Journal of Computer-Aided Molecular Design, **16:** 819–824, 2002.
KLUWER/ESCOM

How to acquire new biological activities in old compounds by computer prediction

V. V. Poroikov* & D.A. Filimonov
Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Pogodinskaya Street 10, 119121, Moscow, Russia

Key words: lead finding, computer prediction, PASS, biological activity spectra, drug efficacy and safety, new drug indication

Summary

Due to the directed way of testing chemical compounds' in drug research and development many projects fail because serious adverse effects and toxicity are discovered too late, and many existing prospective activities remain unstudied. Evaluation of the general biological potential of molecules is possible using a computer program PASS that predicts more than 780 pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, etc. on the basis of structural formulae of compounds, with average accuracy ~85%. PASS applications to both databases of available samples included hundreds of thousands compounds, and small collections of compounds synthesized by separate medicinal chemists are described. It is shown that 880 compounds from Prestwick chemical library represent a very diverse pharmacological space. New activities can be found in existing compounds by prediction. Therefore, on this basis, the selection of compounds with required and without unwanted properties is possible. Even when PASS cannot predict very new activities, it may recognize some unwanted actions at the early stage of R&D, providing the medicinal chemist with the means to increase the efficiency of projects.

Introduction

The major reasons for failures in drug R&D are: (1) low efficacy, (2) non-safe pharmacology and toxicity, and (3) non-appropriate pharmacokinetic properties [1]. Due to the limitations in financial and time expenses, testing facilities, and incomplete knowledge about possible macromolecular targets, the compounds (hits and leads) are usually tested in directed mode: against a certain target, for a particular disease treatment. As a consequence, many projects fail because of the late discovery of dangerous adverse effects and toxicity, and also many prospective biological activities are not found at the early stages of drug discovery process. Sometimes, new actions of old compounds are found during the clinical trials or practical use of medicine, and that becomes a reason for new indication of a drug. Examples are: *acetazolamide*, that was first launched as a diuretic and later on as an antiepileptic agent; *levamisole*, currently used as antihelmintic and immunomodulator; *sildenafil (viagra)*, formerly studied as a remedy with antihypertensive action but now is it widely used for male sexual dysfunction treatment; and others. Conversely, sometimes the launched pharmaceuticals are removed from the market because serious adverse effects are found during their medical use. Such cases may damage the reputation of a company and decrease the cost of its shares at the stock exchange.

In general, finding of new leads among existing drugs provides significant advantages for a pharmaceutical company because their general pharmacology, toxicity and pharmacokinetic properties are already studied in more detail [2, 3].

In the present study we have investigated the possibilities of utilizing computer-aided prediction to estimate the general biological potential of molecules under study. Such prediction might significantly increase the chance of selecting candidates with desirable but

*To whom correspondence should be addressed. E-mail: vvp@ibmh.msk.su

without unwanted effects, thus changing the remedy discovery strategy from 'Fail early, fail fast' to 'Think first, predict, then measure, and not fail' [1].

Methods

Estimation of general biological potential for drug-like compounds on the basis of their structural formulae can be performed with a computer program PASS (Prediction of Activity Spectra for Substances) that predicts more than 780 pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity [4–6]. PASS prediction is based on the SAR analysis of the training set, including 45,466 compounds with experimentally established biological activity spectra. Biological activity is presented in **PASS** qualitatively ('yes' or 'none'), which is explained in particular by the necessity to use information from different sources when forming the training set. For the description of chemical structure in PASS we developed original descriptors called the Multilevel Neighborhoods of Atoms (MNA) [7]. It was shown that MNA descriptors are rather universal to represent various structure–property relationships [6–9]. The mathematical approach used in PASS [6–9] was selected by a special comparison of the quality of prediction from many different methods [10]. In leave one out cross-validation (LOO CV) for PASS 1.608 the average accuracy through all 45,466 compounds of the training set and 783 kinds of biological activity is about 85% [6]. It was also shown that, despite the incompleteness of data in the training set, PASS provides a reasonable accuracy of SAR analysis and predictions [11].

A detailed description of the general PASS approach and opportunity for free prediction is available via the Internet [6]. Using MOL or SD files as an input for the PASS program, one can obtain a list of probable biological activities for any drug-like molecule as an output. For each activity P_a and P_i values are calculated, which can be interpreted either as the probabilities of a molecule belonging to the classes of active and inactive compounds respectively, or as the probabilities of the first and second kind of errors in prediction.

PASS predictions for Top 200 drugs coincided with the known pharmacotherapeutic effects in 93.2% of cases and with the known side-effects in 83.0% of cases [3]. The analysis of PASS predictions for 42,689 compounds in comparison with the results of anti-HIV screening by NCI (NIH, U.S.) demonstrates that using PASS results it is possible to enrich the population of 'actives' in the initial set of compounds from ~2 to ~17 times (at the thresholds $P_a = 10\%$ and $P_a = 90\%$ respectively) [12]. Therefore, PASS can be effectively applied to analyze the large databases of available samples [13–17], as well as to any drug-like compounds synthesized by organic and medicinal chemists. Both types of application are considered below.

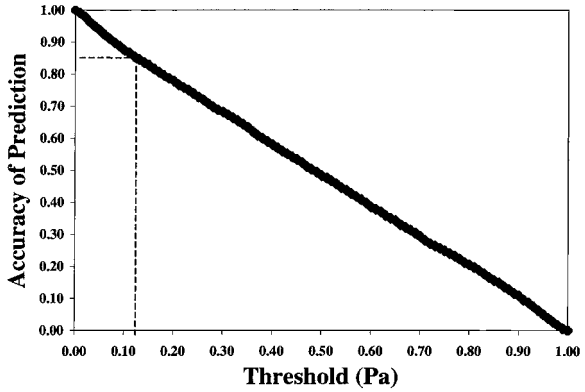


Figure 1. Accuracy of PASS predictions versus different Pa threshold values.

Results

PASS application to Prestwick's compounds

While the number of compounds in the existing libraries constitutes hundreds of thousands [13–16], for the analysis we selected a relatively small collection from Prestwick Chemicals [17]. According to [17], these 880 carefully selected compounds are highly diverse in structure and cover many therapeutic areas – from neuropsychiatry to cardiology, immunology, anti-inflammatory, and more. Over 85% of these compounds are marketed drugs. Therefore, by comparing PASS predictions with the known activities, an additional validation of the program can be performed. This data presented in Figure 1 shows that the average accuracy of prediction approximately corresponds to the one obtained in LOO cross-validation: 85% of the known activities are predicted correctly if $P_a > 12\%$ threshold is chosen.

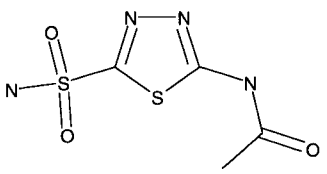
We have verified whether the Prestwick chemical library is really diverse in pharmacological space, as it is stated in [17]. To do so, the statistics of prediction

Table 1. A number of compounds from Prestwick chemical library whose particular activities are predicted at different thresholds.

No	Pa > Pi	Pa > 30%	Pa > 50%	Pa > 70%	Type of Activity
1	382	309	125	31	Toxic
2	357	295	159	63	Spasmolytic
3	318	318	140	46	Nootropic
4	316	268	108	36	Emetic
5	313	313	180	119	Arrhythmogenic
6	312	312	189	69	Neuroprotector
7	311	267	107	37	Antitoxic
8	311	311	148	40	Myocardial ischemia treatment
9	305	305	99	10	Histamine release stimulant
10	303	303	176	25	Fibrinolytic
11	301	301	152	43	Psychosexual dysfunction treatment
12	296	219	101	47	Cardiodepressant
13	295	291	88	24	Male reproductive disfunction treatment
14	295	188	70	12	Embryotoxic
15	294	219	82	22	Antisecretoric
16	291	291	149	60	Lipid metabolism regulator
17	284	250	122	28	Immunosuppressant
18	284	284	73	3	ATPase inhibitor
19	283	220	102	41	Spasmogenic
20	281	281	120	38	Antileishmanial
21	281	281	135	87	Convulsant
22	280	280	102	7	Calmodulin antagonist
23	278	278	103	10	Multiple sclerosis treatment
24	278	219	89	38	Antiasthmatic
				...	
779	13	1	1	0	Prostaglandin F2 alpha agonist
780	10	1	1	1	Antibiotic Rifamycin-like
781	10	1	0	0	Antibiotic Carbapenem-like
782	9	1	0	0	Renin inhibitor

results was analyzed with the computer program PharmaExpert [18]. Some results are presented in Table 1. The data demonstrate that 782 from 783 activities, which can be predicted by PASS [6], are found in the Prestwick compounds at the threshold $P_a > P_i$. The activity 'Urokinase-type plasminogen activator receptor antagonist' is the only one not predicted for any of the 880 Prestwick compounds. Even if we increase the threshold value up to $P_a > 30$, 50 and 70%, the number of activities that are found within the predicted activity spectra for the Prestwick compounds is still significant: 711, 598, and 467, respectively. At $P_a > 50\%$ the most frequent activity is neuroprotector that is predicted for 189 compounds; the next are arrhythmogenic, fibrinolytic, spasmolytic, psychosexual dysfunction treatment, lipid metabolism regulator, etc. (180, 176, 159, 152, 149 compounds, respectively). All major pharmacotherapeutic areas are fairly well covered by the compounds from the Prestwick chemical library. Therefore, this collection really represents molecules diverse in pharmacological space. However, some compounds are predicted as having adverse and toxic effects (arrhythmogenic, convulsant, toxic, spasmogenic, teratogen, etc.), and such predicted unwanted effects have to be also considered during the selection of the most prospective hits.

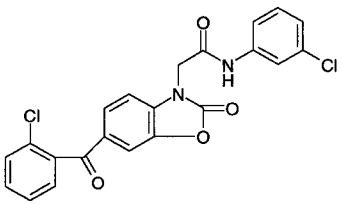
Results of prediction for acetazolamide (Figure 2) illustrate that known biological activities are well predicted. Some new biological activities are also predicted for this molecule: antiarthritic ($P_a = 67.9\%$), bone formation stimulant ($P_a = 59.7\%$), and others. Therefore, acetazolamide may become a new hit or



28 substructure descriptors; 0 new.
12 of 783 possible activities at Pa>40%.

Pa	Pi	Activity
0.901	**0.004**	**Ophthalmic drug**
0.848	**0.003**	**Antiglaucomic**
0.834	**0.001**	**Carbonic anhydrase inhibitor**
0.807	**0.004**	**Diuretic**
0.679	*0.018*	*Antiarthritic*
0.638	**0.004**	**Diuretic inhibitor**
0.597	*0.008*	*Bone formation stimulant*
0.571	**0.003**	**Saluretic**
0.557	0.003	Electrolyte absorption antagonist

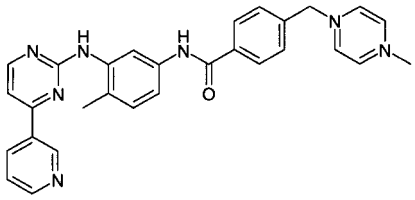
Figure 2. Predicted activity spectrum for acetazolamide for $P_a > 50\%$: activities known from literature are marked in bold, activities that may become a reason for new applications are marked in italic.



43 substructure descriptors; 0 new.
13 of 783 possible activities at Pa>50%.

Pa	Pi	Activity
0.880	**0.007**	**Analgesic, non-opioid**
0.783	*0.016*	*Antiepileptic*
0.753	*0.036*	*Neuroprotector*
0.693	**0.010**	**Analgesic**
0.664	*0.037*	*Nootropic*
0.630	*0.019*	*Anticonvulsant*
0.618	*0.012*	*Alzheimer's disease treatment*
0.578	*0.013*	*Acaricide*
0.586	**0.030**	**Antiinflammatory**
0.555	*0.010*	*Prostaglandin antagonist*
0.547	*0.032*	*Sedative*
0.512	*0.015*	*Uricosuric*
0.502	*0.021*	*Muscle relaxant*

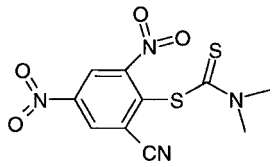
Figure 3. Predicted activity spectrum for the compound (1) from [17] at $P_a > 50\%$: activities known from the literature are marked in bold, activities that may become a reason for new applications are marked in italic.



52 substructure descriptors; 2 new.
52 of 783 possible activities at Pa > Pi.

Pa	Pi	Activity
0.558	0.031	Interleukin 1 antagonist
0.552	0.042	Interleukin antagonist
0.542	0.032	Restenosis treatment
0.496	0.007	Protein kinase inhibitor
0.548	0.087	Myocardial ischemia treatment
0.468	0.007	P38 MAP kinase inhibitor
0.557	0.100	Multiple sclerosis treatment
0.457	0.035	Antihelmintic (nematodes)
0.472	0.083	Angiogenesis inhibitor
0.441	0.118	Cytokine modulator
0.326	**0.008**	**Tyrosine kinase inhibitor**
0.379	0.077	Rhinitis treatment
	...	

Figure 4. Predicted activity spectrum for Gleevec. Activities known from the literature [18] are marked in bold.



31 substructure descriptors; 0 new.
100 of 783 possible activities at Pa > Pi.

Pa	Pi	Activity
0.807	0.005	Hepatoprotectant
0.804	0.006	Hepatic disorders treatment
0.799	0.006	Mediator release inhibitor
0.684	*0.020*	*Toxic*
0.619	*0.018*	*Embryotoxic*
0.604	0.011	Acaricide
0.586	*0.007*	*Mutagenic*
	...	
0.200	**0.148**	**Antituberculosic**
	...	

Figure 5. Predicted activity spectrum for compound published in [19]. Activity known from the literature is marked in bold; predicted toxicity is marked in italic.

even lead for the appropriate pharmacotherapeutic applications. Since no significant adverse/toxic effects are predicted for this molecule even at the threshold $P_a > P_i$, the compound looks very prospective for further testing in the appropriate assays.

PASS applications to small collections of compounds

Even for relatively small number of compounds, synthesized or only designed for the synthesis by some organic/medicinal chemists, PASS predictions often provide valuable information about the biological potential of these molecules. Let us consider several examples that illustrate how to use the results of prediction.

The first example is a compound recently disclosed at the XVIIth International Symposium of Medicinal Chemistry [19]. Its structure and predicted biological activity spectra are presented in Figure 3. It is clear that PASS predicts well the known biological activities (analgesic and anti-inflammatory [19]). New prospective activities are predicted too. Some of them are associated with psycho- and neuropharmacological applications (neuroprotector, antiepileptic, nootropic, anticonvulsant, etc.), while the others, like acaricide, may point at quite different use of this chemical series. It is important to emphasize that from 783 biological activities predicted by PASS only 13 are predicted at $P_a > 50\%$, therefore, (1) PASS predictions significantly reduce the pharmacological space for further testing and (2) one may expect that this molecule will exhibit a relative selectivity.

Another example is presented in Figure 4. This is an antitumor drug Gleevec recently launched by Novartis [20]. Its mechanism of action (tyrosine kinase inhibitor) is predicted with $P_a = 32.6\%$. Such rather moderate probability can be explained by the relative novelty of this molecule: it has two MNA descriptors that are not presented in any of the 45,466 compounds from the training set. However, some new activities are predicted for this compound including interleukin antagonist ($P_a = 55.2\%$), restenosis treatment ($P_a = 54.2\%$), multiple sclerosis treatment ($P_a = 55.7\%$), etc. These activities, if confirmed by the experiment, might lead to new indications for Gleevec.

In some cases a compound under study is very novel in relation to a particular activity: although all its descriptors are found in the compounds from the training set, this activity may not be well predicted by PASS. Such example is shown in Figure 5. For this compound a potent antituberculosis activity was found in the experiment [21], but it is poorly predicted ($P_a \approx P_i$ for this activity) because of high novelty of this molecule as an antituberculosis agent. However, for medical application of this compound PASS prediction recognizes possible problems: toxic action is predicted with $P_a = 68.4\%$, embryotoxic with $P_a = 61.9\%$, mutagenic with $P_a = 58.6\%$, etc. Therefore, even when PASS cannot predict useful pharmacotherapeutic effects, it may recognize some unwanted actions at the early stage of R&D, directing the conclusion 'Fail early, fail fast'. Such application of PASS is possible even for new targets for which we are yet unable to create the training set because of too small a number of known ligands.

Conclusions

In many cases the computer program PASS predicts with reasonable accuracy new biological activities in compounds under study that may provide the reasons either for new medical applications of compounds or, if unwanted effects and toxicity are predicted, for the refusal of such molecules at the early stages of R&D. Since PASS predictions for ~10,000 compounds take a few minutes in a normal PC, PASS can be effectively used for selection of compounds with the required and without unwanted effects in databases of available samples including hundreds of thousands chemical compounds.

However, it is necessary to keep in mind: PASS cannot predict whether a compound becomes a drug, it provides the 'food for thought' for the medicinal chemist.

References

1. Han van de Waterbeemd. In Abstr. XVIIth Symp. Med. Chem., Barcelona, 1–5 Sept. 2002, L10. 2. Wermuth, C.G. Med. Chem. Res. 10 (2001) 431.
2. Wermuth, C.G. Med. Chem. Res. 10 (2001) 431.
3. Poroikov, V., Akimov, D., Shabelnikova, E. and Filimonov, D. SAR and QSAR in Environm. Res. 12 (2001) 327.
4. Poroikov, V. and Filimonov, D., In Kartsev, V.G. and Tolstikov, G.A. (Eds.)., Nitrogen-containing heterocycles and alcaloides, Iridium Press, Moscow, 2001, 1, pp. 149–154.
5. Poroikov, V. and Filimonov, D. In Holtje, H.-D. and Sippl, W. (Eds.), Rational Approaches to Drug Design, Prous Science, Barcelona, 2001, pp. 403–407.
6. http://www.ibmh.msk.su/PASS
7. Filimonov, D., Poroikov, V., Borodina, Yu. and Gloriozova, T.J. Chem. Inf. Comput. Sci., 39, (1999) 666.
8. Anzali, S., Barnickel, G., Cezanne, B., Krug, M., Filimonov, D. and Poroikov, V. J. Med. Chem. 44 (2001) 2432.
9. Lagunin, A., Stepanchikova, A., Filimonov, D. and Poroikov, V. Bioinformatics. 16 (2000) 747.
10. Filimonov, D. Abstr. II Rus. Natl. Congr. 'Man and Drugs', Moscow, 1995, pp. 62–63.
11. Poroikov, V., Filimonov, D., Borodina, Yu., Lagunin, A. and Kos, A.J. Chem. Inform. Comput. Sci. 40 (2000) 1349.

12. Poroikov, V.V., Filimonov, D.A., Ihlenfeldt, W.-D., Gloriozova, T.A., Lagunin, A.A., Borodina, Yu.V., Stepanchikova, A.V. and Nicklaus, M.C. J. Chem. Inform. Comput. Sci. 43 (2003) 228.
13. http://www.chembridge.com
14. http://www.asinex.com
15. http://www.specs.net
16. http://www.akosgmbh.de
17. http://www.prestwickchemical.com
18. Lagunin, A.A., Filimonov, D.A. and Poroikov, V.V. In Abstracts of 6th International Conference on Chemical Structures, Noordwijkerhout. 2002, pp. 25–26.
19. Banoglu, E., Okcelik, B., Kupeli, E., Unlu, S., Yesilada, E. and Sahin, M.F. In Abstr. XVIIth Symp. Med. Chem., Barcelona, 1–5 Sept. 2002, P127.
20. Zimmermann, J., Buchdunger E. and Mantley P. In Abstr. XVIIth Symp. Med. Chem., Barcelona, 1–5 Sept. 2002, L19.
21. Makarov, V.A. and Mollmann, U. In Abstr. XVIIth Symp. Med. Chem., Barcelona, 1–5 Sept. 2002, P.351.