try and Biology of Peptides," J. Meienhofer, Ed., Ann Arbor Science Publishers, Ann Arbor, Mich., 1972, pp 509-512.
(12) D. Regoli, W. K. Park, F. Rioux, and C. S. Chan, *Rev. Can. Biol.*, **30**, 319 (1971).
(13) D. Regoli and W. K. Park, *Can. J. Physiol. Pharmacol.*, **50**, 99 (1972).
(14) C. Pena, J. M. Stewart, and T. C. Goodfriend, *Life Sci.*, **14**, 1331 (1974).
(15) M. C. Khosla, R. R. Smeby, and F. M. Bumpus, *J. Amer. Chem. Soc.*, **94**, 4721 (1972).
(16) R. B. Merrifield, *J. Amer. Chem. Soc.*, **85**, 2149 (1963).
(17) R. F. Furchgott and S. Bhadrakom, *J. Pharmacol. Exp. Ther.*, **103**, 129 (1953).
(18) O. Arunlakshna and H. O. Schild, *Brit. J. Pharmacol.*, **14**, 48 (1959).
(19) G. C. Windridge and E. C. Jorgensen, *J. Amer. Chem. Soc.*, **93**, 6318 (1971).
(20) W. Koenig, R. Geiger, and H. Wissman, German Offen. 2,202,613 (1972); *Chem. Abstr.*, **79**, 137,510 (1973).
(21) J. M. Stewart and J. D. Young, "Solid-Phase Peptide Synthesis," W. H. Freeman, San Francisco, Calif., 1969, pp 48-49.
(22) G. R. Marshall, *Advan. Exp. Med. Biol.*, **2**, 48 (1968).
(23) R. B. Woodward, R. A. Olofson, and H. Mayer, *J. Amer. Chem. Soc.*, **83**, 1010 (1961).
(24) M. Bodanszky and K. W. Funk, *J. Org. Chem.*, **38**, 1296 (1973).
(25) R. K. Olsen, *J. Org. Chem.*, **35**, 1912 (1970).
(26) P. T. Pickens, F. M. Bumpus, A. M. Lloyd, R. R. Smeby, and I. H. Page, *Circ. Res.*, **17**, 438 (1965).

Structure–Activity Relationships in Immunochemistry. 2. Inhibition of Complement by Benzamidines†

Corwin Hansch* and Masafumi Yoshimoto‡

Department of Chemistry, Pomona College, Claremont, California 91711. *Received April 29, 1974*

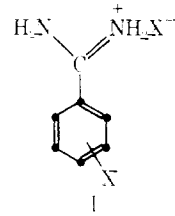
A quantitative structure-activity relationship has been formulated for 108 benzamidine derivatives inhibiting complement. The relationship is expressed as: $\log 1/C = 0.15(\text{MR-1,2}) + 1.07(\text{D-1}) + 0.52(\text{D-2}) + 0.43(\text{D-3}) + 2.43$ where C is the molar concentration causing 50% inhibition of complement, MR is the molar refractivity of substituents, D-1 is an indicator variable which accounts for activity of the moiety PhX where X represents a variety of bridge units, D-2 is an indicator variable for the presence or absence of a pyridine moiety, and D-3 is an indicator variable for the presence of the structural unit NHCOY attached to the second benzene ring where Y may be a variety of different units. The above equation correlates activity of 108 derivatives with a correlation coefficient of 0.935 and a standard deviation of 0.258. This relationship suggests new approaches to the synthesis of complement inhibitors.

As our understanding of the immune response grows, it becomes more feasible to systematically attempt to influence it by chemical means. It was found in our first study in this area[1] that the inhibition of antibody-antigen interactions by haptens could be formulated in a mathematical relationship. It was shown that the inhibition of the antibody-antigen interaction by haptens of the type X-$C_6H_4COO^-$ and X-$C_6H_4AsO_3H^-$ from the studies of Pauling and Pressman[2] could be described quantitatively by equations such as eq 1. In eq 1, E_s is the Taft steric

$$\log K_{rel} = 0.86E_s^{o} + 0.08E_s^{m} - 0.45E_s^{p} - 0.69 \quad (1)$$
$$n \quad r \quad s$$
$$22 \quad 0.974 \quad 0.177$$

parameter and the superscripts o, m, and p refer to the substituents in the ortho, meta, and para positions of the substituted benzoate haptens. Little is lost in dropping the E_s^m since its coefficient is quite small. In eq 1, n represents the number of data points, r is the correlation coefficient, and s is the standard deviation from the regression line. The positive coefficient with E_s^o in this quantitative structure-activity relationship (QSAR) indicates that large substituents in the ortho position make poor haptens and the negative coefficient with E_s^p indicates that the opposite is true for para substituents. Our success in correlating the structurally demanding hapten-antibody interactions encouraged us to attempt other immunochemical QSAR.

The late B. R. Baker and his students carried out an extensive study of the inhibitory action of derivatives of benzamidine (I) on guinea pig complement. This was part of Baker's generalized search for drugs to inhibit mamma-



I

lian proteolytic enzymes. As he pointed out,[3] at least 15 such distinct proteolytic enzymes have been characterized. Since all of these proteolytic enzymes hydrolyze peptide bonds, they must be closely related and their specificity appears to reside mainly in the type of acylated amino acid amide preferred for complexing.[4] Hence, designing an inhibitor for one such enzyme is not an easy problem and requires the study of inhibitors on a variety of proteolytic enzymes. Baker's group was particularly concerned with trypsin, chymotrypsin, and complement. Complement consists of 11 distinct proteins[5,6] which are required for cell lysis brought about *via* antibodies and complement. The function of the antibodies is to identify the invading cell as a foreign organism and activate complement attack which results in cell lysis by means of the proteolytic enzymes. When complement is activated by an antibody, it of course could attack the host's own cells. The antibody circumvents this problem by fixing complement on the surface of the foreign cell.

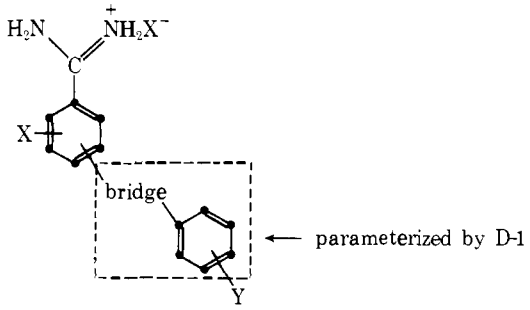
Thus it is apparent that there are several routes open for the inhibition of the rejection of tissue or organ transplants. One might inhibit antibody formation, control formed antibodies with haptens, or inhibit the functioning of the complement system. Baker chose to study complement inhibitors.

Coats[7] has studied a subset of 25 of Baker's benzamidine complement inhibitors. Although he obtained a good

†This research was supported by Grant CA 11110 from the National Institutes of Health and by the Sankyo Co. of Tokyo, Japan.
‡Visiting Scientist from the Sankyo Co.

correlation, it was apparent from the complex nature of the derivatives of I that dummy variables would have to be employed if one wished to embrace all of the reversible inhibitors studied by Baker.[8-12] Indicator variables (dummy variables) provide a means for handling structural features in numerical terms which do not lend themselves to a description by a continuous function.[13]

Method. Table I contains activities and constants used in the formulation of eq 2-6. C in log $1/C$ represents the molar concentration of inhibitor required for 50% inhibition of lyophilized guinea pig complement when assayed in buffer.[8] C for 50% inhibition has been estimated graphically from Baker's data by making a linear plot of activity *vs.* concentration. The benzamidine inhibitors can be characterized by the following general formula.



II

To parameterize the structural features of II, π-1, σ-1, and MR-1 were used to characterize X attached to the parent benzene ring and π-2, σ-2, and MR-2 were used for substituents Y on the second benzene ring. An indicator variable, D-1, was assigned a value of 1.00 for bridge units attached to a second ring [$-O(CH_2)_2O-$, $-(CH_2)_4-$, $-O(CH_2)_3O-$, $-O(CH_2)_3-$, $-O(CH_2)_4O-$, $-O(CH_2)_4$]. D-1 also included the second ring as shown above. For all other cases D-1 was given the value 0.00. In preliminary work with the above parameters it was observed that four instances where a pyridine moiety was attached at the end of side chain were generally underpredicted. Indicator variable D-2 was employed to account for this feature. D-2 pertains only to the pyridine moiety; the connecting $(CH_2)_x$ was characterized by π or MR as needed. Since only four points are involved, D-2 can be omitted without significantly affecting the correlation. D-3 was used to account for the special activating effect of the function 3-$NHC(=O)XC_6H_5$ when attached to the second ring. Five variations of this function were dealt with by assigning D-3 a value of 1.00 for the examples where X = zero, NH, CH_2, $NHCH_2CH_2$, and CH_2O. It is most interesting that when such amide functions are in the 2 or 4 position they do not produce any special effect; their contribution to activity is accounted for in terms of π and MR alone. It was expected that SO_2F might require special parameterization because of its ability to form covalent bonds with nucleophilic functions; this turned out to be unnecessary. Electronic effects were studied by factoring σ into σ_1, σ_2, and σ_3 for each of the three benzene rings (see Table I); however, no electronic role for substituents could be discerned.

Table II lists log P and π values not previously reported. These were measured by the usual procedure.[14] The other values in Table I were taken from the literature or estimated from additivity principles.[15,16] The values for $\pi[NHCO(CH_2)_2C_6H_5] = 0.91$,[17] $\pi(OCH_2C_6H_5) = 1.66$,[18] $\pi(NHCONHC_6H_5) = 0.83$,§ and $\pi(NHCOC_6H_4\text{-}4\text{-}NO_2) = 0.82$,& are known. Two values were used for the SO_2F function. The value of 0.05 has been determined from $CH_3C_6H_4\text{-}4\text{-}SO_2F$** and was used in cases where the SO_2F was on rings not directly attached to lone pair electron-bearing atoms (O, N). It is known[14] that when strong electron-withdrawing groups are, for example, placed on a phenol, log P is much higher than expected from additivity of π constants from the benzene system. When SO_2F is on a benzene ring attached to O or N, a second $\pi(SO_2F)$ is calculated as

$$\log P(CH_3CONHC_6H_4\text{-}4\text{-}SO_2F) - \log P(CH_3CONHC_6H_5) = 2.17 - 1.16 = 1.01$$

The calculation of log P and π is also illustrated as follows.

$$\pi[O(CH_2)_3OC_6H_4\text{-}4\text{-}NO_2] = \pi(CH_3OC_6H_4\text{-}4\text{-}NO_2) + \pi(OCH_3) + \pi(CH_2) = 2.03 - 0.02 + 0.50 = 2.51$$

The value of 2.51 is close to the experimental value of 2.41 (Table II) which again illustrates the additive character of log P.

$$\pi[O(CH_2)_3OC_6H_5] = \pi[O(CH_2)_3OC_6H_4\text{-}4\text{-}NO_2] - \pi(NO_2) = 2.41 - (-0.08) = 2.49$$

where

$$\pi(NO_2) = \log P(CH_3OC_6H_4\text{-}4\text{-}NO_2) - \log P(CH_3OC_6H_5) = 2.03 - 2.11 = -0.08$$

In this case it is necessary to use $\pi(NO_2)$ from the anisole system because the strong electron withdrawal by NO_2 on oxygen increases hydrophobicity.

$$\pi[O(CH_2)_2OC_6H_5] = \pi[O(CH_2)_3OC_6H_5] - \pi(CH_2) = 2.49 - 0.50 = 1.99$$

$$\pi[O(CH_2)_4OC_6H_5] = \pi[O(CH_2)_4OC_6H_4\text{-}4\text{-}NO_2] - \pi(NO_2) = 2.84 - (-0.08) = 2.92$$

$$\pi[O(CH_2)_3C_6H_5] = \pi(OCH_3) + 2\pi(CH_2) + \log P(C_6H_6) = -0.02 + 1.00 + 2.13 = 3.11$$

$$\pi[O(CH_2)_4C_6H_5] = 3.11 + 0.50 = 3.61$$

$$\pi[(CH_2)_4C_6H_5] = \pi[(CH_2)_2C_6H_5] + 2\pi(CH_2) = 2.66 + 1.00 = 3.66$$

$$\pi(CH_2CH_2\text{-}Py) = \pi(C_3H_7\text{-}4\text{-}Py)^{15} - \pi(CH_3) = 2.10 - 0.50 = 1.60$$

$$\pi[(CH_2)_4\text{-}Py] = \pi[(CH_2)_2\text{-}Py] + 2\pi(CH_2) = 1.60 + 1.00 = 2.60$$

$$\pi(NHCOCH_2Br) = \pi(NHCOCH_3) + \pi(Br) = -0.97 + 0.60 = -0.37$$

$$\pi(C_6H_5CH_2CONH-) = \pi(C_6H_5CONH-) = 0.49^{17}$$

$$\pi(C_6H_5OCH_2CONH-) = 0.66^{18}$$

It has been assumed that the same value of π can be employed for ortho, meta, and para substituents and no correction was made for groups adjacent to each other.

§C. Hansch and C. Church, unpublished results.

&C. Hansch and D. Nikaitani, unpublished results.

** C. Hansch and D. Soderberg, unpublished results.

Table I. Constants Used for Deriving Eq 2-6

No.	X	Log 1/C		Δ log 1/C	MR-1,2[c]	π-1,2[c]	D-1[c]	D-2[c]	D-3[c]
		Obsd[a]	Calcd[b]						
1	3,5-$(OCH_3)_2$	2.21[d]	2.521	−0.31	1.47	−0.04	0.0	0.0	0.0
2	2-CH_3	2.25[e]	2.503	−0.25	0.57	0.56	0.0	0.0	0.0
3	3,4-$(CH_3)_2$	2.35	2.609	−0.26	1.04	1.12	0.0	0.0	0.0
4	H	2.39	2.397	−0.01	0.10	0.0	0.0	0.0	0.0
5	3-OH	2.41[e]	2.341	0.07	0.29	−0.67	0.0	0.0	0.0
6	3-$NHCO(CH_2)_2C_6H_5$	2.43[e]	2.901	−0.47	4.40	0.91	0.0	0.0	0.0
7	3-CF_3	2.44	2.532	−0.09	0.50	0.88	0.0	0.0	0.0
8	3-NO_2	2.44[e]	2.426	0.01	0.74	−0.28	0.0	0.0	0.0
9	3-Br	2.47[e]	2.566	−0.10	0.89	0.86	0.0	0.0	0.0
10	3-CH_3	2.51[e]	2.503	0.01	0.57	0.56	0.0	0.0	0.0
11	3-OCH_3	2.52	2.459	0.06	0.79	−0.02	0.0	0.0	0.0
12	3-$CH_2C_6H_5$	2.68	2.891	−0.21	3.00	2.01	0.0	0.0	0.0
13	3,5-$(CH_3)_2$	2.77	2.609	0.16	1.04	1.12	0.0	0.0	0.0
14	3-OC_3H_7	2.82	2.656	0.16	1.71	0.98	0.0	0.0	0.0
15	3-*i*-C_5H_{11}	2.82	2.869	−0.05	2.43	2.30	0.0	0.0	0.0
16	3-OC_4H_9	2.85	2.755	0.10	2.17	1.48	0.0	0.0	0.0
17	3-C_4H_9	2.96	2.792	0.17	1.96	2.00	0.0	0.0	0.0
18	3-$CH{=}CHC_6H_5$	2.95	3.034	−0.08	3.42	2.95	0.0	0.0	0.0
19	3-$OCH_2C_6H_5$	3.02	2.873	0.15	3.22	1.66	0.0	0.0	0.0
20	3-$(CH_2)_2C_6H_5$	3.03	3.007	0.02	3.47	2.66	0.0	0.0	0.0
21	3-OC_6H_{13}	3.04[e]	2.951	0.09	3.09	2.48	0.0	0.0	0.0
22	3-$O(CH_2)_4OC_6H_5$	3.08	3.583	−0.50	0.10	0.0	1.00	0.0	0.0
23	3-$O(CH_2)_2OC_6H_5$	3.10	3.583	−0.48	0.10	0.0	1.00	0.0	0.0
24	3-C_6H_5	3.10	2.842	0.26	2.54	1.96	0.0	0.0	0.0
25	3-$O(CH_2)_3OC_6H_4$-4-COOH	3.11[e]	3.158	−0.05	0.69	−4.36	1.00	0.0	0.0
26	3-OC_5H_{11}	3.12	2.853	0.27	2.63	1.98	0.0	0.0	0.0
27	3-O-*i*-C_5H_{11}	3.14	2.831	0.31	2.63	1.78	0.0	0.0	0.0
28	3-$O(CH_2)_2OC_{10}H_7$-α	3.20[d]	3.875	−0.68	1.64	1.33	1.00	0.0	0.0
29	3-$O(CH_2)_4OC_6H_4$-4-NH_2	3.27	3.489	−0.22	0.54	−1.23	1.00	0.0	0.0
30	3-$(CH_2)_4C_6H_5$	3.31	3.583	−0.27	0.10	0.0	1.00	0.0	0.0
31	3-$O(CH_2)_3OC_6H_4$-4-NO_2	3.32	3.613	−0.29	0.74	−0.28	1.00	0.0	0.0
32	3-$O(CH_2)_3OC_6H_4$-4-NH_2	3.33	3.489	−0.16	0.54	−1.23	1.00	0.0	0.0
33	3-$(CH_2)_2$-4-C_5H_4N	3.34	3.401	−0.06	3.23	1.60	0.0	1.00	0.0
34	3-$O(CH_2)_3OC_6H_5$	3.34	3.583	−0.24	0.10	0.0	1.00	0.0	0.0
35	3-$O(CH_2)_3C_6H_5$	3.38[e]	3.583	−0.20	0.10	0.0	1.00	0.0	0.0

36	3-$(CH_2)_2$-3-C_5H_4N	3.40	3.401	0.00	3.23	1.60	0.0	1.00	0.0
37	3-$(CH_2)_4C_6H_4$-4-NHAc	3.40	3.607	−0.21	1.49	−0.97	1.00	0.0	0.0
38	3-$(CH_2)_2$-2-C_5H_4N	3.44	3.401	0.04	3.23	1.60	0.0	1.00	0.0
39	3-$O(CH_2)_3OC_6H_4$-2-NH_2	3.54	3.489	0.05	0.54	−1.23	1.00	0.0	0.0
40	3-$O(CH_2)_3OC_6H_4$-4-NHAc	3.60	3.607	−0.01	1.49	−0.97	1.00	0.0	0.0
41	3-$(CH_2)_4$-3-C_5H_4N	3.62	3.598	0.02	4.16	2.60	0.0	1.00	0.0
42	3-$O(CH_2)_4C_6H_5$	3.62	3.583	0.04	0.10	0.0	1.00	0.0	0.0
43	3-$O(CH_2)_3OC_6H_4$-3-NHAc	3.70	3.607	0.09	1.49	−0.97	1.00	0.0	0.0
44	3-$O(CH_2)_3OC_6H_3$-3,4-Cl_2	3.77[e]	3.834	−0.06	1.10	1.42	1.00	0.0	0.0
45	3-$O(CH_2)_3OC_6H_4$-3-NH_2	3.77	3.489	0.28	0.54	−1.23	1.00	0.0	0.0
46	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-4-SO_2F	3.77	4.031	−0.26	4.23	0.54	1.00	0.0	0.0
47	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_5$	3.77	3.953	−0.18	3.46	0.49	1.00	0.0	0.0
48	3-$O(CH_2)_3OC_6H_4$-4-OCH_3	3.82	3.646	0.17	0.79	−0.02	1.00	0.0	0.0
49	3-$O(CH_2)_4OC_6H_4$-4-$NHCONHC_6H_4$-4-SO_2F	3.85	4.207	−0.36	4.58	1.84	1.00	0.0	0.0
50	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-OCH_3-5-SO_2F	3.85[e]	4.092	−0.24	4.91	0.52	1.00	0.0	0.0
51	3-$O(CH_2)_3OC_6H_4$-4-Cl	3.89	3.709	0.18	0.60	0.71	1.00	0.0	0.0
52	3-$O(CH_2)_3OC_6H_4$-2-NO_2	3.89	3.613	0.28	0.74	−0.28	1.00	0.0	0.0
53	3-$O(CH_2)_3OC_6H_4$-3-NO_2	3.89	3.613	0.28	0.74	−0.28	1.00	0.0	0.0
54	3-$O(CH_2)_3OC_6H_4$-3-OCH_3	3.90	3.646	0.25	0.79	−0.02	1.00	0.0	0.0
55	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-Cl-6-SO_2F	3.92	4.155	−0.24	4.72	1.25	1.00	0.0	0.0
56	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_5$	3.92	4.023	−0.10	3.81	0.83	1.00	0.0	0.0
57	3-$O(CH_2)_2OC_6H_4$-2-$NHCONHC_6H_3$-2-Cl-5-SO_2F	4.00	4.331	−0.33	5.07	2.55	1.00	0.0	0.0
58	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHCH_2C_6H_4$-4-SO_2F	4.00	4.144	−0.14	5.04	0.88	1.00	0.0	0.0
59	3-$O(CH_2)_3OC_6H_4$-2-NHCONH-C_6H_2-2,4-$(CH_3)_2$-5-SO_2F	4.03	4.416	−0.39	5.49	2.96	1.00	0.0	0.0
60	3-$O(CH_2)_3OC_6H_4$-4-$COOCH_3$	4.05	3.694	0.36	1.29	−0.01	1.00	0.0	0.0
61	3-$O(CH_2)_3OC_6H_3$-3-NO_2-4-CH_3	4.08	3.718	0.36	1.21	0.28	1.00	0.0	0.0
62	3-$O(CH_2)_3OC_6H_4$-3-CF_3	4.09	3.718	0.37	0.50	0.88	1.00	0.0	0.0
63	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-4-CH_3-3-SO_2F	4.09	4.312	−0.22	5.04	2.40	1.00	0.0	0.0
64	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_5$	4.10	3.953	0.15	3.46	0.49	1.00	0.0	0.0
65	3-$O(CH_2)_3OC_6H_4$-2-$NHCOCH_2OC_6H_4$-4-SO_2F	4.10	4.200	−0.10	4.91	1.50	1.00	0.0	0.0
66	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-OCH_3	4.11	4.015	0.10	4.15	0.47	1.00	0.0	0.0
67	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-3-SO_2F	4.12[e]	4.031	0.09	4.23	0.54	1.00	0.0	0.0
68	3-$O(CH_2)_3OC_6H_4$-2-$NHCOCH_2C_6H_4$-4-SO_2F	4.12	4.074	0.05	4.69	0.54	1.00	0.0	0.0
69	3-$O(CH_2)_3OC_6H_4$-3-$COOCH_3$	4.14	3.694	0.45	1.29	−0.01	1.00	0.0	0.0
70	3-$O(CH_2)_3OC_6H_4$-2-$NHCO(CH_2)_2C_6H_4$-4-SO_2F	4.14	4.172	−0.03	5.15	1.04	1.00	0.0	0.0
71	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-NO_2	4.19	4.049	0.14	4.10	0.82	1.00	0.0	0.0
72	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_4$-4-NO_2	4.19	4.049	0.14	4.10	0.82	1.00	0.0	0.0
73	3-$O(CH_2)_3OC_6H_4$-4-$NHCONHC_6H_5$	4.21	4.023	0.19	3.81	0.83	1.00	0.0	0.0
74	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-3-NO_2	4.21	4.049	0.16	4.10	0.82	1.00	0.0	0.0
75	3-$O(CH_2)_3OC_6H_4$-2-$NHCO(CH_2)_4C_6H_4$-4-SO_2F	4.22	4.370	−0.15	6.08	2.04	1.00	0.0	0.0

Table I *(Continued)*

No.	X	Log 1/C: Obsd[a]	Log 1/C: Calcd[b]	Δ log 1/C	MR-1,2[c]	π-1,2[c]	D-1[c]	D-2[c]	D-3[c]
76	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-4-SO_2F	4.22	4.207	0.01	4.58	1.84	1.00	0.0	0.0
77	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-3-SO_2F	4.23	4.668	−0.44	4.58	1.84	1.00	0.0	1.00
78	3-$O(CH_2)_3OC_6H_4$-2-$NHCONH(CH_2)_2C_6H_4$-4-SO_2F	4.23	4.244	−0.01	5.51	1.38	1.00	0.0	0.0
79	3-$O(CH_2)_4OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	4.23	4.492	−0.26	4.23	0.54	1.00	0.0	1.00
80	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_3$-4-Cl-3-SO_2F	4.25	4.226	0.02	5.08	1.59	1.00	0.0	0.0
81	3-$O(CH_2)_4OC_6H_4$-2-$NHCOC_6H_3$-4-CH_3-3-SO_2F	4.25	4.136	0.11	4.69	1.10	1.00	0.0	0.0
82	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_2$-2,4-$(CH_3)_2$-5-SO_2F	4.27	4.241	0.03	5.15	1.66	1.00	0.0	0.0
83	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_2$-2,4-Cl_2-5-SO_2F	4.28	4.281	0.00	5.23	1.96	1.00	0.0	0.0
84	3-$(CH_2)_4C_6H_4$-2-$NHCONHC_6H_4$-3-SO_2F	4.28	4.207	0.07	4.58	1.84	1.00	0.0	0.0
85	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-OCH_3	4.29	4.476	−0.19	4.15	0.47	1.00	0.0	1.00
86	3-$(CH_2)_4C_6H_4$-2-$NHCONHC_6H_4$-4-SO_2F	4.30	4.207	0.09	4.58	1.84	1.00	0.0	0.0
87	3-$O(CH_2)_3OC_6H_4$-4-$NHCOC_6H_4$-4-Cl	4.31[e]	4.078	0.23	3.96	1.20	1.00	0.0	0.0
88	3-$O(CH_2)_3OC_6H_4$-2-$NHCOC_6H_3$-2-CH_3-5-SO_2F	4.31	4.136	0.17	4.69	1.10	1.00	0.0	0.0
89	3-$O(CH_2)_4OC_6H_5$-4-$NHCONHC_6H_3$-2-OCH_3-5-SO_2F	4.31[e]	4.729	−0.42	5.25	1.82	1.00	0.0	1.00
90	3-$O(CH_2)_3OC_6H_4$-4-C_6H_5	4.35[e]	4.029	0.32	2.54	1.96	1.00	0.0	0.0
91	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_4$-3-SO_2F	4.35	4.207	0.14	4.58	1.84	1.00	0.0	0.0
92	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-3-SO_2F	4.35	4.492	−0.14	4.23	0.54	1.00	0.0	1.00
93	3-$O(CH_2)_2OC_6H_4$-3-$NHCOC_6H_4$-3-SO_2F	4.37	4.492	−0.12	4.23	0.54	1.00	0.0	1.00
94	3-$O(CH_2)_3OC_6H_4$-4-CH_3-3-$NHCOC_6H_4$-4-SO_2F	4.37	4.597	−0.23	4.69	1.10	1.00	0.0	1.00
95	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-3-SO_2F	4.51	4.668	−0.16	4.58	1.84	1.00	0.0	1.00
96	3-$O(CH_2)_3OC_6H_4$-3-$NHCOCH_2C_6H_4$-4-SO_2F	4.54	4.535	0.01	4.69	0.54	1.00	0.0	1.00
97	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	4.57	4.492	0.08	4.23	0.54	1.00	0.0	1.00
98	3-$O(CH_2)_3OC_6H_4$-2-$NHCONHC_6H_3$-2-Cl-5-SO_2F	4.60	4.332	0.27	5.08	2.55	1.00	0.0	0.0
99	3-$O(CH_2)_3OC_6H_4$-3-$NHCOCH_2OC_6H_4$-4-SO_2F	4.60	4.680	−0.08	4.91	1.67	1.00	0.0	1.00
100	3-$O(CH_2)_2OC_6H_4$-3-$NHCONHC_6H_4$-4-SO_2F	4.62	4.668	−0.05	4.58	1.84	1.00	0.0	1.00
101	3-$O(CH_2)_4OC_6H_4$-3-$NHCONHC_6H_4$-4-SO_2F	4.64	4.668	−0.03	4.58	1.84	1.00	0.0	1.00
102	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-NO_2	4.68	4.510	0.17	4.10	0.82	1.00	0.0	1.00
103	3-$O(CH_2)_2OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	4.68	4.492	0.19	4.23	0.54	1.00	0.0	1.00
104	3-$O(CH_2)_4OC_6H_4$-2-$NHCONHC_6H_4$-2-Cl-5-SO_2F	4.82	4.332	0.49	5.08	2.55	1.00	0.0	0.0
105	3-$O(CH_2)_3OC_6H_4$-3-$NHCONHC_6H_4$-4-NO_2	4.89	4.522	0.37	4.54	0.55	1.00	0.0	1.00
106	3-$O(CH_2)_3OC_6H_3$-4-CH_3-3-$NHCONHC_6H_4$-4-SO_2F	4.89	4.774	0.12	5.05	2.40	1.00	0.0	1.00
107	3-$O(CH_2)_3OC_6H_4$-3-$NHCONH(CH_2)_2C_6H_4$-4-SO_2F	5.17	4.705	0.47	5.51	1.38	1.00	0.0	1.00
108	3-$O(CH_2)_3OC_6H_4$-3-$NHCOC_6H_4$-4-SO_2F	5.21	4.492	0.72	4.23	0.54	1.00	0.0	1.00

[a] Calculated from results of Baker, *et al.*[8,12] [b] Calculated using eq 2. [c] See section on Method for sources of these constants. [d] In these two compounds only one concentration was tested and inhibition was less than 10% in each case. [e] In the case of these log 1/C values it was necessary to use only one datum in making the extrapolation to the I_{50} concentration since only one concentration had been tested by Baker. In some examples, very low activity or saturation of activity makes linear extrapolation unreliable and, hence, only one data point was employed.

Table II. Partition Coefficients and New π Constants

	Compound	Log P^a	Mp, °C	Substituent	π
I	Benzamidine hydrochloride	−1.59 ± 0.05		$C(NH_2)$=NH·HCl	−3.72[b]
II	4-(4-Nitrophenoxypropoxy)benzamidine hydrochloride	0.82 ± 0.01	186–188	$O(CH_2)_3OC_6H_4$-4-NO_2	2.41[c]
III	4-(4-Nitrophenoxypropoxy)benzamidine hydrobenzenesulfonate	0.83 ± 0.01	196–198		
IV	4-(3-Nitrophenoxypropoxy)benzamidine hydrochloride	0.99 ± 0.09	195–198	$O(CH_2)_3OC_6H_4$-3-NO_2	2.58[c]
V	4-(4-Nitrophenoxybutoxy)benzamidine hydrochloride	1.25 ± 0.07	187–190	$O(CH_2)_4OC_6H_4$-4-NO_2	2.84[c]

[a] In the determination, 0.1 *N* sodium chloride solution was used as the aqueous phase. [b] π = log *P* − log *P*(benzene) = log *P* − 2.13. [c] π = log *P* − log *P*(benzamidine hydrochloride) = log *P* − (−1.59) = log *P* + 1.59.

Since folding effects[14] appeared to be absent in the model compounds of Table II, no correction was made for this factor.

We have combined π and log *P* in certain instances to obtain new log *P* values. This assumes that π_H is in fact ~0.00. Of course π_H is defined in a relative sense as 0.00. However, experience indicates[15] that, in an absolute sense, π_H must be near zero. The MR values in Table I are taken from our recent compilation or were calculated in the same manner.[19] MR has been scaled by 0.1 for convenience in calculation. This also makes MR more equiscalar with respect to π. When one function is attached to a ring, MR for the substituent is simply taken from our tabulation[19] [*e.g.*, $MR(CH_3)$ = 0.1 × 5.65 = 0.57]. MR for the parent ring system = 0.1 × 1.03 = 0.1[MR(H)]. Hence, when two or more substituents are present, 0.1 must be subtracted for the H they replace on the ring [*e.g.*, $MR(CH_3)_3$ = 0.1[3 × 5.65 − 2(1.03)] = 1.50].

σ constants are not available for some of the functions. Values were estimated from similar functions.[19] E_s values are not available for most of the complex substituents and, in retrospect, would seem to be inappropriate in any case.

Compounds II–V in Table II were prepared by the usual methods. Compound I was commercially available. The purity of all compounds was checked by thin-layer chromatography. All compounds gave satisfactory carbon-hydrogen analyses.

Results

The formulation of a QSAR for a very complex set of congeners such as those in Table I is a formidable problem and, just as in solving any complex puzzle, there are a variety of approaches one can take. To begin with, one must make the tentative assumption that all of the molecules are acting in the same way qualitatively. During the course of the study congeners whose activity is grossly different will stand out from the others and they can be deleted from the study or parameters can be formulated to account for their deviant character. Experience in working with substituent constants such as π, σ, etc., is of great help in spotting trends in the data which can then be checked by the formulation of the appropriate linear combination of substituent constants. The solution of problems such as those posed by the structures and activities of Table I is by no means a "plug-in" type problem. Many hypotheses must be tested and rejected or included in more encompassing hypotheses before the results begin to become internally self-consistent. In the present instance, the approach we employed was to attempt to correlate the data in each of the publications by Baker and his students which appeared over a period of several years. From these studies there appeared to be no reason not to merge all of the data. Since there was some uncertainty about the π value and σ constants for the bridge units, D-1 was formulated early in the game. It soon became apparent that D-1 would account for several types of bridges (see Method). D-2 was employed for the pyridine moiety and D-3 for 3-NHCO units in the second ring (see Method). The two "best" mathematical models for the QSAR of the benzamidines are shown in eq 2 and 3. The quality of the correlation in the two equations is so close that it is not possible to say with certainty that the substituent effect is hydrophobic in nature or due to the polarizability of the substituents.

$$\log 1/C = 0.146\ (\pm 0.03)(\text{MR-1,2}) + 1.068\ (\pm 0.13)\ (\text{D-1}) + 0.520\ (\pm 0.28)\ (\text{D-2}) + 0.429\ (\pm 0.14)\ (\text{D-3}) + 2.425\ (\pm 0.12) \quad (2)$$

n	r	s
108	0.935	0.258

$$\log 1/C = 0.211\ (\pm 0.05)\ (\pi\text{-1,2}) + 1.345\ (\pm 0.13)\ (\text{D-1}) + 0.620\ (\pm 0.29)\ (\text{D-2}) + 0.565\ (\pm 0.14)\ (\text{D-3}) + 2.440\ (\pm 0.12) \quad (3)$$

n	r	s
108	0.931	0.267

The degree of collinearity of the variables can be seen in the correlation matrix of Table III. The ambivalent character of eq 2 and 3 highlights the extreme importance of carefully studying and selecting substituents before undertaking a structure-activity study to avoid the frustration which so often results from collinearity in the independent variables.

The range of activity in Table I correlated by eq 2 or 3 is 1000-fold. On the average, eq 2 predicts the effective concentration of any given compound within a factor of ±1.8 (*i.e.*, antilog of s = ±1.8). Considering the extreme variation in chemical structure in which substituents as small as H, OH, and CH_3 must be accommodated in the same equation with substituents such as

This is indeed impressive. The overall *F* statistic for eq 2 is: $F_{4,103}$ = 179.2; $F_{4,60;\ \alpha\ 0.005}$ = 4.1. Equation 2 is a solid

Table III. Correlation Matrix for Collinearity between Variables. Values are r^2

	MR-1,2	π-1,2	D-1	D-2	D-3
MR-1,2	1.00	0.36	0.09	0.00	0.15
π-1,2		1.00	0.05	0.03	0.01
D-1			1.00	0.11	0.07
D-2				1.00	0.01
D-3					1.00

relationship based on 27 data points per variable (on the average).

The relative importance of the variables in eq 2 can be appreciated by following its stepwise development (eq 4-6). Equation 4 is the most important single-variable

$$\log 1/C = 1.28\ (\pm 0.19)\ (\text{D-1}) + 2.80\ (\pm 0.17) \quad (4)$$

n = 108, r = 0.785, s = 0.447

$$\log 1/C = 1.05\ (\pm 0.14)\ (\text{D-1}) + 0.19\ (\pm 0.03)\ (\text{MR-1,2}) + 2.42\ (\pm 0.14) \quad (5)$$

n = 108, r = 0.905, s = 0.308

$$\log 1/C = 0.99\ (\pm 0.13)\ (\text{D-1}) + 0.16\ (\pm 0.03)\ (\text{MR-1,2}) + 0.41\ (\pm 0.15)\ (\text{D-3}) + 2.48\ (\pm 0.12) \quad (6)$$

n = 108, r = 0.927, s = 0.273

equation although the equation linear in MR-1,2 was almost as important. This illustrates the qualitative importance of the second benzene ring which, even with the shortest bridge (four atoms), is probably located outside of the "active site" (where the amidine is bound). Equation 6 is found to be the most important three-variable equation. Generating all possible linear combinations of π-1,2, MR-1,2, σ-1, D-1, D-2, and D-3 yielded 63 equations. The equation with the lowest standard deviation (s = 0.245) contained a term in π-1,2 in addition to the terms in eq 2. While this equation was significant in terms of the F statistic ($F_{1,101}$ = 12.2), it in fact reduced the total variance by only 1.4%. Adding terms in (MR-1,2)^2 or (π-1,2)^2 to eq 2 or 3, respectively, did not reduce the variance in log $1/C$ nor did the use of cross-product terms MR-1,2·D-1; MR-1,2·D-3; D-1·D-3; MR-1,2·D-1·D-3 singly or in combination.

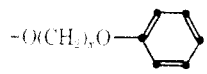
Discussion

Although a final decision cannot be made on MR-1,2 or π-1,2 on statistical grounds, it seems more likely that MR-1,2 is the parameter of choice. The reason for this is that the coefficient with π or log P in simple single-variable QSAR generally falls[20] in the range 0.4–1.2. In the relatively few examples studied where MR is the significant parameter and collinearity is not high, the coefficients with MR are often low. Coats[7] also reached the conclusion that MR is the significant descriptor, not π. However, the choice between MR and π is only a matter of opinion at this stage.

Since the coefficient with the MR term is positive, this suggests that a steric effect in the negative sense of steric hindrance is not involved. The positive term in MR indicates that, roughly speaking, the larger the substituents X and Y, the more effective the inhibitor. When MR-1,2 was factored into two terms, the size of the coefficients was identical within the confidence limits, indicating the same effect from substituents on each ring. This makes it seem unlikely that MR is related to substituents causing conformational changes in complement which are responsible for the inhibition. It seems unlikely (although it is not out of the question) that substituents on the widely separated rings would produce the same kinds of conformational change. Thus the most likely role for MR of the substituents would appear to be a binding of the inhibitor to the complement *via* dispersion forces (rather than desolvation which appears to be heavily associated with π) which results from the polarizability of the substituents. This problem has been analyzed by Agin, Hersh, and Holtzman.[21] Recent studies[18,22] confirm that one can expect to find clear-cut cases where π as well as MR rationalize substituent effects. In fact, this is what one should *a priori* expect. The well-known hydrophobic pockets in enzymes will relate to π and the polar spaces to MR.

Although Coats[7] found a role for σ in a small subset of the data in Table I, we have not uncovered an electronic role for substituents on either ring using the complete set. Coats' role for σ_m may be associated with the fact that a very large number of $O(CH_2)_xOC_6H_4Y$ groups are in the 3 position and D-1 is highly important.

The exact role of D-1 is not unambiguous. Two possible explanations come easily to mind. The large phenyl group attached to a flexible side chain might allow this ring to locate itself in a hydrophobic pocket. Since π for the functions



is roughly 2.0, if one assumes a coefficient of 0.5 with π, this would yield a contribution of about 1 to log $1/C$ (*i.e.*, about the same as 1.068 × 1 of eq 2). A slope of 0.5 with π terms is commonly observed.[20] On the other hand, one could rationalize the role of D-1 by saying that forcing this large group into the complement system produces an important conformational change which is responsible for the inhibition of complement. D-2 is not a very significant variable since only four data points are involved. Dropping this term from eq 2 gives a three-variable equation with r = 0.927 and s = 0.273. However, this indicator variable is significant in that it brings to light an important characteristic of the pyridine ring which could be used to advantage in the design of other derivatives.

D-3 is most interesting in that when NHCO functions are placed on the second ring, they bring about a special interaction in the 3 position which does not occur in the 2 or 4 position. While the difference is not large, it is quite significant and can be employed in the design of better congeners.

It is of interest that no special role could be found for the SO_2F function. This group does not appear to be causing inhibition by irreversible binding as Baker had expected. This may be overstating the case a bit since most of the most active inhibitors contain this function. However, **105** in Table I does not contain the SO_2F group and it is one of the four most active congeners. The activity of the three most active congeners is somewhat underpredicted by eq 2; this might suggest some small special role for the SO_2F group.

Neither terms in (MR-1,2)^2 nor (π-1,2)^2, when added to eq 2 or 3, reduced the variance. This indicates that more potent inhibitors could be made by using substituents with larger MR values. This is easy to do; moreover, since it is MR on ring 2 which is most likely the important descriptor and not π, one is in the advantageous position of being

able to select large polar groups which would not increase the overall lipophilicity of the molecule. This is of extreme importance for the design of drugs for *in vivo* use where one invariably finds that beyond a certain value of log P, activity falls off because of the random walk problem.[23] Guided by eq 2, it should be possible to make inhibitors at least ten times as active as those in Table I without the toxic SO_2F function. With log $1/C$ values in the range of 6-7 and low overall log P values, such compounds might be valuable for *in vivo* transplant studies.

Aside from the practical implications of eq 2, there is a great deal of intellectual satisfaction which can be taken from this correlation. To our knowledge, this constitutes the largest group of congeners correlated in a single equation. The previous "record" was that for a set of 102 antimalarials acting in mice.[24] More important, this work, especially when taken with that on other enzymic studies, emphasizes that one can now organize large amounts of enzyme substrate or inhibitor activity in quantitative fashion. Such equations not only have predictive values, but they enable one to compare the characteristics of different enzymes in numerical terms. We have found very few examples in our laboratory of sets of data on enzyme substrate or inhibitor activity which cannot be correlated using extrathermodynamic techniques.[25,26]

Success in uncovering linear free energy relationships in biochemical structure-activity relationships has (until recently) been the exception rather than the rule. Possibly the principal reason for this is the reluctance of workers to explore parameters other than the well-known Hammett σ constants. As we have become more aware in the last decade or so of the enormous importance of the role of hydrophobic interactions and polarizability in biochemical processes, it is now all too clear that one, in general, cannot expect to rationalize substituent effects purely in terms of electronic interactions of the substituent with the reaction center. One must consider electronic, steric, hydrophobic, and the polarizable effects of substituents.[27]

In any enzymic study one should attempt to formulate extrathermodynamic relationships early in the work. These are of enormous value in avoiding redundancy in the synthesis of congeners.[28] For example, in the present study, eq 5 could have been derived early in the work. Even though this is not a perfect correlation, with it one would have been guided to much better derivatives more quickly. Of course, as more compounds were tested, eq 5 could have been modified to yield still better guidance long before 108 derivatives had been made.

Finally, eq 2 illustrates the great advantage of the use of indicator variables in bringing order to structure-activity studies. Equation 2 can be viewed as the extrathermodynamic approach assisted by what is sometimes termed the Free-Wilson method.[29] No physical constants are employed in the Free-Wilson approch; the whole QSAR is formulated in terms of dummy variables. The disadvantages of this approach are that no physicochemical meaning can be easily attached to any of the variables and one must generally use a relatively low ratio of data points/variable. Equation 2 incorporates the best aspects of the two quantitative approaches to structure-activity studies.

In conclusion, it should be emphasized that while the predictive value of correlations such as eq 2 is important for the design of more selective inhibitors, one must not overlook the fact that eq 2 characterizes "substituent space" in the complement system in the region of the amidine binding site. As complement is studied with more molecular probes, further mapping of this system will be possible.

References

(1) E. Kutter and C. Hansch, *Arch. Biochem. Biophys.*, **135**, 126 (1969).
(2) L. Pauling and D. Pressman, *J. Amer. Chem. Soc.*, **67**, 1003 (1945).
(3) B. R. Baker and E. H. Erickson, *J. Med. Chem.*, **10**, 1123 (1967).
(4) M. Dixon and E. C. Webb, "Enzymes," Academic Press, New York, N. Y., 1958, pp 185-188, 247-266.
(5) H. J. Müller-Eberhard, *Advan. Immunol.*, **8**, 1 (1968).
(6) M. M. Mayer, *Sci. Amer.*, **229**, 54 (Nov 1973).
(7) E. A. Coats, *J. Med. Chem.*, **16**, 1102 (1973).
(8) B. R. Baker and E. H. Erickson, *J. Med. Chem.*, **12**, 408 (1969).
(9) B. R. Baker and M. Cory, *J. Med. Chem.*, **12**, 1049 (1969).
(10) B. R. Baker and M. Cory, *J. Med. Chem.*, **12**, 1053 (1969).
(11) B. R. Baker and M. Cory, *J. Med. Chem.*, **14**, 119 (1971).
(12) B. R. Baker and M. Cory, *J. Med. Chem.*, **14**, 805 (1971).
(13) C. Daniel and F. S. Wood, "Fitting Equations to Data," Wiley-Interscience, New York, N. Y., 1971, pp 55, 169, 203.
(14) T. Fujita, J. Iwasa, and C. Hansch, *J. Amer. Chem. Soc.*, **86**, 5175 (1964).
(15) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
(16) C. Hansch, A. Leo, and D. Nikaitani, *J. Org. Chem.*, **37**, 3090 (1972).
(17) J. Iwasa, T. Fujita, and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965).
(18) C. Hansch and C. Silipo, *J. Med. Chem.*, **17**, 661 (1974).
(19) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
(20) C. Hansch and W. J. Dunn III, *J. Pharm. Sci.*, **62**, 1 (1972).
(21) D. Agin, L. Hersh, and D. Holtzman, *Proc. Nat. Acad. Sci. U. S.*, **53**, 952 (1965).
(22) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970).
(23) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).
(24) P. N. Craig and C. Hansch, *J. Med. Chem.*, **16**, 661 (1973).
(25) J. E. Leffler and E. Grunwald, "Rates and Equilibria of Organic Reactions," Wiley, New York, N. Y., 1963.
(26) N. B. Chapman and J. Shorter, "Advances in Linear Free Energy Relationships," Plenum Press, New York, N. Y., 1972.
(27) C. Hansch in "Structure-Activity Relationships," Vol. I, Section 5, C. J. Cavallito, Ed., Pergamon Press, Oxford, England, 1973, p 75 (International Encyclopedia of Pharmacology and Therapeutics).
(28) C. Hansch, S. Unger, and A. B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
(29) P. N. Craig, *Advan. Chem. Ser.*, **No.** 114, 115 (1972).