Journal of Medicinal Chemistry

Volume 20, Number 4 April 1977

A Manual Method for Applying the Hansch Approach to Drug Design[1]

John G. Topliss

Department of Chemical Research, Schering-Plough Research Division, Bloomfield, New Jersey 07003.
Received August 18, 1976

A procedure is described in which an initial small group of compounds is selected, tested, and ordered according to potency. The potency order in the group is then compared to the tabulated potency order calculated for various parameter dependencies relating to hydrophobic, electronic, and steric effects. From this activity pattern analysis the probable operative parameters can be deduced and a new substituent selection made for the synthesis of potentially more potent analogues. Application of the method is illustrated with a series of examples. It differs from a previously described decision tree, single compound stepwise approach in that it involves the batchwise analysis of small groups of compounds, usually the preferred procedure for logistical reasons if the compounds are relatively easy to synthesize.

A very common problem in drug design is to find the optimum substitution on a benzene ring or on the benzenoid portion of a fused ring system in an active lead compound for maximization of drug potency. A literature survey has shown that some 40% of all reported compounds incorporate an unfused benzene ring.[3] Also more than 50% of drug-oriented patents are concerned with substituted benzenes.[4] Efficient solutions to the problem of optimizing substitution would therefore have much value. With the advent[5] and subsequent development[6] of the Hansch method for structure–activity correlations, a more rational approach to this problem became possible. Subsequently, a number of publications[7-11] have dealt with the most advantageous procedural strategies for substituent selection in applying the Hansch method.

In a previous publication the utilization of operational schemes for analogue synthesis in drug design was described.[2] This method, which is based on the fundamental assumptions of the Hansch approach, involves the stepwise selection for synthesis of new analogues of an active lead compound designed to maximize the chances of synthesizing the most potent compounds in the series as early as possible. The stepwise selection takes the form of a decision tree and does not require multiple regression analysis. Application of the method should be particularly advantageous when analogue synthesis is difficult and slow and test results are relatively rapid. It is also of interest as a way of applying Hansch type principles without the use of statistical procedures and computers.

The present paper describes a procedure, not involving statistical methodology, by which groups of analogues representing different types of substitution on a benzene ring can be selected and synthesized at one time. As such it may be viewed as a manual method for applying the Hansch approach to drug design.

The initial group of analogues selected for synthesis consists of the first five compounds in Table I. This set of five compounds comprises the top section of the operational scheme for aromatic substitution previously described[2] and the compounds should be readily accessible from a synthetic standpoint.

The projected order of potency of these five compounds for various parameter dependencies is listed in Table II. Comparison with the actual experimentally determined potency order allows a possible deduction to be made concerning the probable operative parameters which in turn provides the basis for a new substituent selection from Table III. The new substituent selections are examples of suitable choices and clearly other selections can be made. An examination of Tables I and II reveals that the expected potency order for some parameter dependencies is quite similar. Given the fact that only five compounds are being considered and taking into account some possible uncertainty in the biological data it is apparent that a narrowing of the probable operative parameters to a related group of possibilities may be accomplished rather than a precise identification. These related groups of probable operative parameters are listed in the left-hand column of Table III. It should also be clear that the biological potencies of the initial compound group have to show sufficient spread to permit a meaningful analysis.

The parameter dependencies listed in Table II encompass those most commonly found in Hansch type correlations, viz. linear and parabolic π, σ, and a wide range of $\pi \pm \sigma$. Steric effects, E, are also included where these exert a dominant influence. The choice of coefficients in the various $\pi \pm \sigma$ dependencies reflects a range of relative weightings of the two parameters. Different relative weightings are included only if they produce a significant change in the potency order for the five compounds in Table II. Thus, the potency order for π, σ, and $\pi + \sigma$ is very similar and there would be no point in including columns for $2\pi + \sigma$, $\pi + 2\sigma$, etc. On the other hand, the potency order for π and $-\sigma$ is essentially reversed, hence the opportunity to consider not only $\pi - \sigma$ but $2\pi - \sigma$, $\pi - 2\sigma$, and $\pi - 3\sigma$.

At this stage it is appropriate to illustrate the application of the method with some examples. In reviewing the

Table I. Parameter Values[a]

Substituents	π	σ	$\pi - \sigma$	$2\pi - \sigma$	$\pi - 2\sigma$	$\pi - 3\sigma$	$-\sigma$	$\pi + \sigma$	$2\pi - \pi^2$
H	0	0	0	0	0	0	0	0	0
4-Cl	0.71	0.23	0.48	1.19	0.25	0.02	−0.23	0.94	0.92
3,4-Cl_2	1.25	0.52	0.73	1.98	0.21	−0.31	−0.52	1.77	0.94
4-CH_3	0.56	−0.17	0.73	1.29	0.90	1.07	0.17	0.39	0.81
4-OCH_3	−0.02	−0.27	0.25	0.23	0.52	0.79	0.27	−0.29	−0.04
3-CF_3,4-Cl	1.59	0.66	0.93	2.52	2.91	−0.39	−0.66	2.25	0.65
3-CF_3,4-NO_2	0.60	1.21	−0.61	−0.01	−1.82	−2.97	−1.21	1.81	0.84
4-CF_3	0.88	0.54	0.34	1.22	−0.20	−0.74	−0.54	1.42	0.99
2,4-Cl_2	1.42	0.46	0.96	2.38	0.50	0.04	−0.46	1.88	0.82
c-C_5H_9	2.14	−0.02	2.16	4.30	2.18	2.20	0.02	2.12	−0.30
c-C_6H_{11}	2.51	−0.22	2.73	5.24	2.95	3.17	0.22	2.29	−1.28
4-$CH(CH_3)_2$	1.53	−0.05	1.58	3.11	1.63	1.68	0.05	1.48	0.72
4-$C(CH_3)_3$	1.98	−0.20	2.18	4.16	2.38	2.58	0.20	1.78	0.04
3,4-$(CH_3)_2$	0.99	−0.30	1.29	2.28	1.59	1.89	0.30	0.69	1.00
4-$O(CH_2)_3CH_3$	1.55[b]	−0.32	1.87	3.42	2.19	2.51	0.32	1.23	0.70
4-OCH_2Ph	2.13[b]	−0.42	2.55	4.68	2.97	3.39	0.42	1.71	−0.28
4-$N(C_2H_5)_2$	1.18	−0.83	2.01	3.19	3.84	4.67	0.83	0.35	0.97
4-$N(CH_3)_2$	0.18	−0.83	1.01	1.19	1.84	2.67	0.83	−0.65	0.33
4-NH_2	−1.23	−0.66	−0.57	−1.80	0.09	0.75	0.66	−1.89	−3.97
4-NHC_4H_9	1.45	−0.51	1.96	2.39	2.47	2.98	0.51	0.94	0.80
4-$OCH(CH_3)_2$	1.03[b]	−0.45	1.48	2.51	1.93	2.38	0.45	0.58	1.00
3-CH_3,4-OCH_3	0.54	−0.26	0.80	1.34	1.06	1.32	0.26	0.28	0.79
4-Br	0.86	0.23	0.57	1.49	0.40	1.55	−0.23	1.09	0.98
3-CF_3	0.88	0.43	0.45	1.43	0.02	2.17	−0.43	1.31	0.99
4-C_2H_5	1.02	−0.15	1.17	2.19	1.32	1.47	0.15	0.87	1.00
4-$O(CH_2)_2CH_3$	1.05	−0.25	1.30	2.35	1.55	1.80	0.25	0.80	1.00
3-CH_3,4-Cl	1.29	0.17	1.12	2.41	0.95	0.78	−0.17	1.46	0.92
3-Cl	0.71	0.37	0.34	1.05	−0.03	−0.30	−0.37	1.08	0.92
3-CH_3	0.56	−0.07	0.63	1.19	0.70	0.77	0.07	0.49	0.81
3-OCH_3	−0.02	0.12	−0.14	−0.16	−0.26	−0.38	−0.12	0.10	−0.04
3-$N(CH_3)_2$	0.18	−0.15	0.33	0.51	0.48	0.63	0.15	0.03	0.33
3,5-Cl_2	1.25	0.75	0.50	1.75	−0.25	−1.00	−0.75	2.20	0.94
2-Cl	0.71	0.23	0.48	1.19	0.25	0.02	−0.23	0.94	0.92
2-CH_3	0.56	−0.17	0.73	1.29	0.90	1.07	0.17	0.39	0.81
2-OCH_3	−0.02	−0.27	0.25	0.23	0.52	0.79	0.27	−0.29	−0.04
2-F	0.14	0.06	0.08	0.22	0.02	−0.04	−0.06	0.20	0.28
4-F	0.14	0.06	0.08	0.22	0.02	−0.04	−0.06	0.20	0.28
4-$NHCOCH_3$	−0.97	0.00	−0.97	−1.94	−0.97	−0.97	0.00	−0.97	−2.88
4-$NHSO_2CH_3$	−1.18	0.03	−1.21	−2.39	−1.24	−1.27	−0.03	−1.15	−3.75
4-NO_2	−0.28	0.78	−1.06	−1.34	−1.84	−2.62	−0.78	0.50	−0.64
4-$COCH_3$	−0.55	0.50	−1.05	−1.60	−1.55	−2.05	−0.50	−0.05	−1.40
4-SO_2CH_3	−1.63	0.72	−2.35	−3.98	−3.07	−3.79	−0.72	−0.72	−5.92
4-$CONH_2$	−1.49	0.36	−1.85	−3.34	−2.21	−2.57	−0.36	−1.13	−5.20
4-SO_2NH_2	−1.82	0.57	−2.39	−4.21	−3.01	−3.53	−0.57	−1.25	−6.95

[a] C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, 16, 1207 (1973). [b] Estimated value.

Table II. Potency Order for Various Parameter Dependencies

	Parameters									
Substituents	π	$2\pi - \pi^2$	σ	$-\sigma$	$\pi + \sigma$	$2\pi - \sigma$	$\pi - \sigma$	$\pi - 2\sigma$	$\pi - 3\sigma$	E_4[a]
3,4-Cl_2	1	1-2	1	5	1	1	1-2	3-4	5	2-5
4-Cl	2	1-2	2	4	2	2-3	3	3-4	3-4	2-5
4-CH_3	3	3	4	2	3	2-3	1-2	1	1	2-5
4-OCH_3	4-5	4-5	5	1	5	4	4	2	2	2-5
H	4-5	4-5	3	3	4	5	5	5	3-4	1

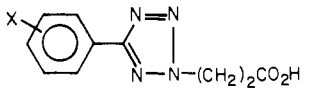
[a] Unfavorable steric effect from 4 substitution.

Table III. New Substituent Selection

Probable operative parameters	New substituent selection
π, $\pi + \sigma$, σ	3-CF_3, 4-Cl; 3-CF_3, 4-NO_2; 4-CF_3; 2,4-Cl_2; 4-c-C_5H_9; 4-c-C_6H_{11}
π, $2\pi - \sigma$, $\pi - \sigma$	4-$CH(CH_3)_2$; 4-$C(CH_3)_3$; 3,4-$(CH_3)_2$; 4-$O(CH_2)_3CH_3$; 4-OCH_2Ph; 4-$N(C_2H_5)_2$
$\pi - 2\sigma$, $\pi - 3\sigma$, $-\sigma$	4-$N(C_2H_5)_2$; 4-$N(CH_3)_2$; 4-NH_2; 4-NHC_4H_9; 4-OH; 4-$OCH(CH_3)_2$; 3-CH_3,4-OCH_3
$2\pi - \pi^2$	4-Br; 3-CF_3; 3,4-$(CH_3)_2$; 4-C_2H_5; 4-$O(CH_2)_2CH_3$; 3-CH_3,4-Cl
	3-Cl; 3-CH_3; 3-OCH_3; 3-$N(CH_3)_2$; 3-CF_3; 3,5-Cl_2
Ortho effect	2-Cl; 2-CH_3; 2-OCH_3; 2-F
Other	4-F; 4-$NHCOCH_3$; 4-$NHSO_2CH_3$; 4-NO_2; 4-$COCH_3$; 4-SO_2CH_3; 4-$CONH_2$; 4-SO_2NH_2

results of these examples criteria for judging the degree of success of the method are (a) rapid identification of the more potent analogues and (b) deduction of the correct parameter dependency as determined by a standard Hansch analysis on a larger series. The first example concerns the antiinflammatory activity of some 5-aryl-

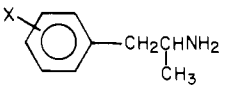
Table IV. Antiinflammatory Activity of 5-Aryltetrazolylpropionic Acids[a]



X	AI[b]	Rank order: Obsd	Rank order: Calcd for E_4
Initial Compound Group			
3,4-Cl_2	6.2	2	2–5
4-Cl	5.9	3	2–5
4-CH_3	3.1	5	2–5
4-OCH_3	4.9	4	2–5
H	8.2	1	1
Second Compound Group			
3-Cl	11.2	1–3	
3-CH_3	7.9	4	
3-CF_3	5.7	5	
3,5-Cl_2	11.2	1–3	
Other Compounds			
3-Br	11.2	1–3	
3-NH_2	0.3	6	

[a] Reference 12. [b] Activity index = ten times the ratio of the mean pleural exudate volumes of the test compound at 1 mmol/kg over 1 mmol/kg of aspirin.

Table V. Inhibition of PNMT by 3- and 4-Substituted Amphetamines[a]



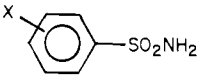
X	pI_{50}[b]	Rank order: Obsd	Rank order: Calcd for $\pi + \sigma$
Initial Compound Group			
3,4-Cl_2	5.10	1	1
4-Cl	3.60	2	2
4-CH_3	3.14	3	3
4-OCH_3	2.57	5	5
H	2.89	4	4
Second Compound Group			
4-CF_3	4.00		

[a] Reference 13. [b] Negative log of the molar concentration of inhibitor needed for 50% inhibition of enzyme activity.

tetrazolylpropionic acids[12] (Table IV). In this series rank ordering according to potency of the initial compound group reveals the parent phenyl compound to be substantially more potent than the others. Reference to Table II indicates the inference to be drawn is that there is probably an adverse steric effect at position 4 and that a second group of compounds should be synthesized and tested with substituents at the 3 and 3,5 positions. Among the indicated second compound group shown in Table IV are three of the four most potent analogues in the 28 compound series.

The method may be further illustrated by reference to a series of 3- and 4-substituted amphetamines which are inhibitors of phenethanolamine *N*-methyltransferase (Table V). Rank ordering by potency of the initial compound group suggests a $\pi + \sigma$ or possibly a π or σ dependency (Table II). The next step would be the synthesis and testing of the second compound group (Table V). Only one compound from this group was reported by Fuller et al.[13] but its potency was in line with the parameter dependency indicated by analysis of the first compound group. The 3,4-Cl_2 analogue included in the first group

Table VI. Sulfonamide Carbonic Anhydrase Inhibitors[a]



X	Log $1/K_1$[b]	Rank order: Obsd	Rank order: Calcd π	Rank order: Calcd $\pi + \sigma$
Initial Compound Group				
3,4-Cl_2	1.40	1	1	1
4-Cl	0.72	2	2	2
4-CH_3	0.42	3	3	3
4-OCH_3	0.35	4	4–5	5
H	0.22	5	4–5	4
Second Compound Group				
3-CF_3,4-NO_2	1.85			

[a] Reference 14. [b] K_1 represents carbonic anhydrase inhibition constant. The value $1/K_1$ is proportional to the magnitude of carbonic anhydrase inhibitory activity.

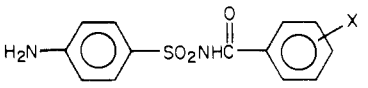
was the most potent compound in the 22 compound series. By means of a conventional Hansch treatment using multiple regression analysis, Fuller et al.[13] established that activity correlated in a positive sense with $\pi + \sigma$. Thus, the conclusion with respect to parameter dependency reached by analysis of the initial five compound group was correct. It is important to recognize that it is not usually necessary to know the precise parameter dependency in order to select analogues which have a good probability of showing increased potency. Thus, the analogue sequence H; 4-Cl; 3,4-Cl_2; 3-CF_3,4-Cl; 3,4-$(CF_3)_2$, having progressively higher values for both $\Sigma\pi$ and $\Sigma\sigma$, would produce progressively more potent analogues for π, σ, or $\pi + \sigma$ dependencies.

Kakeya et al.[14] reported structure–activity studies on a series of 19 substituted sulfonamides which show carbonic anhydrase inhibitory activity. Tabulation of the initial compound group (Table VI) and rank ordering according to biological potency indicated a π or $\pi + \sigma$ relationship (Table II). This is consistent with Kakeya et al. analysis of the entire 19 compound series using regression analysis which showed that activity was $\pi + \sigma$ (primarily σ) related. Of the indicated second compound group only one member, the 3-CF_3,4-NO_2 analogue, had been synthesized. However, this compound was the most potent in the entire series.

It is not actually necessary to have information on all five compounds in the specified initial compound group. Often useful conclusions can be drawn on only four compounds. For example, in a series of 13 *N'*-benzoylsulfanilamides[15] with antibacterial activity against *Escherichia coli*, the 3,4-Cl_2 analogue is not reported. Ordering the four available compounds in the initial compound group as shown in Table VII leads to the conclusion, by reference to Table II, that activity is probably $-\sigma$ dependent. This is the same as the conclusion reached by Cammarata[15] based on a regression analysis of the 13 compound series. In the indicated second compound group taken from Table III, the 3-CH_3,4-OCH_3 analogue had been made and was in the high potency category. Other compounds, which formed part of the reported series and expected to show high potency based on a $-\sigma$ dependency, are the 3-CH_3, 4-C_2H_5, 4-*n*-C_3H_7, and 4-*i*-C_3H_7 analogues. The observed potencies of these are consistent with this projection.

An interesting case is the structure–activity relationships of 9-(X-phenyl)guanine inhibitors of xanthine oxidase reported by Silipo and Hansch.[16] Analysis of the initial five compound group taken from this series of 33 com-

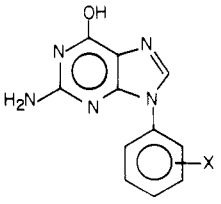
Table VII. In Vitro Activities of Benzoylsulfanilamides against *Escherichia coli*[a]



X	Log $1/C^b$	Rank order: Obsd	Rank order: Calcd for $-\sigma$
Initial Compound Group			
4-Cl	5.10	4	4
4-CH_3	5.40	1-2	2
4-OCH_3	5.40	1-2	1
H	5.25	3	3
Second Compound Group			
3-CH_3,4-OCH_3	5.25		
Other Compounds			
3-CH_3	5.40		
4-C_2H_5	5.62		
4-*n*-C_3H_7	5.18		
4-*i*-C_3H_7	5.40		
3-CF_3	4.65		
4-NO_2	4.50		
4-CN	4.05		

[a] Reference 15. [b] C is the minimum inhibitory concentration against *E. coli*.

Table VIII. 9-(X-Phenyl)guanine Inhibitors of Xanthine Oxidase[a]

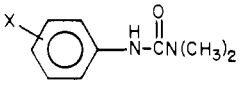


X	Log $1/C^b$	Rank order: Obsd	Rank order: Calcd for E_4
Initial Compound Group			
3,4-Cl_2	5.96	3	2-5
4-Cl	5.74	4-5	2-5
4-CH_3	5.80	4-5	2-5
4-OCH_3	6.30	1-2	2-5
H	6.39	1-2	1
Second Compound Group			
3-Cl	6.57		
3-CH_3	6.62		
3-OCH_3	6.66		
3-CF_3	6.82		

[a] Reference 16. [b] C is the concentration which produces 50% inhibition of xanthine oxidase.

pounds (Table VIII) strongly suggests that there is an adverse steric effect from 4 substitution. Thus, the two most potent members of this group, the 4-unsubstituted and the 4-OCH_3 compounds, have the lowest steric requirements, whereas the two least potent, the 4-CH_3 and 4-Cl analogues, have the highest steric requirements as measured by E_s values. The intermediate potency value of the 3,4-Cl_2 analogue is consistent with the idea of a positive effect on activity of the 3-chloro substituent. Of the second compound group specified in Table III, the 3-Cl, 3-CF_3, and 3-CH_3 analogues had been reported by Silipo and Hansch and they were all more potent than any of the compounds in the initial compound group. From a multiple regression analysis on the entire series, Hansch and Silipo concluded that activity was negatively correlated with substituent size at the 4 and 2 positions and positively

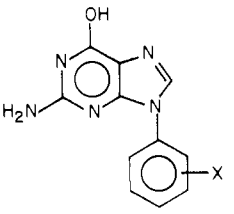
Table IX. Inhibition of Hill Reaction by Phenyldimethylureas[a]



X	Log $1/C^b$	Rank order: Obsd	Rank order: Calcd for $\pi + \sigma$
Initial Compound Group			
3,4-Cl_2	6.70	1	1
4-Cl	5.40	2	2
4-CH_3	4.52	3	3
4-OCH_3	4.30	5	5
H	4.40	4	4

[a] Reference 17. [b] C is the concentration which produces 50% inhibition of the Hill reaction.

Table X. Inhibition of Guanine Deaminase by 9-(X-Phenyl)guanines[a]



X	Log $1/C^b$	Rank order obsd
Initial Compound Group		
3,4-Cl_2	6.03	2
4-Cl	5.42	3-4
4-CH_3	5.45	3-4
4-OCH_3	6.70	1
H	5.00	5
Second Compound Group		
4-OC_2H_5	7.01	
4-$O(CH_2)_3C_6H_5$	7.25	

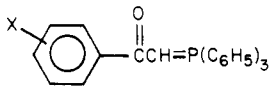
[a] Reference 18. [b] C represents the molar concentration for 50% reversible inhibition of guanine deaminase from rabbit liver when assayed at pH 7.4.

correlated with the combined molecular refractivity of substituents at the 3 and 4 positions.

A series of herbicidal phenyldimethylureas which inhibit photosynthesis and originally studied from a structure–activity standpoint by Hansch and Deutsch[17] contains all five compounds needed for the initial compound group analysis (Table IX). The rank order of these analogues corresponds with the expected order for a $\pi + \sigma$ parameter dependency, consistent with the findings of Hansch and Deutsch arrived at by multiple regression analysis on the 12 member series. The indicated new substituent selection (Table III) could not be checked since these compounds were not reported in the original study.

Even where the rank order of the substituents with respect to activity does not correspond completely to any parameter dependency listed in Table II, it may be still possible to conduct a useful analysis leading to the identification of more potent analogues. This is illustrated by reference to a series of 9-(X-phenyl)guanines which inhibit guanine deaminase (from rabbit liver)[18] (Table X). From Table II no combination of parameters can account for the 4-OCH_3 analogue having the highest activity and 3,4-Cl_2 the second highest. Therefore, it must be concluded that special factors are operative which are not part of the assumptions made in constructing Table II. Leaving out temporarily the 4-OCH_3 compound from consideration, the

Table XI. Hypoglycemic Activity of Phenacyltriphenylphosphoranes[a]



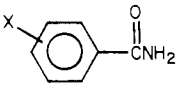
X	Hypoglycemic act.[b]	Rank order: Obsd	Rank order: Calcd for E_4
Initial Compound Group			
3,4-Cl_2	2	3-5	2-5
4-Cl	-21	1-2	2-5
4-CH_3	-2	3-5	2-5
4-OCH_3	-4	3-5	2-5
H	-24	1-2	1
Second Compound Group			
3-Cl	-42		
3-CH_3	-30		
3-OCH_3	-42		
3-CF_3	-22		

[a] Reference 19. [b] Expressed as the percent difference between the mean change in control and treated groups (rats) at 4 h.

remaining four analogues in the group display a relative potency pattern consistent with a $+\pi$ dependency. One consideration is that the 4-OCH_3 analogue activity determination might be in error. However, in order to fit the $+\pi$ dependency idea, the magnitude of the experimental error would have to be very large (from 6.70 to ca. 5.0), which seems improbable. A second consideration would be that the 4-OCH_3 group confers a large activity enhancement through a factor or factors which are not part of the assumptions made in constructing Table II. Assuming this to be the case, indicated new analogues for synthesis would be other 4-alkoxy and also 4-aralkoxy compounds where this same special factor would probably also be operative. Higher π values associated with the new 4-alkoxy compounds might also be expected to have a positive effect on activity. Two compounds of this type, the 4-OC_2H_5 and 4-$O(CH_2)_3C_6H_5$, were reported in the original study and had the third and second highest activities, respectively, in the entire 33 compound series. Other choices, not reported, would be the 4-$O(CH_2)_2CH_3$, 4-$OCH(CH_3)_2$, 4-$OCH_2C_6H_5$, and 4-$O(CH_2)_2C_6H_5$ analogues. Silipo and Hansch[18] concluded, on the basis of a detailed study of the series using multiple regression analysis and surveying 1920 equations, that activity was positively correlated with the π value of the 4-substituent, the molar refractivity of the 3-substituent, and an indicator variable associated with the presence of a 4-alkoxy substituent, and negatively correlated with the steric constant of the 2-substituent. The present analysis, although not providing this degree of understanding of structure–activity relationships, nevertheless pointed the way to some of the most potent compounds in the series in a relatively simple manner while requiring consideration of only a very limited number of compounds.

Another illustration of the use of the method can be made from the publication of Blank et al.[19] on the hypoglycemic activity of a series of phenacyltriphenylphosphoranes. Although the rank order of the initial compound group (Table XI) does not exactly correspond to any parameter dependency given in Table II, it is more consistent with the idea that there is probably some unfavorable steric effect from 4 substitution in view of the rank of the 4-H compound. This leads from Table III to the selection of the second compound group shown in Table XI. Of this latter group data on four compounds

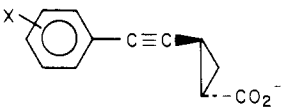
Table XII. Complex Formation between X-$C_6H_4CONH_2$, Alcohol Dehydrogenase, and DPNH[a]



X	Log $K_{ER.I}$[b]	Rank order: Obsd	Rank order: Calcd for $\pi - \sigma$
Initial Compound Group			
4-Cl	-1.93	2	2
4-CH_3	-1.78	1	1
4-OCH_3	-2.20	3	3
H	-2.72	4	4
Second Compound Group			
4-$CH(CH_3)_2$	-1.70		

[a] Reference 20. [b] $K_{ER.I}$ is the dissociation constant for the complex of alcohol dehydrogenase with benzamides and DPNH.

Table XIII. Fibrinolytic Activity of 2-Phenethynylcyclopropanecarboxylates[a]



X	Log $1/C$[b]	Rank order: Obsd	Rank order: Calcd for π
Initial Compound Group			
4-Cl	1.92	1	1
4-CH_3	1.82	2	2
4-OCH_3	1.70	3-4	3-4
H	1.70	3-4	3-4

[a] Reference 21. [b] C is the molar concentration of anion which dissolves a standard clot of human plasma in 24 h.

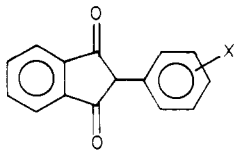
are available and show a definite trend to substantially increased potency. The two most potent compounds in the series are included in this group.

In a series of benzamides which inhibit alcohol dehydrogenase,[20] the 3,4-Cl_2 analogue is missing from the specified initial compound group. Rank ordering of the four available compounds as shown in Table XII suggests (Table II) a $\pi - \sigma$ activity dependency for the series. Of the indicated second compound group (Table III) only the 4-$CH(CH_3)_2$ was reported and this was the most potent member of the series. It is noteworthy that Hansch et al.[20] established by regression analysis on the 15 compound series that activity correlated with a $\pi - \sigma$ function.

A series of 2-phenethynylcyclopropanecarboxylates exhibiting fibrinolytic activity[21] provides another illustration of how, from rank ordering four compounds, it is possible to make a useful judgement concerning probable parameter dependency (Table XIII). By reference to Table II a probable linear dependence of activity on π is indicated. This turns out to be consistent with the conclusion reached by Yoshimoto et al.[21] who found activity to be a linear function of log P by a regression analysis on the nine compound series.

A further example where four compounds of the initial compound group are available is a series of 2-phenyl-1,3-indandiones which are uncouplers of oxidative phosphorylation in rat liver mitochondria.[22] Rank ordering by potency of the four compounds (Table XIV) and reference to Tables I and II suggests that activity is π dependent with the possibility of a small $-\sigma$ effect. This leads (Table III) to the choice of the second compound group which

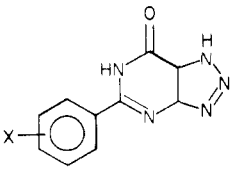
Table XIV. Uncoupling Activities of 2-Phenyl-1,3-indandiones[a]



X	Log $1/C_{50}$[b]	Rank order: Obsd	Rank order: Calcd for $2\pi - \sigma$
Initial Compound Group			
4-Cl	4.20	1-2	1-2
4-CH_3	4.10	1-2	1-2
4-OCH_3	3.75	3-4	3
H	3.74	3-4	4
Second Compound Group			
4-$CH(CH_3)_2$	4.87		
4-$C(CH_3)_3$	5.03		

[a] Reference 22. [b] C_{50} = concentration for 50% uncoupling.

Table XV. Antiallergic Activity of 2-Phenyl-8-azapurin-6-ones[a]



X	Rel act.[b]	Rank order: Obsd	Rank order: Calcd for E_4
Initial Compound Group			
4-Cl	2	2	2-4
4-CH_3	0.8	3-4	2-4
4-OCH_3	1	3-4	2-4
H	4	1	1
Second Compound Group			
H	4	1-3	
3-CH_3	4	1-3	
3-OCH_3	2	5	
3-$N(CH_3)_2$	0.5	4	
3-CF_3	4	1-3	
Third Compound Group			
H	4	2	
2-Cl	0.2	4	
2-CH_3	0.04	5	
2-OCH_3	10	1	
2-F	0.5	3	

[a] Reference 23. [b] Activity relative to disodium cromoglycate (=1) in the rat PCA test following iv administration.

contains two compounds, the 4-$CH(CH_3)_2$ and 4-$C(CH_3)_3$ analogues, where information is available. These both showed greatly enhanced potency over compounds in the initial compound group. By application of multiple regression analysis to the entire 44 member series van den Berg et al.[22] concluded that π was the most important parameter controlling activity with a $-\sigma$ effect playing a secondary role. Ortho substitution was shown to be a negative influence.

Situations will sometimes arise where after following the procedure through the second compound group phase no clear-cut progress toward identifying analogues with improved activity will have been made. In such cases an examination of 2-substituents is indicated as a next step. An example is provided by a series of 2-phenyl-8-azapurin-6-ones showing antiallergic activity.[23] From a study of the relative activities of the initial compound group (Table XV), it is clear that 4 substitution is unfavorable. This suggests an examination of 3-substituents (Table II) for the second compound group. Here no advantage is gained in activity over the parent compound which leads to the 2-substituted analogues in the third compound group. Of this group the 2-OCH_3 compound has exceptional activity compared to the 2-Cl and 2-CH_3 analogues and is considerably more active than the parent compound. These relative activities cannot be explained on the basis of π, σ, and E_s influences and it is clear from this that some special factor is involved in explaining the high potency of the 2-OCH_3 analogue and it would be obvious from this point to explore other 2-alkoxy compounds. The 2-alkoxy compounds are, in fact, the most active group in the entire 44 compound series with the 2-*n*-propoxy compound showing peak activity. Wooldridge and co-workers[23] were able to show that the special factor involved was probably a strong positive activity influence from the hydrogen bonding capacity of the ortho substituents and by a multiple regression technique were able to establish an equation relating activity quantitatively to this positive factor and negatively to the bulk of the 2-substituent. It is interesting to note that the special potential of the 2-OCH_3 is recognized after an examination of ten compounds in the present analysis which can be considered quite efficient considering the unusual aspects involved. Wooldridge and co-workers[23] reached the same stage in their analysis after an attempted correlation using multiple regression analysis with partition, electronic, and steric parameters on an initial ten compound series.

As illustrated in the last example, the examination of 2-substituents is made after first exploring the potential of 4- and 3-substituents. For most series there is a priori no reason to expect that one position of substitution will be superior to another as far as activity is concerned. Thus, the choice of sequence can logically be made on the basis of synthetic and economic considerations which generally favor 4 and 3 over 2 substitution. In the exceptional cases where there is a strong rationale for expecting favorable effects on activity from a particular substituent position the sequence should be modified accordingly.

In Table III the new substituent selection listed for the designation "other" under probable operative parameters requires comment. These are indicated for investigation if the analysis, through the phase of examination of 2 substitution, has not yielded compounds of substantially increased potency. The 4-F analogue, which provides minimal change in π and σ effects compared to the unsubstituted compound, should prove advantageous in the event that the initial analysis indicates the unsubstituted compound may be essentially optimal in terms of π and σ but subject to rather rapid metabolic transformation by 4-hydroxylation. The remaining listed substituents, 4-$NHCOCH_3$, 4-$NHSO_2CH_3$, 4-NO_2, 4-$COCH_3$, 4-SO_2CH_3, 4-$CONH_2$, and 4-SO_2NH_2, are all examples of $-\pi + \sigma$ type substituents which should prove fruitful if increased potency is related to reduced lipophilicity or reduced lipophilicity combined with a $+\sigma$ effect. Alternatively, these substituents may be employed in the 3 position. Cases benefiting from a $-\pi - \sigma$ effect should have been detected through the 4-NH_2 and 4-OH substituents covered earlier.

The foregoing examples amply demonstrate that a manual approach to Hansch analysis can be very useful in efficiently identifying compounds which have a high probability of enhanced potency in situations involving substituent variation on a phenyl ring. The examples considered embrace many different compound series,

biological activities, and various types of parameter dependencies. For space reasons other examples where the method could be usefully applied have not been included. In addition to these, some other series were found in the literature where the activity spread between the members of the initial compound group was too small to permit a meaningful analysis.

A key feature is that, based on the results from only four or five readily available analogues, the correct synthetic direction for increased potency can often be determined. At this stage the parameter dependency can usually be narrowed to a small range of possibilities and further substituents can be chosen which should increase potency no matter what the precise activity–parameter relationship is. This marks an important difference between the strategy outlined in the present approach and that of the standard Hansch method. In the latter method the object is to first determine, utilizing a computer based analysis of the results on eight to twelve compounds, a precise activity–parameter relationship in the form of an equation. This equation is then used to select new analogues which should have improved potency. On the other hand, the manual method does not attempt to precisely identify the activity–parameter relationship but seeks to use a more rapidly obtained approximate determination of this relationship as a stepping stone to the identification of more potent analogues.

In terms of numbers of compounds prepared the position reached after preparation of the second compound group in the manual method is roughly equivalent to that reached after a multiple regression analysis on the first group of compounds made in the standard Hansch analysis.[24] Thus, to the extent that the present manual method can successfully narrow the possible operative parameter dependencies at the end of the first stage, it may represent a more advantageous strategy if the primary goal is to find a readily accessible compound in the maximum potency area in the shortest possible time rather than to determine the exact activity–parameter relationship. Also, computers and statistical procedures are not required thus offering greater simplicity of use for most medicinal chemists.

References and Notes

(1) Presented in part at the 167th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1–5, 1974, and the Fifth International Symposium on Medicinal Chemistry, Paris, July 19–22, 1976.
(2) J. G. Topliss, *J. Med. Chem.*, **15**, 1006 (1972).
(3) C. E. Granito, G. T. Becker, S. Roberts, W. J. Wiswesser, and K. J. Windlinz, *J. Chem. Doc.*, 11 (2), 106 (1971).
(4) P. J. Goodford, *Adv. Pharmacol. Chemother.*, **11**, 51 (1973).
(5) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(6) C. Hansch, "Drug Design", Vol. I, E. J. Ariens, Ed., Academic Press, New York, N.Y., 1971, p 271.
(7) P. N. Craig, *J. Med. Chem.*, **14**, 680 (1971).
(8) C. Hansch, *Cancer Chemother. Rep.*, **56**, 433 (1972).
(9) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, **16**, 1217 (1973).
(10) F. Darvas, *J. Med. Chem.*, **17**, 799 (1974).
(11) R. Wootton, R. Cranfield, G. C. Shappey, and P. J. Goodford, *J. Med. Chem.*, **18**, 607 (1975).
(12) R. T. Buckler, S. Hayao, O. J. Lorenzetti, L. F. Sancilio, H. E. Hartzler, and W. G. Strycker, *J. Med. Chem.*, **13**, 725 (1970).
(13) R. W. Fuller, J. Mills, and M. M. Marsh, *J. Med. Chem.*, 14, 322 (1971).
(14) N. Kakeya, N. Yata, A. Kamada, and M. Aoki, *Chem. Pharm. Bull.*, 17, 2558 (1969).
(15) A. Cammarata, R. C. Allen, J. K. Seydel, and E. Wempe, *J. Pharm. Sci.*, **59**, 1496 (1970).
(16) C. Silipo and C. Hansch, *Farmaco, Ed. Sci.*, **30**, 35 (1974).
(17) C. Hansch and E. W. Deutsch, *Biochim. Biophys. Acta*, **112**, 381 (1966).
(18) C. Silipo and C. Hansch, *Mol. Pharmacol.*, **10**, 954 (1974).
(19) B. Blank, N. W. Di Tullio, L. Deviney, J. T. Roberts, and H. L. Saunders, *J. Med. Chem.*, **18**, 952 (1975).
(20) C. Hansch, K. H. Kim, and R. H. Sarma, *J. Am. Chem. Soc.*, **95**, 6447 (1973).
(21) M. Yoshimoto, K. N. von Kaulla, and C. Hansch, *J. Med. Chem.*, **18**, 950 (1975).
(22) G. van den Berg, T. Bultsma, R. F. Rekker, and W. T. Nauta, *Eur. J. Med. Chem.*, **10**, 242 (1975).
(23) B. J. Broughton, P. Chaplen, P. Knowles, E. Lunt, S. M. Marshall, D. L. Pain, and K. R. H. Wooldridge, *J. Med. Chem.*, **18**, 1117 (1975).
(24) The option exists at this point to transpose into the standard Hansch approach, although the compound mix may not be optimal.

A Statistical-Heuristic Method for Automated Selection of Drugs for Screening

Louis Hodes,* George F. Hazard, Ruth I. Geran, and Sidney Richman

Division of Cancer Treatment, National Cancer Institute, Silver Spring, Maryland 20910.

Received June 30, 1976

A statistical-heuristic method for selecting drugs for animal screening is developed with molecular structure features as predictors of biological activity. The method is intended to work on large amounts of data over varied structures. A trial of this method on a small data set allows some comparison with more sophisticated pattern recognition methods. Problems connected with interdependence among structure predictors are critical in this method and schemes to eliminate redundancy are reviewed. Alternate sets of structure predictors are considered. The discussion here outlines directions to be taken in the near future.

A major activity of the Developmental Therapeutics Program (DTP) in the Division of Cancer Treatment (DCT), National Cancer Institute (NCI), is the development of new drugs useful in the treatment of human cancer. As one means of identifying leads to such drugs, DTP, which subsumed the Drug Research and Development Program (DR&DP), operates an antitumor screening program that involves the testing of compounds in a variety of animal tumor models. Because of the limited capacity for screening, currently roughly 15 000 synthetic compounds per year, and the almost limitless possibilities for obtaining compounds, many approaches to selecting acquisitions or assigning priorities of testing are being explored.

Some of these approaches involve the use of biological test data from previous acquisitions and chemical structure data to create a system for predicting the biological activity of a new compound by examining its chemical structure.[1,2] Chemical structural parts are obvious choices for prediction parameters because of their clear pertinence and easy availability. At NCI automated files currently contain chemical structure on more than 280 000 compounds and