Available online at www.sciencedirect.com
ScienceDirect
European Journal of Medicinal Chemistry 44 (2009) 869–876
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
http://www.elsevier.com/locate/ejmech

Original article

# QSAR analysis of salicylamide isosteres with the use of quantum chemical molecular descriptors

R. Doležal [a], S. Van Damme [b], P. Bultinck [b], K. Waisser [a,*]

[a] Department of Inorganic and Organic chemistry, Charles University, Heyrovského 1203, Hradec Králové 50003, Czech Republic
[b] Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281(S3), Gent 9000, Belgium

Received 4 January 2008; received in revised form 2 April 2008; accepted 21 April 2008
Available online 6 May 2008

**Abstract**

Quantitative relationships between the molecular structure and the biological activity of 49 isosteric salicylamide derivatives as potential antituberculotics with a new mechanism of action against three Mycobacterial strains were investigated. The molecular structures were represented by quantum chemical B3LYP/6-31G* based molecular descriptors. A resulting set of 220 molecular descriptors, including especially electronic properties, was statistically analyzed using multiple linear regression, resulting in acceptable and robust QSAR models. The best QSAR model was found for *Mycobacterium tuberculosis* ($r^2 = 0.92$; $q^2 = 0.89$), and somewhat less good QSAR models were found for *Mycobacterium avium* ($r^2 = 0.84$; $q^2 = 0.78$) and *Mycobacterium kansasii* ($r^2 = 0.80$; $q^2 = 0.56$). All QSAR models were cross-validated using the leave-10-out procedure.


*Keywords:* QSAR; Antituberculotics; Salicylamide isosteres; Molecular descriptors

## 1. Introduction

The number of strategies one can follow in the search for a new drug is fairly limited. Some research methodologies may be based on random synthesis of new compounds, utilizing combinatorial chemistry techniques, followed by a biological test to select active substances. Similarly, various natural materials (e.g. soil, plants, tissues or body fluids) can be screened for biologically active substances. However, the chemical space in which drug like molecules are found is of such dimensions that random or combinatorial synthesis approaches are not likely to be very successful. A systematic optimization of a lead structure is usually needed.

A different approach is based on the so-called structure–activity relationships (SAR). The SAR philosophy stems from the idea that the biological response to a drug exposition depends on the drug's molecular structure so that, on condition that the structure–activity relationship is known, the effect of a molecule can, in principle, be predicted from its structure. Naturally, unraveling SARs is not always obvious.

The development of modern QSAR methods originated from several works, especially those by Hansch, Fujita, Free and Wilson [1–4]. In QSAR (quantitative structure–activity relationships), the molecular structure is described mathematically in terms of so-called molecular descriptors varying from simple descriptors such as the number of atoms or molecular weight to more advanced quantum chemical data, such as reactivity descriptors. A mathematical relationship is then sought between experimentally observed activities on the one hand and the descriptors on the other hand. The motivation behind the search for relevant QSARs is to predict the biological activity for some compound and/or to explain what the biological activity depends on. As such, true QSAR equations can be useful for the prediction as far as the molecular descriptors are more easily acquirable than the biological activity itself. In this regard, calculated molecular descriptors seem to be convenient for QSAR investigations.

In the present study, QSAR models for antituberculous activity were obtained for a 49-membered group of isosteres of

* Corresponding author. Tel.: +42 0495067276; fax: +42 0495067166.
*E-mail address:* waisser@faf.cuni.cz (K. Waisser).

0223-5234/$ - see front matter © 2008 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2008.04.020

salicylamides [5]. The compounds cover a promising group of potential drugs with a new mechanism of action. They can serve as structural templates for inhibitors of two-component regulatory systems in bacteria [6]. The isosteres of salicylamides were tested for in vitro activities against *Mycobacterium tuberculosis*, *Mycobacterium kansasii* and *Mycobacterium avium*. As tuberculosis is still a considerable illness causing the death of more than a million people a year [7], it remains a main target for drug discovery. Analyses of structure–activity relationships of salicylamide derivatives can produce fruitful suggestions for further research of antituberculotics. Using computed properties of the salicylamide derivatives we wish to identify the essential conditions depending on which the compounds elicit a stronger/weaker biological response.

## 2. Materials and methods

### 2.1. Computational chemistry

The molecular set of 49 molecules was drawn in HyperChem 7.52 [8]. As the molecules can be classified into five groups (Fig. 1), conformational analysis was carried out only for the five leading structures (i.e. R = H) using the MM+ molecular mechanics force field [9] and random conformational searching in dihedral angle space, followed by a geometry optimization.

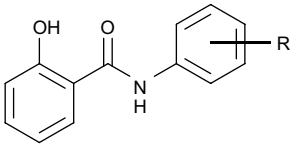
After 1000 iterations, the conformer having the lowest energy was considered to be the global minimum. Within each class, all other molecules were drawn in the same conformation, simply replacing functional groups followed by geometry optimization. The optimized geometries from HyperChem 7.52 were used as initial structures for ab initio calculations in Gaussian 03 [10], employing the B3LYP/6-31G* level of theory [11]. Further geometry optimizations and frequency analyses were carried out to ensure that the located conformers correspond to minima.

### 2.2. Biological activities

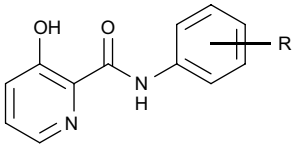
In vitro antimycobacterial activities of compounds **1–5** against *M. tuberculosis* CNCTC My 331/88, *M. kansasii* CNCTC My 235/80 and *M. avium* CNCTC My 330/88 were taken from the literature [5,12]. The antimycobacterial activities of the compounds were determined as minimal inhibitory concentrations (MIC) after incubation at 37 °C for 14 and 21 days. The concentrations of the compounds applied in the assay were 1000, 500, 250, 125, 62.5, 31, 16, 8, 4 and 2 μmol/l, giving rise to a discrete character of MICs.

### 2.3. Molecular descriptors

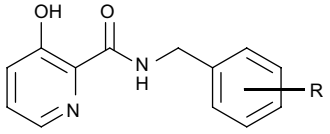
In this research, four groups of descriptors are generated to provide an as complete description of each molecule as



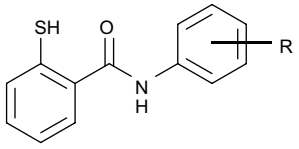

**1**



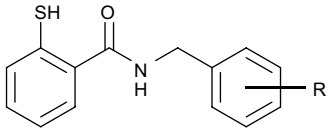

**2**




**3**



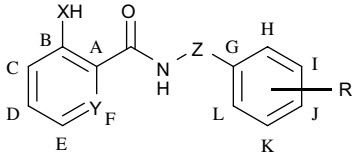

**4**




**5**

| | | R | | | | |
|---|---|---|---|---|---|---|
| **1** | **a** | H | **f** | 4-Br | **k** | 3-$NO_2$ |
| **2** | **b** | 4-$CH_3$ | **g** | 4-F | **l** | 4-$OCH_3$ |
| **3** | **c** | 4-Cl | **h** | 3-F | **m** | 4-$N(CH_3)_2$ |
| **4** | **d** | 3-Cl | **i** | 4-$CF_3$ | **n** | 4-COOEt |
| **5** | **e** | 3, 4-$Cl_2$ | **j** | 4-$NO_2$ | **o** | 4-CN |

Fig. 1. Five groups of isosteres of salicylamides [5].




X = O, S
Y = C, N
Z = $CH_2$, no atom (N binds the ring directly)

Fig. 2. Atom labeling in the salicylamide isosteres.

possible. The first class of descriptors accounts for lipophilic properties, the second one expresses electronic properties and the third one describes steric properties of molecules [13]. The final group is a set of Free–Wilson binary presence/absence descriptors.

The molecular descriptors used are shown in Table S1 (Supplementary data). The final set contains 220 descriptors. When applicable, molecular descriptors were taken immediately from the Gaussian 03 or Hyperchem 7.52 output. Other descriptors were obtained using self-developed programs with the wave function data as input. For a complete description of the quantum chemical descriptors used, the reader is referred to Refs. [14,20]. Throughout Table S1 (Supplementary data), A–L stands for the carbon atoms according to Fig. 2 (except for **2** and **3** group where "F" labels the nitrogen atom). X, Y, Z express substituent variations of the basic skeleton of the salicylamide isosteres.

### 2.4. Statistical computations

Statistical computations were carried out in the ARTE-QSAR program [15]. In the statistical analysis, a systematic search is performed among all descriptors to determine statistically significant relationships between the logarithm of the minimal inhibitory concentration (log MIC) and a selection of a number of descriptors out of the 220 available descriptors. As regression technique, multiple linear regression (MLR) is used because the dependent variables are normally distributed and all have the same variance. As the molecular set contains nearly 50 molecules, according to the Topliss ratio [16] the number of descriptors should not exceed 9. An excessive number of descriptors lead to over-correlated equations that are difficult to interpret in terms of interactions and mechanisms. It is well known that increasing the model complexity always increases the multiple correlation coefficient, but if the model complexity is not well supervised then the predictive power of the model decreases. The output provided in the paper gives the most promising 5-parameter model for each dataset. Multicollinearity between the remaining five descriptors is checked by calculating the variation inflation factors (VIF) [21]. The quality of the multivariate regression models is evaluated by assessing the goodness-of-fit (with $r^2$ the coefficient of determination, $r^2_{adj}$ the adjusted coefficient of determination and $s$ the standard error of the estimate) and the statistical significance (with $F$ the $F$-value of the Fisher test). Internal validation is conducted with leave-10-out cross-validation and is given by $q^2$. The significance of this value is estimated by Y-randomization. Possible outlying molecules are shown in Williams plots, which represent the domain of applicability [15].

## 3. Results

### 3.1. M. tuberculosis

The determination of minimal inhibitory concentration (MIC) was accomplished for 46 compounds in the case of 14-day incubation and for 45 compounds in the case of 21-day incubation. Statistical analysis of the biological activities and the descriptors (Table S2 – Supplementary data) provided the relationships summarized in Table 1 and Fig. 3. In Table 1 $\varepsilon_{HOMO}$ represents the energy of the HOMO. C-Se and D-Se are electrophilic superdelocalizabilities of the "C" and "D" atoms (see Fig. 2). H-am-Sn and H-am-Sn-w are classic and weighted nucleophilic superdelocalizabilities of the hydrogen atom in the CONH moiety.

### 3.2. M. avium

The determination of MIC was accomplished for 47 compounds in the case of 14-day incubation and for 46 compounds in the case of 21-day incubation. Statistical analysis of the biological activities and the descriptors (Table S3 – Supplementary data) provided the relationship summarized in Table 2 and Fig. 4. The descriptor softness is a function of the HOMO/LUMO energies (see Table S1 – Supplementary data). A-Fe and G-Fe are indices of electrophilic atomic frontier electron

Table 1
QSAR for *Mycobacterium tuberculosis*

| *M. tuberculosis* – 14-day incubation | | | | | *M. tuberculosis* – 21-day incubation | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Descriptor | Coefficient | Error | Student *t*-test | VIF | Descriptor | Coefficient | Error | Student *t*-test | VIF |
| Constant | −58.6559 | 4.3614 | −13.4487 | | Constant | −61.9681 | 4.6516 | −13.3218 | |
| $\varepsilon_{HOMO}$ | −11.1436 | 2.2350 | −4.9860 | 1.4946 | $\varepsilon_{HOMO}$ | −14.8063 | 2.4420 | −6.0631 | 1.5022 |
| C-Se | −7.0834 | 0.7672 | −9.2329 | 3.7511 | C-Se | −7.4271 | 0.8187 | −9.0717 | 3.6953 |
| D-Se | −13.4853 | 0.7543 | −17.8782 | 4.7765 | D-Se | −14.0900 | 0.8106 | −17.3828 | 4.5924 |
| H-am-Sn | 0.1190 | 0.0120 | 9.8978 | 2.5892 | H-am-Sn | 0.1211 | 0.0128 | 9.4381 | 2.5884 |
| H-am-Sn-w | 0.0003 | 0.00004 | 6.7438 | 1.4474 | H-am-Sn-w | 0.0003 | 0.00005 | 6.9789 | 1.4362 |
| $r^2 = 0.9152$ $r^2_{adj} = 0.9046$ $s = 0.1563$ $F = 86.3084$ | | | | | $r^2 = 0.9135$ $r^2_{adj} = 0.9024$ $s = 0.1667$ $F = 82.4080$ | | | | |
| $q^2 = 0.8918$ $n = 46$ $t^{0.05}_{crit} = 2.021$ $F^{0.05}_{crit} = 2.450$ | | | | | $q^2 = 0.8853$ $n = 45$ $t^{0.05}_{crit} = 2.023$ $F^{0.05}_{crit} = 2.456$ | | | | |

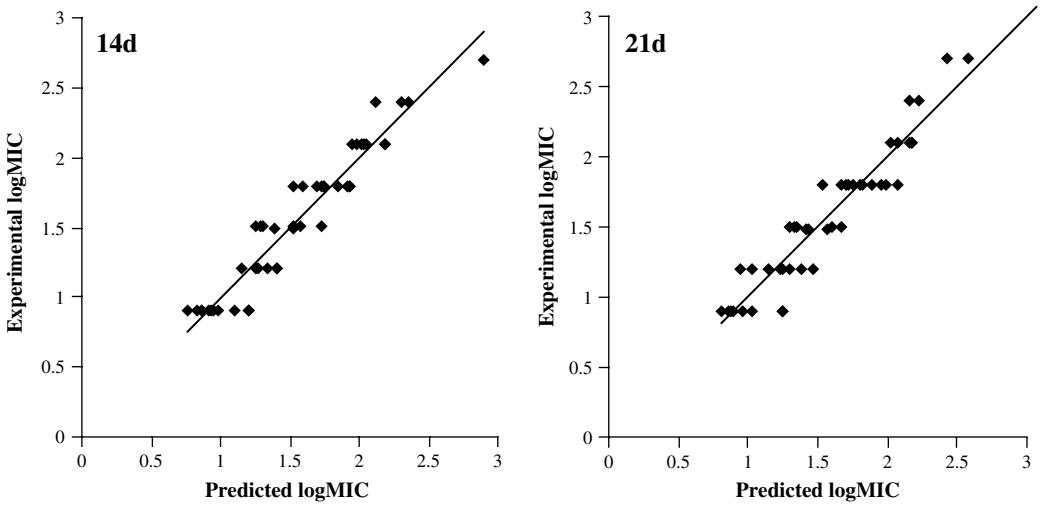



Fig. 3. Experimental and predicted biological activities for *Mycobacterium tuberculosis* (14- and 21-day incubation).

densities of the "A" and "G" atoms (see Fig. 2). B-Fn is the index of nucleophilic atomic frontier electron density of the "B" atom (see Fig. 2). G-char is the Mulliken charge on the "G" atom (see Fig. 2).

### 3.3. M. kansasii

The determination of the MIC was accomplished for 49 compounds in the case of 14-day incubation and for 48 compounds in the case of 21-day incubation. Statistical analysis of the biological activities and the descriptors (Table S4 – Supplementary data) provided the relationship summarized in Table 3 and Fig. 5. The abbreviation El. ext. stands for the electronic spatial extent. O-am-Fn is the index of nucleophilic atomic frontier electron density of the oxygen atom of the CONH function. Rn1-Se-sum is the sum of electrophilic superdelocalizabilities of the atoms forming the aromatic ring at the acyl side. H-am-Sn is the nucleophilic superdelocalizability of the hydrogen atom in the CONH moiety. Sum-Sn-w is the sum of all weighted nucleophilic superdelocalizabilities in the molecule.

## 4. Discussion

Statistical analysis of the achieved QSAR models confirms that the selected sets of molecular descriptors are capable of significantly explaining the variation in the biological activities. However, a high $r^2$ is a necessary but not sufficient condition for assessing the quality of the statistical relationships. Additional statistical parameters need to be evaluated before considering the models suitable for prediction and interpretation.

### 4.1. Diagnostics of the molecular descriptors

The Student *t*-test indicates that the descriptors in every model are significantly different from zero. The presence of multicollinearity is detected by calculating VIF values. These factors measure how much the variances of the estimated regression coefficients are inflated as compared to when the predictor variables are not linearly related. The largest VIF value among all descriptors is used as an indicator of the severity of multicollinearity. A maximum VIF value in excess of 10 is frequently taken as an indication that multicollinearity may be unduly influencing the least squares estimates. In the three models considered, no multicollinearity seems to be present. As the models satisfy the diagnostics of the molecular descriptors, they can be considered as robust and acceptable for further consideration.

### 4.2. Model validation

Since the real utility of a QSAR model lies in its ability to accurately predict the modeled property for new molecules,

Table 2
QSAR for *Mycobacterium avium*

| *M. avium* – 14-day incubation | | | | | *M. avium* – 21-day incubation | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Descriptor | Coefficient | Error | Student *t*-test | VIF | Descriptor | Coefficient | Error | Student *t*-test | VIF |
| Constant | −8.0183 | 1.3782 | −5.8181 | | Constant | −8.0275 | 1.4218 | −5.6459 | |
| Softness | −0.3898 | 0.0632 | −6.1658 | 1.5925 | Softness | −0.4145 | 0.0655 | −6.3278 | 1.5430 |
| A-Fe | −3.4819 | 0.4439 | −7.8441 | 3.9234 | A-Fe | −3.0719 | 0.4580 | −6.7078 | 3.8821 |
| B-Fn | 0.4483 | 0.0397 | 11.3039 | 1.9142 | B-Fn | 0.4775 | 0.0411 | 11.6140 | 1.8642 |
| G-char | 6.6061 | 0.6679 | 9.8912 | 8.2667 | G-char | 6.4158 | 0.6887 | 9.3159 | 8.1222 |
| G-Se | −4.1422 | 0.4766 | −8.6914 | 5.0208 | G-Se | −4.3282 | 0.4923 | −8.7911 | 4.9397 |
| $r^2 = 0.8129$ $r^2_{adj} = 0.7901$ $s = 0.1633$ $F = 35.6276$ | | | | | $r^2 = 0.8376$ $r^2_{adj} = 0.8174$ $s = 0.1684$ $F = 41.2753$ | | | | |
| $q^2 = 0.7563$ $n = 47$ $t^{0.05}_{crit} = 2.020$ $F^{0.05}_{crit} = 2.444$ | | | | | $q^2 = 0.7752$ $n = 46$ $t^{0.05}_{crit} = 2.021$ $F^{0.05}_{crit} = 2.450$ | | | | |

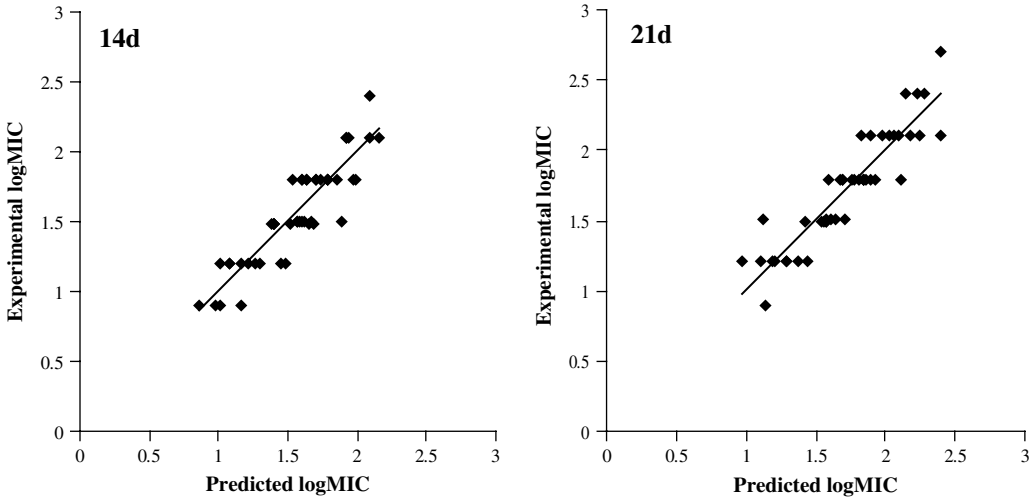



Fig. 4. Experimental and predicted biological activities for *Mycobacterium avium* (14- and 21-day incubation).

a reliable assessment of the predictive power is necessary for a confident application. This is achieved by validating the model, which is performed in this research in two different ways: the leave-10-out cross-validation and the Y-randomization test [17].

The internal validation is performed by a leave-10-out cross-validation. Cross-validation is the most commonly used validation technique where a number of modified datasets are created by deleting, in each case, a small group of compounds (10) from the data. A reduced dataset is obtained from the original set by removing the 10 data points. Based on the reduced dataset, a new QSAR model is derived and its predictive capacity is checked for the 10 molecules left out. The squared difference between the true response and the predicted response for each compound left out is calculated, and makes up the value for $q^2$. The premise being that if a QSAR model has a high average $q^2$ (>0.5) in the validation, we can reasonably conclude that the obtained model is robust. For each of the models built, the cross-validation measures are considerably higher than 0.5.

The Y-scrambling test is a widely used alternative to validate a QSAR model. In this test, the dependent variable vector is randomly shuffled and a new QSAR model is developed using the original independent-variable matrix. The process is repeated 500 times. It is expected that the resulting QSAR models should generally have lower $r^2$ and lower $q^2$ values. The results of the Y-randomization are summarized in Figs. S1–S3 (Supplementary data). If all QSAR models obtained in the Y-randomization test have relatively high $r^2$ and $q^2$ values, it implies that an acceptable QSAR model cannot be obtained for the given dataset by the current modeling method. In Figs. S1–S3 (Supplementary data) it is clearly demonstrated that the true QSAR models always have a significantly larger $r^2$ and $q^2$ compared to the randomly shuffled models.

### 4.3. Model applicability domain

The domains of applicability are represented by the Williams plots of residuals (Figs. S4–S6, Supplementary data). These plots indicate whether one of the molecules is possibly an influential outlier. The Williams plots indicate that in each of the models one molecule could have a significant influence on the fitted regression function. A case is considered influential if its exclusion causes a major change in the fitted regression function. A highly influential molecule was identified in the QSAR for *M. kansasii*. Nevertheless, the marked molecule (i.e. **2a**), when omitted, decreases $r^2$ approximately by 0.01. The obtained QSAR models still preserve statistical significance regarding outliers exclusion. Based on these considerations, one can conclude that there are practically no influential cases in the molecular set.

Table 3
QSAR for *Mycobacterium kansasii*

| *M. kansasii* – 14-day incubation | | | | | *M. kansasii* – 21-day incubation | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Descriptor | Coefficient | Error | Student *t*-test | VIF | Descriptor | Coefficient | Error | Student *t*-test | VIF |
| Constant | −7.3990 | 1.4277 | −5.1825 | | Constant | −7.1413 | 1.5358 | −4.6499 | |
| El. ext. | −0.0001 | +0.000 | −5.9579 | 1.0460 | El. ext. | −0.0002 | 0.0000 | −5.6840 | 1.0498 |
| O-am-Fn | −0.2288 | 0.0469 | −4.8772 | 1.6424 | O-am-Fn | −0.2651 | 0.0504 | −5.2581 | 1.6476 |
| Rn1-Se-sum | −0.3341 | 0.0655 | −5.0979 | 2.0034 | Rn1-Se-sum | −0.3611 | 0.0709 | −5.0957 | 1.9916 |
| H-am-Sn | 0.0959 | 0.0136 | 7.0749 | 1.4506 | H-am-Sn | 0.0824 | 0.0146 | 5.6557 | 1.4325 |
| Sum-Sn-w | +0.000 | +0.000 | 4.3939 | 1.0045 | Sum-Sn-w | +0.000 | +0.000 | 3.7130 | 1.0053 |
| $r^2 = 0.7995$ $r^2_{adj} = 0.7762$ $s = 0.2503$ $F = 34.3022$ | | | | | $r^2 = 0.8023$ $r^2_{adj} = 0.7788$ $s = 0.2683$ $F = 34.0901$ | | | | |
| $q^2 = 0.6394$ $n = 49$ $t^{0.05}_{crit} = 2.017$ $F^{0.05}_{crit} = 2.433$ | | | | | $q^2 = 0.5617$ $n = 48$ $t^{0.05}_{crit} = 2.018$ $F^{0.05}_{crit} = 2.438$ | | | | |

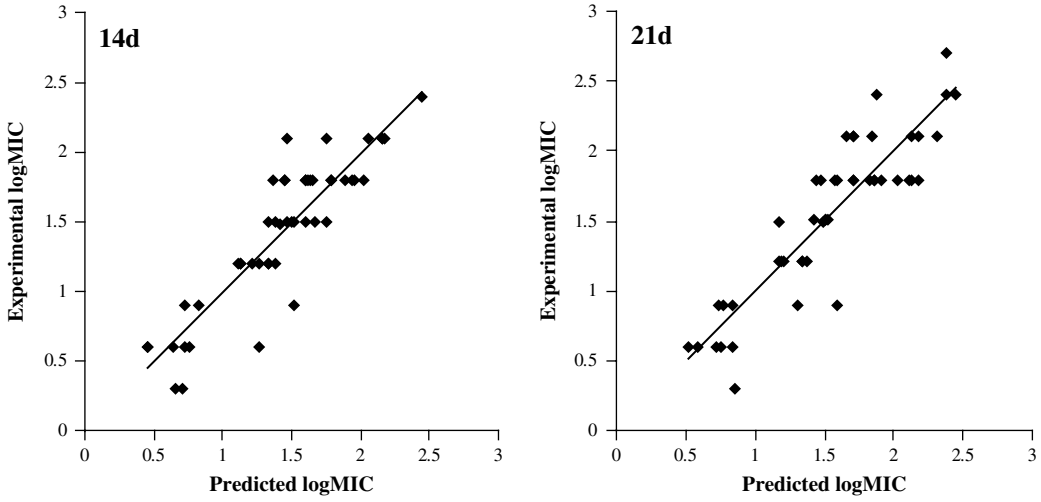



Fig. 5. Experimental and predicted biological activities for *Mycobacterium kansasii* (14- and 21-day incubation).

### 4.4. Mechanistic interpretation of the QSAR models

For *M. tuberculosis*, the QSAR model (Table 1) contains only descriptors relating to electronic properties of the molecules. The $\varepsilon_{HOMO}$ descriptor gives the energy of HOMO orbital and the biological activity (i.e. log MIC) decreases when the $\varepsilon_{HOMO}$ increases. The easier a molecule provides the HOMO electrons, the more the compound is active (i.e. exhibits lower MIC). The C-Se and D-Se descriptors express electrophilic superdelocalizability on the carbon atoms at 3 and 4 positions of the salicylic ring. As the electrophilic superdelocalizability of the two carbon atoms grows, the MIC becomes lower. We found that the C-Se and D-Se quite well mirror the Hammett $\sigma$ constants of substituents on the phenyl ring [18] in the amine moiety (Fig. 6). A significant correlation exists only within the separated five groups of the salicylamide isosteres, so that Hammett $\sigma$ constants cannot substitute for the C-Se and D-Se in general.

The H-am-Sn and H-am-Sn-w descriptors describe the nucleophilic superdelocalizability and the weighted nucleophilic superdelocalizability of the hydrogen atom of the CONH functional group. The weighted superdelocalizabilities descriptors were established according to Eq. (1).

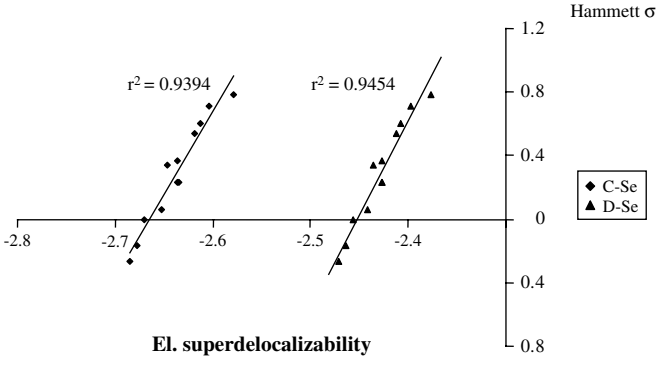



Fig. 6. Correlation of electrophilic superdelocalizabilities C-Se and D-Se for the **1** group of compounds to the Hammett substituent $\sigma$ constants.

$$\mathrm{Se}_a^{\mathrm{w}} = \sum_{\mathrm{MO}}^{\mathrm{occup}} \left[ \left( \frac{1}{\varepsilon_{\mathrm{MO}}} \sum_{i}^{\mathrm{atom\ a}} c_i^2 \right) \left( \frac{1}{\varepsilon_{\mathrm{MO}}} \sum_{i \neq a}^{\mathrm{all\ atoms}} c_i^2 \right) \right]$$

$$\mathrm{Sn}_a^{\mathrm{w}} = \sum_{\mathrm{MO}}^{\mathrm{virt}} \left[ \left( \frac{1}{\varepsilon_{\mathrm{MO}}} \sum_{i}^{\mathrm{atom\ a}} c_i^2 \right) \left( \frac{1}{\varepsilon_{\mathrm{MO}}} \sum_{i \neq a}^{\mathrm{all\ atoms}} c_i^2 \right) \right] \quad (1)$$

The concept of the weighted electrophilic or nucleophilic superdelocalizability descriptors resembles that of previously cited superdelocalizabilities [14] in treating the wave function data. Weighting the superdelocalizability increments overall molecular orbitals (MO) gives information on how the MOs are spread over the molecule. The biological activity increases with the increasing H-am-Sn and H-am-Sn-w descriptors. Thus, for a lower log MIC it is suggested to minimize both the nucleophilic superdelocalizabilities H-am-Sn and H-am-Sn-w on the hydrogen atom of the CONH function. This can be done by electron-donating substituents.

The QSAR model found for *M. avium* shows that the electronic descriptors prevail. Softness is a function of the energy gap between HOMO and LUMO. This global softness determines whether the molecule prefers to react by the orbital control mechanism, according to Klopman [19]. Substituents which make a molecule "softer" decrease the log MIC. The electrophilic index of frontier electron density of the carbon atom at 1 position (A-Fe) and the nucleophilic index of frontier electron density of the carbon atom at 2 position (B-Fn) of the salicylic moiety are related to the HOMO and the LUMO orbital, respectively. The remaining descriptors G-char, meaning the charge of the carbon atom at 1′ position of the ring in the amine moiety, and G-Se, meaning the electrophilic superdelocalizability of the same atom, are both related with the local electron density. The Mulliken charge on the atom "G" should be decreased while the electrophilic superdelocalizability should be increased in order to bring a more active compound.

The QSAR found for *M. kansasii* contains mostly electronic descriptors. The electronic spatial extent (El. ext) is a value which corresponds to the volume of a molecule, computed

as the integral of $\mathbf{r}^2\rho(\mathbf{r})$ overall space. This descriptor is likely to express the fact that substituted molecules are probably biologically more active than unsubstituted ones. The descriptors of the LUMO density on the oxygen atom of the CONH function (O-am-Fn) imply that the more the oxygen atom is able to participate in the charge accommodation, the higher biological activity is reached. The value of O-am-Fn descriptor can be increased upon appropriate electron-withdrawing substitution in the molecule. In the Rn1-Se-sum descriptor, we summed electrophilic superdelocalizabilities of the atoms forming the aromatic ring at the acyl side. This descriptor represents the ability of the conjugate system to provide electrons. Molecules with a higher Rn1-Se-Sum are likely to exhibit a lower log MIC. The nucleophilic superdelocalizability on the hydrogen atom of the CONH function (H-am-Sn descriptor) is associated with the ability of the atom to accommodate another electron in a virtual molecular orbital. The biological activity becomes lower when H-am-Sn grows. The last descriptor included in the QSAR for *M. kansasii* is the sum of weighted nucleophilic superdelocalizabilities of all atoms in a molecule. The positive regression coefficient of the Sum-Sn-w descriptor means that a higher evenness of distribution of virtual molecular orbitals lowers the biological activity.

### 4.5. Practical guidelines for the use of the QSAR models

It has become known that the knowledge of the structure–activity relationship does not imply an easy approach to the prediction of new structures standing out of the original molecular set. Mostly, the subtle interactions between the ligand and the receptor cannot be expected as a simple function of additive structural features suggesting what molecular fragments to be changed. A meaningful application of quantum chemically based QSAR models, which do overcome the non-aditivity of substituent constants, demands reconstructing a new structure with the use of complete set of molecular descriptors for each drug-candidate. As a rule, a set of computer molecular models should be investigated at first in order to design a new potential drug.

The three QSAR models found to elucidate the real mechanism of action of salicylamide isosteres. On the basis of the models, it can be induced that their common mechanism consists in electronic properties of the compounds. Existing differences in the biological activities and the underlying QSAR models suggest that the binding sites may differ in minor structural details. All of the in vitro activities are predicted by electronic properties, indicating that the working mechanism is mainly influenced by electronic features. If the molecular descriptors selected for one QSAR model are used for prediction of the other biological activities, the deterioration of $r^2_{adj}$ is not very large (Table 4). However, it is still possible that the three bacteria work in a slightly different way, as indicated by the difference in the descriptors.

In attempt to design a new potential drug-candidate according to the found QSAR models, whether for in vitro or in silico testing, it is suggested to introduce electron-withdrawing groups into the acyl moiety of salicylamide isosteres.

Table 4
Performance of the QSAR models on prediction of different biological activities

| | $r^2_{adj}$ | | |
|---|---|---|---|
| | *M. tuberculosis* set | *M. avium* set | *M. kansasii* set |
| QSAR of *M. tuberculosis* | 0.90/0.90[a] | 0.67/0.81 | 0.61/0.72 |
| QSAR of *M. avium* | 0.76/0.73 | 0.79/0.81 | 0.76/0.75 |
| QSAR of *M. kansasii* | 0.77/0.73 | 0.70/0.79 | 0.78/0.78 |

[a] (14-day model/21-day model).

Generally, all the antimycobacterial activities studied are enhanced by electron-withdrawing substituents (in the amine part), but the drop in the electron density is found important especially in the acyl side. From the QSAR models it also follows that the electron density on the CONH hydrogen should be somewhat increased. A promising structure of a salicylamide should contain the SH group in the acyl moiety. However, controlling the simultaneous changes of the electron density at different places in the molecule upon a substitution requires assistance of quantum chemical computations which only can give a true picture of the electron state.

## 5. Conclusions

The best QSAR model was found for *M. tuberculosis* ($r^2 = 0.92$; $q^2 = 0.89$), the worst QSAR model was found for *M. kansasii* ($r^2 = 0.80$; $q^2 = 0.56$). The influence of the incubation time is almost negligible. The previous Free–Wilson analyses [5] were marked with low *F*-values (3.52–5.33) and relatively large variances of many regression coefficients. Therefore, the presented QSAR study on salicylamide isosteres with the use of quantum chemical descriptors is a significantly better quantitative description of the observed antimycobacterial activities. The QSAR models may be used for predictions of antimycobacterial activities provided the necessary molecular descriptors of new drug-candidates are known. Generally, all the antimycobacterial activities studied are enhanced by electron-withdrawing substituents.

In our future research, we will focus on obtaining information on other similar compounds to salicylamides and will try to evaluate our QSAR models. The prediction of the biological activity for an external test set is the only way how to ultimately assess the quality of the found QSAR models.

## Acknowledgement

Authors are very grateful to Tom Kuppens from the department of inorganic and physical chemistry at Ghent University for his significant help in quantum chemical computations. The cooperation of authors was enabled by support of the Erasmus/Socrates program and by project No. MSM 0021620822 of the Ministry of Education of the Czech Republic. S.V.D. thanks the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for the Ph.D. grant.

## Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejmech.2008.04.020.

## References

[1] C. Hansch, P.P. Maloney, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[2] C. Hansch, T. Fujita, J. Am. Chem. Soc. 86 (1964) 1616–1626.
[3] S.M. Free, J.W. Wilson, J. Med. Chem. 7 (1964) 395–399.
[4] T. Fujita, T. Ban, J. Med. Chem. 14 (1971) 148–152.
[5] K. Waisser, M. Peřina, P. Holý, M. Pour, O. Bureš, J. Kuneš, V. Klimešová, V. Buchta, P. Kubanová, J. Kaustová, Arch. Pharm. Med. Chem. 336 (2003) 322–335.
[6] M.J. Macielag, J.P. Demers, S.A. Fraga-Spano, D.J. Hlasta, S.G. Johnson, R.M. Kanojia, R.K. Russell, Z. Sui, M.A. Weidner-Wells, H. Werblood, B.D. Foleno, R.M. Goldschmidt, M.J. Loeloff, G.C. Webb, J.F. Barrett, J. Med. Chem. 41 (1998) 2939–2945.
[7] WHO (World Health Organization), Tuberculosis. Fact Sheet No. 104, Revised March 2007, Geneva, 2007.
[8] HyperChem™ Professional 7.52, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA, 2005.
[9] N.L. Allinger, J. Am. Chem. Soc. 99 (1977) 8127–8134.
[10] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, Gaussian 03, Revision B.05, Gaussian, Inc., Pittsburgh PA, 2003.
[11] A.D. Becke, J. Chem. Phys. 98 (1993) 5648–5652.
[12] K. Waisser, O. Bureš, P. Holý, J. Kuneš, R. Oswald, J. Jirásková, M. Pour, V. Klimešová, L. Kubicová, J. Kaustová, Arch. Pharm. Med. Chem. 336 (2003) 53–71.
[13] G.M. Downs, in: P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere (Eds.), Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, New York, 2004, pp. 515–538.
[14] M. Karelson, V.S. Lobanov, A.R. Katritzky, Chem. Rev. 96 (1996) 1027–1043.
[15] S. Van Damme, P. Bultinck, J. Comput. Chem. 28 (2007) 1924–1928.
[16] J.G. Topliss, R.P. Edwards, J. Med. Chem. 22 (1979) 1238–1244.
[17] P. Gramatica, QSAR Comb. Sci. 26 (2007) 694–701.
[18] C. Hansch, A. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, New York, 1979.
[19] G. Klopman, J. Am. Chem. Soc. 90 (1968) 223–234.
[20] M. Karelson, in: P. Bultinck, H. De Winter, W. Langenaeker, J.P. Tollenaere (Eds.), Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, New York, 2004, pp. 641–667.
[21] J. Neter, M.H. Kutner, C.J. Nachtsheim, W. Wasserman, Applied Linear Statistical Models, fourth edition, WCB/McGraw-Hill, 1996.