SAR and QSAR in Environmental Research, 2002 VOL. 13 (3-4), pp. 457–471

Taylor & Francis
Taylor & Francis Group

COMPUTER AIDED PREDICTION OF BIOLOGICAL ACTIVITY SPECTRA: EVALUATING VERSUS KNOWN AND PREDICTING OF NEW ACTIVITIES FOR THIAZOLE DERIVATIVES*

A. GERONIKAKI[a,†], A. LAGUNIN[b], V. POROIKOV[b], D. FILIMONOV[b], D. HADJIPAVLOU-LITINA[a] and P. VICINI[c]

[a]*Aristotelian University of Thessaloniki, School of Pharmacy, Department of Pharmaceutical Chemistry, Thessaloniki 54006, Greece;* [b]*Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, 10, Pogodinskaya Street, Moscow, 119992, Russia;* [c]*University of Parma, Pharmaceutical Department, Parma, Parco Area delle Scienze, 27A, 43100 Italy*

(Received 12 July 2001; In final form 5 December 2001)

Computer aided prediction of biological activity spectra by the computer program PASS was applied to a set of 89 new thiazole derivatives. Experimentally tested activities (NSAID, local anaesthetic and antioxidant) coincide with the experiment in 70.8% cases, that exceeds significantly the random guess-work (~0.1%). Therefore, computer aided prediction using the Prediction of Activity Spectra for Substances (PASS) system (http://www.ibmh.msk.su/PASS) provides a reliable basis for planning of synthesis and experimental study for new compounds. New psychotropic activities are predicted for some compounds from the series under study. In particular, 7, 44 and 55 compounds likely have anxiolytic, anticonvulsant and cognition enhancer effects, respectively. Most of these compounds have the estimated values of probability to be active (P_a) less than 60%. Therefore, if their activity will be confirmed by the experiment, they might occur to be New Chemical Entities.

Keywords: Biological activity spectra; Computer-aided prediction; Thiazole derivatives; NSAID; Local anaesthetic; Antioxidant

INTRODUCTION

During the investigation of new pharmacological substances, each kind of biological activity is not revealed at once. Some of them are discovered later in clinical trials as "side effects". The biological data of each compounds are usually incomplete. The reason for this is the "directed testing". Thus, many existing activities could be possible to be uncovered.

*Presented at CMTPI 2001, Computational Methods in Toxicology and Pharmacology Integrating Internet Resources, July 11–13, 2001, Bordeaux, France.

†Corresponding author. E-mail: geronik@pharm.auth.gr

ISSN 1062-936X print/ISSN 1029-046X online © 2002 Taylor & Francis Ltd
DOI: 10.1080/10629360290014322

TABLE I Structures and predicted activities for thiazole derivatives under study

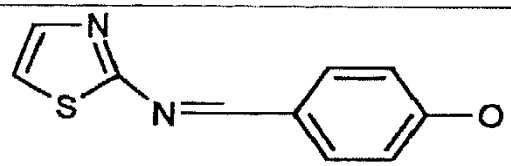
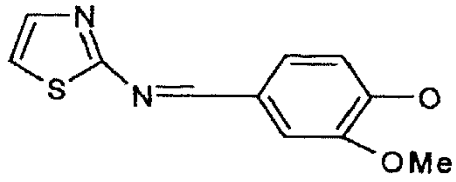
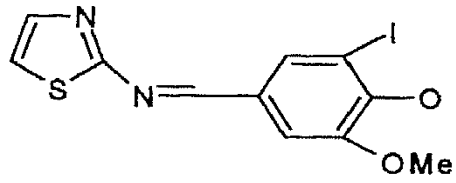
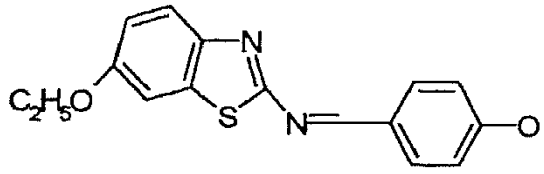
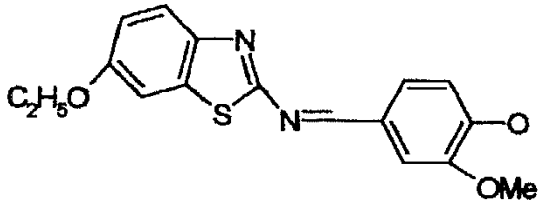
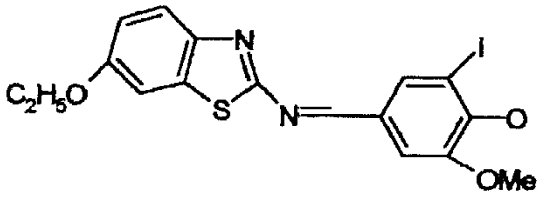
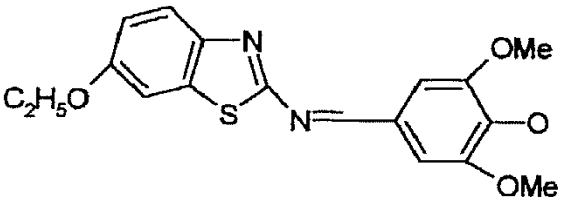
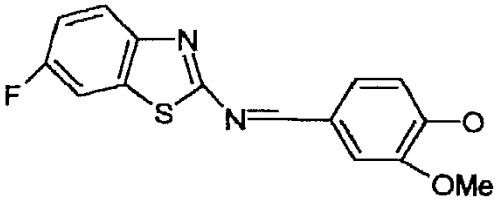
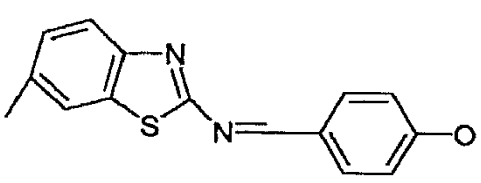
No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
51		Nucleotide metabolism regulator	70	2
		Cognition disorders treatment	66	1
		Fibrinolytic	64	5
		TNF convertase inhibitor	52	1
52		Nucleotide metabolism regulator	68	2
		Fibrinolytic	63	5
		Cognition disorders treatment	58	2
		5 Lipoxygenase inhibitor	53	2
53		Nucleotide metabolism regulator	63	3
54		Nucleotide metabolism regulator	70	2
		Glutamate release inhibitor	59	1
		Cognition disorders treatment	52	3
55		Nucleotide metabolism regulator	69	2
		Glutamate release inhibitor	56	1
56		Nucleotide metabolism regulator	65	3
		Sodium channel blocker	60	2
		Glutamate release inhibitor	53	2
		Histamine release stimulant	52	4
		ATPase inhibitor	52	4
57		Nucleotide metabolism regulator	68	2
		Glutamate release inhibitor	57	1
58		Nucleotide metabolism regulator	70	2
		Glutamate release inhibitor	57	1
		Myocardial ischemia treatment	59	7
		Psychotropic	54	4
59		Nucleotide metabolism regulator	74	1
		Glutamate release inhibitor	59	1
		Antiulcerative	55	2
		Cognition disorders treatment	55	2
		ATPase inhibitor	55	3
		Antimycobacterial	51	1
		Anti-Helicobacter pylori	50	1

TABLE I – *continued*

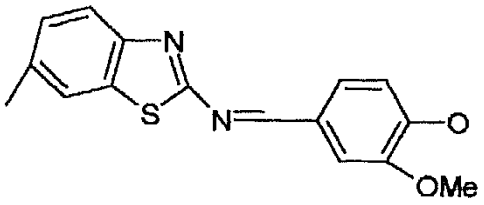
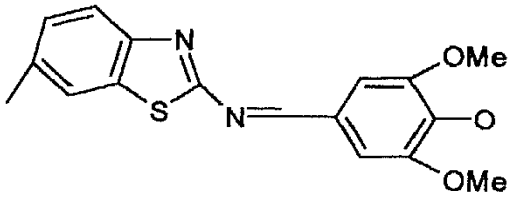
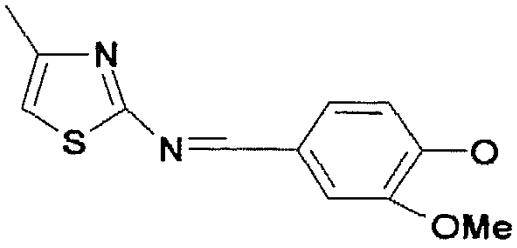
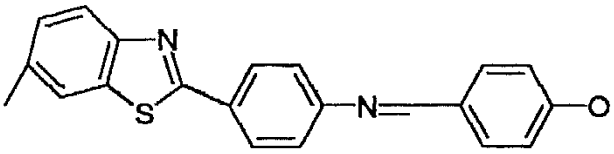
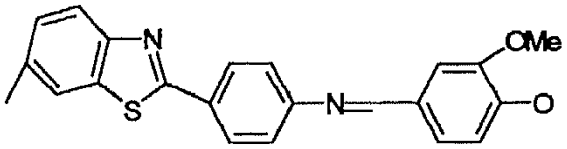
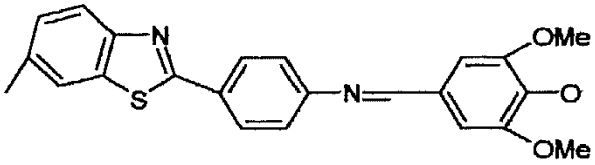
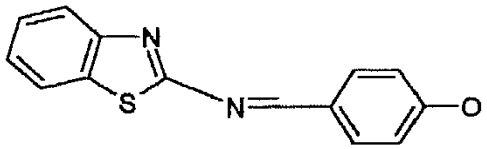
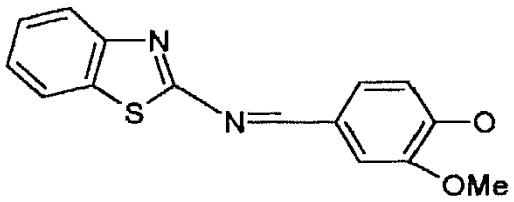
No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
60		Nucleotide metabolism regulator	73	1
		Glutamate release inhibitor	55	1
		Antiulcerative	53	2
61		Nucleotide metabolism regulator	71	2
		Glutamate release inhibitor	55	1
		Antiulcerative	51	3
62		Nucleotide metabolism regulator	70	2
		Cognition disorders treatment	62	2
		Antiulcerative	57	2
		Anti-Helicobacter pylori	54	1
63		Nucleotide metabolism regulator	77	1
		Antiseptic	63	1
		Antituberculosic	60	1
		Antimycobacterial	58	1
		Antiinflammatory	53	4
64		Nucleotide metabolism regulator	76	1
		Antituberculosic	57	1
		Antimycobacterial	55	1
		Antiinflammatory	52	5
65		Nucleotide metabolism regulator	75	1
		Antituberculosic	56	1
		Antimycobacterial	54	1
		Antiinflammatory	52	5
66		Nucleotide metabolism regulator	75	1
		Cognition disorders treatment	64	1
		Glutamate release inhibitor	61	1
		ATPase inhibitor	59	2
		Antiulcerative	56	2
		Antimycobacterial	54	1
		Antitoxic	53	4
		Antituberculosic	53	1
		Anti-Helicobacter pylori	53	1
		Dopamine D3 agonist	50	1
67		Nucleotide metabolism regulator	73	1
		Glutamate release inhibitor	58	1
		Cognition disorders treatment	57	2
		Antiulcerative	54	2
		Antimycobacterial	51	1
		Antituberculosic	51	1

TABLE I – *continued*

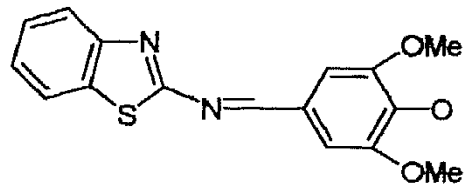
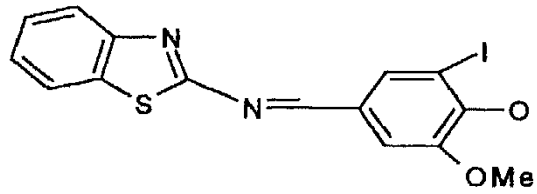
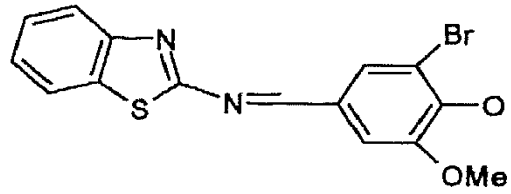
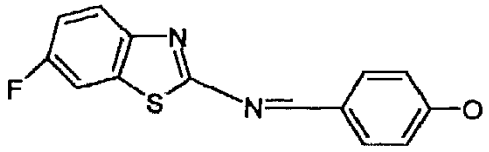
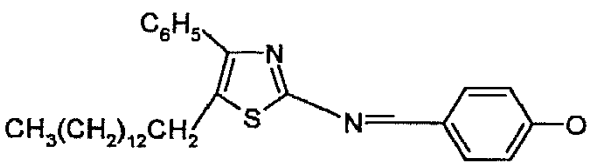
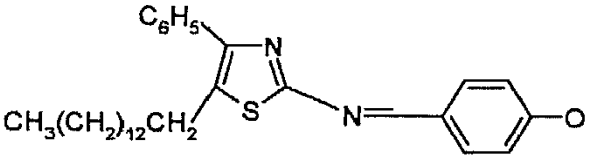
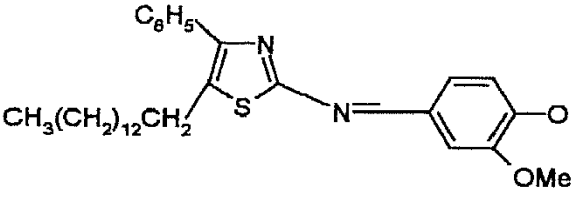
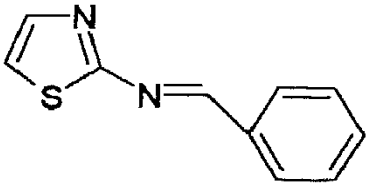
No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
68		Nucleotide metabolism regulator	72	1
		Glutamate release inhibitor	58	1
		Cognition disorders treatment	57	2
		ATPase inhibitor	53	3
		Antiulcerative	52	2
		Antimycobacterial	50	1
69		Nucleotide metabolism regulator	69	2
		ATPase inhibitor	58	2
		Sodium channel blocker	57	2
		Histamine release stimulant	55	3
		Glutamate release inhibitor	53	1
		Free radical scavenger	52	2
70		Nucleotide metabolism regulator	66	3
		Antituberculosic	56	1
		Renal disease treatment	55	1
		Antimycobacterial	54	1
		Antiseptic	55	2
		Glutamate release inhibitor	53	1
		Cognition disorders treatment	53	3
71		Nucleotide metabolism regulator	70	2
		HDL-cholesterol increasing	62	3
		Glutamate release inhibitor	58	1
		Cognition disorders treatment	57	2
		Antiulcerative	55	2
		ATPase inhibitor	52	4
72		Nucleotide metabolism regulator	71	2
		Psychotropic	63	2
		Glutamate release inhibitor	60	1
		Cognition disorders treatment	57	2
		Antiulcerative	51	2
		ATPase inhibitor	52	4
73		Renal disease treatment	65	1
		Cognition disorders treatment	56	2
		Nucleotide metabolism regulator	57	5
74		Renal disease treatment	61	1
		Nucleotide metabolism regulator	55	6
75		Cognition disorders treatment	77	1
		Nucleotide metabolism regulator	73	1
		Antiulcerative	59	1
		TNF convertase inhibitor	58	1
		Fibrinolytic	63	6
		Histamine antagonist	57	1
		Anti-Helicobacter pylori	56	1
		Antituberculosic	56	1
		Antimycobacterial	54	1
		Antidepressant	54	3
		Prostaglandin antagonist	54	3
		Hematopoietic inhibitor	51	1

TABLE I – *continued*

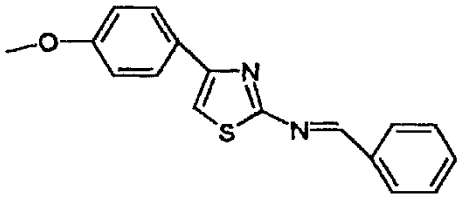
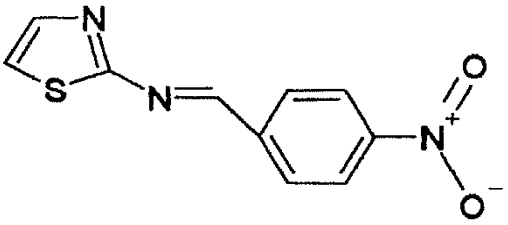
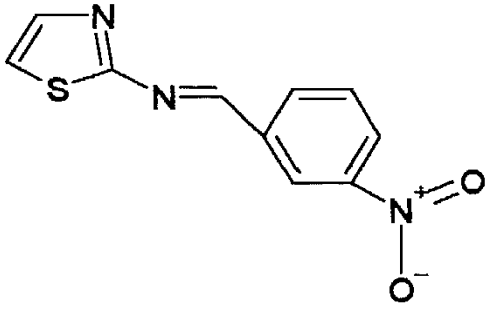
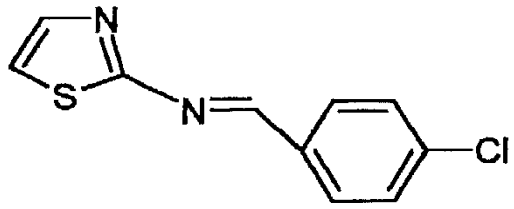
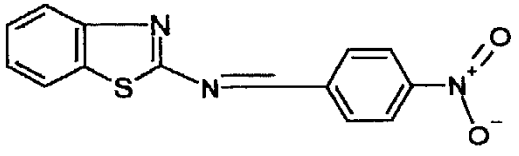
No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
		Antibacterial	50	1
76		Fibrinogen receptor antagonist	82	4
		Nucleotide metabolism regulator	67	3
		Cognition disorders treatment	66	1
		Antiulcerative	63	1
		Anti-Helicobacter pylori	62	1
		ATPase inhibitor	56	2
		Histamine antagonist	55	1
		Antituberculosic	51	1
		Gastric antisecretory	50	1
77		Cognition disorders treatment	68	1
		Nucleotide metabolism regulator	62	4
		Antituberculosic	56	1
		Fibrinolytic	62	7
		Antimycobacterial	55	1
		TNF convertase inhibitor	53	1
		Antiulcerative	51	3
78		Cognition disorders treatment	66	1
		Fibrinolytic	62	6
		Nucleotide metabolism regulator	59	5
		Antituberculosic	55	1
		Antimycobacterial	53	1
		TNF convertase inhibitor	52	1
		Antiulcerative	50	3
79		Cognition disorders treatment	71	1
		Nucleotide metabolism regulator	71	2
		Prostaglandin antagonist	60	1
		HDL-cholesterol increasing	59	4
		Antidepressant	57	3
		Antiulcerative	54	2
		TNF convertase inhibitor	53	1
		Anti-Helicobacter pylori	53	1
		Fibrinolytic	60	9
		Histamine antagonist	53	1
		GABA A receptor agonist	53	2
		Antituberculosic	51	1
		Antimycobacterial	51	1
80		Nucleotide metabolism regulator	70	2
		Cognition disorders treatment	67	1
		ATPase inhibitor	60	1
		Antimycobacterial	59	1
		Antituberculosic	59	1
		Glutamate release inhibitor	56	1
		Antiulcerative	57	2
		Anti-Helicobacter pylori	54	1
		Antischistosomal	53	2
		Histamine antagonist	51	1
		Antidepressant	52	3

TABLE I – *continued*

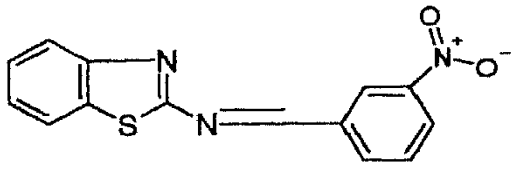
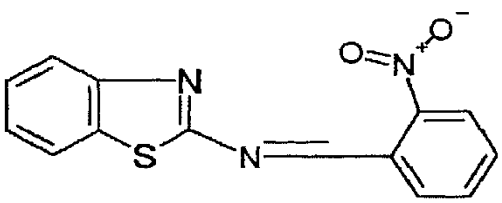
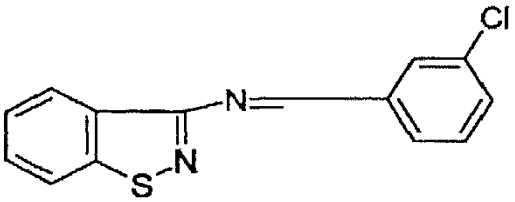
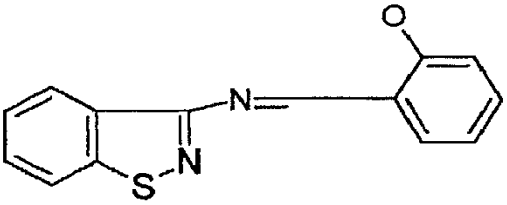
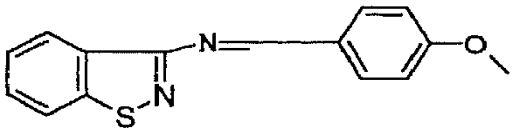
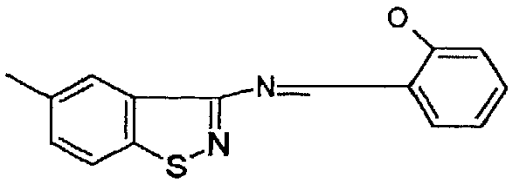
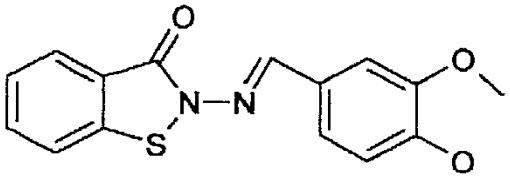
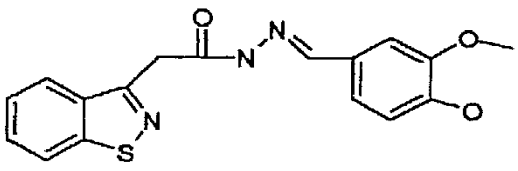
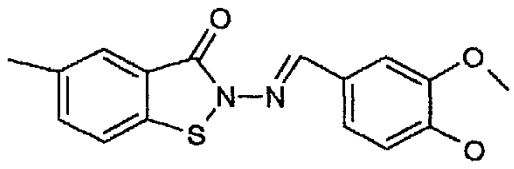
No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
81		Nucleotide metabolism regulator	69	2
		Cognition disorders treatment	65	1
		ATPase inhibitor	60	1
		Antimycobacterial	58	1
		Antituberculosic	58	1
		Antiulcerative	56	2
		Glutamate release inhibitor	55	1
		Anti-Helicobacter pylori	53	1
		Antischistosomal	50	2
82		Cognition disorders treatment	64	1
		Nucleotide metabolism regulator	64	3
		ATPase inhibitor	58	2
		Antidepressant	57	3
		Antiulcerative	56	2
		Antischistosomal	55	1
		Histamine antagonist	51	1
83		Nucleotide metabolism regulator	76	1
		2,3 Oxidosqualene lanosterol cyclase inhibitor	60	2
		Histamine antagonist	57	1
		HDL-cholesterol increasing	56	6
		Antihistaminic	50	2
		Antifungal	50	3
		Alzheimer's disease treatment	51	5
84		Nucleotide metabolism regulator	75	1
		Antifungal	51	2
		Histamine antagonist	50	2
		2,3 Oxidosqualene lanosterol cyclase inhibitor	50	3
85		Nucleotide metabolism regulator	77	1
		Histamine antagonist	51	1
		Alzheimer's disease treatment	52	4
86		Nucleotide metabolism regulator	73	1
		Antifungal	50	3
87		Antiulcerative	67	1
		Cardiovascular analeptic	63	6
		Gastric antisecretory	54	1
		Psychosexual dysfunction treatment	59	6
		Fibrinolytic	59	11
		Antacid	53	8
88		MAO inhibitor	63	1
		Antituberculosic	58	1
		Hypotermic	57	3
		Antimycobacterial	53	1
		Nucleotide metabolism regulator	55	6

TABLE I – *continued*

No	*Structure*	*Predicted Activity Spectra*	P_a,%*	P_i,%*
89		Antiulcerative	67	1
		Gastric antisecretory	53	1
		Cardiovascular analeptic	60	8
		Psychosexual dysfunction treatment	56	7
		Antacid	54	7
		Fibrinolytic	58	12

The firm's research interest and strategy [1] and the used standard biological tests concerning the drug safety assessment [2], restrict the investigations of any lead molecule and new drug candidates.

If we were able to suggest or delineate possible or uncovered existing activities for every new or known compound new priorities and directions in testing could be established.

Basic information about different properties of chemical compounds are contained in its structural formula. At the early stage of drug's R&D, structural formulae of potentially active compounds are available. Therefore, the only real possibility for complex investigation of compounds' biological activity is to develop a computer program, which will be able to predict many kinds of biological activity on the basis of (only) structural formula of chemical compound.

Computer aided structure–activity relationship (SAR) analysis is widely used now for new finding leads and optimisation. Most of the SAR/QSAR/Modelling methods are applicable to one or several types of biological activity within the same chemical series [3–5]. Thus, they are unable to elucidate the general biological "potential" of a molecule under study. Recently, the possibility for simultaneous prediction of many different biological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity within different series of compounds have been demonstrated [6–8] by the computer system PASS (Prediction of Activity Spectra for Substances).

The purpose of the present work is to evaluate the PASS predictive abilities on the set of 89 new thiazole derivatives with known NSAID, local anaesthetic and antioxidant activity, and to estimate which psychotropic effects can be discovered in some compounds of this series.

MATERIALS AND METHODS

Data Set

Thiazolyl group is of great importance in biological systems. Anti-inflammatory, analgesic, antipyretic activities for some thiazolyl and benzoisothiazolyl derivatives are also known [9,10]. The screening tests for these compounds have been chosen on the basis of certain researchers' purposes. It is obvious that known pharmacological actions of tested thiazolyl derivatives do not represent the comprehensive biological activity spectra of the compounds. Therefore, there is not any thiazolyl derivative that has been tested in the battery of all available tests. However, the existing experimental data [11–19] provide the basis for evaluation of PASS prediction and, if the results of evaluation will be satisfactory, new activities can be further discovered for these compounds based on computer prediction.

The structures of the synthesized compounds with numbers 1–50 is presented in our publication [20]; the structures with numbers 51–89 are given in Table I.

PASS Description

The basic elements of PASS include: presentation of biological activity, description of chemical structure, structure–activity relationships knowledge base–SAR base, algorithm of the activity spectra estimation. Current version of PASS has many essential differences from earlier ones [21,6,7]. Here we present just brief description of the program, necessary for general understanding what is PASS and how its results can be interpreted. More detailed description of the approach is available on the web-site [8] and from publications [22–25].

Activity Presentation

A biological activity spectrum should present every activity of a compound despite of the difference in the essential conditions of its experimental determination. On the other hand, a sufficiently large set of the substances can be collected only using many different sources because the information taken from one publication never covers all aspects of a biological action of the described substance. For example, according to Ref. [27] caffeine (CAS No 58-08-2) is "stimulant," "analeptic" and "diuretic". As a result of informational search, it was found that caffeine is: psychotropic, psychostimulant, analeptic, respiratory analeptic, cardiotonic, diuretic, saluretic, immunomodulator, immunosuppressant, spasmolytic, spasmogenic, vasopressor, vasodilator, hypertensive, nucleotide metabolism regulator, cAMP phosphodiesterase inhibitor, teratogenic and embriotoxic agent. Some apparent contradictions in terms may be attributed to the reversing of effects for various doses of caffeine.

Thus, the *biological activity spectrum* may be defined as the "intrinsic" compound property depending only on its structure and physicochemical characteristics. In PASS, biological activities are described qualitatively and compounds could be subdivided into two subsets, "active" and "inactive" ones.

Chemical Structure Description

The 2D structural formulas of a compound are chosen as the basis for the structure description because they correspond our purpose to estimate a biological activity spectrum of any substance even if it just virtually designed, neither synthesized nor experimentally studied. There are many chemical compounds characteristics used as descriptors in SAR analysis: sub-structural fragments, geometrical and topological indexes, physicochemical characteristics, etc. For different kinds of biological activity in various chemical series, particular descriptors occurred more or less significant in the appropriate SAR/QSAR equations. Thus, in the simultaneous prediction of the wide range of the activity kinds for non-congeneric compounds it is necessary to use the description, which would be exactly sufficient to achieve a consistent prediction result but would not be too sensitive to ignore some random regularities. In earlier versions of PASS, the Substructure Superposition Fragment Notation (SSFN) [28] was used. In our paper published recently [22], we described the substructure descriptors called "Multilevel Neighborhoods of Atoms" (MNA). MNA descriptors are better SSFN from the many points of view, and in PASS MNA descriptors are used now.

MNA descriptors are based on the structure representation, that does not specify the bond types and includes hydrogen according to valence and partial charge of atoms. MNA descriptors are generated as recursively defined sequence:

zero-level MNA descriptor for each atom is the mark A of the atom itself;

any next-level MNA descriptor for each atom is the substructure notation $A(D_1D_2...D_i...)$, where D_i is the previous-level MNA descriptor for the i-th immediate neighbours of the atom.

This iterative process may be continued enclosing 2nd, 3rd, etc. neighbourhoods of the each atom. It is important to emphasize that the atom mark may include not only the atom type but also any additional information about atom, for example, about belonging to cycle or chain. A structure of molecule is represented in PASS as a set of the 1st- and 2nd-levels MNA descriptors. In 2nd-level MNA descriptors, we use the indicator "–" of belonging to a chain.

More detailed presentation of MNA descriptors is available from Ref. [22].

Structure Equivalence

The *structure equivalence* is an important feature of the PASS approach. The structures are considered as equivalent if they have the same molecular formula and MNA descriptors' set. Only unique structures are included into SAR base. Since MNA descriptors do not represent the stereochemical peculiarities of a molecule, the compounds, which have only stereochemical differences in the structure, are formally considered as equivalent. Such simplification is obligatory "payment" for the possibility to predict many different activities on the basis of uniform approach.

SAR Base

The prediction is carried out using the structure–activity relationships knowledge-base (SAR base), which is based on analysis of the training set containing biologically active compounds. To include training set into the SAR base the MNA descriptors are generated for each compound in this set. If a structure is not fully determined, i.e. includes undetermined atoms or residua, the compound is not included into the SAR base. If in the SAR base the equivalent structure is found, then new activity spectrum is joined with the existing one. The prediction is based on a SAR analysis of the training set containing more than 43,000 compounds, which have more than 700 kinds of biological action.

Algorithm of the Activity Spectra Estimation

The prediction algorithm was chosen from more than hundred variants tested through several years. Its detailed is published elsewhere [23–26].

The main target of PASS is to predict the activity spectra for a new substance. Therefore, the general principle of PASS algorithm is the excluding the substance from SAR base, if its structure is equivalent to the structure of compound under estimation.

The estimation of activity spectrum is the ranked list of probabilities "to be active" P_a, "to be inactive" P_i, and the type of activity. The ranking is arranged on descending order of $P_a - P_i$; thus, the most probable activities are at the top of predicted spectrum. The spectrum can be shortened at any desirable cut-off value, but $P_a > P_i$ is used by default for selection of likely active compounds.

The probabilities P_a and P_i are considered as the measures of belonging to subsets of "active" and "inactive" compounds, and the probabilities of the 1st and the 2nd kinds of the prediction error, respectively.

TABLE II Comparison of predicted and experimental data for some synthesized and tested compounds

No	*Activity*	*Experimental data, %*	P_a,%	P_i,%	*Coincidence (Exp./Pred.)*
3	LA*	84.4 ± 10.0	33	5	−/+
4	NSAID†	123.3 ± 1.7	–	–	+/−
	LA	103.4 ± 9.0	34	5	+/+
7	NSAID	100.0 ± 2.1	–	–	+/−
8	NSAID	62.5 ± 1.0	–	–	−/−
9	NSAID	50.0 ± 2.5	–	–	−/−
10	NSAID	106.3 ± 6.9	–	–	+/−
11	NSAID	83.3 ± 2.1	–	–	−/−
12	NSAID	116.7 ± 1.9	–	–	+/−
13	NSAID	89.6 ± 2.3	–	–	−/−
	LA	53.9 ± 4.7	43	2	−/+
14	NSAID	91.7 ± 7.7	21	13	+/+
15	NSAID	68.8 ± 1.5	24	10	−/+
16	NSAID	97.5 ± 4.2	30	4	+/−
17	NSAID	89.6 ± 3.8	–	–	−/−
	LA	98.8 ± 8.3	37	4	+/+
18	LA	125.5 ± 5.2	34	4	+/+
19	NSAID	97.9 ± 3.8	22	11	+/+
20	NSAID	111.1 ± 3.8	19	17	+/+
21	NSAID	93.5 ± 6.5	30	4	+/+
22	NSAID	112.5 ± 2.1	30	5	+/+
23	NSAID	74.0 ± 1.9	25	9	−/+
24	NSAID	126.0 ± 4.6	25	10	+/+
25	NSAID	103.5 ± 5.4	25	8	+/+
26	NSAID	70.2 ± 3.8	–	–	−/−
27	NSAID	66.9 ± 4.0	–	–	−/−
28	NSAID	114.4 ± 6.9	23	11	+/+
29	NSAID	97.9 ± 3.4	22	12	+/+
30	NSAID	102.5 ± 3.1	–	–	+/+
	LA	53.9 ± 12.5	37	4	−/+
31	NSAID	84.2 ± 5.2	–	–	−/−
32	NSAID	167.3 ± 2.7	–	–	+/−
34	NSAID	114.2 ± 2.5	19	17	+/+
35	NSAID	125.0 ± 3.3	20	15	+/+
39	NSAID	82.5 ± 3.3	–	–	−/−
40	NSAID	167.3 ± 2.7	–	–	+/−
41	NSAID	108.5 ± 4.0	25	8	+/+
42	NSAID	135.2 ± 2.7	26	8	+/+
43	NSAID	72.9 ± 5.0	24	10	−/+
44	NSAID	88.3 ± 2.1	24	10	−/+
45	NSAID	122.1 ± 2.3	24	9	+/+
46	NSAID	123.3 ± 11.5	25	9	+/+
47	NSAID	74.2 ± 5.2	–	–	−/−
49	NSAID	149.6 ± 6.9	–	–	+/−
50	NSAID	81.3 ± 4.6	–	–	−/−
	AO‡	24.5 ± 2.7	16	15	−/+
51	NSAID	59.2 ± 5.4	25	8	−/+
52	NSAID	56.3 ± 2.5	24	9	−/+
53	NSAID	56.5 ± 2.8	–	–	−/−
54	NSAID	71.8 ± 3.3	–	–	−/−
55	NSAID	42.1 ± 0.8	–	–	−/−
56	NSAID	39.5 ± 1.3	–	–	−/−
57	NSAID	48.9 ± 1.5	–	–	−/−
58	NSAID	39 ± 2.2	–	–	−/−
59	NSAID	47.3 ± 2.0	–	–	−/−
60	NSAID	23.9 ± 1.6	–	–	−/−
61	NSAID	24.3 ± 1.8	–	–	−/−
62	NSAID	35.7 ± 4.5	–	–	−/−
63	NSAID	47.3	27	7	−/+
64	NSAID	23.9	26	7	−/+
65	NSAID	24.3	26	8	−/+
66	NSAID	36 ± 1.4	–	–	−/−

TABLE II – *continued*

No	*Activity*	*Experimental data, %*	P_a,%	P_i,%	*Coincidence (Exp./Pred.)*
67	NSAID	54 ± 0.7	–	–	–/–
68	NSAID	49.3 ± 1.9	–	–	–/–
69	NSAID	75.04 ± 0.6	–	–	–/–
70	NSAID	63.2 ± 0.4	–	–	–/–
71	NSAID	44.8 ± 0.5	–	–	–/–
72	NSAID	62 ± 1.7	–	–	–/–
73	NSAID	64 ± 0.8	–	–	–/–
74	NSAID	58 ± 0.8	–	–	–/–
75	NSAID	32.2	27	7	–/+
76	NSAID	42.3	20	5	–/+
77	NSAID	42	23	10	–/+
78	NSAID	64	24	10	–/+
79	NSAID	50.4	31	4	–/+
80	NSAID	39	–	–	–/–
81	NSAID	47	–	–	–/–
82	NSAID	39	–	–	–/–
83	NSAID	59.2	–	–	–/–
	AO	17.1/92	–	–	–/–
84	NSAID	37.6	–	–	–/–
	AO	5.8/97.1	–	–	–/–
85	NSAID	52	–	–	–/–
	AO	8.9/92.4	–	–	–/–
86	NSAID	46	–	–	–/–
	AO	24.8/88			–/–
87	NSAID	24.6	–	–	–/–
88	NSAID	–	–	–	–/–
89	NSAID	16.3	–	–	–/–

* Relative activity to Procaine (100.0 ± 3.1%) in percentage.
† Relative activity to Indomethacine (100.0 ± 4.8%) in percentage.
‡ Relative activity to Acetylcalicylic acid (100.0 + 2.6%) in percentage.

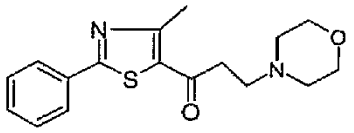
Interpretation of Prediction Results

Depending on the value of calculated probability to be active P_a and inactive P_i, the results of prediction are more or less reliable. However, the reliability is not the only criteria for choosing of prospective compounds. If $P_a > 0.7$, the chance to find the activity in experiment is high, but in many cases the compound may occur to be the close analogue of known pharmaceutical agents. If $0.5 < P_a < 0.7$, the chance to find the activity in experiment is less, but the compound is not so similar to known pharmaceutical agents. If $P_a < 0.5$, the chance to find the activity in experiment is even more less, but if the predicted activity will be confirmed, the compound might occur to be a New Chemical Entity. Thus, one may choose which activities have to be tested in compounds under study, based on compromise between the novelty of pharmacological action and the risk to obtain the negative result in experimental testing.

RESULTS AND DISCUSSION

Examples of predicted biological activity spectra for compounds 51–89 are presented in Table I. Due to the limited space of publication, only biological activities with $P_a > 50\%$ are shown here, but all activities for which $P_a > P_i$ can be considered as probable.

To evaluate the applicability of the approach to the thiazole derivatives, the results of prediction are compared to experimental data for some compounds from the set, studied as



> <ACTIVITY_PREDICTION>

39 substructure descriptors; 1 new.

Pa	Pi	for Activity:
0.619	0.069	Cardiovascular analeptic
0.535	0.050	Antiischemic, cerebral
0.515	**0.047**	**Anticonvulsant**
0.501	0.037	Sodium channel blocker
0.445	0.019	Anesthetic general
0.457	0.040	Mediator release inhibitor
0.536	0.126	Arrhythmogenic
0.453	0.064	Sedative
0.423	0.035	Antiarrhythmic
0.433	0.050	Cardiotonic
0.415	0.033	Bone formation stimulant
0.437	0.060	Spasmolytic, urinary
0.397	0.037	GABA A receptor agonist
0.348	0.013	Xanthine oxidase inhibitor
0.475	0.155	Myocardial ischemia treatment
0.401	0.082	Psychotropic
0.386	0.081	Spasmolytic
0.366	**0.067**	**Anxiolytic**
0.358	0.061	Cyclic GMP phosphodiesterase inhibitor
0.375	0.083	Antisecretoric
0.385	0.102	Analgesic, non-opioid

FIGURE 1 Example of biological activity spectra prediction for 3-morpholin-1-(4-methyl-2-phenyl-5-thiazolyl)-1-propanon.

TABLE III Compounds predicted as anxiolytics, anticonvulsants and cognition enhancers

Activity	*Number of compounds*	*Registry number*
Anxiolytic	7	4, 5, 9, 11, 38, 39, 40
Anticonvulsant	44	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 17, 18, 19, 20, 26, 27, 28, 29, 30, 31, 32, 34, 35, 37, 38, 40, 54, 57, 58, 59, 63, 65, 66, 68, 70, 71, 72, 80, 81, 82
Cognition disorders treatment	55	3, 4, 11, 12, 18, 20, 23, 24, 25, 26, 27, 30, 31, 32, 34, 35, 38, 39, 40, 43, 44, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82

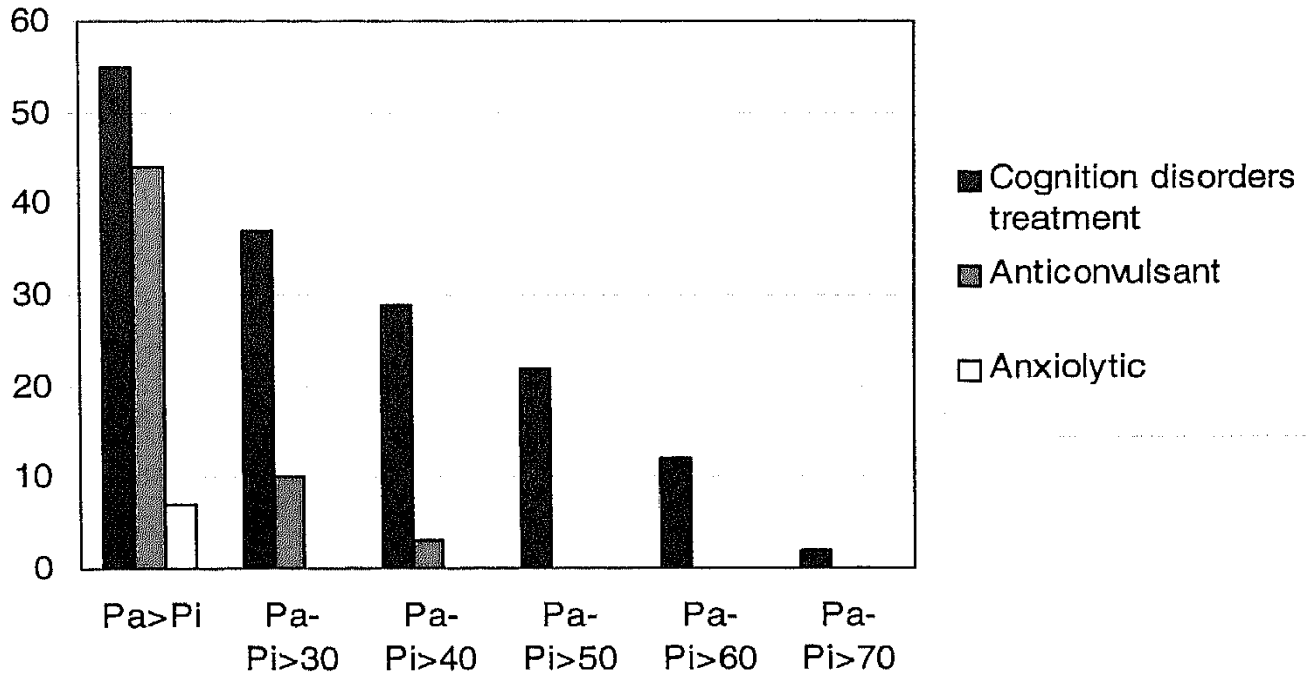


FIGURE 2 Distribution of the predicted psychotropic activities versus $P_a - P_i$ values.

non-steroid antiinflammatory (78 compounds); local anaesthetics (6 compounds); antioxidants (5 compounds) tests. These data are given in Table II.

In the experiment the compound is considered as being active if its potency is no less than the potency of the appropriate reference drug (the experimental errors are taken into account). It is obvious that such criteria may be too strong for evaluation of the results of biological activity prediction, but it is often used in search for new lead compounds by pharmaceutical companies. Predicted activity is considered as significant if its estimated $P_a > P_i$. Coincidence of the experiment and prediction is designated, respectively, by " + /+" (active/active) or " − /−" (inactive/inactive), contradiction between the experiment and prediction is designated by " − /+" (inactive/active) or " + /−" (active/inactive). Experiment coincides with the prediction for 59 out of 78 compounds tested as NSAIDs (75.6%); for 3 out of 6 compounds tested as local anaesthetics (50%); for 5 compounds tested as antioxidant (100%). General concordance that describes the overall percentage of correct predictions (" + /+" and " − /−") equals to $52/89 \approx 70.8\%$. It is less than 85% average accuracy obtained in LOO cross-validation [8], but taking into account that the probability of random guessing any activity for a compound is $1/773 \sim 0.1\%$, the average accuracy of prediction seems to be sufficient. The values of the first kind (+/−) and second kind (−/+) errors are 9.0 and 20.2%, respectively).

Therefore, PASS could be used for prioritising of further pharmacological study of the compounds from this set and screening of new thiazole derivatives.

In particular, we can consider the prediction of such kinds of psychotropic action as anxiolytic, anticonvulsant and cognition enhancer. The example of such prediction for compound No. 5 is given in Fig. 1. It is obvious that this compound might be tested as anticonvulsant ($P_a = 0.515$) and as anxiolytic ($P_a = 0.366$).

The registry numbers of all compounds from the set predicted as anxiolytics, anticonvulsants and cognition enhancers are given in Table III. The distribution of these compounds versus $P_a - P_i$ values is presented in Fig. 2. Most of estimated P_a values are less than 60%, therefore, keeping in mind usual interpretation of PASS predictions (see above), one might suggest that the compounds under study are not very close to well-known

anxiolytics, anticonvulsants and cognition enhancers. Moreover, if their activity will be confirmed by the experiment, some of them may occur to be New Chemical Entities (NCE).

CONCLUSIONS

1. By evaluation of PASS versus 89 recently synthesized thiazole derivatives it is shown that the concordance between the prediction and experiment is 70.8% even if very strong criteria for determination of activity in the experiment is applied.
2. According to the computer prediction, new anxiolytic, anticonvulsant and cognition enhancer agents can be found among the studied thiazole derivatives. Since the estimated probability values for these activities P_a are less than 60%, NCE can be discovered among the compounds from the studied series.

Acknowledgement

We gratefully acknowledge the partial support of this work by the International Association for the promotion of cooperation with scientists from the New Independent States of the Former Soviet Union (grant INTAS-2000-0711).

References

[1] http://www.phrma.org/publications/]">[http://www.phrma.org/publications/]
[2] http://www.fda.gov/]">[http://www.fda.gov/]
[3] Wermuth, C. (1996) *The Practice of Medicinal Chemistry* (Academic Press, New York).
[4] Gundertofte, K. and Jorgensen, F.S. (2000) *Molecular Modeling and Prediction of Bioactivity* (Kluwer Academic Press, Dordrecht).
[5] Holtje, H.-D. and Sippl, W. (2001) *Rational Approaches to Drug Design* (Prous Science, Barcelona).
[6] Filimonov, D.A., Poroikov, V.V., Karaicheva, E.I., *et al.*, (1995) "Computer-aided prediction of biological activity spectra of chemical substances on the basis of their structural formulae: computerized system PASS", *Exp. Clin. Pharmacol. (Rus)* **58**, 56–62.
[7] Filimonov, D.A. and Poroikov, V.V. (1996) "PASS: computerized prediction of biological activity spectra for chemical substances", In: *Bioactive Compound Design: Possibilities for Industrial Use* (BIOS Scientific Publishers, Oxford), pp 47–56.
[8] http://www.ibmh.msk.su/pass]">[http://www.ibmh.msk.su/PASS]
[9] Klose, N., Niedbolla, K., Schwartz, K. and Bottcher, I. (1983) "4,5-Bis (4-methoxyphenyl)-2-arylthio-azole mit antiphlogischer wirkung", *Arch. Pharm.* **316**, 941–951.
[10] Satsangi, R.K., Zaidi, S.M. and Misra, V.S. (1983) "1-(4-Substituted-thiazol-2-yl) hydantoins as antinflammatory and CNS-active agents", *Pharmazie* **38**, 341–342.
[11] Vicini, P., Amoretti, L., Chiavarini, M. and Impicciatore, M. (1990) "Synthesis and local anaesthetic activity of alkyl aminoacyl derivatives of 3-amino-1,2-benzoisothiazoles", *Il Farmaco.* **45**, 933–944.
[12] Vicini, P., Fiscaro, E. and Lugari, M-T. (2000) "Synthesis and correlation between experimental and calculated lipophilic Indice of new 1,2-benzisothiazole derivatives with potential antimicrobial activity", *Arch. Pharm. Pharm. Med. Chem.* **333**, 135–144.
[13] Geronikaki, A., Sotiropoulou, E. and Kourounakis, P.N. (1992) "Synthesis and mass spectra of some thiazole aminoketones with prospective local anaesthetic activity", *Pharmazie* **47**, 298–300.
[14] Mgonzo, R., Geronikaki, A. and Kourounakis, P. (1995) "Synthesis and antiinflammatory activity of some new thiazole derivatives", *Pharmazie* **50**, 505–506.
[15] Geronikaki, A. and Hadjipavlou-Litina, D. (1996). Novel thiazole and benzothiazole derivatives designed as possible lipoxygenase inhibitors. In: *Abstracts of XIV International Symposium on Medicinal Chemisty*, Maastricht (The Netherlands), P-5.21.
[16] Geronikaki, A and Hadjipavlou-Litina, D. (1996) "*In vitro* studies of some 2,4-disubstituted thiazolyl-aminoketones with antiinflammatory activity. Correlation with lipophilicity and structural characteristics", *Arz. Forsh./Drug Res.* **46**, 1134–1137.
[17] Mgonzo, R., Geronikaki, A. and Theophilidis, G. (1996) "Synthesis of new thiazole derivatives and comparative study of their local anaesthetic activity using the evoked action potential in the sciatic nerve", *Res. Commun. Pharmacol. Toxicol.* **1**, 137–148.

[18] Hadjipavlou-Litina, D., Geronikaki, A. and Sotiropoulou, E. (1993) "Antiinflammatory activity of aminoketone derivatives of 2,4-disubstituted thiazoles", *Res. Commun. Chem. Pathol. Pharmacol.* **79**, 355–362.
[19] Hadjipavlou-Litina, D.J. and Geronikaki, A.A. (1997) "Effect of 2,4-disubstituted thiazol-5-yl aminoketones on carrageenin induced inflammation. Biological evaluation and a structure–activity approach", *Res. Commun. Mol. Pathol. Pharmacol.* **96**, 307–318.
[20] Geronikaki, A., Poroikov, V., Hadjipavlou-Litina, D., Filimonov, D., Lagunin, A. and Mgonzo, R. (1999) "Computer aided predicting of biological activity spectra and experimental testing of new thiazole derivatives", *Quant. Struct.-Act. Relat.* **18**, 16–25.
[21] Poroikov, V.V., Filimonov, D.A., Stepanchikova, A.V., *et al.*, (1996) "Opimization of synthesis and pharmacological testing of new compounds based on computerized prediction of their biological activity spectra", *Chim.-Pharm. J. (Rus)* **30**, 20–23.
[22] Filimonov, D.A., Poroikov, V.V., Borodina, Y. and Gloriozova, T. (1999) "Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptions", *J. Chem. Inf. Comput. Sci.* **39**, 666–670.
[23] Poroikov, V.V., Filimonov, D.A., Borodina, Yu.V., Lagunin, A.A. and Kos, A. (2000) "Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds", *J. Chem. Inf. Comput. Sci.* **40**, 1349–1355.
[24] Poroikov, V. and Filimonov, D. (2001) "Computer-aided prediction of biological activity spectra. Application for finding and optimization of new leads", In: Holtje, H.-D. and Sippl, W., eds, *Rational Approaches to Drug Design* (Prous Science, Barcelona), pp 403–407.
[25] Anzali, S., Barnickel, G., Cezanne, B., Krug, M., Filimonov, D. and Poroikov, V. (2001) "Discriminating between drugs and nondrugs by prediction of activity spectra for substances (PASS)", *J. Med. Chem.* **44**, In press (Web Release Date: June 14, 2001).
[26] Poroikov, V., Akimov, D., Shabelnikova, E. and Filimonov, D. (2001) "Top 200 medicines: can new actions be discovered through computer-aided prediction?", *SAR QSAR Environ. Res.* **12**, 327–344.
[27] Negwer, M. (1987) *Organic Chemical Drugs and Their Synonyms* (Akademie, Berlin) **Vol. 1**, Vol. 2, pp. 817–1654; Vol. 3, pp. 1655–2470, pp 1–816.
[28] Avidon, V.V., Pomerantsev, I.A., Rozenblit, A.B. and Golender, V.E. (1982) "Structure–activity relationship oriented language for chemical structure representation", *J. Chem. Inf. Comput. Sci.* **22**, 207–214.