Journal of Computer-Aided Molecular Design, **16:** 611–633, 2002.
KLUWER/ESCOM

QSAR of benzene derivatives: comparison of classical descriptors, quantum theoretic parameters and flip regression, exemplified by phenylalkylamine hallucinogens

Brian W. Clare*
Department of Chemistry, The University of Western Australia, 35 Stirling Highway, Crawley, W.A. 6009, Australia

Received 26 October 2001; accepted in final form 13 October 2002

Key words: benzene, FOPA, hallucinogen, phenylalkylamine, QSAR, quantum, regression, symmetry

Summary

A physical model of electronic effects in the QSAR of benzene derivatives, together with a regression technique for finding predictive equations, is presented. The model is simple, based on the quantum theoretic description of the benzene molecule, and accounts for the variance in activity of hallucinogenic phenylalkylamines as well as a classical description in terms of electronic (atomic charge, orbital energy), hydrophobic (Hansch π) and steric (substituent volume) terms. The new model involves the energies of four π-like near frontier orbitals and the orientations of their nodes. It is less affected by colinearity than the classical approach. This model more than any other illustrates the essential wave mechanical nature of the interaction of a drug with its receptor, as the π-like orbitals involved are standing waves of probability of finding an electron in a given location in the field of the atomic nuclei, and have no classical counterpart.

Introduction

Some years ago [1] a QSAR for 50 of a large group of phenylalkylamine hallucinogens [2] was presented, using orbital energies, atomic charges, dipole moments and the lipophilicities and volumes of substituents as descriptors. An 8 term equation involving the volumes and lipophilicities of the meta and para substituents and their squares, interactions between the meta and para groups, the energy gap between the highest occupied π-like orbital (HOPO) and lowest unoccupied π-like orbital (LUPO), and the difference in charge between the ortho and meta carbon atoms was presented. In that paper it was pointed out that compounds such as 3,4-dimethoxyphenethylamine and 4,5-dimethoxyphenethylamine, being chemically identical, should yield identical predicted activities. To ensure this, substituent descriptor values for the two ortho positions were summed, as were those for the two meta positions, based on the symmetry of the benzene moiety. This procedure is not entirely justifiable, and unless either product terms (as in reference [1]) or indicator variables for vicinal substitution are introduced, enforces the incorrect prediction of 2,3,4-trimethoxyamphetamine and 2,4,5-trimethoxyamphetamine having equal activities. The former of these is not hallucinogenic (activity less than 0.5 mescaline units), while the latter is an hallucinogen of high potency (10 mescaline units).

When a phenylalkylamine molecule lies on the receptor (in the case of the hallucinogens this is probably the 5-HT_{2A} serotonin receptor) it can do so in two ways, given that it is anchored by the ethylamine group. These orientations are related through the rotation of the benzene ring through 180° around the 1–4 carbon axis, and will have the two ortho and two meta groups flipped into each others positions, i.e. $2 \rightarrow 6$ and $3 \rightarrow 5$. The dipole moment will also be reflected in this operation. These orientations will not be equivalent in their interactions with the receptor, given that the latter presumably has no symmetry.

A procedure called flip regression is now proposed whereby a multiple linear regression is carried out on

*Correspondence: E-mail: bwc@theochem.uwa.edu.au

all of the drugs, with each drug independently in both possible orientations. Flipping consists of rotating or reflecting the drug such that, for a phenethylamine, the charge and substituent values for two o-positions, and also the two m-positions, are interchanged. Thus given that there are N drugs, 2^N regressions must be carried out, and that one giving the best fit is selected. To do this rigorously will of course result in a combinatorial explosion as N becomes greater than about 20 to 25, depending on the computational resources available. However two iterative solutions are presented which avoid this, and while they do not always give the best fit, usually come very close to it.

More recently [3] a QSAR on a much larger group of phenylalkylamines was presented, using some new descriptors based on the quantum mechanics of benzene derivatives. These descriptors, the frontier orbital phase angles (FOPA) were subsequently shown [4] to be simply the angles made by the nodes in the π-like orbitals of the molecules with the position of the ethylamine chain. These angles are of course also affected by the binding orientation of the drug, a positive angle being transformed into its negative on flipping. All of these calculations were based on semiempirical calculations: the early complete neglect of differential overlap (CNDO) [5] in the case of the first QSAR [1], and the more recent Austin model 1 (AM1) [6] in the case of the second [3]. The quantum theoretic calculations in the present paper are based on a density functional (DF) approach [7] using a basis set of reasonable quality, made possible by recent advances in computer hardware and molecular modeling software. It is expected that this will greatly improve the accuracy of the calculated parameters, and so also improve the quality of the regression fits, if the model is valid.

The purpose of the present communication is to compare the use of classical descriptors [1] with the FOPA variables [3], and to see whether both can be improved by the flip regression procedure. Secondary goals are to determine whether both types of descriptor can be combined, and whether the more advanced level of theory improves the results.

Both the HOMO and LUMO (lowest unoccupied molecular orbital) of benzene consist of a pair of degenerate π orbitals which can be combined in any proportions to give equally acceptable pairs of orbitals, the nodes of which may have any orientation perpendicular to the ring. Previously [3], it was shown that two particular orthogonal orbitals, designated **P** and **Q** could be arbitrarily selected, such that, with the z axis normal to the ring, the coefficients of the p_z atomic orbitals of the six carbon atoms were given by

$$\mathbf{P}_{\mathrm{H}} = \left(\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, -\frac{1}{\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

and

$$\mathbf{Q}_{\mathrm{H}} = \left(\frac{1}{2}, \frac{1}{2}, 0, -\frac{1}{2}, -\frac{1}{2}, 0\right)$$

for the HOMO, and

$$\mathbf{P}_{\mathrm{L}} = \left(-\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{\sqrt{3}}, -\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

and

$$\mathbf{Q}_{\mathrm{L}} = \left(\frac{1}{2}, -\frac{1}{2}, 0, \frac{1}{2}, -\frac{1}{2}, 0\right)$$

for the LUMO. Then any of the infinity of possible HOMO and LUMO orbitals of benzene may be given by linear combinations of **P** and **Q**: $C_p\mathbf{P} + C_q\mathbf{Q}$, where normalization requires that $C_p^2 + C_q^2 = 1$. It is convenient to introduce angular parameters Θ_{H} and Θ_{L}, such that $C_p = \cos\Theta$ and $C_q = \sin\Theta$. These are the FOPA parameters. A procedure was given [3], to derive for substituted benzene derivatives Θ_{H} and Θ_{L} corresponding to a best fit to the coefficients of the p_z orbitals of the compound to **P** and **Q**, the coefficients of the p_z orbitals on the six carbon atoms of benzene. Later, [4] a physical interpretation of Θ was found. If Φ_{H} and Φ_{L} are that angles which the nodes in the HOMO and LUMO make at the center of the benzene ring with the first carbon of the ring, then for the HOMO, $\Phi_{\mathrm{H}} = \Theta_{\mathrm{H}} + \pi/6$. For the LUMO there are two mutually perpendicular nodes, making angles of $\Phi_{\mathrm{L}} = (\Theta_L - \pi/6)/2$ and $\Phi_{\mathrm{L}}' = (\Theta_L + 5\pi/6)/2$ with the first atom of the ring.

When benzene is substituted, the degeneracies of both the HOMO and LUMO are lifted and the nodes in the orbitals take on a definite orientation in space. Some orbitals near the HOMO and LUMO of the molecule may be regarded approximately as derived from linear combinations of the original degenerate orbitals of benzene. This is an approximation, because in fact, in the strict group-theoretic sense, only linear or planar molecules can have canonical orbitals that are of π symmetry. There are however at least two pairs and sometimes more π-like canonical orbitals that can be regarded as being derived from the pure degenerate π frontier orbitals of benzene. As well as the angles, either the energies of the four orbitals, or alternatively, the splitting in energies of the degenerate orbitals of benzene in formation of the substituted molecule also seem important in explaining biological activities.

The reason that the FOPA quantities affect activity is that they approximate the actual orientation of the nodes of the orbitals in the compound, and the π-like orbitals on aromatic compounds presumably interact with π-like orbitals on the receptor. It is well recognized in quantum chemistry that a correspondence in orbital symmetry between the pair of interacting species is required to produce bonding. That is, for a strong interaction the nodes in the interacting orbitals must nearly coincide. Further, for two orbitals to interact strongly, they must be of similar energy. Applying Koopmans' Theorem, the magnitude of the splitting in energy may be a measure of the difficulty of distorting the interacting orbitals to induce correspondence, that is, the hardness, in the Pearson-Parr [8] sense of the molecular ion resulting from the gain (LUMO) or loss (HOMO) of an electron from the molecule. With too small a splitting, however, one can get a strong interaction between the receptor and, say, the HOPO of the compound in an ideal orientation being offset by an interaction with an almost orthogonal second highest occupied π-like orbital (SHOPO) in the worst possible orientation

Calculations

Activities of the drugs

The activities of the drugs were taken from the literature. [2] Their identities are as given in Table 1. This data has been widely used as a benchmark in QSAR studies. [1, 3, 9, 10, 11, 12, 13, 14, 15] The activity is in mescaline units, defined as the ratio of the active dose of the drug to that of mescaline. Only those compounds were selected to which had been assigned definite activity ranges [2], and were identical with those used previously [3], but with the deletion of the iodine compounds. This was because of the limitations of the basis set used, and the fact that in iodine there are appreciable errors introduced by the relativistic velocities of the innermost core electrons, which are not accounted for in Gaussian 98, the program used.

Molecular orbital calculations

The initial geometries of the compounds were set up using the molecular modeling software PCMODEL [16] and HYPERCHEM [17], and were initially optimized by molecular mechanics calculations to locate as best as possible global energy minima, before being fully optimized using the AM1 [6] Hamiltonian and the program MOPAC [18], version 6. The atoms were numbered so that the atom to which the ethylamine chain was attached was atom 1, then consecutively 2 to 6 around the ring, with substituents attached as in the systematic chemical names of the compounds.

For the purposes of the FOPA calculations and the dipole moment components, the center of the benzene ring was placed at the origin, atom 1 of the ring was on the positive X axis, and the Z axis was normal to the plane of the ring and the remaining 5 ring carbon atoms were numbered consecutively, anticlockwise around the ring. The angles Φ_H and Φ_L were calculated as described recently [4]. They are related to those described earlier [3] by $\Phi_H = \Theta_H + \pi/6$ and $\Phi_L = (\Theta_L - \pi/6)/2$, where Θ_H and Θ_L are the angles defined previously [3]. Please note that these results are misprinted in the recent reference [4]. So defined, the angles are approximations to the angles made by the nodes in near-frontier π-like occupied and unoccupied orbitals with atom 1 of the benzene ring. It should be noted that strictly speaking, there are no π orbitals in compounds that are not planar, as is the case with all of these compounds. There are, however orbitals that are very like π orbirals, in having nodes passing through the plane of the six ring carbon atoms. These are identified by the relatively large value of SSQ, the sum of squares of the coefficients of the $2p_z$ orbitals on atoms 1–6, and the small value of the fitting error α, defined previously [3, 4]. In many, but not all cases the orbitals so identified are the HOMO and LUMO of the molecule.

The energies of these orbitals will be denoted E_H and E_L, and the orbitals will be called the HOPO and LUPO respectively. It is also necessary to identify a second pair of orbitals, corresponding to the other components of the degenerate HOMO and LUMO of benzene. These will be called the SHOPO and SLUPO respectively, and their energies will be denoted E_{SH} and E_{SL}. As before, these will be recognized by the large value of the SSQ and small value of α. There are other π-like orbitals of the wrong symmetry, and these may be recognized by the relatively large error term α in the least squares fitting of the orbitals of the compounds to those of benzene, as described in previous publications.

Liberal limits for SSQ and α were used in identifying the orbitals. In all cases at least one orbital fitting the criteria for each of SHOPO, HOPO, LUPO and SLUPO was found. When there were more than one, the orbitals were identified on the basis of SSQ and α. In most cases (e.g. **67**), the four orbitals for

Table 1. The hallucinogenic phenylalkylamines in the study, and their activities in mescaline units

No.	Designation	Chemical name	A
1	2-TOM	5-Methoxy-4-methyl-2-methylthioamphetamine	4
2	2,5-DMA	2,5-Dimethoxyamphetamine	2.5
3	2C-B	2,5-Dimethoxy-4-bromophenethylamine	16
4	2C-C	2,5-Dimethoxy-4-chlorophenethylamine	10
5	2C-D	2,5-Dimethoxy-4-methylphenethylamine	8
6	2C-E	2,5-Dimethoxy-4-ethylphenethylamine	18
7	2C-G	2,5-Dimethoxy-3,4-dimethylphenethylamine	11
8	2C-G-3	5-(2-Aminoethyl)-4,7-dimethoxyindane	14
9	2C-G-5	3,6-Dimethoxy-4-(2-aminoethyl)benzonorbornane	22
10	2C-G-N	1,4-Dimethoxynaphth-2-ylethylamine	10
11	2C-P	2,5-Dimethoxy-4-(n)-propylphenethylamine	37
12	2C-T	2,5-Dimethoxy-4-methylthiophenethylamine	4
13	2C-T-13	2,5-Dimethoxy-4-(2-methoxyethylthio)phenethylamine	9
14	2C-T-17	2,5-Dimethoxy-4-(s)-butylthiophenethylamine	4
15	2C-T-2	2,5-Dimethoxy-4-ethylthiophenethylamine	16
16	2C-T-21	2,5-Dimethoxy-4-(2-fluoroethylthio)phenethylamine	30
17	2C-T-4	2,5-Dimethoxy-4-(i)-propylthiophenethylamine	21
18	2C-T-7	2,5-Dimethoxy-4-(n)-propylthiophenethylamine	15
19	2C-T-8	2,5-Dimethoxy-4-cyclopropylmethylthiophenethylamine	7
20	2C-T-9	2,5-Dimethoxy-4-(t)-butylthiophenethylamine	4
21	2C-T-F	2,5-Dimethoxy-4-(2-fluoroethylthio)phenethylamine	29
22	3-TASB	3-Ethylthio-4-ethoxy-5-methoxyphenethylamine	2
23	3-TE	3-Methoxy-4-ethoxy-5-methylthiophenethylamine	4
24	3-TM	3,4-Dimethoxy-5-methylthiophenethylamine	4
25	3-TME	3,4-Dimethoxy-5-ethylthiophenethylamine	4
26	3C-BZ	3,5-Dimethoxy-4-benzyloxyamphetamine	3
27	3C-E	3,5-Dimethoxy-4-ethoxyamphetamine	7
28	4-MA	4-Methoxyamphetamine	5
29	4-TASB	3-Ethoxy-4-ethylthio-5-methoxyphenethylamine	4
30	4-TE	3,5-Dimethoxy-4-ethylthiophenethylamine	12
31	4-TME	3-Ethoxy-4-methylthio-5-methoxyphenethylamine	4
32	4-BR-3,5-DMA	3,5-Dimethoxy-4-bromoamphetamine	43
33	5-TASB	3,4-Diethoxy-5-methylthiophenethylamine	2
34	5-TOET	4-Ethyl-2-methoxy-5-methylthioamphetamine	15
35	5-TOM	2-Methoxy-4-methyl-5-methylthioamphetamine	7
36	AL	3,5-Dimethoxy-4-allyloxyamphetamine	11
37	ALEPH-2	2,5-Dimethoxy-4-ethylthioamphetamine	50
38	ALEPH-4	2,5-Dimethoxy-4-(i)-propylthioamphetamine	32
39	ALEPH-7	4-Propylthio-2,5-dimethoxyamphetamine	55
40	ASB	3,4-Diethoxy-5-methoxyphenethylamine	1.3
41	BOB	2,5,β-Trimethoxy-4-bromophenethylamine	20
42	BOD	4-Methyl-2,5,β-trimethoxyphenethylamine	15
43	BOH	β-Methoxy-3,4-methylenedioxyphenethylamine	3
44	CPM	3,5-Dimethoxy-4-cyclopropylmethoxyphenethylamine	4
45	DESOXY	3,5-Dimethoxy-4-methylphenethylamine	4

Table 1. Continued

No.	Designation	Chemical name	A
46	DMCPA	2-(2,5-Dimethoxy-4-methylphenyl)cyclopropylamine	17
47	DMMDA	2,5-Dimethoxy-3,4-methylenedioxyamphetamine	6
48	DOB	2,5-Dimethoxy-4-bromoamphetamine	150
49	DOC	2,5-Dimethoxy-4-chloroamphetamine	133
50	DOEF	2,5-Dimethoxy-4-(2-fluoroethyl)amphetamine	110
51	DOET	2,5-Dimethoxy-4-ethylamphetamine	75
52	DOM	2,5-Dimethoxy-4-methylamphetamine	50
53	DON	2,5-Dimethoxy-4-nitroamphetamine	80
54	DOPR	2,5-Dimethoxy-4-(n)-propylamphetamine	80
55	E	3,5-Dimethoxy-4-ethoxyphenethylamine	6
56	FLEA	N-Hydroxy-N-methyl-3,4-methylenedioxyamphetamine	2.5
57	G	2,5-Dimethoxy-3,4-dimethylamphetamine	12
58	G-3	2,5-Dimethoxy-3,4-trimethyleneamphetamine	20
59	G-5	3,6-Dimethoxy-4-(2-aminopropyl)benzonorbornane	18
60	HOT-17	2,5-Dimethoxy-4-(s)-butylthio-N-hydroxyphenethylamine	3
61	HOT-2	2,5-Dimethoxy-4-ethylthio-N-hydroxyphenethylamine	22
62	HOT-7	2,5-Dimethoxy-N-hydroxy-4-(n)-propylthiophenethylamine	15
63	IP	3,5-Dimethoxy-4-(i)-propoxyphenethylamine	5
64	J	2-Amino-1-(3,4-methylenedioxyphenyl)butane	1.5
65	MAL	3,5-Dimethoxy-4-methallyloxyphenethylamine	6
66	MDA	3,4-Methylenedioxyamphetamine	3
67	MDE	3,4-Methylenedioxy-N-ethylamphetamine	2
68	MDOH	3,4-Methylenedioxy-N-hydroxyamphetamine	2.3
69	MDPH	α,α-Dimethyl-3,4-methylenedioxyphenethylamine	1.5
70	ME	3-Ethoxy-4,5-dimethoxyphenethylamine	1
71	MEM	2,5-Dimethoxy-4-ethoxyamphetamine	9
72	Mescaline	3,4,5-Trimethoxyphenethylamine	1
73	META-DOB	2,4-Dimethoxy-5-bromoamphetamine	4
74	Methyl-J	2-Methylamino-1-(3,4-methylenedioxyphenyl)butane	1.5
75	MMDA	3-Methoxy-4,5-methylenedioxyamphetamine	2
76	MMDA-2	2-Methoxy-4,5-methylenedioxyamphetamine	8
77	MMDA-3A	2-Methoxy-3,4-methylenedioxyamphetamine	6
78	P	3,5-Dimethoxy-4-propoxyphenethylamine	7
79	PARADOT	2,5-Dimethoxy-4-methylthioamphetamine	40
80	Ψ−DOM	2,6-Dimethoxy-4-methylamphetamine	15
81	TA	2,3,4,5-Tetramethoxyamphetamine	6
82	TB	3,5-Dimethoxy-4-butylthiophenethylamine	3
83	TM	3,5-Dimethoxy-4-methylthiophenethylamine	10
84	TMA	3,4,5-Trimethoxyamphetamine	2
85	TMA-2	2,4,5-Trimethoxyamphetamine	10
86	TMA-4	2,3,5-Trimethoxyamphetamine	4
87	TMA-5	2,3,6-Trimethoxyamphetamine	10
88	TMA-6	2,4,6-Trimethoxyamphetamine	10
89	TP	3,5-Dimethoxy-4-propylthiophenethylamine	16

which SSQ was maximal also had α minimal, and these orbitals were used. In other cases the most plausible assignment was made on the basis of SSQ and α The number of cases in which there was ambiguity in the assignment of orbitals was small. In difficult cases the problem may usually be resolved by using a visualization program such as MOLDEN [19] to inspect the orbitals and identify those with a node mostly in the plane of the ring.

In most cases the angle calculated for the SHOPO (Φ_{SH}) was close to 90° from that calculated for the HOPO (mean 93°, standard deviation 26°). For the SLUPO the corresponding mean was 43° and standard deviation 14°.

A single point calculation was run on the compounds at the AM1-optimized geometries using the program Gaussian 98 [20]. Typically this required from less than one to four hours per compound. Bromo compounds were particularly time consuming. To obtain the FOPA descriptors, a molecular orbital calculation is run on the molecule with the ring of interest oriented normal to the z axis. Then the parameters Θ_H and Θ_L are least squares fitted to the calculated coefficients of the p_z orbitals, as described previously [3, 4]. This is done for the ten highest occupied and ten lowest unoccupied orbitals. A program for doing this calculation, run after Gaussian 98, is available on the Internet [21]. The level of calculation was the DF technique B3LYP [22, 23], which gives reasonable accuracy in the calculation of chemical properties, including much of the correlation energy, and the Pople basis set 6-31G* which is of intermediate quality, was used throughout. Hartree Fock calculations using 6-31G* [24] reproduce experimental bond lengths within 0.02 Å on average, across a wide range of compounds, including pathological cases involving bonds between two highly electronegative elements. This is greatly improved by consideration of electron correlation, so the compounds studied here should be very well described at the B3LYP/6-31G* level. The self consistent reaction field (SCRF) procedure COSMO [25] (conductive shielding model) with water as solvent, and the Besler-Merz-Kollman [26] method of fitting atomic charges to calculated electrostatic potential were also used. For the purposes of calculating the orbital angles, the method described previously [4] was employed, with a modification to accomodate the improved basis set: using a split valence basis set such as 6-31G*, there are two coefficients, not one, for each p_z orbital. These often, but by no means always, give approximately the same angle. For the purpose of these calculations, the two coefficients were summed (or equivalently, averaged). This is not rigorously justifiable, but the FOPA calculation is in any case based on regarding the molecule as approximated by benzene, for which the assumption of equal angles is valid by symmetry. The results of the molecular orbital calculations are given in Table 2, and Figure 1 shows contour plots of the HOPO and LUPO for the potent hallucinogen DOM, in both the flipped and unflipped orientation, 0.5 Å above the plane of the ring. The molecule and the angles calculated by the fitting procedure are superimposed. Note that the nodes in the orbitals do not pass precisely through the center of the ring as they do in benzene.

Classical parameters

The volumes of the substituents on C_2–C_6 of the benzene ring were used as steric parameters. These were calculated using the QSAR module of Hyperchem [17], and were obtained by subtracting the volume of benzene from the volume of benzene monosubstituted with the substituent, using the model builder geometries. Similarly, the lipophilicity of a substituent was obtained from the lipophilicity of benzene substituted with the substituent, calculated with the program ClogP [27], minus the lipophilicity of benzene itself. As before [1] the volume or lipophilicity of a ring system fused to the benzene ring was halved and the halves ascribed to both of the two points of attachment. The total lipophilicity was also included, which was not done in the earlier work [1]. The substituent lipophilicities and volumes used are given in Table 3.

The charge descriptor used was similar to that used previously [1], being $Q_{C1} - Q_{C2} + Q_{C3} - Q_{C4} + Q_{C5} - Q_{C6}$, where Q_{C1} for example is the charge on carbon atom 1 of the ring, and was calculated from the Gaussian output, using the electrostatic potential based charges as described by Besler, Merz and Kollman [26]. The components of the dipole moment, and the SHOPO, HOPO, LUPO and SLUPO energies were also obtained from the Gaussian calculation.

Flip regression

Flipping consists, for each compound, of swapping the substituent values for the two ortho substituents of a benzene ring, also the two meta substituents, and changing the sign of the orbital angles, if these are included. Other quantities, such as here the Y and Z components of the dipole moment, can also be negated if so desired. The programs that were written accept

Table 2. Values of descriptors from DF and ClogP calculations

Drug	Φ_H (°)	Φ_L (°)	E_{SH} (eV)	E_H (eV)	E_L (eV)	E_{SL} (eV)	μ_x (D)	μ_y (D)	μ_z (D)	Δ_Q (e)	log P
1	143.1	54.3	−6.581	−5.140	−0.058	0.337	−0.91	−0.11	−1.18	−0.11	2.67
2	146.7	54.7	−6.772	−5.452	−0.065	0.550	−0.21	1.48	0.79	0.30	1.60
3	141.6	54.3	−6.677	−5.597	−0.383	0.354	2.44	1.50	0.19	1.37	1.99
4	141.1	54.8	−6.723	−5.599	−0.353	0.356	2.37	1.55	0.20	0.66	1.91
5	142.8	54.4	−6.554	−5.379	0.042	0.582	−0.42	−1.56	1.27	−0.89	1.79
6	136.4	55.9	−6.563	−5.418	−0.012	0.571	−0.40	−1.40	1.43	0.02	2.32
7	135.2	56.8	−6.348	−5.666	−0.001	0.407	0.38	−2.07	−0.46	0.92	2.24
8	132.9	51.5	−6.285	−5.714	0.078	0.229	0.19	−1.94	−0.37	1.28	2.30
9	133.5	56.7	−6.234	−5.686	−0.001	0.414	−0.06	−1.73	−0.42	1.52	2.62
10	147.4	58.1	−6.652	−5.446	−0.329	1.055	0.97	−3.11	−1.04	0.95	2.46
11	146.9	51.7	−6.559	−5.371	−0.020	0.581	−0.25	−1.22	0.84	0.25	2.85
12	124.4	55.7	−6.086	−5.107	−0.174	0.496	0.21	−2.95	0.69	−0.28	1.86
13	120.6	57.1	−6.069	−5.086	−0.177	0.502	−0.11	−0.56	−1.55	−0.21	1.45
14	119.0	55.6	−6.079	−5.220	−0.232	0.379	0.20	−2.84	1.06	−0.38	3.23
15	122.2	56.1	−6.083	−5.113	−0.168	0.489	0.05	−3.01	0.69	0.11	2.39
16	118.6	55.6	−6.080	−5.179	−0.242	0.408	2.25	−0.29	−1.75	0.13	2.12
17	119.3	56.3	−6.087	−5.215	−0.219	0.391	0.26	−2.92	0.99	0.04	2.70
18	119.9	57.0	−6.071	−5.102	−0.157	0.494	−0.06	−2.95	0.71	−0.24	2.92
19	121.4	56.6	−6.070	−5.107	−0.166	0.485	−0.06	−2.91	0.79	−0.43	2.84
20	120.7	50.9	−6.054	−5.301	−0.382	0.302	1.26	−1.41	−0.81	0.28	3.10
21	129.9	54.8	−6.270	−5.247	−0.165	0.415	1.52	−1.37	0.64	0.24	2.12
22	39.3	88.5	−5.967	−5.757	−0.122	0.142	3.44	−2.17	3.04	0.45	2.37
23	143.6	33.5	−5.998	−5.758	−0.084	0.165	3.39	−0.62	2.84	0.42	1.84
24	32.0	56.8	−6.035	−5.778	−0.107	0.135	3.61	0.69	2.77	0.29	1.31
25	139.1	88.7	−5.982	−5.770	−0.130	0.137	3.71	2.12	2.96	0.31	1.84
26	91.2	43.7	−6.013	−5.889	−0.249	0.100	2.51	1.12	3.40	1.72	2.25
27	75.3	44.9	−5.867	−5.768	0.116	0.499	4.94	0.56	1.49	2.08	1.01
28	94.7	83.4	−6.892	−5.821	−0.066	0.311	−0.67	−0.04	0.94	−0.78	1.51
29	175.5	14.1	−6.179	−5.997	−0.431	0.392	3.93	−1.13	3.09	1.96	2.92
30	173.9	14.8	−6.176	−6.005	−0.449	0.390	4.10	−1.38	2.89	2.02	2.39
31	116.3	14.8	−5.997	−5.971	−0.415	0.397	4.16	−0.81	2.94	2.27	2.39
32	160.0	46.2	−6.061	−5.996	−0.267	0.292	5.06	−1.52	0.50	3.54	2.07
33	140.7	33.4	−5.978	−5.750	−0.075	0.175	3.30	−0.37	2.88	0.32	2.37
34	141.0	55.1	−6.632	−5.200	−0.118	0.311	−0.45	0.93	−1.86	−0.64	3.20
35	143.9	53.4	−6.631	−5.199	−0.018	0.300	−0.52	0.73	−1.70	−1.46	2.67
36	86.1	45.0	−5.949	−5.865	0.021	0.433	3.42	−2.02	1.51	1.46	1.30
37	124.8	56.3	−6.048	−5.066	−0.177	0.520	−0.41	−1.71	−1.64	0.34	2.70
38	120.5	56.8	−6.053	−5.174	−0.238	0.416	−0.18	−1.63	−1.35	0.18	3.01
39	123.2	57.0	−6.041	−5.054	−0.167	0.530	−0.54	−1.66	−1.65	−0.04	3.23
40	93.5	46.3	−5.989	−5.901	0.053	0.414	2.87	−1.25	2.90	1.53	1.23
41	136.7	52.9	−6.830	−5.644	−0.593	0.234	1.13	0.43	−0.32	2.45	1.78
42	137.4	51.9	−6.707	−5.452	−0.217	0.490	−1.67	−1.91	−0.40	−0.57	1.58
43	61.9	20.5	−6.726	−5.590	−0.096	0.166	−1.24	0.51	−0.20	−0.01	0.43
44	102.7	45.7	−5.947	−5.807	0.044	0.410	3.27	−1.94	1.32	1.44	1.15
45	100.0	46.0	−5.879	−5.840	0.075	0.551	2.51	−1.54	0.28	1.20	1.79

Table 2. Continued

Drug	Φ_H (°)	Φ_L (°)	E_{SH} (eV)	E_H (eV)	E_L (eV)	E_{SL} (eV)	μ_x (D)	μ_y (D)	μ_z (D)	Δ_Q (e)	log P
46	120.5	55.1	−6.691	−5.078	0.002	0.613	0.21	1.25	−0.06	−1.07	1.61
47	111.0	55.4	−5.948	−5.515	0.072	0.306	0.03	−2.49	−0.72	0.97	1.55
48	134.4	55.7	−6.682	−5.581	−0.401	0.309	1.89	1.21	0.62	1.54	2.30
49	142.1	56.3	−6.727	−5.578	−0.368	0.335	1.80	1.26	0.60	0.74	2.22
50	144.6	53.3	−6.712	−5.409	−0.154	0.506	1.58	0.17	−1.21	−0.03	1.90
51	139.6	56.6	−6.560	−5.399	−0.002	0.550	−0.62	1.47	1.44	0.19	2.63
52	145.5	53.6	−6.539	−5.335	0.000	0.598	−0.64	−1.49	1.17	−0.68	2.10
53	147.2	78.2	−7.356	−5.884	−0.078	0.334	6.19	−0.35	−1.64	0.22	1.33
54	135.6	57.5	−6.553	−5.380	0.012	0.558	−0.68	1.42	1.51	−0.09	3.16
55	97.8	46.5	−5.993	−5.909	0.049	0.414	2.97	−1.53	2.82	1.67	0.70
56	60.1	18.8	−6.745	−5.491	−0.098	0.202	−0.03	−0.19	−1.02	0.39	0.46
57	139.6	57.1	−6.429	−5.616	−0.011	0.450	−0.10	−0.77	−2.85	1.07	2.55
58	136.0	55.9	−6.331	−5.701	0.050	0.221	−0.45	0.38	−0.63	1.14	2.61
59	131.5	58.2	−6.307	−5.683	−0.043	0.399	−0.71	0.61	−0.69	1.32	2.93
60	127.1	58.9	−6.058	−5.115	−0.191	0.397	0.04	−1.20	0.01	−0.71	4.77
61	125.5	56.6	−6.082	−5.044	−0.155	0.517	−0.03	−1.56	0.19	−0.30	3.93
62	124.1	57.4	−6.074	−5.037	−0.150	0.519	−0.15	−1.51	0.19	−0.65	4.46
63	94.8	48.6	−5.939	−5.793	0.053	0.433	3.01	−0.94	1.40	1.74	1.01
64	55.6	14.2	−6.766	−5.499	−0.036	0.209	−0.95	−0.91	0.86	0.61	1.48
65	96.8	47.5	−5.947	−5.834	0.018	0.421	3.43	−0.73	1.17	1.44	1.15
66	54.0	12.8	−6.734	−5.534	−0.001	0.226	−0.67	−1.26	0.83	0.56	0.95
67	74.3	11.4	−6.616	−5.392	0.094	0.331	1.53	0.17	0.44	0.71	1.62
68	61.6	13.8	−8.206	−5.454	0.063	0.255	−0.32	−0.13	−0.28	0.93	2.49
69	46.3	19.4	−6.718	−5.488	−0.004	0.210	−0.20	1.47	0.83	0.49	1.35
70	97.7	45.4	−5.985	−5.914	0.053	0.420	3.08	−1.25	2.86	1.42	0.70
71	141.7	0.5	−6.425	−5.238	0.203	0.495	−0.92	3.19	1.14	−0.92	1.56
72	73.3	43.8	−6.001	−5.932	0.044	0.406	3.19	1.48	2.86	1.47	0.18
73	129.8	77.5	−6.606	−5.653	−0.138	0.074	−1.07	6.15	0.74	−2.73	2.30
74	59.3	18.3	−6.605	−5.371	0.105	0.335	1.57	0.29	0.28	0.87	1.62
75	120.1	44.4	−6.122	−5.466	0.154	0.291	0.65	2.42	0.84	1.68	0.93
76	134.2	87.2	−6.651	−5.212	0.011	0.404	−1.96	2.46	0.77	−0.59	1.47
77	81.5	88.1	−6.527	−5.580	0.024	0.150	−1.30	−1.32	−0.82	−0.08	1.47
78	75.8	44.2	−5.974	−5.890	0.071	0.433	2.88	1.48	2.83	1.27	1.23
79	141.0	55.2	−6.237	−5.296	−0.138	0.413	0.35	3.06	1.15	0.43	2.17
80	91.4	47.0	−5.933	−5.642	0.069	0.529	−3.55	−0.21	−1.74	−2.92	2.10
81	154.9	56.1	−6.493	−5.867	−0.149	0.228	1.40	−3.36	−0.62	1.09	−0.12
82	0.4	14.9	−6.225	−5.988	−0.430	0.410	3.93	1.34	2.98	1.70	3.45
83	5.0	15.6	−6.192	−6.005	−0.441	0.397	4.27	1.38	2.70	2.28	1.86
84	71.9	46.5	−5.998	−5.876	0.062	0.425	2.83	−1.44	2.93	1.66	0.49
85	139.5	2.6	−6.467	−5.248	0.172	0.489	−0.78	3.09	1.08	−1.05	1.03
86	157.8	47.9	−6.337	−5.706	0.034	0.402	−1.22	−3.14	−0.73	2.21	1.03
87	19.9	34.1	−6.545	−5.533	−0.040	0.000	−3.46	−3.21	−0.85	−0.22	1.03
88	80.4	51.5	−6.008	−5.563	0.258	0.479	−3.22	−3.03	1.30	−3.13	1.59
89	2.1	15.0	−6.231	−5.988	−0.429	0.409	4.02	1.41	2.82	1.68	2.92

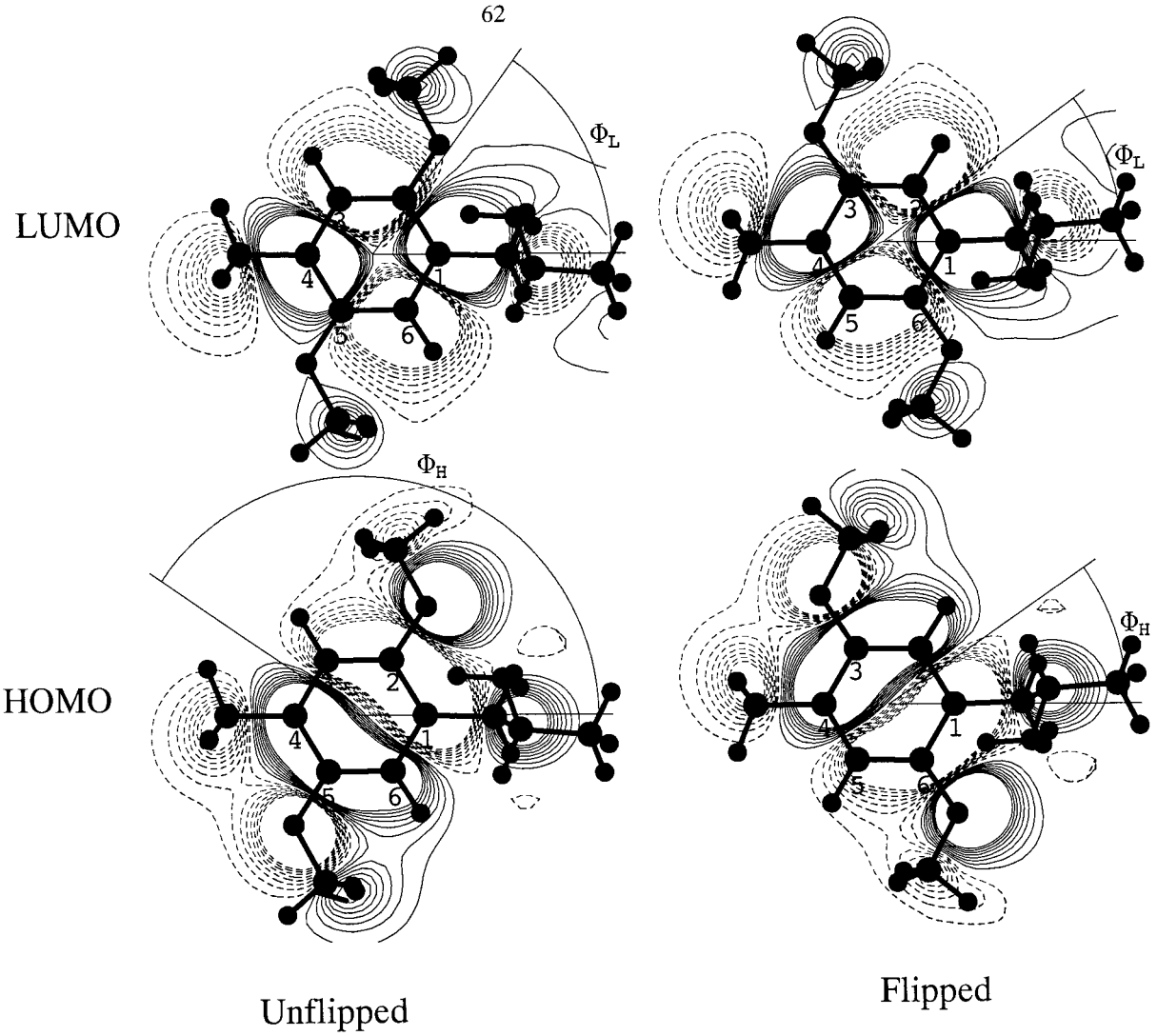

Figure 1. HOPO and LUPO of the hallucinogen DOM plotted 0.5 Å above the plane of the ring, in both standard and flipped orientation, showing the calculated nodal angles Φ_H and Φ_L, in relation to the nodes. Positive contours are solid, and negative contours broken.

the activity and descriptor data from a file, in which the angles, if present, are called PAH for Φ_H and PAL for Φ_L, and interactively accept flags identifying descriptors to be flipped, and on completion write out the flipped variables and the sin and cos of $2\Phi_H$ and $4\Phi_L$. In addition, two variables are appended: the flip status of and the statistical significance of the flip for each compound. The regression analysis is carried out for all of the variables in the file, and assumes that the last variable in the file is the dependent variable. The output file is then analyzed by other modules of the authors MARTHA suite of programs, which are available on the Internet [28].

Four programs were written. In each case, an array INVERT was constructed which initially contained unit values for each compound. If a compound was flipped, the value in the array was negated. The first program, FLIPALL, carried out multiple linear regression on every possible combination of flipped and unflipped compounds, and recorded the combination which gave the largest Fisher F ratio. In a trial of FLIPALL on a data set of 27 carbonic anhydrase inhibitors (to be published) with 15 dependent variables the runtime on a PC having a 700 MHz Pentium III processor, the runtime was 14 hours. This would double for each additional compound, so for the number of compounds

in the present study the use of this program could not be contemplated.

The second program, FLIPREG, started with the compounds as calculated, and unity in each of the elements of INVERT. The regression was carried out with each compound in turn, flipped and unflipped. If F improved on flipping, this was noted. Then all of those compounds for which flipping improved F were flipped. The entire procedure was then iterated until finally no compound flipped. This was then at least a local maximum for F. An alternative criterion to using F, that is calculating the residual for each compound both flipped and unflipped, and taking that which gave the minimum residual, was also implemented. The procedure did not always converge. When one of the criteria failed to converge, the other sometimes succeeded. Failure to converge consisted of the same small number of compounds alternating their flip status on each iteration. In this case the procedure was terminated at 100 iterations. The runtime of the hallucinogen data set, on the same PC as was mentioned above was approximately 2 seconds, and was dominated by screen I/O.

The third program, FLIPANNL, treated the problem as a global optimization in the elements of the INVERT array, with the F value as the objective function. Optimization was accomplished by carrying out 10000 iterations of simulated annealing using the Metropolis algorithm [29, 30]. In both FLIPREG and FLIPANNL the change in the residual for flipping each compound was calculated. This value divided by the standard deviation was treated as a standard Student's *t* value, and a statistical significance was calculated for the flip for each compound. Tests with synthetic data indicated that the three programs frequently gave the same result. When they did not, FLIPANNL always outperformed FLIPREG. FLIPREG was much faster than FLIPANNL, while FLIPALL was practical only for 20–25 compounds. The runtime of FLIPANNL on the hallucinogen data set was approximately 20 seconds on the 700 MHz Pentium PC, much longer than FLIPREG, but sometimes giving a much better F ratio. The runtime of FLIPALL on this data set would have been long even on cosmological time scales. It should be noted in comparing solutions that there are always two equivalent solutions to the problem, in which the INVERT column differs by a factor of −1.

Simulated annealing, as its name implies, is analogous to the actual annealing in a thermal process, where a low energy state is found by slowly lowering the temperature. In the context of maximization of Fishers F, a move consists of randomly changing the sign of one element of the INVERT array. If F increases, the move is always accepted. If F decreases, the move is accepted with probability (exp −ΔF/kT), where ΔF is the change in F, k is a constant and T is a parameter corresponding to temperature, which starts at a suitably high value and is progressively slowly lowered as moves are made. The starting position is chosen at random. The method seeks a global maximum of F among many local maxima. If T is lowered infinitely slowly the global maximum is guaranteed.

Table 3. Volumes and lipophilicities of substituents

Radical	π	Volume (Å^3)
Me	0.588	52.0
Et	1.008	97.6
Pr	1.578	151.0
Bu	2.238	204.6
OMe	−0.081	68.9
OEt	0.448	126.3
OPr	0.977	179.9
OBu	1.542	233.3
SMe	0.559	106.9
SEt	1.088	161.7
SPr	1.617	215.3
SBu	2.146	268.7
Cl	0.713	45.5
Br	0.863	64.1
NO_2	−0.257	65.9
C_2H_4F	1.301	106.8
$OCH_2C(CH_3)=CH_2$	0.892	213.3
O-i-Pr	0.757	162.2
S-s-Bu	1.926	243.4
−C3−	0.519*	58.2*
−C5−	0.663*	86.6*
OCH_2O	−0.322*	31.8*
OCH_2-c-Pr	0.892	193.0
OAllyl	0.493	164.5
OBz	1.687	291.2
SCH_2CH_2F	0.811	170.9
S-t-Bu	1.796	233.2
SCH_2-c-Pr	1.532	234.8
S-i-Pr	1.397	197.3
S-s-Bu	1.926	243.2
$SCH_2CH_2OCH_3$	0.142	244.5
Naphth	0.587*	62.4*

In practice, the solution found is usually close to the global maximum.

The fourth program, FLIP, flipped molecules in the data file either selected from the keyboard, or according to the INVERT column in an output file from one of the other three programs. It writes out the variables, including the angles, but not the significance column, in a format readable by the other programs. It is useful as a utility in both testing and using the other programs.

Table 4. Inflation of R^2 and Q^2 obtained using FLIPANNL on random data

	Unflipped		Flipped	
No. points	R^2	Q^2	R^2	Q^2
100	0.03	−0.08	0.39	0.33
50	0.04	−0.22	0.50	0.35
20	0.06	−0.60	0.55	0.29

QSAR calculations

For studying the QSAR of the hallucinogens a mixed Hansch-Free Wilson approach was employed. There is a wide variety of substituents on the benzene ring, so here a Hansch approach is warranted [31]. For the ethylamine side chain, the range of substituents is much more limited, so a Free Wilson methodology is more appropriate [32]. In the original study [1] an indicator variable was introduced to allow for the presence or absence of an α methyl group, and this was the only variation in the ethylamine side chain. In the FOPA study [3] this was continued in spite of the greater variability of the expanded data set. The present study will be more rigorous.

In the α position of the ethylamine moiety there is a 0, 1 or 2-carbon atom chain. This will be allowed for by a variable L_α, which will be 0 for no substituent, 1 for a methyl and 2 for an ethyl, and the square of this variable will also be used. For these three possibilities this is approximately equivalent to using two properties such as the volume or lipophilicity of the substituent. An indicator variable $I_{2\alpha}$ will have the value 1 if there is a second α methyl, and 0 otherwise. In several compounds there is a β methoxy group, and this is the only β substituent. The indicator variable I_β will have the value 1 if this group is present and 0 otherwise. In those compounds with a hydroxy bound to the amine nitrogen the indicator variable I_{HX} will take the value 1, and will be zero otherwise. In one compound only the ethylamine side chain is elaborated into a cyclopropane ring, and for this compound the indicator I_{cp} will be 1, and 0 for all others. Finally, the presence of an N-methyl or N-ethyl group will be accounted for by a variable L_N and its square, just as for the α methyl or ethyl.

For the benzene ring there are two groups of descriptors. As before [1] there are the classical parameters, the lipophilicities and volumes of the 2–6 position substituents, the HOPO and LUPO orbital energies, the charge descriptor, and the three components of the dipole moment. As in the second [3] study there are the FOPA parameters, the nodal angles Φ_H and Φ_L and the frontier orbital splitting energies Δ_H and Δ_L and the squares of the latter, and also the orbital energies themselves. Three batches of statistical calculations were run: the Free Wilson and classical parameters, the Free Wilson and FOPA parameters, and all of the parameters pooled. Each set of parameters was treated with the FLIPANNL program and the resulting file was analyzed with the Furnival and Wilson 'all possible subsets' regression technique, [33] using the BMDP 9R program [34]. Mallows C_p [35] was used as criterion of goodness of fit, with the penalty increased to 5–10 in order to restrict the number of variables entering the regression equations. Terms with a significance level poorer than 0.05 were removed, except that when either a sin or cos term of an angle was significant, both were included even if the second was not formally significant. This is because the pair of terms is a complete entity in itself, being convertible by trigonometric identities into the form $c\cos(\Phi - \delta)$, where the phase δ determines the optimal direction of the node.

Results

Testing of flip programs

Obviously, applying the flip regression procedure to any set of data will improve the correlations at least slightly, so it is necessary to explore the extent to which this happens with test data of known quality. A set of synthetic data was constructed by creating four independent variables X_1, X_2, X_3, and X_4, with 100 points, all uniformly distributed on the range 0–1. A dependent variable Y was constructed: $Y = 0.7X_1 + 1.3X_2 - 1.0X_3 + 1.5X_4$. This was the data set used in the following tests.

Table 5. Effect of error variance on inflation of R^2 and Q^2 on applying FLIPANNL

% Error	Unflipped		Flipped	
	R^2	Q^2	R^2	Q^2
100	0.03	−0.08	0.39	0.33
50	0.55	0.49	0.69	0.65
20	0.94	0.93	0.95	0.94

Replacing the dependent variable in the test data set with a set of normally distributed data with mean zero and standard deviation 0.7 (the same as that of the variable Y in the test data, and carrying out multiple regression on the entire data set, and on the first 20 and the first 50 points gave the results presented in Table 4. Here R^2 is the square of the multiple correlation coefficient, and Q^2 is the same based on the predicted residuals (the leave one out technique). The unflipped columns are on the raw data, and the flipped columns after swapping the data, X_1 with X_2, and X_3 with X_4. It may be seen that R^2 is very small and Q^2 negative before flipping, and both have become appreciable after swapping. R^2 becomes larger as the number of cases decreases, while Q^2 is approximately constant at 0.3. This illustrates the performance of the method in the absence of true correlation. To claim significance, one would need a Q^2 of better than 0.5, and an R^2 much better than this.

When a normally distributed random number with mean zero and standard deviation sufficient to make the error variance 50% and 20% of the total variance was added to Y, retaining all 100 points, and regressions carried out without and with swapping, the results in Table 5 were obtained. The row for 100% error from Table 4 is also included for comparison. At 50% error there is an appreciable improvement in fit on flipping, but at 20% error variance this has almost disappeared.

Using the program FLIP the last 50 points of the data set were swapped, variable X_1 with X_2 and X_3 with X_4. To separate copies of this data random, normally distributed error was added to Y, with mean zero and variance calculated to make the error variance 0%, 5%, 10%, 20% and 50% of the total variance. Subsets of each of these data sets were randomly selected containing 10, 20, 40, 70 and 100 points, making 25 subsets in all. All of these subsets were analyzed with FLIPANNL and FLIPREG, and the subsets with 10 and 20 points also with FLIPALL. The number of erroneous flips for each of these tests is presented in Tables 6 to 8. The results from FLIPANNL do not significantly differ from those from FLIPREG, and indicate that when 50% of the variance is error only with 70 or 100 points are the results at all satisfactory. With the error at 20% or less both methods gave acceptable results with all point numbers. For the 10 and 20 point samples FLIPALL significantly outperformed FLIPREG and FLIPANNL. With zero error all three methods retrieved the original data exactly.

For 100 points and 50% error variance, only 41 points had significant flips. Of these, 8 were wrong. With 20% error, 52 points were significant and none of these was in error. With 10% and 5% error, 70%

Table 6. FLIPANNL
Number of erroneous flips with 10–100 points and 0–50% error variance

Points	Error variance as percentage of total variance				
	0%	5%	10%	20%	50%
10	0	0	2	2	2
20	0	0	4	5	9
40	0	7	8	6	17
70	0	6	8	10	21
100	0	12	14	16	32

Table 7. FLIPREG
Number of erroneous flips with 10–100 points and 0–50% error variance

Points	Error variance as percentage of total variance				
	0%	5%	10%	20%	50%
10	0	2	4	3	3
20	0	0	4	7	9
40	0	7	8	6	17
70	0	6	8	10	19
100	0	12	14	16	32

Table 8. FLIPALL
Number of erroneous flips with 10–20 points and 0–50% error variance

Points	Error variance as percentage of total variance				
	0%	5%	10%	20%	50%
10	0	0	0	2	0
20	0	0	5	5	9

and 76% points were significant respectively, and only one significant flip was in error.

Flip regression and the FOPA results

Previously [3], an equation was developed relating the observed activities of the phenylalkylamines to the FOPA variables. An updated version of this, incorporating later experience, but not using the flip procedure, is the following:

$$\begin{aligned}\log A \;=\; & C_1 I_{Me} + C_2 \Delta_L + C_3 \cos 2\Phi_H \\ & + C_4 \sin 2\Phi_H + C_5 \cos 4\Phi_L \\ & + C_6 \sin 4\Phi_L + C_7 \end{aligned} \quad (1)$$

	1	2	3	4	5	6	7
C	0.337	1.303	−0.203	−0.389	−0.269	0.156	0.176
σ	0.079	0.264	0.087	0.083	0.069	0.099	0.114
α	0.00005	0.00000	0.02346	0.00001	0.00019	0.12083	0.12673

$N = 89$, $R^2 = 0.585$, $Q^2 = 0.517$, $F = 19.3$, $P = 6\times10^{-14}$, $s = 0.337$, $\Lambda = 1.56$

Here I_{Me} is 1 if there are 1 or 2 methyl groups on the α position, and 0 otherwise, Δ_L is the difference in energy between the SLUPO and LUPO, and Φ_H and Φ_L are the angles the nodes in the HOPO and LUPO respectively make with atom 1 of the benzene ring. The C_i are the regression coefficients of the descriptors, the σ and α are the standard errors and statistical significances of these coefficients respectively. F is the Fisher variance ratio, and P the probability (the statistical significance) based on this. s is the standard error of estimate, and Λ is an indicator of colinearity [36] in the equation. It is defined as $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i}$, where N is the number of variables and the λ_i are the eigenvalues of the correlation matrix of independent variables. Λ is 1 if the independent variables are orthogonal, and a value of greater than 5 indicates excessive colinearity in the equation.

When FLIPANNL is applied prior to the regression, still using AM1 derived angles and energies, the following equation results:

$$\begin{aligned}\log A \;=\; & C_1 I_{Me} + C_2 \Delta_H + C_3 \Delta_L \\ & + C_4 \cos 2\Phi_H + C_5 \sin 2\Phi_H \\ & + C_6 \cos 4\Phi_L + C_7 \sin 4\Phi_L + C_8 \end{aligned} \quad (2)$$

	1	2	3	4	5	6	7	8
C	0.411	0.367	0.957	−0.0693	0.556	−0.337	−0.158	−0.032
σ	0.059	0.086	0.205	0.0671	0.059	0.051	0.065	0.084
α	0.00000	0.00005	0.00001	0.30483	0.00000	0.00000	0.01710	0.70295

$N = 89$, $R^2 = 0.776$, $Q^2 = 0.727$, $F = 40.1$, $P = 8\times10^{-24}$, $s = 0.249$, $\Lambda = 1.56$

It will be seen that the correlation is greatly improved, both in fit and in a predictive sense. The descriptor Δ_H, previously not significant, is now very highly significant.

If instead of the AM1 parameters the DF energies and angles are used, but still using the I_{Me} indicator variable, the result is:

$$\begin{aligned}\log A \;=\; & C_1 I_{Me} + C_2 \Delta_H + C_3 \Delta_L \\ & + C_4 \cos 2\Phi_H + C_5 \sin 2\Phi_H \\ & + C_6 \cos 4\Phi_L + C_7 \sin 4\Phi_L + C_8 \end{aligned} \quad (3)$$

	1	2	3	4	5	6	7	8
C	0.487	0.869	0.118	0.0986	−0.311	−0.124	−0.199	0.071
σ	0.060	0.136	0.054	0.0491	0.046	0.041	0.061	0.088
α	0.00000	0.00000	0.03220	0.04794	0.00000	0.00342	0.00160	0.42561

$N = 89$, $R^2 = 0.804$, $Q^2 = 0.769$, $F = 47.5$, $P = 4\times10^{-26}$, $s = 0.233$, $\Lambda = 1.59$

Free Wilson and FOPA results

An all possible subsets regression of all of the Free Wilson and FOPA variables after application of FLIPANNL led to the elimination of the descriptors I_{cp}, $I_{2\alpha}$ and Δ_H^2, resulting in the equation:

$$\begin{aligned}\log A \;=\; & C_1 I_B + C_2 I_{HX} + C_3 L_\alpha \\ & + C_4 L_N + C_5 L_\alpha^2 + C_6 L_N^2 \\ & + C_7 \Delta_H + C_8 \Delta_L + C_9 \Delta_L^2 \\ & + C_{10} \cos 2\Phi_H + C_{11} \sin 2\Phi_H \\ & + C_{12} \cos 4\Phi_L + C_{13} \sin 4\Phi_L \\ & + C_{14} \end{aligned} \quad (4)$$

	1	2	3	4	5	6	7
C	−0.519	−0.333	0.535	−1.359	−0.309	0.607	0.372
σ	0.148	0.125	0.135	0.441	0.104	0.245	0.062
α	0.00077	0.00935	0.00017	0.00290	0.00411	0.01559	0.00000

	8	9	10	11	12	13	14
C	1.630	−0.727	0.0936	−0.295	−0.0683	−0.175	−0.029
σ	0.370	0.304	0.0502	0.046	0.0436	0.060	0.114
α	0.00003	0.01953	0.06650	0.00000	0.12098	0.00481	0.79822

$N = 89$, $R^2 = 0.824$, $Q^2 = 0.768$, $F = 27.0$, $P = 5\times10^{-23}$, $s = 0.230$, $\Lambda = 7.22$

Table 9 shows the benzene substitution pattern for each drug, and its flip status and significance. The equation appears to have a severe colinearity problem, but examination of the eigenvectors of the correlation matrix reveals that the colinearity involves the terms L_α, L_N, Δ_L and their squares. This colinearity, between descriptors and their squares, has no deleterious effects on the regression, and may be entirely removed

Table 9. Significance α and flip status I of each compound with the three variable choices

No.	Benzene substitution	Angle		Classical		All	
		I	α	I	α	I	α
1	5-Methoxy-4-methyl-2-methylthio	−1.	0.003	−1.	0.451	1.	0.823
2	2,5-Dimethoxy	−1.	0.004	−1.	0.000	−1.	0.224
3	2,5-Dimethoxy-4-bromo	−1.	0.002	−1.	0.000	−1.	0.409
4	2,5-Dimethoxy-4-chloro	−1.	0.002	−1.	0.000	−1.	0.367
5	2,5-Dimethoxy-4-methyl	−1.	0.002	1.	0.006	1.	0.444
6	2,5-Dimethoxy-4-ethyl	1.	0.001	1.	0.006	−1.	0.748
7	2,5-Dimethoxy-3,4-dimethyl	1.	0.001	−1.	0.001	1.	0.072
8	2,5-dimethoxy-3,4-C_3	1.	0.007	−1.	0.001	1.	0.117
9	2,5-Dimethoxy-3,4-norborna	1.	0.001	−1.	0.004	−1.	0.566
10	2,5-Dimethoxy-3,4-benzo	−1.	0.632	−1.	0.001	1.	0.760
11	2,5-Dimethoxy-4-(n)-propyl	1.	0.007	1.	0.001	−1.	0.325
12	2,5-Dimethoxy-4-methylthio	−1.	0.002	−1.	0.002	1.	0.008
13	2,5-Dimethoxy-4-(2-methoxyethylthio)	−1.	0.002	1.	0.000	−1.	0.183
14	2,5-Dimethoxy-4-(s)-butylthio	−1.	0.004	−1.	0.013	1.	0.033
15	2,5-Dimethoxy-4-ethylthio	1.	0.002	1.	0.003	−1.	0.009
16	2,5-Dimethoxy-4-(2-fluoroethylthio)	1.	0.004	1.	0.000	−1.	0.016
17	2,5-Dimethoxy-4-(i)-propylthio	1.	0.003	1.	0.011	−1.	0.026
18	2,5-Dimethoxy-4-(n)-propylthio	1.	0.002	1.	0.004	−1.	0.012
19	2,5-Dimethoxy-4-cyclopropylmethylthio	−1.	0.002	1.	0.007	−1.	0.024
20	2,5-Dimethoxy-4-(t)-butylthio	−1.	0.008	−1.	0.000	1.	0.006
21	2,5-Dimethoxy-4-(2-fluoroethylthio)	1.	0.002	1.	0.000	−1.	0.258
22	3-Ethylthio-4-ethoxy-5-methoxy	1.	0.039	1.	0.000	−1.	0.337
23	3-Methoxy-4-ethoxy-5-methylthio	1.	0.206	1.	0.011	−1.	0.416
24	3,4-Dimethoxy-5-methylthio	−1.	0.265	−1.	0.004	−1.	0.006
25	3,4-Dimethoxy-5-ethylthio	1.	0.019	1.	0.086	1.	0.001
26	3,5-Dimethoxy-4-benzyloxy	1.	0.987	1.	0.004	−1.	0.001
27	3,5-Dimethoxy-4-ethoxy	−1.	0.245	1.	0.124	−1.	0.218
28	4-Methoxy	−1.	0.374	1.	0.324	−1.	0.360
29	3-Ethoxy-4-ethylthio-5-methoxy	1.	0.555	1.	0.000	−1.	0.023
30	3,5-Dimethoxy-4-ethylthio	−1.	0.621	−1.	0.000	1.	0.529
31	3-Ethoxy-4-methylthio-5-methoxy	−1.	0.511	1.	0.000	1.	0.019
32	3,5-Dimethoxy-4-bromo	1.	0.163	1.	0.262	1.	0.953
33	3,4-Diethoxy-5-methylthio	−1.	0.194	1.	0.001	−1.	0.020
34	4-Ethyl-2-methoxy-5-methylthio	−1.	0.002	−1.	0.000	1.	0.102
35	2-Methoxy-4-methyl-5-methylthio	−1.	0.003	−1.	0.000	1.	0.144
36	3,5-Dimethoxy-4-allyloxy	−1.	0.741	−1.	0.017	−1.	0.704
37	2,5-Dimethoxy-4-ethylthio	1.	0.002	1.	0.000	−1.	0.001
38	2,5-Dimethoxy-4-(i)-propylthio	1.	0.003	1.	0.000	−1.	0.001
39	4-Propylthio-2,5-dimethoxy	1.	0.002	1.	0.000	−1.	0.001
40	3,4-Diethoxy-5-methoxy	−1.	0.672	1.	0.000	−1.	0.016
41	2,5-Dimethoxy-4-bromo	1.	0.045	−1.	0.002	1.	0.753
42	4-Methyl-2,5-dimethoxy	1.	0.047	1.	0.002	−1.	0.096
43	3,4-methylenedioxy	−1.	0.032	−1.	0.415	−1.	0.698
44	3,5-Dimethoxy-4-cyclopropylmethoxy	1.	0.262	−1.	0.017	−1.	0.970
45	3,5-Dimethoxy-4-methyl	1.	0.354	1.	0.349	−1.	0.577

Table 9. continued

No.	Benzene substitution	Angle		Classical		All	
		I	α	I	α	I	α
46	2,5-Dimethoxy-4-methyl	−1.	1.000	−1.	1.000	−1.	1.000
47	2,5-Dimethoxy-3,4-methylenedioxy	1.	0.013	1.	0.006	−1.	0.017
48	2,5-Dimethoxy-4-bromo	1.	0.002	1.	0.000	1.	0.264
49	2,5-Dimethoxy-4-chloro	1.	0.002	1.	0.000	1.	0.251
50	2,5-Dimethoxy-4-(2-fluoroethyl)	1.	0.004	1.	0.000	−1.	0.156
51	2,5-Dimethoxy-4-ethyl	1.	0.001	1.	0.000	1.	0.034
52	2,5-Dimethoxy-4-methyl	1.	0.003	1.	0.007	−1.	0.464
53	2,5-Dimethoxy-4-nitro	1.	0.003	1.	0.000	−1.	0.164
54	2,5-Dimethoxy-4-(n)-propyl	1.	0.001	1.	0.001	1.	0.041
55	3,5-Dimethoxy-4-ethoxy	1.	0.427	−1.	0.000	1.	0.091
56	3,4-methylenedioxy	−1.	0.324	−1.	0.884	1.	0.598
57	2,5-Dimethoxy-3,4-dimethyl	−1.	0.001	−1.	0.000	1.	0.010
58	2,5-Dimethoxy-3,4-trimethylene	1.	0.002	−1.	0.000	−1.	0.817
59	2,5-Dimethoxy-3,4-norbornane	1.	0.001	−1.	0.000	1.	0.389
60	2,5-Dimethoxy-4-(s)-butylthio	−1.	0.011	−1.	0.000	1.	0.258
61	2,5-Dimethoxy-4-ethylthio	1.	0.016	1.	0.001	−1.	0.213
62	2,5-Dimethoxy-4-(n)-propylthio	1.	0.015	1.	0.001	−1.	0.203
63	3,5-Dimethoxy-4-(i)-propoxy	1.	0.457	−1.	0.034	1.	0.361
64	3,4-methylenedioxy	1.	0.283	1.	0.457	−1.	0.835
65	3,5-Dimethoxy-4-methallyloxy	1.	0.418	−1.	0.096	1.	0.377
66	3,4-Methylenedioxy	1.	0.002	1.	0.005	1.	0.260
67	3,4-Methylenedioxy	1.	1.000	1.	1.000	−1.	1.000
68	3,4-Methylenedioxy	1.	0.773	1.	0.430	−1.	0.942
69	3,4-methylenedioxy	1.	1.000	−1.	1.000	1.	1.000
70	3-Ethoxy-4,5-dimethoxy	−1.	0.490	1.	0.000	−1.	0.015
71	2,5-Dimethoxy-4-ethoxy	−1.	0.029	1.	0.000	1.	0.014
72	3,4,5-Trimethoxy	1.	0.147	1.	0.007	−1.	0.003
73	2,4-Dimethoxy-5-bromo	−1.	0.001	−1.	0.002	−1.	0.000
74	3,4-methylenedioxy	−1.	0.281	1.	0.588	−1.	0.903
75	3-Methoxy-4,5-methylenedioxy	−1.	0.055	1.	0.615	−1.	0.001
76	2-Methoxy-4,5-methylenedioxy	−1.	0.014	1.	0.000	1.	0.036
77	2-Methoxy-3,4-methylenedioxy	−1.	0.640	1.	0.130	−1.	0.153
78	3,5-Dimethoxy-4-propoxy	−1.	0.223	−1.	0.008	1.	0.001
79	2,5-Dimethoxy-4-methylthio	1.	0.002	1.	0.000	1.	0.001
80	2,6-Dimethoxy-4-methyl	1.	0.770	1.	0.062	−1.	0.354
81	2,3,4,5-Tetramethoxy	−1.	0.007	1.	0.047	1.	0.081
82	3,5-Dimethoxy-4-butylthio	1.	0.318	1.	0.008	−1.	0.006
83	3,5-Dimethoxy-4-methylthio	−1.	0.143	−1.	0.008	1.	0.017
84	3,4,5-Trimethoxy	1.	0.219	1.	0.000	−1.	0.640
85	2,4,5-Trimethoxy	−1.	0.041	1.	0.000	1.	0.047
86	2,3,5-Trimethoxy	−1.	0.057	−1.	0.069	1.	0.027
87	2,3,6-Trimethoxy	−1.	0.044	1.	0.015	1.	0.033
88	2,4,6-Trimethoxy	−1.	0.875	−1.	0.016	−1.	0.285
89	3,5-Dimethoxy-4-propylthio	−1.	0.253	−1.	0.008	1.	0.009

by the substitution for these terms of their first and second degree orthogonal Forsythe polynomials [37, 38], giving:

$$\begin{aligned}\log A \;=\; & C_1 I_\beta + C_2 I_{HX} + C_3 P_1(L_\alpha) \\ & + C_4 P_2(L_N) + C_5 P_1'(L_\alpha) \\ & + C_6 P_2'(L_N) + C_7 \Delta_H \\ & + C_8 P_3(\Delta_L) + C_9 P_3'(\Delta_L) \\ & + C_{10} \cos 2\Phi_H + C_{11} \sin 2\Phi_H \\ & + C_{12} \cos 4\Phi_L + C_{13} \sin 4\Phi_L \\ & + C_{14} \end{aligned} \quad (5)$$

	1	*2*	*3*	*4*	*5*	*6*	*7*
C	−0.519	−0.333	0.156	−0.344	−0.309	0.607	0.372
σ	0.148	0.125	0.060	0.107	0.104	0.245	0.062
α	0.00077	0.00935	0.01140	0.00200	0.00411	0.01550	0.00000

	8	*9*	*10*	*11*	*12*	*13*	*14*
C	0.817	−0.727	0.0936	−0.295	−0.0683	−0.175	0.629
σ	0.133	0.305	0.0503	0.046	0.0436	0.060	0.057
α	0.00000	0.01952	0.06650	0.00000	0.12098	0.00481	0.00000

$N = 89$, $R^2 = 0.824$, $Q^2 = 0.768$, $F = 27.0$, $P = 5\times10^{-23}$, $s = 0.230$, $\Lambda = 1.54$

Here $P_1(x)$ and $P_1'(x)$ are the first and second degree Forsythe polynomials in L_α, and similarly for P_2 and P_3 and P_2' and P_3' in L_N and Δ_L. $P_1(x)$ for example is given by $x - 0.49438$ and $P_1'(x)$ by $x^2 - 1.2294x + 0.0685$, with the numerical coefficients derived as given in the reference. The statistics of this equation are identical to those of Equation 4, but the colinearity has been removed. Figure 2 shows the fit of the observed activities to this model.

The noteworthy features of Equations 4 and 5 are the Free Wilson terms that are significant. A β-methoxy group is deleterious to activity, reducing it by 70%. An N-hydroxy group is also deleterious, reducing activity by 55%. An α-methyl group enhances activity, but extending it to ethyl is detrimental, as shown by the linear and quadratic terms in the length of this chain. Methyl substitution on the amine nitrogen reduces activity, but much of this is regained on extension to ethyl. Adding a second α-methyl group, or elaborating the ethylamine side chain into cyclopropylamine, at least with the only two such compounds in this study, has little detectable effect.

A particularly interesting feature is the quadratic term in Δ_L. This was not apparent when AM1 was used. By this analysis there is an optimum value of this parameter for maximum activity, and this occurs at about 1.1 eV. Although with this choice of descriptors the similar term in Δ_H^2 was not statistically significant,

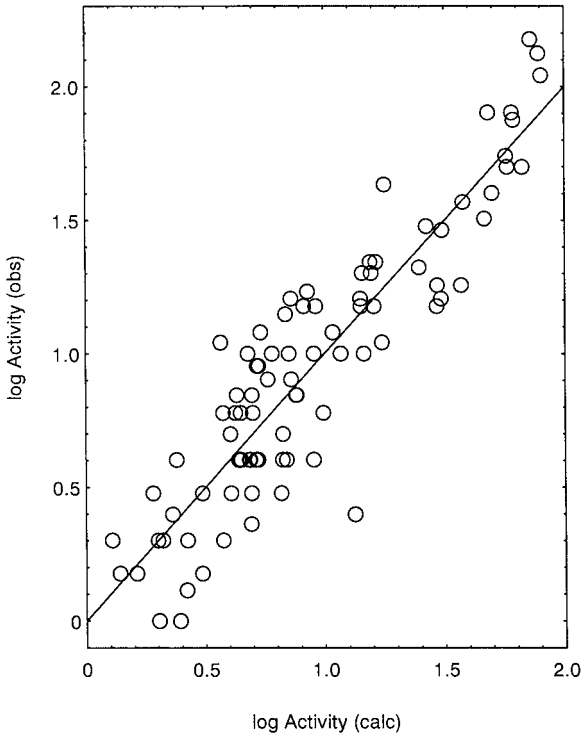

Figure 2. Plot of observed log activities of the hallucinogens against those calculated according to Equation 5.

others among the best 10 models did contain this term, and these gave a maximum activity at about values in the range 1.8–2.4 eV for Δ_H. Thus all of the FOPA parameters tried have been found influential in some significant models. From the trigonometric terms, the optimum value for Φ_H is 144°, and for Φ_L 63°.

Free Wilson and classical variables

Because there is only one 6-substituent in any of these compounds (OMe), there is a perfect linear relationship between V_6 and π_6. For this reason V_6 was removed from consideration. As V_2 is moved into V_6 in flipping, the latter must also be removed. Because a major finding of the first study [1] was the importance of squared terms in π_4 and V_4, and an interaction between the meta and para volume and lipophilicity terms, these quadratic terms, and also the products $\pi_3\pi_4$, $\pi_4\pi_5$,V_3V_4 and V_4V_5 were included in the statistics in the present study. On applying FLIPANNL followed by all possible subsets regression the best equation obtained was:

$$\begin{aligned}\log A \;=\; & C_1 I_{2\alpha} + C_2 I_{cp} + C_3 L_\alpha + C_4 L_\alpha^2 \\ & + C_5 \mu_z + C_6 \Delta_Q + C_7 \pi_2 \end{aligned}$$

$$+ C_8\pi_3 + C_9V_3 + C_{10}\pi_4 + C_{11}\pi_5 + C_{12}V_4^2 + C_{13} \quad (6)$$

	1	2	3	4	5	6	7
C	−0.684	0.931	0.557	−0.278	−0.106	0.202	−4.32
σ	0.193	0.192	0.099	0.073	0.013	0.024	0.86
α	0.00070	0.00001	0.00000	0.00028	0.00000	0.00000	0.00000

	8	9	10	11	12	13
C	0.868	-7.06×10^{-3}	0.560	0.609	-1.19×10^{-5}	0.781
σ	0.133	1.03×10^{-3}	0.046	0.086	1.6×10^{-6}	0.076
α	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

$N = 89$, $R^2 = 0.894$, $Q^2 < -1.0$, $F = 53.6$, $P = 5\times10^{-32}$, $s = 0.177$, $\Lambda = 3.26$

Table 9 gives the flip status and significance for each compound according to this equation, which does not include the interaction terms. These product terms explained the same variance as flipping does in this equation. The equation is in some regards superior to that obtained previously [1], having a larger number of cases and less of a colinearity problem. Its defects are the still appreciable colinearity and the very poor Q^2, which is entirely due to the Free Wilson term I_{cp}. This arises because only one compound has a nonzero value for this variable, so that when this compound is left out a very poor prediction for it is obtained. Even so, its colinearity can be improved as before by using Forsythe polynomials of L_α rather than a simple linear and quadratic term.

Correcting these deficiencies by dropping the I_{cp} term and using Forsythe polynomials rather than the linear and quadratic terms results in the equation:

$$\begin{aligned}\log A = {} & C_1\mu_z + C_2\Delta_Q + C_3I_{2\alpha} + C_4\pi_2 \\ & + C_5\Delta\pi_3 + C_6V_3 + C_7\pi_4 \\ & + C_8\pi_5 + C_9V_4^2 + C_{10}P_1(L_\alpha) \\ & + C_{11}P_1'(L_\alpha) + C_{12} \end{aligned} \quad (7)$$

	1	2	3	4	5	6
C	−0.0986	0.158	−4.98	0.630	-5.19×10^{-3}	0.606
σ	0.0161	0.026	0.97	0.152	1.15×10^{-3}	0.056
α	0.00000	0.00000	0.00000	0.00008	0.00002	0.00000

	7	8	9	10	11
C	0.548	-1.34×10^{-5}	0.181	−0.250	0.823
σ	0.100	1.3×10^{-6}	0.049	0.087	0.078
α	0.00000	0.00000	0.00044	0.00514	0.00000

$N = 89$, $R^2 = 0.845$, $Q^2 = 0.792$, $F = 42.6$, $P = 2\times10^{-27}$, $s = 0.211$, $\Lambda = 2.16$

Among the best 10 subsets of descriptors π_2-π_5 were prominent, usually with very high significance. Usually, π_4 was the most significant term. The π_2 coefficient was always negative and the others positive, indicating that a lipophilic *o*-substituent was detrimental to activity. The steric term V_3 occurred in all of them, always with a negative coefficient. The V_4 term never appeared, but its square frequently did, with a negative coefficient, indicating that bulk here was disadvantageous. While the coefficient of π_4 was always positive and very highly significant, its square when it appeared was negative and of moderate significance, implying that a lipophilic *p*-substituent favored activity only up to a limit.

Only two of the best 10 equations included a significant *m-p* interaction, this being V_4V_5 in both cases, with negative coefficient and moderate significance. In all of the best 10 equations the Z component of the dipole moment appeared with a negative sign. In two, the Y component appears with a positive sign. The charge term occurs in all 10, always with a positive sign and very high significance, implying that the more positively charged are the atoms C_1, C_3 and C_5, and the more negatively charged C_2, C_4 and C_6, the more active the drug. Notably, the Free Wilson parameters are much less prominent than was the case with the FOPA case.

Using all descriptors together

The best equation found was:

$$\begin{aligned}\log A = {} & C_1L_\alpha + C_2L_\alpha^2 + C_3\mu_x + C_4\mu_y \\ & + C_5\mu_z + C_6\pi_2 + C_7\pi_4 + C_8\pi_6 \\ & + C_9\cos2\Phi_H + C_{10}\sin2\Phi_H \\ & + C_{11}\cos4\Phi_L + C_{12}\sin4\Phi_L \\ & + C_{13}V_4V_5 + C_{14}\pi_4^2 + C_{15} \end{aligned} \quad (8)$$

	1	2	3	4	5	6	7	8
C	0.759	−0.417	0.0681	0.0572	0.0813	−1.13	0.764	−4.79
σ	0.125	0.094	0.0161	0.0144	0.0170	0.35	0.118	1.24
α	0.00000	0.00003	0.00006	0.00017	0.00001	0.00183	0.00000	0.00023

	9	10	11	12	13	14	15
C	0.181	−0.0853	−0.0376	−0.159	-3.44×10^{-5}	−0.261	0.513
σ	0.049	0.0533	0.0528	0.068	7.1×10^{-6}	0.064	0.073
α	0.00041	0.11362	0.47803	0.02137	0.00001	0.00010	0.00000

$N = 89$, $R^2 = 0.836$, $Q^2 = 0.706$, $F = 27.0$, $P = 2\times10^{-23}$, $s = 0.223$, $\Lambda = 4.03$

By the use of Forsythes polynomials in L_α and π_4, Λ can be reduced to 2.10, thus there is only a small detrimental colinearity in this equation, though substantially more than in Equation 5. Table 9 shows the flip status and significance for each compound. The terms entering the best 10 equations tended to be

those which had a strong tendency to enter the FOPA and classical equations, except that the components of the dipole moments were all always present, with a moderate statistical significance, and always a positive sign, and the charge difference term never appeared. The difference between E_H and E_L entered only two of the best 10 equations, and then with a low significance.

The remarks made about the π and V terms and their products and squares under the classical results apply here also. In none of the best 10 equations does Δ_L, or its square appear. Three of the equations contained Δ_H and its square, at low to moderate significance, with signs as in the FOPA section. The angular functions appeared in 8 of the 10, but with relatively low significance. The Free-Wilson parameters, as with the classical descriptors, are less prominent that with the FOPA results.

A mixed approach

To obtain the best practical result, two compounds were deleted (**46** and **69**) because each had a feature not present in any of the other compounds (the elaboration of the ethylamine side chain into cyclopropylamine, and the presence of a second α methyl). The inclusion of such atypical cases degrades the Q^2 value obtained. The Free-Wilson plus electronic set of variables was augmented by the inclusion of the variables π_4, V_4 and their squares, which were very significant in the classical treatment. Instead of Δ_H and Δ_L and their squares, the original four orbital energies and their squares were used. The FLIPANNL procedure was applied, followed by a backwards stepwise regression, resulting in the equation:

$$\begin{aligned}\log A \;=\; & C_1 I_\beta + C_2 L_\alpha + C_3 L_\alpha^2 + C_4 \pi_4 \\ & + C_5 \pi_4^2 + C_6 V_4 + C_7 V_4^2 \\ & + C_8 \cos 2\Phi_H + C_9 \sin 2\Phi_H + C_{10} E_H \\ & + C_{11} E_H^2 + C_{12} E_L + C_{13} E_{SL} \\ & + C_{14} E_{SL}^2 + C_{15} \qquad (9)\end{aligned}$$

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*
C 0.415	1.040	−0.517	0.481	−0.223	7.54×10^{-3}	-2.46×10^{-5}	0.110
σ 0.131	0.108	0.078	0.104	0.061	1.00×10^{-3}	5.57×10^{-6}	0.047
α 0.00232	0.00000	0.00000	0.00002	0.00050	0.00001	0.00003	0.02285

9	*10*	*11*	*12*	*13*	*14*	*15*
C 0.343	−21.1	−1.92	−1.048	1.730	−1.822	−57.75
σ 0.028	4.6	0.41	0.184	0.459	0.471	12.75
α 0.00000	0.00002	0.00002	0.00000	0.00034	0.00024	0.00002

$N = 87$, $R^2 = 0.878$, $Q^2 = 0.800$, $F = 37.0$, $P = 3\times10^{-27}$, $s = 0.192$, $\Lambda = 2.13$

The Λ value was obtained using the Forsythe polynomials to replace the linear and quadratic terms as before. This equation is of statistically very high quality except for a small amount of colinearity between π_4 and V_4. This may be removed by deleting the term of lesser significance, π_4 and its square, resulting in an equation with $R^2 = 0.841$, $Q^2 = 0.758$, $F = 32.7$, $P = 1\times10^{-24}$, $s = 0.215$, $\Lambda = 1.51$. Although there is appreciably less colinearity in this equation, equation 9 is probably preferable for most purposes, being superior in both fit and predictivity.

Discussion

The tests on flip regression indicate that it can recover flip status reliably if there is a sufficient number of points and not too much error variance. It produces quite small correlations when applied to random numbers, and these spurious improvements fall off rapidly when the random error decreases with respect to the genuine correlation. It is safe in that failure is indicated by a large fraction of the flips becoming insignificant, and when most flips are significant, those that are significant are nearly always correct.

The QSAR results above indicate that the flip procedure dramatically improves the correlations obtained. This is not, of course, the justification for using it: that is that it is a necessary logical consequence of the physical model. All three flip programs perform well, and the procedure contains a safety feature: when there is more error in the data than the model can support, a large fraction of the flips become statistically insignificant. The technique works best, as may be expected, when the error variance is small and the number of observations large.

Because, especially when the orbital energy terms rather than spitting energies, and when the squared terms are included, the number of fitted parameters is large, the technique is applicable only when a fairly large number of drugs are studied. Sets should number at least 25, and for preference many more. The larger the number of points the greater the number of parameters that can be included in the final equation, and the better the statistical quality of the resulting equations.

It is also apparent from the correlations that the DF procedure is superior to the semiempirical methods, as would be expected if the model were valid. It is to

be expected that the DF results are much more accurate, as the AM1 procedure neglects some large terms, which the DF treatment does not. The errors in the DF treatment are due to truncation of an approximation series (the basis set), and an approximation to the correlation energy, which AM1 ignores altogether.

Between the three sets of descriptors, a comparable amount of variance is explained using the FOPA, the classical variables, or both, given that the number of terms in the equation and the amount of colinearity are also varying. The classical variables are those traditionally used in medicinal chemistry, and correspond to the steric, electronic and hydrophobic parameters of Hansch. The FOPA parameters, however, result in a smaller number of explanatory variables and a lesser degree of colinearity. They also spring from a much more unified physical model, involving apart from the Free Wilson terms, only electronic terms: the energies of the π-like orbitals or the degree of splitting of the benzene frontier orbitals and the orientation of the resulting nodes. In the case of the phenylalkylamine hallucinogens these variables appear to be dominant, and are sufficient by themselves to explain most of the variance, but this cannot apply to nonaromatic drugs and is unlikely to be true of all aromatic classes of drugs.

The physical model on which the FOPA variables are based involves the interaction of π-like orbitals on the receptor with similar orbitals on the drug. The HOMO and LUMO of benzene each consist of a degenerate pair of orbitals with mutually perpendicular nodes in the case of the HOMO, and two pairs of mutually perpendicular nodes at 45° to each other in the case of the LUMO. In substituted benzenes the degeneracy is lifted, and usually two pairs of orbitals are found which differ in energy, but are recognizable as derived from the frontier orbitals of benzene, and frequently their nodal angles differ by close to 90° (HOPO, SHOPO) and 45° (LUPO, SLUPO). In the case of benzene, they differ by exactly this amount, but in DOB (**48**), for example, Φ_H is 134.4° and Φ_L is 55.7°, while Φ_{SH} is 46.0° and Φ_{SL} is 9.2°, where Φ_{SH} and Φ_{SL} are the angles calculated for the nodes in the second highest occupied and second lowest unoccupied π-like orbital. Thus the HOPO and SHOPO nodal angles thus differ by 88.4° and the LUPO and SLUPO by 46.5°. If one of the frontier orbitals is suitably oriented to interact with an orbital on a receptor, the other component of its pair is in the worst possible orientation. It may however be favorably oriented to interact with the other such orbital on the receptor, if this is of suitable energy.

When the squared terms in orbital energy were not included the correlation of activity with energy was alternatively negative and positive in the sequence E_{SH}, E_H, E_L E_{SL}. It was this that led to the apparent significance of the difference terms. Including the terms quadratic in orbital energy resulted in only a small improvement in fit.

While it appears from the hallucinogen data that it was the difference in energy between the pairs of orbitals that is significant, this observation has not generalized to other classes of drugs. Investigation of inhibitors of serine proteases and carbonic anhydrase [39] revealed that better fits and better predictivity could be obtained using the energies of the four π-like orbitals individually, rather than their differences, as predictors. This possibility was explored in developing Equation 9, and together with the use of the volume and lipophilicity parameters for the para substituent, which are the best of the classical parameters, led to the most statistically satisfactory QSAR of this study.

For an orbital on a drug to enter into an interaction with an orbital on a receptor, it is a requirement that the two must be of similar energy. The hallucinogens bind to a protein receptor the detailed structure of which is unknown. It seems likely however that the portion of the receptor that binds to the aromatic ring is a π-electron bearing aminoacid residue such as phenyl (phenylalanine), hydroxyphenyl (tyrosine), imidazolyl or imidazolium (histidine), or indolyl (tryptophan).

As well as depending on the splitting energy, (or the individual orbital energies), the orbital overlap becomes greatest when the nodes on the drug and receptor orbitals most nearly coincide. This will happen twice in a complete rotation in the case of the HOPO, and four times in the case of the LUPO, as the latter has two mutually perpendicular nodes. This is the reason for using the trigonometric functions of $2\Phi_H$ and $4\Phi_L$.

A difficulty in the flip regression procedure is the necessity, in the predictive use of the model, to assign a flip status to the new drug. Unless one is prepared to accept the prediction of one of two possible values this can only be done on the basis of a comparison with the flip status of the drugs of the training set, in this case as given in Table 9 for the three models discussed here. However, if 'activity' is an indicator of energy of binding, it makes sense to select that flip status which gives the highest predicted activity, as the compound will

bind in the orientation which minimizes the energy of the complex.

For all of the DOx drugs (**48**–**54**) the flip is highly significant and of the same sign for both the FOPA and classical cases. (For the mixed model the results are generally unsatisfactory, and most of the drugs have nonsignificant flip status, indicating that this case is not well handled by flip regression, probably because of the large number of variables.) It is presumed from this that all of the DOx drugs lie the same way on the receptor.

For those drugs for which the flip status is significant (e.g. $\alpha < 0.05$) it may be concluded that if the flip status is -1 the drug lies the opposite way on the receptor to the input orientation. When the flip status is not significant no conclusion can be drawn as to which orientation is preferred. Thus for **73**, the 2-methoxy actually takes the 6-position and the 5-bromo the 3 position, with all three regressions. For 6 of the 7 3,4-methylenedioxy compounds, the flip is not significant, suggesting that for this group there is little preference of orientation, using either the classical or FOPA descriptors.

Inspection of Table 9 shows that agreement between the FOPA and classical regression on which compounds are flipped is not particularly good (34 agreements, 14 disagreements). This is not surprising. The physical models on which they are based are quite different, so only one is likely to be valid. The FOPA results are probably to be preferred in the case of the hallucinogens, on the grounds of parsimony.

From the standardized values of the regression coefficients and the standardized values of the descriptors, a contribution to activity from each term in an equation can be calculated. Terms corresponding to the same or related descriptors can be collected. This is done in Table 10, which gives, for each drug, the contribution due to the combined Free-Wilson terms, Φ_H, Φ_L, Δ_H and Δ_L for Equation 4, or equivalently, Equation 5. The calculated logarithm of the activity and the residual are also given. The numbers indicate the contribution of the descriptor to the mean in units of standard deviations. A negative value contributes to the drug having less than average activity. Thus the most active drug, **48** (DOB) has positive contributions from all five sources. It has, along with **6**–**9**, **23**–**25**, **49**–**54**, **58** and **79** a near optimal contribution from its Φ_H value of 134° (**24**, with a value of 32°, flips). Its Δ_L value of 0.71 eV is also quite favorable, but not optimal.

Table 10. Contributions to activity from the Free-Wilson, angle, and energy splitting terms, the calculated log activities and residuals

Drug	F.W.	Φ_H	Φ_L	Δ_H	Δ_L	log A	Resid
1	0.37	−0.52	−0.19	0.47	−0.10	0.60	−0.35
2	0.37	−0.47	−0.20	0.39	0.29	0.40	−0.73
3	−0.08	−0.54	−0.19	0.21	0.45	1.20	0.34
4	−0.08	−0.54	−0.20	0.24	0.42	1.00	0.15
5	−0.08	−0.52	−0.19	0.28	0.17	0.90	0.14
6	−0.08	0.58	0.25	0.26	0.24	1.26	−0.31
7	−0.08	0.57	0.26	−0.08	−0.07	1.04	−0.20
8	−0.08	0.56	0.19	−0.17	−0.69	1.15	0.31
9	−0.08	0.56	0.26	−0.18	−0.06	1.34	0.15
10	−0.08	−0.46	−0.28	0.30	0.56	1.00	0.04
11	−0.08	0.60	0.19	0.29	0.27	1.57	−0.01
12	−0.08	−0.62	−0.22	0.13	0.37	0.60	−0.12
13	−0.08	−0.61	−0.26	0.14	0.38	0.95	0.24
14	−0.08	−0.61	−0.22	0.05	0.29	0.60	−0.05
15	−0.08	0.43	0.25	0.13	0.35	1.20	−0.28
16	−0.08	0.38	0.25	0.08	0.34	1.48	0.05
17	−0.08	0.39	0.26	0.06	0.28	1.32	−0.07
18	−0.08	0.40	0.26	0.13	0.34	1.18	−0.29
19	−0.08	−0.62	−0.24	0.12	0.34	0.85	0.15
20	−0.08	−0.61	−0.10	−0.03	0.38	0.60	−0.11
21	−0.08	0.53	0.24	0.17	0.24	1.46	−0.03
22	−0.08	−0.55	−0.18	−0.43	−0.40	0.30	0.19
23	−0.08	0.60	−0.24	−0.41	−0.43	0.60	−0.05
24	−0.08	0.59	−0.25	−0.40	−0.45	0.60	−0.04
25	−0.08	0.59	−0.19	−0.43	−0.39	0.60	−0.08
26	0.37	−0.17	0.02	−0.49	−0.20	0.48	−0.22
27	0.37	0.11	0.05	−0.51	−0.12	0.85	−0.04
28	0.37	−0.29	−0.36	0.20	−0.14	0.70	−0.13
29	−0.08	0.26	−0.45	−0.45	0.53	0.60	−0.24
30	−0.08	0.05	0.14	−0.46	0.54	1.08	0.04
31	−0.08	−0.59	0.14	−0.57	0.52	0.60	−0.05
32	0.37	0.50	0.08	−0.54	0.20	1.63	0.38
33	−0.08	−0.55	0.26	−0.42	−0.43	0.30	−0.02
34	0.37	−0.54	−0.21	0.47	−0.03	1.18	0.21
35	0.37	−0.51	−0.17	0.47	−0.27	0.85	−0.04
36	−0.08	−0.12	0.05	−0.52	−0.06	1.04	0.48
37	0.37	0.47	0.26	0.14	0.40	1.70	−0.06
38	0.37	0.41	0.26	0.06	0.35	1.51	−0.16
39	0.37	0.45	0.26	0.14	0.40	1.74	−0.02
40	−0.08	−0.27	0.02	−0.52	−0.17	0.11	−0.31
41	−1.11	0.58	0.21	0.29	0.54	1.30	0.11
42	−1.11	0.58	0.19	0.34	0.41	1.18	0.02
43	−1.11	0.37	0.24	0.25	−0.40	0.48	−0.13
44	−0.08	0.07	0.07	−0.48	−0.16	0.60	−0.04

Table 10. Continued

Drug	F.W.	Φ_H	Φ_L	Δ_H	Δ_L	log A	Resid
45	−0.08	0.01	0.07	−0.56	0.06	0.60	−0.09
46	−0.08	−0.61	−0.21	0.60	0.29	1.23	0.30
47	0.37	0.24	0.25	−0.27	−0.47	0.78	−0.22
48	0.37	0.57	0.25	0.22	0.42	2.18	0.32
49	0.37	0.60	0.26	0.26	0.41	2.12	0.23
50	0.37	0.60	0.22	0.37	0.35	2.04	0.14
51	0.37	0.59	0.26	0.27	0.19	1.88	0.09
52	0.37	0.60	0.22	0.30	0.27	1.70	−0.12
53	0.37	0.59	0.08	0.50	−0.06	1.90	0.22
54	0.37	0.57	0.27	0.28	0.18	1.90	0.12
55	−0.08	−0.04	0.09	−0.52	−0.16	0.78	0.20
56	−1.78	0.40	0.22	0.34	−0.31	0.40	0.04
57	0.37	−0.56	−0.26	0.01	0.03	1.08	0.34
58	0.37	0.58	0.25	−0.12	−0.64	1.30	0.14
59	0.37	0.54	0.27	−0.13	−0.00	1.26	−0.21
60	−0.74	−0.62	−0.30	0.11	0.25	0.48	0.20
61	−0.74	0.48	0.26	0.18	0.37	1.34	0.13
62	−0.74	0.46	0.27	0.18	0.37	1.18	−0.03
63	−0.08	−0.10	0.13	−0.48	−0.13	0.70	0.09
64	−0.41	−0.62	−0.45	0.35	−0.44	0.18	0.04
65	−0.08	−0.06	0.11	−0.50	−0.08	0.78	0.15
66	0.37	−0.62	−0.44	0.30	−0.49	0.48	−0.01
67	−0.21	−0.47	−0.43	0.31	−0.46	0.30	0.00
68	−0.29	−0.60	−0.45	1.44	−0.58	0.36	−0.33
69	0.37	−0.60	−0.45	0.32	−0.52	0.18	−0.31
70	−0.08	−0.35	0.04	−0.53	−0.16	0.00	−0.39
71	0.37	−0.54	−0.21	0.29	−0.33	0.95	0.23
72	−0.08	−0.49	0.02	−0.53	−0.17	0.00	−0.30
73	0.37	−0.62	−0.44	0.12	−0.53	0.60	0.22
74	−1.90	0.41	0.21	0.32	−0.48	0.18	−0.04
75	0.37	−0.61	0.06	−0.10	−0.73	0.30	−0.12
76	0.37	−0.60	−0.29	0.47	−0.10	0.90	0.04
77	0.37	−0.02	−0.27	0.11	−0.76	0.78	0.13
78	−0.08	0.10	0.07	−0.52	−0.17	0.85	0.21
79	0.37	0.60	0.24	0.11	0.19	1.60	−0.10
80	0.37	−0.17	0.10	−0.37	0.03	1.18	0.26
81	0.37	−0.34	−0.23	−0.12	−0.14	0.78	0.08
82	−0.08	0.16	−0.45	−0.41	0.55	0.48	−0.34
83	−0.08	0.27	0.16	−0.45	0.54	1.00	−0.16
84	0.37	−0.51	0.08	−0.50	−0.17	0.30	−0.27
85	0.37	−0.56	−0.16	0.31	−0.27	1.00	0.22
86	0.37	−0.29	−0.02	−0.12	−0.16	0.60	−0.22
87	0.37	0.50	0.25	0.16	−1.02	1.00	−0.07
88	0.37	0.00	−0.12	−0.26	−0.51	1.00	0.32
89	−0.08	0.21	0.15	−0.41	0.54	1.20	0.05

For the phenylalkylamine hallucinogens a large fraction of the variance of the activity can be explained using the electronic properties of the ring alone. This may not be the case with all series of drugs, and we have found other series, including carbonic anhydrase inhibitors [40] and trypsin inhibitors [41] in which these electronic terms play a role along with other factors. In these two reports we did not use the flip procedure, and the statistical significance of the results would be greatly improved by doing so. The technique can be applied to any series of drugs based on benzene, and it is probable that the principles involved, but not the numerical procedure used to calculate the angles, could be applied to any series of drugs requiring an aromatic group in the pharmacophore.

The use of flip regression, together with the FOPA parameters constitutes a consistent and unified approach to the QSAR analysis of drugs based on benzene. As presented here there is no selection of a small number of variables from a large pool, so the problem of chance correlation does not arise. Most of the variables go into the equation, and the significance based on the F value is reasonable. The physical model used is molecular orbital theory, and is not restricted to any set of substituent values, and illustrates better than any other QSAR so far reported the essential quantum mechanical nature of the interaction between an aromatic drug and its receptor.

The results of a QSAR that is based on the FOPA and flip techniques are as readily interpretable as the classical. Equation 9 for example indicates that a β-methoxy group on the average increases activity, that a 1-carbon on the α carbon, rather than 0 or 2 is favorable, that a large and hydrophobic, but not too large or too hydrophobic ($\pi_4 \sim 1.1$, $V_4 \sim 147$ Å^3) group is required at the 4 position, that an optimum value for Φ_H is 36°, and that a low E_L and about 5.5 eV for E_H and 0.47 eV for E_{SL} are desirable. Caution should of course be observed with any values that extrapolate beyond the range of the data.

It is easy to find substituents with suitable volume and lipophilicity values, but not so substituents that will give satisfactory orbital energies and nodal angles. Candidate compounds can however be screened fairly rapidly. The quantum theoretic calculations can be time consuming, especially if non-empirical techniques are used, but the flip procedure takes only seconds, using FLIPANNL, which is recommended. It is also apparent from Equations 1–9 that the variables that are important for high activity can be identified only approximately. Several variables usually com-

pete to explain the same variance, and in a data set that is not strictly orthogonal this is unavoidable with classical regression analysis. Techniques like principle component analysis and partial least squares can generate orthogonal data sets, but then do not eliminate irrelevant physical features. Although methods exist for variable selection in these techniques, the methods used here, classical regression with control of colinearity, are probably as satisfactory as any.

With the use of FLIPANNL a stochastic element has been introduced into the calculations. Successive calculations on the same data set do not give identical results. The differences tend to be in the statistical significance of the more marginal terms, and different sets are chosen after successive applications of the flip procedure. The overall quality of the fit however is relatively constant. In general though, the procedure should be tried several times and that giving the largest F value chosen.

To enable routine use of the methods developed herein it would be desirable to build the methods, both statistical and quantum theoretic, into a QSAR package, similar to the Molecular Descriptors package of Karelson [42]. Some extra constraints are necessary to achieve this, in that it is required to specify particular atom numbers to the quantum theoretic package in order to select the benzene ring of interest. As the procedure stands, it is also required to orient the aromatic center in a specific way. Together with the time required for the molecular orbital calculations this makes the FOPA method rather ill-adapted to high throughput procedures and more suitable for obtaining detailed insight into the requirements for a particular kind of pharmacological activity in sizable groups of aromatic drugs. The flip procedure on the other hand is intrinsic to the symmetry of benzene derivatives, and should always be used, except in the simplest cases (e.g., when only p-substituents are considered). MARTHA and GAUSSANG supplement rather than replace the standard packages.

Acknowledgements

I am indebted to Graham S. Chandler and Ian Bytheway, of the Department of Chemistry, the University of W.A. for several helpful discussions, and to Paula J. McLay, of the Australian Government Analytical Laboratories for a reading and criticism of the manuscript.

References

1. Clare, B.W., J. Med. Chem., 33 (1990) 687.
2. Shulgin, A., Shulgin, A., PIHKAL: A Chemical Love Story; Transform Press: Box 13675 Berkeley, California, 1991.
3. Clare, B.W., J. Med. Chem., 41 (1998) 3845.
4. a.) Clare, B.W., THEOCHEM 507 (2000) 157, b.) Clare, B.W., THEOCHEM 535 (2001) 301.
5. Pople, J.A., Beveridge, D.L., Approximate Molecular Orbital Theory; McGraw Hill: New York, 1970.
6. Dewar, M.J.S., Zoebisch, E.G., Healy, E.F., Stewart, J.J.P., J. Amer. Chem. Soc. 107 (1985) 3902.
7. Parr, R. G., Yang, W., Density-functional theory of atoms and molecules, Oxford Univ. Press: Oxford, 1989.
8. Parr, R.G., Pearson, R.G., J. Am. Chem. Soc., 105 (1983) 7512.
9. Mracec, M., Muresan, S., Mracec, M., Simon, Z., Naray-Szabo, G., Quant. Struct.-Act. Relat., 16 (1997) 459.
10. Mracec, M., Mracec, M., Kurunczi, L., Nusser, T., Simon, Z., Naray-Szabo, G., THEOCHEM, 367 (1996) 139.
11. Barfknecht, C.F., Nichols, D.E., J. Med. Chem., 18 (1975) 208.
12. Gupta, S.K., Bindal, M.C., Singh, P., Arzneim. Forsch., 32 (1982) 1223.
13. Domelsmith, L.N., Eaton, T.A., Houk, K.N., Anderson, G.M. III., Glennon, R.A., Shulgin, A.T., Castagnoli, N., Kollman, P.A., J. Med. Chem., 24 (1981) 1414.
14. Mracec, M., Mracec, M., Ann. West Univ. Timisoara, ser. chem., 6 (1997) 15.
15. Mracec, M., Mracec, M., Muresan, S., Bologa, C., Kurunczi, L., Simon, Z., Ann. West Univ. Timisoara, ser. chem., 5 (1996) 125.
16. *PCMODEL V. 5.0*, Available from Serena Software, Box 3076 Bloomington IN 47402-3076 U.S.A.
17. *Hyperchem v. 6.0*, Available from Hypercube, Inc. 1115 NW 4th Street, Gainesville, Florida 32601-4256 U.S.A.
18. Stewart, J.J.P., Q.C.P.E. Bull., 10 (1990) 86.
19. MOLDEN, written by G. Schaftenaar, CAOS/CAMM Center Nijmegen, Toernooiveld, Nijmegen, The Netherlands.
20. Gaussian 98, Revision A.9, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 1998.
21. File Gaussang.zip from the site http://www.chem.uwa.edu.au/research/bclare/
22. Becke, A.D., Phys. Rev., A38 (1988) 3098.
23. Lee, C., Yang, W., Parr, R.G., Phys. Rev. B37 (1988) 785.
24. Hehre, W.J., Radom, L., Schleyer, P. v.R., Pople, J.A., Ab Initio Molecular Orbital Theory, John Wiley & Sons: New York, 1986.
25. Klamt, A., Schüürman, G., J. Chem. Soc. Perkin Trans 2, (1993) 799.

26. Besler, B.H., Merz, K.M., Kollman, P.A., J. Comp. Chem., 11 (1990) 431.
27. *ClogP for Windows, ver 1.0.0.* Available from BioByte Corp. 201 West 4th St., Suite 204, Claremont CA 91711, U.S.A.
28. file martha.zip from the site http://www.chem.uwa.edu.au/research/bclare/
29. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A., Teller, E., J. Chem. Phys., 21 (1953) 1087.
30. Press, W.H., Teukolsky, S.A, Vetterling, W.T., Flannery, B.P., Numerical Recipes in Fortran, Second Edition; Cambridge: 1992, pp. 436–448.
31. Hansch, C., Maloney, P.P., Fujita, T., Muir, R., Nature (London), 194 (1962) 178.
32. Free, S.M., Wilson, J.W., J. Med. Chem., 7 (1964) 395.
33. Furnival, G. M., Wilson R.W., Technometrics, 16 (1974) 499.
34. *BMDP Dynamic, Release 7.* Available from SPSS Inc., Headquarters, 233 S. Wacker Drive, 11th Floor, Chicago. Illinois 60606.
35. Daniel, C., Wood, F. S., Fitting Equations to Data; Wiley: New York, 1971.
36. Chatterjee, S., Price, B., Regression Analysis by Example; 1st Ed, Wiley: New York, 1977; pp. 199–200.
37. Forsythe, G.E., J. Soc. Indust. Appl. Math., 5 (1957) 74.
38. Johnson, K.J., Numerical Methods in Chemistry. Dekker; NewYork: 1980, pp. 262–273.
39. Clare, B.W., Supuran, C.T., Unpublished observations.
40. Supuran, C.T., Clare, B.W., SAR and QSAR in Environmental Research, 12 (2001) 17.
41. Supuran, C.T., Scozzafava, A., Briganti, F., Clare, B.W., J. Med. Chem., 43 (2000) 1793.
42. Karelson, K. Molecular Descriptors in QSAR/QSPR, Wiley-Interscience, New York, 2000.