RESEARCH PAPER

Accepted 30 March 2000
Revised 24 March 2000
Received 17 January 2000
Indian J. Pharm. Sci., 2000, 62 (4) 280-286

Molecular Shape Analysis of Some Antiinflammatory Benzimidazole Derivatives

S.K. CHAKRAVARTI* AND S.C. CHATURVEDI
Department of Pharmacy,
Shri Govindram Seksaria Institute of Technology and Science,
23-Park Road, Indore - 452003 (M.P.)

To study the effect of variation in three dimensional molecular shape on antiinflammatory activity of benzimidazole derivatives and to obtain predictive models, a series of 2-(Substituted pyridinyl) benzimidazoles was subjected to molecular shape analysis. Various shape descriptors and other physicochemical descriptors were calculated and used to derive quantitative structure-activity relationships. A genetic function approximation algorithm was used to search significant relationships. The equations were validated by cross validation method. Y-component of dipole moment and shape parameters like common overlap steric volume and root mean square deviation from the shape reference compound were found to have significant contribution to antiinflammatory activity. It was also found that in this case genetic function approximation method was able to find more significant relationships than stepwise multiple regression method.

In continuation of our study of quantitative structure-activity relationships of non-acidic antiinflammatory compounds, a series of 2-(substituted pyridinyl) benzimidazoles[1] Fig. 1 was selected to analyze the effect of molecular shape[2] and conformational flexibility on antiinflammatory activity.

In place of stepwise multiple regression method, an advanced genetic function approximation algorithm (GFA)[3,4,5] has been used to obtain QSAR models. Genetic algorithms (GA) have their basis in Darwinian model of natural selection and evolution. The general idea behind GAs is the evolutionary creation of a new population of entities from an earlier generation through certain combinations and changes and passing on the fittest offspring to the next generation. GAs work by simulating two important characteristics of natural evolution. The first is the "survival of the fittest" idea and second is the propagation of the attributes of the mating members by the recombination of genetic material and transmission to the offspring. In QSAR, the equations are the members of a population and descriptors are the genetic material. One advantage of this method is that GAs examine and manipulate a set of potential candidate equations instead of a single one and thus expected to lead to generations of equations that become more and more fit through genetic evolution thus achieving the desired characteristics[4,5].

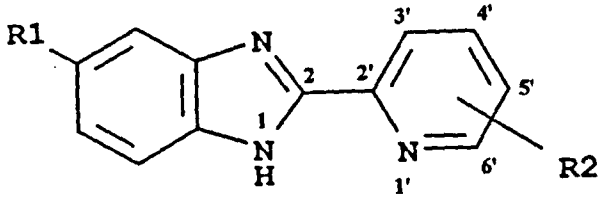

Fig. 1 : 2-(Substituted pyridinyl) benzimidazoles used in the study

EXPERIMENTAL

The antiinflammatory data of 2-(substituted pyridinyl) benzimidazoles were taken from Tsukamoto *et al*[1]. All

* For correspondence
Phone : 91-731-434095
Fax : 91-731-432540

TABLE 1 : SUBSTITUENTS AND ANTIINFLAMMATORY ACTIVITY FOR 2- (SUBSTITUTED PYRIDINYL) BENZIMIDAZOLES

No.	R1	R2	PICPE+	BA++	LOG(BA)
1	H	H	22	0.0429	-1.3671
2	H	3-Me	17	0.0356	-1.4489
3	H	4-Me	-4	-0.0084	-
4	H	5-Me	19	0.0398	-1.4006
5	H	6-Me	44	0.0921	-1.0359
6	H	5-Et	44	0.0982	-1.0077
7	H	6-Et	41	0.0915	-1.0383
8	H	5-n-Bu	-16	-0.0402	-
9	H	5-CH_2OH	-10	-0.0225	-
10	H	6-CH_2OH	-13	-0.0293	-
11	H	5-COOEt	-16	-0.0428	-
12	H	6-COOEt	31	0.0829	-1.0817
13	H	6-Cl	30	0.0689	-1.1618
14	H	6-OMe	47	0.1059	-0.9752
15	H	6-$CONH_2$	13	0.0309	-1.5090
16	H	6-OH	-7	-0.0148	-
17	Me	H	35	0.0732	-1.1353
18	OMe	H	34	0.0766	-1.1159
19	Cl	H	36	0.0827	-1.0826
20	Me	5-Me	7	0.0156	-1.8061
21	Cl	5-Me	-5	-0.0122	-
22	Me	6-Me	42	0.0938	-1.0279
23	OMe	6-Me	40	0.0957	-1.0190
24	Cl	6-Me	39	0.0950	-1.0221
25	OH	6-Me	6	0.0135	-1.8692
26	Me	5-Et	35	0.0831	-1.0806
27	OMe	5-Et	58	0.1469	-0.8329
28	Cl	5-Et	-2	-0.0052	-
29	OH	5-Et	25	0.0598	-1.2232
30	NO_2	5-Et	-5	-0.0134	-
31	NH_2	5-Et	25	0.0596	-1.2249
32	NHAc	5-Et	5	0.0140	-1.8534
33	Me	6-Et	10	0.0237	-1.6247
34	OMe	6-Et	52	0.1317	-0.8804
35	Cl	6-Et	3	0.0077	-2.1117
36	H	5,6-$(Me)_2$	37	0.0826	-1.0829
37	Me	5,6-$(Me)_2$	-19	-0.0451	-
38	Me	6-OMe	32	0.0766	-1.1159
39	Me	6-Cl	36	0.0877	-1.0569
40	Me	6-OH	-6	-0.0135	-

+ Per cent inhibition of paw edema by 1000.0 mg/kg of drug

++ Per cent paw edema inhibition per micromolecule of drug per kg of body weight

the biological activity data (PICPE) has been converted to logarithmic equieffective molar doses (LogBA) for MSA analysis. The substitution and biological activity data of this series is given in Table 1. The software Cerius2 (Biosym/MSI)[6] installed on a Silicon Graphics Workstation has been used to perform all molecular modeling functions. Energy of the molecules were minimized using conjugate gradient algorithm working under Universal Force Field[7].

Molecular Shape Analysis was performed through the following steps[2] :

Generation of conformers:

The purpose of this task was to generate and analyze conformers for each of the structures, then to reduce the number that are likely to relevant to the biological activity. Boltzman jump method was used to generate conformers. The maximum number of conformers which can be generated was set to 200 and a cutoff value of 5 kcal/mol was given. A conformer was accepted if its energy value was less than the cutoff value plus the value of the lowest energy conformer.

Identifying a shape reference compound:

A conformer of the most active molecule (compound No. 27) was selected as the shape reference compound which was used when shape descriptors were calculated. The compound in the series were aligned on the shape reference compound using root mean square rigid alignment.

Measuring the molecular shape commonality:

In this step shape descriptors[2] were calculated to compare the properties that two molecules have in common. Following shape descriptors were calculated which are presented in Table 2.

TABLE 2 : SHAPE DESCRIPTORS CALCULATED FOR 2-(SUBSTITUTED PYRIDINYL) BENZIMIDAZOLES

Comp NO*	DIFFV	COSV	FO	NCOSV	SRVOL	Shape RMS
1	-41.0547	153.8554	0.8836	20.2604	215.1705	0.01824
2	-23.461	152.3601	0.7947	39.3493	215.1705	0.09275
4	-24.3127	170.3264	0.8924	20.5314	215.1705	0.01587
5	-24.081	156.0098	0.8164	35.0797	215.1705	0.02365
6	-7.5178	181.5821	0.8744	26.0706	215.1705	0.10249
7	-7.1685	157.6163	0.7578	50.3856	215.1705	0.02159
12	19.2059	155.5004	0.6635	78.8760	215.1705	0.02463
13	-26.9647	155.1956	0.8246	33.0102	215.1705	0.02524
14	-15.2865	154.1584	0.7712	45.7256	215.1705	0.03029
15	-10.6969	151.1444	0.7392	53.3292	215.1705	0.08326
17	-24.5553	163.7903	0.8593	36.8249	215.1705	0.02322
18	-15.9085	178.4973	0.8958	20.7647	215.1705	0.01471
19	-26.995	163.255	0.8676	24.9204	215.1705	0.01919
20	-7.4818	180.5259	0.8692	27.1628	215.1705	0.01987
22	-7.5512	166.1403	0.8002	41.479	215.1705	0.02237
23	0.7217	180.9314	0.8381	34.9609	215.1705	0.01031
24	-10.43	165.3157	0.8074	39.4248	215.1705	0.02204
25	-16.2075	163.3617	0.8211	35.6012	215.1705	0.02222
26	8.6383	192.3009	0.8592	31.5079	215.1705	0.09668
27	17.9183	215.1705	0.9231	17.9183	215.1705	0
29	0.1884	188.9222	0.8772	26.4367	215.1705	0.09857

Comp NO*	DIFFV	COSV	FO	NCOSV	SRVOL	Shape RMS
31	2.9908	187.9287	0.8614	30.2326	215.1705	0.11595
32	40.1301	202.4414	0.7929	52.8592	215.1705	0.06365
33	9.4605	170.167	0.7575	54.4639	215.1705	0.02024
34	17.2944	184.5512	0.7939	47.9137	215.1705	0.01182
35	6.362	168.7483	0.7617	52.7842	215.1705	0.022
36	-7.5365	170.975	0.8234	36.659	215.1705	0.02832
38	0.7864	164.736	0.7628	51.2209	215.1705	0.03209
39	-10.1281	164.9548	0.8045	40.0875	215.1705	0.02431

* Compounds only with nonzero and positive PICPE are considered

Common overlap steric volume (COSV): The common volume between each individual molecule and the shape reference compound, **Difference volume (DIFFV)**: The difference between the individual molecule and the volume of the shape reference compound, **Common overlap volume ratio (FO)**: The common overlap steric volume descriptor (COSV) divided by the volume of the individual molecule, **Non-common overlap steric volume (NCOSV)**: The volume of the individual molecule and the common overlap steric volume, **RMS to shape reference (Shape_RMS)**: This is root mean square deviation between the volume of individual molecule and the volume of shape reference compound (SRVOL).

Determining other molecular features:

Various molecular descriptors were calculated: Energy of the selected conformation (ENERGY), energy of the most stable conformation (LOWENE), difference between ENERGY and LOWENE (EPENALTY), Molecular surface area (AREA), van der Waals surface area (VWS), connolly surface area (CONNOLLYS), molecular volume (VM), radius of gyration (RADGYR), molecular density (DENSITY), principal moment of inertia (PMI) and its X, Y and Z component (PMIX, PMIY, PMIZ), molecular weight (MW), number of rotatable bonds (ROTLBONDS), number of hydrogen bond acceptors (HB_ACCEPTOR), number of hydrogen bond donors (HB_DONOR), octanol/water partition coefficient (LOGP), molecular refractivity (MR), dessolvation free energy for water (FH20), dessolvation free energy for octanol (FOCT), heat of formation (HF), highest occupied molecular orbital energy (HOMO), lowest unoccupied molecular orbital energy (LUMO), difference between LUMO and HOMO (HOLU), dipole moment (DIPOLE) and its X, Y and Z component (XDIP, YDIP, ZDIP) and partial atomic charges (on atoms C2, N1 and C2').

Construction of a trial QSAR:

This is the final step of MSA. Appropriate conformers were first selected to use in generating a 3D QSAR. All descriptor data were examined to determine the conformer that best match the shape reference data. The data were evaluated using the partial least squares regression (PLS) method[8]. The steps 2 to 5 were repeated until a satisfying QSAR was obtained. In the present work, GFA was used to produce the QSAR equations. GFA consists of following steps[3].

Building the initial population and evolving the population:

The analysis begins by building a population of 300 randomly constructed equations. The populations was then evolved for 5000 generations. For each generation, two better scoring equations were selected as parents. Parts of each parent equation were then used and crossover was performed to create a child equation. Mutation operations were performed on the child at creation. The worst rated equation was then replaced by the new child equation.

Evaluating the QSAR equations:

Each equation of the evolved population was then evaluated using various statistical measures such as number of data points (n), correlation coefficient (r), cross-validated r^2 (cvr^2), bootstrap r^2 (bsr^2), F-test (F), standard deviation (std), squared correlation coefficient (r^2) and predicted sum of squares (PRESS).

RESULTS AND DISCUSSION

When all the compounds (excluding those with PICPE less than or equal to 0.0 because their log values can not be calculated) were subjected to MSA, then the equations which were resulted from GFA analysis were not statistically significant and can not be used for prediction purpose. One of the equations is as follows

LOG(BA) = + (0.015326) ENERGY + (-0.01077) NCOSV+ (9.363007) C2' + (0.0007055) PMI + (0.03681) HOMO + (-2.54968) [1]

n=29, r^2 = 0.360, r=0.600, PRESS = 0.982, cvr^2 = 0.102, bsr^2=0.141, F=15.199, std = 0.288

When compounds with insignificant antiinflammatory activity (PICPE<=10) were deleted from space and MSA was again carried out then the equations obtained from GFA analysis showed significant improvement. Some of the equations with promising statistics from a population of 300 equations are presented below:

LOG(BA) = +(7.216254)C'2+(-2.309939)C2 + (-1.580474) SHAPE_RMS + (0.137197) YDIP + (0.351990) RADGYR + (-2.438280) [2]

n=24, r^2 = 0.759, r=0.872, PRESS=0.320, cvr^2=0.506, bsr^2=0.762, F = 11.346, std = 0.093

LOG (BA) = +(0.245542) YDIP + (0.028022) VWS+ (-0.000495) PMI+ (-0.125789) XDIP+ (-1.343364) SHAPE_RMS + (0.003182) COSV + (-4.374235) [3]

n=24, r^2 = 0.759, r=0.872, PRESS=0.293, cvr^2=0.548, bsr^2=0.762, F = 9.956, std = 0.096

LOG (BA) = +(-2.423520) C2+(0.275788) RADGYR + (0.118282) YDIP + (7.714505) C2' + (-1.634893) SHAPE_RMS + (0.002495) COSV + (-2.535354) [4]

n=24, r^2 = 0.779, r=0.883, PRESS=0.306, cvr^2=0.528, bsr^2=0.781, F = 9.979, std = 0.092

LOG (BA) = +(-2.486700) C2+(0.384184) RADGYR + (0.083567) ZDIP + (0.062174) XDIP + (9.293721) C2' + (-1.356526) SHAPE_RMS + (-2.675697) [5]

n=24, r^2 = 0.759, r=0.872, PRESS=0.273, cvr^2=0.579, bsr^2=0.762, F = 8.967, std = 0.096

LOG(BA) = +(-1.952186) C2 + (0.321337) YDIP + (-1.094580) SHAPE_RMS + (0.041855) VWS + (-0.000642) PMI + (-0.189814) XDIP + (0.904235) FO+ (-5.750052) [6]

n=24, r^2 = 0.806, r=0.898, PRESS=0.270, cvr^2 = 0.583, bsr^2=0.807, F = 9.492, std = 0.089

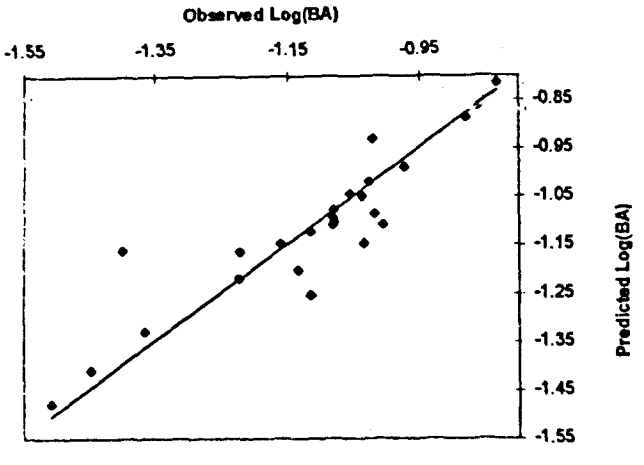

Fig. 2 : Observed vs. predicted antiinflammatory activity using equation [6].The (•) symbols show the corresponding observed and predicted Log (BA) for each compound of the series of 2-(substituted pyridinyl) benzimidazoles

It is evident from the equations [2]-[6] that they have significant bootstrap r^2 and cross validated r^2 values. This shows that these equations have good prediction ability and their r^2 values are not dependent on any one compound. The predicted vs observed activity plot of equation [6] is given in Fig. 2. These equations can be used for prediction of antiinflammatory activity of new benzimidazole derivatives. In fact, in our laboratory these equations have been used in designing and synthesis of some new benzimidazole derivatives with potent antiinflammatory activity. Two of such compounds are shown in Fig. 3. Compound I and compound II has shown an experimental 64 and 55 per cent inhibition of paw edema respectively at a dose of 100 mg/kg whereas ibuprofen and mefenamic acid has shown 45 and 56 per cent inhibition respectively at the same dose. The predicted paw edema inhibition of compound I and compound II using equation [6] was 44 and 52 respectively.

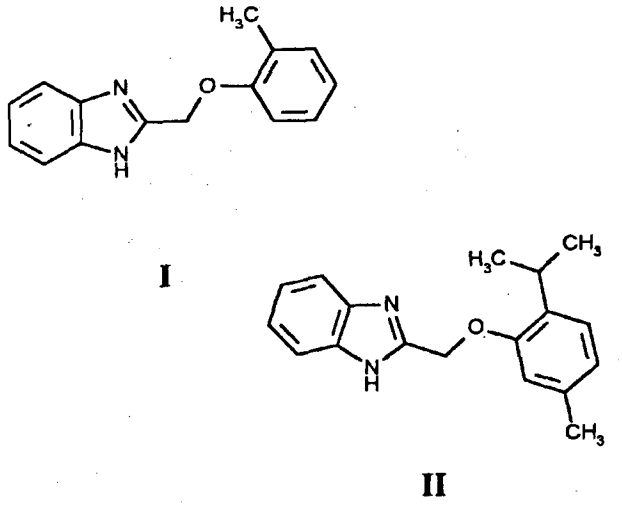

Fig. 3 : Structures of two new benzimidazoles designed using molecular shape analysis and shown comparable or better antiinflammatory activity than compound no. 27 after synthesis

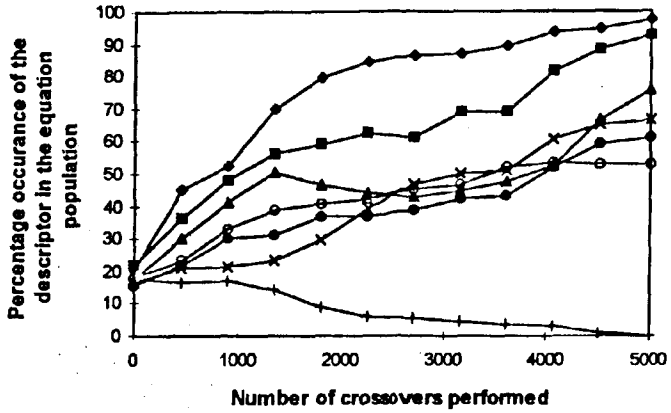

Fig. 4 : Percentage occurrence of the descriptors in the Genetic Function Approximation Analysis (● PMIX, x COSV, ▲ C2', ■ C2, ◆ YDIP, o SHAPE_RMS, + DENSITY)

To show the relative importance of the selected descriptors in the variation of activity, the percentage occurrence of the descriptors in the final population of 300 equations was studied. It was found that YDIP has found place in about 100% of the equations among a population of 300 equations. C2 found place in about 90% of the equations, C2' in about 75% COSV in about 70%, SHAPE_RMS in about 50% and PMIX in about 60% of the equations. These parameters are strong enough to survive through the 5000 generations in GFA analysis and ultimately they found their place in the final population. A plot of percentage occurrence of some descriptors in the equation population during the crossovers of GFA is shown in Fig. 4. It can be seen that powerful parameters like YDIP, C2, C2', COSV, SHAPE_RMS and PMIX gradually increase their number of presence in the population and survive while weak parameters like DENSITY gradually diminish and could not survive through the generations. Powerful parameters are those which can produce desired characteristics in the equations like decrease in the mean squared error between observed and predicted biological activity and increase in r^2. Weak parameters fail to produce desired characteristics. These observations show that relative difference in three dimensional shape of the molecules in the series and particular charge distribution is important in defining the variation of antiinflammatory activity in this series of benzimidazole derivatives.

In our previous paper[9] in which stepwise regression was used for QSAR analysis of the same series, following equation was obtained as one the models

LOG(BA) = 0.16711(±0.10766) YDIP + 0.00611 (±0.00197) VM + 0.02238 (±0.00838) FH20- 0.12376 (±0.16007) HOMO -3.18513 [7]

n = 24, r = 0.759, r^2 = 0.576, F = 29.886, std = 0.120

It is evident from the r^2 value that the equations obtained from GFA are statistically more significant than the equation obtained from stepwise multiple regression.

ACKNOWLEDGEMENTS

Authors wish to thank M/s Torrent Research Center, Bhat, Gandhinagar, for providing access to computational and molecular modeling software facilities. Authors specially thank A.B. Mandhare, Torrent Research Center, Bhat, Gandhinagar, for providing valuable help during this study. Authors also wish to thank the Director, S.G.S.I.T.S., Indore for providing facilities for this project. S.K. Chakravarti is grateful to the CSIR, New Delhi for providing a senior research fellowship.

REFERENCES

1. Tsukamoto, G., Yoshino, K., Kohno, T., Ohtaka, H., Hajim, K. and Ito, K., **J. Med. Chem.,** 1980, 23, 734.
2. Hopfinger, A.J. and Burke, B.J., In; Johnson, M.A. and Maggiora, G.M., Eds; Concepts and Applications of Molecular similarity, John Wiley and Sons, New York, 1990, 173.
3. Rogers, D. and Hopfinger, A.J., **J. Chem. Inf. Comput. Sci.,** 1994, 34, 854.
4. Venkatasubramanian, V., Chan, K. and Caruthers, J.M., In; Reynolds, C.H., Holloway, M.K. and Cox, H.K. Eds; Computer-Aided Molecular Design-Application in Argochemicals, Materials and Pharmaceuticals, ACS Symposium Series, American Chemical Society, Washington, DC, 1995, 397.
5. Holland, J.H., In; Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, 1975 1.
6. Cerius2, Biosym/Molecular Simulations Inc., 9685 Scranton Road, California, USA 92121-3752.
7. Rappe, A.K., Casewit, C.J., Colwell, K.S., Goddard III, W.A. and Skiff, W.M., **J. Am. Chem. Soc.,** 1992, 114, 10024.
8. Wold, S., Marten, H. and Wold, H., In; Kagstrom, R.A., Eds; Matrix Pencils (Lecture Notes in Mathematics), Springer-Verlag: Heidelberg, 1983, 286.
9. Chakravarti, S.K. and Chaturvedi, S.C., **Indian J. Pharm. Sci.,** 1999, 61, 206.