(40) Y. K. Syrkin and M. E. Dyatkina, "Structure of Molecules and the Chemical Bond", Dover Publications, New York, N.Y., 1964.
(41) C. Hansch and C. Silipo, *J. Med. Chem.*, 17, 661 (1974).
(42) L. B. Kier, L. H. Hall, W. J. Murray, and M. Randic, *J. Pharm. Sci.*, 64, 1971 (1975).
(43) L. H. Hall, L. B. Kier, and W. J. Murray, *J. Pharm. Sci.*, 64, 1974 (1975).
(44) W. J. Murray, L. H. Hall, and L. B. Kier, *J. Pharm. Sci.*, 64, 1978 (1975).
(45) L. B. Kier, W. J. Murray, and L. H. Hall, *J. Med. Chem.*, 18, 1272 (1975).

Quantitative Structure–Activity Relationships among Steroids. Investigations of the Use of Steric Parameters

Robert A. Coburn and Alan J. Solo*

Department of Medicinal Chemistry, School of Pharmacy, State University of New York at Buffalo, Buffalo, New York 14214.
Received May 27, 1975

The importance of steric factors in quantitative structure–activity relationships involving steroid hormones is discussed. A variety of steric parameters, such as parachlor, molecular volume, van der Waals volume, and including difference and squared steric terms, is explored in an attempt to find preferred forms for such expressions. Improved correlations for 6-substituted 16-methylene-17α-acetoxy-4,6-pregnadiene-3,20-dione derivatives were found in which activity is related to π and a squared or difference steric factor. The activity of 9α-substituted cortisols correlates well with σ_I and a simple steric factor, provided that the 9α-hydroxylated compound is excluded from the series.

The use of extrathermodynamic linear free-energy relationships in the correlation of biological data from in vivo systems has resulted in growing experimental support.[1] The modification of this method by inclusion of physicochemical, theoretical (quantum chemical), and dummy parameters not derived from linear free-energy relationships represents a widespread stochastic approach to quantitative structure–activity relationships (QSAR).[2]

Relatively few QSAR studies have been reported for steroids. James has correlated the lipophilicity of testosterone and 19-nortestosterone esters with the prolongation of their biological effects.[3-5] However, if one excludes the, thus far, unpublished study mentioned by Ostrenga,[6] the only reports of quantitative correlations of steroid structure with activity are those of Wolff and Hansch on 9α-substituted cortisol derivatives[7] and on 6-substituted 16-methylene-17α-hydroxy-4,6-pregnadiene-3,20-dione acetate derivatives.[8] While these studies suggest that the multiparameter regression technique is of value in the study of steroids, the results to data have been less satisfying than applications in other areas. We discuss below previously ignored factors relating to the steric influence of substituents in QSAR studies and report methods leading to improved QSAR for steroids.

Following the submission of this article for review, a report by Topliss and Shapiro[9] appeared in which structure–activity relationships of 6-substituted 16-methylene-17α-hydroxy-4,6-pregnadiene-3,20-dione acetates were reappraised. In that report improved correlations were obtained by inclusion of a term involving the circumference of the 6-substituent. Although no evidence was offered relating circumference to a linear free-energy steric term, the finding suggests the importance of a steric factor in this correlation. In this investigation we confirm the importance of such a factor by obtaining further improvement in correlations employing a variety of more conventional steric terms. This study was then extended to a second group of compounds. This investigation was undertaken to contribute to an understanding of factors determining the type and optimal form of the steric term to be used in QSAR.

In view of the enormous number of steroid analogues which are known, the paucity of QSAR reported seems surprising. Part of this problem arises from the fact that relatively few large series of steroids, in which only a single substituent is systematically varied, have been prepared and assayed. A further complication arises from the variability of much of the in vivo assay data. For the steroid hormones an important additional complication stems from the interactions with a number of high-affinity relatively specific receptor sites including the hormonal receptors, the active sites of steroid metabolizing enzymes, and carrier proteins in the blood. In addition to these specific receptors, which all have limited capacity but high (and frequently similar) binding affinities, there are nonspecific binding sites which bind large quantities of steroid less tightly. While any of these factors may influence an assay, the classical in vivo assays which involve multiple dosing over a week or longer seem most likely to be influenced by factors which affect transport, rate of metabolism, and interaction with the hormonal receptor(s). This has been generally recognized in the past and was used in a qualitative sense in the design of steroid hormone analogues. To the extent that assays are affected by interaction of the test substance with secondary receptor sites, such as metabolizing enzymes, QSAR must reflect an optimization of transport and a balancing of those factors which maximize interaction with the primary receptor against those which minimize the interaction with the secondary receptors.

Competing processes, differently influenced by variation of substituent properties, would not in themselves be expected to lead to nonlinearity among free-energy correlations. Although the hydrophobic Hansch constant π is a linear free-energy parameter, biological activity is often better represented as a quadratic function of π or log P.[10] In dealing with steric interactions generally two types of parameters have been considered: Taft's constant E_s, or Hancock's modification E_s^c, and the physicochemical constant, molar refractivity (MR). It has been rationalized that either intra- or intermolecular steric interactions may have to be examined.[11] Since, as Hansch has cautioned,[12] the equivalence of molar refractivity with steric requirements can be misleading, we have chosen to investigate several other parameters which may be more directly related to substituent steric influence. [Most substituents found in this study contain π bonds or nonbonded electrons suggesting that polarizability contributes substan-

tially to their MR values. The correlation between MR and MV for these substituents (Table II) is lower than that reported for other collections of substituents (0.91).[12] These are parachor (P),[13-15] Exner's molecular volume (MV),[16] and van der Waals volume (V).[17] For substituents found in this study several parameter values were estimated in order to complete the data set. The use of this variety of related steric parameters served as a check of the validity of both the estimated substituent values and the observed correlations.

The derivation of a linear free-energy steric parameter representing the relationship of substituent geometry to the free energy of intermolecular interactions is obstructed by a lack of knowledge concerning critical variables (orientation and equilibrium intergroup distances in the drug–receptor complex, conformational energy functions in both drug and receptor, etc.) For this reason it would appear more immediately profitable to investigate the use of these available parameters even though they may be only indirectly related to the steric free-energy contribution of the substituents. The use of van der Waals group radii would introduce orientational uncertainties for nonspherically symmetric groups and was therefore examined only briefly, vide infra. Nonlinear relationships involving steric parameters may result from any of a number of causes. Many of the steric parameters being considered are not derived from linear free-energy relationships, and the detailed nature of the steric interactions is unknown. Differences in limitations of critical bulk tolerance among one or more acceptors involved in competing processes may lead to discontinuities in a series of derivatives with respect to their correlation within those interactions. This may then result in an apparent optimal parameter value for the summary effect. Therefore, the above parameters were also examined in quadratic functions (eq i) and expressions relating activity to the absolute magnitude of the difference between substituent steric constant and an ideal value, the latter determined by an iterative empirical method (eq ii). A square term alone (eq iii) was conveniently included and provides a more steeply rising function which might more closely approximate the exponential form of the van der Waals repulsive interaction.

$$\log A = aX^2 + bX + c \quad \text{(i)}$$

$$\log A = a|X_0 - X| + b \quad \text{(ii)}$$

$$\log A = aX^2 + b \quad \text{(iii)}$$

Shapiro et al. reported structure–activity relationships of 6-substituted 16-methylene-17α-acetoxy-4,6-pregnadiene-3,20-dione derivatives.[18] While this report did not include a quantitative relationship, the authors noted that, for undetermined reasons, progestational activity seemed strongly dependent on the size of the 6-substituent. Activity was optimized if the volume of the substituent was approximately that of a methyl group. Furthermore, the critical relationship was the volume of the substituent rather than its bond length. Compounds containing –CN or –CHO at C-6 seemed less active than expected on the basis of the bulk of the substituents. Cross conjugation was offered as a possible explanation for the latter effect.

Wolff and Hansch reexamined Shapiro's data and reported QSAR for these compounds.[8] The most important parameter proved to be π. In contrast to Shapiro's results, Wolff and Hansch found that neither steric nor electronic effects alone could account for as much as 20% of the variance. Equations iv and v were reported to give the best correlations. The results were interpreted as showing no

$$\log A = 0.26\ (\pm 1.01) - 0.07\ (\pm 0.07)\ \mathrm{MR} + 0.97\ (\pm 0.53)\ \pi + 2.84\ (\pm 2.50)\ \mathfrak{F} \quad \text{(iv)}$$

$$n = 13,\ r = 0.892,\ s = 0.489,\ F = 11.6$$

$$\log A = -0.17\ (\pm 0.88) + 1.06\ (\pm 0.52)\ \pi + 3.46\ (\pm 2.68)\ \mathfrak{F} - 1.05\ (\pm 1.11)\ \pi^2 \quad \text{(v)}$$

$$n = 13,\ r = 0.890,\ s = 0.492,\ F = 11.4$$

significant resonance contribution, a minor or no significant steric contribution, an inductive effect which probably correlates with relative affinity for the cytoplasmic hormonal receptor, and a major contribution from π which reflects changes in drug transport and distribution. The large relative uncertainties in the reported regression coefficients and the unexplained deletion of the most active compound in the series, the methyl derivative, seriously detract from these QSAR.

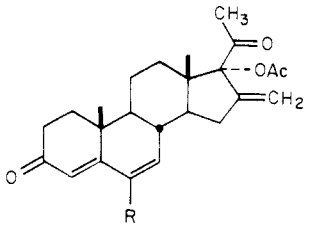
Cooke and Vallance[19] reported that introduction of 6-methyl-6-dehydro substitution into 17α-acetoxyprogesterone resulted in much slower metabolism of the compound by rabbit liver. Recently, Terenius[20] reported that although 6-methyl-6-dehydro-17α-acetoxyprogesterone and 6-chloro-6-dehydro-17α-acetoxyprogesterone are both reported to have about 50 times the progestational activity of progesterone, they have the same affinity as progesterone for the progestational receptor of rabbit uterus. From these data, Terenius speculates that the high in vivo activity of these two compounds derives largely from their resistance to metabolic breakdown. These studies[19,20] imply that the enhanced activities reported by Shapiro for a number of compounds (especially the chloro and methyl derivatives) may also be due largely to their metabolic resistance. Therefore, we decided to reinvestigate these compounds and specifically to reassess the steric influence of substituents through the use of steric parameters in both linear and nonlinear relationships.

Table I contains the variables employed in this study. In addition to those, the squared terms π^2, σ_m^2, MV^2, MR^2, P^2, V^2, and E_s^2 and the difference terms $|V_0 - V_x|$, $(V_0 - V_x)^2$, $|\mathrm{MV}_0 - \mathrm{MV}_x|$, $(\mathrm{MV}_0 - \mathrm{MV}_x)^2$, $|\mathrm{MR}_0 - \mathrm{MR}_x|$, $(\mathrm{MR}_0 - \mathrm{MR}_x)^2$, and ΔV were also employed. [Difference terms represent the difference between the value of a standard substituent (designated subscript zero), identified in reported equations, and the values of the other substituents. The difference term ΔV is defined as $|V_0 - V_x|$ where the value of V_0 (12.2 ± 0.1) was determined from the optimal regression involving π and ΔV in which V_0 was varied from 5.8 to 13.4 in increments of 0.1.] The 357 regression equations involving one (21), two (98), three (233), and four (5) independent variables were examined. (An all-possible-regressions method was adopted via program REGRES3, a driver and compilation routine written for this purpose, which employs the IBM SSP subroutines CORRE, MINV, ORDER, and MULTR for the statistical calculations.) All 14 compounds in Table I were included in each regression.

As Topliss and Costello have pointed out,[21] the possibility of a chance correlation in multiple regression analyses increases with the number of variables examined. However, only by the study of a number of steric parameters in many QSAR's can guidelines be determined for the selection of the optimal parameter in a given situation. A study which reports the use of a single arbitrary steric parameter does not contribute to this understanding.

Since the correlation between independent variables is an important consideration in the selection of variable combinations, these correlation coefficients, r, are given in Table II. Although the use of both π and a parameter

Table I. Progestational Activity and Substituent Constants



R	Log A^a	$\mathcal{F}^a$	$E_s{}^b$	π^a	MR^a	P	V	MV	σ_I	σ_m
CH_3	1.96	−0.04	0.00	0.50	5.65	56.5	13.67	31.48	−0.05	−0.07
Cl	1.89	0.41	0.27	0.71	6.03	56.0	11.65	22.96	0.47	0.37
F	1.74	0.43	0.78	−0.14	0.92	26.5	5.80	15.11	0.52	0.34
Br	1.62	0.44	0.08	0.86	8.88	69.5	14.50	26.19	0.45	0.39
N_3	1.30	0.30	0.63	0.46	10.20	95.0	20.80	35.00	0.42	0.27
OCH_3	1.15	0.26	0.69	−0.02	7.87	76.2	18.87	38.52	0.25	0.12
SCN	1.08	0.36	0.17	0.41	13.40	111.5	25.50	33.50	0.55	0.41
CF_3	1.04	0.38	−1.16	0.88	5.02	86.5	23.21	32.11	0.42	0.43
CN	0.78	0.51	0.21	−0.57	6.33	63.5	14.70	22.67	0.58	0.56
OEt	0.00	0.22	0.62	0.38	12.47	115.9	29.10	54.80	0.25	0.10
H	0.00	0.00	1.24	0.00	1.03	16.8	3.40	14.90	0.00	0.00
CHO	0.00	0.31	−0.67^c	−0.65	6.88	62.5	15.70	25.08	0.30	0.35
OAc	−0.70	0.41	0.62	−0.64	12.47	120.1	29.07	48.43	0.39	0.39
NHAc	−1.00	0.28	−0.58^c	−0.97	14.93	129.5	33.70	48.99	0.28	0.21

[a] Reference 8. [b] Reference 23 and references cited therein. [c] Coplanar conformation assumed.

Table II. Correlation Coefficient Matrix[a]

Const	$\mathcal{F}$	E_s	π	MR	P	V	MV	σ_I	σ_m
$\mathcal{F}$	1.000								
E_s	−0.214	1.000							
π	−0.058	0.020	1.000						
MR	0.179	−0.193	−0.247	1.000					
P	0.220	−0.309	−0.206	0.949	1.000				
V	0.172	−0.375	−0.257	0.908	0.988	1.000			
MV	−0.049	−0.135	−0.215	0.843	0.905	0.921	1.000		
σ_I	0.954	−0.146	0.058	0.186	0.215	0.149	−0.102	1.000	
σ_m	0.939	−0.340	−0.067	0.116	0.168	0.123	−0.171	0.923	1.000

[a] For substituents shown in Table I.

such as V, MV, or P might be considered suspect due to the generally high correlation (r = 0.86–0.88) between lipophilic and steric constants,[22] it can be seen in Table II that for this selection of substituents no such correlation exists.

Among the simple single parameter equations, the variables resulting in the highest correlations were π, MR, P, MV, and V (eq 1–5). Taft's steric constant E_s or $E_s{}^2$ and the electronic constants $\mathcal{F}$, σ_I, σ_m, and $\sigma_m{}^2$ as well as π^2 showed little correlation. A number of single parameter equations involving difference or squared steric terms (eq 6–14) gave higher correlations than the corresponding simple terms. Equations 13 and 14 which involve ΔV and ΔV^2, respectively, gave as good a correlation as that obtained with π (eq 1).

The addition of π^2 or of the electronic terms $\mathcal{F}$, σ_I, σ_m, or $\sigma_m{}^2$ to the π constant in two parameter equations produced no substantial improvement (r = 0.778, 0.765, 0.769, 0.763, and 0.775, respectively). Two parameter equations involving a steric constant and its square gave better correlations than did the corresponding single parameter equations involving either term. However, such quadratic equations proved inferior to two parameter equations involving π and the steric constants E_s, MR, P, MV, and V (r = 0.759, 0.809, 0.830, 0.843, and 0.841, respectively). Again, poorest results were obtained when E_s was employed and better results when P, MV, or V were used. The best correlations involving two parameter equations resulted from combinations of π with difference or squared steric terms (eq 15–24). With the exception of eq 21, which involved a difference term based on the molecular volume of a methyl group, all equations involving difference or squared steric terms represented improvements over the corresponding simple steric terms.

Equation 25 employs the Hansch constant in linear form and the radius (r_s) of a sphere with a van der Waals volume equivalent to that of the substituent, in exponential form. This regression appears to offer little improvement over the simpler eq 17–19 despite the inclusion of an additional parameter.

A number of three parameter equations involving π, π^2, and a steric term or π, an electronic parameter, and a steric term were found to have r values between 0.80 and 0.87. Their F-statistic values ranged from 5.98 to 10.51 and in all cases at least one coefficient failed to meet the t test for significance (α = 0.05). Similar equations involving at least one difference or squared steric term gave better correlations. The four best three parameter equations (eq 26–29) all gave only marginally better correlations than the best two parameter equations and have at least one regression coefficient which fails to satisfy the t test. Similarly, the four parameter equations studied showed slightly better correlations than related two and three parameter equations, but the improvement seemed too little to justify inclusion of the extra terms.

In comparing the results of these studies with the results reported by Wolff and Hansch[8] on the 13 compound subset (which excludes the methyl derivative), one must note that with the exception of the regression equation employing the single parameter π, all of the previously reported

	n	r	s	F	
$\log A = 1.20\ (\pm 0.30)\ \pi + 0.65$ (4.00)	14	0.756	0.654	16.02	(1)
$\log A = -0.10\ (\pm 0.057)\ \mathrm{MR} + 1.60$ (−1.82)	14	0.466	0.884	3.33	(2)
$\log A = -0.014\ (\pm 0.007)\ P + 1.84$ (−1.95)	14	0.491	0.871	3.81	(3)
$\log A = -0.041\ (\pm 0.019)\ \mathrm{MV} + 2.10$ (−2.15)	14	0.527	0.850	4.62	(4)
$\log A = -0.059\ (\pm 0.026)\ V + 1.88$ (−2.28)	14	0.551	0.835	5.22	(5)
$\log A = -0.0077\ (\pm 0.0032)\ \mathrm{MR}^2 + 1.40$ (−2.37)	14	0.565	0.825	5.61	(6)
$\log A = -0.00066\ (\pm 0.00026)\ \mathrm{MV}^2 + 1.55$ (−2.54)	14	0.592	0.806	6.46	(7)
$\log A = -0.0019\ (\pm 0.00061)\ V^2 + 1.56$ (−3.09)	14	0.666	0.746	9.55	(8)
$\log A = -0.19\ (\pm 0.073)\ \lvert\mathrm{MR_{Me}} - \mathrm{MR}\rvert + 1.49$ (−2.60)	14	0.600	0.800	6.77	(9)
$\log A = -0.023\ (\pm 0.0079)\ (\mathrm{MR_{Me}} - \mathrm{MR})^2 + 1.29$ (−2.84)	14	0.633	0.773	8.04	(10)
$\log A = -0.081\ (\pm 0.030)\ \lvert\mathrm{MV_{Me}} - \mathrm{MV}\rvert + 1.54$ (−2.74)	14	0.620	0.785	7.50	(11)
$\log A = -0.042\ (\pm 0.019)\ \lvert\mathrm{MV_F} - \mathrm{MV}\rvert + 1.49$ (−2.17)	14	0.531	0.847	4.72	(12)
$\log A = -0.11\ (\pm 0.028)\ \Delta V + 1.72$ (−3.91)	14	0.748	0.663	15.28	(13)
$\log A = -0.0053\ (\pm 0.0013)\ \Delta V^2 + 1.37$ (−4.13)	14	0.766	0.642	17.08	(14)
$\log A = 1.22\ (\pm 0.28)\ \pi - 0.59\ (\pm 0.36)\ E_s^2 + 0.90$ (4.33) (−1.65)	14	0.810	0.612	10.53	(15)
$\log A = 1.02\ (\pm 0.28)\ \pi - 0.0049\ (\pm 0.0024)\ \mathrm{MR}^2 + 1.06$ (3.59) (−2.00)	14	0.828	0.585	12.02	(16)
$\log A = 1.01\ (\pm 0.26)\ \pi - 0.00008\ (\pm 0.00003)\ P^2 + 1.21$ (3.89) (−2.59)	14	0.857	0.539	15.17	(17)
$\log A = 1.03\ (\pm 0.25)\ \pi - 0.00048\ (\pm 0.00018)\ \mathrm{MV}^2 + 1.22$ (4.06) (−2.67)	14	0.860	0.533	15.64	(18)
$\log A = 0.95\ (\pm 0.25)\ \pi - 0.0013\ (\pm 0.00044)\ V^2 + 1.21$ (3.97) (−2.92)	14	0.871	0.513	17.33	(19)
$\log A = 0.99\ (\pm 0.28)\ \pi - 0.12\ (\pm 0.056)\ \lvert\mathrm{MR_{Me}} - \mathrm{MR}\rvert + 1.13$ (3.60) (−2.24)	14	0.840	0.566	13.22	(20)
$\log A = 0.95\ (\pm 0.31)\ \pi - 0.045\ (\pm 0.026)\ \lvert\mathrm{MV_{Me}} - \mathrm{MV}\rvert + 1.10$ (3.04) (−1.76)	14	0.816	0.603	10.97	(21)
$\log A = 1.07\ (\pm 0.26)\ \pi - 0.030\ (\pm 0.013)\ \lvert\mathrm{MV_F} - \mathrm{MV}\rvert + 1.18$ (4.07) (−2.33)	14	0.845	0.559	13.70	(22)
$\log A = 0.89\ (\pm 0.21)\ \pi - 0.080\ (\pm 0.020)\ \Delta V + 1.37$ (4.17) (−4.08)	14	0.911	0.431	26.76	(23)
$\log A = 0.83\ (\pm 0.23)\ \pi - 0.0037\ (\pm 0.0010)\ \Delta V^2 + 1.11$ (3.56) (−3.68)	14	0.899	0.458	23.11	(24)
$\log A = 0.88\ (\pm 0.25)\ \pi - 0.000007\ (\pm 0.000002)\ e^{5.0(\pm 0.5)r_s}$ (3.47) (−3.04)	14	0.876	0.504	18.16	(25)
$\log A = 0.89\ (\pm 0.23)\ \pi + 0.22\ (\pm 0.70)\ \sigma_m - 0.079\ (\pm 0.021)\ \Delta V + 1.30$ (4.00) (0.31) (−3.84)	14	0.912	0.450	16.41	(26)

	n	r	s	F	
$\log A = 0.90\ (\pm 0.22)\ \pi - 0.079\ (\pm 0.020)\ \Delta V + 0.54\ (\pm 0.77)\ \mathfrak{F} + 1.20$ (4.11) (−3.93) (0.69)	14	0.915	0.442	17.16	(27)
$\log A = 0.89\ (\pm 0.22)\ \pi - 0.31\ (\pm 0.41)\ \pi^2 - 0.077\ (\pm 0.020)\ \Delta V + 1.45$ (4.10) (−0.75) (−3.80)	14	0.916	0.440	17.32	(28)
$\log A = 0.87\ (\pm 0.21)\ \pi - 0.080\ (\pm 0.019)\ \Delta V + 0.71\ (\pm 0.62)\ \sigma_I + 1.12$ (4.16) (−4.13) (1.14)	14	0.921	0.426	18.75	(29)

equations[8] involve one or more parameters displaying regression coefficients whose t statistic fails to show significance at the 95% confidence level. In examining all possible one, two, and three parameter regressions among this data subset with π, π^2, E_s, MR, P, V, MV, $\mathfrak{F}$, σ_I, and σ_m, several improved equations were discovered. The best two parameter equation involved π and MV (eq 1a). Other two parameter equations showing improved correlation for this subset involved π and σ_I, σ_m, or $\mathfrak{F}$, π and V, or P. However, all of the latter equations contained at least one regression coefficient which failed the t test at the 95% confidence level. Among the three parameter equations, six equations involving combinations of π with an electronic term and with a steric term (eq 2a–7a) gave higher correlations than previously reported and passed the t test, at the 95% confidence level, for all correlation coefficients.

In examining the regression data for the full set of 14 compounds, we noted that for our better equations, the compound which most frequently showed the poorest correlation was not that in which the C-6 substituent was methyl[8] but rather was the one in which it was hydrogen. Therefore we decided to investigate the regression equations for the 13 compound subset which excludes the C-6 unsubstituted compound. For this subset, π gives a correlation coefficient of 0.766 compared to 0.751 for the subset lacking the methyl derivative[6] and to 0.756 for the full set of compounds. However, in contrast to the full set or to the previous subset, for this subset π did not give the single parameter regression equation with the best correlation coefficient. For this subset $r = 0.784$ and 0.816 for the single parameter equations containing V and V^2, respectively. For this subset, two parameter equations which gave improved correlations and whose correlation coefficients all passed the t test at the 95% confidence level included eq 1b–6b. Inclusion of a third parameter gave only the marginal improvement shown in eq 7b and 8b.

It is apparent from our results that the choice of MR as the sole parameter investigated for steric influence by Wolff and Hansch[8] was unfortunate since these results show that V and MV consistently give a better correlation. It is only for the 13 compound subset, lacking the methyl derivative, that electronic terms ($\mathfrak{F}$, σ_I, σ_m) appear to be significant. Since the resulting correlations are lower than can be achieved without invoking electronic terms in the other 13 compound subset, the significance of these terms may be questioned. This point is further emphasized by the results obtained on the entire set of 14 compounds where two parameter equations containing difference or squared steric terms gave higher correlations than were previously reported[8] for three parameter equations which included electronic effects. For the entire set of compounds the best of the two parameter equations (eq 23) accounted for 83% of the variance. Addition of an electronic term (eq 29) accounted for less than a 2% gain in the explained variance. This study clearly supports the importance of steric effects for this series. The best single parameter equation containing a squared steric term (eq 14) accounted for 28% more variance than did the best single parameter equation based on a simple steric term (eq 5). For two parameter equations the corresponding improvement amounted to 12% (compare eq 14 and 23).

In an attempt to test the generality of the above findings, we examined the only other study[7] which has attempted to correlate *quantitatively* steroid structure with intensity of effect. Since the various steric parameters for the particular substituents reported in this series correlated better with MR (correlation coefficient for $V \sim 0.967$, for MV ~ 0.833, and for $P \sim 0.980$) than in the previous series, we anticipated that substituting these other parameters for MR would be less likely to effect a significant improvement in correlation here than was found previously. This was confirmed by our finding that V gave a negligible improvement in correlation, as compared to MR for one and two parameter equations and gave a poorer correlation for three parameter equations (difference of 0.011, 0.006, and 0.018, respectively). In all cases MV and P gave poorer

	n	r	s	F	
$\log A = 1.00\ (\pm 0.26)\ \pi - 0.030\ (\pm 0.012)\ \mathrm{MV} + 1.59$ (3.92) (−2.46)	13	0.853	0.534	13.40	(1a)
$\log A = 0.86\ (\pm 0.22)\ \pi - 0.078\ (\pm 0.030)\ \mathrm{MR} + 2.68\ (\pm 0.84)\ \sigma_I + 0.25$ (3.96) (−2.61) (3.21)	13	0.913	0.440	15.04	(2a)
$\log A = 0.97\ (\pm 0.21)\ \pi - 0.029\ (\pm 0.010)\ \mathrm{MV} + 2.37\ (\pm 0.99)\ \mathfrak{F} + 0.76$ (4.60) (−2.84) (2.40)	13	0.913	0.439	15.12	(3a)
$\log A = 0.98\ (\pm 0.20)\ \pi - 0.011\ (\pm 0.0036)\ P + 3.00\ (\pm 0.97)\ \mathfrak{F} + 0.48$ (4.79) (−3.01) (3.09)	13	0.918	0.427	16.16	(4a)
$\log A = 0.90\ (\pm 0.21)\ \pi - 0.028\ (\pm 0.010)\ \mathrm{MV} + 2.09\ (\pm 0.80)\ \sigma_I + 0.73$ (4.33) (−2.83) (2.62)	13	0.920	0.424	16.43	(5a)
$\log A = 0.94\ (\pm 0.20)\ \pi - 0.044\ (\pm 0.013)\ V + 2.89\ (\pm 0.92)\ \mathfrak{F} + 0.48$ (4.76) (−3.28) (3.14)	13	0.925	0.408	18.00	(6a)
$\log A = 0.84\ (\pm 0.18)\ \pi - 0.045\ (\pm 0.012)\ V + 2.63\ (\pm 0.69)\ \sigma_I + 0.48$ (4.62) (−3.73) (3.81)	13	0.941	0.366	23.14	(7a)

Table III. Liver Glycogen Deposition Activity and Regression Data for 9α-Substituted Cortisols

9-Subst	Obsd rel log A^a	Calcd log A^b	Calcd log A^c	$\sigma_I{}^a$	π^a	MR[a]	V^d
F	1.03	1.16	1.06	0.52	−0.17	1.20	6.20
Cl	0.67	0.31	0.22	0.47	0.39	5.96	12.24
Br	−0.52	−0.18	−0.11	0.45	0.60	8.86	14.60
I	−1.00	−1.13	−0.97	0.38	1.00	13.90	20.35
OH	−0.70	0.26	0.17	0.25	−1.16	2.62	8.04
H	0.00	−0.14	0.10	0.0	0.0	1.10	3.40
CH_3	−1.00	−0.84	−1.12	0.0	0.50	5.72	13.67
OCH_3	*e*	−0.44	−0.94	0.25	−0.47	7.24	17.37
OCH_2CH_3	*e*	−1.14	−2.16	0.25	0.03	11.86	27.60
SCN	*e*	−0.99	−1.16	0.55	0.03	15.84	25.50

[a] Reference 7. [b] Calculated by eq 1c. [c] Calculated by eq 2c. [d] Reference 17. [e] No activity detected.[7]

	n	*r*	*s*	*F*	
$\log A = 0.96\ (\pm 0.23)\ \pi - 0.12\ (\pm 0.037)\ \mathrm{MR} + 1.72$ (4.10) (−3.14)	13	0.890	0.486	19.08	(1b)
$\log A = 0.97\ (\pm 0.19)\ \pi - 0.045\ (\pm 0.010)\ \mathrm{MV} + 2.22$ (5.04) (−4.31)	13	0.925	0.405	29.58	(2b)
$\log A = 0.94\ (\pm 0.19)\ \pi - 0.00063\ (\pm 0.00014)\ \mathrm{MV}^2 + 1.51$ (4.91) (−4.41)	13	0.927	0.399	30.66	(3b)
$\log A = 0.95\ (\pm 0.18)\ \pi - 0.018\ (\pm 0.0036)\ P + 2.20$ (5.43) (−4.98)	13	0.939	0.367	37.18	(4b)
$\log A = 0.82\ (\pm 0.16)\ \pi - 0.0018\ (\pm 0.00031)\ V^2 + 1.54$ (5.07) (−5.84)	13	0.952	0.326	48.39	(5b)
$\log A = 0.89\ (\pm 0.16)\ \pi - 0.072\ (\pm 0.012)\ V + 2.15$ (5.63) (−5.88)	13	0.953	0.324	48.95	(6b)
$\log A = 0.92\ (\pm 0.15)\ \pi - 0.80\ (\pm 0.32)\ \pi^2 - 0.00061\ (\pm 0.00012)\ \mathrm{MV}^2$ (5.93) (−2.52) (−5.29) $+ 1.79$	13	0.958	0.322	33.44	(7b)
$\log A = 0.87\ (\pm 0.16)\ \pi - 0.056\ (\pm 0.0092)\ \mathrm{MV} - 1.68\ (\pm 0.62)\ \sigma_m + 3.12$ (5.62) (−6.14) (−2.71)	13	0.959	0.317	34.73	(8b)

correlations than MR. Substitution of difference and squared for simple steric terms resulted in a slightly better correlation for single parameter equations but gave little or no improvement in equations involving σ_I and a steric term.

Neither the regression equations reported by Wolff and Hansch[7] nor those which we explored afforded correlation coefficients higher than 0.864 for any two parameter equation. While the inclusion of a third parameter results in a significant improvement in the correlation coefficient and in the *F* values,[7] all of the equations have at least one coefficient which fails the *t* test for the 95% confidence level. Moreover, as Hansch and Wolff noted,[7] the number of compounds studied is not sufficient to justify the use of a three parameter equation.

In accord with the results reported by Wolff and Hansch,[7] we found that addition of a term containing π afforded the greatest improvement in going from a two to a three parameter equation. However, a review of the π values of the 9α-substituents for this series suggested that the main effect of including a term containing π might be to reduce the predicted activity for the 9α-OH compound. Calculation of the activity of these compounds by the use of Wolff and Hansch's best two parameter (σ_I and MR) equation confirmed that the predicted activity of the 9α-OH compound was too high and that the error of this prediction was the largest in the series.

Deletion of the 9α-OH compound from the series investigated by Wolff and Hansch[7] resulted in considerably improved correlations. For single parameter equations, the regression coefficient for σ_I was almost unchanged, but the regression coefficients for π and MR improved from 0.194 and 0.547 to 0.785 and 0.673, respectively. The coefficients of all single parameter equations still failed the *t* test at the 95% confidence level. However, the two parameter eq 1c and 2c were found to have correlation coefficients almost as high as had previously been found for the best three parameter equations. Moreover only these, from all one, two, or three parameter equations examined in this abbreviated series, were found to satisfy the *F* test at the 95% confidence level, and they are the only equations which we have found in the full or abbreviated series whose coefficients satisfy the *t* test at the 95% confidence level.

$\log A = 2.519\ (\pm 0.635)\ \sigma_I - 0.152\ (\pm 0.0313)\ \mathrm{MR}$ (3.97) (−4.86) $+ 0.030$ (1c)

r	SE	*F*
0.955	0.327	15.63

$\log A = 2.487\ (\pm 0.704)\ \sigma_I - 0.119\ (\pm 0.028)\ V$ (3.53) (−4.30) $+ 0.503$ (2c)

r	SE	*F*
0.944	0.364	12.36

While for the compounds used in the regression, eq 1c gives a slightly better correlation than eq 2c, as is shown in Table III; eq 2c proved superior in predicting low activity for the three compounds reported to be inactive.

Both equations indicate that in this series activity decreases with increasing bulk at the 9α position and increases as the 9α-substituent becomes more electron withdrawing. While these conclusions are in agreement with those of Hansch and Wolff,[7] they bring into question the significance of the π term. These authors concluded that a π containing term was required for a good correlation and suggested that its importance might relate either to the effect of lipophilicity on transport or might reflect a hydrophobic interaction at the active site. An alternative interpretation is that the role of their π term is to lower the predicted activity of compounds containing strongly hydrated 9α-substituents. In an aqueous environment such substituents would have a larger effective bulk and hence a lower than predicted activity. (Calculations by eq 2c of the effective van der Waals volume required to account for the observed activity give a volume comparable to a monohydrated hydroxyl group.) Such an explanation seems to be in accord with the correlation between molecular shape and glucocorticoid activity reported by Weeks et al.[24]

In conclusion, we believe that these studies clearly establish the importance of steric factors in QSAR involving steroid hormones. The nature of the substituents in both studies was such as to permit a clear differentiation between steric parameters and π.[22] Moreover, the studies illustrate the importance of investigating a variety of steric parameters and they suggest the possible importance of difference or squared steric terms. Further studies will be required to establish the generality of the latter effect.

Acknowledgment. The authors are grateful to the SUNY/Buffalo Computing Center for their donation of computing services of a CDC 6400 computer.

References and Notes

(1) W. van Valkenburg, *Adv. Chem. Ser.*, **No.** 114 (1972).
(2) T. Higuchi and S. S. Davis, *J. Pharm. Sci.*, **59**, 1376 (1970).
(3) K. C. James, P. J. Nicholls, and M. Roberts, *J. Pharm. Pharmacol.*, **21**, 24 (1969).
(4) K. C. James, *Experientia*, **28**, 479 (1972).
(5) M. A. Q. Chaudry and K. C. James, *J. Med. Chem.*, **17**, 157 (1974).
(6) J. A. Ostrenga, *J. Med. Chem.*, **12**, 349 (1969).
(7) M. E. Wolff and C. Hansch, *Experientia*, **29**, 1111 (1973).
(8) M. E. Wolff and C. Hansch, *J. Med. Chem.*, **17**, 898 (1974).
(9) J. G. Topliss and E. L. Shapiro, *J. Med. Chem.*, **18**, 621 (1975).
(10) C. Hansch, A. R. Stewart, S. M. Anderson, and D. Bentley, *J. Med. Chem.*, 11, 1 (1967).
(11) C. Hansch in "Drug Design", Vol. 1, E. J. Ariens, Ed., Academic Press, New York, N.Y., 1971, p 327.
(12) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
(13) J. C. McGowan, *J. Appl. Chem.*, **2**, 323 (1952); 4, 41 (1954).
(14) J. C. McGowan, *Nature (London)*, **200**, 1317 (1963).
(15) O. Exner, *Collect. Czech. Chem. Commun.*, **32**, 24 (1967).
(16) O. Exner, *Collect. Czech. Chem. Commun.*, **32**, 1 (1967).
(17) A. Bondi, *J. Phys. Chem.*, **68**, 441 (1964).
(18) G. Teutsch, L. Weber, G. Page, E. L. Shapiro, H. L. Herzog, R. Neri, and E. J. Collins, *J. Med. Chem.*, **16**, 1370 (1973).
(19) B. A. Cooke and D. K. Vallance, *Biochem. J.*, **97**, 672 (1965).
(20) L. Terenius, *Steroids*, **23**, 909 (1974).
(21) J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972).
(22) P. N. Craig, *J. Med. Chem.*, **14**, 680 (1971).
(23) E. Kutter and C. Hansch, *J. Med. Chem.*, **12**, 647 (1969).
(24) C. M. Weeks, W. L. Duax, and M. E. Wolff, *J. Am. Chem. Soc.*, **95**, 2865 (1973).

Semisynthetic Cephalosporins. Synthesis and Structure–Activity Relationships of 7-Sulfonylacetamido-3-cephem-4-carboxylic Acids

R. M. DeMarinis,* J. R. E. Hoover, L. L. Lam, J. V. Uri, J. R. Guarini, L. Phillips, P. Actor, and J. A. Weisbach

Research and Development Division, Smith Kline & French Laboratories, Philadelphia, Pennsylvania 19101.
Received September 22, 1975

The synthesis and in vitro and in vivo activities of a series of 7-sulfonylacetamido-3-cephem-4-carboxylic acids with acetoxymethyl or heterocyclic thiomethyl substituents at the 3 position are described. Lengthening the alkyl chain attached to the sulfonyl group increased gram-positive activity but the effect on gram-negative activity was variable. Other structural changes on the 7-acyl side chain resulted in only minor changes in in vitro activity. The protective effectiveness in infected mice generally paralleled the in vitro activity, except that the butylsulfonyl derivatives were less protective than predicted by in vitro activity. Replacement of the 3-acetoxymethyl by a 3-heterocyclic thiomethyl group resulted in an overall improvement of activity both in vitro and in vivo.

A majority of cephalosporins that possess significant antibacterial activity have on the 7 position an acetamido group to which is attached a heterocyclic or benzene ring. This ring may be attached directly or linked through a heteroatom. These substituents are relatively large and complex by virtue of the attached aromatic ring. An exception to this is cephacetrile which has a simple cyanoacetamido grouping at the 7 position. It seemed reasonable that other simple side chains might impart broad-spectrum antibacterial activity and this led us to investigate derivatives of mercaptoacetic acid. Previously, we reported the broad-spectrum activities of 7-trifluoromethylthioacetamido-3-(1-methyl-1*H*-tetrazol-5-ylthiomethyl)-3-cephem-4-carboxylic acid (SK&F 59962) and its closely related analogues.[1,2] This article presents some of the structure–activity studies that led to this compound. It extends the previous work to additional derivatives of mercaptoacetic acid in which the sulfur atom is oxidized to the sulfonyl state. It also describes the structure–activity relationships observed by altering the length of the alkyl group on the sulfur atom, by substituting for the alkyl group strongly electron-donating or -withdrawing substituents such as amino or trifluoromethyl, and by altering the 7-substituent size through substitution of phenyl for the alkyl group and through alkylation of the α-methylene. Additionally, the effect on activity of varying the 3-substitutent is presented; for each alteration at the 7 position, the 3-acetoxymethyl analogue was compared with one or more 3-heterocyclic thiomethyl analogues.

Chemistry. The cephalosporins were prepared by acylation of 7-aminocephalosporanic acid (7-ACA) or its 3-heterocyclic thiomethyl analogues. The latter were made