The analysis of structure–anticancer and antiviral activity relationships for macrocyclic pyridinophanes and their analogues on the basis of 4D QSAR models (simplex representation of molecular structure)[★][✿]

Victor E. Kuz'min[1,2,✉], Anatoly G. Artemenko[2], Victor P. Lozitsky[3], Eugene N. Muratov[2], Alla S. Fedtchouk[3], Natalia S. Dyachenko[5], Lidiya N. Nosach[5], Tatiyana L. Gridina[3], Larisa I. Shitikova[3], Liubov M. Mudrik[3], Aleksey K. Mescheriakov[1], Vladislav A. Chelombitko[1], Andrey I. Zheltvay[1] and Jean-Jaques Vanden Eynde[4]

[1]*I.I. Mechnikov Odessa National University, Odessa, Ukraine;* [2]*A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences of Ukraine, Odessa, Ukraine;* [3]*Ukrainian Mechnikov Research Anti-Plague Institute, Odessa, Ukraine;* [4]*University of Mons-Hainaut, Mons, Belgium;* [5]*D.K. Zabolotny Institute of Microbiology and Virology, Kyiv, Ukraine*

A new 4D-QSAR approach has been considered. For all investigated molecules the 3D structural models have been created and the set of conformers (fourth dimension) have been used. Each conformer is represented as a system of different simplexes (tetratomic fragments of fixed structure, chirality and symmetry).

The investigation of influence of molecular structure of macrocyclic pyridinophanes, their analogues and certain other compounds on anticancer and antiviral (anti-influenza, antiherpes and antiadenovirus) activity has been carried out by means of the 4D-QSAR. Statistic characteristics for QSAR of PLS (partial least squares) models are satisfactory (R = 0.92–0.97; CVR = 0.63–0.83).

Abbreviations: CCM, chorioallantoic membranes; CNS, central nervous system; CVR, cross-validation coefficient; HSV-1, herpes simplex virus, type 1; MCP, macrocyclic pyridinophanes; PLS, partial least squares; SN, synthetic nucleosides.

Molecular fragments increasing and decreasing biological activity were defined. This information may be useful for design and direct synthesis of novel anticancer and antiviral agents.

Prevention and treatment of oncologic diseases and viral pathology are the top problems. The above problems are interdependent, since oncogenicity of some viruses (including adenoviruses and viruses of the herpes group) has been established. We have shown that the well-known medicinal agent ε-aminocapronic acid, which decreases experimental tumours growth and inhibits metastasis dissemination, has the wide spectrum of antiviral activity including anti-influenza, antiadenoviral and antiherpetic actions [1–3].

New type of MCP-macrocyclic Schiff bases and their analogues that contain different structural fragments are very promising. We have established that some of them have anticancer and/or antiviral activity [4–6].

The structure–anticancer activity relationship of MCP within the topologic model was analyzed [4, 7]. The good agreement between observed and predicted values of activity against some cell cultures was observed for regression models. Thus, even within the simple topologic model it is possible to detect some patterns of relationship between structure of MCP and its activity. Assuming that more adequate QSAR models, taking into account a spatial structure of molecules and their conformational variety, should give more reliable results, we used a novel 4D QSAR approach, based on simplex representation of molecular structure.

MATERIALS AND METHODS

Structures of investigated compounds (most of those compounds were previously synthesized and identified by G.L. Kamalov, R.N. Lozytska, D.N. Kryzhanovsky from the A.V. Bogatsky Physico-Chemical Institute of NAS of Ukraine and I.V. Alekseeva, L.I. Palchikovskaya from the Institute of Molecular Biology and Genetics NAS of Ukraine) are presented in Fig. 1.

Anticancer activity of investigated compounds was tested *in vitro* at the National Cancer Institute (Bethesda, U.S.A.). The panel includes 60 cell lines of the following 9 human malignant tumors: leukemia, CNS cancer, prostate cancer, breast cancer, melanoma, non-small cell lung cancer, colon cancer, ovarian cancer, renal cancer.

The compounds were tested at 5 concentrations, resulting from a series of 10-fold dilutions (10^{-8}–10^{-4} M). Dimethylsulphoxide was used for preliminary dissolving of the compounds. A 48 h continuous drug exposure protocol and a sulforhodamine B (SRB) protein assay is used to estimate cell growth or cell death [8].

We have compared results of the anticancer activity tests of the above macrocycles at the concentration 1×10^{-4} M only, since the highest activity was demonstrated at this concentration. Anticancer activity is expressed as percent of control cell growth. Negative numbers testify that not cell growth, but cell destruction, takes place.

Anti-influenza activity *in vitro* was studied by inhibition of reproduction of A/Hong Kong/1/68 (H3N2) virus in tissue cultures of chorioallantoic membranes (CCM) of 12–14 days old chick embryos. Samples of compounds under examination (structures 1–33, Fig. 1) were dissolved in a special medium at 10^{-3} M concentration. CCM were infected with 1000 TID_{50} (tissue infection dose) of influenza virus. CCM cultures containing studied compounds were called "experimental" and those without them "control". Control and experimental samples were tested on viral infective titres after incubation (24 h, 37°C) [1]. At least 5 experiments were done while investigating each compound. Anti-influenza activity was expressed as $lgTID_{50}$ and
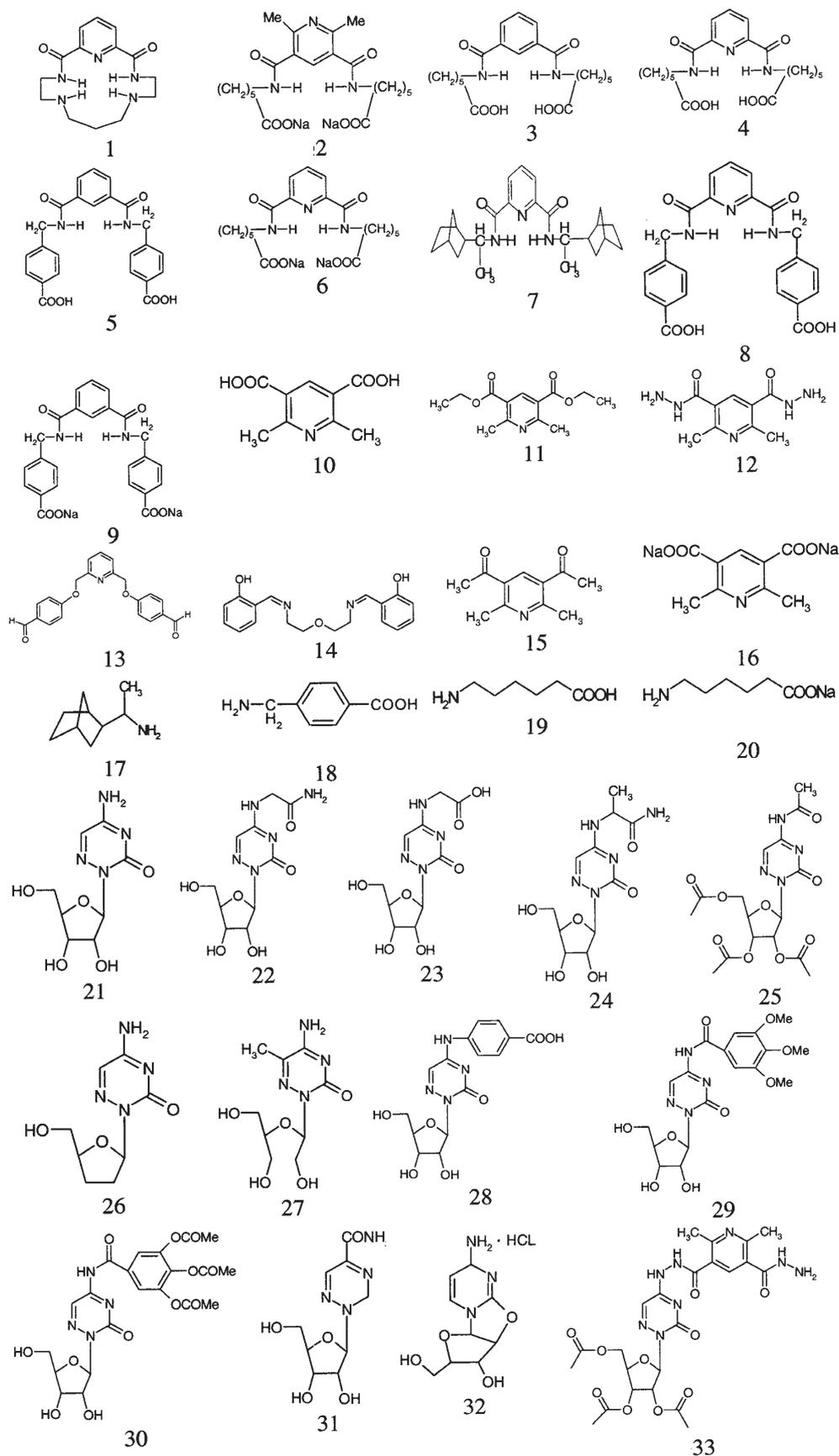
Figure 1. Investigated compounds.
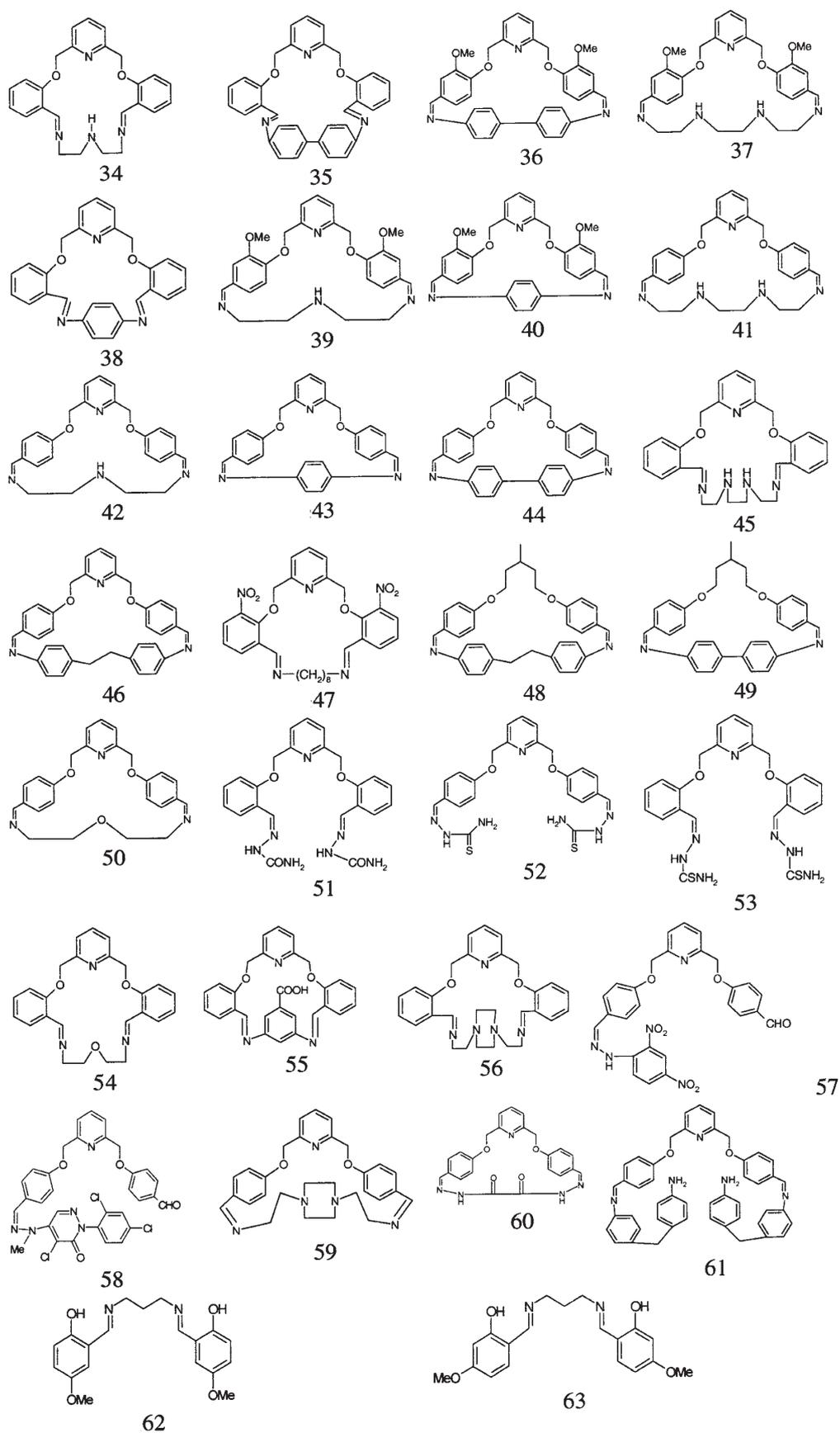
Figure 1. Continued.

reflected suppression of viral replication in "experimental", as compared to "control" samples.

Antiherpetic and antiadenoviral actions of investigated compounds were studied using cytomorphological method, based on calculation of the number of infected cells that contained intranuclear viral DNA inclusions [9]. The inclusions were revealed by the luminescent microscopy method.

A scheme of an experiment allowing determination of antiherpes activity was as follows. Two-day Hep-2 cell culture was infected with herpes simplex virus type 1 (HSV-1) strain US or adenovirus type 5 (Ad 5). Treatment of infected cells by compounds at the non-cytotoxic concentration (10^{-4} M) suggested their action on the adsorption and reproduction of the viruses. Adsorption of viruses was allowed for 2 h on Hep-2 cells in the presence of studied compounds. Then all cells were washed from non-adsorbed viruses. Reproduction of adeno- and herpesviruses was allowed for 48 h in the presence of studied compounds. Infected cells not treated by studied compounds were used as virus control. Then cells were washed and fixed by ethanol. After that infected cells were stained by a 0.01% solution of acridine orange and studied by luminescent microscopy method. Five hundred cells were counted in each of three parallel experiments.

The compounds ability to inhibit adenoviral or herpetical reproduction was estimated as reduction of the percentage of infected cells in treated, as compared with untreated, cell cultures.

RESULTS AND DISCUSSION

4D QSAR approach, based on simplex representation of molecular structure

In the present work a new 4D-QSAR approach has been considered. For all investigated molecules the 3D structural models have been created and sets of conformers (fourth dimension) [10] have been used. Each conformer is represented as a system of different simplexes (tetratomic fragments of fixed structure, chirality and symmetry) (see Fig. 2). An example is presented on Scheme 1.
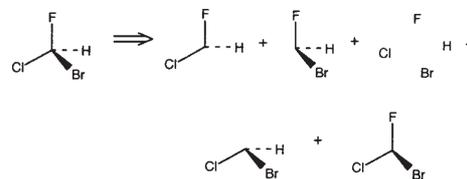
Scheme 1.

The simplex representation of molecular structure for biologically active substances allows unifying the description of spatial structure of compounds with saving of the complete stereochemical information [11–13]. It enables to determine easily common fragments of spatial structure both promoting given biological activity and interfering with it. Thus, molecular structure descriptors in such a model are simplexes (fixed topology and stereochemical configuration) and their various combinations. The molecular design of compounds with the given level of activity is possible by means of generating of the allowed combinations of such types simplexes, which determine investigated property.

The developed approach has not the limitations of CoMFA (comparative molecular field analysis) [14]. In the CoMFA only single fixed conformer is considered for each compound and a choice of such a conformer is often accidental. Moreover, in CoMFA there is a problem of optimal alignment of a set of considered molecules.

The approach developed in this work is similar to HQSAR (holographic QSAR) [15] but has none of its restrictions, connected with only topological representation of molecular structure and lacks ambiguity of descriptors formation when procedure of hashing of molecular holograms is realized.

Simplexes as descriptors of molecular structure

For any molecule there are 11 base types of molecular simplexes with all possible variants

Table 1. All possible basic types of molecular simplexes.

of atoms bonding in a simplex. The molecular graphs of the base types of simplexes and respective examples of molecular fragments are represented in Table 1. It is possible to differentiate all simplexes to right (R), left (L), symmetrical achiral (S), plane achiral (P) (see Scheme 2).

Stereochemical configuration of simplexes is defined by modified Kahn-Ingold-Prelog rules [12].

The structure of each simplex in molecule has been determined, with differences in atom and bond types considered. All atoms in a simplex are differentiated, based on the following characteristics: 1) element; 2) partial charge; 3) lipophility; 4) atomic refraction; 5) electronegativity of atom; 6) donor/acceptor of hydrogen in H-bond; etc.

Based on characteristics, expressed by real values (Nos. 2–5 in the list), preliminary division into a definite discrete number of atom groups, characterized by different ranges of those values, is made.[1]

Chemical bonds are differentiated by its nature: single, double, triple, aromatic.

The subsequent examples (see Fig. 2) demonstrate different systems of the representation of molecules as simplex sets. Thus molecule of acetic acid can be represented as a set of 13 simplexes (3 right, 3 left, 2 symmetrical,

5 plane). The differentiation of atom by element gives 7 types of simplexes, but differentiation of atom by charge gives only 4 types of simplexes.

In the proposed model a number of simplexes of specific type is a descriptor of the molecular structure.

Descriptors calculation in 4D QSAR models

In the 4D QSAR models each molecular structure parameter (D, descriptor) is calculated by summing products of descriptor value for each conformer $(D_k)^2$ and probability of the corresponding conformer (P_k).

$$D = \sum_{k=1}^{N}(D_k \cdot P_k),$$

where N is the number of conformers being considered, D_k is the descriptor value for conformer k.

As it is well known [16], a probability of conformation P_k is defined by its energy:

$$P_k = \left\{1 + \sum_{i \neq k} EXP\left(\frac{-(E_i - E_k)}{RT}\right)\right\}^{-1}, \quad \sum_k P_k = 1,$$

Scheme 2.

---
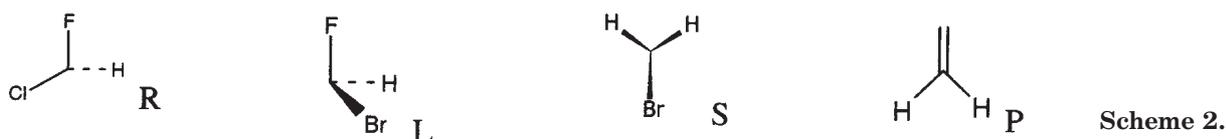
[1]The number of groups (N) is a tuning parameter and can be varied. As a rule N = 3–7.
[2]The conformational search was carried out using software "HyperChem 6.01". Trial-version from http://www.hypercube.com

where E_i, E_k is the energy of conformations i and k. The conformers are analyzed within energy band of 5–7 kcal/mol.

5. Choice of significant structural parameters (descriptors) and QSAR analysis using statistical methods such as regression analysis,

No.	Simplex	Base type	Quantity of simplexes of different kinds[*]				
			R	L	S	P	A
1		5				1	1
2		6	1	1		1	3
3		6	1	1		1	3
4		6				1	1
5		5	1	1	1		3
6		5			1		1
7		6				1	1
Sum total			3	3	2	5	13

No.	Simplex[**]	Base type	Quantity of simplexes of different kinds[*]				
			R	L	S	P	A
1		5				1	1
2		6	1	1		2	4
3		6	1	1		2	4
4		5	1	1	2		4
Sum total			3	3	2	5	13

Figure 2. Examples of fragmentation by element (a) and by atom charge (b) of acetic acid into simplexes using base types 5 and 6 only.

*Kinds of simplexes: **R,** right configuration; **L,** left configuration (chiral, see [12]); **S,** symmetrical (achiral); **P,** plane (achiral); **A,** any (configuration has not been considered); [**]$q_A < = 0$, $q_B > 0$, where q is partial atom charge.

An example of descriptors calculation in 4D QSAR has been demonstrated on Fig. 3. The calculation of descriptor value for one of all possible simplexes of alanine has been shown.

Thus, 4D QSAR procedure consists of the following stages:

1. Formation of training set.

2. Total conformational analysis of each compound and choice of more stable conformers within specified energy diapason (see Fig. 3).

3. Fragmentation of each conformer into simplexes. Identification of type of each simplex and defining its stereochemical configuration (see Fig. 2).

4. Descriptor value calculation for each concrete type of simplex, taking into account probability of its realization in a given compound (see Fig. 3).

PLS (partial least squares), neural networks, etc.

6. Verification of QSAR models using leave-one-out cross-validation procedure and test set.

7. Off-experimental screening of novel compounds. Molecular design of effective agents.

It should be noted that stages 1, 5, 6 and 7, are obligatory for any QSAR research.

Results of the QSAR modeling

As compared to the previously used QSAR approaches, based on topologic modeling of molecular structure [4, 7], a more adequate «structure–activity» model has been constructed with the use of a training set representative for anticancer and various types of

Alanine							
	Set of favourable conformers						
Energy (E_i)	−6.03	−5.86	−5.75	−5.38	−5.12	−4.97	−1.13
Probability(P_i)	0.323	0.242	0.202	0.109	0.069	0.054	$8.18 \cdot 10^{-5}$
Calculation of the number of simplexes for these conformers, exemplified by one possible simplexes							
Right	2	1	2	1	1	1	1
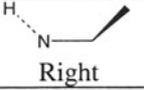
Descriptor value calculation for the considered simplex

$$D = 0.323 \cdot 2 + 0.242 \cdot 1 + 0.202 \cdot 2 + 0.109 \cdot 1 + 0.069 \cdot 1 + 0.054 \cdot 1 + 8.18 \cdot 10^{-5} \cdot 1 = 1.525$$

Figure 3. An example of descriptor value calculation in 4D QSAR.

antiviral activity and using 4D QSAR approach. It is noteworthy that the training set (Fig. 1) encompasses a broad structural variety: different MCP and their acyclic analogues, and also some of synthetic nucleosides (SN) and well-known antiviral agents (ambenum, deiteforin and other) representing various classes of organic compounds. 3D QSAR approaches, such as CoMFA [14], HASL [17], lattice model [18] presumably cannot to be applied for this set. The main reason is that one-valued and optimal alignment of such heterogeneous structures for calculation of molecular field parameters is impossible.

As mentioned above, such limitation does not exist for simplex representation of molecular structure. In the present work the PLS method [19] was used for QSAR solutions. Quite adequate (R = 0.92–0.97) "structure–activity" models (Tables 2, 3) were obtained using PLS, by considering sets of conformers (4D-QSAR) instead of one fixed conformation. It is of note that all models have satisfactory predictive ability (CVR^2 > 0.4, Table 3). Observed and predicted values of

considered types of biological activities are presented in Table 2. The new approach allowed to analyze the contributions to appropriate activity of all possible simplexes for each molecule. These contributions were calculated for each atom. Contribution of an atom is defined as an arithmetic mean of contributions of all simplexes[3], containing this atom. Thus, for each molecule it is possible to analyze an influence of various fragments on considered biological activity. Examples of such analyses for active and inactive molecules are presented in Fig. 4.

Analysis of simplexes combinations allowed to select appropriate molecular fragments (Table 4) which have positive and negative influence on values of each type of considered biological activity. This information may be useful for design and direct synthesis of novel antiviral and anticancer agents. The results suggested to synthesize and investigate the following molecular structures, expected to shown different biological activities.

The results of investigation of these compounds will be discussed in separate publications.

---

[3]In this tasks for description of molecular structure we used the simplexes of types 5 and 6 only.

Table 2. Observed (Obs.) and predicted (Pred.) values of antiviral and anticancer activity of investigated compounds (see Fig.1)

| No. | Antiviral activity | | | | | | No. | Anticancer activity | |
| | Influenza | | HSV-1 | | Adenovirus | | | Obs. | Pred. |
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.			
1	0.2	0.5	57.5	51	0	0	34	–8	19
2	nd	–0.4	23.6	17	0	4	35	42	54
3	–0.45	–0.3	2	6	21	16	36	92	96
4	–0.25	0.0	65.2	10	0	9	37	8	65
5	0.72	0.3	21	35	0	6	38	64	46
6	0.25	0.2	13	10	11	8	39	65	63
7	2.5	3.1	nd	0	0	31	40	84	94
8	0.71	0.5	39.4	37	8	1	41	54	47
9	0	0.2	40	35	12	5	42	54	48
10	1.25	0.6	nd	49	25	20	43	89	79
11	nd	1.5	nd	48	1	17	44	7	76
12	–0.25	–0.5	62	60	31	21	45	23	22
13	nd	1.9	nd	46	nd	25	46	100	88
14	5.33	5.4	nd	43	4	46	47	92	90
15	0.33	0.2	46	50	22	22	48	78	89
16	0.33	0.4	43.3	50	0	16	49	94	88
17	3.5	2.4	nd	33	nd	35	50	77	51
18	0.05	0.5	52.44	48	nd	30	51	29	24
19	0.15	0.7	46	38	nd	28	52	34	29
20	0.75	0.7	27	37	nd	27	53	22	12
21	nd	2.6	nd	41	100	84	54	2	25
22	nd	2.5	nd	41	64	58	55	35	39
23	nd	2.4	nd	38	11	47	56	–12	–6
24	nd	2.3	nd	38	36	31	57	78	85
25	nd	1.8	nd	32	100	92	58	58	35
26	nd	2.5	nd	39	100	81	59	78	92
27	nd	3.2	nd	41	98	103	60	94	99
28	nd	2.6	nd	40	0	–13	61	62	71
29	nd	4.1	nd	39	76	83	62	2	17
30	nd	2.5	nd	30	55	72	63	24	21
31	nd	3.0	nd	44	55	57			
32	nd	2.8	nd	42	100	113			
33	nd	0.1	nd	45	53	56			

nd, no data; Obs., observed activity; Pred., predicted activity. Anti-influenza activity is expressed as $lgTID_{50}$; TID_{50}, 50% tissue infected dose. Anti-HSV-1 and antiadenoviral activities are expressed as percent of reduction of number of infected cell as compared to control. Anticancer activity is expressed as percent of control cell growth.
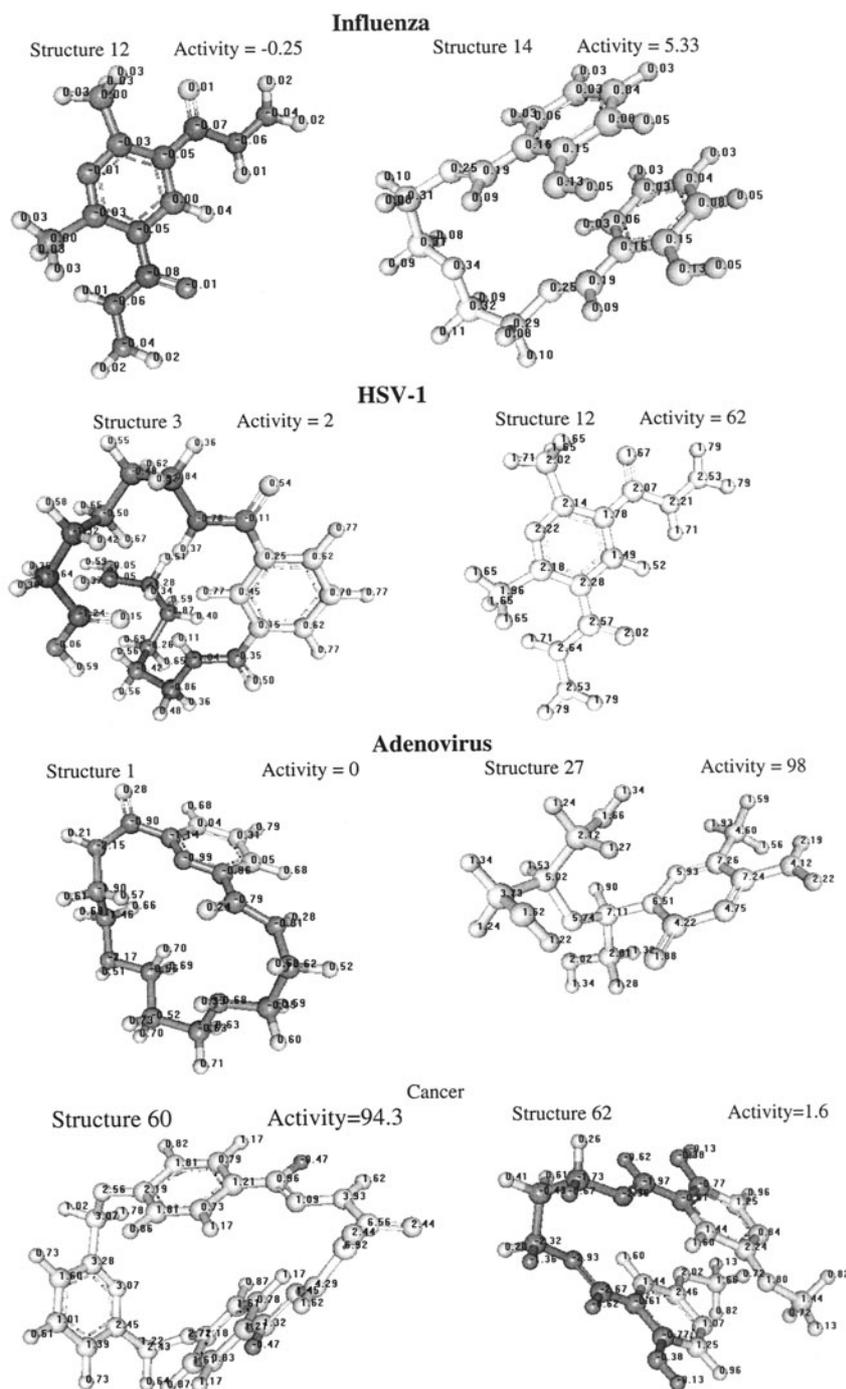
Figure 4. The influence of molecular fragments on value of activity.

Dark colour, negative influence on antiviral activity and positive influence on anticancer activity; light colour — positive influence on antiviral activity and negative influence on anticancer activity.

Table 3. Analysis of antiviral and anticancer activity by PLS-method

Statistical parameter	Anti-influenza	Anti-HSV-1	Antiadenovirus	Anticancer
R	0.960	0.922	0.965	0.924
CVR	0.632	0.730	0.830	0.667

PLS, partial least squares; R, correlation coefficient.

Table 4. Molecular fragments which increase and decrease anticancer and antiviral activity of MCP and SN.

Anti-influenza	Anti-HSV-1	Antiadenovirus	Anticancer
Increase			
(structure)	*(structure)*	*(structure)*	*(structure)*
(structure)	-CH₂-COOH	*(structure)*	*(structure)*
(structure) COOH	*(structure)* COONa	*(structure)*	*(structures)*
(structure)	*(structure)* COOH		*(structure)*
Decrease			
-CO-NH-CH₂-	-CH₂-O-CH₂-	*(structure)*	*(structure)*
(structure)	-CH₂-COONa	-NH-CH₂-CH₂-NH-	-OCH₃
(structure)	*(structure)*	*(structure)*	
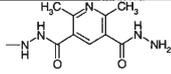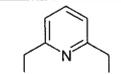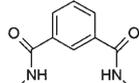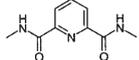
Thus, developed 4D QSAR approach can be used for effective solution of the most different "structure–activity" tasks and molecular design of perspective drugs.

———

The authors express gratitude to colleagues from NCI for fruitful cooperation.

REFERENCES

1. Lozitsky, V.P., Puzis, L.E. & Polyak, R.Ya. (1988) Resistance of mice to reinfection after E-aminocaproic acid treatment of primary influenza virus infection. *Acta Virol.* **32**, 117–122.

2. Dyachenko, N., Nosach, L., Zhovnovataya, V. & Butenko, N. (1994) The effect of some nucleoside analogues on reproduction of adenoviruses. *Antiviral Res.* **23** (Suppl.1), 121.

3. Fedtchouk, A.S., Veveritsa, P.G., Lozitsky, V.P. & Girlya, Yu.I. (1999) Medical cure of recidiving herpes simplex virus infections by means of proteolysis inhibitors. *Antiviral Res.* **41**, A67.

4. Kuz'min, V.E., Lozitsky, V.P., Kamalov, G.L., Lozitskaya, R.N., Zheltvay, A.I., Fedtchouk, A.S. & Kryzhanovsky, D.N. (2000) The analysis of "structure–anticancer activity" relationship in a set of macrocyclic 2,6-bis (2- and 4-formylaryloxymethyl) pyridines Schiff bases. *Acta Biochim. Polon.* **47**, 867–875.

5. Lozitsky, V.P., Kuz'min, V.E., Artemenko, A.G., Lozitskaya, R.N., Fedtchouk, A.S., Gridina, T.L. & Kryzhanovsky, D.N. (2001) The analysis of structure-anti-influenza relationship on the basis molecular lattice model for macrocyclic piridinophanes and their analogs. *Antiviral Res.* **50**, A85.

6. Fedchuk, A.S., Lozitsky, V.P., Gridina, T.L., Lozytska, R.N., Kamalov, G.L., Vanden Eynde, J.J., Dyachenko, N.S., Nosach, L.N. & Povnitsa, O.Y. (2001) Synthesis and antiviral properties of 2,6- and 3,5-disubstituted pyri-

dines and their analogs. *Antiviral Res.* **50**, A86.

7. Kuz'min, V.E., Lozitsky, V.P., Kamalov, G.L., Lozitskaya, R.N., Zheltvay, A.I., Fedtchouk, A.S., Kryzhanovsky, D.N., Artemenko, A.G. & Kuz'mina, A.V. (2000) The analysis of "structure–anticancer activity" relationship in a set of macrocyclic pyridinophanes. *Intl. Y. Med. Biol. Environ.* **28**, 75–82.

8. Monks, A., Scudiero, D. & Skehan, P. (1991) Feasibility of high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *Y. Natl. Cancer Inst.* **83**, 757–766.

9. Nosach, L. & Dyachenko, N. (1982) *Cytopathology of adenoviruses infections;* pp. 1–234, Naukova Dumka, Kiev (in Russian).

10. Vedani, A. & Dobler, M. (2000) Multi-dimensional QSAR in drug design. *Prog. Drug Res.* **55**, 107–135.

11. Kuz'min, V.E. (1994) The structure of chiral molecules. Analysis of the conception of configuration and stereoisomerization mechanism. *Zh. Phys. Khim.* **68**, 1037–1043 (in Russian).

12. Kuz'min, V.E. (1995) The homo- and heterochirality of the dissimmetrical tetrahedrons (chiral simplexes). Stereochemical tunneling. *Zh. Struct. Khim.* **36**, 871–878 (in Russian)

13. Kuz'min, V.E., Chelombitko, V.A., Yudanova, I.V., Stelmakh, I.B. & Rublev, I.S. (1998) Stereochemical analysis by simplex representation of molecules. *Zh. Struct. Khim.* **39**, 549–554 (in Russian).

14. Cramer, R.D., Patterson, D.I. & Bunce, J.D. (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape binding to carrier proteins. *J. Am. Chem. Soc.* **110**, 5959–5967.

15. Seel, M., Turner, D.B. & Wilett, P. (1999) HQSAR — a highly predictive QSAR Technique based on molecular holograms. *QSAR,* **18**, 245–252.

16. Burkert, U. & Allinger, N.I. (1982) *Molecular Mechanics;* pp. 1–364, ACS, Washington D.C.

17. Doweyko, A.M. (1988) The hypothetical active site lattice. An approach to modeling active sites from data on inhibitor molecules. *J. Math. Chem.* **31**, 1396–1406.

18. Kuz'min, V.E., Artemenko, A.G., Kovdienko, N.A., Tetko, I.V. & Livingstone, D.J. (2000) Lattice model for QSAR studies. *J. Mol. Modeling* **6**, 517–526.

19. Rännar, S., Lindgren, F., Geladi, P. & Wold, S. (1994) A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: Theory and algorithm. *J. Chemometrics* **8**, 111–125.