Acta Poloniae Pharmaceutica – Drug Research, Vol. 66 No. 1 pp. 25–35, 2009 ISSN 0001-6837
Polish Pharmaceutical Society

APPLICATION OF THE STATISTICAL METHODS IN SYSTEMATIC STRUCTURE-ACTIVITY RELATIONSHIP ANALYSIS OF THIAZOLE, BENZOTHIAZOLE AND TETRAHYDROISOQUINOLINE DERIVATIVES WITH BIOLOGICAL ACTIVITY

ELŻBIETA O. BRZEZIŃSKA* and JUSTYNA D. STOLARSKA

Department of Analytical Chemistry, Medical University of Łódz, 1 Muszynskiego St., 90-151 Łódź, Poland.

Abstract: The procedure of initial studies to evaluate usefulness of the collected data for performing efficient QSAR analysis of any group of pharmacologically active substances was worked out. Structure-Activity Relationship (SAR) study on the thiazole derivatives with H_1-antihistamine activity, the benzothiazole derivatives with H_3-antihistamine activity and tetrahydroisoquinoline (TIQ) derivatives with β_2-adrenolytic activity, was performed by means of Factor Analysis (FA) and Regression Analysis (RA). The potential drug forms (neutral or ionic) in the particular physiological environment (pH) were used to examine the actual properties of the drug in its action place. It was determined based on their pK_a values. The usefulness of the group cases selected for QSAR studies was evaluated, based on the relationship between their structural variability and the variability of their specific biological activity.

Keywords: structure-activity relationship; factor and regression analysis; thiazoles; benzothiazoles; tetrahydroisoquinolines.

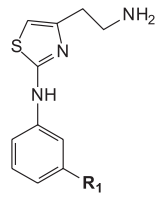
A structure of chemical compounds is a source of information about their biological activity. This information can be used to predict new compounds activity. The investigation of structural properties of the compounds with the known mechanism of pharmacological activity is always the source of new information about the targets and causes of their interactions with a drug. The most effective use of physicochemical data analysis is an optimization of the lead compound structures to obtain some derivatives with strongly required pharmacological profile – the hits.

The biologically active compounds most often represent the group of weak bases or weak acids. The ionization of drug molecules is important with regard to their absorption into the circulation and their distribution within different tissues of the body. Water as the environment of their interactions with biological target is conductive to un-ionized drug transformation into its ionized form. This fact causes changes in many properties and determines the pharmacokinetic and pharmacodynamic phase of drug activity. Differences between values of physicochemical descriptors of un-ionized and ionized species apply to electronic, hydrophobic and steric parameters. The probable drug form is the function of its dissociation constant (pK) and the pH of the environment in which it is located. The identification of a drug form in the particular physiological environment (pH) can be used to examine the actual properties of the drug in its action place.

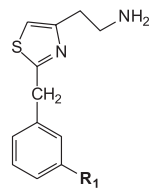
The basis of the effective QSAR analysis is the adequate selection of the representative group of investigated compounds and obtaining their properties in biological action environment.

The QSAR analysis includes most often active substances examined in the *in vitro* biological experimentation. It allows to observe the particular drug-receptor interaction. We can only use the pharmacological data from the *in vivo* experimentation when the strong relationship between behavioral effect and drug-receptor interaction exists. These limitations can be a cause of unfavorable selection of the examined compounds group. The structural variability observed in this group has to correlate with particular biological activity. Only such a group of compounds can be a basis for effective QSAR analysis. Physicochemical data are analyzed by use of many approachable mathematical methods. Well-chosen statistical methods and selected physicochemical data are the source of the precise solution of the problem.

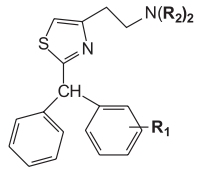
* Corresponding author: e-mail: ebrzezinska@pharm.am.lodz.pl



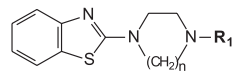
1, 4, 6-7, 11



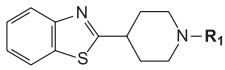
2, 3, 5, 8-10



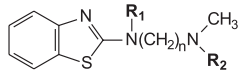
12-19



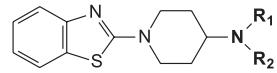
23, 29-30, 34, 38-39, 47-49, 51-54



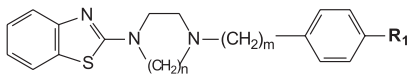
41, 50



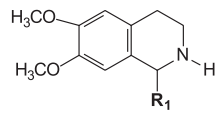
42-46



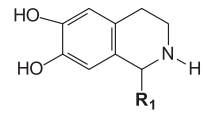
31, 33, 36-37, 40



20-22, 24-28, 32, 35



55-58, 61-63, 66



59-60, 64-65

Comp.	R_1	R_2	n	m	Comp.	R_1	R_2	n	m	Comp.	R_1	R_2	n
1	$-CH_3$				**23**	$-CH_2$-⬡		2		**45**	-H	$-CH_3$	2
2	$-OCH_3$				**24**	$-OCH_3$		3	2	**46**	$-CH_3$	$-C_2H_5$	2
3	-H				**25**	$-OCH_3$		2	2	**47**	⬡		2
4	-H				**26**	$-C(CH_3)_3$		2	2	**48**	$-CH_2(CH_2)_2CH_3$		2
5	-F				**27**	$-C(CH_3)_3$		2	1	**49**	$-CH_2CH_3$		3
6	-Cl				**28**	$-OCH_3$		3	1	**50**	$-C_3H_7$		
7	-F				**29**	$-CH_2CH(CH_3)_2$		2		**51**	$-CH_2$-◁		2
8	$-CH_3$				**30**	$-CH_2(CH_2)_3CH_3$		2		**52**	⬠		2
9	-Br				**31**	-H	-H			**53**	$-CH_2CH{=}CH_2$		2
10	-Cl				**32**	$-CH_3$		2	1	**54**	$-CH(CH_3)_2$		2
11	-Br				**33**	$-CH_3$	$-CH_3$			**55**	$-3,4-C_6H_3(OC_2H_5)_2$		
12	-4-Br	-H			**34**	$-CH_2(CH_2)_4CH_3$			2	**56**	$-3-C_6H_4(OCH_3)$		
13	-H	-H			**35**	$-OCH_3$		2	1	**57**	$-3,4-C_6H_3(OC_2H_5).(OH)$		
14	-H	$-CH_3$			**36**	$-C_2H_5$ $-C_2H_5$				**58**	$-2-C_6H_4(OCH_3)$		
15	-4-F	-H			**37**	$-CH_3$	$-C_3H_7$			**59**	$-3,4,5-C_6H_2(OCH_3)_3$		
16	-4-Cl	-H			**38**	$-CH_2(CH_2)_5CH_3$		2		**60**	$-2-C_6H_4C$		
17	-3-F	-H			**39**	$-CH_2$-⬠		2		**61**	$-C_6H_5$		
18	-3-Br	-H			**40**	$-CH_3$	$-C_2H_5$			**62**	$-2-C_6H_4Cl$		
19	-3-Cl	-H			**41**	$-CH_3$				**63**	$-4-C_6H_4(OCH_3)$		
20	$-C(CH_3)_3$		3	1	**42**	$-CH_3$	$-CH_3$	2		**64**	$-3-C_6H_4NO_2$		
21	-H		3	1	**43**	-H	$-CH_3$	3		**65**	$-3-C_6H_4(OCH_3)$		
22	-H		2	2	**44**	$-CH_3$	$-C_3H_7$	2		**66**	$-3,4,5-C_6H_2(OCH_3)_3$		

Figure 1. Molecular structures of the examined compounds

The possibility of obtaining reliable results of the planned QSAR analysis based on available examples of active compounds was investigated in the present study. The aim of the performed experiments was to work out the procedure of initial studies to evaluate usefulness of the collected data for performing efficient QSAR analysis of any group of pharmacologically active substances. The selected substances make the three sets of similarly built compounds with known biological activity that was determined in particular pharmacological tests. The thiazole, benzothiazole and tetrahydroisoquinoline derivatives represent compounds (see Fig. 1) with H_1-antihistamine (**1-19**), H_3-antihistamine (**20-54**) and β_2-adrenolytic (**55-66**) activity, respectively. The basis for the initial assessment is (i) determination of a probable form (neutral or ionic) of the compounds studied, at specific pH environmental conditions in which the measure of their biological activity was carried out; (ii) evaluation of usefulness of the cases group selected for QSAR studies basing on the relationship between their structural variability and variability of specific biological activity; (iii) setting structural assumptions for cases making up the group of compounds studied, when its assessment appears to be a negative; (iv) selection of the group of physicochemical data, determined for neutral or ionic forms, as an efficient basis in the planned QSAR analysis. It has been assumed that the methods used should have a universal character, allowing to assess the group of cases of any structure and activity.

Calculation

The molecular modeling was carried out using HyperChem 7.0 softwere package. All structures of the studied compounds were geometrically optimized by the use of semi-empirical method AM1 (Austin Model 1) with algorithm Polac-Ribiere, RMS grad = 0,01 kcal/(Å mol) *in vacuo*. The search of conformations was not done.

A set of additional quantum-chemical descriptors: distance between aliphatic and aromatic nitrogen atoms (**AB**); binding energy ($\mathbf{E_b}$); heat of formation ($\mathbf{H_f}$); the HOMO energy ($\boldsymbol{\varepsilon}_{\mathbf{HOMO}}$); the LUMO energy ($\boldsymbol{\varepsilon}_{\mathbf{LUMO}}$); electric charge focused on N-aliphatic atom (**Q**); electric charge focused on N-aromatic atom ($\mathbf{Q_{Ar}}$); dipole moment ($\boldsymbol{\mu}$) have also been obtained for each molecule after geometry optimization procedure for molecules after a low-energy conformation. The molecular descriptors such as: van der Waals surface area ($\mathbf{A_W}$); van der Waals volume ($\mathbf{V_W}$); partition coefficient (**log P**); molecular refractivity (**MR**); polarizability ($\boldsymbol{\alpha}$); the molecular weight logarithm (**log** $\mathbf{M_W}$); hydration energy ($\mathbf{H_h}$) have been calculated applying QSAR properties ChemPlus program (modules for HyperChem). Dissociation constants and distribution coefficients were carried out using ACD/Labs 8.0 program.

The Factor Analysis and the Regression Analysis were carried out using STATISTICA 7.0 program.

RESULTS AND DISCUSSION

The dissociation constant was determined for the group of benzothaizole derivatives by using spectrometric method as described (1).

The investigated substances make the set of 66 differently built compounds with known biological activity. The statistical analysis applied the values of biological activity ($\mathbf{pA_2}$) of the studied compounds determined in particular *in vitro* biological tests and reported previously by researchers who had synthesized the group (2-9). The compounds (see Fig. 1) represent three groups of pharmacological activity: H_1-antihistamine, H_3-antihistamine and β_2-adrenolytic activity. The synthesis method, analytical data and biological activity of the thiazole derivatives with H_1-antihistamine activity (group **I** – compounds **1-19**; $\mathbf{pA_2}$ = 4.0÷6.38) (2, 3); benzothiazole derivatives with H_3-antihistamine activity (group **II** – compounds **20-54;** $\mathbf{pA_2}$ = 4.40÷7.21) (4, 5); and tetrahydroisoquinolone (TIQ) derivatives with β_2-adrenolytic activity (group **III** – compounds **55-66;** $\mathbf{pA_2}$ = 3.0÷5.0) (6-9) have been described in previous papers.

Physicochemical parameters of compounds in groups **I-III** (cases **1-66**) were evaluated for un-ionized molecules: distance between aliphatic and aromatic nitrogen atoms (**AB**); binding energy ($\mathbf{E_b}$); heat of formation ($\mathbf{H_f}$); the HOMO energy ($\mathbf{e_{HOMO}}$); the LUMO energy ($\boldsymbol{\varepsilon}_{\mathbf{LUMO}}$); electric charge focused on N-aliphatic atom (**Q**); electric charge focused on N-aromatic atom ($\mathbf{Q_{Ar}}$); dipole moment ($\boldsymbol{\mu}$) have also been obtained for the each molecule after geometry optimization procedure for molecules after a low-energy conformation. The molecular descriptors such as: van der Waals surface area ($\mathbf{A_W}$); van der Waals volume ($\mathbf{V_W}$); partition coefficient (**log P**); molecular refractivity (**MR**); polarizability ($\boldsymbol{\alpha}$); the molecular weight logarithm (**log** $\mathbf{M_W}$); hydration energy ($\mathbf{H_h}$) (data not shown) and, in addition, for their mono-protonated forms in groups $\mathbf{I_a}$-$\mathbf{III_a}$ (cases **1a-66a**), respectively (data not shown). The dissociation constant ($\mathbf{pK_a}$) and distribution coefficient (**log D)** values were calculated

using a pK ACD/Labs 8.0 and log D ACD/Labs 8.0 programs.

Determination of physicochemical properties of the examined compounds in the particular environment (pH)

It was suggested that the QSAR analysis, carried out by using the molecular descriptors of the compounds in their probable forms (at particular pH), can be the most effective. The identification of the probable drug form in its action place is the function of its dissociation constant (K_a) and the pH of the environment in which it is located. The necessary data were collected by means of the semi-empirical method (pK ACD/Labs 8.0) after examining their credibility with spectrometric method (1). The calculation of the percentage of the investigated compounds ionization in its action place was carried out using an algorithm with $\mathbf{pK_a}$ values of the particular compound and pH of physiological environment.

The determination of the dissociation constant was carried out for the group of 12 benzothiazole derivatives (**29,34,37-38,41-42,46-48,50-51,54**) by using the spectrometric method (1). The results of spectrometric analysis were compared to the calculated pK_a values. Calculations of the $\mathbf{pK_a}$ values were performed using the "*approximated apparent constants*" algorithm.

In the case of di-protonated acid H_2A with two dissociation constants pK_{a1} and $pK_{a2,}$ we can calculate the concentration of ionized and un-ionized forms on the basis of the compound concentration and the pH value (10). The two calculated dissociation constants pK_{a1} and pK_{a2} were used in the calculation of mole fraction of ionized and un-ionized forms at pH = 7.4.

$$\frac{1}{x_2} = 1 + \frac{K_{a1}}{[H_3O^+]} + \frac{K_{a1}K_{a2}}{[H_3O^+]^2}$$

mole fraction of di-protonated form (NH^{2+})

$$\frac{1}{x_1} = \frac{[H_3O^+]}{K_{a1}} + 1 + \frac{K_{a2}}{[H_3O^+]}$$

mole fraction of mono-protonated form (NH^+)

$$\frac{1}{x_0} = \frac{[H_3O^+]^2}{K_{a1}K_{a2}} \quad \frac{[H_3O^+]}{K_{a2}}$$

mole fraction of un-ionized form.

x_1, x_2, x_0 – mole fractions of particular forms of the

Table 1. Factor loadings from FA of group I physicochemical parameters

Parameters	Factor loadings			
	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_3}$	$\mathbf{f_4}$
$\mathbf{E_b}$	-0.852722 *	-0.115138	-0.222003	-0.098817
$\mathbf{A_w}$	0.927587 *	0.169100	0.127025	0.082191
$\mathbf{V_w}$	0.933096 *	0.162564	0.119830	0.107356
$\mathbf{H_h}$	0.340997	0.419823	0.750498 *	-0.100442
logP	0.963459 *	0.060327	0.040425	0.215896
logD	0.946944 *	0.231031	0.016158	0.166810
MR	0.946863 *	0.157870	0.090496	0.114422
$\boldsymbol{\alpha}$	0.944228 *	0.160165	0.087192	0.111564
$\mathbf{logM_w}$	0.908596 *	0.241498	-0.087443	0.178085
$\boldsymbol{\varepsilon_{HOMO}}$	-0.270856	-0.906735 *	0.050417	-0.173589
$\boldsymbol{\varepsilon_{LUMO}}$	0.182945	0.295380	0.291018	-0.278285
$\boldsymbol{\mu}$	0.397231	0.281209	-0.478853	0.129646
$\mathbf{H_f}$	0.818970 *	-0.125824	0.021864	-0.165856
$\mathbf{Q_{Ar}}$	0.435688	0.762368 *	0.173959	-0.018523
Q	0.112170	-0.032709	0.945352 *	0.113770
$\mathbf{pK_a}$	-0.237366	0.634863	0.497372	-0.425958
AB	0.275531	0.054090	0.047177	0.891948

* – factor loadings > 0.70

Table 2. Eigenvalues and the percent of total variance explained by factors $\mathbf{f_{1-4}}$ from FA of group **I** physicochemical parameters

Factor (**f**)	Eigenvalue	% Total variance (TSV)	Cumulative eigenvalue	Cumulative %TSV
1	9.659743	56.82202	9.65974	56.82202
2	2.834637	16.67434	12.49438	73.49635
3	1.727276	10.16045	14.22166	83.65680
4	0.995885	5.85815	15.21754	89.351495

examined compounds. $K_{a1,}$ K_{a2} – dissociation constants of compounds **1-66** calculated by pK ACD/Labs 8.0 program, $[H_3O^+] = [H^+]$ at pH = 7.4

For a concentration of the examined compounds (c) of approximately 2×10^{-5} M [mean experimental concentration (1)], the concentration of particular ionized and un-ionized forms can be calculated as follows:

$$[H_nA] = c/(x_n^{-1})$$

$[H_nA]$ – the concentration of the ionized or un-ionized form

x_n – mole fractions of the particular form

From the obtained concentrations, the percentage participations of particular forms were calculated. In the case of group **I** (thiazole derivatives **1-19** with H_1-antihistamine activity) the participation of the mono-protonated form (at N atom of aliphatic amine) was dominant (94.89-97.00%). It was found for group **II** (benzothiazole derivatives **20-54** with H_3-antihistamine activity) that at pH = 7.4, 17 compounds exist as the mono-protonated form (at N atom of aliphatic amine) (82.85-97.73%), and 15 compounds as un-ionized forms (80.93-98.16%). In group **III** (tetrahydroisoquinoline derivatives **55-66** with β_2-adrenolytic activity) we can calculate four pK_a values at most. In the calculations, two first constants were used – $\mathbf{pK_{a1}}$ and $\mathbf{pK_{a2}}$. At pH = 7.4, all compounds exist as mono-protonated forms at the N aliphatic amine atom (76.39-93.10%).

The ionization of the molecules caused the modification of their physicochemical parameters. We observed the greatest changes in the group of electronic parameters of the investigated compounds. The values of descriptors: $\mathbf{e_{LUMO}}$, **Q**, $\mathbf{H_f}$, **m**, for un-ionized and mono-protonated forms of the particular compounds, showed a difference on a large scale (117-219%). So, the molecular descriptors of the un-ionized (**1-66**) and mono-protonated (**1a-66a**) forms were collected (see Table 1 and 2) to be used in further assays.

The environment of the drug action is very important with regard to its distribution coefficient (**log D**). The parameter was first determined by chromatographic method in the group of 12 compounds (11). Secondly, the obtained data were compared with the calculated (by computer method) **logD** values of the particular compounds. The significant relationship between experimentally (11) obtained logD [by the use of the pK_a from experiment (1)] and calculated by the logD ACD/Labs 8.0 program was found (R = 0.93; n = 12). The other physicochemical parameters, applied in further mathematical assays, were calculated by the HyperChem® 7.0 program. The collected data (not shown) represent all the groups of physicochemical parameters (hydrophobic, electronic and steric) applied in the classic QSAR analysis.

Evaluation of usefulness of the studied active substances group for QSAR analysis basing on the relationship of their structural variability to particular biological activity

The Factor Analysis (FA) belongs to the chemometric methods, useful in multidimensional data assays. The main applications of factor analytic techniques are: to reduce the number of variables and to detect structure in the relationship between variables (to classify variables). Application of FA together with the regression analysis (RA) as statistic methods gives possibility to construct mathematical models useful in predicting the biological activity of new compounds (12, 13). We applied the results of FA and the regression analysis to classify variables and to examine the adequate selection of the representative group of the investigated compounds. The RA based on the FA results can show the way of the possible modification of the group to increase variability of the significant structural properties. The QSAR analysis carried out for an adequately selected group of investigated compounds can give information to predict a similar activity of new compounds. The obtained significant relationships could be then the tool for the new active structures planning. FA should constitute an element of the initial study in order to establish possibilities to

obtain reliable results of the planned QSAR analysis based on available examples of active compounds.

The factorial analysis was carried out with the application of parameters calculated from un-ionized (**1-66**) and mono-protonated (**1a-66a**) structures of the examined compounds in the groups with different activity (**I-III, I_a-III_a**). As a result of the variables (physicochemical data) analysis, the factors (**f**) were determined. Their eigenvalues have described the participation of the particular factors in the explanation of the total structural variance of the observed cases in the groups. The significant factors and the cumulative percent of total structural variance which they explain were found. The factor loadings, as a correlation range (> 0.70) between the factor and variables, determined the significant parameters in the particular factor. In all cases the factor loadings were obtained by rotating the factor solution. The aim of this strategy was to find a simple structure i.e. factors which have high loadings of some variables and low loadings of the others. The simple structure showed the principal groups of the parameters which describe the mutability of the compounds properties related to their structure variability.

For the obtained factors, the measures of the individual cases characteristics (described by parameters representing the particular factor) were determined as factor scores (values). The factor scores forming novel independent variables ($\mathbf{f_{1,2..}}$), were applied in the RA. The biological activity ($\mathbf{pA_2}$) of the studied compounds was used as dependent variables. In the obtained relationships, the correlation coefficients R express the degree to which factors are related to the dependent variable. The coefficient of determination (R-square) is an indicator of how well the model fits to the data and expresses how much of the original variability of the cases' pharmacological potency has been explained by this relationship. In the afore mentioned analysis it was represented as the percent of the Overall Pharmacological Variability (**OPhV**). The evaluation of the analytical value of the studied compounds was based on comparison of these parameters with a range of structural variability explained by the factors – the percent of Total Structural Variance (**TSV**). The results of the analysis permitted to answer the question whether and to which degree the structural variability observed in a given group of active substances was associated with their specific biological activity.

There were 17 physicochemical parameters calculated from un-ionized form structures. Some of them could not be calculated from ionized form. Numerous independent variables describe a given phenomenon with strict inter-correlation among their values. It is recommended to group up these correlated independent variables by FA, before undertaking QSAR analysis of their values, and to attempt to create a novel variable, which is more suitable to describe a given phenomenon as a source of variability. This novel variable can reduce the number of similar variables to one factor (**f**). After the first factor has been extracted, the consecutive factor is extracted from the remaining variability. The consecutive factors are independent (orthogonal) of each other. In this way, the structure in the relationships between variables can be detected, that is, the classification of variables is performed.

The course of FA was presented in detail on the first discussed example related to group **I** (**1-19**).

Table 3. The relationship between H_1-antihistamine activity and the structural properties (factor scores of $\mathbf{f_{1-4}}$) of the un-ionized thiazole derivatives (group **I**)

Independent variables in equation pA_2 =	**R**	**p**	**%OPhV**	**%TSV**
$af_1 + b$	0.87	< 0.00000	75.69	56.82
$af_2 + b$	0.20	< 0.40070	4.00	16.67
$af_3 + b$	0.13	< 0.95700	1.69	10.16
$af_4 + b$	0.29	< 0.22680	8.41	5.86
$af_1 + bf_2 + c$	0.90	< 0.00000	81.00	73.49
$af_1 + bf_3 + c$	0.87	< 0.00001	75.69	66.98
$af_1 + bf_4 + c$	0.92	< 0.00000	84.64	62.68
$af_1 + bf_2 + cf_3 + d$	0.90	< 0.00001	81.00	83.65
$af_1 + bf_2 + cf_4 + d$	0.94	< 0.00000	88.36	79.35
$af_1 + bf_3 + cf_4 + d$	0.92	< 0.00000	84.64	72.84

Table 4. The relationship between H_1-antihistamine activity and the structural properties (factor scores of $\mathbf{f_{1\text{-}4}}$) of the mono-protonated thiazole derivatives (group I_a)

Independent variables in equation pA_2 =	**R**	**p**	**%OPhV**	**%TSV**
$af_1 + b$	0.93	< 0.00000	86.49	53.14
$af_2 + b$	0.05	< 0.84080	0.25	15.52
$af_3 + b$	0.19	< 0.42820	3.61	10.94
$af_4 + b$	0.05	< 0.82380	0.25	7.41
$af_1 + bf_2 + c$	0.93	< 0.00000	86.49	68.66
$af_1 + bf_3 + c$	0.95	< 0.00000	90.25	64.08
$af_1 + bf_4 + c$	0.93	< 0.00000	86.49	60.55
$af_1 + bf_2 + cf_3 + d$	0.95	< 0.00000	90.25	79.60
$af_1 + bf_2 + cf_4 + d$	0.93	< 0.00000	86.49	76.07
$af_1 + bf_3 + cf_4 + d$	0.95	< 0.00000	90.25	71.49

Table 5. The relationships between H_3-antihistamine activity and the structural properties (factor scores of $\mathbf{f_{1\text{-}5}}$) of the un-ionized benzothiazole derivatives (group **II**)

Independent variables in equation pA_2 =	**R**	**p**	**%OPhV**	**%TSV**
$af_1 + b$	0.72	< 0.00000	51.84	48.66
$af_2 + b$	0.02	< 0.90570	0.04	18.34
$af_3 + b$	0.18	< 0.28460	3.24	8.38
$af_4 + b$	0.13	< 0.47150	1.69	7.63
$af_5 + b$	0.15	< 0.33300	2.25	5.53
$af_1 + bf_2 + c$	0.72	< 0.00002	51.84	67.00
$af_1 + bf_3 + c$	0.74	< 0.00000	54.76	57.04
$af_1 + bf_4 + c$	0.73	< 0.00000	53.29	56.29
$af_1 + bf_5 + c$	0.73	< 0.00000	53.29	54.19
$af_1 + bf_2 + cf_3 + d$	0.74	< 0.00002	54.76	75.38
$af_1 + bf_2 + cf_4 + d$	0.73	< 0.00001	53.29	74.63
$af_1 + bf_2 + cf_5 + d$	0.73	< 0.00003	53.29	72.53
$af_1 + bf_3 + cf_4 + d$	0.75	< 0.00000	56.25	64.67
$af_1 + bf_3 + cf_5 + d$	0.76	< 0.00000	57.76	62.57
$af_1 + bf_2 + cf_3 + df_4 + e$	0.75	< 0.00002	56.25	83.01
$af_1 + bf_2 + cf_3 + df_5 + e$	0.76	< 0.00003	57.76	80.91
$af_1 + bf_2 + cf_4 + df_5 + e$	0.74	< 0.00003	54.76	80.16
$af_1 + bf_3 + cf_4 + df_5 + e$	0.77	< 0.00000	59.29	70.20

Examination of the thiazole derivatives with H_1-antihistamine activity

Factor Analysis of the parameters calculated from un-ionized forms of studied compounds – group **I**

In the case of group **I** FA four significant factors were founded ($\mathbf{f_1}$-$\mathbf{f_4}$), which explain 89.51% of the overall variance. The factor loadings and eigenvalues of factors, after rotating the factor solution, are shown in Tables 1 and 2.

In the group of the investigated compounds (**1-19**), $\mathbf{f_1}$ factor explaining 56.82% of the total structural variance (**TSV**) collects some parameters with strong factor loadings. The parameters forming $\mathbf{f_1}$ are: $\mathbf{E_b}$; $\mathbf{A_W}$; $\mathbf{V_W}$; **log P**; **log D**; **MR**; $\boldsymbol{\alpha}$; $\mathbf{\log M_W}$; $\mathbf{H_f}$.

Table 6. The relationships between H_1-antihistamine activity and the structural properties (factor scores of $\mathbf{f_{1\text{-}5}}$) of the mono-protonated benzothiazole derivatives: **20a**, **21a**, **24a**, **28a**, **31a**, **33a**, **36a**, **37a**, **40a**-**46a**, **49a**, **50a** (from group $\mathbf{II_a}$)

Independent variables in equation pA_2 =	**R**	**p**	**%OPhV**	**%TSV**
$af_1 + b$	0.78	< 0.00021	60.84	57.84
$af_2 + b$	0.11	< 0.66810	1.21	14.98
$af_3 + b$	0.32	< 0.20330	10.24	8.95
$af_4 + b$	0.41	< 0.87530	16.81	6.82
$af_1 + bf_2 + c$	0.79	< 0.00002	62.41	72.82
$af_1 + bf_3 + c$	0.85	< 0.00000	72.25	66.79
$af_1 + bf_4 + c$	0.78	< 0.00000	60.84	64.66
$af_1 + bf_2 + cf_3 + d$	0.85	< 0.00002	72.25	81.77
$af_1 + bf_2 + cf_4 + d$	0.79	< 0.00001	62.41	79.64
$af_1 + bf_3 + cf_4 + d$	0.85	< 0.00001	72.25	73.61

Table 7. The relationships between β_2-adrenolytic activity and the structural properties (factor scores of $\mathbf{f_{1\text{-}3}}$) of the un-ionized TIQ derivatives (group **III**)

Independent variables in equation pA_2 =	**R**	**p**	**%OPhV**	**%TSV**
$a + b\ f_1$	0.54	< 0.07161	29.16	50.05
$a + b\ f_2$	0.08	< 0.8052	0.64	20.17
$a + b\ f_3$	0.15	< 0.6334	2.25	14.86
$a + b\ f_1 + c\ f_2$	0.54	< 0.2073	29.16	70.22
$a + b\ f_1 + c\ f_3$	0.56	< 0.1854	31.36	64.91

The $\mathbf{f_1}$ describes the variability of steric and hydrophobic characteristics of compounds. The second factor – $\mathbf{f_2}$ explaining 16.67% of **TSV** was built by descriptors: $\mathbf{e_{HOMO}}$ i $\mathbf{Q_{Ar}}$, that is the electronic characteristics. The third factor – $\mathbf{f_3}$ includes electronic parameters **Q**; $\mathbf{H_h}$ and explains 10.16% of **TSV**. Last factor – $\mathbf{f_4}$, explaining 5.86% of **TSV,** is represented by the spatial descriptor **AB**. Classification of the variables obtained in FA shows the principal groups of parameters, which describe structural variability of the investigated group **I**. **TSV** of compounds **1-19** corresponds to their steric, hydrophobic (and less intense) electronic mutability. First, the factor scores of $\mathbf{f_1}$-$\mathbf{f_4}$ were obtained for all cases of group **I**. Then, the RA was done by the use of **1-19** H_1-antihistamine activity ($\mathbf{pA_2}$ = 4.0÷6.38) as a dependent variable. The results are summarized in Table 3.

The results of this analysis answer the question whether the structural mutability observed in a given group of cases affects particular biological activity.

In the case of the investigated group **I**, we observe strong correlations (**R**) in univariate relationships involving the first factor $\mathbf{f_1}$ and in multivariate relationships involving $\mathbf{f_1}$ and one or two remaining factors. The correlation coefficients **R** for the significant relationships were 0.87-0.94. The indicator of fitness – **%OPhV** in particular relationships was higher than the cumulative structural variability – **%TSV** explained by the factor or factors participating in a given relationship, except for the model involving three factors – $\mathbf{f_{1\text{-}3}}$ but the difference is very small (81% and 84%).

On this basis, we can say that H_1-antihistamine activity in group **I** is related to the structural variability of investigated cases **1-19**. The percent of variances not explained by factors $\mathbf{f_{1\text{-}4}}$ accounts for the remaining random variability that does not affect this activity. The collected cases are a well selected representative group for the given biological activity of group **I**. Group **I** can be used in adequate QSAR analysis, based on physicochemical data from un-ionized forms of compounds **1-19**. This group does not have to be complete.

Factor Analysis of the parameters calculated from mono-protonated forms of studied compounds – group $\mathbf{I_a}$

In the case of group $\mathbf{I_a}$ FA the 4 significant factors with eigenvalues ł 1 were found ($\mathbf{f_1}$-$\mathbf{f_4}$), which explained 86.02% of the overall variance. The first factor $\mathbf{f_1}$ explaining 53.14% of **TSV** was built by the following variables: $\mathbf{E_b}$, $\mathbf{A_W}$, $\mathbf{V_W}$, **log P**, **MR**, **a**, **log** $\mathbf{M_W}$. The factor describes the mutability of the steric and hydrophobic characteristics. The second factor $\mathbf{f_2}$ explains 15.52% of **TSV** and is built on the basis of mutability of electronic parameters: $\mathbf{H_h}$, $\mathbf{e_{LUMO}}$, **Q**. The factor $\mathbf{f_3}$ is represented by one electronic parameter $\mathbf{e_{HOMO}}$ and explains 10.94% of **TSV**. The last factor $\mathbf{f_4}$ explains only 7.41% **TSV** and involves the electronic parameter $\mathbf{Q_{Ar}}$.

The analysis enabled to identify 4 primary specific property areas that can be described as feature variability. These are the steric, hydrophobic and electronic properties. The factor scores were obtained and used as independent variables in regression analysis. In Table 4, the results of the analysis are collected.

In the regression analysis of mono-protonated forms of the thiazole derivatives ($\mathbf{1_a}$-$\mathbf{19_a}$) carried out by the use of factor scores, we have obtained the significant correlation coefficients R = 0.92-0.95. The correlations were higher than those from the group **I** analysis. The **%OPhV** indicators in the particular relationships were higher (86.49-90.25%) than the appropriate **%TSVs** explained by the factor or factors (53.14-79.60%). The differences between **%OPhVs** and **%TSVs** were the biggest. So, the structural mutability observed in group $\mathbf{I_a}$ describes the H_1-antihistamine activity. We can say that the data from mono-protonated structures **1a-19a** are a better source of information about the pharmacological mutability of the investigated thiazole derivatives than the data from un-ionized structures **1-19**. It should be remembered that nearly all thiazole derivatives occurred in the biological study environment in the mono-protonated form.

Examination of the benzothiazole derivatives with H_3-antihistamine activity

Factor analysis of the parameters calculated from un-ionized forms of the studied compounds – group **II**

In the group of 35 examined benzothiazole derivatives of group **II** (un-ionized forms **20-54**) with H_3-antihistamine activity, the FA was carried out. Five significant factors were found. The fifth factor $\mathbf{f_5}$ has eigenvalue < 1, but the **Q** parameter which represents this factor plays a significant role in the ligand binding to H_3-receptor. The obtained factors $\mathbf{f_{1\text{-}5}}$ explain 88.54% of the total structural variance of group **II.** The first factor – $\mathbf{f_1}$ describes 48.66% of **TSV** and includes such steric and hydrophobic parameters as: $\mathbf{E_b}$, $\mathbf{A_W}$, $\mathbf{V_W}$, **log P**, **MR**, **α**, **log** $\mathbf{M_W}$. The second factor – $\mathbf{f_2}$ explains 18.34% of **TSV.** The factor is built by electronic parameters: $\mathbf{e_{HOMO}}$, $\mathbf{e_{LUMO}}$, $\mathbf{Q_{Ar}}$. The factors $\mathbf{f_3}$, $\mathbf{f_5}$ explain 8.38% and 5.53% of **TSV,** respectively. The factors represent some electronic properties of compounds: $\mathbf{H_h}$ and **Q**. The factor $\mathbf{f_4}$ with electronic – $\mathbf{pK_a}$ and spatial – **AB** parameters explains only 7.63% of **TSV.** The factor scores (as independent variables) and H_3-antihistamine activity ($\mathbf{pA_2}$ = 4.40÷7.21) of **20-54** (as a dependent variable) were used in the regression analysis. The results are summarized in Table 5.

In the case of the investigated group **II** it was determined that in the univariate relationship with $\mathbf{f_1}$ and in the multivariate relationships including $\mathbf{f_1}$, the correlation coefficients were significant – **R** = 0.72-0.77. The percent of overall pharmacological variance explained by the proposed models was found as **%OPhV** = 51.84-59.29%, and in particular relationships was smaller than the cumulative structural variability – **%TSV** explained by the factors participating in a given equation (48.66-83.01%), except for the univariate relationship with $\mathbf{f_1}$. The structural variability observed in group **II** is only partially relevant to H_3-antihistamine activity of cases. The investigated group **II** is not a representative group of cases for the adequate QSAR analysis. We have observed some effect of steric and hydrophobic parameters on the H_3-antihistamine activity of **20-54**. The parameter **Q** is probably important too. So we can suggest that group **II** should be supplemented by some cases with different alkyl substitutes of the aliphatic amine, what can increase the steric and hydrophobic properties variability. The electric charge focused on N-aliphatic atom (**Q**) can be also changed. The further investigations of group **II** can be realized because in many cases the **TSV** values are comparable with the pharmacological variability of the group. However, less satisfactory results of the statistical QSAR analysis can be expected as compared to group **I.**

Factor Analysis of the parameters calculated from cases **20a, 21a, 24a, 28a, 31a, 33a, 36a, 37a, 40a-46a, 49a, 50a** mono-protonated at pH = 7.4

The results of the pK_a values calculation of the group **II** have shown that some of the compounds (**20, 21, 24, 28, 31, 33, 36, 37, 40-46,49, 50**) exist as mono-protonated forms at pH = 7.4. We have made a decision to carry out the factor analysis of these compounds group. In the analysis we have used the

data calculated from mono-protonated forms **20a, 21a, 24a, 28a, 31a, 33a, 36a, 37a, 40a-46a, 49a, 50a.**

The first factor $\mathbf{f_1}$ includes such steric, hydrophobic and electronic parameters as: $\mathbf{E_b}$, $\mathbf{A_W}$, $\mathbf{V_W}$, **log P**, **MR**, α, **log** $\mathbf{M_W}$, $\mathbf{e_{LUMO}}$ μ, and explains 57.84% of **TSV**. The second factor $\mathbf{f_2}$ with parameters: $\mathbf{e_{HOMO}}$-$\mathbf{e_{LUMO}}$, **AB,** explains 14.98% of **TSV**. The factors $\mathbf{f_3}$ – with parameter **Q** and $\mathbf{f_4}$ – with parameter $\mathbf{Q_{Ar}}$ represent some electronic properties of the cases and explain 8.95% and 6.82% of **TSV,** respectively. After the determination of factor scores, the regression analysis was carried out.

In the case of the group represented by 17 benzothiazole derivatives, which exist at pH = 7.4 as mono-protonated forms, the significant correlation coefficients in univariate relationship with $\mathbf{f_1}$ and in multivariate relationships including $\mathbf{f_1}$ were 0.78-0.85. The indicator of fitness – **%OPhV** in some relationships was higher than the cumulative structural variability – **%TSV** explained by the factor or factors participating in a given relationship. On this basis, we can say that the H_3-antihistamine activity in the group is related to the structural variability of investigated cases **20a, 21a, 24a, 28a, 31a, 33a, 36a, 37a, 40a-46a, 49a, 50a.**

Application of the case selection criterion enabled to obtain a more representative group. Evidently, the improvement in the analysis quality is also associated with reduction of the cases number, although the conclusion on a better adjustment of the group studied is based on equalization of proportions between the indices of this adjustment – OPhV and TSV. The selected group of cases mentioned above can become the basis for the QSAR analysis.

Examination of the tetrahydroisoquinoline derivatives with β_2-adrenolytic activity

Factor Analysis of the parameters calculated from un-ionized forms of studied compounds – group **III**

In the group of tetrahydroisoquinoline derivatives (TIQ) with β_2-adrenolytic activity (compounds **55-66**) the FA was carried out and three significant factors were determined. The factors explain 85.08% of structural variability **TSV** in the group **III**. Not the whole group of physicochemical descriptors can be used in the analysis, because of the small group of used cases. The first factor $\mathbf{f_1}$ explaining 50.05% of **TSV** was built by hydrophobic parameters: $\mathbf{E_b}$; $\mathbf{A_W}$; **log** $\mathbf{M_W}$. The second factor $\mathbf{f_2}$, which explains 20.17% of variability, represents the mutability of electronic characteristic (ε_{HOMO}-ε_{LUMO}, μ). The third factor $\mathbf{f_3}$ describes the electronic characteristic of compounds and was represented by **Q** parameter. It explains 14.86% of **TSV**. From the determined factors we have calculated the factor scores. Values of factor scores were compared with β_2-adrenolytic activity ($\mathbf{pA_2}$ = 3.0÷5.0) of the investigated compounds (see results in Table 7).

All relationships observed in the regression analysis were insignificant. The correlation coefficients were very low (**R** = 0.54-0.56). The overall pharmacological variability of the investigated compounds was explained by the models at the level 29.16-31.36%. Such results are incomparable with appropriate **%TSV** that was explained by the factors used in the equations. The structural mutability observed in the group **III** is limited and only partially connected with β_2-adrenolytic activity. The examined group of compounds is not representative. It can be expected that further investigations on the QSAR analysis will provide unsatisfactory results. The necessity of enlarging the studied group in relation to variability of compound parameters $\mathbf{A_W}$ and **log** $\mathbf{M_W}$ has been found while analyzing the regression results. Due to $\mathbf{f_3}$ importance, the increase in the number of derivatives differently placed on the nitrogen atom of isoquinoline (effect on **Q**, $\mathbf{A_W}$ and **log** $\mathbf{M_W}$) might be beneficial.

Factor Analysis of the parameters calculated from mono-protonated forms of studied compounds – group $\mathbf{III_a}$

Next, the factor analysis, by the use of the variables calculated from the structures of mono-protonated forms ($\mathbf{55_a}$-$\mathbf{66_a}$), was carried out. Three significant factors with eigenvalues ≥ 1 were obtained. The total structural variability in the group $\mathbf{III_a}$ was 88.22%. The factors: $\mathbf{f_1}$ – with $\mathbf{e_{HOMO}}$ and $\mathbf{e_{LUMO}}$ electronic parameters; $\mathbf{f_2}$ – with **log P** and μ hydrophobic parameters and; $\mathbf{f_3}$ – with **Q** electronic parameter explain: 57.85%, 15.92%, and 14.48% of **TSV,** respectively. The RA was carried out by the use of factor scores and $\mathbf{pA_2}$ values, but the results were worse than those obtained for the cases **55-66**. The significant relationships were not determined (**R** = 0.29-0.55). The structural variability observed in this group $\mathbf{III_a}$ is only partially connected with β_2-adrenolytic activity of the cases. The investigated group $\mathbf{III_a}$ should not be used in planned QSAR analysis. The use of the data calculated from mono-protonated forms of investigated compounds **55-66** did not improve the quality of FA results.

CONCLUSION

It is evident from this study that the FA and the regression analysis can be used to predict the use-

fulness of the active substance groups for carrying out the planned QSAR analysis. Optimum fitting of the group **I** and $\mathbf{I_a}$ to represent of H_1-antihistamine activity was determined. The compounds **1-19** can be pharmacologically optimized on the basis of physicochemical parameters determined from their un-ionized and mono-protonated forms as well. The structural variability of **1-19** is well connected with their pharmacological activity. The groups **I** and $\mathbf{I_a}$ are well selected and can be useful for adequate QSAR analysis.

The benzothiazole derivatives with H_3-antihistamine activity (comp. **20-54**) represent the variability of their biological activity only partially. Probably, a whole group **II** or $\mathbf{II_a}$ will not be useful for adequate QSAR analysis, because of diverse behavior of compounds **20-54** in the physiological environment. The result of the analysis suggests supplementing group **II** by some cases with different alkyl substitutes of the aliphatic amine. It can improve the steric and hydrophobic properties variability. The electric charge focused on N-aliphatic atom (**Q**) can be also changed. We can say that the selected group of cases **20a, 21a, 24a, 28a, 31a, 33a, 36a, 37a, 40a-46a, 49a, 50a** can be useful for adequate QSAR analysis.

The structural variability in the group **III** with β_2-adrenolytic activity is limited and only partially explains variability of biological activity compounds **55-66**. The investigated group of compounds is not well selected. It can be expected that further QSAR study will give unsatisfactory results. The outcome of the study suggests supplementing group **III** by some cases with different substitutes of the aliphatic amine.

Acknowledgments

This study was supported by the Medical University of Łódź, Poland, Research Programme No. 502 13 232.

REFERENCES

1. Brzezińska E., Stolarska, J.: Annales UMCS XVII, 22, 187 (2004).
2. Walczyński K., Guryn R., Zuiderveld O.P., Zhang M.Q., Timmerman H.: Farmaco 55, 569 (2000).
3. Walczyński K., Timmerman H., Zuiderveld O.P., Zhang M.Q., Glinka R.: Farmaco 54, 533 (1999).
4. Walczyński K., Guryn R., Zuiderveld O.P., Zhang M.Q., Timmerman H.: Farmaco 54, 694 (1999).
5. Walczyński K., Guryn R., Zuiderveld O.P, Timmerman H.: Arch. Pharm. Pharm. Med. Chem. 332, 389 (1999).
6. Brzezińska E.: Acta Pol. Pharm. 51, 137 (1994).
7. Brzezińska E.: Acta Pol. Pharm. 53, 25 (1996).
8. Brzezińska E.: Acta Pol. Pharm. 53, 365 (1996).
9. Brzezińska E., Venter, D., Glinka R.: Pharmazie 51, 397 (1996)
10. Hulanicki A.: Reactions of acids and bases in analytical chemistry (in Polish), PWN, Warszawa, 1992
11. Brzezińska E.: Stolarska, J.: J. Planar Chromatogr. 18, 443 (2005).
12. Brzezińska E.: Acta Pol. Pharm. Drug Res. 60, 3 (2003).
13. Leonard J.T., Roy K.: QSAR Comb. Sci. 20, 23 (2004).

Received: 02. 06. 2008