# QSAR Models for Prediction of Binding and Inhibitory Properties of *[(E)-2-R-vinyl]benzene* Derivatives with Therapeutic Effects against *Helicobacter pylori*

**Juan J Galano[1,2] and Javier Sancho[1,2]\***

[1]*Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain*
[2]*Biocomputation and Complex Systems Physics Institute (BIFI), Joint Unit BIFI-IQFR (CSIC), Edificio I+D, Mariano Esquillor, 50018, Zaragoza, Spain*

### Abstract

*Helicobacter pylori* is a gram-negative bacterium that infects the luminal surface of the human gastric epithelium. Around one half of the world's population is thought to be infected by this bacterium, which is able to develop diseases such as peptic ulcer or gastric cancer. Eradication of *Helicobacter pylori* is becoming increasingly difficult due to resistance to common antibiotics. In previous work we have shown that an essential protein, flavodoxin, constitutes a target for the development of novel, specific antibiotics against infection caused by this microorganism, and we have described compounds sharing the [(E)-2-R-vinyl]benzene scaffold that exhibit bactericidal properties against *Helicobacter pylori* cultures. Based on the affinity and activity of 24 such compounds we have now developed QSAR models for affinity and minimal inhibitory concentration that will guide the improvement of antibacterial compounds based in the [(E)-2-R-vinyl]benzene scaffold. The two models show high statistical correlation and predictive capacity. Discovering novel chemicals with specific antimicrobial properties against *Helicobacter pylori*, and presumably not affected by existing resistances, will be of great help for the treatment of the diseases associated with this bacterium.

**Keywords:** *Helicobacter pylori*; Flavodoxin; Antibiotic; Antimicrobial; Inhibitor; QSAR

## Introduction

The classical Quantitative Structure-Activity Relationship (QSAR as we know them today) emerged in the first half of the 60s with the early works of Hansch and Fujita [1], the Free-Wilson method [2], the subsequent modification made by Ban [3] and the later approaches proposed by Kubinyi [4-10] and Topliss [11]. QSAR, assisted by a wealth of organic synthetic techniques, computer technologies and by the discovery of many potential therapeutic targets, has since become a central tool for chemists, biologists and biochemists in search of improving active compounds to use them as drugs for the treatment of a variety of diseases.

*Helicobacter pylori* (*Hp*) is a gram-negative bacteria that establishes life-long infections in the gastric mucosa of infected people [12,13]. Nowadays, about 50% of the global population is believed to be infected by *Hp*, with a lower prevalence in Europe and the United States but much higher in developing countries [14]. The presence of *Hp* in the human gastric mucosa is related to the development of diseases such as type B gastritis, chronic peptic ulcers and gastric neoplasias on individuals infected by that microorganism.

The traditional treatment used worldwide to combat *Hp,* known as the triple therapy (one proton pump inhibitor together with two wide spectrum antibiotics: clarithromycin and a choice of amoxicillin or metronidazole) [15], has unfortunately lost efficacy due to increased resistances. No new drugs have been developed in recent years to treat *Hp* infection, but some selective targets (one of them being the protein flavodoxin: *Hp*-Fld) have been identified [16]. *Hp*-Fld is an electron carrier essential for *Hp* viability [17,18] and it has been used in previous works [19,20] as target for the discovery of novel compounds that could be potential drugs against *Hp*. Four different biological responses (BR) or biological activities were assayed for those novel compounds, which are mainly derived from a common substructure represented in Figure 1. From those data we have gathered a database of 24 congeneric compounds and their respective biological activity values. We describe here two QSAR models for the binding affinity and the therapeutic potency of bactericidal compounds based on the *[(E)-2-R-vinyl]benzene* scaffold (Figure 1), which will guide the future improvement of *Hp*-specific bactericides.

## Materials and Methods

### Data set for QSAR analysis

Four *Hp* flavodoxin inhibitors were identified in previous work [19] and subsequently optimized [20]. Three of them, originally named C1, C2 and C4, were shown to display bactericidal activity against *Hp*. Compounds C1 and C2 are structurally related as they contain the common *[(E)-2-R-vinyl]benzene* substructure (Figure 1). In this work, those two compounds are referred to as I and II to retain their initial numbering. Twenty two additional compounds: 7 and 15 analogues of I and II respectively, were selected from [20]. The numbering assigned to them in that work [20] is retained here for clarity (Table 1). Compounds I, II, plus the additional 22 compounds constitute the database that we have used for the development of the QSAR models.
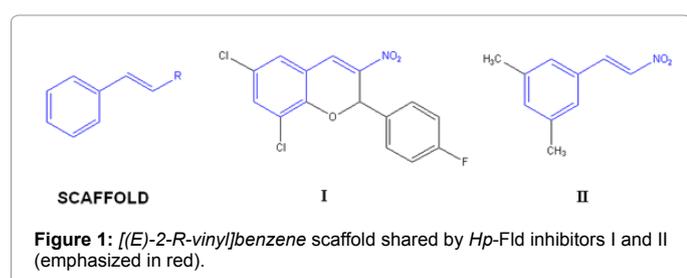
Cluster analysis (CA) was used to assess the structural diversity of such compounds by checking that relatively balanced distributions of compounds for a given cluster number were obtained, and to design the training (80%) and test (20%) series. Agglomerative hierarchical clustering was performed starting with each point as a singleton cluster and then repeatedly merging the two closest clusters until a single all-encompassing cluster remained. The "centroid method"

**\*Corresponding author:** Javier Sancho, Departamento de Bioquímica y Biología Molecular y Celular, Facultad de Ciencias, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain, Tel: (+34) 976 761286; Fax: (+34) 976 762123; E-mail: jsancho@unizar.es

**Figure 1:** *[(E)-2-R-vinyl]benzene* scaffold shared by *Hp*-Fld inhibitors I and II (emphasized in red).

| Cp[a] | Structure | $K_d$ | MIC (μM) | MCC (μM) | TI[b] |
|---|---|---|---|---|---|
| I | | 1 | 1.2 | 6 | 5.0 |
| 1[c] | | 0.43 | 2.4 | 1 | 0.4 |
| 2[c] | | 1.5 | 2.4 | 6 | 2.5 |
| 3[c] | | 1.7 | 2.4 | 6 | 2.5 |
| 4[c] | | 0.78 | 1.2 | 10 | 8.3 |
| 5[c] | | 1.9 | 7.5 | 70 | 9.3 |
| 6[c] | | 25 | 2 | 10 | 5.0 |
| 7[c] | | 1.8 | >150 | 12 | <0.08 |
| II | | 3.5 | 2.4 | 3.5 | 1.5 |
| 31[d] | | 7.9 | 4.75 | 40 | 8.4 |
| 32[d] | | 5.5 | 2.4 | 0.70 | 0.3 |
| 33[d] | | 10 | 8 | 0.01 | 0.001 |
| 34[d] | | 8.8 | 2.4 | 0.1 | 0.04 |
| 35[d] | | 0.67 | 2.4 | 2 | 0.8 |
| 36[d] | | 3.7 | >150 | 100 | <0.7 |
| 37[d] | | 16 | 4.8 | 1 | 0.2 |
| 38[d] | | 19 | 4.8 | 10 | 2.1 |
| 39[d] | | 11.0 | 75 | 100 | 1.3 |
| 40[d] | | 2.8 | 1.06 | 5 | 4.7 |
| 41[d] | | 4.4 | 0.53 | 8 | 15.1 |
| 42[d] | | 7.4 | 0.53 | 20 | 37.7 |
| 43[d] | | 4.1 | 0.53 | 8 | 15.1 |
| 44[d] | | 6.6 | 1.06 | 10 | 9.4 |
| 45[d] | | 40 | 1.06 | 20 | 18.9 |

[a]The same number of compound given in [20]
[b]Values calculated as *MCC/MIC*
[c,d]Derivatives of inhibitor I and II respectively

**Table 1:** Observed values for tested BRs in selected compounds.

was used to merge objects into clusters, and distance between two clusters was defined as the squared Euclidean distance between their respective centroids [21]. The clustering was performed using the STATGRAPHICS 5.1 program [22].

As described elsewhere [20], the biological response variables (BR) tested for each compound were: affinity ($K_d$) of the compound/*Hp*-Fld complex, minimal inhibitory concentration (*MIC),* minimal cytotoxic concentration (*MCC),* and therapeutic index (*TI=MCC/MIC*). Normality for each BR (i.e., the fit to a Normal statistical distribution) was checked and, in cases where it was not met, some compounds were removed to obtain a dataset with Normal distribution. For QSAR analyses each BR was expressed as LogBR.

## Molecular descriptors

A large number of molecular descriptors (MD) can be used in QSAR studies [23] (e.g., the latest version of the DRAGON program [24] calculates over 4800 MDs). The specific biological action of drugs may be described by a combination of hydrophobic, electronic and steric properties. A total of 35 MDs, mostly related to those properties and easily interpretable, have been selected for the current work. We have considered as hydrophobic descriptor the logarithm of the octanol-water partition coefficient (*Log P*) which has been calculated through Ghose-Crippen methodology [25,26] implemented in DRAGON 6. The four electronic descriptors used: total dipole moment (*μ*), total energy (*E_T*), *HOMO* eigenvalue and *LUMO* eigenvalue, were calculated

by quantum mechanical procedures using the MOPAC 6 software [27] with the Austin Model 1 (AM1) semi-empirical Hamiltonian [28] after full geometrical optimization of each molecule. Three additional descriptors were calculated based on the previous ones: electrophilicity index ($\omega$) [29], chemical hardness ($\eta$) and softness ($s$) [30]. The polarizability property (*Pol*) [31] was calculated from the partition method implemented in HYPERCHEM 7.51 software [32]. Steric and structural parameters were obtained from DRAGON 6 software. MDs exhibiting values above 25 were scaled (divided by their standard deviations). The QSAR models here developed are based on orthogonal MDs exhibiting very low inter-correlation coefficients. This was checked by subjecting MDs to an inter-correlation study. Description, classification and other statistic parameters of the MDs used are summarized in Table 2. The final MDs included in the best QSAR models derived were selected using genetic algorithms. The models selected were those displaying the best values of statistical parameters such as correlation coefficient, standard deviation and Fisher statistics.

### QSAR validations

The capability of the QSAR models to predict correctly the BR values within a training set was measured by the coefficient of determination ($R^2$) and the cross-validation coefficient ($q^2$), here determined by leave-one-out (LOO) cross-validation. The possibility that the correlation found between the selected MDs and each dependent variable were just due to statistical chance was ruled out using *Y-scrambling* [33] (thirty iterations were performed). Commonly, models with $R^2 > 0.75$ and $q^2 > 0.5$ are required [34]. The ability of one of the QSAR models to predict correctly the BR values of an external series (not used as training set) was measured through the external coefficient of determination ($0 < R^2_{Ext} < 1$).

### Inhibitor candidates

The sources of the 24 flavodoxin inhibitors selected here for the QSAR analysis, their purities, and the chemical characterizations of those made to order by Maybridge or synthesized at home, are reported in detail in Galano et al. [20].

### Biological activity assays

Table 1 shows the results of the biological responses determined per compound in Galano et al. [20]. The biological assays performed for determining such biological activities are also detailed there.

| DM or BR | Mean | SD | Min | Max | Description |
|---|---|---|---|---|---|
| MW | 280.1 | 63.4 | 167.2 | 448.0 | Molecular Weight |
| RBN | 2.3 | 0.6 | 2.0 | 4.0 | Number of Rotatable Bonds |
| RBF | 0.1 | 0.02 | 0.05 | 0.1 | Rotatable Bond Fraction |
| nAB | 8.5 | 3.4 | 6.0 | 17.0 | Number of Aromatic Bonds |
| nHM | 0.9 | 0.9 | 0.0 | 3.0 | Number of Heavy Atoms |
| nHet | 6.0 | 1.5 | 4.0 | 9.0 | Number of Heteroatoms |
| nX | 2.6 | 1.8 | 0.0 | 6.0 | Number of Halogen Atoms |
| X% | 11.2 | 8.0 | 0.0 | 27.8 | Percentage of Halogen Atoms |
| nCIC | 1.8 | 1.1 | 1.0 | 4.0 | Number of Rings (cyclomatic number) |
| nCIR | 2.3 | 1.7 | 1.0 | 7.0 | Number of Circuits |
| MCD | 0.5 | 0.2 | 0.3 | 0.8 | Molecular Cyclized Degree |
| RCI | 1.1 | 0.1 | 1.0 | 1.2 | Ring Complexity Index |
| ARR | 0.5 | 0.1 | 0.3 | 0.6 | Aromatic Ratio |
| nCar | 8.4 | 3.3 | 6.0 | 16.0 | Number of Aromatic Carbon (sp2) |
| nCbH | 4.6 | 2.2 | 0.0 | 10.0 | Number of Unsubstituted Benzene Carbon (sp2) |
| nCb | 3.8 | 1.6 | 2.0 | 6.0 | Number of Substituted Benzene Carbon (sp2) |
| nHDon | 0.04 | 0.2 | 0.0 | 1.0 | Number of Donor Atoms for H-bonds (N, O) |
| nHAcc | 4.3 | 1.9 | 2.0 | 8.0 | Number of Acceptor Atoms for H-bonds (N,O,F) |
| Qmean | 0.12 | 0.05 | 0.0 | 0.2 | Mean absolute charge (charge polarization) |
| Hy | -0.62 | 0.13 | -0.8 | -0.09 | Hydrophilic factor |
| AMR | 63.9 | 17.7 | 42.3 | 93.9 | Ghose-Crippen molar refractivity |
| ALOGP | 3.6 | 1.0 | 2.0 | 5.4 | Ghose-Crippen oct.-wat. partition coeff. (LogP) |
| ALOGP2 | 13.8 | 7.6 | 4.1 | 28.6 | Squared Ghose-Crippen oct.-wat. partition coeff. |
| SAtot | 290.6 | 61.8 | 206.1 | 409 | Total surface area from P_VSA-like descriptors |
| Vx | 279.6 | 69.2 | 190.5 | 401 | McGowan volume |
| Pol | 23.3 | 7.1 | 15.3 | 36.0 | Polarizability Ghose-Crippen |
| TE | 3711 | 941 | 5199 | 2298 | Total Energy (EV) |
| HOMO | -9.9 | 0.55 | -10.83 | -8.3 | Orbital HOMO Energy |
| LUMO | -1.7 | 0.24 | -2.3 | -1.2 | Orbital LUMO Energy |
| $\mu$ | 4.4 | 1.4 | 1.2 | 7.1 | Total Dipole Moment |
| $\eta$ | 4.1 | 0.23 | 3.4 | 4.4 | Chemical Hardness |
| U | -5.8 | 0.36 | -6.4 | -4.8 | Electrochemical Potential |
| s | 0.2 | 0.01 | 0.2 | 0.3 | Chemical Softness |
| $\omega$ | 4.1 | 0.37 | 3.4 | 5.0 | Electrophilic Index |
| $K_d$ ($\mu$M) | 7.7 | 9.3 | 0.4 | 40.0 | Dissociation Constant |
| MIC ($\mu$M) | 19.4 | 43.4 | 0.5 | 150.0 | Minimal Inhibitory Concentration |
| MCC ($\mu$M) | 18.7 | 29.3 | 0.01 | 100.0 | Minimal Cytotoxic Concentration |
| TI | 5.9 | 8.7 | 0.0 | 37.7 | Therapeutic Index (MCC/MIC) |

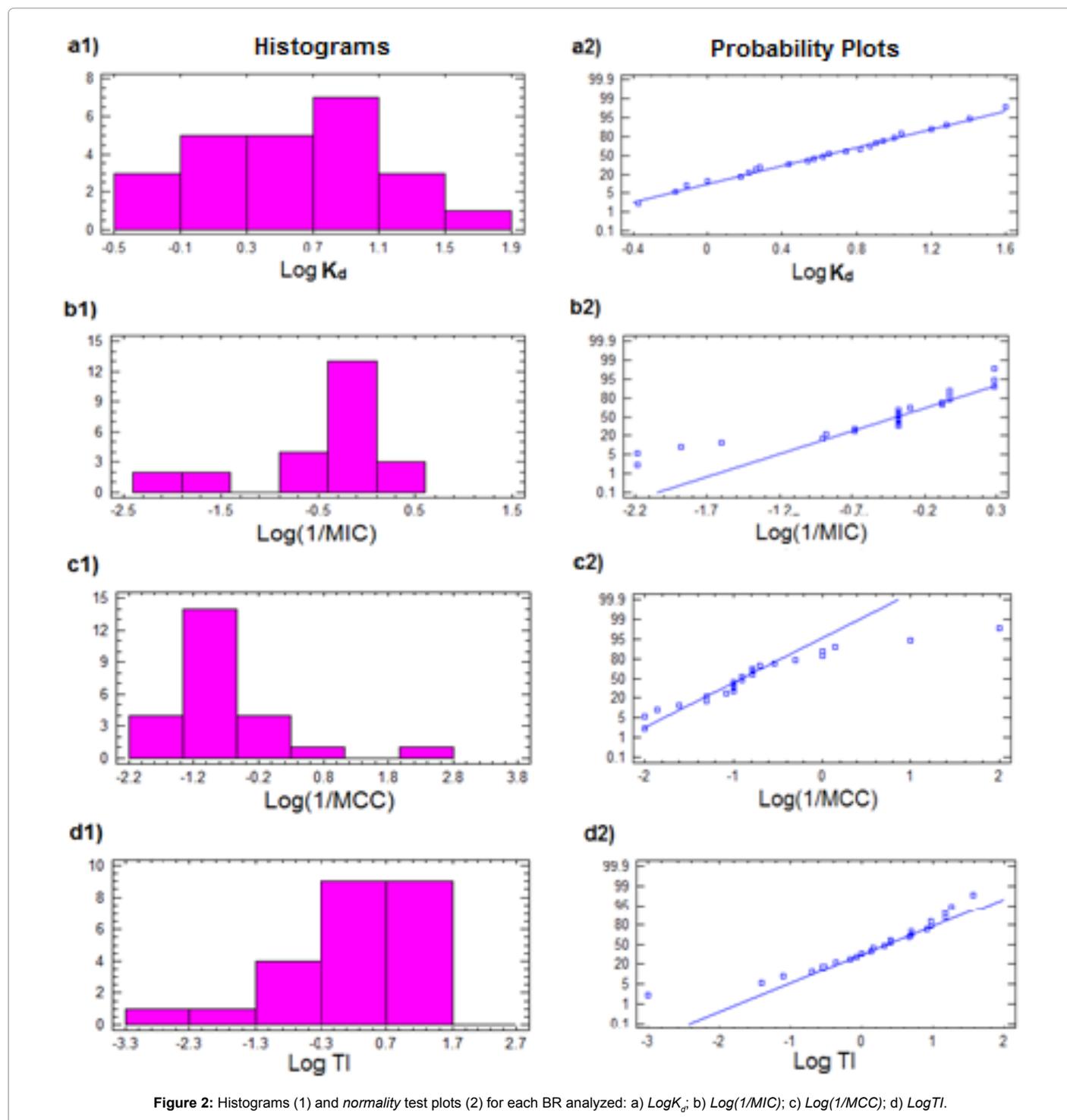**Table 2:** Statistical parameters, value ranges and description of the MDs and BRs analyzed.

## Results

### QSAR study for variants of inhibitors I and II

The observed improvement in the therapeutic index (*TI)* exhibited by some of the derivatives purposely designed in [20] prompted us to apply a classical QSAR approach to facilitate further optimizations. Twenty four structurally congeneric compounds, including I, II (Table 1) constituted the starting point. The values of the molecular descriptors finally selected for the QSAR models are summarized in Table S1.

Hierarchical CA revealed several clusters (Figure S1) evidencing a molecular diversity, which is also reflected in the value ranges of both MDs and BRs shown in Table 2.

**QSAR model for LogK$_d$:** This BR passed the *normality* test: the histogram in Figure 2 displays the typical Gauss bell curve of a *Normal* probability distribution, and the BR values fitted a straight line in the *Normal* probability plot (Figure 2). Typified kurtosis and asymmetry parameters also demonstrated compliance with *normality* (not shown). Six outliers (compounds 6, 35, 38, 40 and 45) were detected using



**Figure 2:** Histograms (1) and *normality* test plots (2) for each BR analyzed: a) *LogK$_d$*; b) *Log(1/MIC)*; c) *Log(1/MCC)*; d) *LogTI*.

standard statistical tests (residuals, standardized residuals, studentized residuals and Cook distances) and were removed. The training set, defined by CA comprised 15 compounds II, 1-3, 5, 7, 31-33, 36, 37, 39, 41, 42 and 44. The remainder compounds: I, 4, 34 and 43 constituted the test set.

The best model found for $LogK_d$ was:

**$Log K_d$** = $0.02(\pm 0.01)\%X + 0.29(\pm 0.12)TE + 1.04(\pm 0.5)\eta - 2.77(\pm 2.08)$ (1)

(s=0.167; F=25.41; SPRESS=0.237)

The model combines three MDs: *%X*, *TE* and *η*, which exhibit low pair correlations (0.265 for *%X* and *TE*; 0.359 for *%X* and *η*; and 0.002 for *TE* and *η*). The $R^2$ (=0.874) indicates that there is a good correlation between the observed $LogK_d$ values and the MDs selected. On the other hand, $q^2$ (=0.747) and $R^2_{ext}$ (=0.924) values are indicative of good internal and external predictability of this model. Besides, *Y-scrambling* suggests there is neither chance correlation nor chance prediction in the model (Figure S2). The observed $K_d$ values determined by ITC and the predicted values are compared in Table 3. Compound 1 is predicted as the most affine one in this series whereas compound 37 is predicted as the least affine one, which fully coincides with the experimental results.

**QSAR model for Log(1/MIC):** This BR did not initially met the *normality* requirement (Figure 2c and 2d), which was confirmed by both typified kurtosis and asymmetry values (not shown). Normality was obtained by removing compounds 7, 36 and 39. After this, compounds 37 and 40-42 and I were eliminated as outliers. Because as many as 8 compounds had to be discarded, we preferred to use the remaining 16 compounds as training set. Thus, no test set was extracted for external validation. The training set, constituted by compounds: II, 1-6, 31-35, 38 and 43-45 lead to the following model:

Log(1/MIC)=-0.25(±0.08)nCb- -8.48(±3.19)Hy + 1.38(±0.61)ω - 10.5(±4.21) (2)

(s=0.156; F=16.175; SPRESS=0.198)

The model combines three MDs: *nCb-*, *Hy* and *ω*, exhibiting low pair correlations (0.074 for *nCb-* and *Hy*; 0.004 for *nCb-* and *ω*; and 0.22 for *Hy* and *ω*). A high correlation was also obtained ($R^2$=0.801), the $q^2$ (=0.682) value warrants a good internal predictability and the *Y-scrambling* analysis discards both chance correlation and chance prediction (Figure S2). The observed *Log(1/MIC)* values and the predicted ones are compared in Table 4. The most potent compounds against *H. pylori*: 43 and 44 were correctly predicted by the model, whereas the least potent one: 33 was predicted as the second less potent one.

**QSAR model for Log(1/MCC):** Compound 33 was removed to ensure the *normality* condition (Figure 2e and 2f). Then, compounds 5, 34 and 45 were removed as outliers. Compounds: I, II, 2-4, 6, 31, 32, 36-41, 43 and 44 were the training set, and compounds 1, 7, 35 and 42 were chosen as test set. The orthogonality between the selected MDs was checked and all pair wise correlations were <=0.3 (not shown). The developed QSAR model is shown in equation 3:

*Log(1/MCC)*=-0.33(±0.07)nCb- -3.71(±0.70)Hy - 0.31(±0.11)TE - 3.07(±0.62) (3)

(s=0.302; F=11.583; SPRESS=0.409)

$R^2$ (=0.743) and $q^2$ (=0.53) values are lower than in the previous QSAR models and $R^2_{ext}$ (=0.512) is quite low, indicating poorer correlation and external predictability for this model.

| Compound | Y Obs. | Y Pred. | Residual |
|---|---|---|---|
| 1[a] | -0.37 | **-0.32** | -0.05 |
| 2 | 0.18 | 0.23 | -0.05 |
| 3 | 0.22 | 0.27 | -0.05 |
| 5 | 0.28 | 0.22 | 0.06 |
| 7 | 0.26 | 0.15 | 0.11 |
| II | 0.54 | 0.77 | -0.23 |
| 31 | 0.90 | 0.85 | 0.05 |
| 32 | 0.74 | 0.67 | 0.06 |
| 33 | 1.00 | 0.79 | 0.21 |
| 36 | 0.57 | 0.67 | -0.10 |
| 37[b] | 1.20 | **1.05** | 0.15 |
| 39 | 1.04 | 0.78 | 0.26 |
| 41 | 0.65 | 0.88 | -0.23 |
| 42 | 0.87 | 0.91 | -0.04 |
| 44 | 0.82 | 0.98 | -0.16 |

[a]Predicted as the highest affinity compound
[b]Predicted as the lowest affinity compound

**Table 3:** Observed and predicted $LogK_d$ values.

| Compound | Y Obs. | Y Pred. | Residual |
|---|---|---|---|
| 1 | -0.38 | -0.43 | 0.05 |
| 2 | -0.38 | -0.30 | -0.08 |
| 3 | -0.38 | -0.31 | -0.07 |
| 4 | -0.08 | -0.07 | -0.01 |
| 5[b] | -0.88 | **-0.96** | 0.08 |
| 6 | -0.30 | -0.52 | 0.22 |
| II | -0.38 | -0.37 | -0.01 |
| 31 | -0.68 | -0.49 | 0.19 |
| 32 | -0.38 | -0.23 | -0.15 |
| 33 | -0.90 | -0.95 | 0.05 |
| 34 | -0.38 | -0.28 | -0.10 |
| 35 | -0.38 | -0.47 | 0.09 |
| 38 | -0.68 | -0.45 | -0.24 |
| 43[a] | 0.28 | **-0.01** | 0.29 |
| 44[a] | -0.03 | **-0.01** | -0.02 |
| 45 | -0.03 | -0.12 | 0.09 |

[a]Predicted as the highest inhibitory activity compound
[b]Predicted as the lowest inhibitory activity compound

**Table 4:** Observed and predicted *Log(1/MIC)* values.

**QSAR model for Log(TI):** We also tried to develop a model for *Log(TI)* but no model with good statistical parameters was found ($R^2$<0.75, $q^2$<0.5 and $R^2_{ext}$<0.75).

## Discussion

### QSAR models for affinity and inhibitory properties of *[(E)-2-R-vinyl]benzene* derivatives

QSAR studies are based on the assumption that similar molecules have similar activities, and summarize existing relationships between chemical structures and biological activity within a dataset of structurally similar chemicals. Therefore, QSARs allow to predict the activities of new compounds using regression models relating a set of predictor variables to a biological response variable. Predictors may be physico-chemical properties or theoretical or structural parameters of compounds, while the response variables generally consist of biological activities such as binding affinity, potency, etc. We have performed classic QSAR analysis on compounds I and II, and analogues thereof, all sharing a *[(E)-2-R-vinyl]benzene* substructure. Several outliers were identified and excluded from the database. Reported sources for outliers in QSAR studies include flexible binding sites, inappropriate

calculation of selected MDs, different binding modes, etc. [35,36]. The specific reasons why those compounds are outliers are at present not known.

Four different BRs were modeled. For two of them, we found either no model with good statistical parameters (QSAR model for *Log(TI)*) or a model that showed a poor correlation and poor external predictability (QSAR model for *Log(1/MCC)*). In contrast, the QSAR models for $LogK_d$ and for *Log(1/MIC)*) appear promising. As shown in eq. 1, the developed model for $LogK_d$ describes the affinity of ligands for *Hp*-Fld by means of electronic and constitutional terms. Equation 1 suggests that decreasing *%X* (percentage of halogen atoms) decreases $LogK_d$, thus increasing the affinity of the compounds for *Hp*-Fld. It should be noticed that although this term is statistically significant, its contribution to the value of the BR is small. Similarly, decreasing *TE* (total energy) or $\eta$ (chemical hardness) lowers the value of $LogK_d$, thus increasing the affinity of the complex. The *TE* is related to the amount and type of bonds present in a given compound. In our context, its contribution to affinity suggests that reducing the number of heteroatoms and/or shifting from fluorine to iodine along the halogen series would increase the affinities of the compounds. On the other hand, $\eta$, whose contribution to $LogK_d$ is greater than that of *TE*, is inversely related to the reactivity of compounds. According to eq. 1, the lower the $\eta$ value (e.g. higher the reactivity) the lower the $LogK_d$ value and hence higher the compound/*Hp*-Fld affinity. The model correctly identifies the more and less affine compounds and its external predictability is high.

The QSAR model developed for *Log(1/MIC)* is given in eq. 2. The model suggests that decreasing *nCb-* (number of substituted benzene carbon) will increase the inhibitory activity. This is in perfect agreement with the above conclusion drawn from $LogK_d$ model which states that decreasing the number of bonds (benzene Carbon substitutions in this case) could increase the affinity of the compounds and consequently its inhibitory potency. Besides, it also could reflect on steric issues related to binding to flavodoxin. Meanwhile, the term *Hy* (hydrophilic factor) has also a negative contribution to this BR, which indicates that more hydrophobic substituents contribute to a greater inhibitory effect. Finally, more electrophilic compounds (greater electrophilicity index: $\omega$) also contribute to a greater inhibitory effect, which might indicate that the *Hp*-Fld binding site is an electron-rich site.

## Conclusions

Two QSAR models for target affinity and inhibitory activities of *[(E)-2-R-vinyl]benzene* derivatives have been developed that show high statistical correlation and predictive capacity. They provide insights about the type of physical-chemical interaction of these inhibitors with their possible biological receptor (the flavodoxin protein) as well as simple structural information about how to optimize such compounds in order to guide the design of variants with improved affinity and improved inhibitory effect towards *Hp* cells. The discovery of novel chemicals with such specific antimicrobial properties against *Hp* is urgently needed at present for the treatment of the diseases associated with these bacteria.

## References

1. Hansch C, Fujita T (1964) *p*-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J Am Chem Soc 86: 1616-1626.

2. free sm jr, wilson jw (1964) A mathematical contribution to structure-activity studies. J Med Chem 7: 395-399.

3. Ban T, Fujita T (1969) Mathematical approach to structure--activity study of sympathomimetic amines. Norepinephrine-uptake inhibition. J Med Chem 12: 353-356.

4. Kubinyi H (1976) Quantitative structure-activity relationships. 2. A mixed approach, based on Hansch and Free-Wilson Analysis. J Med Chem 19: 587-600.

5. Kubinyi H (1976) Quantitative structure-activity relationships. IV. Non-linear dependence of biological activity on hydrophobic character: a new model. Arzneimittelforschung 26: 1991-1997.

6. Kubinyi H (1977) Quantitative structure--activity relationships. 7. The bilinear model, a new model for nonlinear dependence of biological activity on hydrophobic character. J Med Chem 20: 625-629.

7. Kubinyi H (1977) Quantitative structure-activity relationships. V. A simple simple algorithm for Fujita-Ban and Free-Wilson analyses. Arzneimittelforschung 27: 750-758.

8. Kubinyi H, Kehrhahn OH (1976) Quantitative structure-activity relationships. 3.1 A comparison of different Free-Wilson models. J Med Chem 19: 1040-1049.

9. Kubinyi H, Kehrhahn OH (1976) Quantitative structure-activity relationships. 1. The modified Free-Wilson approach. J Med Chem 19: 578-586.

10. Kubinyi H, Kehrhahn OH (1978) Quantitative structure-activity relationships. VI. Non-linear dependence of biological activity on hydrophobic character: calculation procedures for bilinear model. Arzneimittelforschung 28: 598-601.

11. Topliss JG (1972) Utilization of operational schemes for analog synthesis in drug design. J Med Chem 15: 1006-1011.

12. Bruce MG, Maaroos HI (2008) Epidemiology of Helicobacter pylori infection. Helicobacter 13 Suppl 1: 1-6.

13. Marshall BJ, Warren JR (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1: 1311-1315.

14. Everhart JE (2000) Recent developments in the epidemiology of Helicobacter pylori. Gastroenterol Clin North Am 29: 559-578.

15. Malfertheiner P, Mégraud F, O'Morain C, Bell D, Bianchi Porro G, et al. (1997) Current European concepts in the management of Helicobacter pylori infection--the Maastricht Consensus Report. The European Helicobacter Pylori Study Group (EHPSG). Eur J Gastroenterol Hepatol 9: 1-2.

16. Chalker AF, Minehart HW, Hughes NJ, Koretke KK, Lonetto MA, et al. (2001) Systematic identification of selective essential genes in Helicobacter pylori by genome prioritization and allelic replacement mutagenesis. J Bacteriol 183: 1259-1268.

17. Cremades N, Bueno M, Toja M, Sancho J (2005) Towards a new therapeutic target: Helicobacter pylori flavodoxin. Biophys Chem 115: 267-276.

18. Freigang J, Diederichs K, Schäfer KP, Welte W, Paul R (2002) Crystal structure of oxidized flavodoxin, an essential protein in Helicobacter pylori. Protein Sci 11: 253-261.

19. Cremades N, Velázquez-Campoy A, Martínez-Júlvez M, Neira JL, Pérez-Dorado I, et al. (2009) Discovery of specific flavodoxin inhibitors as potential therapeutic agents against Helicobacter pylori infection. ACS Chem Biol 4: 928-938.

20. Galano JJ, Alías M, Pérez R, Velázquez-Campoy A, Hoffman PS, et al. (2013) Improved flavodoxin inhibitors with potential therapeutic effects against Helicobacter pylori infection. J Med Chem 56: 6248-6258.

21. Everitt BS (1993) Cluster Analysis. London: Edward Arnold.

22. Polhemus N (2000) Statgraphics Plus v.5.1, Statistical Graphics Corporation. p. Statistical Package.

23. Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. Weinheim, Germany: Wiley-VCH.

24. Talete srl, DRAGON v.6.0 p. (2010) Software for Molecular Descriptor Calculation.

25. Ghose AK, Crippen GM (1987) Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships I. Partition Coefficients as a measure of hydrophobicity. J Comput Chem 7: 565-577.

26. Ghose AK, Pritchett A, and Crippen GM (1988) Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships 3. Modeling hydrophobic interactions. J Comput Chem 9: 80-90.

27. Stewart JJP (1989) Optimization of parameters for Semiempirical Methods I. Method. J Comput Chem 10: 209-221.

28. Dewar MJ, Zoebisch EG, Healy EF, Stewart JP (1985) Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. J Am Chem Soc 107: 3902-3909.

29. Parr RG, László VS, Shubin L (1999) Electrophilicity Index. J Am Chem Soc 121: 1922-1924.

30. Parr RG and Pearson RG (1983) Absolute hardness: comParrion parameter to absolute electronegativity. J Am Chem Soc 105: 7512-7516.

31. Miller KJ (1990) Additivity Methods in Molecular Polarizability. J Am Chem Soc 112: 8533-8542.

32. Hyperchem v.7.51, (2002) Hypercube Inc.: Gainesville, FL (USA). Molecular Modeling Program.

33. Dogra SK "Y Scrmabling" From QSARWorld-A Strand Life Sciences Web Resource. Oct 06, (2011)

34. Tropsha A, Gramatica P and Gombar VK (2003) The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR & Combinatorial Science 22: 69-77.

35. Kim KH (2007) Outliers in SAR and QSAR: 2. Is a flexible binding site a possible source of outliers? J Comput Aided Mol Des 21: 421-435.

36. Kim KH (2007) Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers? J Comput Aided Mol Des 21: 63-86.