RC/PELCO and OASIS methods I. Methodological comparison. Modeling purine pKa and antitumor activity

ARTICLE *in* EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY · SEPTEMBER 1991

Impact Factor: 3.45 · DOI: 10.1016/0223-5234(91)90192-P

4 AUTHORS, INCLUDING:

Danail Bonchev
Virginia Commonwealth University
263 PUBLICATIONS **4,230** CITATIONS

SEE PROFILE

Eur J Med Chem (1991) 26, 575–592

DARC/PELCO and OASIS methods I. Methodological comparison. Modeling purine pKa and antitumor activity

C Mercier[1], O Mekenyan[2], JE Dubois[1], D Bonchev[2]

[1]*Institut de Topologie et de Dynamique des Systèmes de l'Université Paris VII, associé au CNRS, 1, rue Guy-de-la-Brosse, 75005 Paris, France;*
[2]*Highler School of Chemical Technology, 8010 Bourgas, Bulgaria*

(Received 1 March 1990; accepted 20 December 1990)

Summary — Two QSAR approaches have been compared in detail. The DARC/PELCO method is based on the exhaustive generation of all topochromatic sites around the reference structure and the evaluation of their contribution to the property. The OASIS system is an optimized version of the extended Hansch type physicochemical method which makes use of a large set of molecular descriptors (topological, steric, quantum chemical, etc.). The DARC/PELCO models of a series of purine derivatives have a more specific area of reliable prediction and greater accuracy than OASIS models. The latter, in turn, provide more details concerning the possible biological mechanism. In spite of their differences, the two methods produce similar predictions for the properties under study. A conclusion is drawn about the complementarity of these QSAR approaches for optimizing the lead structure in congeneric series of biologically active molecules.

Résumé — Méthodes DARC/PELCO et OASIS. I. Comparaison méthodologique. Modèles d'évolution des pKa et d'activité antitumorale de purines. *Nous présentons une comparaison approfondie de deux méthodes quantitatives de recherches de relations structure–activité (QSAR). La méthode DARC/PELCO est basée sur la génération exhaustive de tous les sites topochromatiques autour de la structure de référence et de leur contribution à la propriété. Le système OASIS est une version optimisée des méthodes physico-chimiques de type Hansch et utilise un ensemble étendu de descripteurs moléculaires (topologiques, stériques, quantiques...). Les modèles DARC/PELCO d'évolution des pKa et de l'activité antitumorale de quelques séries de dérivés puriniques ont une plus grande précision et un domaine mieux défini de prévisions fiables que les modèles OASIS. Ces derniers en contrepartie, fournissent plus d'élements pour la compréhension du mécanisme d'action. En dépit de leurs différences, les deux méthodes conduisent à des prévisions similaires pour les propriétés étudiées. Nous abordons en conclusion la complémentarité de ces deux approches de QSAR pour l'optimisation de séries de molécules biologiquement actives.*

QSAR / DARC/PELCO / OASIS / purines / pKa / antitumor

Introduction

The basic goal of structure–activity relationship (SAR) studies is to design more active and less toxic molecules. The two major trends in drug design are: lead generation and lead optimization. The first makes use of computational techniques like cluster and discriminating analysis, but their contribution to practical pharmacology is still debatable. This approach attempts to find the common structural features conditioning activity within classes of structurally dissimilar molecules. The common fragments thus found determine new classes of biologically active molecules. The quantitative structure–activity relationship (QSAR) studies contribute to the development of the second approach. Here, biological response is correlated with chemical structure *via* structural or physicochemical parameters called molecular descriptors. The models are obtained by multiple regression analysis on congeneric series of compounds. The critical analysis of QSAR approaches, as well as comparative studies on existing methods should help to determine the most convenient methods for solving a specific QSAR problem. Various correlation methods have been proposed, separating to a greater or lesser degree physical information and biological activity from structural information. They have led to certain correct predictions and have contributed to the elucidation of the mechanisms by which biologically active compounds act. It is nonetheless essential for determining the method best adapted to the solution of a particular QSAR problem.

The DARC/PELCO topological method, dealing with structural environments, was proposed during the

1970s, first by Dubois (1966, 1973) [1, 2]. It was then refined, broadened and computerized by Dubois *et al* and Mercier *et al* [3, 7]. In recent years, it soon extended to the study of the 3-D action of conformation in competition (hydrophobicity) and completed by a method for studying the action of the focus (PULFO) [8–10]. Various applications to complex biological situations [11–17] have confirmed the validity of this method for correlating successive action focuses. It has proved very useful for diverse structure–property correlations.

The PELCO method was compared with those of Hansch and Hansch and Fujita [18, 19] and Free and Wilson [20] and that of molecular connectivity (MC) [21] as early as 1973 by Mercier and Dubois [22] on an antimicrobial halogenophenol population, and by Pacheco *et al* [23]. One can, of course, use the PELCO approach in association with features from certain quantum modelizations [24] or from molecular mechanics. These associations have not yet been formalized in computer programs, while the OASIS system developed by Mekenyan and Bonchev and Mekenyan *et al* [25–30] tends to integrate the various topological, metric, quantum-mechanical and physicochemical descriptors in this way.

In this article, the two methods are systematically compared in a study carried out by the two research teams who developed the DARC/PELCO and OASIS methods. We examine examples involving classes of biologically active compounds. This implies particular evaluation processes related to pK modeling and to other features of purines in association with the current models of the antitumor activity of these products. The aim of a common project of this kind is to take a critical step forward in data validation, to test the operational nature of programs and to compare the prediction values and areas of reliability of these two methods.

On the exploratory level, there were already signs of the notable correlation potential of DARC/PELCO, while there were important interpretative tools with OASIS. These main avenues had to be tested and their complementary character evaluated. Moreover, we had to seek the possible existence of a synergy arising from their eventual combination for certain correlation and modeling strategies. Such validation studies are rare. We therefore present the characteristics of both methods to facilitate their methodological comparison in specific applications both in this article and in another study on PNMT inhibitory potency of benzylamines and amphetamines.

Materials and Methods

Physicochemical methods: the OASIS concept

According to Hansch's physicochemical method, the variations in biological activity caused by substituents in a congeneric series can be attributed to three types of substituent effect: hydrophobic, electronic and steric, described by the hydrophobic index, π, the Hammett constant, σ, and Taft's steric parameter, *Es*, respectively:

$$\log 1/C = -a\pi^2 + b\pi + d\sigma + eEs + f \qquad (1)$$

The OASIS (optimized approach based on structural indices set) method is based on the same assumption and can be regarded as an extended and optimized version of the Hansch approach. More specifically: i) besides substituent constants, a large set of calculable geometric (topological and steric) and electronic indices for both molecules and their fragments is considered in the OASIS method:

$$\begin{aligned}\log 1/C = {} & a\ (\text{topological parameters}) + \\ & b\ (\text{steric parameters}) + \\ & d\ (\text{electronic parameters}) + \\ & e\ (\text{hydrophobic parameters})\end{aligned} \qquad (2)$$

This equation allows for the possibility that more than one component of the different factors contribute to the overall biological action of the molecule. On the other hand, these components can be either local, *ie* referring to atoms of the reference molecule, or global, *ie* describing the molecule as a whole.

Topological indices, derived from different graph-invariants, are used as topological molecular descriptors: the Randic connectivity index [31], Wiener number [32 ,33], Hosoya nonadjacency number [34 ,35], Balaban centric indices [36], Balaban distance connectivity index [37 ,38], Zagreb group indices [39, 40], Bonchev information theoretical indices [41], etc. The atomic distance number [42, 43] and path numbers [44] are used as local topological indices, along with the fragment topological indices [45].

Molecular metrics is characterized by the 3-dimensional distribution of the atoms in one of the stable molecular conformations. Global molecular metrics is assessed [27] by the metric analogues of the Wiener number and its information counterpart, the largest interatomic distance, as well as by other indices based on the matrix of interatomic geometric distances. The geometric analogue of the atomic distance number is used as a local metric parameter, along with Verloop's multidimensional steric indices [46].

Electronic structure is described quantum-mechanically (MO-LCAO) for the molecular ground state. Different semiempirical methods [47–50] are adopted in the OASIS system [51], in addition to a method originally designed for π-electronic structure description, taking into account the σ-skeleton. Various indices are calculated: frontier orbital energies and their differences, dipole moments (global descriptors), atomic charges, σ- and/or π-acceptor and donor superdelocalizability indices, their frontier analogue (local descriptors). Nonquantum-chemical electronic indices are also used. They proceed from a joint consideration of atom connectivity and electron distribution: the extended connectivities of Kier, and Kier and Hall [52, 53], l'Haya electropy and bondtropy indices [54, 55], neighborhood indices of Ray *et al* and Basac *et al* [56, 57], electronic analogues of Balaban distance connectivity and Wiener index [29].

In order to describe the effect of the environment on a reference structure, the physicochemical constants of its substituents [46, 58–60] are used: the Hammett constants, their field and resonance components, molecular refraction, hydrophobic parameter, Taft steric parameter, Charton steric constant, molecular weight, van der Waals volume, etc. They are retrieved from a database of about 1600 substituents. The physicochemical properties of molecules can also be used as global characteristics. Thus, the OASIS system also provides an option to perform extended Hansch analysis.

The molecular descriptors listed in the foregoing are systematized in table I.

ii) In Hansch's approach, the substituent effect is evaluated indirectly, because these constants are obtained from a linear free energy relationship with the substituent effect measured on an external reference population. Thus, these parameters are usually classified as external structural variables [6].

In series of congeneric molecules (causing one and the same type of biological activity), all compounds of the experimental population incorporate the reference (or 'lead') structure as a common substructure. Consequently, in the OASIS approach the reaction sites responsible for biological interaction are assumed to belong to the reference structure only but not to the substituents modifying it. The substituents attachment only modifies the reference structure potency. For this reason, when the local characteristics that condition a specific biological activity are determined, alterations in the electronic (charges, superdelocalizability indices) and geometric (path numbers and other atomic geometric indices) features are only examined for the reference structure and they are therefore internal structural variables.

Indeed, global parameters could also reflect the specific behaviour of molecules. Nonspecific biological interactions are assessed by the hydrophobic parameters along with the descriptors related to the compound as a whole. The latter evaluate changes in the electronic and geometric structure, as well as in the physicochemical properties of the molecules.

The topological, steric and electronic indices calculated in the OASIS system are derived from purely structural information. For this reason, they are also classified as internal structural variables.

Table I. Molecular descriptors used in the OASIS system.

Molecular geometry				*Electronic structure*			*Physicochemical properties*	
Topological indices		*Steric indices*		*Non-quantum chemical indices*	*Quantum-chemical indices*			
Global	*Local*	*Global*	*Local*	*Global*	*Global*	*Local*	*Global*	*Local*
Randic connectivity index	Atomic distance number	Geometric analogues of Wiener index and its information counterparts	Geometric analogue of atomic distance number	Valence connectivity index of Kier and Hall	Energy of frontier orbitals and their difference	π- and/or $\sigma+\pi$ charges and acceptor and donor superdelocalizability indices (including the frontier ones)	Heat of formation	Hammett constants
Wiener number	Path numbers			l'Haja electropy and boundtropy	Total electron energy		Heat of combustion	Verloop sterical indices
Hosoya non-adjacency number	Fragment indices	Maximal distance in molecules		Neighborhood indices of Sarcar, Roy, Basak *et al*	Dipole moments		Boiling point	Molecular refraction
Balaban centric indices					Molar polarizability		Partition coefficient	Taft constants
Balaban distance connectivity index				Electronic analogues of Balaban distance connectivity index and Wiener index			Chromatographic retention indices	Molecular volume
Zagreb group indices								van der Waal's volume
Bonchev information theoretic indices								

After the initial extended set of variables has been specified, the latter is subjected to a rational screening due to: i) the combinatorial complexity in using any stepwise multivariate regression technique; ii) the possibility of detecting in advance the structural features most significant for activity; iii) reduction of possibility for a chance correlation [61].

The initial set of parameters B is partitioned into disjoint subsets (clusters). The clustering procedure is based on the so-called parameter intercorrelation graph IG [62, 63]. Two vertices of this graph i and j which correspond to the parameters B_i and B_j are connected only if the correlation coefficient for the respective parameters is higher than the threshold level X ($O \leq X < 1$):

$$ij \in E(IG(X)) \text{ for } r(B_i, B_j) > X$$

where E is the set of edges of IG(X). In fact, when this procedure is implemented the number of clusters prescribed by the user is obtained by varying the threshold X value (usually in the range 0.5 to 0.7).

The second step of OASIS parameter selection considers the objective property influence. It proceeds from linear correlation between each parameter and biological activity. For each cluster the most significant parameters are selected for the modeling procedure.

Then, a stepwise regression analysis based on singular value decomposition [64] searches for the best structure–activity model. All combinations between the selected representatives of each cluster are examined by taking only one parameter from a cluster in the current model. An additional parameter is included in the model only when there is a statistically significant improvement.

The OASIS system does not have any specific requirements as to the choice of the experimental population to be handled. Rather, it encompasses the set of all available compounds with a known activity. Once the model has been derived, interpolations are carried out within the range of variation of regression model parameters. Unlike other approaches, the OASIS interpolation criterion, however, is sufficient to provide predictions beyond the structural area defined by the population treated. Indeed, the interpolations for compounds having substituents other than those from the experimental series under study are legitimate because in characterizing the compound activity the changes in local electronic and geometric features of reference structure atoms are analyzed but not those of the substituents.

On the other hand, one cannot always classify the OASIS predictions as interpolations or extrapolations. The answer to this question depends on the range of parameter variations included in the ultimate model. Thus the same prediction may be regarded as an interpolation for one parameter and an extrapolation for another one.

In the last stage of the method the validation of the best model is tested by the 'leave-one-out' procedure [65].

Purely structural models. The DARC/PELCO method

The Free–Wilson structural model [20] is based on the additivity concept: each time a substituent group appears in the same position, its contribution to the biological activity is additive and constant in spite of the structural variations in the rest of the molecule (its environment). The contribution of each substituent x_i in position j to the overall average biological activity μ is expressed by a specific constant G_{ij} calculated by means of the least-squares fit:

$$BR = \mu + \Sigma G_{ij} \cdot X_{ij} \qquad (3)$$

where X_{ij} is the existence of substituent X_i in position j.

The Free–Wilson model was next modified by Fujita and Ban [66]. Aside from the fact that the term μ_o is the activity of the parent molecule, interaction terms can be introduced into the correlation.

The DARC/PELCO method [1–7] is one of the most highly developed and optimized structural models. It uses a very general yet refined structural variable which makes several approaches possible, using either topological, fragmentary or global variables [67].

The structural variable is based on simultaneous representation of all the experimental compounds and of the population containing them. Each molecule is represented by an ordered chromatic graph [6] which describes exhaustively the topological and chemical nature of each site (atom, bond). These sites are concentrically organized around the focus (parent molecule) according to priority rules (fig 1).

The generation of the experimental compounds creates an ordered multidimensional space called the hyperstructure HS [6], where the structures studied are located with respect to the affiliated ones (formal molecules related to those of the experimental population) (fig 2). These new structures, whose activity is predictable by interpolation from the experimental structures, constitute the so-called 'Proference'.

Within this vectorial space HS, whose basis is the population trace [6], each compound is characterized by its topochromatic vector (fig 3). This vector directly accounts for the overall topology and chromatism. It includes not only the nature of bonds and atoms, but also the geometrical and stereochemical data of the structure described. This variable is explicit and makes it possible to retrieve the molecular structure directly and non-ambiguously. Furthermore, it is well suited to prediction reliability

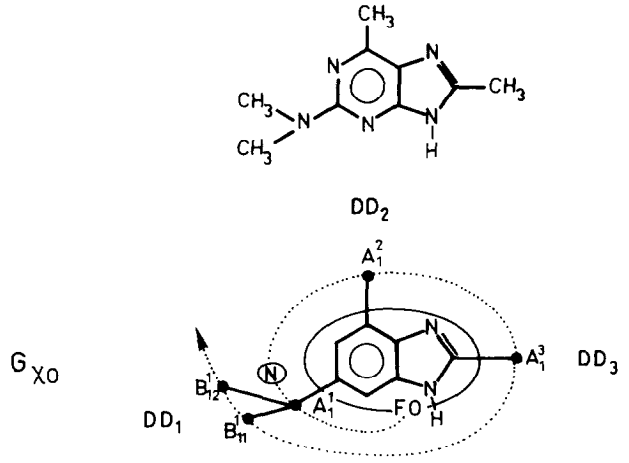


Fig 1. Generation of an ordered chromatic graph $G\chi_0$ modeling a structure. The ordered chromatic graph $G\chi_0$ modeling a target structure (**56**, table III) is concentrically organized around the substructure common to the population called the focus, FO. It is linearly ordered by generating the sites of the environement according to topological and chromatic criteria.

estimation by structural interpolation within the ordered multidimensional space.

The topochromatic vectors constitute the basic structural multidimensional variable used in the DARC/PELCO method. They are built automatically by choosing all the structural elements included within the population trace. They can be transformed into modulated topochromatic vectors by extending or diminishing the dimensions of the structural representation space [6]. The primary topochromatic site s_p, which constitute the basic elements, can be automatically or interactively combined into complex sites. Condensed sites s_c are used to take account of eventual gaps in the population. Interaction sites S_i and equivalence sites s_e are used to search for an optimal SAR: s_i account for some additivity deviations while s_e reflect regularities.

The SAR search method is based on the synchronous generation of information I (biological activity) with that of structure S and hyperstructure HS [6]. The information I of a structure S belonging to an isofocal population is equal to the sum of information I_o of the parent structure (the focus) and of the information contributions of each topochromatic site s, called perturbations $p(s)$. The general topo-information is expressed as:

$$I = I_o + \Sigma p(s) \cdot \zeta(s) \qquad (4)$$

where s is either a primary site s_p, an interaction site s_i or an equivalence site s_e, $\zeta(s)$ the occurrence of site s in the modulated topochromatic vector.

The topo-information relationship search is based on the structural interpolation hypothesis in S/HS/I

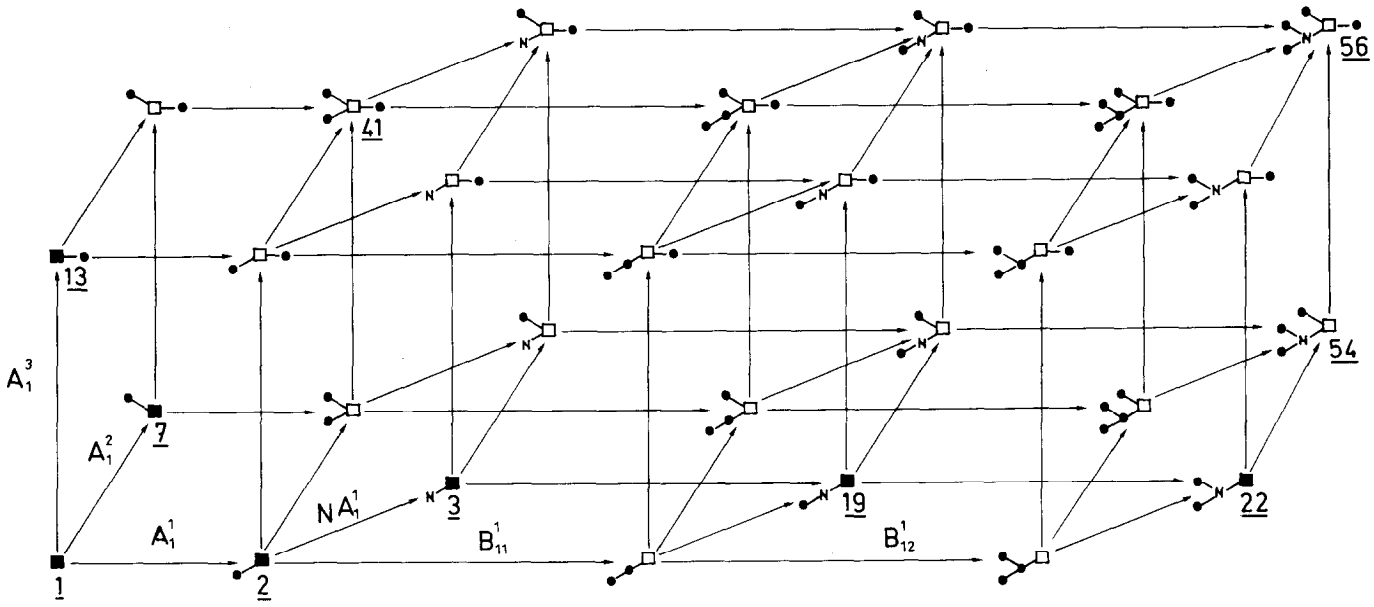
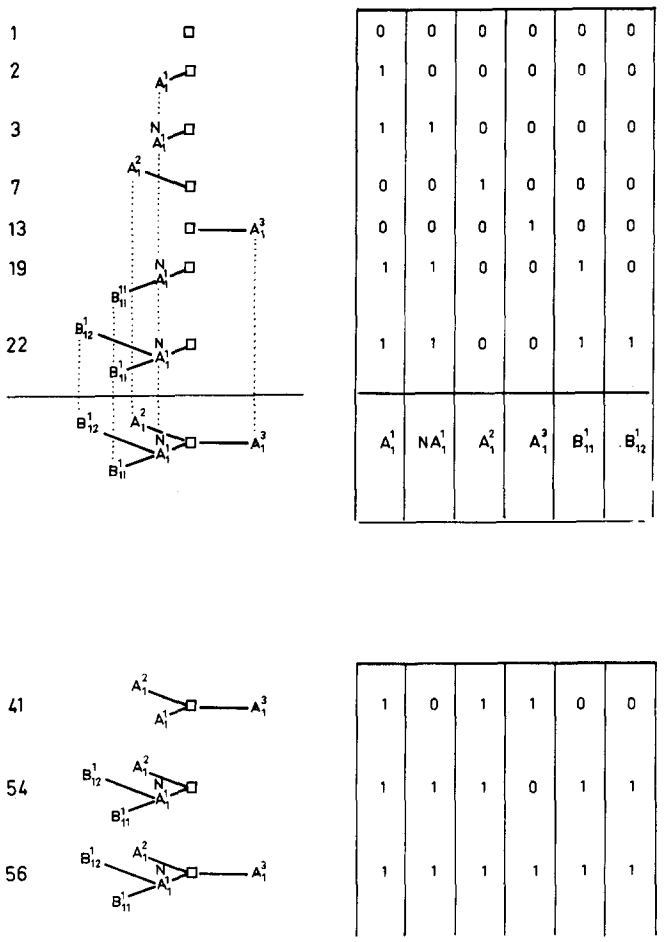


Fig 2. Hyperstructure of the sample population. The new structures (in light) are engendered starting from the seven experimental structures of the sample population (in dark). They belong to the prediction range and constitute the so-called 'Proference'. N: number of experimental structures (tables II and III); ■: experimental structures of the sample population; ❐: related structures belonging to the proference.



1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	1	1	0	0	0	0
7	0	0	1	0	0	0
13	0	0	0	1	0	0
19	1	1	0	0	1	0
22	1	1	0	0	1	1
	A_1^1	NA_1^1	A_1^2	A_1^3	B_{11}^1	B_{12}^1

41	1	0	1	1	0	0
54	1	1	1	0	1	1
56	1	1	1	1	1	1

Fig 3. Sample population trace (left) and topochromatic vectors (right). The sample population trace, obtained by superimposing all the experimental structures under study, constitutes the basic representation of the hyperstructure. In this vectorial space, each structure's graph is represented by a topochromatic vector which indicates the sites used for its generation.

space and its optimization on a heuristic modulation of this space [6].

The best model is derived from an experimental key population chosen so as to ensure maximal structural interpolation of the predictive area [3, 6]. It is based on the detection of singularities and regularities, given the experimental precision [4]. The initial representation space, which contains all the primary sites present in the trace, is progressively modulated by introducing complex sites up to the optimal representation space. At each step, the complex sites to be introduced are deduced from the previous treatment and the new SAR is performed by multiple regression analysis. The most significant parameters are selected by a stepwise regression procedure. Each new parameter introduced is the one whose introduction threshold corresponds to a 99% confidence level. In addition, the criterion of experimental precision eliminates all parameters whose contributions are lower than experimental error. The best model is that whose precision (assessed by s) is identical with experimental precision.

The predictive validity of the optimal correlation is usually checked against the compounds not belonging to the experimental key population.

A particular advantage of the DARC/PELCO method is the determination of the structural area for reliable prediction. This area, called proference, comprises all the structures generated from the population P and not belonging to it. These predictable structures are localized in the hyperstructure HS with respect to the experimental ones (fig 2). Their activity is predicted by means of direct interpolation in HS, each dimension of which refers to a site existing in an experimental compound. Different levels of reliability are determined depending on the extent to which the structure is surrounded by experimental compounds [3]. In some cases, one can use extrapolation and this gives rise to 'pseudoproference' [4]. These predictions are less reliable but are very useful for orienting experimental plans.

The DARC/PELCO method is also well suited for determining structures with optimal activity. It combines the structural features that increase the property the most. It should be interesting to compare the DARC/PELCO and OASIS predictions of structures thus found.

DARC/PELCO and OASIS models for pKa and antitumor activity of purine derivatives

pKa models

Purine derivatives are an important class of potential anticancer drugs. In spite of the fact that nowadays only mercaptopurine and its guanine analogue are used clinically, many purine derivatives are tested for their anticancer properties [68, 69]. In this respect it is worthwhile to study the physicochemical properties and physiological behaviour of purines. An important characteristic of this kind is the pKa. Tomasik *et al* [70], using the linear-free-energy method correlated the pKa values of monosubstituted purines with the Hammett σ_m-constant. Predictions for pKa values of some aminopurines are reported [71] on the basis of a Hammett equation for pyridines. Later, by making use of an extended set of substituent constants, Neiman and Quinn [72] derived a general linear-free-energy relationship in terms of σ_m and σ_p for 45 monosubstituted purine derivatives. They used it to predict the pKa of 33 polysubstituted purines.

This experimental set constitutes a good basis for

comparing DARC/PELCO and OASIS methods. We must note, however, that pKa values are taken from different sources [72].

The test set of 40 purines examined in the present work is presented in table II. They are ordered according to the generation law used to describe their structures in the DARC/PELCO method. The three SO_3H and SO_2CH_3 derivatives of the original set [72] are not considered due to the limitations of the MNDO method in treating molecules having sulfur in a hypervalent state [73]. Similarly, two $N+(CH_3)_3$ derivatives are eliminated in order to avoid the presence of an additional ionized atom. All the pKa values were determined in water except for compound **4**, but they were measured either by potentiometry or spectrophotometry. Some values that seemed unreliable in the original set [72] were taken from elsewhere (see footnotes of table II). Experimental precision varies from 0.01 to 0.1 with an average of about 0.05.

One of the best OASIS models correlating purine pKa values as anions is given below together with the following statistics: the regression coefficient confidence intervals, according to Student's *t*-test at a 95% confidence limit; correlation coefficient, *r*, standard deviation, *s*, Fisher value, *F*, regression coefficient ranges, as well as mean square deviation between experimental and predicted values obtained after the 'leave-one-out' procedure, *s'*.

pKa (anions)

$$= 17.30\ (\pm 1.76) - 26.98\ (\pm 5.20)S^N(4) + 2.26\ (\pm 2.16)q(2)$$
$$16.65 \div 17.95 \quad -25.06 \div -29.07 \quad 1.96 \div 2.81$$

$$- 2.19\ 10^{-4}\ (\pm 1.01\ 10^{-4})FW2(8)$$
$$- 1.77\ 10^{-4} \div -3.00\ 10^{-4} \qquad (5)$$

$R = 0.903$, $s = 0.514$, $F = 53.12$, $s' = 0.560$, $n = 40$.

Two of the best OASIS models for pKa of purines as cations with comparable statistics are:

pKa (cations)

$$= -51.24\ (\pm 11.8) + 100.83\ (\pm 36.48)S^E(7)$$
$$-49.21 \div -55.13 \quad 93.18 \div 118.26$$

$$+ 87.05\ (\pm 44.10)S^E(5) + 1.47\ 10^{-2}\ (\pm 0.94\ 10^{-2})FW(8)$$
$$79.27 \div 101.59 \quad 1.01\ 10^{-2} \div 1.48\ 10^{-2} \qquad (6)$$

$R = 0.876$, $s = 0.623$, $F = 33.09$, $s' = 0.699$, $n = 34$

pKa (cations)

$$= -28.78\ (\pm 10.91) + 189.90\ (\pm 48.29)S^E(5) - 70.48\ (\pm 36.13)S^E(8)$$
$$-26.85 \div -32.42 \quad 184.37 \div 210.28 \quad -64.17 \div -78.67$$

$$+ 1.71\ 10^{-2}\ (\pm 1.10\ 10^{-2})FW(8)$$
$$1.42\ 10^{-2} \div 2.12\ 10^{-2} \qquad (7)$$

$R = 0.830$, $s = 0.722$, $F = 22.07$, $s' = 0.779$, $n = 34$.

The accuracy of the models is clearly not very high. The average deviations between the observed and calculated (according to (5) and (6) pKa values are 0.39 and 0.42, respectively (table II).

On the basis of the OASIS models, we can conclude as follows:

i) the electronic characteristics of the reference structure atoms conditioning purine pKa are: pKa (anions): the acceptor superdelocalizability index at position 4, $S^N(4)$. This parameter plays a predominant role with respect to purine acidity (the ratio between the $S^N(4)$, FW(8) and q(2) coefficients included in the model with the normalized parameters, is 0.76; 0.33; 0.14). The negative sign of the regression coefficients for $S^N(4)$ indicates that the decrease in the electron acceptor character of this atom will increase the anionic pKa (the MNDO quantum-chemical calculations are based on ideal molecular geometry); the charge of atom 2, q(2); pKa(cations): the donor superdelocalizability indices at positions 5, 7 and 8 of the imidazole fragment, $S^e(5)$, $S^e(7)$, $S^e(8)$. The different signs of the corresponding regression coefficients reflect the complicated mechanism of the electron density distribution within the imidazole moiety upon cation formation.

The foregoing conclusions are justified by the well known fact that the acidity and basicity of molecules are conditioned by their electron-acceptor and electron-donor abilities.

ii) The fragment molecular mass at position 8 (FW8), may be assumed to be related indirectly to the electronic factor (through the molar polarizability).

The OASIS system also abtains a series of models in which the calculable electronic and geometric parameters arte combined with the Hammett σ-constants. The accuracy of these models is less than that of the foregoing ones. Unfortunately, in repeating the Neiman and Quinn calculations [72] with the same correlation sample and substituent constants, we could not reproduce their results. We found $r = 0.843$ and $s = 0.643$ instead of $r = 0.931$ and $s = 0.535$ [72]. The negative coefficients of the Hammett substituent constants in all models reflect the well known fact that the electron-withdrawing groups increase purine acidity.

DARC/PELCO models that relate purine acidity to purely structural parameters are given below:

pKa (anions)

$$= 8.96\ (\pm 0.03) + 0.17\ (\pm 0.01)\ \Sigma c,N\text{–}2,6 + 0.42\ (\pm 0.05)\ \Sigma c,N\text{–}8 \qquad (8)$$
$$- 0.40\ (\pm 0.02)\ \Sigma s,X\text{–}2,6 - 0.48\ (\pm 0.01)\ \Sigma o,s,X\text{–}8 + 0.11\ (\pm 0.02)\ \Sigma s$$

$R = 0.999$, $s = 0.05$, $F = 4748$, $n = 40$.

pKa (cations)

$$= 2.43\ (\pm 0.03) + 0.45\ (\pm 0.01)\ \Sigma N + 0.25\ (\pm 0.02)\ \Sigma c,o\text{–}8{\cdot}s\text{–}8 \qquad (9)$$
$$- 0.25\ (\pm 0.02)\ \Sigma o,s\text{–}2,6 - 0.35\ (\pm 0.01)\ \Sigma X$$

$R = 0.999$, $s = 0.07$, $F = 3966$, $n = 34$.

Table II. Observed and calculated pKa values of monosubstituted purines. The observed values are reported by Neiman and Quinn [72] except for the ones specified by a superscript.

N°	R	pKa (anions)			pKa (cations)		
		Observed	Calculated		Observed	Calculated	
			OASIS	DARC/PELCO		OASIS	DARC/PELCO
1	H	8.93	9.46	8.96	2.39[a]	2.40	2.43
2	2-CH_3	9.10	9.42	9.13	-	-	-
3	2-NH_2	9.93	10.22	9.99	3.80	3.95	3.78
4	2-F	8.17	8.23	8.16	-	-	-
5	2-Cl	8.21	7.21	8.16	0.69	1.45	0.66
6	2-C_6H_5	9.60	9.64	9.64	-	-	-
7	6-CH_3	9.02	9.58	9.07	2.60	2.48	2.67
8	6-NH_2	9.83[b]	10.11	9.82	4.25[b]	3.63	4.23
9	6-Cl	7.88	7.84	7.76	0.45	1.35	0.31
10	6-CF_3	7.35	7.46	7.38	-	-	-
11	6-CN	6.88	7.96	6.96	0.30	0.72	0.31
12	6-CHO	8.80	8.60	8.85	2.40[c]	1.25	2.43
13	8-CH_3	9.37	9.50	9.38	2.85	2.84	2.92
14	8-NH_2	9.36	9.58	9.38	4.68	4.36	4.68
15	8-Cl	6.02	6.06	6.08	1.77	1.68	1.72
16	8-C_6H_5	8.09	8.35	8.00	2.68	4.07	2.67
17	8-CF_3	5.12[d]	6.09	5.12	1.00	0.48	1.01
18	8-COOH	9.37	8.06	9.38	-[e]	-	-
19	2-$NHCH_3$	10.32	10.01	10.33	4.01	3.71	4.03
20	2-OCH_3	9.20	9.11	9.19	2.44	2.38	2.43
21	2-SCH_3	8.91	8.55	8.90	1.91	2.16	1.93
22	2-$N(CH_3)_2$	10.22	10.11	10.22	4.02	3.74	4.03
23	6-$NHCH_3$	9.99	10.04	9.99	4.18	3.51	4.23
24	6-OCH_3	9.16	9.35	9.24	2.21	2.26	2.18
25	6-SCH_3	8.74[d]	9.17	8.73	1.63	2.13	1.68
26	6-NHOH	9.83	9.55	9.82	3.80	2.92	3.74
27	6-$N(CH_3)_2$	10.50	10.01	10.50	3.87	3.45	3.98
28	8-$NHCH_3$	9.56	9.18	9.55	4.78	4.27	4.68
29	8-OCH_3	7.73	8.61	7.69	3.14	2.85	3.17
30	8-SCH_3	7.67	8.34	7.69	2.95	2.82	2.92
31	8-CH_2OH	8.79	8.55	8.79	2.62	1.83	2.67
32	8-$N(CH_3)_2$	9.73	9.15	9.72	4.80	4.54	4.68
33	2-OC_2H_5	9.47	9.16	9.47	2.46	2.44	2.43
34	2-SC_2H_5	9.19	8.62	9.18	-	-	-
35	6-OC_2H_5	9.52	9.39	9.41	2.13	2.31	2.18
36	6-SC_2H_5	8.86	9.23	8.90	1.72	2.23	1.68
37	6-$CONHCH_3$	8.90	8.52	8.91	1.00	1.02	1.01
38	6-$NHCONH_2$	9.95	9.75	9.93	2.35	3.09	2.47
39	8-SC_2H_5	7.72	7.15	7.69	3.04	3.08	2.92
40	6-$NHCOOC_2H_5$	9.63	9.84	9.60	2.40	3.90	2.47

[a]This basic pKa value collected from the same sources as the acidic value [74] and taken as a reference in most of the articles, is more reliable than the one reported by Quinn which corresponds to an apparent pKa value; [b]these values [75] are more reliable than those collected by Quinn in [76] which have been taken as equal to the half-neutralization; [c]value collected in [77]; [d]value reported in [78] and [74], respectively; [e]the value reported by Quinn corresponds to the first acidic pK [79] and has been suppressed.

The structural variables are equivalence sites. They are interactively constructed by grouping sites belonging to homogeneous series of the same type and according to their contributions to the pKa estimated in the exploratory correlation [4, 6, 67]. In some cases, these equivalence sites use multiplying factors [14]. The optimal parameters are chosen by stepwise regression.

The results are presented in the form of an 'activity map' called a topo-information diagram (figs 4 and 5).

These two models whose statistical criteria are excellent have a precision comparable to that of the measurements (0.05). The average deviations between the observed pKa values and those calculated according to (8) and (9) are 0.03 and 0.05, respectively (table II).

The structural variables selected in the two models group a large number of modifications that suggest a structural interpretation. The 2 and 6 substituents have a similar influence on pKa, especially for cations, and the 8-substituents lower the pKa of anions to a greater extent. This is coherent with the fact that N(1) is the site of protonation in purine and that the negative charge is shared between N(7) and N(9) [80].

The main contributions to the pKa come from the first environment elements. Amino substituents always enhance both anions and cations pKa's, whereas fluorine, chlorine, trifluoromethylm and cyano groups lower them. This is consistent with their electron donor or attractor properties. The influence of the phenyl group, which can serve either as an electron-source or as an electron-sink, depends both on its position in the purine ring and on the species ofcations or anions. The structural modifications arising from these elements, such as chain-lengthening, generally have little effect. However, the influence of the NH_2, $CONH_2$ and COOEt groups is similar to that of the halogens, directly substituting the purine ring and strongly limiting the pKa of cations.

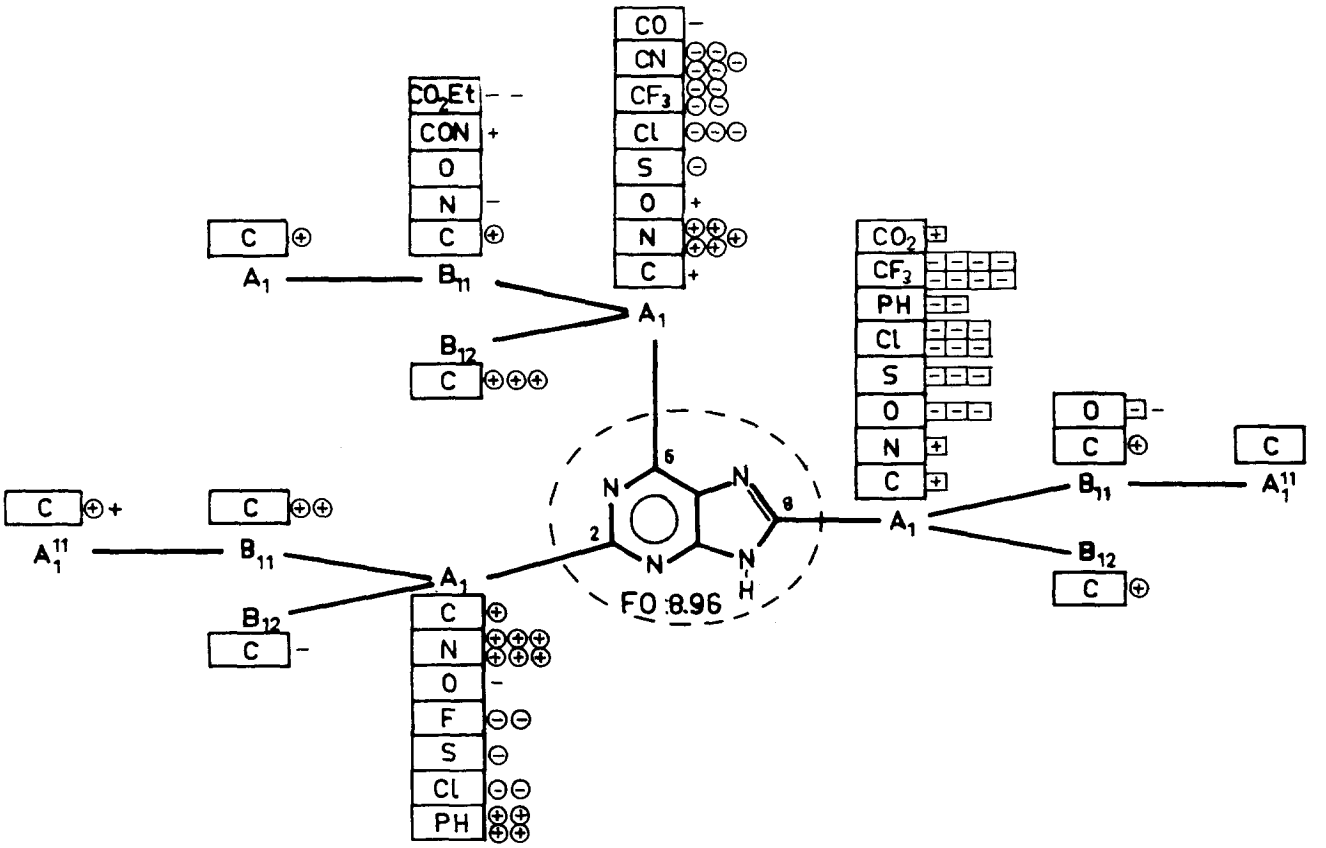


Fig 4. DARC/PELCO model for the pKa of purine anions. The equation of the topo-information relationship is expressed as: pKa = 8.96 + 0.17 $\Sigma c,N$–2,6 + 0.42 $\Sigma c,N$–8 – 0.40 $\Sigma s,X$–2,6 – 0.48 $\Sigma o,s,X$–8 + 0.11 Σs.
⊞: + 0.42; ⊕: + 0.17; +: + 0.11; ⊟ – 0.48; ⊖= – 0.40; –: – 0.11.

Table III. Observed and predicted OASIS and DARC/PELCO pKa values for some two and trisubstituted purines.

N°	Substituent at position			pKa (anions)					pKa (cations)				
	2	6	8	Observed	OASIS Eq/ 5	ΔpKa	DARC/PELCO Eq/ 8	ΔpKa	Observed	OASIS Eq/ 6	ΔpKa	DARC/PELCO Eq/ 9	ΔpKa
41	CH_3	CH_3	CH_3	9.90	9.62	0.28	9.66	0.24	4.49	3.08	1.41	3.42*	1.07
42	CH_3	CH_3	OC_2H_5	8.74	7.90*	0.84	7.97	0.77	4.71	3.78	0.93	3.67*	0.94
43	CH_3	$NHCH_3$	-	-	10.01*	-	10.16	-	5.08	3.65*	1.33	4.48*	0.60
44	Cl	Cl	-	7.06	4.55**	2.51	6.96	0.10	(-1.16)	0.53*	1.69	-1.45	0.29
45	Cl	Cl	Cl	3.96	0.77**	3.19	4.08	0.12	(-3.10)	0.01**	3.11	-2.16	0.94
46	NH_2	-	C_6H_5	9.20	9.05*	0.15	9.03	0.17	3.98	5.49*	1.41	4.03	0.05
47	NH_2	NH_2	-	10.77	10.82*	0.05	10.84	0.07	5.09	5.09**	0.00	5.59	0.50
48	NH_2	NH_2	NH_2	10.79	10.89*	0.10	11.26	0.47	(6.23)	7.22**	0.99	7.84	1.61
49	NH_2	CF_3	-	8.87	8.60	0.27	8.39	0.48	1.85	1.79	0.06	1.68*	0.17
50	NH_2	-	CF_3	6.14	7.30*	1.16	6.15	0.01	2.59	2.06*	0.53	2.37*	0.22
51	NH_2	CF_3	CF_3	5.02	4.36*	0.66	4.55	0.47	0.30	0.22*	0.08	0.28*	0.02
52	NH_2	NH_2	CF_3	7.55	8.28*	0.63	7.00	0.55	3.68	3.03	0.65	4.17	0.49
53	NH_2	-	SCH_3	8.48	9.26	0.78	8.72	0.24	4.40	4.29	0.11	4.28	0.12
54	$N(CH_3)_2$	CH_3	-	10.32	10.24	0.08	10.33	0.01	4.14	3.82	0.32	4.28	0.14
55	$N(CH_3)_2$	C_2H_5	-	10.73	10.25	0.48	10.50	0.23	4.42	3.85	0.43	4.28	0.14
56	$N(CH_3)_2$	CH_3	CH_3	10.83	10.27	0.56	10.75	0.08	4.92	4.24	0.68	4.77	0.15
57	$N(CH_3)_2$	CH_3	C_2H_5	11.11	10.15	0.04	10.92	0.19	5.05	4.50	0.55	4.77	0.28
58	$N(CH_3)_2$	C_2H_5	CH_3	-	10.28*	-	10.67	-	4.90	4.27	0.63	4.77	0.13
59	SCH_3	-	CH_3	9.58	8.63*	0.85	9.32	0.26	2.83	2.61	0.22	2.43	0.40
60	SCH_3	C_2H_5	-	9.35	8.76*	0.59	9.18	0.17	2.50	2.30	0.20	2.18	0.32
61	SCH_3	-	C_2H_5	9.75	8.53*	0.22	9.49	0.26	2.80	2.88	0.08	2.43	0.37
62	SCH_3	CH_3	CH_3	9.70	8.80*	0.90	9.43	0.27	3.04	2.70	0.34	2.67	0.37
63	SCH_3	CH_3	C_2H_5	9.75	8.69*	1.06	9.60	0.15	3.08	2.96	0.12	2.67	0.41
64	SCH_3	C_2H_5	CH_3	9.52	8.83*	0.69	9.60	0.08	3.05	2.73	0.32	2.67	0.38
65	SCH_3	CH_3	CF_3	5.67	5.06*	0.61	5.17	0.50	1.25	0.65	0.60	0.77	0.48
66	SCH_3	-	SCH_3	7.73	7.34*	0.39	7.63	0.10	2.19	2.62	0.43	2.43	0.24
67	SCH_3	CH_3	SCH_3	7.94	7.55	0.39	7.74	0.20	2.74	2.70	0.04	2.67	0.07
68	SC_2H_5	CH_3	-	9.43	8.87	0.56	9.29	0.14	2.67	2.37	0.30	2.18	0.49
69	SC_2H_5	-	CH_3	9.50	8.77	0.73	9.60	0.10	2.71	2.74	0.03	2.43	0.28
70	SC_2H_5	-	C_2H_5	9.62	8.66	0.96	9.77	0.15	2.81	3.01	0.19	2.43	0.38

*pKa values obtained by short-range extrapolations; **pKa values obtained by long-range extrapolations; ( ) observed pKa values far from the range of the studied population (0.30, 4.80).

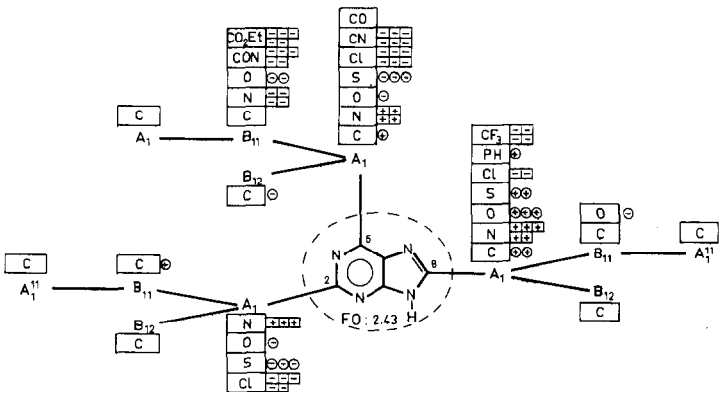


Fig 5. DARC/PELCO model for the pKa of purine cations. The equation of the topo-information relationship is expressed as: pKa = 2.43 + 0.45 ΣN + 0.25 $\Sigma c,o$–8, s–8 – 0.25 $\Sigma o,s$–2,6 – 0.35 Σx. ⊞: + 0.45; ⊕: + 0.25; ⊖: – 0.25; ⊟= – 0.35.

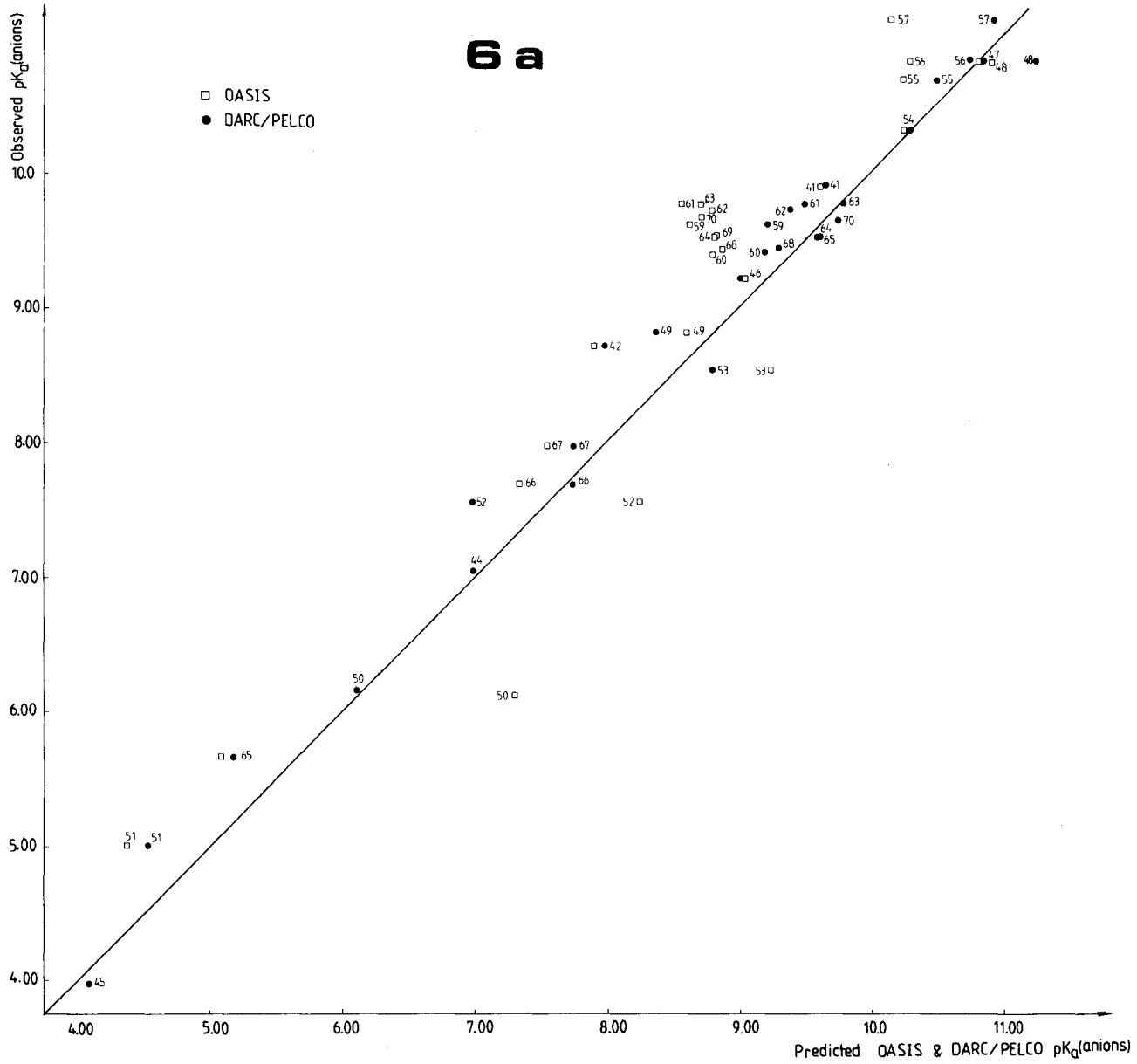
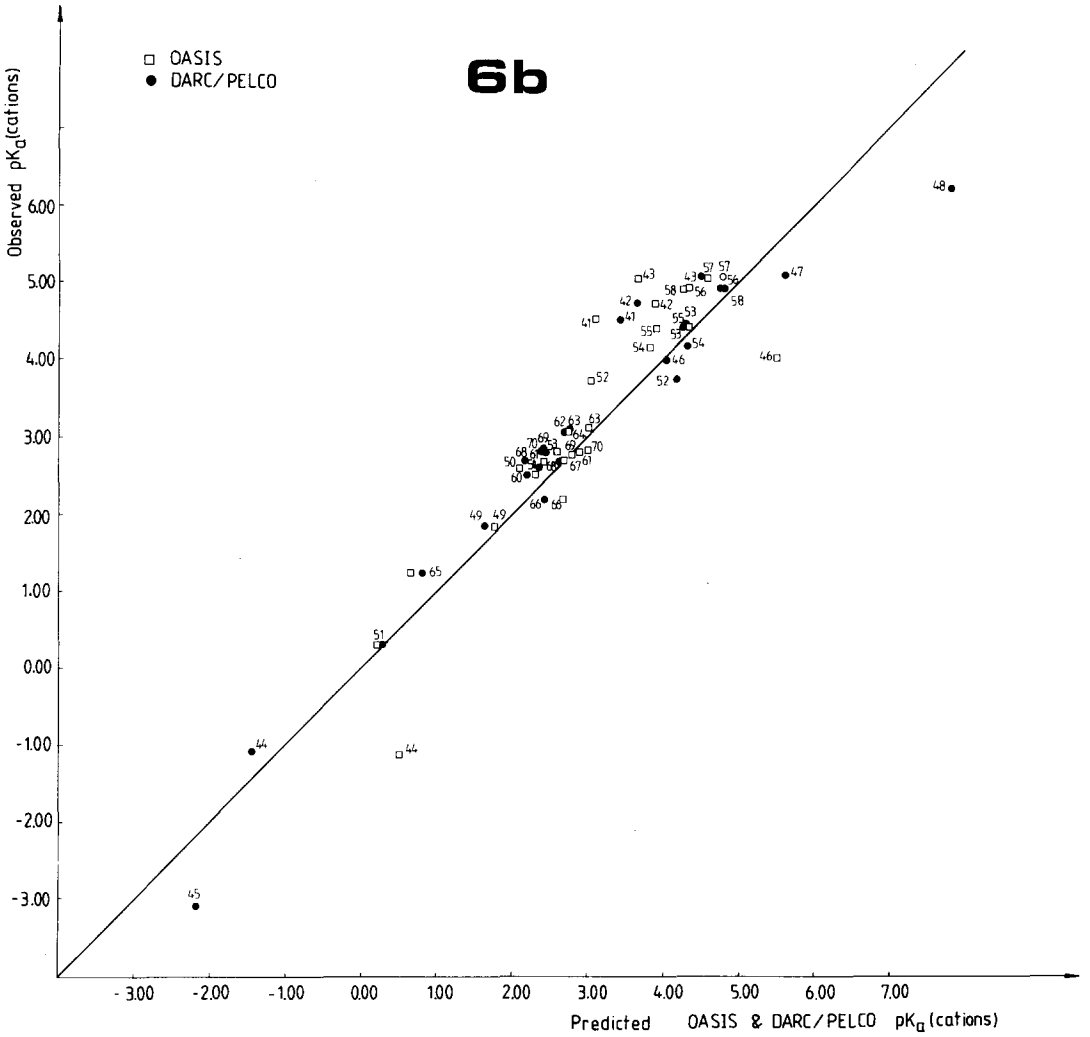


Fig 6. Plot of the DARC/PELCO predicted pKa values *versus* OASIS ones for some two and trisubstituted purines. **a.** pKa(anions). **b.** pKa(cations).

The pKa values of 30 di- and tri-substituted purines are calculated by OASIS and DARC/PELCO models and compared with the observed values taken from [72] (table III). Three of the derivatives given in the original sample are omitted here, because the DARC/PELCO method can predict both, the anionic and cationic pKa's of these compounds by extrapolation alone. The agreement between measured and predicted values according to the OASIS and DARC/PELCO methods is reasonable, considering the variation in the experimental data from different analytical methods. The large deviations between the experimental and predicted pKa values are essentially observed for long-range extrapolations (see the OASIS anionic and cationic values for compounds **44** and **45**). They concern either the range of the OASIS model variables (**44** and **45**) or that of the observed values (cationic values for **44**, **45** and **48**).



The results presented in table III are illustrated also in figures 6a and 6b where the predictions obtained by long-range extrapolations are omitted.

All the pKa (anions) values predicted by the DARC/ PELCO method are obtained by interpolation. This is not the case with the OASIS predictions where eleven purines are obtained by extrapolation close to the range of model variable variation and two by long-range extrapolation (**44** and **45** not included in fig 6a). For this reason, the average deviation between the observed and predicted OASIS pKa(anions) value is much larger (s = 0.66) than with the DARC/PELCO method (s = 0.23). However, the interpolated OASIS predictions are still less precise (0.44 *vs* 0.23). There are only three short-range and three long-range OASIS extrapolations for pKa (cations) *versus* five DARC/PELCO extrapolations. Moreover, the observed pKa values for **44**, **45** and **48** are far from the range of the population studied. For this reason, the mean deviations between the observed values and those predicted by both methods are not satisfactory: s: 0.60 for OASIS and s = 0.40 for DARC/PELCO (see fig 6b). Again, the interpolated DARC/PELCO predictions are more precise (0.20 *vs* 0.45).

The above agreement of the calculated OASIS and DARC/PELCO pKa values prompts us to compare the predictions by both methods for ten other disubstituted purines, given in table IV. In the same table, the pKa values obtained by the OASIS models 8 and 9 are compared with DARC/PELCO ones, calculated by (11) and (12), respectively. The results of the comparison are also illustrated in figures 7a and 7b.

As seen, the DARC/PELCO and OASIS predictions coincide fairly well, with one exception for the anionic pKa value of **78** (not presented in fig 7a), where the OASIS pKa value is again obtained by extrapolation far from the range of variation of model variables.

Table IV. Predicted OASIS and DARC/PELCO pKa values for some purine derivatives.

N°	R	pKa (anions)		pKa (cations)	
		OASIS Eq/ 5	DARC/PELCO Eq/ 8	OASIS Eq/ 6	DARC/PELCO Eq/ 9
71	2-NH_2, 8-OCH_3	9.47	8.71	4.40	4.53
72	2-NH_2, 8-CH_3	10.26	10.40	4.37*	4.28
73	2-NH_2, 8-NH_2	10.32*	10.40	5.99*	6.03
74	2-NH_2, 8-CH_2OH	9.42	9.81	3.31	4.03
75	6-CH_3, 8-NH_2	9.70	9.49	4.42*	4.93
76	6-CH_3, 8-OCH_3	8.78	7.80	2.92	3.43
77	6-CH_3, 8-Cl	5.71*	6.19	1.76	1.98
78	6-NH_2, 8-Cl	1.77**	6.93	2.83	3.53
79	6-NH_2, 8-CH_3	10.14	10.23	4.05	4.73
80	6-Cl, 2-NH_2	9.02	8.78	2.90	1.68

*pKa values obtained by short-range extrapolations; **pKa values obtained by long-range extrapolations.

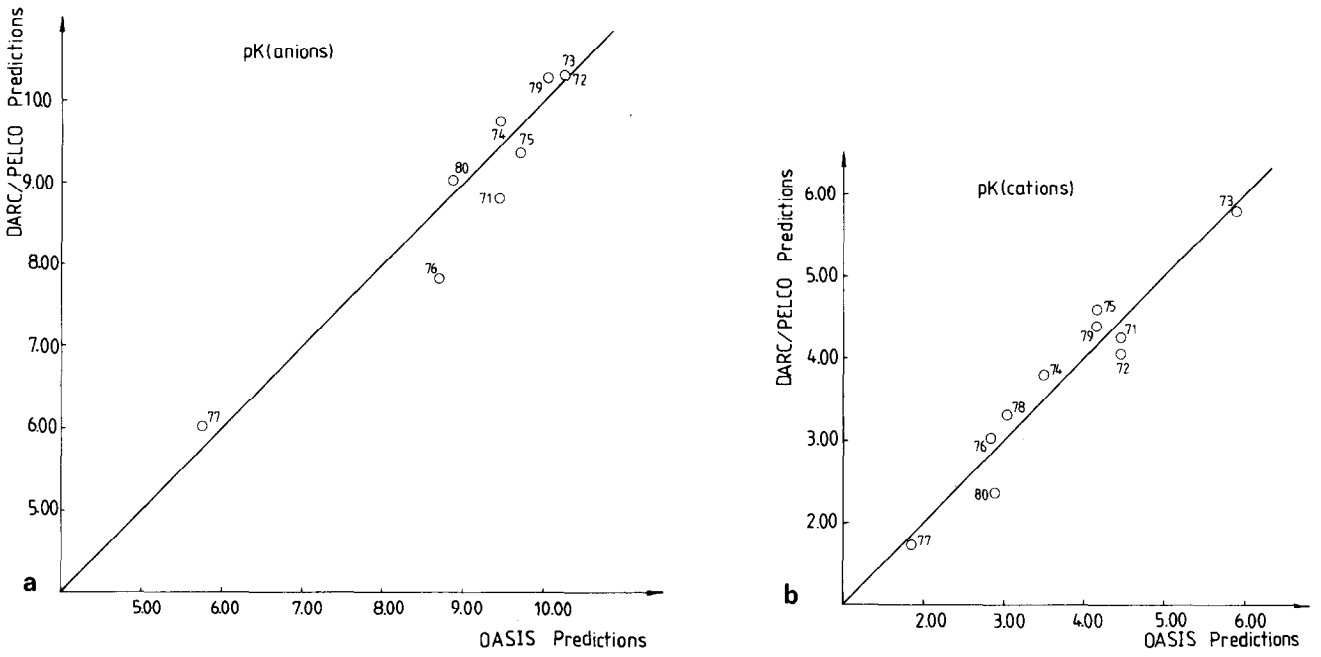


Fig 7. Plot of the DARC/PELCO predicted pKa values *versus* OASIS ones for some two substituted purines. **a.** pKa (anions). **b.** pKa (cations).

Modeling antitumor activity of purines again adenocarcinoma CA 755

DARC/PELCO and OASIS methods are applied in this work to the interaction of 2- and 6-substituted purines with the murine solid tumor adenocarcinoma CA 755. The modeling of this biological system is important because it was proved that a number of purines are carcinostatic and even carcinolytic. Recently, Hansch type QSARs were derived [81] involving the hydrophobic parameter π, the Swain-Lupton field and resonance constants F and R used as electronic parameters and the molar refractivity MR, as a measure of the geometric properties. However, the use of such a restricted set of parameters, as well as the fact that the purine ring system is not typically aromatic, results in models accounting for a small fraction of the biological data variance ($r = 0.811$, $s = 0.376$).

The sample of eighteen 2- and 6-mono- and di-substituted purines is presented in table V. Five of the derivatives listed in the original data set [81] are not included here: three iodo derivatives, owing to the lack of reliable CNDO parameters for iodine, as well as two larger purine derivatives, due to the microcomputer quantum-chemical program limitations concerning the size of the molecule. The activity of purines is described in terms of log $1/C$, where C is the concentration (in mol/kg) which produces a tumor mass regression of 80%. The unsubstituted purine is also eliminated from our test set because there is no definite measure for its biological activity, though it was reported to be inactive.

Below we present the best OASIS model obtained (with normalized parameters) [82]:

$$\log 1/C = 3.69\ (\pm 0.14) + 0.51\ (\pm 0.14)S^E(6) + 0.24\ (\pm 0.14)\pi(6) \qquad (10)$$

$n = 17$, $r = 0.920$, $s = 0.265$, $F = 38.57$, $s' = 0.298$.

The equations including $S^E(6)$ in a quadratic form statistically produce more accurate results (with respect to r, s and F). However, the 95% confidence interval for this parameter is comparable in value to the coefficient itself. Thus, the search for parabolic models requires a larger correlation sample (correlations displaying extrema for the biological system are presented in [82]). The calculated purine activity according to (10) is given in table V.

The preceding OASIS model indicates that: i) biological interaction is not affected by steric hindrance; ii) the essential electronic process conditioning the activity of purines is most likely charge transfer from position 6 on the purine fragment to the biomacromolecule. This assertion is supported by the large positive coefficients related to the $S^E(6)$ parameter in (10) obtained by using CNDO calculation based on idealized molecular geometry). In fact, all purine derivatives with high positive values of the standardized donor superdelocalizability index $S^E(6)$ possess high CA 755 antitumor potency. On the contrary, purines with negative $S^E(6)$ values are relatively inactivate. This result confirms Neiman and Quinn's conclusion [81] that purine CA 755 activity is predominantly determined by the resonance effect of the substituent at position 6 (resonance effects are associated with charge transfer processes); iii) the hydrophobic interactions play a subordinate role in determining CA 755 antitumor activity of purines as can be judged from the small $\pi(6)$ term coefficient.

These OASIS findings justify the following hypothesis. Before charge transfer can take place, the purine ring is fixed to the receptor cavity by means of a preliminary hydrophobic interaction at position 6 of the purine moiety.

The model's predictive ability is estimated by the 'leave-one-out' procedure. All equations obtained after successive elimination of the compounds converge. The deviation intervals of the three coefficients are smaller than the respective 95% confidence intervals according to the t-test of (10): 3.69 ± 0.03, 0.51 ± 0.04 and 0.24 ± 0.05.

The proposed DARC/PELCO model from the same experimental population:

$$\log 1/C = 3.20\ (\pm 0.51)\ \Sigma_{N,0-6} + 4.20\ (\pm 0.50)\ \Sigma_{s,Cl,\ Br,SO2-6} - 0.83\ (\pm 0.36)\ \Sigma_{Br,SO2F-2} \qquad (11)$$

$R = 0.979$, $s = 0.237$, $F = 110$, $s' = 0.278$.

is as precise as the measurements (± 0.20). This model has about the same accuracy (average deviation 0.21) as the OASIS one (average deviation 0.25) but is again more significant ($R = 0.98$ against $R = 0.92$). The activity values calculated according to (11) are listed in table V. The DARC/PELCO model (fig 8) again confirms the importance of the purine position 6. The large positive coefficients of the two first terms, as well as the negative coefficient of the last term of the model, indicate that the structural changes at position 6 increase activity while those at position 2 decrease it.

The mean square deviation, s', between the observed and predicted activities after the 'leave-one-out' procedure is comparable with that obtained for the OASIS model (0.278 against 0.298). The DARC/PELCO and OASIS predicted log $1/C$ values after the 'leave-one-out' procedure are plotted against the observed experimental data in figure 9. Given the low precision of the experimental CA 755 activities, the coincidence between the predictions of both methods and observed potency can be qualified as good.

Table V. The observed CA 755 antitumor activity of the studied purines compared with the calculated and predicted ones.

N°	R	Purines CA 755 antitumor activity, log 1/C				
		Observed	Calculated		Predicted	
			OASIS Eq/ 10	DARC/PELCO Eq/ 11	OASIS Eq/ 10	DARC/PELCO Eq/ 11
1	6-Cl	4.23	4.28	4.20	4.29	4.20
2	6-Br	4.50	4.36	4.20	4.33	4.16
3	6-OCH_3	3.00	3.42	3.20	3.45	3.24
4	6-$OCH_2CH_2CH_3$	3.42	3.73	3.20	3.78	3.16
5	6-$NH-NH_2$	3.22	3.24	3.20	3.25	3.20
6	6-SCH_3	4.51	4.37	4.20	4.35	4.16
7	6-SCH_2CH_3	4.45	4.50	4.20	4.51	4.17
8	6-SO_2CH_3	4.04	4.07	4.20	4.09	4.22
9	6-SO_2NH_2	4.26	4.00	4.20	3.84	4.19
10	2-CH_3, 6-NH_2	3.23	3.06	3.20	3.02	3.19
11	2-Cl, 6-$NHCH_3$	3.37	3.07	3.20	2.99	3.17
12	2-Br, 6-NH_2	2.56	3.12	2.37	3.23	2.13
13	2-NH_2, 6-Cl	4.05	3.77	4.20	3.74	4.22
14	2-NH_2, 6-Br	3.91	3.85	4.20	3.84	4.23
15	2-NH_2, 6-SO_2CH_3	4.04	4.19	4.20	4.21	4.22
16	2,6-$NH-NH_2$	2.77	2.52	3.20	2.41	3.28
17	2-SO_2F, 6-Cl	3.18	3.19	3.37	3.19	3.61

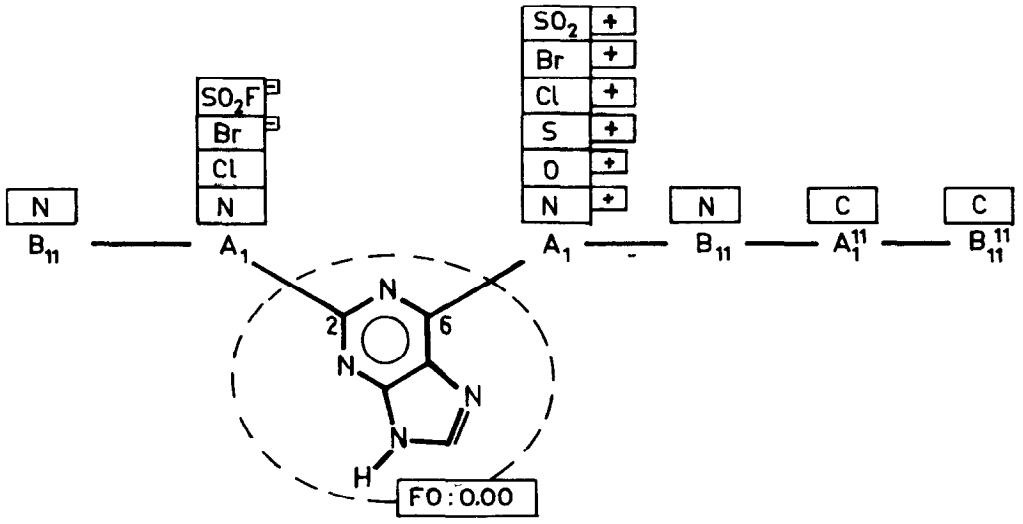


Fig 8. DARC/PELCO model for antitumor activity of purines. The equation of the topo-information relationship is expressed as: log $1/C = 3.20\ [N, O] - 6 + 4.20\ [S, Cl, Br, SO_2F] - 6 - 0.83\ [Br, SO_2F] - 2$. ⊞: 4.20; ⊞: + 3.20; ⊟= − 0.83.

Conclusion

Two QSAR methods with completely different backgrounds are compared in the present work. The DARC/PELCO is one of the best optimized of the purely structural methods. It is based on the exhaustive generation of all topochromatic sites around the reference structure and the evaluation of their contribution to the biological property. On the other hand, the OASIS system can be considered as one of the best optimized versions of the extended Hansch type physicochemical methods. It deals with a large set of the most important and relatively independent structural variables including topological, steric and quantum-chemical descriptors, as well as molecular or fragment physichochemical properties. The OASIS descriptors are attributed either to the reference structure atoms or to the molecules as a whole. Indeed the DARC/PELCO and OASIS variables complement each other in characterizing molecular structure. The OASIS system reflects in detail the molecular electronic structure and metrics, while the DARC/PELCO method describes molecular topology exhaustively and systematically.

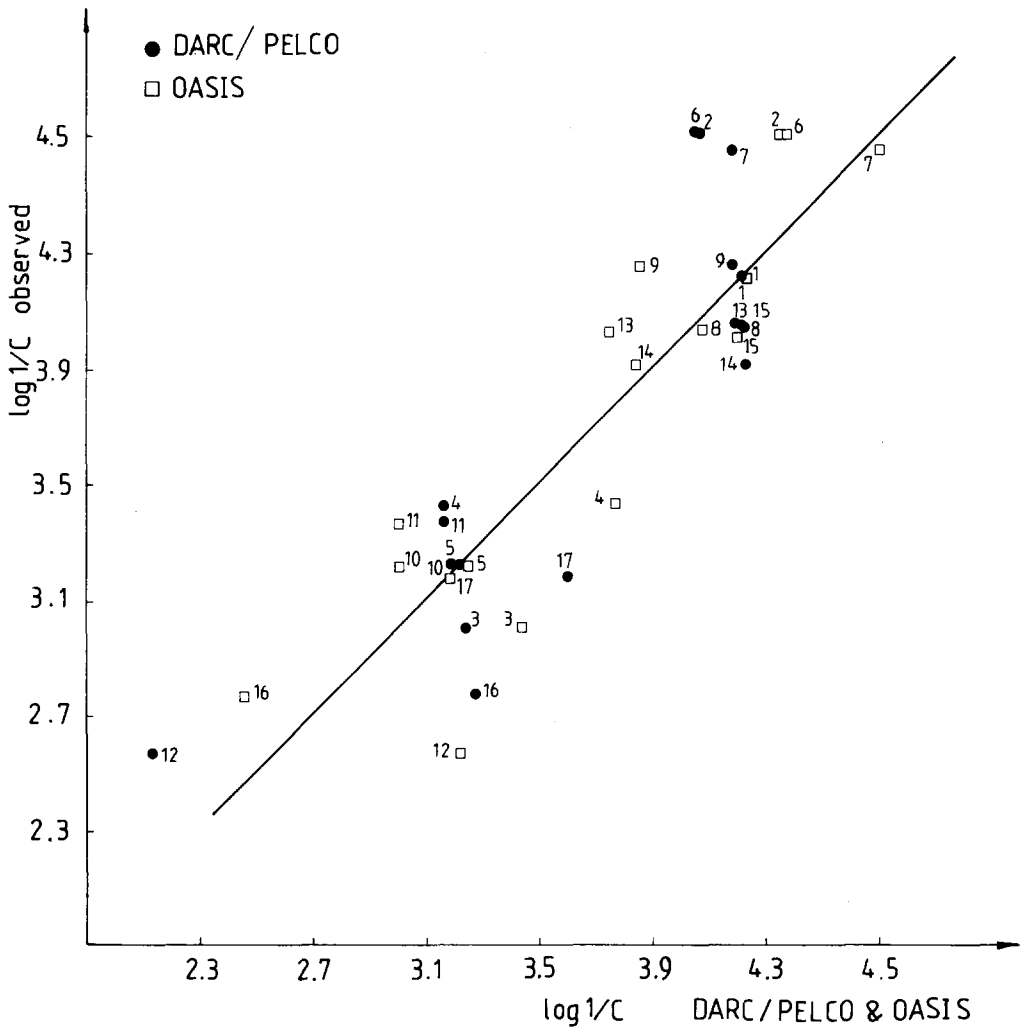


Fig 9. Plot of the observed CA 755 antitumor activity of the studied purine derivative *versus* predicted ones (after 'leave-one-out' analysis) by DARC/PELCO and OASIS models.

Being structurally exhaustive, the DARC/PELCO variables provide higher accuracy of the structure–activity models than the OASIS system. The number of parameters is generally slightly higher in the DARC/PELCO models than in the OASIS ones but the comparison is complex because of their different nature. However, the DARC/PELCO parameters are chosen from a large but limited set of structural elements used to describe the molecules under study. Whereas, in the OASIS system, the parameters are selected from a large and undefined number of variables. This set includes all the calculable molecular descriptors (topological and steric) and electronic indices but is limited for the substituent constants by their availability.

The DARC/PELCO variables, which group some structural modifications according to their regular influence make for a generalization of certain effects and facilitate a structural interpretation. The OASIS model, in turn, provides more details concerning the biological interaction mechanism.

As opposed to the OASIS system, the DARC/PELCO method defines the area of reliable predictions. The activities of the optimized structures according to the DARC/PELCO system can then be compared with the OASIS approach. The results of the present study indicate that, in spite of their different nature, both approaches lead to comparable prediction values for the objective properties being studied.

In conclusion, the DARC/PELCO and OASIS methods are complementary in optimizing the lead structure in congeneric series of biologically active molecules. These conclusions will stimulate further QSAR studies with the combined application of both methods.

References

1 Dubois JE, Laurent D, Vieillard H (1966) *C R Acad Sci* Paris, 263C, 764–767
2 Dubois JE, Laurent D, Aranda A (1973) *J Chim Phys* 11–12, 1608–1624
3 Dubois JE, Mercier C (1979) *C R Acad Sci Sér II* Paris, 289C, 89–92
4 Dubois JE, Sobel Y, Mercier C (1979) *C R Acad Sci Sér II* Paris, 292, 783–788
5 Dubois JE, Sicouri G, Sobel Y, Picchiottino R (1984) *C R Acad Sci Sér II* Paris, 298, 525–528
6 Mercier C, Sobel Y, Dubois JE (1991) *Annals of Mathematical Chemistry*. Vol 2, Gordon and Breach Science Publ (in press)
7 Mercier C, Fabart V, Sobel Y, Dubois JE (1990) *In: Les Systèmes d'Information en Chimie*. Société Française pour le Développement de l'Information en Chimie, Paris 28–38
8 Dubois JE, Doucet JP, Panaye A (1989) *Bull Soc Chim Belg* 98, 31–44
9 Exner O (1988) *In: Correlation Analysis on Chemical Data*, Plenum Press, 173–175
10 Doucet JP, Panaye A, Dubois JE (1983) *J Org Chem* 48, 3174–3182
11 Mercier C, Trouiller G (1989) *In: Quantitative Structure–Activity Relationship in Drug Design* (Fauchère JL, ed) Alan R Liss Inc, NY, 203–207
12 Mercier C, Trouiller G, Dubois JE (1990) *Quant Struct Act Relat* 9, 88–93
13 Dubois JE, Laurent D, Bost P, Chambaud S, Mercier C (1976) *Eur J Med Chem* 11, 225–236
14 Mercier C, Sobel Y, Dubois JE (1981) *Eur J Med Chem* 16, 473–476
15 Sobel Y, Mercier C, Dubois JE (1981) *Eur J Med Chem* 16, 477–479
16 Chrétien JR, Szymoniak J, Dubois JE, Poirier MF, Denicker P (1985) *Eur J Med Chem* 20, 315–325
17 Dubois JE, Mercier C, Panaye A (1986) *Acta Pharm Jugosl* 36, 135–169
18 Hansch C (1976) *J Med Chem* 19, 1–6
19 Hansch C, Fujita T (1964) *J Am Chem Soc* 86, 1616–1626
20 Free SM, Wilson JW (1964) *J Med Chem* 7, 395–399
21 Kier LB, Hall LH (1976) *In: Molecular Connectivity Chemistry and Drug Research*. Academic Press, NY
22 Mercier C, Dubois JE (1979) *Eur J Med Chem* 14, 415–423
23 Duperray B, Chastrette M, Makabeth MC, Pachero H (1976) *Eur J Med Chem* 11, 323–336
24 Donnay RH, Garnier F, Dubois JE (1974) *J Chim Phys* 11–12, 1418
25 Mekenyan O, Bonchev D (1986) *Acta Pharm Jug* 36, 225–237
26 Mekenyan O, Bonchev D (1985) *In: Proceedings of the 3rd International Conference on Chemistry and Biotechnology of Biologically Active Natural Products*, Sofia, Vol III, 385–392
27 Mekenyan O, Peitchev D, Bonchev D, Trinajstic N, Bangov I (1986) *Drug Res* 36, 176–183
28 Mekenyan O, Peitchev D, Bonchev D, Trinajstic N, Dimitrova J (1986) *Drug Res* 36, 629–635
29 Mekenyan O, Bonchev D, Entchev V (1988) *Quant Struct Act Relat* 7, 240–244
30 Mekenyan O, Karabunarliev S, Bonchev D (1991) *J Math Chem* (in press)
31 Randic M (1975) *J Am Chem Soc* 97, 6609–6615
32 Wiener H (1947) *J Am Chem Soc* 69, 17–20
33 Wiener H (1948) *J Phys Chem* 52, 1082–1089
34 Hosoya H (1971) *Bull Chem Soc* Jpn, 44, 2332–2339
35 Hosoya H (1972) *J Chem Docum* 12, 181–183
36 Balaban AT (1979) *Theoret Chem Acta* (Berl) 53, 355– 375
37 Balaban AT (1983) *Pure Appl Chem* 55, 199–206
38 Balaban AT (1982) *Chem Phys Lett* 89, 399–404
39 Gutman I, Trinajstic N (1972) *Chem Phys Lett* 17, 535–538
40 Gutman I, Ruscic B, Trinajstic N, Wilcox Jr CF (1975) *J Chem Phys* 62, 3399–3405
41 Bonchev D (1983) *In: Information Theoretic Indices for Characterization of Chemical Structure*. Research Studies Press, Chichester, UK
42 Polansky DE, Bonchev D (1986) MATCH, 21, 133–186
43 Polansky DE, Bonchev D (1986) MATCH, 21, 341–344
44 Randic M, Wilkins CL (1979) *Chem Phys Lett* 63, 332–336
45 Mekenyan O, Boncev D, Balaban AT (1988) *J Math Chem* 2, 347–375

46 Verloop A, Hoogenstroatem W, Tipker J (1976) *In: Drug Design*, 7, 165–187
47 Dewar MJS, Thiel W (1977) *J Am Chem Soc* 99, 4899– 4907
48 Thiel W (1981) *J Am Chem Soc* 103, 1413–1425
49 Bingham RC, Dewar MJS, Lo DH (1975) *J Am Chem Soc* 97, 1285–1293
50 Kanetti H, QCPE version (submitted)
51 Mekenyan O, Karabunarliev , Bonchev D (1990) *Comput & Chem* 14, 193–200
52 Kier LB (1981) *J Pharm Sci* 70, 930–933
53 Kier LB, Hall LH (1986) *In: Molecular Connectivity in Structure–Activity Analysis*. Research Studies Press, Letchworth, UK
54 Yee WT, Sakamoto KY, I'Haya YJ (1976) Rep Univ Electr Commun 27, 53–71
55 Yee WT, Sakamoto KY, I'Haya YJ (1977) Rep Univ Electr Commun 27, 53–71
56 Ray SK, Basak SC, Roychaudhury C, Roy AB, Ghosh JJ (1981) *Indian J Chem* 20B, 894–897
57 Basak SC, Harris DK, Magnuson VR (1984) *J Pharm Sci* 73, 429–437
58 Hansch C, Leo A (1979) *In: Substituent Constants for Correlation Analysis in Chemistry and Biology*. John Wiley and Sons, NY
59 Hansch C, Leo A, Unger S, Kim K, Nikaitani D, Lien E (1973) *J Med Chem* 16, 1207–1216
60 Taft RW (1952) *J Am Chem Soc* 74, 3120–3128
61 Topliss JG, Edwards RB (1979) *J Med Chem* 22, 1238–1244
62 Motoc I, Balaban AT, Mekenyan O, Bonchev D (1982) *J Math Chem* 13, 369–404
63 Mekenyan O, Karabunarliev St, Bonchev D (1989) *Proceedings of the III-International Conference on Mathematical Chemistry*, Galveston, TX
64 Golub GH, Reinsch C (1970) *Numer Math* 14, 403–430
65 Cramer RD, Bunce JD, Patterson DA, Franck IE (1988) *Quant Struct–Act Relat* 7, 18–27
66 Fujita T, Ban T (1971) *J Med Chem* 14, 148–152
67 Mercier C, Fabart V, Sobel Y, Dubois JE (1991) *J Med Chem* 34, 934–942
68 Lister JH (1971) *In: Purines*. Wiley-Interscience, NY, 9
69 Cone WH (1970) *In: Chemotherapy of Cancer*, Lea and Febiger, Philadelphia, PA, Chap 1
70 Tomasik P, Zalewski R, Chodzinski J (1979) *Chem Zvesti* 33, 105–109
71 Perrin DD (1965) *J Chem Soc* 5590–5595
72 Neiman Z, Quinn FR (1981) *J Pharm Sci* 70, 425–430
73 Stewart JP (1989) *J Comput Chem* 10, 209, 221
74 Albert A, Brown DJ (1954) *J Chem Soc* 2060–2071
75 Albert A, Serjeant EP (1960) *Biochem J* 76, 621
76 Lewin S, Tann NW (1962) *J Chem Soc* 1466–1467
77 Giner-Sorolla A, Zimmerman I, Bendich A (1959) *J Am Chem Soc* 81, 2515–2520
78 Giner-Sorolla A, Bendich A (1958) *J Am Chem Soc* 80, 5744–5749
79 Albert A (1960) *J Chem Soc* 4705–4705
80 Albert A, Serjeant EP (1971) *In: The Determination of Ionization Constant*. 2nd edn Chapman and Hall, London, UK
81 Neiman Z, Quinn FR (1982) *J Pharm Sci* 71, 618–625
82 Mekenyan O, Bonchev D, Rouvray DH, Peitchev D, Bangov I (1991) *Eur J Med Chem* 26, 305–312