Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 2197–2211
Bioorganic & Medicinal Chemistry

# A physically interpretable quantum-theoretic QSAR for some carbonic anhydrase inhibitors with diverse aromatic rings, obtained by a new QSAR procedure

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

[a]School of Biomedical and Chemical Science, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia
[b]Università degli Studi, Laboratorio di Chimica Inorganica e Bioinorganica, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Received 4 November 2004; revised 27 December 2004; accepted 27 December 2004
Available online 22 January 2005

**Abstract**—A QSAR based almost entirely on quantum theoretically calculated descriptors has been developed for a large and heterogeneous group of aromatic and heteroaromatic carbonic anhydrase inhibitors, using orbital energies, nodal angles, atomic charges, and some other intuitively appealing descriptors. Most calculations have been done at the B3LYP/6-31G* level of theory. For the first time we have treated five-membered rings by the same means that we have used for benzene rings in the past. Our flip regression technique has been expanded to encompass automatic variable selection. The statistical quality of the results, while not equal to those we have had with benzene derivatives, is very good considering the noncongeneric nature of the compounds. The most significant correlation was with charge on the atoms of the sulfonamide group, followed by the nodal orientation and the solvation energy calculated by COSMO and the charge polarization of the molecule calculated as the mean absolute Mulliken charge over all atoms.


## 1. Introduction

The carbonic anhydrases (CAs, EC 4.2.1.1) are ubiquitous metalloenzymes in all kingdoms, starting with *Archaea*, *Bacteria*, algae and green plants, and ending with superior animals, including vertebrates.[1–4] Their catalytic activity, the hydration of carbon dioxide to bicarbonate and a proton and their physiological function is essential for all these organisms, as the CA-catalyzed reaction is fundamental in physiology, whereas the three chemical entities participating in it are ubiquitous, abundant and chemically critical in a multitude of processes.[1–3] Thus, this reaction is fundamental for respiration and transport of $CO_2$ between metabolizing tissues and excretion sites, secretion of electrolytes in a variety of tissues and organs, pH regulation and homeostasis, $CO_2$ fixation (for algae and green plants), several metabolic biosynthetic pathways, such as gluconeogenesis, lipogenesis, and ureagenesis, bone resorption, calcification, tumorigenicity (in vertebrates), etc.[1–3] Inhibition (but also activation) of CAs may be exploited clinically in the treatment or prevention of a variety of disorders.[1–3] In consequence, CA inhibitors (CAIs) and to a less extent up to now, CA activators (CAAs) possess a variety of applications in therapy.[1–4] In higher vertebrates, including humans, 15 different CA isozymes or CA-related proteins (CARP) were described so far, with very different subcellular localization and tissue distribution.[1–3] Among the active enzymes, there are several cytosolic forms (CA I-III, CA VII, CA XIII), four membrane-bound isozymes (CA IV, CA IX, CA XII, and CA XIV), mitochondrial form (CA V), as well as a secreted CA isozyme (CA VI).[1,2] Many of these isozymes are important targets for the design of inhibitors with clinical applications, and indeed, recently this group described compounds with interesting applications as antiglaucoma, antiobesity, or antitumor agents among others.[5–7]

Recently we described a series of carbonic anhydrase inhibitors based on benzenesulfonamides and diverse five-membered aromatic rings with sulfonamide substituents and fluorinated tails,[8] which showed interesting

*Keywords*: QSAR; Quantum theoretic; Carbonic anhydrase; Regression.
* Corresponding author. Tel.: +61 8 6488 8562; fax: +61 8 6488 1005 (B.W.C.); fax: +39 055 4573385 (C.T.S.); e-mail addresses: bwc@theochem.uwa.edu.au; claudiu.supuran@unifi.it

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.12.055

applications as topically acting antiglaucoma agents in an animal model of the disease.[1] A QSAR for these compounds has been published that gave excellent statistical qualities, but was based partly on graph theoretic indices and used neural nets for data analysis.[9] These choices preclude any inferences of mechanism of inhibition, or indeed, of any theoretical interpretation whatsoever. The present contribution develops our previous work on benzenesulfonamides,[10] extending it to five-membered aromatic ring systems with heteroatoms. It is hoped that the model developed in this way will reflect the true nature of the complex intermolecular forces zbetween the drugs and their enzyme receptor, and thus suggest ways in which these interactions can be optimized.

## 2. Calculations

All calculations were run on a Dell Inspiron 3800 PC running Microsoft Windows Xp.

This is the first time our flip regression and orbital nodal orientation calculation methods have been used on five-membered aromatic rings, and this was achieved by the same procedures that we used for benzene rings. The center of the ring is placed at the origin, the atom of the ring to which the sulfonamide group is attached is numbered 1 and placed on the positive $X$ axis, and the remaining atoms numbered consecutively anticlockwise around the ring. The calculated angles are then the angles that the nodal planes in the HOPO and LUPO (highest occupied π-like orbital and lowest unoccupied π-like orbital) make with atom 1, measured at the origin. If the ring is numbered in the alternative direction the angle obtained is the negative of this value.

No attempt is made to include steric influences and localized hydrophobic effects, but only interactions that have been shown to be important in other series of drugs, particularly atomic charges and the energies and nodal orientations of π-like near-frontier orbitals.[11–13] Many of the compounds had two or three aromatic rings. In these cases only the ring bearing the sulfonamide moiety was considered.

We have for some time now been examining the influence of such factors on the activity of drugs, and also a method of dealing with symmetry in aromatic drugs.[14] This arises because when a planar aromatic molecule lies on a flat surface or within a flat pocket it can do so in two ways, as shown in Figure 1 for the molecule thiadiazole-2-sulfonamide. If the angle the highest occupied-orbital nodal surface of the aromatic ring makes with the sulfonamide group is $\Phi_H$, and this is to match up with a similar surface in the receptor, the aromatic group can rotate 180° around the aromatic-sulfonamide bond, and the angle then becomes $-\Phi_H$, and the ring will orient itself in whichever way minimizes its binding energy. Which of these options is appropriate in the drugs in a series is decided by considering all possible combinations of $\Phi_H$ and $-\Phi_H$ for all the drugs, and accepting that combination giving the best fit, as measured by the Fisher $F$ for the regression. The parameters used in the fitting are $\sin 2\Phi_H$ and $\cos 2\Phi_H$, and the use of trigonometric identities allows us to calculate, from the coefficients of these two terms, the optimum value of the angle $\Phi_H$. The factor 2 arises because there are two optimal orientations as $\Phi_H$ goes from 0° to 360°. For the lowest unoccupied π-orbital the functions are $\sin 4\Phi_L$ and $\cos 4\Phi_L$, because in this case there are two mutually perpendicular nodes, and thus four optimal orientations in a complete rotation. For each drug in the series the procedure gives a flip status, which may be −1 or +1, and a flip significance. A status of −1

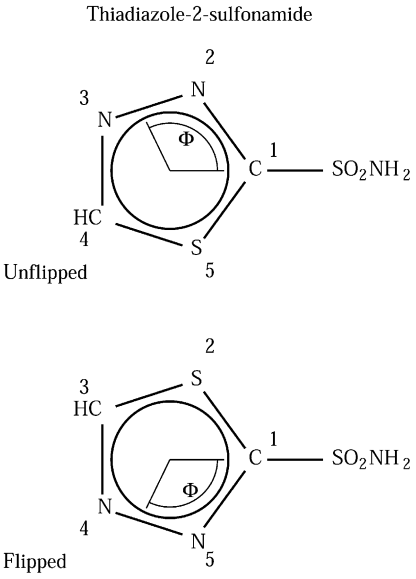



**Figure 1.** The molecule thiadiazole-2-sulfonamide, showing the ring numbering, the effect of flipping and the HOPO angle before and after flipping, with the HOPO angle from Scheme 2.

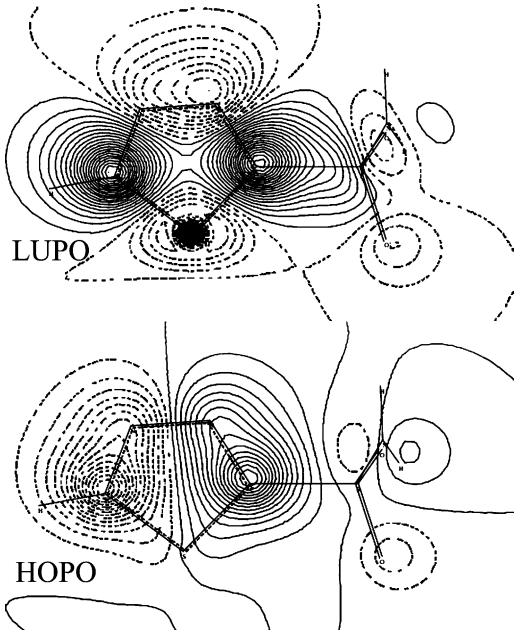
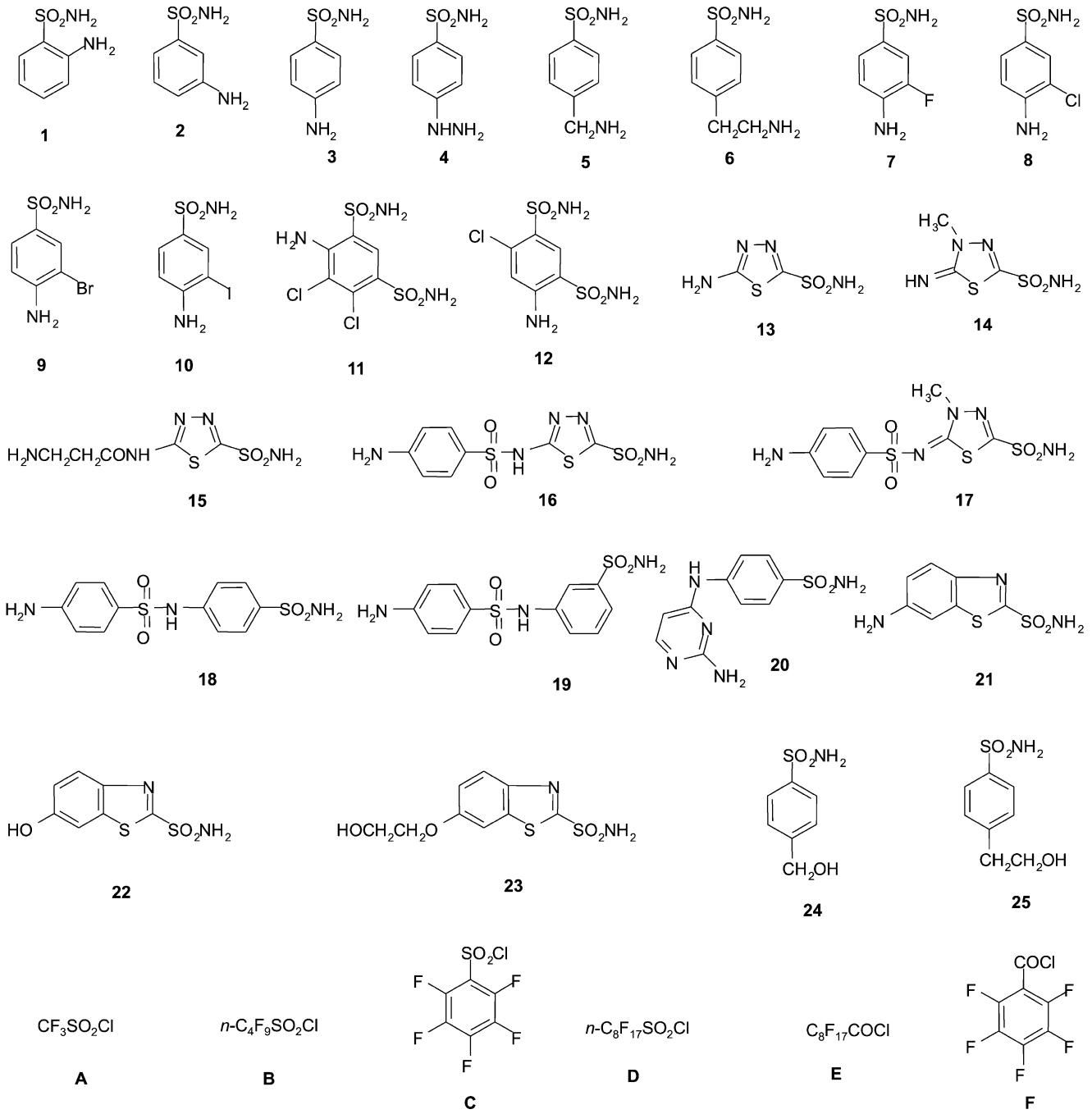



**Figure 2.** The nodes in the HOPO and LUPO of unflipped thiadiazole-2-sulfonamide, 0.5 Å above the ring plane, calculated at HF/6-31G*, showing the orientation of the nodes.

means that the drug has orientated the opposite way to that in which it was introduced, but bear in mind that there are always two equivalent solutions, differing in flip status by a factor of −1 (i.e., flipping all of the drugs changes nothing). This discussion has assumed a symmetric ring system such as benzene, but a similar argument applies to nonsymmetrical parent molecules. The HOPO and LUPO orbitals of the molecule, thiadiazole-2-sulfonamide, calculated at the HF/6-31G* level, are shown in Figure 2, and the output of a NODANGLE calculation for it in Scheme 2.

The flip significance is calculated by flipping each molecule in turn and applying a Student's *t*-test to the change in *F* on flipping. The flip procedure would probably be best not applied in a series where all drugs were based on the same, nonsymmetric ring system, where it could be assumed that the ring system itself had a preferred orientation on the site, although it could happen that the orientation of such a system could be dominated by the substituents rather than the ring components. In this case the flip procedure should lead to varied flip statuses and possibly nonsignificant flips. Here, however, we are




**Scheme 1.** The formulae of the compounds studied. (Reprinted with permission from *J. Med. Chem.* **2000**, *43*, 4542, copyright 2000 American Chemical Society).

```
Thiadiazole-2-sulfonamide
 5 membered ring starting at   1

 Local ring atom coordinates

 1 C  1     1.204     0.000     0.000
 2 N  2     0.403     1.118     0.000
 3 N  3    -0.855     0.848    -0.002
 4 C  4    -1.097    -0.496     0.001
 5 S  5     0.299    -1.415    -0.001

 Occupied orbitals, sums of squares on ring atoms

 HOMO-     9      8      7      6      5      4      3      2      1      0
 PZ   0.0043 0.0028 0.0023 0.0272 0.0456 0.0007 0.0052 0.8715 0.0282 0.8542
 PY   0.2935 0.0067 0.3425 0.0745 0.0224 0.2018 0.0080 0.0240 0.4674 0.0012
 PX   0.2478 0.0026 0.1153 0.0870 0.0193 0.2034 0.0318 0.0080 0.1556 0.0003
  S   0.3035 0.0119 0.1324 0.0627 0.0184 0.1179 0.0105 0.0196 0.2973 0.0015

 Virtual orbitals, sums of squares on ring atoms

 LUMO+     0      1      2      3      4      5      6      7      8      9
 PZ   0.8989 0.0001 0.8022 0.0029 0.0005 0.0029 0.0004 0.0004 0.0125 0.0020
 PY   0.0014 0.1794 0.0083 0.1419 0.0797 0.0359 0.2262 0.1187 0.0936 0.1094
 PX   0.0014 0.3790 0.0044 0.0275 0.0154 0.0666 0.1916 0.0459 0.0748 0.0477
  S   0.0014 0.3898 0.0194 0.3277 0.1218 0.0642 0.3910 0.4941 0.3333 0.2712

 ORB      SCALE  SSQ   FUNC        ANGLE         ENERGY
 HOMO-0    1.66  0.85  0.0026      107.98        -10.366
 HOMO-2    2.02  0.87  0.0851       17.62        -12.145

 ORB      SCALE  SSQ   FUNC        ANGLE         ENERGY
 LUMO+0    0.99  0.90  0.0201       60.02          1.437
 LUMO+2    1.00  0.80  0.0045       13.03          4.290
```

**Scheme 2.** The output of the program NODANGLE for thiadiazole-2-sulfonamide.

applying it to a multiplicity of ring types, and making no such assumption. A situation similar to this arises in classical QSAR of benzene derivatives. If we are seeking, for example, an equation relating activity to $\sum C_i\pi_i$ for a benzene ring where $C_i$ is a constant to be determined and $\pi_i$ is the lipophilicity of the $i$th substituent on the ring, we cannot, as has sometimes been done, treat the substituents independently. Otherwise, we would predict different activities, for example, 3-nitrophenol and 5-nitrophenol, which are the same substance. We handle this by, on flipping, exchanging the properties for the 3- and 5-position substituents, and also the 2- and 6-position. In the past, some authors have avoided the problem by their choice of substances to be studied, usually, for example, using only monosubstituted benzenesulfonamides.[15–18] Our programs allow this problem to be handled in such a way that any substituted benzene derivative can be treated correctly.

For the purpose of treating the present data some extensions have been made to the flip regression software. First, because the simulated annealing procedure does not always find the global minimum, provision has been made for FLIPANNL to automatically repeat the entire procedure a number of times and select the pattern with the highest $F$.

Secondly the model selection procedure has been automated. A series of annealings is carried out and the statistical significances and variance inflation factors for all the variables in the model are computed. If any variance inflation factor (VIF, defined as $1/(1-R_i^2)$, where $R_i$ is the multiple correlation coefficient of the $i$th independent variable regressed on all of the others.) is greater than a specified maximum VIFMAX, the variable or flippable pair of variables with the maximum VIF is removed, and the procedure restarted. This is done to avoid problems of colinearity. If no such variable exists, then if any of the variables have a statistical significance greater than a prescribed maximum SIGMAX the variable or flippable pair of variables with the maximum significance value is removed and the procedure is restarted. The procedure terminates when no variable remains with VIF greater than VIFMAX or significance greater than SIGMAX remains in the equation. Originally VIFMAX was set at 10.0 as statistics texts regard this as the level where colinearity becomes problematic,[19] and SIGMAX was set at the traditional value of 0.05. It soon became apparent that the VIFs obtained were all much less than VIFMAX, so it became advantageous to raise VIFMAX to 30 or 40. This usually did not raise the obtained VIFs beyond 10.0, and also usually resulted in better models. This algorithm is embodied in the program FLIPSTEP, which is part of the authors package MARTHA,[20] available on the Web. The procedure is equivalent to the classical backwards stepwise procedure. The VIF testing can be avoided by simply giving a very large value to VIFMAX.

It has been known for a long time now that a major influence on the activity of CA-inhibitory sulfonamides is the effect of the charge on the atoms of the sulfonamide moiety itself.[21] In addition to this and the orbital contributions we shall include lipophilicity, and some indicators of generalized nonbonded interactions, such as polarizability, solvation energy and mean absolute charge. Steric factors are not considered. We believe that this small group of descriptors will describe much, but

not all of the variation in activity in this group of drugs. To proceed further would require incorporation of our procedures into some form of 3D-QSAR program.

The ring systems **1**–**25** and tails **A**–**F** of Scheme 1 were built with Hyperchem[22] and geometry optimized with molecular mechanics, followed by AM1, to give the lowest energy conformers, except for ring systems **1**, **11**, and **12** with tails **A**, **B**, and **D**, where energy minimization led to a doubtful structure, in which an oxygen of the sulfonamide group appeared to be making a nucleophilic attack on the sulfur of the perfluoroalkylsulfonyl group, giving rise to a trigonal bipyramidal structure with an abnormally long $CF_2$–S bond. This structure could not be confirmed with a DFT optimization, although that calculation did indicate an abnormally flat PE surface in this neighborhood. In these nine cases a higher energy minimum was chosen with the perfluoroalkyl group pointing away from the benzenesulfonamide sulfur, which gave a structure more in keeping with usual experience of organic chemistry. Frequency calculations were run on the abnormal structures only, and were indicative of minima rather than transition states. To optimize all of the structures at the DFT level would have been prohibitively expensive, while it was felt that to use different optimization techniques for different

**Table 1.** Log $K_I$ (nm) of the 150 CA inhibitors against the three isozymes

| Compound | $K_I$CA1 | $K_I$CA2 | $K_I$CA4 | Compound | $K_I$CA1 | $K_I$CA2 | $K_I$CA4 | Compound | $K_I$CA1 | $K_I$CA2 | $K_I$CA4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **A01** | 5.243 | 4.311 | 4.987 | **C01** | 3.143 | 1.380 | 2.833 | **E01** | 3.255 | 2.013 | 3.176 |
| **A02** | 5.275 | 4.272 | 4.979 | **C02** | 3.041 | 1.000 | 2.699 | **E02** | 3.204 | 1.839 | 2.971 |
| **A03** | 5.219 | 4.037 | 4.810 | **C03** | 2.949 | 1.000 | 2.613 | **E03** | 3.146 | 1.820 | 2.929 |
| **A04** | 5.192 | 4.170 | 4.784 | **C04** | 3.000 | 1.204 | 2.699 | **E04** | 3.267 | 2.029 | 2.924 |
| **A05** | 5.139 | 3.699 | 4.389 | **C05** | 2.898 | 1.176 | 2.380 | **E05** | 3.000 | 1.792 | 2.875 |
| **A06** | 4.910 | 2.778 | 4.699 | **C06** | 2.881 | 1.176 | 2.362 | **E06** | 2.978 | 1.740 | 2.716 |
| **A07** | 5.000 | 2.699 | 4.462 | **C07** | 2.851 | 0.954 | 2.146 | **E07** | 3.146 | 1.708 | 2.740 |
| **A08** | 5.255 | 2.863 | 4.623 | **C08** | 3.161 | 2.041 | 2.732 | **E08** | 3.362 | 1.740 | 2.778 |
| **A09** | 5.301 | 3.017 | 4.672 | **C09** | 3.176 | 2.097 | 2.740 | **E09** | 3.398 | 1.839 | 2.792 |
| **A10** | 5.322 | 3.114 | 4.690 | **C10** | 3.244 | 2.176 | 2.829 | **E10** | 3.477 | 1.845 | 2.833 |
| **A11** | 3.740 | 2.633 | 2.860 | **C11** | 3.176 | 1.176 | 2.152 | **E11** | 3.845 | 1.857 | 2.919 |
| **A12** | 2.740 | 1.954 | 2.176 | **C12** | 2.477 | 0.699 | 1.415 | **E12** | 2.908 | 1.544 | 2.114 |
| **A13** | 4.777 | 2.000 | 2.778 | **C13** | 2.322 | −0.523 | 0.778 | **E13** | 2.477 | 0.699 | 1.114 |
| **A14** | 3.301 | 1.380 | 1.505 | **C14** | 2.398 | −0.523 | 0.954 | **E14** | 2.519 | 0.778 | 1.255 |
| **A15** | 3.477 | 1.114 | 1.602 | **C15** | 2.301 | −0.398 | 0.954 | **E15** | 2.477 | 0.301 | 0.903 |
| **A16** | 3.000 | 0.477 | 1.398 | **C16** | 1.380 | 0.000 | 0.903 | **E16** | 2.778 | 0.176 | 1.114 |
| **A17** | 3.000 | 0.699 | 1.519 | **C17** | 1.322 | 0.176 | 0.954 | **E17** | 2.954 | 0.301 | 1.079 |
| **A18** | 3.602 | 1.322 | 1.602 | **C18** | 2.097 | 0.903 | 1.176 | **E18** | 3.380 | 1.041 | 1.398 |
| **A19** | 3.699 | 1.362 | 1.653 | **C19** | 2.176 | 0.903 | 1.204 | **E19** | 3.477 | 1.204 | 1.477 |
| **A20** | 3.903 | 1.398 | 1.591 | **C20** | 2.512 | 1.041 | 1.380 | **E20** | 3.643 | 1.342 | 1.556 |
| **A21** | 1.845 | −0.046 | 0.903 | **C21** | 1.613 | −0.699 | 0.699 | **E21** | 1.653 | 0.000 | 0.954 |
| **A22** | 1.740 | −0.046 | 1.000 | **C22** | 1.398 | −0.523 | 0.778 | **E22** | 1.568 | 0.477 | 1.146 |
| **A23** | 1.699 | 0.000 | 0.845 | **C23** | 1.255 | −0.301 | 0.845 | **E23** | 1.380 | 0.301 | 0.845 |
| **A24** | 5.130 | 3.708 | 4.382 | **C24** | 2.462 | 1.602 | 2.000 | **E24** | 4.176 | 2.740 | 4.041 |
| **A25** | 4.903 | 2.740 | 4.653 | **C25** | 2.477 | 1.544 | 2.013 | **E25** | 4.301 | 2.613 | 4.079 |
| **B01** | 4.342 | 2.398 | 2.954 | **D01** | 3.176 | 1.982 | 3.079 | **F01** | 3.193 | 1.544 | 2.908 |
| **B02** | 4.265 | 2.230 | 2.771 | **D02** | 3.146 | 1.826 | 2.940 | **F02** | 3.146 | 1.279 | 2.813 |
| **B03** | 4.182 | 2.204 | 2.681 | **D03** | 3.057 | 1.792 | 2.881 | **F03** | 2.978 | 1.230 | 2.778 |
| **B04** | 4.176 | 2.255 | 2.778 | **D04** | 3.225 | 2.041 | 2.934 | **F04** | 3.176 | 1.362 | 2.991 |
| **B05** | 4.114 | 2.176 | 2.407 | **D05** | 3.000 | 1.778 | 2.848 | **F05** | 2.991 | 1.301 | 2.544 |
| **B06** | 4.097 | 2.176 | 2.380 | **D06** | 2.985 | 1.732 | 2.672 | **F06** | 2.954 | 1.230 | 2.491 |
| **B07** | 3.029 | 1.991 | 2.301 | **D07** | 3.161 | 2.097 | 2.699 | **F07** | 2.881 | 1.176 | 2.267 |
| **B08** | 4.009 | 2.628 | 2.944 | **D08** | 3.394 | 2.740 | 4.013 | **F08** | 3.033 | 2.097 | 2.771 |
| **B09** | 4.307 | 2.686 | 2.954 | **D09** | 3.407 | 2.833 | 4.097 | **F09** | 3.272 | 2.193 | 2.937 |
| **B10** | 4.369 | 2.813 | 3.017 | **D10** | 3.531 | 2.878 | 4.176 | **F10** | 3.408 | 2.362 | 3.017 |
| **B11** | 3.708 | 1.708 | 2.204 | **D11** | 3.903 | 1.991 | 2.884 | **F11** | 3.279 | 1.580 | 2.623 |
| **B12** | 2.602 | 0.903 | 1.633 | **D12** | 2.949 | 1.505 | 2.190 | **F12** | 2.875 | 1.079 | 1.732 |
| **B13** | 2.556 | 0.301 | 0.954 | **D13** | 2.602 | 0.477 | 1.079 | **F13** | 2.398 | 0.301 | 1.000 |
| **B14** | 2.544 | 0.301 | 1.000 | **D14** | 2.580 | 0.477 | 1.146 | **F14** | 2.431 | 0.176 | 0.903 |
| **B15** | 2.477 | 0.477 | 1.041 | **D15** | 2.477 | 0.301 | 0.954 | **F15** | 2.477 | 0.301 | 1.000 |
| **B16** | 2.954 | 0.301 | 1.230 | **D16** | 2.903 | 0.301 | 1.322 | **F16** | 3.176 | 0.903 | 1.505 |
| **B17** | 3.000 | 0.602 | 1.380 | **D17** | 3.041 | 0.477 | 1.398 | **F17** | 3.204 | 0.845 | 1.462 |
| **B18** | 3.556 | 1.176 | 1.491 | **D18** | 3.477 | 1.079 | 1.462 | **F18** | 3.820 | 1.255 | 1.591 |
| **B19** | 3.602 | 1.301 | 1.556 | **D19** | 3.556 | 1.322 | 1.519 | **F19** | 3.903 | 1.556 | 1.740 |
| **B20** | 3.778 | 1.301 | 1.477 | **D20** | 3.653 | 1.301 | 1.531 | **F20** | 3.886 | 1.431 | 1.792 |
| **B21** | 1.732 | −0.301 | 0.778 | **D21** | 1.690 | −0.222 | 0.845 | **F21** | 2.079 | −0.301 | 0.903 |
| **B22** | 1.462 | −0.301 | 0.845 | **D22** | 1.447 | −0.301 | 0.903 | **F22** | 2.021 | −0.222 | 0.954 |
| **B23** | 1.279 | 0.000 | 0.778 | **D23** | 1.230 | −0.222 | 0.699 | **F23** | 1.813 | −0.155 | 1.041 |
| **B24** | 4.389 | 2.663 | 4.023 | **D24** | 4.176 | 2.732 | 4.037 | **F24** | 3.079 | 1.732 | 2.613 |
| **B25** | 4.320 | 2.585 | 3.964 | **D25** | 4.342 | 2.643 | 4.090 | **F25** | 3.097 | 1.699 | 2.602 |

members of the series could have given misleading results.

Experimental data for CA inhibition by the entire group of compounds, taken from our previous publication[8] are given in Table 1. In all cases except rings **11** and **12**, which have two sulfonamide groups, the sulfonamide group was attached to atom 1 of the ring and the ring was numbered 1–5 or 1–6 anticlockwise, maintaining the orientation given in Scheme 1. Ring **12** was numbered anticlockwise starting at the top sulfonamide group, and ring **11** clockwise starting at the right hand side sulfonamide. This was done in order to make the relation between the sulfonamide and amine groups 1:4 in each case.

There is of course an ambiguity here: which of the two sulfonamide groups should we be comparing with the sulfonamide group of the other ring systems, both with regard to Mulliken charges and nodal orientation? The only rigorous answer is to consider every possible combination, but this would be prohibitively time consuming. It is hoped that the choice made is reasonable, and that if it proves wrong in any particular case, this will be apparent as an outlier.

The tails were attached to the ring systems in an extended conformation and the geometry of the complete molecule was fully optimized at the AM1[23] level with MOPAC, version 6.[24] Polarizabilities were taken from this calculation. A single point AM1 calculation was run on the resulting geometry with MOPAC 93,[25] using the COSMO[26] approach to obtain the solvation energy. Mean absolute atomic charges were calculated from the Mulliken charges for the MOPAC 93 result. A density functional calculation was then run on the same geometry using the B3LYP[27,28] functional, the COSMO option and the 6-31G* basis set, followed by a NODANGLE[29] calculation to obtain the orbital angles, energies, and Mulliken charges on the atoms of the $SO_2NH_2$ group. This calculation was done with GAUSSIAN 2003.[30] The charges cited for the sulfonamide O and H are the sum of the two values for these pairs of atoms. The orbital angles were relative to atom 1 of the ring and the orientations described above. The appropriate orbitals were identified from the NODANGLE output, and in doubtful cases, from a visualization of the orbitals with GAUSSVIEW.[31] Lipophilicities were calculated with Clog*P*.[32] All of the descriptors that were tried are listed in Table 2, and the values of the relevant AM1 and Clog*P* results are given in the Supplementary data, and the DFT results in Table 3.

The statistical calculations were done with the FLIPSTEP and MULTLR modules of MARTHA.[20] All variables were entered in the beginning of the calculation, and were eliminated one by one on the basis of the Student's $t$ value, until the statistical significance of all of the remaining variables was better than 0.05 and no variance inflation factor was greater than 10. Flipping was carried out at every stage of the stepwise regression. The final equation was evaluated with the multiple linear regression facility of the statistical package NCSS.[33] The flip status values tabulated are relative to the input orientation. It should be noted that the five-membered rings are not bilaterally symmetric relative to the sulfonamide group. No correspondence is assumed between the binding orientations of different, or even the same, heterocyclic ring, and it may be possible to infer such relationships from the resulting flip statuses and the input orientations of Scheme 1.

## 3. Results

Because the basis set 6–31G* does not cover iodine, no results were obtained for ring system **10**. Total lipophilicities were taken as the sum of the lipophilicities of the ring system and the tail. Clog*P* was unable to calculate the lipophilicity of rings **14** and **17** because of a

**Table 2.** All descriptors considered for the compounds

| Number | Name | Descriptor |
|---|---|---|
| 1 | $\Phi_H$ | Angle between node in highest occupied π orbital and $SO_2NH_2$ group, DFT (°) |
| 2 | $\Phi_L$ | Angle between node in lowest unoccupied π orbital and $SO_2NH_2$ group, DFT (°) |
| 3 | $E_H$ | Energy of highest occupied π orbital, (HOPO) DFT (eV) |
| 4 | $E_{SH}$ | Energy of second highest occupied π orbital, (SHOPO) DFT (eV) |
| 5 | $E_L$ | Energy of lowest unoccupied π orbital, (LUPO) DFT (eV) |
| 6 | $E_{SL}$ | Energy of second lowest unoccupied π orbital, (SLUPO) DFT (eV) |
| 7 | $Q_M$ | Mean absolute Mulliken charge over all atoms, AM1 |
| 8 | $Q_N$ | Mulliken charge on sulfonamide N, DFT |
| 9 | $Q_O$ | Mulliken charge on sulfonamide O, DFT |
| 10 | $Q_S$ | Mulliken charge on sulfonamide S, DFT |
| 11 | $Q_C$ | Mulliken charge on C attached to sulfonamide, DFT |
| 12 | $Q_H$ | Mulliken charge on sulfonamide H, DFT |
| 13 | $Q_T$ | Total Mulliken charge on sulfonamide group, DFT |
| 14 | $\Pi_{xx}$ | Component of polarizability, longest axis, AM1 (Å$^3$) |
| 15 | $\Pi_{yy}$ | Component of polarizability, intermediate axis, AM1 (Å$^3$) |
| 16 | $\Pi_{zz}$ | Component of polarizability, shortest axis, AM1 (Å$^3$) |
| 17 | $\Delta H_S$ | Difference between heats of formation with and without COSMO, AM1 (kcal) |
| 18 | $\pi_{tail}$ | Clog*P* value for tail |
| 19 | $\pi_{tot}$ | Sum of Clog*P* values for rings and tail |

**Table 3.** Relevant DFT-computed properties of the 144 drugs

| Compound | $\Phi_H$ (°) | $\Phi_L$ (°) | $E_H$ (eV) | $E_{SH}$ (eV) | $E_L$ (eV) | $E_{SL}$ (eV) | $Q_O$ (e) | $Q_N$ (e) | $Q_C$ (e) | $Q_H$ (e) |
|---|---|---|---|---|---|---|---|---|---|---|
| **A01** | 149.6 | 34.8 | −6.944 | −7.285 | −1.170 | −0.719 | −1.071 | −0.845 | −0.168 | 0.806 |
| **A02** | 88.9 | 55.0 | −7.188 | −7.450 | −1.156 | −0.820 | −1.073 | −0.845 | −0.132 | 0.826 |
| **A03** | **88.7** | **44.1** | −6.541 | −7.417 | −0.923 | −0.612 | −1.080 | −0.847 | −0.117 | 0.820 |
| **A04** | **89.7** | **45.4** | −6.377 | −7.248 | −0.769 | −0.425 | −1.084 | −0.848 | −0.120 | 0.818 |
| **A05** | **92.2** | **42.4** | −7.003 | −7.281 | −1.057 | −0.504 | −1.078 | −0.847 | −0.115 | 0.821 |
| **A06** | **91.4** | **43.0** | −6.789 | −7.166 | −0.878 | −0.367 | −1.083 | −0.848 | −0.117 | 0.818 |
| **A07** | 76.0 | 39.7 | −6.699 | −7.355 | −1.165 | −0.630 | −1.071 | −0.846 | −0.110 | 0.824 |
| **A08** | 75.0 | 43.9 | −6.777 | −7.316 | −1.284 | −0.853 | −1.071 | −0.845 | −0.105 | 0.826 |
| **A09** | 74.7 | 44.5 | −6.742 | −7.259 | −1.269 | −0.882 | −1.071 | −0.845 | −0.105 | 0.826 |
| **A11** | 71.3 | 27.1 | −7.181 | −7.270 | −1.981 | −1.512 | −1.036 | −0.846 | −0.099 | 0.842 |
| **A12** | 117.7 | 31.6 | −7.192 | −7.479 | −1.854 | −1.315 | −1.036 | −0.849 | −0.100 | 0.836 |
| **A13** | 112.8 | 59.8 | −6.973 | −9.247 | −2.002 | −0.315 | −1.020 | −0.822 | −0.172 | 0.849 |
| **A14** | 119.4 | 58.0 | −6.947 | −9.256 | −2.016 | −1.232 | −1.001 | −0.826 | −0.152 | 0.854 |
| **A15** | 111.8 | 60.5 | −6.597 | −8.716 | −1.864 | −0.546 | −1.035 | −0.828 | −0.189 | 0.839 |
| **A16** | 114.1 | 57.5 | −6.487 | −8.707 | −1.744 | 0.053 | −1.030 | −0.828 | −0.177 | 0.842 |
| **A17** | 118.5 | 54.1 | −6.413 | −8.407 | −1.754 | −1.245 | −1.013 | −0.830 | −0.153 | 0.847 |
| **A18** | **88.9** | **46.5** | −6.149 | −7.238 | −0.598 | −0.385 | −1.089 | −0.850 | −0.119 | 0.814 |
| **A19** | 31.0 | 43.0 | −6.282 | −7.038 | −0.893 | −0.351 | −1.081 | −0.848 | −0.123 | 0.821 |
| **A20** | **89.0** | **44.3** | −6.036 | −7.155 | −1.314 | −0.348 | −1.085 | −0.849 | −0.117 | 0.816 |
| **A21** | 113.8 | 51.5 | −6.488 | −6.920 | −1.812 | −0.613 | −1.025 | −0.826 | −0.147 | 0.844 |
| **A22** | 123.2 | 52.3 | −6.737 | −7.013 | −1.901 | −0.756 | −1.023 | −0.826 | −0.146 | 0.845 |
| **A23** | 114.3 | 49.1 | −6.147 | −6.833 | −1.654 | −0.434 | −1.030 | −0.828 | −0.149 | 0.840 |
| **A24** | **93.0** | **42.5** | −7.023 | −7.310 | −1.029 | −0.512 | −1.077 | −0.847 | −0.113 | 0.822 |
| **A25** | **91.2** | **43.7** | −6.838 | −7.187 | −0.931 | −0.384 | −1.082 | −0.849 | −0.117 | 0.818 |
| **B01** | 150.3 | 33.1 | −6.927 | −7.286 | −1.214 | −0.777 | −1.071 | −0.843 | −0.170 | 0.802 |
| **B02** | 36.1 | 50.7 | −6.803 | −7.202 | −1.067 | −0.776 | −1.074 | −0.845 | −0.126 | 0.825 |
| **B03** | **90.4** | **46.2** | −6.605 | −7.382 | −1.096 | −0.554 | −1.079 | −0.848 | −0.115 | 0.819 |
| **B04** | **90.4** | **44.2** | −6.406 | −7.243 | −0.793 | −0.420 | −1.082 | −0.848 | −0.119 | 0.817 |
| **B05** | **91.8** | **42.6** | −7.011 | −7.287 | −1.078 | −0.517 | −1.077 | −0.847 | −0.115 | 0.821 |
| **B06** | **88.3** | **47.3** | −6.797 | −7.160 | −0.867 | −0.372 | −1.083 | −0.849 | −0.118 | 0.818 |
| **B07** | 76.2 | 39.8 | −6.695 | −7.353 | −1.227 | −0.629 | −1.071 | −0.846 | −0.110 | 0.824 |
| **B08** | 75.5 | 43.3 | −6.752 | −7.299 | −1.252 | −0.829 | −1.071 | −0.846 | −0.105 | 0.825 |
| **B09** | 75.6 | 43.8 | −6.719 | −7.250 | −1.280 | −0.870 | −1.071 | −0.846 | −0.105 | 0.824 |
| **B11** | 71.9 | 27.9 | −7.167 | −7.266 | −2.000 | −1.522 | −1.035 | −0.846 | −0.099 | 0.842 |
| **B12** | 114.9 | 28.6 | −7.139 | −7.242 | −1.784 | −1.308 | −1.036 | −0.849 | −0.102 | 0.835 |
| **B13** | 112.8 | 58.6 | −6.840 | −9.260 | −1.888 | −0.300 | −1.021 | −0.823 | −0.176 | 0.848 |
| **B14** | 119.3 | 59.6 | −6.969 | −9.369 | −2.068 | −1.295 | −1.001 | −0.826 | −0.152 | 0.855 |
| **B15** | 111.8 | 60.8 | −6.616 | −9.101 | −1.894 | −0.615 | −1.034 | −0.827 | −0.188 | 0.839 |
| **B16** | 114.1 | 57.6 | −6.492 | −8.724 | −1.746 | 0.054 | −1.029 | −0.826 | −0.177 | 0.842 |
| **B17** | 118.5 | 54.1 | −6.424 | −8.836 | −1.756 | −1.269 | −1.012 | −0.830 | −0.154 | 0.846 |
| **B18** | **88.8** | **45.8** | −6.152 | −7.240 | −0.767 | −0.388 | −1.089 | −0.850 | −0.119 | 0.814 |
| **B19** | 31.0 | 43.1 | −6.292 | −7.038 | −0.896 | −0.362 | −1.081 | −0.848 | −0.123 | 0.821 |
| **B20** | **88.9** | **44.3** | −6.031 | −7.151 | −1.308 | −0.346 | −1.085 | −0.849 | −0.117 | 0.816 |
| **B21** | 113.4 | 51.4 | −6.491 | −6.910 | −1.790 | −0.596 | −1.025 | −0.825 | −0.147 | 0.844 |
| **B22** | 123.4 | 52.2 | −6.726 | −6.992 | −1.877 | −0.632 | −1.023 | −0.825 | −0.146 | 0.845 |
| **B23** | 114.0 | 49.1 | −6.121 | −6.824 | −1.644 | −0.417 | −1.030 | −0.828 | −0.149 | 0.840 |
| **B24** | **94.6** | **42.3** | −7.037 | −7.303 | −1.271 | −0.523 | −1.077 | −0.847 | −0.113 | 0.822 |
| **B25** | **91.3** | **43.5** | −6.833 | −7.182 | −0.937 | −0.379 | −1.082 | −0.849 | −0.117 | 0.818 |
| **C01** | 148.0 | 36.0 | −6.745 | −7.201 | −1.075 | −0.579 | −1.081 | −0.863 | −0.155 | 0.819 |
| **C02** | 33.4 | 42.3 | −6.554 | −7.136 | −0.956 | −0.547 | −1.077 | −0.848 | −0.123 | 0.821 |
| **C03** | **88.7** | **47.4** | −6.386 | −7.314 | −0.803 | −0.450 | −1.085 | −0.848 | −0.116 | 0.817 |
| **C04** | **87.8** | **43.8** | −6.141 | −7.210 | −0.737 | −0.428 | −1.087 | −0.848 | −0.120 | 0.816 |
| **C05** | **86.6** | **48.0** | −6.907 | −7.195 | −0.946 | −0.425 | −1.081 | −0.848 | −0.115 | 0.819 |
| **C06** | **87.8** | **46.8** | −6.749 | −7.149 | −0.867 | −0.350 | −1.084 | −0.849 | −0.117 | 0.817 |
| **C07** | 77.8 | 44.2 | −6.478 | −7.272 | −0.981 | −0.539 | −1.077 | −0.847 | −0.111 | 0.822 |
| **C08** | 78.6 | 48.8 | −6.525 | −7.225 | −1.034 | −0.742 | −1.077 | −0.847 | −0.106 | 0.822 |
| **C09** | 78.2 | 51.4 | −6.506 | −7.178 | −1.038 | −0.770 | −1.077 | −0.847 | −0.106 | 0.822 |
| **C11** | 86.2 | 30.8 | −6.847 | −7.237 | −1.898 | −1.473 | −1.039 | −0.848 | −0.096 | 0.838 |
| **C12** | 93.6 | 29.3 | −6.729 | −7.249 | −1.661 | −1.339 | −1.047 | −0.850 | −0.111 | 0.833 |
| **C13** | 113.1 | 58.7 | −6.736 | −9.238 | −1.933 | −0.171 | −1.011 | −0.826 | −0.177 | 0.845 |
| **C14** | 118.7 | 59.3 | −6.689 | −8.931 | −1.958 | −1.425 | −1.009 | −0.829 | −0.152 | 0.850 |
| **C15** | 111.8 | 60.7 | −6.600 | −8.709 | −1.879 | −0.599 | −1.035 | −0.828 | −0.188 | 0.838 |
| **C16** | 114.6 | 60.8 | −6.498 | −7.032 | −2.042 | −0.181 | −1.027 | −0.826 | −0.170 | 0.844 |
| **C17** | 118.0 | 54.9 | −6.855 | −8.395 | −1.775 | −1.167 | −1.014 | −0.830 | −0.153 | 0.847 |
| **C18** | **88.9** | **45.7** | −6.135 | −7.239 | −0.722 | −0.381 | −1.089 | −0.850 | −0.119 | 0.813 |

(*continued on next page*)

**Table 3** (*continued*)

| Compound | $\Phi_H$ (°) | $\Phi_L$ (°) | $E_H$ (eV) | $E_{SH}$ (eV) | $E_L$ (eV) | $E_{SL}$ (eV) | $Q_O$ (e) | $Q_N$ (e) | $Q_C$ (e) | $Q_H$ (e) |
|---|---|---|---|---|---|---|---|---|---|---|
| **C19** | 31.7 | 35.8 | −6.231 | −7.098 | −0.860 | −0.320 | −1.080 | −0.848 | −0.123 | 0.820 |
| **C20** | **88.7** | **44.7** | −5.957 | −7.079 | −1.237 | −0.315 | −1.087 | −0.849 | −0.117 | 0.815 |
| **C21** | 110.0 | 51.5 | −6.239 | −6.851 | −1.756 | −0.502 | −1.028 | −0.827 | −0.148 | 0.842 |
| **C22** | 117.4 | 50.9 | −6.549 | −6.899 | −1.765 | −0.567 | −1.027 | −0.827 | −0.146 | 0.842 |
| **C23** | 113.7 | 49.1 | −6.108 | −6.816 | −1.646 | −0.413 | −1.031 | −0.828 | −0.149 | 0.840 |
| **C24** | **85.6** | **47.4** | −6.904 | −7.218 | −0.990 | −0.429 | −1.081 | −0.848 | −0.113 | 0.818 |
| **C25** | **91.9** | **43.1** | −6.811 | −7.164 | −0.880 | −0.359 | −1.084 | −0.849 | −0.116 | 0.817 |
| **D01** | 150.9 | 37.7 | −6.929 | −7.268 | −1.165 | −0.664 | −1.074 | −0.847 | −0.171 | 0.829 |
| **D02** | 40.9 | 47.9 | −6.897 | −7.226 | −1.048 | −0.678 | −1.074 | −0.846 | −0.128 | 0.824 |
| **D03** | **89.6** | **43.9** | −6.605 | −7.381 | −1.112 | −0.556 | −1.079 | −0.848 | −0.114 | 0.819 |
| **D04** | **90.4** | **44.2** | −6.404 | −7.243 | −0.792 | −0.420 | −1.082 | −0.848 | −0.119 | 0.817 |
| **D05** | **91.7** | **42.6** | −7.016 | −7.292 | −1.085 | −0.526 | −1.077 | −0.847 | −0.115 | 0.822 |
| **D06** | **91.5** | **43.0** | −6.790 | −7.165 | −0.876 | −0.367 | −1.083 | −0.848 | −0.118 | 0.818 |
| **D07** | 76.2 | 39.9 | −6.698 | −7.353 | −1.248 | −0.631 | −1.071 | −0.846 | −0.110 | 0.824 |
| **D08** | 76.6 | 43.0 | −6.730 | −7.308 | −1.296 | −0.839 | −1.071 | −0.846 | −0.105 | 0.824 |
| **D09** | 75.7 | 43.8 | −6.719 | −7.252 | −1.301 | −0.874 | −1.071 | −0.846 | −0.105 | 0.824 |
| **D11** | 72.1 | 28.0 | −7.164 | −7.266 | −2.000 | −1.523 | −1.036 | −0.846 | −0.099 | 0.842 |
| **D12** | 114.7 | 28.6 | −7.137 | −7.422 | −1.783 | −1.308 | −1.036 | −0.849 | −0.102 | 0.835 |
| **D13** | 112.8 | 58.8 | −6.848 | −8.856 | −1.902 | −0.303 | −1.021 | −0.823 | −0.176 | 0.848 |
| **D14** | 119.3 | 59.1 | −6.950 | −9.036 | −2.045 | −1.277 | −1.001 | −0.826 | −0.152 | 0.854 |
| **D15** | 111.8 | 60.9 | −6.620 | −8.730 | −1.899 | −0.625 | −1.034 | −0.828 | −0.188 | 0.840 |
| **D16** | 114.1 | 57.6 | −6.494 | −8.711 | −1.748 | 0.051 | −1.029 | −0.826 | −0.177 | 0.842 |
| **D17** | 118.4 | 54.1 | −6.423 | −8.835 | −1.755 | −1.271 | −1.012 | −0.830 | −0.154 | 0.846 |
| **D18** | **89.1** | **46.5** | −6.151 | −7.240 | −0.766 | −0.388 | −1.089 | −0.850 | −0.119 | 0.814 |
| **D19** | 31.5 | 39.5 | −6.255 | −7.069 | −0.940 | −0.209 | −1.077 | −0.850 | −0.128 | 0.824 |
| **D20** | **89.1** | **44.3** | −6.026 | −7.152 | −1.312 | −0.346 | −1.085 | −0.849 | −0.117 | 0.816 |
| **D21** | 111.4 | 51.1 | −6.380 | −6.900 | −1.758 | −0.637 | −1.026 | −0.826 | −0.148 | 0.843 |
| **D22** | 123.1 | 52.3 | −6.722 | −6.991 | −1.878 | −0.644 | −1.023 | −0.825 | −0.146 | 0.845 |
| **D23** | 114.0 | 49.0 | −6.123 | −6.826 | −1.648 | −0.419 | −1.029 | −0.828 | −0.150 | 0.840 |
| **D24** | **93.4** | **43.0** | −6.970 | −7.263 | −1.119 | −0.471 | −1.078 | −0.848 | −0.115 | 0.821 |
| **D25** | **91.6** | **43.3** | −6.844 | −7.185 | −0.955 | −0.384 | −1.082 | −0.848 | −0.117 | 0.818 |
| **E01** | 151.2 | 25.4 | −6.726 | −7.188 | −1.580 | −0.847 | −1.077 | −0.841 | −0.188 | 0.827 |
| **E02** | 31.8 | 68.9 | −6.554 | −7.135 | −1.450 | −0.852 | −1.077 | −0.848 | −0.122 | 0.822 |
| **E03** | **87.8** | **45.1** | −6.498 | −7.316 | −1.611 | −0.460 | −1.082 | −0.848 | −0.113 | 0.818 |
| **E04** | **87.9** | **45.2** | −6.231 | −7.237 | −0.697 | −0.404 | −1.086 | −0.849 | −0.120 | 0.816 |
| **E05** | **94.5** | **42.5** | −6.873 | −7.202 | −1.101 | −0.405 | −1.081 | −0.848 | −0.116 | 0.819 |
| **E06** | **90.7** | **41.5** | −6.761 | −7.144 | −0.927 | −0.314 | −1.084 | −0.849 | −0.118 | 0.817 |
| **E07** | 74.7 | 42.9 | −6.641 | −7.317 | −1.732 | −0.566 | −1.074 | −0.846 | −0.108 | 0.824 |
| **E08** | 75.4 | 45.9 | −6.641 | −7.255 | −1.800 | −0.750 | −1.074 | −0.846 | −0.103 | 0.823 |
| **E09** | 74.0 | 46.0 | −6.674 | −7.219 | −1.807 | −0.802 | −1.073 | −0.846 | −0.102 | 0.824 |
| **E11** | 168.4 | 34.0 | −7.296 | −7.384 | −2.115 | −1.451 | −1.031 | −0.848 | −0.087 | 0.839 |
| **E12** | 107.0 | 35.3 | −7.050 | −7.396 | −2.108 | −1.330 | −1.040 | −0.850 | −0.100 | 0.834 |
| **E13** | 111.8 | 63.7 | −6.883 | −8.869 | −2.310 | −1.133 | −1.025 | −0.826 | −0.182 | 0.844 |
| **E14** | 117.0 | 66.6 | −6.761 | −8.933 | −2.266 | −1.421 | −1.008 | −0.829 | −0.171 | 0.850 |
| **E15** | 114.2 | 59.8 | −6.612 | −8.679 | −1.989 | −0.512 | −1.029 | −0.827 | −0.168 | 0.841 |
| **E16** | 114.3 | 61.2 | −6.729 | −8.950 | −2.097 | −0.227 | −1.025 | −0.828 | −0.168 | 0.844 |
| **E17** | 118.8 | 40.9 | −6.349 | −8.391 | −1.527 | −0.862 | −1.014 | −0.830 | −0.153 | 0.846 |
| **E18** | **88.7** | **45.9** | −6.153 | −7.241 | −0.749 | −0.390 | −1.089 | −0.850 | −0.118 | 0.814 |
| **E19** | 31.1 | 42.5 | −6.191 | −7.071 | −0.882 | −0.248 | −1.080 | −0.849 | −0.120 | 0.819 |
| **E20** | **88.7** | **44.5** | −5.995 | −7.238 | −1.236 | −0.314 | −1.087 | −0.849 | −0.117 | 0.815 |
| **E21** | 113.0 | 53.4 | −6.324 | −6.873 | −1.925 | −0.446 | −1.026 | −0.827 | −0.146 | 0.845 |
| **E22** | 126.0 | 54.0 | −6.732 | −6.971 | −2.003 | −0.608 | −1.024 | −0.826 | −0.145 | 0.844 |
| **E23** | 114.0 | 49.1 | −6.114 | −6.822 | −1.649 | −0.420 | −1.030 | −0.828 | −0.149 | 0.841 |
| **E24** | **86.2** | **47.1** | −6.890 | −7.242 | −0.959 | −0.427 | −1.081 | −0.848 | −0.114 | 0.819 |
| **E25** | **91.9** | **43.2** | −6.830 | −7.177 | −0.891 | −0.372 | −1.083 | −0.849 | −0.117 | 0.818 |
| **F01** | 152.5 | 43.2 | −6.546 | −7.143 | −0.954 | −0.508 | −1.081 | −0.847 | −0.193 | 0.804 |
| **F02** | 31.0 | 40.8 | −6.406 | −7.075 | −0.855 | −0.490 | −1.081 | −0.848 | −0.120 | 0.820 |
| **F03** | **87.9** | **47.1** | −6.346 | −7.253 | −0.748 | −0.382 | −1.086 | −0.849 | −0.114 | 0.816 |
| **F04** | **87.8** | **46.2** | −6.124 | −7.174 | −0.656 | −0.350 | −1.088 | −0.849 | −0.121 | 0.814 |
| **F05** | **87.2** | **47.4** | −6.731 | −7.196 | −0.868 | −0.398 | −1.084 | −0.848 | −0.117 | 0.818 |
| **F06** | **91.0** | **43.2** | −6.675 | −7.140 | −0.815 | −0.332 | −1.085 | −0.849 | −0.117 | 0.817 |
| **F07** | 76.8 | 43.7 | −6.439 | −7.226 | −1.751 | −0.469 | −1.078 | −0.847 | −0.109 | 0.821 |
| **F08** | 77.8 | 46.5 | −6.490 | −7.190 | −1.800 | −0.678 | −1.077 | −0.847 | −0.104 | 0.821 |
| **F09** | 77.4 | 46.8 | −6.468 | −7.136 | −1.800 | −0.712 | −1.077 | −0.847 | −0.104 | 0.821 |

**Table 3** (*continued*)

| Compound | $\Phi_H$ (°) | $\Phi_L$ (°) | $E_H$ (eV) | $E_{SH}$ (eV) | $E_L$ (eV) | $E_{SL}$ (eV) | $Q_O$ (e) | $Q_N$ (e) | $Q_C$ (e) | $Q_H$ (e) |
|---|---|---|---|---|---|---|---|---|---|---|
| **F11** | 77.9 | 36.0 | −7.005 | −7.135 | −2.041 | −1.477 | −1.036 | −0.849 | −0.095 | 0.836 |
| **F12** | 100.7 | 35.1 | −6.877 | −7.302 | −2.022 | −1.369 | −1.043 | −0.851 | −0.101 | 0.832 |
| **F13** | 111.9 | 64.0 | −6.672 | −8.765 | −2.204 | −1.407 | −1.032 | −0.828 | −0.187 | 0.840 |
| **F14** | 117.3 | 65.8 | −6.585 | −8.899 | −2.115 | −1.387 | −1.012 | −0.831 | −0.175 | 0.846 |
| **F15** | 111.8 | 60.5 | −6.589 | −8.700 | −1.876 | −0.574 | −1.036 | −0.829 | −0.188 | 0.838 |
| **F16** | 112.9 | 57.3 | −6.525 | −8.971 | −1.785 | −0.032 | −1.030 | −0.829 | −0.179 | 0.840 |
| **F17** | 117.9 | 58.0 | −6.690 | −8.358 | −1.854 | −1.034 | −1.015 | −0.831 | −0.153 | 0.845 |
| **F18** | **88.8** | **45.5** | −6.140 | −7.246 | −0.726 | −0.375 | −1.090 | −0.850 | −0.118 | 0.813 |
| **F19** | 30.5 | 42.6 | −6.228 | −7.058 | −0.883 | −0.304 | −1.082 | −0.848 | −0.121 | 0.820 |
| **F20** | **88.8** | **44.4** | −5.960 | −7.190 | −1.200 | −0.306 | −1.088 | −0.850 | −0.117 | 0.815 |
| **F21** | 110.7 | 53.3 | −6.165 | −6.805 | −1.894 | −0.431 | −1.029 | −0.827 | −0.147 | 0.841 |
| **F22** | 120.7 | 49.8 | −6.585 | −6.869 | −1.699 | −0.507 | −1.026 | −0.827 | −0.146 | 0.842 |
| **F23** | 113.6 | 49.0 | −6.091 | −6.812 | −1.643 | −0.407 | −1.031 | −0.832 | −0.149 | 0.840 |
| **F24** | **93.7** | **42.4** | −6.825 | −7.199 | −0.912 | −0.385 | −1.083 | −0.849 | −0.114 | 0.818 |
| **F25** | **88.1** | **46.9** | −6.797 | −7.155 | −0.858 | −0.350 | −1.084 | −0.849 | −0.117 | 0.817 |

Angles in bold are values for the optimized conformers that ideally should be 0° or 90° (HOPO) or 0° or 45° (LUPO), but are not, due to unsymmetric conformation.

missing fragment. Preliminary calculations indicated that the quality of the fit in the remaining compounds was no better using the total lipophilicity than just the tail component, so rings **14** and **17** were retained and only the tail lipophilicity was used, thus enabling us to use 144 of the 150 compounds.

Rings **3**–**6**, **18**, **20**, **24**, and **25** are symmetrically substituted benzenesulfonamides, and disregarding the symmetry of the rest of the molecule would be expected to have $\Phi_H$ of exactly 0°, 90°, or 180° and $\Phi_L$ of 0°, 45°, or 90°. Their divergence from these values gives some idea of the influence of conformation on the calculated values of the angles. In fact, only $\Phi_H$ values of about 90° and $\Phi_L$ values of about 45° were encountered in this data set, and over the 48 compounds with these rings the mean of $\Phi_H$ was 89.8° with standard deviation 2.1°, and the mean for $\Phi_L$ was 44.7° with standard deviation 1.8°. Calculations on molecules with full symmetry imposed were within 0.1° of the expected values. Thus the variations of the angles due to conformational effects and the effects of parts of the molecule remote from the ring are less than 5° in the worst cases.

It was found that $\Delta H_S$, the mean atomic charge, one or more of the orbital energies, both nodal angles and a sulfonamide atomic charge entered the equation for all three isozymes. For the latter, the charge of any of the S, O, N, or H atoms or the C to which the group was attached could be used. Both the charges of O and H could enter, but because of the high correlation between these, this was not a satisfactory equation on colinearity grounds, and was eliminated by the stepwise procedure. The sum of the charges on all the atoms of the sulfonamide group was also satisfactory as a descriptor, but this is not described in what follows.

On the whole, the results for all three isozymes were very similar. Because of the stochastic nature of the simulated annealing procedure successive runs do not give identical results. It was found that rather than a spread of results the same small group of models was obtained repeatedly. These will now be described in detail.

Because the flip procedure invalidates the formal statistical quality parameters, the results were checked by running the regressions 5000 times with the dependent variable randomized. The multiple correlation coefficients $R$ from the randomization tests were compared with those for the actual flip calculation using the Fisher $Z$ (or $v$) statistic,[34] $Z = \frac{1}{2}\ln\frac{1+R}{1-R}$. This statistic is approximately normally distributed, so giving us an approximate statistical significance for the flip regression procedure as a whole. These significances are given for each model in the models below.

For CA I two models were obtained:

$$\begin{aligned}\log K_I = &-0.153(1.9)\cos 2\Phi_H - 0.406(5.8)\sin 2\Phi_H \\ &- 1.465(7.3)\cos 4\Phi_L + 1.228(13.7)\sin 4\Phi_L \\ &+ 0.0214(4.2)\Delta H_S + 0.554(7.9)\pi_{tail} \\ &+ 8.035(10.9)Q_M - 84.95(19.4)Q_N \\ &- 0.0035(3.2)\Pi_{zz} - 0.471(6.4)E_{SH} \\ &- 76.36\end{aligned} \tag{1}$$

$N = 144$, $R^2 = 0.828$, $Q^2 = 0.800$, $s = 0.41$, $\alpha_{rand} = 3 \times 10^{-38}$, $\Phi_H$opt = 34.7°, $\Phi_L$opt = 80.0°, significant autocorrelations: Lag 16: $R = +0.26$, Lag 25: $R = -0.32$.

$$\begin{aligned}\log K_I = &-0.2208(2.7)\cos 2\Phi_H - 0.4274(5.6)\sin 2\Phi_H \\ &- 1.2627(5.3)\cos 4\Phi_L + 1.3262(12.7)\sin 4\Phi_L \\ &+ 0.0210(4.2)\Delta H_S + 0.5637(7.7)\pi_{tail} \\ &+ 0.5422(3.2)E_L + 7.9811(10.8)Q_M \\ &- 95.57(10.9)Q_N - 0.0040(3.6)\Pi_{zz} \\ &+ 26.67(2.9)Q_H - 0.421(5.7)E_{SH} - 106.10\end{aligned} \tag{2}$$

$N = 144$, $R^2 = 0.836$, $Q^2 = 0.802$, $s = 0.41$, $\alpha_{rand} = 1 \times 10^{-36}$, $\Phi_H$opt = 31.3°, $\Phi_L$opt = 78.4°, significant autocorrelations: Lag 16: $R = +0.24$, Lag 25: $R = -0.22$.

Here $N$ is the number of compounds in the regression, $R^2$ is the squared multiple correlation coefficient, $Q^2$ the

same, crossvalidated (i.e., based on the leave one out technique), $s$ is the standard error of estimate, $\alpha_{rand}$ is the statistical significance of the flip regression based on randomizing the dependent variable, $\Phi_H$opt and $\Phi_L$opt are the values of $\Phi_H$ and $\Phi_L$ that minimize $K_I$ (i.e., maximize the activity), calculated from the coefficients of the trigonometric terms, and the numbers in parentheses are the magnitude of the Student's $t$ values for each of the terms. A $t$ value greater than approximately 2 indicates a term that is significant with better than 95% confidence, and larger values are indicative of terms of greater importance in explaining log $K_I$.

For Model 2 the compounds with large positive residuals (greater than 0.6) were **A4**, **A5**, **A6**, **A8**, **A9**, **D12**, **D25**, **E25**, **F17**, **F19**, and **F20**. Those with large negative residuals were **A12**, **B7**, **B15**, **C1**, **C3**, **C7**, **C18**, **C25**, and **F3**. Figure 3 is a plot of log $K_I$ observed against that calculated from Model 2.

For CA II three models were obtained, having 10, 11, and 12 descriptors.

$$\begin{aligned}\log K_I = &-0.425(5.4)\cos 2\Phi_H - 0.880(12.4)\sin 2\Phi_H \\ &- 0.498(2.5)\cos 4\Phi_L - 0.256(3.1)\sin 4\Phi_L \\ &+ 0.030(7.1)\Delta H_S - 1.317(9.3)E_H \\ &+ 0.426(6.1)\pi_{tail} + 7.013(9.3)Q_M \\ &- 70.05(19.6)Q_H - 0.368(3.1)E_{SL} + 47.24\end{aligned} \tag{3}$$

$N = 144$, $R^2 = 0.869$, $Q^2 = 0.848$, $s = 0.40$, $\alpha_{rand} = 3 \times 10^{-55}$, $\Phi_H$opt = 147.9°, $\Phi_L$opt = 6.8°, significant autocorrelations Lag 16: $R = +0.31$, Lag 28: $R = -0.25$.

$$\begin{aligned}\log K_I = &-0.166(2.1)\cos 2\Phi_H + 0.810(11.6)\sin 2\Phi_H \\ &- 0.858(3.9)\cos 4\Phi_L - 0.150(2.0)\sin 4\Phi_L \\ &- 10.29(4.4)Q_C + 0.0354(8.6)\Delta H_S \\ &- 1.324(9.5)E_H + 0.467(6.8)\pi_{tail} \\ &+ 7.53(10.2)Q_M - 38.65(18.0)Q_O \\ &- 0.709(4.8)E_{SL} - 53.41\end{aligned} \tag{4}$$

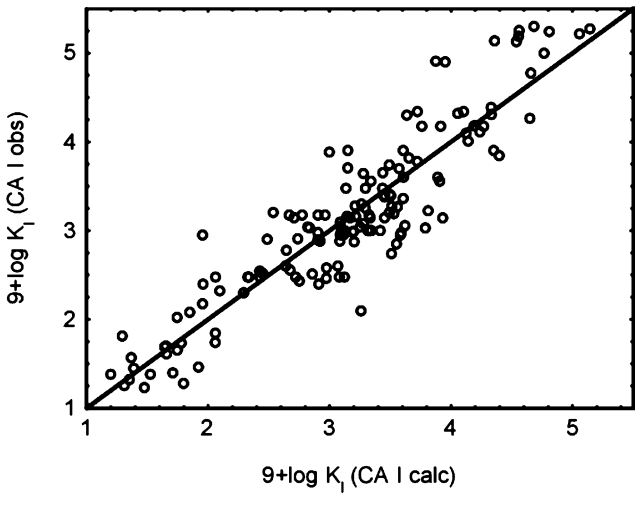



**Figure 3.** Plot of observed versus calculated log $K_I$ for CA I using Model 2.

$N = 144$, $R^2 = 0.876$, $Q^2 = 0.854$, $s = 0.39$, $\alpha_{rand} = 6 \times 10^{-58}$, $\Phi_H$opt = 140.8°, $\Phi_L$opt = 2.5°, significant autocorrelations Lag 16: $R = +0.31$, Lag 29: $R = -0.24$.

$$\begin{aligned}\log K_I = &\ 0.0847(1.1)\cos\Phi_H + 0.846(11.5)\sin 2\Phi_H \\ &- 1.289(5.9)\cos 4\Phi_L - 0.103(1.4)\sin 4\Phi_L \\ &+ 0.0316(6.0)\Delta H_S - 1.191(8.6)E_H \\ &+ 0.531(7.4)\pi_{tail} + 8.125(10.2)Q_M \\ &- 0.0032(3.0)\Pi_{zz} - 34.66(20.3)Q_O \\ &- 0.285(4.2)E_{SH} - 0.469(4.0)E_{SL} - 49.26\end{aligned} \tag{5}$$

$N = 144$, $R^2 = 0.878$, $Q^2 = 0.853$, $s = 0.39$, $\alpha_{rand} = 2 \times 10^{-55}$, $\Phi_H$opt = 132.1°, $\Phi_L$opt = 1.1°, significant autocorrelation Lag 16: $R = +0.22$.

Model 4 is the most satisfactory of any that was obtained. On the basis of Student's $t$ $Q_O$ is the most significant term, followed by $\sin 2\Phi_H$ and $Q_M$. From the coefficients of $\sin 2\Phi_H$ and $\cos 2\Phi_H$ we can calculate an optimal $\Phi_H$, and similarly for $\Phi_L$. The angles for the nodes of the LUPO are a relatively small contributor, along with $Q_C$ and $E_{SL}$. $E_H$, $\Delta H_S$, and $\pi_{tail}$ are moderate contributors. Much the same impression may be obtained pictorially from Figure 4, which shows the partial residuals for Model 4, and this also illustrates some other features of the data. The strongest contributor, $Q_O$, may be seen to be dichotomous. It breaks down into two clusters: the benzenoids and the heterocyclics. To a lesser extent, so does $Q_C$. The angle term $\cos 2\Phi_H$ is also polarized into a large cluster containing most of the molecules and a small one with positive values, containing ring systems **1**, **2**, and **19**, and $E_{SL}$ into a large group and a smaller one with very negative energies, containing ring systems **11**, **12**, **14**, and **17**. The compounds with large positive residuals were **A2**–**A5**, **A24**, **F16**, **F17**, and **F20**. Those with large negative residuals were **B13**, **C3**, **C5**, and **F13**. Figure 5 shows a plot of log $K_I$ observed against that calculated from Model 4.

For CA IV two models, were obtained: having 10 and 11 descriptors.

$$\begin{aligned}\log K_I = &\ 0.396(4.0)\cos 2\Phi_H - 1.091(11.6)\sin 2\Phi_H \\ &- 1.963(7.1)\cos 4\Phi_L + 0.277(3.1)\sin 4\Phi_L \\ &+ 0.0455(9.7)\Delta H_S - 1.409(8.8)E_H \\ &+ 0.279(3.5)\pi_{tail} + 0.548(3.6)E_L \\ &+ 5.72(6.7)Q_M - 33.50(5.7)Q_N - 36.38\end{aligned} \tag{6}$$

$N = 144$, $R^2 = 0.850$, $Q^2 = 0.826$, $s = 0.46$, $\alpha_{rand} = 6 \times 10^{-37}$, $\Phi_H$opt = 55.0°, $\Phi_L$opt = 88.0°, significant autocorrelations: Lag 6: $R = -0.38$, Lag 8: $R = +0.22$, Lag 12: $R = -0.29$, Lag 16: $R = +0.23$, Lag 20: $R = -0.27$.

$$\begin{aligned}\log K_I = &\ 0.414(4.3)\cos 2\Phi_H - 1.085(11.7)\sin 2\Phi_H \\ &- 2.053(7.5)\cos 4\Phi_L + 0.289(3.3)\sin 4\Phi_L \\ &+ 0.0381(6.7)\Delta H_S - 1.221(6.8)E_H \\ &+ 0.364(4.1)\pi_{tail} + 0.430(2.7)E_L \\ &+ 5.049(5.6)Q_M - 34.1(5.9)Q_N \\ &- 0.0025(2.2)\Pi_{xx} - 35.50\end{aligned} \tag{7}$$

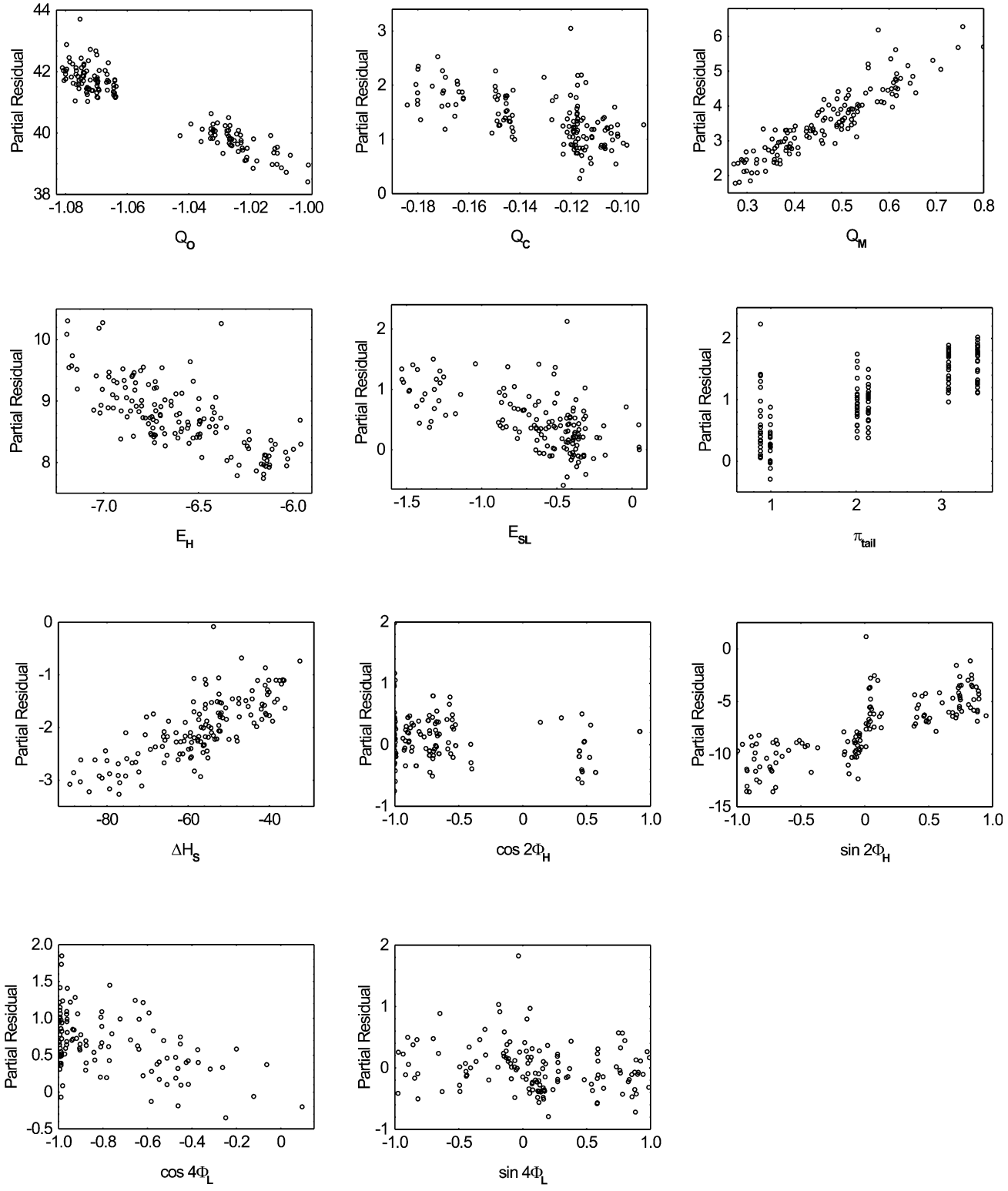



**Figure 4.** Plots of partial residuals for the variables of Model 4, showing the quality of the correlations.

$N = 144$, $R^2 = 0.855$, $Q^2 = 0.831$, $s = 0.45$, $\alpha_{rand} = 4 \times 10^{-37}$, $\Phi_H$opt = 55.4°, $\Phi_L$opt = 88.0°, significant autocorrelations Lag 4: $R = -0.37$, Lag 8: $R = +0.23$, Lag 12: $R = -0.29$, Lag 16: $R = +0.23$, Lag 20: $R = -0.27$.

Figure 6 shows a plot of log $K_I$ observed against that calculated from Model 7.

For Model 7 the terms with large positive residuals were **A2**–**A6**, **A25**, **D25**, and **E17**. The compounds with large negative residuals were **B3**, **B5**, **B6**, **B7**, **B13**, **B20**, **B22**, **C5**, **C24**, **C25**, and **D12**.

Flip statuses, flip significances and residuals for the seven models are reported in the Supplementary data.

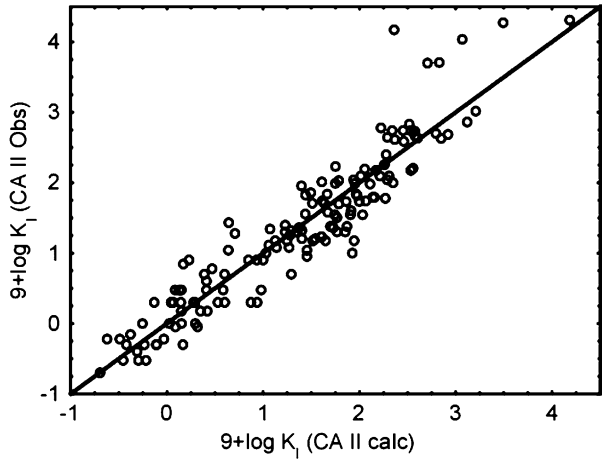



**Figure 5.** Plot of observed versus calculated log $K_I$ for CA II using Model 4.

### 3.1. Analysis of flip status and flip significance

In none of the seven models did any of the symmetrically substituted molecules reach a flip significance as small as 0.1. This is as would be expected. If the values for $\Phi_H$ and $\Phi_L$ were given their ideal values of 90° and 45°, respectively, rather than the calculated values, all of these compounds would have had a flip significance of exactly 1.

In Models 1 and 2, for the nonsymmetrically substituted compounds, 12 compounds differed in flip status, but of these only **F1** did not have both flip significances greater than 0.1. For Model 1 **F1** has flip significance 0.037 and for Model 2 0.054.

For Models 3 and 4, 11 nonsymmetrically substituted compounds differed in flip status. Of these, 7 had both significances less than 0.05. These were **B2**, **D1**, **D2**, **D12**, **E1**, **E12**, and **F2**. For Models 4 and 5, 14 compounds had discordant flip status. Of these seven had both significances less than 0.05. These were **A9**, **A14**, **A19**, **B19**, **D21**, **E1**, and **E12**.

For Models 6 and 7, no nonsymmetrically substituted compounds had differing flip statuses.

Thus among four pairs of models of the same CA isozyme having 96 nonsymmetric compounds there are 14 pairs of compounds differing in flip status with significance better than 0.05. This is rather less than would be expected by chance.

For Models 2 and 4, 45 out of 96 nonsymmetric compound differed in flip status. Of these 12 had both significances better than 0.05. These were **A1**, **A12**, **A14**, **B12**, **B15**, **C1**, **D1**, **D2**, **D13**, **E2**, **E17**, and **F1**. For Models 4 and 7, 33 out of 96 compounds differed in flip status. Of these 21 had both significances better than 0.05. For Models 2 and 7, 39 out of 96 nonsymmetric compounds had discordant flip status. Of these 16 had both significances better than 0.05. Among these three models 61 significant differences in status were observed. This is greatly in excess of expectation, and indicates that there are substantial differences in orientational preference for these drugs between the three isozymes. CA I most nearly resembles CA II and CA IV differs greatly from this pair. Because so many flip statuses have changed in these cases, and because the final flip statuses are indeterminate to a factor of ±1, it is not possible in these cases to compare individual flip statuses. Thus flip statuses are consistent within models for any one isozyme, and by comparing them with the input orientation conclusions may be drawn about relative orientational preference on the receptor, and also the differences between these preferences for different isozymes.

Because the data is structured with the six tails and the 24 rings in sequence, and the tails are not expected to contribute greatly to flip status, an autocorrelation trial on flip status was conducted. The compounds with symmetric rings were removed, leaving 16 rings, and autocorrelation runs were conducted with lags of up to 32, using NCSS. The results gave significant positive autocorrelations with lag 16 in all seven cases, noted under each model. Thus the nature of the tail is not dominant in determining flip status, but did effect some influence, or the autocorrelations would have been higher. In some cases strong negative and positive autocorrelations were also obtained with shorter lags, especially negative autocorrelation with lag 4, indicating that there was some unrecognized systematic variation with the sequence of the rings themselves, which were not intentionally placed in any particular order.

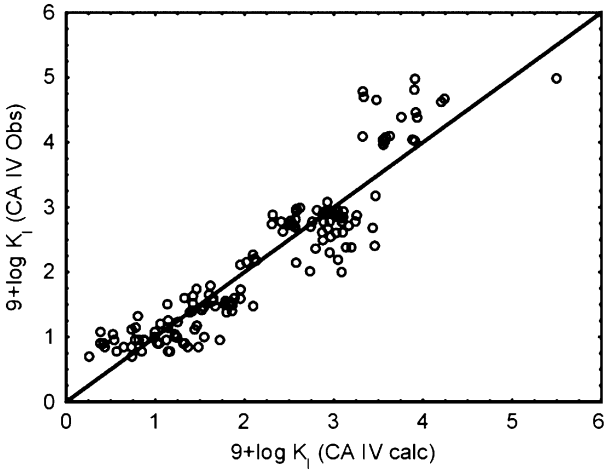



**Figure 6.** Plot of observed versus calculated log $K_I$ for CA IV using Model 7.

## 4. Discussion

The results of this study confirm the outcomes of numerous other studies: the main factor influencing the activity of sulfonamide inhibitors on all three of these isozymes is the charge on the sulfonamide group, either the total charge, or the charge on a particular atom. As shown in the dendrogram in Figure 7, the

charges of the sulfonamide atoms are strongly positively correlated.

The secondary factors differ. For CA I and CA II the second best descriptor is $\Phi_L$, while for CA IV it is $\Phi_H$. The mean atomic charge is important in all three, while $\Pi_{zz}$ is relatively unimportant. $\Delta H_S$ is important for CA IV, but less so for CA I and CA II. $E_H$ is important for CA I and $E_{SH}$ for CA II, but neither is very important for CA IV. Tail lipophilicity is moderately important for CA I and CA II but less so for CA IV. The three best models are of similar quality: $R^2$ is 0.83, 0.86, and 0.84, while $Q^2$ is 0.80, 0.84, and 0.83 for CA I, CA II, and CA IV, respectively. The activities of the drugs against the three isozymes are highly correlated, as may be seen from the dendrogram in Figure 8.

As may be seen from Models 1–7 there is reasonable agreement between models for the optimum angles for a given isozyme, but substantial difference between optimum angles for different isozymes. For CA I and CA IV the agreement between models for the same isozyme in excellent, and between the two there is a discordance of approximately 20° for $\Phi_H$ and 10° for $\Phi_L$, which is appreciable, though not very great. For CA II the optimum angles are very different: approximately 140° for $\Phi_H$ and close to 0° (or equivalently, 90°) for $\Phi_L$. Within the CA II models the agreement between the models is not as good as for the other two isozymes.

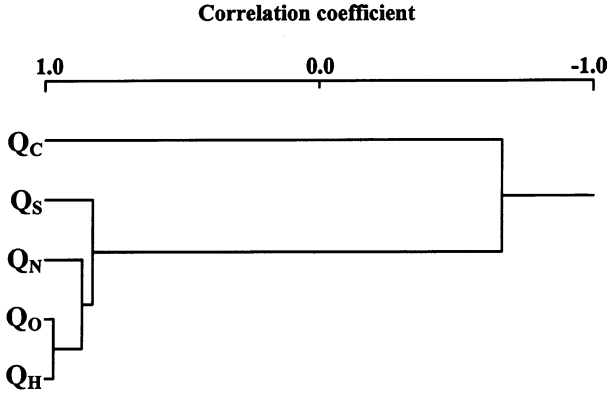



**Figure 7.** Dendrogram of correlations for the B3LYP/6-31G* Mulliken charges of the atoms of the sulfonamide group for the 144 compounds.

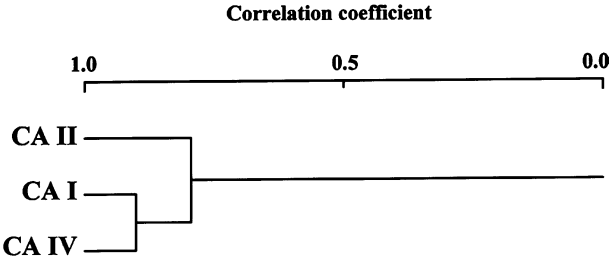



**Figure 8.** Dendrogram of correlations for $\log K_I$ for the three isozymes.

An idea of the conformational dependence of $\Phi_H$ and $\Phi_L$ may be obtained from Table 3. Those angles listed in bold are for structures with bilateral symmetry. If the conformation were such that the symmetry held perfectly, the $\Phi_H$ values would have to be either 0° or 90°, and for $\Phi_L$ 0° or 45°. In the actual conformer calculated, they differ from this by an average 1.8°, maximum 4.6° for $\Phi_H$ and on average 1.6°, maximum 3.5° for $\Phi_L$. This is about the reliability that may be attached to the computed values of the angular descriptors. Small differences in angle cannot be interpreted, owing to the idealized nature of the model used in their calculation.

The reason for the dependence of activity on the orbital angles remains to be explained. The most likely explanation is the interaction of π-like orbitals on the drug with similar orbitals on the receptor. Examination of the structure of crystallized CA reveals no aromatic residues close to the Zn atom to which the sulfonamide coordinates. Other than aromatic amino acids, the π-system that the drug interacts with could be on a guanidinium, carboxylate or amide group. Alternatively, the π-system may be binding the drug at a site remote from the active site, thus preventing it from functioning, influencing activity in this way.

The effect of orbital energy can be understood in terms of the same π-like orbitals on the drug, HOPO, SHOPO, LUPO, SLUPO, as are involved in the angular term interacting with similar orbitals on the receptor. Orbital interactions are strongest when the orbitals are of similar energy. When the orbital energies of the drugs bracket that of the receptor the interaction is a maximum and the linear dependence is small. When the orbital energies of the drug are far from that of the drug dependence is also small. Maximal dependence is obtained when the difference in energy is relatively small and unidirectional. The orbital energies for this set of compounds are only weakly correlated, as shown in the dendrogram in Figure 9. The three isozymes differ in which energy or energies are significantly effective in influencing activity.

The formal statistics of the fit are not nearly as satisfactory as was those of our previous results with thrombin and collagenase inhibitors,[10] that had a similarly wide range of activities, perhaps because in contrast to the present data set, the previous study used a set comprising only benzenoid derivatives, while the present set has also many and diverse five-membered aromatic ring systems. It is the usual experience in QSAR that it is much more difficult to fit a heterogeneous data set than a homogeneous one, and allowing for this, the results of the present study are more encouraging than is usually the case with diverse structures in Hansch-type analyses. There is only one QSAR known that is completely general. That is the QSAR for nonspecific, or baseline toxicity, that relates activity to lipophilicity, and applies to all substances that do not have any other activity that masks it. All other QSARs apply only to limited ranges of compounds. There are two ways of getting high $R^2$ values. The first is to severely limit the diversity of structures studied. This was done, for example, in Refs. 15–18 where attention was restricted to, for example,

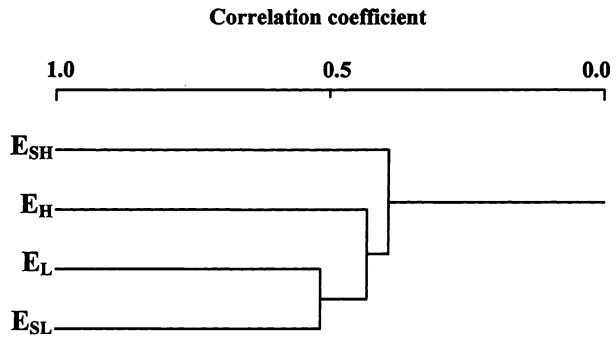



**Figure 9.** Dendrogram of correlations for the four B3LYP/6-31G* orbital energies for the 144 compounds.

*meta* and *para* monosubstituted benzenesulfonamides. The other approach is to use more varied compounds but more descriptors. Activity depends on many factors, and satisfactory results will be obtained only if all relevant factors are included. We have used more descriptors than most workers do, but still not a large number. Those we have chosen are ones that have been useful in a number of studies on aromatic compounds, particularly the nodal orientations. In our earlier work on benzenesulfonamides with a complicated substitution pattern[10] we obtained $R^2$ values of better than 0.98. The present study includes a diverse range of aromatic rings, and the only parameters used to describe them are the orientations of the nodes and the four orbital energies. Our equations can only ever be an approximation, as examples exist of aliphatic sulfonamides with CA inhibitory activity.

The additional descriptors can be easily understood. The mean absolute atomic charge is a measure of localized charge on a molecule that will favorably interact with charges and polarizabilities on neighboring atoms of the receptor. The polarizability of the drug will similarly interact with local charges on the receptor. The solvation energy is a measure of both enthalpic and entropic differences in solvation of the drugs. The tail lipophilicity is a similar indicator of the environment of the tail in the drug–receptor complex.

A number of sources of systematic errors have been pointed out, including the ambiguity of the treatment of the disulfonamides and the wide diversity of the ring systems. Study of the residuals throws little light on any such influences. There seems to be no association of large positive or negative residuals with any particular ring system consistent throughout the six tails. Compounds **A3**–**A6**, the symmetrically substituted benzenesulfonamides, typically gave large positive residuals in all models, but this was not duplicated with the other tails. Compound **D12**, a disulfonamide, gave large residuals in Models 1–3, 6, and 7, and this may have been related to an inappropriate choice of sulfonamide group. Similarly, **A12** gave large residuals in Models 1 and 2, **B11** in Model 1 and **C12** in Model 3.

Obviously there are factors influencing the activity of these drugs that are not addressed by the descriptors that we have employed. These may include steric and local hydrophobic effects. The accuracy of measurements is much better than the standard error of estimate in Models 1–7 would indicate. However, because of the very high significance of the randomization results, and also the large number of drugs studied, it is probable that a large part of the electronic portion of the variance in activity is covered by the factors included in the present contribution.

## Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2004.12.055. Four tables, giving AM1 descriptor values, and the flip significances, flip statuses, and residuals for all seven models.

## References and notes

1. *Carbonic Anhydrase—Its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC (Taylor and Francis Group): Boca Raton, Florida, 2004; pp 1–363, and references cited therein.
2. (a) Supuran, C. T.; Scozzafava, A. *Expert Opin. Ther. Pat.* **2000**, *10*, 575; (b) Supuran, C. T.; Scozzafava, A. *Exp. Opin. Ther. Pat.* **2002**, *12*, 217; (c) Supuran, C. T.; Scozzafava, A.; Casini, A. *Med. Res. Rev.* **2003**, *23*, 146; (d) Scozzafava, A.; Mastrolorenzo, A.; Supuran, C. T. *Exp. Opin. Ther. Pat.* **2004**, *14*, 667.
3. (a) Pastorekova, S.; Parkkila, S.; Pastorek, J.; Supuran, C. T. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 199; (b) Supuran, C. T.; Vullo, D.; Manole, G.; Casini, A.; Scozzafava, A. *Curr. Med. Chem.—Cardiovasc. Hematol. Agents* **2004**, *2*, 49.
4. Supuran, C. T.; Casini, A.; Scozzafava, A. Development of Sulfonamide Carbonic Anhydrase Inhibitors (CAIs). In *Carbonic Anhydrase–Its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC: Boca Raton FL, USA, 2004; pp 67–147.
5. Mincione, F.; Menabuoni, L.; Supuran, C. T. Clinical Applications of Carbonic Anhydrase Inhibitors in Ophthalmology. In *Carbonic Anhydrase—Its Inhibitors and Activators*; Supuran, C. T., Scozzafava, A., Conway, J., Eds.; CRC: Boca Raton, FL, 2004; pp 243–254.
6. Vullo, D.; Franchi, M.; Gallori, E.; Antel, J.; Scozzafava, A.; Supuran, C. T. *J. Med. Chem.* **2004**, *47*, 1272.
7. Švastova, E.; Hulikova, A.; Rafajova, M.; Zat'ovicova, M.; Gibadulinova, A.; Casini, A.; Cecchi, A.; Scozzafava, A.; Supuran, C. T.; Pastorek, J.; Pastorekova, S. *FEBS Lett.* **2004**, *577*, 439.
8. Scozzafava, A.; Menabuoni, L.; Mincione, F.; Briganti, F.; Mincione, G.; Supuran, C. T. *J. Med. Chem.* **2000**, *43*, 4542.
9. Mattioni, B. E.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 94.
10. Supuran, C. T.; Clare, B. W. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 237.
11. Clare, B. W. *J. Med. Chem.* **1998**, *41*, 3845.
12. Clare, B. W. *Theochem* **2000**, *507*, 157.
13. Clare, B. W. *Theochem* **2001**, *535*, 301.
14. Clare, B. W. *J. Comput.—Aided Mol. Des.* **2002**, *16*, 611.
15. Kakeya, N.; Aoki, M.; Kamada, A.; Yata, N. *Chem. Pharm. Bull.* **1969**, *17*, 1010.

16. Hansch, C.; McLarin, J.; Klein, T.; Langridge, R. *Mol. Pharmacol.* **1985**, *27*, 493.
17. Carotti, A.; Raguseo, C.; Campagna, F.; Langridge, R.; Klein, T. E. *QSAR* **1989**, *8*, 1.
18. DeBenedetti, P. G.; Menziani, M. C.; Frassineti, C. *QSAR* **1985**, *4*, 23.
19. Freund, R. J.; Wilson, W. J. In *Regression Analysis: Statistical Modeling of a Response Variable*; Academic: Newyork, 1998; pp 192–201.
20. File martha.zip on site http://www.chem.biomed-chem.uwa.edu.au/staff/homepages/BrianClare.
21. Maren, T. H.; Clare, B. W.; Supuran, C. T. *Roum. Chem. Quart. Rev.* **1994**, *4*, 259.
22. *Hyperchem v. 6.0*, Available from Hypercube, Inc. 1115 NW 4th Street, Gainesville, FL 32601-4256, USA.
23. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
24. Stewart, J. J. P. *QCPE Bull.* **1990**, *10*, 86.
25. Stewart, J. J. P. MOPAC 93.00 **1993**, Fujitsu Ltd, Tokyo, Japan, also Stewart, J. J. P. MOPAC93 Release 2. *QCPE Bull.* **1995**, *15*, 13–14 (Copyright Fujitsu 1993, all rights reserved).
26. Barone, V.; Cossi, M. *J. Phys. Chem.* **1998**, *A102*, 1995.
27. Becke, A. D. *Phys. Rev.* **1988**, *A38*, 3098.
28. Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B37*, 785.
29. File nodangle.zip on site http://www.chem.biomed-chem.uwa.edu.au/staff/homepages/BrianClare.
30. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision B.04*; Gaussian, Inc.: Pittsburgh PA, 2003.
31. Gaussview, www.gaussian.com.
32. *ClogP for Windows, ver 1.0.0*. Available from BioByte Corp. 201 West 4th Street, Suite204, Claremont CA 91711, USA.
33. Hintze, J. *NCSS* (2001) Number Cruncher Statistical Systems, Kaysville, Utah, www.ncss.com.
34. Affi, A. A.; Azen, S. P. In *Statistical Analysis: A Computer Oriented Approach*, 2nd ed.; Academic: New York, 1979; p 140.