Quantitative Structure-Activity Relationships. 3.[1] A Comparison of Different Free-Wilson Models

Hugo Kubinyi* and Ott-Hermann Kehrhahn

Research Institute of Knoll AG, Chemische Fabriken, D-6700 Ludwigshafen/Rhein, Germany. *Received December 24, 1975*

The Fujita-Ban model and the classical Free-Wilson model are shown to be linearly related: the de novo group contributions obtained by one model are linear transformations of those obtained by the other model. An example is given to illustrate this linear dependence. The Fujita-Ban model is characterized by a number of advantages as compared with the classical Free-Wilson model: no transformation of the structural matrix and no symmetry equations are necessary; all group contributions are based on an arbitrarily chosen reference compound, preferably the unsubstituted compound; the constant term, which is the theoretically predicted activity value of the reference compound, and the values of the group contributions are not markedly influenced by addition or elimination of a compound; the problem of linear dependence (the singularity problem) sometimes can be circumvented by preparation of a contracted matrix; if the unsubstituted compound is chosen as reference compound, the group contributions are numerically equivalent to Hansch-derived group contributions; therefore, the Hansch approach and the Fujita-Ban model can be combined to a mixed approach. Taking all these facts into consideration, the Fujita-Ban model is recommended as the most suitable approach for the calculation of de novo group contributions.

The Free-Wilson model[2] is a mathematical approach for the quantitative description of structure-activity relationships. It is based on the assumption that the biological activity of a molecule is the sum of the activity contributions of definite substructures, e.g., of the unsubstituted parent fragment (which is a hypothetical compound bearing no substituents, not even hydrogen) and the corresponding substituents. This additivity concept implies that the activity contributions of the parent fragment and of each substituent are constant, regardless of the structural variations in the rest of the molecule.

The classical Free-Wilson model is expressed by eq 1,

$$\text{biological activity (linear or log values)} = \sum_{ij} G_{ij}X_{ij} + \mu \quad (1)$$

in which μ is the overall average of biological activity values and G_{ij} is the activity contribution of the substituent X_i in position j ($X_{ij} = 1$ if the substituent X_i is in position j; otherwise $X_{ij} = 0$).

Based on the Free-Wilson additivity concept, three different modifications of this model were used in the past: the classical Free-Wilson model (eq 1),[2] the Cammarata model (eq 2),[3,4] and the Fujita-Ban model (eq 3)[5]

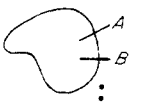
$$\text{biological activity (logarithmic values)} = \sum_{ij} a_{ij}X_{ij} + \mu_{\mathrm{H}} \quad (2)$$

$$\text{biological activity (logarithmic values)} = \sum_{ij} a_{ij}X_{ij} + \mu_0 \quad (3)$$

where a_{ij} = group contribution of the substituent X_i in position j, based on the definition that all $a_{\mathrm{H}} = 0$; μ_{H} = biological activity of the unsubstituted compound (all X_{ij} = H), *observed* value; and μ_0 = biological activity of the unsubstituted compound (all X_{ij} = H), *theoretically predicted* value.

Comparison of the Different Free-Wilson Models. Comparing the Free-Wilson models and the Hansch approach, it was stated in a recently published paper[6] that the Fujita-Ban modification is a linear transformation of the classical Free-Wilson model, but no mathematical evidence was given for this statement.

To provide this evidence and to clear up the relationships between the different Free-Wilson models, it is necessary to go back to the origin and to develop the models step by step. For a group of compounds with substituents A_i, B_j, ... in different positions p, q, ... first a *structural matrix* can be prepared. The structural matrix for a group of *N*,*N*-dimethyl-2-bromophenethylamines[7] is given in Table I; A_i and B_j are used for the meta and para substituents instead of X and Y to avoid confusions with the dependent variables y (note that the meta and para labels are reversed in Table I of ref 3b). Although it is possible to come to the same results using a theoretical example, it is more descriptive to use a real-life example; all aspects of Free-Wilson analysis can be demonstrated with this well-analyzed[3b,6,8,9] example (all conclusions drawn from this example are valid likewise for other examples, especially for compounds with more sites of substitution).



$A = A_1, A_2, A_3, \ldots, A_i$
$B = B_1, B_2, B_3, \ldots, B_j$
⋮

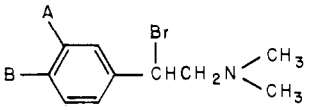
If the Free-Wilson additivity concept is applied to the compounds of Table I, a set of 22 equations of the general form of eq 4 results for the biological activity $y_{k,l}$ of every compound A_kB_l. In eq 4 the α_1, α_2, ..., α_i and β_1, β_2, ..., β_j are the group contributions (i.e., the G_i and a_i values of eq 1–3) of the substituents A_1, A_2, ..., A_i and B_1, B_2, ..., B_j and γ is the biological activity value of the parent fragment. Since the $y_{k,l}$ values are observed values including an experimental error $\epsilon_{k,l}$, this set of equations must be solved by linear multiple regression analysis; α_1, α_2, ..., α_i and β_1, β_2, ..., β_j are now regression coefficients, A_1, A_2, ..., A_i and B_1, B_2, ..., B_j are the independent variables, and γ is the intercept.

$$y_{k,l} = \alpha_k + \beta_l + \gamma \quad (4)$$

However, the matrix given in Table I cannot be solved in its original form due to a linear dependence of the A_i, B_j, ... columns of *each* structural matrix (although there are no linear dependences of the type described by Craig[10]): for each row of a structural matrix the sums of all A_i, all B_j, ... are equal to one (eq 5). This linear dependence is evident because for every site of substitution there is one and only one substituent (including H). Due to this linear dependence of the A_i, B_j, ... columns, every structural matrix is singular and, therefore, has an infinite number of solutions of the general form (c_a and c_b are

$$A_1 + A_2 + \ldots + A_i = B_1 + B_2 + \ldots + B_j = \ldots = 1 \quad (5)$$

Table I. Adrenergic Blocking Potencies of *N,N*-Dimethyl-2-bromophenethylamines. Structural Matrix and Biological Activity Values (Antagonism vs. Adrenaline in the Rat)



Compd no.	Meta substituents						Para substituents						Log 1/C obsd
	A_1 (H)	A_2 (F)	A_3 (Cl)	A_4 (Br)	A_5 (I)	A_6 (CH_3)	B_1 (H)	B_2 (F)	B_3 (Cl)	B_4 (Br)	B_5 (I)	B_6 (CH_3)	
1	1							1					8.16
2	1								1				8.68
3	1									1			8.89
4	1										1		9.25
5	1											1	9.30
6		1					1						7.52
7			1				1						8.16
8				1			1						8.30
9					1		1						8.40
10						1	1						8.46
11			1					1					8.19
12				1				1					8.57
13						1		1					8.82
14			1						1				8.89
15				1					1				8.92
16						1			1				8.96
17			1							1			9.00
18				1						1			9.35
19						1				1			9.22
20						1						1	9.30
21				1								1	9.52
22	1						1						7.46
Sums (n = 22)	6 (p_1)	1 (p_2)	4 (p_3)	5 (p_4)	1 (p_5)	5 (p_6)	6 (q_1)	4 (q_2)	4 (q_3)	4 (q_4)	1 (q_5)	3 (q_6)	191.32 (Σy)

constants with arbitrary values; only two sites of substitution are considered)

$$a_1 = \alpha_1 - c_a \qquad b_1 = \beta_1 - c_b$$
$$a_2 = \alpha_2 - c_a \qquad b_2 = \beta_2 - c_b$$
$$\vdots \qquad\qquad \vdots$$
$$a_i = \alpha_i - c_a \qquad b_j = \beta_j - c_b$$
$$\mu = \gamma + c_a + c_b$$

These equations interrelate any two different solutions a_1, a_2, ..., a_i, b_1, b_2, ..., b_j, μ and α_1, α_2, ..., α_i, β_1, β_2, ..., β_j, γ, respectively; c_a and c_b behave like additional unknowns which make a definite solution impossible. The assignment of arbitrary numbers to c_a and c_b does not solve this problem; c_a and c_b must be eliminated in any other way.

One possible way to do this is to define c_a by an α_i term and c_b by a β_j term, e.g., $c_a = \alpha_1$ and $c_b = \beta_1$. Substitution of these arbitrary assignments into the general, indefinite solution gives the following linear transformed values.

$$a_1 = 0 \qquad b_1 = 0$$
$$a_2 = \alpha_2 - \alpha_1 \qquad b_2 = \beta_2 - \beta_1$$
$$\vdots \qquad\qquad \vdots$$
$$a_i = \alpha_i - \alpha_1 \qquad b_j = \beta_j - \beta_1$$
$$\mu = \gamma + \alpha_1 + \beta_1$$

A new set of equations (eq 6) results if these linear transformed values are used to define the biological activity values $y_{k,l}$. From this set of equations a new matrix (Table

$$y_{k,l} = a_k + b_l + \mu \qquad (6)$$

II) results which corresponds to the original structural matrix (Table I), the only difference being the elimination of the A_1 and B_1 columns; these columns are suppressed because $a_1 = b_1 = 0$. Due to this fact the *calculated* biological activity value of the unsubstituted compound (compound **22**, $A_1 = B_1$ = H) is identical with the intercept μ (eq 7). This new matrix is a Fujita–Ban matrix, based

$$y_H(\text{obsd}) = \mu + \epsilon_H; \quad y_H(\text{calcd}) = \mu \qquad (7)$$

on the *arbitrary* assumption that all a_H (in this case a_1 and b_1) are zero.

From the Fujita–Ban matrix given in Table II, the regression coefficients a_2–a_6 and b_2–b_6 and the intercept μ can be calculated by standard programs of linear multiple regression analysis (for results see Tables V–VII).

Another possible way to eliminate c_a and c_b would be the arbitrary assignment $c_a = \alpha_2$ and $c_b = \beta_5$, which leads to $a_2 = 0$ and $b_5 = 0$. Another matrix results from this assumption, now lacking the A_2 and B_5 columns (the matrix is not given here because the transformation is evident from the foregoing example). Since all group contributions calculated from this matrix are based on $a_2 = b_5 = 0$, μ is now the calculated biological activity value of a reference compound with A_2 and B_5 as substituents; all a_i and b_j values are based on this new reference compound (it does not matter if this A_2B_5 compound is not included in the original structural matrix per se). This modification is in all respects equivalent to the Fujita–Ban model.

Out of a large number of different arbitrary assignments for c_a and c_b leading to different matrices (e.g., $c_a = \alpha_1 + \alpha_2 - \alpha_3$ leads to $a_3 = a_1 + a_2$; or $c_a = 0.4\,\alpha_1 + 0.6\,\alpha_2$ leads to $a_1 = -1.5\,a_2$, etc.), only two special cases will be considered. First, the arbitrary assignment can be made that

$$c_a = \frac{1}{i}(\alpha_1 + \alpha_2 + \ldots + \alpha_i) = \bar{\alpha} \qquad (8)$$

$$c_b = \frac{1}{j}(\beta_1 + \beta_2 + \ldots + \beta_j) = \bar{\beta} \qquad (9)$$

Table II. Fujita-Ban Matrix ($a_1 = b_1 = 0$) for the Compounds of Table I[a]

Compd no.	Meta substituents A_2 (F)	A_3 (Cl)	A_4 (Br)	A_5 (I)	A_6 (CH_3)	Para substituents B_2 (F)	B_3 (Cl)	B_4 (Br)	B_5 (I)	B_6 (CH_3)	Log 1/C obsd
1						1					8.16
2							1				8.68
3								1			8.89
4									1		9.25
5										1	9.30
6	1										7.52
7		1									8.16
8			1								8.30
9				1							8.40
10					1						8.46
11		1				1					8.19
12			1			1					8.57
13					1	1					8.82
14		1					1				8.89
15			1				1				8.92
16					1		1				8.96
17		1						1			9.00
18			1					1			9.35
19					1			1			9.22
20					1					1	9.30
21			1							1	9.52
22											7.46
Sums ($n = 22$)	1 (p_2)	4 (p_3)	5 (p_4)	1 (p_5)	5 (p_6)	4 (q_2)	4 (q_3)	4 (q_4)	1 (q_5)	3 (q_6)	191.32 (Σy)

[a] For better readability all zeros are deleted from the matrix.

Table III. Matrix for the Compounds of Table I, Based on the Arbitrary Assumptions That $a_1 = -a_2 - a_3 - a_4 - a_5 - a_6$ (Derived from Eq 10) and $b_1 = -b_2 - b_3 - b_4 - b_5 - b_6$ (Derived from Eq 11)[a]

Compd no.	Meta substituents A_2 (F)	A_3 (Cl)	A_4 (Br)	A_5 (I)	A_6 (CH_3)	Para substituents B_2 (F)	B_3 (Cl)	B_4 (Br)	B_5 (I)	B_6 (CH_3)	Log 1/C obsd
1	−1	−1	−1	−1	−1	1					8.16
2	−1	−1	−1	−1	−1		1				8.68
3	−1	−1	−1	−1	−1			1			8.89
4	−1	−1	−1	−1	−1				1		9.25
5	−1	−1	−1	−1	−1					1	9.30
6	1					−1	−1	−1	−1	−1	7.52
7		1				−1	−1	−1	−1	−1	8.16
8			1			−1	−1	−1	−1	−1	8.30
9				1		−1	−1	−1	−1	−1	8.40
10					1	−1	−1	−1	−1	−1	8.46
11		1				1					8.19
12			1			1					8.57
13					1	1					8.82
14		1					1				8.89
15			1				1				8.92
16					1		1				8.96
17		1						1			9.00
18			1					1			9.35
19					1			1			9.22
20					1					1	9.30
21			1							1	9.52
22	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	7.46
Sums ($n = 22$)	−5	−2	−1 ($p_i - p_1$)	−5	−1	−2	−2	−2 ($q_j - q_1$)	−5	−3	191.32 (Σy)

[a] For better readability all zeros are deleted from the matrix.

which leads to eq 10 and 11.

$$a_1 + a_2 + \ldots + a_i = \alpha_1 + \alpha_2 + \ldots + \alpha_i - i\bar{\alpha} = 0 \quad (10)$$

$$b_1 + b_2 + \ldots + b_j = \beta_1 + \beta_2 + \ldots + \beta_j - j\bar{\beta} = 0 \quad (11)$$

From eq 10 and 11 any a_i and b_j can be defined as a function of all other a and b values, e.g.

$$a_1 = -a_2 - a_3 - \ldots - a_i \quad (12)$$

$$b_1 = -b_2 - b_3 - \ldots - b_j \quad (13)$$

From eq 12 and 13 a matrix (Table III) results, which is similar but not identical with a classical Free–Wilson matrix. Despite this similarity the statement made by Schaad et al.[11] that eq 10 and 11 (eq 4 of ref 11) are identical with the classical Free–Wilson symmetry equations is incorrect (only in the special case that all p_i are identical and all q_j are identical, the classical Free–Wilson symmetry equations take the form of eq 10 and 11, see below).

Secondly, an even more complex arbitrary assignment must be considered to come to the original Free–Wilson model. If all equations derived from a structural matrix

Table IV. Free-Wilson Matrix for the Compounds of Table I (Derived from Symmetry Eq 21 and 22)[a]

Compd no.	Meta substituents					Para substituents					Log 1/C obsd
	A_2 (F)	A_3 (Cl)	A_4 (Br)	A_5 (I)	A_6 (CH_3)	B_2 (F)	B_3 (Cl)	B_4 (Br)	B_5 (I)	B_6 (CH_3)	
1	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$	1					8.16
2	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$		1				8.68
3	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$			1			8.89
4	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$				1		9.25
5	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$					1	9.30
6	1					$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	7.52
7		1				$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	8.16
8			1			$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	8.30
9				1		$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	8.40
10					1	$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	8.46
11		1				1					8.19
12			1			1					8.57
13					1	1					8.82
14		1					1				8.89
15			1				1				8.92
16					1		1				8.96
17		1						1			9.00
18			1					1			9.35
19					1			1			9.22
20					1					1	9.30
21			1							1	9.52
22	$-1/6$	$-2/3$	$-5/6$	$-1/6$	$-5/6$	$-2/3$	$-2/3$	$-2/3$	$-1/6$	$-1/2$	7.46
Sums ($n = 22$)	0	0	0	0	0	0	0	0	0	0	191.32 (Σy)
			$(p_i - p_1 \cdot p_i/p_1)$					$(q_j - q_1 \cdot q_j/q_1)$			

[a] For better readability all zeros are deleted from the matrix.

(see Table I and eq 4) are added, eq 14 results (p_i, q_j = the number of times each substituent A_i, B_j is present in the matrix; note that $p_1 + p_2 + ... + p_i = q_1 + q_2 + ... + q_j = n$).

$$\sum_{k=1}^{n} y_k = p_1\alpha_1 + p_2\alpha_2 + ... + p_i\alpha_i + q_1\beta_1 + q_2\beta_2 + ... + q_j\beta_j + n\gamma = \Sigma p_i\alpha_i + \Sigma q_j\beta_j + n\gamma \quad (14)$$

From eq 14 a further *arbitrary* assumption can be derived to eliminate c_a and c_b

$$c_a = \frac{1}{n}\Sigma p_i\alpha_i = \overline{p_i\alpha_i} \text{ and } c_b = \frac{1}{n}\Sigma q_j\beta_j = \overline{q_j\beta_j}$$

which leads to

$$p_1a_1 + p_2a_2 + ... + p_ia_i = \Sigma p_ia_i = p_1(\alpha_1 - \overline{p_i\alpha_i}) + p_2(\alpha_2 - \overline{p_i\alpha_i}) + ... + p_i(\alpha_i - \overline{p_i\alpha_i}) = \Sigma p_i\alpha_i - (p_1 + p_2 + ... + p_i)\overline{p_i\alpha_i} = \Sigma p_i\alpha_i - n\overline{p_i\alpha_i} = 0 \quad (15)$$

and likewise

$$\Sigma q_jb_j = \Sigma q_j\beta_j - n\overline{q_j\beta_j} = 0 \quad (16)$$

$\Sigma p_ia_i = 0$ (eq 15) and $\Sigma q_jb_j = 0$ (eq 16) are the well-known Free–Wilson "symmetry equations" for every site of substitution, from which any a_i and b_j can be defined as a weighted function of all other a and b values, e.g.

$$a_1 = -\frac{p_2}{p_1}\cdot a_2 - \frac{p_3}{p_1}\cdot a_3 - ... - \frac{p_i}{p_1}\cdot a_i \quad (17)$$

$$b_1 = -\frac{q_2}{q_1}\cdot b_2 - \frac{q_3}{q_1}\cdot b_3 - ... - \frac{q_j}{q_1}\cdot b_j \quad (18)$$

Since $\mu = \gamma + c_a + c_b = \gamma + \overline{p_i\alpha_i} + \overline{q_j\beta_j}$, eq 14 can be transformed to eq 19

$$\sum_{k=1}^{n} y_k = \Sigma p_i\alpha_i + \Sigma q_j\beta_j + n(\mu - \overline{p_i\alpha_i} - \overline{q_j\beta_j}) = n\mu \quad (19)$$

from which μ can be calculated (eq 20).

$$\mu = \Sigma y/n = \bar{y} \quad (20)$$

Equation 20 is the explanation for the fact that in the original Free–Wilson model, the intercept μ is the overall average of biological activity values $\bar{y}$.

If a_1 and b_1 are defined by symmetry eq 21 and 22, a Free–Wilson matrix results for the compounds of Table I (Table IV; instructions for the preparation of a Free–Wilson matrix from the structural matrix and the corresponding symmetry equations are given in ref 10 and 12).

$$6a_1 + a_2 + 4a_3 + 5a_4 + a_5 + 5a_6 = 0 \quad (21)$$

$$6b_1 + 4b_2 + 4b_3 + 4b_4 + b_5 + 3b_6 = 0 \quad (22)$$

If not a_1 and b_1 but a_2 and b_5 are defined by the other a and b values, another Free–Wilson matrix results, now lacking the a_2 and b_5 columns (the corresponding matrix is not presented here).

These theoretical considerations demonstrate that the definitions used in Fujita–Ban and Free–Wilson analysis are only facilities to get a definite solution and that all solutions are linear transformations of each other solution. For practical proof a_i and b_j values were calculated for the compounds of Table I by six different analyses, based on four different arbitrary assignments.

(I) a Fujita–Ban analysis (matrix given in Table II)

$a_1 = 0, b_1 = 0$

(II) a Fujita–Ban type analysis (no matrix given)

$a_2 = 0, b_5 = 0$

(III) an analysis based on the arbitrary restrictions, eq 10 and 11 (matrix given in Table III)

$a_1 = -a_2 - a_3 - a_4 - a_5 - a_6$ (derived from eq 10)

$b_1 = -b_2 - b_3 - b_4 - b_5 - b_6$ (derived from eq 11)

(IV) an analysis based on the same arbitrary restrictions (no matrix given)

Table V. Group Contributions for the Compounds of Table I, Derived by the Different Analyses I–VI (Group Contributions Derived by the Cammarata Model and the Hansch Approach Are Given for Comparison)

Group contribns, statistical parameters	Fujita-Ban models		Arbitrary model,[a] III,[e] IV	Free-Wilson model,[b] V,[f] VI	Cammarata model		Hansch-derived group contribn[c]
	I[d]	II			Values from ref 3b	Recalcd values	
Meta subst							
a_1 (H)	0.000	0.301	−0.229	−0.252	0.00	0.00	0.00
a_2 (F)[g]	−0.301	0.000	−0.530	−0.553	0.06	0.06	−0.21
a_3 (Cl)	0.207	0.508	−0.022	−0.045	0.52	0.40	0.26
a_4 (Br)	0.434	0.735	0.205	0.182	1.01	0.61	0.40
a_5 (I)[g]	0.579	0.880	0.350	0.327	0.84	0.94	0.62
a_6 (CH_3)	0.454	0.755	0.225	0.202	0.76	0.63	0.49
Para subst							
b_1 (H)	0.000	−1.429	−0.802	−0.623	0.00	0.00	0.00
b_2 (F)	0.340	−1.089	−0.462	−0.283	0.40	0.56	0.33
b_3 (Cl)	0.768	−0.661	−0.035	0.144	0.82	0.99	0.72
b_4 (Br)	1.020	−0.409	0.218	0.397	1.08	1.24	1.07
b_5 (I)[g]	1.429	0.000	0.627	0.806	1.79	1.79	1.41
b_6 (CH_3)	1.256	−0.173	0.454	0.633	1.32	1.50	1.28
Intercept μ	7.821	8.949	8.852	8.696	7.46	7.46	7.80
Correlation coeff r	0.969	0.969	0.969	0.969	*h, i*	*h*	0.966
Std dev s	0.194	0.194	0.194	0.194	0.324[i]	0.247	0.164

[a] Identical values were obtained from analyses III and IV. [b] Identical values were obtained from analyses V and VI. [c] Calculated from eq 23 and 24; compare ref 6 and 9. [d] Matrix given in Table II. [e] Matrix given in Table III. [f] Matrix given in Table IV. [g] Single point determinations. [h] No meaningful r values can be given because different formulas lead to different r values (r values > 1 are obtained if $r^2 = \Sigma(y_{calcd} - \bar{y})^2/\Sigma(y_{obsd} - \bar{y})^2$ is used for the calculation of r). [i] s value calculated from the log $1/C$ values given by Cammarata (we could not reproduce the values $r = 0.911$ and $s = 0.214$ published by Cammarata[3b]).

$a_2 = -a_1 - a_3 - a_4 - a_5 - a_6$ (derived from eq 10)

$b_5 = -b_1 - b_2 - b_3 - b_4 - b_6$ (derived from eq 11)

(V) a Free–Wilson analysis I (matrix given in Table IV)

$a_1 = -{}^1/_6 a_2 - {}^2/_3 a_3 - {}^5/_6 a_4 - {}^1/_6 a_5 - {}^5/_6 a_6$ (derived from eq 21)

$b_1 = -{}^2/_3 b_2 - {}^2/_3 b_3 - {}^2/_3 b_4 - {}^1/_6 b_5 - {}^1/_2 b_6$ (derived from eq 22)

(VI) a Free–Wilson analysis II (no matrix given)

$a_2 = -6a_1 - 4a_3 - 5a_4 - a_5 - 5a_6$ (derived from eq 21)

$b_5 = -6b_1 - 4b_2 - 4b_3 - 4b_4 - 3b_6$ (derived from eq 22)

The results (a_i and b_j values, μ values, correlation coefficients r, and standard deviations s) are given in Table V; all a_i and b_j values are given with three decimal places to minimize rounding errors (for practical purposes two decimal places are sufficient). Group contributions resulting from the Cammarata model (eq 2) and Hansch-derived group contributions are given for comparison; since the values given by Cammarata[3b] seem to be erroneous, recalculated values are given for comparison (for a discussion of the Cammarata model see below). The Hansch-derived group contributions were calculated by appropriate transformation[6] (eq 23, 24) of an equation given by Unger and Hansch[9] (eq 25).

$$a_i = 0.83\pi_i - 0.92\sigma^+_i \quad (23)$$

$$b_j = 1.33\pi_j - 1.89\sigma^+_j \quad (24)$$

$$\log 1/C = 0.83\ (\pm 0.27)\ \pi_m + 1.33\ (\pm 0.20)\ \pi_p - 0.92\ (\pm 0.50)\ \sigma^+_m - 1.89\ (\pm 0.57)\ \sigma^+_p + 7.80 \quad (25)$$

$n = 22, r = 0.966, s = 0.164$

These Hansch-derived group contributions allow the direct comparison of Hansch analysis with the different modifications of Free–Wilson analysis.

The group contributions derived by analyses I–VI (Table V) demonstrate the validity of the foregoing mathematical derivations: all different solutions are simple linear transformations. The a_i, b_j, and μ values found with analyses II–VI can be transformed to the a_i, b_j, and μ values found with analysis I by subtraction of the corresponding a_1 value from all a_i values and the corresponding b_1 value from all b_j values and by addition of the a_1 and b_1 values to the intercept μ. All other linear transformations can be done in a similar manner.

Since all these linear transformed solutions lead to the same set of calculated log $1/C$ values[11] (Table VI), the question arises whether it is necessary to have different modifications of the Free–Wilson model and which modification is the most suitable for practical purposes. To answer this question the statistical parameters should be considered first; all six modifications I–VI give identical r, s, and F values (all $F = 16.99$), while the Cammarata modification gives a larger standard deviation (see Tables V and VI), indicating the better fit of the modifications I–VI.

The Cammarata model[3,4] (eq 2) differs from the Fujita–Ban model (eq 3) in only one respect: instead of the theoretically predicted activity value of the reference compound (= the unsubstituted compound, all X_{ij} = H) the *observed* value is used as the constant term (=μ_H). To demonstrate the differences among classical Free–Wilson analysis, the Fujita–Ban modification, and the Cammarata model, the normal equations for each model must be considered; in the case of a group of compounds with substituents A_i and B_j the normal equations are a set of $(i + j - 2)$ equations which are developed from the corresponding matrix (e.g., Tables II, III, or IV) for the calculation of the a_i and b_j values following the rules of the least-squares method (eq 26) (compare ref 13, p 14). In classical Free–Wilson analysis and in Fujita–Ban an-

Table VI. Calculated Log $1/C$ Values and Deviations Δ between Observed and Calculated Log $1/C$ Values from Different Models (Δ = Log $1/C$ Calcd − Log $1/C$ Obsd; Deviations of ±0.01 May Occur Due to Rounding Errors). For Comparison of the Different Models, the Sums of the Squared Deviations $\Sigma\Delta^2$, the Degrees of Freedom DF, and the Standard Deviations s Are Given

		Free-Wilson model, Fujita-Ban models (analyses I-VI[a])		Cammarata model				Hansch model, values calcd from eq 25[b]	
				Values from ref 3b		Recalcd values			
Compd no.	Log $1/C$ obsd	Log $1/C$ calcd	Δ	Log $1/C$ calcd	Δ	Log $1/C$ calcd	Δ	Log $1/C$ calcd	Δ
1	8.16	8.16	0.00	7.86	−0.30	8.02	−0.14	8.13	−0.03
2	8.68	8.59	−0.09	8.28	−0.40	8.45	−0.23	8.52	−0.16
3	8.89	8.84	−0.05	8.54	−0.35	8.70	−0.19	8.87	−0.02
4	9.25	9.25	0.00	9.25	0.00	9.25	0.00	9.21	−0.04
5	9.30	9.08	−0.22	8.78	−0.52	8.96	−0.34	9.08	−0.22
6	7.52	7.52	0.00	7.52	0.00	7.52	0.00	7.59	0.07
7	8.16	8.03	−0.13	7.98	−0.18	7.86	−0.30	8.06	−0.10
8	8.30	8.26	−0.04	8.47	0.17	8.07	−0.23	8.20	−0.10
9	8.40	8.40	0.00	8.40	0.00	8.40	0.00	8.42	0.02
10	8.46	8.28	−0.18	8.22	−0.24	8.09	−0.37	8.29	−0.17
11	8.19	8.37	0.18	8.38	0.19	8.42	0.23	8.39	0.20
12	8.57	8.60	0.03	8.87	0.30	8.64	0.07	8.53	−0.04
13	8.82	8.62	−0.20	8.62	−0.20	8.66	−0.16	8.62	−0.20
14	8.89	8.80	−0.09	8.80	−0.09	8.85	−0.04	8.79	−0.10
15	8.92	9.02	0.10	9.29	0.37	9.06	0.14	8.93	0.01
16	8.96	9.04	0.08	9.04	0.08	9.08	0.12	9.01	0.05
17	9.00	9.05	0.05	9.06	0.06	9.10	0.10	9.14	0.14
18	9.35	9.28	−0.07	9.55	0.20	9.32	−0.03	9.28	−0.07
19	9.22	9.30	0.08	9.30	0.08	9.34	0.12	9.36	0.14
20	9.30	9.53	0.23	9.54	0.24	9.59	0.29	9.57	0.27
21	9.52	9.51	−0.01	9.79	0.27	9.57	0.05	9.48	−0.04
22	7.46	7.82	0.36	7.46	0.00	7.46	0.00	7.80	0.34
$\Sigma\Delta^2$			0.41		1.26		0.74		0.46
DF			11		12		12		17
s			0.194		0.324		0.247		0.164

[a] Identical log $1/C$ values were obtained from analyses I-VI. [b] Compare ref 9.

$$a_2[A_2A_2] + a_3[A_2A_3] + \ldots + a_i[A_2A_i] + b_2[A_2B_2] + \ldots + b_j[A_2B_j] = [A_2Y]$$
$$a_2[A_3A_2] + a_3[A_3A_3] + \ldots + a_i[A_3A_i] + b_2[A_3B_2] + \ldots + b_j[A_3B_j] = [A_3Y]$$
$$\vdots$$
$$a_2[A_iA_2] + a_3[A_iA_3] + \ldots + a_i[A_iA_i] + b_2[A_iB_2] + \ldots + b_j[A_iB_j] = [A_iY]$$
$$a_2[B_2A_2] + a_3[B_2A_3] + \ldots + a_i[B_2A_i] + b_2[B_2B_2] + \ldots + b_j[B_2B_j] = [B_2Y]$$
$$\vdots$$
$$a_2[B_jA_2] + a_3[B_jA_3] + \ldots + a_i[B_jA_i] + b_2[B_jB_2] + \ldots + b_j[B_jB_j] = [B_jY] \quad (26)$$

alysis the terms $[A_2A_2]$, etc., are defined as

$$[A_2A_2] = \Sigma(A_2 - \bar{A}_2)^2$$
$$[A_2A_i] = [A_iA_2] = \Sigma(A_2 - \bar{A}_2)(A_i - \bar{A}_i)$$
$$[A_iB_j] = [B_jA_i] = \Sigma(A_i - \bar{A}_i)(B_j - \bar{B}_j)$$
$$[A_2Y] = \Sigma(A_2 - \bar{A}_2)(y - \bar{y}), \text{ etc.} \quad (27)$$

and the intercept μ is defined as

$$\mu = \bar{y} - a_2\bar{A}_2 - a_3\bar{A}_3 - \ldots - a_i\bar{A}_i - b_2\bar{B}_2 - \ldots - b_j\bar{B}_j \quad (28)$$

Equations 26–28 must be transformed appropriately if not the A_1 and B_1 columns but any other A_i and B_j columns are absent in the corresponding Free–Wilson or Fujita–Ban matrix (the normal equations presented by Schaad et al.[11] are algebraic transformations of eq 26–28 which lead to a Fujita–Ban solution).

The normal eq 26 and the definitions of the $[A_2A_2]$ ... terms given in eq 27 correspond to the common least-squares calculation procedure used in linear multiple regression analysis. Due to the identity of the normal equations every Free–Wilson matrix and every Fujita–Ban matrix can be solved with standard programs of linear multiple regression analysis; it should be noted that the linear relationship of the solutions obtained from classical Free–Wilson analysis and Fujita–Ban analysis is based on this identity of the normal equations.

However, there is another possible way to solve a Free–Wilson matrix; due to the symmetry restrictions all sums of the columns and therefore all $\bar{A}_1$, $\bar{A}_2$, ..., $\bar{A}_i$, $\bar{B}_1$, ..., $\bar{B}_j$ are zero (see Table IV). This special property of a Free–Wilson matrix leads to a simplification of the $[A_2A_2]$ ... terms and of eq 28.

$$[A_2A_2] = \Sigma A_2^2;\ [A_2A_i] = [A_iA_2] = \Sigma A_2 \cdot A_i$$
$$[A_iB_j] = [B_jA_i] = \Sigma A_i \cdot B_j;\ [A_2Y] = \Sigma A_2(y - \bar{y}) \quad (29)$$
$$\mu = \bar{y} \quad (30)$$

If $\bar{y}$ is subtracted from all y_{ij}, a least-squares calculation procedure for equations containing *no* intercept can be applied, leading to identical solutions.

The calculation procedure of the Cammarata model cannot be reconstructed definitely because only two examples were presented; the first example[3a] is made up of

ten single point determinations (no degree of freedom), and the second example[3b] seems to be erroneous (compare Table V). From the definitions of the model one must assume that Cammarata eliminated the intercept by subtraction of y_H (= the observed biological activity value of the unsubstituted compound; all X_{ij} = H) from all y_{ij} values and that the resulting matrix, identical in all other respects with a Fujita–Ban matrix, was solved by the least-squares method for equations containing no intercept.

However, this calculation procedure is based on the incorrect assumption that the observed activity value of the unsubstituted compound contains no experimental error; since each observed activity value y_{ij} includes an experimental error ϵ_{ij}, the observed activity value y_H includes an error ϵ_H. If this experimental activity value is used as the constant term μ, the error ϵ_H has different weight compared to all other errors ϵ_{ij}. The basic equations of all different Free–Wilson modifications (eq 1–3) contain an intercept; therefore, the simplified calculation procedure (eq 29) for equations containing *no* intercept can only be applied in the special case that all sums of the columns in the corresponding matrix are zero (which is the case in the classical Free–Wilson model but not in the Cammarata model). If the group contributions of all hydrogen substituents are arbitrarily put to zero, which is the definition of the Fujita–Ban model *and* the Cammarata model, the intercept μ is the theoretically predicted activity value of the unsubstituted compound (eq 7) and not the experimental value. It should be noted that also in Hansch analysis the constant term is the theoretically predicted activity value of the unsubstituted compound, provided that all structural parameters ϕ_j are based on $\phi_H = 0$ (e.g., π or σ but not log P or E_s).

Due to this inappropriate calculation procedure used by Cammarata, the classical Free–Wilson model and the Fujita–Ban model always give better and more reliable results than the Cammarata model; only in the case of $\epsilon_H \rightarrow 0$ the Cammarata model gives a comparable solution; if a compound with a substituent occurring only once in the structural matrix (a single point determination) is chosen as reference compound, the Cammarata model and the Fujita–Ban model give identical solutions because the corresponding error term is included in the single point derived group contribution ($y_{calcd} \equiv y_{obsd}$ for single point determinations).

Going back to classical Free–Wilson analysis and Fujita–Ban analysis, next the confidence limits and the t values of each a_i and b_j term found with the different modifications I–VI must be considered (Table VII). It is not surprising that the t values are meaningless (a_i and b_j values significant in one analysis are not significant in another analysis and conversely); since all solutions are linear transformations and the t values are linearly proportional to the a_i and b_j values, the t values are fortuitous values, dependent on the arbitrary assignment which was used as basis of the corresponding model. Somewhat more surprising is the fact that the confidence limits of the Fujita–Ban-derived group contributions seem to be larger than the confidence limits derived by classical Free–Wilson analysis (see Table VII). However, this difference comes only from the "distribution" of the uncertainty of the reference group contribution to the other group contributions in the Fujita–Ban model; in analysis I the confidence limits of a_1 and b_1 (reference substituents) are zero; therefore, the confidence limits of the residual group contributions include the uncertainty of the corresponding group contribution *and* the uncertainty of the reference group contribution; in analysis II all group contributions

Table VII. Confidence Limits and t Values for the Group Contributions Derived with the Different Analyses I–VI

Group contribns a_i (p_i), b_j (q_j)	Confidence limits (t values) for the group contribns derived with analyses			
	I	II	III, IV[a]	V, VI[b]
a_1 (6)	0.00[c] (–)	±0.50 (1.31)	±0.21[d] (2.42)	±0.16[f] (3.46)
a_2 (1)	±0.50 (1.31)	0.00[c] (–)	±0.39[e] (3.00)	±0.45[g] (2.70)
a_3 (4)	±0.29 (1.56)	±0.51 (2.19)	±0.22 (0.22)	±0.20 (0.50)
a_4 (5)	±0.27 (3.55)	±0.50 (3.21)	±0.21 (2.18)	±0.17 (2.35)
a_5 (1)	±0.50 (2.52)	±0.60 (3.21)	±0.39 (1.98)	±0.45 (1.59)
a_6 (5)	±0.27 (3.71)	±0.50 (3.29)	±0.21 (2.39)	±0.17 (2.61)
b_1 (6)	0.00[c] (–)	±0.50 (6.23)	±0.21[d] (8.50)	±0.17[f] (7.89)
b_2 (4)	±0.30 (2.48)	±0.50 (4.75)	±0.21 (4.90)	±0.20 (3.18)
b_3 (4)	±0.30 (5.61)	±0.50 (2.88)	±0.21 (0.37)	±0.20 (1.62)
b_4 (4)	±0.30 (7.45)	±0.50 (1.78)	±0.21 (2.31)	±0.20 (4.46)
b_5 (1)	±0.50 (6.23)	0.00[c] (–)	±0.39[e] (3.55)	±0.45[g] (3.98)
b_6 (3)	±0.33 (8.35)	±0.52 (0.74)	±0.23 (4.28)	±0.24 (5.89)

[a] Identical values were obtained from analyses III and IV. [b] Identical values were obtained from analyses V and VI. [c] By definition. [d] From analysis IV only. [e] From analysis III only. [f] From analysis VI only. [g] From analysis V only.

are based on two single point determinations (a_2 and b_5) which lead to confidence limits for the other group contributions that are even larger than the confidence limits obtained by analysis I. It is not strictly correct but very descriptive to imagine that in analysis I the uncertainties of the single point derived group contributions a_2, a_5, and b_5 are only in the corresponding confidence limits, while in the case of analysis II the uncertainties resulting from the single points a_2 and b_5 are in *all* confidence limits because the confidence limits of a_2 and b_5 were forced to zero by definition.

For that reason it is evident that only well-represented substituents should be taken as reference substituents in Fujita–Ban analysis to minimize the confidence limits of the resulting group contributions. The most meaningful confidence limits seem to result from classical Free–Wilson analysis; in contrast to Hansch analysis the indicative value of the confidence limits is small in Free–Wilson analysis because the confidence limits are only a measure of the relative frequencies p_i, q_j, ... of the substituents A_i, B_j, ... (note that all statistical parameters should be interpreted with care because most log $1/C$ values used in quantitative structure–activity analyses are mean values obtained from a number of experiments; therefore, the total experimental error is not included in the regressions).

Some problems arising from single point determinations have been discussed in the literature;[10] an analysis based on a large number of single point determinations is not very reliable because every single point derived group contribution reflects the experimental error of one compound. Nevertheless, there is no reason to exclude them from the analysis if there are not too much: if compounds **4**, **6**, and **9** (leading to single point determinations for group contributions a_2, a_5, and b_5) are eliminated from Table II, the resulting group contributions a_3, a_4, a_6, b_2–b_4, and b_6 and the intercept μ_0 are identical with those obtained from the original Fujita–Ban matrix; the statistical parameters

are not significantly influenced (n = 19, r = 0.958, s = 0.194, F = 17.33). The situation is more complex in classical Free–Wilson analysis: different symmetry equations, a different matrix, and different group contributions result from the elimination of compounds **4**, **6**, and **9**; of course, these new group contributions are only a new linear transformation of the old values, based on the new intercept (the statistical parameters are identical with those obtained from the corresponding Fujita–Ban analysis).

A further problem of classical Free–Wilson analysis was demonstrated by Craig[10] and Purcell et al.;[12] if two substituents A_1 and B_1 always occur together in a structural matrix (Craig used the term "singularities"), the corresponding group contributions cannot be separated because there are not as many *independent* equations as unknowns for these substituents. Craig pointed out that the use of standard programs of linear multiple regression analysis can force a solution which is only one possible solution out of an infinite number of solutions (this case is comparable to the linear dependences in each structural matrix); Craig recognized that the problem of "ill conditioning" described by Hudson et al.[14] was only caused by linear dependences in their matrix (compounds **3**–**6**; compare ref 10). The mathematical background of linear dependence was discussed by Schaad et al.[11]

Purcell et al.[12] gave the following example.

$A_1B_1C_1$	$A_2B_2C_1$	$A_3B_3C_1$	$A_2B_3C_1$	$A_3B_2C_1$
$A_1B_1C_2$	$A_2B_2C_2$	$A_3B_3C_2$	$A_2B_3C_2$	$A_3B_2C_2$
$A_1B_1C_3$	$A_2B_2C_3$	$A_3B_3C_3$	$A_2B_3C_3$	$A_3B_2C_3$

A set of 15 compounds with different substituents A, B, and C gives after appropriate transformation a Free–Wilson matrix corresponding to a system of 15 equations with six unknowns. Dependent on the computer program used for regression analysis, no solution or a meaningless solution will result for the group contributions a_1 and b_1. While the compounds $A_1B_1C_1$, $A_1B_1C_2$, and $A_1B_1C_3$ must be eliminated in classical Free–Wilson analysis to get a reliable result, this needs not to be done in Fujita–Ban analysis; if, e.g., A_2, B_2, and C_2 are taken as reference substituents ($a_2 = b_2 = c_2 = 0$), the substitution A_1B_1 can be regarded like a single substituent D; the resulting group contribution d corresponds to the group contribution of the A_1B_1 substitution, based on the A_2B_2 substitution. This combination of two inseparable group contributions leads to the following contracted Fujita–Ban matrix.

compd	D	C_1	A_3	B_3	C_3
$A_1B_1C_1$	1	1	0	0	0
$A_1B_1C_2$	1	0	0	0	0
$A_1B_1C_3$	1	0	0	0	1
$A_2B_2C_1$	0	1	0	0	0
$A_2B_2C_2$	0	0	0	0	0
$A_2B_2C_3$	0	0	0	0	1
$A_3B_3C_1$	0	1	1	1	0
.	.				
.	.				
.	.				

Complex linear dependences can arise in an unbalanced group of compounds: in a group of compounds A_iB_j

A_1B_1	A_2B_1	A_3B_1	A_4B_4	A_5B_4	A_6B_4
A_1B_2	A_2B_2	A_3B_2	A_4B_5	A_5B_5	A_6B_5
A_1B_3	A_2B_3	A_3B_3	A_4B_6	A_5B_6	A_6B_6

the substituents A_1–A_3 always occur together with B_1–B_3 and the substituents A_4–A_6 always occur together with B_4–B_6. This corresponds to two linear dependences $A_1 + A_2 + A_3 = B_1 + B_2 + B_3$ and $A_4 + A_5 + A_6 = B_4 + B_5 + B_6$ in the structural matrix. Due to these linear dependences Free–Wilson analysis and Fujita–Ban analysis can only be applied to the first nine compounds and to the second nine compounds separately but not to the complete group of compounds because unreliable solutions will result for the group contributions of those substituents which are linearly dependent in the corresponding matrix (if, e.g., the substituents A_1 and B_4 are taken as reference substituents, the columns A_4–A_6 and B_1–B_3 are linearly dependent in the resulting Fujita–Ban matrix: $A_4 + A_5 + A_6 + B_1 + B_2 + B_3 = 1$).

Addition of a compound A_1B_4 or A_3B_5 to the given example eliminates the problem of linear dependence, but a new problem arises: Free–Wilson analysis or Fujita–Ban analysis forces a solution where all group contributions are influenced by the experimental error of the activity value of the added compound. This example demonstrates that one should select the compounds for synthesis and subsequent structure–activity analysis carefully. If grouping of substituents results from an unbalanced choice of compounds, unreliable solutions may be obtained from Free–Wilson or Fujita–Ban analysis.

The most important disadvantage of Free–Wilson analysis and Fujita–Ban analysis is the usually high number of variables needed to describe all substituents. In some cases this disadvantage can be circumvented by appropriate combination of substituents in one column (some of the following simplifications can only be applied in Fujita–Ban analysis, not in Free–Wilson analysis).

(a) In symmetrically substituted compounds, e.g., ortho,ortho'- or meta,meta'-disubstituted benzenes, the corresponding substituent columns can be combined to one column; a coefficient of 2 is given to such symmetrically disubstituted compounds.[6,11,15] If a substituent gives very similar group contributions in all positions of the benzene ring, all corresponding columns may be combined: coefficients of 2, 3, 4, and 5 are given to di-, tri-, tetra- and penta-substituted benzenes.[11,16]

(b) In homologous series the group contributions of the $-CH_2-$ group can be assumed to be identical; if, e.g., the C_8 compound is used as reference compound, coefficients of 2, 4, 6, ... may be given to the C_{10}, C_{12}, C_{14}, ... compound.

(c) Hansch et al.[17] combined chemically different substituents in one column if these substituents had similar group contributions and eliminated columns if the corresponding group contributions were not significantly different from zero. To our opinion this extreme reduction of the number of variables is a dangerous procedure because of its arbitrariness (no sequential F test should be applied in Free–Wilson and Fujita–Ban analysis because also a group contribution not significantly different from zero is meaningful despite its influence on the other statistical parameters).

(d) If for one definite region of the molecule a Hansch correlation can be obtained for the substituents, while substituents in another position of the molecule must be treated by Free–Wilson analysis, the Fujita–Ban model and the Hansch approach can be combined to a mixed approach,[16,17] e.g., eq 31 (note that the Fujita–Ban model

$$\log 1/C = \Sigma a_i + k_1\pi + k_2\sigma + \mu \qquad (31)$$

was called the "modified Free–Wilson model" in ref 6 and 16). In eq 31 Σa_i is a Free–Wilson part for the substituents X_i, $k_1\pi + k_2\sigma$ is a Hansch part for substituents Y_j, and μ is the theoretically predicted activity value of the unsubstituted parent compound (X = Y = H) or of an arbitrarily chosen reference compound. Other forms of the mixed approach which are applicable in the case of nonadditivity of group contributions, e.g., nonlinear dependence of biological activity from lipophilic character, have been discussed in ref 16.

Discussion

The original Free–Wilson model is based on symmetry equations (originally called "restrictions"[2]). Although Free and Wilson gave an example which demonstrates that these restrictions are *arbitrary* assumptions and that the symmetry results only from the use of average values as reference values, these symmetry equations have led to some speculations: Craig[10] stated that the basic assumption of additivity of group contributions demands these symmetry equations; Franke and Oehme[18] stated that the symmetry equations are meaningless if the logarithms of the biological activity values are used as activity parameters instead of the linear values. Nothing of that is correct; the symmetry equations are only one possible arbitrary assumption to get a definite solution of the matrix. Furthermore, the assumption that in classical Free–Wilson analysis the intercept μ is identical with the biological activity of the parent fragment is not correct; μ is the overall average of biological activity values and nothing else.

Due to the specific properties of the Free–Wilson model, a simple calculation procedure for the a_i and b_j values results (compare eq 29 and 30); since this work is usually done by a computer with standard programs, this simplification of the calculation procedure is of no real value. On the other hand, two disadvantages arise from the symmetry equations: if no special computer program is available which includes the transformation of the structural matrix to the Free–Wilson matrix, this transformation must be done "by hand"; every elimination of a compound or addition of a new compound gives new symmetry equations and a new matrix; since the μ term is the overall average and all a_i, b_j, ... values are based on this value, every addition or elimination of a compound causes a change of μ and therefore a change of all a_i, b_j, ... values. This is a more serious disadvantage because such eliminations of one or more compounds are usually done to detect irregular values (outliers).

The modification of the Free–Wilson model described by Fujita and Ban[5] (eq 3) differs from the original model in three respects: first, all activity contributions are based on the arbitrary assumption that all a_H, b_H, ... are zero; secondly, as a consequence of this definition, the constant term μ_0 obtained by the least-squares method is the theoretically predicted activity value of the unsubstituted compound (all X_{ij} = H; eq 7); thirdly, Fujita and Ban used the logarithms of the biological activity values instead of the linear values since the log of activity is considered to be a free-energy related parameter which is additive. In accordance with Hansch analysis (which is also based on the additivity concept in its nonparabolic form) only log values of biological activity should be used nowadays in Free–Wilson and Fujita–Ban analysis; the arguments given by Purcell and Clayton[19] from a comparison of linear and antilog values are incompatible with the definition of the least-squares method (it should be noted that negative dose values or negative values of relative biological activities can result from the use of linear values).

If for one or more definite sites of substitution no hydrogen substituent is included in the structural matrix, any other substituent may be taken as reference substituent;[10] in that case the constant term μ_0 is the theoretically predicted activity value of a compound bearing these reference substituents (compare analysis II, Table V); all group contributions are based on this reference compound. Due to the arbitrary choice of the reference compound, we cannot agree with the opinion of Fujita et al.[20,21] that the Fujita–Ban model is only applicable if the unsubstituted compound is included.

The definitions of the Fujita–Ban model lead to a number of important advantages.

(1) No symmetry equations and no complex transformations of the structural matrix are necessary; the elimination of one column for every site of substitution (the reference substituent columns) is the only change in going from the structural matrix (Table I) to a Fujita–Ban matrix (Table II).

(2) The matrix is not changed by addition or elimination of one row (addition or elimination of a compound); if single point determinations are eliminated from the matrix, one row (the compound) and one column (the corresponding substituent) are eliminated from the matrix.

(3) The constant term μ_0 and all group contributions a_i, b_j, ... are not markedly influenced by the addition or elimination of a compound. If only single point determinations are eliminated from the matrix, no changes occur in the a_i, b_j, ... and μ_0 values (which is not the case in the original Free–Wilson model).

(4) The set of a_i, b_j, ... and μ_0 values is a simple linear transformation of the values obtained by classical Free–Wilson analysis (see Table V).

(5) The Fujita–Ban group contributions are directly comparable to Hansch-derived group contributions; if all Hansch-derived group contributions are calculated from structural parameters ϕ_j which are based on $\phi_H = 0$, and if the reference compound in the Fujita–Ban analysis is the unsubstituted compound, the group contributions are numerically equivalent[6] (compare Table V; the group contributions derived from analysis I and the Hansch-derived group contributions are nearly identical). Due to this numerical equivalence the Fujita–Ban model and the Hansch approach can be combined to a mixed approach which incorporates the elements of Free–Wilson analysis and Hansch analysis in one model. This mixed approach makes use of the advantages of each model and widens the applicability of Hansch and Free–Wilson analysis.

(6) The problem of linear dependence which leads to unreliable results in classical Free–Wilson analysis sometimes can be circumvented in Fujita–Ban analysis by the preparation of a contracted matrix.

From this comparison of the different Free–Wilson models the following statements and recommendations are derived: only the Fujita–Ban model should be used for the calculation of de novo group contributions; the corresponding matrix is easy to prepare and all group contributions are based on an arbitrarily chosen reference compound, preferably the unsubstituted compound (all X_{ij} = H). If the unsubstituted compound is used as reference compound, the resulting group contributions are directly comparable to Hansch-derived group contributions. In cases where no hydrogen-substituted compound is present, any well-represented substituents may be chosen as reference substituents. The group contributions obtained by classical Free–Wilson analysis are linear transformations of those obtained by Fujita–Ban analysis; however, the preparation of the matrix is more complex and all values change by addition or elimination of a single compound. A simple algorithm for Fujita–Ban analysis which allows the calculation of de novo group contributions without a computer will be presented in a forthcoming paper.[22]

References and Notes

(1) For paper 2, see ref 16.
(2) S. M. Free, Jr., and J. W. Wilson, *J. Med. Chem.*, 7, 395 (1964).
(3) (a) A. Cammarata and S. J. Yau, *J. Med. Chem.*, 13, 93 (1970); (b) A. Cammarata, *ibid.*, 15, 573 (1972).

(4) A. Cammarata and K. S. Rogers in "Advances in Linear Free Energy Relationships", N. B. Chapman and J. Shorter, Ed., Plenum Press, New York, N.Y., 1972, p 403.
(5) T. Fujita and T. Ban, *J. Med. Chem.*, **14**, 148 (1971).
(6) H. Kubinyi and O. H. Kehrhahn, *J. Med. Chem.*, **19**, 578 (1976).
(7) J. D. P. Graham and M. A. Karrar, *J. Med. Chem.*, **6**, 103 (1963).
(8) C. Hansch and E. J. Lien, *Biochem. Pharmacol.*, **17**, 709 (1968).
(9) S. H. Unger and C. Hansch, *J. Med. Chem.*, **16**, 745 (1973).
(10) P. N. Craig, *Adv. Chem. Ser.*, **No. 114**, 115 (1972).
(11) L. J. Schaad, R. H. Werner, L. Dillon, L. Field, and C. E. Tate, *J. Med. Chem.*, 18, 344 (1975).
(12) W. P. Purcell, G. E. Bass, and J. M. Clayton, "Strategy of Drug Design: A Guide to Biological Activity", Wiley-Interscience, New York, N.Y., 1973.
(13) C. Daniel and F. S. Wood, "Fitting Equations to Data", Wiley-Interscience, New York, N.Y., 1971.
(14) D. R. Hudson, G. E. Bass, and W. P. Purcell, *J. Med. Chem.*, **13**, 1184 (1970).
(15) R. T. Buckler, *J. Med. Chem.*, **15**, 578 (1972).
(16) H. Kubinyi, *J. Med. Chem.*, **19**, 587 (1976).
(17) C. Hansch, C. Silipo, and E. E. Steller, *J. Pharm. Sci.*, **64**, 1186 (1975).
(18) R. Franke and P. Oehme, *Pharmazie*, **28**, 489 (1973).
(19) W. P. Purcell and J. M. Clayton, *J. Med. Chem.*, **11**, 199 (1968).
(20) E. Mizuta, N. Suzuki, Y. Miyake, M. Nishikawa, and T. Fujita, *Chem. Pharm. Bull.*, **23**, 5 (1975).
(21) S. Inoue, A. Ogino, M. Kise, M. Kitano, S. Tsuchiya, and T. Fujita, *Chem. Pharm. Bull.*, **22**, 2064 (1974).
(22) H. Kubinyi, *Arzneim.-Forsch.*, submitted for publication.

Clonidine and Related Analogues. Quantitative Correlations

Bruno Rouot, Gérard Leclerc,* Camille-Georges Wermuth,

Laboratoire de Chimie Organique et Thérapeutique (E.R.A. 393), Faculté de Pharmacie, Université Louis Pasteur, 67083 Strasbourg Cedex, France

François Miesch, and Jean Schwartz

Institut de Pharmacologie et de Médecine Expérimentale (E.R.A. 142), Faculté de Médecine, Université Louis Pasteur, 67000 Strasbourg, France. *Received May 29, 1975*

Twenty-two structural derivatives of clonidine [2-(2,6-dichlorophenylimino)imidazolidine] have been synthesized and their main physicochemical parameters (log P, ΔR_M, pK_a) determined. Quantitative correlations between the peripheral α-mimetic action (pithed rats) and physicochemical parameters pointed out the critical role of the steric effect in the ortho positions. On the other hand, attempted quantitative correlations between physicochemical parameters and central hypotensive activity were unsuccessful. These results are discussed in the light of the postulated mechanism of action of clonidine.

In view of our interest in the hypotensive agent clonidine [2-(2,6-dichlorophenylimino)imidazolidine],[2] we describe here quantitative structure–activity relationships (QSAR) obtained by the Hansch method[3] in a series of phenyliminoimidazolidines related to clonidine.

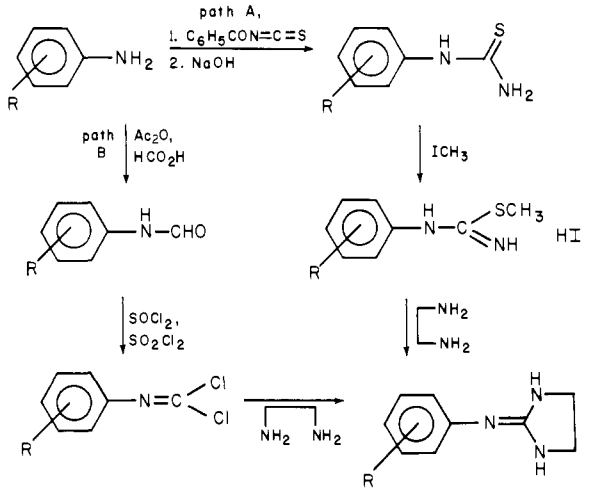
Clonidine is a hypotensive agent widely used therapeutically. It is well known, however, that its overall action on arterial blood pressure is a resultant of peripheral and central effects.[4] Peripheral effects are vasoconstrictive, whereas central effects are hypotensive. The central site of action is located in the medulla oblongata and probably more precisely in the obex[5] or on the ventral surface of the brain stem.[6]

The peripheral mechanism is explained by an α-sympathomimetic action. The central mechanism may be similar but one cannot rule out the possibility that clonidine acts by inhibiting adrenergic presynaptic receptors.

Our work is devoted to the synthesis of compounds related to clonidine, to the determination of their physicochemical parameters (log P, pK_a, ΔR_M), and to their pharmacological evaluation. In particular we have measured the hypertensive effect in the pithed rat (peripheral action). Our aim was to determine the physicochemical parameters which correlate best with the biological activity.

Synthesis of 2-Aryliminoimidazolidines. Aryliminoimidazolidines were generally synthesized by the action of ethylenediamine on the *S*-methylisothiouronium salt derivative (Scheme I, path A) and less commonly from the phenyldichloro isocyanide derivative (Scheme I, path B).

Scheme I



Starting materials were generally commercially available. The 2,6-difluoroaniline was obtained according to the method of Burton and Roe.[7] The 2,6-dimethyl-4-methoxyaniline was synthesized by some useful modifications of the method of Saunders and Watson[8] (see Scheme II).

The use of sulfanilic acid instead of aniline facilitated the isolation of the 2,6-dimethyl-4-methoxyaniline. It is also worthwhile mentioning that the original catalytic high-pressure hydrogenation was replaced by a chemical reduction using sodium hydrosulfite. The main physi-