Chemistry of Heterocyclic Compounds, Vol. 50, No. 3, June, 2014 (Russian Original Vol. 50, No. 3, March, 2014)

PREDICTION OF THE BIOLOGICAL ACTIVITY SPECTRA OF ORGANIC COMPOUNDS USING THE PASS ONLINE WEB RESOURCE

D. A. Filimonov[1], A. A. Lagunin[1], T. A. Gloriozova[1], A. V. Rudik[1], D. S. Druzhilovskii[1], P. V. Pogodin[1,2], and V. V. Poroikov[1,2]*

The freely accessible web resource PASS Online is presented. This resource is designed for the prediction of the biological activity spectra of organic compounds based on their structural formulas for more than 4000 types of biological activity with average accuracy above 95% (http://www.way2drug.com/passonline). The prediction is based on an analysis of the structure-activity relationships in the training set containing information on the structure and biological activity of more than 300000 organic compounds. The possibilities and limitations of this approach are described. Recommendations are given for interpreting the prediction results. Examples are given for the practical use of the PASS Online web resource in order to establish priorities for chemical synthesis and biological testing of substances on the basis of prediction results. The further trends are considered for the using PASS Online as an Internet platform for joint projects of academic researchers for the search and development of new pharmaceutical agents.

Keywords: PASS Online, biological activity, computer prediction, organic compounds, pharmacologically active compounds, web resource.

The development of new and safer drugs is based on the use of existing physical and information resources. More than 60 million organic compounds have been synthesized and are available today as samples for biological activity testing [1]. The library of molecules generated *in silico* consisting of 17 or less non-hydrogen atoms contains 166.4 billion structural formulas [2], which provides for virtually inexhaustible possibilities for the development of methods for the synthesis of new pharmacologically active compounds. There are about 2000 molecular targets for drugs [3] and two million potential pharmacological targets taking account of post-translational modifications, alternative splicing, protein-protein interactions, and various ligand-protein binding sites. The combinatorics of ligand-protein interactions has grown repeatedly from the "one drug–one target" approach to the concept of pleiotropic substances [4]. The search for promising new pharmacologically active compounds involves multiple criteria since in addition to possessing a desired specific

*To whom correspondence should be addressed, e-mail: vladimir.poroikov@ibmc.msk.ru.

[1] V. N. Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, 10/8 Pogodinskaya St., 119121 Moscow, Russia.
[2] N. I. Pirogov Russian National Research Medical University, 1 Ostrovityanova St., 117997 Moscow, Russia.

Translated from Khimiya Geterotsiklicheskikh Soedinenii, No. 3, pp. 483-499, March, 2014. Original article submitted February 7, 2014.

activity, a compound must have a minimum of side effects and toxic properties along with suitable pharmacological characteristics [5]. The experimental testing of dozens of millions of organic compounds for thousands biological activities is obviously unachievable, mandating the need for computer methods for the search and optimization of new pharmacologically active compounds [6]. The presently used computer approaches for drug discovery are divided into structure-based drug design and ligand-based drug design [7].

Despite advances in methods based on the structure of the target, their use for the development of new drugs is limited due to 1) the need for information on the three-dimensional structure of the target macromolecule, which is far from available for all proteins holding interest as pharmacological targets, 2) the dynamics of change in the three-dimensional structure of proteins during function and, especially, upon the formation of protein-protein and protein-ligand complexes, 3) the difficulty of determining the biologically active conformation for conformationally flexible ligands, and 4) the lack of clarity in selecting estimation function of the binding energy (*scoring function*). Thus, methods based on the structure of the ligands have been much more often used for evaluating the biological activity profiles or spectra characterizing the complex action of organic compounds on biological systems. In contrast to classical approaches to the analysis of quantitative structure-activity relationships (QSAR) [8], when QSAR models are constructed and used for prediction relative to compounds belonging to the same chemical class, the evaluation of biological activity spectra is carried out by analyzing the structure-activity relationships for heterogeneous sets using classification methods. In this case, compounds tested under different but, in the experts' opinion, equivalent experimental conditions are designated "active" relative to specific types of biological activity according the expert estimates.

Dozens of examples of the development and application of QSAR models for heterogeneous compound sets are given in the monograph by Franke [9]. The studies cited in this book, however, were not designed for a complex evaluation of biological activity profiles and were oriented primarily for the supramolecular level of biological activity types such as antibacterial, antifungal, antimalarial, cytotoxic, spasmolytic, antihypertensive, analgesic, etc. activities, which reflects the predominant methods for the experimental testing for biological activity in that period.

The first attempts to develop methods for the computer prediction of biological activity based on the structure of compounds belonging to different chemical classes in order to evaluate biological activity profiles were carried out by Avidon and his colleagues [10] in the 1970's based on the analysis of similarity [11, 12] and a substructural approach [13]. The possibilities of predicting biological activity by the analysis of similarity were studied using a database containing 7200 biologically active compounds [14]. This approach corresponded to the traditional logic of chemists and pharmacologists since it is based on the major hypothesis of a structure-activity relationship, namely, that similar structures display similar activity. In this case, similarity was evaluated relative to the resemblance of two compounds in the descriptor space of a Substructure Superposition Fragmental Notation (SSFN). As noted in the cited work, the prediction of biological activity using similarity analysis permits the "assignment" of types of biological activity established for standard structures but differing from the canonical list entailing 58 types of activity to the compounds found [14]. In order to save computing time, a subset of 711 standard structures was selected from the 7200 compounds of the initial data set for this study. In the analysis of a test set of 500 chemical compounds, the authors found that, out of 158 errors of the first kind (the assignment of active compounds as inactive) during substructure analysis, 55 cases (using all 7200 structures of the initial set) and 21 cases (for the subset of standards) were correctly recognized on the basis of the similarity analysis. The authors concluded that a complex approach combining the results for predicting active compounds by substructural analysis and similarity analysis would be useful, which is especially important in the case of compounds from new chemical classes provided that a single analog exists among the standard structure set [14].

At the same time, Golender and Rozenblit [15-18] proposed a logic-structural approach based on image recognition theory for the classification of chemical compounds relative to their biological activity. Some of the work of Golender and Rozenblit was carried out in collaboration with Avidon [19, 20]. Golender and Rozenblit were the first to carry out automatic encoding of the structures of chemical compounds using SSFN descriptors

[10, 11, 21]. The logic-structural approach was realized in the ORAKUL computer program package, which was used for experiments with a data bank of biologically active compounds, containing 5964 chemical compounds with information on 55 types of biological activity [19]. The accuracy for the recognition of active compounds was 91% in leave-one-out cross-validation procedure and 80% for an independent test set [17]. It should be noted that although the authors properly noted the disadvantages of the SSFN language and proposed alternative variants for descriptor languages [17, 20], they also used structure descriptors analogous to the SSFN language in their subsequent OREX system (Optimized Recognizing Expert System) [22].

The work of Avidon, Rozenblit, Golender and their colleagues provided the basis for the large-scale computer prediction of various type of biological activity and the selection of the most promising compounds among those in the USSR State Registry System for experimental testing [23-25]. At the end of the 1980's, we began work on the development of a new computer program for the prediction of the biological activity spectra of organic compounds, which later was called PASS (Prediction of Activity Spectra for Substances) [26]. In the course of this work, we studied various methods for describing chemical structure and the construction of structure-activity relationships [27-33]. In subsequent work [32, 33], we carried out a detailed description of the PASS program algorithm, validation of the developed approach in various computer experiments, and gave examples for the practical application of the PASS program in the search for new pharmacologically active compounds. Since refinement of the PASS program is carried out permanently, we shall briefly examine the current state of this project and present practical recommendations for using the prediction results. A more detailed description of the PASS procedure is given in Supplement 1 (Supplementary information).

COMPUTER PREDICTION OF BIOLOGICAL ACTIVITY SPECTRA USING THE PASS PROGRAM

The PASS computer program [32-34] allows to estimate the probable profile of biological activity of a drug-like organic compound (whose molecular mass ranges from 50 to 1250 Da) based on its structural formula. The estimation is based on an analysis of the structure-activity relationships for a broad training set entailing drug substances, drug-candidates in various stages of clinical and preclinical investigation, pharmaceutical agents and chemical probes, and compounds, for which specific toxicity information is known.

The categorical description of biological activity as "active" or "inactive" is used in the PASS program. The chemical structure is represented by the set of descriptors of multilevel neighborhoods of atoms, MNA [35]. The algorithm for constructing models of the structure-activity relationship based on the compounds from the training set and prediction of activity for new compounds is based on the Bayesian estimates [32, 33].

The average prediction accuracy calculated by the leave-one-out cross-validation procedure for the whole training set and for all represented in it types of biological activity is about 95%. In the leave-one-out cross-validation procedure information on structure and the known types of activity is exluded for each compound sequentially, and the prediction based on the remaining part of the training set is then carried out for it. The procedure is repeated iteratively for all the compounds of the training set and the fraction of properly evaluated types of biological activity is calculated.

Since the training set, in general, cannot contain complete information on the biological activity of the compounds included in it (none of the chemical compounds has been studied for all the possible types of biological activity), we carried out a special study to determine the effect of the incompleteness of the information in the training set on the quality of the prediction [36]. This study was carried out based on a set consisting of about 19000 compounds from the MDDR (MDL Drug Data Report) database (the principal compounds, for which experimental biological activity data are given). A total of 120 different types of activity were represented in this set by not less than 100 compounds. During the course of the computer experiments, the entire set was randomly divided 50 times into two approximately equal subsets, one of which was used as the training set and the other was used as the test set and *vice versa*. Thus, a total of 100 experiments were carried out. In order to model the information incompleteness, 20, 40, 60, and 80% of the information on

structure or biological activity was randomly excluded from the training set. The mean prediction error value was calculated during the study. Exclusion of up to 60% of the information still does not prevent reasonable estimates of biological activity for the test set compounds, i.e., the PASS program algorithm has statistical robustness despite the incompleteness of the information in the training set.

New pharmaceutical agents were found possessing antiulcer [37], antioxidant [38], antimetastatic [39], nootropic [40], anti-inflammatory [41], and other action (see below). Feasibility was demonstrated for use of the PASS program for detecting antihypertensive [42] and anti-inflammatory compounds [43] with multitarget action, establishing new types of biological action for drugs launched for medical use [44, 45], and evaluating the mechanisms of side effects and toxicity of pharmaceuticals [46, 47].

Data on molecular structural formulas given in the Molfile format (Accelrys Inc., http://accelrys.com) (for a single structure) or in the SDfile format (Accelrys Inc., http://accelrys.com) for a structure set [48] are used as the input information into the PASS program.

The PASS user obtains an output information as a list of predicted types of activity with the estimated probability for each type of activity "to be active" *Pa* and "to be inactive" *Pi*, which vary from zero to one. The probabilities *Pa* and *Pi* also indicate the estimated probabilities of first- and second-kind errors, respectively. These terms may be seen as indices of the belonging of the predicted compound to fuzzy classes of active and inactive compounds (Fig. 1). Widely differing criteria for analyzing the prediction results, corresponding to the solution of specific practical problems, may be established using these probabilities.

The chance of experimentally finding a given activity increases with increasing value of *Pa* and decreasing value of *Pi*. We shall subsequently examine a situation, in which the *Pa* value is rather high and much greater than the *Pi* value. If in analyzing a predicted activity spectrum, those types of activity are selected, for which *Pa* > 90%, we risk missing about 90% of actually active compounds but the probability of false-positive predictions is insignificant. For *Pa* > 80%, we already miss only 80% of the active compounds but the probability of false-positive predictions will be higher, etc. Finally, for *Pa* = *Pi*, the probabilities of false-positive and false-negative errors will be equal. This threshold is used as the default for cut-off of the probable types of activity in the prediction (Fig. 1).

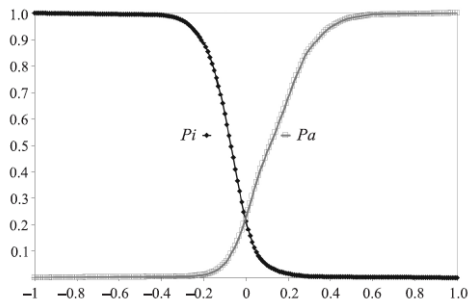


Fig. 1. Evaluation of values of *Pa* and *Pi* (along the ordinate) for antitumor activity as a function of a primary Bayesian estimates (*B* statistics [32, 33], along the abscissa).

We should note that the probability *Pa* primarily reflects the similarity of the structure of a given molecule to the structures of molecules of the most typical active compounds in the corresponding subset of the training set. Thus, as a rule, there is no direct correlation of the values of *Pa* with quantitative activity characteristics. A really active molecule possessing molecular structure atypical for the training set may have a low *Pa* value in the prediction, perhaps even *Pa* < *Pi*. This is obvious from the method of constructing the functions *Pa*(*B*) and *Pi*(*B*): the values of *Pa* for active compounds and of *Pi* for inactive compounds from the training set are distributed with strict homogeneity.

Another important aspect of interpreting the prediction results is related to novelty of the analyzed compound. If we limit ourselves only to activity types predicted with the highest values of *Pa*, the compounds selected by the prediction may prove to be analogs of known pharmacological agents. For example, when $Pa > 0.7$, the chances of finding experimental activity are rather high but the compounds found may be close structural analogs of known drugs. If we select in the range $0.5 < Pa < 0.7$, the chances for detecting experimental activity will be lower but the compounds will be less similar to known pharmaceutical agents. For $Pi < Pa < 0.5$, the chances of detecting experimental activity will be even lower but if the prediction is confirmed, the compound found may prove a parent compound for a new chemical class for the biological activity examined.

If the predicted activity spectrum is broad, the molecular structure of this compound is rather simple and does not consist of any features providing for high selectivity of its biological action.

In some cases, two mutually-exclusive activities are simultaneously predicted for a compound, for example, activities as an agonist and antagonist, stimulator and blocker, and activator and inhibitor of a specific receptor. In this case, the PASS program prediction does not permit differentiation of possible "internal" activities of the compound and only indicates likely affinity relative to a given receptor, enzyme, or transporter.

The PASS program does not permit us to evaluate the effect of stereoisomerism on the display of biological activity since it is impossible to create a representative training set taking account of the three-dimensional structure of the included compounds.

Of course, one must take into account the actual possibilities of experimental testing in analyzing the biological activity spectra predicted by the PASS program. In this case, a general recommendation is to carry out a consecutive study of the different predicted types of biological activity from the most probable to least probable.

We should stress that the PASS program cannot predict whether a specific compound will become a drug since this depends on a series of factors. However, the prediction can help to determine which biological activities should be initially tested and which compounds have the highest probability for displaying the required types of activity.

PASS ONLINE WEB RESOURCE

Since 2000, PASS Online has been functioning as a freely accessible web resource for the prediction of biological activity spectra [49-53]. The computer prediction of biological activity may be carried out both for published and new compounds, which permits filtering out unpromising compounds at the very earliest stages of an investigation.

The user of PASS Online may obtain a prediction either by selecting the structural formula of an organic compound holding interest as a file in the Molfile format (Accelrys Inc., http://accelrys.com) [48] or SMILES code [54] or by directly entering the structural formula in the web browser using the *Marvin applet* [55]. The pharmacological effects and mechanisms of action, for which the calculated probabilities for the display of activity exceeds the probability threshold ($Pa > Pi$, by default) are found in the default list of predicted activity types. Depending on the specifics of the problem to be solved, the user may limit a given list only to those activity types, for which $Pa > 0.3$ or $Pa > 0.7$.

When necessary, the user may also obtain a prediction for side effects of a compound. In interpreting the results of a prediction, the user must take account that the major pharmacological effects and mechanisms of action reflect the interaction of the organic compound with a biological object (a test system or patient's body) observed in 100% (or at least the overwhelming number of) cases, while evaluation of a side effect is often based on clinical observations related to a small number of patients or even a single individual taking the specific pharmaceutical agent. Hence, computer-aided predictions of the major pharmacological action and side effects of an organic compound are not equivalent.

PASS Online provides predictions for over 4000 types of biological activity with mean accuracy of 95%, which is much higher than for other web resources that also predict biological activity profiles using the structural formulas of organic compounds [59], in particular, ChemSpider [56], SuperPred [57], and DRAR-CPI [58].

The number of registered users for PASS Online is now more than 10000 from 90 countries throughout the world and predictions have been obtained for more than 300000 molecules. The usage statistics for PASS Online in 2013 is given in Figure 2.

A total of 49727 predictions were obtained for the PASS Online web resource in 2013. India has the leading position for number of prediction (18379), followed by Russia (13947), Ukraine (4332), Mexico (2880), Poland (1333), and China (1124). PASS Online was also used by researchers in industrially-developed countries: Canada (303 predictions), USA (284), Austria (91), Great Britain (76), Japan (70), Italy (61), and France (54). The much smaller number of compounds sent for prediction by researchers in industrially-developed countries may be attributed to the more stringent requirements for confidentiality of information on the structure of synthesized molecules.

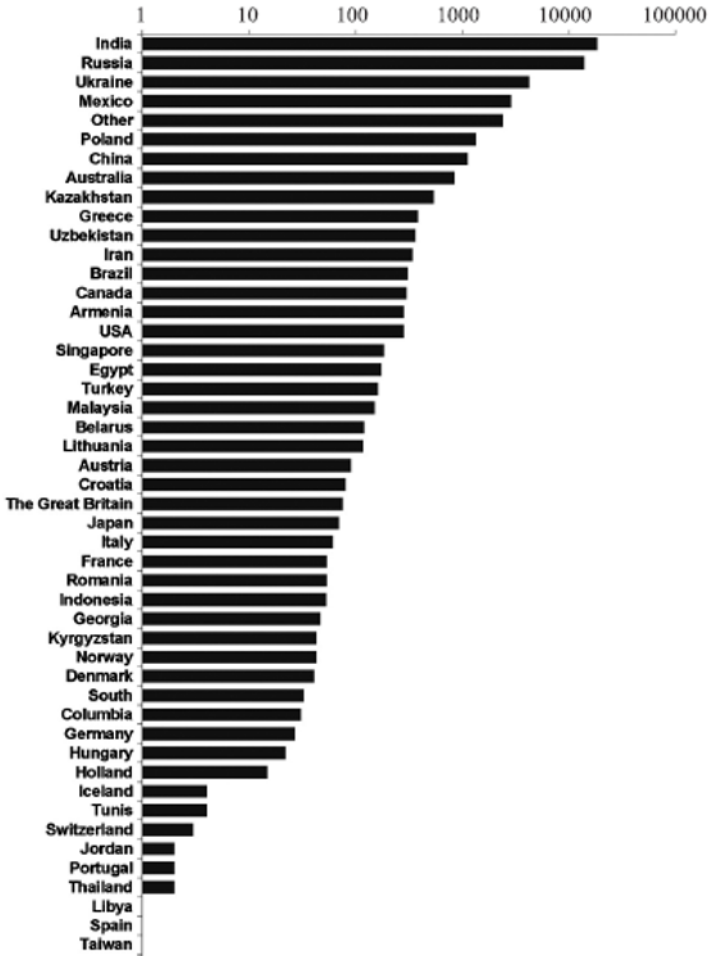


Fig. 2. Usage statistics for the PASS Online web resource for users in various countries in 2013 relative to number of predictions.

At the end of 2013, we carried out a selective survey of PASS Online users, which showed that the major goals for using computer predictions of biological activity spectra are as follows: planning biological testing (124), planning chemical synthesis (86), discovering new types of activity for known compounds (83), and evaluation of safety and usage risks for chemical compounds (35).

Independent (of the PASS program authors) researchers have published more than 150 articles on the results of prediction of biological activity of compounds studied in their laboratories. In more than half of the publications (82 of 152 articles), the prediction results were confirmed by biological testing. In the other studies, the predictions were used to indicate promise for the synthesis of specific compounds. A review of studies published up to 2008 was given in the journal, SAR QSAR Environmental Research [52]. Let us examine some typical recently published examples of the use of the PASS program. A list of publications, in which the results of biological activity prediction were obtained with the PASS program, is given in Supplement 2 (Supplementary information for the article). We should note that these examples were obtained using various versions of the PASS program since work is constantly underway to enlarge the scope and improve the quality of the training set and to develop more accurate methods for analyzing structure-activity relationship.

EXAMPLES FOR THE PRACTICAL APPLICATION OF BIOLOGICAL ACTIVITY SPECTRUM PREDICTIONS OBTAINED USING THE PASS PROGRAM

Screening the library of synthesized compounds. Virtual screening of the library led to the selection of 32 molecules (out of 2648 synthesized compounds), for which the PASS program predicted xanthine oxidase inhibition with $Pa > 0.3$ [60]. The possibility of use in the treatment of gout was also indicated for 21 of these molecules. We should stress that the aim of B-Rao et al. [60] was to obtain a new class of xanthine oxidase inhibitors based on the original scaffold, which might avoid side effects observed for presently available drugs. This accounts for the relatively low estimates of probability of this activity *Pa* obtained in the prediction. In addition to prediction by the PASS program, they carried out docking of the selected molecules to the three-dimensional structure of the xanthine oxidase complex with piraxostat (the structure PDB code 1VDV), which yielded evaluations of the affinity for binding of ligands to the target. B-Rao et al. [60] presented results of the express testing *in vitro* for 24 of the 32 selected compounds: only three compounds inhibited xanthine oxidase at a concentration of 100 μM (compounds **1**-**3**, Fig. 3). The extent of inhibition of xanthine oxidase for the most active structure **3** was 98.3%.

In subsequent study, IC_{50} values were found for these compounds: **1** (9.4), **2** (30.2), and **3** (1.4 μM). Relative to the IC_{50} value, pyrimidine **3** is more active than the reference drug allopurinol (5.7 μM). B-Rao et al. [60] also presented the results of a study of reduction of the uric acid level upon the introduction of pyrimidines **1** and **2** in a hyperuremic rat model induced by the intraperitoneal introduction of potassium oxonate. The reduction in the uric acid level upon the intraperitoneal introduction of pyrimidines **1** and **2** was 84% (dose 100 mg/kg pyrimidine **1**) and 57% (dose 50 mg/kg pyrimidine **2**). This is considerably worse than the corresponding indices for allopurinol (184% for dose 10 mg/kg *per os*).

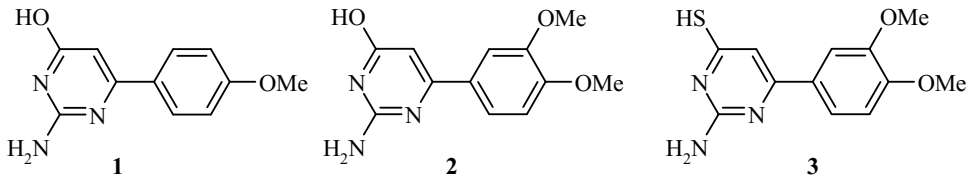
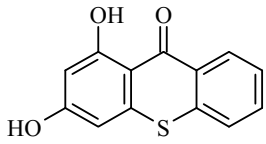


Fig. 3. Xanthine oxidase inhibitors discovered by virtual screening of a library of 2648 synthesized compounds. The PASS prediction for xanthine oxidase inhibitors: *Pa* = 0.314, 0.307, and 0.427 for pyrimidines **1**-**3**, respectively.

TABLE 1. Predicted Activity Spectrum for 1,3-Dihydroxyxanthen-9-one (*Pa* > 0.5)



Predicted pharmacological effects		
Activity	*Pa*	*Pi*
Antineurotoxic	0.859	0.005
Insulin and insulin analogs	0.781	0.004
Fibrinolytics	0.776	0.009
Antihelmintic (Nematodes)	0.629	0.009
Postmenopausal disorders treatment	0.625	0.003
Antidote, cyanide	0.541	0.008
Keratolytic	0.536	0.005
Antihelmintic	0.512	0.005
Predicted mechanisms of action		
Activity	*Pa*	*Pi*
Thioredoxin inhibitor	0.705	0.009
Phosphatidylinositol 3-kinase gamma inhibitor	0.553	0.005
Quercetin 2,3-dioxygenase inhibitor	0.543	0.005
Thioredoxin disulfide reductase inhibitor	0.541	0.007
FMO3 substrate	0.540	0.008
CF transmembrane conductance regulator inhibitor	0.539	0.005
Sulfotransferase substrate	0.505	0.004
Estrogen beta receptor agonist	0.501	0.001
SULFT1A3 substrate	0.500	0.004

These authors note that the use of the PASS program permitted rapid identification of potentially active compounds. In this case, it proved possible to use information not only on prediction of the action sought but also of other types of activity, toxicity, and metabolism. On the other hand, according to B-Rao et al. [60], the use of docking gives more reliable results. This conclusion appears somewhat strange since good scoring function evaluations according to the docking data were obtained for 21 inactive compounds subjected to express testing and for seven specially synthesized analogs of pyrimidine **1** (Tables 1 and 2, respectively, in the work of B-Rao [60]) although IC_{50} values for these compounds exceeded 100 μM. On the whole, the use of the PASS prediction and docking results permitted the selection of 24 of 2648 compounds for experimental testing; 3 of the 24 compounds with the original scaffold proved active. This led to a reduction in the cost of investigation by two orders of magnitude.

Determination of priorities in *in vitro* biological testing. Verbanac et al. [61] developed efficient methods for the synthesis of natural compounds and their derivatives (xanthones and thioxanthones) and the study of their biological activity in *in vitro* experiments and on the basis of *in silico* calculation. The *in vitro* testing showed that 16 synthesized compounds do not possess potent antimicrobial activity relative to bacteria *S. aureus*, *S. pyogenes*, *M. catarrgalis*, *H. influenza*, and *E. coli* and also cytotoxicity relative to cell lines HepG2 hepatocytes and Jurkat T-lymphocytes. A PASS program prediction was carried out to determine the most likely types of biological activity. An example of a prediction for one of these compounds is presented in Table 1.

The authors note that neither antibacterial nor cytotoxic action is predicted for any of the 16 compounds analyzed, which corresponds to the negative *in vitro* testing results. Chlordecone reductase inhibition,

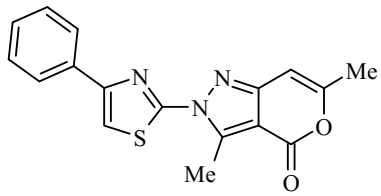
UDP-glucuronesyltransferase inhibition, membrane integrity agonist action, and membrane permeability inhibition were among the most probable types of activity predicted for all 16 compounds. Unfortunately, the authors did not experimentally check the most probable predicted activities.

Determination of priorities in *in vivo* biological testing. Kumar et al. [62] attempted to design and synthesize pyranopyrazole derivatives and carry out computer evaluation and experimental testing of the biological activity of these compounds. The prediction of the biological activity spectra obtained using the PASS computer program showed that most of the synthesized compounds are predicted to display analgesic and anti-inflammatory action. An example of the prediction for one of these compounds is given in Table 2.

The analgesic and anti-inflammatory activities of the synthesized compounds were tested on suitable experimental mouse models. Pain was elicited by the intraperitoneal introduction of acetic acid and an inflammatory response was obtained by the injection of carrageenan into a paw. The control was diclofenac. Experimental testing showed that some of the synthesized compounds display analgesic and anti-inflammatory effects at or above the level of the control. Since one of the possible action mechanisms predicted by the PASS program was cyclooxygenase inhibition, Kumar et al. [62] carried out docking of the compounds studied to the active site of cyclooxygenase II (COX-2, structure code in the Protein DataBase (PDB) 1CX2) and evaluated the binding energy. This analysis led the authors to conclude that the inhibitory properties of the compounds studied relative to COX-2 decrease relative to the nature of the substituent in the series: phenothiazolyl > benzothiazolyl > quinolinyl > pyrimidinyl > OMe > Br > Me > H. We should note, however, that these docking results also require experimental validation.

Evaluation of the multitarget action by prediction and *in vitro* profiling. Shcherbakova [63] designed and synthesized a series of tetrahydroindolo[2,3-*c*]quinolones, which are analogs of the well-known nootropic drug, ambocarb. The biological activity spectra of these compounds were evaluated. The molecular mechanisms for the nootropic action of ambocarb are still unknown. In this study, Shcherbakova attempted to deconvolute the structural formula of this drug into various fragments. These fragments determine the specific chemical classes, to which the compound belongs. Subsequently, prediction of the biological activity spectra was carried out for each fragment (Fig. 4).

TABLE 2. Predicted Activity Spectrum for 3,6-Dimethyl-2-(4-phenyl-1,3-thiazol-2-yl)pyrano[4,3-*c*]pyrazol-4(2*H*)-one



Activity	*Pa*	*Pi*
Analgesic, non-opioid	0.702	0.011
5-Hydroxytryptamine release inhibitor	0.681	0.005
Antineoplastic (Ovarian cancer)	0.677	0.025
Analgesic	0.621	0.022
Antiarthritic	0.606	0.021
Cognition disorders treatment	0.597	0.020
Anti-inflammatory	0.592	0.039
Antiviral (Arbovirus)	0.613	0.067
Complement factor D inhibitor	0.572	0.050
Immunomodulator	0.532	0.033
Immunosuppressant	0.454	0.044
Cyclooxygenase inhibitor	0.400	0.004
HCV IRES inhibitor	0.431	0.050

The prediction results gave the potential molecular targets and led to *in vitro* testing of ambocarb with these targets. The testing was carried out for a set of 227 pharmacological targets supplied by Cerep related to the action of the drug on the central nervous system [64] and for a set of 442 kinases from KINOMEscan® [65]. Ambocarb itself and its individual fragments were found to possess different complex interaction profiles with the molecular targets. Then, taking account of this analysis, new tetrahydroindolo[2,3-*c*]quinolones were designed and synthesized. The *in vivo* testing of these derivatives at National Institute of Neurological Disorders and Stroke in the United States showed that tetrahydroindolo[2,3-*c*]quinolones have a broad spectrum of neuroprotective, nootropic, antipsychotic, anticonvulsive, and analgesic action. These effects depend on the structure of the specific compound and dose.

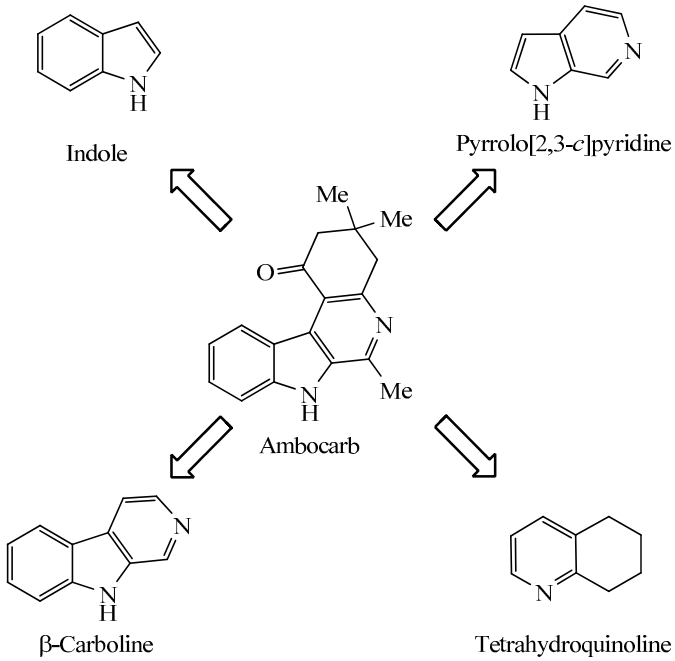


Fig. 4. Deconvolution of ambocarb structure into different fragments.

Unfortunately, Shcherbakova [63] did not give details on the use of computer prediction of the biological activity spectra for the design of new derivatives with multitarget action but, in itself, the approach based on the virtual deconvolution of structures with subsequent synthesis *in silico* appears extremely promising as illustrated in our recent studies [66].

Evaluation of the Biological Activity Spectra of Natural Products. The prediction of pharmacological properties for the phytocomponents from various plant sources has an important place in the applications of the computer prediction of biological activity [67]. This is related to the much greater structural variety of natural products than for synthetic compounds. Let us give, for example, the work of Frearson and Wyatt [68], in which the PASS computer program was used to evaluate the biological activity profile for rare steroid derivatives isolated from *Axinella carteri* marine algae. Properties such as nerve growth factor agonist action were predicted for six of eight of the analyzed compounds with probability $Pa > 0.7$. These authors did not carry out direct experimental testing of the PASS prediction results but literature data on such activity in molecular analogs were presented as indirect evidence for the correctness of the prediction. They concluded that it would be worthwhile to further investigate steroid derivatives from *Axinella carteri* marine algae as potential agents for the treatment of neurodegenerative disease.

Furthermore, actions useful in the treatment of female infertility, urological disorders, osteoporosis, psoriasis, and pruritus were also predicted for some of the derivatives studied with probability $Pa > 0.7$. In this

case, there was also no direct experimental testing of the prediction results required for a definitive conclusion concerning the possession of activity.

CONCLUSIONS

We have presented the current state of work on the development and application of the freely available PASS Online web resource. Already more than 150 articles have been published, in which biological activity spectrum prediction results obtained using PASS Online have been used (Supplement 2, Supplementary information). In more than half of these studies, the computer prediction results have been confirmed either by direct experimental testing *in vitro* or *in vivo* or by referring to published literature data. Unfortunately, the authors of some of the studies given in Supplement 2 draw conclusions about a given activity for compounds synthesized in their laboratories only on the basis of computer predictions, which is incorrect.

Our recent survey of PASS Online users has shown that the overwhelming majority are researchers from academic institutions and universities. The important role of academic research in the search for new drugs is well known [68-70]. Since academic researchers have greater freedom in selecting subjects for study, they often generate new ideas, which later transform into applied projects designed for the development of safer and more efficient drugs. However, experience has shown that the extremely limited possibilities for experimental testing of the predicted types of activity pose a difficult obstacle for many researchers.

On the other hand, a solution to this problem may lie in a joining of forces of organic chemists, molecular and cell biologists, pharmacologists, and toxicologists, whose research is related, to some extent, to the search for new drugs using the Internet [71]. Such an interactive multi-user system of computer support for research using the Internet for the organization of research networks for the analysis of molecular structure - biological activity relationships may be realized on the basis of the PASS Online web resource [53]. This would require the establishment of a knowledge base, which can be improved jointly, and computer system for the prediction of the biological activity spectra of organic compounds. For this purpose, the organization and demarcation of the work space of the users, the introduction of a system for information exchange, monitoring the introduced information so that it is both sufficient and noncontradictory, and distributing tasks within groups of researchers are demanded. These problems may be solved by establishing a program complex constructed using modern tools and providing for client-server interaction. Such work has already been undertaken and on condition of the active participation of the PASS Online users, this web resource will constantly evolve in this direction.

The file containing supplementary information is available on the site http://hgs.osi.lv and contains Supplement 1 "Theoretical Basis for Analysis of Structure-activity Relationships and Prediction of the Activity of New Compounds Using the PASS Computer Program" and Supplement 2 "Review of Publications Using the Prediction of Biological Activity Spectra Obtained by the PASS Online Program".

This work was supported by the Russian Foundation for Basic Research (grant 12-07-00597-a).

REFERENCES

1. http://www.chemnavigator.com/cnc/products/iRL.asp.
2. L. Ruddigkeit, L. C. Blum, and J.-L. Reymond, *J. Chem. Inf. Model.*, **53**, 56 (2013).
3. *Thomson Reuters Integrity*^{SM}, http://integrity.thomson-pharma.com.
4. A. L. Hopkins, *Nat. Chem. Biol.*, **4**, 682 (2008).
5. R. Morphy and Z. Rankovic, *J. Med. Chem.*, **49**, 4961 (2006).

6. I. M. Kapetanovic, *Chem.-Biol. Interact.*, **171**, 165 (2008).
7. C. G. Wermuth (editor), *The Practice of Medicinal Chemistry*, Third Edition, Academic Press, San Diego (2008).
8. C. Hansch, *Acc. Chem. Res.*, **26**, 147 (1993).
9. R. Franke, *Theoretical Drug Design Methods*, Akademie-Verlag, Berlin (1984), p. 382.
10. V. V. Avidon, *Khim.-Farm. Zh.*, **8**, 22 (1974).
11. V. V. Avidon and V. S. Arolovich, *Nauchno-Tekhnicheskaya Informatsiya*, Ser. 2, No. 5, 26 (1975).
12. V. V. Avidon, V. S. Arolovich, V. G. Blinova, S. P. Kozlova, and A. M. Freidina, *Khim.-Farm. Zh.*, **17**, 321 (1983).
13. V. V. Avidon, V. S. Arolovich, S. P. Kozlova, and L. A. Piruzyan, *Khim.-Farm. Zh.*, **12**, No. 5, 88 (1978).
14. V. V. Avidon, V. S. Arolovich, S. P. Kozlova, and L. A. Piruzyan, *Khim.-Farm. Zh.*, **12**, No. 6, 99 (1978).
15. V. E. Golender and A. B. Rozenblit, *Avtomatika i Telemekhanika*, **11**, 99 (1974).
16. V. E. Golender and A. B. Rozenblit, *Computer Methods for Drug Design* [in Russian], Zinatne, Riga (1978).
17. V. E. Golender and A. B. Rozenblit, *Zh. Vses. Khim. Obshch. Im. D. I. Mendeleeva*, **25**, No. 1, 28 (1980).
18. A. B. Rozenblit and V. E. Golender, *Logic-combinatorial Methods in Drug Design* [in Russian], Zinatne, Riga (1984).
19. L. A. Piruzyan, V. V. Avidon, A. B. Rozenblit, V. S. Arolovich, V. E. Golender, S. P. Kozlova, E. M. Mikhailovskii, and E. G. Gavrishchuk, *Khim.-Farm. Zh.*, **11,** 35 (1977).
20. V. V. Avidon, I. A. Pomerantsev, V. E. Golender, and A. B. Rozenblit, *J. Chem. Inf. Comput. Sci.*, **22**, 207 (1982).
21. V. V. Avidon and L. A. Leksina, *Nauchno-Tekhnicheskaya Informatsiya*, Ser. 2, No. 3, 22 (1975).
22. N. Veretennikova, A. Skorova, V. Kudryashova, A. Rozenblit, A. Barkans, Ya. Betinsh, V. Drboglav, L. Gitlina, Ya. Grinfelds, P. Mellis, D. Petersone, and V. Shatokhin, in: F. Chen and G. Schüürmann (editors), *Quantitative Structure-Activity Relationships in Environmental Sciences*, SETAC, Pensacola (1997), p. 115.
23. G. M. Barenboim, L. A. Piruzyan, and A. G. Malenkov, *Vestn. Akad. Nauk SSSR,* No. 2, 50 (1977).
24. L. A. Piruzyan, A. G. Malenkov, and G. M. Barenboim, *Chemical Aspects of Human Activities and Nature Conservation* [in Russian], Preprint, Otdel. Inst. Khim. Fiz. Akad. Nauk SSSR, Chernogolovka (1979).
25. G. M. Barenboim and A. G. Malenkov, *Biologically Active Compounds. New Search Principles* [in Russian], Nauka, Moscow (1986).
26. Yu. V. Burov, L. V. Korol'chenko, and V. V. Poroikov, *Byull. Vses. Nauchn. Tsentra po Bezopasnosti Biologicheski Aktivnykh Veshchestv*, No. 1, 4 (1990).
27. V. V. Poroikov, D. A. Filimonov, and A. P. Budunova, *Nauchno-tekhnicheskaya Informatsiya*, Ser. 2, No. 6, 11 (1993).
28. D. A. Filimonov, V. V. Poroikov, E. I. Karaicheva, R. K. Kazaryan, A. P. Budunova, E. M. Mikhailovskii, A. V. Rudnitskikh, L. V. Goncharenko, and Yu. V. Burov, *Eksperimental'naya i Klinicheskaya Farmakologiya*, **58**, No. 2, 56 (1995).
29. D. A. Filimonov and V. V. Poroikov, in: M. G. Ford, R. Greenwood, G. T. Brooks, and R. Franke (editors), *Bioactive Compound Design: Possibilities for Industrial Use*, BIOS Scientific Publishers, Oxford (1996), p. 47.
30. T. A. Gloriozova, D. A. Filimonov, A. A. Lagunin, and V. V. Poroikov, *Khim.-farm. Zh.*, **32**, No. 12, 33 (1998).
31. A. V. Stepanchikova, A. A. Lagunin, D. A. Filimonov, and V. V. Poroikov, *Curr. Med. Chem.*, **10**, 225 (2003).
32. D. A. Filimonov and V. V. Poroikov, *Ros. Khim. Zh.*, **L**, No. 2, 66 (2006).

33. D. A. Filimonov and V. V. Poroikov, in: A. Varnek and A. Tropsha (editors), *Chemoinformatics Approaches to Virtual Screening*, RSC Publishing, Cambridge (2008), p. 182.
34. D. A. Filimonov, V. V. Poroikov, T. A. Gloriozova, and A. A. Lagunin, *Certificate for the Official Registration of the PASS Computer Program No. 2006613275, September 15, 2006* [in Russian], Federal Service for Intellectual Property, Patents, and Trademarks, Moscow.
35. D. Filimonov, V. Poroikov, Yu. Borodina, and T. Gloriozova, *J. Chem. Inf. Comput. Sci.*, **39**, 666 (1999).
36. V. V. Poroikov, D. A. Filimonov, Yu. V. Borodina, A. A. Lagunin, and A. Kos, *J. Chem. Inf. Comput. Sci.*, **40**, 1349 (2000).
37. V. A. Trapkov, A. P. Budunova, O. A. Burova, D. A. Filimonov, and V. V. Poroikov, *Voprosy Med. Khim.*, **43**, 39 (1997).
38. A. Geronikaki, V. Poroikov, D. Hadjipavlou-Litina, D. Filimonov, A. Lagunin, and R. Mgonzo, *Quant. Struct.-Act. Relat.*, **18**, 16 (1999).
39. V. V. Poroikov, D. A. Filimonov, W.-D. Ihlenfeldt, T. A. Gloriozova, A. A. Lagunin, Yu. V. Borodina, A.V. Stepanchikova, and M. C. Nicklaus, *J. Chem. Inf. Comput. Sci.*, **43**, 228 (2003).
40. A. Geronikaki, J. Dearden, D. Filimonov, I. Galaeva, T. Garibova, T. Gloriozova, V. Krajneva, A. Lagunin, F. Macaev, G. Molodavkin, V. Poroikov, S. Pogrebnoi, F. Shepeli, T. Voronina, M. Tsitlakidou, and L. Vlad, *J. Med. Chem.*, **47**, 2870 (2004).
41. A. Geronikaki, S. Vasilevsky, D. Hadjipavlou-Litina, A. Lagunin, and V. Poroikov, *Khim. Geterotsikl. Soedin.*, 769 (2006). [*Chem. Heterocycl. Compd.*, **42**, 675 (2006)].
42. A. A. Lagunin, O. A. Gomazkov, D. A. Filimonov, T. A. Gureeva, E. A. Dilakyan, E. V. Kugaevskaya, Yu. E. Elisseeva, N. I. Solovyeva, and V. V. Poroikov, *J. Med. Chem.*, **46**, 3326 (2003).
43. A. A. Geronikaki, A. A. Lagunin, D. I. Hadjipavlou-Litina, P. T. Elefteriou, D. A. Filimonov, V. V. Poroikov, I. Alam, and A. K. Saxena, *J. Med. Chem.*, **51**, 1601 (2008).
44. V. Poroikov, D. Akimov, E. Shabelnikova, and D. Filimonov, *SAR QSAR Environ. Res.*, **12**, 327 (2001).
45. S. A. Kryzhanovskii, R. M. Salimov, A. A. Lagunin, D. A. Filimonov, T. A. Gloriozova, and V. V. Poroikov, *Khim.-farm. Zh.*, **45**, No. 10, 25 (2011).
46. V. Poroikov, D. Filimonov, A. Lagunin, T. Gloriozova, and A. Zakharov, *SAR QSAR Environ. Res.*, **18**, 101 (2007).
47. S. M. Ivanov, A. A. Lagunin, A. V. Zakharov, D. A. Filimonov, and V. V. Poroikov, *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, **7**, 40 (2013).
48. A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grier, B. A. Leland, and J. Laufer, *J. Chem. Inf. Comput. Sci.*, **32**, 244 (1992).
49. A. Lagunin, A. Stepanchikova, D. Filimonov, and V. Poroikov, *Bioinformatics*, **16**, 747 (2000).
50. A. V. Sadym, A. A. Lagunin, D. A. Filimonov, and V. V. Poroikov, *Khim.-farm. Zh.*, **36**, No. 10, 21 (2002).
51. A. Sadym, A. Lagunin, D. Filimonov, and V. Poroikov, *SAR QSAR Environ. Res.*, **14**, 339 (2003).
52. A. Geronikaki, D. Druzhilovsky, A. Zakharov, and V. Poroikov, *SAR QSAR Environ. Res.*, **19**, 27 (2008).
53. *PASS Online*, http://www.way2drug.com/passonline.
54. D. Weininger, *J. Chem. Inf. Comput. Sci.*, **28**, 31 (1988).
55. http://www.chemaxon.com.
56. http://www.chemspider.com/.
57. http://bioinformatics.charite.de/superpred/.
58. http://cpi.bio-x.cn/drar/.
59. D. Filimonov, A. Lagunin, A. Rudik, D. Druzhilovsky, S. Ivanov, P. Pogodin, and V. Poroikov, in: *Abstracts of the 19th European Symposium on Quantitative Structure-Activity Relationships*, Vienna, Austria, August 26-30, 2012, p. 60.

60. C. B-Rao, A. Kulkarni-Almeida, K. V. Katkar, S. Khanna, U. Ghosh, A. Keche, P. Shah, A. Srivastava, V. Korde, K. V. S. Nemmani, N. J. Deshmukh, A. Dixit, M. K. Brahma, U. Bahirat, L. Doshi, R. Sharma, and H. Sivaramakrishnan, *Bioorg. Med. Chem.*, **20**, 2930 (2012).
61. D. Verbanac, S. C. Jain, N. Jain, M. Chand, H. Čipčić Paljetak, M. Matijašić, M. Perić, V. Stepanić, and L. Saso, *Bioorg. Med. Chem.*, **20**, 3180 (2012).
62. A. Kumar, P. Lohan, D. K. Aneja, G. K. Gupta, D. Kaushik, and O. Prakash, *Eur. J. Med. Chem.*, **50**, 81 (2012).
63. I. Shcherbakova, *Khim. Geterotsikl. Soedin.*, 6 (2013). [*Chem. Heterocycl. Compd.*, **49**, 2 (2013)].
64. http://www.cerep.fr.
65. http://www.kinomescan.com.
66. P. Eleftheriou, A. Geronikaki, D. Hadjipavlou-Litina, P. Vicini, O. Filz, D. Filimonov, V. Poroikov, S. S. Chaudhaery, K. K. Roy, and A. Saxena, *Eur. J. Med. Chem.*, **47**, 111 (2012).
67. A. Lagunin, D. A. Filimonov, and V. V. Poroikov, *Curr. Pharm. Des.*, **16**, 1703 (2010).
68. J. Frearson and P. Wyatt, *Expert Opin. Drug Discovery*, **5**, 909 (2010).
69. W. Rhodes, *J. Transl. Med.*, **10** (Suppl. 2), A42 (2012).
70. D. M. Huryn, L. O. Resnick, and P. Wipf, *J. Med. Chem.*, **56**, 7161 (2013).
71. S. Ekins and B. A. Bunin, in: S. Cortagere (editor), *In Silico Models for Drug Discovery*, Series: Methods in Molecular Biology, Vol. 993, Humana Press, New York (2013), p. 139.