Nevertheless, the authors believe that this is the first data bank published which covers a reasonable spread of different substituent groups whose parameters are reasonably uncorrelated with each other (Table III) and, furthermore, which has no missing values. The user is not, therefore, forced to choose between omitting either a compound or a parameter when a value is missing, or alternatively of devising a crude and possibly incorrect substitute value.

Experimental Section

Measurement of Partition Coefficients. Partition coefficients were determined between 1-octanol and water and, in the case of ionizable compounds, the aqueous phase was brought to a pH at which the ionization was suppressed to <1%. The 1-octanol (Koch Light, pure) and aqueous phase were mutually saturated before use.

The test compounds were dissolved in the aqueous phase to yield a solution which could be estimated spectrophotometrically, using a Unicam SP800 spectrophotometer. A uv peak having an absorption of 0.5–1.0 was commonly used.

Every partition coefficient was measured using two volume ratios and two concentrations. Each was done in triplicate, giving a total of 12 determinations per compound. The solutions were gently shaken at room temperature at ≃60 strokes per minute. Two of each set of three were shaken for 2 days while the third was left for an extra day to test if equilibrium had been attained. At the end of this time a sample of the aqueous phase was removed, centrifuged for 1 hr at 1000 g, and again estimated spectrophotometrically.

In each case π was calculated as $\pi = \log P_x - \log P_H$ where P_x is the partition coefficient of the test compound and P_H is the partition coefficient of the corresponding unsubstituted compound.[25]

The model compounds used in measuring π values were as follows: *p*-CHO, phenoxyacetic acid; *p*-CF_3, phenylacetic acid; *o*-OPh, NHAc, *p*-OPh, NHAc, SMe, SO_2Me, benzoic acid; *o*-OAc, NMe_2, CHO, *m*-OAc, NMe_2, *p*-OAc, NMe_2, SO_2NH_2, toluene. The model compounds used in measuring π_- values were: *o*-OMe, NHAc, CN, *m*-Ph, OAc, NHAc, CHO, *p*-Ph, NHAc, CHO, SMe, phenol; *o*-OPh, OAc, NMe_2, *p*-NMe_2, anisole; *o*-SMe, *m*-OPh, SMe, SO_2NH_2, *p*-CF_3, OPh, OAc, SO_2Me, SO_2NH_2, aniline.

Acknowledgment. The authors wish to thank Mr. M. Lillis and Mrs. N. Trist for valuable technical assistance in the preparation of this work.

References and Notes

(1) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(2) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, **16**, 1207 (1973).
(3) M. S. Tute, *Adv. Drug Res.*, **6**, 1 (1971).
(4) P. J. Goodford, *Adv. Pharmacol. Chemother.*, **11**, 51 (1973).
(5) R. M. Hyde, F. E. Norrington, S. G. Williams, and R. Wootton, manuscript in preparation.
(6) P. N. Craig, *J. Med. Chem.*, **14**, 680 (1971).
(7) T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964).
(8) C. Hansch and W. J. Dunn, *J. Pharm. Sci.*, **61**, 1 (1972).
(9) C. Hansch and J. M. Clayton, *J. Pharm. Sci.*, **62**, 1 (1972).
(10) A. E. Bird and A. C. Marshall, *Biochem. Pharmacol.*, **16**, 2275 (1967).
(11) L. P. Hammett, *Chem. Rev.*, **17**, 125 (1935).
(12) R. Taft, "Steric Effects in Organic Chemistry", M. S. Newman, Ed., Wiley, New York, N.Y., 1956.
(13) C. G. Swain and E. C. Lupton, *J. Am. Chem. Soc.*, **90**, 4328 (1968).
(14) S. G. Williams and F. E. Norrington, in press.
(15) f_{para} and r_{para} are both defined as 1.000; $f_{meta} = 0.980$; $r_{meta} = 0.347$; $f_{ortho} = 1.248$; $r_{ortho} = 0.863$.
(16) (a) A. Cammarata, *Annu. Rep. Med. Chem.*, 245 (1971); (b) E. Kutter and C. Hansch, *J. Med. Chem.*, **12**, 647 (1969).
(17) R. W. Fuller, M. M. Marsh, and J. Mills, *J. Med. Chem.*, **11**, 397 (1968).
(18) A. Leo, C. Hansch, and C. Church, *J. Med. Chem.*, **12**, 766 (1969).
(19) J. C. McGowan, *J. Appl. Chem.*, **4**, 41 (1954).
(20) A. I. Vogel, *J. Chem. Soc.*, 607 (1948), and following papers.
(21) A. I. Vogel, *J. Chem. Soc.*, 1809 (1948), and following papers.
(22) D. Gilbert, P. J. Goodford, F. E. Norrington, B. C. Weatherley, and S. G. Williams, in press.
(23) G. Barlin and D. D. Perrin, *Q. Rev., Chem. Soc.*, **20**, 75 (1966).
(24) J. Clark and D. D. Perrin, *Q. Rev., Chem. Soc.*, **18**, 295 (1964).
(25) A. Leo, C. Hansch, and D. Elkins, *Chem. Rev.*, **71**, 525 (1971).

Physicochemical–Activity Relationships in Practice. 2. Rational Selection of Benzenoid Substituents

Raymond Wootton,* Robert Cranfield, G. Clive Sheppey, and Peter J. Goodford

Wellcome Research Laboratories, Langley Court, Beckenham, Kent, BR3 3BS, England. *Received November 11, 1974*

A rational method is presented for the selection of substituents to be introduced into a benzenoid ring system of a biologically active compound in order to explore a defined physicochemical parameter space. The method, which may be readily programmed for use on a computer, relies on maintaining a minimum distance between compounds in the multidimensional physicochemical parameter space. The series of compounds produced will then have a well-spread set of minimally correlated physicochemical parameter values and could thus be used for the reliable correlation of the variation in the biological activities of the members of the series with changes in these physicochemical parameters. Some examples of the use of the present method under various conditions are given, and it is compared with alternatives in the literature.

Many papers involving the use of the method of physicochemical–activity relationships (the PAR method) have been published,[1,2] and it is now well established as an aid to the design of biologically active molecules.[3,4] Unfortunately, the method has often been applied to series of compounds which are far from ideal for the purpose. This is usually because the method was applied as an afterthought, and the compounds in the series were not chosen at the onset with the aim of testing rigorously the hypotheses implicit in the method. In these cases the range of values for some of the physicochemical parameters may be small, and the chance of finding a real relationship between activity and these parameters will therefore be correspondingly small. Also, correlations may exist between the physicochemical parameters under test, leading to problems in distinguishing which parameters are truly correlated with activity.[5] These factors often give rise to regression equations which are inadequate for the prime objective of predicting which new members of the series will have higher biological activities. Before the PAR method is applied predictively it

is therefore essential that the range and spread of physicochemical parameter values for the compounds be ascertained and a check of interparameter correlations be made. It may then be necessary to synthesize more compounds in order to improve the range and spread of parameter values, or to decrease correlations, or both.

In an ideal case, however, the principles of optimally exploring the available physicochemical parameter space and minimizing interparameter correlations should be applied as soon as a lead molecule to a new series has been uncovered. This then provides the basis for a rational approach to the problem of which compounds to synthesize in a series in order to predict more active compounds with the minimum of effort and cost. The PAR approach and the classical structure–activity relationships (SAR) approach are complimentary as both require changes in substituent type derived from the same lead molecule.

Topliss[6] has suggested that there are two limiting cases to be considered when embarking on a synthetic program to enhance the activity of a novel compound. The limiting factors involved are the relative times taken to synthesize compounds and to test them for biological effect. When the biological test is rapid compared with chemical synthesis, the chemist may progress stepwise to the region of physicochemical parameter space corresponding to high activity, as each synthesis may be guided by the results for the previous compounds. Topliss[6] has attempted to formalize this case by tabulating the actual substituents to be introduced into the aromatic ring system. If the biological test is slow, however, the chemist may synthesize several compounds in a series before receiving any biological data on which to base further synthetic strategy. Hansch[7] has discussed the factors involved in choosing compounds for synthesis in this "batch" mode and, more recently,[8] has published a method of choosing a set of monosubstituted compounds to explore a defined physicochemical parameter space using a technique known as cluster analysis.[9,10] This method involves the construction of a specified number of subsets of substituents from the total substituents for which the relevant physicochemical parameter values are available, each subset containing those substituents with the most similar parameter values. Selection of one substituent for synthesis from each subset should then result in a series of compounds with a good range and spread for all parameter values and with sets of parameter values which are essentially independent of each other. Unfortunately, problems may arise in the use of this method when synthetic considerations are taken into account as it may not be practicable to synthesize a compound containing any substituents from one or more of the subsets.

In this paper an alternative method is presented for choosing batches of compounds for synthesis, which should result in a more ideal set of monosubstituted compounds, tailored within the limits of synthetic feasibility for any given series. This method may also be readily applied to the case of multiple substitution of the aromatic ring and to the case where several compounds already exist in a particular series which were not chosen with the present criteria in mind and are therefore "poorly distributed". Cluster analysis does not provide a ready answer to the problem of which compounds to synthesize in these cases.

Method. In order to choose a series of compounds with a well-spread set of physicochemical properties it is first necessary to compile a data bank containing the values of these properties for individual substituent groups. This data bank then forms a basis for the definition of the extent of available physicochemical parameter space and provides the limits with which to assess how well the selected compounds fill the total space available. Hansch et al.[11] have now provided complete data for 100 substituents in the parameters π (log relative partition coefficient), σ_m and σ_p (Hammett σ constants for meta- and para-substituted benzoic acids, respectively[12]), $\mathcal{F}$ and $\mathcal{R}$ ("field" and "resonance" components of Hammett's σ derived according to Swain and Lupton[13]), and MR (molar refraction) and MW (molecular weight) of the substituent. The positional dependency of π and $\mathcal{F}$ and $\mathcal{R}$ values was not taken into account in this work. The present study has been limited to 35 substituents (including hydrogen) selected largely on the basis of ease of introduction into a benzene ring by conventional synthetic methods. Positionally dependent π and F and R values and positionally independent MR values have been collated for all 35 substituents.[14]

Thus, for substitution of any one position, the selection of compounds must be made from 35 substituents as opposed to the 90 substituents considered by Hansch, Unger, and Forsythe[8] for the cluster analysis approach. However, with positionally dependent parameter values the much larger number of compounds derived from multiple substitution of the benzene ring may also be dealt with, the parameter values for these compounds being calculated by summing the values for the individual substituents.[4] Because of this the range of possible parameter values is increased and the available physicochemical parameter space becomes larger.

The method to be described here relies on maintaining a preset minimum distance (D) between compounds in the defined multidimensional parameter space. The position of each compound in space is determined by the values of the physicochemical parameters for that compound; i.e., these parameters define the Cartesian coordinates of each compound in the space. Distances between compounds may therefore be calculated as the simple, multidimensional Euclidean distances.[8] However, since the range of values for each physicochemical parameter is different, it is necessary to scale the individual values to lie in equivalent ranges in order to ensure that each parameter is given an equal weighting in the calculation of the intercompound distances. This may be achieved using eq 1 where x_{ik} represents the kth value of the parameter x_i, $\bar{x}_i$ is the mean of the highest and lowest possible values of x_i, and r_i is the range of x_i values. With this procedure the scaled values, x'_{ik}, will lie in the range ± 0.5. A computer may be used with advantage to perform the required calculations as these are considerable with the large numbers of compounds involved.

$$x'_{ik} = (x_{ik} - \bar{x}_i)/r_i \quad (1)$$

To apply the selection procedure in order to select a set of compounds, it is first necessary to choose a starting compound defining a starting point in the scaled multidimensional parameter space. This point would often be the coordinates representing hydrogen, corresponding to the unsubstituted member of the series, but could be the coordinates for any other compound. The distances in space from this starting compound to all other theoretical compounds are then computed, and the second compound for synthesis is proposed as the one closest in space to the starting compound, yet greater than the preset minimum distance D from it. In this way all compounds which are closer to the starting compound than this minimum distance D, i.e., compounds possessing similar predicted physicochemical parameter values, are rejected. The distances of all remaining compounds from the second compound are then computed, and compounds within the minimum distance D from the second compound are also rejected. The

third compound is proposed as the one closest to the center of gravity in space of the two selected compounds, yet greater than the minimum distance D from each of them. The process is continued, rejecting compounds which lie too close in space to selected compounds, and choosing further compounds according to their distance from the center of gravity in space of the selected compounds. Eventually there are no compounds left to choose from, space may be said to be "filled", and the selection process is complete.

The net result is the selection of a subset of compounds from the total of theoretical compounds with the subset possessing a well-spread set of physicochemical parameter values. Also, the method tends to select compounds as far as possible to form a regular array in the physicochemical parameter space leading to small interparameter correlation coefficients. The number of compounds selected is determined by the value of the minimum distance D between compounds and this may be adjusted to give the required number of compounds for any particular series.

The method so far described is in fact an ideal case and forms the basis of a fully automatic selection procedure as there is no doubt at any stage which compound to choose next. In practice it may be that some of the compounds chosen in this way are synthetically impractical or otherwise undesirable for a given series of compounds, and so it becomes necessary to allow some freedom of choice in the selection. This is easily accommodated in an interactive computer program. Now, instead of automatically selecting at each stage the compound nearest the center of gravity of previously selected compounds, all compounds greater than the minimum distance D from those previously selected may be listed out in order of their distance from this center of gravity. Any one of these may then be chosen and included in the set, but for even space filling it is best to choose the first compound on the list that is synthetically practical each time. The use of this interactive approach will then result in a set of compounds, tailored within the limits of synthetic feasibility to a given series. Moreover, provided sufficient compounds are chosen and a realistic minimum distance D between compounds is maintained, a well-spread set of minimally correlated physicochemical parameter values should still result. In practice, the selection procedure can be stopped or started at any number of compounds and this provides a means of improving the spread of parameter values for the case where several compounds already exist in a series but they have been poorly chosen from a space-filling point of view.

Results and Discussion

Before selecting compounds for synthesis it is important to decide how many biological data points (compounds) will be required to enable a reliable regression analysis to be performed when the compounds have been made and tested. The reliability of the analysis will depend upon the number of degrees of freedom for the regression which is determined by the number of data points and the number of parameters under test; the greater the number of parameters tested the greater the number of data points needed. Topliss and Costello[15] have laid down some guidelines concerning this problem which suggest that quite high numbers of data points are needed to reduce chance correlations to acceptable levels. However, this statistical requirement for large numbers of compounds must be offset at the practical level against the synthetic effort required to produce the compounds and some compromise is necessary. In the examples given in this paper, the numbers of compounds selected are regarded as the absolute minimum numbers required to allow regression analyses to be performed. In general, the reliability of the regressions will be improved considerably by the inclusion of extra data points obtained for further compounds.

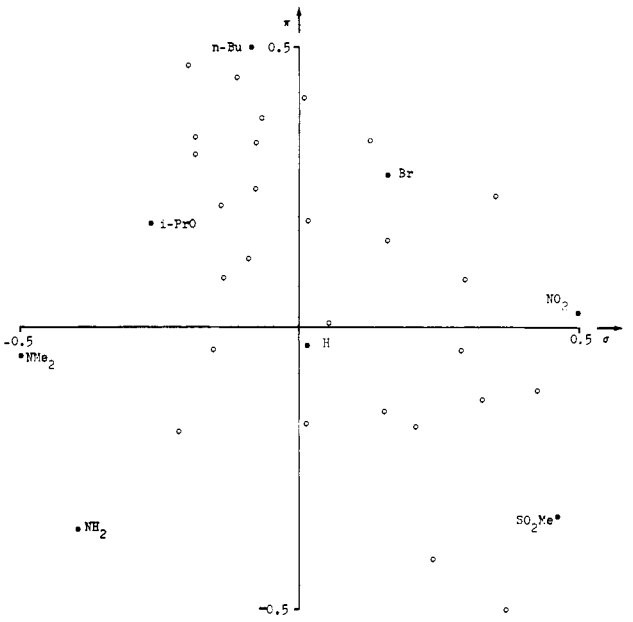

Figure 1. The π and σ values for 35 substituents scaled according to eq 1. The eight selected substituents are indentified by name and shown as solid circles.

It is recommended that the four physicochemical parameters π, F, R, and MR, which provide a description of the hydrophobic, electronic, and steric effects of substituents, should be tested as a standard set in any PAR exercise.[16] However, the synthesis of a sufficient number of compounds to test these four parameters (or five if π^2 is included) may not always be a practical proposition. It may therefore be more prudent, if limited synthetic resources are available, to consider only the parameters π and σ in an initial exploration of a series. In fact, most PAR studies in the past have employed the parameters π and σ and many examples of activity correlations with these parameters have been found. In this case the requirement is for a series of compounds with a well-spread set of uncorrelated π and σ values representing only a two-dimensional parameter space which may be readily visualized in the form of a simple π/σ plot.

Monosubstitution. Initially, to avoid problems of synthetic feasibility arising from multiple substitution, an example of the selection of compounds in a π/σ space for substitution at one position only will be described. Multiple substitution will be dealt with later in the discussion. The π and σ values[14] for 35 para substituents, scaled according to eq 1, are plotted in Figure 1. To select a subset of these substituents with a well-spread set of uncorrelated π and σ values from the 35 available does not require a computer as it is an easy matter to select such a subset by eye. However, this example forms a good illustration of the principles of the present method, and the substituents chosen using the method are also shown in Figure 1.

Hydrogen was chosen as starting substituent and the other substituents were selected in turn, maintaining a minimum distance of 0.3 units in space between substituents. This distance led to the series of eight substituents shown by solid circles in Figure 1 which form a well-spread set in the π/σ parameter space with a low π/σ correlation coefficient (−0.05). Thus, if these eight para substituents could be used to synthesize a series of biologically active compounds, then a well-spread set of uncorrelated π and σ

Table I. Twelve Para Substituents with a Well-Spread Set of Physicochemical Parameter Values

Compd no.	Substituent	Physicochemical properties			
		π	F	R	MR
1	H	0.00	0.00	0.00	0.0
2	NH_2	−1.30	0.04	−0.68	4.2
3	NMe_2	−0.08	0.03	−0.85	14.4
4	NO_2	0.22	1.11	0.16	6.0
5	Br	1.19	0.73	−0.18	7.6
6	SO_2Me	−1.20	0.90	0.22	12.5
7	n-Bu	2.10	−0.06	−0.13	18.7
8	i-PrO	0.85	0.49	−0.72	16.0
9	CO_2Et	0.46	0.55	0.14	16.2
10	F	0.15	0.71	−0.34	−0.4
11	NHCOMe	−0.56	0.47	−0.27	14.6
12	n-AmO	1.97	0.42	−0.58	25.3

Table II. Correlation Matrix for the Physicochemical Parameters of the 12 Selected Para Substituents of Table I

	π	F	R
F	−0.11		
R	−0.08	0.49	
MR	0.51	−0.12	−0.21

parameter values would be assured. The eight compounds produced should then provide sufficient information for a regression analysis of their biological activity against π and σ with five degrees of freedom. In favorable circumstances, particularly where one or two extra compounds were available, perhaps made as synthetic intermediates in the synthesis of the eight specific compounds, it might also be possible to test for a parabolic dependence on π by introducing an extra term in π^2.

On finding a relationship with π and/or σ in the above circumstances then the next stage of the analysis might be to synthesize a few more compounds in order to test the derived regression equation. At the same time it would be possible to explore the effects of changes in the four parameters π, F, R, and MR by selecting further compounds to be well spread in this four-dimensional space. Alternatively, if synthetic resources were more freely available, then the exploration of this four-dimensional space could be treated as the first step in the analysis. Again it may be best initially to consider substitution at one position only, thus avoiding some of the problems of dealing with multiple substitution, such as the much larger numbers of compounds involved and the deviations from additivity of parameter values sometimes found.[11]

The eight substituents selected to produce a good range of π and σ parameter values also provide a reasonable range of F, R, and MR values. This is not surprising due to the relationship of σ with F and R and the general correlation observed between π and MR. These eight substituents were therefore used as a starting set for a selection by computer in the four-dimensional π, F, R, and MR parameter space. An extra four para substituents were chosen to "fill" space more completely by setting the minimum distance D between substituents at 0.42 units. The resulting series of 12 substituents is shown in Table I together with the relevant physicochemical parameter values. The distribution of parameter values along each parameter axis, scaled according to eq 1, is illustrated in Figure 2 and Table II gives an interparameter correlation matrix.

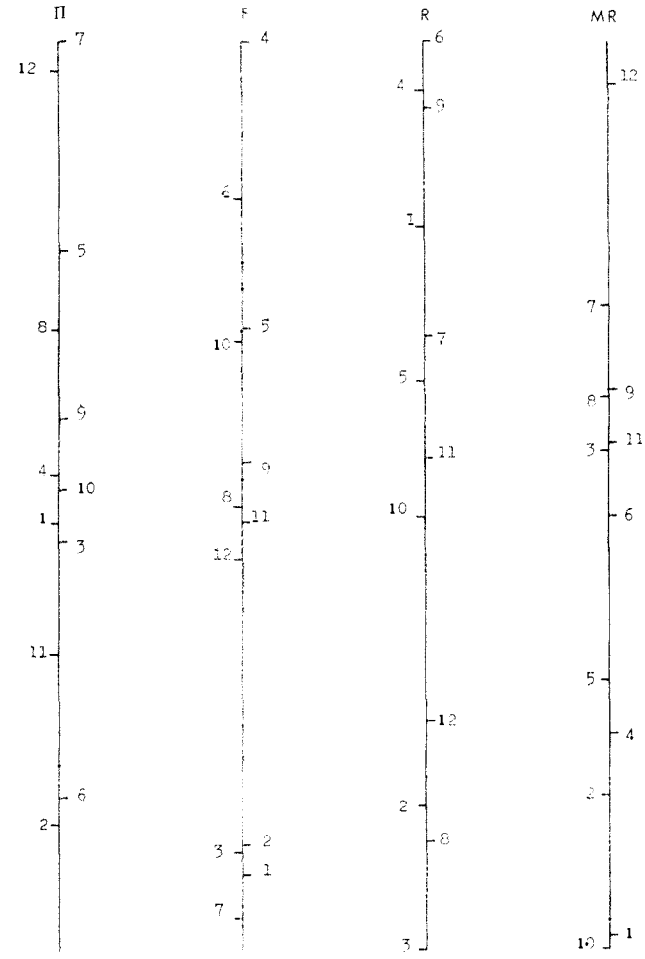

Figure 2. Representation of the spread of physicochemical parameter values for the 12 selected substituents of Table I. Each parameter axis has a range of −0.5 to +0.5 units which corresponds to the maximum ranges possible for the 35 substitutents of the data bank. The parameter values for the 12 selected substituents are scaled within these ranges using eq 1.

It may be seen that the 12 substituents provide a very good range and spread of parameter values along each parameter axis with no seriously high interparameter correlation coefficients. In this case, if these substituents could be used to produce a series of biologically active compounds, then a well-conditioned regression of the variation in activity against the parameters π, F, R, and MR could be per-

Table III. Substitution Patterns and Predicted Physicochemical Parameters for 15 Well-Spread Commercially Available[a] Benzaldehydes

Compd no.	Substituents				Predicted physicochemical parameters				
	R_2	R_3	R_4	R_5	π	F	R	MR_o	$MR_{m,p}$
1					0.00	0.00	0.00	0.0	0.0
2	Br				0.84	0.91	−0.15	7.6	0.0
3			CN		−0.33	0.85	0.18	0.0	5.2
4			NHCOMe		−0.56	0.47	−0.27	0.0	14.6
5			NMe_2		−0.08	0.03	−0.85	0.0	14.4
6			Ph		1.74	0.14	−0.09	0.0	24.3
7	Cl		NMe_2		0.68	0.89	−0.99	4.8	14.4
8		Me	Me		1.12	−0.10	−0.19	0.0	9.4
9		MeO	*n*-AmO		2.09	0.83	−0.75	0.0	31.8
10		MeO	MeO		0.09	0.82	−0.67	0.0	13.0
11		NO_2	Cl		0.84	1.78	−0.11	0.0	10.8
12	MeO		MeO	MeO	−0.24	1.34	−1.10	6.5	13.0
13	EtO	Br		Br	2.09	1.87	−0.50	11.3	15.2
14	EtO		EtO	EtO	1.26	1.17	−0.97	11.3	22.6
15		MeO	MeO	Br	1.05	1.53	−0.73	0.0	20.6

[a] Compounds available from Ralph N. Emanuel Ltd.

formed. Again in favorable circumstances, with a few additional compounds it might also be possible to include a term in π^2. The point should be made at this stage, however, that the 12 para substituents of Table I do not form the only possible series and may not be the best. Alternative series with well-spread sets of minimally correlated parameter values could be chosen to take account of the specific synthetic limitations pertaining in any particular compound type.

Multiple Substitution. The two examples of the selection process presented so far have been limited to substitution at one position only. In general, it will usually be possible to introduce substituents at more than one position in an aromatic nucleus. The selection of suitable substituents in this case becomes a far greater practical problem due to the much larger number of theoretical compounds involved and the fact that synthetic feasibility problems will usually be greater. A rational method of choosing substituents is therefore very valuable, and the principle of maintaining a minimum distance between compounds in the multidimensional parameter space is again applicable.

Two alternative strategies have been devised. In the first it is necessary to define the extent of physicochemical parameter space by deciding on the maximum number of substituents to be used for any compound. When this is done the predicted properties of all theoretically possible compounds derived from the 35 substituents of the data bank[14] may be generated by computer and lists of compounds produced for the chemist to choose from as described in the methods section. In practice it has been found convenient to limit substitution to a maximum of three substituents and two different substituent types in any one compound, thereby restricting the theoretical number of compounds to manageable levels. Even with this restriction, however, it is unlikely that a series of compounds with parameter values covering the whole of space could ever be synthesized as many of the compounds required may be synthetically impractical. Moreover, it is not an easy matter to examine a long list of theoretical compounds and decide which will be the easiest to make.

Because of these factors, an alternative strategy in which synthetic feasibility is considered first may be adopted. In this case the chemist concerned would compile a list of compounds considered to be synthetically feasible. These could then be tested in turn and accepted for synthesis or rejected depending on their distances in space from compounds already included in the synthetic program. In this way a long list of possible compounds may be reduced to a subset possessing a well-spread set of physicochemical parameters with respect to the parameter space defined by the compounds in the original list. A compromise must be reached between compounds desirable from a space-filling point of view and those that are easiest to make, but it must not be forgotten that the chances of establishing a valid regression equation will depend upon the range and spread of parameter values for the chosen compounds. It is therefore extremely important to maintain a realistic minimum distance D between compounds in space as the greater this distance the better the range and spread of parameter values for the resulting series of compounds.

In deciding which compounds to make in a particular series, the chemist is often guided by the availability of appropriate commercially available intermediate compounds. The substitution patterns attainable in the final compounds may depend to a large extent on the substitution patterns in these intermediates. Thus, if a list of commercially available intermediates could be compiled for a given series of compounds, these could be tested, using the method described, to find the best ones from a space-filling point of view. Until recently, the searching of chemical catalogs for the required compounds was a laborious and often haphazard process, but it is now possible to obtain specific searches of this kind carried out using computers.

A computer-generated list of commercially available substituted benzaldehydes has been provided by the Aldrich Chemical Co. Inspection of this list revealed 94 compounds for which the relevant π, F, R, and MR parameter values were available, and, using the computer program described here, it was possible to select from this list a subset of compounds providing a well-spread set of parameter values within the parameter space defined by the 94 compounds. For chemical reasons compounds containing hydroxyl or second formyl function, compounds with two ortho substituents, and compounds with *o*-nitro groups were excluded from selection. Also, many of the benzaldehydes contained ortho substituents and, since ortho substituents may give rise to special intramolecular steric effects, the parameter MR was in this case subdivided into MR at the ortho posi-

Table IV. Correlation Matrix for the Predicted Physicochemical Parameters of the Selected Commercially Available Benzaldehydes of Table III

	π	F	R	MR_o
F	0.27			
R	−0.08	−0.28		
MR_o	0.31	0.51	−0.39	
$MR_{m,p}$	0.56	0.16	−0.52	0.00

tion, MR_o, and a combined MR at the meta and para positions, $MR_{m,p}$.

Fifteen compounds were therefore chosen to allow a regression analysis of the biological activities of the compounds derived from the benzaldehyde intermediates against the parameters π, π^2, F, R, MR_o, and $MR_{m,p}$ with eight degrees of freedom. The selected benzaldehydes are shown in Table III together with their physicochemical parameter values. The range of parameter values is better in F, R, and $MR_{m,p}$ than could be achieved by monosubstitution only, but is not quite so good in π and MR_o. Table IV shows that there are no serious interparameter correlations for this series of 15 compounds. Thus, if these 15 benzaldehydes could be used as synthetic intermediates for the synthesis of a series of biologically active compounds, then a not unreasonable spread of minimally correlated parameter values would be achieved.

In practice, however, it would probably be possible to design a better series of compounds in a specific case by interfacing the particular synthetic problems for a given series of compounds with the computer program. In addition, the use of a common synthetic intermediate such as a substituted benzaldehyde may not always be possible throughout the synthesis of a series of compounds due to steric or electronic effects. In this event, alternative or complimentary lists of well-spread sets of other commercially available substituted intermediates such as benzoic acids or phenols could be produced in a similar manner.

Conclusion

The fundamental hypothesis of the PAR method is that the quantitative differences in biological activity of the compounds in a series of substituted aromatics may be largely explained in terms of the effect of the substituents on the physicochemical properties of the molecules as a whole. It is not possible at this time to obtain quantitative, predictive measures for all the relevant physicochemical properties, but the four parameters π, F, R, and MR have been thought to describe some of the most important general effects of substituents, i.e., hydrophobic, electronic, and steric.[11] However, such factors as directional dipolar effects, hydrogen bonding ability, and sensitivity to metabolism may also play a role in determining the activity for a particular compound but are not considered here.

The method of selecting substituents described here and the cluster analysis approach of Hansch, Unger, and Forsythe[8] both stress the need to provide a good range of uncorrelated parameter values in order to optimize the chances of determining a relationship. By ensuring a good range of parameter values it is hoped that a relationship will emerge, if one exists, essentially by smoothing out the individual deviations from the regression line due to effects not described by the chosen parameters. On the other hand, the method presented by Topliss,[6] involving a stepwise synthesis of compounds, relies heavily on the results for each individual compound as each synthesis is guided by the activity for the previous compound. Thus, although this method has been demonstrated to work satisfactorily in a number of cases in a retrospective manner,[6,17] its prospective use must be viewed with some caution as a compound showing a large deviation from the average behavior for the whole series could seriously disrupt the analysis. Also the method is limited, being designed to deal primarily with the two parameters π and σ (with steric effects brought in to explain anomolies) and to differentiate the effects due to these two parameters. It would be very difficult to use the method to consider the effects of changes in the fuller parameter set of π, F, R, and MR as for this case there are too many variables to be sorted out without the aid of a computer. The method described here and the cluster analysis approach[8] seem, therefore, to provide a more logical and complete answer to the problem of which compounds to synthesize in a series in order to delineate as soon as possible a relationship between biological activity and physicochemical parameters.

Of these two approaches, the method described here provides for the most general case as the simple criterion of maintaining a minimum distance in space between compounds may be used to select compounds in any defined set of circumstances, whether for substitution at one position only or for multiple substitution. It is also possible to consider problems of synthetic feasibility during the actual selection process leading to the selection of series of compounds, tailored to the synthetic requirements pertaining to a given series, yet still possessing a well-spread, minimally correlated set of physicochemical parameter values. The computer program required to aid the selection procedure can be simple or complex depending on the degree of sophistication required with respect to the interaction between chemist and computer when problems of synthetic feasibility are considered. In fact, in individual cases it may be possible to introduce into the program certain simple algorithms to limit compounds from a synthetic feasibility viewpoint, but the complete computer description of synthetic feasibility is obviously not possible at this time. It is encouraging, however, that with the power and availability of modern digital computers[18] it is now easy to consider problems such as this, involving multidimensional spaces and millions of theoretical compounds.

Acknowledgments. The authors wish to thank Mr. M. H. Gimblett and Miss L. Rahr for computational assistance and many of their colleagues at the Wellcome Research Laboratories, in particular, R. M. Hyde, F. E. Norrington, G. Sutherland, B. Weatherley, and S. G. Williams, for their much valued help and discussions.

References and Notes

(1) C. Hansch and T. Fujita, *J. Am. Chem. Soc.*, **86**, 1616 (1964).
(2) C. Hansch in "Drug Design", Vol. 1, E. J. Ariens, Ed., Academic Press, New York, N.Y., 1971, p 271.
(3) R. F. Gould, *Adv. Chem. Ser.*, **No. 114** (1972).
(4) P. J. Goodford, *Adv. Pharmacol. Chemother.*, 11, 51 (1973).
(5) P. N. Craig, *J. Med. Chem.*, 14, 680 (1971).
(6) J. G. Topliss, *J. Med. Chem.*, 15, 1006 (1972).
(7) C. Hansch, *Cancer Chemother. Rep.*, **56**, 433 (1972).
(8) C. Hansch, S. H. Unger, and A. B. Forsythe, *J. Med. Chem.*, 16, 1217 (1973).
(9) R. C. Tryon and D. E. Bailey, "Cluster Analysis", McGraw-Hill, New York, N.Y., 1970.
(10) B. R. Kowalski and C. F. Bender, *J. Am. Chem. Soc.*, **95**, 686 (1973).
(11) C. Hansch, A. Leo, S. H. Unger, K. H. Kim, D. Nikaitani, and E. J. Lien, *J. Med. Chem.*, 16, 1207 (1973).
(12) L. P. Hammett in "Physical Organic Chemistry", 2nd ed, McGraw-Hill, New York, N.Y., 1940, Chapter 7.
(13) C. G. Swain and E. C. Lupton, Jr., *J. Am. Chem. Soc.*, **90**, 4328 (1968).

(14) F. E. Norrington, R. M. Hyde, S. G. Williams, and R. Wootton, preceding paper in this issue.
(15) J. G. Topliss and R. J. Costello, *J. Med. Chem.*, **15**, 1066 (1972).
(16) R. M. Hyde, F. E. Norrington, S. G. Williams, and R. Wootton, manuscript in preparation.
(17) Y. C. Martin and W. J. Dunn, *J. Med. Chem.*, **16**, 578 (1973).
(18) The computer selections given in this paper were all made using interactive computer programs written in the Focal language (PS/8 Focal) on a PDP8-e computer with 12k of core store and DECtape backup store. Copies of these programs are available on request from the authors.

Synthesis and Biological Activity of Luteinizing Hormone-Releasing Hormone and Related Peptides

Derek J. Schafer,* Alan D. Black, John D. Bower,

Reckitt & Colman Pharmaceutical Division, Hull, United Kingdom

and D. Barry Crighton

University of Nottingham School of Agriculture, Sutton Bonington, United Kingdom. *Received November 18, 1974*

Syntheses of the decapeptide luteinizing hormone-releasing hormone, <Glu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH_2 are described. The basic properties of arginine can provide a simple repetitive isolation procedure for arginine-containing peptides. The biological activities of the decapeptide, of a range of fragments and modified fragments, and of two analogs with alteration in the serine at position 4 were measured by in vitro incubation with sheep pituitary slices, measuring the liberated LH by bioassay. None of the compounds of shortened sequence were active, with the exception of <Glu-His-Trp which showed 1% of the activity of the decapeptide in one of four experiments. Neither [Ser(Bu^t)^4]-LH-RH nor [Leu^4]-LH-RH showed significant activity indicating (despite the known activity of [Ala^4]-LH-RH) the importance of this part of the structure for full biological activity.

Following the discovery of the decapeptide structure of luteinizing hormone-releasing hormone (LH-RH) of both porcine[1] and ovine[2] origin, a number of syntheses have been described[3-14] and the structure–activity relationships of the molecule are emerging from the study of synthetic analogs.[15-33] Replacement of single amino acid residues often leads to a dramatic reduction in biological activity, particularly with residues **1**, **3**, and **9** or in replacing glycine at position 6 with L-amino acids. Analogs with a D-amino acid such as D-alanine in position 6 show,[32] in contrast, a remarkably high level of biological activity. Lower but significant activity has resulted by replacing histidine with phenylalanine[19] in position 2, by replacing tyrosine with phenylalanine[23] in position 5, by replacing leucine with isoleucine and other amino-acids[19] in position 7, and by replacing arginine in position 8 with lysine,[19] ornithine,[19] or glutamine.[21] Replacement of serine in position 4 by alanine,[19,22,27] threonine,[19,28] or glutamine[19] gave significantly active analogs, and replacement of the C-terminal glycineamide residue by ethylamido and other groups[19,20,24,28,29] gave analogs with high activity. Smaller peptides or fragments of the decapeptide have generally been inactive[18,33] although the tripeptide amide <Glu-His-Trp-NH_2 was reported as having significant activity,[25] a claim subsequently retracted.[30] There are also conflicting reports about the activity of the corresponding acid.[18,31]

The present work describes our syntheses of LH-RH and of [Ser(Bu^t)^4]- and [Leu^4]-LH-RH and the activity of these compounds and of a range of smaller fragments of LH-RH in releasing luteinizing hormone from ovine pituitary tissue in vitro.

Synthesis. Luteinizing hormone-releasing hormone was synthesized as shown in Charts I–III, using either unprotected serine or *tert*-butyl ether protection for the hydroxy group. In Chart I, the protected heptapeptide **11** corresponding to sequence 4–10 was synthesised by a stepwise active ester approach starting from glycineamide hydrochloride and protecting the arginine side chain with a nitro group. Serine and tyrosine were left unprotected. Benzyloxycarbonyl (Z) groups were used for α-amino protection and were removed by HBr in AcOH. At the heptapeptide stage, hydrogenation removed the nitro and Z groups and tryptophan was introduced using Z-Trp-ONp. Hydrogenation and coupling with <Glu-His-N_3 gave LH-RH. A scheme similar in part to this was adopted by Yanaihara et al.

The dipeptide **1** has been reported as having different melting points, which seem to be best explained by there being two crystalline forms melting at ca. 120° [34] and at ca. 145°,[12] respectively. In our work we obtained initially the form with mp 120°; this was difficult to recrystallize and tended to form a gel. Subsequently the compound crystallized in the higher melting form. The deprotected dipeptide salt **2** analyzed as the dihydrobromide, as did other hydrobromides in this series, possibly by formation of a weak salt with the C-terminal amide group. Countercurrent distribution was used to purify several protected intermediates of Chart I and was carried out either with relatively few transfers using separating funnels (tripeptide **3**, for example) or with more transfers using an automatic (steady state) machine (peptides **11** and **13**). The LH-RH (**15a**) was purified by ion-exchange chromatography on CM-Sephadex C-25 using pyridine–AcOH buffers, followed by partition chromatography on Sephadex LH20. The chromatographically pure decapeptide had the expected amino acid and elemental analyses and optical rotation.

A second approach to the synthesis of LH-RH is shown in Charts II and III. For several stages, use was made of the basic properties of arginine peptides to provide a simple separation of protected peptides from neutral coproducts of the coupling reaction.[35] The approach was based on the similar use of 4-picolyl esters[36-38] and of the basic properties of the histidine side chain[39] when this is present in the peptide. The coupling reaction is carried out with excess acylating agent until no amino component is detected and the product is separated from neutral and acidic coproducts by absorption into an acidic phase. It was found suf-

*Address correspondence to this author at the National Research Development Corp., Kingsgate House, Victoria Street, London, U.K.