CARPATHIAN J. MATH.
20 (2004), No. 2, 275 - 280

Qsar Study On Dipeptide Ace Inhibitors

OLEG URSU, MARINELA DON, GABRIEL KATONA,
LORENTZ JÄNTSCHI and MIRCEA V. DIUDEA

ABSTRACT**.** Quantitative Structure-Activity Relationships (QSARs) establish a mathematical relation between the biological activity of chemical compounds and their molecular structure. They provide quantitative models aimed to accurately predict a certain activity from the structural attributes. This topic has become a well-delimited branch in chemistry and was favored by the progress in computer science. Cluj property indices are used for modeling the ACE inhibition biological activity of a set of 58 dipeptides, taken from the literature. Description of dipeptide molecules is made by using the fragmental property Cluj indices. Four models were taken into consideration: two of them *topological* (dense topological and rare topological) and two others *geometric* (dense geometric and rare geometric). In these models, a *weak dependence on distance* for the potential function (gravitational and Coulombian), in uniform field, and a *strong dependence on distance* for the potential, that generates a non-uniform field, were considered. The indices are calculated as local descriptors of some fragments of the molecule and, a global index is then obtained by summing the fragmental contributions. The statistics were performed by STATISTICA software package. The results are compared to those reported in some previous works.

1. INTRODUCTION

QSPRs/QSARs (Quantitative Structure-Property Relationships/Quantitative Structure-Activity Relationships) link in a quantitative manner the physico-chemical or biological properties of chemicals with their molecular structure.[1]
Some molecular properties (*i.e.*,, those of which numerical value vary with changes in the molecular structure) such as the normal boiling point, critical parameters, viscosity, solubility, retention chromatographic index, are often used for characterizing chemicals in databases. However, a certain property is not always available in tables or other reference sources. It is just the case of newly synthesized compounds. As a consequence, methods of evaluating physico-chemical properties from the structural features of organic molecules become very important.

In this work several correlating results, both QSPRs and QSARs, by using Cluj type topological indices are reported, with the aim to demonstrate the capability of our indices to model the molecular properties or acrivities of organic compounds.

Cluj indices are calculated on the ground of the Cluj matrices[2-9].

Received: 26.09.2004; In revised form: 17.01.2004
2000 *Mathematics Subject Classification.* 92E10
Key words and phrases. *QSAR, PCA, Molecular Similarity*

2. Dipeptide Ace Inhibitors Structure And Qsar Analysis

The set consists of 58 dipeptides and was taken from Cocchi's report[10]. The molecular structure of these peptides was input and optimized by using the MM+ and then by semiempirical *AM1* procedure of the HyperChem Program (HyperCube Inc.). Table 1 includes the dipeptide names by using the one-letter code for aminoacids, the observed ACE inhibitory activity (biological activity, ***BA***, as $\log(1/IC_{50})$, the calculated ***BA*** according to the best model and the corresponding residuals. As above mentioned, ***FPIF*** descriptors take explicitly into account 3D-structural features of the whole molecule of dipeptides[9,11]. For more details about ***FPIF*** descriptors notations see references.

Table 1. ACE Dipeptide Inhibitors. Biological Activity ***BA*** (as $\log(1/IC_{50})$) and ***FPIF*** Descriptors.

	Dipeptide	**BA**	DTfDiM_ p/d2GP_	lnDGsDiE_ 1/p_GE_	DTjDeM_ p/d2GP_	lnDTjDe Ep2/d2AE_	DTsDeP_ 1/d_GP2	LnRGsDe Mp2/d2AE_
1	YG	2.7	12.9270	3.7077	18.3092	4.2863	10.8381	1.9687
2	YA	3.34	14.6581	3.7815	20.0651	4.3323	11.6309	2.0751
3	WG	2.23	14.6052	3.7875	23.5738	4.6148	13.7695	2.2868
4	VY	4.66	17.3311	3.8738	22.7886	4.4010	13.3679	2.0634
5	VW	5.8	19.0595	3.9408	28.1389	4.6983	16.2160	2.3330
6	VP	3.38	15.4466	3.8011	18.2386	4.1798	11.3738	1.9561
7	VG	2.96	10.9823	3.5898	10.9823	3.7426	6.9310	1.4776
8	VF	4.28	16.2311	3.8102	21.5362	4.3734	12.4505	1.9783
9	TG	2	11.1870	3.6232	11.1870	3.7729	6.9310	1.5237
10	SG	2.07	9.7467	3.5535	9.7467	3.7032	6.1003	1.4419
11	RW	4.8	19.6456	4.0138	28.8440	4.7391	15.4172	2.4548
12	RP	3.74	16.1072	3.8811	19.0216	4.2530	10.6092	1.9639
13	RF	3.64	16.9176	3.8919	22.2669	4.4311	11.9059	2.0501
14	RA	3.34	13.6241	3.7729	13.6241	3.9293	7.8329	1.6549
15	QG	2.13	11.6083	3.6867	11.6083	3.8406	6.9736	1.5911
16	PG	1.77	10.3069	3.5804	13.6744	4.0637	8.8699	2.0694
17	MG	2.32	10.8489	3.5758	10.8489	3.7383	6.3875	1.5218
18	LG	2.06	11.2208	3.6132	11.2208	3.7670	6.9604	1.4957
19	LA	3.51	12.9805	3.6951	12.9805	3.8453	7.9698	1.5932
20	KG	2.49	10.5937	3.6004	10.5937	3.7622	6.4371	1.4571
21	KA	3.42	12.3337	3.6830	12.3337	3.8400	7.4004	1.5594
22	IY	5.43	17.9426	3.8994	23.4182	4.4200	13.5252	2.0820
23	IW	5.7	19.6576	3.9641	28.7697	4.7123	16.3099	2.3465
24	IP	3.89	16.0675	3.8290	18.8929	4.2048	11.5543	1.9598
25	IG	2.92	11.6255	3.6270	11.6255	3.7827	7.2701	1.4981
26	IF	3.03	16.8470	3.8374	22.1651	4.3933	12.6226	1.9916
27	HL	2.49	16.7772	3.9310	20.0274	4.3480	11.9353	2.3013
28	HG	2.2	12.3274	3.7617	15.5328	4.2155	9.5921	2.2112
29	GY	3.68	13.8684	3.7366	19.2542	4.3101	11.9490	1.9518
30	GW	4.52	15.6288	3.8161	24.5851	4.6324	14.9982	2.2572

Table 1 continued

31	GV	2.34	11.6946	3.6205	11.6946	3.7747	7.7642	1.5541
32	GT	2.24	11.9008	3.6518	11.9008	3.8039	7.7642	1.5975
33	GS	2.42	10.3017	3.5784	10.3017	3.7305	6.7016	1.5005
34	GR	2.49	12.7478	3.7308	12.7478	3.8960	7.8394	1.6234
35	GQ	2.15	12.4229	3.7182	12.4229	3.8745	7.8769	1.6551
36	GP	3.35	11.9039	3.6453	14.5732	4.0595	9.7856	1.8659
37	GM	2.85	11.6033	3.6081	11.6033	3.7735	7.2250	1.5680
38	GL	2.6	11.9976	3.6453	11.9976	3.8014	7.8470	1.5597
39	GK	2.27	11.3900	3.6338	11.3900	3.7988	7.3105	1.5110
40	GI	2.92	12.4301	3.6596	12.4301	3.8168	8.2285	1.5785
41	GH	2.51	13.2148	3.7892	16.4180	4.2399	10.6400	2.2864
42	GG	2.14	7.4962	3.3811	7.4962	3.5384	4.8851	1.2600
43	GF	3.2	12.7536	3.6623	18.0069	4.2785	10.9713	1.8638
44	GE	2.27	12.5246	3.7377	12.5246	3.8912	7.8769	1.6641
45	GD	2.04	12.2748	3.7181	12.2748	3.8710	7.8470	1.6681
46	GA	2.7	9.2407	3.4873	9.2407	3.6424	6.0288	1.3929
47	FR	3.04	16.8720	3.8970	22.2093	4.4312	11.8141	1.9982
48	FG	2.43	11.8340	3.6309	17.0837	4.2545	9.8682	1.8490
49	EG	2	11.7090	3.7071	11.7090	3.8578	6.9736	1.6004
50	EA	2	13.4590	3.7814	13.4590	3.9283	7.9363	1.6882
51	DG	1.85	11.4944	3.6923	11.4944	3.8389	6.9604	1.6241
52	DA	2.42	13.2588	3.7725	13.2588	3.9116	7.9698	1.7379
53	AY	4.06	15.2848	3.7906	20.6980	4.3463	12.5836	1.9861
54	AW	5	17.0370	3.8650	26.0391	4.6584	15.5628	2.2807
55	AP	3.64	13.3645	3.7105	16.0783	4.1079	10.5120	1.9311
56	AG	2.6	8.8993	3.4652	8.8993	3.6231	5.7322	1.3367
57	AF	3.72	14.1753	3.7207	19.4483	4.3164	11.6299	1.9017
58	AA	3.21	10.6634	3.5622	10.6634	3.7172	6.8536	1.4574

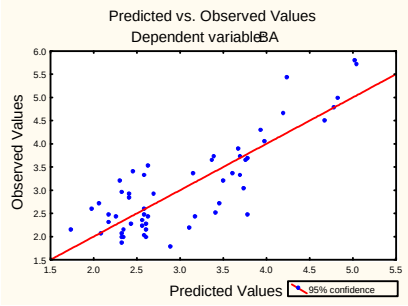
After geometry optimization, ***FPIF*** descriptors were calculated using *TOPOCLUJ* molecular topology software package. Descriptor selection procedures and final QSAR model building were done using *STATISTICA* Software Package[13]. Table 2 collects the statistics of monovariate and multivariate stepwise regression in modeling the ACE inhibiting potency of dipeptides by ***FPIF***. These two models were found to have the best statistical significance, after iterative procedures of descriptors selection using forward stepwise regression which automates the best descriptors selection task.

The quality of the QSAR derived regression models shown in Figure 1, are compared to those reported in literature (see Table 3). The model given by **BA** equation is superior, both in estimation and prediction, to those reported in literature (see Table **3**). Note that the Zaliani's results[12] refer both to a single conformation (i.e., extended) of amino acids and to a library conformation family (i.e., rotameric).

Table 2. Statistics for ACE inhibitors set.

Model 2				
	B	*Std. error*	*t*	*p-level*
Intercept	26.683	6.970	3.828	3.670E-04
DTsDeP_1/d_GP2	0.401	0.145	2.758	8.154E-03
lnRGsDeMp2/d2AE_	-0.910	0.737	-1.234	2.230E-01
lnDTjDeEp2/d2AE_	-3.730	2.197	-1.698	9.588E-02
DTjDeM_p/d2GP_	0.174	0.160	1.083	2.840E-01
lnDGsDiE_1/p_GE_	-4.531	3.248	-1.395	1.693E-01
DTfDiM_p/d2GP_	0.261	0.243	1.074	2.879E-01
Model 1				
Intercept	0.310	0.252	1.231	2.238E-01
DTsDeP_1/d_GP2	0.291	0.025	11.459	4.437E-16

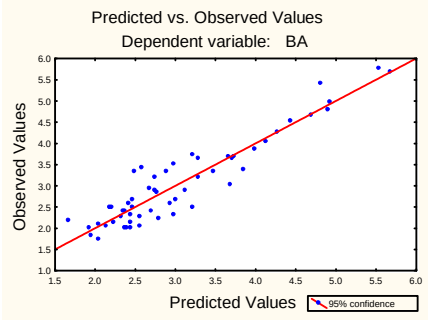
Model 1



$BA_{calc} = 0.310 + 0.291 * \text{DTsDeP\_1/d\_GP2}$

R^2=0.7085; s = 0.5520; F=131.30

Model 2



$BA_{calc} = \mathbf{26.683} + 0.401 * \mathbf{DTsDeP\_1/d\_GP2} - 0.901 * \mathbf{lnRGsDeMp2/d2AE\_} - \mathbf{3.730} * \mathbf{lnDTjDeEp2/d2AE\_} + \mathbf{0.174} * \mathbf{DTjDeM\_p/d2GP\_} - \mathbf{4.531} * \mathbf{lnDGsDiE\_1/p\_GE\_} + \mathbf{0.261} * \mathbf{DTfDiM\_p/d2GP\_}$

R^2=0.8773; s = 0.3759; F=58.429

Figure 1. Plots and derived QSAR regression models.

Table 3. Comparative statistics of QSAR models of 58 ACE inhibitors and 48 sweeteners dipeptides.

	Peptide Set (Reference)	Descriptors per Residue	No. Comp.	R^2 (fitting)
1	ACE (Cocchi et al.)[16]	7	1	0.744
2	ACE (Collantes et al.) [19]	2	nd	0.700
3	ACE (Zaliani et al. -extended)[18]	3	2	0.708
4	ACE (Zaliani et al. -rotameric)[18]	3	6	0.657
5	ACE (***FPIF***) [this work]	2	2	**0.877**
6	Sweeteners (Jonsson et al.)[20]	3	1	nd
7	Sweeteners (Collantes et al.)[16]	2	2	0.847
8	Sweeteners (Zalini et al. -extended)[18]	3	3	0.754
9	Sweeteners (Zalini et al. -rotameric)[18]	3	3	0.704
10	Sweeteners (***FPIF***) [this work]	2	2	**0.851**

Both topology (***T*** - in the index symbol) and geometry (***G***) contribute to the best model. As local property, the atomic mass (***M***) and electronegativity (***E***) modulate the structure-activity relationship.

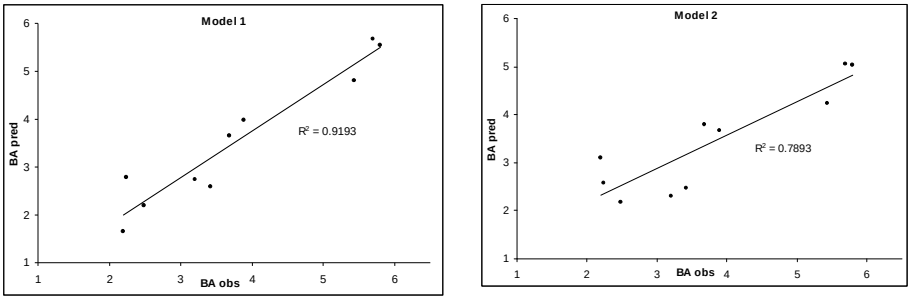


Figure 2. The plot of observed vs. calculated BA for best derived QSAR models.

In general, a model is built up by using a training set of structures (that provides a calibration equation) and further it is validated by using an external prediction set. Thus a random set of 10 dipeptides was taken out to form external prediction dataset with presumably unknown ***BA***. The results of validation proved that derived models are statistically relevant and a very good estimation of the BA (see Figure 2 and Table 4).

Table 4. Predicted and residual values in the validation set

Compound	Obs. BA	Model 1			Model 2		
		Pred. BA	Residual	**CV %**	Pred. BA	Residual	**CV %**
VW	5.80	5.543	0.257	0.238	5.026	0.774	0.394
KG	2.49	2.195	0.295	0.190	2.182	0.308	0.338
KA	3.42	2.581	0.839	0.165	2.462	0.958	0.296

Table 4 continued

IY	5.43	4.807	0.623	0.182	4.243	1.187	0.272
IW	5.70	5.677	0.023	0.249	5.053	0.647	0.399
IP	3.89	3.974	-0.084	0.139	3.670	0.220	0.220
HG	2.20	1.663	0.537	0.231	3.099	-0.899	0.226
GY	3.68	3.653	0.027	0.133	3.785	-0.105	0.226
GT	2.24	2.779	-0.539	0.154	2.568	-0.328	0.281
AA	3.21	2.733	0.477	0.156	2.303	0.907	0.319

As local property, the atomic mass (***M***) occurs five times in the first variable while the electronegativity (***E***) seven times in the second variable. Other occurring properties are the partial charge (***P***) and cardinality (***C***). Clearly, the chemical features play an important role in discriminating vertices (i.e., atoms or atom groups), fragments and whole molecules of dipeptides. They are strongly involved in modeling the biological activity of dipeptide ACE inhibitors.

3. Conclusions

FPIF offer good description and modeling of dipeptides activity, such as the ACE inhibition or bitter tasting. As it is known, a correlation model does not involve a causal relationship between descriptors and a molecular property. The above results demonstrate the usefulness of our descriptors in modeling peptide structures and properties. For other ***FPIF*** modeling examples the reader can consult ref.[9]. These indices can be calculated with *TOPOCLUJ* software package available on request.

References

[1] Free, S. M. and Wilson, J. W., J. Med. Chem. **7** (1964), 395.
[2] Diudea, M.V., Commun. Math. Comput. Chem. (MATCH), **35** (1997), 169-183.
[3] Diudea, M.V., Minailiuc, O.M., Katona, G., Gutman, I., Commun. Math. Comput. Chem. (MATCH), **35** (1997), 129-143.
[4] Diudea, M.V., J.Chem.Inf. Comput. Sci., **37** (1997), 300-305.
[5] Diudea, M.V., Pârv, B., Topan, M.I., J. Serb. Chem. Soc. **62** (1997), 267-276.
[6] Kiss, A.A., Katona, G., Diudea, M.V., Coll. Sci. Papers Fac. Sci. Kragujevac, **19** (1997), 95-107.
[7] Gutman, I., Diudea, M.V., J. Serb. Chem. Soc., **63** (1998), 497-504.
[8] Diudea, M.V., Katona, G., Lukovits, I., Trinajstić, N., Croat. Chem. Acta, **71** (1998), 459-471.
[9] Jäntschi, L., Katona, G., Diudea, M.V., Commun. Math. Comput. Chem. (MATCH), **41** (2000), 151-188.
[10] Cocchi, M., Johansson, E., Quant. Struct.-Act. Relat., **12** (1999), 1-8.
[11] Opris, D., Diudea, M.V., SAR/QSAR Environ.Res. **12** (2001), 159-179.
[12] Zaliani, A., Gancia, E.J., Chem. Inf. Comput. Sci., 39 (1999), 525-533.
[13] StatSoft, Inc. (2001). STATISTICA (data analysis software system), version 6. www.statsoft.com.

Babeş - Bolyai University
Faculty Of Chemistry And Chemical Engineering
Arany János 11, 40028, Cluj, Romania
E-Mail: diudea@chem.ubbcluj.ro