[Chem. Pharm. Bull. 23(2) 437—444 (1975)] UDC 547.821'533.09.03

Structure-Activity Relationships in Immunochemistry. III.[1] Inhibition of Complement by Benzyl Pyridinium Ions[2]

Masafumi Yoshimoto,[3a] Corwin Hansch and Priscilla Y.C. Jow[3]

Department of Chemistry, Pomona College[3]

(Received August 24, 1974)

A quantitative structure-activity relationship for benzyl pyridinium ions inhibiting complement has been formulated. The best equation describing the relationship is $\log 1/C = 0.18\pi\text{-}1 + 0.46\pi\text{-}2 + 1.01\sigma^{+}\text{-}1 + 0.72D\text{-}1 + 2.50$ where C is the molar concentration causing 50% inhibition of complement, π-1 is the hydrophobic constant for substituents on the pyridine moiety and π-2 for substituents on the benzyl moiety, σ^{+}-1 refers to the electronic effects of substituents on the pyridine ring, and the indicator variable D-1 refers to the SO_2F function in the 2-position of the benzene ring. This expression correlates 69 derivatives with a correlation coefficient of 0.939 and a standard deviation of 0.198. The implications of this relationship for the design of inhibitors is discussed.

As our understanding of biochemistry and molecular biology increases, we are presented with more possibilities for chemical alteration and control of the processes of life. Control of the immune response offers the medicinal chemist opportunities in the field of organ transplant as well as the important area of control of the autoimmune response. This latter process appears to be most important in various inflammatory and allergic reactions.

There are a number of routes open to the medicinal chemist for control of the immune response: 1) inhibition of antibody formation; 2) modulation of formed antibodies by means of haptens; 3) inhibition of the complement system.

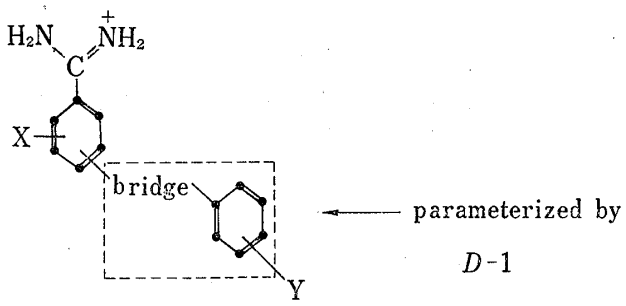
Medicinal chemistry is now in a position to explore any or all of these routes. For example, many anticancer drugs appear to inhibit antibody formation. It has long been established that haptens can easily be designed to prevent the union of antibody with antigen. In the first paper in this series[4] it was shown that the structure-activity relationship for the union of hapten with antibody could be formulated in mathematical terms. It was shown in the second paper in this series[1] that the structure-activity relationship for 108 benzamidines inhibiting complement could be described quantitatively by equation (1). C in equation

$$\log\frac{1}{C} = 0.15\text{MR-}1,2 + 1.07D\text{-}1 + 0.52D\text{-}2 + 0.43D\text{-}3 + 2.42$$

$$n=108,\ r=0.935,\ s=0.258 \qquad (1)$$

(1) is the molar concentration of benzamidine causing 50% inhibition of lyophilized guinea pig complement. MR-1,2 refers to the molar refractivity (scaled by 0.1) of the substituents X and Y of I. The indicator variable D-1 parameterizes a variety of bridges such as $O(CH_2)_xO$ and the benzene ring (D-1 assumes a value of 1 for the presence of a bridge and 0.0 when such a bridge is absent). The indicator variable D-2 accounts for a pyridine ring attached at the end of a side chain. D-3 was used to account for the special activating effect of the function

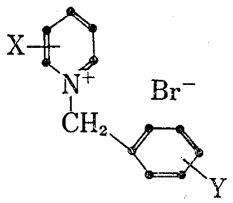
1) Part II: C. Hansch and M. Yoshimoto, *J. Med. Chem.*, **17**, 1160 (1974).
2) This research was supported by Grant CA-11110 from the National Institutes of Health and by the Sankyo Company of Tokyo, Japan.
3) Location: *Claremont, California 91711 U.S.A.*; *a*) Visiting scientist from the Sankyo Company.
4) E. Kutter and C. Hansch, *Arch. Biochem. Biophys.*, **135**, 126 (1969).



parameterized by

D-1

I



II

3-NHC(=O)–X–C_6H_5 attached to the second ring. In equation (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 MR-1,2 in this quantitative structure-activity relationship (QSAR) indicates that the larger and more polarizable substituents make good inhibitors and the coefficients with *D*-1, *D*-2, and *D*-3 indicate that the above-mentioned bridges and special functions are preferable. Our success in correlating the benzamidine inhibitors of complement prompted us to attempt the formulation of other QSAR for complement inhibitors.

In addition to the formulation of the QSAR, we have become interested in the characterization of the different types of space in enzymes in which substituents from substrates or inhibitors might be positioned. From our present knowledge there appear to be three types of space in which a substituent might fall: 1) There is considerable evidence for the existence of hydrophobic pockets;[5,6] interaction of substituents with this type of space can be modeled using the hydrophobic parameter π.[7,8] 2) A substrate or inhibitor may be bound to a macromolecule in such a way that substituents in a given position do not establish contact with elements of the macromolecule but remain in the aqueous phase. Under such conditions no effect of the moiety in terms of dispersion forces or desolvation can be seen.[9,10] 3) Finally, substituents may find themselves not in hydrophobic space or circumambient aqueous space but, instead, in a milieu of polar groups of the macromolecule. We have been exploring the possibility that the interaction of substituents in such polar areas might be modeled by the use of molar refractivity (MR).[1,6,11–13] Unfortunately, unless considerable care[14] is taken in making congeners, one often ends with a set in which π and MR are highly collinear. Under such conditions it is not possible to deduce the character of substituent space.[1] One of the important reasons for studying the interactions of sets of congeners with macromolecules is to learn more about the different kinds of "substituent space." This problem is of considerable concern to us in the present study.

The late B.R. Baker and his students carried out an extensive study of the inhibitory action of derivatives of N-benzylpyridinium bromide (II) on guinea pig complement. This work was part of Baker's generalized search for drugs to inhibit mammalian proteolytic enzymes. As he pointed out,[15] the complement system is a complex mixture of serum pro-

5) B.R. Baker, "Design of Active-Site-Directed Irreversible Enzyme Inhibitors," John Wiley, New York, 1967.
6) C. Hansch and E. Coats, *J. Pharm. Sci.*, **59**, 731 (1970).
7) T. Fujita, J. Iwasa and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964).
8) A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).
9) C. Hansch, E.W. Deutsch and R.N. Smith, *J. Am. Chem. Soc.*, **87**, 2738 (1965).
10) C. Hansch, K.H. Kim and R.H. Sarma, *J. Am. Chem. Soc.*, **95**, 6447 (1973).
11) C. Hansch and C. Silipo, *J. Med. Chem.*, **17**, 661 (1974).
12) C. Silipo and C. Hansch, *Mol. Pharmacol.*, **10**, 954 (1974).
13) C. Silipo and C. Hansch, *Farmaco*, **30**, 35 (1975).
14) C. Hansch, S. Unger and A.B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
15) B.R. Baker and J.A. Hurlbut, *J. Med. Chem.*, **12**, 677 (1969).

TABLE I. Constants Used for Deriving Equations (2) and (3)

No.	X	Y	$\log \frac{1}{C}$ Obsd.[a]	$\log \frac{1}{C}$ Calcd.[b]	$\lvert\Delta \log \frac{1}{C}\rvert$	$\pi-1$	MR−2	$D-1$	σ^+-1	$\pi-2$
1	3-(3,4-$Cl_2C_6H_3OCH_2CONHCH_2$)	H	2.460	2.803	0.34	2.080	0.100	0.0	−0.070	0.0
2	4-CH_3	4-SO_2F	2.480	2.460	0.02	0.560	0.870	0.0	−0.310	0.370
3	H	4-SO_2F	2.550	2.672	0.12	0.0	0.870	0.0	0.0	0.370
4[c]	2,3-(CH=CH)$_2$	4-SO_2F	2.550	2.656	0.11	1.320	0.870	0.0	−0.140	0.130
5	3-CH_3CONH	4-SO_2F	2.640	2.709	0.07	−0.970	0.870	0.0	0.210	0.370
6	3-$C_6H_5OCH_2CONH$	4-SO_2F	2.690	3.002	0.31	0.660	0.870	0.0	0.210	0.370
7	3-$C_6H_5CH_2CONH$	4-SO_2F	2.710	2.971	0.26	0.490	0.870	0.0	0.210	0.370
8	3-C_6H_5CONH	4-SO_2F	2.720	2.780	0.06	0.490	0.870	0.0	0.020	0.370
9[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONHCH_2$)	4-MeO	2.810	2.794	0.02	2.080	0.790	0.0	−0.070	−0.020
10[c]	4-$C_6H_5CH_2$	4-SO_2F	2.890	2.720	0.17	2.010	0.870	0.0	−0.310	−0.370
11	4-C_6H_5	4-SO_2F	2.890	2.842	0.05	1.960	0.870	0.0	−0.180	0.370
12	3-(2-$ClC_6H_4OCH_2CONH$)	4-SO_2F	2.900	3.129	0.23	1.370	0.870	0.0	0.210	0.370
13	3-(L-4-Ts-NH[$PhCH_2$]$CHCONHCH_2$)	4-SO_2F	2.920	2.946	0.03	1.920	0.870	0.0	−0.070	0.370
14	4-$C_6H_5CH_2CH_2$	4-SO_2F	2.960	2.837	0.12	2.660	0.870	0.0	−0.310	0.370
15[c]	4-$C_6H_5(CH_2)_4$	4-SO_2F	3.000	3.017	0.02	3.660	0.870	0.0	−0.310	0.370
16[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	4-NO_2	3.000	2.886	0.11	2.080	0.740	0.0	0.210	−0.430
17	4-$C_6H_5(CH_2)_3$	4-SO_2F	3.020	2.927	0.09	3.160	0.870	0.0	−0.310	0.370
18	3-(3,4-$Cl_2C_6H_3OCH_2CONHCH_2$)	4-SO_2F	3.040	2.975	0.07	2.080	0.870	0.0	−0.070	0.370
19[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	H	3.050	3.085	0.04	2.080	0.100	0.0	0.210	0.0
20[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	4-CH_3O	3.050	3.076	0.03	2.080	0.790	0.0	0.210	−0.020
21	3-$C_6H_5(CH_2)_4$	4-SO_2F	3.050	3.259	0.21	3.660	0.870	0.0	−0.070	0.370
22	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	3-NO_2	3.060	2.886	0.17	2.080	0.740	0.0	0.210	−0.430
23	3-(3-$ClC_6H_4OCH_2CONH$)	4-SO_2F	3.060	3.129	0.07	1.370	0.870	0.0	0.210	0.370
24	3-(4-$ClC_6H_4OCH_2CONH$)	4-SO_2F	3.080	3.129	0.05	1.370	0.870	0.0	0.210	0.370
25	3-(3,4-$Cl_2C_6H_3OCH_2CONHCH_2$)	3-SO_2F	3.100	2.975	0.13	2.080	0.870	0.0	−0.070	0.370
26	H	2-SO_2F	3.150	3.395	0.25	0.0	0.870	1.000	0.0	0.370
27	2-(3,4-$Cl_2C_6H_3CH_2CH_2$)	4-SO_2F	3.170	3.092	0.08	4.080	0.870	0.0	−0.310	0.370
28	2-(3,4-$Cl_2C_6H_3[CH_2]_4$)	4-SO_2F	3.220	3.272	0.05	5.080	0.870	0.0	−0.310	0.370
29	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	3-Cl-4-SO_2F	3.250	3.304	0.05	2.080	1.370	0.0	−0.070	1.080
30	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	4-SO_2F	3.270	3.257	0.01	2.080	0.870	0.0	0.210	0.370
31	3-(L-4-Ts-NH[$PhCH_2$]$CHCONH$)	4-SO_2F	3.280	3.228	0.05	1.920	0.870	0.0	0.210	0.370
32[c]	3-(2,4-$Cl_2C_6H_3OCH_2CONH$)	4-SO_2F	3.300	3.257	0.04	2.080	0.870	0.0	0.210	0.370
33[c]	4-(3,4-$Cl_2C_6H_3O[CH_2]_3$)	4-SO_2F	3.300	3.173	0.13	4.530	0.870	0.0	−0.310	0.370
34[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	4-CH_3	3.310	3.345	0.04	2.080	0.570	0.0	0.210	0.560
35	3-(2,3-$Cl_2C_6H_3OCH_2CONH$)	4-SO_2F	3.310	3.257	0.05	2.080	0.870	0.0	0.210	0.370
36	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	4-Cl-3-SO_2F	3.310	3.304	0.01	2.080	1.370	0.0	−0.070	1.080
37[c]	3-(3,4-$Cl_2C_6H_3[CH_2]_4$)	4-SO_2F	3.310	3.514	0.20	5.080	0.870	0.0	−0.070	0.370
38	3-(3,4-$Cl_2C_6H_3CH_2CH_2$)	4-SO_2F	3.320	3.334	0.01	4.080	0.870	0.0	−0.070	0.370
39	4-(3,4-$Cl_2C_6H_3CH_2CH_2$)	4-SO_2F	3.340	3.092	0.25	4.080	0.870	0.0	−0.310	0.370
40[c]	3-(3,4-$Cl_2C_6H_3OCH_2CONH$)	3-SO_2F	3.360	3.257	0.10	2.080	0.870	0.0	0.210	0.370
41[c]	4-(3,4-$Cl_2C_6H_3[CH_2]_4$)	4-SO_2F	3.380	3.272	0.11	5.080	0.870	0.0	−0.310	0.370
42	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	2-Cl-4-SO_2F	3.430	3.304	0.13	2.080	1.370	0.0	−0.070	1.080
43	H	6-Cl-2-SO_2F	3.430	3.725	0.30	0.0	1.370	1.000	0.0	1.080
44	2-(3,4-$Cl_2C_6H_3[CH_2]_4$)	2-SO_2F	3.640	3.995	0.36	5.080	0.870	1.000	−0.310	0.370
45	3-(3,4-Cl_2PhOCH_2CONH)	3-Cl-4-SO_2F	3.660	3.587	0.07	2.080	1.370	0.0	0.210	1.080
46	3-(3,4-Cl_2PhOCH_2CONH)	4-Cl-3-SO_2F	3.720	3.587	0.13	2.080	1.370	0.0	0.210	1.080
47	2-(3,4-$Cl_2C_6H_3CH_2CH_2$)	2-SO_2F	3.720	3.815	0.10	4.080	0.870	1.000	−0.310	0.370
48	4-(3,4-$Cl_2C_6H_3CH_2CH_2$)	2-SO_2F	3.720	3.815	0.10	4.080	0.870	1.000	−0.310	0.370
49	3-(3,4-$Cl_2C_6H_3CH_2CH_2$)	2-SO_2F	3.720	4.048	0.33	4.030	0.870	1.000	−0.070	0.370
50	3,4-(CH=CH)$_2$	6-Cl-2-SO_2F	3.740	3.709	0.03	1.320	1.370	1.000	−0.140	0.840
51	3-(3,4-Cl_2PhOCH_2CONH)	3-Cl-2-SO_2F	3.770	3.587	0.18	2.080	1.370	0.0	0.210	1.080
52	4-$C_6H_5(CH_2)_4$	6-Cl-2-SO_2F	3.800	4.070	0.27	3.660	1.370	1.000	−0.310	1.080
53	3-$C_6H_5CH_2CH_2$	6-Cl-2-SO_2F	3.800	4.132	0.33	2.660	1.370	1.000	−0.070	1.080
54	4-$C_6H_5(CH_2)_3$	6-Cl-2-SO_2F	3.800	3.980	0.18	3.160	1.370	1.000	−0.310	1.080
55	4-(3,4-$Cl_2C_6H_3[CH_2]_4$)	2-SO_2F	4.000	3.995	0.01	5.080	0.870	1.000	−0.310	0.370
56	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	3-Cl-2-SO_2F	4.070	4.028	0.04	2.080	1.370	1.000	−0.070	1.080
57	2,3-(CH=CH)$_2$	6-Cl-2-SO_2F	4.150	3.709	0.44	1.320	1.370	1.000	−0.140	0.840
58	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	4-Cl-2-SO_2F	4.170	4.028	0.14	2.080	1.370	1.000	−0.070	1.080
59	3-(3,4-Cl_2PhOCH_2CONH)	3-Cl-2-SO_2F	4.180	4.310	0.13	2.080	1.370	1.000	0.210	1.080
60	3-$C_6H_5(CH_2)_4$	6-Cl-2-SO_2F	4.220	4.311	0.09	3.660	1.370	1.000	−0.070	1.080
61	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	6-Cl-2-SO_2F	4.240	4.028	0.21	2.080	1.370	1.000	−0.070	1.080
62	3-(3,4-Cl_2PhOCH_2CONH)	2-SO_2F	4.290	3.980	0.31	2.080	0.870	1.000	0.210	0.370
63	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	5-Cl-2-SO_2F	4.310	4.028	0.28	2.080	1.370	1.000	−0.070	1.080
64	3-(3,4-Cl_2PhOCH_2CONH)	4-Cl-2-SO_2F	4.320	4.310	0.01	2.080	1.370	1.000	0.210	1.080
65	3-(3,4-$Cl_2C_6H_3[CH_2]_4$)	6-Cl-2-SO_2F	4.320	4.566	0.25	5.080	1.370	1.000	−0.070	1.080
66	3-(3,4-Cl_2PhOCH_2CONH)	5-Cl-2-SO_2F	4.330	4.310	0.02	2.080	1.370	1.000	0.210	1.080
67	3-(3,4-$Cl_2PhOCH_2CONHCH_2$)	2-SO_2F	4.380	3.698	0.68	2.080	0.870	1.000	−0.070	0.370
68	3-(3,4-Cl_2PhOCH_2CONH)	6-Cl-2-SO_2F	4.460	4.310	0.15	2.080	1.370	1.000	0.210	1.080
69	3-(3,4-$Cl_2C_6H_3[CH_2]_4$)	2-SO_2F	4.570	4.237	0.33	5.080	0.870	1.000	−0.070	0.370

a) calculated from results of Baker, *et al.*[15,17,18]
b) calculated using equation (3)
c) 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.
Ph=phenyl

teases used for the destruction of foreign cells by lysis and, hence, can also reject foreign organ or tissue transplants. Since the complement system has both "tryptic" and "chymotryptic" properties, it can be inhibited (as measured by complement-antibody-mediated lysis of sheep red cells) by certain inhibitors of trypsin or chymotrypsin. Benzamidines (trypsin-type inhibitors) have been studied[1,16] in this respect. The analysis of Baker's results[15,17,18] with structures of type II constitutes the subject of this paper.

Method

Table I contains activities and the essential constants used in the formulation of equations (2—10). C in log $1/C$ represents the molar concentration of inhibitor required for 50% inhibition which has been estimated graphically from Baker's data by making a linear plot of activity *vs.* concentration. To parameterize the structural features of II, π-1, MR-1, σ-1, and σ^+-1 were used to characterize the effects of X attached to the parent pyridine ring. π-2, MR-2, and σ-2 were used for substituents Y on the benzene ring. During the course of preliminary work, the unusual activity of the 2-SO_2F group Y attached to the benzene ring became apparent. To cast this effect in numerical terms, an indicator variable D-1 was assigned the value of 1.00 for this function; D-1 was given the value of 0.00 for all other cases. It is very interesting that when such fluorosulfonyl functions are in the 3- or 4-position, they do not produce any special effect; their contribution to activity is accounted for in terms of π, MR, and σ. This problem is considered in the section on Discussion.

Table II lists log P and π values not previously reported. These were measured by the usual procedure.[7] The motive for determining these values was to check the additivity principles[8] for the benzyl pyridinium system. Since strong electron-attracting groups are known to have the greatest affect on partition coefficients[7,19] two values were measured for the SO_2F function which has one of the highest σ constants known. The value of 0.37 was determined from using compounds I and II, and 0.13 from compounds V and VI. In earlier studies[20] $\pi_{SO_2F}=0.05$ was found from $CH_3C_6H_4$-4-SO_2F and $\pi_{NO_2}=-0.28$ was

Table II. Partition Coefficients and π Constants

	Compound	log P[a]	Substituent	π
I	4-phenyl-N-benzylpyridinium bromide	0.04±0.05[b]	$C_5H_4\overset{+}{N}CH_2C_6H_5\cdot Br^-$	−2.09[c]
II	4-phenyl-N-(4-fluorosulfonylbenzyl)-pyridinium bromide	0.41±0.03	SO_2F	0.37[d]
III	N-(4-fluorosulfonylbenzyl)pyridinium bromide	−1.44±0.12	$C_5H_4\overset{+}{N}CH_2C_6H_5\cdot Br^-$	−1.81[e]
IV	4-methyl-N-(4-fluorosulfonylbenzyl)-pyridinium bromide	−1.09±0.17	CH_3	0.55[f]
V	N-benzylquinolinium bromide	−0.82±0.07	$C_5H_4\overset{+}{N}CH_2C_6H_5\cdot Br^-$	−2.14[g]
VI	N-(4-fluorosulfonylbenzyl)-quinolinium bromide	−0.69±0.16	SO_2F	0.13[h]
VII	4-phenylpropyl-N-(4-nitrobenzyl)-pyridinium bromide	0.72±0.01	NO_2	−0.43[i]

a) In the determination, 0.1N sodium chloride was used as the aqueous phase and the observed values were extrapolated to infinite dilution.
b) 95% confidence intervals
c) $\pi=\log P-\log P_{C_6H_6}=\log P-2.13$
d) $\pi=\log P-\log P_I=\log P-0.04$
e) $\pi=\log P-\pi_{SO_2F}=\log P-0.37$
f) $\pi=\log P-\pi_{SO_2F}-\pi_{C_5H_4\overset{+}{N}CH_2C_6H_5\cdot Br^-}=\log P-0.37-(-2.01)$: −2.01 is the average value of three cases (I, III, and V)
g) $\pi=\log P-\pi_{(CH=CH)_2}=\log P-1.32$
h) $\pi=\log P-\log P_V=\log P-(-0.82)$
i) $\pi=\log P-(\pi_{C_6H_5(CH_2)_2}+\pi_{CH_2}+\pi_{C_5H_4\overset{+}{N}CH_2C_6H_5\cdot Br^-})=\log P-2.66-0.50-(-2.01)$

16) *a*) B.R. Baker and E.H. Erickson, *J. Med. Chem.*, **12**, 408 (1969); *b*) B.R. Baker and M. Cory, *J. Med. Chem.*, **12**, 1049, 1053 (1969); *c*) B.R. Baker and M. Cory, *J. Med. Chem.*, **14**, 119, 805 (1971).
17) B.R. Baker and J.A. Hurlbut, *J. Med. Chem.*, **12**, 902 (1969).
18) B.R. Baker and M.H. Doll, *J. Med. Chem.*, **14**, 793 (1971).
19) C. Hansch, A. Leo and D. Nikaitani, *J. Org. Chem.*, **37**, 3090 (1972).
20) C. Hansch, A. Leo, S.H. Unger, K.H. Kim, D. Nikaitani and E.J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).

found from nitrobenzene. $\pi_{NO_2}=-0.43$ in Table II. These experimental π values were used in the correlation studies for these strong electron-attracting groups. The other values in Table I were taken from the literature[20] or estimated from additivity principles.[8,19] Other known values are: $\pi_{CH_2NHCOCH_2OPh}=\pi_{NHCOCH_2OC_6H_5}=0.66$,[11] $\pi_{CH_2NHCO(CH_2)_2C_6H_5}=\pi_{NHCO(CH_2)_2C_6H_5}=0.91$,[21] and $\pi_{NHCOC_6H_5}=\pi_{NHCOCH_2C_6H_5}=0.49$.[21]

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. Such corrections would be on the order of 0.2 to 0.3 π units. Since the coefficient of π is about 0.2, these approximations could only lead to small errors.

The MR values in Table I are taken from our recent compilation or were calculated in the same manner.[20] MR has been scaled by 0.1 which makes this vector somewhat more equiscalar with respect to π. Since $MR_H=0.1\times1.03=0.10$ when two substituents are present on a ring, the value of 0.10 for one hydrogen must be subtracted. σ and σ^+ values are not available for some functions. Values for these were estimated from similar functions.[20] The values of σ^+ are taken from the compilation of Leffler and Grunwald.[22]

Compounds I—VII in Table II were prepared by the usual methods. The purity of all compounds was checked by thin-layer chromatography. All new compounds gave satisfactory carbon-hydrogen analyses.

Results

The formulation of a QSAR for a very complex set of congeners such as that in Table I is a difficult problem to solve, but there are a variety of approaches one can take. Many hypotheses must be tested and rejected or included in more encompassing hypotheses before a self-consistent pattern comes into focus. The approach we employed in the present instance was to attempt to correlate the data in each of the publications by Baker and his colleagues; from these studies there appeared to be no reason not to merge all of the data. In order to check the special effect of the fluorosulfonyl group in each of the 2-, 3-, and 4-positions, indicator variables *D*-2-SO_2F, *D*-3-SO_2F, and *D*-4-SO_2F were studied. It soon became apparent that only *D*-2-SO_2F possessed special significance. This variable was employed as *D*-1. A study of σ-1 and σ^+-1 on the pyridine ring and σ-2 on the benzene ring brought out the importance of σ^+-1. The two best QSAR to emerge from a consideration of hundreds of equations are:

$$\log\frac{1}{C} = 0.183(\pm0.04)\pi\text{-}1 + 0.624(\pm0.19)\text{MR-}2 + 1.024(\pm0.29)\sigma^+\text{-}1 + 0.724(\pm0.12)D\text{-}1 + 2.122(\pm0.21)$$
$$n=69,\ r=0.939,\ s=0.199 \qquad (2)$$

$$\log\frac{1}{C} = 0.180(\pm0.04)\pi\text{-}1 + 0.464(\pm0.14)\pi\text{-}2 + 1.008(\pm0.28)\sigma^+\text{-}1 + 0.723(\pm0.12)D\text{-}1 + 2.500(\pm0.13)$$
$$n=69,\ r=0.939,\ s=0.198 \qquad (3)$$

The quality of the correlation of the two equations is identical. This is largely due to the high collinearity between π-2 and MR-2 (see Table III). The figures in parentheses in these equations are the 95% confidence limits. A most significant aspect of this study is that every data point was included in the formulation of equations (2—3). This is also true for equation (1). Such results are rarely found and merit special consideration.

Although one might anticipate that the magnitude of the coefficients with the π-1 and π-2 terms of equation (3) are such that one might expect them to preclude lumping of these two terms, equation (4) suggests that this is possible.

$$\log\frac{1}{C} = 0.192(\pm0.04)\pi\text{-}1,2 + 1.112(\pm0.30)\sigma^+\text{-}1 + 0.131(\pm0.10)\sigma\text{-}2 + 0.752(\pm0.13)D\text{-}1 + 2.468(\pm0.15)$$
$$n=69,\ r=0.931,\ s=0.210 \qquad (4)$$

21) J. Iwasa, T. Fujita and C. Hansch, *J. Med. Chem.*, **8**, 150 (1965).
22) J.E. Leffler and E. Grunwald, "Rates and Equilibria of Organic Reactions," Wiley, New York, N.Y., 1963, p. 204.

Table III. Correlation Matrix for Collinearity Between Variables

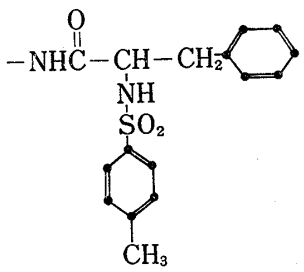
	π-1	π-2	π-1, 2	MR-1	MR-2	MR-1, 2	σ^+-1	σ-2	D-1
π-1	1.00	0.00	0.93	0.18	0.00	0.18	0.29	0.00	0.03
π-2		1.00	0.06	0.00	0.79	0.03	0.00	0.56	0.26
π-1, 2			1.00	0.17	0.04	0.20	0.27	0.05	0.10
MR-1				1.00	0.00	0.97	0.03	0.03	0.01
MR-2					1.00	0.02	0.00	0.66	0.26
MR-1, 2						1.00	0.03	0.00	0.00
σ^+-1							1.00	0.02	0.03
σ-2								1.00	0.27
D-1									1.00

Values are r^2.

Although the quality of the correlation by equation (4) is essentially as good as that of equations (2—3), a term in σ-2 is necessary to achieve this. It is apparent from Table III that there is high collinearity between σ-2 and π-2; hence σ-2 in equation (4) is likely serving as a partial substitute for π-2.

These problems of collinearity highlight the extreme importance of carefully studying and selecting substituents before undertaking a structure-activity study in order to avoid the frustration which so often results from collinearity in the "independent" variables.

The range of activity in Table I correlated by equations (2—4) is 130-fold. On the average, these equations predict the effective concentration within a factor of ± 1.6 (*i.e.*, antilog of $s=1.6$). 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 as gross as:



the correlation is indeed impressive. Our standard for terming equations (2—3) the "best" equations is that of all of the hundreds of equations considered, these equations had the lowest standard deviation (except for equation [10]) with all terms justified by the stepwise application of the F test. The overall F statistic for equation (3) is: $F_{4,64}=118.9$; $F_{4,60,\alpha.001}=5.31$. Equation (3) is a solid relationship based on 17 data points per variable (on the average).

The relative importance of the variables in equation (3) can be appreciated by following its stepwise development:

$$\log\frac{1}{C} = 0.94D\text{-}1 + 3.08$$

$$n=69,\ r=0.813,\ s=0.328 \qquad (5)$$

$$\log\frac{1}{C} = 0.42\pi\text{-}2 + 0.76D\text{-}1 + 2.92$$

$$n=69,\ r=0.851,\ s=0.297 \qquad (6)$$

$$\log\frac{1}{C} = 1.03\sigma^+\text{-}1 + 0.17\pi\text{-}1 + 0.92D\text{-}1 + 2.71$$

$$n=69,\ r=0.897,\ s=0.253 \qquad (7)$$

Equation (5) is the most important single-variable equation (lowest standard deviation), equation (6) is the most significant two-variable result, equation (7) is the most important three-variable equation, and equation (3) the most important four-variable equation. Equation (5) brings out the great importance of the 2-SO_2F function. Equation (6) is a significant improvement over equation (5): $F_{1,66}=15.4$; $F_{1,60,\alpha.001}=12.0$. Equation (7) is an unambiguous improvement over equation (6); $F_{1,65}=26.4$. Going from equation (7) to equation (3), $F_{1,64}=41.6$.

The most important two-variable equation considering MR-1,2 is:

$$\log\frac{1}{C} = 0.13\text{MR-}1,2 + 0.94D\text{-}1 + 2.37$$
$$n=69,\ r=0.897,\ s=0.251 \qquad (9)$$

A still further reduction in the variance was achieved by the addition of a term in $(\pi\text{-}1,2)^2$:

$$\log\frac{1}{C} = 0.39\pi\text{-}1,2 - 0.034(\pi\text{-}1,2)^2 + 1.02\sigma^+\text{-}1$$
$$+ 0.12\sigma\text{-}2 + 0.76D\text{-}1 + 2.24 \qquad n=69.\ r=0.944,\ s=0.191 \qquad (10)$$
$$\text{ideal } \pi\text{-}1,2\text{: } \pi - 1,2_0 = 5.74$$

While equation (10) is a significant improvement over equation (4) ($F_{1,63}=14.1$), it in fact reduces the variance by only 2.5%.

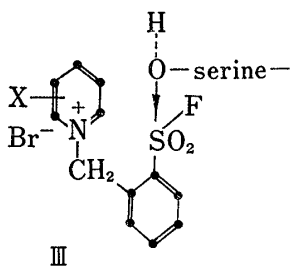
Discussion

At present a distinction cannot be made between equations (2—3). To advance our understanding of the relative importance of MR *vs.* π, more derivatives must be made and tested using cluster analysis as an aid in planning the compounds to be studied.[14]

As mentioned above, a salient feature of this work and the previous study of complement inhibitors is that all of the large number of derivatives tested are well fit and can be utilized to formulate the regression equations. We find this to be quite unusual since one usually finds a small percent of outliers in QSAR work which must be omitted to obtain a high correlation. It seems unlikely that the reason for this is unusually high quality of the biological test results. A possible explanation is that the set of macromolecules which make up complement is not very demanding sterically in the region where the positively charged function binds the inhibitors. The high correlations found with equations (1—3) are taken to mean that all congeners must be acting on the same center in the complement system. Except for the positively charged function, the rest of the molecular features (except 2-SO_2F and σ^+) are quite nonspecific.

Needless to say, the 2-SO_2F function on Y deserves most careful consideration. One of Baker's tenets[5] in the design of irreversible inhibitors was that in achieving specificity *via* covalent bond formation by nucleophilic substitution, the position the reactive function assumed in substituent space is most important. Baker's idea is nicely brought out by the 2-SO_2F function. Whence comes this special character ?

It is well known that in many hydrolases a serine hydroxyl plays a major role in the cleavage of peptide or ester bonds. For example, with chymotrypsin it is known that SER-OH-195 can be sulfonylated by sulfonyl fluoride to cause irreversible inhibition.[23] Structure III suggests how such a reaction might be greatly favored by SO_2F in the *ortho* position of the benzyl moiety. In this view the positively charged N could serve to anchor the inhibitor near the critical serine hydroxyl so that sulfonylation is favorable to the 2-SO_2F but not for the 3- and 4-isomers. One could even speculate that the positive charge of nitrogen might interact with the lone pair of electrons of the serine hydroxyl of the active site to more precisely position the 2-SO_2F function.



III

One should consider the relative importance of the various terms of equation (3), for example, for the design of more potent inhibitors. For equation (5), r^2 is 0.661, indicating

23) *a*) D.E. Fahrney and A.M. Gold, *J. Am. Chem. Soc.*, **85**, 997 (1963); *b*) P.B. Sinyler, B.A. Jeffery, B.W. Mattews and D.M. Blow, *J. Mol. Biol.*, **15**, 175 (1966); *c*) D.M. Blow, J.J. Birletoft and B.S. Hartley, *Nature*, **221**, 337 (1969); *d*) S.S. Wong, K. Quiggle, C. Triplet and L.J. Berliner, *J. Biol. Chem.*, **249**, 1678 (1974).

that D-1 explains 66% of the variance in the data. For equation (6), r^2 is 0.725, indicating that π-2 accounts for 6% of the variance and, for equation (7), r^2 is 0.804, suggesting that σ^+ accounts for 8% of the variance. We find r^2 of 0.881 with equation (3), indicating that 8% of the variance is accounted for by π-1. The above method of accounting is rough and, of course, assumes complete orthogonality of the vectors.

The small role for π-1 is most interesting in view of the large variation (6 powers of 10) in these functions. Since neither π-1 nor MR-1 does much to correlate this enormous amount of variance, it seems most likely that substituents in the 3- and 4-positions of the pyridine ring are placed largely in the aqueous phase and, hence, out of contact with the macromolecules of complement.

It is clear from the above delineation of the roles of the variables that, starting with a 2-SO_2F function, some further activity could be obtained by placing strong electron-withdrawing functions on the pyridine ring. Increasing π-2 and MR-2 offers the possibility for more activity; however, in the light of equation (10), one would want to keep π-1 and MR-1 small. The range in π-2 in Table I is not wide and the coefficient with π-2 in equation (3) or MR-2 in equation (2) is of moderate size so that one could very likely gain 0.5 to 1.0 units of log $1/C$ by proper modification on the benzyl moiety. It is interesting that Baker pushed large groups in X so hard while neglecting what appears to be the more profitable Y position.

The fact that the few derivatives in Table I which do not contain a SO_2F function are well predicted and, MR or π being equal, are grouped with the 3-SO_2F and 4-SO_2F functions brings out the lack of reactivity of SO_2F in these positions; in these positions it has no more activity than a chemically inert function such as Cl. This result confirms Baker's view that, by properly positioning the SO_2F function, one can greatly increase its reaction possibilities with nucleophilic moieties.

Comparing the results of Baker's studies embodied in equation (1) with equations (2—3), we find that in each case it has been necessary to employ indicator variables to account for special interactions of certain functions. Sometimes such reactions can be inferred simply by inspection of the data as with 2-SO_2F; in other instances they are harder to discern. The use of indicator variables can be most helpful in establishing the importance of such reactions.

In summary, it can be said that the good correlations with equations (1—3) establish the fact that substituent constants employed with regression analysis can be used in the design of *in vitro* inhibitors of complement. To move the *in vitro* work to *in vivo* work, the overall lipophilicity of the inhibitors must be more carefully controlled.[24] The present results provide an approach. Since substituents in position X have little hydrophobic effect, one might place highly polar functions on this ring to keep the overall log P value low. Within the limits of these functions and π-1,2_0 one could add lipophilic groups to the benzene ring to gain maximum lipophilic interaction at the receptor site.

24) C. Hansch and J.M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1973).