dec; UV λ_{max} (MeOH) 233 nm (ϵ 4900); λ_{max} (pH 8 buffer) 232 nm ($\epsilon^{t=0}$ 5600). Anal. ($C_8H_{14}N_4O_5$, 246.2) C, H, N.

The free base can also be liberated by dissolving the hydrochloride **2** in ethanolic ammonia. Evaporation, after standing several hours at room temperature, and recrystallization of the residue from ethanol gave **3**, mp 179–180 °C dec.

4-Amino-1-(β-D-arabinofuranosyl)-1,3,5-triazin-2(1*H*)-one (5). A mixture of BSTFA (40 mL) and dry acetonitrile (40 mL) was gently refluxed with **2** (2.0 g, 7.1 mmol) under anhydrous conditions. As the reaction proceeded, the single peak in the gas chromatogram due to **3**·TMS_5 (IRI 2538) was accompanied by increasing concentrations of a second compound having a longer retention time. By mass spectral analysis it was shown to be **5**·TMS_4 (IRI 2650). After refluxing for 53 h, the latter compound produced the only peak in the chromatogram, indicating a complete thermal conversion: MS (**5**·TMS_4) *m/e* (rel intensity) 532 (1.7), 517 (6.0), 442 (1.7), 387 (2.4), 349 (0.7), 315 (19.8), 258 (44.8), 243 (24.3), 217 (61.6), 169 (17.4), 147 (21.4), 73 (100). The solvent was removed under vacuum and the residue taken up in 100 mL of methanol. Methanolysis of the trimethylsilyl groups was accomplished by slow distillation of the methanol solution during 1.5 h. The concentrate was diluted to 100 mL with a fresh portion of methanol, and the distillation was continued for another 1.5 h. After renewing the solution volume a second time with methanol, heat was removed and the solution was allowed to come to room temperature, causing 1.4 g (82%) of white crystals of **5** to separate: mp 223–225 °C dec; $[\alpha]^{24}_D$ 122° (*c* 1.0, H_2O, *t* = 0); $[\alpha]^{24}_D$ 63° (*c* 1.0, H_2O, *t* = 5 h); UV λ_{max} (H_2O, *t* = 0) 243 nm (ϵ 6800); λ_{max} (H_2O, *t* = 1 h) 242 nm (ϵ 7200); λ_{max} (H_2O, *t* = 3 h) 241 nm (ϵ 7800); λ_{max} (pH 2 buffer, *t* = 0) 252 nm (ϵ 3100); λ_{max} (pH 8 buffer, *t* = 0) 242 nm (ϵ 7400); IR (Nujol) 1702, 1665, 1616, 1165, 1082, 822, 799 cm^{-1}; NMR (Me_2SO-d_6) δ 8.22 (s, 1, C_6 H), 6.03 (d, *J* = 4 Hz, 1, $C_{1'}$ H). Anal. ($C_8H_{12}N_4O_5$, 244.2) C, H, N.

Acknowledgment. The authors thank Mr. L. Dudeck and Mr. G. Congleton, Hazleton Laboratories, Vienna, Va., for biological evaluations and Dr. J. A. Kelley of this laboratory for mass spectral measurements.

References and Notes

(1) A preliminary report of this work has been published: J. A. Beisler, M. M. Abbasi, and J. S. Driscoll, *Biochem. Pharmacol.*, **26**, 2469 (1977).
(2) NIH Visiting Postdoctoral Fellow from the National Research Center, Dokki, Cairo, Egypt.
(3) W. A. Creasey in "Antineoplastic and Immunosuppressive Agents", Part II, A. C. Sartorelli and D. G. Johns, Eds., Springer-Verlag, Berlin, 1975, p 232.
(4) J. Skoda, in ref 3, p 348.
(5) W. R. Vogler, D. S. Miller, and J. W. Keller, *Blood*, **48**, 331 (1976).
(6) D. D. Von Hoff, M. Slavik, and F. M. Muggia, *Ann. Intern. Med.*, **85**, 237 (1976).
(7) C. A. Presant, F. Valeriote, and T. J. Vietti, *Cancer Res.*, **37**, 376 (1977).
(8) J. A. Beisler, *J. Med. Chem.*, **21**, 204 (1978).
(9) (a) J. A. Beisler, M. M. Abbasi, and J. S. Driscoll, *Cancer Treat. Rep.*, **60**, 1671 (1976). (b) J. A. Beisler, M. M. Abbasi, J. A. Kelley, and J. S. Driscoll, *J. Med. Chem.*, **20**, 806 (1977).
(10) M. W. Winkley and R. K. Robins, *J. Org. Chem.*, **35**, 491 (1970).
(11) J. A. Beisler, M. M. Abbasi, J. A. Kelley, and J. S. Driscoll, *J. Carbohydr. Nucleosides, Nucleotides*, **4**, 281 (1977).
(12) R. I. Geran, N. H. Greenberg, M. M. Macdonald, A. M. Schumacher, and B. J. Abbott, *Cancer Chemother. Rep., Part 3*, **3**, 7 (1972).
(13) E. Kovats, *Helv. Chim. Acta*, **41**, 1915 (1958).
(14) R. B. Woodward, *Pure Appl. Chem.*, **9**, 49 (1964).

Potential Antitumor Agents. 31. Quantitative Structure–Activity Relationships for the Antileukemic Bis(guanylhydrazones)

William A. Denny and Bruce F. Cain*

Cancer Chemotherapy Research Laboratory, University of Auckland School of Medicine, Auckland, New Zealand.
Received April 4, 1979

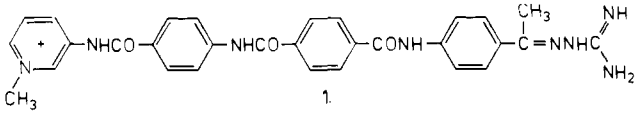
Certain L1210-active bis(guanylhydrazones) have structural and biological properties in common with the DNA minor groove binding, antileukemic, bisquaternary ammonium heterocycles. Monitoring of the DNA binding of the bis(guanylhydrazones), by fluorimetric quantitation of drug displacement of DNA-bound ethidium, shows that, like the bisquaternary salts, these agents bind more strongly to poly[d(A-T)] than poly[d(G-C)]. The drug concentrations necessary to inhibit L1210 DNA-dependent DNA polymerase in vitro by 50% (IC_{50}) are linearly related to measures of drug–DNA binding with no preference for a particular primary sequence of DNA being evident. Mammalian toxicity of the bis(guanylhydrazones) is effectively modeled by a regression equation containing binomial terms in R_m values, used as a measure of agent lipophilic–hydrophilic balance, and the logarithms of the IC_{50} values.

Development of high-activity clinical candidates, from the initial lead of a L1210 leukemia active bisquaternary ammonium heterocycle (BQAH),[1] first relied on the development of qualitative structure–activity relationships.[2] These qualitative relationships demonstrated a need for (1) two strongly basic centers separated by an aromatic framework providing a charge separation of greater than 18Å, (2) a capacity to fit to a slotlike annular site, and (3) agent lipophilic–hydrophilic balance within a critical range.[2] When it was later appreciated that these agents act by interference with DNA-template activity of tumor cells,[3,4] it proved possible to develop QSAR by utilizing measures of agent–DNA interactions as an indicator of site fit.[3] The measure of DNA interaction employed was derived from competition studies between agents and the fluorochrome ethidium for DNA sites.[3,5] A C_{50} value for the drug–DNA interaction was defined as the micromolar drug concentration necessary to displace 50% of DNA-bound ethidium, as monitored by fluorimetry. It can be demonstrated that such C_{50} values are inversely related to drug–DNA association constants.[5] This micromethod, which employs milligrams of drug and only micrograms of DNA, permitted an examination of the importance of drug binding to different DNA primary sequences for antitumor activity.[3] The BQAH agents bind more strongly to adenine–thymine (A-T) rich DNAs than to their guanine–cytosine (G-C) containing counterparts, and there was limited covariance between the C_{50} values observed for agent binding to poly[d(A-T)] and poly[d(G-C)]. Equations from multiple regression analysis demonstrated that the C_{50} values for drug interaction with poly[d(A-T)] could incorporate significantly more of the variance in the biologic data than those obtained from poly(dA)·poly(dT) or poly[d(G-C)]. Interaction of these agents with DNA can

0022-2623/79/1822-1234$01.00/0

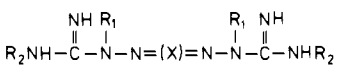
then be specified in two ways: (1) by their level of interaction with a particular DNA and (2) by their discriminatory ability for different DNAs. The latter may be conveniently quantified as the ratio of the C_{50} values measured for two different sequenced DNAs, e.g., C_{50}[poly[d(A-T)]]/C_{50}[poly[d(G-C)]]. There is minimal covariance between the latter discriminatory quotient and the individual C_{50} values for a particular DNA. Development of acceptable regression equations incorporating this discriminant, for the BQAH series, suggests that the antitumor selectivity shown by these agents is a function of their ability to distinguish alternating A-T rich areas, in the DNA of a target cell, in relation to G-C sequences.[3]

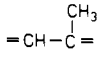
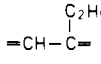
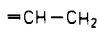
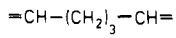
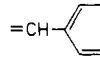
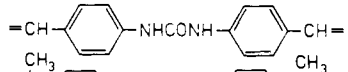
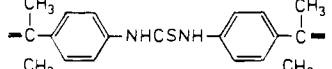
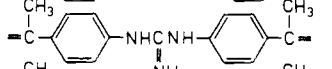
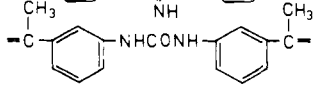
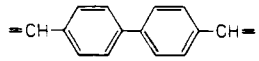
In development of the initial qualitative SAR it was determined that at least one quaternary ammonium function could be acceptably replaced by an alternative strongly basic function, if this was selected from among the resonant amidinium-type bases (amidine, guanidine, imidazoline, 1,4,5,6-tetrahydropyrimidine, guanylhydrazone).[2] As an example from the data base developed,

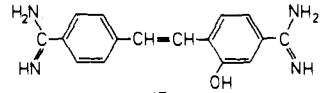


1

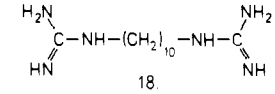
AGENT STRUCTURES



No.	X	R_1	R_2
2		H	H
3		CH_3	H
4		H	CH_3
5		H	H
6		H	H
7		H	H
8		H	H
9		H	H
10		H	H
11		H	H
12		H	H
13		H	H
14		H	H
15		H	H
16		H	H



17



18

1 serves as a conceptual bridge between the BQAH and certain L1210-active aromatic bis(guanylhydrazones) (BGH), such as **12**. It can be noted that such agents, extensively investigated by Mihich et al.,[6-11] meet the criteria required by our qualitative SAR and, additionally, considerable research has shown that DNA is probably the site of action for these drugs.[8,9] The correspondence between the BQAH and the aromatic BGH agents, both possessing L1210 activity, charge separation of greater than 18Å, structural aromatic components permitting lodgement in a slotlike annular site, and DNA as a putative site of action, suggested that the two groups of agents might be considered congeneric and that earlier developed QSAR[3] might then be applicable to the aromatic BGH. Also, if these agent groups were congeneric, then the BGH should demonstrate sequence selectivity of binding to the synthetic, homopolymeric DNAs. This paper details a study of these various possibilities.

Additionally, Mihich and co-workers[9] have examined a series of aliphatic BGH and two further dibasic agents, certain of which have moderate antitumor activity but clearly do not meet the structural requirements of our qualitative SAR.[2] These agents, grouped separately in Table I (**2–9** and **18**), have clearly different biologic properties from their aromatic counterparts (**10–17**), as evidenced by investigations with the two examples afforded clinical trial [DDUG (**12**) and methylglyoxal bis(guanylhydrazone), MGGH (**2**)]. The aliphatic example requires carrier-mediated transport into target cells,[12] whereas DDUG penetrates cells by passive diffusion.[11,13] The latter binds strongly to double-stranded DNAs with much of the binding being attributable to nonionic forces,[9,11] as is the case with the BQAH.[3] In contrast, MGGH binds weakly to DNA at low ionic strengths, indicating mainly weak ionic interactions.[9,14] The aromatic example (**12**) strongly inhibits DNA-dependent DNA polymerase (DDP) in vitro,[9] while MGGH is a poor inhibitor of this enzyme.[9,10] The mode of antitumor action of **2** is unclear but is thought to involve interference with the functioning of cellular polyamines.[8] It was clearly of interest to also examine the DNA binding of the subgroup of aliphatic BGH and the possible applicability of earlier developed QSAR to these.

Agents Examined. These are those tabulated in the paper of Mihich et al.[9] One L1210 inactive guanylhydrazone (compound XVIII of Mihich et al.) was unavailable but, since it clearly should not be considered along with bis(guanylhydrazones), it was not resynthesized. Compounds **6**, **10**, and **16** were prepared in this laboratory as bis(hydrochloride salts) by published procedures,[15,16] while **17** was contributed by May and Baker Ltd., Dagenham, U.K. All remaining compounds were generously provided by Dr. E. Mihich of Roswell Park Memorial Institute.

Data Base. Table I lists the values of all biologic and physicochemical parameters employed. Biologic data are those of Mihich et al.[9] The animal strains employed for the L1210 screening tests (DBA_2), the screening protocols employed (10^6 L1210 cells implanted ip, dosing qd 1–4 ip), and the data interpretation methods employed[9] differ from those employed in our earlier QSAR study ($DBA_2 \times C_3H$ F_1 hybrid mice, 10^5 L1210 cells implanted ip, dosing qd 1–5 ip).[3] The maximum percentage increases in life span (ILS) observed in the former studies might then differ appreciably from those that would have been obtained if

Table I. Biologic[a] and Physicochemical Data for the Bis(guanylhydrazones)

		C_{50} data[b]		log (IC_{50})[c]			log (OD)[d]			log (ILS)[e]		
no.	R_m[f]	d(A-T)[g]	d(G-C)[h]	obsd	calcd[i]	diff	obsd	calcd[j]	diff	obsd	calcd[k]	diff
					aliphatic bis(guanylhydrazones)							
2	-0.27	34	128	2.85	2.75	0.10	1.70	2.09	-0.39	2.00	1.98	0.02
3	0.04	103	379	>3.00	2.99	>0.01	2.18	2.03	0.15	<1.40[l]	1.74	>-0.34
4	-0.08	27	111	2.81	2.70	0.11	2.30	1.94	0.36	<1.40	1.87	>-0.47
5	-0.20	14	51	2.70	2.56	0.14	2.18	1.90	0.27	<1.40	1.89	>-0.49
6	-0.08	112	248	2.87	3.00	-0.13	2.00	2.07	-0.07	<1.40	1.70	>-0.30
7	-0.52	*m*	*m*	>3.00			2.30			<1.40		
8	-0.04	70	81	>3.00	2.91	>0.09	1.87	1.85	0.02	<1.40	1.65	>-0.25
9	-0.25	22	200	2.85	2.66	0.19	1.88	2.16	-0.28	<1.40	1.84	>-0.44
18[n]	0.89	127	191	>3.00	3.03	>-0.03	0.60	0.67	-0.07	2.11	<1.40	<0.71
					aromatic bis(guanylhydrazones)							
10	-0.43	2.9	7.6	1.90	2.23	-0.33	1.60	1.72	-0.12	1.84	1.97	-0.13
11	-1.11	0.30	0.69	1.95	1.74	0.21	1.48	1.35	0.13	2.08	1.97	0.11
12	-0.75	0.43	1.3	1.95	1.82	0.13	1.40	1.51	-0.11	2.47	2.05	0.42
13	-0.10	1.2	3.4	2.11	2.04	0.07	1.40	1.39	0.01	1.97	2.01	-0.04
14	-0.08	0.08	0.20	1.60	1.46	0.14	0.70	0.91	-0.21	1.82	2.16	-0.34
15	-0.06	0.42	1.6	1.60	1.81	0.19	1.48	1.23	0.25	2.10	2.11	-0.01
16	-0.43	0.14	1.0	1.60	1.58	0.02	1.30	1.39	-0.09	1.63	2.28	-0.65
17	-0.46	1.2	6.7	1.60	2.04	-0.44	1.87	1.71	0.16	1.85	2.13	-0.28

[a] All biologic data are taken from ref 9. [b] The C_{50} value is the micromolar drug concentration necessary to displace 50% of DNA-bound ethidium; see ref 3 and 5. [c] IC_{50} is the micromolar concentration needed to effect a 50% decrease in the velocity of the DNA-dependent DNA-polymerase reaction; see ref 9. [d] OD is the optimum dose in (mg/kg)/day, qd 1-4, for anti-L1210 leukemic activity; see ref 9. [e] ILS is the maximum percentage increase in lifespan observed in L1210 leukemia assays. [f] R_m is a relative measure of lipophilic–hydrophilic balance for partition chromatography; see ref 3. [g] d(A-T) = poly[d(A-T)] - poly[d(A-T)]. [h] d(G-C) = poly[d(G-C)] - poly[d(G-C)]. [i] Calculated using eq 1. [j] Calculated using eq 4. [k] Calculated using eq 12 in ref 5. [l] Inactive compound. [m] Compound too insoluble for C_{50} values to be obtained. [n] Synthalin.

the compounds were screened under the conditions employed in our earlier QSAR study.[3] If possible, the biologic data of Table I should be handled in an internally self-consistent manner.

As a relative measure of agent lipophilic–hydrophilic balance, R_m values from partition chromatography were employed. The chromatographic system employed was that used earlier with the BQAH.[3]

Relative binding levels of the compounds to the synthetic polymers poly[d(A-T)] and poly[d(G-C)] were determined by the ethidium-displacement assay.[3,5] A further advantage of the latter method is that it is also applicable to UV-transparent compounds; thus, this paper provides the first data on the DNA-binding characteristics of the hypoglycemic agent synthalin (18), as well as those aliphatic BGH which failed to provide UV-spectral changes in the presence of DNA.[9]

Results and Discussion

DNA Binding. Both the qualitative SAR[2] and summarized subsidiary evidence[3] are compatible with lodgement of the BQAH agents in the minor groove of twin helical DNAs. The same binding site has been later proposed for the antibiotics distamycin A and netropsin.[17] From the correspondence seen between the BQAH and the aromatic BGH, it appears reasonable to postulate the same type of binding for the latter. This view is further supported by the finding that DDUG (12) shows no unwinding of closed circular, superhelical PM2 bacteriophage DNA and, to this criteria, does not bind to DNA by intercalation.[19] Nevertheless, this compound binds very strongly to double-stranded DNAs by a combination of ionic and nonionic forces,[9,11] as do the BQAH.[3] The data of Table I show further that all the BGH studied bind more strongly (lower C_{50} value) to poly[d(A-T)] than to poly[d(G-C)]. This phenomenon has been consistently observed with other minor-groove binders, and it is suggested that this results from steric inhibition of binding by the guanine 2-NH_2 of the G-C base pairs.[3] While there is considerable covariance between the measured C_{50} values for binding of the BGH to both the A-T and G-C polymers ($r^2 = 0.96$), it is noteworthy that the ability of these compounds to discriminate between different DNA polymers, as provided by the ratio of these values, is independent of the absolute magnitude of these values ($r^2 = 0.07$ and 0.003; see Table II). Although binding levels of the BGH to poly[d(A-T)] and poly[d(G-C)] are more closely related than were those of the BQAH (see Table II), the two major types of DNA-association properties, their binding and discriminatory properties, are again quite independent.

Table II. Correlation Matrices (r^2 Matrices) for DNA-Binding Parameters

	log $[1/C_{50}$(G-C)]	log $[C_{50}$(G-C)/C_{50}(A-T)]
(i) for the bisquaternary ammonium heterocycles,[a] 174 compounds		
log $[1/C_{50}$(A-T)]	0.40	0.24
log $[1/C_{50}$(G-C)]		0.14
(ii) for the bis(guanylhydrazones) 2-6 and 8-18, 16 compounds		
log $[1/C_{50}$(A-T)]	0.96	0.07
log $[1/C_{50}$(G-C)]		0.003

[a] See ref 3. [b] In Table I.

DNA-Dependent DNA Polymerase (DDP) Inhibition. As a measure of the interaction of the BGH with calf thymus DNA, the changes in the UV spectrum of the compound, when DNA was added, were formerly employed.[9] It was then observed that there was a qualitative relationship between the magnitude of this interaction and the drug concentration providing 50% inhibition of DDP activity (IC_{50} values). Extending this study, the quantitative relationship between IC_{50} values and the C_{50} values for drug binding to poly[d(A-T)] and poly[d(G-C)] have been examined. Four compounds were not included in the analysis because the IC_{50} values for compounds **3**, **8**, and **18** were not quantitated but recorded as greater than 1000 μM, and, for unclear reasons, because **7** was too insoluble

Table III. Stepwise Development of the Equations Discussed[a]

eq no.	intercept	G-C	R_m	R_m^2	r	s	$F_{1,X}$
4	1.33	−0.24			0.53	0.42	5.7 (14)[b]
	0.96	−0.41	−0.82		0.81	0.31	12.2 (13)
	1.10	−0.37	−0.92	−0.58	0.90	0.23	10.2 (12)
		A-T	R_m	R_m^2	r	s	$F_{1,X}$
5	1.48	−0.20			0.46	0.45	3.7 (14)
	1.18	−0.38	−0.85		0.76	0.34	11.1 (13)
	1.30	−0.37	−1.00	−0.68	0.89	0.25	13.0 (12)
		log (IC_{50})	R_m	R_m^2	r	s	$F_{1,X}$
6	0.81	0.35			0.42	0.46	2.9 (14)
	0.09	0.59	−0.68		0.66	0.39	6.0 (13)
	0.21	0.59	−0.87	−0.74	0.83	0.30	10.4 (12)

[a] Terms employing C_{50} values for a particular DNA species have been abbreviated; e.g., log [$1/C_{50}$(A-T)] is provided as A-T. [b] Degrees of freedom X for the F test.

to allow C_{50} values to be obtained. For the remaining 13 compounds there was a clear relationship between DDP inhibition and DNA binding (eq 1). As expected, because

$$\log (IC_{50}) = -0.49(\pm 0.12) \log [1/C_{50}(\text{A-T})] + 2.00 \quad (1)$$

$$n = 13,\ r = 0.92,\ s = 0.22,\ F_{1,11} = 64.6$$

of covariance between poly[d(A-T)] and poly[d(G-C)] binding for these compounds (Table II), a similar result is obtained using the values for binding to poly[d(G-C)] (eq 2). In contradistinction, no significant relationship

$$\log (IC_{50}) = -0.47(\pm 0.11) \log [1/C_{50}(\text{G-C})] + 1.74 \quad (2)$$

$$n = 13,\ r = 0.91,\ s = 0.24,\ F_{1,11} = 51.5$$

was observed between IC_{50} values and DNA discriminatory properties, defined as the ratio of the two C_{50} values, of the BGH. Therefore, the inhibitory capabilities of these compounds toward DDP appears almost entirely due to their ability to bind to DNA in a general way and not to their discriminatory ability. This is in agreement with earlier observations[11,20] that, in the case of DDUG (12), the inhibition reaction is dependent only on DDUG/DNA template concentrations and is not altered by change of the DNA template. Note that these equations do predict the poor inhibitory activities of the three compounds not included in this analysis. Predicted log (IC_{50}) values for these compounds from eq 1 are: 3, 2.99; 8, 2.91; 18, 3.03; quoted, >3.0.[9]

Although there is considerable variation in lipophilic–hydrophilic balance through the series, no term in R_m could be usefully included in eq 1 and 2. The earlier studies with the BQAH also showed that DNA binding appeared insensitive to changing agent lipophilic character.

Animal Toxicity. From previous experience, it was expected that agent lipophilic–hydrophilic balance would play a role in determining mammalian toxicity. In agreement, eq 3 (where OD = optimal dose) could be

$$\log (\text{OD}) = -0.73(\pm 0.66)R_m^2 - 0.45(\pm 0.51)R_m + 1.68 \quad (3)$$

$$n = 16,\ r = 0.56,\ s = 0.43,\ F_{2,13} = 3.0,\ p < 0.1,$$
$$R_m(\text{optimum}) -0.31\ (-1.13 \text{ to } 0.51)$$

derived for the compounds of Table I, omitting 7 because DNA binding data for this was unavailable. Further addition of a term in binding of the compounds to poly[d(G-C)] provided a considerable improvement (eq 4).

$$\log (\text{OD}) = -0.58(\pm 0.36)R_m^2 - 0.92(\pm 0.34)R_m - 0.37(\pm 0.13) \log [1/C_{50}(\text{G-C})] + 1.10 \quad (4)$$

$$n = 16,\ r = 0.90,\ s = 0.23,\ F_{3,12} = 17.5,$$
$$R_m(\text{optimum}) -0.79\ (-1.29 \text{ to } -0.29)$$

In the latter, the C_{50} binding term is the most important (see Table III for details of stepwise development), but inclusion of all terms is significant at below the 0.01 level.

As expected, from the covariance seen in DNA binding terms a similar equation (eq 5) employing C_{50} values for

$$\log (\text{OD}) = -0.68(\pm 0.39)R_m^2 - 1.00(\pm 0.39)R_m - 0.37(\pm 0.13) \log [1/C_{50}(\text{A-T})] + 1.30 \quad (5)$$

$$n = 16,\ r = 0.89,\ s = 0.25,\ F_{3,12} = 15.5,$$
$$R_m(\text{optimum}) -0.74\ (-1.28 \text{ to } -0.20)$$

binding to poly[d(A-T)] was equally useful.

Once again, attempted use of the ratio of the two C_{50} values, as a DNA binding discriminant, was not successful.

In view of the clear relationships demonstrated earlier between DNA binding and IC_{50} values for inhibition of DDP (eq 1 and 2), it is not surprising the in vivo toxicity can also be modeled in terms of this parameter (eq 6). To

$$\log (\text{OD}) = -0.74(\pm 0.45)R_m^2 - 0.87(\pm 0.44)R_m + 0.59(\pm 0.30) \log (IC_{50}) + 0.21 \quad (6)$$

$$n = 16,\ r = 0.83,\ s = 0.30,\ F_{3,12} = 9.2,$$
$$R_m(\text{optimum}) -0.59\ (-1.21 \text{ to } 0.03)$$

allow comparisons between eq 6 and eq 4 and 5, log (IC_{50}) values of 3.00 have been used for compounds **3**, **8**, and **18** in order to retain the same data base.

The term in log IC_{50} for inhibition of DDP, in eq 6, accepts a greater proportion of the variance in the biologic data than any other term. For this series of basic compounds, in which OD (optimal dose) varies by 50-fold, mammalian toxicity appears highly responsive to the DDP-inhibitory action of the agents.

Tumor Selectivity. While the screening protocols employed for assaying antitumor effectiveness of the BGH were different from those employed in the BQAH study, it was clearly of interest to find if the QSAR derived for the latter could be used in a predictive fashion to at least correctly order the activities of the BGH. Using measured R_m and DNA-binding parameters and the best equation derived for the BQAH (eq 12 in ref 5), the maximum ILS values predicted for the BGH are those provided in Table I. Predictions for the L1210 activity of the aliphatic bis(guanylhydrazones) were, with the exception of MGGH (**2**), clearly seriously in error. Many inactive examples would have been predictd to be active (**3–9**), while the L1210-active diguanide synthalin (**18**) would, because of both its excessive lipophilic nature and its poor DNA discriminatory ability, have been expected to be inactive. As earlier discussed, the aliphatic BGH and particularly synthalin may have quite different modes of action, and such poor ranking of activities would not then be unex-

pected. However, it was correctly predicted that all of the aromatic BGH's examined would be L1210 active, although the rank ordering of these was clearly different from that observed. With the view that such ranking might reflect the different screening protocols employed, it was then desirable to apply regression analysis directly to the biologic data. However, with only eight tumor-active aromatic BGH's available, there are barely enough data points to adequately examine the importance of one variable, and it has been clearly demonstrated that in vivo antitumor activity is modeled successfully only by multivariable analysis. Although no single term could be found significant at the 5% level, in modeling ILS values for the aromatic BGH, the parameter which enters first in a forward, stepwise, multiple linear regression is log $[C_{50}(G\text{-}C)]/[C_{50}(A\text{-}T)]$, the measure of DNA discriminating ability being employed. This latter parameter is also the single most important variable in modeling of the antitumor activities of the BQAH.[3]

While it has been suggested that the antitumor activity of the aromatic BGH's is related to their ability to inhibit DDP in vivo,[8,9] the present study shows that this property is related to the in vivo toxicity of the general class of BGH and may indeed be a useful predictor of this toxicity. However, if the aromatic BGH's are congeneric with the BQAH, then it would be expected that the antitumor selectivity is dependent on in vivo binding to an as yet undefined, alternating A-T rich site(s) in the tumor cell DNA.

Experimental Section

R_m values were determined by the chromatographic method detailed earlier for the BQAH,[3] employing Merck cellulose F_{254} DC sheets as support. UV-absorbing compounds were detected by their fluorescence quenching of the cellulose support. UV-transparent compounds, for example, synthalin (18), were located by spraying with pentacyanoaquoferriate reagent, prepared as follows: Equal volumes of cold 10% aqueous solutions of sodium nitroprusside, potassium ferricyanide, and sodium hydroxide were mixed together, in that order, and the red-orange solution stood at room temperature until the color faded to a clear yellow-green (20–25 min). H_2O (4 volumes) and Me_2CO (3 volumes) were then added, and the solution was used immediately.

Where adequate quantities of the BQH salts were available, weighed amounts were employed to prepare standard aqueous solutions of 2 mM strength. For those agents in short supply, aqueous solutions were prepared and the concentration of these was determined from the UV data quoted by Mihich et al.,[9] final dilutions being made to provide 2 mM solutions.

Acknowledgment. We are grateful to Mrs. E.-M. Falkenhaug for performance of the many C_{50} determinations. We are indebted to Dr. E. Mihich for samples of many BGH, without which this study could not have been undertaken. This work was supported by the Auckland Division of the Cancer Society of New Zealand (Inc.) and, in part, by the Medical Research Council of New Zealand.

References and Notes

(1) G. J. Atwell and B. F. Cain, *J. Med. Chem.*, **10**, 706 (1967).
(2) B. F. Cain, G. J. Atwell, and R. N. Seelye, *J. Med. Chem.*, **12**, 199 (1969).
(3) W. A. Denny, G. J. Atwell, B. C. Baguley, and B. F. Cain, *J. Med. Chem.*, **22**, 134 (1979).
(4) A. C. Counsilman, B. F. Cain, and B. C. Baguley, *Eur. J. Cancer*, **10**, 75 (1974).
(5) B. C. Baguley and E. M. Falkenhaug, *Nucleic Acids Res.*, **5**, 161 (1978).
(6) E. Mihich, *Cancer Res.*, **23**, 1375 (1963).
(7) E. Mihich, *Cancer*, **20**, 880 (1967).
(8) E. Mihich, "Antineoplastic and Immunosuppressive Agents", A. C. Sartorelli and D. G. Johns, Ed., Springer-Verlag, New York, 1975, p 766.
(9) C. Dave, M. J. Ehrke, and E. Mihich, *Chem.-Biol. Interact.*, **16**, 57 (1977).
(10) C. Dave, M. J. Ehrke, and E. Mihich, *Proc. Am. Assoc. Cancer Res.*, **12**, 40 (1971).
(11) C. Dave, M. J. Ehrke, and E. Mihich, *Chem.-Biol. Interact.*, **12**, 183 (1975).
(12) C. Dave and L. Caballes, *Fed. Proc., Fed. Am. Soc. Exp. Biol.*, **32**, 736 (1973).
(13) M. T. Hakala, *Biochem. Pharmacol.*, **20**, 81 (1971).
(14) A. C. Sartorelli, A. T. Iannotti, B. A. Booth, F. H. Schneider, J. R. Bertino, and D. G. Johns, *Biochim. Biophys. Acta*, **103**, 174 (1965).
(15) E. G. Podrebarac, W. H. Nyberg, F. A. French, and C. C. Cheng, *J. Med. Chem.*, **6**, 283 (1963).
(16) G. Cavallini, E. Massarani, D. Nardi, L. Mauri, and D. Mantegazza, *J. Med. Pharm. Chem.*, **4**, 177 (1961).
(17) C. Zimmer, *Prog. Nucleic Acid Res. Mol. Biol.*, **15**, 285 (1975).
(18) B. C. Baguley, unpublished data, this laboratory.
(19) M. J. Waring, *J. Mol. Biol.*, **54**, 247 (1970).
(20) J. Soucek, M. J. Ehrke, and E. Mihich, *Cancer Res.*, **30**, 2187 (1970).

Chance Factors in Studies of Quantitative Structure–Activity Relationships

John G. Topliss* and Robert P. Edwards

Schering-Plough Research Division, Bloomfield, New Jersey 07003. *Received April 2, 1979*

Multiple regression analysis is a basic statistical tool used for QSAR studies in drug design. However, there is a risk of arriving at fortuitous correlations when too many variables are screened relative to the number of available observations. In this regard, a critical distinction must be made between the number of variables screened for possible correlation and the number which actually appear in the regression equation. Using a modified Fortran stepwise multiple-regression analysis program, simulated QSAR studies employing random numbers were run for many different combinations of screened variables and observations. Under certain conditions, a substantial incidence of correlations with high r^2 values were found, although the overall degree of chance correlation noted was less than that reported in a previous study. Analysis of the results has provided a basis for making judgements concerning the level of risk of encountering chance correlations for a wide range of combinations of observations and screened variables in QSAR studies using multiple-regression analysis. For illustrative purposes, some examples involving published QSAR studies have been considered and the reported correlations shown to be less significant than originally presented through the influence of unrecognized chance factors.

During the past decade, quantitative structure–activity relationships (QSAR) have been increasingly used in drug-design studies.[1] Typically, a number of possible independent variables, usually physicochemical parameters relating to a series of compounds, are evaluated for correlation with activity values using multiple-regression analysis.[2] The correlation equation which emerges from this analysis may contain only a small number of inde-

0022-2623/79/1822-1238$01.00/0