Structure-Activity Correlations for Psychotomimetics. 1. Phenylalkylamines: Electronic, Volume, and Hydrophobicity Parameters

Brian W. Clare

School of Mathematical and Physical Sciences, Murdoch University, Murdoch, Western Australia 6150.
Received October 21, 1988

CNDO/2 calculations have been performed on a series of alkyl, alkoxy, and alkylthio derivatives of phenethylamine and phenylisopropylamine. The results of these, of van der Waals volume calculations, and of Hansch type hydrophobicity calculations were correlated with psychotomimetic activity by chemometric methods. Eight parameters, involving seven chemical descriptors, were found to be highly significant. Directional hydrophobicity and volume effects were found, which suggests that steric and hydrophobic interactions in the neighborhood of the receptor site are important. A puzzling but strong interaction effect between meta and para substituents was noted. Electronic terms may be explicable in terms of formation of charge-transfer complexes by accepting, rather than by donating, charge, as has been believed in the past. A charge effect indicates that a charge or dipole is influential at the binding site, or alternatively, a specific reactivity at the meta position is involved.

1. Introduction

In the years since the formulation of the "M substance" hypothesis for schizophrenia,[1-3] it has become apparent that the state induced by the "classical" hallucinogens mescaline, psilocybin, and LSD does not represent a model for schizophrenia in the sense that it does not precisely reproduce the symptoms of that disease.[4] Nevertheless, it seems certain that the two conditions at least share some symptoms in common.[5,6] Young[7] in a comparison of the schizophrenic and hallucinogen-induced states was able to conclude that "The LSD and schizophrenic experiences are similar in more ways than they are different", although he did find differences and also cast doubt on some prevalent opinions on the nature of these differences. In view of this, and of the fact that psychotomimetic substances and related materials have been reported in the body fluids of schizophrenic patients,[8,9] it seems likely that the receptors upon which the hallucinogens act are involved in the schizophrenic process, and thus any light shed upon these receptors will help to illuminate the schizophrenic psychosis itself. Discovery of a quantitative structure-activity relationship (QSAR) for substances which excite these receptors is a first step in learning the nature of these receptors. This is likely to prove difficult because, as evidenced by the variety of subjective responses to different drugs, more than one type of receptor must be involved in producing the psychic effects.

The resemblance or otherwise of the hallucinogen-induced state to schizophrenia, or to other abnormal states of mind, has been hotly debated since it was first proposed. There is no doubt, however, that the drugs powerfully influence emotion, affect, and perception, and a study of their mechanism of action in this area should help illuminate these facets of normal behavior. The receptors upon which they act presumably have some function in normal behavior, and a study of the drugs which affect these receptors may shed some light on the function and identity of the still unidentified natural agonist.

Snyder and Merril,[10] using Hückel molecular orbital theory, first reported a correlation of hallucinogenic activity with a quantum index, finding that high activity is associated with a high-energy HOMO (highest occupied molecular orbital) in a small number of phenylalkylamines, tryptamines, and LSD. Kang and Green[11,12] followed this with an intermediate neglect of differential overlap (INDO) study of LSD, 12 amphetamines, and 8 tryptamines and found a similar correlation. These authors also showed a steric correspondence between LSD and certain conformations of amphetamines and tryptamines. As with most recent QSAR studies, hydrophobicity has been shown to be important.[13] More recently, an extensive investigation by Anderson[14] has examined the electronic, conformational, and hydrophobicity correlations and reached similar conclusions. He did not combine the various effects into a single relationship. Numerous authors have correlated activity with gross molecular features.

I report here a computational study of some 63 compounds of the phenethylamine and phenylisopropylamine (amphetamine) classes, including alkoxy, alkyl, and alkylthio derivatives, and also the methylenedioxy compounds (1,3-benzodioxoles). This represents most of the compounds in these classes for which human data are available and is the largest group of psychotomimetics to be included in a single QSAR to date. Compounds containing bromine and iodine were not included, because of a limitation of the complete neglect of differential overlap (CNDO) program. The inclusion of sulfur compounds precluded the use of the INDO option.

The human data used in this study are principally those of A. T. Shulgin and co-workers, obtained from experiments conducted on themselves and on volunteers.[15] Because it is extremely difficult in most countries to obtain the necessary permission to conduct such research, this work remains largely unreplicated. The trials themselves pose great difficulties in methodology, as the powerful psychic effects of the drugs render a double blind study impossible, and uncontrollable factors such as the expec-

(1) Osmond, H.; Smythies, J. *J. Ment. Sci.* **1952**, *98*, 20.
(2) Hoffer, A.; Osmond, H.; Smythies, J. *J. Ment. Sci.* **1954**, *100*, 29.
(3) Osmond, H.; Hoffer, A. *J. Ment. Sci.* **1959**, *105*, 653.
(4) Angrist, B. *Neurochemistry and Pharmacology of Schizophrenia*; Handbook of Schizophrenia; Henn, F. A., Delisi, L. E., Ed.; Elsevier: Amsterdam, 1987; Vol. 2, Chapter 14.
(5) Guttmann, E. *J. Ment. Sci.* **1936**, *82*, 15.
(6) Tayleur Stockings, G. *J. Ment. Sci.* **1940**, *86*, 29.
(7) Young, B. G. *Br. J. Psychiatry* **1974**, *124*, 64.
(8) Delisi, L. E.; Wyatt, R. J. *Neurochemistry and Pharmacology of Schizophrenia*; Handbook of Schizophrenia; Henn, F. A., Delisi, L. E., Ed.; Elsevier: Amsterdam, 1987; Vol. 2, Chapter 13.
(9) McGeer, E. G.; McGeer, P. L. *J. Ment. Sci.* **1959**, *105*, 1.
(10) Snyder, S. H.; Merril, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *54*, 258.
(11) Kang, S.; Green, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1970**, *67*, 62.
(12) Kang, S.; Green, J. P. *Nature* **1970**, *226*, 645.
(13) Barfknecht, C. F.; Nichols, D. E. *J. Med. Chem.* **1975**, *18*, 208.
(14) Anderson, G. M., III. Structure-Activity Relationships in One-Ring Psychotomimetics. *Diss Abstr. Int. B*, **1982**, *43*, 1500 (Ph.D. Thesis, University of California, San Francisco).
(15) Shulgin, A. T. In *Handbook of Psychopharmacology*, Iversen, L. L., Iversen, S. D., and Snyder, S. H., Ed.; Plenum: New York, 1980; Vol. 11, pp 251, 254.

0022-2623/90/1833-0687$02.50/0

tations of the subject and the heightened suggestibility which the drugs induce greatly influence the course of the intoxication. For these reasons, one must have some reservations about the reliability of the human data.

Because of the highly subjective nature of the drug effects, animal data must also be treated with suspicion. A possible exception is the animal data obtained using the two-lever drug-discrimination test.[16] Of the large number of animal models which have been proposed, the latter is the most appealing. An animal (usually a rat) is trained to recognize and respond to the effect of the drug. Even this is open to question, in that the effect of the drug to which the animal responds remains unknown. One cannot assume that it is the effect which we recognize as "hallucinogenic"; we can only infer this from the correlation of the animal response with the human response to drugs which we know from human experiments to be hallucinogenic. Thus all the animal experiments refer back to the human data, and there can be no net gain in accuracy, only a gain in consistency and in experimental convenience. Animal data considered in this paper comes from the discrimination paradigm and is due mainly to R. A. Glennon and co-workers.[17]

It has long been recognized[18] that a common feature of the sympathomimetic group of hallucinogens (phenylalkylamines, tryptamines, ergolines, β-carbolines, ibogaine) is the occurrence of an aromatic nucleus (benzene or indole) separated from an amine nitrogen by two sp^3 carbon atoms. It is of interest, then, to find a relationship between the characteristics of the aromatic nucleus and the hallucinogenic activity of the compound. This is especially true for the phenylalkylamines because the unsubstituted molecule (i.e. phenethylamine or amphetamine) is devoid of hallucinogenic activity.

Some calculated quantities (e.g. orbital energies and dipole moments) refer to the molecule as a whole (global properties); others (e.g. net charge, superdelocalizabilities, frontier electron densities) refer to particular atoms in the molecule (local properties). Because only the phenethylamine nucleus itself is common to all of the phenylalkylamine hallucinogens, the local properties were considered for the atoms of this nucleus only, and in particular, only for the carbon atoms of the benzene ring. The only other such property to be used is the presence of a hydrogen atom or a methyl group on the α carbon atom.

It is proposed to find an equation which relates hallucinogenic activity to a linear combination of global properties and of local properties for these atoms only. It has been pointed out[19] that for local properties the coefficients for the two ortho positions must be equal, as must those for the two meta positions. The postulated linear model is, numbering the phenylalkylamine system in the usual way

$$\log A = C_2D_2 + C_3D_3 + C_4D_4 + C_5D_5 + C_6D_6$$

where A is the activity of the compound (the reciprocal of the effective dose), the C_i are constants for the series of compounds, and the D_i are local properties (descriptors), for example atom net charges, hydrophobicities, or van der Waals volumes of substituents. By the symmetry of the unchanging skeleton of the compounds in the study, C_2 must equal C_6, as must C_3 equal C_5; otherwise one would have the absurdity of the predicted activity of (2,3,4-trimethoxyphenyl)isopropylamine being not equal to that of the identical 4,5,6 "isomer". Thus, collecting terms, $\log A = C_2(D_2 + D_6) + C_3(D_3 + D_5) + C_4D_4$, or by an obvious change of notation, $\log A = C_o(D_{o1} + D_{o2}) + C_m(D_{m1} + D_{m2}) + C_pD_p$. However, on this model, compounds such as (2,3,4-trimethoxyphenyl)isopropylamine and the corresponding 2,4,5 isomer would have the same activity, in conflict with observation. If local properties are to account for this difference, they can do so only through interaction (product) terms. Such product terms, along with squared terms, make it possible to allow for nonlinearity in the QSAR. It is easily seen that an ortho–para or meta–para interaction does not relieve the problem; an ortho–meta interaction, however, does so. The interaction term is $(D_{o1} + D_{o2})(D_{m1} + D_{m2})$, which when multiplied out gives $D_{o1}D_{m1} + D_{o2}D_{m2} + D_{o1}D_{m2} + D_{o2}D_{m1}$. The first pair of terms, by the symmetry of the problem, must enter the equation with equal coefficients, as must the second pair. The former represents an along-the-ring interaction, the latter an across-the-ring effect. Such terms therefore were considered for net charges of the ring atoms and for substituent hydrophobicities and van der Waals volumes. The difference between isomers could also, of course, be attributale to global properties, for which such a problem does not arise.

At physiological pH, the amines would be partially protonated. It is not known whether the active species is the cation or the free base, but as the amine function is remote from the benzene ring by two saturated carbon atoms, it seems unlikely that there would be much interaction between them that wouuld not be constant between compounds. The pK_a of phenethylamines would be slightly different from phenylisopropylamines, and this difference would be incorporated in the indicator variable for the methyl group.

2. Calculations

Methods. Calculations were done on a Hewlett-Packard Integral PC using an Absoft FORTRAN compiler (F77HPUX3). The statistical and pattern-recognition software was written by the author.

The molecular orbital energies and net charges of the unprotonated amines were calculated by the CNDO/2 method.[20] Van der Waals volumes for substituents were calculated with spherical atoms, van der Waals radii as given by Pauling,[21] and the standard covalent bond lengths of Pople and Beveridge.[22] Allowance was made for overlap between bonded atoms only. Geometries were obtained by molecular mechanics (MMP2)[23] from starting geometries determined by the program EUCLID[24] from standard bond lengths and angles and with side chains in an extended conformation. It will be appreciated that these conformations are for the isolated molecule and that, in the absence of knowledge of the receptor site, the active conformation cannot be calculated. The descriptors chosen for this work are relatively insensitive to conformation.

(16) Appel, J. B.; Rosecrans, J. A. In *Hallucinogens: Neurochemical, Behavioral, and Clinical Perspectives*, Jacobs, B. L., Ed.; Raven: New York, 1984; pp 77–94.
(17) Glennon, R. A.; Rosecrans, J. A.; Young, R. *Med. Res. Rev.* **1983**, *3*, 289.
(18) Shulgin, A. T. In *Handbook of Psychopharmacology*, Iversen, L. L., Iversen, S. D., Snyder, S. H., Ed.; Plenum: New York, 1980; Vol. 11, p 259.
(19) Giles, D. E. (Murdoch University). Private communication, 1988.
(20) Dobosh, P. *QCPE* **1975**, *10*, 281, CNINDO/74. (Modified to accept 70 atoms and 160 orbitals.)
(21) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed., Cornell University: New York, 1960; p 260.
(22) Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill: New York, 1970; p 111.
(23) Allinger, N. L. MMP2(82). Available from QCPE, and also from Molecular Design Ltd., 1122 B Street, Hayward, CA 94541.
(24) Essen, H. *QCPE Bull.* **1983**, *3*, 13.

Hydrophobicities were determined by the method of Hansch and Leo[25] using the values quoted by Rekker.[26] Hydrophobicity is defined as log P, where P is the distribution coefficient for the substance between octanol and water and is obtained in this work by summing contributions from molecular fragments. For the substituents on a benzene ring, individual hydrophobicities are obtained similarly and may be related to the properties of parts of the receptor site, although physically, their significance is less clear than for a whole molecule. They are more likely to relate to local properties of the receptor site than to barriers which the molecule encounters before reaching that site. In the case of the methylenedioxy compounds, half of the hydrophobicity or van der Waals volume was assigned to each of the two attachment points. Calculated hydrophobicities for whole molecules in this study are less reliable than those for individual parts, owing to the effect of steric crowding, which is known to have a large influence with neighboring alkoxy groups on benzene rings. No allowance was made for this effect, as the interactions in these highly substituted compounds would be too complex for reliable estimation. Except where otherwise specified, all energies are in atomic units (Hartrees), charges are in protonic charges, and volumes are in cubic angstroms.

The properties to be considered are as follows: presence or absence of an α-methyl group, energies of the highest occupied and next highest occupied orbitals, energies of the lowest unoccupied and next lowest unoccupied orbitals, energies of the highest occupied and lowest unoccupied pi orbitals, magnitude and X, Y, and Z components of the dipole moment, sum of the hydrophobicities of the substituents on the ortho positions, sum of the hydrophobicities of the substituents on the meta positions, hydrophobicity of the substituent on the para position, total hydrophobicity, sum of van der Waals volumes of substituents on the ortho positions, sum of van der Waals volumes of substituents on the meta positions, van der Waals volume of substituent on the para position, total van der Waals volume, net charge on the vicinal carbon atom, sum of net charges on the ortho carbon atoms, sum of net charges on the meta carbon atoms, net charge on the para carbon atom, and total charge on the benzene ring.

Anderson,[14] and Anderson et al.[27] calculated conformations with the CNDO program and also by an ab initio method and obtained the result that one or both methoxy groups in o-dimethoxybenzene were out of the plane of the benzene ring. They also quoted considerable experimental evidence, from photoelectron spectra, vibrational spectra, and Kerr constant measurements, in support of this conclusion. With use of MMP2,[23] a planar geometry for o-dimethoxybenzene and for anisole is predicted. This program, developed by Allinger, is usually regarded as very reliable.[28] In spite of the evidence from Anderson et al. that MMP2 may fail in the case of o-dimethoxybenzene, this program was used unmodified in the present study, for two reasons. Firstly, it is not clear from Anderson's work how one could incorporate in MMP2 the ability to quantitatively predict the magnitude of this effect, and secondly, the conformation at the receptor site is in any case not necessarily that of minimum energy in vacuum. Even so, the anomaly must be borne in mind when interpreting the present results.

Table I. Nondefault MMP2 Parameters Used To Calculate Bond Lengths, Bond Angles, and Conformations

Torsional Parameters[a]

atom types[e]				V_1	V_2	V_3
5	1	15	2	0.000	0.000	0.530
2	2	15	1	2.800	1.950	−2.800
2	2	2	15	0.000	16.250	0.000
5	2	2	15	0.000	16.250	0.000
1	2	2	15	−1.200	16.250	0.000
6	2	2	15	−2.000	16.250	0.000
1	1	15	2	0.000	0.000	0.400
6	1	6	2	0.500	−0.420	0.000
2	1	6	2	0.000	0.000	0.400
2	2	9	7	0.000	1.900	0.000
2	2	2	9	−0.270	15.000	0.000
5	2	2	9	0.000	15.000	0.000
9	2	2	6	0.000	15.000	0.000

Stretching Parameters[b]

atom types		k_s	L_0
2	15	7.00	1.872
2	9	4.40	1.470
7	9	6.00	1.210

Bending Parameters[c]

atom types[e]			k_b	T_0
	2	15	0.050[d]	
2	2	15	0.700	120.000
1	15	2	1.100	97.500
2	2	9	5.000	120.000
2	9	7	5.000	118.000
7	9	7	5.000	124.000

[a] Torsion equation: $E_t = V_1(1 + \cos W)/2 + V_2(1 - \cos 2W)/2 + V_3(1 + \cos 3W)/2$, where V_1, V_2, and V_3 are the tabulated values and W the dihedral angle. [b] Stretching equation and parameters: $E_s = 71.94k_s(Dl)^2(1 + CS(Dl))$, where E_s is in kcal/mol/A^2, k_s is the stretch constant in md/A, Dl is the difference between the actual bondlength and L_0, the equilibrium bondlength, and CS = −2.00 and is the cubic stretch term. [c] Bending equation: $E_b = 0.021914k_bD^2(1 + SF\cdot D^4)$, where D is actual angle − T_0, E_b is in kcal/deg^2 per mol and k_b is in md A/rad^2. T_0 is the equilibrium angle. SF = 0.007×10^{-5} is the sextic bending term. [d] Out of plane bending. [e] Atom identification (atom type number, atom type): 1, sp^3 carbon; 2, sp^2 alkene carbon; 5, hydrogen; 6, oxygen C–O–C or C–O–H; 7, nitro oxygen; 9, nitro nitrogen; 15, sulfide sulfur.

In order to treat the molecules in this study, some extensions to the MMP2[23] force field were needed. In treating the methylenedioxy compounds, torsional parameters for the dihedral angle O (ether)–C (sp^3)–O (ether)–C (sp^2) had to be included. Values appropriate for the fourth atom being C (sp^3) were used. This would not introduce appreciable error in the geometry of such a compound, owing to the rigid nature of the (methylenedioxy)benzene moiety. For the aromatic thioethers, a number of parameters had to be introduced. Equilibrium bond lengths and bond angles, together with the corresponding stretching and bending constants, were obtained by perturbative configuration interaction using localized orbitals (PCILO)[29] calculations on thioanisole (methylthiobenzene). The torsional constant, however, could not be obtained in this way, as PCILO predicted a nonplanar geometry for this compound, in contrast to experimental data of Dewar et al.[30]

(25) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.
(26) Rekker, R. F. The Hydrophobic Fragmental Constant, Pharmacochemical Library, Elsevier: Amsterdam, 1977, Vol. 1, p 26.
(27) Anderson, G. M., III; Kollman, P. A.; Domelsmith, L. N.; Houk, K. N. *J. Am. Chem. Soc.* **1979**, *101*, 2344.
(28) Clark, T. *A Handbook of Computational Chemistry*, Wiley Interscience: New York, 1985; p 1.
(29) Boca, R. *QCPE* **1980**, *12*, 390, PCILO2. (Modified to accept 60 atoms and 140 orbitals.)
(30) Dewar, P. S.; Ernstbrunner, E.; Gilmore, J. R.; Godfrey, M.; Mellor, J. M. *Tetrahedron* **1974**, *30*, 2455.

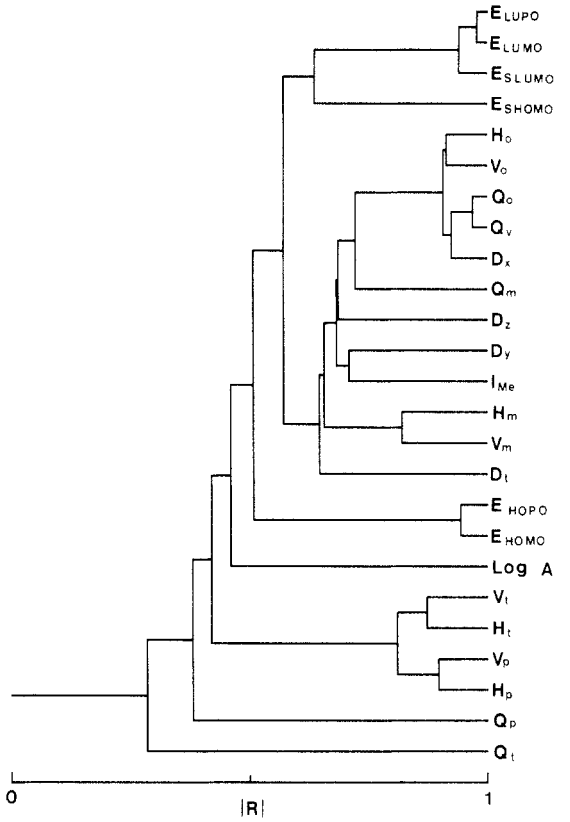


Figure 1. Hierarchical clustering of the correlation matrix of all calculated variables.

Consequently, the torsional parameters were adjusted to reproduce the rotation barrier, calculated from the partition ratio of approximately 0.1 for the nonplanar conformer, quoted by those authors. Remaining parameters were less critical, and values used were those for oxygen. The geometry of the one compound bearing a nitro group (DON) was predicted using bond lengths and angles for nitrobenzene,[31] high force constants (forcing the distances and angles to the values for nitrobenzene), and torsional constants designed to reproduce the experimental rotation barrier of 5.4 kcal/mol.[32] All nondefault MMP2 parameters used in this work are given in Table I. These parameters cannot be recommended for general use but should reproduce geometries of the molecules which are considered here to be of sufficient accuracy for our present purposes, since electronic properties, although dependent on conformation, are not critically so.

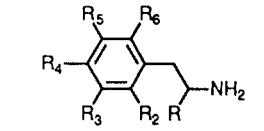
For the purposes of the CNDO calculation, the molecules were oriented so that the origin was midway between atom 1 of the ring (to which the alkylamine moiety is attached) and atom 4 (the para carbon atom). The positive X axis was in the direction of atom 1, and atom 2 lay on the XY plane, its Y coordinate being positive. The Z axis was normal to the ring, the Z coordinate of the amine nitrogen being positive. The compounds included in the study together with their human potencies, taken from the literature, are shown in Table II.

3. Results

The most relevant properties calculated by CNDO are shown in Table III, and the substituent properties are in Table IV. All of the data mentioned herein is available

(31) Sutton, L. E., Ed. *Interatomic Distances Supplement*; Special Publication 18; The Chemical Society: London, 1965; p 127.
(32) Varsanyi, G.; Holly, S.; Imre, L. *Spectrochim. Acta, Part A* **1967**, *23*, 1205.

Table II. Phenalkylamine Hallucinogens: Identity and Activity



Me = methyl
Et = ethyl
Pr = propyl
iPr = isopropyl
Bu = butyl
iBu = isobutyl
Am = 1-pentyl
Bzl = benzyl

drug	R	R_2	R_3	R_4	R_5	R_6	potency, MU	ref[a]
1	H	H	H	MeO	H	H	<1	a
2	H	H	MeO	MeO	H	H	<0.2	a
3	H	H	MeO	MeO	MeO	H	1	a
4	H	MeO	MeO	MeO	H	H	<1	a
5	H	MeO	H	MeO	MeO	H	<1	d
6	H	H	EtO	MeO	MeO	H	1	a
7	H	H	MeO	EtO	MeO	H	7	a
8	H	H	MeO	PrO	MeO	H	6	a
9	H	H	MeO	BuO	MeO	H	2	a
10	H	H	EtO	EtO	MeO	H	1.5	a
11	H	H	EtO	MeO	EtO	H	<1	a
12	H	H	EtO	EtO	EtO	H	<1	a
13	H	H	PrO	MeO	MeO	H	<1	a
14	H	H	MeO	$-OCH_2O-$		H	1	a
15	H	H	$-OCH_2O-$		H	H	1	a
16	H	MeO	H	Et	MeO	H	18	a
17	H	MeO	H	Me	MeO	H	20	a
18	H	H	MeO	MeO	MeS	H	6	b
19	H	H	MeO	MeS	MeO	H	12	b
20	H	H	MeO	MeO	EtS	H	6	b
21	H	H	EtO	MeS	MeO	H	4	b
22	H	H	EtO	MeO	MeS	H	<1	b
23	H	H	MeO	EtO	MeS	H	6	b
24	H	H	MeO	EtS	MeO	H	20	b
25	H	H	EtS	EtO	MeO	H	2	b
26	H	H	EtC	EtS	MeO	H	4	b
27	H	H	EtS	MeO	EtO	H	<1	b
28	H	H	EtO	MeS	EtO	H	<1	b
29	H	H	EtO	EtO	MeS	H	2	b
30	H	H	EtS	EtO	EtO	H	<1	b
31	H	H	EtO	EtS	EtO	H	<1	b
32	H	H	MeO	PrS	MeO	H	16	b
33	H	H	MeO	BuS	MeO	H	3	b
34	Me	H	H	MeO	H	H	5	d
35	Me	MeO	H	H	MeO	H	8	d
36	Me	MeO	H	MeO	H	H	5	d
37	Me	H	MeO	MeO	H	H	0.5	d
38	Me	H	MeO	MeO	MeO	H	2	a
39	Me	MeO	H	MeO	MeO	H	20	d
40	Me	MeO	MeO	MeO	H	H	2	c
41	Me	MeO	MeO	H	MeO	H	4	d
42	Me	MeO	MeO	H	H	MeO	10	c
43	Me	MeO	H	MeO	H	MeO	10	c
44	Me	MeO	H	EtO	MeO	H	20	c
45	Me	MeO	H	PrO	MeO	H	20	d
46	Me	H	Me	BzlO	MeO	H	2	d
47	Me	MeO	MeO	MeO	MeO	H	6	d
48	Me	H	$-OCH_2O-$		H	H	3	d
49	Me	$-OCH_2O-$		MeO	H	H	3	c
50	Me	H	MeO	$-OCH_2O-$		H	2.7	c
51	Me	MeO	H	$-OCH_2O-$		H	10	c
52	Me	MeO	$-OCH_2O-$		H	H	10	c
53	Me	$-OCH_2O-$		H	H	MeO	10	c
54	Me	MeO	$-OCH_2O-$		MeO	H	12	c
55	Me	MeO	MeO	$-OCH_2O-$		H	5	c
56	Me	MeO	H	Me	MeO	H	80	c
57	Me	MeO	H	Et	MeO	H	100	c
58	Me	MeO	H	Pr	MeO	H	80	d
59	Me	MeO	H	Bu	MeO	H	40	c
60	Me	MeO	H	iBu	MeO	H	20	c
61	Me	MeO	H	Am	MeO	H	10	c
62	Me	MeO	H	MeS	MeO	H	40	c
63	Me	MeO	H	iPrS	MeO	H	40	c

[a] (a) ref 34, (b) ref 35, (c) ref 36, (d) ref 37.

Table III. Important Electronic Variables for Compounds of Table II

	energies		charges				dipole
drug	HOPO	LUPO	vicinal	ortho	meta	para	X
1	−0.4267	0.1422	0.0206	0.0116	−0.0830	0.1716	0.141
2	−0.4095	0.1398	0.0396	−0.0753	0.1135	0.1244	1.713
3	−0.4162	0.1388	0.0617	−0.1632	0.3156	0.0719	3.378
4	−0.4194	0.1379	−0.0130	0.1298	0.0393	0.1381	0.055
5	−0.3921	0.1360	−0.0098	0.1206	0.0308	0.1438	0.106
6	−0.4145	0.1403	0.0614	−0.1643	0.3137	0.0713	3.269
7	−0.4131	0.1403	0.0605	−0.1623	0.3136	0.0708	3.126
8	−0.4121	0.1409	0.0604	−0.1626	0.3133	0.0708	3.113
9	−0.4114	0.1412	0.0601	−0.1622	0.3126	0.0713	3.102
10	−0.4113	0.1419	0.0597	−0.1646	0.3128	0.0708	3.004
11	−0.4126	0.1419	0.0606	−0.1665	0.3127	0.0714	3.152
12	−0.4096	0.1432	0.0594	−0.1652	0.3104	0.0704	2.918
13	−0.4142	0.1405	0.0614	−0.1645	0.3136	0.0710	3.242
14	−0.4037	0.1416	0.0586	−0.1559	0.3153	0.0656	1.763
15	−0.4074	0.1380	0.0386	−0.0691	0.1143	0.1173	0.208
16	−0.3991	0.1329	0.0009	0.1002	0.1001	0.0022	−0.535
17	−0.4000	0.1297	0.0039	0.0951	0.1088	0.0018	−0.483
18	−0.4101	0.1092	0.0393	−0.0639	0.1221	0.1366	3.385
19	−0.4203	0.0956	0.0881	−0.2033	0.4461	−0.1014	4.063
20	−0.4104	0.1056	0.0392	−0.0641	0.1200	0.1365	3.261
21	−0.4173	0.0975	0.0878	−0.2041	0.4439	−0.1010	3.960
22	−0.4087	0.1111	0.0392	−0.0650	0.1205	0.1360	3.277
23	−0.4079	0.1106	0.0385	−0.0635	0.1210	0.1352	3.117
24	−0.4362	0.0942	0.0868	−0.2034	0.4452	−0.1015	3.877
25	−0.4082	0.1064	0.0384	−0.0636	0.1188	0.1352	2.995
26	−0.4205	0.0954	0.0866	−0.2040	0.4429	−0.1012	3.781
27	−0.4086	0.1101	0.0393	−0.0655	0.1187	0.1357	3.168
28	−0.4413	0.0984	0.0872	−0.2057	0.4427	−0.0998	3.861
29	−0.4066	0.1119	0.0388	−0.0642	0.1193	0.1344	3.017
30	−0.4064	0.1108	0.0386	−0.0650	0.1178	0.1343	2.910
31	−0.4195	0.0966	0.0866	−0.2050	0.4409	−0.1012	3.689
32	−0.4209	0.0949	0.0865	−0.2035	0.4446	−0.1014	3.813
33	−0.4318	0.0952	0.0866	−0.2035	0.4447	−0.1018	3.773
34	−0.4248	0.1425	0.0199	0.0123	−0.0844	0.1717	0.053
35	−0.4084	0.1360	0.0114	0.0876	0.1386	−0.0274	−0.281
36	−0.4146	0.1423	−0.0294	0.2062	−0.1696	0.1911	−1.479
37	−0.4087	0.1397	0.0386	−0.0743	0.1118	0.1248	1.575
38	−0.4151	0.1389	0.0610	−0.1623	0.3142	0.0719	3.278
39	−0.3911	0.1366	−0.0105	0.1212	0.0295	0.1438	0.139
40	−0.4174	0.1386	−0.0140	0.1294	0.1393	0.1384	−0.086
41	−0.4122	0.1358	0.0277	0.0106	0.3482	−0.1039	−0.560
42	−0.4085	0.1337	−0.0413	0.2894	0.0719	−0.0307	−3.353
43	−0.4108	0.1452	−0.0249	0.3991	−0.2515	0.2087	−3.049
44	−0.3895	0.1384	−0.0112	0.1212	0.0282	0.1421	−0.141
45	−0.3890	0.1388	−0.0112	0.1223	0.0269	0.1420	−0.161
46	−0.4134	0.1293	0.0625	−0.1628	0.3195	0.0661	3.027
47	−0.4070	0.1320	0.0075	0.0403	0.2442	0.0869	1.498
48	−0.4062	0.1387	0.0378	−0.0680	0.1083	0.1173	0.103
49	−0.4156	0.1396	−0.0103	0.1262	0.1385	0.1430	0.050
50	−0.4028	0.1421	0.0579	−0.1549	0.3138	0.0657	1.632
51	−0.3896	0.1352	−0.0130	0.1290	0.0297	0.1365	−1.397
52	−0.4090	0.1375	−0.0141	0.1303	0.0401	0.1349	−0.474
53	−0.4032	0.1339	−0.0387	0.2823	0.0626	−0.0209	−1.885
54	−0.4021	0.1381	0.0053	0.0435	0.2443	0.0841	0.995
55	−0.3980	0.1341	0.0041	0.0516	0.2416	0.0798	−0.047
56	−0.3991	0.1296	0.0034	0.0948	0.1077	0.0019	−0.590
57	−0.3981	0.1333	0.0002	0.1011	0.0985	0.0023	−0.482
58	−0.3971	0.1337	−0.0002	0.1012	0.0982	−0.0003	−0.520
59	−0.3965	0.1339	−0.0005	0.1012	0.0979	−0.0001	−0.525
60	−0.3973	0.1331	0.0009	0.0997	0.1012	−0.0022	−0.579
61	−0.3961	0.1340	−0.0009	0.0990	0.0972	0.0003	−0.540
62	−0.3922	0.1127	0.0170	0.0781	0.1660	−0.0306	0.732
63	−0.3964	0.0928	0.0183	0.0771	0.1673	−0.0309	0.713

as supplementary material. The correlation structure of the complete data set is shown in the dendrogram[33] in Figure 1. As would be expected, Figure 1 shows a high correlation between hydrophobicities and van der Waals volumes. To simplify discussion, the following notation will be employed. Q refers to net charge, as calculated by CNDO, H refers to hydrophobicity, V refers to van der Waals volume, and D refers to dipole moment. E_{HOMO} and E_{LUMO} are the energies of the highest occupied and lowest unoccupied molecular orbitals, respectively. E_{SHOMO} and E_{SLUMO} are the energies of the next highest and next lowest orbitals, and E_{HOPO} and E_{LUPO} refer to the highest occupied and lowest unoccupied π orbitals. A subscript v, o, m, or p refer to the vicinal, ortho, meta, or para atom of the benzene ring. Similarly, a subscript x, y, or z refers to the x, y, or z component of the dipole moment. E_{L-H} is the difference in energy between the LUPO and HOPO, and

(33) Sokal, R. R.; Sneath, P. H. *Principles of Numerical Taxonomy*; Freeman: San Francisco, 1963; p 182.

Table IV. Hydrophobicity and van der Waals Volume Parameters of Substituents

radical	hydrophobicity	volume, A^3
methyl	0.56	18.6
ethyl	1.06	37.3
n-propyl	1.56	56.0
isopropyl	1.36	56.0
n-butyl	2.06	74.8
isobutyl	1.86	74.8
n-pentyl	2.56	93.5
methoxy	−0.02	28.7
ethoxy	0.48	47.5
n-propoxy	0.98	66.2
n-butoxy	1.48	84.9
methylenedioxy	−0.04[a]	13.8[a]
benzyloxy	2.69	112.6
methylthio	0.61	39.9
ethylthio	1.11	58.6
n-propylthio	1.61	77.3
isopropylthio	1.41	77.3
n-butylthio	2.11	96.1
nitro	−0.28	17.7

[a] To each attached ring atom.

I_{Me} is an indicator variable which takes the value 1 if there is a methyl group on the α carbon atom, and 0 otherwise. The total hydrophobicity or volume, or resultant dipole moment, is indicated by a subscript t.

The activity data, selected from the literature,[34-37] was arbitrarily divided into three classes. Activity, as is traditional for these substances, is measured in mescaline units (MU), that is, the ratio of the effective dose of mescaline to the effective dose of the substance in question and is denoted by *A*. Substances less active than mescaline (activity less than 1 MU) were classified as inactive. Indeed, most of these substances showed no activity in any dose tried. Substances having activities in the range from 1 to less than 10 MU were classified as having low activity, and those having activity of 10 MU or more were classified as highly active. An indication of the separability of the data can be obtained by the pattern-recognition techniques of nonlinear mapping (nlm)[38,39] and k nearest neighbors (knn).[40] Figure 2 shows the result of nonlinear mapping, using all the data, unit weighting, and Euclidean distances. It can be seen that there is considerable separation of the classes, especially the highly active compounds from the others. The knn technique was applied to the data for inactive and highly active compounds only, deleting the compounds of low activity, with *k* equal to five and weight inversely proportional to Euclidean distance. As a validation procedure, the data was repeatedly separated at random into a training set (approximately 85% of the data) and a test set (the remainder). In 30 trials, (137 classifications) 93% of the test set was correctly identified. Substances misclassified were compounds **24** (3,5-dimethoxy-4-(ethylthio)phenethylamine) **32** (3,5-dimethoxy-4-(isopropylthio)phenethylamine), both misclassified

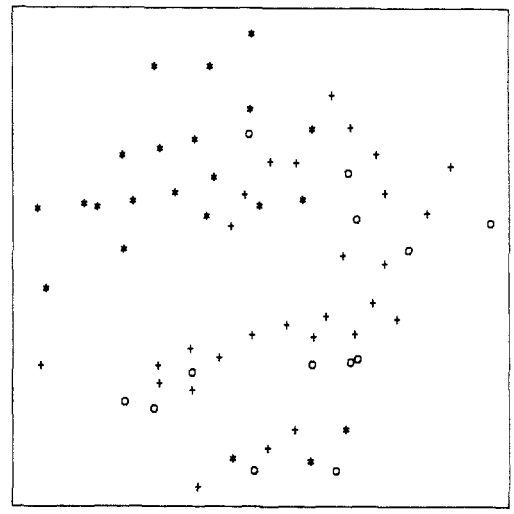


Figure 2. Results of nonlinear mapping calculations: O, inactive; +, low activity; *, high activity. Distances between compounds in variable space are preserved as well as possible in a least-squares sense in mapping from 24 space to 2 space. The activity variable is, of course, excluded. Parameters used: weight = 1.00, distance exponent = 2 (Euclidean).

three times as inactive, and 1 (*p*-methoxyphenethylamine) and **37** ((3,4-dimethoxyphenyl)isopropylamine, misclassified four times), each misclassified as highly active. Application of knn to the compound DON (see below, under regression in section 5) with the classifer trained on the inactive and highly active compounds led to a correct classification of highly active, with the score for the latter being 1.3 and the score for inactive being 0.3.

An examination of univariate statistics for the data is of interest, when compared with the results of a multivariate treatment. One-way analysis of variance between classes for each variable reveals 12 variables which differ between classes with more than 95% confidence. These are shown in Table V. Allowing for the number of descriptors tested (24), only those with probability less than 0.0021 (i.e. the first 10) should be regarded as significant at the 95% level. Electronic terms which are important include HOPO and HOMO energies, charge on the benzene ring adjacent to the amine side chain, and the *X* component of the dipole moment. The first of these is concordant with the work of previous investigators, whereas the last two indicate either the influence of a charge or dipole in the vicinity of the receptor or the influence of the reactivity of the molecule at this point. Hydrophobicity, interestingly, appears through individual substituent terms, rather than the value for the molecule as a whole. A similar situation exists for van der Waals volumes, where only the ortho and meta terms distinguish between the classes. These indications change, however, when we examine the data by multivariate methods, which give a different and more reliable picture of the data as a whole. As would be expected, E_{HOPO} and E_{LUPO} correlate strongly with E_{HOMO} and E_{LUMO}; indeed, except in the case of the sulfur compounds, they are identical.

4. Multiple Linear Regression

4.1 Human Data. With the large number of variables, and their expected interactions, it is obviously impossible to consider all possible models. Of the various ways of selecting variables for inclusion or rejection,[41] none guar-

(34) Gupta, S. P.; Singh, P.; Bindal, M. C. *Chem. Rev.* **1983**, *83*, 633.
(35) Jacob, P., III; Shulgin, A. T. *J. Med. Chem.* **1984**, *27*, 881.
(36) Nichols, D. E.; Glennon, R. A. In *Hallucinogens: Neurochemical, Behavioral, and Clinical Perspectives*, Jacobs, Barry L., Ed.; Raven: New York, 1984; pp 95–142.
(37) Shulgin, A. T. In *Handbook of Psychopharmacology*, Iversen, L. L., Iversen, S. D., Snyder, S. H., Eds.; Plenum: New York, 1980; Vol. 11, pp 243–333.
(38) Kowalski, B. R.; Bender, C. F. *J. Am. Chem. Soc.* **1972**, *94*, 5632.
(39) Kowalski, B. R.; Bender, C. F. *J. Am. Chem. Soc.* **1973**, *95*, 686.
(40) Varmuza, K. *Pattern Recognition in Chemistry*; Springer Verlag: Berlin, 1980; p 62 (Lecture Notes in Chemistry 21).
(41) Draper, N. R.; Smith, H. *Applied Regression Analysis*; John Wiley & Sons: New York, 1966; Chapter 6.

Table V. Univariate Statistics and Analysis of Variance Significant at 95% Level or Better

variable[a]	mean			standard deviation			F^b	prob[c]
	inactive	low A	high A	inactive	low A	high A		
volume (o)	4.415	6.111	27.98	10.78	11.31	14.30	23.666	0.00000
hydro (o)	−0.003	−0.006	−0.021	0.007	0.011	0.013	15.282	0.00000
hydro (m)	0.691	0.199	−0.026	0.622	0.386	0.013	14.675	0.00001
HOMO energy	−0.4118	−0.4117	−0.4013	0.008	0.006	0.010	12.267	0.00003
charge (o)	−0.076	−0.072	0.090	0.111	0.115	0.145	12.127	0.00004
dipole (x)	2.293	1.980	−0.017	1.409	1.668	1.911	10.899	0.00009
methyl	0.077	0.429	0.773	0.277	0.504	0.429	10.487	0.00012
volume (m)	68.42	51.67	29.86	38.44	25.85	14.00	9.791	0.00021
HOPO energy	−0.4135	−0.4121	−0.4019	0.012	0.007	0.012	8.437	0.00059
charge (v)	0.042	0.041	0.007	0.031	0.030	0.036	8.296	0.00066
dipole (y)	−1.388	−0.121	0.749	1.205	1.811	2.161	5.500	0.00641
dipole (z)	2.008	1.803	1.064	0.809	1.132	0.892	4.887	0.01080
hydro (p)	0.192	0.392	0.816	0.364	0.702	0.777	4.081	0.02180

[a] v = vicinal; o = ortho; m = meta; p = para; x, y, z = components of the dipole moment. [b] The Fisher variance ratio. [c] The corresponding probability resulting from one-way analysis of variance (ANOVA) between the three classes.

antees the statistical best choice, and indeed, the statistical best choice is not necessarily the best chemical model. The method used here was the stepwise regression algorithm of Efroymson[42] supplemented by an informal, nonautomated search based on the forward-selection and backward-elimination methods. All of the variables listed, and a large number of interactions and squared terms (approximately 80 in total) were tried in the fitting. The automated search typically either eliminated most of the variables or gave an equation having many statistically insignificant terms, depending on the levels set for Fisher's variance ratio (F) to enter and to leave. The insignificant terms were eliminated manually, one at a time. The more stringent settings resulted in a six-term equation having all terms significant at the 99.999% level.

Cubic terms were tried for total and para hydrophobicity and volume terms but were found to be nonsignificant, in contrast to the results of Nichols et al.[43]

In order to account for the apparent difference of effect of an α-methyl group between 3,4,5 and 2,4,5 series, an interaction between the I_{Me} variable and the hydrophobicity and volume terms was also sought.

After some experimentation, it becomes apparent that the major variables to be considered for the final model were I_{Me}, H_p, E_{HOPO}, E_{LUPO}, D_x, V_p, Q_v, Q_o, Q_m, H_t; the squared terms H_p^2, V_p^2, $E_{L\text{-}H}$; and the interactions H_mH_p and V_mV_p.

More doubtful terms were Q_p, H_m, V_m and the interactions H_oH_p, $I_{Me}V_o$, $I_{Me}H_p$, and V_oV_m. This last term was the only indication of an interaction which could account for the difference in activity between TMA-2 and TMA-3 (compounds **5** and **4**), and was an interaction between adjacent ortho and meta groups, rather than an "across-the-ring" interaction. The two preceding terms could account for the differential effect on the α-methyl.

In the regression, only compounds known to have hallucinogenic activity were included. Those substances classified as inactive, that is, those having activities less than 1 MU, were assigned activities equivalent to twice the highest dose reportedly tried and were used only for calculation of predicted activity, not for the determination of the regression plane.

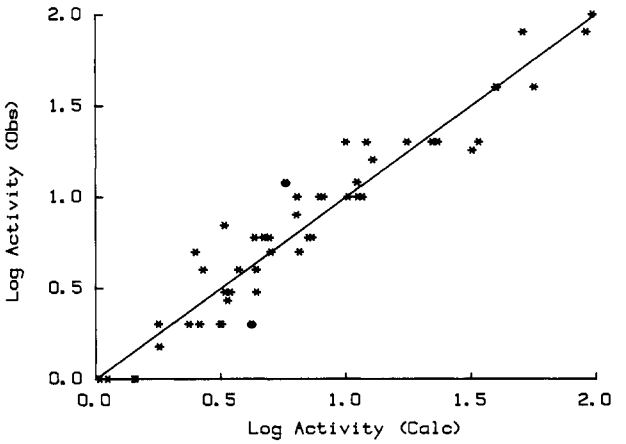


Figure 3. Plot of observed activity against calculated activity for active compounds.

After some consideration, the model finally selected as most plausible and most parsimonious was

$$\log A = C_1I_{Me} + C_2H_p + C_3E_{L\text{-}H} + C_4H_p^2 + C_5H_mH_p + C_6V_mV_p + C_7V_p^2 + C_8Q_{m\text{-}o} + C_9$$

	1	2	3	4	5
C	0.474	1.348	−10.894	−0.270	−1.885
s	0.077	0.112	1.759	0.056	0.256
P	0.00000	0.00000	0.00000	0.00002	0.00000

	6	7	8	9
C	0.000211	−0.000278	−0.676	6.289
s	0.000058	0.000046	0.129	0.963
P	0.00072	0.00000	0.00001	0.00000

$N = 50$ $R = 0.9563$ $S = 0.1647$
$F = 54.858$ $P' = 0.00000$ PRESS = 1.56 (1)

Here C is the regression coefficient corresponding to each term, s is the standard error for each term, P is the probability of obtaining a value this different from zero by chance (the significance, from a t test), N is the number of compounds in the regression, R is the multiple correlation coefficient, S is the standard error of estimate, F is the variance ratio for the regression, and P' is the significance of the regression as a whole (from the F value). PRESS is the "prediction error sum of squares"[43] calculated on the residuals and is a measure of the ability of the equation to predict, based on data which has not been used in the derivation of the relationship. A plot of the relationship given in eq 1 is presented in Figure 3. This shows a good linearity and strong clustering of the points around the fitted line.

(42) Efroymson, M. A. In *Mathematical Methods for Digital Computers*, Ralston, A., and Wilf, H. S., Eds.; John Wiley & Sons: New York, 1962; p 191.
(43) Nichols, D. E.; Shulgin, A. T.; Dyer, D. C. *Life Sci.* **1977**, *21*, 569.
(44) Younger, M. S. *A Handbook of Linear Regression*; Duxbury: North Scituate, 1979; p 483.

Of the terms D_x, Q_m, Q_o, and Q_v, only one could be supported in the model, the nominally best being Q_m, as may be seen from

$$\log A = C_1 I_{Me} + C_2 H_p + C_3 D_x + C_4 E_{L\text{-}H} + C_5 H_p^2 + C_6 H_m H_p + C_7 V_m V_p + C_8 V_p^2 + C_9$$

	1	2	3	4	5
C	0.477	1.269	−0.0823	−12.09	−0.294
s	0.096	0.140	0.0339	2.14	0.068
P	0.00001	0.00000	0.01961	0.00000	0.00009

	6	7	8	9
C	−1.675	0.000183	−0.000234	6.913
s	0.314	0.000082	0.000054	1.175
P	0.00000	0.03089	0.00009	0.00009

$N = 50$ $R = 0.9356$ $S = 0.1990$
$F = 35.969$ $P' = 0.00000$ PRESS = 2.29 (2)

$$\log A = C_1 I_{Me} + C_2 H_p + C_3 Q_m + C_4 E_{L\text{-}H} + C_5 H_p^2 + C_6 H_m H_p + C_7 V_m V_p + C_8 V_p^2 + C_9$$

	1	2	3	4	5
C	0.530	1.393	−1.082	−11.400	−0.278
s	0.078	0.114	0.218	1.796	0.057
P	0.00000	0.00000	0.00001	0.00000	0.00002

	6	7	8	9
C	−1.903	0.000198	−0.000282	6.622
s	0.265	0.000058	0.000047	0.984
P	0.00000	0.00151	0.00000	0.00000

$N = 50$ $R = 0.9543$ $S = 0.1683$
$F = 52.298$ $P' = 0.00000$ PRESS = 1.627 (3)

$$\log A = C_1 I_{Me} + C_2 H_p + C_3 Q_o + C_4 E_{L\text{-}H} + C_5 H_p^2 + C_6 H_m H_p + C_7 V_m V_p + C_8 V_p^2 + C_9$$

	1	2	3	4	5
C	0.430	1.301	1.229	−10.438	−0.276
s	0.086	0.121	0.291	1.906	0.061
P	0.00001	0.00000	0.00013	0.00000	0.00005

	6	7	8	9
C	−1.693	0.000173	−0.000245	5.979
s	0.266	0.000060	0.000048	1.046
P	0.00000	0.00649	0.00001	0.00000

$N = 50$ $R = 0.9490$ $S = 0.1776$
$F = 46.43$ $P' = 0.00000$ PRESS = 1.81 (4)

A marginally better fit was given by the difference of the two charge terms, $Q_{m\text{-}o} = Q_m - Q_o$. A less satisfactory correlation could be obtained with the charge on the vicinal and para carbon atoms. The sign of the coefficient of the charge term alternated in the order vicinal, ortho, meta, to para. This was due to the fact that the charges themselves were strongly correlated in this manner, as shown by the heirarchical dendrogram in Figure 4. The dipole term (in eq 2) suggests that a shift of negative charge in the ring toward the ethylamine moiety favors high activity, while the charge term (in eq 1 and 3) suggests a specific involvement of the meta position, with a high negative charge on one or both of these carbon atoms favoring high hallucinogenic activity.

A linear dependency may be expected for the hydrophobicity and volume terms also, because of their strong mutual correlations, but it was found that the data would support both, and that the total molecular hydrophobicity gave a poorer correlation than that for the para substituent alone, as was reported by Nichols et al.[43]

$$\log A = C_1 I_{Me} + C_2 H_t + C_3 Q_{m\text{-}o} + C_4 V_t + C_5 E_{L\text{-}H} + C_6 H_t^2 + C_7 H_m H_p + C_8 V_t^2 + C_9$$

	1	2	3	4	5
C	0.379	0.930	−0.356	0.0644	−12.30
s	0.105	0.188	0.162	0.0205	2.74
P	0.0086	0.00001	0.0335	0.0322	0.00006

	6	7	8	9
C	−0.194	−1.513	−0.000144	−0.019
s	0.090	0.276	0.000044	2.840
P	0.03656	0.00000	0.00204	0.995

$N = 50$ $R = 0.9004$ $S = 0.2451$
$F = 21.952$ $P' = 0.00000$ PRESS = 3.59 (5)

The HOMO energy has long been believed to be the dominant predictor of hallucinogenic activity, and this would appear to be confirmed by the correlation coefficient of approximately 0.46, as reflected by the cluster level in Figure 1. The logarithm of the activity joins at this level a cluster comprising the HOMO and HOPO energies on one hand and many other variables, including the LUMO and LUPO energies, on the other. The multiple regression however does not support the conclusion that the HOMO is the most important contributor to activity. In eq 6, the

$$\log A = C_1 I_{Me} + C_2 E_{HOPO} + C_3 H_p + C_4 H_p^2 + C_5 H_m H_p + C_6 V_m V_p + C_7 V_p^2 + C_8 Q_{m\text{-}o} + C_9$$

	1	2	3	4	5
C	0.494	3.088	1.612	−0.334	−1.652
s	0.108	3.950	0.146	0.077	0.350
P	0.00005	0.43768	0.00000	0.00012	0.00003

	6	7	8	9
C	0.000241	−0.000305	−0.677	1.614
s	0.000080	0.000063	0.184	1.617
P	0.00431	0.00002	0.00068	0.32417

$N = 50$ $R = 0.9150$ $S = 0.2273$
$F = 26.356$ $P' = 0.00000$ PRESS = 3.15 (6)

$$\log A = C_1 I_{Me} + C_2 E_{LUPO} + C_3 H_p + C_4 H_p^2 + C_5 H_m H_p + C_6 V_m V_p + C_7 V_p^2 + C_8 Q_{m\text{-}o} + C_9$$

	1	2	3	4	5
C	0.536	−10.815	1.472	−0.296	−1.834
s	0.081	1.952	0.111	0.059	0.269
P	0.00000	0.00000	0.00000	0.00001	0.00000

	6	7	8	9
C	0.000203	−0.000290	−0.805	1.825
s	0.000061	0.000048	0.136	0.279
P	0.00187	0.00000	0.00000	0.00000

$N = 50$ $R = 0.9516$ $S = 0.1732$
$F = 49.115$ $P' = 0.00000$ PRESS = 1.67 (7)

$$\log A = C_1 I_{Me} + C_2 E_{HOPO} + C_3 E_{LUPO} + C_4 H_p + C_5 H_p^2 + C_6 H_m H_p + C_7 V_m V_p + C_8 V_p^2 + C_9 Q_{m\text{-}o} + C_{10}$$

	1	2	3	4	5
C	0.4959	7.466	−11.970	1.369	−0.274
s	0.0778	2.898	1.885	0.111	0.056
P	0.00000	0.01380	0.00000	0.00000	0.00002

	6	7	8	9	10
C	−1.888	0.000205	0.000279	−0.726	5.038
s	0.252	0.000057	0.000045	0.131	1.275
P	0.00000	0.00090	0.00000	0.00000	0.00031

$N = 50$ $R = 0.9586$ $S = 0.1624$
$F = 50.379$ $P' = 0.00000$ PRESS = 1.53 (8)

coefficient of the HOPO energy has become nonsignificant, in the presence of the other predictors. Although the LUPO energy is better in this respect, the best result is obtained from the difference between the LUPO and

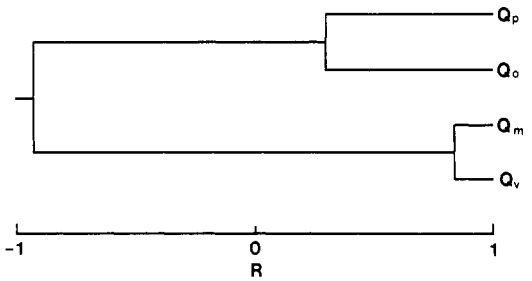


Figure 4. Hierarchical clustering of the correlation matrix for net charge on the carbon atoms of the benzene rings of the compounds in Table II.

HOPO energies. It may be seen from the above that the HOPO and LUPO energies, when both are in the equation, have coefficients which are of opposite sign, and are of comparable magnitude considering the errors, the HOPO energy having again become significant, justifying their combination as $E_{L\text{-}H}$ in eq 1. This matter is discussed at greater length below (section 7). The equation of minimal complexity, in which all coefficients are of extremely high significance, is

$$\log A = C_1 I_{Me} + C_2 H_p + C_3 H_m H_p + C_4 H_p^2 + C_5 V_p^2 + C_6 E_{L\text{-}H} + C_7$$

	1	2	3	4
C	0.4882	1.348	−1.225	−0.333
s	0.071	0.141	0.242	0.068
P	0.00000	0.00000	0.00001	0.00001

	5	6	7
C	−0.000162	−11.41	6.600
s	0.000033	2.226	1.213
P	0.00001	0.00001	0.00000

$N = 50$ $R = 0.9250$ $S = 0.209$
$F = 42.492$ $P' = 0.00000$ PRESS = 2.40 (9)

This equation has lower predictive ability than most of the above and consequently probably does not fully extract the information content of the data.

The statistically best equation which included an interaction between the α-methyl group and a volume or hydrophobicity term was the following:

$$\log A = C_1 I_{Me} + C_2 H_p + C_3 H_m H_p + C_4 E_{L\text{-}H} + C_5 I_{Me} V_o + C_6 H_p^2 + C_7 V_p^2 + C_8$$

	1	2	3	4
C	0.3176	1.2913	−1.1897	−10.779
s	0.0955	0.1351	0.2291	2.118
P	0.0018	0.00000	0.00001	0.00000

	5	6	7	8
C	0.00664	−0.31875	−0.000151	6.259
s	0.00266	0.06429	0.000031	1.153
P	0.01673	0.00001	0.00002	0.00000

$N = 50$ $R = 0.9350$ $S = 0.197$
$F = 41.723$ $P' = 0.00000$ PRESS = 2.35 (10)

This equation is not as good a predictor as eq 1, and its coefficients are not as significant.

4.2 Animal Data. The activities of a number of drugs in the discrimination experiments, taken from the literature,[17] are shown in Table VI. These represent most of the available data for this group of compounds. Only a small number, those for which human data are not also available, are excluded. The doses in Table VI are ED_{50} for generalization to 1 mg/kg DOM (compound **56**), that is, the dose of the test drug which causes 50% of the animals which were trained on 1 mg/kg DOM to respond as

Table VI. Effective Dose of Some Compounds in Table II in Rats Trained on 1 mg/kg DOM (Compound **56**) in Discrimination Procedure

drug	ED_{50}, mg/kg	drug	ED_{50}, mg/kg	drug	ED_{50}, mg/kg
3	14.64	40	7.8	50	3.66
17	1.31	41	16.48	56	0.44
35	5.51	43	3.69	57	0.23
36	4.88	44	6.33	58	0.17
38	6.34	48	1.68	59	0.91
39	3.59				

they would to the training drug. Defining A_R as the reciprocal of this dose, regression of log A against log A_R gave a correlation coefficient of 0.844 and an F value of 34.711, indicating a very highly significant regression, but much scatter. A plot of the relationship is shown in Figure 5.

A backward-elimination regression of the data in Table VI using the variables of eq 1, and retaining terms significant at the 95% level, resulted in eq 11. Thus most

$$\log A_R = C_1 H_p + C_2 V_p^2$$

	1	2
C	1.416	0.254
s	−0.000384	0.000114
P	0.00009	0.00516

$N = 16$ $R = 0.8679$ $S = 0.321$
$F = 19.833$ $P' = 0.00011$ PRESS = 2.28 (11)

of the terms of eq 1 are lost. This is only partly due to the much smaller data set for this regression, as shown by the following equation, derived from the human activities for the same drugs, using the same method.

$$\log A = C_1 I_{Me} + C_2 E_{H\text{-}L} + C_3 Q_{m\text{-}o} + C_4 V_p^2 + C_5 H_m H_p + C_6 V_m V_p + C_7$$

	1	2	3	4
C	0.395	−18.37	−0.979	−0.000433
s	0.129	3.920	0.201	0.000087
P	0.01385	0.00114	0.00090	0.00076

	5	6	7
C	−54.510	0.000301	10.49
s	7.346	0.000126	2.15
P	0.00004	0.04092	0.00087

$N = 16$ $R = 0.9825$ $S = 0.149$
$F = 41.612$ $P' = 0.00000$ PRESS = 0.58 (12)

5. Analysis of Regression Results

On the basis of eq 1, compounds having an α-methyl group are 0.929 standard deviations (0.48 log units) more active than those without, a factor of 3.0 in activity. It is often stated[45] that a methyl group is more effective in 2,4,5-substituted compounds than in their 3,4,5 isomers in increasing potency. Five pairs of compounds in the present data set allow a test of this hypothesis. These are the 3,4,5 compounds **3** and **38** (potency ratio = 2.0), and **14** and **50** (ratio = 2.7), and the three 2,4,5 pairs **17**, **56** (ratio = 4.0), **16**, **57** (ratio = 5.6), and **5**, **39** (compound **5** is inactive, ratio is at least 20). Deleting the last pair and applying the unpaired t test (from any elementary statistics text) one obtains a significance (α) value of 0.107—that is the difference between the means is not significant, even at 90% confidence. Inclusion of the fifth pair at ratio

(45) Nichols, D. E.; Glennon, R. A. In *Hallucinogens: Neurochemical, Behavioral, and Clinical Perspectives*; Jacobs, B. L., Ed.; Raven: New York, 1984; p 107.

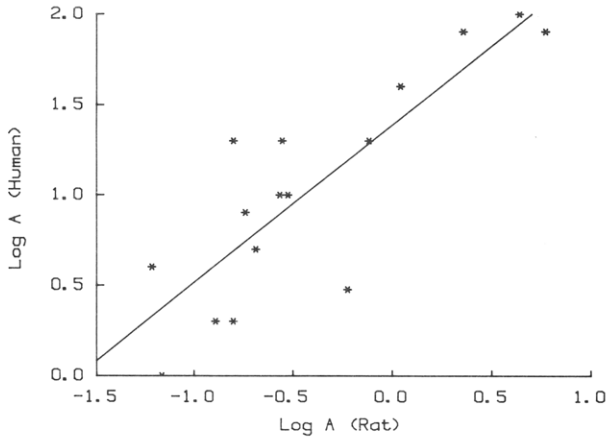


Figure 5. Plot of human activity (Shulgin) against discriminability in rat (Glennon).

values of 20 and 40 does not improve this result. Admittedly, with such a small sample, this is a very insensitive test, and a different result may be obtained if a larger number of drugs were used. Human data for such a test seem to be unavailable.

The effect of charge is confounded with that of dipole moment. A high negative charge on the meta carbon atoms favors high activity. The remaining atoms become implicated by way of the correlation of their charges with those of the meta atoms, the sign of the correlation alternating around the ring. Overall, the charges sum to produce an effect in the X component of the dipole moment which correlates with the hallucinogenic effect of the drug and also with the charge on the meta carbon atoms. Although the highest correlation coefficients are obtained with the charge on this pair of atoms, the difference is not great enough to convincingly demonstrate that these two atoms and not the general field of the molecule is involved.

The effect of frontier orbital energy also leaves some ambiguity. The best correlation is with the energy difference of the HOPO and LUPO. The involvement turns out to be a little unsymmetrical; however, the LUPO seems to play a larger part than the HOPO. A smaller difference between HOPO and LUPO energies favors higher activity, as does a lower LUPO energy.

Both the above effects are linear as far as can be determined from the present data. Volume and hydrophobicity effects, however, are nonlinear. There is an interaction between the meta and para substituents, with respect to both volume and hydrophobicity. The hydrophobicity interaction is highly significant, the volume less so. There is no great influence from either volume or hydrophobicity of groups in the ortho position. These conclusions, of course, are valid only within the limitations of the present data set and must not be extrapolated. Figure 6 shows three-dimensional plots of the volume and hydrophobicity contributions to the activity, as a function of the meta and para components, over the range of the data.

Application of the Fisher transformation[46] to the multiple correlation coefficient of eq 1 gives a 95% confidence interval of 0.922–0.975, so on this basis, only eq 5 explains a insufficient amount of the variance. The other equations do, however, have highly significant explanatory power, based on the P values quoted, calculated from t statistics.

Obviously, in a model containing strongly related variables such as this one, multicollinearity is likely to pose

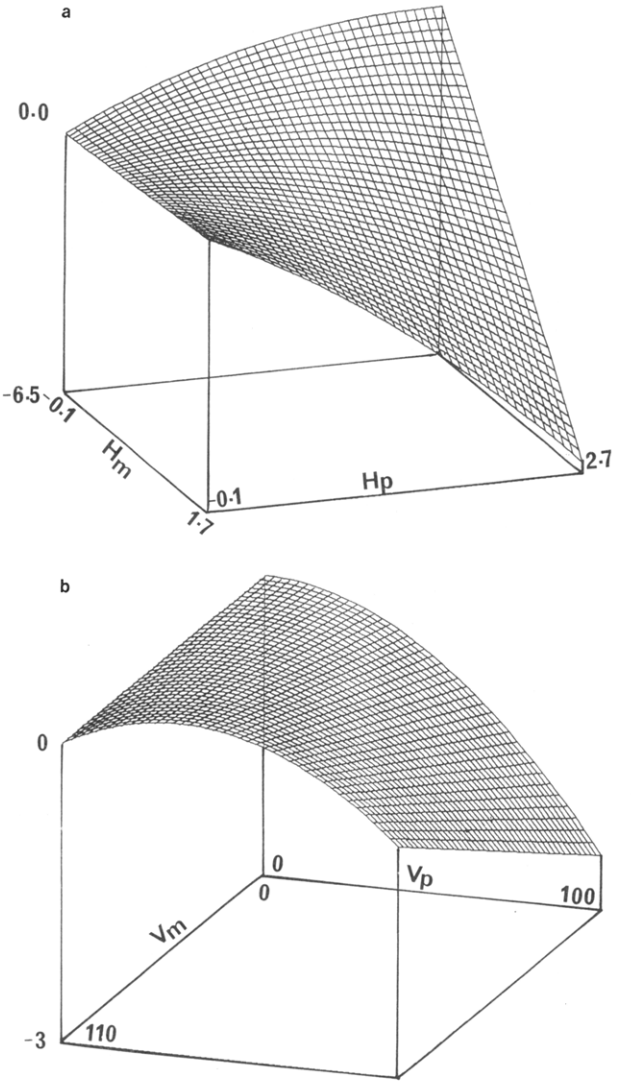


Figure 6. The meta–para interaction in (a) hydrophobicity and (b) van der Walls volume represented as three dimensional plots. The vertical axis is contribution to logarithm of activity from the hydrophobicity or volume terms, calculated from eq 1: (a) terms 2, 4, and 5 and (b) terms 6 and 7.

a problem. All of the models 1–9 were subjected to principal components analysis (PCA). With the exception of model 9, the ratio of the sum of the reciprocals of the eigenvalues of the correlation matrix to the number of parameters fell in the range 9.7–11.6. This is evidence for considerable linear dependence, a value of 5 being more generally acceptable.[47] Equation 9 gave a value of 6.72 for this parameter. Inspection of the PCA results showed that in each case the dominant terms were two eigenvalues of approximately 0.04 and 0.02. These are not excessively small, and the second of them alone is enough to inflate the ratio beyond 5 in each case. The eigenvectors corresponding to them have large coefficients for V_p^2, V_mV_p, and H_p^2 for the smaller of the two, and H_p^2 and H_p for the larger. This indicates that the source of the collinearity lies with these variables, which would be expected in any case. Equation 9 contains only the lesser of these two collinearities. Table VII gives the eigenvalues and eigenvectors of the correlation matrix for the variables in eq 1.

Ridge regression is also a useful technique for analysis of collinear data.[48] Applied to the data and eq 1, it ap-

(46) Afifi, A. A.; Azen, S. P. *Statistical Analysis, A Computer Oriented Approach*; Academic: New York, 1979; p 162.

(47) Chatterjee, S.; Price, B. *Regression Analysis by Example*; John Wiley & Sons: New York, 1977; p 200.

Table VII. Eigenvalues and Eigenvectors for Correlation Matrix of Variables in Eq 1

component	1	2	3	4	5	6	7	8
eigenvalue	3.77	2.03	1.07	0.589	0.332	0.142	0.045	0.021
I_{Me}	0.182	0.551	0.243	0.160	0.731	0.151	−0.120	−0.087
H_p	−0.468	0.233	0.095	0.019	−0.134	−0.416	−0.724	−0.052
H_mH_p	−0.029	−0.499	0.499	0.661	0.146	−0.257	0.042	0.130
V_mV_p	−0.462	−0.227	−0.060	0.193	0.053	0.634	−0.113	−0.526
H_p^2	−0.436	0.338	0.002	0.113	−0.024	−0.391	0.647	−0.334
V_p^2	−0.479	0.213	−0.033	0.140	0.018	0.336	0.134	0.757
Q_d	−0.271	−0.399	−0.434	−0.298	0.644	−0.256	−0.003	0.105
E_{L-H}	0.200	0.153	−0.733	0.616	−0.059	−0.078	−0.104	−0.010

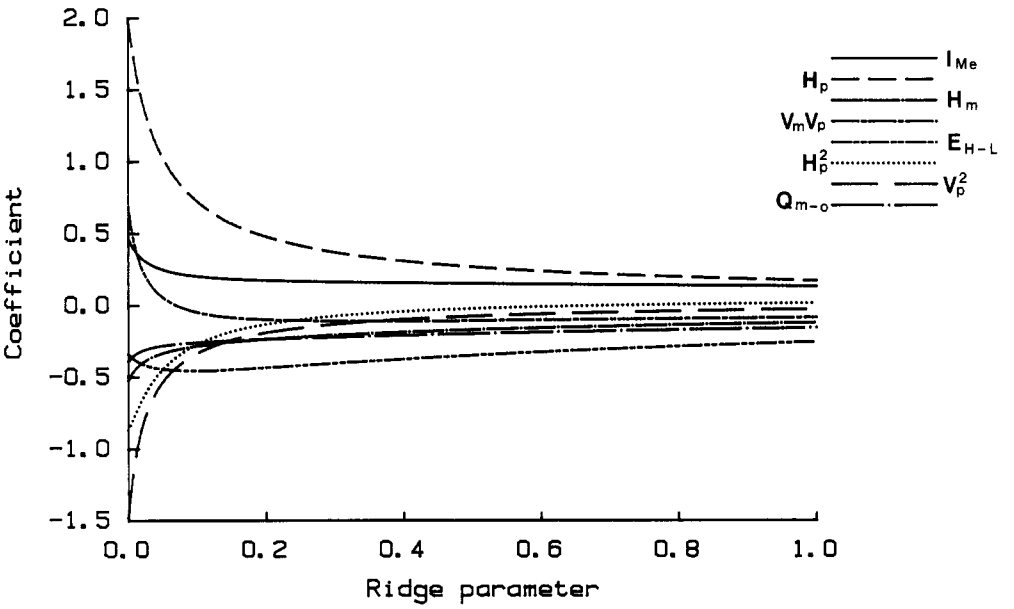


Figure 7. Ridge traces of standardized coefficients of variables in eq 1.

pears that the coefficients which become unstable are those for H_p, H_p^2,, V_mV_p, and V_p^2, largely confirming the principal components analysis results. The corresponding variance inflating factors are also large for these terms. A plot of the ridge trace for the data and eq 1 is presented in Figure 7. On the basis of Figure 7, the volume interaction term must be regarded as doubtful, as its coefficient tends to zero, and the para hydrophobicity linear and quadratic terms are somewhat unstable. The coefficients corresponding to the α-methyl, orbital energy, net charge, and the hydrophobicity interaction terms are very stable. This indicates that the values obtained for these latter terms are not strongly dependent on the choice of data and may therefore be considered reliable.

Inspection of the coefficients and errors in eq 1–10 reveals a variation of only 10% in corresponding terms and error terms which are small compared with the coefficients. This and the ridge regression and principal component results indicate that multicollinearity, although present, is not a major influence in this data set. Of the equations considered, eq 8 is the best predictor, but as it only marginally surpasses eq 1, the latter is to be preferred, having one less term.

When the variables are standardized, a regression equation which contains no constant term can be constructed. Terms involving specific chemical effects can then be collected, as is done in Table VIII. Here, the (standardized) logarithm of the activity is broken into five parts: an orbital energy contribution and contributions from the α-methyl group, charge, volume, and hydrophobicity effects. A positive term tends to make the compound more potent than the mean. These effects were calculated from eq 1. This is possible because no interactions across types of descriptor were found.

It may be seen that compound **3** (mescaline), in addition to lacking an α-methyl group, has a deactivating contribution from orbital energy and charge. These together contribute −0.8 standard deviations to activity. The terms for volume and hydrophobicity partially cancel (as for many of these compounds) but their sum is deactivating. For compounds **6–8**, the orbital and charge effects are similar to those of compound **3**, but the hydrophobicity increasingly outweighs the volume contribution. A high hydrophobicity contribution is obtained with a bulky alkyl, alkoxy, or alkylthio group in the 4-position. This is outweighed by the diminishing volume contribution when the alkyl becomes bulkier than propyl. With the 2,4,5-substitution pattern, the orbital-energy contribution is more favorable (compounds **16**, **17**, **56–61**), whereas in the thioethers (compounds **18–33**, **62**, **63**) it is especially so. With this substitution pattern, charge is also more favorable, the two contributing 0.25 standard deviations to activity. In 2,4-DMA (**36**) and 2,4,6-TMA (**43**) the orbital contribution is at its least, but this is more than compensated by the charge, particularly in the latter compound. In the tetramethoxy compound **47** the orbital energy contribution is small, but not as small as in **36** and **43**, and this is not compensated by charge. This results in low activity for this compound. In DOM (**56**), hydrophobicity and volume both contribute strongly to activity. As the alkyl chain lengthens, the hydrophobicity becomes more favorable, but the volume becomes even more strongly unfavorable, the activity passing through a maximum in DOET (**57**) and declining rapidly, until in DOAM (**61**) it is only $^1/_{10}$ of its maximum value.

(48) Chatterjee, S.; Price, B. *Regression Analysis by Example*; John Wiley & Sons: New York, 1977; Chapter 8.

Table VIII. Breakdown of Hallucinogenic Activity into Components[a]

drug	methyl	orbital	charge	volume	hydro	log activity		
						calcd	obsd	residual
1	−0.534	−0.707	0.363	0.025	−0.979	−0.095	−0.301	−0.206
2	−0.534	−0.293	−0.009	0.364	−0.980	0.100	−1.000	−1.100
3	−0.534	−0.413	−0.390	0.702	−0.982	0.016	0.000	−0.016
4	−0.534	−0.462	0.358	0.364	−0.980	0.202	−0.301	−0.503
5	−0.534	0.155	0.357	0.364	−0.980	0.519	−0.301	−0.820
6	−0.534	−0.409	−0.389	0.923	−0.945	0.151	0.000	−0.151
7	−0.534	−0.379	−0.386	0.373	0.279	0.515	0.845	0.330
8	−0.534	−0.371	−0.386	−0.332	1.277	0.671	0.778	0.108
9	−0.534	−0.362	−0.384	−1.413	2.013	0.498	0.301	−0.196
10	−0.534	−0.375	−0.388	0.740	−0.599	0.253	0.176	−0.077
11	−0.534	−0.402	−0.390	1.145	−0.909	0.287	−0.301	−0.588
12	−0.534	−0.367	−0.385	1.107	−1.477	−0.005	−0.301	−0.296
13	−0.534	−0.364	−0.389	1.144	−0.909	0.307	−0.301	−0.608
14	−0.534	−0.208	−0.380	0.607	−1.041	0.047	0.000	−0.047
15	−0.534	−0.210	−0.002	0.445	−1.038	0.159	0.000	−0.159
16	−0.534	0.073	0.239	0.159	1.335	1.504	1.255	−0.249
17	−0.534	0.121	0.221	0.502	0.457	1.244	1.301	0.057
18	−0.534	0.341	−0.005	0.834	−0.936	0.694	0.778	0.084
19	−0.534	0.413	−0.614	0.552	0.563	1.045	1.079	0.034
20	−0.534	0.411	−0.003	1.054	−0.899	0.864	0.778	−0.085
21	−0.534	0.436	−0.612	0.860	−0.552	0.642	0.602	−0.040
22	−0.534	0.330	−0.005	1.055	−0.899	0.822	−0.301	−1.123
23	−0.534	0.358	−0.003	0.592	−0.827	0.635	0.778	0.143
24	−0.534	0.107	−0.612	0.000	1.493	1.082	1.301	0.219
25	−0.534	0.440	−0.001	0.956	−1.705	0.414	0.301	−0.113
26	−0.534	0.413	−0.610	0.453	−0.536	0.429	0.602	0.174
27	−0.534	0.354	−0.003	1.278	−0.863	0.968	−0.301	−1.269
28	−0.534	−0.090	−0.612	1.168	−1.668	−0.046	−0.301	−0.255
29	−0.534	0.358	−0.002	0.958	−1.705	0.372	0.301	−0.071
30	−0.534	0.385	−0.001	1.323	−2.583	0.122	−0.301	−0.423
31	−0.534	0.409	−0.609	0.905	−2.566	−0.386	−0.301	0.085
32	−0.534	0.415	−0.612	−0.928	2.161	1.107	1.204	0.097
33	−0.534	0.178	−0.612	−2.241	2.566	0.517	0.477	−0.040
34	0.387	−0.673	0.366	0.025	−0.979	0.398	0.699	0.301
35	0.387	−0.189	0.172	0.469	−0.926	0.803	0.903	0.100
36	0.387	−0.453	0.732	0.025	−0.979	0.700	0.699	−0.001
37	0.387	−0.274	−0.005	0.364	−0.980	0.586	−0.301	−0.887
38	0.387	−0.392	−0.387	0.702	−0.982	0.502	0.301	−0.201
39	0.387	0.164	0.359	0.364	−0.980	0.999	1.301	0.302
40	0.387	−0.434	0.226	0.364	−0.980	0.623	0.301	−0.322
41	0.387	−0.265	−0.204	0.469	−0.926	0.570	0.602	0.032
42	0.387	−0.143	0.524	0.469	−0.926	1.009	1.000	−0.009
43	0.387	−0.434	1.093	0.025	−0.979	0.896	1.000	0.104
44	0.387	0.159	0.361	−0.187	0.243	1.345	1.301	−0.044
45	0.387	0.162	0.364	−1.112	1.205	1.367	1.301	−0.066
46	0.387	−0.153	−0.394	−3.708	2.708	0.250	0.301	0.051
47	0.387	−0.075	−0.029	0.702	−0.982	0.850	0.778	−0.072
48	0.387	−0.200	0.007	0.445	−1.038	0.643	0.477	−0.166
49	0.387	−0.417	0.223	0.188	−0.982	0.538	0.477	−0.061
50	0.387	−0.200	−0.376	0.607	−1.041	0.528	0.431	−0.096
51	0.387	0.225	0.369	0.445	−1.038	1.048	1.000	−0.048
52	0.387	−0.233	0.357	0.445	−1.038	0.806	1.000	0.194
53	0.387	−0.035	0.527	0.469	−0.926	1.066	1.000	−0.066
54	0.387	−0.100	−0.025	0.607	−1.041	0.760	1.079	0.319
55	0.387	0.071	−0.011	0.526	−1.041	0.814	0.699	−0.115
56	0.387	0.142	0.222	0.501	0.415	1.708	1.903	0.195
57	0.387	0.085	0.242	0.158	1.335	1.986	2.000	0.014
58	0.387	0.098	0.243	−0.563	1.992	1.960	1.903	−0.057
59	0.387	0.107	0.243	−1.667	2.387	1.599	1.602	0.003
60	0.387	0.107	0.237	−1.667	2.260	1.531	1.301	−0.230
61	0.387	0.113	0.241	−3.143	2.520	0.909	1.000	0.091
62	0.387	0.645	0.124	0.080	0.519	1.752	1.602	−0.150
63	0.387	0.977	0.121	−1.842	1.822	1.603	1.602	−0.001

[a] Each component adds 0.515 (i.e. the standard deviation) times the listed numbers to the mean log activity (0.848). Thus the calculated activity of mescaline (**3**) is given by $10^{[0.848 + 0.515[(-0.534) + (-0.413) + (-0.390) + (0.702) + (-0.982)]]} = 1.038$ MU.

No compound is more than 0.5 log unit (3 standard deviations, a factor of 3) more active than predicted. Eight, however, are less active by more than this amount (**2**, **4**, **5**, **11**, **13**, **22**, **27**, **37**). It should be noted that these substances, except for **37**, are those for which no activity was found experimentally, and so the discrepancies here are lower bounds, and the anomalies may be greater than is apparent. Curiously, one of the largest known anomalies is DMPEA (**2**), which has potential biological significance in the etiology of schizophrenia. Compound **37**, known to be half as active as mescaline and predicted to be three times as active, is the corresponding phenylisopropylamine. There is no obvious correspondence of less than expected activity with steric crowding of alkoxy or alkylthio groups, or with abnormally large or small values of any of the contributing variables, apart from the fact that only one

of the eight is a phenylisopropylamine. All are members of the "inactive" group, which were not used in deriving the regression.

The standard error of estimate of the best regression is 0.1624 in log A. This corresponds to an error in the activity A of 50%. Shulgin et al. claim an error in the determination of A of "not closer than 25%".[49] This should probably not be taken too literally, especially as Cassels and Gomez-Jeria[50] have more recently observed discrepancies, involving a factors of up to 2.5 in the activities of some hallucinogens. In view of this it would appear that within the class of compounds included herein, the accuracy of prediction is as good as can be expected, given the uncertainty of the data.

The animal data presented in the plot of log A (human) versus log A (rat) (Figure 5) do not seem to support the hypothesis that the animal drug-discrimination tests and the human evaluation of the drugs are measuring precisely the same thing. The QSAR's developed from the two sets of data show little resemblance to each other, (eq 11 and 12) and although there is a strong relationship between the effective doses for human and rat, the correlation coefficient for the two is quite poor (0.844), whereas it might have been expected to be as good or better than that for the human QSAR (0.956). In view of the relatively small number of drugs which have been tested using the discrimination technique, however, this conclusion may be premature. Since the legal and ethical restrictions on hallucinogen research involving animals are less severe than for that involving humans, it may be hoped that this will soon be rectified.

As a fairly severe test of the statistical model, a prediction of the activity of DON ((2,5-dimethoxy-4-nitrophenyl)isopropylamine) was made, based on eq 1. The CNDO results are

E_{HOPO}, −0.4348; E_{LUPO}, 0.0614; D_x, 4.878; Q_1, 0.0346; Q_2, −0.0097; Q_3, 0.1903; Q_4, −0.0164; Q_5, −0.0164; and Q_6, 0.1412.

Using these values and the values from Table IV gives an activity of 8.0 MU, compared with an observed value of approximately 70,[51] a result which is disappointing, but it must be noted that the value of the hydrophobicity of the nitro group represents a considerable extrapolation from the compounds for which eq 1 was derived.

6. Discriminant Analysis

It is sometimes maintained[52] that discriminant analysis is inappropriate in structure–activity studies. In the present case, when statistical linear discriminant analysis[53,54] was applied to the full data set, a linear dependency was indicated, aborting the calculation. With selected variables, however, the result was more satisfactory. The variables selected were those which had been useful in the regression analysis, namely, I_{Me}, H_p, H_m, V_p, V_m, E_{L-H}, and Q_{m-o}. With just these variables, and all three compound classes, two discriminants were obtained

$$D_{13} = C_1I_{Me} + C_2H_p + C_3V_p + C_4E_{L-H} + C_5H_m + C_6V_m + C_7Q_{m-o} - F$$

	1	2	3	4
C	−1.571	−0.836	0.0237	60.88
P	9.91	3.2	1.4	0.6
	5	6	7	
C	7.31	−0.0642	1.794	
P	64.5	19.3	1.2	(13a)
	1	2	3	4
C	−1.725	−6.506	0.152	177.5
P	4.8	16.5	4.3	8.4
	5	6	7	
C	13.03	−0.116	10.8	
P	38.0	18.2	9.8	(13b)

Note: F is given by $F = {}^1/_2 \sum C_i(\bar{V}_{ij} + \bar{V}_{ik})$, where i is the variable number, and $\bar{V}$ refers to the mean of the variable for classes j and k, the sum being over all i. The means are shown in Table IX. With two classes, the single discriminant classifies according to whether it is negative or positive, a negative value indicating high activity and a positive value inactivity. With three classes, there are two discriminants, a and b. In discriminant a, the classes j and k are inactive and low activity, and in discriminant b they are inactive and high activity. If both discriminants are positive, the compound is classified as inactive. If a is negative and b is greater than a, the indication of low activity, whereas if a is negative and a is greater than b, the decision is high activity. The discriminant weights are listed above as C for each term, and the discriminating power as p (expressed as a percentage).[53]

This discriminant misclassified 16 out of 63 compounds. In view of the arbitrary nature of the separation of the drugs into classes, a high misclassification rate is to be expected. It is noteworthy that only one inactive compound was classified as highly active (**5**, 2,3,4-trimethoxyphenethylamine), and no highly active compound was classified as inactive. Training the discriminant on the inactive and highly active substances only resulted in a similar function:

$$D_{14} = C_1I_{Me} + C_2H_p + C_3V_p + C_4E_{L-H} + C_5H_m + C_6V_m + C_7Q_{m-o} - F$$

	1	2	3	4
C	−5.32	−7.81	0.2073	149.2
P	10.3	13.5	4.0	4.8
	5	6	7	
C	18.25	−0.229	10.31	
P	36.4	24.5	6.4	(14)

Table IX. Means for Each Class (Inactive, Low Activity, High Activity) of All Drugs in Table II of All Variables Used in Discriminant Analysis

variable	inactive	low act.	high act.
I_{Me}	0.07692	0.4286	0.7727
H_p	0.1923	0.3921	0.8159
V_p	34.75	38.09	41.65
V_m	68.42	51.67	29.86
H_m	0.6915	0.1996	−0.02636
H_mH_p	0.2142	0.07154	−0.01921
V_mV_p	2555	2202	1375
E_{H-L}	0.5403	0.5408	0.5286
Q_{m-o}	0.2600	0.2970	0.03567
H_p^2	0.1591	0.6288	1.243
V_p^2	1304	2130	2453

(49) Shulgin, A. T.; Sargent, T.; Naranjo, C. *Nature* **1969**, *221*, 537.
(50) Cassels, B. K.; Gomez-Jeria, J. S. *J. Psychoact. Drugs* **1985**, *17*, 129.
(51) Gomez-Jeria, J. S.; Cassels, B. K.; Saavedra-Aguilar, J. C. *Eur. J. Med. Chem.* **1987**, *22*, 433.
(52) Wold, S., et al. In *Chemometrics: Mathematics and Statistics in Chemistry*; Series C: Mathematical and Physical Sciences; Kowalski, B. R., Ed.; NATO Advanced Study Institute Series D. Reidel: Dordrecht, The Netherlands, 1983; Vol. 138, p 64.
(53) Massart, D. L.; Dijkstra, A.; Kaufmann, L. *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*; Techniques and Instrumentation in Analytical Chemistry; Elsevier: Amsterdam, 1978; Vol. 1, p 412.
(54) Srivastava, M. S.; Carter, E. M. *An Introduction to Applied Multivariate Analysis*; North-Holland: Amsterdam, 1983; p 247.

A glance at eq 13 and 14 will show that two of the variables, (3 and 4), have low discriminating power in all three functions. Omitting these and training the discriminant on the inactive and highly active compounds only yields the discriminant function

$$D_{15} = C_1 I_{Me} + C_2 H_p + C_3 H_m + C_4 V_m + C_5 Q_{m-o} - F$$

	1	2	3	4	5	
C	-4.99	-2.17	15.06	-0.2078	8.70	
P	13.6	5.3	42.2	31.3	7.6	(15)

This equation results in one misclassification out of 35 compounds, but it must be appreciated that this was on the training set. If, instead of the raw variables, the transformed variables from the regression analysis and all of the data are used, the following two discriminants are obtained:

$$D_{16} = C_1 I_{Me} + C_2 H_p + C_3 E_{L\text{-}H} + C_4 H_m H_p + C_5 V_m V_p + C_6 H_p^2 + C_7 V_p^2 + C_8 Q_{m-o} - F$$

	1	2	3	4	
C	-2.728	-3.474	28.07	3.870	
P	25.1	18.2	0.4	14.5	
	5	6	7	8	
C	0.00072	1.756	-0.00049	-2.96	
P	6.7	21.6	10.6	2.9	(16a)
	1	2	3	4	
C	-5.363	-17.05	119.6	12.96	
P	14.1	40.2	5.3	11.4	
	5	6	7	8	
C	0.00076	4.043	0.00143	3.22	
P	3.4	16.6	6.2	2.74	(16b)

This functions misclassifies 13 out of 63 compounds, but no inactive compound is classified as highly active, and no highly active compound is classified as inactive. When trained on the inactive and highly active set, the following function is obtained:

$$D_{17} = C_1 I_{Me} + C_2 H_p + C_3 E_{L\text{-}H} + C_4 H_m H_p + C_5 V_m V_p + C_6 H_p^2 + C_7 V_p^2 + C_8 Q_{m-o} - F$$

	1	2	3	4	
C	-6.593	-12.38	79.83	8.205	
P	21.8	36.7	4.4	9.1	
	5	6	7	8	
C	0.00111	3.048	0.0100	-0.526	
P	6.2	15.7	5.5	0.6	(17)

Removing the two least powerful discriminating variables from 17 (3 and 8) and training on the inactive and highly active data set yields the following discriminant function:

$$D_{18} = C_1 I_{Me} + C_2 H_p + C_3 H_m H_p + C_4 V_m V_p + C_5 H_p^2 + C_6 V_p^2 - F$$

	1	2	3	
C	-6.91	-13.39	8.84	
P	22.8	39.7	9.8	
	4	5	6	
C	0.000629	3.80	0.000825	
P	3.5	19.6	4.5	(18)

This discriminant, applied to the data on which it was trained, misclassifies one compound (**37**, (3,4-dimethoxyphenyl)isopropylamine). When the data was repeatedly split at random into a training and test set, putting approximately 85% into the training set, the misclassification rate was 9% (15 out of 161 classifications). With the predictor variables of eq 16, the corresponding misclassification rate was 11% (13 out of 118), which does not vary significantly from that for the contracted equation. Misclassified by eq 18 were compounds **16**, **19** (twice), **24** (twice), **32** (three times), **37** (3 times), and **61** (four times). Noteworthy is compound **37** ((3,4-dimethoxyphenyl)isopropylamine) which was also anomalous in the regression and knn results. DMPEA (compound **2**) was correctly classified by the discriminant-analysis procedure. Thus the discriminant analysis of selected variables, particularly when an appropriate transformation is applied, can give satisfactory results. These transformations can be regarded as making the discrimination surface nonplanar. Comparison of the discriminant equations with those from the regression analyses suggests that the descriptors which determine whether a compound is active or inactive are not ranked in the same order as those which determine whether it is of high or low activity.

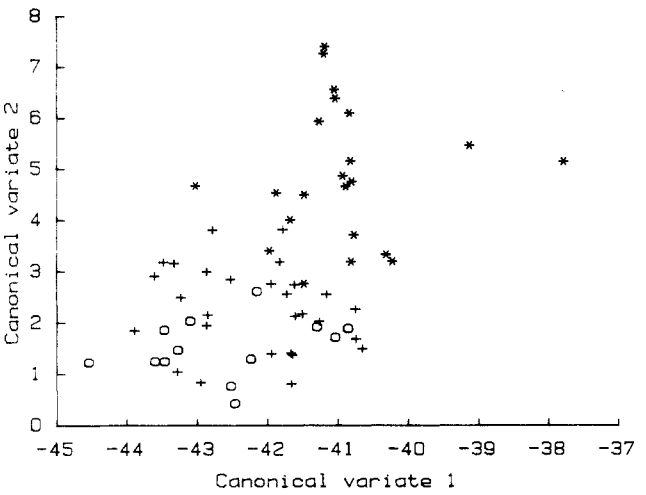


Figure 8. Plot of the two canonical variates from discriminant analysis, showing separation of high activity compounds from inactive ones: O, inactive; +, low activity; *, high activity.

To classify the hallucinogens from this particular group with 90% reliability requires only knowledge of the presence of absence of an α-methyl group, hydrophobicities, and volumes of the groups in the meta and para position. This should not be taken to indicate that the other variables are without effect; merely that within the present context, in the presence of the other variables, and within the variability of this data set their influence is negligible. A plot of canonical variates, using canonical discriminant analysis of the descriptors in eq 17[54] is shown in Figure 8. It will be seen that the highly active compounds are well separated from the inactive ones, and the compounds of low activity are not well separated, as would be expected. Equation 17 applied to the test case DON which was discussed under regression in section 5 gives the value -0.132, resulting in the correct classification of high activity. Equation 18 however incorrectly classified DON as inactive, with the discriminant value of 3.46.

7. Discussion and Conclusions

The pattern-recognition techniques of nonlinear mapping, and to a greater extent, k nearest neighbors, indicate that the calculated variables studied in this paper do indeed discriminate between hallucinogenic and nonhallucinogenic derivatives of phenethylamine. Multiple linear regression proved to be the most useful technique for identifying relevant and eliminating redundant variables, particularly when used in conjunction with hierarchical clustering of the correlation matrix.

Regression is not well-suited to those compounds in the "inactive" class, where there may well be an effective dose, but only a lower bound is known. Discriminant analysis is more appropriate for such dichotomous variables—in this case, highly active or inactive. It proved possible to classify this group, and further select variables by the use of this technique.

As with most QSAR studies, the dominating influence is hydrophobicity. Correlation was obtained not with the hydrophobicity of the molecule as a whole, however, but with that of the substituents in the meta and para positions on the benzene ring. This would be due at least in part to the fact that the calculated "whole molecule" hydrophobicities would be in error due to interaction between alkoxy groups adjacent to each other on the ring. Because no significant interactions between ortho and meta groups were found, the difference between such isomers as TMA-2 and TMA-3 must be due to ring charge or orbital separation, if they are to be explained by the variables studied herein. Comparison of the values for compounds **39** and **40** in Table VIII shows that it is indeed the orbital energy contribution which accounts for the difference in calculated activity between these compounds. This is however less than $^1/_2$ of the difference in observed activity.

A strong interaction, dominant among the effects found, was discovered between both the volumes and hydrophobicities of the groups in the meta and para positions on the benzene ring. The fact that both the substituent hydrophobicity and the substituent volume can be accommodated in the equation is interesting. Both are rather crude measures—one of lipid-seeking and membrane-passing ability, and the other of steric hindrance and receptor site fit. They are, as must be expected from the manner in which they are calculated, strongly correlated.

In summary, activity increases with increasing meta substituent volume at constant para substituent volume. When the meta substituent volume is held constant, activity first increases with para substituent volume, but passes through a maximum, and then decreases. Thus increasing meta substituent volume doubly enhances activity. This is illustrated in Figure 6b. Activity decreases, however, with increasing meta substituent hydrophobicity at constant para substituent hydrophobicity. At constant meta substituent hydrophobicity, activity passes through a maximum as the hydrophobicity of the para substituent increases. The optimum para substituent hydrophobicity decreases with increasing meta substituent hydrophobicity, the latter thus doubly inhibiting activity. This can be seen in Figure 6a.

Inspection of a plot of meta against para volume revealed a fairly uniform spread of points. A similar plot for hydrophobicity shows a cluster of points along the para axis. The few remaining points of high meta hydrophobicity are all compounds of low activity. No points are present in the region where both meta and para hydrophobicity are high, so this region of Figure 6a represents an extrapolation. Interpretation of the large negative values of the hydrophobicity contribution in this region would therefore be unwarranted.

Thus, for high activity, the requirement is for meta substituents of large volume but low hydrophobicity, methoxy being particularly favorable, and a moderate volume and hydrophobicity for the para substituent. The reason for the meta–para interaction remains obscure.

Another puzzle arises from the orbital energy contribution. On first inspection of the correlation matrix, the HOPO energy seems to be the main influence affecting activity. This has been accepted by most workers in this field and has been taken to mean that a charge-transfer complex is formed with some electron acceptor in a site in the brain. This correlation disappears, however, on examination of all variables simultaneously, as in multiple regression. It appears that a better correlation is with the energy difference between the LUPO and HOPO. On examining eq 8, this conclusion is almost irresistible. Such correlations are not unknown in structure–activity work,[55-57] and similar results have been found for carcinogenic substances.[58,59] The physical meaning of the result is, however, elusive.

Liu and Zheng[55] stated that the first step in the reaction must be excitation of the toxic molecule, but how does this excitation occur? It cannot be thermal, as the energy involved is of the order of 1200 kJ/mol. In their study of carcinogenic hydrocarbons, Buu-Hoi and Sung[60] attributed it to excitation by chemiluminescent reactions. In their case, this may be true, but here that would require radiation in the far UV, at 80 nm! Interaction between ground-state and excited-state configurations can also occur without actual excitation, and this effect increases rapidly with decreasing energy separation, but only if the ground and excited state configurations are of the same symmetry. This interaction effect could lead to changes in chemical reactivity, but again, the magnitude of the excitation energy is prohibitive. Radiationless transfer of energy from such reactions was also suggested.

Birks[58] suggested dipole–dipole transfer of excitation energy from molecule to molecule, resulting in an optimum value for the excitation energy. This was not observed in the present study, the coefficient of the term quadratic in excitation energy being extremely nonsignificant when included with a linear term (but not on its own, where it was comparable in significance to the linear term on its own). Thus the optimum energy difference, if it exists, is well below any that were observed. It may also be that the correlation is fortuitous and that it is the LUPO energy which is the influence. At least, the correlation coefficient for that case is not much worse than for the alternative, and with some sets of predictor variables, there was some evidence for this view, and this would be consistent with the formation of a charge-transfer complex by accepting electrons.

A situation well-known in organometallic chemistry where the energy difference might be expected to correlate with compound stability arises in connection with bonding in transition-metal carbonyls. A filled lone pair of σ symmetry on the carbon atom overlaps with an empty σ orbital on the metal. The buildup of charge on the metal is minimized by the metal donating charge from its d orbitals into empty π antibonding orbitals on the carbon, the process being synergistic.[61]

It is difficult to see how an effect such as this could occur with the phenylalkylamines discussed here. The only σ lone pairs on the hallucinogens are those on the oxygen, nitrogen, and sulfur atoms, and the only antibonding π orbitals are those on the benzene ring, which are sterically obstructed by the hydrogen atoms in the edge-on orientation. In the face-on orientation (as in the sandwich compounds) no σ bond formation is possible, and hence,

(55) Liu, R.-Z.; Zheng, S.-J. *Int. J. Quantum Chem.* **1983**, *23*, 1609.
(56) Mehler, E. L.; Gerhards, J. *Mol. Pharmacol.* **1987**, *31*, 284.
(57) Kikuchi, O. *Quant. Struct.-Act. Relat.* **1987**, *6*, 179.
(58) Birks, J. B. *Nature* **1961**, *190*, 232.
(59) Sung, S.-S. *Ann. Soc. Sci. Bruxelles Ser. 2*, **1975**, *89*, 319.
(60) Buu-Hoi, N. P.; Sung, S.-S. *Naturwissenschaften* **1970**, *57*, 135.
(61) Cotton, F. A.; Wilkinson, G. *Advanced Inorganic Chemistry*, 4th ed.; John Wiley & Sons: New York, 1980; p 82.

no synergistic effect of this kind can occur.

Less perplexing is the effect found for the charge on the meta ring atoms, or alternatively, the X component of the dipole moment. This could be due to the influence of a charge or dipole on the receptor, where that portion of the molecule is accommodated, or alternatively, to an influence on the reactivity of the molecule at the meta position, possibly at a site remote from the receptor. This charge or dipole component seems to be the only significant influence of ortho substituents.

Care is needed in the use of statistical and pattern recognition methods on multivariate data. If the number of variables is not small compared with the number of patterns (data points), impressive correlation coefficients can be obtained in the presence of little actual relationship. It is usually stated that the number of points in pattern recognition should exceed the number of variables by a factor of at least 3 or 4.[62,63] In this work, all regressions were done with 50 points, and the final equation involves seven variables and nine adjustable parameters. Thus the ratio here is greater than 5.

A second pitfall is the fact that when one chooses and tests a set of variables from a large pool, apparently significant correlations can arise by chance, with a probability increasing with the number of selections. Topliss and Costello[64] studied the relationship of the number of parameters to the probability of obtaining misleadingly significant correlations. The number of raw variables included in this study was 24. The effective number would be greater than this because of the interactions and squared terms, but this would be offset by the intercorrelations between variables, squared terms, and interaction terms. Topliss and Costello obtained their result from uncorrelated random numbers. From Topliss and Costello, for 60 observations on 20 variables, a correlation coefficient of 0.63 is obtained by chance, with an average of 2.6 variables included in the equation. The correlation coefficient for eq 1 is 0.96 with seven variables included. This, and the extremely low probabilities P' in regressions 1–10, indicates that the correlations derived in this work are not the result of chance. Because negative results are not normally reported, a more insidious source of error would be between authors, rather than within any one set of results.

By using a large data set, a wide variety of substituents and substitution patterns, and multivariate methods, it was possible to detect effects that were not evident in studies published previously. Although it is true that the risk of chance correlations between variables selected from a large set increases rapidly as the number of selected variables increases, this can be minimized by using as many points as possible. By considering more homogeneous (but of course smaller) groups of compounds, correlation coefficients greater than 0.995 were readily obtained, but with only two to four degrees of freedom. These are to be regarded as spurious, due to the causes discussed above.

Relatively little experimental data is available for compound **37** ((3,4-dimethoxyphenyl)isopropylamine), but in view of its persistent misclassification by three different techniques (knn, regression, and discriminant analysis), perhaps its "inactive" status should be reexamined.

Acknowledgment. I am indebted to Graham S. Chandler of the University of Western Australia School of Chemistry for advice on quantum chemistry, to Ronald J. McKay (Murdoch University M.P.S.) and Jennifer M. Hopwood (U.W.A. Maths) for help and discussions on statistical and mathematical matters, to Mark C. Favas (U.W.A. Chem.) for advice and assistance with FORTRAN programming, and to Dr. McKay and Paula J. McLay (Murdoch University M.P.S) for reading and criticizing the manuscript. Finally, I thank the referees for a number of suggestions, which have been implemented in the revised manuscript.

Registry No. **1**, 55-81-2; **2**, 120-20-7; **3**, 54-04-6; **4**, 3937-16-4; **5**, 15394-83-9; **6**, 90132-31-3; **7**, 39201-82-6; **8**, 39201-78-0; **9**, 64778-75-2; **10**, 63918-08-1; **11**, 90109-61-8; **12**, 90109-63-0; **13**, 90132-33-5; **14**, 23693-38-1; **15**, 1484-85-1; **16**, 71539-34-9; **17**, 24333-19-5; **18**, 78335-85-0; **19**, 71539-35-0; **20**, 90132-35-7; **21**, 90109-47-0; **22**, 90132-39-1; **23**, 90132-37-9; **24**, 90109-49-2; **25**, 63918-08-1; **26**, 90109-49-2; **27**, 90132-51-7; **28**, 90109-45-8; **29**, 90132-47-1; **30**, 90132-53-9; **31**, 90109-53-8; **32**, 90109-55-0; **33**, 90109-57-2; **34**, 64-13-1; **35**, 2801-68-5; **36**, 23690-13-3; **37**, 120-26-3; **38**, 1082-88-8; **39**, 1083-09-6; **40**, 1082-23-1; **41**, 23693-14-3; **42**, 20513-16-0; **43**, 15402-79-6; **44**, 16128-88-4; **45**, 123643-24-3; **46**, 123643-25-4; **47**, 23693-26-7; **48**, 4764-17-4; **49**, 23693-20-1; **50**, 13674-05-0; **51**, 23693-18-7; **52**, 23693-19-8; **53**, 23693-22-3; **54**, 15183-13-8; **55**, 15183-26-3; **56**, 15588-95-1; **57**, 22004-32-6; **58**, 63779-88-4; **59**, 63779-89-5; **60**, 89556-70-7; **61**, 63779-90-8; **62**, 61638-07-1; **63**, 123643-26-5.

Supplementary Material Available: Three tables listing the compounds of interest; orbital energies for SHOMO, HOMO, LUMO, and SLUMO, net charges on the benzene ring carbon atoms, and the X, Y, and Z components of the dipole moment; and the Cartesian coordinates of all atoms in each molecule. The conformations of the compounds in Table II, calculated by MMP2 and plotted by the program NAMOD,[65] are shown in a figure (20 pages). Ordering information is given on any current masthead page.

(62) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. In *Chemometrics: Theory and Application*; Kowalski, B. R., Ed.; ACS Symposium Series 52; American Chemical Society: Washington, DC, 1977; Chapter 9.

(63) Otto, P.; Seel, M.; Ladik, K.; Muller, R. *J. Theor. Biol.* **1979**, *78*, 197.

(64) Topliss, J. G.; Costello, R. J. *J. Med. Chem.* **1972**, *15*, 1066.

(65) Beppu, Y. *QCPE* **1979**, *11*, 370, NAMOD.