Eur J Med Chem (1997) 32, 311–319

Carbonic anhydrase inhibitors. Part 41. Quantitative structure–activity correlations involving kinetic rate constants of 20 sulfonamide inhibitors from a non-congeneric series

BW Clare[1]*, CT Supuran[2]

[1]*School of Physical Sciences, Engineering and Technology, Murdoch University, Murdoch, WA 6150, Australia;*
[2]*Laboratorio di Chimica Inorganica e Bioinorganica, Dipartimento do Chimica, Università Delgi Studi di Firenze, Via G Capponi 7, 50121 Florence, Italy*

(Received 12 July 1996; accepted 23 September 1996)

Summary — A quantitative structure–activity relationship study is presented for 20 sulfonamide inhibitors of carbonic anhydrase. These drugs do not form a classical congeneric series, in that the only common factor is the sulfonamide group, which is attached to a variety of substituted aromatic and hetero-aromatic nuclei. The important factors were local factors such as Mulliken charge on atoms of the sulfonamide group, and global factors such as the size and shape of the molecule, its calculated frontier orbital energies, and its lipophilicity. Good correlations were obtained with the equilibrium constant and the kinetic association rate constant, but not with the kinetic dissociation rate constant.

carbonic anhydrase / kinetic study / QSAR / charge / LUMO

Introduction

Sulfonamides with the general formula RSO_2NH_2 constitute a thoroughly investigated class of inhibitors of the zinc enzyme carbonic anhydrase (CA, EC 4.2.1.1) [1–3]. They bind as anions to the Zn(II) ion within the enzyme active site (substituting the metal-bound water molecule) [4, 5] with abnormally high affinities of around 10^6–10^9 M^{-1} (for isozyme CA II, one of the most effective catalysts known [1–3]). This was hard to explain initially, since sulfonamides are relatively poor complexing agents for transition metal ions [6], and even if recently a large number of such complexes have been reported [7], their stability is relatively low [8]. Later it was shown that the strong inhibitory properties of this class of substances can be accounted by several factors. These include firstly, the stability of enzyme–inhibitor complex being stabilized by a large favorable enthalpy change associated with the binding of the sulfonamide to the enzyme [6, 9].

Secondly, the weak coordination bond between the active site Zn(II) ion and the sulfonamido nitrogen is enormously supplemented by the cooperative interactions of the organic moieties of the inhibitor with amino acid side chains from the active site, as proved by X-ray crystallographic determinations of structures of several sulfonamide adducts with diverse CA isozymes [10–12].

Thirdly, kinetic factors control the association rates of the inhibitor with the enzyme, as well as the dissociation of the enzyme-inhibitor adducts. As the rate constants for the reactions of CA with its inhibitors were thoroughly investigated [13–17], some interesting facts emerged.

Thus, in contrast to the metal complexing anions (such as cyanide, cyanate, thiocyanate, sulfide, etc) possessing CA inhibitory properties [18] for which association rates (k_{on}) of the order of magnitude

*Correspondence and reprints
Abbreviations: AM1: Austin model 1; CA: carbonic anhydrase; CNDO: complete neglect of differential overlap; HOMO: highest occupied molecular orbital; LUMO: lowest unoccupied molecular orbital; QSAR: quantitative structure–activity relationships.

of 10^8–10^9 $M^{-1}s^{-1}$ and dissociation rates (k_{off}) of 10^3–10^6 s^{-1} [17] were measured, the sulfonamides have much slower association rates than the anions (k_{on} in the range of (0.00331–31) x 10^6 M^{-1} s^{-1} whereas the dissociation rates are also much slower, but ranging only between 0.01 and 0.05 s^{-1} [3, 17].

$$E + I \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} E...I$$

Thus, the changes in the association rates of sulfonamides with CA control the stability of the E–I complex and hence the potency of the inhibitor, a situation quite uncommon for enzyme–inhibitor reactions [18, 19].

Such factors, as well as the need to design tight binding inhibitors of these enzymes (with a k_{off} around 10^{-3}–10^{-4} s^{-1}) prompted us to hypothesize in an earlier work [3] that a QSAR study in which kinetic instead of thermodynamic constants (such as the inhibition constant, K_I usually employed in QSAR studies) would be used in calculation, might bring novel insights regarding the factors that govern activity in this class of pharmacological agents. Apart from the theoretical interest such calculations might have, sulfonamide CA inhibitors are firmly established in clinical medicine in the treatment or prevention of glaucoma [20], gastroduodenal ulcers [21], mountain sickness [22] and other conditions. The thermodynamic inhibition constant is related to the kinetic parameters by the formula $K_I = k_{off}/k_{on}$.

In this paper we report the first such QSAR study, involving kinetic as well as thermodynamic inhibition constants for 20 sulfonamides for which literature data regarding association and dissociation rates were available [3]. The object of the study is two-fold: first, to enable simple prediction of the potency of as yet unsynthesized substances; and second, to throw light on the mechanism of action of the drugs, and their interaction with the active site of the enzyme. The series of compounds considered is unusual in that it does not form a congeneric series.

The only structure common to all compounds is the sulfonamide moiety. All have an aromatic nucleus, but these are of varying identity. Previous studies of such diverse structures include many on narcosis, employing hydrophobicity [23], and two on hallucinogens [24, 25], using HOMO energy. Other studies on CA inhibitors have used more homogeneous series of drugs [26–29].

The drugs as well as their inhibition and kinetic constants are shown in table I. Mention should be made that some of the compounds included in the study are used clinically. Thus, acetazolamide **O** methazolamide **N**, and ethoxzolamide **T** have been in parenteral use, originally as diuretics [1], but then as antiglaucoma or antiulcer drugs [1, 2, 20, 21]; benzolamide **Q** is a renal drug with a specific pharmacology [30], whereas MK-927 **P** is a member of a new class of water-soluble topically active antiglaucoma drugs recently introduced in clinical medicine [20]. Piretanide **G**, hydrochlorothiazide **H**, chlorothiazide **J**, furosemide **L** and benzothiazide **M** belong to the class of chloruretic or high ceiling diuretics, being widely used in the treatment of cardiac edema and possessing renal mechanisms apart from CA inhibition [17]. Other such compounds (such as sulfanilamide **I**, benzenesulfonamide **K** or inhibitors **R**, **S**) are historically important molecules (CA inhibition by sulfonamides was discovered working with sulfanilamide [31]) or led to the development of very effective classes of such inhibitors (for a recent review, see ref [2]).

Calculations

The molecules were set up using the program PCMODEL [32], and their geometries were fully optimized, first using PCMODEL to explore conformational space, and finally with AM1 [33]. These conformations are of course the conformations of minimum energy in vacuo, not the unknown conformations of the molecules when bound to the enzyme. The AM1 calculations were run with the program MOPAC 93 [34], and CNDO calculations to obtain the atomic charges with the programs CNINDO [35] and GEOMOS [36]. The descriptors calculated are listed in table II.

The linear dimensions of the molecule were calculated from the principal components of the inertial tensor, I_x, I_y and I_z, and the molecular weight W, using the formula:

$$A_x = 2\sqrt{\frac{5(I_y + I_z - I_x)}{2W}}$$

and similar formulae for A_y and A_z. These represent the lengths in angstroms of an ellipsoid with the same principal axes of inertia as the molecule, and with a density of 1 da/Å^3. The van der Waals area and volume were calculated with the GEPOL [37] option of the program ARVOMOL [38]. The octanol/water partition coefficient was calculated with the program Clog*P* for Windows [39]. For consistency, and to facilitate prediction, all values of log *P* used were calculated, although the measured values were available for many of these drugs.

The charges on the sulfonamide group have been shown to be important for CA inhibitors [27–29, 40], and the polarizabilities and frontier orbital energies

Table I. Drugs **A–T** and their inhibition and kinetic constants.

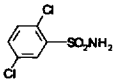
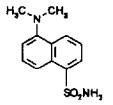
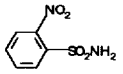
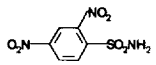
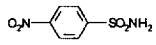
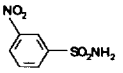
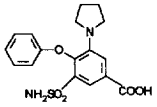
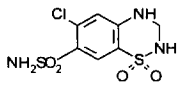
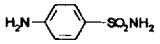
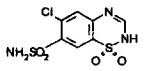
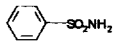
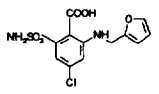
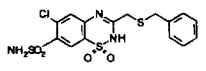
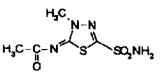
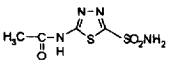
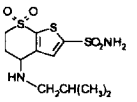
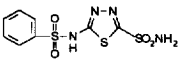
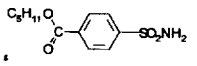
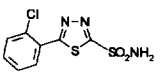
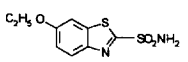
Drug	*Structure*	$K_I \times 10^9$ *M*	$k_{on} \times 10^{-6}$ *L/mol* s^{-1}	k_{off} s^{-1}
A		10 000	0.003	0.03
B		6000	0.002	0.01
C		3000	0.003	0.01
D		3000	0.003	0.01
E		900	0.04	0.04
F		800	0.01	0.008
G		17 000	0.003	0.050
H		2350	0.013	0.030
I		750	0.033	0.024
J		460	0.066	0.030
K		440	0.10	0.044
L		80	0.3	0.024
M		16	0.6	0.010

Table I. *Continued.*

Drug	*Structure*	$K_I \times 10^9$ *M*	$k_{on} \times 10^{-6}$ *L/mol* s^{-1}	k_{off} s^{-1}
N		13	3.5	0.042
O		7	3.0	0.021
P		4	4.4	0.020
Q		0.9	31	0.028
R		0.8	15	0.012
S		0.8	30	0.024
T		0.7	14	0.010

were found to be significant in a previous study [41]. A correlation dendrogram [42] for the entire data set, using the algorithm of Spearman to combine correlations [43], and excluding the dependent variables, is shown in figure 1.

MOPAC has been well tested with hypervalent compounds [44], and performs well. However, because it does not consider *d* orbitals, the charges calculated for the sulfur and oxygen atoms would be unrealistic. The S=O bond is treated as being extremely polar. For this reason, CNDO [45], which does treat *d* orbitals, was used to calculate these quantities.

Results

This study involves selection of descriptors from a pool. Topliss and Edwards have shown that this entails the risk of obtaining formally significant correlations by chance [46]. In order to determine whether

Table II. Descriptors used in this study.

	Descriptor	*Symbol*	*Units*
1	Charge on sulfonamide hydrogen, CNDO	Q_H	e
2	Charge on sulfonamide sulfur, CNDO	Q_S	e
3	Charge on sulfonamide nitrogen, CNDO	Q_N	e
4	Charge on sulfonamide oxygen, CNDO	Q_O	e
5	Energy of highest occupied molecular orbital, AM1	E_H	eV
6	Energy of lowest unoccupied molecular orbital, AM1	E_L	eV
7	Component of polarizability tensor, AM1	Π_{xx}	au
8	Component of polarizability tensor, AM1	Π_{yy}	au
9	Component of polarizability tensor, AM1	Π_{zz}	au
10	Length of axis of molecule	A_x	Å
11	Length of axis of molecule	A_y	Å
12	Length of axis of molecule	A_z	Å
13	van der Waals volume of molecule, ARVOMOL	V_w	Å^3
14	van der Waals area of molecule, ARVOMOL	A_w	Å^2
15	Log of octanol–water partition coefficient, CLog*P*	Log *P*	
16	Dipole moment of molecule, AM1	*D*	Debye

or not stepwise regressions on the data set were invalidated by chance correlations, a stepwise regression by Effroymson's algorithm [47] was carried out on a data set including all of the descriptors, and compared with 1000 similar regressions on the same data set, but with the dependent variable values randomly reassigned [48]. The statistical significance obtained by this procedure was 0.0017 for log k_{on}, and 0.0023 for log K_I. Thus, for these models, the risk of chance correlation in the stepwise regressions is at an acceptable level. No significant model could be obtained for log k_{off}. This is not surprising, given the smaller degree of variation of this parameter over the set of drugs. The models obtained for log k_{on} and log K_I by Effroymsons algorithm were:

$$\log k_{on} = 51.2\, Q_N + 1.22\, E_L + 0.373\, A_x - 0.139\, V_w + 0.125\, A_w + 6.33 \quad [1]$$

$$\log K_I = -55.7\, Q_N - 1.64\, E_L + 0.459\, D - 0.572\, A_x - 0.251\, A_y + 0.625 \log P - 8.56 \quad [2]$$

The data sets were subjected to 'all possible subsets' regression using the program BMDP-9R [49]. The selection criterion was Mallows' C_p [50], with the default penalty of 2.0. This technique resulted in the same optimal subsets for both data sets, identical to that obtained for log k_{on} above, although there were differences in ordering in intermediate models.

Considering only log k_{on}, the one-descriptor equation involving A_x was the best, followed by Q_N, with R^2 of 0.31 and 0.29, respectively. All other one-descriptor equations had R^2 poorer than 0.13. The best two-descriptor equation involved both of these, with R^2 of 0.482, followed by an equation containing A_x and Π_{yy}, with R^2 at 0.476, and all others poorer than 0.42. In the next six poorer equations, Π_{xx} or Π_{zz} substituted for A_x, and Q_O, Q_H or D substituted for Q_N.

The best three-descriptor equations, with R^2 of 0.617, 0.589 and 0.575, lacked a charge term, but involved A_x, A_y, A_w, V_w, Π_{yy} and Π_{zz} – thus being dependent on the size, shape and polarizability of the molecule. The five next-best equations all contained Q_N and A_x, with one of D, E_L, Π_{yy}, Q_H, and Π_{xx}, and had R^2 ranging from 0.570 to 0.548.

The best four-descriptor equations again included charge terms, Q_O, Q_S, Q_O and Q_N, respectively. The best of these, with R^2 of 0.749, included A_x, V_w and A_w. In the other three, the only other substitution was A_y for A_x.

The best five-descriptor equation, with an R^2 of 0.838 and involving Q_N, E_L, A_x, V_w and A_w was the best subset, in terms of its C_p of –1.61. The remainder had R^2 of from 0.794 to 0.807. In the next best E_H substituted for E_L, and in the third best, A_y substituted for A_x. The term Q_N occurred in all of the ten best five-descriptor equations.

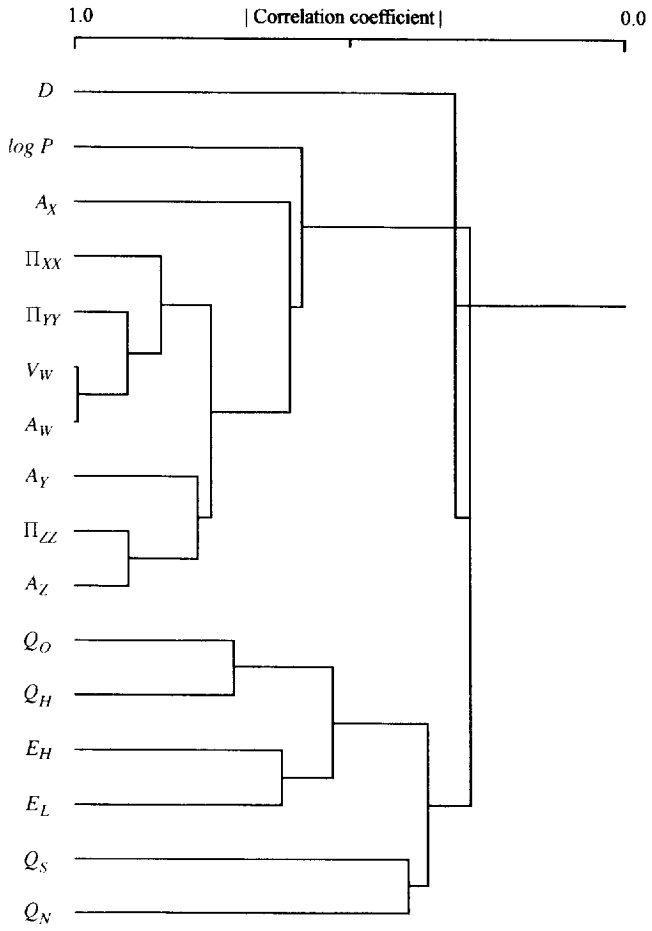


Fig 1. Dendrogram of correlation matrix of descriptors for compounds **A–T**.

The next descriptor to enter was D, with a substantial improvement in R^2, but a deterioration in C_p. Further addition of descriptors led to only small improvements in R^2. Thus the best equations for both dependent variables is analogous to equation [1] for log k_{on}, obtained by the Effroymson algorithm.

A rule-of-thumb in multiple regression analysis is that there should always be at least three times as many cases as variables. On this basis, since there are 20 drugs in this set, we can consider at most six variables in any one equation. The two subsets chosen on the basis of Mallows' C_p have five variables.

The optimal linear regression models were:

$$\log k_{on} = 51.2\ (\pm 11.1)\ Q_N + 1.22\ (\pm 0.40)\ E_L + 0.373\ (\pm 0.111)\ A_x - 0.138\ (\pm 0.028)\ V_w + 0.125\ (\pm 0.026)\ A_w + 6.33\ (\pm 3.15) \quad [3]$$

$R^2 = 0.839$, $Q^2 = 0.669$, $S = 0.705$, $F = 14.5$, $\alpha = 4 \times 10^{-5}$, all $\alpha_i < 0.01$

$$\log K_l = -46.5\ (\pm 10.6)\ Q_N - 0.988\ (\pm 0.38)\ E_L - 0.418\ (\pm 0.106)\ A_x + 0.144\ (\pm 0.027)\ V_w - 0.128\ (\pm 0.025)\ A_w - 3.26\ (\pm 2.98) \quad [4]$$

$R^2 = 0.857$, $Q^2 = 0.701$, $S = 0.669$, $F = 16.8$, $\alpha = 2 \times 10^{-5}$, $\Lambda = 38.0$, and all $\alpha_i < 0.002$ except that for E_L which was 0.021

Here, R^2 is the square of the multiple correlation coefficient, Q^2 is the same quantity based on predicted errors (the leave-one-out technique [51]), S is the standard error of estimate of the equation, F is the Fisher variance ratio, α is the probability (statistical significance) based on this F, and the α_i values are the individual statistical significance of each of the coefficients of the equation, based on a Student's t-test. The numbers in parentheses are standard errors of estimate for each coefficient in the equation. The diagnostic Λ is defined as:

$$\Lambda = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i}$$

where n is the number of descriptors and the λ_i are the eigenvalues of the correlation matrix of descriptors [52]. A value of Λ greater than 5 is taken to indicate that a collinearity problem exists in the equation. The value of 38 suggests that the equation is unreliable. Examination of the eigenvector matrix showed that the problem was entirely due to the very high correlation between V_w and A_w.

A repetition of all the possible subsets regression with the A_w term deleted gave a more satisfactory result in this regard. For one and two variable subsets, the results were identical with those obtained previously. For three variables, a good subset was obtained with Π_{zz}, A_y and A_z ($C_p = 1.89$, $R^2 = 0.666$), with Q_N and then Q_O substituting for A_y with less satisfactory statistics, and the fourth and fifth best equations contained Q_N, A_x and D and E_L respectively.

The best four-descriptor equations were less satisfactory ($C_p = 2.38$, $R^2 = 0.702$), but contained Q_N, Π_{zz}, A_z and one of A_y, E_L, and D in order of increasing C_p (ie, poorer fit).

The best five-descriptor equation ($C_p = 2.08$, $R^2 = 0.757$) contained Q_N, E_L, D, A_x and log P. With less satisfactory fit, E_H substituted for E_L ($C_p = 2.90$). The remaining four of the best six five-descriptor equations contained Q_N, A_z and Π_{zz}, with combinations of E_L, A_y and D.

The best six-descriptor equation was the overall best, with $C_p = 1.37$ and $R^2 = 0.822$. It contained Q_N, E_L, D, A_x, A_y and log P. In the three next best equa-

tions, Π_{zz} substituted for A_y, and then E_H substituted for E_L, followed by A_x for A_y. In the best seven-descriptor equation, A_z and Π_{zz} replaced A_y in the best six-descriptor equation, giving $C_p = 2.78$ and $R^2 = 0.837$. The activity parameters log k_{on} and log K_I are very strongly negatively correlated. The optimal fit for log k_{on} with six variables was:

$$\log k_{on} = 60.3\ (\pm 12.8)\ Q_N + 1.91\ (\pm 0.49)\ E_L + 0.551\ (\pm 0.100)\ A_x + 0.286\ (\pm 0.131)\ A_y - 0.446\ (\pm 0.111)\ D - 0.659\ (\pm 0.212) \log P + 11.30\ (\pm 3.49) \quad [5]$$

$R^2 = 0.823$, $Q^2 = 0.580$, $S = 0.767$, $F = 10.1$, $\alpha = 3 \times 10^{-4}$, all $\alpha_i < 0.01$ except A_y (0.047)

Substitution of the compound variables $\Pi_{xx} + \Pi_{yy} + \Pi_{zz}$ for Π_{xx}, Π_{yy} and Π_{zz}, A_w/V_w for A_w and V_w, and of $E_L - E_H$ for E_L and E_H did not lead to improved models. Similarly, substitution of various sums of Q_S, Q_H, Q_O and Q_N for Q_N did not improve the statistics.

In order to determine whether the fit of equation [5] could be improved by nonlinear transformations of the independent variables, the ACE technique of Breiman and Friedman was applied [48, 53–55]. This method gives plots of optimal smooth nonlinear transformations of the variables, maximizing R^2. All independent, but not the dependent variables were transformed, giving the plots shown in figure 2.

The plots for log P, A_y and to a lesser extent, Q_N suggest piecewise linear transforms [55] with breakpoints at 0.35, 6.0 and –0.24 respectively, and those for A_x, D and E_L possible but weak polynomial transforms. When these were tried, the lower part of the log P and A_y transforms were found nonsignificant, and both parts of the Q_N transform were significant, as were the A_x^2 and A_x^3, but not the A_x terms. The resulting equation was an excellent fit, with an R^2 of 0.945, but had too many terms for the number of points, and was inferior in the Q^2 value of 0.609 (ie, it was a poorer predictor). The best predictor was obtained by retaining only the upper branches of the log P and A_y transforms. This was as follows:

$$\log k_{on} = 66.4\ (\pm 10.8)\ Q_N + 2.11\ (\pm 0.41)\ E_L + 0.571\ (\pm 0.085)\ A_x + 0.555\ (\pm 0.156)\ A_y' - 0.418\ (\pm 0.087)\ D - 0.976\ (\pm 0.227) \log P' + 11.11\ (\pm 2.82) \quad [6]$$

$R^2 = 0.879$, $Q^2 = 0.732$, $S = 0.634$, $F = 15.71$, $\Lambda = 1.76$, $\alpha = 3 \times 10^{-5}$, all $\alpha_i < 0.001$ except A_y' (0.0035), where log P' = log P (log P>0.35), otherwise log $P' = 0.35$, and $A_y' = A_y$ ($A_y > 6.0$) otherwise $A_y' = 6.0$

The fit of the data to this model is shown in figure 3. This model, for both dependent variables, is analogous to equation [2] for log K_I, and its statistics are superior to those for equation [5].

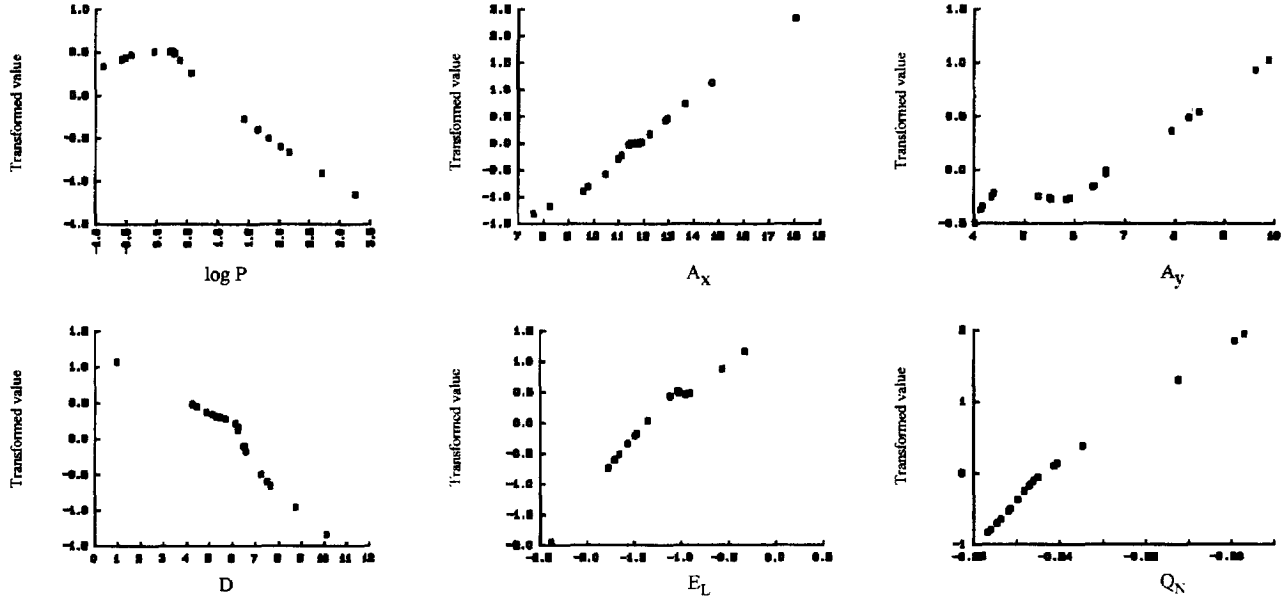


Fig 2. ACE transformation plots for descriptors of equation [5].

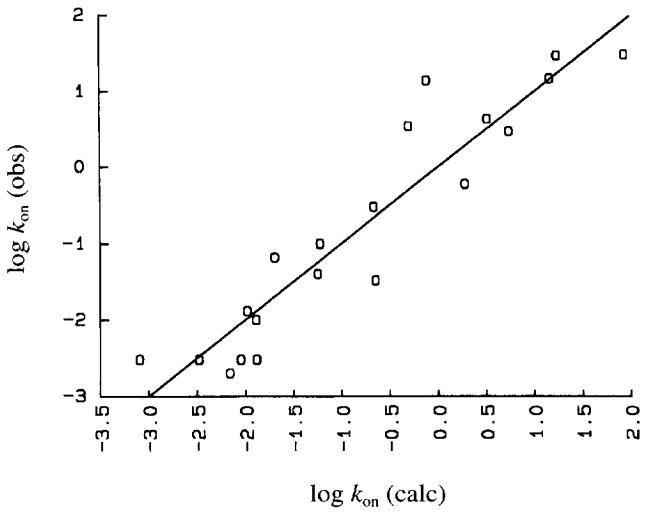


Fig 3. Plot of observed k_{on} against k_{on} calculated by equation [6].

In the five best n-term equations for log k_{on}, for each of n = 1 to 7, the coefficients of E_L, Q_N, A_x, and Π_{zz} were always positive and those of Π_{xx}, D, and log P were always negative. Those of A_y, A_z, and Π_{yy} varied. Thus aromatic sulfonamides that have a low dipole moment, a low negative charge on the sulfonamide N, a high E_L, and which are long, relatively wide ($A_y > 6$ Å) and relatively hydrophilic (log $P < 0.35$) are predicted to have a large value of the association rate constant k_{on}.

A similar treatment was applied to log K_I. The coefficient for A_y was statistically non-significant in the original regression, and the ACE transformation plot of A_y gave a less pronounced discontinuity at 6.0 than it did with log k_{on}, but on applying the same transformations, the lower parts of A_y and log P transforms were again insignificant, leaving the same six significant terms as before. The resulting equation was:

$$\log K_I = -61.0\ (\pm 11.5)\ Q_N - 1.82\ (\pm 0.43)\ E_L - 0.592\ (\pm 0.091)\ A_x - 0.484\ (\pm 0.166)\ A_y' + 0.434\ (\pm 0.093)\ D + 0.919\ (\pm 0.242)\ \log P' - 8.41\ (\pm 3.01) \quad [7]$$

$R^2 = 0.864$, $Q^2 = 0.677$, $S = 0.678$, $F = 13.76$, $\Lambda = 1.76$, $\alpha = 6 \times 10^{-5}$, all $\alpha_i < 0.001$ except A_y' (0.012), E_L (0.001) and log P' (0.002), where log P' = log P (log $P > 0.35$), otherwise log P' = 0.35, and $A_y' = A_y$ ($A_y > 6.0$) otherwise $A_y' = 6.0$

Thus the regression for log K_I gives very similar results to log k_{on}, but with somewhat poorer statistics.

The values of the calculated descriptors found to be most important are given in table III.

Discussion

The activities of the drugs depend on three factors, which may be exemplified for log k_{on}. First is the positive correlation with charge on sulfonamide N, which is a common finding with sulfonamide CA inhibitors. More negative values of the charge on this nitrogen, lead to smaller rate constants, k_{on}. The charge on sulfonamide S or O can substitute for this, but less well. The magnitude of the dipole moment is similarly implicated, the rate constant decreasing with increasing dipole moment.

Second is the positive correlation with E_L, for which E_H can substitute, again less well. Larger values of E_L lead to greater rates. This may be indicative of a charge-transfer interaction between the aromatic part of the drug and some electron donor in the receptor site. The negative sign of E_L for all of the drugs is consistent with this. The positive sign of the correlation suggests that the charge-transfer interaction tends to inhibit the formation of the intermediate complex necessary for the enzyme inhibition. That is, the sulfonamide–enzyme complex with strong charge-transfer bonding is still able to catalyze the CO_2/water reaction. The location of the charge-transfer binding site in the enzyme is perhaps such that the sulfonamide is kept away from the site of enzyme inhibition.

There is some uncertainty about the physical significance of calculated LUMO energies. They are less soundly based, and correlate less well between calculation methods than HOMO energies [56], but calculations on model systems [57] indicate that within limits, they are meaningful. The finding that E_L is an important factor in this group of neutral compounds contrasts with our earlier finding that E_H was significant in the same negative sense, but in a much narrower group of positively charged CA inhibitors.

Third is a geometric factor, in which V_w and A_w appear with opposite signs (negative and positive respectively), which may signify degree of departure from sphericity, and A_x, the length of the longest axis of the molecule, appears with a positive sign. The greater the departure from spherical and the greater the length of the molecule, the larger is k_{on}. If one discounts this observation because of the collinearly between V_w and A_w, one has a positive correlation with the length of the molecule, a less certain positive correlation with its breadth, and a negative correlation with lipophilicity. The latter would indicate that the environment of the active site is generally hydrophilic. The collinearity, being a consequence of the physical nature of the two descriptors, does not necessarily invalidate the regression.

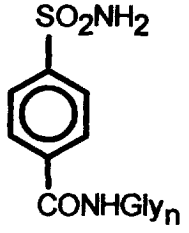
A recent finding illustrating the size dependence, obtained by calculation in our study, was the report of Whitesides' group that in a series of inhibitors

Table III. Values of the most important descriptors.

Drug	Q_N *(e)*	E_L *(eV)*	A_x *(Å)*	A_y *(Å)*	A_z *(Å)*	*Log P*	*D (Debye)*	V_w *(Å³)*	A_w *(Å²)*
A	–0.2520	–0.906	9.59	6.39	2.06	1.65	4.24	153.51	189.05
B	–0.2546	–0.950	11.91	5.84	2.67	1.83	6.49	215.07	250.77
C	–0.2405	–1.476	7.60	6.61	2.54	0.40	6.58	143.26	177.79
D	–0.2346	–2.380	10.99	6.37	2.30	0.30	4.85	163.29	203.50
E	–0.2450	–1.781	11.82	4.14	2.29	0.58	4.44	143.52	176.83
F	–0.2467	–1.495	10.47	5.29	2.28	0.58	5.30	143.48	181.13
G	–0.2537	–1.039	11.11	9.88	4.88	3.27	7.67	297.87	341.83
H	–0.2498	–1.034	11.50	6.62	2.61	–0.50	10.12	195.93	234.70
I	–0.2567	–0.327	9.78	4.01	2.48	–0.57	6.24	135.33	171.90
J	–0.2460	–1.712	11.45	6.63	2.61	–0.41	6.55	191.65	228.16
K	–0.2518	–0.564	8.27	4.17	2.59	0.31	5.13	124.71	157.87
L	–0.2515	–1.021	11.81	8.48	4.37	2.17	5.41	245.95	293.88
M	–0.2470	–1.662	18.05	7.93	3.31	2.03	8.73	317.64	367.63
N	–0.2461	–0.959	11.41	5.51	2.15	0.27	6.13	170.19	215.19
O	–0.2123	–1.358	11.62	4.34	2.17	–0.87	7.26	152.17	193.13
P	–0.2561	–1.568	12.25	9.62	3.18	–0.03	5.47	248.97	314.57
Q	–0.1992	–1.705	13.67	5.92	4.03	0.22	7.53	221.82	270.63
R	–0.2482	–1.120	12.87	8.27	3.56	2.71	0.97	232.53	287.44
S	–0.1969	–1.776	12.94	5.54	1.92	1.43	5.69	192.74	227.90
T	–0.2413	–1.125	14.75	4.39	2.05	2.02	6.26	197.18	239.67

obtained by attaching glycine or $(Gly)_n$ units (n from 2 to 4) to 4-carboxybenzenesulfonamide (compounds of type **U**) the inhibition constants greatly increased with n [58].

The nonlinear dependence on hydrophobicity is probably consistent with Kubinyi's bilinear equation



U: n = 1-4

[59], based on partitioning between multiple compartments, but this cannot be tested in the present model, because it would entail too great a proliferation of fitted parameters. The nonlinear dependence on A_y is less certain, but if it exists, is presumably of steric origin. The effectiveness of the drug decreases with decreasing breadth, but only down to 6 Å, beyond which it becomes constant.

As with any QSAR study, the conclusions of this study cannot be safely extrapolated beyond the range of the data on which it is based. Nonlinearities have been found within that range. No doubt others exist beyond it, and as the 'all possible subsets regression' resuls indicate, even the identity of the relevant parameters must be regarded as tentative. Within the range of the data, reasonable prediction should be possible by equation [6], which predicts approximately 73% of the variance of the association rate constant.

References

1 Maren TH (1967) *Physiol Rev* 47, 595–781
2 Supuran CT (1994) In: *Carbonic Anhydrase and Modulation of Physiologic and Pathologic Processes in the Organism* (Puscas I, ed), Helicon, Timisoara, 29–113
3 Maren TH, Clare BW, Supuran CT (1994) *Roum Chem Quart Rev* 2, 259–282
4 Liljas A, Hakansson K, Jonsson BH, Xue Y (1994) *Eur J Biochem* 219, 1–10
5 Lindskog S, Liljas A (1993) *Cur Opin Struct Biol* 3, 915–920
6 Coleman JE (1975) *Ann Rev Pharmacol* 15, 221–242
7 Alzuet G, Ferrer S, Borras J, Supuran CT (1994) *Roum Chem Quart Rev* 2, 283–300
8 Luca C, Barboiu M, Supuran CT (1991) *Rev Roum Chim* 36, 1169–1173
9 Taylor PW, King RW, Burgen ASV (1970) *Biochemistry* 9, 2368–2645
10 Vidgren J, Liljas A, Walker NPC (1990) *Int J Biol Macromol* 12, 342–344
11 Vidgren J, Svensson A, Liljas A (1993) *Int J Biol Macromol* 15, 97–100
12 Hakansson K, Liljas A (1994) *FEBS Lett* 350, 319–322
13 Kernohan JC (1966) *Biochim Biophys Acta* 118, 405–412
14 Taylor PW, Burgen ASV (1971) *Biochemistry* 10, 3859–3866
15 Olander J, Bosen SF, Kaiser ET (1973) *J Am Chem Soc* 95, 1616–1621
16 King RW, Maren TH (1974) *Mol Pharmacol* 10, 344–348
17 Maren TH (1992) *Mol Pharmacol* 41, 419–426
18 Stewart JJP (1989) *J Comp Chem* 10, 209–220
19 Schloss JV (1988) *Acc Chem Res* 21, 348–353
20 Maren TH (1995) *J Glaucoma* 4, 49–62
21 Puscas I, Supuran CT (1996) In: *Aparelho Digestivo* (Coelho J, ed), MEDSI, Rio de Janeiro, 1704–1734
22 Larson EB, Roach RC, Schoene RB, Hornbein TF (1982) *J Am Med Assoc* 248, 328–332
23 Overton E (1899) *Vierteljahrschr Naturforsch Ges Zuer* 44, 88–135
24 Snyder SH, Merril CR (1965) *Proc Natl Acad Sci* 54, 258
25 Kang S, Green JP (1970) *Proc Natl Acad Sci* 67, 62
26 Carotti I, Raguseo C, Campagna F, Langridge R, Klein TE (1989) *Quant Struct–Act Relat* 8, 1–10
27 De Benedetti PG, Menziani MC (1985) *Quant Struct–Act Relat* 4, 23–28
28 Menziani MC, De Benedetti PG (1992) *Struct Chem* 3, 215–219
29 Menziani MC, De Benedetti PG, Gago F, Richards WG (1989) *J Med Chem* 32, 951–956
30 Maren TH (1982) In: *Orphan Drugs* (Karch FE, ed), Dekker, New York, 89–115
31 Mann T, Keilin D (1940) *Nature* 146, 164–165
32 PCMODEL V. 5.0, Available from Serena Software, Box 3076 Bloomington IN 47402–3076 USA
33 Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107, 3902–3909
34 MOPAC 93.00 (1993) Stewart JJP, Fujitsu Ltd., Tokyo, Japan, also Stewart JJP (1995) *QCPE Bull* 15, 13–14 (copyright Fujitsu 1993, all rights reserved)
35 Dobosh P (1975) *QCPE Bull* 10, 281
36 Rinaldi D, Hoggan PE, Cartier A (1989) *QCPE Bull* 9, 128–129
37 Pascual-Ahuir JL, Silla E (1988) *QCPE Bull* 8, 133–134
38 Pacios LF (1995) *QCPE Bull* 15, 17
39 ClogP for Windows, ver 1.0.0. available from BioByte Corp. 201 West 4th St, Suite 204, Claremont CA 91711, USA
40 Kishida K (1978) *Chem Pharm Bull* 26, 1049–1053
41 Supuran CT, Clare BW (1995) *Eur J Med Chem* 30, 687–696
42 Sokal RR, Sneath PH (1963) *Principles of Numerical Taxonomy*, Freeman, San Francisco, 182
43 Spearman C (1913) *Br J Psychol* 5, 417
44 Stewart JJP (1990) *J Comput-Aided Mol Des* 4, 1–103
45 Pople JA, Beveridge DL (1970) *Approximate Molecular Orbital Theory*, McGraw Hill, New York
46 Topliss JG, Edwards RJ (1979) *J Med Chem* 22, 1238–1244
47 Efroymson MA (1962) In: *Mathematical Methods for Digital Computers* (Ralston A, Wilf HS, eds), Wiley, New York, 191
48 Clare BW (1994) In: *Methods and Principles in Medicinal Chemistry* (Mannhold R, Krogsgaard-Larsen P, Timmerman H, eds), vol 3, VCH, Weinheim, *Advanced Computer-Assisted Techniques in Drug Discovery* (van de Waterbeemd H, ed), pp 284, 289
49 BMDP Dynamic, Release 7 available from BMDP Statistical Software, Inc, 1440 Sepulveda Boulevard, Suite 316, Los Angeles, CA 90025, USA
50 Daniel C, Wood FS (1971) *Fitting Equations to Data*, Wiley, New York
51 Cramer RD III, Bunce JD, Patterson DE (1988) *Quant Struct–Act Relat* 7, 18
52 Chatterjee S, Price B (1977) *Regression Analysis by Example*, Wiley, New York, 200
53 Breiman L, Friedman JH (1985) *J Am Stat Assoc* 80, 580–619
54 Clare BW (1993) *Chemometrics Intelligent Lab Syst* 18, 71–93
55 Frank IE, Lanteri S (1988) *Chemometrics Intelligent Lab Syst* 3, 301–313
56 Clare BW (1994) *Theor Chim Acta* 87, 415–430
57 Clare BW (1995) *THEOCHEM* 337, 139–150
58 Jain A, Whitesides GM, Alexander RS, Christianson DW (1994) *J Med Chem* 37, 2100–2105
59 Kubinyi H (1977) *J Med Chem* 20, 625–629