The Chemistry and Biological Activity
of Nitrogen-Containing Heterocycles and Alkaloids

# NITROGEN-CONTAINING HETEROCYCLES AND ALKALOIDS

*V.G. Kartsev, D. Sc. (Chem.) and G.A. Tolstikov, Acad. (Editors)*

**Volume 1**

Iridium Press
Moscow
2001

# Computer-assisted prediction of biological activity in a search for and optimization of new drugs

Poroikov V.V., Filimonov D.A.

*Orekhovich Institute of Biomedical Chemistry, Russian Academy of Medical Sciences, ul. Pogodinskaya 10, Moscow, 119832 Russia*
*tel.: +7 (095) 245-2753, e-mail: vvp@ibmh.msk.su*

Nowadays, a search for and design of new pharmaceuticals are carried out largely in a purpose-oriented manner: chemical compounds are being tested only for some required types of biological activity, and the properties of revealed structures are further optimized upon synthesis and characterization of their analogs. In this case, some kinds of biological activity—typical of a compound under study but being "side" relative to the adopted line search—remain non-tested. On the other hand, versatile biological activity of typical bioactive compounds is commonplace. Some kinds of bioactivity sometimes manifest themselves as side or toxic effects while others may give grounds for new pharmaceutical applications. For instance, **acetazolamide** was launched as a diuretic in 1954 and then as antiepileptic in 1956; **levamisole** was launched as antihelintic in 1968 and as immunostimulant in 1980; **alprostadil** was launched as antiagregant in 1988 and as erectant in 1994; analgesic action of **aspirin** was discovered in 1899 but its antiaggregant activity was found only in 1971, etc. Therefore, there exists some contradiction between the directed testing of biologically active compounds and a multiplicity of biological actions potentially exhibited by any substance. None of chemical compounds can be experimentally tested for all known types of acivity [1]. Even highthroughput screening does not solve the problem because screening is also carried out in a directed mode toward one or several biological targets of potential drug considered as promising within a given period of time [2]. The only real possibility for complex evaluation of the biological activity of chemicals is to develop new techniques of computer-aided prediction and to apply them for estimation of probable types of biological activity for chemical compounds with the subsequent testing according to the predictions. The most of currently available methods for molecular modeling and structure–activity relationship (SAR) analysis are applied to study the ligand–receptor interaction (one biological target) and for optimization of compounds properties on the basis of SAR for the same chemical class.

Our work aims at development of a computer system that will predict all types of biological activity based a structural formula and using some standard description of chemical structure and a universal mathematical algorithm for SAR analysis. The attempts to develop such a computer system have been made earlier [3–7]. In particular, the feasibility for computer-aided prediction of the biological activity of chemical compounds on the basis of their structural formulae has been studied within the framework of the State System for Registration of New Chemical Compounds Synthesized in the USSR for many years [8]. For some objective and subjective reasons, this problem was not completely solved then, but the obtained data gave a basis for development of such a computer system in the future [9–12].

## Basic principles

We suggested a concept of the spectrum of biological activity as a set of all kinds of biological activity exhibited by a given compound [9, 10]. Accordingly, the biological activity is regarded as an intrinsic property of compound that depends only on its structure. Any 'component' of the activity spectrum of a given compound can be detected under particular experimental conditions (different for various types of activity), the magnitude of a given property strongly depending on experimental conditions.

Based on this concept, we developed the computer system **PASS** (**P**rediction of **A**ctivity **S**pectra for **S**ubstances). This system can predict 565 types of biological activity based only on the structural formula of a chemical compound. The spectra of biological activity involve the main and side pharmacological

effects, biochemical mechanisms of action, mutagenicity, carcinogenicity, teratogenicity, and embryotoxicity [13, 14]. **PASS** analyzes the structure–activity relationship for more than 35 000 compounds from the training set (launched pharmaceuticals and pharmacologically active compounds). The training set is permanently updated with new data taken from both published sources and electronic databases [15], including un-published "grey" information [16].

Biological activity is presented in **PASS** qualitatively ("yes" or "none"), that is explained in particular by the necessity of using information from different sources when forming the training set [13]. For description of chemical structure in **PASS**, we developed original descriptors called the Multilevel Neighbourhoods of Atoms (MNA) [17]. It was shown that the MNA descriptors are rather universal to represent various structure–property relationships [17]. The mathematical approach used in **PASS** [14] was selected by comparing the predicting ability of about a hundred of different methods [18]. It was shown that the adopted approach provides obtaining the statistically robust structure–activity relationships and predictions [19]. It is extremely important because all data included into the training set are always incomplete as concerning the structure biological behavior (not all individual compounds were tested).

We estimated the quality of prediction in leave-one-out cross-validation tests (LOO CV). In this procedure, each compound is removed from the training set and its biological activity is predicted on the basis of the rest of the training set. The results are compared with experimental data, and the accuracy of prediction averaged over ca. 35 000 compounds of the training set, and 565 types of biological activity are calculated. It was shown that the mean accuracy of prediction is about 85% [14] which is sufficient for practice (mean accuracy for random prognosis is only about 0.2%—for 500 types of biological activity).

The results of prediction are presented as a list of probable types of activity with calculated probabilities of being active ($P_a$) and inactive ($P_i$). The $P_a$ and $P_i$ values range between 0 and 1 (Figure 1). As $P_a$ and $P_i$ are calculated independently, their sum is not equal to 1.

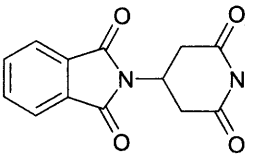
## Applications of pass

The results of **PASS** predictions that are given below were obtained with different versions of the system in past ten years, with permanent control of prediction quality which was found higher for the latest versions of **PASS**.

Comparison of computer-aided prediction of biological activity with the expert estimates (specialists in medicinal chemistry and pharmacology) shows that the computer system provides three times more accurate prediction, with a markedly lower number of cancellations [20]. The economic effectiveness of using PASS in screening exceeds 500% [21]. Blind testing of PASS on a heterogeneous set of 118 compounds from different chemical classes with 58 types of biological activity was carried out by nine scientists from eight countries (Russia, Ukraine, Belarus', UK, Belgium, Italy, Greece, Portugal). The mean accuracy of prediction was found to be 82.6% [22]. It was also ound [23] that the mean accuracy of predicting mutagenicity and carcinogenicity (in order to rule out potentially hazardous compounds at the earlier stages of research) was found to be fairly high.

Currently, the databases (DB) of chemical compounds synthesized earlier and available for screening include hundred thousands of substances (DB of SPECS & BioSPECS have more than 1 200 000 compounds; DB of IBS and ChemDiv have more than 200 000 compounds, etc.). Computer system **PASS** predicts the spectrum of biological activity spectra for 1000 compounds in less than one minute in a personal computer with Pentium processor, which makes computer-assisted screening of the available samples convenient for selection of substances with required properties [24].

Computer-aided prediction is applicable to the compounds prepared by traditional organic synthesis, by combinatorial chemistry [25], and even to the small (di, tri- and tetra-) peptides and peptidomimetics [26, 27]. Since the prediction is based on the structural formula of compound, it can be done already at the stage of synthesis planning.



<ACTIVITY_PREDICTION>
24 Substructure descriptors; 0 new
84 Possible activities

| $P_a$ | $P_i$ | Activity |
|---|---|---|
| 0.781 | 0.006 | **Cytokine modulator** |
| 0.713 | 0.019 | **Sedative** |
| 0.678 | 0.030 | ***Cardiovascular analeptic*** |
| 0.656 | 0.015 | **Angiogenesis inhibitor** |
| 0.439 | 0.007 | ***Neurokinin antagonist*** |
| 0.435 | 0.008 | ***Calpain inhibitor*** |
| 0.433 | 0.009 | ***Oxytocin antagonist*** |
| 0.443 | 0.024 | ***Chemoprotective*** |
| 0.421 | 0.011 | **Tumour necrosis factor antagonist** |
| 0.398 | 0.007 | **Hypnotic** |
| 0.439 | 0.050 | NMDA agonist |
| 0.407 | 0.028 | Bronchodilator |
| 0.430 | 0.059 | Psychotropic |
| 0.417 | 0.054 | **Anxiolytic** |
| 0.370 | 0.007 | Protein kinase C inhibitor |
| 0.428 | 0.068 | Anticonvulsant |
| 0.421 | 0.062 | **Teratogen** |
| 0.361 | 0.008 | Antidiabetic symptomatic |
| ... | ... | ... |

**Fig. 1.** Structural formula and a part of the predicted biological activity spectrum for **thalidomide** (measured activities are marked bold; newly predicted activities recommended for further testing are marked italic).

Some pharmaceuticals are administered in the form of prodrugs. The predicted spectra of the biological activity of prodrugs and their active metabolites imply that the structural formula of drug precursor is sufficient (by 74%) in order to predict the biological action of a given drug [28]. For numerous pharmaceuticals widely used in medical practice, the computer system **PASS** predicts new kinds of activity that seem reasonable to test in experimental and clinical practice [29–32]. In particular, prediction of biological activity spectra for top 200 drugs used in the USA not only coincided with the known effects and mechanisms of action in 93% cases but also indicated new probable applications of some known drugs, in particular, angiogenesis inhibition by myorelaxant drug **carisoprodol**, cognition disorders treatment by antihypertensive drug **ramipril**, multiple sclerosis treatment by antihypertensive drug **amlodipin**, etc. [31–32].

In postgenomic era, the number of known molecular targets is increasingly growing [2]. It can be expected that new targets will not always find appropriate ligands (inhibitors, activators, agonists, antagonists) in amounts sufficient for updating the training set of **PASS**. To overcome the problem, we developed a method for direct estimation of chemical similarity by using the computer program **SIMEST**. This method provides the means of search for 'analogs' in databases based on the structure of even a single ligand [33, 34]. Applying **SIMEST** to more than 200000 compounds from the ChemBridge database, we indetified the substances whose predicted receptor profiles were experimentally confirmed in 75% cases [35].

Our predictions were experimentally confirmed for different classes of chemicals with versatile activity: anti-arrhythmic [36], anticancer [37, 38], antibacterial [39], hepatoprotective [40, 41], antiamnestic [42], anti-inflammatory, antioxidant and local anestetic [43]. The effectiveness of prediction was convincingly demonstrated by discovery of the antiulcer effect in the compounds that were synthesized as potential diuretics. At Novokuznetsk Chemical & Pharmaceutical Institute, about 300 chemical compounds were synthesized. For 20 of them, the antiulcer effect was predicted with **PASS**. Nine compounds have been synthesized and tested, five of which exhibited the potent antiulcer effect comparable to that of reference antiulcer agents. These compounds exhibit a novel chemical structure,

that is, they are New Chemical Entities (NCE) [44]. In case of screening, the expenses for biological testing could be higher by as much as a factor of 15.

In order to extend the applicability range for **PASS**, we exposed our system on Internet [14, 45–47]. With standrard browsers (Netscape or Internet Explorer), the user can send a chemical structure as a MOL-file to the website [14] and obtain the predicted types of biological activity on the display of his own computer. Hundreds of researchers from Russia, Ukraine, Latvia, the USA, Germany, UK, France, Brazil, and some other countries have already applied to **PASS** to get predictions for the biological activity of several thousand compounds.

To our knowlegde, the computer system **PASS** has no foreign analogs. This is because leading pharmaceutical companies are focused only on several lines of pharmacotherapeutic research, thus restricting themselves to several types of biological activity. The idea of registering all synthesized chemical compounds and selecting the most promising ones by computer-assisted prediction could arise only in the former Soviet Union where all research institutes were under the state control. The potentialities of computer-aided prediction have recently drawn the attention of Dr. Marc Nicklaus, Laboratory of Medicinal Chemistry, National Cancer Institute (NCI, National Institute of Health, USA) who proposed cooperation. We predicted the biological activity spectra for 250 000 compounds registered by NCI. It was found that not only known antitumor mechanisms are predicted but also some previously unknown effects. These results are presented on the server of NCI in Internet [48] and are used by American colleagues to select compounds with required types of biological activity.

Since the predicted spectra of biological activity contain the probability estimates for main and side pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity and embryotoxicity, the selection of the most promising compounds among the available samples can be done on the basis of complex criteria. Attention has to be paid to both the presence of required types of biological activity and the absence of undesired side effects and toxicity. Within the **PASS** approach, the problems of finding new leads and optimization of their properties (decrease in side and toxic effects), which are usually solved step by step, are simultaneously fixed at the early stages of research. Moreover, it was shown that the algorithm used in **PASS** can successfully be applied to discriminating the so-called "drug-like" compounds from "drug-unlike" substances [49]. The former result can also be used for "filtering out" compounds with a small probability to become drugs at the early stages of research.

## Conclusions

1. Based on our concept of biological activity spectra as an intrinsic property of chemical structure, we developed the computer program **PASS** which predicts more than 565 probable pharmacological effects and mechanisms of action on the basis of structural formulae of compounds with mean accuracy in leave-one-out cross-validation about 85%.
2. The computer system **PASS** is being used in the process of finding and optimization of new leads. In particular, the computer prediction allowed us to discover new substances with the antiulcer, hepatoprotective, antitumor, antibacterial and antiamnestic activity.
3. In case of essentially new targets, when the number of known ligands is insufficient to create the training set for **PASS**, a search for new ligands can be carried out by using the computer system SIMEST which calculates the 'similarity' chemical compounds.
4. Computer-aided prediction of biological activity spectra allows us to decide which tests have to be used to estimate more precisely the efficiency and safety of given compounds.

*This work was partially supported by (a) the Russian Ministry of Industry, Science, and Technology (project 'Computer-Aided Drug Discovery, supervisor Acad. A.I. Archakov; sub-program 'Research and Development of new Pharmaceuticals by Chemical and Biological Synthesis', supervisor Acad. P.V. Sergeev); (b) program 'Civil Research and Development of Primary Priority'), and (c) the Civil Research and Development Foundation (USA) (grant no. RC1-2064).*

1. G.M. Barenboim, A.G. Malenkov, *Biologicheski aktivnye veshchestva: Novye printsipy poiska* (Biologically Active Compounds: New Principles of Search), Moscow: Nauka, 1986.
2. A.I. Archakov, V.V. Poroikov, *Bioinformatika: Novye biomeditsinskie tekhnologii* (Bioinformatics: New Biomedical Technologies, Moscow: Institute of Biomedical Chemistry, 1999.
3. V.V. Avidon, *Khim.-Farm. Zh.*, (7) 22–25 (1974).
4. V.E. Golender, A.B. Rozenblit, *Avtomatika Telemekhanika* (11) 99–105 (1974).
5. V.V. Avidon, V.S. Arolovich, S.P. Kozlova, L.A. Piruzyan, *Khim.-Farm. Zh.* (5) 88–92 (1978).
6. A.B. Rozenblit, V.E. Golender, *Logiko-kombinatornye metody v konstruirovanii lekarstv* (Logical and Combinatorial Methods in Drug Design), Riga: Zinatne, 1984.
7. A.N. Kochetkov, P.M. Vasiliev, A.G. Breslaukhov, *Abstr. I All-Union Conf. on Theoretical Organic Chem.*, Volgograd, 1991, vol. 2, p. 500.
8. Yu.V. Burov, L.V. Korol'chenko, V.V. Poroikov, *Byull. VNTs BAV* (1) 4–25 (1990).
9. D.A.Filimonov, V.V.Poroikov, E.I. Karaicheva, *et al.*, *Eksp. Klinich. Farmakologiya* **58** (2) 56–62 (1995).
10. V.V.Poroikov, D.A. Filimonov, in *QSAR and Molecular Modelling Concepts, Computational Tools and Biological Applications*, Barcelona: Prous Science Publishers, 1996, p. 49–50.
11. D.A. Filimonov, V.V. Poroikov, in *Bioactive Compound Design: Possibilities for Industrial Use*, Oxford: BIOS Scientific Publishers, 1996, pp. 47–56.
12. V.V. Poroikov, D.A. Filimonov, A.V. Stepanchikova, *et al.*, *Khim.-Farm. Zh.* **30** (9) 20–23 (1996).
13. T.A. Gloriozova, D.A. Filimonov, A.A. Lagunin, V.V. Poroikov, *Khim.-Farm. Zh.*, **32** (12) 32–39 (1998).
14. Available at http://www.ibmh.msk.su/PASS
15. V.V. Poroikov, *Komp'yut. Tekhnol. Meditsine* (2) 91–96 (1998).
16. V.V. Poroikov, in *SCRIP Russian Pharmaceutical Market Guide*, Richmond (UK): PJB Publications, 1997, pp. 159–203.
17. D. Filimonov, V. Poroikov, Yu. Borodina, T. Gloriozova, *J. Chem. Inf. Comput. Sci.* **39** (4) 666–670 (1999).
18. D.A. Filimonov, *Abstr. II All-Rusia Congr. "Man and Drugs"*, Moscow, 1995, pp. 62–63.
19. V.V. Poroikov, D.A. Filimonov, Yu.V. Borodina, A.A. Lagunin, A. Kos. *J. Chem. Inf.. Comput. Sci.* **40** (6) 1349–1355 (2000).
20. V.V. Poroikov, D.A. Filimonov, A.P. Boudunova, *Automatic Documentation and Mathematical Linguistics* **27** (3) 40–43 (1993).
21. V.V. Poroikov, D.A. Filimonov, A.P. Boudunova, *Abstr XIV Int. Symp. on Medicinal Chemistry*, Maastricht (the Netherlands), 1996, P-3.05.
22. Available at http://www.vei.co.uk/chemweb/library/lecture17/slideroom_babaev/transcript.html
23. A.P. Suchkov, D.A. Filimonov, A.V. Stepanchikova, V.V. Poroikov, *Abstr. XI Europ. Symp. on Quantitative Structure–Activity Relationships: Computer-Assisted Lead Finding and Optimization*, Lausanne, 1996, P-32C.
24. V.V. Poroikov, D.A. Filimonov, *Abstr. Int. Workshop "Virtual Screening: New Approaches in Drug Design & Discovery"*, Marburgh (Germany), 1999, p. 58.
25. V.V. Poroikov, D.A. Filimonov, *Abstr. Int..Conf. on Organic Synthesis and Combinatorial Chemistry*, Moscow, 1999, p.L-16.
26. N.B. Martynova, D.A. Filimonov, V.V. Poroikov, *Bioorg. Khim.* **26** (5) 330–339 (2000).
27. V.V. Poroikov, in *Peptidy v kardiologii* (Peptides in Cardiology), ed. O.A. Gomazkov, Moscow: Materik Alpha, 2000, pp. 132–143.
28. Yu.V. Borodina, D.A. Filimonov, V.V. Poroikov, *Khim.-Farm. Zh.* **30** (1) 39–42 (1996).
29. V.V. Poroikov, *Abstr. All-Russia Conf. on Current Problems in Exp. Clinical Pharmacology*, St. Petersburg, 1999, pp. 166–167.
30. V.V. Poroikov, *Khimiya v Rossi* (2) 8–12 (1999).
31. V. Poroikov, D. Filimonov, D. Akimov, *Newslett. QSAR Modelling Soc.* (11) 19–22 (2000).
32. V. Poroikov, D. Akimov, E. Shabel'nikova, D. Filimonov, SAR & QSAR in Envir. Res., in press (2001).
33. Yu. Borodina, D. Filimonov, V. Poroikov, *Quant. Struct.–Activ. Relation.* **17** (5) 459–464 (1998).
34. Yu. Borodina, D. Filimonov, V. Poroikov, *Abstr. XII Europ. Symp. Quant. Struct.–Activ. Relation*, Copenhagen, 1998, P114.

35. Yu. Borodina, *Cand. Sci. (Biol.) Dissertation*, Moscow, Institute of Biomedical Chemistry, 1998.
36. H. Sergeeva, V. Poroikov, *Abstr. XI Noordwijkerhout-Camerino Symp. on Trends in Drug Research*, 1997, p.47.
37. A.V. Pogrebnyak, V.V. Poroikov, V.V. Starykh, D.A. Konovalov, *Rastit. Resursy* **34** (1) 61–64.
38. M.K. Islyaikin, E.A. Danilova, E.V. Kudrik, R.P. Smirnov, A.P. Boudunova, A.S. Kinzirskii, *Khim.-Farm. Zh.* **31** (8) 19–22 (1997).
39. D.A. Maiboroda, E.V. Babaev, L.V.Goncharenko, *Khim.-Farm. Zh.* **32** (6) 24–28 (1998).
40. S.I. Sal'nikova, V.V. Poroikov, I.A. Mazur, *Abstr. of Conf. on Perspectives in Development and Production of Pharmaceuticals in Ukraine*, Odessa, 1993, pp. 101–102.
41. O.B. Flekhter, L.T. Karachurina, V.V. Poroikov, *et al.*, *Bioorg. Khim.* **26** 3, 215–223 (2000).
42. D.A. Filimonov, V.V. Poroikov, A.P. Boudunova, A.V. Rudnitskih, Yu.V. Burov, *Abstr. SCI Conf. on Design of Bioactive Compounds*, Potsdam (Germany), 1995, p. 26.
43. Geronikaki, V. Poroikov, D. Hadjipavlou-Litina, D. Filimonov, A. Lagunin, R. Mgonzo, *Quant. Struct.–Activ. Relationships* **18**, 16–25 (1999).
44. V.A. Trapkov, A.P. Budunova, O.A. Burova, D.A. Filimonov, V.V. Poroikov, *Voprosy Med. Khimii* **43** (1) pp. 41–57.
45. V. Poroikov, D. Filimonov, T. Gloriozova, A. Lagunin, A. Stepanchikova, *II Int. Electronic Conf. on Synthetic Organic Chemistry*, 1998, Poster-e0004.
46. A.A. Lagunin, D.A. Filimonov, A.V. Stepanchikova, V.V. Poroikov, *Abstr. II All-Russia Conf. "Internet and Modern Society"*, St. Petersburg, 1999, p. 65.
47. Lagunin, A. Stepanchikova, D. Filimonov, V. Poroikov, *Bioinformatics* **16** (8) 747–748 (2000).
48. Available at http://cactus.nci.nih.gov/ncidb2/
49. S. Anzali, G. Barnickel, B. Cezanne, M. Krug, D. Filimonov, V. Poroikov, *J. Med. Chem.*, in press (2001).