A QSAR Approach for the Prediction of Stability of Benzoglycolamide Ester Prodrugs

Balasubramanian NARASIMHAN,^a Afaque Mehboob ANSARI,^b Nartaj SINGH,^c Vishnukant Mourya,^d and Avinash Shridhar DHAKE*,*^e*

^a Department of Pharmaceutical Sciences, Guru Jambheshwar University; Hisar 125001, India: ^b BSPS Mandal's College of Pharmacy; Solapur 413003, India: ^c Department of Food Technology, Guru Jambheshwar University; Hisar 125001, India: ^d Govt College of Pharmacy, Osmanpura; Aurangabad 431005, India: and ^e L.B.Rao Institute of Pharmaceutical Education and Research; B.D.Rao College Campus, Khambhat–388620, India. Received October 6, 2005; accepted May 2, 2006

A method consisting of quantitative structure–activity relationship (QSAR) (MLR) was developed to predict the hydrolytic rate constant of 37 benzoglycolamide ester prodrugs. The regression method was used as a calibration model for calculating the hydrolytic rate constant and investigating their linear characteristics. The QSAR study indicated the importance of the descriptors charge on amide nitrogen (AN), lipophilic parameter (log P) and nucleophilic frontal density (NUFD) in contribution to the ester hydrolysis with the correlation coefficient value of 0.908 for the developed MLR model. The models were validated by leave one out (LOO) technique as well as by the calculation of statistical parameters for the developed MLR models.

Key words benzoglycolamide ester; hydrolytic constant; molecular descriptor

Quantitative structure–activity relationship (QSAR) represents an attempt to relate structural descriptors of compounds with their physicochemical properties and biological activities. It is widely used for the prediction of physicochemical properties in chemical, environmental, and pharmaceutical areas.^{1,2)} The main steps involved in this method include the following: data collection, molecular descriptor selection and procurement, correlation model development and finally model evaluation. At present, many types of molecular descriptors have been proposed to describe the structural features of the molecules. $3-5$)

The main focus of these studies is to search for regularity in the manner in which molecular properties change as molecular structure changes.^{6,7)} One such property is the hydrolysis of drug molecules. Double ester prodrugs of penicillins are more susceptible to enzymatic hydrolysis than their simple alkyl esters but have the problem of poor aqueous solubility. This problem of carboxylic acid prodrugs was overcome by the synthesis of benzoglycolamide esters. 8) The literature reveals many reports on QSRR (quantitative structure reactivity relationship).^{9—15)} Chung *et al.*¹⁶⁾ has shown structure reactivity relationships in chemical hydrolysis of prodrugs of nicotinic acid. Similar studies on QSRR were found in literature.^{17—19)} In such studies the reactivity, in terms of rate constants, is determined and then correlated with structural parameters by use of suitable statistical techniques. The aim of this study was to establish quantitative relationship(s) between structure of benzogylcolamide esters⁸⁾ and their basic hydrolytic rate constants. Bundgaard *et al.*8) synthesized benzoglycolamide esters and studied their physicochemical properties and stability, which were based on qualitative observations, which can be fortuitous also. But the correlation of benzoglycolamide ester structure with hydrolytic rate constant has not yet been investigated.

In view of above and in continuation of our studies on correlation of molecular properties with activity, $20-22$ we decided to study the usefulness of QSAR in the prediction of the basic hydrolytic rate constant of benzoglycolamide esters synthesized by Bundgaard *et al.*⁸⁾ The multiple linear regression (MLR) model has been developed as a calibration model for predicting the $-\log \kappa c$ of compounds belonging to test set and investigating their linear characteristics. The present results are important from a point of accurate prediction of $-\log \kappa c$. Our model can predict the basic hydrolytic rate constant of variety of a benzoglycolamide esters and it reduces the need for different time consuming experiments.

Methodology

The main aim of the present work was development of a quantitative model to predict the basic hydrolytic rate constant of different benzoglycolamide esters. In the present work, the hydrolytic rate constants of 44 benzoglycolamide esters synthesized 8) were subjected to MLR with their physicochemical properties. The best MLR model was used to predict the basic hydrolytic rate constants, based on which 7 outliers were removed from the model set and a final set of 37 compounds were selected for MLR and their data is listed in Table 1. In multivariate statistics, it is common to define three types of outliers.²³⁾

- 1. X/Y relation outliers are substances for which the relationship between the descriptors (X variables) and the dependent variables (Y variables) is not the same as in the (rest of the) training data.
- 2. X outliers. Briefly, a substance is an X outlier if the molecular descriptors for this substance do not lie in the same range as the (rest of the) training data.
- 3. Y outliers are only defined for training or test samples. They are substances for which the reference value of response is invalid.

In light of the above guidelines, seven benzoglycolamide esters were considered as outliers because their response values [basic hydrolytic constant] were outside the range in comparison to the other compounds included in the present study. One of the main problems in developing MLR models is choosing the proper inputs (descriptors) for them. There are two different methods of feature selection techniques: objective and subjective methods. The former method selects the relation between the descriptors themselves, whereas the latter method defines the relation between the descriptors and the dependent variable *i.e.*, basic hydrolytic rate constant.

Pearson Correlation Analysis We have employed Pearson's correlation analysis, as an objective feature selection method to classify the descriptors. This technique was adopted for choosing a suitable set of generated descriptors for developing a multiple linear regression model. The best generated MLR model was used to prepare a calibration model, which predicts the basic hydrolytic rate constant of benzoglycolamide esters and illustrates the extension of the linear characteristics of the hydrolytic behavior of these compounds.

Table 1. Training Set and Corresponding Observed, Predicted (MLR), and Residual Basic Hydrolytic Rate Constant

a) log $\kappa o = \log \kappa$ (observed); as reported by Bundgaard *et al.*⁸⁾ *b*) log $\kappa c = \log \kappa$ (calculated).

Descriptor Generation The next step in developing a model is generation of the numerical description of the molecular structures. The numerical descriptors are responsible for encoding important features of the structure of the molecules and can be categorized as hydrophobic, geometric, electronic and topological characters. Descriptors were calculated for each compound in the data set, using the software Dragon²⁴⁾ and Hyperchem.²⁵⁾ Since there was large number of descriptors for each compound, we used Pearson's correlation matrix as a qualitative model, in order to select the suitable descriptors for MLR analysis. The values of descriptors selected for MLR model are presented in Table 2.

Regression Analysis The stepwise multiple linear regression procedure was used for model generation. The stepwise addition method implemented

Table 2. Values of Eight Descriptors of Benzoglycolamide Esters Used in MLR

No.	AN	log P	MR	MV	Pr	NPSSA	MW	NUFD
1	-0.42	0.51	49.11	154.7	406.1	73.4	177.2	0.20
$\overline{\mathbf{c}}$	-0.37	0.86	50.91	166.2	424.7	112.5	193.2	0.20
3	-0.36	1.39	55.54	182.7	464.5	136.4	207.2	0.20
$\overline{\mathbf{4}}$	-0.36	1.91	60.18	199.2	504.2	115.3	221.3	0.20
5	-0.36	1.69	60.14	199.6	501.6	157.6	221.3	0.20
6	-0.36	2.09	64.78	215.8	539.1	175.8	235.3	0.20
7	-0.35	-0.05	59.11	183.5	498.5	96.9	236.2	0.19
8	-0.36	0.25	59.24	182.0	496.3	98.9	236.2	0.20
9	-0.37	-0.48	69.38	208.3	589.3	137.1	283.3	0.20
10	-0.35	1.81	60.30	197.7	502.0	163.3	221.3	0.15
11	-0.31	2.34	64.94	214.2	541.7	122.3	235.3	0.16
12	-0.32	3.40	74.20	247.2	621.3	214.0	263.3	0.16
13	-0.32	2.43	73.65	237.4	599.3	163.9	259.3	0.18
14	-0.32	4.46	83.47	280.2	700.9	246.6	291.4	0.14
15	-0.31	4.20	83.39	281.0	695.7	231.0	291.4	0.16
16	-0.34	0.53	63.87	198.5	536.1	123.0	250.3	0.18
17	-0.32	1.08	61.84	195.2	518.7	145.9	237.3	0.19
18	-0.32	1.61	66.47	211.7	558.5	110.6	251.3	0.15
19	-0.31	1.50	77.19	243.0	649.5	188.6	295.3	0.18
20	-0.31	2.12	77.69	259.9	660.7	170.1	295.3	0.15
21	-0.32	2.26	71.34	234.2	606.3	155.6	279.3	0.20
22	-0.32	3.31	76.84	242.6	629.6	213.4	275.3	0.20
23	-0.33	5.35	97.93	205.5	798.6	285.3	343.5	0.19
24	-0.32	1.88	73.29	237.9	607.2	200.5	264.3	0.20
25	-0.32	1.92	62.57	193.3	511.6	163.7	233.3	0.20
26	-0.31	2.48	67.17	211.0	551.6	177.8	247.3	0.18
27	-0.31	2.04	71.78	228.7	591.7	198.9	261.3	0.18
28	-0.32	1.51	64.16	202.0	531.9	160.2	249.3	0.16
29	-0.31	3.53	76.50	248.6	629.7	215.9	275.3	0.19
30	-0.31	3.51	76.59	251.0	627.0	215.3	275.3	0.19
31	-0.31	0.39	68.76	208.0	566.0	154.3	263.3	0.16
32	-0.32	2.14	70.83	222.1	578.0	187.4	262.3	0.15
33	-0.32	1.56	76.99	237.1	635.1	203.6	292.0	0.19
34	-0.29	1.40	64.05	187.8	531.7	131.1	263.25	0.19
35	-0.32	1.95	68.67	205.5	571.8	147.5	277.3	0.20
36	-0.32	2.26	73.51	230.9	615.2	187.0	291.3	0.16
37	-0.31	1.06	70.67	211.7	584.5	149.2	276.3	0.18

in the SPSS software package²⁵⁾ was used for choosing the descriptors contributing to the hydrolytic rate constant. As a first step, a correlation matrix was performed for all descriptors with $-\log \kappa c$. The eight descriptors, charge on amide nitrogen (AN), lipophilic parameter, (log P), molar Volume (MV), molar refractivity (MR), parachor (Pr), molecular weight (MW), nucleophilic frontal density (NUFD) and nonpolar solvent accessible surface area (NPSSA) which were showing maximum correlation with $-\log \kappa c$ were chosen for further evaluation. The best MLR model consists of three descriptors.

The three descriptors appearing in this model are AN, log P, and NUFD. The main goal of the MLR analysis was developing a model for the prediction of basic hydrolytic rate constant.

Calculation of Statistical Parameters The selected models were validated by the calculation of following statistical parameters^{26,27)}: probable error of the coefficient of correlation (PE), least square error (LSE), Friedman's lack of fit measure (LOF), standard error of prediction (SEP), quality value (Q) . These parameters were calculated from the following equations.

 $PE = 2(1 - r^2)/3\sqrt{n}$

Where, *r*, correlation coefficient and *n*, number of compounds used.

$$
LSE = \sum (Y_{obs} - Y_{calc})^2
$$

Where, Y_{obs} and Y_{calc} are the observed and calculated values.

 $LOF = LSE/{1-(C+d\cdot p/n)}^2$

Where, LSE, least square error; *C*, number of descriptors $+1$; *p*, number of independent parameters; *n*, number of compounds used; *d*, smoothing pa-

Table 3. Pearson Correlation Matrix Constructed to Determine the Interrelationship among the Parameters

	$-\log \kappa c$ AN		log P	MR	MV	MW	NPSSA NUFD	PR
$-\log\kappa c$	1.000							
AN	0.812	1.000						
log P	0.690	0.430	1.000					
MR	0.713	0.600	0.742	1.000				
MV	0.705	0.612	0.658	0.789	1.000			
MW	0.629	0.640	0.527	0.938	0.696	1.000		
NPSSA	0.693	0.513	0.835	0.877	0.745	0.732	1.000	
				NUFD $-0.511 - 0.416 - 0.269 - 0.309 - 0.443 - 0.265 - 0.246$			1.000	
PR	0.700	0.593	0.721	0.996	0.806	0.947	$0.856 - 0.325$	1.000

rameter which controls the bias in the scoring factor between equations with different number of terms and was kept 1.0.

 $\text{SEP}=\sqrt{\text{LSE}}/n$

The Quality value, *Q* is given by

Q-*r*/Se

Where, *Q*, Quality value; *r*, correlation coefficient and Se, standard error. The predictive ability of MLR models was also quantified in terms of Q^2 , which is defined as

$$
Q^2 = 1 - \left\{ \sum (Y_{\text{obs}} - Y_{\text{calc}})^2 / \sum (Y_{\text{obs}} - Y_{\text{mean}})^2 \right\}
$$

The low value of PE, LSE, LOF and SEP and high value of Q and Q^2 are the essential criteria for qualifying the model as the best one.

Results and Discussion

The main goals of the present work were as follows: (1) to accurately predict the basic hydrolytic rate constant of benzoglycolamide esters, (2) to assess the ability of MLR technique in predicting the hydrolytic behavior of a set of benzoglycolamide esters. As can be seen in Table 1, a data set consisting of 37 benzoglycolamide esters with different structures were chosen to develop the model.

Pearson Correlation Matrix Developing a general model requires a diverse set of data, and, thereby a large number of descriptors have to be considered. Descriptors are numerical values that encode different structural features of the molecules. Selection of a set of appropriate descriptors from a large number of them requires a method, which is able to discriminate between the parameters. We have performed Pearson correlation matrix by using SPSS on all descriptors calculated for each molecule. The analysis of the matrix revealed 8 descriptors for the development of MLR model. It is noteworthy that the three descriptors appearing in the MLR model show the largest variance in the Pearson correlation matrix (Table 3). The interrelationship among these parameters are very poor, $r=0.430$, 0.416 and -0.269 between AN and log P, AN and NUFD and NUFD and log P, respectively. Therefore, the Pearson correlation matrix confirms the selection of these descriptors using MLR technique.

Multiple Linear Regression Analysis as Feature Selection and Calibration Model A correlation matrix of $-\log \kappa c$ with selected parameters of MLR model was constructed and is presented in Table 3. A high interrelationship was observed only between PR and MR (r =0.996), and PR and MV $(r=0.806)$ while interrelationship between all other parameters was less than 0.8. Linear models were formed by a stepwise addition of terms. A deletion process was then

Table 4. Best MLR Models for the Prediction of Basic Hydrolytic Constant: (A) 37 Compounds and 1 Parameter, (B) 37 Compounds and 2 Parameters, and (C) 37 Compounds and 3 Parameters

Descriptor	Coefficient		t-Test value	
(Model $1)^{a}$)				
Intercept	1.723	0.377	4.573	
AN	9.358	1.139	8.216	
(Model $2)^{b}$)				
Intercept	0.844	0.341	2.475	
AN	7.280	0.976	7.463	
log P	0.099	0.020	4.953	
(Model $3)^{c}$)				
Intercept	1.081	0.346	3.127	
AN	6.577	0.993	6.621	
log P	0.094	0.019	4.925	
NUFD	-2.542	1.237	-2.055	

a) $n=37$, $r=0.812$, $r^2=0.657$, $r_{ev}^2=0.535$, $F=67.50$, $s=0.175$, $p<0.01$. *b*) $n=37$, $r=0.895$, $r^2=0.799$, $r_{ev}^2=0.691$, $F=68.71$, $s=0.135$, $p<0.01$. *c*) $n=37$, $r=0.908$, *r*²=0.823, *r*_{cv}=0.689, *F*=51.55, *s*=0.129, *p*<0.01.

Table 5. PE, LSE, LOF, SEP, Q and Q^2 Values Calculated for the Derived Models for Modelling Basic Hydrolytic Rate Constant of Benzoglycolamide Esters

Model Descriptor			PE LSE LOF SEP	Ω	O ²
$1 \quad AN$			0.0374 1.06 1.25 0.1692 4.63 0.666		
2 AN, $log P$	0.0217		0.60 0.8021 0.1273 6.60		0.811
3 AN, log P, NUFD 0.0192 0.53 0.8061 0.1197 7.00					0.833

employed where each variable in the model was held out in turn and using the remaining parameters models were generated. Each descriptor was chosen as input for the software package of SPSS and then the stepwise addition method implemented in the software was used for choosing the descriptors contributing to the hydrolytic rate constant of benzoglycolamide esters.

The specifications for the best-selected MLR models are shown in Table 4. The monoparametric model indicated the importance of charge on amide nitrogen (AN) in contribution to basic hydrolytic rate constant (model 1, Table 4). Addition of log P as an additional parameter to AN, significantly increased the correlation from 0.812 to 0.895 (model 2, Table 4). Similarly the addition of NUFD as a third parameter also increased the correlation from 0.895 to 0.908 (model 3, Table 4). Further the low interrelationship (Table 3) between AN, log P and NUFD also favors the model expressed by these descriptors. The MLR model indicated the importance of electronic parameters AN and NUFD, lipophilic parameter log P in contribution to basic hydrolytic rate constant. It should be noted that the addition of other parameters to AN, log P and NUFD does not improve the correlation coefficient.

Cross Validation The models are cross-validated by leave one out (LOO) technique and the $r_{\rm cv}^2$ values are presented in Table 4. Also we have used the following strategy for testing the validity of the predictive power of the selected MLR model. The MLR models presented in Table 4 were cross validated by the calculation of the statistical parameters PE, LSE, LOF, SEP, Q and Q^2 , the values are presented in Table 5. The low values of PE, LSE, LOF, SEP and high values of Q and Q^2 are also in favor of the selected models.

The hydrolytic rate constant of 37 molecules included in

Fig. 1. Plot of the MLR Model Calculated Basic Hydrolytic Rate Constants against the Experimental Values for Benzoglycolamide Esters

Fig. 2. Plot of the Residuals *versus* Experimental Values of Basic Hydrolytic Rate Constant for the Best MLR Model

the study was calculated by the MLR model containing AN, Log P and NUFD, and the values are presented in Table 1. The residual activity [difference between experimentally observed log κ (log κ *o*) and QSAR calculated log κ (log κ *c*)] is 0.30 in case of compound **1** while in others it is less than or equal to 0.2, which indicates the predictability of the MLR model. Figure 1 shows the plot of MLR predicted *versus* experimental values of the basic hydrolytic rate constants of benzoglycolamide esters. To investigate the existence of a systemic error in developing the MLR model, the residuals of MLR predicted values of basic hydrolytic rate constants were plotted against the experimental values in Fig. 2. The propagation of the residuals on both sides of zero indicates that no systemic error exists in the development of MLR model as suggested by Jalali-Heravi and Kyani*.* 29)

Conclusion

An accurate and versatile MLR model was developed for predicting the basic hydrolytic rate constant of some benzoglycolamide esters in alkaline medium. The problem of selecting the appropriate descriptor as input for MLR model was overcome by Pearson correlation matrix, which can be used as a tool for identifying the appropriate descriptors when a large number of them with different features are available The MLR model indicated the importance of electronic parameters AN and NUFD, lipophilic parameter log P in contribution to basic hydrolytic rate constant. The validity of the models have been established by the determination of suitable statistical parameters.

References and Notes

- 1) Katrizky A. R., Fara D. C., Petruklin R. O., Tathom D. B., Maran U., Lomaka A., Karelson M., *Curr. Top. Med. Chem.*, **2**, 1333—1356 (2002).
- 2) Katrizky A. R., Maran U., Lobawov V. S., Karelson M., *J. Chem. Inf. Comput. Sci.*, **4**, 1—18 (2000).
- 3) Kubinyi H., "QSAR-Hansch Analysis and Related Approaches," Wiley-VCH, New York, 1993, pp. 1—117.
- 4) Larsen P. K., Lifjefors T., Madson U. A., "Textbook of Drug Design and Development," 2nd ed., Harwood Academic Publishers, Australia, 1996, pp. 96—103.
- 5) Hansch C., "Comprehensive Medicinal Chemistry," 1st ed., Vol. 4, Pergamon Press, Oxford, 1990, pp. 9—258.
- 6) Oliveria A. M., Custodio F. B., Donnici C. L., Montanari A. C., *Eur. J. Med. Chem.*, **38**, 141—155 (2003).
- 7) Grover M., Gulati M., Singh S., *J. Chromatogr. B*, **708**, 153—159 (1998).
- 8) Nielson N. M., Bundgaard H., *J. Pharm. Sci.*, **77**, 285—298 (1988).
- 9) Nishikawa J., Tori K., *J. Antibiot.*, **34**, 1641—1644 (1981).
- 10) Nishikawa J., Tori K., Takasuka M., Onone H., Narisada M., *J. Antibiot.*, **35**, 1724—1728 (1982).
- 11) Boyd D. B., *J. Med. Chem.*, **26**, 1010—1013 (1983).
- 12) Nielsen N. M., Bundgaard H., *Int. J. Pharm.*, **39**, 75—85 (1987).
- 13) Plotkowaik Z., *Pharmazie*, **42**, 449—451 (1987).
- 14) Plotkowaik Z., *Pharmazie*, **45**, 291—292 (1990).
- 15) Nagata O., Terasaki T., Yada E., Sato H., Tamai I., Tsuji A., *J. Pharm. Sci.*, **78**, 57—61 (1989).
- 16) Chung W. G. N., Mayer J. M., Kakolidon T. A., Testa B., *Int. J. Pharm.*, **63**, 129—134 (1990).
- 17) Kearney A. S., Stella V. J., *J. Pharm. Sci.*, **82**, 69—72 (1993).
- 18) Guenat B. M., Carrupt P. A., Lisa G., Testa B., Rose S., Thomas K., Jenner P., Ventura P., *J. Pharm. Pharmacol.*, **47**, 861—869 (1995).
- 19) Narasimhan B., Belsare D., Pharande P., Mourya V., Dhake A., *Eur. J. Med. Chem.*, **39**, 827—834 (2004).
- 20) Narasimhan B., Kothawade U. R., Pharande D. S., Mourya V. K., Dhake A. S., *Indian J. Chem.*, **42B**, 2828—2834 (2003).
- 21) Gangwal N. A., Narasimhan B., Mourya V. K., Dhake A. S., *Indian J. Het. Chem.*, **12**, 201—204 (2003).
- 22) Karelson M., Perkson A., *Comp. Chem.*, **23**, 49—59 (1999).
- 23) Furusjo E., Svenson A., Rahmberg M., Andersson M., *Chemosphere*, **63**, 99—108 (2006).
- 24) Dragon 5 Evaluation version, 2004, Talete SRL, Milano, Italy.
- 25) Hyperchem 6.0, 1993, Hypercube, Inc., Florida.
- 26) SPSS for windows, Version 10.05, 1999, SPSS Inc., Bangalore, India.
- 27) Mandloi D., Joshi S., Khadikar P. V., Khosla N., *Bioorg. Med. Chem. Lett.*, **15**, 405—411 (2005).
- 28) Pinheiro A. A. C., Borges R. S., Santos L. S., Alves C. N., *J. Struct. Mol. (Theochem)*, **672**, 215—219 (2004).
- 29) Jalali-Heravi M., Kyani A., *J. Chem. Inf. Comput. Sci.*, **44**, 1328— 1335 (2004).