Multi-Targeted Natural Products Evaluation Based on Biological Activity Prediction with PASS

Alexey Lagunin, Dmitry Filimonov and Vladimir Poroikov*

Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., 10, Pogodinskaya Str., Moscow, 119121, Russia

Abstract: Natural products found a wide use in folk medicine. Presently, when routine development of new drugs faced a considerable challenge, they become an inspiration and valuable source in drugs discovery. Rather complex and diverse chemical structures of natural compounds provide a basis for modulation of different biological targets. Natural compounds exhibit a multitargeted action that may lead to additive/synergistic or antagonistic effects. Rational design of more safe and potent pharmaceuticals requires an estimation of probable multiple actions of natural products. Our software PASS can perform such estimation. It predicts with reasonable accuracy over 3500 pharmacotherapeutic effects, mechanisms of action, interaction with the metabolic system, and specific toxicity for drug-like molecules on the basis of their structural formulae. We analyzed PASS predictions utilizing PharmaExpert, which performs selection of compounds with multiple mechanisms of action, analysis of activity-activity relationships and drug-drug interactions. The paper describes an application of PASS and PharmaExpert to the evaluation of biological activity of natural compounds including marine sponge alkaloids, triterpenoids of lupane group, and their derivatives. Proposed computer-aided methods can generate combinatorial libraries of macrolides. They help to select the most promising pharmaceutical leads with the required properties. Case study, based on the analysis of biological activity spectra predicted for St John's Wort constituents, clearly demonstrates capabilities of computational methods in the evaluation of multitargeted actions, additive/synergistic and/or antagonistic effects of natural products.

Keywords: Natural products, computational evaluation, biological activity spectra prediction, PASS, multitargeted action, drug-drug interaction, marine sponge alkaloids, triterpenoids, St John's Wort.

1. INTRODUCTION

Natural products are widely used in a non-traditional medicine [1], especially in China [2], India [3] and Russia [4]. Being created by Mother Nature, natural compounds are specially adapted for their interactions with biological systems [5]; therefore, they are considered as valuable sources for drug discovery. Over 70% of New Chemical Entities (NCEs) introduced into medical practice in 1981-2006 were obtained on the basis of natural products [6]. Rather complex and diverse chemical structures of natural compounds provide the basis for modulation of different biological targets [7]. Multitargeted actions of natural compounds could lead to additive/synergistic or antagonistic effects [8]. Since there are several thousands of known pharmacological targets and natural products exhibit pleiotropic action interacting with multiple targets, therefore computer-aided methods could be extremely useful for natural products evaluation [9].

Generally, natural products research requires the utilization of virtual screening methods to find new lead substances. Currently, three main integrated approaches are in use for achieving this purpose, including: (1) creation of 3D structure database of natural product components with description of biological activity obtained from *in vitro* screening of natural products extracted from ethnopharmacological sources; (2) selection of biologically active material on the basis of hits found by docking in the database of natural product 3D structures; (3) parallel screening of unstudied natural substances, to identify the promising lead compounds [10]. But all these strategies confined to identification of phytochemical lead ethnic biological activity only. It is believed that just a small part of structural diversity exhibited by plant compounds has been seriously explored for its pharmacological potential so far [11]; and, therefore, new *in-silico* approaches are necessary to reveal novel biological activities of known natural products, including their interactions with the known biological targets and related pharmacotherapeutic effects. Predicted pharmacological profiling [12] of natural products could be performed with computer program PASS (Prediction of Biological Activity for Substances).

2. PASS APPROACH

The computer program PASS is the product of ideas that originated some 35 years ago from the development of a National Registration System for New Chemical Compounds in the former USSR [13]. Since all research institutes and universities in the Soviet Union were state-owned, it was intended that all chemists who synthesized new chemical compounds or extracted them from natural sources should register the structural formulae of those compounds, together with the associated information about physico-chemical properties and biological activities. In order to identify probable biological activities and to select the most promising compounds for biological testing there was a proposition that during the registration process, these structures must be analyzed using computer-aided methods.

The first attempts to develop computer-aided methods for the prediction of biological activity from chemical structures were made in the 1970s [14, 15]. Further progress was made in the following decade [16-19], but for various reasons the problem was not completely solved at that time. However, the obtained experience had provided the basis for the later development of PASS [20, 21].

From the beginning, PASS target was the prediction of many kinds of biological activities from the structural formulae of chemical compounds. The PASS algorithm was based on the concept of the "biological activity spectrum", which is an intrinsic property of a compound. It reflects all its different biological activities that arise from interactions with biological entities [20-26]. It should be noted that this definition differs significantly from some other definitions of "biological activity profile" or "biological activity spectrum" published in the literature [27-31], because in PASS the biological activity spectrum represents a theoretical estimate for general biological potential of the compound under study.

* Address correspondence to this author at the Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., 10, Pogodinskaya Street, Moscow, 119121, Russia; Tel.: +7 499 2460920; Fax: +7 499 2450857;
E-mail: vladimir.poroikov@ibmc.msk.ru

1381-6128/10 $55.00+.00

2.1. Biological Activities Predicted by PASS

The latest version of PASS (9.1) predicts 3750 kinds of biological activity with the average prediction accuracy of about 95% (Supplement **1**). PASS could predict 3300 kinds of biological activities in 2007, 2500 activities - in 2005, 541 activities - in 1998, and only 114 activities - in 1996. Thus, the information about new biologically active compounds and novel activities is permanently collected and added into the PASS training set.

In PASS 9.1 the default list of predictable biological activities includes: 261 pharmacotherapeutic actions (e.g., *anxiolytic*); 66 anti-infective actions (e.g., *antileishmanial*); 72 actions blocking a certain process (e.g., *apoptosis antagonist*); 40 actions stimulated a certain process (e.g., *apoptosis agonist*); 140 actions blocking activity of certain endogenous substance (e.g., *acetylcholine antagonist*); 71 actions stimulating activity of certain endogenous substance (e.g., *acetylcholine agonist*); 5 actions blocking a release of a certain endogenous substance (e.g., *cytochrome C release inhibitor*); 9 actions stimulating a release of a certain endogenous substance (e.g., *acetylcholine release stimulant*); 9 actions blocking an uptake of a certain endogenous substance (e.g., *adenosine uptake inhibitor*); 2219 actions inhibiting a certain enzyme (e.g., *12 lipoxygenase inhibitor*); 41 actions stimulating action of a certain enzyme (e.g., ATPase stimulant); 268 actions blocking a certain receptor (e.g., *5 hydroxytrypamine 1 agonist*); 121 actions stimulating a certain receptor (e.g., *5 hydroxytrypamine 1 antagonist*); 28 actions blocking a certain channel (e.g., *chloride channel antagonist*); 5 actions stimulating a certain channel (e.g., *calcium channel agonist*); 28 actions blocking a certain transporter (e.g., *GABA transporter 1 inhibitor*); 128 actions that is a substrate of a certain metabolic enzyme (e.g., *CYP3A4 substrate*); 24 actions inhibiting a certain metabolic enzyme (e.g, *CYP3A4 inhibitor*); 13 actions inducing a certain metabolic enzyme (e.g., *CYP3A4 inducer*); 28 actions inhibiting a certain protein (e.g., *collagen inhibitor*); 8 actions inhibiting an expression of a certain transcription factor (e.g., *transcription factor Rho inhibitor*); 2 actions stimulating an expression of a certain transcription factor (e.g., *TP53 expression enhancer*); 389 actions that cause a certain adverse/toxic effect (e.g., *carcinogen*).

In PASS we describe biological activities qualitatively ("active" or "inactive"). The qualitative presentation allows integrating the information about biologically active compounds collected from many different sources into the general PASS training set. Any property of chemical compounds, which is determined by their structural peculiarities, can be used for prediction by PASS. It is obvious, that the applicability of PASS is broader than the prediction of biological activities. For instance, this approach was successfully used for prediction of such general property of organic molecules as drug-likeness [32].

2.2. Chemical Structure Description

We have chosen 2D structural formulae as the basis for description of chemical structure because this is the only information available at the early stage of research. Thus, using the structural formula as an input data, one can obtain the estimates of biological activity profiles even for virtual molecules, prior to their chemical synthesis and biological testing.

Many different characteristics of chemical compounds can be calculated on the basis of structural formulae [33]. In earliest versions of PASS [20, 21] we used the Substructure Superposition Fragment Notation (SSFN) codes [34]. However, SSFN, like many other structural descriptors [35], reflects rather abstract level of chemical structure than the nature of ligand-target interactions, which are the molecular mechanisms of biological activities. The Multilevel Neighbourhoods of Atoms (MNA) descriptors [36] have certain advantages in comparison with SSFN. These descriptors are based on the molecular structure representation, which includes the hydrogens according to the valences and partial charges of other atoms and does not specify the types of bonds. MNA descriptors are generated as recursively defined sequence:

- zero-level MNA descriptor for each atom is the label *A* of the atom itself;
- any next-level MNA descriptor for the atom is the sub-structure notation $A\left(D_1D_2...D_i...\right)$,

where the descriptor D_i of each successive level is a concatenation of the zero-level descriptor of the atom and, enclosed in parentheses, a lexicographically ordered list of descriptors of the previous level of its nearest neighbors [36].

The atoms label may include not only the atomic type but also any additional information about the atom. In particular, if the atom is not included into the ring, it is marked by "-". The neighbour descriptors $D_1D_2...D_i...$ are arranged in unique lexicographic order. Iterative process of MNA descriptors generation can be continued covering first, second, etc. neighbourhoods of each atom.

The molecular structure is represented in PASS by the set of unique MNA descriptors of the 1st and 2nd levels. We consider that the substances are *equivalent* when they have the same set of MNA descriptors. Since MNA descriptors do not represent the stereochemical peculiarities of a molecule, the substances which structures differ only stereochemically, are formally considered as *equivalent*.

2.3. SAR Base

PASS estimations of biological activity spectra of new compounds are based on the structure-activity relationships data and the knowledgebase (SAR Base), which accumulates the results of the training set analysis. PASS 9.1 training set includes 205873 of the known biologically active substances (drugs, drug-candidates, pharmaceutical leads, and toxic compounds). Since new information about biologically active compounds is discovered regularly, we perform a special informational search and analyse the new information, which is further used for updating and correcting the PASS training set. SAR Base is obtained during the training procedure, which is performed using each new PASS training set.

2.4. Biological Activity Spectrum Estimation

PASS algorithm of the biological activity spectrum prediction is based on Bayesian estimates of the probabilities of a molecule belonging to the classes of active and inactive compounds, respectively. The mathematical method is described in several publications [23-26], and its details will not be discussed in this paper. Only general description necessary for interpretation of prediction results is presented here.

Since the main purpose of PASS is to predict of activity spectra of a new molecules, the general principle of the PASS algorithm is the exclusion from SAR Base the substances, which are *equivalent* to the substance under prediction (see above).

The structural formula of a molecule, whose prediction should be performed by PASS, is presented as a MOL file (for the set of molecules - as a SDF file). The predicted activity spectrum is presented in PASS by the list of activities with the probabilities "to be active" (Pa) and "to be inactive" (Pi) calculated for each activity. The list is arranged in descending order of Pa-Pi; therefore, more probable activities are at the top of the list. Only activities with Pa>Pi are considered as probable for a particular compound. The list can be shortened at any desirable cutoff value, but Pa>Pi is used by default. If the user chooses rather high value of Pa as a cutoff for selection of probable activities, the chance to confirm the predicted activities by the experiment is high too, but many existing activities will be lost. For instance, if Pa>90% is used as a cutoff, then about

90% of real activities will be lost; for Pa>80%, the part of lost activities is 80%, etc.

General PASS interface with representation of prediction results for a set of molecule is shown in Fig. (**1**).

It is necessary to keep in mind that the probability Pa reflects the similarity of a molecule under prediction with the structures of molecules, which are the most typical in a sub-set of "actives" in the training set. Therefore, usually there is no direct correlation between the Pa values and quantitative characteristics of activities.

If the structure of active and potent compound does not resemble any typical structure of "actives" from the training set, then the Pa value calculated for such compound may be rather small (even negative Pa-Pi values could be observed). This may be explained by the following calculation. Assume that the values Pa for "active" and Pi for "inactive" are distributed uniformly, then, for instance, when Pa=0.9, then the corresponding "active" estimates are less in 90% of compounds from the training set and only in 10% these values are higher. If the investigator declines the proposition that the studied compound is active, then it leads to a wrong decision with probability 0.1. In case when Pa<0.5 and Pa > Pi, then the corresponding estimates are higher for more than a half of the "active" compounds from the training set. By declining the proposition that this compound is active, the investigator will make a wrong decision with probability less than 0.5. In such cases the probability to experimentally confirm this kind of activity is small, however, if being confirmed, this structure has a high novelty and may become NCE in more than 50% cases.

If the predicted biological activity spectrum is wide, then the structure of the compound is quite simple and does not contain peculiarities responsible for the selectivity of its biological action.

If it appears that the structure under prediction contains several new MNA descriptors (in comparison with the descriptors from the compounds of the training set), then the structure has low similarity with any structure from the training set, and the results of prediction should be considered as rough estimates.

Based on these criteria, on the basis of a compromise between the novelty of the expected pharmacological action and the risk to obtain the negative result in the experimental testing, one may choose which activities must be tested for the studied compounds. Certainly, one should also take into account a particular interest to some kinds of activities, experimental facilities, etc.

2.5. PASS Validation

Leave one out cross-validation (LOO CV) for the whole PASS 9.1 training set, which includes 205873 substances with 3750 kinds of biological activity, provides the estimates of PASS prediction accuracy during the training procedure. Average accuracy of prediction is about 95.3% according to the LOO CV estimation, while for the different kinds of activity the prediction accuracy varies from 70.7% (antineoplastic, myeloid leukemia) to 99.9% (p21-activated kinase 1 inhibitor).

The accuracy of PASS predictions depends on several factors, of which the quality of the training set seems to be the most important one. A perfect training set must include the comprehensive information about known and possible biological activities for each compound. In other words, the whole *biological activity spectrum* should be thoroughly investigated for each compound included into the PASS training set. Unfortunately, there is no a database with information about biologically active compounds tested against each kind of biological activities. Therefore, the information concerning the known biological activities for any compound is always incomplete.

We investigated the influence of the information's incompleteness on the prediction accuracy for new compounds. About 20000 "principal compounds" from MDDR database (SYMYX

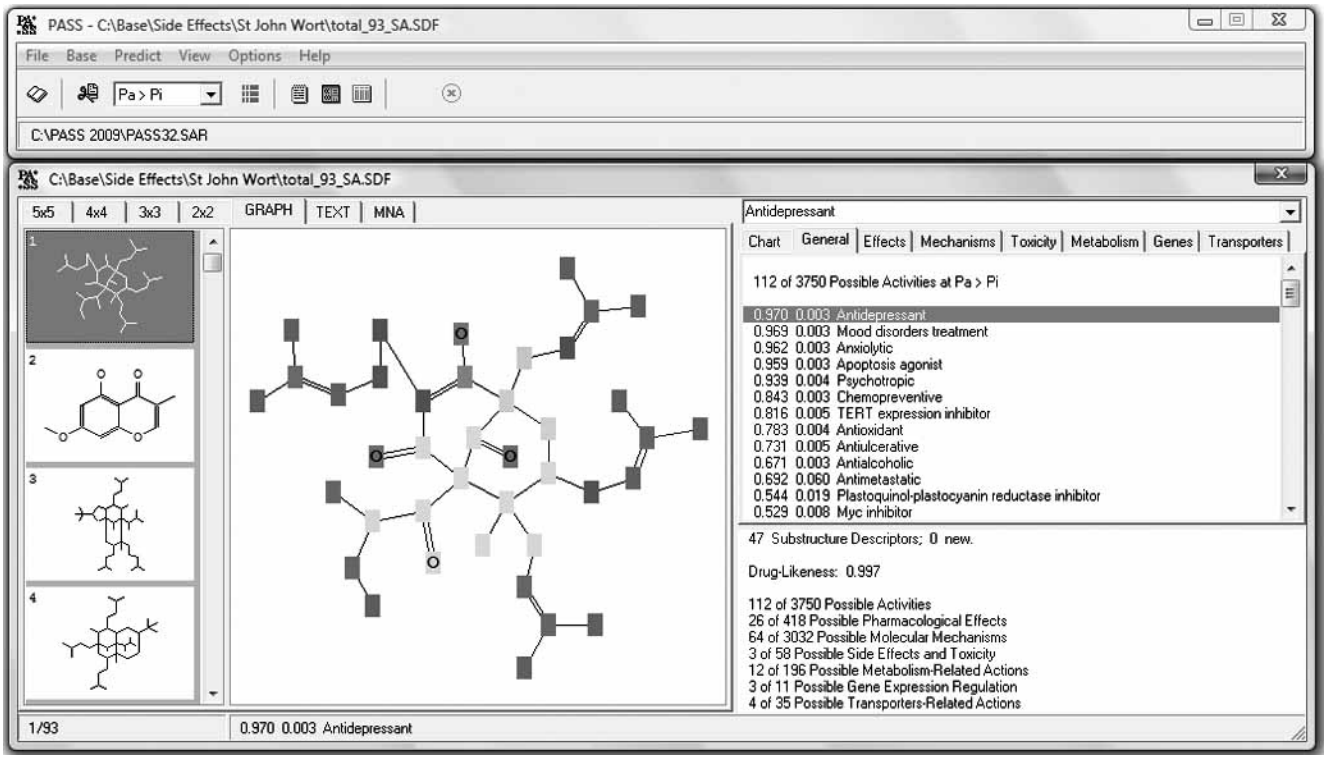


Fig. (1). General PASS interface with representation of prediction results for a set of molecules. Input structural formulae (left); predicted activity spectrum (right) for the selected molecule (central part). "Known activities" - the list of activities for the selected molecule included into the PASS training set. Probable activities are selected at the cutoff value Pa>0.5.

MDL) were used to create the heterogeneous training and evaluation sets. 20, 40, 60, 80% of information was randomly excluded from the training set. Either structural data or biological activity data were removed in two separate computer experiments. In both cases it was shown that even if up to 60% of information is excluded, the results of prediction are still satisfactory [37]. Thus, despite the incompleteness of information in the training set, PASS algorithm is robust enough to achieve the reasonable results of predictions.

PASS predictions were performed for about 250000 molecules from Open NCI database [38]. This information is available at the NCI web-site (http://cactus.nci.nih.gov/ncidb2/) in a searchable mode. One could combine different terms in a query using Boolean operators. For example, with a query "angiogenesis inhibitor AND Pa>0.9 AND Pi<0.2 NOT acid NOT amide" we identified 85 hits. Seven compounds were tested in NCI and four compounds showed an angiogenesis inhibitory activity in the range of 10-100 μM. [38]. Also, on the basis of results of anti-HIV testing of compounds from Open NCI database, we estimated that using PASS predictions one can significantly (up to 17 times) increase the fraction of "actives" in the selected sub-set [38].

PASS INet service (http://www.ibmc.msk.ru/PASS) provides the predictions via Internet for any registered user free of charge. It has been available since 2000 [23, 39, 40]. The input is a MOL file or drawing of a structural formula using Marvin applet. By December the 1st, 2009 the number of registered users exceeded 5000, and over 115000 predictions were made. Based on the prediction results, the researchers select the most prospective substances for chemical synthesis and biological testing. Comparison of PASS prediction results with the experiments provides the independent validation of the approach versus compounds from different chemical series with various kinds of biological activity. Currently, over twenty publications described the coincidence of PASS predictions with the experiment. For example, due to the PASS predictions, new antileishmanial agents were found among 7-substituted 9-chloro and 9-amino-2-methoxyacridines [41], 2 substitution-bearing 6-nitro- and 6-amino-benzothiazoles [42], beta-carboline alkaloids [43]; new anxiolytics were found among quinazolines [44], thiazoles, pyrazoles, isatins, a-fused imidazoles and other chemical series [45]; new anti-inflammatory agents were found among 1-acylaminoalkyl-3,4-dialkoxybenzene derivatives [46], substituted amides and hydrazides of dicarboxylic acids [47]; etc. (for review - see [40]).

Also, on the basis of PASS predictions new antihypertensive and antiinflammatory agents with dual mechanisms of actions were discovered [48, 49], which demonstrated the capability of PASS in finding the multitargeted drugs.

2.6. PharmaExpert: Tool for Analysis of PASS Predictions

PharmaExpert [50] was developed to analyze the biological activity spectra of substances predicted by PASS. This software provides a flexible mechanism for selecting compounds with the required biological activity profiles. Different kinds of biological activities are divided into six classes: mechanisms of actions, pharmacological effects, toxic/adverse effects, metabolic terms, transporter terms and gene expression terms. Mechanism of an action reflects the interactions of biologically active compounds with biological entities at macromolecular level, for example, *acetylcholinesterase inhibitor*, *acetylcholine release inhibitor* or *alpha 1 adrenoreceptor agonist*. The pharmacological effect reflects the pharmacological action or pharmacotherapeutic application of the compound e.g., *antiischemic*, *anxiolytic* or *Alzheimer's disease treatment*. Toxic/adverse effect reflects the specific toxicity (e.g. *mutagenic*, *teratogenic*) or adverse action (*arrhythmogenic*, *anemic*, *nauseant*). Metabolic terms reflect interactions of chemical compounds with metabolic enzymes (e.g., *CYP2D6 inhibitor*, *CYP3A4 substrate*, *CYP 2C9 inducer*). Transporter terms reflect the interaction of compound with the transporters (P-glycoprotein substrate, P-glycoprotein inhibitor, P-glycoprotein inducer, etc.). Gene expression terms reflect the influence of compounds on the expression of certain genes (APOA1 expression enhancer, ErbB-2 expression inhibitor, etc.). PharmaExpert analyzes the relationships between biological activities ("mechanism-effect(s)" and "effect-mechanism(s)"), identifies the probable drug-drug interactions, and searches for compounds acting on multiple targets. The analysis is based on the knowledgebase that is collected from literature during the past 10 years and includes about 8000 "mechanism-effect(s)" and "effect-mechanism(s)" relationships at present time. One example of PharmaExpert analysis is given in Fig. (**2**).

3. PASS APPLICATIONS TO NATURAL PRODUCTS

3.1. Marine Sponge Alkaloids

We applied PASS to more than ninety marine sponges' alkaloids and their synthetic analogs (halitulin, nortoseptins, motuporamines, pyridoacrydines, aaptamines, pyrinodemins, etc.) [51]. Sixty four compounds exhibited cytotoxic activity in tumor cell lines bioassay, and they were suggested for further evaluation *in vivo* by NCI; all others were the novel molecules with unknown biological activities. Predicted antineoplastic activity coincided with the known experimental data in ~80% of cases (51/64). In five cases, when predictions did not correspond to the experiment, the compounds had more than two new MNA descriptors; therefore, for these compounds the reliability of predictions is low (see section 2.4).

In addition to the known antineoplastic activities, PASS predicted some new activities for these molecules. For example, we predicted a possible application of Halitulin for the treatment of psychosexual dysfunction, seborrhea, neurogenic pain and peripheral vascular disease (Fig. (**3**)).

The most probable activities predicted for the analyzed set of marine sponge alkaloids include: interleukin antagonist, cerebral vasodilator, MAP kinase inhibitor, liver fibrosis treatment, 5 HT release stimulant, protein kinase inhibitor, antineoplastic; telomerase inhibitor, antineoplastic alkaloid, antiamyloidogenic, benzodiazepine 1 receptor agonist, etc. [51]. Therefore, marine sponge alkaloids represent a valuable source for finding of new biologically active compounds in many different pharmacotherapeutic areas.

3.2. Hepatoprotective Activity of Triterpenoids

Chemical modification of natural bioactive compounds often brings to the obtaining more effective substances with a different biological activity spectrum [52]. Triterpenoids of the lupane group are considered as suitable parent compounds due to (1) wide spectrum of biological activity including hepatoprotective, antibacterial, anti-inflammatory, antitumor, antiviral and other actions; (2) availability of raw materials (lupeol and betulin) extracted from white birch (*Betula pendula*) [53]. Thus, chemical synthesis of hemisuccinates, hemiphthalates, acetylsalicylates, cinnamates and p-metoxycinnamates of lupeol, betulin and 3-O-acetylbetulin was performed via interaction with corresponding acid anhydrides or acid chlorides. We used PASS predictions to estimate probable biological activity spectra of the synthesized compounds. For all compounds hepatoprotective activity was predicted with the reasonable probability (Pa values varied from 0.75 to 0.80). Through the subsequent biological testing of hepatoprotective action of the synthesized compounds in rats it was shown that all of them have a potent hepatoprotective effect, more potent than the extract of birch bark, betulin and reference drug silibor [53]. Betulin bis hemiphthalate demonstrated the most potent hepatoprotective effect and it was chosen for more detailed studies. Its hepatoprotective action was further confirmed in rat models of hepatitis induced by CCl_4, ethanol and tetracycline [53]. Structural formula and predicted biological activity spectrum of betulin

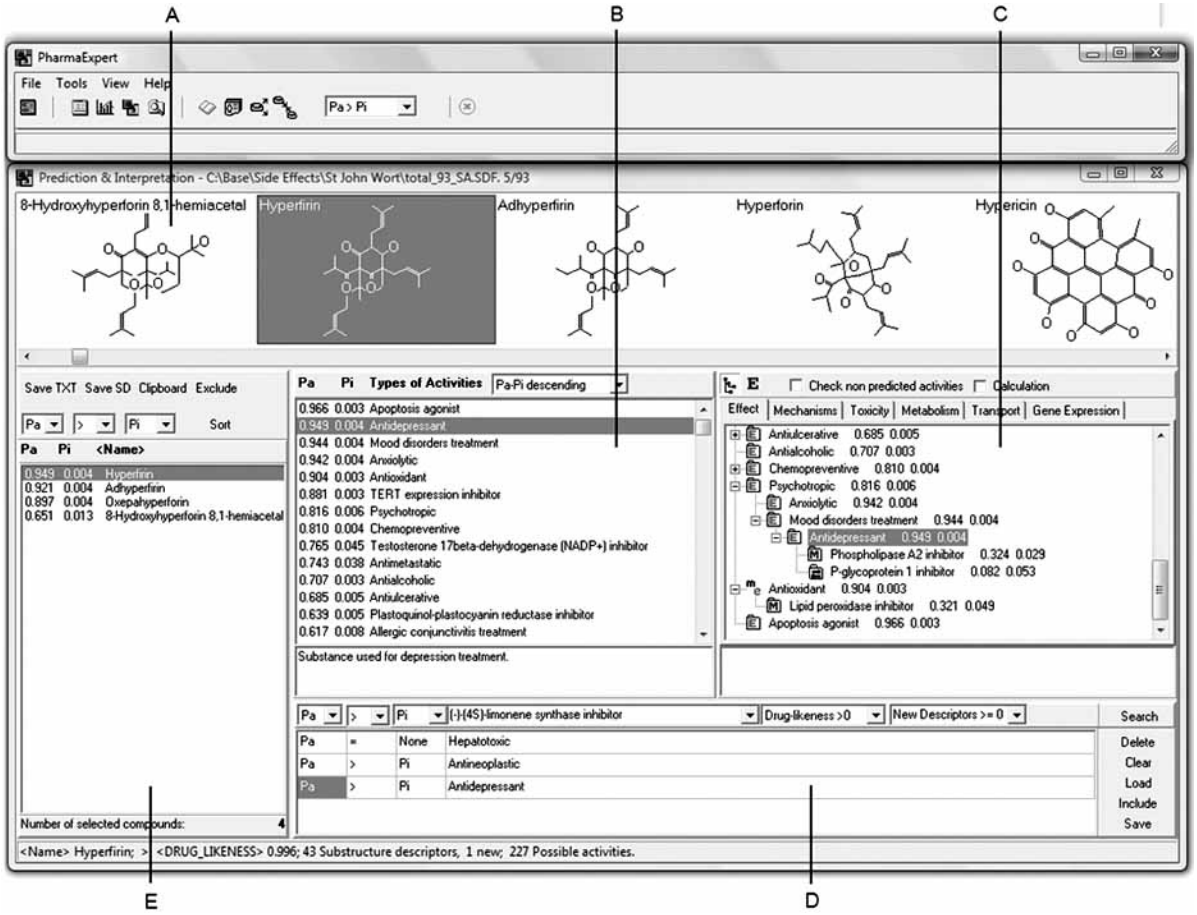
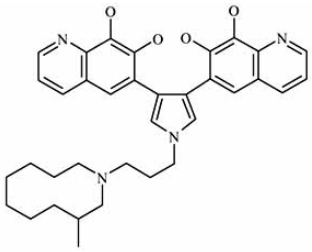


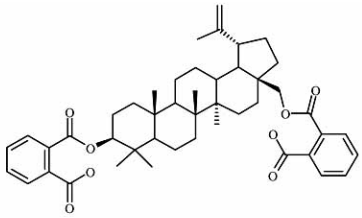
Fig. (2). Results of PASS predictions analysis by PharmaExpert. **A** - structural formulae; **B** - PASS prediction results; **C** - visualization of effect-mechanisms relationships; **D** - query for searching compounds with the required biological activity profile; **E** - search results.



Pa	Pi	Activity
0.81	0.01	Psychosexual dysfunction treatment
0.64	0.02	Calmodulin antagonist
0.64	0.03	Antiseborrheic
0.61	0.02	Antineurogenic pain
0.62	0.04	Cardiovascular analeptic
0.52	0.03	Sigma receptor agonist
0.58	0.14	Vascular (periferal) disease treatment
0.50	0.07	Topoisomerase II inhibitor

Fig. (3). Structural formula and predicted biological activity spectrum for Halitulin (only activities predicted with Pa>0.5 are shown).

bishemiphthalate is shown in Fig. (**4**). As one may see from this data, that no toxic/adverse effects are expected for this compound; thus, betulin bishemiphthalate could be considered as a promising lead substance. Also, in addition to the hepatoprotective effect, some other biological activities are predicted, which could be the reasons for new applications.



Pa	Pi	Activity
0.78	0.01	Hepatoprotector
0.73	0.03	Dermatological
0.71	0.02	Gestagen antagonist
0.71	0.01	Endothelin A receptor antagonist
0.70	0.03	Reverse transcriptase inhibitor
0.63	0.01	Endothelin B receptor antagonist
0.57	0.07	Lipid metabolism regulator
0.56	0.06	14-alpha lanosterol demethylase inhibitor
0.55	0.04	Nerve growth factor agonist
0.53	0.05	Kidney disease treatment

Fig. (4). Structural formula and predicted biological activity spectrum for Betulin bishemiphtalate (only activities predicted with Pa>0.5 are shown).

3.3. Computer-Aided Design of Macrolides with the Required Activity Profiles

Secondary metabolites produced by different organisms display diverse biological activities, thus having a potential of being pharmacological agents [54]. An example of secondary metabolites are macrolides (amphotericin, nystatin, erythromycin, methymycin

etc.), produced mainly by bacteria and some fungi. Macrolides display a wide range of biological activities, including antibacterial, antifungal, antiviral, immunosuppressive, antineoplastic etc. [55]. Macrolides are biosynthetically assembled by the modular polyketide synthase (PKS) enzymes [56] (example of erythromycin synthesis see in Fig. (**5**)).

Mode of action of PKS theoretically leads to an extremely large number of analogues, which can be produced upon combinatorial manipulation with these enzymes. Experimental engineering of all possible variants of a certain PKS system as well as isolation and testing of the resulting macrolides in the laboratory are unfeasible. Due to the achievements in postgenomic technologies, rational approach can be applied to design of the macrolides' structures based on the information on their biosynthesis. Mode of action of PKS can be simulated *in silico*, thus leading to the generation of virtual libraries of macrolide's analogues [57]. Such combinatorial libraries can be analyzed using computer-aided prediction of biological activities, physico-chemical properties, drug-likeness etc., and yielding compounds with the desired properties. This approach provides rational selection of macrolide's molecules with the desired biological activities.

To perform a rational design of new polyketides with the required biological activity spectrum we developed BioGenPharm software, which generates combinatorial libraries of polyketides based on the user-defined input parameters, performs PASS predictions of biological activity spectra for the generated structures, and PharmaExpert selection of molecules with the required properties [58]. Validation of PASS prediction ability for polyketides was carried out against the test set containing 242 natural macrolides from the Dictionary of Natural Products [59]. The mean prediction accuracy was 75.5% [58].

We show applicability of the proposed approach by generation of a virtual library of the erythromycin analogues and by selection of substances with low probability of the hepatotoxic effect. Erythromycin is reasonably well tolerated by patients; the most serious side-effect observed during administration of all pharmaceutical forms of erythromycin is cholestasis [60]. Administration of erythromycin may also cause the increase of serum transaminases and the development of cholestatic hepatitis. The mechanism of toxic liver damage by erythromycin administration remains unclear: it was shown that erythromycin does not inhibit bile acids (taurocholic acid) transport [61]. We have searched for erythromycin analogues, whose hepatotoxic activity was predicted with a minimal probability. The erythromycin analogues library was generated using the following set of parameters: propionate starter, 6 extender modules, and cladinose glycosylation by the 3rd carbon atom, desosamine glycosylation by the 5th carbon atom, hydroxylation by the 6th and 12th carbon atom. Compounds were selected at Pa>0.8 and Pa<0.6 for antibacterial and hepatotoxic activities, respectively.

As a result, we obtained 17 erythromycin analogues which satisfy all the criteria settings. Among them, the lowest probability of hepatotoxic activity and the highest probability of antibacterial activity were observed for a molecule No. 2456. Hepatotoxic activity was predicted for this compound with Pa=56.5%, while for erythromycin hepatotoxicity was predicted with Pa=94.1%; thus, the most perspective erythromycin analogue was a compound No. 2456. The information on domain composition of PKS required for biosynthesis, necessary for the design of corresponding microorganism-producer, was obtained with BioGenPharm program (Fig. (**6**)).

4. IN SILICO ANALISIS OF HYPERICUM PERFORATUM AS A HERBAL MEDICINE

We selected one of the most popular herbal medicines St John's Wort (*Hypericum perforatum*) as an example for the case study. In this study, using PASS and PharmaExpert we estimated *in silico* the most probable pharmacological mechanisms and effects that could be expected for individual components of St John's Wort (SJW), likely interactions between the components in extracts, possible peculiarities in pharmacological action of herbs collected in different areas.

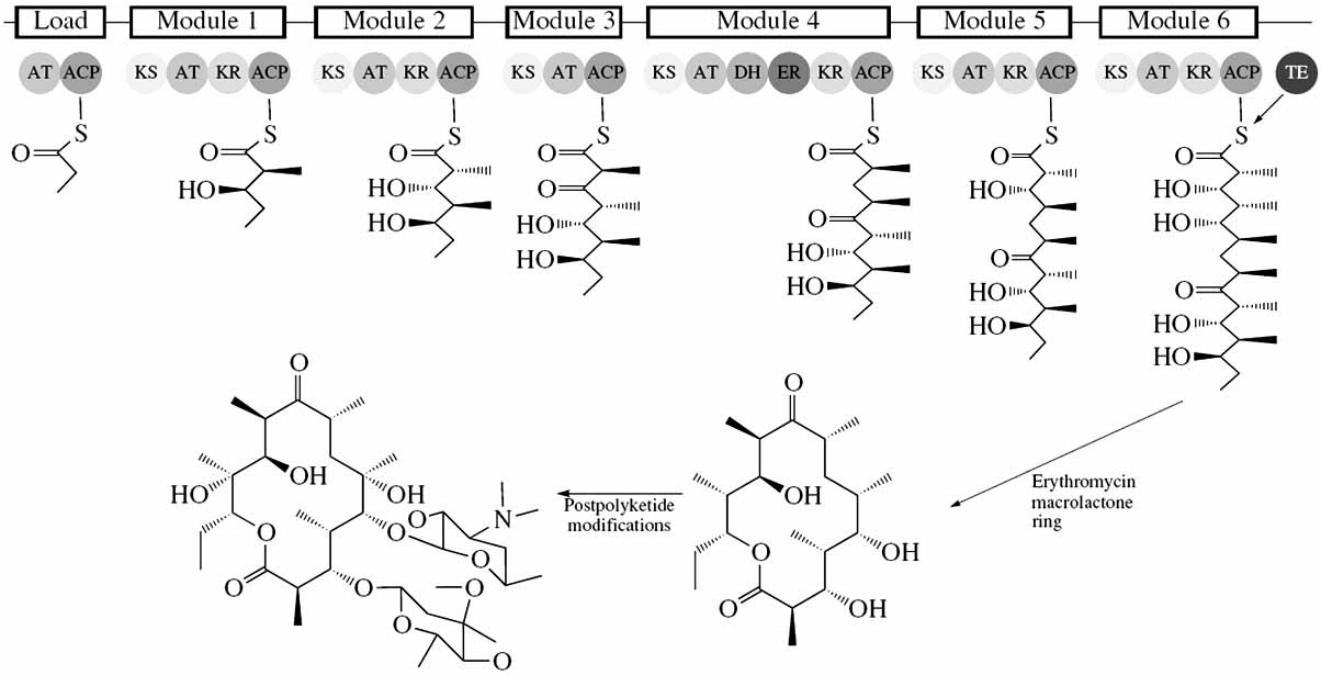
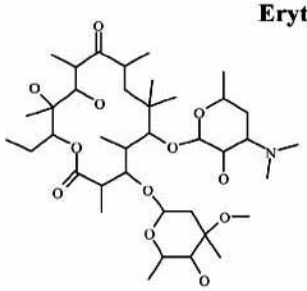


Fig. (5). Erythromycin synthesis by modular polyketide synthases. The three genes EryAI-III encode three proteins of PKS: DEBS1 (the loading module, modules 1, 2) DEBS2 (modules 3, 4), DEBS3 (modules 5, 6, TE domain). Thus, PKS consists of the loading module, six extension modules, and TE domain. Each module includes from three to six domains: AT-acyl transferase, ACP-acyl carrier protein, KS–ketosynthase, KR-ketoreductase, DH-dehydratase, ER–enoyl reductase.

Erythromycin A

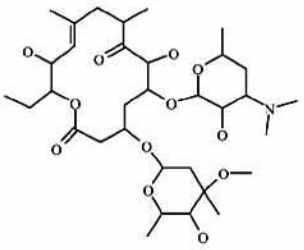


[Loading: Propionate]
[KS-ATmet-**-**-KR-ACP]
[KS-ATmet-**-**-KR-ACP]
[KS-ATmet-**-**-**-ACP]
[KS-ATmet-DH-ER-KR-ACP]
[KS-ATmet-**-**-KR-ACP]
[KS-ATmet-**-**-KR-ACP]

PASS ACTIVITY SPECTRUM at Pa > 0.8

Pa	Pi	
0.991	0.001	Antibiotic Macrolide-like
0.990	0.003	Cytochrome P450 inhibitor
0.982	**0.000**	**Antibacterial**
0.975	0.000	Antibiotic
0.969	0.001	Mucolytic
0.968	0.002	Protein 50S ribosomal subunit inhibitor
0.961	0.000	Transcription factor NF kappa B inhibitor
0.957	0.003	Antiprotozoal (Toxoplasma)
0.955	0.003	Protein synthesis inhibitor
0.951	0.003	HERG channel antagonist
0.943	0.005	QT interval prolongation
0.937	0.000	Antirickettsial
0.941	**0.004**	**Hepatotoxic**
0.932	0.001	Expectorant
0.922	0.002	Transcription factor inhibitor
0.870	0.003	Gastrointestinal motility stimulant
0.853	0.002	Prokinetic
0.844	0.014	Toxic
0.819	0.005	Antidyskinetic

Erythromycin A analog No. 2456



[Loading:Propionate]
[KS-***ATmal***-**-**-KR-ACP]
[KS-ATmet-***DH***-**-KR-ACP]
[KS-ATmet-***DH-ER-KR***-ACP]
[KS-***ATmal***-**-**-**-ACP]
[KS-***ATmal***-**-**-KR-ACP]
[KS-***ATmal***-**-**-KR-ACP]

PASS ACTIVITY SPECTRUM at Pa > 0.8

Pa	Pi	
0.904	0.007	Antiprotozoal (Toxoplasma)
0.888	0.001	Transcription factor NF kappa B inhibitor
0.885	0.009	Toxic
0.869	**0.001**	**Antibacterial**
0.868	0.004	Transcription factor inhibitor
0.845	0.001	Antibiotic
0.839	0.006	HERG channel antagonist
...		
0.565	**0.109**	**Hepatotoxic**

Fig. (6). Structural formulae, predicted biological activities, module composition for PKS type I required for synthesis for erythromycin and its analogue No. 2456 obtained with BioGenPharm software. ATmal - malonyl-specific acyltransferase domain; ATmet - methylmalonyl-specific acyltransferase domain.

SJW is mainly used in clinical practices for the treatment of a mild or moderate depression. In Germany, St John's Wort is the leading treatment against depression, outselling Fluoxetine (Prozac®) [62]. The main advantage of this herb, which promoted its wide use, is less side effects in comparison with traditional antidepressants (Fig. (**7**)).

SJW is also recommended for the treatment of bacterial and viral infections, skin wounds, tumor, peptic ulcers and inflammation; it has antiseptic, choleretic, spasmolytic, analgesic, antioxidant and photosensitive activities [62-64]. These pharmacotherapeutic effects and SJW interactions with the drugs, mainly on CYP3A4 level, is the reason for an intensive study of both constituents of the extract and their biological activity. Using "St John's Wort" as a keyword one can find about 2000 articles in PubMed.

Numerous biologically active constituents were identified in St John's Wort, including naphthodianthrones (e.g. hypericin and its derivatives), phloroglucinols derivatives (e.g. hyperforin) and flavonoids (e.g. rutin, quercetin, quercitrin and biapigenin) [63]. *Hypericum perforatum* is widely distributed herb. It grows in America, Europe, Asia, Africa and Australia. Some of the extract components may be different depending on the region. This may influence on both therapeutic and side effects of SJW extracts.

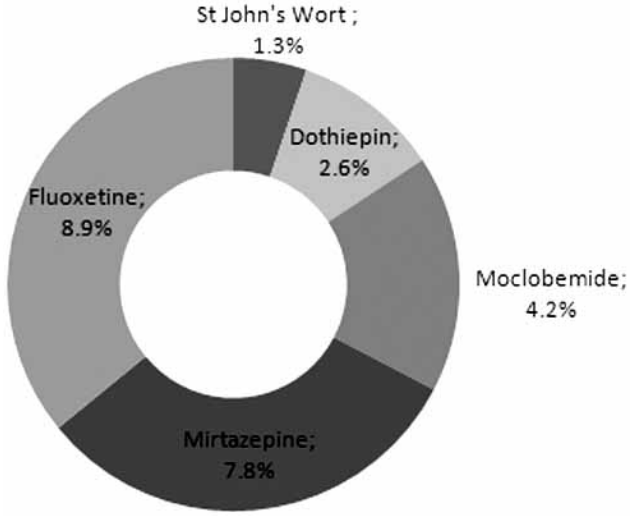


Fig. (7). Average frequencies of side effects observed in patients treated with St John's Wort and traditional antidepressants.

Therefore, the knowledge of probable activity of each component and the interaction between them is important for estimating the clinical usage of St John's Wort obtained from different places. Such information could be obtained by computer prediction of biological activity spectra with PASS.

4.1. General Analysis of Individual Constituents in St John's Wort

We have found available information about ninety three compounds, which were determined in St John's Wort collected from different regions (Supplement **2**). The basic compounds represented in the majority of studied extracts of *Hypericum perforatum* include: hyperforin, hypericin, pseudohypericin, rutin, quercetin, quercitrin, isoquercitrin, biapigenin, alpha-pinene, hyperoside and chlorogenic acid. Information on the known biological activities collected from the main reviews of St John's Wort compositions and contained in PASS training set is also provided in Supplement **2**.

Prediction of biological activity spectra for all 93 compounds was carried out with PASS 9.1. The distribution of predicted activities' number for 93 compounds depending on the predicted probability is shown in Fig. (**8**). Fig. (**8**) shows that 817 from 3750 biological activities, which can be predicted by PASS, are expected with probability Pa>0.7 at least for one constituent from the St John's Wort extract.

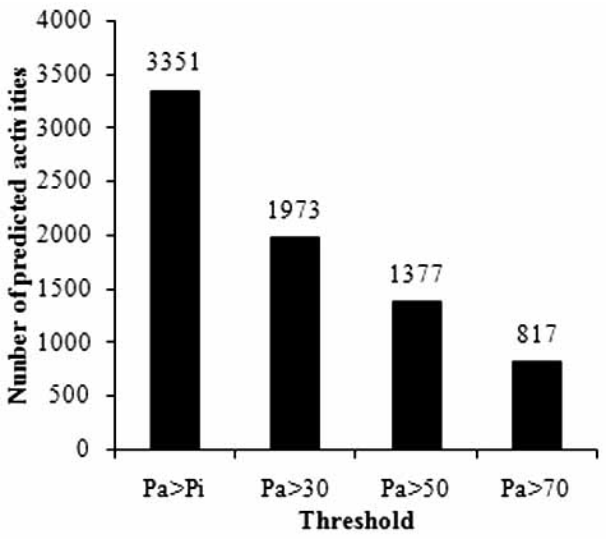


Fig. (8). Number of predicted activities for 93 compounds isolated from St John's Wort *versus* the probability Pa.

The number of compounds with a particular amount of predicted activities (N) is shown in Fig. (**9**). It is clear that at the cutoff Pa>0.7 maximum of distribution is located at N=10. Also, about 38% of compounds have no more than 20 activities predicted with Pa>0.7, however over 45% of compounds have more than 40 predicted activities. It means that only a small part of compounds exhibit some selectivity, while the majority of St John's Wort constituents are multitargeted agents.

4.2. Prediction of Pharmacotherapeutic Effects

Pharmacotherapeutic effects predicted at least for one of the analyzed compounds with Pa>0.5 (analgesic and antiviral activities - with less probability Pa>0.4) are listed in Table **1**. Also, in Table **1** one could find the IDs of compounds designated according to the Supplement **2**. The IDs allow to label the compounds exhibiting particular kinds of activity (IDs are arranged in the descending order of probability of the appropriate activity). Table **1** shows that

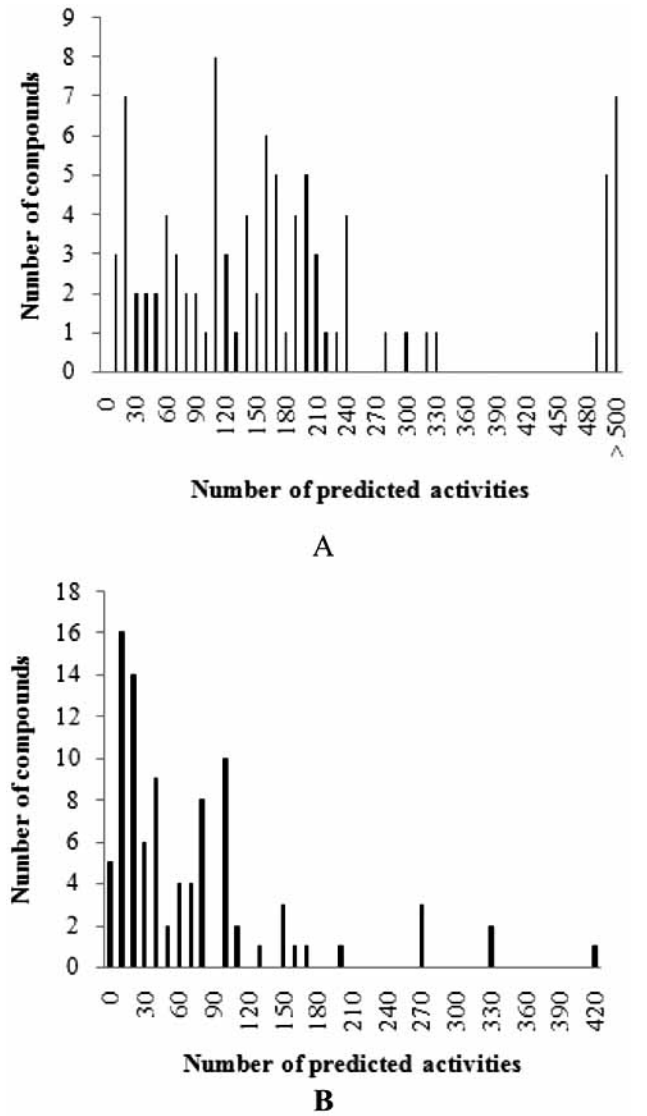


Fig. (9). Amount of compounds from St John's Wort, which have a particular number of predicted activities at different thresholds (A - Pa>0.5; B - Pa>0.7).

the most frequent activities predicted for SJW constituents are antioxidant, antiinflammatory, antineoplastic, choleretic, and antiulcerative.

4.3. Prediction of Side Effects

As was mentioned above, St John's Wort reveals fewer side effects in comparison with the traditional antidepressants. Analyzing the prediction of biological activity spectra for 93 SJW constituents, one may determine the compounds probably causing particular side effects. The most frequent side effects of St John's Wort mentioned in [62] are shown in Table **2**. This data could be compared with the appropriate predictions obtained with the special version of PASS trained using information on the known side effects of drugs [60]. All known side effects have been predicted except for the Gastrointestinal symptoms, which are not covered by this version of PASS (PASS terms used in Table **2** sometimes synonymous to the names of adverse effects from [62]). Moreover, PASS predictions help to identify the compound(s), which likely have a particular side effect. Such information may be useful in further studies of individual components of SJW as a potential pharmaceutical leads.

Table 1. Pharmacotherapeutic Effects Predicted with Pa>50% and the IDs of Corresponding Compounds

No.	Activity	IDs
1	Analgesic*	58, 61
2	Antibacterial	23, 80, 24, 48, 30
3	Antidepressant	7, 19, 1, 5, 3, 6, 13, 4
4	Antiinflammatory	55, 20, 21, 52, 54, 30, 32, 24, 22, 53, 49, 79, 80, 43, 27, 88, 76, 23, 35, 81, 50, 48, 68, 51, 78, 39, 91, 69, 77, 63, 31, 86, 28, 56, 38, 14, 92, 60, 93, 65, 42, 57, 82, 29
5	Antineoplastic	66, 43, 71, 61, 38, 72, 35, 24, 23, 33, 25, 80, 48, 16, 8, 42
6	Antioxidant	30, 20, 21, 32, 80, 22, 53, 52, 49, 79, 48, 27, 76, 78, 5, 91, 7, 81, 50, 29, 6, 51, 31, 92, 24, 82, 23, 54, 1, 83, 4, 55, 33, 93, 77, 87, 28, 56, 86, 14, 19, 41, 13, 85, 57, 2, 15, 36, 58, 84, 12
7	Antiseptic	25, 89, 84, 67, 86, 37, 74, 47
8	Antiulcerative	54, 85, 1, 41, 7, 5, 6, 36, 27, 55, 68, 35, 15, 86, 77, 89
9	Antiviral*	78, 87
10	Choleretic	50, 81, 51, 77, 86, 76, 27, 68, 88, 92, 2, 18, 93, 84, 56, 31
11	Photosensitizer	18
12	Spasmolytic	89, 85, 91, 29, 28, 58, 44

*Predicted with Pa>0.4

**ID is the compound number in Supplement 2

Table 2. Known Side Effects of St John's Wort in Comparison with PASS Predictions

Side Effect	ID*	PASS Prediction	N**	IAP***
Dizziness	89, 37, 47, 74, 39, 75, 59, 84, 11, 70, 34, 30, 92	Dizziness	223	74.9
Dry mouth	89, 86, 75, 59, 47, 37, 74, 77	Xerostomia	62	87.6
Gastrointestinal symptoms		Not Predictable		
Headache	89, 52, 47, 74, 37, 30, 59, 75, 85, 84, 21, 79, 86, 34, 70, 11, 53, 49, 22, 91, 77, 32, 39, 65, 44, 62, 78, 48, 20, 56, 40, 38	Headache	224	75
Insomnia	52, 30, 14, 32, 79, 49, 22, 53, 31, 82, 59, 75, 92, 56, 11, 70, 34, 74, 37, 47, 93, 86, 89, 69, 77, 57, 84, 91, 48, 76, 83, 65, 78, 21, 62, 87, 28, 20, 38, 2, 80, 45, 16	Insomnia	86	74.3
Photophobia	18	Photosensitizer	59	94.5
Restlessness	59, 75, 86, 84, 47, 74, 37, 89, 77, 45, 34, 11, 70, 60, 56, 50, 81, 14, 73, 65, 44, 51, 67, 57, 64, 46, 82, 93, 92	Akathisia	46	75.7
Skin reactions	41, 36, 85, 67, 74, 47, 37, 34, 11, 70, 62	Skin irritative effect	1030	96.7
Tiredness Fatigue	91, 78, 65, 39, 63, 75, 59, 44, 84, 40, 56, 77, 28, 72, 86, 38	Lassitude	24	76.9
Tremor	30, 32, 89, 22, 53, 49, 79, 52, 37, 47, 74, 48, 80, 76, 20, 75, 59, 14, 21, 92, 11, 34, 70, 31, 82, 93, 84, 57, 91, 58, 23, 83, 87, 78, 56, 28, 24, 86, 25, 62, 26, 77, 33, 17, 16, 67, 73, 27, 18	Tremor	84	74.7
Vertigo	37, 47, 74, 75, 59, 85, 65, 11, 70, 34, 62, 61, 44, 68, 38, 69, 42, 41, 84, 63, 60, 86, 26, 35, 89, 45, 25, 77, 67, 39, 64, 73, 66, 91, 52, 46, 48, 40, 72, 8, 24, 71, 51, 23, 78, 56, 57, 76, 36, 80, 16, 18, 82, 90, 32, 81, 50, 21, 49, 22, 17, 53, 58, 93, 92, 43, 79, 31, 87, 2, 12, 28, 20	Vertigo	60	65.4

*ID is the compound number in Supplement 2.

**N is the number of compounds with the appropriate side effect in PASS training set.

*** IAP is the Invariant Accuracy of Prediction calculated by LOO CV during the training procedure.

4.4. Prediction of Probable Interactions Between the St John's Wort Constituents

Each compound in St John's Wort has a unique profile of biological activity. The total number of predicted activities for ninety three compounds grouped according to the activity types is represented in Fig. (**10**).

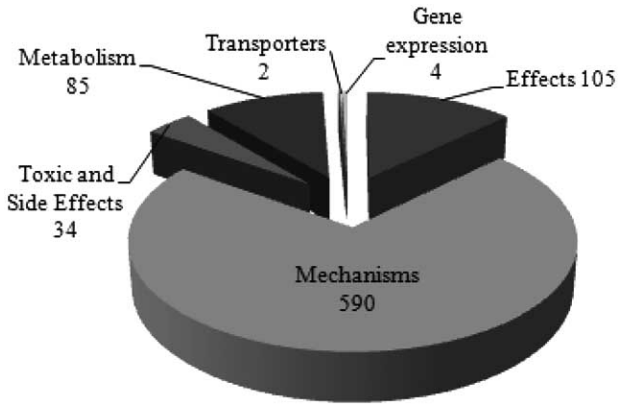


Fig. (10). Distribution of predicted activities for constituents of St John's Wort through the different types of biological activity.

When the St John's Wort is administered to patients in clinics or to experimental animals in preclinical studies, the individual compounds interact with each other pharmacologically that could lead to both positive and negative manifestations. The analysis of predicted biological activity spectra for individual compounds, which can be performed on the basis of knowledge about mechanism-effect relationships, reveals the possible interactions. If either a particular mechanism of action or a few mechanisms of action causing the same pharmacotherapeutic of toxic/side effect were predicted for several compounds, then such interaction could lead to additive/synergistic activity. If several compounds have different mechanisms, which lead to an opposite to pharmacotherapeutic toxic/side effect, then such interaction could lead to the antagonistic activity. The total number of predicted interaction at cutoff Pa>0.7 is given in Fig. (**11**). Fig. (**11**) shows that multiple potential interactions could be found at each of four activity types (mechanisms, pharmacotherapeutic effects, toxic/side actions, and metabolism).

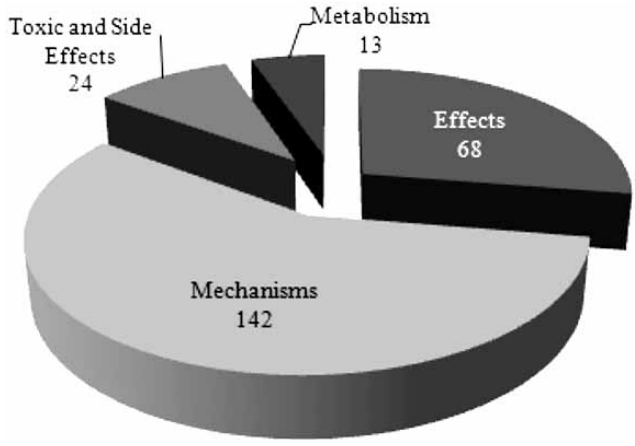


Fig. (11). Possible interactions between the different St John's Wort constituents.

All possible additive/synergistic interactions and their causes identified by PharmaExpert at cutoff Pa>0.7 are presented in Table **1** (Supplement **3**). As one can see from this data, the basic known pharmacotherapeutic effects could be caused by additive/synergistic action of different components of St John's Wort. Based on this analysis, several new pharmacotherapeutic effects caused by additive or synergistic action were also identified (e.g. atherosclerosis treatment, antipsoriatic, neuroprotector, etc.). These findings should be checked in further experimental studies.

We have identified several additive/sinergistic interactions leading to toxic/effects (Table **2**, Supplement **3**) and interactions with metabolic enzymes (Table **3**, Supplement **3**) of St John's Wort compounds. Data on the components responsible for toxic/side effects (Table **2**, Supplement **3**) can be used either for excluding these components from the extracts or for collecting the *Hypericum perforatum* in the regions, where the fraction of these components in the herb is either absent or small. Data on additive/synergistic or antagonistic effects at the level of metrabolism should also be taken into account at dosage determining, particularly, in combination therapy.

We also found 142 possible interactions of St John's Wort constituents at the level of molecular mechanisms of action. This data is partially given in Table **4** (Supplement **3**). Interaction at the level of molecular mechanisms could also lead to additive/ synergistic or antagonistic effects.

4.5. Prediction of Likely Different Actions in St John's Worts Collected in Different Regions

The analysis of predicted biological activity spectra may reveal the features in clinical manifestations of plant extracts prepared using raw materials collected in different geographical regions. This could be illustrated by comparison of essential oils' compositions from *Hypericum perforatum* collected in Southeastern Serbia [65] and Lithuania [66]. The essential oil compositions from Southeastern Serbia and Lithuania contain 14 and 19 compounds, respectively. Six compounds with IDs 11, 35, 43, 46, 47 and 59 (Supplement **2**) are identical.

Table **3** demonstrates the difference between the predicted activities for the samples from the Southeastern Serbia and Lithuania. While the total numbers of mechanisms of action and toxic effects are similar for both sets, the significant difference is observed in therapeutic effects and interactions with drug-metabolizing enzymes. More detailed analysis of predicted toxic/side effects showed that, despite the qualitative identity of the effects predicted for the both sets, the number of compounds that could reveal these effects is different (Table **4**). Hematotoxic, hyperthermic, neurotoxic effects and the inhibition of thrombo-cytopoiesis were predicted for more compounds from essential oil from Serbia. At the same time emetic and reproductive dysfunction effects were predicted for more compounds from essential oil from Lithuania.

The detailed analysis of prediction of the studied subsets interactions with drug-metabolic enzymes shows that the majority of predicted interactions are the same (Table **5**). CYP17 inhibition and CYP2C19 induction only were not predicted for the samples from Southeastern Serbia. The number of compounds predicted as the inducers of CYP2B6 and inhibitors of CYP2B6 and CYP2C8 for essential oil composition from Serbia exceeded those from Lithuania.

Prediction of biological activity spectra for the two sets of compounds and further analysis of possible interactions between the constituents determined the difference in additive/synergistic pharmacotherapeutic effects for essential oil compositions from Serbia and Lithuania. Table **6** shows that the essential oil compositions from Southeastern Serbia and Lithuania have 25 identical additive/synergistic pharmacotherapeutic effects.

Table 3. Number of Predicted of Biological Activity Types for SJW Essential Oil Constituents from Southeastern Serbia and Lithuania

Region	N	Effects	Mechanisms	Toxic/side Effects	Metabolic Effects	Transporters	Gene Interaction
Serbia	14	58	354	14	21	-	3
Lithuania	19	72	356	14	37	1	4
Serbia*	14	32	78	17	20	-	1
Lithuania*	19	36	47	15	14	-	-

N is the number of compounds in the studied set.
* number of activities that could lead to additive/synergistic effects.

Table 4. Predicted Toxic/Side Effects for Essential Oil Compositions from Southeastern Serbia and Lithuania*

No.	Serbia	Lithuania
1	Bronchoconstrictor (2): 74, 47	Bronchoconstrictor (2): 60, 47
2	Convulsant (2): 74, 47	Convulsant (1): 47
3	Emetic (4): 74, 47, 75, 59	Emetic (5): 47, 44, 65, 59, 61
4	Eye irritation, high (2): 74, 47	Eye irritation, high (3): 67, 47, 41
5	Eye irritation, moderate (2): 74, 47	Eye irritation, moderate (2): 47,41
6	Hematotoxic (9): 70, 34, 11, 75, 59, 74, 47, 71, 69	Hematotoxic (5): 11, 60, 68, 59, 47
7	Hypercholesterolemic (5): 72, 71, 75, 59, 73	Hypercholesterolemic (5): 60, 40, 59, 65, 63
8	Hyperthermic (7): 59, 75, 74, 47, 11, 34, 70	Hyperthermic (4): 67, 59, 47, 11
9	Neurotoxic (4): 74, 47, 75, 59	Neurotoxic (2): 59, 47
10	Reproductive dysfunction (2): 74, 47	Reproductive dysfunction (4) : 60, 65, 47, 41
11	Skin irritation, high (5): 47, 74, 11, 70, 34	Skin irritation, high(4): 67, 47, 11, 41
12	Skin irritation, moderate(5): 47, 74, 11, 70, 34	Skin irritation, moderate (3): 41, 47, 11
13	Skin irritative effect (2): 74, 47	Skin irritative effect (3): 41, 67, 47
14	Thrombocytopoiesis inhibitor (4): 75, 59, 47, 74	Thrombocytopoiesis inhibitor (2): 59,47

*Number of predicted activities related to the particular effect is given in brackets; IDs of the appropriate compounds are in accordance with Supplement 2.

Table 5. Prediction of Interactions with Drug-Metabolic Enzymes for Essential Oil Compositions from Southeastern Serbia and Lithuania

No.	Serbia	Lithuania
1	-	CYP17 inhibitor (1): 60
2	CYP2A6 inhibitor (2): 75, 59	CYP2A6 inhibitor (2): 40, 59
3	CYP2B6 inducer (11): 59, 75, 70, 34, 11, 47, 74, 72, 73, 46, 71	CYP2B6 inducer (7): 60, 59, 11, 47, 46, 67, 65
4	CYP2B6 inhibitor (9): 47, 74, 70, 11, 34, 75, 59, 69, 35	CYP2B6 inhibitor (8): 47, 11, 40, 59, 44, 62, 35, 42
5	-	CYP2C19 inducer (1): 64
6	CYP2C8 inhibitor (7): 70, 34, 11, 74, 47, 75, 59	CYP2C8 inhibitor (4): 11, 40, 47, 59
7	CYP2C9 inhibitor (2): 74, 47	CYP2C9 inhibitor (1): 47
8	CYP2D2 inhibitor (2): 75, 59	CYP2D2 inhibitor (1): 59
9	CYP2E1 inducer (2): 74, 47	CYP2E1 inducer (1): 47
10	CYP2E1 inhibitor (4): 74, 47, 75, 59	CYP2E1 inhibitor (3): 47, 41, 59

Table 6. Predicted Additive/Synergistic Therapeutic Effects for Essential oil Compositions from Southeastern Serbia and Lithuania

No.	Serbia	Lithuania
1	Allergic conjunctivitis treatment	Allergic conjunctivitis treatment
2	Alopecia treatment	Alopecia treatment
3	Ankylosing spondylitis treatment	Ankylosing spondylitis treatment
4	Antidote, cyanide	-
5	Antidyskinetic	Antidyskinetic
6	-	Antiepileptic
7	Antihypoxic	-
8	Antiinflammatory	Antiinflammatory
9	Antiinflammatory, intestinal	-
10	Antiinflammatory, pancreatic	Antiinflammatory, pancreatic
11	Antimetastatic	Antimetastatic
12	Antimutagenic	-
13	Antineoplastic	Antineoplastic
14	-	Antineoplastic (gastric cancer)
15	-	Antineoplastic (lymphoma)
16	-	Antineoplastic (non-Hodgkin's lymphoma)
17	-	Antineoplastic (non-small cell lung cancer)
18	-	Antineoplastic (ovarian cancer)
19	Antineurotic	-
20	Antineurotoxic	Antineurotoxic
21	-	Antipruritic
22	Antipsoriatic	Antipsoriatic
23	Antiseborrheic	Antiseborrheic
24	-	Antisecretoric
25	-	Antiulcerative
26	Antiviral (Arbovirus)	Antiviral (Arbovirus)
27	Apoptosis agonist	Apoptosis agonist
28	Cardiovascular analeptic	Cardiovascular analeptic
29	Carminative	Carminative
30	Cholesterol synthesis inhibitor	Cholesterol synthesis inhibitor
31	Cytoprotectant	-
32	Fibrinolytic	Fibrinolytic
33	Gaucher disease treatment	Gaucher disease treatment
34	-	Hepatoprotectant
35	-	Hypolipemic
36	Leukopoiesis stimulant	Leukopoiesis stimulant
37	Mucositis treatment	Mucositis treatment

(Table 6) Contd….

No.	Serbia	Lithuania
38	Proliferative diseases treatment	Proliferative diseases treatment
39	Psychosexual dysfunction treatment	Psychosexual dysfunction treatment
40	Sclerosant	Sclerosant
41	Sialagogue	-
42	Spermicide	Spermicide
43	Vasoprotector	Vasoprotector

At the same time, seven additive/synergistic therapeutic effects (antidote, cyanide; antihypoxic; antiinflammatory, intestinal; anti-mutagenic; antineurotic; cytoprotectant and sialagogue) were predicted for the sample from Serbia only. The other eleven additive/synergistic therapeutic effects were predicted for the sample from Lithuania only (Table **6**). Such information could help to estimate the clinical potential of plant extracts collected from different geographical regions.

5. FUTURE TASKS

In this review we presented some examples of how computer-aided prediction of biological activity spectra could be used for the analysis of pharmacotherapeutic potential of natural products. More examples can be found in our papers [67-73]. The prediction results for the individual constituents identified in herbs, sponges, microorganisms etc. could reveal the hidden potentials of natural products by identification of active substances that caused the effects known from ethnopharmacology and molecular mechanisms of their action. The predicted biological activity spectra provide the basis for new rational combinations of medicinal plants with improved efficacy and safety, as well as the recommendations for the personalization of therapy. This information can also be used for selecting new pharmaceutical leads with the required biological activity profiles [12, 50, 74-76].

The accuracy of biological activity spectra prediction strongly depends on three issues: (1) robustness of the algorithm for the analysis of structure-activity relationships, (2) its orientation on prediction of biological activity of novel molecules, and (3) quality of the training set. The first two issues were thoroughly investigated and implemented in PASS. To provide more reliable estimates, we are permanently collecting new information about biologically active compounds, add it to the training set and update the PASS knowledgebase. Further development of the algorithm [77] and the increase of quantitative data available for the training set provide the possibility to switch from SAR to QSAR analysis.

Certainly, for a more detailed analysis of the multitargeted action of compounds and their interactions, it is necessary to apply the computational methods simulating the behavior of regulatory pathways under treatment by individual pharmaceutical agents and their combinations. Such methods are currently developed by our team [78] and by some other researchers [79, 80].

ABBREVIATIONS

ATP = Adenosine triphosphate
DNP = Dictionary of natural products
HIV = Human immunodeficiency virus
IAP = Invariant accuracy of prediction
LOO CV = Leave one out cross-validation
MDDR = MDL drug data report
MNA = Multilevel neighborhoods of atoms
NCE = New chemical entities
NCI = National Cancer Institute
PASS = Prediction of Biological Activity Spectra
PKS = Polyketide synthase
QSAR = Quantitative structure-activity relationships
SDF = Structure-data file
SJW = St John's Wort
SAR = Structure-activity relationships
SSFN = Substructure superposition fragment notation

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publishers Web site along with the published article.

REFERENCES

[1] Samuelsson G. Drugs of natural origin: a textbook of pharmacognosy. 5th ed. Stockholm: Swedish Pharmaceutical Press 2004.
[2] Kong DX, Li XJ, Zhang HY. Where is the hope for drug discovery? Let history tell the future. Drug Discov Today 2009; 14: 115-9.
[3] Vaidya ADB, Devasagayam TPA. Current status of herbal drugs in India: an overview. J Clin Bioch Nutr 2007; 41: 1-11.
[4] Domarew CA, Holt RR, Snitkoff GG. A study of Russian phytomedicine and commonly used herbal remedies. J Herbal Pharmacother 2002; 2: 31-48.
[5] Ertl P, Roggo S, Schuffenhauer A. Natural product-likeness score and its application for prioritization of compound libraries. J Chem Inf Model 2008; 48: 68-74.
[6] Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod 2007; 70: 461-477.
[7] Harley AL. Natural products in drug discovery. Drug Discov Today 2008; 13: 894-901.
[8] Morphy R, Rankovic Z. Designing multiple ligands - medicinal chemistry strategies and challenges. Cur Pharm Des 2009; 15: 586-600.
[9] Rollinger JM, Langer T, Stuppner H. Strategies for efficient lead structure discovery from natural products. Curr Med Chem 2006; 13: 1491-507.
[10] Rollinger JM, Stuppner H, Langer T. Virtual screening for the discovery of bioactive natural products. In: Petersen F, Amstutz R, Eds. Progress in Drug Research. Series: Natural compounds as drugs. Vol. 1. Basel, Switzerland: Birkhäuser Verlag 2008; Vol. 65: pp. 212-249.
[11] Ehrman TM, Barlow DJ, Hylands PJ. Phytochemical databases of Chinese herbal constituents and bioactive plant compounds with known target specificities. J Chem Inf Model 2007; 47: 254-63.
[12] Rollinger JM. Accessing target information by virtual parallel screening - The impact on natural product research. Phytochem Lett 2009; 2: 53-8.
[13] Burov YuV, Poroikov VV, Korolchenko LV. National system for registration and biological testing of chemical compounds: facilities for new drugs search. Bull Natl Cent Biol Active Compds (Rus) 1990; 1: 4-25.

[14] Avidon VV. Criteria for the comparison of chemical structures and principles of construction of an information language for a logical information system for biologically active compounds. Pharm-Chem J (Rus) 1974; 8: 22-5.
[15] Golender VE, Rozenblit AE. Computer Methods for Drug Design. Riga: Zinatne 1978.
[16] Avidon VV, Arolovich VS, Kozlova SP, Piruzian LA. Statistical study of information file on biologically active compounds. II. Choice of decision rule for biological activity prediction. Pharm-Chem J (Rus) 1978; 12: 88-93.
[17] Avidon VV, Arolovich VS, Kozlova SP, Piruzian LA. Statistical investigation of large volumes of data with respect to the biological activity of compounds III. Selection of a determinant for predicting biological activity. Pharm-Chem J (Rus) 1978; 12: 99-106.
[18] Avidon VV, Arolovich VS, Blinova VG, Freidina AM. Statistical investigation of the data file on biologically active compounds. V. Allowance for the novelty of the chemical structure in the prediction of the biological activity by an improved method of substructural analysis. Pharm-Chem J (Rus) 1983; 17: 59-62.
[19] Golender VE, Rosenblit AB. Logical and Combinatorial Algorithms for Drug Design, UK: Research Studies Press, Wiley & Sons 1983.
[20] Poroikov VV, Filimonov DA, Boudunova AP. Comparison of the results of prediction of the spectra of biological activity of chemical compounds by experts and the PASS system. Automat Doc Math Linguist 1993; 27: 40-3.
[21] Filimonov DA, Poroikov VV, Karaicheva EI, Kazarian RK, Budunova AP, Mikhailovskii EM, *et al*. Computer-aided prediction of biological activity spectra of chemical substances on the basis of their structural formulae: computerized system PASS. Exp Clin Pharmacol (Rus) 1995; 58: 56-62.
[22] Filimonov DA, Poroikov VV. PASS: computerized prediction of biological activity spectra for chemical substances. In: Bioactive Compound Design: Possibilities for Industrial Use. Oxford, UK: BIOS Scientific Publishers 1996; pp. 47-56.
[23] Lagunin A, Stepanchikova A, Filimonov D, Poroikov V. PASS: prediction of activity spectra for biologically active substances. Bioinformatics 2000; 16: 747-8.
[24] Stepanchikova AV, Lagunin AA, Filimonov DA, Poroikov VV. Prediction of biological activity spectra for substances: Evaluation on the diverse set of drugs-like structures. Cur Med Chem 2003; 10: 225-33.
[25] Poroikov V, Filimonov D. PASS: Prediction of Biological Activity Spectra for Substances. In: Predictive Toxicology. Helma C, Ed. Boca Raton: Taylor & Francis 2005; pp. 459-478.
[26] Filimonov DA, Poroikov VV. Probabilistic approach in activity prediction. In: Chemoinformatics Approaches to Virtual Screening. Varnek A, Tropsha A, Eds. Cambridge, UK: RSC Publishing 2008; pp. 182-216.
[27] Lewi PJ. Spectral mapping, a technique for classifying biological activity profiles of chemical compounds. Arzneimittelforschung 1976; 26: 1295-300.
[28] Battistini A, Affabris E, Fiorucci G, Coccia EM, Romeo G, Marziali G, *et al*. Spectrum of biological activity of interferons. Annali dell'Istituto Superiore di Sanità 1990; 26: 227-53.
[29] Gringorten JL, Sohi SS, Masson L. Activity spectra of Bacillus thuringiensis delta-endotoxins against eight insect cell lines. In Vitro Cell Dev Biol Anim 1999; 35: 299-303.
[30] Fliri AF, Loging WT, Thadeio PF, Volkmann RA. Biological spectra analysis: linking biological activity profiles to molecular structure. Proc Natl Acad Sci USA 2005; 102: 261-6.
[31] Rana A. Benzothiazoles: a new profile of biological activities. Indian J Pharm Sci 2007; 69: 10-7.
[32] Anzali S, Barnickel G, Cezanne B, Krug M, Filimonov D, Poroikov V. Discriminating between drugs and nondrugs by Prediction of Activity Spectra for Substances (PASS). J Med Chem 2001; 44: 2432-7.
[33] Gasteiger J. Handbooks of Cheminformatics: From Data to Knowledge. Weinheim: Wiley-VCH 2003.
[34] Avidon VV, Pomerantsev IA, Rozenblit AB, Golender VE. Structure-activity relationship oriented languages for chemical structure representation. J Chem Inf Comput Sci 1982; 22: 207-14.
[35] Todeschini R, Consonni V. Handbook of Molecular Descriptors. Weinheim: Wiley-VCH 2000.
[36] Filimonov D, Poroikov V, Borodina Yu, Gloriozova T. Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. J Chem Inf Comput Sci 1999; 39: 666-70.
[37] Poroikov VV, Filimonov DA, Borodina YuV, Lagunin AA, Kos A. Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds. J Chem Inform Comput Sci 2000; 40: 1349-55.
[38] Poroikov VV, Filimonov DA, Ihlenfeldt WD, Gloriozova TA, Lagunin AA, Borodina YuV, *et al*. PASS biological activity spectrum predictions in the enhanced open NCI database browser. J Chem Inform Comput Sci 2003; 43: 228-36.
[39] Sadym A, Lagunin A, Filimonov D, Poroikov V. Prediction of biological activity spectra *via* Internet. SAR QSAR Environ Res 2003; 14: 339-47.
[40] Geronikaki A, Druzhilovsky D, Zakharov A, Poroikov V. Computer-aided predictions for medicinal chemistry *via* Internet. SAR QSAR Environ Res 2008; 19: 27-38.
[41] Di Giorgio C, Delmas F, Filloux N, Robin M, Seferian L, Azas N, *et al*. *In vitro* activities of 7-substituted 9-chloro and 9-amino-2-methoxyacridines and their bis- and tetra-acridine complexes against *Leishmania infantum*. Antimicrob Agents Chemother 2003; 47: 174-80.
[42] Delmas F, Di Giorgio C, Robin M, Azas N, Gasquet M, Detang C, *et al*. *In vitro* activities of position 2 substitution-bearing 6-nitro- and 6-aminobenzothiazoles and their corresponding anthranilic acid derivatives against *Leishmania infantum* and *Trichomonas vaginalis*. Antimicrob Agents and Chemother 2002; 46: 2588-94.
[43] Di Giorgio C, Delmas F, Ollivier E, Elias R, Balansard G, Timon-David P. *In vitro* activity of the beta-carboline alkaloids harmane, harmine, and harmaline toward parasites of the species *Leishmania infantum*. Exp Parasitol 2004; 106: 67-74.
[44] Goel RK, Kumar V, Mahajan MP. Quinazolines revisited: search for novel anxiolytic and GABAergic agents. Bioorg Med Chem Lett 2005; 15: 2145-8.
[45] Geronikaki A, Babaev E, Dearden J, Dehaen W, Filimonov D, Galaeva I, *et al*. Design of new anxiolytics: from computer prediction to synthesis and biological evaluation. Bioorg Med Chem 2004; 12: 6559-68.
[46] Labanauskas L, Brukstus A, Udrenaite E, Bucinskaite V, Susvilo I, Urbelis G. Synthesis and anti-inflammatory activity of 1-acylaminoalkyl-3,4-dialkoxybenzene derivatives. Il Farmaco 2005; 60: 203-7.
[47] Dolzhenko AV, Kolotova NV, Koz'minykh VO, Vasilyuk MV, Kotegov VP, Novoselova GN, *et al*. Substituted amides and hydrazides of dicarboxylic acids. Part 14. Synthesis and antimicrobial and antiinflammatory activity of 4-antipyrylamides, 2-thiazolylamides, and 1-triazolylamides of some dicarboxylic acids. Pharm Chem J 2003; 37: 149-51.
[48] Lagunin AA, Gomazkov OA, Filimonov DA, Gureeva TA, Dilakyan EA, Kugaevskaya EV, *et al*. Computer-aided selection of potential antihypertensive compounds with dual mechanisms of action. J Med Chem 2003; 46: 3326-32.
[49] Geronikaki AA, Lagunin AA, Hadjipavlou-Litina DI, Elefteriou PT, Filimonov DA, Poroikov VV, *et al*. Computer-aided discovery of anti-inflammatory thiazolidinones with dual cyclooxygenase/lipoxygenase inhibition. J Med Chem 2008; 51: 1601-9.
[50] Poroikov V, Lagunin A, Filimonov D. Pharmaexpert: diseases, targets and ligands - three in one. Proceedings of the 15th European symposium on structure-activity relationships (QSAR) and molecular modeling. Istanbul, Turkey, September 05-10, 2004. Sener EA, Yalcin I, Eds. Ankara (Turkey): CADD & D Society 2005; pp. 514-515.
[51] Dembitsky VM, Gloriozova TA, Poroikov VV. Novel antitumor agents: marine sponge alkaloids, their synthetic analogues and derivatives. Mini-Rev Med Chem 2005; 5: 319-36.
[52] Newman DJ. Natural products as leads to potential drugs: an old process or the new hope for drug discovery? J Med Chem 2008; 51: 2589-99.
[53] Flekhter OB, Karachurina LT, Poroikov VV, Nigmatullina LR, Baltina LA, Zarudii FC, *et al*. Synthesis of ether triterpenoides of lupan group and their hepatoprotective activity. Rus J Bioorg Chem 2000; 26: 215-23.
[54] Rollinger JM, Schuster D, Danzl B, Schwaiger S, Markt P, Schmidtke M, *et al*. *In silico* target fishing for rationalized ligand discovery exemplified on constituents of *Ruta graveolens*. Planta Med 2009; 75: 195-204.

[55] Omura S. Macrolide antibiotics: chemistry, biology, and practice. 2nd ed. New York: Academic Press 2002.

[56] Weissman KJ. Polyketide biosynthesis: understanding and exploiting modularity. Phil Trans R Soc Lond 2004; 362: 2671-90.

[57] Zotchev SB, Stepanchikova AV, Sergeyko AP, Sobolev BN, Filimonov DA, Poroikov VV. Rational design of macrolides by virtual screening of combinatorial libraries generated through *in silico* manipulation of polyketide synthases. J Med Chem 2006; 49: 2077-87.

[58] Sergeyko AP, Stepanchikova AV, Sobolev BN, Zotchev SB, Filimonov DA, Lagunin AA, *et al.* Computer-aided design of polyketides with the required properties. Biomed Chem (Rus) 2007; 53: 522-31.

[59] Buckingham J. Dictionary of Natural Products. UK: Chapman & Hall/CRC 2007.

[60] Aronson JK. Meyler's Side Effects of Drugs. 15th ed. London: Elsevier 2006.

[61] Kostrubsky VE, Strom SC, Hanson J, Urda E, Rose K, Burliegh J, *et al.* Evaluation of hepatotoxic potential of drugs by inhibition of bile-acid transport in cultured primary human hepatocytes and intact rats. Toxicol Sci 2003; 76: 220-8.

[62] Di Carlo G, Borrelli F, Ernst E, Izzo AA. St John's wort: Prozac from the plant kingdom. Trends Pharmacol Sci 2001; 22: 292-7.

[63] Erdelmeier CAJ, Koch E, Hoerr R. *Hypericum perforatum* - St. John's Wort chemical, pharmacological and clinical aspects. Stud Nat Prod Chem 2000; 22: 643-716.

[64] Kurkin VA. Pharmacognosy. Samara, Russia: SSU 2007.

[65] Smelcerovic A, Spiteller M, Ligon AP, Smelcerovic Z, Raabe N. Essential oil composition of *Hypericum L.* species from Southeastern Serbia and their chemotaxonomy. Biochem Syst Ecol 2007; 35: 99-113.

[66] Radusiene J, Judzentiene A, Bernotiene G. Essential oil composition and variability of *Hypericum perforatum L.* growing in Lithuania. Biochem Syst Ecol 2005; 33: 113-24.

[67] Poroikov VV, Filimonov DA. How to acquire new biological activities in old compounds by computer prediction. J Comput-Aid Mol Des 2002; 16: 819-24.

[68] Pogrebnyak AV, Poroikov VV, Staryh VV, Konovalov DA. Computer assisted analysis of antitumor activity of sesquiterpene lactones from the family of Asteraceae. Plant Res (Rus) 1998; 34: 61-4.

[69] Dembitsky VM, Levitsky DO, Gloriozova TA, Poroikov VV. Acetylenic aquatic anticancer agents and related compounds. Nat Prod Commun 2006; 1: 773-812.

[70] Sergeiko AP, Poroikov VV, Hanus LO, Dembitsky VM. Cyclobutane-containing alkaloids: origin, synthesis, and biological activities. Open Med Chem J 2008, 2: 26-37.

[71] Dembitsky VM, Gloriozova TA, Poroikov VV. Natural peroxy antincancer agents. Mini-Rev Med Chem 2007; 7: 571-89.

[72] Devillers J, Dore JC, Guyot M, Poroikov V, Gloriozova T, Lagunin A, *et al.* Prediction of biological activity profiles of cyanobacterial secondary metabolites. SAR QSAR Environ Res 2007; 18: 629-43.

[73] Goel RK, Lagunin AA, Poroikov VV. PASS - assisted exploration of new therapeutic potential of natural products. Med Chem Res 2010; in press.

[74] Phillipson JD. Phytochemistry and pharmacognosy. Phytochemistry 2007; 68: 2960-72.

[75] Sardari S, Shokrgozar MA, Ghavami G. Cheminformatics based selection and cytotoxic effects of herbal extracts. Toxicol in Vitro 2009; 23: 1412-21.

[76] Pors K, Goldberg FW, Leamon CP, Rigby AC, Snyder SA, Falconer RA. The changing landscape of cancer drug discovery: a challenge to the medicinal chemist of tomorrow. Drug Discov Today 2009; 14: 1045-50.

[77] Filimonov DA, Zakharov AV, Lagunin AA, Poroikov VV. QNA based "Star Track" QSAR approach. SAR QSAR Environ Res 2009; 20: 679-709.

[78] Koborova ON, Filimonov DA, Zakharov AV, Lagunin AA, Ivanov SM, Kel A, *et al.* In silico method for identification of promising anticancer drug targets. SAR QSAR Environ Res 2009; 20: 755-66.

[79] Kel A, Voss N, Valeev T, Stegmaier P, Kel-Margoulis O, Wingender E. ExPlain: finding upstream drug targets in disease gene regulatory networks. SAR QSAR Environ Res 2008; 19: 481-94.

[80] Lopez A, Parsons AB, Nislow C, Giaever G, Boone C. Chemical-genetic approaches for exploring the mode of action of natural products. Prog Drug Res 2008; 66(237): 239-71.

Received: January 22, 2010 **Accepted: February 2, 2010**