*Eur. J. Med. Chem. 33 (1998) 489–500*



# Carbonic anhydrase inhibitors – Part 47: Quantum chemical quantitative structure–activity relationships for a group of sulfanilamide Schiff base inhibitors of carbonic anhydrase

Claudiu T. Supuran[a], Brian W. Clare[b,c]*

[a]*Laboratorio di Chimica Inorganica e Bioinorganica, Dipartimento di Chimica, Università Delgi Studi di Firenze, Via G. Capponi 7, 50121 Firenze, Italy*
[b]*Division of Science, Murdoch University, Murdoch, W.A. 6150, Australia*
[c]*Department of Chemistry, University of Western Australia, Nedlands, W.A. 6009, Australia*

(Received 14 August 1997; accepted 28 January 1998)

**Abstract** – A series of sulfanilamide Schiff base inhibitors of CA I and CA II have been studied by the semi-empirical AM1 method. The charges on the atoms of the sulfonamide group, and the dipole moment, have been calculated by four methods: a standard vacuum calculation, a solution calculation by the COSMO method, a solution calculation with, in addition, the charges and dipole moments calculated by fitting to the calculated electrostatic potential, and a calculation by the older CNDO method. The data were subjected to a classical multiple regression analysis with care to avoid the possibility of chance correlation or collinearity. The ACE technique was also used to allow for nonlinearity. A number of statistically significant equations were derived, which were mostly consistent with previous studies. While the ESP-based charges gave the best equations, the improvement was not sufficient to convincingly exclude the other methods. This is the first study which evidenced by means of QSAR calculations isozyme-specific features of the two isozymes CA I and CA II, in their interaction with sulfonamide inhibitors. © Elsevier, Paris

**ACE / carbonic anhydrase inhibitor / charge / collinearity / QSAR / quantum chemical**

## 1. Introduction

A QSAR study on a group of sulfonamide inhibitors of carbonic anhydrase (both CA I and CA II) is presented. The predictor variables are all based on calculation, either quantum chemical or, for hydrophobicity and $pK_a$, empirical. The predictors chosen have mostly been reported as relevant in the past.

De Benedetti [1] et al. have reported correlations of CA inhibitory activity of benzenesulfonamides with total charge on the $SO_2NH_2$ moiety, calculated by the CNDO method, and also with HOMO and LUMO orbital energy of the anionic forms. They also obtained similar correlations with aliphatic and bridged diaryl sulfonamides. The regression coefficient for LUMO energy was remarkably similar in the two cases. The same group [2] have also reported correlations with charge on the O and $NH_2$ components of the $SO_2NH_2$ group, for a series of heterocyclic sulfonamides. This study also used CNDO charges. Menziani and De Benedetti [3, 4] have also published a molecular modelling study in which they examine the binding of thiadiazole sulfonamides to CA, finding HOMO energy, total charge on $SO_2NH_2$ and local dipole character of the thiadiazole ring to be important contributors to binding energy. This study used the more recent AM1 molecular orbital method.

Because it has frequently been reported that the activity of CA inhibitors is related to charge on the sulfonamide group, an attempt has been made to determine how this charge can best be calculated. In particular, modern semi-empirical methods such as AM1 and PM3 do not use *d* orbitals on sulfur. This results in very high Mulliken charges on sulfur and

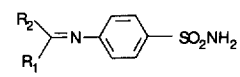
*Correspondence and reprints
*Abbreviations:* ACE, Alternating Conditional Expectations; AM1, Austin Model 1; CA, Carbonic Anhydrase; CNDO, Complete Neglect of Differential Overlap; COSMO, Conductive Shielding Model; ESP, Electrostatic Potential; HOMO, Highest Occupied Molecular Orbital; LUMO, Lowest Unoccupied Molecular Orbital; QSAR, Quantitative Structure–Activity Relationships; MLR, Multiple Linear Regression; PCA, Principal Component Analysis; SCRF, Self-Consistent Reaction Field

the atoms bonded to it. Although atomic charge is not a well-defined quantity, these charges are probably not realistic. Electrostatic potential is a quantum theoretic observable, and an alternative to Mulliken charge is to calculate the electrostatic potential at points on the surface of the molecule, and assign charges to the nuclear positions so as to best fit this potential. This still gives rather high charges. The older semi-empirical method CNDO, though usually considered obsolete, does use *d* orbitals on S, and gives much lower atomic charges on the atoms of the $SO_2NH_2$ group, and thus might be more realistic. Further, most quantum theoretic calculations on drug molecules have been done on isolated molecules in vacuum. It may be more realistic to carry out SCRF calculations in a medium of dielectric constant approximating that of water, a procedure which is within the capabilities of modern programs.

## 2. Calculations

The formulae of the compounds considered, and their activities in inhibiting CA I and CA II are given in *table I*. The preparation of the compounds and the determination of their activities has been described previously [5, 6].

**Table I.** Carbonic anhydrase inhibitors considered, with their activities against CA I and CA II.




| No. | $R_1$ | $R_2$ | $C_I$CA I (x $10^6$ M) | $C_I$CA II (x $10^8$ M) |
|---|---|---|---|---|
| 1 | Phenyl | H | 18 | 27 |
| 2 | 2-Hydroxyphenyl | H | 35 | 41 |
| 3 | 2-Nitrophenyl | H | 9 | 21 |
| 4 | 4-Chlorophenyl | H | 25 | 28 |
| 5 | 4-Hydroxyphenyl | H | 14 | 19 |
| 6 | 4-Methoxyphenyl | H | 13 | 19 |
| 7 | 4-Dimethylaminophenyl | H | 10 | 8 |
| 8 | 4-Nitrophenyl | H | 13 | 5 |
| 9 | 4-Cyanophenyl | H | 4 | 11 |
| 10 | 3-Methoxy-4-hydroxyphenyl | H | 5 | 8 |
| 11 | 3,4-Dimethoxyphenyl | H | 7 | 3 |
| 12 | 3-Methoxy-4-acetoxyphenyl | H | 3 | 10 |
| 13 | 2,3-Dihydroxy-5-formylphenyl | H | 4 | 2 |
| 14 | 2-Hydroxy-3-methoxy-5-formylphenyl | H | 5 | 3 |
| 15 | 3,4,5-Trimethoxyphenyl | H | 5 | 3 |
| 16 | 3-Methoxy-4-hydroxy-5-bromophenyl | H | 12 | 4 |
| 17 | 2-Furyl | H | 3 | 5 |
| 18 | 5-Methyl-2-furyl | H | 3 | 4 |
| 19 | Pyrol-2-yl | H | 5 | 2 |
| 20 | Imidazol-4(5)-yl | H | 1 | 12 |
| 21 | 2-Pyridyl | H | 2 | 9 |
| 22 | 3-Pyridyl | H | 4 | 8 |
| 23 | 4-Pyridyl | H | 4 | 5 |
| 24 | Styryl | Me | 14.4 | 0.39 |
| 25 | 4-Methoxystyryl | Me | 9.4 | 0.12 |
| 26 | 4-Dimethylaminostyryl | Me | 0.6 | 0.10 |
| 27 | 3,4,5-Trimethoxystyryl | Me | 1.3 | 0.24 |
| 28 | Styryl | Ph | 20.9 | 0.56 |
| 29 | 4-Methoxystyryl | Ph | 19.0 | 1.50 |
| 30 | 4-Dimethylaminostyryl | Ph | 16.0 | 1.69 |
| 31 | 3,4,5-Trimethoxystyryl | Ph | 10.7 | 2.35 |
| 32 | 3,4,5-Trimethoxystyryl | $4\text{-MeOC}_6H_4$ | 12.5 | 1.27 |
| 33 | 3-Nitrostyryl | $4\text{-MeOC}_6H_4$ | 6.3 | 0.65 |
| 34 | 3,4,5-Trimethoxystyryl | $4\text{-H}_2NC_6H_4$ | 10.6 | 0.85 |
| 35 | 3,4,5-Trimethoxystyryl | $4\text{-PhC}_6H_4$ | 25.0 | 2.48 |

The molecules were constructed and optimized with PCMODEL [7], obtaining the conformation of lowest energy by exploring the potential energy surface with the dihedral driver. These conformers were then further optimized with the MOPAC 93 [8] AM1 Hamiltonian [9]. The molecules were then aligned with the origin in the center of the sulfanilamide benzene ring, the positive $x$-axis in the direction of the 1-carbon atom (the atom bonded to the sulfonamide S), the positive $y$-axis in the direction of the center of the 2-3 C–C bond, and the $z$-coordinate of the sulfonamide nitrogen was made positive.

Quantum chemical indices were calculated using MOPAC 93, and the AM1 Hamiltonian. HOMO and LUMO energies and Mulliken charges of atoms of the $SO_2NH_2$ group were calculated both in vacuo and in a medium of dielectric constant 78.4 by the COSMO method [10], and in the latter medium, ESP-based charges were also calculated by the method of Merz and Besler [11]. Charges for the two oxygens, $Q_O$ and the two hydrogens, $Q_H$ were summed, as were charges $Q_o$ and $Q_m$ and electrophilic superdelocalizabilities $S_o^E$ and $S_m^E$ on the two o- and m- positions on the sulfanilamide benzene ring. Dipole moments $D$ and their components along the coordinate axes $D_x$, $D_y$ and $D_z$ were calculated. The principal moments of the polarizability tensor $P_{xx}$, $P_{yy}$ and $P_{zz}$ were calculated by MOPAC 93 using the method of Kurtz and Korambath [12]. Dimensions of the inertial ellipsoid, $L_x$, $L_y$ and $L_z$ were calculated as described previously [13]. Values of the hydrophobicity were obtained with *C*log*P* [14], and an estimate of the p$K_a$ values of the sulfonamide and Schiff base moiety, p$K_1$ and p$K_2$, with PALLAS [15]. In some of the compounds, a final value of log P could not be obtained, because of a missing fragment, which was the same in every case. This was handled by using the partial result given, and defining an indicator variable, $I_p$, with the value 0 if the $R_2$ in *table I* was H or Me, and 1 if it was phenyl or substituted phenyl, i.e. 0 if the fragment is absent and 1 if it is present. The resulting regression coefficient could not be regarded as an estimate of the hydrophobicity contribution of the missing fragment, as $I_p$ may also contain contributions from the fragment unrelated to hydrophobicity. Molecular surface area and volume were calculated by ARVOMOL [16] using the GEPOL option. Superdelocalizabilities, and local dipole indices, defined as the mean of the absolute difference in charge of each bonded pair of atoms, were calculated from the AM1 solution (SCRF) results, using a program written by the authors.

It will be assumed that, in general, AM1 is to be preferred over CNDO and an SCRF calculation is to be preferred over a vacuum calculation. The issue to be decided is whether the atomic charges calculated by any one method is superior to those calculated by another. Elsewhere, we have shown [17] that ESP based charges relate better to actual charge distributions in molecules than do Mulliken charges, but that there is considerable variation in the correlation between Mulliken and ESP-based charges. There has been no previous comparison of the methods in an actual QSAR. In the analysis of the results, four sequences were tried. In the first, designated Solution, solution Mulliken charges, frontier orbital energies, dipole moments, and superdelocalizabilities were used. In the second, designated Solution ESP, ESP-based charges were substituted for Mulliken charges on the atoms of the $SO_2NH_2$ moiety, and ESP-based dipole moments for the dipole moment based on the expectation value of the dipole moment operator. In the third sequence, designated Vacuum, vacuum Mulliken charges, and expectation dipole moments and frontier orbital energies were used, and in the fourth series, CNDO-based $SO_2NH_2$ charge and dipole parameters and AM1 vacuum frontier orbital energies were used. Electrophilic superdelocalizabilities [18] $S_o^E$, $S_m^E$ and $S_p^E$ and local dipole index [19] $D_l$, were calculated from the AM1 solution calculations throughout.

## 3. Collinearity and chance correlation

QSAR investigations are plagued by twin problems: chance correlation, in which statistical significance is inflated by the multitude of possible models, and collinearity, in which misleading predictivity is obtained by an inadequate exploration of model space. The first of these was explored by Topliss et al. [20, 21], in studies based on correlations with random numbers. It can be estimated to some extent by repeated random reassignment of activity values to the matrix of independent variables, and repetition of the selection process [22, 23]. If these 'nonsense' regressions frequently lead to correlations as good as the original one, then the original one is nonsignificant.

The second problem, collinearity, can be attacked in several ways. One of these is Livingstone and Rahr's CORCHOP [24] procedure, in which variables which correlate with other variables are systematically eliminated. This, if applied too severely, can lead to loss of information. More seriously, it can fail to detect multicollinearities which involve linear combinations of three or more variables. Other methods involve the formation of orthogonal linear combinations of the original independent variables, and the use of these new variables, fewer in number than the original but involving all of them. The problem with these methods is that the meaning of the new variables may be obscure.

The method chosen here is to obtain the activity as a function of a subset of the original variables, by the use of the 'all possible subsets' algorithm of Furnival and Wilson [25], as implemented in the 9R module of the BMDP package [26]. This results in the 10 best subsets for each subset size, up to the total number, using Mallow's $C_p$ as a criterion, with the default penalty of 2.0 [27]. Each subset in the overall best 5 selected on the basis of Mallow's $C_p$ is then tested for collinearity by computing the $\Lambda$ statistic [28], defined as

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

where $n$ is the number of variables in the equation and the $\lambda_i$ are the eigenvalues of the correlation matrix of the independent variables. If $\Lambda < 5.0$, the subset is considered free from correlation problems, and the equation is accepted. If not, the eigenvector matrix is examined. For the eigenvectors corresponding to excessively small eigenvalues (those which take $\Lambda$ beyond 5.0), the variables with the highest loadings are those causing the collinearity problem in the respective eigenvector (collinearity). If there is more than one collinearity, and one variable is common to more than one of them, this variable is dropped. If not, the variable with the highest loading is dropped, or alternatively, two or more variables are combined (if they are commensurate) to form a new variable, the original ones being dropped. This is guided by the physical nature of the variables, and their coefficients in the regression. The all possible subsets and $\Lambda$ calculation are then repeated, until no collinearities remain. Terms with statistical significance $\alpha$ greater than 0.05 are dropped from the equation. The result of this procedure is a number of equations involving usually a small subset of the original variables, and free from collinearity.

## 4. Results

The values of the calculated descriptors are given in *tables III–IX*, which are available over the Internet, [29] and the correlations between the atomic charges on the atoms of the sulfonamide group in the dendrograms in *figures 1–4*. Of these, *figure 2* with $Q_S$ and $Q_H$ showing a measure of positive correlation with each other and negative correlation with $Q_N$, $Q_O$ and $Q_C$ is nearest to what is expected, and most in line with the methanesulfonamide study [17]. This immediately suggests that it is the solution ESP charge results which are physically most reasonable, if all of the atomic charges are considered.

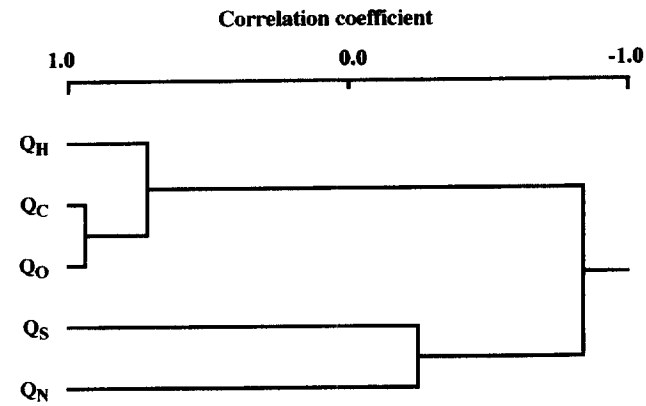



**Figure 1.** Dendrogram of correlations between atomic charges: solution calculation.

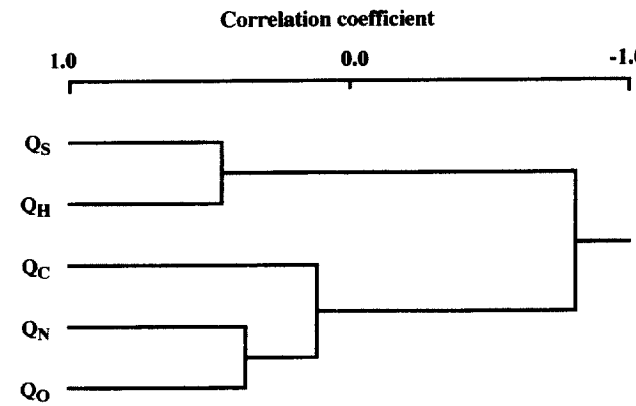



**Figure 2.** Dendrogram of correlations between atomic charges: solution ESP calculation.

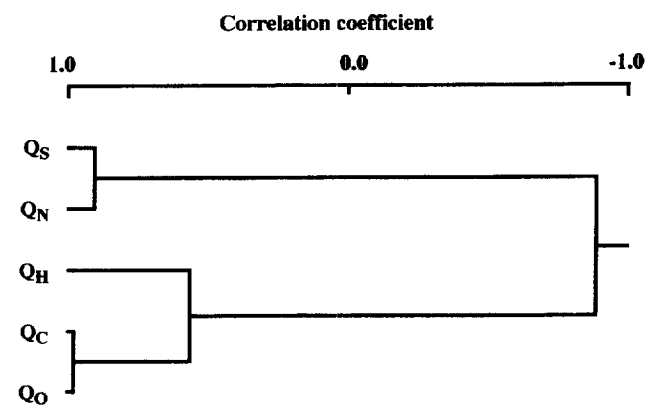



**Figure 3.** Dendrogram of correlations between atomic charges: vacuum calculation.

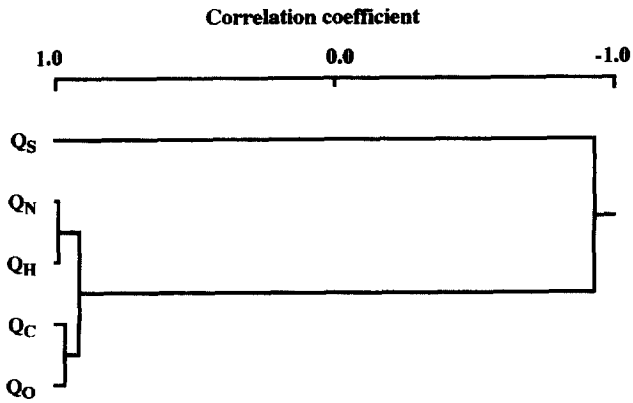



**Figure 4.** Dendrogram of correlations between atomic charges: CNDO calculation.

The probability of chance correlation in the regressions of CA inhibitory activity on the calculated descriptors was estimated as described previously [22, 23]. The results are presented in *table II.* All of the probabilities are well within acceptable limits.

*4.1. CA I*

*4.1.1. Solution*

On applying the all possible subsets and principal components procedure, the best two equations found were:

$$\log K_i = C_1 D_x + C_2 D_z + C_3 D_1 + C_4 Q_H + C_5 \log P + C_6 S_m^E + C_7 (S_p^E - 0.5 S_o^E) + C_8 \quad (1)$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| C | 0.1003 | -0.0949 | 3.133 | -736.7 | 0.3893 | 80.81 | -907.7 | 472.5 |
| σ | 0.0189 | 0.0203 | 0.635 | 95.9 | 0.0473 | 16.52 | 148.0 | 60.1 |
| α | 0.00001 | 0.00007 | 0.00004 | 0.00000 | 0.00000 | 0.00004 | 0.00000 | 0.00000 |

$N = 35$, $R^2 = 0.838$, $Q^2 = 0.683$, $F = 19.0$, $s = 0.190$, $\Lambda = 3.66$, where N is the number of compounds, $R^2$ the squared multiple correlation coefficient, $Q^2$ is $R^2$ based on prediction residuals (the 'leave one out' technique), F the Fisher variance ratio and s the standard error of estimate, and $K_i$ is $10^6$ $K_I$ for CA I. The value of $Q^2$ is the best obtained for the CA I solution series.

$$\log K_i = C_1 D + C_2 P_{xx} + C_3 Q_H + C_4 \log P + C_5 pK_1 + C_6 pK_2 + C_7 \quad (2)$$

**Table II.** Estimated probability of chance correlation in the multiple regression analyses.

| | CA I | CA II |
|---|---|---|
| Solution | 0.00024 | 0.00015 |
| Solution ESP | 0.00021 | 0.00075 |
| Vacuum | 0.00007 | 0.00005 |
| CNDO | 0.00014 | 0.00082 |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| C | -0.0727 | -0.00790 | -286.0 | 0.2986 | 0.5931 | 0.0629 | 160.8 |
| σ | 0.0205 | 0.00130 | 97.1 | 0.0506 | 0.2260 | 0.0295 | 58.4 |
| α | 0.00140 | 0.00000 | 0.00644 | 0.00000 | 0.01388 | 0.04197 | 0.01020 |

$N = 35$, $R^2 = 0.765$, $Q^2 = 0.584$, $F = 15.0$, $s = 0.222$, $\Lambda = 3.00$. The descriptors log P, $pK_1$ and $P_{xx}$ have a strong tendency to enter the equation, always with a positive coefficient, and $Q_H$, with a negative coefficient.

*4.1.2. Solution ESP*

Application of the all possible subsets MLR/PCA technique to the solution ESP results for CA I gives equations (3) and (4).

$$\log K_i = C_1 Q_S + C_2 D_1 + C_3 \log P + C_4 pK_1 + C_5 S_p^E + C_6 \quad (3)$$

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| C | -11.74 | 1.713 | 0.1956 | 1.0062 | 172.8 | 59.07 |
| σ | 3.65 | 0.621 | 0.0457 | 0.1571 | 62.0 | 13.0 |
| α | 0.00320 | 0.00990 | 0.00019 | 0.00000 | 0.00927 | 0.00009 |

$N = 35$, $R^2 = 0.762$, $Q^2 = 0.573$, $F = 18.5$, $s = 0.218$, $\Lambda = 2.52$.

$$\log K_i = C_1 Q_S + C_2 P_{xx} + C_3 \log P + C_4 pK_1 + C_5 \quad (4)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | -12.43 | -.00374 | 0.2601 | 0.9130 | 25.5 |
| σ | 3.78 | 0.00129 | 0.0531 | 0.1574 | 9.9 |
| α | 0.00259 | 0.00711 | 0.00003 | 0.00000 | 0.01535 |

$N = 35$, $R^2 = 0.736$, $Q^2 = 0.609$, $F = 20.9$, $s = 0.226$, $\Lambda = 2.84$.

Application of the ACE technique of Breiman and Friedman [30] to this led to the transformation plots of *figure 5*, and a piecewise linear fit of the first two variables to the equation

$$\log K_i = C_1 Q_S^* + C_2 P_{xx}^* + C_3 \log P + C_4 pK_1 + C_5 \quad (5)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | -24.53 | -0.00340 | 0.2461 | 0.6939 | 60.1 |
| σ | 4.53 | 0.00120 | 0.0395 | 0.1296 | 11.9 |
| α | 0.00001 | 0.00818 | 0.00000 | 0.00001 | 0.00002 |

$N = 35$, $R^2 = 0.813$, $Q^2 = 0.729$, $F = 33.2$, $s = 0.189$, $\Lambda = 2.48$, where $Q_S^* = Q_S$ if $Q_S > 2.667$ and $Q_S^* = 2.667$ otherwise,and $P_{xx}^* = P_{xx}$ if $P_{xx} > 300$, and $P_{xx}^* = 300$ otherwise. This was the best $Q^2$ obtained for the CA I solution ESP series.

As for the Mulliken case, the descriptors log P and $pK_1$ strongly tend to enter the equation with a positive coefficient. Here, however, $Q_S$ rather than $Q_H$ tends to enter, with a negative coefficient.

*4.1.3. Vacuum*

Application of the all possible subsets MLR/PCA technique to the CA I vacuum results leads to equation (6).

$$\log K_i = C_1 D_x + C_2 D_z + C_3 P_{xx} + C_4 Q_H + C_5 \log P + C_6 pK_1 + C_7 (S_p^E - 0.5S_o^E) + C_8 \quad (6)$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| C | 0.1111 | -0.1248 | -.00792 | -291.1 | 0.3675 | 0.7025 | -401.2 | 143.1 |
| σ | 0.0280 | 0.0308 | 0.00129 | 70.4 | 0.0432 | 0.2210 | 178.8 | 36.5 |
| α | 0.00048 | 0.00039 | 0.00000 | 0.00031 | 0.00000 | 0.00210 | 0.04191 | 0.00055 |

$N = 35$, $R^2 = 0.811$, $Q^2 = 0.624$, $F = 16.9$, $s = 0.200$, $\Lambda = 4.30$.

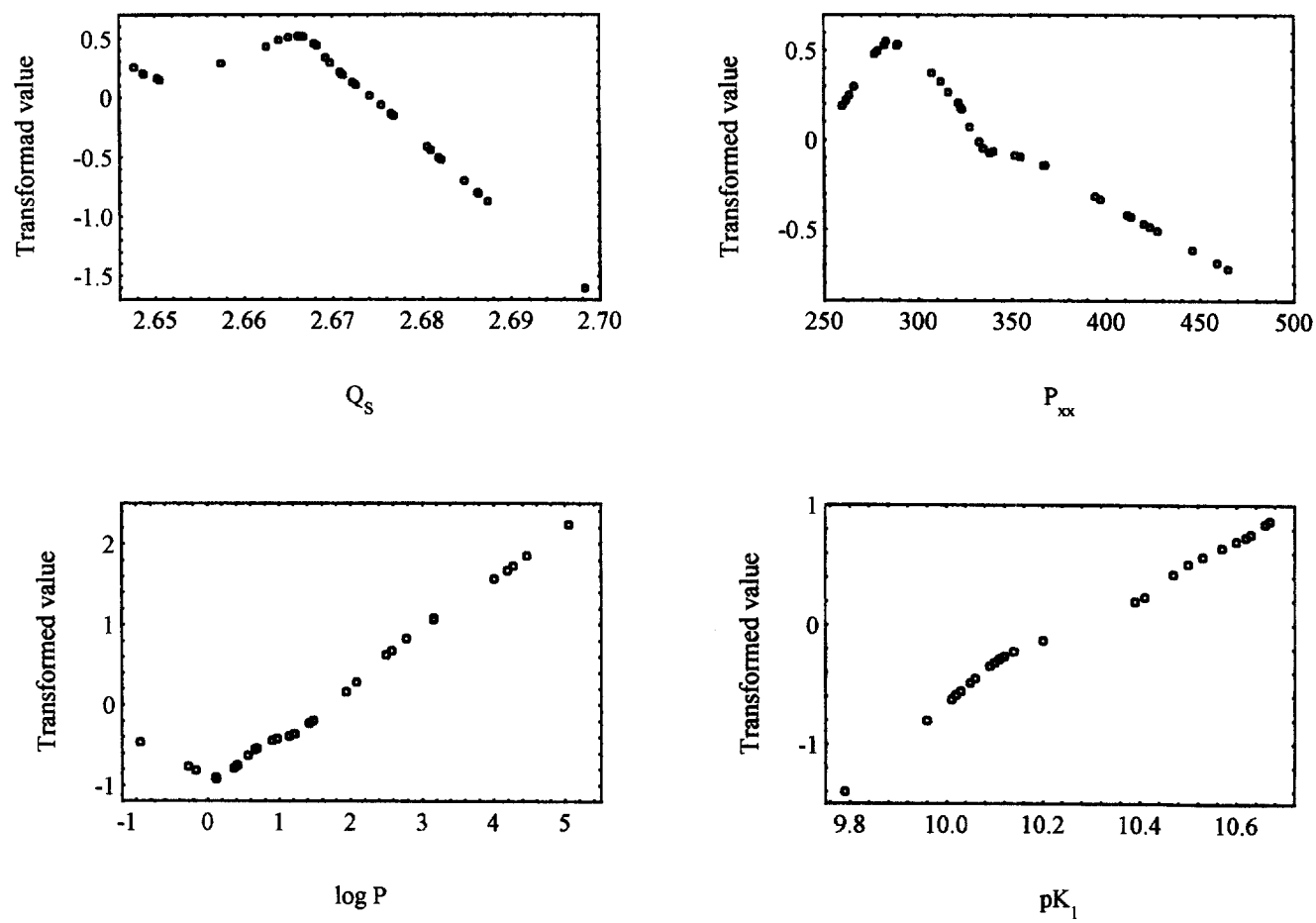



**Figure 5.** ACE transformation plot for variables of equation (4), leading to equation (5).

*4.1.4. CNDO*

Application of the all possible subsets MLR/PCA technique to the CA I CNDO results leads to equations (7) and (8).

$$\log K_i = C_1 D + C_2 E_L + C_3 D_1 + C_4 \log P + C_5 pK_1 + C_6 S_o^E + C_7 \quad (7)$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| C | −0.0671 | 0.2912 | 3.214 | 0.2982 | 1.050 | 144.44 | 56.7 |
| σ | 0.0257 | 0.2030 | 0.7381 | 0.0449 | 0.1731 | 28.60 | 13.8 |
| α | 0.01447 | 0.16287 | 0.00017 | 0.00000 | 0.00000 | 0.00003 | 0.00033 |

$N = 34$, $R^2 = 0.783$, $Q^2 = 0.540$, $F = 16.2$, $s = 0.215$, $\Lambda = 2.45$.

$$\log K_i = C_1 D + C_2 D_1 + C_3 \log P + C_4 pK_1 + C_5 S_o^E + C_6 \quad (8)$$

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| C | −0.0620 | 2.855 | 0.2936 | 0.9842 | 137.5 | 54.04 |
| σ | 0.0259 | 0.707 | 0.0456 | 0.1700 | 27.7 | 13.9 |
| α | 0.02369 | 0.00038 | 0.00000 | 0.00000 | 0.00005 | 0.00057 |

$N = 34$, $R^2 = 0.767$, $Q^2 = 0.578$, $F = 18.4$, $s = 0.219$, $\Lambda = 2.50$. This was the best $Q^2$ obtained for the CA I CNDO series.

As before, log P and $pK_1$ entered always with a positive coefficient. The dipole moment D now had a strong tendency to enter with a negative coefficient, and $D_1$, also with a negative coefficient.

In terms of $Q^2$, the effectiveness of the best equation for each series as a predictor was: Solution ESP > Solution > Vacuum > CNDO.

*4.2. CA II*

*4.2.1. Solution*

Application of the all possible subsets MLR/PCA technique to the CA II solution results leads to equations (9)–(12).

$$\log K_{ii} = C_1 Q_S + C_2 D_1 + C_3 I_p + C_4 (S_p^E - 0.5\, S_o^E) + C_5 \quad (9)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | 184.3 | 3.070 | 1.455 | −1506.6 | −538.0 |
| σ | 62.4 | 0.805 | 0.241 | 186.6 | 349.6 |
| α | 0.00609 | 0.00063 | 0.00000 | 0.0000 | 0.13436 |

$N = 35$, $R^2 = 0.783$, $Q^2 = 0.702$, $F = 30.7$, $s = 0.308$, $\Lambda = 2.82$. $K_{ii}$ is $10^8$ $K_I$ for CA II.

Analysis of this equation by the ACE technique led to the transformation plots of *figure 6*, and the equation

$$\log K_{ii} = C_1 Q_S + C_2 D_1{}^* + C_3 I_p + C_4 (S_p^E - 0.5\, S_o^E)^* + C_5 \quad (10)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | 156.6 | 3.599 | 1.390 | −1605.8 | −453.3 |
| σ | 55.9 | 0.788 | 0.206 | 188.1 | 313.2 |
| α | 0.00885 | 0.00008 | 0.00000 | 0.00000 | 0.15814 |

$N = 35$, $R^2 = 0.817$, $Q^2 = 0.745$, $F = 37.9$, $s = 0.283$, $\Lambda = 2.54$, where $D_1{}^* = D_1$ if $D_1 > 0.62$ and $D_1{}^* = 0.62$ otherwise, and $(S_p^E - 0.5\, S_o^E)^* = (S_p^E - 0.5\, S_o^E$ if $(S_p^E - 0.5\, S_o^E) > 0.01325$ and $(S_p^E - 0.5\, S_o^E)^* = 0.01325$ otherwise.

$$\log K_{ii} = C_1 D_1 + C_2 Q_H + C_3 I_p + C_4 (S_p^E - 0.5\, S_o^E) + C_5 \quad (11)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | 3.441 | −312.12 | 1.5494 | −1541.8 | 196.4 |
| σ | 0.850 | 104.52 | 0.2580 | 196.4 | 62.2 |
| α | 0.00033 | 0.00558 | 0.00000 | 0.00000 | 0.00362 |

$N = 35$, $R^2 = 0.788$, $Q^2 = 0.715$, $F = 28.2$, $s = 0.318$, $\Lambda = 2.98$.

Analysis of this equation by the ACE technique led to the transformation plots of *figure 7*, and the equation

$$\log K_2 = C_1 D_1{}^* + C_2 Q_H{}^2 + C_3 I_p + C_4 (S_p^E - 0.5\, S_o^E)^* + C_5 \quad (12)$$

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | 3.992 | −238.8 | 1.447 | −1617.8 | 96.7 |
| σ | 0.845 | 85.5 | 0.220 | 196.8 | 28.3 |
| α | 0.00005 | 0.00899 | 0.00000 | 0.00000 | 0.00185 |

$N = 35$, $R^2 = 0.821$, $Q^2 = 0.755$, $F = 34.2$, $s = 0.295$, $\Lambda = 2.63$, where $D_1{}^* = D_1$ if $D_1 > 0.62$ and $D_1{}^* = 0.62$ otherwise, and $(S_p^E - 0.5\, S_o^E)^* = (S_p^E - 0.5\, S_o^E)$ if $(S_p^E - 0.5\, S_o^E) > 0.01325$ and $(S_p^E - 0.5\, S_o^E)^* = 0.01325$ otherwise.

This equation gave the best $Q^2$ of the CA II solution series. In contrast to the CA I results, log P and $pK_1$ had little tendency to enter the equations. The strongest contenders here were $Q_S$, $D_1$ and $I_p$, all with positive coefficients. $Q_H$ could replace $Q_S$.

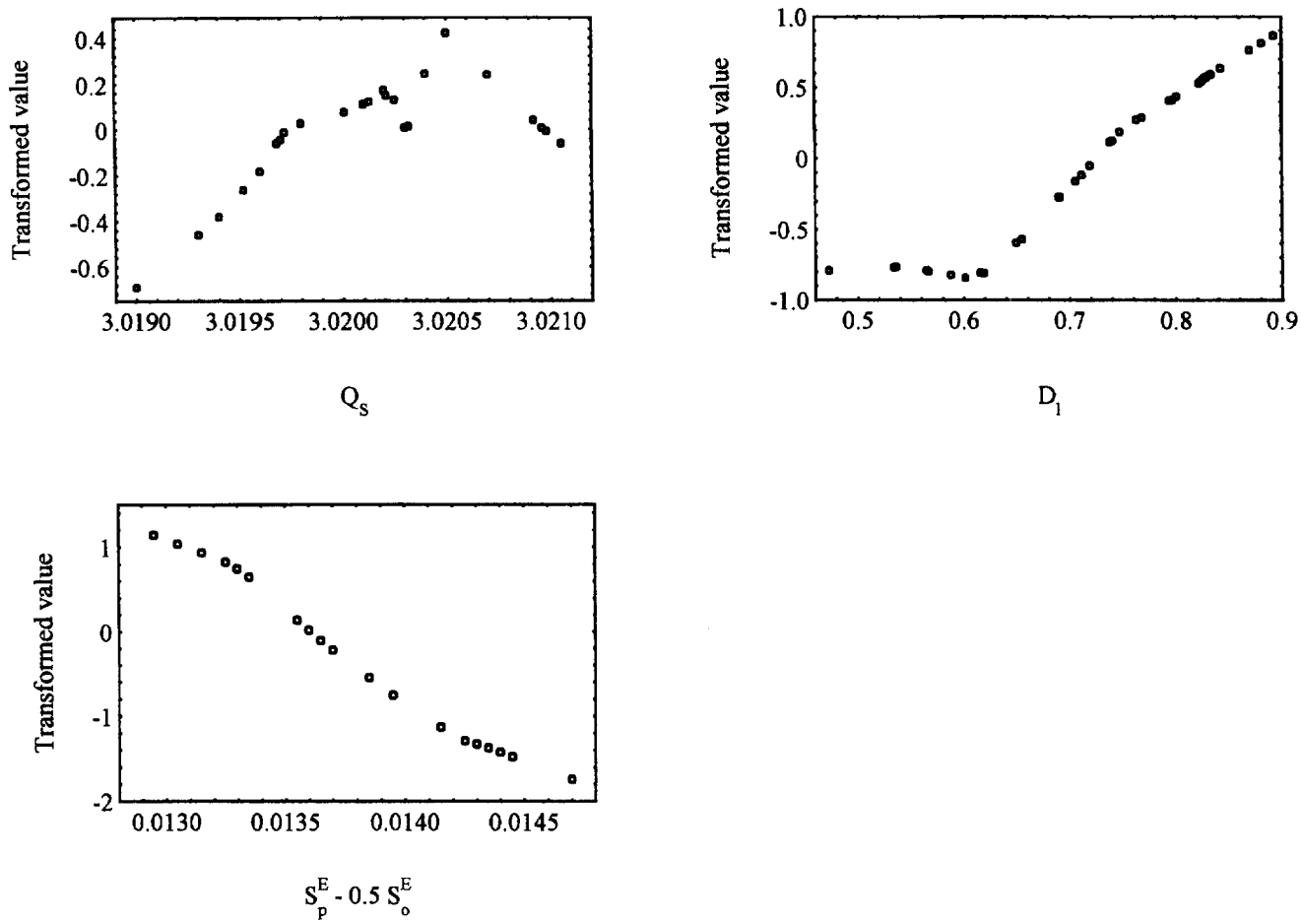



**Figure 6.** ACE transformation plot for variables of equation (9), leading to equation (10).

*4.2.2. Solution ESP*

Equations similar to (11) were obtained with an extra term, $D_1$, $P_{zz}$, or $P_{xx}$, and an equation with two extra terms, D and $D_1$. In each case, these extra terms had $\alpha > 0.05$. Analysis of one of these equations by the ACE technique led to the transformation plots of *figure 8* and the equation:

$$\log K_{ii} = C_1 D_x + C_2 Q_H + C_3 Q_H^2 + C_4 I_p + C_5 (S_p^E - 0.5\ S_m^E) + C_6 (Q_m + Q_p) + C_7 \quad (13)$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| C | -0.0651 | -9916.1 | 6114.2 | 1.1921 | -603.5 | -113.6 | 4007.0 |
| σ | 0.0212 | 3916.9 | 2407.7 | 0.2413 | 58.7 | 21.1 | 1592.5 |
| α | 0.00470 | 0.01725 | 0.01694 | 0.00003 | 0.00000 | 0.00001 | 0.0179 |

$N = 35$, $R^2 = 0.854$, $Q^2 = 0.794$, $F = 27.5$, $s = 0.274$, $\Lambda = 67423$.

The massive collinearity in this equation is entirely due to the correlation between $Q_H$ and $Q_H^2$. Because of the functional relationship between these two variables, the equation is not invalidated by the collinearity, but it did become necessary to carry out the calculation in double precision to avoid numerical instability. As may be seen from the α values, $Q_H$ and $Q_H^2$ are the weakest contenders for inclusion in the equation, but their inclusion does substantially improve the predictivity. This equation gave the best $Q^2$ for the CA II solution ESP results.

*4.2.3. Vacuum*

Application of the all possible subsets MLR/PCA technique to the CA II vacuum results leads to equations (14) and (15).

$$\log K_{ii} = C_1 Q_H + C_2 Q_p + C_3 Q_m + C_4 S_o^E + C_5 \quad (14)$$

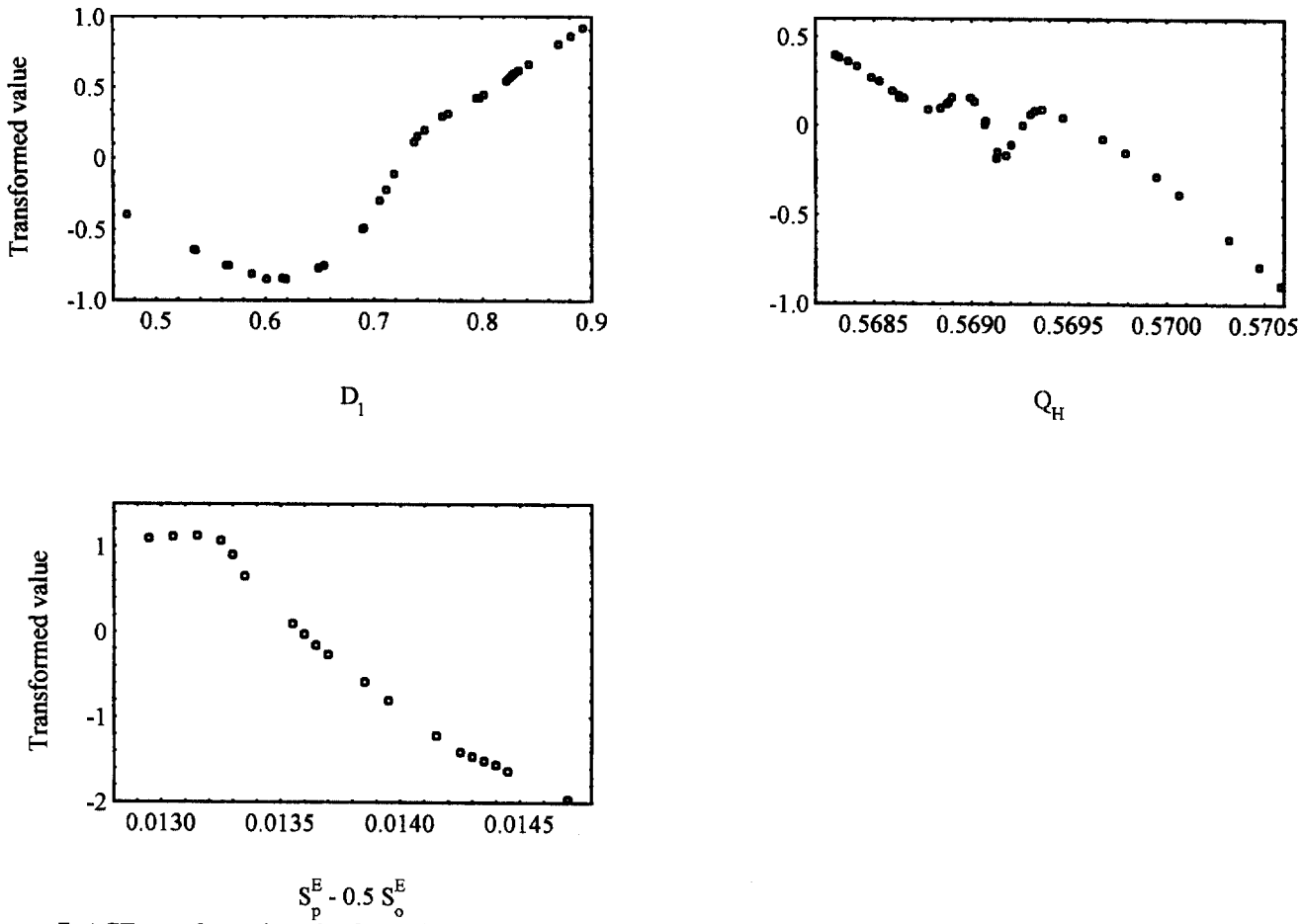



**Figure 7.** ACE transformation plot for variables of equation (11), leading to equation (12).

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| C | -179.3 | -112.1 | -37.7 | 185.6 | 169.0 |
| σ | 65.1 | 16.2 | 15.5 | 47.1 | 34.9 |
| α | 0.00983 | 0.00000 | 0.02088 | 0.00045 | 0.00004 |

$N = 35$, $R^2 = 0.786$, $Q^2 = 0.735$, $F = 27.1$, $s = 0.324$, $\Lambda = 3.15$.

$$\log K_{ii} = C_1 Q_C + C_2 Q_H + C_3 I_p + C_4 (Q_m + Q_p) + C_5 (E_L - E_H) + C_6 \quad (15)$$

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| C | 119.6 | -427.3 | 1.254 | -80.8 | 0.9839 | 275.6 |
| σ | 0.00000 | 0.00000 | 0.00009 | 0.00148 | 0.00034 | 0.00000 |
| α | 0.00000 | 0.00000 | 0.00009 | 0.00148 | 0.00034 | 0.00000 |

$N = 35$, $R^2 = 0.829$, $Q^2 = 0.774$, $F = 27.7$, $s = 0.294$, $\Lambda = 3.31$.

This was the best equation for CA II in the vacuum series. Again, log P and $pK_I$ showed little tendency to enter the equation, the dominant contenders being $Q_H$, and $Q_m$, $Q_p$ and their sum, always with negative coefficients. The hardness, $\eta = (E_L - E_H)/2$, with a positive coefficient, was a term in the best equation.

*4.2.4. CNDO*

Application of the all possible subsets MLR/PCA technique to the CA II CNDO results leads to equation (16).

$$\log K_{ii} = C_1 D_x + C_2 D_1 + C_3 Q_H + C_4 \log P + C_5\ pK_I + C_6 (S_p^E - 0.5\ S_m^E) + C_7 \quad (16)$$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| C | 0.1633 | 6.577 | -541.0 | 0.2439 | 0.9760 | -633.6 | 152.3 |
| σ | 0.0430 | 1.285 | 109.8 | 0.0721 | 0.2386 | 86.6 | 30.3 |
| α | 0.00076 | 0.00002 | 0.00004 | 0.00222 | 0.00035 | 0.00000 | 0.00003 |

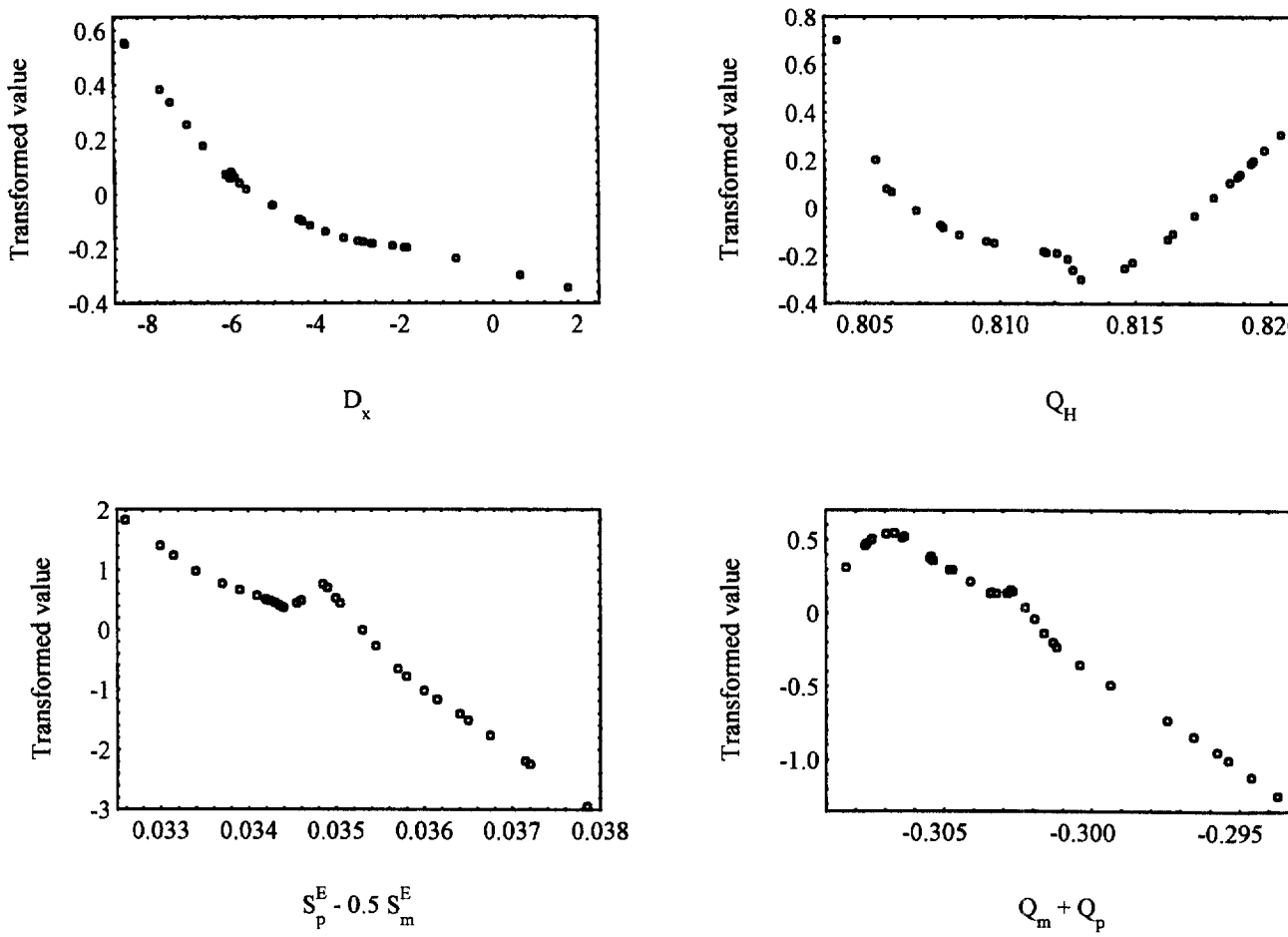



**Figure 8.** ACE transformation plots leading to equation (13).

N = 34, $R^2$ = 0.828, $Q^2$ = 0.725, F = 20.9, s = 0.308, $\Lambda$ = 4.75.

This equation gave the best $Q^2$ for the CA II CNDO series. It should be noted that log P has entered (with a positive coefficient), unlike any of the other CA II results. The CNDO series however gave the poorest predictivity (this equation) of the CA II results. The order of predictivity was: Solution ESP > Vacuum > Solution > CNDO.

## 5. Discussion

While the results support the expectation that the solution ESP calculation gives the best description of the charges on the sulfonamide group, the improvement in correlation over the other series is not sufficient to be definitive. In the description of the dependence on structure of activity of these compounds as inhibitors of both CA I and CA II the equations are consistent within themselves and with other studies, and indicate some differences between the actions of these inhibitors on CA I and CA II.

Wherever $Q_H$ occurs in equations (1)–(16), its coefficient is negative. The coefficient of $Q_S$ is negative in the case of CA I and positive in the case of CA II. The coefficient of $pK_1$ is positive wherever it occurs in equations (1)–(16). These findings are consistent with many other studies, which indicate that activity of sulfonamide CA inhibitors is strongly dependent on the charges on and the acidity of the $SO_2NH_2$ group [1–3, 31–34]. Our findings in a non-congeneric series of sulfonamide inhibitors of CA II, [35] were that $Q_N$, which might be expected to correlate negatively with

both $Q_S$ and $Q_H$, was implicated with a negative coefficient. The charges in that study were calculated with CNDO. Given that the correlations between the charges on the atoms of the $SO_2NH_2$ group vary greatly with the method of calculation, the dependence of activity on calculated charge is complex.

The coefficient of log $P$ is always positive in the equations for CA I inhibition, in contrast to the results of Kakeya et al. [36–39] and Hansch et al. [40] in studies of benzenesulfonamides [41]. It is also positive in Equation 16, the only CA II QSAR in which it appears. log $P$ in all of these studies covered a similar range.

Wherever $P_{xx}$ occurs in equations (1)–(16), its coefficient is negative, and $P_{yy}$ when it appears has a positive coefficient. This contrasts with our findings in a group of positively charged pyridinium thiadiazolesulfonamide CA I inhibitors [42], but agrees in the finding of anisotropy. The positive coefficient of $E_L - E_H$ in equation (15) is however consistent with this study.

The local dipole index $D_1$, an important contributor to activity against both isozymes, is a measure of separation of charge in the molecule. Its large numerical value is due in part to the high charges on the sulfonamide moiety, but it correlates very poorly ($R^2 = 0.30$) with the $D_1$ calculated for only this group, or with the individual bond dipoles. Thus the variation in $D_1$ measures variation in charge separation mainly in parts of the molecule other than the sulfonamide moiety. Its positive coefficient in all of the equations in which it appears, implies that separation of charge in these parts of the molecule reduces the activity of the drug.

The indicator variable $I_p$ had a strong tendency to enter equations for activity of CA II inhibitors, but not CA I. Its coefficient was positive. It is noteworthy that $I_p$ and log $P$ did not occur together in any equation. This would suggest that the missing fragment value in the *C*log $P$ calculations is rather small, and that $I_p$ appears for reasons other than its contribution to lipophilicity. Thus having a second phenyl on the Schiffs base nitrogen is detrimental to activity of the compounds as inhibitors of CA II, but not CA I.

The negative coefficient of the magnitude of the dipole moment in several of our equations contrasts with its positive coefficient in the non-congeneric series. It appears in equations for CA I inhibitory activity, but not that for CA II. The component $D_x$, however, appears in both, but with positive sign in CA I, and negative sign in CA II inhibition (equation (13), (16) is atypical). This may be related to the observation that $Q_S$ appears with negative coefficient in CA I inhibition, but positive in CA II inhibition (the S is strongly positively charged. There is more negative charge on the O and N bonded to the S than on the C, so the local dipole is in the $-x$ direction and increases with the charge on S).

The two isozymes behave oppositely to an increase in this component ($D_x$), which strongly suggests that this is due to differences in the active site architecture of the two isozymes, and as a consequence, differences in their interactions with the inhibitor molecule. Thus, CA I, generally possessing a 10–100 times lower affinity for sulfonamides as compared to CA II [43] differs primarily in two respects from CA II (from a structural point of view, not considering differences in catalytic power, susceptibility to inhibitors, etc, which are addressed in detail in [44]): (i) the presence of a larger number of His residues within the active site cavity; and (ii) the absence of a histidine cluster at its entrance [44]. Both these factors influence the interaction of these isozymes with the sulfonamide inhibitors, in ways which are probably reflected by the calculated parameters mentioned above.

Human CA I possesses four histidine residues within its active site, i.e., His 64, His 67, His 200 or His 243 [45], and at least two of them have an important role in catalysis, as His 200 stabilizes the E–S (and presumably also the E–I) complex [44, 46], whereas His 64 and probably also His 67 act as proton shuttling groups between the active site and the medium (generally the rate-limiting step in catalysis [43–45] involving CAs). In contrast to CA I, isozyme CA II possesses only one such residue, His 64, which acts as an extremely efficient proton shuttling group, assuring the very high catalytic efficiency of this isozyme [47]. Due to the polar character of these residues, they also contribute in a large extent to the hydrogen bond network within the active site cavity, and as a consequence, to the interaction of these enzymes with the inhibitors/substrates/activators [44]. According to the present finding, the positive sign of $D_x$ for CA I (and its negative sign for CA II) might reflect this type of discrimination of the two active sites for the different distribution of charges in the inhibitor molecule, which might be due to the presence of a different number of histidine residues within them. Mention should be made that the general shape as well as the sequence homology of other amino acid residues of the active sites of HCA I and HCA II are very similar (except for the characteristics mentioned above) [43, 44].

The second factor we mentioned, i.e., the presence of a histidine cluster at the entrance of the active site of CA II (and which is absent in CA I) might explain why a second phenyl on the Schiff base nitrogen is detrimental to activity for compounds acting as CA II inhibitors (since an interaction between the hydrophobic phenyl with the highly hydrophilic region at the entrance of the active site, consisting of His 3, His

4, His 10, His 15 and His 17 would not be favoured [44]), whereas the same group in CA I does not reduce activity, since CA I possesses much more hydrophobic residues in this region [44]. We assumed (based on X-ray crystallographic studies of CA–sulfonamide adducts [48, 49]) that the second phenyl of the Schiff base is oriented towards the outside of the active site, since obviously, the inhibitor should bind with the $SO_2NH^-$ moiety directly to the Zn(II) ion, as all sulfonamides investigated up to now [44, 48, 49].

The $pK_a$ of the sulfonamide group is very important for both CA I and CA II inhibition, and has much greater effect than the $pK_a$ of the Schiffs base N, as may be seen from their coefficients in equations (2)–(8) and (16). The influence of the Schiff base basicity on the activity of CA II appears opposite to that of the acidity of the sulfonamide, and is much weaker.

A large number of correlations with varied descriptors, obtained in many studies, have reported to date. The present results are not consistent with all of these, and the differences may be due in part to the calculation methods employed. Their internal consistency however implies that they are mainly due to differences between the classes of inhibitors, and between the CA isozymes considered. It is not possible to select any one descriptor as being dominant. This is prevented by the intercorrelations between the calculated descriptors, many of which are inherent in the chemistry, rather than being accidents of compound selection. It is not possible to point to the definite superiority of any one of the quantum chemical procedures used, on the basis of the results presented here, although the ESP-based charges did result in marginally better predictions. One should also note that the present is the first study in which isozyme-specific features have emerged from QSAR calculation, which might be useful for the design of isozyme-specific CA inhibitors.

## References


[1] De Benedetti P.G., Menziani M.C., Frassineti C., Quant. Struct. Act. Relat. 4 (1985) 23–28.

[2] De Benedetti P.G., Menziani M.C., Cocchi M., Frassineti C., Quant. Struct. Act. Relat. 6 (1987) 51–53.

[3] Menziani M.C., De Benedetti P.G., Struct. Chem. 3 (1992) 215–219.

[4] Menziani M.C., De Benedetti P.G., Gago F., Richards W.G., J. Med. Chem. 32 (1989) 951–956.

[5] Supuran C.T., Nicolae A., Popescu A., Eur. J. Med. Chem. 31 (1996) 431–438.

[6] Supuran C.T., Popescu A., Ilisiu M., Costandache A., Banciu M.D., Eur. J. Med. Chem. 31 (1996) 439–447.

[7] PCMODEL V. 5.0, Available from Serena Software, P.O. Box 3076, Bloomington, IN 47402–3076, USA.

[8] Stewart J.J.P., MOPAC 93.00, Fujitsu Ltd., Tokyo, Japan, 1993; also Stewart J.J.P., QCPE Bull. 15 (1995) 13–14 (Copyright Fujitsu 1993, all rights reserved).

[9] Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P., J. Amer. Chem. Soc. 107 (1985) 3902–3909.

[10] Klamt A., Shuurmann G., Perkin Trans. (1993) 799–805.

[11] Merz K.M., Besler B.H., J. Comp. Chem. 11 (1990) 431–439.

[12] Kurtz H.A., Stewart J.J.P., Dieter K.M., J. Comp. Chem. 11 (1990) 82–87.

[13] Clare B.W., Supuran C.T., J. Pharm. Sci. 83 (1994) 768–773.

[14] *ClogP* for Windows, version 1.0.0, available from BioByte Corp., 201 West 4th St., Suite 204, Claremont, CA 91711, USA.

[15] Pallas for Windows, version 3.1, $pK_a$ Prediction Module, pKalc 3.1, available from CompuDrug Chemistry Ltd., H-1395 Budapest, P.O. Box 62, 405 Hungary.

[16] Pacios L.F., QCPE Bull. 15 (1995) 16.

[17] Clare B.W., Supuran C.T., Theochem., 428 (1998) 109–121.

[18] Brown R.E., Simas A.M., Theor. Chim. Acta 62 (1982) 1–16.

[19] Kikuchi O., Quant. Struct. Act. Relat. 6 (1987) 179–184.

[20] Topliss J.G., Costello R.J., J. Med. Chem. 15 (1972) 1066–1068.

[21] Topliss J.G., Edwards R.J., J. Med. Chem. 22 (1979) 1238–1244.

[22] Clare B.W., in: Mannhold R., Krogsgaard-Larsen P., Timmerman H. (Eds.), Methods and Principles in Medicinal Chemistry, Vol. 3 in: Advanced Computer-Assisted Techniques in Drug Discovery (Van de Waterbeemd H., Ed.), VCH, Weinheim, 1994, pp. 284, 289.

[23] Clare B.W., Chemom. Intell. Lab. Syst. 18 (1993) 71–93.

[24] Livingstone D.L., Rahr E., Quant. Struct. Act. Relat. 8 (1989) 103–108.

[25] Furnival G.M., Wilson R.W., Technometrics 16 (1974) 499–511.

[26] BMDP Dynamic, Release 7, available from BMDP Statistical Software Inc., 1440 Sepulveda Boulevard, Suite 316, Los Angeles, CA 90025, USA.

[27] Daniel C., Wood F.S., Fitting Equations to Data, Wiley, New York, 1971.

[28] Chatterjee S., Price B., Regression Analysis by Example, 1st ed., Wiley, New York, 1977, pp. 199–200.

[29] File cai_47.zip at site ftp://central.murdoch.edu.au/pub/chem/clare.

[30] Breiman L., Friedman J.H., J. Amer. Statist. Assoc. 80 (1985) 580–619.

[31] Hansch C., McLarin J., Klein T., Langridge R., Mol. Pharmacol. 27 (1985) 493–498.

[32] Hansch C., Fujita T., J. Amer. Chem. Soc. 86 (1964) 1616–1623.

[33] Roblin R.O., Clapp J.W., J. Amer. Chem. Soc. 72 (1950) 4890–4892.

[34] Maren T.H., Conroy C.W., J. Biol. Chem. 268 (1993) 26233–26239.

[35] Clare B.W., Supuran C.T., Eur. J. Med. Chem. 32 (1997) 311–319.

[36] Kakeya N., Aoki M., Kamada A., Yata N., Chem. Pharm. Bull. 17 (1969) 1010–1018.

[37] Kakeya N., Yata D., Kamada A., Aoki M., Chem. Pharm. Bull. 17 (1969) 2000–2007.

[38] Kakeya N., Yata N., Kamada A., Aoki M., Chem. Pharm. Bull. 17 (1969) 2558–2564.

[39] Kakeya N., Yata N., Kamada A., Aoki M., Chem. Pharm. Bull. 18 (1970) 191–194.

[40] Hansch C., McLarin J., Klein T., Langridge R., Mol. Pharmacol. 27 (1985) 493–498.

[41] Maren T.H., Clare B.W., Supuran C.T., Roum.Chem. Quart. Rev. 4 (1994) 259–282.

[42] Supuran C.T., Clare B.W., Eur. J. Med. Chem. 30 (1995) 687–696.

[43] Supuran C.T., in: Puscas I. (Ed.), Carbonic Anhydrase and Modulation of Physiologic and Pathologic Processes in the Organism, Helicon, Timisoara, 1994, pp. 29–111.

[44] Briganti F., Mangani S., Orioli P. et al., Biochemistry, in press.

[45] Kumar V., Kannan K.K., J. Mol. Biol. 241 (1994) 226–232.

[46] Xue Y., Vidgren J., Svensson A. et al., Proteins 15 (1993) 80–87.

[47] Eriksson A.E., Jones T.A., Liljas A., Proteins 4 (1988) 274–282.

[48] Vidgren J., Liljas A., Walker N.P.C., Int. J. Biol. Macromol. 12 (1990) 342–344.

[49] Hakansson K., Liljas A., FEBS Lett. 350 (1994) 319–322.