*SAR and QSAR in Environmental Research*
2001, Vol. 12, pp. 327–344
Reprints available directly from the publisher
Photocopying permitted by license only



# TOP 200 MEDICINES: CAN NEW ACTIONS BE DISCOVERED THROUGH COMPUTER-AIDED PREDICTION?*

V. POROIKOV[a,b,†], D. AKIMOV[b], E. SHABELNIKOVA[b] and D. FILIMONOV[a]

[a]*Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, 10, Pogodinskaya Street, Moscow, 119832, Russia;* [b]*Medical and Biological Faculty of the Russian State Medical University, 1, Ostrovityanova Street, Moscow, 117869, Russia*

*(Received 30 June 2000; In final form 31 March 2001)*

Computer-aided prediction of the biological activity spectra by the program PASS was applied to a set of 130 pharmaceuticals from the list of the Top 200 medicines. The known pharmacological effects were found in the predicted activity spectra in 93.2% of cases. Additionally, the probability of some supplementary effects was also predicted to be significant, including angiogenesis inhibition, bone formation stimulation, possible use in cognition disorders treatment, multiple sclerosis treatment, *etc.* These predictions, if confirmed experimentally, may become a cause for a new application of pharmaceuticals from the Top 200 list. Most of known side and toxic effects were also predicted by PASS. PASS predictions at earlier R&D stages may thus provide a basis for finding new "leads" among already launched drugs and may help direct more attention to those particular effects of pharmaceuticals in clinical use which become apparent only in a small part of the population and require additional precautions.

*Keywords*: Biological activity spectra; Top 200 medicines; Side effect; Toxicity; Computer-aided prediction; PASS

## INTRODUCTION

The process of new drug research is based on the existing knowledge of the pharmacological effects that may prove useful for the treatment of certain

*Presented at the Symposium *Bioinformatics and Computer-Aided Drug Discovery* (April 12–13, 2000), Moscow, Russia.

†Corresponding author. e-mail: vvp@ibmh.msk.su

disorders and the classes of molecules that are capable of producing such effects. The firm's research strategy [1] and the test standards for the drug safety assessment [2] restrict the investigation of any lead compound and drug-candidate. Besides, the information on the biological activity of each compound is always incomplete because of the concentration on only directed testing, which does not cover many possibly existing but non-tested activities. If one were able to suggest the probable activities for any compound, the directions of and priority in their testing could be established.

Computer aided structure-activity relationship (SAR) analysis is now widely used for new lead finding and optimization. Most of the SAR/QSAR/Modeling methods are only applicable to one or several types of biological activity within the same chemical series [3–5] and, therefore are unable to elucidate the general biological "potential" of a molecule under study. Recently, the possibility for simultaneous prediction of many pharmacological effects, their mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity has been demonstrated [6–8]. The total complex of these biological actions is called the biological activity spectrum of the compound [8]. The mean accuracy obtained with the computer program PASS [8] in leave one out cross-validation for more than 30,000 compounds with more than 500 activities from the training set was about 85% [8]. Recently, it was shown in experiments with the principal compounds from the MDDR database [9] that the approach used in PASS provided highly robust estimates for structure-activity relationships [10]. Thus, PASS ones trained can be applied to predict the activities for a new compound. In particular, PASS can be applied to already approved pharmaceuticals. Finding a new "lead" among the launched drugs is more attractive because their safety is already proven, in many cases, by previous clinical practice. On the other hand, a computer-generated "side-action" hypothesis may help focus the attention on those particular effects of pharmaceuticals in clinical use, which become apparent only in a minority of the population and require additional precautions.

The purpose of this work is to present the predictions of biological activity spectra for compounds from the list of the Top 200 pharmaceuticals [11], which are especially significant for the today's healthcare. Such computer-aided predictions may open new horizons in the study of some of the most widely used pharmaceuticals and, when compared with the experimentally established activities give an additional evaluation of the reliability for the computer program PASS.

## MATERIALS AND METHODS

### Data Set

The list of the Top 200 pharmaceuticals, which are the best-selling drugs in the U.S., is available from the Internet [11]. Based on their international nonproprietary and trade names, we collected the structural formulae of the substances and created the database under the ISIS/Base system (MDL Information Systems, Inc.) [9]. Information about the pharmacological main and side effects and specific toxicity of the Top 200 pharmaceuticals was collected from Refs. [11, 12]. Eighty-five percent of the pharmaceuticals had a molecular weight between 160–480 D [13] (typical drug-like compounds). The other 15% had a molecular weight between 129 and 780 D. Molecules both too small and too large were eliminated, such as: Potassium sodium chloride (K-Dur-20), Klor-Con, Potassium Chloride (ETHHEX)), Insulin (Humulin N, Humulin 70/30, Humulin R), Calcitoninum (Miacalcin Nasal), Polumyxin (Neomycin/Polymyx/HC). Mixtures of different substances or compounds with unidentified structural formulae (Premarin, Prempro) were not included in the database either. We excluded duplicates from the database which, are the same substances present in the list of Top 200 drugs under several trade names: Levothyroxine (67) – Synthroid (Knoll) and Levoxyl (Jones Pharma); Amoxicillin (10) – Amoxicillin (Teva Pharma), Trimox (Apothecon) and Amoxil (SK Beecham); Hydrocodone w/APAP (59) – Hydrocodone w/APAP (Watson, Mallincrodt, Qualitest), *etc.* The final database consisted of 130 drug substances that were used in further computer experiments.

### Prediction of Biological Activity Spectra

The computer program PASS (Prediction of Activity Spectra for Substances) [6–8] predicts simultaneously several hundred biological activity types, in particular main and side pharmacological effects, mechanisms of their action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity. The prediction is based on a SAR analysis of the training set containing more than 35000 compounds, which show more than 500 kinds of biological activity. The basic elements of PASS include the representation of biological activity, the description of the chemical structures, the structure-activity relationship data and knowledge base (SAR base), and the algorithm of the activity spectra estimation.

### *Biological Activity*

A biological activity is the result of the chemical compound's interaction with a biological entity. In a clinical study, a biological entity is represented by a human organism. In the preclinical testing, it is the experimental animals (*in vivo*) and the experimental models (*in vitro*). A biological activity depends on the peculiarities of the compound (the structure and physicochemical properties), the biological entity (species, sex, age, *etc.*), and the mode of treatment (dose, route, *etc.*). Each biologically active compound possesses a wide spectrum of the different effects. Some of them are useful in the treatment of definite diseases but the others cause the various side and toxic effects. The total complex of activities caused by the compound in the biological entities can be called "*biological activity spectrum of a substance*".

A *biological activity spectrum* should present every activity of a compound despite of the differences in the essential conditions of their experimental determination. On the other hand, a sufficiently large set of substances can be collected only using many different sources because the information taken from a unique publication never covers all aspects of a biological action of the described substance. Thus, the "*biological activity spectrum*" may be defined as the "intrinsic" compound property depending only on its structure and physicochemical characteristics, and can be identified only qualitatively.

The biological activities are described qualitatively as the above mentioned biological activity spectra of compounds. Each compound has a list of those activities, which it possesses under some conditions.

### *Chemical Structure Description*

In our paper published recently [14], we described the substructure descriptors called "Multilevel Neighborhoods of Atoms" (MNA). MNA descriptors are based on a structure representation that does not specify the bond types and includes hydrogen atoms according to valence and partial charge of atoms. MNA descriptors are generated as a recursively defined sequence: zero-level MNA descriptor for each atom is the mark $A$ of the atom itself; any next-level MNA descriptor for each atom is the substructure notation $A(D_1D_2..D_i...)$, where $D_i$ is the previous-level MNA descriptor for the $i$-th immediate neighbors of the atom.

This iterative process can be continued including 2nd, 3rd, *etc.*, neighborhoods of each atom. It is important to emphasize that the atom mark may include not only the atom type but also any additional information about the atom, for example, its belonging to a certain ring or

chain. A structure of a molecule is represented in PASS as a set of the 1st- and 2nd-level MNA descriptors. In 2nd-level MNA descriptors we use the mark "-" to define the belonging to a chain.

The *structure equivalence* is an important feature of the PASS approach. The structures of compounds are considered as equivalent if they have the same molecular formulae and the same set of MNA descriptors. Only unique structures are included into the SAR base. Since MNA descriptors do not represent the stereochemical peculiarities of a molecule, the compounds that differ only in stereochemistry, are formally considered as equivalent ones.

### *SAR Base*

The prediction is carried out using the SAR base, which is created based on an analysis of the training set(s) containing the biologically active compounds. The SAR base includes the activity types and the MNA descriptor dictionaries, the substance structure and an activity description sample, and the SAR information.

To include the training set into the SAR base, the MNA descriptors are generated for each compound in this set. If a structure is not completely defined, *i.e.*, includes undetermined atoms or residuae, the compound is not included in the SAR base. If an equivalent structure is found in the SAR base then new activity spectrum is combined with the existing one. The SAR data and knowledge are generated by the *training procedure* described below.

In PASS (version 1.41), as presented in this work, the SAR base contains more than 35000 substances. To create this SAR base, the training set containing drugs and well-known biologically active compounds was prepared. In the different sources the biological activities are described by different terms. Therefore the activity spectra were standardized and combined for all equivalent compounds from many sources. The MNA descriptor's dictionary contains about 36000 items. The number of different activities exceeds 800, but many of them are represented by less than 3 compounds in the training set. The total available "activity spectrum", *i.e.*, the list of predictable types of biological activities includes more than 500 predictable biological activities. This list is modified permanently and its current version can be found on our web-site [8].

### *Algorithm of the Activity Spectra Estimation*

The prediction algorithm was chosen from more than a hundred variants tested through the several past years. It is based on the following

SAR data:

$n$ is the total number of compounds in the SAR base;
$n_i$ is the number of compounds, containing the descriptor $i$ in the structure description;
$n_j$ is the number of compounds, containing the activity $j$ in the activity spectrum;
$n_{ij}$ is the number of compounds, containing both the descriptor $i$ and the kind of activity $j$.

To estimate the activity spectrum the MNA descriptor set is generated for the compound under prediction. If this compound has an equivalent structure in the SAR base, that structure is excluded from the SAR base during the prediction by default. In a such case the calculations are performed by using $n-1$, $n_i-1$, when the compound contains the descriptor $i$ in the structure description, $n_j-1$, when it contains the activity $j$ in the activity spectrum, $n_{ij}-1$, when it contains both the descriptor $i$ and the activity $j$. For each of the predictable activities $j$ the initial estimation $t_j$ is calculated as:

$$s_j = Sin(\Sigma_i ArcSin(r_i * (2 * p_{ij} - 1))/m),$$
$$s_{0j} = Sin(\Sigma_i ArcSin(r_i * (2 * p_j - 1))/m),$$
$$t_j = (1 + (s_j - s_{0j})/(1 - s_j * s_{0j}))/2,$$

where the summation is taken over the MNA descriptors of the compound under prediction, and $m$ is that number, $r_i = n_i/(n_i + 0.5/m)$ is the regulating factor, $p_j = n_j/n$ is the estimation of the *a priori* probability of the activity $j$, and $p_{ij} = n_{ij}/n_i$ is the estimation of the conditional probability of the activity $j$ for the MNA descriptor $i$.

The result of the prediction for a new compound is presented by the *activity spectrum*, which is the ranked list of the probabilities "to be active" $Pa$, "to be inactive" $Pi$, and the type of activity. The ranking is done in descending order of $Pa-Pi$; thus, more probable activities are at the top of predicted spectrum. A compound is considered active if $Pa-Pi$ exceeds the cutoff value, *e.g.*, by default $Pa-Pi < 0.0$.

The probabilities $Pa$ and $Pi$ are functions of the initial estimation, and defined by the equations:

$$A_j(Pa) = t_j,\ I_j(Pi) = t_j,$$

where the functions $A_j$, $I_j$ are the SAR knowledge obtained as the final result of the *training procedure*, which is described below.

### *Training Procedure*

The SAR data $n$, $n_i$, $n_j$, $n_{ij}$ are generated. For each kind of activity $j$, for each $p$ of the $n_j$ active compound, and for each $q$ of the $n-n_j$ inactive compounds in the SAR base the leave-one-out initial estimations, *i.e.*, after exclusion this compound from the SAR base, $t_{jp}$ and $t_{jq}$ are calculated. The $n_j$ estimates $t_{jp}$ for the active compounds are sorted in ascending order; the $n-n_j$ estimates $t_{jq}$ for the inactive compounds are sorted in descending order. The functions $A_j$, $I_j$ are calculated as the conditional expectations:

$$A_j(F) = \Sigma_p Pr(p-1,\ n_j-1,\ F)t_{jp},$$
$$I_j(F) = \Sigma_q Pr(q-1,\ n-n_j-1,\ F)t_{jq},$$

where $Pr(m,n,F) = C_n^m F^m(1-F)^{n-m}$ is the binomial distribution, $C_n^m = n!/m!(n-m)!$ is the binomial coefficient, $F$ is in the range [0, 1]. It is clear that $A_j$ and $I_j$ are the estimations of the quantiles of the probability distributions of the initial estimation. Thus, the probabilities $Pa$ and $Pi$ are also the measures of belonging to subsets of "active" and "inactive" compounds, but can also be seen as the probabilities of the 1st and 2nd kinds of the prediction error, respectively. All of these possible interpretations of the probabilities $Pa$ and $Pi$ are equivalent and useful.

A test of PASS in blind mode by 9 scientists from 8 countries has shown that the mean accuracy of prediction is 82.6% for 108 non-congeneric compounds with 46 different types of biological activity [15]. New lead compounds with antiulcer, antitumor, hepatoprotective, antibacterial and other actions were found based on the predicted activity spectra [8].

### *Data Preparation*

A data set of 130 drug substances, including 20 mixtures, *e.g.*, Hydrocodone w/APAP (59),[1] Amoxicillin/Clavulanate (9), *etc.* (see Tab. I), was prepared. The structure of each component of each mixture was represented separately. We exported the structural formulae of the drugs' components from ISIS/Base as SDfile [8, 9], and predicted the activity spectra for them with a cutoff condition $Pa-Pi>0.0$. Only 2 components, Norgestimate (91), and Nitrofurantoin (87) are not found in the SAR base (they are marked in Tab. I by an asterisk *). The predictions of the activity spectra for all other substances were made with the exclusion of appropriate information from the SAR base (see above).

[1]The numbers in the brackets correspond to the numbers in Table I.

TABLE I Results of activity prediction for 130 substances from the list of Top 200 drugs. The activity types are given in accordance with Ref. [11]; *Pa* is the estimation of the probability "to be active"

| | *International Nonproprietary and Trade Names* | *Activity Types* | *Pa* |
|---|---|---|---|
| 1 | Acetaminophen/Codeine | Analgesic, general | 0.960 |
| | | Analgesic-narcotic | 0.957 |
| | | Analgesic-non-narcotic | 0.412 |
| | | Antitussive/Expectorant | 0.975 |
| 2 | Albuterol; Albuterol Aerosol | Antiasthmatic | N/P |
| | | Beta(2)-receptor agonist | 0.897 |
| | | Bronchodilator | 0.618 |
| 3 | Alendronate; Fosamax | Calcium metabolism | 0.978 |
| 4 | Allopurinol | Xanthine oxidase inhibitor | 0.625 |
| 5 | Alprazolam | Antianxiety | 0.908 |
| | | Sedative/Hypnotic | 0.891 |
| 6 | Amitriptyline | Antidepressant | 0.970 |
| 7 | Amlodipine/Benazepril; Lotrel | ACE-inhibitor | 0.522 |
| | | Antianginal | 0.650 |
| | | Antihypertensive | 0.751 |
| 8 | Amlodipine; Norvasc | Antihypertensive | 0.751 |
| | | Calcium channel blocker | 0.915 |
| 9 | Amoxicillin/Clavulanate; Augmentin | Penicillin | 0.703 |
| 10 | Amoxicillin; Trimox, Amoxil | Penicillin | 0.703 |
| 11 | Amphetamine; Adderall | CNS, stimulant | 0.838 |
| 12 | Atenolol | Antianginal | 0.429 |
| | | Antihypertensive | 0.409 |
| | | Beta(1)-receptor blocker | 0.549 |
| 13 | Atorvastatin; Lipitor | Antihyperlipidemic | 0.809 |
| | | HGM-CoA reductase inhibitor | 0.475 |
| 14 | Azithromycin; Zithromax, Zithromax Susp | Macrolide | 0.865 |
| 15 | Beclomethasone; Vancenase AQ DS | Antiasthmatic | N/P |
| | | Glucocorticoid | 0.785 |
| 16 | Benazepril; Lotensin | ACE-inhibitor | 0.522 |
| | | Antihypertensive | 0.522 |
| 17 | Bisoprolol/HCTZ; Ziac | Antihypertensive | 0.856 |
| | | Beta(1)-receptor blocker | 0.730 |
| | | Diuretic | 0.974 |
| 18 | Budesonide; Rhinocort | Corticosteroid | 0.712 |
| 19 | Bupropion; Wellbutrin SR, Zyban | Antidepressant | 0.207 |
| 20 | Buspirone; BuSpar | Antianxiety | 0.946 |
| 21 | Carisoprodol; Rela, Soma | Skeletal muscle relaxant | 0.865 |
| 22 | Cefprozil; Cefzi | Cephalosporin | 0.633 |
| 23 | Cefuroxime; Ceftin | Cephalosporin | 0.711 |
| 24 | Cephalexin | Cephalosporin | 0.682 |
| 25 | Cetirizine; Zyrtec | Antihistaminic | 0.797 |
| | | H1-receptor antagonist | 0.552 |
| 26 | Cimetidine | Antiulcerative | 0.781 |
| | | H2-receptor antagonist | 0.814 |
| 27 | Ciprofloxacin; Cipro | Anti-infective | 0.326 |
| | | Quinolone/Derivative | 0.631 |
| 28 | Cisapride; Propulsid | Acid/Peptic disorder | N/P |
| | | Gastrointestinal agent | 0.685 |
| 29 | Clarithromycin; Biaxin | Macrolide | 0.954 |

TABLE I (Continued)

| | *International Nonproprietary and Trade Names* | *Activity Types* | *Pa* |
|---|---|---|---|
| 30 | Clonazepam; Klonopin | Anticonvulsant | 0.850 |
| 31 | Clonidine | Antihypertensive | N/P |
| 32 | Clotrimoxazole/Betamethasone; Lotrisone | Antifungal | 0.707 |
| | | Corticosteroid | 0.877 |
| 33 | Cyclobenzaprine | Skeletal muscle relaxant | 0.502 |
| 34 | Desogestrel/Ethinyl Estradiol; Desogen | Contraceptive | 0.968 |
| 35 | Diazepam; Valium, Valrelease | Antianxiety | 0.954 |
| | | Anticonvulsant | 0.884 |
| | | Skeletal muscle relaxant | 0.826 |
| 36 | Digoxin; Lanoxin | Antiarrhythmic | 0.397 |
| | | Cardiac glycoside | 0.945 |
| 37 | Diltiazem; Cardizem CD | Antianginal | 0.815 |
| | | Antiarrhythmic | 0.490 |
| | | Antihypertensive | 0.728 |
| | | Calcium channel blocker | 0.927 |
| 38 | Divalproex; Depakote | Anticonvulsant | 0.714 |
| 39 | Doxazosin; Cardura | Alpha(1)-receptor antagonist | 0.734 |
| | | Antihypertensive | 0.450 |
| 40 | Enalapril; Vasotec | ACE-inhibitor | 0.592 |
| | | Antihypertensive | 0.680 |
| 41 | Erythromycin; Ery-Tab | Acne product | 0.538 |
| | | Macrolide | 0.969 |
| 42 | Estradiol | Estrogen | 0.763 |
| 43 | Famotidine; Pepcid | Acid/Peptic disorder | 0.994 |
| | | H2-receptor antagonist | 0.770 |
| 44 | Felodipine; Plendil | Antihypertensive | 0.769 |
| | | Calcium channel blocker | 0.954 |
| 45 | Fexofenadine; Allegra | Antihistaminic | 0.874 |
| | | H1-receptor antagonist | 0.363 |
| 46 | Fluconazole; Diflucan | Antifungal | 0.983 |
| 47 | Fluoxetine; Prozac | Antidepressant | 0.856 |
| 48 | Fluticasone; Flonase | Corticosteroid – inhalation/nasal | 0.968 |
| | | Topical steroid | 0.803 |
| 49 | Fluvastatin; Lescol | Antihyperlipidemic | 0.895 |
| | | HGM-CoA reductase inhibitor | 0.732 |
| 50 | Fosinopril; Monopril | ACE-inhibitor | 0.944 |
| | | Antihypertensive | 0.838 |
| 51 | Furosemide; Furosemide Oral, Lasix | Antihypertensive | 0.351 |
| | | Diuretic | 0.816 |
| 52 | Gabapentin; Neurontin | Anticonvulsant | 0.639 |
| 53 | Gemfibrozil | Antihyperlipidemic | 0.572 |
| 54 | Glimepiride; Amaryl | Blood glucose regulator | 0.582 |
| 55 | Glipizide | Blood glucose regulator | 0.669 |
| 56 | Glyburide; Azuglucon | Blood glucose regulator | 0.677 |
| 57 | Guaifenesin/Phenylpropanolamine | Antitussive/Expectorant | 0.531 |
| | | Cold remedy | 0.542 |
| 58 | Hydrochlorothiazide; Esidrix, Dichlotride | Antihypertensive | 0.856 |
| | | Diuretic | 0.974 |
| 59 | Hydrocodone w/APAP | Analgesic, general | 0.967 |
| | | Analgesic-narcotic | 0.961 |
| | | Antitussive/Expectorant | 0.886 |

TABLE I (Continued)

| | *International Nonproprietary and Trade Names* | *Activity Types* | *Pa* |
|---|---|---|---|
| 60 | Ibuprofen | Analgesic, general | 0.558 |
| | | Analgesic-non-narcotic | 0.614 |
| | | Antiarthritic | 0.522 |
| | | Anti-inflammatory | 0.812 |
| 61 | Ipratropium; Atrovent | Antihypertensive | 0.423 |
| | | Bronchodilator | N/P |
| 62 | Isosorbide Mononitrate; Imdur | Antianginal | 0.986 |
| 63 | Lansoprazole; Prevacid | Antiulcerative | 0.978 |
| 64 | Latanoprost; Xalatan | Antiglaucomic | 0.683 |
| 65 | Levofloxacin; Levaquin | Antibacterial | 0.627 |
| | | Quinolone/Derivate | 0.606 |
| 66 | Levonorgestrel/Ethinyl Estradiol; Triphasil | Contraceptive | 0.952 |
| 67 | Levothyroxine; Synthroid, Levoxyl | Thyroid hormone | 0.594 |
| 68 | Lisinopril/HCTZ; Zestoretic | ACE-inhibitor | 0.579 |
| | | Diuretic | 0.974 |
| 69 | Lisinopril; Prinivil, Zestril | ACE-inhibitor | 0.575 |
| | | Antihypertensive | 0.488 |
| 70 | Loracarbef; Lorabid | Antibacterial, miscellaneous | 0.762 |
| | | Cephalosporin | 0.606 |
| 71 | Loratadine/Pseudoephedrine; Claritin D 12HR, Claritin D 24HR | Antihistaminic | 0.513 |
| | | Cold remedy | 0.597 |
| 72 | Loratadine; Claritin | Antihistaminic | 0.513 |
| 73 | Lorazepam; Tavor, Norlormetazepam, Temesta | Antianxiety | 0.964 |
| | | Sedative/Hypnotic | 0.850 |
| 74 | Losartan/HCTZ; Hyzaar | Angiotensin 2 receptor antagonist | 0.978 |
| | | Diuretic | 0.974 |
| 75 | Losartan; Cozaar | Angiotensin 2 receptor antagonist | 0.992 |
| | | Antihypertensive | 0.993 |
| 76 | Lovastatin; Mevacor | Antihyperlipidemic | 0.956 |
| | | HGM-CoA reductase inhibitor | 0.936 |
| 77 | Medroxyprogesterone; Cycrin, Provera | Antineoplastic | 0.550 |
| | | Contraceptive | 0.860 |
| | | Progestin | 0.756 |
| 78 | Metformin; Glucophage | Blood glucose regulator | 0.807 |
| 79 | Methylprednisolone | Antiarthritic | N/P |
| | | Corticosteroid | 0.519 |
| 80 | Metoprolol; Toprol-XL | Antianginal | 0.575 |
| | | Antihypertensive | 0.534 |
| | | Beta(1)-receptor blocker | 0.805 |
| 81 | Mometasone; Elocon | Anti-inflammatory | 0.975 |
| | | Corticosteroid | 0.705 |
| 82 | Mupirocin; Bactroban | Dermatologic/Miscellaneous | 0.408 |
| 83 | Nabumetone; Relafen | Antiarthritic | 0.583 |
| 84 | Naproxen; Naprosyn, Equiproxen | Analgesic, general | 0.517 |
| | | Analgesic-non-narcotic | 0.629 |
| | | Antiarthritic | 0.409 |
| 85 | Nefazodone; Serzone | Antidepressant | 0.881 |

TABLE I (Continued)

| | *International Nonproprietary and Trade Names* | *Activity Types* | *Pa* |
|---|---|---|---|
| 86 | Nifedipine; Procardia XL, Adalat CC | Antianginal | 0.754 |
| | | Antihypertensive | 0.860 |
| | | Calcium channel blocker | 0.989 |
| 87 | Nitrofurantoin*; Macrobid | Antiseptic, urinary tract | 0.614 |
| 88 | Nitroglycerin; Nitrostat | Antianginal | 0.995 |
| | | Antihypertensive | 0.857 |
| 89 | Nizatidine; Axid | Antihistaminic | 0.985 |
| | | Antiulcerative | 0.856 |
| | | H2-receptor antagonist | 0.907 |
| 90 | Norethindrone/Ethinyl Estradiol; Ortho-Novum | Contraceptive | 0.956 |
| 91 | Norgestimate*/Ethinyl Estradiol; OrthoTri-Cyclen | Contraceptive | 0.896 |
| 92 | Omeprazole; Prilosec | Acid/Peptic disorder | 0.987 |
| 93 | Oxaprozin; Daypro | Anti-inflammatory | 0.811 |
| 94 | Oxycodone/APAP; Roxicet | Analgesic, general | 0.950 |
| | | Analgesic-narcotic | 0.935 |
| 95 | Paroxetine; Paxil | Antidepressant | 0.718 |
| 96 | Penicillin V Potassium; Veetids | Penicillin | 0.674 |
| 97 | Phenytoin; Dilantin | Anticonvulsant | 0.897 |
| 98 | Pravastatin; Pravachol | Antihyperlipidemic | 0.926 |
| | | HGM-CoA reductase inhibitor | 0.873 |
| 99 | Prednisone | Antiarthritic | N/P |
| | | Antineoplastic | 0.436 |
| | | Corticosteroid | 0.508 |
| | | Ocular anti-inflammatory | 0.900 |
| 100 | Promethazine; Phenergan | Anesthetic | 0.786 |
| | | Antiemetic | 0.636 |
| | | Antihistaminic | 0.917 |
| | | Antitussive/Expectorant | 0.820 |
| | | Sedative/Hypnotic | 0.915 |
| 101 | Propoxyphene N/APAP | Analgesic, general | 0.939 |
| | | Analgesic-narcotic | 0.625 |
| 102 | Propranolol | Antianginal | 0.487 |
| | | Antiarrhythmic | 0.759 |
| | | Antimigraine | N/P |
| | | Beta-receptor blocker | 0.881 |
| 103 | Quinapril; Accupril | ACE-inhibitor | 0.644 |
| | | Antihypertensive | 0.594 |
| 104 | Ramipril; Altace | ACE-inhibitor | 0.802 |
| | | Antihypertensive | 0.619 |
| 105 | Ranitidine | Antihistaminic | 0.994 |
| | | Antiulcerative | 0.986 |
| | | H2-receptor antagonist | 0.970 |
| 106 | Risperidone; Risperdal | Antipsychotic/Antimanic | 0.970 |
| 107 | Salmeterol; Servent | Bronchodilator | 0.621 |
| 108 | Sertraline; Zoloft | Antidepressant | 0.448 |
| | | Serotonin reuptake inhibitor | 0.218 |
| 109 | Sildenafil Citrate; Viagra | (cGMP)-specific PDE-5 inhibitor | 0.839 |
| 110 | Simvastatin; Zocor | Antihyperlipidemic | 0.982 |
| | | HGM-CoA reductase inhibitor | 0.975 |

TABLE I (Continued)

| | *International Nonproprietary and Trade Names* | *Activity Types* | *Pa* |
|---|---|---|---|
| 111 | Sumatriptan; Imitrex | 5-HT1 receptor agonist | 0.985 |
| | | Antimigraine | 0.988 |
| 112 | Tamoxifen | Anti-estrogen | 0.953 |
| | | Antineoplastic | 0.444 |
| 113 | Temazepam | Sedative/Hypnotic | 0.969 |
| 114 | Terazosin; Hytrin | Alpha(1)-receptor antagonist | 0.711 |
| | | Antihypertensive | 0.461 |
| | | Diuretic | N/P |
| 115 | Tetracycline; Sumycin | Tetracycline | 0.586 |
| 116 | Timolol; Timoptic XE | Antiglaucomic | 0.566 |
| | | Antihypertensive | 0.446 |
| | | Antimigraine | N/P |
| | | Beta-receptor blocker | 0.531 |
| 117 | Tobramycin/Dexamethasone; Tobradex | Aminoglycoside | 0.871 |
| 118 | Tramadol; Ultram | Analgesic-non-narcotic | 0.877 |
| 119 | Trazodone | Analgesic, general | 0.545 |
| | | Antianxiety | 0.720 |
| | | Antiarthritic | N/P |
| | | Antidepressant | 0.933 |
| 120 | Tretinoin; Retin-A | Antineoplastic | 0.638 |
| | | Dermatologic/Miscellaneous | 0.954 |
| 121 | Triamcinolone; Azmacort | Antiasthmatic | N/P |
| | | Corticosteroid | 0.698 |
| | | NSAID | N/P |
| 122 | Triamterene/HCTZ | Antihypertensive | 0.856 |
| | | Diuretic | 0.974 |
| 123 | Trimeth/Sulfameth; Trimethoprim/Sulfa | Antimicrobial | 0.775 |
| 124 | Troglitazone; Rezulin | Antihyperglycemic | 0.596 |
| 125 | Valsartan; Diovan | ACE-inhibitor | N/P |
| | | Angiotensin 2 receptor antagonist | 0.370 |
| 126 | Venlafaxine; Effexor | Antidepressant | 0.312 |
| 127 | Verapamil | Antianginal | 0.700 |
| | | Antiarrhythmic | 0.753 |
| | | Antiarthritic | N/P |
| | | Antihypertensive | 0.610 |
| | | Calcium channel blocker | 0.559 |
| 128 | Warfarin; Coumadin | Anticoagulant | 0.785 |
| 129 | Zafirlukast; Accolate | Antiasthmatic | N/P |
| | | Peptide leukotriene receptor antagonist | 0.457 |
| 130 | Zolpidem; Ambien | Sedative/Hypnotic | 0.970 |

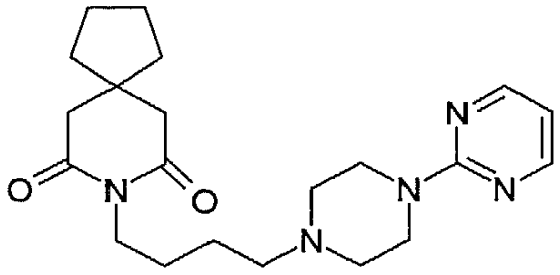
## RESULTS AND DISCUSSION

### Prediction of Known Activities

A comparison of the prediction results with biological activities known for each of 130 pharmaceuticals from the list of Top 200 drugs is presented in Table I. The names of activities are given in accordance with Ref. [11]. For

the majority of the compounds all known activities are predicted. There are fifteen compounds including Albuterol (2), Beclomethasone (15), Cisapride (28), *etc.*, for which some activities are not predicted. These activities are marked as "NP" in Table I. The mean accuracy of prediction, calculated for all 130 compounds and 236 activities is 93.2%.

Figure 1 presents, as an example, the predicted biological activity spectrum for Buspirone (20). All information about this compound and its




```
PASS 1.41 - Prediction of Activity Spectra for Substances
    Copyright (C) 1998-2000 V.V.Poroikov, D.A.Filimonov & Associates

    Chemical Structure File: buspiron.mol
>  <ACTIVITY_PREDICTION>
  34  Substructure descriptors;  0  new.
Exclude Original PASS Structure 25891 with activities:
              5 Hydroxytryptamine 1 agonist
              5 Hydroxytryptamine 1A agonist
              5 Hydroxytryptamine agonist
              Antidepressant
              Antiemetic
              Antipsychotic
              Anxiolytic
              Chemoprotective
              Hypotermic
              Psychosexual dysfunction treatment
              Psychotropic
  81  Possible activities at Pa > Pi.
  Pa     Pi     for Activity:
0.981  0.003  5 Hydroxytryptamine 1A agonist
0.969  0.006  Psychotropic
0.946  0.006  Anxiolytic
0.933  0.005  5 Hydroxytryptamine agonist
0.925  0.004  5 Hydroxytryptamine 1 agonist
0.851  0.006  Antipsychotic
0.836  0.007  Psychosexual dysfunction treatment
0.801  0.008  Antidepressant
0.709  0.007  Alpha adrenoreceptor antagonist
0.623  0.010  Neurotrophic factor
0.597  0.008  Dopamine antagonist
0.579  0.007  Dopamine D2 antagonist
0.589  0.037  Rhinitis treatment
0.571  0.020  Cognition disorders treatment
0.573  0.038  Cytokine modulator
0.566  0.044  Hypotermic
0.501  0.019  Sodium channel blocker
```

FIGURE 1 The predicted biological activity spectrum for Buspiron. Only activities with $Pa > 0.5$ are shown.

activities (5 Hydroxytryptamine 1 agonist, 5 Hydroxytryptamine 1A agonist, 5 Hydroxytryptamine agonist, Antidepressant, Antiemetic, Antipsychotic, Anxiolytic, Chemoprotective, Hypotermic, Psychosexual dysfunction treatment, Psychotropic) was excluded from the training set during the prediction. Its main activity "Tranquilizer (minor)" known from Ref. [11] corresponds to the activity "Anxiolytic" (position 3 in the prediction list, Pa = 0.946). New prospective activity (Neurotrophic factor) is predicted for Buspirone (20) with Pa = 0.623.

## Prediction of New Activities in Some Drugs from Top 200 List

Figure 2 illustrates the percentage of the known activity distribution *versus* the values of the calculated probability to be active. As can be seen from Figure 2, about 89% of the predicted activities that coincided with the known activities had Pa values exceeding 50% while 63% of these activities had Pa values of more than 70%. However, in many cases additional (previously unknown) activities were predicted with rather high probability, which sometimes exceeded the probability values calculated for the known ones.

For instance, Buspirone (20) (see Fig. 1) has the highest values of probability *Pa* in prediction for the known activities, but there are many exceptions, the examples of which are shown in Figure 3. New prospective effects are likely to be found in some pharmaceuticals from the Top 200 list, in particular: Antiparkinsonian for Sertraline (108); Cognition disorders treatment for Ramipril (104); Multiple sclerosis treatment for Amlodipine (8) and Carisoprodol (21); Bone formation stimulant for Oxaproin (93);

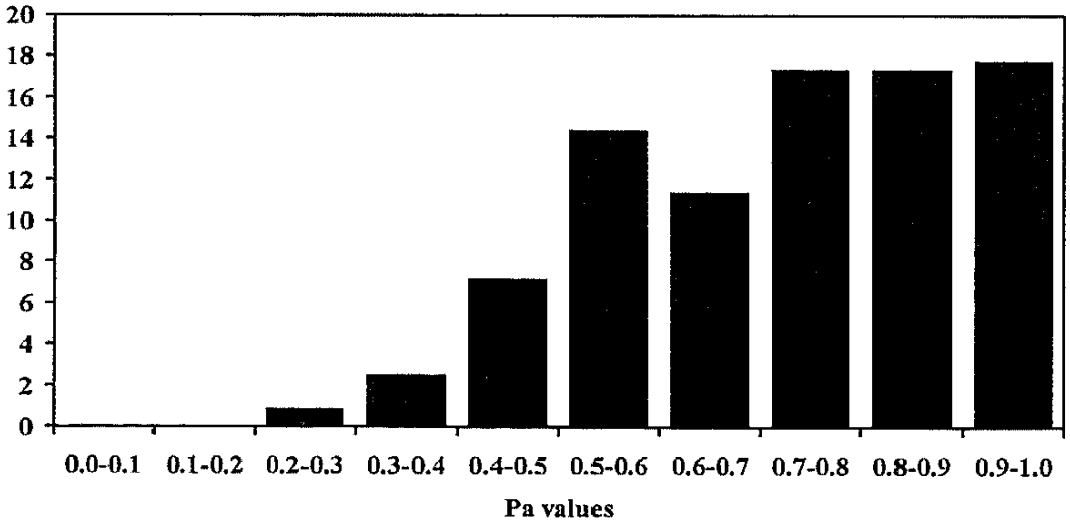
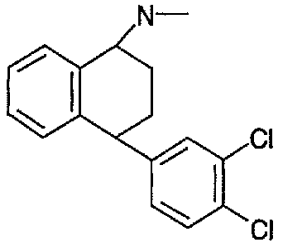
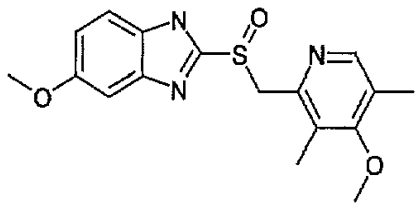


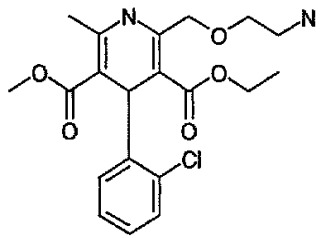

FIGURE 2 Distribution of the known activities *versus* the calculated Pa values.



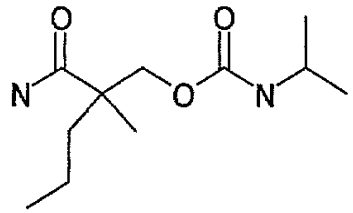
**Sertraline**
Adrenergic transmitter uptake inhibitor (0.770)
Antiparkinsonian (0.609)
Leukopoiesis inhibitor (0.582)
Cocain dependency treatment (0.560)
Acute neurologic disorders treatment (0.541)



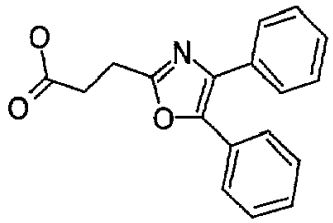
**Omeprazole**
TNF-alpha release inhibitor (0.658)
Atherosclerosis treatment (0.541)



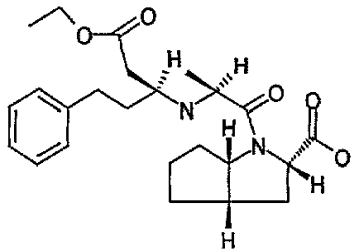
**Amlodipine**
Antineoplastic enhancer (0.608)
Multiple sclerosis treatment (0.508)



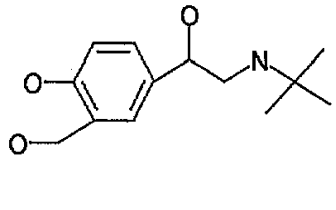
**Carisoprodol**
Angiogenesis inhibitor (0.569)
Multiple sclerosis treatment (0.549)



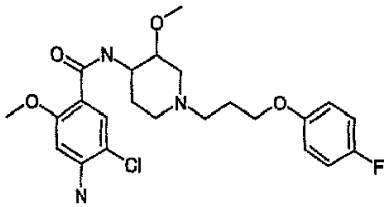
**Oxaprozin**
Bone formation stimulant (0.785)
Interleukin 1 antagonist (0.640)



**Ramipril**
Multiple sclerosis treatment (0.589)
Cognition disorders treatment (0.562)
Antiarthritic (0.454)



**Albuterol**
Antiobesity (0.784)



**Cisapride**
Irritable Bowel syndrome therapy( 0.720)
Rhinitis treatment (0.524)

FIGURE 3 Examples of biological activities predicted de novo for some pharmaceuticals from the Top 200 list, which may become a reason for a new application. Pa values are given in brackets.

Angiogenesis inhibitor for Carisoprodol (21); *etc.* These new predicted activities have to be tested experimentally, because they may become a reason for a new application of some approved drugs.

## Prediction of Side Effects in Some Drugs from Top 200 List

On the other hand, predicted biological activity spectra may include some side and toxic actions for pharmaceuticals from the Top 200 list. These data are summarized in Table II. Among the toxic effects predicted by PASS are teratogenicity and embryotoxicity, carcinogenicity; and among the side effects are abort induction, ulcerogenicity, and cardiotoxicity.

Eighty three percent of the predicted side/toxic effects coincided with the data known from literature.

Just one overprediction of teratogenicity and embryotoxicity was obtained for Prednisone (99); and three overpredictions of cardiotoxic effects were obtained for Cyclobenzaprine (33), Tramadol (118) and Amitriptyline (6). Taking into account that the equivalent structures and their biological activities were removed from the training set during the prediction, one

TABLE II Prediction of side and toxic actions for some drugs from Top 200 list

| *Action* | *Pharmaceuticals* | *Prediction* | *Experiment* |
|---|---|---|---|
| Teratogen | Sertraline | + | + |
| Embryotoxic | Furosemide | + | + |
| | Prednisone | + | + |
| | Norgestimate/Ethyl Estradiol | + | − |
| | Beclomethasone | + | + |
| | Clotrimoxazole/Betamethasone | + | + |
| | Medroxyprogesterone | + | + |
| | Desogestrel/Ethyl Estradiol | + | + |
| | Triamcinolone | + | + |
| | Temazepam | + | + |
| | Estradiol | + | + |
| | Methylprednisolone | + | + |
| | Tobramycin/Dexamethasone | + | + |
| | Mometasone | + | + |
| | Budesonide | + | + |
| | Tetracycline | + | + |
| | Lorazepam | + | + |
| Carcinogenic | Nitrofurantoin | + | + |
| Abort | Propoxyphene N/APAP | + | + |
| Inducer | Tamoxifen | + | + |
| Ulcerogenic | Nabumetone | + | + |
| Cardiotoxic | Cyclobenzaprine | + | − |
| | Tramadol | + | − |
| | Amitriptyline | + | − |

might conclude that the average prediction accuracy is reasonable to recommend PASS use for identification of such effects at the early stages of R & D.

## CONCLUSIONS

(1) The known biological actions of pharmaceuticals from the Top 200 list were predicted with the computer program PASS in 93% of the cases. 83% of predicted side and toxic effects were known for the pharmaceuticals from the Top 200 list previously. Despite the exclusion of each equivalent compound from the training set during the prediction, the mean prediction accuracy is even better than in LOO cross-validation.

(2) High accuracy of PASS predictions provides the evidence that this program can be effectively used in search for new lead compounds with desirable activities. In particular, some new prospective biological actions were predicted for the pharmaceuticals from the Top 200 list, including possible also as: angiogenesis inhibitor, bone formation stimulant, cognition disorders treatment, multiple sclerosis treatment, *etc.*

### *References*

[1] Walker, S. (1994). The future of the Japanese pharmaceutical industry. – How can R & D be really international. *Pharma Japan*, **1424**, 7–10.
[2] Maggon, K. K. and Mechkovski, A. (1992). Total quality management. *Drug News and Perspect*, **5**, 261–570.
[3] Wermuth, C. (1996). *The practice of Medicinal Chemistry*. Academic Press.
[4] Van de Waterbeemd, H. (1996). *Structure-property Correlations in Drug Research*. Academic Press.
[5] Kubinyi, H., Folkers, G. and Martin, Y. C. (1997). *3D QSAR in Drug Design. Vol. II and Vol. III*. Kluwer, The Netherlands.
[6] Filimonov, D. A. and Poroikov, V. V. (1996). PASS: Computerized prediction of biological activity spectra for chemical substances. In: *Bioactive Compound Design: Possibilities for Industrial Use*. BIOS Scientific Publishers, Oxford, pp. 47–56.
[7] Poroikov, V. V., Filimonov, D. A., Stepanchikova, A. V. *et al.* (1996). Opimization of synthesis and pharmacological testing of new compounds based on computerized prediction of their biological activity spectra. *Chim.-Pharm. J.* (*Rus*), **30**(9), 20–23.
[8] [http://www.ibmh.msk.su/PASS]
[9] [http://www.mdli.com]
[10] Poroikov, V. V., Filimonov, D. A., Borodina, Yu. V., Lagunin, A. A. and Kos, A. (2000). Robustness of biological activity spectra predicting by computer program PASS for noncongeneric sets of chemical compounds. *J. Chem. Inf. Comput. Sci.*, **40**(6), 1349–1355.
[11] [http://www.rxlist.com]

[12] Drug Information for Healthcare Professionals. US Pharmacopeia, 2000.
[13] Ghose, A. K., Viswanadhan, V. N. and Wendolovski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.*, 1, 55–68.
[14] Filimonov, D. A., Poroikov, V. V., Borodina, Yu. V. and Gloriozova, T. A. (1999). Chemical similarity assessment through multilevel neighborhoods of atoms: Definition and comparison with the other descriptors. *J. Chem. Inf. Comput. Sci.*, **39**, 666–670.
[15] [http://www.vei.co.uk/chemweb/library/lecture17/slideroom_babaev/transcript.html]