Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 13 (2005) 4597–4621
Bioorganic & Medicinal Chemistry

Review

# An approach toward the problem of outliers in QSAR

Rajeshwar P. Verma* and Corwin Hansch

*Department of Chemistry, Pomona College, Claremont, CA 91711, USA*

Received 18 March 2005; revised 3 May 2005; accepted 3 May 2005
Available online 13 June 2005

**Abstract**—Compounds that have unexpected biological activity and are unable to fit in a QSAR model are known as outliers. These are valuable in defining the limitations under which compounds act by a common molecular mechanism modeled by one or more descriptors, and also in defining the experimental limitations of the biological test data. Thus, the outliers should be submitted to particular attention to see if the reason for their peculiarity can be determined. Separating these outliers from the main data set and formulating another QSAR can resolve the problem. Our result shows that these outliers may be acting by a different mechanism or interacting with the receptor in different modes.


**Contents**



*Keywords*: Hydrophobicity; Molar refractivity; Molar volume; Outlier; Structure–activity relationships.
* Corresponding author. Tel.: +1 909 607 4249; fax: +1 909 607 7726; e-mail: rverma@pomona.edu

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.05.002

## 1. Introduction

In the past 43 years, the use of QSAR, since its advent,[1] has become increasingly helpful in understanding chemical–biological interactions in the drug-design process and in pesticide research, as well as in the areas of toxicology.[2] It is useful in elucidating the mechanisms of chemical–biological interaction in various biological structures, particularly enzymes, membranes, organelles, and cells.[2,3] It has also been utilized for the evaluation of absorption, distribution, metabolism, and excretion (ADME) phenomena in many organisms and whole animal studies.[4] The QSAR approach employs extra-thermodynamically derived and computational-based descriptors to correlate biological activity in isolated receptors, cellular systems, and in vivo. QSAR models can stand alone, augment other graphical approaches, or be examined in tandem with equations of a similar mechanistic genre to establish their authenticity and reliability.[5]

The typical problem in developing a QSAR model is the presence of outliers. Outliers are those compounds which have unexpected biological activity and are unable to fit in a QSAR model. These are valuable in defining the limitations under which compounds act by a common molecular mechanism modeled by one or more descriptors, and also in defining the experimental limitations of the biological test data. Thus, the outliers should be submitted to particular attention to see if the reason for their peculiarity can be determined. QSAR models that have no or very few outliers are known as good models. On the other hand, those QSAR models that have a large number of outliers are bad models. The analyses of outliers indicate that these compounds may be acting by a different mechanism or interacting with the receptor in different modes.

In the present review, we would like to present QSAR (quantitative structure–activity relationship) studies on different sets of compounds which have a large number of outliers. Separating outliers from the main data set and formulating another QSAR can resolve the problem. The new subsets of compounds (outliers), which may be acting by a different mechanism or interacting with the receptor in different modes, have been supported by the different nature of their QSAR. The splitting of the main data set for the development of meaningful QSAR is shown in Figure 1. A search from SciFinder Scholar (2004 Edition) of the Chemical Abstract for the last five years reveals that the average number of publications on QSAR is about 1080 articles/reviews each year. Presently, more than 12,350 articles/reviews have already been published on QSAR; surprisingly, not a single attempt was made toward the problem of outliers.

## 2. Materials and methods

All the data have been collected from the literature (see individual QSAR for respective references). $C$ is the molar concentration of a compound and $\log 1/C$ is the dependent variable that defines the biological parameter for QSAR equations. Physicochemical descriptors are autoloaded, and multi-regression analyses (MRA) used to derive the QSAR were executed with the C-QSAR program.[6] The parameters used in this review have already been discussed in detail along with their application.[2] Briefly, $C\log P$ is a calculated partition coefficient in octanol/water and is a measure of hydrophobicity, and $\pi$ is the hydrophobic parameter for the substituents usually measured for substituents attached to benzene. Calculated molar refractivity (CMR) is the calculated molar refractivity for the whole molecule. MR is calculated from the Lorentz–Lorenz equation and is described as follows: $(n^2 - 1/n^2 + 2)(\mathrm{MW}/\delta)$, where $n$ is the refractive index, MW is the molecular weight, and $\delta$ is the density of a molecule. MR is dependent on the volume and polarizability. It can be used for a substituent or for the whole molecule. Number of valence electrons (NVE) is a parameter[7–10] that was found to be another approach to understanding polarizability and is calculated by simply summing up the valence electrons in a molecule, that is, H = 1, C = 4, Si = 4, N = 5, P = 5, O = 6, S = 6, and halogens = 7. MgVol is the molar volume for the whole molecule and is calculated by C-QSAR program using the method of McGowan.[11] Hammett $\sigma$, $\sigma^-$, and $\sigma^+$ constants are electronic parameters that apply to substituent effects on aromatic systems.[2] $B1$, $B5$, and $L$ are Verloop's sterimol parameters for substituents.[12] $B1$ is a measure of the width of the first atom of a substituent, $B5$ is an attempt to define the width of the whole substituent, and $L$ is the substituent length. The indicator variable $I$ is assigned the value of 1 or 0 for special features with special effects that cannot be parametrized and has been explained wherever used. Each regression equation includes 95% confidence limits for each term in parentheses.

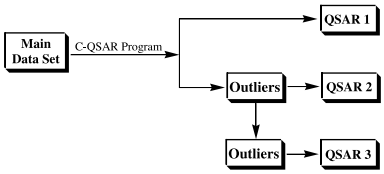



**Figure 1.** Scheme for the splitting of main data set and development of QSAR.

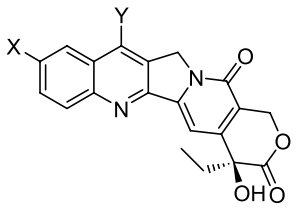
In QSAR equations, $n$ is the number of data points, $r$ is the correlation coefficient, $r^2$ is the goodness of fit, $q^2$ is the goodness of prediction, and $s$ is the standard deviation. All the QSAR reported here are derived by us and were not given with the original data sets taken from the literature as referenced. In this review, we are presenting the experimental data from various types of molecules, which we find a large enough data set to understand the problem with outliers in QSAR.

## 3. Results and discussion

### 3.1. Splitting the main data set into three sub data sets

#### 3.1.1. Camptothecin derivatives

**3.1.1.1. Cytotoxic activity of camptothecin derivatives I in vitro against H460 human non-small lung carcinoma cell line.** Data from Dallavalle et al.[13] (Tables 1a–c).



**I**

Dallavalle et al.[13] studied the cytotoxicity of camptothecin derivatives **I** in vitro against H460 human NSCLC cell line. From these results, we were unable to derive a good QSAR due to the presence of 15 outliers out of 33 compounds, which is not acceptable. The presence of a large number of outliers may be due to their interaction with the receptors in different modes. Thus, we removed all the 15 outliers from the main data set of QSAR 1a to make it useful. Now these 15 outliers were subjected to the formulation of QSAR 1b, which gave 5 outliers. These 5 outliers were removed from QSAR 1b, which was further used in the derivation of QSAR 1c. Finally, the main data set of 33 compounds was split into three sub data sets (see Fig. 1), which gave three QSAR equations 1a–1c. It is interesting to note here that all three QSAR equations 1a–1c have good statistics, and do not have any outlier.

**Table 1a.** Biological and physicochemical constants used to derive QSAR equation 1a for the cytotoxic activity of camptothecin derivatives (**Ia**–**Ir**) against H460 human non-small lung carcinoma cell line

| No. | X | Y | log 1/*C* (Eq. 1a) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | | |
| **Ia** | H | H (camptothecin) | 6.48 | 6.58 | −0.10 | 0.90 | 9.52 |
| **Ib** | OH | $C_2H_5$ (SN-38) | 7.10 | 7.48 | −0.39 | 1.97 | 10.60 |
| **Ic** | OH | CH=NOCH$_3$ | 6.70 | 6.74 | −0.04 | 1.04 | 11.25 |
| **Id** | H | CH=NOCH$_2$CH=CH$_2$ | 7.22 | 7.38 | −0.16 | 1.77 | 12.00 |
| **Ie** | H | CH=NOC(CH$_3$)$_3$ | 7.82 | 7.54 | 0.28 | 2.23 | 12.49 |
| **If** | $OCH_3$ | CH=NOC(CH$_3$)$_3$ | 7.15 | 7.52 | −0.37 | 2.51 | 13.11 |
| **Ig** | H | CH=NOC(CH$_3$)$_2$CH$_2$OH | 6.47 | 6.45 | 0.02 | 0.78 | 12.64 |
| **Ih** | H | CH=NOCH$_2$CH$_2$N(CH$_3$)$_2$ | 6.52 | 6.74 | −0.22 | 1.05 | 12.86 |
| **Ii** | H | CH=NOCH$_2$COOH | 5.63 | 5.77 | −0.14 | 0.21 | 11.75 |
| **Ij** | H | CH=NOCH$_2$CONH(CH$_2$)$_3$NH$_2$ | 5.28 | 5.18 | 0.10 | −0.28 | 13.73 |
| **Ik** | H | CH=NOCH$_2$CH$_2$-morpholinyl | 7.10 | 6.77 | 0.33 | 1.07 | 13.76 |
| **Il** | H | CH=NOCH$_2$C$_6$H$_5$ | 7.52 | 7.49 | 0.03 | 1.99 | 13.61 |
| **Im** | H | C(CH$_3$)=NOCH$_2$C$_6$H$_5$ | 6.89 | 6.75 | 0.14 | 3.74 | 14.07 |
| **In** | H | CH=NOCH$_2$C$_6$H$_4$-(*p*-CH$_3$) | 7.70 | 7.53 | 0.17 | 2.49 | 14.07 |
| **Io** | H | CH=NOCH$_2$C$_6$H$_4$-(*p*-NO$_2$) | 7.77 | 7.36 | 0.41 | 1.74 | 14.22 |
| **Ip** | H | CH=NOCH$_2$C$_6$F$_5$ | 7.55 | 7.52 | 0.03 | 2.50 | 13.69 |
| **Iq** | H | CH=NOCH$_2$C$_6$H$_4$-(*p*-C$_6$H$_5$) | 6.55 | 6.63 | −0.08 | 3.88 | 16.12 |
| **Ir** | H | CH=NOC(C$_6$H$_5$)$_3$ | 5.84 | 5.86 | −0.02 | 4.78 | 18.63 |

**Table 1b.** Biological and physicochemical constants used to derive QSAR equation 1b for the cytotoxic activity of camptothecin derivatives (**Is**–**Iab**) against H460 human non-small lung carcinoma cell line

| No. | X | Y | log 1/*C* (Eq. 1b) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | | |
| **Is** | H | CH=NOH | 7.49 | 7.49 | 0.00 | 1.15 | 10.63 |
| **It** | H | CH=NOCH$_3$ | 7.40 | 7.48 | −0.09 | 1.00 | 11.10 |
| **Iu** | H | CH=NOCH$_2$CH(CH$_2$)O | 7.70 | 7.59 | 0.11 | 0.52 | 12.00 |
| **Iv** | OH | CH=NOC(CH$_3$)$_3$ | 6.89 | 6.82 | 0.06 | 2.28 | 12.64 |
| **Iw** | H | CH=NOC(CH$_3$)$_2$COOC(CH$_3$)$_3$ | 6.85 | 6.90 | −0.05 | 2.41 | 14.53 |
| **Ix** | H | CH=NO(CH$_2$)$_2$-3-(N–CH$_3$)piperidinyl | 7.00 | 7.06 | −0.06 | 2.03 | 14.54 |
| **Iy** | H | CH=NOC$_6$H$_5$ | 6.80 | 6.79 | 0.01 | 2.37 | 13.15 |
| **Iz** | H | CH=NOCH$_2$C$_6$H$_4$-(*p*-NH$_2$) | 7.60 | 7.53 | 0.07 | 0.77 | 13.98 |
| **Iaa** | H | CH=NOCH$_2$-9-anthracenyl | 6.72 | 6.70 | 0.02 | 4.34 | 16.98 |
| **Iab** | H | CH=NOCH$_2$-4-pyridyl | 7.52 | 7.60 | −0.07 | 0.50 | 13.40 |

**Table 1c.** Biological and physicochemical constants used to derive QSAR equation 1c for the cytotoxic activity of camptothecin derivatives (**Iac**–**Iag**) against H460 human non-small lung carcinoma cell line

| No. | X | Y | log 1/*C* (Eq. 1c) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | | |
| **Iac** | H | CH=NOCOC$_6$H$_5$ | 5.75 | 5.70 | 0.05 | 1.94 | 13.64 |
| **Iad** | H | C(C$_6$H$_5$)=NOH | 5.29 | 5.32 | −0.03 | 2.95 | 13.15 |
| **Iae** | H | CH=NOCH$_2$CH$_2$NH$_2$ | 6.31 | 6.36 | −0.05 | 0.15 | 11.93 |
| **Iaf** | H | CH=NOCH$_2$CH$_2$-1-uracylyl | 6.30 | 6.47 | −0.17 | −0.16 | 14.08 |
| **Iag** | H | CH=NOCH$_2$-2-imidazolyl | 6.68 | 6.48 | 0.20 | −0.18 | 12.83 |

$$\log 1/C = 1.23(\pm 0.22)C\log P - 2.11(\pm 0.39) \times \log(\beta \times 10^{C\log P} + 1) + 5.53(\pm 0.29)$$
$$n = 18, \quad r^2 = 0.913, \quad q^2 = 0.888, \quad s = 0.240 \quad (1a)$$

optimum $C\log P = 2.30$

$\log \beta = -2.16$

$$\log 1/C = -0.42(\pm 0.09)C\log P - 1.07(\pm 0.50)\mathrm{CMR} + 0.04(\pm 0.02)\mathrm{CMR}^2 + 14.60(\pm 3.33)$$
$$n = 10, \quad r^2 = 0.969, \quad q^2 = 0.823, \quad s = 0.081 \quad (1b)$$

inversion point for CMR = 12.76 (11.98–13.39)

$$\log 1/C = -0.37(\pm 0.18)C\log P + 6.42(\pm 0.28)$$
$$n = 5, \quad r^2 = 0.937, \quad q^2 = 0.844, \quad s = 0.158 \quad (1c)$$

The hydrophobicity of the molecules correlates with their cytotoxic activity in a bilinear fashion in Eq. 1a, which suggests that the activity of compounds (**Ia**–**Ir**) first increases linearly with an increase in hydrophobicity to an optimum $C\log P$ of 2.30 and then decreases linearly. Eq. 1b is an allosteric correlation with molar refractivity, followed by a linear correlation with negative hydrophobicity. An allosteric correlation with CMR suggests that the activity of compounds (**Is**–**Iab**) first decreases with an increase in CMR up to the inversion point for CMR = 12.76 and then increases. Eq. 1c indicates a negative effect of hydrophobicity for the molecules (**Iac**–**Iag**). QSAR equations 1a–1c explain 91.3%, 96.9%, and 93.7% of the variance, respectively, in the cytotoxicity data of camptothecin derivatives **I**. It is of interest to note here that all the three QSAR equations 1a–1c are different in nature, suggesting that the compounds associated with these equations may be interacting with the receptor in different modes.

The term 'allosteric' (from Greek origin) means 'another shape'. A mathematical model for allosteric interactions in proteins which are oligomeric in nature was first proposed by Monod et al.,[14] which are oligomeric in nature and the model was applied with considerable success on several enzymes or metabolically active proteins. The early work in this fascinating field has already been reviewed.[15–17] Allosteric effects have been considered to occur when the interaction between ligand and receptor results in a change in the structure of the receptor. The first precise understanding came from the X-ray analysis of the exact location of the ligand on the hemoglobin subunits.[16] Hemoglobin is an incredibly complex substance that can, apparently, be composed of possibly many subunits.[18] Recently, the allosteric effect has been illustrated for single receptor or enzymes.[19,20] Bender et al.[21] have inferred an allosteric interaction via QSAR.

Until the use of QSAR to uncover such a reaction, establishing an allosteric reaction was a tedious process. In an allosteric correlation, at first the activity decreases as the value of dependent parameter increases, but then at a specific point it turns around and increases. Obviously, a change in the mechanism of the reaction occurs at the inversion point. Thus, allosteric interactions are mainly due to a change in the structure of the receptor. The result of an allosteric reaction is a U- or V-shaped curve. This is easy to see by a simple plot of the data when a single parameter is involved. The way one detects such a reaction from the QSAR model can be illustrated by Eq. i with $C\log P$ (for inverted parabola):

$$\log 1/C = -aC\log P + bC\log P^2 \pm \text{constant} \quad (i)$$

This equation implies that at first, as $C\log P$ increases, the activity decreases; but then at a certain point, the exponential term takes over and the activity starts to increase with further increase in $C\log P$ value. Similarly, Eq. ii represents the allosteric bilinear correlation with $C\log P$ (inverted bilinear):

$$\log 1/C = -aC\log P + b\log(\beta \times 10^{C\log P} + 1) \pm \text{constant} \quad (ii)$$

In terms of CMR, Eqs. i and ii may be represented by Eqs. iii and iv, respectively.

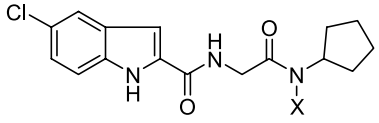
$$\log 1/C = -a\mathrm{CMR} + b\mathrm{CMR}^2 \pm \text{constant} \quad (iii)$$

$$\log 1/C = -a\mathrm{CMR} + b\log(\beta \times 10^{\mathrm{CMR}} + 1) \pm \text{constant} \quad (iv)$$

There is a practical as well as a theoretical side to allosteric reactions. In designing a new drug or studying environmental toxicology and trying to gain perspective via QSAR, one must have a considerable range in the properties of the parameters or one may obtain a false impression of the predictive properties of a QSAR. Christopoulos[22] reviewed extensively the importance of allosteric interactions for drug discovery and their complexity. For in-depth knowledge about the use of QSAR in allosteric reaction, the interested reader is to refer to earlier publications.[23–27]

### 3.1.2. 5-Chloroindolyl derivatives

**3.1.2.1. Inhibition of glycogen phosphorylase A (GPA, EC 2.4.1.1) by 5-chloroindolyl derivatives II.** Data from Wright et al.[28] (Tables 2a–c).




**II**

From the data of Wright et al.,[28] we again are unable to derive a good QSAR due to the presence of 16 outliers out of 37 compounds. This data set was also divided into three sub data sets according to Figure 1, which gave QSAR equations 2a–2c with good statistics.

$$\log 1/C = -4.96(\pm 2.61)\mathrm{CMR} + 0.20(\pm 0.12)\mathrm{CMR}^2 + 36.54(\pm 14.44)$$
$$n = 21, \quad r^2 = 0.855, \quad q^2 = 0.819, \quad s = 0.193 \quad (2a)$$

inversion point for CMR = 12.38 (11.85–14.30)

$$\log 1/C = 28.01(\pm 6.42)C\log P - 4.19(\pm 0.95)C\log P^2 - 39.39(\pm 10.74)$$
$$n = 10, \quad r^2 = 0.940, \quad q^2 = 0.866, \quad s = 0.249 \quad (2b)$$

optimum $C\log P = 3.34$ (3.28–3.40)

**Table 2a.** Biological and physicochemical constants used to derive QSAR equation 2a for the inhibition of glycogen phosphorylase A (GPA) by 5-chloroindolyl derivatives (**IIa–IIu**)

| No. | X | log 1/*C* (Eq. 2a) | | | CMR |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIa** | $CH_2CN$ | 7.52 | 7.41 | 0.12 | 9.58 |
| **IIb** | $CH_2CH_2OH$ | 7.24 | 7.25 | −0.01 | 9.72 |
| **IIc** | Allyl | 7.02 | 6.97 | 0.05 | 10.01 |
| **IId** | $CH_2CH_2CN$ | 6.85 | 6.93 | −0.07 | 10.04 |
| **IIe** | $CH_2C(Me)_2OH$ | 6.82 | 6.44 | 0.39 | 10.65 |
| **IIf** | $CH_2CH_2OMe$ | 6.70 | 6.80 | −0.10 | 10.18 |
| **IIg** | $CH(Me)CH_2CH_2OH$ | 6.68 | 6.44 | 0.24 | 10.65 |
| **IIh** | $CH_2CH_2OCH_2CH_2OH$ | 6.64 | 6.33 | 0.30 | 10.80 |
| **IIi** | $CH(CH_2OH)_2$ | 6.46 | 6.67 | −0.22 | 10.34 |
| **IIj** | $CH_2CH_2CH_2OMe$ | 6.43 | 6.44 | 0.00 | 10.65 |
| **IIk** | $CH_2CONMe_2$ | 6.39 | 6.27 | 0.11 | 10.90 |
| **IIl** | $CH_2CH(NH_2)CH_2OH$ | 6.29 | 6.50 | −0.21 | 10.55 |
| **IIm** | $CH_2COOEt$ | 6.28 | 6.41 | −0.14 | 10.68 |
| **IIn** | Cyclopentyl | 6.18 | 6.35 | −0.17 | 10.78 |
| **IIo** | $(CH2)_3O$-isopropyl | 6.17 | 5.96 | 0.21 | 11.58 |
| **IIp** | $CH_2CH(OH)CH_2O$-isopropyl | 6.03 | 5.92 | 0.11 | 11.73 |
| **IIq** | $CH_2CH_2O$-isopropyl | 5.96 | 6.16 | −0.20 | 11.11 |
| **IIr** | $CH_2CH(NH_2)CH_2OMe$ | 5.92 | 6.21 | −0.29 | 11.02 |
| **IIs** | $CH_2CH_2OPh$ | 5.85 | 5.84 | 0.01 | 12.23 |
| **IIt** | $CH_2CH_2$-(4-morpholinyl) | 5.82 | 5.91 | −0.09 | 11.77 |
| **IIu** | $CH_2CH_2OCH_2Ph$ | 5.80 | 5.85 | −0.06 | 12.70 |

**Table 2b.** Biological and physicochemical constants used to derive QSAR equation 2b for the inhibition of glycogen phosphorylase A (GPA) by 5-chloroindolyl derivatives (**IIv–IIae**)

| No. | X | log 1/*C* (Eq. 2b) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIv** | Methyl | 7.26 | 7.41 | −0.15 | 3.31 |
| **IIw** | $CH_2CH(OH)CH_3$ | 7.80 | 7.39 | 0.40 | 3.42 |
| **IIx** | $CH_2CH_2CH_2OH$ | 7.38 | 7.42 | −0.04 | 3.35 |
| **IIy** | $CH_2CH(OH)CH_2OH$ | 7.27 | 7.38 | −0.11 | 3.25 |
| **IIz** | $CH_2C(Me)(OH)CH_2OH$ | 7.08 | 7.02 | 0.06 | 3.65 |
| **IIaa** | $CH_2COOH$ | 6.59 | 6.93 | −0.34 | 3.68 |
| **IIab** | $CH_2CH_2NMe_2$ | 5.01 | 5.13 | −0.11 | 4.08 |
| **IIac** | $CH_2CH(OH)CH_2OMe$ | 7.00 | 7.06 | −0.06 | 3.05 |
| **IIad** | Ethyl | 6.68 | 6.37 | 0.31 | 3.84 |
| **IIae** | $CH_2CONH_2$ | 5.34 | 5.29 | 0.05 | 2.63 |

**Table 2c.** Biological and physicochemical constants used to derive QSAR equation 2c for the inhibition of glycogen phosphorylase A (GPA) by 5-chloroindolyl derivatives (**IIaf–IIak**)

| No. | X | log 1/*C* (Eq. 2c) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIaf** | $CH_2CH_2CH_2NMe_2$ | 5.30[a] | 6.48 | −1.18 | 4.28 |
| **IIag** | *n*-Butyl | 7.09 | 6.99 | 0.10 | 4.90 |
| **IIah** | *n*-Propyl | 6.47 | 6.55 | 0.08 | 4.37 |
| **IIai** | $CH_2CH_2NH_2$ | 5.80 | 5.59 | 0.21 | 3.18 |
| **IIaj** | $CH_2CH_2CH_2NH_2$ | 5.52 | 5.79 | −0.27 | 3.43 |
| **IIak** | $CH_2CH_2CH_2CH_2OH$ | 5.41 | 5.37 | 0.04 | 2.91 |

[a] Data point not included in the derivation of Eq. 2c.

$$\log 1/C = 0.81(\pm 0.41)C\log P + 3.01(\pm 1.55)$$
$$n = 5, \quad r^2 = 0.931, \quad q^2 = 0.831, \quad s = 0.215 \qquad (2c)$$

outlier: **IIaf**

The GPA inhibitory activity of the molecules (**IIa–IIu**) correlates with their molar refractivity in an inverted parabolic fashion, as represented by Eq. 2a. Thus, the activity of these compounds first decreases with an increase in molar refractivity up to the inversion point for CMR = 12.38 and then increases. Eq. 2b is a parabolic correlation with *C*log *P*, which suggests that the activity of compound (**IIv–IIae**) first increases with an increase in hydrophobicity to an optimum *C*log *P* of 3.34 and then decreases. It is interesting that in Eq. 2c we obtained a linear correlation with *C*log *P* with posi-

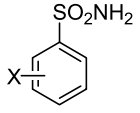
**Table 3a.** Biological and physicochemical constants used to derive QSAR equation 3a for the inhibition of human carbonic anhydrase II (hCA II) by sulfonamide derivatives (**IIIAa–IIIAl**)

| No. | X | log 1/*C* (Eq. 3a) | | | CMR |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIIAa** | 2-$NH_2$ | 5.58 | 5.51 | 0.06 | 4.30 |
| **IIIAb** | 4-NCS | 5.44 | 5.65 | −0.21 | 5.48 |
| **IIIAc** | 4-$CH_2OH$ | 5.24 | 5.45 | −0.20 | 4.55 |
| **IIIAd** | 3-F,4-$NH_2$ | 5.52 | 5.51 | 0.01 | 4.31 |
| **IIIAe** | 3-I,4-$NH_2$ | 5.76 | 5.74 | 0.02 | 5.60 |
| **IIIAf** | 4-$NH_2$ | 5.62 | 5.51 | 0.11 | 4.30 |
| **IIIAg** | 4-$CH_2NH_2$ | 5.40 | 5.43 | −0.03 | 4.76 |
| **IIIAh** | 4-$CH_2CH_2NH_2$ | 5.51 | 5.53 | −0.02 | 5.23 |
| **IIIAi** | 4-$CH_2CH_2OH$ | 5.52 | 5.46 | 0.07 | 5.01 |
| **IIIAj** | 2-$CF_3$,4-$NH_2$,5-$SO_2NH_2$ | 6.47 | 6.15 | 0.32 | 6.05 |
| **IIIAk** | 4-2-($NH_2$-pyrimidin-4-ylamino) | 7.05 | 7.14 | −0.09 | 6.76 |
| **IIIAl** | 4-NHNHC(=S)$NH_2$ | 6.54 | 6.58 | −0.04 | 6.39 |

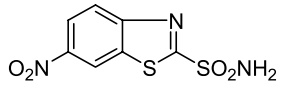
tive coefficient. QSAR equations 2a–2c explain 85.5%, 94.0%, and 93.1% of the variance, respectively, in the GPA inhibitory activity data of the molecules (**IIa–IIu**), (**IIv–IIae**), and (**IIaf–IIak**), respectively. It is of interest to note here that all the three QSAR equations 2a–2c are different in nature, which suggests that the compounds associated with these equations may be interacting with the receptor in different modes.

### 3.1.3. Sulfonamide derivatives

**3.1.3.1. Inhibition of human carbonic anhydrase II (hCA II) by sulfonamide derivatives (IIIA and IIIB).** Data from Innocenti et al.[29] (Tables 3a–c).




**IIIA**




**IIIB**

**Table 3b.** Biological and physicochemical constants used to derive QSAR equation 3b for the inhibition of human carbonic anhydrase II (hCA II) by sulfonamide derivatives (**IIIAm–IIIAr**)

| No. | X | log 1/*C* (Eq. 3b) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIIAm** | 3-$NO_2$ | 6.39 | 6.36 | 0.03 | 0.58 |
| **IIIAn** | 2-$NHNH_2$ | 7.10 | 7.07 | 0.02 | −0.19 |
| **IIIAo** | 4-$NO_2$ | 6.40 | 6.36 | 0.04 | 0.58 |
| **IIIAp** | 4-$CH_3$ | 6.24 | 6.14 | 0.09 | 0.80 |
| **IIIAq** | 3-Br,4-$NH_2$ | 6.21 | 6.42 | −0.21 | 0.51 |
| **IIIAr** | 4-$NHNH_2$ | 7.09 | 7.07 | 0.02 | −0.19 |

From the data of Innocenti et al.,[29] we derived the following three QSAR equations by splitting into three sub data sets on the basis of Figure 1:

$$\log 1/C = -4.04(\pm 1.93)\mathrm{CMR} + 0.43(\pm 0.18)\mathrm{CMR}^2 + 15.01(\pm 5.12)$$
$$n = 12, \quad r^2 = 0.937, \quad q^2 = 0.882, \quad s = 0.156 \qquad (3a)$$

inversion point for CMR = 4.75 (4.23–4.98)

$$\log 1/C = -0.93(\pm 0.34)C\log P + 6.89(\pm 0.18)$$
$$n = 6, \quad r^2 = 0.935, \quad q^2 = 0.886, \quad s = 0.117 \qquad (3b)$$

$$\log 1/C = 2.24(\pm 1.22)\mathrm{MgVol} + 3.20(\pm 1.80)$$
$$n = 4, \quad r^2 = 0.969, \quad q^2 = 0.892, \quad s = 0.091 \qquad (3c)$$

outlier: **IIIAs**

Eq. 3a is an allosteric parabolic correlation with molar refractivity, which suggests that the inhibitory activity of the compounds (**IIIAa–IIIAl**) toward hCA II first decreases with an increase in molar refractivity up to the inversion point of CMR = 4.75 and then increases. Eq. 3b indicates the negative hydrophobic effects for the compounds (**IIIAm–IIIAr**). In Eq. 3c, we obtained a linear correlation with molar volume for the compounds (**IIIAs–IIIAv** and **IIIB**). QSAR equations 3a–3c explain 93.7%, 93.5%, and 96.9% of the variance, respectively, in the hCA II inhibitory data of the molecules (**IIIAa–IIIAl**), (**IIIAm–IIIAr**), and (**IIIAs–IIIAv** and **IIIB**), respectively.

**Table 3c.** Biological and physicochemical constants used to derive QSAR equation 3c for the inhibition of human carbonic anhydrase II (hCA II) by sulfonamide derivatives (**IIIAs–IIIAv** and **IIIB**)

| No. | X | log 1/*C* (Eq. 3c) | | | MgVol |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **IIIAs** | 2-$COOCH_3$ | 4.13[a] | 6.46 | −2.33 | 1.45 |
| **IIIAt** | 4-COOH | 6.06 | 6.14 | −0.09 | 1.31 |
| **IIIAu** | 3-Cl,4-$NH_2$ | 6.25 | 6.16 | 0.09 | 1.32 |
| **IIIAv** | 2-Cl,4-$NH_2$,5-$SO_2NH_2$ | 7.02 | 7.01 | 0.01 | 1.70 |
| **IIIB** | 2-$SO_2NH_2$-6-$NO_2$-benzothiazole | 6.60 | 6.62 | −0.02 | 1.52 |

[a] Data point not included in the derivation of Eq. 3c.

**3.1.3.2. Inhibition of human carbonic anhydrase II isozyme (hCA II) by sulfonamide derivatives IV (Fig. 2).** Data from Vullo et al.[30] (Tables 4a–c).

The data of Vullo et al.[30] were divided into three sub data sets on the basis of outliers (see Fig. 1), which gave the following QSAR equations:

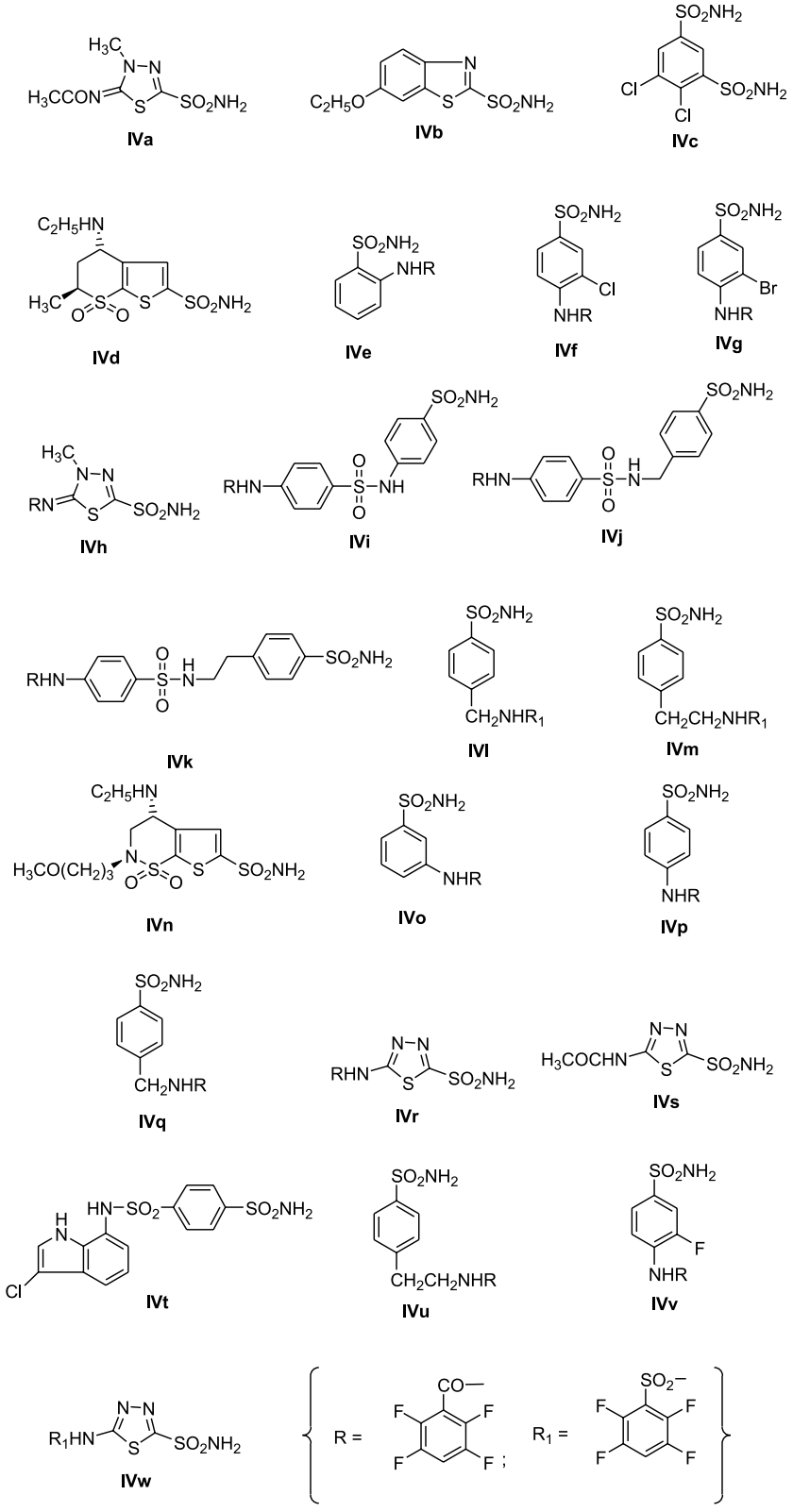



**Figure 2.** Structure of sulfonamide derivatives (**IVa**–**IVw**) used in the derivation of Eqs. 4a–4c.

**Table 4a.** Biological and physicochemical constants used to derive QSAR equation 4a for the inhibition of human carbonic anhydrase II isozyme (hCA II) by sulfonamide derivatives (**IVa–IVm**) (Fig. 2)

| Compound | $\log 1/K_i$ (Eq. 4a) | | | $C\log P$ |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **IVa** | 7.85 | 7.63 | 0.23 | 0.09 |
| **IVb** | 8.10 | 8.21 | −0.11 | 2.05 |
| **IVc** | 7.42 | 7.51 | −0.09 | 0.24 |
| **IVd** | 8.05 | 8.12 | −0.07 | −0.43 |
| **IVe** | 7.42 | 7.28 | 0.14 | 0.92 |
| **IVf** | 7.12 | 7.27 | −0.16 | 0.88 |
| **IVg** | 7.08 | 7.29 | −0.22 | 1.00 |
| **IVh** | 8.85 | 8.73 | 0.12 | 2.43 |
| **IVi** | 7.72 | 7.69 | 0.03 | 1.61 |
| **IVj** | 7.82 | 7.56 | 0.26 | 1.47 |
| **IVk** | 7.89 | 7.90 | −0.02 | 1.80 |
| **IVl** | 7.92 | 7.84 | 0.08 | 1.74 |
| **IVm** | 8.05 | 8.24 | −0.20 | 2.07 |

**Table 4b.** Biological and physicochemical constants used to derive QSAR equation 4b for the inhibition of human carbonic anhydrase II isozyme (hCA II) by sulfonamide derivatives (**IVn–IVr**) (Fig. 2)

| Compound | $\log 1/K_i$ (Eq. 4b) | | | $C\log P$ |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **IVn** | 8.52 | 8.51 | 0.01 | 0.33 |
| **IVo** | 7.68 | 7.81 | −0.14 | 0.92 |
| **IVp** | 7.80 | 7.81 | −0.02 | 0.92 |
| **IVq** | 7.89 | 7.75 | 0.14 | 0.98 |
| **IVr** | 8.68 | 8.68 | 0.00 | 0.20 |

**Table 4c.** Biological and physicochemical constants used to derive QSAR equation 4c for the inhibition of human carbonic anhydrase II isozyme (hCA II) by sulfonamide derivatives (**IVs–IVw**) (Fig. 2)

| Compound | $\log 1/K_i$ (Eq. 4c) | | | MgVol |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **IVs** | 7.92[a] | 11.08 | −3.16 | 1.34 |
| **IVt** | 7.82 | 7.70 | 0.12 | 2.44 |
| **IVu** | 7.92 | 8.08 | −0.16 | 2.31 |
| **IVv** | 8.82 | 8.89 | −0.07 | 2.05 |
| **IVw** | 9.15 | 9.05 | 0.11 | 2.00 |

[a] Data point not included in the derivation of Eq. 4c.

$$\log 1/K_i = -1.14(\pm 0.45) C\log P + 2.64(\pm 0.82) \times \log(\beta \times 10^{C\log P} + 1) + 7.58(\pm 0.21)$$
$$n = 13, \quad r^2 = 0.887, \quad q^2 = 0.802, \quad s = 0.182 \qquad (4a)$$

inversion point for $C\log P = 0.84$

$$\log \beta = -0.96$$
$$\log 1/K_i = -1.86(\pm 0.48) C\log P + 8.91(\pm 0.36)$$
$$n = 5, \quad r^2 = 0.954, \quad q^2 = 0.894, \quad s = 0.113 \qquad (4b)$$

$$\log 1/K_i = -3.07(\pm 2.00)\mathrm{MgVol} + 15.19(\pm 4.41)$$
$$n = 4, \quad r^2 = 0.956, \quad q^2 = 0.788, \quad s = 0.169 \qquad (4c)$$

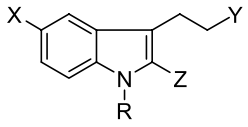
outlier: **IVs**

The hydrophobicity of the molecules (**IVa–IVm**) correlates with their inhibitory activity to the hCA II in an inverted bilinear fashion in Eq. 4a. The inversion point of this equation is $C\log P = 0.84$. In Eq. 4b, we obtained a negative effect of hydrophobic term of the ligands (**IVn–IVr**). Eq. 4c is a linear correlation with the molar volume of the compounds (**IVs–IVw**). QSAR equations 4a–4c explain 88.7%, 95.4%, and 95.6% of the variance, respectively, in the hCA II inhibitory data of the molecules (**IVa–IVm**), (**IVn–IVr**), and (**IVs–IVw**), respectively.

### 3.2. Splitting the main data set into two sub data sets

#### 3.2.1. Tryptamine derivatives

**3.2.1.1. Binding affinity of 1-R,2-Z,3-($CH_2CH_2$-Y),5-X-tryptamines V to 5-HT6 receptor in HeLa cells.** Data from Russel et al.[31] (Tables 5a and b).




**V**

This data set was divided into two sub data sets according to Figure 1, which gave QSAR equations 5a and 5b.

$$\log 1/K_i = 1.01(\pm 0.28) C\log P - 6.57(\pm 1.96) \times \log(\beta \times 10^{C\log P} + 1) + 4.45(\pm 0.92)$$
$$n = 14, \quad r^2 = 0.870, \quad q^2 = 0.825, \quad s = 0.268 \qquad (5a)$$

optimum $C\log P = 4.13$

$$\log \beta = -4.87$$

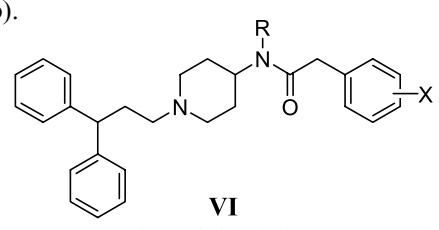
$$\log 1/K_i = 1.38(\pm 0.84)\sigma_X^+ + 1.60(\pm 0.66) I_Y + 7.12(\pm 0.85)$$
$$n = 8, \quad r^2 = 0.944, \quad q^2 = 0.841, \quad s = 0.294 \qquad (5b)$$

outlier: **Vs**

Eq. 5a is a bilinear correlation with hydrophobicity of the molecules (**Va–Vn**). The optimum value of $C\log P$ for this equation is 4.13. On the other hand, $\sigma^+$ of X-substituents makes the best prediction for the binding affinity in Eq. 5b; this suggests that electron-withdrawing X-substituents of compounds (**Vo–Vw**) will favor their activity via resonance. The indicator variable $I_Y$ takes the value of 1 for Y = $NMe_2$.

#### 3.2.2. Phenylacetamide derivatives

**3.2.2.1. Binding affinity to human CCR5 receptor by phenylacetamides VI.** Data from Burrows et al.[32] (Tables 6a and b).




**VI**

This data set was also divided into two sub data sets according to Figure 1, which gave QSAR equations 6a and 6b.

$$\log 1/C = -0.36(\pm 0.09)\pi_{X\text{-}4} + 0.92(\pm 0.21) B1_{X\text{-}4} + 4.88(\pm 0.33)$$
$$n = 33, \quad r^2 = 0.888, \quad q^2 = 0.866, \quad s = 0.209 \qquad (6a)$$

**Table 5a.** Biological and physicochemical constants used to derive QSAR equation 5a for the binding affinity of 1-R,2-Z,3-($CH_2CH_2$-Y),5-X-tryptamines (**Va**–**Vn**) to 5-HT6 receptor in HeLa cells

| No. | X | Y | Z | R | $\log 1/K_i$ (Eq. 5a) | | | *C*log *P* |
|---|---|---|---|---|---|---|---|---|
| | | | | | Observed | Calculated | $\Delta$ | |
| **Va** | OMe | $NH_2$ | H | $SO_2Ph$ | 7.40 | 7.59 | −0.20 | 3.16 |
| **Vb** | OMe | $NMe_2$ | H | $SO_2Ph$ | 8.64 | 8.09 | 0.55 | 3.84 |
| **Vc** | OMe | $NMe_2$ | H | $SO_2Ph$ (2-Cl) | 7.96 | 7.94 | 0.02 | 4.55 |
| **Vd** | OMe | $NMe_2$ | H | $SO_2Ph$ (3-Cl) | 8.10 | 7.94 | 0.16 | 4.55 |
| **Ve** | OMe | $NMe_2$ | H | $SO_2Ph$ (4-Cl) | 7.77 | 7.94 | −0.17 | 4.55 |
| **Vf** | OMe | $NMe_2$ | H | $SO_2$-(2-thienyl) | 8.08 | 7.88 | 0.20 | 3.51 |
| **Vg** | OMe | Piperidinyl | H | $SO_2Ph$ | 6.46 | 6.45 | 0.01 | 5.20 |
| **Vh** | H | $NMe_2$ | H | $SO_2Me$ | 6.21 | 6.13 | 0.08 | 1.66 |
| **Vi** | OH | $NMe_2$ | H | $SO_2Ph$ | 7.72 | 7.84 | −0.12 | 3.45 |
| **Vj** | OH | $NMe_2$ | H | COPh | 7.27 | 7.57 | −0.30 | 3.13 |
| **Vk** | CN | $NMe_2$ | H | $SO_2Ph$ | 7.48 | 7.74 | −0.26 | 3.33 |
| **Vl** | OMe | $NMe_2$ | Me | $SO_2Ph$ | 7.92 | 8.13 | −0.21 | 4.29 |
| **Vm** | OMe | $NMe_2$ | Me | H | 7.05 | 7.04 | 0.01 | 2.58 |
| **Vn** | H | $NMe_2$ | COOEt | H | 7.70 | 7.46 | 0.24 | 3.01 |

**Table 5b.** Biological and physicochemical constants used to derive QSAR equation 5b for the binding affinity of 1-R,2-Z,3-($CH_2CH_2$-Y),5-X-tryptamines (**Vo**–**Vw**) to 5-HT6 receptor in HeLa cells

| No. | X | Y | Z | R | $\log 1/K_i$ (Eq. 5b) | | | $\sigma_X^+$ | $I_Y$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Observed | Calculated | $\Delta$ | | |
| **Vo** | OMe | $NMe_2$ | H | $SO_2Ph$(4-Me) | 7.35 | 7.64 | −0.29 | −0.78 | 1 |
| **Vp** | OMe | $NMe_2$ | H | $SO_2Ph$(4-OMe) | 7.59 | 7.64 | −0.05 | −0.78 | 1 |
| **Vq** | OMe | $NMe_2$ | H | $SO_2$(2-naphthyl) | 8.01 | 7.64 | 0.37 | −0.78 | 1 |
| **Vr** | OMe | $NMe_2$ | H | COPh | 7.60 | 7.64 | −0.04 | −0.78 | 1 |
| **Vs** | OMe | Pyrrolidinyl | H | $SO_2Ph$ | 7.36[a] | 6.04 | 1.32 | −0.78 | 0 |
| **Vt** | OMe | 4-Me-piperazinyl | H | $SO_2Ph$ | 6.31 | 6.04 | 0.27 | −0.78 | 0 |
| **Vu** | OMe | Morpholinyl | H | $SO_2Ph$ | 5.77 | 6.04 | −0.27 | −0.78 | 0 |
| **Vv** | H | $NMe_2$ | H | $SO_2Ph$ | 8.54 | 8.71 | −0.17 | 0.00 | 1 |
| **Vw** | H | $NMe_2$ | 3-(3-OMe-$CH_2$Ph)-1,2,4-oxadiazole-5-yl | H | 8.89 | 8.71 | 0.18 | 0.00 | 1 |

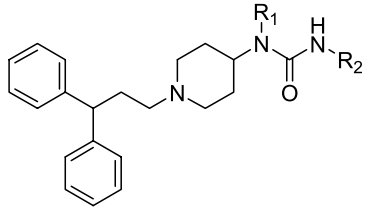
[a] Data point not included in the derivation of Eq. 5b.

$$\log 1/C = -0.44(\pm 0.07)\pi_{X\text{-}4} + 6.95(\pm 0.06)$$
$$n = 5, \quad r^2 = 0.993, \quad q^2 = 0.981, \quad s = 0.038 \qquad (6b)$$

The most important parameter for both Eqs. 6a and 6b is the hydrophobic term $\pi_{X\text{-}4}$, which shows that only substituents in position 4 of the phenyl ring contact the hydrophobic space on the receptor, whereas the negative coefficient shows that there is a need for a less lipophilic substituent. The $B1_{X\text{-}4}$ term (the sterimol smallest width) appearing in Eq. 6a confirms a positive steric effect for position 4 substituents of the phenyl ring.

### 3.2.3. Ureas

**3.2.3.1. Binding affinity to human CCR5 receptor by ureas VII.** Data from Burrows et al.[32] (Tables 7a and b).



**VII**

From this data set, we obtained two QSAR equations 7a and 7b (see Fig. 1).

$$\log 1/C = 3.25(\pm 1.22)\mathrm{MgVol} - 6.49(\pm 4.77)$$
$$n = 8, \quad r^2 = 0.876, \quad q^2 = 0.812, \quad s = 0.297 \qquad (7a)$$

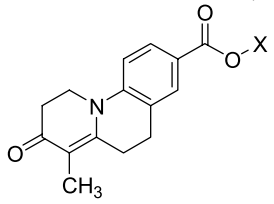
$$\log 1/C = 0.46(\pm 0.18)\mathrm{CMR} + 2.50(\pm 0.67)$$
$$n = 5, \quad r^2 = 0.958, \quad q^2 = 0.848, \quad s = 0.070 \qquad (7b)$$

outlier: **VIIi**

There is a linear correlation between binding affinity of the molecules and their MgVol in Eq. 7a, which suggests that the binding affinity of the molecules (**VIIa**–**VIIh**) increases with increase in their MgVol. On the other hand, there is a linear corelation between the binding affinity of the molecules and their CMR in Eq. 7b, which suggests that the binding affinity of the molecules (**VIIi**–**VIIn**) increases with an increase in their molar refractivity.

### 3.2.4. 8-Substituted-benzo[*c*]quinolizin-3-ones

**3.2.4.1. Inhibition of human 5A-reductase 1 expressed by CHO 1827 cells by 8-substituted benzo[*c*]quinolizin-3-ones VIII.** Data from Ferrali et al.[33] (Tables 8a and b).



**VIII**

**Table 6a.** Biological and physicochemical constants used to derive QSAR equation 6a for the binding affinity to human CCR5 receptor by phenylacetamides (**VIa–VIag**)

| No. | R | X | log 1/*C* (Eq. 6a) | | | $\pi_{X\text{-}4}$ | $B1_{X\text{-}4}$ |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | $\Delta$ | | |
| **VIa** | $CH_3$ | H | 6.11 | 5.79 | 0.32 | 0.00 | 1.00 |
| **VIb** | $CH_3$ | 2-Cl | 5.44 | 5.79 | −0.35 | 0.00 | 1.00 |
| **VIc** | $CH_3$ | 3-Cl | 5.66 | 5.79 | −0.14 | 0.00 | 1.00 |
| **VId** | $CH_3$ | 4-Cl | 6.10 | 6.27 | −0.17 | 0.71 | 1.80 |
| **VIe** | $CH_3$ | 3,4-$Cl_2$ | 6.11 | 6.27 | −0.16 | 0.71 | 1.80 |
| **VIf** | $CH_3$ | 2-F | 5.72 | 5.79 | −0.07 | 0.00 | 1.00 |
| **VIg** | $CH_3$ | 3-F | 5.85 | 5.79 | 0.06 | 0.00 | 1.00 |
| **VIh** | $CH_3$ | 4-F | 6.18 | 6.06 | 0.12 | 0.14 | 1.35 |
| **VIi** | $CH_3$ | 3,4-$F_2$ | 6.16 | 6.06 | 0.10 | 0.14 | 1.35 |
| **VIj** | $CH_3$ | 3-$OCH_3$ | 6.17 | 5.79 | 0.37 | 0.00 | 1.00 |
| **VIk** | $CH_3$ | 4-$OCH_3$ | 6.24 | 6.12 | 0.12 | −0.02 | 1.35 |
| **VIl** | $CH_3$ | 3,4-$(OCH_3)_2$ | 6.19 | 6.12 | 0.07 | −0.02 | 1.35 |
| **VIm** | $CH_3$ | 3,5-$(OCH_3)_2$ | 5.57 | 5.79 | −0.22 | 0.00 | 1.00 |
| **VIn** | $CH_3$ | 2,4,5-$(OCH_3)_3$ | 5.96 | 6.12 | −0.16 | −0.02 | 1.35 |
| **VIo** | $CH_3$ | 4-Br | 6.24 | 6.35 | −0.12 | 0.86 | 1.95 |
| **VIp** | $CH_3$ | 4-$OCH_2C_6H_5$ | 5.46 | 5.51 | −0.06 | 1.66 | 1.35 |
| **VIq** | $CH_3$ | 4-$C_6H_5$ | 5.64 | 5.73 | −0.10 | 1.96 | 1.71 |
| **VIr** | $CH_3$ | 4-$CF_3$ | 6.43 | 6.38 | 0.05 | 0.88 | 1.99 |
| **VIs** | $CH_3$ | 4-$NHCOCH_3$ | 6.17 | 6.46 | −0.30 | −0.97 | 1.35 |
| **VIt** | $CH_3$ | 4-$SO_2NH_2$ | 7.05 | 7.40 | −0.36 | −1.82 | 2.04 |
| **VIu** | $CH_3$ | 4-$SO_2N(CH_3)_2$ | 7.34 | 7.02 | 0.32 | −0.78 | 2.03 |
| **VIv** | $CH_3$ | 4-$SCH_3$ | 6.25 | 6.21 | 0.04 | 0.61 | 1.70 |
| **VIw** | $CH_3$ | 4-$COOCH_3$ | 6.20 | 6.38 | −0.18 | −0.01 | 1.64 |
| **VIx** | $CH_3$ | 4-OH | 6.33 | 6.36 | −0.03 | −0.67 | 1.35 |
| **VIy** | $CH_3$ | 4-$NO_2$ | 6.82 | 6.54 | 0.29 | −0.28 | 1.70 |
| **VIz** | $C_2H_5$ | 4-$SO_2NH_2$ | 7.42 | 7.40 | 0.02 | −1.82 | 2.04 |
| **VIaa** | $C_2H_5$ | 4-$SO_2CH_3$ | 7.12 | 7.33 | −0.21 | −1.63 | 2.03 |
| **VIab** | $C_2H_5$ | 4-$NO_2$ | 6.96 | 6.54 | 0.42 | −0.28 | 1.70 |
| **VIac** | Cy-$C_3H_5$ | 4-$SO_2NH_2$ | 7.48 | 7.40 | 0.08 | −1.82 | 2.04 |
| **VIad** | Cy-$C_3H_5$ | 4-$SO_2CH_3$ | 7.29 | 7.33 | −0.03 | −1.63 | 2.03 |
| **VIae** | Cy-$C_3H_5$ | 4-$NO_2$ | 6.51 | 6.54 | −0.03 | −0.28 | 1.70 |
| **VIaf** | Allyl | 4-$SO_2CH_3$ | 7.43 | 7.33 | 0.11 | −1.63 | 2.03 |
| **VIag** | Allyl | 4-$NO_2$ | 6.74 | 6.54 | 0.21 | −0.28 | 1.70 |

**Table 6b.** Biological and physicochemical constants used to derive QSAR equation 6b for the binding affinity to human CCR5 receptor by phenylacetamides (**VIah–VIal**)

| No. | R | X | log 1/*C* (Eq. 6b) | | | $\pi_{X\text{-}4}$ |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | $\Delta$ | |
| **VIah** | $CH_3$ | 4-$OCF_3$ | 6.54 | 6.50 | 0.04 | 1.04 |
| **VIai** | $CH_3$ | 4-CN | 7.22 | 7.20 | 0.02 | −0.57 |
| **VIaj** | $C_2H_5$ | 4-$OCF_3$ | 6.51 | 6.50 | 0.01 | 1.04 |
| **VIak** | $C_2H_5$ | 4-CN | 7.18 | 7.20 | −0.02 | −0.57 |
| **VIal** | Allyl | 4-$OCF_3$ | 6.46 | 6.50 | −0.04 | 1.04 |

This data set was divided into two sub data sets according to Figure 1, which gave QSAR equations 8a and 8b.

$$\log 1/C = 3.32(\pm 0.55)C\log P - 0.31(\pm 0.06)C\log P^2 - 0.15(\pm 0.06)\mathrm{CMR} - 1.27(\pm 0.43)$$
$$n = 17, \quad r^2 = 0.944, \quad q^2 = 0.874, \quad s = 0.120 \quad (8a)$$

optimum $C\log P = 5.31\,(5.17\text{–}5.49)$

**Table 7a.** Biological and physicochemical constants used to derive QSAR equation 7a for the binding affinity to human CCR5 receptor by ureas (**VIIa–VIIh**)

| No. | $R_1$ | $R_2$ | log 1/*C* (Eq. 7a) | | | MgVol |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | $\Delta$ | |
| **VIIa** | $CH_3$ | Cy-$C_6H_{11}$ | 5.74 | 5.47 | 0.27 | 3.68 |
| **VIIb** | $CH_3$ | 3-CN–$C_6H_4$ | 5.31 | 5.55 | −0.24 | 3.70 |
| **VIIc** | $CH_3$ | 3-$CH_3$–$C_6H_4$ | 5.23 | 5.51 | −0.28 | 3.69 |
| **VIId** | Allyl | $C_6H_5$ | 5.68 | 5.83 | −0.15 | 3.79 |
| **VIIe** | $C_2H_5$ | 4-$CH_3$–$C_6H_4$ | 6.49 | 5.97 | 0.53 | 3.83 |
| **VIIf** | Allyl | $CH_2C_6H_4$(3-$CH_3$) | 6.62 | 6.74 | −0.12 | 4.07 |
| **VIIg** | Allyl | $CH_2C_6H_4$(4-$OCH_3$) | 6.96 | 6.93 | 0.02 | 4.13 |
| **VIIh** | $C_2H_5$ | $CH_2C_6H_4$(4-$SO_2CH_3$) | 7.31 | 7.34 | −0.03 | 4.25 |

**Table 7b.** Biological and physicochemical constants used to derive QSAR equation 7b for the binding affinity to human CCR5 receptor by ureas (**VIIi–VIIn**)

| No. | $R_1$ | $R_2$ | log 1/*C* (Eq. 7b) | | | CMR |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | |
| **VIIi** | $CH_3$ | 3,4-$Cl_2$–$C_6H_3$ | 6.43[a] | 7.20 | −0.77 | 14.30 |
| **VIIj** | $CH_3$ | 4-F–$C_6H_4$ | 6.72 | 6.76 | −0.04 | 13.33 |
| **VIIk** | $CH_3$ | $CH_2C_6H_4$ | 7.00 | 6.97 | 0.03 | 13.78 |
| **VIIl** | $C_2H_5$ | $CH_2C_6H_4$ | 7.21 | 7.18 | 0.03 | 14.24 |
| **VIIm** | $C_2H_5$ | $CH_2C_6H_4$(3-$CH_3$) | 7.46 | 7.39 | 0.06 | 14.70 |
| **VIIn** | $C_2H_5$ | $CH_2C_6H_4$(4-$OCH_3$) | 7.38 | 7.46 | −0.08 | 14.86 |

[a] Data point not included in the derivation of Eq. 7b.

**Table 8a.** Biological and physicochemical constants used to derive QSAR equation 8a for the inhibition of human 5A-reductase 1 expressed by CHO 1827 cells by 8-substituted benzo[*c*]quinolizin-3-ones (**VIIIa–VIIIq**)

| No. | X | log 1/*C* (Eq. 8a) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | | |
| **VIIIa** | $C_2H_5$ | 6.66 | 6.55 | 0.11 | 3.85 | 8.08 |
| **VIIIb** | $CH_2C_6H_5$ | 7.01 | 6.91 | 0.10 | 5.15 | 10.13 |
| **VIIIc** | $C_3H_6OH$ | 5.16 | 5.20 | −0.05 | 2.83 | 8.70 |
| **VIIId** | $C_6H_4$(4-$CH_3$) | 6.99 | 6.92 | 0.07 | 5.33 | 10.13 |
| **VIIIe** | $C_6H_4$(4-*t*-$C_4H_9$) | 6.25 | 6.15 | 0.10 | 6.66 | 11.52 |
| **VIIIf** | $C_6H_4$(4-$OCH_3$) | 6.94 | 6.83 | 0.10 | 4.86 | 10.28 |
| **VIIIg** | $C_6H_4$(4-$OC_2H_5$) | 6.68 | 6.83 | −0.15 | 5.39 | 10.74 |
| **VIIIh** | $C_6H_4$(4-F) | 7.03 | 6.97 | 0.06 | 5.11 | 9.68 |
| **VIIIi** | $C_6H_4$(3-F) | 6.80 | 6.97 | −0.17 | 5.11 | 9.68 |
| **VIIIj** | $C_6H_4$(2-F) | 6.86 | 6.93 | −0.07 | 4.91 | 9.68 |
| **VIIIk** | $C_6H_4$(4-$CF_3$) | 6.83 | 6.78 | 0.05 | 5.96 | 10.17 |
| **VIIIl** | $C_6H_4$(3-$CF_3$) | 6.57 | 6.78 | −0.21 | 5.96 | 10.17 |
| **VIIIm** | $C_6H_4$(4-$NHCOCH_3$) | 6.15 | 6.16 | −0.01 | 3.90 | 11.00 |
| **VIIIn** | $C_6H_4$(4-$COOCH_3$) | 6.89 | 6.80 | 0.09 | 5.05 | 10.78 |
| **VIIIo** | $C_6H_4$(4-$CONHC_3H_7$) | 6.57 | 6.58 | 0.00 | 4.80 | 11.92 |
| **VIIIp** | $C_6H_4$[4-$CON(C_2H_5)_2$] | 6.33 | 6.44 | −0.10 | 4.62 | 12.39 |
| **VIIIq** | $C_6H_4$(4-CONH-*t*-$C_4H_9$) | 6.65 | 6.55 | 0.09 | 4.98 | 12.39 |

**Table 8b.** Biological and physicochemical constants used to derive QSAR equation 8b for the inhibition of human 5A-reductase 1 expressed by CHO 1827 cells by 8-substituted benzo[*c*]quinolizin-3-ones (**VIIIr–VIIIy**)

| No. | X | log 1/*C* (Eq. 8b) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **VIIIr** | $CH_3$ | 6.93 | 6.82 | 0.12 | 3.32 |
| **VIIIs** | $CH(CH_3)_2$ | 6.14 | 6.14 | −0.01 | 4.16 |
| **VIIIt** | $C(CH_3)_3$ | 6.00 | 5.97 | 0.03 | 4.56 |
| **VIIIu** | $C_6H_5$ | 6.03 | 5.97 | 0.05 | 4.83 |
| **VIIIv** | $C_6H_4$(2-$CF_3$) | 7.38 | 7.37 | 0.00 | 5.96 |
| **VIIIw** | $C_6H_4$(4-$NH_2$) | 6.57 | 6.52 | 0.04 | 3.65 |
| **VIIIx** | $C_6H_4$(3-$COOCH_3$) | 6.02 | 6.07 | −0.05 | 5.05 |
| **VIIIy** | $C_6H_4$(4-$CONH_2$) | 6.44 | 6.63 | −0.18 | 3.53 |

$$\log 1/C = -0.97(\pm 0.30)C\log P + 3.43(\pm 0.86) \times \log(\beta \times 10^{C\log P} + 1) + 9.99(\pm 1.15)$$
$$n = 8, \quad r^2 = 0.969, \quad q^2 = 0.918, \quad s = 0.118 \qquad (8b)$$

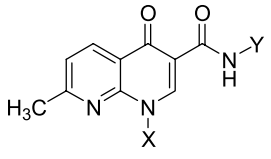
inversion point for $C\log P = 4.70$

$\log \beta = -5.10$

Eq. 8a is a parabolic correlation with *C*log *P*, followed by a linear correlation with negative CMR, whereas Eq. 8b is an allosteric bilinear correlation with *C*log *P*. Thus, the hydrophobic parameter of the molecules is the most important parameter for both Eqs. 8a and 8b.

### 3.2.5. 1,8-Naphthyridine derivatives

**3.2.5.1. Binding affinity of 1,8-naphthyridine derivatives IX to cannabinoid receptor (CB2).** Data from Ferrarini et al.[34] (Tables 9a and b).



**IX**

We derived two QSAR equations 9a and 9b from the data of Ferrarini et al.[34] (Fig. 1)

**Table 9a.** Biological and physicochemical constants used to derive QSAR equation 9a for the binding affinity of 1,8-naphthyridine derivatives (**IXa–IXp**) to cannabinoid receptor (CB2)

| No. | X | Y | $\log 1/K_i$ (Eq. 9a) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | Δ | | |
| **IXa** | H | Cyclohexyl | 6.00 | 5.77 | 0.23 | 1.49 | 8.10 |
| **IXb** | H | Benzyl | 6.00 | 6.28 | −0.28 | 1.42 | 8.47 |
| **IXc** | Ethylmorph | Cyclohexyl | 7.00 | 6.86 | 0.14 | 1.40 | 11.23 |
| **IXd** | Ethylmorph | Morph | 6.00 | 6.29 | −0.29 | −1.00 | 10.92 |
| **IXe** | Ethylmorph | CH2-cyclohexyl | 6.93 | 6.58 | 0.35 | 2.02 | 11.70 |
| **IXf** | Ethylmorph | *N*-$CH_3$-Pipz | 6.00 | 6.14 | −0.14 | 0.05 | 11.51 |
| **IXg** | Ethylmorph | Benzyl | 6.32 | 6.46 | −0.14 | 1.33 | 11.60 |
| **IXh** | Ethylmorph | Cyclopentyl | 7.30 | 6.99 | 0.31 | 0.84 | 10.77 |
| **IXi** | Ethylmorph | Isopentyl | 7.30 | 7.06 | 0.24 | 1.35 | 10.95 |
| **IXj** | Ethylmorph | 4-Cl-benzyl | 6.00 | 6.04 | −0.04 | 2.05 | 12.09 |
| **IXk** | Benzyl | Cyclohexyl | 8.00 | 7.85 | 0.15 | 3.99 | 11.08 |
| **IXl** | 2-F-Benzyl | Cyclohexyl | 7.36 | 7.89 | −0.53 | 4.13 | 11.10 |
| **IXm** | 4-F-Benzyl | Cyclohexyl | 8.26 | 7.89 | 0.37 | 4.13 | 11.10 |
| **IXn** | 4-F-Benzyl | Benzyl | 7.19 | 7.53 | −0.35 | 4.07 | 11.47 |
| **IXo** | *n*-Hexyl | Cyclohexyl | 8.10 | 8.01 | 0.09 | 4.08 | 10.89 |
| **IXp** | *n*-Butyl | Cyclohexyl | 7.76 | 7.87 | −0.11 | 3.02 | 9.96 |

**Table 9b.** Biological and physicochemical constants used to derive QSAR equation 9b for the binding affinity of 1,8-naphthyridine derivatives (**IXq–IXw**) to cannabinoid receptor (CB2)

| No. | X | Y | $\log 1/K_i$ (Eq. 9b) | | | *C*log *P* |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | Δ | |
| **IXq** | Ethylmorph | 4-Me-cyclohexyl | 7.52 | 7.58 | −0.06 | 1.92 |
| **IXr** | Ethylmorph | Cycloheptyl | 7.66 | 7.55 | 0.10 | 1.96 |
| **IXs** | Benzyl | Benzyl | 6.00 | 6.34 | −0.34 | 3.93 |
| **IXt** | Benzyl | 4-Cl-benzyl | 6.00 | 5.89 | 0.11 | 4.64 |
| **IXu** | 2-F-Benzyl | Benzyl | 6.22 | 6.25 | −0.02 | 4.07 |
| **IXv** | *n*-Hexyl | Benzyl | 6.49 | 6.28 | 0.21 | 4.01 |
| **IXw** | *n*-Butyl | Benzyl | 6.00[a] | 6.94 | −0.94 | 2.95 |

[a] Data point not included in the derivation of Eq. 9b.

$$\log 1/K_i = 0.34(\pm 0.11)C\log P + 7.93(\pm 2.98)\mathrm{CMR} - 0.39(\pm 0.15)\mathrm{CMR}^2 - 33.30(\pm 14.70)$$

$$n = 16, \quad r^2 = 0.888, \quad q^2 = 0.785, \quad s = 0.308 \quad (9a)$$

optimum CMR = 10.13 (9.92–10.37)

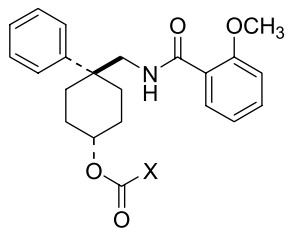
$$\log 1/K_i = -0.62(\pm 0.22)C\log P + 8.77(\pm 0.80)$$

$$n = 6, \quad r^2 = 0.936, \quad q^2 = 0.884, \quad s = 0.213 \quad (9b)$$

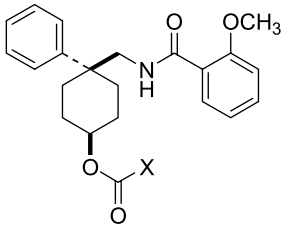
outlier: **IXw**

### 3.2.6. Benzamide derivatives

**3.2.6.1. Inhibition of Kv1.3 (voltage-gated potassium channel) in CHO cells by benzamide derivatives (XA and XB).** Data from Miao et al.[35] (Tables 10a and b).



**XA (*trans*)**



**XB (*cis*)**

QSAR equations 10a and 10b were derived from this data set on the basis of the outliers (Fig. 1).

$$\log 1/C = 1.77(\pm 0.55)C\log P - 0.17(\pm 0.06)C\log P^2 - 0.32(\pm 0.10)\mathrm{CMR} + 6.49(\pm 1.80)$$

$$n = 20, \quad r^2 = 0.898, \quad q^2 = 0.845, \quad s = 0.094 \quad (10a)$$

optimum *C*log *P* = 5.17 (4.83–5.75)

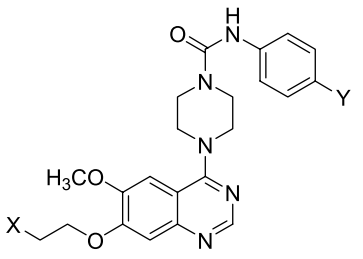
$$\log 1/C = -0.57(\pm 0.24)\mathrm{CMR} + 13.78(\pm 3.05)$$

$$n = 8, \quad r^2 = 0.848, \quad q^2 = 0.732, \quad s = 0.186 \quad (10b)$$

outlier: **XBl**

### 3.2.7. 4-Piperazinylquinazoline derivatives

**3.2.7.1. Inhibition of PDGFR (platelet-derived growth factor receptor) phosphorylation in MG63 cells and in the presence of plasma by 4-piperazinylquinazolines XI.** Data from Pandey et al.[36] (Tables 11a and b).



**XI**

We derived the following two QSAR equations 11a and 11b from the data of Pandey et al.[36]

**Table 10a.** Biological and physicochemical constants used to derive QSAR equation 10a for the inhibition of Kv1.3 (voltage-gated potassium channel) in CHO cells by benzamide derivatives (**XAa**–**XAi** and **XBa**–**XBk**)

| No. | X | log 1/*C* (Eq. 10a) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | | |
| **XAa** | $NHCH_3$(*trans*) | 7.24 | 7.11 | 0.12 | 3.72 | 11.24 |
| **XAb** | $NHCH(CH_3)_2$(*trans*) | 7.12 | 7.11 | 0.01 | 4.56 | 12.17 |
| **XAc** | $NHCH_2CH_2CH_3$(*trans*) | 7.30 | 7.15 | 0.15 | 4.78 | 12.17 |
| **XAd** | $NHCH_2CH{=}CH_2$(*trans*) | 7.15 | 7.11 | 0.05 | 4.49 | 12.14 |
| **XAe** | NH(2-$CH_3$-allyl)(*trans*) | 7.05 | 7.02 | 0.02 | 4.89 | 12.61 |
| **XAf** | $NHCH_2CH_2CH_2CH_3$(*trans*) | 7.05 | 7.03 | 0.02 | 5.31 | 12.63 |
| **XAg** | $NH(CH_2)_2OH$(*trans*) | 6.51 | 6.58 | −0.07 | 3.15 | 11.86 |
| **XAh** | $NH(CH_2)_2OCH_3$(*trans*) | 6.85 | 6.86 | −0.01 | 3.92 | 12.32 |
| **XAi** | $NH(CH_2)_3OCH_3$(*trans*) | 6.84 | 6.83 | 0.01 | 4.23 | 12.78 |
| **XBa** | $NH_2$(*cis*) | 6.92 | 6.94 | −0.02 | 3.17 | 10.78 |
| **XBb** | $NHCH_2CH_2CH_3$(*cis*) | 7.00 | 7.15 | −0.15 | 4.78 | 12.17 |
| **XBc** | $NHCH_2C(CH_3){=}CH_2$(*cis*) | 6.92 | 7.02 | −0.10 | 4.89 | 12.61 |
| **XBd** | $NHCH_2CH_2CH_2CH_3$(*cis*) | 6.85 | 7.03 | −0.17 | 5.31 | 12.63 |
| **XBe** | $NH(CH_2)_2OCH_3$(*cis*) | 6.85 | 6.86 | −0.01 | 3.92 | 12.32 |
| **XBf** | $NH(CH_2)_3OCH_3$(*cis*) | 6.89 | 6.83 | 0.06 | 4.23 | 12.78 |
| **XBg** | $NHC_6H_5$(*cis*) | 6.92 | 6.80 | 0.12 | 5.49 | 13.29 |
| **XBh** | $NHCH_2C_6H_5$(*cis*) | 6.61 | 6.65 | −0.04 | 5.49 | 13.75 |
| **XBi** | $NH(CH_2)_2C_6H_5$(*cis*) | 6.38 | 6.45 | 0.07 | 5.82 | 14.21 |
| **XBj** | $NH(CH_2)_3C_6H_5$(*cis*) | 6.27 | 6.19 | 0.08 | 6.20 | 14.68 |
| **XBk** | $N(CH_2CH{=}CH_2)_2$(*cis*) | 6.69 | 6.67 | 0.02 | 5.86 | 13.51 |

**Table 10b.** Biological and physicochemical constants used to derive QSAR equation 10b for the inhibition of Kv1.3 (voltage-gated potassium channel) in CHO cells by benzamide derivatives (**XAj**–**XAl** and **XBl**–**XBq**)

| No. | X | log 1/*C* (Eq. 10b) | | | CMR |
|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | |
| **XAj** | $N(CH_3)CH_2CH{=}CH_2$(*trans*) | 6.49 | 6.56 | −0.06 | 12.61 |
| **XAk** | $NHCH_2C_6H_5$(*trans*) | 5.77 | 5.90 | −0.13 | 13.75 |
| **XAl** | $N(CH_2CH{=}CH_2)_2$(*trans*) | 6.08 | 6.04 | 0.04 | 13.51 |
| **XBl** | $NHCH_3$(*cis*) | 6.82[a] | 7.34 | −0.51 | 11.24 |
| **XBm** | $NHCH(CH_3)_2$(*cis*) | 6.89 | 6.81 | 0.08 | 12.17 |
| **XBn** | $NHCH_2CH{=}CH_2$(*cis*) | 6.89 | 6.82 | 0.06 | 12.14 |
| **XBo** | $N(CH_3)CH_2CH{=}CH_2$(*cis*) | 6.84 | 6.56 | 0.28 | 12.61 |
| **XBp** | $NH(CH_2)_2OH$(*cis*) | 7.02 | 6.99 | 0.04 | 11.86 |
| **XBq** | $OCH_2CH{=}CH_2$(*cis*) | 6.64 | 6.95 | −0.31 | 11.93 |

[a] Data point not included in the derivation of Eq. 10b.

**Table 11a.** Biological and physicochemical constants used to derive QSAR equation 11a for the inhibition of PDGFR (platelet-derived growth factor receptor) phosphorylation in MG63 cells, and in the presence of plasma by 4-piperazinylquinazolines **XIa**–**XIp**

| No. | X | Y | log 1/*C* (Eq. 11a) | | | *C*log *P* |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | |
| **XIa** | $OCH_3$ | CN | 6.10 | 5.97 | 0.13 | 2.92 |
| **XIb** | Morpholin-1-yl | CN | 5.90 | 6.14 | −0.24 | 3.15 |
| **XIc** | Piperidin-1-yl | CN | 7.01 | 7.04 | −0.03 | 4.36 |
| **XId** | 4-OH-piperidin-1-yl | CN | 5.49 | 5.49 | 0.00 | 2.28 |
| **XIe** | 4-Oxo-piperidin-1-yl | CN | 6.22 | 6.09 | 0.13 | 3.09 |
| **XIf** | Thiomorpholin-1-yl | CN | 6.52 | 6.69 | −0.17 | 3.89 |
| **XIg** | Pyrrolidin-1-yl | CN | 6.69 | 6.63 | 0.06 | 3.81 |
| **XIh** | $OCH_3$ | $OCH(CH_3)_2$ | 6.67 | 6.57 | 0.10 | 3.73 |
| **XIi** | Morpholin-1-yl | $OCH(CH_3)_2$ | 6.75 | 6.74 | 0.01 | 3.96 |
| **XIj** | 1,2,3-Triazol-1-yl | $OCH(CH_3)_2$ | 6.42 | 6.21 | 0.21 | 3.24 |
| **XIk** | 11,2,3-Triazol-2-yl | $OCH(CH_3)_2$ | 6.28 | 6.49 | −0.20 | 3.62 |
| **XIl** | 1,2,3-Triazol-1-yl | CN | 5.33 | 5.61 | −0.28 | 2.44 |
| **XIm** | 1,2,3-Triazol-2-yl | CN | 5.66 | 5.89 | −0.23 | 2.81 |
| **XIn** | 1,1-Di-oxide-thiomorpholin-1-yl | CN | 5.53 | 5.42 | 0.11 | 2.18 |
| **XIo** | 1,1-Di-oxide-thiomorpholin-1-yl | $OCH(CH_3)_2$ | 6.17 | 6.02 | 0.15 | 2.99 |
| **XIp** | Tetrazol-2-yl | $OCH(CH_3)_2$ | 6.46 | 6.21 | 0.25 | 3.25 |

**Table 11b.** Biological and physicochemical constants used to derive QSAR equation 11b for the inhibition of PDGFR (platelet-derived growth factor receptor) phosphorylation in MG63 cells, and in the presence of plasma by 4-piperazinylquinazolines **XIq–XIw**

| No. | X | Y | log 1/*C* (Eq. 11b) | | | CMR |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | |
| **XIq** | 1,4-Dioxa-8-aza-spiro[4.5]dec-8-yl | CN | 6.65 | 6.71 | −0.05 | 15.49 |
| **XIr** | Piperazin-1-yl | CN | 5.11[a] | 5.86 | −0.76 | 14.34 |
| **XIs** | 4-$CH_3$-piperazin-1-yl | CN | 6.18 | 6.20 | −0.02 | 14.80 |
| **XIt** | Imidazol-1-yl | CN | 5.29 | 5.36 | −0.08 | 13.65 |
| **XIu** | Pyrrol-1-yl | CN | 5.60 | 5.52 | 0.09 | 13.87 |
| **XIv** | Piperidin-1-yl | $OCH(CH_3)_2$ | 6.77 | 6.72 | 0.05 | 15.50 |
| **XIw** | Tetrazol-1-yl | $OCH(CH_3)_2$ | 5.84 | 5.83 | 0.01 | 14.30 |

[a] Data point not included in the derivation of Eq. 11b.

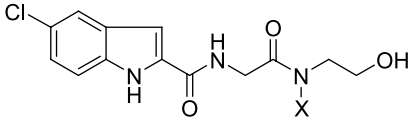
$$\log 1/C = 0.75(\pm 0.16)C\log P + 3.79(\pm 0.52)$$
$$n = 16, \quad r^2 = 0.881, \quad q^2 = 0.852, \quad s = 0.178 \quad (11a)$$

$$\log 1/C = 0.73(\pm 0.11)\mathrm{CMR} - 4.65(\pm 1.60)$$
$$n = 6, \quad r^2 = 0.989, \quad q^2 = 0.969, \quad s = 0.070 \quad (11b)$$

outlier: **XIr**

### 3.2.8. 5-Chloroindolyl derivatives

**3.2.8.1. Inhibition of glycogen phosphorylase A (GPA, EC 2.4.1.1) by 5-chloroindolyl derivatives XII.** Data from Wright et al.[28] (Tables 12a and b).




**XII**

The data of Wright et al.[28] were divided into two sub data sets on the basis of outliers (Fig. 1), resulting in two QSAR equations 12a and 12b.

$$\log 1/C = 0.62(\pm 0.13)C\log P + 4.10(\pm 1.10)\mathrm{CMR} - 0.23(\pm 0.06)\mathrm{CMR}^2 - 13.03(\pm 5.60)$$
$$n = 20, \quad r^2 = 0.881, \quad q^2 = 0.819, \quad s = 0.172 \quad (12a)$$

optimum CMR = 8.92 (8.48–9.25)

$$\log 1/C = -1.52(\pm 0.39)C\log P + 0.38(\pm 0.10)C\log P^2 + 7.74(\pm 0.32)$$
$$n = 7, \quad r^2 = 0.967, \quad q^2 = 0.926, \quad s = 0.131 \quad (12b)$$

inversion point for $C\log P$ = 2.00 (1.85–2.18)

outlier: **XIIw**

### 3.2.9. 6-Substituted-9-(4-methylbenzyl)-2-trifluoromethyl-9*H*-purines

**3.2.9.1. Plaque inhibition in M-HeLa cells by 6-X-9-(4-methylbenzyl)-2-trifluoromethyl-9*H*-purines XIII.** Data from Kelley et al.[37] (Tables 13a and b).

**Table 12a.** Biological and physicochemical constants used to derive QSAR equation 12a for the inhibition of glycogen phosphorylase A (GPA) by 5-chloroindolyl derivatives (**XIIa–XIIt**)

| No. | X | log 1/*C* (Eq. 12a) | | | *C*log *P* | CMR |
|---|---|---|---|---|---|---|
| | | Observed | Calculated | *Δ* | | |
| **XIIa** | Cyclopentyl | 7.24 | 7.04 | 0.20 | 3.11 | 9.72 |
| **XIIb** | Cyclononyl | 7.09 | 7.08 | 0.01 | 5.34 | 11.47 |
| **XIIc** | Cycloheptyl | 7.52 | 7.19 | 0.33 | 4.23 | 10.65 |
| **XIId** | Cyclooctyl | 7.49 | 7.23 | 0.27 | 4.79 | 11.01 |
| **XIIe** | Cyclohexyl | 7.00 | 7.17 | −0.16 | 3.67 | 10.18 |
| **XIIf** | Cyclodecyl | 6.68 | 6.83 | −0.15 | 5.90 | 11.93 |
| **XIIg** | Phenyl | 6.92 | 6.83 | 0.09 | 3.04 | 10.09 |
| **XIIh** | *n*-Butyl | 6.92 | 7.20 | −0.28 | 3.22 | 9.43 |
| **XIIi** | Cyclobutyl | 6.89 | 6.81 | 0.07 | 2.55 | 9.26 |
| **XIIj** | *n*-Propyl | 6.82 | 6.93 | −0.11 | 2.69 | 8.97 |
| **XIIk** | Cyclopropylmethyl | 6.80 | 6.85 | −0.05 | 2.61 | 9.30 |
| **XIIl** | Cyclopropyl | 6.70 | 6.64 | 0.06 | 2.22 | 8.83 |
| **XIIm** | *i*-Propyl | 6.68 | 6.79 | −0.12 | 2.47 | 8.97 |
| **XIIn** | (2-Furyl)methyl | 6.59 | 6.62 | −0.04 | 2.46 | 9.77 |
| **XIIo** | Benzyl | 6.49 | 6.68 | −0.19 | 3.29 | 10.55 |
| **XIIp** | 3-Tetrahydrothiophenyl | 6.47 | 6.60 | −0.14 | 2.65 | 10.06 |
| **XIIq** | Methyl | 6.24 | 6.10 | 0.15 | 1.64 | 8.04 |
| **XIIr** | Cyclododecyl | 6.02 | 6.04 | −0.02 | 7.02 | 12.86 |
| **XIIs** | (3-Pyridyl)methyl | 5.96 | 5.90 | 0.06 | 1.79 | 10.34 |
| **XIIt** | 3-Sulfolanyl | 5.89 | 5.87 | 0.01 | 2.50 | 10.94 |

**Table 12b.** Biological and physicochemical constants used to derive QSAR equation 12b for the inhibition of glycogen phosphorylase A (GPA) by 5-chloroindolyl derivatives (**XIIu–XIIab**)

| No. | X | log 1/*C* (Eq. 12b) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | Δ | |
| **XIIu** | $CH[(CH_2)_2]_2SO_2$ | 7.92 | 7.92 | 0.00 | −0.11 |
| **XIIv** | 1-(2,3-Dihydro)indanyl | 7.31 | 7.33 | −0.02 | 3.70 |
| **XIIw** | 3-Tetrahydrofuryl | 7.26[a] | 6.22 | 1.04 | 1.98 |
| **XIIx** | *i*-Butyl | 6.74 | 6.68 | 0.07 | 3.09 |
| **XIIy** | $CH[(CH_2)_2]_2O$ | 6.60 | 6.43 | 0.18 | 1.27 |
| **XIIz** | *t*-Butyl | 6.54 | 6.51 | 0.02 | 2.87 |
| **XIIaa** | (2-Imidazolyl)methyl | 6.37 | 6.52 | −0.15 | 1.12 |
| **XIIab** | (2-THF)methyl | 6.26 | 6.36 | −0.10 | 2.60 |

[a] Data point not included in the derivation of Eq. 12b.

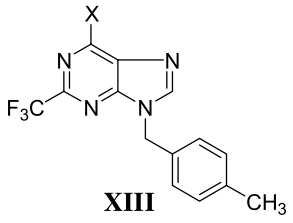
**Table 13a.** Biological and physicochemical constants used to derive QSAR equation 13a for the plaque inhibition in M-HeLa cells by 6-X-9-(4-methylbenzyl)-2-trifluoromethyl-9*H*-purines (**XIIIa–XIIIl**)

| No. | X | log 1/*C* (Eq. 13a) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | Δ | |
| **XIIIa** | $NHCH_3$ | 6.25 | 6.34 | −0.09 | 3.60 |
| **XIIIb** | H | 4.96 | 4.92 | 0.04 | 2.70 |
| **XIIIc** | $SCH_3$ | 6.22 | 6.57 | −0.35 | 3.91 |
| **XIIId** | $N(CH_3)CH_2CH_3$ | 7.00 | 6.66 | 0.34 | 4.18 |
| **XIIIe** | $N(CH_3)C_3H_5$-cyclo | 7.00 | 6.67 | 0.33 | 4.23 |
| **XIIIf** | $N(CH_3)CH_2CH_2CH_3$ | 6.52 | 6.56 | −0.04 | 4.70 |
| **XIIIg** | $N(CH_3)CH_2CH{=}CH_2$ | 6.52 | 6.66 | −0.14 | 4.42 |
| **XIIIh** | $N(CH_3)CH_2SCH_3$ | 6.52 | 6.55 | −0.03 | 3.88 |
| **XIIIi** | $N(CH_3)CH_2CH_2CH_2CH_3$ | 6.15 | 6.08 | 0.08 | 5.23 |
| **XIIIj** | $N(CH_3)C_6H_5$ | 5.72 | 5.82 | −0.10 | 5.42 |
| **XIIIk** | $N(CH_3)CH_2CH_2OH$ | 5.43 | 5.35 | 0.08 | 2.91 |
| **XIIIl** | $N(CH_3)CH_2COOCH_3$ | 6.05 | 6.16 | −0.11 | 3.43 |

**Table 13b.** Biological and physicochemical constants used to derive QSAR equation 13b for plaque inhibition in M-HeLa cells by 6-X-9-(4-methylbenzyl)-2-trifluromethyl-9*H*-purines (**XIIIm–XIIIr**)

| No. | X | log 1/*C* (Eq. 13b) | | | CMR |
|---|---|---|---|---|---|
| | | Observed | Calculated | Δ | |
| **XIIIm** | $N(CH_3)_2$ | 7.52 | 7.26 | 0.26 | 8.42 |
| **XIIIn** | $NH–C_3H_5$-cyclo | 5.45 | 5.80 | −0.36 | 8.74 |
| **XIIIo** | $N(CH_3)CH_2COOH$ | 4.68 | 4.35 | 0.34 | 9.07 |
| **XIIIp** | $N(CH_3)OH$ | 5.05[a] | 8.65 | -3.60 | 8.11 |
| **XIIIq** | $N(CH_3)OCH_3$ | 6.52 | 6.58 | −0.05 | 8.57 |
| **XIIIr** | $N(CH_3)C(O)CH_3$ | 4.84 | 5.03 | −0.19 | 8.92 |

[a] Data point not included in the derivation of Eq. 13b.




**XIII**

This data set was also divided into two sub data sets on the basis of outliers (see Fig. 1) to derive QSAR equations 13a and 13b.

$$\log 1/C = 5.86(\pm 1.61)C\log P - 0.68(\pm 0.20)C\log P^2 - 5.91(\pm 3.22)$$
$$n = 12, \quad r^2 = 0.897, \quad q^2 = 0.843, \quad s = 0.214 \quad (13a)$$

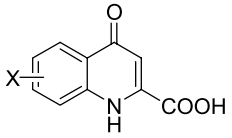
optimum $C\log P = 4.30\,(4.17–4.47)$

$$\log 1/C = -4.46(\pm 2.07)\mathrm{CMR} - 44.83(\pm 18.13)$$
$$n = 5, \quad r^2 = 0.940, \quad q^2 = 0.765, \quad s = 0.340 \quad (13b)$$

outlier: **XIIIp**

### 3.2.10. Kynurenic acid derivatives

**3.2.10.1. Apparent dissociation constant of X-kynurenic acid XIV with NMDA receptor.** Data from Leeson et al.[38] (Tables 14a and b).




**XIV**

From the data of Leeson et al.,[38] we derived the following two QSAR equations 14a and 14b on the basis of Figure 1:

**Table 14a.** Biological and physicochemical constants used to derive QSAR equation 14a for the apparent dissociation constant of X-kynurenic acid (**XIVa**–**XIVaa**) with NMDA receptor

| No. | X | log 1/*KB* (Eq. 14a) | | | $\sigma_X$ | $L_{X-5}$ | $B1_{X-5}$ | $B1_{X-7}$ | $B5_{X-7}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | Observed | Calculated | $\Delta$ | | | | | |
| **XIVa** | H | 3.81 | 3.80 | 0.01 | 0.00 | 2.06 | 1.00 | 1.00 | 1.00 |
| **XIVb** | 5-F | 4.39 | 4.0 | 0.38 | 0.34 | 2.65 | 1.35 | 1.00 | 1.00 |
| **XIVc** | 5-Cl | 4.43 | 4.54 | −0.11 | 0.37 | 3.52 | 1.80 | 1.00 | 1.00 |
| **XIVd** | 5-Br | 4.70 | 4.70 | 0.00 | 0.39 | 3.82 | 1.95 | 1.00 | 1.00 |
| **XIVe** | 5-I | 5.12 | 4.97 | 0.15 | 0.35 | 4.23 | 2.15 | 1.00 | 1.00 |
| **XIVf** | 5-$CF_3$ | 4.46 | 4.93 | −0.47 | 0.43 | 3.30 | 1.99 | 1.00 | 1.00 |
| **XIVg** | 5-Me | 4.68 | 4.56 | 0.12 | −0.07 | 2.87 | 1.52 | 1.00 | 1.00 |
| **XIVh** | 5-$C_3H_7$ | 3.40 | 3.84 | −0.45 | −0.06 | 4.92 | 1.52 | 1.00 | 1.00 |
| **XIVi** | 5-$C_4H_9$ | 3.31 | 3.43 | −0.12 | −0.08 | 6.17 | 1.52 | 1.00 | 1.00 |
| **XIVj** | 7-F | 4.49 | 4.20 | 0.30 | 0.34 | 2.06 | 1.00 | 1.35 | 1.35 |
| **XIVk** | 7-Cl | 5.15 | 5.02 | 0.14 | 0.37 | 2.06 | 1.00 | 1.80 | 1.80 |
| **XIVl** | 7-Br | 5.10 | 5.28 | −0.19 | 0.39 | 2.06 | 1.00 | 1.95 | 1.95 |
| **XIVm** | 7-Me | 4.47 | 4.48 | −0.01 | −0.07 | 2.06 | 1.00 | 1.52 | 2.04 |
| **XIVn** | 7-$C_2H_5$ | 3.82 | 3.74 | 0.09 | −0.07 | 2.06 | 1.00 | 1.52 | 3.17 |
| **XIVo** | 7-CH=$CH_2$ | 3.58 | 3.90 | −0.32 | 0.06 | 2.06 | 1.00 | 1.60 | 3.09 |
| **XIVp** | 7-$NO_2$ | 4.11 | 4.09 | 0.03 | 0.71 | 2.06 | 1.00 | 1.70 | 2.44 |
| **XIVq** | 5,7-$Cl_2$ | 5.52 | 5.76 | −0.23 | 0.74 | 3.52 | 1.80 | 1.80 | 1.80 |
| **XIVr** | 5,7-$Br_2$ | 5.80 | 6.19 | −0.39 | 0.78 | 3.82 | 1.95 | 1.95 | 1.95 |
| **XIVs** | 5-Br,7-Me | 5.68 | 5.39 | 0.29 | 0.32 | 3.82 | 1.95 | 1.52 | 2.04 |
| **XIVt** | 5-Br,7-$C_2H_5$ | 4.72 | 4.64 | 0.08 | 0.32 | 3.82 | 1.95 | 1.52 | 3.17 |
| **XIVu** | 5-Me,7-Br | 5.59 | 6.05 | −0.46 | 0.32 | 2.87 | 1.52 | 1.95 | 1.95 |
| **XIVv** | 5-$C_2H_5$,7-Br | 5.74 | 5.62 | 0.13 | 0.32 | 4.11 | 1.52 | 1.95 | 1.95 |
| **XIVw** | 5,7-$Me_2$ | 5.15 | 5.24 | −0.09 | −0.14 | 2.87 | 1.52 | 1.52 | 2.04 |
| **XIVx** | 5-I,7-Me | 6.00 | 5.66 | 0.34 | 0.28 | 4.23 | 2.15 | 1.52 | 2.04 |
| **XIVy** | 5-$C_2H_5$,7-Cl | 6.00 | 5.35 | 0.65 | 0.30 | 4.11 | 1.52 | 1.80 | 1.80 |
| **XIVz** | 5-Cl,7-$C_2H_5$ | 4.42 | 4.48 | −0.06 | 0.30 | 3.52 | 1.80 | 1.52 | 3.17 |
| **XIVaa** | 5-I,7-Cl | 6.39 | 6.19 | 0.20 | 0.72 | 4.23 | 2.15 | 1.80 | 1.80 |

**Table 14b.** Biological and physicochemical constants used to derive QSAR equation 14b for the apparent dissociation constant of X-kynurenic acid (**XIVab**–**XIVag**) with NMDA receptor

| No. | X | log 1/*KB* (Eq.14b) | | | *C*log *P* |
|---|---|---|---|---|---|
| | | Observed | Calculated | $\Delta$ | |
| **XIVab** | 5-$C_2H_5$ | 5.26[a] | 3.73 | 1.53 | 1.20 |
| **XIVac** | 7-I | 4.33 | 3.95 | 0.37 | 1.32 |
| **XIVad** | 7-$CF_3$ | 3.52 | 3.56 | −0.04 | 1.11 |
| **XIVae** | 7-SMe | 2.67 | 2.88 | −0.21 | 0.74 |
| **XIVaf** | 5-Me,7-I | 5.05 | 4.88 | 0.17 | 1.82 |
| **XIVag** | 5-Cl,7-I | 5.00 | 5.29 | −0.29 | 2.04 |

[a] Data point not included in the derivation of Eq. 14b.

$$\log 1/KB = 1.90(\pm 0.52)B1_{X\text{-}5} + 2.53(\pm 0.56)B1_{X\text{-}7} - 0.66(\pm 0.24)B5_{X\text{-}7} - 0.75(\pm 0.68)\sigma_X - 0.35(\pm 0.19)L_{X\text{-}5} + 0.85(\pm 0.74)$$
$$n = 27, \quad r^2 = 0.892, \quad q^2 = 0.831, \quad s = 0.308 \quad (14a)$$

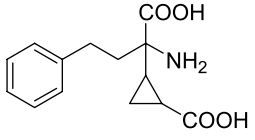
$$\log 1/KB = 1.85(\pm 0.95)C\log P + 1.51(\pm 1.40)$$
$$n = 5, \quad r^2 = 0.928, \quad q^2 = 0.756, \quad s = 0.315 \quad (14b)$$

outlier: **XIVab**

Sterimol parameters $B1_{X\text{-}5}$, $B1_{X\text{-}7}$, $B5_{X\text{-}7}$, and $L_{X\text{-}5}$ followed by $\sigma_X$, are the important parameters for QSAR 14a. On the other hand, Eq. 14b is a linear correlation with *C*log *P*.

### 3.2.11. Amino acids

**3.2.11.1. Displacement of [$^3$H]glutamate binding of compound XV to ACPD-sensitive glutamate receptor.** Data from Ornstein et al.[39] (Tables 15a and b).




**XV**

The data of Ornstein et al.[39] were divided into two sub data sets on the basis of outliers (see Fig. 1), which gave the following QSAR:

$$\log 1/C = 1.03(\pm 0.29)B1_{X\text{-}3} + 0.59(\pm 0.36)B1_{X\text{-}5} - 0.21(\pm 0.09)L_{X\text{-}3} - 0.28(\pm 0.14)MR_{X\text{-}4} + 0.79(\pm 0.19)I + 5.38(\pm 0.43)$$
$$n = 36, \quad r^2 = 0.869, \quad q^2 = 0.807, \quad s = 0.172 \quad (15a)$$

$$\log 1/C = 0.64(\pm 0.24)L_{X\text{-}3} + 4.59(\pm 0.72)$$
$$n = 7, \quad r^2 = 0.904, \quad q^2 = 0.812, \quad s = 0.200 \quad (15b)$$

outlier: **XVan**

Sterimol parameters $B1_{X\text{-}3}$, $B1_{X\text{-}5}$, and $L_{X\text{-}3}$ are the most important parameters for QSAR 15a, following the $MR_{X\text{-}4}$. The indicator variable *I* takes the value of 1

**Table 15a.** Biological and physicochemical constants used to derive QSAR equation 15a for the displacement of [$^3$H]glutamate binding of compounds (**XVa**–**XVaj**) to ACPD-sensitive glutamate receptor

| No. | X | log 1/*C* (Eq. 15a) | | | $L_{X\text{-}3}$ | $B1_{X\text{-}3}$ | $MR_{X\text{-}4}$ | $B1_{X\text{-}5}$ | *I* |
|---|---|---|---|---|---|---|---|---|---|
| | | Observed | Calculated | Δ | | | | | |
| **XVa** | H | 6.49 | 6.54 | −0.04 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVb** | 2-F | 6.51 | 6.54 | −0.03 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVc** | 3-F | 6.80 | 6.78 | 0.02 | 2.65 | 1.35 | 0.10 | 1.00 | 0 |
| **XVd** | 4-F | 7.66 | 7.33 | 0.33 | 2.06 | 1.00 | 0.09 | 1.00 | 1 |
| **XVe** | 2-Cl | 6.59 | 6.54 | 0.05 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVf** | 3-Cl | 7.00 | 7.06 | −0.06 | 3.52 | 1.80 | 0.10 | 1.00 | 0 |
| **XVg** | 4-Cl | 6.39 | 6.40 | −0.01 | 2.06 | 1.00 | 0.60 | 1.00 | 0 |
| **XVh** | 3-Br | 7.16 | 7.15 | 0.01 | 3.82 | 1.95 | 0.10 | 1.00 | 0 |
| **XVi** | 4-Br | 6.17 | 6.32 | −0.15 | 2.06 | 1.00 | 0.89 | 1.00 | 0 |
| **XVj** | 2-Me | 6.24 | 6.54 | −0.29 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVk** | 3-Me | 7.05 | 6.90 | 0.15 | 2.87 | 1.52 | 0.10 | 1.00 | 0 |
| **XVl** | 4-Me | 6.23 | 6.41 | −0.18 | 2.06 | 1.00 | 0.56 | 1.00 | 0 |
| **XVm** | 2-$CF_3$ | 6.32 | 6.54 | −0.22 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVn** | 3-$C_3H_7$ | 6.24 | 6.48 | −0.24 | 4.92 | 1.52 | 0.10 | 1.00 | 0 |
| **XVo** | 3-$C_6H_5$ | 6.29 | 6.40 | −0.11 | 6.28 | 1.71 | 0.10 | 1.00 | 0 |
| **XVp** | 4-$C_6H_5$ | 5.82 | 5.87 | −0.04 | 2.06 | 1.00 | 2.54 | 1.00 | 0 |
| **XVq** | 2-OMe | 6.77 | 6.54 | 0.23 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVr** | 2-OH | 6.66 | 6.54 | 0.12 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVs** | 3-OH | 6.82 | 6.76 | 0.07 | 2.74 | 1.35 | 0.10 | 1.00 | 0 |
| **XVt** | 4-OH | 6.52 | 6.49 | 0.04 | 2.06 | 1.00 | 0.29 | 1.00 | 0 |
| **XVu** | 2-$OC_6H_5$ | 6.44 | 6.54 | −0.09 | 2.06 | 1.00 | 0.10 | 1.00 | 0 |
| **XVv** | 3-$OC_6H_5$ | 6.66 | 6.39 | 0.26 | 4.51 | 1.35 | 0.10 | 1.00 | 0 |
| **XVw** | 3-$SO_2Me$ | 6.96 | 7.17 | −0.21 | 4.11 | 2.03 | 0.10 | 1.00 | 0 |
| **XVx** | 4-COOH | 6.43 | 6.37 | 0.06 | 2.06 | 1.00 | 0.69 | 1.00 | 0 |
| **XVy** | 3-$NH_2$ | 6.70 | 6.75 | −0.05 | 2.78 | 1.35 | 0.10 | 1.00 | 0 |
| **XVz** | 2,3-$F_2$ | 6.89 | 6.78 | 0.11 | 2.65 | 1.35 | 0.10 | 1.00 | 0 |
| **XVaa** | 2,4-$F_2$ | 7.20 | 7.33 | −0.13 | 2.06 | 1.00 | 0.09 | 1.00 | 1 |
| **XVab** | 2,5-$F_2$ | 6.54 | 6.74 | −0.20 | 2.06 | 1.00 | 0.10 | 1.35 | 0 |
| **XVac** | 3,4-$F_2$ | 7.47 | 7.57 | −0.10 | 2.65 | 1.35 | 0.09 | 1.00 | 1 |
| **XVad** | 3,5-$F_2$ | 7.68 | 7.77 | −0.09 | 2.65 | 1.35 | 0.10 | 1.35 | 1 |
| **XVae** | 2,4-$Cl_2$ | 6.66 | 6.40 | 0.26 | 2.06 | 1.00 | 0.60 | 1.00 | 0 |
| **XVaf** | 3,4-$Cl_2$ | 7.17 | 6.92 | 0.25 | 3.52 | 1.80 | 0.60 | 1.00 | 0 |
| **XVag** | 3,5-$Cl_2$ | 7.54 | 7.53 | 0.01 | 3.52 | 1.80 | 0.10 | 1.80 | 0 |
| **XVah** | 2,5-$(OMe)_2$ | 6.77 | 6.74 | 0.03 | 2.06 | 1.00 | 0.10 | 1.35 | 0 |
| **XVai** | 3,4-$(OMe)_2$ | 6.34 | 6.31 | 0.02 | 3.98 | 1.35 | 0.79 | 1.00 | 0 |
| **XVaj** | 3,5-$(OMe)_2$ | 6.96 | 6.71 | 0.25 | 3.98 | 1.35 | 0.10 | 1.35 | 0 |

**Table 15b.** Biological and physicochemical constants used to derive QSAR equation 15b for the displacement of [$^3$H]glutamate binding of compounds (**XVak**–**XVar**) to ACPD-sensitive glutamate receptor

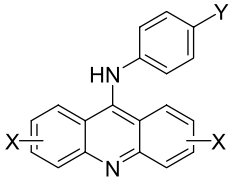
| No. | X | log 1/*C* (Eq. 15b) | | | $L_{X\text{-}3}$ |
|---|---|---|---|---|---|
| | | Observed | Calculated | Δ | |
| **XVak** | 3-$CF_3$ | 6.85 | 6.72 | 0.14 | 3.30 |
| **XVal** | 3-OMe | 7.03 | 7.16 | −0.13 | 3.98 |
| **XVam** | 4-OMe | 5.85 | 5.92 | −0.07 | 2.06 |
| **XVan** | 3-SMe | 6.39[a] | 7.36 | −0.98 | 4.30 |
| **XVao** | 3-SOMe | 5.62 | 5.92 | −0.30 | 2.06 |
| **XVap** | 3-COOH | 7.11 | 7.11 | 0.00 | 3.91 |
| **XVaq** | 2,3,4,5,6-$F_5$ | 6.43 | 6.30 | 0.13 | 2.65 |
| **XVar** | 2,6-$Cl_2$ | 6.15 | 5.92 | 0.23 | 2.06 |

[a] Data point not included in the derivation of Eq. 15b.

for X = 4-F or 3,5-$F_2$. On the other hand, Eq. 15b is a linear correlation with $L_{X\text{-}3}$.

### 3.2.12. 9-Anilinoacridine derivatives

**3.2.12.1. Association constant of 9-anilinoacridine derivatives (XVI) to DNA poly(dA–dT) in 0.01 M ionic strength buffer.** Data from Baguley et al.[40] (Tables 16a and b).




**XVI**

This data set was divided into two sub data sets according to Figure 1, which gave QSAR equations 16a and 16b.

$$\log K = 5.69(\pm 1.76)\mathrm{CMR} - 0.25(\pm 0.08)\mathrm{CMR}^2 + 0.40(\pm 0.17)I - 25.96(\pm 9.79)$$
$$n = 17, \quad r^2 = 0.873, \quad q^2 = 0.773, \quad s = 0.154 \quad (16a)$$

optimum CMR = 11.47(11.29–11.71)

$$\log K = -0.33(\pm 0.10)C\log P + 8.39(\pm 0.53)$$
$$n = 10, \quad r^2 = 0.873, \quad q^2 = 0.765, \quad s = 0.121 \quad (16b)$$

outliers: **XVIr**, **XVIv**

**Table 16a.** Biological and physicochemical constants used to derive QSAR equation 16a for the association constant of 9-anilinoacridine derivatives (**XVIa–XVIq**) to DNA poly(dA–dT) in 0.01 M ionic strength buffer

| No. | X | Y | log *K* (Eq. 16a) | | | CMR | *I* |
|---|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | | |
| **XVIa** | H | $NHCOCH_3$ | 6.30 | 6.16 | 0.14 | 10.07 | 0 |
| **XVIb** | 3-$NH_2$ | NHCOPH | 6.71 | 6.78 | −0.08 | 12.48 | 1 |
| **XVIc** | H | $NHSO_2CH_3$ | 6.18 | 6.38 | −0.21 | 10.44 | 0 |
| **XVId** | 3-$NH_2$ | $NHSO_2CH_3$ | 6.76 | 6.93 | −0.17 | 10.81 | 1 |
| **XVIe** | 3-$NHCOCH_3$ | $NHSO_2CH_3$ | 6.72 | 6.62 | 0.10 | 11.77 | 0 |
| **XVIf** | 3-$CH_3$ | $NHSO_2CH_3$ | 6.52 | 6.56 | −0.05 | 10.90 | 0 |
| **XVIg** | 3-$OCH_3$ | $NHSO_2CH_3$ | 6.57 | 6.60 | −0.03 | 11.06 | 0 |
| **XVIh** | 2-$NH_2$ | $NHSO_2CH_3$ | 6.76 | 6.54 | 0.23 | 10.81 | 0 |
| **XVIi** | 4-$NH_2$ | $NHSO_2CH_3$ | 6.71 | 6.54 | 0.17 | 10.81 | 0 |
| **XVIj** | 3,6-$(NH_2)_2$ | $NHSO_2CH_3$ | 7.02 | 7.02 | 0.00 | 11.18 | 1 |
| **XVIk** | H | $NHSO_2PH$ | 6.20 | 6.38 | −0.18 | 12.49 | 0 |
| **XVIl** | 3-$NH_2$ | $NHSO_2PH$ | 6.71 | 6.56 | 0.15 | 12.85 | 1 |
| **XVIm** | H | $SO_2NH_2$ | 5.96 | 6.09 | −0.13 | 9.97 | 0 |
| **XVIn** | 3-$NH_2$ | $SO_2NH_2$ | 6.61 | 6.73 | −0.11 | 10.34 | 1 |
| **XVIo** | 3,6-$(NH_2)_2$ | $SO_2CH_3$ | 6.99 | 6.93 | 0.06 | 10.81 | 1 |
| **XVIp** | H | $NO_2$ | 5.48 | 5.53 | −0.05 | 9.34 | 0 |
| **XVIq** | 3-$NH_2$ | $NO_2$ | 6.43 | 6.28 | 0.15 | 9.71 | 1 |

**Table 16b.** Biological and physicochemical constants used to derive QSAR equation 16b for the association constant of 9-anilinoacridine derivatives (**XVIr–XVIac**) to DNA poly(dA–dT) in 0.01 M ionic strength buffer

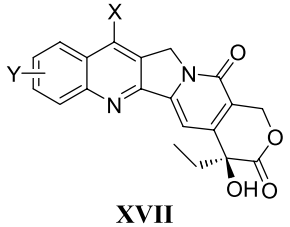
| No. | X | Y | log *K* (Eq. 16b) | | | *C*log *P* |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | *Δ* | |
| **XVIr** | H | H | 5.85[a] | 6.47 | −0.62 | 5.86 |
| **XVIs** | 3-$NH_2$ | H | 6.49 | 6.65 | −0.16 | 5.31 |
| **XVIt** | 3,6$(NH_2)_2$ | H | 6.89 | 6.83 | 0.06 | 4.76 |
| **XVIu** | H | $CH_3$ | 6.28 | 6.30 | −0.03 | 6.36 |
| **XVIv** | H | $NH_2$ | 6.30[a] | 6.87 | −0.57 | 4.64 |
| **XVIw** | 3-$NH_2$ | $NH_2$ | 7.08 | 7.05 | 0.03 | 4.08 |
| **XVIx** | 3,6-$(NH_2)_2$ | $NH_2$ | 7.06 | 7.23 | −0.17 | 3.53 |
| **XVIy** | H | NHCOPH | 6.18 | 6.30 | −0.13 | 6.37 |
| **XVIz** | 3-$NHCH_3$ | $NHSO_2CH_3$ | 6.87 | 6.78 | 0.09 | 4.90 |
| **XVIaa** | 3-$NHCOOCH_3$ | $NHSO_2CH_3$ | 7.00 | 6.87 | 0.13 | 4.65 |
| **XVIab** | 3-Cl | $NHSO_2CH_3$ | 6.63 | 6.59 | 0.04 | 5.48 |
| **XVIac** | 3-Br | $NHSO_2CH_3$ | 6.67 | 6.54 | 0.13 | 5.63 |

[a] Data points not included in the derivation of Eq. 16b.

Eq. 16a is a parabolic correlation with molar refractivity, which suggests that the activity of compounds (**XVIa–XVIq**) first increases with an increase in molar refractivity to an optimum CMR of 11.47 and then decreases. The indicator variable $I = 1$ for those compounds having X = 3-$NH_2$ group. Positive coefficient of *I* suggests that the presence of X = 3-$NH_2$ group in the molecules is favorable. Eq. 16b is a linear correlation with negative *C*log *P*, which suggests that the activity of compounds (**XVIr–XVIac**) decreases with the increase of their hydrophobicity. The results of Eqs. 16a and 16b suggest that these two subsets of compounds must be interacting DNA with two different mechanisms.

### 3.2.13. Camptothecin derivatives

**3.2.13.1. Inhibition of growth of PC-3 human prostate carcinoma cells by camptothecin derivatives XVII.** Data from Manikumar et al.[41] (Tables 17a and b).




**XVII**

This data set was also divided into two sub data sets according to Figure 1, which gave QSAR equations 17a and 17b with good statistics.

$$\log 1/C = 0.71(\pm 0.16)C\log P - 1.39(\pm 0.52) \times \log(\beta \times 10^{C\log P} + 1) + 6.18(\pm 0.33)$$
$$n = 11, \quad r^2 = 0.959, \quad q^2 = 0.920, \quad s = 0.099 \quad (17a)$$

optimum *C*log *P* = 3.38

**Table 17a.** Biological and physicochemical constants used to derive QSAR equation 17a for the inhibition of growth of PC-3 human prostate carcinoma cells by camptothecin derivatives (**XVIIa–XVIIk**)

| No. | X | Y | log 1/*C* (Eq. 17a) | | | *C*log *P* |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | Δ | |
| **XVIIa** | H | H | 6.85 | 6.81 | 0.04 | 0.90 |
| **XVIIb** | $C_2H_5$ | 10-OH | 7.45 | 7.55 | −0.10 | 1.97 |
| **XVIIc** | $C_2H_5$ | H | 7.45 | 7.52 | −0.07 | 1.93 |
| **XVIId** | $C_2H_5$ | 10-$OCH_2O$-11 | 8.00 | 7.84 | 0.16 | 2.44 |
| **XVIIe** | Cy-$C_5H_9$ | 10-$OCH_2O$-11 | 8.05 | 8.13 | −0.09 | 3.48 |
| **XVIIf** | Cy-$C_6H_{11}$ | 10-$OCH_2O$-11 | 7.94 | 7.98 | −0.04 | 4.04 |
| **XVIIg** | $C_6H_5$ | 10-$OCH_2O$-11 | 8.09 | 8.14 | −0.04 | 3.30 |
| **XVIIh** | 4-$CH_3$–$C_6H_4$ | 10-$OCH_2O$-11 | 8.08 | 8.07 | 0.01 | 3.80 |
| **XVIIi** | 4-Cl–$C_6H_4$ | 10-$OCH_2O$-11 | 8.09 | 7.99 | 0.10 | 4.02 |
| **XVIIj** | 4-F–$C_6H_4$ | 10-$OCH_2O$-11 | 8.20 | 8.14 | 0.06 | 3.45 |
| **XVIIk** | 4-$CF_3$–$C_6H_4$ | 10-$OCH_2O$-11 | 7.87 | 7.91 | −0.04 | 4.19 |

**Table 17b.** Biological and physicochemical constants used to derive QSAR equation 17b for the inhibition of growth of PC-3 human prostate carcinoma cells by camptothecin derivatives (**XVIIl–XVIIp**)

| No. | X | Y | log 1/*C* (Eq. 17b) | | | *C*log *P* |
|---|---|---|---|---|---|---|
| | | | Observed | Calculated | Δ | |
| **XVIIl** | *n*-$C_4H_9$ | H | 7.31 | 7.30 | 0.01 | 2.98 |
| **XVIIm** | *sec*-$C_4H_9$ | H | 6.77 | 6.88 | −0.11 | 2.85 |
| **XVIIn** | *tert*-$C_4H_9$ | H | 6.51 | 6.46 | 0.05 | 2.72 |
| **XVIIo** | $CH_2C_6H_5$ | H | 7.29 | 7.24 | 0.05 | 2.96 |
| **XVIIp** | H | 10-$OCH_2O$-11 | 8.09[a] | 2.22 | 5.87 | 1.42 |

[a] Data point not included in the derivation of Eq. 17b.

$$\log \beta = -3.37$$

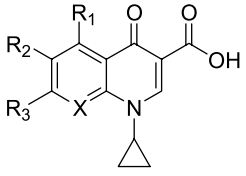
$$\log 1/C = 3.24(\pm 1.97)C\log P - 5.68(\pm 2.37)$$
$$n = 4, \quad r^2 = 0.962, \quad q^2 = 0.734, \quad s = 0.095 \quad (17b)$$

outlier: **XVIIp**

The hydrophobicity of the molecules correlates with their inhibitory activity in a bilinear fashion in Eq. 17a, whereas there is a linear correlation in Eq. 17b for the compounds (**XVIIa–XVIIk**) and (**XVIIl–XVIIp**), respectively.

### 3.2.14. Quinolone derivatives

**3.2.14.1. Inhibition of topoisomerase II by quinolone derivatives XVIII.** Data from Stanton et al.[42] (Tables 18a and b).




**XVIII**

This data set was also divided into two sub data sets according to Figure 1 that gave QSAR equations 18a and 18b with good statistics.[43]

$$\log 1/C = 12.38(\pm 4.11)\text{CMR} - 0.65(\pm 0.22)\text{CMR}^2 - 0.91(\pm 0.21)I - 57.63(\pm 19.51)$$
$$n = 18, \quad r^2 = 0.901, \quad s = 0.167, \quad q^2 = 0.847 \quad (18a)$$

optimum CMR = 9.48 (9.35–9.60)

$I = 1$ for the presence of piperazine substituent at position −7.

$$\log 1/C = -2.33(\pm 1.01)C\log P - 1.33(\pm 0.88)$$
$$n = 6, \quad r^2 = 0.910, \quad s = 0.281, \quad q^2 = 0.800 \quad (18b)$$

By comparing Eqs. 18a and 18b, it is clear that these two subsets of quinolone derivatives **XVIIIa–XVIIIr** and **XVIIIs–XVIIIx** may be interacting with the receptor in two different modes.

### 3.2.15. Sulfonamide derivatives

**3.2.15.1. Inhibition of human carbonic anhydrase I isozyme (hCA I) by sulfonamide derivatives XIX (Fig. 2).** Data from Vullo et al.[30] (Tables 19a and b)

The following two QSAR were obtained with good statistics by splitting this data set into two sub data sets according to Figure 1:

$$\log 1/K_i = 0.36(\pm 0.13)C\log P - 0.24(\pm 0.05)\text{CMR} + 7.43(\pm 0.35)$$
$$n = 17, \quad r^2 = 0.872, \quad q^2 = 0.833, \quad s = 0.140 \quad (19a)$$

$$\log 1/K_i = 0.19(\pm 0.09)\text{NVE} - 15.63(\pm 10.96)$$
$$n = 4, \quad r^2 = 0.974, \quad q^2 = 0.830, \quad s = 0.315 \quad (19b)$$

outlier: **XIXr**

Eqs. 19a and 19b are for the sulfonamide derivatives **XIXa–XIXq** and **XIXr–XIXv**. NVE is another approach to understanding polarizability.

**Table 18a.** Biological and physicochemical constants used to derive QSAR equation 18a for the inhibition of topoisomerase II by quinolone derivatives (**XVIIIa–XVIIIr**)

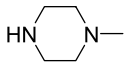
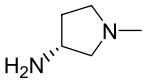
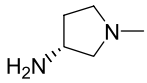
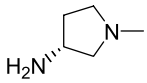
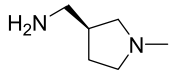
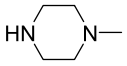
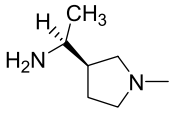
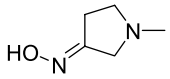
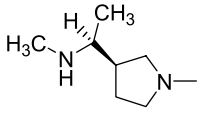
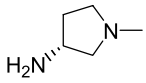
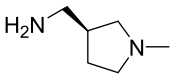
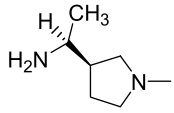
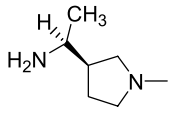
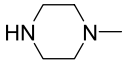
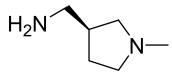
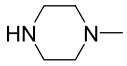
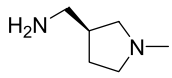
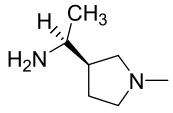
| No. | $R_1$ | $R_2$ | $R_3$ | X | log 1/*C* (Eq. 18a) | | | CMR | *I* |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Observed | Calculated | *Δ* | | |
| **XVIIIa** | H | F |  | CH | −0.18 | −0.26 | 0.08 | 8.72 | 1 |
| **XVIIIb** | H | F |  | N | 0.35 | 0.42 | −0.07 | 8.50 | 0 |
| **XVIIIc** | $CH_3$ | F |  | N | 0.94 | 0.87 | 0.07 | 8.97 | 0 |
| **XVIIId** | F | H |  | CF | 0.54 | 0.67 | −0.13 | 8.73 | 0 |
| **XVIIIe** | H | F |  | CF | 0.96 | 0.98 | −0.03 | 9.20 | 0 |
| **XVIIIf** | H | H |  | CF | −0.08 | −0.26 | 0.17 | 8.72 | 1 |
| **XVIIIg** | $NH_2$ | H |  | CF | 0.97 | 0.85 | 0.12 | 10.01 | 0 |
| **XVIIIh** | H | F |  | CF | 0.94 | 0.79 | 0.15 | 8.86 | 0 |
| **XVIIIi** | H | F |  | $C(OCH_3)$ | 0.00 | 0.02 | −0.01 | 10.72 | 0 |
| **XVIIIj** | H | F |  | CH | 0.52 | 0.66 | −0.14 | 8.72 | 0 |
| **XVIIIk** | H | F |  | CH | 1.06 | 0.98 | 0.08 | 9.18 | 0 |
| **XVIIIl** | H | F |  | CH | 0.78 | 1.02 | −0.24 | 9.64 | 0 |
| **XVIIIm** | H | H |  | C(Cl) | 0.73 | 0.77 | −0.04 | 10.12 | 0 |
| **XVIIIn** | H | H |  | C(Cl) | 0.06 | 0.07 | −0.01 | 9.19 | 1 |
| **XVIIIo** | H | H |  | C(Cl) | 0.96 | 1.01 | −0.05 | 9.66 | 0 |
| **XVIIIp** | H | F |  | $C(OCH_3)$ | −0.14 | 0.11 | −0.25 | 9.33 | 1 |
| **XVIIIq** | $NH_2$ | H |  | CF | 0.95 | 1.03 | −0.08 | 9.55 | 0 |
| **XVIIIr** | H | H |  | CF | 1.38 | 1.02 | 0.36 | 9.64 | 0 |

**Table 18b.** Biological and physicochemical constants used to derive QSAR equation 18b for the inhibition of topoisomerase II by quinolone derivatives (**XVIIIs–XVIIIx**)

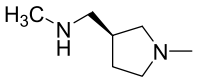
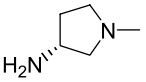
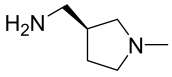
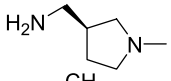
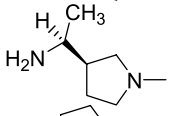
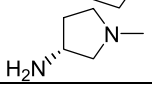
| No. | $R_1$ | $R_2$ | $R_3$ | X | $\log 1/C$ (Eq. 18b) | | | $C\log P$ |
|---|---|---|---|---|---|---|---|---|
| | | | | | Observed | Calculated | $\Delta$ | |
| **XVIIIs** | H | F |  | CF | 0.67 | 0.90 | −0.23 | −0.96 |
| **XVIIIt** | H | F |  | C(Cl) | −0.04 | −0.34 | 0.30 | −0.42 |
| **XVIIIu** | H | H |  | CF | 1.54 | 1.59 | −0.05 | −1.26 |
| **XVIIIv** | F | H |  | CF | 1.56 | 1.24 | 0.32 | −1.11 |
| **XVIIIw** | H | F |  | $C(OCH_3)$ | −0.19 | −0.10 | −0.09 | −0.53 |
| **XVIIIx** | H | H |  | C(Cl) | −0.24 | 0.01 | −0.24 | −0.57 |

**Table 19a.** Biological and physicochemical constants used to derive QSAR equation 19a for the inhibition of human carbonic anhydrase I isozyme (hCA I) by sulfonamide derivatives (**XIXa–XIXq**) (Fig. 2)

| No. | Compound | $\log 1/K_i$ (Eq. 19a) | | | $C\log P$ | CMR |
|---|---|---|---|---|---|---|
| | | Observed | Calculated | $\Delta$ | | |
| **XIXa** | Acetazolamide (**IVs**) | 6.05 | 5.97 | 0.07 | −0.98 | 4.65 |
| **XIXb** | Methazolamide (**IVa**) | 6.11 | 6.17 | −0.06 | 0.09 | 5.42 |
| **XIXc** | Dichlorophenamide (**IVc**) | 5.92 | 6.05 | −0.13 | 0.24 | 6.15 |
| **XIXd** | **IVe** | 5.82 | 6.00 | −0.18 | 0.92 | 7.37 |
| **XIXe** | **IVo** | 5.77 | 6.00 | −0.23 | 0.92 | 7.37 |
| **XIXf** | **IVp** | 6.01 | 6.00 | 0.01 | 0.92 | 7.37 |
| **XIXg** | **IVq** | 6.01 | 5.91 | 0.10 | 0.98 | 7.84 |
| **XIXh** | **IVu** | 6.05 | 5.84 | 0.20 | 1.11 | 8.30 |
| **XIXi** | **IVv** | 6.12 | 5.87 | 0.25 | 0.56 | 7.39 |
| **XIXj** | **IVf** | 5.97 | 5.87 | 0.10 | 0.88 | 7.86 |
| **XIXk** | **IVg** | 5.73 | 5.84 | −0.11 | 1.00 | 8.15 |
| **XIXl** | **IVh** | 6.57 | 6.50 | 0.07 | 2.43 | 7.53 |
| **XIXm** | **IVi** | 5.20 | 5.35 | −0.15 | 1.61 | 11.12 |
| **XIXn** | **IVj** | 5.21 | 5.19 | 0.03 | 1.47 | 11.59 |
| **XIXo** | **IVk** | 5.22 | 5.19 | 0.03 | 1.80 | 12.05 |
| **XIXp** | **IVl** | 6.07 | 6.09 | −0.03 | 1.74 | 8.21 |
| **XIXq** | **IVm** | 6.12 | 6.10 | 0.03 | 2.07 | 8.67 |

**Table 19b.** Biological and physicochemical constants used to derive QSAR equation 19b for the inhibition of human carbonic anhydrase I isozyme (hCA I) by sulfonamide derivatives (**XIXr–XIXv**) (Fig. 2)

| No. | Compound | $\log 1/K_i$ (Eq. 19b) | | | NVE |
|---|---|---|---|---|---|
| | | Observed | Calculated | $\Delta$ | |
| **XIXr** | Ethoxzolamide (**IVb**) | 7.60[a] | 0.35 | 7.25 | 86 |
| **XIXs** | Dorzolamide (**IVd**) | 4.30 | 4.44 | −0.14 | 108 |
| **XIXt** | E7070 (**IVt**) | 7.51 | 7.78 | −0.27 | 126 |
| **XIXu** | **IVr** | 6.60 | 6.30 | 0.31 | 118 |
| **XIXv** | **IVw** | 7.89 | 7.78 | 0.10 | 126 |

[a] Data point not included in the derivation of Eq. 19b.

**3.2.15.2. Inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives XX (Fig. 3).** Data from Garaj et al.[44] (Tables 20a and b) This data set was split into two sub data sets on the basis of outliers (see Fig. 1), which gave two QSAR equations 20a and 20b with good statistics.

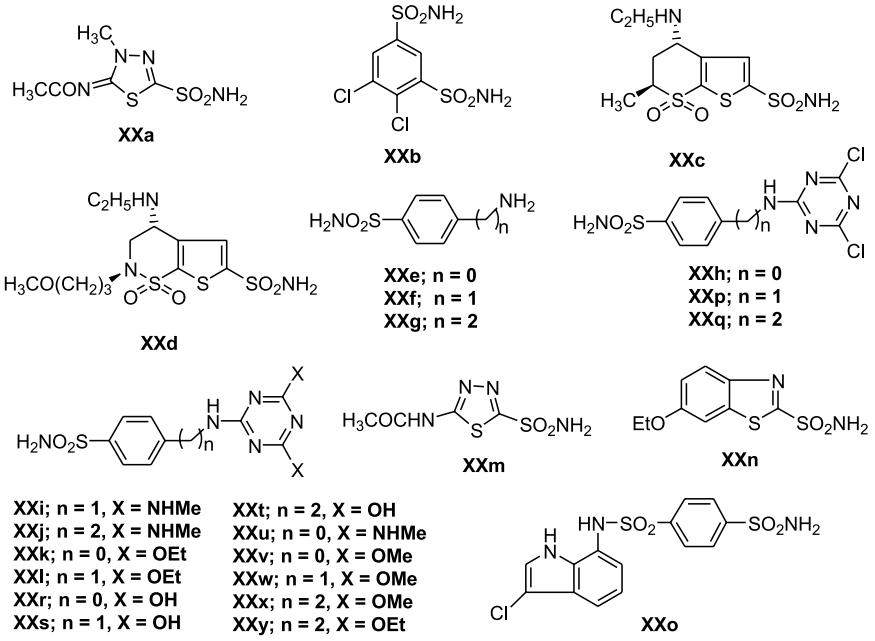



**Figure 3.** Structure of sulfonamide derivatives (**XXa**–**XXy**) used in the derivation of QSAR equations 20a and 20b.

**Table 20a.** Biological and physicochemical constants used to derive QSAR equation 20a for the inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives (**XXa**–**XXl**) (Fig. 3)

| Compound | $\log 1/K_i$ (Eq. 20a) | | | $C\log P$ |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **XXa** | 7.57 | 7.61 | −0.04 | 0.09 |
| **XXb** | 7.30 | 7.75 | −0.45 | 0.24 |
| **XXc** | 7.28 | 7.14 | 0.14 | −0.43 |
| **XXd** | 7.43 | 7.84 | −0.41 | 0.33 |
| **XXe** | 6.53 | 7.01 | −0.48 | −0.57 |
| **XXf** | 6.99 | 6.85 | 0.14 | −0.74 |
| **XXg** | 7.48 | 7.16 | 0.32 | −0.40 |
| **XXh** | 9.82 | 8.70 | 1.13 | 1.27 |
| **XXi** | 8.89 | 8.58 | 0.30 | 1.15 |
| **XXj** | 8.96 | 9.18 | −0.22 | 1.80 |
| **XXk** | 9.92 | 10.16 | −0.23 | 2.86 |
| **XXl** | 9.47 | 9.66 | −0.19 | 2.32 |

$$\log 1/K_i = 0.92(\pm 0.26)C\log P + 7.53(\pm 0.35)$$
$$n = 12, \quad r^2 = 0.856, \quad q^2 = 0.809, \quad s = 0.473 \quad (20a)$$

$$\log 1/K_i = -0.73(\pm 0.34)C\log P + 1.61(\pm 0.57) \times \log(\beta \times 10^{C\log P} + 1) + 6.88(\pm 0.25)$$
$$n = 11, \quad r^2 = 0.873, \quad q^2 = 0.804, \quad s = 0.198 \quad (20b)$$

inversion point for $C\log P = 0.84$;

$\log \beta = -0.92$;

outliers: **XXt**, **XXx**

Eqs. 20a and 20b are for the sulfonamide derivatives **XXa**–**XXl** and **XXm**–**XXy**. QSAR 20b is an allosteric bilinear correlation with $C\log P$ being similar as given in Eqs. 4a and 8b.

**Table 20b.** Biological and physicochemical constants used to derive QSAR equation 20b for the inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives (**XXm**–**XXy**) (Fig. 3)

| Compound | $\log 1/K_i$ (Eq. 20b) | | | $C\log P$ |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **XXm** | 7.60 | 7.60 | 0.00 | −0.98 |
| **XXn** | 7.47 | 7.24 | 0.23 | 2.05 |
| **XXo** | 7.62 | 7.50 | 0.12 | 2.37 |
| **XXp** | 6.91 | 6.69 | 0.22 | 0.74 |
| **XXq** | 6.86 | 6.81 | 0.05 | 1.38 |
| **XXr** | 6.85 | 6.78 | 0.07 | 1.31 |
| **XXs** | 6.44 | 6.69 | −0.24 | 0.76 |
| **XXt** | 6.26[a] | 8.11 | −1.85 | 1.41 |
| **XXu** | 6.82 | 6.99 | −0.17 | 1.69 |
| **XXv** | 6.82 | 7.06 | −0.24 | 1.80 |
| **XXw** | 6.81 | 6.76 | 0.05 | 1.26 |
| **XXx** | 6.79[a] | 8.55 | −1.76 | 1.91 |
| **XXy** | 7.92 | 8.0 | −0.08 | 2.97 |

[a] Data points not included in the derivation of Eq. 20b.

**3.2.15.3. Inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives XXI (Fig. 4).** Data from Cecchi et al.[45] (Tables 21a and b).

We derived two QSAR equations 21a and 21b from this data set by splitting it into two sub data sets on the basis of outliers (Fig. 1):

$$\log 1/K_i = 0.06(\pm 0.01)\mathrm{CMR} + 7.07(\pm 0.15)$$
$$n = 12, \quad r^2 = 0.900, \quad q^2 = 0.869, \quad s = 0.091 \quad (21a)$$

$$\log 1/K_i = 0.44(\pm 0.31)\mathrm{MgVol} + 6.95(\pm 0.82)$$
$$n = 5, \quad r^2 = 0.870, \quad q^2 = 0.719, \quad s = 0.193 \quad (21b)$$

Eqs. 21a and 21b are for the sulfonamide derivatives **XXIa**–**XXIl** and **XXIm**–**XXIq**.

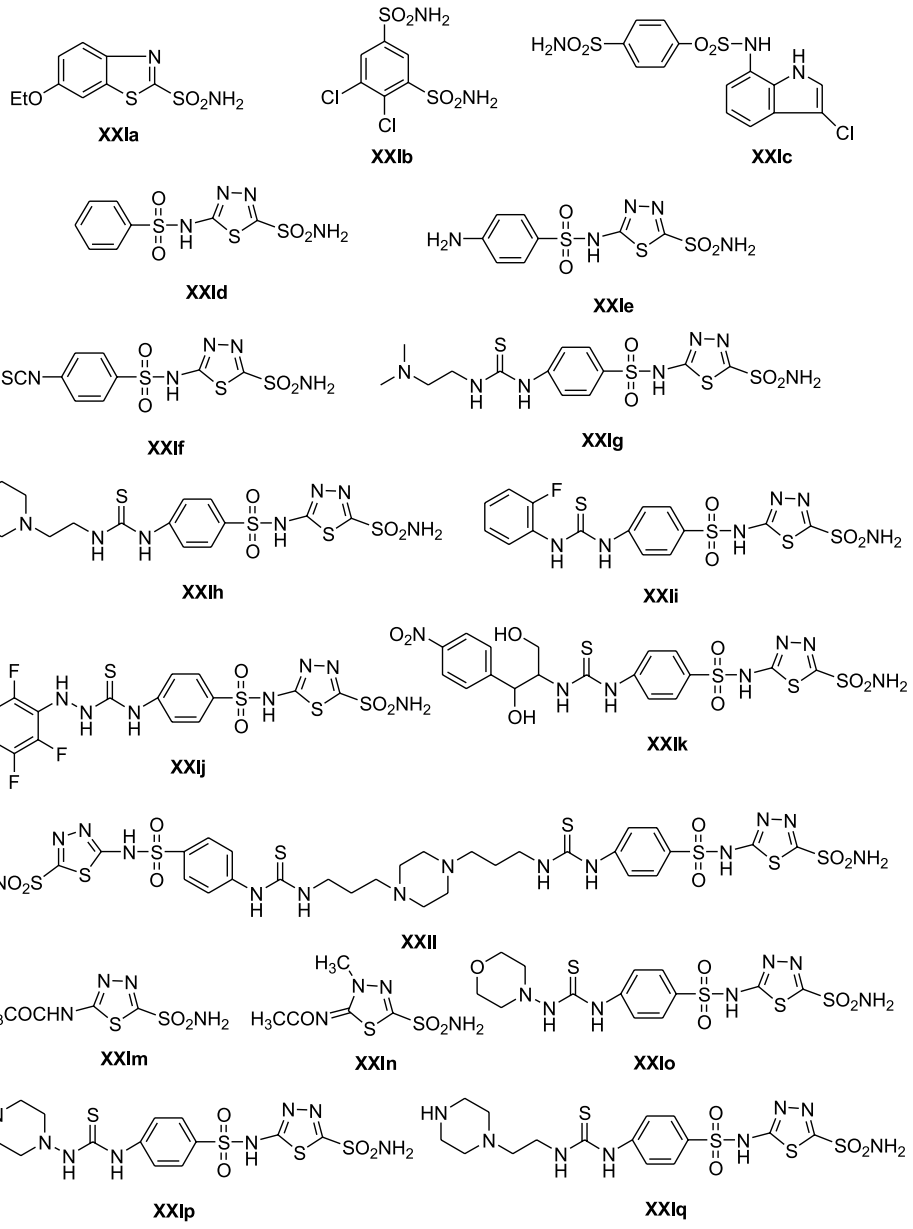



**Figure 4.** Structure of sulfonamide derivatives (**XXIa**–**XXIq**) used in the derivation of QSAR equations 21a and 21b.

**Table 21a.** Biological and physicochemical constants used to derive QSAR equation 21a for the inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives (**XXIa**–**XXIl**) (Fig. 4)

| Compound | $\log 1/K_i$ (Eq. 21a) | | | CMR |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **XXIa** | 7.47 | 7.41 | 0.05 | 6.30 |
| **XXIb** | 7.30 | 7.41 | −0.11 | 6.15 |
| **XXIc** | 7.62 | 7.58 | 0.04 | 9.29 |
| **XXId** | 7.38 | 7.46 | −0.08 | 7.07 |
| **XXIe** | 7.42 | 7.48 | −0.06 | 7.44 |
| **XXIf** | 7.64 | 7.54 | 0.10 | 8.62 |
| **XXIg** | 7.89 | 7.70 | 0.19 | 11.39 |
| **XXIh** | 7.72 | 7.74 | −0.02 | 12.29 |
| **XXIi** | 7.74 | 7.71 | 0.03 | 11.69 |
| **XXIj** | 7.70 | 7.74 | −0.04 | 12.12 |
| **XXIk** | 7.74 | 7.84 | −0.09 | 13.98 |
| **XXIl** | 8.34 | 8.36 | −0.02 | 23.34 |

**Table 21b.** Biological and physicochemical constants used to derive QSAR equation 21b for the inhibition of human carbonic anhydrase IX isozyme (hCA IX) by sulfonamide derivatives (**XXIm**–**XXIq**) (Fig. 4)

| Compound | $\log 1/K_i$ (Eq. 21b) | | | MgVol |
|---|---|---|---|---|
| | Observed | Calculated | $\Delta$ | |
| **XXIm** | 7.60 | 7.54 | 0.06 | 1.34 |
| **XXIn** | 7.57 | 7.60 | −0.03 | 1.48 |
| **XXIo** | 8.00 | 8.28 | −0.28 | 3.00 |
| **XXIp** | 8.49 | 8.36 | 0.14 | 3.18 |
| **XXIq** | 8.52 | 8.42 | 0.10 | 3.33 |

## 4. Overview

The present review on the problem of outliers contains 21 biological data sets for various types of molecules. These data sets were split into 46 sub data sets on

the basis of outliers (Fig. 1), resulting in the development of 46 individual QSAR and surprisingly 28 of them have the most important factor, that is, the hydrophobic parameter. We obtain bilinear $C\log P$ term in three QSAR (Eqs. 1a, 5a, and 17a). Optimum $C\log P$ for these QSAR is 2.30, 4.13 and 3.38, respectively. Parabolic $C\log P$ term was obtained in four QSAR (Eqs. 2b, 8a, 10a, and 13a). Optimum $C\log P$ for these QSAR is 3.34, 5.31, 5.17 and 4.30, respectively. Four QSAR (Eqs. 2c, 11a, 17b, and 20a) are linear in $C\log P$ with positive coefficient, whereas six QSAR (Eqs. 1c, 3b, 4b, 9b, 16b, and 18b) are linear in $C\log P$ with negative coefficient. The hydrophobic parameter for the substituents ($\pi$) is present in two QSAR (Eqs. 6a and 6b). It is interesting that three QSAR (Eqs. 4a, 8b, and 20b) are bilinear allosteric, whereas one QSAR (Eq. 12b) is an allosteric parabolic in terms of $C\log P$. Inversion point of $C\log P$ for these QSAR is 0.84, 4.70, 0.84, and 2.00, respectively.

The second important parameter for this review is CMR and is present in 16 QSAR. The role of CMR is brought out by four of the QSAR (Eqs. 9a, 12a, 16a, and 18a) where we obtain a parabolic CMR term. Optimum CMR for these QSAR is 10.13, 8.92, 11.47 and 9.48, respectively. Three QSAR (Eqs. 7b, 11b, and 21a) are linear in CMR with positive coefficient, whereas two QSAR (Eqs. 10b and 13b) are linear in CMR with negative coefficient. It is interesting to note here that three QSAR (Eqs. 1b, 2a, and 3a) bring out an allosteric reaction in terms of CMR. Inversion point of CMR for these QSAR is 12.76, 12.38 and 4.75, respectively.

MgVol term is present in four QSAR (Eqs. 3c, 4c, 7a, and 21a). Sterimol steric terms $B1$, $B5$ and $L$ also appear in four QSAR (Eqs. 6a, 14a, 15a, and 15b). NVE is a parameter that was found to be another approach to understanding polarizability and is present in one QSAR (Eq. 19b). Hammett electronic parameters $\sigma$ and $\sigma^+$ are also present in two QSAR (Eqs. 14a and 5b). Indicator variables appear in four QSAR (Eqs. 5b, 15a, 16a, and 18a).

## 5. Conclusion

In this review, 21 biological data sets for various types of molecules were divided into two or three subsets, as described in Figure 1, resulting in the formulation of 46 QSAR. We found that the QSAR results of the split subsets for a particular data set are always different in nature, which suggests that congeners associated with these QSAR may be acting by a different mechanism or interacting with the receptor in different modes.

In conclusion, we can say that the outliers may be acting by a different mechanism or interacting with the receptor in different modes. Thus, QSAR of the outliers should be considered before the synthesis of the next derivatives.

## References and notes


1. Hansch, C.; Maloney, P. P.; Fujita, T.; Muir, R. M. *Nature* **1962**, *194*, 178.
2. Hansch, C.; Leo, A. In *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
3. Selassie, C. D.; Garg, R.; Kapur, S.; Kurup, A.; Verma, R. P.; Mekapati, S. B.; Hansch, C. *Chem. Rev.* **2002**, *102*, 2585.
4. Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. *Bioorg. Med. Chem.* **2004**, *12*, 3391.
5. Selassie, C. D.; Mekapati, S. B.; Verma, R. P. *Curr. Top. Med. Chem.* **2002**, *2*, 1357.
6. C-QSAR Program, BioByte Corp., 201W. 4th, Suit 204, Claremont, CA 91711, USA. www.biobyte.com.
7. Hansch, C.; Steinmetz, W. E.; Leo, A. J.; Mekapati, S. B.; Kurup, A.; Hoekman, D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 120.
8. Hansch, C.; Kurup, A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1647.
9. Verma, R. P.; Kurup, A.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 237.
10. Verma, R. P.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 2355.
11. Abraham, M. H.; McGowan, J. C. *Chromatographia* **1987**, *23*, 243.
12. Verloop, A. In *The Sterimol Approach to Drug Design*; Marcel Dekker: New York, 1987.
13. Dallavalle, S.; Ferrari, A.; Biasotti, B.; Merlini, L.; Penco, S.; Gallo, G.; Marzi, M.; Tinti, M. O.; Martinelli, R.; Pisano, C.; Carminati, P.; Carenini, N.; Beretta, G.; Perego, P.; De Cesare, M.; Pratesi, G.; Zunino, F. *J. Med. Chem.* **2001**, *44*, 3264.
14. Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
15. Koshland, D. E., Jr.; Hamadani, K. *J. Biol. Chem.* **2002**, *277*, 46841.
16. Perutz, M. G. In *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*; Cambridge University: Cambridge, 1990.
17. Changeux, J.-P.; Edelstein, S. J. *Neuron* **1998**, *21*, 959.
18. Royer, W. E.; Knapp, J. E.; Strand, K.; Heaslet, H. A. *Trends Biochem. Sci.* **2001**, *26*, 297.
19. Changeux, J.-P.; Edelstein, S. J. *Curr. Opin. Neurobiol.* **2001**, *11*, 369.
20. Hurley, J. H.; Dean, A. M.; Sohl, J. L.; Koshland, D. E., Jr.; Stroud, R. M. *Science* **1990**, *249*, 1012.
21. Bender, W.; Staudt, M.; Trankle, C.; Mohr, K.; Holzgrabe, U. *Life Sci.* **2000**, *66*, 1675.
22. Christopoulos, A. *Nat. Rev. Drug Disc.* **2002**, *1*, 198.
23. Hansch, C.; Garg, R.; Kurup, A. *Bioorg. Med. Chem.* **2001**, *9*, 283.
24. Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. *Chem. Rev.* **2002**, *102*, 783.
25. Hansch, C.; Garg, R.; Kurup, A.; Mekapati, S. B. *Bioorg. Med. Chem.* **2003**, *11*, 2075.
26. Mekapati, S. B.; Kurup, A.; Verma, R. P.; Hansch, C. *Bioorg. Med. Chem.* **2005**, *13*, 3737.
27. Verma, R. P. *Lett. Drug Des. Discovery* **2005**, *2*, 205.
28. Wright, S. W.; Rath, V. L.; Genereux, P. E.; Hageman, D. L.; Levy, C. B.; McClure, L. D.; McCoid, S. C.; McPherson, R. K.; Schelhorn, T. M.; Wilder, D. E.; Zavadoski, W. J.; Gibbs, E. M.; Treadway, J. L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 459.
29. Innocenti, A.; Antel, J.; Wurl, M.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5703.

30. Vullo, D.; Scozzafava, A.; Pastorekova, S.; Pastorek, J.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2351.
31. Russel, M. G. N.; Baker, R. J.; Barden, L.; Beer, M. S.; Bristow, L.; Broughton, H. B.; Knowles, M.; McAllister, G.; Patel, S.; Castro, J. L. *J. Med. Chem.* **2001**, *44*, 3881.
32. Burrows, J. N.; Cumming, J. G.; Fillery, S. M.; Hamlin, G. A.; Hudson, J. A.; Jackson, R. J.; McLaughlin, S.; Shaw, J. S. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 25.
33. Ferrali, A.; Menchi, G.; Occhiato, E. G.; Danza, G.; Mancina, R.; Serio, M.; Guarna, A. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 145.
34. Ferrarini, P. L.; Calderone, V.; Cavallini, T.; Manera, C.; Saccomanni, G.; Pani, L.; Ruiu, S.; Gessa, G. L. *Bioorg. Med. Chem.* **2004**, *12*, 1921.
35. Miao, S.; Bao, J.; Garcia, M. L.; Goulet, J. L.; Hong, X. J.; Kaczorowski, G. J.; Kayser, F.; Koo, G. C.; Kotliar, A.; Schmalhofer, W. A.; Shah, K.; Sinclair, P. J.; Slaughter, R. S.; Springer, M. S.; Staruch, M. J.; Tsou, N. N.; Wong, F.; Parsons, W. H.; Rupprecht, K. M. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1161.
36. Pandey, A.; Volkots, D. L.; Seroogy, J. M.; Rose, J. W.; Yu, J.-C.; Lambing, J. L.; Hutchaleelaha, A.; Hollenbach, S. J.; Abe, K.; Giese, N. A.; Scarborough, R. M. *J. Med. Chem.* **2002**, *45*, 3772.
37. Kelley, J. L.; Linn, J. A.; Selway, J. W. T. *Eur. J. Med. Chem* **1990**, *25*, 131.
38. Leeson, P. D.; Baker, R.; Carling, R. W.; Curtis, N. R.; Moore, K. W.; Williams, B. J.; Foster, A. C.; Donald, A. E.; Kemp, J. A.; Marshall, G. R. *J. Med. Chem.* **1991**, *34*, 1243.
39. Ornstein, P. L.; Bleisch, T. J.; Arnold, M. B.; Wright, R. A.; Johnson, B. G.; Tizzano, J. P.; Helton, D. R.; Kallman, M. J.; Schoepp, D. D.; Herin, M. *J. Med. Chem.* **1998**, *41*, 358.
40. Baguley, B. C.; Ferguson, L. R. *Chem.-Biol. Interact.* **1985**, *56*, 145.
41. Manikumar, G.; Wadkins, R. M.; Bearss, D.; Von Hoff, D. D.; Wani, M. C.; Wall, M. E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5377.
42. Stanton, D. T. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1423.
43. Verma, R. P. *Bioorg. Med. Chem.* **2005**, *13*, 1059.
44. Garaj, V.; Puccetti, L.; Fasolis, G.; Winum, J.-Y.; Montero, J.-L.; Scozzafava, A.; Vullo, D.; Innocenti, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5427.
45. Cecchi, A.; Winum, J.-Y.; Innocenti, A.; Vullo, D.; Montero, J.-L.; Scozzafava, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5775.

## Biographical sketches

**Rajeshwar P. Verma** was born in 1966 in Barh (India). He received his M.Sc. (1988) and Ph.D. (1992) in chemistry from Magadh University, Bodh-Gaya. He spent a year at the same university as a postdoctoral fellow with Professor K. S. Sinha. He joined Roorkee University (Now IIT Roorkee) as a research associate and worked with Professor S. M. Sondhi (1993–1997). He also worked as a Lecturer of Chemistry at Gurukula Kangri University, Hardwar (1994–1995). He won a Research Associateship Award in December 1994 from the Council of Scientific and Industrial Research, New Delhi (India). In 1997, he moved to Pomona College to join the renowned QSAR research group of Professor Corwin Hansch and Cynthia Selassie, working as a postdoctoral research associate. Dr. Verma's research interests include the following: Isolation, characterization, and synthesis of natural products derived from medicinal plants; chemistry of isothiocyanates; synthesis of biologically important heterocycles and phenolic/active hydrogen compounds; application of principles of quantitative structure–activity relationships (QSAR) to the study of antifolates, multidrug resistance, free-radical-mediated toxicity of phenolic/active hydrogen compounds, and computer-assisted drug design.

**Corwin Hansch** received his undergraduate education at the University of Illinois and his Ph.D. in organic chemistry from New York University in 1944. After working with Du Pont Co., first on the Manhattan Project and then in Wilmington, DE, he joined the Pomona College faculty in 1946. He has remained at Pomona except for two sabbaticals: one at the Federal Institute of Technology in Zurich, Switzerland, with Professor Prelog and the other at the University of Munich with Professor Huisgen. The Pomona group published the first paper on the QSAR approach relating chemical structure with biological activity in 1962. Since then, QSAR has received widespread attention. Dr. Hansch is an honorary fellow of the Royal Society of Chemistry and has recently received the ACS Award for Computers in Chemical and Pharmaceutical Research for 1999.