Journal of Medicinal Chemistry

VOLUME 7, NUMBER 6 NOVEMBER 6, 1964

The Use of Substituent Constants in the Analysis of the Structure–Activity Relationship in Penicillin Derivatives

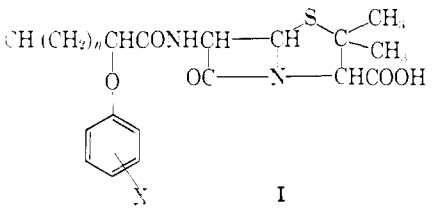
CORWIN HANSCH AND A. RUTH STEWARD

Department of Chemistry, Pomona College, Claremont, California

Received May 9, 1964

A substituent-constant analysis has been made correlating the relative activities of a series of 22 penicillin derivatives by means of two parameters σ and π. Correlations have been obtained for both *in vitro* and *in vivo* experiments with two strains of *Staphylococcus aureus*. From this analysis it is apparent that the primary effect of side-chain substituents on biological activity results from a modification of the lipophilic character of the penicillins. Electronic and negative steric effects appear to be of minor importance.

We have recently found[1,2] that by using two substituent constants σ and π the effect of substituents on the biological activity of a parent molecule can be rationalized. This paper is concerned with the extension of this approach to the excellent work of Gourevitch, Hunt, and Lein[3] on substituted penicillins. These investigators studied the activity of derivatives of I.



I

In I, X represents one or more substituents on the phenoxy ring as indicated in Table I. The side chain was varied from methyl (n = 0) to n-butyl (n = 3).

It is assumed that substituents X in I might modify the parent molecule in three primary ways: steric, electronic, and lipophilic. The latter two effects can be estimated by means of the Hammet[4] constant σ and our recently developed[1,2] constant π. σ is a measure of the electronic effect of X and π is defined as: $\pi = \log P_X - \log P_H$, where P_X is the partition coefficient between 1-octanol and water of the given derivative, and P_H is the partition coefficient of the parent compound. For analysis of the penicillins we have used values of π obtained from the phenoxyacetic acid.[1] We have assumed that the rate at which biologically active molecules make their way to the sites of action in a particular cell in a given organism is highly dependent on π and that the reaction at the site of action can be treated as pseudo first order from the kinetic point of view. With these assumptions and considering steric effects to be constant, we have developed[1,2] eq. 1. In

$$\log 1/C = -k\pi^2 + k'\pi + \rho\sigma + k'' \quad (1)$$

eq. 1, C represents the concentration of drug giving an equivalent biological response. In the present instance this is the CD_{50} with mice and the minimum inhibitory concentration with the *in vitro* tests on bacteria.

Equations 2–4 were generated from a least-squares fit of the data in Table I to eq. 1. The correlation coefficient is represented by r and s is the standard deviation. All of the points in Table I were used to derive the constants except two. Compounds **21** and **22** in Table I were omitted in deriving the numerical coefficients in these equations. The 2-chloro-4-phenyl derivatives was omitted because a precise value was not given for its activity. The acetamido derivative was omitted because it is quite susceptible to hydrolysis and was therefore not expected to give consistent results in the *in vivo* tests. The results of the regression analysis as highlighted by $\Delta \log 1/C$ in Table I are, on the whole, quite satisfying. Of the 22 derivatives, only the 4-acetamido is badly predicted, and this was anticipated. The correlation coefficient of 0.91 for all of the compounds except **21** and **22** is quite good considering the uncertainties involved in the biological testing and the necessary assumptions made in deriving eq. 1.

	r	s	
$\log 1/C = 0.053\pi^2 - 0.610\pi + 0.019\sigma + 5.751$	0.918	0.192	(2)
$\log 1/C = 0.055\pi^2 - 0.613\pi + 5.756$	0.918	0.187	(3)
$\log 1/C = -0.445\pi + 5.673$	0.909	0.191	(4)

Comparison of eq. 2 with eq. 3 in which the σ-term has been dropped shows the great advantage in the use of the substituent constants π and σ to separate the electronic effect of substituents. In this instance it is quite clear that the electronic effects of the groups attached to the phenoxy ring are not important except in so far as they affect the partition coefficient of the

(1) C. Hansch, R. Muir, T. Fujita, P. P. Maloney, F. Geiger, and M. Streich, *J. Am. Chem. Soc.*, **85**, 2817 (1963).
(2) C. Hansch and T. Fujita, *ibid.*, **86**, 1616 (1964).
(3) A. Gourevitch, G. A. Hunt, and J. Lein, *Antibiot. Chemotherapy*, **10**, 121 (1960).
(4) H. H. Jaffé, *Chem. Rev.*, **53**, 191 (1953).

TABLE I

EFFECTIVE CONCENTRATION FOR THE ACTIVITY OF PENICILLIN DERIVATIVES ON *Staphylococcus aureus* IN MICE

	Function	$\Sigma\sigma$	$\Sigma\pi$	log 1/C Obsd.	log 1/C Calcd.	Δ log 1/C
1.	H	0	0	5.86	5.67	0.19
2.	4-Cl	0.23	0.74	5.79	5.34	0.45
3.	4-OCH_3	−0.27	−0.04	5.69	5.69	0.00
4.	α-Et	0	0.50	5.54	5.45	0.09
5.	4-NO_2	0.78	0.06	5.53	5.65	0.12
6.	2-Cl	0.23	0.59	5.40	5.41	0.01
7.	3-CF_3	0.42	1.09	5.38	5.19	0.19
8.	2,5-Cl_2	0.60	1.35	5.24	5.07	0.17
9.	α-Pr	0	1.00	5.03	5.23	0.20
10.	3,5-$(CH_3)_2$	−0.14	1.02	5.03	5.22	0.19
11.	3-CF_3, 4-NO_2	1.26	1.15	5.03	5.16	0.13
12.	α-Bu	0	1.50	5.01	5.01	0.00
13.	2,4-Cl_2	0.46	1.33	4.97	5.08	0.11
14.	2,4-Br_2	0.46	1.77	4.87	4.88	0.01
15.	2,3,6-Cl_3	0.83	1.94	4.72	4.81	0.09
16.	4-Cyclohexyl	−0.15	2.52	4.70	4.55	0.15
17.	4-*t*-Bu	−0.20	1.71	4.67	4.91	0.24
18.	3,4,5-$(CH_3)_3$	−0.31	1.54	4.65	4.99	0.34
19.	4-*t*-Amyl, α-Et	−0.20	2.71	4.57	4.47	0.10
20.	Cl_5	1.43	3.44	4.25	4.14	0.11
21.	4-$NHCOCH_3$	−0.02	−0.79	4.82	6.02	1.20
22.	2-Cl, 4-C_6H_5	0.22	2.50	<4.37	4.56	>0.19

molecule in question. This would appear to mean that the function of the phenoxy side chain is that of affording lipotropic character to the rest of the molecule. It may also play an important role in sterically hindering hydrolysis of the lactam ring.

Another important consequence of the good correlation obtained with eq. 3 is that steric effects of substituents on the phenoxy ring are shown to be of minor importance, at least within the limits set by such substituents as phenyl, 4-*t*-amyl, and pentachloro. This also appears to be true for the alkyl groups (CH_3 through C_4H_9) on the side chain. Such a point is not always easy to differentiate from other substituent effects and again illustrates the usefulness of the ρ–σ–π analysis.[2]

As we pointed out earlier,[2] when the compounds in a given biologically active series have $\Sigma\pi$-values which are relatively far from the ideal value, one can expect a pseudo-linear relation between $\Sigma\pi$ and the logarithm of the biological response. A comparison of eq. 3 and eq. 4 shows that in this case we are operating rather far from the ideal $\Sigma\pi$-value (the π^2 term in eq. 3 is not statistically significant). Thus, eq. 4 can be taken as describing adequately the situation for these derivatives; it is with this equation that the results in Table I were obtained.

The most interesting aspect of eq. 4 is the negative coefficient associated with π. This would indicate the much more active derivatives could be obtained by using substituents in I which have negative π-values. Unfortunately only one such function was tested, the 4-acetamido (π for $-OCH_3$ is too small to be significant). Of the groups one might consider to be suitable for decreasing the lipotropic character of I, a function of low metabolic susceptibility, is CH_3SO_2– ($\pi \sim -0.5$).[1]

A decrease in lipophilic character could also be accomplished by omitting the aromatic ring and employing instead an aliphatic group with a highly branched side chain. Such a branched chain as $CH_3O(CH_2)_nC(CH_3)_2$ might have the added advantage of showing the hydrolysis of both the amide side chain and the lactam ring. It has been recognized for some time that one of the main weaknesses of the penicillins is their liability to inactivation through the opening of the lactam ring by the penicillinase which is produced by the bacteria. The side chain, $C_6H_5OCH(CH_3)$–, considered by Gourevitch, *et al.*,[3] would have a π-value of about 2.5. The above aliphatic chain with $n = 1$ would have a value of about 0.8 and a value of 1.3 with $n = 2$. Heterocyclic functions such as pyridine could also be used to advantage. It would seem well worthwhile to investigate these as well as other hydrophilic functions.

The derivatives in Table I were also tested in three other *in vitro* tests in addition to the *in vivo* mouse test. In one, the Smith strain of *Staphylococcus aureus* was used and in another the more resistant BLK strain of this organism was employed. Using the constants from Table I and the data in Table II we have obtained

TABLE II

MINIMUM INHIBITORY CONCENTRATIONS FOR *in Vitro* ACTION OF PENICILLINS AGAINST *Staphylococcus aureus*

		BLK strain log 1/C		Smith strain log 1/C		Smith strain with serum log 1/C	
	Function	Obsd.	Calcd.[a]	Obsd.	Calcd.[b]	Obsd.	Calcd.[c]
1.	H	4.77	4.78	6.66	6.54	6.36	6.48
2.	4-Cl	4.80	4.78	7.00	6.77	6.40	6.08
3.	4-OCH_3	4.20	4.71	6.09	6.46	6.39	6.51
4.	α-Et	5.08	4.74	6.67	6.67	6.07	6.22
5.	4-NO_2	5.41	4.96	7.01	6.76	6.71	6.41
6.	2-Cl	4.50	4.79	7.00	6.74	6.40	6.16
7.	3-CF_3	4.84	4.80	6.73	6.87	5.84	5.90
8.	2,5-Cl_2	4.54	4.82	7.03	6.94	6.15	5.77
9.	α-Pr	5.10	4.70	6.39	6.75	5.50	5.97
10.	3,5-$(CH_3)_2$	5.10	4.67	6.39	6.72	5.50	5.96
11.	3-CF_3, 4-NO_2	4.87	4.98	6.78	7.07	5.88	5.84
12.	α-Bu	4.21	4.65	6.71	6.79	5.81	5.73
13.	2,4-Cl_2	4.24	4.78	7.64	6.90	5.84	5.79
14.	2,4-Br_2	4.92	4.73	7.12	6.91	5.62	5.58
15.	2,3,6-Cl_3	5.18	4.80	6.47	7.00	4.97	5.49
16.	4-Cyclohexyl	4.25	4.49	6.75	6.71	5.25	5.28
17.	4-*t*-Bu	3.92	4.58	6.72	6.75	5.83	5.64
18.	3,4,5-$(CH_3)_3$	5.11	4.57	6.71	6.72	5.81	5.72
19.	4-*t*-Amyl, α-Et	4.86	4.45	6.75	6.67	5.55	5.21
20.	Cl_5	4.63	4.73	6.83	6.92	4.73	4.85
21.	4-$NHCOCH_3$	4.53	4.81	6.42	6.26	6.42	6.91
22.	2-Cl, 4-C_6H_5	4.58	4.58	7.07	6.81	5.28	5.28

[a] These values were obtained using eq. 7. [b] Obtained with eq. 5. [c] Obtained with eq. 9.

eq. 5 and 6 for the Smith strain and eq. 7 and 8 for the results with the BLK strain. In eq. 5–8, C is the minimum inhibitory concentration for the *in vitro* tests using a heart infusion broth medium. The correlations with these equations are extremely poor (see Table III), and about the only conclusions one can draw are that π is of little consequence and that a positive value for σ might be of some importance.

	r	s	
$\log 1/C = -0.083\pi^2 + 0.292\pi + 0.250\sigma + 6.543$	0.439	0.314	(5)
$\log 1/C = 0.230\sigma + 6.724$	0.347	0.310	(6)
$\log 1/C = -0.017\pi^2 - 0.058\pi + 0.245\sigma + 4.777$	0.344	0.400	(7)
$\log 1/C = 0.189\sigma + 4.669$	0.233	0.392	(8)

TABLE III

COMPARISON OF OBSERVED LOG (C'/C) WITH VALUES OBTAINED FROM EQUATION 19

	Function	log C'/C Obsd.	log C'/C Calcd.	$\Delta\Delta$ log $1/C$
1.	H	−0.30	−0.07	0.23
2.	4-Cl	−0.60	−0.69	0.09
3.	4-OCH_3	+0.30	+0.05	0.25
4.	α-Et	−0.60	−0.45	0.15
5.	4-NO_2	−0.30	−0.34	0.04
6.	2-Cl	−0.60	−0.58	0.02
7.	3-CF_3	−0.89	−0.96	0.07
8.	2,5-Cl_2	−0.89	−1.16	0.27
9.	α-Pr	−0.89	−0.79	0.10
10.	3,5-$(CH_3)_2$	−0.89	−0.76	0.13
11.	α-Bu	−0.89	−1.07	0.18
12.	3-CF_3, 4-NO_2	−0.89	−1.23	0.34
13.	2,4-Cl_2	−1.80	−1.11	0.69
14.	2,4-Br_2	−1.49	−1.33	0.16
15.	*t*-Bu	−0.89	−1.11	0.22
16.	Cyclohexyl	−1.49	−1.43	0.06
17.	3,4,5-$(CH_3)_3$	−0.89	−1.00	0.11
18.	2,3,6-Cl_3	−1.49	−1.51	0.02
19.	4-*t*-Amyl, α-Et	−1.19	−1.47	0.28
20.	4-C_6H_5, 2-Cl	−1.80	−1.53	0.27
21.	Cl_5	−2.10	−2.07	0.03

In the third *in vitro* experiment, the heart infusion broth was mixed with equal parts of a filtered sterilized human serum. The results we find for the derivatives in Table II are dramatically different from those obtained without serum.

$$\log 1/C = 0.023\pi^2 - 0.534\pi - 0.042\sigma + 6.476 \quad r = 0.859 \quad s = 0.280 \quad (9)$$

$$\log 1/C = 0.019\pi^2 - 0.526\pi + 6.466 \quad 0.858 \quad 0.273 \quad (10)$$

$$\log 1/C = -0.468\pi + 6.437 \quad 0.857 \quad 0.267 \quad (11)$$

Equations 9–11 come from the work using the Smith strain and serum. Here the correlation coefficient is reasonable although not as good as for the *in vivo* experiments. The coefficients in eq. 9–11 are rather close to those in eq. 2–4 and thus indicate parallel mechanisms of action for the *in vivo* and *in vitro* experiments.

The role of the serum raises immediate interest and we have attempted to evaluate this from the substituent constant point of view.

It is apparent from the date of Gourevitch, *et al.*,[3] that a higher concentration of the penicillins was almost always needed to bring about the minimum inhibitory effect in the presence of serum. As they pointed out, this would indicate that the penicillins were adsorbed onto serum protein which thus reduced the effective concentration. It is well known[5] that serum albumin has a great affinity for molecules and ions of all sorts. If this is true, such adsorption should be susceptible to substituent-constant analysis. The equilibrium between adsorbed and unadsorbed penicillin can be formulated as in eq. 12. The equilibrium constant for the

$$\underset{\text{Pe}}{\text{penicillin}} + \underset{\text{SP}}{\text{serum protein}} \rightleftharpoons \underset{\text{Pe-SP}}{\text{penicillin-serum protein}} \quad (12)$$

adsorption of a given penicillin, X, on serum protein could be expressed as in eq. 13. Assuming the sites

$$K_X = [\text{Pe-SP}]/[\text{Pe}][\text{SP}] \quad (13)$$

for the adsorption of penicillin on serum protein to be in large excess over the number of penicillin molecules, [SP] can be assumed constant and eq. 13 can be written as

$$K_X' = [\text{Pe-SP}]/[\text{Pe}] \quad (14)$$

For the case with serum, let $C = [\text{Pe–SP}] + [\text{Pe}] \cong [\text{Pe–SP}]$. The justification for assuming that $C \cong [\text{Pe–SP}]$ is that eq. 9–11 give high correlations while eq. 5 and 6 do not.

For the case without serum, let $C' = [\text{Pe}']$. If we represent equivalent biological response by R, then

$$R = R_{\text{Pe}} + R_{\text{Pe-SP}} = R_{\text{Pe}'} \quad (15)$$

Assuming that response is proportional to concentration, we can define response in each of the systems as follows.

$$\begin{aligned} R_{\text{Pe}'} &= a[\text{Pe}'] = aC' \\ R_{\text{Pe}} &= a[\text{Pe}] \\ R_{\text{Pe-SP}} &= b[\text{Pe-SP}] \cong bC \end{aligned} \quad (16)$$

Substitution of eq. 16 into 15 yields

$$\begin{aligned} a[\text{Pe}] + b[\text{Pe-SP}] &= a[\text{Pe}'] \\ a[\text{Pe}] + bC &= aC' \\ [\text{Pe}] = \frac{aC' - bC}{a} &= C' - \frac{bC}{a} \end{aligned} \quad (17)$$

Taking the logarithm of eq. 17 and 14 and substituting log [Pe] from 17 into 14 gives rise to eq. 18 (assuming $C \cong [\text{Pe–SP}]$).

$$\log K_X' = \log C - \log\left(C' - \frac{bC}{a}\right) = \log \frac{C}{C' - \frac{bC}{a}} \quad (18)$$

$$\log \frac{1}{K_X'} = \log\left(\frac{C'}{C} - \frac{b}{a}\right) \text{ or } \log\left(\frac{1}{K_X'} + \frac{b}{a}\right) = \log \frac{C'}{C}$$

Subtracting eq. 5 from eq. 9 we obtain

$$\log (C'/C) = 0.106\pi^2 - 0.826\pi - 0.293\sigma - 0.067 \quad r = 0.911 \quad (19)$$

Equating expressions 19 and 18 gives eq. 20. In eq.

$$\log\left(\frac{1}{K_X'} + \frac{b}{a}\right) = 0.106\pi^2 - 0.826\pi - 0.293\sigma - 0.067 \quad (20)$$

$$-\log\left(\frac{1}{K_X'} + \frac{b}{a}\right) = -0.106\pi^2 + 0.826\pi + 0.293\sigma + 0.067$$

20, b/a can be regarded as the ratio of two activity coefficients and K_X' as a binding constant. The correlation of 0.911 is better than that obtained with eq. 9 (0.859). This is the best justification for the validity of subtracting eq. 5 from eq. 9. The very poor rationalization of the variance in the data by eq. 5 ($r^2 = 0.2$) causes one to hesitate in using this equation as a base from which to estimate the effect of substituents on K_X'. The good correlation coefficient appears to justify the effort. The importance of the various terms in eq. 19 or 20 are illustrated in eq. 21–23. An F

$$\log (C'/C) = -0.505\pi - 0.244\sigma - 0.233 \quad r = 0.891 \quad (21)$$

$$= 0.078\pi^2 - 0.774\pi - 0.140 \quad 0.879 \quad (22)$$

$$= -0.535\pi - 0.258 \quad 0.868 \quad (23)$$

test shows the π^2-term in 19 to be significant at the 0.90 level and the σ-term to be significant at the 0.95 level. In comparing eq. 23 and 19, an F test indicates the sum of the contributions of the π^2- and σ-terms to be significant at the 0.99 level. The observed and calculated values for log (C'/C) using eq. 19 are shown in Table III.

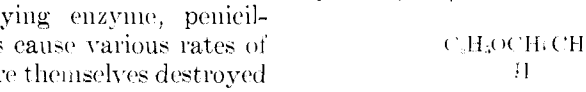
(5) J. T. Edsall, *Advan. Protein Chem.*, **3**, 463 (1947).

The problem of using *in vitro* tests on penicillin derivatives to evaluate them for *in vivo* use is quite complicated because the penicillins induce the bacteria to produce a penicillin-destroying enzyme, penicillinase. The various derivatives cause various rates of production of this enzyme and are themselves destroyed at different rates, depending on their structure.[6,7] To further complicate the picture, bacteria also produce proteinases capable of destroying the activity of penicillinase.[8]

The poor correlations obtained with the *in vitro* experiments as shown by eq. 5-8 are, in part, surprising. In the case of the BLK strain which produces penicillinase, the poor correlation is probably the result of metabolism by this enzyme, the rate of which might be quite dependent on steric factors. The problem with the Smith strain is different since this nonresistant organism does not produce penicillinase. In this example part of the poor correlation is due to the small amount of variance in the biological response to be explained by the regression analysis ($S^2_{\log 1/C} = 0.104$, while for the *in vivo* experiments $S^2_{\log 1/C} = 0.198$). However, this would not account for all of the unexplained variance. This unknown quantity must be ascribed to unaccountable steric or metabolic factors.

Gourevitch, *et al.*,[3] made the observation that the linear correlation between log (C'/C) and the molecular weight of the side chain was 0.644. Although this is quite significant, it does not "explain" most of the variance in the data ($r^2 = 0.42$). It would appear that in so far as the molecular weight of substituents is proportional to lipophilic character, the approach of Gourevitch, *et al.*, is valid. However, a very great improvement in "explaining" the variance comes about with the use of eq. 20. For this correlation $r^2 = 0.83$. In the former case 58% of the variance is unaccounted for, while in the latter case only 17% is unexplained.

A point of particular interest is the fact[3] that when the side chain attached to the amino group in I is α-phenoxyethyl (II) instead of phenoxymethyl (III),



II III

one obtains a more active derivative with II than with III. This would seem to be at odds with the fact that II has a higher π-value than III and should therefore be less active according to eq. 4 or 11. The fact that the activities for α-Et, α-Pr, and α-Bu are nicely accounted for by eq. 4 and 11 makes this substituent effect one of considerable importance.

Since this positive effect cannot be rationalized by considering its electronic contribution (σ is of very little consequence even for such groups as NO_2 and CF_3), its lipophilic character (π), or its fit to the site of action, one is left with little to consider except metabolic effects. Good evidence[9] has been obtained to show that branched side chains are much more resistant to metabolic action than normal ones. Thus, in the case of side chain II, the expected decrease in activity due to the added lipophilic character of the methyl group is apparently more than offset by the increased metabolic stability of the penicillins with branched chains.

The close correspondence of eq. 2-4 with eq. 9-11 indicates that the *in vitro* test with serum is indeed a good model for predicting activity of penicillin in mice. The fact that human serum was actually used indicates the similarity between the binding power of serum protein in mice and humans for penicillins.

Acknowledgment.—This work was supported under Research Grant GM-07492-04 from the National Institutes of Health.

(6) R. P. Novick, *Biochem. J.*, **83**, 229 (1962).
(7) J. M. T. Hamilton-Miller, *ibid.*, **87**, 209 (1963).
(8) M. Kogut, M. R. Pollick, and E. J. Tridgell, *ibid.*, **62**, 391 (1956).
(9) S. H. Sharman, *Nature*, **201**, 704 (1964).