Eur J Med Chem (1995) 30, 171–175

Short communication

A QSAR study of the biological activities of some benzimidazoles and imidazopyridines against *Bacillus subtilis*

H Ertepinar[1], Y Gök[2], Ö Geban[1], S Özden[3]

[1]*Middle East Technical University, Faculty of Education, Department of Science Education, 06531 Ankara;*
[2]*Karadeniz Technical University, Faculty of Arts and Sciences, Department of Chemistry, Trabzon;*
[3]*Ankara University, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, 06100 Ankara, Turkey*

(Received 22 July 1994; accepted 17 October 1994)

Summary — A set of benzimidazole (I) and imidazopyridine (II) derivatives that have previously been tested for their antibacterial activities against *Bacillus subtilis* were analyzed using the quantitative structure–activity relationship (QSAR) method. The activity contributions for structural and substituent effects were determined from the correlation equations, which were derived using all possible combination and stepwise regression techniques. The best equation was chosen among the other equations by considering the various statistical criteria. The resulting QSAR revealed that the activity contributions of benzimidazoles and imidazopyridines against *B subtilis* depend almost entirely on their relative lipophilic character as defined by their octanol/water partition coefficients, log *P*. The ideal lipophilic character for the investigated imidazole nuclei has been found to be about 4.9.

QSAR / benzimidazole / imidazopyridine / antibacterial activity / *Bacillus subtilis*

Introduction

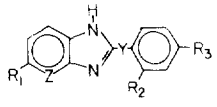
In the last decade, several methods have emerged for the evaluation and prediction of the biological effect of chemicals in drug development. One of the methods that have shown greatest promise is the use of quantitative structure–activity relationships (QSAR). The purpose of this work is to investigate the QSAR of some benzimidazole and imidazopyridine derivatives against *Bacillus subtilis*.

The imidazole nucleus and its related structures are known to play a crucial role in the structure and functioning of a number of biologically important molecules. Among them, benzimidazole is one of the most important heterocyclic rings which showed different biological activities, for example, antibacterial [1–5], antifungal [1–7], antiparasitic [8, 9] and antiviral [10]. Most imidazo[4,5-*b*] and [4,5-*c*]pyridines are biologically active. A number of 2*n*-alkylimidazo[4,5-*b*] and [4,5-*c*]pyridines showed herbicidal [11] antibacterial [12] and antihistaminic [13] activities.

Over the last few years, Pedini *et al* [1, 4] have studied the antibacterial and antimycotic activities of a series of 103 thienyl- and furyl-substituted benzimidazoles and benzoxazoles. They selected a set of 16 representative compounds of the available set of 103, and used linear PLS modeling in order to establish quantitative relationships between the antibacterial activities of a number of benzimidazoles and benzoxazoles [14]. They obtained a straightforward interpretation of the structural features relevant to the activities and the prediction of a possible optimal structure.

Şener *et al* have studied the QSAR of antibacterial benzimidazoles, benzoxazoles and oxazolopyridines against *Klebsiella pneumoniae* [15] and the QSAR of antifungal benzoxazoles and oxazolo[4,5-*b*]pyridines against *Candida albicans* [16, 17]. According to the predictions obtained from QSAR analysis, they concluded that the oxazolo[4,5-*b*]pyridine ring system with a benzyl moiety substituted at position 2 was the most favorable structure over the other heterocyclic nuclei against *K pneumoniae*. They used hydrophobic, electronic, steric and structural parameters and quantum chemical parameters [17] for the QSAR analysis of the benzoxazoles and oxazolopyridines against *C albicans*. From the derived QSAR, they concluded that a more potent compound would possess an oxazolo pyridine ring system substituted with an electron-withdrawing group at position 5 and benzyl moiety at position 2 [16]. In another work [17], they found that the antifungal activity of these compounds against *C albicans* highly correlated with the decreasing order of ε_{LUMO}, MW, R (resonance effect) and ε_{HOMO}.

In the QSAR analysis of previous works [15–17], antibacterial 2-phenyl-substituted benzimidazoles and imidazopyridine derivatives were not investigated. Furthermore, *K pneumoniae* and *C albicans* were used as dependent variables. The QSAR of the previously synthesized 2,5-disubstituted benzimidazoles (I) [18, 19] and 2-substituted imidazo[4,5-*b*]pyridines (II) [20] against *B subtilis* are therefore analyzed in this work.



I, Z: CH Y: - - - or CH_2
II, Z: N Y: - - -

The substituents R_1, R_2 and R_3 in the compounds are electron-donating and electron-withdrawing groups. The antibacterial activity contributions for either ring systems and substituent effects have been determined from the correlation equations.

Chemistry

The physicochemical parameters investigated include the electronic parameter pK_a (dissociation constant), the lipophilic parameter log*P* (octanol/water partition coefficient), molar refractivity (MR) for the steric effects and/or dispersion interaction due to substituents and log MW. The pK_a and log*P* values were determined experimentally. For pK_a measurements, an Orion 601 A digital pH meter and Ingold combined electrode were used. The pK_a values were determined by direct titration in the solvent mixture of ethanol/water (50%). For the determination of the partition coefficient *P* of the compounds, the octanol/water system is used [21, 22]. MR values were taken from the literature [23]. The antibacterial activities of these compounds against *B subtilis* were determined previously [19, 20, 24] using the tube dilution method and given as minimum inhibitory concentration (MIC, μg/ml). The dilution methods are very reproducible and accurate. For these systems the precision of reproducibility is given in the range of 96–97% [25]. The potency has been defined as log1/*C* in the QSAR analysis where *C* is the molar MIC value of the compound and is used as the dependent variable in the QSAR study. The values of all these variables are given in table I.

Table I. Compounds and parameters used in equations [1]–[11].

Compound	R_1	R_2	R_3	*Z*	*Y*	*logP*	*MR*	*logMW*	pK_a	*log1/C (obs)*	*log1/C*[a] *(calc)*	*Difference*
1	CH_3	OCH_3	OCH_3	CH	–	3.78	78.41	2.43	7.25	4.08	4.22	–0.14
2	NO_2	CH_3	H	CH	–	3.42	71.06	2.40	4.87	4.06	4.09	–0.03
3	H	H	CH_3	CH	–	3.67	64.73	2.32	6.90	4.29	4.18	0.11
4	H	OCH_3	H	CH	–	3.02	66.95	2.35	7.17	4.02	3.90	0.12
5	NO_2	CH_3	CH_3	CH	–	4.10	75.68	2.43	5.29	4.39	4.31	0.08
6	Cl	H	Br	CH	CH_2	5.38	77.61	2.51	5.42	4.41	4.39	0.02
7	Cl	H	CH_3	CH	CH_2	5.21	74.38	2.41	7.09	4.31	4.40	–0.09
8	Cl	H	NH_2	CH	CH_2	3.20	74.15	2.41	7.47	4.01	3.99	0.02
9	Cl	H	Cl	CH	CH_2	5.18	74.76	2.44	4.86	4.35	4.41	–0.06
10	NO_2	H	OCH_3	CH	CH_2	3.48	77.93	2.45	4.26	4.06	4.11	–0.05
11	H	H	Cl	N	–	2.95	63.10	2.36	7.46	3.81	3.86	–0.05
12	H	H	CH_3	N	–	3.02	62.72	2.32	7.88	3.92	3.90	0.02
13	H	H	C_2H_5	N	–	3.42	67.34	2.35	7.80	3.95	4.09	–0.14
14	H	H	$C(CH_3)_3$	N	–	4.82	75.58	2.40	7.84	4.60	4.41	0.19
15	H	H	F	N	–	2.52	57.99	2.33	7.52	3.63	3.61	0.02

[a]Calculated using equation [2].

Results and discussion

For the QSAR study, the multiple regression analysis technique was employed. The regression analysis were performed using the SPSS computer program [26]. Regression equations were sought to relate the biological activity against *B subtilis* to possible combinations of the parameters log$P$, MR, log MW, pK_a and the squares of log$P$ and MR. All possible combinations of parameters were considered except that squared terms were only allowed in equations containing the corresponding linear terms. This technique gave a total of 34 possible equations. Only those parameters thought to be probable significant predictors of activity were included in the analysis in order to avoid bias and it is therefore clear that the equation of choice should be the most statistically significant equation. The best equation in the present study was equation [2] (table II) due to the relative importance of the various statistical criteria [27–30].

The best one-term, two-term, three-term, four-term, five-term and six-term equations are listed in table II. All are significant at the 1% level. Clearly the best one-term equation is significant according to the defined criteria, but the additional term $(\log P)^2$ in equation [2] has an F statistic which is significant at the 1% level; equation [2] also has a higher r value and a smaller s value than equation [1]. In fact, all the possible three-term, four-term, five-term and six-term equations inevitably have slightly better r values than equation

Table II. The best equations.

Number	*Equation*	n	r	s	F
The best one-term equation					
1	$\log 1/C = 0.244\ (\pm 0.039)\ \log P + 3.198$	15	0.865	0.135	38.56
The best two-term equation					
2	$\log 1/C = -0.132\ (\pm 0.041)\ (\log P)^2 + 1.316\ (\pm 0.336)\ \log P + 1.131$ $\log P_0 = 4.9$	15	0.930	0.103	38.16
The best three-term equations					
3	$\log 1/C = -0.155\ (\pm 0.052)\ (\log P)^2 - 0.006\ (\pm 0.007)$ MR + $1.532\ (\pm 0.444)\ \log P + 1.059$	15	0.933	0.105	24.76
4	$\log 1/C = -0.135\ (\pm 0.042)\ (\log P)^2 - 0.632\ (\pm 0.702)$ logMW + $1.371\ (\pm 0.344)\ \log P + 2.490$	15	0.935	0.104	25.31
The best four-term equations					
5	$\log 1/C = -0.134\ (\pm 0.044)\ (\log P)^2 - 0.005\ (\pm 0.031)\ pK_a -$ $0.716\ (\pm 0.923)$ logMW + $1.363\ (\pm 0.364)\ \log P + 2.733$	15	0.935	0.109	17.30
6	$\log 1/C = -0.130\ (\pm 0.077)\ (\log P)^2 + 0.001\ (\pm 0.017)$ MR – $0.755\ (\pm 1.687)$ logMW + $1.327\ (\pm 0.650)\ \log P + 2.772$	15	0.935	0.109	17.27
7	$\log 1/C = -0.155\ (\pm 0.054)\ (\log P)^2 \pm 0.004\ (\pm 0.027)\ pK_a -$ $0.005\ (\pm 0.008)$ MR + $1.528\ (\pm 0.466)\ \log P + 1.015$	15	0.933	0.110	16.91
The best five-term equations					
8	$\log 1/C = -0.121\ (\pm 0.093)\ (\log P)^2 - 0.008\ (\pm 0.037)\ pK_a +$ $0.003\ (\pm 0.021)$ MR – $1.073\ (\pm 2.358)$ logMW + $1.250\ (\pm 0.780)\ \log P + 3.588$	15	0.935	0.115	12.50
9	$\log 1/C = -1.500 \times 10^{-5}\ (\pm 0.001)\ MR^2 + 0.004\ (\pm 0.029)\ pK_a -$ $0.154\ (\pm 0.069)\ (\log P)^2 + 1.523\ (\pm 0.594)\ \log P - 0.003\ (\pm 0.156)$ MR + 0.950	15	0.933	0.116	12.18
10	$\log 1/C = 5.796 \times 10^{-5}\ (\pm 0.001)\ MR^2 - 0.132\ (\pm 0.086)\ (\log P)^2 -$ $0.770\ (\pm 1.801)$ log MW + $1.341\ (\pm 0.731)\ \log P - 0.007\ (\pm 0.155)$ MR + 3.059	15	0.935	0.115	12.44
The best six-term equation					
11	$\log 1/C = 8.550 \times 10^{-5}\ (\pm 0.001)\ MR^2 - 0.008\ (\pm 0.040)\ pK_a -$ $0.123\ (\pm 0.102)\ (\log P)^2 - 1.110\ (\pm 2.550)$ log MW + $1.267\ (\pm 0.857)\ \log P -$ $0.008\ (\pm 0.164)$ MR + 4.047	15	0.935	0.122	9.27

[2], but they contain at least one coefficient with a poor F statistic ($p > 0.01$). The poor F statistics for the coefficients of equations [3]–[11] are given in table III. Therefore, equation [2] is considered to be the most statistically significant equation. The overall F statistic for this equation is statistically significant at the $p < 0.0001$ level, and the individual F statistics for its coefficients are all significant at the $p < 0.01$ level. Thus, this equation has a multiple correlation coefficient of 0.930 and explains 86.49% of the variation in the biological data. On the other hand, the use of more than three independent variables for this data set (n = 15) is not justified due to a significantly increased probability of obtaining chance correlation [31].

Table III. The poor F statistics ($p > 0.01$) for the coefficients of variables in the equations [3–11].

	F	p
Equation 3		
$(\log P)^2$	8.984	0.0121
MR	0.586	0.4600
Equation 4		
log MW	0.811	0.3872
Equation 5		
$(\log P)^2$	9.220	0.0125
pK_a	0.022	0.8841
log MW	0.602	0.4559
Equation 6		
$(\log P)^2$	2.844	0.1226
MR	0.007	0.9372
log MW	0.200	0.6640
log$P$	4.173	0.0683
Equation 7		
$(\log P)^2$	8.088	0.0174
pK_a	0.017	0.8980
MR	0.395	0.5438
Equation 8		
$(\log P)^2$	1.681	0.2271
pK_a	0.042	0.8422
MR	0.028	0.8715
log MW	0.207	0.6598
log$P$	2.570	0.1433
Equation 9		
MR^2	0.000	0.9894
pK_a	0.015	0.9037
$(\log P)^2$	4.990	0.0524
log$P$	6.583	0.0304
MR	0.000	0.9848
Equation 10		
MR^2	0.003	0.9593
$(\log P)^2$	2.323	0.1618
log MW	0.183	0.6790
log$P$	3.360	0.1000
MR	0.002	0.9666
Equation 11		
MR^2	0.005	0.9438
pK_a	0.040	0.8461
$(\log P)^2$	1.446	0.2636
log MW	0.189	0.6749
log$P$	2.183	0.1778
MR	0.003	0.9607

An examination of the intervariable correlations involving all the variables in table IV demonstrates the relatively large correlations between MR and log MW (0.884) in equations [6], [8], [10] and [11]; log MW and pK_a (–0.652) in equations [5], [8] and [11]; log$P$ and log MW (0.687) in equations [4]–[6], [8], [10] and [11], and log$P$ and MR (0.684) in equation [3] and [6–11]. High correlations produce a good relationship in a set of predictor variables (collinearity). When high collinearity exists, the regression analyses using the given set of independent variables cannot be performed effectively [32].

The stepwise multiple regression method was also applied to our system and the result verified the best fit of equation [2] which was obtained by the best possible regression method. The stepwise development of the best equation is given in table V.

To verify the validity of equation [2], the cross-validation is applied to the original data set and the resulting Press (predicted residual sum of squares) was calculated. The calculated overall Press is 0.15793 and SSY (sum of the squares of the response values of the total observations) is 0.81293. The overall Press is smaller than the SSY of the given set. This proves that the best fitted equation predicts better than chance and can be considered as 'statistically significant' [27, 33]. The ratio Press/SSY for the given set is 0.19427. For a reasonable QSAR model, Press/SSY should be smaller than 0.4 [33].

Table IV. Correlation matrix of all variables used in QSAR analysis.

	log1/C	*logP*	*MR*	*log MW*	pK_a
log 1/C	1.000				
log$P$	0.865	1.000			
MR	0.689	0.684	1.000		
log MW	0.548	0.687	0.884	1.000	
pK_a	–0.318	–0.347	–0.516	–0.652	1.000

Table V. The stepwise development of equation [2].

Equation	*n*	*r*	*s*	*F*
$\log 1/C = 0.244\ (\pm 0.039)\ \log P + 3.198$	15	0.865	0.135	38.56
$\log 1/C = 1.316\ (\pm 0.336)\ \log P - 0.132\ (\pm 0.041)\ (\log P)^2 + 1.131$	15	0.930	0.103	38.16

The parabolic relationship of activity against *B subtilis* ($\log 1/C$) on the lipophilicity of the benzimidazoles and imidazopyridine derivatives suggests that $\log 1/C$ initially increases as the lipophilicity increases up to an ideal $\log P$ value, $\log P_o$ (~ 4.9), and then it decreases after this point. It is expected that the compounds with $\log P$ values of ~ 4.9 have the highest $\log 1/C$ values.

References

1 Pedini M, Bistocchi GA, De Meo G, Ricci A (1987) *Il Farmaco Ed Sci* 42, 541–547
2 Badawey EAM, Rida SM, Soliman FSG, Kappe T (1989) *Monatsh Chem* 120, 1159–1164
3 Habib NS, Abdel-Hamid S, El-Hawash M (1989) *Il Farmaco Ed Sci* 44, 1225–1232
4 Pedini M, De Meo G, Ricci A, Bastianini L, Jacquignon P (1990) *Il Farmaco Ed Sci* 45, 303–312
5 Badawey EAM, Gohar YM (1992) *Il Farmaco Ed Sci* 47, 489–496
6 Güneş HS, Coşar G (1992) *Arzneim-Forsch Drug Res* 42, 1045–1048
7 Baillie AC, Buss AD, Parsons JH, Russell PE (1987) Eur Pat EP 219, 192; *Chem Abstr* (1987) 107, 134305q
8 Dubey R, Abuzar S, Sharma S, Chatterjee RK, Katiyar JC (1985) *J Med Chem* 26, 1748–1750
9 Kumar P, Agarwal SK, Bhakuni DS (1990) *Ind J Chem* 29B, 1077–1080
10 Dominianni SJ (1990) *Oppi Briefs* 22, 106–108
11 Röchling H, Büchel KH (1971) *Chem Ber* 104, 344–347
12 Brown DJ, Danckwerts L, Grigg GW, Iwai Y (1979) *Aust J Chem* 32, 453–458
13 Özden S (1978) Thesis for Assoc Prof, Ankara University, Turkey
14 De Meo G, Pedini M, Ricci A *et al* (1990) *Il Farmaco Ed Sci* 45, 313–330
15 Yalçın İ, Şener E (1993) *Int J Pharm* 98, 1–8
16 Şener E, Yalçın İ, Sungur E (1991) *Quant Struct-Act Relat* 10, 223–228
17 Türker L, Şener E, Yalçın İ, Akbulut U, Kayalıdere İ (1990) *Sci Pharm* 58, 107–113
18 Gümüş F, Özden S, Özden T, Abbasoglu U (1986) *J Fac Pharm Ankara* 16, 1–13
19 Özden S, Özden T, Gümüş F, Akın A (1987) *J Fac Pharm Ankara* 17, 1–10
20 Özden S, Özden T, Gümüş F, Akın S (1985) *J Fac Pharm Ankara* 15, 79–90
21 Biagi GL, Barbaro AM, Guerra MC, Forti CG, Fracasso ME (1974) *J Med Chem* 17, 28–33
22 Fujita T, Iwasa J, Hansch C (1964) *J Am Chem Soc* 86, 5175–5183
23 Hansch C, Leo A, Unger SH, Kim KH, Nikaitani D, Lien EJ (1973) *J Med Chem* 16, 1207–1216
24 Özden S, Özden T, Gümüş F, Akın S (1986) *J Fac Pharm Ankara* 16, 37–44
25 Lennette EH, Balows A, Hausler WJ, Shadomy HJ (1985) *In: Manual of Clinical Microbiology* (4th ed). American Society for Microbiology, Washington, DC, USA, 972–977
26 Norusis MJ (1991) *In: The SPSS Guide to Data Analysis for SPSS/PC+* (2nd ed). SPSS Inc, Chicago, IL, USA
27 Rawlings JO (1988) *Applied Regression Analysis: A Research Tool*. Wadsworth and Brooks / Cale, Pacific Grove, CA, USA, 189
28 Darper NR, Smith H (1981) *Applied Regression Analysis*. John Wiley and Sons Inc, New York, USA, 294–342
29 Howell DC (1987) *Statistical Methods for Psychology*. PWS-Kent Publishing Company, Boston, MA, USA, 463–507
30 Anderson TW, Stanley LS (1978) *An Introduction to the Statistical Analysis of Data*. Houghton Mifflin Company, Boston, MA, USA, 619–637
31 Topliss JG, Costello RJ (1972) *J Med Chem* 15, 1066–1068
32 Norman HI, Hull CH, Jenkins JG, Steinbrenner K, Bent DH (1975) *SPSS Statistical Package for the Social Sciences*. McGraw Hill Inc, New York, USA, 340–341
33 Wold S (1991) *Quant Struct-Act Relat* 10, 191–193