Effect of Substitution on the Antimycobacterial Activity of 2-(Substituted benzyl)sulfanyl Benzimidazoles, Benzoxazoles, and Benzothiazoles— A Quantitative Structure–Activity Relationship Study

Oldřich Pytela[a] and Věra Klimešová*,[b]

[a] *Institute of Organic Chemistry and Technology, Faculty of Chemical Technology, University Pardubice; Studentsktá 573, 532 10 Pardubice, Czech Republic: and* [b] *Department of Inorganic and Organic Chemistry, Faculty of Pharmacy, Charles University; Heyrovského 1203, 500 05 Hradec Králové, Czech Republic.*

Received August 4, 2010; accepted November 10, 2010; published online November 11, 2010

A set of 1160 minimum inhibitory concentration (MIC) values evaluating effect of substitution on the antimycobacterial activity of the previously published 2-(substituted benzyl)sulfanyl benzimidazoles, benzoxazoles, and benzothiazoles has been analyzed by the methods of multidimensional analysis (exploratory analysis, 2D-nonlinear mapping (NLM), principal component analysis (PCA), factor analysis (FA), multiple linear regression (MLR)). The antimycobacterial activity of 2-(subst. benzyl)sulfanyl derivatives of benzimidazole (BIM), 5-methylbenzimidazole (5-Me-BIM), benzoxazole (BOZ), and benzothiazole (BTZ) increased in the order of BTZ<BOZ~BIM<5-Me-BIM. The sensitivity of particular strains towards these compounds decreased in the order of *Mycobacterium kansasii* 6509/96, *M. avium* My 330/88, *M. kansasii* My 235/80, and *M. tuberculosis* My 331/88. In general, derivatives with 3-$CSNH_2$, 2,4-$(NO_2)_2$, 4-$CSNH_2$, 3,5-(NO_2), and partially 4-NO_2 substituents possess the highest antimycobacterial activity. The effect of substitution was also described quantitatively with good correlation factor *R* of 0.79—0.88. The log MIC values depended with a negative slope on the Hammett substituent constants σ or molar refractions MR and, for the given set of substituents, were dominantly raised with increasing log *P* value and partially lowered with $(\log P)^2$ or $\sigma \times \Delta \log P$. The derivatives featuring high polarizability, low lipophilicity and electron-withdrawing subtituents showed the highest antimycobacterial activity. The dependence on the steric substituent constant *v* was not statistically significant and, therefore, the *ortho* effect was most probably not important.

Key words quantitative structure–activity relationship; benzimidazole; benzothiazole; benzoxazole; 2-benzylsulfanyl derivative; antimycobacterial activity

Tuberculosis (TB) belongs to leading cause of infectious disease mortality in the world. In spite of the fact that drugs for treating tuberculosis have been available for more than half a century, the total number of deaths and TB cases is still rising due to population growth. TB annual incidence rates have peaked globally in 2003—2004, however they are falling very slowly in all World Health Organization (WHO) regions, and stagnating in Eastern Europe. According to the statistical data of WHO, 2 billion people are infected with *Mycobacterium tuberculosis*, the microbes that cause TB, almost 9.2 million new TB cases occurred annually and 1.7 million people died.[1] A serious threat to TB control worldwide becomes multidrug-resistant strains (MDR-TB) and even more serious is the emergence of extensively drug-resistant TB (XDR-TB) that has been confirmed in more than 57 countries. This increase of MDR-TB and emergence of XDR-TB provide the rationale to search for new antimycobacterial drugs.[2—4]

Previously, we have reported the synthesis of 2-(substituted benzyl)sulfanyl derivatives of benzimidazole,[5] 5-methylbenzimidazole,[6] benzoxazole,[7] and benzothiazole[8] (X=H, 4-NO_2, 3-NO_2, 2-NO_2, 3,5-$(NO_2)_2$, 2,4-$(NO_2)_2$, 4-CN, 3-CN, only). General structure of these compounds is shown in Fig. 1.

Benzyl moiety has been modified by electron-donating and electron-withdrawing substituents at various positions on benzene ring (see Table 1). The compounds were evaluated *in vitro* against *Mycobacterium tuberculosis*, *M. avium* and two strains of *M. kansasii*. The results of the antimycobacterial screening revealed that compounds exhibited the significant *in vitro* activity against obligate strain *M. tuberculosis* and in particular, against opportunistic non-tuberculous mycobacteria (*M. avium*, *M. kansasii*). *M. avium* is naturally resistant toward a number of antitubercoulosis drugs and disseminated infections become the serious health problem.

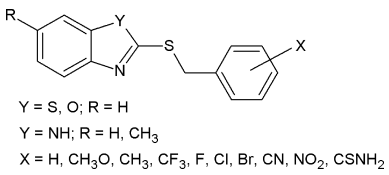

Fig. 1. General Structure of 2-(Subst.benzyl)sulfanyl Benzimidazoles, Benzoxazoles, and Benzothiazoles

In the present work, a quantitative structure–activity relationship (QSAR) analysis of the 2-(subst. benzyl)sulfanyl derivatives of benzazoles is reported. We attempted to discover and quantitatively describe the structure–activity relationships for our model compounds and their observed antimycobacterial activity. The results of the study can lead to design the most effective compounds and aid to find a general relationship between their structure and antimycobacterial activity.

Experimental

Chemistry The synthesis and structural features of 2-(subst. benzyl)sulfanyl benzimidazole (BIM),[5] 5-methylbenzimidazole (5-Me-BIM),[6] benzoxazole (BOZ),[7] and some benzothiazole (BTZ)[8] derivatives (benzazoles) have already been published.

Biology The antimycobacterial activity of the model compounds was evaluated *in vitro* against *Mycobacterium tuberculosis* CNCTC My 331/88, *Mycobacterium kansasii* CNCTC My 235/80, *Mycobacterium kansasii* 6509/96 and *Mycobacterium avium* CNCTC My 330/88 using the mi-

* To whom correspondence should be addressed. e-mail: vera.klimesova@faf.cuni.cz

Table 1. Measured $\log MIC_p$ Values, Hammett Substituent Constants σ, Charton Steric Constants υ, and Calculated $\Delta \log P$ and ΔMR

X	$\log MIC_p$	σ	σ^-	σ_{ortho}	σ^-_{ortho}	υ	$\Delta \log P$	ΔMR
H	1.93E+00	0	0	0	0	0	0	0
4-Cl	2.56E+00	0.22	0.24	0.22	0.24	0	0.518	4.805
3-Cl	2.56E+00	0.37	0.37	0.37	0.37	0	0.518	4.805
2-Cl	2.97E+00	0.22	0.24	0.576	0.576	0.55	0.518	4.805
4-F	1.84E+00	0.06	0.15	0.06	0.15	0	0.139	0.216
3-F	1.79E+00	0.34	0.34	0.34	0.34	0	0.139	0.216
2-F	2.41E+00	0.06	0.06	0.425	0.425	0.27	0.139	0.216
4-Br	2.06E+00	0.22	0.26	0.22	0.26	0	0.792	7.623
3-Br	2.58E+00	0.37	0.37	0.37	0.37	0	0.792	7.623
4-Me	2.42E+00	−0.14	−0.14	−0.14	−0.14	0	0.467	5.041
3-Me	2.70E+00	−0.06	−0.06	−0.06	−0.06	0	0.467	5.041
4-OMe	1.42E+00	−0.28	−0.12	−0.28	−0.12	0	−0.253	6.463
3-OMe	2.41E+00	0.1	0.1	0.1	0.1	0	−0.253	6.463
4-NO_2	2.16E+00	0.81	1.25	0.81	1.25	0	−0.047	7.324
3-NO_2	1.61E+00	0.71	0.71	0.71	0.71	0	−0.047	7.324
2-NO_2	8.13E−01	0.81	1.25	1.4	1.4	0.35	−0.047	7.324
2-F-6-Cl	2.07E+00	0.28	0.39	1.001	1.001	0.82	0.658	5.021
3,4-Cl_2	2.38E+00	0.59	0.59	0.59	0.59	0	1.036	9.609
3,4-F_2	1.50E+00	0.4	0.4	0.4	0.4	0	0.279	0.433
3,5-$(NO_2)_2$	1.06E+00	1.42	1.42	1.42	1.42	0	−0.093	14.649
2,4-$(NO_2)_2$	1.08E−01	1.62	2.5	2.21	2.21	0.35	−0.093	14.649
2-F-6-NO_2	7.02E−01	0.87	1.4	1.825	1.825	0.62	0.093	7.541
4-CF_3	1.68E+00	0.53	0.53	0.53	0.53	0	0.883	5.973
3-CF_3	1.92E+00	0.46	0.46	0.46	0.46	0	0.883	5.973
3,5-$(CF_3)_2$	1.80E+00	0.92	0.92	0.92	0.92	0	1.766	11.947
4-CN	9.48E−01	0.71	0.99	0.71	0.99	0	0.036	5.737
3-CN	6.14E−01	0.62	0.62	0.62	0.62	0	0.036	5.737
4-$CSNH_2$	5.03E−02	0.38	0.7	0.38	0.7	0	−0.347	16.693
3-$CSNH_2$	−3.26E−01	0.25	0.25	0.25	0.25	0	−0.347	16.693

Table 2. Calculated HOMO and LUMO Energies

X	5-Me-BIM		BIM		BOZ		BTZ	
	E_{HOMO}	E_{LUMO}	E_{HOMO}	E_{LUMO}	E_{HOMO}	E_{LUMO}	E_{HOMO}	E_{LUMO}
H	−8.148	−0.309	−8.243	−0.330	−8.508	−0.490	−8.367	−0.534
4-Cl	−8.228	−0.545	−8.326	−0.565	−8.594	−0.703	−8.454	−0.601
3-Cl	−8.220	−0.497	−8.319	−0.515	−8.586	−0.661	−8.445	−0.593
2-Cl	−8.150	−0.480	−8.242	−0.495	−8.509	−0.637	−8.366	−0.553
4-F	−8.229	−0.537	−8.327	−0.555	−8.596	−0.696	−8.456	−0.603
3-F	−8.231	−0.523	−8.331	−0.543	−8.599	−0.688	−8.457	−0.604
2-F	−8.174	−0.496	−8.263	−0.509	−8.530	−0.649	−8.388	−0.579
4-Br	−8.244	−0.607	−8.343	−0.625	−8.612	−0.760	−8.472	−0.637
3-Br	−8.228	−0.519	−8.327	−0.537	−8.594	−0.680	−8.453	−0.599
4-Me	−8.128	−0.303	−8.220	−0.323	−8.487	−0.479	−8.346	−0.517
3-Me	−8.134	−0.281	−8.228	−0.303	−8.494	−0.463	−8.351	−0.523
4-OMe	−8.125	−0.270	−8.216	−0.283	−8.482	−0.445	−8.345	−0.514
3-OMe	−8.129	−0.325	−8.273	−0.366	−8.488	−0.500	−8.398	−0.564
4-NO_2	−8.448	−1.407	−8.556	−1.421	−8.833	−1.506	−8.694	−1.395
3-NO_2	−8.391	−1.224	−8.501	−1.233	−8.777	−1.302	−8.637	−1.209
2-NO_2	−8.294	−1.195	−8.401	−1.214	−8.661	−1.301	−8.523	−1.180
2-F-6-Cl	−8.147	−0.675	−8.246	−0.689	−8.511	−0.814	−8.367	−0.690
3,4-Cl_2	−8.282	−0.714	−8.384	−0.734	−8.652	−0.857	−8.512	−0.732
3,4-F_2	−8.310	−0.768	−8.414	−0.788	−8.684	−0.912	−8.543	−0.783
3,5-$(NO_2)_2$	−8.602	−1.958	−8.720	−1.969	−9.006	−2.024	−8.870	−1.941
2,4-$(NO_2)_2$	−8.564	−2.080	−8.683	−2.103	−8.951	−2.171	−8.821	−2.060
2-F-6-NO_2	−8.279	−1.381	−8.385	−1.395	−8.657	−1.487	−8.521	−1.371
4-CF_3	−8.344	−0.880	−8.447	−0.897	−8.721	−1.016	−8.581	−0.888
3-CF_3	−8.308	−0.736	−8.414	−0.753	−8.685	−0.879	−8.544	−0.753
3,5-$(CF_3)_2$	−8.456	−1.160	−8.567	−1.177	−8.848	−1.281	−8.708	−1.154
4-CN	−8.321	−0.903	−8.423	−0.920	−8.694	−1.030	−8.555	−0.908
3-CN	−8.297	−0.723	−8.398	−0.735	−8.670	−0.855	−8.529	−0.735
4-$CSNH_2$	−8.279	−0.910	−8.430	−0.772	−8.620	−1.036	−8.514	−0.933
3-$CSNH_2$	−8.322	−0.757	−8.411	−0.588	−8.615	−0.847	−8.378	−0.772

cromethod for the determination of the minimum inhibitory concentration (MIC). All strains were obtained from the Czech National Collection of Type Cultures (CNCTC), with the exception of *M. kansasii* 6509/96, which was a clinical isolate. The MIC values of studied compounds are within a range of 2—1000 μmol/l for all tested strains, most often between 8—62 μmol/l. The MIC values were determined after incubation at 37 °C for 7 (*M. kansasii* My 235/80, *M. kansasii* My 6 509/96 only), 14, and 21 d. The MIC values for the used benzazoles were reported previously (see above).

Dataset and Parameters The MIC values were ordered in a matrix with 29 rows (substituents) and 40 columns (benzazoles×strains×time of incubation). For the MIC values marked with >, where the minimal inhibition concentration was not exactly determined, the matrix element represents double of the value (max. 1000). Due to such modifications leading to a statistical uncertainty of real values, the data set was analyzed entirely. The matrix elements were further logarithmized providing the log MIC values and the entire matrix as source matrix **SM**. For some calculations, the matrix was modified by the subtraction of MIC values for hydrogen substituent from all elements in the given column. This matrix was indicated as **ΔSM**.

The Hammett substituent constants σ_m, σ_p, and σ_p^-,[9] the substituent constants σ_{ortho},[10] and the steric substituent constants υ[11] were taken from literature. The substituent constants for $CSNH_2$ group were taken from the work of Waisser *et al.*[12,13] Four sets of substituent constants were compiled for data analysis (see Table 1). Whereas the substituent constants σ_p and σ_p^- (sets marked as σ and σ^-) were used for the substituents appended at the position 2 of the phenyl ring, known substituent constants[10] were used in the sets marked as σ_{ortho} and σ^-_{ortho}.

The energies E_{HOMO} and E_{LUMO} (Table 2) were quantum-chemically calculated on a PC computer using software HyperChem Suite for Windows (release 5.1) The semi-empirical AM1 (Austin Model 1) method was used for all molecular modelling calculations. The most stable conformations of the molecules were found by the molecular dynamics, by simulated heating to 7000 K followed by the Monte Carlo method (293 K, 500 steps). Calculations of log *P* values (Table 1) were carried out on the software HyperChem Suite for Windows (release 5.1), the module ChemPlus 1.6 using atomic parameters derived by Ghose *et al.*[14] and Viswanadhan *et al.*[15] The molar refraction (MR) was estimated by the same method as log *P*. Atomic contributions for these calculations were published by Ghose and Crippen[16] and Viswanadhan *et al.*[15] The log *P* and MR values were transformed to the increments Δ log *P* and ΔMR by the subtraction of the value corresponding to unsubstituted derivative (Table 1).

Statistical Analysis Several statistic methods were tested for the data treatment. However, the methods of exploratory analysis (EA), methods with latent variables (principal component analysis PCA and factor analysis FA), 2D-nonlinear mapping (NLM), and methods based on multiple linear regression (MLR) showed to be the most suitable. Polygons (EA), where the ray length corresponds with the log MIC values, provided the best insight into the data structure. Two-dimensional data projection in NLM has been calculated by nonlinear regression. The algorithm NIPAL[17] in Exner variation[18] has been used for the calculation of the latent variables in PCA. This algorithm involves modification of the coordinate origin for each column of the matrix of manifest variables during the calculation. Thus, it is possible to efficiently analyze non-standardized data with the same scale but with different origin in this way. The Thomphson's iterative process of correlation matrix (standardized data) has been used for the calculation of FA factors. Score matrix **T** and loadings matrix **P** were calculated by the data decomposition using methods with latent variables. Whereas the score matrix-derived parameters are in the following text marked with the subscript *t*, the loading matrix parameters have the subscript *p*. The most suitable explanatory variables have been selected by calculating all regressions with *k*, $(k-1), (k-2), \cdots$ of different explanatory variables from *n*-proposed (k=3, n=15 in our case) according criterion of the minimal residual standard deviation. Statistical significance of regression parameters was tested throughout the performed regressions, statistical nonsignificant parameters were omitted and the calculation was repeated until valid parameters were obtained. Program OPstat[19] was employed for all statistical calculations.

Results and Discussion

Exploratory Analysis Polygons characterizing the effect of substitution on the log MIC values are shown in Fig. 2. The polygon's rays represent benzazoles, strains, and time of incubation (7, 14, 21 d). Derivatives with the following substitution types showed the minimum polygon's area and, therefore, the highest antimycobacterial activity: 3,5-$(NO_2)_2$ (except for BTZ characterized by the area with longer rays), 2,4-$(NO_2)_2$ (except for 5-Me-BIM, My 330/88), 4-$CSNH_2$ (unambiguously the most effective), and partially also 4-NO_2 and 3-$CSNH_2$ (mainly BTZ).

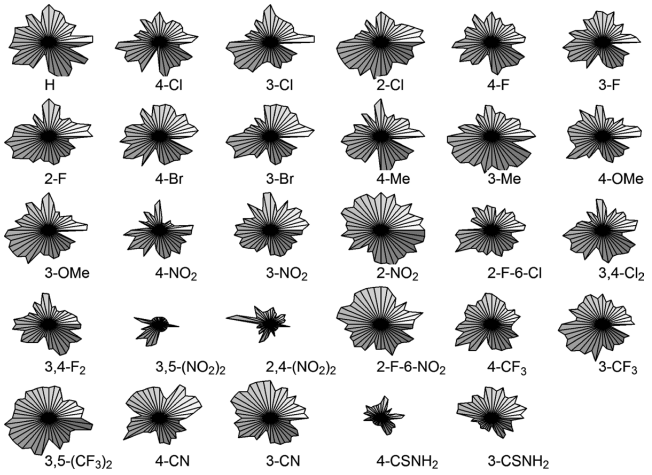

Fig. 2. Polygons Characterizing the Effect of Substitution on the log MIC Values

The polygon's rays correspond with the benzazoles, strains, and time of incubation (7, 14, 21 d).

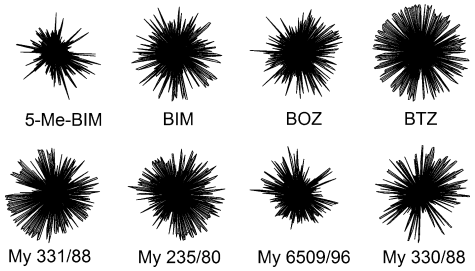

Fig. 3. Polygons Showing the Dependence of the log MIC Values on Benzazole Type (Upper) and on the Tested Strain (Lower)

The upper polygon's rays correspond with the substituents, strains, and time of incubation (14, 21 d). The lower polygon's rays correspond with the benzazoles, substituents, and time of incubation (14, 21 d).

Polygons showing the dependence of the log MIC values on benzazole type are shown in Fig. 3. The polygon's rays correspond with the substituents, strains, and time of incubation (14, 21 d). The polygon's area visually increase in the order of 5-Me-BIM<BIM~BOZ<BTZ. Hence, 5-methylbenzimidazole derivatives showed the highest antimycobacterial activity.

Figure 3 also depicts the polygons showing the dependence of log MIC values on the tested strain. The polygon's rays correspond with benzazoles, substituents, and time of incubation (14, 21 d). Fourth quadrants of the polygon correspond to the particular benzazoles in the order of 5-Me-BIM, BIM, BOZ, and BTZ. The polygon's area visually increases and, on the contrary, the sensitivity of the particular strains decreases in the order of *M. kansasii* 6509/96, *M. avium* My 330/88, *M. kansasii* My 235/80, and *M. tuberculosis* My 331/88. The shape of the polygons revealed a partial selectivity of the strains toward particular benzazoles. From this point of view, *M. kansasii* My 235/80 was the most sensitive strain toward BOZ (3rd quadrant), while *M. kansasii* 6509/96

showed sensitivity toward 5-Me-BIM (1st quadrant).

2D-Nonlinear Mapping Nonlinear mapping applied on the **ΔSM** matrix showed distribution of points in the area of two variables ×1 and ×2 (see Fig. 4). Application of implicit function around the points produced the clusters graphically. The point descriptors in the main cluster were omitted for clarity. From the obtained pattern is clear, that derivatives indicated as antimycobacterially active (3-$CSNH_2$, 2,4-$(NO_2)_2$, 4-$CSNH_2$, 3,5-$(NO_2)_2$, and partially 4-NO_2) are located in the right part of the diagram. Thus, the dimension ×1 represents apparently antimycobacterial activity. The substituents 2-NO_2, 2-F-6-NO_2, 3-CN, and 4-CN create own cluster in the upper part of the diagram. The low lipophilicity in a combination with the high substitution constant is the most probable reason why these substituents differ from others.

Analysis of Latent Variable The factor analysis (FA) seems to be the most suited tool for the analysis of the relation between properties (here substituents). The factor analysis of the transposed matrix $\mathbf{SM}^T$ based on the correlation matrix (standardized data) explained with one factor 56.6%, with two factors 65.8%, with three factors 76.0%, with four factors 79.2%, and with five factors 82.2% of the data variability. From these values is clear that for the description of the intrinsic data structure (relation between substituents) one or max. two factors are real. With respect to the uncertainty of some MIC values, the explained variability can be considered as sufficient.

The substituents, in the plane made of first two factor loadings, create an elongated band along the quadrant axis with two separated clusters as well as some isolated points. It is possible to clearly distinguish a main cluster created by the same substituents as in NLM, a cluster created by substituents 2-NO_2, 2-F-6-NO_2, 3-CN, and 4-CN, and more or less isolated points of 3-$CSNH_2$, 2,4-$(NO_2)_2$, 4-$CSNH_2$, and 3,5-$(NO_2)_2$. In comparison with NLM, a dimension reduction through FA does not provide new information. However, the NLM gave a better decomposition and insight into the data structure.

Multiple Linear Regression MLR.

Analysis of Potential Explanatory Variables The substituent constants σ, σ^-, σ_{ortho}, and σ^-_{ortho}, the steric substituent constants υ, $\Delta \log P$, ΔMR, energies E_{HOMO}, E_{LUMO}, and $E_{HOMO}-E_{LUMO}$ (see Dataset and parameters) were chosen as potential characteristics for the interpretation of the effect of substitution on the antimycobacterial activity of studied benzazoles. For the further data treatment, it was necessary to gain only one characteristic of the frontier orbital energy. These characteristics were calculated by PCA from the energies E_{HOMO}, E_{LUMO}, and $E_{HOMO}-E_{LUMO}$ (Table 2) as first latent variable (score). The explained variability of 98.8% for E_{HOMO}, 99.5% for E_{LUMO}, and 99.7% for $E_{HOMO}-E_{LUMO}$ showed that the effect of substitution is similar and can be described by global characteristics $E_{HOMO,t}$, $E_{LUMO,t}$, and $E_{HOMO}-E_{LUMO,t}$.

The relation between selected characteristics as explanatory variables has been analyzed by the factor analysis. Distribution of the characteristics in the area of first two rotated factors is shown in Fig. 5. A causal relation between the energies of frontier orbitals and the substituent constants is obvious. A saturation of the factors with these properties can also be clearly seen. As expected, the energies of frontier orbitals decrease with increasing values of the substitution constant. In fact, all of these quantities describe the same property. The second dimension is generated by the coefficient of distribution-derived characteristics, mainly $(\Delta \log P)^2$. The residual characteristics (ΔMR and υ) have a low communality with the first two factors and, therefore, they describe different properties. Whereas the ΔMR quantity is reciprocally related to $E_{HOMO,t}$, the steric substituent constant does not relate with other characteristics. With respect to the mutual relation of some characteristics and in order to suppress a possible multicollinearity, the characteristics of υ, $\Delta \log P$, $(\Delta \log P)^2$, and ΔMR in a combination with one of the substitution constant sets $-\sigma$, σ^-, σ_{ortho}, and σ^-_{ortho} were used for the multiple regression.

Construction of the Dependent Variable Vector For the multiple linear regression, it is necessary to gain only one dependent variable which characterize the dependence of the

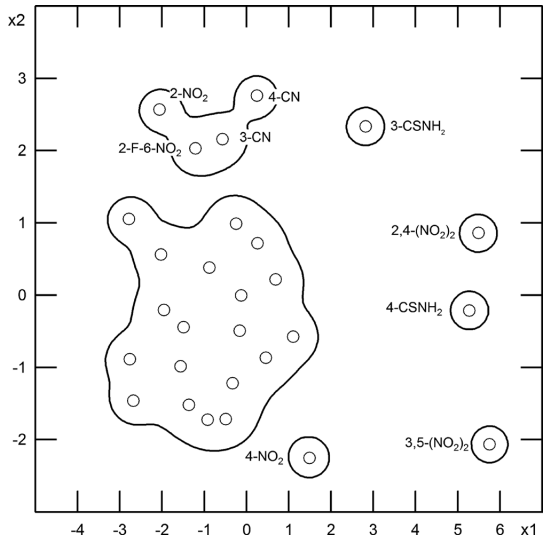

Fig. 4. Substituent Distribution in Two-Dimensional Projection Calculated by NLM

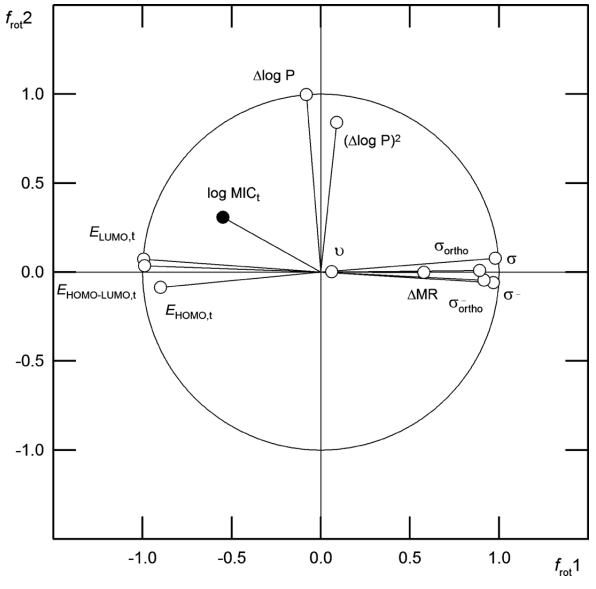

Fig. 5. Characteristic Distribution of Particular Subst. Derivatives in the Area of the First Two Rotated Factors (FA)

log MIC values on the substitution. Methods with latent variables are generally well suited for such purpose. The required quantity can be obtained as 1. score of the matrix in which the rows correspond to the substitution (subst. derivatives are objects) or as 1. loadings of the matrix in which the columns correspond to the substitution (substitution as a property). As we analyze the relation between activity and substitution, we have used the second approach as methodically correct.

Application of the PCA method on the transposed nonstandardized matrix $\mathbf{SM}^{T}$ provided loadings of the first principal component marked as $\log MIC_p$. These values represented 59.3% of the data variability (see also Analysis of Latent Variable).

The loadings calculated by the factor analysis seem to be a possible alternative (see Analysis of Latent Variable). However, this way calculated vector showed to be unsuitable for the multiple linear regression. The resulting loadings obtained by the FA calculation are significantly affected by the columns with high communality and the loadings have also cluster structure. As a result, the regression consists of a connecting line between the centre of large compact cluster and four remote points.

Interpretation of the Effect of Substitution Statistically closest dependence of the $\log MIC_p$ on the explanatory variables can be expressed by Eq. 1.

$$\log MIC_p = (1.78\pm0.17) - (0.61\pm0.19)\sigma^{-} + (1.78\pm0.43)\Delta\log P - (0.90\pm0.34)(\Delta\log P)^2$$
$$n=29,\quad s=0.567,\quad R=0.784 \qquad (1)$$

A similar result has been obtained with the set of substituent constants σ_{ortho} (s=0.567), σ (s=0.574), and σ^{-}_{ortho} (s=0.578). The analysis through Jackknife residuals of the dependence (1) indicated three outliers (4-NO_2, 3-OMe and 3-$CSNH_2$). After the outliers have been removed, the following closer relationship (2) has been obtained:

$$\log MIC_p = (1.77\pm0.14) - (0.70\pm0.14)\sigma^{-} + (1.73\pm0.36)\Delta\log P - (0.82\pm0.26)(\Delta\log P)^2$$
$$n=26,\quad s=0.395,\quad R=0.887 \qquad (2)$$

Again, a similar result has been obtained with the set of substituent constants σ_{ortho} (s=0.438), σ (s=0.426) and σ^{-}_{ortho} (s=0.422).

A bit less close is the relationship (3) which indicated a mutual interaction between the substitution and the coefficient of distribution.

$$\log MIC_p = (1.73\pm0.18) - (0.64\pm0.20)\sigma^{-} + (1.67\pm0.45)\Delta\log P - (1.44\pm0.64)\sigma\times\Delta\log P$$
$$n=29,\quad s=0.587,\quad R=0.767 \qquad (3)$$

The analysis through Jackknife residuals of the dependence (3) indicated the same three outliers as in Eq. 1 and, after their removal, closer relationship (4) has been obtained.

$$\log MIC_p = (1.75\pm0.15) - (0.74\pm0.15)\sigma^{-} + (1.64\pm0.35)\Delta\log P - (1.38\pm0.47)\sigma\times\Delta\log P$$
$$n=26,\quad s=0.402,\quad R=0.882 \qquad (4)$$

The last found significant dependence is represented by Eq. 5.

$$\log MIC_p = (2.12\pm0.22) + (0.82\pm0.23)\Delta\log P - (0.10\pm0.02)\Delta MR$$
$$n=29,\quad s=0.593,\quad R=0.750 \qquad (5)$$

After removal of only one outlier (3-OMe), the relationship can be expressed by Eq. 6.

$$\log MIC_p = (2.04\pm0.21) + (0.93\pm0.22)\Delta\log P - (0.10\pm0.02)\Delta MR$$
$$n=28,\quad s=0.555,\quad R=0.788 \qquad (6)$$

The above-mentioned equations unambiguously imply that the log MIC value decreases with increasing electron-withdrawing character of the substituent and with increasing molecule polarizability characterized by ΔMR. The influence of the lipophilicity is more complicated. In general, the log MIC values increase with the increasing $\Delta\log P$ values (less lipophilic molecules are more efficient) and this effect is dominant for this set of substituents. This finding also explains why the derivatives with thioamide group showed the highest antimycobacterial activity. The high lipophilicity among with strong electron-withdrawing substituent (*e.g.* 3,5-$(CF_3)_2$) caused a partial decrease of the log MIC values (increasing antimycobacterial activity). This is a consequence of parabolic character of the dependence of log MIC on $\Delta\log P$ and respective interaction of substitution effects and lipophilicity ($\sigma\times\Delta\log P$). The most probable reason can be seen in an easier penetration of these derivatives through the lipophilic membrane. However, these derivatives are not as active as less lipophilic derivatives bearing electron-withdrawing substituents. The dependence on the steric substituent constant υ was not statistically significant and, therefore, the *ortho* effect was most probably not important.

Conclusion

The multidimensional analysis methods proved to be a suitable tool for the evaluation of biological data loaded with high level of uncertainty on condition that the data is somehow structured, large and complete (1160 values in our case). Employing exploratory analysis, 2D-nonlinear mapping, methods with latent variables and multiple linear regression resulted in the following conclusions.

The antimycobacterial activity of 2-(subst. benzyl)sulfanyl derivatives of benzimidazole (BIM), 5-methylbenzimidazole (5-Me-BIM), benzoxazole (BOZ), and benzothiazole (BTZ) increased in the order of BTZ<BOZ~BIM<5-Me-BIM. The sensitivity of particular strains toward these compounds decreased in the order of *M. kansasii* 6509/96, *M. avium* My 330/88, *M. kansasii* My 235/80, and *M. tuberculosis* My 331/88. Whereas *M. kansasii* My 235/80 was the most sensitive strain toward BOZ, *M. kansasii* 6509/96 showed the highest sensitivity toward 5-Me-BIM.

In general, derivatives with 3-$CSNH_2$, 2,4-$(NO_2)_2$, 4-$CSNH_2$, 3,5-$(NO_2)_2$, and partially 4-NO_2 substituents possess the highest antimycobacterial activity. Derivatives with 2-NO_2, 2-F-6-NO_2, 3-CN, and 4-CN substituents showed nonsystematic variations in the MIC values that depended on the benzazole type as well as on the used strain and time of incubation. The low lipophilicity in a combination with the high substitution constant is the most probable reason why these substituents differ from others.

The effect of substitution was also described quantitatively with good correlation factor R of 0.79—0.88. The log MIC values depended with a negative slope on the Hammett substituent constants σ or molar refractions MR and, for the

given set of substituents, were dominantly raised with increasing $\log P$ value and partially lowered with $(\log P)^2$ or $\sigma \times \Delta \log P$. The derivatives featuring high polarizability, low lipophilicity and electron-withdrawing subtituents showed the highest antimycobacterial activity. The dependence on the steric substituent constant v was not statistically significant and, therefore, the *ortho* effect was most probably not important.

Acknowledgements This work was supported by the Ministry of Education, Youth and Sport of the Czech Republic (MSM 0021620822 and MSM 0021627501) and the Ministry of Public Health of the Czech Republic (IGA NS10367-3). We are indebted to PharmDr. Karel Palát, PhD (Department of Inorganic and Organic Chemistry, Faculty of Pharmacy, Charles University) for his help by finding of physicochemical and quantum-chemical parameters.

References

1) World Health Organization. http://www.who.int/tb, Tuberculosis Facts 2009.
2) Mitschison D. A., *Am. J. Resp. Crit. Care Med.*, **171**, 699—706 (2005).
3) Janin Y. L., *Bioorg. Med. Chem.*, **15**, 2479—2513 (2007).
4) O'Brien R. J., Spigelman M., *Clin. Chest. Med.*, **26**, 327—340 (2005).
5) Klimešovtá V., Kočí J., Pour M., Stachel J., Waisser K., Kaustovtá J., *Eur. J. Med. Chem.*, **37**, 409—418 (2002).
6) Klimešovtá V., Kočí J., Waisser K., Kaustovtá J., *Farmaco*, **57**, 259—265 (2002).
7) Klimešovtá V., Kočí J., Waisser K., Kaustovtá J., Möllmann U., *Eur. J. Med. Chem.*, **44**, 2286—2293 (2009).
8) Kočí J., Klimešovtá V., Waisser K., Kaustovtá J., Dahse H.-M., Möllmann U., *Bioorg. Med. Chem. Lett.*, **12**, 3275—3278 (2002).
9) Exner O., "Correlation Analysis of Chemical Data," Plenum Press, New York, 1988.
10) Pytela O., Liška J., *Collect. Czech. Chem. Commun.*, **59**, 2005—2021 (1994).
11) Charton M., "Steric Effect in Drug Design," Springer-Verlag, Berlin, 1983, pp. 57—91.
12) Waisser K., Macháček M., Labvoua J., Hrbata J., Držata J., *Collect. Czech. Chem. Commun.*, **53**, 2957—2961 (1988).
13) Polášek M., Waisser K., Bouček T., *Collect. Czech. Chem. Commun.*, **56**, 2964—2968 (1991).
14) Ghose A. K., Pritchett A., Crippen G. M., *J. Comput. Chem.*, **9**, 80—90 (1988).
15) Viswanadhan V. N., Ghose A. K., Revankar G. R., Robins R. K., *J. Chem. Inf. Comput. Sci.*, **29**, 163—172 (1989).
16) Ghose A. K., Crippen G. M., *J. Chem. Inf. Comput. Sci.*, **27**, 21—35 (1987).
17) Wold S., Geladi P., Esbensen K., Ohman J., *J. Chemometrics*, **1**, 41—56 (1987).
18) Exner O., *Collect. Czech. Chem. Commun.*, **41**, 1516—1540 (1976).
19) Pytela O., OPstat, version 6.23; http://webak.upce.cz/~koch/cz/veda/OPgm.htm.