STUDIA UBB CHEMIA, LXIII, 1, 2018 (p. 177-188)
(RECOMMENDED CITATION)
DOI:10.24193/subbchem.2018.1.13

C-C CHEMOKINE RECEPTOR TYPE 3 INHIBITORS: BIOACTIVITY PREDICTION USING LOCAL VERTEX INVARIANTS BASED ON THERMAL CONDUCTIVITY LAYER MATRIX

CLAUDIU N. LUNGU*

ABSTRACT. A series of compounds with known inhibitory activity for C-C chemokine receptor type 3 (CCR_3) was considered in order to build a predictive model useful in further development of novel CCR_3 inhibitors. Model was built using topological descriptors (Cluj indices included) and multiple linear regression. Principal component analysis was applied in order to enhance the model. Errors were taken into consideration and discussed. Finally, vertex invariants based on thermal conductivity layer matrix proved to be a valuable tool in bioactivity prediction of CCR_3 inhibitors.

Keywords: *Topological descriptors; QSAR; Regeression model; CCR3 inhibitors.*

INTRODUCTION

C-C chemokine receptor type 3 (CCR_3), recently designated cluster for differentiating CD193, is highly expressed in eosinophils [1]. Its function is the accumulation and activation of eosinophils at the site of an immune stimulus (allergy, parasitic infection, etc.). A chemokine inhibitor will prohibit this proinflammatory effect. A series of broad spectrum chemokine inhibitors (BSCI) were designed. Remarkably, peptide 3'', a dodecapeptide section of chemokine (C-C-motif) ligand 2 (CCL2) was proven to be a functional inhibitor of many chemokines [2]. The crucial peptide sequence responsible for its

* *Department of Chemistry, Faculty of Chemistry and Chemical Engineering, Babes-Bolyai University, 400028 Cluj, Romania. E-mail: lunguclaudiu5555@gmail.com.*

effect was identified as the tripeptide AcNH-Trp-Val-Gln-OH. Peptide mimetics were fashioned mainly a range of 3-acylaminoglutarimides, with low nM BSCI potencies [3]. A particular interest exists for cyclic peptide NR58-3.14. CCR_3 was shown to inhibit HIV replication [4]. In this respect, there is an entry co-receptor for HIV-1. This gene receptor and other chemokine receptors genes form a cluster on chromosomal region 3p21 [5]. A number of small molecules with inhibitor effect on CCR_3 (expressed as IC50, nM) were studied. In this work, a prediction model based on computed data regarding molecules that act as inhibitors of CCR_3 was carried out, in order to be used as a tool for further development of these compounds.

RESULTS AND DISCUSSION

A set of 41 compounds with inhibitory effect on CCR_3 was selected from PubChem (Table 1), on which topological indices [6], encoding fragmental topo-chemical information [7-9], were computed by means of layer matrices [11-14]. The fragmental properties were chosen the mass M and thermal conductivity Tc (Table 2).

In order to select the descriptors for model building, a monovariate correlation between each descriptor and the dependent variable (IC_{50}) was performed: X[LM[Tc]] has r^2=0.335 followed by C[LM[Tc]] r^2=0.240 and C[LM[Mass]] r^2=0.203. The other descriptors have r^2 values below 0.1. A multiple regression model [13] was computed using all mass and thermal conductivity descriptors. Dependent variable was chosen IC_{50}. Pearson correlation (r) retrieved was r=0.632, with r^2=0.399, Spearman rank correlation (p)=0.673, cross validated squared (q^2) =0.399, y=0.399x+12.11. Based on PCA, principal components (PC) 1-8 for variables listed in Table 2 were computed. Results are shown in Table 3.

In order to build a model using descriptors listed in Table 1 and PC showed in Table 3, a multi-variate correlation between target variable and descriptors was performed. Conclusive results are shown in Table 4.

A model using correlation (entry 5) (Table 4) was built with IC_{50} (nM) as dependent variable. This solution was chosen mainly because of its correlation with the target variable (r^2=0.999) and the reduced number of descriptors (4). The plot of correlation between predicted IC_{50}p and IC_{50}o experimentally determined (observed) is shown in Fig. 1. In Table 5 descriptors contribution to the model is listed. Model equation is y=0.029+0.998*IC_{50}o. Person correlation square (r^2)=0.998,Speareman rank correlation(p)=0.931, Cross validated square (q^2)=0.998.

Table 1. Compounds (in Smiles) used in building of the model, $IC_{50}o$ (observed), $IC_{50}p$ (predicted)

No	Compound	$IC_{50}o$	$IC_{50}p$
1	ON(=O)C(=CC=1)C=CC=1CC(C(OC)=O)NC(=O)C=2C=CC=C3C=CC=CC=23	5.000	4.987
2	FC(=CC=1)C=CC=1CC(CC2)CCN2CC(C3)NCCN3C(=O)NC(C=4)=CC=CC=4OC	30.000	29.986
3	FC(=CC=1)C=CC=1CC(CC2)CCN2CC(C3)CCCN3C(=O)NC(C=4)=CC=CC=4OC	50.000	50.006
4	O=C(NC=1C=C(C=CC=1)C(C)=O)NCCCN(C(C(CCCC)O)C2)CCC2CC3=CC=CC=C3	80.000	80.016
5	ClC(=CC=1)C=CC=1CC(CC2)CCN2CCCNC(=O)NC(=C3)C=CC=C3C#N	20.000	20.051
6	O=C(NC=1C=C(C=CC=1)C(C)=O)NCCCN(C(C(CC)O)C2)CCC2CC3=CC=CC=C3	8.000	7.981
7	FC(=CC=1)C=CC=1CC(CC2)CCN2CCCNC(=O)NC(C=3)=CC=CC=3C#N	80.000	80.053
8	FC(=CC=1)C=CC=1CC(CC2)CCN2CC(C3)OCCN3C(=O)NC(C=4)=CC=CC=4OC	50.000	49.996
9	O=C(NC=1C=C(C=CC=1)C(C)=O)NCCCN(C(CCC)C2)CCC2CC3=CC=CC=C3	100.000	100.017
10	ClC(C=C1)=CC=C1CC(C2)CC3CCC2N3CC(C(C)O)NC(=O)NC(C=4)=CC(OC)=C(OC)C=4OC	8.200	8.192
11	ClC(C=C21)=CC=C2OC3=CC=C(Cl)C=C3C1C(=O)NC(CC4)CCN4(CC)CC=5CCCCCCC=5	0.580	0.652
12	ClC(C=C1)=CC=C1CC(C2)CC3CCC2N3CC(C(C)O)NC(=O)NC(C=4)=CC=CC=4S(=O)(=O)C	12.000	11.963
13	ClC(C=C1)=CC=C1CC(C2)CC3CCC2N3CC(C)NC(=O)NC(C=4)=CC(OC)=C(OC)C=4OC	11.000	11.003
14	ClC(C=C1)=CC=C1CC(C2)CC3CCC2N3CC(C)NC(=O)NC(C=4)=CC=CC=4S(=O)(=O)C	65.000	64.980
15	FC(=CC=1)C=CC=1CC(C2)CC3CCC2N3CCCNC(=O)NC(C=4)=CC=CC=4C(=O)C	47.000	46.957
16	FC(C=C1)=CC=C1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(C=4)=CC=CC=4C(=O)C	0.364	0.347
17	ClC=1C=CC=C(S2)C=1N=C2NC(=O)NC3CCCCC3CN(C4)CCCC4CC=5C=CC(F)=CC=5	0.090	0.084
18	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC=4C=CC=CC=4C	100.000	99.998
19	FC(=CC=1)C=CC=1NC(=O)NC2CCCCC2CN(C3)CCCC3CC=4C=CC(F)=CC=4	5.000	5.004
20	BrC(=CC=1)C=CC=1NC(=O)NC2CCCCC2CN(C3)CCCC3CC=4C=CC(F)=CC=4	50.000	49.967
21	FC(=C1)C(C)=CC=C1NC(=O)NC2CCCCC2CN(C3)CCCC3CC=4C=CC(F)=CC=4	35.000	34.963
22	FC(=CC=1)C=CC=1CC2CCCCN2NC(=O)NC(C=3)=CC=CC=3C(=O)C	2.600	2.564
23	FC(=CC=1)C=CC=1CC(CC2)CCN2CC3=C(C(NC)=O)C=CC=C3NC(=O)NC(=C4)C=CC=C4C(=O)C	30.000	29.987
24	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC=4C=CC=CC=4	17.000	16.946
25	FC(C=C1)=CC=C1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(=CC=4)C=CC=4OC	1.000	0.996
26	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(C=4)=CC=CC=4N5N=NN=C5C	1.400	1.334
27	FC(C(=C1)C(C)=O)=CC=C1NC(=O)NC2CCCCC2CN(C3)CCCC3CC=4C=CC(F)=CC=4	0.015	-0.017
28	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC4=CC=C5NN=CC5=C4	0.045	0.053
29	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(=C4)C=C(C=5N(C)N=NN=5)C=C4C6=NN=NN6C	0.042	0.009
30	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(S4)=NC(C)=C4C(=O)C	0.030	0.071
31	FC(C=C1)=CC=C1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(C=C4)=CC=C4C5=NN=NN5C	0.700	0.667
32	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC4=CC=C5NC=CC5=C4	0.400	0.364
33	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(C=C4(C#N))CC=C4N5C=CC=N5	4.000	4.049
34	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC=4C=CC=CC=4OC	1.400	1.394
35	FC(C=C1)=CC=C1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(=C4)C=C(C(C)=O)C=C4C(=O)C	0.007	0.049
36	FC(=CC=1)C=CC=1CC(C2)CCCN2CCCNC(=O)NC(C=3)=CC=CC=3C(=O)C	2.500	2.509
37	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCCCC3NC(=O)NC(C=4)=CC=CC=4C5=NN=NN5C	0.010	-0.003
38	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCN(C(C)=O)CC3NC(=O)NC(S4)=NC(C)=C4C(=O)C	7.000	7.054
39	FC(C=C1)=CC=C1CC(C2)CCCN2CC3CS(=O)(=O)CC3NC(=O)NC(S4)=NC(C)=C4C(=O)C	0.900	0.987
40	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CNCCC3NC(=O)NC(S4)=NC(C)=C4C(=O)C	0.100	0.150
41	FC(=CC=1)C=CC=1CC(C2)CCCN2CC3CCNCC3NC(=O)NC(S4)=NC(C)=C4C(=O)C	0.075	0.093

Table 2. Descriptors used in building of the model: C-Centrality, X-Centrocomplexity, CS-Colum Sum, PDS-Product Distance Sum, LM-Layer matrix, [M]-Mass, [Tc]- Thermal conductivity

No.	C[LM[M]]	C[LM[Tc]]	CS[LM[M]]	CS[LM[Tc]]	PDS[LM[TM[M]]	PDS[LM[Tc]]	X[LM[M]].10^{-3}	X[LM[Tc]].10
1	2.193	2.698	10080	758.572	55964	4036.730	10.212	126.836
2	1.925	2.195	13024	1032.060	93054	7345.680	10.219	125.679
3	1.925	2.192	12960	1073.330	92670	7593.310	10.219	125.585
4	2.100	2.368	15330	1309.410	114510	9889.530	10.211	125.105
5	1.720	2.011	11107	860.470	82491	6187.130	10.208	125.732
6	2.004	2.271	13662	1149.440	99072	8421.480	10.212	125.318
7	1.724	2.008	10643	860.476	77707	6187.190	10.213	125.732
8	1.924	2.195	13088	1032.060	93438	7345.680	10.219	125.679
9	1.944	2.202	12736	1114.600	92362	8172.440	10.213	124.865
10	2.203	2.508	18241	1336.520	135797	9688.480	10.212	126.324
11	2.411	2.807	17640	1393.240	117398	9161.840	1.002	126.600
12	2.160	2.493	16975	1174.060	123295	8114.060	10.219	126.623
13	2.072	2.368	16275	1219.120	121519	8844.720	10.213	126.023
14	2.031	2.352	15081	1064.390	109761	7350.250	10.220	126.310
15	1.897	2.165	12960	1073.330	95809	7822.950	10.218	126.407
16	2.115	2.419	14586	1228.130	102979	8573.730	10.217	126.139
17	2.150	2.496	16870	1219.200	122363	8690.510	1.002	127.098
18	2.088	2.403	12832	1114.60	85200	7397.140	10.219	125.827
19	2.031	2.356	13056	1073.330	88632	7108.270	10.217	126.346
20	2.019	2.351	15008	1073.360	107786	7108.560	1.002	126.346
21	2.075	2.372	13860	1149.440	95676	7833.250	10.216	126.278
22	2.013	2.420	9315	731.473	53579	4200.480	10.213	126.021
23	2.275	2.569	18354	1470.680	136021	10777.300	10.214	126.380
24	2.042	2.348	12059	1039.780	78691	6766.310	10.219	125.877
25	2.030	2.317	13761	1149.440	96222	7930.010	10.218	125.831
26	2.193	2.493	17353	1336.530	130849	9820.180	10.222	126.191
27	2.155	2.473	15680	1264.270	113277	8924.710	10.215	126.579
28	2.115	2.458	14586	1184.280	102743	8146.820	10.223	126.765
29	2.444	2.813	23607	1664.250	189543	12992.800	1.002	126.598
30	2.112	2.449	15334	1140.510	108185	7941.050	10.216	126.355
31	2.125	2.429	17353	1336.530	134811	10000.800	10.222	126.262
32	2.116	2.420	14518	1228.130	102105	8558.250	10.223	126.603
33	2.175	2.473	18202	1470.680	140565	11222.300	10.221	126.365
34	2.128	2.455	13761	1149.440	93554	7762.290	10.218	125.832
35	2.242	2.571	17353	1431.970	128555	10461.100	10.215	126.367
36	1.829	2.119	11430	928.848	82195	6612.890	10.212	125.634
37	2.193	2.519	17353	1336.530	130909	9781.480	10.222	126.263
38	2.280	2.634	18241	1288.900	131569	9199.060	10.215	126.300
39	2.175	2.543	17185	1083.870	120554	7588.320	10.222	126.867
40	2.111	2.464	15402	1096.660	108635	7650.860	10.216	126.520
41	2.111	2.452	15402	1096.660	108639	7648.280	10.216	126.520

Table 3. Principal components used in building the model.
PC 1-8 for the descriptors in Table 2

No.	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
1	2.008	-4.022	-0.198	-1.783	-0.018	-0.092	0.272	-0.023
2	2.162	0.629	0.040	0.459	-0.263	-0.139	-0.056	0.041
3	2.229	1.184	-0.213	0.084	-0.088	0.011	-0.055	-0.007
4	0.226	2.994	-0.303	-1.2542	-0.482	-0.029	0.097	0.045
5	3.988	0.188	-0.005	1.622	-0.060	-0.210	0.095	-0.060
6	1.188	0.991	0.545	0.408	-1.332	-0.086	-0.001	0.058
7	4.535	1.101	-0.914	0.472	0.902	0.012	0.133	-0.023
8	2.283	0.966	-0.259	0.105	0.088	-0.141	-0.052	0.039
9	2.693	2.968	-0.841	-1.116	-0.396	-0.113	0.046	0.018
10	-2.184	0.457	0.700	-0.014	0.227	-0.299	-0.101	-0.014
11	-3.772	-1.335	-2.212	-0.615	-1.150	0.369	-0.159	-0.022
12	-1.124	-0.571	0.395	-0.075	0.728	-0.455	-0.128	-0.029
13	-0.476	0.588	0.542	0.413	-0.072	-0.300	-0.050	-0.034
14	0.872	0.442	-0.504	-0.546	1.278	-0.409	-0.053	-0.047
15	1.823	0.372	-0.218	0.630	1.186	0.525	0.080	0.057
16	-0.236	-0.188	0.637	0.091	-0.429	0.341	-0.044	-0.005
17	-2.263	-1.455	-2.342	1.327	0.259	0.320	0.024	0.098
18	1.711	1.121	-1.010	-1.993	0.673	0.298	-0.040	-0.019
19	1.101	-0.967	0.336	0.242	0.001	0.288	-0.065	-0.045
20	0.202	-0.247	-3.236	0.717	0.117	-0.171	-0.124	-0.075
21	0.658	-0.065	-0.001	-0.367	0.408	0.366	-0.087	0.017
22	3.576	-2.531	-0.070	-0.629	-0.890	-0.269	0.120	0.025
23	-2.886	1.046	0.514	-0.721	0.482	0.312	-0.034	0.024
24	1.851	-0.530	0.138	-0.295	-0.631	0.206	-0.112	0.009
25	0.862	0.049	0.562	0.421	-0.755	0.126	-0.118	-0.023
26	-1.927	0.463	0.816	0.071	-0.160	-0.103	0.004	0.012
27	-1.137	-0.520	0.662	0.116	0.266	0.371	-0.002	-0.001
28	-0.447	-1.140	0.528	0.154	0.504	0.485	0.034	-0.030
29	-7.242	1.021	-1.646	0.449	-0.434	-0.440	0.259	-0.001
30	-0.277	-0.677	0.539	0.126	0.006	-0.165	-0.042	-5.680
31	-1.768	0.642	0.843	0.573	0.102	-0.055	0.091	-0.051
32	-0.446	-0.730	0.594	0.225	0.262	0.654	-0.024	0.030
33	-2.756	1.057	0.952	0.425	0.237	0.378	0.133	-0.012
34	0.421	-0.360	0.530	-0.355	-1.004	0.045	-0.062	-0.020
35	-2.595	0.330	0.902	-0.214	-0.117	0.408	0.085	-0.041
36	3.160	-0.072	0.327	1.269	-0.753	-0.116	0.060	-0.001
37	-2.015	0.306	0.823	0.041	-0.097	-0.089	0.059	-0.044
38	-2.356	-0.073	0.642	-0.701	-0.038	-0.566	-0.057	0.006
39	-1.151	-1.416	0.447	-0.045	0.883	-0.706	-0.080	0.083
40	-0.257	-1.019	0.478	0.120	0.275	-0.284	-0.004	0.021
41	-0.231	-0.998	0.481	0.158	0.286	-0.275	-0.036	0.045

Table 4. Model correlation in multi-variate regression

No.	Pearson r^2	Descriptors	Descriptors
1	0.999	6	C[LM[Mass]],PC2,PC3,PC7, PDS[LM[Tc]],X[LM[Mass]]
2	0.999	6	CS[LM[Tc]],C[LM[Mass]],PC2,PC3,PDs[LM[Tc]],X[LM[Mass]]
3	0.999	6	CS[LM[Mass]],C[LM[Mass],PC2,PC3,PDS[LM[Tc]], X[LM{Mass]]
4	0.999	5	C[LM[Mass]],PC2,PC3,PDS[LM[Tc]],X[LM[Mass]]
5	0.999	4	PC2,PC3,PDS[LM[Tc]], X[LM[Mass]]
6	0.798	3	PC2,PC3,PDS[LM[Tc]]
7	0.680	2	PC2, PDS[LM[Tc]]
8	0.627	2	CS[LM[Tc]], PC2
9	0.615	2	PC2,PDS[LM[Mass]]
10	0.371	1	PC2
11	0.172	1	PC3
12	0.074	1	PDS[LM[Mass]]

Table 5. Tolerance, VIF (value of inflation) and r^2 calculated for the descriptors used in the model

No.	Descriptors	r^2	Tolerance	VIF
1	PC2	0.777	0.222	4.485
2	PC3	0.936	0.063	15.645
3	PDS[LM[Tc]]	0.860	0.140	7.149
4	X[LM[Mass]]	0.944	0.056	17.860

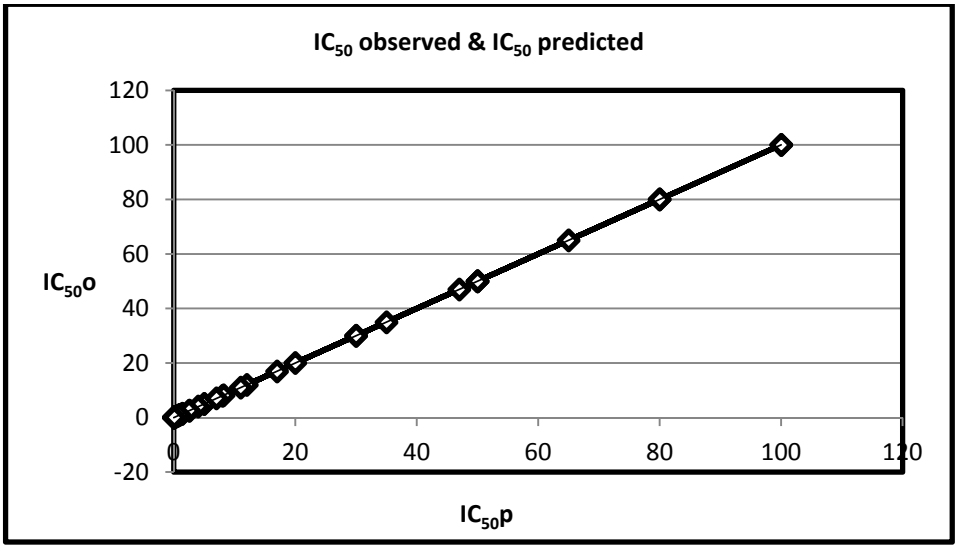

Figure 1. Scatter plot of IC_{50} observed ($IC_{50}o$) vs IC_{50} predicted ($IC_{50}p$)

The results obtained by using simple descriptors alone (without PC values) are rather modest. Best r^2 value described in literature is 0.960 on a series of 4-benzylpiperidinealkylurease CCR3 antagonist with a model obtained by using CoMFA methodology [10]. However, the cited study used a vast set of descriptors. The approach by local vertex invariants based on thermal conductivity descriptors, within Cluj descriptors, has the advantage of using a small number of descriptors which encode both topological and physico-chemical properties of a ligand. Correlation obtained supports the supposition that thermal conductivity and topological information plays a distinct role in bioactivity of these compounds. Furthermore, by using PCA, a significant increase in predictive performance of the model is obtained. Certainly, this model is not the best CCR3-interaction estimator but regardless of collinearity and multi-collinearity introduced by using PCA, the model is a good predictor of bioactivity of the studied compounds, as demonstrated below. In Table 6, standard errors are listed. Standard error is a measure of divergence of estimator value from the expected value.

Average standard deviation computed was 19.665. Percentage error and percent difference error, calculated for observed and predicted values are listed in Table 6.

Multicollinearity has no impact on the regression model and association statistics such as r^2, F ratio and p values. It also should not generally have an impact on prediction made using the overall model. Tolerance and VIF are two main tests in detecting multicollinearity. Results are represented in Tables 7 and 8.

Fig. 2 shows data dimensionality, although r^2=0.999 variations are observed. For two variables 1-100% interval is violated. In this case there is a lack of accuracy and precision.

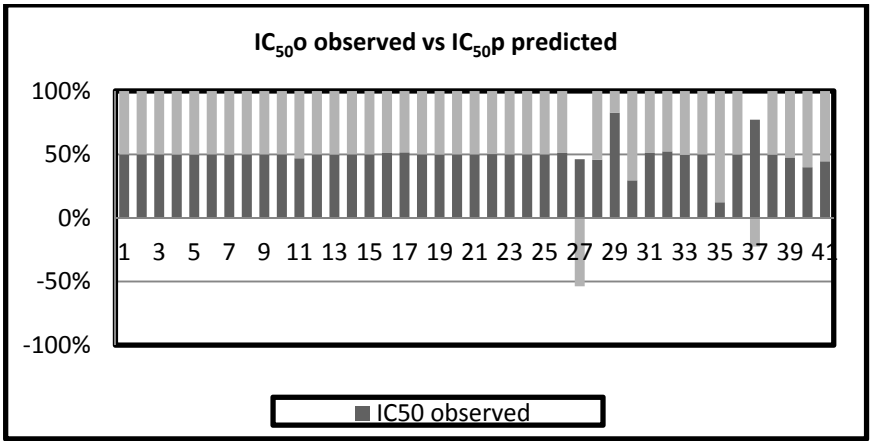

Figure 2. $IC_{50}o$ vs $IC_{50}p$ % stacked column.

Table 6. Percentage error, percent difference error, calculated for observed and predicted values

No.	$IC_{50}o$	$IC_{50}p$	%error	%difference	%error predicted	Error SD
1	5.000	4.987	0.256	0.128	0.637	15.170
2	30.000	29.986	0.045	0.022	0.113	-29.986
3	50.000	50.005	0.011	0.005	0.027	-50.005
4	80.000	80.015	0.019	0.009	0.048	-80.015
5	20.000	20.051	0.256	0.128	0.635	-20.051
6	8.000	7.980	0.240	0.120	0.596	-7.980
7	80.000	80.052	0.067	0.032	0.163	-80.052
8	50.000	49.996	0.007	0.003	0.018	-49.996
9	100.000	100.017	0.017	0.008	0.042	-100.017
10	8.200	8.191	0.099	0.049	0.246	-8.191
11	0.580	0.651	12.408	5.841	28.979	-0.651
12	12.000	11.963	0.305	0.152	0.757	-11.963
13	11.000	11.002	0.024	0.012	0.060	-11.002
14	65.000	64.980	0.030	0.015	0.075	-64.980
15	47.000	46.957	0.091	0.045	0.227	-46.957
16	0.364	0.347	4.568	2.337	11.597	-0.347
17	0.090	0.083	6.7664	3.501	17.371	-0.083
18	100.000	99.998	0.002	0.001	0.004	-99.998
19	5.000	5.004	0.082	0.041	0.205	-5.004
20	50.000	49.967	0.065	0.032	0.161	-49.967
21	35.000	34.962	0.107	0.053	0.265	-34.962
22	2.600	2.563	1.390	0.700	3.473	-2.563
23	30.000	29.987	0.043	0.021	0.106	-29.987
24	17.000	16.945	0.318	0.159	0.790	-16.945
25	1.000	0.996	0.364	0.182	0.905	-0.996
26	1.400	1.333	4.718	2.416	11.987	-1.334
27	0.015	-0.017	215.900	1357.810	6736.000	0.017
28	0.045	0.053	17.901	8.215	40.756	-0.053
29	0.042	0.008	79.458	65.918	327.016	-0.008
30	0.030	0.071	137.365	40.717	201.994	-0.071
31	0.700	0.666	4.750	2.433	12.070	-0.666
32	0.400	0.363	9.124	4.780	23.713	-0.363
33	4.000	4.048	1.215	0.604	2.996	-4.048
34	1.400	1.393	0.461	0.231	1.147	-1.393
35	0.007	0.049	604.554	75.141	372.770	-0.049
36	2.500	2.508	0.340	0.170	-0.843	-2.508
37	0.010	-0.003	129.130	182.208	903.923	0.003
38	7.000	7.054	0.773	0.385	1.910	-7.054
39	0.900	0.987	9.681	4.617	22.904	-0.987
40	0.100	0.150	50.025	20.008	99.258	-0.150
41	0.075	0.093	24.295	10.832	53.736	-0.093

Table 7. PCA and topological descriptors combined model VIF and Tolerance

Tolerance calculus: Variable	R^2	Tolerance (1-R^2)	VIF(1/Tolerance)
PDS[LM[Tc]]	0.983	0.016	62.282
C[LM[Mass]]	0.990	0.009	110.192
X[LM[Mass]]	0.988	0.011	89.015
PC2	0.967	0.032	30.700
PC3	0.983	0.016	59.234
PC4	0.973	0.026	37.813
PC7	0.323	0.676	1.478
PC8	0.005	0.994	1.005

Table 8. PCA and topological descriptors combined model validation

No	Descriptors	p	Q^2	R^2	Adjuat R^2	St error	F ratio
1	PDS[LM[Tc]], C[LM[Mass]]	0.002	0.220	0.220	0.200	2.341	11.041
2	PDS[LM[Tc]], X[LM[Mass]]	0.258	0.032	0.032	0.007	18.172	1.313
3	PDS[LM[Tc]], PC2	0.812	0.681	0.680	0.672	3.424	83.251
4	PDS[LM[Tc]], PC3	0.469	0.188	0.188	0.167	7.885	9.076
5	PDS[LM[Tc]], PC4	0.339	0.168	0.168	0.147	8.321	7.904

Random errors evaluation is represented in Fig. 3; there are no randomization errors. However, there is a violation in data dimensionality for variables 27 and 37.

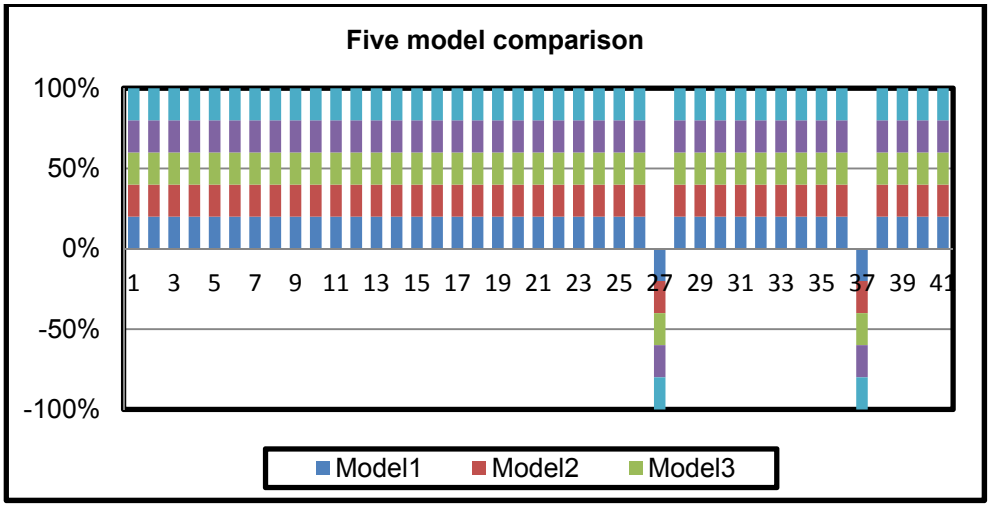

Figure 3. % stacked column plot of identical multiple regression models.

METHODS

Data set used to build the regression model was curated: 3D structures were built using Smiles formulas, compounds resulted were energetically minimized, only compounds with IC50(nM) equal or >100 nM were kept. After this process 41 compounds (Table 1) were retained. Compounds were randomly divided in a training set (21 compounds) and a test set (20 compounds). All structures were converted to "hin" format. The model was built by using TopoCluj topological descriptors. Layer matrices of two physico-chemical properties were used: mass and thermal conductivity of respective compounds (as included in data base of TopoCluj) Centrocomplexity, colum sum, product distance sum, as vertex invariants, were used to compute mass and thermal conductivity layer matrix indices. In order to increase the coefficient of correlation, principal components analysis (PCA) was performed [15]; PC 1-8 were computed. Dependent variable IC_{50} was predicted using combined PC and topological descriptors, applying the multiple linear regression method. The model was internally and externally validated. Contribution of descriptors to the model was evaluated based on variance, inflation factor (VIF) [16] and tolerance [17]; these were computed based on r^2. To find if a value is "accurate", a comparison with the acceptable value must be made; in this respect, percentage error (% error) was used. To find if a value is "precise", the average of IC_{50} predicted value was calculated. Average of the deviations was calculated to find the uncertainty. Also, VIF, Tolerance, p, q^2, r^2, adjust r^2, standard error, F ratio were computed in order to assess the model multicolinearity and validate the model.

Error analysis, in respect to observed and predicted IC_{50}, was performed using the percent error and percent difference, calculated by formulas:

$$\%\,\text{error} = 100\,\text{x}\,\frac{|(\text{value}\,2)| - |(\text{value}\,1)|}{|\,\text{value}\,2\,|}$$

$$\%\,\text{difference} = 100\,\text{x}\,\frac{\dfrac{|(\text{value}\,1)| - |(\text{value}\,2)|}{|(\text{value}\,1)| + |(\text{value}\,2)|}}{2}$$

where value 1 is IC_{50}o observed and value2 is IC_{50}p predicted. The two values are interchangeable due to allowance of sign neglection.

A percentage representation of predicted and observed values is used to show data dimensionality. Ideally, none predicted or observed values should pass the 1-100% limit. A perfect model has both observed and predicted values represented as same quantities (percent), i.e., 50/50%.

Randomized errors were also computed. In this respect, 5 models using same descriptors, same methods of model generation and evaluation were used. Models were compared using a stacked column plot.

CONCLUSIONS

Bioactivity prediction using local vertex invariants based on thermal conductivity layer matrix is a good methodology in predicting activity of C-C chemokine receptors type 3 inhibitors.

PCA is a valuable tool in optimizing a regression model. Accuracy and precision errors affecting data dimensionality were detected in the model. Such errors may occur even at high values of r^2. Random errors were uncommon in this type of model and were not observed.

ACKNOWLEDGMENTS

The author declares no conflict of interest.

REFERENCES

1. K.D. Dyer, K.E. Garcia-Crespo, K. E. Killoran and H. F. Rosenberg, *Journal of Immunological Methods*, **2011**, *369* (1-2), 91.
2. J. Reckless, D.J. Grainger, *Biochemical Journal* **1999**, *340* (3), 803.
3. D.J. Fox, J. Reckless, S.G. Warren, D.J. Grainger, *Journal of Medicinal Chemistry*, **2002**, *45* (2), 360.
4. D.J. Grainger, A.M. Lever, *Medicinal Chemistry,* **2005**, *2* (1), 23.
5. Entrez Gene, CCR3 chemokine (C-C motif) receptor 3.
6. H. González-Díaz, S. Vilar, L. Santana, E Uriarte, *Current Topics in Medicinal Chemistry*, **2007**, *7* (10), 1015.
7. L. Jäntschi, G. Katona, M.V. Diudea *MATCH Communications in Mathematical and Computer Chemistry* **2000**, *41*, 151.
8. P.V. Khadikar, S. Karmarkar, V.K. Agrawal, J. Singh, A. Shrivastava, I. Lukovits, M.V. Diudea, *Letters in Drug Design & Discovery*, **2005**, *2* (8), 606.
9. O.M. Minailiuc, G. Katona, M.V. Diudea, M. Strunje, A. Graovac, I. Gutman, *Croatica Chemical Acta*, **1998**, *71* (3), 473.
10. Jain V1, Pandey A, Gupta S, Mohan CG, *Journal of Molecular Modeling*, **2010**,*16* (4), 669.
11. M.V. Diudea, M. Topan, A. Graovac, *Journal Chemical Information and Computer Science,* **1994**, *34* (5), 1072.

12. M.V. Diudea, O. Ursu, *Indian Journal of Chemistry A,* **2003**, *42* (6), 1283.
13. M.V. Diudea, O.M. Minailiuc, G. Katona, I. Gutman, Szeged matrices and related numbers, *MATCH Communications in Mathematical and Computer Chemistry,* **1997**, *35*, 129.
14. M.H. Kutner, C.J. Nachtsheim, and J. Neter, “Applied Linear Regression Models”, 4th ed., McGraw-Hill/Irwin, Boston, 2004, p. 25.
15. H. Abdi, L.J. Williams, “Principal component analysis”. Wiley Interdisciplinary Reviews: Computational Statistics. **2010**, 2 (4), 433.
16. A.F. Zuur, E.N. Ieno, C.S. Elphick, *Methods in Ecology and Evolution*. **2010**, *1*, 3.
17. S. Menard, Applied Logistic Regression Analysis: Sage University Series on Quantitative Applications in the Social Sciences. Thousand Oaks, CA. Sage, 1995.