Modeling the Anticancer Action of Some Retinoid Compounds by Making Use of the OASIS Method

D. Bonchev,* C. F. Mountain,† W. A. Seitz,‡ and A. T. Balaban§

M.D. Anderson Cancer Center, Houston, Texas 77030, Texas A&M University, Galveston, Texas 77553, and Organic Chemisitry Department, Polytechnic Institute, Bucharest 77206, Roumania

Received September 17, 1992

The powerful OASIS (optimized approach based on structural indices set) approach is applied to the anticancer activity of a series of vitamin A analogs. The best three- and four-variable models obtained via the OASIS technique have correlation coefficients of 0.973 vs. 0.990 and standard deviations $s^2 = 0.11$ and 0.05, respectively. The models incorporate the hydrophobicity factor log P, two geometric parameters (topological indices and/or 3-D steric ones), and the molecular dipole moment. For a set of 15 compounds studied here, the activity measured by ED_{50} was well correlated by models with approximately equal contribution of the through cell membrane transport and the geometric drug–receptor correspondence while weak nonspecific electronic interaction was also found to play some role. Comparison to previous treatments of this data is given and extension to larger sets is discussed.

Introduction

The use of theoretical methods in the search for new compounds with potential anticancer activity has been ongoing for quite some time with considerable, but mixed, success. QSAR (quantitative structure–activity relationships) methods constitute one of the widely used approaches in which drug activity is related, albeit statistically, to certain molecular descriptors, structural indices which quantify molecular structure. Graph theory is of great help in constructing various molecular descriptors called topological indices, such as the molecular connectivity indices of Randić[1] and Kier and Hall,[2,3] the topological distance of the molecule (known as the Wiener index[4]), the distance connectivity index of Balaban,[5] information–theoretic indices of Bonchev.[6] Although generally successful in drug design, the topological QSAR approach has so far had only moderate success in cancer research because it accounts for the geometric aspects of the drug–receptor interaction (and some physicochemical properties) but not for the electronic ones. On the other hand, studies using quantum mechanical indices in developing structure–activity relationships also have not been very successful in describing the anticancer action of chemical compounds. The main obstacles in the design of anticancer agents are perhaps related to the difficulties in assessing the right target of interaction. Yet, we propose that a SAR approach combining quantum chemistry with traditional chemical structure information (expressed in a graph–theoretical language) may overcome most of these difficulties.

Such a novel methodology, the OASIS approach, was developed in 1985–86 and successfully applied to quantitative assessments of drug or toxic activity of structurally related compounds.[7-14] The OASIS method has been developed into a computer package[14] yielding quantitative predictions. Its accuracy has previously been favorably compared with a well-known competitive method DARC-PELCO, from Dubois' laboratory in France.[15,16]

Retinoids (vitamin A metabolites and synthetic analogs) have been the focus of much interest over the last 20 years during which more than 1000 new compounds have been synthesized in the search for highly active anticancer drugs with low toxicity.[17-20] These compounds have many applications such as acting as chemoprevention agents, carcinogenesis inhibitors, or reverters for malignant and premalignant cell disorders.[17] There is much interest in retinoids for the treatment of "oral leukoplakia, bronchial metastasis, laryngeal papillomatosis, cervical dysplasia, preleukemia, superficial bladder cancer, and active promyelocytic leukemia...".[21]

A correlation between the topological structure and biological activity of some retinoids (TOC assay, i.e. reversion of the keratinization of hamster tracheal cells in organic cells) was published by Balaban et al.[22] Another correlation with the same group of compounds, originally studied by Newton et al.,[23] was recently reported by Niculescu-Duvaz et al. proceeding from the minimal steric difference method.[24,25]

Related to the foregoing, the present study aims to examine the applicability of the OASIS approach to retinoids and to draw some preliminary conclusions concerning the factors (geometric, electronic, and hydrophobic ones) influencing the anticancer potency of these compounds.

OASIS—The Optimal Approach Based on Structural Indices Set

The OASIS method is a second-generation QSAR approach which may be viewed partly as a generalized Hansch approach.[7-10] It deals with the molecule as a whole (but not with the "lead" molecule substituents only) and makes combined use of (i) the geometric description of the three-dimensional molecular structure, (ii) a large set of topological (graph–theoretic) descriptors termed topological indices, (iii) a detailed quantum-chemical characterization (at the SCF level) of the electronic structure and, (iv) a multivariate treatment.

The following topological indices are used: the Randić connectivity index,[1] the total distance of the graph (the Wiener index),[4] the Hosoya nonadjacency number,[26] the

* Author to whom all correspondence should be sent. On a sabbatical leave from the Higher Institute of Chemical Technology, Burgas 8010, Bulgaria.
† M.D. Anderson Cancer Center.
‡ Texas A&M University.
§ Polytechnic Institute.

0022-2623/93/1836-1562$04.00/0

Balaban centric and distance connectivity indices,[5,27] several of the Bonchev information-theoretic indices,[6,28] the Zagreb group indices,[29] the valence connectivities of Kier and Hall,[2,3] and I'Haya's electropy.[30] Their computational formulae are briefly outlined below.

In describing molecular structure, graph theory usually deals with the so-called hydrogen-depleted (or skeletal) graphs in which the hydrogen atoms are not taken into consideration. Let the number of vertices (atoms) and edges (bonds) in such a graph be denoted by N and E, respectively. Also, let a_i denote the vertex degree which is simply number of the nearest neighboring vertices to vertex i. Several topologial indices are based on these atomis graph variants:

the Zagreb group indices[29]

$$M1 = \sum_i a^2_i \quad (1)$$

$$M2 = \sum_{ij} a_i a_j \quad (2)$$

the Randić molecular connectivity index[1]

$$\chi = \sum_{ij} (a_i a_j)^{-1/2} \quad (3)$$

where the summation in eqs 2 and 3 is taken over all bonds $\{ij\}$.

The Kier and Hall valence connectivity[2,3] index is obtained by the substitution of χ^v for a in eq 3

$$\chi^v = (z^v - h)/(z - z^v - 1) \quad (4)$$

here z and z^v stand for the total number of electrons and the number of valence electrons of the atom, respectively, and h is the number of hydrogens bonded to that atom.

Several other indices used in the OASIS computational package are based on graph distances. d_{ij}, the distance between the graph vertices i and j, is an integer equal to the number of edges (bonds) traversed along the shortest path between them. The sum of the distances between all vertices in the graph is termed graph distance but is better known as the Wiener index[4]

$$W = 1/2 \sum_{ij} d_{ij} = \sum_i d_i \quad (5)$$

The second equality in eq 5 involves d_i, the sum of the distances from vertex i to all other vertices in the graph; d_i is also called the vertex distance or distance degree. The Balaban[5] distance connectivity index J makes use of eq 3, where the distance degrees are substituted for the respective vertex degrees

$$J = [E/(\mu + 1)] \sum_{ij} (d_i d_j)^{-1/2} \quad (6)$$

where the cyclomatic number μ stands for the number of graph cycles and the summation is again taken over all bonds $\{ij\}$. The Balaban[27] centric index $D2$ is the mean-square distance between all vertices in the graph.

$$D2 = [N(N-1)]^{-1} (\sum_{ij} d^2_{ij})^{-1/2} \quad (7)$$

In a recent paper[22] Balaban modified his J index to a new index MJ so as to take into account heteroatoms and multiple bonds. The graph distance between two adjacent atoms connected by a multiple bond has been divided by the bond multiplicity b (for single, double, triple, and aromatic bonds b is 1, 2, 3, and 1.5, respectively). A parametrized formula has been used for the distance degree (vertex distance) of heteroatoms

$$d'_i = 0.5 Y_i d_i = 0.5(1.1191 + 0.0160 Z_i - 0.0537 G_i) d_i \quad (8)$$

where Z_i is the atomic number of the heteroatom and G_i is the number of the group in the short form of Mendeleev's periodic system. $Y_i = 1$ for carbon, 0.925 for oxygen, and 0.963 for nitrogen.

The Hosoya nonadjacency number[26]

$$Z = \sum_k p(k) \quad (9)$$

counts the number of ways in which k nonadjacent edges are chosen from a graph. By definition $p(0) = 1$, and trivially $p(1) = E$, the number of graph edges. The summation extends to $[p/2]$, where the Gaussian brackets represent the largest integer not exceeding the real number they close.

The electropy index[30]

$$\epsilon = \ln(n!/\Pi_i n_i!) \quad (10)$$

is a measure of the degree of freedom of distributing n electrons among different skeletal subsets upon the formation of the molecule. The carbon σ-skeleton forms one such subset unless it does not incorporate a heteroatom. CH_3, CH_2, and CH groups are regarded in different subsets; the same holds for isolated π-bonds. The conjugated π-electrons, however, form a common group.

Several of the Bonchev information-theoretic indices[6,28] are also included in the OASIS computational scheme. They all are calculated by the Shannon formula for the information entropy of a finite set of N elements partitioned, according to a certain criterion, into k different subsets each one having N_i elements

$$I = -\sum_i (N_i/N) \log_2 (N_i/N), \text{ bits per element} \quad (11)$$

More specifically, the information-theoretic analogs of the topological indices of Randić, I_χ, Wiener, I_W, and Hosoya, I_Z, are used. The distributions needed for the calculations are constructed by partitioning the χ, W, and Z into contributions of the different edges (bonds), distances, and $p(k)$ numbers, respectively (see eqs 3, 5, and 9, above).

The 3D molecular geometry is characterized by the distribution of atoms in stable conformations. The steric indices used are the Wiener number 3D-analog, WG, and its information-theoretic analogue, I_{WG}, the largest interatomic distance, L_{max}, and some characteristic intramolecular distances taken from the matrix of interatomic Euclidean distances.[11-13] The WG and I_{WG} indices are calculated by eqs 5 and 11, respectively, by making use of the metric interatomic distances instead of the integer graph distances. The minimal steric difference parameter MTD, which accounts for the similarity between the molecules in the series, is not included in the OASIS computational scheme, but for the sake of comparison we made use of the MTD values of compounds **1–19** as calculated in the recent publication of Niculescu-Duvaz et al.[24]

Physicochemical properties are also calculated by the OASIS method. Included here are the n-octanol/water partition coefficient P (or hydrophobicity factor $\log P$), as specified by the atomic increments approach of Ghose and Crippen,[31] as well as molecular volume V_M and molecular refraction MR, calculated as described by Hansch and Leo.[32]

The quantum mechanical SCF electronic structure is obtained for the ground state within the MO-LCAO (PM3, AM1, MNDO, INDO, or CNDO/2) approximation. Molecular geometry is optimized (usually by AM1) for better comparability of the compounds within the series under examination. Several quantum chemical quantities are calculated, such as atomic charges, donor/acceptor superdelocalizabilities (both global and frontier orbital ones), atomic and volume polarizabilities, heat of formation, the frontier molecular orbital (HOMO and LUMO) energies and their gap, molecular dipole moment, ionization potential, and bond orders.

The large number of parameters discussed in the foregoing can help provide a more adequate description of molecular structure, and hence, more reliable structure–activity models may result. However, the large number of screened variables increases the risk for arriving at chance correlations. Several means are utilized to reduce that risk. Indeed, in applying the OASIS method, the researcher may avoid the testing of some descriptors on the basis of previous experience (permanent insignificance in a wide range of test samples). Another mitigating factor is the high collinearity of some variables which produces a smaller effective number of independent variables to be screened. Finally, one should avoid the examination of all possible regression models by careful selection of the statistical treatment.

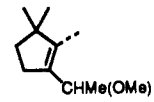
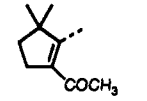
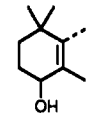
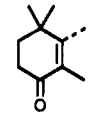
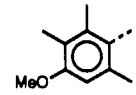
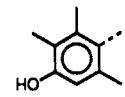
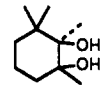
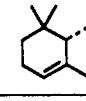
OASIS adopts the multiple regression analysis (MRA) procedure known as "forward/backward" or "addition/deletion" algorithm. The procedure starts with the highest one-variable correlation. Then two steps are performed consecutively in an iterative way. Each new variable is incorporated into the structure–activity model after showing the highest significance as evaluated by the partial F-test. The same test is performed once again in the second step of each iteration for the variables included into the current model during the previous stages of this multistep technique. The procedure terminates when the same model results after two consecutive iterations. A considerably smaller number of regressions is thus examined, which further reduces the change factor effect.

The last point of importance to mention concerning OASIS modeling is that the latter starts with one-variable models, all preliminary screened variables of which are tested by making use of four different functions: linear, quadratic, logarithmic, and exponential ones. Each variable is then used in the multistep regression procedure only with the function providing the best one-variable statistics. Previous experience has shown the logarithmic transform to be slightly better than or comparable to the linear one, the quadratic and exponential transforms ranging far behind. The advantage of the logarithmic function is evident when dealing with descriptors whose values differ by several order of magnitude (e.g. like the Hosoya index). This transform, however, is less universal than the linear one because it cannot treat variables whose values alternate in sign (in the event that all values are negative one could use absolute values).

Background

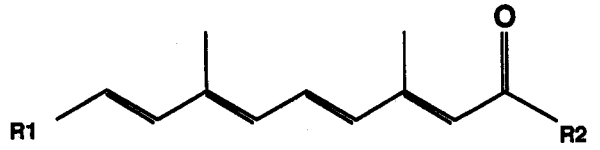
The very recent work of Balaban et al.[22] and the first part of the very extensive study of Niculescu-Duvaz et al.[24] deal with a series of 14 and 19 retinoids, respectively, for which reliable data for the anticancer activity were available.[23] This gives a good basis for comparison of the possible improvements provided by the OASIS method.

Table I: The Examined Series of Retinoids with Their Anticancer Activities,[23] Along with the Calculated Minimal Steric Differences, MTD, and Modified Balaban J Indices, MJ

no.	R1	R2	$-\log ED_{50}$	MTD	MJ
1		H	9.70	16	4.040
2		Me	9.52	16	4.199
3	same as for 2	Et	9.30	16	4.168
4		H	9.30	16	4.292
5		H	9.15	17	4.249
6		H	9.15	17	4.408
7		H	8.70	17	3.978
8	same as for 7	Me	8.52	18	3.995
9		H	8.30	18	4.805
10	same as for 9	Et	7.70	18	4.790
11		Et	7.00	18	4.945
12		Me	6.70	19	5.098
13		H	6.52	19	5.603
14		H	6.00	19	7.052
15		H	<9.00	17	3.710
16		H	<8.00	18	3.985
17		H	>8.00	19	6.301
18		H	6.00	19	5.162
19		H	10.52	16	4.183

In Table I we present the basis set of compounds. The numbering of compounds which will be used throughout this paper is taken from ref.[22] Table I also contains the measured activities $-\log ED_{50}$ of the compounds,[23] as well as the values of the best parameters from papers:[22,24] the modified Balaban J index (MJ) and the minimal steric difference MTD of Simon. The MJ values of compounds

15–19 have been calculated in this study. The substituents R1 and R2 are specified in Table I.



A careful examination of the data presented in the two reference papers has shown some inconsistency with the basic work of Newton et al.[23] Thus, compounds 11 and 12 from ref 22 were not found among the compounds reported by Newton et al. In addition, an erroneous activity was given for compound 13 in ref 24. Thus, the precise comparison with the results of these authors was impossible. Nevertheless, we repeated the calculations of these two groups of authors with their own data in order to check whether the QSAR models they reported would be confirmed by the OASIS procedure which (as a QSAR technique of a second generation) makes use of more rigorous statistical criteria.

The outcome of this re-examination of studies[22,24] was that no single model including more than one variable was found to be statistically significant; e.g. the three-variable models given in ref 22 were shown to be insignificant above a 64% level while those with four variables were above 51% level. (One should recall here that regressions with a significance lower than 90% level are usually regarded as not reliable and are discarded.) The modified Balaban J index, however, proved to be significant and was included to supplement the topological indices of use in the OASIS modeling. The three-variable model presented by eq 5 in ref 24 was also discarded by the OASIS modeling technique, which found two of these variables to be insignificant. The only significant parameter which was also used later in our models was the minimal steric difference parameter MTD.

Variable Screening and Chance Factor Evaluation

In what follows we report the results obtained for the test series of retinoid compounds given in Table I with the emphasis put on the series 1–14 compounds, due to the imprecise data for the biological activity of compounds 15–17. Extension of this basic series to 15 compounds by including retinoid acid was also performed. As said before, to ensure comparability between molecules in the set, all molecular descriptors used in the correlations were calculated by the OASIS method, including the physicochemical properties and quantum chemical electronic parameters.

The number of variables to be screened was reduced following the OASIS strategy for reducing the chance factor effect. Thus, molecular volume and volume polarizability are less sensitive to variations in molecular architecture at a constant number of atoms. M2 and I_Z topological indices, as well as the total electronic energy and ionization potential of the drug molecule, have shown so far little significance in drug–activity relationships. Last, but not least, all local electronic parameters (atomic charges, superdelocalizabilities, and polarizabilities, as well as bond orders) were found to be almost constant along the polyenic side chain and its terminal carbocyclic (or ester) functional group. They vary more significantly in the end cyclic fragment but no direct comparison was possible, due to the large variety of cycle sizes (six-, five-, and three-membered cycles) and substituents. Thus, 20 variables were selected to be screened: M1, χ, χ^v, J, MJ, W, WG, Z, D2, ϵ, L_{max}, I_χ, I_W, I_{WG}, log P, MR, μ, ΔH_f, E_{HOMO}, $\Delta E_{HOMO-LUMO}$. For most of them the logarithmic transforms were found to provide a better fit than the linear ones.

The 14 topological indices and the two physicochemical properties (log P and MR) have been found to intercorrelate highly; almost 50% of the 120 pairs of variables were classified as intercorrelating ones, according to the Topliss and Edwards criterion[33] $r \geq 0.80$. However, these 16 indices do not correlate with the four global molecular electronic descriptors listed above. Thus, before MRA was applied, the number of *independent* variables to be screened was reduced to $(0.5 \times 16) + 4 = 12$. This reduced number cannot eliminate completely the chance factor for the retinoid series to be studied. By making use of the Topliss and Edwards standard tables and figures[33] one can find that from 14 or 15 observations and 12 variables there is a 6–8% probability that some correlation with $r^2 \geq 0.9$ could emerge from chance factor alone. However, such numerical estimates of the risk for arriving at a chance correlation should be used with caution. The Topliss and Edwards estimates have been obtained for completely independent variables (random numbers) while a considerable residual bias remains after correction for collinearity of real data samples (e.g. one hardly should neglect the intercorrelations with $r = 0.50$–0.80). Klopman and Kalos have shown[34] a smaller chance effect when dealing with real experimental data (a direct comparison is, however, not possible, due to the more stringent statistical criteria used). Another study[35] has shown a reduced chance effect even for simulations with random numbers. By making use of the tables and figures of the latter work, the risk for obtaining by chance three-variable and four-variable models of the retinoid series under study was found to be less than the acceptable 1% limit. Another arguement in favor of the nonchance character of our models is the fact that the same geometric and hydrophobicity descriptors were found to be significant in the three-variable and four-variable models.

The Basic Series with 14 Retinoids

One- and Two-Variable Correlations. A surprisingly good linear correlation was obtained with the minimal steric difference parameter MTD, which accounts for the geometric similarity of the compounds:

$$-\log ED_{50} = -17.251(\pm 2.086)MTD + 57.524(\pm 5.959) \quad (12)$$

$$n = 14,\ r = 0.9223,\ s^2 = 0.251,\ F = 68.4,\ \alpha = 99\%$$

Another topological parameter, the modified J index, MJ, also manifests a satisfying correlation having $r = 0.88$, $s^2 = 0.38$, and $F = 41.3$ at a 99% level of significance. The next three correlating variables are the information-theoretic analog of the Randić molecular connectivity index I_χ with $r = 0.75$, the hydrophobicity factor log P with $r = 0.68$, and the energy gap between the highest occupied and the lowest unoccupied molecular orbitals (HOMO-LUMO Gap) with $r = 0.67$.

No significant two-variable correlations were obtained, due to the specific regression procedure used which

Table II. Comparison of the Experimental and Calculated (according to eq 13) Anticancer Activities of the Retinoids under Study, Their Refraction Indices MR, and the Hosoya Topological Indices Z

no.	exp.	calcd	diff	MR	Z
1	9.70	9.36	0.34	102.90	41 890
2	9.52	9.67	−0.15	100.97	28 842
3	9.30	9.49	−0.19	105.77	45 497
4	9.30	9.16	0.14	97.34	24 465
5	9.15	9.29	−0.14	97.95	25 167
6	9.15	8.87	0.28	97.34	25 167
7	8.70	8.87	−0.17	95.66	29 346
8	8.52	8.37	0.15	100.32	50 813
9	8.30	8.14	0.16	104.96	46 939
10	7.70	7.29	0.41	114.43	128 220
11	7.00	7.56	−0.56	109.77	76 283
12	6.70	6.92	−0.22	103.58	55 709
13	6.52	6.84	−0.32	80.31	5 818
14	6.00	5.72	0.28	77.85	3 572
15	<9.00	11.16	−2.16	96.56	16 657
16	<8.00	10.64	−2.64	96.70	16 657
17	>8.00	8.20	−0.20	83.86	3 572
18	6.00	7.91	−1.91	81.91	5 818
19	10.52	10.17	0.35	96.29	16 657

switches on three-variable or one-variable correlations if they were found to be more significant.

Three-Variable Correlations. This is the major class of correlations found for the series of compounds under study. Evidently, the drug–receptor interaction is more complex and at least three parameters are required to describe its different aspects. As shown below, besides the topological and steric parameters describing the drug–receptor geometric fitness, the hydrophobicity factor log P mirroring the through cell membrane drug transport proved to be of significance in most of the best fitted models. Below we discuss the three-variable correlations obtained with a correlation coefficient higher than 0.96.

The highest correlation includes two topological parameters: the modified J parameter and the Hosoya index Z along with molecular refraction MR, with the highest weight being found for MR and the lowest one for Z. The logarithmic transforms of the three variables were again found to provide a better fit than the linear ones:

$$-\log ED_{50} = 49.889(\pm 10.369) \log MR - 17.849(\pm 1.641) \log MJ - 6.283(\pm 1.168) \log Z - 51.173(\pm 15.842) \quad (13)$$

$$n = 14, r = 0.9728, s^2 = 0.108, F = 58.9, \alpha = 95\%$$

It should be recalled that the only regressions allowed are those in which the variables do not intercorrelate or intercorrelate weakly (the highest intercorrelation allowed varying between 0.80 and 0.85). Thus, the following intercorrelations were obtained for the above three variables: MJ/MR − 0.48, MJ/Z − 0.03, MR/Z − 0.81. The quality of the derived model of retinoid biological activity may also be demonstrated by the direct comparison between the experimental and calculated activities given in Table II.

As seen in Table II, the activity of the 14 compounds included in the basic series of retinoids (1–14) is reproduced fairly well. Thus, for eight compounds the difference is less than 0.20, while for the remaining six it is within the 0.20–0.56 range. The activity of retinoic acid (**19**) is also correctly assessed; such is also the case with the activity prediction of **17**. The potency of the three other compounds (**15**, **16**, and **18**) is, however, ill-reproduced. Similar difficulties are faced in the next regressions, as well. A possible explanation could be the lack of conjugation between the side chain and the terminal ring in compounds **15** and **16**. On the other hand, compound **18** is regarded as inactive and it is located at the low bound of the biological activity scale.

The other five correlations having r close to 0.96 include the hydrophobicity parameter log P with the highest weight. Most of them also incorporate, although with the lowest weight, the D2 centric topological parameter. The third parameter is also a geometric one (the global electronic indices when incorporated into the models produce poorer correlations). Included here are the Bonchev information–theoretic analog of the Randić molecular connectivity index I_χ, the Kier and Hall molecular connectivity χ^v, the modified Balaban J index MJ, the three-dimensional Wiener index WG, and its information–theoretic analog I_{WG}. Once again, most of these structural indices were found to fit the best when making use of logarithmic rather than linear transforms. The intercorrelation between any pair of these parameters is less than 0.85. The two next best regression equations follow:

$$-\log ED_{50} = 5.013(\pm 0.617) \log(\log P) + 7.314(\pm 1.102) \log KH - 3.513(\pm 0.436)D2 + 7.870(\pm 1.890) \quad (14)$$

$$n = 14, r = 0.9645, s^2 = 0.141, F = 44.4, \alpha = 99\%$$

$$-\log ED_{50} = 4.506(\pm 0.667) \log(\log P) + 4.017\ (\pm 0.644) \log I_\chi - 2.012(\pm 0.380)D2 + 10.481(\pm 1.931) \quad (15)$$

$$n = 14, r = 0.9606, s^2 = 0.156, F = 39.8, \alpha = 95\%$$

The use of regressors like log(log P) might look a bit awkward. However, the physical meaning of the variables used is not changed by this fit (e.g. log(log P) still measures hydrophobicity), similar to the usage of −log ED_{50} in medicinal chemistry, instead of the inhibiting concentration itself. Confining oneself to strictly linear regressors produces slightly worse statistics but qualitatively the same result.

Four-Variable Correlations. A number of statistically significant regressions are obtained with four variables. They differ qualitatively from those with three variables by the incorporation of the calculated (by the MNDO quantum chemical method) molecular dipole moment μ, regarded as a global electronic index. The dipole moment has as an additional advantage its very low intercorrelation with the other examined indices (less than 0.62 and usually less than 0.50). In this way, stable mathematical models of the retinoid biological action are built in which one parameter (log P) describes the drug molecule transport through cell membranes, another parameter (dipole moment) is related to the electronic drug–receptor interaction, and two parameters describe the optimal correspondence in the drug–receptor surface or spatial geometry. Two molecular connectivity indices proved again their applicability: the information–theoretic analog of Randić's index I_χ, and the Kier and Hall molecular connectivity χ^v. A third topological index, the centric index D2, is also included in two correlations. Two more geometrical, but not topological, indices contribute as well: the geometric (or 3-D) Wiener index WG, and its information–theoretic analog I_{WG}.

Table III. Comparison between Experimental and Calculated (according to eq 16) Anticancer Activities of the Retinoids under Study, Their Partition Coefficients log P, 3-D Wiener information indices I_{WG}, and dipole moments μ

no.	exp.	calcd	diff	log P	I_{WG}	μ
1	9.70	9.28	0.42	6.58	10.280	1.760
2	9.52	9.71	−0.19	7.63	10.188	1.829
3	9.30	9.38	−0.08	8.12	10.335	1.897
4	9.30	9.56	−0.26	6.63	9.984	2.377
5	9.15	9.14	0.01	6.10	10.093	1.923
6	9.15	8.95	0.20	6.63	9.978	1.475
7	8.70	8.84	−0.14	4.91	10.084	2.478
8	8.52	8.31	0.21	5.08	10.234	2.313
9	8.30	8.33	−0.03	6.57	10.029	1.853
10	7.70	7.61	0.09	7.22	10.334	1.975
11	7.00	7.09	−0.09	7.34	10.178	1.007
12	6.70	6.87	−0.17	4.20	10.398	1.268
13	6.52	6.50	0.02	4.69	9.077	1.816
14	6.00	5.98	0.02	3.69	8.898	2.230
15	<9.00	10.50	−1.50	7.29	10.136	2.121
16	<8.00	10.09	−2.09	7.09	10.034	2.051
17	>8.00	6.67	1.33	4.13	8.897	2.515
18	6.00	7.61	−1.61	5.37	9.144	2.286
19	10.52	10.49	0.03	7.47	10.035	2.303

As it will be demonstrated below with the best fitted regressions of this kind, the four-variable models are also considerably improved quantitatively as compared to the best three-variable ones, having correlation coefficients of almost 0.99 and standard deviations s^2 near 0.05:

$$-\log ED_{50} = 4.692(\pm0.368)\log(\log P) + 20.306(\pm1.892)\log I_{WG} + 1.296(\pm0.275)\log\mu - 2.853(\pm0.269)D2 - 28.428(\pm3.546) \quad (16)$$

$$n = 14,\ r = 0.9890,\ s^2 = 0.049,\ F = 100.7,\ \alpha = 90\%$$

$$-\log ED_{50} = 4.322(\pm0.433)\log(\log P) + 3.847(\pm0.418)\log I_\chi + 1.229(\pm0.318)\log\mu - 1.597(\pm0.268)D2 + 7.420(\pm1.477) \quad (17)$$

$$n = 14,\ r = 0.9854,\ s^2 = 0.065,\ F = 75.3,\ \alpha = 90\%$$

The inspection of the parameter weights in models 16 and 17 shows hydrophobicity log P to have the highest weight followed by the I_{WG} and I_χ geometric parameters in second place, the dipole moment μ being third, and the second geometric parameter D2 fourth. In the next best models (not shown here) hydrophobicity comes second after the I_χ topological parameter, the dipole moment is again third, and the second geometric parameter WG has again the least weight. One may conjecture on this basis that the hydrophobic and geometric factors are of prime and approximately equal importance for the biological activity of the retinoids under study, while the electrostatic interaction is also involved but is not dominant.

Some more evidence on the potencies of the derived four-variable correlations is presented in Table III. It demonstrates the high accuracy with which model 16 reproduces the anticancer activity of the 14 retinoids and predicts that of retinoic acid.

The Basic Series of Retinoids Expanded with Retinoic Acid

The expansion of the original series of 14 compounds with the most active compound of this type, retinoic acid (19), improved somewhat the statistics of all models obtained in the previous section, as will be briefly outlined below.

One- and Two-Varible Correlations:

$$-\log ED_{50} = -18.255(\pm2.073)MTD + 60.440 \quad (18)$$

$$n = 15,\ r = 0.9255,\ s^2 = 0.275,\ F = 77.6,\ \alpha = 99\%$$

The comparison with model 12 shows an increase in r from 0.9223 to 0.9255, as well as in F from 68.4 to 77.6 while the standard deviation s^2 is slightly higher (0.275 vs 0.250). No two-variable correlation was found at a level of significance of 80% or higher.

Three-Variable Correlations. Four regressions with a correlation coefficient equal to or larger than 0.97 were obtained vs only one in the case of $n = 14$ compounds. The Fisher criterion gains in average 15 points, and the standard deviation is slightly larger. The two models with the best statistics are shown below

$$-\log ED_{50} = 53.851(\pm9.158)\log MR - 18.175(\pm1.576)\log MJ - 6.784(\pm0.998)\log Z - 56.594(\pm14.335) \quad (19)$$

$$n = 15,\ r = 0.9765,\ s^2 = 0.105,\ F = 75.2,\ \alpha = 95\%$$

$$-\log ED_{50} = 5.120(\pm0.531)\log(\log P) + 7.355(\pm1.053)\log KH - 3.584(\pm0.381)D2 + 8.080(\pm1.741) \quad (20)$$

$$n = 15,\ r = 0.9709,\ s^2 = 0.130,\ F = 60.4,\ \alpha = 99\%$$

In the first of the above regressions, molecular refraction contributes with the highest weight while in the next four models the highest contribution comes invariably from the hydrophobicity factor. The lowest weight is that of the second topological parameter D2 or, in case of model 19, this is the Hosoya index Z. The agreement with the experimental activities of the compounds under examination is high. The discrepancy between the experimental data and the calculated ones is, with the exception of one or two compounds, less than 0.50 including here five to nine compounds for which this difference is less than 0.20.

Four-Variable Correlations. We present below the two best fits which contain the same variables as the respective models with 14 compounds. As seen, they show very high correlation and very low standard deviation at a high level of significance.

$$-\log ED_{50} = 4.709(\pm0.317)\log(\log P) + 20.326(\pm1.788)\log I_{WG} + 1.299(\pm0.260)\log\mu - 2.863(\pm0.239)D2 - 28.434(\pm3.366) \quad (21)$$

$$n = 15,\ r = 0.9911,\ s^2 = 0.044,\ F = 138.9,\ \alpha = 95\%$$

$$-\log ED_{50} = 4.506(\pm0.418)\log(\log P) + 3.848(\pm0.418)\log I_\chi + 1.261(\pm0.316)\log\mu - 1.693(\pm0.249)D2 + 7.729(\pm1.441) \quad (22)$$

$$n = 15,\ r = 0.9869,\ s^2 = 0.065,\ F = 93.8,\ \alpha = 90\%$$

The Basic Series of 14 Compounds Expanded with All Five Additional Compounds. For the same reason that the correct prediction of the activities of compounds 15, 16, and 18 failed for the series with 14 and 15 compounds, no two-, three-, or four-variable models of

the anticancer activity of the series of 19 retinoid compounds were obtained at a level of significance of 90% or higher. The only statistically significant regression obtained is that with a single variable, the minimal steric difference parameter MTD, for which $r = 0.90$, $s^2 = 0.34$, $F = 73.4$, $\alpha = 99\%$.

Discussion

Although incorporating parameters with insufficient statistical significance, the models reported in refs 22 and 24 may be compared quantitatively with our best fitted models. Thus, in ref 22 dealing with the series of 14 compounds, the highest correlations found for models with three and four variables are 0.934 and 0.938, respectively. Our best models of these two types are given by eqs 13 and 16, with correlation coefficients 0.973 and 0.990, respectively. This comparison provides evidence for the greater potential of the OASIS technique for modeling the anticancer activity of retinoid compounds.

The detailed OASIS modeling of retinoid activity presented in the previous section provides opportunities for some useful conclusions. Due to the complex nature of the retinoid molecule-receptor interaction, it is not likely that one- or two-variable models could be of importance. Most probably, models with more variables are needed to describe adequately this interaction. The three- and four-variable models we derived in best agreement with the measured activities outline some important tendencies. Thus, the three-parameter regressions seem to require one hydrophobic and two geometric parameters. The latter could be selected out of a set of topological indices with similar action, such as the modified Balaban J index, several information–theoretic indices, and the centric D2 index. Nontopological indices are also relevant, and first of all the three-dimensional analog of the Wiener topological index. The presence of two geometric parameters in the OASIS models perhaps indicates the importance of the geometric correspondence between the retinoid molecule and the receptor surface.

Our best four-parameter correlations include, although with a low weight, the molecular dipole moment, which may be regarded as an indication for some role of the electrostatic drug–receptor interaction. Thus, an optimal sample of four parameters seems to be the most appropriate for describing the anticancer action of the studied series of retinoids: a hydrophobic, an electronic, and two geometric parameters.

A special comment is needed for the failure of correct prediction of the activity of compounds **15**, **16**, and **18**. Besides the drastic difference in the electronic systems of compounds **15** and **16** with the rest of the series, we would like to point out here that these compounds were found to be outliers in the examined models dealing with the series with 19 compounds. This is perhaps an indication for the incompleteness of the set of variables used in this study. Another possible explanation is the specificity of the OASIS concept for quantitative structure–activity correlations. The potential of this approach could be maximally utilized when applied to a series of generic compounds differing in the nature, number, and location of the substituents to the basic molecular fragment which is also biologically active. This imposes some more rigorous requirements to the selection of compounds in order to deal with a homogeneous set of data thus providing a larger predictive area of compounds with unknown activity. These requirements were not completely met in the test series of 19 compounds, which differ considerably in their end cyclic fragments. Work on sets of retinoids which better satisfy these requirements is under way in our laboratory to explore compounds of recent experimental interest.

Acknowledgment. Danail Bonchev gratefully acknowledges the financial support from the Robert Welch Foundation and the Clifton F. Mountain Foundation. The authors are indebted to Dr. X. Liu, Texas A&M University, for his assistance in adapting some of the OASIS computer programs to the IBM RISC system. The reviewers' comments contributed to a significant improvement of the manuscript.

References

(1) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
(2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies Press: Chichester, England, 1986.
(4) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
(5) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
(6) Bonchev, D. *Information-Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press: Chichester, England, 1983.
(7) Mekenyan, O.; Bonchev, D. OASIS Method for Predicting Biological Activity of Chemical Compounds. *Acta Pharm. Yugosl.* **1986**, *36*, 225–237.
(8) Mekenyan, O.; Karabunarliev, S.; Bonchev, D. The OASIS Concept for Predicting Biological Activity of Chemical Compounds. *J. Math. Chem.* **1990**, *4*, 207–215.
(9) Mekenyan, O.; Bonchev, D.; Rouvray, D. H.; Peitchev, D.; Bangov, I. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. IV. The In Vivo Interaction of 2- and 6-Substituted Purines with Mirine Solid Tumor Adenocarcinoma CA 755. *Eur. J. Med. Chem.* **1991**, *26*, 305–312.
(10) Mekenyan, O.; Bonchev, D.; Enchev, V. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. V. Toxicity of Phenols to Algae "Lemna Minor". *Quant. Struct. Act. Relat.* **1988**, *7*, 240–244.
(11) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Bangov, I. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. I. Interaction of Substituted Pyridines with anti-3-Azopyridine Antibody. *Drug Design* **1986**, *36*, 176–183.
(12) Mekenyan, O.; Bonchev, D.; Trinajstić, N.; Peitchev, D. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. II. A Generalized Concept for Biological Interactions. *Drug Design* **1986**, *36*, 421–424.
(13) Mekenyan, O.; Peitchev, D.; Bonchev, D.; Trinajstić, N.; Dimitrova, J. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. III. Interaction of Benzoates with Anti-p-(p'-Azophenylazo) Benzoate Antibody. *Drug Design* **1986**, *36*, 629–634.
(14) Mekenyan, O.; Karabunarliev, S.; Bonchev, D. The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds. *Comput. Chem.* **1990**, *14*, 193–200.
(15) Mercier, C.; Sobel, Y.; Dubois, J.-E. In *Mathematical Chemistry*; Bonchev, D.; Rouvray, D. H., Eds.; Gordon and Breach: London, 1991; Vol. 1, pp 199–256.
(16) Mercier, C.; Mekenyan, O.; Dubois, J.-E.; Bonchev, D. DARC/PELCO and OASIS Methods. I. Methodological Comparison Modeling Purines pKa and Antitumor Activity. *Eur. J. Med. Chem.* **1991**, *26*, 575–592.
(17) Sporn, M. B.; Roberts, A. B.; Goodman, D. S., Eds. *The Retinoids*; Academic Press: New York, 1984; Vols. 1, 2.
(18) Marks, R., Ed.; *Retinoids in Cutaneous Malignancy*; Blackwell Scientific: Oxford, Boston, 1991.
(19) Dawson, M. I.; Okamura, W. H., Eds.; *Chemistry and Biology of Synthetic Retinoids*; CRC Press: Boca Raton, FL, 1990.
(20) Lotan, R. Vitamin A Analogs (Retinoids) as Biological Response Modifiers. In *Nutrition, Growth, and Cancer*; Tryliates, N., Ed.; Alan R. Liss Inc.: New York, 1988; pp 261–271.
(21) Lippman, S. M.; Meyskens, F. L. Retinoids as Anticancer Agents. In ref 20, pp 229–244.
(22) Balaban, A. T.; Catana, C.; Dawson, M.; Niculescu-Duvaz, I. Applications of Weighted Topological Index J for QSAR of Carcinogenesis Inhibitors (Retinoic Acid Derivatives). *Rev. Roum. Chim.* **1990**, *35*, 997–1003.

(23) Newton, D. L.; Henderson, W. R.; Sporn, M. B. Structure–Activity Relationships of Retinoids in Hamster Tracheal Organ Culture. *Cancer Research* **1980**, *40*, 3413–3425.
(24) Niculescu-Duvaz, I.; Simon, Z.; Voikuletz, N. In ref 19, pp 575–606.
(25) Voiculetz, N.; Balaban, A. T.; Niculescu-Duvaz, I.; Simon, Z. *Modeling of Cancer Genesis and Prevention*; CRC Press: Boca Raton, FL, 1991.
(26) Hosoya, H. Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
(27) Bonchev, D.; Balaban, A. T.; Mekenyan, O. Generalization of the Graph Center Concept, and Derived Topological Indices. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 106–113.
(28) Bonchev, D.; Trinajstić, N. Information Theory, Distance Matrix and Molecular Branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.
(29) Gutman, I.; Ruscić, B.; Trinajstić, N.; Wilcox, C. W. Graph Theory and Molecular Orbitals. XII. Acyclic Polyenes. *J. Chem. Phys.* **1975**, *62*, 3399–3405.
(30) Yee, W. T.; Sakamoto, K.; I'Haya, Y. J. Information Theory of Molecular Properties. I. A Theoretical Study of the Information Content of Organic Molecules. *Rep. Univ. Electro-Commun.* **1976**, *27*, 53–67.
(31) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure–Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35.
(32) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley Interscience: New York, 1979.
(33) Topliss, J. G.; Edwards, R. P. Chance Factors in Studies of Quantitative Structure-Activity Relationships. *J. Med. Chem.* **1979**, *22*, 1238–1244.
(34) Klopman, G.; Kalos, A. N. Causality in Structure–Activity Studies. *J. Comput. Chem.* **1985**, *6*, 492–506.
(35) Mekenyan, O. G.; Dimitrov, D. N.; Bonchev, D. G.; Karabunarliev, S. H. Some Quantitative Estimates of the Chance Correlations in Regression Models. *Quant. Struct.-Act. Relat.* Submitted.