\ \mathrm{RMSE}_{\mathrm{Test}} = 1.928 \quad (6)$$

The 95% confidence intervals of regression coefficients are given within parentheses. The resultant model was of good statistical qualities: 58.2% explained variance and 53.5% predicted variance. The descriptors TOT/1950, TOT/2586, etc. are added energy of both electrostatic interaction energy and van der Waals interaction energy at points 1950, 2586, etc. By using receptor data to develop a QSAR model, the goodness of fit can be evaluated between a candidate structure and a postulated pseudoreceptor. The view of the receptor surface model (RSM) with embedded most active compound **16** and the interaction points is shown in Figure 3. The RSM represents essential information about the hypo-

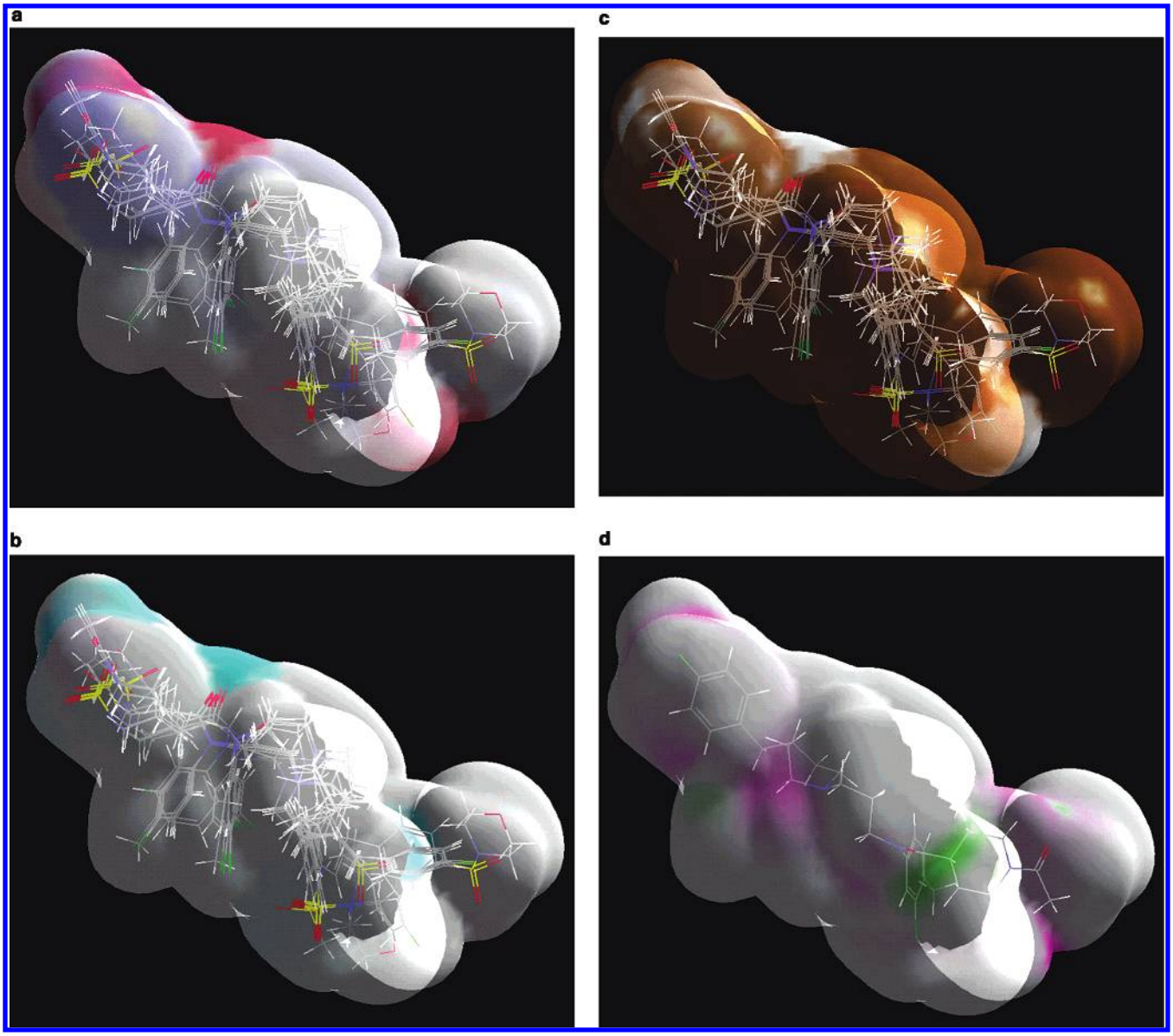


Figure 4. (a) The 10 best active compounds (**13, 14, 15, 16, 25, 29, 33, 34, 35, 57**) embedded into the receptor surface model mapped with charge: the red areas are positively charged, blue areas are negatively charged, and white areas are neutral. **(b)** The 10 best active compounds (**13, 14, 15, 16, 25, 29, 33, 34, 35, 57**) embedded into the receptor surface model mapped with hydrogen bonding: the purple areas act as hydrogen bond donors, light blue areas act as hydrogen bond acceptors, and white areas do not have hydrogen bonding activity. **(c)** The 10 best active compounds (**13, 14, 15, 16, 25, 29, 33, 34, 35, 57**) embedded into receptor surface model mapped with hydrophobicity: the brown areas are hydrophobic. **(d)** The best active compound **16** embedded into the receptor surface model mapped with total energy: the magenta color represents favorable interaction sites, while the green-colored regions represent unfavorable sites for binding of the molecule on the receptor surface.

thetical receptor site as a three-dimensional surface with associated properties mapped onto the surface model. The location and magnitude of a descriptor can be used as a guideline to improve the activity of molecules. The surface represents information about the steric nature of the receptor site and the associated properties of interest, such as hydrophobicity, partial charge, and hydrogen-bonding propensity. Figure 4(a) shows the 10 best active compounds embedded into the receptor surface model mapped with charge: the red areas are positively charged, the blue areas are negatively charged, and the white areas are neutral. Similarly, Figure 4(b) shows the 10 best active compounds embedded into the receptor surface model mapped with hydrogen bonding: the purple areas act as hydrogen bond donors, the light blue areas act as hydrogen bond acceptors, and the white areas do not have hydrogen bonding activity. Figure 4(c) shows the receptor surface model mapped with hydrophobicity: the brown areas are hydrophobic. Figure 4(d) shows the best active compound **16** embedded into the receptor surface model mapped with total energy: the magenta color represents negative energy values as favorable interaction sites, while the green-colored regions represent positive energy values that are unfavorable sites for binding of the molecule on the receptor surface. This color coding of the ligand−receptor interactions can offer a qualitative way of examining compounds, by introducing them into the virtual receptor and visually inspecting the favorable/unfavorable interactions; substituents that increase or decrease the binding affinity can be easily recognized, and one can make easily simple but accurate structure−activity

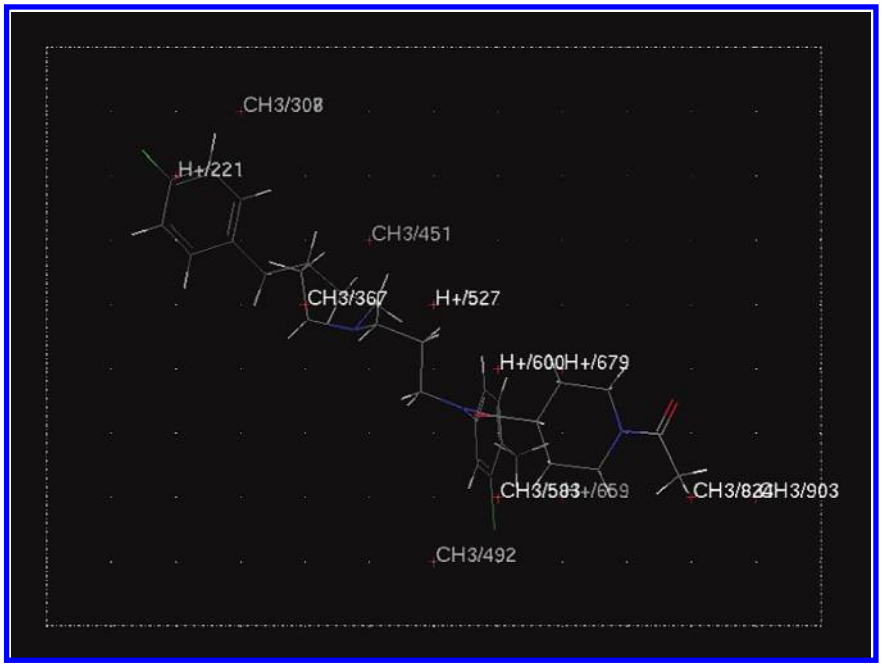


Figure 5. The view of the most active compound **16** in the molecular field and the interaction points.

estimations. Though we got statistically acceptable results for the training set, the prediction results of the test set are not equally encouraging. However, on deletion of three compounds (**22**, **52**, and **76**) from the test set, r^2_{Pred} value rises to 0.458. The calculated and leave-one-out cross-validation values of binding affinity according to eq 6 are given in Table 2. Figure 1(k),(l) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation test, and the results (Q^2) are presented in Table 8.

Molecular Field Analysis. The field generated was of the rectangular type. The probes used were H^+ and CH_3. The charge method used was Gasteiger, and the energy cutoff was kept at −30 to +30 Kcal. The QSAR equation was generated using the genetic partial least-squares (G/PLS) method. The number of iterations was set to 5000, and the number of components was set to 3 to obtain the final equation. The mutation probabilities were set to the system defaults. The final result was obtained with the number of components at three optimized by cross-validation. The view of aligned training set molecules in the field is shown in Figure 2. The following equation was obtained from the molecular field analysis:

$$\begin{aligned} pK_i = {} & 0.011H^+/221 + 0.025H^+/527 + 0.014H^+/600 + \\ & 0.023H^+/659 - 0.001H^+/679 - 0.015CH_3/307 - \\ & 0.008CH_3/308 - 0.025CH_3/367 - 0.010CH_3/451 + \\ & 0.014CH_3/492 + 0.023CH_3/583 + 0.031CH_3/824 + \\ & 0.026CH_3/903 + 3.143 \end{aligned}$$

$$n = 67,\ R_a^2 = 0.719,\ R^2 = 0.774,\ R = 0.880,$$

$$\text{LSE} = 0.216,\ \text{AVRES} = 0.375,\ \text{RMSE} = 0.465,$$

$$Q^2 = 0.660,\ \text{SDEP} = 0.570,\ S_{\text{PRESS}} = 0.641,$$

$$\text{PRESS} = 21.8,\ \text{bsr}^2 = 0.774(\pm 0.003),$$

$$n_{\text{Test}} = 22,\ r^2_{\text{Pred}} = 0.327,\ \text{AVRES}_{\text{Test}} = 0.766,\ \text{RMSE}_{\text{Test}} = 0.971 \quad (7)$$

In eq 5, $H^+/221$, $H^+/527$,... and so on are the probes and their numbering (corresponding to spatial positions as shown in Figure 5); i.e., these represent interactions at points 221 by H^+, 527 by H^+, etc. Equation 7 is of good statistical qualities: it shows 71.9% explained variance while leave-one-out cross-validation R^2 is found to be 66.0%. The view of molecular field and the interaction points is shown in Figure 5. The developed model was subjected to a randomization test (Table 7) at a 99% confidence level. The calculated and leave-one-out cross-validation values of binding affinity according to eq 7 are given in Table 2. Figure 2(m),(n) shows scatter plots of observed versus leave-15%-out and leave-25%-out cross-validation test, and the results (Q^2) are also presented in Table 8. The predictive R^2 value for the test set was found to be 0.327. However, on deletion of two compounds (**20** and **63**) from the test set, r^2_{Pred} value rises to 0.500.

The comparative statistics of all the models developed are presented in Table 9.

4. CONCLUSIONS

The Hansch analysis of the CCR5 antagonist 3-(4-benzylpiperidin-1-yl)propyl congeners unveiled the importance of the lipophilicity of the *meta* substituents of the R_2 position and the *para* substituents of the R_2 position for the binding affinity. The presence of electron-donating substituents at the *para* position of the R_2 position is conducive for the binding affinity. The best model with acceptable statistical quality was derived from the MSA, showing the importance of the relative negative charge (RNCG), where substituents with a high RNCG value have more binding affinity than the unsubstituted piperidine and phenylamino (R_1 position) congeners. The relative negative charge surface area (RNCS) is detrimental (e.g. R_2 = 3,4-Cl_2) for the activity. An increase in the length of the molecule in the Z dimension (Lz) is conducive (e.g. R_3 = sulfonylmorpholino), while an increase in the area of the molecular shadow in the XZ plane (Sxz) is detrimental (e.g. R_1 = *N*-*c*-hexylmethyl-

5-oxopyrrolidin-3-yl) for the binding affinity. The presence of a chiral center makes the molecule less active (e.g. R_1 = 5-oxopyrrolidin-3-yl or piperidin-3-yl). An increase in the van der Waals area, the molecular volume, and the difference between the volume of the individual molecule and the shape reference compound are conducive (e.g. $R_3 = (CH_3)_2NSO_2-$) for the binding affinity. Substituents with higher JursFPSA_2 values (fractional charged partial surface area) like the *N*-methylsulfonylpiperidin-4-yl (R_1 position) group have better binding affinity than the substituents such as 4-chlorophenylamino (R_1 position). Unsubstituted piperidines with less JursFNSA_1 values have lower binding affinity than the 4-chlorophenyl substituted compounds.

The receptor surface model gives information about the steric nature of the receptor site and the associated properties of interest, such as hydrophobicity, partial charge, and hydrogen-bonding propensity. The MFA derived equation shows interaction energies at different grid points.

The graphical representations of the beneficial and nonbeneficial interactions allow medicinal chemist to design new structures. All the 3D features (i.e., electrostatic, hydrophobic, and hydrogen bonding) can be easily identified from the contour maps derived from the 3D-QSAR, and the substitutional requirements can also be identified for the favorable receptor-drug interaction.

The developed models were validated through the leave-one-out, leave-15%-out, and leave-25%-out cross-validation techniques. The developed models were also subjected to randomization test (99% confidence level). Although the MSA derived models had excellent statistical qualities both for the training as well as test sets, RSA and MFA results for the test sets are not comparable statistically with the MSA derived models. This shows the capability of flexible modeling techniques and warns one to use the commercial modeling packages with adequate validation strategies.

Abbreviations: quantitative structure−activity relationships (QSAR); acquired immuno deficiency syndrome (AIDS); human immunodeficiency virus (HIV); linear free energy related (LFER).

ACKNOWLEDGMENT

One of the authors (J.T.L.) thanks the AICTE, New Delhi for a QIP fellowship. K.R. thanks the AICTE, New Delhi for a financial grant under the Career Award for Young Teachers scheme.

REFERENCES AND NOTES

(1) Campiani, G.; Ramunno, A.; Maga, G.; Nacci, V.; Fattorusso, C.; Catalanotti, B.; Morelli, E.; Novellino, E. Nonnucleoside HIV-1 reverse transcriptase (RT) inhibitors: past, present, and future perspectives. *Curr. Pharm. Des*. **2002**, *8*, 615−657.

(2) http://www.unaids.org/Unaids/EN/Resources/Publications.

(3) Jiang, S.; Zhao, Q.; Debanth, A. K. Peptide and non-peptide HIV fusion inhibitors. *Curr. Pharm. Des*. **2002**, *8*, 563−580.

(4) Sanders, R. W.; Dankers, M. M.; Busser, E.; Caffrey, M.; Moore, J. P.; Berkhout, B. Evolution of the HIV-1 envelope glycoproteins with a disulfide bond between gp120 and gp41. *Retrovirology* **2004**, *1*, 3−13.

(5) Markovic, I.; Clouse, K. A. Recent Advances in Understanding the Molecular Mechanisms of HIV-1 Entry and Fusion: Revisiting Current Targets and Considering New Options for Therapeutic Intervention *Curr. HIV Res.* **2004**, *2*, 223−234.

(6) Mager, P. P. The active site of HIV-1 protease. *Med. Res. Rev*. **2001**, *21*, 348−351.

(7) Farber, J. M.; Berger, E. A. HIV's response to a CCR5 inhibitor: I'd rather tighten than switch. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1749−1751.

(8) Richman, D. D. HIV chemotherapy. *Nature* **2001**, *410*, 995−1001.

(9) Kazmierski, W.; Bifulco, N.; Yang, H.; Boone, L.; DeAnda, F.; Watson, C.; Kenakin, T. Recent progress in discovery of small-molecule CCR5 chemokine receptor ligands as HIV-1 inhibitors. *Bioorg. Med. Chem*. **2003**, *11*, 2663−2676.

(10) Tsamis, F.; Gavrilov, S.; Kajumo, F.; Seibert, C.; Kuhmann, S.; Ketas, T.; Trkola, A.; Palani, A.; Clader, J. W.; Tagat, J. R.; McCombie, S.; Baroudy, B.; Moore, J. P.; Sakmar, T. P.; Dragic, T. Analysis of the mechanism by which the small-molecule CCR5 antagonists SCH-351125 and SCH-350581 inhibit human immunodeficiency virus type 1 entry. *J. Virol*. **2003**, *77*, 5201−5208.

(11) Xu, G.; Kannan, A.; Hartman, T. L.; Wargo, H.; Watson, K.; Turpin, J. A.; Buckheit, R. W., Jr.; Johnson, A. A.; Pommier, Y.; Cushman, M. Synthesis of substituted diarylmethylenepiperidines (DAMPs), a novel class of anti-HIV agents. *Bioorg. Med. Chem.* **2002**, *10*, 2807−2816.

(12) Stevens, M.; Pannecouque, C.; DeClercq, E.; Balzarini, J. Novel human immunodeficiency virus (HIV) inhibitors that have a dual mode of anti-HIV action. *Antimicrob. Agents Chemother.* **2003**, *47*, 3109−3116.

(13) Debnath, A. K. Generation of predictive pharmacophore models for CCR5 antagonists: study with piperidine- and piperazine-based compounds as a new class of HIV-1 entry inhibitors. *J. Med. Chem*. **2003**, *46*, 4501−4515.

(14) Xu, Y.; Liu, H.; Niu, C.; Luo, C.; Luo, X.; Shen, J.; Chen, K.; Jiang, H. Molecular docking and 3D QSAR studies on 1-amino-2-phenyl-4-(piperidin-1-yl)-butanes based on the structural modeling of human CCR5 receptor. *Bioorg. Med. Chem.* **2004**, *12*, 6193−6208.

(15) Song, M.; Breneman, C. M.; Sukumar, N. Three-dimensional quantitative structure−activity relationship analyses of piperidine-based CCR5 receptor antagonists. *Bioorg. Med. Chem*. **2004** *12*, 489−499.

(16) Leonard, J. T.; Roy, K. Classical QSAR modeling of CCR5 receptor binding affinity of substituted benzylpyrazoles. *QSAR Comb. Sci.* **2004**, *23*, 387−398.

(17) Leonard, J. T.; Roy, K. Classical QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners. *QSAR Comb. Sci*. **2004**, *23*, 23−35.

(18) Roy, K.; Leonard, J. T. QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners using molecular connectivity and E-sate parameters. *Bioorg. Med. Chem.* **2004**, *12*, 745−754.

(19) Leonard, J. T.; Roy, K. QSAR modeling of anti-HIV activities of alkenyldiarylmethanes using topological and physiochemical descriptors. *Drug Des. Dicovery* **2003**, *18*, 165−180.

(20) Roy, K.; Leonard, J. T. QSAR by LFER Model of Cytotoxicity Data of Anti-HIV 5-Phenyl-1-phenylamino-1*H*-imidazole Derivatives Using Principal Component Factor Analysis and Genetic Function Approximation. *Bioorg. Med. Chem.* **2005**, *13*, 2967−2973.

(21) Roy, K.; Leonard, J. T. Topological QSAR modeling of cytotoxicity data of anti-HIV 5-phenyl-1-phenylamino-imidazole derivatives using GFA, G/PLS, FA and PCRA techniques. *Indian J. Chem*. **2005**, *44A*, accepted.

(22) Imamura, S.; Nishikawa, Y.; Ichikawa, T.; Hattori, T.; Matsushita, Y.; Hashiguchi, S.; Kanzaki, N.; Iizawa, Y.; Baba, M.; Sugihara, Y. CCR5 antagonist as anti-HIV-1 agents. Part 3: Synthesis and biological evaluation of piperidin-4-carboxamide derivatives. *Bioorg. Med. Chem.* **2005**, *13*, 397−416.

(23) Imamura, S.; Kurasawa, O.; Nara, Y.; Ichikawa, T.; Nishikawa, Y.; Iida, T.; Hashiguchi, S.; Kanzaki, N.; Iizawa, Y.; Baba, M.; Sugihara, Y. CCR5 antagonist as anti-HIV-1 agents. Part 2: Synthesis and biological evaluation of *N*-[3-(4-benzylpiperidin-1-yl)propyl]-*N*,*N*′-diphenylureas. *Bioorg. Med. Chem*. **2004**, *12*, 2295−306.

(24) Imamura, S.; Ishihara, Y.; Hattori, T.; Kurasawa, O.; Matsushita, Y.; Sugihara, Y.; Kanzaki, N.; Iizawa, Y.; Baba, M.; Hashiguchi, S. CCR5 Antagonist as Anti-HIV-1 agents. 1. Synthesis and Biological Evaluation of 5-Oxopyrrolidine-3-carboxamide Derivatives. *Chem. Pharm. Bull.* **2004**, *52*, 63−73.

(25) Hansch, C.; Fujita, T. ρ σ π Analysis: A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc*. **1964**, *86*, 1616−1626.

(26) Kubinyi, H. In *Methods and Principles in Medicinal Chemistry*; Wolff, M. E., Ed.; VCH: Weinheim, 1993; Vol. 1, pp 497−571.

(27) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(28) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.

(29) Franke, R.; Gruska, A. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, pp 113−163.

(30) Cerius2 version 4.8 is a product of Accelrys, Inc., San Diego, U.S.A., http://www.accelrys.com/cerius2.
(31) Hopfinger, A. J.; Tokarsi, J. S. In *Practical Applications of Computer-Aided Drug Design*; Charifson, P. S., Ed.; Marcel Dekker: New York, 1997; pp 105−164.
(32) Hahn, M. Receptor Surface Models. 1. Definition and Construction. *J. Med. Chem*. **1995**, *38*, 2080−2090.
(33) Rogers, D.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure−activity relationships and quantitative structure−property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(34) Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. Quantitative structure-antitumor activity relationships of camptothecin analogues: cluster analysis and genetic algorithm-based studies. *J. Med. Chem*. **2001**, *44*, 3254−3263.
(35) Wold, S. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, pp 195−218.
(36) SPSS is a statistical software of SPSS Inc., IL.
(37) Snedecor, G. W.; Cochran, W. G. *Statistical Methods*; Oxford & IBH Publishing Co. Pvt. Ltd.: New Delhi, 1967.
(38) Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1995; Vol. 2, pp 312−317.
(39) Debnath, A. K. In *Combinatorial Library design and Evaluation*; Ghose, A. K., Viswanadhan, V. M., Eds.; Marcel Dekker: New York, 2001; pp 73−129.
(40) Patrick, G. L. *An Introduction to Medicinal Chemistry*; Oxford University Press Inc.: New York, 2001.

CI050205X